



İSTATİSTİK

MALİYE UZAKTAN EĞİTİM

DR. ÖĞR. ÜYESİ LEYLA İŞBİLEN YÜCEL

İSTANBUL ÜNİVERSİTESİ AÇIK VE UZAKTAN EĞİTİM FAKÜLTESİ

İSTANBUL ÜNİVERSİTESİ AÇIK VE UZAKTAN EĞİTİM FAKÜLTESİ

MALİYE LİSANS PROGRAMI



İSTATİSTİK

DR. ÖĞR. ÜYESİ LEYLA İŞBİLEN YÜCEL

Yazar Notu

Elinizdeki bu eser, İstanbul Üniversitesi Açık ve Uzaktan Eğitim Fakültesi'nde okutulmak için hazırlanmış **bir ders notu niteliğindedir.**

ÖNSÖZ

Teknolojinin hayatımızın her alanına girmeye başlamasıyla her birimiz adeta veri ve bilgi bombardımanı altında kalmış bulunuyoruz. Üstelik bu bilgi kirliliği denilebilecek durum, ancak ve ancak bilginin doğru yönetilebilmesiyle aşılabilecek bir durumdur. Verinin doğru okunması, amaçlarımıza hizmet edip edemeyeceğinin ortaya konması, kısacası bu veri deryasından kendimize nasıl fayda sağlayacağımızı bize gösterecek olan başlıca kaynak istatistik okuryazarlığıdır. Dünyada istatistiğin kullanımının her geçen gün bireyler ve firmalar tarafından giderek daha da fazla kullanılması istatistiği yükselen bir değer haline getirmiştir. Örneğin ASA (American Statistical Association) 2013 yılını dünya İstatistik yılı ilan etmiştir. North Carolina Üniversitesi'nden Marie Davidian'ın 19 Aralık 2013'te JASA (Journal of American Statistical Association) dergisinde yayınlanan "The International Year of Statistics: A Celebration and A Call to Action" adlı makalesinde istatistiğin dünyada yükselen bir değer olmasından bahsetmektedir. İstatistiğin doğru kullanımı ve gücü hakkında son yıllarda en büyük katkıyı 2012 Amerika devlet başkanlığı seçimlerinden önce yaptığı tutarlı tahminlerle Nate Silver yapmıştır. "The Signal and the Noise: Why So Many Predictions Fail But Some Don't?" adlı kitabı bir anda en çok satanlar arasına girerek, tüm siyasilerin ve halkın ilgi odağı haline gelmiştir.

Ülkemizde çok uzak değil, bundan 10-15 yıl önce istatistik denilince akla, futboldaki özet sonuçlar veya sokakta yapılan anketler gelmekteydi. Oysa günümüzde tıpkı dünyada olduğu gibi istatistik hayatın her alanına girmiş ve etkin bir biçimde kullanılmaya başlanmıştır.

İÇİNDEKİLER

ÖNSÖZ.....	I
İÇİNDEKİLER.....	II
KISALTMALAR	VI
YAZAR NOTU	VII
1. BÖLÜM İSTATİSTİĞE GİRİŞ, TEMEL KONULAR VE KAVRAMLAR, İSTATİSTİĞİN TARİHTEN GÜNÜMÜZE GELİŞİMİ.....	1
1.1. İstatistik Nedir?	7
1.2. İstatistiğin Temel Konuları.....	8
1.3. İstatistiğin Tarihçesi	10
1.3.1. Tarihten Günümüze Öne Çıkan İstatistiksel Çalışmalar.....	12
1.4. Türkiye İstatistik Kurumu Hakkında Genel Bilgiler	12
1.5. Günümüzde İstatistik.....	13
2. VERİLERİN DERLENMESİ, DÜZENLENMESİ VE SUNUMU	19
2.1. Verilerin Derlenmesi	25
2.1.1. Veri Kaynakları.....	25
2.1.2. Verilerin Elde Edilmesi	25
2.2. Verilerin Düzenlenmesi.....	27
2.3. Grafikler.....	28
2.3.1. Basit Serinin Sunumu İçin Nokta Diyagramı	28
2.3.2. Frekans Serisinin Sunumu İçin Çizgi Diyagramı	29
2.3.3. Sınıflanmış Serinin Sunumu İçin Histogram	31
2.3.4. Box-Whisker Grafiği	36
2.3.5. Daire (pasta) Grafiği.....	36
3. MERKEZİ EĞİLİM ÖLÇÜLERİ.....	43

3. 1. Merkezi eğilim ölçüleri	49
3.1.1. Analitik (duyarlı) ortalamalar	49
3.1.2. Analitik olmayan (duyarsız) ortalama	50
3.1.1.2. Tartılı Ortalama (Ağırlıklı Ortalama)	54
3.1.1.3. Kareli Ortalama.....	55
3.1.1.4. Geometrik Ortalama	57
3.1.1.5. Harmonik Ortalama	60
4. MERKEZİ DEĞİŞİM ÖLÇÜLERİ.....	67
4. 1. Merkezi Değişim Ölçüleri	73
4.1.2. Standart Sapma	75
4.1.3. Değişim Katsayısı	81
4.2. Chebyshev Teoremi.....	82
5. MOMENTLER.....	88
5.1. Momentler	94
5.1.1. Sıfıra Göre Momentler (basit moment)	94
5.1.2. Ortalamaya Göre Momentler	94
5.2. Momentler Yoluyla Çarpıklık Ölçümü	95
5.3. Momentler Yoluyla Basıklık Ölçümü	96
5.4. König Teoremi.....	102
6. ÇEKİLİŞLERLE İLGİLİ BAĞIMLILIK VE BAĞIMSIZLIK KAVRAMLARI, OLASILIĞIN KLASİK VE FREKANS TANIMI, KOŞULLU OLASILIK, OLASILIK FONKSİYONU, DAĞILIM FONKSİYONU.....	107
6.1. Çekilişlerle İlgili bağımlılık ve Bağımsızlık kavramları	113
6.2. Olasılığın Klasik ve Frekans Tanımı	114
6.2.1. Klasik yaklaşım.....	114
6.2.2. Frekans yaklaşımı	114

6.3. Koşullu Olasılık.....	116
6.4. Kesikli Rd'nin Olasılık Fonksiyonu	120
6.5. Dağılım Fonksiyonu:	122
7. BİNOM DAĞILIMI VE POISSON DAĞILIMI	130
7.1. Bazı Özel Kesikli Dağılımlar	136
7.1.1. Binom Dağılımı	136
7.1.2. Poisson Dağılımı.....	140
8. SÜREKLİ DAĞILIMLAR- NORMAL DAĞILIM.....	148
8.1. SÜREKLİ DAĞILIMLAR.....	154
8.2. Sürekli Bir Fonksiyonun Olasılık Yoğunluk Fonksiyonu Olabilmesi İçin Gerekli Şartlar	155
8.3. NORMAL DAĞILIM	155
8.4. Z tablosunun (normal dağılım tablosu) özellikleri ve okunuşu.....	160
9. GÜVEN ARALIKLARI: İKİ ANAKÜTLE	175
9.1. ORTALAMALAR ARASI FARKLAR İÇİN GÜVEN ARALIĞI	181
9.2. İKİ ORAN FARKI İÇİN GÜVEN ARALIĞI.....	182
9.3. Örneklem Büyüklüğünün Belirlenmesi ve Hata.....	188
10. HİPOTEZ TESTLERİ	194
10.1. HİPOTEZ TESTLERİ	200
10.2. Hipotezlerin İfade Edilmesi.....	200
10.3. I. Tür ve II. Tür Hata	203
10.4. Anakütle ortalamasına ilişkin Hipotez Testleri	204
11. GENEL TEKRAR	218
11.1. Soru Çözümü	224
12. REGRESYON ANALİZİ II	235

12.1.REGRESYON DOĞRUSUNUN ÖZELLİKLERİ VE HATA (ARTIK) KAVRAMI	241
12.1.Regresyon Doğrusunun Özellikleri	242
12.2. Regresyon Denklemi Parametrenin Hipotez Testi ve Güven Aralığı.....	244
13. Kİ-KARE TESTLERİ	258
13.1. Kİ-KARE TESTLERİNİ NEDEN KULLANIRIZ?.....	264
13.2. Kİ-KARE BAĞIMSIZLIK TESTİ.....	265
11.3. Kİ-KARE HOMOJENLİK TESTİ.....	269
11.4. Kİ-KARE UYUM TESTİ.....	271
11.5. KONTENJANS KATSAYISI	273
14. ENDEKS KAVRAMI	280
14.1. ENDEKS (İNDEKS)	286
14.2. BASİT VE BİLEŞİK ENDEKSLER.....	291
14.3. LASPEYRES ENDEKSİ.....	296
14.4. PAASCHE ENDEKSİ.....	298
14.5. FİŞHER ENDEKSİ	299
KAYNAKLAR.....	306

KISALTMALAR

YAZAR NOTU

1. BÖLÜM İSTATİSTİĞE GİRİŞ, TEMEL KONULAR VE KAVRAMLAR, İSTATİSTİĞİN TARİHTEN GÜNÜMÜZE GELİŞİMİ

Bu Bölümde Neler Öğreneceğiz?

- İstatistik nedir?
- İstatistiğin temel konuları (betimsel ve çıkarımsal istatistik)
- İstatistiğin tarihçesi
- Tarihten günümüze önemli istatistik çalışmaları
- TÜİK hakkında genel bilgiler
- Günümüzde istatistik

Bölüm Hakkında İlgi Oluşturan Sorular

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
İstatistiğe giriş ve temel konular, istatistiğin tarihsel gelişimi, ülkemizde istatistik kurumları	İstatistik biliminin tanıtılması, amacı ve kullanım alanları ve katkıları	Metinler, çözümlü problemler, alıştırmalar ve interaktif materyaller ile konuların daha kolay anlaşılması sağlanacaktır.

Anahtar Kavramlar

- İstatistik
- Tesadüfi olay
- Betimsel istatistik
- Çıkarımsal istatistik
- Status-devlet
- Bernoulli
- Gauss
- Laplace
- Moivre
- Adolphe Quotelet
- Karl Pearson
- TÜİK

Giriş

1.1. İstatistik Nedir?

Tanım 1: İstatistik, belirli bir amaç için veri toplama, tablo ve grafiklerle özetleme, sonuçları yorumlama, sonuçların güven düzeylerini açıklama, örneklemelerden elde edilen (hesaplanan) sonuçları kitle için genelleme (çıkarsama), karakteristikler arasındaki ilişkileri araştırma, çeşitli konularda geleceğe ilişkin projeksiyon yapma, deney tasarlama ve düzenleme vb. kapsayan bir yöntemler topluluğudur.¹

Tanım 2: İstatistik verileri işler ve özetler, araştırmacılara yol gösterir. Değişkenler arasındaki ilişkileri incelememize, tahminler ve öngörüler yapmamıza, doğru karar vermemize yarar.²

Tanım 3: Belirli amaç ya da amaçlar doğrultusunda gözlenen yığın olaylardan sayısal verilerin işlenerek, ilgili olayların oluşturduğu yığınların bilimsel olarak incelenmesinde kullanılan teknik ve yöntemler bilimi olarak tanımlanabilir.³

Tanım 4: İstatistik; gözlem ve ölçme sonucunda elde edilen sayısal (rakamsal) verilerin tablolar veya grafikler halinde sunulması ve bunların karar alma sürecinde nasıl kullanılacağına ilişkin yöntemler sunar. İstatistiğe konu olacak olaylar gözlemlenebilir ve ölçülebilir olmalıdır. İstatistikler belirli ölçümler yapılarak elde edilmişlerdir yani istatistikler rastgele bulunmuş sayılar değildirler. Belirli bir kitleyi oluşturan birimlerin gözlemlenmesiyle elde edilen, ölçülebilen veya sayılabilen nicel bilgileri ifade etmektedirler. Bu nedenle; logaritma cetveli, telefon rehberi, milli piyango sonuçları birer istatistik değildirler.

Tanım 5: İstatistik; olasılığa dayalı yani sonucun baştan kesin olarak bilinmediği rastlantısal (tesadüfi) durumlar üzerinde çalışırken, belirsizlik altında karar vermeyi sağlar.

Tanım 6: Verilerin derlenmesi, işlenmesi ve yorumlanmasına yönelik yöntemlerin tümü. (İşleme ve yorum verileri temsil edebileceği düşünülen kuramsal dağılım ve

¹ (Özer Serper, *Uygulamalı İstatistik 1 -2*)

² (Necmi Gürsakal, *Betimsel İstatistik*, syf. 6)

³ (Murat Açıköğretim Yayınları, *İstatistik*)

modellere geiři de ermek zere matematiksel istatistik kapsamına girer)⁴. İstatistiksel arařtırmalara birkaç rnek verecek olursak, rneėin, bir arařtırmacı, hamile iken diyetisyen kontrolnde iyi ve kaliteli beslenmenin doėacak bebeėin genel saėlık lmlerinde, dzensiz ve bilinsiz beslenen bir annenin bebeėine kıyasla olumlu bir etkisi var mıdır? Fazla kilolu olmanın yařam sresi zerinde gerekten olumsuz etkileri var mıdır? Sigara kanser iliřkisinin boyutu nelerdir? Kanser olmayı tetikleyen diėer faktrler neler olabilir ve risk sıralaması nasıldır?

1.2. İstatistiėin Temel Konuları

İstatistik konuları bakımından ikiye ayrılır; Betimsel (tasviri) istatistik ve ıkarımsal istatistik.

1.2.1. Betimsel İstatistik

Verilerin toplanması, yorumlanıp zetlenmesi gibi konularla ilgilenir. Tanımsal istatistik, verilerin bir takım sayısal ve(ya) grafiksel yntemlerin kullanımı ile bilgiye dnřtrlmesini ve karar verme srecinde kullanılmasını saėlar.⁵

İstatistik, gemiři ve iinde bulunulan durumu tanımlayarak bir veri kmesine iliřkin zet deėerler ve grafikler ortaya koyduėunda betimsel istatistik (descriptive statistics) adını alır. zetleme ve verilerin yoėunlařtırılması diėer bir deyiřle verilerin hacimleri azaltılarak kullanım deėerlerinin arttırılması betimsel istatistiėin konusuna girer. Tablolar ve grafikler yardımı ile verilerin zetlenmesi ve ok sayıda sayıdan oluřan bir veri grubunun “ortalama” gibi tek bir sayıya indirgenmesi yine bu alan iin geerlidir. Kısaca betimsel istatistik bir veri kmesinde bulunan bilgiyi sayısal ve grafiksel yntemleri kullanarak zetler ve sunar.⁶

Betimsel amalı istatistik kitledeki tm birimlerden ilgili deėiřken ya da deėiřkenler iin veri toplandıėında bunları kullanarak kitleyi zetlemeyi (betimlemeyi)

⁴ (Byk Larousse, Cilt: 10, Syf: 5872)

⁵ Onur zsoy, İktisatılar ve İřletmeciler iin İstatistik, Excel Uygulamalı, 2010, Ankara, Syf: 4.

⁶ Necmi Grsakal, Betimsel İstatistik, Syf:20.

amaçlar. Bu ise frekans dağılımı oluşturularak, grafikler çizilerek ya da parametreler (kitle ortalaması ve varyansı) hesaplanarak yapılır.⁷

1.2.2. Çıkarımsal İstatistik

Tümevarımsal amaçlı istatistik kitle(ler)den rastgele seçilen örneklem(ler)den toplanan verileri kullanarak kitle(ler) nin parametrelerini tahmin etmeyi veya parametrelerle ilgili olan savların doğru olup olmadığının araştırılmasını amaçlar. Yanlı örneklem gözlemlendiğinde belirsizlikler içeren kitle hakkındaki önermelere geçmek için geliştirilen süreçle istatistiksel sonuç çıkarır.⁸

Çıkarımsal istatistik, araştırma sürecinin betimsel istatistiğin bıraktığı yerden devam ettirilmesi suretiyle, çalışmanın amacına uygun bir şekilde daha ileri tekniklerle (olasılık başta olmak üzere) karar verme ve kitleye dair çıkarsamalar yapma sürecidir. Bu çıkarsamalar örnekleme dayalı olarak yapıldıkları için belirli bir hata düzeyinde ifade edilirler. (Kitle **N** adet birimden oluşan teorik topluluktur, aslında tüm birimlere ulaşmak olasıdır ancak zaman ve maliyet kısıdından dolayı genellikle istatistiksel çalışmalar kitleyi iyi bir biçimde temsil eden daha küçük alt gruplar yani örneklemeler üzerinden yürütülür.

Örneklemin büyüklüğü **n** ile gösterilir. O halde **N** büyüklüğündeki bir kitleden **n** büyüklüğünde tane mümkün örneklem çekilebilir($C_n^N = \frac{N!}{(N-n)!n!}$). Bu örneklemeler büyüklük yani hacim olarak birbirinin aynı fakat içerdikleri birimler bakımından birbirinden farklıdır, kombinasyonun doğası gereği seçilen birim yerine konmamaktadır, bu nedenle mümkün örneklem sayısı hesaplanırken permütasyon yerine kombinasyon kullanılmaktadır.)

Örnekleme dayanan çıkarımlar mutlak doğru olamayacağından “**olasılık**” kelimesi kullanılarak ifade edilir. Çıkarımsal istatistiğin genel olarak kullandığı yöntem “tümevarım” dır. Önceleri betimsel amaçlar taşıyan istatistik, 17-18 ve 19. yy lar boyunca Bernoulli, Gauss, Laplace, Moivre gibi ünlü matematikçilerin olasılık teorisini geliştirmeleriyle çıkarımsal istatistiğin temellerini atmışlardır.⁹

⁷ Fikri Akdeniz, Olasılık ve İstatistik, Syf: 288.

⁸ Fikri Akdeniz, Olasılık ve İstatistik, Syf: 288.

⁹ Özer Serper, Uygulamalı İstatistik 1- 2.

****Her şey sabit, belirli ve kontrol altında olsaydı, yani değişkenlik olmasaydı
istatistiğe gerek kalmazdı!!!*

Yukarıdaki cümle istatistiğin özünü ifade etmektedir. Her şey bir ihtiyacı gidermek için ortaya çıkıp kabul gördüğüne göre, istatistik de, gerçek hayatı olduğu gibi ele alabilmeyi sağlamıştır. Olasılık teorisinin katkılarıyla belirsizliği ölçebilmeyi, geleceğe dair tahminler ve çıkarımlar yapabilmeyi olanaklı kılmıştır. Yaşamın belirsizlikleri insanoğlunu her zaman 2 kere 2'nin dört etmeyeceği gerçeğiyle karşı karşıya bırakmıştır. Özellikle sosyal bilimlerde olayları salt pozitif bilimlerle açıklamak yetersiz kalmaktadır. İstatistik böyle durumlarda ölçümlemeyi sağlayabilen yöntem ve teknikler bütünü olarak öne çıkmaktadır.

1.3. İstatistiğin Tarihçesi

“İstatistik kelimesi Almanya’da Achenwall tarafından 1748’de kullanılmaya başlanmışsa da, istatistik çalışmaları çok eski çağlara kadar uzanır. Tacitus, imparatorluk zenginlikleri üzerinde Augustus’un geniş bir anket yaptığını yazar. Böylece askerler, gemiler her çeşit gelir kaynakları sayılmış ve kamu gelirleri tespit edilmiştir.

Bütün Ortaçağ boyunca ve 17. Yy’e kadar istatistik sadece tasviri olarak kalmıştır. Bu alanda iki okul ortaya çıktı: Göttingen Üniversitesi profesörü ve Almanlar tarafından istatistiğin babası olarak bilinen Achenwall’in temsil ettiği tasviri okul ve bazı sosyal olayların yaklaşık düzenliliğinden öngörüler ve kanunlar çıkarmayı deneyen siyasi matematikçiler okulu. 18. Yy’de Fransız Desparcieux ve İsviçreli Wargentin toplum olaylarını öngörmenin pratik önemini gösteren ve çok gelişmekte olan sigorta sanayinin hareket noktasını teşkil eden ilk ölüm tablolarını (mortalite tablosu) düzenlemişlerdi.

Fransa’da ilk resmi istatistik bürosu 17. Yy’in sonlarında kurulmuş ve içişleri bakanı Chaptal 1801’de genel nüfus sayımını gerçekleştirmişti. Türkiye’de ilk resmi istatistik bürosu 1933’te kuruldu.

Jacques Bernoulli ve özellikle Laplace, tasviri sayısal bilgilere matematik bilgileri ilave ederek sonuçların değerlendirilmesinde olasılık hesap imkânlarının kullanılmasını araştırdılar. Birincisi ünlü “büyük sayılar kanunu” nu ortaya attı, ikincisi ise “analitik olasılık teorisi” adlı eserinde karmaşık sebepli doğal olayların incelenmesinde olasılık hesaplarından sağlanacak faydaları açıkça belirtti. Quotelet, çevresinde bütün diğer insanların

yer alacağı hayali bir “ortalama insan” bulmak maksadıyla bu yöntemin uygulanışını insanların fiziki, ahlaki ve düşünsel özelliklerine uyguladı.”¹⁰

“Quetelet bütün sosyal olayları birey etrafında toplamıştır. Daha ileri giderek istatistiği bütün sosyal alanlara hatta ahlaki olaylara da yaymıştır. İlk “sosyal kriminolog” olarak adlandırılmaktadır. 1836 yılında Quetelet, “Toplum suçu hazırlar, Suçlu ise ancak bir araçtır” demiştir. Bugün Quetelet modern istatistiğin kurucusu olarak kabul edilmekle birlikte, ileri sürdüğü sistem ve vardığı sonuçlar reddedilmektedir. Quetelet’ten sonraki gelişmeler istatistik yöntembiliminde matematiğin daha yaygın bir şekilde uygulanmasını sağlamış ve matematiksel istatistik disiplininin meydana çıkmasına yol açmıştır.

İstatistik Türklere de yabancı değildir. Selçuklular döneminde İran’da İlhanlılar döneminde Hintte nüfus sayımları düzenlenmişti. Öte yandan Osmanlı İmparatorluğu’nun yönetim sistemi nüfus ve arazi hakkında düzenli biçimde bilgi toplamasını gerektiriyordu. Bu nedenle imparatorluk zamanında 30-40 yıl gibi aralıklarla nüfus sayımları ve arazi yazımları yapılmaktaydı. Bu istatistiksel işlemler sırf vergi toplanması amacıyla yapıldığı için sadece vergi mükellefleri göz önünde bulunduruluyordu. İmparatorluğun gerileme devrinde bu sayımların arkası kesilmiş, modern tekniklere dayanan istatistikler de 20. Yy’e kadar ülkeye girememiştir. Gerçi bu gibi istatistiklerin düzenlenmesi yolunda 19. Yy’de bazı çalışmalar yapılmıştır. Örneğin orduyu modernleştirmek amacıyla askere alınabilecek erkeklerin sayısı öğrenilmek istenmiş ve böylece ilk sayım 1831’de gerçekleştirilmiştir. Ancak bu sayım yalnız erkekleri dikkate aldığı gibi ülkenin bütünü de kapsamamıştır. Nitekim sayım yapılan yerler Rumeli ve Anadolu sancakları ve kasabalarıdır. Dolayısıyla bu sayım bugünkü anlamıyla gerçek bir nüfus sayımı olmaktan uzak kalmıştır. 1844’te yapılan ikinci nüfus sayımının amacı ise kimlik belgesi verilecek vatandaşların belirlenmesi idi. Bu nedenle kadınlar da erkeklerle birlikte sayılmıştır. Ne var ki askerlik korkusuyla birçok erkek sayım dışında kalmıştır.

