TC. ANADOLU ÜNİVERSİTESİ YAYINI NO: 1448
AÇIKÖĞRETİM FAKÜLTESİ YAYINI NO: 771

İSTATİSTİK

Yazarlar
Prof.Dr. Ali Fuat YÜZER (Ünite 1, 2)
Prof.Dr. Embiya AĞAOĞLU (Ünite 3, 13)
Prof.Dr. Hüseyin TATLIDİL (Ünite 4, 5, 6)
Prof.Dr. Ahmet ÖZMEN (Ünite 7, 8, 9, 14)
Prof.Dr. Emel ŞIKLAR Ünite (10, 11, 12)

Editör
Prof.Dr. Ali Fuat YÜZER

ANADOLU ÜNİVERSİTESİ

UZAKTAN ÖĞRETİM TASARIM BİRİMİ

Genel Koordinatör
Prof.Dr. Levend Kılıç

Genel Koordinatör Yardımcısı
Doç.Dr. Müjgan Bozkaya

Öğretim Tasarımcısı
Yard.Doç.Dr. Melih Zeytinoğlu

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

Televizyon Programları Yöneticisi
Prof. Yalçın Demir

Dil ve Yazım Danışmanları
Yard.Doç.Dr. Hülya Pilancı
Okt. Aydın Fındıkoğlu
Okt. Meral Aşkar

Ölçme Değerlendirme Sorumlusu
Öğr.Gör. Reha Akgün

Kitap Koordinasyon Birimi
Yard.Doç.Dr. Feyyaz Bodur
Uzm. Nermin Özgür

Kapak Düzeni
Prof. Tevfik Fikret Uçar

Dizgi
Açıköğretim Fakültesi Dizgi Ekibi

İstatistik

ISBN
975 - 06 - 0183 - 1

6. Baskı
Bu kitap ANADOLU ÜNİVERSİTESİ Web-Ofset Tesislerinde 60.000 adet olarak basılmıştır.
ESKİŞEHİR, Ekim 2009

İçindekiler

Sunuş

Anadolu Üniversitesi uzaktan eğitim uygulayan İktisat ve İşletme Fakültelerinde yürütülen istatistik dersleri kapsamına ve uzaktan öğretim koşullarına göre hazırlanan bu kitap, istatistiğin temel konularının ele alındığı ondört üniteden oluşmuştur.

Kitabın hazırlanmış amaçları uyarınca konular işlenirken ilgili kavramlar daha çok sezgiye dayalı yaklaşımlarla verilmeye özen gösterilerek, teorik anlatımlardan kaçınılmaya çalışılmıştır. Bu nedenle her ünitede, ele alınan kavram ve tekniklere ilişkin yeterli sayıda çözümlü örneğe yer verilmiştir.

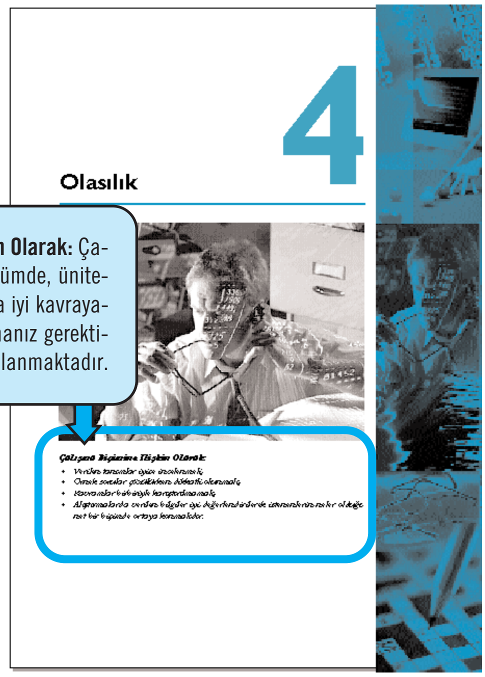
Metin içerisinde **sıra sizde** başlığı altında, işlenen konuyla doğrudan ilgili alıştırmalarla öğrenilenlerin, basit de olsa, günlük yaşamdaki bazı sorunların çözümünde kullanılarak verilen kavram ve tekniklerin pekiştirilmesi amaçlanmıştır.

Ünitelerin sonunda **kendimizi sınayalım** başlığı altında, ilgili ünitenin yanı sıra, önceki ünitelerde öğrenilenleri de sınamaya yönelik, cevapları ünitenin sonunda bulunan ve sizleri sınava hazırlamayı amaçlayan, çoktan seçmeli test türü sorular yer almaktadır. Ancak çözüm için gerekli çabayı harcamadan cevaplara bakmayınız. Unutmayınız ki öğrenmek keşfetmek demektir. Özveri, sabır ve çaba ister. Öğrenebilmek için kaleminizi kullanınız. Eğer karşılaştığınız problemlerin çözümünde güçlüklerle karşılaşıyorsanız, ilgili konuları tekrar tekrar gözden geçiriniz. Örnek çözümleri ve alıştırmaları yeniden çözünüz. Başardığınızı göreceksiniz.

Elinizdeki kitap, geniş bir ekibin uzun süren çalışmaları sonucunda ortaya çıkan bir üründür. Bu ekibin oluşturulması ve çalışmasında her türlü olanağı sağlayan Anadolu üniversitesi Rektörü Prof.Dr.Engin ATAÇ'a ve genel koordinatör Prof.Dr.Levend KILIÇ'ın şahsında kitabın hazırlanmasında tüm emeği geçenlere, editör ve yazarlar olarak teşekkür ederiz.

Prof.Dr. Ali Fuat YÜZER
Mart 2003

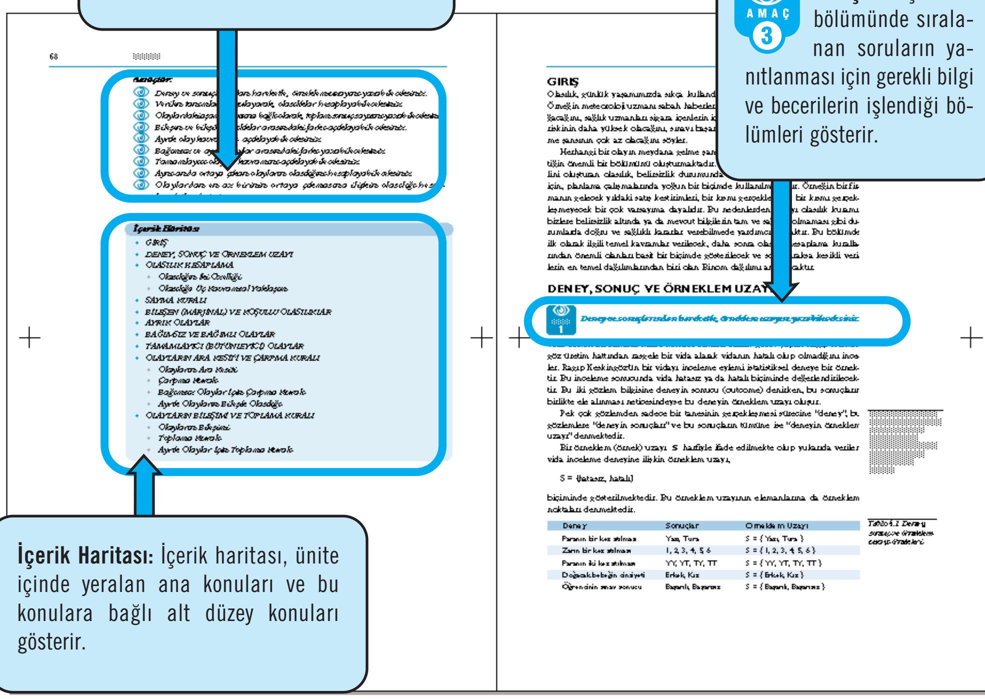
Kendi kendine öğrenme ilkelerine göre hazırlanmış olan bu kitabın işlevlerini öğrenmek için hazırlanan "Kullanım Kılavuzu", konuları anlamanızda ve sınavlara hazırlanmanızda sizlere fayda sağlayacaktır.



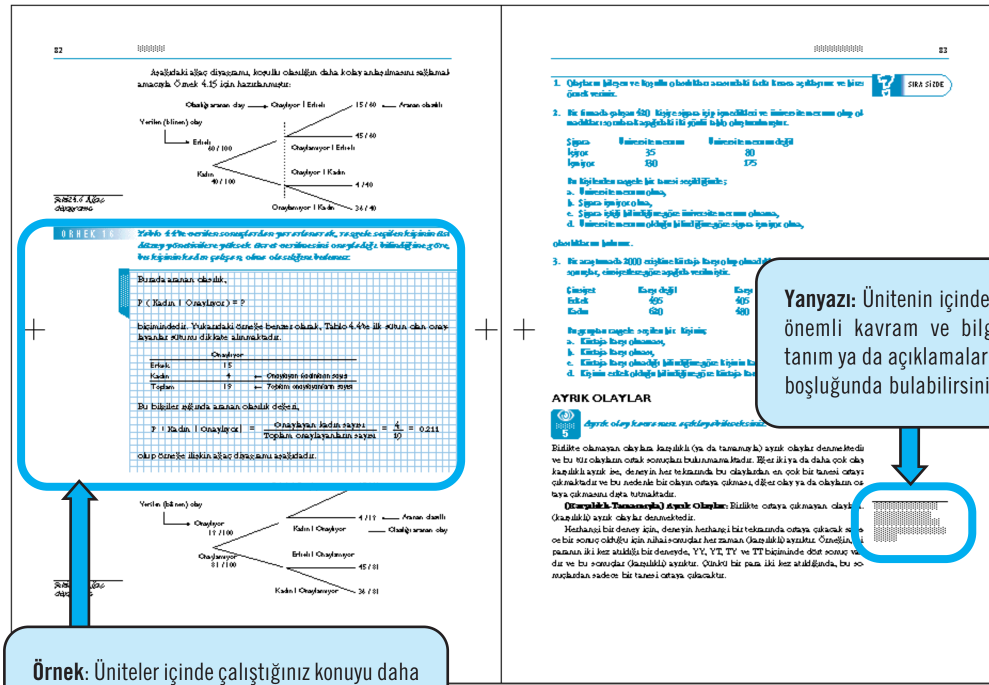
Çalışma Biçimine İlişkin Olarak: Çalışma biçimiyle ilgili bölümde, ünitede yeralan konuları daha iyi kavrayabilmeniz için neler yapmanız gerektiği maddeler halinde sıralanmaktadır.

Amaçlarımız: Amaçlarımız bölümünde, okuduğunuz ünite sonunda kazanacağınız bilgi ve beceriler sunulmaktadır.

Amaç: Amaçlarımız bölümünde sıralanan soruların yanıtlanması için gerekli bilgi ve becerilerin işlendiği bölümleri gösterir.



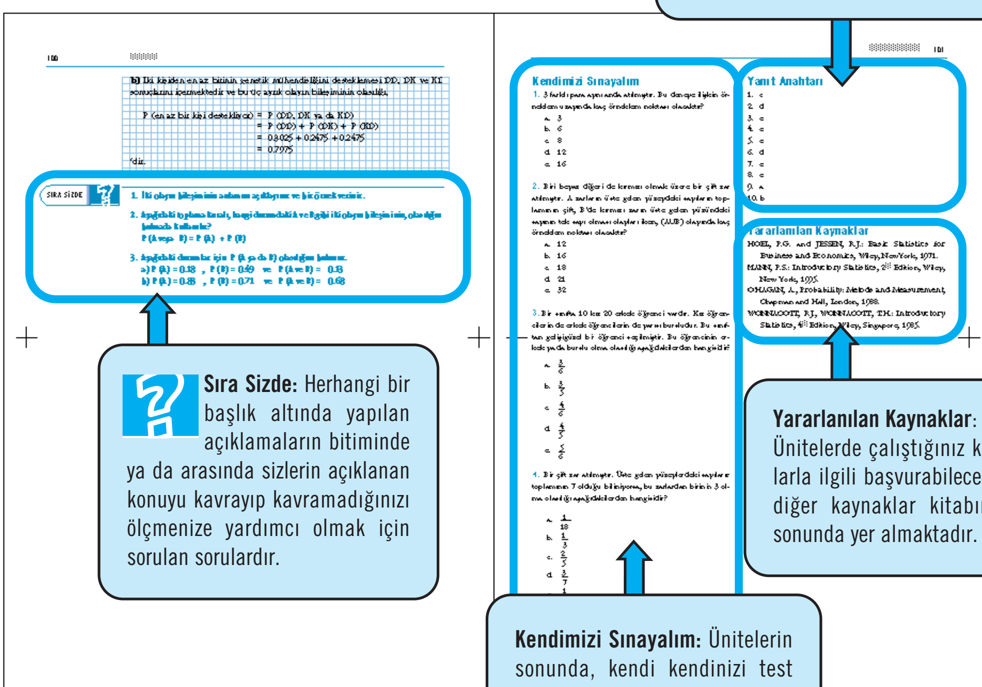
İçerik Haritası: İçerik haritası, ünite içinde yeralan ana konuları ve bu konulara bağlı alt düzey konuları gösterir.



Yanyazı: Ünitenin içinde yer alan bazı önemli kavram ve bilgilere yönelik tanım ya da açıklamaları sayfanın yan boşluğunda bulabilirsiniz.

Örnek: Üniteler içinde çalıştığınız konuyu daha iyi kavramanız, bilgi ve beceri kazanmanızı sağlayacak, çok sayıda örnek problem ve çözümleri bulabilirsiniz.

Yanıt Anahtarı: Kendimizi sınayalım bölümlerinde yanıtladığınız çoktan seçmeli soruların yanıtları kitabınızın sonunda sunulmuştur.



Sıra Sizde: Herhangi bir başlık altında yapılan açıklamaların bitiminde ya da arasında sizlerin açıklanan konuyu kavrayıp kavramadığınızı ölçmenize yardımcı olmak için sorulan sorulardır.

Yararlanılan Kaynaklar: Ünitelerde çalıştığınız konularla ilgili başvurabileceğiniz diğer kaynaklar kitabınızın sonunda yer almaktadır.

Kendimizi Sınayalım: Ünitelerin sonunda, kendi kendinizi test edebilmenizi amaçlayan çoktan seçmeli sorular sunulmuştur. Bu sorular, sınavda karşılaştığınız sorularla aynı türdendir.

Temel Kavramlar

Çalışma Biçimine İlişkin Olarak:

- *Tüm kavramlar dikkatle gözden geçirilmeli,*
- *Kavramlar arası ilişkilere dikkat edilmelidir.*

Amaçlar:

- *İstatistiğin temel kavramlarını açıklayabileceksiniz.*
- *Veri derleme kavramını açıklayabilecek ve veri derleme türlerini sınıflandırabileceksiniz.*

İçerik Haritası

- *GİRİŞ*
- *BİRİM, DEĞİŞKEN VE İSTATİSTİK KÜTLESİ (ANA KÜTLE)*
  - *Birim*
  - *Değişken (Özellik)*
  - *İstatistik Kütlesi (Ana Kütle)*
- *VERİ DERLEME*
  - *Birim Seçimi*
  - *Değişken ve Şıkların Belirlenmesi*
  - *Kütlenin Sınıflandırılması*
  - *Veri Derleme Türleri*

GİRİŞ

İstatistik sözcüğü farklı yaklaşımlara göre değişik anlamlar taşır. Günlük dilde istatistik ya da istatistikler denildiğinde, belirli bir olaya ilişkin derlenmiş sayısal bilgiler akla gelir. Örneğin, dışalım, dışsatım, turizm, inşaat istatistikleri ve benzerleri gibi.

Metodoloji açısından istatistik sözcüğü, istatistiğe konu olabilen olayların gözlenerek ilgili verilerin derlenmesi, işlenmesi, analizi ve yorumlanmasında kullanılan tekniklerin tümünü ifade eder.

XX. yüzyılın başlarında istatistik alanındaki gelişmeler, istatistik sözcüğüne teknik içerikli yeni bir anlam kazandırmıştır. Bağlı olarak istatistik sözcüğü, hakkında bilgi edinilmek istenen ve ana kütle olarak isimlendirilen yığına ilişkin sayısal karakteristikleri (parametreleri) tahminleyebilmek amacıyla, ilgili kütleden belirli kurallara göre seçilen istatistik birimlerinin oluşturduğu ve örneklem adı verilen topluluğa ilişkin sayısal karakteristikler anlamında da kullanılmaktadır.

İstatistik de tüm diğer bilim dalları gibi olayları konu alır. Olay varsa istatistik vardır. Ancak her olay da istatistiğe konu oluşturamaz. İstatistik yığın olaylarla ilgilenir. Yığın olay, bir olaylar kümesinde tek bir olayın diğerlerini, bağlı olarak da ait olduğu kümeyi temsil edemeyen olaylardır. Eğer bir olaylar kümesinde tek bir olay, tüm olaylar kümesini temsil edebiliyorsa, bu tür olaylara tipik olay denir. Ancak istatistik tipik olaylarla ilgilenmez. Örneğin, ideal koşullar altında ve uygun bir laboratuvar ortamında iki hidrojen ve bir oksijen atomu bir araya getirilirse, su elde edilir. Bu deney aynı koşullar altında kaç kez tekrarlanırsa tekrarlansın, her deneyin sonucunda su elde edilecektir. Görüleceği gibi, bu örnekte tek bir deney ilgili deneyler kümesini temsil edebilmektedir. Dolayısıyla bu olay tipik olaydır. Ancak günlük yaşamdaki olaylar bu örnekteki olaya benzemez. Örneğin, firmaların yıllık ciroları, trafik kazaları, evlenmeler, boşanmalar, doğumlar, ölümler ve benzeri gibi her gün karşılaşılan olaylar, birer yığın olay niteliğindedir.

Bir olaylar kümesinde tek bir olay kümedeki diğer olayları temsil edemiyorsa, bu tür olaylara yığın olay denir ve istatistik yığın olayları konu alır.

İstatistik, belirli amaç ya da amaçlar doğrultusunda gözlenen yığın olaylardan derlenen sayısal verilerin işlenerek, ilgili olayların oluşturduğu yığınların bilimsel olarak incelenmesinde kullanılan teknik ve yöntemler bilimi olarak tanımlanabilir.

İstatistik, belirlenen amaç ya da amaçlar doğrultusunda gözlenen yığın olaylardan derlenen sayısal verilerin işlenerek, ilgili olayların oluşturduğu yığınların bilimsel olarak incelenmesinde kullanılan teknik ve yöntemler bilimidir.

Tanımdan da anlaşılabileceği gibi, çeşitli etkenlerin etkisini taşıyan yığının, ilgilenilen özellik ya da özelliklerinin aldığı değerler, rakamlarla ifade edilebilmelidir.

Günümüzde, istatistik, deney ya da gözlemlere dayalı tüm bilim dallarında, geniş bir uygulama alanına sahiptir.

Daha önce de değinildiği gibi, istatistik yığın olayların gözlenerek incelenmesi ve analizinde kullanılan teknikler topluluğudur. İlgilenilen olayın kavranabilmesi ve yapılacak deney ya da gözlemlerin sayısal olarak analiz edilebilmesi için öncelikle deney ya da gözlemlere konu olan olayın ilgilenilen özellik ya da özelliklerinin belirlenmesi, sonra da bunların sayılması ya da ölçülmesi gerekir. Bu aşamaya, verilerin toplanması ya da verilerin derlenmesi adı verilir.

Veriler derlenirken, ilgilenilen kütleye ilişkin birimler sayılır ya da ölçülürken, öte yandan da bu birimlerin ilgilenilen özellik ya da özellikleri açısından hangi şıklara sahip olduğu belirlenir ve kaydedilir.

Yukarıda değinilen kütle, birim, özellik ve şık kavramları, izleyen kesimlerde yeterli ayrıntıyla ele alınacaktır.

BİRİM, DEĞİŞKEN VE İSTATİSTİK KÜTLESİ (ANA KÜTLE)

AMAÇ 1 *İstatistiğin temel kavramlarını açıklayabileceksiniz.*

Birim

Yığın olay niteliğindeki her olaya birim adı verilir. Kolaylıkla anlaşılabileceği gibi tüm canlı ve cansız varlıklar birer istatistik birimidir. Ancak, maddesel bir varlığa sahip olmayan olaylar ve sosyal kurumlar da birer istatistik birimi olabilirler. Bir olayın birim olabilmesi için, ölçülmeye ya da sayılmaya elverişli olması gerekir. Ölçülemeyen ya da sayılamayan nesneler ve olaylar istatistiksel anlamda birim oluşturamazlar. Örneğin; insan, bina, araba ve hayvan gibi canlı ve cansız varlıklar istatistik birimleridir. Öte yandan, doğum, ölüm, evlenme, iflas ve trafik kazası gibi olaylar da birim oluşturabilirler. Ancak sevinçler, korkular, rüyalar ve renkler sayılamadıkları ya da ölçülemedikleri için birim olamazlar.

Bir nesne ya da olayın, birim olabilmesi için, ölçülmesi ya da sayılması yeterlidir.

Birim Türleri

Birimler farklı ölçütlere göre sınıflandırılabilirler. İzleyen paragraflarda birimlerin maddesel bir varlığa sahip olup olmamalarına, ömür sürelerine, doğal olup olmadıklarına ve gerçek ya da varsayımsal oluşlarına göre sınıflandırılarak kısaca ele alınacaktır.

Maddesel Bir Varlığa Sahip Olan ya da Olmayan Birimler

Eğer birimler insan, araba ve benzeri gibi canlı ya da cansız maddesel bir varlığa sahipse, bu tür birimlere, maddesel varlığa sahip birimler adı verilir. Eğer birimler, doğum, ölüm, trafik kazası ve benzeri gibi olay niteliğindeyse bu tür birimlere de maddesel varlığa sahip olmayan birimler adı verilir.

Sürekli ya da Ani Birimler

İstatistiğin ilgi alanına giren olaylar, doğal olarak, sınırlı bir ömre sahiptir. Belirli bir zaman aralığı içinde herhangi bir anda gözlenebilen istatistik birimlerine sürekli birimler adı verilir. Örneğin; insan, bina, ticari bir kuruluş ve benzerleri gibi. Bu tür birimler varlıklarını sürdürdükleri süre içinde herhangi bir anda gözlemlenebilirler. Dolayısıyla bu tür birimler, istenilen bir zamanda yapılacak bir sayım için uygun bir ortam oluştururlar. Maddesel bir varlığa sahip birimler sürekli birimlerdir.

Öte yandan; evlenme, boşanma, trafik kazası gibi bir olay ya da bir fiil biçiminde ortaya çıkan birimler, oldukça kısa ömürlüdürler. Ani birimler olarak isimlendirilen bu tür birimler, zaman içinde dağılmış olarak ortaya çıkarlar. Kolaylıkla anlaşılabileceği gibi, "ani birimler" maddesel bir varlığa sahip olmayan birimlerdir.

Doğal ya da Doğal Olmayan Birimler

Nitelikleri açısından bir bütün oluşturan, parçalanmaları yada birleştirilmeleri halinde niteliklerini kaybeden birimlere "doğal birim" adı verilir. Örneğin bir canlı parçalandığında, canlı olma niteliğini kaybeder ve her parça da daha küçük bir canlı oluşturmaz. Bir TV alıcısı ya da bir otomobil için de durum aynıdır. Öte yandan, iki öğrenci bir araya getirilerek, daha uzun boylu, daha ağır ya da daha zeki bir öğrenci oluşturulamaz. Kolaylıkla anlaşılabileceği gibi, doğal birimler bir araya getirilerek ya da parçalanarak aynı nitelikte birimler elde edilemezler.

Nitelikleri açısından bir bütün olma özelliği göstermeyen birimlere doğal olmayan birim adı verilir. Bu tür birimlerin, birleştirildikleri ya da parçalandıkları zaman, nitelikleri değişmez. Örneğin bir arsa bir kaç parçaya bölünürse, daha küçük arsalar ortaya çıkar. Arsanın, arsa olma niteliği değişmez.

Gerçek ya da Varsayımsal Birimler

Gerçekte var olan birimlere "gerçek birim" adı verilir. Bir birimin gerçek birim olabilmesi için mutlaka maddesel bir varlığa sahip olması gerekmez. Örneğin; ev, arsa, insan, bisiklet gibi maddesel bir varlığa sahip birimler gerçek birimlere örnek oluştururken, doğum, ölüm, evlenme, iflas gibi olay ya da fiil biçiminde ortaya çıkan birimler de gerçek birimlerdir. Bir birimin, gerçek birim olabilmesi için, ortaya çıkmış olması yeterlidir.

Öte yandan kuramsal olarak oluşturulabilecek birimler de söz konusudur. Örneğin on öğrenci arasından, üçer öğrenciden oluşacak her grup da bir birim olarak görülebilir. Bu tür birimlere de "varsayımsal birimler" adı verilir.

Değişken (Özellik)

İstatistik birimlerinin sahip oldukları özellikler birer değişken olarak görülebilir. Örneğin bir öğrenci grubu göz önüne alındığında, bu öğrencilerin doğum yerleri, yaşları, ağırlıkları ve boy uzunlukları aynı değildir. Öğrenciden öğrenciye değişir. Bu özelliklerin farklı ortaya çıkış biçimlerine, başka bir anlatımla değişkenlerin aldıkları değerlere ise "şık" adı verilir.

İstatistik birimlerinin sahip olduğu özelliklere "değişken", değişkenlerin aldıkları değerlere ise "şık" adı verilir.

Yukarıda sözü edilen öğrenci grubunun boy uzunlukları ve ağırlıklarıyla ilgilenilecekse, boy uzunluğu ve ağırlık "değişkenleri (özellikleri)", her öğrencinin ayrı ayrı boy uzunluğu ve ağırlığı da bu değişkenlere ilişkin "şıkları" oluşturur.

Değişken (Özellik) Türleri

Değişkenler de farklı ölçütler esas alınarak sınıflandırılabilir. Ancak bu ünitede değişkenler zaman, mekan ve maddesel değişkenler başlıkları altında sınıflandırılacaktır.

Eğer bir değişkenin şıkları mekana göre oluşuyorsa, bu tür değişkenlere "mekan değişkeni", zamana göre oluşuyorsa bu tür değişkenlere de "zaman değişkeni" adı verilir. Mekan ve zaman değişkenleri dışındaki tüm değişkenlere ise "maddesel değişken" adı verilir. Örneğin; doğum yeri ve üniversitelerin bulundukları şehirler mekan değişkenlerine, doğum yılı ve üniversitelerin kuruluş yılları da zaman değişkenlerine örnekler oluşturur. Öte yandan, insanların medeni durumu, bir işletmedeki birim değişken maliyetler ve bir sınıfta belirli bir dersten alınan notlar da maddesel değişkene örnekler oluşturur.

İstatistik Kütlesi (Ana Kütle)

Yığın olay niteliğinde ve aynı cins birimlerin oluşturduğu topluluğa "istatistik kütlesi" ya da "ana kütle" adı verilir. Ancak, bir istatistik kütlesinden söz edebilmek için, öncelikle kütleyi oluşturan birimlerin, aynı genel nedenlerin etkisinde olması gereklidir. Ayrıca kütle, istatistik birimlerinin toplamından farklı bir yapıya da sahip olmamalıdır. Bir ülkede yaşayan insanlar, belirli bir bölgedeki evler, bir yıl süresince belirli bir yerleşim merkezinde gözlenen doğumlar, ölümler, trafik kazaları, istatistik kütlesi için örnekler oluşturur. Ancak, Anadolu Üniversitesi, öğrenci ve öğretim üyelerinden oluşmuş bir istatistik kütlesi olarak değerlen-

Yığın olay niteliğindeki aynı cins birimlerin oluşturduğu topluluğa "ana kütle" adı verilir.

dirilemez. Üniversite, öğrenci ve öğretim üyelerinin toplamından farklı, tüzel kişiliğe sahip bir varlıktır.

Kütle Türleri

İstatistik kütlelerini de, kütleyi oluşturan birimlerin niteliklerine göre sınıflandırmak mümkündür. Böyle bir sınıflandırma aşağıda ana çizgileriyle ele alınmıştır.

Gerçek ya da Varsayımsal Kütleler

Gerçek birimlerin oluşturdukları kütlelere, "gerçek kütle" adı verilir. Bir üniversitenin öğrencileri, bir yerleşim merkezinde bir yılda gözlenen trafik kazaları ve doğum olaylarının oluşturdukları kütleler, gerçek kütlelere örnek oluştururlar.

Henüz oluşmamış, ancak oluşturulması mümkün olan kütlelereyse "varsayımsal kütle" adı verilir. Kolaylıkla görülebileceği gibi varsayımsal kütleler, varsayımsal birimlerim oluşturduğu kütlelerdir. Örneğin, 30 kişilik bir sınıftan rasgele seçilecek 5 kişilik bir grup için $C_{30}^{5} = 142.506$ farklı seçim yapılabilir. 142.506 farklı 5 kişilik grupların oluşturduğu kütle varsayımsal bir kütledir.

Sonlu ya da Sonsuz Kütleler

Eğer bir kütledeki birimler sonlu sayıdaysa başka bir anlatımla sayılabiliyorsa, bu tür kütlelere "sonlu (belirli)", kütleyi oluşturan birim sayısı sayılamıyorsa, bu tür kütlelere de "sonsuz (belirsiz)" kütle adı verilir. Örneğin, bir ülkede yaşayan insanların sayısı sayılabileceğinden bu ülkede yaşayan insanların oluşturduğu kütle sonlu bir kütledir. Marmara Denizinde yaşayan canlılarsa sayılamayacakları için sonsuz bir kütle oluştururlar.

Sürekli ya da Süreksiz Kütleler

Parçalandıkları ya da birleştirildikleri zaman, niteliklerini kaybettikleri için, doğal birimlerden oluşan kütleler süreksiz, parçalandıkları ya da birleştirildiklerinde, niteliklerini kaybetmedikleri için de doğal olmayan birimlerden oluşan kütlelerse, sürekli kütleler oluştururlar.

Zaman ve mekan birimleri doğal birimler olmadıkları için, her zaman sürekli kütleleri oluştururlar.

SIRA SİZDE

1. **Günümüzde istatistiğin hangi nedenlerle geniş bir uygulama alanına sahip olduğunu açıklayınız.**

2. **İstatistik ne tür olaylarla ilgilenir, nedenleriyle açıklayınız.**

3. **Maddesel varlığı olmayan trafik kazası, doğum, evlenme ve grev gibi olayların niçin birim olabildiklerini açıklayınız.**

VERİ DERLEME

AMAÇ 2

Veri derleme kavramını açıklayabilecek ve veri derleme türlerini sınıflandırabileceksiniz.

Veri derleme; belirlenen amaçlar doğrultusunda gözlemlenecek birimlerin ölçülmesi ya da sayılması, sonra da bunların, ilgilenilen değişkenlere göre, hangi şıklara sahip olduğunun belirlenmesi ve kaydedilmesi işlemlerini içerir.

Yukarıdaki tanımdan da anlaşılabileceği gibi, belirlenen amaçlar doğrultusunda istatistiksel bir çalışma başlatılırken, öncelikle araştırma konusuna uygun birimin ve ilgilenilen değişken ya da değişkenlerin dikkatli bir biçimde belirlenmesi gerekir.

Birim Seçimi

Belirlenen amaç ya da amaçlar doğrultusunda, ilgilenilen yığın olayın tanımlanmasıyla "birim seçme" işlemi gerçekleştirilmiş olur. Başka bir anlatımla, kimlerin ya da nelerin gözleneceği belirlenir. Ancak birim seçilirken, amaca uygunluk ve uygulanabilirlik özelliklerinin öncelikle göz önünde bulundurulması gerekir. Bunun için de birim belirlenirken, birim tanımının kesin, amaca uygun ve uygulamaya elverişli olması gerekir.

Tanımın kesin olması, uygulamacılarca ilgilenilen yığın olaya ilişkin aynı şeyin anlaşılması, başka bir anlatımla kuşkulara yer açmayacak biçimde açık olmasıdır.

Tanımın kesin olmasının yanı sıra, tanımın amaca uygunluğu ve kolaylıkla uygulanabilirliği de gözden uzak tutulmamalıdır.

Değişken ve Şıkların Belirlenmesi

Açıktır ki, bir kütleyi oluşturan istatistik birimleri üzerinde bir çok değişken tanımlanabilir. Veri derlenirken sadece belirlenen amaçlar doğrultusundaki değişkenler göz önünde tutulmalıdır. Uygulamalarda fazla ayrıntı sorunlara neden olabilir.

Öte yandan uygulamalarda gözlem sayısı kesinlikle sonlu bir sayı olacaktır. Ayrıca, gözlemlere bağlı olarak ilgili değişkenlerin alacakları değerler de (şıklar da) ilgili değişken sürekli ya da süreksiz olsun, sonlu olacaktır.

Şıklar belirlenirken, gözden uzak tutulmaması gereken önemli bir nokta da gözlemlerde kullanılan ölçü biriminin araştırmanın doğasına uygun olması gereğidir. Örneğin; ayçiçek yağı üreten bir firmanın, aylık üretimi için kilo, şişe, teneke ya da ton makul ölçülerken, bir sarrafın bir günde sattığı bilezikler için gram uygun bir ölçü olur.

Kütlenin Sınıflandırılması

Bir istatistiksel araştırma planlanırken, araştırmanın nerede, kimlerle ve nelerle gerçekleştirileceği, ne kadar zamanda tamamlanacağı ve araştırma için ayrılan kaynaklar, ayrıca gözlem sayısının sonlu olması, kütlenin mekan ve zaman açısından sınırlandırılmasını zorunlu kılar.

Başarılı bir sınırlandırma uygulamacılara büyük kolaylık sağlar.

Veri Derleme Türleri

Veri derleme süreci kabaca, sürekli ya da ani ve kısmi ya da genel olmak üzere iki başlık altında toplanabilir.

Ani ya da Sürekli Veri Derleme

Eğer gözlemlenecek kütledeki birimler sürekli karakterdeyse, istenilen bir anda gözlenmeye hazır olan bu tür birimlerin gözlenmesi ya da kaydedilmesi işlemlerine "ani veri derleme" denir. Nüfus sayımları ve iş yeri sayımları bu tür veri derlemeye örnek oluşturur.

Eğer ilgilenilen kütle ani birimlerden oluşmuşsa (bu tür birimler zamana yayıldığından), belli bir zaman aralığında gözlenmeleri ve kaydedilmeleri gerekir. Bu tür işlemlere "sürekli veri derleme" denir. Belirli bir bölgede ve zaman aralığında evlenmeler, boşanmalar, trafik kazaları, doğumlar ve ölümlere ilişkin derlenen veriler, bu tür veri derlemeye örnek oluşturur.

Genel ya da Kısmi Veri Derleme

Hakkında bilgi edinilmek istenen kütlenin tamamının gözlenmesine "genel veri derleme" adı verilir. Genel nüfus ve tarım sayımları birer genel veri derlemedir. Açıktır ki, bu tür veri derleme hem pahalı hem de güçtür. Öte yandan, genel veri derlemede bilgi edinilmek istenen kütlenin zaman içindeki değişim hızı da önem taşır. Eğer kütlenin değişimi, araştırmada öngörülen zaman içinde sonuçları etkileyebilecek düzeydeyse, genel veri derleme kendisinden beklenen yararları sağlayamayacağı için tercih edilmemelidir. Ayrıca gözlem ya da deneyler, gözlendikleri anda fiziksel zararlara uğruyorsa, böyle durumlarda da genel veri derleme uygulanamaz. Örneğin, yeni bir teknolojiyle üretilen top mermilerinin hedef üzerindeki etkilerinin denenmesi gibi. Harcanan her mermi yok olacağından, istenilen sonuçlara ne derece ulaşıldığı, ancak üretilen mermilerin bir kısmının denenmesiyle araştırılabilir. Elbette ki genellemelerin yapılabilmesi için, denenecek mermilerin belirli kurallara göre seçilmesi gerekir.

Hakkında bilgi edinilmek istenen kütleyi oluşturan birimler arasından, belirlenen amaçlar doğrultusunda yalnızca bir kısmının seçilip gözlenmesine, "kısmi veri derleme" adı verilir.

Kısmi veri derleme, genel veri derlemenin pahalı oluşu, zaman alışı, gözlem birimlerinin fiziksel zarara uğraması gibi nedenlerle yapılmak istenmediği zaman uygulanır.

Anımsanacağı gibi belirlenen amaçlar doğrultusunda hakkında bilgi edinilmek istenen yığının tümüne ana kütle (ya da sadece kütle) adı verilir. Bir ana kütleden uygun tekniklerle seçilen birimlerin oluşturduğu topluluğaysa "örneklem" adı verilir.

Ana kütleden uygun tekniklerle seçilen birimlerin oluşturduğu alt topluluğa "örneklem" adı verilir.

Belirlenen amaçlar uyarınca bir örneklem oluşturulurken, örneği oluşturmak için seçilen tekniğe göre de kısmi veri derleme, rassal ve iradi olama üzere iki kısma ayrılır. Konu örnekleme bölümünde ayrıntılarıyla ele alınacağından, burada sadece tanımlarla yetinilecektir.

Eğer örneklem seçilirken, ana kütledeki birimlerin hepsine örnekleme girebilmek için eşit şans verilirse oluşturulan örnekleme, "rassal örneklem" adı verilir. Eğer, bir örneklem oluşturulurken, kütledeki tüm birimlere eşit seçilme şansı verilmez, örneğe girmesi mümkün birimler arasında fark gözetilirse, "iradi örnekleme" yapılmış olur.

SIRA SİZDE

1. Veri derleme kavramını açıklayınız.

2. Birim seçilirken dikkat edilmesi gereken noktaları açıklayınız.

3. Bir istatistik kütlesi hangi nedenlerle sınırlanmak zorundadır, açıklayınız.

Kendimizi Sınayalım

1. Aşağıdakilerden hangisi istatistik birimi olarak **alınamaz?**

a. Coğrafi bölge
b. Doğum
c. Koku
d. Boykot
e. Aile

2. Aşağıdakilerden hangisi ani birimdir?

a. Öğrenci
b. Aile
c. Derslik
d. Kavga
e. Evli çiftler

3. Aşağıdakilerden hangisi sürekli bir değişkendir?

a. Medeni Hal
b. Doğum yılı
c. Bir ülkede yaşayanların sayısı
d. Ülkelerin yüzölçümü
e. Apartmanların daire sayısı

4. Aşağıdakilerden hangisi maddesel bir değişkendir?

a. Doğum tarihi
b. Medeni hal
c. Ebeveyn
d. İşletmelerin kuruluş yeri
e. Günün saatlerine göre ortalama sıcaklık

5. Birimlerle ilgili aşağıda verilen ifadelerden hangisi **yanlıştır?**

a. Tüm olaylar istatistik birimi oluştururlar.
b. Canlı ve cansız varlıklar istatistik birimi olabilirler.
c. Birimin mutlaka maddesel bir varlığa sahip olması gerekmez.
d. Sayılamayan ya da ölçülemeyen olaylar ya da nesneler birim olamazlar.
e. Maddesel bir varlığa sahip birimler, sürekli birimlerdir.

6. Aşağıdakilerden hangisi doğal birim **değildir?**

a. Otomobil
b. Öğrenci
c. Banknot
d. Kitap
e. Uzunluk

7. Kütlelere ilişkin aşağıdaki ifadelerden hangisi doğrudur?

a. Doğal birimlerden oluşan kütleler süreklidir.
b. Bir istatistik kütlesi, istatistik birimlerinin toplamından farklı bir yapıya sahip olabilir.
c. Zaman birimlerinden oluşan kütleler, süreksiz kütlelerdir.
d. Mekan birimlerinden oluşan kütleler, sürekli kütlelerdir.
e. Doğal olmayan birimlerden oluşan kütleler süreksiz kütlelerdir.

8. Aşağıdaki olaylardan hangisi ani veri derlemeye konu oluşturur?

a. Nüfus sayımı
b. Belirli bir yerde ve zaman aralığındaki doğumlar
c. Belirli bir yerde ve zaman aralığındaki iflaslar
d. Belirli bir yerde ve zaman aralığındaki boşanmalar
e. Belirli bir yerde ve zaman aralığındaki grevler

9. Belirlenen amaçlar doğrultusunda hakkında bilgi edinilmek istenen yığının tümüne ne ad verilir?

a. Topluluk
b. Örneklem
c. Grup
d. Örnek
e. Ana Kütle

10. Aşağıdakilerden hangisi varsayımsal kütledir?

a. Bir kütüphanedeki kitaplar
b. Bir üniversitedeki öğrenciler
c. Belirli bir bölgede ve zaman aralığında oluşan trafik kazaları
d. 30 kişilik bir sınıftan rasgele seçilecek farklı dörder kişilik öğrenci grupları
e. Bir fakültedeki öğretim üyeleri

Yanıt Anahtarı

1. c
2. d
3. d
4. b
5. a
6. e
7. d
8. a
9. e
10. d

Yararlanılan Kaynaklar

AVRAOĞLU, Zeki: **İstatistik**, 2. Baskı, Ankara İktisadi ve Ticari İlimler Akademisi Yayını 14, Ankara, 1977.

ÇÖLMEKÇİ, Necla: **Temel İstatistik İlke ve Teknikleri**, 2. Baskı, Bilim Teknik Yayınevi, Eskişehir, 1994.

GÜRTAN, Kenan: **İstatistik ve Araştırma Metodları**, İstanbul Üniversitesi Yayınları, No 2265, İstanbul, 1977.

ÖZMEN, Ahmet: **Uygulamalı Araştırmalarda Örnekleme Yöntemleri**, Anadolu Üniversitesi Yayınları, No 1257, Eskişehir, 2000.

SERPER, Özer: **Uygulamalı İstatistik I**, Filiz Kitabevi, İstanbul, 1986.

TURANLI, M. GÜRİŞ, S. AYAYDIN, A.: **İstatistik Temel Kavramlar ve Uygulamalar**, M. Ü. Nihad Sayar Eğitim Vakfı Yayınları, No 452-685, İstanbul, 1993.

RONALD AYLMER FISHER (1890-1962)

Başlangıçta ilgi alanı astronomi idi. Airy'nin hatalar teorisi üzerinde çalışırken, istatistiğe ilgisi artmıştır.
1919'da rassallık kavramına ilişkin çalışmaları sırasında varyans analizini, 1921'de de Likelihood (Benzerlik) kavramını geliştirmiştir.
1921'de kuramsal yapıyı önde tutarak, istatistiğe ilişkin yeni bir tanım vermiştir.
Sonraki yıllarda küçük örneklemler için uygun yöntemler geliştiren Fisher, modern istatistiğin kurucularından biri olarak görülür.

İstatistik Serileri (Frekans Dağılımları)

Çalışma Biçimine İlişkin Olarak

- *Ünite dikkatle gözden geçirilmeli,*
- *Örnekler dikkatle incelenmelidir.*

Amaçlar

- *Gözlem değerlerinden hareketle istatistik serileri oluşturabileceksiniz.*
- *İstatistik serilerinin grafiklerini çizebileceksiniz.*

İçerik Haritası

GİRİŞ

Derleme sonucunda elde edilen veriler, bir veri yığını oluşturur. Böyle bir yığından belirlenen amaçlar doğrultusunda ihtiyaç duyulan bilgilerin elde edilebilmesi, ancak verilerin belirli esaslara göre düzenlenmesiyle mümkün olur. Bunun için de ilk akla gelen, verileri büyüklüklerine göre sıralamaktır. Böyle bir sıralama sonucu elde edilen rakamlar dizisine, istatistik serisi adı verilir.

Geniş anlamda istatistik serileri, gözlem değerlerinin büyüklüklerine göre sıralanmasıyla oluşturulur.

Bir istatistik serisi gözlem amaçlarına uygun değişken ya da değişkenlerin aldıkları değerlerden (şıklardan) oluştuğu için ilgilenilen yığın olayın gerek yapısı, gerek değişimlerine ilişkin ayrıntılı ipuçları içerir. Bu nedenle de istatistik serileri, ilgili yığın olayın kavranması açısından etkin bir araç olup, istatistik analizlere temel oluşturur.

SERİ TÜRLERİ

AMAÇ 1

Gözlem değerlerinden hareketle istatistik serileri oluşturabileceksiniz.

Değişik ölçütler temel alınarak istatistik serileriyle ilgili farklı sınıflandırmalar yapmak mümkündür. Ancak bu ünitede, zaman serileri ayrı bir ünitede ele alınacağından, zaman ve mekan serilerine kısaca değinilecek, dağılma serileri de yeterli ayrıntıyla ele alınacaktır.

Zaman ve Mekan Serileri

Eğer gözlem sonuçları yıl, ay, hafta, gün ya da saat gibi bir zaman değişkeninin şıklarına göre sıralanırsa, oluşturulan seriye “zaman serisi” adı verilir. Yıllara göre ülke nüfusları ve belirli bir noktada günün saatlerine göre trafik yoğunluğu, bu tür serilere örnek olarak gösterilebilir. Aşağıdaki tabloda, zaman serilerine örnek olarak, Eskişehir ilindeki aylara göre ortalama sıcaklıklar verilmiştir:

Aylar	Ortalama Sıcaklık (C°)
OCAK	-1.5
ŞUBAT	1.3
MART	4.9
NİSAN	10.4
MAYIS	15.1
HAZİRAN	18.8
TEMMUZ	21.4
AĞUSTOS	21.2
EYLÜL	17.1
EKİM	12.0
KASIM	6.7
ARALIK	2.2

Tablo 2.1 *Eskişehir ilindeki aylara göre ortalama sıcaklıklar (72 yıllık gözlem ortalamaları).*

Kaynak: *Türkiye İstatistik Yıllığı 2000.*

Eğer gözlem sonuçları ülke, bölge, şehir ya da köy gibi bir mekan (yer) değişkeninin şıklarına göre sıralanırsa, elde edilen seriye "mekan serisi" adı verilir. Şehirlere göre elektrik tüketimi, bölgelere göre tahıl üretimi bu tür serilere örnek olarak gösterilebilir.

Aşağıdaki tabloda, mekan serisine örnek olarak, bazı illerin denizden yükseklikleri verilmiştir:

Tablo 2.2 *Bazı illerin denizden yükseklikleri.*

Kaynak: *Türkiye İstatistik Yıllığı 2000.*

İller	**Denizden Yükseklik (m)**
ANKARA	891
BALIKESİR	147
ÇANAKKALE	6
DİYARBAKIR	677
ESKİŞEHİR	801
GAZİANTEP	855
İZMİR	29
KARS	1755
MUĞLA	646
RİZE	9
SİVAS	1285
TRABZON	30
VAN	1661
ZONGULDAK	137

Dağılma Serileri

Gözlem sonuçlarının maddesel bir değişkenin şıklarına göre sıralanmasıyla oluşturulan serilere, "dağılma serileri" adı verilir. Dağılma serileri ana çizgileriyle nicel ve nitel dağılma serileri olmak üzere ikiye ayrılır. Ancak bu ünitede, işlemlere elverişli olması nedeniyle sadece nicel dağılma serileri ele alınacaktır. Nicel dağılma serileri de basit seriler (diziler), frekans serileri ve sınıflandırılmış (gruplandırılmış) seriler olmak üzere üç alt başlık altında incelenebilir.

Eğer gözlem sonuçları, bir zaman değişkeninin şıklarına göre sıralanırsa zaman, mekan değişkeninin şıklarına göre sıralanırsa mekan, zaman ve mekan değişkenlerinin dışında bir değişkenin şıklarına göre sıralanırsa dağılma serileri elde edilir.

Eğer derlenen veriler ilgilenilen konunun dışında başka bir yönde, örneğin; gözlem sırasına göre sıralanmışsa, bu sıralamaya "liste" adı verilir. Açıktır ki, derlenen verilerden ihtiyaç duyulan bilgilerin bir liste yardımıyla elde edilmesi veri sayısı arttıkça giderek zorlaşır. Çünkü, her aşamada listedeki sonuçların tekrar tekrar gözden geçirilmesi gerekir.

Konunun kolaylıkla anlaşılabilmesi için, derlenen verilerden hareketle sırasıyla basit, frekans ve sınıflandırılmış serilerin elde edilmeleri aşağıdaki örnek temel alınarak gösterilecektir.

ÖRNEK 1

Tablo 2.3'de bir doğum evinde doğan 100 bebeğin ağırlıkları, doğum sırasına göre verilmiştir:

Tablo 2.3 *Bebek ağırlıkları (kg) (Liste).*

Doğum Sırası	Ağırlık (kg)	Doğum Sırası	Ağırlık (kg)	Doğum Sırası	Ağırlık (kg)	Doğum Sırası	Ağırlık (kg)
001	2.0	**026**	3.0	**051**	2.3	**076**	1.8
002	2.5	**027**	2.0	**052**	2.8	**077**	2.8
003	2.6	**028**	3.3	**053**	2.5	**078**	2.7
004	1.7	**029**	3.5	**054**	2.7	**079**	2.8
005	2.6	**030**	2.6	**055**	1.7	**080**	1.9
006	2.8	**031**	3.5	**056**	2.7	**081**	3.0
007	2.5	**032**	1.7	**057**	2.0	**082**	2.5
008	1.5	**033**	2.8	**058**	3.0	**083**	2.7
009	2.5	**034**	3.1	**059**	2.4	**084**	3.2
010	2.7	**035**	2.3	**060**	2.2	**085**	2.6
011	2.3	**036**	3.1	**061**	2.6	**086**	2.1
012	3.0	**037**	2.9	**062**	2.5	**087**	2.8
013	2.4	**038**	2.5	**063**	1.6	**088**	2.3
014	1.9	**039**	2.5	**064**	2,8	**089**	2.7
015	3.2	**040**	2.7	**065**	2.5	**090**	3.2
016	2.2	**041**	2.6	**066**	3.0	**091**	2.6
017	3.4	**042**	2.2	**067**	2.8	**092**	1.9
018	2.7	**043**	2.8	**068**	2.7	**093**	3.1
019	3.5	**044**	2.1	**069**	1.9	**094**	2.5
020	1.8	**045**	2.1	**070**	2.6	**095**	2.8
021	3.5	**046**	2.4	**071**	2,4	**096**	2.7
022	2.5	**047**	2,8	**072**	3.1	**097**	2.6
023	2.8	**048**	2.5	**073**	2.2	**098**	2.5
024	2.3	**049**	2.7	**074**	3.1	**099**	2.9
025	2.9	**050**	2.6	**075**	2.5	**100**	2.3

Tablo 2.3'de doğan bebeklerin ağırlıkları doğum sırasına göre kaydedildiğinden, oluşturulan tablo bir liste niteliğindedir.
Şimdi bu listeden yararlanarak, 3.2 kg'ın üzerinde kaç bebeğin doğduğu araştırılsın.

ÇÖZÜM

Verilen listenin incelenmesiyle gözlem (doğum) sırasına göre, 017, 019, 021, 028, 029 ve 031'inci sırada doğan bebeklerin 3.2 kg'ın üzerinde olduğu görülür. Ancak bu sonuca ulaşabilmek için, listenin en az bir kez baştan sona kadar gözden geçirilmesi gerekir.

Eğer liste belirlenen amaçlar doğrultusunda düzenlenirse, başka bir anlatımla bir frekans dağılımı oluşturulursa, istenilen sonuçlara daha kısa zamanda ulaşılabilir. Örneğin 3.2 kg'dan daha ağır doğan bebek sayısına, tüm veriler gözden geçirilmeden kolaylıkla ulaşılabilir. Böyle bir sıralama sonucu elde edilen istatistik serisine "basit seri" adı verilir.

100 bebeğin ağırlıkları hafiften ağıra doğru sıralanarak oluşturulan basit seri aşağıda Tablo 2.4'de verilmiştir.

Tablo 2.4 Bebek ağırlıkları (Basit Seri).

Ağırlık (kg)									
1.5	1.9	2.2	2.4	2.5	2.6	2.7	2,8	2,9	3,1
1.6	2.0	2.3	2.5	2.5	2.6	2.7	2,8	3,0	3,2
1.7	2.0	2.3	2.5	2.5	2.6	2.7	2,8	3,0	3,2
1.7	2.0	2.3	2.5	2.5	2.6	2.7	2,8	3,0	3,2
1.7	2.1	2.3	2.5	2.5	2.6	2.7	2,8	3,0	3,3
1.8	2.1	2.3	2.5	2.6	2.7	2.7	2,8	3,0	3,4
1.8	2.1	2.3	2.5	2.6	2.7	2.8	2,8	3,1	3,5
1.9	2.2	2.4	2.5	2.6	2.7	2.8	2,8	3,1	3,5
1.9	2.2	2.4	2.5	2.6	2.7	2.8	2,9	3,1	3,5
1.9	2.2	2.4	2.5	2.6	2.7	2.8	2,9	3,1	3,5

Tablo 2.4 yardımıyla, yapılan gözlemler çerçevesinde en ağır doğan bebeklerin sayısının 4 olduğu bir bakışta kolaylıkla görülebilir.

Eğer en ağır doğan bebekler değil de örneğin 2.5 kg doğan bebeklerin sayısıyla ilgilenilirse, 2.5 kg doğan bebek sayısı tablodan tek tek sayılarak elde edilebilecektir.

Açıkça görülebileceği gibi gözlem sayısı arttıkça, istenilen bilgilere ulaşmak da giderek zorlaşacaktır.

Tablo 2.4 incelendiğinde, gözlem değerlerindeki tekrarlar dikkat çekecektir. Verilerin daha kolay kavranması açısından, gözlem değerlerinin yanına gözlem değerinin kaç kez tekrarlandığı kaydedilerek oluşturulan seriye "frekans serisi", tekrarlara da "frekans" adı verilir.

Tüm bu sözü edilenler doğrultusunda oluşturulan frekans serisi aşağıda verilmiştir:

Ağırlık (kg) X	**Frekans f**
1.5	1
1.6	1
1,7	3
1.8	2
1.9	4
2.0	3
2.1	3
2.2	4
2.3	6
2.4	4
2.5	14
2.6	10
2.7	11
2.8	12
2.9	3
3,0	5
3.1	5
3.2	3
3.3	1
3.4	1
3.5	4
Toplam Frekans	100

Yukarıdaki frekans serisinden, örneğin 2.5 kg doğan bebeklerin sayısının 14 olduğu bir bakışta görülebilmektedir.

Frekans serilerinin basit serilere göre kavranmaları daha kolay olmakla birlikte, yine de ayrıntılıdır. Açıktır ki, gözlem sayısı arttıkça bu tür serilerinde kavranmaları giderek zorlaşır.

Deney ya da gözlem sayıları çok iken, deney ya da gözlem sonuçlarının belirli aralıklar (sınıflar) içinde kalan şıklara göre düzenlenmesiyle oluşturulan istatistik serisine sınıflandırılmış ya da gruplandırılmış seri adı verilir. Örneğin bir doğum evinde dünyaya gelen 100 bebeğin ağırlıkları için farklı büyüklükteki sınıflara göre, aşağıdaki gibi frekans dağılımları oluşturulabilir:

Ağırlık Sınıfları (kg)	**Frekanslar f**	**Ağırlık Sınıfları (kg)**	**Frekanslar f**
1.50 - 1.75	5	1.2 - 1.7	2
1.75 - 2.00	6	1.7 - 2.2	15
2.00 - 2.25	10	2.2 - 2.7	38
2.25 - 2.50	10	2.7 - 3.2	36
2.50 - 2.75	35	3.2 - 3.7	9
2.75 - 3.00	15		**100**
3.00 - 3.25	13		
3.25 - 3.50	2		
3.50 - 3.75	4		
	100		

Sınıflandırılmış seriler oluşturulurken dikkat edilmesi gereken önemli bir nokta, eğer sürekli bir değişkene ilişkin gözlem değerleri sınıflandırılıyorsa, her sınıfın üst sınırıyla onu izleyen sınıfın alt sınırı arasındaki farkın sonsuz küçük olacak şekilde oluşturulması gereğidir. Örneğin yukarıdaki sınıflandırılmış serilerin ilkinde 2.50 - 2.75 sınıfı göz önüne alınsın. 2.50 dahil olmak üzere 2.75'den küçük tüm gözlem değerleri bu sınıf içinde, 2.75 dahil olmak üzere 3.00'den küçük tüm gözlem değerleri de izleyen sınıf içinde yer almalıdır.

Eğer gözlem değerleri sürekli olmayan (kesikli) bir değişkene ilişkinse, örneğin 100 öğrencinin istatistik dersinden aldığı notlar aşağıdaki gibi sınıflandırılabilir:

Not Sınıfları (Puanlar)	**Frekanslar f**
0-15	1
16-31	10
32-47	33
48-63	34
64-79	10
80-95	10
96 ve daha çok	2
	100

Bir sınıfın alt ve üst sınırları arasındaki farka, "sınıf aralığı" ya da "sınıf büyüklüğü" adı verilir ve h ile gösterilir. Yukarıdaki not dağılımı örneğinde sınıf büyüklüğü 15 puandır. Dikkat edilirse aynı seride son sınıf, 96 ve daha çok olarak yer almıştır. Bu durum en büyük puanın 100 olmasından kaynaklanmıştır. Eğer sınıflandırmaya aynı sistematikle devam edilmiş olsaydı, son sınıfın 96 - 111 biçiminde olması gerekirdi. Başlangıç ve bitiş sınırları belirtilmeyen bu tür sınıflara "açık sınıflar" denir. Açık sınıfların kullanılmaları halinde, en küçük ya da en büyük değer bilinemeyeceğinden bazı hesaplamalarda ve grafik çizimlerinde güçlüklerle karşılaşılacaktır. Herhangi bir zorunluluk olmadıkça, açık sınıfların kullanılmasından kaçınılmalıdır.

Kuramda sınıfların oluşturulmasına ilişkin kesin bir kural yoktur. Sınıf sayısını doğrudan araştırmacı belirler. Ancak sınıf sayısının, karşılaşılan özel problemin yapısına ve araştırmanın amaçlarına uygun bir biçimde belirlenmesi gerekir.

Eğer sınıflama yapılırken sınıf aralığı dar seçilirse, sınıf sayısı artar ve frekans dağılımının kavranması giderek zorlaşır. Aksi durumdaysa, sınıf sayısı azalır. Ancak dağılıma ilişkin bazı ayrıntılar gizli kalır.

Uygulamalarda bir frekans dağılımına ilişkin sınıf sayısının 7 - 20 ya da 10 - 30 arasında olmasının uygun sonuçlar verdiği görülmüştür.

Gerçekte verilerin sınıflandırılması çok sayıdaki verinin kavranmasını büyük ölçüde kolaylaştırırken, bazı bilgi kayıplarına da neden olur. Örneğin 100 bebeğin ağırlıklarına ilişkin frekans serisinden 2.6 kg doğan bebek sayısının doğrudan 10 bebek olduğu görülebilir. Buna karşın aynı veri sınıflandırıldığında, sınıflandırılmış seride doğrudan kaç bebeğin 2.6 kg olarak doğduğu görülemez. İlgili veriye ilişkin ilk sınıflandırılmış seri göz önüne alındığında 2.6 kg, 2.50 - 2.75 sınıfının içinde yer almakta ve bu sınıfın frekansı da 35 olarak görülmektedir. Gerçekte 2.50 - 2.75 sınıfında, 2.50 ve 2.7499... kg doğan tüm bebekler yer almaktadır.

Sınıflandırılmış serilerde sınıfları temsil edebilecek değişken değerinin ne olacağı da bir sorun olarak ortaya çıkar. Uygulamada gözlem değerlerinin ilgili sınıf içinde düzgün dağıldığı varsayılarak sınıf orta noktaları, ilgili sınıfa ilişkin değişken değeri olarak kabul edilir.

Konuyla ilgili olarak aşağıdaki örneği dikkatlice gözden geçiriniz.

ÖRNEK 2

Aşağıda verilen frekans dağılımında, ilgili sınıflara karşı gelen değişken değerlerini belirleyiniz.

Sınıflar	**f**
0 – 4	4
4 – 8	10
8 – 12	17
12 – 16	25
16 – 20	14
20 – 24	6
24 – 28	4
	80

ÇÖZÜM

Sınıf orta noktaları ilgili sınıflara ilişkin değişken değeri olacağından, her sınıfın orta noktası ilgili sınıfa değişken değeri olarak atanır.

Sınıflar	f	Sınıf Orta Noktaları X
0 – 4	4	(0+4) / 2 = 2
4 – 8	10	(4+8) / 2 = 6
8 – 12	17	(8+12) / 2 = 10
12 – 16	25	(12+16) / 2 = 14
16 – 20	14	(16+20) / 2 = 18
20 – 24	6	(20+24) / 2 = 22
24 – 28	4	(24+28) / 2 = 26
	80	

Birikimli Seriler

Bir frekans dağılımında, her sınıfın frekansına bir önceki sınıfın frekansı eklenerek oluşturulan seriye "birikimli seri", bu tür oluşturulan frekanslara da "birikimli frekanslar" adı verilir.

Birikimli seriler, küçükten büyüğe ya da büyükten küçüğe doğru oluşturulabilirler. Eğer birikimli seriler küçükten büyüğe doğru oluşturulmuşsa "-den az", büyükten küçüğe doğru oluşturulmuşsa "-den çok" olarak isimlendirilirler.

Bir frekans dağılımında, her sınıfın frekansı kendinden önceki sınıfın frekansına eklenerek oluşturulan seriye "birikimli seri", bu tür frekanslara da "birikimli frekanslar" denilir.

ÖRNEK 3

Bir doğum evinde doğan 100 bebeğe ilişkin sınıflandırılmış seriyi ele alarak küçükten büyüğe ve büyükten küçüğe doğru birikimli serileri oluşturunuz.

ÇÖZÜM

Ağırlık Sınıfları (kg)	Frekanslar f	(-den az)	(-den çok)
1.50 - 1.75	5	5	95 + 5 = 100
1.75 - 2.00	6	6 + 5 = 11	85 + 6 = 95
2.00 - 2.25	10	10 + 11 = 21	79 + 10 = 89
2.25 - 2.50	10	10 + 21 = 31	69 + 10 = 79
2.50 - 2.75	35	35 + 31 = 66	34 + 35 = 69
2.75 - 3.00	15	15 + 66 = 81	19 + 15 = 34
3.00 - 3.25	13	13 + 81 = 94	6 + 13 = 19
3.25 - 3.50	2	2 + 94 = 96	4 + 2 = 6
3.50 - 3.75	4	4 + 96 = 100	4
	100		

Birikimli seriler, uygulamada, genellikle gözlem değerlerinin büyüklüklerine göre kaçıncı sırada yer aldıklarının belirlenmesinde kullanılır. Yukarıdaki örnekte bebeklerin ağırlık sınıfları göz önüne alınırsa -den az serisi yardımıyla, 66 bebeğin ağırlıklarının 2.75 kg'dan daha az olduğu bir bakışta görülebilir. Ayrıca den çok serisi yardımıyla da 34 bebeğin 2.75 kg'dan daha ağır doğduğu doğrudan görülebilir.

Konuyu yeterince pekiştirebilmek için aşağıdaki örneği dikkatlice gözden geçiriniz.

ÖRNEK 4

Sınıflar	-den az
0 – 5	4
5 – 10	6
10 – 15	16
15 – 20	31
20 – 25	51
25 – 30	63
30 - 35	70

serisi verilmiştir.

a. Serideki toplam gözlem sayısını belirleyiniz.

b. Sayısal değeri 25'den küçük gözlem sayısını belirleyiniz.

c. Sayısal değeri 15 ve 15'den büyük, 30'dan küçük gözlem sayısını belirleyiniz.

ÇÖZÜM

a. Verilen seriye göre sayısal değeri 35'den küçük olan gözlem sayısı 70 olduğundan, toplam gözlem sayısı 70'dir.

b. Sayısal değeri 25'den küçük gözlem sayısı 51'dir.

c. Sayısal değeri 30'dan küçük gözlem sayısı 63 ve sayısal değeri 15'den küçük gözlem sayısı 16 olduğundan, sayısal değerleri 15 ile 30 arasındaki gözlem sayısı 63 - 16 = 47 olarak elde edilir.

-den az ve -den çok serileri, frekans serileri için de oluşturulabilir.

ÖRNEK 5

Aşağıda verilen frekans serisi için -den az ve -den çok serilerini oluşturunuz.

X	f
5	3
10	5
15	8
20	6
25	3
30	5
	30

ÇÖZÜM

X	f	(-den az)	(-den çok)
5	3	3	30
10	5	8	27
15	8	16	22
20	6	22	14
25	3	25	8
30	5	30	5
	30		

Bileşik Seriler

Birimlerin birden fazla değişkene göre dağılımlarını bir arada gösteren serilere "bileşik seri" adı verilir. Bir bileşik seri oluşturulurken, ilk sütunda bir değişkenin gözlem değerleri büyüklük sırasına göre yazılırken, diğer sütunlarda da ilgili değişkenlerin ilk değişkene göre durumları yer alır.

Birden fazla değişkenin dağılımlarını bir arada gösteren serilere "bileşik seri" adı verilir.

ÖRNEK 6

Bir sınıftan rasgele seçilen 5 öğrencinin boy uzunlukları ve ağırlıkları aşağıdaki gibidir:

Öğrenci Gözlem No	Uzunluk (m) X	Ağırlık (kg) Y
1	1.72	68
2	1.68	70
3	1.80	76
4	1.74	73
5	1.76	71

İlgili frekans dağılımını oluşturunuz.

ÇÖZÜM

Verilen problemde birim öğrencidir. Boy uzunluğu ve ağırlık ise aynı birim üzerinde tanımlanmış iki farklı değişkendir. Bu duruma göre ilgili frekans dağılımı, bir bileşik seri biçiminde oluşturulmalıdır.
Uzunluk bağımsız, ağırlık da bağımlı değişken olarak alındığında, istenilen frekans dağılımı aşağıdaki gibi olmalıdır.

Uzunluk (m) X	Ağırlık (kg) Y
1.68	70
1.72	68
1.74	73
1.76	71
1.80	76

Eğer ağırlık bağımsız, uzunluk da bağımlı değişken olarak alınırsa, aynı veriye ilişkin bileşik seri;

Ağırlık (kg) Y	Uzunluk (m) X
68	1.72
70	1.68
71	1.76
73	1.74
76	1.80

biçiminde oluşturulur.

SIRA SİZDE

1. **Frekans dağılımı kavramını açıklayınız.**

2. **Bir radar tarafından geliş sırasına göre 50 arabanın hızı (km/saat) aşağıdaki gibi ölçülmüştür:**

82.7	105.6	127.5	107.3	112.2	131.0	105.5	87.9	114.1	116.7
88.4	99.8	101.8	86.5	92.4	80.6	82.4	97.6	125.0	137.0
130.2	87.5	95.6	83.9	94.2	117.3	95.5	120.4	85.7	121.5
103.5	83.0	96.2	95.1	99.9	97.1	134.2	136.3	133.9	81.4
91.3	93.6	88.8	136.0	108.2	129.3	139.1	83.6	97.5	113.1

Yukarıdaki veri kümesini kullanarak sınıf büyüklüğü 5 km/saat ve ilk sınıf da 80 – 85 olacak biçimde bir frekans dağılımı oluşturunuz (NOT: Hızın sürekli bir değişken olduğuna dikkat ediniz).

3. **Bir X değişkenine ilişkin 50 gözlem değeri gözlem sırasına göre aşağıda verilmiştir:**

11	5	18	38	13	7	32	20	14	6
3	16	36	19	27	9	16	47	9	20
12	11	24	21	42	43	19	49	28	15
9	3	13	42	5	12	27	32	41	48
6	7	44	4	23	29	41	8	40	5

Yukarıdaki veri kümesini kullanarak ve ilk sınıf 2 – 6 olacak şekilde bir frekans dağılımı oluşturunuz (NOT: X değişkeninin kesikli bir değişken olduğuna dikkat ediniz).

SERİLERİN GRAFİKLE GÖSTERİLMESİ

AMAÇ 2

İstatistik serilerinin grafiklerini çizebileceksiniz.

Dağılma serilerine ilişkin grafikler gözden geçirilirken önce frekans ve sınıflandırılmış serilerin, sonra da birikimli ve bileşik serilerin grafiklerine değinilecektir.

Frekans Serilerinin Grafikle Gösterilmesi

Anımsanacağı gibi, frekans serileri biri gözlem değerleri diğeri de gözlem değerlerine karşı gelen frekansları gösteren iki sütundan oluşur. Frekans serilerinde frekanslar gözlem değerlerine göre değiştiğinden gözlem değerleri yatay eksende, frekanslarsa dik eksende gösterilir. Grafik, yatay eksende belirlenen değerlerden uzunlukları ilgili frekanslar kadar olan dik doğru parçalarıyla oluşturulur. Bu tür grafiklere "çubuk grafik" adı verilir.

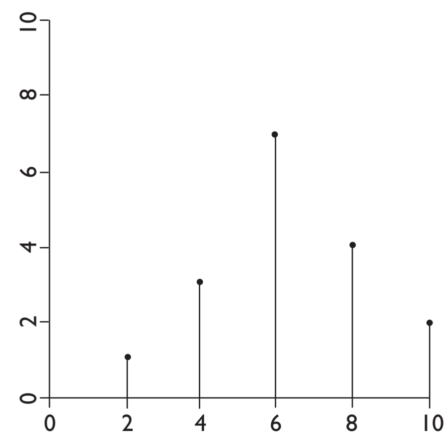
ÖRNEK 7

X değişkenine ilişkin gözlem sonuçları aşağıda bir frekans serisi halinde verilmiştir. Verilen serinin grafiğini çiziniz.

X	f
2	1
4	3
6	7
8	4
10	2
	17

ÇÖZÜM

Gözlem değerleri yatay, frekanslar da dik eksende gösterilerek grafik aşağıdaki gibi oluşturulur:



Şekil 2.1 *Çubuk grafik.*

Sınıflandırılmış Serilerin Grafikle Gösterilmesi

Sınıflandırılmış seriler, "histogram" ya da "frekans poligonu" adı verilen grafiklerle gösterilirler. Adı geçen grafiklerin çizimi, aşağıda ayrıntıları ile ele alınmıştır.

Histogram

Histogram ve frekans poligonu sınıflandırılmış serilerin grafikleridir.

Histogram; alanı ilgili sınıfın frekansına ve tabanı da ilgili sınıfın aralığına eşit, birbirine bitişik dikdörtgenlerden oluşan bir grafik gösterimdir.

Bir histogram çizilmeden önce, sözü edilen dikdörtgenlerin uzunluklarının ayarlanması gerekir. Bunun için frekanslar sınıf aralığına bölünerek, dikdörtgenlerin alanları ilgili sınıfların frekanslarına eşit hale getirilir.

Konuyla ilgili olarak aşağıdaki örneği dikkatlice gözden geçiriniz.

ÖRNEK 8 ***Sürekli bir X değişkenine ilişkin gözlem sonuçları aşağıdaki seriyle verilmiştir:***

Sınıflar	frekanslar
0 – 4	12
4 – 8	16
8 – 12	20
12 – 16	24
16 – 20	20
20 – 24	8
	100

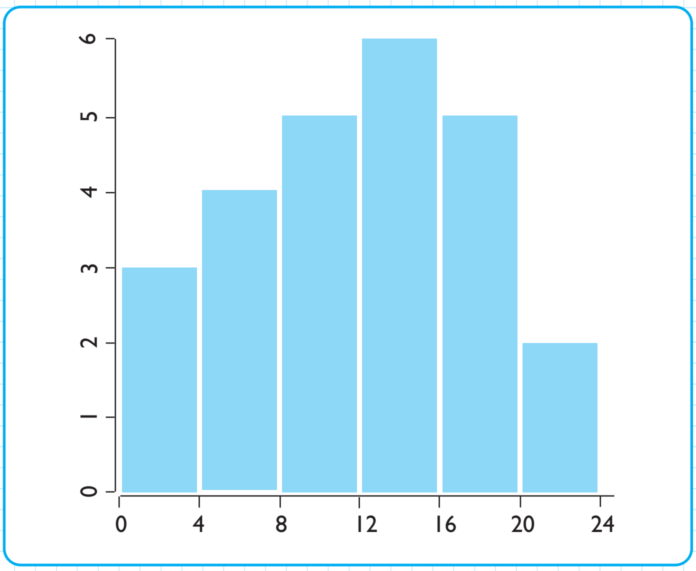
Verilen serinin histogramını çiziniz.

ÇÖZÜM

Histogramın çizilebilmesi için öncelikle frekansların ayarlanması gerekir. Ayarlanmış frekansların elde edilişleri aşağıda gösterilmiştir.

Sınıflar	f	Sınıf Aralıkları h	Ayarlanmış Frekanslar f / h
0 – 4	12	4	12 / 4 = 3.0
4 – 8	16	4	16 / 4 = 4.0
8 – 12	20	4	20 / 4 = 5.0
12 – 16	24	4	24 / 4 = 6.0
16 – 20	20	4	20 / 4 = 5.0
20 – 24	8	4	8 / 4 = 2.0
	100		

Birinci ve son sütundan yararlanılarak histogram aşağıdaki gibi çizilir:



Şekil 2.2 *Eşit aralıklı sınıflar için histogram.*

Eğer verilen seride sınıf aralıkları eşit değilse, histogram yine aynı yöntemle oluşturulur.

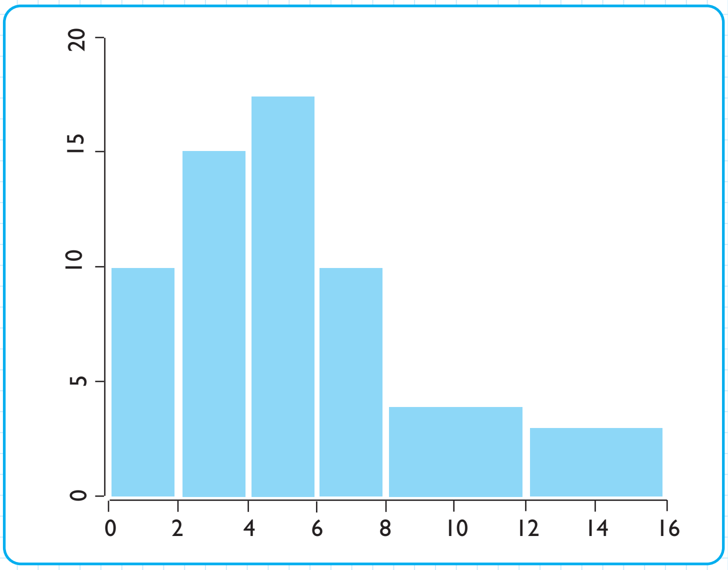
ÖRNEK 9

Aşağıdaki serinin histogramını çiziniz.

Sınıflar	f
0 – 2	20
2 – 4	30
4 – 6	36
6 – 8	20
8 – 12	16
12 – 16	12
	128

ÇÖZÜM

Sınıflar	f	h	f / h
0 – 2	20	2	10.0
2 – 4	30	2	15.0
4 – 6	36	2	18.0
6 – 8	20	2	10.0
8 – 12	16	4	4.0
12 – 16	12	4	3.0
	134		



Şekil 2.3 *Farklı büyüklükteki sınıflar için histogram.*

Frekans Poligonu

Frekans poligonu, histogramın tepe orta noktalarının birleştirilmesiyle elde edilen, sınıflandırılmış serilere ilişkin, diğer bir grafik türüdür. Histogramın tepe orta noktaları, ilgili sınıflara ilişkin değişkenlerin değerlerini ifade ettiğinden, frekans poligonu, değişkenlerin değerlerine göre oluşturulmuş bir grafiktir. Konuyla ilgili bir örnek aşağıda verilmiştir.

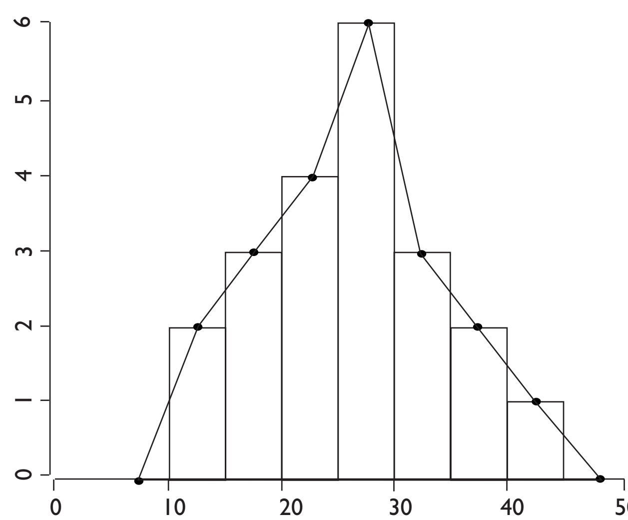
ÖRNEK 10 ***Aşağıda verilen serinin frekans poligonunu çiziniz.***

Sınıflar	frekanslar
10 – 15	10
15 – 20	15
20 – 25	20
25 – 30	30
30 – 35	15
35 – 40	10
40 – 45	5
	105

ÇÖZÜM

Öncelikle ayarlanmış frekanslar oluşturularak histogram, sonra da histogramın tepe orta noktaları birleştirilerek frekans poligonu elde edilir.

Sınıflar	f	X	h	f / h
10 – 15	10	12.5	5	2
15 – 20	15	17.5	5	3
20 – 25	20	22.5	5	4
25 – 30	30	27.5	5	6
30 – 35	15	32.5	5	3
35 – 40	10	37.5	5	2
40 – 45	5	42.5	5	1
	105			



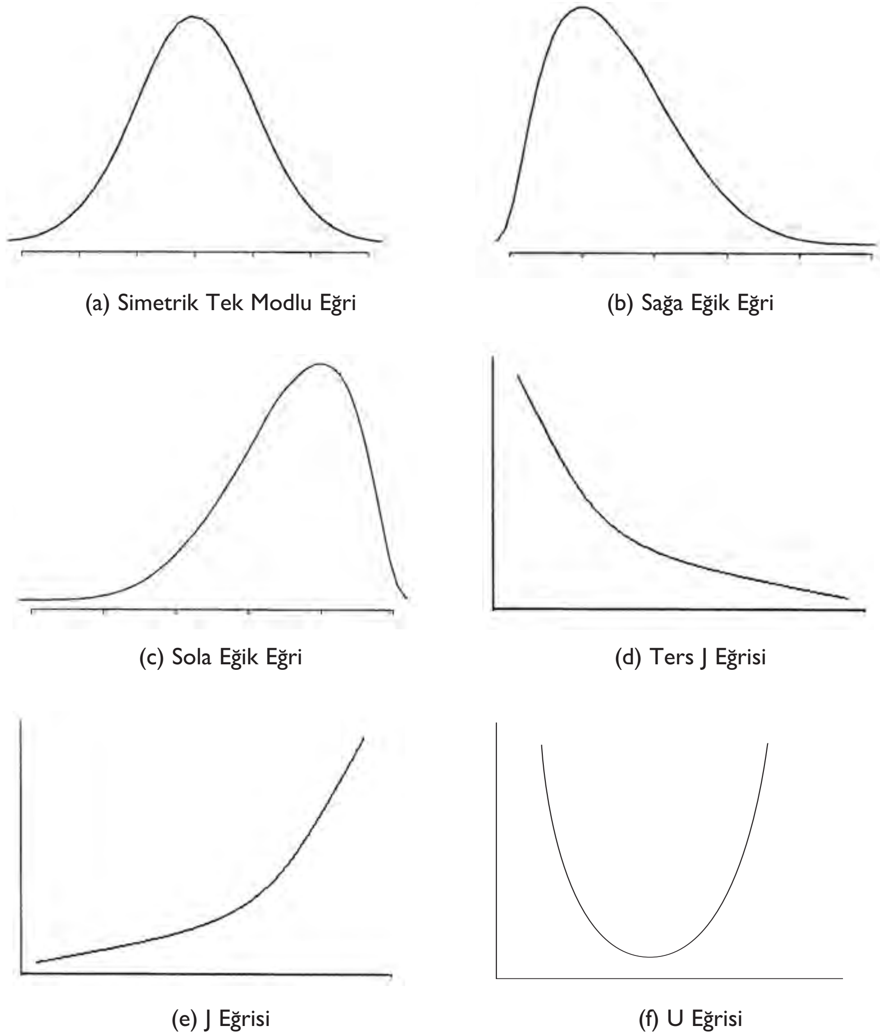
Şekil 2.4 *Frekans poligonu.*

Frekans poligonun altında kalan alan frekanslar toplamına eşittir.

Histogram ve frekans poligonunun sürekli değişkenler için uygun grafikler olduğuna dikkat ediniz.

Eğer gözlem sayısı artarken sınıf aralığı sonsuz küçültülürse, frekans poligonu bir frekans eğrisi şekline dönüşür.

Uygulamada sıkça karşılaşılan frekans eğrileri Şekil 2.5'de gösterilmiştir.



Şekil 2.5 *Bazı frekans eğrileri.*

Birikimli Serilerin Grafikle Gösterilmesi

Birikimli serilerin grafikleri çizilirken sınıflar yatay, birikimli frekanslarsa dik eksende gösterilir.

-den az serilerinin grafikleri çizilirken, koordinat sisteminde sınıf üst sınırlarıyla ilgili sınıfa karşı gelen birikimli frekansların belirledikleri noktalar birleştirilerek grafik oluşturulur.

-den çok serilerinin grafikleri oluşturulurken, sınıf alt sınırlarıyla ilgili sınıfa karşı gelen birikimli frekansların belirledikleri noktalar birleştirilerek çizim tamamlanır.

-den az eğrisi ilk sınıfın frekansından başlayarak sürekli artan, -den çok eğrisi ise son sınıfın frekansına kadar sürekli azalan bir eğridir.

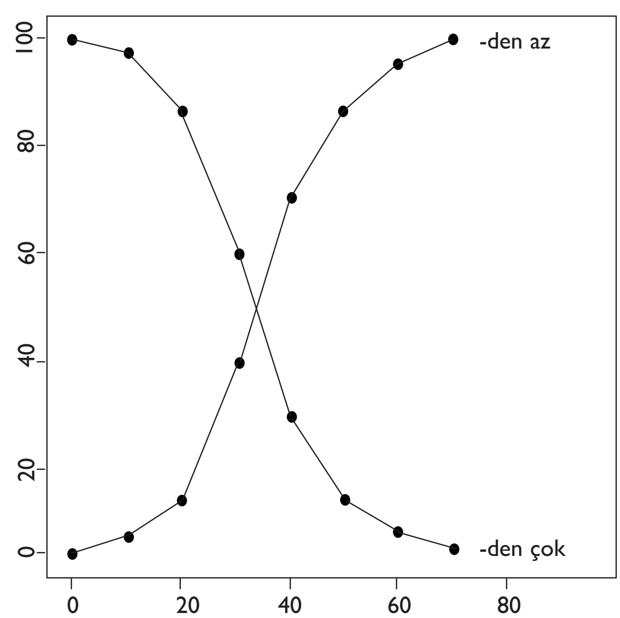
Konuyla ilgili bir örnek aşağıda verilmiştir.

ÖRNEK 11 ***Aşağıda verilen seri için -den az ve -den çok serilerini oluşturarak, elde ettiğiniz serilerin grafiklerini çiziniz.***

Sınıflar	Frekanslar
0 – 10	3
10 – 20	12
20 – 30	25
30 – 40	30
40 – 50	15
50 – 60	10
60 – 70	5
	100

ÇÖZÜM

Sınıflar	Frekanslar	-den az	-den çok
0 – 10	3	3	100
10 – 20	12	15	97
20 – 30	25	40	85
30 – 40	30	70	60
40 – 50	15	85	30
50 – 60	10	95	15
60 – 70	5	100	5
	100		



Şekil 2.6 *Birikimli serilerin grafikleri.*

Bileşik Serilerin Grafikle Gösterilmesi

Bileşik serilerin grafikleri oluşturulurken, ilk değişkenin değerleri yatay, diğer değişkenin değerleriyse dik eksende yer alır. Bu değerlere koordinat sisteminde karşı gelen noktalar belirlenerek grafik elde edilir.

Bileşik serilerin grafiklerine serpilme diyagramı adı da verilir.

Aşağıda bileşik serilerin grafiklerinin çizilmesine ilişkin bir örnek verilmiştir.

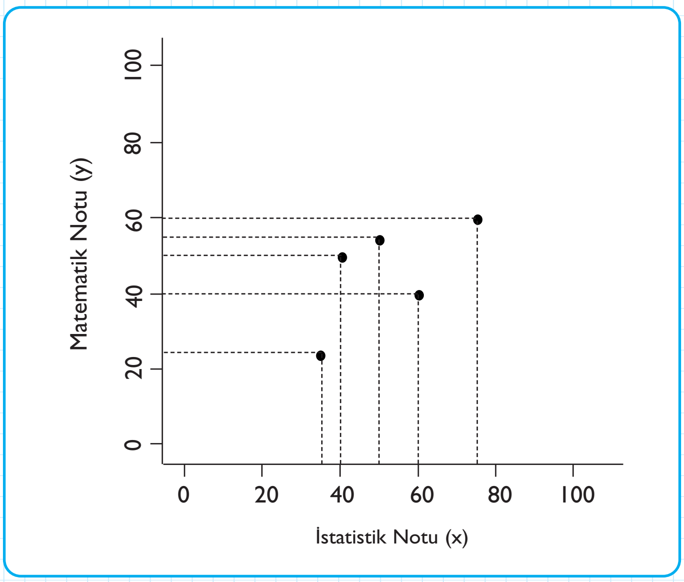
ÖRNEK 12

Aşağıda 5 öğrencinin istatistik ve matematik derslerinden aldıkları notlar bir bileşik seri biçiminde verilmiştir.

İstatistik Notu X	Matematik Notu Y
35	25
40	50
50	55
60	40
75	60

Verilen serinin grafiğini çiziniz.

ÇÖZÜM



Şekil 2.7 *Serpilme diyagramı.*

SIRA SİZDE

1. Bir histogram oluşturulurken, dikkat edilmesi gereken noktaları belirleyiniz.

2. Bir sınıftaki 50 öğrencinin boyları (m) gözlem sırasına göre aşağıda verilmiştir:

1.62	1.56	1.86	1.65	1.79	1.72	1.80	1.84	1.90	1.94
1.74	1.92	1.89	1.64	1.71	1.66	1.79	1.76	1.77	1.83
1.58	1.75	1.56	1.78	1.87	1.76	1.92	1.76	1.73	1.78
1.63	1.68	1.78	1.76	1.55	1.78	1.64	1.87	1.70	1.68
1.74	1.83	1.67	1.88	1.93	1.81	1.70	1.83	1.68	1.79

a. Verilen veri kümesini kullanarak ve ilk sınıf 1.53 – 1.58 olacak biçimde eşit aralıklı bir frekans dağılımı oluşturunuz.
b. Oluşturduğunuz frekans dağılımının histogramını çiziniz.

3. Bir karayolunda meydana gelen hasarlar arası uzaklıklar (km) aşağıdaki gibidir:

21.23	9.23	13.27	23.34	75.47	56.25	62.51	60.64	30.21	50.26
8.45	76.48	18.83	68.23	22.78	47.94	67.29	21.71	37.58	36.45
63.71	17.68	52.91	31.96	59.45	45.26	28.75	40.17	42.95	47.90
12.36	25.57	43.41	71.13	64.32	37.93	70.36	31.29	72.27	71.80
19.44	54.21	34.84	57.65	48.19	53.81	43.34	10.24	45.90	18.45

a. Yukarıdaki veri kümesini kullanarak, uygun göreceğiniz bir sınıf aralığına göre frekans dağılımını oluşturunuz.
b. Oluşturduğunuz frekans dağılımının frekans poligonunu çiziniz.

Kendimizi Sınayalım

1. Birimlerin iki değişkene göre dağılımlarını bir arada gösteren serilere ne ad verilir?

a. Basit seri
b. Frekans serisi
c. Sınıflandırılmış seri
d. Birikimli seri
e. Bileşik seri

2.

Sınıflar	Frekanslar
0 – 5	2
5 – 10	4
10 – 15	7
15 – 20	13
20 – 25	3
25 – 30	1
	30

Yukarıda verilen serinin, frekans eğrisinin görünümü aşağıdakilerden hangisidir?

a. Sağa eğik
b. Sola eğik
c. U
d. J
e. Ters J

3.

X	f
2	1
4	3
6	7
8	7
10	10
12	5
14	2
	35

Yukarıda verilen seri için **-den az serisi** oluşturulmak istendiğinde, birikimli frekanslar aşağıdakilerden hangisidir?

a. -den az	b. -den az	c. -den az	d. -den az	e. -den az
4	1	35	1	1
11	4	34	3	4
18	11	31	10	11
28	18	24	17	18
33	28	17	27	28
35	33	7	32	34
40	35	2	35	35

4. Bir frekans dağılımına ilişkin **-den az serisi** aşağıda verilmiştir.

Sınıflar	-den az
2 – 8	6
8 – 14	10
14 – 20	17
20 – 26	22
26 – 32	34
32 – 38	37
38 – 44	45
44 – 50	46

Yukarıda verilen seriye ilişkin gözlenen frekanslar aşağıdakilerden hangisidir?

a. f	b. f	c. f	d. f	e. f
2	6	6	3	6
8	4	4	7	4
7	8	7	7	3
5	3	5	5	6
4	11	12	8	7
10	9	3	9	5
8	5	8	4	5
4	1	1	14	9

5. Bir frekans dağılımına ilişkin **-den çok serisi** aşağıda verilmiştir.

Sınıflar	-den çok
13,5 – 17,5	56
17,5 – 21,5	51
21,5 – 25,5	43
25,5 – 29,5	39
29,5 – 33,5	32
33,5 – 37,5	25
37,5 – 41,5	12
41,5 – 45,5	10

Yukarıda verilen seriye ilişkin gözlenen frekanslar aşağıdakilerden hangisidir?

a.

f
9
4
3
7
11
13
2
10

b.

f
10
2
13
7
7
4
8
5

c.

f
7
9
4
4
3
6
12
5

d.

f
5
8
4
7
7
13
2
10

e.

f
6
9
2
1
12
13
2
10

6.

Gözlem No	İşletme Notu	Ekonomi Notu
1	70	90
2	50	70
3	80	60
4	50	40
5	60	50

5 öğrencinin işletme ve ekonomi derslerinden aldıkları notları gösteren yukarıdaki tabloya ilişkin aşağıdaki ifadelerden hangisi doğrudur?

a. Tablodaki veriler basit bir seri oluşturur.
b. Tablodaki veriler bir frekans serisi oluşturur.
c. Tablodaki veriler sınıflandırılmış bir seri oluşturur.
d. Tablodaki veriler bileşik bir seri oluşturur.
e. Tablodaki veriler bir istatistik serisi oluşturmaz.

7. Gözlem sonuçlarının maddesel bir değişkenin şıklarına göre sıralanmasıyla oluşturulan serilere ne ad verilir?

a. Zaman serisi
b. Bileşik seri
c. Mekan serisi
d. Birikimli seri
e. Dağılma serisi

8. Serilere ilişkin aşağıdaki ifadelerden hangisi doğrudur?

a. Bölgelere göre tahıl üretimi zaman serileri için uygun bir örnektir.
b. Derlenen veriler ilgilenilen konunun dışında bir temele göre sıralanmışsa, bu sıralamaya liste adı verilir.
c. Gözlem değerlerinin yanına, gözlenen değerlerin tekrar sayısı yazılarak oluşturulan seriye, sınıflandırılmış seri denir.
d. Eğer sınıflandırma yapılırken sınıf aralığı küçük seçilirse, ilgili dağılıma ilişkin bazı ayrıntılar gizli kalır.
e. Uygulamalarda 20 ile 50 arası sınıf sayısı, en uygun sınıf sayısıdır.

9. Aşağıda bir frekans dağılımına ilişkin -den az ve -den çok serileri birlikte verilmiştir.

Sınıflar	-den az	-den çok
0 – 4	1	50
4 – 8	5	49
8 – 12	12	45
12 – 16	22	38
16 – 20	34	28
20 – 24	44	16
24 - 28	50	6

Yukarıdaki tabloya göre, sayısal değeri 20'den küçük gözlem sayısı kaçtır?

a. 44
b. 38
c. 34
d. 22
e. 16

10. Serilere ilişkin aşağıdaki ifadelerden hangisi **yanlıştır?**

a. -den az serileri, her sınıfın frekansına bir önceki sınıfın frekansı eklenerek oluşturulur.
b. Basit serilerin grafiklerine histogram denir.
c. Bileşik serilerin grafiklerine serpilme diyagramı denir.
d. Bir seriye ilişkin frekans poligonunun altında kalan alan, seriye ilişkin frekanslar toplamına eşittir.
e. Bir seriye ilişkin histogramda, dikdörtgenlerin alanları toplamı seriye ilişkin frekanslar toplamına eşittir.

Yanıt Anahtarı

1. e
2. b
3. b
4. c
5. d
6. e
7. e
8. b
9. c
10. b

Yararlanılan Kaynaklar

ÇÖMLEKÇİ, Necla: **Temel İstatistik İlke ve Teknikleri**, 2. Baskı, Bilim Teknik Yayınevi, Eskişehir, 1994.

GÜRTAN, Kenan: **İstatistik ve Araştırma Metodları**, İstanbul Üniversitesi Yayınları, No 2265, İstanbul, 1977.

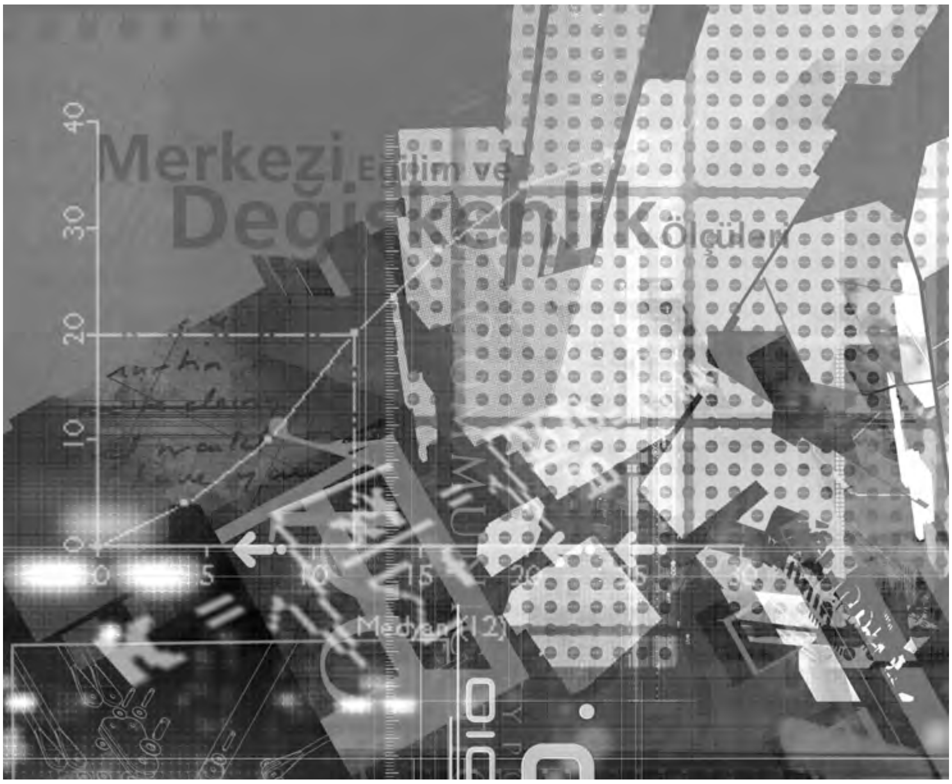
HARPER, W.M.: **Statistics**, 4. ed., Pitman Pub. Comp., 1988.

JOHNSON, Robert: **Elementary Statistics**, 6. ed., PSW-KENT Pub. Comp., Boston, 1992.

MELNYK, M.: **Principle of Applied Statistics**, Pergamon Press Inc., New York, 1974.

3

Merkezi Eğilim ve Değişkenlik Ölçüleri



Çalışma Biçimine İlişkin Olarak:

- *Önceki ünitelerde verilen kavramları yeniden gözden geçirmeli,*
- *Verilen örnekleri dikkatle incelemelidir.*

Amaçlar:

- *Merkezi eğilim ölçüleri kavramını açıklayabilecek ve istatistik serilerinin ortalamalarını hesaplayabileceksiniz.*
- *Değişkenlik ölçüleri kavramını açıklayabilecek ve istatistik serilerine ilişkin değişkenlik ölçülerini hesaplayabileceksiniz.*

İçindekler

GİRİŞ

Merkezi eğilim ölçüleri, adının da çağrıştıracağı gibi bir veri kümesinin ortasını belirleme eğiliminde olan sayısal değerlerdir. Ortalama terimi genelde bu ölçülerle ilgilidir.

Bu ünitede öncelikle merkezi eğilim ölçüleri (ortalamalar) ele alınacak, sonra da değişkenlik ölçülerine yer verilecektir.

MERKEZİ EĞİLİM ÖLÇÜLERİ (ORTALAMALAR)

AMAÇ 1

Merkezi eğilim ölçüleri kavramını açıklayabilecek ve istatistik serilerinin ortalamalarını hesaplayabileceksiniz.

Bir ortalama ile, nüfus, hız, ışık yılı, ısı ve benzeri gibi ölçülebilen ya da sayılabilen bir olay ya da nesneye ilişkin derlenen veri kümesini temsil edebilen, tek bir değer hesaplanır. Ancak bir kaç tane olan merkezi eğilim ölçülerinin her biri, aynı veri kümesi için farklı bir tablo çizer.

Geniş anlamda ortalama, bir istatistik serisindeki gözlem değerlerinin, etrafında toplanma eğilimi gösterdiği değer olarak tanımlanır.

Konuya açıklık kazandırması açısından aşağıdaki örneği göz önüne alalım.

9 dairelik bir apartmanda oturan ailelerin aylık gelirleri YTL olarak aşağıdaki gibi olsun.

Ortalama, bir seride en küçük değerle en büyük değer arasında yer alır. (X_{min} < ortalama < X_{max})

520, 580, 670, 700, 700, 700, 860, 1000, 1200

Bu ailelerin normal geliri nedir sorusunun cevabı, muhtemelen gelirlerin ortalamasıdır biçiminde olacaktır.

İzleyen kesimlerde örnekte sözü edilen normal gelirin hesaplanmasında kullanılan merkezi eğilim ölçüleri, başka bir anlatımla ortalamalar ayrıntılarıyla gözden geçirilecektir.

Ana çizgileriyle ortalamalar, duyarlı ve duyarlı olmayan ortalamalar olmak üzere, iki ana başlık altında incelenebilir.

Duyarlı Ortalamalar

Duyarlı ortalamalar, serideki tüm gözlem değerlerinden etkilenen ortalamalardır.

Bu ünitede duyarlı ortalamalardan sadece aritmetik, geometrik ve kareli ortalamalar ele alınacaktır.

Aritmetik Ortalama

Aritmetik ortalama, bir seriyi oluşturan gözlem değerleri toplamının, gözlem sayısına oranı olarak tanımlanır.

Seriyi oluşturan gözlem değerleri $x_1, x_2, ..., x_n$ aritmetik ortalama da $\bar{x}$ ile gösterilirse tanım uyarınca,

Bir seride aritmetik ortalama, seriyi oluşturan gözlem değerleri toplamı gözlem sayısına bölünerek hesaplanır.

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

olarak hesaplanır.

En kolay hesaplanan ve en çok kullanılan ortalama, aritmetik ortalamadır. Eğer ne tür olduğu belirtilmeden bir ortalamadan söz ediliyorsa, muhtemelen kastedilen aritmetik ortalamadır.

ÖRNEK 1 ***Yukarıda verilen 9 dairelik apartmanda oturan ailelerin normal gelirini, aritmetik ortalama kullanarak hesaplayınız.***

ÇÖZÜM

X, ailelerin aylık gelirlerini göstermek üzere gelirlerden oluşan basit seri aşağıdaki gibi olacaktır :

x (YTL)
520
580
670
700
700
700
860
1000
1200
$\sum x =$ **6930**

Ailelerin toplam geliri 6930 YTL olduğundan tanım doğrultusunda, toplam gelir aile sayısına bölünerek ortalama gelir,

$$\bar{x} = \frac{\sum x}{n} = \frac{6930}{9} = 770 \text{ YTL}$$

olarak hesaplanır.

ÖRNEK 2 ***Aşağıda verilen basit serinin aritmetik ortalamasını hesaplayınız.***

x
10
13
16
20
26
85

ÇÖZÜM

Gözlenen değerlerin toplamı 85 ve gözlem sayısı da 5 olduğundan

$$\bar{x} = \frac{\sum x}{n} = \frac{85}{5} = 17$$

olarak hesaplanır.

Öte yandan frekans serilerinde her gözlem değeri frekansı kadar tekrarlandığından, aritmetik ortalama hesaplanırken gözlem değerleri frekanslarıyla çarpılarak toplanır ve bu sonuç frekanslar toplamına bölünür.

Aşağıdaki örneği dikkatle inceleyiniz.

ÖRNEK 3

Aşağıda verilen frekans serisinin aritmetik ortalamasını hesaplayınız.

x	f
10	2
12	3
15	6
19	4
21	1
	16

ÇÖZÜM

Verilen seri, 2 tane 10, 3 tane 12, 6 tane 15, 4 tane 19 ve 1 tane de 21 değerinden oluşmuştur.
16 gözlem değerinin toplamı,

$$\sum x = \underbrace{10 + 10}_{2} + \underbrace{12 + 12 + 12}_{3} + \underbrace{15 + 15 + 15 + 15 + 15 + 15}_{6} + \underbrace{19 + 19 + 19 + 19}_{4} + \underbrace{21}_{1}$$

$$\sum x = 2(10) + 3(12) + 6(15) + 4(19) + 21$$
$$= 20 + 36 + 90 + 76 + 21$$
$$= 243$$

olarak elde edilir.
Bu toplam, gözlem değeri frekanslar ile çarpılarak aşağıdaki gibi kolaylıkla elde edilebilir.

x	f	xf
10	2	10. 2 = 20
12	3	12. 3 = 36
15	6	15. 6 = 90
19	4	19. 4 = 76
21	1	21. 1 = 21
	16	**243**

Hesaplanan gözlem değerleri toplamı $\left(\sum xf\right)$, frekanslar toplamına $\left(\sum f\right)$ bölünerek aritmetik ortalama,

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{243}{16} = 15{,}1875$$

olarak hesaplanır.
Örnekten de görülebileceği gibi, frekans serilerinde aritmetik ortalama,

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_n f_n}{f_1 + f_2 + \ldots + f_n} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i}$$

ile hesaplanır.

ÖRNEK 4 ***Aşağıda verilen serinin aritmetik ortalamasını hesaplayınız.***

x	f
10	12
15	20
20	25
25	25
30	15
35	3
	100

ÇÖZÜM

x	f	xf
10	12	120
15	20	300
20	25	500
30	25	625
35	15	450
30	3	105
	100	**2100**

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{2100}{100} = 21$$

olarak elde edilir.

Aritmetik ortalama sınıflandırılmış serilerde de frekans serilerinde olduğu gibi hesaplanır. Ancak dikkat edilmesi gereken, değişken değerleri olarak sınıf orta noktalarının alınmasıdır.

ÖRNEK 5 ***Aşağıda verilen serinin aritmetik ortalamasını hesaplayınız.***

Sınıflar	f
10 -14	4
14 - 18	5
18 -22	8
22 - 26	6
26 -30	2
	$\sum f = \mathbf{25}$

ÇÖZÜM

Sınıflar	f	x	xf
10 - 14	4	12	48
14 - 18	5	16	80
18 - 22	8	20	160
22 - 26	6	24	144
26 - 30	2	28	56
	$\sum f = \mathbf{25}$		$\sum xf = \mathbf{488}$

Buradan,

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{488}{25} = 19{,}52$$

olarak elde edilir.
Ancak dikkat etmek gerekir ki, sınıflamadaki kayıplar nedeniyle, sınıflandırılmış serilerde aritmetik ortalama, yaklaşık olarak hesaplanabilmektedir.

