T.C. ANADOLU ÜNİVERSİTESİ YAYINI NO: 2806
AÇIKÖĞRETİM FAKÜLTESİ YAYINI NO: 1764

İSTATİSTİK-II

Yazarlar
Prof.Dr. Ahmet ÖZMEN (Ünite 1-3)
Prof.Dr. Emel ŞIKLAR (Ünite 4)
Prof.Dr. Hasan DURUCASU (Ünite 5)
Yrd.Doç.Dr. Mahmut ATLAS (Ünite 6, 7)
Yrd.Doç.Dr. Fikret ER (Ünite 8)

Editörler
Prof.Dr. Emel ŞIKLAR
Yrd.Doç.Dr. Ali ÖZDEMİR

ANADOLU ÜNİVERSİTESİ

Bu kitabın basım, yayım ve satış hakları Anadolu Üniversitesine aittir.
“Uzaktan Öğretim” tekniğine uygun olarak hazırlanan bu kitabın bütün hakları saklıdır.
İlgili kuruluştan izin almadan kitabın tümü ya da bölümleri mekanik, elektronik, fotokopi, manyetik kayıt veya başka şekillerde çoğaltılamaz, basılamaz ve dağıtılamaz.

Copyright © 2013 by Anadolu University
All rights reserved
No part of this book may be reproduced or stored in a retrieval system, or transmitted in any form or by any means mechanical, electronic, photocopy, magnetic tape or otherwise, without permission in writing from the University.

UZAKTAN ÖĞRETİM TASARIM BİRİMİ

Genel Koordinatör
Doç.Dr. Müjgan Bozkaya

Genel Koordinatör Yardımcısı
Arş.Gör.Dr. İrem Erdem Aydın

Öğretim Tasarımcıları
Doç.Dr. T. Volkan Yüzer
Öğr.Gör. Orkun Şen

Grafik Tasarım Yönetmenleri
Prof. Tevfik Fikret Uçar
Öğr.Gör. Cemalettin Yıldız
Öğr.Gör. Nilgün Salur

Dil Yazım Danışmanları
Hatice Çalışkan
Funda Gürbüz

Grafikerler
Gülşah Karabulut
Aysun Şavlı
Ayşegül Dibek
Hilal Küçükdağaşan

Kitap Koordinasyon Birimi
Uzm. Nermin Özgür

Kapak Düzeni
Prof. Tevfik Fikret Uçar
Öğr.Gör. Cemalettin Yıldız

Dizgi
Açıköğretim Fakültesi Dizgi Ekibi

İstatistik-II

ISBN
978-975-06-1471-2

1. Baskı

Bu kitap ANADOLU ÜNİVERSİTESİ Web-Ofset Tesislerinde 90.000 adet basılmıştır.
ESKİŞEHİR, Ocak 2013

İçindekiler

Önsöz

Anadolu Üniversitesi uzaktan eğitim uygulayan İşletme ve İktisat Fakültelerinde yürütülen İstatistik II dersi için hazırlanan bu kitap temel istatistiğin konularının ele alındığı sekiz üniteden oluşmuştur.

Kitabın uzaktan öğretim koşullarına göre hazırlanması temel amaç olarak benimsenmiştir. Ünitelerle ilgili amaçlar ve anahtar kavramlara yer verildikten sonra, metin içinde sıra sizde başlığı altında işlenen konuyla ilgili sorularla, verilen kavram ve tekniklerin daha iyi özümsenmesi amaçlanmıştır.

Ünitelerin sonunda kendimizi sınayalım başlığı altında öğrenciler, ilgili ünitede öğrendiklerinin küçük bir sınamasını yapa bileceklerdir.

Bu kitabın hazırlanması için emeği geçen çalışanlara editör ve yazarlar olarak teşekkür ederiz.

Editörler
Prof.Dr. Emel ŞIKLAR
Yrd.Doç.Dr. Ali ÖZDEMİR

1

Amaçlarımız

Bu üniteyi tamamladıktan sonra;

- Tamsayım ve örnekleme kavramlarını ayırt edebilecek,
- Örneklemeye başvurmanın yararlarını açıklayabilecek,
- Örnekleme planı hazırlayabilecek,
- Bir örnek araştırma için örnekleme uygulaması yapabilecek ve istenen bilgileri üretebilecek bilgi ve becerilere sahip olacaksınız.

Anahtar Kavramlar

- Tamsayım
- Çerçeve
- Parametre
- Değişken
- Örnekleme
- Örneklem
- İstatistik

İçindekiler

Örnekleme ve Örnekleme Dağılımları

TAMSAYIM VE ÖRNEKLEME

Anahtar kelimeler başlığı altında verilen kavramlar örnekleme konusunu açıklayabilmek için de bilinmesi gereken kavramlardır. Bu kavramlar İstatistik 1 kitabında açıklandığı için bu üniteye sadece tamsayım ve örnekleme kavramlarıyla ilgili hatırlatıcı bilgiler verilerek başlanmıştır.

DİKKAT

Bu ünitedeki konuları kolayca anlayabilmeniz için İstatistik 1 kitabında açıklanan temel kavramları tekrar okuyunuz.

Bilindiği gibi istatistiksel araştırmaların amacı tanımlanan **evren**in özellikleriyle ilgili bilgiler üretmektir. Bu bilgiler ya tamsayım sonucu elde edilen veri kümesinin (evren veri kümesinin) çözümlenmesiyle ya da örneklemden elde edilen veri kümesinin (örneklem veri kümesinin) çözümlenmesiyle üretilebilir.

Hakkında araştırma yapılacak birimler topluluğuna **evren** denir.

Tamsayım

Planlanan bir istatistiksel araştırma için tanımlanan sonlu evrenin bütün birimleri üzerinden araştırmaya konu olan değişkenler itibarıyla veri derleniyorsa yapılan işleme **tamsayım** denir. Tamsayım sonucu elde edilen veri kümesinin çözümlenmesiyle elde edilen bilgiler (parametre değerleri) veri derleme ve çözümleme hatası işlenmemiş ise kesin ve doğru bilgilerdir.

Tamsayım sonucu elde edilen veriler kullanılarak hesaplanan sayısal değerlere parametre denir.

Tamsayım genellikle sonlu ve küçük hacimli evrenlere uygulanır. Bununla birlikte büyük hacimli sonlu evrenlerin bütün birimlerine ulaşabilmek olanaklıysa karşılaşılan özel problemin çözümü için mümkün bütün verilerin elde edilmesine gereksinim varsa tamsayım yapılmalıdır.

ÖRNEK 1

Yeni bir ücret sisteminin uygulandığı 50 işçisi olan bir işletmede, işçilerin bu yeni ücret sisteminden memnuniyetleri araştırılmak istenmektedir. Tamsayım yapılabilir mi?

Çözüm:

Burada evren hacmi N=50 işçiden oluşmaktadır ve küçük hacimli bir evrendir. İşçilerin her birine ulaşmak ve onlardan memnuniyetleri konusunda veri elde etmek hem kolaydır hem de çok zaman almaz. Bu nedenle istenen araştırma için tamsayım yapılabilir.

ÖRNEK 2

Bir banka şube müdürü, genel müdürlükte izleyen gün yapılacak bir toplantı için, şubesinin, toplantı öncesindeki son iş günü itibarıyla mevduat durumuna ilişkin bilgiye ihtiyaç duymaktadır. Şubenin 150.000 mevduat müşterisi bulunmaktadır. İhtiyaç duyulan bilgi için tamsayım yapılabilir mi?

Çözüm:

Evren hacmi N=150.000 müşteridir. Büyük hacimli bir evrendir. Ancak, şubenin bilgisayar donanımına sahip olması ve mevduat hesaplarıyla ilgili verilerin veritabanında bulunması nedeniyle, evren hacmi büyük bile olsa çok kısa zamanda gerekli olan bilgilerin elde edilmesi, tamsayım yapılması mümkündür.

Ancak tanımlanan evrenin bütün birimleri üzerinden veri derlemek veya tamsayım yapmak her zaman izleyen kısımda açıklanacak çeşitli nedenlerle mümkün olamaz, parametre değerleri hesaplanamaz. Böyle bir durumda evrenin özellikleriyle ilgili bilgiler ve genellemeler örnekleme uygulamasıyla elde edilebilir.

Yapılmakta olan tamsayımların çoğu devlet tarafından planlanmakta ve icra edilmektedir. Bu tamsayımlar ülkenin nüfusundaki, ticaretindeki, sanayiindeki vb. bilinmek istenen gelişmelere ilişkin önemli bilgileri üretmek için yapılmak zorundadır. Genel nüfus sayımı, genel seçmen yazımı tamsayım için bilinen örneklerdir.

Örnekleme - Örneklem

Örneklem, evrenden belirli yöntemlerle seçilmiş olan ve seçildiği evreni temsil ettiği düşünülen birimler kümesidir.

Tanımlanan evrenden onu ilgilenilen değişkenler bakımından temsil eden sınırlı sayıda birimin belirli yöntemler kullanılarak seçilmesi işlemine örnekleme, seçilen birimlerin oluşturduğu topluluğa **örneklem** denir.

ÖRNEK 3

Bir anaokulu işletmecisinin 5 ayrı bölgedeki okullarında 1000 öğrencisi bulunmaktadır. Bu işletmeci öğrencilerine uyguladığı beslenme programıyla ilgili öğrenci ailelerinin görüşlerini almak amacıyla bir araştırma planlıyor.

Araştırmanın Amacı: Uygulanan beslenme programıyla ilgili ailelerin görüşlerinin alınması.

Araştırmanın Evreni: 5 bölgedeki okullarda okuyan öğrencilerin ailelerinin oluşturduğu topluluktur.

Örnekleme: Her bölgedeki okuldan 20'şer olmak üzere toplam n=100 aile seçiliyor. Ailelerin seçilmesi işlemine örnekleme; seçilen 100 ailenin oluşturduğu topluluğa örneklem adı verilir.

Değişken: Öğrencilerin annelerinin beslenme programıyla ilgili görüşleri değişkendir.

Gözlem Birimi-Örnekleme Birimi: Öğrencilerin anneleri gözlem birimi, aileler örnekleme birimidir.

Evreni temsil eden, onun bir modeli olan örneklemden elde edilen veri kümesi kullanılarak yapılacak çözümlemenin sonucu olan bilgi (örneklem istatistiği), evren bilgisi anlamında kullanılabilir. Başka bir deyişle, örneklem istatistiği, değeri bilinmeyen evren parametre değeri hakkında genelleme yapmak amacıyla kullanılabilir. Örneklemede önemli olan, eğer evren doğru, net tanımlanmış ise örneklemin araştırmaya konu olan değişkenler itibarıyla evreni temsil edip etmemesi konusudur. İyi, temsili bir örneklem evrenin sadece birim sayısı bakımından küçük, özellikleri bakımından benzeri ve modeli olan örneklemdir. Eğer ilgilenilen değişken bakımından evrendeki ve örneklemdeki birimler benzer dağılım gösteriyorsa oluşturulan örneklem temsili bir örneklemdir.

ÖRNEK 4

Bir fakültede kayıtlı olan ve hacmi N=2000 öğrenci olan bir evren tanımlayalım. Bunların %40 'ı erkek ve erkeklerin de %80 'inin yaşları 20-24 arasında değişiyor olsun. Diyelim ki n=200 öğrenciden oluşan bir örneklem seçilsin ve bu öğrencilerin %40 'ı (80 öğrenci) erkek ve bunların %80 'inin (64 öğrencinin) yaşları 20-24 arasında değişiyorsa bu örneklem cinsiyet ve yaş aralığı değişkenleri bakımından temsili örneklemdir.

SIRA SİZDE 1

- **Bir kitapçıya gelen müşterilerin kitap okuma alışkanlığı araştırılmak isteniyor. Tamsayım yapılabilir mi, tartışınız.**
- **Örneklem istatistiğinin değeri her zaman parametre değeri için bilgi niteliği taşır mı, tartışınız.**
- **Örneklemenin temel amacı nedir, açıklayınız.**

ÖRNEKLEME YAPMAYI GEREKLİ KILAN NEDENLER

KİTAP

Bu ünitenin izleyen kısımlarındaki konular için A. F. Yüzer, E. Şıklar, E. Ağaoğlu, H. Tatlıdil, A. Özmen, Editör: A. F. Yüzer (2011). İstatistik, Ünite 7, Eskişehir, Anadolu Üniversitesi Yayını İsimli kitaptan yararlanılmıştır. Editör ve ünite yazarının izni alınmıştır.

Evrenin sonsuz evren olması

Tanımlanan evrenin sonsuz evren olması durumunda tamsayım mümkün olmaz. Çünkü incelenecek birimler X rassal değişkeninin teorik olasılık dağılımının türettiği sonuçlardır. Dolayısıyla incelenecek birim sayısında, gözlem değeri sayısında da bir sınır yoktur. Bu nedenle örnekleme uygulamasına başvurmak kaçınılmaz.

ÖRNEK 5

Bir bisküvi fabrikasında üretilen paketlerin planlanan ağırlıkta üretilip üretilmediği, sistemin doğru çalışıp çalışmadığı araştırılmak isteniyor. Bu amaçla üretilen paketler arasından 250 paket seçiliyor. Tamsayım yapılıp yapılmayacağını açıklayınız.

Çözüm:

Bisküvi paketleri bir üretim sürecinin çıktıları niteliğindedir. Üretim sürdükçe bisküvi paketleri evrenine yeni paketler dahil olacaktır. Bu nedenle bisküvi paketleri evreni sonsuz bir evrendir. İlgili araştırma için örnekleme uygulaması zorunludur.

Evrenin sonlu evren olması

Daha önce de açıklanmış olduğu gibi, tanımlanan evren sonlu evren olduğunda evrenle ilgili bilgiler hem tamsayım uygulayarak hem de örneklemeye başvurularak elde edilebilir. Bu durumda tamsayım mı, örnekleme mi uygulanacağına karar verebilmek için aşağıdaki kriterler değerlendirilir.

KİTAP

Neter J., Wasserman W., Whitmore G. A. (1993). Applied Statistics 4. Edition, Allyn and Bacon

Maliyet: Örnekleme bütçesi, örneklemeyi tamsayıma tercih etmede en önemli belirleyicidir. Örnekleme tamsayıma göre daha az maliyetle bilgi üretme imkânı sağlar. Öte yandan eğer evren hacmi küçükse veya tamsayım yapmak bütçe olanaklarıyla da mümkünse tamsayım tercih edilmelidir. Burada dikkat edilmesi gereken nokta tamsayım yapma maliyetinin, elde edilecek bilginin değerinden küçük olması gerekir. Aksi durumda örneklemeye başvurmak uygun olacaktır.

ÖRNEK 6

Belirli bir bölge için planlanan bir siyasi araştırmada, partilerin bugünkü oy dağılımı hakkında bilgi edinmek amaçlanmış olsun. Tamsayıma mı örneklemeye mi başvurursunuz?

Çözüm:

Bu araştırmadaki evren ilgili bölgedeki seçmen sayısıdır, sonlu evrendir. İstenen bilgilerin üretilmesi amacıyla tamsayım yapmaya karar verilirse üstlenilmesi gereken maliyet bölgedeki genel seçim maliyetine eş değer olacaktır. Bölgede yapılmış genel seçim harcamalarına bakıldığında bunun araştırma yapacak kişi ya da kuruluş tarafından karşılanması oldukça zor görünebilir. Kaldı ki genel seçimlerde bütün seçmenler oy kullanmamaktadırlar. Tamsayım yapmanın imkânsız ve maliyetli olduğu bu gibi durumlarda en akılcı yol adı geçen araştırma için örneklemeye başvurmak olmalıdır.

Zaman: Bir araştırma sonunda ulaşılacak bilgiye duyulan ihtiyacın zaman sınırları, araştırmanın tamsayımla mı yoksa örneklemeyle mi yapılacağına karar verirken değerlendirilecek diğer önemli bir etkendir. Örnekleme, tamsayıma göre daha kısa zamanda ve yeterli ayrıntıda bilgi elde etme olanağı verir. Örneklemenin bu özelliği ve üstünlüğü bilgiye çok hızlı gereksinim olduğu durumlarda bilhassa önemlidir.

Hem bir tamsayımdan hem de bir örneklemden elde edilecek bilgi için gerekli olan zaman bir alternatif maliyet üstlenmeyi de gerektirir. Çünkü bilgi elde etme süresine bağlı olarak verilecek kararın erken ya da geç oluşu kazanç kadar kayıplara da neden olabilir.

Örneğin; seçmenlerin oy verme tercihleri üzerinde pek çok faktörün etkisi bulunmaktadır. Partileri, bugünkü oy dağılımını belirlemeye yönelik bir araştırma uzun bir zamana yayılırsa, seçmenlerin araştırmaya başlandığı günkü görüşleriyle araştırmanın sonuçlandığı zamandaki görüşleri arasında önemli farklılıklar oluşabileceğinden üretilen bilginin değeri ve geçerliliği giderek azalacaktır. Bu nedenle örneklemeye başvurmak önem arz eder.

Doğru veri elde etme: Her ne kadar tamsayım yapılınca kesin, doğru bilgiye ulaşılır denilse de tamsayımın yapılabilmesi için gerekli olan sayıda veri derleme aracı ve istenen özelliklere sahip, veri derleme hatası yapmayacak gözlemci ya da görüşmeci bulmak ya da yetiştirmek oldukça zor hatta olanaksızdır. Bu nedenlerle örnekleme uygulamaları tamsayıma göre daha doğru veri derleme ve daha doğru bilgi üretme imkânı verir.

İncelenecek birimlerin fiziksel zarara uğraması: Tanımlanan evrende yer alan birimler, veri derlemek ya da ölçüm yapmak amacıyla fiziksel zarara uğratılıyorsa örneklemeye başvurmak zorunludur. Örneğin bir savunma sanayi kuruluşunda belirlenen bir gün içinde üretilen mermilerin içerisindeki hatalı mermi oranının belirlenmesi için yapılacak bir araştırmada tamsayım benimsenirse gerekli verilerin derlenmesi amacıyla üretilen tüm mermilerin patlatma testine tabi tutulması gerekir. Bu durumda üretimin amacı gerçekleşmemiş olur. Bu anlamsız bir testtir. Bu gibi durumlarda örneklemeden yararlanmak kaçınılmaz olur.

Evreni oluşturan birimlerin değişkenliği: Evreni oluşturan birimler araştırmaya konu olan değişkenler bakımından heterojen olduğunda mümkün ise tamsayım yapmak, değil ise büyük hacimli örneklem seçmek gerekir.

SIRA SİZDE 2

• Evren hacmi küçük, parasal imkânların yeterli olduğu bir araştırma için tamsayım mı yoksa örnekleme mi tercih edersiniz, açıklayınız.
• 1,5 milyon öğrencinin olduğu Açıköğretim Sisteminin değerlendirileceği bir araştırma için gerekli olan zaman ve parasal imkânlar yeterli ise tamsayım mı yoksa örnekleme mi uygularsınız, tartışınız.

ÖRNEKLEM İÇİN BİRİM SEÇME YÖNTEMLERİ

Örnekleme girecek birimlerin seçiminde kullanılan yöntemler keyfî seçim yöntemi ve rassal seçim yöntemi şeklinde sınıflandırılmaktadır.

Keyfî Seçim

Örneklem oluşturulurken tanımlanan evreni oluşturan birimler arasında fark gözetilir, yani bütün birimlere bilinen bir olasılıkla seçilme şansı verilmez ise bu türden birim seçimine keyfî seçim adı verilir. Bu seçim yönteminde araştırmacı, hangi birimlerin örnekleme seçileceğini bilerek ve isteyerek belirler. Örneğin yaşadığınız ildeki öğretmenlerin sorunlarını belirlemek amacıyla yapılacak bir araştırma için her bir öğretmen evi lokalinde birlikte oturduğunuz öğretmenler arasından tanıdığınız öğretmenleri seçmek suretiyle bir örneklem oluşturursanız yapmış olduğunuz seçim keyfî seçimdir.

Rassal Seçim

Sonlu Evrenlerde Rassal Örneklem Seçimi

Sonlu evrenlerde rassal birim seçim imkânı veren iki seçim uygulaması bulunmaktadır. Bunlar kura seçimi ve sistematik seçimdir.

KİTAP

Serper Ö. (2004). Uygulamalı İstatistik 2, Genişletilmiş 5. Baskı, Bursa: Ezgi Kitabevi

Kura Seçimi

Rassal birim seçimi için kura usulü uygulanacak ise aşağıdaki adımlar izlenir:

- Tanımlanan evrenle ilgili oluşturulacak güncel çerçevedeki bütün birimlere birden N'e kadar numara verilir. Bu numaralar fişlere yazılır ve bir torbaya veya bir kaba atılır.
- Fişler iyice karıştırıldıktan sonra n tane fişin çekilmesi işlemine başlanır. Çekilen fiş her çekilişten sonra torbaya iade edilir veya edilmez. Çekilen fiş torbaya iade ediliyorsa birim seçimine iadeli seçim, iade edilmiyorsa iadesiz seçim adı verilir.
- Seçilen n sayıdaki birim örneklemi oluşturur.

Bu birim seçim uygulamasıyla evreni oluşturan çerçevede yer alan birimler arasında örneklemde yer almaları bakımından ayrıcalık yapılmamış ve eşit seçilme olasılığı tanınmış olur.

Birim seçimi iadeli yapıldığında aynı birim tekrar tekrar örnekleme seçilmiş olabilir. Bu durumda örnekleme kuramının önemli koşullarından biri olan bağımsızlık koşulu sağlanmış olur ve herhangi bir birimin seçilmesi bir başka birimin seçilmesi olasılığını etkilememiş olur.

DİKKAT

Birim seçimi iadesiz yapıldığında bağımsızlık koşulu sağlanmamış olur.

Gerçekte uygulanan örnekleme planlarında iadeli seçim genellikle uygulanmaz. Birim seçimi iadesiz yapıldığında, seçilen bir birimin tekrar seçilmesi engellenmiş olur. Ancak, evren hacmi N çok büyük, örneklem hacmi n küçük olduğunda her birimin seçiminin diğerinden bağımsız olduğu ve iadeli seçimdeki bağımsızlık koşulunun sağlanabilecek olduğu varsayılır.

Belirlenen sonlu evrende yer alması gereken birim oluşturulacak çerçevede yer almıyorsa veya birden fazla yer alıyorsa kura seçiminin rassallığı etkilenir. N hacimli sonlu bir evrende rassal iadeli seçimle Nn tane farklı örneklem seçmek mümkün iken aynı evrende aynı hacimli iadesiz seçim uygulandığında $C_N^n = \frac{N!}{(N-n)!.n!}$ tane farklı örneklem seçmek mümkün olur.

Sistematik Seçim

- Güncel çerçevedeki birimler birden N'ye kadar numaralandırılır.
- Örneklem hacmi belirlenir.
- k = N / n oranı hesaplanır. Bu oran "büyütme faktörü" olarak isimlendirilir.
- 1, 2,....., k adet sayı arasından rassal olarak bir sayı çekilir. Çekilen sayı a ile gösterilsin. a, örnekleme girecek birinci birimin sıra numarasını gösterir.
- a'ıncı, a + k'ıncı,....., a + (n - 1)k'ıncı sıra nolu birimlerin seçilmesiyle n hacimli örneklem oluşturulur.

Hem kura usulü seçimde hem de sistematik seçimde seçilecek bir birimin belirlenen n hacimli örneklemde yer alması olasılıkları aynı (n / N) olmasına rağmen olası örneklemlerden birinin incelenen örneklem olma olasılıkları farklıdır. Bu olasılıklar kura usulü (iadesiz) seçimde 1 / C_N^n olduğu hâlde, sistematik seçimde örneklemi oluşturabilme şansına sahip kombinasyonların her biri için eşit 1 / k, diğerlerininkinde ise 0 (sıfır) dır.

Olasılıklı örneklemenin üç önemli üstünlüğü vardır:

- Örneklemden elde edilen verilerden hesaplanan istatistikler evren parametreleri hakkında genelleme yapmak üzere kullanılabilir.
- Örneklem hatasının büyüklüğü hakkında bilgi elde edilebilir.
- Keyfî seçimde söz konusu olabilecek yanlılık (sistematik hata) giderilmiş olur.

Olasılıklı örneklem oluşturma prensibi esas olmak üzere, uygulamada ya birim seçim işlemini kolaylaştırmak ya da evreni temsil edecek daha iyi bir örneklemin oluşturulmasını sağlamak üzere çeşitli rassal örnekleme yöntemleri geliştirilmiştir. Bu yöntemler izleyen başlıklar altında açıklanmıştır.

Bu örnekleme yöntemlerinden herhangi birini bir örnekleme uygulaması için seçerken yöntemlerin etkinlik ve doğruluk kriterlerine göre değerlendirilmesi gerekir. Örnekleme yöntemlerinin etkinlikleri farklılık gösterir. Etkinlik, örnekleme maliyeti ve doğruluğu arasındaki dengeyi yansıtan bir kavramdır. Doğruluk ise ölçülecek özelliğin belirsizliği ile ilgili düzeyi gösterir. Doğruluk ile örnekleme hataları arasında ters ilişki varken maliyetle aynı yönde ilişki vardır. Yani daha çok maliyet daha doğru bilgi, daha doğru bilgi daha az hatalı karar ve tahmin demektir.

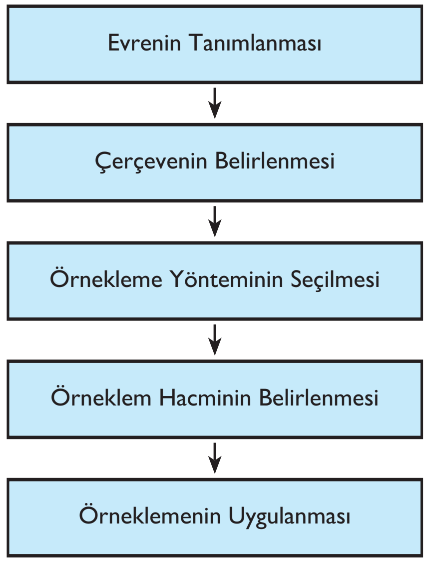
ÖRNEKLEME SÜRECİNİN AŞAMALARI

Genel olarak örnekleme süreci 5 aşamadan oluşmaktadır. Birbirleriyle ve bir araştırma sürecinin diğer aşamalarıyla sıkı sıkıya ilişki içinde olan bu aşamalar Şekil 1.1' de gösterilmiştir.

Evrenin Tanımlanması

Örnekleme süreci evrenin tanımlanmasıyla başlar ve bir araştırma sürecinde araştırmacının ilk yapacağı işlerden biridir.

Evren, araştırmacı tarafından belirlenen bir tanıma uyan ve hakkında bilgilerin üretileceği, çıkarımların yapılacağı birimlerden oluşan topluluktur. Evrenin ayrıntılı bir biçimde tanımlanmasıyla hangi birimlerin araştırma kapsamına alınacağı, hangilerinin alınmayacağı belirlenmiş olur. Bu, başka bir ifadeyle örneklemde yer alabilecek ve yer alamayacak birimlerin belirlenmesi anlamına gelir.



Şekil 1.1

Örnekleme sürecinin aşamaları.

Kaynak: *N. K. Malhotra, Marketing Research An Applied Orientation, Prentice Hall International Inc., 1996*

Evrenin tanımlanması genel olarak örnekleme birimi, gözlem birimi, yer ve zaman kavramlarıyla yapılmaktadır. Araştırmanın konusunu tanımlayan değişkenlerle ilgili verilerin derlendiği birimlere gözlem birimi adı verilir. Örnekleme birimi ise örnekleme seçilecek birimlerdir.

ÖRNEK 7

Araştırmanın konusu: Eskişehir Merkez İlçesinde ikamet eden 250.000 ailede hangi marka bulaşık deterjanı kullanıldığını araştırmak.

Evren: Eskişehir Merkez İlçesinde araştırma yapıldığı tarihte ikamet etmekte olan ailelerin oluşturduğu topluluktur. Sonlu bir evrendir. Evren hacmi 250.000 ailedir.

Örnekleme Birimi: Her bir aile.

Gözlem Birimi: Ailede çamaşır yıkama görevini üstlenen veya deterjan satın alma tercihinde bulunan kişidir.

Yer: Eskişehir Merkez İlçesi

Zaman: Araştırmanın yapıldığı tarih.

Örnekten anlaşılabileceği gibi önce örnekleme aile seçilecek sonra ailedeki çamaşır yıkayan, deterjan tercihini yapan kişiden veri derlenecektir.

Bu örnekte olduğu gibi her araştırmada gözlem birimi ve örnekleme birimi ayrımı olmayabilir. Örnekleme seçilen ve veri derlenen birim aynı olabilir. Bu durumda sadece birim kavramı kullanılır.

ÖRNEK 8

Araştırmanın Konusu: 2010-2011 öğretim yılında Anadolu Üniversitesi İktisat Fakültesi Kamu Yönetimi Bölümünde öğrenim gören öğrencilerin kitap okuma alışkanlığını araştırmak.

Evren: 2010-2011 öğretim yılında Kamu Yönetimi bölümünde kayıtlı olan öğrenciler topluluğudur.

Birim: Kamu Yönetimi bölümünde kayıtlı olan her bir öğrenci hem gözlem birimi hem de örnekleme birimidir. Çünkü örnekleme seçilecek birim de veri derlenecek birim de öğrencilerdir.

DİKKAT

Örnekleme ve gözlem birimi aynı olduğu zaman araştırmalarda sadece birim kavramı kullanılmaktadır.

Gözlem birimi ve örnekleme birimi ayrımına gidilmesinin nedeni gözlem birimleri ile ilgili bir çerçevenin temin edilmesinin veya hazırlanmasının zor, maliyetli ve çok zaman alacak olmasıdır. Örnekleme birimi birden çok gözlem birimini kapsayacak şekilde de tanımlanabilir. Örnek 3 üzerinden açıklamak gerekirse beslenme programıyla ilgili öğrencinin hem annesinin hem de babasının görüşlerine başvurulabilir. Bu durumda gözlem birimi öğrencinin hem annesi hem de babası olur.

Evrenin tanımlanması yukarıda yapılan açıklamalarda olduğu gibi her zaman kolay olmayabilir. Örnek 8 üzerinden açıklama yapılacak olursa evren tanımı yapılırken örneğin öğrenci evleri, yabancı uyruklu olup Eskişehir Merkez İlçesinde ikamet edenler, tek kişilik yaşamın yaşandığı konutlar, araştırmada ifade edilen aile tanımı içine alınacak mı yoksa alınmayacak mı karar verilmelidir. Gözlem birimi olarak ailedeki anne mi kız mı yoksa annenin yardımcısı olarak çalışan kişi mi seçilecektir?

Bir araştırmanın evrenini tanımlarken açıklık, kesinlik, amaca uygunluk ve örnekleme uygulaması için güçlük yaratmaması gibi ilkelerin de göz önünde bulundurulması gerekir.

Çerçevenin Belirlenmesi

Çerçeve sonlu bir evrenin bütün birimlerinin kayıtlı olduğu bir listedir, tablodur veya cetveldir. Nüfus kayıtları, seçmen kütükleri, tapu ve sicil kayıtları, ticaret ve sanayi odaları üye listeleri, ekonomik büyüklüklerine göre sanayi kuruluşlarının listesi, telefon rehberi, öğrenci kayıt listeleri, su, elektrik abonelik listeleri vb. çerçeve olarak kullanılabilecek araçlardır.

DİKKAT

Sonsuz evrenler için yapılacak örnekleme uygulamalarında çerçeve söz konusu olmaz.

Örneklemeye başlamadan önce amaca uygun bir çerçevenin var olup olmadığı, yoksa sağlanıp sağlanamayacağı öncelikle araştırılmalıdır. Araştırmaya uygun bir çerçevenin var olması durumunda bu çerçevenin güncel olup olmadığının araştırılması da önemli bir konudur. Dikkat edilmelidir ki çerçeve olmadan ne tamsayım ne de örnekleme yapılabilir.

Bir çerçeve yoksa yeni bir çerçevenin hazırlanması problemiyle karşılaşılır. Yeni bir çerçevenin hazırlanmasında çerçeve maliyeti ve kapsam hatası özellikle göz önünde tutulmalıdır. Bazen tanımlanan evrenin bazı birimleri çerçevede yer almadığı gibi tanımlanan evrenin dışında kalması gereken birimler de çerçevede yer alabilir ya da bazı birimler tekrar tekrar çerçevede yer alabilir. Bu özellikteki çerçevelerde kapsam hatası işlenmiş olur. Güncel çerçeve bulmak zordur. Kapsam hatası işlenen mevcut çerçevelerin de güncelleştirilmesi uzun zaman alır ve maliyetli olur. Bu nedenlerle uygulamada güncel olmayan bir çerçevenin kabul edilebilirliği onun güncelleştirilmesinin maliyeti ve sağladığı zaman tasarrufu ile ilişkilendirilerek belirlenir. Çerçeve, kabul edilebilir bir çerçeve hatası düzeyinde evren birimlerinin çok büyük kısmını kapsamalıdır. Şüphesiz amaç evren tanımında yer alan bütün birimleri kapsayan bir çerçeve elde etmek veya oluşturmaktır.

Örnekleme Yönteminin Seçimi

Örneklemeye girecek birimlerin belirlenmesine imkan veren yöntemlere örnekleme yöntemleri denir. Bu yöntemler örneklem için birim seçiminde uygulanan usulün keyfî ya da rassal oluşuna göre iki sınıfa ayrılır. Birinci durumda olasılıklı olmayan örnekleme, ikinci durumdaysa olasılıklı örnekleme söz konusu olur. Örnekleme yönteminin seçimiyle ilgili en önemli karar bir örnekleme planında ne tür bir örnekleme yöntemi uygulanacağıdır. Bu konu örnekleme yöntemleri başlığı altında ayrıntılı bir biçimde ele alınacaktır.

Örneklem Hacminin Belirlenmesi

Örneklem hacmi, örnekleme girecek birimlerin sayısını gösterir ve “n” simgesiyle ifade edilir. Bu sayının ne olacağına ilişkin kesin yanıt vermek mümkün değildir. Ancak, bu sorunun yanıtlanabilmesi için aşağıda açıklanan faktörlere ilişkin yapılacak nitel değerlendirmelere ve nicel yöntemlere başvurulur.

Malhotra N. K. (1996). Marketing Research An Applied Orientation, Prentice Hall International Inc.

KİTAP

Nitel Değerlendirmede Esas Olan Faktörler

- **Evrenin homojenliği:** Ele alınan evrenin ilgilenilen değişken bakımından homojen ya da heterojen olması örneklem hacminin belirlenmesine etki eder. Eğer evrenin bütün birimleri ilgilenilen değişken itibarıyla aynı değere sahipse, bir birimin incelenmesi amaca ulaşmak için yeterlidir. Ancak birimlerin özellikler bakımından farklılığı arttıkça evreni temsil edebilecek bir örneklem oluşturabilmek için örneklem hacminin de giderek büyümesi gerekir.
- **Araştırmada verilecek kararın önemi:** Önemli kararlar için olabildiğince çok veriye ve ayrıntılı bilgiye gereksinim vardır. Bu gibi durumlar büyük hacimli bir örneklem üzerinde araştırma yapmayı gerekli kılar. Ancak örneklem hacmi arttıkça maliyet ve gereksinim duyulan zaman ve nitelikli personel sayısı da artar. Burada dikkat edilmesi gereken husus, bir yandan küçük hacimli örneklem oluşturmak suretiyle bu örneklemin evreni temsil etmesi bakımından yetersiz kalmasını engellemek, diğer taraftan da gereksiz yere çok büyük hacimli örneklem seçerek zaman ve maliyet yönünden kayba uğramamak için uygun büyüklükte bir örneklem hacmini belirlemektir. Örneklem hacmi arttıkça örnekleme seçilecek her yeni birimin alınacak kararın, yapılacak tahminin doğruluğuna katkısının azalabileceğini dikkate almak gerekir.
- **Araştırmanın yapısı:** Araştırmanın doğası da örneklem hacmi üzerinde etkilidir. Uygulamada genellikle nitel araştırmalarda küçük hacimli örneklemlerde nicel araştırmalarda ise örneğin betimsel araştırmalarda daha büyük hacimli örneklemlerle çalışılır. Ayrıca araştırmalarda değişken sayısı arttıkça örneklem hacminin artırılması bilginin niteliği açısından ihtiyaç olur. Örneğin çok değişkenli analiz teknikleri ve yöntemlerinin kullanıldığı araştırmalarda örneklem hacmi büyük olmalıdır.

Örneklem Hacminin Belirlenmesinde Nicel Yöntemler

Karşılanabilecek Maliyeti Esas Alan Yöntem: Örneklem hacmi n, araştırma bütçesine bağlı olarak,

$$n = \frac{C - c_0}{C_t}$$

eşitliği ile hesaplanır. Burada,

C = Araştırma bütçesini,

c_0= Araştırmanın sabit maliyetini,

C_t= Örnekleme birimi için değişken maliyeti gösterir.

ÖRNEK 9

Araştırma bütçesinin ₺2.200 ile sınırlı olduğu bir araştırmada, sabit maliyet ₺800 ve örnekleme seçilecek her örnekleme birimi için maliyet ise ₺5'dır. Bu bütçeyle oluşturulabilecek örneklem hacmi en fazla ne olabilir?

Çözüm:

$$n = \frac{C - c_0}{C_t} = \frac{2.200 - 800}{5} = 280 \text{ birim}$$

Örneklem hacmi en fazla 280 birim olmalıdır.

- Kabul Edilebilir Hata Düzeyini Esas Alan Yöntem: Örneklem istatistiğinin dağılımının normal olduğu varsayımı altında bu yöntemle örneklem hacminin belirlenmesi için aşağıdaki eşitlikten yararlanılır.

Kabul Edilebilir Hata Düzeyi $(\bar{X} - \mu) = d$ olduğunda;

$$n = \frac{z^2 \sigma^2}{d^2}$$

Kabul Edilebilir Hata Düzeyi $(p - \pi) = d$ olduğunda;

$$n = \frac{z^2 \left[\pi(1 - \pi)\right]}{d^2}$$

Bu eşitliklerde

n = Örneklem hacmini

d = $(\bar{X} - \mu)$ veya $(p - \pi)$ araştırmacının belirlediği kabul edilebilir değeri

z = Belirlenen 1- α güven düzeyinde standart normal dağılım tablo değerini

σ = Evren standart sapmasını

π = Evren oranını

gösterir.

Örneğin, kabul edilebilir hata düzeyi $d = (\bar{X} - \mu)$ esas alındığında örneklem hacminin

$$n = \frac{z^2 \sigma^2}{d^2}$$

eşitliği ile hesaplanabilmesi için araştırmacının σ anlamlılık düzeyini ve d^2 değerini belirlemesi ve evren varyansı σ^2 hakkında bilgiye sahip olması gerekir. Evren varyansı σ^2 genellikle bilinmez. Bu durumda, σ^2 ile ilgili bilgi geçmiş yıllarda yapılmış olan aynı ya da benzer konudaki çalışmalardan elde edilebileceği gibi bir pilot çalışmadan ya da en büyük değerli gözlem değeri x_{max} ve en küçük değerli gözlem değeri x_{min} biliniyorsa ve X rassal değişkeni normal dağılıyorsa α=0.01 için

$$S = \frac{x_{max} - x_{min}}{6}$$

tahmincisi kullanılarak da hesaplanabilir.

ÖRNEK 10

Bir araştırmacı X ilinin merkez ilçesinde ikamet eden ailelerin ortalama aylık mutfak harcama tutarını tahminlemek istiyor. Ayrıca bu tahminlemede 0.05 anlamlılık düzeyinde ₺10'lik bir yanılgı payı amaçlıyor. Örneklem hacmi ne olmalıdır? Benzer amaçla bu il merkezinde yapılan araştırmalardan ailelerin aylık mutfak giderleriyle ilgili standart sapmanın ₺150 olduğu öğrenilmiştir.

Çözüm:

d = ₺10

z = 0.95, α= 0.05

σ = ₺150

$$n = \frac{(1.96)^2 150^2}{10^2} = \frac{2.200 - 800}{5} = 864.36 \rightarrow 865 \text{ birim}$$

en az 865 aile rassal olarak seçilmelidir.

Örneklemin Seçimi

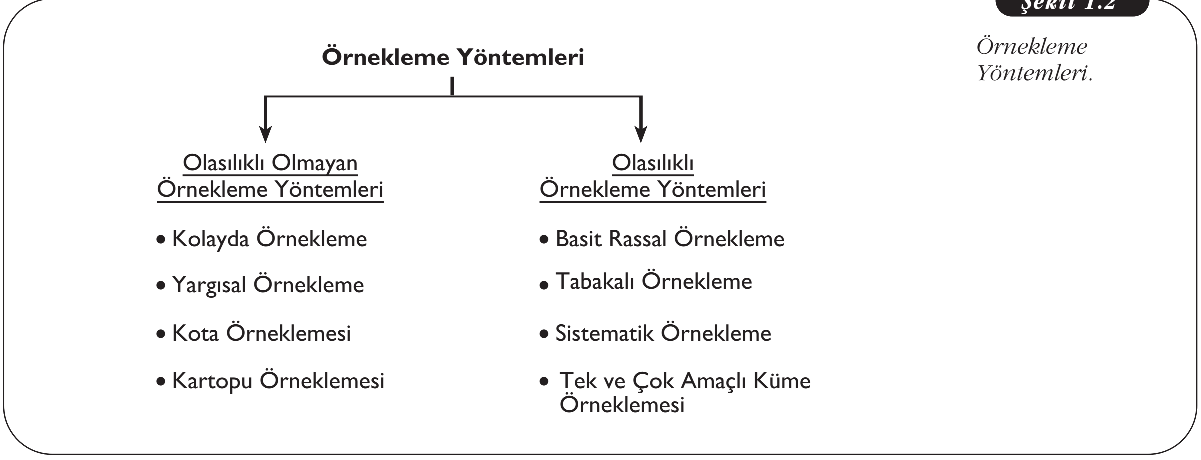
Örnekleme sürecinin bu son aşamasında örnekleme girecek birimler keyfî veya rassal seçim uygulamalarıyla seçilirler. Seçilen birimlerden gerekli veriler derlenir. Bu aşamada yapılacak işlemleri; uygun özellikte büro ve çalışma ortamı ile nitelikli işgörenlerin temini gibi sayabiliriz. Önceki aşamalarda yapılan yanlış uygulamalar ve dikkatsizlikler bu aşamada büyük sorunların yaşanmasına neden olur.

ÖRNEKLEME YÖNTEMLERİ

Örnekleme yöntemleri evrenden örnekleme birim seçiminde uygulanan usule göre Şekil 1.2'de gösterildiği gibi olasılıklı olmayan örnekleme yöntemleri ve olasılıklı örnekleme yöntemleri gibi iki sınıfa ayrılır.

KİTAP

N. K. Malhotra, Marketing Research An Applied Orientation, Prentice Hall International Inc., 1996



Şekil 1.2

Örnekleme Yöntemleri.

Olasılıklı Olmayan Örnekleme Yöntemleri

Araştırmayı planlayan ya da örnekleme uygulamasını yapan kişi ya da grubun istekleri ve değer yargıları örnekleme seçilecek birimlerin ve örneklem hacminin belirlenmesinde etkili oluyorsa yapılan örnekleme olasılıklı olmayan örneklemedir.

Olasılıklı olmayan örnekleme, birim seçiminin keyfî olarak yapıldığı örneklemedir.

Bu örnekleme yöntemleri, örneklem için birim seçiminde keyfî seçim usulünün uygulandığı örnekleme yöntemleridir. Örneklem oluşturulurken, tanımlanan evreni oluşturan birimler arasında fark gözetilir ve bütün birimlere, bilinen bir olasılıkla seçilme şansı verilmezse yapılan seçim keyfî seçimdir. Keyfî seçimle oluşturulan olasılıklı olmayan örneklemenin evreni temsil etmeyeceği anlamına gelmez. Ancak olasılık kuramının uygulanamayacağı anlamına gelir. **Olasılıklı olmayan örnekleme** uygulandığında örneklemin evreni temsil etme olasılığı bilinemez. Oysa bu bilgi araştırmacılar için çok önemlidir.

Temsili örneklem oluşturma bakımından olasılıklı olmayan örnekleme yönteminin başarısı, örnekleme uygulamasını yürüten kişi ya da grubun araştırma konusuyla ilgili deneyimine, tanımlanan evrenin özellikleri hakkındaki öncül bilgilerine ve bu evrenin ilgilenilen özelliklerinin homojenliğine bağlıdır.

Temsili örneklem oluşturma ve uygulama kolaylığı sağlaması amacıyla çeşitli olasılıklı olmayan örnekleme yöntemleri geliştirilmiştir. Uygulamada sıkça kullanılan ve aşağıda incelenen bu yöntemlerin ortak özellikleri:

- Örneklem için birim seçimi keyfîdir.
- Örneklem hacmi keyfî olarak belirlenir.
- Örneklemden hesaplanan istatistikler evren parametreleri hakkında genelleme amacıyla kullanılamaz.

Kolayda Örnekleme

Burada amaç, araştırma konusu ile ilgili ve kolayca ulaşılabilir olan birimlerden bir örneklemin oluşturulmasıdır. Araştırma konusu ile ilgili olan ve doğru yerde, doğru zamanda bulunan birimler arasından keyfî olarak birimler seçiliyorsa yapılan örneklemeye kolayda örnekleme denir. Kolayda örnekleme gönüllülük esasına göre katılan birimlerden oluşur.

ÖRNEK 11

Eskişehir'de yaşayan insanların hayat pahalılığı ile ilgili görüşlerini öğrenmek amacıyla bir araştırma planlanıyor. Hayat pahalılığının yaşandığı yer olarak haftanın herhangi bir günü Eskişehir'de kurulan semt pazarlarına alışverişe gelenler arasından keyfî olarak belirlenen ve araştırmanın amacıyla ilgili mülakata katılmayı kabul eden kişilere görüşleri soruluyor. Mülakata katılanların oluşturduğu topluluk kolayda örneklemedir. Bu örnekte doğru yer Eskişehir'de kurulan semt pazarlarıdır. Doğru zaman ise semt pazarının kurulduğu gündür. Bu araştırmada bir gün içinde veriler derlenebilir ve bu veriler çözümlenerek istenen bilgi üretilebilir.

Uygun görülen sokaktan, uygun görülen zamanda gelip geçen bireylerle görüşme yapılması ya da bir konferansa katılan belirli sayıdaki katılımcıdan araştırma konusuyla ilgili görüşlerinin alınması, birer kolayda örnekleme uygulamasıdır. Bu örnekleme uygulamasında örnekleme birimlerine kolayca ulaşılabilir, ilgilenilen değişkenlerle ilgili veriler kolayca derlenebilir ve birimlerle iş birliği sağlanabilir.

En kısa zamanda ve en az maliyetle bilgi üretilmesine ihtiyaç duyulduğu durumlarda kolayda örnekleme yöntemi bir seçenektir.

Bu örnekleme yönteminde en önemli sorun, seçilen örneklemin seçildiği evreni ne kadar temsil edebildiğidir. Kolayda örnekleme uygulaması ile oluşturulan örneklem, birim seçimindeki yanlılık nedeniyle tanımlanan bir evreni temsil etmeyebilir. Betimleyici ve ilişki araştırıcı araştırmalarda kolayda örnekleme uygun bir yöntem değildir. Kolayda örnekleme fokus gruplar, soru kâğıtlarının (anket formlarının) ön testi veya pilot çalışmalar için kullanılabilir.

Yargısal Örnekleme

Bu örnekleme de bir tür kolayda örneklemedir. **Yargısal örnekleme,** örneklemin araştırmacının ya da örneklemecinin kişisel arzu, düşünce ve deneyimlerine göre seçilmiş olduğu örneklemedir.

Evrenden temsili örneklem oluşturacağına inanılan kriterlere dayalı olarak birimlerin seçilmesi işlemine **yargısal örnekleme** denir.

Bu yöntemin kolayda örneklemeden farkı örnekleme birim seçimi için araştırmacının uzman fikirleriyle belirlediği ölçütler kullanması ve bu ölçütlerin temsilî bir örneklem oluşturacak ölçütler olduğuna inanıyor olmasıdır.

ÖRNEK 12

A Üniversitesinin sorunlarını araştırmak amacıyla bu üniversitenin üst düzey yöneticilerinden seçim yapılması yargısal örnekleme için bir örnektir. Çünkü üniversitenin üst düzey yöneticileri üniversite sorunlarını en iyi bilen kişilerdir. Bu düşünceyle seçimin bu kişiler arasından yapılması temsilî bir örneklem oluşturabilir.

Kolayda örneklemede olduğu gibi yargısal örneklemede de örnekleme birimlerine kolayca ulaşılabilir ve verilerin çok hızlı biçimde derlenmesi mümkün olur. Yargısal örnekleme pazarlama araştırmasında, kamuoyu araştırmalarında ve biyolojik araştırmalarda yaygın olarak kullanılmaktadır.

Eğer evreni oluşturan birimler araştırmaya konu olan değişkenler bakımından homojen ise kolayda ve yargısal örnekleme uygulamaları temsilî örneklem oluşturma imkânı verir. Bu örnekleme uygulamalarının maliyeti, uzman çalıştırılacağı için kolayda örneklemeye göre daha yüksektir.

Kota Örneklemesi

Örneklem için birim seçiminin keyfî olarak yapıldığı yöntemlerden biri de kota örneklemesidir. Tanımlanan sonlu evren heterojen özelliklere sahip birimlerden oluşuyorsa olasılıklı olmayan örnekleme yöntemleri grubundan kota örneklemesi temsili örneklem oluşturma amacıyla tercih edilmelidir. Bu yöntemin başarıyla uygulanabilmesi için tanımlanan sonlu evrenle ilgili bir çerçevenin var olması, ilgili evrenin homojen veya heterojen özelliğe sahip olup olmadığının sorgulanabilmesi için evren hakkında öncül bilgilere sahip olunması, evrenin heterojen olduğuna karar verilmiş ise hangi kritere göre heterojen birimlerden oluşan bu evrenin homojen birimlerden oluşacak tabakalara ayırmada kullanılacak kriterin belirlenmesi ve tabaka hacimlerinin bilinmesi gerekir.

Kota örneklemesi sürecindeki adımlar aşağıdaki gibidir:

- Evren hacmi N ve tabaka hacimleri N_h, (Tabaka sayısı h=1, 2,...) belirlenir.
- Örneklem hacmi n keyfî olarak belirlenir.
- Her tabakanın, evren hacmi içindeki oranı N_h / N belirlenir.
- Her tabakada keyfî seçimle $n_h = (N_h / N).n$ sayıda birim seçilir ve bu seçilen birimler örneklemi oluşturur.

ÖRNEK 13

Anadolu Üniversitesi yönetimi verdiği açıköğretim hizmetlerinden memnun olan öğrencilerin oranını belirlemek amacıyla bir araştırma planlamıştır. Araştırmayı gerçekleştirecek grup kota örneklemesi uygulamayı düşünmektedir.

Çözüm:

Cinsiyet tabakalama kriterine göre öğrencilere ilişkin bilgiler;

$N = 600.000$ Evren hacmi.

$N_E = 400.000$ Erkek öğrenci hacmi.

$N_K = 200.000$Kız öğrenci hacmi.

$n = 3.000$ Örneklem hacmi.

Erkek öğrenci tabakasından (N_E) seçilecek öğrenci hacmi:

$n_E = (N_E / N). n = (400.000/600.000). 3000 = 2000$ olarak bulunur.

Benzer hesaplama kız öğrenciler tabakası için de yapılırsa:

$n_K = 1000$ öğrenci bulunur.

Erkek ve kız öğrenci tabakalarından sırasıyla 2000 ve 1000 öğrenci keyfî seçimle seçilmek suretiyle $n = 2000 + 1000 = 3000$ hacimli örneklem seçilmiş olur. Bu örneklemdeki birimler üzerinden gerekli veriler derlenir ve istenilen bilgi üretilir.

Kota örneklemesi kolayda ve yargısal örneklemeye göre daha temsilî örneklem oluşturma çalışmasıdır. Ancak bu örnekleme yönteminin uygulanması sonucu oluşturulan örneklemin evreni temsil etmesinin garantisi yoktur. Çünkü kota örneklemesi uygulamasında belirlenen tabakalardaki birimlerin homojen özellikli birimlerden oluştuğunun, tabaka oranlarının doğruluğunun garantisi yoktur. Ayrıca tabakalardan birimler keyfî olarak seçildiği için yanlılık söz konusu olabilir. Bu örnekleme uygulaması sonucu oluşturulan örneklemden elde edilen bilgiler evren bilgisi için genelleme amacıyla kullanılamaz.

Kartopu Örneklemesi

Kartopu örneklemesi, özellikle bir çerçevenin mevcut olmaması ya da oluşturulmasının imkânsız olduğu durumlarda faydalı bir örneklemedir. Bu yöntemde örnekleme süreci tanımlanan evrende yer alan bir bireyin genellikle rassal olarak seçilmesiyle başlar. Belirlenen bu birey örnekleme giren birinci birimdir. Bu bireyden aynı evren tanımında yer alan tanıdığı bir bireyin olup olmadığı öğrenilir. Varsa bu bireye ulaşılır. Böylece örneklemde yer alacak ikinci birime ulaşılmış olur. Benzer şekilde bu süreç, referanslarla keyfî olarak belirlenen hacimde örnekleme ulaşılıncaya kadar sürdürülür.

ÖRNEK 14

Bir bölgedeki uyuşturucu madde kullananlar üzerinde bir araştırma yapılacak olsun. Bu bölgede uyuşturucu kullananlarla ilgili bir liste bulmak mümkün değildir. Bölgede bir ya da iki uyuşturucu kullanan tanımlanabilirse kartopu örnekleme süreci başlar. Örnekleme seçilmiş olan bu kişi ya da kişilere uyuşturucu kullanan arkadaşları ya da tanıdıklarının olup olmadığı sorulur. Varsa adresleri öğrenilir, bu kişilere ulaşılır ve bunlar da bu örnekleme seçilirler. Bu süreç keyfî olarak belirlenen n hacimli örneklem oluşturulncaya kadar sürdürülür.

Çete üyeleri ve bir ülkeye yasal olamayan yollarla girmiş kişilerle ilgili araştırmalarda, bir kentte İnternet üzerinden alışveriş yapanlarla ilgili araştırmalarda kartopu örneklemesi uygulanır. Bu örnekleme, endüstriyel ürün alan ve satanlar hakkında yapılacak araştırmalarda da kullanılabilir. Bu yöntem uygulandığında temsilî örneklem oluşturmak olanaklıdır. Kartopu örneklemesinin maliyeti ve örneklem değişkenliği düşüktür.

Tüm olasılıklı olmayan örnekleme yöntemlerinde örnekleme girecek birimlerin seçiminin keyfî olması tek yönlü hatalara neden olur. Bu tür hatalardan kaçınmak için izleyen kısımlarda ele alınacak olan olasılıklı örnekleme yöntemleri tercih edilmelidir.

SIRA SİZDE 3

- **Kolayda örnekleme mi yargısal örnekleme mi daha temsilî örneklem oluşturur?**
- **Evrenin birimleri ilgilenilen özellik bakımından heterojen ise hangi olasılıklı olmayan örnekleme yöntemi kullanılır? Açıklayınız.**

Olasılıklı Örnekleme Yöntemleri

Olasılıklı örnekleme, ilgilenilen evrendeki her örnekleme birimine hesaplanabilir ve sıfırdan farklı bir olasılıkla seçilme imkanı veren örneklemedir.

Rassal örnekleme yöntemleri olarak da alınan bu örnekleme yöntemleri örnekleme planlarında yaygın olarak uygulanır. Bu tür örneklemede örnekleme girecek birimlerin seçimi rassal olarak yapılır. Rassal seçim evrenden örnekleme girecek birimleri seçerken herhangi bir ayrıcalığın uygulanmadığı seçimdir.

Örneklem için birim seçiminde rassal seçimin uygulandığı yöntemlere **olasılıklı örnekleme** yöntemleri denir.

Basit Rassal Örnekleme

Sonlu Evrenlerde Basit Rassal Örnekleme

K İ T A P

Serper Ö. (2004). Uygulamalı İstatistik 2, 5. Baskı, Bursa. Trochim W. M. (2001). Research Methods Knowledge Base, Cornell University

Örnekleme planlarında uygulanan en temel olasılıklı örnekleme basit rassal örneklemedir. Basit rassal örnekleme hacmi N olan sonlu bir evrenden birbirinden farklı ve n hacimli oluşturulabilecek C_N^n sayıdaki olası örneklemlerin her birine incelenecek örneklem olması bakımından eşit şans tanıyan örnekleme yöntemidir. Bu tanımda belirtilen özellikleri taşıyan C_N^n sayıdaki mümkün örneklemlerin her birine basit rassal örneklem denir. Bu örnekleme yöntemi evrendeki bütün birimlere hacmi n olarak belirlenen örnekleme girmeleri bakımından bilinen ve birbirine eşit $\frac{n}{N}$ seçilme olasılığı sağlar.

ÖRNEK 15

Hizmet içi eğitim programına katılan bir firmanın 4 personelinin yapılan sınav sonuçları incelenecektir. Sınav sonuçları 60, 40, 50 ve 70 puandır. Sonlu evreni oluşturan bu 4 personel, A, B, C ve D olarak simgelendirelim ve 2 personellik

$$C_N^n = \frac{N!}{(N-n)!.n!} = C_4^2 = \frac{4!}{(4-2)!.2!} = 6$$

tane mümkün farklı örneklem aşağıdaki gibi oluşturulabilir.

BİRİMLER	MÜMKÜN ÖRNEKLEMLER
A	A, B
B	A, C
C	A, D
D	B, C
	B, D
	C, D

Tablo 1.1
4 birimlik evrenden iadesiz seçimle oluşturulabilecek 2 hacimli mümkün örneklemler.

Bu mümkün farklı 6 rassal örneklemden birinin incelenen örneklem olması olasılığı $\frac{1}{C_N^n} = \frac{1}{C_4^2} = \frac{1}{6}$ olur. Burada herhangi bir birimin yapılacak bir rassal seçimde seçilmesi olasılığı $\frac{1}{N} = \frac{1}{4} = \mathbf{0{,}25}$; herhangi bir birimin n = 2 hacimlik örneklemde yer alması olasılığı da $\frac{n}{N} = \frac{2}{4} = \mathbf{0{,}5}$ olacaktır.

DİKKAT

Örnekleme uygulamalarında mümkün örneklemler oluşturulmaz; oluşturulabileceği varsayılır. Sadece bu mümkün örneklemlerin birisi oluşturulur ve araştırma bu örneklem üzerinden yapılır.

Sonlu bir evrenden iadesiz seçimle n hacimli bir rassal örneklem oluşturmak için aşağıdaki adımlar izlenir:

- Güncel çerçeve temin edilir ya da hazırlanır.
- Örneklem hacmi belirlenir.
- Çerçevede yer alan n sayıdaki birime tanımlayıcı numara ya da işaret verilir.
- Evrendeki her birime eşit $\frac{1}{N}$ seçilme şansı vermek suretiyle örnekleme girecek birinci birim rassal seçim araçları kullanılarak belirlenir.
- Geriye kalan (N - 1) birimin her birine yine eşit şans vermek suretiyle ikinci birim seçilir. Bu birimin seçilmesi olasılığı 1 / (N-1) olur.
- Bu birim seçim süreci n hacimli örneklem seçilinceye kadar tekrarlanır.

Açıklanan seçim sürecinde her çekilişte seçilen birim incelendikten sonra evrene iade edilmediği için bu seçim sürecine iadesiz rassal seçim süreci adı verilir. Eğer basit rassal örnekleme planlarında önceki çekilişte seçilen birim incelendikten sonra evrene iade ediliyorsa başka bir ifadeyle birimler tekrar tekrar seçilme şansına sahipse bu seçim sürecine iadeli rassal seçim süreci adı verilir. Bu seçim sürecinde evren hacmi çekilişten çekilişe değişmez. Sonlu bir evrenden iadeli seçimle bir rassal örneklem seçilirse sonsuz evrenden basit rassal örnekleme yapılıyormuş gibi bir anlam ifade eder. Bu çekiliş sürecinde evrendeki her birimin yapılacak çekilişlerin her birinde birbirine eşittir. $\frac{1}{N}$ olan seçilme olasılığına sahiptir ve birbirini izleyen çekilişler bağımsızdır. Sonlu evrenlerde evren hacminin büyük ya da küçük oluşu iadeli ya da iadesiz seçimler için önemli farklılıklar gösterir. Evren hacmi büyük, örnekleme oranı $\frac{n}{N}$ küçük olduğu zaman iadeli ve iadesiz örneklemler benzer özellikler gösterirler. Çünkü iadeli çekiliş uygulandığında önceki çekilişlerde seçilmiş olan bir birimin yeniden örnekleme seçilmiş olma olasılığı çok küçüktür. Ancak evren hacmi küçükse iadeli ve iadesiz rassal seçimlerle oluşturulan aynı hacimli örneklemler için hesaplanan örneklem istatistikleri ile evren parametreleri karşılaştırılırsa iadesiz basit rassal seçimle oluşturulan örneklem iadeli olana göre daha az hatayla tahminleme imkânı sağlar. Bu özellik nedeniyle de iadesiz rassal seçim uygulamada genellikle başvurulan yöntem olmaktadır.

İlgilenilen özellik bakımından evrenin homojen olması durumunda basit rassal örnekleme tercih edilmesi gereken bir yöntemdir. Örnekleme planlarında basit rassal örnekleme yöntemlerinin tercihlerini etkileyen önemli sınırlayıcılar vardır. Bunlardan birincisi güncel bir çerçeve oluşturma ya da hazırlama güçlüğüdür. İkincisi evrenin birimleri geniş bir coğrafi alana yayılmışsa basit rassal örnekleme uygulaması çok zaman alır ve veri derleme maliyeti giderek artar. Üçüncüsü evren homojen değilse basit rassal örneklem sonuçlarının başarısı diğer olasılıklı örnekleme yöntemleri sonuçlarının başarısından düşüktür.

Sonsuz Evrenlerde Basit Rassal Örnekleme

KİTAP

Sonsuz evrenlerde basit rassal örnekleme ile ilgili bilgiler için Neter J., Wasserman W., Whitmore G. A. (1993). Applied Statistics, (Boston: Fourth Edition, Allyn and Bacon) adlı kitaptan yararlanmıştır.

Sonsuz evren "aynı koşullar altında işleyen bir sürecin sonuçlarının oluşturduğu topluluktur" şeklinde tanımlanmıştı ve bir bisküvi fabrikasındaki bisküvi paketleme süreci örnek verilmişti. Bisküvi paketleme süreci sonucu olan her bir paket bir birim, paketleme süreci devam ettikçe yeni bisküvi paketleri evrene dahil olduğu için "bu sürecin sonuçları olan paketler sonsuz evreni oluşturur" denmişti. Örnekten de anlaşılabileceği gibi sonsuz evrendeki bir başka ifadeyle aynı koşullar altında işleyen bir sürecin bütün sonuçları listelenemez. Bütün birimlerle ilgili bir çerçeve hazırlanamaz, bunun yerine bu birimlerin ilgilenilen bir değişken X için bir teorik olasılık dağılımı tarif edilebilir. Tarif edilen bu teorik olasılık dağılımına sonsuz evren adı verilir.

Sonsuz evrenin birimlerinin X değişkeni için ölçümlenen değerleri (gözlem değerleri) bu birimler için X değişkeninin gerçekleşen değerleridir. Eğer sonsuz evrenin birimleri kararlı (aynı koşullar altında meydana gelen) bir sürecin sonuçları, birimleri ise n sayıdaki birimin ilgilenilen X değişkeni bakımından aldığı x_1 x_2,..., x_n değerleri (gözlem değerleri) X_1, X_2,... X_n rassal değişkenlerinin birer gerçekleşmesi olduğu düşünülür. Buna göre, bir süreç tarafından türetilen birbirinden bağımsız ve benzer olasılık dağılımına sahip n sayıdaki X_1, X_2,... X_n rassal değişkenlerinin oluşturduğu topluluğa sonsuz evrenden seçilmiş basit rassal örneklem denir. Bu rassal değişkenlerin teorik olasılık dağılımına sonsuz evren adı verilir.

ÖRNEK 16

Bir bisküvi fabrikasının paketleme sürecinin planlanan ağırlıkta üretimi gerçekleştirip gerçekleştirmediğini incelemek olsun. Bisküvi paketleme süreci aynı koşullar altında işleyen kararlı bir süreçtir. Sürecin sonuçları olan paketler sonsuz evrenin birimleridir. İncelenen değişken üretilen ve üretilecek olan paketlerin ağırlığıdır. Bu üretim sürecinin sonucu olan paketlerin seçilen n tanesinin ölçüm ağırlıkları olan x_1 x_2,..., x_n *sırasıyla* X_1, X_2,... X_n *rassal değişkenlerinin gerçekleşmeleridir. Buna göre paketleme sürecinden rassal olarak seçilen n tane paketin ölçülen ağırlıkları kullanılarak hesaplanan olasılık dağılımı* X_1, X_2,... X_n *teorik olasılık dağılımını ifade eden sonsuz evrenden seçilmiş basit rassal örneklemi oluşturduğu söylenebilir.*

Tabakalı Örnekleme

Tanımlanan evrenin birimleri araştırmaya konu olan değişkenler bakımından heterojen ise önemli farklılıklar gösteriyorsa **tabakalı örnekleme** temsilî örneklem oluşturabilmek için tercih edilmelidir. Tabakalı örnekleme evren birimlerinin tabakalara ayrıldığı ve her tabakadan rassal seçimle örneklemin oluşturulduğu örneklemedir.

Tabakalı örnekleme evren birimlerinin tabakalara ayrıldığı ve her tabakadan rassal seçimle örneklemin oluşturulduğu örneklemedir.

Tabakalı örnekleme üzerinde araştırma yapılacak evren ilgilenilen deişkenler yönünden heterojen olduğunda evren parametre tahminine ilişkin varyansın olabildiğince küçük olmasını sağlayan örneklemedir.

Tabakalı örnekleme 4 aşamalı bir süreçtir.

- Tabakalama kriterinin belirlenmesi.
  Tabakalı örnekleme uygulaması yapacak araştırmacı önce incelenecek değişkenler açısından önemli farklılıklar gösteren N hacimli evrenin birimlerini homojen birimlerden oluşacak tabakalara ayırmada kullanılacak kriter belirler. Burada önemli olan belirlenecek kriterin tabakalar içindeki birimleri olabildiğince homojen, tabakalar arasında ise birimlerin heterojen olmasını sağlayacak kriter olmasıdır. Aynı zamanda bu kriterin uygulama ve ölçme kolaylığı da sağlamak suretiyle maliyet artırmadan tahminleme hatasını azaltması gerekir. Tabakalama kriteri ilgilenilen parametre ile sıkı sıkıya ilişki içindedir. Belirlenen tabakalama kriterinin uygunluğu örneklem değişkenliğinin etkinliği üzerinde olumlu yönde etkilidir. Bu nedenle asimetrik bölünmeye sahip evrenlerde tabakalı örnekleme uygulamasını tercih etmek bir zorunluluktur. Tabakalama amacıyla kullanılabilecek kriterlere demografik özellik, tüketici türü, sosyoekonomik sınıf, meslek grubu, firma büyüklüğü, coğrafi yerleşim yeri, fakülte türü vb. örnek olarak gösterilebilir.
- Tabakaların oluşturulması.
  Belirlenen tabakalama kriteri itibarıyla N hacimli bir evren daha homojen, L sayıda ve hacimleri N_1, N_2,...., N_L olan tabakalara ayrılır. Bu aşamada önemli olan tanımlanan evrendeki her bir birimin yalnız bir tabakaya ait olması ve hiçbir birimin açıkta kalmamasının sağlanmasıdır. Başka bir ifadeyle

  $$\mathbf{N_1} + \mathbf{N_2} + \dots + \mathbf{N_L} = \sum_{h=1}^{L} N_h = N$$

  olmalıdır. Tabaka sayısı L arttıkça tabakaların homojenliği de artacağından tabaka varsaynsları giderek küçülecek ve buna bağlı olarak da tahminlerin güvenilirliği giderek artacaktır. Tabaka sayısının artması maliyetleri yükseltir ve uygulama zorluğu yaratır. Bu nedenlerle tabaka sayısı L belirlenirken tabaka sayısının yaratacağı maliyet, uygulama zorluğu ve elde edilecek tahminlerin güvenilirliği birlikte değerlendirilmelidir. Deneyimler ve uygulamalar tabaka sayısının 6 ‘dan fazla olmamasını önermektedir.
- Tabakalardan birimlerin seçilmesi.
  Her tabakadan basit rassal seçimle sırasıyla n1, n2,..., nL hacimli alt örneklemler oluşturulur. Alt örneklem hacimleri toplamı örneklem hacmine eşittir. Başka bir ifadeyle n örneklem hacmini göstermek üzere

  $$\mathbf{n_1} + \mathbf{n_2} + \dots + \mathbf{n_L} = \sum_{h=1}^{L} n_h = n$$

  olmalıdır.
- Verilerin derlenmesi.
  Oluşturulan alt örneklem birimleri üzerinden veriler derlenir, bu veriler kullanılarak araştırma amaçları için gerekli olan istatistikler hesaplanır ve bu istatstiklere dayanarak istatistiksel çıkarımlar yapılır.

Daha önce de vurgulandığı gibi tabakalar içi homojenlik arttıkça tabakalar içi varyanslar küçülür. Bu da ilgili evren parametre tahminleyicisinin varyansını küçültür. Bu sonuca göre heterojen evrenlerde aynı örneklem hacmi için basit rassal örnekleme uygulamasının örnekleme hatası, tabakalı örneklemenin örnekleme hatasından büyük olur. Heterojen evrenler için tabakalı örnekleme yöntemi daha etkindir.

Tabakalı örneklemenin diğer bir üstünlüğü ilgilenilen evrenin yanısıra her tabaka içinde ayrı bilgi elde etme olanağı sağlamasıdır. Uygulamada evrene göre tabakalar için çerçeve oluşturmak daha kolay olabilir. Ancak sağladığı kolaylıklara rağmen tabakalı örneklemenin bazı güçlükleri de vardır. Örneğin tabaka hacimleri ve bunların toplamı olan evren hacminin bilinmesi gerekir. Bu her zaman mümkün olamamaktadır. Ayrıca ilgilenilen evrenin homojenliğinin sorgulanması için de bu evren hakkında pek çok öncül bilgiye gereksinim vardır. Bu öncül bilgilerin yetersizliği ve geçersizliği oluşturulacak örneklemin temsil niteliğini olumsuz yönde etkiler. Tabakalı örnekleme sürecinin adımlarıyla ilgili yukarıdaki kuramsal bilgileri örnek araştırma üzerinde uygulayalım.

ÖRNEK 17

Amaç: Bir üniversitede görev yapan 1000 öğretim elemanının bilgisayar kullanım alışkanlığını araştırmak.
Evren: 1000 öğretim elemanının oluşturduğu topluluktur. Sonlu evrendir, hacmi N= 1000 kişidir.

Örnekleme Yöntemi: Evren, sonlu evren olduğu için araştırmacı tamsayım da uygulayabilir, örneklemeye de başvurabilir. Araştırmacı örnekleme yapmayı gerekli kılan nedenleri değerlendirmiş ve örnekleme yapmaya karar vermiştir. Araştırmacının gözlemlerine ve öncül bilgilerine göre öğretim elemanlarının bilgisayar kullanım alışkanlıkları bakımından heterojendir. Çünkü yaşları genç olan araştırma görevlileri, öğretim görevlileri daha sık ve daha farklı amaçlarla bilgisayar kullanırken profesörler daha az süreli ve daha sınırlı amaçlarla bilgisayar kullanmaktadırlar. Bu değerlendirmeye göre araştırmacı tabakalı örnekleme yöntemini örnekleme amacıyla tercih etmiş ve uygulama adımlarını aşağıdaki gibi izlemiştir:

- Öğretim elemanlarının ünvan türü kriterine göre bilgisayar kullanma alışkanlığı bakımından homojen tabakalar ayrılabileceği düşünülmüştür.
- Tabakalar öğretim üyesi dışındakiler (araştırma görevlisi, öğretim görevlisi), yardımcı doçent doktor (Yrd. Doç. Dr.), doçent doktor (Doç. Dr.) ve profesör doktor (Prof. Dr.) şeklinde belirlenmiştir. Yönetimden elde edilen bilgilere göre tabaka hacimleri öğretim üyesi dışındakiler $N_1 = 400$, Yrd. Doç. Dr. $N_2 = 300$, Doç. Dr. $N_3 = 200$ ve Prof. Dr. $N_4 = 100$ birim olduğu görülmüştür.
- $n = 100$ olarak belirlenen örneklem hacmi tabaka hacimlerinin (N_h) evren hacmi N içindeki paylarıyla orantılı olarak dağıtıldı:

  $$\mathbf{n_h} = (\frac{N_h}{N}).\mathbf{n}$$

  $$\mathbf{n_1} = (\frac{N_1}{N}).\mathbf{n} = \mathbf{(400/1000).100 = 40}$$

  Benzer şekilde h = 2, 3, 4 için hesaplama yapılırsa $n_2 = 30$, $n_3 = 20$, $n_4 = 10$ hacimlik örneklemler rassal olarak seçildi ve

  $n_1 = 40 + n_2 = 30 + n_3 = 20 + n_4 = 10 = n = 100$

  hacimlik tabakalı örneklemler oluşturuldu.
- Oluşturulan örneklemdeki öğretim elemanlarından birinci elden veri derleme yöntemiyle veriler derlenip çözümleme yapılır ve gerekli bilgiler üretilebilir.

Sistematik Örnekleme

Örneklem için birim seçimi aşağıda ele alınan bir sistematiğe uygun olarak yapıldığı örnekleme sürecine sistematik örnekleme adı verilir. Bu yöntemin sınırlayıcıları ilgili evrene ilişkin bir çerçevenin var olup olmaması veya birimlerin doğal bir sıraya sahip olup olmamasıdır.

Bir sistematik örneklem oluşturmak için aşağıdaki adımlar izlenir:

- Evrendeki birimler 1'den N'ye kadar numaralandırılır.
- Araştırma için yeterli olacak örneklem hacmi n belirlenir.
- $\mathbf{k} = \frac{N}{n}$ büyütme faktörü hesaplanır. Bu oran örnekleme aralığını gösterir.
- 1 ile k arasında bir tam sayı rassal olarak seçilir. Bu sayı a ile gösterilirse a örnekleme girecek birinci birimin sıra numarası olur.
- a'ıncı birimi k aralıklarıyla izleyen a + k'ıncı, a + 2k'ıncı,..., a + (n - 1) k'ıncı sıra nolu birimler örnekleme seçilir ve n hacimli sistematik örneklem oluşturulur.
- Oluşturulan örneklemden elde edilen veriler kullanılarak gerekli istatistikler hesaplanır.

Bütün bu adımları daha açık bir şekilde bir örnek üzerinden gösterelim:

ÖRNEK 18

- *Evrendeki birim sayısı n = 1000 öğretim elemanıdır. Bu öğretim elemanlarının unvan türü, soyadı sırası itibarıyla 1'den N'ye kadar listeler üzerinde numaralandırma yapıldı.*
- *Diyelim ki n = 100 birimlik bir örneklem seçileceği tasarlandı. k = 1000 / 100 = 10, hesaplandı. Bunun anlamı, her 10'uncu birim örnekleme alınacak demektir.*
- *1, 2,..., 10 arasından rassal olarak bir sayı seçildi. Seçilen sayı 4 olsun. A = 4. Sıradaki öğretim elemanı örnekleme girecek birinci birimdir.*
- *4. Sıra nolu öğretim elemanından başlayarak her 10. (4, 14, 24, 34,...) sıra nolu öğretim elemanı örnekleme alınarak n = 100 birimlik rassal örneklem oluşturulmuş olur.*

Tablo 1.2
Örnek problem için sistematik örnekleme uygulama tablosu.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1000

Sistematik örneklemede uygulanacak sistematiğin belirlenmesi her zaman yukarıdaki hesaplamalarda olduğu gibi yapılmayabilir. Evreni oluşturan birimlere sıra numarası verilemiyorsa veya numaralandırma çok zaman alıyor ve masraflı oluyorsa ve fakat sonlu veya sonsuz evren birimleri rastgele dizilişlere veya gelişlere sahip ise sistematik örnekleme uygulanabilir. Örneğin bir üniversitenin 20.000 öğrencisi üzerinde yapılacak bir araştırma için haftanın seçilen bir öğretim günü saat 8:30'da başlayıp her yarım saat aralıklarla üniversite kampüs kapısından giren öğrencinin örnekleme alınması uygulaması da bir sistematik seçimdir. Bir süpermarketten ayrılan her k'ıncı müşteriyle görüşme yapılarak yürütülen araştırmalar da bu örneklemin uygulandığı araştırmalardır. Sistematik örnekleme evrendeki her birimin örneklemde yer alma olasılığı bakımından basit rassal örneklemeye benzer ve bu olasılık n / N dir. Ancak mümkün örneklemlerden herhangi birinin incelenen örneklem olması olasılığı basit rassal örneklemede 1 / C_N^n olduğu hâlde sistematik örneklemede belirlenen sistematiğe göre örneklem olma şansına sahip kombinasyonların her biri için eşit 1 / k diğerlerininki için 0'dır.

Sistematik örnekleme uygulaması sonucu oluşturulacak örneklemin temsil niteliği evren birimlerinin sıralandırılması ve k aralığı ile ilişkilidir. Bu sıralandırma araştırmaya konu olan değişkenlerle ilişkilendirilerek yapılırsa sistematik örnekleme basit rassal örneklemeye göre daha temsilî örneklem oluşturma imkânı verir.

Örneğin firmaların düşük ciroya sahip olandan büyük ciroya sahip olana doğru sıralanması, otellerin yıldız sayıları bakımından büyükten küçüğe doğru sıralanması, öğrencilerin küçük boyludan büyük boyluya doğru sıralandırılması durumunda sistematik örnekleme her gruptan birimin örnekleme girmesini temin edebilir. Ancak örnekleme aralığının uygun şekilde belirlenememesi hâlinde sistematik örneklemenin basit rassal örneklemeye olan yukarıda açıklanan üstünlüğü ortadan kalkar. Çünkü belirli özelliğe sahip birimlerin gereksiz oranda örnekleme girmesi söz konusu olabilir. Bu durum örneklemin temsil niteliğini olumsuz yönde etkiler.

Sistematik örnekleme önceki rassal örnekleme yöntemlerine göre daha az maliyetli ve uygulaması kolaydır.

Çerçevenin doğal yapısında tekrarlamalar varsa sistematik örnekleme kullanılmamalıdır. Örneğin, veriler aylık olarak düzenlenmiş ve $k = 12$ alınmışsa her yılın aynı ayı örnekleme gireceğinden bu tür bir uygulama tek yönlü hatalara neden olabilir.

Tek Aşamalı ve Çok Aşamalı Küme Örneklemesi

Tabakalı örneklemede olduğu gibi evrenin birimleri küme adı verilen gruplara ayrılır. Bu gruplar genellikle doğal olarak vardır. Her küme bir örnekleme birimi olarak tanımlanır. Kümeler arasından rassal olarak belirli sayıda küme seçilir ve seçilen kümelerdeki gözlem birimlerinin tamamı örneklemi oluşturur. Örneğin bir organize sanayi bölgesinde faaliyette bulunan iş yerlerinde çalışan işgörenler hakkında bir araştırma planlandığında bu organize sanayi bölgesinde her bir iş yerinde çalışan işgörenler bir küme olarak tanımlanabilir. Hanehalkı geliriyle ilgili bir araştırmada her mahalledeki hanehalkı topluluğu bir küme olarak tanımlanır. Örneklerden de anlaşılabileceği gibi kümeler genellikle bir coğrafi kritere göre tanımlanmaktadır.

Küme örneklemesi bir ve daha fazla kümeleme aşamaları ile de uygulanabilir. Bir kümeleme aşaması ile gözlem birimlerine ulaşılıyorsa tek aşamalı kümeleme; iki veya daha fazla kümeleme aşaması ile gözlem birimlerine ulaşılıyorsa çok aşamalı kümeleme adı verilir.

Bu örnekleme yöntemleri evrendeki birimlerin homojen, hacimlerinin çok büyük ve geniş bir coğrafi alana yayılmış olmaları ya da örnekleme girecek birimlere ilişkin bir çerçeve oluşturmanın mümkün olmadığı durumlarda tercih edilmesi gereken yöntemlerdir.

Tek aşamalı (küme) örneklemesi sürecinde aşağıdaki adımlar izlenir.

- İlgilenilen evrendeki birimler genellikle coğrafi kritere göre kümelere ayrılır. Bu, birinci düzey kümelemedir. Kümeler doğal olarak bir mekanda var olan birimlerden oluşur. Küme sayısı "M" simgesiyle gösterilir. Üniversiteler, banka şubeleri, lojistik firmaları, kamu kurumları, ortaöğretim okulları birer kümedir. Örneğin ortaöğretim kurumlarını ele alalım. Bu okulların öğrencileri, sınıfları, öğretmenleri kümeleri oluşturur.
- Kümeler arasından rassal seçimle "m" sayıda küme seçilir.
- Seçilen kümelerdeki birimlerin toplamı tek aşamalı küme hacmini gösterir.

Tanımlanan birinci aşama kümelerine, benzer kritere göre ikinci, üçüncü ve n'inci aşama kümelere ayrılır ve son kümeleme aşamasındaki kümeler arasından rassal seçimle m sayılı küme seçilir ve seçilen kümelerdeki birimlerden örneklem oluşturulursa yapılan örneklemeye çok aşamalı küme örneklemesi denir.

Tek ve çok aşamalı küme örneklemesi ile ilgili kuramsal açıklamaları bir araştırma örneği üzerinden açıklayalım.

ÖRNEK 19 *Anadolu Üniversitesi'nde 2010-2011 öğretim yılında örgün öğretim yapan lisans programlarına kayıtlı olan öğrencilerin kendilerine sunulan eğitim-öğretim ortamlarından memnun olup olmadıkları araştırılmak isteniyor.*

Evren: Bu araştırmanın evreni N=26000 olan öğrencilerin oluşturduğu topluluktur. Sonlu ve büyük hacimli bir evrendir. Evren hacminin büyük olması, bütün öğrencilere ulaşmanın güçlükleri gibi nedenlerle araştırma için örneklemeye başvurulması düşünülmüştür.

Örnekleme Yöntemi: Öğrenciler iki büyük yerleşkedeki öğretim birimlerinde öğrenim görmektedirler. Araştırmacı, öğrencilerin eğitim-öğretim ortamlarından memnuniyetlerini değerlendirmeleri bakımından homojen olduğu öncül bilgisine sahiptir. Bu değerlendirmelere göre araştırma için uygun örnekleme yöntemi olarak önce tek aşamalı sonra iki aşamalı örnekleme yöntemi seçilmiştir.

Tek aşamalı örnekleme uygulamasının adımları:

- Öğrenciler lisans öğrenimi yapan program türü kriterine göre kümelere ayrılmıştır. Kümeler İktisadi ve İdari Bilimler Fakültesi, Fen Fakültesi, Güzel Sanatlar Fakültesi, Mühendislik Fakültesi, Edebiyat Fakültesi, Turizm Yüksekokulu vb. öğrencileri topluluğu şeklinde tanımlanmıştır. Bu tanıma göre küme sayısı M=6' dır.
- M=6 lisans programı arasından m=2 program rassal olarak seçilir. Varsayalım ki seçilen 2 program İktisadi ve İdari Bilimler Fakültesi ve Mühendislik Fakültesi olsun. Seçilen birinci fakültedeki öğrencilerin tamamı N_1=5000 ve ikinci fakültedeki öğrencilerin tamamı N_4=4000 ise örneklem hacmi $n=N_1+N_4=5000+4000=9000$ öğrenci örnekleme seçilmiş olur.

İki aşamalı örnekleme uygulaması benimsendiğinde birinci aşamada seçilen İktisadi ve İdari Bilimler ve Mühendislik Fakülteleri bölüm türü kriterine göre tekrar kümelere ayrılırsa ikinci aşama kümeleme yapılmış olur. İkinci aşamada oluşturulan bölümlerden (kümelerden) rassal seçimle m2 sayıda bölüm seçilir ve bu bölümlerdeki bütün öğrenciler örneklemi oluşturur.