19. Yy’de ülke nüfusu hakkında bilgi verecek sonucu sayım 1874’te başlamış ve araya giren Rus savaşı nedeniyle 1884’e kadar sürmüştür. Nüfus kütüklerini düzenleme ve halka nüfus tezkeresi verme amaçlarıyla başlatılan bu sayımın uygulanmasında da bazı aksaklıklar görülmüştür. Birçok teknik hata içeren ve kısmi nitelikteki bu sayımlar yanında ölüm ve doğumlar gibi nüfus hareketlerini suçları ithalat ve ihracatı belirlemek amacıyla çalışmalar yapılmıştır. Ne var ki bu çabalar da sonuçsuz kalmıştır.”¹¹

¹⁰ *Meydan Larousse, Cilt: 6, Syf: 511-512)*

¹¹ *(Özer Serper, Uygulamalı İstatistik 1 – 2)*

1.3.1. Tarihten Günümüze Öne Çıkan İstatistiksel Çalışmalar

- Antikçağ'da Çin'de, Mısır'da, Yunanistan'da nüfus ve bu nüfusun maddi yaşam koşullarına ilişkin sayısal verilere sahip olma gereksinimi izlerine rastlanmıştır.
- Roma'da düzenli nüfus sayımlarına rastlanır. Roma'da önce 5 yıl, sonra da 10 yıl ve 15 yıl gibi düzenli aralıklarla yapılan ve “census” adı verilen bu sayımlar 600 yıl kadar sürmüştür. Her Romalı kendi ve babasının adını, servetini, arazisini, kölelerinin sayısını bildirmek zorunda idi.¹²
- 1662'de Graunt erkek çocuklarının doğumlarının kız çocuklarının doğumlarına oranı gibi değişmezlikleri ortaya çıkardı.
- 1742'de Edmond Halley çağdaş aktüer (hayat sigortacılığı) çalışmalarının temelini oluşturan bir hayat tablosu yaptı.
- 1767'de Süssmilch erkek doğum oranı ve 20 yaşına dek erkek çocuk oranındaki gelişme üzerine önemli çalışmalar yayınladı.
- Pierre Simon De Laplace (1749-1827) “Çözümsel Olasılık Kuramı” (1812) adlı eserinde, nedenlerinin tümünü bilip tek tek çözümlenemeyecek kadar karmaşık olan doğal olayların incelenmesinde bu kuramdan sağlanacak yararları ortaya koydu. Önceleri betimsel amaçlar taşıyan istatistik, 17-18 ve 19. yy lar boyunca Bernoulli, Gauss, Laplace, Moivre gibi ünlü matematikçilerin olasılık teorisini geliştirmeleriyle çıkarımsal istatistiğin temellerini atmışlardır.¹³
- Adolphe Quetelet (1796-1874) yöntemin uygulama alanını, canlı varlıkların antropometrik, psikolojik ve toplumsal açıdan incelenmelerini sağlamıştır. Onun girişimiyle 1885'te Londra'da kurulan Uluslar arası İstatistik Enstitüsü'nün öncüsü olan ilk uluslar arası istatistik kongresi 1853'te Brüksel'de toplanmıştır.
- Karl Pearson (1857-1936) Biyoistatistik'i buldu. Bu yıllarda istatistik ile iktisadın arasındaki ilişkilerin fark edilmesiyle ekonometri doğmuş oldu.¹⁴

1.4. Türkiye İstatistik Kurumu Hakkında Genel Bilgiler

“Cumhuriyet döneminde istatistik işlerine büyük ölçüde önem verilmeye başlanmıştır. Nitekim 1926'da Merkezi İstatistik Dairesi kurulmuştur. Kurumun adı adeta

¹² Özer Serper, Uygulamalı İstatistik 1-2.

¹³ A.g.e.

¹⁴ (Büyük Larousse, Cilt: 10, Syf: 5873)

yazboz tahtası yapılarak 1930'da İstatistik Umum Müdürlüğü, 1945'te İstatistik Genel Müdürlüğü, 1952'de İstatistik Umum Müdürlüğü, 1960'ta İstatistik Genel müdürlüğü şekline çevrilmiş ve 1962'de Devlet İstatistik Enstitüsü adında karar kılınmıştır. 2005 yılında ise Türkiye İstatistik Kurumu (TÜİK) kurulmuş olup, kurum istatistik konseyi ve Türkiye istatistik kurumu başkanlıklarından oluşmuştur. TÜİK başkanlığı 5429 sayılı Türkiye İstatistik kanununun uygulanması sağlamak ve kanunla kendisine verilen görevleri yerine getirmek üzere kurulmuş olup, merkez ve taşra örgütlerine sahiptir. Merkez örgütü ana hizmet danışma ve yardımcı hizmet birimlerinden oluşmaktadır. Taşra örgütü ise bölgesel düzeydeki tüm istatistiksel faaliyetleri yürütmek ve yerel birimlerle koordinasyonu sağlamakla görevli bölge müdürlüklerini kapsamaktadır.”¹⁵

“Enstitü gerekirse Ankara dışında geçici veya sürekli bürolar kurma ve mahalli muhabirler kullanma yetkisine sahiptir. Görevleri: her çeşit çalışmalara ilişkin istatistikleri düzenler, istatistikle ilgili olarak kamu kuruluşları arasında koordinasyon sağlar, çözüm ve incelemeler yapar, düzenlenen istatistikleri yayınlar, istatistik yıllığı çıkarır, sayımlar ve anketler hazırlar ve sonuçlarını yayınlar, kamuoyu çalışmaları yapan bilimsel kurumların çalışmalarında yardım eder, yurt içi ve yurt dışı istatistik eğitim işleri yapar.”¹⁶

Kurumun adı 2005 yılında Türkiye İstatistik Kurumu (TÜİK) olarak değiştirilmiştir ve halen bu isimle anılmaktadır.

1.5. Günümüzde İstatistik

ASA (American Statistical Association) 2013 yılını dünya İstatistik yılı ilan etmiştir. North Carolina Üniversitesi'nden Marie Davidian'ın 19 Aralık 2013'te JASA (Journal of American Statistical Association) dergisinde yayınlanan “The International Year of Statistics: A Celebration and A Call to Action” adlı makalesinde istatistiğin dünyada yükselen bir değer olmasından bahsetmektedir.

Son yıllarda gazetelerde istatistikle ilgili başlıklar atılmıştır, birkaç örnek vermek gerekirse;

-“For today's graduate, just one word: Statistics”, The New York Times, 2009.

¹⁵ A.g.e.

¹⁶ *Meydan Larousse, Cilt: 6, Syf: 512.*

-“The allure of the statistics field grows”, The Boston Globe, 2012

-“What are the odds that stats would be this popular?”, The New York Times, 2012.

-“The numbers guy”, The Wall Street Journal, 2013.

Öne çıkan yazılar bunlar olmakla beraber, istatistik biliminin yükselişi çok sayıda makaleye konu olmuştur.

İstatistiğin doğru kullanımı ve gücü hakkında son yıllarda en büyük katkısı 2012 Amerika devlet başkanlığı seçimlerinden önce yaptığı tutarlı tahminlerle Nate Silver yapmıştır. “The Signal and the Noise: Why So Many Predictions Fail But Some Don’t?” adlı kitabı bir anda en çok satanlar arasına girerek, tüm siyasilerin ve halkın ilgi odağı haline gelmiştir.

Ülkemizde çok uzak değil, bundan 10-15 yıl önce istatistik denilince akla, futboldaki özet sonuçlar veya sokakta yapılan anketler gelmekteydi. Oysa günümüzde tıpkı dünyada olduğu gibi istatistik hayatın her alanına girmiş ve etkin bir biçimde kullanılmaya başlanmıştır.

Uygulamalar

Uygulama Soruları

Bu Bölümde Ne Öğrendik Özeti

Bölüm Soruları

1. İstatistik nedir?
2. Betimsel istatistik nedir?
3. Çıkarımsal istatistik nedir?
4. İstatistiğin kullanım amaçları nelerdir?
5. İstatistiğin doğuşu ve tarihsel gelişimi, dünyadaki son durumu ile ilgili gelişmeler nelerdir?

2. VERİLERİN DERLENMESİ, DÜZENLENMESİ VE SUNUMU

Bu Bölümde Neler Öğreneceğiz?

- Veri kaynakları
- Verilerin elde edilmesi
- Anket
- Verilerin düzenlenmesi
- Grafiklerle sunum

Bölüm Hakkında İlgi Oluşturan Sorular

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
İstatistiksel verinin elde edilmesi, düzenlenmesi ve grafiklerle sunumu	Betimsel istatistik çalışmalarının temelleri öğrenilecek	Metinler, çözümlü problemler, alıştırmalar ve interaktif materyaller ile konuların daha kolay anlaşılması sağlanacaktır.

Anahtar Kavramlar

- Birincil veri
- İkincil veri
- Veri kaynakları
- Anket
- Dizi
- Range (açıklık)
- Nokta diyagramı
- Çizgi diyagramı
- Histogram
- Sıklık poligonu
- Box-Whisker grafiği
- Daire (pasta) grafiği

Giriş

2.1. Verilerin Derlenmesi

Veriler, istatistik çalışmanın temelini oluşturmaktadır. İstatistiksel bir analizin doğruluğu ve güvenilirliği öncelikle verinin kalitesine (çalışmanın amacına uygun ve doğru bir şekilde toplanmış güvenilir verilere) bağlıdır.

2.1.1. Veri Kaynakları

Verilerin elde edilme süreçleri çeşitlilik göstermektedir. Şayet veriyi direkt olarak kaynağından sağlama imkanı var ise böyle elde edilen verilere ***birincil veri*** denir. Çeşitli kurum ve kuruluşların (DPT-Devlet Planlama Teşkilatı, SGK, TOBB, v.b.) yayınladıkları haftalık, aylık, yıllık, v.b. bültenlerden elde edilen veriler de birincil veri kapsamındadır. Mesleki kuruluşlar (Mimarlar Odası, Makine Mühendisleri Odası, v.b.), medya kuruluşlarının ya da araştırma firmalarının yayınladıkları veriler ise daha ziyade ***ikincil veri*** kapsamındadır. Uluslar arası veri kaynaklarının başında ise OECD (Organization for Economic Cooperation and Develoopment), WB (World Bank), BM (UN: United Nations), IMF (International Money Fund) gelmektedir.

2.1.2. Verilerin Elde Edilmesi

İstatistiksel araştırmaların yapılma nedenleri genellikle şu şekilde sıralanabilir:

1. Durum tespiti
2. Geleceğe dönük projeksiyon yapma

Durum tespitinden kasıt, bir kurumun, firmanın, v.b. içinde bulunduğu durumun bilimsel olarak ortaya konmasıdır. Her geçen gün artan rekabet şartları, eskilerin tecrübe, basiret ve sezgiye dayalı yönetim anlayışlarını yetersiz kılmıştır. Artık yapılması gereken böylesi bir ortamda firmanın ne durumda olduğu, hangi şartlarda rekabet etmekte olduğu ve gelecekteki hedeflerine ulaşabilmesi için bugünden neler yapılması gerektiğini bilimsel olarak ortaya koymaktır. Bunun da olmazsa olmaz tek şartı elbette çeşitli ölçümler yapmaktır. Yani veri toplayıp istatistiksel analizler yapmaktır. Firmanın mevcut şartlarının incelenmesi ve ortaya konması durum tespiti, geleceğe dönük yapılması gerekenlerin ortaya konması ise, projeksiyon yapıldığı anlamına gelir.

Verinin toplanması denince ilk akla gelen ve en çok kullanılan yöntem ***anket*** yöntemidir. Anketlerin uygulanma şekilleri farklılık arz eder, örneğin anketörler vasıtasıyla yüzyüze görüşme yaparak olabileceği gibi, e-posta, posta v.b. yöntemlerle de anket formları

göndermek suretiyle veri elde edilebilir. Ancak posta yoluyla yapılan anketlere büyük olasılıkla geri dönüş yapılmamaktadır veya en iyi ihtimalle yarım yamalak cevaplanmış, bir sürü cevapsız soru bırakılmış anket formlarıyla karşılaşmaktadır. Bu nedenle en sağlam veri toplama şekli yüzyüze görüşme yaparak anket yapmaktır denebilir. Gelişmiş ülkelerde posta yoluyla anket cevaplama oranı oldukça yüksektir, her ne konuda olursa olsun bir anket muhakkak bir iyileştirme çabasına hizmet etmektedir, böyle düşünüldüğünde, anketi cevaplayanın ne kadar önemli olduğu anlaşılabilir. Ülkemizde ise bırakın postayla anket cevaplamayı, bazılarının, beş dakikalığına yolunu çeviren anketörleri bile azarladıklarını görmekteyiz. Halbuki artık siyasi partiler bile araştırma şirketlerine anketler yaptırarak halkın nabzını tutmakta, neredeyse seçim sonuçlarıyla aynı denecek yakınlıkta tespitler yapmaktadırlar. Bu sayede politik planlamalar daha gerçekçi yapılabilmekte, kaynaklar daha etkin (gereksiz israf olmadan) kullanılabilmektedir. Örneğin daha küçük bir örnek verelim, bir deterjan fabrikası anket yaptırıyor olsun. Bu ankete ne kadar fazla katılım olur ve dürüstçe cevaplar verilirse, firma yöneticisi de bu doğrultuda üretime geçecek, böylece milli servetimiz heba olmayacaktır. Bir anketten ne olur ki? dememek gerekir. Sonuçta gökten para yağmadığına göre, batan bir tek fabrika bile olsa, bu durum, sadece o fabrika sahibinin iflası gibi görülmemelidir. Orada çalışan binlerce insan işsiz kaldığında, onların aileleri, akrabaları, komşuları, v.b. zaman içinde dalga dalga bu durumdan etkileneceklerdir. O halde işini zamanında doğru düzgün yapmaya çalışan, halkın fikirlerine değer veren ve kaynaklarını etkin bir biçimde kullanmak isteyen herkese bunu bir yardım talebi gibi kabul ederek yardım etmeliyiz. Bakış açımız gerçekten de böyle olmalıdır, bu sadece basit bir örnektir ama binlercesi düşünülebilir.

Anketin öneminden bahsettikten sonra, anketin nasıl uygulanması gerektiğinden bahsedelim. Adım adım özetlersek;

1. Anket konusunun ve amacının belirlenmesi
2. Anketin uygulanacağı birimlerin belirlenmesi (genel olarak herkes mi, yoksa sadece bayanlar mı, ya da sadece 20-25 yaş arası gençler mi, v.b.)
3. Örneklem yöntemi karar verilir. (bu konu örneklemenin konusudur ve çeşitli yöntemler vardır, örneğin basit rastgele örneklem, sistematik örneklem, tabakalı örneklem, küme örnekleme, v.b.)
4. Çerçevenin belirlenmesi (çerçeve anket uygulanacak birimlerin listesi gibi düşünülebilir, böylece çalışmanın sınırları çizilmiş olur, örneğin anket bir ilçede uygulanacaksa her mahalleden kaç kişiye anket uygulanacağını belirlemek çerçeveyi çizmek anlamına gelir)
5. Anket yüzyüze görüşme yöntemi ile uygulanacaksa, bu işi en iyi ve doğru şekilde yapabilecek, konuyla ilgili bilgilendirilmiş anketörlerin seçilmesi. (örneğin inşaat malzemeleri

üreten bir firma, yapı malzemelerinin tercihi konusunda anket yaptırmak istediğinde, malzemelerin özelliklerini ve ne işe yaradığını bilmeyen konudan bihaber anketörlere görev vermek sizce ne kadar doğru olur?)

6. Anketin sahada uygulanması ve verilerin toplanması. (anketin uygulanması başlı başına uzmanlık isteyen bir durumdur. Öyle ki, anketörün tutum ve tavrı, soru sorma stili, ses tonu, giydiği kıyafete kadar özenle tespit edilmelidir. Örneğin bir diş macunu firmasının anketinde, 32 diş pırıl pırıl parlayan çok güzel dişlere sahip bir anketöre bu işi yaptırırsanız, belki haftada bir diş fırçalayan biri, günde üç kez diş fırçalamakta olduğunu beyan edebilir, çünkü anketörün görünümü karşısında ezilmiş ve kendini kötü hissettiği için de maalesef yanlış beyanda bulunmuştur.)

Verilerin anket yoluyla toplanması genel olarak bu biçimdedir. Örnekleme konusu ilerleyen haftalarda ele alınacaktır.

2.2. Verilerin Düzenlenmesi

Aşağıdaki örnek veriyi kullanarak verilerin düzenlenmesi ve özetlenmesi konularını ele alacağız:

Örnek Veri:

19	13	5	4	1	9	9	7	11	11
12	1	1	17	13	18	5	2	12	14
16	18	20	19	19	14	17	15	2	3
6	3	12	16	13	13	11	10	10	8
10	4	2	5	15	10	14	10	1	7

Örnek veri seti, 1 ile 20 arasında değer alan 50 adet gözlemden oluşmaktadır.

Dizi:

Bir veri setinin dizi haline getirilmesi demek, ya küçükten büyüğe, ya da büyükten küçüğe sıralamaktır. Genellikle küçükten büyüğe sıralama yapılır. Diziye bakarak en küçük değer ile en büyük değer arasındaki farkı belirleyerek, böylece gözlemlerin hangi aralıkta dağıldıklarını görebiliriz. Dizideki en küçük değere $X_{\min}$, en büyük değere $X_{\max}$, aralarındaki farka ise “açıklık” yani “range”, “değişim aralığı” da denebilir.

1	1	2	2	3	3	4	4	5	5
---	---	---	---	---	---	---	---	---	---

5	6	6	6	7	7	7	7	8	8
8	9	9	9	9	10	10	10	10	10
10	11	11	11	11	12	12	13	13	13
14	14	15	16	16	17	18	18	19	20

$X_{min}=1$

$X_{max}=20$

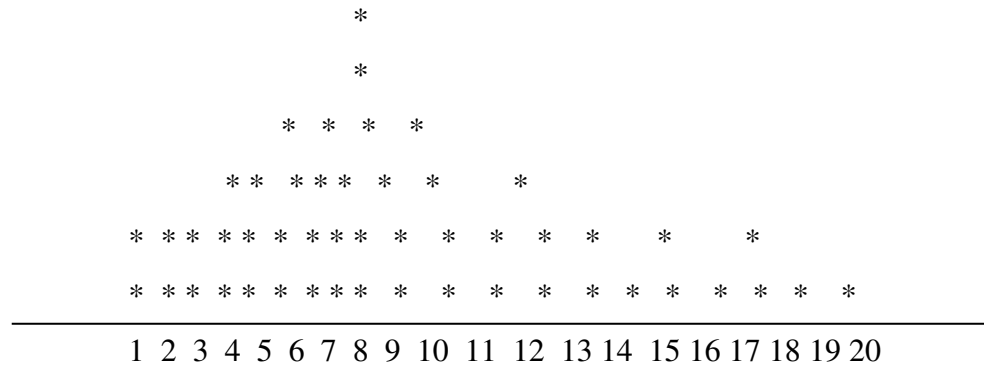
$range=20-1=19$

2.3. Grafikler

2.3.1. Basit Serinin Sunumu İçin Nokta Diyagramı

Verinin değişim aralığını göstermek üzere ölçeklendirilmiş bir doğru üzerinde her bir gözlemin bir nokta ile temsil edildiği bir grafikdir.

Şekil.3.1. Nokta Diyagramına Bir Örnek



Gözlem değerlerinin daha çok hangi değerler etrafında toplandıklarını, ne yönde ve ne büyüklükte dağıldıklarını ve de uç değerleri görmek açısından faydalı bir düzenleme olan nokta diyagramında aynı değere sahip birden fazla gözlem söz konusu olduğunda, noktalar bu değer üzerinde dikey olarak sıralanırlar. Örnek verinin nokta grafiğine bakıldığında verinin daha çok sol tarafta toplandığını, gözlemlerin büyük bölümünün 5-11 aralığına düştüğünü görebiliyoruz.

Örnek: Aşağıdaki verinin nokta diyagramını çizerek yorumlayınız.

103	98	111	106	107	112	107	114
-----	----	-----	-----	-----	-----	-----	-----

106	100	104	96	100	105	118	121
117	101	110	109	107	99	145	101
109	111	105	101	108	102	129	108

Önce veri setini bir dizi haline getirmeliyiz, yani küçükten büyüğe doğru sıralamalıyız:

96	98	99	100	100	101	101	101
102	103	104	105	105	106	106	107
107	107	108	108	109	109	110	111
111	112	114	117	118	121	129	145

* *

** ***** *

* ** ***** * ** * *

95	100	110	120	130	140	145
----	-----	-----	-----	-----	-----	-----

Nokta diyagramına baktığımızda verilerin 107 dolayında yaklaşık olarak simetrik dağıldıklarını, 129 ve 145 uç değerlerini görüyoruz.

2.3.2. Frekans Serisinin Sunumu İçin Çizgi Diyagramı

Veri kümesine bakarak verideki her bir değerin kaç defa tekrar ettiğini görmek mümkündür. Grafıksel bir yola başvurmaksızın, her bir gözlem değerini frekansı (sıklığıyla) birlikte gösteren bir tablo düzenlersek, frekans (sıklık) dağılımını elde etmiş oluruz. Böyle serilerin grafik sunumunda çizgi diyagramı kullanılır.

Örnek veri setimizin frekans dağılımını oluşturalım:

X_i (değerler)	f_i (frekanslar yani sıklıklar)
1	2
2	2
3	2
4	2
5	3
6	3
7	4
8	3
9	4
10	6
11	4
12	2
13	3
14	2
15	1
16	2
17	1
18	2
19	1
20	1

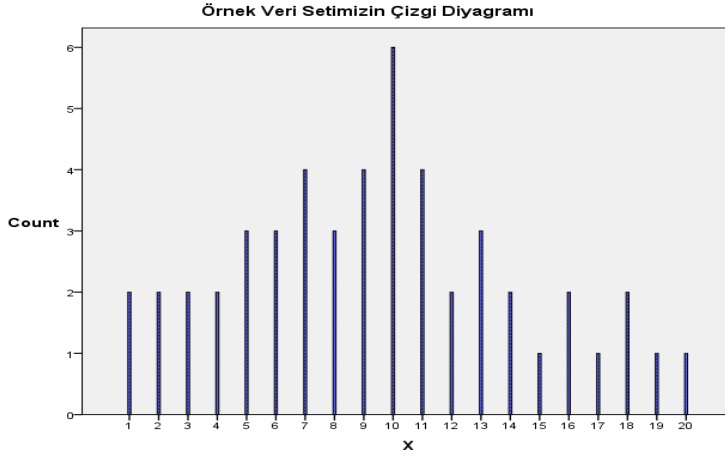
$$\text{Toplam } \sum_{i=1}^{20} f_i = n = 50$$

Sıklıkların toplamı yani $\sum_{i=1}^k f_i$, gözlem sayısı n 'e eşittir. Burada k , sınıf sayısıdır.

İleride sınıflanmış seri kavramını göreceğiz. Buradaki sınıftan kasıt, farklı değerler olarak anlaşılmalıdır. Yani şimdilik gerçek anlamda bir sınıftan bahsetmiyoruz.

Frekans dağılımının grafiksel gösteriminde çoğunlukla çizgi diyagramına başvurulur.

Şekil.3.2. Çizgi Diyagramına Bir Örnek



2.3.3. Sınıflanmış Serinin Sunumu İçin Histogram

Gözlem sayısı çoğaldığında, her bir verinin çizgi diyagramında olduğu gibi tek gösterilmesi güçleşmektedir. Bu nedenle, büyük veri seti söz konusu olduğunda verilerin sınıflandırılmaları yoluna gidilir. Verinin belli sınıflara bölünerek bu aralıklara karşılık gelen frekanslarıyla birlikte gösterildiği tabloya, sınıflandırılmış frekans dağılımı denir. Sınıflandırılmış frekans dağılımını oluşturmak için izlenmesi gereken adımlar şöyledir:

1. Verinin değişim aralığı (açıklığı- range'i) hesaplanır.

2. Değişim aralığı eşit uzunlukta alt aralıklara yani sınıflara bölünür. Sınıfların üst sınırı ile alt sınırı arasındaki fark o sınıfın sınıf genişliğini verir. Sınıf orta noktası, sınıfın alt sınırı ile üst sınırının ortalamasıdır. Sınıf sayısını bulmak için aşağıdaki formüllerden herhangi biri kullanılabilir;

k, sınıf sayısı ve n, veri setindeki gözlem sayısını ifade etmek üzere;

$$k = \sqrt{n}$$

$$k = 1 + 3.3 \log n$$

$$k = 5 \log n$$

Değişim aralığı, sınıf sayısı ve sınıf genişliği arasındaki ilişki şöyledir;

$$\text{sınıfgenişliği} = \frac{\text{range}}{k}$$

3. Her bir sınıf aralığına kaç gözlemin düştüğü sayılır. Bu sayılar o sınıfın frekansını verir.

Örnek: Örnek veri setimizi sınıflandırılmış frekans serisi haline getirelim.

İlk olarak verinin değişim aralığını buluyoruz:

$$X_{\max}=20$$

$$X_{\min}=1$$

$$\text{Range}=19$$

Sonra, k sınıf sayısını buluyoruz:

$k=1+3.3\log(50)=6.6$ bulduk, fakat k tamsayı olmak zorundadır, o halde $k=7$ alalım.

Sınıf genişliği= $19/7=2.714$ bunu da 3'e yuvarlayalım.

İlk sınıfı bulmak için, $X_{\min}$ değerine sınıf genişliğini ekleyelim. Böylece ilk sınıfın üst sınırını da bulmuş olacağız.

<u>Sınıflar</u>	<u>Sıklıklar</u>
1-4 den az	6
4-7 den az	8
7-10 dan az	11
10-13 den az	12
13-16 dan az	6
16-19 dan az	5
19-22 den az	2
Toplam=50	

Histogram:

Histogram, dikey ekseninde sıklıkların, yatay ekseninde sınıf aralıklarının bulunduğu, yan yana dikdörtgen kutucuklardan oluşan bir grafik türüdür. Her bir kutu bir sınıfa tekabül eder ve kutunun alanı, ilgili sınıfın sıklığını verir.