Aritmetik Ortalamanın Özellikleri

- Aritmetik ortalama duyarlı bir ortalamadır ve serideki aşırı değerlerden doğrudan etkilenir.

ÖRNEK 6

Aritmetik ortalama aşağıdaki serilerin hangisinde daha temsilidir?

x	f	u
23	1	23
25	25	25
26	26	26
28	28	100
102	**80**	**174**

ÇÖZÜM

$$\bar{x} = \frac{\sum x}{n} = \frac{102}{4} = 25.5$$

$$\bar{y} = \frac{\sum y}{n} = \frac{80}{4} = 20.0$$

$$\bar{u} = \frac{\sum u}{n} = \frac{174}{4} = 43.5$$

Görüleceği gibi serideki bir tek değerin değişmesi bile, ortalamayı etkilemektedir. x serisinin ortalaması seriyi oluşturan gözlem değerlerine oldukça yakın, başka bir anlatımla temsil yeteneği daha yüksektir. y ve u serilerinde ise gözlem değerlerindeki aşırı kıymetlerin büyüklüğüne bağlı olarak, ortalamaların temsil yeteneği azalmıştır.

- Gözlem değerlerinin aritmetik ortalamadan cebirsel sapmalarının toplamı sıfırdır.

Başka bir anlatımla $\left(\sum_{i=1}^{n}(x_i - \bar{x}) = 0\right)$ olur.

Bu özellik aşağıdaki örnek üzerinde gösterilmiştir :

ÖRNEK 7

x	x - $\bar{x}$
10	10 - 30 = -20
20	20 - 30 = -10
30	30 - 30 = 00
40	40 - 30 = 10
50	50 - 30 = 20
150	$\sum(x - \bar{x}) = 00$

$$\bar{x} = \frac{\sum x}{n} = \frac{150}{5} = 30$$

Eğer verilen seri bir frekans serisiyse, her gözlem değerinden aritmetik ortalama çıkartılır ve ilgili gözlem değerinin frekansıyla çarpıldıktan sonra, toplam değer hesaplanır. İşlemler aşağıdaki örnek üzerinde gösterilmiştir :

ÖRNEK 8

x	**f**	**xf**	**(x - $\bar{x}$)**	**(x - $\bar{x}$) f**
50	1	50	50 - 71.875 = -21.875	-21.875
60	3	180	60 - 71.875 = -11.875	-35.625
70	6	420	70 - 71.875 = -1.875	-11.250
80	4	320	80 - 71.875 = 8.125	32.500
90	2	180	90 - 71.875 = 18.125	36.250
	16	1150		-68.750
				68.750
				$\sum(x - \bar{x}) = 00{,}000$

$$\bar{x} = \frac{\sum xf}{f} = \frac{1150}{16} = 71.875$$

- Gözlem değerlerinin aritmetik ortalamadan cebirsel sapmalarının kareleri toplamı minimumdur.
  Bu özellik de aşağıdaki örnek üzerinde gösterilmiştir:

ÖRNEK 9

x	$(x-\bar{x})$	$(x-\bar{x})^2$
10	-20	400
20	-10	100
30	00	000
40	10	100
50	20	400
		$\sum(x-\bar{x})^2 = 1000$

$\bar{x} = 30$

Verilen X serisinde, aritmetik ortalamadan daha küçük (25) ya da daha büyük bir değer (40) çıkartılırsa sonuçlar,

x	$x - 25$	$(x-25)^2$	$(x-40)$	$(x-40)^2$
10	-20	225	-30	900
20	-5	25	-20	400
30	5	25	-10	100
40	15	225	00	000
50	25	625	10	100
		$\sum(x-25)^2 = 1125$		$\sum(x-40)^2 = 1500$

$$\sum(x-25)^2 = 1125 \text{ ve } \sum(x-40)^2 = 1500$$

olarak elde edilir.

Aritmetik ortalamadan cebirsel sapmaların kareleri toplamı 1000'dir. Ancak bu ortalamadan küçük (25) ve büyük (40) değerler çıkartıldığında, görüleceği gibi cebirsel sapmaların kareleri toplamı 1000'den büyük çıkmaktadır.

Tartılı Aritmatik Ortalama

Eğer bir seriyi oluşturan gözlem değerleri arasında önem derecesine göre farklar varsa ve bu farklar ortalama hesabında göz önüne alınmak isteniyorsa, böyle durumlarda tartılı ortalama hesaplanır.

Gözlem değerleri arasındaki önem derecesine göre farklar, ortalama hesaplanırken göz önüne alınmak istenirse, tartılı ortalama hesaplanmalıdır.

t, tartıyı $\bar{x}_t$'de tartılı aritmetik ortalamayı göstermek üzere, tartılı aritmetik ortalama basit serilerde,

$$\bar{x}_t = \frac{\sum xt}{\sum t}$$

frekans ve sınıflandırılmış serilerde ise,

$$\bar{x}_t = \frac{\sum xtf}{\sum tf}$$

eşitlikleriyle hesaplanır.

ÖRNEK 10

İktisadi ve İdari Bilimler Fakültesi İşletme Bölümü'ndeki birinci sınıf öğrencisinin güz döneminde aldığı dersler, başarı notları, başarı notlarının katsayıları ve kredi değerleri aşağıda verilmiştir:

Dersler	Başarı Notları	Katsayılar	Kredi Değerleri
Genel Matematik I	AA	4.0	5
Türkçe I	AB	3.7	2
Makro Ekonomi I	CC	2.0	3
Genel işletme	BC	2.7	3
A.İ.İ.T	AB	3.7	2
			15

Öğrencinin dönem not ortalamasını katsayı cinsinden hesaplayınız.

ÇÖZÜM

Dersler	Başarı Notları	Katsayılar	Kredi Değerleri	xt
Genel Matematik I	AA	4.0	5	20.0
Türkçe I	AB	3.7	2	7.4
Makro Ekonomi I	CC	2.0	3	6.0
Genel işletme	BC	2.7	3	8.1
A.İ.İ.T	AB	3.7	2	7.4
			15	**48.9**

$$\bar{x}_t = \frac{\sum xt}{\sum t} = \frac{48,9}{15} = 3.26$$

olarak hesaplanır.

Tartılı ortalamalarda tartıları, gözlem değerlerini önem derecesine göre farklı kılan değerler oluşturur. Tartı kavramıyla ilgili olarak aşağıdaki örneği dikkatle gözden geçiriniz.

ÖRNEK 11

Matematik, İstatistik, Fizik, Kimya ve Biyoloji bölümlerinden oluşan bir Fen Fakültesinde, tüm bölümlerin birinci sınıflarına güz döneminde verilen Genel Matematik I dersinin birinci ara sınav sonuçlarına ilişkin bölüm başarı ortalamaları aşağıda verilmiştir:

Bölümler	Bölümlerin Başarı Ortalamaları $\bar{x}$	Bölümlerin Öğrenci Sayıları f
Matematik	70	70
İstatististik	65	60
Fizik	68	50
Kimya	50	40
Biyoloji	50	25

Genel Matematik I dersinin fakülte düzeyindeki başarı ortalamasını bulunuz.

ÇÖZÜM

Bölümler	Başarı Ortalaması $\bar{X}$	Öğrenci Sayısı f	xf
Matematik	70	70	4900
İstatististik	65	60	3900
Fizik	68	50	3400
Kimya	50	40	2000
Biyoloji	50	25	1250
		245	**15450**

$$\bar{x}_t = \frac{\sum xf}{\sum f} = \frac{15450}{245} = 63{,}06$$

olarak hesaplanır.

Oranların ortalaması, ortalamaların ortalaması ve bazı bileşik indeksler, tartılı ortalama kullanılarak hesaplanır.

Uygulamada ortalamaların ortalaması, oranların ortalaması ve bazı bileşik indeksler tartılı ortalama kullanılarak hesaplanır.

Geometrik Ortalama

Geometrik ortalama, seriyi oluşturan gözlem değerlerinin çarpımının gözlem değeri sayısına eşit mertebeden kökü olarak tanımlanır. Eğer seriyi oluşturan gözlem değerleri x_1, x_2, ..., x_n ile ve geometrik ortalama da G ile gösterilirse geometrik ortalama,

$$G = \sqrt[n]{x_1 \,.\, x_2 \,.\, \dots \,.x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

eşitliği ile hesaplanır. Ancak seriyi oluşturan gözlem değerlerinin sayısı arttığında, geometrik ortalamayı yukarıdaki formül yardımıyla hesaplamak güçleşir. Böyle durumlarda geometrik ortalama logaritma yardımıyla aşağıdaki eşitlikle hesaplanır.

Bir serinin geometrik ortalaması, serideki gözlem değerleri çarpımının, gözlem değeri sayısına eşit mertebeden kökü alınarak hesaplanır.

$$\log G = \frac{1}{n}\sum_{i=1}^{n} \log x_i$$

Görüleceği gibi geometrik ortalamanın logaritması, gözlem değerlerinin logaritmalarının aritmetik ortalamasına eşittir.

ÖRNEK 12

Aşağıdaki basit serinin geometrik ortalamasını hesaplayınız.

x
2
5
8
20

ÇÖZÜM

Geometrik ortalamanın tanımı doğrultusunda,

$$G = \sqrt[4]{2.5.8.20} = \sqrt[4]{1600} = 6.32$$

ya da,

$$\log G = \frac{1}{4}\left(\log 2 + \log 5 + \log 8 + \log 20\right)$$

$$= \frac{1}{4}\left(0.30103 + 0.69897 + 0.90309 + 1.30103\right)$$

$$= \frac{1}{4}\left(3.20412\right)$$

$$= 0.8010$$

$$G = 6.32$$

olarak elde edilir.

Frekans serilerinde ve sınıflandırılmış serilerde geometrik ortalama,

$$\log G = \frac{1}{\sum_{i=1}^{n} f_i} \sum_{i=1}^{n} f_i \log x_i$$

eşitliğiyle hesaplanır.

Eğer bir seriyi oluşturan gözlem değerleri, bir önceki gözlem değerine bağlı olarak değişiyor ve değişimin hızı belirlenmek isteniyorsa bu durumda geometrik ortalama hesaplanır.

Uygulamada milli gelir, nüfus, bileşik faiz ve bazı bileşik indekslerin hesaplanmasında geometrik ortalama kullanılır.

Kareli Ortalama

Kareli ortalama, seriyi oluşturan gözlem değerlerinin karelerinin toplamının gözlem sayısına oranının kare kökü olarak tanımlanır.

Kareli ortalama K ile gösterilirse kareli ortalama,

$$K = \sqrt{\frac{x_1^2 + x_1^2 + \ldots + x_n^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$$

eşitliğiyle hesaplanır.

ÖRNEK 13

Aşağıda verilen basit serinin kareli ortalamasını hesaplayınız.

x
1
3
5
7
8
10

Kareli ortalama, seriyi oluşturan gözlem değerlerinin kareleri toplamının, gözlem sayısına oranının kare kökü alınarak hesaplanır.

ÇÖZÜM

x	x^2
1	1
3	9
5	25
7	49
8	64
10	100
	248

olarak elde edilir ve kareli ortalama,

$$K = \sqrt{\frac{x^2}{n}} = \sqrt{\frac{248}{6}} = 6.4291$$

olarak hesaplanır.

Frekans ve sınıflandırılmış serilerde kareli ortalama,

$$K = \sqrt{\frac{\sum x^2 f}{\sum f}}$$

eşitliğiyle hesaplanır.

ÖRNEK 14

Aşağıdaki serinin kareli ortalamasını hesaplayınız.

Sınıflar	f
0 - 4	1
4 - 8	4
8 - 12	8
12 - 16	5
16 - 20	2
	20

ÇÖZÜM

Hesaplamalar aşağıdaki gibidir:

Sınıflar	**f**	**x**	**x^2**	**x^2f**
0 - 4	1	2	4	4
4 - 8	4	6	36	144
8 - 12	8	10	100	800
12 - 16	5	14	196	980
16 - 20	2	18	324	648
	20			**2576**

Kareli ortalama,

$$K = \sqrt{\frac{x^2 f}{\sum f}} = \sqrt{\frac{2576}{20}} = 11.3490$$

olarak elde edilir.

Görüleceği gibi, kareli ortalama da tüm gözlem değerlerinin büyüklüklerinden etkilenen, duyarlı bir ortalamadır.

Duyarlı Olmayan Ortalamalar

Duyarlı olmayan ortalamalar, seriyi oluşturan tüm gözlem değerlerinin büyüklüklerinden etkilenmeyen ortalamalardır.

İzleyen paragraflarda duyarlı olmayan ortalamalardan sadece medyan ve mod ele alınacaktır.

Medyan

Bir istatistik serisinde tam ortaya düşen ve dolayısıyla seriyi iki eşit kısma bölen gözlem değerine medyan denir.

ÖRNEK 15 ***Aşağıdaki basit serinin medyanını hesaplayınız.***

Bir serinin medyanı, ilgili seriyi tam eşit iki kısma bölen gözlem değeridir.

x
10
12
15
17
20

ÇÖZÜM

Verilen seriyi tam ortadan ikiye bölen gözlem değeri 3. gözlem değeri olan 15'dir.

Med = 15

Görüleceği gibi, seride bu değerden küçük ve büyük olmak üzere 2'şer gözlem değeri bulunmaktadır.

Ünitenin başında verilen 9 ailenin aylık gelirler (YTL) serisini tekrar göz önüne alalım.

520, 580, 670, 700, 700, 700, 860, 1000, 1200

Aritmetik ortalama kullanarak aynı apartmanda oturan ailelerin aylık ortalama geliri 770 YTL bulunmuştu. Görüleceği gibi, 3 ailenin aylık geliri aritmetik ortalamadan büyük, 6 ailenin de aylık geliri aritmetik ortalamadan küçüktür.

Bu grubun aylık gelirinin, gelirler büyüklük sırasına konduğunda tam ortadaki ailenin geliri tarafından temsil edilmesi istenebilir. Bu durumda ortalama gelir, medyan kullanılarak hesaplanmalıdır.

Aylık gelirler serisini tam eşit iki kısma bölen gelir, başka bir anlatımla ilgili serinin medyanı 5. gözlem değeri olan 700 YTL'dir. Dikkat edilirse 700 YTL'den küçük 3, büyük de 3 gelir düzeyi vardır.

Eğer söz konusu apartmanda 9 değil de 10 aile ikamet ediyor olsaydı, bu durumda orta aile (5,5. aile) söz konusu olmayacaktır. Böyle durumlarda medyan tam ortaya düşen iki gözlem değerinin aritmetik ortalaması alınarak hesaplanır.

ÖRNEK 16

Aşağıda 8 gözleme ilişkin sonuçlar, gözlem sırasına göre verilmiştir :

2, 7, 3, 8, 7, 3, 4, 10

Gözlem değerlerine ilişkin medyanı hesaplayınız.

ÇÖZÜM

Öncelikle gözlenen değerler büyüklük sırasına konmalıdır (Bir istatistik serisi oluşturulmalıdır).

x
2
3
3
4
7
7
8
10

Görüleceği gibi, verilen seride tam ortaya 4 ve 7 olmak üzere 2 değer düşmektedir. Yukarıdaki açıklamalar doğrultusunda medyan, bu iki gözlem değerinin aritmetik ortalaması olacaktır.

$$\text{Med} = \frac{4+7}{2} = \frac{11}{2} = 5{,}5$$

olarak hesaplanır.

Süreksiz serilerde medyanın hangi sıradaki gözlem değeri olduğu, n serideki gözlem sayısını göstermek üzere, $\frac{n+1}{2}$ ile bulunur.

Yukarıdaki 8 gözlemde oluşan örnekte medyan, $\frac{8+1}{2} = 4.5$ sıradaki gözlem değeridir. 4 ve 5. gözlem değerleri sırasıyla 4 ve 7 olduğundan medyan bu değerlerin aritmetik ortalaması alınarak hesaplanmıştır.

Buna göre 7 gözlem değerinden oluşan bir seride medyan, $\frac{7+1}{2} = 4.$ gözlem değeri, 100 gözlem değerinden oluşan bir seride ise medyan, $\frac{100+1}{2} = 50.5$

gözlem değeri, başka bir anlatımla 50 ve 51. gözlem değerlerinin aritmetik ortalaması alınarak hesaplanacaktır.

Frekans serilerinde de medyanın kaçıncı gözlemin değeri olduğu, $\frac{n+1}{2}$ ile elde edilir. Hangi gözlem değerinin bu sırada yer aldığı, birikimli frekanslar yardımıyla kolaylıkla bulunur. İşlemler aşağıdaki örnekte gösterilmiştir.

ÖRNEK 17 ***Aşağıda verilen frekans serisinin medyanını hesaplayınız.***

x	f
10	2
12	3
15	6
17	5
20	1
	17

ÇÖZÜM

Hangi sıradaki gözlem değerinin medyan değeri olduğunu bulabilmek için, kolaylık açısından öncelikle -den az ya da -den çok serilerinden birisi oluşturulur. Bu örnekte den az serisinden yararlanılmıştır.

x	f	den az
10	2	2
12	3	5
15	6	11
17	5	16
20	1	17
	17	

Toplam gözlem sayısı 17 olduğundan medyan değeri,

$\frac{n+1}{2} = \frac{18}{2} = 9$. sıradaki gözlem değerine eşit olacaktır.den az serisinden 9.sıradaki gözlem değerinin 15 olduğu bir bakışta görülür. Buradan,

Med = 15

sonucuna ulaşılır.

Sınıflandırılmış serilerde de medyan yine birikimli frekanslar yardımıyla hesaplanır. Ancak, sınıflandırılmış serilerde seriyi iki eşit kısma bölen gözlem değeri bir sınıf içinde yer alacaktır. Medyan değerini içinde bulunduran sınıfa medyan sınıfı adı verilir. Medyan sınıfı, frekanslar toplamının yarısını içinde bulunduran sınıftır.

Medyan sınıfı belirlendikten sonra medyan,

l_a : medyan sınıfının alt sınırı,
N : frekanslar toplamı $(\sum f)$,
f_a : medyan sınıfına kadar olan sınıfların frekansları toplamı,
f_m : medyan sınıfının frekansı,
h_m: medyan sınıfının büyüklüğü,
olmak üzere,

$$Med = l_a + \frac{\frac{N}{2} - f_a}{f_m} \cdot h_m$$

eşitliğiyle hesaplanır.

Ancak dikkat etmek gerekir ki elde edilen sonuç, sınıflama nedeniyle yaklaşık olacaktır.

Sınıflandırılmış serilerde medyan sınıfının bulunması ve medyanın hesaplanması aşağıdaki örnek üzerinde gösterilmiştir.

Aşağıdaki serinin medyanını hesaplayınız. **ÖRNEK 18**

Sınıflar	f
10 - 14	3
14 - 18	4
18 - 22	8
22 - 26	6
26 - 30	1
	22

ÇÖZÜM

Öncelikle medyan sınıfını bulabilmek için den az serisi oluşturulur.

Sınıflar	f	-den az
10 - 14	3	3
14 - 18	4	7
(18 - 12)	8	15
12 - 16	6	21
16 - 20	1	22
	22	

Bu tür sınıflandırılmış serilerde değişken sürekli olduğundan, medyan $\frac{N}{2} = \frac{22}{2} = 11.$ gözlem değeri olacaktır. -den az serisinden 11. gözlem değerinin (18 – 22) sınıfında olduğu kolaylıkla görülür. (18 – 22) sınıfı, medyan sınıfıdır.

Medyan sınıfı belirlendikten sonra,

$l_a = 18,$

$\frac{N}{2} = 11,$

$f_a = 7,$

$f_m = 8,$

$h_m = 4,$

değerleri yukarıda verilen eşitlikte yerine konarak,

$$\text{Med} = 18 + \frac{11 - 7}{8} \cdot 4$$
$$= 18 + 2$$
$$= 20$$

olarak elde edilir. (Bulunan değerin medyan sınıfının içinde kaldığına dikkat ediniz.)

Medyanı grafik yardımıyla da hesaplamak mümkündür. Bunun için -den az ya da den çok eğrilerinden birisinin grafiği çizilir. Sonra dik eksende frekanslar toplamının yarısı belirlenir ve bu noktadan yatay eksene bir paralel çizilir. Bu doğrunun birikimli serinin grafiğini kestiği noktanın apsis değeri medyanı belirler.

ÖRNEK 19 ***Aşağıdaki verilen serinin medyanını grafik yardımıyla bulunuz.***

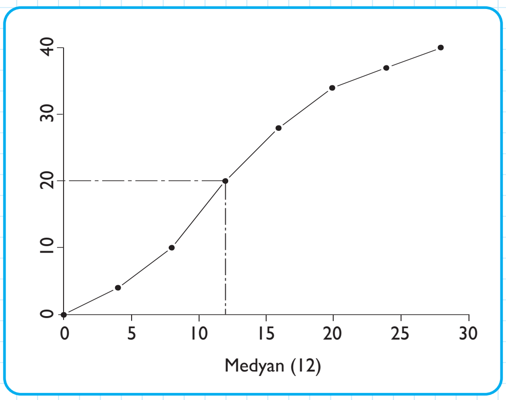
Sınıflar	f
0 - 4	4
4 - 8	6
8 - 12	10
12 - 16	8
16 - 20	6
20 - 24	3
24 - 28	3
	40

ÇÖZÜM

Öncelikle birikimli serilerden birisi, örneğin den az serisi oluşturulur.

Sınıflar	f	-den az
0 - 4	4	4
4 - 8	6	10
8 -12	10	20
12- 16	8	28
16 - 20	6	34
20 - 24	3	37
24 - 28	3	40
	40	

Oluşturulan birikimli serinin grafiği çizilir.



Şekil 3.1 *Medyanın grafik yardımıyla elde edilmesi.*

Frekanslar toplamının yarısı 20 olduğundan dik eksende bu noktadan yatay eksene çizilen paralelin -den az eğrisini kestiği noktanın apsisi belirlenir.

Belirlenen değer 12 olduğundan verilen serinin medyanı,

Med = 12

olarak elde edilir.

Medyan uygulamada, ilgilenilen seride aşırı kıymetlerin varlığı ya da açık (alt ya da üst sınırı belli olmayan) sınıfların bulunması durumunda uygun sonuçlar veren bir merkezi eğilim ölçüsüdür.

Mod

Bir seride en çok tekrarlanan değere mod adı verilir. Tanım uyarınca basit serilerde ve frekans serilerinde mod, en çok tekrarlanan gözlem değerinin belirlenmesi ile kolayca hesaplanır.

Bir serinin modu, seride en çok tekrarlanan değerdir.

Ünitenin başında verilen 9 ailenin aylık gelirler serisini (YTL) tekrar göz önüne alalım.

520, 580, 670, 700, 700, 700, 860, 1000, 1200

Bu gelir grubunda ortalama gelirin en çok tekrarlanan gelir düzeyi tarafından temsil edilmesi istenebilir. Bu durumda 9 aileye ilişkin ortalama gelir, tanım uyarınca mod hesaplanarak elde edilir. En çok tekrarlanan gelir düzeyi 700 YTL olduğundan yukarıdaki seri için,

Mod = 700 YTL

olarak hesaplanır.

Dikkat edilecek olursa, seride mod değerinden küçük 3 ve büyük de 3 gelir düzeyi vardır.

Daha önce de değinildiği gibi, mod ve medyan gibi duyarlı olmayan ortalamalar göz önüne alındığında seride aşırı kıymetlerin oluşu, bu ortalamaların sonucunu etkilemeyecektir. Örneğin ilk gelir düzeyi 100 YTL ya da son gelir düzeyinin 3.000 YTL olması mod ve medyan değerlerini etkilemeyecek ancak duyarlı bir ortalama olan aritmetik ortalamayı doğrudan etkileyecektir. Aşağıdaki örnekleri dikkatle inceleyiniz.

ÖRNEK 20

Aşağıdaki serinin modunu hesaplayınız.

x
10
12
14
17
20

ÇÖZÜM

Bu seride her gözlem değeri yalnız bir kez tekrarlandığından, serinin modu yoktur.

ÖRNEK 21 *Aşağıda verilen serinin modunu hesaplayınız.*

x
10
12
12
12
14
15

ÇÖZÜM

Bu basit seride en çok tekrarlanan gözlem değeri 12 olduğundan,

Mod = 12

olarak hesaplanır.

ÖRNEK 22 *Aşağıdaki frekans serinin modunu hesaplayınız.*

x	f
10	2
13	5
14	8
16	7
19	3
	25

ÇÖZÜM

Verilen frekans serisinde 14 değeri 8 kez gözlenmiştir. En çok tekrarlanan gözlem değeri 14 olduğundan serinin modu,

Mod = 14

olarak kolaylıkla elde edilir.

Eğer modu hesaplanmak istenilen seri sınıflandırılmış bir seriyse, en büyük frekans bir gözlem değerine değil bir sınıfa karşı gelecektir.

En çok tekrarlanan gözlem değerini içinde bulunduran sınıfa mod sınıfı ya da modal sınıf adı verilir.

Mod sınıfı belirlendikten sonra mod,

l_a : mod sınıfının alt sınırı,

Δ_1 : mod sınıfının frekansıyla ondan bir önceki sınıfın frekansları arasındaki mutlak fark,

Δ_2 : mod sınıfının frekansıyla ondan bir sonraki sınıfın frekansları arasındaki mutlak fark,

h : sınıf aralığı,

olmak üzere,

$$\text{Mod} = l_a + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot h$$

eşitliğiyle hesaplanır.

ÖRNEK 23

Aşağıda verilen serinin modunu hesaplayınız.

Sınıflar	f
10 - 14	2
14 - 18	4
(18 - 22)	7
22 - 26	5
26 - 30	3
	21

ÇÖZÜM

Verilen seride, en büyük frekans 7'dir. Bu nedenle mod sınıfı (18 – 22) olacaktır. Mod sınıfı belirlendikten sonra,

$l_a = 18$

$\Delta_1 = 7 - 4 = 3$

$\Delta_2 = 7 - 5 = 2$

$h = 4$

değerleri, yukarıda verilen eşitlikte yerlerine konarak,

$$Mod = 18 + \frac{3}{3+2} \cdot 4$$

$$Mod = 20$$

olarak hesaplanır. (Mod değerinin mod sınıfı içinde kaldığına dikkat ediniz.)

Bazan bir seride aynı maksimum frekansa sahip iki ya da daha çok gözlem değeri ya da sınıf bulunabilir. Böyle durumlarda, ilgili seri frekans serisiyse sınıflandırılarak, sınıflandırılmış seriyse farklı bir sınıf aralığı kullanarak yeniden sınıflandırmak suretiyle modun hesaplanması mümkün olur.

ÖRNEK 24

Aşağıdaki verilen serinin modunu hesaplayınız.

Sınıflar	f
10 - 20	4
20 - 30	7
30 - 40	22
40 - 50	18
50 - 60	22
60 - 70	15
70 - 80	7
80 - 90	5
	100

ÇÖZÜM

Sınıflar	f
10 - 30	11
(30 - 50)	40
50 - 70	37
70 - 90	12
	100

olarak elde edilir. Buradan da,

$$\text{Mod} = 30 + \frac{29}{29 + 3} \,.\, 20$$
$$= 30 + 18.125$$
$$= 48.125$$

olarak hesaplanır.

Sınıflandırılmış serilerde mod, grafik yardımıyla da kolaylıkla bulunabilir. Bunun için önce verilen serinin histogramı çizilir. Histogram üzerinde mod sınıfına ilişkin dikdörtgenin üst köşeleriyle, komşu dikdörtgenlerin üst köşeleri birer doğruyla birleştirilir. Bu doğruların kesişme noktasının apsis değeri, serinin modunu gösterir.

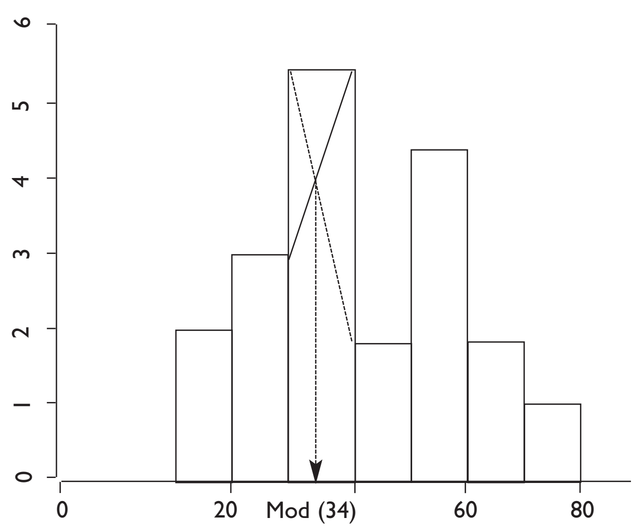
Aşağıdaki örnekte modun grafik yardımıyla bulunması gösterilmiştir.

ÖRNEK 25 ***Aşağıdaki serinin modunu grafik yardımıyla bulunuz.***

Sınıflar	f
10 - 20	10
20 - 30	30
30 - 40	50
40 - 50	20
50 - 60	40
60 - 70	20
70 - 80	10
	180

ÇÖZÜM

Sınıflar	f	f/h
10 - 20	10	1
20 - 30	30	3
30 - 40	50	5
40 - 50	20	2
50 - 60	40	4
60 - 70	20	2
70 - 80	10	1
	180	



Şekil 3.2 *Modun grafik yardımıyla ede edilmesi.*

Verilen serinin modu grafik yardımıyla,

Mod = 34

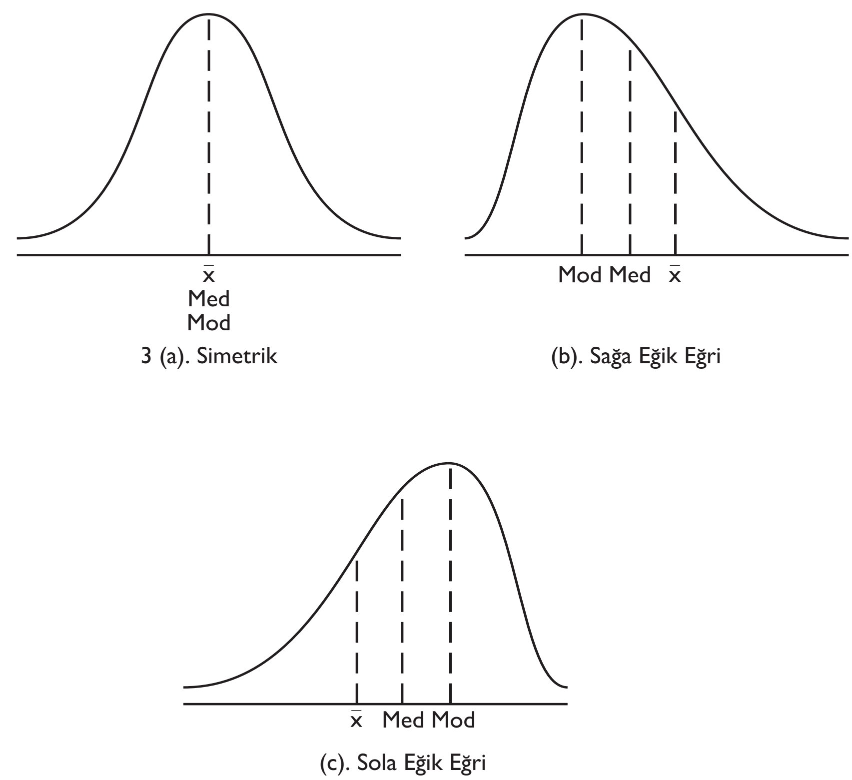
olarak bulunur.

Mod kıymet olarak serideki gözlem değerlerinin büyük bir kısmına uyduğundan, ortalamalar arasında en temsili olanıdır. Ancak, matematiksel işlemlere uygun bir ortalama değildir. Ayrıca U, J ve ters J serileri için de anlamlı bir ortalama değildir.

Mod, U, J ve ters J şeklindeki frekans eğrileri için uygun bir ortalama değildir

Serinin Simetri Durumuna Göre Ortalamalar Arasındaki İlişki

Tek modlu ve eğik serilerde medyan aritmetik ortalama ve mod arasında yer alır. Eğer seri simetrikse, aritmetik ortalama, mod ve medyan birbirine eşit olur. Serinin simetri durumuna göre ortalamalar arasındaki ilişkiler Şekil 3.3.'de gösterilmiştir.



__Şekil 3.3__ Serinin simetri durumuna göre ortalamalar arasındaki ilişkiler.

SIRA SİZDE

1. Bir öğencinin istatistik dersine ilişkin, birinci, ikinci ara sınavlar ve dönem sonu sınavından aldığı notlar aşağıda verilmiştir:

Sınıflar	Puan
1. Ara Sınav	70
II. Ara Sınav	60
Dönem Sonu	50

Ayrıca, başarı notunu birinci ara sınav % 15, ikinci ara sınav % 25 ve dönem sonu sınavı da % 60 oranında etkilemektedir.
Öğrencinin başarı puanını hesaplayınız.

2. Aşağıda verilen seri için uygun ortalamayı hesaplayınız.

Sınıflar	f
5 - 10	3
10 - 15	6
15 - 20	7
20 - 25	5
25 ve daha çok	4
	25

3. U, J, ve ters J eğrileri için modun neden uygun bir ortalama olamayacağını açıklayınız.

DEĞİŞKENLİK ÖLÇÜLERİ

AMAÇ 2

Değişkenlik ölçüleri kavramını açıklayabilecek ve istatistik serilerine ilişkin değişkenlik ölçülerini hesaplayabileceksiniz.

İstatistik serilerinin incelenmesinde ve karşılaştırılmasında ortalama gerekli bir ölçüdür. Ancak tek başına yeterli değildir. Gerçekte ortalamaları eşit olan seriler, biribirinden çok farklı olabilir.

Aşağıdaki ortalamaları aynı olan x ve y serilerini göz önüne alalım :

Bir seriyi oluşturan gözlem değerlerinin değer itibariyle biribirinden ya da herhangi bir ortalamadan uzaklıkları esas alınarak oluşturulan ölçülere, değişkenlik ölçüleri adı verilir.

x	y
30	2
32	14
35	20
36	44
37	90
$\bar{X} = 34$	$\bar{Y} = 34$

Görüleceği gibi x serisinde gözlem değerleri y serisine göre ortalamaya daha yakın konumlanmıştır.

Bu basit örnekten de görülebileceği gibi, bir ortalama değer bir frekans dağılımını karakterize etmede yetersiz kalmaktadır. Bu nedenle bir frekans dağılımının özellikleri araştırılırken, ortalama değerin yanı sıra, gözlem değerlerinin ortalama etrafındaki yayılışına ilişkin ölçülere de ihtiyaç vardır.

Ana çizgileriyle, bir seriyi oluşturan gözlem değerlerinin değer itibariyle biribirinden ya da herhangi bir ortalamadan uzaklıkları, seriyi oluşturan gözlem değerlerinin nasıl yayıldığını, başka bir anlatımla ilgili serinin değişkenliğini ifade eder.

Bu ünitede istatistikte sıkça kullanılan belli başlı değişkenlik ölçüleri ele alınacaktır.

Değişim Aralığı

Değişkenlik ölçülerinin en basiti olan değişim aralığı, bir serideki en büyük değer ile en küçük değer arasındaki fark olarak tanımlanır.

Değişim aralığı kısaca D.A. ile gösterilirse,

$$D.A. = x_{max} - x_{min}$$

olarak ifade edilir.

Değişim aralığı, bir serideki en büyük gözlem değerinden en küçük gözlem değeri çıkartılarak hesaplanır.

Yukarıdaki ortalamaları aynı olan x ve y serilerini göz önüne alınırsa, bu serilere ilişkin değişim aralıkları,

$$D.A.(x) = 37 - 30 = 7$$
$$D.A.(y) = 90 - 2 = 88$$

olarak hesaplanır.

Değişim aralığı, farklı sayıda gözlem değeri içeren ve farklı ölçü birimlerine göre oluşturulmuş serilerin karşılaştırılmalarında kullanılamaz. Bu değişkenlik ölçüsü, uygulamada eşit sayıda küçük örneklemlerin değerlendirildiği alanlarda, örneğin istatistik kalite kontrolünde sıkça kullanılmaktadır.

Standart Sapma

Standart sapma, bir seriyi oluşturan gözlem değerlerinin aritmetik ortalamadan farklarının kareli ortalaması alınarak hesaplanır.

Standart sapma, bir seriyi oluşturan gözlem değerlerinin aritmetik ortalamadan farklarının kareli ortalaması olarak tanımlanır ve σ (sigma) ile gösterilir.

Basit serilerde standart sapma,

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n}}$$

eşitliği yardımıyla hesaplanır.

ÖRNEK 26 ***Aşağıda verilen basit serinin standart sapmasını hesaplayınız.***

x
1
4
5
7
9
10

ÇÖZÜM

x	$x - \bar{x}$	$(x - \bar{x})^2$
1	1 - 6 = -5	25
4	4 - 6 = -2	4
5	5 - 6 = -1	1
7	7 - 6 = 1	1
9	9 - 6 = 3	9
10	10 - 6 = 4	16
		56

$n = 6$
$\bar{x} = 6$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} = \sqrt{\frac{56}{6}} \cong 3.055$$

olarak elde edilir.

Frekans serilerinde ve sınıflandırılmış serilerde standart sapma,

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 f_i}{\sum_{i=1}^{n} f_i}}$$

ile hesaplanır.

Aşağıda verilen frekans serisinin standart sapmasını hesaplayınız. **ÖRNEK 27**

x	**f**
2	1
4	3
5	6
8	4
9	2
	16

ÇÖZÜM

x	**f**	**xf**	$(x - \bar{x})$	$(x - \bar{x})^2$	$(x - \bar{x})^2 \cdot f$
2	1	2	-3.875	15.0156	15.0156
4	3	12	-1.875	3.5156	10.5468
5	6	30	-0.875	0.7656	4.5936
8	4	32	2.125	4.5156	18.0624
9	2	18	3.125	9.7656	19.5312
	16	**94**			**67.7496**

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{94}{16} = 5.875$$

$$\sigma = \sqrt{\frac{67.7496}{16}} = \sqrt{4.23435} \cong 2.0577$$

olarak elde edilir.

ÖRNEK 28 ***Aşağıda verilen sınıflandırılmış serisinin standart sapmasını hesaplayınız.***

Sınıflar	f
0 - 2	2
2 - 4	4
4 - 6	6
6 - 8	5
8 - 10	3
	20

ÇÖZÜM

Sınıflar	x	f	xf	$(x - \bar{x})$	$(x - \bar{x})^2$	$(x - \bar{x})^2 \cdot f$
0 - 2	1	2	2	-4.3	18.49	36.98
2 - 4	3	4	12	-2.3	5.29	21.16
4 - 6	5	6	30	-0.3	0.09	0.54
6 - 8	7	5	35	1.7	2.89	14.45
8 - 10	9	3	27	3.7	13.69	41.07
		20	**106**			**114.20**

$$\bar{x} = \frac{106}{20} = 5.3$$

$$\sigma = \sqrt{\frac{114.2}{20}} = \sqrt{5.71} \cong 2.39$$

olarak elde edilir.

Standart sapma, uygulamada matematiksel işlemlere elverişli olması nedeniyle en çok kullanılan değişkenlik ölçüsüdür.

Bazan σ yerine değişkenlik ölçüsü olarak σ^2 kullanılır. σ^2 'ye varyans adı verilir. (Standart sapma, varyansın pozitif kare köküdür.)

Standart sapmayla ilgili bazı özellikler, aşağıda ispatsız olarak verilmiştir :

- Kareli ortalamanın karesiyle aritmetik ortalamanın karesi arasındaki fark, varyansa eşittir. Başka bir ifadeyle,

$$K^2 - (\bar{x})^2 = \sigma^2$$

'dir.

- Bir seriyi oluşturan gözlem değerlerinin her birine sabit bir sayı eklenir ya da çıkartılırsa, serinin standart sapması değişmez.
- Bir seriyi oluşturan gözlem değerlerinin tümü c gibi bir sayıyla çarpılırsa elde edilen serinin standart sapması, ilk serinin standart sapmasının c katı olur.

Değişim Katsayısı

Buraya kadar ele alınan değişkenlik ölçüleri, mutlak değişkenlik ölçüleridir. Bu nedenle farklı ölçü birimlerine göre oluşturulan serilerin değişkenlikleri, bu ölçülerle karşılaştırılamaz. Ayrıca mutlak değişkenlik ölçüleri, seriyi oluşturan gözlem değerlerinin büyüklüklerinin de etkisi altındadır.

Konuya açıklık kazandırması açısından, aşağıdaki örneği göz önüne alalım.

x	f
10	43
11	48
13	58
14	63
17	73
65	**285**

x serisi için $\bar{x}$ = 13 ve $\sigma_x = 2.7386$
y serisi için $\bar{y}$ = 57 ve $\sigma_y = 11.9373$'tür.

Farklı seriler değişim katsayısı yardımıyla karşılaştırılabilir.

Görüleceği gibi, $\sigma_y > \sigma_x$'dir. Ancak bu sonuç, y serisindeki gözlem değerlerinin x serisine göre daha büyük olmasından kaynaklanmış olabilir.

Eğer, sadece standart sapmalarla bu iki seri karşılaştırılırsa, y serisindeki değişkenliğin x serisine göre daha büyük olduğu ifade edilecektir.

Eğer karşılaştırılan serilerin standart sapmaları ilişkin oldukları serilerin ortalama değerinin bir yüzdesi olarak ifade edilirse, karşılaştırmalarda ölçü birimlerindeki farklılıklar ve gözlem değerlerinin büyüklüğünden oluşan sakıncalar, giderilebilir. Bu yaklaşımla hesaplanan değişkenlik ölçüsüne, değişim katsayısı adı verilir ve kısaca D.K. ile gösterilir.

$$D.K.(x) = \frac{\sigma}{\bar{x}} \cdot 100$$

olarak formüle edilir.

Yukarıda verilen serilere ilişkin değişkenlik, serilerin değişim katsayılarıyla bulunursa,

$$D.K.(x) = \frac{2.7386}{13} \cdot 100 \cong \%21.066$$

$$D.K.(y) = \frac{11.9373}{57} \cdot 100 \cong \%20.9426$$

olarak elde edilir.

Görüleceği gibi, gerçekte X serisindeki değişkenlik, Y serisine göre daha fazladır.

SIRA SİZDE

1. Aşağıda verilen serinin,
  a. Değişim aralığını,
  b. Standart sapmasını,
hesaplayınız.

x	f
12	2
14	6
18	7
20	3
26	2
	20

2. Aşağıda verilen iki seriden hangisinde değişkenliğin daha çok olduğunu belirleyiniz.

x_1 (kg)	f	x_2 (lt)	f
0 - 4	4	0 - 10	7
4 - 8	4	10 - 20	12
8 - 12	8	20 - 30	20
12 - 16	6	30 - 40	11
16 - 20	5	40 - 50	5
20 - 24	3		**55**
	30		

3. Aşağıda verilen serinin varyansını hesaplayınız.

Sınıflar	f
0 - 5	1
5 - 10	4
10 - 15	10
15 - 20	20
20 - 25	15
25 - 30	10
30 - 35	5
	65

Kendimizi Sınayalım

1.

Sınıflar	f
10 - 16	4
16 - 22	3
22 – 28	7
28 – 34	15
34 – 40	8
40 – 46	5
46 - 52	2
	44

Yukarıda verilen serinin aritmetik ortalaması kaçtır?

a. 20.12
b. 27.13
c. 29.42
d. 30.86
e. 32.15

2. 5 birimden oluşan bir basit seride gözlem değerlerinin toplamı $\sum x = 30$ olduğuna göre, serinin aritmetik ortalaması kaçtır?

a. 3
b. 4
c. 5
d. 6
e. 15

3. 15 gözlem değerinden oluşan bir basit serinin aritmetik ortalaması 50 ise bu serideki gözlem değerlerinin toplamı $\sum x$ kaçtır?

a. 60
b. 150
c. 600
d. 750
e. 840

4. Bir öğrencinin Olasılık dersinden birinci, ikinci ara sınav ve final notları aşağıdaki tabloda verilmiştir. (Sonucu birinci ara sınav %10, ikinci ara sınav %20 ve final notu da %70 oranında etkileyecektir.)

Sınıflar	Puan
1. Ara Sınav	75
II. Ara Sınav	50
Final	40

Buna göre bu öğrencinin başarı notu kaçtır?

a. 45.5
b. 50.5
c. 55.0
d. 60.0
e. 65.0

5.

Sınıflar	f
5 - 10	1
10 - 15	4
15 – 20	6
20 –25	10
30 - 35	5
30 - 35	3
35 - 40	1
	30

Yukarıda verilen serinin medyanı kaçtır?

a. 15
b. 18
c. 20
d. 22
e. 27

6. Bir seri için kareli ortalama K = 10 ve aritmetik ortalama $\bar{x} = 6$ olarak hesaplanmıştır. Bu serinin standart sapması kaçtır?

a. 4
b. 8
c. 50
d. 64
e. 72

7. Kareli ortalaması K= 20 ve aritmetik ortalaması $\bar{x} = 12$ olan bir serinin tüm gözlem değerleri 3 ile çarpılarak yeni bir seri oluşturulmuştur. Yeni serinin varyansı kaçtır?

a. 12
b. 16
c. 20
d. 144
e. 256

8.

Sınıflar	f
0 - 4	4
4 - 8	7
8 - 12	12
12 - 16	3
16 - 20	4
20 - 24	12
24 - 28	3
28 - 32	2
	44

Yukarıda verilen serinin modu kaçtır?

a. 47/12
b. 23/4
c. 12
d. 50/3
e. 18

9. 4, 3, 5, 7, 2, 8, 2, 6, 4
serisinin medyanı kaçtır?

a. 2
b. 3
c. 4
d. 5
e. 6

10. Aritmetik ortalaması $\bar{x} = 100$ ve varyansı $\sigma^2 = 144$ olan bir serinin değişim katsayısı yüzde kaçtır?

a. 1.44
b. 12
c. 14.4
d. 28
e. 56

Yanıt Anahtarı

1. d
2. d
3. d
4. a
5. d
6. b
7. d
8. d
9. d
10. b

Yararlanılan Kaynaklar

ÇÖMLEKÇİ, Necla: **Temel İstatistik İlke ve Teknikleri,** 2. Baskı, Bilim Teknik Yayınevi, Eskişehir, 1994.

FOX, William: **Social Statistics Using Micro Case,** Micro Case Corp., Washington, 1992.

GÜRTAN, Kenan: **İstatistik ve Araştırma Metodları,** İstanbul Üniversitesi Yayınları, No 2265, İstanbul, 1977.

NEWBOLD, Paul: (Çeviren: Ümit Şenesen), **İşletme ve İktisat İçin İstatistik,** Literatür Yayınları, İstanbul, 2000.

YATES, D. , MOORE D. , McCABE G. **The Practice of Statistics,** W.H. Freeman, New York, 1999.

4

Olasılık

Çalışma Biçimine İlişkin Olarak:

- *Verilen tanımlar iyice incelenmeli,*
- *Örnek sorular çözülürken dikkatli olunmalı,*
- *Kavramlar birbiriyle karıştırılmamalı,*
- *Alıştırmalarda verilen bilgiler iyi değerlendirilerek istenenlerin neler olduğu net bir biçimde ortaya konmalıdır.*

Amaçlar:

- *Deney ve sonuçlarından hareketle, örneklem uzayını yazabileceksiniz.*
- *Verilen tanımları uygulayarak, olasılıklar hesaplayabileceksiniz.*
- *Olaylardaki aşama sayısına bağlı olarak, toplam sonuç sayısını yazabileceksiniz.*
- *Bileşen ve bileşik olasılıklar arasındaki farkı açıklayabileceksiniz.*
- *Ayrık olay kavramını açıklayabileceksiniz.*
- *Bağımsız ve ayrık olaylar arasındaki farkı yazabileceksiniz.*
- *Tamamlayıcı olaylar kavramını açıklayabileceksiniz.*
- *Aynı anda ortaya çıkan olayların olasılığını hesaplayabileceksiniz.*
- *Olaylardan en az birinin ortaya çıkmasına ilişkin olasılığı hesaplayabileceksiniz.*

İçerik Haritası

GİRİŞ

Olasılık, günlük yaşamımızda sıkça kullandığımız, yararlandığımız bir kavramdır. Örneğin meteoroloji uzmanı sabah haberlerinde o gün % 80 olasılıkla yağmur yağacağını, sağlık uzmanları sigara içenlerin içmeyenlere oranla kansere yakalanma riskinin daha yüksek olacağını, sınavı başarısız geçmiş bir öğrenci o dersten geçme şansının çok az olacağını söyler.

Herhangi bir olayın meydana gelme şansını ölçmeyle ilgilenen olasılık, istatistiğin önemli bir bölümünü oluşturmaktadır. İstatistiğin çıkarsama (öngörü) temelini oluşturan olasılık, belirsizlik durumunda sağlıklı kararlar vermeyi sağladığı için, planlama çalışmalarında yoğun bir biçimde kullanılmaktadır. Örneğin bir firmanın gelecek yıldaki satış kestirimleri, bir kısmı gerçekleşecek bir kısmı gerçekleşmeyecek bir çok varsayıma dayalıdır. Bu nedenlerden dolayı olasılık kuramı, bizlere belirsizlik altında ya da mevcut bilgilerin tam ve sağlıklı olmaması gibi durumlarda doğru ve sağlıklı kararlar verebilmede yardımcı olacaktır. Bu bölümde ilk olarak ilgili temel kavramlar verilecek, daha sonra olasılık hesaplama kurallarından önemli olanları basit bir biçimde gösterilecek ve son olaraksa kesikli verilerin en temel dağılımlarından biri olan Binom dağılımı anlatılacaktır.

DENEY, SONUÇ VE ÖRNEKLEM UZAYI

AMAÇ 1 *Deney ve sonuçlarından hareketle, örneklem uzayını yazabileceksiniz.*

Vida üreten bir firmada kalite kontrol uzmanı olarak görev yapan Ragıp Keskingöz üretim hattından rasgele bir vida alarak vidanın hatalı olup olmadığını inceler. Ragıp Keskingöz'ün bir vidayı inceleme eylemi istatistiksel deneye bir örnektir. Bu inceleme sonucunda vida hatasız ya da hatalı biçiminde değerlendirilecektir. Bu iki gözlem bilgisine deneyin sonucu (outcome) denirken, bu sonuçların birlikte ele alınması neticesindeyse bu deneyin örneklem uzayı oluşur.

Pek çok gözlemden sadece bir tanesinin gerçekleşmesi sürecine "deney", bu gözlemlere "deneyin sonuçları" ve bu sonuçların tümüne ise "deneyin örneklem uzayı" denmektedir.

Pek çok gözlemden sadece bir tanesinin gerçekleşmesi sürecine "deney", bu gözlemlere deneyin sonuçları ve bu sonuçların tümüneyse "deneyin örneklem uzayı" adı verilir.

Bir örneklem (örnek) uzayı **S** harfiyle ifade edilmekte olup yukarıda verilen vida inceleme deneyine ilişkin örneklem uzayı,

S = {hatasız, hatalı}

biçiminde gösterilmektedir. Bu örneklem uzayının elemanlarına da örneklem noktaları denmektedir.

Tablo 4.1 *Deney, sonuç ve örneklem uzayı örnekleri.*

Deney	Sonuçlar	Örneklem Uzayı
Paranın bir kez atılması	Yazı, Tura	S = { Yazı, Tura }
Zarın bir kez atılması	1, 2, 3, 4, 5, 6	S = { 1, 2, 3, 4, 5, 6 }
Paranın iki kez atılması	YY, YT, TY, TT	S = { YY, YT, TY, TT }
Doğacak bebeğin cinsiyeti	Erkek, Kız	S = { Erkek, Kız }
Öğrencinin sınav sonucu	Başarılı, Başarısız	S = { Başarılı, Başarısız }

Bir deneyin örneklem uzayı Venn ya da ağaç diyagramı çizilerek de oluşturulabilmektedir. Venn diyagramı, bir deneyin tüm olası sonuçlarının (kare, dikdörtgen ya da daire gibi) bir resimle gösterilmesidir. Ağaç diyagramındaysa her bir sonuç, ağacın bir dalıyla ifade edilmektedir. Venn ve ağaç diyagramları olasılık kavramlarının, görsel ifade yoluyla kolay anlaşılmasına yardımcı olmaktadır.

ÖRNEK 1 ***Paranın bir kez atılması deneyinin Venn ve ağaç diyagramlarını çiziniz.***

ÇÖZÜM

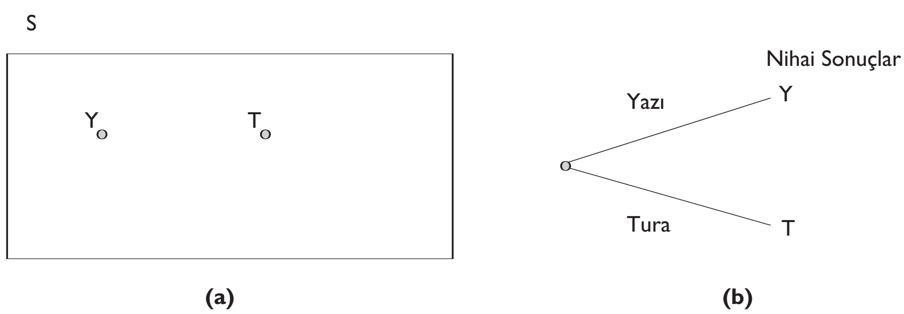
Bu deneyin iki olası sonucu yazı ve tura olup örneklem uzayı Y = Yazı, T = Tura olmak üzere,

S = {Y, T}

biçimindedir.

Bu örneğin Venn diyagramı olarak bir dikdörtgen çizilir ve bu dikdörtgen içerisinde iki sonucu göstermek üzere iki nokta konarak yazı ve tura işaretlenir. Örneklem uzayını belirtmek üzere de dikdörtgenin dışına **S** harfi yazılır (Şekil 4.1 a).

Bu örneğin ağaç diyagramıysa aynı noktadan başlayan iki dal çizilmekte ve dallardan biri yazıyı diğeri turayı ifade etmek üzere dalların sonuna da sonuçlar (Y ve T) yazılmaktadır (Şekil 4.1 b).



***Şekil 4.1** Paranın bir kez atılması deneyinin (a) Venn ve (b) Ağaç diyagramı..*

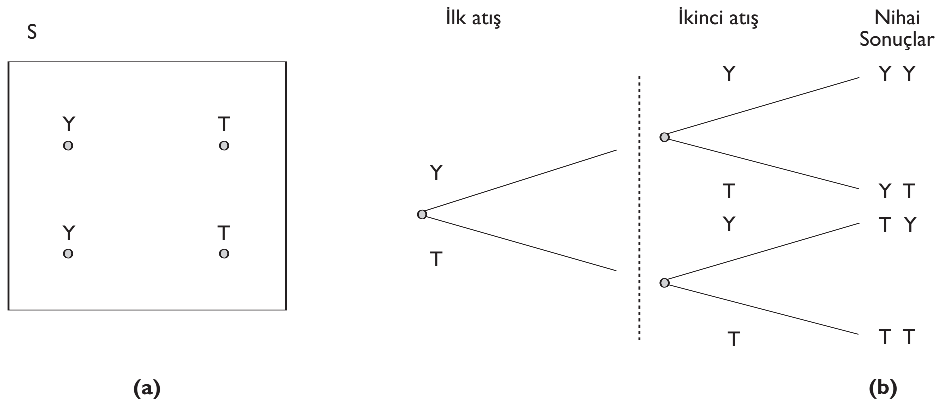
ÖRNEK 2 ***Paranın iki kez atılması deneyinin Venn ve ağaç diyagramlarını çiziniz.***

ÇÖZÜM

Bu deney, paranın ilk ve ikinci atılışında yazı ya da tura gelme durumuna göre iki bölümde şekillenir. ilk para atılışında yazı geldiğinde ikincisinde yazı ya da tura gelebilecektir. Yani yazı geldiğinde YY (birinci ve ikinci atışta yazı), tura geldiğindeyse YT (birinci atışta yazı, ikinci atışta tura) sonucuyla karşılaşılaşılacaktır. Bu durumun tersi de düşünülebilir. İlk atış tura, ikinci atış yazı (TY) gelebilir. Öte yandan ilk atış tura iken, ikinci atışda tura (TT) gelebilir. Sonuç olarak iki kez atılan para deneyinin örneklem uzayı,

S = {YY, YT, TY, TT}

biçimindedir ve bu deneye ilişkin Venn ve ağaç diyagramları da aşağıdaki Şekil 4.2'deki gibidir.



***Şekil 4.2** Paranın iki kez atılması deneyinin (a) Venn ve (b) Ağaç diyagramı.*

ÖRNEK 3

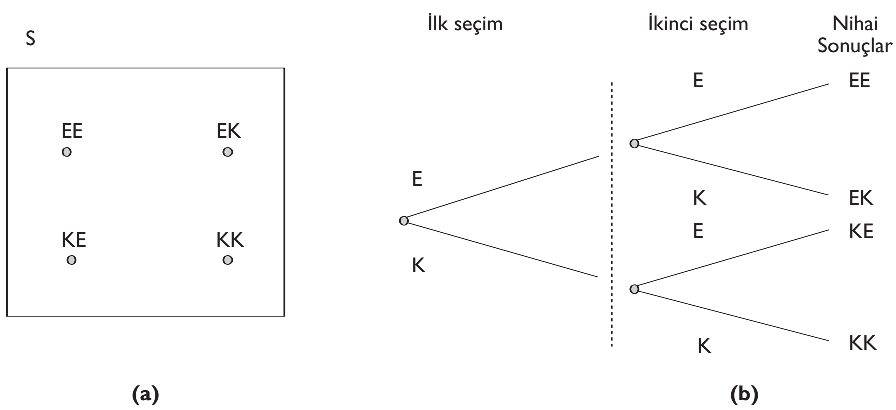
Bir işyerinde çalışan personel arasında rasgele iki tanesinin seçildiği ve cinsiyetlerinin (E = Erkek, K= Kadın) kaydedildiği düşünülsün. Bu deneyin tüm sonuçlarını yazınız, Venn ve ağaç diyagramlarını çiziniz.

ÇÖZÜM

Bu deney de paranın iki kez atılması deneyiyle aynıdır. Çünkü para atma deneyindeki Y ve T biçimindeki iki sonuç, bu deneyde E ve K olarak görülecektir. Aşağıdaki Venn ve ağaç diyagramlarında da (Şekil 4.3) görüleceği gibi bu deneyin sonuçları da EE, EK, KE, KK biçiminde dört noktayla ifade edilmektedir, örneklem uzayı da,

S = {EE, EK, KE, KK}

biçiminde yazılmaktadır.



***Şekil 4.3** İki personel seçilmesi deneyinin (a) Venn ve (b) Ağaç diyagramı.*

Basit ve Bileşik Olaylar

Olay, bir deneyin bir ya da daha çok sonucundan oluşur.

Olay: Bir olay, bir deneyin bir ya da daha çok sonucunun kümesidir.

Olay, basit ya da bileşik olabilmektedir. Basit olaya aynı zamanda "elementer olay" denirken, bileşik olaya da "katışık olay" denir.

Olay: Bir olay, bir deneyin bir ya da daha çok sonucunun kümesidir.

Basit olay:
Bir deneyin sadece ve sadece bir nihai sonucunu içeren olaya basit olay denmekte ve genellikle E_i biçiminde gösterilmektedir.

Basit Olay

Herhangi bir deneyin nihai sonuçlarına basit olay denir. Yani bir basit olay sadece ve sadece bir tane sonuç içermekte ve E_1, E_2, E_3, ... ya da A, B, C, ...biçimindeki harflerle gösterilmektedir.

ÖRNEK 4

Yukarıda verdiğimiz örnek 4.3. deki iki personel seçimi deneyinde, elde edilen (EE EK KE KK) verdiğimiz dört nihai sonucun her biri bu deneyin basit olaylarıdır ve sırasıyla E_1, E_2, E_3 ve E_4 biçiminde gösterilir.

$$E_1 = (EE),\quad E_2 = (EK),\quad E_3 = (KE),\quad E_4 = (KK)$$

Bileşik olay:
Bir deneyin birden çok sonucundan oluşan kümeye bileşik ya da katışık olay denmektedir.

Bileşik Olay

Bir bileşik olay birden çok sonuçtan oluşmaktadır.

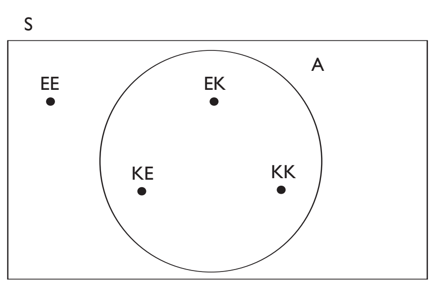
Bileşik olaylar A, B, C, D,.... ya da A_1, A_2, A_3, ..., B_1, B_2, B_3,.... biçiminde gösterilmektedir.

ÖRNEK 5

Bir işyerinde çalışan personel arasından rasgele iki personelin seçilmesi ve cinsiyetlerinin kaydedilmesi biçimindeki, örnek 4.3. tekrar düşünülsün ve A olayı, en çok bir erkeğin seçilmiş olduğu durum olarak tanımlansın. A olayı hiç erkek olmaması ya da bir erkek olması durumunda gerçekleşecektir ve aşağıdaki gibi gösterilecektir.

$$A = \{EK, KE, KK\}$$

A olayı, birden çok sonuçlu olduğu için bir bileşik olaydır. Bu olayın Venn diyagramı yardımıyla grafiksel gösterimi aşağıdadır.



Şekil 4.4 *A Bileşik olayının Venn diyagramı.*

ÖRNEK 6

Bir grup insandan bir kısmı, genetik kopyalamayı, olumlu bulup desteklemekte, geri kalanı karşı çıkmaktadır. Bu gruptan rasgele iki kişi seçilmiş ve genetik kopyalamaya ilişkin görüşleri sorulmuştur. Kaç farklı sonuç söz konusudur? Bu deneye ilişkin Venn ve ağaç diyagramlarını çiziniz. Aşağıda verilen olayların içerdiği sonuçları listeleyiniz ve bu olayların basit mi yoksa bileşik mi olduğunu belirtiniz.

a) Her iki kişi de genetik kopyalamayı destekliyor.
b) En çok bir kişi genetik kopyalamaya karşıdır.
c) Kesinlikle bir kişi genetik kopyalamayı destekliyor.

ÇÖZÜM

D = Genetik kopyalamayı destekliyor, K = Genetik kopyalamaya karşı olmak üzere dört sonuç aşağıdadır.

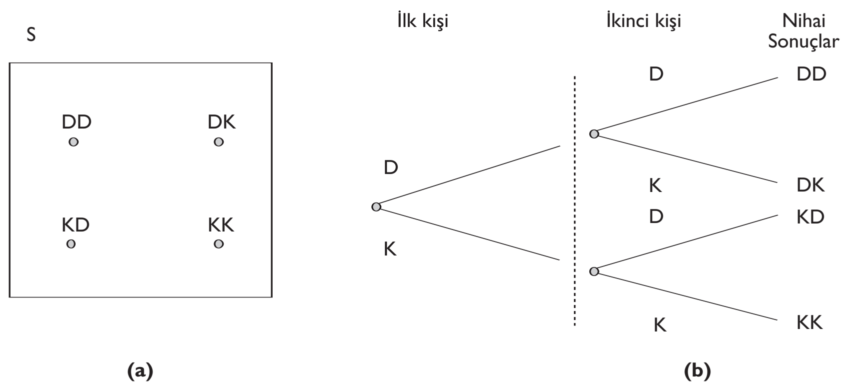
DD = Her iki kişi de genetik kopyalamayı destekliyor.

DK = Birinci kişi genetik kopyalamayı desteklerken ikincisi karşıdır.

KD = Birinci kişi genetik kopyalamaya karşıyken ikincisi destekliyor.

KK = Her iki kişi de genetik kopyalamaya karşıdır.

Bu deneyin dört nihai sonucuna ilişkin Venn ve ağaç diyagramı şöyledir.



***Şekil 4.5** Deneyinin (a) Venn ve (b) Ağaç diyagramı.*

a) "Her iki kişi de genetik kopyalamayı destekliyor" olayının gerçekleşmesi A ile ifade edilmek üzere, A = {DD} biçiminde gösterilir ve bu olay, deneyin dört nihai sonucundan sadece bir tanesini içerdiği için "basit olaydır".

b) "En çok bir kişi genetik kopyalamaya karşı" olayının gerçekleşmesi iki kişinin ya da iki kişiden birinin genetik kopyalamayı desteklemesi durumlarında söz konusudur. B ile ifade edilen olay, B = {DD, DK, KD} sonuçları nedeniyle (birden çok sonuç) "bileşik olaydır".

c) "Kesinlikle bir kişi genetik kopyalamayı destekliyor" olayının gerçekleşmesi, iki kişiden birisinin genetik kopyalamayı desteklerken, diğer kişinin karşı olduğu durumda söz konusudur. C ile ifade edilen ve C = {DK, KD} biçiminde gösterilen olay birden çok sayıda sonuç içermesi nedeniyle "bileşik olaydır".

SIRA SİZDE

1. Deney, sonuç, örneklem uzayı, basit olay, bileşik olay kavramlarını açıklayınız.

2. Aşağıdaki istatistiksel deneyler için basit olayları S örneklem uzayını yazınız.
   a. Bir zarın bir kez atılması,
   b. Bir paranın üç kez atılması,
   c. Bir paranın bir kez ve bir zarın bir kez atılması.

3. Devlet bütçesindeki açığın kapatılabilmesi için zenginlerden alınan verginin arttırılmasını bazı seçmenler isterken, bazıları karşı çıkmaktadır. Seçmenler arasından rasgele üç kişi seçilip görüşleri sorulacak olursa, toplam kaç olası sonuçla karşılaşılır? Bu deneyin ağaç diyagramını çiziniz ve sonuçları S örneklem uzayında gösteriniz.

OLASILIK HESAPLAMA

AMAÇ 2

Verilen tanımları uygulayarak, olasılıklar hesaplayabileceksiniz.

Olasılık: Olasılık, bir olayın meydana gelme şansının sayısal bir ölçüsüdür.

Olasılık (probability) bir olayın meydana gelme, ortaya çıkma şansını ifade eder ve P ile gösterilir. E_i ile gösterilen bir basit olayın olasılığı P (E_i), A bileşik olayının olasılığıysa P (A) biçiminde gösterilmektedir.

Olasılığın İki Özelliği

Olasılığın iki önemli özelliği şunlardır:

1. **Bir olayın olasılığı her zaman sıfır ve bir aralığında yer alır.**
Olay ister basit, isterse bileşik olsun meydana gelme olasılığı hiçbir zaman sıfırdan az, birden çok olamaz. Matematiksel notasyonlarla bu özellik şöyle ifade edilir:

$0 \leq P\,(E_i) \leq 1$

$0 \leq P\,(A) \leq 1$

$$0 \leq P\,(E_i) \leq 1$$

$$0 \leq P\,(A) \leq 1$$

Meydana gelmeyen bir olayın olasılığı sıfır olup, bu tür olaya olanaksız adı verilir. Ortaya çıkma, meydana gelme olasılığı bir olan bir olaya kesin olay adı verilir ve aşağıdaki biçimde gösterilir.

P (M) = 0 ; M olanaksız olay için

P (C) = 1 ; C kesin olay için

2. **Bir deneydeki tüm basit olayların olasılıkları toplamı $\sum P\,(Ei)$ biçiminde gösterilir ve her zaman birdir.**
Bu özellik nedeniyle,

$\sum P\,(E_i) = P\,(E_1) + P\,(E_2) +$
$P\,(E_3) + \ldots\ldots\ldots\ldots\ldots = 1$

$$\sum P\,(E_i) = P\,(E_1) + P\,(E_2) + P\,(E_3) + \ldots\ldots\ldots\ldots\ldots = 1$$

eşitliği yazılabilmektedir. Bu özellikten yararlanarak paranın bir kez atılması deneyi için

P (Y) + P (T) = 1

Paranın iki kez atılması deneyi için

P (Y Y) + P (Y T) + P (T Y) + P (T T) = 1

Süper Ligde oynayan bir futbol takımının maç sonucu içinse

P (Galibiyet) + P (Mağlubiyet) + P (Beraberlik) = 1

eşitlikleri yazılabilir.

Olasılığa Üç Kavramsal Yaklaşım

Olasılığa üç kavramsal yaklaşım: **1)** klasik olasılık, **2)** olasılığın göreli sıklık kavramı ve **3)** öznel olasılık kavramıdır. Olasılığın bu üç kavramının açıklamaları aşağıdadır.

Klasik Olasılık

Sonuçların ortaya çıkma olasılıkları aynı ise buna eşit olasılıklı (benzer) sonuçlar denir. Klasik olasılık kuralı, tüm sonuçları eşit olasılıklı olan deneylerin sonuçlarına ilişkin olasılıkları hesaplamada kullanılmaktadır.

Klasik olasılık kuralına göre bir deneydeki basit bir olayın olasılığı 1'in tüm sonuç sayısına bölünmesine eşittir.

Bu ifadeden de anlaşılmaktadır ki, bir deneyin tüm nihai sonuçlarının olasılıklar toplamı 1'dir ve tüm nihai sonuçlar eşit olasılıklıdır. Öte yandan, A bileşik olayının olasılığıysa, A olayında içerilen sonuç sayısının toplam sonuç sayısına bölünmesiyle elde edilmektedir.

Eşit olasılıklı sonuçlar: İki ya da daha çok sonucun (ya da olayın) ortaya çıkma olasılığı aynıysa bunlara eşit olasılıklı sonuç (ya da olay) denir.

Klasik Olasılık Kuralı

$$P(E_i) = \frac{1}{\text{Deneyin toplam sonuç sayısı}}$$

$$P(A) = \frac{\text{A olayında içerilen sonuç sayısı}}{\text{Deneyin toplam sonuç sayısı}}$$

Klasik olasılık kuralı:

$$P(E_i) = \frac{1}{\text{Deneyin toplam sonuç sayısı}}$$

$$P(A) = \frac{\text{A olayında içerilen sonuç sayısı}}{\text{Deneyin toplam sonuç sayısı}}$$

ÖRNEK 7

Paranın bir kez atılması deneyinde bir yazı ve bir tura elde edilmesi olasılığını bulunuz.

ÇÖZÜM

Bu deneyde yazı ve tura olmak üzere iki sonuç bulunmaktadır ve bu sonuçlar eşit olasılıklıdır. Bu nedenle,

$$P(\text{Yazı}) = \frac{1}{\text{Toplam sonuç sayısı}} = \frac{1}{2} = 0.5$$

$$P(\text{Tura}) = \frac{1}{2} = 0.5$$

sonuçları elde edilir.

ÖRNEK 8

Zarın bir kez atılması deneyinde çift sayı elde edilmesi olasılığını bulunuz.

ÇÖZÜM

Bu deneyde 1, 2, 3, 4, 5 ve 6 olmak üzere altı sonuç bulunmaktadır ve tüm sonuçlar eşit olasılıklı sonuçlardır. A bileşik olayı 2, 4 ve 6 gelmesi biçiminde tanımlanırsa,

$$A = \{2, 4, 6\}$$

örneklem uzayındaki toplam altı sonucun üç tanesi A olayınca içerilmiş olur ve A olayının olma olasılığı,

$$P(A) = \frac{\text{A olayınca içerilen sonuç sayısı}}{\text{Deneyin toplam sonuç sayısı}} = \frac{3}{6} = 0.5$$

olarak bulunur.

ÖRNEK 9 ***Bir derneğin 60'ı erkek ve 40'ı kadın olmak üzere toplam 100 üyesi bulunmaktadır. Bu üyeler arasında bir tanesi dernek başkanı olmak için rasgele seçilecektir. Bir kadın üyenin dernek başkanı seçilme olasılığı nedir?***

ÇÖZÜM Seçim rasgele olacağı için derneğin 100 üyesinin de seçilme olasılığı aynıdır. Yani bu deneyde toplam olarak 100 tane eşit olasılıklı sonuç vardır. Burada istenense 40 kadın üyeden bir tanesinin seçilmesidir. Bu da,

$$P\left(\text{Bir kadının dernek başkanı seçilmesi}\right) = \frac{40}{100} = 0.4$$

biçiminde bulunur.

Olasılığın Göreli Sıklık Kavramı

İlk olarak aşağıdaki olasılıkların hesaplanmak istendiğini düşünülsün.

1. Bir otomobil fabrikasınca bundan sonra üretilecek otomobilin kusurlu olma olasılığı,
2. Rasgele seçilmiş bir ailenin yıllık gelirinin 5.000.000.000 TL'den fazla olması olasılığı,
3. Bir hastanede bundan sonra doğacak çocuğun cinsiyetinin kız olması olasılığı,
4. 80 yaşındaki birinin en az bir yıl daha yaşaması olasılığı,
5. Hileli bir paranın atılması sonucunda yazı gelmesi olasılığı,
6. Civalı bir zarın atılması sonucunda 1 gelmesi olasılığı.

Bu deneylerdeki sonuçlar eşit olasılıklı olmadığı için, yukarıda sıralanan olaylara ilişkin olasılıklar klasik olasılık hesaplama kuralıyla hesaplanamaz. Örneğin fabrikada bundan sonra üretilecek ilk araba kusurlu ya da kusursuz olabilir. Ancak burada kusurlu ya da kusursuz sonuçlarının elde edilmesi olasılıkları eşit değildir.

Yukarıda olduğu gibi, sonuçları eşit olasılıklı olmayan deneylerde, deney defalarca tekrar edilerek veri üretilmektedir. Böylesi durumlarda olasılıkları hesaplamak için ya eski verilerden yararlanılmakta ya da deney çok kez tekrarlanmak suretiyle yeni veri üretilmektedir. Bu verilerden yararlanarak bir olaya ilişkin (yaklaşık) olasılık değeri için göreli sıklıklardan yararlanılmaktadır. Bu yönteme "olasılığın göreli sıklık kavramı" adı verilmektedir. Çünkü deneyin tekrarlanması sonucunda göreli sıklıklar elde edilmekte ve bunlardan yararlanılarak da olasılıklar hesaplanmaktadır. Deneyin her tekrarından sonra göreli sıklıklar değişeceğinden, olasılıklar da değişecektir. Ancak bu olasılıkların değişiminin azaltılması, örneklem hacminin artırılması yoluyla sağlanabilmektedir.

Yaklaşık Olasılık İçin Göreli Sıklık

Eğer bir deney **n** kez tekrarlanmış ve **f** kez bir A olayı gözlenmiş ise olasılığın göreli sıklık kavramına göre olasılık,

$$P(A) = \frac{f}{n}$$

biçiminde bulunur.

Yaklaşık olasılık için göreli sıklık: Eğer bir deney **n** kez tekrarlanmış ve **f** kez bir A olayı gözlenmiş ise olasılığın göreli sıklık kavramına göre olasılık.

$$P(A) = \frac{f}{n}$$

biçiminde bulunur.

ÖRNEK 10

Bir otomobil fabrikasında üretilen otomobillerden rasgele 500 tanesi seçilmiş ve 10 tanesinin kusurlu olduğu görülmüştür. Kusurlu üretim yapmanın da rasgele olduğunu düşünerek, ilk üretilecek otomobilin kusurlu olması olasılığı nedir?

ÇÖZÜM

Örneklemdeki (seçilen) otomobil sayısına n = 500, kusurlu otomobil sayısına f = 10 denecek olursa, göreli sıklık kuralı gereğince olasılık,

$$P\left(\text{İlk üretilecek otomobil}\right) = \frac{f}{n} = \frac{10}{500} = 0.02$$

elde edilir. Bu olasılık 500 otomobilden elde edilen göreli sıklıktan hesaplanmış özel bir değerdir. Aşağıdaki Tablo 4.2'de bu örnek için sıklık ve göreli sıklık dağılımları verilmiştir.

Otomobil	f	Göreli Sıklık
Kusursuz	490	490 / 500 = 0.98
Kusurlu	10	10 / 500 = 0.02
Toplam	500	1.00

Tablo 4.2 *Otomobil örneğinin sıklık ve göreli sıklık dağılımları.*

Bu tablodaki göreli sıklık sütunu yaklaşık olasılıklar sütunu olarak kullanılmaktadır. Bu sütundan,

P (İlk üretilecek otomobil kusurlu) = 0.02

P (İlk üretilecek otomobil kusursuz) = 0.98

değerleri bulunur. Burada unutulmaması gereken, göreli sıklıkların gerçek olasılıklar değil sadece yaklaşık olasılıklar olduğudur. Göreli sıklıklardan elde edilen olasılıkların gerçek olasılıklar olabilmesi için deneyin çok (sonsuz) kez tekrarlanması gerekir ki buna "Büyük Sayılar Yasası" adı verilir.

Büyük Sayılar Yasası

Bir deney çok (sonsuz) kez tekrarlanırsa, bir olayın göreli sıklıkları kuramsal olasılığa yaklaşır.

Büyük Sayılar Yasası: Bir deney çok (sonsuz) kez tekrarlanırsa, bir olayın göreli sıklıkları kuramsal olasılığa yaklaşır.

ÖRNEK 11

Ayşe, Ankara'da rasgele seçilen bir ailenin ev sahibi olma olasılığını belirlemek istemektedir. Bu olasılığı acaba nasıl belirleyecektir?

ÇÖZÜM

Ankara'dan rasgele seçilmiş bir aile için ev sahibi olma ya da olmama gibi iki sonuç bulunmaktadır. Bu iki olay eşit olasılıklı değildir. Çünkü; Ankara'da ikamet edenlerin ne kadarının ev sahibi olduğu bilinmemektedir. Bu nedenle klasik olasılık kuralı uygulanamamaktadır. Böylesi durumlarda aynı deney çok kez tekrarlanarak olasılık değeri (yaklaşık olarak) göreli sıklıklardan hesaplanmaktadır. Ayşe'de bu durumu bildiği için Ankara'dan rasgele 1.000 aileyi seçerek bunlardan 670 tanesinin ev sahibi, 330 tanesinin ise ev sahibi olmadığını belirledi. Bu sonuçlar ışığındaki gibi,

n = örneklem hacmi = 1.000

f = ev sahibi olanların sayısı = 670

olmak üzere olasılık değerleri,

$$P\left(\text{Rastgele seçilen ailenin ev sahibi olması}\right) = \frac{f}{n} = \frac{670}{1,000} = 0.670$$

olarak buldu.

Öznel Olasılık Kavramı

Çoğu kez, ne sonuçları eşit olasılıklı, ne de veri üretmek için tekrarlanabilen deneylerle karşılaşabiliriz. Böylesi durumlarda olayların olma olasılıkları, klasik olasılık kuralı ya da göreli sıklık kavramı kullanılarak hesaplanamamaktadır. Örneğin, İstatistiğe Giriş dersini alan Ahmet'in dönem sonunda o dersten A alarak geçme (başarılı olma) olasılığı nedir? Sorusuna cevap vermek gerçekten güçtür. Çünkü Ahmet bu dersten geçebilmek için test sınavına (sınavlarına) bir kez girecek ve sınavdaki başarı durumuna göre A notu alacak ya da alamayacaktır. Bu olay için söz konusu olan A notu alma ya da almama gibi iki sonuç bulunmakla birlikte, bu sonuçların ortaya çıkması eşit olasılıklı değildir. Bu gibi durumlarda düşünülen (öngörülen) olasılığa öznel olasılık denmektedir. Bu olasılık bireyin değer yargısına, deneyimine, düşüncesine göre değişmektedir. Gerçekten de Ahmet bu dersten A notu alma olasılığını yüksek görürken, dersin hocası daha düşük görebilir.

Öznel olasılık: Bir olay için öznel değer yargısına, deneyim, bilgi ve düşünceye dayalı olasılıktır.

Öznel olasılık keyfi bir değer olup, öngörüde bulunan kişinin deneyiminden, yanlılığından ve beğenisinden etkilenir.

SIRA SİZDE

1. **Üç olasılık yaklaşımını kısaca açıklayınız ve bu üç yaklaşım için birer örnek veriniz.**

2. **Aşağıdakilerden hangilerinin olaylara ilişkin olasılıklar olamayacağını nedenleri ile birlikte söyleyiniz.**
   1 / 5 0.97 -3.5 1.56 5 / 3 0.0 -2 / 7 1.0

3. **Çoktan seçmeli bir test sınavında sorular için beş seçenek bulunmaktadır. Herhangi bir sorunun cevabı rasgele işaretlenecek olursa; cevabın**
   a) Doğru olma olasılığını,
   b) Yanlış olma olasılığını bulunuz.

SAYMA KURALI

AMAÇ 3 ***Olaylardaki aşama sayısına bağlı olarak, toplam sonuç sayısını yazabileceksiniz.***

Bu bölümde şu ana kadar karşılaşılan deneylerde az sayıda sonuç olduğu için, sonuçların listelenmesinde herhangi bir sorunla karşılaşılmadı. Ancak, sonuç sayısının çok olması durumunda sonuçları listelemek çok da kolay olmamaktadır. Böylesi durumlarda toplam sonuç sayısını bulmak için sayma kuralından yararlanılmaktadır.

Sayma kuralı: Eğer bir deneyde; ilk aşamada m tane, ikinci aşamada n tane ve üçüncü aşamada k tane sonuç olmak üzere üç aşama bulunuyorsa, bu deneyin toplam sonuç sayısı = m . n . k 'dır.

Sayma Kuralı
Eğer bir deneyde; ilk aşamada m tane, ikinci aşamada n tane ve üçüncü aşamada k tane sonuç olmak üzere üç aşama bulunuyorsa, bu deneyin toplam sonuç sayısı = m . n . k 'dır.

Sayma kuralı, deneydeki aşama sayısının üçten az ya da çok olması durumunda da kullanılabilmektedir.

ÖRNEK 12

Bir paranın üç kez atılması deneyini düşünelim. Bu deneyde üç aşama bulunmaktadır; ilk kez atış, ikinci kez atış ve üçüncü kez atış. Her aşamanın da yazı ve tura olarak iki sonucu olacaktır. Bu nedenle sayma kuralına göre toplam sonuç sayısı = 2 . 2 . 2 = 8 olacaktır. Deney sonuçları açık olarak; YYY, YYT, YTY, YTT, TYY, TYT, TTY, TTT biçimde olacaktır.

ÖRNEK 13

Bir otomobil satıcısı, kredili otomobil satışlarında, sabit ve değişen faiz oranlarıyla 36, 48 ve 60 ay vade uygulamaktadır. Otomobil satıcısının kaç farklı satış yapması söz konusudur ?

ÇÖZÜM

Bu deneyde iki aşama bulunmaktadır. Aşamalardan ilki sabit ve değişen faiz oranlı (iki sonuçlu) faiz uygulama aşaması, ikincisiyse 36, 48 ve 60 aylık (üç sonuçlu) vade süresi uygulama aşamasıdır. Bu durumda,

Toplam sonuç sayısı = 2 . 3 = 6 ' dır.

ÖRNEK 14

Süper ligdeki bir takımın sezon boyunca 16 maç oynayacağı ve her maçın da galibiyet, malubiyet ve beraberlik biçiminde üç sonucu bulunduğu bilinmektedir. Bu durumda bir takım için sezon boyunca,

Toplam sonuç = 3 . 3 . 3 . 3 . 3 . 3 . 3 . 3 . 3 . 3 . 3 . 3 . 3 . 3 . 3 . 3 = 3^{16} = 43.046.721

farklı sonuç söz konusudur.

SIRA SİZDE

1. **Kusursuz bir zar 2 kez atılmıştır. Toplam sonuç sayısını bulunuz.**
2. **Kusursuz bir para 5 kez atılmıştır. Toplam sonuç sayısını bulunuz.**
3. **Kusursuz bir para n kez atılmıştır. Toplam sonuç sayısını bulunuz.**

BİLEŞEN (MARJİNAL) VE KOŞULLU OLASILIKLAR

AMAÇ 4 ***Bileşen ve bileşik olasılıklar arasındaki farkı açıklayabileceksiniz.***

Bir firmada çalışan 100 kişiye, üst düzey yöneticilere çok yüksek ücretler ödenmesini onaylayıp onaylamadıkları sorulmuş ve aşağıdaki tabloda verilen sonuçlar elde edilmiştir:

Tablo 4.3
Çalışanların cevaplarına ilişkin iki yönlü sınıflama.

Cinsiyet	Onaylıyor	Onaylamıyor
Erkek	15	45
Kadın	4	36

Yukarıdaki tabloda 100 çalışan; cinsiyet (erkek ya da kadın) ve görüş (onaylıyor ya da onaylamıyor) özelliklerine (değişken – karakteristik) göre sınıflanmıştır. Bu sınıflamaya (dağılım) çapraz tablo (contingency table) adı verilmekte olup, sayıların bulunduğu kutulara da göze ya da hücre (cell) adı verilmektedir.

Dikkat edilecek olursa, tabloda dört gözede iki karakteristiğe ait sıklıklar bulunmaktadır. Örneğin; bu gözelerin ilkinde bulunan 15 çalışan, erkek ve yüksek ücret verilmesini onaylayanlar, olmak üzere iki karakteristiği ifade etmektedir.

Yukarıdaki tabloya satır ve sütun toplamlarının eklenmesi sonucunda tablo, aşağıdaki biçime dönüşmektedir.

Tablo 4.4
Çalışanların iki yönlü sınıflaması.

Cinsiyet	Onaylıyor	Onaylamıyor	Toplam
Erkek	15	45	60
Kadın	4	36	40
Toplam	19	81	100

Çalışanlar arasından rasgele bir çalışan seçildiğinde, bu çalışan sadece cinsiyet ya da görüş karakteristiklerinden birine göre de sınıflanabilir. Eğer tek karakteristik dikkate alınacak olursa; seçilen çalışan, erkek olabilir, kadın olabilir, onaylıyor olabilir ya da onaylamıyor olabilir. İşte bu dört karakteristik ya da olayın olasılıklarına bileşen (marjinal) olasılık adı verilmektedir. Bu olasılıklara bileşen ya da basit olasılıklar denmesinin nedeni, bu olasılıkların satır ya da sütun toplamlarının genel toplama bölünmesiyle bulunmasıdır.

Bileşen olasılık:
Basit olasılık olarak da bilinen bileşen olasılık, herhangi başka olay dikkate alınmaksızın sadece bir olaya ilişkin olasılıktır.

Bileşen Olasılık: Basit olasılık olarak da bilinen bileşen olasılık, herhangi başka olay dikkate alınmaksızın, sadece bir olaya ilişkin olasılıktır.

Yukarıdaki Tablo 4.4'e ilişkin dört bileşen olasılık şöyle hesaplanır:

$$P\left(\text{Erkek}\right) = \frac{\text{Erkeklerin sayısı}}{\text{Tüm çalışanların sayısı}} = \frac{60}{100} = 0{,}60$$

Yukarıda belirtildiği gibi, erkek çalışanlara ilişkin (satır) toplam değerlerinin genel toplama bölünmesiyle elde edilen bu değer gibi, öteki üç bileşen olasılık da kolaylıkla bulunur.

$$P\left(\text{Kadın}\right) = \frac{40}{100} = 0.40$$

$$P\left(\text{Onaylıyor}\right) = \frac{19}{100} = 0.19$$

$$P\left(\text{Onaylamıyor}\right) = \frac{81}{100} = 0.81$$

Elde edilen bileşen olasılıkların da eklenmesiyle aşağıdaki tablo oluşturulur:

Cinsiyet	Onaylıyor (A)	Onaylamıyor (B)	Toplam	
Erkek (E)	15	45	60	P (E) = 0.60
Kadın (K)	4	36	40	P (K) = 0.40
Toplam	19	81	100	
	P (A) = 0.19	P (B) = 0.81		

***Tablo 4.5** Bileşen olasılıklı iki yönlü sınıflama.*

Bu durumda, seçilen kişinin, üst düzey yöneticilere yüksek ücret verilmesini onaylıyor olması olasılığı nedir? Bu olasılık şöyle ifade edilebilir:

"Verildiğinde" biçiminde okunur

P (Onaylıyor | Erkek)

Olasılığı hesaplanacak olay

Gerçekleşmiş olay

P (Onaylıyor | Erkek) biçiminde ifade edilen olasılığa, "onaylamanın koşullu olasılığı" denmektedir ve bu gösterim "çalışanın erkek olduğu bilindiğinde (verildiğinde) seçilen çalışanın onaylama olasılığı" olarak okunmaktadır.

Koşullu Olasılık: Koşullu olasılık bir olayın oluştuğunun bilinmesi durumunda diğer olayın olma olasılığıdır. Örneğin A ve B iki olay olmak üzere A olayının koşullu olasılığı,

P (A | B)

biçiminde gösterilir ve B olayı olduğunda A olayının olması olasılığı biçimde okunur.

Koşullu olasılık: Koşullu olasılık bir olayın oluştuğunun bilinmesi durumunda diğer olayın olma olasılığıdır. Örneğin A ve B iki olay olmak üzere A olayının koşullu olasılığı,

P (A | B)

biçiminde gösterilir ve B olayı olduğunda A olayının olması olasılığı biçimde okunur.

ÖRNEK 15

Yukarıda Tablo 4.4'te verilmiş olan 100 çalışana ilişkin sonuçlardan P (Onaylıyor | Erkek) koşullu olasılığını bulunuz.

ÇÖZÜM

P (Onaylıyor | Erkek) koşullu olasılığında rasgele seçilen bir çalışanın erkek olduğu biliniyor ve bu kişinin onaylama olasılığının bulunması isteniyor. Burada ilk olarak Tablo 4.4'ün birinci satırı ele alınmış, seçilmiş olan bu kişinin de, 60 tane erkek çalışandan biri olduğu düşünülerek, bu satır tekrar yazılmıştır.

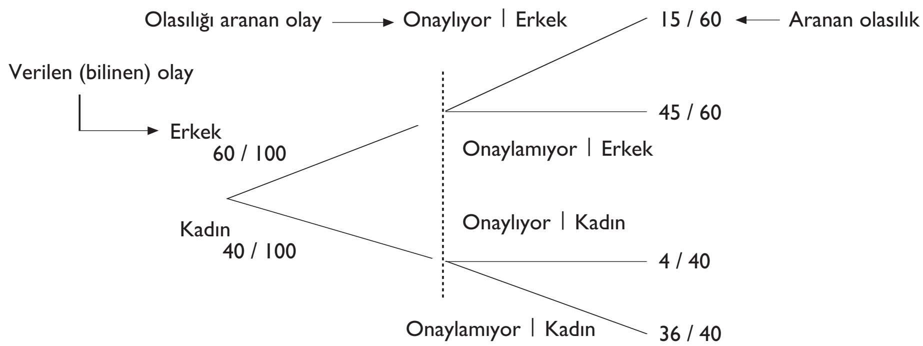
Cinsiyet	Onaylıyor	Onaylamıyor	Toplam
Erkek	15	45	60
	Onaylayan erkekler		*Çalışan toplam erkek*

Bu bilgiler ışığında aranan koşullu olasılık;

$$P\left(\text{Onaylıyor} \mid \text{Erkek}\right) = \frac{\text{Onaylayan erkek sayısı}}{\text{Toplam erkek sayısı}} = \frac{15}{60} = 0.25$$

olarak bulunur. Bu hesaplamadan da görüleceği gibi koşullu olasılık hesaplanırken gerçekleşen olayın değeri (erkek çalışanlar) paydaya, olasılığı bulunmak istenen olayın değeri (onaylayan erkek) paya yazılmaktadır.

Aşağıdaki ağaç diyagramı, koşullu olasılığın daha kolay anlaşılmasını sağlamak amacıyla Örnek 4.15 için hazırlanmıştır:



Şekil 4.6 *Ağaç diyagramı.*

ÖRNEK 16 ***Tablo 4.4'te verilen sonuçlardan yararlanarak, rasgele seçilen kişinin üst düzey yöneticilere yüksek ücret verilmesini onayladığı bilindiğine göre, bu kişinin kadın çalışan, olma olasılığını bulunuz.***

ÇÖZÜM

Burada aranan olasılık,

P (Kadın | Onaylıyor) = ?

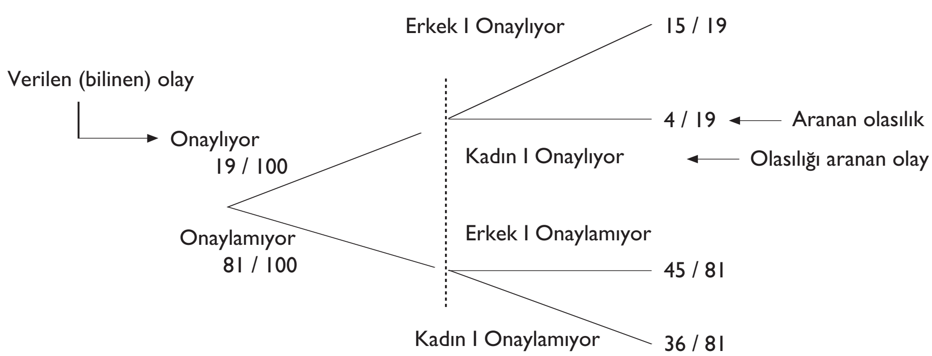
biçimindedir. Yukarıdaki örneğe benzer olarak, Tablo 4.4'te ilk sütun olan onaylayanlar sütunu dikkate alınmaktadır.

	Onaylıyor	
Erkek	15	
Kadın	4	← *Onaylayan kadınların sayısı*
Toplam	19	← *Toplam onaylayanların sayısı*

Bu bilgiler ışığında aranan olasılık değeri,

$$P\left(\text{Kadın} \mid \text{Onaylıyor}\right) = \frac{\text{Onaylayan kadın sayısı}}{\text{Toplam onaylayanların sayısı}} = \frac{4}{19} = 0.211$$

olup örneğe ilişkin ağaç diyagramı aşağıdadır.



Şekil 4.7 *Ağaç diyagramı.*

SIRA SİZDE

1. **Olayların bileşen ve koşullu olasılıkları arasındaki farkı kısaca açıklayınız ve birer örnek veriniz.**

2. **Bir firmada çalışan 420 kişiye sigara içip içmedikleri ve üniversite mezunu olup olmadıkları sorularak aşağıdaki iki yönlü tablo oluşturulmuştur.**

Sigara	Üniversite mezunu	Üniversite mezunu değil
İçiyor	35	80
İçmiyor	130	175

Bu kişilerden rasgele bir tanesi seçildiğinde ;
a. Üniversite mezunu olma,
b. Sigara içmiyor olma,
c. Sigara içtiği bilindiğine göre üniversite mezunu olmama,
d. Üniversite mezunu olduğu bilindiğine göre sigara içmiyor olma,

olasılıklarını bulunuz.

3. **Bir araştırmada 2000 erişkine kürtaja karşı olup olmadıkları sorulmuş ve elde edilen sonuçlar, cinsiyetlere göre aşağıda verilmiştir.**

Cinsiyet	Karşı değil	Karşı
Erkek	495	405
Kadın	620	480

Bu gruptan rasgele seçilen bir kişinin;
a. Kürtaja karşı olmaması,
b. Kürtaja karşı olması,
c. Kürtaja karşı olmadığı bilindiğine göre kişinin kadın olması,
d. Kişinin erkek olduğu bilindiğine göre kürtaja karşı olma olasılıklarını bulunuz.

AYRIK OLAYLAR

AMAÇ 5 *Ayrık olay kavramını açıklayabileceksiniz.*

Birlikte olamayan olaylara karşılıklı (ya da tamamıyla) ayrık olaylar denmektedir ve bu tür olayların ortak sonuçları bulunmamaktadır. Eğer iki ya da daha çok olay karşılıklı ayrık ise, deneyin her tekrarında bu olaylardan en çok bir tanesi ortaya çıkmaktadır ve bu nedenle bir olayın ortaya çıkması, diğer olay ya da olayların ortaya çıkmasını dışta tutmaktadır.

(Karşılıklı-Tamamıyla) Ayrık Olaylar: Birlikte ortaya çıkmayan olaylara, (karşılıklı) ayrık olaylar denmektedir.

(Karşılıklı - Tamamıyla) Ayrık olaylar: Birlikte ortaya çıkmayan olaylara, (karşılıklı) ayrık olaylar denmektedir.

Herhangi bir deney için, deneyin herhangi bir tekrarında ortaya çıkacak sadece bir sonuç olduğu için nihai sonuçlar her zaman (karşılıklı) ayrıktır. Örneğin, bir paranın iki kez atıldığı bir deneyde, YY, YT, TY ve TT biçiminde dört sonuç vardır ve bu sonuçlar (karşılıklı) ayrıktır. Çünkü bir para iki kez atıldığında, bu sonuçlardan sadece bir tanesi ortaya çıkacaktır.

ÖRNEK 17

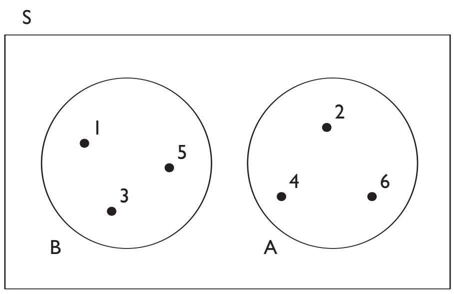
Zarın bir kez atılması deneyine ilişkin olarak aşağıdaki olaylar düşünülecek olursa;

A = çift sayı elde edilmesi = {2, 4, 6}
B = tek sayı elde edilmesi = {1, 3, 5}
C = 5'den küçük sayı elde edilmesi = {1, 2, 3, 4}

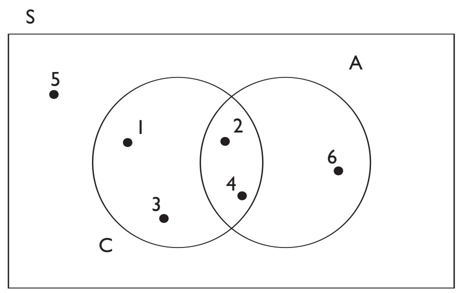
bu olaylara ilişkin olarak,
a) A ve B olayları ayrık mıdır?
b) A ve C olayları ayrık mıdır?

ÇÖZÜM

Aşağıdaki Şekil 4.8 ve 4.9'dan da görüleceği gibi, A ve B olaylarının hiçbir ortak noktası yoktur. Zarın bir kez atılması durumunda A ve B olaylarından sadece bir tanesine ilişkin bir sonuç ortaya çıkacaktır. Bu nedenle A ve B olayları ayrık olaylardır. Öte yandan, A ve C olaylarının ortak noktasıda bulunmaktadır. Başka bir anlatımla zarın bir kez atılması durumunda 2 ya da 4 sayılarının da gelebileceği, bu iki noktanın, her iki olayın da elemanı olması, bir gerçektir. Bu yönüyle A ve C olayları ayrık olmayan olaylardır.



Şekil 4.8 *Ayrık A ve B olayları.*



Şekil 4.9 *Ayrık olmayan A ve C olayları.*

ÖRNEK 18

Büyük bir firmada çalışanlardan rasgele bir kişi seçilmiş ve aşağıdaki iki olay tanımlanmış olsun.

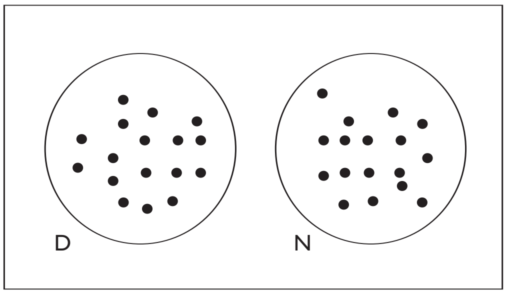
D = seçilen kişi üniversite mezunudur
N = seçilen kişi üniversite mezunu değildir

Tanımlanan D ve N olayları ayrık mıdır?

ÇÖZÜM

Yukarıda tanımlanan D olayı üniversite mezunu çalışanları, N olayıysa üniversite mezunu olmayan çalışanları göstermektedir. Bu olaylar aşağıdaki şekilde gösterilmiştir.

S



Şekil 4.10 Ayrık D ve N olayları.

D ve N olaylarının tanımından ve şekilden de anlaşılacağı gibi bu iki olayda farklı çalışanlar ifade edilmektedir. İki olayın ortak sonucu bulunmadığı için D ve N olayları ayrık olaylardır.

SIRA SİZDE

1. **Bir çift zar atıldığında; A olayı zarlardan birinin 4, B olayıysa zarlar toplamının 11 olma olasılığı olsun. A ve B olaylarının ayrık olaylar olduğunu gösteriniz.**
2. **Bir para 3 kez atılmıştır. En az iki tura gelme olayı A, B ise üç atışta da aynı yüzün gelme olaylarıyken, A ve B olaylarının ayrık olaylar olup olmadığını araştırınız.**
3. **İki farklı para aynı anda atılmıştır. A olayı, iki paranın da yazı gelme, B olayıysa ikinci paranın tura gelme olayı olsun. A ve B olaylarının ayrık olmadığını gösteriniz.**

BAĞIMSIZ VE BAĞIMLI OLAYLAR

AMAÇ 6 *Bağımsız ve ayrık olaylar arasındaki farkı yazabileceksiniz.*

Bağımsız olaylar söz konusu olduğunda, bir olayın ortaya çıkması öteki olayın ortaya çıkma olasılığını değiştirmemektedir.

Bağımsız Olaylar: Eğer bir olayın ortaya çıkması öteki olayın ortaya çıkma olasılığını etkilemiyorsa, bu iki olaya bağımsız olaylar denir. Başka bir anlatımla, A ve B olaylarının bağımsız olabilmeleri için,

$$P(A \mid B) = P(A) \text{ veya } P(B \mid A) = P(B)$$

koşulu sağlanmalıdır.

Bağımsız olaylar: Eğer bir olayın ortaya çıkması öteki olayın ortaya çıkma olasılığını etkilemiyorsa, bu iki olaya bağımsız olaylar denir. A ve B olaylarının bağımsız olabilmeleri için,

$P(A \mid B) = P(A)$ veya $P(B \mid A) = P(B)$

koşulu sağlanmalıdır.

Bu koşullardan biri doğruysa ötekinin de doğru olduğu ya da koşullardan biri yanlışsa ötekinin de yanlış olduğu kolaylıkla gösterilebilir. Sonuç olarak bir olayın ortaya çıkması, bir başka olayın ortaya çıkma olasılığını etkiliyorsa, bu olaylara bağımsız olmayan ya da bağımlı olaylar denmekte ve yukarıdaki koşulların ikisi de sağlamamaktadır.

$$P(A \mid B) \neq P(A), \quad P(B \mid A) \neq P(B)$$

ÖRNEK 19

Tablo 4.4'te verilmiş olan 100 çalışana ilişkin örnekte; kadın (K) ve onaylıyor (O) olayları tanımlanmış olsun. Bu olaylar bağımsız mıdır?

ÇÖZÜM

Eğer K ve O olayları için,

P (K) = P (K | O)

koşulu sağlanıyorsa bu olaylar bağımsız, sağlanmıyorsa bağımlı olaylardır.

Tablo 4.4'te verilenlerden yararlanarak,

P (K) = 40 / 100 = 0.40 ve P (K | O) = 4 / 19 = 0.211

olasılıkları hesaplanabilmektedir. Bu iki olasılık aynı olmadığı için bu olaylar bağımsız olaylar değildir. Bu olayların bağımlı olmalarının nedeniyse üst düzey yöneticilere yüksek ücret verilmesini onaylayan ve onaylamayan erkek çalışanların yüzdelerinin, kadın çalışanlardan farklı olmasıdır. Bu örneğe ilişkin olarak P (O) ve P (O | K) olasılıklarının da aynı olmadığı kolaylıkla görülebilir.

ÖRNEK 20

Bir kutuda, I. Makinede üretilmiş 60, II. Makinede üretilmiş 40 olmak üzere toplam 100 kaset bulunmaktadır. I. Makinede üretilen 60 kasetin 9 tanesi, tüm kasetlerinde 15 tanesi bozuktur. Bu durumda rasgele seçilen bir kasetin bozuk olması D ve bu kasetin I. Makinede üretilmiş olması A olayını göstermektedir. A ve D olayları bağımsız mıdır?