Tek ve çok aşamalı örnekleme uygulamasında tanımlanan kümeler örnekleme birimi olarak benimsendiğinden basit rassal örneklemede olduğu gibi evrenle ilgili bir çerçeveye gerek yoktur. Sadece seçilen kümelerle ilgili çerçeveye gereksinim vardır. Bu durum örnekleme uygulamasında zaman, maliyet tasarrufu yanında uygulama kolaylığı sağlamaktadır.

Eğer birimler kümeler arasında homojen değilse seçilen kümelerdeki birimlerden oluşacak örneklemin evreni temsil niteliği tartışılır çünkü evreni oluşturan her türden birim örnekleme girmemiş olur.

Tek aşamalı ve çok aşamalı örnekleme yöntemleri örnekleme maliyetini azaltarak onun etkinliğini artırırken tabakalı örnekleme doğruluğu artırmaktadır. Küme örneklemesinde kümelerdeki birimlerin mümkün oldukça heterojen olması, tabakalı örneklemede tanımlanan tabakaların ise mümkün oldukça homojen olması istenir.

SIRA SİZDE 4 **Olasılık örnekleme ile olasılıklı olmayan örneklem yöntemleri arasındaki temel fark nedir, açıklayınız.**

ÖRNEKLEME DAĞILIMLARI

Daha önce de değinildiği gibi tanımlanan evrene ilişkin sayısal karakteristiklere parametre adı verilir ve parametre genel olarak θ (theta) simgesiyle gösterilir. Tam sayım yapılamadığı durumlarda araştırmacılar istatistiksel tahminleme ve karar verme (istatistiksel çıkarım) problemleri ile karşılaşırlar. Bu çıkarımlar örneklem istatistiklerine dayanır. Örneklem istatistiklerinin genel gösterimi simgesiyle yapılır. Örneklem istatistiği bilindiği gibi rassal olarak seçilen n hacimli örneklemden elde edilen $x_1, x_2, ..., x_n$ gözlem değerlerinin kullanılmasıyla hesaplanan karakteristiklerin genel adıdır.

Örnekleme sürecinde rassal olarak seçilen n hacimli bir örneklem için hesaplanan istatistik sadece ait olduğu örneklem için bilgi niteliğindedirler. Çünkü incelenen n hacimli bir örneklem aynı hacimli ve fakat farklı birimlerden oluşabilecek mümkün örneklemlerden sadece birisidir ve mümkün örneklemlerin her biri için hesaplanacak istatistikler birbirinden farklı ve evren parametre değerlerine eşit ($\theta=\hat{\theta}$), büyük ($\theta > \hat{\theta}$) veya küçük ($\theta < \hat{\theta}$) olabilir. Bu nedenle örneklem istatistiklerinden yararlanarak evren parametreleri hakkında tahminleme ve karar verme sürecinde rassal olarak seçilen n hacimli bir örneklemin hesaplanan istatistiğinden değil; o istatistiğin mümkün örneklemlerde alacağı değerlerin dağılımından ve bu dağılımın özelliklerinden yararlanılır.

Uygulamada, n hacimli bir tek örneklem seçilir ve bu örneklem için tahminlenecek veya karar verilecek parametre hakkında bilgi üreten istatistik hesaplanır. n hacimli mümkün örneklemler seçilmez, örneklem istatistikleri hesaplanmaz ve bu istatistiğin dağılımı oluşturulmaz. Mümkün örneklemler seçilmiş gibi düşünülerek bu istatistiğin varsayımsal dağılımından yararlanmak suretiyle evren parametreleri hakkında çıkarımlar yapılabilir. Bu durum, örneklem istatistiğinin aynı hacimli örneklemden örnekleme farklı değerler alan rassal değişken olduğu esasına dayanır. Buna göre bir evrenden rassal olarak seçilen n sayıdaki örneklemin gözlem değerleri $x_1, x_2, ..., x_n$ sırasıyla $X_1, X_2, ..., X_n$ rassal değişkenlerinin gerçekleşen değerleridir. Bu gözlem değerlerinden hesaplanan istatistikler de bir rassal değişkendir. Yapılan açıklama bağlamında örnekleme dağılımı, bir rassal değişken olan $\hat{\theta}$ örneklem istatistiğinin olasılık dağılımıdır şeklinde tanımlanabilir.

Çeşitli amaçlar için örneklemeye karar verildiği zaman dikkatlerin en çok odaklandığı parametreler tek evren aritmetik ortalaması (μ), tek evren oranı (π),iki evren ortalaması arasındaki fark ($\mu_1 - \mu_2$) ve iki evren oranı arasındaki fark ($\pi_1-\pi_2$) olmaktadır. Bu nedenle bu ünitenin izleyen kısımlarında bu parametreler hakkında bilgi üreten örneklem istatistiklerinin sırasıyla örneklem aritmetik ortalaması ve örneklem oranı p, iki örneklem aritmetik ortalaması arasındaki fark ve iki örneklem oranı arasındaki fark ($p_1 - p_2$) 'nin örnekleme dağılımları ve özellikleri incelenecektir.

Tek Örneklem İstatistiğine İlişkin Örnekleme Dağılımı

Ortalamanın ($\bar{X}$'nın) Örnekleme Dağılımı

Bir örneklem istatistiği olan örneklem aritmetik ortalaması ($\bar{X}$) rassal bir değişkendir. $\bar{X}$ rassal değişkeninin olasılık dağılımına, ortalamanın örnekleme dağılımı adı verilir. Bir başka anlatımla tanımlanan evrenden n hacimli bir rassal örneklem değil de aynı hacimli C_N^n (veya N^n) sayıdaki mümkün rassal örneklemlerin seçildiğini ve her mümkün örneklem için $\bar{X}$ hesaplandığını varsaydığımızda $\{\bar{X}_i\}$ dan oluşan bir frekans dağılımı elde edilebilir. Bu dağılıma ortalamanın örnekleme dağılımı adı verilir. Bir örnek üzerinde $\bar{X}$ nın örnekleme dağılımını oluşturalım.

ÖRNEK 20

Bir banka şubesinde çalışan 4 işgören bir üst ünvana yükselme (terfi) için açılan hizmet içi eğitim programına katılmışlardır. Bu kişilerin simgesel isimleri ve eğitim sonunda açılan sınavda aldıkları puanlar aşağıda verilmiştir. Evren ortalamasını hesaplayınız ve $\overline{X}$ nın örnekleme dağılımını oluşturunuz.

Eğitime Katılanlar (Birimler)	**Başarı Puanları X_i**
A	90
B	80
C	60
D	70
	300

Çözüm:

Evren ortalaması μ, tamsayım yapıldığında, gözlem değerleri $x_1, x_2, \ldots, x_N$ olarak gösterildiğinde ve yukarıdaki veriler kullanıldığında;

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{\mathbf{300}}{\mathbf{4}} = \mathbf{75\,puan}$$

şeklinde hesaplanır. Hesaplama hatası yapılmamış ise 75 puan kesin, doğru olan ortalama başarıyı gösterir.

Tamsayım yapılamadığı durumlarda açıktır ki yukarıdaki başarı puanları (x_i değerleri) derlenememiş ve μ=75 puan bilgisi hesaplanamamış olur. Bu durumda μ hakkında tahminleme ve karar verme (istatistiksel çıkarım) problemleriyle karşılaşılır. Bu türden problemlerin çözümlenebilmesi için μ hakkında bilgi üreten örneklem istatistiği $\overline{X}$ nın örnekleme dağılımının özellikleriyle ilgili bilgilere gereksinim vardır.

$\overline{X}$ ların örnekleme dağılımının oluşturulması için aşağıdaki adımlar izlenir.

- N=4 birimlik evrenden iadeli veya iadesiz seçimle belirlenen n hacimli mümkün örneklemler seçilir. Örneğin n=2 için iadesiz seçimle oluşturulabilecek mümkün örneklem sayısı;

$$C_N^n = \frac{N!}{(N-n)!.n!} \qquad C_4^2 = \frac{\mathbf{4!}}{(\mathbf{4}-\mathbf{2})!.\mathbf{2!}} = \mathbf{6\,adet}$$

olup aşağıdaki tabloda verilmiştir.

Tablo 1.3
N=4 hacimli evrenden n=2 hacimli mümkün örneklemlerin $\overline{X}_i$ larının larının örnekleme dağılımı.

Örneklem No	Mümkün Örneklemler	Örneklem Gözlem Değerleri	$\overline{X}_i$
1	A, B	90, 80	85
2	A, C	90, 60	75
3	A, D	90, 70	80
4	B, C	80, 60	70
5	B, D	80, 70	75
6	C, D	60, 70	65
			450

$\bar{X}_i$ lar serisinin dağılımına, $\bar{X}$ nın örnekleme dağılımı adı verilir. Bu dağılımın ortalaması $\mu_{\bar{x}}$ simgesiyle gösterilir ve aşağıdaki gibi hesaplanır:

$$\mu_{\bar{x}} = \frac{\sum_{i=1}^{C_N^n} X_i}{C_N^n} = \frac{450}{6} = 75\,\text{puan}$$

Görüldüğü gibi evren ortalaması μ, $\bar{X}$ nın örnekleme dağılımının ortalaması $\mu_{\bar{x}}$ ya eşittir ve

$\mu = \mu_{\bar{x}} = 75$ puan

şeklinde yazılır. Tablo 1.3'teki $\bar{X}_i$ lar frekans dağılımı olarak düzenlendiğinde n=2 hacimli farklı örneklem ortalamalarının olasılığı Tablo 1.4'teki gibi gösterilmiş olur.

$\bar{X}_i$	n_i	Örneklem Ortalamalarının Elde Edilmesi Olasılığı
65	1	1/6 = 0,167
70	1	1/6 = 0,167
75	2	2/6 = 0,333
80	1	1/6 = 0,167
85	1	1/6 = 0,166*
	6	1,000

Not: *: Olasılıklar toplamını 1'e eşitlemek için düzeltme yapılmıştır.

Tablo 1.4
N=4 hacimli evrenden n=2 hacimli mümkün örneklemlerin $\bar{X}_i$ larının frekans dağılımı.

Tablodaki bilgilere göre, örneğin $\bar{X}$ = 75 puan değerini elde etme olasılığı %33,3'tür bilgisi üretilebilir.

Örnekleme girecek birimlerin seçimi iadeli yapılmış olsaydı, N=4 birimden, n=2 birimlik Tablo 1.5'teki $N^n = 4^2 = 16$ farklı örneklem oluşturulurdu. 16 farklı örneklem için yukarıdaki işlemler yapılırsa ların örnekleme dağılımı oluşturulabilir. Bu durumda da

$\mu = \mu_{\bar{x}} = 75$ puan

olduğu görülebilir.

	A	B	C	D
A	A, A	A, B	A, C	A, D
B	B, A	B, B	B, C	B, D
C	C, A	C, B	C, C	C, D
D	D, A	D, B	D, C	D, D

Tablo 1.5
N=4 birimlik evrenden iadeli seçimle oluşturulabilecek mümkün örneklemler.

Özetle;

- Her örneklem hacmi için bir istatistiğe ilişkin örnekleme dağılımı olduğu düşünülür.
- Örnekleme birim seçimi iadeli de yapılsa iadesiz de yapılsa hesaplanan örneklem ortalamalarının dağılımı evren ortalamasına eşit olur.

Ancak, hiçbir araştırmada istatistiksel çıkarım amacıyla yukarıda açıkladığımız işlemler yapılmaz. Bunun yerine, n hacimli tek bir örneklem seçilir, bunun X ortalaması hesaplanır ve örnekleme dağılımı ile ilgili yukarıda açıklanan bilgilerden yararlanılarak μ hakkında çıkarım yapılır.

$\bar{X}$'nın Dağılımının Özellikleri

$\bar{X}$ rassal değişkeninin örnekleme dağılımının özellikleri, bu dağılımın ortalaması μ (evren ortalaması, $\bar{X}$ nın örnekleme dağılımının ortalaması $\mu_{\bar{x}}$ ya eşit olduğu için $\mu_{\bar{x}}$ yerine μ kullanılmıştır) ve standart sapması $\sigma_{\bar{x}}$ (standart hata) ile açıklanır. Standart hatanın karesi ise varyans olarak isimlendirilir ve $\sigma^2_{\bar{x}}$ simgesi ile gösterilir.

$\bar{X}$'nın Dağılımının Ortalaması

$\bar{X}$ nın örnekleme dağılımının ortalaması veya aynı anlama gelen, $\bar{X}$ nın beklenen değeri $E(\bar{X})$ şeklinde gösterilirse;

$$E(\bar{X}) = \mu$$

yazılabilir. Bu sonuca göre $\bar{X}$ nın örnekleme dağılımının ortalaması ile ilgili aşağıdaki değerlendirmeler yapılabilir;

- Örneklem hacmi n arttıkça $\bar{X}$ nın örnekleme dağılımının ortalaması evren ortalamasına yaklaşır. Örneklem hacmi yeterli büyüklüğe ulaştığında $\bar{X}$ nın örnekleme dağılımı normal olur.
- Evrenin dağılım şekli çarpık bir dağılım gösterse bile, örneklem hacmi arttıkça $\bar{X}$ nın dağılımı normal dağılıma yaklaşır.

Örneğin, bankadaki işgörenlerin girdiği sınavla ilgili örnek ele alındığında ve rassal olarak seçilen n=2 hacimli örneklemdeki birimler A ve D birimleri olduğunda örneklem ortalaması;

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{\sum_{i=1}^{2} X_i}{2} = \frac{90+70}{2} = 80 \text{ puan}$$

olarak hesaplanır ve μ nün tahmini

$E(\bar{X}) = \mu = 80$ puan şeklinde yazılabilir.

$\bar{X}$'nın Dağılımının Standart Hatası

$\bar{X}$ nın standart sapması veya aynı anlama gelecek şekilde, $\bar{X}$ nın örnekleme dağılımının standart hatası $\sigma_{\bar{x}}$ simgesiyle gösterilir ve standart hata olarak da isimlendirilir.

Standart hata $\sigma_{\bar{x}}$ ortalamanın örnekleme dağılımının değişkenliğini gösterir. Yani, mümkün örneklem ortalamalarının ($\bar{X}_i$ lar) evren ortalamasından farklarının ($\bar{X}_i - \mu$ = hata) ortalama ölçüsüdür. Standart hatanın karesi ($\sigma^2_{\bar{x}}$), $\bar{X}$ nın dağılımının varyansını ifade eder.

Evren, sonsuz bir evren ise veya evren sonlu fakat basit rassal örneklemede birim seçimi iadeli seçimle yapılıyorsa veya örneklem hacmi n ≥ 30 birim veya örnekleme oranı $\frac{n}{N} < 0,05$ ise $\sigma_{\bar{x}}$ ve $\sigma^2_{\bar{x}}$ ve aşağıdaki eşitlikler yardımıyla hesaplanır:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \; , \; \sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$$

Bu eşitliklerden $\sigma_{\bar{x}}$ nın özellikleri ile ilgili aşağıdaki değerlendirmeler yapılabilir.

- Standart hata evren standart sapması σ ya ve örneklem hacmi n'e bağlıdır. Bir başka ifadeyle evren değişkenliği σ büyük ise herhangi bir örneklem hacmi için $\sigma_{\bar{x}}$ da büyük olur.
- Örneklem hacmi arttıkça, örneklem istatistiğinden yararlanarak μ hakkında daha az hatalı, daha güvenilir bilgi üretmek mümkün olur.

- Örneklem hacminin kare kökü ile $\sigma_{\bar{x}}$ arasında ters yönde ilişki vardır. Yani örneklem hacmini artırdıkça $\sigma_{\bar{x}}$ küçülür. Ancak örneklem hacmini artırarak $\sigma_{\bar{x}}$ yı düşürmeye çalışmak örnekleme başvurmayı gerekli kılan nedenlerden dolayı bazı güçlüklere yol açar. Örneğin n=100 birim iken $\bar{X}$ standart sapmasını yarıya indirebilmek için örneklem hacmi 4 kat artırılmalıdır.

Evren standart sapması genellikle bilinmediğinden $\sigma_{\bar{x}}$ hesaplanırken σ yerine onun yansız bir taminleyicisi olan örneklem standart sapması s kullanılır. Bu durumda standart hata $s_{\bar{x}}$ simgesiyle gösterilir ve $s_{\bar{x}} = \frac{s}{\sqrt{n}}$ eşitliğiyle hesaplanır.

Örneklem standart sapması,

$$s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}}$$

şeklinde hesaplanır.

Eğer ilgilenilen evren sonlu bir evren ve örnekleme oranı n/N ≥ 0,05 ise standart hata hesaplanırken $\sqrt{\frac{N-n}{N-1}}$ şeklindeki bir çarpan, düzeltme faktörü olarak kullanılır ve standart hata hesaplanması

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

veya

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

eşitlikleriyle yapılır.

Basit rassal örneklemede örneklem hacmi arttıkça $\bar{X}$ nın örnekleme dağılımı normal dağılıma yaklaşır. Bu sonuca, istatistikte önemli bir yeri olan aşağıdaki teorem yardımıyla ulaşılır:

Merkezî Limit Teoremi

Evrenin dağılım şekli ne olursa olsun, basit rassal örneklem hacmi büyüdükçe $\bar{X}$ nın örneklem dağılımı normal dağılıma yaklaşır. Bu dağılımın ortalaması μ (mu), varyansı σ^2 / n dir. Örneklem hacmi n için yeterli büyüklük, kesin olmamakla birlikte uygulamada n ≥ 30 birim olarak kabul edilmektedir.

Eğer $\bar{X}$ ortalaması μ ve varyansı σ^2 olan normal dağılımlı bir evrenden seçilmiş n hacimlik basit bir rassal örneklemin ortalaması ise $\bar{X}$ nın örnekleme dağılımı ortalaması μ, varyansı σ^2 / n olan bir normal dağılımdır.

Bir rassal değişken olan $\bar{X}$ nın dağılımı normal olduğunda bu değişkenin aldığı $\bar{X}_i$ değerleri,

$$z_i = \frac{\bar{X}_i - \mu}{\sigma / \sqrt{n}}$$

eşitliğiyle standartlaştırılır. Böylece, normal dağılımın özellikleri kullanılarak örneklem aritmetik ortalamasından evren aritmetik ortalaması μ hakkında bilgi üretmek kolaylaşır.

Normal dağılan bir evrenden, rassal olarak seçilebilecek birbirinden farklı n<30 birimlik mümkün bütün örneklemlerin seçildiğini, her örneklem için $(\bar{X}_i)$ ları ve onların $\frac{\bar{x}-\mu}{s_{\bar{x}}}$ standart değerlerini hesaplandığını düşünelim. Değerler aralığı $-\infty < \left(\frac{\bar{x}-\mu}{s_{\bar{x}}}\right) < \infty$ olan istatistiğin dağılımı (n-1) serbestlik derecesi (sd = n-1) ile t dağılımı adı verilen sürekli bir dağılım gösterir ve bu istatistik

$$t = \frac{\bar{x}-\mu}{s_{\bar{x}}} \quad Burada, s_{\bar{x}} = \frac{s}{\sqrt{n-1}}$$

şeklinde hesaplanır.

Serbestlik derecesi s.d = n - k eşitliği ile ifade edilir ve örneklem hacminden tahminlenecek parametre sayısı (k) arasındaki farkı gösterir. Burada μ tahminleneceği için k = 1 alınır ve s.d = n - 1 yazılır.

t dağılımı ortalaması sıfır olan tek modlu ve simetrik bir dağılımdır. Dağılımın şekli standart normal dağılıma benzer fakat değişkenliği daha büyüktür. Bu değişkenlik serbestlik derecesi ile ters orantılıdır. Örneklem hacmi artarken (sd = n - 1) büyür, t değerinin hesaplanmasında $s_{\bar{x}}$ nın kullanılması nedeniyle ortaya çıkan değişkenlik küçülür ve t dağılımı standart normal dağılım (z dağılımına) yaklaşır. t örnekleme dağılımının özelliklerinden yararlanarak evren ortalaması μ ile ilgili bilgilerin nasıl üretileceği de izleyen ünitede örneklerle açıklanacaktır.

ÖRNEK 21 *Otomobil lastiği üreticisi bir fabrikanın yöneticisi ürettikleri lastiklerin ortalama ömrünü lastiklerin katettiği km olarak tahmin etmek istiyor. Bu amaçla rassal olarak 100 lastik seçilmiş ve bu lastiklerin ortalama ömrünün $\bar{X}$ = 40.000 km ve standart sapmasının s=1.500 km olduğunu tespit etmiştir. Yönetim, ürettikleri lastiklerin 35.000 km ömürlü olmasını planlamıştır.*

Bu bilgileri kullanarak;

- $\bar{X}$ nın örnekleme dağılımının ortalaması nedir? Hesaplayınız.
- İstenen tahminleme yapılırken işlenebilecek hata nedir? Hesaplayınız.
- $\bar{X}$ nın standart z değerini hesaplayınız.

Çözüm:

- $E(\bar{X}) = \mu = 30000$ km
- n=100 lastik *($n \geq 30$ birim)* olduğu için standart hata (evren standart sapması bilinmediği için)

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{\mathbf{1500}}{\sqrt{\mathbf{100}}} = \mathbf{150\,km}$$

hesaplanır. Üretilen lastiklerin tümünün ömrünü yukarıdaki verilere göre tahminlerken işlenebilecek hata düzeyi 150 km dir bilgisi elde edilebilir.

$$Z = \frac{\bar{x}-\mu}{s_{\bar{x}}} = \frac{\mathbf{40000-35000}}{\mathbf{150}} = \frac{\mathbf{5000}}{\mathbf{150}} = \mathbf{33{,}3}$$

Örneklem Oranı p'nın Örnekleme Dağılımı

Örnekleme planlarında ele alınan evrenin araştırılmak istenen özelliklerinin bazıları iki sonuçlu olmaktadır. Örneğin bir fabrikada üretilen ürünler, hatalı ya da hatasız ürün, bir fakültedeki öğrenciler, başarılı ya da başarısız öğrenci olmak üzere iki grupta toplanabilir. Bu iki sonuçtan birinde örneğin A sonucunda başarılı öğrenci yer alan birimlerin oranıyla ilgilenilebilir. Bu durumda evren oranı evrenin birimleri içindeki ilgilenilen türden özelliğe sahip olanların oranı biçiminde tanımlanır.

ÖRNEK 22

Y sınıfındaki öğrencilerin genel başarı durumu aşağıda verilmiştir. Bu sınıfın başarılı öğrenci oranı nedir?

ÖĞRENCİ ADI	BAŞARI DURUMU
A	Başarılı
B	Başarısız
C	Başarılı
D	Başarılı

Örnekte sınıftaki başarılı öğrenci oranı, evren oranıdır ve π ile gösterilir. Bu evrendeki ilgilenilen türden özelliğe sahip (başarılı) birim (öğrenci) sayısı R ile gösterilirse evren oranı π,

$$\pi = \frac{R}{N}$$

Eşitliği ile hesaplanır. Burada R = 0, 1, 2,..., N değerlerini alabileceği için π nin değer aralığı $0 \leq \pi \leq 1$ olur. Sınıftaki başarısız öğrenci sayısı (ilgilenilmeyen türden özelliğe sahip birim sayısı) N-R olduğu için başarısız öğrenci oranı Q,

$$Q = \frac{N-R}{N} = \mathbf{1} - \pi$$

olur.

Yukarıdaki örnekte başarılı öğrenci sayısı, R=3 olduğu için

$$\pi = \frac{\mathbf{3}}{\mathbf{4}} = \mathbf{0{,}75}$$

olarak bulunur. Bu sonuca göre sınıftaki öğrencilerin %75'i başarılıdır. Bu kesin bir sonuçtur.

Tamsayım yapılamadığı zaman R bilinemez ve π hesaplanamaz. Örnekleme planlarında π parametresi hakkında bilgi, bu parametre hakkında bilgi üreten örneklem istatistiklerinden yararlanılarak üretilebilir.

Hacmi n olan bir basit rassal örneklemden, bu örneklemin seçildiği evrenin π parametresi hakkında bilgi üretebilmek için iki örneklem istatistiği söz konusudur. Birincisi, hacmi n olan bir basit rassal örneklemdeki ilgilenilen türden özelliğe sahip olan birimlerin sayısıdır ve r ile gösterilir. İkincisi, ilgilenilen türden özelliğe sahip olan örneklemdeki birimlerin oranıdır. Örneklem oranı p simgesiyle gösterilir ve

$$\mathbf{p} = \frac{r}{n}$$

eşitliği ile hesaplanır. Burada r = 0, 1, 2,..., n değerlerini alabilir. r'nin değerlerine bağlı olarak p de $0 \leq p \leq 1$ aralığında bir değer alır. Örneklemi oluşturan birimler arasında ilgilenilen türden sonuca sahip olmayan birimlerin oranıysa q ile gösterilir. Bu sonuca sahip birimlerin sayısı n- r olduğu için,

$$q = \frac{n-r}{n} = \mathbf{1} - p$$

olur.

Yukarıda verilen örnekte ele alınan N=4 birimlik bir evrenden basit rassal örneklemeyle hacmi n=2 olan bir örneklem seçildiğinde ve örneklemdeki birimler öğrenci B ve C olduğunda başarılı öğrenci oranı,

$$p = \frac{r}{n} = \frac{\mathbf{1}}{\mathbf{2}} = \mathbf{0{,}5}$$

olarak hesaplanmış olur.

Öte yandan, evren oranına ilişkin varyans,

$$\sigma^2 = \pi\ (1-\pi)$$

şeklinde ifade edilir. Varyansın karekökü de standart sapmayı verdiğinden,

$$\sigma = \sqrt{\pi(\mathbf{1}-\pi)}$$

şeklinde yazılır.

Örneklem varyansı ve standart sapması da benzer şekilde sırasıyla

$$s^2 = p\ (1-p)$$

ve

$$s = \sqrt{p(\mathbf{1}-p)}$$

olarak gösterilir.

İki sonuçlu bir evrenden, mümkün bütün n hacimli basit rassal örneklemlerin seçildiğini ve her örneklem için p oranının hesaplandığı varsayıldığında p_i oranlarından oluşan bir dağılım elde edilir. Bir evrenden seçilebilecek aynı hacimli mümkün bütün örneklemler için hesaplanan örneklem oranlarının oluşturduğu dağılıma oranların örnekleme dağılımı adı verilir.

Örnekleme planlarında tanımlanan evrenden rassal olarak n hacimli sadece tek bir örneklem oluşturulur ve bu örneklem için p oranı hesaplanır. P rassal değişkeninin çekilmesi mümkün bütün n hacimli örneklemlerde aldığı değerlerin dağılımına “oranların örnekleme dağılımı” adı verilir.

Ortalama ve Varyans

Evren oranı π hakkında araştırılmak istenin bilgi n hacimli tek bir örneklem için hesaplanan p istatistiğine değil, bir rassal değişken olan p istatistiğinin örnekleme dağılımının özelliklerinden yararlanılarak üretilir. Bu dağılımın özellikleri dağılımın aritmetik ortalaması ve varyansıyla belirlenebilir.

Sonsuz bir evrenden seçilen n hacimli basit rassal örneklem için hesaplanan p oranının örnekleme dağılımının aritmetik ortalaması μ_p evren oranı π ye eşittir. Bu durum örneklem oranı p’nin evren oranı π nin yansız (sistematik hata içermeyen) tahminleyicisi olduğunu göserir. Bu sonuca göre,

$$E(p) = \pi$$

yazılır.

Sonsuz evren ya da örnekleme oranı n/N < 0,05 olan bütün sonlu evrenlere uygulanan basit rassal örnekleme planlarında örneklem oranı p nin dağılımının varyansı σ_p^2 ve standart hatası da σ_p ile gösterilir.

Eğer evren varyansı σ^2 biliniyorsa,

$$\sigma_p^2 = \frac{\pi(1-\pi)}{n}, \qquad \sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

Eşitlikleriyle, evren varyansı σ^2 bilinmiyorsa,

$$s_p^2 = \frac{p(1-p)}{n}, \qquad s = \sqrt{\frac{p(1-p)}{n}}$$

Eşitlikleriyle hesaplanır.

Dağılım Şekli ve Merkezî Limit Teoremi

Oranların örnekleme dağılımının şekli eğer E(p)=π< 0,5 ise sağa çarpık, E(p)=π>0,5 ise sola çarpık ve E(p) = π = 0,5 ise simetrik bir dağılım gösterir. Kolaylıkla görülebileceği gibi E(p) = π nin değeri 0 ve 1 e yaklaşırken dağılımın çarpıklığı artar.

Merkezî limit teoremine göre bir örnekleme planında seçilen basit rassal örneklemin hacmi n büyürken örneklem oranı p'nin örnekleme dağılımı normal dağılıma yaklaşır. Uygulamada n.$\pi \geq 5$ ve n(1-π)$\geq$5 koşullarını birlikte sağlayan örneklem büyüklüğü, yeterli örneklem büyüklüğü olarak kabul edilir. Aynı teoreme göre rassal örneklem hacmi $n \geq 30$ birim olması ve evren oranı π nin 0 ya da 1 e yakın değerler almaması koşuluyla oranların örnekleme dağılımı normal dağılıma yaklaşır. Bu koşulları sağlayan oranların örnekleme dağılımıyla ilgili problemlerin çözümlerinde normal dağılımın özelliklerinden yararlanılır. Bu amaçla p rassal değişkeninin standartlaştırılmış Z değişkeni

$$\mathbf{Z} = \frac{p-\pi}{\sigma_p}$$

şeklinde yazılır. Bu standart değişken kullanılarak evren oranı hakkında bilgi üretmek mümkün olur.

SIRA SİZDE 5

- **Örnekleme dağılımı kavramını açıklayınız.**
- **Merkezî limit teoremi istatistiğe ne tür kolaylıklar getirmiştir, açıklayınız.**

ÖRNEK 23

Bir bankada ayda 8.000 işlem yapılıp fiş düzenlenmektedir. Banka yönetimi ilgili ayda düzenlenen fişler içindeki hatalı işlem oranını tahminlemek amacıyla 100 fişi rassal olarak seçiyor. Seçilen fişlerin 15 tanesinin hatalı düzenlendiği belirlenmiştir. Bu verileri kullanarak istenen tahminleme yapılırken işlenecek hata düzeyi nedir?

Çözüm:

Bilindiği gibi tahminleme yaparken işlenebilecek hata düzeyini belirleme imkânı veren istatistik standart hatadır. Bu;

$$s_p = \sqrt{\frac{p(1-p)}{n}}$$

eşitliği ile hesaplanır. Burada;

$$p = \frac{r}{n} = \frac{15}{100} = 0{,}15 \; ve \; q = 1 - p = 1 - 0{,}15 = 0{,}85$$

$$s_p = \sqrt{\frac{0{,}15.0{,}85}{100}} = \sqrt{\frac{0{,}127}{100}} = 0{,}032$$

Bu bilgiye göre, bu bankada ilgili ayda düzenlenen fişler arasındaki hatalı fiş oranını tahminlerken işlenecek hata düzeyi s_p = 0,032 fiş olacaktır.

İki Örneklem İstatistiği Arasındaki Farkın Örnekleme Dağılımı

Tanımlanan N_1 ve N_2 hacimli iki ayrı evrenden birbirinden bağımsız ve rassal sırasıyla n_1 ve n_2 hacimli farklı mümkün örneklemler seçilir, bu örneklemler için $\{\hat{\theta}_{1i}\}$ ve $\{\hat{\theta}_{2i}\}$ istatistiklerinden oluşan iki ayrı örneklem istatistiği dağılımı elde edilir ve bu iki dağılımın hesaplanan istatistik değerlerinin ikişerli kombinasyonları için farklar hesaplanırsa meydana gelen dağılıma iki örneklem istatistiği arasındaki farkın örnekleme dağılımı adı verilir.

Aşağıda iki örneklem ortalaması ($\bar{X}_1$ - $\bar{X}_2$) ve iki örneklem oranı (p_1-p_2) arasındaki farkın örnekleme dağılımıyla ilgili bilgiler verilmiştir.

($\bar{X}_1$ - $\bar{X}_2$)'nin Örnekleme Dağılımı

Ortalamaları μ_1 ve μ_2 olan iki ayrı evrenin ortalamaları arasındaki (μ_1- μ_2) farka ilişkin istatistiksel çıkarım yapabilmek için bu parametre hakkında bilgi üreten ($\bar{X}_1$ - $\bar{X}_2$) istatistiğine ve bu istatistiğin örnekleme dağılımının özellikleriyle ilgili bilgilere gereksinim vardır.

Tanımlanan N_1 ve N_2 hacimli iki ayrı evrenden rassal ve birbirinden bağımsız olarak n_1 ve n_2 hacimli mümkün örneklemler için oluşturulabileceği varsayılan $\{\bar{X}_{1i}\}$ ve $\{\bar{X}_{2i}\}$ örnekleme dağılımların oluşturulan ortalamaların ikişerli kombinasyonları için farkların hesaplanacağı düşünülürse bu farkların meydana getireceği teorik dağılıma ortalama farkların örnekleme dağılımı adı verilir. Hesaplanan bu farklar, kombinasyondan kombinasyona değişik değerler aldığı için herhangi iki ortalama arasındaki farkın ($\bar{X}_1$ - $\bar{X}_2$) bir rassal değişkenin gerçekleşen değeri olduğu varsayılır.

($\bar{X}_1$ - $\bar{X}_2$)'nin Örnekleme Dağılımının Özellikleri

Diğer örnekleme dağılımlarında olduğu gibi ($\bar{X}_1$ - $\bar{X}_2$) nın örnekleme dağılımının özellikleri de bu dağılımın ortalaması $\mu_{\bar{x}_1-\bar{x}_2}$ ve standart hatası $\sigma_{\bar{x}_1-\bar{x}_2}$ veya varyansı $\sigma^2_{\bar{x}_1-\bar{x}_2}$ ile açıklanabilir.

($\bar{X}_1$ - $\bar{X}_2$)'nin Örnekleme Dağılımının Ortalaması

$\bar{X}_1$ ve $\bar{X}_2$'nın birer rassal değişken olduğu ayrıca birbirinden bağımsız iki rassal değişkenin ortalamaları arasındaki farkın da bir rassal değişken olduğu açıklanmıştı. Bu bilgilere göre farklı evrenlerden birbirinden bağımsız ve rassal olarak seçilen n_1 ve n_2 hacimli rassal örneklemlerin $\bar{X}_1$ ve $\bar{X}_2$ ortalamaları arasındaki ($\bar{X}_1$ - $\bar{X}_2$) farkın beklenen değeri iki evren ortalaması arasındaki (μ_1- μ_2) farkına eşittir. Bu durum aşağıdaki gibi gösterilir:

$$E(X_1) - E(X_2) = E(X_1 - X_2) = \mu_{x_1 - x_2} = (\mu_1 - \mu_2)$$

($\bar{X}_1$ - $\bar{X}_2$)'nin Örnekleme Dağılımının Standart Hatası

($\bar{X}_1$ - $\bar{X}_2$) nın standart hatası bu istatiastiğin örnekleme dağılımının değişkenliğini gösteren bir ölçüdür. Standart hata adı verilen bu ölçü tanımlanan evrenin standart sapmaları σ_1 ve σ_2 biliniyorsa bu çalışmada $\sigma_{\bar{x}_1-\bar{x}_2}$, bilinmiyorsa $s_{\bar{x}_1-\bar{x}_2}$ simgesiyle gösterilecektir.

Tanımlanan her iki evrenin ilgilenilen bir değişken itibarıyla aldığı değerlerin dağılımı normal ise veya her iki evrenden $n_1 \geq 30$ ve $n_2 \geq 30$ birim olmak üzere birbirinden bağımsız örneklemler seçilmiş ise ($\bar{X}_1$ - $\bar{X}_2$)'nın örnekleme dağılımı normal olur ve standart hatası;

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

veya

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

eşitlikleriyle hesaplanır. Hangi eşitlikle hesaplanırsa hesaplansın, hesaplanan değer değişkenin birim türüyle ifade edilir. Burada;

σ_1 : Birinci evrenin standart sapması
σ_2 : İkinci evrenin standart sapması
s_1 : Birinci evrenden seçilen örneklemin standart sapması
s_2 : İkinci evrenden seçilen örneklemin standart sapması
n_1 : Birinci evrenden seçilen örneklemin hacmi
n_2 : İkinci evrenden seçilen örneklemin hacmi

($\bar{X}_1$ - $\bar{X}_2$) nın dağılımının normal olması durumunda bu dağılımın $Z = \frac{(\bar{x}_1 - \bar{x}_2)}{s_{\bar{x}_1-\bar{x}_2}}$ standart değerlerinin dağılımı olan standart normal dağılımın özelliklerine sahip olur ve bu özelliklerden yararlanılarak (μ_1- μ_2) parametresiyle ilgili çıkarımlar kolayca yapılır. Yukarıda yapılan bütün açıklamalar sonsuz evrenler için de geçerlidir. Bu açıklamaların ışığında yapılacak istatistiksel çıkarımlar hakkında bilgiler ikinci ve üçüncü ünitelerde açıklanacaktır.

Ortalamaları μ_1 ve μ_2 olan normal dağılıma sahip evrenlerden sırasıyla n_1<30 ve n_2<30 birim olmak üzere bağımsız rassal örneklemler seçilmiş ise ($\bar{X}_1$ - $\bar{X}_2$) nın örnekleme dağılımının ortalaması yine

$$E(X_1) - E(X_2) = E(X_1 - X_2) = \mu_{x_1-x_2} = (\mu_1 - \mu_2)$$

olur. Bu özellikleri taşıyan dağılımın standart hatası ise

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

eşitliği ile hesaplanır.

Örneklem hacimleri $n_1<30$ ve $n_2<30$ birim olması durumunda $(\mu_1 - \mu_2)$ için çıkarım yapmak amacı olduğunda yukarıda belirtilen z istatistiğinin dağılımından değil;

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{s_{\bar{x}_1 - \bar{x}_2}}$$

istatistiğinden yararlanılır. Burada serbestlik derecesi n_1+n_2-2 kullanılır. Çünkü $(n_1-1) + (n_2-1) = n_1+n_2-2$ olur.

ÖRNEK 24

İstatistik dersini aynı öğretim üyesinden alan A ve B bölümlerinin öğrencileri arasından rassal olarak sırasıyla 40 ve 50 öğrenci rassal olarak seçilmiştir. A bölümünden seçilen öğrencilerin ortalama başarı puanı 75, standart sapması 10; B bölümünden seçilen öğrencilerin ise ortalaması 65, standart sapması 12 puan olarak hesaplanmıştır. İki bölümün öğrencilerinin ortalama başarıları arasındaki farkın standart hatası nedir?

Çözüm:

$n_1 = 40$ $\bar{X}_1 = 75$ puan $s_1 = 10$ puan

$n_2 = 50$ $\bar{X}_2 = 65$ puan $s_2 = 12$ puan

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{100}{40} + \frac{144}{50}} = \sqrt{2{,}5 + 2{,}88} = 2{,}32$$

(p_1-p_2)'nin Örnekleme Dağılımı

N_1 ve N_2 hacimli iki farklı evrenden rassal ve birbirinden bağımsız n_1 ve n_2 hacimli mümkün örneklemler için hesaplanan istatistiklerin sırasıyla $\{p_{1i}\}$ ve $\{p_{2i}\}$ örnekleme dağılımları oluşturulsun. Bu iki dağılımın hesaplanan istatistiklerinin ikişerli kombinasyonları arasındaki farkların teorik dağılımına $(p_1 - p_2)$'nin örnekleme dağılımı adı verilir. Bu ünitede ve izleyen iki ünitede $(p_1 - p_2)$'nin örnekleme dağılımıyla ilgili bilgiler $n_1 \geq 30$ ve $n_2 \geq 30$ olması ve örneklemlerin rassal ve bağımsız seçilmesi durumları için geçerli olacaktır.

Tanımlanan iki evrenden büyük hacimli bağımsız rassal örneklemler seçilmiş ise $(p_1 - p_2)$'nin örnekleme dağılımı yaklaşık normal dağılım özelliğine sahiptir. Bu durumda $(p_1 - p_2)$'nin örnekleme dağılımının özellikleri aşağıdaki gibi açıklanabilir:

$(p_1 - p_2)$'nin örnekleme dağılımının ortalaması $\mu_{p_1 - p_2}$ simgesiyle gösterilir ve

$$E\,(p_1 - p_2) = \mu_{p_1 - p_2} = \pi_1 - \pi_2$$

yazılır. Yani $(p_1 - p_2)$ nin örnekleme dağılımının beklenen değeri $\pi_1 - \pi_2$ ye eşit alınır.

$(p_1 - p_2)$'nin örnekleme dağılımının standart hatası;

$$s_{p_1-p_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$ Evren oranları π_1 ve π_2 bilinmiyorsa

veya

$$\sigma_{p_1-p_2} = \sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}$$ Evren oranları π_1 ve π_2 biliniyorsa

eşitlikleriyle hesaplanır.

(p_1 - p_2) nin örnekleme dağılımı yaklaşık normal dağılımı özelliği taşıdığı durumda da π_1 - π_2 parametresiyle ilgili çıkarımlar için z istatistiğinden yararlanılabilir. Bu istatistik;

$$Z = \frac{p_1 - p_2}{s_{p_1-p_2}}$$

eşitliği ile hesaplanır.

ÖRNEK 25

İki farklı sosyal sınıfta bir reklam filmini beğenenlerin oranları arasında fark olup olmadığı araştırılmak isteniyor. Bu amaçla her iki sınıftan reklam filmini izleyenler arasından sırasıyla n1 = 400, n2 = 300 kişi rassal olarak seçiliyor. Birinci sosyal sınıftan seçilen kişilerin 220 si, ikinci sosyal sınıftan seçilen kişilerin 195 i reklamı beğendiklerini ifade etmişlerdir.

a. $\pi_1 - \pi_2$ *nedir?*

b. $p_1 - p_2$*'nin örnekleme dağılımının standart hatası nedir?*

Çözüm:

n_1 = 400 r_1 = 220 $$p_1 = \frac{r_1}{n_1} = \frac{220}{400} = 0{,}55$$

n_2 = 300 r_2 = 195 $$p_2 = \frac{r_2}{n_2} = \frac{195}{300} = 0{,}65$$

$$s_{p_1-p_2} = \sqrt{\frac{0{,}65(1-0{,}65)}{300} + \frac{0{,}55(1-0{,}55)}{400}} \cong 0{,}037$$

E (p_1 - p_2) = (0,65 - 0,55) = 0,10 = π_1 - π_2

Özet

AMAÇ 1

Tamsayım ve örnekleme kavramlarını ayırt etmek

Bir araştırmada tanımlanan evren sonlu evren ise gerekli bilgilerin üretilebilmesi için tamsayım yapılabilir veya örneklemeye başvurulur. Araştırma için gerekli zamana, ekonomik imkânlara ve araçlara sahip olunduğunda tamsayım yapılmalıdır. Çünkü tamsayım sonucu elde edilen verileri kullanarak hesaplanan bilgiler kesin bilgilerdir. Tanımlanan evren sonsuz evren ise örnekleme zorunludur. Tamsayım uygulamasının imkânsız, örneklemeye başvurmanın gerekli olduğu durumlarda örneklemeye başvurmak kaçınılmazdır.

AMAÇ 2

Örneklemeye başvurmanın yararlarını açıklamak

Örneklemeye başvurulduğunda araştırmacı zaman ve ekonomik tasarruf sağlar. Ayrıca tamsayım yapmayı engelleyen diğer nedenlerin varlığında örnekleme araştırma yapmaya imkan verir.

AMAÇ 3

Örnekleme planı hazırlamak

Örnekleme planı 5 aşamalı bir süreçtir. Bu aşamalarda araştırmacı araştırma yapacağı evreni tanımlar. Bu evren sonlu evren olduğunda evrenle ilgili güncel bir çerçeve hazırlar veya belirli bir kaynaktan nasıl temin edilebileceğini belirler. Daha sonra olasılıklı ve olasılıklı olmayan örnekleme yöntemlerinden hangisini temsilî bir örneklem oluşturma amacıyla araştırmasında kullanacağına karar verir. Seçilecek örneklem hacminin ne olacağını belirler ve örnekleme uygulaması sonucu oluşturduğu örneklemdeki birimler üzerinden araştırmaya konu olan değişkenler itibarıyla veriler derlenir.

AMAÇ 4

Bir örnek araştırma için örnekleme uygulaması yapmak ve istenen bilgileri üretmek

Örneklemden derlenen veriler için araştırmada istenen bilgileri üreten istatistikler; örneklem aritmetik ortalaması, örneklem oranı vb. gibi istatistikler hesaplanır. Bu istatistikler ve bu istatistiklerle ilgili dağılımın özellikleri kullanılarak bu istatistiklerin bilgi ürettiği parametreler; evren aritmetik ortalaması μ, evren oranı π için gerekli çıkarım bilgileri üretilebilir.

Kendimizi Sınayalım

1. Bir örneklemin özelliklerine ilişkin değerlere ne ad verilir?

a. Örnek istatistiği
b. Evren
c. Anlamlı fark
d. Örnekleme
e. Parametre

2. Bir evrenden rastgele seçilen birden fazla örneğin sonuçlarının birbirinden farklı olduğu gözlenmiştir. Bu farklılığın nedeni aşağıdakilerden hangisidir?

a. Beklenen frekans
b. Yöntem farklılığı
c. Örnek değişkenliği
d. Örnek istatistiği
e. Parametre

3. İki sonuçlu bir evrenden, mümkün n hacimli basit rassal örneklemlerin seçildiğini ve her iki örneklem için p oranı hesaplandığı varsayımı altında, lerin dağılımına (i=1, 2,..., n) ne ad verilir?

a. Oranların örnekleme dağılımı
b. Örneklem istatistiği
c. Binom dağılımı
d. Normal dağılım
e. Ortalamaların örnekleme dağılımı

4. Aşağıdakilerden hangisi olasılıklı örnekleme yöntemlerinden biri **değildir?**

a. Tabakalı örnekleme
b. Basit rassal örnekleme
c. Kartopu örneklemesi
d. Küme örneklemesi
e. Sistematik örnekleme

5. X rassal değişkeni, ortalaması (μ) 50 ve standart sapması (σ) 10 olmak üzere normal dağılmıştır. Buna göre, hacmi n=100 olan örneklemin ortalaması $\bar{X}$ =55 değerinin standart normal değeri (z) kaçtır?

a. 0,5
b. 1
c. 1,5
d. 2
e. 5

6. Örneklem oranının değer aralığı aşağıdakilerden hangisidir?

a. $0 \leq \pi \leq 1$
b. $-1 < \pi < 0$
c. $-1 < \pi < 1$
d. $\pi < -1$
e. $\pi > 1$

7. Bir örneklemin gözlem değerleri için hesaplanan karakteristik değerlere ne ad verilir?

a. Ortalama
b. İstatistik
c. Frekans
d. Anlamlı fark
e. Parametre

8. Aşağıdakilerden hangisi, tamsayım yapmayı engelleyen nedenlerden biri **değildir?**

a. Maliyet
b. Ölçüm için birimlerin tahrip edilmesi olasılığı
c. Evren hacminin küçük olması
d. Zaman
e. Evren hacminin sonsuz sayıda olması

9. Tabaka hacimleri sırasıyla 50, 250 ve 200 birimden oluşan bir evrenden kota örneklemesiyle 50 birimlik örneklem oluşturulmak istenmektedir. Bu örneklem hacmi tabaka hacimlerinin evren içindeki paylarıyla orantılı olarak dağıtılırsa üçüncü sıradaki tabakadan seçilecek birim sayısı kaç olmalıdır?

a. 4
b. 15
c. 20
d. 35
e. 40

10. Evren hacmi küçük olduğunda, örnekleme seçilen bir birimin diğerlerinin seçilme şansını etkilememesi için aşağıdaki seçim yöntemlerinden hangisi kullanılır?

a. Rassal iadeli
b. Rassal iadesiz
c. Sistematik
d. Keyfî ve sistematik birlikte
e. Keyfî

Kendimizi Sınayalım Yanıt Anahtarı

1. a	Yanıtınız yanlış ise “Tamsayım ve Örnekleme” konusunu yeniden gözden geçiriniz.
2. c	Yanıtınız yanlış ise “Olasılıklı Örnekleme ve Örnekleme Dağılımı” konusunu yeniden gözden geçiriniz.
3. a	Yanıtınız yanlış ise “Oranların Örnekleme Dağılımı” konusunu yeniden gözden geçiriniz.
4. c	Yanıtınız yanlış ise “Olasılıklı ve Olasılıklı Olmayan Örnekleme” konusunu yeniden gözden geçiriniz.
5. e	Yanıtınız yanlış ise “Ortalamanın Örnekleme Dağılımının Özelliklerini” konusunu yeniden gözden geçiriniz.
6. a	Yanıtınız yanlış ise “Oranların Örnekleme Dağılımının Özellikleri” konusunu yeniden gözden geçiriniz.
7. b	Yanıtınız yanlış ise “Örneklem ve Örneklemenin Amaçları” konusunu yeniden gözden geçiriniz.
8. c	Yanıtınız yanlış ise “Örnekleme Başvurma Nedenlerini” konusunu yeniden gözden geçiriniz.
9. c	Yanıtınız yanlış ise “Kota Örneklemesi uygulaması” ile ilgili açıklamaları konusunu yeniden gözden geçiriniz.
10. a	Yanıtınız yanlış ise “Keyfî ve Rassal Seçim” konusunu yeniden gözden geçiriniz.

Sıra Sizde Yanıt Anahtarı

Sıra Sizde 1

- Tamsayım yapılamaz. Çünkü kitapçıya gelen müşteriler evreni sonsuz evrendir.
- Örnekleme giren birimlerin rassal olarak seçilmesi ve örneklemin temsil örneklem olması durumunda örneklem istatistikleri evren parametreleri için bilgi niteliğindedir.
- Örneklemin temel amacı temsili örneklem oluşturmaktır.

Sıra Sizde 2

- Tamsayım yapmayı imkânsız kılan nedenler söz konusu olmadıkça ve kesin bilgi istendiği sürece tamsayım yapılır.
- Her ne kadar tamsayım için gerekli koşullar mümkün ise de örneklemeye başvurulur. Çünkü, öğrencilerin bazılarına ulaşmak imkânsız olabilir, ulaşılsa bile bazıları bilgi vermeyebilir.

Sıra Sizde 3

- Yargısal örnekleme daha temsilî örneklem oluşturur. Çünkü, bu örnekleme uygulamasında örnekleme birim seçimini yapacak kişi incelenecek evrenle ilgili temsilî örneklem oluşturma bakımından öncül bilgilere sahiptir.
- Kota örneklemesi seçilir.

Sıra Sizde 4

- Temel fark, olasılıklı olmayan örnekleme yöntemlerinde birim seçimi keyfî yapılırken bu örneklem uygulamaları ile oluşturulan örneklem için hesaplanan istatistiklerin evren parametreleri hakkında genelleme yapmak amacıyla kullanılmamasıdır. Olasılıklı örnekleme uygulandığında birim seçimi rassal yapılır ve örneklem istatistikleri parametreler hakkında genelleme amacıyla kullanılır.

Sıra Sizde 5

- N sonlu bir evrenden rassal olarak n hacimli bir örneklem değil de mümkün aynı hacimli bütün örneklemleri seçtiğimiz ve her örneklem için bir istatistik hesapladığımızda meydana gelen dağılıma örnekleme dağılımı denir.
- Bu teorem herhangi bir örnekleme uygulamasında hesaplanan n hacimli bir örneklem için hesaplanan istatistiğin dağılım şekli ve özellikleri ile ilgili ispatlanmış bilgileri verdiği için istatistiksel çıkarımlar bu teoreme dayandırılmaktadır.

Yararlanılan ve Başvurulabilecek Kaynaklar

Fink, A. (1995). **How To Sampling in Surveys,** London: Sage Publication

Gürsakal, N. (2010). **Betimsel İstatistik,** Bursa: Dora Kitabevi

Hirsch, W. Z. (1963). **Introduction to Modern Statistics,** New York: The Macmillan Company

Malhotra, N. K. (1996). **Marketing Research An Applied Orientation,** New Jersey: Prentice Hall International

Neter, J., Wasserman, W., Whitmore, G. A. (1993). **Applied Statistics,** Boston: Simon and Schuster

Serper, Ö., Aytaç, M. (2000). **Örnekleme,** Bursa: Ezgi Kitabevi

Serper, Ö. (2004). **Uygulamalı İstatistik 2,** Bursa: Ezgi Kitabevi

Trochim, W. M. (2001). **Research Methods Knowledge Base,** Cornell University

Tryfos, P. (1996). **Sampling Methods for Applied Research,** New York: John Wiley and Sons Inc.

2

Amaçlarımız

Bu üniteyi tamamladıktan sonra;

- İstatistiksel tahminleme tanımı yapabilecek ve bazı evren parametreleri için tahminleme sürecinin aşamalarını açıklayabilecek,
- Tek evren aritmetik ortalaması, oranı ve iki evren aritmetik ortalaması, oranına ilişkin nokta ve aralık tahminlemesi uygulamalarını açıklayabilecek

bilgi ve becerilere sahip olacaksınız.

Anahtar Kavramlar

- Evren
- Parametre
- Örneklem
- Örneklem İstatistiği
- Tahminleme, Tahminleyici, Tahmin
- Nokta ve Aralık Tahminlemesi

İçindekiler

İstatistiksel Tahminleme

GİRİŞ

Bu ünitede istatistiksel tahminleme ve istatistiksel tahminleme türleriyle ilgili teorik bilgiler açıklanacak, sonra tek evren aritmetik ortalaması, oranı, iki evren aritmetik ortalaması arasındaki farka ve iki evren oranı arasındaki farka ilişkin parametre tahminlemesi uygulamalarına yer verilecektir.

KİTAP

Bu ünitenin izleyen kısımlarındaki konular için A. F. Yüzer, E. Şıklar, E. Ağaoğlu, H. Tatlıdil, A. Özmen, Editör: A. F. Yüzer (2011). İstatistik, Ünite 8 ve 9, Eskişehir Anadolu Üniversitesi Yayını isimli kitaptan, editör ve ünite yazarlarının izni alınarak yararlanılmıştır.

İSTATİSTİKSEL TAHMİNLEME

Günlük yaşamın hemen hemen her alanında parametre tahminlemesiyle karşılaşılır. Bir araştırma sürecinin en önemli aşaması olan örneklemeyle tahminleme birbirinin ayrılmaz parçasıdır.

Bir tanım vermek gerekirse **tahminleme,** tanımlanan evrenden seçilen rassal örneklemden hesaplanan istatistikler yardımıyla, bu evrenin uyduğu dağılımın parametre değerlerini araştırmaktır.

Bir rassal örneklemden hesaplanan istatistikler yardımıyla, ilgili evrenin parametre değerinin araştırılmasına **tahminleme** denir.

Hem örneklem için hem de evren için bilgi üreten istatistiğe ilişkin formülasyona "tahminleyici", örneklem gözlem değerlerinin bir tahminleyiciye uygulanmasıyla hesaplanan değere ise "tahmin" adı verilir. Tahminleyici, tahminin nasıl hesaplanacağını gösterir. Tahminse sayısal bir değerdir. Örneğin,

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

formülü ile tanımlanan tek örneklem aritmetik ortalaması, evren aritmetik ortalaması için bir tahminleyicidir. Bu formül yardımıyla hesaplanacak değerse evren aritmetik ortalamasının bir tahminidir.

Parametreleri, istatistiklerden hareketle, bir sayı veya bu sayıyı kapsayan bir aralığın sınırlarıyla tahminlemek mümkündür. Birinci durumda nokta tahminlemesi, ikinci durumdaysa aralık tahminlemesi söz konusu olur.

SIRA SİZDE 1

- **İstatistiksel tahminleme nedir?**
- **Tahmin ve tahminleyici kavramları arasındaki farkı açıklayınız.**

İSTATİSTİKSEL TAHMİNLEME TÜRLERİ

İstatistiksel tahminleme, nokta ve aralık tahminlemesi olarak sınıflandırılır.

Nokta Tahminlemesi

Bir rassal örneklemden hesaplanan istatistiğin değerini, ilgili evren parametre değerine eşit kabul eden tahminleme sürecine **nokta tahminlemesi** denilir.

Tek bir rassal örneklemden hesaplanan $\hat{\theta}$ istatistiğinin değerini, bu istatistiğin bilgi ürettiği θ parametresinin değerine eşit kabul eden tahminleme sürecine **"nokta tahminlemesi"** denir. Bu tahminleme bir atıcının hedefe tek bir atış yapması eylemine benzer. Tek bir atışın hedefteki gözlenen sonucuna bakarak, gözlenen sonuç tam isabet olsa bile atıcının iyi olduğunu söylemek doğru olmaz, tekrarlanan atışlardaki gözlenen sonuçlara bakmak gerekir. Bu benzetmeden hareketle nokta tahminlemesi tek bir örneklemden hesaplanan istatistiğin değerine dayanarak değil, bu istatistiğin örnekleme dağılımını inceleyerek yapılır. Bu nedenle nokta tahminlemesi sürecinde önemli olan bir rassal değişken olan $\hat{\theta}$ istatistiğinin dağılımının ortalaması $\mu_{\hat{\theta}}$ nın veya aynı anlama gelen $\hat{\theta}$ istatistiğinin beklenen değeri $E(\hat{\theta})$ nın evren parametre değerine eşit alınmasıdır. Bu özelliği taşıyan tahminleyiciler için nokta tahmini genel olarak,

$$E(\hat{\theta}) = \theta$$

şeklinde gösterilir.

Bir $\hat{\theta}$ istatistiğinin örnekleme dağılımının beklenen değeri tahminlenecek evren parametresine eşit ise $\hat{\theta}$ istatistiği (tahminleyici) θ nın yansız tahminleyicisidir denir ve yukarıdaki eşitlik ile gösterilir. Yapılan açıklamalardan anlaşılabileceği gibi tahminleyicilerin yansız olabilmesi için örneklemin rassal seçimle oluşturulması yani evrendeki her birimin örneklemde yer alması için sıfırdan farklı, hesaplanabilir bir olasılığa sahip olması gerekir.

Bir tahminleyicinin iyi bir tahminleyici olup olmadığını belirlerken kullanılan kriterlerin başında yansızlık kriteri gelir. Bu kritere göre tahminleyicinin örnekleme dağılımı tahminlenecek parametrenin yakınında serpilme gösteriyorsa bu tahminleyici yansızdır denir. Eğer $\hat{\theta}$ parametresi yanlı tahminleyici ise bu gösterim,

$$E(\hat{\theta}) \neq \theta$$

şeklinde yapılır.

Burada yanlılık miktarı yani hata,

$$\text{Hata} = E(\hat{\theta}) - \theta$$

olur.

Yansız tahminleyicilerde önemli olan tahmin hatalarının cebirsel toplamının dolayısıyla ortalamasının sıfıra eşit olmasıdır.

Bir tahminleyicinin iyi bir tahminleyici olup olmadığını belirlerken kullanılan ikinci kriter tahminleyicinin etkinliğidir. Evren parametresine daha yakın örnekleme dağılımına sahip olan tahminleyici diğer tahminleyicilere göre daha etkin tahminleyicidir denir.

Bir yansız tahminleyicinin etkinliği; bu tahminleyicinin örnekleme dağılımının standart hatası (veya varyansı) ile ölçülür. Etkinlik, iki yansız tahminleyici arasında tercih yaparken kullanılan bir kriterdir. Eğer aynı örneklem hacmi esas alınarak hesaplanan iki tahminleyici yansız tahminleyiciler ise bunlardan standart hatası küçük olan, diğerine göre daha etkin tahminleyicidir. Bazı durumlarda az miktarda yanlılığa sahip, standart hatası küçük olan tahminleyici yansız fakat standart hatası büyük bir tahminleyiciye tercih edilebilir.

Nokta tahminlemesi tahminlenmesi istenen parametre için tek bir değer belirler. Ancak bu tahminlemenin sonucu olan tahminin parametre değerine ne kadar yakın olduğunu belirlemek mümkün değildir.

Püskülcü, H. ; İkiz, F. (1986). İstatistiğe Giriş, Ege Üniversitesi Basımevi, İkinci Baskı, İzmir. KİTAP

Tek Evren Parametresine İlişkin Nokta Tahminlemesi

Tek evren parametresine ilişkin nokta tahminleme sürecinde aşağıdaki aşamalar izlenir:

- Tanımlanan evrenden, belirlenen n birimlik (hacimli) bir basit rassal örneklem seçilir.
- Bu örneklemdeki birimler üzerinden ilgilenilen değişken itibarıyla veriler derlenir.
- Bu veriler kullanılarak tahminlenecek parametre θ için bilgi üretecek $\hat{\theta}$ istatistiği hesaplanır.
- Son olarak bu istatistiğin değerinden ve dağılımının özelliklerinden yararlanarak θ parametresi için tahminleme yapılır.

Bu kısımda tek evren ortalaması μ, tek evren oranı π, iki evren ortalaması ve son olarak iki evren oranı arasındaki farkın nokta tahminlemesi ile ilgili bilgiler verilmiştir.

Evren Aritmetik Ortalaması μ'nün Nokta Tahminlemesi

Bilindiği gibi μ için bilgi üreten istatistik, tahminleyici $\bar{X}$dır. Tanımlanan evrenden basit rassal örneklemeyle n hacimli bir örneklem oluşturulur ve $(\bar{X})$ hesaplanır ise hesaplanan $\bar{X}$ birinci ünitede açıklandığı gibi bir rassal değişkendir. Eğer evren sonsuz ise veya evren sonlu fakat örneklem hacmi büyük veya örnekleme oranı $n/N < 0{,}05$ ise $\bar{X}$nın örnekleme dağılımının ortalaması $\mu_{\bar{X}}$ veya aynı anlama gelen $\bar{X}$ nın beklenen değeri $E(\bar{X})$ daima evren ortalaması μ ye eşittir ve onun yansız tahminidir. Bu bilgiye göre μ nün nokta tahminlemesi,

$$E(\bar{X}) = \mu$$

şeklinde yazılır.

Tanımlanan pek çok evrenin ortalama bilgisi için genellikle örneklem aritmetik ortalaması en küçük varyansa sahip ve daima yansız tahminleyici olduğundan diğer ortalama (medyan gibi) tahminleyicilerine tercih edilen bir tahminleyicidir.

Yapılan açıklamaları bir örnek üzerinde görelim.

ÖRNEK 1

Bir telefon işletme yönetimi idaresi abonelerinin 2011 yılındaki ortalama aylık telefon ödeme tutarlarını tahminlemek istemektedir. Bu amaçla rassal olarak 80 abone seçilmiş ve abonelerin ortalama aylık ödeme tutarının ₺45 ve standart sapmasının da ₺2 olduğunu hesaplamıştır. Bütün abonelerin aylık ortalama ödeme tutarını tahmin ediniz.

Çözüm:

$\bar{X}$ = ₺45

s = ₺2

n = 80 abone

μ = ?

80 birimlik rassal örneklemin aritmetik ortalaması $\bar{X}$= ₺45, $\bar{X}$evren ortalaması μ için bilgi üreten bir istatistiktir ve bir yansız tahminleyicidir. Buradan telefon abo-

nelerinin aylık ortalama ödeme tutarı μ nün nokta tahmini $\bar{X}$nın beklenen değerine eşit olarak alınır ve;

$$E(\bar{X}) = \mu_{\bar{x}} = \mu = ₺45$$

şeklinde yazılır. Bu nokta tahminlemesiyle gerçekte evren ortalaması μ nün ₺45 ye yakın bir değer olduğu yorumu yapılır.

Evren Oranı π'nin Nokta Tahminlemesi

Örnekleme uygulamalarının pek çoğunda evren oranı hakkında bilgi elde edilmesi istenir. Bu durumda π için bilgi üreten istatistik örneklem oranı p olur.

Evren oranı π'ye ilişkin nokta tahminlemesi bir rassal örnekleme planında oluşturulan n hacimli örneklem için, r bir binom rassal değişkeni olmak üzere, hesaplanan $p = \frac{r}{n}$ oranının değerini π'ye eşit alan bir tahminleme sürecidir. Burada,

$$p = \frac{r}{n}$$

eşitliği tahminleyici, hesaplanacak p değeri de tahmindir. p tahminleyicisi binom dağılımının parametresi olan π için bir tahminleyicidir. Sonsuz evrenden veya sonlu evrenden $n/N < 0,05$ olmak üzere seçilecek rassal bir örneklem için hesaplanan ve bir rassal değişken olan p istatistiğinin örnekleme dağılımının ortalaması μ_p veya başka anlatımla p istatistiğinin örnekleme dağılımının beklenen değeri E(p) evren oranı π ye eşittir. Buradan p 'yi π nin yansız ve küçük varyanslı nokta tahminlemesi

$$E(p) = \pi$$

şeklinde yazılır.

ÖRNEK 2

Bir TV film yapımcısı, gösterime giren filmini beğenenlerin oranını tahminlemek istiyor. Bu amaçla filmi izleyenler arasından rassal olarak 100 kişi seçiyor ve bunların 65'inin filmi beğendiklerini öğreniyor. İstenen tahmini nokta tahmini olarak yapınız.

Çözüm:

n = 100 kişi

r = 65 filmi beğenen kişi sayısı

$p = \frac{r}{n} = \frac{65}{100} = 0,65$ örneklemdeki kişiler içinde filmi beğenenlerin oranı

Örneklem oranı p = 0,65 değerini, evren oranı π nin yansız tahmini olarak alan

$$E(p) = \mu = 0,65$$

şeklindeki tahminleme π nin nokta tahminlemesidir. Bu bilgiye göre filmi izleyenler arasındaki beğenenlerin oranının %65 olabileceği söylenebilir.

İki Evren Parametresi Arasındaki Farkın Nokta Tahminlemesi

İki evren parametresi arasındaki farkın nokta tahminleme sürecinde aşağıdaki aşamalar izlenir:

- Tanımlanan iki evrenden, belirlenen n_1 ve n_2 birimlik (hacimli) basit rassal örneklemler seçilir.

- Bu örneklemlerdeki birimler üzerinden ilgilenilen değişken itibarıyla veriler derlenir.
- Bu veriler kullanılarak tahminlenecek parametre (θ_1 - θ_2) için bilgi üretecek $\hat{\theta}_1$ - $\hat{\theta}_2$ istatistiği hesaplanır.
- Son olarak bu istatistiğin değerinden ve dağılımının özelliklerinden yararlanarak ilgili parametre için tahminleme yapılır.

Bu kısımda iki evren ortalaması arasındaki fark ve iki evren oranı arasındaki fark parametreleri hakkında nokta tahminlemesinin nasıl yapılacağına ilişkin bilgiler verilecektir.

İki Evren Ortalaması Arasındaki Fark (μ_1- μ_2)'nin Nokta Tahminlemesi

Dağılımı normal, ortalamaları ve olan iki evrenden μ_1 ve μ_2 hacimli veya dağılımları nasıl olursa olsun örneklem hacimleri $n_1 \geq 30$, $n_2 \geq 30$ birim olmak üzere birbirinden bağımsız ve rassal örneklemler seçilirse μ_1 - μ_2 parametresi, bu parametre için bilgi üreten $\bar{X}_1$ - $\bar{X}_2$ istatistiğinin beklenen değeri E ($\bar{X}_1$ - $\bar{X}_2$) ne eşittir ve nokta tahminlemesi;

$$E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2$$

şeklinde yazılır.

Bu duruma göre ($\bar{X}_1$ - $\bar{X}_2$) istatistiği (μ_1 - μ_2) parametresi için yansız ve varyansı en küçük olan etkin bir tahminleyicidir.

ÖRNEK 3

A ve B satış eğitimlerini alan satış görevlilerinin sattıkları ortalama ürün sayıları arasındaki μ_1 - μ_2 fark tahminlenmek isteniyor. Bu amaçla A ve B eğitimini alan satış görevlileri arasından rassal olarak 50 şer kişi seçiliyor. A eğitimini alan 50 kişinin ortalama ürün satış sayısı 45 adet, B eğitimini alanların ise 35 adettir. İstenen nokta tahmin değeri nedir?

Çözüm:

A eğitimini alan	B eğitimini alan
n_1 = 50 kişi	n_2 = 50 kişi
$\bar{X}_1$ = 45 adet	$\bar{X}_2$ = 35 adet
	$E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2$ 45 - 35 = 10 adet

B eğitimini alan satış görevlilerinin satış adedi daha azdır.

İki Evren Oranı Arasındaki Fark (π_1 - π_2)'nin Nokta Tahminlemesi

İki evren oranı (iki binom parametresi) arasındaki farkın nokta tahminlemesi, iki evren ortalaması arasındaki farkın nokta tahminlemesine benzer. Birbirinden bağımsız $n_1 \geq 30$, $n_2 \geq 30$ birimlik rassal örneklemler seçilirse bu örneklemlerden hesaplanan p_1 ve p_2 örneklem oranları arasındaki farkı gösteren (p_1 - p_2) istatistiğinin örnekleme dağılımının ortalaması iki evren oranı arasındaki fark parametresine eşittir ve yansız nokta tahminidir. Bu bilgiye göre parametresinin nokta tahmini

$$E(p_1 - p_2) = \pi_1 - \pi_2$$

şeklinde yazılır.

ÖRNEK 4

Bir reklam filmini izleyen kız ve erkek öğrencilerin filmi beğenenlerinin oranları arasındaki fark nokta tahmini olarak tahminlenmek isteniyor. Bu amaçla bu filmi izleyen kız öğrenciler arasından 60, erkek öğrenciler arasından 40 kişi rassal olarak seçiliyor. Kız öğrencilerin 45'inin erkek öğrencilerin ise 24'ünün filmi beğendiği tespit ediliyor. İstenen tahminlemeyi yapınız.

KIZ	**ERKEK**
n_1 = 60 kişi	n_2 = 40 kişi
r_1 = 45 kişi	r_2 = 24 kişi
$p_1 = \frac{r_1}{n_1} = \frac{45}{60} = 0,75$	$p_2 = \frac{r_2}{n_2} = \frac{24}{40} = 0,60$

$$E(p_1 - p_2) = \pi_1 - \pi_2 = 0,75 - 0,60 = 0,15$$

Filmi beğenen kız öğrencilerin oranı ile filmi beğenen erkek öğrencilerin oranı arasındaki fark %15'tir; filmi beğenen kız öğrencilerin oranı daha yüksektir.

SIRA SİZDE 2

- **Tanımlanan evren sonsuz evren ise örneklem aritmetik ortalamasının dağılım şekli nedir?**
- **Dağılım şekli bilinmeyen iki evrenden seçilen birbirinden bağımsız rassal örneklemelerin ($n_1 \geq 30$, $n_2 \geq 30$ olmak üzere) ortalamaları arasındaki farkın dağılım şekli hakkında bilgi veriniz.**
- **Yansız tahmin nedir?**

Aralık Tahminlemesi

Daha önce açıklandığı gibi, bir tahminleme sürecinde, küçük varyansa (ya da standart hataya) sahip olan tahminleyicinin tercih edilmesi, önemli bir kriterdir. Standart hatanın küçüklüğü tahminin güvenilirliğiyle ilgilidir. Güvenilir tahmin, tanımlanan evrenden seçilen aynı hacimli farklı örneklemlerde büyük ölçüde farklılık göstermeyen tahmindir. Nokta tahminlemesi tahminin güvenilirliği hakkında bilgi veremediği, başka bir ifadeyle bu tahminleme tahminin parametre değerine ne kadar yakın olduğu bilgisini veremediği için sınırlı bir tahminlemedir. Tahminlemesi yapılacak parametre değeri ile tahmin değeri arasındaki olası fark için bir olasılık ifadesinin, güven düzeyinin kullanılmasına imkân veren aralık tahminlemesi geliştirilmiştir. Bu özelliği nedeniyle **aralık tahminlemesi,** nokta tahminlemesine tercih edilir.

Bilinmeyen evren parametresini, istenen bir olasılıkla, simetrik bir aralıkta tahminleme sürecine, **aralık tahminlemesi** denilir.

Bir parametrenin örneklem istatistiğine dayanarak, örneklemenin planlama aşamasında araştırmacı tarafından belirlenen bir olasılığa (güven düzeyine) göre simetrik bir aralıkta belirlenmesi çalışmasına aralık tahminlemesi denir.

Bu tanıma göre θ parametresinin aralık tahminlemesinin gösterimi genel olarak;

$$A < \theta < Ü$$

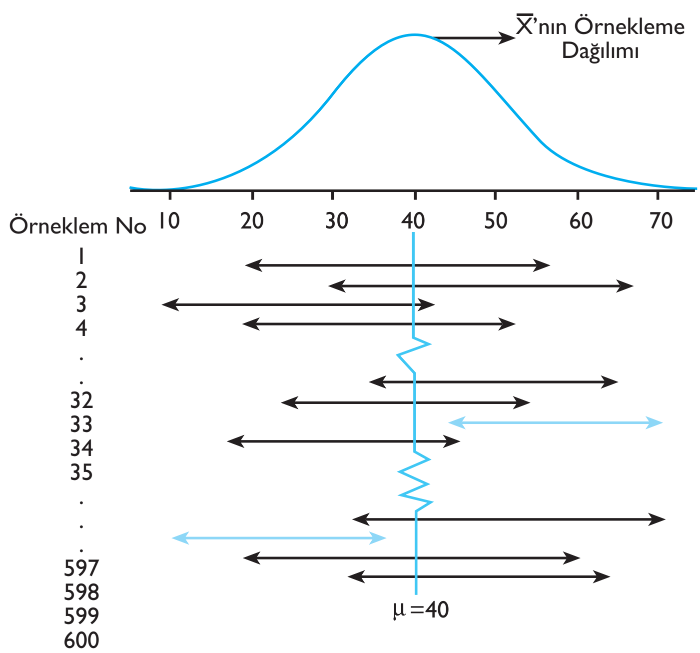
şeklinde yazılır. Burada A ve Ü tahminlenecek parametre değerini kapsayacak alanı belirleyen sınır değerleridir. Güven sınırları adı verilen bu sınırlardan A, alt sınır; Ü, üst sınır olarak tanımlanır.

Örneklem istatistiklerinin değerleri ve standart hataları örneklemden örnekleme değiştiğinden güven aralığının sınır değerleri değişir; güven aralığı genişler ya da daralır. Bu nedenle bazı güven aralıkları parametre değerini kapsar, bazıları kapsamaz.

Bu nedenle tahminleme sürecinde alt ve üst sınırlar birer rassal değişken, bu sınırların belirlediği aralık da rassal aralık niteliğindedir. Aralık tahminlemesi sürecinde, araştırmacı tarafından önceden belirlenen olasılık düzeyi ya da **güven düzeyi** (G.D.) parametre değerini kapsayan güven aralığının tahminlendiği olasılığı ifade eder ve 1-α ile gösterilir. Bir başka ifadeyle güven düzeyi tahminlenecek güven aralıklarının parametre değerini içine alma oranıdır. G.D.= 1-α değeri büyük seçilirse tahminlenen aralığın θ'yı kapsayan bir aralık olma olasılığı artmış fakat tahminlerin güvenilirliği, kesinliği azalmış olur.

Güven aralığının tahmininde parametre değerini kapsayan olasılığa **güven düzeyi** denilir.

Şekil 2.1



Evren ortalamasının mümkün örneklemlerdeki aralık tahminlerinin dağılımı.

Aralık tahminlemesinde güven aralığının mümkün olduğu ölçüde dar tutulması arzu edilir. Çünkü dar aralığın sınırları parametre değerine daha yakındır. Bu aralık güven düzeyine ve örneklem hacmine bağlıdır. Örneğin güven düzeyi %99'dan %95'e düştüğünde daha dar güven aralığı elde edilir. Belirlenen örneklem hacmi için hesaplanan standart hatanın küçük olması durumunda da güven aralığı daralır.

Tek Evren Parametresinin Aralık Tahmini

Bu tahminleme sürecinde şu aşamalar izlenir:

- Güven Düzeyi G.D.=1-α araştırmacı tarafından belirlenir,
- Hacmi n olan bir rassal örneklem seçilir,
- Bu örneklem için $\hat{\theta}$ istatistiği ($\bar{X}$, s, p istatistikleri) hesaplanır,
- $\hat{\theta}$ istatistiğinin standart hatası $\sigma_{\hat{\theta}}$ ($\sigma_{\bar{X}}$, $s_{\bar{X}}$, σ_p, S_p) hesaplanır.
- $\hat{\theta}$ 'nın ($\bar{X}$, p istatistiklerinin) dağılım şekliyle ilgili bilgilerden yararlanarak $A < \theta < Ü$ güven aralığı tahminlenir. Burada; A ve *Ü* güven aralığın sırasıyla, alt ve üst sınır değerlerini gösterirler.

Bu kısımda tek evren ortalaması μ, tek evren oranı π, iki evren ortalaması arasındaki fark ve iki evren oranı arasındaki farkın aralık tahminlemesi ile ilgili bilgiler verilmiştir.

Evren Aritmetik Ortalaması μ'nün Aralık Tahminlemesi

Örneklem hacmi n, $\bar{X}$'nin dağılım şeklini belirleyen önemli bir belirleyici olduğu için, μ'nün aralık tahminlemesiyle ilgili açıklamalar örneklem hacminin yeterli büyüklükte olup olmaması durumuna göre iki alt başlıkta açıklanmıştır.

Büyük Örneklemlerde μ'nün Aralık Tahminlemesi

Bilindiği gibi büyük örneklem hacmi için yeterli büyüklük, $n \geq 30$ birim kabul edilmektedir.

Rassal örneklem hacmi yeterli büyüklükteyse, evren dağılım şekli ne olursa olsun, $\bar{X}$'nın örnekleme dağılımı, ortalaması μ ve varyansı $\frac{\sigma^2}{n}$ olan normal dağılıma uyar. $\bar{X}$ rassal değişkeninin dağılımı normal dağılım özelliğine sahip olduğunda, μ nün aralık tahminlemesi,

σ biliniyorsa;

$$\underbrace{\bar{X} - z.\sigma_{\bar{X}}}_{A} < \mu < \underbrace{\bar{X} + z.\sigma_{\bar{X}}}_{Ü}$$

σ bilinmiyorsa;

$$\underbrace{\bar{X} - z.s_{\bar{X}}}_{A} < \mu < \underbrace{\bar{X} + z.s_{\bar{X}}}_{Ü}$$

Burada,

$\bar{X}$: Örneklem istatistiğinin değerini

$\sigma_{\bar{X}}$: bilindiğinde standart hata

$s_{\bar{X}}$: bilinmediğinde ve örneklem hacmi büyük olduğunda nın tahminleyicisidir

z: 1-α güven düzeyi için standart normal dağılım tablosundan belirlenecek değerdir

$\bar{X}$'nın standart hatası, örnekleme oranı $n/N < 0.05$ ise ve örneklemin seçiminde iadeli seçim uygulanırsa, ya da tanımlanan evren sonsuz hacimliyse,

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

formülüne göre hesaplanır. Buna karşılık örneklemin seçiminde iadesiz seçim uygulanıyor ya da örnekleme oranı $n/N \geq 0{,}05$ olduğunda, $\sigma_{\bar{X}}$ yukarıdaki eşitliğe

$\sqrt{\frac{N-n}{N-1}}$ çarpanı ilave edilerek hesaplanır.

Evren aritmetik ortalaması μ'nün aralık tahminlemesiyle ilgili yukarıda yapılan açıklamalar evren standart sapması σ 'nın bilindiği durum için geçerlidir. Ne var ki gerçek yaşamda karşılaştığımız problem çözümlemelerinde σ genellikle bilinmez. Bu durumda standart hata, σ yerine onun tahmini olan örneklem standart sapması s kullanılarak tahminlenir ve $s_{\bar{X}}$ simgesiyle gösterilir. Bu durumda rassal örneklem hacmi yeterli büyüklükte olduğunda, 1-α güven düzeyi için, μ'nün güven aralığı,

$\bar{X} - \sigma$ z. $s_{\bar{X}} < \mu < \bar{X}$ +z. $s_{\bar{X}}$

şeklinde tahminlenir.

Ortalamanın standart hatasının tahmini; örnekleme birim seçiminde iadeli seçim uygulanır, örnekleme oranı $n/N \leq 0.05$ veya tanımlanan evren sonsuz evren olursa;

$$s_{\bar{X}} = \frac{s}{\sqrt{n}}$$

formülüne göre hesaplanır. Buna karşılık birim seçimi iadesiz yapılır ya da örnekleme oranı $\frac{n}{N} \geq 0{,}05$ olursa standart hata s_x,

$$s_{\bar{X}} = \frac{s}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

formülüyle hesaplanır.

Yukarıdaki formüllerdeki z değeri, araştırmacı tarafından başlangıçta belirlenen 1-α güven düzeyine bağlı olarak standart normal eğri alanları tablosundan bulunur. Bu tablo kitabımızın sonunda yer almaktadır. Bu tablodan alınan örnek bazı 1-α güven düzeyleri için z değerleri aşağıdaki tabloda verilmiştir.

Bazı (1-α) değerleri için $Z\frac{\alpha}{2}$ tablo değerleri.

1-α	$Z\frac{1-\alpha}{2}$	1-α	$Z\frac{1-\alpha}{2}$
0,6827	1,00	-	-
0.95	1.960	0.60	0.842
0.99	2.576	0.80	1.282
0.90	1.645	0.50	0.674

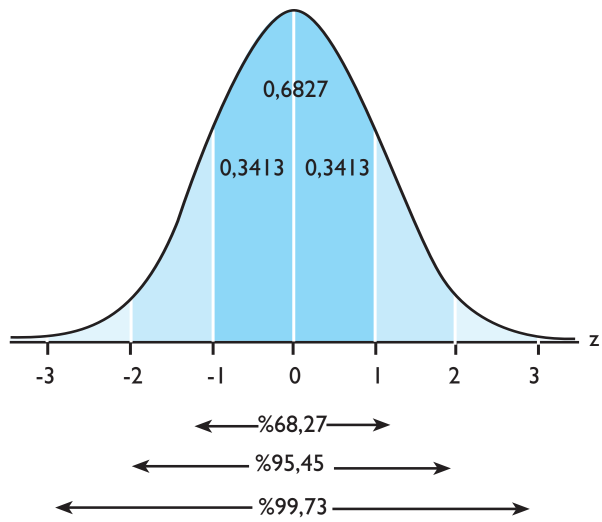
z tablo değerlerinin nasıl belirleneceğine ilişkin bilgiler aşağıda açıklanmıştır. Bilindiği gibi ortalamaları ve standart hataları farklı bütün $\bar{X}$ rassal değişkenleri $\frac{\bar{X}-\mu}{\sigma_{\bar{X}}}$ dönüşümüyle tek bir ortalamaya ve varyansa sahip tek bir rassal değişken ile tanımlanmış olur. $\bar{X}$ rassal değişkeninin dağılımı normal ise standart rassal değişken olarak adlandırılan ve z simgesiyle gösterilen rassal değişken aşağıdaki gibi yazılır:

$$Z = \frac{\bar{X}-\mu}{\sigma_{\bar{X}}}$$

Bu standart rassal değişken ortalaması $\bar{Z} = 0$ ve varyansı 1 olan normal dağılıma sahiptir. Bu dağılıma standart normal dağılım denir ve grafiği aşağıda verilmiştir:

Şekil 2.2

Standart normal dağılımı eğrisi.