Kutunun alanı=taban*yükseklik

Burada taban, sınıf genişliğidir, yani ilgili sınıfın üst sınırı ile alt sınırı arasındaki farktır. Bu değeri “h” ile ifade edersek, yükseklik şuna eşittir:

Kutunun yüksekliği = f_i/h ’tır.

Bu durumda kutunun alanı= $h*(f_i/h) = f_i$ ’dir yani ilgili sınıfın sıklığı (ilgili sınıftaki birim sayısı) dır.

Histogram çizerken şu adımlar izlenir:

1. Sınıf sayısı (k) belirlenir. Hatırlanacağı üzere k’ye elde etmek için çeşitli formüllerimiz vardı.

$$k = \sqrt{n}$$

$$k = 1 + 3.3 \log n$$

$$k = 5 \log n$$

2. Sınıf aralığının genişliği belirlenmeli. Peki bu işlem nasıl yapılır? Bir çubuk düşünelim. Bu çubuğun boyu “range” olsun yani “Xmax-Xmin” demek istiyoruz. k’ya yani sınıf sayısını da bulduğumuza göre, bu çubuğu kaç parçaya böleceğimiz belli demektir. Çubuğu k tane parçaya böleceğiz. O halde “range/k” bize çubuğu böldüğümüz k tane parçanın bir tanesinin uzunluğunu verecektir. Bu uzunluk sınıf genişliğidir, yani sınıfın üst sınırı ile alt sınırı arasındaki farktır yani histogramda ilgili sınıfa ait kutunun taban uzunluğudur.

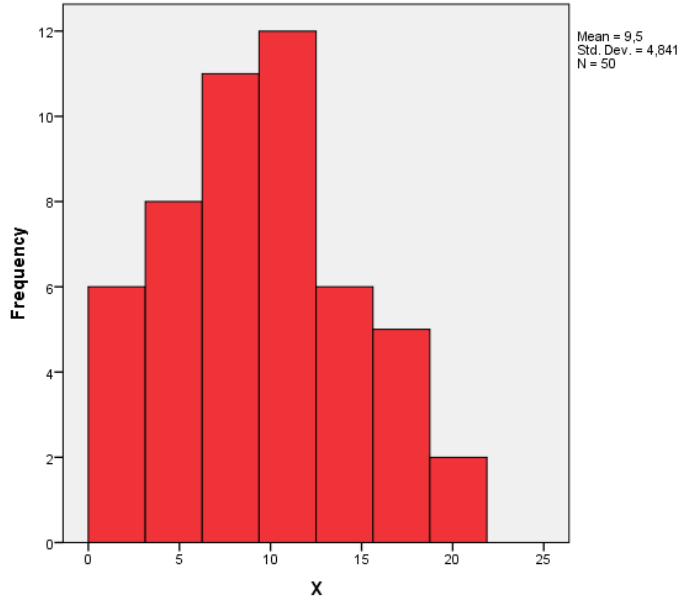
Örnek: Örnek veri setimizin histogramını çizelim.

k=7 alarak, sınıf genişliği 2.714 (bu değeri range/k’dan bulduk) olarak alınsaydı sınıflar aşağıdaki gibi olurdu:

<u>Sınıflar</u>	<u>Sıklıklar</u>
1-3,714	6
3,714-6,428	8
6,428-9,142	11
9,142-11,856	10
11,856-14,57	7
14,57-17,284	4
17,284-20	4

Son sınıfın üst sınırı, 19.998 olarak bulduk ancak bu tamamen virgülden sonra 3 basamakla yetinmemizden kaynaklandı, şayet en azından 5 basamak alsaydık, 20'yi kapsayan bir değer bulacaktık ama böyle olduğunu bildiğimiz için son sınıfın üst sınırını “20” diyerek kapattık. Görüldüğü gibi sınıf genişliğini yuvarlamadan da sınıflandırılmış sıklık dağılımını elde edebiliriz. Gözlemlerimiz tamsayı oldukları için gözlem değerleri sınıf sınırlarıyla çakışmadı. Sıklıklar yazılırken örneğin ilk sınıfın üst sınırı 3.714 olduğundan, bu sınıfa sadece “1,2,3” dâhil edildi, 4 tabii ki alınmadı. 2. Sınıfın sıklıkları yazılırken “4,5 ve 6” altında, tabii ki 7 alınmadı, vb.

Şekil.3.3. Histograma Bir Örnek

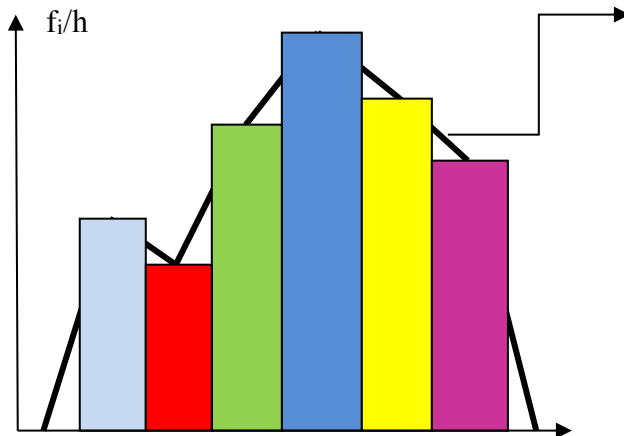


Bu uygulamadan da görüldüğü üzere, bir frekans serisini sınıflandırılmış frekans serisine dönüştürmenin birden fazla yolu vardır (yuvarlama yaparak veya yuvarlama yapmadan yahut k 'nın formül seçiminden kaynaklı farklı sınıf sayıları ortaya çıkabilir, v.b.). *Dikkat edilmesi gereken hususlar ise, veri setindeki tüm gözlemlerin sınıflar tarafından kapsanması, hiçbir verinin açıkta kalmamasıdır. Ayrıca, her bir veri sadece bir sınıfa ait olmalıdır, bir veri birden fazla sınıfta yer alamaz.*

2.3.3.1. Sıklık Poligonu

Sıklık poligonu frekans poligonu olarak da adlandırılır. Histogramda dikdörtgen kutuların üst kısmında sınıf ortalarının birleştirilmesiyle elde edilir.

Şekil.3.4. Sıklık Poligonuna Bir Örnek



Sınıf ortalarını birleştiren bu kırık çizgiye sıklık poligonu denir.

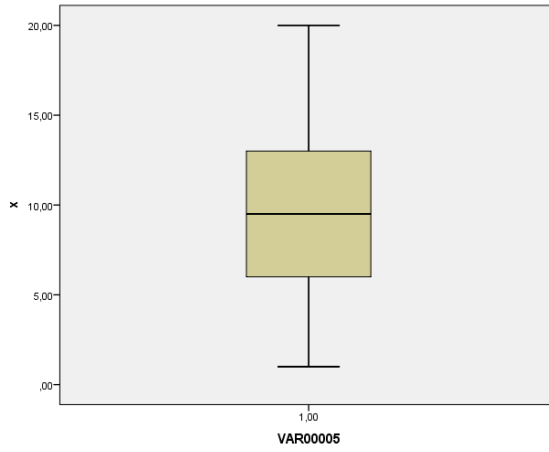
sınıflar

3.3.4. Box-Whisker Grafiđi

Bu grafik türünde beş adet istatistiksel özet bir arada görülür. Bunlar; Xmax, Xmin, Q1 (1. Çeyrek yani küçükten büyüğe sıralanmış bir seriyi %25'ten kesen değeri), Q2 (2. Çeyrek, asıl adı medyandır, küçükten büyüğe sıralanmış bir seriyi %50'den kesen değeri), Q3 (3. Çeyrek yani küçükten büyüğe sıralanmış bir seriyi %75'ten kesen değeri).

Aşağıdaki şekilde örnek veri setimizin SPSS'te çizilmiş Box-Whisker grafiđi yer almaktadır.

Şekil.3.5. Box-Whisker Grafiđine Bir Örnek



3.3.5. Daire (pasta) Grafiđi

Daire şeklindedir. Bu grafiđi çizmek için her bir sınıfın sıklığının toplam sıklık içindeki oranı bilinmelidir. Toplam sıklık 360° olmak üzere, sınıflar sıklıkları oranında birer dilimle temsil edilirler.

Örnek: İstanbul'da 10 milyon yetişkin bulunduğunu varsayalım. Bu kişilerden 3 milyonu ilk okul mezunu, 4 milyonu ortaokul mezunu, 2 milyonu lise mezunu, 1 milyonu üniversite mezunu ise daire grafiđini çizin.

Çözüm:

Oran orantıyla sınıfların kaç derecelik açıyla temsil edilmesi gerektiğini hesaplayacağız.

$$\begin{array}{ccc}
 10 \text{ milyon kiři} & & 360^0 \text{ ise} \\
 & \nearrow & \\
 & \searrow & \\
 3 \text{ milyon kiři} & & \alpha
 \end{array}$$

$$3 \cdot 360 = 10 \cdot \alpha \quad \alpha = 108^0$$

$$\begin{array}{ccc}
 10 \text{ milyon kiři} & & 360^0 \text{ ise} \\
 & \nearrow & \\
 & \searrow & \\
 4 \text{ milyon kiři} & & \alpha
 \end{array}$$

$$4 \cdot 360 = 10 \cdot \alpha \quad \alpha = 144^0$$

$$\begin{array}{ccc}
 10 \text{ milyon kiři} & & 360^0 \text{ ise} \\
 & \nearrow & \\
 & \searrow & \\
 2 \text{ milyon kiři} & & \alpha
 \end{array}$$

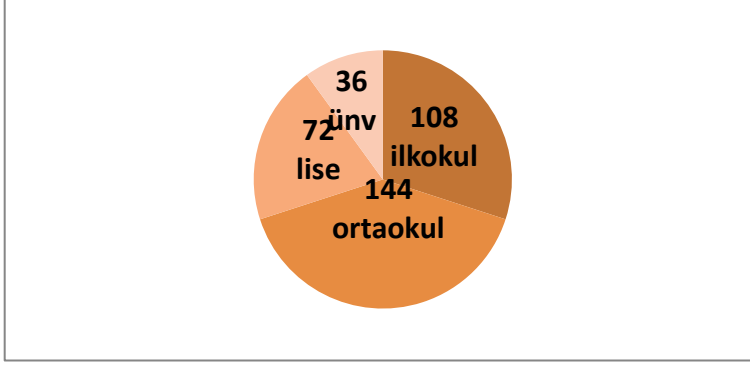
$$2 \cdot 360 = 10 \cdot \alpha \quad \alpha = 72^0$$

$$\begin{array}{ccc}
 10 \text{ milyon kiři} & & 360^0 \text{ ise} \\
 & \nearrow & \\
 & \searrow & \\
 1 \text{ milyon kiři} & & \alpha
 \end{array}$$

$$1 \cdot 360 = 10 \cdot \alpha \quad \alpha = 36^0$$

Örneęe ilişkin pasta diyagramı ařağıdaki gibidir:

Şekil.3.6. Daire Grafiğine Bir Örnek



Yukarıdaki daire grafiğine bakıldığında, 144 derecelik kısım ortaokul mezunlarına, 108 derecelik kısım ilkokul, 72 derecelik kısım lise, 36 derecelik kısım üniversite mezunlarına aittir.

Uygulamalar

Uygulama Soruları

Bu Bölümde Ne Öğrendik Özeti

Bölüm Soruları

1. Birincil veri kaynaklarına örnekler veriniz.
2. İkincil veri kaynaklarına örnekler veriniz.
3. İstatistiksel araştırma yapmanın gerekçeleri neler olabilir?
4. Anketin önemi ve uygulanmasındaki adımları kısaca belirtiniz.
5. Bir frekans (sıklık) dağılımında toplam sıklıklar bize neyi verir?
6. Çizgi diyagramı ne tür serilerde kullanılır?
7. Histogram ne tür serilerde kullanılır?
8. Histogramda sınıf genişliği nasıl elde edilir?
9. Box-Whisker grafiğinin ortadaki çizgi ne anlama gelir?
10. Pasta (daire) grafiğinde dilimlerin büyüklüğü nasıl hesaplanır?

3. MERKEZİ EĞİLİM ÖLÇÜLERİ

Bu Bölümde Neler Öğreneceğiz?

- Merkezi eğilim ölçüleri
- Duyarlı ortalamalar
- Duyarsız ortalamalara giriş

Bölüm Hakkında İlgi Oluşturan Sorular

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Merkezi eğilim ölçüleri	Duyarlı ortalamaların öğrenilmesi, duyarsız ortalamalara giriş	Metinler, çözümlü problemler, alıştırmalar ve interaktif materyaller ile konuların daha kolay anlaşılması sağlanacaktır.

Anahtar Kavramlar

- Merkezi eğilim ölçüleri
- Analitik (duyarlı) ortalamalar
- Analitik olmayan (duyarsız) ortalamalara giriş
- Aritmetik ortalama
- Aritmetik ortalamanın özellikleri
- Tartılı aritmetik ortalama
- Kareli ortalama
- Geometrik ortalama
- Harmonik ortalama

Giriş

3. 1. Merkezi eğilim ölçüleri

Merkezi eğilim ölçüleri bir seri hakkında bilgi edinmemizi ve çeşitli kıyaslamalar yapmamıza olanak sağlayan ölçütlerdir. Verilerin daha çok hangi değerlere eğilimli olduklarını bulmak için merkezi eğilim ölçülerini kullanırız. Bu ölçüler analitik olan (duyarlı) ve analitik olmayan (duyarsız) ortalamalardır. Analitik ortalamalar; aritmetik ortalama, tartılı ortalama, kareli ortalama, harmonik ortalama ve geometrik ortalama. Analitik olmayan ortalamalar ise mod, medyan ve kartillerdir.

3.1.1. Analitik (duyarlı) ortalamalar

Analitik ortalamalara “duyarlı” denmesinin nedeni; veri kümesindeki tüm gözlem birimlerini hesaba katmasıdır. Bu hesaplama yöntemi, özellikle uç değerler (uç değer bir diğer ifade ile sapan değer, serideki değerlerden çok farklı olan değerlerdir) söz konusu olduğunda ortalamayı uç değerlere doğru çekerek doğru tahmin yapmayı zorlaştırırlar. Bu tip durumlarda ya uç değerler hesaplamadan dışlanır, ya da analitik olmayan ortalama türlerinden faydalanılır. Bu duruma en güzel örnek, bir sınıfın ortalama notunu hesaplarken en yüksek ve en düşük değer dışlanarak hesaba alınmamasıdır.

Örneğin örneklemimiz aşağıdaki gibi olsun ve yukarıdaki duyarlı olma durumunu aritmetik ortalamayı kullanarak basitçe anlatmaya çalışalım:

20 23 31 42 55

Bu örneklemin aritmetik ortalaması tüm değerlerin toplamının gözlem sayısı n 'e bölünmesiyle 34,2 olarak hesaplanır. Ortalamaların taşınması gereken en önemli özellik, gözlemleri iyi bir şekilde temsil etmesidir (ortalama aslında serinin tek bir rakamla özetlenmesidir) ve bulduğumuz bu değer örneklemdaki değerleri iyi kötü temsil etmektedir.

Şimdi örnekleme bir uç değer ekleyelim, yeni örneklemimiz şöyle olsun:

20 23 31 42 55 1000

Bu örneklemin aritmetik ortalaması 195,166'dır. Görüldüğü gibi uç değer ortalamayı kendine doğru çekmiş, bu durum ne küçük gözlemleri ne de uç değeri temsil edebilecek bir ortalamanın hesaplanmasına neden olmuştur. İşte böyle durumlarda, tüm gözlemleri hesaba katmayan analitik olmayan yani bir diğer adıyla duyarsız ortalamalar tercih edilmelidir.

3.1.2. Analitik olmayan (duyarsız) ortalama

Analitik olmayan ortalamalara “duyarsız” denmesinin nedeni, özellikle uç değerlerin olduğu durumlarda bu gibi değerleri göz ardı etmesi, tüm gözlemleri hesaba almaması ve sıralamaya dayanmasından kaynaklanmaktadır. Bu sıralamanın ne anlama geldiğini ilerideki sayfalarda göreceğiz.

3.1.1.1. Aritmetik Ortalama ($\bar{X}$, μ)

Örneklemede $\bar{X}$, kitlede μ ile gösterilir. Veri kümesindeki tüm gözlemlerin toplamının toplam gözlem sayısına bölümüdür.

Basit seride:

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

Sıklık serisinde:

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

Sınıflanmış seride:

$$\bar{X} = \frac{\sum_{i=1}^k f_i m_i}{\sum_{i=1}^k f_i} = \frac{\sum_{i=1}^k f_i m_i}{n}$$

k, sınıf sayısıdır. m_i ise i. sınıfın orta noktasıdır (mid point)

Kitlede Gösterimi:

$$\mu = \frac{\sum_{i=1}^N X_i}{N}$$

$$\mu = \frac{\sum_{i=1}^k f_i X_i}{N}$$

$$\mu = \frac{\sum_{i=1}^k f_i m_i}{N}$$

Örnek:

Aşağıda bir öğrencinin 5 dersten aldığı notlar verilmiştir. $\bar{X} = ?$

4,5,7,8,10

$\bar{X} = 6.8$

Örnek:

Bir kavşakta bir ayda meydana gelen trafik kazalarının dağılımı aşağıdaki gibidir. Günlük ortalama kaza sayısını hesaplayınız.

X_i	f_i
0	14
1	7
2	4
3	2
4	3
$\sum f_i = 30$	

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

Günlük ortalama kaza sayısı yaklaşık olarak 1'dir.

Örnek:

Aşağıdaki serinin aritmetik ortalamasını hesaplayınız.

Sınıflar	f_i	m_i	$f_i m_i$
1-3	4	2	8
3-5	8	4	32
5-7	3	6	18
$\sum f_i = 15$		$\sum f_i m_i = 58$	

Sınıf orta noktaları sınıfın alt sınır ile üst sınırının ortalamasıdır, yani ikiye bölümdür. $m_1 = (1+3)/2 = 2$ v.b.

Burada biraz duralım, sınıfları tek tek ele alalım. İlk sınıfımız 1-3 aralığıdır ve bu aralıkta 4 gözlem vardır. Fakat biz bu 4 değerini tam olarak ne olduklarını bilmiyoruz, sadece [1,3) aralığında olduğunu biliyoruz değil mi? O halde standart bir ortalama bulabilmek için

hepimizin ortak bir yaklaşım sergilemesi gerekir. Yani şunu anlatmak istiyoruz, diyelim ki Ali'ye göre $X_1= 1$, $X_2= 1.2$, $X_3=2.8$, $X_4=2.9$ olsun. İkinci sınıftaki değerleri de şöyle olsun; $X_5= 3.1$, $X_6= 4$, $X_7= 4.1$, $X_8= 4.3$, $X_9= 4.4$, $X_{10}= 4.6$, $X_{11}= 4.8$, $X_{12}= 4.9$ olsun. Son sınıftakiler de şöyle olsun; $X_{13}= 5.2$, $X_{14}= 5.9$, $X_{15}= 6.8$

Şimdi Ali'nin serisinin ortalaması tüm bu değerlerin toplamının toplam gözlem sayısına yani 15'e bölümüdür. Ali'nin serisini açıkça bir daha yazalım:

1	1.2	2.8	2.9	3.1	4	4.1	4.3	4.4
4.6	4.8	4.9	5.2	5.9	6.8			

Ortalamayı bulalım:

$$\bar{X} = \frac{\sum_{i=1}^{15} X_i}{15} = 60/15=4$$

Umut'un serisi de şöyle olsun:

1.1	1.3	1.7	2.4	3	4	4.2	4.5	4.6
4.7	4.8	4.9	5	7	7.9			

Bu ortalamayı bulalım:

$$\bar{X} = \frac{\sum_{i=1}^{15} X_i}{15} = 61,1/15=4.0733333 \sim 4.073 \text{ bulduk.}$$

Demek ki herkes bu aralıkta sonsuz tane değer ataması yapabilir ve bunun sonucunda da herkesin bulacağı ortalama değeri farklı olacaktır. İşte bu durumun önüne geçebilmek için sınıf orta noktalarını yani m_i 'leri X_i gibi düşünüyoruz, yani sanki o değerlerin hepsi sınıf orta değerini alıyormuş gibi varsayıyoruz, aksi halde aynı ortalama değerini bulamayız ve herkes kendine göre bir değer bulur.

Sınıflanmış serinin basit seri hali şöyledir:

2	2	2	2	4	4	4	4
4	4	4	4	6	6	6	

Ortalamayı bulalım;

$$\bar{X} = \frac{\sum_{i=1}^k f_i m_i}{\sum_{i=1}^k f_i} = \frac{\sum_{i=1}^k f_i m_i}{n} = 3,8666666 \sim 3,867$$

Aritmetik Ortalamanın Özellikleri

1. En çok bilinen ve en sık kullanılan ortalama dır. Her türlü veride hesaplanabilir ve tek bir değ er alır. Sıfır içeren serilerde harmonik ortalama hesaplanamaz, geometrik ortalama hesaplanabilir ama anlamsız olur. Seride negatif değ er var ise, geometrik ortalama hesaplanamaz. Bazı serilerde birden fazla sayıda mod bulunabilir. **Ancak aritmetik ortalama her türlü veri seti için hesaplanabilir ve tek bir tanedir.**

2. $\bar{X}$ duyarlı bir ortalama dır. Çünkü bütün verileri hesaba alır. Extrem (uç, outlier) değ erler olduğ unda bu değ erlerden etkilenmesi, aritmetik ortalamanın zayıf yönüdür. Böyle durumlarda $\bar{X}$ yerine, duyarsız ortalama da denilen, sıralamaya dayanan ve uç değ erlerden etkilenmeyen medyan tercih edilebilir.

3. Bir seride $\bar{X}$ 'dan sapmaların toplamı sıfırdır.

$$\sum_{i=1}^N (X_i - \mu) = 0$$

$$\sum_{i=1}^n (X_i - \bar{X}) = 0$$

4. Açık uçlu dağılımlarda hesaplanamaz. Açık uçlu dağılıma bir örnek verelim:

Sınıflar	f_i	m_i	$f_i m_i$
1-3	4	2	8
3-5	8	4	32
5-...	3	?	?

$$\sum f_i = 15$$

Son sınıfın üst sınırı açık bırakılmış, yani oradaki değerin ne olduğu bilinmeden sınıf ortası bulunamayacağı için aritmetik ortalama da hesaplanamaz. Sınıf genişlikleri eşit olmak zorunda değildir, yani biz kendiliğimizden oranın “7” olduğunu iddia edemeyiz.

3.1.1.2. Tartılı Ortalama (Ağırlıklı Ortalama)

Bazen veride yer alan bazı gözlemlere diğerlerinden daha fazla önem verilir. Böyle durumlarda ortalama hesaplanırken bu özel değerlere verilen önemi yansıtmak için belirli tartılar tahsis edilir.

Örnek: İstatistik dersine ilişkin vize notunun ağırlığı finalin üçte biri olsun. Öğrenci vizeden 80, finalden 60 almış ise, bu öğrencinin başarı notu tartılı aritmetik ortalama ile hesaplanır. Tartılı aritmetik ortalamanın formülü şöyledir:

$$\bar{X} = \frac{\sum t_i X_i}{\sum t_i}$$

Örneğimize dönersek;

$t_1=1$ vizenin tartısı

$t_2=3$ finalin tartısı (bir anlamda önemi)

$$\bar{X} = \frac{(1*80) + (3*60)}{1+3} = 65$$

Örnek: Aşağıdaki tabloda bir öğrencinin ders notları ve derslerin ağırlıkları yer almaktadır (yani her ders aynı öneme sahip değildir) öğrencinin not ortalamasını hesaplayınız.

Ders	Ağırlık	Puan	Ağırlıklı Puan
Fizik	2	70	140
Kimya	2	80	160
Matematik	2	85	170
Türkçe	1	90	90
Tarih	1	65	65

$$\bar{X} = \frac{\sum t_i X_i}{\sum t_i}$$

$$= (140 + 160 + 170 + 90 + 65) / (2 + 2 + 2 + 1 + 1) = 625/8 = 78.125 \text{ 'tir.}$$

Normal ortalaması (yani her dersin eşit öneme sahip olduğu durumda);

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = (70+80+85+90+65)/ 5 = 390/5=78 \text{ 'dir.}$$

3.1.1.3. Kareli Ortalama

Kareli ortalama serideki değerlerin karelerinin toplamının gözlem sayısına bölünerek karekökünün alınmasıyla hesaplanır.

Basit Seride:

$$KO = \sqrt{\frac{\sum X^2}{n}}$$

Sıklık Serisinde:

$$KO = \sqrt{\frac{\sum f_i X^2}{n}}$$

Sınıflanmış Seride:

$$KO = \sqrt{\frac{\sum f_i m_i^2}{n}}$$

Örnek: Aşağıdaki serinin kareli ortalamasını hesaplayınız.