ÇÖZÜM

Soruda verilen bilgilerden,

P (D) = 15 | 100 = 0.15 ve P (D | A) = 9 / 60 = 0.15

olasılıkları bulunur. P (D) = P (D | A) olduğu için, A ve D olayları bağımsız olaylardır.

Bu örnekte, I. ve II. Makinede üretilen bozuk kaset yüzdesi aynı olduğu için, (9 / 60 = 6 / 40) her iki makinenin de bozuk kaset üretme olasılığı 0.15 olduğu için, bu olaylar bağımsız olaylardır.

Gerçekten de soruda verilen bilgiler kullanılarak aşağıdaki tablo oluşturulabilir.

Tablo 4.6 *İki yönlü sınıflama tablosu.*

Makine	Bozuk (D)	Sağlam (S)	Toplam
Makine I (A)	9	51	60
Makine II (B)	6	34	40
Toplam	15	85	100

Bu tablodan yararlanarak da yukarıdaki olasılıklar kolaylıkla bulunabilmekte ve bu olasılıklardan da A ve D olaylarının bağımsız olaylar oldukları görülmektedir.

P (D) = 15 / 100 = 0.15 ve P (D | A) = 9 / 60 = 0.15

SIRA SİZDE

1. **Kusursuz bir para 3 kez atılmıştır. X, Y ve Z olayları sırasıyla ilk iki atışın yazı gelme, üçüncü atışın tura gelme ve üç atışta iki tura gelme olayları iken,**
   a. X ve Y olaylarının,
   b. Y ve Z olaylarının,
   bağımsız olaylar olup olmadığını araştırınız.

2. **Bir firmada çalışan 420 kişiye sigara içip içmedikleri ve üniversite mezunu olup olmadıkları sorularak aşağıdaki iki yönlü tablo oluşturulmuştur.**

Sigara	Üniversite mezunu	Üniversite mezunu değil
İçiyor	35	80
İçmiyor	130	175

"Sigara içme" ve "Üniversite mezunu olmama" olayları bağımsız olaylar mıdır?

3. **Bir araştırmada 2000 erişkine kürtaja karşı olup olmadıkları sorulmuş ve elde edilen sonuçlar, cinsiyetlere göre aşağıda verilmiştir.**

Cinsiyet	Karşı değil	Karşı
Erkek	495	405
Kadın	620	480

"Kadın" ve "Karşı değil" olayları bağımsız olaylar mıdır?

TAMAMLAYICI (BÜTÜNLEYİCİ) OLAYLAR

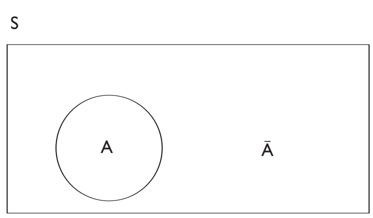
AMAÇ 7 ***Tamamlayıcı olaylar kavramını açıklayabileceksiniz.***

İki ayrık olay bir deneyin tüm sonuçlarını içeriyorsa bu olaylara, tamamlayıcı ya da bütünleyici olaylar denmektedir. Tamamlayıcı olaylar her zaman ayrık olaylardır.

Tamamlayıcı Olaylar: A olayının tamamlayıcısı $\bar{A}$ ile gösterilmekte ve A tamamlayanı denmektedir. $\bar{A}$ olayı, bir deneyde A olayı tarafından içerilmeyen tüm sonuçları içermektedir.

Aşağıdaki Şekil 4.11'deki Venn diyagramından da anlaşılacağı gibi A ve $\bar{A}$ olayları birbirlerinin tamamlayıcısıdır.

Tamamlayıcı olaylar: A olayının tamamlayıcısı $\bar{A}$ ile gösterilmekte ve A tamamlayanı denmektedir. $\bar{A}$ olayı, bir deneyde A olayı tarafından içerilmeyen tüm sonuçları içermektedir.



***Şekil 4.11** İki Tamamlayıcı olayın Venn diyagramı.*

Tamamlayıcı olayların bir deneyin tüm sonuçlarını içermesi nedeniyle, tamamlayıcı olayların olasılıklar toplamı 1'dir.

$$P(A) + P(\bar{A}) = 1 \; ; \; P(A) = 1 - P(\bar{A}) \; ; \; P(\bar{A}) = 1 - P(A)$$

Bu eşitlikler nedeniyle herhangi bir olayın olma olasılığı biliniyorsa, bu olayın tamamlayıcısının olasılığı, verilen olasılığın 1'den çıkartılmasıyla bulunabilmektedir.

ÖRNEK 21 ***Beş çamaşır makinesinden iki tanesi bozuktur. Bu makinelerden bir tanesinin rasgele seçilmesi deneyinde tamamlayıcı olaylar ve bunların olasılıkları nedir?***

ÇÖZÜM

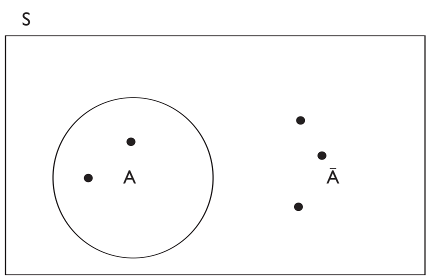
Bu deney için iki tamamlayıcı olay;

A = seçilen makine bozuktur
$\bar{A}$ = seçilen makine bozuk değildir

biçimindedir. Deneydeki beş makinenin üçü sağlam, ikisi bozuk olduğu için yukarıda tanımlanan olayların olasılıkları,

$$P(A) = 2 / 5 = 0.40 \quad \text{ve} \quad P(\bar{A}) = 3 / 5 = 0.60$$

olup, olasılıklar toplamı 1'dir. Bu örneğin Venn diyagramı aşağıda verilmiştir:



Şekil 4.12 *Örneğin Venn diyagramı.*

ÖRNEK 22 ***Bir üniversitede 90 tanesi (aynı zamanda) doktora derecesine de sahip 120 profesör bulunmaktadır. Bu üniversiteden, bir profesörün rasgele seçilmesi deneyinde, tamamlayıcı olaylar ve olasılıklar nedir?***

ÇÖZÜM

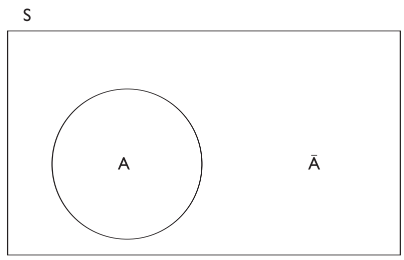
Bu deneyde tamamlayıcı olaylar;

A = seçilen profesörün doktora derecesine sahip olması
$\bar{A}$ = seçilen profesörün doktora derecesine sahip olmaması

biçimindedir. Bunlara ilişkin olasılıklar;

$$P(A) = 90 / 120 = 0.75 \quad \text{ve} \quad P(\bar{A}) = 1 - 0.75 = 0.25$$

olup, Venn diyagramı aşağıda verilmiştir:



Şekil 4.13 Örneğin Venn diyagramı.

SIRA SİZDE

1. **Zarın bir kez atılması deneyinde A olayı, sonucun 3'ten küçük bir sayı gelmesidir. A olayının ortaya çıkma olasılığı nedir? A olayının tamamlayıcı olayı ve olasılığı nedir?**

2. **DİE tarafından yapılan bir araştırmada 7183 kişinin birden çok gelir getirici işi olduğu, bunlardan 4054 tanesininse erkek olduğu bulunmuştur. Bu 7183 kişiden rasgele bir tanesinin seçilmesi durumunda ortaya çıkacak iki tamamlayıcı olay ve bunların olasılıkları nedir?**

3. **Erişkinlerin her gün gazete okuma olasılığı 0.65'tir. Bu olayın tamamlayıcı olayı ve olasılığı nedir?**

OLAYLARIN ARA KESİTİ VE ÇARPMA KURALI

AMAÇ 8 *Aynı anda ortaya çıkan olayların olasılığını hesaplayabileceksiniz.*

Bu bölümde iki olayın ara kesiti (ortak bölge) ve olayların ara kesitinin olasılığını hesaplamada kullanılan çarpma kuralı verilecektir.

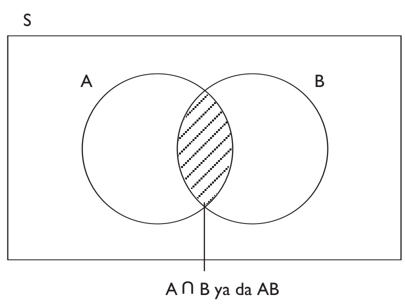
Olayların Ara Kesiti

İki olayın ara kesiti, aynı sonuçlar iki olayda da ortaksa oluşmaktadır.

OLAYLARIN ARA KESİTİ: Bir örneklem uzayında A ve B olayları tanımlanmış olsun. A ve B'nin ara kesiti, hem A hem de B'de yer alan sonuçları ifade eder, A ve B biçiminde gösterilir.

Olayların ara kesiti: Bir örneklem uzayında A ve B olayları tanımlanmış olsun. A ve B'nin ara kesiti, hem A hem de B'de yer alan sonuçları ifade eder, A ve B biçiminde gösterilir.

A ve B olaylarının ara kesiti (kesişimi) $A \cap B$ ya da AB olarak aşağıdaki şekilde gösterilmiştir.



Şekil 4.14 A ve B olaylarının ara kesiti.

Bileşik olasılık:
İki olayın ara kesitinin olasılığına, bunların bileşik olasılığı adı verilir ve P (A ve B) olarak gösterilir.

Çarpma kuralı:
A ve B olaylarının ara kesitinin olasılığı
P (A ve B) = P (A) P (B | A) ‘dır.

Çarpma Kuralı

A ve B gibi birlikte ortaya çıkan olayların olasılığına bileşik olasılık adı verilir ve P (A ve B) biçiminde gösterilir.

İki olayın ara kesitinin olasılığı, bir olayın bileşen olasılığıyla ile ikinci olayın koşullu olasılığından elde edilir ve bu kurala çarpma kuralı denir.

A ve B olayının bileşik olasılığı P ($A \cap B$) ya da P (AB) olarak da gösterilir.

ÖRNEK 23

Aşağıdaki tabloda bir firmada çalışanların cinsiyetlerine ve üniversite mezunu olup olmadıklarına göre dağılımları verilmiştir.

Tablo 4.7 *İki-Yönlü Sınıflama Tablosu.*

Cinsiyet	Üniversite mezunu	Üniversite mezunu değil	Toplam
Erkek	7	20	27
Kadın	4	9	13
Toplam	11	29	40

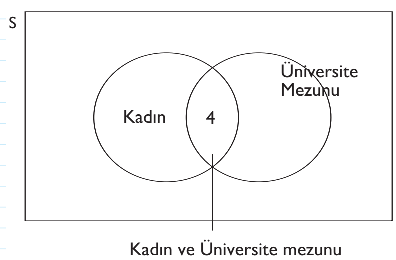
Çalışanlardan bir tanesi işçi yönetim komitesi için rasgele seçilecek olursa, seçilenin kadın ve üniversite mezunu olma olasılığı nedir?

ÇÖZÜM

İstenen olasılık, “Kadın (K)” ve “Üniversite mezunu (M)” olaylarının ara kesiti olup,

P (K ve M) = P (K) P (M | K)

eşitliğinden hesaplanmaktadır. Aşağıdaki şekilde de görüleceği gibi,



Şekil 4.15 *İki olayın ara kesiti.*

toplam 40 çalışanın 13 tanesi kadındır ve olasılığı P (K) = 13 / 40 ‘dır. Aynı biçimde P (M | K) olasılığını bulmak da kolaydır. Çünkü seçilecek kişi 13 kadından biridir ve bunlar arasından sadece 4 tanesi üniversite mezunudur. Bu nedenle

P (M | K) = 4 / 13

değerinden yararlanarak K ve M bileşik olasılığı,

P (M ve K) = P (K) P (M | K) = (13 / 40) (4 / 13) = 0.100

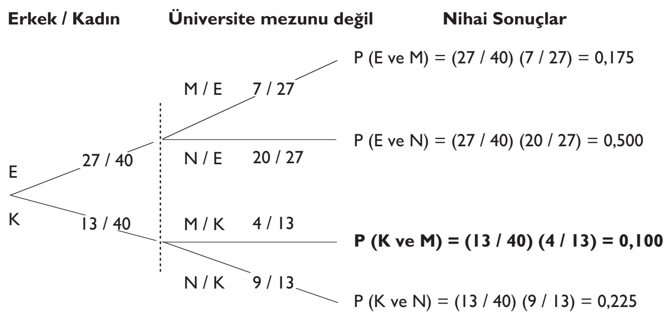
olarak bulunur.

Bu örnekte istenen olasılık, çarpma kuralı kullanılmadan da bulunabilir. Çünkü yukarıda verilmiş bulunan tablo ve şekilden de kolayca anlaşılacağı gibi, toplam 40 çalışandan “Kadın” ve “Üniversite mezunu” özelliklerini taşıyan sadece 4 çalışan bulunmaktadır. Yani, aranan bileşik olasılık P (M ve K) = 4 / 40 = 0.10'dur.

Benzer biçimde; “Erkek (E)” ve “Üniversite mezunu değil (N)” olayları da dikkate alındığında yukarıdaki tabloya ilişkin diğer bileşik olasılıklar kolayca elde edilebilir.

P (E ve M) = P (E) P (M | E) = (27 / 40) (7 / 27) = 0.175
P (E ve N) = P (E) P (N | E) = (27 / 40) (20 / 27) = 0.500
P (K ve N) = P (K) P (N | K) = (13 / 40) (9 / 13) = 0.225

Aşağıdaki ağaç diyagramı, dört bileşik olasılığı göstermektedir. Örnekte bulunması istenen olasılık koyu renkle gösterilmiştir.



Şekil 4.16 *Ağaç diyagramı.*

ÖRNEK 24

Bir kutuda 4 tanesi bozuk, toplam 20 kaset bulunmaktadır. Bu kasetlerden (seçilen yerine konulmaksızın) iki tanesi rasgele seçilmiştir. Seçilen iki kasetin de bozuk olma olasılığı nedir?

ÇÖZÜM

Bu deneyle ilgili olarak, önce olaylar tanımlanacak olursa olaylar;

S_1 = İlk seçilen kasetin sağlam olması
B_1 = İlk seçilen kasetin bozuk olması
S_2 = İkinci seçilen kasetin sağlam olması
B_2 = İkinci seçilen kasetin bozuk olması

biçimindedir. Burada istenen B_1 ve B_2 olaylarının, bileşik olasılıkları olup,

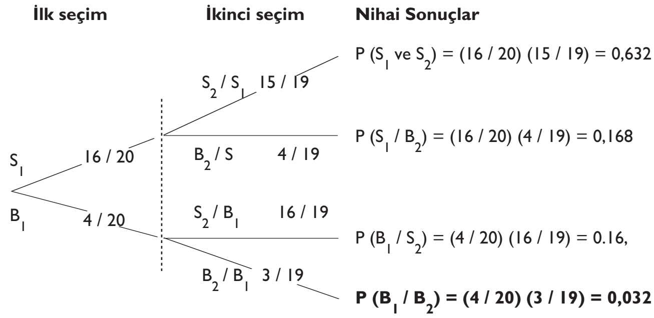
$$P(B_1 \text{ ve } B_2) = P(B_1)\, P(B_2 \mid B_1)$$

eşitliğinden elde edilecektir. Toplam 20 kasetin 4 tanesi bozuk olduğu için ilk seçilen kasetin bozuk olma olasılığı $P(B_1) = 4 / 20$ 'dir. İlk seçilen kasetin bozuk olduğu bilindiğine göre, kutuda üçü bozuk 19 tane kaset kalmıştır. O halde, ikinci seçilen kasetin de bozuk olma olasılığı $P(B_2 \mid B_1) = 3 / 19$ 'dur ve istenen olasılıksa,

$$P(B_1 \text{ ve } B_2) = P(B_1)\, P(B_2 \mid B_1) = (4 / 20)\ (3 / 19) = 0.032$$

olarak bulunur.

Aşağıdaki şekilde, istenen olasılık $P(B_1 \text{ ve } B_2)$ koyu renkli olarak belirtilerek dört nihai sonuç (bileşik olasılık) verilmiştir.



***Şekil 4.17** İki kaset seçimi.*

Koşullu olasılık: Eğer A ve B, P (A) ≠ 0 ve P (B) ≠ 0 durumunda iki olay ise bunlara ilişkin koşullu olasılıklardır.

$$P\left(B \mid A\right) = \frac{P\,(A \text{ ve } B)}{P\,(A)}$$

ve

$$P\,(B) = \frac{P\,(A \text{ ve } B)}{P\,(B)}$$

Alt bölüm, 4.4'de verilmiş olan koşullu olasılık konusu hatırlandığındave bir A olayının olma olasılığı biliniyorsa, yine A ve B olayının bileşik olasılığı biliniyorsa, bu bilgilerden yararlanarak A olayı bilindiğinde B olayının koşullu olasılığını bulmak kolaydır.

Koşullu Olasılık: Eğer A ve B, P (A) ≠ 0 ve P (B) ≠ 0 durumunda iki olay ise bunlara ilişkin koşullu olasılıklardır.

$$P\left(B \mid A\right) = \frac{P\,(A \text{ ve } B)}{P\,(A)} \quad \text{ve} \quad P\,(B) = \frac{P\,(A \text{ ve } B)}{P\,(B)}$$

ÖRNEK 25

Mühendislik fakültesinden rasgele seçilen bir öğrencinin kız olma olasılığı 0.20, bu öğrencinin bilgisayar mühendisliği öğrencisi ve kız öğrenci olmasının bileşik olasılıysa ise 0.03 'tür. Seçilen öğrencinin kız olduğu bilindiğinde, bu öğrencinin bilgisayar mühendisliği öğrencisi olmasının koşullu olasılığını bulunuz.

ÇÖZÜM

Önce örnekle ilgili iki olay tanımlanacak olursa;

A = Seçilen öğrenci kız
B = Seçilen öğrenci bilgisayar mühendisliği öğrencisi

biçimindedir. Bu bilgiler ışığında P (A) = 0.20 ve P (A ve B) = 0.03 kullanılarak,

sonucuna ulaşılır ki bu değer seçilen öğrencinin kız olduğu bilindiğinde bu öğrencinin bilgisayar mühendisliği öğrencisi olması (koşullu) olasılığıdır.

Bağımsız Olaylar İçin Çarpma Kuralı

Burada verilecek olan çarpma kuralı, iki olayın bağımsız olması durumunda kullanılmaktadır. A ve B gibi iki olayın bağımsız oldukları düşünüldüğünde,

$$P\,(A) = P\,(A \mid B) \text{ ve } P\,(B) = P\,(B \mid A)$$

yazılabilmektedir. Bu durumda P (B | A) yerine P (B) yazılacak olursa, yukarıda verilen A ve B olaylarının bileşik olasılığına ilişkin formül,

$$P\,(A \text{ ve } B) = P\,(A)\; P\,(B \mid A) = P\,(A)\, P\,(B)$$

biçimine dönüşür.

Bağımsız Olaylar İçin Çarpım Kuralı: A ve B gibi bağımsız iki olayın bileşik olasılığı

P (A ve B) = P (A) P (B)

'dir.

Bağımsız olaylar için çarpım kuralı: A ve B gibi bağımsız iki olayın bileşik olasılığı

P (A ve B) = P (A) P (B)

dir.

ÖRNEK 26

Bir iş hanında iki tane yangın dedektörü bulunmaktadır. Bir yangın sırasında bu dedektörlerden herhangi birinin çalışmaması olasılığı 0.02' dir. Bir yangın sırasında her iki dedektörün de çalışmama olasılığını bulunuz.

ÇÖZÜM

Bu örnekte iki yangın dedektörü birbirinden bağımsızdır. Çünkü, yangın sırasında dedektörlerden birinin çalışmaması ötekini etkilememektedir. Bu durumda aşağıdaki iki olay tanımlanabilir.

A = İlk dedektörün çalışmaması
B = İkinci dedektörün çalışmaması

A ve B olayları bağımsız olduklarından, bunların bileşik olasılığı,

P (A ve B) = P (A) P (B) = (0.02) (0.02) = 0.0004

olarak bulunur.

Aşağıdaki örnekte de görüleceği gibi çarpma kuralı, ikiden çok olay olması durumunda da bileşik olasılığın bulunmasında kullanılmaktadır.

ÖRNEK 27

Penisilinin hastada alerji yapması olasılığı 0.20' dir. Bu ilacın üç hastaya verildiği bir durumda;
a) Üç hastaya da alerji yapması
b) En az bir hastaya alerji yapmaması
olasılıklarını bulunuz.

ÇÖZÜM

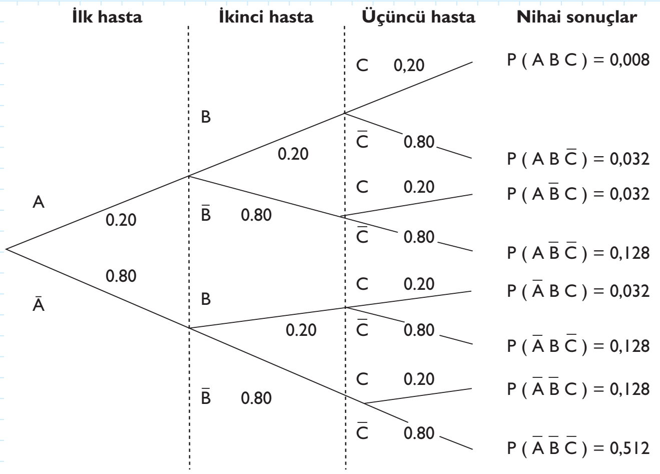
Bu örnekte, A, B ve C, sırasıyla penisilin verilen birinci, ikinci ve üçüncü hastaları göstersin.

a) Penisilin verilen üç hastada alerji yapması, bu üç hastanın bileşik olasılığından elde edilecektir. Hastalardan birinin alerji olması ötekileri etkilemeyeceği için A, B ve C olayları birbirinden bağımsızdır. Yani,

P (A ve B ve C) = P (A) P (B) P (C) = (0.20) (0.20) (0.20) = 0.008

istenen olasılık değeridir.

$\bar{A}$, $\bar{B}$ ve $\bar{C}$ tanımlanmış olan A, B ve C olaylarının tamamlayıcıları olmak üzere bu deneyin tüm sonuçları ağaç diyagramı üzerinde kolayca gösterilebilir:



Şekil 4.18 *Ağaç diyagramı.*

b) Burada da,

G = Üç hasta da alerjiktir
H = En az bir hasta alerjik değildir

olayları tanımlanmış olsun. Tanımlanan G ve H olayları tamamlayıcı olaylardır. Bu nedenle,

P (G) = P (A ve B ve C) = 0.008

sonucundan ve tamamlayıcı olay kuralından yararlanarak,

P (H) = 1 - P (G) = 1 – 0.008 = 0.992

değeri bulunur.

Ayrık Olayların Bileşik Olasılığı

Ayrık olayların bileşik olasılığı: İki ayrık olayın bileşik olasılığı her zaman sıfırdır ve bu durum, A ve B ayrık olaylar ise P (A ve B) = 0 biçiminde gösterilir.

Önceki tartışmalardada dile getirildiği gibi, ayrık olaylar birlikte meydana gelmezler. Bu nedenle, iki ayrık olayın bileşik olasılığı sıfırdır.

Ayrık Olayların Bileşik Olasılığı: İki ayrık olayın bileşik olasılığı her zaman sıfırdır ve bu durum, A ve B ayrık olaylarsa P (A ve B) = 0 biçiminde gösterilir.

ÖRNEK 28

Otomobil kredisi almak için gerekli başvuru formunun doldurulmasına ilişkin aşağıdaki iki olay tanımlanmış olsun;

O = Kredi başvurusu onaylandı
R = Kredi başvurusu reddedildi

O ve R'nin bileşik olasılığı nedir?

ÇÖZÜM

Örnekte tanımlanan O ve R olayları ayrık olaylar olduğu için bileşik olasılığı sıfırdır.

(P (O ve V) = 0)

SIRA SİZDE

1. **İki olayın ara kesitinin anlamını açıklayınız ve bir örnek veriniz.**

2. **İki olayın bileşik olasılığının anlamını açıklayınız ve bir örnek veriniz.**

3. **Aşağıda verilen değerler için, A ve B olaylarının bileşik olasılığını bulunuz.**
   a) P (B) = 0.59 ve P (A | B) = 0.77
   b) P (A) = 0.28 ve P (B | A) = 0.15

OLAYLARIN BİLEŞİMİ VE TOPLAMA KURALI

AMAÇ 9

Olaylardan en az birinin ortaya çıkmasına ilişkin olasılığı hesaplayabileceksiniz.

Bu alt bölümde olayların bileşimi (bileşik olaylar) incelenecek ve bileşik olaylarda olasılık bulmada kullanılan toplama kuralından söz edilecektir.

Olayların Bileşimi

A ve B gibi iki olayın bileşimi; A'da ya da B'de ya da A ve B'de birlikte yer alan tüm sonuçları içermektedir.

Olayların Bileşimi: Aynı örneklem uzayında tanımlı A ve B olaylarının bileşimi A'da ya da B'de ya da A ve B'de birlikte yer alan tüm olayların bileşkesi olup A ya da B biçiminde gösterilir.

Olayların bileşimi: Aynı örneklem uzayında tanımlı A ve B olaylarının bileşimi A'da ya da B'de ya da A ve B'de birlikte yer alan tüm olayların bileşkesi olup A ya da B biçiminde gösterilir.

ÖRNEK 29

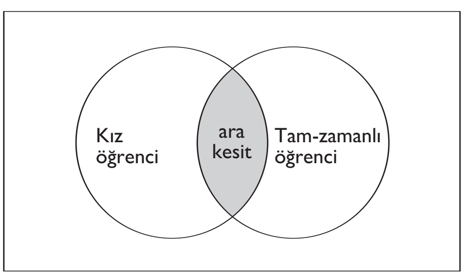
ABD üniversitelerinde 12.439 milyon öğrenci öğrenim görmektedir. Bunlardan 6.868 milyonu kız öğrenci, 7.211 milyonu tam zamanlı öğrenci ve 3.786 milyonu kız ve tam zamanlı öğrencilerdir. "Kız" ve "Tam zamanlı öğrenci" olaylarının bileşimini tanımlayınız.

Burada; "Kız öğrenci" ve "Tam zamanlı öğrenci" olaylarının bileşimi, kız ya da tam zamanlı ya da her iki grupta da yer alan öğrencilerden oluşmaktadır. Bu öğrencilerin sayısı ise 6.868 + 7.221 – 3.786 = 10.303 milyondur. Burada en önemli nokta, iki olayın içerdiği toplam 6.868 + 7.221 = 14.089 milyon kişiden 3.786 milyonunun her iki grupta (olay) yer alması, yani iki olayın ara kesiti olması nedeniyle toplamdan çıkartılmasıdır. Aşağıdaki tablo ve grafikten de görüleceği gibi bu değerlerin toplamdan çıkartılmaması iki kez sayılmasına yani tekrara neden olacaktır.

Cinsiyet	Tam-zamanlı	Yarı-zamanlı	Toplam
Erkek	3.435	2.136	5.571
Kız	3.786 *	3.082	6.868
Toplam	7.211	5.218	12.439

* iki kez tekrarlanan

Tablo 4.8 *İki yönlü sınıflama tablosu.*



Şekil 4.19 *"Kız öğrenci" ve "Tam-zamanlı öğrenci" olaylarının bileşimi.*

Toplama kuralı: A ve B olaylarının bileşiminin olasılığı,
P (A veya B) = P (A) + P (B) - P (A ve B)
biçiminde bulunmaktadır.

Toplama Kuralı

Olayların bileşimine ilişkin olasılık hesaplamada kullanılan yönteme, toplama kuralı denir ve aşağıdaki gibi tanımlanır.

TOPLAMA KURALI: A ve B olaylarının bileşiminin olasılığı,

$$P\,(A \text{ veya } B) = P\,(A) + P\,(B) - P\,(A \text{ ve } B)$$

biçiminde bulunmaktadır.

Bu eşitlik gereğince A ve B olaylarının bileşen olasılıklarının toplamından bunların bileşik olasılıklarının çıkartılmasıyla bu olayların bileşiminin olasılığı bulunur.

ÖRNEK 30

Bir üniversite rektörü, tüm öğrencilerin etik konusunda bir dersi (zorunlu) almasının yararlı olacağını düşünmektedir. Bu konuda öğretim elemanı ve öğrencilerden oluşan toplam 300 kişiye düşüncesini sormuş ve elde edilen sonuçlardan aşağıdaki tablo oluşturulmuştur.

Tablo 4.9 *İki yönlü sınıflama tablosu.*

	Görüş			
Sorulan	Katılıyor	Karşı	Çekimser	Toplam
Öğretim elemanı	45	15	10	70
Öğrenci	90	110	30	230
Toplam	135	125	40	300

Bu gruptan rasgele seçilen birinin öğretim elemanı ya da katılıyor olma olasılığını bulunuz.

ÇÖZÜM

İlk olarak aşağıdaki olaylar tanımlanacak olursa;

A = Seçilen kişi öğretim elemanı
B = Seçilen kişi düşünceye katılmakta

biçimindedir. Tablo bilgilerinden yararlanarak bu olaylara ilişkin olasılıklar,

$$P\,(A) = 70 / 300 = 0.233$$
$$P\,(B) = 135 / 300 = 0.450$$
$$P\,(A \text{ ve } B) = P\,(A)\,P\,(B \mid A) = (70 / 300)\,(45 / 70) = 0.150$$

olarak bulunur. Aranan olasılık değeriyse toplama kuralından yararlanılarak,

$$P\,(A \text{ veya } B) = P\,(A) + P\,(B) - P\,(A \text{ ve } B) = 0.233 + 0.450 - 0.150 = 0.533$$

elde edilir.

Aynı sonucu toplama kuralı kullanmadan sadece tablo bilgilerinden yararlanarak bulmak da olanaklıdır.

$$45 + 15 + 10 + 90 = 160$$
$$P\,(A \text{ veya } B) = 160 \mid 300 = 0.533$$

ÖRNEK 31

Yapılan bir araştırmadan 7225 kişinin birden çok işi olduğu bulunmuştur. Bu kişilerden 4115 tanesi erkek, 1742 tanesi bekar ve 905 tanesiyse erkek ve bekardır. Rasgele seçilen birinin erkek ya da bekar olma olasılığını bulunuz.

ÇÖZÜM

İstenen olasılık değerini bulmak için olayları tanımlamak ve bu olaylara ilişkin olasılıkları yazmak gerekir.

E = Rasgele seçilen kişi erkek
B = Rasgele seçilen kişi bekar

$$P(E) = 4.155 / 7.225 = 0.570$$
$$P(B) = 1.742 / 7.225 = 0.241$$
$$P(E \text{ ve } B) = 905 / 7.225 = 0.125$$

Bu sonuçların toplama kuralında kullanılmasıyla aranan olasılık değeri,

$$P(E \text{ veya } B) = P(E) + P(B) - P(E \text{ ve } B) = 0.570 + 0.241 - 0.125 = 0.686$$

olarak bulunur.

Ayrıca örnekte verilen bilgiler **(koyu renkli)** kullanılarak aşağıdaki tablo oluşturulabilir ve aranan olasılık bu bilgilerden yararlanılarak da bulunabilir.

Cinsiyet	Bekar	Evli	Toplam
Erkek	905	3.210	4.115
Kadın	837	2.273	3.110
Toplam	1.742	5.483	7.225

Tablo 4.10 *İki yönlü sınıflama tablosu.*

$$P(E) = 4.155 / 7.225 = 0.570$$
$$P(B) = 1.742 / 7.225 = 0.241$$
$$P(E \text{ ve } B) = 905 / 7.225 = 0.125$$
$$P(E \text{ veya } B) = P(E) + P(B) - P(E \text{ ve } B) = 0.570 + 0.241 - 0.125 = 0.686$$

Ayrık Olaylar İçin Toplama Kuralı

Olasılık konusunun önceki alt bölümlerinde, iki ayrık olayın bileşik olasılık değerinin sıfır olduğu söylenmişti. Bu nedenle A ve B ayrık olaylar olmak üzere P (A ve B) değeri sıfır olduğu için, toplama kuralında kullanılan formülden çıkartılmakta ve iki ayrık olayın bileşiminin olasılığı, bu iki olayın bileşen olasılıklarının toplamından elde edilmektedir.

Ayrık Olaylar İçin Toplama Kuralı: A ve B ayrık olaylarının bileşiminin olasılığı

$$P(A \text{ veya } B) = P(A) + P(B)$$

'dir.

Ayrık olaylar için toplama kuralı: A ve B ayrık olaylarının bileşiminin olasılığı

$P(A \text{ ya da } B) = P(A) + P(B)$

dir.

ÖRNEK 32

Kamuoyunun ırk ayrımına karşı düşüncesinin ortaya konmasını amaçlayan bir araştırmada; erkek ve kadınlardan oluşan 300 kişi ile görüşülmüştür. Görüşlerin ırk ayrımına "karşı" , "destekliyor" ve "çekimser" olarak kaydedilmesi sonucunda aşağıdaki tablo oluşturulmuştur.

Tablo 4.11. *İki yönlü sınıflama tablosu.*

Cinsiyet	Karşı	Destekliyor	Çekimser	Toplam
Erkek	45	15	12	72
Kadın	85	115	28	228
Toplam	130	130	40	300

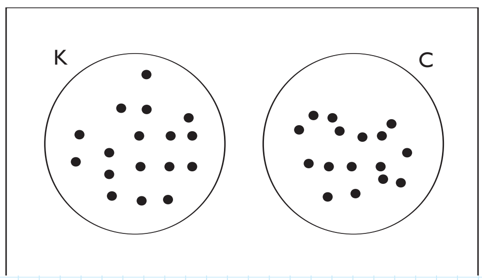
Bu gruptan rasgele seçilen bir kişinin ırk ayrımına karşı ya da çekimser olma olasılığı nedir ?

ÇÖZÜM

İlk olarak aşağıdaki olaylar tanımlanır;

K = Çekilen kişi ırk ayrımına karşı
C = Çekilen kişi ırk ayrımı konusunda çekimser

Bu örnek için tanımlanan olaylar, aşağıdaki Şekil 4.20'den de görüleceği gibi ayrık olaylardır. Çünkü seçilen kişi ırk ayrımına ya karşı olacaktır ya da çekimser olacaktır. Yani her iki görüşte olması söz konusu değildir.



Şekil 4.20 *K ve C ayrık olayları.*

Yukarda verilen bilgilerden,

$P(K) = 130 / 300 = 0.43$
$P(C) = 40 / 300 = 0.133$
$P(K \text{ veya } C) = P(K) + P(C) = 0.43 + 0.133 = 0.563$

elde edilir.

Toplama kuralı formülü ikiden çok olay olması durumlarında da kullanılabilmektedir.

ÖRNEK 33

Zarın iki kez atıldığı bir deneyde, iki atış sonunda toplam 5, ya da 7 ya da 10 elde edilmesi olasılığını bulunuz.

ÇÖZÜM

Bir zarın iki kez atılması deneyinde toplam 36 sonuç bulunmaktadır. Eşit olasılıklı bu sonuçlar Tablo 4.12'de verilmiştir.

Tablo 4.12 Bir zarın iki kez atılmasından elde edilen sonuçlar.

İlk atış \ İkinci atış	1	2	3	4	5	6
1	(1. 1)	(1. 2)	(1. 3)	(1. 4)	(1. 5)	(1. 6)
2	(2. 1)	(2. 2)	(2. 3)	(2. 4)	(2. 5)	(2. 6)
3	(3. 1)	(3. 2)	(3. 3)	(3. 4)	(3. 5)	(3. 6)
4	(4. 1)	(4. 2)	(4. 3)	(4. 4)	(4. 5)	(4. 6)
5	(5. 1)	(5. 2)	(5. 3)	(5. 4)	(5. 5)	(5. 6)
6	(6. 1)	(6. 2)	(6. 3)	(6. 4)	(6. 5)	(6. 6)

İki sayının toplamının 5, ya da 7 ya da 10 olduğu olaylar tabloda işaretlenmiştir. Burada “toplam 5”, “toplam 7” ve “toplam 10” biçimindeki üç olay ayrık olaylardır. Toplamı 5 olan dört sonuç, toplamı 7 olan altı sonuç ve toplamı 10 olan üç sonuç bulunduğu için aranan olasılık,

$$P\ (5 \text{ veya } 7 \text{ veya } 10) = P\ (5) + P\ (7) + P\ (10)$$
$$= (4 / 36) + (6 / 36) + (3 / 36) = 13 / 36 = 0.361$$

olarak bulunur.

ÖRNEK 34

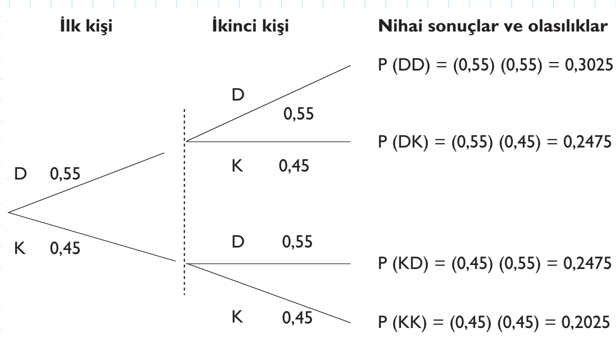
Kamuoyunun %55’inin genetik mühendisliğini desteklediği, %45’ininse karşı olduğu bilinmektedir. Rasgele iki kişi seçilerek ve bu kişilerin genetik mühendisliğini destekleyip desteklemedikleri öğrenilmek istenmektedir.
a) Bu deneyin ağaç diyagramını çiziniz.
b) İki kişiden en az birinin genetik mühendisliğini desteklemesi olasılığını bulunuz.

ÇÖZÜM

a) Deneyin olayları tanımlanacak olursa;

D = Genetik mühendisliğinin desteklenmesi
K = Genetik mühendisliğine karşı olunması

biçimindedir. Bu deneyde dört nihai sonuç bulunmaktadır.



Şekil 4.21 Ağaç diyagramı.

b) İki kişiden en az birinin genetik mühendisliğini desteklemesi DD, DK ve KD sonuçlarını içermektedir ve bu üç ayrık olayın bileşiminin olasılığı,

$$\begin{aligned} P\ (\text{en az bir kişi destekliyor}) &= P\ (DD,\ DK \text{ ya da } KD) \\ &= P\ (DD) + P\ (DK) + P\ (KD) \\ &= 0.3025 + 0.2475 + 0.2475 \\ &= 0.7975 \end{aligned}$$

'dir.

SIRA SİZDE

1. İki olayın bileşiminin anlamını açıklayınız ve bir örnek veriniz.

2. Aşağıdaki toplama kuralı, hangi durumdaki A ve B gibi iki olayın bileşiminin, olasılığını bulmada kullanılır?
P (A veya B) = P (A) + P (B)

3. Aşağıdaki durumlar için P (A ya da B) olasılığını bulunuz.
a) P (A) = 0.18 , P (B) = 0.49 ve P (A ve B) = 0.13
b) P (A) = 0.83 , P (B) = 0.71 ve P (A ve B) = 0.68

Kendimizi Sınayalım

1. 3 farklı para aynı anda atılmıştır. Bu deneye ilişkin örneklem uzayında kaç örneklem noktası olacaktır?

a. 3
b. 6
c. 8
d. 12
e. 16

2. Biri beyaz diğeri de kırmızı olmak üzere bir çift zar atılmıştır. A zarların üste gelen yüzeydeki sayıların toplamının çift, B'de kırmızı zarın üste gelen yüzündeki sayının tek sayı olması olayları iken, $(A \cup B)$ olayında kaç örneklem noktası olacaktır?

a. 12
b. 16
c. 18
d. 36
e. 40

3. Bir sınıfta 10 kız 20 erkek öğrenci vardır. Kız öğrencilerin de erkek öğrencilerin de yarısı bursludur. Bu sınıftan gelişigüzel bir öğrenci seçilmiştir. Bu öğrencinin erkek ya da burslu olma olasılığı aşağıdakilerden hangisidir?

a. $\frac{3}{6}$
b. $\frac{3}{5}$
c. $\frac{4}{6}$
d. $\frac{4}{5}$
e. 5

4. Bir çift zar atılmıştır. Üste gelen yüzeylerdeki sayıların toplamının 7 olduğu biliniyorsa, bu zarlardan birinin 3 olma olasılığı aşağıdakilerden hangisidir?

a. $\frac{1}{18}$
b. $\frac{1}{3}$
c. $\frac{2}{5}$
d. $\frac{3}{7}$
e. $\frac{1}{2}$

5. 52 kartlık bir oyun kağıdı destesinden peş peşe iki kart çekilmiştir. Birinci kart ikili iken, ikinci kartın da ikili olma olasılığı aşağıdakilerden hangisidir?

a. $\frac{1}{221}$
b. $\frac{7}{360}$
c. $\frac{11}{101}$
d. $\frac{7}{52}$
e. $\frac{11}{52}$

6. Bir sınıftaki öğrencilerin %30'u matematik, %20'si istatistik ve %10'u da hem matematik hem de istatistik derslerinden başarısızdır. Bu sınıftan gelişigüzel seçilen bir öğrencinin matematik ya da istatistik derslerinin birinden başarısız olma olasılığı aşağıdakilerden hangisidir?

a. 0.20
b. 0.30
c. 0.35
d. 0.40
e. 0.45

7. Aynı sınıftan iki öğrencinin Türk Dili dersi sınavından başarılı olma olasılıklarının sırasıyla 0.6 ve 0.8 olduğu biliniyorsa bu iki öğrenciden birisinin başarılı olma olasılığı aşağıdakilerden hangisidir?

a. 0.30
b. 0.48
c. 0.50
d. 0.75
e. 0.92

8. 1'den 15'e kadar (15 dahil) olan tam sayılar arasından rasgele seçilen bir sayının 3 ve 5 ile bölünebilen bir sayı olma olasılığı aşağıdakilerden hangisidir?

a. $\frac{1}{15}$
b. $\frac{1}{3}$
c. $\frac{3}{5}$
d. $\frac{10}{15}$
e. $\frac{13}{15}$

9. Bir A dersi sınavına giren 200 öğrencinin cinsiyetlerine göre başarı durumu aşağıdaki tabloda gösterilmiştir.

Cinsiyet \ Başarı	BAŞARILI	BAŞARISIZ	TOPLAM
ERKEK	80	40	120
KIZ	60	20	80
TOPLAM	140	60	200

Sınava giren öğrenciler arasından rasgele seçilen bir öğrencinin erkek ve başarılı olma olasılığı aşağıdakilerden hangisidir?

a. $\frac{1}{5}$
b. $\frac{2}{5}$
c. $\frac{3}{5}$
d. $\frac{2}{3}$
e. $\frac{7}{10}$

10. Bir futbol takımının ilk üç maçı kazanmasına ilişkin olasılıklar sırasıyla 0.3, 0.5 ve 0.8 olarak belirlenmiştir. Bu takımın üç maçı da kazanma olasılığı aşağıdakilerden hangisidir?

a. 0.10
b. 0.12
c. 0.15
d. 0.18
e. 0.20

Yanıt Anahtarı

1. c
2. d
3. e
4. b
5. a
6. d
7. e
8. a
9. b
10. b

Yararlanılan Kaynaklar

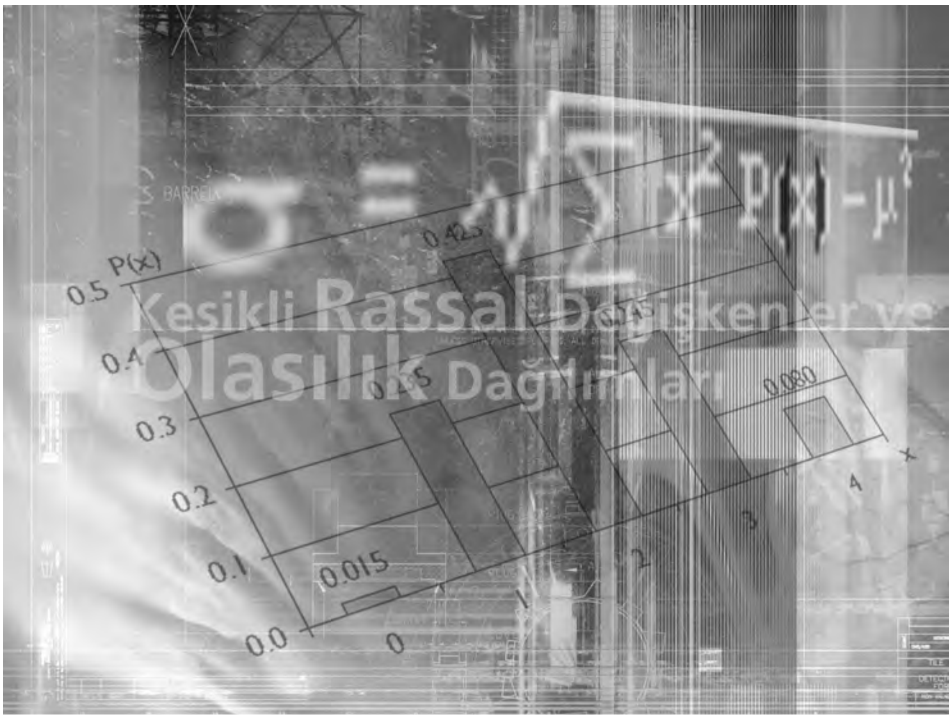
HOEL, P.G. and JESSEN, R.J.: **Basic Statistics for Business and Economics**, Wiley, NewYork, 1971.

MANN, P.S.: **Introductory Statistics,** 2^{nd} Edition, Wiley, New York, 1995.

O'HAGAN, A., **Probability: Metods and Measurement**, Chapman and Hall, London, 1988.

WONNACOTT, R.J., WONNACOTT, T.H.: **Introductory Statistics,** 4^{th} Edition, Wiley, Singapore, 1985.

5 Kesikli Rassal Değişkenler ve Olasılık Dağılımları



Çalışma Biçimine İlişkin Olarak:

- *Tanımlar ve kavramlar iyice incelenmeli,*
- *Örnekler tam olarak anlaşılmalı,*
- *Sorulan sorularda nelerin istendiği kesin olarak anlaşıldıktan sonra çözüme geçilmeli,*
- *Alıştırmaların tümü mutlaka çözülmelidir.*

Amaçlar:

- *Rassal değişken kavramını açıklayabileceksiniz.*
- *Kesikli rassal değişkenlerin ortalama ve standart sapmasını hesaplayabileceksiniz.*
- *Kesikli rassal değişkenlerin olasılık dağılımını oluşturabileceksiniz.*
- *Faktöriyel kavramını açıklayabilecek, kombinasyon hesaplayabileceksiniz.*
- *Kesikli rassal değişkenlerin önemli dağılımlarından Binom dağılımını kullanarak ilgili olasılıkları hesaplayabilecek, dağılımın ortalama ve standart sapmasını hesaplayabileceksiniz.*
- *Kesikli rassal değişkenlerin diğer önemli bir dağılımı olan Poisson dağılımını kullanarak ilgili olasılıkları, dağılımın ortalama ve standart sapmasını hesaplayabileceksiniz.*

İçerik Haritası

- *GİRİŞ*
- *RASSAL DEĞİŞKENLER*
  - *Kesikli Rassal Değişken*
  - *Sürekli Rassal Değişken*
- *KESİKLİ BİR RASSAL DEĞİŞKENİN OLASILIK DAĞILIMI*
- *KESİKLİ BİR RASSAL DEĞİŞKENİN ORTALAMASI VE STANDART SAPMASI*
  - *Kesikli bir rassal değişkenin ortalaması*
  - *Kesikli bir rassal değişkenin standart sapması*
  - *Standart Sapmanın Yorumu*
- *FAKTÖRİYELLER VE KOMBİNASYONLAR*
  - *Faktöriyeller*
  - *Kombinasyonlar*
- *BİNOM (İKİ TERİMLİ) OLASILIK DAĞILIMI*
  - *Binom Deneyi*
  - *Binom Olasılık Dağılımı ve Binom Formülü*
  - *Başarı Olasılığı ve Binom Dağılımının Biçimi*
  - *Binom Dağılımının Ortalama ve Standart Sapması*
- *POİSSON OLASILIK DAĞILIMI*

GİRİŞ

Bir önceki bölümde olasılık kavramından ve kurallarından söz edilmişti. Bu bölümdeyse olasılık kavramı genişletilerek olasılık dağılımlarından söz edilecektir. Hatırlanacağı gibi bir önceki bölümde, bir istatistiksel deneyin birden çok sonucu olduğu ve bu sonuçlardan hangisinin gerçekleşeceğinin önceden bilinmediği, ancak bazı belirsizlikler altında kestirilebileceği vurgulanmıştı. Örneğin loto oynayan bir kişi kazanıp kazanamayacağını önceden bilmemektedir. Eğer kazanamayacağını bilse kuşkusuz oynamazdı. Ancak, loto oynayan kişi sadece (çok net olmasa da) kazanma şansının olduğunu bilmektedir.

Bu bölümde, bir istatistiksel deneyin, çok kez tekrarlanması durumunda ne tür sonuçların elde edileceği konu edilecektir. Yukarıda verilmiş olan loto oyuncusu örneğinde, kişinin sürekli loto oynaması durumunda, (ortalama) kazanma ya da kaybetme olasılıklarının hesaplanması üzerinde durulacaktır.

İzleyen alt bölümlerde, rassal değişken ve türleri açıklanacak, olasılık dağılımlarıyla bu dağılımların ortalama ve standart sapmalarının bulunması ele alınacak ve son olarak da kesikli rassal değişkenler için önemli dağılımlardan olan Binom ve Poisson dağılımları incelenecektir.

RASSAL DEĞİŞKENLER

AMAÇ 1

Rassal değişken kavramını açıklayabilecek ve kesikli rassal değişkenler için olasılık dağılımını oluşturabileceksiniz.

Aşağıdaki tabloda 2000 ailenin sahip oldukları otomobil sayılarına göre sıklık ve göreli sıklık dağılımları verilmiştir:

Tablo 5.1 Sahip Olunan Otomobil Sayılarına Göre Ailelerin Sıklık ve Göreli Sıklık Dağılımları.

Otomobil Sayısı	Sıklık	Göreli Sıklık
0	30	30 / 2000 = 0.015
1	470	470 / 2000 = 0.235
2	850	850 / 2000 = 0.425
3	490	490 / 2000 = 0.245
4	160	160 / 2000 = 0.080
	N = 2000	Toplam = 1.000

Bu gruptan rassal bir aile seçilmiş ve ailenin sahip olduğu otomobil sayısı X ile gösterilmiş olsun. Yukarıdaki tablonun ilk sütununda da görüleceği gibi X (değişken)'in alabileceği beş değer (0 , 1, 2 , 3 , 4) bulunmaktadır ve X 'in değeri seçilen aileye göre değişim göstermektedir. Yani bu değer rassal deneyin sonuçlarına bağlı ve bu X değişkenine "rassal değişken" ya da "şans değişkeni" adı verilmektedir.

Rassal Değişken: Bir deney ya da gözlemin şansa bağlı sonucu bir değişkenin aldığı değer olarak düşünülürse, olasılık ve istatistikte böyle bir değişkene rassal değişken adı verilir.

Bir rassal değişken, aşağıda açıklanacağı gibi kesikli (discrete) ya da sürekli (continuous) olabilmektedir.

Kesikli Rassal Değişken

Bir kesikli değişken, değerleri sayımla elde edilen değişkendir. Başka bir deyişle bir kesikli değişkenin birbirini izleyen değerleri arasında belirli boşluklar vardır.

Kesikli Rassal Değişken: Genel anlamda bir rassal değişken sayılabilir değerler alıyorsa, bu değişkene kesikli rassal değişken denir.

Yukarıdaki tabloda verilmiş olan, sahip olunan otomobil sayısı kesikli rassal değişkene örnektir. Çünkü X ile gösterilmiş bulunan kesikli rassal değişken sadece sayılabilir 0 ,1, 2, 3 ve 4 değerlerini alabilmektedir.

Aşağıda kesikli rassal değişken için bazı örnekler verilmiştir.

1. Bir galerinin herhangi bir ayda satmış olduğu otomobil sayısı.
2. Herhangi bir günde bir tiyatroya gelen izleyici sayısı.
3. Bir kişinin sahip olduğu ayakkabı sayısı.
4. Bir para üç kez atıldığında yazı gelme sayısı.
5. Bir ailenin çocuk sayısı.

Sürekli Rassal Değişken

Değerleri ölçüm ya da tartımla elde edilen, bir başka anlatımla sayımla elde edilemeyen bir değişkene sürekli rassal değişken denir. Sürekli bir rassal değişkenin değerleri aralıklar halinde tanımlanır.

Sürekli Rassal Değişken: alacağı herhangi bir değerle, bir ya da daha fazla aralıkta tanımlanan değişkene, sürekli rassal değişken denir.

Bir aralıkta sonsuz sayıda değer olacağı için, sürekli rassal bir değişkenin alabileceği değer sayısı da sonsuz kabul edilir ve bu değerlerin sayılması olanaksızdır denir. Örneğin bir pilin ömrü 40 , 40.25 ya da 40.247 saat olabilmektedir. Ancak bilinmektedir ki bir pilin ömrü en çok 200 saattir. Bu örnek için X, rassal seçilen bir pilin ömrü olmak üzere, X'in alabileceği değerler 0 ile 200 arasında olacaktır. Aşağıda gösterildiği gibi X, 0 ile 200 arasında sonsuz sayıda değer alabileceği için sürekli rassal bir değişkendir.

Bu aralıktaki herhangi bir nokta X ile gösterilen bir pilin ömrü olabilmektedir. Bu aralıkta sonsuz sayıda nokta olacağından, bu noktaların temsil ettiği değerler sayılamayacak sonsuzluktadır.

Aşağıda sürekli rassal değişken için bazı örnekler verilmiştir.

1. Bir kişinin boy uzunluğu.
2. Sınavda bir sorunun çözülme süresi.
3. Bir bebeğin ağırlığı.
4. Bir evin değeri (fiyatı).
5. Bir şişe sütün ağırlığı.

Daha çok kesikli rassal değişkenlere ve dağılımlarına ayrılan bu bölümü izleyen bölümde, sürekli rassal değişkenler ve dağılımları ayrıntılı olarak verilecektir.

SIRA SİZDE

1. Rassal değişken, kesikli rassal değişken ve sürekli rassal değişken kavramlarını açıklayınız. Kesikli ve sürekli rassal değişken için birer örnek veriniz.

2. Aşağıdaki rassal değişkenleri kesikli ve sürekli olarak sınıflayınız.

a.Bir sınıftaki öğrenci sayısı

b.Bir kutu biranın hacmi

c.Bir çiftlikteki inek sayısı

d. **Bir evin yaşı**
e. **Bir kitapta en az bir hata olan sayfa sayısı**
f. **Bir doktorun bir hastayı muayene süresi**

3. **Aşağıdaki rassal değişkenlerden hangilerinin kesikli hangilerinin sürekli olduğunu söyleyiniz.**
   a. **Bir banka şubesinde herhangi bir günde açılan hesap sayısı**
   b. **Bir maratonu koşma süresi**
   c. **Bir konser biletinin fiyatı**
   d. **Rassal seçilen bir kutudaki çürük yumurta sayısı**
   e. **Bir futbol maçının sonucu**
   f. **Rassal seçilen bir paketin ağırlığı**

KESİKLİ BİR RASSAL DEĞİŞKENİN OLASILIK DAĞILIMI

AMAÇ 2

Kesikli rassal değişkenlerin olasılık dağılımını oluşturabileceksiniz.

X kesikli bir rassal değişken olmak üzere, X'in olasılık dağılımı, X'in alabileceği değerlere göre olasılıklarının nasıl dağıldığını açıklamaktadır.

Kesikli Bir Rassal Değişkenin Olasılık Dağılımı, rassal değişkenin alabileceği değerle bunlara ait olasılıkların listesidir.

Aşağıdaki Örnek 5.1'de; kesikli bir rassal değişkenin olasılık dağılımı kavramını açıklanmaktadır.

ÖRNEK 1

Tablo 5.1'de verilmiş olan ailelerin sahip oldukları otomobil sayılarına ilişkin sıklık ve göreli sıklık dağılımı tekrar yazılacak olursa,

Otomobil Sayısı (x)	Sıklık	Göreli Sıklık	Olasılık (P(x))
0	30	0.015	0.015
1	470	0.235	0.235
2	850	0.425	0.425
3	490	0.245	0.245
4	160	0.080	0.080
	N = 2000	Toplam = 1.000	Σ P(x) = 1.00

Tablo 5.2 *Sahip Olunan Otomobil Sayılarına Göre Ailelerin Sıklık ve Göreli Sıklık Dağılımları.*

biçimindedir. X rassal seçilen bir ailenin sahip olduğu otomobil sayısı olmak üzere, X'in olasılık dağılımını yazınız.

ÇÖZÜM

Bir önceki bölümde (Bölüm 4) bir deneyden ya da örneklemden elde edilen göreli sıklıkların, yaklaşık olasılıklar gibi kullanılabildiğinden söz edilmişti. Ancak bir kütle için göreli sıklıkların bilinmesi, sonuçların gerçek (kuramsal) olasılıklarını vermektedir. Bu nedenle Tablo 5.2'de verilmiş olan göreli sıklıklar kullanılarak X kesikli rassal değişkeninin olasılık dağılımı doğrudan yazılabilmektedir (Tablo 5.2, son sütun).

Kesikli bir rassal değişkenin olasılık dağılımı aşağıdaki iki özelliği taşır.

1. x değişkenin alabileceği her bir değerin olasılığı 0 ile 1 arasında olup $0 \leq P(x) \leq 1$ biçiminde gösterilir.
2. x değişkenin alabileceği tüm değerlerin olasılıklar toplamı 1'dir ve $\Sigma P(x) = 1$ olarak gösterilir.

Olasılık Dağılımının İki Özelliği, kesikli bir rassal değişkenin olasılık dağılımı aşağıdaki iki özelliği taşır.

1. $0 \leq P(x) \leq 1$; x'in her değeri için
2. $\Sigma P(x) = 1$

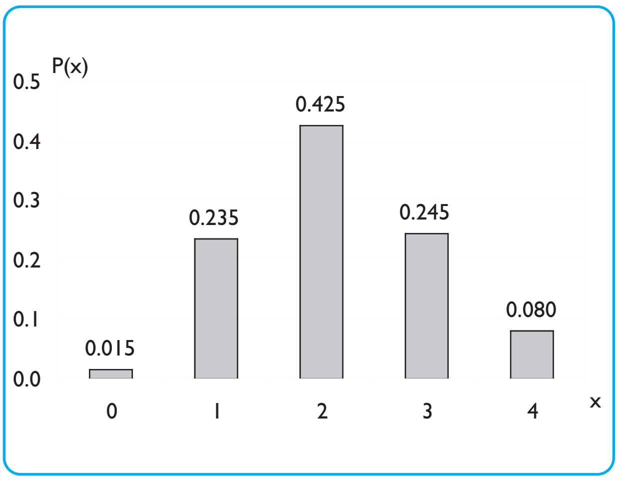
Bu iki özellik, bir olasılık dağılımının (kesinlikle) sağlamak zorunda olduğu iki koşul olarak da bilinmektedir. Bu nedenle yukarıdaki tabloda verilmiş tüm olasılıklar 0 ile 1 arasındadır ve olasılıklar toplamı da 1'dir. Bu durumda yazılmış bulunan P(x) değerleri X'in olasılık dağılımını oluşturur.

Tabloda, örneğin, rassal seçilen bir ailenin iki otomobili olma olasılığı 0.425 olup,

$$P(x = 2) = 0.425$$

biçiminde gösterilir. Ayrıca, seçilen bir ailenin ikiden çok otomobile sahip olma olasılığı sorulduğunda ise olasılıklar toplanmaktadır.

$$P(x > 2) = P(x = 3) + P(x = 4) = 0.245 + 0.080 = 0.325$$



Şekil 5.1 Tablo 5.2.'de Verilen Olasılık Dağılımının Grafiksel Gösterimi.

Kesikli bir rassal değişkenin olasılık dağılımı, matematiksel bir formül, bir tablo ya da bir grafik biçiminde gösterilebilmektedir. Aşağıdaki grafikte, yatay eksen X'in aldığı değerler, dikey eksense bu değerlere karşılık gelen olasılıklar (yükseklik) olmak üzere Tablo 5.2.'nin değerleri kullanılarak çizilmiştir. Bu tür grafiklere çizgi ya da hat grafiği adı verilir.

ÖRNEK 2

Aşağıdaki tablolarda bazı x değerleri ve bunlara ilişkin olasılıklar listelenmiştir. Bu tabloların her birinin geçerli bir olasılık dağılımı olup olmadığını araştırınız.

a)	x	P(x)	b)	x	P(x)	c)	x	P(x)
	0	0.08		2	0.25		7	0.70
	1	0.11		3	0.34		8	0.50
	2	0.39		4	0.28		9	-0.20
	3	0.27		5	0.13			

ÇÖZÜM

a) Bu tabloda verilmiş olan tüm olasılıklar 0 ile 1 arasında oldukları için olasılık dağılımının birinci koşulunu sağlamaktadır. Ancak olasılıklar toplamı (Σ P(x) = 0.08 + 0.11 + 0.39 + 0.27 = 0.85) 1 olmadığı için ikinci koşul sağlanmamakta, yani bu tablo, geçerli bir olasılık dağılımı göstermemektedir.
b) Bu tabloda verilmiş olan olasılıkların tümünün 0 ile 1 arasında olması ve olasılıklar toplamının (Σ P(x) = 0.25 + 0.34 + 0.28 + 0.13 = 1.00) 1 olması nedeniyle, iki koşulda sağlandığı için bu, geçerli bir olasılık dağılımı göstermektedir.
c) Bu tabloda verilmiş olan olasılıklar toplamı (Σ P(x) = 0.70 + 0.50 - 0.20 = 1.00) 1 olduğu halde, olasılıklardan bir tanesinin negatif olması nedeniyle ilk koşul sağlanmamakta ve bu tablo da geçerli bir olasılık dağılımı göstermemektedir.

ÖRNEK 3

Eski verilerden yararlanılarak bir makinenin birer hafta süresince yaptığı arıza sayıları listelenerek aşağıda verilmiştir.

Haftalık arıza	0	1	2	3
Olasılık	0.15	0.20	0.35	0.30

a) Olasılık dağılımını grafiksel olarak gösteriniz.
b) Bu makinenin verilen bir hafta içerisinde aşağıdaki arıza sayılarına ilişkin olasılıkları bulunuz.
i) Kesinlikle iki
ii) Sıfır iki arası
iii) Birden çok
iv) En çok bir

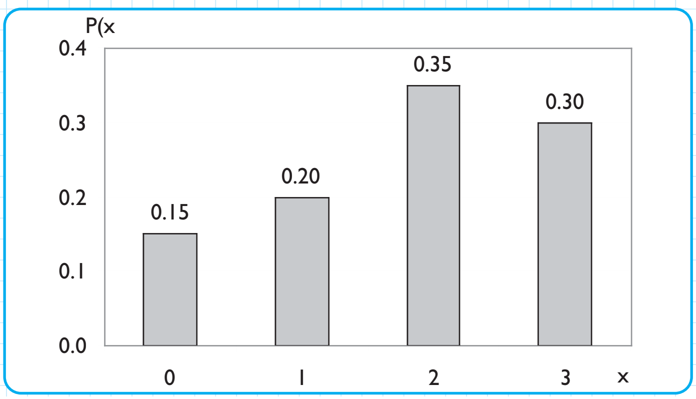
ÇÖZÜM

Yukarda verilmiş olan bilgilerden yararlanılarak, X: verilen bir hafta içerisinde makinenin arıza sayılarını göstermek üzere olasılık dağılımı,

x	P(x)
0	0.15
1	0.20
2	0.35
3	0.30
	Σ P(x) = 1.00

Tablo 5.3 *Arıza Sayılarının Olasılık Dağılımı.*

biçiminde yazılır.
a) Olasılık dağılımı bilgilerinden yararlanarak olasılık dağılım grafiği aşağıdaki biçimde çizilir.



Şekil 5.2 *Tablo 5.3.'deki Olasılık Dağılımının Grafiği.*

b) Yukarıda verilmiş olan Tablo 5.3'den yararlanarak istenen olasılıklar bulunur.

i) Kesinlikle iki arıza olma olasılığı;

P (Kesinlikle iki arıza) = $P(x = 2) = 0.35$

ii) Sıfır-iki arıza olma olasılığı (0, 1 ve 2 arıza durumlarının toplamıdır).

P (0 - 2 arıza) = $P(0 \leq x \leq 2) = P(x = 0) + P(x = 1) + P(x = 2)$
$= 0.15 + 0.20 + 0.35 = 0.70$

iii) Birden çok arıza olma olasılığı (2 ve 3 arıza durumlarının toplamıdır).

P (Birden çok arıza) = $P(x > 1) = P(x = 2) + P(x = 3)$
$= 0.35 + 0.30 = 0.65$

iv) En çok bir arıza olma olasılığı (0 ve 1 arıza durumlarının toplamıdır)

P (En çok bir arıza) = $P(x \leq 1) = P(x = 0) + P(x = 1)$
$= 0.15 + 0.20 = 0.35$

ÖRNEK 4

Yapılan bir araştırmaya göre; üniversite öğrencilerinin % 60'ının matematik derslerini sevmedikleri (fobi) ve sınavlarından korktukları elde edilmiştir. x matematik derslerini sevmeyen öğrenci sayısını göstermek üzere, bu gruptan rassal seçilen iki öğrenci için deneyin olasılık dağılımını yazınız.

ÇÖZÜM

Deney için tanımlanması gereken iki olay;

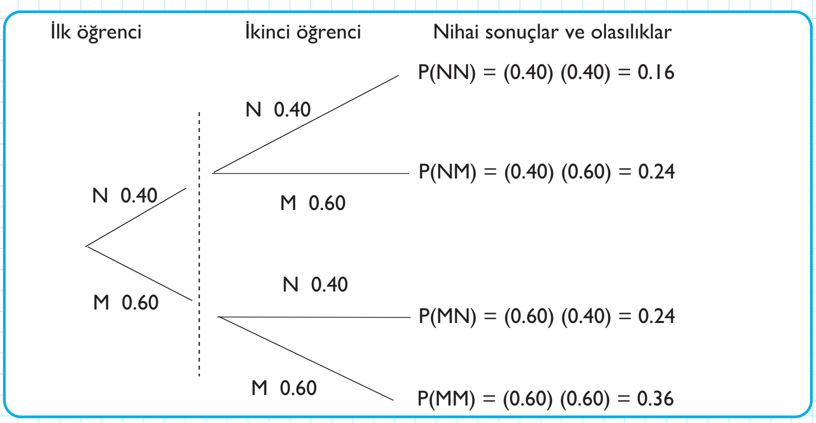
N = Seçilen öğrencide matematik fobisi yok
M = Seçilen öğrencide matematik fobisi var

biçimindedir.

Aşağıdaki şekilden de görüleceği gibi bu deneyin dört olası sonucu bulunmaktadır (NN - her iki öğrencide de matematik fobisi yok , NM - ilk öğrencide matematik fobisi yok ikincide var, MN – ilk öğrencide matematik fobisi var ikincide yok, MM - her iki öğrencide de matematik fobisi var). Yukarıda verilen bilgilerden P (M) = 0.60 olduğu bilinmektedir ve P (N) = 1 - P (M) = 1 - 0.60 = 0.40 olacağı kolaylıkla görülür. Bu durumda deneyin sonuçları,

$P(x = 0) = P(NN) = 0.16$
$P(x = 1) = P(NM \text{ ya da } MN) = P(NM) + P(MN) = 0.24 + 0.24 = 0.48$
$P(x = 2) = P(MM) = 0.36$

biçiminde ifade edilir.



***Şekil 5.3** Ağaç Diyagramı.*

Yukarıda verilmiş olan olasılık değerlerinden, olasılık dağılımı tablo biçiminde de yazılabilir.

x	P(x)
0	0.16
1	0.48
2	0.36
$\sum P(x) =$	1.00

Tablo 5.4 *Matematik Fobisi Bulunan Öğrencilerin Olasılık Dağılımı.*

SIRA SİZDE

1. Kesikli bir rassal değişkenin olasılık dağılımının ne olduğunu açıklayınız. Kesikli bir rassal değişkenin olasılık dağılımının hangi üç farklı yolla ifade edildiğini söyleyiniz.

2. Kesikli bir rassal değişkenin olasılık dağılımının iki özelliğini (koşullarını) kısaca açıklayınız.

3. Aşağıdaki üç tabloda bir dizi x değeri ve bunlara ilişkin olasılıklar verilmiştir. Bunlardan hangileri olasılık dağılımının koşullarını sağlamaktadır.

a) x	P(x)	b) x	P(x)	c) x	P(x)
5	-0.36	1	0.16	0	0.15
6	0.48	2	0.24	1	0.00
7	0.62	3	0.49	2	0.35
8	0.24			3	0.50

KESİKLİ BİR RASSAL DEĞİŞKENİN ORTALAMASI VE STANDART SAPMASI

AMAÇ 3

Kesikli rassal değişkenlerin ortalama ve standart sapmasını hesaplayabileceksiniz.

Kesikli Bir Rassal Değişkenin Ortalaması

Kesikli bir rassal değişkenin ortalaması μ ile gösterilir ve bu değer aynı zamanda olasılık dağılımının da ortalamasıdır. X kesikli değişkeninin ortalamasına, beklenen değer (expected value) adı verilmekte ve E(x) biçiminde gösterilmektedir. Kesikli bir rassal değişkenin ortalaması, deneyin çok kez tekrarlanması durumunda ortaya çıkacak sonuçların ortalama değeridir. Örneğin bir otomobil galerisinin haftalık ortalama satışı 2.4 otomobildir ifadesinde, bu galerinin gelecek haftada kesinlikle 2.4 otomobil satacağı anlamı çıkarılmamalıdır. Uzun haftaların kayıtlarından (araba satışları bazı haftalar sıfır, bazen bir, bazen iki ya da daha çok olabilir) galerinin ortalama satış miktarı 2.4 otomobil olarak bulunmuştur.

Kesikli bir rassal değişkenin ortalaması, x değişken değerlerinin kendilerine karşılık gelen olasılıklarla çarpılıp toplanması işlemiyle hesaplanır.

Kesikli bir rassal değişkenin ortalaması, her x ile buna karşı gelen P(x) olasılıklarının çarpımlarının toplamı alınarak elde edilir.

$$\mu = \Sigma\, x \;\; P(x) = E(x)$$

ÖRNEK 5 ***Örnek 5.3'de bir makinenin bir haftada yaptığı arıza sayılarına ilişkin aşağıdaki olasılık dağılımının ortalamasım bulunuz.***

x	P(x)
0	0.15
1	0.20
2	0.35
3	0.30

ÇÖZÜM

Yukarıda belirtildiği gibi bu makinenin bir hafta süresince yapacağı ortalama arıza sayısını (beklenen değer) bulmak için, X 'in aldığı değerlerle bu değerlere karşılık gelen olasılıkların çarpımlarının toplanması gerekmektedir. Bu yolla oluşturulan tablo aşağıdadır.

***Tablo 5.5** Arıza Sayılarına İlişkin Olasılık Dağılımının Ortalamasının Hesaplanması.*

x	P(x)	x P(x)
0	0.15	0(0.15) = 0.00
1	0.20	1(0.20) = 0.20
2	0.35	2(0.35) = 0.70
3	0.30	3(0.30) = 0.90
		Σ x P(x) = 1.80

Elde edilen $\mu = \Sigma\, x\ P(x) = 1.80$ değerinin anlamı, bu makinenin incelenen sürede, haftada 1.8 kez arıza yapmasının beklendiğidir. Ancak bulunan bu değer ortalama (kuramsal) bir değer olup, gelecek haftalarda da makine bazen hiç arıza yapmayacak, bazen 1, bazen 2 ya da daha çok kez arıza yapacaktır.

Kesikli Bir Rassal Değişkenin Standart Sapması

Kesikli bir rassal değişkenin standart sapması, olasılık dağılımının yayılmasının (saçılmasının) bir ölçüsü olup σ ile gösterilir. Standart sapma değerinin büyük olması, x değerlerinin ortalama etrafında geniş bir aralıkta değerler aldığını gösterirken, küçük standart sapma değeri bu aralığın dar olduğunu, gözlenen x değerlerinin ortalamaya çok yakın değerler aldığını ifade eder. Kesikli bir rassal değişkenin standart sapması,

$$\sigma = \sqrt{\sum (x-\mu)^2\ P(x)} = \sqrt{\sum x^2\, P(x) - \mu^2}$$

eşitliğinden hesaplanır.

Kesikli bir rassal değişken olan X 'in olasılık dağılımının yayılma ölçüsü olan standart sapma,

$$\sigma = \sqrt{\sum x^2\, P(x) - \mu^2}$$

eşitliğinden hesaplanmaktadır.

İstatistiksel değerlendirmelerde, kesikli bir rassal değişkenin standart sapması olduğu kadar (hatta daha çok) standart sapmanın karesi alınarak bulunan ve σ^2 ile gösterilen varyans değeri de kullanılmaktadır.

ÖRNEK 6

Bir elektrik firması bilgisayar parçaları üreterek satışa sunmaktadır. Üretilen her parça tek tek kalite kontrolünden geçirildikten sonra piyasaya sürülmektedir. Ancak tüm bu titiz kontrollere karşın, az sayıda da olsa bazı bozuk (arızalı) parçalar da gözden kaçabilmektedir. x arızalı parça sayısını göstermek üzere, 400 parçalık bir sevkiyatın olasılık dağılımı aşağıda verilmiştir. x değerinin standart sapmasını bulunuz.

x	0	1	2	3	4	5
P(x)	0.02	0.20	0.30	0.30	0.10	0.08

ÇÖZÜM

x ‘in standart sapmasının elde edilmesi için gerekli hesaplamalardan aşağıdaki tablo oluşturulmuştur.

x	P(x)	x P(x)	x^2	x^2 P(x)
0	0.02	0.00	0	0.00
1	0.20	0.20	1	0.20
2	0.30	0.60	4	1.20
3	0.30	0.90	9	2.70
4	0.10	0.40	16	1.60
5	0.08	0.40	25	2.00
		Σ x P(x) = 2.50		Σ x^2 P(x) = 7.70

Tablo 5.6 *Standart Sapma İçin Gerekli Hesaplamalar.*

x rassal değişkeninin standart sapmasının bulunması için şu adımlar izlenir :

Adım 1: Kesikli rassal değişkenin ortalaması hesaplanır.

$\mu = \Sigma x\ P(x) = 2.50$
(400 parça arasında ortalama arızalı parça sayısı)

Adım 2: $\Sigma\ x^2\ P(x)$ değerinin hesaplanması.
x rassal değişkeninin alabileceği değerlerin kareleri alınıp, olasılık değerleri olan P(x) değeriyle çarpılması ve çarpımlarının toplanmasıyla elde edilmektedir.

$\Sigma\ x^2\ P(x) = 7.70$

Adım 3: Bulunan μ ve $\Sigma\ x^2\ P(x)$ değerlerinin formülde yerine konarak x ‘in standart sapması hesaplanır.

$$\sigma = \sqrt{\Sigma\, x^2\, P(x) - \mu^2} = \sqrt{(7.70) - (2.5)^2} = \sqrt{1.45} \cong 1.20$$

Sonuç olarak 400 parçadan oluşan sevkiyatın, 1.2 standart sapma değeriyle ortalama 2.5 tanesi arızalıdır.

ÖRNEK 7

Yeni bir mutfak aleti üreterek piyasaya sürmeyi planlayan bir firmanın finansal kaynaklar bölümü; bu ürüne yüksek talep olursa yılda 4.5 trilyon TL, normal talep olursa 1.2 trilyon TL kar edeceklerini, düşük talep olursa yılda 2.3 trilyon TL zarar edeceklerini hesaplamıştır. Bu üç talep beklentisine ilişkin olasılıklar sırasıyla 0.32, 0.51 ve 0.17 olduğu göre,

a) x yıllık karı göstermek üzere, x‘in olasılık dağılımını yazınız.
b) x ‘in ortalama ve standart sapma değerlerini bulunuz.

ÇÖZÜM

a) Firmanın zarar etmesi durumu negatif kar olarak gösterilerek olasılık dağılımı aşağıdaki gibi elde edilir.

x	P(x)
4.5	0.32
1.2	0.51
-2.3	0.17

b) x rassal değişkeninin ortalama ve standart sapma değerlerinin hesaplanmasında kullanılacak bilgiler aşağıda Tablo 5.7'de verilmiştir.

Tablo 5.7 *Ortalama ve Standart Sapma İçin Gerekli Bilgiler.*

x	P(x)	x P(x)	x^2	x^2 P(x)
4.5	0.32	1.440	20.25	6.4800
1.2	0.51	0.612	1.44	0.7344
-2.3	0.17	-0.391	5.29	0.8993
		S x P(x) = 1.661		S x^2 P(x) = 8.1137

Bu bilgilerden $\mu = \Sigma$ x P(x) = 1.661 trilyon TL ortalama kar ve

$$\sigma = \sqrt{\sum x^2 P(x) - \mu^2} = \sqrt{(8.1137) - (1.661)^2} = 2.314 \text{ trilyon TL.}$$

standart sapma değeri bulunur.

Standart Sapmanın Yorumu

Kesikli bir rassal değişkenin standart sapması da öteki veri kümelerine benzer biçimde yorumlanır. Örneğin birden büyük bir değer olmak üzere Chebyshev teoremine göre eğri altında kalan alanın en az $[\,1 - (1 / k^2)\,]$. kadarı ortalama etrafında ± k standart sapma sınırları arasında kalmaktadır. Örneğin k = 2 alınırsa, toplam alanın % 75'i $\mu - 2\sigma$ ile $\mu + 2\sigma$ sınırları arasında yer alır. Bu teorem gereğince, Örnek 6' da $\mu = 2.50$ ve $\sigma = 1.20$ olarak elde edilen değerlerden,

$$\mu - 2\sigma = 2.50 - 2(1.20) = 0.10$$
$$m + 2\sigma = 2.50 + 2(1.20) = 4.90$$

sonuçları bulunur ve 400 parçalık sevkiyatların en az %75'inin arızalı parça sayısının 0.10 ile 4.90 arasında olacağı söylenir.

SIRA SİZDE

1. Aşağıda verilmiş bulunan olasılık dağılımlarının ortalama ve standart sapmalarını bununuz.

a) x	P(x)	b) x	P(x)
0	0.12	6	0.36
1	0.27	7	0.26
2	0.43	8	0.21
3	0.18	9	0.17

2. x, bir kitaptan rassal seçilmiş bir sayfadaki yazım hatalarının sayısı olmak üzere, aşağıda verilmiş olan olasılık dağılımının ortalama ve standart sapmasını bulunuz.

x	0	1	2	3	4
P(x)	0.73	0.16	0.06	0.04	0.01

3. Elektrik malzemeleri satılan bir mağazada yapılan incelemede bir günde satılan elektrik prizi sayısına ilişkin olasılık dağılımı aşağıdadır. Olasılık dağılımının ortalama ve standart sapmasını bulunuz, bulduğunuz ortalama değerini yorumlayınız.

Satılan priz	0	1	2	3	4	5	6
Olasılık	0.05	0.12	0.23	0.30	0.16	0.10	0.04

FAKTÖRİYELLER VE KOMBİNASYONLAR

AMAÇ 4

Faktöriyel kavramını açıklayabilecek, kombinasyon hesaplayabileceksiniz.

Faktöriyeller

“ ! ” sembolü faktöriyel işareti olup, verilen değerden 1’e kadar tüm pozitif tamsayıların çarpımından oluşur. Örneğin 7!, yedi faktöriyel olarak okunur ve 7’den 1’e kadar tüm pozitif tam sayıların çarpımıne eşittir.

Faktöriyeller, n! (n faktöriyel) n’den 1’e kadar tüm pozitif tam sayıların çarpımını ifade eder ve

$n! = n\,(n-1)\,(n-2)\ldots..3.\ 2.\ 1$

biçiminde gösterilir. Bu tanım gereğince 0! = 1’dir.

ÖRNEK 8

7! değerini bulunuz.

ÇÖZÜM

7! değerini bulmak için 7’den 1’e kadar tüm tamsayıların çarpılması gerekir.

$7! = 7 . 6 . 5 . 4 . 3 . 2 .1 = 5.040$

ÖRNEK 9

10! değerini bulunuz.

ÇÖZÜM

Burada da yine 1 ‘den 10 ‘a kadar tüm tamsayıların çarpımından 10! değeri elde edilir.

$10! = 10 . 9 . 8 . 7 . 6 . 5 . 4 . 3 . 2 .1 = 3.628.800$

ÖRNEK 10

(12 – 4) ! değerini bulunuz.

ÇÖZÜM

Burada ilk olarak parantez içindeki işlem yapılır ve daha sonra faktöriyel değeri bulunur.

$(12 - 4)! = 8! = 8 . 7 . 6 . 5 . 4 . 3 . 2 .1 = 40.320$

ÖRNEK 11

(5 – 5) ! değerini bulunuz.

ÇÖZÜM

Yukarıda da belirtilmiş olduğu gibi burada da önce parantez içerisindeki işlem yapılır ve daha sonra farktöriyel değeri hesaplanır.

$(5 - 5)! = 0! = 1$

ÖRNEK 12 ***15 ! değerini bulunuz.***

ÇÖZÜM

$15! = 15 \,.\, 14 \,.\, 13 .\, ...\, 2 \,.\, 1 = 1.307.674.368.000$

(Not: Faktöriyel değerlerinin kolay bulunmasını sağlayan hazır tablolar geliştirilmiş olup. bu amaçla bazı hesap makinelerinde ! fonksiyon anahtarı (tuşu) bulunmaktadır.)

Kombinasyonlar

Bir süre sonra, bir grup içerisinden az sayıda birimin çekilmesi konusu incelenecektir. Örneğin bir öğrenci dört soruluk bir sınavda iki soruyu cevaplayacaktır. Ya da bir fakültedeki 20 profesör arasından üç kişilik bir komite seçilecektir. Bu örneklerde ortak sorun, seçimin kaç farklı yolla yapılabileceğidir. Örneğin, sınava giren öğrenci dört soru arasından iki soruyu altı farklı şekilde seçebilir. Bunlar (1 ve 2) (1 ve 3) (1 ve 4) (2 ve 3) (2 ve 4) (3 ve 4) biçiminde listelenebilir. Listedeki tüm olası seçimlere bir kombinasyon denmektedir. Bu altı kombinasyonda farklı sorular bulunmaktadır. Kombinasyonlarda seçilme sırası önemli değildir. Yukarıdaki örnekler için (1 ve 2) ile (2 ve 1) soru seti eşdeğer (aynı) dir.

Kombinasyonların Gösterimi, Kombinasyonlar; n eleman arasından x tanesinin seçilme yolları sayısını vermekte ve toplam kombinasyon sayısı,

$\binom{n}{x}$ ya da C_n^x

biçiminde gösterilerek " n eleman arasından her seferinde x tanesinin seçilmesinde kombinasyon sayısı" olarak okunmaktadır.

x tane elemanın, arasından seçilecek toplam eleman sayısı n olarak düşünüldüğünde;

ifadesinden yararlanılmaktadır.

n = toplam eleman sayısı

$\binom{n}{x}$ = n eleman arasında her seferinde x eleman seçilmesinde kombinasyon sayısı

x = her seferinde seçilecek eleman sayısı

Kombinasyonların Sayısı, n tane farklı eleman arasından x tanesinin seçimindeki kombinasyon sayısı

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!}$$

formülünden bulunmaktadır. Burada n! , x! , (n - x) ! faktöriyellerdir.
Kombinasyon formülünde,

$$n! = n\,(n-1)\,(n-2)........\,2\,.\,1$$
$$x! = x\,(x-1)\,(x-2)........\,2\,.\,1$$
$$(n-x)! = (n-x)\,(n-x-1)\,(n-x-2)........\,2\,.\,1$$

olup, burada n değeri x değerinden büyük ya da en azından eşit olmak zorundadır. Aksi durumda x elemanın n 'den küçük bir gruptan seçilmesi söz konusu olmayacaktır.

ÖRNEK 13

Yukarıda verilmiş olan dört soru arasından iki tanesinin seçilmesi deneyinde, farklı seçim sayısını kombinasyon formülünü kullanarak bulunuz.

ÇÖZÜM

Bu örnekte

n = toplam soru sayısı = 4
x = seçilecek soru sayısı = 2

olduğu için, bir öğrencinin seçebileceği farklı küme sayısı,

$$\binom{4}{2} = \frac{4!}{2!\,(4-2)!} = \frac{4!}{2!\,2!} = \frac{4\,.\,3\,.\,2\,.\,1}{1\,.\,2\,.\,1\,.\,2} = 6$$

olarak bulunur.

ÖRNEK 14

Beş kişi arasından üç kişinden oluşacak bir jüri, kaç farklı yolla seçilebilir?

ÇÖZÜM

Burada da n = 5 ve x = 3 olduğu için kombinasyon formülünden,

$$\binom{5}{3} = \frac{5!}{3!\,(5-3)!} = \frac{5!}{3!\,2!} = \frac{120}{6\,.\,2} = 10$$

elde edilir. Eğer A, B, C, D ve E ile ifade edilecek olunursa, olası jüriler; ABC, ABD, ABE, ACD, ACE, ADE, BCD, BCE, BDE, CDE biçiminde gösterilir.

SIRA SİZDE

1. **Bir üniversitenin İstatistik Bölümünde 15 öğretim üyesi görev yapmaktadır. Bunlar arasından rassal olarak iki tanesi, fakülte komitesine üye olarak seçilecektir. Kaç farklı seçim yapılacağını bulunuz.**

2. **Bir ilçenin kaymakamı, gelecek hafta, ilçesinde bulunan 12 ilköğretim okulundan 3 tanesini ziyaret edecektir. Kaç farklı olası seçim yapacağını bulunuz.**

3. **Bir yatırımcı, gıda sektöründe faaliyet gösteren ve hisseleri menkul kıymetler borsasında işlem gören 20 firmadan 5 tanesinin hisse senetlerinden almak istemektedir. Yatırımcı kaç farklı seçim yapabilecektir.**

BİNOM (İKİ TERİMLİ) OLASILIK DAĞILIMI

AMAÇ 5

Kesikli rassal değişkenlerin önemli dağlımlarından Binom dağılımını kullanarak ilgili olasılıkları hesaplayabilecek, dağılımın ortalama ve standart sapmasını hesaplayabileceksiniz.

Binom olasılık dağılımı, X'in kesikli rassal değişken olması durumunda en yaygın kullanılan dağılımlardan biridir. Binom olasılık dağılımı, n tekrarlı bir deneyde x kez istenen sonuç gelmesi durumunda, olasılıkların bulunması amacıyla kullanılmaktadır. Örneğin bir fabrikada üretilen TV setlerinin arızalı olma olasılığı % 5 olarak verilmişse bu fabrikada üretilen TV setlerinden üç tanesinin seçilmesi durumunda, bunlardan bir tanesinin arızalı olma olasılığının bulunmasında kullanılmaktadır.

Binom olasılık dağılımının uygulanabilmesi için X değişkeninin iki sonuçlu (kesikli) bir rassal değişken olması gerekir. İki sonuçlu (kesikli) rassal değişkenin anlamı, deneyin her tekrarından sonra bu iki sonuçtan birinin ortaya çıkmasıdır.

Binom dağılımı, aşağıda ayrıntılı bir biçimde verilecek dört koşulu sağlayan binom deneylerine uygulanmaktadır ki bu deneydeki her tekrara deneme, Bernoulli denemesi ya da sınaması (trial - Bernoulli trial) adı verilmektedir. Örneğin deney, bir paranın bir kez atılması olarak tanımlanırsa ve bu deney 10 kez tekrarlanırsa buradaki her tekrara (atış) bir deneme adı verilir ve bu deneyde toplam 10 deneme yapılmış olur.

Binom Deneyi

Eğer bir deney aşağıdaki dört koşulu sağlıyorsa bu deneye binom deneyi denmektedir.

1. n tane özdeş deneme vardır. Yani verilen deney n kez özdeş (aynı) koşullarda tekrarlanmaktadır.
2. Her denemenin sadece ve sadece iki sonucu vardır. Bu sonuçlara genellikle başarı ya da başarısızlık denmektedir.
3. p başarı olasılığı, q ise başarısızlık olasılığı olmak üzere $p + q = 1$ dir. p ve q olasılıkları her deneme için aynıdır.
4. Bir denemenin sonucu öteki denemenin sonucunu etkilememektedir. Yani denemeler bağımsızdır.

Bir Binom Deneyinin Koşulları: Bir binom deneyi aşağıdaki dört koşulu sağlamak zorundadır.

1. n tane özdeş deneme olmalıdır.
2. Her denemenin sadece iki olası sonucu olmalıdır.
3. İki sonucun olasılıkları hep aynı olmalıdır.
4. Denemeler birbirinden bağımsız olmalıdır.

Yukarıda da belirtildiği gibi, bir denemenin, başarı ve başarısızlık olarak ifade edilen iki sonucu vardır. Ancak burada başarı, gerçek başarı ya da arzulanan bir sonuç anlamına gelmemektedir. Aynı biçimde başarısızlık da istenmeyen, bir durum olmayıp, başarı ve başarısızlık sadece, iki farklı sonucu ifade etmek için kullanılmaktadır. Sonuç olarak, karşılaşılan özel problemde istenen sonuca başarı, istenmeyen sonucaysa başarısızlık denmektedir.

ÖRNEK 16

Hatasız bir paranın 10 kez atıldığı bir deney, bir binom deneyi midir?

ÇÖZÜM

Aşağıda ayrıntılı bir biçimde gözden geçirileceği gibi, hatasız bir paranın 10 kez atılması deneyi gerekli dört koşulu da sağladığı için bir binom deneyidir.

1. Burada aynı türden (özdeş) 10 deneme (atış) vardır ve 10 denemenin tümü de de aynı koşullardadır.
2. Her denemenin yazı ve tura olmak üzere iki olası sonucu vardır ve burada yazı gelmesi başarı, tura gelmesi ise başarısızlık olarak nitelenmektedir.
3. Herhangi bir atışta yazı (başarı) gelme olasılığı $1 / 2$, tura (başarısızlık) gelme olasılığı da $1 / 2$ olup,

$$p = P(Y) = 1 / 2 \quad \text{ve } q = P(T) = 1 / 2$$

olasılıklar toplamı 1 'dir ve bu değerler bütün denemelerde aynıdır.

4. Denemeler birbirinden bağımsızdır ve herhangi bir denemede yazı gelmesi, izleyen denemenin sonucunu etkilememektedir.

Sonuç olarak bir paranın 10 kez atılması deneyi bir binom deneyidir.

ÖRNEK 17

Bir firma tarafından üretilen TV setlerinin % 5'inin kusurlu (arızalı) olduğu bilinmektedir. Bu firmaca üretilen TV setlerinden rassal üç tanesinin seçilmesi ve kalite kontrol uzmanlarınca dikkatli bir biçimde incelenmesi deneyi, bir binom deneyi midir?

ÇÖZÜM

1. Bu örnekte üç tane aynı türden (özdeş) deneme vardır.
2. Her denemede (arızalı ve arızasız) iki sonuç vardır ve bu sonuçlar başarı ve başarısızlık olarak değerlendirilmektedir.
3. TV setinin arızalı (başarı) olma olasılığı $p = 0.05$, arızasız (başarısız) olma olasılığı $q = 0.95$ olup olasılıklar toplamı 1'dir.
4. Denemeler birbirinden bağımsızdır. Çünkü incelenen herhangi bir TV setinin arızalı olması daha sonra incelenecek olanların arızalı ya da arızasız olmasını etkilememektedir.

Bu dört koşulun sağlanması nedeniyle bu deney de bir binom deneyidir.

Binom Olasılık Dağılımı ve Binom Formülü

Bir binom deneyinde n denemede elde edilen başarı sayısı X ile ifade ediliyorsa, X rassal değişkenine binom rassal değişkeni, dağılımınaysa binom olasılık dağılımı ya da kısaca binom dağılımı denmektedir. Binom dağılımı, n denemeden x başarılı sonucun elde edildiği binom deneyinde olasılık hesaplamak amacıyla kullanılmaktadır. Burada X 'in kesikli rassal değişken olduğu unutulmamalıdır. Nitekim yukarıda incelenmiş olan TV seti örneğinde başarılı sonuç sayısı 0, 1, 2 ve 3'den bir tanesi olacaktır.

Binom Formülü, Bir binom deneyinde, n denemeden x tane başarılı sonuç elde edilmesinin olasılığı, aşağıdaki binom formülüyle bulunmaktadır.

Burada; n = toplam deneme sayısı
p = başarılı sonuç elde edilme olasılığı
$q = 1 - p$ = başarısız sonuç elde edilme olasılığı
x = başarılı sonuç sayısı
$n - x$ = başarısız sonuç sayısıdır.

$$P(x) = \binom{n}{x} p^x \cdot q^{n-x}, \; x = 0,1, \ldots, n$$

ÖRNEK 18 ***Yukarıda verilen ve arızalı olma olasılığı % 5 olan TV seti örneğinde, rassal seçilmiş olan 3 TV setinden sadece bir tanesinin arızalı olma olasılığı nedir?***

ÇÖZÜM

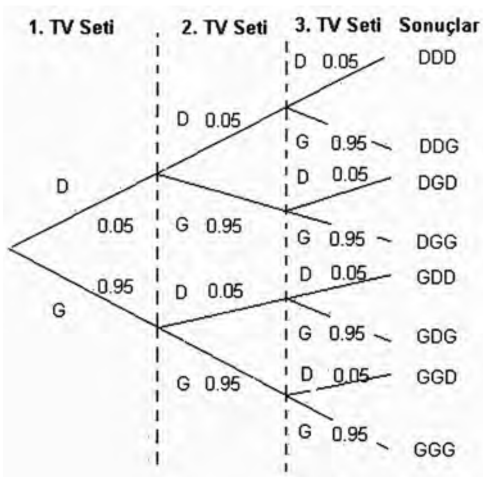
Burada ilk olarak,

D = Seçilmiş bir TV setinin arızalı olması
G = Seçilmiş bir TV setinin arızalı olmaması

olayları tanımlanır. Aşağıdaki ağaç diyagramından da görüleceği gibi burada ortaya çıkacak 8 olası sonuçtan sadece üç tanesiyle ilgilenilmektedir. Bunlar,

DGG, GDG, GGD

dir.



***Şekil 5.4** Üç TV setine ilişkin ağaç diyagramı.*

Burada rassal seçilen bir TV setinin arızalı olma olasılığı p = P(D) = 0.05 ve arızasız olma olasılığı da q = P(D) = 0.95 dir. Çok geniş bir kütleden seçildiği düşünüldüğünde, seçimler birbirinden bağımsız olacaktır. Burada ilgilenilen üç durumun (sadece bir tanesinin arızalı olması) olasılıkları, eski bilgilerden yararlanılarak kolaylıkla bulunur.

$$P(DGG) = P(D)\ P(G)\ P(G) = (0.05)\ (0.95)\ (0.95) = 0.0451$$
$$P(GDG) = P(G)\ P(D)\ P(G) = (0.95)\ (0.05)\ (0.95) = 0.0451$$
$$P(GGD) = P(G)\ P(G)\ P(D) = (0.95)\ (0.95)\ (0.05) = 0.0451$$

Burada DGG, D, G ve G olaylarının arakesiti ya da bu üç olasılığın P(DGG) biçiminde bileşik olasılığıdır. Bu sonucun ortaya çıkma olasılığı (bileşik olasılık), çarpma kuralıyla elde edilmiştir.

Öteki iki durum için de benzer yorumlar yapılabilir. Sonuç olarak istenen olasılık,

$$\begin{aligned} P\,(3 \text{ TV setinden 1 tanesinin arızalı olması}) &= P\,(DGG \text{ ya da } GDG \text{ ya da } GGD) \\ &= P\,(DGG) + P\,(GDG) + P\,(GGD) \\ &= 0.0451 + 0.0451 + 0.0451 \\ &= 0.1353 \end{aligned}$$

olarak bulunur.

Şimdi ise aynı sonuç binom formülü ile bulunacaktır.

n = toplam deneme sayısı = 3
x = başarılı sonuç sayısı = 1
$n - x$ = başarısız sonuç sayısıdır = 3 – 1 = 2
p = P (başarılı) = 0.05
q = P (başarısız) = 0.95

olmak üzere binom formülünden,

3 denemeden 1 başarılı sonuç elde etme olasılığı
Başarılı sonuç sayısı
Başarısız sonuç sayısı

$$P(x = 1) = \binom{3}{1} (0.05)^1 \ (0.95)^2 = 3(0.05)(0.9025) = 0.1353$$

Başarı olasılığı
Başarısızlık olasılığı

elde edilir.

ÖRNEK 19

Yüksek kalitede hizmet sunan bir kargo şirketinin, paketlerinden sadece % 2'sini belirlenen sürede yerine ulaştıramadığı bilinmektedir. Bir müşteri 10 tane paketi bu kargo firmasına getirerek, belirli bir sürede üzerlerinde yazılı adreslere ulaştırılmasını istemiştir.

a) Bu paketlerden bir tanesinin belirlenen sürede yerine ulaşmama olasılığı nedir?

b) Bu paketlerden en çok bir tanesinin belirlenen sürede yerine ulaşmama olasılığı nedir?

ÇÖZÜM

Burada paketin yerine ulaşmaması başarı, ulaşmasıysa başarısızlık olarak tanımlanırsa,

n = toplam paket sayısı = 10
p = P (başarılı) = 0.02
q = P (başarısız) = 1 – p = 1 – 0.02 = 0.98

değerleri yazılır.

a) Sadece bir paketin ulaşmaması durumuyla ilgilenildiğinden,

x = başarılı sonuç sayısı = 1
$n - x$ = başarısız sonuç sayısıdır = 10 – 1 = 9

değerleri de kullanılarak istenen olasılık bulunur.

$$P(x = 1) = \binom{10}{1} (0.02)^1 \ (0.98)^9 = \frac{10!}{1! \ 9!} (0.02)^1 \ (0.98)^9$$

$$= (10)(0.02)(0.8337) = 0.1667$$

b) En çok bir paketin yerine ulaşmaması durumuyla ilgilenildiğindeyse x = 0 ve x = 1 olmaktadır. Bu durumda

$$P(x \leq 1) = P(x = 0) + P(x = 1)$$

$$= \binom{10}{0}(0.02)^{0}\ (0.98)^{10} + \binom{10}{0}\ (0.02)^{1}\ (0.98)^{9}$$

$$= (1)(1)(0.8171) + (10)(002)(0.8337) = 0.9838$$

sonucu elde edilmektedir.

ÖRNEK 20

Bir araştırma sonucunda 6 yaşından küçük çocuklu, evli kadınların % 60'ının ev hanımı olmadıkları bulunmuştur. Bu gruptan evli üç kadın rassal olarak seçilmiştir. x, ev hanımı olmayan kadın sayısını göstermek üzere , üç kadının da ev hanımı olmama olasılığını bulunuz, x rassal değişkeninin olasılık dağılımını yazarak grafiğini çiziniz.

ÇÖZÜM

x, 3 birimlik kadın örnekleminde ev hanımı olmayan kadın sayısı, n - x ise ev hanımı kadın sayısı olmak üzere,

n = toplam kadın sayısı = 3
p = P (ev hanımı olmayan kadın) = 0.60
q = P (ev hanımı kadın) = 1 - p = 1 - 0.60 = 0.40

bilgilerinden yararlanarak istenen olasılık bulunur. Burada x rassal değişkeni 0, 1 , 2 ve 3 değerlerini alacak ve istenen olasılık , bu olasılık değerleri arasından seçilecektir.

$$P(x = 0, 1, 2, 3) = P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3)$$

$$= \binom{3}{0}(0.60)^0\,(0.40)^3 + \binom{3}{1}(0.60)^1\,(0.40)^2 + \binom{3}{2}(0.60)^2\,(0.40)^1$$

$$+ \binom{3}{3}(0.60)^3\,(0.40)^0$$

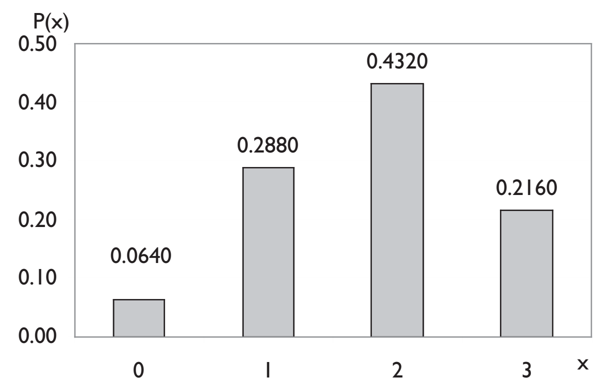
$$= 0.0640 + 0.2880 + 0.4320 + 0.2160 = 1$$

Soruda istenen olasılık P(x = 3) = 0.2160 dır.

Yukarıda elde edilen olasılık değerlerinden yararlanarak x 'in olasılık dağılımını ve olasılık dağılımının grafiği aşağıda verilmiştir.

x	P(x)
0	0.0640
1	0.2880
2	0.4320
3	0.2160
Toplam	1.0000

Tablo 5.8 *x 'in Olasılık Dağılımı.*



Şekil 5.5 *x 'in Olasılık Dağılımının Grafiği.*

ÖRNEK 21

Bir araştırma sonucunda tüketicilerin % 20'sinin indirim yapan marketlerden alışveriş yaptıkları bulunmuştur. Bu tüketiciler arasından rassal seçilen 6 kişi için aşağıdaki değerleri bulunuz.

a) 3 tüketicinin indirim yapan marketlerden alışveriş yapma olasılığını.

b) En çok 2 tüketicinin indirim yapan marketlerden alışveriş yapma olasılığını.

c) En az 3 tüketicinin indirim yapan marketlerden alışveriş yapma olasılığını.

d) Tüketicilerden 1 - 3 tanesinin indirim yapan marketlerden alışveriş yapma olasılığını.

e) x rassal değişkeni indirim yapan marketlerden alışveriş yapan tüketicilerin sayısını göstermek üzere x 'in olasılık dağılımını yazınız ve olasılık dağılımının grafiğini çiziniz.