Standart normal eğri altındaki alan 1 'e yani %100 'e eşittir. Standart normal dağılım, normal dağılımda olduğu gibi simetrik bir dağılımdır. $\bar{z}$= 0'ın solundaki ve sağındaki alanlar 0,5 tir (yani tüm alanın %50 'sidir). $\bar{z}$'ın solundaki z = -1, -2, -3 değerlerinin önündeki - (eksi) işaretinin cebirsel bir anlamı yoktur. Bu - (eksi) işareti söz konusu z değerlerinin $\bar{z}$'ın solunda kaldığını, bu z değerlerine karşı gelen örneklem istatistiklerinin ortalama değerden küçük değerler olduğu anlamına gelir. Şekildeki $z=\mp1$, $z=\mp2$, $z=\mp3$ değerleri sırasıyla G.D. = 0,6827, G.D. = 0,9545 ve G.D. = 0,9973 güven düzeyleri için Ek 1'de verilen standart normal eğri alanları tablosundan yararlanılarak belirlenmiştir. Standart normal dağılımın özelliklerini esas alarak düzenlenmiş olan standart normal eğri alanları tablosu dağılımları normal, ortalamaları ve standart hataları farklı bütün örnekleme dağılımları için istatistiksel tahminleme ve karar vermeye ilişkin çözümlemelerde büyük kolaylık sağlar. Örneğin G.D. = 1-α = 0,6827 (veya α = 1 - 0,6827 = 0,3173) olarak belirlenmiş ise z değeri,

$$Z = \left(\frac{1-\alpha}{2}\right) = Z\left(\frac{0,6827}{2}\right) = Z\left(0,3413\right)$$

veya

$$Z\left(0,5 - \frac{\alpha}{2}\right) = Z\left(0,5 - \frac{0,3173}{2}\right) = Z\left(0,3413\right)$$

için tablo değeri z tablosundan kolaylıkla belirlenebilir. Burada 0,3413 değeri $\bar{z}$ = 0 ile z = 1,00 veya $\bar{z}$ = 0 ile z = -1,00 arasındaki eğri altındaki alanın oransal değeridir. Çünkü standart normal dağılım simetrik dağılımdır. Standart normal eğri altındaki ile z = 1,00 veya ile z = -1,00 arasındaki alanlar birbirine eşittir. 0,3413 değeri (veya en yakın değer) ekte verilen standart normal eğri alanları tablosundan aranır. 0,3413 değerinin bulunduğu satırın solundaki z sütununa karşı gelen değer 1,0 ile 0,3413 değerinin bulunduğu sütunda z değerleri satırındaki değer 0,00 toplanır ve z = 1,0+0,00 = 1,00 değeri belirlenmiş olur. Aranan z değeri $\bar{z}$'ın solunda ise z = 1,00 'ın önüne eksi (-) işareti konulur. Örneğin G.D. = 1-α = 0,95 için z değeri, yapılan açıklamalara benzer şekilde,

$$Z\left(\frac{1-\alpha}{2}\right)=Z\left(\frac{0,95}{2}\right)=Z\left(0,4750\right)$$

veya

$$Z\left(0,5-\frac{\alpha}{2}\right)=Z\left(0,5-\frac{0,05}{2}\right)=Z\left(0,4750\right)$$

standart normal eğri alanları tablosundan 0,4750 alanına karşı gelen z değeri z (0.4750) = $\pm$ 1,96 olarak tablodan belirlenir.

ÖRNEK 5

Konserve bezelye üreten bir fabrikadan, üretim sırasında 64 konserve kutusundan oluşan rassal bir örneklem seçilmiştir. Bu konserve kutularının ortalama ağırlığı 492 gr. ve standart sapma 12 gr. olarak belirlenmiştir. Üretilen konservelerin ortalama ağırlığını %95 güven düzeyinde tahmin ediniz.

Çözüm:
n = 64 konserve
$\bar{X}$ = 492 gr.
s = 12 gr.
G.D. = 1 - α = 0.95 dolayısıyla α = 0.05 olur.

Örnekte evrenin değişkenliği ve bölünme şekli hakkında bilgi bulunmamaktadır.

Örneklem hacmi n = 64 birim (n>30 birimdir, ayrıca evren hacmi sonsuzdur) olduğu için $\bar{X}$'nın örnekleme bölünmesi normaldir (Merkezi Limit Teoremi). Bu bilgilere göre z = $\pm$ 1.96 alınır. Buna göre

$$s_{\bar{X}}=\frac{12}{\sqrt{64}}=\frac{12}{8}=1.5 \text{ gr.}$$

ve güven aralığı

$\bar{X}$ - z.s_x< μ < $\bar{X}$ - z.s_x

492 - 1.96(1.5) < μ < 492 + 1.96(1.5)

489.06 < μ <494.94

olarak hesaplanır.

Bu bilginin anlamı; üretilen konserve kutularının ortalama ağırlığı %95 güvenle 489.06 gr. ile 494.94 gr. arasındadır. Başka bir ifadeyle 489.06 - 494.94 gr. aralığının üretilen konservelerin gerçek ortalama ağırlığını kapsayacağına %95 güvenilebilir. Diğer bir deyişle aynı hacimli 100 mümkün rassal örneklem seçilmiş olduğunda, bu örneklemlerin %95'inin aritmetik ortalamalarının güven aralığı evren aritmetik ortalamasını kapsar; %5'ininki kapsamaz.

Küçük Örneklemlerde μ'nün Aralık Tahminlemesi

Örneklem hacmi küçük (n < 30 birim) olduğu zaman, örneklem ortalamalarının standart değerleri yukarıda açıklandığı gibi normal dağılıma sahip olmaz. Bu durumda μ için aralık tahminlemesi, tanımlanan evrenin normal dağılıma sahip olup olmadığının bilinmesine bağlıdır.

Normal dağılan bir evrenden, rassal olarak seçilebilecek birbirinden farklı n < 30 birim hacimli, mümkün bütün örneklemlerin seçildiğini, her örneklem için $\bar{X}$ ve onların $(\bar{X} - \mu)/ s_{\bar{x}}$ standart değerlerinin hesaplandığını düşünelim. Değer aralığı $-\infty$ ve $+\infty$ olan $(\bar{X} - \mu)/ s_{\bar{x}}$ istatistiği υ = n-1 serbestlik derecesinde t dağılımı adı verilen sürekli bir dağılıma uyar ve

$$t = \frac{\left(\bar{X} - \mu\right)}{s_{\bar{X}}}$$

şeklinde gösterilir.

t dağılımı ortalaması sıfır olan tek modlu ve simetrik bir dağılımdır. Bu dağılımın şekli normal dağılımın şekline benzer fakat değişkenliği daha fazladır.

Bilindiği gibi örneklenen evrenin dağılımı normal olduğunda, $\bar{X} - \mu)/\sigma_{\bar{x}}$ istatistiği standart normal dağılıma sahip olur. Evren değişkenliği σ'nın bilinmediği durumlarda $\sigma_{\bar{x}}$'nin yerine $s_{\bar{x}}$'nin ikame edilmesi; bu standart değişkenin üzerinde ilave bir değişkenliği tanımlar ve t istatistiği bu değişkenliği doğru bir şekilde dikkate alır. Örneklem hacmi artarken, serbestlik derecesi υ = n-1 büyür, $s_{\bar{x}}$'nin kullanılması nedeniyle ortaya çıkan değişkenlik küçülür ve t dağılımı standart normal dağılıma (z dağılımına) yaklaşır. Bu nedenle z dağılımına, t dağılımının özel hâlidir denir.

Yukarıda verilen bilgilerin ışığında ilgilenilen evren normal dağılıma sahip olduğunda, μ için güven aralığı büyük örneklerdekine benzer şekilde aşağıdaki gibi ifade edilir. Tek fark z yerine t istatistiğinin kullanılmasıdır.

Tanımlanan evren normal dağılıyorsa, μ için 1-α güven aralığı

$$\bar{X} - ts_{\bar{x}} < \mu < \bar{X} + ts_{\bar{x}}$$

şeklinde belirlenir.

Burada;

$\bar{X}$: Örneklem istatisiği

$s_{\bar{x}}$: Örneklem ortalamasının standart hatası $\sigma_{\bar{x}}$'nın tahminleyicisidir ve

$$s_{\bar{X}} = \frac{s}{\sqrt{n-1}}$$

tahminleyicisiyle tahminlenir.

t: 1-α güven düzeyinde ve n-1 serbestlik derecesine bağlı olarak t değerleri tablosundan bulunur. t tablosu örneği Ek 2'de verilmiştir.

Ek 2'deki tablodan yararlanarak 1-α güven düzeyi için t tablo değerini bir örnekle belirleyelim. Örneğin G.D. = 1-α = 0,95 olarak belirlenmiş ise α = 0,05 olur. Aralık tahminlemesi örneklem istatistiğine göre simetrik bir aralıkta belirleneceği ve t dağılımı da simetrik bir dağılım olduğu için t tablo değerini belirlemek amacıyla önce $t\frac{\alpha}{2} = t\frac{0,05}{2} = t\left(0,025\right)$ belirlenir. Sonra serbestlik derecesi belirlenir.

Örneğin belirlenen serbestlik derecesi 10 olsun. Tablo 2 'de serbestlik derecesi sütunundan 10, olasılık satırından ise 0,025 belirlenir. Bu iki sütunun kesiştiği noktadaki değer 2,228, t (0,025; 10) = 2,228 tablo değeri olarak belirlenmiş olur.

ÖRNEK 6

Bir cep telefonu üreticisi, üretilen cep telefonlarındaki bataryaların bir kez şarj edildikten sonra ortalama ne kadar süre kullanıldığını tahmin etmek istemektedir. Bu amaçla rassal olarak 17 telefon bataryası seçilmiş ve bunların ortalama 135 saat kullanılabildiği ve standart sapmanın 28 saat olduğu belirlenmiştir. %99 güven düzeyi için istenilen tahminlemeyi yapınız.

Çözüm:

n = 17 batarya

$\bar{X}$ = 135 saat

s = 28 saat

G.D. = 0.99

α = 0.01

Üretilen cep telefonu bataryalarının ortalama kullanım süresinin normal bölünmeye sahip olabileceğini, çok fazla çarpık olamayacağını biliyoruz. Bu nedenle

$$\bar{X} - ts_{\bar{x}} < \mu < \bar{X} + ts_{\bar{x}}$$

güven aralığı uygulanabilir. Belirlenen G.D = 1 - α =0.99 ve serbestlik derecesi υ = n - 1 = 17 - 1 = 16 dır.

%99 güven düzeyi (α = 0.01) ve υ = 16 serbestlik derecesi için t Tablo Değeri t0.01;16 = 2.921 elde edilir.

Ortalamanın standart hatası

$$s_{\bar{X}} = \frac{s}{\sqrt{n-1}} = \frac{28}{\sqrt{16}} 7 \text{ saat}$$

bulunur.

Örneklem ortalaması, tahmininden yararlanılarak %99 güven aralığı,

135 - 2.921 (7) < μ < 135 + 2.921 (7)

114.553 < μ < 155.447

biçiminde oluşturulur.

Üretilen cep telefonlarındaki bataryaların bir kez şarj edildikten sonra ortalama kullanım süresi %99 güvenle 114.553 saat ile 155.447 saat arasında bir değerdir. Bir başka ifadeyle aynı hacimli 100 mümkün rassal örneklem seçilmiş olduğunda, bu örneklemlerin %99'unun aritmetik ortalamalarının güven aralığı evren aritmetik ortalamasını kapsar; %1'ininki kapsamaz.

Evren dağılımı normal değilse, verilerin bir matematiksel dönüşümle normale yaklaştırabileceği düşüncesinden hareketle μ'nün güven aralığının tahminlenmesinde, (yukarıda açıklanan uygulama) dönüştürülmüş verilere uygulanır.

Örneğin logaritmik dönüşüm bu amaçla sıkça kullanılır. Çünkü dönüştürülmüş logaritmik veriler, dönüştürülmemiş verilere göre daha az çarpıktır.

Evren Oranı π'nin Aralık Tahminlemesi

Bu parametre için aralık tahmini prosedürü, bu ünitede örneklem hacmi büyük ($n \geq$ 30 birim) ya da n/N < 0.05 olduğu durum için uygulanmıştır. Daha önce de açıklandığı gibi, büyük örneklemler için örneklem oranı, p'nin örnekleme dağılımı, yaklaşık normal dağılım gösterir. Aynı zamanda biliniyor ki bu dağılımın ortalaması p, standart hatası

$$s_p = \sqrt{\frac{p(1-p)}{n}}$$

dır.

Buradan, rassal örneklem hacmi yeterli büyüklükteyse, örneklem oranlarının standart değerlerinin,

$$Z = \frac{p-\pi}{s_p}$$

standart normal dağılıma sahip olduğu söylenebilir ve evren ortalaması aralık tahmininde olduğu gibi 1-α güven düzeyinde evren oranı Π'nin aralık tahmini evren aritmetik ortalamasının tahminlenmesinde izlenen aşamalara benzer şekilde,

$$p\text{-}z.s_p < \Pi < p+z.s_p$$

hesaplanır.

ÖRNEK 7

Bir bölgede yaşayan kişiler arasından A gazetesi okuru olanların oranı belirlenmek isteniyor. Bu amaçla rassal olarak 200 kişiden oluşan bir örneklem seçiliyor ve seçilen 200 kişi arasından 58 kişinin A gazetesi okuru olduğu belirleniyor. Bu bölgedeki A gazetesi okurlarının oranını %95 güven düzeyinde tahmin ediniz.

Çözüm:

n = 200 kişi

r = 58 kişi

1 - α = 0.95

α = 0.05

π = ?

z= $\mp$= 1.96 belirlenir. Çünkü n = 200 > 30 birim olduğu için p istatistiğinin örnekleme dağılımı ile bu istatistiğin standart değerlerinin dağılımı normal olur. Bu nedenle z değeri standart normal eğri alanları tablosundan bakılmıştır.

$$p = \frac{r}{n} = \frac{58}{200} \; 0.29$$

1 - p = 0,71

$$s_p = \sqrt{\frac{(0.29)(0.71)}{200}} = 0.032$$

hesaplanır ve π

0.29 1.96(0.032) < π < 0.29 + 1.96(0.032)

0.29 0.06272 < π < 0.29 + 0.06272

0.22728 < π <0.35272

olarak tahminlenir.

Bu bölgedeki A gazetesi okurlarının oranı %95 güvenle, %22.73 ile %35.27 arasında bir değer olarak elde edilir. Bir başka ifadeyle aynı hacimli 100 mümkün rassal örneklem seçilmiş olduğunda, bu örneklemlerin %95'inin aritmetik ortalamalarının güven aralığı evren aritmetik ortalamasını kapsar; %5'ininki kapsamaz.

SIRA SİZDE 3

- **Tek bir rassal örneklem ortalamasının dağılımının normal olabilmesi için gerekli koşullar nelerdir?**
- **Uygulamada nokta tahminlemesine mi yoksa aralık tahminlemesine mi güvenirsiniz? Nedenleriyle açıklayınız.**

İki Evren Parametresi Arasındaki Farkın Aralık Tahminlemesi

Bu kısımda aralık tahminlemesi yapılacak parametreler iki evren ortalaması arasındaki fark $\mu_1 - \mu_2$ ve iki evren oranı arasındaki fark π_1 - π_2 olacaktır.

Evren Ortalamaları Arasındaki Fark (μ_1 - μ_2)'nin Aralık Tahminlemesi

Tanımlanan bazı araştırmalarda iki evren ortalaması arasındaki farkın tahminlenmesi istenmektedir. Örneğin 2010 yılı fen lisesi ve anadolu lisesi mezunlarının ÖSYM sınavı başarı ortalamalarının karşılaştırılması, iki farklı firma tarafından üretilen otomobil lastiklerinin ortalama ömürlerinin (km olarak) karşılaştırılması araştırma konusu yapılabilir. Birinci örnekte fen lisesinden 2010 yılında mezun olan ve ÖSYM sınavına giren öğrenciler topluluğu tanımlanan evrenlerden birincisi; anadolu liselerinden 2010 yılında mezun olan ve ÖSYM sınavına giren öğrenciler topluluğu ise ikinci evrendir.

Araştırmaya konu olan iki evrenden birincisinin ortalaması μ_1 ve standart sapması σ_1 ikincinin ortalaması μ_2 ve standart sapması σ_2 ile gösterilir. Her iki evrenden birbirinden bağımsız n_1 ve n_2 hacimli rassal örneklemler seçilir ve bu örneklemler için $\bar{X}_1$, s_1, $\bar{X}_2$, s_2 istatistikleri hesaplanır.

İki evren ortalaması arasındaki fark $\mu_1 - \mu_2$ (veya $\mu_2 - \mu_1$) parametresi hakkında tahminleme yapabilmek için bu parametre hakkında bilgi üreten $\bar{X}_1$ - $\bar{X}_2$ (veya $\bar{X}_2$ - $\bar{X}_1$) istatistiğinin örnekleme dağılımının özellikleri hakkında bilgiye ihtiyaç vardır.

Tanımlanan iki evrenin araştırmaya konu olan değişken itibarıyla dağılımları nasıl olursa olsun bu evrenlerden birbirinden bağımsız $n1 \geq 30$ ve $n2 \geq 30$ birim olmak üzere rassal örneklemler seçilirse merkezî limit teoremi gereğince hem $\bar{X}_1$ ve $\bar{X}_2$ nın hem de $\bar{X}_1$ - $\bar{X}_2$ istatistiklerinin örnekleme dağılımları normal olur.

Yukarıda açıklanan bilgilerin ışığında $\bar{X}_1$ - $\bar{X}_2$ istatistiğinin bilgi ürettiği $\mu_1 - \mu_2$ parametresinin 1-α güven düzeyinde aralık tahminlemesi rassal örneklemler birbirinden bağımsız ve örneklem hacimleri yeterli büyüklükte ise aşağıdaki şekilde yapılır:

$$(\bar{X}_1 - \bar{X}_2) - z.\, \sigma_{\bar{X}_1-\bar{X}_2} \;\; \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + z.\, \sigma_{\bar{X}_1-\bar{X}_2}$$

Burada standart hata,

$$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

formülüyle hesaplanır. Standart hata formülünde yer alan evren varyansları σ_1^2 ve σ_2^2 bilinmiyorsa bunların tahminleyicileri olan örneklem varyansları s_1^2 ve s_2^2 istatistikleri kullanılarak nın tahmini,

$$s_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

şeklinde ifade edilmiş olur. Bu durumda güven aralıkları,

$(\bar{X}_1 - \bar{X}_2) - z.\ s_{\bar{X}_1-\bar{X}_2} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + z.\ s_{\bar{X}_1-\bar{X}_2}$

şeklinde yazılır. $\bar{X}_1 - \bar{X}_2$ nın örnekleme dağılımı $n_1 \geq 30$ ve $n_2 \geq 30$ olması durumunda merkezî limit teoremi gereğince normal dağılıma sahip olduğu için bu yaklaşım önemli hataya sebep olmaz.

Ayrıca örneklem hacimleri yeterli büyüklükte olduğunda veya evrenlerin standart sapmaları σ_1 ve σ_2 bilindiği durumlarda evren standart sapmalarının eşit olup olmadığına bakılmaksızın,

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{X}_1-\bar{X}_2}}$$

veya $\sigma_{\bar{X}_1-\bar{X}_2}$ nın tahmini $s_{\bar{X}_1-\bar{X}_2}$ kullanılarak hesaplanan,

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1-\bar{X}_2}}$$

Standart değerlerinin dağılımı ortalaması 0 ve varyansı 1 olan normal dağılım gösterir. Bu nedenle $\mu_1 - \mu_2$ parametresinin yukarıdaki aralık tahminlemesi ifadelerinde z değeri kullanılmaktadır.

ÖRNEK 8

Aynı sektörde faaliyette bulunan iki firmanın çalışanlarının haftalık ortalama ücretleri arasındaki farkın %99 güvenle tahminlenmesi istenmektedir. Bu amaçla birinci firmadan rassal olarak 100 çalışan, ikinci firmadan ise rassal olarak 120 çalışan seçilmiştir. Birinci firmadan seçilen çalışanların haftalık ortalama ücreti ₺475 standart sapması ₺10; ikinci firmadan seçilenlerin ise haftalık ortalama ücreti ₺575 standart sapması ise ₺12 olduğu belirlenmiştir. İstenen tahminlemeyi yapınız.

Çözüm:

Birinci Firma	İkinci Firma
n_1 = 100 Çalışan	n_2 = 120 Çalışan
$\bar{X}_1$= ₺475	$\bar{X}_2$= ₺575
s_1 = ₺10	s_2 = ₺12

$n_1 > 30$ birim, $n_2 > 30$ birim olduğundan $\bar{X}_1 - \bar{X}_2$ nın örnekleme dağılımı ve bu dağılımın standart değerlerinin dağılımı normal dağılıma sahip olduğunda güven aralıkları,

$(\bar{X}_2 - \bar{X}_1) - z.\, s_{\bar{X}_2-\bar{X}_1} < \mu_1 - \mu_2 < (\bar{X}_2 - \bar{X}_1) + z.\, s_{\bar{X}_2-\bar{X}_1}$

şeklinde tahminlenir.

$$(575-475)-2{,}58\cdot\left(\sqrt{\frac{10^2}{100}+\frac{12^2}{120}}\right)<\mu_2-\mu_1<(575-475)+2{,}58\cdot\left(\sqrt{\frac{10^2}{100}+\frac{12^2}{120}}\right)$$

$96{,}173 < \mu_2 - \mu_1 < 103{,}827$

Burada z (0,4950) = 2,58 değeri standart normal eğri alanları tablosundan 1-α = %99 güven düzeyi için belirlenmiş bir değerdir.

Bu sonuca göre iki firmanın haftalık çalışanlarına ödediği ücretlerin ortalamaları arasındaki fark %99 güvenle ₺97,265 ile ₺102,735 arasında bir değerdir. Güven aralıklarının alt ve üst sınır değerleri pozitif olduğundan ikinci firmada çalışanların haftalık ortalama ücretleri, birinci firmadakilerden daha yüksektir. Başka bir deyişle bu iki firmadan n_1 = 100, n_2 = 120 hacimli mümkün bütün örneklemler seçilmiş ve $\bar{X}_2 - \bar{X}_1$ istatistikleri hesaplanmış olsa ve yukarıdaki şekilde güven aralıkları tahminlenirse tahminlenen güven aralıklarının %99'u iki evren ortalaması arasındaki farkı kapsar; %1'i kapsamaz yorumu yapılabilir.

Küçük Örneklemlerde ($\mu_1 - \mu_2$)'nin Aralık Tahminlemesi

Ortalamaları μ_1 ve μ_2, dağılımları normal veya normale yakın olan standart sapmalarının $\sigma_1 = \sigma_2 = \sigma$ olduğu varsayımı yapılabilen evrenlerden birbirinden bağımsız ve hacimleri $n_1 < 30$ ve $n_2 < 30$ birimlik rassal örneklemler seçilirse örneklemlerden yararlanılarak hesaplanan istatistiğinin dağılımının standart hatası,

$$s_{\bar{X}_1-\bar{X}_2}=\sqrt{\frac{n_1\cdot s_1^2+n_2\cdot s_2^2}{n_1+n_2-2}}\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

eşitliğiyle tahminlenir. Bu durumda $\bar{X}_1 - \bar{X}_2$ nın örnekleme dağılımının standart değerlerinin dağılımı $n_1 + n_2 - 2$ serbestlik derecesinde bilinen dağılımlardan t dağılımının özelliklerini taşır. t istatistiği,

$$t=\frac{\bar{X}_1-\bar{X}_2}{s_{\bar{X}_1-\bar{X}_2}}$$

şeklinde yazılır.

Yukarıdaki bilgiler kullanılarak 1-α güven düzeyinde $\mu_1 - \mu_2$ parametresi için güven aralıkları evrenlerin dağılımları normal veya normale yakın, standart sapmaları birbirine eşit olduğu varsayılırsa ve küçük hacimli örneklemler birbirinden bağımsız ise,

$(\bar{X}_1 - \bar{X}_2) - t.\, s_{\bar{X}_1-\bar{X}_2} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + s_{\bar{X}_1-\bar{X}_2}$

şeklinde ifade edilir.

Eğer $\sigma_1 = \sigma_2 = \sigma$ varsayımının gerçekleşmediği durumda örneklem hacimlerini mümkün olduğu kadar birbirine eşit veya yakın belirlemek suretiyle yukarıda açıklanan şekilde çözümleme yapmak uygun yol olur.

İki Evren Oranı Arasındaki Fark (π_1 - π_2)'nin Aralık Tahminlemesi

Bu tahminleme de iki evren ortalaması arasındaki farkın aralık tahminlemesine benzer. Tanımlanan iki evrenden birbirinden bağımsız $n_1 \geq 30$ ve $n_2 \geq 30$ birim olmak üzere rassal örneklemler seçilir. Her örneklem için,

$$p_1 = \frac{r_1}{n_1} \text{ ve } p_2 = \frac{r_2}{n_2}$$

istatistikleriyle iki örneklem arasındaki fark $p_1 - p_2$ (veya $p_2 - p_1$) istatistiği hesaplanır. $p_1 - p_2$ istatistiği, örnekleme dağılımı ile ilgili bilgiler açıklanırken belirtildiği gibi bir rassal değişkendir. Bilindiği gibi $p_1 - p_2$ istatistiği $\pi_1 - \pi_2$ parametresi için bilgi üreten bir istatistiktir ve onun yansız bir tahminleyicisidir. Merkezî limit teoremine göre iki evrenden birbirinden bağımsız, büyük hacimli rassal örneklemler seçilirse $p_1 - p_2$ istatistiğinin örnekleme dağılımı yaklaşık normal dağılım gösterir. Bu dağılımın standart hatası σ_{p1-p2} tanımlanan evrenlerin π_1 ve π_2 oranları bilinmediği ve tahminlenmesi gerektiği için hesaplanamaz; evrenlerden seçilen örneklemlerden hesaplanan p_1 ve p_2 istatistiklerinin örnekleme dağılımlarının

$$s_{p_1} = \sqrt{\frac{p_1(1-p_1)}{n_1}} \quad ve \quad s_{p_2} = \sqrt{\frac{p_2(1-p_2)}{n_2}}$$

standart hatalarından yararlanılarak tahminlenir.

σ_{p1-p2} nin tahminleyicisi s_{p1-p2} simgesiyle gösterilir ve aşağıdaki gibi yazılır:

$$s_{p_1-p_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

İki evren oranı arasındaki fark $\pi_1 - \pi_2 = 0$ olduğu ve bu iki evrenden rassal olarak seçilen örneklemler birbirinden bağımsız ve hacimleri yeterli büyüklükte ise örneklem istatistiğinin örnekleme dağılımına ilişkin

$$Z = \frac{p_1 - p_2}{s_{p_1-p_2}}$$

standart dağılım normal dağılım özelliğini taşır.

Yukarıdaki bilgilere dayanarak 1-α güven düzeyinde $\pi_1 - \pi_2$ nin güven aralıkları aşağıdaki gibi yazılır:

$$(p_1 - p_2) - z.s_{p1-p2} < \pi_1 - \pi_2 < (p_1 - p_2) + z.s_{p1-p2}$$

Burada

$$Z = Z\left(\frac{1-\alpha}{2}\right) \text{ veya } Z = Z\left(0,5 - \frac{\alpha}{2}\right)$$

z, standart normal eğri alanları tablosundan belirlenir.

ÖRNEK 9

Bir firma iki ayrı coğrafi bölgedeki fabrikalarında üretilen ürünleri içerisindeki defolu ürün oranları arasındaki farkı tahminlemek istemektedir. Bu amaçla birinci bölgedeki fabrikada üretilen ürünler arasından rassal ve birbirinden bağımsız olarak 180 ürün, ikinci fabrikada üretilenler arasından ise 150 ürün seçilmiştir. Birinci fabrikadan seçilen ürünlerin 15 tanesi, ikinci fabrikadan seçilen ürünlerin ise 12 tanesinin defolu ürün olduğu tespit edilmiştir. Güven düzeyi = 1-α = 0,95 için istenen tahminlemeyi yapınız.

Çözüm:

Birinci Fabrika	**İkinci Fabrika**
n_1 = 180	n_2 = 150
r_1 = 20	r_2 = 12
$p_1 = \frac{r_1}{n_1} = \frac{20}{180} = 0{,}11$	$p_2 = \frac{r_2}{n_2} = \frac{12}{150} = 0{,}08$

$(0{,}11 - 0{,}08) - 1{,}96 \,.\, 0{,}032 < \pi_1 - \pi_2 < (0{,}11 - 0{,}08) + 1{,}96 \,.\, 0{,}032$

$0{,}033 < \pi_1 - \pi_2 < 0{,}093$

Fabrikaların defolu ürün oranları arasındaki fark %95 güvenle %3,3 ile %9,3 arasında bir değerdir. 0,033 güven aralığının alt sınır değeri; 0,093 ise güven aralığının üst sınır değeridir. Sınır değerlerinin her ikisi de pozitif olduğu için birinci fabrikanın defolu ürün oranı π_1, ikinci fabrikanın defolu ürün oranı den büyüktür yorumu yapılır. Güven aralıklarının her ikisi de negatif işaretli olsaydı yorumu yapılırdı. Güven aralıklarının her ikisinin de farklı işarette olması durumunda ise kesin bir yorum yapılamaz. Yukarıdaki güven aralıkları için bir başka yorum ise tanımlanan fabrikalardan n_1 = 180 ve n_2 = 150 hacimli mümkün bütün örneklemlerin çekildiği ve hesaplanabilecek mümkün $p_1 - p_2$ istatistikleri için güven aralıkları tahminleneceği düşünülürse bu güven aralıklarının %95 i $\pi_1 - \pi_2$ parametre değerini kapsar; %5 i kapsamaz şeklinde olur.

SIRA SİZDE 4

- **Aralık tahminlemesi nedir?**
- **Dağılımı normal olan bir evrenden küçük hacimli rassal örneklem seçilmiş olduğunda örneklem ortalamasının örnekleme dağılımının standart hatası nasıl tahminlenir?**
- **%95 güvenle yapılan iki evren oranına ilişkin bir aralık tahminlemesinin aşağıdaki sonucunu yorumlayınız:**
  $0{,}10 < \pi_1 - \pi_2 < 0{,}14$

Özet

AMAÇ 1

İstatistiksel tahminleme tanımı yapmak ve bazı evren parametreleri için tahminleme sürecinin aşamalarını açıklamak

Tanımlanan evrenden rassal ve birbirinden bağımsız olarak seçilen örneklemden hesaplanan istatistikler yardımıyla bu evrenin sahip olduğu dağılımın parametre değerini araştırmaya tahminleme denir. Tahminlenecek parametre için bilgi üreten istatistiğin hesaplanmasında kullanılan formülasyona tahminleyici, örneklem gözlem değerlerinin bu tahminleyiciye uygulanması suretiyle hesaplanan sayısal değere tahmin adı verilir. Parametreler örneklem istatistiklerinden hareketle bir sayı olarak tahminleniyorsa yapılan tahminlemeye nokta tahminlemesi, parametre değerini kapsayan bir aralığın sınırlarıyla tahminleniyorsa aralık tahminlemesi söz konusu olur.

Evren parametresine ilişkin tahminleme sürecinin aşamaları aşağıdaki gibi gösterilir:

- Tanımlanan evrenden/evrenlerden belirlenen hacimde rassal örneklem/örneklemler seçilir.
- Bu örneklemdeki/örneklemlerdeki birimler üzerinden araştırmaya konu olan değişkenler itibarıyla veriler derlenir.
- Bu veriler kullanılarak tahminlenecek parametre için bilgi üretecek istatistikler hesaplanır.
- Bu istatistiğin/istatistiklerin dağılımının özelliklerinden yararlanılarak istenen parametre için tahminleme yapılır.

AMAÇ 2

Tek evren aritmetik ortalaması, oranı ve iki evren aritmetik ortalaması, oranına ilişkin nokta ve aralık tahminlemesi uygulamalarını örneklemek

Tahminlenecek parametre tek evrene ilişkin parametre olduğunda tek evren parametresiyle ilgili nokta ve aralık tahminlemesi yapılır. Tek evren parametresiyle ilgili nokta tahminlemesi yapılırken seçilecek örneklemin rassal örneklem olması gerekir. Örneklemden hesaplanan ve tahminlenecek parametre için bilgi üreten istatistiklerin birer rassal değişken olduğu düşünülür. Bir rassal değişken olan $\hat{\theta}$ örneklem istatistiğinin örnekleme dağılımı normal ise ve merkezî limit teoreminin gerekleri sağlanmış ise örneklem istatistiği $\hat{\theta}$ nın beklenen değeri $E(\hat{\theta})$ evren parametresi $\hat{\theta}$ ya eşit alınıyorsa nokta tahminlemesi yapılmış olur. Bu tahminleme $E(\hat{\theta}) = \hat{\theta}$ şeklinde yazılır. Örneklem istatistiklerinden yararlanarak evren parametreleri 1-α güven düzeyinde örneklem istatistiğine göre simetrik bir aralıkta tahminleniyorsa yapılan tahminleme aralık tahminlemesidir. Aralık tahminlemesinin genel gösterimi, $\hat{\theta} - z\,\sigma_{\hat{\theta}} < \theta < \hat{\theta} + z\,\sigma_{\hat{\theta}}$ şeklinde yapılır.

Kendimizi Sınayalım

1. Bir evrenin aritmetik ortalamasının güven sınırlarını %99,34 güvenle tahmin edebilmek için kullanılacak standart normal dağılım tablo değeri (z) kaçtır?

a. 2,66
b. 2,68
c. 2,70
d. 2,72
e. 2,74

2. Her birinde 50'şer işçi olan iki grubun ortalama verimlilik değerleri arasındaki farkın güven düzeyi 0,95 için tahminlenmek istendiğinde, z veya t tablolarından okunacak değer nedir?

a. t = 2,750
b. t = 2,042
c. z = 1,960
d. z = 1,645
e. z = 1,276

3. Hacimleri n_1 = 15, n_1 = 12 olan birbirinden bağımsız rassal örneklemlerin seçildiği evrenlerin dağılımları normal ise bu iki evrenin ortalamaları arasındaki farkın %95 güven düzeyi için tahminlenmesi istendiğinde, z veya t tablosundan okunacak değer nedir?

a. 1,706
b. 2,052
c. 2,58
d. 1,96
e. 2,06

4. Normal dağılıma sahip bir evrenden rassal olarak seçilen 100 birimlik bir örneklemin ortalaması 56, standart hatanın tahmini değeri 1,2'dir. Evren ortalaması 53 ise ortalamanın dönüştüğü z değerinin sağında kalan bölgenin oranlanmış alanı kaçtır?

a. 0,0062
b. 0,4893
c. 0,4938
d. 0,5000
e. 0,9893

5. Bir evrenden seçilen 100 birimlik rassal bir örneklemin ortalaması 20, standart sapması 5 olarak bulunmuştur. Evren ortalaması %95,44 güvenle hangi aralıkta yer alır?

a. 19-21
b. 15-25
c. 17-23
d. 18-22
e. 17-22

6. Kız ve erkek öğrencilerin başarı ortalamaları arasındaki farkın tahminlenmesini konu alan bir araştırmada tahminlenecek parametre nedir?

a. $\pi_1 - \pi_2$
b. $\bar{X}_1 - \bar{X}_2$
c. p_1 - p_2
d. $\mu_1 - \mu_2$
e. $\sigma_1 - \sigma_2$

7. Bir lojistik firmasında çalışan personelin iş yeri memnuniyetinin araştırılması istenmektedir. Bu amaçla 100 personel rassal olarak seçilmiştir. Seçilen personelin 68'inin iş yerinden memnun olduğu belirlenmiştir. Bu bilgilere göre bu firma çalışanlarının iş yerinden memnuniyet oranını tahminleyiniz.

a. 68
b. 32
c. 0,68
d. 0,32
e. 0,72

8. Ortalaması 50 kg ve varyansı 23 kg olan 50 birimlik bir örneklemin, evren ortalaması %95 güvenle hangi aralıkta yer alır?

a. 40,6 - 59,4
b. 45,08 - 59,34
c. 45,2 - 54,8
d. 48,66 - 51,34
e. 49,74 - 50,26

9. Bir milletvekili, kendi seçim bölgesinde partisinin oy oranını tahminlemek istiyor. Bu amaçla rassal olarak 750 seçmen seçiyor ve bunların 495'inin kendi partisine oy vereceğini tespit ediyor. Bu bilgilere göre istenen tahminin %95 güven sınırları nedir?

a. $0,627 < \pi < 0,693$
b. $0,600 < \pi < 0,702$
c. $0,598 < \pi < 0,637$
d. $0,640 < \pi < 0,712$
e. $0,602 < \pi < 0,701$

10. Evren parametresine yakın örnekleme dağılımına sahip olan tahminleyiciye ne ad verilir?

a. Yansız tahminleyici
b. Etkin tahminleyici
c. Tutarlı tahminleyici
d. Yeterli tahminleyici
e. Yanlı tahminleyici

Kendimizi Sınayalım Yanıt Anahtarı

1. d	Yanıtınız yanlış ise “z ve t Tablolarının Kullanımı” konusunu yeniden gözden geçiriniz.
2. c	Yanıtınız yanlış ise “z ve t Tablolarının Kullanımı” konusunu yeniden gözden geçiriniz.
3. e	Yanıtınız yanlış ise “Evren Ortalamaları Arasındaki Farkın Aralık Tahminlemesi” konusunu yeniden gözden geçiriniz.
4. a	Yanıtınız yanlış ise “Tek Evren Ortalamasına İlişkin Aralık Tahminlemesi ve z Tablosunun Kullanımı” konusunu yeniden gözden geçiriniz.
5. a	Yanıtınız yanlış ise “Tek Evren Ortalamasına İlişkin Aralık Tahminlemesi” konusunu yeniden gözden geçiriniz.
6. d	Yanıtınız yanlış ise “İki Evren Parametresi Arasındaki Farka İlişkin Tahminleme” konusunu yeniden gözden geçiriniz.
7. c	Yanıtınız yanlış ise “Tek Evren Oranına İlişkin Nokta Tahminlemesi” konusunu yeniden gözden geçiriniz.
8. d	Yanıtınız yanlış ise “Tek Evren Ortalaması Aralık Tahminlemesi” konusunu yeniden gözden geçiriniz.
9. a	Yanıtınız yanlış ise “Tek Evren Oranının Aralık Tahminlemesi” konusunu yeniden gözden geçiriniz.
10. a	Yanıtınız yanlış ise “Nokta Tahminlemesi” konusunu yeniden gözden geçiriniz.

Sıra Sizde Yanıt Anahtarı

Sıra Sizde 1

- Rassal örneklem için hesaplanan istatistiklerden yararlanarak bu istatistiklerin bilgi ürettiği parametre değerlerinin araştırılmasına tahminleme denir.
- Örneklem gözlem değerinin bir tahminleyiciye uygulanmasıyla hesaplanan sayısal değere tahmin; tahminin hesaplanmasında kullanılan formülasyona tahminleyici denir.

Sıra Sizde 2

- Evren sonsuz ise örneklem ortalamasının örnekleme dağılımı normal olur.
- Dağılım şekli bilinmeyen iki evrenden birbirinden bağımsız büyük hacimli örneklemler seçilirse, örneklemlerin ortalamaları arasındaki farkın dağılımı normal olur.
- Bir tahminleyicinin örnekleme dağılımı tahminlenecek parametrenin yakınında serpilme gösteriyorsa, bu tahminleyici yansızdır denir.

Sıra Sizde 3

- Tek bir rassal örneklem ortalamasının dağılımının normal olabilmesi için; a) evrenin dağılımının normal olması ve evren değişkenliğinin biliniyor olması b) evrenin dağılımının bilinmemesi durumunda rassal örneklem hacminin yeterli büyüklükte olması gerekir.
- Nokta tahminlemesi tahminin güvenilirliği hakkında bilgi veremediği için bu bilgiyi veren aralık tahminlemesi tercih edilir.

Sıra Sizde 4

- Evren parametresinin rassal örneklemenin planlanan aşamasında araştırmacı tarafından belirlenen güven düzeyinde örneklem istatistine göre simetrik bir aralıkta belirlenmesi çalışmalarına aralık tahminlemesi denir.
- Söz konusu standart hata $S_{\bar{X}} = \dfrac{S}{\sqrt{n-1}}$ eşitliği ile hesaplanır.
- Bu iki evren oranı arasındaki fark %95 güvenle 0,10 ile 0,14 arasında bir orandır. Başka bir ifadeyle mümkün bütün örneklemler için güven aralıkları tahminlenmiş olsa, bu güven aralıklarının %95’i iki evren oranı arasındaki farkı kapsar, %5’i kapsamaz şeklinde yorum yapılır.

Yararlanılan ve Başvurulabilecek Kaynaklar

Çömlekçi, N. (2005). **Temel İstatistik İlke ve Teknikleri,** (4. Baskı), İstanbul, Bilim Teknik Yayınevi.

Neter, J., Wasserman, W., Whitmore, G. A., (1993). **Applied Statistics,** 4. Baskı, Boston, Allyn and Bacon.

Özmen, A.,(2000). **Uygulamalı Araştırmalarda Örnekleme Yöntemleri,** Eskişehir, Anadolu Üniversitesi Yayını.

Püskülcü, H., İkiz, F. (1986). **İstatistiğe Giriş,** 2. Baskı, İzmir, Ege Üniversitesi Basımevi.

Serper, Ö. (2004). **Uygulamalı İstatistik 2,** 5. Baskı, Bursa, Ezgi Kitabevi.

Yüzer, A. F. (Ed.), (2011). **İstatistik,** (8. Baskı) Eskişehir, Anadolu Üniversitesi Yayını.

3

Amaçlarımız

Bu üniteyi tamamladıktan sonra;

- Hipotez, istatistiksel hipotez ayrımını ifade edebilecek,
- İstatistiksel hipotezlerin test aşamalarını açıklayabilecek,
- Tek evren parametresiyle ilgili hipotez testi uygulayabilecek,
- İki evren parametresiyle ilgili hipotez testi uygulayabilecek,
- İkiden fazla evren ortalamasına ilişkin karşılaştırma, F testi uygulayabilecek

bilgi ve becerilere sahip olacaksınız.

Anahtar Kavramlar

- Evren
- Parametre
- Örneklem
- Örneklem İstatistiği
- Tahminleme
- İstatistiksel Hipotez

İçindekiler

İstatistik-II → İstatistiksel Karar Alma →

İstatistiksel Karar Alma

GİRİŞ

Araştırmacılar tam sayım yapamayınca, örneklemeye başvurunca evren parametrelerinin daha önceden bilinen değerinde değişiklik olup olmadığı, parametrenin belirlenen standart değerinde farklılık meydana gelip gelmediği veya iki veya daha fazla evren parametreleri arasındaki farklılığa ilişkin karar verilmesi problemiyle karşılaşırlar. Bu türden kararların verilmesi amacıyla istatistiksel hipotez testleri kullanılır.

Bu ünitede, istatistiksel hipotez testi ve test sürecinin aşamalarıyla ilgili teorik bilgiler açıklanacak ve uygulamada sıkça kullanılan tek evren parametreleri ve iki evren parametreleri ile ikiden fazla evren ortalamasının karşılaştırılmasıyla ilgili hipotez testi uygulamalarına yer verilecektir.

Bu ünitenin izleyen kısımlarındaki konular için A. F. Yüzer, E. Şıklar, E. Ağaoğlu, H. Tatlıdil, A. Özmen, Editör: A. F. Yüzer (2011). *İstatistik*, Ünite 8 ve 9, Eskişehir Anadolu Üniversitesi Yayını isimli kitaptan, editör ve ünite yazarlarının izni alınarak yararlanılmıştır.

KİTAP

İSTATİSTİKSEL HİPOTEZ VE İSTATİSTİKSEL HİPOTEZ TESTİ

Hipotez, karşılaşılan özel duruma ilişkin bir önermedir. **İstatistiksel hipotez** ise bir dağılımın evren parametresine ilişkin bir önermedir.

Genel olarak **hipotez**, karşılaşılan özel duruma ilişkin bir önermedir. **İstatistiksel hipotez**, bir araştırmada ilgilenilen bir ya da daha fazla parametrenin değeri hakkında ileri sürülen ve doğruluğu, geçerliliği bu parametre hakkında bilgi üreten istatistikden ve bu istatistiğin örnekleme dağılımıyla ilgili bilgilerden yararlanarak araştırılabilen önermelerdir. İstatistiksel hipotezleri diğer hipotezlerden ayıran özellik, bu hipotezlerin bir frekans dağılımının parametre değerine ait olmasıdır. Bazı istatistiksel hipotez örnekleri aşağıda verilmiştir.

Örnekler:

1. Günlük ortalama üretimi 750 kg olan bir ilaç fabrikasında, uygulanan yeni üretim tekniği, ortalama üretimi arttırmıştır.
2. Bir üretim sürecinde üretilen tereyağı paketleri ortalama 500 gr ağırlığındadır.
3. Bir yerleşim yerinde ikamet eden ailelerin %10'u alışverişlerini süper marketlerden yapmaktadır.

Yukarıdaki örneklerden de anlaşılabileceği gibi evren parametre değerleri hakkındaki hipotezler (önermeler) parametre değerleri hakkında, daha önceden bilinen, belirlenen standart bir değer ya da varsayımsal bir değer olabilir. Birinci örnekte, ilaç üretim yönteminin ortalama üretim düzeyi olan 750 kg bilinen bir değerdir. İkinci örnekteki tereyağı paketlerinin planlanan ağırlığı olan 500 gr standart bir değerdir. Son örnekteki süper marketlerden alışveriş yapan ailelerin oranı olan %10 varsayımsal bir değerdir ya da daha önce aynı konuda yapılan araştırmalarda elde edilen bir bilgidir.

Bir istatistiksel hipotez, doğru ya da yanlış olabilir. Çünkü bu bir önermedir. Gerçeği öğrenebilmek için, evren parametresi θ'nın değerini hesaplamak gerekir. Bu da tamsayım yapmayı gerektirir. Ancak, örnekleme yapmayı gerektiren nedenlerden dolayı bu, her zaman mümkün değildir. Bu durumda istatistiksel hipotezlerin geçerliliği ya da doğruluğu konusunda karar verebilmek için bu hiptezlerin, tanımlanan evrenden seçilen örneklemin gözlem değerinden hesaplanan örneklem istatistiğinden ($\hat{\theta}$'dan), bu istatistiğin ($\hat{\theta}$'nın) örnekleme dağılımının özelliklerinden yararlanarak test edilmesi gerekir. İstatistiksel hipotez testi, örneklem istatistiklerini kullanarak, bir hipotezin doğru olup olmadığını ortaya koymaya yönelik yapılan çalışmalardır.

Daha önce de belirtildiği gibi evrenden rassal örneklem alınmış olsa bile, örneklem gözlemlerinden hesaplanan bir istatistiğin, bu istatistiğin bilgi ürettiği parametre hakkında ileri sürülen değere (θ_0)'a eşit olması beklenemez. Yani örneklem istatistikleri aynı hacimli farklı örneklemlerde farklı değerler alabildiği için $\hat{\theta} - \theta_0 > 0$, $\hat{\theta} - \theta_0 = 0$ ya da $\hat{\theta} - \theta_0 < 0$ gibi farklar olabilir. Bu nedenle, istatistiksel test sonucu verilecek kararın, güvenilir olduğu konusunda kesin karar verilemez. Fakat, olasılık kuramından yararlanarak, bir istatistiksel hipotezin yapılacak bir testle ne derece güvenle (ne derece hatayla) kabul ya da reddedileceğini belirlemek olanaklı olmaktadır. Burada önemli olan ($\hat{\theta} - \theta_0$) farkının istatistiksel olarak anlamlı olup olmadığını belirlemektir. Başka bir anlatımla, farkların gerçek değişmeyi mi, farklılığı mı açıkladığı, yoksa rassal olarak mı meydana geldiğini ortaya koymaktır. Anlamlı farklılık belirlenmişse hipotez, belirli bir hata payıyla reddedilir. Tersi durumda kabul edilir.

SIRA SİZDE 1

- **İstatistiksel hipotez nedir?**
- **İstatistiksel hipotez testinin konusu nedir?**

HİPOTEZ TESTİ TÜRLERİ

Hipotez testleri, ilgilenilen değişkenin ölçülmesinde benimsenen ölçeğe bağlı olarak, parametrik hipotez testleri ve parametrik olmayan hipotez testleri şeklinde sınıflandırılırlar. Parametrik testler değişkenlerinin ölçülmesinde eşit aralıklı ya da oranlı ölçeğin kullanıldığı hipotez testleridir. Çünkü; bu iki ölçekte de elde edilen veriler üzerinden aritmetik işlemler yapmak mümkündür. Parametrik hipotez testlerindeki hipotezde, θ parametresinin önceden bilinen θ_0 değerine eşit ya da bundan büyük, küçük ya da farklı olduğu ileri sürülebilir.

Parametrik testler evren sayısının tek ya da iki oluşuna ve iki evrenin varlığında, bu evrenlerden rassal olarak seçilen örneklemlerin bağımsız ya da bağımlı oluşuna bağlı olarak sınıflandırılırlar. En önemli parametrik testler z ve t testleridir.

Bu ünitede tek evren ortalaması ve iki evren ortalaması arasındaki farka ilişkin z ve t testleri ile tek evren oranı ve iki evren oranı arasındaki farka ilişkin z testi ile ikiden fazla evren ortalamasının karşılaştırılmasına imkân veren F testi uygulamalarına yer verilmiştir.

Parametrik olmayan testler, evren dağılımı nasıl olursa olsun uygulanabilen testlerdir. Bu testlerde, parametrelerle ilgilenilmeyip hipotezler ilgili değişkenin belirli bir nitel özelliğine göre oluşturulur. Bu ünitede parametrik olmayan testler inceleme konusu yapılmamıştır.

Parametrik olmayan testlerle ilgili ihtiyaç duyulan bilgi için Özer Serper'in Uygulamalı İstatistik II (Bursa: 5. Baskı, Ezgi Kitabevi, 2004) adlı kitaptan yararlanabilirsiniz.

KİTAP

HİPOTEZ TESTİ SÜRECİNİN ADIMLARI

Evren parametre değerleri hakkında ileri sürülen iddiaların test edilmesinde başka bir ifadeyle istatistiksel ifadelerin testinde aşağıdaki adımlar izlenir.

Adım 1: Hipotezlerin İfade Edilmesi

İstatistiksel hipotezlerin testinde, iki hipotez söz konusudur. Bunlar; "sıfır hipotezi" ve "karşıt hipotez (alternatif hipotez)" olarak isimlendirilirler. Bu aşamada, sıfır hipotezinin ve karşıt hipotezin nasıl ifade edileceğine karar verilir.

Sıfır hipotezi (H$_0$), ilgili evren parametresinin bilinen değerinde, herhangi bir farklılığın beklenmediğinin ifade edildiği hipotezdir.

Sıfır hipotezi H_0 simgesiyle gösterilir ve hangi hipotezin test edileceğinin ifade edildiği hipotezdir. H_0 hipotezinde test süreci tamamlanıncaya kadar örneklem istatistiği $\hat{\theta}$ değeriyle, θ parametresinin değeri hakkında ileri sürülen θ_0 değeri arasındaki farkın örnekleme hatasından kaynaklanabileceği, bu iki değer arasında gerçekte anlamlı bir farklılık olmadığı, farklılığın istatistiksel olarak, sıfır olduğu; parametrenin önceden belirlenmiş, bilinen değerinde hiçbir farklılığın (etkinin) beklenmediğinin ifade edildiği hipotezdir.

Bu açıklamaların ışığında H_0 hipotezi,

$$H_0 : \theta = \theta_0$$

şeklinde yazılır.

H_0 hipotezi test edilecek hipotez olduğu için test sonucunda verilecek karar da bu hipoteze ilişkin karar olur. H_0 hipotezine ilişkin verilecek karar H_0 kabul veya H_0 red şeklinde olur.

Karşıt hipotez (H$_1$), ilgili evren parametresinin bilinen değerinde, istatistiksel olarak anlamlı farkların beklendiğini ifade eden hipotezdir.

H_0 hipotezinin test edilebilmesinde, bu hipotezden farklı bir hipotezin de ifade edilmesi gerekir. H_1 simgesiyle gösterilen bu hipoteze "**karşıt hipotez**" adı verilir. H_1 hipotezi, H_0 hipotezinin belirli bir olasılıkla reddedilmesi durumunda kabul edilen ve genellikle araştırma hipotezinin ifade edildiği hipotezdir. Karşıt hipotez, parametrenin önceden belirlenmiş, bilinen değerinde anlamlı farklılığın ya da etkinin beklendiğinin ifade edildiği hipotezdir. Bir başka ifadeyle, sıfır hipotezini çürüten bir hipotezdir. Bu hipotez araştırmanın amacına bağlı olarak, aşağıdaki üç farklı şekilden birisiyle ifade edilmiş olur:

$$H_1 : \theta \neq \theta_0$$

veya

$$H_1 : \theta > \theta_0$$

veya

$$H_1 : \theta < \theta_0$$

H_1: $\theta \neq \theta_0$ ifadesi, evren parametresinin belirlenen (ya da bilinen) θ_0 değerinden, her iki yöndeki (hem küçük hem de büyük yöndeki) anlamlı farklılık gösteren örneklem istatistikleri test sonucu verilecek kararı etkileyeceği anlamına gelir. İkinci ifade, H_1: $\theta > \theta_0$ test sonucunda verilecek kararın, evren parametre değerinde, sadece büyük yöndeki anlamlı farklılık gösteren örneklem istatistiklerinden etkileneceği; son ifade ise sadece küçük yöndeki anlamlı farklılığın ($\hat{\theta} - \theta_0$) verilecek kararı etkileyeceği anlamına gelir.

Hipotez testlerinde H_1 hipotezi, testin örnekleme dağılımındaki yönünü ya da H_0 hipotezinin reddedileceği bölgenin (red bölgesinin) yerini belirleyen hipotezdir. Red bölgesi, H_0 hipotezinin reddedilmesine (H_1 hipotezinin kabul edilmesine) neden olan örneklem istatistiği $\hat{\theta}$'nın dağılımında (ya da test istatistiği) ilgili değerler aralığıdır. Kabul bölgesi ise H_0 hipotezinin kabul edilmesine (H_1 hipotezinin reddedilmesine) neden olan örneklem istatistiği $\hat{\theta}$ (test istatistiği) ile ilgili değerler aralığıdır. Hipotez testleri, H_1 hipotezinin ifade ediliş şekline göre: "iki yönlü test", "tek yönlü üst kuyruk testi" ve "tek yönlü alt kuyruk testi" olarak isimlendirilirler. Bu testlere ilişkin hipotezlerin ifade edilmiş biçimi aşağıda verilmiştir.

İki Yönlü Testlerde Hipotezler:

$H_0 : \theta = \theta_0$

$H_1 : \theta \neq \theta_0$

Tek Yönlü Üst Kuyruk Testlerinde Hipotezler:

$H_0 : \theta = \theta_0$

$H_1 : \theta > \theta_0$

Tek Yönlü Alt Kuyruk Testlerinde Hipotezler:

$H_0 : \theta = \theta_0$

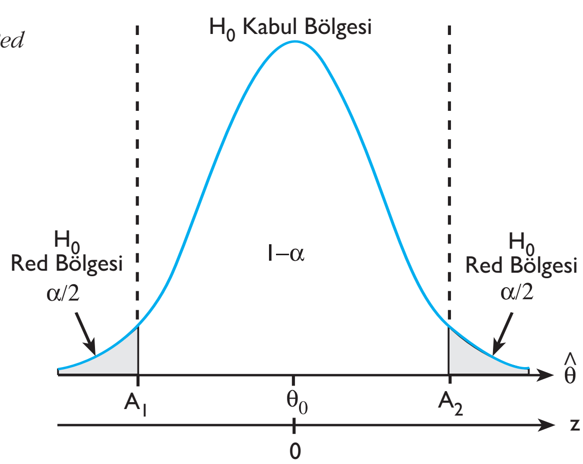
$H_1 : \theta < \theta_0$

şeklinde belirlenir.

Yukarıdaki her hipotez takımında kullanılan isim, H_1 hipotezinde θ için verilen değerler aralığını açıklamaktadır. Bu durum, örneklem istatistiğinin $\hat{\theta}$'nın dağılımının normal dağılıma sahip olduğu kabul edilerek aşağıdaki şekillerle açıklanmıştır. Örneğin iki yönlü hipotezlerde, H_1 hipotezi Şekil 3.1'de görüldüğü gibi θ_0'ın her iki tarafındaki θ ile ilgili değerleri kapsamaktadır. Başka bir ifadeyle, örneklem istatistiği $\hat{\theta}$'nın belirli bir A_1 değerinden küçük ya da belirli bir A_2 değerinden büyük olan değerleri H_1 hipotezi yönünde, H_0 hipotezinin red bölgesinde yer alan değerlerdir.

Şekil 3.1

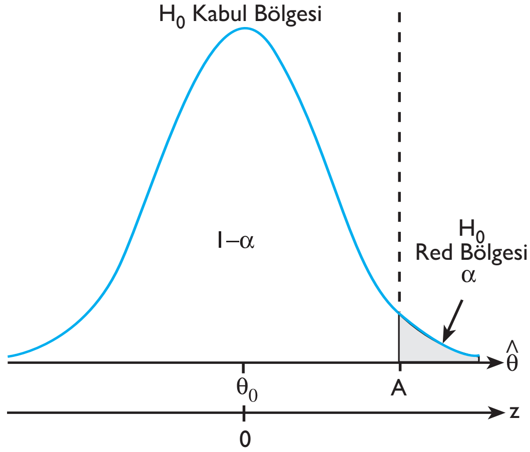
İki Yönlü Testlerde Red Bölgeleri.



Tek yönlü üst kuyruk testlerinde, H_1 hipotezi, θ_0'dan büyük olan θ ile ilgili değerleri içerdiği için, bu isim verilmiştir. Tek yönlü üst kuyruk testlerinde, H_1 hipotezi, Şekil 3.2'de görüldüğü gibi $\hat{\theta}$'nın θ_0'dan büyük olmak üzere, belirli bir A değerinden büyük değerleri H_1 hipotezinin kabul edilmesi yönünde, H_0 hipotezinin red bölgesinde yer almaktadır.

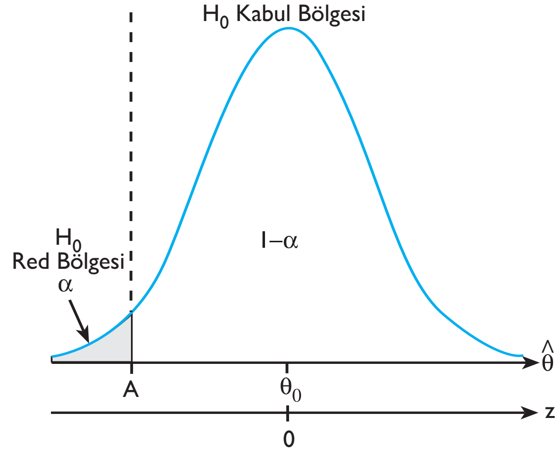
Şekil 3.2

Tek Yönlü Üst Kuyruk Testlerinde Red Bölgesi.



Şekil 3.3

Tek Yönlü Alt Kuyruk Testlerinde Red Bölgesi.



Tek yönlü alt kuyruk testlerindeyse tek yönlü üst kuyruk testinin tam tersine, (Şekil 3.3'te görüldüğü gibi) θ_0'ın solunda ve $\hat{\theta}$'nin A'dan küçük olan değerleri, H_0 hipotezinin red bölgesinde yer alan değerlerdir.

Hipotez testlerinde kabul ya da red edilen hipotez H_0'dır.

DİKKAT

Adım 2: Anlamlılık Düzeyinin Belirlenmesi

Bir istatistiksel hipotez testinde daha önce açıklandığı gibi ya sıfır hipotezinin reddedilmesi ya da kabul edilmesi şeklinde karar verilir. Bu iki karar arasında seçim yaparken örneklem istatistiğinden yararlanıldığı için, hatalı karar verme riski vardır. Çünkü; aynı evrenden rassal olarak seçilen, aynı hacimli farklı örneklemler için hesaplanan istatistikler, örneklemden örnekleme değişen değerler aldığından, evren parametre değerinden farklılık göstermektedirler.

Hipotez testlerinde, sıfır hipotezinin yanlışlıkla reddedilmesi ya da kabul edilmesi sonucu işlenen hataya "yorumlama (çıkarsama) hatası" adı verilir. İki tür yorumlama hatası vardır: Bunlar; gerçekte doğru olan sıfır hipotezinin reddedilmesi durumunda işlenen hatayla, gerçekte yanlış olan sıfır hipotezinin kabul edilmesi durumunda işlenen hatadır. Gerçekte doğru olan sıfır hipotezinin reddedilmesi durumunda işlenen hataya, **I. Tip hata ya da α tipi hata** adı verilir. Araştırmalarda α tipi hata işlemenin maksimum olasılığına "testin **anlam**lılık **düzeyi**" denir. Anlamlılık düzeyinin belirlenmesi, doğru olan sıfır hipotezinin, örneklemden elde dilen bilgilere dayanarak reddedilmesi olasılığını belirleyen α'nın seçilmesi işlemidir. α anlamlılık düzeyi, araştırmacı tarafından, hipotezler ifade edilip veri derlemeye başlamadan önce seçilmesi etik gerekliliktir. Sosyal bilim araştırmalarında α için genellikle %5 veya %1 değerleri seçilmektedir. Yapılan bu seçimle birlikte, doğru olan H_0 hipotezinin reddedilme olasılığı, belirlenmiş olur. Bu olasılık örnekleme dağılımıyla ilişkilendirilerek kullanılır. Bu durumda, α anlamlılık düzeyi, doğru olan sıfır hipotezinin reddedilmesi olasılığına eşit olan, örnekleme dağılımındaki oransal alanı göstermiş olur. Örnekleme dağılımında, doğru olan sıfır hipotezinin, reddedilmesi olasılığına eşit olan oransal alana "red bölgesi" denir. Örnekleme dağılımının bu bölgesi, sıfır hipotezi doğru olduğunda, beklenmeyen örneklem ista-

α tipi hata yapmanın maksimum olasılığına testin **anlam düzeyi** adı verilir.

tistiği değerlerini temsil eder. Örnekleme dağılımında, red bölgesini tanımlamadan önce, örnekleme dağılımını tanımlamak gerekir. Örneklem istatistiğinin normal dağılımlı olması durumu için red ve kabul bölgeleri Şekil 3.1, 3.2 ve 3.3'te gösterilmiştir. Şekillerdeki A, A_1 ve A_2 noktaları red bölgelerinin başlangıç noktalarıdır.

H_0 doğruyken test sonucunda reddedilirse α (I. tip) tipi hata, H_0 doğru değilken test sonucunda kabul edilirse **β (II. tip) tipi hata** gerçekleşmiş olur.

Diğer taraftan, sıfır hipotezi gerçekte yanlış olabilir ve araştırmacı yanlış olan bu hipotezi kabul ederse yine hatalı karar vermiş olur; bu tür hataya **II. Tip hata ya da β tipi hata** denir. Bu türden hata yapmanın maksimum olasılığı da β ile gösterilir.

İstatistiksel uygulamalarda α tipi hatadan daha çok sakınılır ve genellikle sadece α tipi hata kontrol edilir.

Araştırmada, H_0 hipotezinin doğru olduğuna inanan araştırmacı, α anlamlılık düzeyini çok küçük bir değer olarak seçebilir. H_0 hipotezinin kabul edilmesi riskli ise büyük kayıplara neden oluyorsa, α olasılığı büyük tutulmalıdır.

Örneklem hacmi sabit olduğunda, α tipi hata işlemenin azalması (ya da artması), β tipi hata işleme olasılığının artmasına (ya da azalmasına) neden olur.

Adım 3: Örneklemin Seçilmesi, Verilerin Derlenmesi ve Test İstatistiğinin Belirlenmesi

Bir araştırma planında, hipotezlerin ifade edilmesiyle araştırmanın genel çerçevesi ortaya konur, problem ve değişkenler tanımlanmış olur. İfade edilen hipotezlerin test edilmesi için, α anlamlılık düzeyi belirlendikten sonra, belirlenen evrenden, hangi hacimde rassal örneklem/örneklemler seçileceği kararlaştırılır. Daha sonra da ilgili evrenden belirlenen hacimde rassal örneklem/örneklemler seçilerek tanımlanan değişkenler hakkında veriler derlenir. Bu veriler kullanılarak, test edilecek parametre hakkında bilgi üreten örneklem istatistikleri hesaplanır.

Daha önce de belirtildiği gibi evrenden rassal örneklem alınmış olsa bile, hesaplanan örneklem istatistiğinin evren parametresi hakkında, önceden bilinen, belirlenen değere eşit olması beklenmez. Bu durumda şu soru akla gelebilir: Örneklem istatistiğinin değeriyle bu istatistiğin bilgi ürettiği parametrenin sıfır hipotezinde ifade edilen değeri arasında nasıl bir farklılık vardır? Başka bir ifadeyle, sıfır hipotezi doğruysa anlamsız bir farklılığı veren bir örneklem istatistiği elde etmek mümkün müdür?

Hipotez testlerinde, **örneklem** istatistiğinin dağılımının bilinmesi zorunludur.

Bu sorunun yanıtlanabilmesi için sıfır hipotezinin test edilebilmesinde, **örneklem** istatistiğinin dağılımının özelliklerinin bilinmesine ve bu özelliklere bağlı olarak belirlenen uygun test istatistiğine gereksinim vardır.

Test istatistiği, örneklem istatistiğinin değeriyle evrenin, sıfır hipotezinde ifade edilen değeri arasındaki farkın, standartlaştırılmış değeri olarak tanımlanır. Başka bir ifadeyle test istatistiği, örneklem istatistiği θ_0 ile $\hat{\theta}$ arasındaki farkı standart hata birimiyle ifade eden ölçüdür ve $\dfrac{\hat{\theta}-\theta_0}{\sigma_{\hat{\theta}}}$ şeklinde ifade edilir. Bu test istatistiği örneklemin sıfır hipotezine ne kadar uyduğunu gösterir. Bu nedenle de test istatis tiği test sonunda verilecek kararın dayandırıldığı bir örneklem istatistiğidir.

Bir örneklem istatistiğinin değeri, bu örneklem istatistiğinin dağılımının bir değeridir. Mümkün her örneklem istatistiğinin değeri için, bir test istatistiği değeri hesaplanabileceğine göre, test istatistiği örnekleme dağılımından söz edilebilir. Test istatistikleri genellikle normal dağılım (z dağılımı), t dağılımı vb. gibi bilinen dağılımlara uyar.

Hipotez testi türleriyle ilgili bilgiler verilirken açıklandığı gibi, hipotez testleri için de uygun test istatistiğinin seçilmesi konusunda ilgilenilen değişkenlerin ölçülmesinde kullanılan ölçek türü, örneklem hacmi, örneklem sayısı (örneklem sayısı iki olduğunda örneklemlerin bağımsız ya da bağımlı olması) gibi hususların bilinmesi gerekir.

İzleyen bölümde, bazı parametrelere ilişkin hipotezlerin testinde z, t ve F test istatistiklerinin seçilme gerekçeleri ve uygulamalarına yer verilmiştir.

Adım 4: İstatistiksel Kararın Verilmesi

İstatistiksel karar vermekle eş anlamlı olan hipotez testi, aslında α anlamlılık düzeyinde H_0 hipotezinin kabul edilmesi ya da reddedilmesi kararıdır. Bu kararın verilebilmesi için bir ölçütün belirlenmesi gerekir. Test istatistiğinin, kritik değeri olarak isimlendirilen bu ölçüt, istatistiğin örnekleme dağılımında, red ve kabul bölgelerini birbirinden ayıran bir değerdir. Test istatistiğinin kritik değeri, bir örnekleme dağılımında, red bölgesinin başlama noktasını gösteren değerdir. Kritik değer, seçilen α anlamlılık düzeyinde, H_1 hipotezinin ifade ediliş biçimine ve örneklem istatistiğinin dağılım şekline bağlıdır. İzleyen açıklamalar $\hat{\theta}$ örneklem istatistiğinin ve bu istatistiğin standart değeri olan

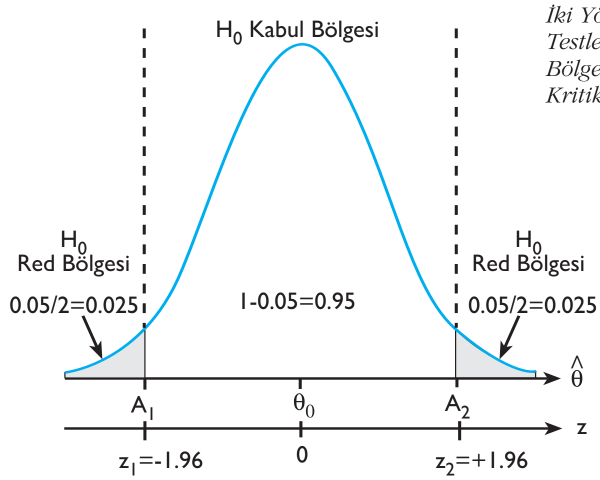
$$z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$$

test istatistiğinin standart normal dağılıma sahip olduğu kabul edilerek yapılmıştır. Açıklamalarda α = 0.05 seçilmiştir.

Eğer karşıt hipotez $H_1 : \theta \neq \theta_0$ şeklinde ifade edilmişse red bölgesi Şekil 3.4'te gösterildiği gibi $\hat{\theta}$ istatistiğine ilişkin normal dağılımın her iki ucunda simetrik olarak tanımlanmış olur ve her red bölgesinin alanı oransal olarak $\alpha/2$ = 0.05/2 = 0.025'tir. Buna bağlı olarak kritik değerler $\hat{\theta}$ istatistiğine ilişkin normal dağılımın her iki kuyruğundaki, θ_0'a göre simetrik, A_1 ve A_2 değerleri olmaktadır.

Şekil 3.4

İki Yönlü Testlerde Red Bölgeleri ve Kritik Değerler.



Ancak istatistiksel hipotez testlerinde, $\hat{\theta}$ örneklem istatistiği yerine bu istatistiğin standartlaştırılmış değeri kullanılmaktadır. Bu durumda kritik değerler A_1 ve A_2'nin

$$z_1 = \frac{A_1 - \theta_0}{\sigma_{\hat{\theta}}} \quad \text{ve} \quad z_2 = \frac{A_2 - \theta_0}{\sigma_{\hat{\theta}}}$$

standart değerleri olur. A_1 örneklem değeri θ_0'ın solunda (A_1 θ_0'dan küçük değerli) olduğu için z_1 negatif değer olarak ifade edilir. A_2 örneklem değeri θ_0'ın sağında (A_2 θ_0'dan büyük değerli) olduğu için, z_2 pozitif değer olarak ifade edilir. z=0'a göre simetrik olan z_1 ya da z_2 kritik değerleri α = 0.05 anlamlılık düzeyi için Ek-1'de verilen Standart Normal Eğri Alanları Tablosundan yararlanılarak belirlenir. Bu kritik değerler z tablo değerleri (z_{tab}) şeklinde ifade edilirse,

$$z_{tab} = z\left(0.5 - \frac{\alpha}{2}\right) = \left(0.5 - \frac{0.05}{2}\right) = z(0.4750) = \pm 1.96\text{'dır.}$$

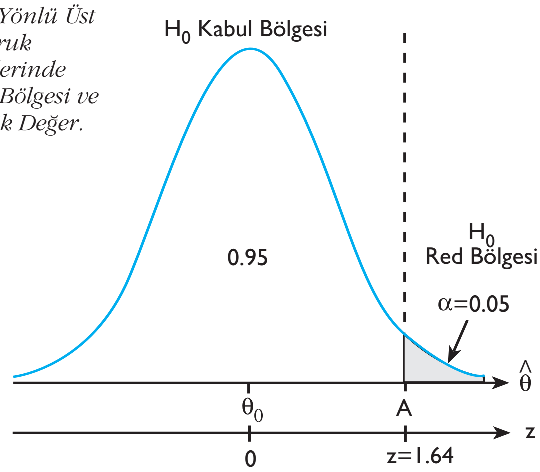
z_{tab}= ±1.96 değeri standart normal dağılımda %47.5'lik oransal alana karşı gelen örneklem istatistiğinin standart değeridir. İki yönlü testte H_0 hipotezinin reddedilmesi için,

$$z_{hes} = \left|\frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}\right| > z_{tab} = 1.96$$

koşulunun sağlanması gerekir. Tersi durumda H_0 hipotezi kabul edilir.

Şekil 3.5

Tek Yönlü Üst Kuyruk Testlerinde Red Bölgesi ve Kritik Değer.



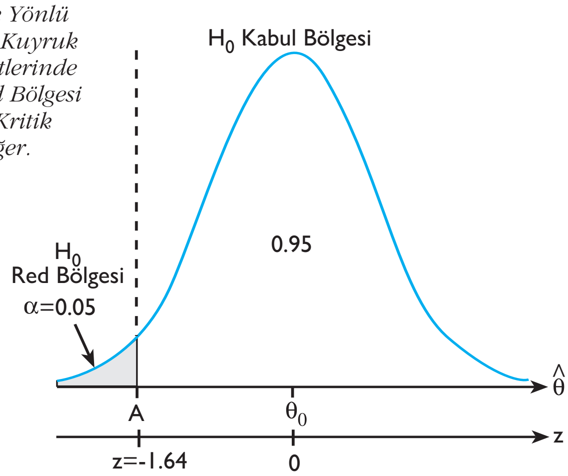
Eğer $H_1 : \theta = \theta_0$ H_0 hipotezinin red bölgesi, $\hat{\theta}$ istatistiğine ilişkin normal dağılımın üst kuyruğunda, $H_1 : \theta < \theta_0$ şeklinde ifade edilmişse $\hat{\theta}$ istatistiğine ilişkin normal dağılımın alt kuyruğunda tanımlanmış alan α=0.05 olur. Buna bağlı olarak kritik değerler sırasıyla (Şekil 3.5'te gösterildiği gibi) $\hat{\theta}$ istatistiğine ilişkin normal dağılımın üst kuyruğundaki A değeri ya da bunun standart değeri

$$z = \frac{A - \theta_0}{\sigma_{\hat{\theta}}}$$

Şekil 3.6'da gösterildiği gibi olur. Bu z değerleri birbirinin simetriğidir. Üstkuyruk testinde, z pozitif altkuyruk testinde, z negatif işaretlidir. z değerleri α=0.05 için Ek-1'de verilen Standart Normal Eğri alanları tablosundan yararlanılarak belirlenirler.

Şekil 3.6

Tek Yönlü Alt Kuyruk Testlerinde Red Bölgesi ve Kritik Değer.



$$z_{tab} = z(0.5 - 0.05) = z(0.4500) = 1.64$$

z_{tab} = 1.64 değeri, standart normal dağılımda, %45'lik oransal alana karşı gelen, örneklem istatistiğinin standart değeridir. Tek yönlü üst kuyruk testi söz konusu olduğunda z_{tab} = 1.64, tek yönlü alt kuyruk söz konusu olduğunda z_{tab} = 1.64 alınır. Bu bilgilere göre, H_0 hipotezinin reddedilmesi için tek yönlü üstkuyruk testinde;

$$z_{hes} = \left(\frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}\right) > (z_{tab} = 1.64)$$

olmalıdır.

Tek yönlü alt kuyruk testinde H_0'ın reddedilebilmesi için

$$z_{hes} = \left(\frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} \right) < (z_{tab} = -1.64)$$

koşulunun sağlanması gerekir. Tersi durumda H_0 hipotezi kabul edilir.

H_0 hipotezinin reddedilmesi yönündeki kararlar, örneklem değeri $\hat{\theta}$ ile evren parametresi arasında, α anlamlılık düzeyinde anlamlı bir farklılığın var olduğunu, H_0 hipotezinin kabul edilmesi durumundaysa varolan farklılığın örnekleme hatasından kaynaklandığı anlamına gelir.

Adım 5: Probleme İlişkin Kararın Verilmesi

Hipotez testlerinde önemli olan, istatistiksel kararın, araştırma problemine ilişkin karara dönüştürülmesidir. Bu konu örnek problemler üzerinde açıklanmıştır.

SIRA SİZDE 2

- **I. Tip hata ne demektir?**
- **Bir hipotez sınaması hangi durumda çift yönlü olarak, hangi hipotezle ve nasıl ifade edilir?**

TEK EVREN PARAMETRESİYLE İLGİLİ HİPOTEZ TESTLERİ

Bu testlerde karar verilirken örneklem istatistiğinin değeriyle bu istatistiğin bilgi ürettiği parametrenin bilinen ya da belirlenen değeri (θ_0) karşılaştırılır. İzleyen bölümlerde, uygulamada sıkça karşılaşılan tek evren parametresiyle ilgili olarak, evren ortalamasına ve evren oranına ilişkin testler ele alınmıştır.

Evren Ortalamasına İlişkin Hipotez Testi

Bu teste, tanımlanan evrenden rassal olarak seçilen bir örneklem için hesaplanan $\bar{X}$ değeriyle, bu örneklemin seçildiği evrenin aritmetik ortalaması μ ile ilgili, önceden belirlenen (ya da bilinen) μ_0 gibi bir değer arasındaki farklılığın istatistiksel olarak anlamlı olup olmadığı araştırılır. Bir başka ifadeyle, örneklem ortalaması $\bar{X}$'nın belirli bir standart hata ile evren parametresiyle ilgili bilinen veya standart değer olarak belirlenen μ_0'dan ne kadar uzaklıkta (farklı) olduğunun ölçülmesidir.

Evren ortalamasına ilişkin hipotez testi uygulamada, sıkça kullanılan önemli bir parametrik testtir. Bu hipotez testlerine ilişkin açıklamalar, örneklem hacminin büyük olması ($n \geq 30$ birim) ve örneklem hacminin küçük olması ($n < 30$ birim) durumları için iki alt başlıkta, aşağıdaki örnek problemler üzerinde ayrıntılı olarak ele alınmıştır.

Evren Ortalamasına İlişkin Büyük Örneklem Testi

Bu test türünde:

- Örneklem rassal olarak seçilir.
- Örneklem hacminin yeterli büyüklükte ($n \geq 30$) birimden oluştuğu ya da evrenin dağılımı normal ve değişkenliğinin biliniyor olması gereklidir.
- $H_0 : \mu = \mu_0$ hipotezi, seçilecek bir α anlamlılık düzeyi için test edilir.

ÖRNEK 1

Margarin üreten bir fabrikada 250 gramlık paketler hâlinde üretim yapılması öngörülmektedir. Margarin paketlerinin ağırlığını kontrol amacıyla rassal olarak 100 paket seçilmiş ve seçilen bu paketler için ortalama ağırlık 244 gram, standart sapma 18 gram olarak saptanmıştır. α = 0.05 anlam düzeyinde, paketlerin ağırlığının öngörüldüğü gibi belirlenen standartta olduğu söylenebilir mi? Karar veriniz.

Çözüm 1:

Adım 1: Hipotezlerin ifade edilmesi

Margarin paketlerinin öngörülen, belirlenen standart ağırlığı 250 gramdır. Bu nedenle sıfır hipotezi, üretilen margarin paketlerinin ağırlığının $\mu_0 = 250$ gram olduğu yönündedir. Paketlerin 250 gramdan hem küçük hem de büyük yöndeki anlamlı ağırlık farklılıkları paketlemenin öngörüldüğü, planlandığı gibi gerçekleşmediğini göstereceği için yapılacak test iki yönlü test olmalıdır. Bu açıklamalara göre, Hipotezler:

$H_0 : \mu = 250$ gr.

$H_1 : \mu \neq 250$ gr.

şeklinde ifade edilmiştir. Test edilecek μ parametresi hakkında bilgi üreten istatistik $\bar{X}$ olduğundan testin red bölgesi H_1 hipotezine göre şekil 3.7'de gösterildiği gibi $\bar{X}$'nın örnekleme dağılımının her iki kuyruğunda tanımlanmıştır.

Adım 2: Anlamlılık Düzeyinin Belirlenmesi

Bu örnekte araştırmacı $H_0 : \mu = \mu_0 = 250$ gram olarak ifade edilen istatistiksel hipotezin gerçekte doğru olması durumunda bu hipotezin yanlışlıkla reddedilmesi olasılığını (riskini) α=0.05 olarak belirlemiştir. Karar verici α=0.05 belirlemekle mümkün örneklem istatistiklerinin %5'inin bu istatistiğin dağılımının red bölgesi içinde yer almasını benimsemektedir.

Problemde doğru olan H_0 hipotezinin reddedilmesi olasılığı α = 0.05 olarak belirlenmiştir. Red bölgeleri, ortalamanın örnekleme dağılımının her iki kuyruğunda tanımlandığı için, red bölgelerinin her birinin oransal büyüklüğü, $\alpha/2 = 0.05/2 = 0.025$'tir.

Adım 3: Örneklemin Seçilmesi, Verilerin Derlenmesi ve Test İstatistiğinin Hesaplanması

Testin yapılabilmesi için hacmi n=100 paketten oluşan rassal bir örneklem seçilmiştir. Örneklemdeki paketler tek tek tartılmış ve derlenen veriler kullanılarak μ parametresi hakkında bilgi üreten örneklem istatistiği $\bar{X}$ ile test için gerekli aşağıdaki istatistikler hesaplanmıştır.

$n = 100$ paket

$\bar{X} = 244$ gr.

$s = 18$ gr.

$\mu_0 = 250$ gr.

$\alpha = 0.05$

$$s_{\bar{\mu}} = \frac{s}{\sqrt{n}} = \frac{18}{\sqrt{100}} = 1.8$$

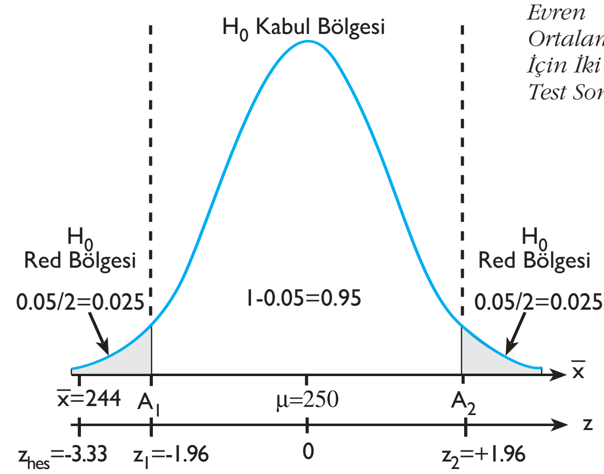
Bu örnekte tanımlanan evren sonsuzdur. Evrenin dağılım şekli ve değişkenliği hakkında bilgi bulunmamaktadır. Örneklem hacmi n=100 paket olup (n ≥ 30 birim olduğundan) örneklem ortalaması $\overline{X}$'nın örnekleme dağılımı normal dağılım gösterir. Bu nedenle tek evren ortalaması μ'ye ilişkin bu testte kullanılması gereken test istatistiği

$$z = \frac{\overline{X} - \mu_0}{s_{\overline{X}}}$$

olur. $\mu = \mu_0$ olduğu zaman z test istatistiğinin dağılımı standart normal dağılım özelliklerine sahiptir. Burada σ bilinmediği ve n ≥ 30 birim olduğu için $s_{\overline{X}}$, $\sigma_{\overline{X}}$'nın yansız tahminleyicisi olarak kullanılmıştır. Bu bilgilere göre test istatistiğinin değeri

$$z = \frac{\overline{X} - \mu_0}{s_{\overline{X}}} = \frac{244 - 250}{1.8} = -3.33$$

olarak hesaplanmış olur.



Şekil 3.7

Evren Ortalaması μ İçin İki Yönlü Test Sonuçları

Adım 4: İstatistiksel Kararın Verilmesi

$\overline{X}$'nın örnekleme dağılımı normal olduğu için, bu dağılımı oluşturan değerlerin standart değerleri olan z test istatistiğinin örnekleme dağılımı, standart normal dağılım gösterir. İki yönlü bir test olduğu için testin red bölgesi şekil 3.7'de gösterildiği gibi $\overline{X}$'nın dağılımının her iki kuyruğunda tanımlanmıştır ve oransal büyüklükleri $\alpha/2 = 0.05/2 = 0.025$'tir. Buna göre, dağılımın sınır değerleri $z_{tab}(0.5-\alpha/2) = z_{tab}(0.5-0.025) = z_{tab}(0.4750) = \pm 1.96$'dır yani alt kuyruk bölgesi için $z_1 = -1.96$ ve üst kuyruk bölgesi için $z_2 = 1.96$ olacaktır. Bu değerlere kritik değerler adı da verilmektedir.

Hesaplanan test istatistiği mutlak değer olarak $z_{hes} = |3.33| > z_{tab} = 1.96$ (veya $z_{hes} = -3.33 < z_{tab} = -1.96$) olduğundan H_0 hipotezi reddedilir, dolayısıyla H_1 kabul edilir. Ayrıca bu karar Şekil 3.7'de $z_{hes} = -3.33$ standart değerinin z_1'in solundaki red bölgesinde yer aldığını belirterek de açıklanabilir.