Sınıflar	f_i	m_i^2	$f_i m_i^2$
1-3	4	4	16
3-5	8	16	128
5-7	3	36	108

$$KO = \sqrt{\frac{\sum f_i m_i^2}{n}} = \sqrt{\frac{16+128+108}{15}} = 4.09878 \sim 4.099$$

3.1.1.4. Geometrik Ortalama

Geometrik ortalama tüm verileri hesaba katan duyarlı bir ortalama değildir. Genellikle oransal olarak artan serilerde bu artışı düzlemek (smooth etmek) amacıyla tercih edilir. Tüm gözlemlerin çarpılarak, gözlem sayısı kadar dereceden kök alınmasıyla hesaplanır.

Geometrik Ortalama $\leq$ Aritmetik Ortalama

$X_1, X_2, \dots, X_n$ veri kümesinin geometrik ortalaması;

$$GO = \sqrt[n]{X_1 X_2 \dots X_n} = (X_1 X_2 \dots X_n)^{1/n} = \left(\prod_{i=1}^n X_i \right)^{1/n}$$

$$\log GO = \frac{1}{n} (\log X_1 + \log X_2 + \dots + \log X_n)$$

Basit seride:

$$\log GO = \frac{\sum_{i=1}^n \log X_i}{n}$$

Sıklık serisinde:

$$\log GO = \frac{\sum_{i=1}^k f_i \log X_i}{n}$$

Sınıflanmış seride:

$$\log GO = \frac{\sum_{i=1}^k f_i \log m_i}{n}$$

Örnek: Aşağıdaki serinin geometrik ortalamasını bulunuz.

X_i

4

5

7

8

16

$$GO = \sqrt[5]{4 * 5 * 7 * 8 * 16} = 7.09$$

$$\log GO = \frac{1}{5} (\log 4 + \log 5 + \log 7 + \log 8 + \log 16)$$

logX_i

0,602

0,699

0,845

0,903

1,204

$$\sum_{i=1}^5 \log X_i = 4,253$$

$$\log GO = \frac{1}{5} * 4,253 = 0,8506$$

$$GO = 10^{0,8506} = 7,09$$

Örnek: Aşağıdaki sıklık serisinin geometrik ortalamasını hesaplayınız.

X_i	f_i	$\log X_i$	$f_i \log X_i$
2	3	0,301	0,903
3	2	0,477	0,954
4	1	0,602	0,602
5	4	0,698	2,795

$$\sum f_i = 10$$

$$\sum f_i \log X_i = 5,255$$

$$\log GO = \frac{\sum_{i=1}^k f_i \log X_i}{n} = 5,255/10 = 0,5255$$

$$GO = 10^{0,5255} = 3,35$$

Örnek: Aşağıdaki sınıflanmış serinin geometrik ortalamasını hesaplayınız.

Sınıflar	f_i	m_i	$\log m_i$	$f_i \log m_i$
1-3	3	2	0,301	0,903
3-5	3	4	0,602	1,806

$$\begin{array}{ccccccc} 5-7 & & 4 & & 6 & & 0,778 & & 3,112 \\ \hline \end{array}$$

$$\sum f_i = 10$$

$$\sum_{i=1}^k f_i \log m_i = 5,821$$

$$\log GO = \frac{\sum_{i=1}^k f_i \log m_i}{n} = 0,5821$$

$$GO = 10^{0,5821} = 3,82$$

3.1.1.5. Harmonik Ortalama

Oransal olarak belirtilebilen değişkenlerin ortalamalarının hesaplanmasında harmonik ortalama kullanılır. Sıfır değerli ya da negatif işaretli değişkenler olduğunda hesaplanamaz.

Basit seride:

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

Sıklık serisinde:

$$HO = \frac{\sum_{i=1}^k f_i}{\sum_{i=1}^k \frac{f_i}{X_i}}$$

Sınıflanmış seride:

$$HO = \frac{\sum_{i=1}^k f_i}{\sum_{i=1}^k \frac{f_i}{m_i}}$$

Örnek:

X_i	$\frac{1}{X_i}$
4	0,25
5	0,2
7	0,14
8	0,12
16	0,06

$$\sum_{i=1}^k \frac{1}{X_i} = 0,77$$

$$HO = \frac{5}{0,77} = 6,49$$

Örnek: Harmonik ortalamasını hesaplayınız.

X_i	f_i	$\frac{f_i}{X_i}$
2	3	1,5
3	2	0,67
4	1	0,25
5	4	0,80

$$\sum f_i = 10 \quad \sum_{i=1}^k \frac{f_i}{X_i} = 3,22$$

$$HO = 10/3,22 = 3,11$$

Örnek: Harmonik ortalamasını hesaplayınız.

Sınıflar	f_i	m_i	$\frac{f_i}{m_i}$
1-3	3	2	1,5
3-5	3	4	0,75
5-7	4	6	0,67
$\sum f_i = 10$			$\sum_{i=1}^k \frac{f_i}{m_i} = 2,92$

$$HO = 10 / 2,92 = 3,42$$

Uygulamalar

Uygulama Soruları

Bu Bölümde Ne Öğrendik Özeti

Bölüm Soruları

1. Duyarlı ve duyarlı olmayan ortalama ne demektir? Örnek veriniz.
2. Aşağıda bir sınıftaki öğrencilerin aldıkları notların dağılımı mevcuttur. Sınıfın ortalama başarısı nedir?

Sınıflar	Öğrenci sayısı
0-20	10
20-40	15
40-60	20
60-80	5
80-100	2

3. Bir öğrencinin vizeden aldığı not 50 ve finalden aldığı not 52 'dir. Bu sınıfta geçme notu 60 ise öğrenci ilgili dersten geçebilir mi? (Vize ağırlığı %30)

4. $\sum_{i=1}^n (X_i - \bar{X}) = 0$ ifadesini 4 gözlemden oluşan bir seriden yola çıkarak doğrulayınız.

5. Toplam frekans değeri 15 olan bir tane sınıflandırılmış, bir de sıklık serisi oluşturarak bu serilerde aritmetik ortalama, kareli ortalama, geometrik ortalama ve harmonik ortalama değerlerini hesaplayarak bu 4 değer arasındaki sıralamayı belirtin.

4. MERKEZİ DEĞİŞİM ÖLÇÜLERİ

Bu Bölümde Neler Öğreneceğiz?

- Merkezi değişim ölçüleri
- Chebyshev Teoremi

Bölüm Hakkında İlgi Oluşturan Sorular

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Merkezi değişim ölçüleri, Chebyshev teoremi	Serilerin yayılımlarını incelemek amacıyla merkezi değişim ölçüleri öğrenilecektir.	Metinler, çözümlü problemler, alıştırma ve interaktif materyaller ile konuların daha kolay anlaşılması sağlanacaktır.

Anahtar Kavramlar

- Sapma
- Varyans
- Standart sapma
- Sheppard düzeltmesi
- Değişim katsayısı
- Chebyshev teoremi

Giriş

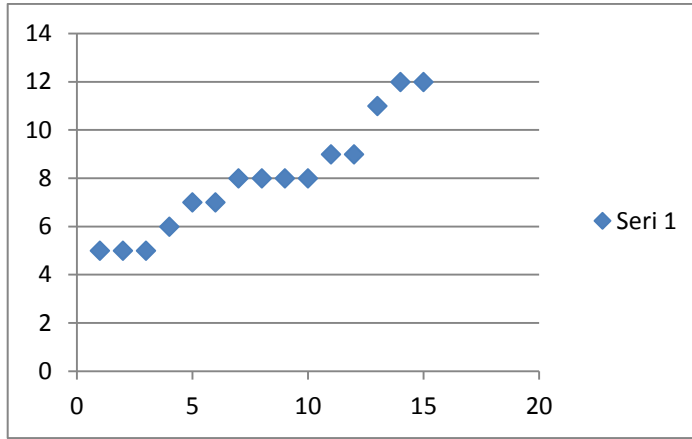
4. 1. Merkezi Değişim Ölçüleri

Bir serideki gözlemlerin (birimlerin) birbirinden ya da herhangi bir ortalama değerden uzaklıklarının çeşitli ölçümlerine merkezi değişim ölçüleri denir. Bir seriyi özetlemede ortalamalar tek başına yeterli bir veri değildirler. Ortalamaların yanısıra değişim ölçülerine de ihtiyaç vardır. Örneğin aşağıda ortalamaları birbiriyle aynı ama dağılımları farklı üç seri göreceksiniz:

Şekil.4.1. Seri 1

Seri1: 5 5 5 6 7 7 8 8 8 8 9 9 11 12 12

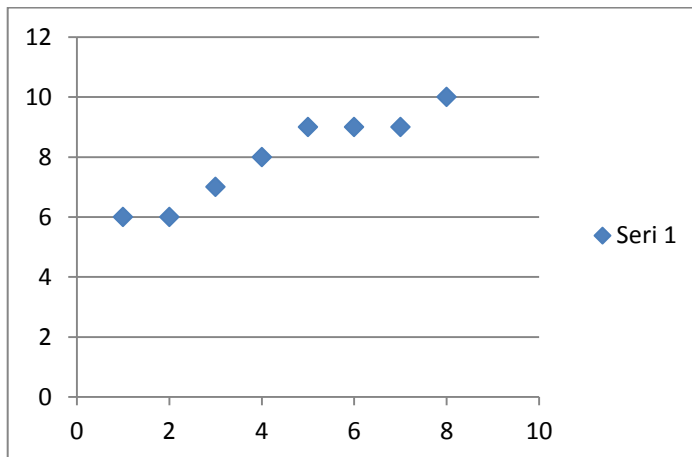
$$\bar{X}_1 = 8$$



Şekil.7. 2. Seri 2

Seri2: 6 6 7 8 9 9 9 10

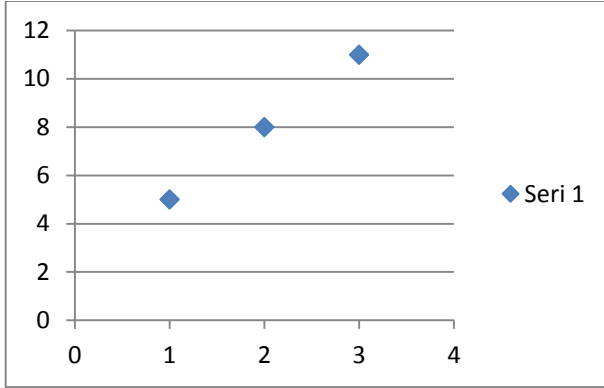
$$\bar{X}_2 = 8$$



Şekil.7. 3. Seri 3

Seri3: 5 8 11

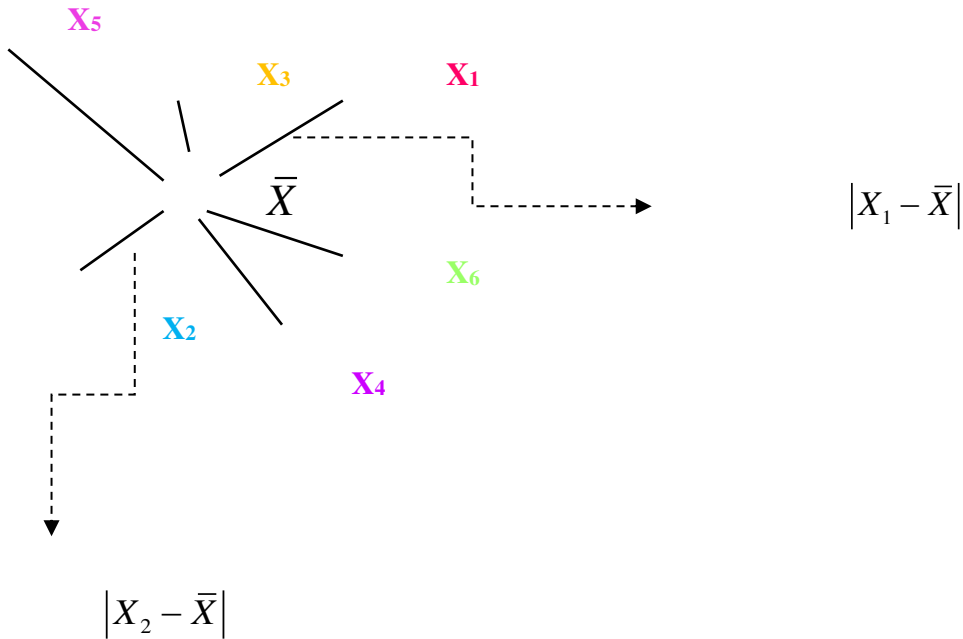
$$\bar{X}_3 = 8$$



Her üçünün de ortalaması birbiriyle aynı olmasına rağmen, veri sayıları, değişimleri, dağılımları birbirinden farklıdır. Öyleyse artık değişim ölçülerinin neler olduklarına ve nasıl hesaplandıklarına ilişkin açıklamalarımıza geçebiliriz.

Verilerin dağılımını ölçmede, gözlem değerlerinin ortalamadan sapmalarını kullanabiliriz. Bu sapmalar yani $|X_i - \bar{X}|$ ne kadar büyükse, X_i gözlemi ortalamadan o denli uzakta demektir.

Şekil.7. 4. Birimlerin Ortalamadan Uzaklıkları (sapmaları)



7.1.1. Varyans

Şekil. 7.4'e bakalım. Sizce bu sapmalar ne anlama geliyor ve nasıl ölçülecek?

Mesela, $\bar{X}$ 'dan sapmalar ile ölçülebilir mi? Hayır ölçülemez, çünkü ortalamadan sapmaların toplamı daima sıfırdır.

$$\sum (X_i - \bar{X}) = 0$$

Öyleyse biz de sapmaların karelerini alıp işe başlarız yani aritmetik ortalama değil de, kareli ortalamayı kullanırız. Sapmaların kareli ortalaması bulunduğunda, istatistik teorisinde ve uygulamalarda en yaygın kullanılan dağılma ölçüsü olan **standart sapmayı** yani σ 'yı elde ederiz. Standart sapmanın karesine yani σ^2 'ye ise **varyans** denir. Varyans değişimin ölçüsüdür.

Varyans kısa yoldan şöyle hesaplanabilir:

$$\sigma^2 = KO^2 - \bar{X}^2$$

4.1.2. Standart Sapma

Standart sapma, varyansın kareköküdür.

Standart sapmanın normal yoldan bulunuşu aşağıdaki gibidir:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}}$$

kitlenin standart sapması

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

örneklemin standart sapması

σ^2 kitle varyansıdır, s^2 ise örneklem varyansıdır.

7.1.2.1.Sıklık serisinde standart sapma hesabı

Kitlede:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N f_i (X_i - \mu)^2}{\sum f_i}}$$

Örneklemede:

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

7.1.2.2. Sınıflanmış seride standart sapma hesabı

Kitlede:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N f_i (m_i - \mu)^2}{\sum f_i}}$$

Örneklemede:

$$s = \sqrt{\frac{\sum_{i=1}^n f_i (m_i - \bar{X})^2}{n - 1}}$$

Sheppard düzeltmesi:

Sınıfları temsilen sınıf ortalarının gelişigüzel tam orta nokta olarak kabul edilmesi (aslında belki verilerin sınıftaki dağılımı tam ortalarda olmayabileceği halde), standart sapma hesabında hataya neden olur. Bunu Sheppard düzeltmesiyle gidermek mümkündür.

$$\sigma^* = \sqrt{\sigma^2 - \frac{c^2}{12}}$$

σ^* düzeltilmiş standart sapmadır. c , ortak sınıf genişliğidir. Standart sapma bulunduğundan sonra bu düzeltme yapılırsa, sınıf ortalarını tam orta nokta kabul etmekten kaynaklanan hatanın giderilmiş olduğu düşünülmektedir.

Varyansın hesaplanmasıyla ilgili birkaç örnek yapalım:

Örnek: Aşağıdaki basit seride varyans ve standart sapmayı hesaplayınız.

4 5 8 15 18 22

Bu veri kümesini **kitle** gibi düşünürsek kullanmamız gereken formül şudur:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}}$$

Öncelikle kitle ortalamasını yani μ 'yü hesaplayalım. Basitçe tüm gözlemleri toplayarak gözlem sayısına bölüyoruz.

$$\mu = \frac{4+5+8+15+18+22}{6} = 12$$

sonra formüle bakıyoruz, dediğimiz gibi formüller konuşurlar ve bize sırasıyla ne yapmamız gerektiğini söylerler. Şimdi kitle ortalaması μ 'yü 12 olarak hesapladık. Her bir gözlem değerini tek tek bu ortalama değerden saptıracağız. Bunun için bir kolon oluşturalım:

$X_i - \mu$

-8

-7

-4

3

6

10

Sonra bu değerlerin tek tek karelerini alıp toplayacağız ve bulduğumuz toplamı n'e yani 6'ya bölerek varyansı bulmuş olacağız.

$(X_i - \mu)^2$

64

49

16

9

36

100

$$\sigma^2 = \frac{64+49+16+9+36+100}{6} = \frac{274}{6} = 45.667$$

Standart sapma ise varyansın karekökü olduğuna göre;

$\sigma = 6.758$ 'dir.

Bu veri kümesini **örneklem** gibi düşünürsek kullanmamız gereken formül şudur:

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

$$s^2 = 274/5 = 54.8$$

$s = 7.403$ olarak elde edilir.

Uyarı!!! örneklem varyansında paydada (n-1) kullandığımız için, kitle varyansına göre bir miktar daha büyük bir değer buluyoruz. Böyle yapmamızın nedeni, örneklem üzerinden çalışıldığında bir miktar hataya göz yummak demektir. Halbuki kitle üzerinden yapılan hesaplarda yani parametre hesaplarken tamsayım söz konusu olduğu için böyle düzeltmelere ve önlemlere gerek yoktur, zira hesaplanan değerler kesindir.

Örnek: Aşağıdaki sıklık serisinde varyansı hesaplayınız.

<u>X_i</u>	<u>f_i</u>
4	2
6	1
8	4
10	3

Bu seri aslında şu basit serinin toplulaştırılmış halidir:

4 4 6 8 8 8 8 10 10 10

Şimdi varyansı bulalım. İlk yapmamız gereken şey ortalamayı bulamaktır. İsterseniz yine bu veri kümesini bir örneklem gibi kabul edelim;

$f_i X_i$

8

6

32

30

$$\bar{X} = \frac{\sum_{i=1}^4 f_i X_i}{10} = \frac{76}{10} = 7.6$$

Formülümüz şöyle;
$$s = \sqrt{\frac{\sum_{i=1}^n f_i (X_i - \bar{X})^2}{n - 1}}$$

Şimdi tüm X değerlerini ortalamadan saptıracağız:

$X_i - \bar{X}$

$4 - 7.6 = -3.6$

$6 - 7.6 = -1.6$

$8 - 7.6 = 0.4$

$10 - 7.6 = 2.4$

Şimdi bu farkların karelerini alalım:

$(X_i - \bar{X})^2$

12.96

2.56

0.16

5.76

Şimdi bu karesel ifadeleri sıklıklarla ağırlıklandıracağız yani çarpacağız;

$f_i (X_i - \bar{X})^2$

$2 * (12.96) = 25.92$

$1 * (2.56) = 2.56$

$4 * (0.16) = 0.64$

$3 * (5.76) = 17.28$

Artık hepsini toplayıp (n-1)'e bölebiliriz ve böylece örneklem varyansını bulmuş oluruz:

$$s^2 = \frac{25.92 + 2.56 + 0.64 + 17.28}{9} = \frac{46.4}{9} = 5.156$$

s= 2.27'dir.

Varyansın Formülünü İnceleyelim

Varyans kelimesi “variation” yani değişim anlamına gelmektedir. İstatistiksel olarak yorumlarsak, bir veri kümesindeki gözlemlerin ortalamadan uzaklıklarının bir ölçüsüdür diyebiliriz. Varyansı büyük bir seride birimlerin ortalamadan farklı (ortalamaya uzak) değerlere sahip olduğunu görürüz. Varyansın küçük olması ise gözlemlerin ortalama dolayında dağılım gösterdiği durumlarda meydana gelir. Bu anlatılanlardan şunu anlamalıyız, eğer bir serideki gözlem birimleri ortalamadan uzakta çok farklı değerler alıyorsa bu serinin varyansı (değişimi, yayılımı), ortalama dolayında bir dağılıma sahip olan bir veri kümesinin varyansına göre daha büyüktür. Formülün paydasında “n” vardır, yani birim sayısı arttıkça varyans küçülmektedir, varyansın yani değişimin küçülmesi ise ileriki konularda anlatacağımız parametre tahminlerinin daha isabetli daha az hata ile yapılmasını sağlayacak bir unsurdur.

4.1.3. Değişim Katsayısı

Bir seriyi özetlemede sadece ortalamanın yeterli olamayacağından bahsetmiştik. Değişim katsayısı, ortalama ile varyansı birlikte kullanarak serilerin değişimlerini ölçmeyi ve farklı türden olsalar bile kıyaslama imkânı sağlayan bir ölçüttür. Değişim katsayısının formülü aşağıdaki gibidir:

$$DK = \frac{\sigma}{\mu} * 100 \quad \text{veya,} \quad DK = \frac{\sigma}{\bar{X}} * 100$$

Değişim katsayısı ölçü birimlerinden bağımsızdır. Oransal yapısından dolayı, pay ve paydasında aynı cins ve büyüklükten değerler birbirini götürceği için basit seri, sıklık serisi ve sınıflanmış seriler arasındaki cins ve büyüklük farklılığını ortadan kalkar.

DK'sı küçük olan serilerin, diğerlerine göre daha az değişken olduğu (daha homojen yani ortalamaya yakın dağılıma sahip) birimlerden oluştuğu söylenir. Çünkü standart sapma küçüldükçe DK da küçülür. Standart sapmanın küçük olması ise ancak ortalama etrafındaki saçılımın $\bar{X}$ 'ya doğru çekilmesiyle mümkündür birimler $\bar{X}$ 'dan uzaklaştıkça standart sapma da büyümektedir.

Örnek: İki hisse senedi olsun. Bu hisselerin 1 aylık ortalama getirileri ve standart sapmaları aşağıdaki gibi ise, siz bir portföy yöneticisi olarak **riski seven** bir yatırımcıya hangi hisse senedini tavsiye edersiniz?

$$\mu_1=70 \quad \sigma_1=6$$

$$\mu_2=25 \quad \sigma_2=3$$

$$DK_1 = \frac{6}{70} * 100 = 8,57$$

$$DK_2 = \frac{3}{25} * 100 = 12$$

DK sı büyük olanı yani 2. hisse senedini tercih etmelidir. Şayet riski sevmeyen yatırımcı olsaydı, ona değişimi görece daha az olan ilk hisse senedini almasını tavsiye ederdik. Borsada riskli senetlerin getirisi de kaybı da büyüktür. Ama riski seven yatırımcı “risk yoksa getiri de yoktur” mantığına sahip olduğu için bu onun tercihidir. Riski düşük olan senetler her ne kadar güvenli olsa da, maalesef getirisi de düşüktür. Elbette bu bir tercih meselesidir.

4.2. Chebyshev Teoremi

Ortalama ve standart sapmanın birlikte kullanıldığı bir durum da, verilerin yüzde kaçının hangi aralıkta bulunduğunu ölçmeye yarayan Chebyshev Teoremi’dir. Ortalaması μ , standart sapması σ olan bir veri kümesindeki gözlemlerin $\mu \pm k\sigma$ aralığına düşenlerin

oranı en az $1 - \frac{1}{k^2}$ ’dir. **Burada k, 1’den büyük bir sayıdır.**

Örnek: Bir lisedeki öğrencilerin IQ ortalaması 105, standart sapması 9’dur. Öğrencilerin en az % kaç 85,2 ile 124,8 arasında IQ’ya sahiptir?

$$\mu - k\sigma = 85,2$$

$$\mu + k\sigma = 124,8$$

Buradan $k=2,2$ elde edilir.

$$1 - \frac{1}{(2,2)^2} = 0,79$$

Yorum: Öğrencilerin en az %79'u verilen aralıkta IQ skoruna sahiptir.

Uygulamalar

Uygulama Soruları

Bu Bölümde Ne Öğrendik Özeti

Bölüm Soruları

12, 15, 4, 8,4 rakamlarından oluşan serinin varyans ve standart sapma değerlerini hesaplayınız.

2. Risk almak istemeyen bir yatırımcı için hangi hisse senedini önerirsiniz?

A hisse senedi ortalama getiri 14 , standart sapması 9

B hisse senedi ortalama getiri 12 standart sapması 4

3. Aşağıdaki sınıflanmış seride varyans, standart sapma ve değişim katsayısını hesaplayınız:

Sınıflar	fi
10-20	2
20-30	4
30-40	7
40-50	2

4. Yukarıdaki soru için düzeltilmiş varyansı hesaplayınız. (Sheppard düzeltmesi)

5. Sheppard düzeltmesi ne tür serilerde yapılır? Neden ihtiyaç duyulur?

6. Aritmetik ortalaması 10, varyansı 11 olan bir seride değişim katsayısı yüzde kaçtır?

7. Aşağıdaki sıklık serisinde değişim katsayısını hesaplayınız.

Xi	fi
10	4
11	2
14	3
16	1

5. MOMENTLER

Bu Bölümde Neler Öğreneceğiz?