ÇÖZÜM

İstenen olasılık değerlerinin bulunabilmesi için

n = toplam tüketici sayısı = 6
x = indirim yapan marketlerden alışveriş yapan kişi sayısı
= 0, 1, 2, 3, 4, 5, 6
p = P (indirim yapan marketlerden alışveriş yapma) = 0.20
q = P (indirim yapan marketlerden alışveriş yapmama) = 1 - p = 1 - 0.20
= 0.80

değerlerine gereksinim vardır.

a) 3 tüketicinin indirim yapan marketlerden alışveriş yapma olasılığı,

$$P(x = 3) = \binom{6}{3}(0.20)^3 (0.80)^3 = 0.0819$$

b) En çok 2 tüketicinin indirim yapan marketlerden alışveriş yapma olasılığı,

$$P(x \leq 2) = P(x = 0) + P(x = 1) + P(x = 2)$$

$$= 0.2621 + 0.3932 + 0.2458 = 0.9011$$

c) En az 3 tüketicinin indirim yapan marketlerden alışveriş yapma olasılığı,

$$P(x \geq 3) = P(x = 3) + P(x = 4) + P(x = 5) + P(x = 6)$$

ya da

$$1 - P(x>2) = 1 - 0.9011 = 0.0989$$

$$= 0.0819 + 0.0154 + 0.0015 + 0.0001 = 0.0989$$

d) Tüketicilerden 1 - 3 tanesinin indirim yapan marketlerden alışveriş yapma olasılığı,

$$P(1 \leq x \leq 3) = P(x = 1) + P(x = 2) + P(x = 3)$$

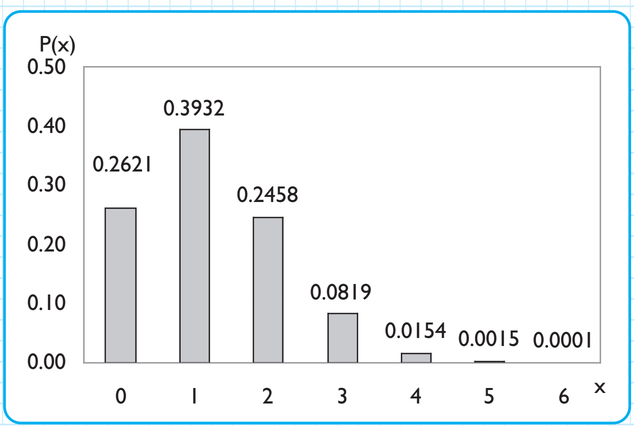
$$= 0.3932 + 0.2458 + 0.0819 = 0.7209$$

dur.

e) x rassal değişkenin olasılık dağılım ve olasılık dağılım grafiği aşağıdaki gibidir.

***Tablo 5.9** x 'in Olasılık Dağılımı.*

x	P(x)
0	0.2621
1	0.3932
2	0.2458
3	0.0819
4	0.0154
5	0.0015
6	0.0001
Toplam	1.0000



***Şekil 5.6** x 'in Olasılık Dağılımının Grafiği.*

Başarı Olasılığı ve Binom Dağılımının Biçimi

n deneme durumunda,

1. Eğer p = 0.50 ise binom dağılımı simetrik,
2. Eğer p , 0.50'den küçük ise binom olasılık dağılımının sağa doğru çarpık,
3. p , 0.50'den büyük ise binom olasılık dağılımının sola doğru çarpık,

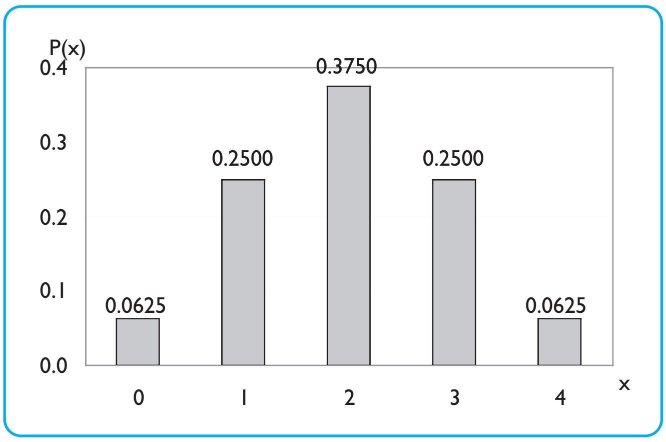
olduğu gösterilebilir.

Bu durumlar aşağıda verilmiştir:

1. n = 4 ve p = 0.50 olarak alınacak olursa, x olasılık dağılımı ve olasılık dağılımının simetrik grafiği aşağıdaki gibidir.

Tablo 5.10 *n = 4 ve p = 0.50 için x'in Olasılık Dağılımı*

x	P(x)
1	0.2500
2	0.3750
3	0.2500
4	0.0625
Toplam	1.0000

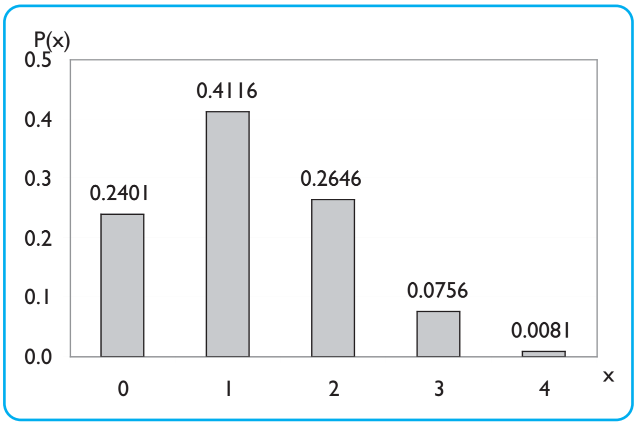


Şekil 5.7 *x 'in Olasılık Dağılımının Grafiği.*

2. n = 4 ve p = 0.30 (0.50'den küçük) olarak alınacak olursa, x 'in olasılık dağılımı ve sağa doğru çarpık grafiği aşağıdaki gibidir.

Tablo 5.11 *n = 4 ve p = 0.30 için x 'in Olasılık Dağılımı.*

x	P(x)
0	0.2401
1	0.4116
2	0.2646
3	0.0756
4	0.0081
Toplam	1.0000

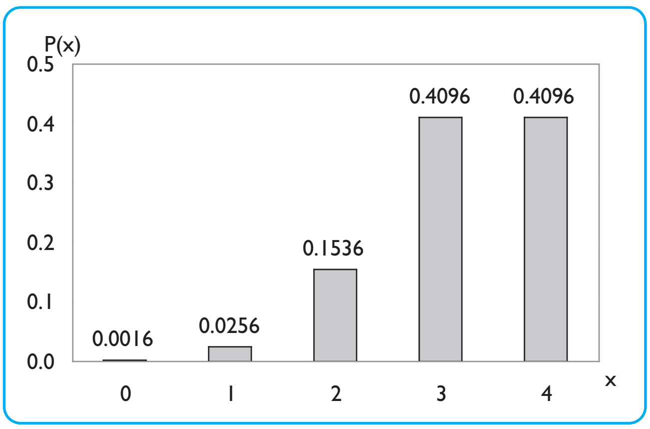


Şekil 5.8 x 'in Olasılık Dağılımının Grafiği.

3. n = 4 ve p = 0.80 (0.50'den büyük) olarak alınacak olursa, x 'in olasılık dağılımı ve sola doğru çarpık grafiği aşağıdaki gibidir.

Tablo 5.12 n = 4 ve p = 0.80 için x 'in Olasılık Dağılımı.

x	P(x)
0	0.0016
1	0.0256
2	0.1536
3	0.4096
4	0.4096
Toplam	1.0000



Şekil 5.9 x 'in Olasılık Dağılımının Grafiği.

Binom Dağılımının Ortalama ve Standart Sapması

Daha önce, kesikli bir rassal değişkenin olasılık dağılımının ortalama ve standart sapmasının nasıl hesaplanacağına değinilmişti. Buradaysa kesikli rassal değişkenin binom dağılımına sahip olması durumunda, ortalama ve standart sapmanın elde edilmesinde kullanılan, daha uygun ve basit formüller incelenecektir.

Binom Dağılımının Ortalama ve Standart Sapması, Bir binom dağılımının ortalama ve standart sapması,

$$\mu = n\,p$$

ve

$$\sigma = \sqrt{n\,p\,q}$$

biçiminde olup, burada n toplam deneme sayısı, p başarı olasılığı ve q ise başarısızlık olasılığıdır.

Aşağıdaki örnekte, bir binom dağılımı için ortalama ve standart sapmanın hesaplanması verilecektir.

ÖRNEK 22

Yapılan bir araştırmayla bir kasabadaki erişkinlerin % 58'inin psikolojik sorunu olduğu bulunmuştur. Bu kasabadan rassal 25 erişkin seçilmiştir. x , bu örneklemdeki psikolojik sorunu olan kişi sayısını göstermek üzere, x 'in olasılık dağılımının ortalama ve standart sapmasını bulunuz.

ÇÖZÜM

25 denemesi olan bu deneyde, psikolojik sorunu olan ve olmayan erişkinler olmak üzere iki sonuç bulunmaktadır. Burada başarı olarak düşünülen sonuç p = 0.58 ve başarısızlık olarak değerlendirilen sonuç ise q = 0.42 dir. Bu örnekte binom olasılık dağılımına ilişkin ortalama ve standart sapma formülleri kullanılmadan da istenen değerleri bulmak olanaklı ancak yorucudur. Oysa ki yukarıda verilmiş olan formüllerden yararlanılarak ortalama ve standart sapma değerleri sırasıyla,

$$\mu = n\,p = 25\ (0.58) = 14.50$$

$$\sigma = \sqrt{n\,p\,q} = \sqrt{(25)(0.58)(0.42)} = 2.47$$

olarak kolayca bulunabilir. Bu değerlerin anlamı; seçilen 25 kişiden 2.47 standart sapmayla ortalama 14.50 tanesinin psikolojik sorunlu olması beklenmektedir.

SIRA SİZDE

1. Aşağıdaki kavramları kısaca açıklayınız.
   a) Bir binom deneyi
   b) Bir deneme
   c) Bir binom rassal değişkeni

2. Aşağıdakilerden hangilerinin binom deneyi olduğunu söyleyiniz.
   a) Bir zarın çok kez atılarak sonuçlarının okunması.
   b) Bir zarın çok kez atılarak sonuçlarının tek sayı mı yoksa çift sayı mı olduğunun okunması.
   c) Bir ülkedeki tüm seçmenlerin % 54'ünün mevcut iktidar partisini desteklediği bilinmektedir. Bu kitleden (seçmenler) rassal seçilen az sayıda seçmene, mevcut iktidar partisini destekleyip desteklemediklerinin sorulması.

3. x , binom dağılımı gösteren kesikli bir rassal değişken olmak üzere, binom formülünden yararlanarak aşağıdaki olasılıkları bulunuz.
   a) n = 8 ve p = 0.60 için P(x = 5)
   b) n = 4 ve p = 0.30 için P(x = 3)
   c) n = 6 ve p = 0.20 için P(x = 2)

POİSSON OLASILIK DAĞILIMI

AMAÇ 6

Kesikli rassal değişkenlerin diğer önemli bir dağılımı olan Poisson dağılmını kullanarak ilgili olasılıkları, ortalama ve standart sapmasını hesaplayabileceksiniz.

Fransız matematikçi Simeon D. Poisson'un adıyla anılan Poisson olasılık dağılımı, binom dağılımı gibi X'in kesikli bir rassal değişken olması durumunda (yaygın) kullanılan dağılımlardan biridir. Örneğin bir kavşakta trafik kazası olması ayda birkaç kez rastlanan bir olaydır. Burada istenen, gelecek ay o kavşakta iki trafik kazası olması olasılığıdır. Bu örnek Poisson olasılık dağılımına uygundur ve her kaza olması; meydana gelme ya da tekrar olma (occurrence) biçiminde ifade edilir. Bu durumda Poisson dağılımının, rassal ve bağımsız olaylı deneylerde kullanıldığı söylenebilir. Kaza örneğinde olduğu gibi, Poisson dağılımında olaylar rassaldır, herhangi bir sıra izlemedikleri gibi önceden kestirilmeleri de olanaklı değildir. Burada olayların bağımsızlığının anlamı, bir olayın bir kez meydana gelmesi ve kendisini izleyen olayların meydana gelmesi ya da gelmemesi üzerinde etkisinin bulunmamasıdır. Olayların meydana gelişi, hep bir aralıkta ele alınır (trafik örneğinde bir ay gibi). Bu aralık bir zaman aralığı, bir uzay aralığı olabileceği gibi bir hacim aralığı da olabilmektedir. İncelenen bir aralıkta olayın tekrarı rassal ve bağımsızdır. Eğer verilen bir aralıkta tekrar sayısının ortalaması biliniyorsa, Poisson olasılık dağılımı kullanılarak, x ile gösterilen tekrar sayısına ilişkin herhangi bir değerin olasılığı hesaplanabilmektedir.

Poisson Olasılık Dağılımının Uygulanma Koşulları, Poisson olasılık dağılımının uygulanabilmesi için aşağıdaki üç koşulun sağlanması gerekir.

1. x kesikli rassal değişkendir.
2. Tekrarlar rassaldır.
3. Tekrarlar bağımsızdır.

Konuya açıklık kazandırılması açısından aşağıda, Poisson olasılık dağılımının uygulanabileceği bazı örnekler ele alınmıştır..

1. Bir hastanenin acil servisine belirli bir zaman aralığında (bir saat, bir gün) gelen hasta sayısı. Burada hasta gelişleri (tekrar) rassaldır ve gelen hasta sayısı 0, 1, 2,.... olabilir. Hasta gelişleri (tekrar) bağımsızdır. Çünkü gelişler tek tektir ve gelen iki hasta arasında ilişki yoktur.
2. Bir makinede üretilecek 100 parçadan, kusurlu parça sayısı da Poisson dağılımına uygundur. Çünkü burada bir hacim aralığı (100 parça) söz konusu olup, kusurlu parça sayıları (tekrar) rassal ve bir parçanın kusurlu olması, bir diğerinden bağımsızdır.
3. 5 metre uzunluğunda bir demir çubuktaki hava kabarcıkları (kusur) inceleniyor olsun. Bu örnekte aralık bir uzay aralığı olup hava kabarcığı sayısı rassaldır ve bu hava kabarcıkları birbirinden bağımsızdır.

Bu örneklere benzer bir biçimde olan aşağıdaki örnekler de Poisson olasılık dağılımına uygundur.

1. Bir otoyolda bir haftalık süredeki kaza sayısı.
2. Bir manava bir saatlik sürede gelen müşteri sayısı.
3. Bir mağazada bir haftalık sürede satılan TV seti sayısı.

Öte yandan bir doktorun muayenehanesine gelen hasta sayısı bunlardan farklıdır. Çünkü gelecek hastalar daha önce randevu aldıklarından rassal bir sayı olmayıp, kaç kişinin geleceği daha önceden (yaklaşık olarak) bilinmektedir. Aynı biçimde bir hava alanından kalkacak ya da bu hava alanına inecek uçak sayısı da rassal değildir ve önceden bilinmektedir. Bu nedenle rassal olma koşulu sağlanmadığı için bu tür verilere Poisson olasılık dağılımı uygulanamamaktadır.

Poisson olasılık dağılımında ortalama tekrar (meydana gelme) sayısı λ (Lamda) ile, verilen aralıktaki tekrar sayısı da x ile gösterilmektedir. Poisson olasılık dağılımı kullanılarak, λ ortalama tekrar sayısı biliniyorken, verilen bir aralıkta x tekrarlanma sayısının olasılığı elde edilmektedir.

Poisson Olasılık Dağılımı Formülü, Poisson olasılık dağılımına göre, bir aralıkta x tekrarın gözlenmesi olasılığı,

$$P(x) = \frac{\lambda^x \ e^{-\lambda}}{x!}$$

eşitliğiyle bulunmaktadır. Burada λ verilen aralıkta ortalama tekrar sayısıdır (e=2.71828).

Bir aralıktaki ortalama tekrar sayısı λ , Poisson olasılık dağılımının parametresi ya da kısaca Poisson parametresi olarak bilinir. Yukarıdaki formülden de anlaşılacağı gibi, x tekrar sayısının olasılığının bulunabilmesi için, sadece λ değerinin bilinmesi yeterlidir. Çünkü formüldeki $e^{-\lambda}$ değeri, ya hesaplanmakta ya da hazır tablolardan bulunmaktadır.

ÖRNEK 23

Yapılan bir araştırmadan 18-24 yaş grubundaki tüketicilerin ayda ortalama 6.9 kez alışverişe çıktıları bulunmuştur. Poisson olasılık dağılımına uyduğu düşünülen rassal değişken için, 18-24 yaş grubunun ayda 5 kez alışverişe çıkması olasılığını bulunuz.

ÇÖZÜM

Ortalama alışveriş sayısı olan 6.9 dağılımın ortalaması ve olasılığı bulunması istenen tekrar sayısı x ise 5 alınarak istenen olasılık değeri, Poisson dağılımı formülünden elde edilir.

$$P(x=5) = \frac{\lambda^x \ e^{-\lambda}}{x!} = \frac{(6.9)^5 \ e^{-6.9}}{5!}$$

$$= \frac{(15640.31349) \ (0.001008)}{120} = 0.1314$$

ÖRNEK 24

Bir çamaşır makinesi, ayda ortalama, üç kez sıkma arızası yapmaktadır. Poisson olasılık dağılımından yararlanarak bu makinenin gelecek ay

a) İki kez arızalanması,

b) En çok bir kez arızalanması

olasılıklarını bulunuz.

ÇÖZÜM

Ayda ortalama üç kez sıkma arızası olduğuna göre λ = 3 dür. Bu durumda;

a) Gelecek ay iki kez sıkma arızası olma olasılığı;

$$P(x = 2) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{(3)^2 e^{-3}}{2!} = \frac{(9)(0.049787)}{2} = 0.2240$$

olarak bulunur.

b) Gelecek ay, en çok bir sıkma arızası ifadesiyle; hiç arıza olmaması ve sadece bir arıza olması kastedilmektedir.

$$P\text{ (En çok bir arıza)} = P(x \leq 1) = P\text{ (0 ya da 1)} = P(x = 0) + P(x = 1)$$

olarak elde edilir.

Poisson olasılık dağılımında λ ve x 'in aralıkları aynı olmalıdır. Aksi takdirde eşitliğin sağlanması için λ ortalamasının tekrar tanımlanması gerekir.

ÖRNEK 25

Bir firma yeni ürettiği bir ürünün pazar bulabilmesi için, bu ürünü alanlardan beğenmeyenlere, 7 günlük süre içerisinde ürünü geri getirdikleri takdirde, paralarının iadesi kampanyası başlatmıştır. Geçen süre içerisinde satılan 10 üründen 2 tanesinin parasının, iade edildiği görülmüştür. Poisson olasılık dağılımından yararlanarak gelecekte satılacak 40 üründen 6 tanesinin parasının iade edilmesi olasılığını bulunuz.

ÇÖZÜM

Burada önemli bir sorunla karşılaşılmaktadır. Bu sorun, yukarıda değinildiği gibi; ortalama değerin aralığı ile x'in aralığıyla farklı olmasıdır. Çünkü λ = 2 değeri 10 satıştan elde edilmişken gelecekte yapılacak 40 satıştan 6 tanesine para iadesi sorulmaktadır. Bu durumda x = 6 aynı kalacak, ancak λ ortalama değeri, istenen aralık için tekrar tanımlanacak, bu değer de λ = 8 olacaktır. Ortalamanın yeniden tanımlanmasının ardından Poisson olasılık dağılımından yararlanarak istenen olasılık;

$$P(x = 6) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{(8)^6 e^{-8}}{6!} = \frac{(262144)(0.000335)}{720} = 0.1220$$

bulunur.

Aslında Poisson olasılık dağılımı, nadir karşılaşılan ya da tekrarlanan (olasılığı çok küçük) olaylar için kullanılmaktadır. Oysa yukarıdaki örnekte olasılık değeri P = 2 / 10, tekrar sayısı n = 40 ve olasılığı bulunması istenen tekrarlanma sayısı x = 6 olarak düşünüldüğünde, deney bir binom deneyi olarak düşünülür ve istenen olasılık binom dağılımı formülünden de elde edilebilir.

$$P(x = 6) = \binom{40}{6} = (0.20)^6 (0.80)^{34} = \frac{40!}{6!\ 34!} = (0.20)^6 (0.80)^{34} = 0.1246$$

Bu duruma, binom dağılımı yaklaşımında Poisson dağılımının kullanılması adı verilir ve özellikle n sayısının çok büyük olması durumunda binom dağılımıyla olasılık bulmanın zaman kaybettirmesini ortadan kaldırmak amacıyla kullanılır.

ÖRNEK 26

Bolkazanç bankasının Kızılay Şubesinde her gün ortalama iki tane yeni hesap açtırıldığı bilinmektedir. Verilen bir günde,
a) 6 yeni hesap
b) En çok 3 yeni hesap
c) En az 7 hesap
açtırılması olasılıklarını bulunuz.

ÇÖZÜM

Önce, formülde kullanılacak değerler tanımlanmalıdır.

λ= Her gün açılan ortalama yeni hesap sayısı.
x = Verilen günde açılacak yeni hesap sayısı.

Bu bilgiler ışığında Poisson olasılık dağılım formülü kullanılarak istenen olasılıklar;

a)

$$P(x = 6) = \frac{\lambda^x \, e^{-\lambda}}{x!} = \frac{(2)^6 \, e^{-2}}{6!} = 0.0120$$

b) $P(x \leq 3) = P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3)$
$= 0.1353 + 0.2707 + 0.2707 + 0.1804 = 0.8571$

c) $P(x \geq 7) = 1 - P(x < 7)$
$= 1 - \{ P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3)$
$+ P(x = 4) + P(x = 5) + P(x = 6) \}$
$= 0.0045$

olarak bulunur.

ÖRNEK 27

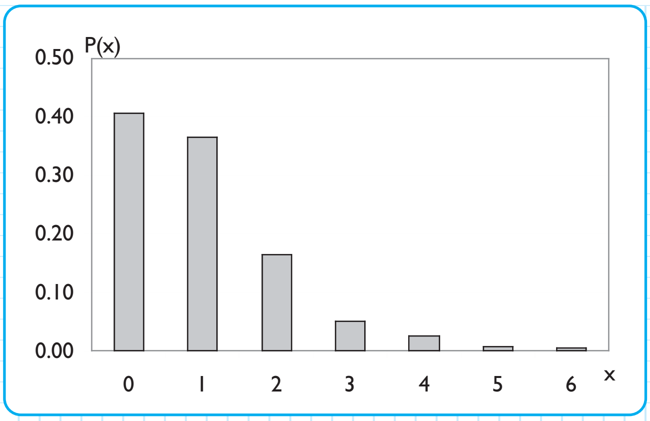
Bir otomobil galerisinde günde ortalama 0.9 otomobil satılmaktadır. x,verilen bir günde satılan otomobil sayısını göstermek üzere, Poisson olasılık dağılımını bulunuz ve olasılık dağılımının grafiğini çiziniz.

ÇÖZÜM

Poisson olasılık dağılımı için gerekli değeri $\lambda = 0.9$ bilinmektedir. Ancak satılan otomobil sayısı x ise 0, 1, 2, 3, 4, ..., olabilecektir. Böylesi durumlarda x değerinin sayısı, bulunan olasılık değerine bakılarak belirlenmektedir. Olasılık değerinin ihmal edilebilecek düzeyde olması durumunda (yaklaşık sıfır) olasılığı bulunan x değeri durdurulmaktadır. Bu düşünce ışığında oluşturulan Poisson olasılık dağılımı ve olasılık dağılımının grafiği aşağıda verilmiştir.

Tablo 5.13 $\lambda = 0.9$ için Olasılık Dağılımı.

x	P(x)
0	0.4066
1	0.3659
2	0.1647
3	0.0494
4	0.0111
5	0.0020
6	0.0003



Şekil 5.10 *Olasılık Dağılımının Grafiği.*

Poisson Olasılık Dağılımının Ortalaması

Poisson dağılımda ortalama ve varyans parametrelerinin her ikisi de λdır. Standart sapma, varyansın pozitif kare kökü olduğundan bu da $\sqrt{\lambda}$ 'dir.

$$\mu = \lambda$$
$$\sigma^2 = \lambda$$
$$\sigma = \sqrt{\lambda}$$

Örneğin, yukarıdaki Örnek 5.27 için bu değerler;

$$\mu = \lambda = 0.9$$
$$\sigma^2 = \lambda = 0.9$$

otomobil olmaktadır.

SIRA SİZDE

1. Poisson olasılık dağılımının uygulanabilmesi için sağlanması gereken koşullar nelerdir?

2. Poisson dağılımının parametresi nedir ? Ne anlama gelmektedir?

3. Poisson formülünden yararlanarak aşağıdaki olasılıkları bulunuz.
a) $\lambda = 4$ için $P(x \leq 1)$
b) $\lambda = 5.3$ için $P(x = 8)$

Kendimizi Sınayalım

1. Dayanıklı tüketim malı satan bir mağazanın son 100 iş günündeki günlük satışları aşağıdaki tabloda verilmiştir.

Satış sayıları	2	3	4	5	6
Gün sayıları	12	21	34	19	14

Yukarıdaki tabloya göre X, günlük satışı göstermek üzere, P(x<4) olasılığı kaçtır?

a. 0.04
b. 0.17
c. 0.21
d. 0.33
e. 0.50

2. X, rassal değişkeninin olasılık dağılımı, aşağıdaki tabloda verilmiştir.

X	P(x)
0	0.12
1	0.27
2	0.43
3	0.18

Yukarıdaki tabloya göre, X' in ortalaması ve standart sapması kaçtır?

a. μ = 1.670
σ = 0.906
b. μ = 2.472
σ = 0.906
c. μ = 1.670
σ = 1.127
d. μ = 1.528
σ = 1.417
e. μ = 1.740
σ = 2.180

3. Bir firma piyasa koşullarının olumsuzluğu nedeniyle 16 çalışanından 2 tanesini rastgele seçerek işlerine son verecektir. Yapılabilecek farklı seçim sayısı kaçtır?

a. 40
b. 70
c. 100
d. 120
e. 130

4. X, p = 0.3 ve q = 0.7 olmak üzere Binom dağılmış bir rassal değişkendir. X' in varyansı 2.1 ise n kaçtır?

a. 2
b. 4
c. 5
d. 7
e. 10

5. A gazetesi okurlarının 0.60'ı okudukları gazeteyi güvenilir bulmaktadır. Bu gazetenin okurları arasından seçilen 3 okuyucunun da gazeteyi güvenilir **bulmama** olasılığı nedir?

a. 0.064
b. 0.216
c. 0.310
d. 0.370
e. 0.421

6. Bir basketbolcunun serbest atışdaki başarısı 0.85' dir. Basketbolcunun deneyeceği 3 atıştan, ikisini sayıya dönüştürme olasılığı nedir?

a. 0.216
b. 0.325
c. 0.442
d. 0.467
e. 0.518

7. Bir havayolu şirketinin merkezine hergün ortalama 9.7 şikayet telefonu gelmektedir. Herhangi bir günde 7 şikayet gelmesi olasılığı nedir? (Yol gösterme: Poisson dağılımından yararlanınız.)

a. 0.0982
b. 0.1740
c. 0.2234
d. 0.2401
e. 0.3547

8. Ekonomik kriz yaşayan ülkemizde çok sayıda küçük ve orta boy işletme kapanmaktadır. Konuyla ilgili olarak Eskişehir'de yapılan bir araştırmada hergün ortalama 1.3 işyerinin kapandığı belirlenmiştir.
Verilen bir günde Eskişehir'de 3' den az işyerinin kapanma olasılığı nedir?

a. 0.0294
b. 0.1714
c. 0.3740
d. 0.5343
e. 0.8571

9. X, ortalaması 1.8 olan Poisson dağılmış bir rassal değişken olduğuna göre standart sapması nedir?

a. 0.04
b. 0.17
c. 0.21
d. 1.34
e. 2.18

10. Bir A firmasında istenilen verimin elde edilmesi ve sürdürülebilmesine ilişkin olarak çalışanlara uygulanan bir anket sonucunda, çalışanların 0.50' si çalışma koşullarının en önemli faktör olduğunu belirtmişlerdir.
A firması çalışanlarından rastgele seçilen 10 kişiden en çok 5 tanesinin çalışma koşullarının **en önemli** etken olduğunu işaretleme olasılığı nedir?

a. 0.2051
b. 0.2725
c. 0.3770
d. 0.6230
e. 0.7730

Yanıt Anahtarı

1. d
2. a
3. d
4. e
5. a
6. b
7. a
8. e
9. d
10. d

Yararlanılan Kaynaklar

HOEL, P.G. and JESSEN, R.J.: **Basic Statistics for Business and Economics**, Wiley, NewYork, 1971.

MANN, P.S.: **Introductory Statistics,** 2nd Edition, Wiley, New York, 1995.

O'HAGAN, A., **Probability: Metods and Measurement**, Chapman and Hall, London, 1988.

WONNACOTT, R.J., WONNACOTT, T.H.: **Introductory Statistics,** 4th Edition, Wiley, Singapore, 1985.

PIERRE SIMON LAPLACE (1749-1827)

1700'lü yılların sonlarında astronomi alanındaki çalışmalarıyla dikkat çekti. İzleyen yıllarda integral, sonlu uzaylar ve differansiyel denklemler üzerinde çalıştı.
1812'de, 1779 yılında yayınladığı bir makalesini temel alarak olasılık kavramı üzerinde çalışmaya başladı. Sonraki yıllarda Gauss ve Legendre'nin de ilgilendiği en küçük kareler yöntemine ilişkin çalışmasını tamamladı. 1819'da olasılık konusunda yazdığı "Theorie des Probabilités" adlı kitabı yayınlandı.

Sürekli Rassal Değişkenler ve Normal Dağılım

6

Çalışma Biçimine İlişkin Olarak

- *Sorulan sorularda nelerin istendiği kesin olarak anlaşıldıktan sonra çözüme geçilmeli,*
- *Alıştırmaların tümü mutlaka çözülmelidir.*

Amaçlar:

- *Sürekli olasılık dağılımı kavramını açıklayabilecek ve normal dağılımı ana çizgileriyle inceleyecek, normal dağılıma ilişkin olasılıkları hesaplayabileceksiniz.*
- *Normal dağılım eğrisi altındaki alan biliniyor iken Z ve X değerlerini belirleyebileceksiniz.*
- *Binom ve normal dağılım arasındaki ilişkiyi açıklayabileceksiniz.*

İçindekiler

GİRİŞ

Bir önceki bölümde kesikli rassal değişkenler ve dağılımlar incelenmişti. Bu bölümdeyse herhangi bir aralıkta çok sayıda değerler alabilen sürekli bir rassal değişken konu edilecektir.

Sürekli bir rassal değişkenin olası değerleri sonsuz ve sayılamaz olarak kabul edilmektedir. Örneğin sabah evden işe giderken harcanan zaman, sürekli bir rassal değişkendir. Çünkü bu zaman en az 5 dakika, en çok ise 130 dakika olmak üzere, bu aralıktaki yüzlerce hatta saniyeler dikkate alınacak olursa milyonlarca değer, zamanı gösteren X sürekli rassal değişkenin değeri olmaktadır. Bu örnekte X zaman değişkeninin alabileceği değerler 5-130 dakika arasında olmakta ve bu aralığa (daha sonra ayrıntılı verileceği gibi) tanım aralığı adı verilmektedir.

Sürekli bir rassal değişkenin uygulanacağı pek çok olasılık dağılımı bulunmakla birlikte, burada sadece normal olasılık dağılımı ve binom dağılımına yaklaşan normal dağılımla karşılaştırmalarda (test amacıyla) kullanılan Student t dağılımı incelenecektir.

SÜREKLİ OLASILIK DAĞILIMI

AMAÇ 1

Sürekli olasılık dağılımı kavramını açıklayabilecek ve normal dağılımı ana çizgileriyle inceleyecek, normal dağılıma ilişkin olasılıkları hesaplayabileceksiniz.

Yukarıda belirtilmiş olduğu gibi, değerleri sayılamayan bir rassal değişkene, "sürekli bir rassal değişken" denilmektedir. Başka bir anlatımla, sürekli bir rassal değişken bir aralıkta (veya aralıklarda) her değeri alabilmektedir. Çünkü bir aralıkta bu değişkenin alabileceği sonsuz sayıda değer olduğu varsayılmakta ve bu değerlerin sayılamayacak çoklukta olduğu kabul edilmektedir. Örneğin; (bir önceki bölümde verilmiş olan) bir pilin ömrü, kişilerin boy uzunluğu, bir sınavın tamamlanma süresi, yeni doğmuş bebeklerin ağırlığı gibi değişkenler, sürekli rassal değişkenlerdir.

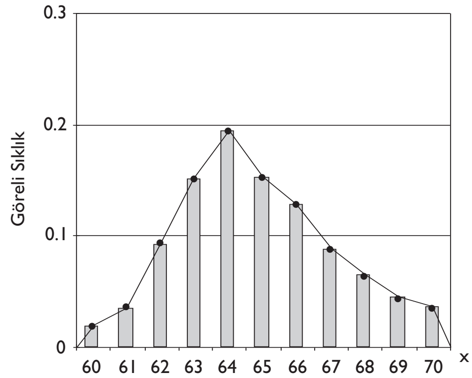
Bir üniversitede öğrenim gören 5000 erkek öğrencinin ağırlıkları X rassal değişken olarak düşünülerek kg cinsinden aşağıda verilmiştir:

__Tablo 6.1__ Erkek öğrencilerin ağırlıklarına ilişkin göreli sıklık dağılımı.

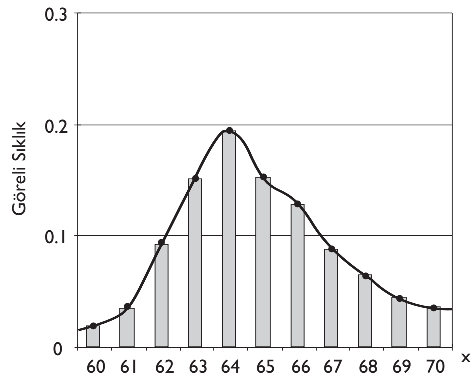
Ağırlık x	f	Göreli Sıklık
60-61	9	0.018
61-62	170	0.034
62-63	460	0.092
63-64	750	0.150
64-65	970	0.194
65-66	760	0.152
66-67	640	0.128
67-68	440	0.088
68-69	320	0.064
69-70	220	0.044
70-71	180	0.036
	N = 5,000	**Toplam= 1.000**

Yukarıda verilmiş olan göreli sıklıklar, belirlenmiş sınıfların yaklaşık olasılıkları olarak kullanılabilmektedir.

Aşağıda Şekil 6.1.'de göreli sıklık dağılımının histogram ve poligonu, Şekil 6.2.'deyse bu verilere ilişkin düzeltilmiş (smooted) poligonu verilmiştir. Düzeltilmiş poligon, x sürekli rassal değişkenin olasılık dağılımı eğrisine bir yaklaşımdır.



Şekil 6.1 *Ağırlıkların histogram ve poligonu.*



Şekil 6.2 *Ağırlıkların olasılık dağılım eğrisi.*

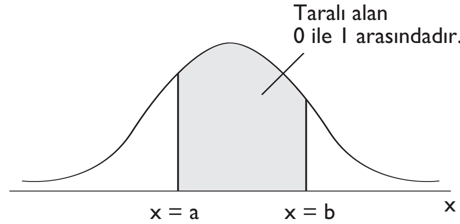
Yukarıdaki tabloda sınıf genişliği 1 birim olarak alınmıştır. Bu genişlik 1 birimden fazla alınacak olursa, bu kez de önce göreli sıklık yoğunlukları yeniden elde edilir ve yoğunluk değerlerinin grafiği çizilerek dağılım eğrisi bulunur. Bir sınıfın göreli sıklık yoğunluğu, bir sınıftaki göreli sıklığın sınıf genişliğine bölünmesiyle elde edilmektedir. Göreli sıklık yoğunlukları, bir histogramdaki dikdörtgenlerin alanları toplamını 1.0 yapmak için bulunmaktadır. Sürekli bir rassal değişkenin olasılık dağılım eğrisine, olasılık yoğunluk fonksiyonu da denmektedir.

Sürekli bir rassal değişkenin olasılık dağılımı aşağıdaki iki özelliği sağlamalıdır.

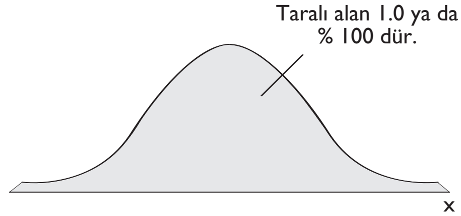
1. Bir aralıkta herhangi bir değer alan x 'in olasılığı 0 – 1 arasındadır.

2. x 'in aldığı tüm değerlerin olasılıkları toplamı 1'dir.

İlk özellik, x sürekli bir rassal değişkenin olasılık dağılım eğrisi altındaki alan, 0 ve 1 arasındaki alandır (Şekil 6.3). İkinci özellikse, sürekli bir rassal değişkenin olasılık dağılım eğrisi altında kalan alan toplamının her zaman 1.0 veya % 100 olduğudur (Şekil 6.4).



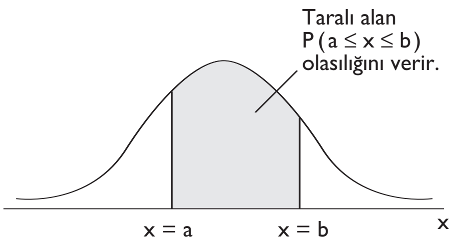
Şekil 6.3 *Eğri altındaki alan iki nokta arasındadır.*



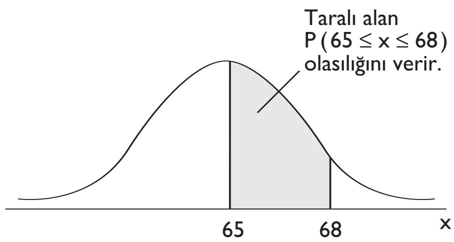
Şekil 6.4 *Bir olasılık dağılım eğrisi altındaki toplam alan.*

Sürekli bir rassal değişkenin bir aralıkta aldığı varsayılan değerlerin olasılığı, aşağıdaki Şekil 6.5.'den de anlaşılacağı gibi, bir aralığın iki limiti arasında ve eğri altındaki alandır. Şekildeki (eğri altında) a ve b arasındaki taralı (gri zeminli) alan, X 'in a ve b arasında olma olasılığını vermektedir. Bu da,

$P\ (a \leq x \leq b)$ = Eğri altında a ve b noktaları arasındaki alan biçiminde gösterilmektedir. Burada x; a'ya eşit ya da büyük, b'ye eşit ya da küçüktür.



Şekil 6.5 *Olasılığın eğri altındaki alan olarak ifadesi.*



Şekil 6.6 *(65-68) kg aralığındaki x 'in olasılığı.*

Erkek öğrencilerin ağırlıkları örneğine dönülecek olursa, bu gruptan rassal seçilen bir öğrencinin ağırlığının 65 – 68 kg arasında olması olasılığı (yukarıdaki Şekil 6.6.'da verildiği gibi)ağırlıkların dağılım eğrisi altındaki alan olarak gösterilmektedir.

$$P\ (65 \leq x \leq 68)$$

olarak gösterilen olasılık değerinde, x, 65 kg'a eşit ya da büyük, 68 kg'a eşit ya da küçüktür.

Sürekli bir olasılık dağılımında olasılık, her zaman bir aralık için hesaplanmaktadır. Nitekim bu olasılık yukarıdaki grafikte taralı olan 65 – 68 kg aralığı için bulunmuştur.

X gibi bir sürekli rassal değişkenin alabileceği tek bir değerin olasılığı her zaman sıfırdır. Çünkü; verilen bir noktanın alanı sıfırdır. Örneğin; yukarıdaki erkek öğrencilere ilişkin örnekte, rassal seçilen bir erkek öğrencinin ağırlığının 67 kg olma olasılığı sıfırdır ve,

$$P\ (x = 67) = 0$$

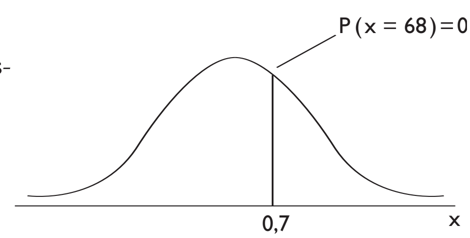
biçiminde gösterilir. Bu durum grafik olarak da aşağıdaki biçimde gösterilir.

Genel ifadesiyle, a ve b, X 'in aldığı iki değer olduğunda, bu değerlere ilişkin olasılıklar;

$$P\ (a) = 0 \quad ve \quad P\ (b) = 0$$

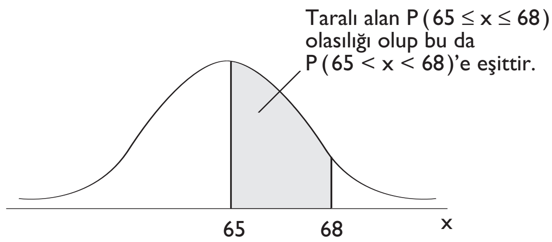
dir. Buradan da sürekli bir rassal değişken için,

$$P\ (a \leq x \leq b) = P\ (a < x < b)$$



Şekil 6.7 *x 'in tek bir değerinin olasılığı sıfırdır.*

yazılabilmektedir. Bir başka ifadeyle, verilen sınır değerlerine eşitlik (olasılık değeri sıfır olduğundan) sonucu değiştirmemektedir (*Eğer aralık, "65 ve 68 arasında" biçiminde ifade edilirse anlamı "65 < x < 68" dir. Eğer aralık, "65'den 68'e kadar" biçiminde ifade edilirse "65 ≤ x ≤ 68" dir*). Erkek öğrencilerin ağırlıklarına ilişkin örnek için de bu özellik grafik olarak aşağıda verilmiştir.

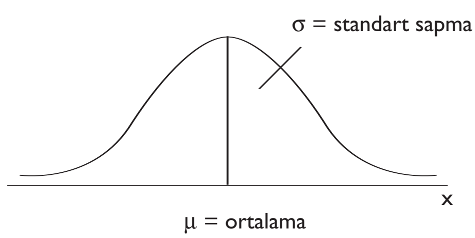


Şekil 6.8 *65 ve 68 kg arasındaki ve 65'den 68 kg'a kadarın olasılığı.*

Normal Dağılım

Normal dağılım, sürekli bir rassal değişkenin uyduğu önemli olasılık dağılımından biridir. Bu nedenle günlük yaşamda karşılaşılan pek çok değişken (kesinlikle ya da yaklaşık olarak), normal dağılır. Örneğin insanların boy uzunlukları, ağırlıkları, sınav sonuçları, paketlerin ağırlıkları, bir şişedeki sütün miktarı, ampul - TV seti - pil gibi nesnelerin ömrünün hep (yaklaşık) normal dağıldığı kabul edilir.

Normal olasılık dağılımı ya da normal eğri (curve), çan biçiminde bir şekille gösterilmektedir. Şekil 6.9'da gösterilmiş olan bu eğrinin ortalaması μ ve standart sapması σ dir. Normal dağılım gösteren sürekli bir rassal değişkene, "normal rassal değişken" denir (ancak çan eğrisine benzeyen her şekil normal dağılım eğrisi değildir).



Şekil 6.9 μ *ortalama ve* σ *standart sapmalı normal dağılım.*

Normal Olasılık Dağılımı

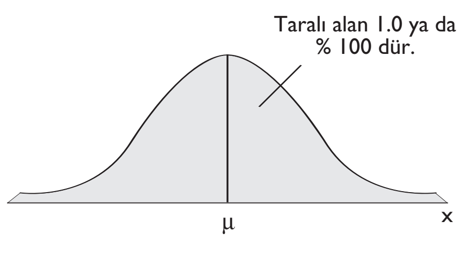
Çizildiğinde çan şeklinde bir eğrisi olan normal olasılık eğrisinde;

1. Eğri altındaki toplam alan 1.0'dır.

2. Eğri ortalamaya göre simetriktir.

3. Eğrinin iki ucu (kuyruğu) sonsuza gitmektedir.

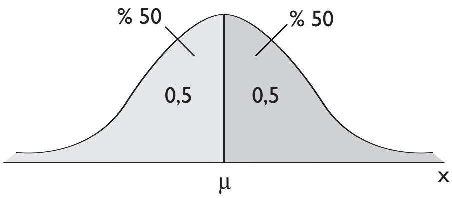
Normal dağılımın taşımak zorunda olduğu bu üç özelliğin açıklaması aşağıdadır:



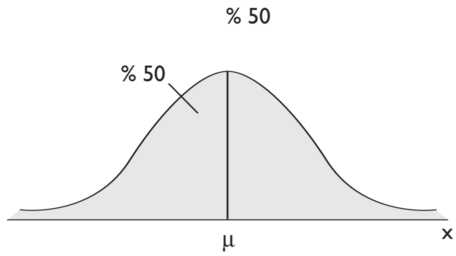
Şekil 6.10 *Bir normal eğri altındaki toplam alan.*

1. Normal dağılım eğrisi altındaki toplam alan 1.0 ya da % 100 dür.

2. Bir normal dağılım eğrisi ortalamaya göre simetrik olup, eğri altındaki toplam alanın yarısı ortalamanın sağında, yarısıysa solunda yer alır.



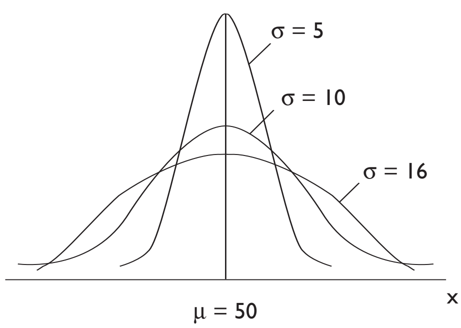
Şekil 6.11 *Bir normal eğri ortalamaya göre simetriktir.*



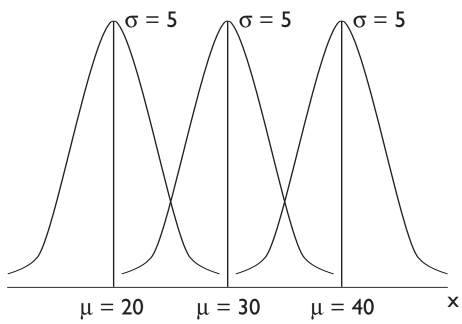
Şekil 6.12 *Bir normal eğrinin iki kuyruğu.*

3. Bir normal eğrinin iki kuyruğu yatay eksene asimptottur (dokunmayacak ve kesmeyecek bir biçimde tanımsız olarak -istendiği kadar- uzayabilmektedir). Bir normal dağılım eğrisi hiçbir zaman yatay eksene dokunmamakla birlikte; (μ - 3σ) dan küçük ve (μ + 3 σ) dan büyük aralıkta eğri altında kalan alanın sıfır olduğu düşünülmektedir.

Normal dağılımın parametreleri; μ (ortalama) ve σ (standart sapma) 'dır. Bu iki parametrenin verilmesi halinde, bir normal dağılım eğrisi altındaki, herhangi bir aralığa karşılık gelen alan bulunabilmektedir. Ancak normal dağılım eğrisi tek olmayıp, bir ailedir. Çünkü her μ ve σ seti için farklı normal dağılım söz konusudur. Bu parametrelerden μ, yatay eksen üzerinde bir normal dağılımın merkezini belirtirken, σ 'da dağılımın yayılımını ifade etmektedir. Aşağıdaki Şekil 6.13'de, ortalaması aynı, ancak standart sapması farklı üç değişik normal dağılım eğrisi verilmiştir. Yine aşağıda Şekil 6.14'deyse standart sapmaları aynı, ancak ortalamaları farklı üç normal dağılım eğrisi verilmiştir.



Şekil 6.13 *Ortalama aynı standart sapma farklı üç normal dağılım eğrisi.*



Şekil 6.14 *Standart sapma aynı ortalama farklı üç normal dağılım eğrisi.*

Normal olasılık dağılımının matematiksel eşitlikle gösterimi,

$$f(x) = \frac{1}{\sigma \, . \, \sqrt{2\,\pi}} \, e^{-(1/2)\,[(x\,-\,\mu)\,/\,\sigma]^2}$$

biçimindedir.

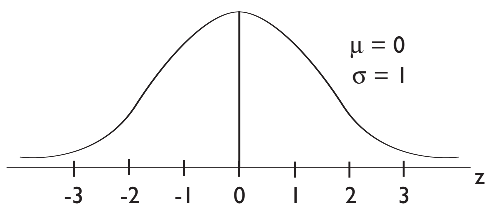
Ancak izleyen alt bölümde bu eşitlik (fonksiyon) kullanılmayacak, onun yerine bu eşitlik kullanılarak oluşturulan "Standart Normal Dağılım Tablosu" 'ndan yararlanılacaktır (Ek: Tablo 1).

Standart Normal Dağılım

Standart normal dağılım, normal dağılımın $\mu = 0$ ve $\sigma = 1$ olduğu özel bir durumudur.

STANDART NORMAL DAĞILIM

$\mu = 0$ ve $\sigma = 1$ olan normal dağılıma, "standart normal dağılım" denir.



Şekil 6.15 *Standart normal dağılım eğrisi.*

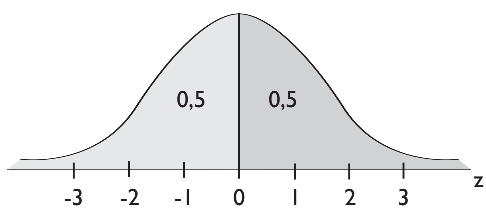
Yandaki Şekil 6.15.'den de görüleceği gibi, standart normal dağılımda rassal değişken z ile gösterilmektedir. Standart normal dağılımın birimi olan z değerlerine z skorları, standart birimler ya da standart skorlar da denir.

z DEĞERLERİ YA DA z SKORLARI

Standart normal eğrinin yatay ekseni üzerinde işaretlenmiş birimlere *z* değerleri ya da *z* skorları denir. *z* 'nin özel bir değeri; ortalama ve *z* ile ifade edilen noktanın, standart sapma cinsinden uzaklığıdır.

Şekil 6.15'de *z* değerlerini gösteren yatay eksende, ortalamanın sağındaki *z* değeri pozitif, solundakilerse negatiftir. Yatay eksen üzerindeki bir noktanın *z* değeri, ortalamayla o nokta arasındaki uzaklığın standart sapma cinsinden değeridir. Örneğin, $z = 2$ 'nin anlamı, sağ tarafta o noktanın ortalamaya iki standart sapma uzaklıkta olduğudur. Aynı biçimde $z = -2$ 'nin anlamıyla sol tarafta yine iki standart sapma uzaklıkta olduğudur.

Ek'te verilmiş olan standart normal dağılım tablosu, standart normal eğri altında; $z = 0$ ile 0.00 'dan 3.09'a kadar olan *z* değerleri arasındaki alanları vermektedir. Bu tablonun okunmasına, standart normal dağılımın ortalaması olan $z = 0$ noktasından başlanmaktadır. Daha önce de söz edildiği gibi, normal dağılım eğrisi altındaki toplam alan 1.0 'dir ve simetriklik nedeniyle ortalamanın her iki tarafındaki alan da 0.5 yani % 50 'dir.



Şekil 6.16 *Standart normal eğri altındaki alan.*

Yukarıda ortalamanın solundaki z değerlerinin negatif olduğu söylenmişti. Ama, burada unutulmaması gereken şey, alan kavramı nedeniyle sol taraftaki eğri altında kalan alanın da pozitif olduğudur.

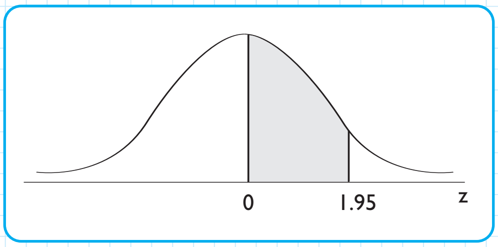
Standart normal eğri altında, iki nokta arasındaki alan z değerinin bu aralık içerisindeki değerleri alabilme olasılığıdır. Aşağıdaki örnek 6.1 – 6.4 ile Ek'de verilen tablolardan yararlanılarak, standart normal eğri altındaki alanların bulunması açıklanmaktadır.

ÖRNEK 1

Standart normal eğri altında; z = 0 ile z = 1.95 arasındaki alanı bulunuz.

ÇÖZÜM

Aşağıdaki şekilden de görüleceği gibi, burada aranan alan z = 0 ile z = 1.95 noktaları arasındaki alandır.
Bu istenen alanın sayısal değeri, Ek'te verilmiş olan Standart Normal Dağılım Tablosundan yararlanılarak bulunmaktadır. Tabloda verilen z değerinin ondalık noktanın solundaki haneyle sağındaki ilk hane (1.9), tablonun ilk sütunu olarak verilen değerlerden seçilir. Ondalık noktasının sağındaki ikinci haneyse (0.05) tablonun ilk satırından seçilir ve seçilmiş olan sütunla satırın kesişim noktasındaki değer, aranan alan değeridir. Bu değer 0.4744 olarak bulunur. Ayrıca,



Şekil 6.17 *z = 0 ile z = 1.95 arasındaki alan.*

$$P\,(0 < z < 1.95) = 0.4744$$

olarak da gösterilebilir. Unutulmamalıdır ki, sürekli bir rassal değişkenin tek bir değeri alma olasılığı sıfırdır. Başka bir ifadeyle,

$$P\,(z = 0) = 0 \quad \text{ve} \quad P\,(z = 1.95) = 0$$

dır ve yine yukarıda ayrıntılarıyla verilmiş olduğu gibi,

$$P\,(0 < x < 1.95) = P\,(0 \leq x \leq 1.95) = 0.4744$$

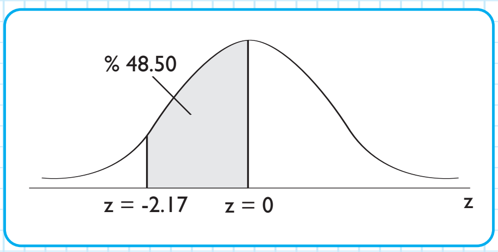
dür.

ÖRNEK 2

Standart normal eğri altındaki z = - 2.17 den z = 0'a kadar olan alanı bulunuz.

ÇÖZÜM

Normal dağılım, ortalamaya göre simetrik olduğu için z = - 2.17 'den z = 0'a kadar olan alan, z = 0 'dan z = 2.17'ye kadar olan alanla eşittir. Bu nedenle sadece z = 0 ile pozitif z değerleri arasındaki alanları veren Ek'deki tablo değerlerinden, bu durum için de yararlanılabilmektedir.
Tabloda;2.1 değeri ilk sütundan, 0.07 değeriyse ilk satırdan bulunarak, bu satır ve sütunun kesiştikleri yerde bulunan 0.4850 değeri, z = - 2.17 ile z = 0 noktaları arasında kalan alandır. Ayrıca bu değer, z değerinin verilen iki sınır değeri arasındaki alanda bulunma olasılığıdır.



Şekil 6.18 *z = -2.17 ile z = 0 noktaları arasındaki alan.*

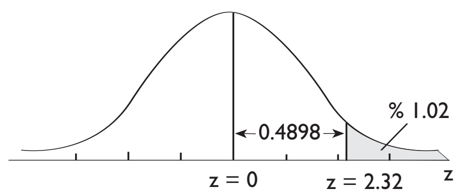
- 2.17 'den 0'a kadar olan alan =
$P\,(-2.17 \leq z \leq 0) = 0.4850$
olarak elde edilir.

ÖRNEK 3

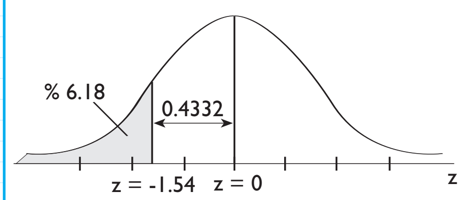
Standart normal eğri altındaki alanları bulunuz.
a) z = 2.32'nin sağındaki alanı
b) z = - 1.54'ün solundaki alanı

ÇÖZÜM

a) Daha önce de söylendiği gibi, normal dağılım tablosundaki değerler $z = 0$ ile başlamakta ve verilen noktayla ortalama ($z = 0$) arasındaki alanı vermektedir. Buradaysa verilen noktanın sağında kalan alan sorulmaktadır.
Bu durumda yine verilen noktayla ortalama arasındaki alan tablodan bulunup, sağ tarafın toplam değeri olan 0.5'den çıkartılırsa, istenen değer elde edilir.
İstenen alan : $P(z \geq 2.32) = 0.500 - 0.4898 = 0.0102$ olarak elde edilir.



***Şekil 6.19** z = 2.32 değerinin sağındaki alan.*



***Şekil 6.20** z = -1.54 değerinin solundaki alan*

b) Aynı biçimde, burada da önce $z = 0$ ile $z = -1.54$ arasındaki alan bulunur ve bulunan alan değeri 0.5 değerinden çıkartılırsa, istenen alan değerine ulaşılır.
İstenen alan : $P(z < -1.54) = 0.5000 - 0.4382 = 0.0618$
olarak hesaplanır.

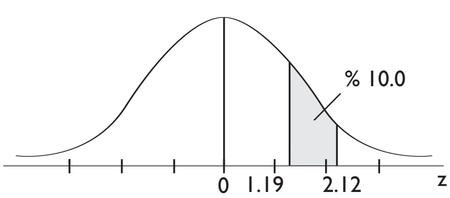
ÖRNEK 4

Standart normal eğri altındaki aşağıda verilen olasılıkları bulunuz.
a) $P(1.19 < z < 2.12)$
b) $P(-1.56 < z < 2.31)$
c) $P(z > -0.75)$

ÇÖZÜM

a) Aşağıdaki şekilde de görülebileceği gibi, $P(1.19 < z < 2.12)$ olasılık değeri $z = 1.19$ ile $z = 2.12$ noktaları arasındaki alandır.
İstenen olasılık değerini (alanı) bulmak için önce ortalamayla $z = 2.12$ arasındaki alan bulunur: $P(z < 2.12) = 0.4830$. Daha sonra ortalamayla $z = 1.19$ arasındaki alan bulunur: $P(z < 1.19) = 0.3830$. Bu değer ilk bulunan değerden çıkartılmak suretiyle istenen (alan) olasılık değeri elde edilir.

$$P(1.19 < z < 2.12) = 0.4830 - 0.3830 = 0.1000$$



***Şekil 6.21** P (1.19 < z < 2.12) değeri.*

(*Dikkat:* Eğer verilen her iki nokta da ortalamanın bir tarafındaysa, ilk olarak ortalamayla bu noktalar arasındaki alanlar bulunur. Daha sonraysa küçük alandan büyük alandan çıkartılır.)

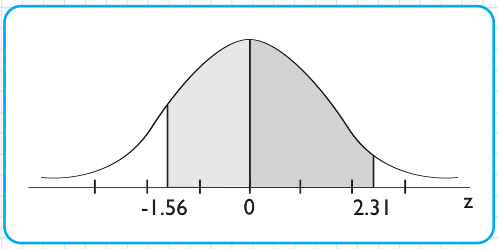
b) Burada istenilen olasılık değeri, standart normal eğri altında z = -1.56 ile z = 2.31 noktaları arasındaki alandır. Aşağıdaki şekilden de görüleceği gibi verilen iki nokta, ortalamanın iki farklı tarafındadır. Bu nedenle istenen olasılık (alan) değerinin bulunabilmesi için önce ortalamaysa z = -1.56 ve z = 2.31 arasındaki alanlar bulunur, daha sonraysa bu alanlar toplanır.

-1.56 ile 0 arasındaki alan $P(-1.56 \leq z \leq 0) = 0.4406$

0 ile 2.31 arasındaki alan $P(0 \leq z \leq 2.31) = 0.4896$

-1.56 ile 2.31 arasındaki alan $P(-1.56 \leq z \leq 2.31) = 0.4406 + 0.4896 = 0.9302$

(*Dikkat:* Eğer verilen iki nokta ortalamanın farklı taraflarındaysa ilk olarak ortalamayla bu noktalar arasındaki alanlar bulunur. Sonra bu iki alan toplanır.)

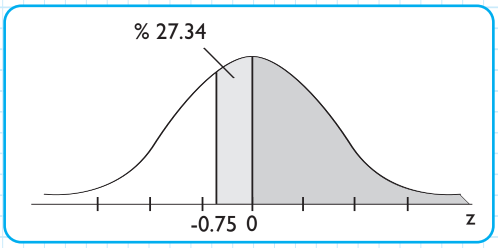


Şekil 6.22
P (-1.56≤ z≤ 2.31) değeri.

c) $P(z > -0.75)$ değeri de verilen noktanın sağındaki tüm alandır. Burada istenen alan iki bölümden oluşmaktadır. İlki, verilen noktayla ortalama arasında kalan alan,

$P(-0.75 \leq z \leq 0) = 0.2734$

ikincisiyse ortalamanın sağındaki (tüm) alandır.



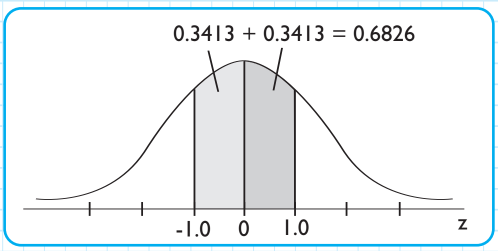
Şekil 6.23
P (z > -0.75) değeri.

$P(z > 0) = 0.50$

Bu iki alan değerinin toplanması sonucunda (0.2734 + 0.5000 = 0.7734) istenen olasılık değeri % 77.34 olarak bulunur.

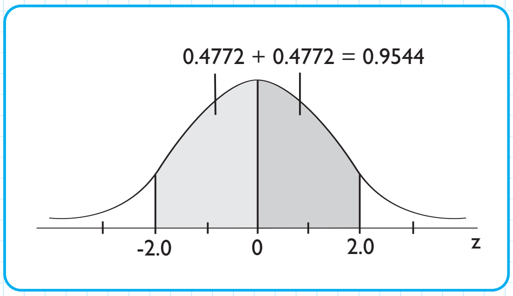
Daha önce de değinilmiş olduğu gibi, standart sapmaya ilişkin üç (ampirik)kuralın doğruluğu, simetriklik özelliği gösteren normal dağılım için, tablo değerinden yararlanılarak gösterilebilir.

1) Ortalamanın bir standart sapma sağında ve solunda kalan noktalar arasındaki alan, toplam alanın % 68.26'sı olarak bulunur. Burada söz edilen alan z = -1.0 den z = 1.0 'e kadar olan alandır. Yandaki şekilden de görüleceği gibi ortalama ile z = 1.0 noktası arasındaki alan % 34.13 'dür. Simetrik bir dağılım olduğu için z = -1.0 ile ortalama arasındaki alan da aynı (%34.13) olacağından toplam alan % 68.26 dır.



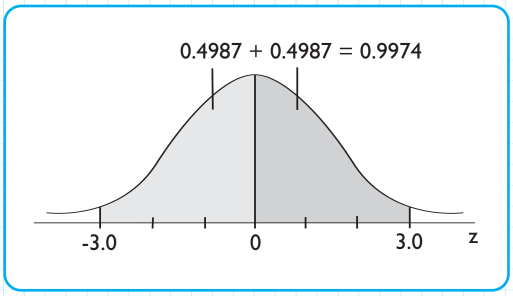
Şekil 6.24 *Bir standart sapma sınırları içerisindeki alan.*

2) Ortalamanın iki standart sapma sağında ve solunda kalan noktalar arasındaki alan, toplam alanın % 95.44'ü olarak bulunur. Burada da (yukarıdakine benzer biçimde), z = -2.0 ile z = 2.0 noktaları arasındaki alanın bulunması için önce ortalamayla z = 2.0 noktası arasında kalan alan bulunur ve bulunan değerin iki katı (simetriklik özelliği) alınarak istenen alan değeri (0.4772 + 0.4772 = 0.9544) elde edilir.



Şekil 6.25 *İki standart sapma sınırları içerisindeki alan.*

3) Ortalamanın üç standart sapma uzaklığındaki sınırlar arasındaki alanı toplam alanın % 99.74'ü olarak bulunur. Bulunması istenen alan z = -3.0 dan z = 3.0'e kadar olan alandır. Ortalamayla z = 3.0 noktası arasındaki alan % 49.87 olduğu için toplam alan % 99.74 olarak bulunur.



Şekil 6.26 *Üç standart sapma sınırları içerisindeki alan.*

Bu özellik nedeniyle, Ek 1.'de verilmiş olan standart normal dağılım tablosunda z = 0 dan z = 3.09'a (ya da z = -3.09'dan z = 0'a)kadar olan değerler için olasılık (alan) değeri bulunabilmektedir.

ÖRNEK 5

Standart normal eğri için aşağıdaki olasılıkları bulunuz.

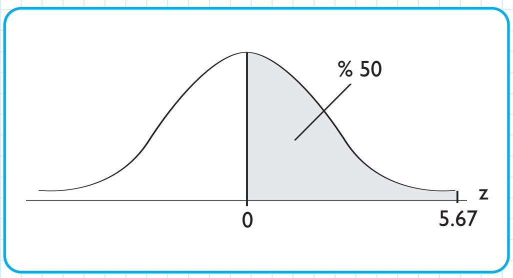
a) P (0 < z < 5.67)

b) P (z < -5.35)

ÇÖZÜM

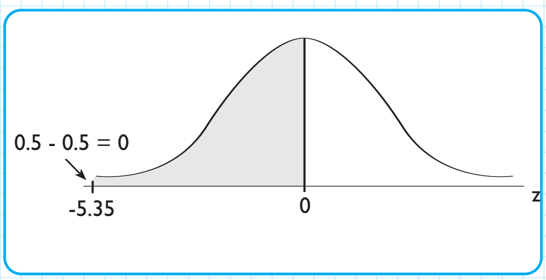
a) Standart normal eğri için istenen bu olasılığın bulunmasında, standart normal dağılım tablosundan yararlanırken bir sorunla karşılaşılmaktadır. Bu sorun tabloda olasılık değeri olarak bulunabilecek en son değerin z = 3.09 olmasıdır. Bu durumda ortalamanın sağında kalan toplam alan (% 50.0) sorunun cevabı olmaktadır.

0 ile 5.67 arasındaki alan = P (0 < z < 5.67)= 0.50



Şekil 6.27 *z = 0 ile z = 5.67 sınırları arasındaki alan.*

b) P (z < -5.35)değeri için de yine standart normal dağılım tablosunun kullanımında benzer sorunla karşılaşılmaktadır. Ortalamanın solundaki toplam alan % 50.0 'dir. z = -5.35 ile ortalama arasındaki alan da yaklaşık % 50.0 'dir. Bu durumda istenen olasılık değeri sıfırdır.



Şekil 6.28 *z = - 5.35 'in solundaki alan.*

- 5.35 ile 0 arasındaki alan = P (z < -5.35)= 0.5 – 0.5 = 0.0

Normal Dağılımın Standartlaştırılması

Yukarıdaki alt bölümlerde, Ek 1.'de verilen standart normal dağılım tablosundan yararlanılarak standart normal eğri altındaki çeşitli alanların bulunması incelendi. Ancak gerçek uygulamalarda, standart normal olmayan sürekli bir rassal değişken, sıfırdan farklı ortalama ve birden farklı standart sapma değeriyle normal dağılım göstermektedir. Böylesi durumlarda Ek 1.'de verilen tablonun kullanılabilmesi için, normal dağılım gösteren sürekli bir rassal değişkenin, bir dönüştürme neticesinde standart normal dağılımlı bir değişkene çevrilmesi gerekmektedir. Bu amaçla normal dağılım gösteren x rassal değişkeninin, standart normal dağılım gösteren z rassal değişkenine dönüştürülmesi yapılacaktır ve bu dönüştürmeye standartlaştırma adı verilmektedir.

x DEĞERİNİN z DEĞERİNE DÖNÜŞTÜRÜLMESİ

x normal dağılım gösteren rassal bir değişkenin herhangi bir değerinin z değeri cinsinden ifadesinde

$$z = \frac{x - \mu}{\sigma}$$

formülünden yararlanılmaktadır. Burada μ ve σ, sırasıyla ilgili normal dağılımının ortalama ve standart sapmasıdır.

Standartlaştırmada önce x rassal değişkeninin ortalama ve standart sapması hesaplanmakta, daha sonraysa x değerinden ortalama çıkartılarak, fark değeri standart sapmaya bölünmektedir.

ÖRNEK 6

x, ortalaması 50 ve standart sapması 10 olan bir normal dağılım göstermektedir. Standartlaştırma formülünden yararlanarak aşağıdaki x değerlerini z değerlerine dönüştürünüz.

a) x = 55

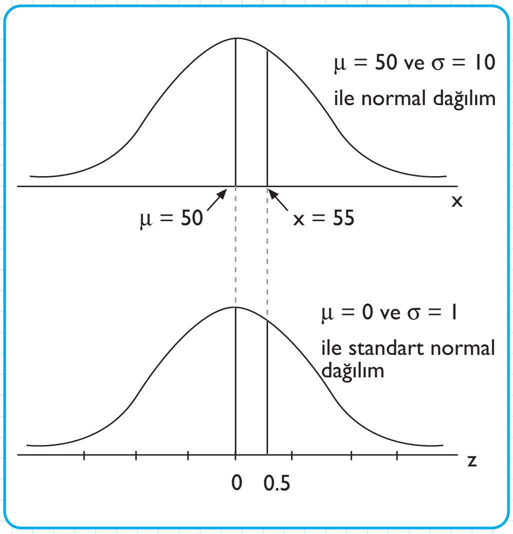
b) x = 35

ÇÖZÜM

a) Yukarıda verilmiş olan formül kullanılarak x = 55 değerinin z cinsinden değeri,

$$z = \frac{x - \mu}{\sigma} = \frac{55-50}{10} = 0.50$$

olarak bulunur. Aşağıdaki şekilden de görüleceği gibi x rassal değişkenin dağılımıyla z rassal değişkenin dağılımı arasında, standart sapmalar açısından fark yoktur. Çünkü x dağılımında x = 55 noktası μ = 50 ortalamasının 1 / 2 standart sapma sağında iken, z = 0.5 değeri yine ortalaması sıfır ve standart sapması bir olan z standart normal dağılımında ortalamanın 1 / 2 standart sapma sağındadır.

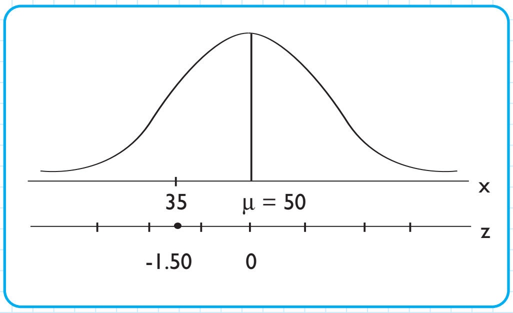


Şekil 6.29 *x = 55'in z değeri*

b) x = 35 değerine karşılık gelen z değeri de yine aynı biçimde,

$$z = \frac{x - \mu}{\sigma} = \frac{35-50}{10} = 1.50$$

olarak bulunur. Bulunan z değeri gibi, verilmiş olan x = 35 değeri de ortalamadan küçük olduğu için ortalamanın solunda bir noktadır. Ancak x değeri negatif olmadığı halde z değeri negatiftir. Bunun nedenide z dönüştürmesinde dağılım ortalamasının sıfır noktasına taşınmış olmasıdır.



Şekil 6.30 *x = 35'in z değeri.*

Normal dağılımlı bir x değişkeninin iki değeri arasındaki alanın bulunması için önce, her iki x değeri de z değerine dönüştürülmekte, daha sonraysa standart normal eğri altındaki iki z değeri arasındaki alan bulunmaktadır. Bu alan, aynı zamanda verilmiş olan x 'ler arasındaki alandır.

ÖRNEK 7

x sürekli rassal değişkeni 25 ortalama ve 4 standart sapmayla normal dağılmaktadır. Aşağıda verilen noktalar arasındaki alanı bulunuz.
a) x = 25 ve x = 32 arası
b) x = 18 ve x = 34 arası

ÇÖZÜM

Verilen normal dağılımda $\mu = 25$ ve $\sigma = 4$ dür.
a) Burada ilk adım, verilmiş olan x = 25 ve x = 32 değerlerinin standart normal z değerlerine dönüştürülmesidir.

$$z = \frac{x - \mu}{\sigma} = \frac{25-25}{4} = 0.00$$

$$z = \frac{x - \mu}{\sigma} = \frac{32-25}{4} = 1.75$$

İkinci adım z = 0.00 ile z = 1.75 noktaları arasındaki alanın Ek 1.'de verilmiş olan tablodan bulunmasıdır. Bu değer,

$$P(25 < x < 32) = P(0 < z < 1.75) = 0.4599$$

olarak bulunur.

b) x = 18 ve x = 34 değerleri de yine standart normal değerlere dönüştürülür.

$$x = 18 \text{ için } z = \frac{18\text{-}25}{4} = -1.75$$

$$x = 34 \text{ için } z = \frac{34\text{-}25}{4} = -2.25$$

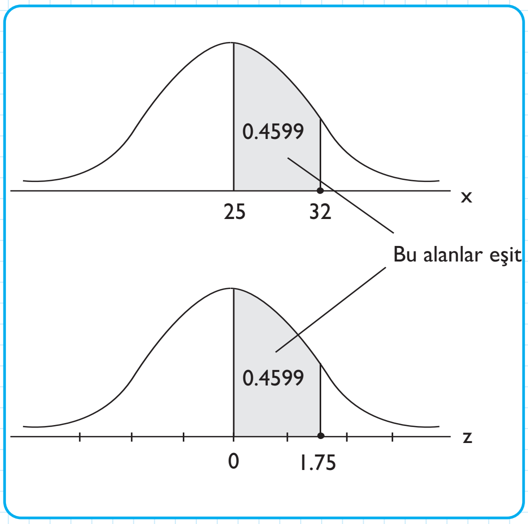
olarak hesaplanır.

Daha sonraysa aşağıdaki şekilden de görüleceği gibi z = -1.75 ile z = 2.25 noktaları arasındaki alan bulunur. Burada verilen noktalar, ortalamanın solunda ve sağında olduğu için, toplam alan, bulunacak iki alanın toplamından oluşacaktır.

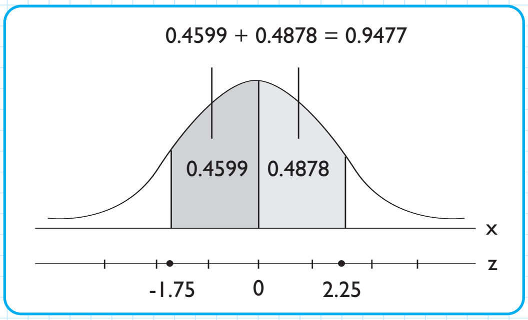
$$P(18 < x < 25) = P(-1.75 < z < 0) = 0.4599$$

$$P(25 < x < 34) = P(0 < z < 2.25) = 0.4878$$

$$P(18 < x < 34) = 0.4599 + 0.4878 = 0.9477$$



Şekil 6.31 *x = 25 ve x = 32 arasındaki alan.*



Şekil 6.32 *x = 18 ve x = 34 arasındaki alan.*

ÖRNEK 8

X rassal değişkeni, 40 ortalama ve 5 standart sapmayla normal dağılım göstermektedir. Aşağıdaki olasılıkları normal dağılım için bulunuz.

a) P (x > 55)

b) P (x < 49)

ÇÖZÜM

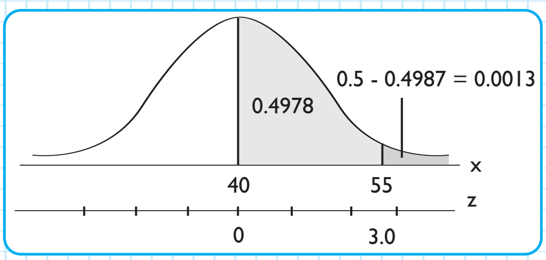
Normal dağılımda μ = 40 ve σ = 5 olarak verilmiştir.

a) x rassal değişkenin 55'den büyük değer alması bu şıkta istenen olasılık, normal dağılım eğrisinden (Ek 1.'de verilmiş standart normal dağılım tablosundan yararlanılarak) bulunacaktır. Bunun için x = 55 değerinin z değerine dönüştürülmesi gerekir.

$$x = 55 \text{ için } z = \frac{55\text{-}40}{5} = -3.00 \ ; \ P(0 < z < 3.00) = 0.4987$$

Bu değer, yandaki şekilden de görülebileceği gibi, eğrinin sağ uç noktasındaki küçük bir alanı dışta bırakmaktadır ve burası istenen olasılıktır.

Sonuç olarak istenen olasılık değeri,



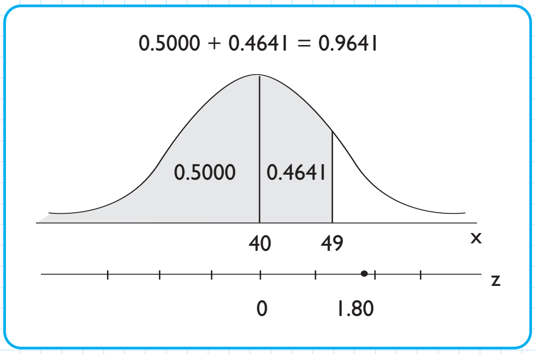
Şekil 6.33 *P (x > 55) değeri.*

$$P(x > 55) = P(z > 3.00) = 0.5 - 0.4987 = 0.0013$$

olarak bulunur.

b) Burada da yine x = 49 değerinin z değerine dönüştürülmesinden,

$$x = 49 \text{ için } z = \frac{49-40}{5} = 1.80$$



Şekil 6.34 *P (x < 49) değeri.*

elde edilir. Yandaki şekilde de görülebileceği gibi burada x değeri için üst sınır verilmiş, alt sınır için hiçbir kısıt getirilmemiştir. Bu nedenle, ortalamanın sol tarafındaki alan (0.5) ve z = 0 ile z = 1.80 arasında kalan alan toplanarak, toplam alan (olasılık) elde edilmektedir.

$$P(x < 49) = P(z < 1.80) = 0.50 + 0.4641 = 0.9641$$

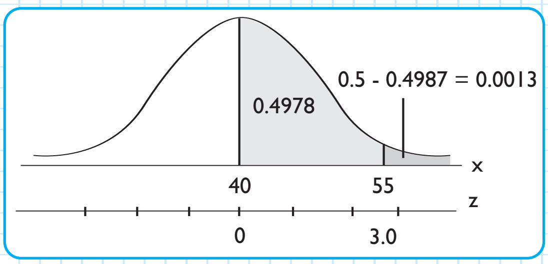
ÖRNEK 9 ***x sürekli rassal değişkeni $\mu = 50$ ve $\sigma = 8$ ile normal dağılım göstermek üzere, $P(30 \leq x \leq 39)$ değerini bulunuz.***

ÇÖZÜM

$\mu = 50$ ve $\sigma = 8$ olmak üzere, verilen x = 30 ve x = 39 değerlerinin (her ikisi de ortalamanın solunda) z cinsinden değerleri,

$$x = 30 \text{ için } z = \frac{30-50}{8} = -2.50$$

$$x = 39 \text{ için } z = \frac{39-50}{8} = -1.38$$



Şekil 6.35 *P (30 ≤ x ≤ 39) değeri.*

olarak bulunur. Ortalamayla bu iki nokta arasındaki alanlar bulunarak, büyük alandan küçük alanın çıkartılmasıyla istenen olasılık (alan) değeri elde edilir.

$$P(30 \leq x \leq 39) = P(-2.50 \leq z \leq -1.38) = 0.4938 - 0.4162 = 0.0776$$

ÖRNEK 10

x sürekli rassal değişkeni $\mu = 80$ ve $\sigma = 12$ ile normal dağılımını göstermek için, normal dağılım eğrisi altında kalan aşağıdaki alanları bulunuz.

a) x = 70 den x = 135'e kadar
b) x = 27'nin sol tarafı

ÇÖZÜM

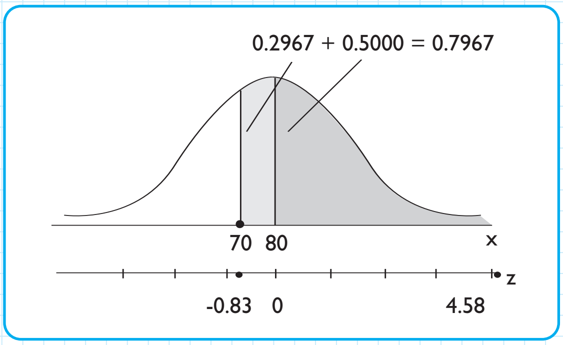
Verilen x noktalarının z cinsinden değerleri,

$$x = 70 \text{ için } z = \frac{70-80}{12} = -0.83$$

$$x = 135 \text{ için } z = \frac{135-80}{12} = 4.58$$

olarak bulunur. Bu noktalar ortalamanın iki tarafında yer aldığı için, tablodan elde edilecek alan (olasılık) değerleri toplanır.

$P\ (70 \le x \le 135\ )$
$= P\ (-\ 0.83 \le z \le 4.58)$
$= 0.2967 + 0.5000$
$= 0.7967$



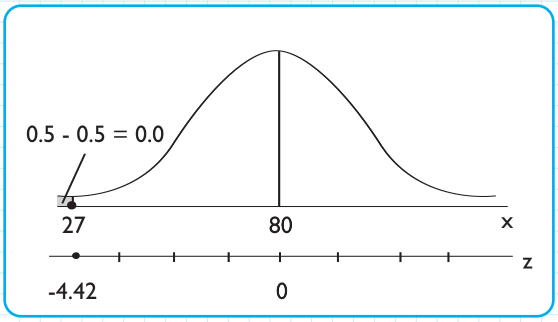
***Şekil 6.36** x = 70 ile x = 135 arasındaki alan.*

b) x = 27'nin sol tarafındaki alanı ya da x 'in 27'den küçük değer alma olasılığını bulmak için z değeri bulunarak sonuca gidilir.

x = 27 için

$$z = \frac{27-80}{12} = -4.42$$

$P\ (x < 27) = P\ (z < -\ 4.42)$
$= 0.5 - P\ (-\ 4.42 < z < 0)$
$= 0.5 - 0.5 = 0.0$



***Şekil 6.37** x = 27' nin solundaki alan.*

Normal Dağılım Uygulamaları

Buraya kadarki alt bölümlerde; normal dağılımdan, normal dağılımın standart normal dağılıma dönüştürülerek, normal dağılım eğrisi altındaki alanın bulunmasından söz edildi. Aşağıdaysa normal dağılımın gerçek verilerle uygulanması konu edilecektir.

ÖRNEK 11

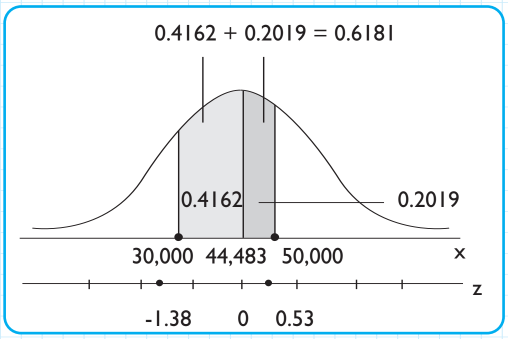
ZENGİN ülkesinde aile başına düşen gelir 44.483 dolar ortalama ve 10.500 dolar standart sapmayla normal dağılım göstermektedir. Bu ülkeden rassal seçilen bir ailenin 30.000 – 50.000 dolar arasında gelire sahip olma olasılığını bulunuz.

ÇÖZÜM

x rassal değişkeni, μ = 44.483 ve σ = 10.500 dolar değerleriyle normal dağılmaktadır. Aşağıdaki şekilden de görüleceği gibi verilen sınır değerleri, ortalamanın solunda ve sağında yer almaktadır. Ortalamayla sınır değerleri arasındaki bu olasılıklar (alan), z dönüşümüyle elde edilir.

Bulunan z değerleri kullanılarak standart normal dağılım tablosundan gerekli olasılık değerlerine ulaşılır ve sonuç bulunur.

P (30.000 ≤ x ≤ 44.483) = P (- 1.38 ≤ z ≤ 0 = 0.4162
P (44.483 ≤ x ≤ 50.000) = P (0 ≤ z ≤ 0.53) = 0.2019
P (30.000 ≤ x ≤ 50.000) = P (-1.38 ≤ z≤ 0.53) = 0.6181



Şekil 6.38
x = 30,000 – 50.000 arasındaki alan.

Sonuç olarak rassal seçilen bir ailenin gelirinin 30.000 – 50.000 dolar arasında olma olasılığı % 61.81 dir.

ÖRNEK 12

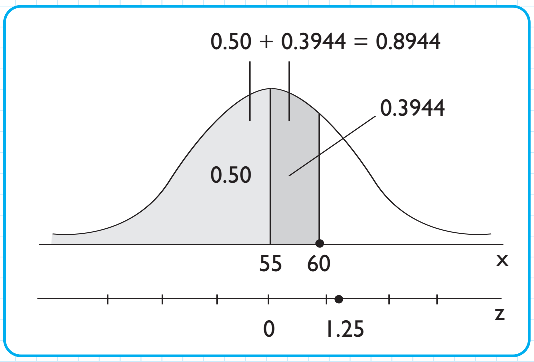
Oyuncak üreten bir firmada, bir işçinin, oyuncak bir yarış otomobilini monte etme süresi; 55 dakika ortalama ve 4 dakika standart sapmayla normal dağılmaktadır. İşyerinin saat 17:00'de kapandığı bilindiğine göre, saat 16:00'da montaja başlayan bir işçinin kapanış saatine kadar işi bitirme olasılığını bulunuz.

ÇÖZÜM

Oyuncak bir yarış otomobilinin montaj süresi; μ = 55 dakika ve σ = 4 dakikayla normal dağılım göstermektedir. Aşağıdaki şekilden de görüleceği gibi, burada istenen, montajın en çok 60 dakikada bitmesidir.
Standart normal dağılım tablo değerinden yararlanabilmek için gerekli z değeri,

$$x = 60 \text{ için } z = \frac{60-55}{4} = 1.25$$

olarak bulunduktan sonra istenen olasılık hesaplanır. Bunun için önce ortalamanın solundaki tüm alan (0.5) ve ortalamanın sağındaki z = 1.25'e kadar olan alan (0.3944) toplanır ve sonuç değeri,



Şekil 6.39 *x = 60 noktasının solundaki alan.*

$$P(x \leq 60) = P(z \leq 1.25) = 0.5 + 0.3944 = 0.8944$$

elde edilir.

ÖRNEK 13

Meşrubat üreten bir firmada, üretilen sodaların 12 cl olması gerekmektedir. Ancak otomatik makinelerce yapılan şişelemede, şişe içerisindeki soda miktarı bazen 12 cl'den çok ya da az olabilmektedir. Firmaca üretilen sodaların incelenmesinde şişe içindeki soda miktarının 12 cl ortalama ve 0.015 cl standart sapmayla normal dağıldığı bulunmuştur.

a) Rassal seçilen bir şişe içerisindeki soda miktarının 11.97 ile 11.99 cl arasında olma olasılığı nedir ?
b) Şişelerin 12.02 ile 12.07 cl arasında, soda bulundurma yüzdesi nedir?

ÇÖZÜM

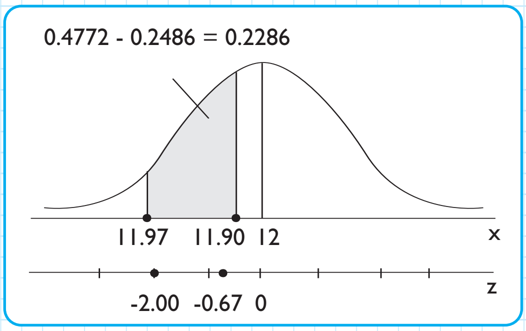
X rassal değişkeni şişe içerisindeki soda miktarını göstermekte, μ = 12 cl ve σ = 0.015 cl ile normal dağılmaktadır.

a) Daha önceki örneklerde olduğu gibi ilk olarak verilen x değerleri z değerlerine dönüştürülür.

$$x = 11.97 \text{ için } z = \frac{11.97 - 12.00}{0.015} = -2.00$$

$$x = 11.99 \text{ için } z = \frac{11.99 - 12.00}{0.015} = -0.67$$

Aşağıdaki şekilden de görüleceği gibi, ortalamanın solundaki iki alanın farkı alınarak, istenen olasılık değeri bulunur.



Şekil 6.40
x = 11.97 ile
x = 11.99
arasındaki alan.

$$P(11.97 \le x \le 12.00) = P(-2.00 \le z \le 0.00) = 0.4772$$
$$P(11.99 \le x \le 12.00) = P(-0.67 \le z \le 0.00) = 0.2486$$
$$P(11.97 \le x \le 11.99) = P(-2.00 \le z \le -0.67) = 0.2286$$

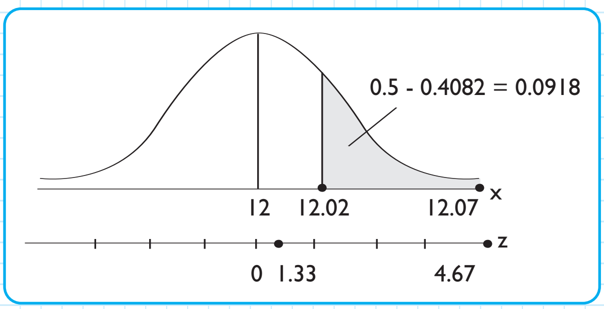
Sonuç olarak şişelerin % 22.86'sındaki soda miktarı, 11.97 cl ile 11.99 cl arasındadır.

b) Yukarıdaki şıkka benzer biçimde z değerleri,

$$x = 12.02 \text{ için } z = \frac{12.02 - 12.00}{0.015} = 1.33$$

$$x = 12.07 \text{ için } z = \frac{12.07 - 12.00}{0.015} = 4.67$$

olarak bulunur ve aşağıdaki şekilden de anlaşılacağı gibi, ortalamanın sağındaki iki alan değerinin farkı alınarak, sorunun cevabı elde edilir.



Şekil 6.41
x = 12.02 ile
x = 12.07
arasındaki alan.

$P\ (12.00 \le x \le 12.02) = P\ (0.00 \le z \le 1.33) = 0.4082$

$P\ (12.00 \le x \le 12.07) = P\ (0.00 \le z \le 4.67) = 0.5000$

$P\ (12.02 \le x \le 12.07) = P\ (1.33 \le z \le 4.67) = 0.0918$

Sonuç olarak şişelerin % 9.18'inde 12.02 cl ile 12.07 cl soda vardır.

ÖRNEK 14

Hesap makinesi üreten bir firma, ürettiği aletlerin ömürlerinin 54 ay ortalama ve 8 ay standart sapmayla normal dağılım gösterdiğini bulmuştur. Firma yetkilileri bir kampanya başlatarak ilk 36 ay içerisinde arızalanan hesap makinelerini yenisiyle değiştirme yükümlülüğüne girmiştir. Rassal satılan bir hesap makinesinin yenisiyle değiştirilme olasılığı nedir?

ÇÖZÜM

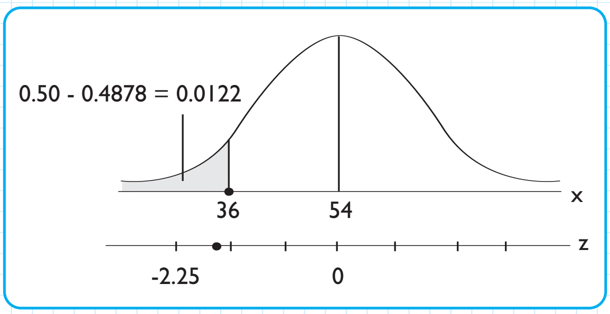
Bu örnekte $\mu = 54$ ve $\sigma = 8$ aydır. Verilmiş olan $x = 36$ ay değerinin z cinsinden değeri,

$$x = 36 \text{ için } z = \frac{36 - 54}{8} = -2.25$$

dir. Aşağıdaki şekilden de anlaşılacağı gibi,

$$P\ (x < 36) = P\ (z < -2.25) = 0.50 - 0.4878 = 0.0122$$

değerinden satılan bir hesap makinesinin yenisiyle değiştirilme olasılığının % 1.22 olduğu bulunur.



Şekil 6.42
x = 36 'nın solundaki alan.

SIRA SİZDE

1. X rassal değişkeni bir maratonun koşulma süresi olmak üzere, 195 dakika ortalama ve 21 dakika standart sapmayla normal dağılım göstermektedir. Bu maratona katılan bir atlet rassal seçildiğinde bu atletin maratonu;
a) 150 dakikadan önce tamamlamış olma,
b) 205 ile 245 dakika arasında tamamlamış olma,
olasılıklarını bulunuz.

SIRA SİZDE

2. ABD'de ilköğretim okullarında görev yapan öğretmenlerin yıllık maaşları; 35.104 dolar ortalama ve 3.200 dolar standart sapmayla normal dağılım göstermektedir. Bu gruptan rassal seçilen bir öğretmenin,
a) 38.700 dolardan çok kazanması,
b) 32.625 ile 38.830 dolar arasında kazanması,
olasılıklarını bulunuz.

3. İstanbul-Ankara otoyolunu kullanan özel otomobillerin hızları saatte; 69 mil ortalama ve 3.5 mil standart sapmayla normal dağılmaktadır. Bu otoyolu kullanan özel otomobillerden;
a) Saatte 61 – 66 mil yapanların,
b) Saate 65 – 74 mil yapanların,
yüzdelerini bulunuz.

NORMAL DAĞILIM EĞRİSİ ALTINDAKİ ALAN BİLİNİYORKEN z VE x DEĞERLERİNİN BULUNMASI

AMAÇ 2

Normal dağılım eğrisi altındaki alan biliniyor Z ve X değerlerini belirleyebileceksiniz.

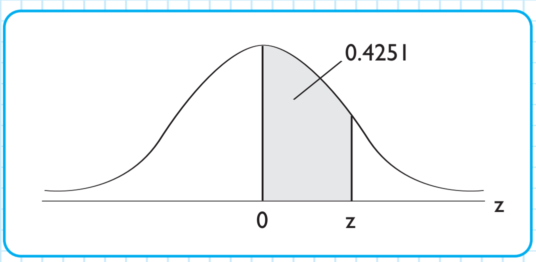
Buraya kadar ele alınan kesimlerde, verilen bir z ya da x aralığına düşen (normal dağılım eğrisi altındaki) alanın bulunması konu edildi. Burada ise tersi bir durum ele alınacak; normal dağılım eğrisi altındaki alan biliniyorken, bu alanı sınırlayan z ya da x değerlerinin bulunması ele alınacaktır.

ÖRNEK 15 ***Standart normal eğri altında 0 ile bir z değeri arasında kalan alan 0.4251 dir. Bu alanın üst sınırı olan pozitif z değerini bulunuz.***

ÇÖZÜM

Aşağıdaki şekilde de görüleceği gibi, verilen alan, sağ tarafta 0 ile z sınırları arasındadır ve değeri 0.4251 dir.

z değerinin bulunmasında da yine, Ek 1.'de verilen standart normal dağılım tablosundan yararlanılacaktır. Yukarıdaki örneklerde verilen, bir z değerine karşılık gelen alanın bulunmasında, tablonun ilk sütunundaki tam sayıyla kesir noktasının sağındaki ilk sayıya ve tablonun ilk satırından ya kesir noktasının sağındaki ikinci sayıya bakılıyordu. Burada ya önce, tablo içerisinden verilen alan değeri bulunacak, daha sonra ya da o değerin bulunduğu satır ve sütun başlarındaki z değeri yazılarak istenen z değeri elde edilecektir.



Şekil 6.43
z değerinin bulunması.

Bu örnekte 0.4251 değeri, tablo içerisinden bulunduktan sonra, bu değerin satır başında 1.4, sütun başındaysa 0.04 değerleri kaydedilir. Bu iki değerin birleştirilmesi sonucunda aranan z değeri (1.44 olarak) elde edilir.

ÖRNEK 16

Standart normal dağılım eğrisi altında, sağ uçtaki (kuyruk) 0.005 alanını belirleyen z değerini bulunuz.

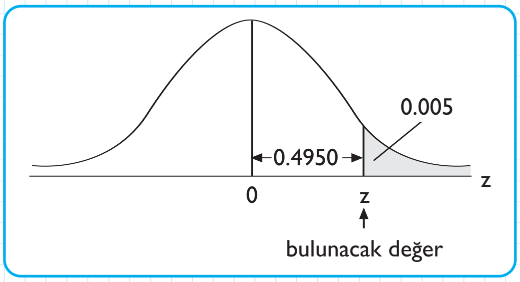
ÇÖZÜM

Burada verilen alan, sağ uçta yer alan bir alandır. Yani alanın üst sınırı belirsiz (sonsuz) olacağı için, alt sınırın bulunması gerekir. Ancak, daha önce de belirtildiği gibi, Ek 1.'de verilen normal dağılım tablosundaki alan değerleri, ortalama (z = 0) ile verilen z değerleri arasındadır. Bu nedenle öncelikle dağılımın sağında yer alan bu noktayla z = 0 noktası arasındaki alanın bilinmesi gerekir. Normal dağılımın simetrik olması ve ortalamanın sağındaki toplam alanın 0.5 olduğu bilgilerinden hareketle, 0 ile z noktası arasındaki alan,

$$0.5000 - 0.0050 = 0.4950$$

olarak bulunur.

Tablo içerisindeki alan değerleri arasından 0.4950 değeri yaklaşık olarak (0.4949 ve 0.4951)belirlenir ve ilişkin z değeri 2.58 (ya da 2.57)olarak elde edilir. Sonuç olarak z = 2.58 noktasının sağında kalan alan, toplam alanın 0.005'idir.



Şekil 6.44
z değerinin bulunması.

ÖRNEK 17

Standart normal eğri altında sol uçtaki 0.05 alanının z değerini bulunuz.

ÇÖZÜM

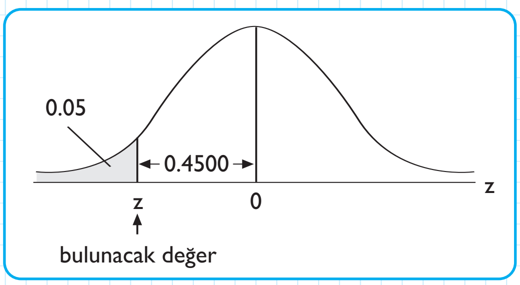
Bir önceki örnekte olduğu gibi, burada da ortalamanın sol ucunda yer alan bir alan verilmiştir. Bu alanın sol tarafındaki nokta belirsiz olup sağ tarafındaki noktanın bulunması gerekir. Yine ortalamayla aranan nokta arasındaki alan bulunarak çözüme ulaşılacaktır.

Söz konusu alan,

$$0.500 - 0.050 = 0.450$$

dir ve bu alanı soldan sınırlayan nokta, (z değeri)–1.65 olarak (yaklaşık değer) elde edilir.

Önceki alt bölümlerde, standart normal olmayan, normal dağılım eğrisi altında verilen x noktaları arasında kalan alanın bulunmasında, μ ve σ değerlerinden yararlanılarak verilen x değerine karşılık gelen z değerleri bulunmuş ve standart normal dağılım tablosu kullanılarak çözüme ulaşılmıştı. Buradaysa ise z = (x - μ) / σ eşitliği kullanılarak, yani, μ ve σ değerleri bilinen ve alan değeri verilmiş normal dağılımın x değerinin bulunması ele alınacaktır.



Şekil 6.45
z değerinin bulunması.

NORMAL DAĞILIM İÇİN BİR x DEĞERİNİN BULUNMASI

μ ve σ değerleri, bilinen bir normal eğri altında, ortalamayla x noktası arasındaki alan için x değeri,

$$x = \mu + z\,\sigma$$

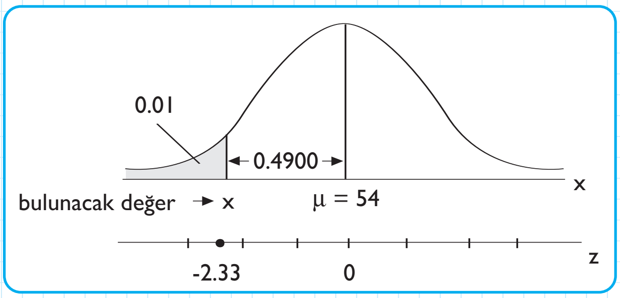
biçiminde bulunur.

ÖRNEK 18

Bir firma tarafından üretilen hesap makinelerinin ömürleri 54 ay ortalama ve 8 ay standart sapmayla normal dağılmaktadır. Firma yetkilileri satışları artırmak için bir garanti süresi uygulamak istemektedir. Ancak garanti kapsamında değiştirilecek hesap makinesinin, toplam satışın % 1'inden daha fazla olmasını da istenmemektedir. Garanti uygulanacak bu süreyi bulunuz.

ÇÖZÜM

Aşağıdaki şekilde de görüleceği gibi, normal dağılımın parametreleri (μ = 54 ay ve σ = 8 ay olarak) bilinmektedir. Burada istenen, hesap makinesi satın alanlara uygulanacak garanti süresinin bulunmasıdır.



Şekil 6.46
z değerinin bulunması.

İlk olarak ortalamayla verilen nokta arasındaki alan 0.5000 – 0.0100 = 0.4900 elde edilir. Ek 1.'de verilen tablodanda bu alan değerine karşılık gelen z = – 2.33 değeri bulunur ve yukarıda verilmiş olan eşitlik kullanılarak aranan x değeri,

$$x = \mu + z\,\sigma = 54 + (-2.33)(8) = 54 - 18.64 = 35.36$$

elde edilir. Sonuç olarak üretici firma (yaklaşık olarak) 35 aydan önce kullanım dışı kalan hesap makinelerini garanti kapsamına alıp, yenisiyle değiştirmeyi kabul ederse, toptan satışın sadece % 1'ini değiştirecektir.

ÖRNEK 19

Bilgin Fen Lisesinin, öğrenci seçmek amacıyla, geçen yıllarda uyguladığı sınav sonuçlarının 904 puan ortalama ve 153 puan standart sapmayla normal dağıldığı ve bu lisenin yöneticilerinin her yıl yapılan sınav sonucunda ilk % 10'a giren öğrencilerin kaydını yaptığı bilinmektedir. Bilgin Fen Lisesi sınavına hazırlanan Arif Hırslı'nın (puanların geçen yıllarla aynı olacağı varsayımı altında)bu liseye girebilmesi için en az kaç puan alması gerekir?

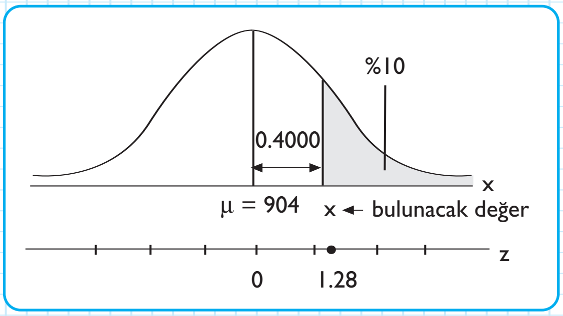
ÇÖZÜM

x rassal değişkeni, $\mu = 904$ ve $\sigma = 153$ parametreleriyle normal dağılım göstermektedir. Burada istenen, ortalamanın sağ ucunda yer alacak % 10 'luk alanın alt sınırının belirlenmesidir. Bu sınırla ortalama arasındaki alan 0.5000 – 0.1000 = 0.4000 olacağından bu alana karşılık gelen z değeri (tablodan) 1.28 bulunur.

Bu değer eşitlikte yerine konarak, aranan x değeri,

$$x = \mu + z\,\sigma = 904 + (1.28)\,(153)$$
$$= 904 + 195.84 \cong 1100$$

olarak bulunur. Yani, Arif Hırslı'nın Bilgin Fen Lisesine girebilmesi için, en az 1100 puan alması gerekmektedir.



Şekil 6.47
z değerinin bulunması.