Şekil 3.7 deki A_1 ve A_2 değerleri aşağıdaki gibi hesaplanır:

$A_1 = \mu_0 - z \,.\, s_{\overline{X}} = 250 - 1{,}96 \,.\, 1{,}8 = 246{,}47$

$A2 = \mu_0 + z \,.\, s_{\overline{X}} = 250 + 1{,}96 \,.\, 1{,}8 = 253{,}53$

Adım 5: Probleme İlişkin Kararın Verilmesi

H_0 hipotezinin reddedilmesi, üretilen margarin paketlerinin ortalama ağırlığının 250 gram olmadığını, bu değerden anlamlı bir biçimde farklılık olduğunu, paketleme üretim sisteminin planlandığı şekilde üretim yapmadığını gösterir.

ÖRNEK 2

Bir firmanın geliştirdiği yeni sistemin ortalama paketleme süresini ürün başına 15 dakikanın altına indirdiği iddia edilmektedir. Bu iddiayı araştırmak amacıyla paketleme esnasında rassal olarak seçilen 225 ürünün yeni sistemde ortalama paketlenme süresi 13 dakika ve standart sapması 4.2 dakika olarak belirlenmiştir. Yeni sistemle ilgili iddia hakkında α *= 0.01 anlam düzeyinde karar veriniz.*

Çözüm 2:

Adım 1: Hipotezlerin İfade Edilmesi

Burada verilecek karar, "önceki üretim sisteminin bilinen üretim başına ortalama paketleme süresi μ_0 = 15 dakika, yeni üretim sistemi uygulamaya girince düşmüş müdür?" başka bir ifadeyle "$\bar{X}$ ve μ_0 arasında belirlenen bir anlamlılık düzeyi için azalma yönünde anlamlı bir farklılık var mıdır?" sorusunun yanıtı olmalıdır. Bu bilgilere göre hipotezler:

$H_0 : \mu = 15$ dk.

$H_1 : \mu < 15$ dk.

şeklinde ifade edilecektir.

Bu hipotez testi H_1 hipotezinin ifade ediliş şekline göre tek yönlü alt kuyruk testidir. Çünkü $H_1 : \mu < 15$ dakika olarak ifade edilmiştir. Testin red bölgesi $\bar{X}$'nın örnekleme dağılımının sol ucunda (μ_0'ın solunda) tanımlanmıştır.

Adım 2: Anlamlılık Düzeyinin Belirlenmesi

Örnekte doğru olan H_0 hipotezinin reddedilmesi olasılığı α = 0.01 olarak belirlenmiştir. H_0 hipotezinin kabul bölgesinin oransal büyüklüğü Şekil 3.8'de gösterildiği gibi 1 - α = 0.99 iken red bölgesinin oransal büyüklüğü α = 0.01'dir.

Adım 3: Örneklemin Seçilmesi, Verilerin Derlenmesi ve Test İstatistiğinin Hesaplanması

Yukarıda ifade edilen testin yapılabilmesi için yeni paketleme üretim sürecinden rassal olarak n = 225 (n ≥ 30) ürün seçilmiş, bu ürünlerin her birinin paketleme süresi ölçülmüştür. Bu ölçüm sonuçları (veriler) kullanılarak hesaplanan istatistikler ve veriler aşağıda gösterilmiştir.

$\mu_0 = 15$ dk

$\bar{X} = 13$ dk

n = 225 ürün

s = 4.2 dk.

$\alpha = 0.01$

$$s_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{4.2}{\sqrt{225}} = 0.28$$

Yukarıdaki veri ve bilgilerden yararlanılarak test istatistiğinin önce belirlenmesi sonra bu istatistiğin hesaplanması gerekir. Örneklem hacmi n=225 ürün (n ≥ 30 birim) olduğu için $\bar{X}$'nın örnekleme dağılımı normaldir. $\mu = \mu_0$ = 15 dakika olduğu zaman $\bar{X}$ = 13 dakika değerinin standart hata ($s_{\bar{X}}$) birim cinsinden bilinen μ_0 = 15 dakika değerinden ne kadar farklılık gösterdiğini ölçen test istatistiği z test istatistiği olur. z test istatistiği aşağıdaki gibi yazılır ve hesaplanır;

$$z = \frac{\bar{X} - \mu_0}{s_{\bar{X}}} = \frac{13 - 15}{0.28} = -7.142$$

Burada

$$s_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{4.2}{\sqrt{225}} = 0.28$$

olarak hesaplanmıştır.

Adım 4: İstatistiksel Kararın Verilmesi

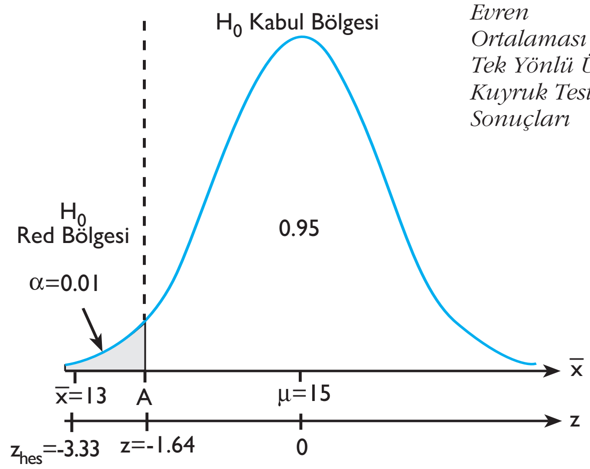
İstatistiksel kararın verilebilmesi için hesaplanan test istatistiği z_{hes} = –7.142 ile bu istatistiğin karşılaştırılacağı α=0.01 anlamlılık düzeyindeki kritik değerin z_{tab} (0.5–0.01) = z_{tab}(0.4900) = 2.33 standart normal eğri alanları tablosundan belirlenmesi gerekir. Bu bilgilere göre istatistiksel karar,

$$z_{hes} = -7.142 < z_{tab}(0.4900) = -2.33$$

veya

$$z_{hes} = |7.142| > z_{tab}(0.4900) = 2.33$$

olduğu için H_0 hipotezi red, H_1 hipotezi kabul edilir yönünde karar verilir. Şekil 3.8'deki A değeri;



Şekil 3.8

Evren Ortalaması İçin Tek Yönlü Üst Kuyruk Testi Sonuçları

$$A = \mu_0 - z \,.\, s_{\bar{X}} = 15 - 2{,}33 \,.\, 0{,}28 = 14{,}35$$

bulunur.

Adım 5: Probleme İlişkin Kararın Verilmesi

Bu firmanın geliştirdiği yeni sistem, ortalama paketleme süresini ürün başına 15 dakikanın altına indirmektedir.

Evren Ortalamasına İlişkin Küçük Örneklem Testi

Araştırmaların bir çoğunda araştırmaya ayrılan para, zaman ve diğer imkânların sınırlı olması gibi nedenlerle, örneklem hacmini, daha önceki açıklamalarımızda belirtilen büyüklükte (genellikle n ≥ 30 birim) olmayabilir. Örneğin; çok nadir görülen bir hastalıkla ilgili araştırmada vaka sayısını, uzun süren deneylere dayanan araştırmalarla ve maliyeti yüksek olan laboratuar çalışmalarıyla örneklem hacmini arttırmak çok güçtür. Örneklem hacminin az olduğu bu gibi durumlarda, küçük örneklemler için geliştirilmiş test yöntemlerine başvurulur. Bu bölümde, tek evren ortalaması için kurulan hipotezlerin, küçük örneklemler (n < 30 birim) kullanılarak, nasıl test edileceği konusu ele alınmıştır.

Önceki bölümde açıklanan tek evren ortalamasına ilişkin büyük örneklem testinde, sıfır hipotezinin testi için örnekleme dağılımı olarak, normal dağılım kullanılmıştı. Çünkü örneklem hacmini en az 30 birim olması ya da evren dağılımının normal ve değişkenliği σ'nın biliniyor olması durumları, göz önüne alınmıştı.

Evren standart sapması bilindiğinde, ortalamanın örnekleme dağılımı ortalaması μ ve standart sapması (standart hatası) $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$ olan, normal dağılımı gösterir.

Genellikle σ bilinmez. Araştırmacı tek evren ortalamasına ilişkin hipotez testi için σ yerine onun tahmini olan örneklem standart sapması s'yi kullanarak ortalamanın örnekleme dağılımının standart hatası ($s_{\overline{X}}$) tahminlenir. Bu durumda, ortalamanın standart hata tahmini ($s_{\overline{X}}$) aşağıdaki gibi yazılır:

$$s_{\overline{X}} = \frac{s}{\sqrt{n}}$$

ve büyük bir hata işlenmemiş olur.

Örneklem hacminin küçük olması durumunda, σ yerine s'nin kullanılması istatistiksel test üzerinde etkili olur. Çünkü; σ yerine s'nin kullanılması durumunda tahmin edilen istatistik $\overline{X} - \mu_0 / s_{\overline{X}}$ standart normal dağılım göstermemekte, dolayısıyla büyük örneklemlerde izlenen yöntem geçerli olmamaktadır. Normal dağılıma sahip ve değişkenliği bilinmeyen bir evrenden seçilen 30'dan daha az birim içeren bir rassal örneklemin aritmetik ortalamasının örnekleme dağılımının $\overline{X} - \mu_0 / s_{\overline{X}}$ standart değerlerinin dağılımı n–1 serbestlik derecesiyle t dağılır. t istatistiği,

$$t = \frac{\overline{X} - \mu_0}{s_{\overline{X}}}$$

şeklindedir. Burada $s_{\overline{X}}$, örneklem ortalamasının standart hata tahminini gösterir ve

$$s_{\overline{X}} = \frac{s}{\sqrt{n-1}}$$

eşitliği ile hesaplanır.

t dağılımı da normal dağılım gibi simetrik bir dağılımdır ve örneklem hacmi büyüdükçe normal dağılıma yaklaşır.

Küçük örneklem kullanılarak yapılan tek evren ortalamasına ilişkin hipotez testleri, kullanılan test istatistiği dışında tek evren ortalamasına ilişkin büyük örneklem testlerine benzemektedir. Aşağıdaki örnek problem üzerinde bu testin uygulanış biçimi test sürecinin adımları itibarıyla açıklanmıştır.

Tek evren ortalamasına ilişkin büyük örneklem testinde olduğu gibi küçük örneklem testinde de örneklem aritmetik ortalaması $\overline{X}$ ile evrenin ortalaması hakkında daha önceden bilinen ya da belirlenen bir değer μ_0 arasındaki farkın istatistiksel olarak anlamlı olup olmadığı araştırılır.

ÖRNEK 3

Bir su dolum sisteminde doldurulan damacanalardaki su miktarının 22 litre olması planlanmıştır. Damacanalardaki su miktarının belirlenen standartta olup olmadığını kontrol etmek amacıyla rassal olarak 17 damacana seçilerek, bunlardaki ortalama su miktarı 21.7 litre ve standart sapması 0.8 litre olarak hesaplanmıştır. Su dolum sisteminde damacanaların olması gerektiği gibi doldurulup doldurulmadığı, dolum sisteminin doğru çalışıp çalışmadığı hakkında 0.01 anlamlılık düzeyinde karar veriniz.

Çözüm 3:

Adım 1: Hipotezlerin İfade Edilmesi

Damacanalarda olması gereken su miktarı standardı 22 litre olarak belirlenmiştir. Buna göre sıfır hipotezi, damacanalardaki ortalama su miktarının 22 litreye eşit olduğu yönündedir. Bu iddiayı ortalama 22 litreden hem az hem de fazla su miktarı farklılıkları çürütecektir. Bu nedenle yapılacak test iki yönlü test olacak ve hipotezler,

H_0 : μ = 22 lt.

H_1 : $\mu \neq$ 22 lt.

şeklinde yazılacaktır.

Testin red bölgeleri $\overline{X}$'nın örnekleme dağılımının her iki kuyruğunda tanımlanmalıdır.

Adım 2: Anlamlılık Düzeyinin Belirlenmesi

Problemde doğru olan H_0 hipotezinin reddedilmesi olasılığı α = 0.01 olarak belirlenmiştir. H_0 hipotezinin kabul bölgesinin oransal büyüklüğü $1-\alpha$= 0.99 iken red bölgelerinin her birinin oransal büyüklüğü α = 0.01/2 olur.

Adım 3: Örneklemin Seçilmesi, Verilerin Derlenmesi ve Test İstatistiğinin Hesaplanması

Su dolum sisteminde doldurulan damacanalar arasından rassal olarak n=17 damacana seçilmiş ve her birindeki su miktarı ölçülmüştür. Örneklemdeki damacanalardaki su miktarı ile ilgili hesaplanan bilgiler ve bazı veriler aşağıda verilmiştir:

n = 17 damacana

$\overline{X}$ = 21.7 lt

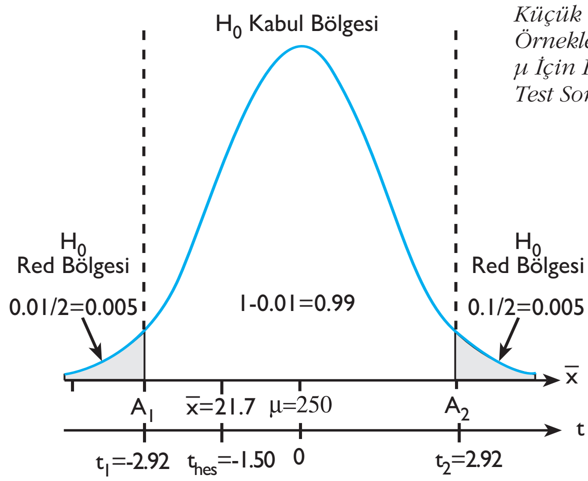
s = 0.8 lt

μ_0 = 22 lt

α = 0.01

Şekil 3.9

Küçük Örneklemlerde μ İçin İki Yönlü Test Sonuçları.



Örneklem hacmi n=17 damacanadır. Kullanılması gereken test istatistiği küçük örneklem t testidir. t test istatistiği,

$$t = \frac{\overline{X} - \mu_0}{s_{\overline{X}}} = \frac{21.7 - 22}{0.2} = -1.5$$

biçiminde yazılır ve hesaplanır.

$$s_{\overline{X}} = \frac{s}{\sqrt{n-1}} = \frac{0.8}{\sqrt{16}} = 0.2$$

şeklinde hesaplanır.

Adım 4: İstatistiksel Kararın Verilmesi

H_0 hipotezi iki yönlü test edilecektir. Red bölgeleri, t dağılımının hem alt kuyruğunda hem de üst kuyruğunda tanımlanmıştır. Bu durum Şekil 3.9'da gösterilmiştir.

α = 0.01 ve n = 17 için t tablo değeri (t_{tab})

t_{tab} = t ($1-\alpha/2$; ν = n–1) = 2.921 olarak belirlenir.

t_{hes} = |1.5|>t_{tab} = 2.921 olduğundan H_0 hipotezi kabul edilir, dolayısıyla H_1 hipotezi reddedilir. Şekil 3.9'da görüldüğü gibi hesaplanan test istatistiği değeri, kabul bölgesinde yer almaktadır.

Adım 5: Probleme İlişkin Kararın Verilmesi

Yukarıda verilen istatistiksel karara göre, su dolum sisteminde damacanalar olması gerektiği gibi 22 litre olarak doldurulmaktadır. Örneklem aritmetik ortalaması $\overline{X}$ = 21.7 litre ile evren aritmetik ortalaması μ_0 = 22 arasında anlamlı bir farklılık bulunmamıştır. $\overline{X} - \mu_0$ =21.7 –22 = –0.03 litrelik fark rassal bir farktır. Bu fark, "istatistiksel olarak sıfır (0)'dır" diye yorumlanır.

Şekil 3.9'da A_1 ve A_2;

$A_1 = \mu_0 - t \,.\, s_{\overline{X}} = 22 - 2{,}92 \,.\, 0{,}2 = 21{,}416$

$A2 = \mu_0 + t \,.\, s_{\overline{X}} = 22 + 2{,}92 \,.\, 0{,}2 = 22{,}584$

bulunur.

Evren Oranına İlişkin Hipotez Testi

Pek çok araştırmada ilgilenilen değişken, iki şıklı ya da iki şıkka indirgenmiş değişken olabilir. Örneğin Anadolu Üniversitesi öğrencileri, cinsiyet değişkeni bakımından erkek kadın, başarı değişkeni bakımından da başarılı başarısız olmak üzere iki şıklıdır.

Daha önce de açıklandığı gibi bir evrenin, ilgilenilen iki şıklı bir değişkeninin herhangi bir şıkkına sahip birimlerinin oranına "evren oranı" denir ve $\prod$ simgesiyle gösterilir. Ünitenin bu kesiminde, evren oranı $\prod$'nin değeri hakkında ileri sürülen bir önermenin, nasıl test edileceği konusu ele alınmıştır. Tek evren oranına ilişkin test olarak isimlendirilen bu testte, örneklem oranı p ile evren oranı $\prod$'nin iddia edilen değeri $\prod_0$ arasındaki farkın istatistiksel olarak anlamlı olup olmadığı araştırılır. Örneklem hacmi yeterli büyüklükte olduğunda (n ≥ 30 birim), evren oranı $\prod$ hakkındaki testler daha önce açıklanan evren ortalaması μ için büyük örneklem testlerindekilere benzer şekilde yapılır. Ancak, test için örneklem istatistiği olarak örneklem oranı p ve bu istatistiğin örnekleme dağılımı kullanılır. n ≥ 30 olduğunda, örneklem oranı p'nin örnekleme dağılımı, yaklaşık normal dağılıma sahip olur. Bu durumda p örneklem istatistiği dağılımına ilişkin standart değerlerin dağılımının da normal olacağı açıktır.

Eğer örneklem oranı p ile evren oranı Π'nin ileri sürülen Π_0 değeri arasındaki farkın, istatistiksel olarak anlamlı olup olmadığı araştırılıyorsa evren oranına ilişkin test uygulanır ve test **istatistiği** $z = \frac{p - \Pi_0}{\sigma_p}$ 'dir.

Evren oranı $\prod$'ye ilişkin testlerde örneklem hacmi büyük olduğunda, aşağıdaki z **istatistiği** kullanılır:

$$z = \frac{p - \Pi_0}{\sigma_p}$$

Burada, σ_p örneklem oranı, p'nin örnekleme dağılımının standart sapmasını gösterir ve

$$\sigma_p = \sqrt{\frac{\Pi_0(1 - \Pi_0)}{n}}$$

eşitliği ile hesaplanır.

ÖRNEK 4

Özel bir dershane yetkilisi, üniversiteye giriş sınavına hazırlık için dershanesine gelen öğrencilerden, üniversitede istediği bölümü kazananların oranının %80'den fazla olduğunu iddia etmektedir. Bu iddiayı araştırmak amacıyla söz konusu dershaneye giden ve üniversiteyi kazanan öğrenciler arasından rassal olarak seçilen 120 öğrenciye istedikleri bölümü kazanıp kazanmadıkları sorulmuş ve 102 öğrencinin üniversitede istediği bölümü kazandığı öğrenilmiştir. Yetkili iddiasında haklı mıdır? 0.05 anlam düzeyinde karar veriniz.

Çözüm 4:

Adım 1: Hipotezlerin İfade Edilmesi

Bu problemde sıfır hipotezi, söz konusu dershaneye giderek üniversitede istediği bölümü kazananların oranının 0.80'den farklı olmadığı şeklindedir. Karşıt (araştırma) hipotezi ise söz konusu dershaneye giderek üniversitede istediği bölümü kazananların oranının 0.80'den fazla olduğu yönündeki tek taraflı hipotezdir. Bu açıklamalara göre hipotezler;

$H_0 : \Pi = 0.80$

$H_1 : \Pi > 0.80$

şeklinde ifade edilir.

Adım 2: Anlamlılık Düzeyinin Belirlenmesi

Problemde doğru olan H_0 hipotezinin reddedilmesi olasılığı $\alpha = 0.05$ olarak belirlenmiştir. Red bölgesinin oransal büyüklüğü, 0.05'tir ve Şekil 3.10'da gösterildiği gibi p'nin örnekleme dağılımının üst kuyruğunda tanımlanmıştır.

Adım 3: Örneklemin Seçilmesi, Verilerin Derlenmesi ve Test İstatistiğinin Hesaplanması

Seçilen 120 öğrenciden oluşan rassal örneklemde üniversitede istediği bölümü kazananların sayısı 102 kişidir. Bu durumda test edilecek evren oranı Π hakkında bilgi üreten örneklem oranı p;

$$p = \frac{r}{n} = \frac{102}{120} = 0.85$$

dir. Örneklem hacmi n = 120 (n ≥ 30) olduğu için p'nin örnekleme dağılımı normaldir. Buna göre kullanılması gereken test istatistiği z testi uygulanmalıdır:

$$z = \frac{p - \Pi_0}{\sigma_p}$$

olur. Örnek için standart hata,

$$\sigma_p = \sqrt{\frac{\Pi_0(1-\Pi_0)}{n}} = \sqrt{\frac{(0.80)(0.20)}{120}} = 0.036$$

şeklinde hesaplanır ve

$$z = \frac{0.85 - 0.80}{0.036} = 1.38$$

olarak elde edilir.

Adım 4: İstatistiksel Kararın Verilmesi

Hesaplanan test istatistiği $z_{hes} = 1.38 <$ $z_{tab}(0.05–0.05)$ $z_{tab}(0.4500) = 1.645$ olduğundan H_0 hipotezi kabul edilir. Dolayısıyla H_1 hipotezi reddedilir.

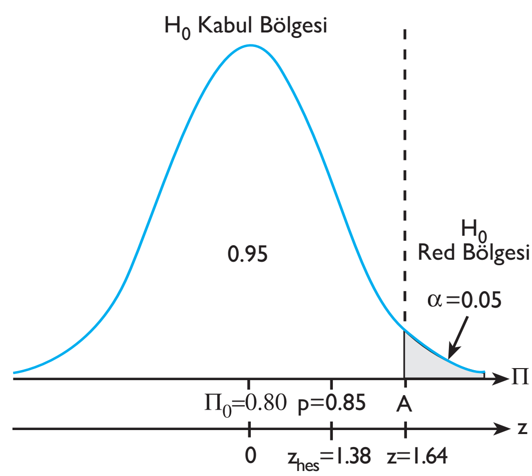
Şekil 3.10'da A değeri;

$A = \pi_0 + z \,.\, s_p = 0{,}8 + 1{,}64 \,.\, 0{,}036 = 0{,}859$

bulunur.

Şekil 3.10

Evren Oranı Π İçin Test Sonuçları.



Adım 5: Probleme İlişkin Kararın Verilmesi

Bu istatistiksel kararın anlamı, söz konusu özel dershane yetkilisinin iddiası yanlıştır, üniversiteye giriş sınavına hazırlık için dershanesine gelen öğrencilerden, üniversitede istediği bölümü kazananların oranının %80'den fazla olmadığı kararına varılmıştır.

ÖRNEK 5

Bir yemek firması, yapmış oldukları yemeklerden memnun olmayan müşteri oranının %10'dan az olduğunu ileri sürmektedir. Konuyu araştırmak amacıyla, bu yemek firmasını tercih eden müşteriler arasından rassal olarak 150 kişi seçilmiş ve 12 kişinin yemeklerden memnun olmadığı öğrenilmiştir. İddianın geçerliliğini 0.01 anlam düzeyinde araştırınız.

Çözüm 5:

Adım 1: Hipotezlerin İfade Edilmesi

Bu örnekte sıfır hipotezi, söz konusu yemek firmasını tercih eden müşteriler arasından yemeklerden memnun olmayanların oranının 0.10'dan farklı olmadığı şeklindedir. Karşıt (araştırma) hipotezi ise yemek firmasını tercih eden müşteriler arasından yemeklerden memnun olmayanların oranının 0.10'dan az olduğu yönündeki tek taraflı hipotezdir. Bu hipotezler;

$H_0 : \Pi = 0.10$

$H_1 : \Pi < 0.10$

şeklinde ifade edilir.

Adım 2: Anlamlılık Düzeyinin Belirlenmesi

Problemde doğru olan H_0 hipotezinin reddedilmesi olasılığı $\alpha = 0.01$ olarak belirlenmiştir. Red bölgesinin oransal büyüklüğü, 0.01'dir ve red bölgesi oranlar örnekleme dağılımının alt kuyruğunda tanımlanmıştır. Bu durum Şekil 3.11'de gösterilmiştir.

Adım 3: Örneklemin Seçilmesi, Verilerin Derlenmesi ve Test İstatistiğinin Hesaplanması

Rassal olarak seçilen 150 müşteri arasından yemeklerden memnun olmayan müşteri sayısı 12 dir. Bu durumda örneklem oranı;

$$p = \frac{r}{n} = \frac{12}{150} = 0.08$$

dir. Örnek için standart hata,

$$\sigma_p = \sqrt{\frac{\Pi_0(1-\Pi_0)}{n}} = \sqrt{\frac{(0.10)(0.90)}{150}} = 0.024$$

gibi hesaplanır.

Örneklem hacmi n = 150 ($n \geq 30$) olduğu için p'nin örnekleme dağılımı normaldir. Buna göre kullanılması gereken test istatistiği,

$$z = \frac{p-\Pi_0}{\sigma_p}$$

olur ve evren oranına ilişkin tek yönlü alt kuyruk z testi uygulanır:

$$z = \frac{0.08 - 0.10}{0.024} = -0.83$$

olarak hesaplanır.

Adım 4: İstatistiksel Kararın Verilmesi

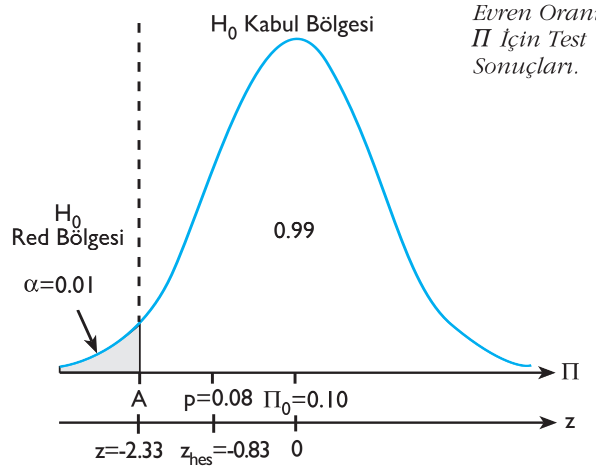
Hesaplanan test istatistiği $z_{hes} = |0.83| < z_{tab} = 2.33$ olduğundan H_0 hipotezi kabul edilir (dolayısıyla H_1 hipotezi reddedilir).

Adım 5: Probleme İlişkin Kararın Verilmesi

Bu kararın anlamı; yemek firmasının yapmış olduğu yemeklerden memnun olmayan müşteri oranının %10'dan az olmadığıdır.

Şekil 3.11

Evren Oranı Π İçin Test Sonuçları.



SIRA SİZDE 3

- **Örneklem hacmi n ≥ 30 birim olduğunda, tek evren ortalamasına ilişkin bir testte, hangi test istatistiği kullanılır? Nedenini açıklayınız.**
- **Evren oranına ilişkin bir testte, test istatistiği nasıl hesaplanır?**
- **Şekil 3.11'deki A'nın değeri nedir?**

İKİ EVREN PARAMETRESİYLE İLGİLİ HİPOTEZ TESTLERİ

İki Evren Ortalaması Arasındaki Farka İlişkin Hipotez Testi

İki evrenin dağılımı normal olduğunda veya dağılımları hakkında bilgi sahibi olmadığımız evrenlerden rassal olarak seçilen örneklemlerin hacimleri yeterli büyüklükte olduğu zaman iki evren ortalaması arasındaki fark parametresi $(\mu_1 - \mu_2)$'ye ilişkin testler tek evren ortalamasının testine benzer şekilde yapılır. Ancak bu hipotez testinde hipotezler aşağıdaki gibi ifade edilir:

H_0: $\mu_1 - \mu_2 = 0$	H_0: $\mu_1 - \mu_2 = 0$	H_0: $\mu_1 - \mu_2 = 0$
H_1: $\mu_1 - \mu_2 \neq 0$	H_1: $\mu_1 - \mu_2 > 0$	H_1: $\mu_1 - \mu_2 < 0$

$\mu_1 - \mu_2 = 0$ ifadesi ile $\mu_1 = \mu_2$ aynı anlama gelen ifadelerdir. Bu nedenle yukarıdaki hipotezler bir başka şekilde;

H_0: $\mu_1 = \mu_2$	H_0: $\mu_1 = \mu_2$	H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 \neq \mu_2$	H_1: $\mu_1 > \mu_2$	H_1: $\mu_1 < \mu_2$

gibi de ifade edilebilirler.

Tanımlanan evrenlerin dağılımları normal, değişkenlikleri biliniyor ise bu evrenlerden rassal olarak seçilen bağımsız örneklemlerin ortalamaları arasındaki fark $\overline{X}_1 - \overline{X}_2$ istatistiğinin örnekleme dağılımı normal olur ve bu istatistiğin standart değerlerinin dağılımı da normaldir. Bu durumda kullanılacak test istatistiği

$$z = \frac{\overline{X}_1 - \overline{X}_2}{\sigma_{\overline{X}_1 - \overline{X}_2}}$$

olur. Burada, $\sigma_{\overline{X}_1-\overline{X}_2}$,

$$\sigma_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

eşitliği ile hesaplanır.

Birbirinden bağımsız rassal örneklemlerin hacimleri $n_1 \geq 30$ ve $n_2 \geq 30$ olması durumunda bu örneklemlerin seçildiği evrenlerin dağılım şekline bakılmaksızın yapılacak testlerde kullanılacak test istatistiği,

$$z = \frac{\overline{X}_1 - \overline{X}_2}{s_{\overline{X}_1-\overline{X}_2}}$$

olur. Burada $s_{\overline{X}_1-\overline{X}_2}$,

$$s_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

eşitliği ile hesaplanır.

$\mu_1 - \mu_2 = 0$ olduğunda z test istatistiğinin dağılımı standart normal dağılım gösterir.

Bu parametreyle ilgili test sürecinin aşamaları diğer testlerdeki gibidir. Bu aşamalar aşağıdaki örnekler üzerinde gösterilmiştir:

ÖRNEK 6

Bir lojistik firması araçları için A veya B firmaları tarafından üretilen lastiklerden satın almayı düşünmektedir. Firma, A firmasının daha fazla kilometre yol kat ettiği bilgisine sahiptir. Lastik satın alma kararını vermeden A firmasının ürettiği lastiklerle ilgili sahip olduğu bilginin doğruluğunu 0.05 anlam düzeyinde araştırmak amacıyla A firmasının ürettiği lastiklerden 100 adet, B firmasının ürettiği lastiklerden 150 adet lastiği rassal olaral seçmiş ve aynı yol koşullarında denemiştir. Deneme yapılan lastiklerin kat ettiği mesafeyle ilgili derlenen veriler kullanılarak aşağıdaki bilgiler elde edilmiştir:

A FİRMASININ LASTİKLERİ	B FİRMASININ LASTİKLERİ
$\overline{X}_1$ = 45000 *KM*	$\overline{X}_1$ = 40000 *KM*
s_1 = 2000 KM	s_2 = 2500 KM
n_1 = 100 adet	n_2 = 150 adet

Çözüm 6:

Adım 1: Hipotezlerin İfade Edilmesi

H_0 hipotezinde A ve B firmaların tarafından üretilen lastiklerin kat ettiği ortalama mesafenin (km olarak) farklılık göstermediği; H_1 hipotezinde ise A firmasının ürettiği lastiklerin B firması tarafından üretilen lastiklerin kat ettiği ortalama mesafeden daha büyük olduğu hipotezi ifade edilir. Buna göre;

H_0: $\mu_1 = \mu_2$ veya H_0: $\mu_1 - \mu_2 = 0$

H_1: $\mu_1 > \mu_2$ veya H_1: $\mu_1 - \mu_2 > 0$

şeklinde ifade edilmiş olur. H_1 hipotezinin ifade ediliş şeklinden anlaşılabileceği gibi yapılacak test tek yönlü üst kuyruk testidir. Yani H_0 hipotezinin red bölgesi $\bar{X}_1 - \bar{X}_2$ örnekleme dağılımının veya bu dağılıma ilişkin standart dağılımın üst kuyruğunda tanımlanmıştır.

Adım 2: Anlamlılık Düzeyinin Belirlenmesi

Araştırmacı, $\mu_1 = \mu_2$ olduğunda α=0.05 olarak belirlemiştir. Bir başka anlatımla araştırmacı H_0 hipotezi gerçekte doğru olduğunda bu hipotezin yanlışlıkla reddedilmesi riskini 0.05 olarak belirlemiştir demektir. α=0.05 örnekleme dağılımının red bölgesinin oransal büyüklüğünü göstermektedir.

Adım 3: Örneklemin Seçilmesi, Verilerin Derlenmesi ve Test İstatistiğinin Hesaplanması

Birbirinden bağımsız olarak seçilen n_1=100 ve n_2=150 birimlik rassal örneklemlerden elde edilen veriler ve hesaplanan gerekli istatistikler yukarıda verilmiştir.

Dağılım şekilleri hakkında bilgi sahibi olmadığımız iki evrenden birbirinden bağımsız olarak seçilen rassal örneklemler büyük hacimli olduğu için $\bar{X}_1 - \bar{X}_2$ örneklem istatistiğinin dağılım şekli normal olur. Bu durumda kullanılacak test istatistiği,

$$z = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{45000 - 40000}{285.17} = 17.5$$

olur ve aşağıdaki gibi hesaplanır:

$$z = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}}$$

Burada $s_{\bar{X}_1 - \bar{X}_2}$,

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{2000^2}{100} + \frac{2500^2}{150}} = 285.77$$

şeklinde hesaplanmıştır.

Adım 4: İstatistiksel Kararın Verilmesi

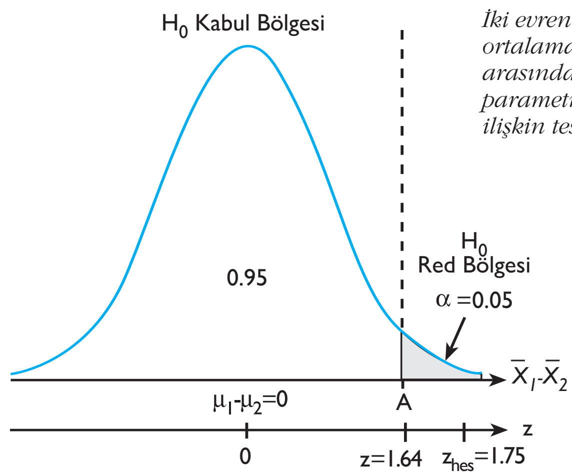
İstatistiksel kararın verilebilmesi için hesaplanan test istatistiğinin (z_{hes}) değeriyle α=0.05 anlamlılık düzeyindeki standart normal eğri alanları tablosundan belirlenen kritik değer (z_{tab}) karşılaştırılır. α=0.05 anlamlılık düzeyinde tek tönlü test için $z_{tab}(0.5–0.05) = z_{tab}(0.4500) = 1.64$ değeri ekte verilen standart normal eğri alanları tablosundan belirlenir.

$$z_{hes} = 17.5 > z_{tab} = 1.64$$

olduğu için H_0 hipotezi reddedilir, dolayısıyla H_1 hipotezi kabul edilir.

Şekil 3.12

İki evren ortalaması arasındaki fark parametresine ilişkin test sonucu.



Adım 5: Probleme İlişkin Kararın Verilmesi

A firmasının ürettiği lastiklerin kat ettiği mesafe B firmasının ürettiği lastiklerden anlamlı şekilde farklılık göstermektedir. A firmasının lastikleri aynı koşullarda daha fazla mesafe kat etmektedir. Başka bir ifadeyle, A firmasının lastikleri daha uzun ömürlüdür sonucuna varılır.

İki Evren Ortalaması Arasındaki Farka İlişkin Küçük Örneklem Testi

Dağılım şekillerinin normal veya normale yakın ve standart sapmaları birbirine eşit () olduğu bilinen iki evrenden birbirinden bağımsız $n_1 < 30$ ve $n_2 < 30$ birimlik rassal örneklemler seçilmiş ise $H_0 = \mu_1 - \mu_2 = 0$ hipotezinin test edilmesi sürecinde aynı adımlar izlenir ancak kullanılması gereken test istatistiği $n_1 + n_2 - 2$ serbestlik derecesinde,

$$t = \frac{\overline{X}_1 - \overline{X}_2}{s_{\overline{X}_1 - \overline{X}_2}}$$

olur. Burada $s_{\overline{X}_1 - \overline{X}_2}$,

$$s_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} . \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

eşitliğiyle tahminlenir.

ÖRNEK 7

İki ayrı ilde aynı iş kolunda faaliyette bulunan iş yerlerinde çalışan işgörenlerin aylık ortalama ücretleri arasındaki farklılık olup olmadığı araştırılmak istenmektedir. Bu amaçla A ilindeki iş yerlerinde çalışanlar arasından rassal olarak 15, B ilindeki iş yerlerinde çalışanlar arasından 15 işgören rassal olarak seçilmiştir. İşgörenlerden derlenen verilerden hesaplanan bilgiler aşağıda çözümleme sürecinin Adım 3 başlığı altında verilmiştir. $\alpha = 0.01$ *için istenen araştırmaya ilişkin kararı veriniz.*

Çözüm 7:

Adım 1: Hipotezlerin İfade Edilmesi

H_0: $\mu_1 - \mu_2 = 0$

H_1: $\mu_1 - \mu_2 \neq 0$

şeklinde ifade edilir. Bu hipotezlerde sırasıyla A ve B illerinde aynı iş kolunda faaliyette bulunan iş yerlerinde çalışan personelin aylık ortalama ücretleri arasında fark olmadığının ifade edildiği H_0 hipotezi farklılık olduğu durumunun ifade edildiği H_1 hipoteziyle test edilmiştir. H_1 hipotezinin ifade ediliş şekline göre, yapılacak test çift yönlü test adını alır. Bu testte red bölgeleri Şekil 3.13'te gösterildiği gibi $\overline{X}_1 - \overline{X}_2$ örneklem istatistiğinin dağılımının $n_1 + n_2 - 2$ veya serbestlik derecesinde t dağılımının her iki kuyruğunda tanımlanmış olur.

Adım 2: Anlamlılık Düzeyinin Belirlenmesi

Örnek problemde araştırmacı H_0: $\mu_1 - \mu_2 = 0$ hipotezini $\alpha = 0.01$ anlamlılık düzeyinde kontrol etmeyi belirlemiştir. Buna göre t dağılımının her iki kuyruğunda tanımlanan red bölgelerinin her birinin oransal büyüklükleri $\alpha/2 = 0.01/2 = 0.005$ olur.

Adım 3: Örneklemin Seçilmesi, Verilerin Derlenmesi ve Test İstatistiğinin Hesaplanması

Her iki evrenden birbirinden bağımsız seçilen rassal küçük örneklemler için hesaplanan istatistikler aşağıda verilmiştir:

n_1 = 15 kişi	n_2 = 15 kişi
$\overline{X}_1$ = ₺2000	$\overline{X}_1$ = ₺1800
s_1 = ₺100	s_2 = ₺120

Bu istatistikler kullanılarak yapılacak test için belirlenecek test istatistiği t test istatistiğidir ve aşağıdaki gibi hesaplanır:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{s_{\overline{X}_1 - \overline{X}_2}} = \frac{2000 - 1800}{41.84} = 4.781$$

Burada $s_{\overline{X}_1 - \overline{X}_2}$,

$$s_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \sqrt{\frac{15.100^2 + 15.150^2}{15 + 15 - 2}} \cdot \sqrt{\frac{1}{15} + \frac{1}{15}} = 41.84$$

şeklinde hesaplanmıştır.

Adım 4: İstatistiksel Kararın Verilmesi

Bu kararın verilebilmesi için hesaplanan test istatistiğinin değeri (t_{hes}) ile belirlenen α = 0.01 anlamlılık düzeyinde t dağılımına ilişkin ekte verilen t değerler tablosundan belirlenen t tablo değeri (t_{tab}) ile karşılaştırılır. Bu örnekte uygulanacak test çift yönlü test olduğu için $t_{tab} = t_{\alpha/2}$; s.d = $n_1 + n_2 - 2$ = t(0.005); s.d = 28 = 2.763 belirlenir. t_{tab} = 2.763 değeri Ek 2'de verilen t tablosunda serbestlik derecesi sütununda 28 serbestlik derecesine karşılık gelen satır ile olasılık düzeyi sütunundaki $\frac{\alpha}{2} = \frac{0.01}{2} = 0.005$ sütununun kesiştiği yerdeki değerdir. Bu bilgilere göre,

$t_{hes} = 4.781 > t_{tab} = 2.763$

olduğu için istatistiksel karar H_0 hipotezinin reddedilmesi; H_1 hipotezinin kabul edilmesi yönünde verilir.

Adım 5: Probleme İlişkin Kararın Verilmesi

Yukarıdaki istatistiksel karara göre A ilindeki iş yerlerinde çalışanların aylık ortalama ücretleri B ilindeki çalışanlardan farklıdır. Farklılık A ilindeki çalışanların aylık ortalama ücretlerinin daha yüksek olduğu yönündedir.

İki Evren Oranı Arasındaki Farka İlişkin Hipotez Testi

Bu bölümde iki evren oranı arasındaki farka ilişkin test sadece büyük örneklemler için uygulanmıştır. Bu test süreci iki evren ortalaması arasındaki farka ilişkin test sürecine benzemektedir. Ancak bu hipotez testinde hipotezler aşağıdaki hipotez takımlarından herhangi birisiyle ifade edilirler:

$H_0: \Pi_1 = \Pi_2$	$H_0: \Pi_1 = \Pi_2$	$H_0: \Pi_1 = \Pi_2$
$H_1: \Pi_1 \neq \Pi_2$	$H_1: \Pi_1 > \Pi_2$	$H_1: \Pi_1 < \Pi_2$

veya $\Pi_1 = \Pi_2$ ifadesi ile $\Pi_1 - \Pi_2 = 0$ ifadesi aynı anlama geldiği için bu hipotezler,

$H_0: \Pi_1 - \Pi_2 = 0$	$H_0: \Pi_1 - \Pi_2 = 0$	$H_0: \Pi_1 - \Pi_2 = 0$
$H_1: \Pi_1 - \Pi_2 \neq 0$	$H_1: \Pi_1 - \Pi_2 > 0$	$H_1: \Pi_1 - \Pi_2 < 0$

şeklinde de ifade edilebilir.

Bu test türünde her iki evrenden de rassal olarak seçilecek örneklemlerin hacimleri yeterli büyüklükte olduğu zaman kullanılacak standart test istatistiği $\Pi_1 - \Pi_2 = 0$ olduğu durumda yaklaşık normal dağılıma sahip olan z test istatistiğidir. Bu istatistik,

$$z = \frac{p_1 - p_2}{s_{p_1 - p_2}}$$

gibi ifade edilir. Burada $s_{p_1 - p_2}$ örneklem istatistiğinin standart hatasıdır ve

$$s_{p_1 - p_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

eşitliği ile hesaplanır.

ÖRNEK 8

YGS sonucuna göre fen lisesi mezunu öğrenciler içerisinde lisans programlarını tercih etme başarısına sahip olan öğrencilerin oranının anadolu lisesi mezunlarının oranından yüksek olduğu iddia edilmektedir. Söz konusu olan iddianın araştırılması için araştırmacı her iki ortaöğretim kurumlarından mezun olan ve YGS'ye giren öğrenciler arasından birbirinden bağımsız ve rassal olarak sırasıyla $n_1 = 80$ ve $n_2 = 70$ öğrenci seçmiştir. Seçilen 80 fen lisesi mezununun 64'ü ve 70 anadolu lisesi mezununun ise 52'sinin lisans programlarını tercih etme başarısına sahip olduğunu belirlemiştir. İddianın araştırılması için aşağıdaki çözümleme sürecinin aşamaları izlenmiştir.

Çözüm 8:

Adım 1: Hipotezlerin İfade Edilmesi

Söz konusu iddianın araştırılması için test edilecek hipotezler aşağıdaki gibi ifade edilir. Uygulanacak test tek yönlü üst kuyruk testi olmalıdır; testin red bölgesi $p_1 - p_2$ örneklem istatistiğinin dağılımının üst kuyruğunda tanımlanmıştır. Çünkü fen lisesi mezunlarıyla ilgili Π_1 oranının anadolu lisesi mezunlarıyla ilgili Π_2 oranından büyük olduğu yönünde bir araştırma hipotezi söz konusudur.

$H_0: \Pi_1 = \Pi_2$

$H_1: \Pi_1 > \Pi_2$

Adım 2: Anlamlılık Düzeyinin Belirlenmesi

Araştırmacı, $H_0: \Pi_1 = \Pi_2$ hipotezini kontrol edebilmek için α anlamlılık düzeyini 0.05 olarak belirlemiştir. Buna göre tek yönlü üst kuyruk testinde tanımlanan red bölgesinin oransal büyüklüğü %5 olarak belirlenmiş olur. Başka bir anlatımla yukarıdaki H_0 hipotezi doğru olduğunda bu hipotezin reddedilmesi olasılığı %5 olarak belirlenmiş demektir.

Adım 3: Örneklemin Seçilmesi, Verilerin Derlenmesi ve Test İstatistiğinin Hesaplanması

Örnekte verilere göre test için gerekli olan istatistikler aşağıda hesaplanmıştır:

$n_1 = 80$ $\qquad$ $n_2 = 70$

$r_1 = 64$ $\qquad$ $r_2 = 52$

$$p_1 = \frac{r_1}{n_1} = \frac{64}{80} = 0.80 \qquad p_2 = \frac{r_2}{n_2} = \frac{52}{70} = 0.74$$

$n_1 \geq 30$ ve $n_2 \geq 30$ birim olduğu için $p_1 - p_2$ örneklem istatistiğinin dağılım şekli normal olur ve bu istatistiğin $\frac{p_1 - p_1}{s_{p_1-p_2}}$ standart değerlerinin dağılımı da standart normal dağılım özelliğini gösterir. Bu nedenle söz konusu testin uygulanması için kullanılacak test istatistiği

$$z = \frac{p_1 - p_1}{s_{p_1-p_2}}$$

olur ve

$$z = \frac{0.80 - 0.74}{0.069} = 0.87$$

şeklinde hesaplanır. Burada $s_{p_1-p_2}$,

$$s_{p_1-p_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} = \sqrt{\frac{0.8 \cdot 0.2}{80} + \frac{0.74 \cdot 0.26}{70}} = 0.069$$

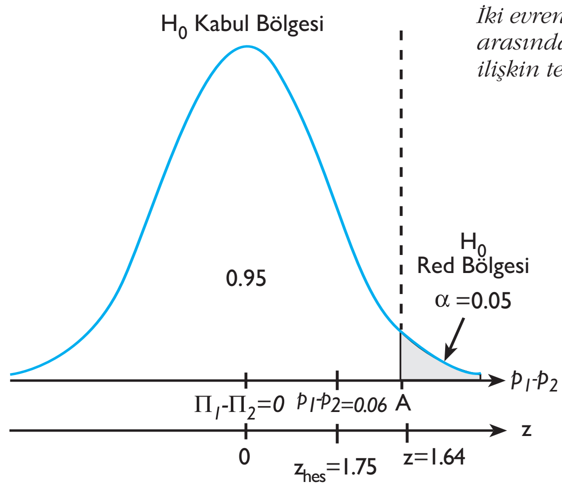
şeklinde hesaplanmıştır.

Adım 4: İstatistiksel Kararın Verilmesi

Anlamlılık düzeyi $\alpha = 0.05$ olarak belirlendiği için z tablo değeri $z_{tab}(0.5–0.05) = z_{tab}(0.4500) = 1.64$ olarak belirlenmiş olur. Bu değer hesaplanan test istatistiği ile karşılaştırılırsa

$$z_{hes} = 0.87 < z_{tab}(0.4500) = 1.64$$

olduğundan H_0 hipotezi kabul, H_1 hipotezi reddedilir. Bu durum Şekil 3.14'te gösterilmiştir.



Şekil 3.13

İki evren oranı arasındaki farka ilişkin test sonucu.

Adım 5: Probleme İlişkin Kararın Verilmesi

Fen ve anadolu lisesi mezunlarının YGS sonuçlarına göre lisans programlarını seçme başarısına sahip olanların oranları arasında anlamlı bir farklılık görülmediği, $p_1 - p_2 = 0.80 – 0.74 = 0.06$ farkının rassal bir fark olduğu söylenebilir.

SIRA SİZDE 4

- **İki evren ortalaması arasındaki farka ilişkin çift yönlü bir test uygulamasında hipotezler nasıl ifade edilir, test edilecek hipotez nedir?**
- **Örneklem hacmi yeterli büyüklükte olmadığında, seçilen rassal örneklemin ait olduğu evrenin dağılımının normal olduğu bilindiğinde iki evren arasındaki farka ilişkin bir testte kullanılacak test istatistiği nedir? Belirtiniz.**
- **İki evren ortalaması arasındaki farka ilişkin bir küçük örneklem testinde örneklem hacimleri 15 ve 10 birim ise serbestlik derecesi kaçtır?**

İKİDEN FAZLA EVREN ORTALAMASININ KARŞILAŞTIRILMASI, TEK YÖNLÜ VARYANS ÇÖZÜMLEMESİ - F TESTİ

Bu ünitenin önceki kısımlarında iki evren ortalaması arasındaki hipotez sınamasının z ve t test uygulamaları ile nasıl yapılacağı örnek 3.12 ve 3.13'te ayrıntılı olarak incelenmiştir. Örnek 3.12'de A ve B firmalarının ürettikleri lastiklerin ortalama ömürleri karşılaştırılırken örnek 3.13'te iki ilde aynı iş kolunda faliyette bulunan iş yerlerinde çalışan işgörenlerin aylık ortalama ücretleri karşılaştırılmıştır. Örneklerden de anlaşılabileceği gibi gruplar bir bağımsız değişkenin ölçme düzeylerine karşılık gelmektedir. Örnek 3.12'de bağımsız değişken firma türü, ölçme düzeyleri (yani gruplar) ise A ve B firmalarıdır. Örnek 3.13'te ise bağımsız değişken il türüdür, grup sayısı ise iki ille sınırlıdır. Örnek 3.12'de araştırmaya konu olan bağımlı değişken üretilen lastiklerin ömrü iken Örnek 3.13'te işgören aylık ücretleridir.

Örnek 3.12'ye dönecek olursak ortalama ömürleri karşılaştırılacak lastik firması sayısı ikiden fazla olduğunda bir başka anlatımla ikiden fazla grubun (evrenin) ortalamasının karşılaştırılması amaç olduğunda tek yönlü varyans çözümlemesi (ANOVA) uygulanır. Çünkü karşılaştırılması düşünülen ikiden fazla evrenin ortalamasının değişik kombinasyonlarda ikişerli karşılaştırılmasını z ve t testleri ile yapmaya çalışmak uygun olmaz. Bunun nedenlerini aşağıdaki gibi saymak mümkündür:

- Ortalamaları karşılaştırılacak evren (grup) sayısı, örneğin r=4 olduğunda 6 tane z veya t testi uygulaması gerektirir.
- 6 tane z veya t testi uygulanmış bile olsa elde edilen bilgilerden grupların tamamı için genelleme yapılamaz.
- z ve t test uygulaması ile ikişerli grup (evren) ortalamalarının karşılaştırılması yanlış karar verme riskini (olasılığını) artırır. Örneğin anlamlılık düzeyi α=0,05 (güven düzeyi 1–α=0,95) olması durumunda uygulanancak 6 tane testin tamamı için doğru karar verebilme olasılığı $0,95^6$ = 0,735 olur. Yani hiç olmazsa bir yanlış karar verme olasılığı 1–0.735 = 0,265 olur. Bu yanlış karar verme olasılığı grup sayısı arttıkça da artacaktır. Örneğin aynı güven düzeyinde grup sayısı 6 olduğunda yapılacak ikişerli grup karşılaştırma sayısı 15 olur ve 15 tane test için doğru karar verme olasılığı 0.95^{15} = 0,4633 olur. Hiç olmazsa bir yanlış karar verme olasılığı ise 1–0,4633 = 0,5367 bulunur. Görüldüğü gibi yanlış karar verme olasılığı 0,265'ten 0,5367'ye yükselmiştir.

Yukarıda açıklanan nedenlerden dolayı bir bağımsız değişkenin ikiden fazla grubuna (ölçme düzeylerine) ait bir bağımlı değişkenin ortalamalarının dağılımı farklı mıdır? sorusunun yanıtını z veya t testi uygulamaları ile değil; tek yönlü varyans çözümlemesi - F testi ile aramak gerekir.

Tek Yönlü Varyans Çözümlemesi Modeli

Modelin Tanıtılması

Tek yönlü varyans çözümlemesi modelleri, gözlem değerlerinin kategorik bir bağımsız değişkene göre gruplandırıldığı ve grupların evren ortalamalarının birbirine eşitliğinin test edildiği modellerdir.

Tek yönlü varyans çözümlemesi modeli

$$X_{ij} = \mu_j + \varepsilon_{ij}$$

şeklinde yazılır.

Burada;

X : İlgilenen bağımlı değişkeni

j : Grup sayısını (Bağımsız değişkenin ölçme düzeyi sayısını) , j=1,2,........, r

i : Birimlerin numarasını i=1,2,......., n_j

X_{ij} = j'inci gruptaki i'inci birime ait gözlem değerini

μ_j = X değişkenin j'inci grup için ortalamasını

ε_{ij} = j'inci gruptaki i'inci birime ait hata terimini gösterir.

Modeldeki bu simgeleri bir örnek üzerinde açıklayalım:

ÖRNEK 9

Üç farklı eğitim yönteminin uygulandığı okullardaki öğrencilerin dönem sonu başarı ortalamalarının farklı olup olmadığı araştırmak isteniyor, bu amaçla bu eğitim yöntemlerinin uygulandığı okulların her birinden, birbirinden bağımsız rassal olarak 5'er öğrenci seçiliyor. Seçilen öğrencilerin 10 puan üzerinden değerlendirilen dönem sonu başarı puanları ile ilgili derlenen veriler aşağıdaki tabloda verilmiştir.

Tablo 3.1 *Eğitim yöntemi türü itibarıyla örneklem başarı puanları dağılımı.*

	Eğitim Yöntemi Türü		
	A	**B**	**C**
	4	6	6
	6	6	5
	5	8	2
	5	10	5
	10	10	7
Toplam	**30**	**40**	**25**
	$n_1 = 5$	$n_2 = 5$	$n_3 = 5$

Bu örnekte bağımsız değişken eğitim yöntemi türüdür. A, B, C eğitim yöntemleri bu değişkenin gruplarıdır, ölçme düzeyleridir. Grup sayısı r = 3'tür. A, B ve C eğitim yöntemleriyle eğitimin yapıldığı incelemeye alınan okullardaki öğrenci toplulukları evrenleri oluşturur. Bu okullardaki öğrencilerin tamamının oluşturduğu topluluk büyük(bütünleşik) evren, genel evren olarak isimlendirilebilir. Bu kısımda büyük evren yerine genel evren kavramı kullanılmıştır. Büyük evrenin ortalaması μ, hacmi ise N_T simgesiyle gösterilebilir.

Varyans çözümlemeleri ile ilgili açıklamaların yapıldığı kimi alanyazında bağımsız değişkene faktör, bağımsız değişkenin ölçme düzeylerine faktör düzeyleri denmektedir. Bu ünitede faktör düzeyleri yerine grup ifadesi benimsenmiştir. 3.15'nolu örnekte öğrencilerin başarı puanları bağımlı değişkendir.

Tek yönlü varyans çözümlemesinde bağımlı değişkenin nicel, bağımsız değişkeninde kategorik değişken olması şarttır. Ayrıca bağımsız değişkenin yukarıda belirtildiği gibi ikiden fazla ölçme düzeyine sahip olması gerekir.

Yukarıdaki örnekte her gruptan (evrenden) gözlenen örneklem birim sayısı birbirine eşit olup n_{ij} = 5'tir. Gruplardaki örneklem birim sayılarının eşit olması şart değildir. Ancak eşit olması veya birbirine yakın olması durumunda daha doğru bir bilgi üretebileceği açıktır.

Modelin Varsayımları

- X rassal değişkeni bakımından grupların x_{ij} değerlerinin olasılık dağılımı normaldir.
- x_{ij} değerleri μ_j ve ε_{ij} değerlerinin toplamıdır. Çünkü x_{ij} bir rassal değişkendir.
- Hata terimleri ε_{ij} birbirinden bağımsızdır, dağılımı normaldir ve ortalaması sıfırdır.
- Grupların varyansları birbirine eşittir, yani ortak varyansa sahiptir ve σ^2 simgesi ile gösterilir.

Özetle, x_{ij}'ler birbirinden bağımsız, ortalamaları μ_j ve varyansları σ^2 olan normal dağılıma sahiptir.

Tek Yönlü Varyans Çözümlemesi - F Testi Sürecinin Sürecinin Aşamaları

Bu çözümleme sürecinde de daha önceki test sürçlerindeki adımlar izlenir.

Adım 1: Hipotezlerin İfade Edilmesi

H_0 hipotezinde bağımlı değişken itibarıyla grupların ortalamaları arasında fark olmadığı, grup ortalamalarının birbirine eşit olduğu ifade edilirken H_1 hipotezinde grupların en az ikisinin ortalamalarının farklılık gösterdiği ifade edilir. Buna göre Örnek 3.15 için hipotezler;

H_0: $\mu_1 = \mu_2 = \mu_3$

H_1: $\mu_1 \neq \mu_2 \neq \mu_3$

gibi ifade edilirler.

Burada;

μ_1 = A Eğitim yöntemi ile eğitim alan bütün öğrencilerin ortalama başarı puanını

μ_2 = B Eğitim yöntemi ile eğitim alan bütün öğrencilerin ortalama başarı puanını

μ_3 = C Eğitim yöntemi ile eğitim alan bütün öğrencilerin ortalama başarı puanını

göstermektedir.

Adım 2: Anlamlılık Düzeyinin Belirlenmesi

Araştırmacı diğer hipotez sınamalarında olduğu gibi hipotezler ifade edildikten sonra veri derleme aşamasına geçmeden önce α anlamlılık düzeyini belirlenmesi etik bir gerekliliktir. Örneğimiz için anlam düzeyi α = 0,05 olarak belirlenmiştir.

Adım 3: Örneklemin Seçilmesi, Verilerin Derlenmesi ve Test İstatistiğinin Hesaplanması

Test sürecinin bu aşamasını Örnek 3.15 üzerinden açıklayalım. A, B, ve C eğitim yöntemleri ile öğrenim gören öğrenci gruplarının (alt evrenlerin) her birinden bağımsız ve rassal olarak $n_1 = n_2 = n_3 = 5$ öğrenci seçilmiş ve bu öğrencilerin başarı puanları ile ilgili Tablo 3.1'deki veriler derlenmiştir. Derlenen bu veriler kullanılarak aşağıdaki istatistikler hesaplanmıştır.

Örneklem Ortalamalarının Hesaplanması

- Grup içi örneklem ortalamalarının hesaplanması

Her grup için örneklem ortalaması $\overline{X}_J$ simgesi ile gösterilir ve

$$\overline{x}_J = \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j}$$

eşitliği ile hesaplanır.

Burada n_j j'inci grubun örneklem hacmini gösterir.

$\overline{X}_J$, μ_j'nin yansız tahminleyicisidir.

Örnek 3.15 için $\overline{X}_J$'ler Tablo 3.1'deki verilerden yararlanarak aşağıdaki gibi hesaplanır:

$$\overline{x}_1 = \frac{30}{5} = 6\,\text{puan}$$

$$\overline{x}_2 = \frac{40}{5} = 8\,\text{puan}$$

$$\overline{x}_3 = \frac{25}{5} = 5\,\text{puan}$$

bulunur.

- Gruplar arası genel örneklem ortalamasının hesaplanması

Gruplardaki toplam örneklem birim sayısı $n_T = \sum n_j$ simgesi ile gösterilirse genel ortalama

$$\overline{\overline{x}} = \frac{\sum_{j=1}^{r} \sum_{i=1}^{n_j} x_{ij}}{n_T}$$

eşitliği ile hesaplanır.

Örnek 3.15 için bu ortalama hesaplanırsa

$$\overline{\overline{x}} = \frac{30+40+25}{15} = \frac{95}{15} = 6,33$$ puan bulunur.

Değişkenliğin Hesaplanması

Tek yönlü varyans çözümlemesinde gruplar arası, gruplar içi ve toplam değişkenlik hesaplanır.

- Gruplar arası değişkenliğin (varyansın) hesaplanması

Bu değişkenlik, grupların örneklem ortalamaları ile genel ortalama arasındaki farkın her grubun örneklem hacmi ile ağırlıklandırılması ve gruplar arası serbestlik derecesine bölünmesiyle bulunur. Gruplar arası değişkenlik S_B^2 simgesi ile gösterilir ve aşağıdaki eşitlik yardımıyla hesaplanır.

$$S_B^2 = \frac{\sum_{j=1}^{r} n_j \left(\bar{x}_j - \bar{\bar{x}}\right)^2}{r-1} = \frac{SSB}{r-1}$$

Burada;

$\bar{x}_J$: j'inci grup için örneklem ortalamasını

$\bar{\bar{x}}$: Bütün gruplar için genel örneklem ortalamasını

n_j : j'inci gruptaki gözlem (birim) sayısını

r : Grup sayısını

$r - 1$: Gruplar arası serbestlik derecesini

$(\bar{x}_j - \bar{\bar{x}})$: grup örneklem ortalamalarının genel ortalamadan sapmalarını gösterir.

$\sum_{j=1}^{r} n_j(\bar{x}_j - \bar{\bar{x}})^2$: Gruplar arası kareler toplamını ifade eder ve SSB simgesi ile gösterilir.

Örnek 3.15 için S_B^2'nin değeri aşağıdaki gibi hesaplanır.

$\sum_{j=1}^{r} n_j(\bar{x}_j - \bar{\bar{x}})^2 = 5\,(6 - 6{,}33)^2 + 5\,(8 - 6{,}33)^2 + 5\,(5 - 6{,}33)^2 = 23{,}33$

$r - 1 = 3 - 1 = 2$

$S_B^2 = 23{,}33/2 = 11{,}67$ bulunur.

- Gruplar içi değişkenliğin hesaplanması

Bu değişkenlik her gruptaki gözlem değerlerinin kendi grup ortalamasından sapmalarının toplamına ilişikin değişkenliği gösterir ve S_E^2 simgesi ile ifade edilir ve aşağıdaki eşitlik yardımıyla hesaplanır:

$$S_E^2 = \frac{\sum_{j=1}^{r} \sum_{i=1}^{n_j} \left(x_{ij} - \bar{x}_j\right)^2}{n_T - r} = \frac{SSE}{n_T - r}$$

Burada;

$\bar{x}_J$: j'inci grup için örneklem ortalamasını

r : Grup sayısını

n_T : Gruplardaki toplam gözlem değeri sayısı

$n_T - r$: Gruplar içi serbestlik derecesi

$(x_{ij} - \bar{x}_j)$: Gruplardaki gözlem değerlerinin kendi grup ortalamasından sapmasını

$\sum_{j=1}^{r} \sum_{i=1}^{n_j} \left(x_{ij} - \bar{x}_j\right)^2$: Her gruptaki gözlem değerlerinin kendi grup ortalamasından sapmalarının toplamını ifade eder, gruplar içi kareler toplamı olarak isimlendirilir ve SSE simgesiyle gösterilir.

Örnek 3.15 için gruplar içi değişkenlik;

A Eğitim yöntemi için:

$$\sum_{i=1}^{5} (x_{ij} - \bar{x}_1)^2 = (4–6)^2 + (6–6)^2 + (5–6)^2 + (5–6)^2 + (10–6)^2 = 22$$

B Eğitim yöntemi için:

$$\sum_{i=1}^{5} (x_{ij} - \bar{x}_2)^2 = (6–8)^2 + (6–8)^2 + (8–8)^2 + (10–8)^2 + (10–8)^2 = 16$$

C Eğitim yöntemi için:

$$\sum_{i=1}^{5} (x_{ij} - \bar{x}_3)^2 = (6–5)^2 + (5–5)^2 + (2–5)^2 + (5–5)^2 + (7–5)^2 = 14$$

$$SSE = \sum_{j=1}^{r} \sum_{i=1}^{n_j} \left(x_{ij} - \bar{x}_j\right)^2 = 22 + 16 + 14 = 52$$

$$n_T - r = 15 - 3 = 12$$

$$S_E^2 = 52/12 = 4,33$$

olur.

F Test İstatistiğinin Hesaplanması

Daha önce açıklanmış olduğu gibi ikiden fazla grup ortalamasının karşılaştırılması amaç olduğunda değişkenlik gruplar arası ve gruplar içi varyanslarla belirleniyordu ve ortalamalardan sapmaların kareleri toplamı değişmeyi gösteriyordu. Grup içi varyans S_E^2 her grubun gözlem değerlerinin kendi ortalaması etrafında değişkenliğini göstermektedir. Bu nedenle, bu değişkenlik grup farklılıklarından etkilenmez. Bu değişkenlik grup içindeki rassal değişmeyi gösterir. Öte yandan gruplar arası değişkenlik S_B^2 ise sadece gözlem değerlerinden gözlem değerlerine rassal değişmeyi değil; bir gruptan öteki gruba farklılığı ölçer. Eğer gruptan gruba farklılık yoksa S_B^2, S_E^2'ye yaklaşır. Eğer gerçekten farklılık varsa S_B^2, S_E^2'den daha büyük olur. Bu iki varyansın birbirine oranını $\left(\frac{S_B^2}{S_E^2}\right)$ ifade eden istatistiğin dağılımı bilinen dağılımlardan F dağılımının özelliklerine sahip olur.

Varyans çözümlemelerinde F test istatistiği kullanılır ve bu istatistik aşağıdaki eşitlikle ifade edilir:

$$F = \frac{S_B^2}{S_E^2}$$

F dağılımının dağılım şekli, Şekil 3.14'te gösterilmiştir. F dağılımının şekli 2 tane serbestlik derecesine bağlıdır. Bu serbestlik dereceleri;

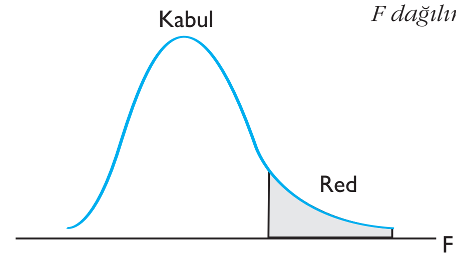
Gruplar arası serbestlik derecesi: $sd_1 = r - 1$
Gruplar içi serbestlik derecesi : $sd_2 = n_T - r$

şeklinde gösterilir.

F istatistiği 0 ve $+\infty$ arasında değerler alır. Sürekli bir bölünmedir. F, pozitif değerler aldığından fonksiyonun eğrisi ordinat ekseninin sağ tarafında kalır ve asimetrik bir görünüme sahiptir. Serbestlik derecesi arttıkça dağılımın şekli normal dağılıma yaklaşır.

Şekil 3.14

F dağılımı



Örnek 3.15 için test istatistiği ve serbestlik dereceleri aşağıdaki gibi hesaplanmıştır:

$$F = \frac{S_B^2}{S_E^2} = \frac{11,67}{4,33} = 2,69$$

$sd_1 = 3 - 1 = 2$
$sd_2 = 15 - 3 = 12$

Yukarıdaki hesaplamalarla ilgili bilgiler aşağıdaki Tablo 3.2 ve 3.3'te gösterilmiştir:

Tablo 3.2 *Varyans çözümlemesi tablosunun formüllerle gösterimi.*

Kaynak	Kareler Toplamı (SS)	Serbestlik Dereceleri	Varyans	F
Gruplar Arası	$SSB = \sum_{j=1}^{r} n_j \left(\bar{x}_j - \bar{\bar{x}}\right)^2$	r - 1	S_B^2	$\frac{S_B^2}{S_E^2}$
Gruplar İçi	$SSE = \sum_{j=1}^{r} \sum_{i=1}^{n_j} \left(x_{ij} - \bar{x}_j\right)^2$	n_T - r	S_E^2	
Toplam	$\sum_{j=1}^{r} \sum_{i=1}^{n_j} \left(x_{ij} - \bar{\bar{x}}\right)^2$	n_T - 1		

Tablo 3.3 *Örnek 3.15 için tek yönlü varyans çözümlemesi sonuçları.*

Kaynak	Kareler Toplamı	Serbestlik Dereceleri	Varyans	F
Gruplar Arası	23,33	2	11,67	2,69
Gruplar İçi	52	12	4,33	
Toplam	75,33	14		

Tablo 3.2 ve Tablo 3.3'ten anlaşılabileceği gibi SSB+SSE'nin toplamı, kareler toplamının toplamını verir. Bu durum;

SST = SSB + SSE

şeklinde yazılır. Toplam değişkenliği gösteren kareler toplamının toplamı (SST), grupların x_{ij} gözlem değerlerinin genel ortalamadan ($\bar{\bar{x}}$) farklılıklarını ölçer.

Adım 4: İstatistiksel Kararın Verilmesi

Eğer gerçekten grup ortalamaları arasında anlamlı farklılık varsa gruplar arası varyans S_B^2, S_E^2'den anlamlı şekilde büyük olur. Dolayısıyla hesaplanan F istatistiğinin değeri (F_{hes}) α anlamlılık düzeyinde sd_1 ve sd_2 serbestlik derecelerinde (F_α ; sd_1 ; sd_2) F tablo değerinden (F_{tab}) büyükse H_0 hipotezi red; H_1 hipotezi kabul edilir. Aksi durumda H_0 hipotezi kabul edilir.

Örnek 3.15 için H_0 hipoteziyle ilgili kararı verebilmek için F_{hes} = 2,69 değeriyle α = 0,05 ; sd_1 = 3–1 = 2 ; sd_2 = 15–3 = 12 için F tablo değerinin karşılaştırılması gerekir. Ek 3'te α = 0,05 ve α = 0,01 için F dağılımı tabloları verilmiştir. α = 0,05 için düzenlenen F tablosunun bir kesiti Tablo 3.4'te yer almaktadır. Bu tabloda $sd_1 = 2$ sütunu ile $sd_2 = 12$ satırının kesiştiği yerdeki değer F tablo değeri olarak alınır.

$$F_{tab} = 3{,}88$$

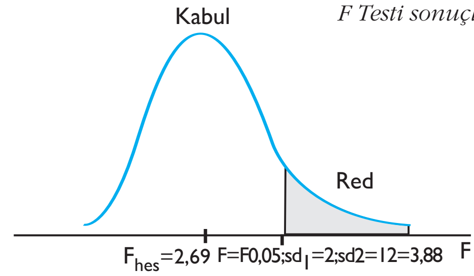
olarak belirlenmiştir. Bu bilgilere göre;

$$F_{hes} = 2{,}69 < F_{0{,}05\,;\,sd_1=2\,;\,sd_2=12} = F_{tab} = 3{,}88$$

olduğundan H_0 hipotezi kabul; H_1 hipotezi ise reddedilir. Bu sonuçlar şekil 3.15'te gösterilmiştir.

Şekil 3.15

F Testi sonuçları.



	sd_1	1	2	3	...	6	...	8	...
sd_2									
1		161	200	216	...	234	...	239	...
2		18,51	19	19,16	...	19,33	...	19,37	...
...		...	...	...	...	...	...	...	...
11		4,84	3,97	3,59	...	3,09	...	2,95	...
12		4,75	3,88	3,49	...	3	...	2,85	...
13		4,67	3,8	3,41	...	2,92	...	2,77	...

Tablo 3.4
α=0.05 için F dağılım tablosunun bir kesiti.

Adım 5: Probleme İlişkin Kararın Verilmesi

Eğitim yöntemi türlerinin ortalama başarı puanları arasında anlamlı bir fark bulunmamaktadır; örneklem ortalamalarının farklılığı rasal farklılığı göstermektedir.

SIRA SİZDE 5

- **Tek yönlü varyans çözümlemesinde, bağımsız değişkenin ölçme düzeyi dört(4) ise gruplar arası serbestlik derecesi kaç olur?**
- **Gruplar içi değişkenlik neyi ölçer ve nasıl hesaplanır?**

Özet

AMAÇ 1

Hipotez, istatistiksel hipotez ayrımını ifade etmek

Genel olarak hipotez, karşılaşılan özel duruma ilişkin bir önermedir. İstatistiksel hipotez, bir araştırmada ilgilenilen bir ya da daha fazla parametrenin değeri hakkında ileri sürülen ve doğruluğu, geçerliliği bu parametre hakkında bilgi üreten istatistikden ve bu istatistiğin örnekleme dağılımıyla ilgili bilgilerden yararlanarak araştırılabilen önermelerdir. İstatistiksel hipotez bir dağılımın parametre değerine ilişkin hipotezdir.

AMAÇ 2

İstatistiksel hipotezlerin test aşamalarını açıklamak

İstatistiksel hipotezlerin test edilmesi sürecinin aşamalarını aşağıdaki gibi saymak mümkündür: 1) Hipotezlerin ifade edilmesi 2) Anlamlılık düzeyinin belirlenmesi 3) Rassal örneklemin seçilmesi, verilerin derlenmesi, test istatistiğinin hesaplanması 4) İstatistiksel kararın verilmesi 5) Probleme ilişkin kararın açıklanması.

İstatistiksel hipotezler sıfır hipotezi, istatistiksel hipotez (H_0) ve karşıt hipotez, araştırma hipotezi (H_1) şelinde ifade edilirler. H_0 hipotezinde test edilecek parametrenin bilinen, belirlenen değerinde herhangi bir değişikliğin olmadığı durum ifade edilir. H_1 hipotezinde ise H_0 hipotezini çürütecek bir hipotezdir. Bu hipotez test edilecek hipotezin yönünü tanımlar. Test sonucu verilecek kararı parametre değerinden hem küçük hem de büyük yöndeki anlamlı örneklem istatistiği farklılıkları etkileyecek ise H_1 hipotezi çift yönlü; verilecek kararı evren parametresinden ya küçük ya da büyük yöndeki anlamlı örneklem istatistiği farklılıkları etkileyecekse H_1 hipotezi tek yönlü olarak tanımlanacaktır. Test sürecinin ikinci aşa- α anlamlılık düzeyini belirler. Bu değer örneklem istatistiğinin dağılımında sıfır hipotezinin red bölgesini tanımlar. Bir başka anlatımla α, sıfır hipotezi gerçekte doğru iken yanlışlıkla bu hipotezin reddedilmesi olasılığını gösterir. Üçüncü aşamada ise test edilecek parametreyle ilgili bilgi üreten örneklem istatistiğinin dağılımının şekliyle ilgili bilgilerden yararlanılarak test istatistiği belirlenir. Bu ünitede z ve t test istatistikleri ile F test istatistiği uygulamalı olarak gösterilmiştir. Dördüncü ve beşinci aşamalarda sırasıyla istatistiksel karar ve bu kararın probleme ilişkin anlamı ortaya konur. Karar sürecinde hesaplanan test istatistiğinin değeri ile belirlenen _ anlamlılık düzeyindeki z, t ve F tablo değerleri karşılaştırılır.

AMAÇ 3

Tek evren parametresiyle ilgili hipotez testi uygulamak

Bu bölümde, uygulamada sıkça karşılaşılan tek evren aritmetik ortalaması ile tek evren oranına ilişkin parametreler hakkında hipotez sınamaları uygulanmıştır. Tek evren ortalamasına ilişkin testte örneklem hacminin küçük ve büyük olması durumları için testler, test sürecinin aşamalarına uygun şekilde ayrı ayrı örneklerle açıklanmıştır. Tek evren oranına ilişkin test uygulaması ise sadece büyük örneklem hacimleri için yapılmıştır.

AMAÇ 4

İki evren parametresiyle ilgili hipotez testi uygulamak

Bu bölümde, uygulamada sıkça karşılaşılan iki evren aritmetik ortalaması arasındaki fark ile iki evren oranı arasındaki fark parametrelerine ilişkin hipotez sınamaları uygulanmıştır. İki evren ortalamasının farkına ilişkin testte örneklem hacminin küçük ve büyük olması durumları için testler, test sürecinin aşamalarına uygun şekilde ayrı ayrı örneklerle açıklanmıştır. İki evren oranının farkına ilişkin test uygulaması ise sadece büyük örneklem hacimleri için yapılmıştır.

AMAÇ 5

İkiden fazla evren ortalamasına ilişkin karşılaştırma, F Testi uygulamak

İkiden fazla evren ortalamasının karşılaştırılması istendiğinde t ve z testleri uygun olmamaktadır. Çünkü ikiden fazla evren ortalamasının karşılaştırılması t ve z testleri ile ikişerli kombinasyonlar hâlinde yapılması durumunda doğru karar verme riski artmaktadır. Bu nedenle bu tür karşılaştırmalar için F testi uygulanmaktadır. Bu test sürecinde de diğer test süreçlerindeki aşamalar izlenir, kullanılan F test istatistiği $F = \frac{S_B^2}{S_E^2}$ şeklinde yazılır. Örnek uygulamalarda hesaplanan test istatistiği ile berlirlenen anlamlılık düzeyinde ve hesaplanan serbestlik derecelerinde F tablo değeri ile karşılaştırma yapılır. Hesaplanan F istatistiğinin değeri tablodaki F değerinden büyükse sıfır hipotezi reddedilir.

Kendimizi Sınayalım

1. Örnek büyüklüğü 12 birim, anlam düzeyi 0.05 olan tek yönlü bir sınamadaki kritik t değeri kaçtır?

a. 1.782
b. 1.796
c. 1.812
d. 2.179
e. 2.201

2. ve 3. sorular aşağıdaki bilgilere göre cevaplandırılacaktır.

Pil üreten bir işletmede, ürünlerin ortalama ömrünü artırmak amacıyla bir üretim yöntemi uygulanacaktır. Bu amaçla uygulanacak yeni yöntem sınanmak istenmektedir.

2. Bu sınamada kurulacak sıfır hipotezi nedir?

a. Yeni yöntem pillerin ortalama ömrünü artırır.
b. Yeni yöntem ortalama pil üretimini artırır.
c. Yeni yöntem pillerin ortalama ömrünü değiştirmez.
d. Yeni yöntem pillerin ortalama ömrünü değiştirir.
e. Yeni yöntem ortalama pil üretimini azaltır.

3. Bu sınamada kurulacak alternatif hipotez nedir?

a. Yeni yöntem ortalama pil üretimini artırır.
b. Yeni yöntem ortalama pil üretimini azaltır.
c. Yeni yöntem ortalama pil üretimini değiştirir.
d. Yeni yöntem pillerin ortalama ömrünü artırır.
e. Yeni yöntem pillerin ortalama ömrünü değiştirir.

4. Aşağıdaki alternatif hipotezlerden hangisi çift yönlü bir sınama gerektirir?

a. A firmasının günlük ortalama üretim miktarı B firmasının günlük ortalama üretim miktarından fazladır.
b. A marka malın ortalama dayanıklılık süresi B marka malın ortalama dayanıklılık süresinden azdır.
c. A makinesinin aylık ortalama verimliliği B makinesinin aylık ortalama verimliliğinden farklıdır.
d. A marka ampülün ortalama ömrü B marka ampülün ortalama ömründen kısadır.
e. A marka malın aylık satış oranı B marka malın aylık satış oranından fazladır.

5. 0.02 anlam düzeyinde sınanan bir hipotez için, doğru olan sıfır hipotezini reddederek hatalı karar verme olasılığı kaçtır?

a. 0.01
b. 0.02
c. 0.05
d. 0.98
e. 0.99

6. ve 7. sorular aşağıdaki bilgilere göre cevaplandırılacaktır.

Bir işletmede, üretilen ampüllerin 450 saat olan dayanma süresini artırmak için, yeni bir hammaddenin kullanımı düşünülmektedir. Bu hammadde kullanılarak 1000 ürün üretilmiş ve ortalama dayanma süresi 462 saat olarak hesaplanmıştır. Hammaddenin olumlu sonuç verip vermediği %95 güvenle sınanacaktır.

6. Bu sınamada örnekleme dağılımının red bölgesi aşağıdakilerden hangisidir?

a. Sağ uçta, %2.5'lik alan
b. Sol uçta, %2.5'lik alan
c. Sağ uçta, %5'lik alan
d. Sol uçta, %5'lik alan
e. Sağ uçta, %10'luk alan

7. Bu sınamadaki alternatif hipotez nedir?

a. $\mu \neq 450$
b. $\mu > 462$
c. $\mu = 450$
d. $\mu \neq 462$
e. $\mu > 450$

8. Üniversiteye giriş sınavlarında kız öğrencilerin ortalama başarısının (μ_1) erkek öğrencilerin ortalama başarısından (μ_2) daha büyük olduğu iddia edilmektedir. Bu iddianın araştırılmasında sıfır hipotezi nedir?

a. $H_0 : \mu_1 > \mu_2$
b. $H_0 : \mu_1 < \mu_2$
c. $H_0 : \mu_1 - \mu_2 \neq 0$
d. $H_0 : \mu_1 - \mu_2 = 0$
e. $H_0 : \mu_1 - \mu_2 < 0$

9. Üniversiteye giriş sınavlarında kız öğrencilerin erkek öğrencilerden daha başarılı olduğu yönündeki iddianın araştırılması için her iki grup öğrenciden sırasıyla 16 ve 18 öğrenci rassal olarak seçilmiştir. Bu araştırmada serbestlik derecesi kaçtır?

a. 32
b. 34
c. 36
d. 2
e. 18

10. Bir tüketici organizasyonu 3 farklı firma tarafından üretilen kalem pillerin ortalama ömürlerinin karşılaştırılmasını amaçlamaktadır. Bu amaçla her firmanın ürettiği kalem pillerden rassal olarak 7'şer tane seçilmiş ve araştırma için gerekli olan veriler derlenmiştir. Derlenen veriler için hesaplanan bazı istatistikler ve veriler aşağıdaki tabloda verilmiştir.
Tablodaki verileri kullanarak hesaplanacak olan test istatistiğinin değeri nedir?

Kaynak	Sapma Kareler Toplamı
Gruplar Arası	45
Gruplar İçi	10

a. 55
b. 35
c. 4,5
d. 40,2
e. 0,22

Kendimizi Sınayalım Yanıt Anahtarı

1. b	Yanıtınız yanlış ise "Evren Ortalamasına İlişkin Küçük Örneklem Testi" ile ilgili bilgileri yeniden gözden geçiriniz.
2. c	Yanıtınız yanlış ise "Evren Ortalamasına İlişkin Hipotez Testi" ile ilgili bilgileri yeniden gözden geçiriniz.
3. d	Yanıtınız yanlış ise "Hipotezlerin İfade Edilmesi" ile ilgili bilgileri yeniden gözden geçiriniz.
4. c	Yanıtınız yanlış ise "Hipotezlerin İfade Edilmesi" ile ilgili bilgileri yeniden gözden geçiriniz.
5. b	Yanıtınız yanlış ise "Hipotez Testlerindeki Hata Türleriyle" ile ilgili bilgileri yeniden gözden geçiriniz.
6. c	Yanıtınız yanlış ise "Hipotez Testlerinde Red Bölgesinin Belirlenmesi" ile ilgili bilgileri yeniden gözden geçiriniz.
7. e	Yanıtınız yanlış ise "Hipotezlerin Formüle Edilmesi" ile ilgili bilgileri yeniden gözden geçiriniz.
8. d	Yanıtınız yanlış ise "Ortalama Farkların Testinde Hipotezlerin İfade Edilmesi" ile ilgili bilgileri yeniden gözden geçiriniz.
9. a	Yanıtınız yanlış ise "Küçük Örneklem Testi" ile ilgili bilgileri yeniden gözden geçiriniz.
10. d	Yanıtınız yanlış ise "F Testi" ile ilgili bilgileri yeniden gözden geçiriniz.

Sıra Sizde Yanıt Anahtarı

Sıra Sizde 1

- İstatistiksel hipotez, bir dağılımın evren parametresine ilişkin bir önermedir.
- İstatistiksel hipotez testi, örneklem istatistiklerini kullanarak, bir hipotezin doğru olup olmadığını ortaya koymaya yönelik yapılan çalışmalardır.

Sıra Sizde 2

- H_0 hipotezi gerçekten doğru ise bu hipotezin yanlışlıkla reddedilmesi olasılığıdır.
- Bir testte verilecek olan kararı evren parametresinden hem küçük hem de büyük yöndeki anlamlı farklılıklar etkileyecek ise bu test çift yönlü test ile test edilir. H_1 hipotezi ile ifade edilir ve genel gösterimi $H_1 : \theta \neq \theta_0$ şeklinde yapılır.