- Sıfıra göre ve ortalamaya göre momentler
- Momentler yoluyla çarpıklık ve basıklık ölçümleri
- König Teoremi

Bölüm Hakkında İlgi Oluşturan Sorular

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Sıfıra göre ve ortalamaya göre momentler	Momentler yoluyla çarpıklık ve basıklık ölçümelerini öğreneceğiz.	Metinler, çözümlü problemler, alıştırma ve interaktif materyaller ile konuların daha kolay anlaşılması sağlanacaktır.

Anahtar Kavramlar

- Sıfıra göre momentler
- Ortalamaya göre momentler
- Çarpıklık
- Basıklık
- König Teoremi

Giriş

5.1. Momentler

Moment, bir rastlantı değişkeninin sıfıra göre veya ortalamaya göre çeşitli kuvvetlerinin beklenen değeridir.

Sıfıra göre demek, sıfırdan sapmaların kuvvetlerinin beklenen değeri demektir. Ortalamaya göre moment ise, ortalamadan sapmaların kuvvetlerinin ortalamalarıdır.

5.1.1.Sıfıra Göre Momentler (basit moment)

$$M_R = \frac{\sum X_i^R}{N} \quad M_R = \frac{\sum f_i X_i^R}{\sum f_i} \quad M_R = \frac{\sum f_i m_i^R}{\sum f_i}$$

Sıfıra göre 1. Moment aritmetik ortalamaya eşittir.

$$M_1 = \frac{\sum X_i}{N} = \bar{X}$$

Sıfıra göre 2. Moment, kareli ortalamanın karesine eşittir.

$$M_2 = \frac{\sum X_i^2}{N} = KO^2$$

Sıfıra göre 3. Moment aşağıdaki gibi hesaplanır:

$$M_3 = \frac{\sum X_i^3}{N}$$

Sıfıra göre 4. Moment aşağıdaki gibi hesaplanır:

$$M_4 = \frac{\sum X_i^4}{N}$$

5.1.2.Ortalamaya Göre Momentler

$$\mu_R = \frac{\sum (X_i - \bar{X})^R}{N} \quad \mu_R = \frac{\sum f_i (X_i - \bar{X})^R}{\sum f_i}$$
$$\mu_R = \frac{\sum f_i (m_i - \bar{X})^R}{\sum f_i}$$

Basit seride ortalamaya göre 1. Moment sıfıra eşittir.

$$\mu_1 = \frac{\sum (X_i - \bar{X})^1}{N} = 0$$

Ortalamaya göre 2. Moment varyansı verir.

$$\mu_2 = \frac{\sum (X_i - \bar{X})^2}{N} = \sigma^2$$

Ortalamaya göre 3. Moment aşağıdaki gibidir:

$$\mu_3 = \frac{\sum (X_i - \bar{X})^3}{N}$$

Ortalamaya göre 4. Moment aşağıdaki gibidir:

$$\mu_4 = \frac{\sum (X_i - \bar{X})^4}{N}$$

Momentler temel olarak şunu ölçerler:

1. momentler; merkezi konum ölçüsü (ortalama)
2. momentler; varyans
3. momentler; çarpıklık
4. momentler; basıklık.

5.2. Momentler Yoluyla Çarpıklık Ölçümü

Çarpıklık ve basıklık ölçüleri bir dizideki gözlem değerlerinin saçılımının (dağılımının) şeklini ortaya koyarlar. Bu ölçüler yorumlanırken normal dağılımın özellikleri referans alınır. Çünkü Normal dağılım eğrisi tam simetrik ve normal (ideal) bir basıklığa sahiptir. Çarpıklık ölçüsü serinin frekans (sıklık) dağılımının simetrik dağılımdan uzaklaşma derecesini gösterirken, basıklık ölçüsü gözlemlerin normal dağılım eğrisinin ne kadar altında ya da üstünde dağılıma sahip olduğunu gösteren ölçülerdir.

Momentler yoluyla çarpıklık ve basıklık aşağıdaki gibi hesaplanır. Çarpıklık katsayısı α_3 'tür.

$$\alpha_3 = \frac{\mu_3}{\sigma^3}$$

$\alpha_3 < 0$ ise sola çarpık

Veya;

$$\text{Çarpıklık} = \frac{3(\bar{X} - \text{Medyan})}{s}$$

$$\text{Çarpıklık} = \frac{(\bar{X} - \text{Mod})}{s}$$

Burada s, standart sapmadır. Şayet;

$\alpha_3 = 0$ ise simetrik. Simetrik dağılım gösteren serilerde merkezi eğilim ölçüleri, dağılımın tam ortasında yer alır. Gözlemlerin %50 'si merkezi eğilim ölçüsünden büyük, diğer yarısı ise küçüktür.

$\alpha_3 > 0$ ise sağa çarpıktır.

$\alpha_3 < 0$ ise sola çarpıktır.

5.3. Momentler Yoluyla Basıklık Ölçümü

Basıklık katsayısı α_4 'tür.

$$\alpha_4 = \frac{\mu_4}{\sigma^4} = \frac{\mu_4}{(\text{var yans})^2}$$

$\alpha_4 < 3$ ise normalden basık

$\alpha_4 = 3$ ise normal basıklıkta veya sivrilikte (ideal sivrilik, mesela normal dağılımın basıklığı 3'tür)

$\alpha_4 > 3$ ise normalden sivri

Örnek: Aşağıdaki serinin sıfıra göre ve ortalamaya göre momentlerini hesaplayınız. Çarpıklık ve basıklığı araştırınız.¹⁷

Basit momentler (yani sıfıra göre momentler) aşağıdaki gibidir:

	X_i	X_i^2	X_i^3	X_i^4
	1	1	1	1
	3	9	27	81
	4	16	64	256
	6	36	216	1296
	10	100	1000	10000
	12	144	1728	20736
toplam	36	306	3036	32370
ortalama	6	51	506	5395
basit momentler	M1	M2	M3	M4

Ortalamaya göre momentler aşağıdaki gibidir:

	$X_i - X_{ort}$	$(X_i - X_{ort})^2$	$(X_i - X_{ort})^3$	$(X_i - X_{ort})^4$
	-5	25	-125	625
	-3	9	-27	81
	-2	4	-8	16
	0	0	0	0
	4	16	64	256
	6	36	216	1296

toplam	0	90	120	2274
ortalama	0	15	20	379
basit momentler	μ_1	μ_2	μ_3	μ_4

$$\mu_2 = \sigma^2 = 15$$

$$\sigma = \sqrt{15} = 3,872$$

$$\alpha_3 = \frac{\mu_3}{\sigma^3} = \frac{20}{3,872^3} = 0,34 > 0 \text{ sağa çarpık seri}$$

$$\alpha_4 = \frac{\mu_4}{\sigma^4} = \frac{379}{(15)^2} = 1,68 < 3 \text{ normalden basıktır}$$

Örnek: Aşağıdaki serinin sıfıra göre ve ortalamaya göre momentlerini hesaplayınız. Çarpıklık ve basıklığı araştırınız.¹⁸

Sıfıra göre momentler aşağıdaki gibidir:

	i	i	iXi	i^2	iXi^2	i^3	iXi^3	i^4	iXi^4
				4	2	8	4	6	8
			8	9	4	7	62	1	86
			6	6	4	4	56	56	024
			2	6	52	16	512	296	072
to plam		0	2		3		1		1
					82		954		0630

¹⁸ A.g.e.

or			4	1	9	5
talama		,1	9,1	7,7	31,5	
		1	2	3	4	

Ortalamaya göre momentler aşağıdaki gibidir:

	f_i	$(X_i - \bar{X})$	$f_i(X_i - \bar{X})^2$	$f_i(X_i - \bar{X})^3$	$f_i(X_i - \bar{X})^4$
	2,1	-6,3	41,323	-9,261	27,783
	1,1	-6,6	21,7	-1,331	7,986
	0,1	-0,4	0,04	0,001	0,004
	,9	3,3	5,27	,859	8,013
toplam	0		5,8	2,24	8,354
ortalama	0		,29	0,612	7,9177
	μ		μ^2	μ^3	μ^4

$$\mu_2 = \sigma^2 = 2,29$$

$$\sigma = \sqrt{2,29} = 1,51$$

$$\alpha_3 = \frac{\mu_3}{\sigma^3} = \frac{0,612}{1,51^3} = 0,177 > 0 \quad \text{sağa çarpık}$$

$$\alpha_4 = \frac{\mu_4}{\sigma^4} = \frac{7,9177}{(2,29)^2} = 1,52 < 3 \quad \text{Normalden basık bir seri}$$

Örnek: Aşağıdaki serinin sıfıra göre ve ortalamaya göre momentlerini hesaplayınız. Çarpıklık ve basıklığı araştırınız.¹⁹

Sıfıra göre momentler aşağıdaki gibidir:

Sınıflar	i	i^2	i^3	i^4	i^5	i^6	i^7	i^8	i^9
0									
-2	1	1	1	1	1	1	1	1	1
-4	3	9	27	81	243	729	2187	6561	19683
-6	5	25	125	625	3125	15625	78125	390625	1953125
-8	7	49	343	2401	16807	117649	823543	5764801	40353607
		Toplam	0	56	4	2	744	1	7088
		Ortalama		8,5		71,5		068	
			1	2	3	4			

Ortalamaya göre momentler aşağıdaki gibidir:

Sınıflar	i	i^2	i^3	i^4	i^5	i^6	i^7	i^8	i^9
0									
-2	1	4	8	6	2	64	128	12	

¹⁹ A.g.e.

2				-				-	-	1
-4		3	2	2	4	4	8	8	6	
4										
-6		5		0	0	0	0	0	0	
6				1		2		4	8	
-8		7		0	4	0	8	0	0	
		to			2	5	-	-	6	
		plam		0	4	6	64	96	08	
		or				3		-	3	
		talama		0		,5		6	8	

$$\mu_2 = \sigma^2 = 3,5$$

$$\sigma = \sqrt{3,5} = 1,87$$

$$\alpha_3 = \frac{\mu_3}{\sigma^3} = \frac{-6}{1,87^3} = -0,917 < 0$$

$$\alpha_4 = \frac{\mu_4}{\sigma^4} = \frac{38}{(3,5)^2} = 3,102 > 3$$

Seri, normalden sivri ve sola çarpık bir dağılıma sahiptir.

σ^2 'nin sağlaması:

$$\sigma^2 = KO^2 - \bar{X}^2 = M_2 - M_1^2 = 28,5 - 25 = 3,5$$

Örnek: Aşağıdaki gruplandırılmış serinin Pearson çarpıklık katsayısını moda dayalı olarak hesaplayıp yorumlayınız.

<u>X_i</u>	<u>f_i</u>
1-3	1
3-5	2
5-7	4
7-9	3

Çözüm: Serinin modu şöyle hesaplanır;

$$\text{Mod} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} * c = 5 + \frac{2}{2+1} * 2 = 6.33$$

Moda dayalı Pearson çarpıklık ölçütü;

$$\text{Çarpıklık} = \frac{(\bar{X} - \text{mod})}{s} = \frac{5.8 - 6.33}{1.87} = -0.28$$

Yorum: Seri hafif bir biçimde sola çarpıktır.

5.4. König Teoremi

König teoremi, basit momentlerden ortalamaya göre momentlere geçişi sağlar. İki moment türü arasındaki ilişkiyi aşağıdaki gibi açıklamaktadır:

$$\mu_2 = M_2 - M_1^2$$

$$\mu_3 = M_3 - 3M_1M_2 + 2M_1^3$$

$$\mu_4 = M_4 - 4M_1M_3 + 6M_1^2M_2 - 3M_1^4$$

Uygulamalar

Uygulama Soruları

Bu Bölümde Ne Öğrendik Özeti

Bölüm Soruları

1. Sıfıra göre ve aritmetik ortalamaya göre birinci, ikinci, üçüncü ve dördüncü momentler (sadece uygun olanlar) hangi ölçütlere karşılık gelir?

2. Aşağıda yer alan seriden hareketle sıfıra göre 1. 2. 3. ve 4. moment değerlerini hesaplayınız.

Notlar	Öğrenci sayısı
0-20	4
20-40	12
40-60	8
60-80	6
80-100	2

3. Yukarıdaki serinin aritmetik ortalamadan sapmalara göre ilk üç momentini hesaplayınız.

4. König teoremini bu seri için doğrulayınız.

5. Aşağıdaki seride α_3 ve α_4 hesaplayarak yorumlayınız:

Sınıflar	f_i
50-60	1
60-70	3
70-80	7
80-90	4
90-100	2

**6. ÇEKİLİŞLERLE İLGİLİ BAĞIMLILIK VE BAĞIMSIZLIK
KAVRAMLARI, OLASILIĞIN KLASİK VE FREKANS TANIMI,
KOŞULLU OLASILIK, OLASILIK FONKSİYONU, DAĞILIM
FONKSİYONU**

Bu Bölümde Neler Öğreneceğiz?

- 6.1.Çekilişlerle ilgili bağımlılık ve bağımsızlık kavramları
- 6.2.Olasılığın tanımları
- 6.3.Koşullu olasılık
- 6.4.Olasılık fonksiyonu
- 6.5.Dağılım fonksiyonu

Bölüm Hakkında İlgi Oluşturan Sorular

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
İadeli ve iadesiz çekilişlerde bağımlılık bağımsızlık durumlarının incelenmesi, olasılığın farklı tanımları, koşullu olasılık, kesikli rd nin olasılık fonksiyonu ve dağılım fonksiyonu	İadeli ve iadesiz çekilişlerde bağımlılık bağımsızlık durumlarının incelenmesi, olasılığın farklı tanımları, koşullu olasılık, kesikli rd nin olasılık fonksiyonu ve dağılım fonksiyonunun öğrenilmesi.	Metinler, çözümlü problemler, alıştırmalar ve interaktif materyaller ile konuların daha kolay anlaşılması sağlanacaktır.

Anahtar Kavramlar

- İadeli çekiliş
- İadesiz çekiliş
- Bağımsızlık
- Bağımlılık
- Olasılığın klasik tanımı
- Olasılığın frekans tanımı
- Koşullu olasılık
- Olasılık fonksiyonu
- Dağılım fonksiyonu

Giriş

6.1. Çekilişlerle İlgili bağımlılık ve Bağımsızlık kavramları

Çekilişler iadeli ise denemeler bağımsızdır, iadesiz ise bağımlıdır. Top çekme deneyinde çekilen toplar yerine iade edildiğinde her çekiliş sanki ilk çekilişmiş gibi olur yani olasılıklar değişmez, eğer çekilen top yerine konmazsa her defasında top eksileceği için her çekilişte olasılıklar değişir yani olaylar bağımlı hale gelirler.

Örnek: Bir torbada 10 tane top olsun. 3 mavi, 4 turuncu, 3 kırmızı renkte olsunlar. Peş peşe 3 top çekilsin. Bu topların iadeli ve iadesiz durumları ayrı ayrı ele alarak , MTK renkte olması ihtimallerini hesaplayınız.

Çözüm: Önce iadeli durumu ele alalım. İlk topun mavi olması ihtimali 3/10'dur. İkincinin turuncu olması 4/10, üçüncünün kırmızı olması 3/10'dur. Yani sorulan olasılık şöyle bulunur:

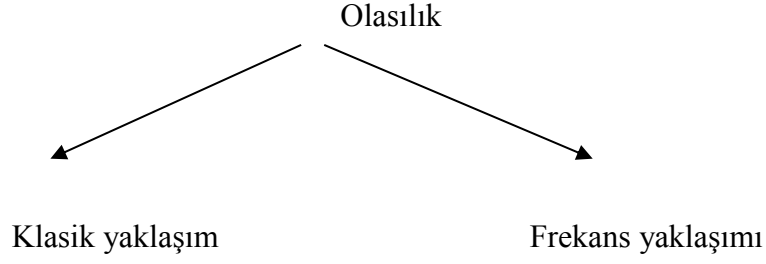
$$\frac{3}{10} * \frac{4}{10} * \frac{3}{10} = 36/1000 = 0.036$$

İadesiz durumda,

$$\frac{3}{10} * \frac{4}{9} * \frac{3}{8} = 0.05$$

Uyarı!!! Burada olasılıkları neden toplamıyoruz da çarpıyoruz sizce? Bunun sebebi, olayın ancak peş peşe 3 top çektikten sonra tamamlanmasıdır. Tıpkı bir şehirden bir başka şehre aktarmalı olarak giderken kullanabileceğimiz ulaşım araçları sayısını çarpmamız gibi. Toplama yapmak için olayların her basamağının birbirinden ayrı ve tamamlanmış olması yani ayrık olması gerekir.

6.2. Olasılığın Klasik ve Frekans Tanımı



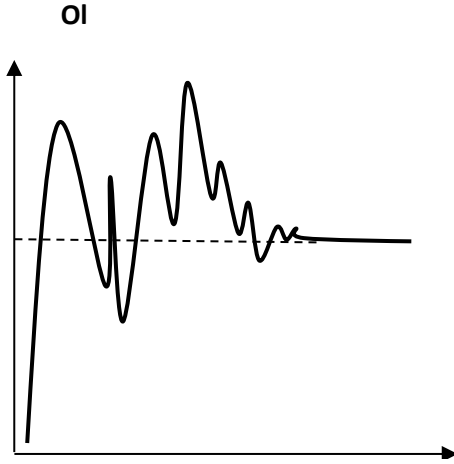
6.2.1. Klasik yaklaşım

Bir örneklem uzayında n tane mümkün sonuç olsun. A olayı S örneklem uzayının bir alt kümesi olmak üzere, A 'nın k tane elemanı olsun. Bu durumda klasik yaklaşıma göre $P(A)=k/n$ 'dir.

6.2.2. Frekans yaklaşımı

Bir paranın Yazı gelme olasılığı bilindiği üzere %50'dir. Ancak, diyelim ki bir parayı 10 defa atsak 5 defa Yazı 5 defa Tura gelecek diye bir zorunluluk yoktur. Onu da yazı veya dokuzu tura biri yazı, v.b. çok farklı sayıda örneklemeler ortaya çıkabilir. Fakat, deneme sayısı arttığında yani 500 defa, 1000 defa, 10000 defa atış yapılsa, olasılık bir noktadan sonra stabil hale gelir ve bu noktada olasılık, klasik yaklaşımın öngördüğü 0.5'tir. Bu durum aşağıdaki şekilde anlatılmak istenmiştir:

Şekil.6.1. Olasılığın Klasik ve Frekans Yaklaşımı



***n=10 defa atış sonucunda şu sonuçlar elde edilmiş olsun:

TTYYYYYYTTY

Bu örnekte $P(\text{Yazı})=7/10$, $P(\text{Tura})=3/10$ 'dur.

***Bir başka örnek şöyle olsun:

YYYYTTYYYT

Bu örnekte $P(\text{Yazı})=8/10$, $P(\text{Tura})=2/10$ 'dur.

TTYTTTTTTT

Bu örnekte $P(\text{Yazı})=1/10$, $P(\text{Tura})=9/10$ 'dur.

n=1000 defa olsun,

Örneğin $P(\text{Yazı})=468/1000$, $P(\text{Tura})=532/1000$ 'dir.

Halen klasik yaklaşımın öngördüğü olasılığa ulaşamadığına göre, deneme sayısı daha da arttırılmalıdır.

Söz gelimi $n=5000$ olsa, $P(\text{Yazı})=2500/5000$, $P(\text{Tura})=2500/5000$ olacaktır. (Tabii kesin bir durum değil, belki de olasılığın stabil (durağan) hale gelmesi 10000 defa deneme gerektirebilir)

Bazen de, bir olayın mümkün sonuçları eşit olasılığa sahip olmayabilir. Mesela atılan zar hileli ise ve çoğunlukla “2” geliyorsa, bu durumda zarın üst yüzüne gelen sayıların olasılıklarını belirlemek için çok sayıda deneme yapmak gerekir. Örneğin hileli bir para düşünelim, bu durumda $P(Y)=P(T)=0.5$ olması beklenemez. İşte böyle durumlarda;

$P(A)=(n \text{ denemede } A \text{’nin ortaya çıkma sayısı})/n$ ‘dir.

Örnek: Hileli bir para 1000 kez atılıyor. Sonuçlar kaydediliyor. 320 defa Tura, 680 defa Yazı geliyor. Bu durumda $P(\text{Yazı})=0.68$, $P(\text{Tura})=0.32$ ’dir.

Örnek: Bir paranın 3 kez atılması deneyinde B olayı; en az bir kez Tura gelmesi olsun. $P(B)=?$

$n(S)=2^3=8$ olası sonuç vardır.

$S=\{YYY, \textcolor{red}{YYT}, \textcolor{red}{YTY}, \textcolor{red}{TYY}, \textcolor{red}{TTY}, \textcolor{red}{TYT}, \textcolor{red}{YTT}, \textcolor{red}{TTT}\}$

$P(B)=7/8$ ’dir. Çünkü bahsedilen durumu sağlayan 7 sonuç vardır.

C olayı en fazla 2 Yazı gelme olayı ise $P(C)=?$

Yani 0,1 ya da 2 yazı gelmesi demektir.

$S=\{YYY, \textcolor{red}{YYT}, \textcolor{red}{YTY}, \textcolor{red}{TYY}, \textcolor{red}{TTY}, \textcolor{red}{TYT}, \textcolor{red}{YTT}, \textcolor{red}{TTT}\}$

$P(C)=7/8$ ’dir.

6.3. Koşullu Olasılık

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Yukarıdaki koşullu olasılığın sözel ifadesi şudur; B olayının gerçekleştiği biliniyorken, A olayının meydana gelme olasılığı. Sağ tarafta kalan olay, koşulu yaratan olaydır. Şayet A ile B olayları bağımsız iseler bu defa koşullu olasılık;

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) * P(B)}{P(B)} \quad \text{olur,}$$

P(B)'ler birbirini götürür ve; $P(A/B) = P(A)$ olur.

Örnek:

A ve B iki tedavi yöntemidir. Tedavilerin kilolarına göre tasnif edilmiş insanlardaki başarı olasılıkları aşağıdaki tabloda özetlenmiştir:

	Zayıf	Normal	Kilolu	Obes	Toplam
A	0,05	0,10	0,15	0,05	0,35
B	0,05	0,20	0,25	0,15	0,65
Toplam	0,10	0,30	0,40	0,20	1,00

a) A yöntemiyle başarı olasılığı nedir?

$$P(A)=0,35$$

b) B yöntemiyle başarı olasılığı nedir?

$$P(B)=0,65$$

c) Seçilen bir kişinin kilolu olduğu biliniyorken, A yöntemiyle tedavi olma olasılığı nedir?

$$P(A/\text{kilolu})= P(A \cap \text{kilolu})/P(\text{kilolu})= 0.15/0.4=0.375$$

d) Seçilen bir kişinin zayıf olduğu biliniyorken B yöntemiyle tedavi olma olasılığı nedir?

$$P(B/\text{zayıf})= P(B \cap \text{zayıf})/P(\text{zayıf})=0.05/0.10 = 0.5$$

e) A yöntemiyle tedavi olduğu biliniyorken bu kişinin obez olma olasılığı?

$$P(\text{obez}/A)= P(\text{obez} \cap A)/P(A)= 0.05/0.35 =0.143$$

$$\mathbf{f)} P(\text{normal}/B)= P(\text{normal} \cap B)/P(B)= 0.2/0.65 =0.308$$

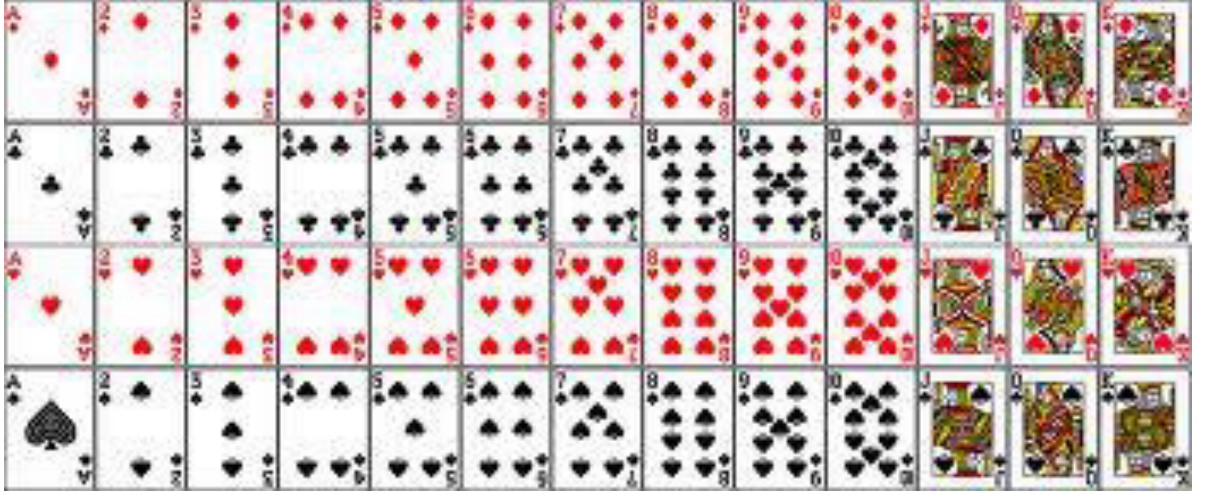
$$\mathbf{g)} P(A/\text{zayıf})= P(A \cap \text{zayıf})/P(\text{zayıf})= 0.05/0.10 = 0.5$$

$$\mathbf{h)} P(B/\text{normal})= P(B \cap \text{normal})/P(\text{normal}) = 0.20/ 0.30 = 0.667$$

Örnek: 52'lik desteden rastgele bir kart çekiliyor. Bu kartın kırmızı olduğu biliniyor iken As olması ihtimali nedir?