SIRA SİZDE

1. **Standart normal eğri altında aşağıdaki alanları veren z değerlerini bulunuz.**
   a) **0 ile arasında 0.4772 alan olan pozitif z değeri.**
   b) **0 ile arasında 0.4785 alan olan negatif z değeri.**
   c) **Sol uçtaki 0.3565'lik alan.**
   d) **Sağ uçtaki 0.1530'luk alan.**

2. **Standart normal eğri altında, aşağıdaki alanları veren z değerlerini bulunuz.**
   a) **Sağ uçtaki 0.0500'lik alan.**
   b) **Sol uçtaki 0.0250'lik alan.**
   c) **Sol uçtaki 0.0100'lik alan.**
   d) **Sağ uçtaki 0.0050'lik alan.**

3. **X sürekli rassal değişkeni 200 ortalama ve 25 standart sapmayla normal dağılmaktadır.**
   a) **Normal eğri altında, solunda yaklaşık 0.6330'luk alan bulunan x değerini bulunuz.**
   b) **Normal eğri altında, sağında yaklaşık 0.05'lik alan bulunan x değerini bulunuz.**
   c) **Normal eğri altında, sağında 0.8051'lik alan bulunan x değerini bulunuz.**
   d) **Normal eğri altında, solunda 0.0150'lik alan bulunan x değerini bulunuz.**
   e) **Normal eğri altında, m ile arasında 0.4525'lik alan bulunan m 'den küçük x değerini bulunuz.**
   f) **Normal eğri altında, m ile arasında 0.4800'lik alan bulunan m 'den büyük x değerini bulunuz.**

BİNOM DAĞILIMINA NORMAL DAĞILIM YAKLAŞIMI

AMAÇ 3 *Binom ve normal dağılım arasındaki ilişkiyi açıklayabileceksiniz.*

Bir önceki bölümde binom dağılımıyla ilgili verilenler hatırlanacak olursa,

1. Binom dağılımı kesikli rassal değişkenlere uygulanmaktadır,
2. Binom deneyinin her tekrarında; başarı ve başarısızlık olarak adlandırılan iki olası sonuç bulunmaktadır,
3. İki olası sonuca ilişkin olasılıklar her tekrar için aynıdır,
4. Denemeler birbirinden bağımsızdır,

biçimindedir n denemede x başarılı sonuç elde etme olasılığı, aşağıdaki binom formülüyle hesaplanmaktadır.

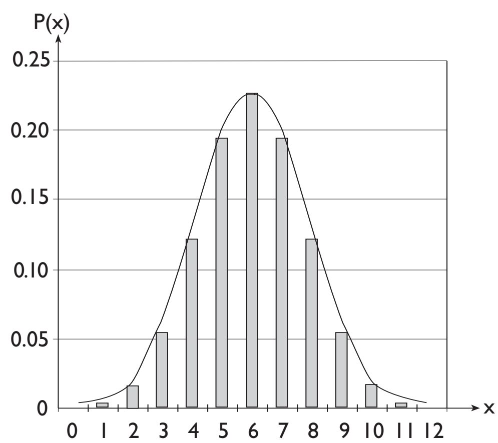
$$P(x) = \binom{n}{x} p^x q^{n-x}$$

n deneme sayısının çok büyük olması durumunda, binom formülüyle olasılık hesaplamak oldukça güç olduğundan, böylesi durumlarda, binom formülüyle bulunacak kesin olasılık değerleri yerine, normal dağılımdan yararlanarak, yaklaşık olasılıklar elde edilebilmektedir. Ancak, n çok büyük ve p olasılık değerinin de 0.5'e yakın olması durumunda, normal dağılımdan elde edilen yaklaşık olasılık değerleri de kesin değerlere eşit olmaktadır. Çünkü $p = 0.5$ olması durumunda binom dağılımı da simetrik olduğundan, yine simetrik bir dağılım olan normal dağılıma, daha çok benzeyecektir.

BİNOM DAĞILIMINA NORMAL DAĞILIM YAKLAŞIMI

$n \cdot p > 5$ ve $n \cdot q > 5$ olması durumunda, binom dağılımına yaklaşım amacıyla, normal dağılım kullanılabilmektedir.

Aşağıdaki tablo ve şekilden de görüleceği gibi $n = 12$ ve $p = 0.5$ ($n \cdot p > 5$ ve $n \cdot q > 5$)için binom dağılımının biçimi, normal dağılma çok yakın olmaktadır.



Şekil 6.48 *n = 12 ve p = 0.5 için binom olasılık dağılımının histogramı.*

Tablo 6.2 *n = 12 ve p = 0.5 için Binom Olasılık Dağılımı*

x	P (x)
0	0.0002
1	0.0029
2	0.0161
3	0.0537
4	0.1208
5	0.1934
6	0.2256
7	0.1934
8	0.1208
9	0.0537
10	0.0161
11	0.0029
12	0.0002

ÖRNEK 20

Bir araştırmaya göre Ankara'da yaşayan erişkinlerin % 50'sinin en az bir kredi kartı bulunmaktadır. Bu gruptan rassal seçilen 30 erişkinden 19 tanesinde en az bir kredi kartı bulunması olasılığı nedir?

ÇÖZÜM

Bu soruda istenen olasılık değeri binom formülü kullanılarak bulunabilir. Çünkü örnekte,

$n = 30, \; p = 0.5; \; x = 19, \; q = 1 - p = 1 - 0.5 = 0.5$ ve $n - x = 30 - 19 = 11$

dir. Bu değerler formülde yerine konacak olursa,

$$P(x = 19) = \binom{30}{19}(0.5)^{19}(0.5)^{11} = 0.0509$$

değeri elde edilir. Ayrıca bu örnekte $n\,p = 30\,(0.5) = 15$ ve $n\,q = 30\,(0.5) = 15$ olduğu için, istenen olasılık değeri normal dağılım yardımıyla da elde edilebilmektedir. Bu yaklaşımda üç aşama izlenmektedir.

Aşama 1. Binom dağılımı için μ ve σ 'nin hesaplanmasıdır.
Normal dağılımın kullanılabilmesi için dağılımın ortalama ve standart sapmasının bilinmesi gerekir. Binom dağılımı için bu parametreler,

$$\mu = n\,.\,p = 30\,(0.5) = 15$$

$$\sigma = \sqrt{n\,.\,p\,.\,q} = \sqrt{(30)(0.5)(0.5)} = 2.7386$$

biçiminde bulunur.

Aşama 2. Kesikli rassal değişkenin sürekli bir rassal değişkene dönüştürülmesidir. Binom dağılımı kesikli rassal değişkenlere, normal dağılıma sürekli değişkenlere uygulanan dağılımlardır. Bu nedenle binom dağılımının normal dağılıma yaklaşımını sağlamak amacıyla "süreklilik düzeltmesi" yapılmaktadır.

SÜREKLİLİK İÇİN DÜZELTME FAKTÖRÜ
Normal dağılımın binom dağılımına yaklaşımını sağlamak için n denemede x başarılı sonuç sayısına $\pm\, 0.5$ değeri eklenir.

Verilen örnekte, $x = 19$ olduğu için, süreklilik düzeltmesi sonucunda 18.5 ve 19.5 değerleri elde edilir. Bu durumda, binom dağılımında $P\,(x = 19)$olasılık değeri yerine, normal dağılımda $P\,(18.5 \leq x \leq 19.5)$ olasılık değeri bulunacaktır.

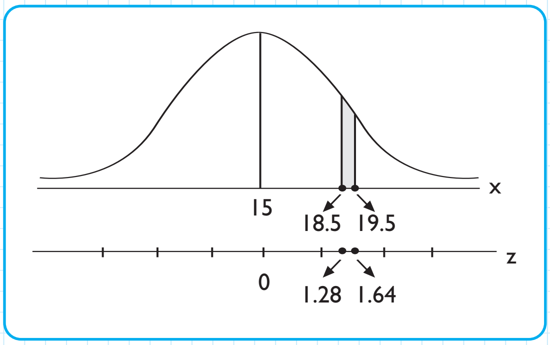
Aşama 3. Normal dağılım kullanılarak istenen olasılığın hesaplanmasıdır.
Daha önceki örneklere benzer biçimde standart normal dağılım tablosundan yararlanabilmek için, x sınır değerlerine karşılık gelen z değerlerinin bulunması gerekir.

$$x = 18.5 \text{ için } z = \frac{18.5 - 15}{2.7386} = 1.28$$

$$x = 19.5 \text{ için } z = \frac{19.5 - 15}{2.7386} = 1.64$$

z değerinin bulunmasının ardından iki alan değeri bulunur ve büyük alandan küçük alan çıkartılarak aranan olasılık değerine ulaşılır.

$P(0 \leq z \leq 1.64) = 0.4495$;
$P(0 \leq z \leq 1.28) = 0.3997$
$P(1.28 \leq z \leq 1.64) = 0.4495 - 0.3997 = 0.0498$



Şekil 6.49 *x = 18.5 ve x = 19.5 arasındaki alan.*

Normal dağılım yaklaşımı sonucunda elde edilen (yaklaşık) olasılık değeriyle binom formülünden elde edilmiş olan kesin olasılık değerleri arasında (0.0509 – 0.0498 = 0.0011) çok küçük bir fark bulunmaktadır ve bu fark da ihmal edilebilecek düzeydedir. Süreklilik düzeltmesi, hep normal dağılım yaklaşımının kullanılmasında uygulanmaktadır. Yukarıda eşitlik durumda süreklilik verilmişti. Ancak; bazen binom dağılımında istenen olasılık bir aralık olabileceği gibi, eşitsizlik durumları da olabilmektedir. Örneğin $P(7 \leq x \leq 12)$ olasılık değerinin normal dağılım yaklaşımında aranan olasılık değeri $P(6.5 \leq x \leq 12.5)$, $P(x \geq 9)$ için $P(x \geq 8.5)$ ve $P(x \leq 10)$ içinse $P(x \leq 10.5)$ olarak bulunmaktadır.

ÖRNEK 21

Yapılan bir pazar araştırması neticesinde, çamaşır makinesi kullanan ev hanımlarından % 63'ünün yerli malı çamaşır makinesini tercih ettikleri bulunmuştur. Bu gruptan, rassal seçilen 100 ev hanımından, 55 – 60 tanesinin, yerli malı çamaşır makinesi tercih etme olasılığını bulunuz.

ÇÖZÜM

İki sonuçlu (binom) bu deneyde,

$n = 100$; $p = 0.63$, $q = 1 - p = 1 - 0.63 = 0.37$

dir ve istenen olasılık $P(55 < x < 60)$ 'dir. Burada $n\,p > 5$ ve $n\,q > 5$ olması nedeniyle istenen olasılık değeri, normal dağılım yaklaşımıyla bulunabilir. Ancak; burada aranacak olasılık $P(54.5 \leq x \leq 60.5)$ biçimindedir.
İlk olarak μ ve σ değerleri hesaplanacak olursa,

$$\mu = n\,p = 100\,(0.63) = 63$$

$$\sigma = \sqrt{n\,p\,q} = \sqrt{100\,(0.63)\,(0.37)} = 4.8280$$

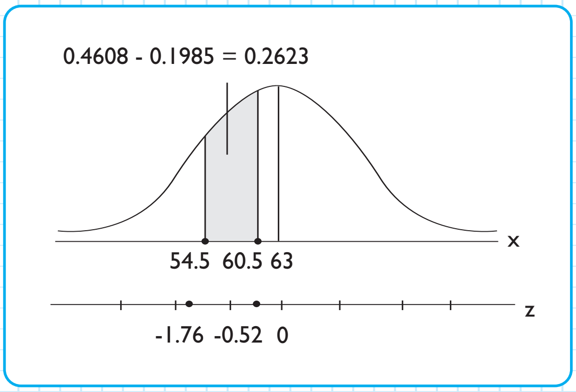
biçimindedir. Daha sonra gerekli z değerleri hesaplanır.

$$x = 54.5 \text{ için } z = \frac{54.5 - 63}{4.8280} = -1.76$$

$$x = 60.5 \text{ için } z = \frac{60.5 - 63}{4.8280} = -0.52$$

Bu değerlerin kullanımı sonucunda ortalamanın solunda yer alan iki alan bulunur ve büyük alandan küçük alanın çıkartılması sonucunda istenen olasılık değerine ulaşılır.

$P(-1.76 \leq z \leq 0) = 0.4608$
$P(-0.52 \leq z \leq 0) = 0.1985$
$P(-1.76 \leq z \leq -0.52) = 0.2623$



Şekil 6.50 *x = 54.5 ve x = 60.5 arasındaki alan.*

ÖRNEK 22

18 yaşın üzerindeki nüfusu hedef alan bir kamuoyu araştırması sonucunda, milli piyangodan ikramiye çıkacağına inananların oranı % 54 olarak bulunmuştur. Bu kitleden rassal seçilen 100 kişiden 60 ya da daha fazla kişinin piyangodan ikramiye çıkacağına inanması olasılığını bulunuz.

ÇÖZÜM

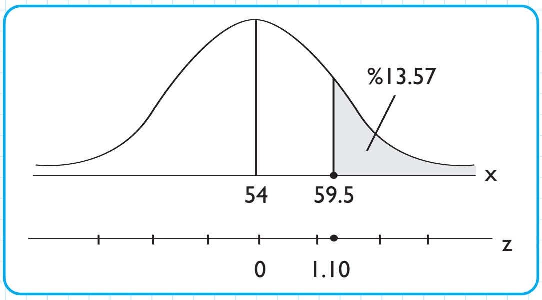
Yukarıdaki örneklerde olduğu gibi, binom deneyine uyan bu deney de, normal dağılım yaklaşımıyla çözülebilir.

$n = 100$; $p = 0.54$, $q = 1 - p = 1 - 0.54 = 0.46$

$\mu = n\,p = 100\,(0.54) = 54$

$\sigma = \sqrt{n\,p\,q} = \sqrt{100\,(0.54)\,(0.46)} = 4.9840$

Bu değerlerden yararlanarak $P(x \geq 59.5)$olasılık değeri standart normal dağılım tablosundan elde edilir.



Şekil 6.51 *$x \geq 59.5$ 'in olasılık değeri.*

$x = 59.5$ için $z = \dfrac{59.5 - 54}{4.9840} = 1.70$

$P(x \geq 59.5) = P(z \geq 1.10) = 0.5 - P(z < 1.10) = 0.5 - 0.3643 = 0.1357$

SIRA SİZDE

1. **Hangi koşullarda normal dağılım binom dağılımına yaklaşım amacıyla kullanılır?**
2. **Bir binom dağılımında $n = 25$ ve $p = 0.40$ olarak verilmiştir.**
   a) **Binom formülünü kullanarak $P(8 \leq x \leq 12)$ değerini bulunuz.**
   b) **Normal dağılım yaklaşımından yararlanarak $P(8 \leq x \leq 12)$ değerini bulunuz.**
3. **Bir binom dağılımı için $n = 120$ ve $p = 0.60$ 'dır. x 120 denemedeki başarılı sonuç sayısını göstermek üzere;**
   a) **Binom dağılımının ortalama ve standart sapmasını bulunuz.**
   b) **Normal dağılım yaklaşımıyla $P(x \leq 70)$ değerini bulunuz.**
   c) **Normal dağılım yaklaşımıyla $P(67 \leq x \leq 71)$ değerini bulunuz.**

Kendimizi Sınayalım

1. Standart normal dağılımda, 1.5 standart sapma sınırları (μ - 1.5σ ve μ + 1.5σ) arasında kalan alan nedir?

a. 0.0220
b. 0.4332
c. 0.8664
d. 0.8814
e. 0.9212

2. Z, standart normal dağılmış bir rassal değişken olduğuna göre, P (z < -2.04) olasılığı nedir?

a. 0.0207
b. 0.0603
c. 0.0968
d. 0.1841
e. 0.2178

3. Z, standart normal dağılmış bir rassal değişken olduğuna göre, P (z > -0.78) olasılığı nedir?

a. 0.4713
b. 0.7823
c. 0.8234
d. 0.8873
e. 0.9564

4. X, sürekli rassal değişken, μ = 25 ve σ = 6 olmak üzere normal dağılmıştır. Bu bilgilere göre, P(22 < x < 33) olasılığı nedir?

a. 0.2178
b. 0.3336
c. 0.4599
d. 0.5564
e. 0.5997

5. X, sürekli rassal değişkeni, ortalaması 100 ve varyansı 225 olmak üzere normal dağılmıştır. Bu bilgilere göre, P (115 < x < 130) olasılığı nedir?

a. 0.0437
b. 0.0948
c. 0.1056
d. 0.1359
e. 0.1443

6. Tek tip vida üreten otomatik makinelerden üretilen vidaların boyları; ortalaması 3.0 cm. ve standart sapması da 0.009 cm. olmak üzere normal dağılmaktadır. Üretilen vidaların boyu 2.98 cm.' den kısa olanlarla 3.02 cm.' den uzun olanlar kusurlu oldukları için kullanılmamaktadır.
Bu makinede üretilen vidaların yüzde kaçı kusurludur?

a. 1.17
b. 2.10
c. 2.64
d. 3.00
e. 3.15

7. Bir A bölgesinde 20 000 konuta ilişkin aylık elektrik enerjisi tüketimi, ortalaması 1 650 kilovat saat ve standart sapması da 320 kilovat saat olmak üzere normal dağılmaktadır.
Bu bölgede kaç konutun elektrik tüketim miktarının 900 - 1.300 kilovat saat arasında olması beklenir?

a. 1.854
b. 2.566
c. 2.700
d. 2.850
e. 3.204

8. Otomobillerin yağ değişimini yapan bir servis istasyonunda servis süresinin, ortalaması 15 dakika ve standart sapması da 2.4 dakika olmak üzere normal dağıldığı bilinmektedir. Servis sorumlusu, daha çok müşteri çekebilmek amacıyla, bir kampanya başlatmak istemektedir. Kampanya süresince, belirlenen süreden fazla bekleyen müşterilerden servis ücretinin yarısı alınacaktır. Ancak, servis maliyeti dikkate alınarak, yarı ücret alınacak müşteri sayısının, toplam müşteri sayısının % 5' inden fazla **olması** istenmektedir. Buna göre öngörülecek bekleme süresi kaç dakikadır?

a. 12
b. 15
c. 19
d. 22
e. 25

9. Yapılan bir anket sonucunda belirli bir bölgedeki aile reislerinin 0.90 'ının kendilerini ekonomik yönden fakir gördükleri belirlenmiştir.
Bu bölgeden rasgele seçilen 10 aile reisinden 8 ya da **daha fazlasının** kendini fakir görme olasılığı nedir?

a. 0.1427
b. 0.2218
c. 0.3715
d. 0.5219
e. 0.9290

10. X, ortalaması 3 ve varyansı da 9 olan normal dağılmış bir rassal değişken iken $P(0<x<1)$ olasılığı aşağıdakilerden hangisidir?

a. 0.0960
b. 0.1223
c. 0.1544
d. 0.2218
e. 0.3441

Yanıt Anahtarı

1. c
2. a
3. b
4. e
5. d
6. c
7. b
8. c
9. e
10. a

Yararlanılan Kaynaklar

HOEL, P.G. and Jessen, R.J.: **Basic Statistics for Business and Economics,** Wiley, New York, 1971.

MANN, P.S.: **Introductory Statistics,** 2nd Edition, Wiley, New York, 1995.

O'HAGAN, A.: **Probability: Metods and Measurement,** Chapman and Hall, London, 1988.

WONNACOTT, R.J., WONNACOTT, T.H.: **İntroductory Statistics,** 4th Edition, Wiley, Singapore, 1985.

GAUSS (1777-1855)

Yoksul ve eğitimsiz bir ailenin çocuğu olan Gauss 25 yaşından önce, matematik ve astronomi alanındaki çalışmalarıyla ün kazanmıştır. Sayılara ve hesaplamaya karşı erken yaşlarındaki tutkusu, cebir, analiz, geometri, olasılık, hata teorisi, astronomi, haritalama, jeodezi, jeomagnetizma, elektromagnetizma ve aktüerya gibi farklı dallarda başarı kazanmasına neden olmuştur.

Gauss, tüm zamanların en büyük bilim ustası olarak gösterilmektedir.

7

Örnekleme

Çalışma Biçimine İlişkin Olarak

- *Kavramlar ve bu kavramlar arasındaki ilişkiler dikkatle incelenmeli,*
- *Örnekler ve örnek çözümleri dikkatle incelenmeli, sorunlarla karşılaşılırsa kavramsal açıklamalara geri dönülmelidir.*

Amaçlar:

- *Tamsayım ve örneklem kavramlarını açıklayabileceksiniz.*
- *İstatistiksel araştırmalarda, örneklemenin, tamsayıma tercih edilmesindeki nedenleri, kavrayarak açıklayabileceksiniz.*
- *Bir örnekleme plan sürecinin aşamalarını belirleyebileceksiniz.*
- *Örnekleme yöntemlerini sınıflandırabileceksiniz.*
- *Örnekleme dağılımı kavramını açıklayabileceksiniz.*
- *Örnekleme uygulamalarında işlenecek hata türlerini açıklayabileceksiniz.*

İçerik Haritası

- *GİRİŞ*
- *TAMSAYIM VE ÖRNEKLEM*
- *ÖRNEKLEME YAPMAYI GEREKLİ KILAN NEDENLER*
- *ÖRNEKLEME SÜRECİNİN AŞAMALARI*
    - *Ana Kütlenin Tanımlanması*
    - *Çerçevenin Belirlenmesi*
    - *Örnekleme Yönteminin Seçimi*
    - *Örneklem Hacminin Belirlenmesi*
    - *Örneklemin Seçimi*
- *ÖRNEKLEME YÖNTEMLERİ*
    - *Olasılıklı Olmayan Örnekleme Yöntemleri*
    - *Olasılıklı Örnekleme Yöntemleri*
- *ÖRNEKLEME DAĞILIMI*
    - *Örneklem Ortalaması $\overline{X}$ 'nın Örnekleme Dağılımı*
    - *Örneklem Oranı p'nin Örnekleme Dağılımı*
- *ÖRNEKLEMEDE HATA KAVRAMI VE STANDART HATA*
    - *Örnekleme Hatası - Standart Hata*
    - *Örnekleme Dışı Hatalar*

GİRİŞ

Genel olarak bir istatistiksel araştırma sürecinin ilk aşamasında, ilgili ana kütle tanımlanır. Sonraki aşamalarda, tanımlanan ana kütlenin ilgilenilen parametreleri hakkında bilgi üretilmeye çalışılır. Tanımlanan ana kütle hakkında bilgi üretmede başvurulabilecek ilk yöntem, bu ana kütle tanımı içinde yer alan bütün birimler üzerinden, değişken ya da değişkenlerle ilgili veri derlemek, tamsayım yapmaktır. Ancak tamsayım, çeşitli nedenlerle her zaman mümkün olmaz. Bu durumda, istenilen bilginin üretilebilmesi, ancak, tanımlanan ana kütleden, onu temsil edebilecek sınırlı sayıda birimin, yani bir örneklemin seçilmesi ve bu örneklem birimleri üzerinden, gereken verilerin derlenmesiyle mümkün olur.

Uygulamada, örneklemden elde edilen, verilerden hesaplanan örneklem istatistikleri yardımıyla, ana kütle parametreleri hakkında bilgi üretilebilir.

Örnekleme, bir araştırma sürecinin bütün aşamalarıyla iç içelik gösteren en önemli aşamasıdır ve araştırmacıların vazgeçemedikleri önemli uğraşlarından biridir. Bu ünitede, öncelikle, örneklemenin temel kavramları üzerinde durulacak, uygulamada yaygın olarak, neden örneklemeye başvurulur, sorusu yanıtlanacaktır. Sonra da bir örnekleme sürecinde izlenecek aşamalar tanıtılacaktır. Daha sonra olasılıklı olmayanla, olasılıklı örnekleme yöntemleri ve önemli türleri hakkında bilgi verilerek, ne tür bir örneklem seçilecektir sorusuna yanıt aranacaktır. Seçilen örneklemler için, hesaplanan örneklem istatistikleri dağılımlarının özellikleriyle ilgili bilgiler verilecektir. Bu ünite, örneklemede hata kavramına ilişkin açıklamalarla tamamlanmış olacaktır.

TAMSAYIM VE ÖRNEKLEM

AMAÇ 1 ***Tamsayım ve örneklem kavramlarını açıklayabileceksiniz.***

Ulusal yayın yapan ve habercilikte de iddialı bir TV kanalının yöneticileri, İç Anadolu bölgesinde, kanallarının etkinliğini belirleyebilmek amacıyla, bir araştırma yapılmasını kararlaştırmıştır. Konuyla ilgili öngörülen bütçe ve sonuçların iki ay içinde alınması koşuluyla bir araştırma firmasıyla anlaşmaya varılmıştır.

Araştırmayı üstlenen firma, sınırlı zaman ve parasal koşullar altında, araştırmayı gerçekleştirebilmek için, nasıl bir yol izlemelidir?

İlk akla gelen, bölgedeki tüm TV izleyicileriyle birebir görüşme yapmaktır. Ancak, ne zaman, ne de öngörülen bütçe, genel olarak elverişli olmaz.

Ancak bu ve benzeri sorunların üstesinden, konuyla ilgili geliştirirlen istatistiksel teknikler yardımıyla gelinebilir.

Araştırmaların pek çoğundaki amaç, bir ana kütlenin parametreleri hakkında bilgi elde etmektir. Geniş anlamıyla ana kütle, bir araştırma için verilen tanım çerçevesinde yer alan bütün birimlerin oluşturduğu kümedir. Ana kütle bazan evren olarak da isimlendirilir. Ana kütlenin özelliklerini belirleyen sayısal karakteristiklere “parametre” adı verilir. Parametre değerleriyse örneğin, bir sınıftaki öğrencilerin dönem ortalama başarı puanı, belirli bir marka ürüne bir coğrafi bölgede bağımlı olan tüketici oranı gibi sayılardır. Ana kütle parametreleri hakkındaki bu türden bilgiler, tam sayım yapılarak ya da seçilen bir örneklemin incelenmesiyle elde edilebilir. Tam sayım, sonlu bir ana kütlenin bütün birimlerinin incelenmesi ya da sayılması işlemidir.

Geniş anlamda ana kütle, bir araştırmada tanım çerçevesinde yer alan tüm birimlerin oluşturduğu kümedir.

Ana kütlenin özelliklerini belirleyen sayısal karakteristiklere, parametre adı verilir.

Tam sayım, sonlu bir kütlenin tüm birimlerinin sayılması işlemidir.

Tamsayım genellikle küçük hacimli ana kütlelere uygulanır. Bununla birlikte, büyük hacimli ana kütlelerde bütün birimlere ulaşabilmek olanaklıysa ya da araştırmada ana kütlenin bütün birimlerinden veri derlenmesi isteniyorsa, yine tamsayım yapılır. Başka bir ifadeyle, karşılaşılan özel problemin çözümü için mümkün bütün verilerin elde edilmesine gereksinim varsa ve bu mümkünse, 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 ana kütle hacmi N= 50 işçiden oluşmaktadır ve küçük hacimli bir ana kütledir. İşçilerin her birine ulaşmak ve onlardan memnuniyetleri konusunda veri elde etmek kolaydır. Bu nedenle tamsayım yapılabilir.

ÖRNEK 2

Bir bankanın şube müdürü genel müdürlükte yapılacak bir toplantı için, şubesinin, toplantı öncesindeki son iş günü itibariyle mevduat durumuna ilişkin bilgiye ihtiyaç duymaktadır. Şubenin 150.000 mevduat müşterisi bulunmaktadır. Adı geçen şubenin yeterli bilgisayar donanımına sahip olduğu ve mevduat hesaplarına ilişkin bir veri tabanının da olduğu bilindiğine göre, ihtiyaç duyulan bilgi, tamsayım yapılarak elde edilebilir mi?

ÇÖZÜM

Ana kütle hacmi N = 150.000 müşteridir. Büyük hacimli bir ana kütledir. Ancak şubenin bilgisayar donanımına sahip olması ve mevduat hesaplarıyla ilgili verilerin veri tabanında bulunması nedeniyle, ana kütle hacmi büyük bile olsa, çok kısa zamanda gerekli bilginin elde edilmesi, başka bir anlatımla tamsayım yapılması mümkündür.

Bir kütlenin ilgilenilen özelliklerini yansıtması amacıyla, sözü edilen kütleden belirli yöntemlerle seçilmiş birimlerin oluşturduğu topluluğa örneklem adı verilir.

Ana kütlenin ilgilenilen özelliklerini yansıtması amacıyla, ana kütleden belirli yöntemlerle seçilmiş birimlerin oluşturduğu topluluğa, örneklem adı verilir. Bir başka ifadeyle örneklem, ana kütlenin bir alt kümesidir.

ÖRNEK 3

26.000 öğrencisi bulunan bir üniversitede, öğrencilerin kendilerine sunulan hizmetleri yeterli bulup bulmadıklarını belirlemek amacıyla, bir araştırma planlanmış ve seçilen 500 öğrenciden görüşleri alınmıştır. Bu araştırmada ana kütle hacmi nedir, örneklem nedir, neden tamsayımdan kaçınılmıştır, açıklayınız.

ÇÖZÜM

Ana kütle hacmi N = 26.000 öğrencidir. Büyük hacimli sonlu bir ana kütledir. 500 öğrenciden oluşan topluluk örneklemdir. 26.000 öğrencinin görüşüne baş vurmak, onlara ulaşmak gerçekte çok zaman alır. Öğrenciler her gün okula gelmeyebilirler. Onlara ulaşmak mümkün olsa bile görüşlerini çeşitli nedenlerle bildirmeyebilirler. Bu nedenlerle tamsayım oldukça zordur.

ÖRNEK 4

Bir fabrikada üretilen bisküvi paketlerinin, planlanan ağırlıkta üretilip üretilmediğinin araştırılması amacıyla, üretilen paketler arasından 250 paket seçilmiştir. Bu araştırmada ana kütle ve örneklem nedir, belirtiniz. Ayrıca tamsayımın yapılıp yapılamayacağını açıklayınız.

ÇÖZÜM

Ana kütle sonsuzdur. Açıktır ki bu tür ana kütleler üzerinde tam sayım yapılamaz, örnekleme zorunludur. Örneklem, 250 bisküvi paketinden oluşan topluluktur.

En iyi örneklem ana kütleyi iyi temsil eden örneklemdir. Eğer ilgilenilen özellikler bakımından ana kütle ve örneklemdeki birimler benzer dağılım gösteriyorsa, oluşturulan örneklem ana kütle için temsili bir örnektir denir. Örneğin, tanımlanan ana kütle hacmi 1000 kişi olsun, bunların %50'si erkek ve erkeklerin de %30'u 60 yaşın üzerinde olduğu varsayılsın. Bu ana kütleden 200 kişilik bir örneklem seçilmiştir. Seçilen 200 kişiden oluşan bu örneklemin temsili olabilmesi için, 200 kişiden %50'sinin erkek ve onların da %30'unun 60 yaşın üzerinde olması beklenir.

Örneklemler kendi başlarına bir anlam ifade etmezler. Onların önemi, seçildikleri ana kütleyi temsil etmedeki gücünde yatmaktadır. Bunun bir ölçüsü de örneklem istatistiğinin doğruluk ve güvenilirlik derecesidir.

Bir örneklem yardımıyla ilgilenilen kütleye ilişkin genelleme yapma sürecine, örnekleme denir.

Örneklemenin temel amacı temsili örneklem oluşturmaktır. Örnekleme, tanımlanan ve ilgilenilen özellikleri bakımından, hakkında genellemelerin yapılması düşünülen bir ana kütleden belirli yöntemlerle sınırlı sayıda birimin seçilmesi, bu birimlerden oluşan örneklemin ilgilenilen özellik (özellikler) açısından incelenmesi ve gerekli istatistiklerin hesaplanması ve bu istatistiklerin ana kütle hakkında genelleme yapmak amacıyla kullanılması işlemleridir. Bir başka şekilde ifade etmek gerekirse örnekleme, örneklem'e dayanan bir araştırmayı planlama ve yönetme sürecidir. Bu sürecin aşamalarıyla ilgili bilgiler izleyen kesimlerdeki "örnekleme sürecinin aşamaları" başlığı altında verilmiştir.

SIRA SİZDE

1. Büyük hacimli ana kütlelere tamsayım uygulanabilir mi?

2. Örneklemenin temel amacı nedir?

3. Tamsayımın yapılamadığı durumlarda parametre değerleri hesaplanabilir mi?

ÖRNEKLEME YAPMAYI GEREKLİ KILAN NEDENLER

AMAÇ 2

İstatistiksel araştırmalarda, örneklemenin, tamsayıma tercih edilmesindeki nedenleri, kavrayarak açıklayabileceksiniz.

Hakkında araştırma yapılacak ana kütle sonsuz bir kütle olduğu zaman, tamsayım imkansız olduğu için ilgilenilen özelliklere ilişkin bilgi ancak bir örneklem yardımıyla elde edilebilir. Daha önce de değinildiği gibi ilgilenilen bu ana kütle sonlu bir ana kütleyse, istenilen bilgi mümkünse, tamsayım yapılarak, değilse, yine uygun bir örneklem yardımıyla elde edilebilir. Ancak, bir araştırma için gerekli olan bütün verileri derlemek mümkünse ve isteniyorsa açıktır ki, örneklemeye ihtiyaç yoktur. Uygulamalarda, aşağıdaki nedenlerden dolayı örnekleme, tamsayıma tercih edilmektedir.

Maliyet: Örnekleme bütçesi, örneklemi tamsayıma tercih etmede en önemli belirleyicidir. Örnekleme, tamsayıma göre daha az maliyetle bilgi üretme imkanı sağlar.

Tamsayımın yapılıp yapılmayacağına karar verilirken örnekleme bütçesi ,önemli sınırlayıcılardan biridir.

ÖRNEK 5

Belirli bir bölge için planlanan bir siyasi araştırmada, partilerin bugünkü oy dağılımı hakkında bilgi edinmek amaçlanmış olsun. Bu araştırmadaki ana kütle, ilgili bölgedeki seçmen sayısıdır. İstenen bilginin üretilmesi amacıyla, tamsayım yapmaya karar verilirse, üstlenilmesi gereken maliyet, bölgedeki genel seçim maliyetine eşdeğer olacaktır. Adı geçen bölge için yapılmış genel seçim harcamalarına bakıldığında, bunun bir araştırmacı kişi ya da kuruluş tarafından karşılanması oldukça zor görülmektedir. Kaldı ki, genel seçimlerde bütün seçmenler oy kullanmamakta ve kullanılan bazı oylar da geçersiz sayılmaktadır. Tamsayım yapmanın imkansız ve maliyetli olduğu bu gibi durumlarda, en akılcı yol, adı geçen araştırma için uygun bir örneklemden yararlanmaktır.

Öte yandan, eğer ana kütle hacmi küçükse ve tamsayım yapmak bütçe olanaklarıyla da mümkünse, tamsayım tercih edilmelidir. Burada dikkat edilmesi gereken nokta, tamsayım yapma maliyeti elde edilecek bilginin değerinden küçük olması gereğidir. Aksi durumda örneklemeye başvurmak uygun olacaktır.

Zaman: Bir araştırma sonunda ulaşılacak bilgiye duyulan ihtiyacın zaman sınırları, tamsayımla mı yoksa örneklemeyle mi yapılacağına karar vermek, 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 zaman bilhassa önemlidir.

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

ÖRNEK 6

Bilindiği gibi, seçmenlerin oy verme tercihleri üzerinde, pek çok faktörün etkisi bulunmaktadır. Partilerin, bu günkü oy dağılımını belirlemeye yönelik bir araştırma uzun bir zamana yayılırsa, seçmenlerin araştırmaya başlandığı günkü görüşleriyle, araştırmanın sonuçlandığı zamandaki görüşleri arasında bazı farklar oluşabileceğinden, üretilen bilginin değeri de giderek azalacaktır.

Doğru Veri Elde Etme: Tamsayım yapılan araştırmalarda, örnekleme hatası işlenmez. Ancak, örneklemede hata kavramı başlığı altında, ileride açıklanacak olan, örnekleme dışı bazı hatalar da söz konusu olabilir. Örneklemeye başvurulduğu zaman, hem örnekleme hataları hem de örnekleme dışı hatalar işlenmiş olabilir. Bu nedenle tamsayım, örneklemden daha tutarlı veri elde etme imkanı verir. Ancak, tamsayım yapılabilmesi için gerekli olan sayıda ve istenen özelliklere sahip, veri derleme hatası yapmayacak gözlemci ya da görüşmeci bulmak ya da yetiştirmek oldukça zor, hatta olanaksızdır. Bu nedenlerle örnekleme uygulamalarında büyük önem taşımaktadır.

Örneğe Giren Birimlerin Fiziksel Zarara Uğraması: Tanımlanan ana kütlede yer alan birimler, veri derlemek ya da ölçüm yapmak amacıyla fiziksel zarara uğratılıyorsa, açıktır ki örneklemeye başvurmak zorunludur. Örneğin, bir savunma sanayii kuruluşunda, üretilen mermilerin içerisindeki hatalı mermi oranının belirlenmesi için yapılacak bir araştırmada tam sayım benimsenirse, gerekli bilgilerin derlenmesi amacıyla üretilen tüm mermilerin teste tabi tutulması gerekir. Bu durumda üretimin amacı gerçekleşmemiş olur. Bu anlamsız bir testtir. Bu gibi durumlarda örneklemeden yararlanmak kaçınılmaz olur.

SIRA SİZDE

1. **Tamsayım yapmayı engelleyen nedenleri açıklayınız.**
2. **Ana kütle hacmi küçük, parasal imkanların yeterli olduğu bir araştırma için, tam sayım mı yoksa uygun bir örneklem mi tercih edersiniz, açıklayınız.**
3. **35.000.000 seçmenin olduğu bir ülkede yapılacak bir kamuoyu yoklaması için, örnekleme mi ya da tam sayım mı yaparsınız?**

ÖRNEKLEME SÜRECİNİN AŞAMALARI

AMAÇ 3 ***Bir örnekleme planı sürecinin aşamalarını belirleyebileceksiniz.***

Genel olarak örnekleme süreci altı aşamadan oluşmaktadır. Birbirleriyle ve bir araştırma sürecinin diğer aşamalarıyla sıkı sıkıya ilişki içinde olan bu aşamalar, aşağıda ele alınmıştır :

Ana Kütlenin Tanımlanması

Örnekleme süreci öncelikle ana kütlenin tanımlanmasıyla başlar ve bir araştırma sürecinde araştırmacının ilk önce yapacağı işlerden biridir.

Ana kütle, belirli bir tanıma uyan ve hakkında bilgilerin üretileceği, çıkarsamaların yapılacağı birimlerden, daha açık bir ifadeyle nesnelerden, olaylardan, kurumlardan ya da bireylerden oluşan topluluktur. Ana kütlenin, ayrıntılı bir biçimde tanımlanmasıyla, hangi birimlerin örneklemde yer alabileceği, hangilerininde yer alamayacağı belirlenir.

Örneklemede verilerin derlendiği birimlere, "gözlem birimi" adı verilir.

Örneklemede, araştırmanın konusuyla ilgili verilerin derlendiği birimlere "gözlem birimi" adı verilir. Bu birimler aynı zamanda örnekleme seçilen birimler de olabilir. Bu durumda gözlem birimiyle örnekleme birimi aynı şeydir. Gözlem birimlerine ilişkin örneklemeyle ilgili olarak izleyen paragraflarda tanımlanacak olan bir çerçevenin temin edilemediği ya da (bir çerçeve oluşturmanın çok masraflı ve çok zaman alabileceği) durumlarda, örnekleme uygulamasını kolaylaştırmak, zaman ve maliyet tasarrufu sağlamak gibi nedenlerle örnekleme birimi kavramına kuramda yer verilir. Kısaca, örnekleme giren birimlere örnekleme birimi adı verilir. Yeri gelmişken vurgulamak gerekir ki örnekleme birimi birden fazla gözlem birimini kapsayacak şekilde de tanımlanabilir.

ÖRNEK 7

Bir il merkezindeki tüketicilerin, bulaşık deterjanı tercih nedenlerini belirlemek için bir araştırma planlanmış olsun. Araştırma, ilgili il merkezinde, 18 yaşının üstündeki bayanlar arasından, bir örneklem seçilerek gerçekleştirilebilir. Bu durumda, gözlem birimi ve örnekleme birimi (aynı birim), 18 yaşın üstündeki bayanlar olur. Alternatif olarak araştırma, hane halklarını örnekleme birimi olarak tanımlayarak da gerçekleştirilebilir. Bu durumda, hane halkları örnekleme birimi, seçilen hane halklarındaki 18 yaşın üstündeki bayanlarsa gözlem birimi olur. Son durumda örnekleme birimi ve gözlem birimi aynı değildir. Öte yandan, ana kütleyi oluşturan birimler yer ve zaman bakımından da tanımlanarak sınırlandırılabilir. Hangi coğrafi sınırlar içindeki birimlerin ana kütle tanımı içinde yer alacağının belirlenmesine "yer bakımından sınırlandırma", hangi zaman noktasındaki ya da aralığındaki birimlerin ana kütle tanımı içinde yer alacağının belirlenmesine de "ana kütlenin zaman bakımından sınırlandırılması" denir. Tüketicilerin bulaşık deterjanı tercihini etkileyen nedenlerin belirlenmesi konulu araştırmada sözü edilen il Eskişehir ise, yer Eskişehir il merkezi olur. Araştırma Mayıs-2002 tarihinde yapılacaksa zaman Mayıs-2002 olur.

Bir araştırmanın amaçlarına ulaşabilmesi için ana kütle tanımlanmasında; açıklık, kesinlik, araştırma amacına uygunluk ve uygulama güçlüğü yaratmaması gibi ilkelerin göz önünde tutulması gerekir.

Çerçevenin Belirlenmesi

Sonlu bir ana kütlede tüm birimlerin yer aldığı listeye, çerçeve adı verilir.

Örneklem oluşturma sürecinin ikinci aşaması, çerçevenin belirlenmesidir. Çerçeve, sonlu bir ana kütlenin bütün birimlerinin yer aldığı bir listedir. İdari bakımdan uzun bir geçmişe sahip ülkelerde nüfus kayıtları, seçmen kütükleri, tapu ve sicil kayıtları ticaret ve sanayi odaları üye listeleri, ekonomik büyüklüklere göre sanayi kuruluşlarının sıra listesi, telefon rehberi, öğrenci kayıt listeleri v.b. çerçeve olarak kullanılabilecek araçlardır.

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 tam sayım ne de örnekleme yapılabilir.

Bir çerçeve yoksa ya da oluşturulamıyorsa bu durumda yeni bir çerçevenin hazırlanması problemiyle karşılaşılır.Yeni bir çerçevenin hazırlanmasında, çerçeve maliyeti ve kapsam hatası özellikle göz önünde tutulmalıdır. Bazen, tanımlanan ana kütlenin bazı birimleri, çerçevede yer alamayacağı gibi, tanımlanan ana kütlenin dışında kalması gereken birimler de çerçevede yer alabilir ya da bazı birimler tekrarlanarak çerçevede yer alabilir. Böyle bir özellikteki çerçevenin kullanılması, çerçeve (kapsam) hatasına neden olur. Uygulamada kabul edilebilir çerçeve hatası, sağladığı maliyet tasarrufuyla ilişkilendirilerek belirlenir.

Çerçeve, kabul edilebilir bir çerçeve hatası düzeyinde, ana kütle birimlerinin çok büyük bir kısmını kapsamalıdır. Şüphesiz amaç, ana kütle tanımında yer alan bütün birimleri kapsamasıdır.

Örnekleme Yönteminin Seçimi

Örnekleme girecek birimlerin belirlenmesine imkan veren yöntemlere, örnekleme yöntemleri denir. Örnekleme yöntemleri, örneklem için birim seçiminde uygulanılan usulün keyfi ya da rassal (tesadüfi) 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 (olasılıklı, olasılıklı olmayan) uygulanacağıdır. Bu konu, izleyen bölümlerde, örnekleme yöntemleri başlığı altında, ayrıntılı bir biçimde ele alınacaktır.

Örneklem Hacminin Belirlenmesi

Örneklem hacmi, örnekleme girecek birimlerin sayısını gösterir. 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.

Nitel Değerlendirmede Esas Olan Faktörler

- **Ana Kütlenin Homojenliği:** Ele alınan ana kütlenin ilgilenilen değişken/özellik bakımından homojen ya da heterojen olması örneklem hacminin belirlenmesine etki eder. Eğer ana kütlenin bütün birimleri ilgilenilen değişken itibarıyla aynı değere sahipse, bir birimin incelenmesi, amaca ulaşmak için yeterlidir. Ancak heterojenlik arttıkça, ana kütleyi temsil edebilecek bir örneklem oluşturabilmek için, örneklem hacminin de giderek büyümesi gerekir.
- **Araştırmada Verilecek Kararın Önemi:** Önemli kararlar için olabildiğince çok ve ayrıntılı bilgiye gereksinim vardır. Böyle durumlar, büyük hacimli bir örneklem üzerinden 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 ana kütleyi 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.
- **Araştırmanın Yapısı:** Araştırmanın doğası da örneklem hacmi üzerinde etkilidir. Uygulamada genellikle nitel araştırmalarda küçük hacimli örneklemlerle, nicel araştırmalardaysa daha büyük hacimli örneklemlerle çalışılır.
- **Benzer Çalışmalarda Kullanılan Örneklem Hacimleri:** Örneklem hacmi, benzer çalışmalarda kullanılan örneklem hacimlerinin ortalamalarından yararlanarak da belirlenebilir. Özellikle, olasılıklı olmayan örnekleme yöntemleri kullanıldığı zaman, bu tür yaklaşım, örneklem hacmi konusunda kabaca fikir verir.
- **Kaynaklarla İlgili Sınırlayıcılar:** Örneklem hacminin belirlenmesiyle ilgili karar, zaman ve parasal imkanlarla sınırlıdır. Ancak veri derleme konusunda iyi yetişmiş eleman bulma hususu da bu bağlamda önemli sınırlayıcıdır.

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 8

Araştırma bütçesinin 2.200.000.000 TL. ile sınırlı olduğu bir araştırmada, sabit maliyet 800.000.000 TL. ve örnekleme seçilecek her örnekleme birimi için maliyet ise 5.000.000 TL. dır. Bu bütçeyle oluşturulabilecek örneklem hacmi en fazla ne olabilir?

ÇÖZÜM

$$n = \frac{C - c_0}{C_t} \frac{2200.000.000 - 800.000.000}{5.000.000} = 280$$

gözlem birimi olmalıdır.

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

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

Burada,

z , Araştırmacı tarafından belirlenen anlamlılık düzeyi (hata düzeyi) için standart normal dağılım tablo değeri,

σ^2 , Ana kütle varyansı,

$h = z \,.\, \sigma_{\bar{x}}$ Araştırmacı tarafından α anlamlılık düzeyi için belirlenen kabul edilebilir yanılgı payını gösterir. (Bu bir örneklem istatistiğinin değeriyle ilgili parametre değeri arasındaki mutlak fark olarak belirlenebileceği gibi, ilgilenilen parametrenin oransal bir değeri olarak da ifade edilebilir.)

Örneklem hacminin yukarıdaki eşitlikle hesaplanabilmesi için araştırmacının α anlamlılık düzeyini ve h^2 değerini belirlemesi ve ana kütle varyansı σ^2 hakkında bilgiye sahip olması gerekir. Ana kütle 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_i rassal değişkeni normal dağılıyorsa, $\alpha = 0.01$ için

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

tahminleyicisi kullanılarak da hesaplanabilir.

ÖRNEK 9 ***Bir araştırmacı, X ilinin merkez ilçesinde ikamet eden ailelerin ortalama aylık mutfak harcama tutarını tahminlemek istiyor. Ayrıca bu tahminlemede 0.05 anlamlılık düzeyinde en fazla 10 milyon TL'lik bir yanılgı payı amaçlıyor. Örneklem hacmi ne olmalıdır? Benzer amaçla bu il merkezinde yapılan araştırmalardan ailelerin aylık mutfak giderleriyle ilgili standart sapmanın 30 milyon TL olduğu öğrenilmiştir.***

ÇÖZÜM

$\alpha = 0.05$
$\sigma = 30$ milyon TL.
$z_{0.05} = 1.96$
$h = z \,.\, \sigma_{\bar{x}} = 10$ milyon TL'dir. Bu verilerden hareketle

$$n = \frac{(1.96)^2 \,.\, (30)^2}{(10)^2} = 34 \text{ aile}$$

olarak elde edilir.

Örneklemin Seçimi

Örnekleme sürecinin bu son aşamasında, örnekleme girecek birimler seçilerek, veri derlenir. Bu, uygun özellikte büro ve çalışma ortamıyla nitelikli iş görenlerin teminini gerektirir. Önceki aşamalarda alınan yanlış kararlar ve dikkatsizlikler bu aşamada büyük sorunların yaşanmasına neden olur.

SIRA SİZDE

1. **Örnekleme sürecinin aşamalarını sıralayınız.**
2. **Anadolu Üniversitesi'nin Açıköğretim Fakülteleri'ne kayıtlı olan öğrencilere ilişkin yapılacak bir araştırmada, anakütle nedir?**
3. **Açıköğretim Fakülteleri öğrencilerine ilişkin çerçeve bulunabilir mi, bulunabilirse güncel midir? Açıklayınız.**

ÖRNEKLEME YÖNTEMLERİ

AMAÇ 4 ***Örnekleme yöntemlerini sınıflandırabileceksiniz.***

Örnekleme yöntemlerini olasılıklı olmayan ve olasılıklı yöntemler şeklinde sınıflandırılabilecek ve bu yöntemlerin önemli türleri hakkında bilgi sahibi olacaksınız.

Olasılıklı Olmayan Örnekleme Yöntemleri

Araştırmayı planlayan ya da (örnekleme uygulamasını yapan) kişi yada grubun istekleri ve değer yargıları, örnekleme seçilecek birimlerin ve örneklem hacminin belirlenmesinde etkili oluyorsa, yapılan örneklemeye "olasılıklı olmayan örnekleme" denir.

Olasılıklı olmayan örnekleme, örneklem için birim seçiminde, keyfi seçim usulünün uygulandığı örneklemedir. Örneklem oluşturulurken, tanımlanan ana kütleyi oluşturan birimler arasında fark gözetilir ve bütün birimlere, bilinen bir olasılıkla seçilme şansı verilmezse, söz konusu olan seçim, keyfi seçimdir.

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

Sosyal araştırmalarda sıkça kullanılan olasılıklı olmayan örnekleme yöntemlerinden bazıları aşağıda ele alınmıştır.

Kolayda Örnekleme

Bu örneklemede amaç, kolayca ulaşılabilir birimlerin seçilmesiyle örneklemin oluşturulmasıdır. Örneğin; 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.

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 ilgili ana kütleyi ne kadar temsil edebildiğidir. Betimleyici ve ilişki araştırıcı araştırmalarda kolayda örnekleme uygun bir yöntem değildir.

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. Bu yöntemin kolayda örneklemeden farkı, örneklem seçimi için araştırmacının belirli ölçütler belirlemesi ve bu ölçütlerin temsili bir örneklem oluşturacak ölçütler olduğuna inanıyor olmasıdır. Kolayda örneklemede olduğu gibi bu örneklemede de örnekleme birimlerine kolayca ulaşılabilir ve verilerin çok hızlı biçimde derlenmesi mümkün olur. Bir A üniversitesinin sorunlarını araştırmak amacıyla bu üniversitenin üst düzey yöneticilerinin örnekleme seçilmesi yargısal örnekleme için bir örnektir. Çünkü üniversite üst düzey yöneticilerinin bu araştırma için temsili bir örneklem olacağı düşünülebilir.

Yargısal örnekleme, pazarlama araştırmasında, kamuoyu araştırmalarında ve biyolojik araştırmalarda yaygın olarak kullanılmaktadır.

Kota Örneklemesi

Örneklem için birim seçiminin keyfi olarak yapıldığı yöntemlerden biri de kota örneklemesidir. Bu yöntemin başarıyla uygulanabilmesi için tanımlanan ana kütleye ilişkin bir çerçevenin var olması, ilgili ana kütle hakkında öncül bilgiye sahip olunması, ana kütlenin tabakalara bölünebilmesi ve tabaka hacimlerinin bilinmesi gerekir.

Kota örneklemesi sürecinde aşamalar aşağıdaki gibidir :

- Ana kütle hacmi N ve tabaka hacimleri N_h, (h = 1, 2, ...) belirlenir.
- Örneklem hacmi n keyfi olarak belirlenir.
- Her tabakanın, ana kütle hacmi içindeki $\frac{N_h}{N}$ oranı belirlenir.
- Her tabakadan keyfi seçimle $n_h = \frac{N_h}{N} . n$ sayıda birim seçilir ve bu seçilen birimler örneklemi oluşturur.

ÖRNEK 10 ***Anadolu Üniversitesi yönetimi, verdiği açıköğretim hizmetlerinden memnun olan öğrencilerinin oranını belirlemek amacıyla bir araştırma planlamıştır. Bu araştırmayı gerçekleştirecek grup araştırmayı, kota örneklemesi uygulamak suretiyle tamamlamayı düşünmektedir.***

ÇÖZÜM

Cinsiyet tabakalama kriterine göre öğrencilere ilişkin bilgiler;

N	= 600.000	Ana kütle hacmi.
N_E	= 400.000	Erkek öğrenci hacmi
N_K	= 200.000	Kız öğrenci hacmi
n	= 3000	Örneklem hacmi

Erkek öğrenciler tabakasından (N_E) seçilecek öğrenci hacmi;

$$n_E = \frac{N_E}{N} . n = \frac{400.000}{600.000} . 3000 = 2000 \text{ öğrenci}$$

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 keyfi seçimle seçmek suretiyle n = 2000 +1000 hacimli örneklem seçilmiş olur.
Bu örneklem üzerinden ilgili veriler derlenir ve istenilen bilgi üretilir.

Kartopu Örneklemesi

Kartopu örneklemesi, özellikle bir çerçevenin mevcut olmaması ya da oluşturulmasının imkansız olduğu durumlarda, faydalı bir örneklemedir. Bu yöntemde, örnekleme süreci tanımlanan evrende yer alan bir bireyin, genellikle rassal olarak seçilmesiyle başlar. Belirlenen bu birey örneklemeye giren birinci birimdir. Bu bireyle aynı ana kütle tanımında yer alan tanıdığı bir bireyin olup olmadığı araştırılır. Varsa, bu bireye ulaşılır. Böylece örneklemde yer alacak olan ikinci birim belirlenmiş olur. Bu süreç keyfi olarak belirlenen n hacimli örneklem oluşturuluncaya kadar sürdürülür.

ÖRNEK 11

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

Çete üyeleri ve bir ülkeye yasal olmayan yollarla girmiş olan kişilerle ilgili araştırmalarda da kartopu örneklemesi uygulanır. Kartopu örneklemesi endüstriyel ürün alan ve satanlar hakkında yapılacak araştırmalarda da kullanılabilir. Bu yöntem uygulandığında temsili örneklem oluşturmak olanaklıdır. Kartopu örneklemesinin örneklem oluşturma maliyeti ve örneklem varyansı düşüktür.

Tüm olasılıklı olmayan örnekleme yöntemlerinde, örneğe girecek birimlerin seçiminin keyfi olması, tek yönlü hatalara (sistematik hatalara) neden olur. Bu tür hatalardan kaçınmak için, izleyen paragraflarda ele alınacak olan olasılıklı örnekleme yöntemleri kullanılır.

Olasılıklı Örnekleme Yöntemleri

Rassal örnekleme yöntemleri olarak da anılan bu yöntemler, örnekleme planlarında yaygın olarak uygulanır. Bu tür örneklemede, örnekleme girecek birimlerin seçimi rassal olarak yapılır. Rassal seçim, ana kütleden, örnekleme girecek birimleri seçerken, herhangi bir ayrıcalığın uygulanmadığı seçimdir. Rassal seçim yapma imkanı veren çeşitli seçim uygulamaları bulunmaktadır. Bunlar, kura usulü, rassal sayılar tablosu, rassal sayı türeten bilgisayar programları ve sistematik seçim olarak sayılabilir.

Rassal seçim için kura usulü uygulamasında önce, oluşturulan çerçevedeki bütün birimlere numara verilir. Sonra bu numaralar fişlere yazılır ve bir torbaya atılır. Fişler iyice karıştırıldıktan sonra, çekilen fiş torbaya iade edilmeksizin (ya da iade edilerek) n tane fiş seçilir. Çıkan fişlerdeki numaralara karşı gelen birimler, rassal örneklemi oluşturur. Rassal seçim uygulamaları için kullanılan rassal sayılar tablosu ve rassal sayı türeten bilgisayar programları kura usulüyle seçimin güçlüklerini ortadan kaldıran rassal seçim araçlarıdır. Bu araçların birim seçim sürecinde nasıl kullanılacağına değinilmemiştir.

Eğer rassal seçim için, sistematik seçim uygun görülürse, aşağıdaki aşamalar izlenir:

- $k = N/n$ 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+2k'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 sırasıyla $1/C_N^n$ ve 1/k' dır.

Rassal seçim ve rassal seçim usullerini açıkladıktan sonra, olasılıklı örnekleme tanımını yapmak kolaylaşmaktadır. En basit tanımıyla, örneklem için birim seçiminde rassal seçimin uygulandığı yöntemlere "olasılıklı örnekleme yöntemleri" adı verilir. Geniş anlamıyla olasılıklı örnekleme, ilgili ana kütledeki her örnekleme birimine hesaplanabilir ve sıfırdan farklı bir olasılıkla seçilme imkanı veren örneklemedir. Tanımda belirtilen özelliklere sahip örneklemse "olasılıklı örneklem" olarak isimlendirilir.

Olasılıklı örneklemenin üç önemli üstünlüğü vardır :

- Örneklemden elde edilen verilerden hesaplanan istatistikler, ana kütle parametreleri hakkında genelleme yapmak amacıyla kullanılabilir.
- Örnekleme hatasının büyüklüğü hakkında bilgi elde edilebilir.
- Keyfi seçimde söz konusu olabilecek yanlılık (sistematik hata) giderilmiş olur.

Olasılıklı örneklem oluşturma prensibi esas olmak üzere, uygulamada ya birim seçim işlemini kolaylaştırmak ya da ana kütleyi temsil edecek daha iyi bir örneklemin oluşturulmasını sağlamak üzere, çeşitli örnekleme yöntemleri geliştirilmiştir. İzleyen kesimlerde, uygulamada en çok kullanılan olasılıklı örnekleme yöntemleri açıklanacaktır.

Basit Rassal Örnekleme

Örnekleme planlarında, örnekleme girecek birimlerin seçiminde, rassal seçimin kolaylıkla uygulanabileceği olasılıklı örnekleme yöntemi, basit rassal örneklemedir.

Basit rassal örnekleme, hacmi N olan sonlu bir ana kütleden, 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 örneklemlerin her birine, "basit rassal örneklem" denir. Bu örnekleme yöntemi, ana kütledeki bütün birimlere, hacmi n olarak belirlenen örnekleme girmeleri bakımından, bilinen ve birbirine eşit (n/N) seçilme olası sağlar.

Uygulamalarda, basit rassal örneklemin seçimi için aşağıdaki süreç izlenir:

- Güncel çerçeve temin edilir ya da hazırlanır. Seçilecek örneklemin temsili olabilmesi için, çerçevenin güncel olması çok önemlidir.
- Örneklem hacmi belirlenir.
- Çerçevede yer alan N sayıdaki birime, onları tanımlayıcı numara ya da işaret verilir.
- Ana kütledeki her birime eşit, 1/N seçilme şansı vermek suretiyle, örnekleme girecek birinci birim seçilir. Seçim için kura usulü, rassal sayılar tablosu ya da rassal sayı türeten bilgisayar programları kullanılır.
- Geriye kalan N-1 birimin her birine eşit şans vermek suretiyle ikinci birim seçilir.
- Bu birim seçim süreci, n hacimli örneklem seçilinceye kadar tekrarlanır.

Açıklanan seçim sürecinde, her çekilişte seçilen birim incelendikten sonra, ana kütleye iade edilmediği için, bu seçim sürecine "iadesiz rassal seçim süreci" adı verilir. Eğer, basit rassal örnekleme planlarında, önceki çekilişte seçilen birim, incelendikten sonra ana kütleye iade ediliyorsa, başka bir ifadeyle, birimler tekrar tekrar seçilme şansına sahipse, bu seçim sürecine "iadeli rassal seçim süreci" adı verilir. İadeli rassal seçim sürecinde, ana kütle hacmi çekilişten çekilişe değişmez.

Sonlu bir ana kütleden iadeli çekilişle bir basit rassal örneklem seçilirse, sonsuz ana kütleden basit rassal örnekleme yapıyormuş gibi bir anlam ifade eder. Bu çekiliş sürecinde ana kütlenin her birimi, yapılacak çekilişlerin her birinde birbirine eşittir, 1/N olan seçilme olasılığına sahiptir ve birbirini izleyen çekilişler bağımsızdır.

Sonlu ana kütlelerde, kütle hacminin büyük ya da küçük oluşu, iadeli ve iadesiz seçim için önemli farklılıklar gösterir. Ana kütle hacmi büyük, örneklem oranı (n/N) çok küçük olduğu zaman, iadeli ve iadesiz örneklemler benzer özellik gösterirler. Çünkü; iadeli çekiliş uygulandığında, önceki çekilişlerde seçilmiş olan bir birimin yeniden örnekleme seçilmesi olasılığı çok küçüktür. Ancak, ana kütle hacmi küçükse, iadeli ve iadesiz rassal çekilişle oluşturulan aynı hacimli örneklemler için hesaplanan örneklem istatistikleriyle ana kütle parametreleri karşılaştırılırsa, iadesiz basit rassal çekilişle oluşturulan örneklem, iadeli olana göre daha az hatayla tahminleme imkanı sağlar. Bu özellik nedeniyle de iadesiz rassal çekiliş, uygulamada genellikle başvurulan yöntem olmaktadır.

Araştırmalarda ilgilenilen özellik açısından, ana kütlenin homojen olması durumunda, basit rassal örnekleme, tercih edilmesi gereken bir yöntemdir.

Örnekleme planlarında, basit rassal örnekleme yönteminin tercihini etkileyen önemli sınırlayıcılar vardır. Bunlardan birincisi, güncel bir çerçeve oluşturmak ya da hazırlamak oldukça zordur. İkincisi, tanımlanan ana kütlenin birimleri geniş bir coğrafik alana yayılmışsa, basit rassal örnekleme uygulaması çok zaman alır ve veri derleme maliyeti giderek artar. Üçüncüsü, eğer tanımlanan ana kütle homojen değilse, basit rassal örneklem sonuçlarının başarısı diğer olasılıklı örnekleme yöntemleri sonuçlarının başarısından düşüktür.

Tabakalı Örnekleme

Örnekleme planlarında amaç, ana kütleyi, ilgilenilen değişken(ler) açısından en iyi temsil edebilecek örneklemi oluşturmaktır. Başka bir anlatımla, ana kütle parametre tahminine ilişkin varyansın, olabildiğince küçük olmasını sağlamaktır. Üzerinde araştırma yapılacak ana kütle, ilgilenilen değişken(ler) yönünden heterojen olduğunda, bu imkanı veren örnekleme yöntemi tabakalı örnekleme yöntemidir.

Tabakalı örnekleme dört aşamalı bir süreçtir. Bu aşamalar aşağıdaki gibi sıralanabilir:

- Öncelikle, incelenecek özellikler açısından önemli farklılıklar gösteren N hacimli bir ana kütlenin birimlerini, birbirine daha çok benzeyen birimlerden oluşacak alt kütlelere, başka bir ifadeyle, tabakalara ayırmada kullanılacak, tabakalama değişkenleri belirlenmelidir. Burada dikkat edilmesi gereken nokta, tabakalama değişkenleri seçilirken, seçilecek değişkenlerin, bir tabakadaki birimlerinin olabildiğince homojen, farklı tabakalardaki birimlerinin de olabildiğince heterojen olmasını sağlayarak, aynı zamanda, uygulama ve ölçme kolaylığı da yaratmak suretiyle maliyeti arttırmadan, tahminleme hatasını azaltması gereğidir. Tabakalama amacıyla kullanılabilecek değişkenlere, demografik özellik, tüketici türü, sosyoekonomik sınıf, meslek grubu, firma büyüklüğü, coğrafik yerleşim yeri vb. örnek olarak gösterilebilir.

- Belirlenen tabakalama değişkeni itibariyle, N hacimli bir ana kütle daha homojen L sayıda ve hacimleri N_1 , N_2 , ... , N_L olan alt tabakalara ayrılır. Bu aşamada önemli olan, tanımlanan ana kütledeki her bir birimin yalnız bir tabakaya ait olması ve hiçbir birimin açıkta kalmamasının sağlanmasıdır. Başka bir ifadeyle,

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

olmalıdır.

Tabaka sayısı L arttıkça tabakaların homojenliği de artacağından tabaka varyansları 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 altıdan fazla olmamasını, çoğu zaman iki ya da üç tabakanın yeterli olabileceğini göstermektedir.

- Her tabakadan basit rassal seçimle sırasıyla n_1 , n_2 , ... , n_L hacimli alt örneklemler oluşturulur. Alt örneklem hacimleri toplamı, örneklem hacmine eşittir. Başka bir ifadeyle, n örneklem hacmini göstermek üzere,

$$n_1 + n_2 + \ldots + n_L = \sum_{h=1}^{L} n_h$$

olmalıdır.

- Nihayet, oluşturulan alt örneklem birimleri üzerinden derlenen bilgiler kullanılarak, araştırma amaçları için gerekli olan istatistikler hesaplanır ve bu istatistiklere dayanarak, istatistiksel çıkarsamalar yapılır.

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

Tabakalı örneklemenin diğer bir üstünlüğü, ilgilenilen ana kütlenin yanı sıra, her tabaka için de ayrı bilgi elde etme olanağı sağlamasıdır. Uygulamada ana kütleye göre tabakalar için çerçeve oluşturmak daha kolay olabilir. Ancak sağladığı bu kolaylıklara rağmen, tabakalı örneklemenin bazı güçlükleri de vardır. Öncelikle tabakalı örnekleme uygulaması için, tabaka hacimleri ve bunların toplamı olan ana kütle hacminin bilinmesi gerekir. Bunu kolayca belirlemek her zaman mümkün olamamaktadır. Ayrıca ilgilenilen ana kütlenin homojen olup olmadığının tespit edilebilmesi için de bu ana kütle hakkında pek çok öncül bilgiye de gereksinim vardır.

Sistematik Örnekleme

Örneklem için birim seçiminin aşağıda ele alınan bir sistematiğe uygun olarak yapıldığı örnekleme sürecine sistematik örnekleme adı verilir. Bu yöntemin sınırlayıcıları, ilgili ana kütleye ilişkin bir çerçevenin var olması ve birimlerin doğal bir sıraya sahip olmasıdır.

Bir sistematik örneklem oluşturma sürecinde şu aşamalar izlenir:

- Ana kütledeki birimler 1'den N'e kadar numaralandırılır.
- Araştırma için yeterli olacak örneklem hacmi (n) belirlenir.
- (k=N/n) örneklem aralığı belirlenir.
- 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 ilgilenilen istatistikler hesaplanır.

Sistematik örnekleme yönteminde bir birimin seçilecek n hacimli örneklemde yer alması olasılığı, basit rassal örneklemede olduğu gibi (n/N)'dir.

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

Öte yandan, ilgili ana kütleye ilişkin çerçevenin yapısı hakkında bilgi sahibi olmaksızın da sistematik örnekleme uygulanabilir. Bu durumda, ana kütle birimlerinin doğal bir sıraya sahip olması gerekir. Örneğin; bir süper marketten ayrılan her k'ıncı müşteriyle görüşme yapılarak yürütülen araştırmalar gibi.

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

Küme Örneklemesi

Bu örnekleme yöntemi, ana kütle hacminin çok büyük, birimlerin geniş bir coğrafik alana yayılmış olması ya da örnekleme girecek birimlere ilişkin bir çerçeve oluşturmanın mümkün olmadığı zaman tercih edilmesi gereken bir yöntemdir. Bu teknikte seçim, kümeler arasından yapılır.

Küme örneklemesi sürecinde aşağıdaki aşamalar izlenilir:

- İlgili ana kütledeki birimler (eğer mümkünse) coğrafik bir ölçüte göre kümelere ayrılır. Kümeler, okullar, üniversiteler, şehirler, devletler gibi doğal olarak var olan birimlerden oluşur.
- Kümeler arasından rassal olarak belirli sayıda küme seçilir.
- Seçilen kümelerdeki birimler üzerinden veri derlenir. Seçilen kümelerdeki birimlerin sayısı, örneklem hacmini gösterir.

Kümeleme örneklemesi, kümelemenin ve örneklemenin herhangi bir aşamada yapılması şeklinde de uygulanabilir. Bu tür küme örneklemesine “aşamalı küme örneklemesi” adı verilir.

ÖRNEK 12

60 mahallesi olan bir il merkezinde partilerin seçim öncesindeki oy dağılımı araştırılacak olsun. Araştırmacı her mahalledeki seçmenleri birer küme olarak tanımlar, mahalleler arasından rassal olarak mahalle seçer ve seçilen mahallelerden de yine rassal olarak belirli sayıda sokak seçerse ve seçilen sokaklardaki bütün seçmenleri incelemeye alırsa iki aşamalı küme örneklemesi uygulamış olur.
Rassal seçim aşama sayısı, benzer şekilde arttırıldığında, çok aşamalı küme örneklemesi uygulanmış olur.

SIRA SİZDE

1. **Örnekleme yöntemlerini sınıflandırınız.**
2. **Olasılıklı örneklemeyle olasılıklı olmayan örnekleme arasındaki temel farklılık nedir, açıklayınız.**
3. **Sistematik örneklemede örnekleme girecek birimler nasıl seçilir, açıklayınız.**

ÖRNEKLEME DAĞILIMI

AMAÇ 5 ***Örnekleme dağılımı kavramını açıklayabileceksiniz.***

Buraya kadar yapılan açıklamalarda, bir istatistiksel araştırmada önemli amaçlardan birinin de ele alınan ana kütlenin ilgilenilen özelliklerini belirlemek olduğu, tamsayımın mümkün olmadığı durumlardaysa bunun uygun bir örneklem yardımıyla nasıl gerçekleştirilebileceği ve uygun örnekleminde nasıl seçilebileceği konuları ele alındı.

Daha önce de değinildiği gibi, bir ana kütleye ilişkin sayısal karakteristiklere parametre adı verilir ve bir parametre genel olarak θ simgesiyle gösterilir. Daha açık bir anlatımla, parametre tam sayımla elde edilen $x_1, x_2, \ldots, x_N$ ölçümlerinin kullanılmasıyla hesaplanan ve ana kütle hakkında bilgi üreten sayısal karakteristiklerin genel adıdır. Tamsayımın yapılamadığı durumlarda hesaplanamayan θ parametresine ilişkin bilgi, θ'ya ilişkin bilgi üreten örneklem istatistiklerinden yararlanılarak elde edilebilir. Örneklem istatistikleri genel olarak $\hat{\theta}$ simgesiyle gösterilir. Örneklem istatistiği ya da sadece istatistik, rassal olarak seçilen n hacimli örneklemden elde edilen $x_1, x_2, \ldots, x_N$ gözlem değerlerinin kullanılmasıyla hesaplanan, sayısal karakteristiklerin genel adıdır.

Örnekleme sürecinde rassal olarak seçilen n hacimli bir örneklem için hesaplanan istatistikler, başlangıçta kendi başına bir anlam ifade etmezler ve sadece ait oldukları örneklem için bilgi niteliğindedirler. Çünkü incelenen bu n hacimli örneklem birbirinden farklı aynı hacimli mümkün örneklemlerden sadece biridir ve her mümkün farklı örneklem için hesaplanan istatistikler birbirinden farklı değerlere sahip olabilirler. Bu nedenle, örneklem istatistiklerinden yararlanarak, ana kütle parametreleri hakkında bilgi üretme sürecinde, bir örnekleme planında rassal olarak seçilen n hacimli bir örneklemin hesaplanan istatistiklerinden değil, o istatistiklerin mümkün örneklemlerde alacağı değerlerin dağılımından, bu dağılımın özelliklerinden ve şeklinden yararlanılır. Bir başka ifadeyle, örneklem istatistiklerinden yararlanmak suretiyle ana kütle parametreleri hakkında bilgi nasıl üretilir sorusunun yanıtı için bilinmesi gereken en önemli kavram, örneklem istatistiğinin örnekleme dağılımı ya da sadece örneklem dağılımı kavramıdır.

N hacimli bir ana kütleden rassal olarak seçilebilecek n hacimli mümkün bütün örneklemlerin seçildiği ve her örneklem için $\hat{\theta}$ istatistiği hesaplandığı varsayıldığında $\hat{\theta}_1 , \hat{\theta}_2 , \ldots , \hat{\theta}_{C_N^n}$ istatistikleri hesaplanmış olur. Hesaplanmış olan istatistiklerinin dağılımına, bu istatistiğin örnekleme dağılımı adı verilir.

Örnekleme planlarında ilgili ana kütleden rassal olarak çekilen n hacimli sadece bir tek örneklem gözönüne alınır ve bu örnekleme ilişkin istatistik hesaplanır. Bu durumda örnekleme dağılımı tanımı, oluşturulacak n hacimli basit rassal örneklemden elde edilecek olan x_1 , x_2 , ... , x_n gözlem değerlerinin , X_1 , X_2 , ... , X_n rassal değişkenlerinin gözlenen değerleri olduğu ve bu gözlem değerlerini kullanarak hesaplanan $\hat{\theta}$ istatistiğinin de bir rassal değişken olduğu gerçeğinden kaynaklanmaktadır. Buna göre örnekleme dağılımı, bir rassal değişken olan $\hat{\theta}$ istatistiğinin olasılık dağılımıdır.

Çeşitli amaçlar için örnekleme yapmaya karar verildiği zaman, dikkatlerin en çok odaklaştığı parametreler ana kütle aritmetik ortalaması μ ve ana kütle oranı Π olmaktadır. Bu nedenle, izleyen bölümlerde bu parametreler hakkında bilgi üreten örneklem istatistiklerinin, sırasıyla örneklem aritmetik ortalaması $\overline{X}$ 'nın ve örneklem oranı p'nin örnekleme dağılımları ve özellikleri ele alınacaktır.

Örneklem Ortalaması $\overline{X}$ 'nın Örnekleme Dağılımı

Ana kütle aritmetik ortalaması μ, ilgili ana kütleye ilişkin önemli bir sayısal karakteristiktir. μ, N hacimli sonlu ana kütlelerde tam sayım yapıldığında gözlem değerleri x_1 , x_2 , ... , x_n olarak gösterildiğinde,

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

eşitliğiyle hesaplanır.

ÖRNEK 13

X sınıfında 4 öğrenci bulunsun. Bu öğrencilerin simgesel isimleri ve başarı puanları Tablo-1 de verilmiştir. Ana kütle aritmetik ortalamasını hesaplayınız.

Öğrenci Adı	Başarı Puanı
A	90
B	80
C	60
D	70

$$\mu = \frac{90 + 80 + 60 + 70}{4} = 75 \text{ puan}$$

Tamsayımın yapılamadığı ve örneklemeye başvurulduğu durumlarda açıktır ki μ hesaplanamaz. Bu durumda μ hakkında bilgi üretme sorunuyla karşılaşılır. Bu sorunun çözümlenebilmesi için, rassal olarak seçilen n hacimli bir örneklem için hesaplanan örneklem aritmetik ortalaması ve bu istatistiğin dağılımından yararlanılır. Örneklemden elde edilen gözlem değerleri x_1 , x_2 , ... , x_n ise,

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

formülü ile hesaplanır.

Yukarıdaki örnek tekrar göz önüne alınsın. N=4 öğrenci olan ana kütleden basit rassal örneklemeyle A ve D isimli öğrencilerin seçilmesiyle n=2 öğrenci olan bir rassal örneklem seçildiğinde, örneklemdeki öğrencilerin başarı puanı ,

$$\overline{X} = \frac{90 + 70}{2} = 80$$

olarak bulunur.

ÖRNEK 14

Örnek 1'de verilen sonlu ana kütleden n=2 birim olan birbirinden farklı mümkün tüm basit rassal örneklemleri oluşturunuz ve her örneklem için aritmetik ortalamayı hesaplayınız.

4 birimden 2 birim, farklı şekillerde seçilir. Sonuçlar aşağıdaki tabloda verilmiştir:

Örneklem No	Örneklem Birimleri	Gözlem Değerleri	Örneklem Ortalamaları
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

Yukarıdaki örnekte ilgilenilen ana kütledeki birim sayısı az olduğu için $\overline{X}$'ın örnekleme dağılımı kolaylıkla oluşturulabilir. Ancak, uygulamada, ilgilenilen ana kütleden elde edilecek mümkün tüm örneklemleri ve bunlara ilişkin aritmetik ortalamaları hesaplamak hem çok külfetli hem de anlamsızdır. Bunun yerine $\overline{X}$ bir rassal değişken olarak alınıp onun kuramsal dağılımından yararlanılır. Bir tanım vermek gerekirse, bir ana kütleden aynı hacimde seçilebilecek mümkün her örneklem için farklı değerler alabilen $\overline{X}$ rassal değişkeninin dağılımına, $\overline{X}$'ın örnekleme dağılımı adı verilir.

Belirlenen her örneklem hacmi için, $\overline{X}$'ın farklı örnekleme dağılımı vardır. $\overline{X}$'ın örnekleme dağılımı, $\overline{X}$ rassal değişkeninin ortalaması ve standart sapması (standart hata) ile belirlenir.

Ortalama Ve Standart Hata

Örneklem hacmi arttıkça, $\overline{X}$'ın örnekleme dağılımının ortalaması ana kütle aritmetik ortalaması μ'ye yaklaşır. $\overline{X}$'ın beklenen değeri (ya da aynı anlama gelen $\overline{X}$'ın örnekleme dağılımının ortalaması) $E(\overline{X})$ ile gösterilirse,

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

olur.

$\overline{X}$'ın örnekleme dağılımının standart sapması (standart hata), örneklem hacmi arttıkça küçülür. $\overline{X}$'ın standart sapması $\sigma_{\overline{X}}$ simgesiyle gösterilir ve $\overline{X}$'ın örnekleme dağılımının değişkenliğini ölçer. $\sigma_{\overline{X}}$, basit rassal örneklemede,

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

eşitliği ile hesaplanır. Standart hatanın karesi, örnekleme dağılımının varyansını ifade eder ve $\sigma_{\bar{x}}^2$ simgesiyle gösterilir. Eşitlikten de anlaşılabileceği gibi, standart hata, ana kütle standart sapması σ'ya ve örneklem hacmi n'e bağlıdır. Bir başka ifadeyle, ana kütle değişkenliği büyükse, herhangi bir n hacimli örneklem için $\bar{X}$ 'ın örneklem dağılımının değişkenliği büyük olur. Örneklem hacmi büyüdükçe daha doğru ve güvenilir bilgi üretme olanağı da artar. Öte yandan, standart hata ile örneklem hacminin kare kökü arasında ters bir ilişki olduğu için, örneklem hacmini arttırmak suretiyle standart hatayı azaltmak bazı güçlüklere yol açar. Örneğin; örneklem hacmi n=100 birim iken, $\bar{X}$ 'ın standart sapmasını yarıya indirebilmek için örneklem hacmi 4 kat arttırılmalıdır.

Ana kütle standart sapması genellikle bilinmediğinden, $\sigma_{\bar{x}}$ hesaplanırken σ yerine onun yansız bir tahminleyicisi olan örneklem standart sapması s kullanılır. Bu durumda standart hata $\sigma_{\bar{x}}$ simgesiyle değil $S_{\bar{x}}$ simgesiyle gösterilir ve

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

ile hesaplanır.

Örneklem standart sapması da,

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

şeklinde hesaplanır.

Eğer ilgilenilen ana kütle sonlu bir ana kütle ve örnekleme oranı $n/N \geq 0.05$ ise, standart hata hesaplanırken $\sqrt{\frac{N-n}{N-1}}$ şeklindeki bir çarpan, düzeltme faktörü olarak kullanılır.

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

Merkezi Limit Teoremi

Ana kütlelerin dağılım şekli ne olursa olsun, basit rassal örneklem hacmi büyüdükçe, $\bar{X}$ 'ın örnekleme dağılımı normal dağılıma yaklaşır. Bu dağılımın ortalaması μ, varyansı σ^2 / n' dir. Örneklem hacmi n için yeterli büyüklük, kesin olmamakla birlikte uygulamada $n \geq 30$ birim olarak kabul edilmektedir.

Eğer $\bar{X}$, ortalaması μ ve varyansı σ^2 olan normal dağılımlı bir ana kütleden seçilmiş n hacimlik bir basit rassal örneklemin ortalamasıysa, $\bar{X}$ 'ın örnekleme dağılımı ortalaması μ, varyansı σ^2 / n olan bir normal dağılımdır.

$\bar{X}$ rassal değişkeninin dağılımı normal olduğundan,

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

eşitliğiyle standart değişkene dönüştürülür. Böylece, normal dağılımın özellikleri kullanılarak ana kütle aritmetik ortalaması μ hakkında, bilgi üretmek kolaylaşır.

Örneklem Oranı p'nın Örnekleme Dağılımı

Örnekleme planlarında ele alınan ana kütlenin 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 birisi örneğin başarılı öğrenci A diğeride $\bar{A}$ ile gösterilsin.

İki sonuçlu ana kütleler üzerinde yapılan örnekleme uygulamalarında, sözü edilen bu iki sonuçtan birinde, örneğin A sonucunda, yer alan birimlerin oranıyla ilgilenilebilir. Bu durumda ana kütle oranı, ana kütlenin birimleri içindeki ilgilenilen türden özelliğe sahip olanların oranı biçiminde tanımlanır.

ÖRNEK 15

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

Öğrenci Adı	Başarı Durumu
A	Başarılı
B	Başarısız
C	Başarılı
D	Başarılı

Örnekte sınıftaki başarılı öğrenci oranı, ana kütle oranıdır ve Π ile gösterilir. Bu ana kütledeki ilgilenilen türden özelliğe sahip (başarılı) birim (öğrenci) sayısı R ile gösterilirse, ana kütle oranı Π,

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

eşitliği ile hesaplanır. Burada R = 0,1,2...,N değerlerini alabileceği için, Π'nin değer aralığı $0 \le \Pi \le 1$ olur. Sınıftaki başarısız öğrenci sayısı (ilgilenilmeyen türden özelliğe sahip birim sayısı) N-R olduğu için, başarısız öğrenci oranı Q,

$$Q = \frac{N - R}{N} = 1 - \Pi$$

olur.

Yukarıdaki örnekte başarılı öğrenci sayısı, R=3 olduğu için

$$\Pi = \frac{3}{4} = 0.75$$

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

Tam sayım yapılamadığı zaman R bilinemez ve Π hesaplanamaz. Örnekleme planlarında Π parametresi hakkında bilgi, bu parametre hakkında bilgi üreten örneklem istatistiklerinden yararlanarak üretilebilir.

Hacmi n olan bir basit rassal örneklemden, bu örneklemin seçildiği ana kütlenin 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

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

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

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

olur.

Yukarıda verilen örnekte ele alınan N = 4 birimlik bir ana kütleden, basit rassal örneklemeyle hacmi n = 2 öğrenci olan bir örneklem seçildiğinde örneklemdeki başarılı öğrenci oranı,

$$p = \frac{r}{n} = \frac{1}{2} = 0.5$$

olarak hesaplanmış olur.
Öte yandan, anakütle oranına ilişkin varyans,

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

şeklinde ifade edilir. Varyansın pozitif kare kökü de standart sapmayı verdiğinden,

$$\sigma = \sqrt{\Pi(1 - \Pi)}$$

şeklinde yazılır.
Örneklem varyansı ve standart sapması da benzer şekilde sırasıyla

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

ve

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

olarak gösterilir.

İki sonuçlu bir ana kütleden, 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 ana kütleden seçilebilecek aynı hacimli, mümkün bütün örneklemler için, hesaplanan örneklem oranlarının oluşturduğu dağılıma, oranların örnekleme dağılımı adı verilir.

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

Ortalama ve Varyans

Ana kütle oranı Π hakkında araştırılmak istenen 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 ana kütleden seçilen n hacimli basit rassal örneklem için hesaplanan p oranının örnekleme dağılımının aritmetik ortalaması μ_p ana kütle oranı Π’ye eşittir. Bu durum örneklem oranı p’nin ana kütle oranı Π’nin yansız (sistematik hata içermeyen) tahminleyicisi olduğunu gösterir. Bu sonuca göre,

$$E(p) = \Pi$$

yazılır.

Sonsuz ana kütlelere ya da örnekleme oranı $n/N \leq 0.05$ olan bütün sonlu ana kütlelere uygulanan basit rassal örnekleme planlarında, örneklem oranı p’nin dağılımının varyansı σ_p^2 ve standart hatası da σ_p ile gösterilir.

Eğer, ana kütle varyansı σ^2 biliniyorsa,

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

eşitlikleriyle, ana kütle varyansı σ^2 bilinmiyorsa,

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

eşitlikleriyle hesaplanır.

Uygulanan bir basit rassal örnekleme planında ilgilenilen ana kütle sonlu ve örnekleme oranı $n/N<0.05$ ise, örneklem oranı p’nin örnekleme dağılımının varyansı σ_p^2 ve standart hatası σ_p ;
ana kütle varyansı biliniyorsa,

$$\sigma_p^2 = \frac{\Pi(1-\Pi)}{n} \cdot \frac{N-n}{N-1} \quad , \quad \sigma_p = \sqrt{\frac{\Pi(1-\Pi)}{n}} \sqrt{\frac{N-n}{N-1}}$$

ana kütle varyansı bilinmiyorsa,

$$s_p^2 = \frac{p(1-p)}{n} \cdot \frac{N-n}{n} \quad , \quad s_p = \sqrt{\frac{p(1-p)}{n}} \cdot \sqrt{\frac{N-n}{N}}$$

eşitlikleri kullanılarak hesaplanır.

Dağılım Şekli Ve Merkezi Limit Teoremi

Oranların örnekleme dağılımının şekli eğer $E(p) = \Pi < 0.5$ ise sağa çarpık, $E(p) = \Pi > 0.5$ ise sola çarpık ve $E(p) = \Pi = 0.5$ ise simetrik bir dağılım gösterir. Kolaylıkla görülebileceği gibi, $E(p) = \Pi$'nin değeri 0 ve 1'e yaklaşırken dağılımın çarpıklığı artar.

Merkezi 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 $n(1-\Pi) \geq 5$ koşullarını birlikte sağlayan örneklem büyüklüğü, yeterli örneklem büyüklüğü olarak kabul edilir. Aynı teoreme göre rassal örneklem hacmi $n \geq 30$ birim olması ve ana kütle oranı Π'nin 0 ya da 1'e yakın değerler almaması koşullarıyla oranların örnekleme dağılımı, normal dağılıma yaklaşır. Bu koşulları sağlayan oranların örnekleme dağılımıyla ilgili problemlerin çözümlerinde normal dağılımın özelliklerinden yararlanılır. Bu amaçla p rassal değişkeni standartlaştırılmış Z değişkeni

$$Z = \frac{p - \Pi}{\sigma_p}$$

şeklinde yazılır. Bu standart değişken kullanılarak, ana kütle oranı hakkında bilgi üretmek mümkün olur.

SIRA SİZDE

1. Örnekleme dağılımı kavramını açıklayınız.

2. Merkezi Limit Teoremi istatistiğe ne tür kolaylıklar getirmiştir, açıklayınız.

3. $E(p) < 0.45$ ise, oranların örnekleme dağılımının şekli nasıldır?

ÖRNEKLEMEDE HATA KAVRAMI VE STANDART HATA

AMAÇ 6

Örnekleme uygulamalarında işlenecek hata türlerini açıklayabileceksiniz.

İstatistiksel araştırmalarda iki tür hatadan söz edilir. Bunlar, örnekleme hatası ve örnekleme dışı hatalardır.

Örnekleme Hatası - Standart Hata

Örneklem istatistikleri, hacmi sonsuz ya da N birimlik bir ana kütlenin sadece n biriminden (n<N) oluşan örneklemlerden derlenen veriler kullanılarak hesaplandığı için, genellikle (bir dereceye kadar) hata içerirler. Bu nedenle, örneklem istatistiklerinin, ana kütle parametrelerine eşit çıkması beklenmez.

Örneklemden örnekleme değişen değerler alan istatistiklerin ana kütle parametre değerlerine göre gösterdikleri sapmalara ($\hat{\theta} - \theta$'lara), örnekleme hatası adı verilir. Herhangi bir sapma pozitifse üst tahminleme, negatifse alt tahminleme yapılıyor demektir ve hata söz konusu olur. Sapmaların sıfır olması durumunda, yapılacak tahmin sapmasızdır ve bu durumda hata söz konusu olmaz. Örnekleme hataları sıfır, negatif ya da pozitif değerler alabildiği için, bunların ortalaması kareli ortalamayla hesaplanır. Bu değer, ilgilenilen örneklem istatistiğinin standart hatasını gösterir. Standart hata aslında bir örneklem istatistiğine ait dağılımın değişkenliğinin bir ölçüsüdür ve başka bir deyişle standart sapmadır.

Standart hata, hakkında bilgi elde edilecek ana kütle parametresine, ana kütlenin dağılım şekline, örnekleme birim seçim usulüne ve örneklem hacminin büyüklüğüne göre farklı tahminleyicilerle hesaplanır. Bu tahminleyiciler örnekleme dağılımı başlığı altında açıklanmıştır. Hesaplanan standart hata değeri, örneklem istatistiğinin değerinin ana kütle parametresi değerinden ortalama olarak ne kadar saptığını gösterir. Bu değerin sıfıra yaklaşması örneklem istatistiğinden yararlanarak üretilen bilginin güvenilir olduğunu gösterir. Standart hata giderilemeyeceğine göre, bu hatanın kabul edilebilir bir düzeyin üstüne çıkmaması istenir.

Standart hatanın değeri, örneklem hacmiyle ters yönde, ana kütle değişkenliğiyle doğru yönde ilişkilidir. Ana kütle değişkenliği veri durumunda olduğu için, standart hatayı azaltabilmek amacıyla örneklem hacmini arttırmak gerekir.

Örnekleme Dışı Hatalar

İster örneklemeden, ister tam sayım sonucu oluşmuş olsun, veri derleme sürecinde işlenen hatalara, örnekleme dışı hatalar denir. Bu hatalar;

- Veri derleme ve kaydetme yönteminden,
- Yanlış anlamaya yol açan soruların varlığından,
- Çerçevenin kapsam hatası içermesinden ya da aynı örnekleme birimin çerçevede birden fazla yer almasından,
- Örnekleme seçilen birimlerin bir kısmından bilgi derlenememesi nedeniyle, ortaya çıkarılabilir.

Hataya neden olan bu unsurlar ortadan kaldırılırsa, örnekleme dışı hata söz konusu olmaz.

SIRA SİZDE

1. Örnekleme sonucu işlenecek hatalar nelerdir?

2. Örnekleme uygulaması sonucu işlenebilecek bütün hatalar istenirse giderilebilir mi?

3. Örnekleme hatasını azaltabilmek için ne yapılmalıdır?

Kendimizi Sınayalım

1. Aşağıdakilerden hangisi, örnekleme yapmayı gerekli kılan nedenlerden **biri değildir?**

a. Maliyet
b. Zaman
c. Kesin sonuç elde etme
d. Doğru veri elde etme
e. Ölçüm için birimlerin fiziksel zarara uğraması

2. Örneklemenin **temel** amacı aşağıdakilerden hangisidir?

a. Ana kütleyi temsil eden bir örneklem oluşturmak.
b. Ana kütleyi temsil eden, mümkün bütün örneklemleri oluşturmak.
c. Ana kütlenin bütün birimlerini incelemek.
d. Örneklemdeki bir birimin incelenmesi.
e. Örneklemdeki birimler arasından, birimler seçmek.

3. Aşağıdakilerden hangisi, bir ana kütlenin tanımlanması için kullanılan bileşenlerden **biri değildir?**

a. Yer
b. Ana kütle hacmi
c. Zaman
d. Örnekleme birimi
e. Gözlem birimi

4. Bir örnekleme planında, kota örneklemesi uygulanması benimsenerek, ilgilenilen ana kütle A, B ve C tabakalarına ayrılmıştır. Tabaka hacimleri $N_A = 1000$, $N_B = 2000$: ve $N_C = 3000$ birimdir. Örneklem hacmi $n = 600$ birim olarak belirlendiğine göre, B tabakasından seçilecek birim sayısı kaçtır?

a. 186
b. 200
c. 210
d. 222
e. 250

5. Aşağıdakilerden hangisi, tanımlanan bir anakütleden rassal olarak seçilen $n = 36$ hacimli, mümkün bütün örneklemler için hesaplanan, p istatistiğinin meydana getirdiği dağımdır?

a. Oranların örnekleme dağılımı
b. Ortalamaların örnekleme dağılımı
c. Örneklem oranları arasındaki farkın örnekleme dağılımı
d. Örneklem ortalamaları arasındaki farkın örnekleme dağılımı
e. Hacmi $n = 36$ olan örneklemin oranı

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. Ana kütlenin tümüne ulaşılamadığı durumlarda, ana kütleyle ilgili bir yargı elde etmek amacıyla, üzerinde istatistiksel değerlerin hesaplandığı gruba ne ad verilir?

a. Kontrol grubu
b. Örnekleme
c. Örneklem
d. Veri
e. Parametre

8. Aşağıdakilerden hangisi ana kütle hacmi büyük olduğunda, tam sayım yapmayı engelleyen nedenlerden **biri değildir?**

a. Parametrelerin kesin değerinin bulunması
b. Masrafları karşılayacak yeterli paranın olmaması
c. Yetenekli işgücü bulunmaması
d. Çok zaman alması
e. Her zaman pratik olarak mümkün olmaması

9. Anadolu Üniversitesi öğrencilerinin boy uzunluğuyla ilgili bir araştırma yapmak için sadece Fen Fakültesi öğrencilerinden rassal olarak belirli sayıda öğrenci seçilmiştir. Fen Fakültesi öğrencilerinin oluşturduğu topluluğa ne ad verilir?

a. Hedef kütle
b. Örneklem hacmi
c. Örneklenen ana kütle
d. Gözlem birimi
e. Örnekleme birimi

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

a. İadeli
b. İadesiz
c. Sistematik
d. Karma
e. Keyfi

Yanıt Anahtarı

1. c
2. a
3. b
4. b
5. a
6. a
7. c
8. a
9. c
10. a

Yararlanılan Kaynaklar

FINK, Arlene: **How to Sampling in Surveys,** Sage Publication, London, 1995.

GÜRSAKAL, Necmi: **Bilgisayar Uygulamalı İstatistik I,** Marmara Kitapevi, Bursa, 1997.

MALHOTRA, Naresh K: **Marketing Research An Applied Orientation** 2nd Edition, Prentice-Holl International, Inc, New Jersey, 1996.

NETER,J; WASSERMAN,W, WHİTMORE, G.A.: **Applied Statistics,** Simon and Schuster, Inc, Boston, 1993.

SERPER, Özer; AYTAÇ, Mustafa: **Örnekleme,** Ezgi Kitapevi, Bursa, 2000.

SERPER, Özer: **Uygulamalı İstatistik II,** Filiz Kitapevi, İstanbul, 1986.

TRYFOS, Peter: **Sampling Methods for Applied Research,** John Wiley and Sons Inc., New York, 1996.

İstatistiksel Tahminleme

8

Çalışma Biçimine İlişkin Olarak:

- *Verilen tanımlar iyice özümsenmeli,*
- *Örnek sorular ve çözümleri dikkatle incelenmeli,*
- *Kavramlar arasındaki farklılıklar göz önüne alınmalı,*
- *İstenenlerin neler olduğu net bir biçimde ortaya konulmalıdır.*

Amaçlar:

- *İstatistiksel tahminleme tanımını kavrayarak, tahminleme sürecinin aşamalarını gerçekleştirebileceksiniz.*
- *Örneklem istatistiklerini kullanarak nokta ve aralık tahminlemesi yapabileceksiniz.*

İçerik Haritası

GİRİŞ

Örnekleme konusu gözden geçirilirken de değinildiği gibi, örneklemler ana kütle parametreleri hakkında bilgi üretmek amacıyla seçilirler. Seçilen örneklemler için hesaplanan istatistikler, bu parametreler hakkında bilgi üretirler. Örneklem istatistiklerinden yararlanarak ana kütle parametreleri hakkında genelleme yapma süreci, istatistiksel yorumlama olarak ifade edilir.

Ana kütle parametreleri hakkında yorumlama iki aşamada gerçekleştirilir. Bunlardan ilki bilmeyen parametre değerinin tahminlenmesi, diğeri de bu parametrenin değeri hakkında karar vermedir. İstatistiksel tahminleme bu ünitede, karar vermeyse hipotez sınamaları adı altında izleyen ünitede ele alınacaktır.

Bu ünitede önce istatistiksel tahminleme ve istatistiksel tahmin kavramları açıklanacak sonra da istatistiksel tahminleme türleri, nokta tahminlemesi ve aralık tahminlemesi ele alınarak ana çizgileriyle açıklanacaktır.

İzleyen paragraflarda kolaylık açısından, ana kütle parametresi yerine "parametre", örneklem istatistiği yerine de "istatistik" ifadeleri kullanılacaktır.

İSTATİSTİKSEL TAHMİNLEME

AMAÇ 1

İstatistiksel tahminleme tanımını kavrayarak, tahminleme sürecinin aşamalarını gerçekleştirebileceksiniz.

Tahminleme, tanımlanan ana kütleden seçilen rassal örneklemden hesaplanan istatistikler yardımıyla, ilgili ana kütlenin parametre değerini araştırmaktır.

2500 işçi çalıştıran bir işyeri, çalışanlarının ulaşım sorununa kolaylık sağlaması açısından işyerine geliş ve dönüş için servis hizmeti vermeyi planlamaktadır. Bunun için de çalışanların ikametgahlarıyla işyeri arasındaki ortalama uzaklığa ilişkin bilgiye ihtiyaç duyulmaktadır. Neler yapılabilir?

Günlük yaşamın hemen hemen her alanında parametre tahminlemesiyle karşılaşılır. Bu nedenle, tahminleme araştırmaların önemli bir evresini oluşturur. Bir araştırma sürecinin en önemli aşaması olan örneklemeyle tahminleme biribirinin ayrılmaz birer parçasıdır.

Bir tanım vermek gerekirse, tahminleme, tanımlanan ana kütleden seçilen rassal örneklemden hesaplanan istatistikler yardımıyla, bu ana kütlenin uyduğu dağılımın parametre değerlerini araştırmaktır denilebilir.

Tanımdan da anlaşılabileceği gibi, bir tahminleme sürecinde aşağıdaki aşamalar söz konusudur :

- Tanımlanan ana kütleden önceden belirlenen n hacimli rassal bir örneklem seçilir.
- Bu gözlem değerleri kullanılarak tahminlenecek parametre için bilgi üretecek istatistikler hesaplanır.
- Parametre için bilgi üreten istatistiğin örnekleme dağılımının özelliklerinden yararlanılarak parametre değeri tahminlenir.

Ana kütle için bilgi üreten istatistiğe ilişkin formülasyona tahminleyici, tahminleyici yardımıyla hesaplanan değere de tahmin denilir.

Hem örneklem için, hem de ana kütle için bilgi üreten istatistiğe ilişkin formülasyona "tahminleyici" örneklem, gözlem değerlerinin bir tahminleyiciye uygulanmasıyla hesaplanan değereyse "tahmin" adı verilir. Tahminleyici, tahminin nasıl yapılacağını gösterir. Tahminse sayısal bir değerdir. Örneğin,

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

formülü ile tanımlanan örneklemin aritmetik ortalaması, ana kütle aritmetik ortalaması için bir tahminleyicidir. Bu formül yardımıyla hesaplanan değerse, ana kütle aritmetik ortalamasının bir tahminidir.

Parametreleri, istatistiklerden hareketle, bir sayı ve bu sayıyı kapsayan bir aralığın sınırlarıyla tahminlemek mümkündür. Birinci durumda nokta tahminlemesi, ikinci durumdaysa aralık tahminlemesi söz konusu olur. Nokta ve aralık tahminlemesi izleyen bölümlerde ayrıntılı olarak ele alınacaktır.

SIRA SİZDE

1. **İstatistiksel tahminleme nedir?**
2. **Bir istatistiksel tahminleme sürecinde hangi aşamalar izlenir?**
3. **Tahmin ve tahminleme kavramları arasındaki farkı açıklayınız.**

İSTATİSTİKSEL TAHMİNLEME TÜRLERİ

AMAÇ 2

Örneklem istatistiklerini kullanarak, nokta ve aralık tahminlemesi yapabileceksiniz.

İstatistiksel tahminleme, nokta ve aralık tahminlemesi olmak üzere iki ana başlık altında sınıflandırılır. İzleyen paragraflarda bu tahmin türleri yeterli ayrıntıda ele alınmıştır.

Nokta Tahminlemesi

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

Bir rassal örneklemden hesaplanan $\hat{\theta}$ istatistiğinin değerini ilgili ana kütle parametresi θ değerine eşit kabul eden tahminleme sürecine, “nokta tahminlemesi” denir. Bu süreçte, dikkat edilmesi gereken nokta, örneklem seçilmeden önce istatistik rassal bir değişkendir ve ilgili ana kütle parametresinin bir nokta tahminleyicisi durumundadır. Örneklem seçildikten sonra bu tahminleyici yardımıyla hesaplanan istatistik, bir sayıdır ve ilgili parametrenin nokta tahminidir.

Tahminleme sürecinde önce, tanımlanan ana kütleden, belirlenen n birimlik (hacimli) bir basit rassal örneklem seçilir. Bu örneklemdeki birimler üzerinden ilgilenilen değişken itibariyle veriler derlenir. Sonra bu veriler kullanılarak tahminlenecek parametre θ için tahminleyici olan $\hat{\theta}$ istatistiği hesaplanır. Bu istatistiğin değeri θ parametresi için bir tahmin olarak kabul edilir.

Ana Kütle Aritmetik Ortalaması μ ’nün Nokta Tahminlemesi

Bir örneklemin aritmetik ortalaması $\bar{X}$’nin hesaplanabilmesi için $X_1, X_2, \ldots\ldots X_n$ gözlem değerlerine gereksinim vardır. Örneklem oluşturulmadan önce aynı ana kütleden seçilebilecek aynı hacimli farklı örneklemlerde farklı değerler alabilecek olan $\bar{X}$, bir rassal değişkendir ve μ’nün nokta tahminleyicisi olarak isimlendirilir. Örneklem oluşturulduktan sonra ($X_1, X_2, \ldots\ldots X_n$ gözlem değerleri elde edildikten sonra) hesaplanan $\bar{X}$ değeri, μ’nün nokta tahminidir.

Gerçekte N birimlik bir anakütleden n birimlik örneklemler C_N^n kadar farklı biçimde seçilebileceğinden, ilgilenilen θ parametresi için C_N^n tane nokta tahmini söz konusu olur.

Bir rassal değişken olan $\bar{X}$’nin değerinin bu istatistiğin bilgi ürettiği μ’nün değerine eşit olan tahminlemeye μ’nün nokta tahminlemesi, hesaplanan $\bar{X}$ değerine de μ’nün nokta tahmini denir.

Yapılan açıklamaları bir örnek üzerinde görelim.

ÖRNEK 1

Bir telefon idaresi, abonelerinin 2001 yılındaki ortalama aylık telefon ödeme tutarlarını tahminlemek istemektedir. Bu amaçla rassal olarak 80 abone seçilmiş ve bu abonelerin ortalama aylık ödeme tutarının 45.000.000 TL ve standart sapmasının da 2000 TL. olduğu hesaplanmıştır. Aylık ödeme tutarını tahmin ediniz.

ÇÖZÜM

$\bar{X}$ = 45.000.000 TL

s = 2000 TL

n = 80

$\hat{\mu}$ = ?

80 birimlik rassal örneğin aritmetik ortalaması, anakütle ortalamasının bir nokta tahmini olduğundan, aylık ortalama ödeme tutarına ilişkin tahmin 45.000.000 TL. olacaktır.

Ancak bu nokta tahminlemeyle, gerçekte ana kütle ortalaması μ'nün 45.000.000 TL.'ye yakın bir değer olduğu yorumu yapılabilir.

Ana kütle ortalaması μ bilinmezken, $\bar{X} = \mu$ olan tek bir örneklem düşünülebilir mi? Düşünmemelisiniz.