Sıra Sizde 3

- Örneklem hacmi $n \geq 30$ birim ise veya $n \,.\, \pi \geq 5$ ve $n \,.\, (1 - \pi) \geq 5$ koşullarını sağlıyorsa örneklem oranı p istatistiğinin örnekleme dağılımı normal olur. Bu durumda evren oranı π'ye ilişkin test için kullanılması gereken test istatistiği $z = \dfrac{p - \pi_0}{\sigma_p}$ olur. Burada standart hata $\sigma_p = \sqrt{\dfrac{\pi_0 - (1 - \pi_0)}{n}}$ eşitliğiyle hesaplanır.
- A = 0,10 - 2,33 . 0,024 = 0,044'tür.

Sıra Sizde 4

- Hipotezler

  $H_0 : \mu_1 - \mu_2 = 0$, $H_1 : \mu_1 - \mu_2 \neq 0$ veya $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 \neq \mu_2$

  şeklinde ifade edilirler. Test edilecek hipotez H_0 hipotezidir.
- Örneklem hacmi küçük olduğunda evrenin dağılım şekli biliniyor ve normalse.
- t test istatistiği kullanılır.
- Serbestlik derecesi sd = n_1+n_2-2 = 15+10–2=23 olur.

Sıra Sizde 5

- Grup sayısı r = 4 olduğu için, gruplar arası serbestlik derecesi; $sd_1 = r - 1 = 4 - 1 = 3$ olur.
- Gruplar içi değişkenlik her gruptaki gözlem değerlerinin kendi grup ortalamasından sapmalarının toplamına ilişkin değişkenlği gösterir. Bu değişkenlik grup içi sapmaların karelerinin toplamlarının grup içi serbestlik derecesine bölünmesiyle hesaplanır.

Yararlanılan ve Başvurulabilecek Kaynaklar

Canküyer, E., Aşan, Z. (2001). **Parametrik Olmayan İstatistiksel Teknikler**, Eskişehir: Anadolu Üniversitesi Yayınları, No: 1266.

Çömlekçi, N. (2001). **Bilimsel Araştırma Yöntemi ve İstatistiksel Anlamlılık Sınamaları**, İstanbul: Bilim Teknik Yayınevi.

Çömlekçi, N. (2005). **Temel İstatistik İlke ve Teknikleri**, Bilim Teknik Yayınevi, İstanbul.

Fink, A. (1995). **How To Sampling in Surveys**, London: Sage Publication.

Gürsakal, N. (1997). **Bilgisayar Uygulamalı İstatistik 1**, Bursa: Marmara Kitabevi.

Hinkle, D. E., Wiersma, W., Jurs S. G. (1998). **Applied Statistics For The Behavioral Sciences**, Boston.

Işığıçok, E. (2011). **Altı Sigma Kara Kuşaklar İçin Hipotez Testleri Yol Haritası**, Marmara Kitabevi Yayınları, Bursa.

Malhotra, N. K. (1996). **Marketing Research An Applied Orientation**, New Jersey: Prentice Hall International.

Nakip, M. (2003). **Pazarlama Araştırmaları Teknikler ve SPSS Destekli Uygulamalar**, Seçkin Yayıncılık, Ankara.

Neter, J., Wasserman, W., Whitmore, G. A. (1993). **Applied Statistics**, Boston: Simon and Schuster.

Özmen, A., Özdamar, K., Odabaşı, Y., Hoşcan, Y., Bir, A. A., Kırcaaliiftar, G., Uzuner, Y. (1999). **Sosyal Bilimlerde Araştırma Yöntemleri**, Eskişehir: TC. Anadolu Üniversitesi Yayınları No: 1081; Açıköğretim Fakültesi Yayınları No: 601.

Püskülcü, H., İkiz, F. (1986). **İstatistiğe Giriş**, (2. Baskı), İzmir: E. Ü. Mühendislik Fakültesi Yayın No: 601, Ege Üniversitesi Basımevi.

Serper, Ö., Aytaç, M. (2000). **Örnekleme**, Bursa: Ezgi Kitabevi.

Serper, Ö. (1986). **Uygulamalı İstatistik 2**, İstanbul: Filiz Kitabevi.

Tryfos, P. (1996). **Sampling Methods for Applied Research**, New York: John Wiley and Sons Inc.

Tull, D. D.,Hawkins, D. I. (1993). **Marketing Research Measurement and Method**, 6th Edition, New York: MacMillan Publishing Company.

İSTATİSTİK-II

4

Amaçlarımız

Bu üniteyi tamamladıktan sonra;

- Sayısal olmayan değişkenler arasındaki ilişkinin varlığını test edebilecek,
- Farklı örneklemlerin aynı evrenden seçilip seçilmediğini test edebilecek,
- n hacimli bir örneklemin, ilgili evreni, iyi temsil edip edemediğini belirleyebilecek,
- Sayısal olmayan iki değişken arasındaki ilişkinin derecesini belirleyebilecek bilgi ve becerilere sahip olacaksınız.

Anahtar Kavramlar

- Kontenjans Tablosu
- Serbestlik Derecesi

İçindekiler

Ki-Kare Testi

GİRİŞ

Daha önce de belirtildiği gibi, istatistikte değişkenler sayısal (nicel) değişkenler ve sayısal olmayan (nitel) değişkenler olmak üzere iki grupta sınıflandırılmaktadır. Günümüzde yapılan birçok araştırmada sayısal olmayan değişkenlerin dikkate alındığı gözlemlenmektedir. Örneğin, insanların medeni durumlarıyla seçtikleri meslek grupları arasındaki bir ilişki incelenmek istendiğinde, medeni durumun ve meslek grubunun rakamlarla ifade edilmesi olası değildir. Medeni durum "evli", "bekâr", "boşanmış" ve "dul" şeklinde gösterilirken meslek grupları da "serbest meslek", "devlet memurluğu", "işçi" vb. şeklinde gruplandırılabilir.

İşte sayısal olmayan değişkenler arasındaki herhangi bir ilişkinin var olmadığını ileri sürerek (H_0 hipotezi) bu hipotezin red edilip edilemeyeceğinin incelenmesinde uygulanan Ki-Kare testi'dir.

Bir örneklemin gözlemlenmesi sonucunda elde edilen frekans dağılımının binom, Poisson, normal vb. gibi genellenmiş bir dağılıma uygun olup olmadığına karar verebilmek için kullanılan test yine Ki-kare testi olacaktır. Diğer yandan iki ya da daha fazla örneklemin aynı evrenden seçilip seçilmedikleri konusunda karar verilirken de ki-kare testinden yararlanılır.

1900 yıllarında Karl Pearson tarafından bulunan ve ismi de onun tarafından verilen bu istatistiksel testin uygulanmasında önce, ki-kare'nin ve serbestlik derecesinin nasıl hesaplanacağının bilinmesi gerekir. Bunlar bağımsızlık, homojenlik ve uygunluk testleri için ünite bölümlerinde ayrı ayrı gösterilecektir.

KİTAP

Bu ünitenin izleyen kısımlarındaki konular için A. F. Yüzer, E. Şıklar, E. Ağaoğlu, H. Tatlıdil, A. Özmen, Editör: A. F. Yüzer (2011). *İstatistik,* Ünite 8 ve 9, Eskişehir Anadolu Üniversitesi Yayını İsimli kitaptan, editör ve ünite yazarlarının izni alınarak yararlanılmıştır.

Kİ-KARE BAĞIMSIZLIK TESTİ

Bir seçim sonrası bir il merkezindeki yerel basın, seçime katılan partilerin aldıkları oylarla seçmenlerin eğitim düzeyleri arasında, göz ardı edilemeyecek, bir ilişkinin varlığını ileri sürmektedir. Oy dağılımına ilişkin farklı görüş taşıyan A partisi yöneticileri, yerel basının bu konuda ne kadar haklı olduğunu belirlemek amacıyla bir araştırmanın yaptırılmasını kararlaştırmıştır. Araştırma, yeterli görülen bir örneklem üzerinden gerçekleştirilecektir.

Bu ve benzeri problemlerin çözümlenmesinde uygun bir istatistiksel teknik de aşağıda yeterli ayrıntıyla ele alınacak olan ki-kare testidir.

İki ya da daha çok sınıflı nitel değişkenler arasındaki bağımsızlık, **ki-kare bağımsızlık testi**yle araştırılır.

İki ya da daha fazla sınıflı iki nitel değişken arasında bağımsızlık olup olmadığını incelemek için, **ki-kare bağımsızlık testi**ne başvurmak gerekir. Bu test yapılırken kontenjans tablosundan yararlanılmaktadır. Bu tablo, incelenen iki değişkenin şıklarına düşen gözlenen frekansların yazıldığı, yatay (satırlar) ve düşey (sütunlar) bantlardan oluşan, çift yönlü tablodur. Ki-kare bağımsızlık ve homojenlik testlerini yapabilmek üzere hazırlanacak kontenjans tablosunun yapısı Tablo 4.1'de gösterilmiştir.

Tablo 4.1
Kontenjans tablosunun yapısı.

	2. Değişkenin Şıkları							
1. Değişkenin Şıkları	1	2	3	....	j	....	c	Toplam
1	n_{11}	n_{12}	n_{13}	....	n_{1j}	....	n_{1c}	n_1
2	n_{21}	n_{22}	n_{23}	....	n_{2j}	....	n_{2c}	n_2
3	n_{31}	n_{32}	n_{33}		n_{3j}		n_{3c}	n_3
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
i	n_{i1}	n_{i2}	n_{i3}	....	n_{ij}	....	n_{ic}	n_i
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
R	n_{r1}	n_{r2}	n_{r3}	....	n_{rj}	....	n_{rc}	n_r
Toplam	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot 3}$	....	$n_{\cdot j}$	....	$n_{\cdot c}$	$n_{..} = n$

Aralarında bağıntı bulunduğu düşünülen birinci değişkenin r şıkkı (satır), ikinci değişkenin c şıkkı (sütun) varsa rxc TABLOSU olarak da isimlendirilen tablo oluşturulur. Satır ve sütunların kesiştikleri yerlerde bulunan gözelerdeyse ilgili frekanslar kaydedilir.

Ki-kare bağımsızlık testinin nasıl uygulandığını bir örnek yardımıyla açıklayalım.

ÖRNEK 1

Yapılan bir çalışmada katılımcıların eğitim düzeyleri ve sigara içme alışkanlıkları sorgulanarak, bu iki değişken arasında bir bağıntı bulunup bulunmadığı, diğer bir ifadeyle iki değişkenin birbirinden bağımsız olup olmadığı belirlenmeye çalışılsın. Bu amaçla 300 kişiyi kapsayan bir örneklem üzerinde yapılan gözlem sonuçları aşağıdaki tablo ile verilmiştir.

Tablo 4.2
Gözlenen frekanslar.

Sigara İçme Alışkanlığı	Eğitim Düzeyi			**TOPLAM**
	İlk	Orta	Yüksek	
Sigara İçiyor	45	55	40	140
Sigara İçmiyor	45	65	50	160
TOPLAM	90	120	90	300

Sigara içme alışkanlığına ilişkin eğitim düzeyinin etkili olup olmadığını $\alpha = 0.01$ anlamlılık düzeyinde araştırınız.

Çözüm: Tabloda yer alan sayılar "gözlenen frekanslardır". Sigara içme alışkanlığı üzerinde eğitim düzeyinin etkisi olup olmadığını test edebilmek için (bağımsızlık testini yapabilmek için), izlenmesi gereken adımları sırasıyla yerine getirelim:

1. Adım: Hipotezlerin ifade edilmesi
   Sıfır Hipotezi (H_0): Sigara içme alışkanlığı ile eğitim düzeyi birbirinden bağımsız değişkenlerdir. Bu iki değişken arasında bir ilişki (bağıntı) yoktur.
   Karşıt Hipotez (H_1): Sigara içme alışkanlığı ile eğitim düzeyi arasında bir ilişki (bağıntı) vardır.
2. Adım: İstatistiksel test
   İki sayısal olmayan değişken arasındaki ilişkinin varlığını araştıran bir test olan x^2 (ki-kare) bağımsızlık testi olmalıdır.
3. Adım: Anlamlılık düzeyinin belirlenmesi
   α= 0.01 olarak belirlenmiştir.
4. Adım: H_0'ın red bölgesinin belirlenmesi
   Bunun için hesaplanan test istatistiği, belli bir anlamlılık düzeyine ve v = (r-1) (c-1) **serbestlik derecesi**ne göre "x^2 değerleri tablosu"ndan bulunan "kritik değer" ile karşılaştırılır. Örneğimiz için serbestlik derecesi v = (2-1) (3-1) = 2 olup α = 0.01 düzeyinde x^2 tablosundan bulunan kritik değer x_k^2 = 9.21'dir. Eğer hesaplanan x^2 istatistiğinin değeri tablodan bulunan x_k^2 kritik değerden büyük çıkarsa H_0 red edilecektir.
5. Adım: x^2 test istatistiğinin hesaplanması
   Test istatistiği

$$x^2 = \sum \frac{(G - B)^2}{B}$$

formülüyle elde edilebilir. Formülde,
G = Gözlenen frekansları
B = Beklenen frekansları
ifade etmektedir. Test istatistiğinin hesaplanabilmesi için öncelikle **beklenen (kuramsal) frekanslar**ın hesaplanması gerekmektedir. Herhangi bir gözenin beklenen frekansı bulunurken, o gözenin yer aldığı satır toplam frekansıyla sütunun toplam frekansı çarpılıp genel toplam frekansa bölünmektedir. Örneğimiz için, beklenen frekansları, ilk gözeden başlamak üzere sırasıyla hesaplayalım:

B_{11} (birinci satır birinci sütunda yer alacak frekans)
B_{11} = (birinci satır toplamı x birinci sütun toplamı) / (genel toplam)
= (140 x 90) / (300) = 42
B_{12} = (birinci satır toplamı x ikinci sütun toplamı) / (genel toplam)
= (140 x 120) / (300) = 56

Aynı yöntemle hesaplanan beklenen frekansları ve gözlenen frekansları kontenjans tablosunda gösterelim (G = Gözlenen frekanslar, B = Beklenen frekanslar).

Kontenjans tablolarında **serbestlik derecesi,** satır ve sütun sayılarından birer çıkartılarak, bunların çarpılması suretiyle elde edilir.

Kuramsal (beklenen) frekanslar, ilgili gözenin yer aldığı satır toplamıyla sütun toplamı çarpılarak genel toplama bölünmek suretiyle elde edilir.

Tablo 4.3 Kontenjans tablosu (Gözlenen ve beklenen frekanslar).

Sigara İçme Alışkanlığı	Eğitim Düzeyi						TOPLAM
	İlk		Orta		Yüksek		
	G	B	G	B	G	B	
Sigara İçiyor	45	42	55	56	40	42	140
Sigara İçmiyor	45	48	65	64	50	48	160
TOPLAM	90		120		90		300

Test istatistiği:
x^2 = (45 - 42)2 / (55 - 56)^2 / (56) + (40 - 42)^2 / (42) + (45 - 48)^2 / (48) + (65 -64)^2 / (64) + (50 - 48)^2 / (48) = 0.58

6. Adım: İstatistiksel Karar

İstatistiksel karar verilirken, red bölgesinin tanımı gereği, $x^2 > x_k^2$ olduğunda sıfır hipotezi red edilir. $x^2 \leq x_k^2$ olduğundaysa sıfır hipotezi reddedilemez. Sıfır hipotezinin kabul edilmesi, değişkenlerin birbirinden bağımsız olduğu (diğer bir ifadeyle değişkenler arasında bir ilişki bulunmadığı) anlamı taşır. Buna göre örneğimizde,

$$x^2 = 0.58 \quad x_k^2 = 9.21 \qquad \text{ve} \qquad x^2 < x_k^2$$

olduğundan H_0 hipotezi red edilecektir. Başka bir anlatımla, sigara içme alışkanlığı ile eğitim düzeyi arasında anlamlı bir ilişki yoktur.

SIRA SİZDE 1

- **Ki-kare testi hangi durumlarda yapılır? Açıklayınız.**
- **Ki-kare bağımsızlık testinde hipotezler nasıl ifade edilir? Açıklayınız.**
- **Ki-kare bağımsızlık testinde serbestlik derecesi nasıl hesaplanır?**

Kİ-KARE HOMOJENLİK TESTİ

Farklı örneklemlerin aynı evrenden seçilip seçilmediği, **ki-kare homojenlik** testiyle araştırılır.

Ki-kare homojenlik testi ana çizgileriyle iki ya da daha fazla bağımsız örneklemin, aynı anakütleden seçilip seçilmediğinin araştırılmasında kullanılır. Testin uygulanması, ki-kare bağımsızlık testinde olduğu gibidir. Yine nitel değişkenlerle ve aynı örneklem istatistiğiyle çalışır. Ancak, dikkat edilmelidir ki, bağımsızlık testinde ele alınan değişkenler arasında bir ilişkinin varlığı araştırılırken, homojenlik testinde iki ya da daha fazla bağımsız örneklemin aynı evrenden seçilip seçilmediği araştırılmaktadır.

ÖRNEK 2

Bir fabrika; A ve B olmak üzere iki farklı teknik uygulanarak üretilen ürünlerin yıpranma sürelerini (kısa sürede, orta sürede, uzun sürede) belirlemek amacıyla, bu ürünlerle ilgili iki rassal örneklem oluşturmuştur. A tekniğiyle üretilen ürünlerden seçilen örneklemde 60 ürün, B tekniğiyle üretilen ürünlerden seçilen örneklemde ise 80 ürün bulunmaktadır. Veriler aşağıdaki tabloda belirtilmiştir. Seçilen örneklemlerin aynı anakütleye ait olup olmadığını, %5 anlamlılık düzeyinde test ediniz.

GÖZLENEN FREKANSLAR			
Yıpranma Süresi	A Tekniği	B Tekniği	**TOPLAM**
Kısa Sürede	30	30	60
Orta Sürede	45	35	80
Uzun Sürede	35	25	60
TOPLAM	110	90	200

Çözüm: Çözüm adımları aşağıdaki gibidir:

1. Adım: Hipotezlerin oluşturulması
   Sıfır Hipotezi (H_0): İki örneklem de aynı anakütleden seçilmiştir.
   Karşıt Hipotez (H_1): Örneklemler farklı anakütlelerden seçilmiştir.
2. Adım: İstatistiksel test
   İki örneklemin aynı anakütleden gelip gelmediği test edileceğinden, ilgili test, ki-kare homojenlik testi olmalıdır.
3. Adım: Anlamlılık Düzeyi
   $\alpha = 0.05$
4. Adım: H_0'ın red bölgesinin belirlenmesi
   Hesaplanan test istatistiği v = (3-1) (2-1) = 2 serbestlik derecesi ve $\alpha = 0.05$ anlamlılık düzeyi için ki-kare tablosundan bulunan kritik değer x_k^2 = 5.99'dur. Eğer hesaplanan x^2 istatistiğinin değeri x_k^2 = 5.99'dan büyük çıkarsa H_0 red edilecektir.
5. Adım: x^2 test istatistiğinin hesaplanması
   Hatırlanacağı gibi, test istatistiğinin hesaplanabilmesi için, öncelikle beklenen frekansların hesaplanması gerekir. Homojenlik testinde de herhangi bir gözenin beklenen frekansı, bağımsızlık testindeki gibi, ilgili gözenin yer aldığı satır toplam frekansıyla sütun toplam frekansı çarpılıp, genel toplam frekansına bölünerek elde edilir. İlgili kontenjans tablosu aşağıdaki gibidir:

Yıpranma Süresi	A Tekniği		B Tekniği		TOPLAM
	G	B	G	B	
Kısa Sürede	30	33	30	27	60
Orta Sürede	45	44	35	36	80
Uzun Sürede	35	33	25	27	60
TOPLAM	110		90		200

Test istatistiği:
$x^2 = (30 - 33)^2 / (33) + (30 - 27)^2 / (27) + (45 - 44)^2 / (44) + (35 - 36)^2 / (36) + (35 - 33)^2 / (33) + (25 - 27)^2 / (27) = 0.9$
olarak elde edilir.

6. Adım: İstatistiksel Karar
   Hatırlanacağı gibi, x2 < x_k^2 ise H_0 hipotezi kabul edilir. Örneğimizde $x^2 = 0.9$ ve x_k^2 = 5.99 olduğundan, H_0 kabul edilecektir. Başka bir anlatımla ilgili örneklemler aynı anakütleden seçilmiştir.

SIRA SİZDE 2

- **Bağımsızlık ve homojenlik testleri hangi açılardan birbirlerinden farklıdır? Açıklayınız.**
- **Bağımsızlık ve homojenlik testlerinde beklenen frekanslar nasıl hesaplanır? Açıklayınız.**

Kİ-KARE UYGUNLUK (İYİ UYUM) TESTİ

n birimlik örneklemin çekildiği, evreni, iyi temsil edip edemeyeceği, **ki-kare uygunluk testi** ile araştırılır.

Ki-kare uygunluk testinin esası, n hacimli (birimlik) bir örneklemin anakütleyi iyi temsil edip edemeyeceğini araştırmak oluşturur. Bu testte, yine x^2 değişkeninin doğası gereği, gözlenen ve beklenen frekanslardan yararlanılır. Testin nasıl yapılacağı, özellikle beklenen frekansların nasıl hesaplanacağı, aşağıdaki örnek yardımıyla açıklanmaya çalışılmıştır.

ÖRNEK 3

Bir üniversitede ortak ders olarak tüm fakültelerde verilen İngilizce dersini alan ve başarılı olan öğrencilerden rassal olarak seçilen 150 öğrencinin fakültelere dağılımı aşağıda verilmiştir .

Fakülteler	**Başarılı**
İktisadi ve İdari Bilimler Fakültesi	24
Fen Fakültesi	28
Mühendislik Fakültesi	30
Hukuk Fakültesi	20
Eğitim Fakültesi	26
İletişim Fakültesi	22

Bu verilere göre fakülteler için İngilizce dersi başarısının aynı oranda olup olmadığını $\alpha = 0.01$ anlamlılık düzeyinde araştırınız.

Çözüm: Çözüm adımları aşağıdaki gibidir:

1. Adım: Hipotezlerin oluşturulması
   H_0: Tüm fakülteler için İngilizce dersinin başarı oranları aynıdır. (İngilizce başarısında fakülteler açısından farklılık yoktur)
   H_1: En az bir fakülte için İngilizce dersinin başarı oranı farklıdır.
2. Adım: İstatistiksel test
   x^2 uygunluk (iyi uyum) testi
3. Adım: Anlamlılık Düzeyi
   $\alpha = 0.01$
4. Adım: H_0'ın red bölgesinin belirlenmesi
   Hatırlanacağı gibi, red bölgesi, hesaplanan x^2 değerinin öngörülen anlamlılık düzeyi ve belirlenen serbestlik derecesine göre, x^2 tablosundan bulunan kritik değeriyle karşılaştırılarak belirlenir. k sınıf sayısını belirtmek üzere, serbestlik derecesi v = k-1'den; 6 sınıf olduğu için 6 - 1 = 5 olarak belirlenir. 5 serbestlik derecesi ve $\alpha = 0.01$ anlamlılık düzeyi için kritik x_k^2 değeri, x^2 tablosundan 15.08 olarak bulunur. Red bölgesi $x^2 > 15.08$ olarak belirlenir.
5. Adım: Ki-kare istatistiğinin hesaplanması
   Sıfır hipotezinde, tüm fakülteler için İngilizce dersinin başarı oranlarının aynı olduğu ileri sürüldüğü için, altı farklı fakülte için genel oran 1 / 6 olacaktır. Dolayısıyla her bir fakülte için "beklenen frekans" 150.(1/6) = 25 olur. Bunun anlamı, fakülteler arasında başarı oranı açısından farklılık olmadığı ve her fakülteden 25 öğrencinin başarılı olmasının beklenmesidir.

Fakülteler	Başarılı	
	G	B
İktisadi ve İdari Bilimler Fakültesi	24	25
Fen Fakültesi	28	25
Mühendislik Fakültesi	30	25
Hukuk Fakültesi	20	25
Eğitim Fakültesi	26	25
İletişim Fakültesi	22	25

Test istatistiği $x^2 = \sum \frac{(G - B)^2}{B}$ formülüyle

$x^2 = (24 - 25)^2 / (25) + (28 - 25)^2 / (25) + (30 - 25)^2 / (25) + (20 - 25)^2 / (25) + (26 - 25)^2 / (25) + (22 - 25)^2 / (25) = 2.8$

olarak elde edilir.

6. Adım: İstatistiksel Karar

$x^2 = 2.8$

$x_k^2 = 15.08$

olarak bulunmuştur. Bu sonuçlara göre,

$x^2 < x_k^2$

olduğundan H_0 kabul edilecektir. Başka bir anlatımla bu üniversitenin tüm fakülteleri için ingilizce dersinin başarı oranları arasında önemli bir farklılık yoktur.

Ki-kare uygunluk testi hangi amaçla yapılır?

SIRA SİZDE 3

KONTENJANS KATSAYISI

Ki-kare bağımsızlık testiyle iki değişken arasındaki ilişkinin varlığıyla ilgili karar verilebiliyordu. Oysa ki bazı hâllerde, iki değişken arasındaki ilişkinin kuvveti hakkında da bilgi sahibi olmak istenebilir. İşte **kontenjans katsayısı** rxc kontenjans tablolarından (r > 2 ve c > 2) hesaplanan x^2 değerinin gösterdiği ilişki düzeyini saptamak amacıyla kullanılan bir katsayıdır. İki değişken arasında bir ilişki bulunmuyorsa c = 0 değeri verir. Buna karşılık iki değişken arasında en üst düzeydeki ilişki katsayısı her zaman 1 çıkmaz, 1'e çok yakın bir değer olur. c ile gösterilen kontenjans katsayısının formülü,

Sayısal olmayan iki değişken arasındaki ilişkinin derecesi, **kontenjans katsayısıyla** belirlenir.

$$c = \sqrt{\frac{x^2}{x^2 + n}}$$

şeklindedir.

ÖRNEK 4

Yapılan bir araştırmada, Z ilinde yaşayanların gelir düzeyleri (düşük, orta, yüksek) ile kullandıkları araçların yakıt özellikleri (benzin, dizel, LPG) arasında anlamlı bir ilişki olup olmadığı incelenmek istenmiştir. Bu amaçla rassal olarak seçilen 200 kişiden elde edilen verilerle 0.01 anlamlılık düzeyinde ki-kare bağımsızlık testi yapılarak; x^2 değeri 42.93 olarak hesaplanmış ve söz konusu iki değişken arasında anlamlı bir ilişki olduğu sonucuna ulaşılmıştır. Şimdi kontenjans katsayısıyla bu ilişkinin derecesini araştıralım.

Çözüm: $x^2 = 42.93$ ve n = 200

olduğuna göre,

$$c = \sqrt{\frac{42.93}{42.93 + 200}}$$

$$c = 0.42$$

elde edilir. Bu durumda, orta düzeyde bir ilişkinin olduğu konusunda karar verilebilir.

- **Kontenjans katsayısı, hangi amaçla kullanılır?**
- **Kontenjans katsayısının değeri, hangi durumda sıfır olur?**

SIRA SİZDE 4

Özet

AMAÇ 1

Sayısal olmayan değişkenler arasındaki ilişkinin varlığını test etmek

İstatistikte değişkenler sayısal (nicel) değişkenler ve sayısal olmayan (nitel) değişkenler olmak üzere iki grupta sınıflandırılmaktadır. İşte sayısal olmayan değişkenler arasındaki herhangi bir ilişkinin var olmadığını ileri sürerek (H_0 hipotezi), bu hipotezin red edilip edilemeyeceğinin incelenmesinde uygulanan Ki-Kare Bağımsızlık Testi'dir.

AMAÇ 2

Farklı örneklemlerin aynı evrenden seçilip seçilmediğini test etmek

İki ya da daha fazla bağımsız örneklemin, aynı anakütleden seçilip seçilmediğinin araştırılmasında Ki-kare Homojenlik Testi kullanılır.

AMAÇ 3

n hacimli bir örneklemin, ilgili evreni, iyi temsil edip edemediğini belirlemek

n hacimli (birimlik) bir örneklemin anakütleyi iyi temsil edip edemeyeceğini araştırılmasında Ki-kare Uygunluk Testi kullanılır.

AMAÇ 4

Sayısal olmayan iki değişken arasındaki ilişkinin derecesini belirlemek

rxc kontenjans tablolarından ($r > 2$ ve $c > 2$) hesaplanan x^2 değerinin gösterdiği ilişki düzeyini saptamak amacıyla kullanılan kontenjans katsayısı sayısal olmayan iki değişken arasındaki ilişkinin derecesini belirler.

Kendimizi Sınayalım

1. Sayısal olmayan değişkenler arasında herhangi bir ilişkinin (bağıntının) var olup olmadığının test edilmesinde aşağıdakilerden hangisi kullanılır?

a. Küçük örneklem testi
b. Ki-kare uygunluk testi
c. Ki-kare bağımsızlık testi
d. Büyük örneklem testi
e. Ki-kare homojenlik testi

2. Ki-Kare bağımsızlık testinde serbestlik derecesi aşağıdakilerden hangisi ile hesaplanır?

a. v = (r-1) + (c-1)
b. v = (r+1) (c+1)
c. v = r / (c-1)
d. v = (r-1) (c-1)
e. v = (r+1) - (c+1)

3. - 6. sorular aşağıdaki bilgilere göre cevaplandırılacaktır.

Bir araştırma sonucunda elde edilen gözlem sonuçları aşağıdaki tabloda gösterilmiştir.

	Başarı Durumu		
Barınma Şekli	Başarılı	Başarısız	**TOPLAM**
Ailenin yanında	55	35	**90**
Devlet Yurdunda	64	46	**110**
Özel Yurtta	61	39	**100**
TOPLAM	**180**	**120**	**300**

3. Yukarıdaki tabloya göre, sıfır hipotezinin doğru ifadesi aşağıdakilerden hangisidir?

a. Öğrencilerin başarı durumları ile barınma şekli birbirinden bağımsızdır.
b. Öğrencilerin başarı durumları ile barınma şekli birbirine bağımlıdır.
c. Devlet yurdunda kalan öğrenciler daha başarılıdır.
d. Öğrencilerin başarıları birbirinden farklıdır.
e. Ailenin yanında kalan öğrenciler başarısızdır.

4. Yukarı verilen tabloya ilişkin serbestlik derecesi kaçtır?

a. 1
b. 2
c. 3
d. 4
e. 5

5. Yukarıda verilen tabloda devlet yurdunda kalan ve başarılı olan öğrencilere ait beklenen frekans değeri kaçtır?

a. 46
b. 58
c. 64
d. 62
e. 66

6. Yukarıda verilen tabloya göre, 0.05 anlamlılık düzeyinde test yaparken H_0 hipotezinin red bölgesi aşağıdakilerden hangisidir?

a. $x^2 > 3.84$
b. $x^2 > 0.10$
c. $x^2 > 7.81$
d. $x^2 > 5.99$
e. $x^2 > 10.59$

7. Kontenjans katsayısının sıfır çıkması hangi anlama gelir?

a. İki değişken arasında üst düzey bir ilişki olduğunu belirtir.
b. İki değişken arasında orta düzey bir ilişki olduğunu belirtir.
c. İki değişken arasında bir ilişki olmadığını belirtir.
d. İki değişken arasında alt düzey bir ilişki olduğunu belirtir.
e. İki değişkenin arasında tam bir ilişki olduğunu belirtir.

8. Farklı örneklemlerin aynı evrenden seçilip seçilmediğinin belirlenmesi için yapılan test aşağıdakilerden hangisidir?

a. Küçük örneklem testi
b. Ki-kare uygunluk testi
c. Ki-kare bağımsızlık testi
d. Büyük örneklem testi
e. Ki-kare homojenlik testi

9. Ki-kare uygunluk testi hangi amaçla yapılır?

a. n hacimli bir örneklemin evreni iyi temsil edip edemeyeceğini araştırmak için
b. İki değişken arasındaki ilişkinin kuvveti hakkında da bilgi sahibi olmak için
c. Farklı örneklemlerin aynı evrenden seçilip seçilmediğini araştırmak için
d. İki nitel değişken arasında bir ilişki olup olmadığını araştırmak için
e. Evrendeki değişkenlerin homojen olup olmadığını araştırmak için

10. 50 kişiden elde edilen verilerin değerlendirildiği bir araştırmada ki-kare istatistiğinin değeri 0.70 olarak hesaplanmıştır. Bu bilgilere göre kontenjans katsayısınn değeri nedir?

a. 0.20
b. 0.11
c. 0.04
d. 0.55
e. 0.38

Kendimizi Sınayalım Yanıt Anahtarı

1. c	Yanıtınız yanlış ise "Ki-Kare Bağımsızlık Testi" konusunu yeniden gözden geçiriniz.
2. d	Yanıtınız yanlış ise "Ki-Kare Bağımsızlık Testi" konusunu yeniden gözden geçiriniz.
3. a	Yanıtınız yanlış ise "Ki-Kare Bağımsızlık Testinde Hipotezlerin İfadesi" konusunu yeniden gözden geçiriniz.
4. b	Yanıtınız yanlış ise "Ki-Kare Bağımsızlık Testi" konusunu yeniden gözden geçiriniz.
5. e	Yanıtınız yanlış ise "Ki-Kare Bağımsızlık Testi" konusunu yeniden gözden geçiriniz.
6. d	Yanıtınız yanlış ise "Ki-Kare Bağımsızlık Testi" konusunu yeniden gözden geçiriniz.
7. c	Yanıtınız yanlış ise "Kontenjans Katsayısı" konusunu yeniden gözden geçiriniz.
8. e	Yanıtınız yanlış ise "Ki-Kare Homojenlik Testi" konusunu yeniden gözden geçiriniz.
9. a	Yanıtınız yanlış ise "Ki-Kare Uygunluk Testi" konusunu yeniden gözden geçiriniz.
10. b	Yanıtınız yanlış ise "Kontenjans Katsayısı" konusunu yeniden gözden geçiriniz.

Sıra Sizde Yanıt Anahtarı

Sıra Sizde 1

- Sayısal olmayan değişkenler arasındaki herhangi bir ilişkinin var olmadığını ileri sürerek (H_0 hipotezi), bu hipotezin red edilip edilemeyeceğinin incelenmesinde uygulanan Ki-Kare testi'dir. Bir örneklemin gözlemlenmesi sonucunda elde edilen frekans dağılımının binom, Poisson, normal vb. gibi genellenmiş bir dağılıma uygun olup olmadığına karar verebilmek için kullanılan test yine Ki-kare testi olacaktır. Diğer yandan iki ya da daha fazla örneklemin aynı evrenden seçilip seçilmedikleri konusunda karar verilirken de ki-kare testinden yararlanılır.
- Ki-Kare bağımsızlık testinde sıfır hipotezi (H_0), sayısal olmayan değişkenler arasındaki herhangi bir ilişkinin var olmadığını ileri sürerek; karşıt hipotez (H_1) ise sayısal olmayan değişkenler arasındaki bir ilişkinin varolduğunu ileri sürerek ifade edilir.
- Ki-Kare bağımsızlık testinde serbestlik derecesi hesaplanırken; aralarında bağıntı bulunduğu düşünülen birinci değişkenin r şıkkı (satır), ikinci değişkenin c şıkkı (sütun) varsa serbestlik derecesi, v = (r-1) (c-1) ile hesaplanır.

Sıra Sizde 2

- Ki-Kare bağımsızlık testinde ele alınan değişkenler arasında bir ilişkinin varlığı araştırılırken, ki-kare homojenlik testinde iki ya da daha fazla bağımsız örneklemin aynı evrenden seçilip seçilmediği araştırılmaktadır.
- Ki-Kare bağımsızlık ve homojenlik testlerinin ikisinde de herhangi bir gözenin beklenen frekansı, ilgili gözenin yer aldığı satır toplam frekansıyla sütun toplam frekansı çarpılıp, genel toplam frekansına bölünerek elde edilir.

Sıra Sizde 3

Ki-kare uygunluk testi, n hacimli (birimlik) bir örneklemin anakütleyi iyi temsil edip edemeyeceğini araştırmak için yapılır.

Sıra Sizde 4

- Kontenjans katsayısı; rxc kontenjans tablolarından ($r > 2$ ve $c > 2$) hesaplanan x^2 değerinin gösterdiği ilişki düzeyini saptamak amacıyla kullanılan bir katsayıdır.
- İki değişken arasında bir ilişki bulunmadığında kontenjans katsayısı sıfır (c=0) olur.

Yararlanılan ve Başvurulabilecek Kaynaklar

Canküyer, E., Aşan, Z. (2001). **Parametrik Olmayan İstatistiksel Teknikler,** Eskişehir: Anadolu Üniversitesi Yayınları, No 1266.

Newbold, P. (2000). **İşletme ve İktisat İçin İstatistik,** Çeviren: Ümit Şenesen. Literatür Yayıncılık.

Serper, Ö. (1986). **Uygulamalı İstatistik 2,** İstanbul: Filiz Kitabevi.

Yüzer, A. F. (1996) **Olasılık ve İstatistik,** Eskişehir: Anadolu Üniversitesi Yayınları, No 911

5

Amaçlarımız

Bu üniteyi tamamladıktan sonra;

- İki değişken arasındaki ilişkiyi açıklayan doğrusal modeli kurabilecek,
- İki değişken arasındaki ilişkinin derecesini belirleyebilecek

bilgi ve becerilere sahip olacaksınız.

Anahtar Kavramlar

- Regresyon Analizi
- Basit Doğrusal Regresyon Analizi
- En Küçük Kareler Tekniği
- Varyansın(σ^2) Tahmini
- Basit Doğrusal Regresyonda Aralık Tahmini
- Regresyon Katsayısının Anlamlılık Testi
- Korelasyon Analizi
- Korelasyon Katsayısı
- Belirlilik Katsayısı
- Korelasyon Katsayısının Anlamlılık Testi

İçindekiler

Regresyon ve Korelasyon Analizi

GİRİŞ

Regresyon teriminin kökeni Sör Francis Galton'a dayanır. Galton bir değişken üzerinde meydana gelen değişimin başka bir ya da daha fazla değişken tarafından açıklanıp açıklanamayacağı konusunda çalışan ilk araştırmacılardandır. 1885 yılında kalıtım konusunda çalışırken babalar ve oğulların boyları konusunda yaptığı araştırmasında, oğulların boylarının ortalamaya doğru yönlendiğini (İngilizce karşılığı "regressed") belirterek regresyon kelimesinin temelini oluşturmuştur. Günümüzde ise regresyona, değişkenler arasındaki ilişkinin modellenmesi işlemlerinin tümünü içeren geniş bir anlam yüklenmiştir.

Korelasyon ise birlikte değişim gösterme eğilimindeki değişkenler arasındaki ilişkinin derecesini temsil eden bir katsayıdır. Bu katsayı değişkenler arasındaki ilişkinin derecesinin yanı sıra ilişkinin yönünüde belirler. İki değişken arasında yüksek korelasyon olması bile, iki değişkenden birinin diğerinin nedeni olabileceğini göstermez. Korelasyon analizi iki değişken arasındaki nedensellik için kullanılamadığından, nedensellik araştırması için diğer farklı istatistik tekniklerinin kullanılması gerekir.

REGRESYON ANALİZİ

Regresyon analizinde iki veya daha fazla değişken arasındaki ilişkinin hangi matematiksel modelle ifade edileceği araştırılır. Sözü edilen iki değişken arasındaki ilişki genellikle neden-sonuç ilişkisi biçiminde ortaya çıkar: Gelir, Harcama; Yaş, Boy; Gübre miktarı, Verim miktarı; Toprak kalitesi, Verim miktarı; Çalışma süresi, Kıdem tazminatı ikilileri arasında neden-sonuç ilişkisi bulunmaktadır. Bu değişkenlerden sonuç durumundaki değişken bağımlı değişken neden durumundaki değişken ise bağımsız değişken olarak isimlendirilir. Regresyon analizinde bağımlı değişken üzerinde oluşan değişimlerin açıklanmasına çalışılır. Örneğin yukarıdaki ikililerin ilkinde, bir bireyin "harcaması" *bağımlı değişken* olarak ele alınmıştır. *Bağımsız değişken* ise neden durumunda olan "gelir" değişkenidir. Genellikle bağımlı değişken Y sembolüyle bağımsız değişken ise X'le ifade edilir.

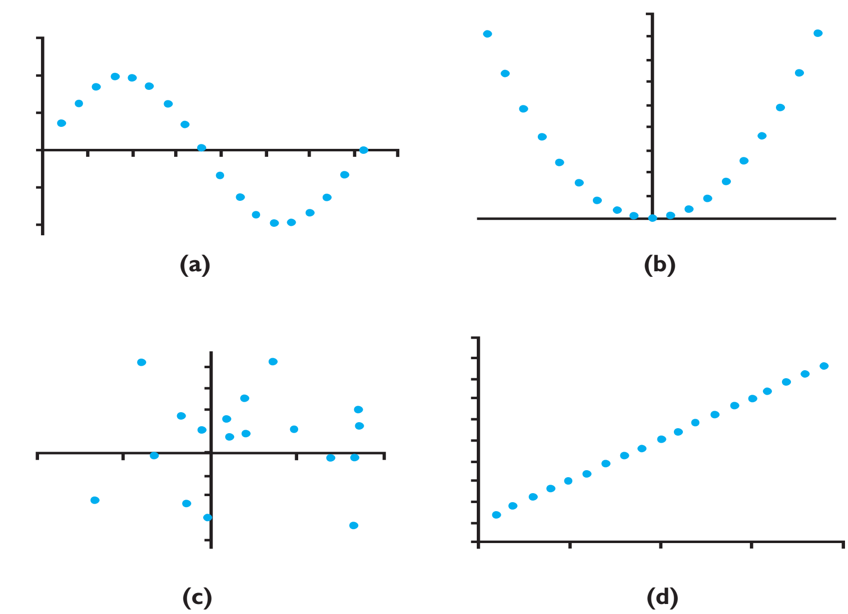
Değişkenler arasındaki ilişkiyi göstermenin ilk yolu grafik yöntemidir. X ve Y gözlem ikilileri bir grafik üzerinde birer nokta hâlinde gösterilir. İşaretlenen bu noktaların oluşturduğu şekil **serpilme diyagramı** olarak isimlendirilir. Farklı (X,Y) ikilileri için Şekil 5.1'dekine benzer görünümler elde edilebilir. Yatay eksen X; düşey eksen Y değişkeni olmak üzere (X,Y) koordinatlı noktalardan oluşan Şe-

X ve Y gözlem ikililerinin grafik üzerinde birer nokta hâlinde işaretlendiği şekil **serpilme diyagramı** olarak isimlendirilir.

kil 5.1'de verilen grafikler birer serpilme diyagramıdır. Şekil 5.1'de iki değişken arasında aldıkları değerlere göre gözlemlenebilecek 4 farklı serpilme diyagramı örneklenmiştir. Serpilme diyagramı üzerinde bağımlı ve bağımsız değişken arasındaki ilişkinin yapısı gözlenebilmektedir. Şekil 5.1(a) ve 5.1(b)'de iki değişken arasında doğrusal olmayan ilişki, 5.1(d)'de doğrusal ilişki görülmektedir. Şekil 5.1(c)'de ise iki değişken arasında ilişki olmadığı söylenebilir. Noktaların oluşturduğu şekle bakarak ilişkinin yönü ve derecesini tahmin etmek de mümkündür. Bunun için noktaların en dışta kalanları birleştirilerek bir şekil elde edilir. Söz konusu şeklin durumuna göre ilişkinin derecesi hakkında tahminde bulunulur. Eğer şekil dar bir elipse benziyorsa ilişki kuvvetlidir, elips genişledikçe ilişki zayıflar.

Şekil 5.1

Serpilme Diyagramları



Basit Doğrusal Regresyon

Regresyon analizinde bağımsız değişken sayısı bir olması durumunda *basit regresyon analizi;* birden fazla bağımsız değişken olması durumunda ise *çoklu regresyon analizi* söz konusu olmaktadır. Bu ünitede sadece basit doğrusal regresyon analizine yer verilecektir.

Serpilme diyagramı sadece ilişki tipi üzerinde bir fikir verebilmektedir. Bu ilişkinin matematiksel bir ifadesinin geliştirilmesi gerekir. Bu denklem elde edildiğinde, bağımlı değişken Y bağımsız değişken X olmak üzere, Y'nin değerlerinin tahmini için X'in değerleri kullanılır. X, Y değişkenleri arasındaki doğrusal ilişkinin ifadesinde kullanılan eşitliğe doğrusal regresyon modeli ya da kısaca regresyon denklemi denir.

Basit doğrusal regresyon modeli,

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad i=1,2,3,\ldots,N$$

şeklinde stokastik (olasılıklı) bir modeldir. Model'de

Y_i ; bağımlı değişkeninin i'inci gözlemlenen değerini,
X_i ; bağımsız değişkeninin i'inci gözlemlenen değerini,
ε_i ; hata terimi,
α : bilinmeyen sabit parametre,
β : bilinmeyen regresyon parametresidir.

α, doğrusal modelin sabit terimidir ve X=0 olduğunda regresyon doğrusunun dikey eksen Y'yi kestiği noktayı göstermektedir. β ise doğrusal modelin eğimini vermektedir ve regresyon analizinde, bağımsız değişken X'deki bir birimlik değişmenin, bağımlı değişken Y'de ne kadarlık bir değişmeye yol açacağını gösteren regresyon parametresidir.

$Y = \alpha + \beta X + \varepsilon$ şeklindeki

doğrusal model, X ve Y değişkenlerinin evrenleri için geçerlidir. İstatistiksel çalışmaların çoğunda olduğu gibi, regresyon analizinde de anakütleye ilişkin verilerin tümüne ulaşılamadığından bu evrenden seçilen örneklem verileriyle analiz yapılır. Örneklem verilerinden hareketle anakütle parametreleri α ve β nın tahminlerini elde edebilmek için EKK yönteminden yararlanılabilir. Bunun için öncelikle, gözlem ikililerini bir serpilme diyagramında gösterdiğimizi varsayalım. Serpilme diyagramı incelendiğinde doğrusal bir eğilim görülüyorsa, Y'nin X'e göre matematik fonksiyonunun doğrusal olduğuna (kesin olmasa da) karar verilebilir. Ancak gözlem noktaları arasından çok sayıda doğrusal fonksiyon geçirilebilir. Bu doğrusal fonksiyonlardan en uygunu Y_i gözlem değerlerine en yakın kuramsal (tahmin) $\hat{Y}_i$ değerini veren doğrusal fonksiyon olacaktır. Bir başka ifadeyle belirli bir X değeri için, elimizde iki ordinat değeri olacaktır; birincisi gözlem değeri, ikincisiyse bu noktanın doğru ya da eğri üzerinde teorik olarak hesaplanacak ordinat değeridir. İşte, $\hat{Y}_i$ kuramsal değerlerle Y_i gözlem değerleri arasındaki farklar hata terimlerini oluşturur.

$e_i = Y_i - \hat{Y}_i$ şeklinde hesaplanan hata terimleri "pozitif" ya da "negatif" ya da "sıfır" değerlerine sahip olurken, bu farkların cebirsel toplamı sıfıra eşittir.

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right) = 0$$

EKK Yönteminin esası α ve β'nın tahminleri olan a ve b'yi söz konusu farkların kareleri toplamını minimum yani

EKK Yöntemi $(Y_i-\hat{Y}_i)$ artıklarının kareleri toplamını en küçük yapabilen α ve β'nın tahminlerinin elde edilmesine ilişkin bir yöntemdir.

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \min$$

olacak şekilde belirlemektir.

α ve β'nın EKK tahminleri, yukarıdaki en küçük kareler yöntemi için,

$$\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n}\left(Y_i - a - bX_i\right)^2$$

ifadesinin a ve b'ye göre türevleri alınıp sıfıra eşitlenerek:

$$\frac{\partial e}{\partial a} = -2\sum_{i=1}^{n}\left(Y_i - a - bX_i\right) = 0$$

$$\frac{\partial e}{\partial b} = -2\sum_{i=1}^{n}\left(Y_i - a - bX_i\right) = 0$$

bulunur, elde edilen bu iki eşitlikten;

$$\sum Y_i = an + b\sum X_i$$

$$\sum X_i Y_i = a\sum X_i + b\sum X_i^2$$

denklemleri bulunur. Bu denklemler "normal denklemler" olarak isimlendirilir. Doğru denklemleri ve katsayılarının en küçük kareler koşuluna uygun olarak hesaplanması, bu iki denklemin çözümüyle gerçekleşebilir. Normal denklemlerdeki diğer değerler n, ΣX_i, ΣY_i, ΣX_i^2 ve $\Sigma X_i Y_i$ dır ve denklemlerdeki X_i ve Y_i değerleri "sıfır orijinine" göre ifade edilmişlerdir. Bu değerleri hesapladıktan sonra basit doğrusal regresyon modelindeki α ve β nın tahminleri olan a ve b'yi normal denklemlere dayanarak kolaylıkla çözmek mümkündür. a ve b gibi iki bilinmeyenli iki denklem sisteminde bu katsayılar hesaplandığında Y'nin X'e göre tahmin edilen doğrusal regresyon denklemi,

$$\hat{Y} = a + bX$$

şeklinde ifade edilir.

Bu denklemde araştırmacı regresyon denklemini kullanarak herhangi bir X değeri için $\hat{Y}$ tahmin değerini hesaplayabilir. Bu şekilde denklemde X yerine ilgilendiği değeri yazan araştırmacı Y'nin modele göre beklenen değerini hesaplamış olur. Tersine, Y bağımlı değişkeninin istenen belirli bir değerini üretecek bir X bağımsız değişkeni için beklenen değeri de hesaplanan Y yerine ilgilendiği değeri yazan araştırmacı, denklemi çözerek X sonucunu elde edecektir.

Tahmin edilen basit doğrusal regresyon modelindeki b katsayısının sıfırdan farklı olması Y ile X arasında bir bağıntı olduğunu göstermektedir. b katsayısının işareti ise X ile Y arasındaki ilişkinin yönünü belirler. Bu işaret pozitif ise iki değişken arasında aynı yönlü negatif ise ters yönlü bir ilişkiden söz edilir.

ÖRNEK 1

Bir araştırmacı bir mağaza için rastgele seçtiği 10 günde gözlem yapmış ve günler itibarıyla gelen günlük müşteri sayısı ve elde edilen ciro (₺1 000) değerlerini gözlemlemiştir. Bu veriler Tablo 5.1'de yer almaktadır.

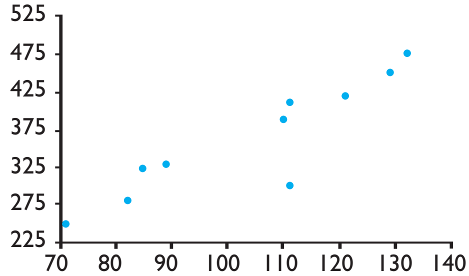
a. Serpilme diyagramını çiziniz?
b. Cironun, müşteri sayısına göre, basit doğrusal regresyon denkleminin tahminini, en küçük kareler tekniğiyle elde ediniz.

	Gün1	Gün2	Gün3	Gün4	Gün5	Gün6	Gün7	Gün8	Gün9	Gün10
Müşteri	71	82	111	85	89	110	111	121	129	132
Ciro	250	280	301	325	328	390	410	420	450	475

Tablo 5.1 *Regresyon Modeli İçin Gözlem Verisi Tablosu*

a.

Şekil 5.2



Ciroya ilişkin serpilme diyagramı

Kartezyen koordinat sisteminde X ve Y'ye ait verileri işaretlediğimizde serpilme diyagramını çizmiş oluruz. Grafik incelendiğinde iki değişken arasındaki ilişkinin doğrusal olduğunu görebiliriz.

b. Basit doğrusal regresyon denklemi $\hat{Y}$ = a + bX'i oluşturabilmek için a ve b katsayılarını normal denklemler yardımıyla hesaplayalım. Normal denklemler;

$$\sum Y_i = an + b\sum X_i$$

$$\sum X_i Y_i = a\sum X_i + b\sum X_i^2$$

dir. Bu amaçla; $\sum X_i\ \sum Y_{i,}\ \sum X_i Y_{i,}\ \sum X_i^2$, değerleri hesaplanmış ve Tablo 5.2 oluşturulmuştur. Tablonun ilk 2 sütunu günlük müşteri sayısı (X) ve günlük ciro miktarı (Y) değişkenlerinin gözlem değerlerini göstermektedir. Tablo 5.2'nin sütunlarının altında yer alan sayılar ise ilgili sütunun değerleri toplamını göstermektedir.

X	Y	XY	X^2
71	250	17750	5041
82	280	22960	6724
111	301	33411	12321
85	325	27625	7225
89	328	29192	7921
110	390	42900	12100
111	410	45510	12321
121	420	50820	14641
129	450	58050	16641
132	475	62700	17424
$\sum$X=1041	**$\sum$Y=3629**	**$\sum$XY=390918**	**$\sum X^2$ =112359**

Tablo 5.2 *Regresyon Denklemi Katsayıları Hesaplama Tablosu*

$3629 = 10a + 1041b$

$390918 = 1041a + 112359b$

Bu iki denklemde yer alan a ve b katsayıları;

a = 20,175; b = 3,29

olarak hesaplanır. Buna göre tahmin edilen regresyon doğrusu denklemi;

$\hat{Y} = 20{,}175 + 3{,}29X$

şeklinde elde edilir.

DİKKAT

Ünite boyunca hesaplamalarda hassasiyet sağlamak amacıyla olabildiğince çok ondalıklı hane kullanılırken, hesaplama sonuçlarının gösteriminde yuvarlatılmış daha az sayıda ondalık haneyle yetinilmiştir.

Regresyon katsayısı 3,29 bulunduğundan, müşteri sayısı bir birim değiştiğinde ciroda 3,29 birimlik değişme gerçekleşecektir. Regresyon denklemi yardımıyla araştırmacı gözlemlediği herhangi bir X değeri için Y'nin alacağı değeri tahmin edebilir. Örneğin, mağazayı 100 kişi ziyaret ettiğinde işletmenin tahmini cirosunun bilinmesi istenebilir. Bu durumda X=100 olarak gözlemlenmiş olduğundan Y'nin tahmini, $\hat{Y} = 20{,}175 + 3{,}29X = 20{,}175 + 3{,}29 \times 100 = 349{,}402$ olarak hesaplanır. Benzer biçimde, regresyon denkleminden Y'nin belirli bir değerini veren X değeri de tahmin edilebilir. 600'lük bir ciro elde etmek için kaç müşterinin mağazaya gelmesine ihtiyaç olduğu, Y=600 için X'in modele göre beklenen değeri $600 = 20{,}175 + 3{,}29X$ denkleminden $X = 176{,}117 \cong 177$ olarak bulunur.

Normal denklemlerde X ve Y değerleri yerine bunların aritmetik ortalamalarından sapmaları olan x ve y değerlerinin konulmasıyla;

$$\sum y_i = an + b\sum x_i$$

$$\sum x_i y_i = a\sum x_i + b\sum x_i^2$$

denklemleri elde edilir. Tanım gereğince $x_i = X - \bar{X}$ ve $y_i = Y - \bar{Y}$ olduğundan aritmetik ortalamanın temel özelliklerine göre (arimetik ortalamadan cebirsel sapmaların toplamı sıfırdır.) $\sum x_i = 0$ ve $\sum y_i = 0$'dır. Böylece son iki eşitlikten

$$b = \frac{\sum x_i y_i}{\sum x_i^2}$$

ya da

$$b_{yx} = \frac{\sum x_i y_i}{\sum x_i^2}$$

elde edilir. a'nın 0'a eşit olması, ortalamalar orijinine göre, regresyon doğrusunun, değişkenlerin ortalamalarıyla tanımlanan bir noktadan geçtiğini ortaya koymaktadır. Bu durumda regresyon denklemi

$\hat{y} = bx_i$ veya $\hat{y} = b_{yx}x_i$

şeklinde yazılabilir.

Bir regresyon doğrusu ister sıfır orijinine göre yazılsın ister ortalamalar orijinine göre yazılsın eğimi değişmez. Bu nedenle her iki orijine göre hesaplanan b katsayısı aynıdır. Buna karşılık a ise;

$$a = \bar{Y} - b\bar{X}$$

eşitliğinden hesaplanır. Veriler için en iyi doğru denklemi

$$\hat{Y} = a + bX$$

şeklinde yazılabilir.

ÖRNEK 2

Örnek 1'deki veriler için Y'nin X'e göre regresyon denklemini, ortalamalar orijinine göre, en küçük kareler tekniğiyle elde edelim.

Basit doğrusal regresyon denklemini ortalamalar orijinine göre yazabilmek için, ilk olarak X ve Y serisinin aritmetik ortalamalarını ve x, y, xy ve x^2'leri elde edelim.

$$\bar{X} = \frac{\sum X_i}{n} = \frac{1041}{10} = 104,1 \qquad \bar{Y} = \frac{\sum Y_i}{n} = \frac{3629}{10} = 362,9$$

x_i	y_i	x_iy_i	x_i^2
-33,1	-112,9	3736,99	1095,61
-22,1	-82,9	1832,09	488,41
6,9	-61,9	-427,11	47,61
-19,1	-37,9	723,89	364,81
-15,1	-34,9	526,99	228,01
5,9	27,1	159,89	34,81
6,9	47,1	324,99	47,61
16,9	57,1	964,99	285,61
24,9	87,1	2168,79	620,01
27,9	112,1	3127,59	778,41
$\sum xi=0$	**$\sum yi=0$**	**$\sum xiyi=13139,1$**	**$\sum x_i^2=3990,9$**

$$a = 0$$

$$b = b_{yx} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{13139,1}{3990,9} = 3,29$$

Bu durumda, ortalamalar orijinine göre, Y'nin X'e göre regresyon denklemi,

$$\hat{y} = 3,29x$$

şeklinde yazılır.

Yukarıdaki açıklamalarda Y değişkeni bağımlı değişken, X değişkeniyse bağımsız değişken kabul edilmişti. X değişkeni bağımlı Y değişkeni bağımsız değişken olduğundaysa doğrusal regresyon denklemi;

$$\hat{X} = a + bY$$

olarak ifade edilecektir. Bu durumda a ve b

$$b = \frac{\sum x_i y_i}{\sum y_i^2}$$

ya da

$$b_{xy} = \frac{\sum x_i y_i}{\sum y_i^2}$$

ve

$$a = \overline{X} - b\overline{Y}$$

formülleri yardımıyla bulunur. b_{xy} ve b_{yx}'in formülleri incelendiğinde, her ikiside daima aynı işareti taşır fakat aynı değerde değildir.

SIRA SİZDE 1

X bağımsız ve Y bağımlı değişken olmak üzere bir araştırmanın gözlemlenen değerleri izleyen tabloda verilmiştir. Verilere ilişkin serpilme diyagramını çiziniz. Serpilme diyagramı üzerindeki noktaların doğrusal eğilim içinde olup olmadıklarını gözleyiniz. b katsayısını ve a katsayısını hesaplayıp regresyon denklemini elde ediniz.

Gözlem	1	2	3	4	5	6	7	8	9	10	11	12
X	12	22	36	34	23	18	18	27	30	38	24	18
Y	10	12	14	16	18	11	17	19	12	18	14	19

Varyansın (σ^2) Tahmini

Basit doğrusal regresyon modelinde α ve β'nın tahminlerine ek olarak, aralık tahminlerinde ve hipotez testlerinde σ^2'nın tahminine gereksinim vardır.

σ^2, ε_i hata terimlerinin ortak varyansıdır. ε_i'in tahmini e_i hata terimi olduğundan e_i'lerin varyansı da σ^2'in bir tahmini olacaktır. Hataların kareler toplamı (HKT),

$$HKT = \sum e_i^2$$

$$HKT = \sum (Y_i - \hat{Y}_i)^2$$

yazılabilir. HKT'in serbestlik derecesine bölümüyle elde edilen

$$HKO = \frac{HKT}{n - k}$$

hata kareler ortalaması, (bir başka ifadeyle hataların varyansı), σ^2'in bir tahminidir. Basit doğrusal regresyon modeliyle hataların hesaplanmasında, α ve β'nın tahminleri a ve b kullanıldığından, serbestlik derecesi (n-2) olarak yazılır.

HKO'un kare kökü alındığındaysa denklemin standart hatası elde edilir ve $\hat{\sigma}$ ile gösterilir.

$$\hat{\sigma} = \sqrt{\text{HKO}}$$

$$\hat{\sigma} = \sqrt{\frac{\sum (Y_i - \hat{Y}_1)^2}{n-2}}$$

ÖRNEK 3

Örnek 1'deki verileri kullanarak elde edilen regresyon denklemine dayanarak yapılacak tahminlerin standart hatasını hesaplayalım.

$\hat{\sigma}$'ı hesaplayabilmek için öncelikle, $\hat{Y}$ 'ları daha sonra da sırasıyla $(Y_i - \hat{Y}_i)$, $(Y_i - \hat{Y}_i)^2$ ları hesaplayalım.

$\hat{Y} = 20{,}175 + 3{,}29x$

$\hat{Y}_i$	$(Y_i - \hat{Y}_i)$	$(Y_i - \hat{Y}_i)^2$
253,93	-3,93	15,41
290,14	-10,14	102,83
385,62	-84,62	7159,94
300,02	24,98	624,12
313,19	14,81	219,44
382,32	7,68	58,92
385,62	24,38	594,56
418,54	1,46	2,13
444,88	5,12	26,24
454,75	20,25	409,90
		9213,5

buradan,

$$\hat{\sigma} = \sqrt{\frac{\sum (Y_i - \hat{Y}_1)^2}{n-2}}$$

$$\hat{\sigma} = \sqrt{\frac{9213{,}5}{10-2}} = \sqrt{1151{,}69} = 33{,}94$$

olarak elde edilir.

Basit Doğrusal Regresyonda Aralık Tahmini

İstatistiksel çıkarsamalarda yapılan tahminlerin, gerçek değerlerle genellenmesi aralık tahminleriyle yapılır. Regresyon çözümlemesi, örneklem verileriyle yapıldığından, elde edilen a ve b'lerin anakütle parametreleri α ve β'ya ilişkin aralık tahminlerinde kullanılması söz konusudur.

α için güven aralığı;

$$P(a - t_\alpha . s_a \leq \alpha \leq a + t_\alpha . s_a) = 1 - \alpha$$

şeklinde verilir. s_a, a'nın standart hatasıdır ve

$$s_a = \hat{\sigma}\sqrt{\frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2}}$$

formülüyle hesaplanır, t_α ise α anlamlılık düzeyi n-2 serbestlik derecesinde t tablosundan bulunan değerdir.

β için güven aralığı;

$$P(b - t_\alpha . s_b \leq \beta \leq a + t_\alpha . s_b) = 1 - \alpha$$

$$s_b = \frac{\hat{\sigma}}{\sqrt{\sum (X_i - \overline{X})^2}}$$

şeklinde hesaplanır.

ÖRNEK 4

Örnek 1'de b=3,29 olarak hesaplanmıştır. Bu sonuca göre, β katsayısının %95 güven aralığını hesaplayalım.

b= 3,29

$$s_b = \frac{\hat{\sigma}}{\sqrt{\sum (X_i - \overline{X})^2}} = \frac{33,94}{\sqrt{\sum 3990,9}} = \frac{33,94}{63,17} = 0,54$$

ve ν=8 serbestlik derecesinde t tablo değeri 2,306'dır. Buna göre β için %95 güven aralığı

$$3,29 \pm (2,306)\,(0,54)$$
$$P(3,29 - 1,25 \leq \beta \leq 3,29 + 1,25) = 0,95$$
$$P(2,04 \leq \beta \leq 4,54) = 0,95$$

olarak hesaplanır.

Regresyon katsayısı β'nın 0,95 olasılıkla alabileceği değerler 2,04 ile 4,54 arasında olacaktır.

Regresyon Katsayısının Anlamlılık Testi

Basit doğrusal regresyon analizinde, bir bağımlı bir bağımsız değişken olması nedeniyle, test edilecek parametreler α ve β olacaktır. Daha önce açıklandığı gibi, α'nın tahmini a regresyon sabitidir ve b ise β'nın tahmini olup regresyon katsayısıdır.

Basit doğrusal regresyon modelindeki regresyon katsayısına ilişkin yapılan test, regresyon doğrusunun anlamlılığını da test etmektedir. Şöyle ki:

$H_o : \beta = 0$
$H_1 : \beta \neq 0$

hipotezleri

$$t_h = \frac{b}{s_b}$$

istatistiğinden yararlanılarak test edilir. α anlam düzeyinde n-2 serbestlik derecesinde t tablosundan bulunan değer hesaplanan t test istatistiğinden büyükse H_o: β=0 hipotezi kabul edilir ve regresyon doğrusu anlamlı değildir. Diğer bir ifadeyle Y'deki değişimler X'deki değişmelerden kaynaklanmamaktadır (X ve Y arasında doğrusl bir ilişki yoktur). t istatistiğinin değeri t tablo değerinden büyükse H_o reddedilir, yani regresyon doğrusu anlamlıdır. Elde edilen doğrusal regresyon modeli amaca uygun olarak kullanılabilir.

ÖRNEK 5

Örnek 1'de elde edilen regresyon katsayısının α=0,05 anlam düzeyinde anlamlılık testini yapalım.

H_o: β=0
H_1: $\beta \neq 0$

Hipotezlerini test edelim $t_h = \frac{b}{s_b}$ örneklem küçük olduğu için benimsenen t, ν= n-2 = 8 serbestlik derecesinde t dağılır.

$$t_h = \frac{b}{s_b} = \frac{3,29}{0,54} = 6,09$$

α=0,05 ve 8 serbestlik derecesi ile $t_{0,05}$= 2,306 olduğundan $t_h > t_{0,05}$'dir ve H_0 reddedilecektir, b katsayısı istatistiksel olarak anlamlıdır.

KORELASYON ANALİZİ

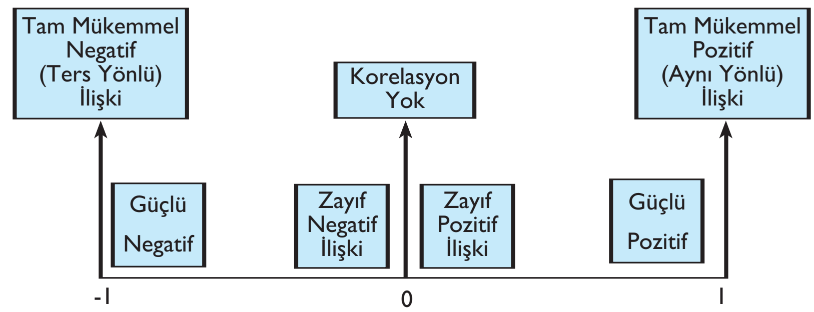
Şu ana kadar çalışmalarımız regresyon analizi başlığı altında biri açıklayıcı (bağımsız) diğeri açıklanan (bağımlı) olmak üzere iki değişken arasındaki doğrusal ilişkiye ait denklem üzerinde yoğunlaşmıştır. Bundan sonraki kesimde, iki değişken arasındaki doğrusal ilişkinin derecesi incelenecektir.

Basit Doğrusal Korelasyon Katsayısı

İki nicel değişken arasındaki ilişkinin derecesini ve yönünü belirlemek için *r* ile gösterilen *Pearson korelasyon katsayısı* veya *basit korelasyon katsayısı* hesaplanır. Bu katsayı, $-1 \leq r \leq 1$ aralığında değer alır. Pozitif korelasyon katsayısı değişkenlerden birinin değeri arttığında diğerinin de değerinin arttığını; negatif korelasyon katsayısı ise değişkenlerden birinin değeri artarken diğerinin değerinin azaldığını belirtir. $r = \pm 1$ olduğunda, söz konusu iki değişken mükemmel/tam ilişki içindedir. Buna karşılık $r = 0$ olması iki değişkenin hiçbir ilişki içinde olmadıklarını gösterir. r -0,50 ya da +0,50 etrafında bir değer aldığında ise değişkenler arasında orta düzeyli bir ilişkinin varlığından söz edilir. Korelasyon en az iki değişken için tanımlansa da, bir değişkenin kendisi ile korelasyonu +1'dir. Korelasyon katsayısı r'nin belirttiği ilişkinin derece ve yönü Şekil 5.3'te özetlenmektedir.

Şekil 5.3

Pearson Korelasyon Katsayısının İlişki Yön ve Dereceleri



Basit doğrusal korelasyon katsayısı iki değişken arasındaki doğrusal ilişkinin derecesini ve yönünü belirler.

X ve Y gibi iki değişken arasındaki **basit doğrusal korelasyon katsayısı** farklı formüller yardımıyla hesaplanabilir. Bunlardan birincisi S_x, S_y sırasıyla *X* ve *Y* değişkenlerinin standart sapması olmak üzere, *X* ve *Y* değişkenleri arasındaki. Pearson

korelasyon katsayısının $r = \dfrac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)s_x s_y}$ olarak hesaplanmasıdır. İkinci olarak doğrudan iki değişkene ilişkin verilerden hareketle $r = \dfrac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$

formülüyle hesaplanabilir.

Korelasyon katsayısı hesaplarken yararlanabileceğimiz bir diğer formül ise şu şekildedir:

$$r = \mp\sqrt{b_{yx} b_{xy}}$$

Formüldeki;

b_{yx} = Y'nin X'e göre regresyon katsayısı

b_{xy} = X'in Y'ye göre regresyon katsayısıdır.

Burada dikkat edilmesi gereken nokta regresyon katsayıları pozitifse korelasyon katsayısı pozitif, eğer her iki regresyon katsayısının işareti negatifse korelasyon katsayısı da negatif olacaktır.

ÖRNEK 6

Örnek 1'deki X ve Y değişkenlerine ilişkin Pearson korelasyon katsayısını standart sapma değerlerini kullanarak bulunuz.

X ve Y değişkenlerinin standart sapmalarını kullanarak korelasyon katsayısının hesaplanabilmesi için izleyen tabloda verildiği gibi sırasıyla $(X_i - \bar{X})$, $(Y_i - \bar{Y})$ ve $(X_i - \bar{X})(Y_i - \bar{Y})$'nın hesaplanması gerekir.

Müşteri	Ciro
X_i	Y_i
71	250
82	280
111	301
85	325
89	328
110	390
111	410
121	420
129	450
132	475
21,058	76,3551

$X_i-\bar{X}$	$Y_i-\bar{Y}$	$(X_i-\bar{X})(Y_i-\bar{Y})$
-33,1	-112,9	3736,99
-22,1	-82,9	1832,09
6,9	-61,9	-427,11
-19,1	-37,9	723,89
-15,1	-34,9	526,99
5,9	27,1	159,89
6,9	47,1	324,99
16,9	57,1	964,99
24,9	87,1	2168,79
27,9	112,1	3127,59
		13139,10

$n = 10$,

$\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y}) = 13139,10$ olarak hesaplanmıştır. Korelasyon katsayısı için gerekli olan X ve Y değişkenlerinin standart sapmaları $s_x = 21,058$ ve $s_y = 76,355$'dır. Bu değerler formülde yerine konularak bulunur. Hesaplanan bu korelasyon katsayısına göre; iki değişken arasında güçlü aynı yönlü (pozitif) ilişki olduğu söylenebilir.

SIRA SİZDE 2

Örnek 1'deki X ve Y değişkenlerine ilişkin Pearson korelasyon katsayısını aşağıdaki eşitliği kullanarak hesaplayınız.

$$r = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2\sum_{i=1}^{n}(Y_i-\bar{Y})^2}}$$

Belirlilik Katsayısı

Belirlilik katsayısı, Bağımlı değişkende meydana gelen değişmelerin ne kadarının bağımsız değişkendeki değişmelerle açıklanabileceğini gösterir.

Değişkenler arasındaki ilişkinin derecesi korelasyon katsayısı yardımıyla belirlenirken bu katsayının değerine bağlı olarak zayıf, orta ve güçlü ilişki gibi nitelemeler söz konusu olur. Bağımlı değişkende meydana gelen değişmelerin ne kadarının bağımsız değişkendeki değişmelerle açıklanabileceğini belirlemek amacıyla **belirlilik katsayısı** kulanılır. Belirlilik katsayısı Pearson korelasyon katsayısı r'nin karesinin alınması ile hesaplanır ve 0 ile 1 arasında değerler alır, r^2 sembolü ile gösterilir.

$0 \leq r^2 \leq 1$

$r^2 = 1$ ise Y'deki değişimin %100'ünün X bağımsız değişkeni tarafından açıklanabildiği kabul edilir.

$r^2 = 0$ ise X bağımsız değişkeni, Y bağımlı değişkenini hiç açıklayamıyor demektir.

ÖRNEK 7 *Örnek 1'de verilen günlük elde edilen ciro değişkenin ne kadarlık bir bölümünün günlük müşteri sayısı ile açıklanabileceğini bulunuz.*

Bu sorunun cevaplanması için korelasyon katsayı değerine ihtiyaç bulunmaktadır. Bu değerde önceki örnekten r=0,908 olarak bulunmuştu. Bu sayının karesi belirlilik katsayısı olacağından $(0{,}908)^2=0{,}82$ olarak hesaplanır. Söz konusu mağazanın günlük cirosuna, günlük müşteri sayısının etkisinin %82 olduğu söylenir. Geriye kalan %18 ise müşteri sayısına bağlı olamayan kısımdır.

Korelasyon Katsayısı Anlamlılık Testi

Evren korelasyon katsayısı ρ (ro) ile sembolize edilir. Evren korelasyon katsayısı ρ=0 olan bir ana kütleden seçilen örneklemlerin r katsayıları normal dağılıma sahiptir. Bu nedenle,

H_0: ρ= 0

H_1: ρ ≠ 0

hipotezleri test edilir.

Hipotezlerin kurulmasının ardından örneklem korelasyon katsayısı yardımıyla $t=\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ istatistiği hesaplanır.

Daha sonra belirlenen anlam düzeyi ve n-2 serbestlik derecesine göre t tablosundan kritik değer belirlenir. Belirlenen bu kritik değer yardımıyla (-kritik değer, +kritik değer) aralığı tanımlanır. Hesaplanan t değeri, söz konusu bu aralıkta yer aldığında H_0 hipotezi kabul, bu aralıkta yer almadığında ise reddedilir. H_0'ın reddedilmesinin anlamı hesaplanmış korelasyon katsayısının istatistiksel olarak anlamlı olduğudur.

ÖRNEK 8 *Örnek 1'de verilen probleme ilişkin olarak ana kütle korelasyon katsayısının değeri 0'a eşit midir? %5 anlam düzeyine göre test ediniz.*

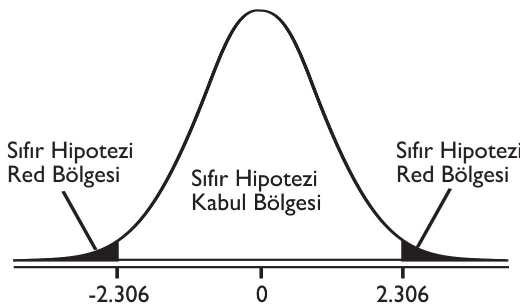
İlk olarak

H_0: ρ = 0 (Evren korelasyonu 0'dır.)

H_1: ρ ≠ 0 (Evren korelasyonu 0 değildir.)

hipotezleri yazılır.

Daha sonra tablodan gerekli kritik değerler elde edilir. t dağılımı tablosu yardımıyla %5 anlam düzeyi ve n-2=10-2=8 serbestlik derecesi için elde edilen kritik değerler ile belirlenen H_0 hipotezi red bölgeleri Şekil 5.4'teki gibidir.



Şekil 5.4

%5 Anlam Düzeyi ve 8 Serbestlik Derecesi ile Korelasyon Katsayısı Anlamlılık Testinde H_0'ın Kabul ve Red Bölgeleri

Sonunda test için gerekli olan gözlemlenen t istatistiği değeri hesaplanır.

Bu değer $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \frac{0,908\sqrt{8}}{\sqrt{1-0,908^2}} = 6,129$ olarak bulunur. Hesaplanan bu 6,129 değeri Şekil 5.4'te gösterilen kritik değerler aralığında, yani H_0'ın kabul bölgesinde yer almadığından "İlgilenilen değişkenler arasındaki korelasyon sıfırdır." hipotezi reddedilir. Değişkenler arasında ana kütle düzeyinde de ilişki bulunmaktadır.

SIRA SİZDE 3

10 kişilik bir sınıftaki öğrencilerin istatistik ve matematik derslerindeki başarı puanları tabloda verilmiştir. Bu sınıftaki öğrencilerin istatistik ve matematik derslerindeki başarı seviyeleri arasında bir ilişki olabileceği düşünülmektedir. Bu ilişkinin belirlenmesi için, öğrencilerin izleyen tabloda verilen istatistik ve matematik sınavları başarı puanları incelenecektir:

Öğrenci	1	2	3	4	5	6	7	8	9	10
İstatistik Başarı Puanı	75	80	95	60	40	70	65	25	90	85
Matematik Başarı Puanı	85	80	90	50	40	55	50	20	95	80

a. İstatistik ve matematik dersleri başarı puanlarıyla ilgili korelasyon katsayısını hesaplayınız.
b. İstatistik dersi başarı puanının ne kadarlık kısmının matematik dersi başarı puanı tarafından açıklanabileceğini hesaplayınız.
c. Korelasyon katsayısının istatistiksel olarak anlamlı olup olmadığını %5 anlamlılık düzeyine göre test ediniz.

Özet

AMAÇ 1

İki değişken arasındaki ilişkiyi açıklayan doğrusal modeli kurmak

Regresyon analizi ilişki içinde bulunan değişkenler arasındaki ilişkinin doğasını belirlemek ve bu ilişkiyi kullanarak o konu ile ilgili tahminler (estimation) ya da kestirimler (prediction) yapabilmek amacıyla kullanılan istatistiksel bir tekniktir. Regresyon analizinde bağımlı değişken üzerinde oluşan değişimlerin açıklanmasına çalışılır. Regresyon analizinin yalnızca bir açıklayıcı değişkenle yapılması düşünülemez. Gerçekten de regresyon analizi birden fazla bağımsız değişken üzerinde de yapılabilir. Bir bağımsız değişken olması durumunda *basit doğrusal regresyon analizi;* birden fazla bağımsız değişken olması durumunda *çoklu doğrusal regresyon analizi* söz konusu olmaktadır. X, Y değişkenleri arasındaki doğrusal ilişkinin ifadesinde kullanılan eşitliğe doğrusal regresyon modeli ya da kısaca regresyon denklemi denir. Doğrusal regresyon denkleminin tahmini için kullanılabilen farklı teknikler arasında en çok bilineni en küçük kareler tekniğidir. Basit doğrusal regresyon modeli,

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

olarak yazılır. En küçük kareler tekniğinde, basit doğrusal regresyon modelinde yer alan ana kütle parametreleri α ve β için eldeki veri yardımıyla

$$a = \frac{\sum Y_i - b\sum X_i}{n} \text{ ve } b = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - \left(\sum X_i\right)^2}$$

tahminleri hesaplanır. Böylelikle bulunan a ve b parametre tahminleri doğru denkleminde yerlerine konarak basit doğrusal regresyon denklemi

$\hat{Y} = a + bX$ olarak yazılır.

AMAÇ 2

İki değişken arasındaki ilişkinin derecesini belirlemek

İki ya da daha fazla ve en az aralıklı ölçeğe uygun şekilde ölçümlenmiş değişkenler arasındaki ilişkinin derecesini belirlemek için r ile gösterilen Pearson korelasyon katsayısı ile hesaplanır. Korelasyon katsayısı, $-1 \leq r \leq 1$ aralığında değer alır. Pozitif korelasyon katsayısı değişkenlerden birinin değeri arttığında diğerinin de değerinin arttığını; negatif korelasyon katsayısı ise değişkenlerden birinin değeri artarken diğerinin değerinin azaldığını belirtir. $r = \pm 1$ olduğunda, söz konusu iki değişken mükemmel/tam ilişki içindedir. Buna karşılık $r = 0$ olması iki değişkenin hiçbir ilişki içinde olmadıklarını gösterir.

Kendimizi Sınayalım

1. Aşağıdakilerden hangisi değişkenler arasındaki ilişkinin yapısı hakkında genel bir bakışa imkan sağlar?

a. Histogram
b. Serpilme diyagramı
c. Frekans poligonu
d. Normal eğri
e. Pasta grafik

2. Hangi yaklaşımda bağımlı ve bağımsız değişkenler arasındaki doğrusal ilişkinin matematiksel fonksiyonu ortaya konmaya çalışılır?

a. Hipotez testi
b. Aralık Tahmini
c. Regresyon analizi
d. Korelasyon analizi
e. Varyans analizi

3. Bağımlı değişkenle (Y_i) gözlem değerleriyle regresyon denkleminden hesaplanan ($\hat{Y}_i$) değeri arasındaki farka ne ad verilir?

a. Hata Terimi
b. Varyans
c. Korelasyon
d. Katsayı
e. Parametre

4. $\hat{Y} = 2 + 5X$ Regresyon denkleminde (b) regresyon katsayısı aşağıdakilerden hangisinde doğru olarak ifade edilmiştir?

a. X'deki bir birimlik değişim Y'de 5 birimlik değişime neden olur.
b. X'deki bir birimlik artış Y'de 1 birimlik artışa neden olur.
c. X'de 1 birimlik azalma Y'de 1 birimlik azalışa neden olur.
d. X'de bir birimlik değişim Y'de 2 birimlik değişime neden olur.
e. X'in değerindeki değişimin Y'nin değerine etkisi yoktur.

Aşağıdaki tabloyu inceleyiniz. 5-6 ve 7. Sorular aşağıdaki bilgilere göre cevaplandırılacaktır.

X_i	Y_i	$\hat{Y}_i = a + bX_i$
3	12	12
5	16	16
11	28	28
7	20	20
1	8	8

5. Yukarıdaki tablodaki verilerden oluşturulan regresyon modelinde a = 6 olduğunda, regresyon doğrusunun eğimi aşağıdakilerden hangisidir?

a. -1
b. 0
c. 1
d. 2
e. 4

6. Tablo verilerine göre HKT değeri nedir?

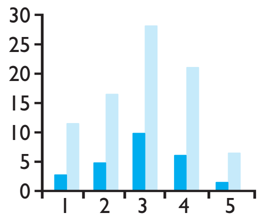
a. 4,6
b. 16,8
c. 0
d. -16,8
e. -4,6

7. Tablo verilerine göre tahminlerin standart hatası(σ) nedir?

a. 3,2
b. 6,4
c. 3
d. 6
e. 0

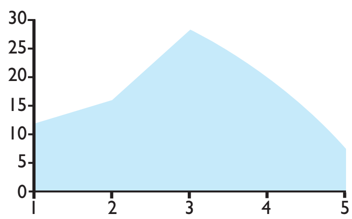
8. Aşağıdakilerden hangisi serpilme diyagramı örneğidir?

a.

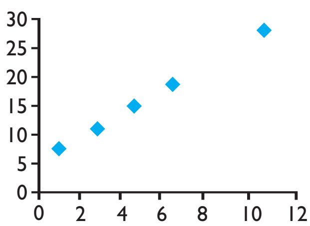


b.

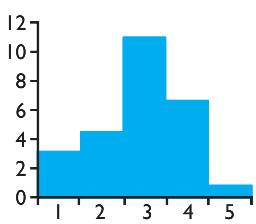
c.



d.



e.