Çözüm: “A: kırmızı olması”, “B:As olması” şeklinde tanımlanıyor, buna göre $P(B/A)$ olasılığı soruluyor.

Şekil. 6.2. İskambil Destesinde Bulunan Kağıtlar



Toplam 52 kart vardır. 4 gruptan oluşur; sinek ve maça (siyah renkliler), kupa ve karo (kırmızı renkliler) A harfiyle gösterilen kağıdın özel adı “As” tır. Her grupta birer tane bulunur, 1 yerine geçer. Her grupta toplam 13 kağıt vardır. J: Vale (11 demektir), Q : Kız (12 demektir), K: Papaz (13 demektir). Yukarıdaki şekilde yukarıdan aşağı doğru gruplar karo, sinek, kupa ve maçadır.

Örneğimize dönersek; $P(A)=26/52$

$P(B) = 4/52$

$P(A \cap B)= 2/52$

O halde $P(B/A)= \frac{2/52}{26/52} = 2/26$ ’dır.

Uyarı!!! Burada A ile B’nin kesişimini niçin 2/26 değil de 2/52 aldık? Çünkü B olayı As olması şeklinde tanımlanmıştır. Dolayısıyla sadece kırmızı asların bulunduğu grubu esas alamayız, sinek ve maça asları da as olduğuna göre destenin tümünü esas almalıyız.

Örnek: $P(\text{Karo/sekiz})=?$

Çözüm: $(1/52)/(4/52)=0,25$

Örnek: maça veya papaz?

Çözüm: $P(\text{maça} \cup \text{papaz})=P(\text{maça})+P(\text{papaz})-(\text{maça} \cap \text{papaz}) = (13/52)+(4/52)-(1/52)$

Örnek: $P(\text{karo/ papaz})=1/4$

6.4. Kesikli Rd'nin Olasılık Fonksiyonu

Kesikli bir rd nin olasılık dağılımına olasılık fonksiyonu denir. X'in alabileceği değerlerin bu değerleri alma olasılıklarıyla birlikte yazılmış haline olasılık fonksiyonu denir. Herhangi bir fonksiyonun kesikli rd için olasılık fonksiyonu olabilmesi için örneklem uzayındaki tüm birimlerin olasılıklarının $[0,1]$ aralığında olması ve tüm olasılıkların toplamının 1'e eşit olması gerekmektedir.

Örnek: Bir olayın aşağıda gösterildiği gibi 4 mümkün sonucu olsun. X rd kesikli olduğuna göre, $P(e_4)=?$

$$P(e_1)=0.3$$

$$P(e_2)=0.4$$

$$P(e_3)=0.2$$

Çözüm: $P(e_4)=1-(0.3+0.4+0.2)=0.1$

Örnek: İki zar atılması deneyinde X rd. Üste gelen yüzlerin çarpımı şeklinde tanımlı olsun. X rd'nin olasılık fonksiyonunu yazınız.

Çözüm: Öncelikle S örneklem uzayını oluşturup X rd nin aldığı değerleri bulacağız. Sonra bu değerleri alması olasılıklarını bularak, iki sütun haline, solda x değerleri sağda ise $P(x)$ 'ler olmak üzere X rd'nin olasılık fonksiyonunu yazacağız.

	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	8	10	12
3	3	6	9	12	15	18
4	4	8	12	16	20	24
5	5	10	15	20	25	30
6	6	12	18	24	30	36

<u>x</u>	<u>P(X=x)</u>
1	1/36
2	2/36
3	2/36
4	3/36
5	2/36
6	4/36
8	2/36
9	1/36
10	2/36
12	4/36
15	2/36
16	1/36
18	2/36
20	2/36
24	2/36
25	1/36
30	2/36
36	1/36

X 'in olasılık fonksiyonu bu şekildedir, solda x değerleri, sağda da bu değerleri hangi olasılıklarla almakta olduğunu yazdığımızda X'in olasılık fonksiyonunu yazmış oluruz. Ancak burada bir sağlama yapalım, olasılık fonksiyonu olma şartlarından biri, her bir olasılığın $[0,1]$ arasında olmasıydı, baktığımızda bu koşulun sağlandığını görüyoruz. Peki $\sum P(x) = 1$ mi? Evet, toplam olasılık 1'e eşittir, o halde yaptığımız işlem doğrudur.

6.5. Dağılım Fonksiyonu:

X rd'nin dağılım fonksiyonu $F(X)$ ile gösterilir. X rd nin x 'e eşit veya daha küçük olması olasılığıdır. Bu birikimli olasılığa dağılım fonksiyonu denir.

$$F(X) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$$

Örnek: Hilesiz bir zar atılsın. X rd üste gelen sayılar olsun. $F(X)$ 'i yazınız.

Çözüm:

Dağılım fonksiyonu, olasılık fonksiyonunun birikimli halidir, o nedenle öncelikle $f(x)$ 'i oluşturmamız.

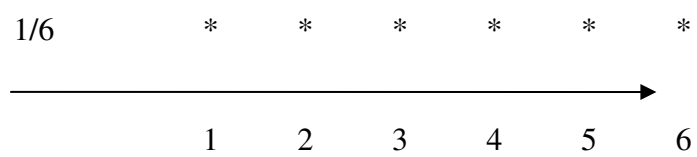
$$S = \{1, 2, 3, 4, 5, 6\}$$

x_i	1	2	3	4	5	6
$f(x_i)$	1/6	1/6	1/6	1/6	1/6	1/6

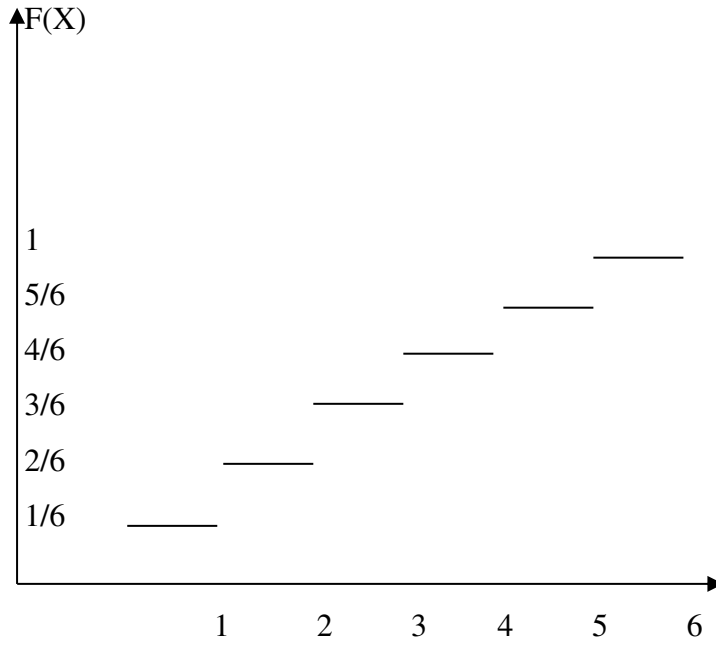
$$F(X) = \begin{cases} 0 & x_i < 1 \\ 1/6 & 1 \leq x_i < 2 \\ 2/6 & 2 \leq x_i < 3 \\ 3/6 & 3 \leq x_i < 4 \\ 4/6 & 4 \leq x_i < 5 \\ 5/6 & 5 \leq x_i < 6 \\ 1 & 6 \leq x_i \end{cases}$$

Şekil. 11.3. Olasılık Fonksiyonunun Grafiği

$f(x_i)$



Şekil.11.4. Dağılım Fonksiyonunun Grafiği



Örnek:

x_i	-3	0	2	3
$f(x_i)$	0.2	0.1	0.4	c

- $c=?$
- $P(X>0)=?$
- $P(X=-2)=?$
- $F(X)=?$

Çözüm:

- $f(x_i)=P(X=x_i)$ ve toplam $f(x_i)=1$ olmak zorundadır, o halde $c= 1-(0.2+0.1+0.4) = 0.3$ 'tür.
- $P(X>0)= P(2)+P(3) = 0.4 + 0.3 =0.7$
- $P(X=-2) = 0$

d.

$F(x) = \left\{ \begin{array}{l} \\ \\ \\ \\ \end{array} \right.$	0	$x_i < -3$
	0.2	$-3 \leq x_i < 0$
	0.3	$0 \leq x_i < 2$
	0.7	$2 \leq x_i < 3$
	1	$3 \leq x_i$

Uygulamalar

Uygulama Soruları

Bu Bölümde Ne Öğrendik Özeti

Bölüm Soruları

1. Bir torbada 4 Sarı, 5 Yeşil, 3 Mavi top vardır. Peş peşe 3 top çekiliyor. İadesiz çekiliş şartı ile $P(\text{SYM})=?$

2. Bir torbada 4 Sarı, 5 Yeşil, 3 Mavi top vardır. Peş peşe 3 top çekiliyor. İadeli çekiliş şartı ile $P(\text{YYM})=?$

3. Bir iskambil destesinden rastgele bir kart çekiliyor. Kartın Kız olduğu biliniyorken maça olması ihtimali nedir?

4. Bir iskambil destesinden rastgele bir kart çekiliyor. Kartın siyah olduğu biliniyorken 4'lü olması ihtimali nedir?

5. iskambil destesinden bir kart çekiliyor:

- a. $P(\text{As U Kupa})$ b. $P(\text{Karo U Sinek})$ c. $P(6/ \text{Siyah})$ d. $P(\text{Sinek U Kız})$

7. BİNOM DAĞILIMI VE POISSON DAĞILIMI

Bu Bölümde Neler Öğreneceğiz?

7.2.Binom Dağılımı

7.3.Poisson Dağılımı

Bölüm Hakkında İlgi Oluşturan Sorular

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Binom Dağılımı ve Poisson Dağılımı	Kesikli rd lerin başlıca dağılım fonksiyonlarından Binom dağılımı ve Poisson dağılımı öğrenilecektir.	Metinler, çözümlü problemler, alıştırmalar ve interaktif materyaller ile konuların daha kolay anlaşılması sağlanacaktır.

Anahtar Kavramlar

- Binom dağılımı
- p
- q
- n
- npq
- poisson dağılımı
- λ

Giriş

7.1. Bazı Özel Kesikli Dağılımlar

Bu bölümde kesikli rd lerin başlıca dağılım fonksiyonlarından Binom ve Poisson dağılımlarına yer vereceğiz.

7.1.1. Binom Dağılımı

Binom dağılımı, temeli Bernoulli denemelerine dayanır. Bernoulli denemeleri sadece iki sonucu olan ve bir defa tekrarlanan deneylerdir. Örneğin, bir paranın bir defa atılması, bir sınavın sonucunda geçme veya kalma, bir ürünün defolu veya sağlam olması, v.b. sadece 1 kez tekrar eden ve iki sonucu olan denemelerdir. Binom rd. ise, mümkün iki sonucu olan Bernoulli denemelerinin 1’den fazla sayıda tekrar edilmesiyle ortaya çıkmıştır. Yani, bir paranın 2 defa yada daha çok sayıda atılması, defolu olup olmama durumunun sadece 1 üründe değil 1’den fazla üründe incelenmesi, birden fazla sayıda dersten geçme veya kalma durumu,v.b.

Binom rd mümkün 2 sonucu olan bir rd dir demiştik. Bu sonuçların olasılıkları “p” ve “1-p” ile gösterilir. “1-p” ise kısaca “q” ile gösterilir. Bu durumda “p+q=1” dir. Şayet bir denemenin 2’den fazla sayıda olası sonucu bulunuyorsa, artık o rd’nin Binom dağıldığı söylenemez.

Binom Dağılımının Özellikleri

1. 2 sonuçlu n adet özdeş denemeden oluşur.
2. Denemeler birbirinden bağımsızdır. Örneğin bir paranın 2 kez atılması deneyinde, ikinci atışta Yazı gelmesi, birinci atışın sonucuna bağlı değildir. Her bir deneme ilk deneme gibidir, birbirinden bağımsızdır. Deneylerin sonuçları bir sonraki denemenin sonucunu etkilemez.
3. İlgilenilen olasılık “p” ile ifade edilir, onun tümleyeni ise “1-p” dir, kısaca “q” ile ifade edilir. “p+q=1” dir.

Binom Dağılımının Olasılık Dağılım Fonksiyonu

$$P(X = x) = \binom{n}{x} p^x (q)^{n-x}$$

Binom dağılımının olasılık fonksiyonunda yer alan $\binom{n}{x}$, C_x^n , yani n'in x'li

kombinasyonu demektir ve $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ 'dir.

Binom dağılan X rd'nin beklenen değeri; $E(X)=np$, Varyansı; $V(X)=npq$ 'dur.

Bu durum özet bir şekilde şöyle ifade edilir: $X \sim \text{Bin}(np, \sqrt{npq})$

Sözel olarak açıklarsak, böyle bir gösterimle karşılaştığımızda, "*X rd; np ortalama ve $\sqrt{npq}$ sapma ile Binom dağılmaktadır*" deriz.

Örnek: Bir ütü fabrikasında üretilen ütülerin %15'i defoludur.

- a) Tesadüfen seçilen 5 ütünün 3'ünün bozuk olması ihtimali nedir?
- b) Tesadüfen seçilen 4 ütünün 2'sinin bozuk olması ihtimali nedir?

Çözüm:

a) Bu sorunun bir binom sorusu olduğunu nasıl anlıyoruz? Öncelikle deneyin mümkün sonuçlarına bakıyoruz. Defolu veya sağlam, yani sadece 2 sonucu var. Peki, deneme 12'den çok defa mı tekrarlanıyor? Evet, 5 ütü seçildiğine göre demek ki n deneme sayısı 5'tir. O halde koşullar binoma uyuyor demektir. Bize sorulan ise $P(X = 3)$ 'tür.

$$P(X = x) = \binom{n}{x} p^x (q)^{n-x} \text{ 'tür.}$$

$$\text{b) } P(X = 2) = \binom{4}{2} 0.15^2 (0.85)^2$$

Örnek: Bir fabrikadaki ürünlerin %10'unun defolu olduğu biliniyor. Tesadüfen seçilen 5 ürünün;

- a) 3'ünün defolu olması ihtimali nedir?
- b) 1'inin sağlam olması?

Çözüm:

$$\text{a) } P(X = 3) = \binom{5}{3} 0.10^3 (0.90)^2 = 0.0081$$

b) 1'inin sağlam olması demek, 4'ünün defolu olması demektir.

$$P(X = 4) = \binom{5}{4} 0.10^4 (0.90)^1 = 0.00045$$

Veya öyle de hesaplanabilirdi;

$$P(X = 1) = \binom{5}{1} 0.90^1 (0.10)^4 = 0.00045$$

Örnek: Bir satın alma uzmanı mal alırken, satın alacağı bir seri malda %5'ten fazla defolu ürün olması halinde satın almaktan vazgeçmektedir. Bir seride 200 adet ürün olduğu biliniyorken, tesadüfen seçilen 10 ürünün kaçının defolu olduğu tespit edilmek isteniyor.

a) Seçilen 10 üründen 6'sının defolu olma ihtimali nedir?

b) Seçilen 10 üründen 2 veya daha fazlasının defolu olma ihtimali nedir?

c) $E(X)=?$

d) $V(X)=?$

Çözüm:

$$\text{a) } P(X = 6) = \binom{10}{6} 0.05^6 (0.95)^4 = 0.00000267$$

$$\text{b) } P(X \geq 2) = 1 - P(X < 2) = P(X = 0) + P(X = 1)$$

$$P(X = 0) = \binom{10}{0} 0.05^0 (0.95)^{10} = 0.59$$

$$P(X = 1) = \binom{10}{1} 0.05^1 (0.95)^9 = 0.315$$

$$P(X \geq 2) = 1 - (0.059 + 0.315) = 0.095$$

$$\text{c) } E(X) = np = 10 \cdot 0.05 = 0.5$$

$$\text{d) } V(X) = npq = 10 \cdot 0.05 \cdot 0.95 = 0.475$$

Örnek: $X \sim \text{Binom}(1, \sqrt{0.8})$ ise;

$$\text{a) } P(X=1) = ?$$

$$\text{b) } P(X=2) = ?$$

Çözüm:

$$np = 1$$

$npq = \sqrt{0.8}$ olduğuna göre; $q = 0.8$ 'dir. bu durumda $p = 1 - 0.8 = 0.2$ 'dir. $np = 1$ olduğuna göre, $n = 5$ 'tir.

$$\text{a) } P(X = 1) = \binom{5}{1} 0.2^1 0.8^4$$

$$\text{b) } P(X = 2) = \binom{5}{2} 0.2^2 0.8^3$$

Örnek: $n=25$, $p=0.5$ ise $V(X)=?$

Çözüm:

$$E(X)=np=12.5$$

$$V(X)=npq=6.25$$

7.1.2. Poisson Dağılımı

Fransız matematikçi *Siméon Denis Poisson* tarafından 1837’de bulunmuştur. Kesikli bir dağılım olan Poisson dağılımı, sürekli bir uzayda (bir zaman dilimi, bir mekan) meydana gelen olayların sayılarını tanımlamakta kullanılır.

Örneğin,

- İstanbul’da bir hafta içinde meydana gelen trafik kazalarını sayısı
- Bir telefon santraline 5 dakika boyunca gelen aramaların sayısı
- Bir markette kasiyerin yarım saatlik bir zaman diliminde hizmet verdiği müşteri sayısı
- Bir kitabın bir sayfasındaki yazım hatalarının sayısı
- Bir limanda bir gün içinde yük boşaltmak için gelen gemilerin sayısı

Poisson dağılımının olasılık fonksiyonu şöyledir:

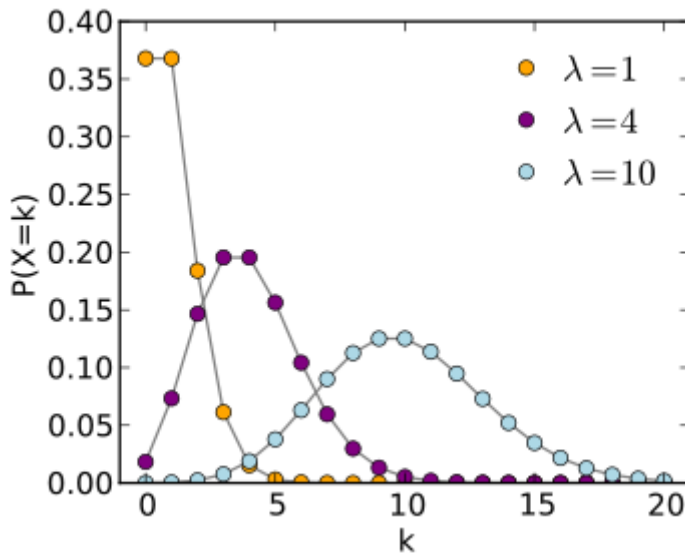
$$p(X = x) = \frac{e^{-\lambda} \times \lambda^x}{x!}$$

$$E(X) = V(X) = \lambda$$

$$e \cong 2,718$$

Görüldüğü üzere, dağılımın beklenen değeri ve varyansı birbirine eşittir ve λ dır. λ büyüdükçe, Poisson dağılımı normal dağılıma yaklaşır. Aşağıda Poisson dağılımının olasılık dağılım fonksiyonunun grafiği görülmektedir.

Şekil.7.1. Poisson Olasılık Dağılım Fonksiyonu²⁰



Örnek: Bir markette bir kasiyerin bir saat içinde ortalama 15 müşteriye hizmet verdiği bilinmektedir. Bu bilgiye dayanarak;

- Bir saat boyunca hiç müşteri gelmemesi?
- En fazla 1 müşteri gelmesi?
- Birden fazla müşteri gelmesi? Olasılıklarını hesaplayınız.

²⁰ http://upload.wikimedia.org/wikipedia/commons/1/16/Poisson_pmf.svg

Çözüm:

a. λ =ortalama=15

bize sorulan $P(X=0)$ 'dır.

$$p(X = 0) = \frac{e^{-2} \times 2^0}{0!}$$

$$\text{b. } P(X \leq 1) = P(X=0) + P(X=1) = \frac{e^{-15} \times 15^0}{0!} + \frac{e^{-15} \times 15^1}{1!}$$

$$\text{c. } P(X > 1) = 1 - P(X=0) = 1 - \frac{e^{-2} \times 2^0}{0!} + \frac{e^{-2} \times 2^1}{1!}$$

Örnek: Bir kitapta yapılan yazım hatalarının sayısı sayfa başına ortalama 2'dir.

a. Bir sayfada 5 yazım hatası olma olasılığı nedir?

b. En çok 1 yazım hatası olması?

c. Hiç yazım hatası olmaması?

Çözüm:

$$\text{a. } p(X = 5) = \frac{e^{-2} \times 2^5}{5!}$$

b. $P(X \leq 1) = P(X=0) + P(X=1) = \frac{e^{-2} \times 2^0}{0!} + \frac{e^{-2} \times 2^1}{1!}$

c. $p(X = 0) = \frac{e^{-2} \times 2^0}{0!}$

Uygulamalar

Uygulama Soruları

Bu Bölümde Ne Öğrendik Özeti

Bölüm Soruları

. Poisson dağılımı ve binom dağılımı arasındaki farkları X rastlantı değişkenini baz alarak belirtiniz.

2. Bir kitapta yapılan yazım hatalarının sayısı sayfa başına ortalama 3'dir. Bir sayfada 4 yazım hatası olma olasılığı nedir?

3. Yukarıdaki soruda hiç hata ile karşılaşılma olasılığını hesaplayınız

4. Bir fabrikadaki ürünlerin %15'unun defolu olduğu biliniyor. Tesadüfen seçilen 6 üründen 3'ünün defolu olması ihtimali nedir?

5. 4. Soruda ürünlerin hepsinin defolu olması olasılığı nedir?

8. SÜREKLİ DAĞILIMLAR- NORMAL DAĞILIM

8.1.Bu bölümde İstatistik I dersinde anlatılmış olan rastlantı değişkenlerinin olasılık dağılımları konusuna, sürekli rassal değişkenler dikkate alınarak devam edilmiştir.Normal dağılım ve standart normal dağılım konuları dersimizin konusunu oluşturacaktır.

Bölüm Hakkında İlgi Oluşturan Sorular

- Normal dağılım neden en sık kullanılan dağılımdır?
- Standart normal dağılımı neden kullanırız

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Normal dağılım, Standart normal dağılım	Günlük hayatta, bir çok alanda karşılaşılabilecek çeşitli olasılık problemlerinin çözümü	Çözümlü problemler ve tanımlar

Anahtar Kavramlar

- Rastlantı Değişkeni
- Normal Dağılım
- Standartlaştırma
- Standart Normal Dağılım

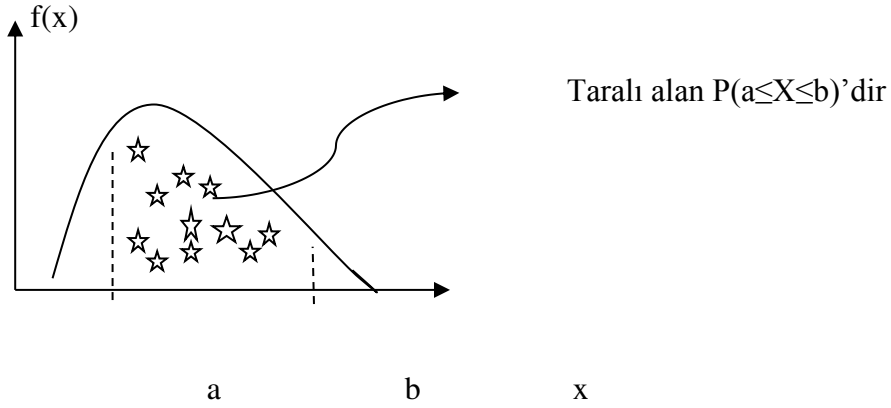
Giriş

İstatistik I dersinde olasılık dağılımının tanımı, kesikli rastlantı değişkeninin olasılık dağılımları; Binom, Poisson, Hipergeometrik ... dağılımlar konularına yer verilmişti. Bu derste ise rastlantı değişkeninin sürekli olması hali ele alınacak ve özel bir olasılık dağılımı olan normal dağılım anlatılacaktır.