Ana Kütle Oranı π'nin Nokta Tahminlemesi

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

Ana kütle 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 değeri π'ye eşit olan bir tahminleme sürecidir. Burada $p = \frac{r}{n}$ eşitliği tahminleyici, hesaplanacak p değeri de tahmindir.

ÖRNEK 2

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

ÇÖZÜM

n = 100 kişi

r = 65 kişi

$$p = \frac{r}{n} = \frac{65}{100} = 0,65$$

olur. Örneklem oranı p = 0,65 değeri ana kütle oranı p'nin yansız tahminidir.

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 ana kütleden seçilen aynı hacimli farklı örneklemlerde büyük ölçüde farklılık göstermeyen tahmindir. Nokta tahminlemesi tahminin güvenilirliği hakkında bilgi veremediği için sınırlı bir tahminlemedir.

Güvenilirliği somut bir şekilde ortaya koymak için güven aralığı kavramı geliştirilmiştir. Bu kavram, tahminlenecek parametre değerini kapsayacak alanı belirleyen bir çift sınır değeri kullanmaktadır.Güven sınırları adı verilen bu sınırlardan biri alt sınır (A), değeri üst sınır (Ü) olarak tanımlanır.

Bilinmeyen anakütle parametresini, istenen bir olasılıkla, bir aralık içinde tahminleme sürecine, aralık tahminlemesi denilir.

Bir parametrenin aralık tahminlemesi (ya da güven aralığı) tahminlenecek parametrenin içinde yer alabileceği ve bu parametrenin nokta tahmini etrafında simetrik olan bir çift sınır değerinin, örneklemin planlanması aşamasında belirlenen bir olasılığa (güven düzeyine)göre hesaplanması sürecidir. Bu tanıma göre θ parametresinin aralık tahminlemesinin gösterimi, genel olarak;

$$A \leq \theta \leq Ü$$

şeklinde yapılır.

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 sebeple 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.) doğru aralık tahminlemesinin yapıldığı parametre değerini kapsayan güven aralığının tahminlendiği olasılığı ifade eder ve 1-α ile gösterilir. G.D=1-α değeri büyük seçilirse tahminlenen aralığın θ'yı kapsayan bir aralık olma olasılığı artmış, fakat tahminlerin güvenilirliği azalmış olur.

Güven aralığının tahmininde parametre değerini kapsayan olasılığa güven düzeyi denilir.

Bir aralık tahminlemesi sürecinde, güven sınırları A ve Ü'nin tahminlenebilmesi için şu adımlar izlenir:

- Güven düzeyi G.D=1-α belirlenir. Uygulamada genellikle güven düzeyi G.D=1- α, % 95 ya da % 99 olarak seçilmektedir. Güven düzeyi tanımlanan ana kütleden, n hacimli, mümkün pek çok farklı örneklemler çekilirse bu, örneklemler için hesaplanan güven aralıklarının, parametreyi kapsama olasılığını gösterir. Bu nedenle, parametreyi kapsayan güven aralıkları, doğru güven aralığı olarak isimlendirilir.Burada α anlamlılık düzeyini gösterir. Güven düzeyi örneğin % 95 seçildiğinde, anlamlılık düzeyi $\alpha = 1 - 0.95 = 0.05$ olur.
- Hacmi n olan bir basit rassal örneklem seçilir. Tahminlenecek parametre için, bilgi üretecek $\hat{\theta}$ ($\bar{X}$, s, p...) istatistiği hesaplanır.
- Bu istatistiğin dağılımıyla ilgili bilgilerden yaralanarak güven aralığı oluşturulur.

Hesaplanan $\hat{\theta}$ istatistiğinin dağılımı normalse, güven sınırlarının belirlenebilmesi için aşağıdaki eşitliklerden yararlanılır.

Alt Sınır,

$$A = \hat{\theta} - z_{\frac{\alpha}{2}} \cdot s_{\hat{\theta}}$$

ve

Üst Sınır,

$$Ü = \hat{\theta} + z_{\frac{\alpha}{2}} \cdot s_{\hat{\theta}}$$

Bu eşitliklerde:

$\hat{\theta}$: Örneklem istatistiğinin değerini,

$z_{\frac{\alpha}{2}}$: Belirlenen güven düzeyi için, standart normal dağılım tablo değerini,

$\sigma_{\hat{\theta}}$: σ bilindiğinde tahminin standart hatasını gösterir.

$S_{\hat{\theta}}$: σ bilinmediğinde (σ yerine onun yansız tahminini s kullanıldığında) standart hata tahminini gösterir.

Belirlenen güven düzeyi için ve $\hat{\theta}$ istatistiğinin örnekleme dağılımının normal olduğu varsayımı altında θ parametresi için aralık tahminlemesi

$$\hat{\theta} - z_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\theta}} < \theta < \hat{\theta} + z_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\theta}} \qquad (\sigma \text{ bilindiğinde})$$

ve

$$\hat{\theta} - z_{\frac{\alpha}{2}} \cdot s_{\hat{\theta}} < \theta < \hat{\theta} + z_{\frac{\alpha}{2}} \cdot s_{\hat{\theta}} \qquad (\sigma \text{ bilinmediğinde})$$

şeklinde ifade edilir.

Aralık tahminlemesinde güven aralığının mümkün olduğu ölçüde dar tutulması arzu edilir.Çünkü dar aralığın sınırları parametre değerine daha yakındır. Bu güven düzeyine ve örneklem hacmine bağlıdır. 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.

Ana Kütle Aritmetik Ortalamasının Aralık Tahminlemesi

Ana kütle aritmetik ortalaması μ için 1-α güven sınırlarının ya da güven aralığının belirlenmesi işlemlerine μ'nün aralık tahminlemesi denir. Bu tahminleme sürecinde şu aşamalar izlenir:

- G.D = 1-α belirlenir,
- Hacmi n olan bir rassal örneklem seçilir,
- Bu örneklem için $\bar{X}$ ve s istatistikleri hesaplanır,
- $\bar{X}$'nın standart hatası $\sigma_{\bar{X}}$ hesaplanır ya da $s_{\bar{X}}$ tahminleyicisi yardımıyla tahminlenir,
- $\bar{X}$'nın dağılım şekliyle ilgili bilgilerden yararlanarak

$$\bar{X}_A \leq \mu \leq \bar{X}_Ü$$

güven aralığı tahminlenir. Burada; $\bar{X}_A$ ve $\bar{X}_Ü$ aralığın sırasıyla, alt ve üst sınır değerlerini gösterirler.

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, ana kütle bölünme 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 olduğundan, Z standart rassal değişkeni,

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

ile dönüştürülür.

İstatistiksel yorumlama amacıyla yapılacak çözümlemelerde büyük kolaylıklar sağlayan bu dönüştürmeden yararlanarak birer rassal değişken olan $\bar{X}_A$ ve $\bar{X}_Ü$ güven sınırları

$$z_A = \frac{\bar{X}_A - \mu}{\sigma_{\bar{X}}} \qquad z_Ü = \frac{\bar{X}_Ü - \mu}{\sigma_{\bar{X}}}$$

eşitlikleri yardımıyla hesaplanabilir. Burada $z_A = -z_{\frac{\alpha}{2}}$ ve $z_Ü = +z_{\frac{\alpha}{2}}$ olmak üzere belirlenecek güven aralığının alt sınır değeri,

$$\bar{X}_A = \mu - z_{\frac{\alpha}{2}}\,\sigma_{\bar{X}}$$

ve üst sınır değeri de

$$\bar{X}_Ü = \mu + z_{\frac{\alpha}{2}}\,\sigma_{\bar{X}}$$

olur.

Daha önce de açıklandığı gibi $\bar{X}$, μ'nün yansız tahminleyicisi olduğuna göre, bilinmeyen anakütle parametresi μ için güven aralığı

$$\bar{X} - z_{\frac{\alpha}{2}} \,.\, \sigma_{\bar{X}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \,.\, \sigma_{\bar{X}}$$

şeklinde oluşturulur.

Yukarıdaki formüllerde z değeri, araştırmacı tarafından başlangıçta belirlenen 1-α güven düzeyine bağlı olarak standart normal eğri alanları tablosundan bulunur. Bazı 1-α güven düzeyleri için z değerleri aşağıdaki tabloda verilmiştir.

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

Tablo 8.1 *Bazı (1-α) değerleri için $Z\frac{\alpha}{2}$ tablo değerleri.*

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

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

formülüne göre, buna karşılık örneklemin seçiminde iadesiz seçim uygulanıyor ya da örnekleme oranı n/N > 0.05 olduğunda, $\sigma_{\overline{X}}$, yukarıdaki eşitliğe $\frac{\sqrt{N-n}}{\sqrt{N-1}}$ çarpanı ilave edilerek hesaplanır.

Ana kütle aritmetik ortalaması μ 'nün aralık tahminlemesiyle ilgili yukarıda yapılan açıklamalar ana kütle 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 $s_{\overline{X}}$ simgesiyle gösterilir. $\overline{X}$'nın örnekleme dağılımı $n \geq 30$ birim olduğunda normal dağılıma uyduğu için, bu yaklaşım önemli bir hataya neden olmaz. Rassal örneklem hacmi yeterli büyüklükte olduğunda, 1 - α güven düzeyi için, μ 'nün güven aralığı

$$\overline{X} - z_{\frac{\alpha}{2}} \cdot s_{\overline{X}} < \mu < \overline{X} + z_{\frac{\alpha}{2}} \cdot s_{\overline{X}}$$

şeklinde bulunur.

Ortalamanın standart hatasının tahmini; örnekleme birim seçiminde iadeli seçim uygulanır, örnekleme oranı n/N<0.05 olur, tanımlanan ana kütle sonsuz olursa

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

formülüne göre hesaplanır. Buna karşılık birim seçimi iadesiz yapılır ya da örnekleme oranı n/N>0.05 olursa,

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

formülüyle hesaplanır.

ÖRNEK 3

Bir ekmek fabrikasının üretim sürecinde, 100 ekmeklik bir rassal örneklem seçilmiştir. Bu ekmeklerin ortalama ağırlığı 375 gr ve standart sapma s=15 gr olarak hesaplanmıştır. Üretilen ekmeklerin ortalama ağırlığını % 95 güven düzeyiyle tahminleyiniz.

ÇÖZÜM

$n = 100$ Ekmek
$\bar{X} = 375$ gr
$s = 15$ gr
$G.D = 1 - \alpha = 0.95$

dolayısıyla

$\alpha = 0.05$

olur.

Problemde ana kütlenin değişkenliği ve bölünme şekli hakkında bilgi bulunmamaktadır.

Örneklem hacmi n=100 birim (n>30 birim) olduğu için $\bar{X}$ 'nin örnekleme bölünmesi normaldir (Merkezi Limit Teoremi). Bu bilgilere göre $Z_{\frac{\alpha}{2}} = \pm 1.96$ dır. Ana kütle hacmi sonsuzdur. Buna göre

$$s_{\bar{X}} = \frac{15}{\sqrt{100}} = \frac{15}{10} = 1.5 \text{ gr.}$$

ve güven aralığı

$$\bar{X} - z_{\frac{\alpha}{2}} \cdot s_{\bar{X}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \cdot s_{\bar{X}}$$

$$375 - 1.96(1.5) < \mu < 375 + 1.96(1.5)$$

$$372.06 < \mu < 377.94$$

olarak hesaplanır.

Üretilen ekmeklerin ortalama ağırlığı % 95 güvenle 372.06 gr ile 377.94 gr arasında bir değerdir. Başka bir ifadeyle 372.06 – 377.94 gr aralığının üretilen ekmeklerin gerçek ortalama ağırlığını kapsayacağına % 95 güvenebilirsiniz.

Küçük Örneklemlerde μ 'nün Aralık Tahminlemesi

Örneklem hacmi küçük (n<30 birim) olduğu zaman, örneklem ortalamalarının

$$z = \frac{\bar{X} - \mu}{s_{\bar{X}}}$$

standart değerleri yukarıda açıklandığı gibi normal dağılıma sahip olmaz. Bu durumda μ için aralık tahminlemesi, tanımlanan ana kütlenin, normal dağılıma sahip olup olmadığının bilinmesine bağlıdır.

Normal dağılan bir ana kütleden, 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 $+\infty$ $(\bar{X} - \mu) / s_{\bar{X}}$ olan istatistiği $\upsilon = n-1$ serbestlik derecesinde t dağılımı adı verilen sürekli bir dağılıma uyar ve

$$t = \frac{\bar{X} - \mu}{s_{\bar{X}}}$$

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 ana kütlenin dağılımı normal olduğunda, $(\bar{X} - \mu) / \sigma_{\bar{X}}$ istatistiği standart normal dağılıma sahip olur. Ana kütle değişkenliği σ 'nın bilinmediği durumlarda $\sigma_{\bar{X}}$ 'nin yerine $s_{\bar{X}}$ 'nin ikame edilmesi; bu standart değişken üzerinde ilave bir değişkenliği tanımlar ve t istatistiği bu değişkenliği doğru bir şekilde dikkate alır. Örneklem hacmi artarken, serbestlik derecesi $\upsilon = n - 1$ büyür, $s_{\bar{X}}$ 'nin kullanılması nedeniyle ortaya çıkan değişkenlik küçülür ve t dağılımı standart normal dağılıma yaklaşır.

Yukarıda verilen bilgilerin ışığında ilgilenilen ana kütle normal dağılıma sahip olduğunda, μ için güven aralığı büyük örneklerdekine benzer şekilde aşağıdaki gibi ifade edilir. Tek fark z yerine t istatistiğinin kullanılmasıdır.
Tanımlanan ana kütle normal dağılıyorsa, μ için $1 - \alpha$ güven aralığı

$$\bar{X} - t\, s_{\bar{X}} < \mu < \bar{X} + t\, s_{\bar{X}}$$

şeklinde belirlenir.
Burada;
t: $1 - \alpha$ güven düzeyinde ve $n - 1$ serbestlik derecesine bağlı olarak t değerleri tablosundan bulunur. t tablosu örneği ekte verilmiştir.

$s_{\bar{X}}$: Ortalamanın standart hata tahminidir ve

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

tahminleyiciyle tahminlenir.

ÖRNEK 4

Otomobil lastiği üreticisi bir fabrikanın yönetim organı üretilen lastiklerin ortalama ömrünü km olarak tahminlemek istemektedir. Bu amaçla rassal olarak 26 lastik seçilmiş ve bu lastiklerin ortalama ömrünün 30.000 km ve standart sapmasının da 1500 km olduğunu tespit edilmiştir. % 99 güven düzeyi için istenen tahminlemeyi yapınız.

ÇÖZÜM

n = 26 lastik
$\bar{X}$ = 30.000 Km
s = 1500 Km
G.D. = 0.99
α = 0.01

Üretilen lastiklerin ortalama ömrünün normal bölünmeye sahip olabileceğini, çok fazla çarpık olamayacağını biliyoruz. Bu nedenle

$$\bar{X} - t\, s_{\bar{X}} < \mu < \bar{X} + t\, s_{\bar{X}}$$

güven aralığı uygulanabilir. Belirlenen G.D = $1 - \alpha$ = 0.99 ve serbestlik derecesi

$\upsilon = n - 1 = 26\text{-}1 = 25$ dir.

% 99 güven düzeyi (α =0.01) ve υ = 25 serbestlik derecesi için t Tablo Değeri $t_{0.01;25}$ = 2.787 elde edilir.

Ortalamanın standart hatasıysa

$$s_{\bar{x}} = \frac{s}{\sqrt{n-1}} = \frac{1500}{\sqrt{25}} = 300 \text{ km}$$

bulunur.

Örneklem ortalaması, $s_{\bar{x}}$ tahmininden yararlanılarak % 99 güven aralığı,

$$30000-2.787(300) < \mu \; 30.000 + 2.787\,(300)$$

$$29163.9 < \mu < 30836.1$$

biçiminde oluşturulur.

Üretilen bu otomobil lastiklerinin ortalama ömrü % 99 güvenle 29163.9 km ile 30836.1 km arasında bir değerdir.

Ana kütlenin 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.

Ana Kütle Oranının Aralık Tahminlemesi

Bu parametre için aralık tahmini prosedürü, örneklem hacmi büyük ($n \geq 30$ birim) ya da $n/\,N \leq 0.05$ olduğu hallerde uygulanır. 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 anakütle ortalaması aralık tahmininde olduğu gibi, $1-\alpha$ güven düzeyinde ana kütle oranı π'nin aralık tahmini için güven aralığı,

$$p - z_{\frac{\alpha}{2}} \cdot s_p < \pi < p + z_{\frac{\alpha}{2}} \cdot s_p$$

şeklinde hesaplanır.

ÖRNEK 5

Bir bölgedeki sigara içen insanların arasında kanser hastası olanların oranının belirlenmesi amaçlanmaktadır. Bu amaçla rassal seçimle 1500 kişilik bir örneklem oluşturulmuş ve 1500 kişinin içinde 375 kişinin kanser hastası olduğu belirlenmiştir. Bu bölgedeki sigara içen insanların içindeki kanserli hasta oranını 1 – a = 0.95 güven düzeyinde tahmin ediniz.

ÇÖZÜM

n = 1500 sigara içen

r = 375 sigara içen ve kanser hastası olan kişi sayısı

π = ?

$1 - \alpha = 0.95$

$\alpha = 0.05$

$Z_{\alpha/2} = \pm 1.96$

$$p = \frac{r}{n} = \frac{375}{1500} = 0.25$$

$1 - p = 0.75$

$$sp = \sqrt{\frac{0.25 \text{ x } 0.75}{n}} = 0.011$$

$0.25-1.96(0.011) < \pi < 0.25 + 1.96\,(0.011)$

$0.25-0.02156 < \pi < 0.25 + 0.02156$

$0.22844 < \pi < 0.27156$

olarak elde edilir.

Bu bölgede sigara içenler arasındaki kanserli hasta oranı % 95 güvenle, % 22.84 ile % 27.16 arasında bir değer olarak elde edilir.

SIRA SİZDE

1. **Örneklem ortalamasının dağılımının normal olabilmesi için gerekli koşullar nelerdir?**
2. **Standart hata, ne ölçüsüdür? Bu hatayı tamamen ortadan kaldırabilir misiniz, kaldıramazsanız nedenleriyle açıklayınız.**
3. **Uygulamada nokta tahminlemesine mi, yoksa aralık tahminlemesine mi güvenirsiniz? Nedenleriyle açıklayınız.**

Kendimizi Sınayalım

1. Bir ana kütlenin aritmetik ortalamasının güven sınırlarını % 99.34 güvenle tahmin edebilmek için kullanılacak standart hata katsayısı $\left(z_{\frac{\alpha}{2}}\right)$ kaçtır?

a. 2.66
b. 2.68
c. 2.70
d. 2.72
e. 2.74

2. Normal dağılıma sahip bir ana kütleden rasgele seçilen 226 birimlik örneğin ortalaması 16. standart sapmasıysa 4 tür. Buna göre, ana kütle ortalaması, % 99.30 güvenle hangi aralıkta yer alır?

a. 14.20 – 16.72
b. 14.28 – 17.72
c. 15.26 – 16.74
d. 15.28 – 16.72
e. 15.28 – 17.72

3. Ana kütle ortalamasının 0.99 güvenle tahminlenebilmesi için rassal olarak seçilen 15 birimlik örneğin ortalaması 160. standart hatası ise 16 olarak hesaplanmıştır. Tahminlemede kullanılacak tablo değeri kaçtır?

a. –1.5
b. 2.718
c. 2.977
d. 2.947
e. 3.105

4. Bir ilaç firmasında belirli bir ilacın günde ortalama kaç kg. üretildiği tahmin edilmek istenmektedir. Bu amaçla ilacın 81 günlük üretimi incelenmiş ve ortalama üretimin 892 kg., standart sapmasınınsa 27 kg. olduğu belirlenmiştir. α = 0.05 için, istenen tahmin aşağıdakilerden hangisidir?

a. 876.15 – 896.88
b. 886.12 – 897.88
c. 989.88 – 886.12
d. 912.15 – 926.10
e. 985.10 – 986.18

5. Aralık tahminlemesinin nokta tahminlemesine tercih edilmesinin nedeni aşağıdakilerden hangisidir?

a. Aralık tahminlemesi, tahminin güvenilirliğini belirleme imkânı verir.
b. Nokta tahminlemesi, tahminin güvenilirliğini belirleme imkânı verir.
c. Aralık tahminlemesi parametre değerini kapsar.
d. Nokta tahminlemesiyle parametre değeri arasında fark daima sıfırdır.
e. Aralık tahminlemesi sadece örneklem istatistiğine ilişkindir.

6. Seçim yapılacak bir ilçede, partilerin oy oranlarının tahminlenmesi amacıyla oluşturulacak rassal örneklem için, uygun örneklem aşağıdakilerden hangisidir?

a. Geçen seçimde oy kullanmış bütün seçmenler örnekleme alınır.
b. İlçedeki seçmen listelerinden rasgele isimler seçilir.
c. Her seçmen eşiyle birlikte örnekleme alınır.
d. Yalnızca okuryazar seçmenler örnekleme alınır.
e. İlçede oturan bütün seçmenler örnekleme alınır.

7. Normal dağılıma sahip bir ana kütleden rasgele seçilen 81 birimlik örneklemler için ortalamanın standart hatası 0.25 olarak hesaplanmıştır. İlgili dağılımın standart sapması kaçtır?

a. 2.25
b. 4.50
c. 5.06
d. 12.25
e. 20.25

8 ve 9. sorular aşağıdaki bilgilere göre cevaplandırılacaktır.

Bir üniversitede sigara içen öğrenci oranı tahminlenmek istenmektedir. Bu amaçla rassal seçimle n = 1 000 öğrenciden oluşan bir örneklem oluşturulmuştur. Bu örneklemden 300 öğrencinin sigara içtiği belirlenmiştir.

8. Bu üniversitede sigara içen öğrenci oranına ilişkin, yaklaşık olarak % 99 güven düzeyi için üst sınır değeri kaçtır?

a. 0.263
b. 0.284
c. 0.337
d. 0.420
e. 0.990

9. Yukarıdaki bilgilere göre, ortalama hata düzeyi kaçtır?

a. 0.0048
b. 0.0100
c. 0.0370
d. 0.2570
e. 0.3280

10. Bir sınıfın ortalama başarı düzeyini tahminlemek amacıyla rassal olarak 16 öğrenci seçilmiş ve bu öğrencilerin ortalama başarısının 4 üzerinden 2.224 olduğu hesaplanmıştır. % 95 güven düzeyi için yapılacak bir tahminlemede kullanılacak serbestlik derecesi kaçtır?

a. 0.95
b. 2.24
c. 15
d. 16
e. 24

Yanıt Anahtarı

1. d
2. d
3. c
4. b
5. a
6. b
7. a
8. c
9. c
10. c

Yararlanılan Kaynaklar

FINK, Arlene: **How to Sampling in Surveys,** Sage Publications, London, 1995.

GÜRSAKAL, Necmi: **Bilgisayar Uygulamalı İstatistik I,** Marmara Kitabevi,Bursa, 1997.

MALTHORA, Naresh K.: **Marketing Research An Applied Orientation,** 2^{nd} Edition, Prentice-Hall International Inc, New Jersey, 1996.

NETER, J; WASSERMAN, W, WHITMORE, G.A.: **Applied Statistics,** Simon and Schuster Inc, Boston, 1993.

ÖZMEN, A., ÖZDAMAR, K., ODABAŞI, Y., HOŞCAN, Y., BİR, A.A., KIRCAALİİFTAR, G., UZUNER, Y., **Sosyal Bilimlerde Araştırma Yöntemleri,** TC. Anadolu Üniversitesi Yayınları No:1081; Açıköğretim Fakültesi Yayınları No:601, Eskişehir, 1999.

PÜSKÜLCÜ Halis, İKİZ Fikret: **İstatistiğe Giriş,** (2. Baskı), E.Ü. Mühendislik Fakültesi Yayın No: 601, Ege Üniversitesi Basımevi, İzmir, 1986.

SERPER, Özer, AYTAÇ Mustafa: **Örneklem,** Ezgi Kitabevi, Bursa, 2000.

SERPER, Özer: **Uygulamalı İstatistik II,** Filiz Kitabevi, İstanbul, 1986.

TRYFOS, Peter: **Sampling Methods for Applied Research,** John Wiley & Sons Inc., New York, 1996.

TULL, Donald S., HAWKINS, Del I.: **Marketing Research Measurement and Method,** 6^{th} Edition, MacMillan Publishing Company, New York, 1993.

9

Hipotez Testleri

Çalışma Biçimine İlişkin Olarak

- *Bu üniteyi kolayca anlayabilmeniz için Örnekleme ve İstatistiksel Tahminleme isimli üniteler özümsenmiş olmalı,*
- *Kavramlar ve bu kavramlar arasındaki ilişkiler dikkatle incelenmeli,*
- *Örnekler ve örnek çözümleri dikkatle incelenmeli, sorunlarla karşılaşılırsa kavramsal açıklamalara geri dönülmelidir.*

Amaçlar:

- *İstatistiksel hipotez ve istatistiksel hipotez testi kavramlarını açıklayabileceksiniz.*
- *Parametrik ve parametrik olmayan teknikler arasından seçim yaparken, dikkat edilecek kriterleri açıklayabileceksiniz.*
- *Bir istatistiksel test sürecinin aşamalarında, hangi işlemlerin yapılacağını sıralayabileceksiniz.*
- *Anakütle aritmetik ortalamasına ve anakütle oranına ilişkin hipotez testi uygulamalarını yapabileceksiniz.*

İçerik Haritası

GİRİŞ

Örnekleme teorisi, anakütle parametrelerinin tahminlenmesi yanında, istatistiksel hipotezlerin test edilmesine de imkan vermektedir. Yorumsal istatistikte geleneksel karar alma işlemi olarak hipotez testi, örneklem bilgilerinden yararlanarak bu örneklemin çekildiği anakütlenin bir ya da daha fazla parametresi hakkında yorum yapma konularını içerir. Burada, örneklemdeki gözlem değerleri kullanılarak hesaplanan istatistiğin değeriyle, bu istatistiğin bilgi ürettiği parametrenin önceden belirlenmiş, bilinen değeri arasındaki farklılığın, istatistiksel olarak anlamlı olup olmadığı belirlenir. Farklılık varsa, bu farklılığın öneminin, sıfır hipotezini reddetmek için yeterli olup olmadığına karar verilir. Eğer sözkonusu farklılık anlamlı bir farklılıksa sıfır hipotezi reddedilir, tersi durumda kabul edilir.

Genellikle, parametrenin önceden belirlenmiş, bilinen değerinin değişmediğinin ifade edildiği sıfır hipotezine ilişkin karar verebilmek için, örneklem bilgilerinin olasılığa dayanarak genelleştirilmesi gerekir. Bu durum, ilgilenilen parametre hakkında bilgi üreten istatistiğin örnekleme dağılımının bilinmesini gerektirir.

Bu ünitede; anakütle aritmetik ortalaması ve anakütle oranına ilişkin hipotezlerin test edilmesinde, olasılık ve örnekleme dağılımı kavramlarının, nasıl uygulanacağı gösterilmiştir. Bu amaçla ünitede, önce hipotez ve hipotez testi kavramları açıklanmış, sonra da bir hipotez testi sürecinin aşamaları sırasıyla açıklanarak, hipotez türleri hakkında bilgi verilmiştir. Daha sonra da tek anakütle aritmetik ortalamasına ve oranına ilişkin hipotez testleri, örnekler üzerinde, ayrıntılı bir şekilde açıklanmıştır.

İSTATİSTİKSEL HİPOTEZ VE İSTATİSTİKSEL HİPOTEZ TESTİ

AMAÇ 1 ***İstatistiksel hipotez ve istatistiksel hipotez testi kavramlarını açıklayabileceksiniz.***

F çimento fabrikası ürünlerini, üzerinde ortalama 50 kg. yazan torbalarla, pazarlamaktadır. Z inşaat firması, fabrika yetkililerine başvurarak, son alınan 100 torbalık bir partinin ortalamasının 47.5 kg. olduğunu bildirerek, zarara uğradığını öne sürmüş ve fabrika yetkililerinden açıklama istemiştir. Fabrika yetkilileri, eğer kendi ürünlerini kullanmaya devam ederlerse, başka partilerin ortalamasının 50 kg. dan fazla çıkabileceğini ve zaman içinde giderek farkın sıfırlanacağını belirterek, bilinçli bir eksik (ya da fazla) doldurmanın söz konusu olmadığını ifade etmişlerdir.

Fabrika yetkililerinin savunmalarında ne kadar haklı oldukları (ya da olmadıkları) istatistiksel tekniklerle araştırılabilir. Bu ünitede, benzer problemlerin çözümlerinde kullanılan teknikler, yeterli ve ayrıntılarıyla ele alınmıştır.

Hipotez, karşılaşılan özel duruma 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(ler) hakkında bilgi üreten istatistik(ler)den ve bu istatistik(ler)in örnekleme dağılımıyla ilgili bilgilerden yararlanarak araştırılabilen önermelerdir. İstatistiksel hipotezler bir ya da daha fazla anakütle parametre değeriyle ilgili olabilirler. İstatistiksel hipotezleri diğer hipotezlerden ayıran özellik, bu hipotezlerin bir frekans dağılımına ait olmasıdır. Bazı istatistiksel hipotez örnekleri aşağıda verilmiştir.

İstatistiksel hipotez, herhangi bir ana kütle parametresine ilişkin olarak ileri sürülen ve doğruluğu olasılık kurallarıyla araştırılabilen önermedir.

Örnekler:

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

Anakütle parametreleri hakkındaki hipotezler (önermeler), parametre değer(ler)i hakkında, daha önceden bilinen bir düzey, standart bir değer ya da varsayımsal bir değer olabilir. Birinci örnekte, ilk 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 değerdir.

Bir istatistiksel hipotez, doğru ya da yanlış olabilir. Çünkü bu bir önermedir. Gerçeği öğrenebilmek için, anakütle 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 hipotezlerin, tanımlanan anakütleden seçilen örneklemin gözlem değerlerinden 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. Yorumsal istatistikte hipotez testi, örneklem gözlem değerlerinden yararlanarak, bu örneklemin seçildiği anakütlenin durumu hakkında yorum yapmaktır.

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

SIRA SİZDE

1. İstatistiksel hipotez nedir?

2. İstatistiksel hipotez testinin konusu nedir?

3. İstatistiksel hipotez neden doğru ya da yanlış olabilir açıklayınız.

HİPOTEZ TESTİ TÜRLERİ

AMAÇ 2

Parametrik ve parametrik olmayan teknikler arasında seçim yaparken, dikkat edilecek kriterleri açıklayabileceksiniz.

Hipotez testleri, ilgilenilen değişken(ler)in 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 ölçekle de elde edilen veriler üzerinde aritmetik işlemler yapmak mümkündür. Parametrik hipotez testlerinde, hipotezde bilinen bir olasılık fonksiyonundaki θ parametresinin önceden bilinen, θ_0 değerine eşit ya da bundan büyük, küçük ya da farklı olduğu ileri sürülebilir.

Parametrik testler örneklem sayısının tek ya da iki oluşuna ve iki örneklemin varlığında, bu örneklemlerin bağımsız ya da bağımlı oluşuna bağlı olarak sınıflandırılırlar. En önemli parametrik testler z ve t testleridir. Bu ünitede tek anakütle (ya da tek örneklem) ortalamasına ilişkin z ve t testleriyle tek anakütle (ya da tek örneklem) oranına ilişkin z testi ayrıntılı olarak incelenmiştir.

Parametrik olmayan testler, anakütle dağılımı nasıl olursa olsun uygulanabilen testlerdir. Bu testlerde, parametrelerle ilgilenilmeyip, hipotezler, ilgili değişkenin belirli bir nitel özelliğine göre oluşturulur.

Parametrik olmayan testler, değişkenlerinin ölçülmesinde, sınıflayıcı ya da sıralayıcı ölçeğin kullanıldığı hipotez testleridir. Bu tür testlerde, parametrik testlerde olduğu gibi, anakütlenin (örneklemin) tek ya da iki oluşuna ve iki anakütle (iki örneklem) sözkonusu olduğunda da örneklemlerin bağımsız ve bağımlı oluşuna göre sınıflandırılırlar.

Parametrik olmayan hipotez testlerine ilişkin “Ki-Kare Testi” ayrı bir ünitede incelenmiştir.

SIRA SİZDE

1. Değişkenlerin ölçülmesinde benimsenen ölçeğe bağlı olarak, istatistiksel hipotezler nasıl sınıflandırılır?

2. Değişkenlerin ölçülmesindeki, eşit aralıklı ölçeğin kullanılmasında, hangi test türü kullanılır?

3. Parametrik olmayan testlere hangi durumlarda başvurulur?

HİPOTEZ TESTİ SÜRECİNİN ADIMLARI

AMAÇ 3

Bir istatistiksel test sürecinin aşamalarında, hangi işlemlerin yapılacağını sıralayabileceksiniz.

Anakütle parametre değerleri hakkında ileri sürülen iddiaların test edilmesinde aşağıdaki adımlar izlenir:

Hipotezlerin İfade Edilmesi

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

Sıfır hipotezi (H_0), ilgili ana kütle parametresinin bilinen değerinde herhangi bir farklılığın beklenmediğini ifade eden hipotezdir.

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

$$H_0 : \theta = \theta_0$$

Karşıt hipotez (H_1), ilgili ana kütle parametresinin bilinen değerinde istatistiksel olarak anlamlı farkların beklendiğini ifade eden hipotezdir.

şeklinde ifade edilir.

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 incelendiği hipotezdir. Karşıt hipotez, parametrenin önceden belirlenmiş, bilinen değerinde bazı farklılığın ya da etkinin beklendiğinin ifade edildiği hipotezdir. Bu hipotez araştırmanın amacına bağlı olarak, aşağıdaki üç farklı şekilden birisiyle ifade edilmiş olur:

$$H_1 : \theta \neq \theta_0$$
$$H_1 : \theta > \theta_0$$
$$H_1 : \theta < \theta_0$$

Birinci ifadede, H_0 hipotezi; verilecek kararın, anakütle parametre değerinden, her iki yöndeki (hem küçük hem de büyük yöndeki), anlamlı farklılıklardan etkileneceğini ifade eder. İkinci ifadede, verilecek kararın, anakütle parametre değerinde, sadece büyük yöndeki anlamlı sapmadan etkileneceğini, son ifadedeyse sadece küçük yöndeki anlamlı farklılığın, verilecek kararı etkileyeceğini ifade eder.

Hipotez testlerinde H_1 hipotezi, testin yönünü ya da H_0 hipotezinin reddedileceği bölgenin yerini belirleyen hipotezdir. Red bölgesi, H_0 hipotezinin reddedilmesine (H_1 hipotezinin kabul edilmesine) neden olan örneklem istatistiği $\hat{\theta}$ (ya da test istatistiği) ile ilgili değerler aralığıdır. Kabul bölgesiyse, 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: "İki yönlü test", "tek yönlü üstkuyruk testi" ve "tek yönlü alt kuyruk testi" olarak isimlendirilirler. Bu testlere ilişkin hipotezlerin ifade ediliş biçimi aşağıda verilmiştir.

İki Yönlü Testlerde Hipotezler:

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta \neq \theta_0$$

Tek Yönlü ÜstKuyruk Testlerinde Hipotezler:

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta > \theta_0$$

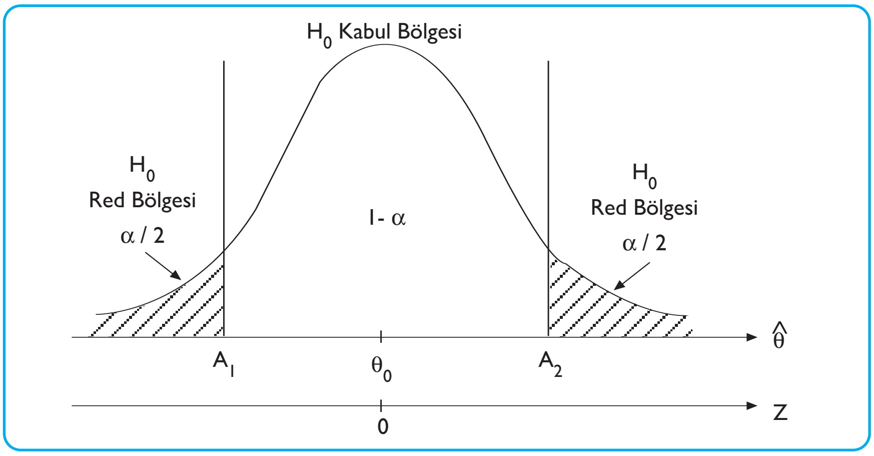
Tek Yönlü AltKuyruk Testlerinde Hipotezler:

$H_0 : \theta = \theta_0$

$H_1 : \theta < \theta_0$

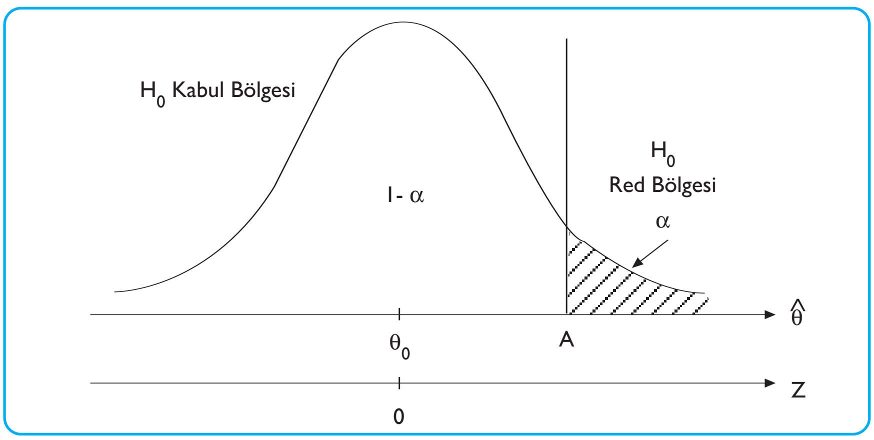
şeklinde belirlenir.

Yukarıdaki her hipotez takımında kullanılan isim, H_1 hipotezinde θ için verilen değerler aralığını açıklamaktadır. Bu durum, örneklem istatistiğinin $\hat{\theta}$'nın normal dağılıma sahip olduğu kabul edilerek aşağıdaki şekillerle açıklanmıştır. Örneğin iki yönlü hipotezlerde H_1 hipotezi Şekil 9.1'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 9.1 *İki Yönlü Testlerde Red Bölgeleri.*

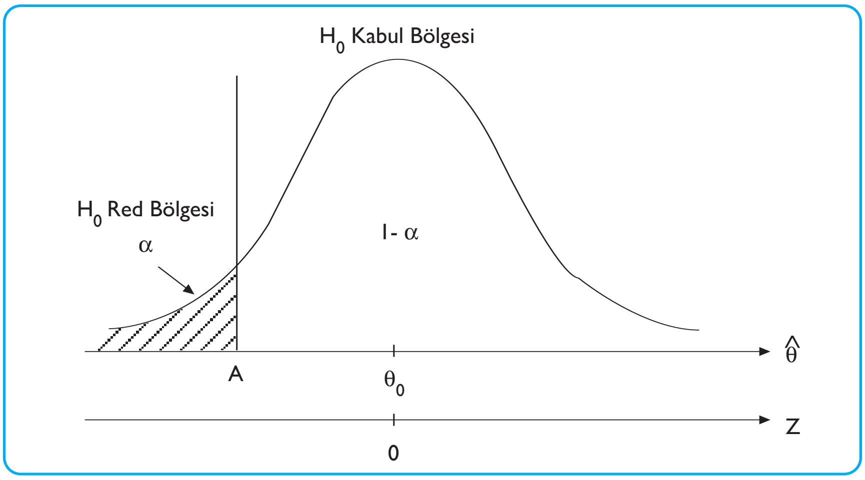
Tek yönlü üstkuyruk testlerinde, H_1 hipotezi, θ_0'dan büyük olan θ ile ilgili değerleri içerdiği için, bu isim verilmiştir. Tek yönlü üstkuyruk testlerinde, H_1 hipotezi, (Şekil 9.2'de görüldüğü gibi) $\hat{\theta}$'nın, θ_0'dan büyük olmak üzere, belirli bir A değerinden büyük değerleri H_1 hipotezi yönünde, H_0 hipotezinin red bölgesinde yer almaktadır.



Şekil 9.2 *Tek Yönlü Üst Kuyruk Testlerinde Red Bölgesi.*

Tek yönlü altkuyruk testlerindeyse tek yönlü üst kuyruk testinin tam tersine, (Şekil 9.3'de 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.



Şekil 9.3 *Tek Yönlü Alt Kuyruk Testlerinde Red Bölgesi.*

Anlamlılık Düzeyinin Belirlenmesi

Bir istatistiksel hipotez testinde 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 anakütleden rassal olarak seçilen, aynı hacimli farklı örneklemler için hesaplanan istatistikler, örneklemden örnekleme değişen değerler aldığından, anakütle 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 anlamlılık düzeyi" denir. Anlamlılık düzeyinin belirlenmesi, doğru olan sıfır hipotezinin, örneklemden elde edilen bilgilere dayanarak reddedilmesi olasılığını belirleyen α'nın seçilmesidir. α anlamlılık düzeyi, araştırmacı tarafından, hipotezler ifade edilip veri derlemeye başlamadan önce seçilmelidir. 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 istatistiği 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 9.1, 9.2 ve 9.3'te gösterilmiştir. Şekillerdeki A, A_1 ve A_2 noktaları red bölgelerinin başlangıç noktalarıdır.

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

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.

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.

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

Araştırmalarda 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.

Verilerin Derlenmesi

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

Test İstatistiğinin Seçilmesi

Daha önce de belirtildiği gibi, anakütleden rassal örneklem alınmış olsa bile, hesaplanan örneklem istatistiğinin anakütle 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?

Bu sorunun yanıtlanabilmesi, için, sıfır hipotezinin test edilebilmesinde, örneklem istatistiğinin dağılımının bilinmesine ve uygun test istatistiğine gereksinim vardır.

Hipotez testlerinde, örneklem istatistiğinin dağılımının bilinmesi zorunludur.

Test istatistiği, örneklem istatistiğinin değeriyle anakütlenin, sıfır hipotezinde ifade edilen değeri arasındaki farkın, standartlaştırılmış değeri olarak tanımlanır. Başka bir ifadeyle test istatistiği, örneklem istatistiği $\hat{\theta}$ ile θ_0 arasındaki farkı standart hata birimiyle ifade eden ölçüdür. Test istatistiği örneklemin sıfır hipotezine ne kadar uyduğunu gösterir. Bu nedenle de test istatistiği 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 ya da Ki-Kare dağılımı v.b. gibi bilinen dağılımlara uyar.

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

Bu ünitede, bazı parametrelere ilişkin hipotezlerin testinde, z ve t test istatistiklerinin seçilme gerekçeleri ve uygulamalarına yer verilmiştir.

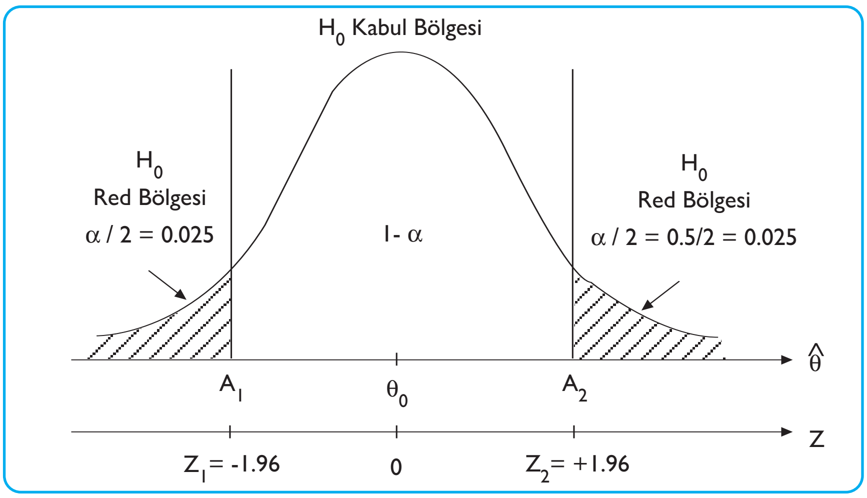
İstatistiksel Kararın Verilmesi

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, $\hat{\theta}$ istatistiğinin 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 düzeyine, 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 normal dağılıma sahip olduğu kabul edilerek yapılmıştır. Açıklamalarda $\alpha = 0.05$ seçilmiştir.

Eğer karşıt hipotez $H_1 : \theta \neq \theta_0$ şeklinde ifade edilmişse, red bölgesi Şekil 9.4'te gösterildiği gibi $\hat{\theta}$ istatistiğine ilişkin normal dağılımın her iki ucunda simetrik olarak tanımlanmış olur ve her birinin 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 9.4 İki Yönlü Testlerde Red Bölgeleri ve Kritik Değerler.

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

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

standart değerleri olur. A_1 örneklem değeri θ_0'ın solunda (A_1 θ_0'dan küçük değerli) olduğu için, z_1 negatif ve 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_1 ya da z_2 kritik değerleri $\alpha = 0.05$ anlamlılık düzeyi için Ek-1'de verilen Standart Normal Eğri Alanları Tablosundan yararlanılarak belirlenir. Z tablo değeri (z_{tab}),

$z_{tab} = z_{0,5-0,025} = z_{0.4750} = 1.96$ dır.

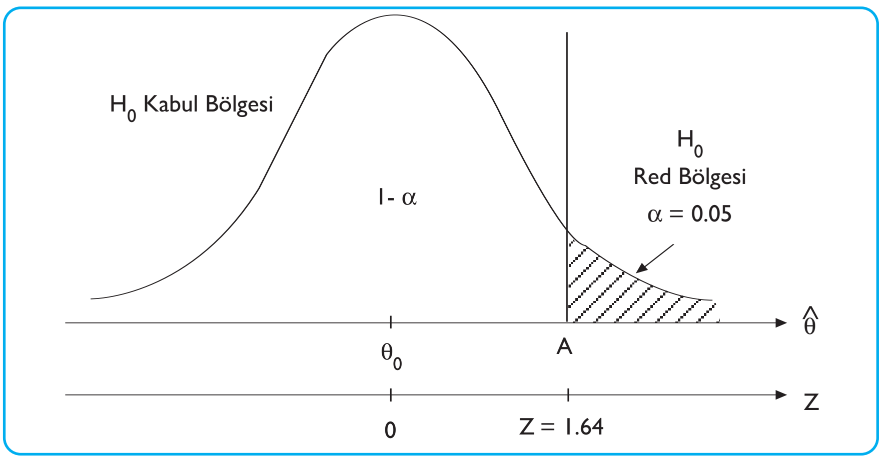
$z_{tab} = \pm 1.96$ değeri standart normal dağılımda %47.5'lik oransal alana karşı gelen örneklem istatistiğinin standart değeridir. İki yönlü testte H_0 hipotezinin reddedilmesi için,

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

koşulunun sağlanması gerekir. Tersi durumda H_0 hipotezi kabul edilir.

Eğer $H_1 : \theta > \theta_0$ ya da $H_1 : \theta < \theta_0$ şeklinde ifade edilmişse, H_0 hipotezinin red bölgesi, birinci durumda, $\hat{\theta}$ istatistiğine ilişkin normal dağılımın üstkuyruğunda, ikinci durumda altkuyruğunda tanımlanmışsa alan $\alpha = 0.05$ olur. Buna bağlı olarak kritik değerler sırasıyla (Şekil 9.5'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 9.5 *Tek Yönlü Üst Kuyruk Testlerinde Red Bölgesi ve Kritik Değer.*

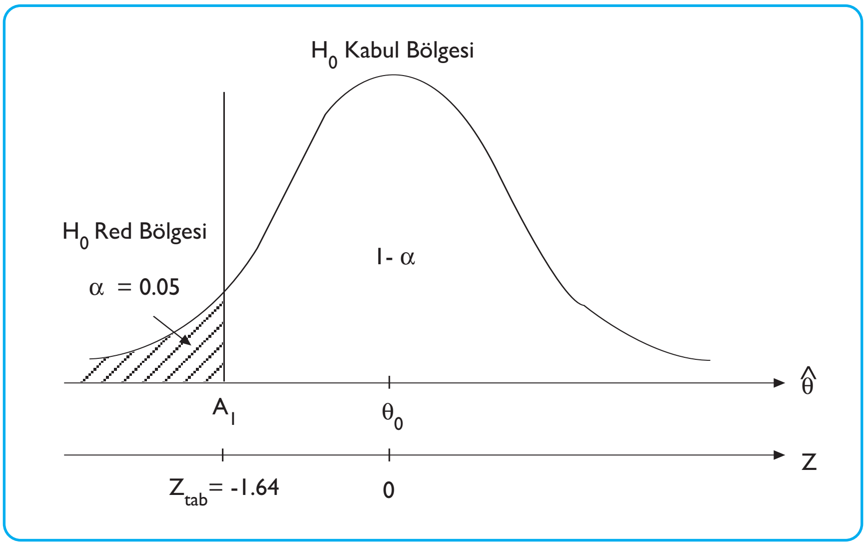
Şekil 9.6'da gösterildiği gibi olur. Bu z değerleri birbirinin simetriğidir. Üstkuyruk testinde, z pozitif, altkuyruk testinde, z negatif işaretlidir. z değerleri $\alpha = 0.05$ için Ek-1'de verilen Standart Normal Eğri alanları tablosundan yararlanılarak belirlenirler.

$z_{tab} = z_{0.5-0.05} = z_{0.4500} = 1.64$ tür.

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

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

olmalıdır.



Şekil 9.6 *Tek Yönlü Alt Kuyruk Testlerinde Red Bölgesi ve Kritik Değer.*

Tek yönlü altkuyruk testinde H_0'ın reddedilebilmesi için,

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

koşulunun sağlanması gerekir. Tersi durumda H_0 hipotezi kabul edilir.

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

Probleme İlişkin Kararın Verilmesi

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

SIRA SİZDE

1. **Hangi hipotez test edilecek hipotezdir?**
2. **Hangi durumda iki yönlü teste başvurursunuz?**
3. **I. Tip hata ne demektir, bu hatanın büyüklüğü nasıl belirlenir?**

TEK ANAKÜTLE PARAMETRESİYLE İLGİLİ HİPOTEZ TESTLERİ

AMAÇ 4

Anakütle aritmetik ortalamasına ve anakütle oranına ilişkin hipotez testi uygulamalarını yapabileceksiniz.

Pek çok araştırmada, tek bir anakütlenin bir parametresinin değerine ilişkin, hipotezlerin ileri sürüldüğü görülmektedir. Başka bir ifadeyle, bir anakütlenin ilgilenilen bir değişkeni hakkında, bilinen ya da belirlenen bir standarta göre yorumların yapıldığı görülmektedir. Aşağıdaki hipotezler tek anakütle parametresiyle ilgili hipotez testine örnek verilebilir: A ürününün reklamını beğenenlerin oranı en az %45'tir. Günlük ortalama üretimi 100 Ton olan bir üretim sürecinde yapılan değişiklik, günlük ortalama üretim miktarını arttırmıştır (vb.) gibi. Bu tür hipotezlerin testinde, tanımlanan bir anakütlenin ilgilenilen bir değişkenine ilişkin önceden belirlenen (ya da bilinen) bir parametre değerinin (θ_0'ın) değişmediği şeklindeki sıfır hipotezi test edilir. Böylece, verilen karara göre, karşıt hipotezde (araştırma hipotezinde) ileri sürülen iddianın, kabul edilip edilmeyeceği ortaya çıkar.

Bu testlerde karar verilirken örneklem istatistiğinin değeriyle bu istatistiğin bilgi ürettiği parametrenin bilinen ya da belirlenen θ_0 değeri karşılaştırılır.

Ünitenin izleyen bölümlerinde, uygulamada sıkça karşılaşılan tek anakütle parametresiyle ilgili olarak, anakütle ortalamasına ilişkin testlerle anakütle oranına ilişkin testler ele alınmıştır.

Anakütle Ortalamasına İlişkin Hipotez Testleri

Bu testlerde, tanımlanan anakütleden rassal olarak seçilen bir örneklem için hesaplanan $\overline{X}$ değeriyle, bu örneklemin seçildiği anakütlenin aritmetik ortalaması μ ile ilgili, önceden belirlenen (ya da bilinen) μ_0 gibi bir değer arasındaki farklılığın istatistiksel olarak anlamlı olup olmadığı araştırılır. Belirlenen farklılığın, sıfır hipotezini reddetmek için, yeterli olup olmadığına karar verilir.

Anakütle ortalamasına ilişkin hipotez testleri uygulamada, sıkça kullanılan önemli parametrik testlerdir.

Bu hipotezlerin test edilmesine ilişkin açıklamalar, örneklem hacminin büyük olması ($n \geq 30$ birim) ve örneklem hacminin küçük olması ($n < 30$ birim) durumları için, alt başlıklar altında, aşağıdaki örnek problemler üzerinde ayrıntılı olarak ele alınmıştır.

Anakütle Ortalamasına İlişkin Büyük Örneklem Testi

Bu test türünde :

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

ÖRNEK 1

Bir peynir üretim sürecinde, üretimin 500 gr.'lık paketler halinde gerçekleştirilmesi planlanmıştır. Üretimin planlandığı gibi gerçekleşip gerçekleşmediğini kontrol amacıyla rassal olarak 100 paket seçilmiş ve bu paketler için ortalama ağırlık 495 gr., standart sapma da 20 gr. olarak hesaplanmıştır. $\alpha = 0.05$ anlam düzeyi için, üretimin planlandığı gibi gerçekleştiği söylenebilir mi? Karar veriniz.

ÇÖZÜM

1. Adım: Hipotezlerin ifade edilmesi

Peynir paketlerinin belirlenen ortalama ağırlığı (standart ağırlık) 500 gr.dır. Bu nedenle, burada sıfır hipotezi, üretilen peynir paketlerinin ortalama ağırlığının 500 gr. olduğu yönündedir. Bu iddiayı, 500 gr.'dan hem küçük, hem de büyük yöndeki anlamlı ağırlık farklılıkları çürütecektir. Başka bir ifadeyle, bu anlamlı farklılıklar üretimin planlandığı gibi gerçekleşmediğini gösterecektir. Buna göre yapılacak test, iki yönlü test olup, hipotezler:

$H_0 : \mu = 500$ gr.
$H_1 : \mu \neq 500$ gr.

şeklinde ifade edilmelidir.

2. Adım: İstatistiksel test

Bu örnekte tanımlanan anakütleler sonsuz olurlar. Anakütlenin dağılımı ve değişkenliği hakkında bilgi yoktur. Örneklem hacmi $n = 100$ pakettir ve $n \geq 30$ olduğundan, (daha önce açıklanmış olduğu gibi) örneklem aritmetik ortalamasının örnekleme dağılımı, normal dağılımdır. Kullanılması gereken test istatistiğide örneklem aritmetik ortalamasının standart değeri olan z istatistiğidir. Bu nedenle burada z testi uygulanmalıdır.

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

Anakütle standart sapması bilinmediği için $\sigma_{\overline{X}}$ 'nın tahmini olan $s_{\overline{X}}$ kullanılmalıdır.

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

3. Adım: Anlamlılık düzeyinin belirlenmesi

Problemde doğru olan hipotezinin reddedilmesi olasılığı α, %5 seçilmiştir. Red bölgeleri, ortalamanın örnekleme dağılımının her iki kuyruğunda tanımlandığı için, red bölgelerinin her birinin oransal büyüklüğü, $\frac{\alpha}{2} = 0.05/2 = 0.025$ dir.

4. Adım: H_0'ın red bölgesinin belirlenmesi

H_1 hipotezi, testin red bölgesinin yönünü belirlediğine göre, bu testte red bölgesi örneklem ortalamasının, örnekleme dağılımının simetrik olması nedeniyle hem alt kuyruğunda hem de üst kuyruğunda tanımlanmıştır. Bu durum Şekil 9.7'de gösterilmiştir.

5. Adım: Test istatistiğinin hesaplanması

n = 100 paket

$\bar{X}$ = 495 gr.

s = 20 gr.

μ_0 = 500 gr.

α = 0.05

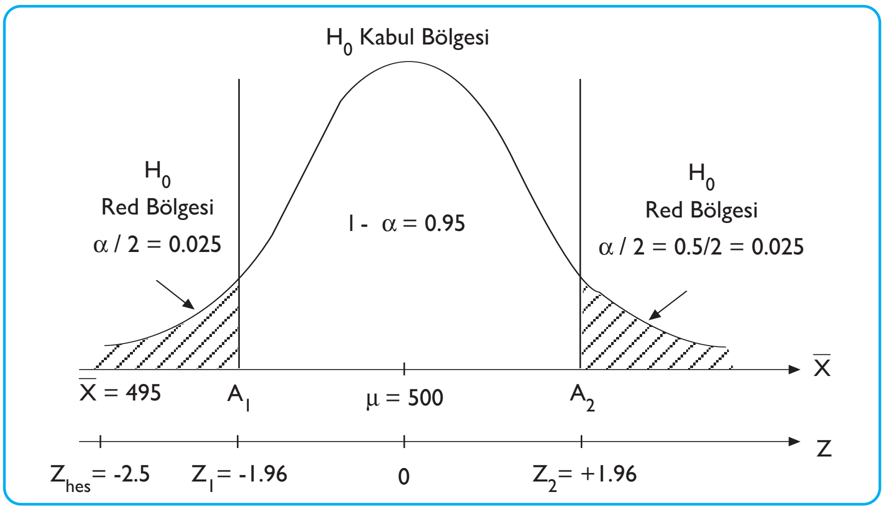
$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{20}{\sqrt{100}} = 2$$

ve

$$z = \frac{\bar{X} - \mu_0}{s_{\bar{x}}} = \frac{495 - 500}{2} = -2.5$$

olarak hesaplanır.

$\bar{x}$ '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 red bölgesi Şekil 9.7'de gösterildiği gibi bu dağılımın her iki kuyruğunda tanımlanmıştır ve oransal büyüklükleri $\alpha/2 = \frac{0,05}{2} = 0.025$ dir. Buna göre, bu dağılımın alt kuyruk bölgesinde tanımlanan red bölgesinin sınır değeri $z_1 = -1.96$, üstkuyruk bölgesinde tanımlanan red bölgesinin sınır değeri $z_2 = 1.96$ olacaktır.



Şekil 9.7 *Anakütle Ortalaması μ İçin İki Yönlü Test Sonuçları*

Hesaplanan test istatistiği $z_{hes} = |2.5| > z_{tab} = 1.96$ olduğundan H_0 hipotezi reddedilir, dolayısıyla H_1 kabul edilir. Ayrıca aynı karar Şekil 9.7'de $z_{hes} = -2.5$ standart değerinin Z_i eksenindeki red bölgesinde yer aldığını söylemek suretiyle de açıklanabilir.

H_0 hipotezinin reddedilmesi, üretilen peynir paketlerinin ortalama ağırlığının 500 gr. olmadığını, üretim sisteminin planlandığı gibi üretimi gerçekleştirmediğini gösterir.

ÖRNEK 2

Bir firmanın pazarlama yöneticisi, üniversite öğrencilerinin aylık ortalama gazlı içecek tüketimlerinin en az 20 lt. olabileceğini düşünmektedir. Eğer yöneticinin bu düşüncesi doğrusa üniversite öğrencilerine yönelik yeni stratejiler geliştirilecektir. Bu amaçla, rassal olarak seçilen, 1000 üniversite öğrencisi üzerinden veriler derlenmiş ve bu öğrencilerin ortalama gazlı içecek tüketiminin 22 lt. ve standart sapmasının da 8 lt. olduğu hesaplanmıştır. Yöneticinin düşüncesinin doğru olup olmadığına $\alpha = 0.01$ anlam düzeyini kullanarak karar veriniz.

ÇÖZÜM

1. Adım: Hipotezlerin ifade edilmesi

Burada verilecek karar, ortalama gazlı içecek tüketiminin 20 lt.'den fazla olmadığıdır. Araştırma hipoteziyse 20 lt. ve daha fazla olduğudur.

Buna göre hipotezler:

$H_0 : \mu = 20$ lt.

$H_1 : \mu > 20$ lt.

şeklinde ifade edilmelidir.

2. Adım: İstatistiksel test

Örneklem hacmi $n = 1000$ öğrenci ($n > 30$ birim) olduğu için $\bar{x}$ 'nın örnekleme dağılımı normaldir. $\bar{x}$ 'nın standart sapması ya da standart hata $s_{\bar{x}}$ birimi cinsinden $\bar{x} = 22$ değerinin bilinen $\mu = 20$ değerinden ne kadar farklılık gösterdiğini ölçmek için standartlaştırılmış z değişkeni kullanılır. z standartlaştırılmış test istatistiği olarak ifade edilir.

$$z = \frac{\bar{x} - \mu_0}{s_{\bar{x}}}$$

$\bar{x}$ 'nın örnekleme dağılımı normal olduğu için $\mu = \mu_0 = 20$ olduğu zaman, z standart normal dağılıma sahip, rassal değişkendir. Bu nedenle, bu hipotezlerin testi için z testi uygulanmalıdır.

3. Adım: Anlamlılık düzeyinin belirlenmesi

Firma yöneticisi, doğru olan H_0 hipotezinin reddedilmesi olasılığı α'yı, 0.01 olarak seçmiştir. α'nın küçük seçilmiş olması yöneticinin, düşüncesinde çok fazla kararlı olmadığını gösterir. Çünkü H_0 hipotezinin kabul bölgesinin oransal büyüklüğü $1 - \alpha = 0.99$'dur. Red bölgesinin büyüklüğü $\alpha = 0.01$'dir. Bu durum Şekil 9.8'de gösterilmiştir.

4. Adım: H_0'ın red bölgesine karar verilmesi

Bu testte $H_1 > 20$ lt. olarak ifade edildiği için H_0 hipotezinin red bölgesi (Şekil 9.8'de gösterildiği gibi) dağılımın üst kuyruğunda tanımlanmıştır. Yani hipotez, tek yönlü üst kuyruk testiyle test edilecektir.

5. Adım: Test istatistiğinin hesaplanması

$\mu_0 = 20$ lt.

$\overline{X} = 22$ lt.

$s = 8$ lt.

$n = 1000$ öğrenci

$\alpha = 0.01$

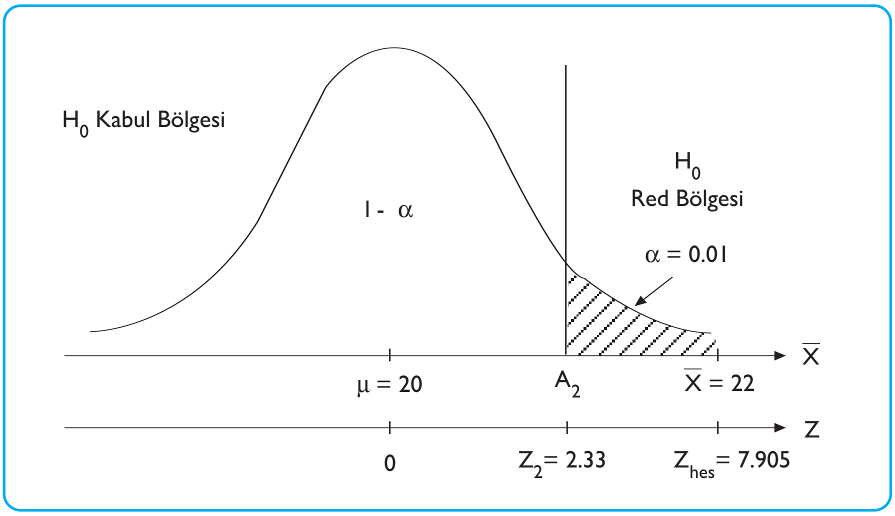
$$s_{\overline{x}} = \frac{s}{\sqrt{n}} = \frac{8}{\sqrt{1000}} = \frac{8}{31.62} = 0.253$$

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

$$z = \frac{22 - 20}{0.253} = 7.905$$

$$z_{hes} = 7.905 > z_{tab} = 2.33$$

olduğu ya da hesaplanan $z_{hes} = 7.905$ değeri (Şekil 9.8'de görüldüğü gibi) red bölgesinde kaldığı için H_0 hipotezi red (dolayısıyla H_1 hipotezi kabul) edilir.



Şekil 9.8 *Anakütle Ortalaması İçin Tek Yönlü ÜstKuyruk Testi Sonuçları.*

Yukarıdaki istatistiksel karara göre, öğrencilerin aylık ortalama gazlı içecek tüketimi 20 lt. den fazladır. Yönetici, üniversite öğrencilerine yönelik yeni stratejiler geliştirmelidir.

Anakütle Ortalamasına İlişkin Küçük Örneklem Testi

Araştırmaların bir çoğunda araştırmaya ayrılan para, zaman ve diğer imkanların sınırlı olması gibi nedenlerle, örneklem hacmini, daha önceki açıklamalarımızda belirtilen büyüklükte (genellikle $n \geq 30$ birim) sağlamak mümkün 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 anakütle ortalaması için kurulan hipotezlerin, küçük örneklemler (n < 30 birim) kullanılarak, nasıl test edileceği konusu ele alınmıştır.

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

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

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

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

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 $\bar{x} - \mu_0 / s_{\bar{x}}$ standart normal dağılım göstermemekte, dolayısıyla büyük örneklemlerde izlenen yöntem geçerli olmamaktadır. Normal dağılıma sahip ve değişkenliği bilinmeyen bir anakütleden seçilen 30'dan daha az birim içeren bir örneklemin aritmetik ortalaması, n – 1 serbestlik derecesiyle t dağılır. t istatistiği,

Küçük örneklem testlerinde test istatistiği, $t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}}$ 'dir.

$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}}$$

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

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

eşitliği ile hesaplanır.

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

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

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

ÖRNEK 3

Bir okuldan geçmiş yıllarda mezun olan öğrencilerin ortalama mezuniyet puanı 66 puandır. Bu yıl mezun olan öğrenciler arasından 26 öğrenci rassal olarak seçilmiş ve bunların ortalama mezuniyet puanının 70 puan ve standart sapmasının 10 puan olduğu hesaplanmıştır. Geçmiş yıllarda mezun olan öğrencilerin ortalama mezuniyet puanıyla bu yıl mezun olanların ortalama puanları arasında farklılık var mıdır? $\alpha = 0.01$ için test ediniz.

ÇÖZÜM

1. Adım: Hipotezlerin ifade edilmesi

Bu problemde, iki dönemdeki mezuniyet ortalamaları arasında bir farklılık olmadığı yönündeki sıfır hipotezi test edilecektir. Buna göre hipotezler

$H_0 : \mu = 66$ puan

$H_1 : \mu \neq 66$ puan

şeklinde ifade edilmelidir.

2. Adım: İstatistiksel test

Örneklem hacmi n = 26 öğrencidir ve öğrencilerin geçmiş yıllardaki mezuniyet puanlarının dağılımı normal dağılıma sahiptir. 26 öğrencinin oluşturduğu örneklem küçük örneklemdir. H_0 hipotezinin testi için küçük örneklem testlerinde kullanılan t test istatistiği

$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}}$$

kullanılır. Bu nedenle de t testi uygulanmalıdır.

3. Adım: Anlamlılık düzeyinin belirlenmesi

Probleme ilişkin testte, $\mu = \mu_0 = 66$ puan olduğunda, testin kontrolünün olasılık düzeyi %1 olarak düşünülmüştür.

4. Adım: H_0'ın red bölgesinin belirlenmesi

H_0 hipotezi, iki yönlü testle test edilecektir. Hipotezin red bölgeleri, t dağılımının her iki kuyruğunda tanımlanmıştır. Şekil 9.9'daki taralı alanlar H_0 hipotezinin red bölgeleridir.

5. Adım: Test istatistiğinin hesaplanması

$n = 26$

$\bar{x} = 70$

$s = 10$

$\mu_0 = 66$

$\upsilon = 26 - 1 = 25$

$$s_{\bar{x}} = \frac{s}{\sqrt{n-1}} = \frac{10}{\sqrt{25}} = 2$$

ve

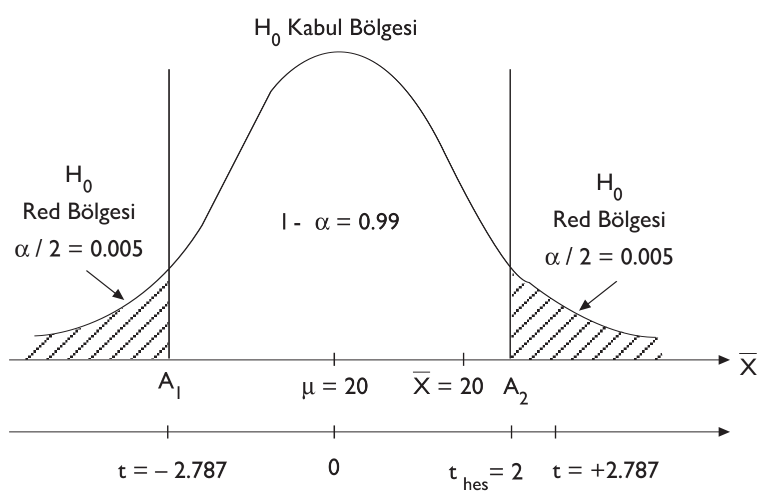
$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} = \frac{70 - 66}{2} = 2$$

olarak elde edilir.

Test edilecek hipotez iki yönlüdür. $\alpha = 0.01$ ve $n = 26$ için t tablo değeri (t_{tablo})

$t_{tablo} = t\,(1 - \alpha/2\ ;\ \upsilon = n - 1) = 2.787$ olarak belirlenir.
$(t_{hes} = 2) < (t_{tablo} = |2.787|)$

Olduğundan, istatistiksel karar H_0 kabul, H_1 red edilir şeklinde olacaktır. Şekil 9.9'da görüldüğü gibi hesaplanan test istatistiği değeri, kabul bölgesinde yer almaktadır.



***Şekil 9.9** Küçük Örneklemlerde μ İçin İki Yönlü Test Sonuçları.*

Yukarıdaki istatistiksel karara göre, eski ve yeni öğrencilerin mezuniyet puanları arasındaki farklılık, örnekleme hatasından kaynaklanmaktadır; eski ve yeni öğrencilerin ortalama mezuniyet puanları arasında anlamlı bir fark yoktur.

Anakütle Oranına İlişkin Test

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 anakütlenin, ilgilenilen iki şıklı bir değişkeninin, herhangi bir şıkkına sahip birimlerinin oranına "anakütle oranı" denir ve Π simgesiyle gösterilir. Ünitenin bu kesiminde, anakütle oranı Π'nin değeri hakkında ileri sürülen bir önermenin, nasıl test edileceği konusu ele alınmıştır. Tek anakütle oranına ilişkin test olarak isimlendirilen bu testin, örneklem oranı p ile anakütle oranı Π'nin iddia edilen değeri Π_0 arasındaki farkın istatistiksel olarak anlamlı olup olmadığı araştırılır. Örneklem hacmi yeterli büyüklükte olduğunda ($n > 30$ birim), anakütle oranı Π hakkındaki testler daha önce açıklanan anakütle ortalaması μ için büyük örneklem testlerindekilere benzer şekilde yapılır. Ancak, test için örneklem istatistiği olarak örneklem oranı p ve bu istatistiğin örnekleme dağılımı kullanılır. $n \geq 30$ 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.

Anakütle oranı Π'ye ilişkin testlerde örneklem hacmi büyük olduğunda, standartlaştırılmış z test 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.

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

ÖRNEK 4

Bir yönetici, işletmesinde toplam kalite yönetimi uygulamayı düşünmektedir. Eğer işçilerinin en az %60'ı bu düşünceden yanaysa yönetici düşüncesini uygulamaya karar verecektir. Bu amaçla rassal olarak 750 işçi seçilmiş ve bunların 495'inin toplam kalite yönetimi düşüncesini benimsediği tespit edilmiştir. Yönetici, toplam kalite yönetimine geçmeli midir? $\alpha = 0.01$ anlam düzeyi kullanarak karar veriniz.

ÇÖZÜM

1. Adım: Hipotezlerin ifade edilmesi

Bu örnekte; sıfır hipotezi, toplam kalite yönetimi uygulamasını düşünenlerin oranı %60 ve daha azdır şeklindedir. Sıfır hipotezi tek yönlü karşıt (araştırma) hipotezi ile test edilecektir. Hipotezler:

$H_0 : \Pi = 0.60$
$H_1 : \Pi > 0.60$

şeklinde ifade edilir.

2. Adım: İstatistiksel test

$H_1 : \Pi > \Pi_0 = 0.60$ olduğundan, H_0'ın kabul ya da reddi için uygulanacak test ana kütle oranına ilişkin tek yönlü üst kuyruk testi olmalıdır.

3. Adım: Anlamlılık düzeyinin belirlenmesi

Red bölgesi, (Şekil 9.10'da gösterildiği) gibi oranların örnekleme dağılımının üstkuyruğunda tanımlanmıştır. Anlamlılık düzeyi $\alpha = 0.01$ benimsenmiştir. Testin red bölgesinin oransal büyüklüğü 0.01'dir.

4. Adım: H_0'ın red bölgesinin belirlenmesi
Örneklem oranının %60'a eşit ve küçük olması durumunda H_0 hipotezinin reddedilmesi söz konusu değildir.

5. Adım: Test istatistiğinin hesaplanması
Test istatistiğinin hesaplanması problemde belirtildiği gibi, 750 işçinin 495'i toplam kalite yönetimi uygulamasından yana düşünceye sahiptir yani örneklem oranı p:

$$p = \frac{r}{n} = \frac{495}{750} = 0.66$$

dır.

Örneklem oranı $p = 0.66$ ile anakütle oranı Π hakkındaki hipotez değeri $\Pi_0 = 0.60$ arasındaki fark, istatistiksel olarak anlamlı bir farklılık mıdır? Başka bir ifadeyle örneklem oranı p=0.66 olması örnekleme hatasından mı kaynaklanmıştır? Bu sorunun yanıtlanabilmesi için, test istatistiğinin hesaplanması gerekir.

n = 750 işçi ($n \geq 30$ birim) olduğu için, örneklem oranının örnekleme dağılımı, normal dağılım gösterir. Buna göre uygulanacak test istatistiği:

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

olur. Örnek için, standart hata:

$$\sigma_p = \sqrt{\frac{\Pi_0(1-\Pi_0)}{n}} = \sqrt{\frac{(0.60)(0.40)}{750}} = 0.0179$$

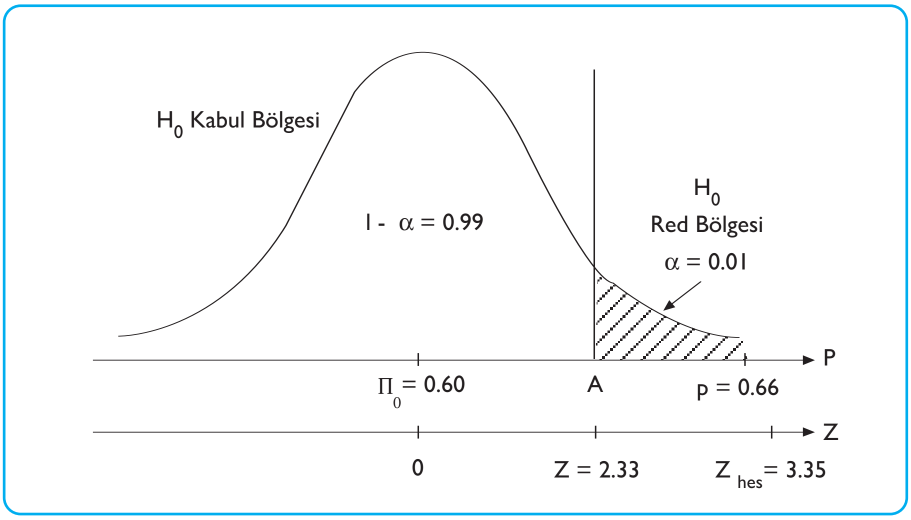
ve

$$z = \frac{0,66 - 0.60}{0.0179} = 3.35$$

olarak elde edilir.

Test istatistiğinin dağılımı normal olduğu ve $\alpha = 0.01$ için kritik z_{tablo} değeri (standart normal eğri alanları tablosundan) $z_{tablo} = 2.33$ olarak belirlenir. Bu değer standart normal dağılımın oransal olarak 0.495'lik alanına karşı gelmektedir. z test istatistiğinin hesaplanan değeri $z_{hes} = 3.35$, $z_{tablo} = 2.33$ değerinden büyük olduğu ($z_{hes} = 3.35 > z_{tablo} = 2.33$) için $H_0 : \Pi = 0.60$ hipotezi reddedilir (dolayısıyla H_1 hipotezi kabul edilir). Örneklem oranı p = 0.66 değeri istatistiksel olarak anlamlı bir farklılığı göstermektedir; örneklem oranı p = 0.66 değeri için standart değer $z_{hes} = 3.35$ red bölgesinde yer almaktadır.

Bu kararın anlamı: Toplam kalite yönetimini benimseyen işçinin oranı %60'tan büyüktür; yönetici toplam kalite yönetimine geçme düşüncesini uygulamalıdır.



Şekil 9.10 *Anakütle Oranı Π İçin Test Sonuçları.*

ÖRNEK 5

Pazarlama yöneticisi, Üniversite öğrencileri arasında günde 2 ve daha fazla gazlı içecek tüketenlerin oranının en fazla %40 olduğuna inanmaktadır. Eğer bu doğruysa yönetici, üniversite öğrencilerine yönelik pazarlama stratejileri uygulamaya karar verecektir. Bu amaçla, üniversite öğrencileri arasından rassal olarak 120 öğrenci seçilmiş ve onların 42'sinin günde 2 ve daha fazla gazlı içecek içtiği tespit edilmiştir. α = 0.05 için yönetici üniversite öğrencilerine yönelik pazarlama stratejileri uygulamalı mıdır? Karar veriniz.

ÇÖZÜM

1. Adım: Hipotezlerin ifade edilmesi

Bu örnekte günde 2 ve daha fazla gazlı içecek tüketen öğrenci oranının 0.40 ve daha fazla olduğu hipotezi test edilecektir. Araştırma hipotezi bu oranın 0.40 'dan az olduğu şeklindedir. Buna göre hipotezler

$$H_0 : \Pi = 0.40$$
$$H_1 : \Pi < 0.40$$

şeklinde ifade edilir.

2. Adım: İstatistiksel test

H_1 hipotezi $\Pi < 0.40$ şeklinde ifade edildiğinden, H_0'ın kabul ya da reddi için uygulanacak test, ana kütle oranına ilişkin tek yönlü alt kuyruk testi olmalıdır.

3. Adım: Anlamlılık düzeyinin belirlenmesi

Testin red bölgesi, oranların örnekleme dağılımının altkuyruğundadır. Anlam düzeyi $\alpha = 0.05$ seçildiği için, testin red bölgesinin oransal büyüklüğü 0.05'tir. Bu %5'lik bölgede yer alan örneklem oranları H_0 hipotezinin reddedilmesine neden olacak örneklem istatistikleridir.

4. Adım: H_0'ın red bölgesinin belirlenmesi

$z_{hes} > z_{tablo}$ için H_0 reddedilecektir.

5. Adım: Test istatistiğinin hesaplanması

120 öğrencinin içinde günde 2 ve daha fazla gazlı içecek içen öğrenci sayısı 42 kişidir. $p = \frac{r}{n} = \frac{42}{120} = 0.35$ tir. Örneklem büyük örneklemdir. Örneklem oranı p'nin dağılımı normaldir ve standart hatası,

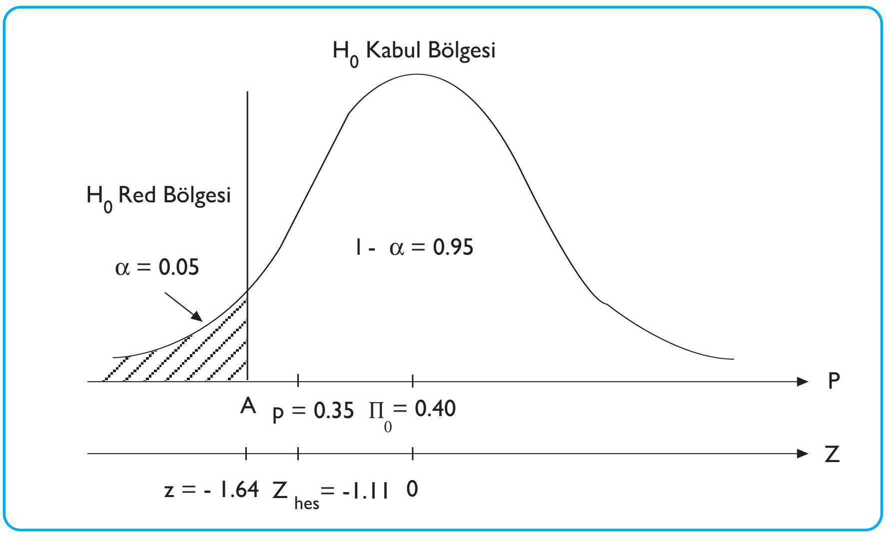
$$\sigma_p = \sqrt{\frac{\Pi_0\left(1-\Pi_0\right)}{n}} = \sqrt{\frac{(0.40)(0.60)}{120}} = \sqrt{\frac{0.24}{120}} = 0.045$$

ve test istatistiği:

$$z = \left|\frac{0.35-0.40}{0.045}\right| = 1.11$$

olarak elde edilir.

Hesaplanan z test istatistiğinin değeri, $z_{hes} = 1.11$ $\alpha = 0.05$ anlam düzeyinde standart normal dağılım tablo değerinden ($z_{tablo} = 1.64$) küçük olduğu için, H_0 hipotezi kabul edilirse H_1 hipotezi reddedilir. Burada $z_{tablo} = 1.64$ değeri standart normal dağılımda 0.45'lik alana karşı gelen değerdir.



Şekil 9.11 *Anakütle Oranı Π İçin Test Sonuçları.*

Bu kararın anlamı, üniversite öğrencileri arasında günde 2 ve daha fazla gazlı içecek tüketen öğrencilerin oranı %40 tan az değil fazladır. Yöneticinin öğrenciler için ayrı bir pazarlama stratejisi planlamasına gerek yoktur.

SIRA SİZDE

1. Örneklem hacmi $n \geq 30$ birim olduğunda, tek anakütle ortalamasına ilişkin bir testte, hangi test istatistiği kullanılır? Nedenini açıklayınız.

2. Tek anakütle ortalamasına ilişkin bir testte, ne zaman test istatistiği kullanılır?

3. Anakütle oranına ilişkin bir testte, test istatistiğinin kullanılmasının koşulları nelerdir?

Kendimizi Sınayalım

1. İstatistik dersine ait notların ortalamasının, 80'den büyük olup olmadığı, sınanmak istenmektedir. Bu sınamada kurulacak sıfır hipotezi nedir?

a. $\mu > 80$
b. $\mu = 80$
c. $\mu \neq 80$
d. $\mu < 81$
e. $\mu > 81$

2. 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

3. Bir hipotez sınamasında red bölgesinin yönünü aşağıdakilerden hangisi belirler?

a. H_0
b. H_1
c. H_0 ve H_1
d. α hatası
e. β hatası

4. Bir işyerinde, her birinde 50'şer işçiden oluşan iki grup, verimlilik bakımından karşılaştırılacaktır. Tek yönlü bir hipotez sınaması yapıldığında, 0.05 olasılık düzeyinde tablodan okunacak kritik değer aşağıdakilerden hangisidir?

a. $t = 2.750$
b. $t = 2.042$
c. $z = 1.960$
d. $z = 1.645$
e. $z = 1.276$

5. Çift yönlü bir hipotez sınamasında t değeri 2.861 ise, örnek büyüklüğü ve anlam düzeyi aşağıdakilerin hangisinde birlikte verilmiştir?

a. 19 - 0.01
b. 19 - 0.02
c. 20 - 0.01
d. 20 - 0.02
e. 21 - 0.05

6. Normal dağılıma sahip bir ana kütlenin ortalamasının 53 olup olmadığının sınanmasında, rasgele seçilen bir örneğin ortalaması 56, standart hatanın tahmini değeri 1.3 tür. Ortalamanın dönüştüğü z değerinin sağında kalan bölgenin oranlanmış alanı kaçtır?

a. 0.0107
b. 0.4893
c. 0.4898
d. 0.5000
e. 0.9893

7 ve 8. sorular aşağıdaki bilgilere göre cevaplandırılacaktır.

Bir işletmede, üretilen ampullerin 450 saat olan dayanma süresini artırmak için, yeni bir hammaddenin kullanımı düşünülmektedir. Bu hammadde kullanılarak 1 000 ürün üretilmiş ve ortalama dayanma süresi 462 saat olarak hesaplanmıştır. Hammaddenin olumlu sonuç verip vermediği %95 güvenle sınanacaktır.

7. 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

8. Bu sınamadaki alternatif hipotez nedir?

a. $\mu \neq 450$
b. $\mu > 462$
c. $\mu = 450$
d. $\mu \neq 462$
e. $\mu > 450$

9. Boyları 170 cm'den uzun olan erkeklerin ağırlık ortalamasının 72 kg olup olmadığı sınanmak istenmektedir. Bu sınamadaki sıfır hipotezi nedir?

a. $\mu \neq 170$
b. $\mu \neq 72$
c. $\mu > 170$
d. $\mu = 72$
e. $\mu = 170$

10. Ana kütle ortalamasının 200 olup olmadığının %99 güvenle test edilmesi için seçilen 15 birimlik rassal örneklemin ortalaması 160, standart sapması 60 tır. Örnek ortalamasına karşı gelen test istatistiğinin değeri kaçtır?

a. -1
b. -1.5
c. -2
d. -2.5
e. -3

Yanıt Anahtarı

1. b
2. b
3. b
4. d
5. c
6. b
7. c
8. e
9. d
10. d

Yararlanılan Kaynaklar

CANKÜYER, Ersoy, AŞAN, Zerrin: **Parametrik Olmayan İstatistiksel Teknikler**, Anadolu Üniversitesi Yayınları, No:1266, Eskişehir, 2001.

ÇÖMLEKÇİ, Necla: **Bilimsel Araştırma Yöntemi ve İstatistiksel Anlamlılık Sınamaları**, Bilim Teknik Yayınevi, İstanbul, 2001.

FINK, Arlene: **How to Sampling in Surveys**, Sage Publications, London, 1995.

GÜRSAKAL, Necmi: **Bilgisayar Uygulamalı İstatistik I**, Marmara Kitabevi, Bursa, 1997.

HINKLE, Dennir E; WIERSMA, Williams; JURS, Stephen G: **Applied Statics For The Bhavioral Sciences**, Boston, 1998.

MALTHORA, Naresh K.: **Marketing Research An Applied Orientation,** 2nd Edition, Prentice-Hall International Inc, New Jersey, 1996.

NETER, J, WASSERMAN, W, WHITMORE, G.A.: **Applied Statistics**, Simon and Schuster, Inc, Boston, 1993.

ÖZMEN, A., ÖZDAMAR, K., ODABAŞI, Y., Hoşcan, Yaşar., BİR, A. Atıf, KIRCAALİİFTAR, G., UZUNER, Yıldız.: **Sosyal Bilimlerde Araştırma Yöntemleri**, TC. Anadolu Üniversitesi Yayınları, No:1081, Açıköğretim Fakültesi Yayınları, No: 601, Eskişehir, 1999.

PÜSKÜLCÜ, Halis, İKİZ, Fikret: **İstatistiğe Giriş,** (2. Baskı), E.Ü. Mühendislik Fakültesi Yayın No: 601, Ege Üniversitesi Basımevi, İzmir, 1986.

SERPER, Özer; Aytaç Mustafa: **Örnekleme**, Ezgi Kitabevi, Bursa, 2000.

SERPER, Özer: **Uygulamalı İstatistik II**, Filiz Kitabevi, İstanbul, 1986.

TRYFOS, Peter: **Sampling Methods for Applied Research**, John Wiley & Sons Inc., New York, 1996.

TULL, Donald S., HAWKINGS, Del I.: **Marketing Research Measurement and Method**, 6th Edition, MacMillan Publishing Company, New York, 1993.

10

Ki-Kare Testi

Çalışma Biçimine İlişkin Olarak

- *Üniteyi anlayabilmeniz için, örnekleme ve hipotez testleri üniteleri, yeterince özümsenmiş olmalı,*
- *Örnekler dikkatlice incelenmeli, sorunlarla karşılaşılırsa ilgili ünitelere dönülmelidir.*

Amaçlar

- *Sayısal olmayan değişkenler arasındaki ilişkinin varlığını test edebileceksiniz.*
- *Farklı örneklemlerin aynı ana kütleden seçilip seçilmediğini test edebileceksiniz.*
- *n hacimli bir örneklemin, ilgili ana kütleyi, iyi temsil edip edemediğini belirleyebileceksiniz.*
- *Sayısal olmayan iki değişken arasındaki ilişkinin derecesini belirleyebileceksiniz.*

İçerik Haritası

- *GİRİŞ*
- *Kİ-KARE BAĞIMSIZLIK TESTİ*
- *Kİ-KARE HOMOJENLİK TESTİ*
- *Kİ-KARE UYGUNLUK (İYİ UYUM) TESTİ*
- *KONTENJANS KATSAYISI*

GİRİŞ

Daha önce de belirtildiği gibi, istatistikte değişkenler, sayısal (nicel) değişkenler ve sayısal olmayan (nitel) değişkenler olmak üzere iki grupta sınıflandırılmaktadır. Günümüzde yapılan bir çok 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", "bekar", "boşanmış" ve "dul" şeklinde gösterilirken meslek grupları da "serbest meslek", "devlet memurluğu", "işçi", vb. şeklinde gruplandırılabilir.

İşte sayısal olmayan değişkenler arasında herhangi bir ilişkinin varolmadığını ileri sürerek (H_0 hipotezi), bu hipotezin red edilip edilemeyeceğinin incelenmesinde uygulanan test 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 ana kütleden seçilip seçilmedikleri konusunda karar verilirken de ki-kare testinden yararlanılır.

1900 yıllarında Karl Pearson tarafından bulunan ve ismi de onun tarafından verilen bu istatististiksel 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.

Kİ-KARE BAĞIMSIZLIK TESTİ

AMAÇ 1

Sayısal olmayan değişkenler arasındaki ilişkinin varlığını test edebileceksiniz.

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.

İki ya da daha çok sınıflı nitel değişkenler arasındaki bağımsızlık, ki-kare bağımsızlık testiyle araştırılır.

Bu ve benzeri problemlerin çözümlenmesinde uygun bir istatistiksel teknik de, aşağıda yeterli ayrıntıyla ele alınacak olan ki-kare testidir.

İki ya da daha fazla sınıflı iki nitel değişken arasında bağımsızlık olup olmadığını incelemek için, ki-kare bağımsızlık testine 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 10.1'de gösterilmiştir.

Tablo 10.1 Kontenjan tablosunun yapısı.

I. Değişkenin Şıkları	2. 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	$n_{.1}$	$n_{.2}$	$n_{.3}$	....	$n_{.j}$	....	$n_{.c}$	$n_{..} = n$

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

Bayan Televizyon izleyicilerinin öğrenim düzeyleri ve TV programlarından tercih ettikleri türler sorgulanarak, bu iki değişken arasında bir bağıntı bulunup bulunmadığını, başka bir anlatımla, iki değişkenin birbirinden bağımsız olup olmadığını, ortaya koymaya çalışılsın. Bu amaçla, 200 kişiyi kapsayan bir örneklem üzerinde yapılan gözlem sonuçları aşağıdaki tablo ile verilmiştir:

Tablo 10.2 Gözlenen frekanslar.

Tercih edilen TV program türü	Öğrenim Düzeyi İLK	ORTA	YÜKSEK	TOPLAM
DİZİ FİLM	50	20	10	80
EĞLENCE	20	30	10	60
MAGAZİN	20	10	30	60
TOPLAM	90	60	50	200

Tercih edilen TV program türüne ilişkin öğrenim düzeyinin etkili olup olmadığını α=0.01 anlamlılık düzeyinde araştırınız.

ÇÖZÜM

Tabloda yer alan sayılar "gözlenen frekanslardır". Tercih edilen TV programı türü üzerinde öğrenim 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): Bayan TV izleyicilerinin öğrenim düzeyiyle TV programı, birbirinden bağımsız değişkenlerdir. Bu iki değişken arasında bir ilişki (bağıntı) yoktur.

Karşıt Hipotez (H_1): Öğrenim düzeyiyle TV programı arasında bir ilişki (bağıntı) vardır.

2. Adım: İstatistiksel test

İki sayısal olmayan değişken arasındaki ilişkinin varlığını araştıran bir test olan χ^2 (ki-kare) bağımsızlık testi olmalıdır.

3. Adım : Anlamlılık düzeyinin belirlenmesi

χ^2 = 0.01 olarak belirlenmiştir.

4. Adım: H_0'ın red bölgesinin belirlenmesi

Bunun için hesaplanan test istatistiği, belli bir anlamlılık düzeyine ve $\nu = (r–1)(c–1)$ serbestlik derecesine göre "χ^2 değerleri tablosu"ndan bulunan "kritik değer" ile karşılaştırılır. Örneğimiz için serbestlik derecesi $\nu = (3–1)\ (3–1) = 4$ olup $\alpha = 0.01$ düzeyinde χ^2 tablosundan bulunan kritik değer χ^2_k=13'tür. Eğer hesaplanan χ^2 istatistiğinin değeri tablodan bulunan χ^2_k kritik değerden büyük çıkarsa H_0 red edilecektir.

Kontenjans tablolarında serbestlik derecesi, satır ve sütun sayılarından birer çıkartılarak, bunların çarpılması suretiyle elde edilir.

5. Adım: χ^2 test istatistiğinin hesaplanması

Test istatistiği

$$\chi^2 = \sum \frac{(G - B)^2}{B}$$

formülüyle elde edilebilir. Formülde,

G= Gözlenen frekansları
B= Beklenen frekansları

ifade etmektedir. Test istatistiğinin hesaplanabilmesi için öncelikle beklenen (kuramsal) frekansların 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:

Kuramsal (beklenen) frekanslar, ilgili gözenin yer aldığı satır toplamıyla sütun toplamı çarpılarak genel toplama bölünmek suretiyle elde edilir.

B_{11} (birinci satır ve birinci sütunda yer alacak frekans)
B_{11} = (birinci satır toplamı x birinci sütun toplamı) / (genel toplam)
= (80 x 90) / (200) = 36
B_{12} = (birinci satır toplamı x ikinci sütun toplamı) / (genel toplam)
= (80 x 60) / (200) = 36

Aynı yöntemle hesaplanan beklenen frekansları ve gözlenen frekansları kontenjans tablosunda gösterelim (G=Gözlenen frekanslar, B=Beklenen freakanslar).

Tercih edilen Tv program türü	Öğrenim Düzeyi						TOPLAM
	İLK		ORTA		YÜKSEK		
	G	B	G	B	G	B	
DİZİ FİLM	50	36	20	24	10	20	80
EĞLENCE	20	27	30	18	10	15	60
MAGAZİN	20	27	10	18	30	15	60
TOPLAM	90		60		50		200

Tablo 10.3 *Kontenjan tablosu (Gözlenen ve beklenen frekanslar).*

Test istatistiği:

$$\chi^2 = (50–36)^2/(36) + (20–24)^2/(24) + (10–20)^2/(20) + (20–27)^2/(27) + (30\text{-}18)^2/(18) + (10–15)^2/(15) + (20–27)^2/(27) + (10–18)^2/(18) + (30–15)^2/(15) = 42.93$$

6. Adım: İstatistiksel Karar

İstatistiksel karar verilirken, red bölgesinin tanımı gereği, $\chi^2 > \chi^2_k$ olduğunda sıfır hipotezi red edilir, $\chi^2 \leq \chi^2_k$ olduğundaysa sıfır hipotezi reddedilemez. Sıfır hipotezinin red edilmesi, değişkenlerin birbirinden bağımsız olmadığı (diğer bir ifadeyle, değişkenler arasında ilişki bulunduğu) anlamını taşır. Buna göre örneğimizde,

$\chi^2 = 42.93$ $\quad\chi^2_k = 13.28$ $\quad$ ve $\quad\chi^2 > \chi^2_k$

olduğundan H_0 hipotezi red edilecektir. Başka bir anlatımla, bayan TV izleyicilerinin öğrenim düzeyiyle izledikleri program türleri arasında ilişki vardır.

SIRA SİZDE

1. Ki-kare testi hangi durumlarda yapılır? Açıklayınız.

2. Ki-kare bağımsızlık testinde hipotezler nasıl ifade edilir? Açıklayınız.

3. Ki-kare bağımsızlık testinde serbestlik derecesi nasıl hesaplanır?

Kİ-KARE HOMOJENLİK TESTİ

AMAÇ 2

Farklı örneklemlerin aynı ana kütleden seçilip seçilmediğini test edebileceksiniz.

Farklı örneklemlerin aynı ana kütleden 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 uygulaması, ki-kare bağımsızlık testinde olduğu gibidir. Yine nitel değişkenlerle ve aynı örneklem istatistiğiyle çalışılır. Ancak, dikkat edilmelidir ki, bağımsızlık testinde ele alınan değişkenler arasında bir ilişkinin varlığı araştırılırken, homojenlik testinde bağımsız örneklemlerin aynı ana kütleden seçilip seçilmediği araştırılmaktadır.

ÖRNEK 2

Bölgesel satış yapan bir üretim işletmesi, 2 yeni ürün geliştirerek piyasaya sürmüştür. Tüketicilerin bu ürünlerle ilgili görüşlerini (beğendikleri, beğenmedikleri ya da ilgisiz kaldıkları) belirlemek amacıyla, birinci ve ikinci ürünlerle ilgili olarak iki rassal örneklem oluşturulmuştur. İlk ürünle ilgili birinci örneklemde 100 tüketiciyle, ikinci ürünle ilgili ikinci örneklemde de 150 tüketiciyle görüşülmüştür. Veriler aşağıdaki tabloda belirtilmiştir. Seçilen örneklemlerin, aynı anakütleye ait olup olmadığını, %5 anlamlılık düzeyinde test ediniz.

Tüketici görüşleri; fiyat, kalite, kolay ulaşabilme vb. gibi objektif ölçütlerle ve piyasadaki benzer ürünlerle mukayese sonucu oluşmuştur.

GÖZLENEN FREKANSLAR				
Ürünler	Tüketici Görüşleri			Toplam
I. Ürün	60	30	10	100
II. Ürün	80	50	20	150
Toplam	140	80	30	250

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 ret bölgesinin belirlenmesi

Hesaplanan test istatistiği $\nu = (2\text{-}1)(3\text{-}1) = 2$ serbestlik derecesi ve $\alpha = 0.05$ anlamlılık düzeyi için ki-kare tablosundan bulunan kritik değer, $\chi^2_k = 5.99$'dur. Eğer hesaplanan χ^2 istatistiğinin değeri, $\chi^2_k = 5.99$'dan büyük çıkarsa, H_0 ret edilecektir.

5. Adım: χ^2 test istatistiğinin hesaplanması

Hatırlanacağı gibi, test istatistiğinin hesaplanabilmesi için, öncelikle, beklenen frekansların hesaplanması gerekir. Homojenlik testinde de her hangi 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:

Ürünler	Tüketici Görüşleri						Toplam
	Beğenen		Beğenmeyen		İlgisiz		
	G	B	G	B	G	B	
I. Ürün	60	56	30	32	10	21	100
II. Ürün	80	84	50	48	20	18	150
Toplam	140		80		30		250

Test istatistiği:

$$\chi^2 = (60–56)^2/(56) + (30–32)^2/(32) + (10–12)^2/(12) + (80–84)^2/(84) + (50\text{-}48)^2/(48) + (20–18)^2/(18) = 1.04$$

olarak elde edilir.

6. Adım: İstatistiksel karar

Hatırlanacağı gibi, $\chi^2 < \chi^2_k$ ise H_0 hipotezi kabul edilir. Örneğimizde $\chi^2 = 1.04$ ve $\chi^2_k = 5.99$ olduğundan, H_0 kabul edilecektir. Başka bir anlatımla, ilgili örneklemler aynı anakütleden seçilmiştir.

SIRA SİZDE

1. Bağımsızlık ve homojenlik testleri hangi açılardan birbirinden farklıdır? Açıklayınız.

2. Bağımsızlık ve homojenlik testlerinde beklenen frekanslar nasıl hesaplanır? Açıklayınız.

3. Serbestlik derecesi ν = 10 ve α = 0.01 için χ^2_k nedir?

Kİ-KARE UYGUNLUK (İYİ UYUM) TESTİ

AMAÇ 3

n hacimli bir örneklemin, ilgili ana kütleyi, iyi temsil edip edemediğini belirleyebileceksiniz.

n birimlik örneklemin çekildiği, anakütleyi, iyi temsil edip edemeyeceği, ki-kare uygunluk testi ile araştırılır.

Ki-kare uygunluk testinin esasını, n hacimli (birimlik) bir örneklemin anakütleyi iyi temsil edip edemeyeceğini araştırmak oluşturur. Bu testte, yine χ^2 değişkeninin doğası gereği, gözlenen ve beklenen frekanslardan yararlanılır. Testin nasıl yapılacağı, özellikle beklenen frekansların nasıl hesaplanacağı, aşağıdaki örnek yardımıyla açıklanmaya çalışılmıştır.

ÖRNEK 3

Belirli bir bölgede, Z marka margarin kullanan aile oranı, 3/8 olarak öngörülmektedir. Her anketör, rassal olarak seçilen 5 aileyle görüşmek üzere, 200 anketör kullanılarak ilgili bölgede bir anket düzenlenmiş ve anket sonuçları aşağıdaki frekans dağılımıyla verilmiştir:

Görüşülen 5 aile içinde Z marka margarin kullanan aile sayısı X	Anketör Sayısı f
0	18
1	57
2	69
3	42
4	11
5	3
	200

Elde edilen bu sonuçlar için,

$$P(x) = \begin{cases} C_5^x \left(\frac{3}{8}\right)^x \left(\frac{5}{8}\right)^{5-x} & , x = 0, 1, 2, 3, 4, 5 \\ 0 & , \text{diğer durumlarda} \end{cases}$$

formunda bir binom dağılımı öngörülmektedir. Öngörülen dağılımın, ele alınan problem için, uygun bir model olup olmadığını α= 0.05 anlamlılık düzeyi için test ediniz.

ÇÖZÜM

1. Adım: Hipotezlerin oluşturulması

H_0: X rassal değişkeni, n = 5 ve $p = \frac{3}{8}$ parametre değerleriyle binom dağılmıştır. (Binom dağılımında q+p=1 olduğunu hatırlayınız.)

H_1: X rassal değişkeni, n = 5 ve $p = \frac{3}{8}$ parametre değerleriyle binom dağılmamıştır.

2. Adım: İstatistiksel test

χ^2 uygunluk (iyi uyum) testi

3. Adım: Anlamlılık düzeyi

$\alpha = 0.05$

4. Adım: H_0'ın red bölgesinin belirlenmesi

Hatırlanacağı gibi, red bölgesi, hesaplanan χ^2 değerinin öngörülen anlamlılık düzeyi ve belirlenen serbestlik derecesine göre, χ^2 tablosundan bulunan kritik değeriyle karşılaştırılarak belirlenir. Uygunluk testinde serbestlik derecesi, sınıf sayısından ilgili dağılımın parametre sayısının çıkartılmasıyla elde edilir. Ancak, uygunluk testlerinde örneklem hacmi değişken olduğundan, örneklem hacmi de bir parametre gibi değerlendirilir. Gerçekte binom dağılımının n ve p olmak üzere iki parametresi vardır. Artı toplam frekans da bir parametre olarak görüleceğinden, serbestlik derecesi, sınıf sayısı eksi 3 olarak belirlenecektir. Buna göre, $\nu = 6–3 = 3$ olur. 3 serbestlik derecesi ve $\alpha = 0.05$ anlamlılık düzeyi için, kritik χ^2_k değeri, χ^2 tablosundan, 7.81 olarak bulunur. Red bölgesi, $\chi^2 > 7.81$ olarak belirlenir.