9. Bir araştırmada X ve Y değişkenleri için gözlem değerleri aşağıdaki tabloda verilmiştir.

X	Y
1	2
3	4
4	6
7	8
9	10
12	12
45	14
50	16
80	18
50	20

X ve Y değişkenleri arasındaki korelasyon katsayısı nedir?

a. -0,92
b. -0,45
c. 0,02
d. 0,55
e. 0,88

10. 9. Sorudaki gözlem verilerinden hareketle hesaplanacak belirlilik katsayısı nedir?

a. 0,81
b. 0,21
c. 0.12
d. 0,42
e. 0,77

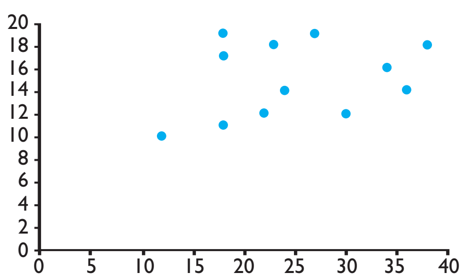
Kendimizi Sınayalım Yanıt Anahtarı

1. b Yanıtınız yanlış ise "Regresyon Analizi" konusunu yeniden gözden geçiriniz.
2. c Yanıtınız yanlış ise "Regresyon Analizi" konusunu yeniden gözden geçiriniz.
3. a Yanıtınız yanlış ise "Basit Doğrusal Regresyon Analizi" konusunu yeniden gözden geçiriniz.
4. a Yanıtınız yanlış ise "Basit Doğrusal Regresyon Analizi" konusunu yeniden gözden geçiriniz.
5. d Yanıtınız yanlış ise "Basit Doğrusal Regresyon Analizi" konusunu yeniden gözden geçiriniz.
6. c Yanıtınız yanlış ise "Varyansın (σ_2) Tahmini" konusunu yeniden gözden geçiriniz.
7. e Yanıtınız yanlış ise "Basit Doğrusal Regresyon Analizi" konusunu yeniden gözden geçiriniz.
8. d Yanıtınız yanlış ise "Basit Doğrusal Regresyon Analizi" konusunu yeniden gözden geçiriniz.
9. e Yanıtınız yanlış ise "Korelasyon Analizi" konusunu yeniden gözden geçiriniz.
10. e Yanıtınız yanlış ise "Korelasyon Analizi" konusunu yeniden gözden geçiriniz.

Sıra Sizde Yanıt Anahtarı

Sıra Sizde 1

Verilere ilişkin serpilme diyagramı çizildiğinde aşağıda verilen grafik elde edilir.



Serpilme diyagramındaki noktaların doğrusal eğilim içinde oldukları gözlenmektedir. Önerilen eşitlik yardımıyla b katsayısının hesaplanabilmesi için öncelikle x ve y değişkenlerinin ortalamaları bulunmalıdır.

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{12+22+...+18}{12} = \frac{300}{12} = 25$$

$$\bar{Y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{10+12+...+19}{12} = \frac{180}{12} = 15$$

Ardından ara hesaplamalar için gereken aşağıdaki tablo oluşturulur. Tablonun altındaki sayıların ilgili sütunların toplamları olduğu unutulmamalıdır. Tablodan elde edilen değerlerle

$$b = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{85}{730} = 0{,}116$$

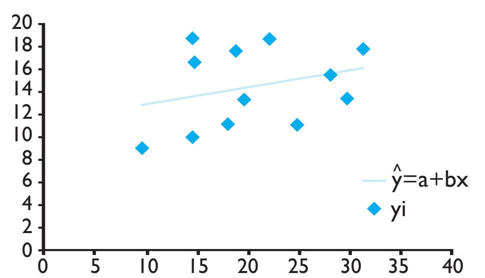
olarak bulunur. Daha sonra bu b değeri kullanılarak

$$a = \frac{\sum y_i - b\sum x_i}{n} = \frac{180-(0{,}116).(300)}{12} = 12{,}089$$

olarak elde edilir.

X_i	Y_i	$X_i-\bar{X}$	$Y_i-\bar{Y}$	$(X_i-\bar{X})(Y_i-\bar{Y})$	$(X_i-\bar{X})^2$
12	10	-13	-170	2210	169
22	12	-3	-168	504	9
36	14	11	-166	-1826	121
34	16	9	-164	-1476	81
23	18	-2	-162	324	4
18	11	-7	-169	1183	49
18	17	-7	-163	1141	49
27	19	2	-161	-322	4
30	12	5	-168	-840	25
38	18	13	-162	-2106	169
24	14	-1	-166	166	1
18	19	-7	-161	1127	49
300	180			85	730

$\hat{Y} = a + bX = 12{,}089 + 0{,}116X$ biçiminde regresyon denklemi elde edilir. Bu doğru serpilme diyagramı üzerinde gösterildiğinde ise izleyen grafik elde edilir.



Sıra Sizde 2

Eşitliğin sağ tarafındaki değerlerin elde edilmesi amacıyla izleyen tablo oluşturulur.

X_i	Y_i	$X_i-\bar{X}$	$Y_i-\bar{Y}$	$(X_i-\bar{X})(Y_i-\bar{Y})$	$(X_i-\bar{X})$	$(Y_i-\bar{Y})^2$
71	250	-33,1	-112,9	3736,99	1095,61	12746,41
82	280	-22,1	-82,9	1832,09	488,41	6872,41
111	301	6,9	-61,9	-427,11	47,61	3831,61
85	325	-19,1	-37,9	723,89	364,81	1436,41
89	328	-15,1	-34,9	526,99	228,01	1218,01
110	390	5,9	27,1	159,89	34,81	734,41
111	410	6,9	47,1	324,99	47,61	2218,41
121	420	16,9	57,1	964,99	285,61	3260,41
129	450	24,9	87,1	2168,79	620,01	7586,41
132	475	27,9	112,1	3127,59	778,41	12566,41
				13139,10	3990,90	52470,90

İlgili değerler eşitlikte yerlerine konarak

$$r = \frac{13139,10}{\sqrt{(3990,90).(52470,90)}} = 0,908$$

olarak bulunur.

Sıra Sizde 3

a. Pearson korelasyon katsayısının hesaplanması için kullanılacak

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

formülüne uygun olarak, gerekli ara değerleri içeren bir tablo oluşturulur.

X_i	Y_i	$X_i-\bar{X}$	$Y_i-\bar{Y}$	$(X_i-\bar{X})(Y_i-\bar{Y})$	$(X_i-\bar{X})^2$	$(Y_i-\bar{Y})^2$
75	85	6,5	20,5	133,25	42,25	420,25
80	80	11,5	15,5	178,25	132,25	240,25
95	90	26,5	25,5	675,75	702,25	650,25
60	50	-8,5	-14,5	123,25	72,25	210,25
40	40	-28,5	-24,5	698,25	812,25	600,25
70	55	1,5	-9,5	-14,25	2,25	90,25
65	50	-3,5	-14,5	50,75	12,25	210,25
25	20	-43,5	-44,5	1935,75	1892,25	1980,25
90	95	21,5	30,5	655,75	462,25	930,25
85	80	16,5	15,5	255,75	272,25	240,25
				4692,5	4402,5	5572,5

$$r = \frac{4692,5}{\sqrt{4402,5x5572,5}} = \frac{46925}{4953,073} = 0,947$$

Korelasyon değerine göre; iki değişken arasında aynı yönlü güçlü korelasyon olduğu söylenebilir.

b. Belirlilik katsayısı korelasyon katsayısı değerinin karesi olan (0,947)2=0,897 olacaktır. Dolayısıyla bu sınıftaki öğrencilerin istatistik dersinden elde ettikleri başarı puanlarına matematik dersinden elde ettikleri başarı puanlarının etkisinin %89,7 olduğu söylenebilir. Bu, oldukça büyük bir değer olduğundan, bu sınıftaki 10 öğrencinin matematik dersindeki başarılarının istatistik dersindeki başarılarına olumlu (pozitif) yönde güçlü bir etkisi olduğu söylenebilir.

c. H_0: $\rho = 0$
H_1: $\rho \neq 0$
hipotezleri kurulur.

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0,947\sqrt{10-2}}{\sqrt{1-0,897}} = \frac{2,6785}{0,3209} = 8,3468$$

bulunur.

8,3468 olarak hesaplanan t değeri Şekil 5.5'teki sıfır hipotezi red bölgesinde kaldığından, "İlgilenilen değişkenler arasındaki korelasyon sıfırdır" hipotezi reddedilir. Başka bir ifadeyle iki değişken arasında istatistiksel olarak anlamlı bir ilişki söz konusudur.

Yararlanılan ve Başvurulabilecek Kaynaklar

Cornillon, Pierre-André ve Matzner-Løber, Éric (2007), **Regression Theorie et Applications,** Springer-Verlag France, Paris.

Çömlekçi, Necla (1994), **Temel İstatistik İlke ve Teknikleri,** Bilim Teknik Yayınevi Yayını, Eskişehir.

Durucasu, Hasan (2003), **Excel ile Matris Uygulamaları,** Birlik Ofset Yayını, Eskişehir.

Şıklar, Emel (2000), **Regresyon Analizine Giriş,** Anadolu Üniversitesi Yayını, Eskişehir.

Yüzer, Ali Fuat (1996), **Olasılık ve İstatistik,** Anadolu Üniversitesi Yayını, Eskişehir.

6

Amaçlarımız

Bu üniteyi tamamladıktan sonra;

- Zaman serisini tanımlayabilecek,
- Zaman serisini etkileyen faktörleri açıklayabilecek,
- Zaman serisini etkileyen ana faktörü tanımlayabilecek,
- Hareketli ortalamalar tekniği ile trendi araştırabilecek,
- En küçük kareler tekniği ile doğrusal trend analizi yapabilecek,
- Doğrusal trend denklemi ile öngörüler yapabilecek

bilgi ve becerilere sahip olacaksınız.

Anahtar Kavramlar

- Zaman Serisi
- Öngörü
- Trend
- Konjonktür
- Mevsimsel Dalgalanmalar
- Düzensiz Dalgalanmalar
- Hareketli Ortalama
- Dalga Uzunluğu
- Dalga Şiddeti
- Normal Denklemler
- Sadeleştirilmiş Normal Denklemler
- Standart Hata
- Enterpolason
- Ekstrapolasyon

İçindekiler

Zaman Serileri Analizi

GİRİŞ

İstatistiksel bir olayda, gözlem değerlerinin zaman değişkeninin konumlarına göre aldığı değerlerin sıralanmasıyla elde edilen verilere **"zaman serisi"** denir. Zaman serileri genellikle eşit zaman aralıklarında toplanan gözlem değerlerinden oluşur; burada, bir filmin bir birini izleyen karelerine benzeyen eşit aralıklı gözlem değerleri, zamanın bir fonksiyonu olarak düşünülür: Y=f (zaman).

Bir olayın zaman değişkenine göre aldığı değerlerin alt alta sıralanması ile elde edilen veri setine **zaman serisi** denir.

Günlük yaşamda birçok olay zamana bağlı olarak incelebilmektedir. Bir işletmenin yıllara göre üretim miktarları serisi veya belli bir ürünün satış miktarı kayıtları birer zaman serisi oluşturur. Zaman serisi, yalnız bir çeşit ürünün (küçük boy buzdolabı) üretim miktarından oluşabileceği gibi, farklı ürünlerin (farkı büyüklükte buzdolapları) toplam üretim miktarından da oluşabilir.

Zaman serilerinde zaman birimi, yıl olabileceği gibi yılın alt zaman birimleri olan ay, hafta veya gün, hatta saat bile olabilir. Zaman serileri gözlem değerlerinin elde edilmesinde benimsenen yaklaşım, zaman değişkenini türüne göre adlandırılmasıdır. Örneğin gözlem değerleri aylık olarak elde edilmiş ise "*aylık zaman serisi*", yıllık olarak elde edilmiş ise "*yıllık zaman serisi*" adını almaktadır. Bir zaman serisi iki sütundan (en az) oluşmaktadır. Sütunların birinde daima zaman değişkeni, diğerinde ise gözlem değerleri (Y_i) yazılır.

Aylık zaman serisi		Yıllık zaman serisi	
Aylar	**A-malı Üretim mik. (Ton)**	**Yıllar**	**B-malı ihracat mik. (adet)**
Ocak	2610	2007	22.416
Şubat	3012	2008	31.600
Mart	3600	2009	74.200
Nisan	4118	2010	70.440
Mayıs	4502	2011	98.888

Zaman serisi analizi ise herhangi bir zaman serisine düzensiz görünüm veren dalgalanma veya hareketlerin, neden kaynaklandığını bularak zaman serisini bileşenlerine ayırmak, bunların gelecekte alacakları değerleri öngörmek ve bileşenleri birleştirerek belirli bir *öngörü* değerine ulaşmakla ilgilidir. Zaman serisi analizi ile gelecek hakkında öngörüler yapılırken geçmişteki hareketlerin gelecekte de aynı eğilim içinde bulunacağı varsayılır. Zaman serileri analizi ile elde edilecek öngörüler, gerek ekonomide gerekse işletme temelinde günlük kararlar alırken ve geleceğe dönük kısa veya uzun süreli planlar yaparken büyük yararlar sağlar.

Zaman serisi analizi, her bir değişkenin zaman içindeki oluşum sürecinin matematiksel modelinin belirlenmesidir.

SIRA SİZDE 1

Zaman serileri neden analiz edilir?

Zaman serisi değerleri devamlı derleme ile kaydedilen değerlerdir, tesadüfi olarak seçilen örnek değerler değildir. Daha doğrusu, ana kütleyi oluşturan verilerden eksiksiz bir kesit (genellikle son yıllardan veya aylardan oluşan) alınarak analiz yapılacak zaman serisi oluşturulur. Zaman serisi değerleri daha çok geleceğe dönük öngörü için kullanılır. İstatistiksel çıkarsama tekniklerinden olan aralık tahmini ve hipotez testleri zaman serilerinde uygulanmaz.

DİKKAT

Zaman serisi analizinde, verilerin aynı ölçü birimi ile ifade edilmiş olması gerekir. Bu amaçla verilere fiziksel bir ölçü birimi uygulanabileceği gibi parasal değer cinsinden de ifade edilebilir.

Zaman serisinin grafiği çizilirken, zaman değişkeni yatay eksende, gözlem değerleri değişkeni de dikey eksende yer alır.

Zaman serileri genel olarak kartezyen koordinatlı bir grafikle gösterilir. Bir **zaman serisinin grafiği** çizilirken gözlem değerleri (Y_i) ordinat ekseninde ve zaman da daima apsis ekseninde gösterilir. Örnek 1'de un tüketimi zaman serisinin grafiği kartezyen koordinat diyagramında gösterilmiştir.

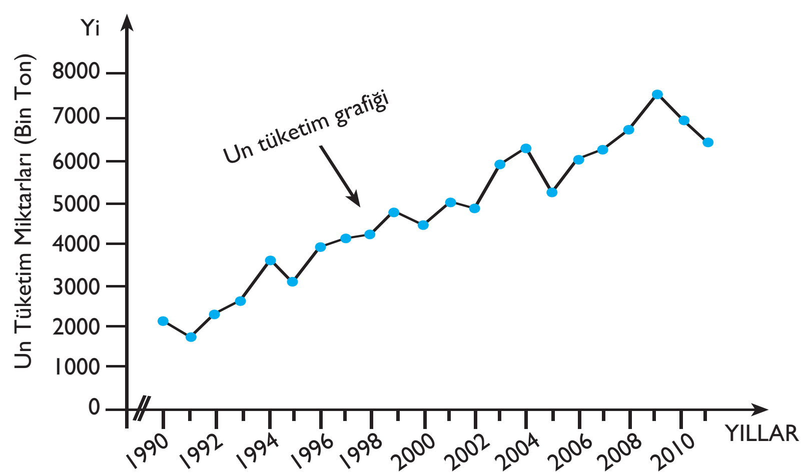
ÖRNEK 1

1990-2011 yılları arasında bir ülkenin un tüketimi (Bin Ton) bir zaman serisi olarak aşağıda verilmiş ve Şekil 6.1' de grafiği çizilmiştir.

Yıllar	Un Tük Mik (Yi).	Yıllar	Un Tük. Mik. (Yi)
1990	2.510	2001	5.020
1991	1.890	2002	4.900
1992	2.360	2003	5.900
1993	2.700	2004	6.330
1994	3.380	2005	5.070
1995	3.140	2006	6.160
1996	3.990	2007	6.300
1997	4.030	2008	6.800
1998	4.170	2009	7.640
1999	4.790	2010	6.960
2000	4.500	2011	6.470

Şekil 6.1

Yıllara göre un tüketimi serisinin grafiği.



Zaman serisinin grafiği nasıl çizilir?

SIRA SİZDE 2

Ekonomik, sosyolojik, demografik gibi birçok konuda zaman serileri vardır ve bunlardan bazıları aynı, bazıları farklı faktörlerden etkilenmektedir. Bu seriler arasında bir iktisatçı veya bir işletmeci için asıl önemli olan ve üzerinde durulması gereken "ekonomik zaman serileri" dir. Bu nedenle izleyen bölümde ekonomik zaman serilerini etkileyen faktörlerin açıklanmasına ağırlık verilecektir.

ZAMAN SERİLERİNİ ETKİLEYEN FAKTÖRLER

Zaman serileri içinde en önemlileri ekonomik olaylarla ilgili olanlardır. Üretim, tüketim, alış, satış, ithalat-ihracat vb. seriler analiz edilmeden büyük bir anlam taşımazlar. Zaman serisinin grafiği çizildiğin de görülebilir ki gözlem değerleri birtakım **dalgalanmalar** gösterir. Söz konusu bu dalgalanmalar, ele alınan bu olayı etkileyen çeşitli faktörlerin sonuçlarıdır. Faktörlerden biri gözlem değerlerinin artması yönünde etki yaparken bir diğeri azalma yönünde etkili olabilir. Zaman serilerinin değerlerinde görülen bu dalgalanmaların *dört faktörün* etkisinden kaynaklandığı varsayılmaktadır. Bu dört faktör:

Zaman serilerinde gözlenen **dalgalanmalar** zaman serisini etkileyen faktörlerin etkisiyle meydana gelebileceği gibi ekonomik, sosyal ve psikolojik etkenlerin birleşiminin bir sonucu da olabilir.

T: Trend (uzun devre eğilimi, ana eğilim),
M: Mevsimlik Dalgalanmalar,
K: Konjonktürel Dalgalanmalar,
D: Düzensiz (Rassal) Dalgalanmalar olarak kabul edilmektedir.

Bir araştırmacı ele aldığı zaman serisini analiz etmek istediğinde bu faktörlerden her birinin hangi yönde ve şiddette etkin olduğunu araştırmak zorundadır. Bu ise zaman serisinin gerçekçi verilere dayanmasına bağlıdır.

Zaman serisinin gerçek gözlem değerleri (Y_i) ile yukarıda sayılan faktörler arasındaki *matematiksel ilişki:*

$$Y_i = T_i + K_i + M_i + D_i$$

şeklinde "*toplamsal ilişki*" kurulabileceği yönünde görüşler olmakla birlikte, bu konuda genel kabul görmüş yaklaşım "*çarpımsal ilişki*"yi kullanmak şeklindedir. Çarpımsal ilişki:

Aylık Zaman serilerinde $\rightarrow Y_i = T_i \times K_i \times M_i \times D_i$

şeklinde ki bu çarpımsal ilişkide, zaman serisinin gözlem değerlerinin, seriyi etkileyen faktörlerin çarpımlarından oluştuğu varsayılmaktadır.

Eğer gözlem değerleri *yıllık verilerden* oluşuyorsa mevsimlik dalgalanmaların etkisini taşımayacağından, çarpımsal ilişki:

Yıllık Zaman serilerinde $\rightarrow Y_i = T_i \times K_i \times D_i$ şekline dönüşür.

Burada,
Y_i : Zaman serisinin i dönemindeki gerçek gözlem değerini,
T_i : Trendin i - dönemindeki etkisini,
K_i : Konjonktürün i - dönemindeki etkisini,
M_i : Mevsimin i - dönemindeki etkisi,
D_i : Düzensiz dalgalanmaların i - dönemindeki etkisini simgelemektedir.

Örneğin, Örnek 1'deki zaman serisinde 2000 yılındaki un tüketim miktarı: Y_{2000}=4.500.000 tonluk gözlem değeri,

$Y_{2000} = T_{2000} \times K_{2000} \times D_{20000}$ =4.500.000 ton demektir.

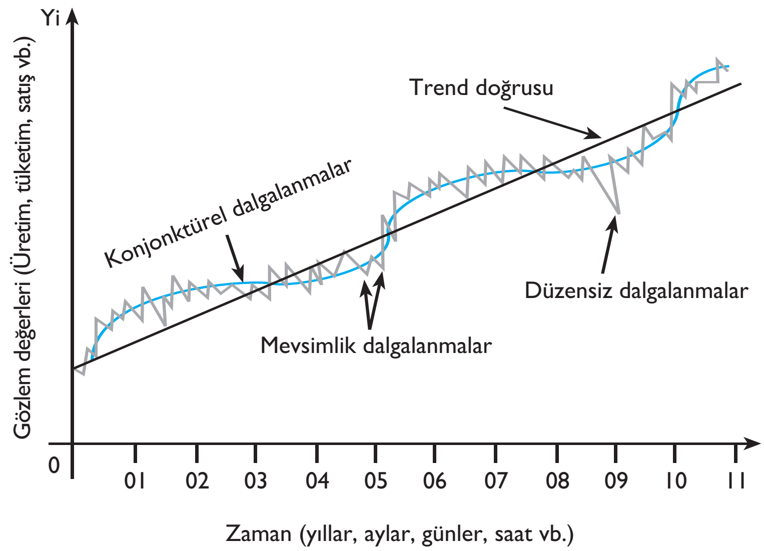
SIRA SİZDE 3

Zaman serisini (Yıllık) etkileyen faktörler nelerdir?

Zaman serisinin dalgalanmaları (Şekil 6.2'de) farklı uzunluk, yön ve şiddettedir; bunlar dört faktörün (T, K, M ve D) etkisinden kaynaklanmaktadır. Şekil 6.2'de varsayımsal bir ekonomik zaman serisi üzerinde etkin olan bütün faktörler bir arada gösterilmiştir.

Şekil 6.2

Zaman serisini etkileyen faktörler.



Zaman serisini etkileyen ilk iki (mevsimlikte ilk üç) faktör sistematik bileşen olarak isimlendirilir, düzensiz dalgalanmalar ise rassal hata bileşeni adını alır. Söz konusu faktörlerin özellikleri sırasıyla açıklanmaya çalışılacaktır.

Trend (Ana Eğilim-Uzun Devre Eğilimi)

Bir zaman serisinde uzun dönem içinde görülen artma veya azalma **trend (ana eğilim)** adını alır.

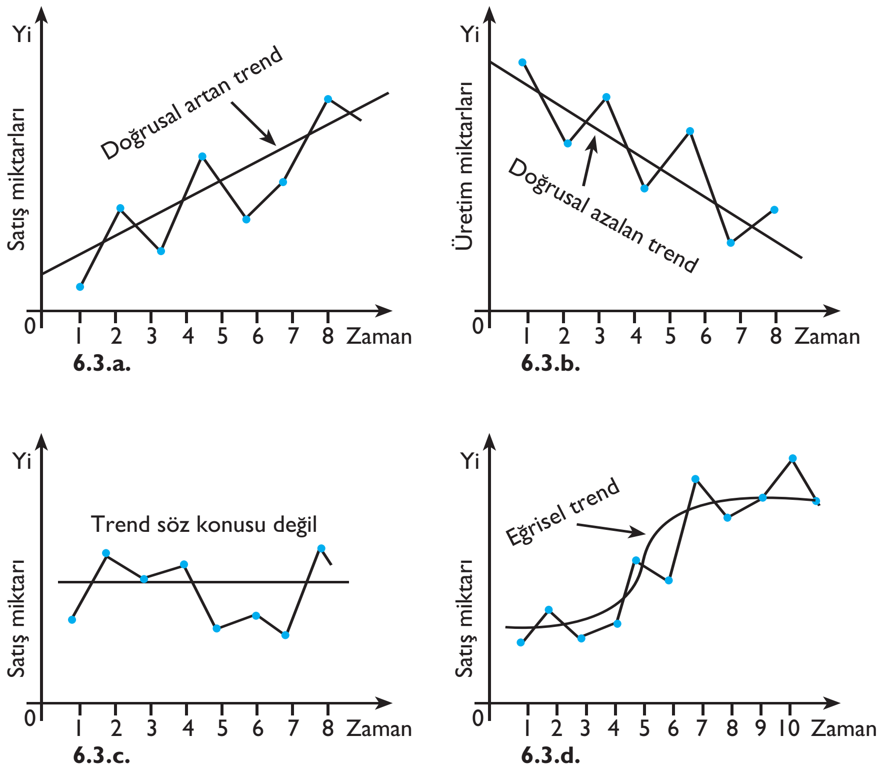
Zaman serilerindeki dalgalanmalara neden olan faktörler, ona belirli bir yön verir. Bir zaman serisinin uzun bir dönemde belli bir yöne doğru gösterdiği eğilime **trend** veya **ana eğilim** adı verilir. Trend, zamana bağlı değişken üzerinde genel eğilime neden olan uzun dönemli etkileri açıklar. Bu etkiler genel olarak nüfus artışları, sermaye birikimindeki artışlar, enerji kaynaklarının genişlemesi, teknik ve teknolojik gelişmeler, tüketici zevk ve alışkanlıklarındaki değişmeler ve fiyat değişmelerindeki etkiler olabilir.

Trend, bağlı olduğu faktörlere göre artan veya azalan bir yönde olabilir. Bu artış ve azalışa etki eden faktörlerin şiddetine göre ilerleyiş, bazen hızlı bazen yavaş olabilir. Dolayısıyla *trend, doğrusal* olabileceği gibi *eğrisel* de olabilir. *Şekil 6.3 (a, b ve d)*'de yer alan trend türleri, uygulamada sık karşılaşılan türlerdir.

Zaman içinde artış veya azalış göstermeyip aşağı yukarı aynı düzeyde kalan (Şekil6.3c.) zaman serilerinin trendi yoktur.

Şekil 6.3

Trend türleri.



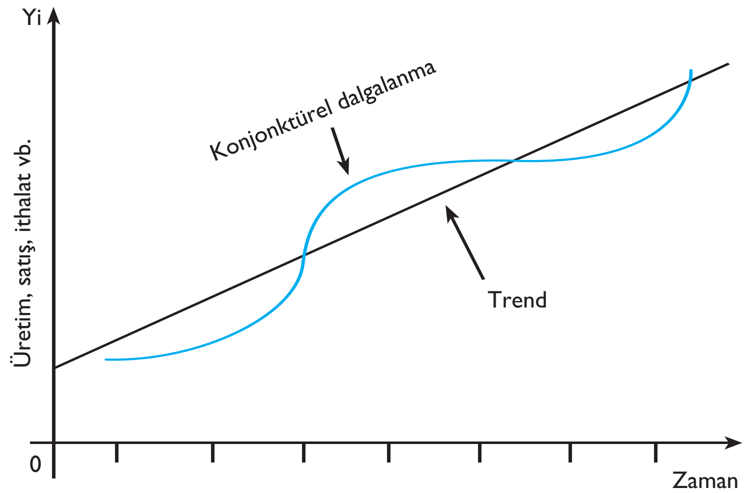
Konjonktürel Dalgalanmalar

Zaman serisini etkileyen diğer bir faktör, konjonktürel dalgalanmalardır. Zaman serisinin trend doğrusu veya eğrisi etrafındaki uzun dönem dalgalanmalarına konjonktürel dalgalanmalar denir. İş dünyasında yatırım artışlarının, üretim artışlarına ve üretim artışlarının gelir artışlarına yol açmasıyla ekonomik durumda bir süreliğine bir gelişme görülür. Bu gelişme genel ekonomide olabileceği gibi bir sektörde de (inşaat sektörü gibi) görülebilir. Gelişmenin maksimum aşamasında bir kriz ortaya çıkar. Sonra bir düşüş başlar. İzleyen aşamada işler belli bir düzeyde bir süre hareketsiz kalır. Daha sonra işlerde yeniden bir kımıldanma ve canlanma görülmeye başlar. Bu aşamalar 3-15 yılda bir tekrarlanır gider. Konjonktürel dalgalanmayı gösteren örnek bir grafik Şekil 6.4'te verilmiştir.

Konjonktürel dalgalanmalar periyodik değil, ancak döngüsel dalgalanmalardır. Bu dalgalanmalar ekonomi ve işletmeyle ilgili değişkenler üzerinde aynı şiddette olmasa da aynı yönde etki ederler. Konjonktürün artma yönündeki etkisi, trendin artış eğilimini hızlandırır. Buna karşın, konjonktürün azalma yönünde etkisi, tendin artış hızını yavaşlatır, hatta tamamen ortadan kaldırabilir.

Şekil 6.4

Konjonktürel dalgalanmalar.



SIRA SİZDE 4

Konjonktürel dalgalanmalar ne demektir?

Mevsimlik Dalgalanmalar

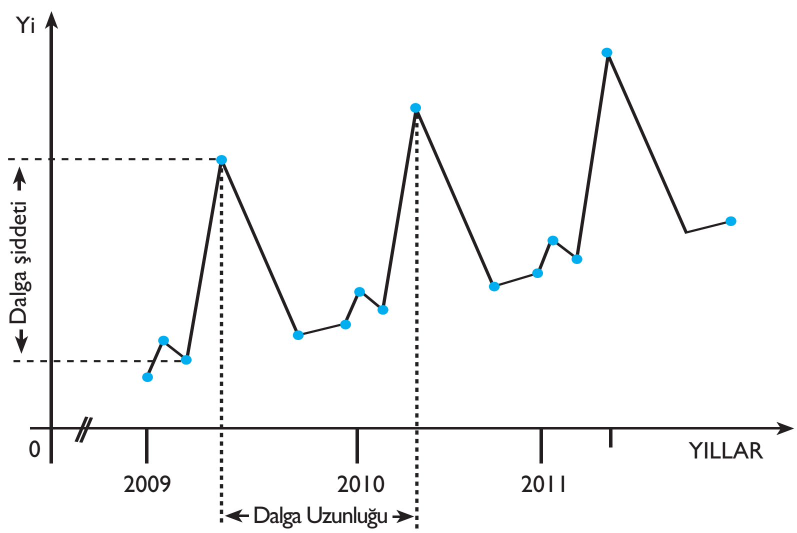
Zaman serisinde, bir yıllık aralarla düzenli bir şekilde tekrarlanan dalgalanmalara **mevsimlik dalgalanmalar** adı verilir.

Zaman serisini etkileyen diğer bir faktör, mevsimlik dalgalanmalardır. Bir zaman serisinde mevsim etkisiyle bir yıllık aralarla düzenli bir şekilde tekrarlanan dalgalanmalara, **mevsimlik dalgalanmalar** adı verilir. Mevsimlik dalgalanmaları gösteren zaman serisinin kartezyen koordinat grafiği örneği Şekil 6.5'te verilmiştir. Mevsimlik dalgalanmalarda birbirini izleyen iki maksimum veya minimum nokta arasındaki zaman farkına *dalga uzunluğu* adı verilir. Zaman serilerinde mevsimsel dalgalanmaların dalga uzunluğu hep aynı ve 12 aydır. Mevsimlik dalgalanmalar, dalga uzunluklarının birbirine eşit olması nedeniyle periyodik, tekrar tekrar meydana gelmesi nedeniyle de döngüsel özelliğe sahiptir.

Mevsimlik dalgalanmaların maksimum ve minimum noktaları arasındaki yükseklik farkına *dalga şiddeti* adı verilir. Mevsimlik dalgalanmaların dalga uzunluğu ve dalga şiddeti Şekil 6.5'te görülmektedir.

Şekil 6.5

Mevsimlik dalgalanmalar.



Mevsimlik dalgalanmalar genellikle doğal ve sosyoekonomik nedenlerden ortaya çıkar. Bir malın üretim, satış, tüketim ve fiyatında hava koşulları ve alışkanlıklar nedeniyle mevsimlik değişmeler meydana gelebilir. Örneğin ülkemizde süt üretimi yaz aylarında artarken kış aylarında azalır. Doğal gaz tüketimi kış aylarında artarken yaz aylarında azalır. Kısacası bazı malların tüketimi, yaz aylarında en düşük, kış aylarında en yüksek düzeye ulaşmaktadır. Kuşkusuz bu durumun tersi de mümkündür. Hemen ekleyelim ki ekonomik faaliyetler mevsimlik dalgalanmalardan özel bir şekilde ve farklı derecede etkilenir.

Düzensiz Dalgalanmalar

Düzensiz dalgalanmalar, beklenmedik olayların zaman serileri üzerinde etkisiyle meydana gelen dalgalanmalardır. Düzensiz dalgalanmalara, ne zaman ve ne şekilde ortaya çıkacağı bilinmeyen su baskını, deprem, don gibi doğal afetler ile grev, seçimler, savaşlar gibi beklenmeyen olaylar neden olabilir.

Rassal nedenlerle veya geçici olarak ortaya çıkan dalgalanmalara **düzensiz dalgalanmalar** denir.

TREND ANALİZİ

Bir zaman serisinin uzun bir dönemde belli bir yöne doğru gösterdiği ana eğilim olarak tanımladığımız trendin bulunmasının iki temel amacı vardır. Trend analizinin birinci amacı, zaman serisini konjonktürel dalgalanmalar ve düzensiz dalgalanmaların etkisinden arındırarak, sadece uzun dönem hareketlerin etkisi altındaki seri değerlerini ortaya çıkarmaktır. Trend analizinin ikinci amacı ise seriyi öngörü amacıyla analiz etmektir. Zaman serisinin öngörü amacıyla analizi, bir zaman serisini etkileyen faktörlerin belirlenmesi, yapılan belirlemeden geçmişin açıklanması ve gelecek dönemler için öngörü yapılması ve bu öngörülerin karar verme ve planlama faaliyetleri için kullanıma sunulması çalışmalarıdır.

Trend analizinde, trendi bulunacak zaman serisinin elden geldiğince çok sayıda veriyi kapsaması, ancak bu veriler içinde aşırı değerlerin bulunmaması arzu edilir. Uygulamada en az “10-15 yıllık” veriyle trend hesaplanır. Çünkü 5-6 yıldan çok kısa bir dönemde trend yeterince ortaya çıkamaz. Öte yandan 30-40 yıldan çok veriyle de çalışmamak gerekir. Çünkü böylesi uzun bir dönemde trend değişmiş olabilir.

Trend analizinde çeşitli tekniklerden yararlanılmaktadır. Bunlardan en çok bilinen ve yaygın kullanıma sahip olan; Hareketli Ortalamalar Tekniği ve En Küçük Kareler Tekniği (EKKT)’dir. Şimdi bunları sırayla açıklamaya çalışalım.

SIRA SİZDE 5

Trend analiz teknikleri nelerdir?

Hareketli Ortalamalar Tekniği

Hareketli ortalamalar tekniğinin esası, zaman serisinin gözlem değerlerini belirli büyüklükte kümeler hâlinde toplama, her küme için aritmetik ortalama hesaplamak ve bu ortalamaları ilgili kümede tam ortaya düşen değerin yerine koymaktır. Hareketli ortalamalar tekniği, zaman serisinin grafiği çizildiğinde açıkça görülen konjonktürel ve mevsimlik dalgalanmaların etkisini ortadan kaldırmak amacıyla kullanılır.

Hareketli ortalamalar hesabı, zaman serisi boyunca hareket eden aritmetik ortalamaya dayanır. Zaman serisinin değerleri, belirli kümeler hâlinde toplanır ve her bir küme için aritmetik ortalama hesaplanır. Hesaplanan ortalama, ilgili kümenin tam ortasına düşen yıla karşılık yazılır. Küme, her defasında bir aşağıya kaydırılarak serinin sonuna kadar ortalama hesaplama işlemi sürdürülür.

Hareketli ortalamaların hesaplanmasıyla asıl zaman serisi gözlem değerleri yerine, ortalamalardan meydana gelmiş yeni seri geçmiş olur; bu yeni serinin ilgilenilen olayın trendini gösterdiği kabul edilir.

DİKKAT

Bir zaman serisinde, bu koşulların üçünün de bulunması, özellikle dalga uzunluklarının eşit olması ender görülür.

Bir zaman serisi trendinin belirlenmesinde hareketli ortalamalar tekniği kullanılabilmesi için;

- *Trendi doğrusal* olmalı,
- *Dalga uzunlukları* eşit olmalı,
- *Dalga şiddetleri* aynı olmalıdır.

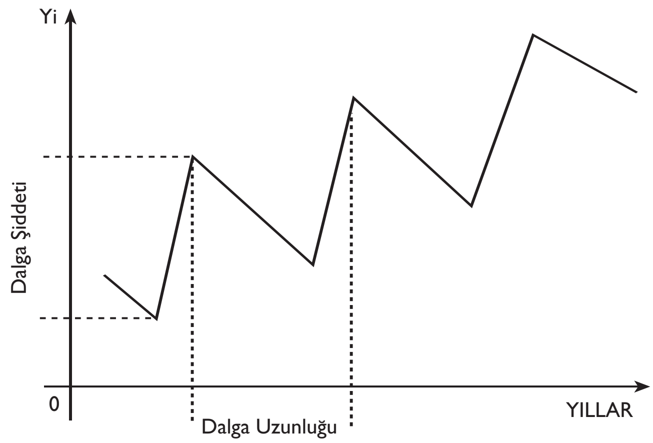
Zaman serisinde birbirini izleyen iki maksimum (veya minimum) nokta arasındaki zaman genişliği, *dalga uzunluğu* olarak tanımlanır. Aynı dalgada maksimum ve minimum noktalar arasındaki yükseklik farkı da *dalga şiddeti* olarak tanımlanır (Şekil 6.6).

SIRA SİZDE 6

Hareketli ortalamalar tekniği hangi amaçla kullanılır?

Şekil 6.6

Dalga uzunluğu ve dalga şiddeti.



Hareketli ortalamaların "kaç gözlem değeri" üzerinden hesaplanacağı, başka bir ifadeyle, hareketli ortalamanın "*kaçarlı*" olacağının belirlenmesi önemlidir. Çünkü hareketli ortalama hesabında farklı değerler için farklı ortalama bulunur. Hesaplanan ortalamalar her bir kümede tam ortaya düşen gözlem değerinin yerine geçirilir. Hareketli ortalama kümeleri *tek* sayıda gözlem değerinden oluşuyor ise (3, 5, 7...gibi) trend değeri aşağıdaki gibi hesaplanır. Örneği; "*Üçerli hareketli ortalama*" şu şekilde hesaplanır;

Gözlem değeri (Y_i)	$\hat{Y}$ (3'erli H.O.)
Y_1	——
Y_2	$\hat{Y}_2 = \dfrac{Y_1 + Y_2 + Y_3}{3}$
Y_3	$\hat{Y}_3 = \dfrac{Y_2 + Y_3 + Y_4}{3}$
"	"
Y_{n-1}	$\hat{Y}_{n-1} = \dfrac{Y_{n-2} + Y_{n-1} + Y_n}{3}$
Y_n	——

Tek sayıda değer üzerinden hareketli ortalama hesaplandığında kümelerin gözlem değeri sayısı **"k"** olmak üzere **(k-1)** kadar gözlem değeri (baştan ve sondan eşit miktarda olmak üzere) kaybedilir. Üçerli hareketli ortalamada **(k=3)** bir baştan ve bir sondan olmak üzere **2** gözlem değeri kaybedilmiştir. Dörderli hareketli ortalamada **(k=4)**, iki baştan ve iki sondan olmak üzere toplam **4** gözlem değeri kaybedilir. Beşerli hareketli ortalamada ise yine toplam dört gözlem değeri kaybedilir. Hareketli ortalamalarda, kümenin gözlem değeri sayısı arttıkça, trend daha doğrusal çıkar. Buna karşın hareketli ortalama kümesinin gözlem değeri sayısı arttıkça kaybedilen gözlem değeri sayısı da artmaktadır. Bu da gerçek gözlem değerlerinden uzaklaşma olumsuzluğunu doğurur.

ÖRNEK 2

Aşağıda verilen zaman serisinde yıllara göre tahıl üretim miktarı verilmiştir. Trendini hareketli ortalamalar tekniği ile belirleyiniz.

Yıllar	***Tahıl üretim mik. (bin ton) Y_i***
2003	*21*
2004	*44*
2005	*93*
2006	*69*
2007	*84*
2008	*131*
2009	*103*
2010	*127*
2011	*175*

Çözüm: Serinin grafiği (Şekil 6.7) incelendiğinde dalga uzunlukları aynı ve 3'er yıl olduğu görülmektedir. Bu nedenle hareketli ortalama 3'erli olacaktır. 3'erli hareketli ortalama hesabı aşağıda verilmiştir.

Yıllar	**Y_i**	**$\hat{Y}_i$ (3'erli H.O.)**
2003	21	———
2004	44	$\hat{Y}_{04} = \dfrac{21+44+93}{3} = 52,67$
2005	93	$\hat{Y}_{05} = \dfrac{44+93+69}{3} = 68,67$
2006	69	$\hat{Y}_{06} = \dfrac{93+69+84}{3} = 82,00$
2007	84	$\hat{Y}_{07} = \dfrac{69+84+131}{3} = 94,67$
2008	131	$\hat{Y}_{08} = \dfrac{84+131+103}{3} = 106,00$
2009	103	$\hat{Y}_{09} = \dfrac{131+103+127}{3} = 120,33$
2010	127	$\hat{Y}_{10} = \dfrac{103+127+175}{3} = 135,00$
2011	175	———

Hesaplanan 3'erli hareketli ortalama değerleri Şekil 6. 7'de görüldüğü gibi koordinat diyagramında işaretlenir ve doğru parçalarıyla birleştirilerek trend çizilir.

Şekil 6.7

Örnek 2'deki zaman serisinin ve 3'erli hareketli ortalama trend grafiği.

Bu örnekte hareketli ortalamalar k=3'erli olduğu için baştan 1 ve sondan 1 olmak üzere toplam 2 değer kaybedilmiştir. Kaybedilen değerler 2003 ve 2011 yıllarına ilişkindir. Şekil 6. 7'de görüldüğü gibi trend, zaman serisinin grafiğine göre daha doğrusala yakın gözükmektedir.

Kümenin *çift* sayıda (**4, 6, 8,** gibi) değerden oluşması hâlinde, tam ortaya bir gözlem değeri düşmez ve böylece hesaplanan ortalamanın hangi yıla ait olduğu belli olmaz. Bundan kurtulmak için küme bir fazla yıl alınarak oluşturulur. Ancak birinci ve sonuncu yıl değerlerine (1/2) tartısı verilir. Dörderli hareketli ortalama aşağıdaki örnekteki gibi hesaplanır.

ÖRNEK 3

Aşağıda verilen zaman serisinde yıllara göre domates ihracat miktarları (milyon ton) verilmiştir. Trendini, hareketli ortalamalar tekniği ile belirleyiniz.

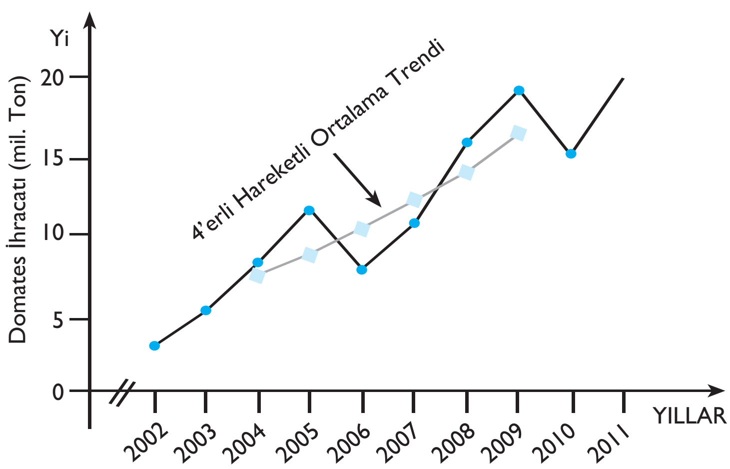
Çözüm: Verilen ihracat zaman serisini grafiğinde (Şekil 6. 8) görülebileceği gibi dalga uzunluğu 4-yıldır. Bu nedenle trend, 4'erli hareketli ortalama ile bulunacaktır. Dörderli hareketli ortalamada, kümeler 5 yıldan oluşturulacak ancak birinci ve beşinci yıl değerlerinin yarısı hesaba katılacaktır.

Yıllar	Y_i	$\hat{Y}_i$ (4'erli H.O.)
2002	3	——
2003	5	——
2004	8	$\hat{Y}_{04} = \dfrac{3/2+5+8+12+7/2}{4} = 7{,}50$
2005	12	$\hat{Y}_{05} = \dfrac{5/2+8+12+7+11/2}{4} = 8{,}75$

2006	7	$\hat{Y}_{06} = \dfrac{8/2+12+7+11+16/2}{4} = 10,50$
2007	11	$\hat{Y}_{07} = \dfrac{12/2+7+11+16+19/2}{4} = 12,38$
2008	16	$\hat{Y}_{08} = \dfrac{7/2+11+16+19+15/2}{4} = 14,25$
2009	19	$\hat{Y}_{09} = \dfrac{11/2+16+19+15+20/2}{4} = 16,38$
2010	15	——
2011	20	——

Hesaplanan 4'erli hareketli ortalama değerleri Şekil 6. 8'de görüldüğü gibi koordinat diyagramında işaretlenir ve doğru parçaları ile birleştirilerek trend çizilir.

Şekil 6.8



Domates ihracatı zaman serisinin ve 4'erli hareketli ortalama trendi grafiği

Şekil 6.8'de görüldüğü gibi trend, zaman serisinin grafiğine göre doğrusala daha yakın görülmektedir. Bu örnekte hareketli ortalamalar k=4'erli olduğu için, baştan 2 ve sondan 2 olmak üzere, toplam 4 değer kaybedilmiştir. Kaybedilen değerler 2002, 2003, 2010 ve 2011 yıllarına ilişkindir.

Hareketli ortalama tekniği, en küçük kareler tekniği gibi trend için matematiksel bir fonksiyon vermez. Bunun sonucu olarak da bu teknikle geleceğe dönük öngörüler yapmak mümkün değildir. Buna karşın konjonktürel dalgalanmalar elendiğinde sadece trend ortaya çıkmış olur.

En Küçük Kareler Tekniği (EKKT)

Zaman serileri analizinde, zaman değişkeni ile gözlem değerleri arasındaki fonksiyonel ilişki *en küçük kareler* tekniği ile araştırılır. Bu fonksiyonel ilişki *doğrusal* olabileceği gibi *eğrisel de* olabilir. Elde edilen fonksiyona ilişkin doğru veya eğrinin, zaman serisinin grafiğine uygun olması ve

Y_i: gözlem değerleri
$\hat{Y}$: Trend değerleri (teorik değerleri),
göstermek üzere,

$\Sigma(Y\text{-}\hat{Y})^2$= minimum

koşulunu sağlaması gerekir. Zaman serisini en iyi temsil edecek *fonksiyon tipinin seçilmesi* çok önemlidir. Bunun için önce, zaman (X) apsis ekseninde, gözlem değerleri (Y_i) ordinat ekseninde işaretlenerek serinin serpilme diyagramı çizilir. Serpilme diyagramına bakılarak, konjonktürel ve düzensiz dalgalanmalara rağmen, uzun devrede nasıl bir eğilim gösterdiği izlenir. Serpilmenin genel seyrinin bir *doğruya mı, yoksa eğriye mi, uyduğuna* karar verilir. *Eğriye uydu ise* bükülme noktalarının sayısına göre eğrinin derecesi belirlenir. Kısacası, serpilme diyagramındaki noktaların dağılımına göre trendin uygun fonksiyon türü belirlenirken

- Gelişim yönünü değiştirmeyen, devamlı olarak artışı (veya azalışı) sabit gibi görünenlere *doğrusal fonksiyon:* $\hat{Y} = a + bX$
- Artış (veya azalış) oranı sabit gibi görünenlere *üstel fonksiyon:* $\hat{Y} = ab^X$ veya $\hat{Y} = aX^b$
- Bükülme noktalarının sayısına göre çeşitli derecelerden *parabolik fonksiyonlar:* $\hat{Y} = a + bX + cX^2$ veya $\hat{Y} = a + bX + cX^2 + dX^3$ kullanılır.

Bu ünitede; genellikle temel istatistik düzeyinde anlatılan doğrusal trend analizi üzerinde durulacaktır.

KİTAP

Eğrisel trend analizi için bkz. Seper, Ö., (2004), "Uygulamalı İstatistik 2", Ezgi Kitabevi, Bursa.

Doğrusal *trend denklemi:* $\hat{Y}$ = **a + bX** şeklinde yazılır.
Söz konusu trend denkleminde gerçek gözlem değerleri Yi olmak üzere;
$\hat{Y}$: Trend değerlerini (teorik değerleri),
a: Trend doğrusunun Y (ordinat) eksenini kestiği noktayı,
b: Zamanda bir birimlik değişmenin $\hat{Y}$'de meydana getireceği değişim miktarını, aynı zamanda trend doğrusunun eğimini,
X: Bağımsız değişken durumundaki zamanı göstermektedir.
Bu doğru denkleminde **"a"** ve **"b"** katsayılarını;

$\Sigma(Y\text{-}\hat{Y})^2$= minimum

şartını sağlayacak şekilde *en küçük kareler tekniği* ile belirleyebilmek için aşağıdaki *"normal denklemlerin"* bir arada çözülmesi gerekir.

Normal denklemler;
$\boldsymbol{\Sigma Y = na + b\,\Sigma X}$
$\boldsymbol{\Sigma XY = a\,\Sigma X + b\,\Sigma X^2}$

KİTAP

Normal denklemlerin nasıl elde edildiğini bkz. Atlas, M., (2007) İstatistik II Çözümlü Örnekler, Akofset, Eskişehir.

Bu iki denklem taraf tarafa toplanarak **“a”** ve **“b”** katsayıları:

$$b = \frac{\sum XY - \frac{\left(\sum X\right)\left(\sum Y\right)}{n}}{\sum X^2 - \frac{\left(\sum X\right)^2}{n}} \quad ve \ a = \overline{Y} - b\overline{X}$$

formülleri yardımıyla bulunur.

Hemen belirtelim ki zaman serisinin trendini belirlerken uygun fonksiyon türünü belirlemede kararsızlığa düşüldüğünde; çeşitli fonksiyonların her biri için standart hata hesaplamak ve en küçük *standart hata*ya sahip olan fonksiyonu tercih etmek gerekir. Standart hata hesabı ilerleyen kısımda açıklanacaktır.

$\hat{Y}$ = a + bX doğrusal fonksiyonunda “a” ve “b” değerleri normal denklemlerle hesaplanırken: Yılları, aynen hesaplamalara dahil etmek zorunlu değildir. X’ler zamanı göstermekle birlikte, hesaplamalarda yıllar yerine 0 veya 1 ile başlayan ve aralarındaki farklar korunarak düzenlenen rakamlar yazılabilir.

Seriyi oluşturan gözlem sayısı *“tek yıllık (n-tek sayı)”* ise ortada bir yıl vardır ve tam ortaya düşen yıla (yani medyan yıla) X=0 değeri verilir. Buna orijinin medyan yıla kaydırılması işlemi denir. Bu durumda X’ ler ...-3, -2, -1, 0, 1, 2, 3, ... olur. Dolayısıyla bu X-değerlerinin tek kuvvetlerinin toplamı daima *sıfır* **(Σ X=0)**’dır.

Seriyi oluşturan gözlem sayısı *“çift yıllık (n-çift sayı)”* ise tam ortaya düşen bir yıl yoktur. Bu durumda tam ortaya düşen iki yıl birlikte ele alınarak *birincisi* için **X= -1**, *ikincisi* için **X= +1** değeri verilir. Böylece bir birini takip eden iki yıl için X’ lerin arasında 2 birimlik farklılık oluşmuş olur. Bu nedenle, önceki yıllara ikişer azaltarak, sonraki yıllara ikişer çoğaltarak X değerleri verilir. Böylece X’ ler ...., -5, -3, -1, 1, 3, 5, .. olur. Bunun sonucu olarak X’ lerin tek dereceli toplamları yine *sıfır* **(Σ X=0)** olur.

“X” değerlerinin toplamının *sıfıra* **(Σ X=0)** eşit olması bu şekilde sağlandıktan sonra *sadeleştirilmiş normal denklemlere* ulaşılır.

Normal denklemler		Sadeleştirilmiş normal denklemler
$\Sigma Y = na + b\Sigma X$	⟶	$\Sigma Y = na$
$\Sigma XY = a\Sigma X + b\Sigma X^2$	⟶	$\Sigma XY = b\Sigma X^2$

Böylece sadeleştirilmiş normal denklemlerden “a” ve “b” katsayıları,

$$a = \frac{\sum Y}{n} \quad , \quad b = \frac{\sum XY}{\sum X^2}$$

formülleri yardımıyla bulunur.

DİKKAT

Daha önceki sayfada normal denklemlerin taraf tarafa toplanması yardımıyla yazılan “a” ve “b” katsayılarının formülleri, (X=0 olması ile;

$$b = \frac{\sum XY - \frac{(0)\cdot\sum Y}{n}}{\sum X^2 - \frac{(0)^2}{n}} = \frac{\sum XY}{\sum X^2}$$

$$a = \overline{Y} - b(0) = \overline{Y} = \frac{\sum Y}{n}$$

şeklinde yazılır.

Bu formüllerle bulunan **"a"** ve **"b"** değerleri yerine konarak doğrusal trend denklemine ulaşılmış olur.

$\hat{Y} = a + bX$

Trend denkleminde X değerleri yerine konularak teorik trend değerleri ($\hat{Y}$) hesaplanır. *Gerçek gözlem değerleri* toplamı (ΣY_i) ile *teorik trend değerleri* toplamı ($\Sigma \hat{Y}_i$) bir birine eşittir: $\Sigma Y_i = \Sigma \hat{Y}_i$

Ancak uygulamada bazen yuvarlamalar nedeniyle çok küçük bir fark doğabilir, bu durum uygulama için önemli bir sorun yaratmaz.

Teorik olarak anlatılan en küçük kareler tekniği ile trend analizi, örnekler yardımıyla açıklanmaya çalışılacaktır.

ÖRNEK 4

2001 ile 2011 yılları arasında bir bisiklet üretim işletmesinin satış gelirleri (Milyon ₺) zaman serisi olarak aşağıda verilmiştir.

a. *Verilen zaman serisinin trend denklemini en küçük kareler tekniği ile belirleyiniz.*

b. *Bulacağınız trend denklemi yardımıyla, teorik trend değerlerini hesaplayarak, zaman serisi grafiğinde gösteriniz.*

c. *2001-2011 yılları arasında söz konusu gelirleri elde eden işletme, gelecek **2012 ve 2013** yıllarında ne kadar satış gelirine ulaşabilir, öngörünüz.*

Yıllar	Y_i *(Satışlar-Milyon ₺)*
2001	*35*
2002	*42*
2003	*53*
2004	*55*
2005	*60*
2006	*70*
2007	*80*
2008	*99*
2009	*116*
2010	*127*
2011	*120*

Çözüm: Yıllara göre bisiklet satış gelirlerinden, 11 yıllık bir zaman serisi oluşturulmuştur. Bir başka ifade ile yıllık bisiklet satış gelirlerinden oluşan bir ana kütleden, son 11 yıllık bir kesit alınmıştır.

a. Koordinat diyagramındaki serpilme noktaları (Şekil 6.9) doğru parçaları ile birleştirildiğinde, zaman serisinin uzun dönemde genel bir artış gösterdiği görülmektedir. Bu nedenle zaman serinin *trendi* için en uygun fonksiyon, **doğrusal fonksiyondur:** $\hat{Y} = a + bX$

Bu fonksiyonda **"a"** ve **"b"** yi bulabilmek için, öncelikle zaman serisinde X değerleri belirlenir. Şöyle ki gözlem sayısı **n=11**- *tek sayı* olduğundan serinin tam ortasına gelen yıla **X=0** değeri verilir ve diğer X değerleri buna bağlı olarak aşağıdaki (üçüncü sütunda) gibi oluşturulur.

Yıllar	Y_i	X_i	XY	X_i^2
2001	35	-5	-175	25
2002	42	-4	-168	16
2003	53	-3	-159	9
2004	55	-2	-110	4
2005	60	-1	-60	1
2006	70	0	0	0
2007	80	1	80	1
2008	99	2	198	4
2009	116	3	348	9
2010	127	4	508	16
2011	120	5	600	25
	ΣY=857	ΣX=0	ΣXY=1062	ΣX_i^2=110

ΣX=0 olduğundan, sadeleştirilmiş normal denklemler kullanılarak **"a"** ve **"b"** katsayıları aşağıdaki gibi hesaplanır.

$$a = \frac{\sum Y}{n} = \frac{857}{11} = 77,91 \quad ve \quad b = \frac{\sum XY}{\sum X^2} = \frac{1062}{110} =$$

Buna göre doğrusal trend denklemi: $\hat{Y}$ = **77,91+ 9,65 X** şeklinde yazılır.

b. Bulunan fonksiyon yardımıyla teorik trend değerlerini hesaplarken ilgili yılların X değerlerini bu denklemde sırasıyla yerine koymak yeterli olacaktır. Buna göre, zaman serisindeki yıllar için trend değerleri aşağıdaki gibi hesaplanır.

2001 için; $\hat{Y}_{01}$= 77,91 + 9,65 (-5) = 29,66
2002 için; $\hat{Y}_{02}$= 77,91 + 9,65 (-4) = 39,31
2003 için; $\hat{Y}_{03}$= 77,91 + 9,65 (-3) = 48,96
2004 için; $\hat{Y}_{04}$= 77,91 + 9,65 (-2) = 58,61
2005 için; $\hat{Y}_{05}$= 77,91 + 9,65 (-1) = 68,26
2006 için; $\hat{Y}_{06}$= 77,91 + 9,65 (0) = 77,91
2007 için; $\hat{Y}_{07}$= 77,91 + 9,65 (1) = 87,56
2008 için; $\hat{Y}_{08}$= 77,91 + 9,65 (2) = 97,21
2009 için; $\hat{Y}_{09}$= 77,91 + 9,65 (3) = 106,86
2010 için; $\hat{Y}_{10}$= 77,91 + 9,65 (4) = 116,51
2011 için; $\hat{Y}_{11}$= 77,91 + 9,65 (5) = $\underline{126,15}$
857,00

Teorik trend değerleri nasıl hesaplanır?

SIRA SİZDE 7

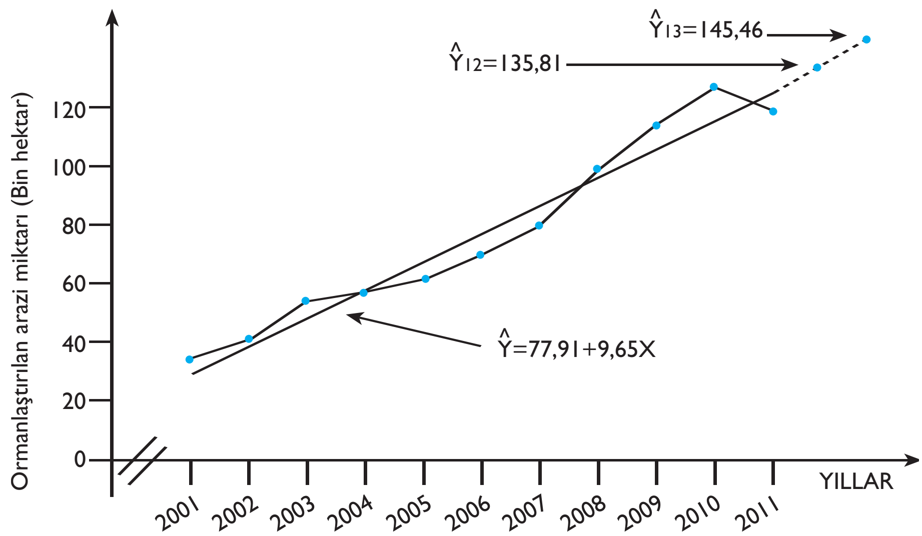
Hesaplanan teorik trend değerleri ile gerçek gözlem değerleri her bir yıl için genellikle farklı olmakla birlikte, bunların toplamı daima birbirine eşittir.

ΣY = 857 = Σ$\hat{Y}$= 857

Hesaplanan teorik trend değerlerine göre çizilen trend doğrusu Şekil 6.9'da gösterilmiştir.

Şekil 6.9

Örn. 4'teki zaman serisinin ve $\hat{Y}$ = 77,91+ 9,65 X trend doğrusunun grafiği.



c. Bulunan doğrusal trend denkleminden ($\hat{Y}$ = 77,91 + 9,65 X) yararlanarak 2012 (*bir sonraki yıl*) teorik değerini öngörmek için, denklemde X yerine X=6 (*bir sonraki X değeri*) konması gerekir. Böylece 2012 yılı öngörüsü;

$\hat{Y}_{12}$= 77,91+ 9,65 **(6)** = ₺135,81 milyon olarak bulunur.

Bu durumda bisiklet üretim işletmesi önümüzdeki yıl, yani 2012 yılında ₺135,81 milyon'lik satış gelirine (Şekil 6.9) ulaşabilecek demektir.
Benzer şekilde 2013 yılı öngörü değeri de: $\hat{Y}_{13}$= 77,91 + 9,65 **(7)** = ₺145,46 milyon'lik satış geliri olarak öngörülür (Şekil 6.9).

Görüldüğü üzere bulduğumuz trend denkleminde ($\hat{Y}$ = 77,91+ 9,65 X) gelecek her yıla karşı gelecek X değeri, denklemde yerine konularak geleceğe dönük öngörüler yapılabilir.

Şimdi de gözlem sayısı çift yıl olan bir zaman serisi için trend analizi çalışması yapalım.

ÖRNEK 5

Aşağıda 2002-2011 yılları arasında demir cevheri üretim miktarları verilmiştir. Bu zaman serisinin;

a. *En küçük kareler tekniği ile doğrusal trend denklemini bulunuz.*

b. *Bulunan trend denklemi yardımıyla, teorik trend değerlerini hesaplayarak, zaman serisi grafiğinde gösteriniz.*

c. *2012 yılı demir cevheri üretim miktarını öngörünüz.*

Yıllar	*Demir cevheri üretim mik. (milyon ton) (Y_i)*
2002	*20*
2003	*40*
2004	*50*
2005	*90*
2006	*110*
2007	*150*
2008	*130*
2009	*110*
2010	*170*
2011	*190*

Çözüm:

a. Demir cevheri üretimiyle ilgili olarak 10 yıllık bir zaman serisi verilmiştir. Bir başka ifade ile yıllara göre demir cevheri üretim miktarlarından oluşan bir ana kütleden 10 yıllık bir kesit, örnek olarak alınmıştır. Serinin grafiğine (Şekil 6.10) bakıldığında, konjonktürün etkisiyle dalgalanmalar görülmekle birlikte, genel eğilim üretim miktarının uzun dönemde arttığı şeklinde görülmektedir. Bu nedenle, zaman serisinin trendi için en uygun fonksiyon, doğrusal fonksiyondur: $\hat{Y} = a + bX$

Burada "a" ve "b" yi kısaltılmış normal denklemler yardımıyla bulabilmek için zaman serisinde bağımsız değişken X değerlerinin belirlenmesi gerekir. Gözlem sayısı n=10, *çift sayı* olduğundan serinin tam ortasına gelen iki yıl birlikte ele alınarak birincisine X=-1, ikincisine X=+1 değerleri verilerek ikişer ikişer azaltılıp, artırılacaktır.

Yıllar	Y_i	X_i	XY	X_i^2
2002	20	-9	-180	81
2003	40	-7	-280	49
2004	50	-5	-250	25
2005	90	-3	-270	9
2006	110	-1	-110	1
2007	150	1	150	1
2008	130	3	390	9
2009	110	5	550	25
2010	170	7	1190	49
2011	190	9	1710	81
	$\Sigma Y=1060$	$\Sigma X=0$	$\Sigma XY=2900$	$\Sigma X_i^2=330$

ΣX=0 olduğundan, kısaltılmış normal denklemler kullanılarak **"a"** ve **"b"** sabitleri aşağıdaki gibi hesaplanır.

$$a = \frac{\sum Y}{n} = \frac{1060}{10} = 106 \quad ve \quad b = \frac{\sum XY}{\sum X^2} = \frac{2900}{330} = 8,79$$

Buna göre doğrusal trend denklemi: $\hat{\mathbf{Y}}$ **= 106 + 8,79 X** şeklinde yazılır.

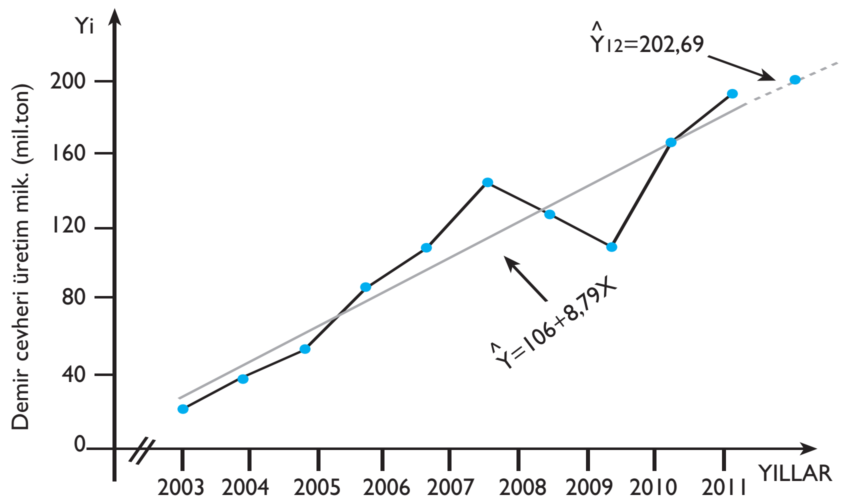
b. Her bir yıla karşılık gelen X değerleri bu fonksiyonda yerine konarak teorik $\hat{Y}$ değerleri aşağıdaki gibi hesaplanır.

2002 için;	$\hat{Y}_{02}$= 106 + 8,79 (-9) = 26,89
2003 için;	$\hat{Y}_{03}$= 106 + 8,79 (-7) = 44,47
2004 için;	$\hat{Y}_{04}$= 106 + 8,79 (-5) = 62,05
2005 için;	$\hat{Y}_{05}$= 106 + 8,79 (-3) = 79,63
2006 için;	$\hat{Y}_{06}$= 106 + 8,79 (-1) = 97,21
2007 için;	$\hat{Y}_{07}$= 106 + 8,79 (1) = 114,79
2008 için;	$\hat{Y}_{08}$= 106 + 8,79 (3) = 132,37
2009 için;	$\hat{Y}_{09}$= 106 + 8,79 (5) = 149,95
2010 için;	$\hat{Y}_{10}$= 106 + 8,79 (7) = 167,53
2011 için;	$\hat{Y}_{11}$= 106 + 8,79 (9) = $\underline{185,11}$
	$\Sigma\hat{Y}$=1060,00

En küçük kareler metodu ile bulunan trend doğrusu, zaman serisinin grafiği üzerinde Şekil 6.10'da görülmektedir.

Şekil 6.10

Örnek 5'teki zaman serisinin ve $\hat{Y}$= 106 + 8,79 X trend doğrusunun grafiği



c. Bulunan trend denkleminden 2012 yılının demir cevheri üretim miktarı öngörülürken denklemdeki X yerine, son X değerinin *iki fazlası* olan 11 değeri konarak öngörülür. Böylece 2012 yılı için demir cevheri üretim miktarı;

$\hat{Y}_{2012}$= 106 + 8,79 (11) = 202,69 milyon ton olarak öngörülür.

Zaman serisinde, trend denkleminin belirlenmesinde temel amaç, ileriye dönük öngörü yapmaktır. Bunun yanında öngörüler ileriye olmasının yanında geriye de (bazen), hatta zaman serisinin içindeki boşlukları doldurmak için içeriye de yapılabilir. Değeri öngörülecek yıl zaman serisinin içinde ise yapılacak olan öngörüye *"enterpolasyon"* denir. Buna karşın değeri öngörülecek yıl, incelenen zaman serisinin öncesi veya sonrasında herhangi bir yıl ise yapılan öngörüye *"ekstrapolasyon"* adı verilir.

Standart Hata

Bir zaman serisinin, en küçük kareler tekniği ile bulunan trend doğrusu, genellikle gözlem değerlerinin (serpilme noktalarının) arasından geçer. Zaman serisinde gözlem değerleri toplamı ile teorik trend değerleri toplamı birbirine eşit olmakla ($\Sigma Y_i = \Sigma \hat{Y}_i$) birlikte, yıllık değerleri çoğu kez birbirinden farklılık ($Y_i \neq \hat{Y}_i$) göstermektedir. Bu farklar, trend denklemi ile yapılan öngörünün hatasını gösterir. Bu hatalara *"Öngörü Hataları"* adı verilir ve öngörü hatalarının ortalama ölçüsü *"Öngörülerin Standart Hatası"*, standart sapmaya benzer şekilde aşağıdaki formülle hesaplanır.

$$S_Y = \sqrt{\frac{\sum\left(Y - \hat{Y}\right)^2}{n}}$$

Verilen standart hata-$\mathbf{S_Y}$ formülü gözlem sayısı $\mathbf{n \geq 30}$ olması hâlinde geçerlidir. Ancak zaman serisi analizinde çoğunlukla 30'dan az veri ile analiz yapılmaktadır. Bu nedenle $\mathbf{n < 30}$ olması hâlinde öngörülerin standart hatası;

$$S_Y = \sqrt{\frac{\sum\left(Y - \hat{Y}\right)^2}{n-2}}$$ formülünden yararlanılarak hesaplanır.

Fonksiyon tipi seçilirken zaman serisinin grafiği çizilir ve bu grafiğin genel gidişinin bir doğruya mı yoksa eğriye mi uyduğuna bakılır. Ancak bu yolla her zaman açık ve kesin bir sonuca varmak olanaksızdır. Bu gibi durumlarda isabetli karar vermeyi kolaylaştırmak için farklı fonksiyonlar için ayrı ayrı hesaplanan *öngörülerin standart hatalarından* en küçük standart hataya sahip fonksiyon, uygun fonksiyon olarak belirlenir.

ÖRNEK 6

Bir önceki örnekteki (Örnek 5.) demir cevheri verileri için tahminlerin standart hatasını hesaplayalım.

Y_i	$\hat{Y}$	$(Y_i-\hat{Y})$	$(Y_i-\hat{Y})^2$
20	26,89	-6,89	47,47
40	44,47	-4,47	19,98
50	62,05	-12,05	145,20
90	79,63	10,37	107,54
110	97,21	12,79	163,58
150	114,79	35,21	1239,74
130	132,37	-2,37	5,62
110	149,95	-39,95	1596,00
170	167,53	2,47	6,10
190	185,11	4,89	23,91
1060	1060		3355,14

$$S_Y = \sqrt{\frac{\sum\left(Y-\widehat{Y}\right)^2}{n-2}} = \sqrt{\frac{3355,14}{10-2}} = \sqrt{419,39} = 20,48 \quad \text{milyon ton.}$$

Bu sonuca göre gözlem değerleri ile teorik trend değerleri arasında ortalama 20,48 milyon tonluk farklılık vardır. Bir başka ifade ile ilgili zaman serisi doğrusal trend denklemi ile temsil edilerek, ortalama 20,48 milyon ton hata yapılmıştır.

SIRA SİZDE 8

Zaman serilerinde öngörü hatalarının ortalama ölçüsü nedir?

Özet

AMAÇ 1

Zaman serisini tanımlamak

İstatistiksel bir olayda, gözlem değerlerinin zaman değişkeninin konumlarına göre aldığı değerlerin sıralanmasıyla elde edilen verilere "zaman serisi" denir. Zaman serilerinde zaman birimi, yıl olabileceği gibi yılın alt zaman birimleri olan ay, hafta veya gün, hatta saat bile olabilir. Bir zaman serisi iki sütundan oluşmaktadır. Sütunların birinde daima zaman değişkeni, diğerinde ise zaman bağlı olarak değişen, gözlem değerleri yazılır. Zaman serisi analizi ise herhangi bir zaman serisine düzensiz görünüm veren dalgalanma veya hareketlerin neden kaynaklandığını bularak zaman serisini bileşenlerine ayırmak, bunların gelecekte alacakları değerleri öngörmek ve bileşenleri birleştirerek belirli bir öngörü değerine ulaşmakla ilgilidir.

AMAÇ 2

Zaman serisini etkileyen faktörleri açıklamak

Zaman serisi analizinde, gözlem değerlerinde meydana gelen dalgalanmaların dört faktörün etkisinden kaynaklandığı varsayılmaktadır. Bu dört faktör, "Trend, Mevsimlik Dalgalanmalar, Konjonktürel Dalgalanmalar ve Düzensiz (Tesadüfi) dalgalanmalar" olarak sayılabilir.

AMAÇ 3

Zaman serisini etkileyen ana faktörü tanımlamak

Zaman serisinin uzun dönemde belli bir yöne doğru gösterdiği ana eğilim trend olarak tanımlanır. Trend analizi ise zaman serisi, üzerinde etkili olan diğer faktörlerin (K ve D) etkisinden arındırarak, trendi (ana eğilimi) ortaya çıkarmaktır. Trend analizinde çeşitli tekniklerden yararlanılmaktadır. Bunlardan en çok bilinen ve yaygın kullanıma sahip olan; Hareketli Ortalamalar Tekniği ve En Küçük Kareler Tekniği'dir.

AMAÇ 4

Hareketli ortalamalar tekniği ile trendi araştırmak

Hareketli Ortalamalar Tekniği, zaman serisinin grafiği çizildiğinde açıkça görülen konjonktürel ve mevsimlik dalgalanmaların etkisini ortadan kaldırmak amacıyla kullanılır. Hareketli ortalamalar tekniği zaman serisi boyunca hareket eden aritmetik ortalamaya dayanır. Zaman serisinin değerleri belirli kümeler hâlinde toplanır ve her bir küme için aritmetik ortalama hesaplanır.

AMAÇ 5

En küçük kareler tekniği ile doğrusal trend analizi yapmak

Zaman serileri analizinde, zaman değişkeni ile gözlem değerleri arasındaki fonksiyonel ilişki en küçük kareler tekniği ile araştırılır. Bu fonksiyonel ilişki doğrusal olabileceği gibi eğrisel de olabilir.

AMAÇ 6

Doğrusal trend denklemi ile öngörüler yapmak

Zaman serisinde, trend denkleminin belirlenmesinde temel amaç, ileriye dönük öngörü yapmaktır. Doğrusal trend denklemi $\hat{Y} = a + bX$ şeklinde yazılır ve "a" ve "b" değerleri normal denklemlerle hesaplanır. Bu denklemde her yıla karşı gelecek X değeri, denklemde yerine konularak geleceğe dönük öngörüler yapılabilir.

Kendimizi Sınayalım

1. Aşağıdakilerden hangisinde toplanacak veriler "zaman serisi" **oluşturmaz?**

a. Yıllara göre ithalat miktarı
b. Yıllara göre ithalat tutarı
c. Aylara göre üretim miktarı
d. Hanelere göre harcama miktarı
e. Saat başı trafik yoğunluğu

2. Aşağıdaki tüketim ürünlerinin hangisinin tüketimi üzerinde mevsim etkisi **görülmez?**

a. Dondurma
b. Yakıt (gaz, kömür, vb.)
c. Un
d. Grip ilacı
e. Ayakkabı

3. En küçük kareler tekniğinde, "en küçük" yapılmaya çalışılan aşağıdakilerin hangisidir?

a. $\sum(Y-\hat{Y})$
b. $\sum(Y-\hat{Y})^2$
c. $\sum Y_i - \sum \hat{Y}_i$
d. $Y_i - \hat{Y}_i$
e. $\overline{Y} - \hat{Y}$

4. Trend analizi: zaman değişkenine ile bağımlı değişen arasındaki ilişkinin fonksiyonel şeklini ortaya koymaya çalışır. Bu tanıma göre aşağıdakilerden hangisi trend fonksiyonu **olamaz?**

a. $\hat{Y} = a + bX$
b. $\hat{Y} = a + bX + cZ$
c. $\hat{Y} = a + bX + cX^2$
d. $\hat{Y} = ab^X$
e. $\hat{Y} = a + bX + cX^2 + dX^3$

5 ve 6. Soruları aşağıdaki zaman serisine göre cevaplayınız.

Yıllar	Satışlar (kg)
2007	4
2008	6
2009	7
2010	11
2011	17

5. Verilen zaman serisinin doğrusal trend denklemi aşağıdakilerin hangisinde doğru yazılmıştır?

a. $\hat{Y} = 9 + 3{,}1X$
b. $\hat{Y} = 9{,}1 + 3X$
c. $\hat{Y} = 10 + 3{,}1X$
d. $\hat{Y} = 10{,}1 + 3X$
e. $\hat{Y} = 9 + 3X$

6. Doğrusal trend denklemi yardımıyla 2012 yılı satış miktarının öngörüsü nedir?

a. $\hat{Y}_{12} = 17{,}1$
b. $\hat{Y}_{12} = 17{,}3$
c. $\hat{Y}_{12} = 18$
d. $\hat{Y}_{12} = 18{,}1$
e. $\hat{Y}_{12} = 18{,}3$

7, 8 ve 9. Soruları aşağıdaki zaman serisine göre cevaplayınız.

Yıllar	Satışlar (kg)
2008	3
2009	9
2010	11
2011	21

7. Verilen zaman serisinin doğrusal trend denklemi aşağıdakilerin hangisinde doğru yazılmıştır?

a. $\hat{Y} = 11 + 2{,}8X$
b. $\hat{Y} = 12 + 2{,}8X$
c. $\hat{Y} = 11 + 3{,}1X$
d. $\hat{Y} = 11 + 3{,}8X$
e. $\hat{Y} = 12 + 3{,}8X$

8. Doğrusal trend denklemi yardımıyla 2012 yılı satış miktarının öngörü değeri nedir?

a. $\hat{Y}_{12} = 21$
b. $\hat{Y}_{12} = 22$
c. $\hat{Y}_{12} = 23$
d. $\hat{Y}_{12} = 24$
e. $\hat{Y}_{12} = 25$

9. Verilen seri için “tahminlerin standart hatası” aşağıdakilerin hangisinde doğru hesaplanmıştır?

a. S_Y = 1,27
b. S_Y = 1,37
c. S_Y = 2,27
d. S_Y = 2,37
e. S_Y = 3.37

10. Aşağıda verilen zaman serisi için 3'erli hareketli ortalamalar hesaplandığında 2006 yılı trend değeri kaçtır?

Yıllar	Üretim (metre)
2005	1500
2006	2000
2007	2800
2008	1900
“	”

a. $\hat{Y}_{12}$= 2000
b. $\hat{Y}_{12}$= 2050
c. $\hat{Y}_{12}$= 2100
d. $\hat{Y}_{12}$= 2200
e. $\hat{Y}_{12}$= 2300

Kendimizi Sınayalım Yanıt Anahtarı

1. d	Yanıtınız yanlış ise “Giriş” konusunu yeniden gözden geçiriniz.
2. c	Yanıtınız yanlış ise “Mevsimlik Dalgalanmalar” konusunu yeniden gözden geçiriniz.
3. b	Yanıtınız yanlış ise “En Küçük Kareler Tekniği” konusunu yeniden gözden geçiriniz.
4. b	Yanıtınız yanlış ise “Trend Analizi” konusunu yeniden gözden geçiriniz.
5. a	Yanıtınız yanlış ise “Trend Analizi” konusunu yeniden gözden geçiriniz.
6. e	Yanıtınız yanlış ise “Trend Analizi” konusunu yeniden gözden geçiriniz.
7. a	Yanıtınız yanlış ise “Trend Analizi” konusunu yeniden gözden geçiriniz.
8. e	Yanıtınız yanlış ise “Trend Analizi” konusunu yeniden gözden geçiriniz.
9. d	Yanıtınız yanlış ise “Standart Hata” konusunu yeniden gözden geçiriniz.
10. c	Yanıtınız yanlış ise “Hareketli Ortalamalar” konusunu yeniden gözden geçiriniz.

Sıra Sizde Yanıt Anahtarı

Sıra Sizde 1
Zaman serisi analizi, herhangi bir zaman serisine düzensiz görünüm veren dalgalanma veya hareketlerin neden kaynaklandığını bularak zaman serisini bileşenlerine ayırmak, bunların gelecekte alacakları değerleri öngörmek ve bileşenleri birleştirerek belirli bir öngörü değerine ulaşmakla ilgilidir.

Sıra Sizde 2
Zaman serisinin grafiği çizilirken zaman değişkeni (Yıl, Ay vb.) yatay eksende, gözlem değerleri (Yi) değişkeni dikey eksende işaretlenir. Bu iki değişkenin koordinat diyagramında kesişme noktaları işaretlenir. İşaretlenen noktalar doğru parçaları ile birleştirilirse zaman serisinin grafiği çizilmiş olur.

Sıra Sizde 3
Yıllık bir zaman serisini etkileyen faktörler trend, konjonktürel dalgalanmalar ve düzensiz dalgalanmalar olmak üzere üç tanedir. Çünkü gözlem değerleri yıllık verilerden oluşuyorsa mevsimlik dalgalanmaların etkisini taşımayacaktır.

Sıra Sizde 4
Zaman serisinin trend doğrusu veya eğrisi etrafındaki uzun dönem dalgalanmalarına konjonktürel dalgalanmalar denir. İş dünyasında yatırım artışlarının, üretim artışlarına ve üretim artışlarının gelir artışlarına yol açmasıyla ekonomik durumda bir süreliğine bir gelişme görülür. Gelişmenin maksimum aşamasında bir kriz ortaya çıkar. Sonra bir düşüş başlar. Daha sonra işlerde yeniden bir kımıldanma ve canlanma görülmeye başlar. Bu aşamalar 3-15 yılda bir tekrarlanır gider.

Sıra Sizde 5
Trend analizi teknikleri; Hareketli Ortalamalar ve En Küçük Kareler Teknikleri'dir.

Sıra Sizde 6
Hareketli ortalamalar tekniği, zaman serisinin grafiği çizildiğinde açıkça görülen konjonktürel ve mevsimlik dalgalanmaların etkisini ortadan kaldırmak amacıyla kullanılır.

Sıra Sizde 7
Teorik trend değerleri hesaplanırken bulunan trend denkleminde X değerlerini sırasıyla yerine koyup hesaplama yapmak yeterlidir.

Sıra Sizde 8
Bir zaman serisinde gerçek gözlem değerleri ile teorik trend değerleri arasındaki farkın ortalaması standart hatadır. Dolayısıyla zaman serisini trend denklemi ile temsil etmenin ortalama hatası standart hatadır.

Yararlanılan ve Başvurulabilecek Kaynaklar

Atlas, M. (2007), **İstatistik II (Çözümlü Örnekler),** Ak Ofset, Eskişehir.

Çömlekçi, N. (1998), **Temel İstatistik İlke ve Teknikleri** (Gözden geçirilmiş 3. Baskı), Bilim Teknik Yayınevi, İstanbul.

Gürsakal, N. (2009), **Çıkarımsal İstatistik** (4. Baskı), Dora Yayıncılık, Bursa.

Kazmier, L.J. and Pohl, N.F. (1987), **Basic Statistics for Business and Economics,** McGraw-Hill Book Co., New York.

Mann, P.S. (1995), **Statistics for Business and Economics,** J. Wiley, New York.

Newbold, P. (2001), **İşletme ve İktisat için İstatistik,** Çev: Ümit Şenesen, Literatür Yayıncılık, İstanbul.

Özmen, A. (2008), **İstatistik (5. Basım)** Editör: Ali Fuat Yüzer, Anadolu Üniversitesi Yayın No: 1448, Açıköğretim Fakültesi Yayın No: 771, Eskişehir.

Render, B. and Stair, R.M.Jr. (1988), **Qantative Analysis for Management** (Third Edition), Allyn and Bacon Inc., Boston.

Serper, Ö. (2004), **Uygulamalı İstatistik 2 (Genişletilmiş 5. Baskı),** Ezgi Kitapevi, Bursa.

Taha, H. (2000). **Operations Research an Introduction,** (6.Basımdan Çeviri: Yöneylem Araştırması) Çeviren ve Uyarlayan: Ş.Alp Baray ve Şakir Esnaf, Literatür Yayınları, İstanbul.

Turban, E. and Meredity, J.R., 1988, **Fundamentals of Management Science,** Fourth Edition, Business Publications, Inc., Plano Texas.

7

Amaçlarımız

Bu üniteyi tamamladıktan sonra;

- İndeks kavramını açıklayıp, indeks sayısı okuyabilecek,
- Basit fiyat (miktar) indeksi hesaplayıp yorumlayabilecek,
- Bileşik indeks hesaplama tekniklerini açıklayabilecek,
- Enflasyon hesabında Laspeyres fiyat indeksi hesabını açıklayabilecek,
- Başlıca fiyat indekslerini açıklayabilecek bilgi ve becerilere sahip olacaksınız.