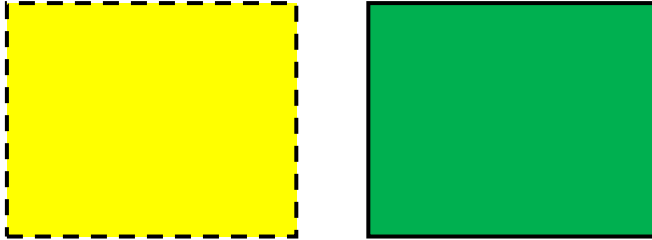
8.1. SÜREKLİ DAĞILIMLAR

Sürekli rastlantı değişkenlerinin(r.d.) alacağı değerler, belirli bir aralıkta yer alır. Örneğin X rastlantı değişkeninin a ve b gibi iki değer arasında bulunması olasılığı $P(a \leq X \leq b)$ şeklinde ifade edilir.

Olasılık yoğunluk fonksiyonu $f(x)$ ile gösterilir, kesikli olmayan (sürekli) bir eğridir. $f(x)$ 'in altında kalan alan X rastlantı değişkeninin olasılıklarını verir.



Sürekli rastlantı değişkenleri söz konusu olduğunda $P(a < X < b) = P(a \leq X \leq b)$ 'dir. Eşitlikler dahil olsa da olmasa da sonuç değişmez. Çünkü söz konusu bir aralıktır ve sonsuz değer içermektedir. Halbuki kesikli rd lerde eşitliklerin dahil olmaması olasılığı doğrudan değiştirebilmektedir ve eşitliklerin dahil olup olmaması önemlidir. Sürekli rd'lerde olasılıklar bir alana tekabül etmekte olduğundan, eşitliklerin dahil olup olmaması olasılığı değiştirmez. Bu durumu şu şekilde açıklayabiliriz. Aşağıda gördüğünüz iki karenin alanı birbiriyle aynıdır ve eşittir:



Kesikli rd lerde olasılıklar toplanırken, sürekli rd’lerde olasılıklar alan hesaplayarak bulunur. Belirli bir aralıktaki olasılık değeri ise bir alana tekabül etmektedir ve söz konusu aralıkta belirli integralin hesaplanmasıyla bulunur.

$$P(a \leq X \leq b) = \int_a^b f(x) dx \text{ 'dir.}$$

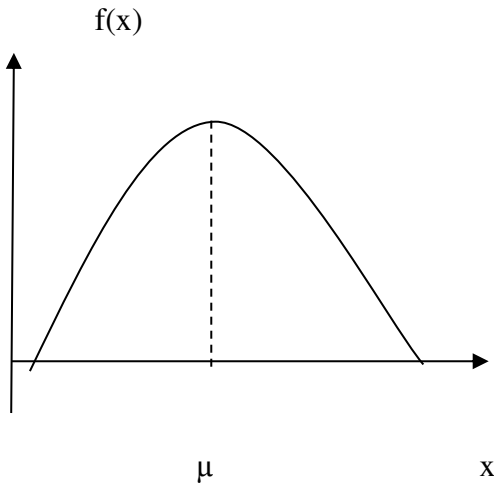
8.2. Sürekli Bir Fonksiyonun Olasılık Yoğunluk Fonksiyonu Olabilmesi İçin Gerekli Şartlar

1) $\forall x_i$ için $f(x_i) \sim [0,1]$, Yani X rd’nin alacağı her bir x değerini alma olasılığı $[0,1]$ aralığında olmak zorundadır. Zaten, olasılığın tanımından da bilindiği üzere, hiçbir olayın olasılığı negatif bir sayı veya 1’den büyük bir sayı olamaz.

2) $\int_{-\infty}^{\infty} f(x) dx = 1$ 'dir. Yani $f(x)$ ’in tanım aralığındaki toplam olasılık yani $f(x)$ eğrisinin altında kalan alan 1’e eşittir.

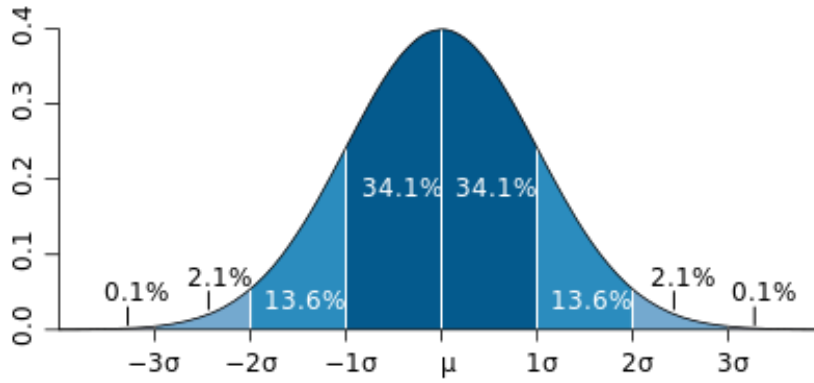
8.3. NORMAL DAĞILIM

İstatistikte sürekli dağılım denildiğinde ilk akla gelen en temel sürekli dağılım “normal dağılım”dır. Normal dağılım çan eğrisi biçimindedir.



μ normal dağılımın ortalamasıdır. Eğriyi iki eşit parçaya bölmektedir. Eğrinin altındaki toplam alan 1'e eşit olmak üzere, μ 'nın sağındaki ve solundaki alanlar birbirine eşit ve 0.5'tir. Normal dağılım simetrik bir yapıya sahiptir. Yani çarpıklık sıfırdır. Normal dağılımda; ortalama=medyan=mod'dur. Basıklığı ise idealdir, yani 4.moment ile hesaplanan α_4 basıklık katsayısı 3'e eşittir. Zaten diğer dağılımların basıklıkları değerlendirilirken, basık veya sivrilik durumlarına normal dağılım ile kıyaslama yapılarak karar verilir. Şayet $\alpha_4 < 3$ ise dağılım normalden basık, $\alpha_4 > 3$ ise normalden daha sivri bir dağılım olduğu söylenir.

Verilerin ortalama dolayındaki dağılımları aşağıdaki gibidir:

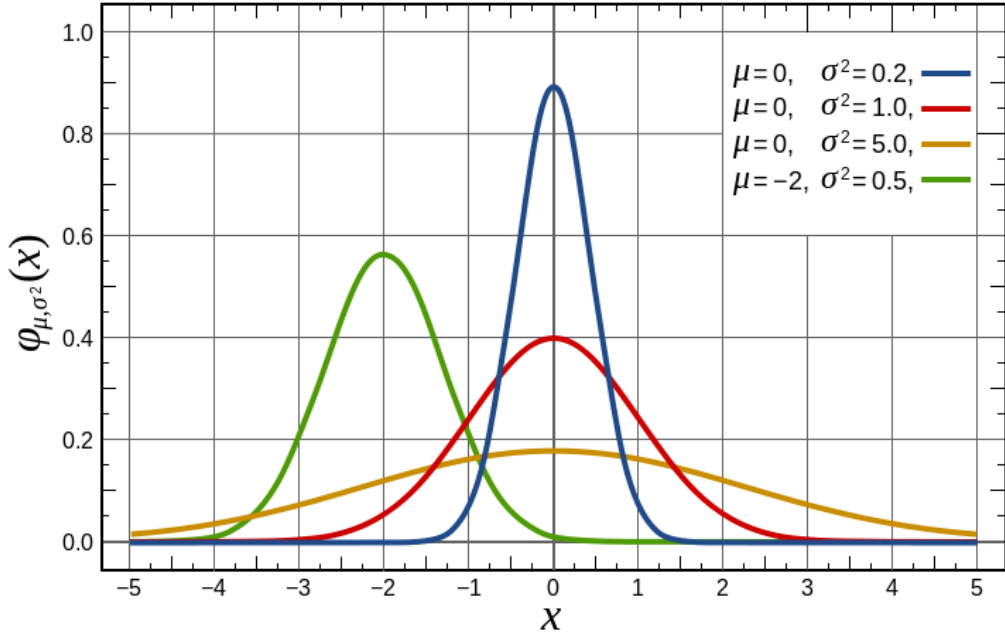


Verilerin %68.2'si (bu değer 2×34.1 'den elde edilmiştir) ortalamanın 1 standart sapma dolayında bulunur. %95.4'ü (bu değer 2×47.7 'den elde edilmiştir) ortalamanın 2 standart sapma dolayında bulunur. %99.6'sı (bu değer 2×49.8 'den elde edilmiştir) ortalamanın 3 standart sapma dolayında bulunur.

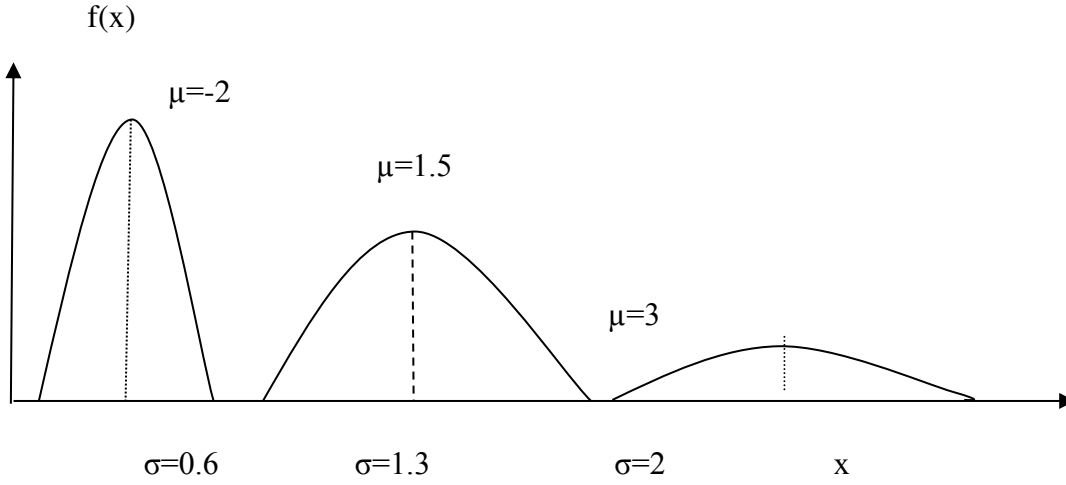
8.3.1 Normal Dağılımın Olasılık Yoğunluk Fonksiyonu

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

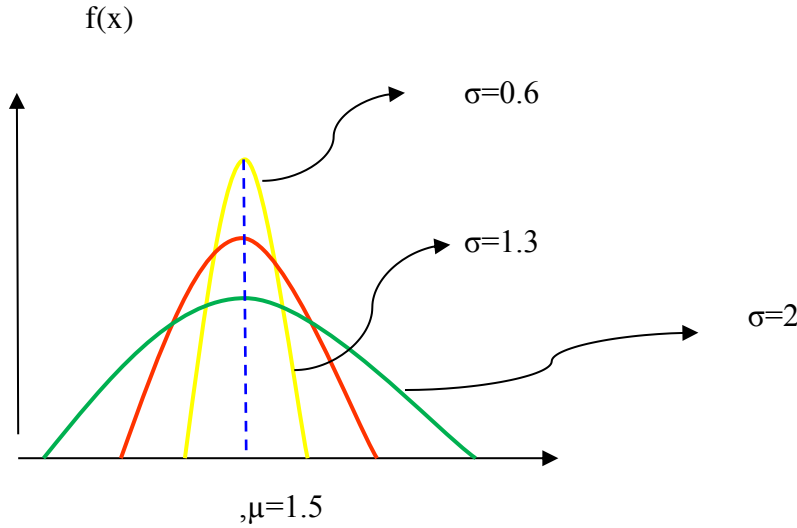
Burada $e=2.718$, $\pi=3.14$, σ : standart sapmadır.



Dağılımların yayılımları (değişimleri) standart sapma yani σ ile belirlenir. Aşağıdaki şekilde farklı varyanslara sahip normal dağılımlar görülmektedir:



veya daha açık bir şekilde ifade edersek, ortalaması aynı fakat varyansları farklı normal dağılım eğrilerinin grafiklerine bakalım:



X sürekli rd'nin olasılığını bulmak demek, verilen aralıkta belirli integral almaktır demiştik.

Örneğin $P(2 < X < 5) = \int_2^5 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$ 'tir. Bu olasılık değerinin hesaplanması

mümkün olmakla birlikte, bizi oldukça karışık ve alınması zor integrallerle karşı karşıya bırakmaktadır. Ancak olasılıklar bu integraller alınmadan da hesaplanabilmektedir. Bunu sağlayan ise “*standart normal dağılım*” dır.

Normal dağılımın standart hale getirilmesiyle yani normal dağılan X rd'nin standardize edilmesiyle standart normal dağılım bulunmuştur. z standart normal değişken, normal dağılan X rd'nin ortalamasından sapması olarak standart sapmasına bölünmesiyle elde edilmektedir. Standart normal dağılım, normal dağılımla aynı özelliklere sahiptir, ancak, ortalaması sıfır, varyansı 1'dir.

$X \sim \text{Normal}(\mu, \sigma^2)$ iken; yapılan $z = \frac{X - \mu}{\sigma}$ dönüşümüyle elde edilen z standart değişkeni; $z \sim \text{Standart Normal}(0,1)$ dağılmaktadır. Standart normal dağılımın olasılık yoğunluk fonksiyonu aşağıdaki gibidir:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z)^2}$$

$f(z)$ o.y.f., $f(x)$ normal o.y.f'da $\mu=0$, $\sigma=1$ konularak elde edilmiştir. Bu dönüşüm sayesinde artık olasılıkların bulunması son derece kolaylaşmıştır. X'ten z'ye dönüşüm

yapılarak X'in çeşitli değerleri için son derece basit hale gelen integrallerin alınmasıyla z tablosu oluşturulmuştur. Aşağıda X rd ve z std. değişkenine ilişkin o.y.f. ları görülmektedir:

$X \sim \text{Normal}(\mu, \sigma^2)$	$z = \frac{X - \mu}{\sigma}$	$z \sim \text{Standart Normal}(0,1)$
$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	dönüşümü yapılarak	$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z)^2}$

Örnek 1.1: $X \sim \text{Normal}(1,4)$ olmak üzere, $P(2 < X < 3) = ?$

1.çözüm:

$$\int_2^3 \frac{1}{\sqrt{2\pi} \cdot 2} e^{-\frac{1}{2}\left(\frac{x-1}{2}\right)^2} dx$$

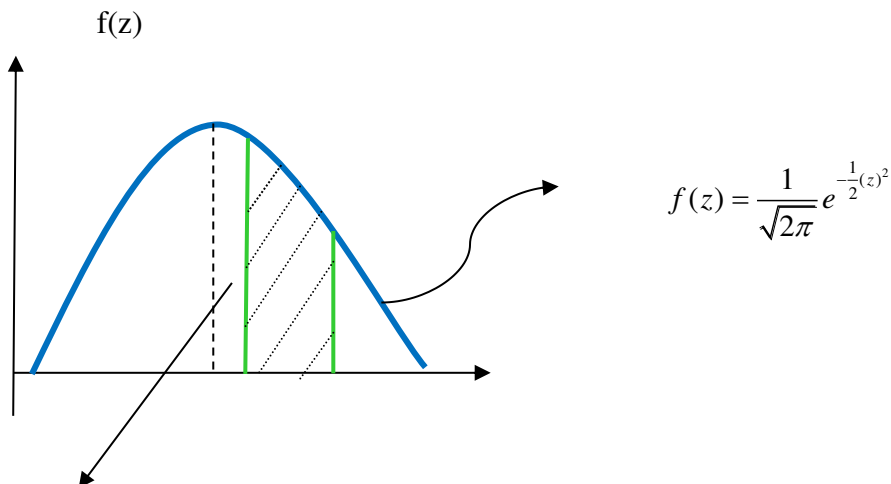
2.çözüm:

$$z_1 = \frac{2-1}{2} = 0.5$$

$$z_2 = \frac{3-1}{2} = 1$$

$$P(2 < X < 3) = \int_2^3 \frac{1}{\sqrt{2\pi} \cdot 2} e^{-\frac{1}{2}\left(\frac{x-1}{2}\right)^2} dx = P(0.5 < z < 1) \text{ yani; standart normal dağılım eğrisinin}$$

altında (0.5,1) arasındaki alandır. Bu durum aşağıda görülmektedir:



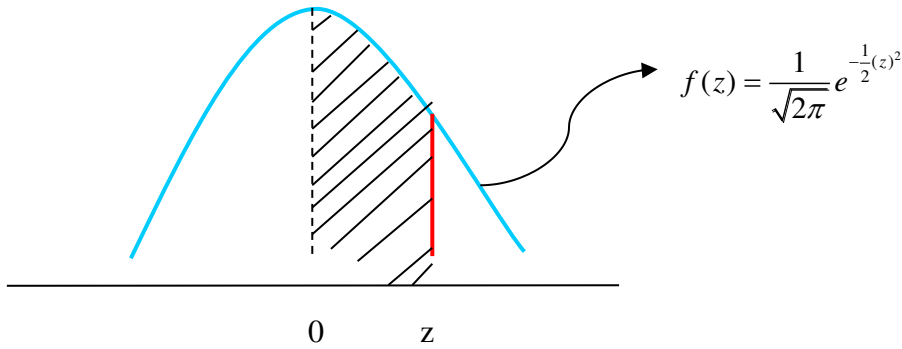
0 0.5 1 z

z tablosuna bakıldığında (0,0.5) arasındaki alanın büyüklüğü yani $P(0 < x < 0.5) = 0.1915$ 'tir. $P(0 < x < 1) = 0.3413$ 'tür. O halde $P(0.5 < x < 1)$ olasılığını bulmak için $P(0 < x < 1)$ 'den $P(0 < x < 0.5)$ çıkarılmalıdır.

$$P(0.5 < x < 1) = 0.3413 - 0.1915 = 0.1498$$

8.4. Z tablosunun (normal dağılım tablosu) özellikleri ve okunuşu

Z tablosu, X rd nin çeşitli değerleri için hesaplanan belirli integrallerden oluşan bir sayılar tablosudur. Tablonun en başında şöyle bir şekil vardır:



Bu şekil, z tablosundaki değerlerin 0 ile z arasındaki alana karşılık geldiği anlamına gelmektedir. ve tablo sayılarla şu şekilde devam eder:

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
--	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
.1	.0398	.0438	.0478							.0753
.2	.0793									.1141
.3	.1179									
.4										
.5										
.6										
.7										
.8										
.9										.3389
.0										.3621
.1										.3830
.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015

.5	.4938	.4940	.4941							.4952
.6										.4964
.7										.4974
.8									.4980	.4981
.9								.4985	.4986	.4986
.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

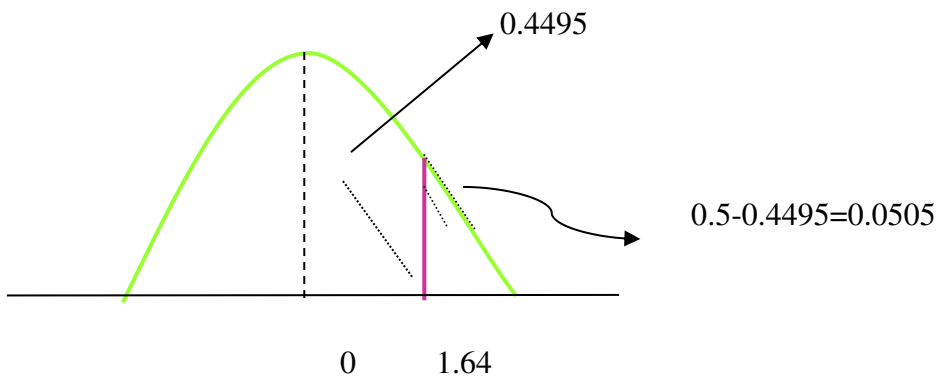
z tablosunu okurken, en soldaki sütun z'nin tam kısmını, en üst satır ,se ondalık kısmını göstermektedir. Örneğin $z=1.26$ değerine bakarken, en soldaki sütundan 1.2'yi bulup, 0.06'lık küsuratı en üst satırdan bularak, ikisinin kesiştiği yere bakacağız. Bulduğumuz bu değer, z tablosunun en başında görülen grafik uyarınca, $z=0$ ile $z=1.26$ arasındaki alandır. Yani $P(0 < z < 1.26) = 0.3962$ 'dir.

Örnek 1.2 : $P(-1.33 < z < 1.33) = ?$

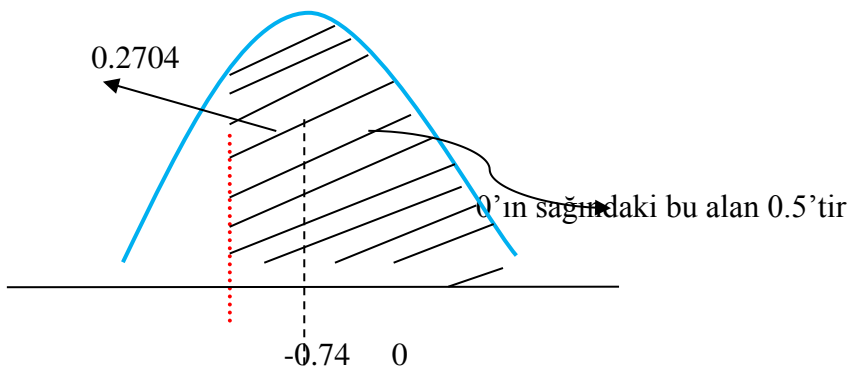
Z tablosunda en soldaki sütundan 1.3'ü buluyoruz, en üst satırdan da küsuratı yani 0.03'ü buluyoruz, ikisinin kesiştiği yerdeki değer yani $P(0 < z < 1.33) = 0.4082$ olarak buluyoruz.

Fakat bize sorulan $P(-1.33 < z < 1.33)$ olasılığıdır yani bizim bulduğumuz bu alanın iki katıdır, bu da 0.8164'tür.

Örnek 1.3: $P(z > 1.64) = ?$ Z tablosunda en soldaki sütundan 1.6'yı buluyoruz, en üst satırdan da küsuratı yani 0.04'ü buluyoruz, ikisinin kesiştiği yerdeki değer yani $P(0 < z < 1.64) = 0.4495$ 'tir. Fakat bize sorulan $P(z > 1.64)$ olasılığıdır yani bizim bulduğumuz bu alanı 0.5'ten çıkarmalıyız



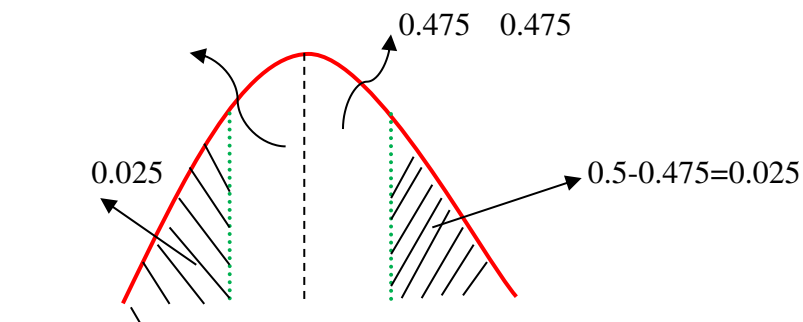
Örnek 1.4: $P(z > -0.74) = ?$



$$P(z > -0.74) = 0.2704 + 0.5 = 0.7704$$

Örnek 1.5: $P(|z| > 1.96) = ?$

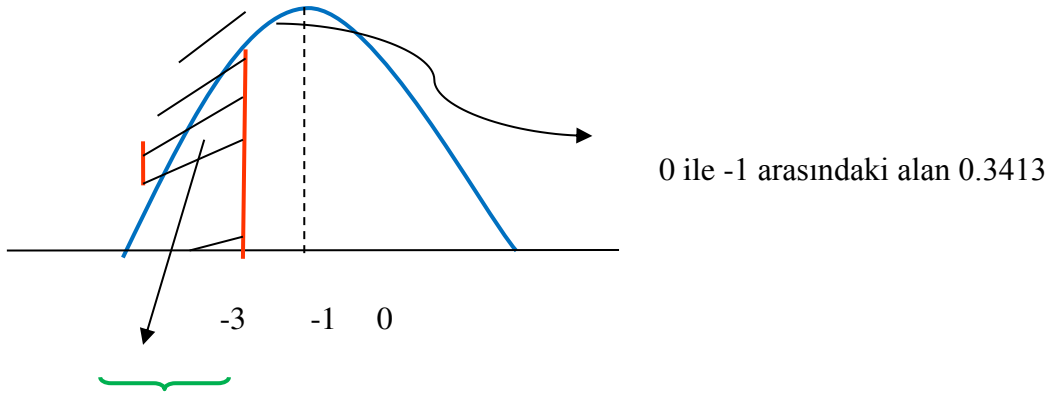
$P(z < -1.96) + P(z > 1.96)$ demektir.



$$-1.96 \quad 0 \quad 1.96$$

$$P(|z| > 1.96) = P(z < -1.96) + P(z > 1.96) = 2 \times 0.025 = 0.05 \text{ tir.}$$

Örnek 1.6: $P(-3 < z < -1) = ?$



$$P(-3 < z < -1) = 0.4987 - 0.3413 = 0.1574$$

Bu değer, 0 ile 3.0 arasındaki alandır yani $P(0 < z < 3) = 0.4987$ dir.