Ki-kare uygunluk testinde serbestlik derecesi, ilgili dağılımın parametre sayısına bir eklenip sınıf sayısından çıkartılmak suretiyle hesaplanır.

5. Adım: Ki-kare istatistiğinin hesaplanması

Beklenen frekansların, ilgili sınıfa ilişkin olasılık toplam frekansın çarpımı olduğu anımsanırsa, ilgili olasılıklar,

$$P(x) = \begin{cases} C_5^x \left(\frac{3}{8}\right)^x \left(\frac{5}{8}\right)^{5-x} & , x = 0, 1, 2, 3, 4, 5 \\ 0 & , \text{diğer durumlarda} \end{cases}$$

olasılık fonksiyonu yardımıyla, x'e sırasıyla 0, 1, 2, 3, 4 ve 5 değerleri verilerek, aşağıdaki gibi hesaplanır:

X	f	Olasılıklar P(x)
0	18	0.0954
1	57	0.2861
2	69	0.3433
3	42	0.2060
4	11	0.0618
5	3	0.0074
	$\sum f = 200$	1.0000

Kuramsal frekansların bulunabilmesi için bu olasılıklar, frekansların toplamıyla çarpılır.

x	f	Olasılıklar P(x)	$P(x).\sum f$
0	18	0.0954	19.08
1	57	0.2861	57.22
2	69	0.3433	68.66
3	42	0.2060	41.20
4	11	0.0618	12.36
5	3	0.0074	1.48
	$\sum f = 200$	1.0000	200.00

Ki-kare uygunluk testinde herhangi bir sınıfa ilişkin kuramsal frekans, ilgili sınıfın olasılığıyla toplam frekans çarpılarak elde edilir.

$$\chi^2 = \sum \frac{(G-B)^2}{B}$$

eşitliği uyarınca,

$$\chi^2 = \frac{(18-19.08)^2}{19.08} + \frac{(57-57.22)^2}{57.22} + \frac{(69-68.66)^2}{68.66} + \frac{(42-41.20)^2}{41.20} + \frac{(11-12.36)^2}{12.36} + \frac{(3-1.48)^2}{1.48}$$

$$\chi^2 = 0.0611 + 0.0008 + 0.0017 + 0.0015 + 0.1496 + 1.5611 = 1.7758$$

olarak elde edilir.

6. Adım: İstatistiksel Karar

$\chi^2 = 1.7758$
$\chi^2_k = 7.81$

olarak bulunmuştur. Bu sonuçlara göre,

$\chi^2 < \chi^2_k$

olduğundan H_0 kabul edilecektir. Başka bir anlatımla, eldeki frekans dağılımı, $n = 5$ ve $p = \frac{3}{8}$ için binom dağılmış bir anakütleden çekilmiş bir örneklemdir.

SIRA SİZDE

1. Ki-kare uygunluk testi, hangi amaçla yapılır? Açıklayınız.

2. Ki-kare uygunluk testind,e serbestlik derecesi nasıl hesaplanır?

3. Ki-kare uygunluk testinde, H_0'ın ret ve kabul bölgesi nasıl belirlenir? Açıklayınız.

KONTENJANS KATSAYISI

AMAÇ 4

Sayısal olmayan iki değişken arasındaki ilişkinin derecesini belirleyebileceksiniz.

Sayısal olmayan iki değişken arasındaki ilişkinin derecesi, kontenjans katsayısıyla belirlenir.

Ki-kare bağımsızlık testiyle, iki değişken arasındaki ilişkinin varlığıyla ilgili karar verebiliyordu. Oysa ki bazı hallerde, iki değişken arasındaki ilişkinin kuvveti hakkında da bilgi sahibi olmak istenebilir. İşte kontenjans katsayısı r x c kontenjans tablolarından ($r > 2$ ve $c > 2$) hesaplanan χ^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,

$$c = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

şeklindedir.

ÖRNEK 4

Örnek 1'de bayan televizyon izleyicilerinin öğrenim düzeyleri ve TV programlarından tercih ettikleri türler sorgulanmış ve bu iki değişken arasında bağıntı bulunup bulunmadığı (ilişki olup olmadığı) test edilmişti. Şimdi kontenjans katsayısıyla ilişkinin derecesini araştıralım.

ÇÖZÜM

$\chi^2 = 42.93$ 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.

SIRA SİZDE

1. Kontenjans katsayısı, hangi amaçla kullanılır?

2. Kontenjans katsayısının değeri, hangi durumda sıfır olur?

3. Gözlem sayısı 150 olan bir araştırmada χ^2 test istatistiği 52.17 olduğuna göre, kontenjans katsayısının değeri ne olacaktır? Hesaplayınız.

Kendimizi Sınayalım

(1 - 6. sorular aşağıdaki bilgilere göre cevaplandırılacaktır.)

Bir araştırma sonucunda elde edilen gözlem sonuçları aşağıdaki tabloda gösterilmiştir.

	Sigara İçme Alışkanlığı		
Cinsiyet	İçenler	İçmeyenler	Toplam
Bayan	35	20	55
Erkek	25	20	45
Toplam	60	40	100

1. Yukarıdaki tabloya göre, sıfır hipotezinin doğru ifadesi aşağıdakilerden hangisidir?

a. Sigara içme alışkanlığıyla cinsiyet birbirine bağımlıdır.
b. Sigara içme alışkanlığıyla cinsiyet birbirinden bağımsızdır.
c. Bayanlarda sigara içme alışkanlığı daha yaygındır.
d. Erkeklerde sigara içme alışkanlığı daha yaygındır.
e. Bayanlarla erkekler aynı miktarda sigara içerler.

2. Yukarıda verilen tabloya ilişkin serbestlik derecesi kaçtır?

a. 1
b. 2
c. 3
d. 4
e. 5

3. Yukarıda verilen tablonun, ilk satırındaki ikinci gözenin ki-kare değerine katkısı kaçtır?

a. 1,5
b. 1,18
c. 0,40
d. 0,38
e. 0,18

4. Yukarıda verilen tabloya göre, sigara içme alışkanlığıyla cinsiyet arasında bir bağıntı olup olmadığını, 0,05 anlam düzeyinde test ederken, H_0 hipotezinin red bölgesi aşağıdakilerden hangisidir?

a. $\chi^2 > 9{,}21$
b. $\chi^2 > 6{,}64$
c. $\chi^2 > 5{,}99$
d. $\chi^2 > 3{,}84$
e. $\chi^2 > 2{,}71$

5. Yukarıda verilen tabloya göre, χ^2 test istatistiğinin değeri kaçtır?

a. 0.42
b. 0.67
c. 0.95
d. 1.15
e. 2.5

6. Yukarıda verilen tabloya ilişkin kontenjans katsayısının değeri kaçtır?

a. 0
b. 0.006
c. 0.017
d. 0.07
e. 0.5

7. Gözlem sayısı 100 olan bir araştırmada kontenjans katsayısı 0,40 olarak bulunmuştur. Buna göre, χ^2 istatistiğinin değeri kaçtır?

a. 10.4301
b. 12.3205
c. 19.0476
d. 20.1313
e. 22.1864

8. Bir araştırmada χ^2 istatistiğinin değeri 14.06 ve kontenjans katsayısı (c) 0.6 olarak hesaplanmıştır. Bu bilgilere göre, gözlem sayısı (n) kaçtır?

a. 15
b. 25
c. 40
d. 45
e. 50

9. Bir X rassal değişkenine ilişkin frekans dağılımı ve X'lere karşı gelen olasılıklar [P(x) değerleri] aşağıdaki tabloda verilmiştir.

X	f	P(x)
0	6	0.1667
1	15	0.5000
2	8	0.3000
3	1	0.0333
	30	**1.0000**

Yukarıdaki tabloya göre, X = 2 için kuramsal frekans kaçtır?

a. 4.87
b. 8.14
c. 9.00
d. 16.13
e. 32.05

10. X rassal değişkeninin frekans dağılımı aşağıdaki tabloda verilmiştir.

Sınıflar	f
0 – 5	2
5 – 10	6
10 – 15	11
15 – 20	15
20 – 25	14
25 – 30	8
30 – 35	3
35 – 40	1
	60

X rassal değişkeninin normal dağıldığı biliniyorken yukarıdaki örneklemin, dağılımı iyi temsil edip etmediğini araştırmak için ki-kare uygunluk testi uygulanacaktır. Yukarıdaki tabloya ve bu bilgilere göre serbestlik derecesi kaçtır?

a. 5
b. 6
c. 7
d. 8
e. 9

Yanıt Anahtarı

1. b
2. a
3. e
4. d
5. b
6. d
7. c
8. b
9. c
10. a

Yararlanılan Kaynaklar

CANKÜYER, Ersoy, AŞAN Zerrin: **Parametrik Olmayan İstatistiksel Teknikler**, Anadolu Üniversitesi Yayınları, No:1266, Eskişehir, 2001

NEWBOLD, Paul: **İşletme ve İktisat İçin İstatistik**, Çeviren: Ümit Şenesen, 4. Basım, Literatür Yayıncılık, 2000.

SERPER, Özer: **Uygulamalı İstatistik II**, 4. Baskı, Ezgi Kitabevi, Bursa, 2000

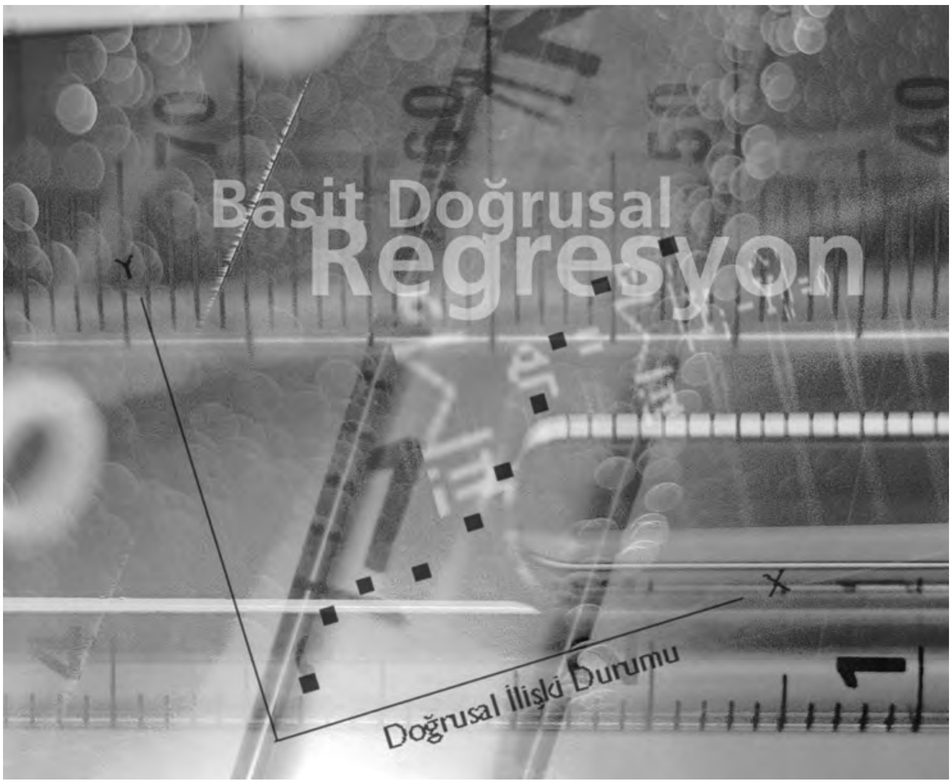
YÜZER, Ali Fuat: **Olasılık ve İstatistik**, Anadolu Üniversitesi Yayınları, No: 911, Eskişehir, 1996.

KARL PEARSON (1857 - 1936)

Temel ilgi alanı genetiktir. 1892'de "The Grammar of Science" adlı kitabı yayınlandı.İzleyen yıllarda kalıtım ve evrim süreçlerine ilişkin çalışmaları sırasında istatistikle ilgilendi. Regresyon ve korelasyon konularındaki önemli katkılarının yanı sıra, kuramda kendi adıyla anılan ve gözlem değerlerinin olasılık dağılımlarına ilişkin Pearson eğri sistemini ve 1912 yılında da Ki-kare testini geliştirdi.

11

Basit Doğrusal Regresyon



Çalışma Biçimine İlişkin Olarak:

- *Kavramlar ve bu kavramlar arasındaki ilişkiler dikkatle incelenmeli,*
- *Örnekler ve örnek çözümleri dikkatle incelenmeli, sorunlarla karşılaşılırsa ilgili ünitelere geri dönülmelidir.*

Amaçlar:

- *Serpilme diyagramı yardımıyla, değişkenler arasındaki ilişkinin, ne tür bir fonksiyonla ifade edilebileceğini araştırabileceksiniz.*
- *Basit doğrusal regresyon modelinde yer alan katsayıları, en küçük kareler tekniğine göre hesaplayabileceksiniz.*
- *Basit doğrusal regresyon modeliyle elde edilen kestirimlerin standart hatasını hesaplayabileceksiniz.*
- *Basit doğrusal regresyonda, parametrelerin nokta kestiriminden sonra genelleme yapabilmek için, parametrelere ilişkin aralık kestirimi yapabileceksiniz.*
- *Basit doğrusal regresyon denkleminde elde edilen parametre kestirimlerinin, istatistiksel olarak anlamlılığını test edebileceksiniz.*

İçerik Haritası

GİRİŞ

Regresyon (bağlanım); sözlük anlamıyla, bir şeyi başka bir şeye bağlama işi ve biçimidir. Bilimsel olarak regresyon terimi bir değişkenle başka bir (ya da birden çok) değişken arasında ilişki kurma işini ve ilişkinin biçimini anlatır.

İstatistiksel anlamda, iki değişken arasındaki ilişki, bunların değerlerinin karşılıklı değişmeleri arasında bir bağlılık şeklinde anlaşılır. (X) değişkeninin değerleri değişirken, buna bağlı olarak (Y) değişkeninin değerleri de değişiyorsa, bu iki değişken arasında bir ilişki olduğu söylenebilir. Örneğin, pancar üretimi arttığında fiyatı düşüyorsa ya da azaldığında fiyatı yükseliyorsa, insanların boy uzunluğuyla birlikte ağırlığı da artıyorsa bunlar, değişkenler arasında ilişki olduğunu gösterir. Aslında değişkenler arasındaki bu ilişki neden-sonuç ilişkisidir. İşte değişkenler arasındaki neden sonuç ilişkisinin matematiksel bir fonksiyonla ifade edilmesi regresyon analizinin konusunu oluşturmaktadır. Regresyon, bir bağımlı (açıklanan) değişken, diğeri de bağımsız (açıklayıcı) değişken olarak en az iki değişken arasındaki ortalama ilişkinin matematik bir fonksiyon şeklinde ifade edilmesidir. Bu fonksiyona regresyon denklemi adı verilmektedir. Bu ünitede ilk olarak, regresyon çözümlemesinde kullanılan serpilme diyagramı hatırlatılacak, daha sonraysa basit doğrusal regresyon modeli ele alınacak, son olarak da basit doğrusal regresyon modeline ilişkin katsayıların, en küçük kareler yöntemiyle, kestirimlerinin elde edilmesi ele alınacaktır.

Regresyon, değişkenler arasındaki ortalama ilişkinin matematiksel bir fonksiyonla ifade edilmesidir.

SERPİLME DİYAGRAMI

AMAÇ 1

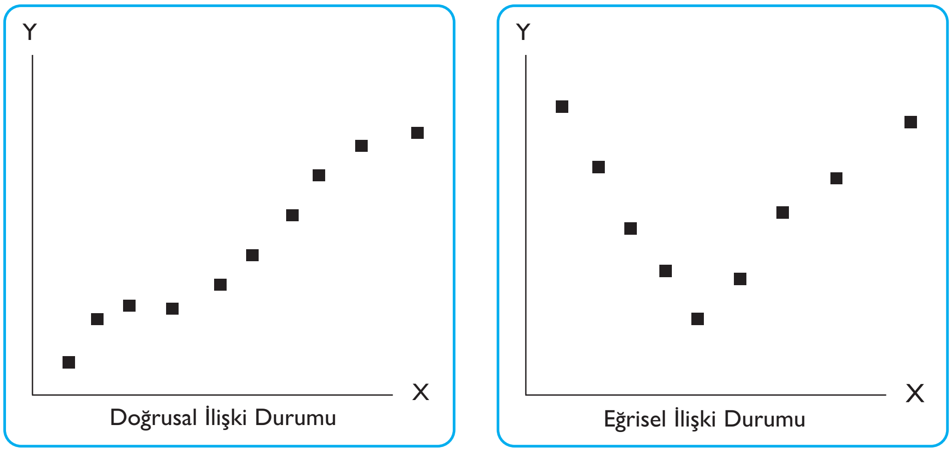
Serpilme diyagramı yardımıyla, değişkenler arasındaki ilişkinin, ne tür bir fonksiyonla ifade edilebileceğini araştırabileceksiniz.

Bir marketler zinciri E şehrinin farklı semtlerinde şubeler açmayı hedeflemektedir. Ancak, planlama bölümü, açılacak şube sayısının belirlenebilmesinde adı geçen şehirde hane başına perakende satışların, hane başına harcanabilir gelirle olan ilişkisine ihtiyaç duymaktadır.

İhtiyaç duyulan ilişki, ilgili bölgeden derlenecek veriler ışığında, uygun istatistiksel teknikler uygulanarak, bir matematiksel model halinde ifade edilebilir.

Bu ünitede, iki değişken arasındaki ilişkinin matematiksel bir modelle ifade edilme süreci, kitabın amaçları uyarınca sadece basit doğrusal regresyon düzeyinde ele alınmış ve konuya ilişkin kavramlar örneklerle pekiştirilmeye çalışılmıştır.

İki değişken arasındaki ilişkinin ne tür bir fonksiyon tipine uyduğu, yaklaşık olarak serpilme diyagramı çizerek belirlenebilir. Değişkenlerin arasındaki ilişkiyi göstermenin en iyi yolu, ilişkinin derecesini sayısal olarak belirlemektir. İlişkiyi göstermenin diğer bir yolu da grafik yöntemidir. X ve Y gözlem ikilileri bir grafik üzerinde birer nokta halinde gösterilsin. İşaretlenen bu noktaların oluşturduğu şekil anımsanacağı gibi "serpilme diyagramı" olarak isimlendirilir.



Şekil 9.1 *Serpilme Diyagramı.*

Serpilme diyagramı, değişkenler arasındaki ilişki tipinin belirlenmesine yardımcı olur.

Serpilme diyagramında noktaların durumu ve genel seyri, iki değişken arasında ilişki olup olmadığını ve varsa ilişkinin ne tür bir fonksiyon tipine uyduğunun belirlenmesinde yardımcı olur.

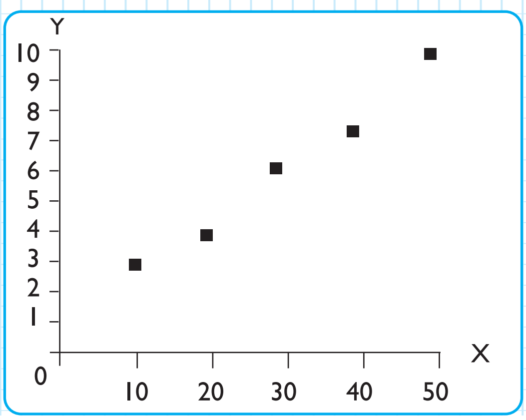
Serpilme diyagramı yalnız ilişkinin olup olmadığını ve fonksiyonel şeklini göstermekle kalmaz, ilişkinin derecesi hakkında da bilgi verir. Bunun için, noktaların en dışta kalanları birleştirilerek, bir şekil elde edilir. Söz konusu şeklin durumuna göre ilişkinin derecesi hakkında tahminde bulunulur. Eğer şekil, oldukça dar bir elipse benziyorsa, ilişki kuvvetlidir. Elips genişledikçe ilişki zayıflar.

ÖRNEK 1

Eskişehir ilinde satış yapan bir mağaza ürünlerini, yerel radyodaki reklamlarla tanıtmaktadır. Firmanın 6 hafta süresince belirli bir ürün için harcadığı reklam tutarı ve satılan ürün sayısı aşağıdaki tabloyla verilmiştir. Serpilme diyagramını çizelim.

Reklam harcaması (X) (Milyon TL)	Satışlar (Y) (Adet)
10	3
20	4
30	6
40	7
50	10

ÇÖZÜM



Şekil 9.2 *Reklam Harcamalarına İlişkin Serpilme Diyagramı.*

Kartezyen koordinat sisteminde X ve Y 'ye ait verileri işaretlediğimizde, iki değişken arasındaki ilişkinin doğrusal olduğunu görebiliriz.

SIRA SİZDE

1. **Serpilme diyagramıyla ne belirlenir?**

2. **Serpilme diyagramı, iki değişken arasında ilişki olup olmadığı ve fonksiyonel şekli dışında, başka ne hakkında bilgi verir?**

3. **Bir A ülkesinde, 1995-2000 yılları arasındaki erkek nüfus artış hızı, yıllara göre aşağıdaki tabloda verilmiştir. Serpilme diyagramını çiziniz.**

YILLAR	Erkek Nüfus Artış Hızı (%)
1995	18.50
1996	19.00
1997	20.00
1998	20.40
1999	21.60
2000	22.90

BASİT DOĞRUSAL REGRESYON

AMAÇ 2

Basit doğrusal regresyon modelinde yer alan katsayıları, en küçük kareler tekniğine göre hesaplayabileceksiniz.

Regresyon analizinde bağımsız (açıklayıcı) değişken sayısı bir olduğunda basit regresyon modelinden, iki ya da daha fazla olduğundaysa çoklu regresyon modelinden söz edilir. Örneğin enflasyon oranıyla para arzı arasındaki ya da hem para arzı hem de kamu harcamaları arasındaki ilişkinin araştırılmasında olduğu gibi. Regresyon analizinde değişkenler arasındaki ilişkinin doğrusal olup olmadığı da önemlidir. Dolayısıyla değişkenler arasındaki ilişki doğrusal olduğunda doğrusal regresyon modeli, doğrusal olmadığındaysa doğrusal olmayan regresyon modeli söz konusu olur. Kitabın amaçları doğrultusunda burada, sadece basit doğrusal regresyon konusuna yer verilecektir.

Basit Doğrusal Regresyon Modeli

Basit doğrusal regresyon modeli

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad i = 1, 2, ..., N$$

şeklinde stokastik(olasılıklı) bir modeldir. β_0 ve β_1 bilinmeyen regresyon katsayılarıdır. ε_i , i 'inci gözleme karşılık gelen hata terimidir. X (bağımsız) ve Y (bağımlı) değişkenlerinin anakütlelerini oluşturan ve bu değişkenler için akla gelebilecek bütün değerlere sahip olunması uygulamada imkansız olduğundan, söz konusu değişkenler için örneklemeye başvurulur. Böylece β_0 ve β_1 parametrelerinin kestirimi olan b_0 ve b_1 bulunabilir ve kestirimi elde edilen ilişki,

$$Y_i = b_0 + b_1 X_i + e_i$$

şeklinde yazılır.

β_0 , doğrusal modelin sabit terimidir ve X = 0 olduğunda regresyon doğrusunun dikey eksen Y 'yi kestiği noktayı göstermektedir. β_1 ise doğrusal modelin eğimini vermektedir ve regresyon analizinde, bağımsız X'deki bir birimlik değişmenin, bağımlı değişken Y'de, ne kadarlık bir değişmeye yol açacağını gösteren regresyon katsayısıdır. b_0 ve b_1 ise anakütle regresyon katsayılarının (β_0 ve β_1' in) kestirimleridir.

Basit Doğrusal Regresyon Denkleminin Kestirimi

İki değişken arasında gerçek doğrusal bir ilişki varsa regresyon denklemi, β_0 ve β_1 'in $Y_i - \hat{Y}_i$ artıklarını küçük yapabilen kestiricilerin elde edilmesiyle bulunur. Varyans kavramı göz önünde tutulduğunda bunun bir ölçüsü artıkların kareler toplamıdır. Bu toplamı en küçükleyen β_0 ve β_1 kestiricilerinin elde edilmesine ilişkin kestirim yöntemi, **En Küçük Kareler (EKK) Yöntemi**'dir. β_0 ve β_1 ve σ^2' in kestirimlerinde sıkça bu yöntem kullanılmakla birlikte, bir başka kestirim yöntemi de en çok olabilirlik yöntemidir. Burada sadece yaygın olarak kullanılan, EKK yöntemiyle regresyon modelinin kestirimine yer verilecektir.

Katsayıların En Küçük Kareler (EKK) Kestirimleri

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

doğrusal ilişkisi, X ve Y değişkenlerinin anakütleleri için geçerlidir. İstatistiksel çalışmaların çoğunda olduğu gibi, regresyon analizinde de anakütleye ilişkin verilerin tümüne ulaşılamadığından, bu anakütleden seçilen örnek verileriyle analiz yapılır. Örnek verilerinden hareketle anakütle parametreleri olan β_0 ve β_1 'in kestirimlerini elde edebilmek için en küçük kareler yönteminden yararlanılabilir. Bunun için, öncelikle, gözlem ikililerini bir serpilme diyagramında gösterdiğimizi varsayalım. Serpilme diyagramı incelendiğinde doğrusal bir eğilim görülüyorsa, Y'in 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:

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right) = 0$$

En küçük kareler yönteminin esası β_0 ve β_1 'in kestirimleri olan b_0 ve b_1 'i söz konusu farkların kareleri toplamını minimum, yani

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{Y}\right)^2 = \text{min.}$$

olacak şekilde belirlemektir.

β_0 ve β_1 'in EKK kestirimleri, yukarıdaki en küçükleme yöntemi için

$$\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2 = \sum_{i=1}^{n}\left(Y_i - b_0 - b_1 X_i\right)^2$$

ifadesinin b_0 ve b_1 'e göre türevleri alınıp sıfıra eşitlenerek:

$$\frac{\partial e}{\partial b_0} = -2\sum\left(Y_i - b_0 - b_1 X_i\right) = 0$$

$$\frac{\partial e}{\partial b_1} = -2\sum X_i\left(Y_i - b_0 - b_1 X_i\right) = 0$$

bulunur, elde edilen bu iki eşitlikten,

$$\sum Y_i = b_0 n + b_1 \sum X_i$$

$$\sum X_i Y_i = b_0 \sum X_i + b_1 \sum X_i^2$$

denklemleri elde edilir. 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, $\sum X_i$, $\sum Y_i$, $\sum X_i^2$ ve $\sum X_i Y_i$ dır ve denklemlerdeki X_i ve Y_i değerleri "sıfır orijinine" göre ifade edilmişlerdir. Bu değerleri seri kıymetlerine dayanarak hesapladıktan sonra, basit doğrusal regresyon modelindeki β_0 ve β_1 'in kestirimleri olan b_0 ve b_1 'i normal denklemlere dayanarak kolaylıkla çözmek mümkündür. b_0 ve b_1 gibi iki bilinmeyenli iki denklem sisteminde bu katsayılar hesaplandığında Y 'nin X 'e göre doğrusal regresyon denklemi

$Y=b_0+b_1X$ Basit doğrusal regresyonda, doğru denkleminin parametreleri (b_0,b_1) en küçük kareler (EKK) tekniği ile hesaplanır.

$$\widehat{Y} = b_0 + b_1 X$$

şeklinde ifade edilecektir.

ÖRNEK 2

Örnek 1'de verdiğimiz probleme tekrar dönelim. Eskişehir ilinde satış yapan bir mağaza ürünlerini, yerel radyodaki reklamla tanıtmaktadır. Firmanın 5 hafta süresince, belirli bir ürün için harcadığı reklam tutarı ve satılan ürün sayısı tabloda verilmiştir. Satış miktarının, reklam harcamalarına göre, basit doğrusal regresyon denkleminin kestirimini, en küçük kareler tekniğiyle elde ediniz.

Reklam harcaması (X) (Milyon TL)	Satışlar (Y) (Adet)
10	3
20	4
30	6
40	7
50	10

ÇÖZÜM

Basit doğrusal regresyon denklemi $\hat{Y} = b_0 + b_1X$ 'i oluşturabilmek için b_0 ve b_1 katsayılarını normal denklemler yardımıyla hesaplayalım. Normal denklemler:

$$\sum Y = b_0 n + b_1 \sum X$$

$$\sum XY = b_0 \sum X + b_1 \sum X^2$$

'dir. Bu durumda, ΣX, ΣY, ΣX^2 ve ΣXY değerlerinin hesaplanması gerekir.

X	Y	XY	X^2
10	3	30	100
20	4	80	400
30	6	180	900
40	7	280	1600
50	10	500	2500
$\sum X = 150$	$\sum Y = 30$	$\sum XY = 1070$	$\sum X^2 = 5500$

$$30 = 5b_0 + 150b_1$$
$$1070 = 150b_0 + 5500b_1$$

Bu iki denklemin b_0 ve b_1 katsayıları;

$$b_0 = 0.9 \ ; \quad b_1 = 0.17$$

olarak hesaplanır. Buna göre, regresyon doğrusu denklemi,

$$\hat{Y} = 0.9 + 0.17X$$

şeklinde elde edilir. Regresyon katsayısı 0.17 bulunduğundan, X bağımsız değişkenindeki bir birimlik değişme, Y bağımlı değişkeninde 0.17 birimlik değişmeye neden olacaktır.

Normal denklemlerde X ve Y değerleri yerine bunların aritmetik ortalamalarından sapmaları olan x ve y değerlerinin konulmasıyla;

$$\sum y_i = b_0 n + b_1 \sum x_i$$

$$\sum x_i y_i = b_0 \sum x_i + b_1 \sum x_i^2$$

denklemleri elde edilir. Tanım gereğince $x_i = X - \bar{X}$ ve $y_i = Y - \bar{Y}$ olduğundan aritmetik ortalamanın temel özelliklerinden birincisine göre (aritmetik ortalamadan cebirsel sapmaların toplamı sıfırdır.) $\Sigma x_i = 0$ ve $\Sigma y_i = 0$ 'dır. Böylece son iki eşitlikten

$$b_0 = 0$$

ya da

$$b_1 = \frac{\sum x_i y_i}{\sum x_i^2}$$

Regresyon doğru denklemleri, sıfır ya da ortalamalar orijinine göre hesaplanabilir.

elde edilir. b_0'in 0'a eşit olması, ortalamalar orijinine göre, regresyon doğrusunun, değişkenlerin ortalamalarıyla tanımlanan bir noktadan geçtiğini ortaya koymaktadır. Bu durumda regresyon denklemi

$$\hat{y} = b_1 x_i \text{ veya } \hat{y} = b_{yx} x_i$$

şeklinde yazılabilir.

Bir regresyon doğrusu ister sıfır orijinine, ister ortalamalar orijinine göre yazılsın, eğimi değişmez. Bu nedenle her iki orijine göre hesaplanan b_1 katsayısı aynıdır. Buna karşılık b_0 ise

$$\Sigma Y_i = nb_0 + b_1 \Sigma X_i$$

eşitliğinden hareketle elde edilir. Eşitliğin her iki tarafı n ile bölündüğünde

$$\frac{\sum Y_i}{n} = \frac{nb_0}{n} + \frac{b_1 \sum X_i}{n}$$

$$\overline{Y} = b_0 + b_1 \overline{X}$$

ve

$$b_0 = \overline{Y} - b_1 \overline{X}$$

sonucuna ulaşılır. ve bu şekilde elde edildiğine göre, veriler için en iyi doğru denklemi;

$$\hat{Y} = b_0 + b_1 X$$

şeklinde yazılabilir.

ÖRNEK 3

Örnek 2'deki veriler için, Y'nin X'e göre regresyon denklemini, ortalamalar orijinine göre, en küçük kareler tekniğiyle elde edelim.

ÇÖZÜM

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.

$$\overline{X} = \frac{\sum X_i}{n} = \frac{150}{5} = 30 \qquad \overline{Y} = \frac{\sum Y_i}{n} = \frac{30}{5} = 6$$

x_i	y_i	$x_i y_i$	x_i^2
-20	-3	60	400
-10	-2	20	100
0	0	0	0
10	1	10	100
20	4	80	400
0	0	170	1000

$$b_0 = 0$$

$$b_1 = b_{yx} = \frac{\sum x_i y_i}{\sum X_i^2} = \frac{170}{1000} = 0.17$$

Bu durumda, ortalamalar orijinine göre, Y'nin X'e göre regresyon denklemi,

$$\hat{Y} = 0{,}17X$$

şeklinde yazılır.

Yukarıdaki açıklamalarda Y değişkeni bağımlı değişken, X değişkeniyse bağımsız değişken kabul edilmişti. X değişkeni bağımlı Y değişkeni bağımsız değişken olduğundaysa doğru denklemi;

$$\hat{X} = b_0 + b_1 Y$$

olarak ifade edilecektir. Bu durumda b_0 ve b_1

$$b_1 = \frac{\sum x_i y_i}{\sum y_i^2}$$

ya da

$$b_{xy} = \frac{\sum x_i y_i}{\sum y_i^2}$$

ve

$$b_0 = \bar{X} - b_1 \bar{Y}$$

formülleri yardımıyla bulunur. β_0 ve β_1 'in kestirimleri olan b_0 ve b_1 'in ve b_{yx} formülleri incelendiğinde, her ikisi de daima aynı işareti taşır, fakat aynı değerde değildir.

ÖRNEK 4

Örnek 1'de verdiğimiz reklam harcaması ve satılan ürün sayısı problemi için bu kez

i) X'in Y 'ye ve ortalamalar orijinine göre, regresyon denklemini,

ii) X'in Y'ye ve sıfır orijinine göre, regresyon denklemini hesaplayalım.

ÇÖZÜM

i) X'in Y'ye ve ortalamalar orijinine göre regresyon katsayısı

$$b_{xy} = \frac{\sum x_i y_i}{\sum y_i^2} = \frac{170}{30} = 5.66$$

olup, regresyon denklemi

$$\hat{X} = 5.66\ Y$$

şeklinde elde edilir.

ii) X'in Y'ye göre ve sıfır orijinine göre, regresyon denklemini yazabilmemiz için b_0 'ı hesaplamamız gerekir.

$$\overline{X} = b_0 + b_1\overline{Y}$$

$$b_0 = \overline{X} - b_1\overline{Y}$$

$$b_0 = 30 - 5.66(6)$$

$$b_0 = -3.96$$

ve regresyon denklemi

$$\widehat{X} = -3.96 + 5.66\,Y$$

şeklinde elde edilir.

SIRA SİZDE

1. Bir öğretim üyesi, öğrencilerin A dersinden aldıkları final notlarının, öğrencilerin vize notlarına bağlı olduğunu düşünmektedir. Öğretim üyesinin bu düşüncesiyle oluşturulabilecek bir doğrusal regresyon denklemindeki bağımlı ve bağımsız değişkenler neler olacaktır?

2. Bağımlı değişkenin sahip olduğu değerle, bu bağımlı değişken için, doğrusal regresyon denkleminden elde edilen tahmin değeri arasındaki farka ne ad verilir?

3. Aşağıda (X) ve (Y) değişkenleri için gözlem değerleri verilmiştir. (X) bağımsız değişken ve (Y) de bağımlı değişken olarak alınırsa, en küçük kareler yöntemine göre regresyon denklemi ne olacaktır?

X	Y
3	2
5	4
7	6
9	5

VARYANSIN (σ^2) KESTİRİMİ

AMAÇ 3

Basit doğrusal regresyon modeliyle elde edilen kestirimlerin standart hatasını hesaplayabileceksiniz.

Basit doğrusal regresyon modelinde, β_0 ve β_1 'in kestirimlerine ek olarak, aralık kestirimlerinde ve hipotez testlerinde gerekli olan σ^2'in kestirimine de gereksinim vardır.

σ^2, ε_i hata terimlerinin ortak varyansıdır. ε_i'in kestirimi e_i hata terimi olduğundan e_i'lerin varyansı da σ^2'in bir kestirimi olacaktır. Hataların kareler toplamı,

$$HKT = \sum e_i^2$$

$$= \sum\left(Y_i - \widehat{Y}_i\right)^2$$

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 kestirimidir. Basit doğrusal regresyon modeliyle hataların hesaplanmasında, β_0 ve β_1 'in kestiricileri b_0 ve b_1 kullanıldığından, serbestlik derecesi (n-2) olarak yazılır.

HKO 'un kare kökü alındığındaysa denklemin standart hatası elde edilir ve $\hat{\sigma}$ ile gösterilir.

$$\hat{\sigma} = \sqrt{HKO}$$

$$= \sqrt{\frac{\sum \left(Y_i - \hat{Y}_i\right)^2}{n-2}}$$

ÖRNEK 5 ***Örnek 1'deki verileri kullanarak, regresyon denklemine dayanarak yapılacak kestirimlerin standart hatası ($\hat{\sigma}$)'ı hesaplayalım.***

ÇÖZÜM

$\hat{\sigma}$ 'ı hesaplayabilmek için öncelikle, $\hat{Y}$ 'ları daha sonra da $\left(Y - \hat{Y}\right)$ ve $\left(Y - \hat{Y}\right)^2$ 'leri hesaplayalım.

$$\hat{Y} = 0{,}9 + 0{,}17X$$

$\hat{Y}$	$\left(Y - \hat{Y}\right)$	$\left(Y - \hat{Y}\right)^2$
2.6	0.4	0.16
4.3	-0.3	0.09
6	0	0
7.7	-0.7	0.49
9.4	0.6	0.36
		1.1

buradan,

$$\hat{\sigma} = \sqrt{\frac{1{,}1}{5-2}} = \sqrt{0.36} = 0.6$$

olarak elde edilir.

SIRA SİZDE

1. **ε_i hata terimlerinin ortak varyansı nedir?**
2. **Basit doğrusal regresyon modelinde serbestlik derecesi ne olacaktır?**
3. **$\hat{\sigma}$ 'in kestirimi nerelerde kullanılır?**

BASİT DOĞRUSAL REGRESYONDA ARALIK KESTİRİMİ

AMAÇ 4 ***Basit doğrusal regresyonda, parametrelerin nokta kestiriminden sonra genelleme yapabilmek için, parametrelere ilişkin aralık kestirimi yapabileceksiniz.***

İstatistiksel çıkarsamalarda yapılan kestirimlerin, gerçek değerlerle genellenmesi aralık kestirimleriyle yapılır. Regresyon çözümlemesi, örneklem verileriyle yapıldığından, elde edilen b_0 ve b_1 'lerin anakütle parametreleri, β_0 ve β_1'e ilişkin aralık testlerinde de elde edilmelidir.
b_0 katsayısı için t örnekleme dağılımı yardımıyla β_0 için güven aralığı

$$P\,(b_0 - t_\alpha \,.\, s_{b_0} \leq \beta_0 \leq b_0 + t_\alpha \,.\, s_{b_0}) = 1 - \alpha$$

şeklinde verilir. S_{b0}, b_0 'in standart hatasıdır ve

$$s_{b_0} = \hat{\sigma}\sqrt{\frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2}}$$

formülüyle hesaplanır, t_α ise α anlamlılık düzeyi ve n-2 serbestlik derecesinde t tablosundan bulunan değerdir.
β_1 için güven aralığı

$$P\,(b_1 - t_\alpha \,.\, s_{b_1} \leq \beta_1 \leq b_1 + t_\alpha \,.\, s_{b_1}) = 1 - \alpha$$

formülüyle elde edilir. s_{b_1} , b_1 'in standart hatasıdır ve

$$s_{b_1} = \frac{\hat{\sigma}}{\sqrt{\sum (X_i - \overline{X})^2}}$$

şeklinde hesaplanır.

ÖRNEK 6

Örnek 2'de örnek 1'in verilerini kullanarak $b_1 = 0.17$ olarak hesaplanmıştır. Bu sonuca göre, β_1 katsayısının %95 güven aralığını hesaplayalım.

ÇÖZÜM

$b_1 = 0{,}17$

$$s_{b_1} = \frac{\hat{\sigma}}{\sqrt{\sum (X_i - \overline{X})^2}} = \frac{0.6}{\sqrt{1000}} = \frac{0.6}{31.6} = 0.019$$

ve ν=3 serbestlik derecesinde t tablo değeri 3.182'dir. Buna göre β_1 için %95 güven aralığı

$0.17 \pm (3.182)(0.019)$
$P(0.109 < \beta_1 < 0.23) = 0.95$

olarak hesaplanır.
Regresyon katsayısı β_1 'in 0.95 olasılıkla alabileceği değerler 0.109 ile 0.23 olacaktır.

SIRA SİZDE

1. **b_0 katsayısının güven aralığı oluşturulurken, hangi örnekleme dağılımı kullanılır?**
2. **$b_1 = 0.48$, $s_{b_1} = 0.002$, $\alpha = 0.05$, $\nu = 5$ için t tablo değeri t = 2.571 iken, β_1 katsayısının %95 güven aralığını hesaplayınız?**
3. **Basit doğrusal regresyonda, parametreler hakkında genellemeler yapabilmek için, hangi tekniğe ihtiyaç duyulur?**

REGRESYON KATSAYILARININ ANLAMLILIK TESTLERİ

AMAÇ 5

Basit doğrusal regresyon denkleminde elde edilen parametre kestirimlerinin, istatistiksel olarak anlamlılığını test edebileceksiniz.

Basit doğrusal regresyon modelinde regresyon katsayısına ilişkin test, regresyon doğrusunun anlamlılığını da test eder.

Basit doğrusal regresyon analizinde, bir bağımlı bir bağımsız değişken olması nedeniyle, test edilecek parametreler β_0 ve β_1 olacaktır. Daha önce açıklandığı gibi, β_0 'ın kestirimi b_0 regresyon sabitidir ve b_1 ise β_1 'in kestirimi olup regresyon katsayısıdır.

Basit doğrusal regresyon modelindeki regresyon katsayısına ilişkin yapılan test, regresyon doğrusunun anlamlılığını da test etmektedir. Şöyle ki:

$H_0 : \beta_1 = 0$
$H_1 : \beta_1 \neq 0$

hipotezleri

$$t_h = \frac{b_1}{s_{b_1}}$$

istatistiğinden yararlanılarak test edilir. α anlam düzeyinde n-2 serbestlik derecesinde t tablosundan bulunan değer, hesaplanan t test istatistiğinden büyükse $H_0 : \beta_1 = 0$ hipotezi kabul edilir ve regresyon doğrusu anlamlı değildir. Diğer bir ifadeyle Y'deki değişimler X'deki değişmelerden kaynaklanmamaktadır (X ve Y arasında doğrusal bir ilişki yoktur.). t istatistiğinin değeri t tablo değerinden büyükse H_0 reddedilir, yani, regresyon doğrusu anlamlıdır. Elde edilen doğrusal regresyon modeli amaca uygun olarak kullanılabilir.

ÖRNEK 7 ***Aynı örneğimiz için regresyon katsayısının anlamlılık testini yapalım.***

ÇÖZÜM

$H_0 : \beta_1 = 0$
$H_1 : \beta_1 \neq 0$

Hipotezlerini test edelim. $t_h = \frac{b_1}{s_{b_1}}$, örneklem küçük olduğu için benimsenen t, ν = n-2 = 5-2 = 3 serbestlik derecesiyle t dağılır.

$$t = \frac{b_1}{s_{b_1}} = \frac{0.17}{0.019} = 8.94$$

$\alpha = 0.05$ ve n-2 = 5-2 = 3 serbestlik derecesi ile $t_{0.05} = 3.182$ olduğundan $t > t_{0.05}$ dir ve H_0 reddedilecektir, b_1 kestirimi istatistiksel olarak anlamlıdır.

SIRA SİZDE

1. **Regresyon denklemi Y = 5.08 + 1.58 X ve katsayıların standart hataları $s(b_0) = 0.29$, $s(b_1) = 0.067$ olarak verilmiştir. Katsayıların anlamlılığını, %5 anlam düzeyine göre test ediniz.**

2. **Basit doğrusal regresyon modelinde katsayının istatistiksel olarak anlamlı olmasının pratik anlamı nedir?**

3. **Regresyon katsayılarının testi için hipotezler nasıl kurulur?**

Kendimizi Sınayalım

1. X ve Y değişkenleri arasındaki ilişkiyi gösteren matematiksel fonksiyona ne ad verilir?

a. Regresyon denklemi
b. Korelasyon denklemi
c. Anlamlılık testi
d. Hipotez testi
e. Bağımlı değişken

2. Değişkenler arasındaki ilişkinin gösterilmesinde kullanılan grafik yöntemi aşağıdakilerden hangisidir?

a. Histogram
b. Standart normal eğri
c. Serpilme diyagramı
d. Kök-yaprak diyagramı
e. Poligon

3-4 ve 5. sorular aşağıdaki bilgilere göre cevaplandırılacaktır.

x : 40, 20, 22, 14
y : 12, 15, 14, 18 bileşik serisi verilmiştir.

3. Yukarıdaki verilere göre, $\sum XY$ değeri kaçtır?

a. 750
b. 825
c. 1250
d. 1340
e. 2340

4. Yukarıda verilen serinin basit doğrusal regresyon denklemini aşağıdakilerden hangisidir?

a. $\hat{Y} = 50 + 8.2X$
b. $\hat{Y} = 18.35 - 0.15X$
c. $\hat{Y} = 18.35 + 0.15X$
d. $\hat{Y} = 50 - 8.2X$
e. $\hat{Y} = 0.15 + 18.35X$

5. Yukarıda verilen bileşik seri için ortalamalar orjinine göre regresyon denklemi aşağıdakilerden hangisidir?

a. $\hat{Y} = 18.35x$
b. $\hat{Y} = 0.15x$
c. $\hat{Y} = -0.15x$
d. $\hat{Y} = -18.35x$
e. $\hat{Y} = 8.2x$

6. Bağımlı değişken (Yi) ile bu değişken için regresyon denkleminden elde edilen tahmin değeri ($\hat{Y}_i$) arasındaki farka ne ad verilir?

a. Standart hata
b. Tahmin hatası
c. Standart sapma
d. Varyans
e. Regresyon katsayısı

7. $\sum \left(Y_i - \hat{Y}\right)^2 = 2,2$, n = 6 olan basit doğrusal regresyon denklemiyle elde edilen tahminlerin standart hatası kaçtır?

a. 0.36
b. 0.55
c. 0.60
d. 0.75
e. 0.91

8-9 ve 10. sorular aşağıdaki bilgilere göre cevaplandırılacaktır.

Regresyon denklemi $\hat{Y} = 1.1 + 0.81X$, $s(b_1) = 0.09$ ve n= 6 olarak verilmiştir.

8. Regresyon katsayısının 0.01 anlam düzeyinde test edilmesi istediğinde, test istatistiğinin değeri aşağıdakilerden hangisidir?

a. t = 9
b. t = 3
c. t = 2.9
d. t = 1.9
e. t = 0.9

9. Yukarıdaki verilere göre, hipotez testinde sıfır hipotezinin red bölgesi aşağıdakilerden hangisidir?

a. $|t| > |3.747|$
b. $|t| > |4.604|$
c. $|t| > |2.776|$
d. $t < 3.182$
e. $t < 2.132$

10. Yukarıdaki verilere göre, regresyon denklemi $\hat{Y} = 1.1 + 0.81X$ olduğunda, (b) regresyon katsayısı aşağıdakilerden hangisinde doğru olarak ifade edilmiştir?

a. X'deki bir birimlik artma (ya da azalma) Y'de 0.81 birimlik artışa (ya da azalmaya) neden olur.
b. X'in değeri sıfır olduğunda Y'nin değeri de sıfır olur.
c. Y'de bir birimlik azalma X'de de bir birimlik azalmaya neden olur.
d. Y'de 1.1 birimlik azalma X'de 2 birimlik artışa neden olur.
e. X'deki artış Y'yi etkilemez.

Yanıt Anahtarı

1. a
2. c
3. d
4. b
5. c
6. b
7. d
8. a
9. b
10. a

Yararlanılan Kaynaklar

ORHUNBİLGE, Neyran: **Uygulamalı Regresyon ve Korelasyon Analizi**, İ. Ü. İşletme Fakültesi, No : 267, İstanbul, 1996.

ÇÖMLEKÇİ, Necla: **Temel İstatistik İlke ve Teknikleri**, Bilim Teknik Yayınevi, Eskişehir, 1989.

NEWBOLD, Paul: **İşletme ve İktisat için İstatistik,** Çeviren : Ümit Şenesen, 4. Basım, Literatür Yayıncılık, 2000.

SERPER, Özer: **"Uygulamalı İstatistik II",** 4. Baskı, Ezgi Kitabevi, Bursa, 2000.

12

Korelasyon

Çalışma Biçimine İlişkin Olarak

- *Kavramlar ve bu kavramlar arasındaki ilişkiler dikkatle incelenmeli,*
- *Örnekler ve örnek çözümleri dikkatle incelenmeli, sorunlarla karşılaşılırsa ilgili ünitelere geri dönülmeli,*
- *Regresyon konusu yeterince özümsenmelidir.*

Amaçlar

- *İki değişken arasındaki ilişkinin yönünü ve derecesini belirleyebileceksiniz.*
- *Bağımlı değişkendeki değişmenin yüzde kaçının bağımsız değişken tarafından açıklandığını belirleyebileceksiniz.*
- *Anakütlede korelasyon olup olmadığı yolundaki hipotezleri, belirli anlam düzeyinde test edebileceksiniz.*

İçerik Haritası

- *GİRİŞ*
- *KORELASYON KATSAYISI*
- *BELİRLİLİK KATSAYISI*
- *KORELASYON KATSAYISININ ANLAMLILIK TESTİ*

GİRİŞ

Bir önceki ünitede iki değişken arasındaki ilişkinin doğrusal regresyon modeliyle gösterimi üzerinde durulmuştu. Bu ünitedeyse iki değişken arasındaki ilişkinin yönü ve derecesinin belirlenmesi konusu **"korelasyon analizi"** ele alınacaktır. Bu tip analizin aracı korelasyon katsayısıdır. Hemen belirtelim ki korelasyon ve regresyon birbirleriyle yakın ilişkileri olan konulardır.

KORELASYON KATSAYISI

AMAÇ 1 ***İki değişken arasındaki ilişkinin yönünü ve derecesini belirleyebileceksiniz.***

İki değişken arasındaki doğrusal ilişkinin yönü ve derecesi korelasyon katsayısıyla ölçülür. Korelasyon katsayısı r ile gösterilir.

Turistik bir bölgede, elde edilen turizm geliriyle yatak sayısı arasındaki ilişkinin derecesi bölgeye, yeni yatırımların yapılmasına gösterge oluşturacaktır. Bölgeye yatırım yapılmasını isteyen bir grup, yatak sayısı arttırılırsa, daha çok turist geleceğini ileri sürmektedir. Bu ve benzer iddiaların geçerliliği yine uygun istatistiksel teknikler kullanılarak araştırılır.

Bu ünitede değişkenler arasındaki ilişkinin derecesi ve yönünün belirlenmesinde kullanılan istatistiksel teknikler ele alınmıştır.

Korelasyon katsayısı neden sonuç ilişkilerinin kurulmasında yeterli olmaz.

İki değişken arasındaki doğrusal ilişkinin derecesi, " r " simgesiyle gösterilen korelasyon katsayısıyla ölçülür. Korelasyon katsayısı iki değişkenin değişimlerinde, ne dereceye kadar uygunluk olduğunu belirler. Fakat hiç bir şekilde neden - sonuç ilişkisi kurmaz.

Aslında bir çok durumda, modelin değişkenlerinden hangisinin bağımsız değişken, hangisinin bağımlı değişken olduğu bilinmez. İşte bu gibi durumlarda, ilişkinin derecesinin belirlenmesinde oransal bir ölçü olan, "korelasyon katsayı-sı"ndan yararlanılır.

Korelasyon katsayısı r, –1 ile +1 arasında değerler alır ($-1 \leq r \leq +1$).

Korelasyon katsayısının alabileceği en küçük değer –1, en büyük değerse +1 olur, başka bir anlatımla korelasyon katsayısı r,

$$-1 \leq r \leq +1$$

arasında değer alır.

Korelasyon katsayısının işareti pozitifse, değişkenlerden birinin değeri artarken (azalırken) diğerinin de arttığını (azaldığını) gösterir. Korelasyon katsayısının işareti negatifse, değişkenlerden birinin değeri artarken (azalırken) diğerinin değerinin azaldığını (arttığını) gösterir. Yani ters yönlü bir ilişki söz konusudur.

r = 0 olduğundaysa değişkenler arasında doğrusal bir ilişkinin bulunmadığı söylenebilir.

r 'nin +1'e eşit olması, değişkenler arasında pozitif ve tam doğrusal bir ilişkinin varlığını ortaya koyar.

r 'nin -1'e eşit olmasıysa, değişkenler arasında negatif ve tam doğrusal bir ilişkiyi belirler. Değişkenler arasındaki ilişki kuvvetlendikçe ±1'e, zayıfladıkça da sıfıra yaklaşan bir korelasyon katsayısı elde edilir.

Korelasyon katsayısı,

$$r = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}} = \frac{\sum \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum \left(X_i - \overline{X}\right)^2 \sum \left(Y_i - \overline{Y}\right)^2}}$$

ile hesaplanır.

ÖRNEK 1 ***Öğrencilerin istatistik dersinde ara sınavdan aldıkları notlarla dönem sonu sınavından aldıkları notlar arasında bir ilişki olduğu düşünülmektedir. Bu ilişkinin yönünü ve derecesini belirleyelim.***

İstatistik Dersi Ara Sınav Notları X	İstatistik Dersi Dönem Sonu Sınav Notları Y
45	83
54	78
55	80
68	72
30	45
48	26
300	384

ÇÖZÜM

Öğrencilerin ara sınav notlarıyla dönem sonu notları arasındaki ilişkinin derecesini korelasyon katsayısıyla belirleyebiliriz.

$$\bar{X} = \frac{\sum X_i}{n} = \frac{300}{6} = 50 \qquad \bar{Y} = \frac{\sum Y_i}{n} = \frac{384}{6} = 64$$

olduğundan,

$$x_i = X_i - \bar{X} \qquad y_i = Y_i - \bar{Y}$$

dönüşümüyle,

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
–5	19	-95	25	631
4	14	56	16	196
5	16	80	25	256
18	8	144	324	64
–20	–19	380	400	361
–2	–38	76	4	1444
		641	794	2682

$$r = \frac{\sum x_i\, y_i}{\sqrt{\sum x_i^2 \sum y_i^2}} = \frac{641}{\sqrt{794.2682}} = \frac{641}{1459} = 0.439$$

olarak hesaplanır.

Öğrencilerin, istatistik dersiyle ilgili, ara sınav notlarıyla dönem sonu sınav notları arasında, pozitif yönde, kuvvetli olmayan bir ilişki söz konusudur.

Korelasyon katsayısı, regresyon katsayılarından da yararlanılarak aşağıdaki eşitlikle hesaplanabilir.

$$r = \pm \sqrt{b_{yx} \cdot b_{xy}}$$

Korelasyon katsayısı r, regresyon katsayıları yardımıyla da hesaplanabilir $\left(r = \pm \sqrt{b_{yx} \cdot b_{xy}}\right)$.

formüldeki,

b_{yx} = Y'nin X'e göre regresyon katsayısı
b_{xy} = X'in Y'ye göre regresyon katsayısıdır.

Burada, dikkat edilmesi gereken nokta, eğer regresyon katsayıları pozitifse r pozitif, eğer her iki regresyon katsayısı da negatifse r negatif olacaktır. Eğer regresyon katsayılarından biri pozitif, diğeri de negatifse, değişkenler arasında ilişki yoktur.

ÖRNEK 2

Örnek 1'de yer alan veriler için, korelasyon katsayısını, regresyon katsayılarından yararlanarak hesaplayalım.

ÇÖZÜM

Bu yaklaşımla, korelasyon katsayısını hesaplayabilmek için öncelikle, regresyon katsayılarını elde edelim :

$$b_{yx} = \frac{\sum y_i\, x_i}{\sum x_i^2} = \frac{641}{794} = 0.807$$

$$b_{xy} = \frac{\sum y_i\, x_i}{\sum y_i^2} = \frac{641}{2682} = 0.239$$

olarak elde edilir. Bu sonuçlardan,

$$r = +\sqrt{0.807 \;.\; 0.239} = +\sqrt{0.912}$$

$$r = +\,0.439$$

olarak elde edilir (Her iki regresyon katsayısının da pozitif olduğuna dikkat ediniz).

Çözümlerden de görüleceği gibi, korelasyon katsayısı hangi yaklaşımla hesaplanırsa hesaplansın, aynı sonuca ulaşılacaktır. Burada dikkat edilmesi gereken husus şudur: Regresyon katsayılarının işaretiyle korelasyon katsayısının işareti aynı olacaktır.

ÖRNEK 3

Alfa üretim işletmesinin belirli bir döneme ilişkin üretim miktarıyla birim değişken maliyetleri aşağıdaki gibidir:

Üretim Miktarı (Bin Ton) X	Birim Değişken Maliyetler (Milyon TL) Y
1	11
2	9
3	8
4	7
5	5
15	40

Regresyon katsayılarından yararlanarak korelasyon katsayısını hesaplayalım.

ÇÖZÜM

Öncelikle regresyon katsayılarını bulalım. Bunun için ilgili hesaplamalar aşağıda verilmiştir:

X_i	Y_i	$x_i = X_i - \bar{X}$	$y_i = Y_i - \bar{Y}$	$x_i y_i$	x_i^2	y_i^2
1	11	–2	3	–6	4	9
2	9	–1	1	–1	1	1
3	8	0	0	0	0	0
4	7	1	–1	–1	1	1
5	5	2	–3	-6	4	9
15	**40**	**0**	**0**	**-14**	**10**	**20**

$\bar{X} = 3$ $\bar{Y} = 8$

bu sonuçlardan hareketle regresyon katsayıları,

$$b_{yx} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{-14}{10} = -1.4$$

$$b_{xy} = \frac{\sum x_i y_i}{\sum y_i^2} = \frac{-14}{20} = -0.7$$

olarak hesaplanır. Buradan,

$$r = \pm\sqrt{b_{yx}\,.\,b_{xy}} = \pm\sqrt{(-1.4)(-0.7)} = -0.9899$$

olarak elde edilir (her iki regresyon katsayısının da negatif olduğuna dikkat ediniz).

Üretim hacmiyle birim değişken maliyetler arasında, ters yönde, kuvvetli bir ilişki vardır.

SIRA SİZDE

1. Korelasyon katsayısı hangi amaçla kullanılır?

2. Korelasyon katsayısının alabileceği en küçük ve en büyük değer nedir?

3. X ve Y değişkenleri için korelasyon katsayısı r = 0,85 bulunduğuna göre, sonucu yorumlayınız.

BELİRLİLİK KATSAYISI

AMAÇ 2

Bağımlı değişkendeki değişmenin yüzde kaçının bağımsız değişken tarafından açıklandığını belirleyebileceksiniz.

Belirlilik katsayısı r^2, bağımlı değişkendeki değişimin ne kadarının bağımsız değişkence açıklandığını ifade eder.

Regresyon denkleminin verilere olan uyumunun sağlanıp sağlanmadığının bir göstergesi de açıklanabilen değişimin toplam değişime olan oranıdır. Bu oran, belirlilik katsayısı olarak isimlendirilir ve bağımlı değişkendeki değişimin ne kadarının bağımsız değişkence açıklanabildiğini gösterir. Korelasyon katsayısının karesine eşit olan belirlilik katsayısının alabileceği en küçük ve en büyük değerler sıfırla artı birdir.

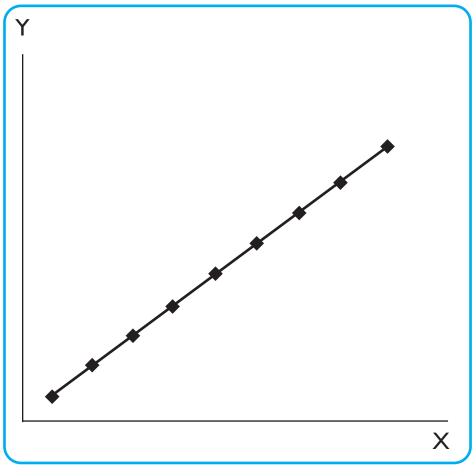
$$0 \leq r^2 \leq 1$$

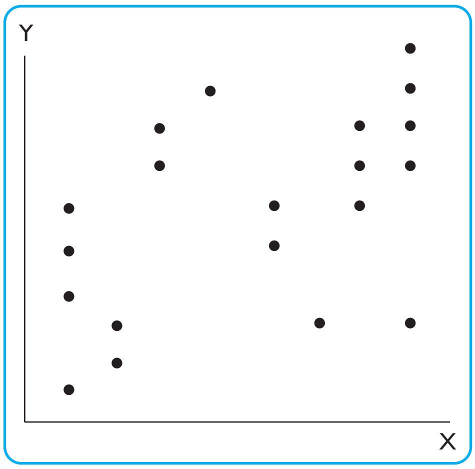
Örneklem verilerinden hareketle, kestirimi yapılan regresyon doğrusu, serpilme diyagramında, noktalar arasından geçmektedir. Belirlilik katsayısı noktaların doğruya yakınlık derecesini gösterir. Fakat sadece şekle bakarak bunu görmek olası değildir. Bu nedenle, belirlilik katsayının sayısal olarak elde edilmesi gerekir.

$r^2 = 1$ ise Y'deki değişimin %100'ünün X bağımsız değişkeni tarafından açıklanabildiği kabul edilir ve serpilme diyagramında tüm noktalar regresyon doğrusu üzerindedir.

Belirlilik katsayısı r^2, 0 ile 1 arasında değerler alır.

$r^2 = 0$ ise X bağımsız değişkeni, Y bağımlı değişkenini hiç açıklayamıyor demektir.





Şekil 12.1 *$r_2 = 1$ olması durumu.*

Şekil 12.2 *$r_2 = 0$ olması durumu (hiç ilişki yok).*

Özetle r^2 sıfırdan küçük ve birden büyük olamaz.

ÖRNEK 4

Örnek 1'deki veriler için belirlilik katsayısını hesaplayalım.

ÇÖZÜM

Belirlilik katsayısının korelasyon katsayısının karesine eşit değer olduğuna göre,

$r^2 = (0.439)^2$
$r^2 = 0.192$

olarak elde edilir. Belirlilik katsayısının değeri sıfıra yakın çıkması nedeniyle, dönem sonu sınavında alınan notla, ara sınav notlarından çok az etkilenmiştir.

SIRA SİZDE

1. **Belirlilik katsayısının bire yakın olması durumunda serpilme diyagramında noktalar nasıl yer alır?**

2. **Belirlilik katsayısı ne amaçla kullanılır?**

3. **X bağımsız değişkeni Y bağımlı değişkenini hiç açıklamıyorsa r^2 hangi değeri alacaktır?**

KORELASYON KATSAYISININ ANLAMLILIK TESTİ

AMAÇ 3

Anakütlede korelasyon olup olmadığı yolundaki hipotezleri, belirli anlam düzeyinde test edebileceksiniz.

Anakütle ilişki katsayısı $\rho=0$ olan bir anakütleden seçilen örneklemlerin r katsayıları normal dağılıma sahiptir. Bu nedenle,

$$H_0 : \rho = 0$$
$$H_1 : \rho \neq 0$$

hipotezleri formüle edildikten sonra

$$t = \frac{r - \rho}{s_r} = \frac{r}{s_r}$$

buradan

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

test istatistiği hesaplanır.

Korelasyon katsayısının hesaplanması, çoğu zaman, küçük örneklemlere dayandığı için formülde, $\rho = 0$ hipotezine göre yapılacak testte bu formülü kullanmak mümkün olur. Formülde s_r, r'nin standart hatasıdır. $n-2$ serbestlik dereceli α anlam düzeyinde tablodan bulunan t değeriyle t istatistiğinin değeri karşılaştırılıp, istatistiksel olarak anlamlı olup olmadığına karar verilir.

ÖRNEK 5

Örnek 1'de hesaplanan $r = 0.439$ değerinin anlamlılığını $\alpha = 0.05$ için sınayalım.

ÇÖZÜM

Hipotezler

$$H_0 : \rho = 0$$
$$H_1 : \rho \neq 0$$

Anlamlılık düzeyi $\alpha = 0.05$

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \qquad (n < 30)$$

$$t = \frac{0.439\sqrt{6-2}}{\sqrt{1-0.192}} = \frac{0.878}{0.898} = 0.97$$

$t_{0.05} = 2.776 > t = 0.97$ olduğundan H_0 kabul edilecektir. Anakütlede korelasyon olmadığı yolundaki H_0 hipotezi %5 anlam düzeyinde kabul edilecektir.

SIRA SİZDE

1. **Anakütle ilişki katsayısı p = 0 olan bir anakütleden çekilen r katsayıları nasıl bir dağılıma sahiptir?**

2. **s_r neyi gösterir?**

3. **$t_{0.05} = 2.776 < t = 3.105$ olarak verilmiştir. Sonucu yorumlayınız.**

Kendimizi Sınayalım

1. X ve Y değişkenleri arasındaki ilişkinin derecesini, aşağıdakilerden hangisi belirler?

a. Regresyon katsayısı
b. t istatistiği
c. Korelasyon katsayısı
d. Aritmetik Ortalama
e. Otokorelasyon katsayısı

2. Aşağıdaki korelasyon katsayılarından hangisi, X ve Y değişkenleri arasında, ters yönde tam doğrusal bir ilişkinin varlığını gösterir?

a. $r = 1$
b. $r = 0.5$
c. $r = 0$
d. $r = -0.5$
e. $r = -1$

3. Y'nin X'e göre regresyon katsayısı $b_{yx} = 0.86$, X'in Y'ye göre regresyon katsayısı $b_{yx} = 0.90$ ise korelasyon katsayısı (r) kaçtır?

a. $r = 0.88$
b. $r = 0.78$
c. $r = 0$
d. $r = -0.78$
e. $r = -0.88$

4. Korelasyon katsayısıyla ilgili aşağıdaki ifadelerden hangisi **yanlıştır?**

a. $-1 \leq r \leq 1$
b. $-1 \leq r^2 \leq 1$
c. $b_{yx} > 0$ ise $r > 0$'dır.
d. $b_{yx} < 0$ ise $r < 0$'dır.
e. $r^2 = 0$

5. X ve Y değişkenleri için korelasyon katsayısı $r = 0.75$ ise aşağıdaki ifadelerden hangisi doğrudur?

a. X ve Y değişkenleri arasındaki ilişki zayıftır.
b. X ve Y değişkenleri arasında ilişki yoktur.
c. X ve Y değişkenleri arasında ters yönlü ve zayıf bir ilişki vardır.
d. X ve Y değişkenleri arasındaki doğru yönlü ve kuvvetli bir ilişki vardır.
e. X ve Y değişkenleri arasında doğru yönlü ve çok zayıf bir ilişki vardır.

6. Belirlilik katsayısının kullanım amacı aşağıdakilerden hangisidir?

a. X ve Y değişkenleri arasındaki regresyonu açıklamak
b. X ve Y değişkenleri arasında hata miktarını göstermek
c. Bağımlı değişkendeki değişimin ne kadarının bağımsız değişkence açıklanabildiğini göstermek
d. Tahminlerin hesaplamasında kullanmak
e. Standart hatayı azaltmak

7. X ve Y değişkenleri arasındaki ilişkiyi gösteren korelasyon katsayısı $r = 0,82$ olduğunda, belirlilik katsayısı aşağıdakilerden hangisidir?

a. $r^2 = 0.75$
b. $r^2 = 0.67$
c. $r^2 = 0$
d. $r^2 = -0.67$
e. $r^2 = -1$

8. X bağımsız değişkeni Y bağımlı değişkenini hiç açıklamıyorsa r^2 değeri kaçtır?

a. $r^2 = -1$
b. $r^2 = -0.5$
c. $r^2 = 0$
d. $r^2 = 0.5$
e. $r^2 = 1$

9. Korelasyon katsayısının anlamlılık sınamasında sıfır hipotezinin ifadesi aşağıdakilerden hangisidir?

a. $H_0 : p = 0$
b. $H_0 : p > 0$
c. $H_0 : p < 0$
d. $H_1 : p = 0$
e. $H_1 : p < 0$

10. Örneklem büyüklüğü $n = 8$, $r = 0.72$ değerinin anlamlılığını sınamak için hesaplanan test istatistiğinin değeri aşağıdakilerden hangisidir?

a. $t = 3.55$
b. $t = 2.55$
c. $t = 2$
d. $t = 1.764$
e. $t = -3$

Yanıt Anahtarı

1. c
2. e
3. a
4. b
5. d
6. c
7. b
8. c
9. a
10. b

Yararlanılan Kaynaklar

ORHUNBİLGE, Neyran: **Uygulamalı Regresyon ve Korelasyon Analizi**, İ. Ü. İşletme Fakültesi, No: 267, İstanbul, 1996.

ÇÖMLEKÇİ, Necla: **Temel İstatistik İlke ve Teknikleri**, Bilim Teknik Yayınevi, Eskişehir, 1989.

NEWBOLD, Paul: **İşletme ve İktisat için İstatistik**, Çeviren: Ümit Şenesen, 4. Basım, Literatür Yayıncılık, 2000.

SERPER, Özer: **Uygulamalı İstatistik II**, 4. Baskı, Ezgi Kitabevi, Bursa, 2000.

13

İndeksler

Çalışma Biçimine İlişkin Olarak

- *Ortalamalar konusunu yeniden gözden geçirilmeli,*
- *Tanımları dikkatle okunmalı,*
- *Örnekleri dikkatle gözden geçirilmelidir.*

Amaç

İndeks kavramını açıklayabilecek, basit ve bileşik indeksler hesaplayabileceksiniz.

İçerik Haritası

- *GİRİŞ*
- *İNDEKSLER*
  - *Mekan ve Zaman İndeksleri*
  - *Basit ve Bileşik İndeksler*

GİRİŞ

Özellikle, ağır bir enflasyonun yaşandığı ülkemizde, indeks sözcüğü kulağa pek yabancı gelmemektedir. Ekonomik haber ve yorumlarda, örneğin; geçen yılın X ayındaki toptan eşya fiyatları indeksiyle içinde bulunulan yılın X ayındaki toptan eşya fiyatları indeksi arasındaki yüzde şu kadar artışın ya da azalışın gerçekleşmesinden hareketle, ülke ekonomisinin geldiği nokta ve geleceğine ilişkin olumlu ya da olumsuz tablolar ve yaşanan enflasyonun kaçınılmaz sonuçları toplumun beklentilerini önemli ölçüde etkilemektedir.

Bir ülkedeki sosyo ekonomik nitelikteki eğilimlerin belirlenebilmesi ve bu konuda verilecek kararların yerindeliği, ancak doğru, güvenilir ve zamanında ulaşılabilir istatistiksel bilgilerle gerçekleştirilebilir. Sözü edilen bilgilerin üretilmesinde önemli bir istatistiksel araç da indekslerdir.

Bu ünitede indeks konusu, yeterli ayrıntıyla ele alınmıştır.

İNDEKSLER

AMAÇ 1

İndeks kavramını açıklayabilecek, basit ve bileşik indeksler hesaplayabileceksiniz.

İndeks, bir istatistiksel olaya ilişkin gözlem değerlerinin, zaman ya da mekana göre gösterdiği oransal değişimler, olarak tanımlanır. Tanım uyarınca, indekslerde biri temel, diğeri de karşılaştırılan (cari) değer olmak üzere, iki değer söz konusudur. İndeks hesaplanırken karşılaştırılan değer paya, temel değer paydaya yazılır ve kıyaslamayı daha basit ifade edebilmek için sonuç 100 ile çarpılır. Böylece temel değer 100 olmak üzere diğer değerlerdeki değişimlerin, temel değere göre kaç olacağı belirlenir.

İndeks, bir istatistiksel olaya ilişkin gözlem değerlerinin zaman ya da mekana göre gösterdiği oransal değişimlerin ölçüsüdür.

Tüm bu sözü edilenler, I indeksi, x_0 temel değeri x_i ve de i. gözlem değerini göstermek üzere,

$$I = \frac{x_i}{x_0} . 100$$

biçiminde formüle edilir.

ÖRNEK 1

Aşağıda, belirli bir bölgedeki buğday üretimi yıllara göre verilmiştir. 1990 yılı değerini temel alarak indeks değerlerini hesaplayınız. 1993, 1994 ve 1997 yıllarına ilişkin indeks değerlerini yorumlayınız.

Yıllar	Buğday Üretimi (1000 ton)
1990	450
1991	375
1992	400
1993	525
1994	450
1995	400
1996	325
1997	300
1998	375
1999	405

ÇÖZÜM

1990 yılı değeri temel kabul edildiğinden, (1990 = 100) olarak gösterilir ve her yılın değeri temel değere bölünüp, 100 ile çarpılarak, indeks değerleri hesaplanır.
İlgili işlemler aşağıda gösterilmiştir:

Yıllar	**Buğday Üretimi** (1000 ton)	**İndeks** (1990 = 100)
1990	450	(450 / 450) . 100 = 100.00
1991	375	(375 / 450) . 100 = 83.33
1992	400	(400 / 450) . 100 = 88.89
1993	525	(525 / 450) . 100 = 116.67
1994	450	(450 / 450) . 100 = 100.00
1995	400	(400 / 450) . 100 = 88.89
1996	325	(325 / 450) . 100 = 72.22
1997	300	(300 / 450) . 100 = 66.67
1998	375	(375 / 450) . 100 = 83.33
1999	405	(405 / 450) . 100 = 90.00

Bu sonuçlar, 1993 yılı buğday üretiminde 1990 yılına göre % 16.67'lik bir artışın, 1994 yılındaysa 1990 yılına göre herhangi bir artış ya da azalışın olmadığını ve 1997 yılındaysa 1990 yılına göre % 33.33'lük bir azalmanın olduğunu ifade etmektedir.

Kuramda, indeks türüne ve kapsamına uygun farklı yöntemler geliştirilmiştir. Kolaylık açısından, hesaplama yöntemleri indeks türleriyle birlikte ele alınacaktır.
Bu ünitede indeksler, mekan ve zaman indeksleri ve basit ve bileşik indeksler olarak iki ana başlık altında ele alınacaktır.

Mekan ve Zaman İndeksleri

Ana çizgileriyle, eğer, indeksler bir mekan serisinden hareketle hesaplanıyorsa bu tür indekslere "mekan", bir zaman serisine dayandırılıyorsa bu tür indekslere de "zaman indeksleri" adı verilir.
İzleyen bölümlerde mekan ve zaman indeksleri ayrıntılarıyla ele alınacaktır.

Mekan İndeksleri

Üretim ve fiyat gibi olaylara ilişkin değerlerin bölgeler, şehirler, kasabalar, köyler vb. gibi bir mekan içindeki oransal değişimlerin ölçüsüne, "mekan indeksi" adı verilir.
Mekan indekslerinin hesabında temel değer olarak, seriyi oluşturan değerlerin aritmetik ortalaması alınır. İndeks sayıları seriyi oluşturan değerlerin aritmetik ortalamaya bölünerek 100 ile çarpılması suretiyle elde edilir. Başka bir anlatımla, mekan indeksleri

Mekan indeksleri ile verilen mekan serisinin aritmetik ortalamaya göre değişimleri araştırılır.

$$I = \frac{x_i}{\bar{x}} \, . \, 100$$

eşitliği ile hesaplanır. Böylece, seriyi oluşturan değerlerin aritmetik ortalamaya göre değişimleri belirlenir.

ÖRNEK 2

Aşağıda, 6 yerleşim bölgesi için 2000 yılı Ocak ayına ilişkin bir X maddesinin fiyatları verilmiştir. Mekan indeksini hesaplayınız ve ilk ikisiyle ilişkin sonuçları yorumlayınız.

İller	Fiyatlar (milyon TL)
Ankara	69
Antalya	75
Kırklareli	70
Eskişehir	79
Konya	75

ÇÖZÜM

Öncelikle, verilen serinin aritmetik ortalaması hesaplanır, sonra da gözlem değerleri aritmetik ortalamaya bölünerek, indeks sayıları elde edilir.

İller	Fiyatlar (milyon TL) x	İndeks
Ankara	69	(69 / 73,6) . 100 = 93.75
Antalya	75	(75 / 73,6) . 100 = 101.90
Kırklareli	70	(70 / 73,6) . 100 = 95,10
Eskişehir	79	(79 / 73,6) . 100 = 107.33
Konya	75	(75 / 73,6) . 100 = 101.90

$\bar{x} = 73{,}6$

Elde edilen sonuçlara göre, Ankara'da X maddesinin fiyatları verilen 6 ilin ortalama fiyatına göre % 6.25 (100 – 93.75 = 6.25) daha düşük, Antalya'daysa % 1.9 daha yüksektir.

Zaman İndeksleri

Üretim ve fiyat gibi istatistiksel olaylara ilişkin değerlerin yıl, ay, hafta, gün vb. gibi, zaman içindeki oransal değişimlerinin ölçüsüne, "zaman indeksi" adı verilir.

Zaman indeksleri ile verilen zaman serisinin zaman içindeki oransal değişimleri araştırılır.

Zaman indeksleri, uygulamada en çok kullanılan indeks türüdür.

Zaman indeksleri, sabit ve değişken esaslı (zincirleme) indeksler olarak iki alt başlık altında toplanabilir.

Sabit Esaslı İndeksler

Sabit esaslı zaman indekslerinin hesabında temel prensip, devrelerden birinin değeri temel kabul edilerek, diğer devrelerin kıymetlerinin, seçilen temel devre kıymetinin yüzdesi olarak ifade edilmesidir.

Gözlem değerleri x_i, temel devre değeri de x_0 ile gösterilirse sabit esaslı indeks,

$$I = \frac{x_i}{x_0} \cdot 100$$

eşitliğiyle hesaplanır.

ÖRNEK 3 ***Turistik bir bölgede, önemli bir müzeyi ziyaret eden turist sayıları, aşağıda verilmiştir. 1996 yılı değerini temel kabul ederek, sabit esaslı indeks sayılarını hesaplayınız ve 1997 ile 2001 yılı indeks sayılarını yorumlayınız.***

Yıllar	Turist Sayısı (000)
1996	16
1997	15
1998	16
1999	17
2000	14
2001	20

ÇÖZÜM

İlgili hesaplamalar aşağıdaki gibidir:

Yıllar	Turist Sayısı (000)	İndeks
1996	16	(16 / 16) . 100 = 100.00
1997	15	(15 / 16) . 100 = 93.75
1998	16	(16 / 16) . 100 = 100.00
1999	17	(17 / 16) . 100 = 106.25
2000	14	(14 / 16) . 100 = 87.50
2001	20	(20 / 16) . 100 = 125.00

1997 yılında, 1996 yılına göre, müzeyi ziyaret eden turist sayısı (100 - 93.75 = 6.25), % 6.25 oranında azalmış, 2001 yılındaysa % 25 oranında artmıştır.

Değişken Esaslı (Zincirleme) İndeksler

Eğer indeks hesabında her değer, bir önceki dönemin değeriyle karşılaştırılmak istenirse, oluşturulan indekslere "değişken esaslı" ya da "zincirleme indeks" adı verilir.

x_i, i. gözlem değerini gösteren değişken esaslı indeks,

$$I = \frac{x_i}{x_{i-1}} \cdot 100$$

eşitliği ile hesaplanır.

ÖRNEK 4 ***Örnek 3'teki verileri kullanarak değişken esaslı indeks sayılarını hesaplayınız ve 1997 ile 1998 yıllarına ilişkin sonuçları yorumlayınız.***

ÇÖZÜM

İlgili işlemler aşağıdaki gibidir:

Yıllar	Turist Sayısı (000) x_i	İndeks
1996	16	–
1997	15	(15 / 16) . 100 = 93.75
1998	16	(16 / 15) . 100 = 106.66
1999	17	(17 / 16) . 100 = 106.25
2000	14	(14 / 17) . 100 = 82.35
2001	20	(20 / 14) . 100 = 142.85

Sabit ve değişken esaslı indeksler, zaman serileri için hesaplanabilir.

1996 yılı için 1995 yılına ilişkin değerler bilinemediğinden değişken esaslı indeks hesaplanamaz. Bu nedenle tabloda yeri boş bırakılmıştır.
1997 yılında müzeyi ziyaret eden turist sayısı 1996 yılına göre %6.25 oranında azalmış, 1998 yılındaysa 1997 yılına göre % 6.66 oranında artmıştır.

Basit ve Bileşik İndeksler

İndeksler kapsadıkları maddelere göre dikkate alındığında basit ve bileşik indeksler olmak üzere iki grupta incelenebilirler.

İzleyen bölümde basit ve bileşik indeksler ayrıntılarıyla ele alınmıştır.

Basit İndeksler

Tek bir madde için hesaplanan indekse basit indeks denir.

Tek bir maddeyi kapsayan indekslere “basit indeksler” adı verilir. Basit indekslerle, bir maddeye ilişkin fiyat ya da miktardaki oransal değişimler araştırılır.

Basit indeks, ilgili maddenin fiyatındaki oransal değişimlerin hesaplanması amacıyla oluşturuluyorsa “basit fiyat indeksi”, miktarındaki oransal değişimlerin belirlenmesi amacıyla oluşturuluyorsa “basit miktar indeksi” adını alır.

p_0 temel devre fiyatı ve p_1 de i. devre fiyatı olmak üzere basit fiyat indeksi,

$$I = \frac{p_i}{p_0} . 100 ,$$

q_0 temel devre miktarı ve q_i de i. devre miktarı olmak üzere basit miktar indeksi,

$$I = \frac{q_i}{q_0} . 100$$

eşitlikleriyle hesaplanır.

ÖRNEK 5

Belirli bir maddenin 2000 yılı fiyatı 175.000 TL ve 2001 yılı fiyatı ise 225.000 TL olsun. 2000 yılı fiyatını temel kabul ederek, adı geçen maddenin 2001 yılındaki fiyat artış oranını bulunuz.

ÇÖZÜM

İlgili işlemler aşağıda gösterilmiştir:

2000	2001
p_0	p_1
175.000	225.000

$$I = \frac{p_1}{p_0} = \frac{225.000}{175.000} \,.\, 100 = \%\ 128.57$$

İlgili maddenin fiyatı, 2001 yılında 2000 yılına göre % 28.57 oranında artış göstermiştir.

ÖRNEK 6

Bir A bölgesinde narenciye üretimi 2000 yılında 2500 ton, 2001 yılında ise 2400 ton olarak gerçekleşmiştir. 2000 yılına göre 2001 yılındaki narenciye üretimindeki düşüş oranını bulunuz.

ÇÖZÜM

İlgili işlemler aşağıda gösterilmiştir:

2000	2001
q_0	q_1
2.500	2.400

$$I = \frac{q_1}{q_0} = \frac{2400}{2500} \,.\, 100 = \%96$$

2001 yılında A bölgesinde narenciye üretiminde 2000 yılına göre %4 oranında düşüş kaydedilmiştir.

Bileşik İndeksler

İki ya da daha çok maddeyi kapsayan indekslere “bileşik indeks” adı verilir. Bileşik indeksle, indeksin kapsadığı maddelere ilişkin fiyat ya da miktarların zaman içindeki oransal değişimleri araştırılır.

İki ya da daha çok maddeyi kapsayan indekslere bileşik indeksler adı verilir.

Bileşik indeks hesabında kullanılan teknikler, “basit toplam indeks”, “basit indekslerin tartısız aritmetik ortalaması” ve “basit indekslerin tartılı aritmetik ortalaması (ya da tartılı toplam indeks)” olmak üzere üç ana başlık altında toplanır.

Sözü edilen teknikler aşağıda ayrıntılarıyla ele alınmıştır.

Basit Toplam İndeks

Basit toplam indeks hesaplanırken, indekse girecek maddelerin indeksi hesaplanarak (cari) devredeki fiyatları toplanır ve temel devre fiyatlar toplamına bölünerek, sonuç 100 ile çarpılır.

Temel devre fiyatları toplamı $\sum p_0$ ve cari devre fiyatları toplamı da $\sum p_1$ ile gösterilirse, indeks,

$$I = \frac{\sum p_1}{\sum p_0} . 100$$

eşitliğiyle hesaplanır.

ÖRNEK 7

Aşağıda 4 maddenin 2001 ve 2002 yılları Nisan ayı kg fiyatları verilmiştir. 2001 yılına ilişkin basit toplam indeksi hesaplayınız.

Maddeler	**Fiyatlar**	
	2001	**2002**
	p_0	**p_1**
Beyaz Peynir	2.725.660	3.831.389
Süt	724.348	1.026.073
Yoğurt	1.373.302	1.531.306
Tereyağı	6.751.502	8.553.892

Kaynak: *D.İ.E. Aylık İstatistik Bülteni, Haziran, 2002.*

ÇÖZÜM

Maddeler	**Fiyatlar**	
	2001	**2002**
	p_0	**p_1**
Beyaz Peynir	2.725.660	3.831.389
Süt	724.348	1.026.073
Yoğurt	1.373.302	1.531.306
Tereyağı	6.751.502	8.553.892
	$\sum p_0$ = 11.574.812	$\sum p_1$ = 14.942.660

$$I = \frac{\sum p_1}{\sum p_0} . 100$$

$$= \frac{14.942.660}{11.574.812} . 100$$

$$= 129,096$$

2002 Nisan ayında 2001 yılı Nisan ayına göre verilen 4 maddenin fiyatı, ortalama olarak % 29,096 oranında artış göstermiştir.

Basit İndekslerin Tartısız Aritmetik Ortalaması

Bu teknikte cari yıldaki fiyatlar, temel devre fiyatlarına bölünür ve aritmetik ortalama hesaplanır. Elde edilen sonuç 100 ile çarpılarak, indeks sayısı elde edilir. Başka bir anlatımla, bu teknikte temel yıla göre her madde için hesaplanan basit indekslerin aritmetik ortalaması 100 ile çarpılır.

$$I = \frac{\sum\left(\frac{p_1}{p_0}\right)}{n} \cdot 100$$

ÖRNEK 8

Örnek 7'de verilen 4 madde için 2002 yılına ilişkin basit indekslerin tartısız aritmetik ortalamasını hesaplayınız.

ÇÖZÜM

Maddeler	Fiyatlar 2001 p_0	2002 p_1	p_1/p_0
Beyaz Peynir	2.725.660	3.831.389	(3.831.389 / 2.725.660) = 1.41
Süt	724.348	1.026.073	(1.026.073 / 724.348) = 1.42
Yoğurt	1.373.302	1.531.306	(1.531.306 / 1.373.302) = 1.12
Tereyağı	6.751.502	8.553.892	(8.553.892 / 6.751.502) = 1.27
	$\sum p_0 = 11.574.812$	$\sum p_1 = 14.942.660$	$\sum(p_1 / p_0) = 5.22$

$$I = \frac{\sum\left(\frac{p_1}{p_0}\right)}{n} \cdot 100$$

$$= \frac{100}{4}(5.22)$$

$$= 130.5$$

olarak elde edilir.

Elde edilen sonuca göre verilen 4 maddenin fiyatı, 2002 yılı Nisan ayında 2001 yılı Nisan ayına göre ortalama olarak % 30.5 oranında artmıştır.

Basit İndekslerin Tartılı Aritmetik Ortalaması

Eğer bir bileşik fiyat indeksinin kapsadığı maddelerin fiyatları arasında önem derecesi açısından farklar söz konusuysa ve indeks hesabında bu farkların da göz önüne alınması istenirse, bu durumda indekslerin tartılı olarak hesaplanması gerekir.

Tartı, ilgili maddelerin oransal fiyatı için kullanılır. t tartıyı gösterirken basit indekslerin tartılı ortalaması,

$$I = \frac{\sum\left(\frac{p_1}{p_0}\right) \cdot t}{\sum t} \cdot 100$$

eşitliği ile hesaplanır.

Uygulamada kullanılan tartılar indeksin türüne göre değişir. Genel olarak tartılar, ilgili maddelerin üretilen ya da tüketilen miktarıyla, fiyatının çarpılması sonucunda elde edilir. Ancak, tartılar oluşturulurken "temel yıl fiyatı" esas alınır. Başka bir ifadeyle,

$$t = p_0 \cdot q$$

olur.

Yukarıdaki eşitliğe dikkat edilirse, tartı hesabında p_0 temel devre fiyatıdır. Ancak q'nun hangi devreye ait olduğu belirtilmemiştir. Tartı hesaplanırken, q yerine q_0 ya da q_1 değerlerinden birisi konabilir. Açıktır ki, bu durumda farklı sonuçlar elde edilecektir. Başka bir anlatımla, iki farklı indeks hesaplanacaktır.

Sözü edilen indeksler aşağıda ayrı ayrı ele alınacaktır.

Temel Devre Miktarlarının Kullanılması (Laspeyres İndeksi)

$$t = p_0 \cdot q$$

eşitliğinde, temel devre miktarı olarak q_0 alınırsa,

$$t = p_0 \cdot q_0$$

olacaktır. Bu tartı, basit indekslerin tartılı aritmetik ortalaması formülünde yerine konursa,

$$I = \frac{\sum \left(\frac{p_1}{p_0}\right) \cdot t}{\sum t} \cdot 100$$

$$= \frac{\sum \left(\frac{p_1}{p_0}\right) \cdot p_0 \cdot q_0}{\sum p_0 \cdot q_0} \cdot 100$$

olur ve gerekli kısaltmalar yapılırsa,

$$I = \frac{\sum p_1 \cdot q_0}{\sum p_0 \cdot q_0} \cdot 100$$

sonucu elde edilir.

Bu indekse "Laspeyres (Lasper okunur) indeksi" adı verilir.

ÖRNEK 9

Aşağıda 4 maddenin belirli bir bölgedeki 2000 ve 2001 yıllarına ilişkin fiyat ve miktarları verilmiştir. 2000 yılını temel devre kabul ederek, 2001 yılı için Laspeyres fiyat indeksini hesaplayınız ve elde ettiğiniz sonucu yorumlayınız.

Maddeler	2000		2001	
	Fiyat (TL)	Miktar (ton)	Fiyat (TL)	Miktar (ton)
Portakal	700.000	125	850.000	145
Mandalina	850.000	130	950.000	155
Elma	500.000	150	700.000	130
Armut	650.000	100	800.000	135

ÇÖZÜM

İlgili işlemler aşağıdaki gibidir:

Maddeler	2000		2001			
	Fiyat (TL)	**Miktar** (ton)	**Fiyat** (TL)	**Miktar** (ton)		
	p_0	q_0	p_1	q_1	$p_1 \cdot q_0$ (000)	$p_0 \cdot q_0$ (000)
Portakal	700.000	125	850.000	145	106.250	87.500
Mandalina	850.000	130	950.000	155	123.500	110.500
Elma	500.000	150	700.000	130	105.000	75.000
Armut	650.000	100	800.000	135	80.000	65.000
					414.750	**338.000**

$$I = \frac{\sum p_1 \cdot q_0}{\sum p_0 \cdot q_0} \cdot 100$$

$$= \frac{414.750}{338.000} \cdot 100$$

$$= 122.70$$

2001 yılı için verilen 4 maddenin Laspeyres fiyat indeksi 122.70'tir. 2001 yılında verilen 4 maddenin fiyatında 2000 yılına göre % 22.70'lik bir artış gerçekleşmiştir.

İndeks Devresi Miktarlarının Kullanılması (Paasche İndeksi)

Eğer $t = p_0 \cdot q$ eşitliğinde q indeks devresi miktarı olarak alınırsa,

$$t = p_0 \cdot q_1$$

olur. Bu ifade basit indekslerin tartılı aritmetik ortalaması formülünde yerine konulursa,

$$I = \frac{\sum \left(\frac{p_1}{p_0}\right) \cdot t}{\sum t} \cdot 100$$

$$= \frac{\sum \left(\frac{p_1}{p_0}\right) \cdot p_0 \cdot q_1}{\sum p_0 \cdot q_1} \cdot 100$$

olur ve gerekli kısaltmalar yapılırsa,

$$I = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_1} \cdot 100$$

olarak elde edilir.

Bu indekse "Paasche (Paaşe okunur) indeksi" adı verilir.

ÖRNEK 10

Örnek 9'da verilen 4 madde için Paasche fiyat indeksini hesaplayınız ve elde ettiğiniz sonucu yorumlayınız.

ÇÖZÜM

Çözüme ilişkin hesaplamalar aşağıdaki gibidir:

Maddeler	2000		2001			
	Fiyat (TL)	**Miktar** (ton)	**Fiyat** (TL)	**Miktar** (ton)		
	p_0	q_0	p_1	q_1	$p_1 \cdot q_1$ (000)	$p_0 \cdot q_1$ (000)
Portakal	700.000	125	850.000	145	123.250	101.500
Mandalina	850.000	130	950.000	155	147.250	131.750
Elma	500.000	150	700.000	130	91.000	65.000
Armut	650.000	100	800.000	135	108.000	87.750
					469.500	**386.000**

$$I = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_1} \cdot 100$$

$$= \frac{469.500}{386.000} \cdot 100$$

$$= 121.63$$

2001 yılında verilen 4 maddenin fiyatları 2000 yılına göre % 21.63 oranında yükselmiştir.

Fisher İndeksi

Laspeyres ve Paasche indeksleri, kullanılan farklı tartılar nedeniyle, doğal olarak, farklı sonuçlar verir. Söz konusu farklar; Laspeyres indeksinde tartı olarak temel devre miktarı kullanılması nedeniyle, fiyat artışlarını, olduğundan fazla, Paasche indeksi de fiyat artışlarını olduğundan az gösterir.

Fisher, fiyatlardaki artış ya da azalışların gerçeklere daha yakın hesaplanabilmesi için, Laspeyres ve Paasche indekslerinin geometrik ortalamasının alınmasını önermiştir. Bu teknikle hesaplanan indekse Fisher indeksi adı verilir.

Fisher indeksi, Laspeyres ve Paasche indekslerinin geometrik ortalaması alınarak hesaplanır.

Fisher indeksi,

$$I = \sqrt{\frac{\sum p_1 \cdot q_0}{\sum p_0 \cdot q_0} \cdot \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_1}} \cdot 100$$

eşitliğiyle hesaplanır.

Fisher indeksine "İdeal indeks" adı da verilir.

ÖRNEK 11

Örnek 9'da verilen 4 madde için Fisher indeksini hesaplayınız ve elde ettiğiniz sonucu yorumlayınız.

ÇÖZÜM

Çözüme ilişkin hesaplamalar aşağıdaki gibidir:

Maddeler	p_0	q_0	p_1	q_1	$p_1 \cdot q_0$ (000)	$p_0 \cdot q_0$ (000)	$p_1 \cdot q_1$ (000)	$p_0 \cdot q_1$ (000)
Portakal	700.000	125	850.000	145	106.250	87.500	123.250	101.500
Mandalina	850.000	130	950.000	155	123.500	110.500	147.250	131.750
Elma	500.000	150	700.000	130	105.000	75.000	91.000	65.000
Armut	650.000	100	800.000	135	80.000	65.000	108.000	87.750
					414.750	**338.000**	**469.500**	**386.000**

$$I = \sqrt{\frac{\sum p_1 \cdot q_0}{\sum p_0 \cdot q_0} \cdot \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_1}} \cdot 100$$

$$= \sqrt{\frac{414.750}{338.000} \cdot \frac{469.500}{386.000}} \cdot 100$$

$$= \sqrt{(1.2270) \cdot (1.2163)} \cdot 100$$

$$= 122.16$$

2001 yılında verilen 4 maddenin fiyatı 2000 yılına göre, %22.16 oranında yükselmiştir.

Dikkat edilecek olursa, Laspeyres, Paasche ve Fisher indekslerinin hesaplanmasında aynı veriler kullanılmıştır. Sonuçlar karşılaştırılacak olursa artış, Laspeyres indeksi için %22.70, Paasche indeksi için %21.63 ve Fisher indeksi için %22.16 bulunmuştur. Görüleceği gibi, Laspeyres ve Paasche indekslerinin geometrik ortalaması olan Fisher indeksine ilişkin indeks sayısı, Paasche ve Laspeyres indeks sayılarının arasında yer almıştır.

Bilindiği gibi, indeksler, temel kabul edilen bir yıla göre hesaplanırlar. Ancak uygulamada, temel yıl olarak rasgele bir yıl seçilmemelidir. Temel yıl, ekonomik özellikler açısından fazla hareketli olmayan (normal) bir yıl olmalıdır. Aksi durumda gerçekleri iyi yansıtmayan sonuçlar elde edilecektir. Örneğin; fiyatların, herhangi bir nedenle yüksek olduğu bir yıl, temel yıl olarak seçilirse, buna göre hesaplanan fiyat indekslerinde fiyatlar düşüyormuş gibi, aksi durumdaysa yükseliyormuş gibi görünecektir. Bu nedenle, harp ve kriz yılları, indeks hesabında temel yıl olarak kullanılmamalıdır.

Temel yıl zaman içerisinde değiştirilebilir. Gerçekte temel yıl eskidikçe, indeks sayıları büyür ve karşılaştırmalar da giderek zorlaşır.

SIRA SİZDE

1. İndeks kavramını açıklayınız.

2. Aşağıda verilen seri için değişken esaslı indeksi hesaplayarak 1987 ve 1991 yıllarına ilişkin indeks değerlerini yorumlayınız.

Yıllar	X
1985	30
1986	40
1987	25
1988	34
1989	27
1990	30
1991	34
1992	37

3. Aşağıda 3 maddenin 2001 ve 2002 yıllarına ilişkin fiyat ve miktarları verilmiştir. 2001 yılını temel alarak 2002 yılına ilişkin Paasche, Laspeyres ve Fisher indekslerini hesaplayınız. Hangi indeksin gerçeğe daha yakın sonuç verdiğini nedenleriyle açıklayınız.

Maddeler	2001		2002	
	Fiyat (TL)	**Miktar** (ton)	**Fiyat** (TL)	**Miktar** (ton)
Pirinç	1.362.170	750	1.935.186	600
Makarna	530.109	800	1.029.163	850
Bulgur	675.301	400	1.119.285	450

Kendimizi Sınayalım

1. 1994 yılında bir X maddesinin kilosu 650.000 TL ve 1998 deyse 975.000 TL'dir. 1994 yılı temel devre olmak üzere, 1998'deki fiyat artış oranı aşağıdakilerden hangisidir?

a. 110
b. 140
c. 150
d. 155
e. 160

2. Aşağıdaki tabloda 3 maddenin 1998 ve 2000 yılı fiyatları verilmiştir.

Maddeler	Fiyatlar	
	1998	2000
A	500	750
B	450	680
C	600	900

1998 yılını temel yıl kabul ederek, 2000 yılına ilişkin basit toplam indeks değeri aşağıdakilerden hangisidir?

a. 114.22
b. 132.50
c. 141.10
d. 150.32
e. 154.33

3. Aşağıdaki tabloda iki maddenin 2000 ve 2001 yıllarına ilişkin fiyat ve miktarları verilmiştir.

	2000		2001	
Maddeler	**Fiyat** (TL)	**Miktar** (ton)	**Fiyat** (TL)	**Miktar** (ton)
Portakal	575	10	625	11
Mandalina	645	15	715	17

Yukarıdaki tabloya göre 2000 yılı temel yıl kabul edildiğinde, 2001 yılına ilişkin Laspeyres fiyat indeksinin değeri aşağıdakilerden hangisidir?

a. 90.85
b. 92.15
c. 110.04
d. 110.45
e. 112.09

4. Aşağıdaki tabloda iki maddenin 2000 ve 2001 yıllarına ilişkin fiyat ve miktarları verilmiştir.

	2000		2001	
Maddeler	**Fiyat** (TL)	**Miktar** (ton)	**Fiyat** (TL)	**Miktar** (ton)
Patates	10.000	5	15.000	5
Soğan	12.000	6	18.000	8

Yukarıdaki tabloya göre 2000 yılı temel kabul edilerek hesaplanacak Paasche fiyat indeksinin değeri aşağıdakilerden hangisidir?

a. 110.40
b. 122.18
c. 132.40
d. 150.00
e. 157.14

5. Aşağıdaki tabloda bir A maddesine ilişkin 5 yıllık fiyatlar verilmiştir.

Yıllar	**Fiyat** (TL)
1995	450.000
1996	500.000
1997	517.000
1998	522.000
1999	530.000

Yukarıdaki tabloya göre, 1997 yılı değişken esaslı indeks değeri aşağıdakilerden hangisidir?

a. 103.40
b. 105.42
c. 108.34
d. 110.04
e. 112.43

6. Beş maddeye ve istenilen bir yıla ilişkin indeksler hesaplanmak istenmektedir. Bu amaçla yapılan bazı hesaplama sonuçları aşağıda verilmiştir.

$\sum p_1 \cdot q_0 = 4.784.000$

$\sum p_0 \cdot q_0 = 3.853.000$

$\sum p_1 \cdot q_1 = 4.560.000$

$\sum p_0 \cdot q_1 = 3.760.000$

Bu sonuçlara göre, Paasche indeks değeri aşağıdakilerden hangisidir?

a. 114,4116
b. 121,2766
c. 124,1312
d. 125,1810
e. 127,1419

7. Üç maddeye ve istenilen bir yıla ilişkin Laspeyres fiyat indeksi değeri 122.13 ve Paasche fiyat indeksi değeri ise 124.17 olarak hesaplanmıştır. Fisher fiyat indeks değeri, aşağıdakilerden hangisidir?

a. 122.18
b. 122.43
c. 122.56
d. 123.15
e. 124.03

8. 6 maddeye ve istenilen bir yıla ilişkin Laspeyres fiyat indeksi 140.18, Fisher fiyat indeksiyse 142.78578 olarak hesaplanmıştır. Paasche fiyat indeksi aşağıdakilerden hangisidir?

a. 127.15000
b. 132.18754
c. 140.14075
d. 141.13715
e. 145.44859

9. 4 maddeye ve istenilen bir yıla ilişkin Paasche fiyat indeksi 120.15, Fisher fiyat indeksiyse 124.25 olarak hesaplanmıştır. Laspeyres fiyat indeksi aşağıdakilerden hangisidir?

a. 123.1000
b. 125.1476
c. 128.4899
d. 129.1411
e. 130.1541

10. Dört maddeye ilişkin bazı hesaplama sonuçları aşağıda verilmiştir.

$\sum p_0 \cdot q_0 = 710$

$\sum p_1 \cdot q_0 = 838$

$\sum p_1 \cdot q_1 = 738$

$\sum p_0 \cdot q_1 = 630$

Bu sonuçlara göre, Fisher fiyat indeksinin değeri aşağıdakilerden hangisidir?

a. 115.42143
b. 117.58467
c. 118.41100
d. 120.13451
e. 121.42138

Yanıt Anahtarı

1. c
2. d
3. c
4. d
5. a
6. b
7. d
8. e
9. c
10. b

Yararlanılan Kaynaklar

BOWEN, Earl K., STARR Martin K.: **Basic Statistics for Business and Economics**, McGraw-Hill Inc., 1994.

GÜRTAN, Kenan: **İstatistik ve Araştırma Metodları**, İstanbul Üniversitesi Yayınları, No 2265, İstanbul, 1977.

HARNETT, Donald L.: **Statistical Methods**, 3rd Edition, Addison Wesley Pub. Comp., London, 1982.

McCLAVE James T., BENSON P.George, SINCICH Terry: **Statistics for Business and Economics**, 7th Edition, Prentice-Hall International Inc., 1998.