Anahtar Kavramlar

- İndeks
- Basit İndeks
- Fiyat İndeksi
- Temel Devre
- Miktar İndeksi
- Bileşik İndeks
- İndeks Ortalama
- Ortalamalar İndeksi
- Laspeyres İndeksi
- Paasche İndeksi
- Enflasyon
- Fisher İndeksi
- Mal Sepeti
- ÜFE
- TÜFE

İçindekiler

İstatistik-II → İndeksler →

- GİRİŞ
- BASİT İNDEKSLER
- BİLEŞİK İNDEKSLER

İndeksler

GİRİŞ

Belli bir değişkenin veya değişkenler grubunun, zaman veya mekân içinde ortaya çıkan değişimlerini göstermek amacıyla hesaplanan, oransal ölçülere **indeks** (indeks sayısı) denir. İndeks bir başka ifadeyle kıyaslamayı kolaylaştıran oransal bir ölçüdür. İndeksler verilen bir dönemin, başka bir dönemle karşılaştırılmasını sağlayacak olan fiyatların, miktarların veya tutarların (fiyat ile miktar çarpımı) genellikle yüzdeler (%) şeklinde ifade edilmiş oranlarıdır.

İndeks, fiyatın, miktarın veya birlikte değişim değerinin iki ayrı zaman noktasındaki karşılaştırılmasında kullanılan oransal ölçüdür.

İndeksler; değişkenliğe neden olan faktörlere, temel devreye, ölçülen değişkene, indeksin kapsadığı mal ve hizmet (madde) sayısına göre farklı açılardan sınıflanır:

i. İndeksler zaman serilerinde ve mekân serilerinde hesaplanabilir. Ancak mekân indeksleri yaygın olarak kullanılmamaktadır. Bu nedenle bundan sonraki açıklamalar zaman indeksleriyle ilgili olacaktır.

ii. İndeks kavramı ilk olarak fiyatlar için kullanılmış, böylece indeks paranın satın alma gücündeki değişmelerin ölçme aracı olmuştur. Bu nedenle bundan sonraki açıklamalar daha çok fiyat indeksleriyle ilgili olacaktır.

KİTAP

İndekslerin sınıflanması ile ilgili olarak Necmi GÜRSAKAL'ın Betimsel İstatistik (Dora-Basım Yayım, Bursa-2010) kitabında bilgi bulabilirsiniz.

İndeksler her zaman iki sayı ile hesaplanır. Bunlardan biri *kıyaslanan*, diğeri değişimin kıyaslanacağı *temel değer*dir. Değişimi araştırılacak sayıya *"cari değer"*, karşılaştırılacak değere de *"temel değer"* adı verilir. Temel değer paydaya, kıyaslanan değer de paya yazılır. Oransal kıyaslamayı kolaylaştırmak için bölme işleminin sonucu *100* (bazen 1000) ile çarpılır. Böylece temel değer %100 kabul edilerek diğerlerinin buna göre yüzde "kaç" değişim gösterdiği hesaplanmış olur. Elde edilen sonuçlar *"yüzde (%)"* şeklinde sunulur.

Günümüzde indeksler çok yaygın uygulama alanına sahiptir. Özellikle üretim, tüketim, dış ticaret, para ve kredi, ücret ve toplu sözleşmeler, fiyat hareketlerinin analizi ve yorumu için indeksler kullanılmaktadır. Örneğin; Türkiye İstatistik Kurumu (TUİK) tarafından hesaplanan, her ayın 3'ü ile 5'i arasında yayımlanan fiyat indeksleri (enflasyon oranları) ve İstanbul Menkul Kıymetler Borsası (İMKB)'ında işlem gören hisse senedi fiyat indeksleri kamuoyunun bildiği, gelişmesini izlediği ve yorumlarının yapıldığı indekslerdir.

İndeks hesabında; tek bir malın fiyatında veya miktarında meydana gelen oransal değişimlerin belirlenmesi söz konusu ise *basit indeks*, birden çok malın fiyatında birlikte değişim veya fiyat ve miktarlarındaki birlikte değişim halinde *bileşik indeks* söz konusudur.

SIRA SİZDE 1

Basit ve bileşik indeks arasındaki farkı açıklayınız.

BASİT İNDEKSLER

Basit indeks, tek bir malın fiyatında miktarında ve değer tutarında (fiyat*miktar) zaman içinde meydana gelen oransal değişmeyi gösteren bir ölçüdür. Böylece bir malın temel alınan yıla (aya) göre fiyatında (miktarında) meydana gelen oransal değişme belirlenir.

Genel olarak basit indeks hesabında aşağıdaki adımlar izlenir.

- **Adım 1:** Araştırmaya konu olan malın *fiyatıyla (miktarıyla)* ilgili zaman serisi elde edilir.
- **Adım 2:** Temel devre belirlenir.
- **Adım 3:** Her bir devre için indeks değeri aşağıdaki formülle hesaplanır.

Basit indeks formülü; $\dot{I}_{i/0} = \dfrac{X_i}{X_0} \times 100$ şeklinde yazılır.

Formülde; $\mathbf{X_0}$**:** Temel devre değerini (fiyat veya miktar),

$\mathbf{X_i}$**:** Diğer devrelere ait değerleri, (i=0,1,2,3,...,n) temsil etmektedir.

Basit fiyat indeksi; İndeks hesabında zaman serisi bir malın yıllara göre *fiyatından* oluşuyor ise verilen formülde X yerine *"p (=fiyat)"* konularak *"basit fiyat indeksi":*

$$\dot{I}_{i/0} = \frac{p_i}{p_0} \times 100 \quad \text{formülü ile hesaplanır.}$$

Basit miktar indeksi; İndeks hesabında zaman serisi bir malın yıllara göre *miktarından* oluşuyor ise, verilen formülde X yerine *"q (=miktar)"* konularak *"basit miktar indeksi":*

$$\dot{I}_{i/0} = \frac{q_i}{q_0} \times 100 \quad \text{formülü ile hesaplanır.}$$

Formüllerden de anlaşılacağı gibi basit indeks hesaplanırken, herhangi bir devrenin değeri (fiyatı veya miktarı) temel devre değerine bölünmekte ve sonuç 100 ile çarpılmaktadır.

Sabit Esaslı İndeks (S.E.İ.)

Sabit esaslı indeks, zaman serisinde belirli bir devreyi temel alarak serinin bütün değerlerini bunun yüzdesi olarak göstermek suretiyle hesaplanan indekstir.

Sabit esaslı indeks hesabında *temel devre*nin seçimi çok önemlidir. Temel devrenin *enflasyon*, deflasyon (durgunluk) ve devalüasyon gibi iktisadi olayların aşırılık kazanmadığı, istikrarlı bir dönemden seçilmesi gerekir. Bu nedenle ekonomik kriz ve savaş yılları temel devre olarak seçilmez. Bunun yanında temel devrenin, hesaplama yapılan yıllara çok uzak olmaması gerekir. Uygulamada, temel devre olarak seçilecek istikrarlı bir yıl bulmak çoğu zaman zor olduğundan genelde *zaman serisini ilk yılı* temel devre olarak seçilir.

SIRA SİZDE 2

Sabit esaslı indeksde temel devre nasıl belirlenir?

Sabit esaslı fiyat ve miktar indeksleri aşağıdaki formüllerle hesaplanır.

Sabit Esaslı Fiyat İndeksi (SEFİ)	Sabit Esaslı Miktar İndeksi (SEMİ)
$\dot{I}_{i/0}=\frac{p_i}{p_0}\times 100$	$\dot{I}_{i/0}=\frac{q_i}{q_0}\times 100$

Formüllerdeki; p_0: Temel devre fiyatını, q_0: Temel devre miktarını
p_i: **i.** Devre fiyatını, q_i: **i.** Devre miktarını, temsil eder.

ÖRNEK 1

Aşağıdaki zaman serisinde yıllara göre Eskişehir'de uygulanmış kurşunsuz benzin (1-litre) fiyatları (₺) verilmiştir. 2006 yılını temel devre kabul ederek sabit esaslı fiyat indekslerini hesaplayalım.

Hesaplamalar 2006, 2007 ve 2008 yıları için gösterilirse:
Sabit esaslı fiyat indeksi;

$$2006\,\text{S.E.F.İ.:}\quad \dot{I}_{0/0}=\frac{p_0}{p_0}\times 100=\frac{2,84}{2,84}\times 100=\%100,00$$

$$2007\,\text{S.E.F.İ.:}\quad \dot{I}_{1/0}=\frac{p_1}{p_0}\times 100=\frac{3,06}{2,84}\times 100=\%107,75$$

$$2008\,\text{S.E.F.İ.:}\quad \dot{I}_{2/0}=\frac{p_2}{p_0}\times 100=\frac{2,76}{2,84}\times 100=\%97,18$$

Hesaplamalar bir bütün olarak aşağıdaki gibidir.

Yıllar	i (devre)	Kurşunsuz benzin fiyatı (p_i)	S.E.F.İ. (%)
2006	0	2,84	(2,84/2,84) ×100 = **100,00**
2007	1	3,06	(3,06/2,84) ×100 = **107,75**
2008	2	2,76	(2,76/2,84) ×100 = **97,18**
2009	3	3,66	(3.66/2,84) ×100 = **128,87**
2010	4	3,89	(3,89/2,84) ×100 = **136,97**
2011	5	4,32	(4,32/2,84) ×100 = **152,11**

Hesaplanan sabit esaslı fiyat indekslerine göre;

- Eskişehir'de 2007 yılı benzin fiyatı, 2006 benzin fiyatına göre %7,75 oranında artmıştır.
- Eskişehir'de 2008 yılı benzin fiyatı, 2006 benzin fiyatına göre %2,82 oranında *azalmıştır.*
- Eskişehir'de 2009 yılı benzin fiyatı, 2006 benzin fiyatına göre %28,87 oranında artmıştır.
- Eskişehir'de 2010 yılı benzin fiyatı, 2006 benzin fiyatına göre %36,97 oranında artmıştır.
- Eskişehir'de 2011 yılı benzin fiyatı, 2006 benzin fiyatına göre %52,11 oranında artmıştır.

Değişken Esaslı İndeks (D.E.İ.)

İndeks hesabında temel devre **sabit** kabul edilebileceği gibi değişken de kabul edilebilir. Eğer temel devre her bir indeks hesabı için değişiyorsa **"değişken esaslı indeks"** söz konusudur. Değişken esaslı indeks hesabında her değer, kendinden bir önceki değere oranlanır. Bu nedenle değişken esaslı indekse *"zincirleme indeks"* de denilmektedir.

İndeks hesabında kesrin paydası sabit kaldığında **"sabit esaslı indeks"**, değiştiğinde **"değişken esaslı indeks"** söz konusudur.

Değişken esaslı fiyat ve miktar indeksleri aşağıdaki formüllerle hesaplanır.

Değişken Esaslı Fiyat İndeksi (DEFİ)	Değişken Esaslı Miktar İndeksi (DEMİ)
$\dot{I}_{i/i-1} = \frac{p_i}{p_{i-1}} \times 100$	$\dot{I}_{i/i-1} = \frac{q_i}{q_{i-1}} \times 100$

Formüllerdeki; p_0: **i.** Devre fiyatını, q_i: **i.** Devre miktarını,
p_{i-1}: **i-1.** Devre fiyatını, q_{i-1}: **i-1.** Devre miktarını, temsil eder.

Değişken esaslı indeks formüllerinde **"i"** yerine **"1"** konulduğunda sabit esaslı indekse benzediği görülür. Bu nedenle **"0"** ve **"1"** nolu devrelerin sabit ve değişken esaslı indeksleri birbirine eşittir.

SIRA SİZDE 3

Değişken esaslı indeks nasıl hesaplanır?

ÖRNEK 2

Karşılaştırma kolaylığı sağlamak için Örnek 1'deki benzin fiyat serisini ele alalım ve değişken esaslı fiyat indekslerini hesaplayalım.

Hesaplamalar 2006, 2007 ve 2008 yılları için gösterilirse:
Değişken esaslı fiyat indeksi;

$$2006\,\text{D.E.F.İ.}: \dot{I}_{0/0} = \frac{p_0}{p_0} \times 100 = \frac{2,84}{2,84} \times 100 = \%100,00$$

$$2007\,\text{D.E.F.İ.}: \dot{I}_{1/0} = \frac{p_1}{p_0} \times 100 = \frac{3.06}{2,84} \times 100 = \%107,75$$

$$2008\,\text{D.E.F.İ.}: \dot{I}_{2/1} = \frac{p_2}{p_1} \times 100 = \frac{2,76}{3,06} \times 100 = \%90,20$$

Hesaplamalar bir bütün olarak aşağıdaki gibidir.

Yıllar	i (devre)	Kurşunsuz benzin fiyatı (p_i)	S.E.F.İ. (%)
2006	0	2,84	(2,84/2,84) ×100 = **100,00**
2007	1	3,06	(3,06/2,84) ×100 = **107,75**
2008	2	2,76	(2,76/3,06) ×100 = **90,20**
2009	3	3,66	(3,66/2,76) ×100 = **132,61**
2010	4	3,89	(3,89/3,66) ×100 = **106,28**
2011	5	4,32	(4,32/3,89) ×100 = **111,05**

DİKKAT

Benzin fiyatları için hesaplanan değişken esaslı indekse benzer şekilde, elde edilecek benzin tüketim miktarları serisi için de miktar indeksleri hesaplanarak yorumlanabilir.

Görülebileceği gibi ilk iki devrenin (0 ve 1. devrenin) sabit esaslı fiyat indeksleri (S.E.F.İ.) ile değişken esaslı fiyat indeksleri (D.E.F.İ.) bir birine eşittir. Hesaplanan değişken esaslı fiyat indekslerine göre;

- Eskişehir'de 2007 yılı benzin fiyatı, 2006 benzin fiyatına göre %7,75 oranında artmıştır.
- Eskişehir'de 2008 yılı benzin fiyatı, 2007 benzin fiyatına göre %9,80 oranında *azalmıştır.*
- Eskişehir'de 2009 yılı benzin fiyatı, 2008 benzin fiyatına göre %32,61 oranında artmıştır.

- Eskişehir'de 2010 yılı benzin fiyatı, 2009 benzin fiyatına göre %6,28 oranında artmıştır.
- Eskişehir'de 2011 yılı benzin fiyatı, 2010 benzin fiyatına göre %11,05 oranında artmıştır, denir.

İndekslerin Birinden Diğerine Geçiş

Aralarındaki matematiksel bağlantılar nedeniyle sabit ve değişken esaslı indekslerden birine ait indeks değerleri bilindiğinde, bunlardan yararlanarak diğeri hesaplanabilir.

i. Sabit esaslı indekslerin bilinmesi hâlinde, bir devrenin sabit esaslı indeksi bir önceki devrenin sabit esaslı indeksine oranlanır ve sonuç 100 ile çarpılarak, ilgili devrenin *değişken esaslı indeksine* ulaşılır.

Sabit esaslı indeksten, değişken esaslı indekse geçiş formülü aşağıdaki gibi yazılır.

S.E.İ. → D.E.İ. geçiş: $\text{İ}_{i/i-1} = \dfrac{\text{İ}_{i/0}}{\text{İ}_{i-1/0}} \times 100$

ÖRNEK 3

2006-2011 yılları arasında benzin fiyatları için hesaplanan sabit esaslı fiyat indekslerinden (S.E.F.İ.) yararlanarak değişken esaslı fiyat indekslerini (D.E.F.İ.) hesaplayalım.

Hesaplamalar 2007 ve 2008 yılları için gösterilirse:

2007 D.E.F.İ.= $\text{İ}_{1/0} = \dfrac{107,75}{100} \times 100 = \%107,75$

2008 D.E.F.İ.= $\text{İ}_{2/1} = \dfrac{97,18}{107,75} \times 100 = \%90,20$

Hesaplamalar bir bütün olarak aşağıdaki gibidir.

Yıllar	i	S.E.F.İ. (%)	Hesaplanan D.E.F.İ. (%)
2006	0	100,00	(100,00/100,00) ×100 = **100,00**
2007	1	107,75	(107,75/100,00) ×100 = **107,75**
2008	2	97,18	(97,18/107,75) ×100 = **90,20**
2009	3	128,87	(128,87/ 97,18) ×100 = **132,61**
2010	4	136,97	(136,97/128,87) ×100 = **106,28**
2011	5	152,11	(152,11/136,97) ×100 = **111,05**

Sabit esaslı fiyat indeksleri (S.E.F.İ.) yardımıyla hesaplanan değişken esaslı fiyat indeksleri (D.E.F.İ.), daha önce gerçek gözlem değerleri ile bulunan değişken esaslı fiyat indeksleri (D.E.F.İ.) ile aynıdır. Dolayısıyla yorumlanması da aynı şekilde olacaktır.

ii. Değişken esaslı indekslerin bilinmesi hâlinde, bir devrenin değişken esaslı indeksi ile bir önceki devrenin sabit esaslı indeksi çarpılır ve sonuç 100'e bölünerek ilgili devrenin *sabit esaslı indeksine* ulaşılmış olur.

Değişken esaslı indeksten, sabit esaslı indekse geçiş formülü aşağıdaki gibi yazılır.

D.E.İ. → S.E.İ. geçiş: $\text{İ}_{i/0} = \dfrac{\text{İ}_{i/i-1} \times \text{İ}_{i-1/0}}{100}$

ÖRNEK 4

2006-2011 yılları arasında benzin fiyatları için hesaplanan değişken esaslı fiyat indekslerinden (D.E.F.İ.) yararlanarak sabit esaslı fiyat indekslerini (S.E.F.İ.) hesaplayalım.

Hesaplamalar 2007 ve 2008 yılları için gösterilirse:

$$2007\,\text{S.E.F.İ.} = \quad \text{İ}_{1/0} = \frac{\text{İ}_{1/0} \times \text{İ}_{0/0}}{100} = \frac{107,75 \times 100}{100} = \%107,75$$

$$2008\,\text{S.E.F.İ.} = \quad \text{İ}_{2/0} = \frac{\text{İ}_{2/1} \times \text{İ}_{1/0}}{100} = \frac{90,20 \times 107,75}{100} = \%97,18$$

Hesaplamalar bir bütün olarak aşağıdaki gibidir.

Yıllar	i	D.E.F.İ. (%)	**Hesaplanan S.E.F.İ. (%)**
2006	0	100,00	(100,00×100,00)/100 = **100,00**
2007	1	107,75	(107,75×100,00)/100 = **107,75**
2008	2	90,20	(90,20×107,75)/100 = **97,18**
2009	3	132,61	(132,61×97,18)/100 = **128,87**
2010	4	106,28	(106,28×128,87)/100 = **136,97**
2011	5	111,05	(111,05×136,97)/100 = **152,11**

Değişken esaslı fiyat indeksleri (D.E.F.İ.) yardımıyla hesaplanan sabit esaslı fiyat indeksleri (S.E.F.İ.) daha önce gözlem değerleri yardımıyla bulunan sabit esaslı fiyat indeksleri (S.E.F.İ.) ile aynıdır. Dolayısıyla aynı şekilde yorumlanır.

SIRA SİZDE 4

Sabit ve değişken esaslı indekslerin birinden diğerine geçiş, neden yapılır?

BİLEŞİK İNDEKSLER

Basit indekslerde olduğu gibi, bileşik indekslerde, **sabit ve değişken esaslı** indeks olarak hesaplanabilir.

Uygulamada **bileşik indeks hesabı** için 50 ile 200 arasındaki mal ve hizmet sayısı yeterli olabilmektedir.

Günlük yaşamımızda çok sayıda mal ve hizmet üretilmekte ve tüketilmektedir. Bu nedenle tek malın fiyatı (miktarı) yerine çok sayıda malın fiyatı (miktarı) ile ilgilenebiliriz. İşte bileşik indeksler; birbiriyle ilişkili birden çok malın fiyatında ve/veya miktarında meydana gelen oransal değişmelerin belirlenmesinde kullanılmaktadır. Bileşik fiyat indeksinin, büyük gruplar hâlinde mal ve hizmetlerin genel fiyat seviyesini karşılaştırmak için hesaplanmaktadır. Söz konusu karşılaştırma, alınan devreye göre **"sabit esaslı"** veya **"değişken esaslı"** olabilir.

Bileşik indeks hesabına konu olacak mal ve hizmet sayısının yeterli düzeyde olması gerekir. Çünkü az sayıda mal ve hizmete dayanan indeks temsilî olamayacağı gibi, gereğinden çok mal ve hizmet olması da hesaplamayı zorlaştırır. Bütün mal ve hizmetlerin yerine, bu grubu en iyi temsil edebilecek tür ve sayıda mal ve hizmeti indeks hesabına dahil etmek en uygun yoldur.

Bileşik indeks hesabında kullanılan teknikler aşağıdaki gibi sıralanır.

i. İndeks ortalaması tekniği,
ii. Ortalamalar indeksi tekniği,
iii. Laspeyres ve Paasche indeksleri
iv. Fisher (İdeal) indeks tekniği,

Bileşik indeks hesabında amaç, indekse dahil edilecek bütün mal ve hizmetlerin eş zamanlı fiyatlarındaki (miktarlarındaki) gelişmeyi temsil edecek bir ortalama bulmak olduğundan, bu tür indekslerin hazırlanmasında *"ortalamalar"*dan yararlanılır. "İndeks ortalaması" ve "ortalamalar indeksi" tekniklerinin ortak yönü, elde edilen bileşik indekslerin *"tartısız bileşik indeks"* niteliğinde olmasıdır. "Laspeyres, Paasche" ve bunlara bağlı olarak hesaplanan "Fisher (İdeal)" İndeksleri ise *"tartılı bileşik indeks"* niteliğindedir.

SIRA SİZDE 5

Bileşik indeks tekniklerinden hangileri "tartılı" indekslerdir?

İndeks Ortalamaları Tekniği

Yıllara göre birden çok mal ve hizmetin fiyatlarındaki (miktarlarındaki) oransal değişme "indeks ortalama tekniği" ile hesaplanabilir. Bunun için öncelikle, her bir malın ayrı ayrı basit fiyat (miktar) indeksleri hesaplanır. Sonra, her bir yıl için hesaplanan bu indekslerin ortalamaları alınır. Bu ortalamalar "bileşik indeks" kabul edilir. Ortalamaların hesaplanmasında "aritmetik ortalama" ve "geometrik ortalama" kullanılır.

Hesaplanan bileşik indeks, sabit esaslı basit indekslerin aritmetik ortalaması ise *"sabit esaslı bileşik indeks (S.E.B.İ.)"*, değişken esaslı basit indekslerin aritmetik ortalaması ise *"değişken esaslı bileşik indeks (D.E.B.İ.)"* adını alır.

ÖRNEK 5

Aşağıda, yıllara göre üç kahvaltılık yiyecek maddesinin fiyatları verilmiştir. Bu üç malın fiyatlarındaki (₺/kg) oransal değişimi; sabit esaslı bileşik indekslerini "indeks ortalaması tekniği" ile hesaplayalım.

Yıllar	Siyah Zeytin p_i=Fiyat (₺/kg)	Beyaz Peynir p_i=Fiyat (₺/kg)	Reçel (Çilek) p_i=Fiyat (₺/kg)
2007	8,50	6,70	2,80
2008	10,50	7,30	3,10
2009	11,00	6,10	2,90
2010	12,40	9,80	3,50
2011	14,50	11,50	4,50

Öncelikle her bir malın sabit esaslı basit fiyat indeksleri:

$\text{İ}_{i/0} = \frac{p_i}{p_0} \times 100$ formülü kullanılarak hesaplanır.

Buradan hareketle üç kahvaltılık maddenin, sabit esaslı basit fiyat indekslerinin hesaplanmış ve aşağıda verilmiştir.

Yıllar	i	S.E.B.F.İ. (%)	S.E.B.F.İ. (%)	S.E.B.F.İ. (%)
2007	0	(8,5/8,5) ×100 = **100,00**	(6,7/6,7) ×100 = **100,00**	(2,8/2,8) ×100 = **100,00**
2008	1	(10,5/8,5) ×100 = **123,53**	(7,3/6,7) ×100 = **108,96**	(3,1/2,8) ×100 = **110,71**
2009	2	(11,0/8,5) ×100 = **129,41**	(6,1/6,7) ×100 = **91,04**	(2,9/2,8) ×100 = **103,57**
2010	3	(12,4/8,5) ×100 = **145,88**	(9,8/6,7) ×100 = **146,27**	(3,5/2,8) ×100 = **125,00**
2011	4	(14,5/8,5) ×100 = **170,59**	(11,5/6,7) ×100 = **171,64**	(4,5/2,8) ×100 = **160,71**

Her bir mal için hesaplanan basit indekslerin, her yıl için aritmetik ortalamaları hesaplanarak" sabit esaslı bileşik fiyat indeksleri" elde edilir.

Yıllar	Aritmetik ortalamaya göre **Sabit Esaslı Bileşik Fiyat İndeksleri S.E.B.F.İ. (%)**
2007 İndeksi:	$\text{İ}_{07} = \frac{100{,}00 + 100{,}00 + 100{,}00}{3} = \%100{,}00$
2008 İndeksi:	$\text{İ}_{08} = \frac{123{,}53 + 108{,}96 + 110{,}71}{3} = \%114{,}40$
2009 İndeksi:	$\text{İ}_{09} = \frac{129{,}41 + 91{,}04 + 103{,}57}{3} = \%108{,}01$
2010 İndeksi:	$\text{İ}_{10} = \frac{145{,}88 + 146{,}27 + 125{,}00}{3} = \%139{,}05$
2011 İndeksi:	$\text{İ}_{11} = \frac{170{,}59 + 171{,}64 + 160{,}71}{3} = \%167{,}65$

Söz konusu üç kahvaltılık malın fiyatlarında temel devre 2007'e göre "ortalama olarak" sırasıyla 2008'de %14,40; 2009'da %8,01; 2010'da %39,05 ve 2011'de %67,65 oranında artış görülmüştür.

ÖRNEK 6

Bir önceki örnekte (Örnek 5) verilen üç kahvaltılık yiyecek maddesinin bileşik indeksini birde temel devrenin değişken olmasına göre yani değişken esaslı olarak hesaplayalım.

Öncelikle her bir malın değişken esaslı basit fiyat indeksleri:

$$İ_{i/i-1} = \frac{p_i}{p_{i-1}}.100 \text{ formülü kullanılarak hesaplanır.}$$

Söz konusu üç kahvaltılık malın, değişken esaslı basit fiyat indekslerinin hesaplanmış hâli aşağıda verilmiştir. 2011

Yıllar	İ	D.E.B.F.İ. (%)	D.E.B.F.İ. (%)	D.E.B.F.İ. (%)
2007	0	(8,5/8,5) ×100 = **100,00**	(6,7/6,7) ×100 = **100,00**	(2,8/2,8) × 100 = **100,00**
2008	1	(10,5/8,5) ×100 = **123,53**	(7,3/6,7) ×100 = **108,96**	(3,1/2,8) × 100 = **110,71**
2009	2	(11,0/10,5) ×100 = **104,76**	(6,1/7,3) ×100 = **83,56**	(2,9/3,1) × 100 = **93,55**
2010	3	(12,4/11,0) ×100 = **112,73**	(9,8/6,1) ×100 = **160,66**	(3,5/2,9) × 100 = **120,69**
2011	4	(14,5/12,4) ×100 = **116,94**	(11,5/9,8) ×100 = **117,35**	(4,5/3,5) × 100 = **128,57**

Her bir madde için hesaplanan indekslerin her yıl için; aritmetik ortalaması hesaplanarak aritmetik ortalamaya göre değişken esaslı bileşik fiyat indeksine, ulaşılır.

Yıllar	Aritmetik ortalamaya göre **Değişken Esaslı Bileşik Fiyat İndeksleri D.E.B.F.İ. (%)**
2007 İndeksi:	$İ_{07} = \frac{100,00+100,00+100,00}{3} = \%100,00$
2008 İndeksi:	$İ_{08} = \frac{123,53+108,96+110,71}{3} = \%114,40$
2009 İndeksi:	$İ_{09} = \frac{104,76+83,56+93,55}{3} = \%93,96$
2010 İndeksi:	$İ_{10} = \frac{112,73+160,66+120,69}{3} = \%131,36$
2011 İndeksi:	$İ_{11} = \frac{116,94+117,35+128,57}{3} = \%120,95$

Değişken esaslı bileşik indekslere göre, üç kahvaltı maddesinin fiyatlarında *bir önceki yılın fiyatlarına* göre ortalama olarak; 2008 yılında %14,40 oranında artış, 2009 yılında %6,04 oranında *azalış*, 210 yılında %31,36 oranında artış ve 2011 yılında % 20,95 oranında artış görülmüştür. İndeks ortalaması ile bileşik *fiyat* indeksi üzerinde durduk. Benzer şekilde, bir miktar serisi için de bileşik *miktar* indeksleri hesaplanabilir.

İndeks ortalama tekniği, kapsadığı mal ve hizmetler arasında önem farkını hesaba katmadığından, fiyatlarında aşırı değişme olan mal ve hizmetlerin etkisi altında kalır. Örneğin, verilen üç kahvaltılık maddeden reçel, alışkanlığımız nedeniyle zeytin ve peynir kadar önemli değildir. Oysa hesaplamalarda, onlarla eşit önemde kabul edilmiştir. İndeks ortalama tekniği taşıdığı bu sakıncalar nedeniyle çoğunlukla tercih edilmez.

Ortalamalar İndeks Tekniği

Adından da anlaşılacağı gibi, burada öncelikle her bir yıl için *ortalama fiyat* (miktar) hesaplanır. Böylece her bir yıl için birden çok malın fiyatlarının (miktarlarının) ortalamasından oluşan yeni bir zaman serisi elde edilir. Ortalamaların hesaplanmasında aritmetik ortalama (geometrik ortalama) kullanılır. Ortalamalardan oluşan serinin, sabit veya değişken esaslı olarak hesaplanan indeksleri, bileşik indeks kabul edilir.

Ortalamalar indeksinde hesaplanan ortalama bileşik indeks hâline dönüştürülürken, sabit veya değişken esaslı olmasına göre "sabit esaslı bileşik indeks (S.E.B.İ.)" veya "değişken esaslı bileşik indeks (D.E.B.İ.)" adını alır.

ÖRNEK 7

Üç otomobil üreticisi firmanın yıllara göre üretim ***miktarları (adet)*** *aşağıda verilmiştir. Bu veriler için ortalamalar indeksi tekniği ile sabit esaslı bileşik indeksleri hesaplayalım.*

Yıllar	A-Firması q_i=üretim mik.	B-Firması q_i=üretim mik.	C-Firması q_i=üretim mik.
2007	6762	4313	1285
2008	6244	4256	1182
2009	7769	4934	1493
2010	8257	5584	2035
2011	9306	5550	2157

Öncelikle üç üretici firmanın her bir yıl için ortalama otomobil üretim miktarları hesaplanır.

$$\text{2007 Aritmetik Ortalaması} = \frac{6762+4313+1285}{3} = 4120 \text{ adet otomobil}$$

$$\text{2008 Aritmetik Ortalaması} = \frac{6244+4256+1182}{3} = 3894 \text{ adet otomobil}$$

$$\text{2009 Aritmetik Ortalaması} = \frac{7769+4934+1493}{3} = 4732 \text{ adet otomobil}$$

$$\text{2010 Aritmetik Ortalaması} = \frac{8257+5584+2035}{3} = 5292 \text{ adet otomobil}$$

$$\text{2011 Aritmetik Ortalaması} = \frac{9306+5550+2157}{3} = 5671 \text{ adet otomobil}$$

Ortalamalardan oluşan seri değerleri, 2007 yılı temel devre kabul edilerek sabit esaslı bileşik *miktar* indeksleri (S.E.B.M.İ.), basit indeks hesabında olduğu gibi hesaplanır.

Yıllar	Aritmetik ortalamaya göre **Sabit Esaslı Bileşik Miktar İndeksleri S.E.B.M.İ. (%)**
2007 İndeksi:	$\dot{I}_{07}=\dfrac{4120}{4120}\times 100=\%100,00$
2008 İndeksi:	$\dot{I}_{08}=\dfrac{3894}{4120}\times 100=\%94,51$
2009 İndeksi:	$\dot{I}_{09}=\dfrac{4732}{4120}\times 100=\%114,85$
2010 İndeksi:	$\dot{I}_{10}=\dfrac{5292}{4120}\times 100=\%128,45$
2011 İndeksi:	$\dot{I}_{11}=\dfrac{5671}{4120}\times 100=\%137,65$

Bu sonuçlara göre, otomobil üreticisi üç firmanın üretim miktarları, ortalama olarak 2007 yılına göre 2008'de %5,49 oranında *azalırken*, 2009'da %14,85 oranında 2010'da %28,45 oranında ve 2011'de %37,65 oranında *artış* göstermiştir.

ÖRNEK 8

Otomobil üreticisi firmaların yıllık ortalama üretim miktarlarından (Örnek 7) yararlanarak "sabit esaslı bileşik indeks" hesaplandığı gibi, "değişken esaslı bileşik indeks" de hesaplanabilir. Yıllara göre aritmetik ortalama ve değişken esaslı bileşik miktar indeksi (D.E.B.M.İ.) değerleri aşağıda verilmiştir.

Yıllar	İ	Aritmetik Ort.	Aritmetik ortalamaya göre **D.E.B.M.İ. (%)**
2007	0	4120	$\dot{I}_{07}=\dfrac{4120}{4120}\times 100=\%100,00$
2008	1	3894	$\dot{I}_{08}=\dfrac{3894}{4120}\times 100=\%94,51$
2009	2	4732	$\dot{I}_{09}=\dfrac{4732}{3894}\times 100=\%121,52$
2010	3	5292	$\dot{I}_{10}=\dfrac{5292}{4732}\times 100=\%111,83$
2011	4	5671	$\dot{I}_{11}=\dfrac{5671}{5292}\times 100=\%107,16$

Bu sonuçlara göre üç firmanın otomobil üretim miktarları ortalama olarak 2009, 2010 ve 2011yıllarında bir önceki yıllara göre sırasıyla %21,52; %11,83 ve %7,16 oranında artış göstermiştir. Buna karşın 2008 yılında, bir önceki yıla (2007) göre %5,49 oranında azalma göstermiştir.

SIRA SİZDE 6

İndeks ortalama tekniği ile ortalamalar indeks tekniği arasındaki farklılığı açıklayınız.

Laspeyres ve Paasche İndeksleri

"İndeks ortalama" ve "ortalamalar indeks" tekniklerinin ortak yönü "tartısız" olması yani kapsadığı maddelerin (malların veya hizmetlerin) önem farkını dikkate almamasıdır. Oysa günlük yaşamda her maddenin fiyatına (miktarına) aynı önemi vermediğimiz açıktır. Maddelerin önemi aynı olmadığından, bileşik indeks hesabında, bu önem farklarını dikkate alan tartıların kullanılması gereği ortaya çıkmaktadır. Tartılı indeks hesabında *Laspeyres* ve *Paasche indeksleri* kullanılır. Bu iki indeks, tartı olarak kullandıkları devre itibari ile birbirinden ayrılırlar. Laspeyres indeksinde temel devre değerleri tartı olarak kullanılır iken Paasche'de indeksi hesaplanan yıl değerleri tartı olarak kullanılmaktadır. Laspeyres ve Paasche indekslerinin her ikisi de fiyat ve miktardaki oransal değişmeyi ölçmek için kullanılabilir.

Laspeyres ve Paasche indeksleri arasındaki farlılık nedir?

SIRA SİZDE 7

Laspeyres İndeksi

Etienne Laspeyres tarafından 18. yüzyılın sonlarında geliştirilmiş bir teknik olan Laspeyres indeksi, fiyat ve miktar için hesaplanabilir. Laspeyres fiyat indeksi temel devre miktarını tartı olarak kullanırken Laspeyres miktar indeksi, temel devre fiyatını tartı olarak kullanmaktadır. Özellikle fiyatlardaki *enflasyonu* ölçmeye yarayan Laspeyres indeksi aşağıdaki formüllerle hesaplanır.

Laspeyres fiyat indeksi ($_p\dot{I}_L$)	Laspeyres miktar indeksi ($_q\dot{I}_L$)
$_p\dot{I}_L = \frac{\sum p_i \cdot q_0}{\sum p_0 q_0} \times 100$	$_q\dot{I}_L = \frac{\sum p_0 \cdot q_i}{\sum p_0 q_0} \times 100$

Bu formüllerde;

p_0 : Temel devre fiyatını,
q_0 : Temel devre miktarını,
p_i : i - Devresi fiyatını,
q_i : i - Devresi miktarını, temsil etmektedir.

Hemen ekleyelim ki, yukarıdaki formüllerde **"i=0"** uygulandığında sonucun %100'e eşit olacağı açıktır. Bunun anlamı, esas devrede Laspeyres fiyat ve miktar indekslerinin daima "%100"e eşit olduğudur.

ÖRNEK 9

Seçilen, dört "yer-altı zenginliğinin" 2009, 2010 ve 2011 birim satış fiyatları (₺) ve satış miktarları (Ton) aşağıda verilmiştir. Bu verilerden yararlanarak 2010 ve 2011 yılı Laspeyres fiyat indekslerini hesaplayalım.

Yıllar	i	Ham Petrol		Kömür		Demir		Bakır	
		Fiyat(p_i)	Mik.(q_i)	Fiyat(p_i)	Mik.(q_i)	Fiyat(p_i)	Mik.(q_i)	Fiyat(p_i)	Mik.(q_i)
2009	0	2	100	20	100	15	90	60	5
2010	1	3	108	24	110	20	85	65	6
2011	2	5	110	38	90	22	80	75	14

2009 yılı temel devre kabul edilerek (indeksi %100) 2010 ve 2011 yılı Laspeyres fiyat indeksi ($_p\dot{I}_L$), aşağıdaki gibi hesaplanır.

2009 indeksi, $$_p\dot{I}_L = \frac{\sum p_0 \times q_0}{\sum p_0 \times q_0} \times 100 = \%100,00 \quad \text{(daima)}$$

2010 indeksi, $_p\dot{I}_L = \dfrac{\sum p_1 \times q_0}{\sum p_0 \times q_0} \times 100$

$$= \frac{(3\times100)+(24\times100)+(20\times90)+(65\times5)}{(2\times100)+(20\times100)+(15\times90)+(60\times5)}\times100 = \%125,32$$

2011 indeksi, $_p\dot{I}_L = \dfrac{\sum p_2 \times q_0}{\sum p_0 \times q_0} \times 100$

$$= \frac{(5\times100)+(38\times100)+(22\times90)+(75\times5)}{3850}\times100 = \%172,86$$

Yıllar	i	Laspeyres Fiyat İndeksi-($_q\dot{I}_L$) (%)
2009	0	100,00
2010	1	125,32
2011	2	172,86

Bu sonuca göre, söz konusu dört yer-altı zenginliğinin fiyatlarında temel devre 2009'a göre *"ortalama olarak"* 2010'da %25,32 oranında ve 2011'de %72,86 oranında artış meydana gelmiştir.

Paasche İndeksi

Paasche fiyat indeksi, Laspeyres fiyat indeksinden sonra geliştirilmiştir. Paasche indeksi, Laspeyres'e benzemekle birlikte, tartısı farklılık gösterir. Paasche fiyat indeksinde, indeksi hesaplanacak yılın miktarı tartı olarak kullanılmaktadır. Paasche fiyat indeksine miktar tartı iken Paasche miktar indeksine de indeksi hesaplanacak yılın fiyatı tartı olarak kullanılır. Paasche fiyat ve miktar indeksleri, aşağıdaki formüllerle hesaplanabilir.

Paasche fiyat indeksi ($_p\dot{I}_P$)	Paasche miktar indeksi ($_q\dot{I}_P$)
$_p\dot{I}_P = \dfrac{\sum p_i \times q_i}{\sum p_0 \times q_i} \times 100$	$_q\dot{I}_P = \dfrac{\sum p_i \times q_i}{\sum p_i \times q_0} \times 100$

Bu formüllerde;

p_0 : Temel devre fiyatını, q_0 : Temel devre miktarını,

p_i : i. Devresi fiyatını, q_i : **i.** Devresi miktarını, temsil etmektedir.

ÖRNEK 10

2009, 2010 ve 2011 yılı için üç gıda maddesinin birim fiyatları (₺/kg) ve tüketim miktarları (kg) aşağıda verilmiştir. Bu verilerden hareketle 2010 ve 2011 yılı Paasche fiyat indeksini hesaplayalım.

Yıllar	i	Zeytinyağı		Ay çiçek yağı		Mısır yağı	
		Fiyat(p_i)	Miktar(q_i)	Fiyat(p_i)	Miktar(q_i)	Fiyat(p_i)	Miktar(q_i)
2009	0	11	550	8	7800	12	300
2010	1	14	560	9	7850	15	450
2011	2	22	500	11	9300	16	480

2009 yılı temel devre kabul edilerek (indeksi %100) 2010 ve 2011 yılı Paasche fiyat indeksi ($_p\dot{I}_L$), aşağıdaki gibi hesaplanır.

2009 indeksi, $_{p}\dot{I}_{P} = \dfrac{\sum p_0 \times q_0}{\sum p_0 \times q_0} \times 100 = \%100,00$ (daima)

2010 indeksi,

$$_{p}\dot{I}_{P} = \frac{\sum p_1 \times q_1}{\sum p_0 \times q_1} \times 100 = \frac{(14 \times 560) + (9 \times 7850) + (15 \times 450)}{(11 \times 560) + (8 \times 7850) + (12 \times 450)} \times 100 = \%114,63$$

2011 indeksi,

$$_{p}\dot{I}_{P} = \frac{\sum p_2 \times q_2}{\sum p_0 \times q_2} \times 100 = \frac{(22 \times 500) + (11 \times 9300) + (16 \times 480)}{(11 \times 500) + (8 \times 9300) + (12 \times 480)} \times 100 = \%141,23$$

Yıllar	i	Paasche Fiyat İndeksi - $_{p}\dot{I}_{P}$ (%)
2009	0	100,00
2010	1	114,63
2011	2	141,23

Paasche fiyat indeksinde her bir yıl için farklı tartı kullanıldığından, temel devreye göre kıyaslama yapmak sakıncalı olabilmektedir. Yine de Paasche indeksi %100'den olan farklılığına göre yorumlanabilir. Buna göre, gıda fiyatlarının ortalama olarak 2010'da %14,63 oranında ve 2011'de %41,23 oranında arttığı söylenir.

Paasche miktar indeksi hesabında her bir yıl için farklı tartı kullanıldığından temel devreye göre kıyaslama yapmak sakıncalı olabilmektedir.

Buraya kadar Laspeyres ve Paasche indekslerini, örnekler yardımıyla açıklamaya çalıştık. Şimdi bu açıklamalarımızı bir örnek üzerinde birleştirelim.

ÖRNEK 11

Endüstriyel üretimin önemli göstergelerinden olan metal çeşitlerinin satış miktarları (Ton) ve fiyat (₺/Ton) seviyeleri aşağıda verilmiştir. Bu verilerden yaralanarak

a. Laspeyres fiyat indekslerini,
b. Paasche fiyat indekslerini,
c. Laspeyres miktar indekslerini,
d. Paasche miktar indekslerini, hesaplayınız.

Yıllar	i	Bakır		Pik demir		Kurşun	
		p_i-fiyat	q_i-miktar	p_i-fiyat	q_i-miktar	p_i-fiyat	q_i-miktar
2007	0	p_0= 1329	q_0= 97	p_0= 213	q_0= 4057	p_0= 603	q_0= 26
2008	1	p_1= 1308	q_1= 96	p_1= 246	q_1= 3473	p_1= 514	q_1= 21
2009	2	p_2= 1278	q_2= 89	p_2= 284	q_2= 3739	p_2= 480	q_2= 25
2010	3	p_3= 1354	q_3= 100	p_3= 310	q_3= 3817	p_3= 528	q_3= 22
2011	4	p_4= 1405	q_4= 125	p_4= 318	q_4= 4094	p_4= 662	q_{11}= 38

a. Laspeyres fiyat indeksleri **($_{p}\dot{I}_{L}$):**

2007 indeksi,

$$_{p}\dot{I}_{L} = \frac{\sum p_0 \times q_0}{\sum p_0 \times q_0} \times 100 = \frac{(1329 \times 97) + (213 \times 4057) + (603 \times 26)}{(1329 \times 97) + (213 \times 4057) + (603 \times 26)} \times 100 = \%100,00$$

2008 indeksi,

$$_p\dot{I}_L = \frac{\sum p_1 \times q_0}{\sum p_0 \times q_0} \times 100 = \frac{(1308 \times 97) + (246 \times 4057) + (514 \times 26)}{1.008.732} \times 100 = \%112,84$$

2009 indeksi,

$$_p\dot{I}_L = \frac{\sum p_2 \times q_0}{\sum p_0 \times q_0} \times 100 = \frac{(1278 \times 97) + (284 \times 4057) + (480 \times 26)}{1.008.732} \times 100 = \%127,75$$

2010 indeksi,

$$_p\dot{I}_L = \frac{\sum p_3 \times q_0}{\sum p_0 \times q_0} \times 100 = \frac{(1354 \times 97) + (310 \times 4057) + (528 \times 26)}{1.008.732} \times 100 = \%139,06$$

2011 indeksi,

$$_p\dot{I}_L = \frac{\sum p_4 \times q_0}{\sum p_0 \times q_0} \times 100 = \frac{(1405 \times 97) + (318 \times 4057) + (662 \times 26)}{1.008.732} \times 100 = \%143,11$$

Bu sonuçlara göre söz konusu üç metalin fiyatlarında, temel devre 2007'ye göre ortalama olarak 2008'de %12,84; 2009'da %27,75; 2010'da %39,06 ve 2011'de %43,11 oranında artış olmuştur.

b. Paasche fiyat indeksleri **($_p\dot{I}_P$):**

2007 indeksi, $$_p\dot{I}_P = \frac{\sum p_0 \times q_0}{\sum p_0 \times q_0} \times 100 = \frac{1.008.732}{1.008.732} \times 100 = \%100,00$$

2008 indeksi,

$$_p\dot{I}_P = \frac{\sum p_1 \times q_1}{\sum p_0 \times q_1} \times 100 = \frac{(1308 \times 96) + (246 \times 3473) + (514 \times 21)}{(1329 \times 96) + (213 \times 3473) + (603 \times 21)} \times 100 = \%112,58$$

2009 indeksi,

$$_p\dot{I}_P = \frac{\sum p_2 \times q_2}{\sum p_0 \times q_2} \times 100 = \frac{(1278 \times 89) + (284 \times 3739) + (480 \times 25)}{(1329 \times 89) + (213 \times 3739) + (603 \times 25)} \times 100 = \%127,73$$

2010 indeksi,

$$_p\dot{I}_P = \frac{\sum p_3 \times q_3}{\sum p_0 \times q_3} \times 100 = \frac{(1354 \times 100) + (310 \times 3817) + (528 \times 22)}{(1329 \times 100) + (213 \times 3817) + (603 \times 22)} \times 100 = \%138,69$$

2011 indeksi,

$$_p\dot{I}_P = \frac{\sum p_4 \times q_4}{\sum p_0 \times q_4} \times 100 = \frac{(1405 \times 125) + (318 \times 4094) + (662 \times 38)}{(1329 \times 125) + (213 \times 4094) + (603 \times 38)} \times 100 = \%141,62$$

Paasche fiyat indeksi hesabında her bir yıl için indeksi hesaplanacak devrenin miktarı tartı (değişken tartı) olarak kullanıldığından, bu indeks değerlerini kıyaslamak pek sağlıklı olmamaktadır. Yine de Paasche indeksi %100'den olan farklılığa göre, söz konusu üç metalin fiyatlarında ortalama olarak 2008'de %12,58 oranında, 2009'da %27,73 oranında, 2010'da %38,6 oranında ve 2011'de %41,62 oranında artış görüldüğü söylenebilir.

c. Laspeyres miktar indeksi ($_q İ_L$):

2007 indeksi, $_q İ_L = \dfrac{\sum p_0 \times q_0}{\sum p_0 \times q_0} \times 100 = \dfrac{1.008.732}{1.008.732} \times 100 = \%100,00$

2008 indeksi,

$$_q İ_L = \frac{\sum p_0 \times q_1}{\sum p_0 \times q_0} \times 100 = \frac{(1329 \times 96) + (213 \times 3473) + (603 \times 21)}{1.008.732} \times 100 = \%87,24$$

2009 indeksi,

$$_q İ_L = \frac{\sum p_0 \times q_2}{\sum p_0 \times q_0} \times 100 = \frac{(1329 \times 89) + (213 \times 3739) + (603 \times 25)}{1.008.732} \times 100 = \%92,17$$

2010 indeksi,

$$_q İ_L = \frac{\sum p_0 \times q_3}{\sum p_0 \times q_0} \times 100 = \frac{(1329 \times 100) + (213 \times 3817) + (603 \times 22)}{1.008.732} \times 100 = \%95,09$$

2011 indeksi,

$$_q İ_L = \frac{\sum p_0 \times q_4}{\sum p_0 \times q_0} \times 100 = \frac{(1329 \times 125) + (213 \times 4094) + (603 \times 38)}{1.008.732} \times 100 = \%105,19$$

Bu sonuçlara göre söz konusu üç metalin üretim miktarlarında, temel devre 2007'ye göre ortalama olarak 2008'de %12,76 oranında; 2009'da %7,83 oranında ve 2010'da %4,81; oranında azalış olmuştur. Buna karşın yine 2007'ya göre 2011'de %5,19 oranında artış olmuştur

d. Paasche miktar indeksi ($_q İ_P$):

2007 indeksi, $_q İ_P = \dfrac{\sum p_0 \times q_0}{\sum p_0 \times q_0} \times 100 = \dfrac{1.008.732}{1.008.732} \times 100 = \%100,00$

2008 indeksi,

$$_q İ_P = \frac{\sum p_1 \times q_1}{\sum p_1 \times q_0} \times 100 = \frac{(1308 \times 96) + (246 \times 3473) + (514 \times 21)}{(1308 \times 97) + (246 \times 4057) + (514 \times 26)} \times 100 = \%87,04$$

2009 indeksi,

$$_q İ_P = \frac{\sum p_2 \times q_2}{\sum p_2 \times q_0} \times 100 = \frac{(1278 \times 89) + (284 \times 3739) + (480 \times 25)}{(1278 \times 97) + (284 \times 4057) + (480 \times 26)} \times 100 = \%92,16$$

20010 indeksi,

$$_q İ_P = \frac{\sum p_3 \times q_3}{\sum p_3 \times q_0} \times 100 = \frac{(1354 \times 100) + (310 \times 3817) + (528 \times 22)}{(1354 \times 97) + (310 \times 4057) + (528 \times 26)} \times 100 = \%94,84$$

2011 indeksi,

$$_q İ_P = \frac{\sum p_4 \times q_4}{\sum p_4 \times q_0} \times 100 = \frac{(1405 \times 125) + (318 \times 4094) + (662 \times 38)}{(1405 \times 97) + (318 \times 4057) + (662 \times 26)} \times 100 = \%104,09$$

Paasche miktar indeksi de Paasche fiyat indeksinde olduğu gibi indeks hesabında her devre için farklı tartılar kullanıldığından, kıyaslamak pek sağlıklı olmamaktadır. Örneğimizde yer alan üç metalin üretim miktarlarında ortalama olarak

2008'de %12,96 oranında, 2009'da %7,84 oranında ve 2010'da %5,16 oranında azalış olmuştur. Buna karşın 2011'de %4,09 oranında artış olmuştur

Laspeyres fiyat indeksleri, fiyat artışlarını olduğundan fazla, Paasche fiyat indeksleri ise olduğundan az göstermektedir. Çünkü fiyatı düşen bir malın talebi artacak, fiyatı yükselen bir malında talebi azalacaktır. Laspeyres fiyat indeksinde tartı olarak temel devre miktarı alındığından, fiyatı artan mallara olduğundan fazla, fiyatı düşen mallara da olduğundan az tartı verilmiş olmaktadır. Bu nedenle Laspeyres indeksinin gerçeğin üstünde sonuçlar vermesi beklenmektedir. Diğer taraftan Paasche fiyat indeksinde, indeksi hesaplanan devrenin miktarı tartı olarak kabul edilmektedir. Bu durumda da fiyatı düşmüş maddelere, fiyatı yükselenlere göre daha büyük tartı verilmiş olmaktadır. Dolayısıyla Paasche fiyat indeksinin gerçeğin altında sonuçlar vermesi beklenmektedir. Bu açıklamalardan Laspeyres indeksinin her zaman Paasche indeksinden daha yüksek çıkacağı sonucu çıkarılmamalıdır. Laspeyres indeksi Paasche indeksinden daha yüksek çıkabileceği gibi daha düşük veya eşit de olabilmektedir.

Fiyat (miktar) indeks hesabında amaç, fiyatların zaman içindeki oransal değişimini görmektir. Böyle bir durumda göz önünde bulundurulan tek değişken fiyat (miktar) olmaktadır. Bu bakımdan Laspeyres fiyat indeksinin tek değişkeni fiyat (miktar)'tır. Oysa Paasche fiyat (miktar) indeksinde, tartılarda fiyatlar gibi değişken durumundadır. Bir başka ifade ile indeksin devresinin değişmesi ile tartı olarak alınan miktar da değişmektedir. Buna bağlı olarak bu miktarların her indeks devresinde yeniden belirlenmesi gerekmektedir.

Sonuç olarak Laspeyres fiyat (miktar) indeksi, 10-15 yılda bir tartıların yenilenmesi koşuluyla sakıncalarının az olması ve kolay hesaplanabilmesi nedeniyle Paasche indeksinden daha üstün görünmektedir. Özellikle Laspeyres indeksinde temel devre fiyatını ve miktarını bilmek paydayı hesaplamak için yeterli olmaktadır. Böylece, devreler itibariyle miktar verilerini bulmada görülen sıkıntı ortadan kalkmaktadır. Laspeyres fiyat indeksi hesabında yalnız fiyatların elde edilmesi yeterli olmaktadır.

SIRA SİZDE 8

Laspeyres ve Paasche indekslerinden hangisi daha etkin kullanıma sahiptir? Neden?

Fisher (İdeal) İndeksi

Laspeyres ve Paasche indekslerinin taşıdıkları bazı sakıncaları gidermek üzere Amerikalı İktisatçı İrving Fisher tarafından "İdeal İndeks" adı verilen bir indeks geliştirilmiştir. İdeal indeks, Laspeyres ve Paasche indekslerinin *geometrik ortalaması*dır ve dolayısıyla bu iki indeks değerinin arasında bir değere sahiptir. Bu yönüyle İdeal indeksin gerçeği daha iyi temsil edebileceğini söylemek mümkündür. Ancak her devre için farklı tartı kullanan Paasche indeksine dayandığından, bu indeks de farklı tartı kullanmanın sakıncalarını taşır. Bu nedenle uygulamada az kullanılan indeks hesaplama tekniğidir. İdeal fiyat ve miktar indeksleri aşağıdaki formüller ile hesaplanır.

İdeal Fiyat İndeksi: $\; {}_{p}\dot{I}_{I} = \sqrt{{}_{p}\dot{I}_{L} \times {}_{p}\dot{I}_{P}}$

İdeal miktar İndeksi: $\; {}_{q}\dot{I}_{I} = \sqrt{{}_{q}\dot{I}_{L} \times {}_{q}\dot{I}_{P}}$

ÖRNEK 12

*Üç metalin fiyat ve miktarı için daha önce hesaplanan (Örnek 11'de) Laspeyres ve Paasche indekslerinden yararlanarak, İdeal **fiyat** ve **miktar** indekslerini hesaplayalım,*

*İdeal **fiyat** indeksleri ($_p\dot{I}_I$);*

2007 indeksi: $_p\dot{I}_I = \sqrt{(100,00)\times(100,00)} = \%100,00$

2008 indeksi: $_p\dot{I}_I = \sqrt{(112,84)\times(112,58)} = \%112,71$

2009 indeksi: $_p\dot{I}_I = \sqrt{(127,75)\times(127,73)} = \%127,74$

2010 indeksi: $_p\dot{I}_I = \sqrt{(139,06)\times(138,69)} = \%138,87$

2011 indeksi: $_p\dot{I}_I = \sqrt{(143,11)\times(141,62)} = \%142,36$

*İdeal **miktar** indeksleri ($_q\dot{I}_I$);*

2007 indeksi: $_q\dot{I}_I = \sqrt{(100,00)\times(100,00)} = \%100,00$

2008 indeksi: $_q\dot{I}_I = \sqrt{(87,24)\times(87,04)} = \%87,14$

2009 indeksi: $_q\dot{I}_I = \sqrt{(92,17)\times(92,16)} = \%92,16$

2010 indeksi: $_q\dot{I}_I = \sqrt{(95,09)\times(94,84)} = \%94,96$

2011 indeksi: $_q\dot{I}_I = \sqrt{(105,19)\times(104,09)} = \%104,64$

Fiyat İndeksleri

İndeks uygulamalarında fiyat indeksleri, miktar indekslerinden daha çok tercih edilmektedir. Çünkü fiyatlardaki oransal değişmeler, ekonomik yönden büyük öneme sahiptir. Buna ek olarak fiyat verilerini elde etmek daha kolay olmakta ve karar vermek için çoğunlukla fiyatlar yeterli kabul edilmektedir. Uygulamada en önemli bileşik fiyat indeksleri "tüketici fiyatları indeksi (TÜFE)" ve "üretici fiyatları indeksi (ÜFE)"dir. Şimdi bunları sırasıyla açıklamaya çalışalım.

Tüketici Fiyatları İndeksi (TÜFE)

Tüketici fiyatları indeksi, tüketicilerin ve özellikle ücretli kesimin satın alma gücünde zaman içinde meydana gelen oransal değişmelerin ve perakende fiyat hareketlerinin seyrinin bir göstergesidir. Tüketiciler tarafından tüketim alışkanlıklarında oluşan mal ve hizmetlerin fiyatlarında zaman içinde meydana gelen değişmeler, hayat standardının devamı için çok önemlidir. Doğal olarak tüketim alışkanlıkları, bir aileden diğerine farklılıklar gösterir. Bu farklılıkların tüketilen mal ve hizmetlerden veya onların aile bütçesi içindeki oransal öneminden kaynaklanmaktadır.

Tüketici fiyatları indeksi, belli bir grup hane halkının tüketim alışkanlıklarını temsil eden belirli bir grup mal ve hizmetin, ortalama parkende satış fiyatlarında zaman içinde meydana gelen değişmelerini ölçmeyi hedefleyen bir araçtır.

Tüketici fiyatları indeksi belirli bir coğrafi alanda, belirli bir sosyoekonomik grubun, belirli bir hayat standardını sürdürebilmesi için fiyat değişmelerinin ne ölçüde etkili olduğunu ortaya koymaktadır. Belirli bir hayat standardının devam ettirilebilmesi, belirli miktarda mal ve hizmetin tüketilmesine bağlıdır. Tüketilmesi gerekli mal ve hizmetlerin miktarı sabit varsayılsa bile fiyatlar, değişken olduğundan, fiyatlarda meydana gelen her değişme, yaşam standardının devamı için gerekli olan para miktarını etkileyecektir.

Tüketim malları sepeti, özellikle ücretli ailelerin belli bir ihtiyacı karşılamak amacıyla satın aldığı bütün mal ve hizmetler yerine çoğunlukla tükettiği mal ve hizmetlerden oluşmaktadır.

Tüketici fiyatları indeksindeki değişme ancak belirli bir hayat standardını korumak için yapılması gereken toplam harcama tutarındaki değişmeyi gösterir. Bu nedenle "**tüketim malları sepetin**deki mal ve hizmetlerin miktar ve kaliteleri bir birini izleyen devrede sabit bırakılmakta, buna karşın tüketicilerin sepeti aynı mal ve hizmetlerle doldurabilmesi için bunlara ödedikleri fiyatlarda meydana gelen değişmeler izlenmektedir.

Tüketici fiyatları indeksi hesaplayabilmek için fiyatları takip edilecek malların (çay, un, yağ, zeytin, vb.) ve hizmetlerin (elektrik, telefon, internet, vb.) meydana getirdiği topluluğa *"mal sepeti"* denir. İndeks hesabında tüketilen bütün mal ve hizmetlerin düzenli olarak fiyat hareketlerini izlemek çok zordur. Tercihen bunlar çoğunlukla tüketilen mal ve hizmetlerle sınırlandırılarak 50 ile 450 arası mal ve hizmet temel alınmaktadır. Sepete dahil edilen mal ve hizmetler; miktar, tür ve kalite olarak açıkça tanımlanır ve indeks hesaplama süresince değiştirilmez. Böylece tüketici fiyatları indeksi, tüketicilerin satın alma gücünde zaman içinde meydana gelen değişmenin göstergesi olarak kullanılır.

Günümüzde tüketici fiyatları indeksi, ekonomik kararların alınmasında, ücret ayarlamalarında, mahkemeler tarafından nafaka bağlanmasında, nafakanın çoğaltılması veya azaltılmasında ve özel okulların yıllık ücretlerinin belirlenmesinde vb. kullanılmaktadır.

Türkiye'de tüketici fiyatları indeksi *Laspeyres fiyat indeksi tekniği* kullanılarak Türkiye İstatistik Kurumu (TÜİK) ve İstanbul Ticaret Odası (İ.T.O.) tarafından hesaplanmaktadır. Türkiye İstatistik Kurumu (TÜİK) 2003=100 yılını temel alarak 423 mal ve hizmet içeren sepetten veri toplayarak TÜFE hesaplanmakta ve her ayın 3'ü ile 5'i arasında yayımlanmaktadırlar. Benzer bir mal sepeti kullanarak İstanbul Ticaret Odası da (İ.T.O.) 1995=100 yılını temel alarak TÜFE hesaplanmakta ve yayımlanmaktadırlar.

İNTERNET

www.tuik.gov.tr/PreTablo.do?tb_id=21&ust_id=7

Üretici Fiyatları İndeksi (ÜFE)

Üretici fiyatları indeksi toplam satışa konu olan malların fiyatlarını zaman içinde karşılaştırarak fiyat değişikliklerini ölçer. İndeks aynı maddenin, aynı miktar ve kalite içeriğinden sadece fiyat değişikliklerini yansıtır.

Üretici fiyatları indeksi (Ü.F.E.) genel olarak toptan satışa konu olan malların toptan fiyatlarındaki oransal değişmenin bir göstergesi olarak ifade edilebilir. Ekonomik hayatın önemli bir göstergesi olan, üretici fiyatları indekslerinin hesaplamasına temel oluşturacak veriler, genellikle toptancılardan, imalatçılardan ve borsalardan elde edilir. Üretici fiyatları indeksinin hazırlık çalışmalarında, indekse girecek malların doğru seçilmesine ve fiyatın takip edileceği bir standardın olmasına dikkat edilmelidir.

Türkiye'de üretici fiyatları indeksi uluslararası standartlara uygun olarak hesaplanmaktadır. Böylece sektörel fiyat hareketlerini ortaya koyabilmesinin yanında, uluslararası kıyaslamaya da olanak sağlanmaktadır.

Türkiye'de üretici fiyatları indeksi hesaplanmasında *Laspeyres fiyat indeksi tekniği* kullanılmaktadır. Üretici fiyatları indeksi genel Türkiye indeksi olarak kabul edilmektedir. Ülkedeki genel fiyat seviyesi, paranın satın alma gücündeki genel değişme, ticari konularda ve ücret toplu sözleşmelerinde resmî kaynak olarak üretici fiyatları indeksinden (Ü.F.E.) yararlanılmaktadır.

SIRA SİZDE 9

Türkiye genelinde fiyatlardaki değişim (enflasyon) hangi indeks ile ölçülmektedir?

Ekonomik, sosyal ve kültürel açıdan hızlı ve büyük bir değişim yaşandığından, zaman içinde teknolojinin de getirdiği yeniliklerle üretim ve tüketim ihtiyaçlarımız değişmektedir. Bu değişim çerçevesinde kimi mal ve hizmetler yerini yenilerine bırakmaktadır. Tüketim ve üretim yapısındaki bu değişikliklerin indekslere yansıtılabilmesi ve böylece güncel olması gerekir. Uluslararası tavsiyeler de dikkate alınarak Türkiye İstatistik Kurumu (TÜİK), tüketici ve üretici fiyatları indeksini 5 yılda bir yenileme yoluna gitmektedir.

Türkiye'de üretici fiyatları indeksi (ÜFE); Türkiye İstatistik Kurumu (TÜİK) tarafından temel-2003=100 yılı alınarak 751 mal ve hizmet için aylık ve yıllık olarak hesaplanmakta ve her ayın 3'ü ile 5'i arasında yayımlanmaktadır. Ayrıca İstanbul Ticaret Odası (İ.T.O.) tarafından da temel-1995=100 yılı alınarak üretici fiyatları indeksi hesaplanmakta ve yayımlanmaktadır.

www.tuik.gov.tr/VeriBilgi.do?tb_id=18&ust_id=6

İNTERNET

Özet

İndeks kavramını açıklamak, indeks sayısı okumak

İndeks, bir değişkenin veya değişkenler grubunun, zaman veya mekân içinde ortaya çıkan değişmelerini göstermek amacıyla hesaplanan, oransal ölçülerdir. İndeksler verilen bir dönemin, başka bir dönemle karşılaştırılmasını sağlayacak olan fiyatların, miktarların veya tutarların (fiyat ile miktar çarpımı) genellikle yüzdeler (%) şeklinde ifade edilmiş oranlarıdır. İndeksler her zaman iki sayı ile hesaplanır. Bunlardan biri kıyaslanan, diğeri değişimin kıyaslanacağı temel değerdir. Değişimi araştırılacak sayıya "cari değer", karşılaştırılacak değere de "temel değer" adı verilir. Temel değer paydaya, cari değer de paya yazılır. Oransal kıyaslamayı kolaylaştırmak için bu bölme işleminin sonucu 100 ile çarpılır.

İndeks hesabında; tek bir malın fiyatında (veya miktarında) meydana gelen oransal değişmelerin belirlenmesi söz konusu ise basit indeks, birden çok malın fiyatlarında ki birlikte değişme veya fiyat ve miktarlarındaki birlikte değişme hâlinde bileşik indeks söz konusudur.

Basit fiyat (miktar) indeksi hesaplamak ve yorumlamak

Basit indeks, tek bir malın fiyatında, miktarında ve değer tutarında zaman içinde meydana gelen oransal değişmeyi ölçen bir orandır. Böylece bir malın temel alınan yıla (veya aya) göre fiyatında (veya miktarında) meydana gelen oransal değişme belirlenir.

Bileşik indeks hesaplama tekniklerini açıklamak

Bileşik indeks; birbiriyle ilişkili birden çok malın fiyatında (ve/veya miktarında) meydana gelen oransal değişmelerin belirlenmesinde kullanılmaktadır. Bileşik indeks hesabına konu olacak mal ve hizmet sayısının yeterli düzeyde olması gerekir. Uygulamada bileşik indeks hesabı için 50 ile 200 arasındaki mal ve hizmet sayısının yeterli olacağı kabul edilmektedir. Bileşik indeks hesabında: i. İndeks ortalaması tekniği, ii. Ortalamalar indeksi tekniği, iii. Laspeyres ve Paasche indeksleri ve iv. Fisher (İdeal) indeks tekniği, kullanılır. Bileşik indeks hesabında amaç, indekse dahil edilecek bütün mal ve hizmetlerin eş zamanlı fiyatlarındaki (veya miktarlarındaki) gelişmeyi temsil edecek bir ortalama oran bulmaktır. Tartılı indeks olan Laspeyres indeksi, özellikle fiyatlardaki enflasyonu ölçmeye yarar.

Enflasyon hesabında Laspeyres fiyat indeksi hesabını açıklamak

Laspeyres fiyat indeksi temel devre miktarını tartı olarak kullanırken Laspeyres miktar indeksi, temel devre fiyatını tartı olarak kullanmaktadır. Özellikle fiyatlardaki **enflasyonu** ölçmeye yarayan Laspeyres fiyat indeksi aşağıdaki formülle hesaplanır.

Laspeyres fiyat indeksi ($_p\dot{I}_L$)
$_p\dot{I}_L = \frac{\sum p_i \cdot q_0}{\sum p_0 q_0} \times 100$

Türkiye'de tüketici fiyatları indeksi **Laspeyres fiyat indeksi tekniği** kullanılarak Türkiye İstatistik Kurumu (TÜİK) ve İstanbul Ticaret Odası (İ.T.O.) tarafından hesaplanmaktadır. Türkiye İstatistik Kurumu (TÜİK) 2003=100 yılını temel alarak 423 mal ve hizmet içeren sepetten veri toplayarak TÜFE hesaplanmakta ve her ayın 3'ü ile 5'i arasında yayımlanmaktadırlar.

Başlıca fiyat indekslerini açıklamak

İndeks uygulamalarında fiyat indeksleri, miktar indekslerinden daha çok tercih edilmektedir. Çünkü fiyatlardaki oransal değişmeler, ekonomik yönden büyük öneme sahiptir. Uygulamada en önemli bileşik fiyat indeksi "tüketici fiyatları indeksi (TÜFE)" ve "üretici fiyatları indeksi (ÜFE)"dir. Tüketici fiyatları indeksi, tüketicilerin ve özellikle ücretli kesimin satın alma gücünde zaman içinde meydana gelen oransal değişmelerin ve parkende fiyat hareketlerinin seyrinin bir göstergesidir. Üretici fiyatları indeksi toplam satışa konu olan malların fiyatlarını zaman içinde karşılaştırarak fiyat değişikliklerini ölçer.

Türkiye'de tüketici fiyatları indeksi ve üretici fiyatları indeksleri Laspeyres fiyat indeksi tekniği kullanılarak; Türk İstatistik Kurumu (TUİK) tarafından hesaplanmakta ve her ayın 3'ü ile 5'i arasında yayımlanmaktadırlar.

Kendimizi Sınayalım

1. A-Malının birim satış fiyatından (₺/ton) oluşan zaman serisi aşağıda verilmiştir.

Yıllar	i	A-malı fiyatı(p_i)
2009	0	400
2010	1	462
2011	2	510

A-malı için 2011 yılı sabit esaslı fiyat indeksi (S.E.F.İ.) aşağıdakilerden hangisidir?

a. %110,5
b. %115,5
c. %127,5
d. %135,5
e. %137,5

2. Aşağıda verilen zaman serisi, doğal gaz fiyatlarından (₺/m^3) oluşmaktadır.

Yıllar	i	Doğal Gaz Fiyatı (p_i)
2008	0	171,14
2009	1	198,25
2010	2	215,30

Doğal gaz için 2010 yılı değişken esaslı fiyat indeksi (D.E.F.İ.) aşağıdakilerden hangisidir?

a. %108,6
b. %115,8
c. %125,8
d. %128,6
e. %135,8

3. Aşağıda verilen zaman serisi, elektrik tüketim miktarlarından (kws) oluşmaktadır.

Yıllar	i	Elektrik Tüketim Miktarı (q_i)
2009	0	3670
2010	1	3880
2011	2	3500

Elektrik tüketim miktarı için 2011 yılı sabit esaslı miktar indeksi (S.E.M.İ.) aşağıdakilerden hangisidir?

a. %90,2
b. %95,4
c. %104,8
d. %105,7
e. %110,8

4. Aşağıda verilen zaman serisinde, bir gıda malının fiyat indeksi verilmiştir.

Yıllar	i	D.E.F.İ (%)
2009	0	100
2010	1	110
2011	2	90

Hesaplanan indeksten 2011 yılı indeksi (%90) aşağıdakilerin hangisinde doğru okunmaktadır?

a. Gıda malının 2011 yılı fiyatı 2009 yılı fiyatına göre %10 azalmıştır.
b. Gıda malının 2011 yılı fiyatı 2009 yılı fiyatına göre %10 artmıştır.
c. Gıda malının 2011 yılı fiyatı 2010 yılı fiyatına göre %10 artmıştır.
d. Gıda malının 2011 yılı fiyatı 2010 yılı fiyatına göre %10 azalmıştır.
e. Gıda malının 2011 yılı fiyatı 2010 yılı fiyatına göre %20 azalmıştır.

5. Aşağıda verilen zaman serisinde sabit esaslı fiyat indeksleri (S.E.F.İ.) verilmiştir.

Yıllar	i	S.E.F.İ (%)
2008	0	100
2009	1	120
2010	2	125
2011	3	150

Bunların yardımıyla 2009 yılı değişken esaslı fiyat indeksi (D.E.F.İ.) aşağıdakilerin hangisidir?

a. %100
b. %105
c. %110
d. %115
e. %120

6. Aşağıda verilen zaman serisinde sabit esaslı fiyat indeksleri (S.E.F.İ.) verilmiştir.

Yıllar	i	S.E.F.İ (%)
2008	0	100
2009	1	120
2010	2	125
2011	3	150

Bunların yardımıyla 2010 yılı değişken esaslı fiyat indeksi (D.E.F.İ.) aşağıdakilerin hangisidir?

a. %101,16
b. %102,16
c. %103,16
d. %104,16
e. %105,16

7. Aşağıda üç mal için, ayrı ayrı sabit esaslı fiyat indeksleri (S.E.F.İ.) hesaplanmış olarak verilmiştir.

Yıllar	i	A-Malı S.E.F.İ. (%)	B-Malı S.E.F.İ. (%)	C-Malı S.E.F.İ. (%)
2009	0	100	100	100
2010	1	119	124	129
2011	2	115	152	125

Buna göre "indeks ortalamalar tekniği" ile hesaplanacak 2010 yılı sabit esaslı bileşik fiyat indeksi, aşağıdakilerin hangisidir?

a. %124
b. %125
c. %129
d. %135
e. %139

8. Borsada işlem gören A,B,C ve D şirketlerinin yıllara göre hisse senetlerinin birim fiyatları ve işlem miktarları aşağıda verilmiştir.
Bu ferilerden yararlanarak hesaplanacak 2010 yılı Laspeyres fiyat indeksi aşağıdakilerin hangisidir?

Yıllar	i	A-Şirketi		A-Şirketi		A-Şirketi		A-Şirketi	
		Fiyat	Mik.	Fiyat	Mik.	Fiyat	Mik.	Fiyat	Mik.
2009	0	16	150	30	200	15	160	18	200
2010	1	20	160	22	180	20	200	20	210
2011	2	24	180	18	300	23	220	22	180

a. %100,39
b. %101,39
c. %102,39
d. %103,39
e. %104,39

9. Borsada işlem gören A,B,C ve D şirketlerinin yıllara göre hisse senetlerinin birim fiyatları ve işlem miktarları aşağıda verilmiştir.

Yıllar	i	A-Şirketi		B-Şirketi		C-Şirketi		D-Şirketi	
		Fiyat	Mik.	Fiyat	Mik.	Fiyat	Mik.	Fiyat	Mik.
2009	0	16	150	30	200	15	160	18	200
2010	1	20	160	22	180	20	200	20	210
2011	2	24	180	18	300	23	220	22	180

Bu ferilerden yararlanarak hesaplanacak 2010 yılı Pasche fiyat indeksi aşağıdakilerin hangisidir?

a. %100,21
b. %101,21
c. %102,21
d. %103,21
e. %104,21

10. Üretici fiyatları indeksi (ÜFE) ve türetici fiyatları indeksi (TÜFE) hesabında etkin olarak kullanılan bileşik indeks hesaplama tekniği aşağıdakilerden hangisidir?

a. İndeks ortalamalar tekniği
b. Ortalamalar indeks tekniği
c. Laspeyres indeks tekniği
d. Paasche indeks tekniği
e. Fisher indeks tekniği

Kendimizi Sınayalım Yanıt Anahtarı

1. c	Yanıtınız yanlış ise "Sabit Esaslı İndeks" konusunu yeniden gözden geçiriniz.
2. a	Yanıtınız yanlış ise "Değişken Esaslı İndeks" konusunu yeniden gözden geçiriniz.
3. b	Yanıtınız yanlış ise "Sabit Esaslı İndeks" konusunu yeniden gözden geçiriniz.
4. d	Yanıtınız yanlış ise "Değişken Esaslı İndeks" konusunu yeniden gözden geçiriniz.
5. e	Yanıtınız yanlış ise "İndekslerin Birinden Diğerine Geçiş" konusunu yeniden gözden geçiriniz.
6. d	Yanıtınız yanlış ise "İndekslerin Birinden Diğerine Geçiş" konusunu yeniden gözden geçiriniz.
7. a	Yanıtınız yanlış ise "İndeks Ortalamalar tekniği" konusunu yeniden gözden geçiriniz.
8. b	Yanıtınız yanlış ise "Laspeyres ve Paasche İndeksleri" konusunu yeniden gözden geçiriniz.
9. e	Yanıtınız yanlış ise "Laspeyres ve Paasche İndeksleri" konusunu yeniden gözden geçiriniz.
10. c	Yanıtınız yanlış ise "Fiyat indeksleri" konusunu yeniden gözden geçiriniz.

Sıra Sizde Yanıt Anahtarı

Sıra Sizde 1

İndeks hesabında; tek bir malın fiyatında (veya miktarında) meydana gelen oransal değişmelerin belirlenmesi söz konusu ise basit indeks, birden çok malın fiyatlarındaki birlikte değişme veya fiyat ve miktarlarındaki birlikte değişme hâlinde bileşik indeks söz konusudur.

Sıra Sizde 2

Sabit esaslı indeks hesabında temel devrenin seçimi çok önemlidir. Temel devrenin enflasyon, deflasyon (durgunluk) ve devalüasyon gibi iktisadi olayların aşırılık kazanmadığı, istikrarlı bir dönemden seçilmesi gerekir. Uygulamada, temel devre olarak seçilecek istikrarlı bir yıl bulmak çoğu zaman zor olduğundan genelde zaman serisini ilk yılı temel devre olarak seçilmektedir.

Sıra Sizde 3

Değişken esaslı indeks hesabında; her devrenin değeri, kendinden bir önceki devrenin değerine oranlanır, oransal kıyaslamayı sağlamak için de sonuç yüz ile çarpılır.

Sıra Sizde 4

Bir kaynakta (makale, kitap, gazete vb.) hesaplanmış, ancak orijinal verileri bulunmayan indekslerle karşılaşılabilir. İşte böyle durumlarda aralarındaki matematiksel bağlantılar nedeniyle sabit ve değişken esaslı indekslerden birine ait indeks değerleri bilindiğinde bunlardan yararlanarak diğerini hesaplamak mümkündür.

Sıra Sizde 5

Laspeyres, Paasche ve bunlara bağlı olarak hesaplanabilen Fisher (İdeal) İndeksleri "tartılı bileşik indeks" niteliğindedir.

Sıra Sizde 6

İndeks ortalamalar tekniğinde her bir malın ayrı ayrı basit fiyat (veya miktar) indeksleri hesaplanır. Sonra, her bir yıl için hesaplanan bu indekslerin ortalamaları alınır. Bu ortalamalar "bileşik indeks" kabul edilir. Ortalamalar indeks tekniğinde ise öncelikle her bir yıl için ortalama fiyat (veya miktar) hesaplanır. Sonra da ortalamalardan oluşan serinin, hesaplanan indeksleri, bileşik indeks kabul edilir.

Sıra Sizde 7

Laspeyres ve Paasche indeksleri tartı olarak kullandıkları devre itibari ile birbirinden ayrılırlar. Laspeyres indeksinde temel devre değerleri tartı olarak kullanılır iken Paasche'de indeksi hesaplanan yıl değeri tartı olarak kullanılmaktadır.

Sıra Sizde 8

Laspeyres indeksi, sakıncalarının az olması ve kolay hesaplanabilmesi nedeniyle Paasche indeksinden daha üstün görünmektedir. Özellikle Laspeyres indeksinde temel devre fiyatını ve miktarını bilmek paydayı hesaplamak için yeterli olmaktadır. Böylece, devreler itibariyle miktar verilerini bulmada görülen sıkıntı ortadan kalkmaktadır.

Sıra Sizde 9

Türkiye'de genel fiyat seviyesi, paranın satın alma gücündeki genel değişme (enflasyon), üretici fiyatları indeksi (ÜFE) ile ölçülmektedir.

Yararlanılan ve Başvurulabilecek Kaynaklar

Atlas, M. (2007), **İstatistik II (Çözümlü Örnekler),** Ak Ofset, Eskişehir

Çömlekçi, N. (1998), **Temel İstatistik (İlke ve Teknikleri),** Gözden geçirilmiş 3. Baskı, Bilim Teknik Yayınevi, İstanbul

Gürsakal, N. (2010), **Betimsel İstatistik (İstatistik I),** Dora Basın Yayın Dağıtım, Bursa

Kazmier, L.J. and Pohl, N.F. (1987), **Basic Statistics for Business and Economics,** McGraw-Hill Book Co., New York

Mann, P.S. (1995), **Statistics for Business and Economics,** J. Wiley, New York

Newbold, P. (2001), **İşletme ve İktisat için İstatistik,** Çev: Ümit Şenesen, Literatür Yayıncılık, İstanbul

Serper, Ö. (2004), **Uygulamalı İstatistik 1 (Genş. 5. Baskı),** Ezgi Kitapevi, Bursa

Ünver, Ö. (1995), **Uygulamalı İstatistik Yöntemler (Giriş),** Siyasal Kitabevi, Ankara

8

Amaçlarımız

Bu üniteyi tamamladıktan sonra;

- Karar problemlerinde strateji tablosu oluşturabilecek,
- Belirsizlik altında karar verebilecek,
- Risk altında karar verebilecek,
- Karar ağacı çizebilecek

bilgi ve becerilere sahip olacaksınız.

Anahtar Kavramlar

- Strateji Tablosu
- Belirsizlik Altında Karar Verme
- İyimserlik Ölçütü
- Kötümserlik Ölçütü
- Hurwicz'in Genelleştirilmiş İyimserlik Ölçütü
- Risk Altında Karar Verme
- En İyi Beklenen Değer
- Karar Ağacı

İçindekiler

İstatistik-II

Karar Teorisi

Karar Teorisi

GİRİŞ

Temel olarak birden fazla seçenek içerisinden seçim yapma işlemine karar verme adı verilir. Günümüzde işletmeler hızlı bir büyüme içerisindedirler. Doğal olarak büyüme ile birlikte eskiye nazaran daha karmaşık kararların yönetim tarafından alınması beklenmektedir. Bireyler gündelik yaşantılarında onlarca kararı içgüdülerine dayalı olarak alabilmektedirler. Örneğin bir caddeden karşıdan karşıya geçme eylemi içgüdülerimize bağlı olarak yaptığımız gündelik bir karar verme sürecidir. Benzer bir şekilde günlük yaşantıda ihtiyaç duyulan gereksinimlerin elde edilmesinde de farklı seçenekler arasından seçim yapma işlemi, bir başka deyişle karar verme işlemi yürütülür. Örneğin bir lokantada ana menüden yemek seçme işlemi bir karar verme işlemidir. Kendi istek ve beğenilerinize göre çeşitli seçenekleri inceler, fiyat, damak tadınıza uygunluk gibi kriterleri de göz önüne alarak kararınızı verir ve yemeğinizi sipariş edersiniz. İşletmeler yeni açacakları mağaza veya fabrikaları nereye kurmaları gerektiği kararını da vermek durumunda olabilmektedirler. Kimi durumlarda ise yöneticilerin hangi ürünü ne miktarda ve yılın hangi dönemlerinde üretmeleri gerektiği kararını vermeleri gerekebilir. Karar verme teknikleri bu tür durumlarda yöneticilerin en uygun kararı vermesinde yardımcı olmak amacı ile geliştirilmiş tekniklerdir. Özellikle işletmelerin başarısı yöneticilerinin tutarlı kararlar verebilmesine de bağlıdır.

Karar verme sürecinde karar probleminin zaman içerisinde oluşturacağı sonuçlardan etkilenen sorumlu kişiye karar verici adı verilir. Karar verici, bir kişi olabileceği gibi bir grup veya bir kurum da olabilir. Karar sürecinde önceden saptanan ve karar verici için belirgin özelliği olumlu olan sonuç ise amacı oluşturur. Belirlenen amaca ulaşmada etkin olan en az iki eylem biçimi varken bu eylem biçimlerinin seçiminde karar vericinin içinde yer alabileceği çeşitli koşullar veya ortamlar etkili olabilmektedir. Ek olarak karar vericilerin çözümünü araştırdıkları karar problemine ilişkin kapsam, çevre ve paydaşları çok iyi bir şekilde değerlendirmeleri gerekmektedir.

Çeşitli eylem seçeneklerinden belirlenen kararın etkin ve kararı uygulayacaklar arasında mümkün olduğu kadar yüksek bir kabul görmesi beklenir. Karar kendi yargı birimlerine göre iyi olarak nitelenmelidir. İyi bir karar, benzer problemler ile karşı karşıya kalan iki farklı yöneticinin aynı seçenekler ve aynı ortamlar altında aynı kararı vermesi ile özdeşleştirilebilir. Karar alıcının çevresindekilerinde, alınan kararı iyi olarak nitelendirmeleri beklenir. Karar vericilerin etkin ve rasyonel olmaları, problemin mali boyutunu iyi analiz etmiş olmaları, geleceğe dönük bir analiz yapmış olmaları gerekir. Karar verme, genel olarak bir problem çözümleme süreci olarak adlandırılabilir.

Bu ünitede karar probleminin temel bileşenleri ele alınmaktadır. Bu bileşenler irdelendikten sonra karar vericinin içinde bulunabileceği farklı durumlara ilişkin karar problemi çözüm süreçlerinde kullanılan teknikler örneklerle açıklanacaktır. Ünite kapsamında karar teorisinde kullanılan tüm tekniklerin gösterimi yerine en çok kullanılan bazı teknikler ele alınmıştır.

KARAR PROBLEMİNİN BİLEŞENLERİ

Bireyler veya işletmeler karar verme durumu ile karşı karşıya kaldıklarında, en iyi karara ulaşabilmek için, karar probleminin çeşitli bileşenlerini en doğru şekilde tanımlamalıdır. Karar verme sürecinde, doğru ve eksiksiz bir şekilde tanımlanması beklenen çeşitli karar bileşenleri bulunmaktadır. Bu bileşenler aşağıdaki gibi sıralanabilir.

Karar Alternatifleri: Üzerinde çözüm araştırılan karar verme probleminde, karar vericinin uygulayabileceği farklı karar seçenekleri "karar alternatifleri"ni oluşturur. Finansal yatırım yapmak isteyen bir kişi elinde bulunan parayı farklı yatırım seçeneklerinde değerlendirebilir. Örneğin, yatırımcı elindeki para ile altın alımında bulunabilir, belirli bir faiz oranı ile bankaya yatırabilir veya borsada çeşitli hisse senetlerine yatırım yapabilir. Yeni bir fabrika kurmak isteyen bir yönetici fabrikasını mevcut fabrikanın yanında mı yoksa farklı bölgelerdeki alternatiflerden birisinde mi kuracağı kararını vermek durumunda olabilir. Yeni bir otomobil geliştiren üretici, ön cam silecek lastiklerini kendi fabrikası içerisinde mi üreteceği yoksa dışarıdan temin yoluna mı gideceği kararını vermek durumundadır. Hemen hemen tüm işletme eylemlerinde farklı karar alternatiflerine rastlamak mümkündür. Karar verici, üzerinde çalıştığı problem için ortaya çıkabilecek tüm karar alternatiflerini araştırmalı ve mümkün olan tüm karar seçeneklerini problemine eklemelidir.

Doğal durumlar gelecekte ortaya çıkması mümkün olan gerçek olayları gösterir.

Doğal Durumlar: Doğal durumlar genellikle bir kişi veya kuruluşun eylemi olarak ortaya çıkmayan ancak karar vericinin verdiği kararın sonuçlarını doğrudan etkileyen durumları temsil etmek üzere tespit edilirler. **Doğal durumlar**, karar verici çeşitli karar alternatiflerden birisini seçtikten sonra ortaya çıkabilecek ve karar vericinin üzerinde etkisi olmayan ancak problemde yer alan karar seçeneklerinin sonuçlarını etkileyebilecek farklı senaryoları gösterirler. Doğal durumlar atmosferik olaylar olabileceği gibi politik durum, ekonomik durum, işçilerin psikolojik durumları gibi farklı durumlarda olabilirler. Örneğin, finansal bir yatırım kararını verecek olan yönetici, yatırımın getirisini beklediği dönemdeki bölgesel ekonomik durum hakkında bilgi sahibi olmak isteyecektir. Yatırım yapılacak dönem içerisinde ekonomik bir durgunluk beklentisinin olup olmaması, yatırımcının parasını borsada, faizde veya altında değerlendirip değerlendirmeme kararında belirleyici olacaktır. Böyle bir durumda doğal durumlar, "ekonomik durgunluk beklentisi var" ve "ekonomik durgunluk beklentisi yok" şeklinde düzenlenecektir. Bir başka örnekte ise bölünmüş yol inşaatı yapmakta olan bir firmanın işi zamanında bitirmesi için inşaat süresi içerisinde "uygun hava koşullarının ortaya çıkması" veya "uygun hava koşullarının ortaya çıkmaması" doğal durumlar olarak tanımlanabilir. Doğal durumlar, karar vericinin belirlediği karar seçenekleri üzerinde direkt veya dolaylı sonuçlara yol açacak durumlardır ve belirlenmeleri zor olmakla beraber, karar verme süreci için çok önemli bir bileşendir. Kimi durumlarda karar verici, doğal durumların ortaya çıkma olasılıklarını da belirleyebilir.

Sonuçlar: Belirli bir karar alternatifinin seçilmesi sonrasında ortaya çıkacak olan kazanç veya maliyet değeri sonuç olarak adlandırılır. Sonuç değerleri seçilen karar alternatifi ve ortaya çıkan doğal duruma göre farklı değerlere sahip olacaktır. Genellikle parasal ifadeler ile tanımlanırlar. Kimi durumlarda karar verici karar eylemleri sonucunda elde edeceği kazanca göre hesaplama yapabilirken kimi diğer durumlarda ise ortaya çıkacak maliyetlere göre hesaplama yapılabilir. Örneğin, bir senelik bir süre sonucunda ₺1000'lik bir altın yatırımı ekonomik durgunluk olmaması durumunda ₺1050'na ulaşabilir, elde edilen bu ₺1050 değeri sonuç değeri olarak ele alınacaktır. Sonuç değerlerinin farklı karar alternatifleri ve ortaya çıkabilecek farklı doğal durumlara göre tek tek hesaplanması gereklidir.

Strateji Tablosu (pay-off): Karar alternatifleri, doğal durumlar ve sonuçları bir araya getiren tabloya strateji tablosu adı verilir. **Strateji tablosu** kazanç yapılı olarak kurulabileceği gibi maliyet yapılı olarak da kurulabilir. Strateji tablosunun satırlarında farklı karar alternatifleri bulunur. m tane karar alternatifinin bulunduğu bir karar probleminde karar seçenekleri d_1, d_2, d_3, ..., d_m ile gösterilirler. Strateji tablosunun sütunlarında ise doğal durumlar yer alır. N tane doğal durumun bulunduğu karar probleminde doğal durumlar S_1, S_2, ..., S_n ile gösterilir. Strateji tablosunun her bir hücresinde ise karar alternatifi ve doğal durum ikilisi sonucu ortaya çıkan kazanç/maliyet değerleri yer alır. Sonuç değerleri, i seçilen karar alternatifini ve j ortaya çıkan doğal durumu göstermek üzere $ö_{ij}$ ile gösterilir. Tablo 8.1.'de bir strateji tablosu örneği gösterilmiştir. Karar probleminin çözülebilmesi için strateji tablosunun en doğru şekilde oluşturulması gereklidir. Eğer araştırmacı/karar verici, ortaya çıkması muhtemel olan doğal durumların gerçekleşme olasılıklarını tespit edebiliyor ise bu olasılıklarda strateji tablosuna doğal durumların altında veya üstünde yer alacak şekilde eklenir.

Karar alternatifleri, doğal durumlar ve sonuçları bir araya getiren tabloya **strateji tablosu** adı verilir.

Tablo 8.1
Strateji Tablosu.

	Doğal Durumlar					
		S_1	S_2	S_3	...	S_n
Karar Alternatifleri	d_1	$ö_{11}$	$ö_{12}$	$ö_{13}$	...	$ö_{1n}$
	d_2	$ö_{21}$	$ö_{22}$	$ö_{23}$	...	$ö_{2n}$
	d_3	$ö_{31}$	$ö_{32}$	$ö_{33}$	...	$ö_{3n}$
	.	.	.	.	...	.
	.	.	.	.	...	.
	d_m	$ö_{m1}$	$ö_{m2}$	$ö_{m3}$	...	$ö_{mn}$

Karar probleminin tüm bileşenleri doğru ve eksiksiz olarak belirlendikten sonra farklı teknikler yardımı ile strateji tablosundan faydalanılarak en iyi karar alternatifinin hangisi olduğuna karar verilir. İzleyen kesimlerde strateji tablosunun çözümünde kullanılan teknikler ele alınacaktır.

ÖRNEK 1

Küçük bir yatırımcı elinde bulunan ₺5000'sini bir seneliğine borsada değerlendirmek istemektedir. Bu amaç ile araştırma yapan yatırımcı üç farklı sektörden birer hisse senedini (Hisse A, Hisse B ve Hisse C) yatırıma uygun bulmuştur. Yatırımcı araştırması sırasında ekonominin genel durumuna ilişkin ortaya çıkabilecek üç farklı durum olduğunu belirlemiştir. Bu durumlar sırasıyla kötü ekonomi, durgun ekonomi ve iyi ekonomi olarak adlandırılmıştır. Farklı sektörlerin farklı ekonomik koşullarda değişik tepkiler verdiğini de geçmiş deneyimlerinden bilen yatırımcımız, hangi hisse senedine yatırım yapacağı kararını vermek durumundadır.

Bu problemde yatırım yapılacak hisse senedinin hangisi olduğu kararı verilmek istenmektedir. Karar verici olan küçük yatırımcının kendisine en yüksek kazancı getirmesi beklenen hisse senedini seçmesi doğal beklentidir. Problemde tespit edilen üç farklı hisse senedi karar alternatifleridir. Ekonomi için beklenen üç farklı durum ise doğal durumları oluşturmaktadır. Strateji tablosunu oluşturmak için, karar vericinin geçmiş kayıtlara bakarak farklı ekonomik koşullar altında ilgilenilen hisse senetlerinin nasıl bir finansal yapı izlediğini incelemesi gerekir. Örneğin A hisse senedi, ekonomi durgun olduğunda ₺5000'lik yatırım ile geçmiş kayıtlara göre bir yıllık süre içerisinde ₺5250'lik bir değere yükseliyor ise yatırımcının bu hisse senedinden kazancı ₺250 olacaktır. Aynı şekilde her hisse senedi ve doğal durum ikilisi için benzer analizler yapılarak elde edilebilecek kazançlar hesaplanır. Hesaplanan kazanç değerlerine göre yıl sonunda yatırımcının elinde olan hisse senedi fiyatları strateji tablosunda bir araya getilir. Yatırımcımızın elde edebileceği kazançlara ilişkin örnek bir strateji tablosu oluşturularak Tablo 8.2.'de gösterilmiştir. Bu strateji tablosuna göre yatırımcı, yatırım kararını B hisse senedinden yana kullanırsa ve ekonomik durum iyi olursa yatırımcımızın hisse senedi ₺5500'lik bir değere ulaşacaktır. Ancak yatırımcı tercihini veya kararını A hisse senedinden yana kullanırsa ve ekonomik durum kötü olursa zarara uğrayacak, seneyi ₺4750 ile tamamlayacaktır. Dolayısıyla yatırımcının karar verirken tüm seçenekleri dikkatli bir şekilde incelemesi gerekmektedir.

Tablo 8.2
Hisse senedi yatırım problemi için strateji tablosu.

		Doğal Durumlar		
		Kötü Ekonomi	Durgun Ekonomi	İyi Ekonomi
Karar Alternatifleri	Hisse A'yı al	4750	5050	5250
	Hisse B'yi al	4800	5020	5500
	Hisse C'yi al	5200	5080	5050

SIRA SİZDE 1

Bir karar verici günlük işlerini yürütebilmek için bilgisayar almak istemektedir. Ancak son teknolojik gelişmeleri inceleyen karar verici günlük işlerini yürütebilecek yeteneklere sahip bir dizüstü bilgisayar, bir tablet bilgisayar ve modern cep telefonu alternatiflerini geliştirmiştir. Karar vericinin karar alternatifleri nelerdir? Karar vericinin seçiminden mutluluğunu etkileyecek etmenler olabilir mi? Olabilir ise bu etmenleri açıklayınız.

BELİRSİZLİK ALTINDA KARAR VERME

Karar vericinin, doğal durumların ortaya çıkışlarına ilişkin herhangi bir olasılık değerine sahip olmadığı duruma belirsizlik adı verilir. Belirsizlik durumunda karar verme işlemi yürütülür ise karar verme sürecine belirsizlik altında karar verme adı verilir. Bu tür durumlarda karar verici farklı kriterlere göre her bir karar alternatifinin kendisine olan getirisini hesaplayarak, içinde bulunduğu duruma en uygun karar alternatifini seçecektir. Belirsizlik altında karar vermek durumunda olan bir karar verici, karar probleminin doğal durumlarını bilmekle beraber, hangi durumun gerçekten ortaya çıkabileceğini veya hangi durumun hangi olasılık ile ortaya çıkabileceğini bilmemektedir. Dolayısıyla belirsizlik altında karar verme durumunda kullanılan tekniklere göre verilen bir karar için, karar vericinin nasıl bir riske girdiğini hesaplaması mümkün değildir. Karar verici, kararını problemde ortaya çıkan ve strateji tablosunda gösterilen sonuç değerlerine göre verecektir. Belirsizlik altında karar verme durumunda bulunan bir karar vericinin kullanabileceği teknikler;

iyimserlik ölçütü, kötümserlik ölçütü ve Hurwicz'in genelleştirilmiş iyimserlik ölçütüdür. Bu ölçütler genel olarak karar vericinin içinde bulunduğu psikolojik durumu yansıtacak şekilde tasarlanmış ölçütlerdir.

İyimserlik Ölçütü

Karar vericinin iyimser olduğu durumlarda kullanılan ölçüttür. Her ne olursa olsun "hep iyi durumlar benim için ortaya çıkar" felsefesini benimseyen karar vericilerin uyguladığı bir tekniktir. **İyimserlik ölçütü** yaklaşımında, karar verici hangi karar alternatifini seçerse seçsin, strateji tablosunun satırlarında yer alan her karar alternatifinin en yüksek kazanç veya en düşük maliyet değerleri ile karşı karşıya kalacağını düşünür. Her karar alternatifinin en uygun değerlerinden de en büyük kazanç veya en küçük maliyete sahip kararın kendisi için geçerli olacağını kabul eder. İyimserlik ölçütünde karar verici, kazanç yapılı problemlerde "kazanabileceğim en büyük kazançlardan en büyüğünü seçerim" felsefesi ile hareket ederken maliyet yapılı problemlerde "ortaya çıkabilecek en küçük maliyetlerden en küçüğünü seçerim" felsefesi ile hareket etmektedir.

İyimserlik ölçütünde "hep iyi durumlar benim için ortaya çıkar" felsefesi benimsenir.

İyimserlik ölçütüne göre karar vericinin benimseyeceği karar alternatifi strateji tablosuna göre, kazanç yapılı problemlerde,

$$\underset{i}{\text{EnBüyük}}\{\underset{j}{\text{EnBüyük}}(ö_{ij})\}$$

işlemi ile tespit edilirken, maliyet yapılı problemlerde ise

$$\underset{i}{\text{EnKüçük}}\{\underset{j}{\text{EnKüçük}}(ö_{ij})\}$$

işlemi ile yürütülür. Bu işlemlere göre, eğer karar verici kazanç yapılı bir problem ile çalışıyor ise her karar alternatifi için strateji tablosunun satırlarında yer alan en büyük değerleri seçer ve bu değerler arasından en büyük değerli olan karar alternatifi en iyi karar olarak kullanılır. Eğer karar verici maliyet yapılı bir problem ile çalışıyor ise her karar alternatifi için strateji tablosunun satırlarında yer alan en küçük değerleri seçer ve bu değerler arasından en küçük değerli olan karar alternatifi en iyi karar olarak kullanılır.

ÖRNEK 2

Bir karar probleminde dört adet karar alternatifi ve üç adet doğal durum bulunmaktadır. Karar vericinin problemine ilişkin strateji tablosu kazanç yapılı olarak hazırlanmış ve Tablo 8.3'te sunulmuştur.

Tablo 8.3
Strateji tablosu.

		Doğal Durumlar		
		Durum 1	Durum 2	Durum 3
Karar Alternatifleri	Karar 1	900	850	1200
	Karar 2	1000	920	880
	Karar 3	850	1120	900
	Karar 4	660	875	950

İyimserlik ölçütünü kullanarak karar vericinin seçmesi gereken karar alternatifini belirleyelim.

Karar verici, ilk olarak strateji tablosunun sonuna yeni bir sütun ekler. Bu sütunda problem kazanç yapılı olduğundan dolayı, her bir karar içim ortaya çıkan en yüksek değerler yer alır. Daha sonra bu sütunda yer alan en yüksek kazanç değerlerinden en yükseğine sahip olan değeri veren karar en uygun karar olarak belirlenir. Tablo 8.4'te bahsedilen değerler gösterilmiştir.

Tablo 8.4
Her bir karar alternatifi için en büyük kazanç değerleri.

	Doğal Durumlar				
Karar Alternatifleri		Durum 1	Durum 2	Durum 3	En Büyük
	Karar 1	900	850	1200	1200
	Karar 2	1000	920	880	1000
	Karar 3	850	1120	900	1120
	Karar 4	660	875	950	950

Her bir karar alternatifi için belirlenen en büyük değerlerden (1200, 1000, 1120, 950) en büyüğü ise iyimserlik ölçütüne göre en iyi kararı verecektir. Tablo 8.4.'ün son sütununda yer alan değerlerden en büyüğü 1200'e eşittir ve bu değer 1 numaralı karar alternatifi için elde edilmiştir. Dolayısıyla iyimserlik ölçütüne göre karar vericinin seçmesi gereken karar alternatifi 1 numaralı karardır.

ÖRNEK 3

Bir inşaat firmasının yürütmek istediği 3 farklı inşaat projesi bulunmaktadır. Ancak firmanın elindeki olanaklar bu projelerden sadece bir tanesine yeteceği için firma yöneticilerinin en az maliyeti verecek olan projeyi seçmesi gerekmektedir. Firma yöneticileri tüm inşaat projelerini dikkatli bir şekilde incelemiş ve özellikle iklim koşullarının proje maliyetleri üzerinde büyük etkileri olabileceğini tespit etmişlerdir. Hava koşullarının uygun olması ve uygun olmaması durumlarına göre her proje için maliyetler ve Tablo 8.5'te gösterilmiştir.

Tablo 8.5
İnşaat proje maliyetleri için strateji tablosu (₺1.000.000 katları olarak).

	Doğal Durumlar		
Karar Alternatifleri		İklim Koşulu İyi	İklim Koşulu Kötü
	Proje 1	17	24
	Proje 2	15	28
	Proje 3	21	22

İyimserlik ölçütünü kullanarak karar vericinin seçmesi gereken karar alternatifini belirleyelim.

Karar verici ilk olarak strateji tablosunun sonuna yeni bir sütun ekler. Bu sütunda problem maliyet yapılı olduğundan dolayı, her bir karar alternatifinin uygulanması durumunda ortaya çıkacak olan en küçük değerler yer alır. Daha sonra bu sütunda yer alan en küçük maliyet değerleri içinden en küçüğüne sahip olan değeri veren karar en uygun karar olarak belirlenir. Tablo 8.6'da bahsedilen değerler gösterilmiştir.

Doğal Durumlar				
		İklim Koşulu İyi	İklim Koşulu Kötü	En Küçük
Karar Alternatifleri	Proje 1	17	24	17
	Proje 2	15	28	15
	Proje 3	21	22	21

Tablo 8.6
İnşaat proje maliyetleri için en küçük maliyet değerleri.

Her bir karar alternatifi için belirlenen en küçük değerler (17, 15, 21) içinden en küçüğü ise iyimserlik ölçütüne göre en iyi kararı verecektir. Tablo 8.6'nın son sütununda yer alan değerlerden en küçüğü 15'e eşittir ve bu değer 2 numaralı proje için elde edilmiştir. Dolayısıyla iyimserlik ölçütüne göre karar vericinin seçmesi gereken karar alternatifi 2 numaralı projenin gerçekleştirilmesi kararıdır. Ancak bu problemde dikkat edilmesi gereken bir nokta ödeme tablosunun dikkatli incelenmesi ile göze çarpabilir. Dikkat edilirse Proje 2 uygulamaya konulduğunda iklim koşulları kötü olursa firmanın göze aldığı maliyet 28 olacaktır. 28 değeri strateji tablosunun en yüksek değeridir. Dolayısıyla iyimserlik ölçütünün kullanılması, her ne kadar en düşük maliyet beklentisini getirmekteyse de aynı zamanda karar vericinin en kötü durum ile karşılaşması ile de sonuçlanabilir.

Kötümserlik Ölçütü

Karar vericinin kötümser olduğu durumlarda kullanılan ölçüttür. Her ne olursa olsun "hep kötü durumlar benim için ortaya çıkar" felsefesini benimseyen karar vericilerin uyguladığı bir tekniktir. **Kötümserlik ölçütü** yaklaşımında karar verici hangi karar alternatifini seçerse seçsin, strateji tablosunun satırlarında yer alan her bir karar alternatifinin en düşük kazanç veya en yüksek maliyet değerleri ile karşı karşıya kalacağını düşünür. Kötümserlik ölçütünde karar verici, kazanç yapılı problemlerde "kazanabileceğim en küçük kazançlardan en büyüğünü seçerim" felsefesi ile hareket ederken maliyet yapılı problemlerde "ortaya çıkabilecek en büyük maliyetlerden en küçüğünü seçerim" felsefesi ile hareket etmektedir.

Kötümserlik ölçütünde "hep kötü durumlar benim için ortaya çıkar" felsefesi benimsenir.

Kötümserlik ölçütüne göre karar vericinin benimseyeceği karar alternatifi strateji tablosuna göre, kazanç yapılı problemlerde,

$$\underset{i}{\text{EnBüyük}}\{\underset{j}{\text{EnKüçük}}(ö_{ij})\}$$

işlemi ile tespit edilirken, maliyet yapılı problemlerde ise

$$\underset{i}{\text{EnKüçük}}\{\underset{j}{\text{EnBüyük}}(ö_{ij})\}$$

işlemi ile tespit edilir. Bu işlemlere göre, eğer karar verici kazanç yapılı bir problem ile çalışıyor ise her bir karar alternatifi için strateji tablosunun satırlarında yer alan en düşük değerleri seçer ve bu değerler arasından en büyüğüne sahip değeri veren karar alternatifini en iyi karar olarak kullanır. Eğer karar verici maliyet yapılı bir problem ile çalışıyor ise her bir karar alternatifi için strateji tablosunun satırlarında yer alan en büyük değerleri seçer ve bu değerler arasından en küçüğüne sahip değeri veren karar alternatifini en iyi karar olarak kullanır.

ÖRNEK 4

Bir karar probleminde üç adet karar alternatifi ve üç adet doğal durum bulunmaktadır. Karar vericinin problemine ilişkin strateji tablosu kazanç yapılı olarak hazırlanmış ve Tablo 8.7'de sunulmuştur.

Tablo 8.7
Strateji tablosu.

	Doğal Durumlar			
		Hakem 1	Hakem 2	Hakem 3
Karar Alternatifleri	Yöntem 1	45	42	49
	Yöntem 2	52	46	51
	Yöntem 3	55	58	63

Kötümserlik ölçütünü kullanarak karar vericinin seçmesi gereken karar alternatifini belirleyelim.

Karar verici ilk olarak strateji tablosunun sonuna yeni bir sütun ekler. Bu sütunda problem kazanç yapılı olduğundan dolayı, her bir karar için ortaya çıkan en küçük değerler yer alır. Daha sonra bu sütunda yer alan en küçük kazanç değerlerinden en büyüğüne sahip olan değeri veren karar en uygun karar olarak belirlenir. Tablo 8.8'de bahsedilen değerler gösterilmiştir.

Tablo 8.8
Her bir karar alternatifi için en küçük kazanç değerleri.

	Doğal Durumlar				
		Hakem 1	Hakem 2	Hakem 3	En Küçük
Karar Alternatifleri	Yöntem 1	45	42	49	42
	Yöntem 2	52	46	51	46
	Yöntem 3	52	55	59	52

Her bir karar alternatifi için belirlenen en küçük değerlerden (42, 46, 52) en büyüğü ise kötümserlik ölçütüne göre en iyi kararı verecektir. Tablo 8.8'in son sütununda yer alan değerlerden en büyüğü 52'e eşittir ve bu değer 3 numaralı karar alternatifi için elde edilmiştir. Dolayısı ile kötümserlik ölçütüne göre karar vericinin seçmesi gereken karar alternatifi 3 numaralı (Yöntem 3) karardır.

Hurwicz'in Genelleştirilmiş İyimserlik Ölçütü

Daha önce bahsedilen iyimserlik ve kötümserlik ölçütleri, karar verme durumu ile karşı karşıya kalan karar vericinin benimseyebileceği iki uç durumu temsil etmektedir. Karar vericinin, karar probleminin doğasına ve içinde bulunduğu psikolojik duruma ilişkin daha esnek davranmasına olanak vermek amacı ile iyimserlik ve kötümserlik ölçütlerini bir arada kullanan bir ölçüt Hurwicz tarafından önerilmiştir. Hurwicz'in genelleştirilmiş iyimserlik ölçütünde karar vericinin iyimserlik düzeyini belirleyen bir katsayı bulunmaktadır. Karar vericinin iyimserlik düzeyi α ile gösterilir. Hurwicz'in iyimserlik düzeyi olan α'ın 0 ile 1 aralığında değerler aldığı varsayılır. Eğer α 1'e eşit ise karar verici iyimser, 0'a eşit ise karar verici kötümserdir.

Hurwicz'in genelleştirilmiş iyimserlik ölçütüne göre karar vericinin benimseyeceği karar, kazanç yapılı problemlerde,

$$\underset{i}{\text{EnBüyük}}\{\alpha \underset{j}{\text{EnBüyük}}(ö_{ij}) + (1 - \alpha)\ \underset{j}{\text{EnKüçük}}\ (ö_{ij})\}$$

işlemi ile tespit edilirken, maliyet yapılı problemlerde ise

$$\underset{i}{\text{EnKüçük}}\{\underset{j}{\text{EnKüçük}}(ö_{ij}) + (1 - \alpha)\ \underset{j}{\text{EnBüyük}}\ (ö_{ij})\}$$

işlemi ile tespit edilir. Yukarıdaki ifadeler incelendiğinde Hurwicz'in genelleştirilmiş iyimserlik ölçütüne göre karar verilirken strateji tablosunun sonuna üç adet sütun eklenir. Eklenen ilk sütunda her bir karar alternatifi için en büyük değerler yer alırken, eklenen ikinci sütunda ise en küçük değerler yer alır. Eklenen son sütun ise problemin kazanç veya maliyet yapılı olmasına göre önerilen formüllerden bir tanesi kullanılarak oluşturulur.

ÖRNEK 5

Bir karar probleminde dört adet karar alternatifi ve üç adet doğal durum bulunmaktadır. Karar vericinin problemine ilişkin strateji tablosu kazanç yapılı olarak hazırlanmış ve Tablo 8.9'da sunulmuştur (Hatırlanırsa bu strateji tablosu daha önce Örnek 8.2'de de kullanılmıştır).

Tablo 8.9
Strateji tablosu.

		Doğal Durumlar		
		Durum 1	Durum 2	Durum 3
Karar Alternatifleri	Karar 1	900	850	1200
	Karar 2	1000	920	880
	Karar 3	850	1120	900
	Karar 4	660	875	950

Karar vericinin yüzde 65 iyimser olduğu varsayımı altında Hurwicz'in genelleştirilmiş iyimserlik ölçütünü kullanarak karar vericinin seçmesi gereken karar alternatifini belirleyelim.

Karar verici ilk olarak strateji tablosunun sonuna üç adet sütun ekler. Sütunlarda sırası ile her karar alternatifi için en büyük, en küçük ve formül sonuçları yer alır. Problem kazanç yapılı olduğu için son sütunda yer alan değerlerden en büyüğüne sahip karar alternatifi en iyi karar alternatifi olarak benimsenir. Tablo 8.10'da gerekli hesaplamalar gösterilmiştir.

Tablo 8.10
Hurwicz'in genelleştirilmiş iyimserlik ölçütü için gerekli olan işlemler tablosu.

		Doğal Durumlar		
		En Büyük	En Küçük	$0.65\ \underset{j}{\text{EnBüyük}}\ (ö_{ij}) + (1\text{-}0.65)\ \underset{j}{\text{EnKüçük}}\ (ö_{ij})$
Karar Alternatifleri	Karar 1	1200	850	0.65 (1200) + (1-0.65) (850) = 1077.50
	Karar 2	1000	880	0.65 (1000) + (1-0.65) (880) = 958.00
	Karar 3	1120	850	0.65 (1120) + (1-0.65) (850) = 1025.50
	Karar 4	950	660	0.65 (950) + (1-0.65) (660) = 848.50

Her bir karar alternatifi için formül ile belirlenen değerler (1077.50, 958, 1025.50, 848.50) içinden en büyüğü 1077.50 değeridir. Dolayısıyla Hurwicz'in genelleştirilmiş iyimserlik ölçütüne göre karar vericinin seçmesi gereken karar alternatifi 1 numaralı karar alternatifidir.

SIRA SİZDE 2

İzleyen strateji tablosunda bir karar vericinin farklı karar alternatifleri için kazanç değerleri gösterilmiştir. İyimserlik ve kötümserlik ölçütlerine göre karar verici için en iyi karar nedir?

Doğal Durumlar			
Karar Alternatifleri		İyi	Kötü
	Proje 1	45	65
	Proje 2	75	97
	Proje 3	120	140

RİSK ALTINDA KARAR VERME

Kimi durumlarda karar problemi için tanımlanan doğal durumların ortaya çıkma olasılıkları belirlenebilir. Örneğin, bir yatırımcı yatırım dönemi içerisinde ekonomide durgunluk yaşanıp yaşanmama olasılığını geçmiş deneyimlerine dayanarak tespit edebilir. Böyle bir durumda, yatırımcının tespit ettiği olasılık değerini karar verme sürecinde kullanmak istemesi en doğal eylem olacaktır. Doğal durumların ortaya çıkma olasılıklarının biliniyor olmasının karar vericiye vereceği en büyük avantaj, karar vericinin benimsediği karar alternatifine göre elde etmeyi beklediği kazanç/maliyet için ne kadarlık bir riske girdiğini hesaplayabiliyor olmasıdır. Risk ortamında doğal durumlar için olasılıkların belirlenmesi süreci çok önemlidir. Ortaya çıkma olasılığı düşük olan bir durum için yüksek bir olasılık değerinin atanması, karar vericinin kararının yanlış olmasına neden olacaktır. Bilindiği gibi, olasılık değerleri 0 ile 1 arasında değerler almaktadır ve 0'a yakın olasılık değerlerine sahip olayların gerçekleşme şanslarının az, 1'e yakın olasılık değerlerine sahip olayların gerçekleşme şanslarının ise yüksek olduğu söylenmektedir. Risk ortamında karar verme sürecinde, strateji tablosu içerisinde yer alan tüm doğal durumların gerçekleşme şansına ilişkin bir olasılık değerinin mutlaka tanımlanması ve tüm doğal durumların olasılıklar toplamının da 1'e eşit olması gereklidir.

Doğal durumların ortaya çıkma/gerçekleşme olasılıkları belirlendikten sonra bu olasılıklar strateji tablosuna eklenirler. Daha sonra da ünitenin izleyen kesiminde verilen ölçütlerden bir tanesi kullanılarak, en iyi kararın hangisi olduğu tespit edilir.

En İyi Beklenen Değer Ölçütü

Risk ortamında karar verme durumunda en çok kullanılan ölçüt, en iyi beklenen değer ölçütüdür. Genellikle EMV kısaltması ile gösterilir. En iyi beklenen değerin hesaplanması için, her bir karar alternatifinde yer alan sonuçlar ilgili doğal durum olasılıkları ile çarpılır, daha sonra bu çarpım değerleri her satır için toplanarak ilgili karar alternatifi için beklenen kazanç veya maliyet değeri hesaplanmış olur. Daha sonra karar alternatifleri için hesaplanan beklenen değerlerin kazanç yapılı problemlerde en büyük değerli karar alternatifi, maliyet yapılı karar problemlerinde ise en küçük değerli karar alternatifi en iyi karar olarak tanımlanır.

En iyi beklenen değer ölçütünde, j'inci doğal durumun ortaya çıkma olasılığı P_j ile gösterilirse, i'inci karar alternatifinin uygulanması durumunda beklenen kazanç/maliyet,

$$B(d_i) = \sum_{j=1}^{n} P_j ö_{ij}, \qquad i = 1,\ldots,m$$

eşitliği yardımı ile hesaplanır. Eşitlik yardımı ile her bir karar alternatifinin uygulanması durumunda ortaya çıkması beklenen kazanç/maliyetler, strateji tablosunun son sütununda hesaplanarak gösterilir. Daha sonra kazanç yapılı karar problemlerinde, her bir karar alternatifi için tespit edilen $B(d_i)$ değerlerinden en büyüğüne sahip olan karar alternatifi en iyi karar alternatifi olarak belirlenir. Eğer maliyet yapılı karar problemi söz konusu ise her karar alternatifi için tespit edilen $B(d_i)$ değerlerinden en küçüğüne sahip olan karar alternatifi en iyi karar alternatifi olarak belirlenir.

ÖRNEK 6

Bir karar verici gayrimenkul yatırımı yapmak istemektedir. Karar verici önümüzdeki dönem için ekonominin %65 olasılık ile iyi durumda, %35 olasılıkla ise kötü durumda olacağını tahmin etmektedir. Yatırımcı üç farklı gayrimenkul türüne yatırım yapma seçeneği olduğunu tespit etmiştir. Yatırımcının farklı yatırım seçenekleri ve doğal durumlar için beklediği kazanç değerleri (₺ olarak) Tablo 8.11'de gösterilmiştir.

Tablo 8.11 *Gayrimenkul yatırımı strateji tablosu*

		Doğal Durumlar	
		İyi Ekonomik Durum P(E+)=0.65	Kötü Ekonomik Durum P(E-)=0.35
Karar Alternatifleri	Depo alımı	40000	5000
	Ofis alımı	90000	10000
	Apartman dairesi alımı	50000	40000

En iyi beklenen değer ölçütüne göre karar vericinin kararını vermesine yardımcı olalım.

Tablo 8.11 incelendiğinde karar probleminde karşılaşılabilecek tüm doğal durumların tespit edildiği, bu doğal durumların gerçekleşme şanslarını göstermek üzere doğal durumların ortaya çıkış olasılıklarınında verildiği görülmektedir. Dolayısıyla bu karar probleminde karar verici risk altında karar vermek durumundadır. En iyi karar alternatifini tespit etmek amacı ile her karar alternafinin beklenen değeri hesaplanır. Hesaplanan değerler Tablo 8.12'de sunulmuştur.

Tablo 8.12 *Gayrimenkul yatırımı strateji tablosu ve beklenen değerler*

		Doğal Durumlar		
		İyi Ekonomik Durum P(E+)=0.65	Kötü Ekonomik Durum P(E-)=0.35	Beklenen Değerler
Karar Alternatifleri	Depo Alımı	40000	5000	0.65 (40000)+ 0.35 (5000) = 27750
	Ofis Alımı	90000	10000	0.65 (90000) + 0.35 (10000)= 62000
	Apartman dairesi alımı	50000	40000	0.65 (50000)+ 0.35 (40000)=46500

Karar vericinin bu yatırım probleminde kazanç ile ilgilendiğinden dolayı yatırımcının hesaplanan beklenen değerler (27750, 62000, 46500) içinden en büyüğünü seçmesi gerekir. Dolayısıyla bu karar probleminde en yüksek beklenen değer ₺62000'dir ve yatırımcımızın seçmesi gereken karar alternatifi (yatırım alternatifi) ofis alımıdır.

ÖRNEK 7

Bir elektronik mağazası işletmesi yeni mağaza açmayı planlamaktadır. Firma yetkililerinin yaptıkları bir Pazar araştırmasına göre yeni mağazanın açılabileceği üç farklı şehir olduğu (Eskişehir, Adana ve Trabzon) tespit edilmiştir. Firma inşaat maliyetlerini de araştırmış ve inşaat yapılması planlanan dönemde demir, çimento fiyatlarında bir yükselme olabileceği öngörüsüne sahiptir. Malzeme fiyatlarının artması ve artmaması durumuna göre hesaplanan maliyetlere ilişkin strateji tablosu oluşturularak Tablo 8.13'te gösterilmiştir. En iyi beklenen değer ölçütüne göre karar verici kangi karar alternatifini seçmelidir?

Tablo 8.13
Risk altında karar problemi için strateji tablosu

		Doğal Durumlar	
		Durum 1	Durum 2
		P(D1)=0.70	P(D2)=0.30
Karar Alternatifleri	Eskişehir	125000	165000
	Adana	175000	185000
	Trabzon	145000	162000

Tablo 8.13 incelendiğinde doğal durumların ortaya çıkma olasılıkları verildiği için karar vericinin risk altında karar verme durumunda olduğu gözlenebilir. Dolayısıyla her satır için beklenen değer, sonuç değerlerinin doğal durumların ortaya çıkma olasılıkları ile çarpılması ve bu değerlerin toplanması yolu ile bulunabilir. Eskişehir karar alternatifi için beklenen değer 0.70 (125000) + 0.30 (165000) = 137000, Adana karar alternatifi için beklenen değer 0.70 (175000) + 0.30 (185000) = 178000 ve Trabzon karar alternatifi için beklenen değer 0.70 (145000) + 0.30 (162000) = 150100 olacaktır. Bu değerlerden en küçüğü olan 137000 değeri Eskişehir karar alternatifi için ortaya çıktığından dolayı karar vericinin en iyi kararı Eskişehir alternatifidir.

En Büyük Olasılık Ölçütü

Daha önce belirsizlik ortamında ele alınan iyimserlik ve kötümserlik ölçütlerinin karar vericinin içinde bulunduğu psikolojik durumu yansıtan ölçütler olduğu belirtilmişti. Benzer şekilde bir karar probleminde risk altında karar verme süreci ile karşı karşıya kalan karar verici, eğer tüm doğal durumlar üzerinden hesaplama yapmak yerine ortaya çıkma şansı bir başka deyişle gerçekleşme olasılığı en yüksek olan doğal duruma göre hareket ederse, en büyük olasılık ölçütüne göre karar vermiş olur. Böyle bir durumda karar verici yalnızca en yüksek olasılıklı doğal durumun sonuçları ile ilgilenir.

Eğer karar problemi kazanç yapılı ise en yüksek olasılığa sahip doğal durum için karar alternatiflerinin sonuç değerlerinden en yükseğini veren karar alternatifi, en iyi karar olarak kabul edilir. Karar problemi maliyet yapılı ise en yüksek olasılığa sahip doğal durum sonuçları içinde en küçük değeri veren karar alternatifi, en iyi karar olarak ele alınır. En büyük olasılık ölçütü, hesaplaması en kolay ölçütlerden biri olmakla beraber, yalnızca bir tek doğal durumu gözönüne aldığından dolayı, diğer ölçütlere göre daha zayıf bir karar verme ölçütüdür. Unutulmamalıdır ki bir olayın olasılığının (doğal durumun olasılığının) düşük olması bu olayın gerçekleşmeyeceği (doğal durumun ortaya çıkmayacağı) anlamı taşımaz. Bunun aksine, olayın gerçekleşme olasılığının var olduğunu ancak diğer alternatiflere göre ortaya çıkma şansının düşük olduğunu gösterir.

ÖRNEK 8

Yeni bir elektronik alet için üç farklı dizayn düşünülmektedir. Üründen elde edilecek kazançların pazarın genel durumuna ve dizayna göre değişeceği düşünülmektedir. Karar verici üç farklı pazar koşulunun ortaya çıkabileceğini düşünerek Tablo 8.14'te yer alan kazanç yapılı strateji tablosunu oluşturmuştur. En büyük olasılık ölçütünü kullarak karar vericinin kararını vermesinde yardımcı olalım.

Tablo 8.14
Dizayn-Pazar koşulları strateji tablosu

		Pazar Koşulları		
		I P(I)=0.65	II P(II)=0.15	III P(III)=0.20
Karar Alternatifleri	Klasik Dizayn	56000	57000	69000
	Modern Dizayn	53000	42000	54000
	Alternatif Dizayn	65000	40500	41000

Tablo 8.14'te yer alan Pazar koşul olasılıkları incelendiğinde en büyük olasılık değerinin 0.65'e eşit olduğu görülmektedir. Dolayısıyla en büyük olasılık değerine sahip olan I numaralı Pazar koşulunun ortaya çıkacağı, karar verici tarafından varsayılır. I numaralı Pazar koşulunun gerçekleşmesi durumunda elde edilecek en yüksek kazanç değeri olan 65000, alternatif dizayn için elde edildiğinden, bu problemde en büyük olasılık ölçütüne göre en iyi karar alternatif dizayndır. Ancak strateji matrisi dikkatli bir şekilde incelendiğinde tablodaki en büyük kazanç değerinin 69000 olduğu ve bu değerin III numaralı Pazar koşulunun ortaya çıkması ve Klasik dizayn seçilmesi hâlinde elde edilebileceği görülmektedir. III numaralı Pazar koşulunun ortaya çıkma olasılığı 0.20 gibi çok da küçük sayılmayacak bir değerdir. Dolayısıyla bu problemde en iyi beklenen değer ölçütüne göre de hesaplama yapılması çok daha faydalı olacaktır. Eğer en iyi beklenen değer ölçütüne göre hesaplamalar yürütülürse en iyi kararın, 58750 değeri ile Klasik dizayn için elde edileceği hesaplanabilir.

TAM BİLGİNİN BEKLENEN DEĞERİ

Karar vericinin karar alternatiflerinden birini seçmeden önce, ortaya çıkabilecek doğal durumlardan hangisinin gerçekten olacağını öğrenebileceğini varsayalım. Eğer karar verici hangi doğal durumun ortaya çıkabileceğini kesinlikle öğrenirse, ilgili doğal durum içerisinde en yüksek kazanç veya en düşük maliyeti veren karar seçeneğini belirleyerek, en iyi kararı vermiş olur. Tam bilginin beklenen değeri, böyle bir bilginin (hangi doğal durumun kesinlikle ortaya çıkacağı bilgisinin) elde edilmesi ile karar vericinin elde edeceği ilave kazancı temsil etmektedir. Genellikle EVPI kısaltması ile gösterilir.

Tam bilginin beklenen değeri tanımı, hangi doğal durumun ortaya çıkacağının kesinlikle öğrenilmesiyle elde edilebilecek en iyi beklenen değer ile hangi doğal durumun kesinlikle ortaya çıkacağı bilgisi öğrenilmeden önceki en iyi beklenen değer arasındaki fark olarak verilebilir.

Eğer karar verici hangi doğal durumun ortaya çıkacağını kesinlikle öğrenmek isterse bu bilgiye ödeyeceği en büyük miktarı bilmek zorundadır. Örneğin, karar vericinin bir karar probleminde karara ilişkin en iyi beklenen değeri ₺25000 ve doğal durumlardan hangisinin kesinlikle ortaya çıkacağını bilmesi durumunda en iyi beklenen kazanç değeri ₺42000 ise hangi doğal durumun ortaya çıkacağını kesinlikle öğrenmek için karar verici en fazla ₺42000 - ₺25000 = ₺17000 kadarlık bir ödemeyi göz önüne alabilecektir. ₺17000'lik değerden daha az yapılan her ödeme miktarında karar verici, eski duruma göre daha iyi durumda olacaktır. Eğer bir dış kaynak hangi doğal durumun ortaya çıkacağını söylemek için ₺10000 talep ederse karar verici bu miktarı ödemelidir. Sonuç olarak bilgi sayesinde ekstradan kazanmayı beklediği ₺17000'ndan ₺10000 çıkartılırsa, karar verici ilk durum olan ₺25000'lik kazanca göre ₺7000'lık daha ekstra kazanç sağlamaktadır.

Örneğin, dört doğal durumun olduğu bir karar probleminde karar vericinin doğal durumlar için belirlediği olasılıklar sırasıyla 0.25, 0.20, 0.35 ve 0.20 olsun. Bu karar vericinin üç farklı karar alternatifi olduğunu varsayalım. Örnek strateji tablosu Tablo 8.15'te sunulmuştur. Bu karar problemi kazanç yapılı olarak oluşturulmuştur. Karar verici öncelikle strateji tablosunun sonuç değerlerini belirler ve her bir karar alternatifi için beklenen değerleri hesaplar. Tablo 8.15'e göre beklenen değerlerden en yüksek olanı 18000 değeridir. Dolayısıyla karar vericinin vermesi gereken en iyi karar 3 numaralı karar alternatifidir. En iyi beklenen değer ölçütüne göre karar vericinin 3 numaralı karar alternatifini seçmesi hâlinde beklenen kazanç değeri 18000 olacaktır.

Tablo 8.15
Tam bilginin beklenen değeri için örnek bir strateji tablosu

		Doğal Durumlar				
		Durum 1 P(D1)=0.25	Durum 2 P(D2)=0.20	Durum 3 P(D3)=0.35	Durum 4 P(D4)=0.20	Beklenen Değerler
Karar Alternatifleri	Karar 1	11000	12000	16000	18000	14350
	Karar 2	14000	24000	5000	14000	12850
	Karar 3	40000	16000	8000	10000	18000

Bir dış kaynaktan hangi doğal durumun ortaya çıkacağı bilgisine ulaşılması mümkün olduğunu varsayalım. Eğer karar verici doğal durumu kesinlikle bilirse o zaman o doğal durumda en büyük kazancı veya en küçük maliyeti veren alternatifi seçecektir. Dolayısıyla karar verici bu aşamada "her doğal durum kesinlikle ortaya çıkarsa, hangi karar alternatifini seçerdim" araştırması yapar. Birinci doğal durum gerçekten ortaya çıkarsa elde edilecek en büyük kazanç 40000'dir, benzer şekilde diğer durumlar için elde edilecek en büyük kazanç değerleri sırasıyla 24000, 16000, ve 18000'dir. Şimdi bu değerlerin her birini ilgili doğal durum olasılıkları ile çarpıp, çarpımların genel toplamını alarak, tam bilginin varlığında elde edilmesi beklenen kazaç değeri hesaplanabilir. Bu değer

$$0.25\ (40000) + 0.20\ (24000) + 0.35\ (16000) + 0.20\ (18000) = 24000$$

olacaktır. Hatırlanırsa tam bilgiyi elde etmeden önce karar vericinin beklenen kazanç değeri 18000'dir. Dolayısıyla karar verici için tam bilginin beklenen değeri

$$\text{Tam Bilginin Beklenen Değeri} = 24000 - 18000 = 6000$$

olacaktır. Karar verici tam bilgi altında 6000 daha fazla kazanç elde edecektir. Dolayısıyla tam bilgiyi elde etmek için karar vericinin harcamayı göze alacağı en yüksek rakam 6000'dir. Bu değerden daha fazla bir harcama yapılması daha önce kazanmayı beklediği 18000'den daha az kazanca yol açacaktır.

ÖRNEK 9

Yeni bir elektronik alet için üç farklı dizayn düşünülmesine ilişkin karar problemi Örnek 8.8'de incelenmiştir. Hatırlanırsa karar verici en iyi beklenen değer ölçütüne göre hesaplamalarını yürütürse, en iyi kararın 58750 değeri ile Klasik dizayn için elde edileceği görülmektedir. Karar vericinin hangi Pazar koşulunun gerçekten ortaya çıkacağını bilmesi durumuna ilişkin tam bilginin beklenen değerini hesaplayalım ve yorumlayalım. Tablo 8.16'da strateji tablosu hatırlatma amaçlı olarak tekrar sunulmuştur.

Tablo 8.16 *Dizayn-Pazar koşulları strateji tablosu*

		Pazar Koşulları		
		I P(I)=0.65	II P(II)=0.15	III P(III)=0.20
Karar Alternatifleri	Klasik Dizayn	56000	57000	69000
	Modern Dizayn	53000	42000	54000
	Alternatif Dizayn	65000	40500	41000

Karar vericinin gerçekleştirmesi beklenen ilk işlem her Pazar koşulu için en büyük kazanç değerlerini belirlemektir. Bu değerler sırası ile 65000, 57000 ve 69000 değerleridir. Daha sonra her bir en büyük değer ilgili Pazar koşulunun ortaya çıkma olasılıkları ile çarpılıp genel toplam hesaplanır. Tam bilgi durumunun ortaya çıkması hâlinde karar vericinin elde edebileceği beklenen kazancı gösteren bu değer,

$$0.65\ (65000) + 0.15\ (57000) + 0.20\ (69000) = 64600$$

olacaktır. Daha sonra tam bilgi durumunun gerçekleşmesi hâlinde beklenen değer ile tam bilgi elde edilmeden önceki beklenen değer farkı,

$$64600 - 58750 = 5850$$

olacaktır. Dolayısıyla karar verici için tam bilginin beklenen değeri 5850 olacaktır. Karar vericinin bu aşamada tam bilgiyi elde etmek için bir çaba gösterip göstermeyeceğine karar vermesi beklenir. Karar verici tam bilgi durumunu göz ardı ederse 58750 kazanç beklentisi içerisinde iken tam bilgi sayesinde en çok 5850'lik bir kazanca sahip olabilir. Önemli olan karar vericinin tam bilgi için ödeyeceği miktar ile tam bilginin beklenen değeri arasındaki farkın, yapılacak olan ekstra çabaya deyip deymeyeceğine karar vermektir.

SIRA SİZDE 3

Bir karar verme probleminde karar verici tam bilgiyi elde etmeden önce beklenen kazanç değerinin 21000 olduğunu tespit etmiştir. Ancak bir dış kaynak karar vericiye hangi doğal durumun kesinlikle ortaya çıkacağı bilgisini verebileceğini ve bu bilginin karar verici için bir maliyeti olacağını bildirmiştir. Karar verici böyle bir bilgiyi elde ettiği zamanki beklenen kazanç değerini hesaplamış ve 21250 olarak belirlemiştir. Tam bilginin beklenen değeri ne kadardır? Sizce karar verici tam bilgiyi elde etme yoluna başvurmalımıdır?

KARAR AĞACI

Karar verme durumu ile karşı karşıya kalan bir karar vericinin kullanabileceği bir başka araçta, karar problemi için karar ağacının oluşturulmasıdır. Karar ağaçları karar probleminin grafiksel olarak gösterimidir. Ayrıca karar ağaçları karar vericinin içinde bulunduğu karar verme probleminde ortaya çıkabilecek tüm durumları, senaryoları, bir arada görebileceği bir grafiksel yaklaşımdır. Pek çok kişi, derlenen verilerin kendilerine tablolar şeklinde sunulması yerine, grafiklerle sunulması durumunda daha kolay, daha ayrıntılı ve net bilgiyi elde edebildiklerini belirtmektedirler. Özellikle birden fazla kararın ardı ardına alınmasını gerektiren durumlarda strateji tablosu yetersiz kalacaktır. Bu tür ardışık karar verme problemlerinde, karar problemi için karar ağacının oluşturulması en iyi kararların belirlenmesi işlemini kolaylaştıracaktır. Karar ağaçları, risk altında karar verme durumunda bulunan karar vericiler tarafından oluşturulurlar.

Karar ağaçları temel olarak izleyen bileşenlerden oluşur.

Karar Düğümü: Karar düğümünde karar vericinin çeşitli karar alternatiflerinden birisini seçmesi beklenir. Strateji tablosunun satırlarında yer alan karar alternatifleri bu düğümde kullanılır. Karar düğümleri kare sembolü ile gösterilir. Bu karar düğümü karar ağacının başlangıç düğümüdür. Bu düğümden sağa doğru karar alternatifi sayısı kadar doğrular (dallar) çizilir. Her bir karar alternatifini temsil eden çizginin üzerine o karar alternatifinin adı yazılır.

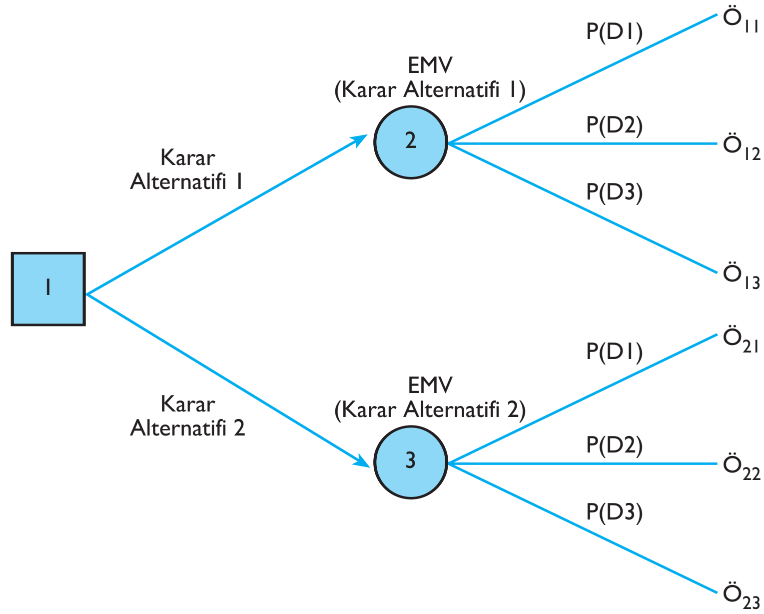
Şans Düğümü: Daha önce strateji tablosunda incelediğimiz doğal durumları göstermek amacı ile şans düğümü oluşturulur. Şans düğümleri (veya şans çatalları) daire sembolü ile gösterilir. Şans düğümünden sağa doğru doğal durum sayısı kadar doğrular (dallar) çizilir. Şans düğümünden sağa doğru çizilen doğrulara şans dalı adı verilir. Şans dalının sonunda ise strateji tablosunun hücrelerinde yer alan sonuç değerleri yer alacaktır. Eğer birden fazla karar verme söz konusu ise bir şans dalından sonra bir başka karar düğümü yer alabilir. Şans dalının üzerinde ise o şans dalının temsil ettiği doğal durumun ortaya çıkma olasılığı yer alır. Her şans düğümünün üzerinde ise o şans dalına ulaşmak için seçilmesi gereken karar alternatifi için hesaplanan EMV değerleri (en iyi beklenen değerler) yer alacaktır.

Şekil 8.1'de iki adet karar alternatifi ve üç adet doğal duruma sahip bir karar problemi için oluşturulmuş karar ağacı örneği yer almaktadır.

Şekil 8.1 kullanılarak karar ağacının nasıl oluşturulacağını ele alalım. Karar ağacı çizimine en sola bir karar düğümü çizilmesi ile başlanır. Şekil 8.1'de 1 numara ile gösterilen kare, bir karar düğümüdür. Daha sonra bu düğümden karar alternatifi sayısı kadar dal sağa doğru yönlendirilir. Şekil 8.1'e göre bu karar probleminde iki adet karar alternatifi bulunmaktadır. Her karar alternatifinin sonuna bir şans düğümü oluşturulur. Şekil 8.1'de 2 ve 3 numara ile gösterilen daireler şans düğümleridir. Bu şans düğümleri (ya da şans çatallarında) karar problemindeki doğal durum sayısı kadar dalı içerir. Şekil 8.1'e göre bu karar probleminde üç adet doğal durum olduğundan dolayı her bir karar alternatifi için üç adet şans dalı bulunmaktadır. Her şans dalı sonunda ise strateji tablosunun sonuç değerleri yer alır. Şekil 8.1.'e göre öi1 sonuç değeri, karar vericinin 1 numaralı kararı vermesi ve 1 numaralı doğal durumun ortaya çıkması ile elde edilecek kazanç/maliyet değeridir. Benzer şekilde diğer sonuç değerleri de strateji tablosundan karar ağacına aktarılır. Şekil 8.1.'in oluşturulması için kullanılan strateji matrisi Tablo 8.17'de sunulmuştur.

Şekil 8.1

Bir Karar Ağacı örneği



Tablo 8.17

Şekil 8.1'in çizilmesi için oluşturulan strateji tablosu

		Doğal Durumlar			
		Durum 1 P(D1)	Durum 2 P(D2)	Durum 3 P(D3)	Beklenen Değerler
Karar Alternatifleri	Karar Alternatifi 2	ö11	ö12	ö13	EMV(Karar1)
	Karar Alternatifi 1	ö21	ö22	ö23	EMV(Karar2)

Karar ağaçlarında en iyi karar alternatifinin hangisi olduğunu göstermek üzere karar düğümünden çıkan en iyi karar alternatifine bir ok işareti konulur. Şekil 8.1'de verilen karar ağacı örneğinde en iyi karar alternatifi 1 numaralı karar alternatifi olarak örneklenmiş ve ok işareti bu karar dalına eklenmiştir.

Karar ağaçları çizimleri kolay olan ve karar vericiye en iyi karar alternatifinin seçiminde görsel olarak yardımcı olan grafiklerdir. Karar vericinin içinde bulunduğu karar probleminin modellenmesinde önemli bir araçtır. Karar ağacının en soldan en sağa kadar ilerleyerek bulunan her bir dal sırası o karar problemi için bir senaryo olacaktır. Dolayısıyla bir karar ağacı karar probleminde ortaya çıkabilecek tüm senaryoları bir arada göstermeye yarayan bir grafiksel gösterimdir. Ancak karar ağaçlarının en büyük dezavantajı, karar alternatifi ve doğal durum sayısının artması durumunda karar ağacının karmaşık bir yapı alabilmesidir.

ÖRNEK 10

Bir firma hali hazırda sahip oldukları ürün gamına iki yeni ürünü daha eklemeyi düşünmektedir. Firma yöneticileri ürün gamına iki ürünü birden ekleme, iki üründen birisini ekleme veya hiçbirini eklememe karar seçenekleriyle karşı karşıyadır. Ürünlerin başarıları müşterilerin gösterecekleri tepkilere bağlıdır. Daha önceden yapılan çalışmalara göre müşterilerin ürünlere gösterdikleri tepkiler 0.30 olasılıkla "iyi", 0.50 olasılıkla "orta" ve 0.20 olasılıkla "kötü" olarak sınıflandırılmaktadır. Tablo 8.18'de firma tarafından hazırlanan tahmini kazanç değerleri strateji tablosu sunulmuştur.

Tablo 8.18
Ürün piyasaya sürme kararı strateji tablosu (10000'in katları)

		Doğal Durumlar		
		İyi P(İ)=0.30	Orta P(O)=0.50	Kötü P(K)=0.20
Karar Alternatifleri	Hiçbir ürünü ekleme	0	0	0
	Sadece birinci ürünü ekle	125	65	30
	Sadece ikinci ürünü ekle	105	60	30
	Her iki ürünü de ekle	220	110	40

En iyi beklenen değer ölçütüne göre firma yöneticilerinin karar vermesinde yardımcı olarak karar ağacını oluşturalım.

En iyi beklenen değer ölçütüne göre çözüm yapılacağından her bir karar alternatifi için beklenen kazanç değerleri hesaplanır. Tablo 8.19'da her bir karar alternatifi için beklenen değerler gösterilmiştir.

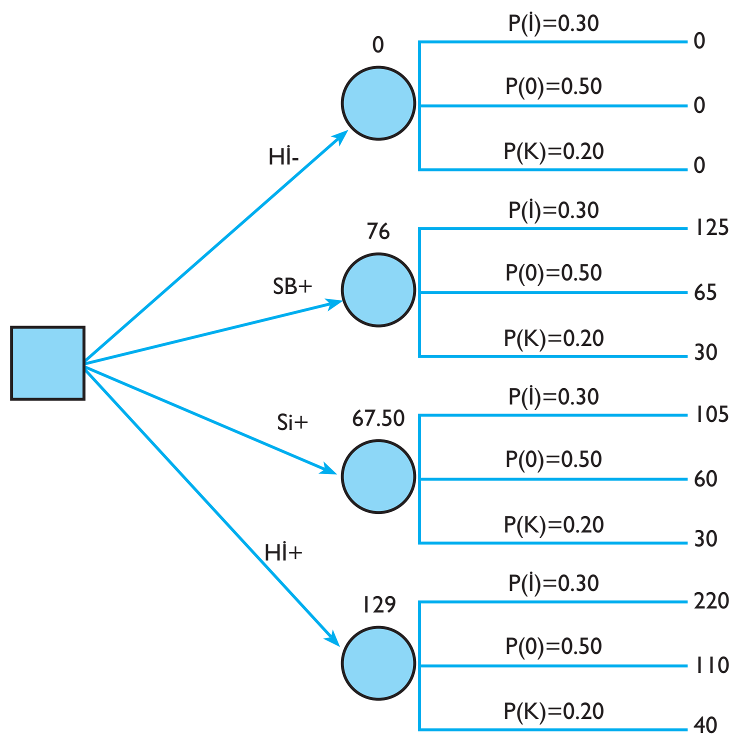
Tablo 8.19
Ürün piyasaya sürme kararı strateji tablosu ve beklenen değerler (10000'in katları)

		Doğal Durumlar			
		İyi P(İ)=0.30	Orta P(O)=0.50	Kötü P(K)=0.20	Beklenen değerler
Karar Alternatifleri	Hiçbir ürünü ekleme (Hİ-)	0	0	0	0
	Sadece birinci ürünü ekle (SB+)	125	65	30	76.00
	Sadece ikinci ürünü ekle (Sİ+)	105	60	30	67.50
	Her iki ürünü de ekle (Hİ+)	220	110	40	129.00

Beklenen değerler incelendiğinde firma yöneticileri için en iyi karar en büyük beklenen değer olan 129'a sahip her iki ürününde ürün gamına eklenmesi karar alternatifi olduğu görülmektedir. Ayrıca bu karar probleminde tam bilginin beklenen değeri sıfıra eşittir. Problemin karar ağacı Şekil 8.2'de gösterilmiştir.

Şekil 8.2

Tablo 8.19 için karar ağacı



Şekil 8.2'de yer alan karar ağacı incelendiğinde okun gösterdiği her iki ürünün ürün gamına eklenmesi seçeneğinin 129 beklenen kazanç değeri ile en iyi karar olduğu görülmektedir.

Özet

AMAÇ 1

Karar problemlerinde strateji tablosu oluşturmak

Bir karar probleminin çözümüne ilişkin birden fazla karar seçeneği ile karşı karşıya kalan bir karar vericinin, her bir karar alternatifi için elde edeceği kazanç veya maliyetleri belirlemesi gerekir. Eğer karar vericinin karar alternatifleri çeşitli dış etmenlerden etkileniyor ise bu etmenlerin beklenen kazanç ve maliyet değerleri üzerindeki etkisi de araştırılmalıdır. Örneğin, hafta sonunu geçirecek bir aktivite planlayan bir karar verici için hafta sonu boyunca hava koşullarının uygun olup olmayacağı bilgisi karar alternatiflerinizin sonuçlarını değiştirebilir, “yağmurlu bir havada pikniğe giderseniz ıslanırsınız gibi”. İyi bir karar vericinin karar problemine ilişkin tüm karar alternatiflerini belirlemesi ve bu karar alternatifleri üzerinde etkisi olabilecek doğal durumları belirleyerek strateji tablosu altında bir araya getirmesi gerekir.

AMAÇ 2

Belirsizlik altında karar vermek

Karar verici strateji tablosunu oluştururken tespit ettiği doğal durumlar hakkında mümkün olan en çok bilgiyi toplamalıdır. Ancak kimi durumlarda doğal durumların neler olabileceği bilgisi dışında bir bilgiye ulaşılamaz. Dolayısıyla karar alternatifleri üzerinde etki yaratabilecek hangi doğal durumun gerçekleşeceği tam olarak bilinemediğinden, bir başka deyişle belirsiz olduğundan bu tür karar verme problemlerinde belirsizlik altında karar verme söz konusu olacaktır. Belirsizlik altında karar verme teknikleri genel olarak karar vericinin içinde bulunduğu psikolojik yapıyı yansıtacak şekilde hazırlanmış tekniklerdir.

AMAÇ 3

Risk altında karar vermek

Karar vericinin oluşturduğu strateji tablosundaki doğal durumların gerçekleşme şanslarına ilişkin olasılık değerleri tespit edilebilirse karar verici herhangi bir karar alternatifini uygulamaya koyduğunda ne kadarlık bir riski de göz önüne aldığını tespit edebilir. Doğal durumların olasılıklarının bilindiği karar problemlerinde karar verici risk altında karar vermek durumundadır.

AMAÇ 4

Karar ağacı çizmek

Bazı durumlarda karar verici karar probleminde karşılaşabileceği bütün farklı durumları (senaryoları) bir grafik üzerinde görmek isteyebilir. Bu amaç ile karar ağaçları oluşturulur. Karar ağaçlarının çizimleri kolay olmakla beraber, karar vericinin karar problemine ilişkin tüm detayları görmesine olanak tanıdığı için çok kullanışlıdır. Ancak karar problemindeki karar alternatif sayısı ve doğal durum sayısının fazla olması durumunda karar ağaçları karmaşık bir yapıya da kolaylıkla sahip olabilir.

Kendimizi Sınayalım

1. Aşağıdakilerden hangisi karar verme sürecinin bileşeni **değildir?**

a. karar alternatifi
b. mevsim İndeksi
c. doğal durum
d. doğal durumların ortaya çıkma olasılığı
e. beklenen değer

2. Bir karar probleminde karar vericinin tercih etmesi gereken beş adet karar alternatifi vardır. Karar verici, her bir alternatif için beklenen değerleri tespit etmiştir. Karar alternatif sırasına göre bahsedilen değerler 12, 21, 17, 28 ve 30'dur. En iyi beklenen değer ölçütüne göre karar verici hangi kararı vermelidir.

a. ikinci karar alternatifini
b. üçüncü karar alternatifini
c. beşinci karar alternatifini
d. verilen değerler ile hesaplama yapılamaz
e. tüm kararları uygulayabilir

3. Birden fazla seçenek içerisinden seçim yapma işlemine ne ad verilir?

a. sonuç
b. karar verme
c. karar alternatifi
d. doğal durum
e. karar verici

4. Karar vericinin kontrolü altında olmayan ve karar probleminde yer alan karar alternatiflerinin sonuçlarını etkilediği düşünülen karar bileşenine ne ad verilir?

a. sonuç
b. en yüksek değerli karar alternatifi
c. karar verici
d. doğal durum
e. olasılık

5. Karar vericinin içinde bulunduğu genel psikolojik duruma göre daha esnek davranmasına olanak veren karar ölçütü aşağıdakilerden hangisidir?

a. karar ağacı
b. kötümserlik ölçütü
c. en büyük olasılık ölçütü
d. en iyi beklenen değer ölçütü
e. Hurwicz'in genelleştirilmiş iyimserlik ölçütü

6. Aşağıda yer alan seçeneklerden hangisi kesinlikle **yanlıştır?**

a. doğal durumlar gelecekte ortaya çıkması mümkün olan gerçek olayları gösterir
b. iyimserlik ölçütünde her zaman iyi durumların ortaya çıkacağı benimsenir
c. doğal durumların ortaya çıkma olasılıkları biliniyorsa risk ortamında karar verme söz konusudur
d. strateji tabloları kazanç yapılı kurulur
e. karar ağaçları, karar problemlerinin grafiksel gösterim biçimidir.

7. Bir araştırmada izleyen strateji tablosu elde edilmiştir. Bu strateji tablosundan faydalanarak iyimserlik ölçütüne göre en iyi karar aşağıdaki seçeneklerden hangisidir?

		Doğal Durumlar		
		I	II	III
Karar Alternatifleri	Alternatif 1	50	80	54
	Alternatif 2	45	75	65
	Alternatif 3	90	82	110

a. birinci alternatif
b. ikinci alternatif
c. üçüncü alternatif
d. hem ikinci hem de üçüncü alternatif
e. hesaplanamaz

8. Bir araştırmada izleyen strateji tablosu kazanç yapılı olarak elde edilmiştir. Bu strateji tablosundan faydalanarak kötümserlik ölçütüne göre en iyi karar aşağıdaki seçeneklerden hangisidir?

		Doğal Durumlar		
		I	II	III
Karar Alternatifleri	Alternatif 1	120	110	55
	Alternatif 2	95	70	65
	Alternatif 3	65	45	115

a. birinci alternatif
b. ikinci alternatif
c. üçüncü alternatif
d. hem ikinci hem de üçüncü alternatif
e. hesaplanamaz

9. Karar vericinin karar probleminde risk ortamında kullanılan teknikleri uygulayabilmesi için aşağıdakilerden hangisine kesinlikle ihtiyacı vardır?
a. doğal durumlar
b. karar alternatifleri
c. strateji tablosu
d. doğal durum olasılıkları
e. kötümserlik ölçütü

10. Aşağıdakilerden hangisi karar ağacı bileşeni **değildir?**
a. alternatif düğüm
b. karar düğümü
c. şans düğümü
d. karar dalı
e. doğal durum olasılıkları

Kendimizi Sınayalım Yanıt Anahtarı

1. b	Yanıtınız yanlış ise "Karar Problemi Bileşenleri" konusunu yeniden gözden geçiriniz.
2. c	Yanıtınız yanlış ise "En İyi Beklenen Değer" konusunu yeniden gözden geçiriniz.
3. b	Yanıtınız yanlış ise "Karar Problemi Bileşenleri" konusunu yeniden gözden geçiriniz.
4. d	Yanıtınız yanlış ise "Karar Problemi Bileşenleri" konusunu yeniden gözden geçiriniz.
5. e	Yanıtınız yanlış ise "Belirsizlik Altında Karar Verme" konusunu yeniden gözden geçiriniz.
6. d	Yanıtınız yanlış ise "Ünite Girişi" başlıklı konusunu yeniden gözden geçiriniz.
7. e	Yanıtınız yanlış ise "İyimserlik Ölçütü" konusunu yeniden gözden geçiriniz.
8. b	Yanıtınız yanlış ise "Kötümserlik Ölçütü" konusunu yeniden gözden geçiriniz.
9. d	Yanıtınız yanlış ise "Risk Altında Karar Verme" konusunu yeniden gözden geçiriniz.
10. a	Yanıtınız yanlış ise "Karar Ağacı" konusunu yeniden gözden geçiriniz.

Sıra Sizde Yanıt Anahtarı

Sıra Sizde 1

Karar vericinin karar alternatifleri dizüstü bilgisayar, tablet bilgisayar ve cep telefonudur. Bu karar alternatiflerinin karar vericinin genel memnuniyeti üzerinde farklı etkileri olabilir. Ancak burada üzerinde durulması gereken nokta her karar alternatifini etkileyebilecek ve karar vericinin genel memnuniyetine direk etki edecek faktörlerin, bir başka deyişle doğal durumların bulunması gerekir. Karar vericimiz bugüne yönelik bir alım yapacağı için karar verme tekniklerine ihtiyaç duymaksızın genel memnuniyetini en büyükleyecek ürünü seçebilir. Bu problemde doğal durumların oluşturulması çok zordur.

Sıra Sizde 2

Bu karar probleminde iyimserlik ve kötümserlik ölçütlerine göre karar verilmek istenmektedir. Dolayısı ile her bir karar alternatifi için en küçük ve en büyük değerleri belirleyelim. Bahsedilen bu değerler izleyen Tablo'da sunulmuştur.

		Doğal Durumlar			
		İyi	Kötü	En Büyük	En Küçük
Karar Alternatifleri	Proje 1	45	65	65	45
	Proje 2	75	125	125	75
	Proje 3	65	140	140	65

En iyi kararı tespit etmek amacıyla iyimserlik ölçütü kullanıldığında, her bir karar alternatifinin en büyük değerlerinden en büyüğü olan 140'a sahip proje 3 en iyi karar olarak benimsenirken; kötümserlik ölçütü kullanıldığında her bir karar alternatifinin en küçük değerlerinden en büyüğü olan 75'e sahip proje 2 en iyi karar olarak benimsenecektir. Görüldüğü gibi karar vericinin iki farklı psikolojik durumu, iki farklı karar alternatifinin en iyi karar alternatifi olabileceği sonucunu ortaya çıkarmıştır. Dolayısıyla karar vericinin içinde bulunduğu psikolojik durumu iyi değerlendirip karar vermesi tavsiye edilmelidir.

Sıra Sizde 3

Bu karar probleminde karar verici tam bilginin beklenen değerini hesaplamak için gerekli tüm bilgiye sahiptir. Tam bilginin beklenen değerini hesaplayabilmemiz için, tam bilgi altında en iyi beklenen değer ile tam bilgiyi elde etmeden önceki en iyi beklenen değer arasındaki cebirsel farka bakmalıyız. Karar verici için bu değerler sırasıyla 21250 ve 21000'dir. Bu iki değer arasındaki fark sadece 250'dir. Eğer karar problemimizdeki kazanç değeri parasal bir değer ise kolaylıkla görülebilir ki tam bilgiyi elde etmenin karar verici için çok büyük bir kazanç getirmesi beklenemez. Ek olarak tam bilgi için bir ödeme yapılması da söz konusudur. Dolayısıyla tam bilgiden beklenen değer ile hesaplanan 250'den daha da az bir kazanç elde edilme olasılığı bulunmaktadır. Karar verici tam bilgiye ihtiyaç duymaksızın tespit ettiği en iyi kararı uygulamalıdır.

Yararlanılan ve Başvurulabilecek Kaynaklar

Aitchison, J. (1970).Choice Against Chance: An Introduction to Statistical Decision Theory, Addison-Wesley, United Kingdom.

Aladağ, Z. (2011). **Karar Teorisi,** Umuttepe Yayınları, Kocaeli.

Bağırkan, Ş. (1983). **Karar Verme,** Der Yayınları, İstanbul.

Bunn, D. (1982). **Analysis of Optimal Decisions,** John Wiley, USA.

De Groot, M. (1970), **Optimal Statistical Decisions.** Wiley Classics Library. 2004, USA.

Erdoğmuş, Ş. (2003). **Karar Kuramına Giriş,** Ders Notları, Eskişehir.

French, S. (1986). **Decision Theory,** John Wiley, USA.

Goodwin, P. and Wright, G. (2004). **Decision Analysis for Management Judgment,** 3rd edition, Chichester: Wiley, USA.

Gürsakal, N. (2000). **Sosyal Bilimlerde Araştırma Yöntemleri,**Vipaş Yayınları, Bursa.

Harnett, D.L. (1982). **Statistical Methods,** Addison Wesley, USA.

Huber, P.J. (2011). **Data Analysis What Can Be Learned from the Past 50 Years,** Wiley, USA.

Kara, İ. (1985). **Karar ve Oyun kuramıyla İlgili Başlangıç Bilgiler (Ders Notları),** Anadolu Üniversitesi Yayınları, No 65, Eskişehir.

Lind, D., Marchal, W.G., and Wathen, S.A. (2005). **Statistical Techniques in Business And Economics,** McGraw-Hill Irwin, USA.

Lindley, D.V. (1971). Making Decisions, John Wiley, USA.

Schlaifer, R. (1969). Analysis of Decisions Under Uncertainity, McGraw-Hill, USA.

Sullivan, M. (2005). **Fundamentals of Statistics,** Pearson Prentice Hall, USA.

Taylor III, B.W. (2010). **Introduction to Management Science,** Tenth Edition, Pearson, USA.

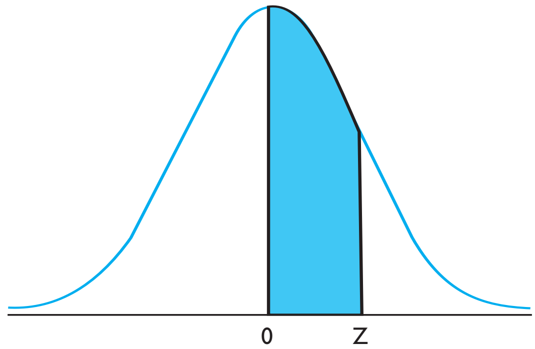
Turanlı, M. (1988). **Pazarlama Yönetiminde Karar Alma,** Beta Basım, İstanbul.

Şıklar, E.İ. (2001). **Karar Kuramı,** Yayımlanmamış Ders Notları, Eskişehir.

Winkler, R.L. (1972). Introduction to Bayesian Inference and Decision, Rinehart and Winston, USA.

Ek-1

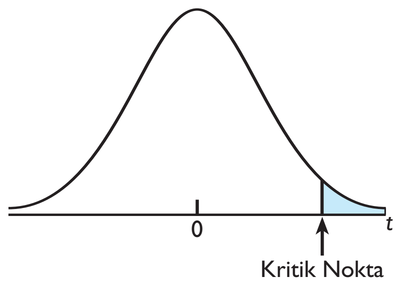
İSTATİSTİK-II SORULARININ CEVAPLANMASINDA GEREKLİ OLABİLECEK TABLOLAR



Tablo 1: *Normal Eğri Alanları*

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990

Ek-2

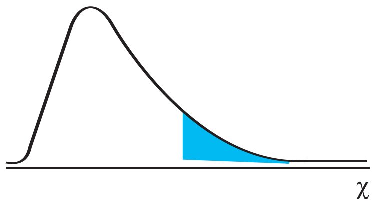


Tablo 2: *Kritik t Değerleri Tablosu*

(Anlamlılık Düzeyi)												
sd	**0,45**	**0,40**	**0,35**	**0,30**	**0,25**	**0,20**	**0,15**	**0,10**	**0,05**	**0,025**	**0,01**	**0,005**
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750
∞	0,1257	0,2533	0,3853	0,5244	0,6745	0,8416	1,0364	1,2816	1,6449	1,9600	2,3264	2,5759

sd; *Serbestlik Derecesi*

Ek-3



Tablo 3: Ki-kare Tablosu

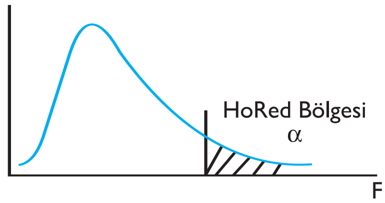
sd \ α	0,995	0,990	0,975	0,950	0,050	0,025	0,010	0,005	0,001
1	0,000039	0,0002	0,0010	0,0039	3,8415	5,0239	6,6349	7,8794	10,8276
2	0,0100	0,0201	0,0506	0,1026	5,9915	7,3778	9,2103	10,5966	13,8155
3	0,0717	0,1148	0,2158	0,3518	7,8147	9,3484	11,3449	12,8382	16,2662
4	0,2070	0,2971	0,4844	0,7107	9,4877	11,1433	13,2767	14,8603	18,4668
5	0,4117	0,5543	0,8312	1,1455	11,0705	12,8325	15,0863	16,7496	20,5150
6	0,6757	0,8721	1,2373	1,6354	12,5916	14,4494	16,8119	18,5476	22,4577
7	0,9893	1,2390	1,6899	2,1673	14,0671	16,0128	18,4753	20,2777	24,3219
8	1,3444	1,6465	2,1797	2,7326	15,5073	17,5345	20,0902	21,9550	26,1245
9	1,7349	2,0879	2,7004	3,3251	16,9190	19,0228	21,6660	23,5894	27,8772
10	2,1559	2,5582	3,2470	3,9403	18,3070	20,4832	23,2093	25,1882	29,5883
11	2,6032	3,0535	3,8157	4,5748	19,6751	21,9200	24,7250	26,7568	31,2641
12	3,0738	3,5706	4,4038	5,2260	21,0261	23,3367	26,2170	28,2995	32,9095
13	3,5650	4,1069	5,0088	5,8919	22,3620	24,7356	27,6882	29,8195	34,5282
14	4,0747	4,6604	5,6287	6,5706	23,6848	26,1189	29,1412	31,3193	36,1233
15	4,6009	5,2293	6,2621	7,2609	24,9958	27,4884	30,5779	32,8013	37,6973
16	5,1422	5,8122	6,9077	7,9616	26,2962	28,8454	31,9999	34,2672	39,2524
17	5,6972	6,4078	7,5642	8,6718	27,5871	30,1910	33,4087	35,7185	40,7902
18	6,2648	7,0149	8,2307	9,3905	28,8693	31,5264	34,8053	37,1565	42,3124
19	6,8440	7,6327	8,9065	10,1170	30,1435	32,8523	36,1909	38,5823	43,8202
20	7,4338	8,2604	9,5908	10,8508	31,4104	34,1696	37,5662	39,9968	45,3147
21	8,0337	8,8972	10,2829	11,5913	32,6706	35,4789	38,9322	41,4011	46,7970
22	8,6427	9,5425	10,9823	12,3380	33,9244	36,7807	40,2894	42,7957	48,2679
23	9,2604	10,1957	11,6886	13,0905	35,1725	38,0756	41,6384	44,1813	49,7282
24	9,8862	10,8564	12,4012	13,8484	36,4150	39,3641	42,9798	45,5585	51,1786
25	10,5197	11,5240	13,1197	14,6114	37,6525	40,6465	44,3141	46,9279	52,6197
26	11,1602	12,1981	13,8439	15,3792	38,8851	41,9232	45,6417	48,2899	54,0520
27	11,8076	12,8785	14,5734	16,1514	40,1133	43,1945	46,9629	49,6449	55,4760
28	12,4613	13,5647	15,3079	16,9279	41,3371	44,4608	48,2782	50,9934	56,8923
29	13,1211	14,2565	16,0471	17,7084	42,5570	45,7223	49,5879	52,3356	58,3012
30	13,7867	14,9535	16,7908	18,4927	43,7730	46,9792	50,8922	53,6720	59,7031
40	20,7065	22,1643	24,4330	26,5093	55,7585	59,3417	63,6907	66,7660	73,4020
50	27,9907	29,7067	32,3574	34,7643	67,5048	71,4202	76,1539	79,4900	86,6608
100	67,3276	70,0649	74,2219	77,9295	124,3421	129,5612	135,8067	140,1695	149,4493

sd; *Serbestlik Derecesi*

α; *Anlam Düzeyi*

Ek-4

Anlamlılık düzeyi (α = 0,05)



Tablo 4: *F Tablo Değerleri*

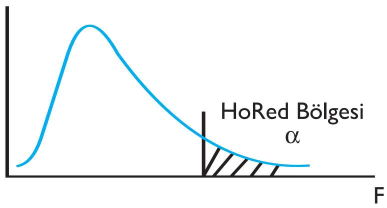
Serbestlik Derecesi Sd_1											
Sd_2	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**12**	**24**	**∞**
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,68	4,53	4,37
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,79	2,61	2,41
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,69	2,51	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,20	2,01	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,18	1,98	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,16	1,96	1,71
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,12	1,91	1,66
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,09	1,89	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,00	1,79	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	1,92	1,70	1,39
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	1,88	1,65	1,33
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,85	1,63	1,28
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,83	1,61	1,26
∞	3,84	3,00	2,61	2,37	2,22	2,10	2,01	1,94	1,75	1,52	1,00

Sd_1: Gruplararası serbestlik derecesi

Sd_2: Gruplariçi serbestlik derecesi

Ek-5

Anlamlılık düzeyi (α = 0,01)



Tablo 5: *F Tablo Değerleri*

	Serbestlik Derecesi Sd_1										
Sd_2	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**12**	**24**	**∞**
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	9,89	9,47	9,02
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,72	7,31	6,88
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,47	6,07	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,67	5,28	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,11	4,73	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,71	4,33	3,91
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,40	4,02	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,16	3,78	3,36
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	3,96	3,59	3,17
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	3,80	3,43	3,01
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,67	3,29	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,55	3,18	2,75
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,46	3,08	2,65
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,37	3,00	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,30	2,92	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,23	2,86	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,17	2,80	2,36
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,12	2,75	2,31
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,07	2,70	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,03	2,66	2,21
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	2,99	2,62	2,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	2,96	2,58	2,13
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	2,93	2,55	2,10
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	2,90	2,52	2,07
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	2,87	2,49	2,04
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	2,84	2,47	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,66	2,29	1,81
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,50	2,12	1,60
80	6,96	4,88	4,04	3,56	3,26	3,04	2,87	2,74	2,42	2,03	1,50
100	6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,37	1,98	1,43
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,34	1,95	1,38
∞	6,64	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,19	1,79	1,0
Sd_1: Gruplararası serbestlik derecesi											
Sd_2: Gruplariçi serbestlik derecesi											