14

Zaman Serisi Çözümlemesi

Çalışma Biçimine İlişkin Olarak

- *Verilen tanımlar iyice özümsenmeli,*
- *Örnek sorular ve çözümleme süreci dikkatle incelenmeli,*
- *Kavramlar arasındaki farklılıklar belirlenmeli,*
- *İstenenlerin neler olduğu net bir biçimde ortaya konulmalıdır.*

Amaçlar

- *Zaman serilerini grafikle gösterebileceksiniz.*
- *Zaman serilerini etkileyen temel ve yanıltıcı faktörleri açıklayabileceksiniz.*
- *Zaman serisi çözümlemesi kavramını açıklayabilecek ve hareketli ortalamalar hesaplayabileceksiniz.*
- *Mevsimsel olmayan ve mevsimsel serilerin, betimsel ve öngörü amaçlarıyla çözümlenmesinde, bileşenlerine ayırma modellerini uygulayabileceksiniz.*

İçerik Haritası

GİRİŞ

İktisatçılar ve iş idarecileri, bir mal ya da hizmetin üretilmesi, pazarlanması ve müşteri memnuniyetinin araştırılması aşamalarında, zaman içinde değişen değerler alan pek çok değişkenle ilgilenmek zorunda kalırlar. Bir işletmenin satış tutarı, likidite düzeyi, işgücü devir hızı, bir ülkeye gelen yabancı turist sayısı, bir ülkenin enflasyon düzeyi ve benzerleri, bu tür değişkenlere örnek olarak verilebilir. Anımsanacağı gibi, zamanla ilişkili olarak tanımlanan bu değişkenler hakkında yapılan ölçümleri, zamana göre sıralanmış olarak gösteren serilere "zaman serisi" denir.

Ekonomik, sosyal, psikolojik vb. çeşitli nedenlerin, zamanla ilişkili değişkenler üzerindeki etkisi, yön ve şiddetinin farklı olması nedeniyle, zaman serisi gözlem değerlerinde bazı değişmeler gözlenir. Bu değişmeler zaman serilerini etkileyen faktörler ya da bileşenler olarak ifade edilirler ve Trend Bileşeni, Mevsimsel Bileşen, Konjonktürel Bileşen ve Rassal Bileşen olarak isimlendirilirler.

Zaman serileri, yukarıda sözü edilen bu değişmeler nedeniyle, çözümlenmeden bir anlam ifade etmezler. En basit anlamda, zaman serisi çözümlemesi, serinin özelliklerini açıklamaktır. Zaman serilerinin özelliklerini açıklayabilmek için, bu serilerin rassal bileşenin yanında, diğer üç bileşenden hangi(ler)inin etkisinde olduğunu belirlemek ve bu bileşenlerin etkilerini tahminlemek gerekir. Zaman serisi çözümlemesinin ikinci, belki de en önemli amacı öngörü yapmaktır. Öngörü amacıyla zaman serisi çözümlemesi; seriyi etkileyen bileşenlerin, belirlenmesi belirlenen bileşenlerinin etkilerinin tahminlenmesi ve hesaplanan tahminlerden yararlanarak, serinin gelecek dönemlerine ilişkin öngörü çalışmalarını kapsar.

Bu ünitede önce zaman serisi ve zaman serisi çözümlemesiyle ilgili tanım ve açıklamalar yapılacak, daha sonra da zaman serilerinin açıklanması amacıyla çözümlemelerde kullanılan "Bileşenlere Ayırma Yöntemi" tanıtılacak ve örnek uygulamalara yer verilecektir.

ZAMAN SERİSİNİN TANIMI VE GRAFİKLE GÖSTERİLMESİ

AMAÇ 1 *Zaman serilerini grafikle gösterebileceksiniz.*

Zaman Serisi Tanımı

Zaman değişkeniyle ilişkili bir değişken hakkında, elde edilen gözlem değerlerini zamana göre sıralanmış olarak gösteren serilere, "zaman serisi" denir. Bu tanım genel bir tanımdır. Zaman serilerini konu alan pek çok çalışmada, serilerin gözlem değerlerinin eşit aralıklı zaman noktalarında elde edilmiş olduğu görülmektedir. Eşit aralıklı zaman noktaları (başka bir ifadeyle zaman değişkeninin şıkları), günler (günlük hava sıcaklığında olduğu gibi), aylar (aylık satış miktarlarında olduğu gibi) ve yıllar (yıllık ihracat tutarlarında olduğu gibi) olabilir. Zaman serisi çözümlemelerinde zaman değişkeninin şıkları genellikle $t = 1, 2,...., n$ ile ifade edilmektedir. Buna göre bir zaman serisi, eşit aralıklı $t = 1, 2,...., n$ zaman noktalarında Y değişkeniyle ilgili elde edilen $y_1, y_2, ..., y_t , ... y_n$ gözlem değerlerini zamana göre sıralanmış olarak gösteren seri olarak tanımlanır. Kuramda kesikli zaman serisi tanımı olarak bilinen bu tanım, bu ünitede zaman serisi tanımı olarak benimsenmiştir.

Zaman serileri gözlem değerlerinin elde edilmesinde benimsenen yaklaşım, zaman serilerinin zaman değişkeninin şıklarına göre isimlendirilmesidir. Örneğin gözlem değerleri zaman değişkeninin ay şıkkına göre elde edilmişse "aylık zaman serisi" (Tablo 14.1'de gösterildiği gibi), zaman değişkeninin yıl şıkkına göre elde edilmişse "yıllık zaman serisi"adı verilir.

Tablo 14.1
Türkiye'nin (1989-2001 dönemi) ihracat tutarı yıllık zaman serisi.

Kaynak:
http://tcmbf40.tcmb.gov.tr/cbt.html

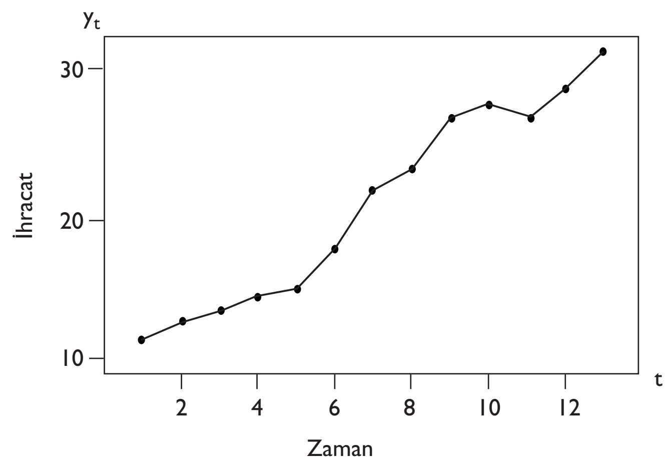
Yıllar	t	y_t İhracat Tutarı (milyon $)	Yıllar	t	y_t İhracat Tutarı (milyon $)
1989	1	11.625	1996	8	23.224
1990	2	12.959	1997	9	26.261
1991	3	13.593	1998	10	26.974
1992	4	14.715	1999	11	26.578
1993	5	15.345	2000	12	27.775
1994	6	18.106	2001	13	31.187
1995	7	21.637			

Zaman serileri iki şekilde oluşturulur:

- İlgilenilen zamana bağlı değişkenin, belirlenen eşit zaman aralıkları itibariyle almış olduğu değerlerin, toplamı ya da ortalaması alınır. Örneğin; Tablo 14.1'deki 1997 yılına ilişkin ihracat tutarı 26.261 milyon ABD Doları, 01 Ocak 1997 - 31 Aralık 1997 zaman döneminde yapılan ihracat tutarlarının toplamını gösterir.
- İlgilenilen zamana bağlı değişkenin belirlenen eşit zaman aralıklarında bir tek ölçümü yapılır. Örneğin; bir bankanın aylık vadesiz mevduat zaman serisini oluşturan gözlem değerleri, ayların son işgünü itibariyle bu bankanın vadesiz mevduat hesabındaki bakiyeleri gösterir.

Zaman Serisinin Grafikle Gösterilmesi

Zaman serileri genel olarak "kartezyen koordinatlı" bir grafikle gösterilir. Grafiğin apsis ekseninde zaman değişkeninin şıkları (ya da bu şıklara karşı gelen kod numaraları), ordinat ekseninde bu şıklar itibariyle Y değişkeninin aldığı değerler, gözlem değerleri y_t yer alır. Belirlenen eşit aralıklı t zaman noktaları (t = 1, 2...., n) ile bu zaman noktalarında zamana bağlı y değişkeninin aldığı y_1, y_2, ..., y_t ,..., y_n gözlem değerlerini eşleştirmek suretiyle kartezyen koordinat sistemi üzerinde işaretlenen noktaların meydana getirdiği şekle, kartezyen grafik ya da serpilme diyagramı adı verilir. Tablo 14.1'de verilen zaman serisinin kartezyen grafiği örnek olarak Şekil 14.1'de gösterilmiştir.



Şekil 14.1
Türkiye'nin (1989-2001 dönemi) ihracat zaman serisinin grafiği.

Kartezyen grafikteki noktaların oluşturduğu şekle bakarak, zaman değişkeniyle ihracat tutarı değişkeni arasındaki ilişkiyi betimlemek mümkündür. Başka bir anlatımla, zaman serisinin kartezyen grafiğini incelemek suretiyle, izleyen kısmında açıklanacak olan, zaman serisini etkileyen temel bileşenlerden hangilerinin, bu seriyi etkilediğini, görsel olarak belirlemek mümkündür.

SIRA SİZDE

1. Zaman serisi nedir?

2. Zaman serileri kaç şekilde oluşturulur?

3. Kartezyen grafik nedir?

ZAMAN SERİLERİNİ ETKİLEYEN FAKTÖRLER

AMAÇ 2

Zaman serilerini etkileyen temel ve yanıltıcı faktörleri açıklayabileceksiniz.

Zaman serilerini etkileyen faktörleri, Temel Faktörler (Bileşenler) ve Yanıltıcı Faktörler başlıkları altında ele almak mümkündür.

Zaman Serisini Etkileyen Temel Faktörler (Bileşenler)

Zaman serilerinin gözlem değerlerinde, zaman içinde azalma ya da artma şeklinde, bazı değişmeler gözlenir. Ekonomik, sosyal, psikolojik vb. çeşitli nedenlerin zaman serisi gözlem değerleri üzerindeki, yön ve şiddetinin farklı olmasından ileri gelen bu değişmeler; Trend (T), Mevsimsel Değişmeler (M), Konjonktürel Değişmeler (K) ve Rassal Değişmeler (R) olarak sayılabilir. Bu değişmelere, genel olarak, Zaman Serisi Bileşenleri ya da temel faktörleri adı verilir.

Bir zaman serisinin bileşenleri; Trend (T), Mevsimsel değişmeler (M), Konjonktürel değişmeler (K) ve Rassal değişmeler (R)'den oluşur.

Zaman serileri üzerinde etkili olabilen bu bileşenlerin her birinin etkilerini araştırmaya geçmeden önce, sözü edilen bileşenlerin özellikleri ele alınacaktır.

Trend Bileşeni: Zaman serisi gözlem değerinin uzun zaman döneminde (en az 7 yıl) artma ya da azalma yönünde gösterdiği genel eğilime "trend" adı verilir. Bu eğilimi açıklayan bileşene de "Trend bileşeni" denir. Trend bileşeni, zamana bağlı değişken üzerindeki genel eğilime neden olan uzun dönemli etkileri açıklar. Bu etkileri genel olarak, demografik özelliklerdeki, coğrafi dağılımdaki, kişi başı-

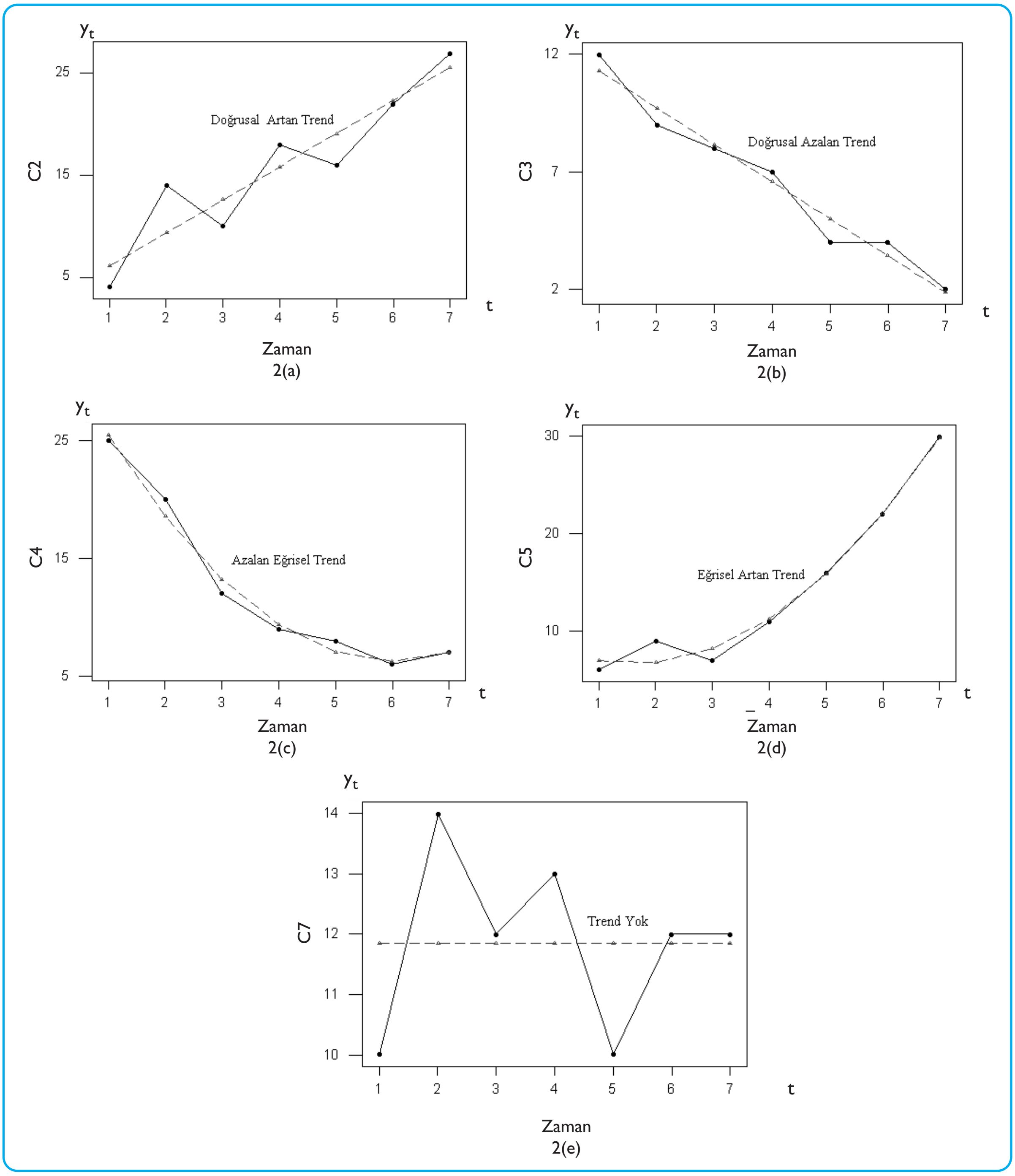
Bir zaman serisinde gözlem değerlerinin uzun dönemde artma ya da azalma yönünde gösterdiği genel eğilime Trend adı verilir.

na gelirdeki, teknolojik gelişmelerdeki, tüketici zevk ve alışkanlıklarındaki değişmelerdeki ve fiyat değişmelerindeki etkiler olarak sıralamak mümkündür.

Yukarıda belirtilen etkilerin şiddetine bağlı olarak, artış ve azalış yönündeki değişmeler, bazan artabilir ya da yavaşlayabilir, yani trend aynı kalmaz. Trend Şekil 14.2a, b, c, d'de gösterildiği gibi doğrusal ya da eğrisel olabilir. Şekil 14.2'de yer alan trend türleri uygulamada sık karşılaşılan türlerdir.

Zaman içinde artış ya da azalış göstermeyen, hemen hemen aynı düzeyde kararlılık gösteren serilerin trendi yoktur. Şekil 14.2e trendinin olmadığı durumu göstermektedir.

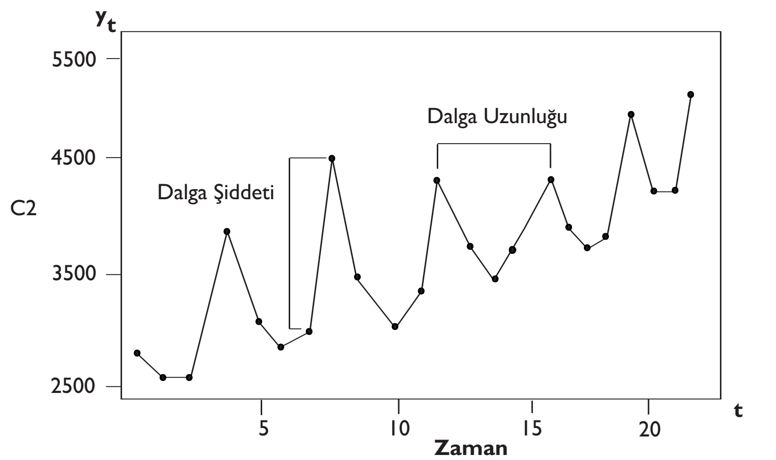
Şekil 14.2 *Çeşitli trend türleri*



Mevsimsel Bileşen : Mevsimsel bileşen birbirini izleyen yılların, mevsimlerin, çeyrek yılların, ayların ya da günlerin aynı zaman noktalarında zaman serisi gözlem değerlerindeki bir artma ve bir azalma şeklindeki düzenli değişmeleri, mevsimsel değişmeleri açıklar. Mevsimsel değişmeler, genellikle iklimle, saatle ya da geleneklerle ilişkilidir. Örneğin; ülkemiz deniz turizmi için önemli bir ülke olduğundan, yaz mevsimlerinde ülkemize gelen yabancı turist sayısı artar, kış mevsimlerinde azalır; İşe gidiş ve işten dönüş saatlerinde şehir içi toplu taşıma araçlarında yolcu sayıları artar, diğer saatlerde azalır; cumartesi ve pazar günleri tatil olduğu için büyük alışveriş merkezlerinin satışları artar, hafta içi günlerde azalır.

Biribirini izleyen yıl, mevsim vb. aynı zaman noktalarında artma ya da azalma şeklindeki düzenli değişmelere, mevsimsel değişmeler adı verilir.

Mevsimsel değişme gösteren C2 zaman serisinin kartezyen grafiği örneği Şekil 14.3'de verilmiştir. Çeyrek yıllık gözlem değerlerinden oluşan bu serinin kartezyen grafiği incelendiğinde, serinin rassal dalgalanmaların yanında trend ve birbirini izleyen yılların IV. çeyreklerinde maksimum, II. çeyreklerinde minimum değerler alma eğilimine sahip olan mevsimsel bileşenin etkisi altında olduğu söylenebilir.



Şekil 14.3 *Çeyrek yıllık bir örnek zaman serisinin kartezyen grafiği.*

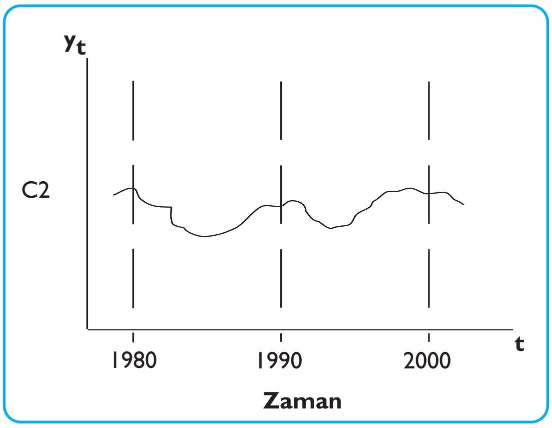
Birbirini izleyen iki mevsimsel değişmenin maksimum noktaları arasındaki zaman aralığına dalga uzunluğu adı verilir ve L simgesiyle gösterilir. Genellikle mevsimsel değişmelerin dalga uzunlukları birbirine eşittir. Şekil 14.3 incelendiğinde mevsimsel değişmelerin dalga uzunluklarının L = 4 çeyrek yıl olduğu görülmektedir. Aylık zaman serilerinde L = 12 aydır.

Mevsimsel değişmeler, dalga uzunluklarının birbirine eşit olması nedeniyle periyodik, tekrar tekrar meydana gelmiş olmaları nedeniyle de döngüsel özelliğe sahip değişmelerdir.

Bir mevsimsel değişmenin maksimum ve minimum noktası arasındaki yükseklik farkına dalga şiddeti adı verilir. Mevsimsel değişmelerin dalga şiddetleri, Şekil 14.3'de görüldüğü gibi, farklı ya da eşit olabilir.

Eğer bir zaman serisindeki mevsimsel değişmelerinin dalga şiddetleri trend etkisinin belirlediği genel eğilimden bağımsızsa, bu serinin, eşit mevsimsel dalga şiddetine sahip mevsimsel değişme gösterdiği , bağımlıyla eşit olmayan dalga şiddetine sahip mevsimsel değişme gösterdiği söylenir.

Mevsimsel değişmelerin dalga uzunluğunun ve dalga şiddetinin doğru olarak belirlenmesi, zaman serisi çözümlemelerinde en önemli konulardandır. Çünkü; bu durum, çözümleme amacıyla kullanılan yöntem türünün ve model tipinin belirlenmesine etki eder. Mevsimsel değişmeler düzenli değişmeler olduğundan, herhangi bir zaman dönemi için etkileri daha kolay tahminlenebilir.



Şekil 14.4
Konjonktürel değişme.

Uzun vadede trend etrafında artma ya da azalma şeklinde tekrarlanan değişmelere, konjonktürel değişmeler adı verilir.

Konjonktürel Bileşen: Ekonomi ve iş idaresi alanlarıyla ilgili değişkenlerde genellikle, sabit bir oranda artış ya da azalış görülmez. Trend düzeyi etrafında, iki ile on yıl ya da daha fazla yıl zaman aralıklarıyla, herhangi bir dönemde, artma ya da azalma şeklinde tekrarlanabilen değişmeler gözlenir. Konjonktürel değişme adı verilen bu değişmelerin etkisini açıklayan bileşene "konjonktürel bileşen" denir. Örneğin, ülkemizde yaşanan Şubat-2001 kriziyle birlikte pek çok ekonomi ve iş idaresiyle ilgili değişken azalan ya da artan değerler alma eğilimine girmiştir. Yatırımlar azalmış, hatta durmuş, üretim azalmış, dolayısıyla gelirler azalmış, özetle ekonomi durgunluk dönemine girmiştir. Bu eğilimin yönü, alınan çok yönlü tedbirlerle artan ya da azalan yönde eğilime dönüştürülmeye çalışılmaktadır. Bu durum, yeniden; yatırımlarda, üretimde ve gelirde artışlara yol açacak, pek çok değişkenin zaman içindeki değerlerinde artışlara ya da azalışlara sebep olabilecektir. Ancak, unutulmamalıdır ki, uzun bir zaman dönemi içinde, herhangi bir zamanda yeni benzer ya da başka türlü bir krizle karşılaşılabilir. Açıklanan türden değişmeler tekrarlanır gider. Konjonktürel değişme gösteren örnek bir grafik, Şekil 14.4'te verilmiştir.

Konjonktürel bileşenin açıkladığı değişmeler periyodik olmayan, ancak döngüsel olan değişmelerdir. Bu değişmeler, ekonomi ve iş idaresiyle ilgili değişkenler üzerinde aynı şiddette olmasa da aynı yönde etki ederler. Konjonktürün artma yönündeki etkisi, trendin artış eğilimini hızlandırır. Buna karşılık, konjonktürün azalma yönündeki etkisi trendin artış hızını yavaşlatır, hatta tamamen durdurabilir.

Konjonktürel değişmelerin sadece ekonomik faktörlerdeki değişmelerden meydana gelmesi gerekmez. Örneğin; iklim koşullarındaki değişmeler, tarım ürünlerinin üretim miktarlarında konjonktürel değişmelere neden olabilir. Yine modadaki değişmeler, belirli bir ürünün satışları üzerinde döngüsel değişmelere neden olabilir.

Rassal Bileşen: Zaman serilerindeki düzensiz değişmelere "rassal değişme" adı verilir. Rassal değişmeler, beklenmedik olayların zaman serileri üzerindeki etkisiyle meydana gelen değişmelerdir. Örneğin; deprem, siyasal karışıklıklar, savaş, grev ve lokavt, rakip firmaların politikalarındaki değişiklikler v.b. gibi etkiler, rassal değişmelere neden olur. Rassal Bileşen, zaman serileri üzerinde trendin, mevsimsel bileşenin ve konjonktürel bileşenin etkisi ayrıştırıldıktan sonra geride kalan etkiyi açıklayan bileşendir.

Yanıltıcı Faktörler

Bazı aylık zaman serilerinde bir ayın gözlem değeri, ilgili değişkenin o ay içinde almış olduğu değerlerin toplamından oluşmaktadır. Bu tür zaman serilerinin gözlem değerleri ayların gün sayılarının 28 (bazan 29), 30 ve 31 olmasından ya da bayram ve hafta sonu tatilleri nedeniyle, ayların işgünü sayılarının farklı olmasından etkilenmektedir.

Ayların gün ve işgünü sayılarındaki faklılıklar zaman serilerinde mevsimsel değişme varmış gibi izlenim yarattıkları için, zaman serilerini etkileyen önemli yanıltıcı faktörler olarak bilinirler. Aylık satış miktarlarına ilişkin zaman serileri, bu yanıltıcı faktörlerden etkilenen serilere örnek olarak verilebilir.

Gözlem değerleri para birimiyle ifade edilmiş olan zaman serileri, enflasyon nedeniyle meydana gelen fiyat değişikliklerinin etkisi altında kalırlar. Fiyat değişiklikleri, bu tür zaman serilerinde trend etkisi varmış gibi izlenim yaratan bir başka yanıltıcı faktör olarak bilinmektedir. Yıllık öğrenci yurt ücretleri, zaman serisi fiyat değişikliklerinin etkisinde olan serilere örnek olarak verilebilir.

Zaman serilerini çözümlemeye geçmeden önce, eğer gerekliyse, serileri ayların gün ve işgünü sayılarının farklı olması ya da fiyat değişiklikleri gibi yanıltıcı faktörlerin etkisinden arındırmak gerekir. Aksi halde çözümleme sonucu, hatalı bilgi üretilmiş olur. Uygulamalarda zaman serisi gözlem değerlerinin yanıltıcı faktörlerin etkisinden arındırılması amacıyla, gün sayısı, işgünü sayısı ya da fiyat değişikliklerine ilişkin uygun düzeltmelerin yapılması gerekir.

Gün sayısı ve işgünü sayısına ilişkin düzeltilmiş gözlem değerlerinin hesaplanması için sırasıyla, aşağıdaki eşitliklerden yararlanılır:

$$GSD = \frac{\text{Ortalama Bir Aydaki Gün Sayısı}}{\text{Düzeltme Yapılacak Aydaki Gün Sayısı}} \cdot y_t$$

$$\text{İ}GSD = \frac{\text{Ortalama Bir Aydaki İşgünü Sayısı}}{\text{Düzeltme Yapılacak Aydaki İşgünü Sayısı}} \cdot y_t$$

Burada,

GSD = Gün sayısı bakımından düzeltmeyi,
İGSD = İşgünü sayısı bakımından düzeltmeyi,
y_t = Düzeltme yapılacak ayın gözlem değerini göstermektedir.

Zaman serilerini fiyat değişikliklerinin etkisinden arındırmak için yapılacak düzeltmeyse, cari fiyatlarla verilmiş gözlem değerlerinin sabit fiyatlara dönüştürülmesi suretiyle yapılır. Bu amaçla, cari fiyatlarla verilmiş gözlem değerleri toptan eşya fiyatları indeksi, tüketici fiyatları indeksi v.b. gibi uygun indekse bölünür.

SIRA SİZDE

1. Trende neden olan etkileri açıklayınız.

2. Dalga uzunluğu nedir?

3. Zaman serilerini etkileyen yanıltıcı faktörler nelerdir?

ZAMAN SERİSİ ÇÖZÜMLEMESİ

AMAÇ 3

Zaman serisi çözümlemesi kavramını açıklayabilecek ve hareketli ortalamalar hesaplayabileceksiniz.

Zaman Serisi Çözümlemesi Tanımı

Zaman serisi gözlem değerleri üzerinde rassal değişmelerin yanında diğer zaman serisi bileşenlerinden hangilerinin etkili olduğunun belirlenmesi çalışmalarına, "zaman serisi çözümlemesi" denir. Bu tanım kapsamında, zaman serisi çözümlemesinin amacı, serinin hangi bileşenlerin etkisinde olduğunu belirlemek ve her bileşenin etkisini tahminlemek, (başka bir anlatımla) ilgili zaman serisinin özelliklerini açıklayabilmektir.

Zaman serisi çözümlemesinin ikinci ve en önemli amacı, serinin öngörü amacıyla çözümlenmesidir. Zaman serisi gözlem değerleri arasında bağımlılığın, iç bağımlılığın, var olması ve serideki değişmelere neden olan etkilerin gelecekte de aynı şekilde devam edeceği varsayımıyla, bir zaman serisinin geçmiş dönem gözlem değerlerini kullanarak, gelecek dönem öngörü değerlerini elde etmek mümkün olabilmektedir. Zaman serileri çözümlemesini, bağımsız gözlem değerlerinden meydana gelen serilerin çözümlemesinden ayıran bu önemli özellik, öngörü amacıyla zaman serisi çözümlemesi tanımı yapma gereğini ortaya çıkarmaktadır. Zaman serilerinin öngörü amacıyla çözümlemesi, bir zaman serisini etkileyen unsurların belirlenmesi, yapılan belirlemeden yararlanarak geçmişin açıklanması ve istatistiksel açıdan normale göre gerçekleşen durumun değerlendirilmesi, belirlenen unsurların gelecekte de seriyi aynı şekilde etkilemeye devam edeceği varsayımı altında, gelecek dönemler için öngörüler yapılması ve bunların karar verme ve planlama faaliyetleri için kullanıma sunulması çalışmalarıdır.

Karar verme ve planlama faaliyetlerine her geçen gün artan gereksinim, zaman serilerinin öngörü amacıyla çözümlemesini önemli bir konu haline getirmektedir.

Ana çizgileriyle çözümlemelerde, zaman serisi bileşenleriyle ilgilenilmektedir. Bu ünitenin izleyen bölümlerinde, zaman serisi çözümlemelerinin bileşenlere ayırma yöntemiyle nasıl inceleneceği konusu, uygulamalı olarak ele alınacaktır.

Zaman Serisi Çözümlemesinde Hareketli Ortalamalar

Hareketli ortalamalar zaman serilerini mevsimsel ve rassal bileşenlerin etkisinden arındırmak suretiyle, bu serilerin genel eğilimini elde etmek amacıyla başvurulan bir istatistiksel yaklaşımdır.

Hareketli ortalamaların hesaplanmasındaki aşamalar, aşağıda ayrıntılarıyla ele alınmıştır.

- Zaman serisinin ilk gözlem değerinden başlamak üzere k sayıda $y_1, y_2, ..., y_k$ gözlem değerinden oluşan bir küme belirlenir. Bu kümede yer alan gözlem değerleri için aritmetik ortalama $\bar{x}_1$ hesaplanır. Burada k, oluşturulan kümedeki gözlem değeri sayısını gösterir. Hesaplanacak hareketli ortalamalar serisini etkilediği için k'nın değerini belirlemek önemlidir. Aylık serilerde mevsimsel dalga uzunluğu L = 12, mevsimlik ve çeyrek yıllık serilerde dalga uzunluğu L = 4 değerleri, k değeri olarak alınır. k = 3 ya da daha küçük hareketli ortalamalar, genellikle aylık zaman serilerini rassal bileşenin etkisinden arındırmak amacıyla kullanılır. Çünkü rassal etkiler üç aydan uzun sürmez. Hareketli ortalamalar, belirlenen k değeriyle isimlendirilirler. Örneğin; k = 3 ise 3'erli hareketli ortalamalar (3'erli H.O), k = 4 için 4'erli hareketli ortalamalar (4'erli H.O) gibi.
- İlk k sayıdaki gözlem değerleri kümesinin ilk gözlem değeri y_1 kümeden çıkarılır. (k+1)'inci sıradaki gözlem değeri y_{k+1} kümeye dahil edilerek k gözlem değerinden oluşan ikinci bir küme oluşturulur ve bu küme için de aritmetik ortalama $\bar{x}_2$ hesaplanır.
- Bu işlemler benzer şekilde sürdürülerek k tek sayıysa n-(k-1) sayıda, k çift sayıysa n-k sayıda gözlem değerleri kümeleri oluşturulur ve n-(k-1) ya da n-k sayıda aritmetik ortalama hesaplanmış olur.

Hesaplanan aritmetik ortalamalar (bir başka ifadeyle hareketli ortalamalar), ait oldukları kümenin tam ortalamasına karşı gelecek şekilde yazılır. Gözlem sayısı değeri tek olduğunda, hesaplanan ortalamalar hesaplandıkları kümenin bir terimi-ne karşı gelir. 3’erli H.O’lar Tablo 14.2’de gösterilmiştir. Örneğin, t=2 zaman dönemine karşılık gelen 3’erli H.O aşağıdaki gibi hesaplanmıştır.

$$\frac{y_1 + y_2 + y_3}{3} = \frac{10 + 31 + 43}{3} = 28$$

t	y_t	3’erli H.O
1	10	-
2	31	28
3	43	30
4	16	
5	11	
6	33	
7	45	
8	17	
9	13	21.,33
10	34	

Tablo 14.2 *Üçerli hareketli ortalamalar.*

k çiftse, örneğin 4’erli hareketli ortalamaların uygulanması halinde hesaplanan hareketli ortalamalar ait oldukları kümedeki herhangi bir gözlem değerine (ya da zaman dönemine) karşı gelmezler. Örneğin; Tablo 14.3’de 3’üncü sutünda yer alan ilk dört gözlem değerinden oluşan birinci kümenin hesaplanan aritmetik ortalaması $\frac{10 + 31 + 43 + 16}{4} = 25$ y_2 ile y_3 gözlem değerleri arasında t = 2.5 zaman noktasına karşı gelmektedir. Hesaplanan hareketli ortalamaların ait oldukları kümedeki herhangi bir gözlem değerine karşı gelmesini sağlamak amacıyla, yeniden 2’şerli hareketli ortalamaları alınır. Bu hareketli ortalamalara “merkezileştirilmiş hareketli ortalamalar” adı verilir. Hesaplanan merkezileştirilmiş ortalamalar Tablo 14.3’de gösterilmiştir.

T		4’erli H.O	2’şerli H.O Merkezileştirilmiş H.O
1	10		-
2	31		-
		25	
3	43		25.12
		25.25	
4	16		25.50
		25.75	
5	11	.	.
		.	.
6	33	.	.
		.	.
7	45	.	.
		.	.
8	17		.
		27.25	.
9	13		-
10	34		-

Tablo 14.3 *Dörderli hareketli ortalamalar.*

Örneğin birinci ve ikinci dörderli hareketli ortalamalar $\bar{x}_1 = 25$ ve $\bar{x}_2 = 25{,}25$ dir. Buna göre birinci merkezileştirilmiş hareketli ortalama

$$\frac{\bar{x}_1 + \bar{x}_2}{2} = \frac{25 + 25.25}{2} = 25.12$$

işlemiyle bulunmuştur.

Hareketli ortalamalar, mevsimsel ve rassal bileşenlerin etkisinden arındırılmış trend ve konjonktürel bileşenlerin etkisini gösteren değerlerdir. Zaman serilerinin çözümlenmesi sürecinde k çift sayıysa merkezileştirilmiş hareketli ortalamalar, k tek sayıysa hareketli ortalamalar kullanılır.

SIRA SİZDE

1. Zaman serisi çözümlemesi kavramını açıklayınız.

2. Hareketli ortalamaların hesaplanmasında izlenen aşamaları sıralayınız.

3. Merkezileştirilmiş hareketli ortalamalara, hangi özellikteki serilerde gereksinilir? Açıklayınız.

ZAMAN SERİSİ ÇÖZÜMLEMESİNDE BİLEŞENLERE AYIRMA YÖNTEMİ

AMAÇ 4

Mevsimsel olmayan ve mevsimsel serilerin, betimsel ve öngörü amaçlarıyla çözümlenmesinde, bileşenlerine ayırma modellerini uygulayabileceksiniz.

Genel Açıklamalar

Bu yöntemin esası, zaman serisi bileşenlerinin, serinin gözlem değerleri üzerindeki etkilerini ayrı ayrı tahminlemek ve bu tahminlerden yararlanarak, serinin özelliklerini açıklamaya imkan veren bilgiler üretmektir. Ayrıca; eğer, çözümlenecek zaman serisinin parametreleri zaman içinde değişmiyorsa (seriyi etkileyen bileşenlerin etkileri gelecekte de seriyi aynı şekilde etkilemeye devam ederse), bu tahminler, nokta öngörülerini hesaplamak amacıyla da kullanılabilir.

Bileşenlere ayırma yöntemi, geçmişte yaygın olarak kullanılmış olan betimsel bir yöntem olup, kuramsal yönü oldukça zayıftır.

Yönteme İlişkin Modeller

Bir araştırmacı bir zaman serisini çözümlemek istediğinde, zaman serisi bileşenlerinin etkilerini araştırmak durumundadır. Çünkü; zaman serileri, rassal bileşenin yanında, diğer üç bileşenin değişik kombinasyonlarının etkisinde, bunların adeta ortak bir sonucu niteliğindedir. Bu durum, zaman serisinin herhangi bir t dönemindeki y_t gözlem değeriyle yukarıda belirtilen zaman serisi bileşenleri arasında,

$$y_t = T_t \,.\, M_t \,.\, K_t \,.\, R_t \qquad \text{çarpımsal}$$

ya da

$$y_t = T_t + M_t + K_t + R_t \qquad \text{toplamsal}$$

şeklinde bir ilişkiye yer veren modelle açıklanmaktadır.

Burada

y_t = Zaman serisinin t zaman dönemindeki gözlem değerini,
T_t = Trend bileşeni'nin (ya da faktörün) t zaman dönemindeki etkisini,
M_t = Mevsimsel bileşenin (ya da faktörün) t zaman dönemindeki etkisini,
K_t = Konjonktürel bileşenin (ya da faktörün) t zaman dönemindeki etkisini,
R_t = Rassal bileşenin (ya da faktörün) t zaman dönemindeki etkisini,

göstermektedir.

Çözümlenmesi istenen zaman serisi, trend artarken (artan ya da azalırken) azalan dalga şiddetine sahip mevsimsel değişmeler gösteriyorsa ve seriyi açıklayan parametreler zaman içinde değişmiyorsa, bu zaman serisinin çözümlenmesi için çarpımsal modelin kullanılması uygundur. İlgilenilen zaman serisinin parametreleri, zaman içinde değişmiyorsa ve seriyi etkileyen mevsimsel değişmelerin dalga şiddetleri birbirine eşitse, (başka bir ifadeyle trend etkisinden bağımsızsa) toplamsal model tercih edilir. Ancak, gerçek yaşamda karşılaşılan serilerin pek çoğunda mevsimsel değişmelerin dalga şiddetleri farklılık gösterdiği için, zaman serisi çözümlemelerinde, genellikle çarpımsal model kullanılmaktadır. Bu nedenle bu ünitede çarpımsal modele ilişkin kuramsal açıklamalara ve uygulamalarına yer verilmiştir.

Bileşenlere Ayırma Yöntemiyle Çözümlemede Aşamalar

Serinin Yanıltıcı Faktörlerin Etkisinden Arındırılması

Kimi zaman, serilerinde trend ya da mevsimsel bileşenin etkisi varmış gibi izlenim yaratan ve daha önce açıklanan yanıltıcı faktörlerin etkisi olabilir. Zaman serilerini çözümlemeye geçmeden önce, serilerin eğer gerekliyse, söz konusu etkilerden arındırılması gerekir. Bu amaçla yine daha önce sözü edilen (gün sayısı, işgünü sayısı ya da fiyat değişmeleri bakımından) uygun düzeltme işlemlerine başvurulur. Gerekli olan düzeltme işlemlerine başvurmadan yapılacak çözümlemelerde, zaman serisi bileşenlerine ilişkin tahminler hatalı ve bu tahminleri kullanarak elde edilecek öngörüler tutarsız olabilir.

Çarpımsal Modelin Uygulanması

Zaman serilerinin mevsimsel olmayan ve mevsimsel seriler olarak sınıflandırılması, zaman serisi çözümlemelerinde kullanılan en önemli bir sınıflandırmadır. Bu nedenle, yukarıda açıklanan çarpımsal modellerin uygulamaları, bu sınıflandırmaya uygun olarak, ayrı ayrı başlıklar altında yapılmıştır.

Mevsimsel Olmayan Serilerin Çözümlenmesinde Çarpımsal Modelin Uygulanması

Modele İlişkin Açıklamalar

Zaman serileri, R_t rassal bileşenin yanında diğer üç bileşenin de mutlaka etkisinde olmayıp, onların değişik kombinasyonlarının etkisinde olabilmektedir. Örneğin; sık karşılaşılan yıllık zaman serileri, mevsimsel bileşenin etkisini göstermezler. Mevsimsel olmayan serilerin açıklanmasında

$$y_t = T_t \, . \, K_t \, . \, R_t$$

çarpımsal model uygulanabilir. Bu model türü tercih edildiğinde, zaman serisini etkileyen T_t , K_t ve R_t bileşenlerine ilişkin t_t , k_t ve r_t tahminlerinin ayrı ayrı hesaplanması gerekir. Bu durumda zaman serisinin t dönemine ait tahmin değeri,

$$y'_t = t_t \,.\, k_t \,.\, r_t$$

olur.

Eğer, ilgilenilen zaman serisinin gözlem değerleri üzerinde, mevsimsel bileşeninin etkisi söz konusu değilse, seri, yıllık zaman serisiyse, öngörü amacıyla yapılacak çözümlemede kullanılacak çarpımsal model,

$$y_t = T_t + \varepsilon_t$$

şeklinde ifade edilir ve y_t'nin nokta öngörü değeri $y'_t = t_t$ olur.

Modelde Yer Alan Bileşenlerin Tahminlenmesi Sürecinde İzlenen Aşamalar

Trend Bileşeninin Tahminlenmesi: Çözümlenen $y_1, y_2, \ldots , y_n$ serisinin t dönemindeki T_t bileşenini tahminleyebilmek için, bu serinin gözlem değerlerine, genellikle

$$T_t = \beta_0 + \beta_1 t \qquad \text{doğrusal trend}$$

ya da

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2 \qquad \text{eğrisel trend}$$

eşitlikleri uygulanır ($y_1, y_2, \ldots , y_n$ gözlem değerleri yanıltıcı faktörlerin etkisinden arındırılmış olduğu varsayılmıştır). Bu uygulamayla zaman ve gözlem değerleri arasında fonksiyonel bir ilişki kurulmuş olur.

Burada β_0, β_1 ve β_2 tahminlenecek model parametrelerini göstermektedir. Bu iki modelden hangisinin, T_t'in tahminlenmesi için kullanılacağına karar verebilmek amacıyla, serinin kartezyen grafiği incelenir. Serinin trendi Şekil 14.2a ve b örneklerine benziyorsa, T_t 'in tahminlenmesi için, doğrusal modelin Şekil 14.2c ve d örneklerine benzemesi durumunda eğrisel modelin tercih edilmesi önerilir. Trend bileşeni T_t 'in t_t tahmini için doğrusal model

$$t_t = b_0 + b_1 t$$

eğrisel model

$$t_t = b_0 + b_1 t + b_2 t^2$$

eşitlikleri ile ifade edilir. Burada b_0, b_1 ve b_2 sırasıyla β_0, β_1 ve β_2 'nin tahminlerini gösterirler ve aşağıda verilen “En Küçük Kareler Normal Denklemlerinin” çözümlenmesiyle hesaplanırlar.

Doğrusal Trend modeline ilişkin En Küçük Kareler Normal Denklemleri

$$\Sigma y_t = nb_0 + b_1 \Sigma t$$

$$\Sigma y_t\, t = b_0 \Sigma t + b_1 \Sigma t^2$$

Eğrisel Trend modeline ilişkin En Küçük Kareler Normal Denklemleri

$\Sigma y_t = nb_0 + b_1 \Sigma t + b_2 \Sigma t^2$

$\Sigma y_t\, t = b_0 \Sigma t + b_1 \Sigma t^2 + b_2 \Sigma t^3$

$\Sigma y_t\, t^2 = b_0 \Sigma t^2 + b_1 \Sigma t^3 + b_2 \Sigma t^4$

eşitlikleriyle ifade edilmektedir.

Ancak, bilgisayar desteğinin olmadığı durumlarda, hesaplamalarda kolaylık sağladığı için, yukarıdaki En Küçük Kareler Normal Denklemlerinden değil, onların sadeleştirilmiş yazılımlarından yararlanılır. En Küçük Kareler Normal Denklemlerinde sadeleştirmeler zaman dönemlerine (daha önce açıklanmış olduğu gibi) t = 1, 2, ..., n kodları vermek yerine, serinin medyan dönemine sıfır kodu verilmesi ve diğer dönemlere ilişkin kodlamalarda, $\Sigma t = 0$ olacak şekilde, düzenleme yapılmasıyla sağlanmış olur.

Serinin terim sayısı n tek sayıysa medyan dönem vardır ve bu dönem t = 0 olarak kodlanırsa, diğer dönemler .... -3, -2 ,-1, 0, 1, 2, 3 .... kodlanmış ve dolayısıyla $\Sigma t = 0$ elde edilmiş olur.

Serinin terim sayısı n çift sayıysa medyan dönem yoktur. En ortadaki iki dönemin aritmetik ortalaması medyan dönem olarak kabul edilir. Böylece iki zaman noktasının aralığı iki birim uzunluk olarak tanımlanmış olur. Bu tanıma göre serinin en ortasındaki iki dönemden ilkine t = -1 diğerine t = 1 kodu verilerek diğer t dönemleri ... -5, -3, -1, 1, 3, 5, ... şeklinde kodlanmış olur. Burada da amaç $\Sigma t = 0$'a eşit kılmaktır.

Zaman değişkeninin şıklarına verilecek kodların cebirsel toplamlarının sıfıra eşitlenecek şekilde düzenlenmesi, En Küçük Kareler Denklemlerini sadeleştirir. Çünkü n bilinmeyenli denklem sistemi n-1 bilinmeyenli denklem sistemine dönüştürülmüş olur.

Sadeleştirilmiş normal denklemler aşağıda verilmiştir.

Doğrusal Modele İlişkin Sadeleştirilmiş Normal Denklemler

$\Sigma y_t = nb_0$

$\Sigma y_t t = b_1 \Sigma t^2$

eşitlikleriyle,

Eğrisel Modele İlişkin Sadeleştirilmiş Normal Denklemlerse,

$\Sigma y_t = nb_0 + b_2 \Sigma t^2$

$\Sigma y_t\, t = b_1 \Sigma t^2$

$\Sigma y_t\, t^2 = b_0 \Sigma t^2 + b_2 \Sigma t^4$

eşitlikleriyle ifade edilir.

Tahmin edilen trend doğrusu ya da eğrisinin zaman serisini açıklama gücü $\Sigma(y_t - t_t)^2$ = minimum koşulunu sağlamasıyla ilişkilidir. y_t gözlem değerleriyle t_t trend tahmin değerleri arasında

$\Sigma y_t = \Sigma t_t$ ve $y_t \neq t_t$

ilişkisi vardır. Gözlem değerleriyle trend tahmin değerleri arasındaki farklara $(y_t - t_t)$, tahmin hataları adı verilir. Tahmin hatalarının ortalama ölçüsü standart hata, serideki gözlem değeri sayısı $n \geq 30$ ise,

$$s_y = \frac{\sqrt{\Sigma(y_t - t_t)^2}}{n}$$

Buna karşılık, $n < 30$ olduğunda

$$s_y = \sqrt{\frac{\Sigma(y_t - t_t)^2}{n - k}}$$

eşitliği ile hesaplanır. Burada k, uygulanan trend denklemindeki tahminlenen parametre sayısını gösterir. Bu sayı, doğrusal model için $k = 2$, eğrisel model için $k = 3$ alınır.

Bir zaman serisi trendinin tahmininde tercih edilecek denklem türüne karar verebilmek için, trendi her iki denkleme göre tahmin etmek ve standart hatalarını karşılaştırmak gerekir. Standart hatası küçük olan diğerine tercih edilir.

Bu ünitede yer alan örnek çözümlerde, sadece bir trend modelinin uygulamasına yer verileceği için, seriyi daha iyi açıklayan trend denkleminin belirlenmesi kartezyen grafiğin incelenmesiyle yapılacak, standart hata ölçüsünden yararlanılmayacaktır.

ÖRNEK 1 ***Tablo 14.1'de verilen zaman serisinin trend bileşenini tahminleyiniz.***

ÇÖZÜM

Tablo 14.1'de verilen zaman serisinin Şekil 14.1'de verilen grafiği incelendiğinde serinin doğrusal artan bir eğilim gösterdiği söylenebilir. Seride yanıltıcı faktörlerin etkisi olmadığı varsayılmıştır.

Bu görsel tespit nedeniyle, T_t'nin tahminlenebilmesi için seri değerlerine

$$T_t = \beta_0 + \beta_1 t$$

doğrusal model uygulanmıştır. Modelin β_0 ve β_1 parametrelerinin tahminleri b_0 ve b_1, En Küçük Kareler Normal Denklemlerinden yararlanarak aşağıdaki gibi hesaplanır: İlgili değerler Tablo 14.4'den elde edilmiştir.

$$\Sigma y_t = 269.987$$
$$\Sigma t = 91$$
$$\Sigma t^2 = 819$$
$$\Sigma y_t t = 2197.06$$

Bu değerler normal denklemlerde yerlerine konarak,

$$269.987 = 13b_0 + 91b_1$$
$$2197.06 = 91b_0 + 819b_1$$

$$-7(269.978) = -7(13)b_0 + (-7)(91)b_1$$
$$2197.06 = 91b_0 + 819b_1$$

$$-1889.,846 = -91b_0 - 637b_1$$
$$2197.06 = 91b_0 + 819b_1$$

$$\frac{307.214}{182} = \frac{0 + 182b_1}{182}$$

$$b_1 = 1.688$$

olarak elde edilir. b_1 değeri ilk eşitlikte yerine konarak,

$$269.987 = 13b_0 + 91.1688$$
$$b_0 = 8.955$$

olarak elde edilir. Sonuçlara göre T_t'nin doğrusal tahmin modeli t_t

$$t_t = b_0 + b_1 t = 8.955 + 1.688t$$

olarak elde edilir. b_1 = 1,688 pozitif bir değer olduğu için serinin trendi, artan eğilime sahiptir (b_1 trend doğrusunun eğimini gösterir).

Tablo 14.4 *Doğrusal trend tahminleri.*

t	ihracat = y_t	t^2	t . y_t	t_t = 8.955 + 1.688t
1	11.625	1	11.625	10.642
2	12.959	4	25.918	12.33
3	13.593	9	40.779	14.018
4	14.715	16	58.86	15.705
5	15.345	25	76.725	17.393
6	18.106	36	108.636	19.081
7	21.637	49	151.459	20.768
8	23.224	64	185.792	22.456
9	26.261	81	236.349	24.143
10	26.974	100	269.74	25.831
11	26,587	121	292.457	27.519
12	27.775	144	333.3	29.206
13	31.186	169	405.418	30.894
Toplam	**269.987**	**819**	**2197.058**	**269.986**

Bu tahmin modelinde t için t = 1, 2, ..., 13 değerlerini uygulamak suretiyle Tablo 14.4'teki tahmin değerleri t_t hesaplanmış olur. Örneğin, Tablo 14.4'te t = 2 için hesaplanan t_2 = 12.33 trend tahmin değeri,

$$t_2 = 8.955 + 1.688\ (2) = 12.33$$

şeklinde elde edilmiştir.

Konjonktürel Bileşenin Tahminlenmesi: Bu tahminleme için, serinin önce, trend bileşeninden arındırılması gerekir. Bu arındırma y_t gözlem değerini t_t tahmin değerlerine bölmek ($\frac{y_t}{t_t}$) suretiyle yapılır.

$$y_t{}' = t_t \cdot k_t \cdot r_t$$

olduğundan

$$\frac{y_t}{t_t} = \frac{t_t \cdot k_t \cdot r_t}{t_t} = k_t \cdot r_t$$

trend bileşeninden arındırılmış seri değerleri, bir başka ifadeyle Konjonktürel ve rassal bileşenlerin k_t . r_t tahminleri elde edilmiş olur. Yıl ya da daha uzun zaman aralıklarıyla yapılan ölçümlerden meydana gelen zaman serilerinde K_t konjonktürel bileşeninin, tahmin değeri k_t genellikle R_t rassal bileşeninin tahmin değeri r_t ile birlikte $k_t \cdot r_t = \frac{y_t}{t_t}$ eşitliğiyle tahmin edilir. Tahmin edilmiş olan $k_t \cdot r_t$ tahminleri sanki, K_t bileşeninin k_t tahminiymiş gibi yorumlanırlar. $k_t \cdot r_t = 1$ olduğunda zaman, serisinin t dönemindeki gözlem değeri üzerinde konjonktürel etkinin görülmediği yorumu yapılır. $k_t \cdot r_t > 1$ ya da $k_t \cdot r_t < 1$ olması durumlarında serinin t dönemindeki gözlem değeri üzerinde konjonktürel etkinin olduğu yorumu yapılır. Birinci durumdaki etki, serinin trendin üzerinde değer almasına, ikinci durumdaki etkiyse trendin altında değer almasına neden olur. Varolan etkinin düzeyi $k_t \cdot r_t - 1$ farkı alınarak belirlenir ve yorum kolaylığı sağlamak için bu fark 100 ile çarpılır.

ÖRNEK 2

Tablo 14.1'deki zaman serisinin konjonktürel bileşeninin etkisini tahminleyiniz.

ÇÖZÜM

Tablo 14.1'de verilen seri için $k_t \cdot r_t$ tahmin değerleri Tablo 14.5'te verilmiştir. Örneğin; t = 2 için $y_2 = 12.959$ ve $t_2 = 12.33$ olduğuna göre, trend bileşenin etkisinden arındırılmış seri değeri,

$$\frac{y_t}{t_t} = \frac{y_2}{t_2} = \frac{12.959}{12.33} = 1.051$$

bulunmuş olur. Bu 1.05 değeri t = 2 için Konjonktürel ve rassal bileşenlerin t dönemindeki $k_2 \cdot r_2$ tahmin değerini verir.

Tablo 14.5: *$k_t \cdot r_t$ tahminleri*

t	ihracat = y_t	$t_t = 8.955 + 1.688t$	$y_t/t_t = k_t \cdot r_t$
1	11.625	10.642	1.092
2	12.959	12.33	1.051
3	13.593	14.018	0.97
4	14.715	15.705	0.937
5	15.345	17.393	0.882
6	18.106	19.081	0.949
7	21,637	20.768	1.042
8	23.224	22.456	1.034
9	26.261	24.143	1.088
10	26.974	25.831	1.044
11	26.587	27.519	0.966
12	27.775	29.206	0.951
13	31.186	30.894	1.009

Tablo 14.5 incelendiğinde, ihracat değişkeni konjonktür etkisiyle örneğin, 1990 yılında (t = 2 için) (1.051 - 1) . 100 = %5.1 oranında normalin (trend değerinin) üzerinde değer almış, buna karşılık 1991 yılında (t = 3 için) (0.970 - 1) . 100 = %-3 oranında normalin (trend değerinin) altında değer almıştır denir. Genel olarak $k_t \cdot r_t$ tahminleri 1'e çok yakın değerler olduğundan incelenen dönemde ihracat değişkeni üzeride konjonktürün etkisi yoktur yorumu yapılabilir.

Öngörü Değerlerinin Türetilmesi: Daha önce de açıklanmış olduğu gibi, Konjonktürel ve Rassal Bileşenlere ilişkin güvenilir öngörüler türetilemediği için, öngörü amacıyla çözümlemelerde genellikle bu iki bileşenle ilgilenilmez ve çarpımsal öngörü modellerinde bu iki bileşene yer verilmez. Bu durumda t+1 ön dönem için öngörü modeli

$$y_{t+1} = T_{t+1} + \varepsilon_{t+1}$$

olur. Doğrusal Trend bileşeni t dönemi için

$$T_t = \beta_0 + \beta_1 t$$

ile açıklandığından bu öngörü modeli aşağıdaki gibi de yazılabilir:

$$y_{t+1} = \beta_0 + \beta_1 t_{t+1} + \varepsilon_{t+1}$$

Trend bileşeni T_t 'nin tahmini t_t

$$t_t = b_0 + b_1 t$$

eşitliğiyle ifade edildiğinden t+1 ön dönem öngörü değeri

$$y'_{t+1} = t_{t+1} = b_0 + b_1(t + 1)$$

eşitliğiyle hesaplanmış olur.

ÖRNEK 3

Tablo 14.1'deki zaman serisi için 2002 yılı öngörü değerini hesaplayınız.

ÇÖZÜM

İlgilendiğimiz örnek için 2002 yılı t + 1 = 13 + 1 = 14 için

y'_{14} öngörü değeri

$$y'_{14} = b_0 + b_1(t + 1) = 8.955 + 1.688(14) = 32.587$$

olur. Bu 2002 yılı için ihracat öngörü değerini ifade eder. Bu tahmin modelinde t yerine sırasıyla t + 2 = 13 + 2 = 15, t + 3 = 13 + 3 = 16 ve ... yazarak diğer dönemler için öngörüler yapılabilir.

Mevsimsel Serilerin Çözümlenmesinde Çarpımsal Modelin Uygulanması

Modele İlişkin Açıklamalar

Bilindiği gibi, çeyrek yıllık, mevsimlik ve aylık zaman serileri, genellikle mevsimsel bileşenin etkisini gösterirler. Anımsanacağı gibi, mevsimsel serileri açıklamak amacıyla kullanılan çarpımsal model genel olarak,

$$y_t = T_t \,.\, M_t \,.\, K_t \,.\, R_t$$

ve bu modele ilişkin tahmin modeli de,

$$y'_t = t_t \,.\, m_t + k_t \,.\, r_t$$

gibi gösterilir. Ancak, öngörü amacıyla yapılacak çözümlemelerde de mevsimsel olmayan çarpımsal modeller için açıklanan nedenle, Konjonktürel ve Rassal bileşenlere yer verilmemektedir. Buna göre, t dönemi için öngörü modeli,

$$y'_t = t_t \cdot m_t$$

eşitliğiyle ifade edilmiş olur.

Modelde Yer Alan Bileşenlerin Tahminlenmesi Sürecinde İzlenen Aşamalar

Yukarıdaki çarpımsal modeller kullanılarak, mevsimsel serilerin çözümlenmesinde aşağıdaki sıra izlenir.

Mevsimsel Bileşenin Tahminlenmesi: Bu tahminleme 4 aşamada gerçekleştirilir:

I. Öncelikle serinin mevsimsel dalga uzunluğu, L belirlenir. Mevsimsel değişmelerin dalga uzunlukları mevsimlik (ya da çeyrek yıllık) serilerde L = 4, aylık serilerde L = 12 olup, genellikle birbirine eşittir. Bazı serilerde mevsimsel değişmelerin her birinin dalga uzunlukları eşit olmayabilir. Bu durumda ortalama dalga uzunluğu, L değeri olarak anılabilir.

II. Serinin hareketli ortalamaları (ya da merkezileştirilmiş hareketli ortalamaları) hesaplanır. Eğer k tek sayıysa merkezleştirilmiş hareketli ortalamalara gereksinim duyulmaz. Hesaplanan hareketli ortalamalar mevsimsel ve rassal bileşenlerden ($m_t \cdot r_t$) arındırılmış, trend ve konjonktürel bileşenlerin etkisini gösteren seri değerleridir. Bu durum matematiksel olarak t dönemi için,

$$\frac{y_t}{m_t \cdot r_t} = \frac{t_t \cdot m_t \cdot k_t \cdot r_t}{m_t \cdot r_t} = t_t \cdot k_t$$

eşitliğiyle ifade edilir.

III. y_t gözlem değerleri, hareketli ortalamalara (ya da merkezileştirilmiş hareketli ortalamalara) bölünerek, $M_t \cdot R_t$ bileşenlerinin tahmini $m_t \cdot r_t$ aşağıdaki eşitlik kullanılarak hesaplanır.

$$\frac{y_t}{t_t \cdot k_t} = m_t \cdot r_t$$

IV. Son işlem, mevsimsel bileşenin tahminlenmesidir. Bunun için, $M_t \cdot R_t$ bileşenlerinin $m_t \cdot r_t$ tahminlerinden r_t etkisinin arındırılması ve m_t tahmininin elde edilmesi gerekir. Mevsimsel bileşeni tahminlemek için aynı zaman dönemine karşı gelen $m_t \cdot r_t$ tahmin değerleri gruplandırılarak ortalamaları hesaplanır. Örneğin; birbirini izleyen yılların aynı aylarının, mevsimlerinin ya da çeyrek yıllarının $m_t \cdot r_t$ tahmin değerleri gruplandırılır ve ortalamaları hesaplanır. Bu ortalamaların sayısı t = 1, 2 .... L olup $\overline{m}_t$ simgesiyle gösterilir. $\sum_{t=1}^{L} \overline{m}_t = L$ olmalıdır. Eğer bu eşitlik sağlanmıyorsa, eşitliğin sağlanması ve m_t tahminlerinin elde edilmesi için $m_t = \overline{m}_t \cdot \dfrac{L}{\sum_{t=1}^{L} \overline{m}_t}$ düzeltmesi yapılır.

t = 1, 2 ..., L sayıda düzeltilmiş m_t. tahminleri, örneğin; bir yılın 12 ayı, 4 mevsimi (ya da 4 çeyrek yılı) için mevsimsel değişmeyi ifade eden tahminler olarak alınır.

Mevsimsel Bileşenden Arındırılmış Seri Değerlerinin Tahminlenmesi: Bu tahminler d_t ile gösterilir ve

$$d_t = \frac{y_t}{m_t} = \frac{t_t \,.\, m_t \,.\, r_t}{m_t}$$

eşitliğinden hesaplanır.

Eğer m_t tahmin değeri 1'den küçükse buna karşı gelen mevsimsel değişmeden arındırılmış gözlem değeri d_t, gerçek gözlem değeri y_t'den büyük değerli, m_t tahmin değeri 1'den büyükse d_t, y_t değerinden küçük olur. Mevsim etkisinden arındırılmış gözlem değerleri d_t trend değerlerine yaklaşırlar. Bir başka ifadeyle $m_t > 1$ olduğunda mevsimsel bileşen t dönemindeki gözlem değerinin aynı dönemdeki normal değerden (normal değer $T_t \,.\, M_t$ dir) büyük değer almasına neden olurken, $m_t < 1$ olduğunda tersi söz konusu olur.

Trend Bileşeninin Tahminlenmesi: Trend bileşeninin t dönemine ilişkin T_t bileşenin t_t tahmini için mevsimsel olmayan serilere uygulanan çözümleme süreci aynen izlenir. Buradaki tek farklılık, trend eşitliklerinin gerçek y_t gözlem değerlerine değil, mevsimsel bileşenin etkisinden arındırılmış olan d_t serisine uygulanacak olmasıdır.

Konjonktürel ve Rassal Bileşenlerin Tahminlenmesi: Mevsim etkisinden arındırılmış d_t değerlerini kullanarak $K_t \,.\, R_t$ bileşenlerinin $k_t \,.\, r_t$ ya da k_t ve r_t tahminleri için mevsimsel olmayan serilere uygulanan çözümleme süreci aynen uygulanır.

Öngörü Değerlerinin Hesaplanması: Daha önce de açıklandığı gibi, bileşenlere ayırma yöntemiyle yapılan öngörü amaçlı çözümlemelerde konjonktürel ve rassal bileşenlerle ilgilenilmemektedir. Bu durumda, t + 1 ön dönem için, y_{t+1}'in öngörü değeri $y'_{t+1} = t_{t+1} \,.\, m_{t+1}$ eşitliğiyle hesaplanır.

Tüm yukarıda sözü edilenlere ilişkin uygulama, aşağıdaki örnekte gösterilmiştir.

ÖRNEK 4

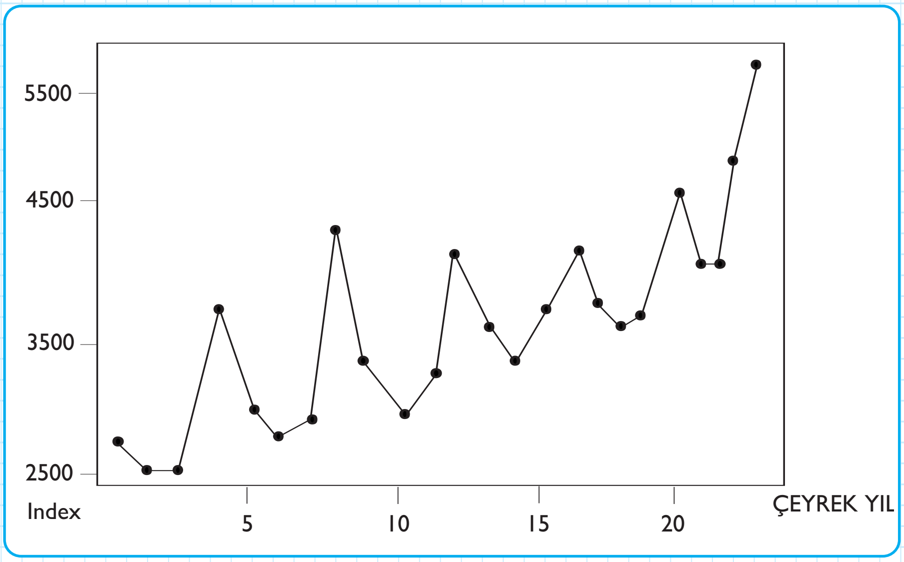
Aşağıdaki seriyi çarpımsal model kullanarak çözümleyiniz.

Tablo 14.6 *Çeyrek yıllık zaman serisi.*

Çeyrek Yıllar	y_t	Çeyrek Yıllar	y_t
1989Q1	2803.27	1992Q1	3549.95
1989Q2	2573.86	1992Q2	3303.34
1989Q3	2568.12	1992Q3	3701.31
1989Q4	3679.43	1992Q4	4160.02
1990Q1	2994.33	1993Q1	3673.27
1990Q2	2745.46	1993Q2	3477.28
1990Q3	2858.71	1993Q3	3561.98
1990Q4	4360.79	1993Q4	4632.53
1991Q1	3378.66	1994Q1	3826.36
1991Q2	2904.93	1994Q2	3830.78
1991Q3	3208.71	1994Q3	4815.19
1991Q4	4101.16	1994Q4	5633.54

ÇÖZÜM

Mevsimsel bileşenin tahminlenmesi (Mevsim İndeksinin Tahminlenmesi): Şekil 14.5 incelendiğinde, serinin rassal bileşenin yanında, artan trend ve dalga uzunluğu L = 4 olan değişen dalga şiddetli mevsimsel bileşene sahip olduğu görülmektedir. Serinin yanıltıcı faktörlerin etkisinde olmadığı varsayılmıştır.



Şekil 14. 5 *Tablo 14.6'da verilen zaman serisinin grafiği.*

Bu bilgilere göre serinin açıklanmasında kullanılabilecek çarpımsal model,

$$y_t = T_t \,.\, M_t \,.\, K_t \,.\, R_t$$

şeklinde ifade edilir. Mevsimsel dalga uzunluğu L = 4 olduğundan, mevsimsel bileşenin tahminlenmesi için serinin önce, dörderli (L = k = 4) hareketli ortalamaları, sonra da merkezileştirilmiş hareketli ortalamaları hesaplanmıştır. Tablo 14.7'de gösterilmiş olan merkezileştirilmiş ortalamalar, trend ve konjonktürel bileşenlerin tahmin değerlerini ($t_t \,.\, k_t$) ya da mevsimsel ve rassal bileşenlerden ($m_t \,.\, r_t$) arındırılmış değerlerini gösterirler. ($m_t \,.\, r_t$) tahminleri,

$$\frac{y_t}{t_t \,.\, k_t} = m_t \,.\, r_t$$

eşitliği kullanılarak hesaplanır.

Tablo 14.7 *Merkezileştirilmiş hareketli ortalamalar ve mevsim indeksleri.*

Çeyrek Yıllar	y_t	Merkezileştirilmiş Hareketli Ortalama (HO) $t_t \cdot k_t$	$\frac{y_t}{HO} = m_t \cdot r_t$	Mevsim İndeksi
1989Q1	2803.27			0.97
1989Q2	2573.86			0.88
1989Q3	2568.12	2930.06	0.8764	0.92
1989Q4	3679.43	2975.39	1.2366	1.2
1990Q1	2994.33	3033.16	0.9871	0.97
1990Q2	2745.46	3154.65	0.8702	0.88
1990Q3	2858.71	3287.86	0.8692	0.92
1990Q4	4360.79	3355.84	1.2994	1.2
1991Q1	3378.66	3419.52	0.9880	0.97
1991Q2	2904.93	3430.82	0.8467	0.88
1991Q3	3208.71	3419.78	0.9382	0.92
1991Q4	4101.16	3490.99	1.1747	1.2
1992Q1	3549.95	3602.37	0.9854	0.97
1992Q2	3303.34	3671.30	0.8997	0.88
1992Q3	3701.31	3694.07	1.0019	0.92
1992Q4	4160.02	3731.23	1.1149	1.2
1993Q1	3673.27	3735.56	0.9833	0.97
1993Q2	3477.28	3777.20	0.9205	0.88
1993Q3	3561.98	3855.40	0.9238	0.92
1993Q4	4632.53	3918.73	1.1821	1.2
1994Q1	3826.36	4119.56	0.9288	0.97
1994Q2	3830.78	4401.34	0.8703	0.88
1994Q3	4815.19			0.92
1994Q4	5633.54			1,2

Mevsimsel bileşenin tahminlenmesiyle ilgili son işlem olarak, Tablo 14.7'deki $m_t \cdot r_t$ tahminlerinden, r_t bileşeninin etkisinin arındırılması ve m_t mevsimsel bileşeninin tahmin edilmesi için, aynı çeyrek yıllara karşı gelen $m_t \cdot r_t$ tahmin değerleri, Tablo 14.8'de gösterildiği gibi gruplandırılmış ve her grup için $\overline{m}_t$ ortalamalar hesaplanmıştır.

Tablo 14.8 *Çeyrek yıllar itibarıyla gruplandırılmış $m_t \cdot r_t$ ve m_t tahminleri.*

Yıllar	Çeyrek Yıllar 1	2	3	4
1989	-----	------	0.8764	1.2366
1990	0.9871	0.8702	0.8692	1.2994
1991	0.9880	0.8467	0.9382	1.1747
1992	0.9854	0.8997	1.0019	1.1149
1993	0.9833	0.9205	0.9238	1.1821
1994	0.9288	0.8703	-----	------
Ortalama = $\overline{m}_t$	0.97452	0.88148	0.9219	1.20154

$$\sum_{t=1}^{L} \overline{m}_t = L = 4$$

olmalıdır.

$$\sum_{t=1}^{L} \overline{m}_t = 0{,}9746 + 0{,}8816 + 0{,}9220 + 1{,}2020 = 4{,}0002$$

hesaplanan bu değer 4'e eşit kabul edildiği için mevsimsel bileşen tahminlerinde düzeltme yapma gereği duyulmamıştır. Her bir çeyreğe ilişkin hesaplanan grup ortalamaları $\overline{m}_t$ 'lar, Tablo 14.7'de olduğu gibi, yılların ilgili çeyrekleri için, mevsim bileşen tahminleri (mevsim indeksleri) m_t olarak alınmıştır.

Serinin Mevsimsel Değişmenin Etkisinden Arındırılması: t dönemi için, mevsimsel değişmeden arındırılmış değer, d_t simgesiyle gösterilir ve $d_t = \frac{y_t}{m_t} = \frac{t_t \,.\, m_t \,.\, r_t}{m_t}$ eşitliğiyle hesaplanır. Örneğin, t = 1 için Tablo 14.9'daki d_t = 2889.97 değeri $d_t = \frac{2803.27}{0.97} = 2889.97$ şeklinde hesaplanmıştır.

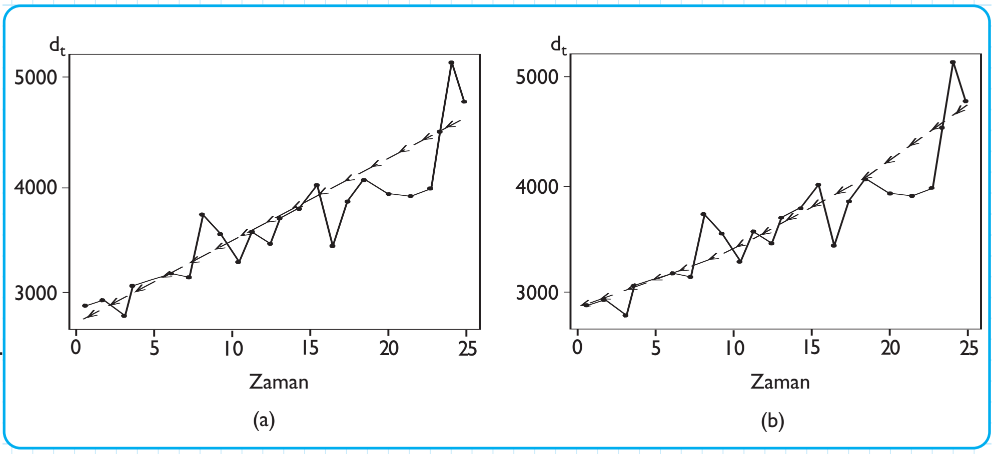
Trend Bileşeninin Tahminlenmesi : Bu aşamada T_t bileşenin t_t tahmini hesaplanır. Hesaplama için, mevsim etkisinden arındırılmış d_t seri değerleri kullanılır. Bu değerler,

$$d_t = \frac{y_t}{m_t} = \frac{t_t \,.\, m_t \,.\, k_t \,.\, r_t}{m_t} = t_t \,.\, k_t \,.\, r_t$$

bileşenlerinin tahmin değerlerini gösterir. Trend bileşeninin t dönemine ilişkin t_t tahminin değerini hesaplamaya geçmeden önce, d_t serisinin kartezyen grafiği Şekil 14.6'da a ve b incelenir. Bu incelemeyle, d_t serisini en iyi açıklayacak trend denklemi belirlenir. Şekil 14.6'da a ve b incelendiğinde uygun olan trend tahmin modelinin,

$$t_t = b_0 + b_1 t + b_2 t^2$$

eğrisel modeli olabileceğine karar verilmiştir.



Şekil 14.6
d_t serisinin grafikleri.

Bu tahmin modeli aşağıdaki gibi sadeleştirilmiş, En Küçük Kareler Normal Denklemleri yardımıyla $t_t = 3525 + 37.2t + 0.504t^2$ olarak tahminlenmiştir. Gerekli olan veriler Tablo 14.9'dan elde edilmiştir. Bu eğrisel tahmin denkleminde, t yerine t = -23, -21, ..., 21, 23 değerleri yazılırsa, Tablo 14.10'da verilen t_t tahminleri elde edilmiş olur.

Sadeleştirilmiş En Küçük Kareler Normal Denklemleri

$\Sigma d_t = nb_0 + b_2\ \Sigma t^2$ $\quad$ $86910,3 = 24b_0 + 4600b_2$

$\Sigma d_t\ t = b_1\ \Sigma t^2$ $\quad$ $171178,48 = 4600b_1$

$\Sigma d_t\ t^2 = b_0\ \Sigma t^2 + b_2\ \Sigma t^4$ $\quad$ $17011448 = 4600b_0 + 1583320b_2$

$b_1 = 37,2$
$b_0 = 3525$
$b_2 = 0,504$

Tablo 14.9 En Küçük Kareler normal denklemleri veri tablosu.

	y_t	t	$y_t/m_t=t_t.k_t=d_t$	t^2	t^3	t^4	$d_t \cdot t$	$d_t \cdot t^2$
	2803.27	-23	2889.97	529	-12167	279841	-66469.31	1528794.13
	2573.86	-21	2924.85	441	-9261	194481	-61421.85	1289858.85
	2568.12	-19	2791.44	361	-6859	130321	-53037.36	1007709.84
	3679.43	-17	3066.19	289	-4913	83521	-52125.23	886128.91
	2994.33	-15	3086.94	225	-3375	50625	-46304.10	694561.50
	2745.46	-13	3119.84	169	-2197	28561	-40557.92	527252.96
	2858.71	-11	3107.30	121	-1331	14641	-34180,30	375983.30
	4360.79	-9	3633.99	81	-729	6561	-32705.91	294353.19
	3378.66	-7	3483.16	49	-343	2401	-24382.12	170674.84
	2904.93	-5	3301.06	25	-125	625	-16505.30	82526.50
	3208.71	-3	3487.73	9	-27	81	-10463.19	31389.57
	4101.16	-1	3417.63	1	-1	1	-3417.63	3417.63
	3549.95	1	3659.74	1	1	1	3659.74	3659.74
	3303.34	3	3753.80	9	27	81	11261.40	33784.20
	3701.31	5	4023.17	25	125	625	20115.85	100579.25
	4160.02	7	3466.68	49	343	2401	24266.76	169867.32
	3673.27	9	3786.87	81	729	6561	34081.83	306736.47
	3477.28	11	3951.46	121	1331	14641	43466.06	478126.66
	3561.98	13	3871.72	169	2197	28561	50332.36	654320.68
	4632.53	15	3860.44	225	3375	50625	57906.60	868599.00
	3826.36	17	3944.70	289	4913	83521	67059.90	1140018.30
	3830.78	19	4353.16	361	6859	130321	82710.04	1571490.76
	4815.19	21	5233.90	441	9261	194481	109911.90	2308149.90
	5633.54	23	4694.62	529	12167	279841	107976.26	2483453.98
Toplam	**86342,98**	**0**	**86910.36**	**4600**	**0**	**1583320**	**171178.48**	**17011437.48**

Konjonktürel Bileşenin Tahmin Edilmesi: Bu bileşenin tahmin edilmesinde ve tahminlerin yorumlanmasında izlenen aşamalar mevsimsel olmayan serilerin çözümlenmesine benzer şekilde yapılır. Örneğin, t = -23 dönemi (1989 I'inci çeyreği) için konjonktürel ve rassal bileşen tahmini,

$$k_t \cdot r_t = \frac{y_t}{t_t \cdot m_t} = 0.98$$

olarak hesaplanır. $k_t \cdot r_t = 0.98$ değerinden yararlanarak 1989 yılı I'inci çeyreğine ilişkin gözlem değeri üzerindeki konjonktürel değişmenin etkisi hakkında yorum yapılır. Yorum: 1989 yılı I'inci çeyreğinde, çeyrek yıllık zaman serisi konjonktür etkisiyle (0.98 - 1).100 = %-1 oranında normalin (trend ve mevsimsel etkinin) altında değer almıştır.

Tablo 14.10
Trend, konjonktürel ve rassal bileşen tahminleri.

y_t	t	$y_t/m_t=t_t.k_t=d_t$	y_t=3525+37.2t+0.504t^2	m_t	t_t . m_t	k_t . r_t
2803.27	-23	2889.97	2935.38	0,97	2847.32	0.98
2573.86	-21	2924.85	2965.46	0,88	2609.60	0.98
2568.12	-19	2791.44	2999.56	0,92	2759.60	0.93
3679.43	-17	3066.19	3037.70	1.20	3645.24	1.00
2994.33	-15	3086.94	3079.87	0.97	2987.48	1.00
2745.46	-13	3119.84	3126.08	0.88	2750.95	0.99
2858.71	-11	3107.30	3176.31	0.92	2922.21	0.97
4360.79	-9	3633.99	3230.58	1.20	3876.69	1.12
3378.66	-7	3483.16	3288.88	0.97	3190.21	1.05
2904.93	-5	3301.06	3351.20	0.88	2949.06	0.98
3208.71	-3	3487.73	3417.57	0,92	3144,16	1.02
4101.16	-1	3417.63	3487.96	1.20	4185.55	0.97
3549.95	1	3659.74	3562.39	0.97	3455.51	1.02
3303,34	3	3753.80	3640.84	0.88	3203.94	1.03
3701.31	5	4023.17	3723.33	0.92	3425.47	1.08
4160.02	7	3466.68	3809.85	1.20	4571.82	0.90
3673.27	9	3786.87	3900.41	0.97	3783.39	0,97
3477.28	11	3951.46	3994.99	0.88	3515.59	0,98
3561.98	13	3871.72	4093.61	0.92	3766.12	0,94
4632.53	15	3860.44	4196.25	1.20	5035.51	0.91
3826.36	17	3944.70	4302.94	0.97	4173.85	0.91
3830.78	19	4353.16	4413.65	0.88	3884.01	0.98
4815.19	21	5233.90	4528.39	0.92	4166.12	1.15
5633.54	23	4694.62	4647.17	1.20	5576.60	1.01

Öngörü değerlerinin türetilmesi:
Bu amaçla $y'_{t+1} = t_{t+1} \,.\, m_{t+1} = (b_0 + b_1(t+1) + b_2t^2(t+1)\, m_{t+1}$ öngörü modelinden yararlanılır.

$$t_{t+1} = 3525 + 37{,}2\,(t+1) + 0{,}504\,(t+1)^2$$

t+1, 1995 yılı 1. çeyreği olduğundan bu dönemin kod değeri 25'tir. Buna göre,

$$t_{1995.\,Q1} = 3525 + 37.2.(25) + 0.504.(25^2) = 4770$$

$$y'_{1995.Q1} = 4770.0.97 = 4626.9$$

Burada 0.97, birinci çeyrek yıl için mevsim bileşenin tahmin değerini ifade eder.

SIRA SİZDE

1. **Bileşenlerine ayırma yöntemiyle çözümlemede, çözümleme aşamalarını sıralayınız.**
2. **Eğer ilgilenilen zaman serisinin gözlem değerleri üzerinde mevsimsel bileşenin etkisi söz konusu değilse, öngörü amacıyla kullanılacak çarpımsal modeli belirleyiniz.**
3. **Çarpımsal modellerde mevsimsel bileşenin tahminlenmesindeki aşamaları sıralayınız.**

Kendimizi Sınayalım

1. Aşağıdakilerden hangisi, zaman serilerinin **temel** bileşeni **değildir?**

a. Trend Bileşeni
b. Rassal Bileşen
c. Zaman Bileşeni
d. Konjonktürel Bileşen
e. Mevsimsel Bileşen

2. Bir zaman serisinin trend tahmini $t_t = 25 + 0.8t$ 'dir. Aşağıdakilerden hangisi, bu trend tahmini için **söylenemez?**

a. Trend doğrusaldır.
b. Trend artan eğilim gösterir.
c. Trend eğriseldir.
d. Trendin eğimi pozitiftir.
e. Trendin eğimi 0,8'dir.

3. Yıllık zaman serilerinde hangi zaman serisi bileşeninin etkisi **görülmez?**

a. Trend bileşeninin
b. Mevsimsel bileşenin
c. Konjonktürel bileşenin
d. Rassal bileşenin
e. Konjonktürel ve Rassal bileşenin

4. Aşağıdakilerden hangisi, mevsimsel değişmenin dalga uzunluğunu ifade eder?

a. Birbirini izleyen iki mevsimsel değişmenin maksimum ya da minimum noktaları arasındaki zaman aralığı
b. Birbirini izleyen mevsimler arasındaki zaman aralığı
c. Bir mevsimsel değişmenin maksimum ve minimum noktaları arasındaki fark
d. Serideki veri sayısı
e. En yüksek ve en düşük mevsimsel değişme arasındaki fark

5. Zaman serilerinde etkisi **mutlaka** görülen bileşen aşağıdakilerden hangisidir?

a. Trend
b. Konjonktürel
c. Mevsimsel
d. Rassal
e. İş gün sayısının farklılığı

6. Bir zaman serisinin hareketli ortalamaları hesaplandığında, seri hangi bileşenlerin etkisinden arındırılmış olur?

a. Trend – Mevsimsel
b. Trend – Konjonktürel
c. Mevsimsel
d. Trend
e. Mevsimsel – Rassal

7 ve 8. sorular aşağıdaki tabloya göre cevaplandırılacaktır.

7.

t	y_t
-3	12
-1	15
1	13
3	16

Yukarıda verilen zaman serisinin tahmin edilecek doğrusal trend denkleminde eğimi veren parametre tahmininin değeri kaçtır?

a. 0.05
b. 0.12
c. 0.50
d. 0.80
e. 1.1

8. Yukarıda verilen tabloya göre, zaman serisi için tahmin edilen trend denklemi $t_t = 14 + 0{,}5t$ ise serinin $t = 1$ için trend tahmin değeri kaçtır?

a. 12.50
b. 14.25
c. 14.50
d. 15.10
e. 15.50

9. Yıllık bir zaman serisinin bileşenleri çarpımsal model kullanılarak tahmin edilmek istenmektedir. $t = 1$ için gözlem değeri 11.625, trend tahmin değeri 10.642 hesaplanmıştır. Aynı döneme ilişkin konjonktürel bileşenin tahmin değeri kaçtır?

a. 0.09
b. 1.009
c. 1.092
d. 1.15
e. 1.18

10. Bir zaman serisinin t dönemine ilişkin mevsimsel bileşen tahmini 0.92 dir. Bu değeri, aşağıdakilerden hangisi doğru olarak ifade eder?

a. Mevsimsel bileşen t dönemindeki gözlem değerinin normalden %92 büyük değer almasına neden olmuştur.
b. Mevsimsel bileşen t dönemindeki gözlem değerinin normalden % 8 küçük değer almasına neden olmuştur.
c. Mevsimsel bileşen t dönemindeki gözlem değerinin normalden %8 büyük değer almasına neden olmuştur.
d. Mevsimsel bileşen t dönemindeki gözlem değerinin normalden %0,8 büyük değer almasına neden olmuştur.
e. Mevsimsel bileşen t dönemindeki gözlem değerinin normalden 0,02 küçük değer almasına neden olmuştur.

Yanıt Anahtarı

1. c
2. c
3. b
4. a
5. d
6. e
7. c
8. c
9. c
10. b

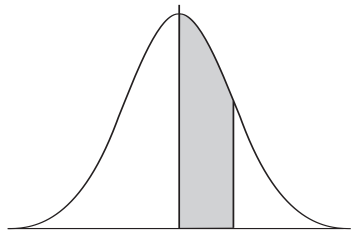
Yararlanılan Kaynaklar

BOWERMAN, Bruce L, O'Connell: **Forecasting and Time Series An Applied Approach**, 3rd Edition, Wadsworth Inc., California, 1993.

GÜRTAN, Kenan: **İstatistik ve Araştırma Metodları**, İ.Ü. Yayın No: 1941, Sermet Matbaası, İstanbul, 1974.

SERPER, Özer: **Uygulamalı İstatistik II**, Filiz Kitapevi, İstanbul, 1986.

TRYFOS, Peter: **Sampling Methods for Applied Research**, John Wiley and Sons Inc, NewYork, 1996.



EK 1: NORMAL EĞRİ ALANLARI TABLOSU

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.000000	0.003989	0.007989	0.011967	0.015953	0.019939	0.023922	0.027903	0.031881	0.035856
0.1	0.039828	0.043795	0.047758	0.051717	0.055670	0.059618	0.063559	0.067495	0.071424	0.075345
0.2	0.079260	0.083166	0.087064	0.090954	0.094835	0.098706	0.102568	0.106420	0.110261	0.114092
0.3	0.117911	0.121719	0.125516	0.129300	0.133072	0.136831	0.140576	0.144309	0.148027	0.151732
0.4	0.155422	0.159097	0.162757	0.166402	0.170031	0.173645	0.177242	0.180822	0.184386	0.187933
0.5	0.191462	0.194974	0.198468	0.201944	0.205402	0.208840	0.212260	0.215661	0.219043	0.222405
0.6	0.225747	0.229069	0.232371	0.235653	0.238914	0.242154	0.245373	0.248571	0.251748	0.254903
0.7	0.258036	0.261148	0.264238	0.267305	0.270350	0.273373	0.276373	0.279350	0.282305	0.285236
0.8	0.288145	0.291030	0.293892	0.296731	0.299546	0.302338	0.305106	0.307850	0.310570	0.313267
0.9	0.315940	0.318589	0.321214	0.323814	0.326391	0.328944	0.331472	0.333977	0.336457	0.338913
1.0	0.341345	0.343752	0.346136	0.348495	0.350830	0.353141	0.355428	0.357690	0.359929	0.362143
1.1	0.364334	0.366500	0.368643	0.370762	0.372857	0.374928	0.376976	0.378999	0.381000	0.382977
1.2	0.384930	0.386860	0.388767	0.390651	0.392512	0.394350	0.396165	0.397958	0.399727	0.401475
1.3	0.403199	0.404902	0.406582	0.408241	0.409877	0.411492	0.413085	0.414656	0.414207	0.417736
1.4	0.419243	0.420730	0.422196	0.423641	0.425066	0.426471	0.427855	0.429219	0.430563	0.431888
1.5	0.433193	0.434478	0.435744	0.436992	0.438220	0.439429	0.440620	0.441792	0.442947	0.444083
1.6	0.445201	0.446301	0.447384	0.448449	0.449497	0.450529	0.451543	0.452540	0.453521	0.454486
1.7	0.455435	0.456367	0.457284	0.458185	0.459071	0.459941	0.460796	0.461636	0.462462	0.463273
1.8	0.464070	0.464852	0.465621	0.466375	0.467116	0.467843	0.468557	0.469258	0.469946	0.470621
1.9	0.471284	0.471933	0.472571	0.473197	0.473810	0.474412	0.475002	0.475581	0.476148	0.476705
2.0	0.477250	0.477784	0.478308	0.478822	0.479325	0.479818	0.480301	0.480774	0.481237	0.481691
2.1	0.482136	0.482571	0.482997	0.483414	0.483823	0.484222	0.484614	0.484997	0.485371	0.485738
2.2	0.486097	0.486447	0.486791	0.487126	0.487455	0.487776	0.488089	0.488396	0.488696	0.489989
2.3	0.489276	0.489556	0.489830	0.490097	0.490358	0.490613	0.490863	0.491106	0.491344	0.491576
2.4	0.491802	0.492024	0.492240	0.492451	0.492656	0.492857	0.493053	0.493244	0.493131	0.493613
2.5	0.493790	0.493963	0.494132	0.494297	0.494457	0.494614	0.494766	0.494915	0.495060	0.495201
2.6	0.495339	0.495473	0.495603	0.495731	0.495855	0.495975	0.496093	0.496207	0.496319	0.496427
2.7	0.496533	0.496636	0.496736	0.496833	0.496928	0.497020	0.497110	0.497197	0.497282	0.497365
2.8	0.497445	0.497523	0.497599	0.497673	0.497744	0.497814	0.497882	0.497948	0.498012	0.498074
2.9	0.498134	0.498193	0.498250	0.498305	0.498359	0.498411	0.498462	0.498511	0.498559	0.498605
3.0	0.498650	0.498694	0.498736	0.498777	0.498817	0.498856	0.498893	0.498930	0.498965	0.498999

EK 2: "t" TABLOSU

Serbestlik derecesi	P = 0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.05	0.02	0.01
1	0.158	0.325	0.510	0.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.654
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	3.449
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	0.269	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
α	0.12566	0.25335	0.38532	0.52440	0.67449	0.84162	1.03643	1.28155	1.64485	1.95996	2.32634	2.57582

EK 3: χ^2 TABLOSU

α / r	0.995	0.990	0.975	0.950	0.050	0.025	0.010	0.005	0.001
1	0.00003927	0.000157	0.000982	0.003932	3.841455	5.023903	6.634891	7.879400	10.827360
2	0.010025	0,020100	0.050636	0.102586	5.991476	7.377779	9.210351	10.596530	13.915004
3	0.071723	0.114832	0.215795	0.351846	7.814725	9.348404	11.344882	12.838073	16.265959
4	0,206984	0.297107	0.484419	0.710724	9.487728	11.143262	13.276699	14.860166	18.466226
5	0.411751	0.554297	0.831209	1.145477	11.070483	12.832492	15.086317	16.749648	20.514651
6	0.675733	0.872083	1,.237342	1.635380	12.591577	14.449355	16.811872	18.547513	22.457479
7	0.989251	1.239032	1,689864	2.167349	14.067127	16.012774	18.475324	20.277738	24.321296
8	1.344403	1,646506	2,179725	2.732633	15.507312	17.534545	20.090159	21.954861	26.123931
9	1.734911	2.087889	2,700389	3.325115	16.918960	19.022778	21.666048	23.589275	27.876731
10	2.155845	2.558199	3.246963	3.940295	18.307029	20.483201	23.209287	25.188055	29.587885
11	2.603202	3.053496	3.815742	4.574809	19.675153	21.920023	24.725022	26.756864	31.263507
12	3.073785	3.570551	4.403778	5.226028	21.026055	23.336660	26.216964	28.299660	32.909230
13	3.565042	4.106900	5.008738	5.891861	22.362027	24.735581	27.688184	29.819318	34.527367
14	4.074659	4.660415	5.628724	6.570632	23.684782	26.118935	29.141163	31.319425	36.123867
15	4.600874	5.229356	6.262123	7.260935	24.995797	27.488365	30.577951	32.801491	37.697774
16	5.142164	5.812197	6.907664	7.961639	26.296221	28.845325	31.999861	34.267053	39.251776
17	5.697274	6.407742	7.564179	8.671754	27.587100	30.190983	33.408717	35.718378	40.791109
18	6.264766	7.014903	8.230737	9.390448	28.869321	31.526410	34.805237	37.156386	42.311948
19	6.843923	7.632698	8.906514	10.117006	30.143505	32.852337	36.190775	38.582122	43.819365
20	7.433811	8.260368	9.590772	10.850799	31.410420	34.169581	37.566272	39.996856	45.314218
21	8.033602	8.897172	10.282907	11.591316	32.670558	35.478859	38.932232	41.400943	46.796271
22	8.642681	9.542494	10.982330	12.338009	33.924460	36.780678	40.289448	42.795664	48.267624
23	9.260383	10.195689	11.688534	13.090505	35.172460	38.075609	41.638334	44.181385	49.727643
24	9.886199	10.856349	12.401146	13.848422	36.415026	39.364060	42.979781	45.558363	51.178969
25	10.519647	11.523951	13.119707	14.611396	37.652489	40.646498	44.314014	46.927966	52.618738
26	11.160218	12.198177	13.843881	15.379163	38.885130	41.923138	45.641636	48.289777	54.051136
27	11.807655	12.878468	14.573373	16.151395	40.113266	43.194521	46.962837	49.645035	55.475080
28	12.461281	13.564666	15.307854	16.927876	41.337152	44.460790	48.278166	50.993559	56.891756
29	13.121067	14.256406	16.047051	17.708381	42.556948	45.722279	49.587829	52.335495	58.300642
30	13.786682	14.953464	16.790756	18.492667	43.772954	46.979218	50.892181	53.671868	59.702212
40	20.706577	22.164201	24.433058	26.509296	55.758487	59.341679	63.690771	66.766047	73.402900
50	27.990825	29.706725	32.357385	34.764236	67.504805	71.420194	76.153802	79.489839	86.660312
100	67.327533	70.064995	74.221882	77.929442	124.342101	129.561252	135.806891	140.169714	149.448789

Sözlük

A

Alfa (α) Tipi Hata (I. Tip Hata): Doğru bir hipotezin (H_0 doğru iken) test sonucunda reddedilmesi halinde işlenecek hata.

Ana Kütle: Yığın olay niteliğinde ve aynı cins birimlerin oluşturduğu topluluk.

Anlam(lılık) Düzeyi: Doğru bir hipotezin test sonucunda reddedilmesinin maksimum olasılığı.

Aralık Tahminlemesi: Araştırılan ana kütle parametresini istenilen bir olasılıkla bir aralık içinde tahminleme tekniği.

Aritmetik Ortalama: Bir seriyi oluşturan değerler toplamının, gözlem sayısına oranı.

Ayrık Olaylar: Birlikte ortaya çıkamayan olaylar.

B

Bağımsız Olaylar: Ortaya çıkıp çıkmaması, başka bir olayın ortaya çıkmasını ya da çıkmamasını zorunlu hale getiremeyen olaylar.

Basit İndeks: Tek bir madde için hesaplanan indeks.

Basit Olay: Örneklem uzayında tek örneklem noktası olan olay.

Basit Seri: Gözlem değerlerinin büyüklüklerine göre küçükten büyüğe ya da büyükten küçüğe doğru sıralanmasıyla oluşturulan seri.

Belirlilik Katsayısı: Regresyon analizinde bağımlı değişkendeki değişimin ne kadarının bağımsız değişkence açıklandığını belirleyen katsayı.

Beta (β) Tipi Hata (II. Tip Hata): Doğru olmayan bir hipotezin (H_0 doğru değilken) test sonucunda kabul edilmesi halinde işlenecek hata.

Bileşik İndeks: İki ya da daha çok maddeyi kapsayan indeks.

Bileşik Olay: Örneklem uzayında birden çok örneklem noktası olan olay.

Bileşik Seri: Birimlerin birden fazla değişkene göre dağılımını bir arada gösteren seri.

Birikimli Seri: Bir frekans dağılımında her sınıfın frekansına bir önceki sınıfın frekansı eklenerek oluşturulan seri.

D

Dağılma Serisi: Gözlem sonuçlarının maddesel bir değişkenin şıklarına göre sıralanmasıyla oluşturulan seri.

Değişim Aralığı: Bir serideki en büyük değer ile küçük değer arasındaki fark.

Değişken: İstatistik birimlerinin sahip olduğu özellikler.

Değişken Esaslı İndeks: Bir zaman indeksinde her değerin, bir önceki dönemin değerine göre oransal değişimleri.

Duyarlı Olmayan Ortalama: Serideki aşırı değerlerden etkilenmeyen ortalama.

Duyarlı Ortalama: Bir seriyi oluşturan tüm değerlerden etkilenen ortalama.

F

Fisher İndeksi: Laspeyres ve Paasche indekslerinin geometrik ortalaması.

Frekans Serisi: Bir serideki, gözlem değerlerinin kaç kez tekrarlandığını, ilgili gözlem değerlerinin yanına, kaydedilerek oluşturulan seri.

G

Geometrik Ortalama: Bir seriyi oluşturan değerlerin çarpımının, gözlem değerleri sayısına eşit mertebeden kökü.

H

Hareketli Ortalama: Zaman serilerinin mevsimsel ve rassal bileşenlerin etkilerinden arındırmak suretiyle, bu serilerin genel eğilimini elde etmek amacıyla başvurulan teknik.

Histogram: Sınıflandırılmış bir seriye ilişkin, alanı ilgili sınıfın frekansına ve tabanı da ilgili sınıfın aralığa eşit, birbirine bitişik dikdörtgenlerden oluşan grafik.

İ

İndeks: Bir istatiksel olaya ilişkin gözlem değerlerinin zaman ya da mekana göre gösterdiği oransal değişimler.

İstatistik:
1. Belirlenen amaç ya da amaçlar doğrultusunda gözlenen yığın olaylardan derlenen verilerin işlenerek, ilgili olayların oluşturulduğu yığınların bilimsel olarak incelenmesinde kullanılan yöntemler bütünü.
2. Örnekleme ilişkin sayısal karateristikler.

İstatistik Birimi: Ölçülebilen ya da sayılabilen yığın olay niteliğindeki her olay.

İstatistik Serisi: Gözlem değerlerinin büyüklüklerine göre oluşturulan dizi.

İstatistiksel Hipotez: Herhangi bir ana kütle parametresine ilişkin olarak ileri sürülen ve doğruluğu olasılık kurallarıyla araştırabilen önerme.

K

Kareli Ortalama: Bir seriyi oluşturan değerlerin karelerinin toplamının gözlem sayısına oranı.

Karşıt (H_1) Hipotez: İlgilenilen ana kütle parametresinin bilinen değerinde, istatiksel olarak anlamlı farkların beklendiğini ifade eden hipotez.

Kontenjans Katsayısı: Sayısal olmayan değişkenler arasındaki ilişkinin derecesini gösteren katsayı.

Korelasyon Katsayısı: İki değişken arasındaki doğrusal ilişkinin yönünü ve derecesini belirleyen (gösteren) katsayı.

L

Laspeyres İndeksi: Ağırlıklı indekslerde ağırlık katsayılarının başlangıç zamanındaki (ya da yerindeki) değerlerden seçilmesiyle oluşturulan indeks.

M

Medyan: Bir seride, seriyi iki eşit kısma bölen değer.

Mekan İndeksi: Mekan serisi ile verilen bir istatistiksel olayın aritmetik ortalamaya göre değişimleri.

Mekan Serisi: Gözlem sonuçlarının ülke, bölge, şehir ya da köy gibi bir mekan (yer) değişkeninin şıklarına göre sıralanmasıyla oluşturulan seri.

Mod: Bir seride en çok tekrarlanan değer.

N

Nokta Tahminlemesi: Bir rassal örneklemden elde edilen istatistiğin değerini, ilgili ana kütle parametre değerine eşit kabul eden tahminleme tekniği.

O-Ö

Olay: Bir deneyin bir ya da daha çok sonucunu içeren küme.

Ortalama: Bir istatistik serisindeki gözlem değerlerinin, etrafında toplanma eğilimi gösterdiği değer.

Örneklem: Ana kütleden uygun tekniklerle seçilen birimlerden oluşan alt topluluk.

Örneklem Noktası: Örneklem uzayının her bir elemanı.

Örneklem Uzayı: Deney ya da gözlem sonuçlarının tümünü içeren küme.

P

Paasche İndeksi: Ağırlıklı indekslerde ağırlık katsayılarının indeks devresindeki (ya da yerindeki) değerlerden seçilmesiyle oluşturulan indeks.

R

Regresyon: Değişkenler arasındaki ortalama ilişkinin matematiksel bir fonksiyonla ifade edilmesi.

S-Ş

Sabit Esaslı indeks: Bir zaman indeksinde devrelerden birisinin değerinin temel kabul edilerek, diğer devrelerin değerlerinin seçilen temel devreye göre oransal değişimi.

Sıfır (H_0) Hipotezi: İlgilenilen ana kütle parametresinin bilinen değerinde, herhangi bir farkın beklenmediğini ifade eden hipotez.

Sınıflandırılmış Seri: Deney ya da gözlem sonuçlarının belirli aralıklar içinde kalan şıklara göre düzenlenmesiyle oluşturulan seri.

Standart Sapma: Bir seriyi oluşturan değerlerin aritmetik ortalamadan farklarının kareli ortalaması.

Şık: Bir değişkenin aldığı değer.

T

Tahminleme: Bir rassal örneklemden elde edilen istatistikler yardımıyla, örneğin çekildiği ana kütleye ilişkin parametre değerlerini araştırmak.

Tamamlayıcı Olay: Bir olayın ortaya çıkmaması ile tanımlanan olay.

Tartılı Ortalama: Gözlem değerleri arasındaki önem derecelerini içeren ortalama.

Tipik Olay: Ait olduğu kümedeki tüm olayları tek başına temsil edebilen olay.

Trend: Zaman serilerine ilişkin gözlem değerlerinin uzun dönemde artma ya da azalma yönünde gösterdiği genel eğilim.

Y

Yığın Olay: Ait olduğu kümedeki olayları tek başına temsil edemeyen olay.

Z

Zaman İndeksi: Zaman serisi ile verilen bir istatistiksel olayın, zaman içindeki oransal değişimleri.

Zaman Serisi: Gözlem sonuçlarının yıl, ay, hafta, gün ya da saat gibi bir zaman değişkeninin şıklarına göre sıralanmasıyla oluşturulan seri.

Dizin

A

B

C-Ç

D

K

L

M

N

O-Ö

P-R

S-Ş

T

U-Ü

V