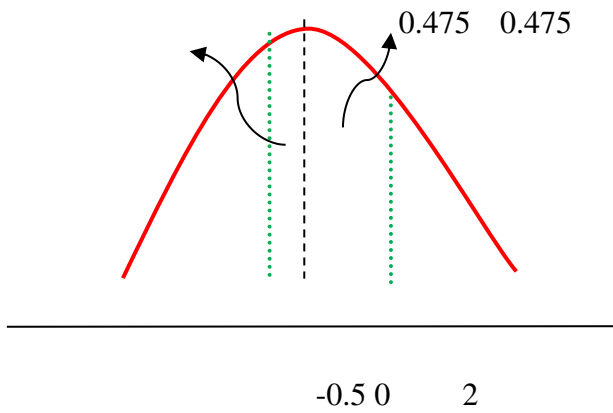
Örnek 1.7: $X \sim \text{Normal}(12, 16)$ veriliyor. $P(10 < X \leq 20) = ?$

X rd'nin ortalaması 12, varyansı 16'dır. Bu olasılık aşağıdaki integralin alınmasıyla bulunur.

$$\int_{10}^{20} \frac{1}{\sqrt{2\pi} \cdot 4} e^{-\frac{1}{2} \left(\frac{x-12}{4} \right)^2} dx$$

Fakat bu integrali almak zor olduğundan z tablosunu kullanacağız. Bunun için de öncelikle X leri z'lere dönüştüreceğiz.

$$P\left(\frac{10-12}{4} < Z < \frac{20-12}{4}\right) = P(-0.5 < Z < 2) = 0.6687$$

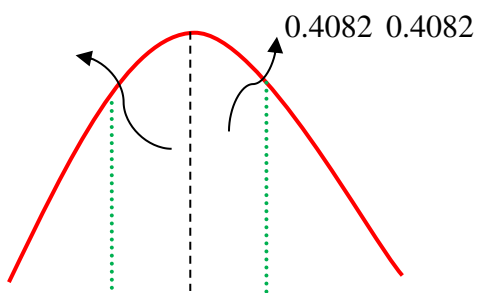


Örnek1.8: $X \sim \text{Normal}(100, 225)$ veriliyor. $P(80 < X < 120) = ?$

$$\mu = 100$$

$$\sigma = 15$$

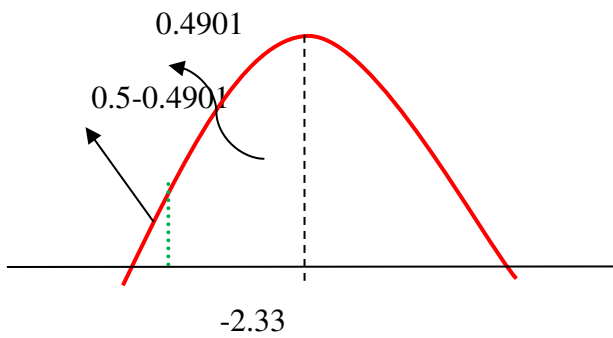
$$P\left(\frac{80-100}{15} < Z < \frac{120-100}{15}\right) = P(-1.33 < Z < 1.33) = 2 \times 0.4082 = 0.8164$$



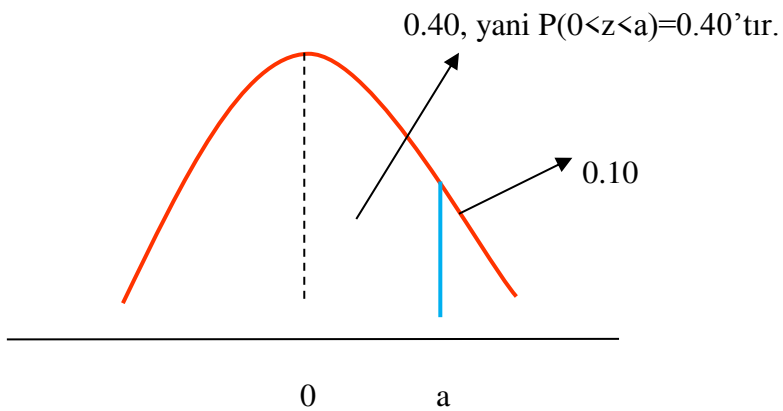
-1.33 0 1.33

Örnek 1.9: $X \sim \text{Normal}(27,9)$ veriliyor. $P(X < 20) = ?$

$$P\left(z < \frac{20 - 27}{3}\right) = P(z < -2.33) = 0.5 - 0.4901 = 0.0099$$



Örnek 1.10: $P(z > a) = 0.10$ ise $a = ?$



Z tablosuna bakarak 0.40 olasılıđına denk gelen z deđerini bulacađız. Tabloya bakıyoruz, fakat tamı tamına 0.40 deđerini bulunmamakta. Böyle durumlarda 0.40'ı kapsayan en dar aralıđı göz önüne alarak, *interpolasyon* denilen tekniđi uygulamak suretiyle, 0.40'a denk gelen z tablo deđerini kendimiz hesaplıyoruz. 0.40'ı kapsayan en dar aralık şöyledir:

1.28	0.3997
a	0.40
1.29	0.4015

İnterpolasyon şu şekilde uygulanır:

$$\frac{1.28 - a}{a - 1.29} = \frac{0.3997 - 0.40}{0.40 - 0.4015}$$

Bu denklemi çözünce, a'yı elde etmiş olacağız.

$$0.00192 - 0.0015a = 0.0003a - 0.000387 \text{ 'den, } a = 1.2816$$

Yani $P(z > 1.2816) = 0.10$ 'dur.

Örnek 1.11: $P(z > a) = 0.26$ ise $a = ?$

0.64	0.2389
a	0.24
0.65	0.2422

İnterpolasyon uyguladığımızda;

$$\frac{0.64 - a}{a - 0.65} = \frac{0.2389 - 0.24}{0.24 - 0.2422}$$

$a=0.6433$ 'tür.

$$P(z > 64.33) = 0.26$$

Uygulamalar

Uygulama Soruları

Bu Bölümde Ne Öğrendik Özeti

Bu bölümde normal dağılım ve standart normal dağılım, normal dağılım tablolarının okunması konuları öğrenildi

Bölüm Soruları

A) $X \sim \text{Normal}(50,9)$ olmak üzere;

- 1) $P(X < a) = 0.8413$ ise $a = ?$
- 2) $P(X > a) = 0.025$ ise $a = ?$
- 3) $P(X > a) = 0.95$ ise $a = ?$
- 4) $P(41 < X < a) = 0.8630$ ise $a = ?$
- 5) $P(X > a) = 0.01$ ise $a = ?$

B) $X \sim \text{Normal}(11,4)$ olmak üzere;

- 6) $P(10 < X < 12) = ?$
- 7) $P(6 < X < 10) = ?$
- 8) $P(13 < X < 16) = ?$
- 9) $P(7.8 < X < 12.6) = ?$
- 10) $P(X > 13.24) = ?$

Cevaplar:

- 1) $a = -1$
- 2) $a = 1.96$
- 3) $a = -1.645$
- 4) $a = 53.282$
- 5) $a = 2.33$
- 6) 0.383
- 7) 0.3023
- 8) 0.1525
- 9) 0.7373
- 10) 0.1315

9. GÜVEN ARALIKLARI: İKİ ANAKÜTLE

Bu Bölümde Neler Öğreneceğiz?

Bu bölümde önceki bölümde öğrendiğimiz tek anakütle parametresinin tahmini konusu, iki anakütle olması halinde parametrelerin farkları dikkate alınarak değerlendirilecektir.

Bölüm Hakkında İlgi Oluşturan Sorular

- İtalyan ve İsveçli kadınların ortalama boy uzunlukları arasında fark olup olmadığını araştırmak için 80'er kişi ile görüşülmüş ortalama boy uzunluklarının sırasıyla 172 cm ve 180 cm, boy uzunluklarını standart sapması sırasıyla 5 ve 8 cm'dir. İki ülke kadınlarının ortalama boy uzunlukları arasındaki fark anlamlı mıdır?
- A ilinde görüşülen 100 kişiden 68'i, B ilinde görüşülen 120 kişiden 54'ü seçimlerde X partisini destekleyeceğini söylemiştir. A ve B illerinde X partisinin desteklenme oranı arasında fark var mıdır?

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

İki anakütle oran ve ortalama farklarına ilişkin güven aralıkları	İki anakütle olması halinde ortalamaları (ya da oranları) dikkate alınarak aralarında istatistiksel açıdan fark olup olmadığını araştırmak.	Metinler, çözümlü problemler.
---	---	-------------------------------

Anahtar Kavramlar

- Ortalamalararası farklar
- Oranlararası farklar
- Güven aralığı

Giriş

İncelenen anakütle sayısının iki olması halinde, anakütlelerin ortalamalarının (ya da oranlarının) karşılaştırılması, aralarındaki farkın istatistiksel açıdan farklı olup olmaması araştırmacının ilgi odağı olabilir. Önceki bölümdekine benzer yaklaşımla söz konusu farklar için bir güven aralığı oluşturulur.

*Örnek istatistikleri farkı $\mp$ tablo değeri * s tan dard hata*

9.1. ORTALAMALAR ARASI FARKLAR İÇİN GÜVEN ARALIĞI

İki anakütlenin ortalamaları arasındaki farkın tahmini için güven aralığı söz konusu olduğunda güven aralığı şöyle bulunur:

$$P\left(\left(\bar{X}_1 - \bar{X}_2\right) - Z_{\alpha/2} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \left(\bar{X}_1 - \bar{X}_2\right) + Z_{\alpha/2} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = 1 - \alpha$$

Eğer anakütle varyansları bilinmiyorsa bu durumda;

$$P\left(\left(\bar{X}_1 - \bar{X}_2\right) - Z_{\alpha/2} \times \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \left(\bar{X}_1 - \bar{X}_2\right) + Z_{\alpha/2} \times \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right) = 1 - \alpha$$

Her bir anakütleyle ilişkin ortalama ve standart sapma değerleri kullanılarak yukarıdaki eşitsizlikler kullanılır. Örnek sayılarının toplamının 30'dan küçük olması halinde t tablosuna başvurulur ancak bu duruma ilişkin örnekler bu dersin konusu dışında bırakılmıştır.

ÖRNEK:

İktisat fakültesi işletme ve iktisat bölümlerinin matematik dersindeki ortalama başarı notları değerlendirilmek isteniyor. İşletme bölümünde seçilen 40 öğrencinin test sonucunda elde ettiği ortalama başarı notu 86 ve standart sapması 12, İktisat bölümünden seçilen 35 öğrencinin ortalama başarı notu 72 ve standart sapması 14'tür. İşletme ve iktisat bölümlerinde okutulan matematik dersinin ortalama başarı notları arasındaki farkın güven aralığını %99 olasılıkla belirleyiniz.

$$\begin{array}{lll} \bar{X}_1 = 86 & S_1 = 12 & n_1 = 40 \\ \bar{X}_2 = 72 & S_2 = 14 & n_2 = 35 \end{array}$$

$$P\left((86 - 72) - 2.58 \times \sqrt{\frac{12^2}{40} + \frac{14^2}{35}} \leq \mu_1 - \mu_2 \leq (86 - 72) + 2.58 \times \sqrt{\frac{12^2}{40} + \frac{14^2}{35}}\right) = 0.99$$

$$P(6.18 \leq \mu_1 - \mu_2 \leq 21.82) = 0.99$$

İki anakütle farkını %99 güvenle gösteren bu eşitsizlikte, 6.18- 21.82 aralığı 0'ı kapsamadığı için iki bölüm arasında başarı notları açısından fark olduğu söylenebilir.

ÖRNEK:

A ülkesinde yaşayan kadınlarla B ülkesinde yaşayan kadınların boyları karşılaştırılmak isteniliyor. A ülkesinde yaşayanların kadınlardan rastgele 120 kişi seçiliyor, ortalama boy uzunlukları 172 cm ve standart sapmaları 20 cm olarak hesaplanıyor. B ülkesinde yaşayan 100 kişi ile görüşülüyor ortalama boy uzunluğu 180 cm ve standart sapma da 22 cm olarak belirleniyor. İki ülkede yaşayanların boy uzunluğunu %95 güvenle değerlendiriniz.

$$P\left(\left(\bar{X}_1 - \bar{X}_2\right) - Z_{\alpha/2} \times \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \left(\bar{X}_1 - \bar{X}_2\right) + Z_{\alpha/2} \times \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right) = 1 - \alpha$$

$$(180 - 172) - 1.96 \times \sqrt{\frac{20^2}{120} + \frac{22^2}{100}} \leq \mu_1 - \mu_2 \leq (180 - 172) + 1.96 \times \sqrt{\frac{20^2}{120} + \frac{22^2}{100}}$$

$$8 \pm 5.62$$

$$2.40 \leq \mu_1 - \mu_2 \leq 13.62$$

Aralık sıfırı içermediği için iki ülkede yaşayan kadınların ortalama boy uzunlukları arasında anlamlı bir fark olduğu söylenebilir.

9.2. İKİ ORAN FARKI İÇİN GÜVEN ARALIĞI

Örnek oranları $\hat{p}_1$ ve $\hat{p}_2$ gösterilirse örnek oranları arasındaki farktan hareketle anakütle oranları arasındaki farkın güven sınırları aşağıdaki gibi olur.

$$\left(\hat{p}_1 - \hat{p}_2\right) - Z_{\alpha/2} \times \sqrt{\frac{\hat{p}_1 \cdot \hat{q}_1}{n_1} + \frac{\hat{p}_2 \cdot \hat{q}_2}{n_2}} \leq P_1 - P_2 \leq \left(\hat{p}_1 - \hat{p}_2\right) + Z_{\alpha/2} \times \sqrt{\frac{\hat{p}_1 \cdot \hat{q}_1}{n_1} + \frac{\hat{p}_2 \cdot \hat{q}_2}{n_2}}$$

Örnek:

Bir seçim öncesinde A partisine B ve C illerinde verilecek oy oranları değerlendirilmek istenilmektedir. B ve C illerinden 1000'er kişi ile görüşülmüş sırasıyla A partisine oy vereceklerin sayısının 825 ve 760 olduğu saptanmıştır. Buna göre oy oranlarının farkının %95'lik güven aralığını bulunuz.

$$p_1 = \frac{825}{1000} = 0.825 \quad p_2 = \frac{760}{1000} = 0.760$$

$$S_{\text{standart sapma}} = \sqrt{\frac{p_1 \cdot q_1}{n_1} + \frac{p_2 \cdot q_2}{n_2}} = \sqrt{\frac{0.825 \cdot (1 - 0.825)}{1000} + \frac{0.760 \cdot (1 - 0.760)}{1000}} = 0.018$$

$$(0.82 - 0.760) - 1.96 \times 0.018 \leq P_1 - P_2 \leq (0.82 - 0.760) + 1.96 \times 0.018$$

$$0.029 \leq P_1 - P_2 \leq 0.10$$

%95 güvenle iki şehir arasındaki oy oranlarının farkı yukarıdaki gibidir. Oy oranları sıfırı kapsamadığı için istatistiksel açıdan farklıdır.

ÖRNEK:

Bir fabrikada üretilen malların kusurlu oranı araştırılmaktadır. 200 mamülden 35 tanesi kusurludur. **90%** güven aralığında ana kütledeki kusurlu oranını nedir?

$$\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

$$0.175 - 1.645 \cdot \sqrt{\frac{0.175 \cdot (1 - 0.175)}{200}} \leq p \leq 0.175 + 1.645 \cdot \sqrt{\frac{0.175 \cdot (1 - 0.175)}{200}}$$

$$0.1308 \leq p \leq 0.2192$$

ÖRNEK:

ABC lisesinin 400 mezunundan 32 'si üniversiteye gitmektedir. Üniversiteye gidenler için 95% güven aralığında tahminde bulununuz

$$\hat{p} - Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \leq p \leq \hat{p} + Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

$$0.08 - 1.96 \cdot \sqrt{\frac{0.08 \cdot (1 - 0.08)}{400}} \leq p \leq 0.08 + 1.96 \cdot \sqrt{\frac{0.08 \cdot (1 - 0.08)}{400}}$$

$$0.053 \leq p \leq 0.107$$

ÖRNEK:

Bir fabrikada üretilen 100 ürünün ortalama ağırlığı 1040 gr standart sapması 25 gr bulunmuştur. Bu imalat prosesinde üretilen ürünlerin ortalama ağırlığı %99 güvenle hangi aralıktadır?

$$P\left(\bar{X} - z_{\alpha/2} \frac{\sigma_x}{\sqrt{n}} \leq \mu_x \leq \bar{X} + z_{\alpha/2} \frac{\sigma_x}{\sqrt{n}}\right) = 1 - \alpha$$

$$1040 - 2.58 \frac{25}{\sqrt{100}} \leq \mu \leq 1040 + 2.58 \frac{25}{\sqrt{100}}$$

ÖRNEK:

Ampül üreten bir firma ampüllerin dayanıklılık süresini değerlendirmek istiyor. 100 tane rassal olarak seçilen ampülün standart sapması 140 saat, ortalama dayanıklılık süreleri de 1280 saat bulunuyor. %90 güvenle anakütle ortalama dayanıklılık süresi nedir?

$$P\left(\bar{X} - Z_{\alpha/2} \frac{S_x}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\alpha/2} \frac{S_x}{\sqrt{n}}\right) = 1 - \alpha$$

$$1280 - 1.65 \times \frac{140}{\sqrt{100}} \leq \mu \leq 1280 + 1.65 \times \frac{140}{\sqrt{100}}$$

Firmanın ürettiği ampullerin ortalama ömrü, 0.95 olasılıkla yaklaşık 1257 ile 1303 saat arasındadır.

ÖRNEK:

A ilacını kullanan 60 hasta ortalama 8 günde iyileştirmektedir, B ilacını kullanan 50 hastanın iyileştirme süresi ise 7gündür. Her iki ilacın iyileştirme sürelerine ilişkin hesaplanan sapma değerleri sırasıyla 0.2 ve 0.3 gündür. İki ilacın tedavi süresi arasında istatistiksel açıdan önemli bir fark var mıdır?

(%94 güven düzeyi)

$$(8 - 7) - 1.88\sqrt{\frac{0.2^2}{60} + \frac{0.3^2}{50}} \leq \mu_1 - \mu_2 \leq (8 - 7) + 1.88\sqrt{\frac{0.2^2}{60} + \frac{0.3^2}{50}}$$

$$0.906 \leq \mu_1 - \mu_2 \leq 1.093$$

Aralık sıfır değerini kapsamamaktadır, iki ilacın tedavi süreleri açısından farklı etkileri vardır.

ÖRNEK:

Bir toplulukta günde 4 bardaktan fazla demli çay içen erkek ve kadınların oran farkları için güven aralığı oluşturulacaktır. Görüşülen 100 kadından 45'i ve 100 erkekten 80'i günde 4 bardaktan fazla demli çay içtiğini söylemiştir. Kadın ve erkeklerin oranı arasında istatistiksel açıdan önemli bir fark var mıdır?

$$(0.80 - 0.45) - 1.96\sqrt{\frac{0.45 * 0.55}{100} + \frac{0.85 * 0.15}{100}} \leq p_1 - p_2 \leq (0.80 - 0.45) + 1.96\sqrt{\frac{0.45 * 0.55}{100} + \frac{0.85 * 0.15}{100}}$$

$$0.2632 \leq p_1 - p_2 \leq 0.4363$$

Aralık sıfırı kapsamamaktadır, erkeklerin oranı ile kadınların oranı arasında istatistiksel açıdan fark vardır.

ÖRNEK:

İki farklı cinsteki atların katıldıkları yarışta parkuru 3 dakikanın altında tamamlayanların oran farklı değerlendirilmek istenmektedir. 130 Arap atından 105'i, 50 İngiliz atından 35'i yarış 3 dakikanın altında bir sürede tamamlamıştır. Kitle oran farkı için %70 güven düzeyinde aralığı oluşturarak, iki cins at arasında fark var mıdır değerlendiriniz.

$$\hat{p}_1 = \frac{35}{50} = 0.7$$

$$\hat{p}_2 = \frac{105}{130} = 0.81$$

$$(0.81 - 0.7) - 1.04 \sqrt{\frac{0.7 * 0.3}{50} + \frac{0.81 * 0.19}{130}} \leq p_1 - p_2 \leq (0.81 - 0.7) + 1.04 \sqrt{\frac{0.7 * 0.3}{50} + \frac{0.81 * 0.19}{130}}$$

$$0.1009 \leq p_1 - p_2 \leq 0.1103$$

Aralık sıfırı kapsamamaktadır, Arap atlarının daha iyi olduğu sonucuna varılır.

ÖRNEK:

Bir kurumda çalışan 120 kadından 105'i , 130 erkekte 100'ü greve katılmıştır. Anakütle oranları farkı için %92 güven düzeyini oluşturunuz. Greve katılan kadın ve erkeklerin oranları arasında önemli bir fark var mıdır?

$$\hat{p}_1 = \frac{105}{120} = 0.87 \quad \hat{p}_2 = \frac{100}{130} = 0.76$$

$$(0.87 - 0.76) - 1.76 \sqrt{\frac{0.87 * 0.13}{120} + \frac{0.76 * 0.24}{130}} \leq p_1 - p_2 \leq (0.87 - 0.76) + 1.76 \sqrt{\frac{0.87 * 0.13}{120} + \frac{0.76 * 0.24}{130}}$$

$$0.0249 \leq p_1 - p_2 \leq 0.1951$$

ÖRNEK:

Yapılan bir araştırmada Türkiye'de düzenli kitap okuma oranının kadın ve erkek okuyucular arasında önemli bir fark taşıyıp taşımadığı test edilmek istenmiştir. Bu araştırmaya göre 200 kadından 150'si, 200 erkekte 60'ı düzenli kitap okumaktadır. Kitle oran farkları için güven sınırlarını %86 olasılıkla belirleyiniz.

$$\hat{p}_1 = \frac{150}{200} = 0.75$$

$$\hat{p}_2 = \frac{60}{200} = 0.3$$

$$(0.75 - 0.3) - 1.48 \sqrt{\frac{0.75 * 0.25}{200} + \frac{0.3 * 0.7}{200}} \leq p_1 - p_2 \leq (0.75 - 0.3) + 1.48 \sqrt{\frac{0.75 * 0.25}{200} + \frac{0.3 * 0.7}{200}}$$

$$0.3841 \leq p_1 - p_2 \leq 0.4345$$

9.3 .Örneklem Büyüklüğünün Belirlenmesi ve Hata

Anakütle parametresine ilişkin güven aralıklarının belirlenmesinde şimdiye kadar kullandığımız eşitsizliklerde de dikkatimizi çekeceği üzere örnek büyüklüğünün önemi vardır. Örneklemen çok küçük olması sonuçların kitle için sağlıklı genelebilmesini güçleştirir.

$\bar{X} \mp e$ aralık tahminini göstermek üzere, “e” hata olarak isimlendirilir.

$e = \text{tablo değeri} * s \text{ tan dart hata}$

$$e = Z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$$

$$n = (Z_{\alpha/2} * \frac{\sigma}{e})^2$$

Şeklinde olacaktır. Eğer oranlara üzerinden hesaplanacaksa yine benzeri yaklaşımla minimum örnek sayısına şöyle ulaşılır:

$$e = Z_{\alpha/2} * \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$n = (Z_{\alpha/2}^2 * \frac{\hat{p}\hat{q}}{e^2})$$

ÖRNEK

Bir topluluktan rastgele seçilen 100 kişiden 58’i alkol kullanmadığını söylemiştir. Bu toplulukta alkol kullanmayanların oranı için %88 güven aralığını oluşturunuz. Elde edilen güven aralığı ne kadar hata içermektedir? Bu hatayı aynı güven düzeyinde 0.08 olarak belirlenseydi kaç birimlik örneğe ihtiyaç duyulurdu?

$$\hat{p} = \frac{58}{100} = 0.58 \quad \hat{q} = 0.42$$

$$0.58 - 1.56\sqrt{\frac{0.58 \times 0.42}{100}} \leq p \leq 0.58 + 1.56\sqrt{\frac{0.58 \times 0.42}{100}}$$

$$0.58 - 0.07644 \leq p \leq 0.58 + 0.07644$$

□ *hata* □ *hata*

$$0.50356 \leq p \leq 0.6564$$

hata = 0.08 *olursa*

$$n = z_{\alpha/2}^2 \frac{pq}{hata^2} = 1.56^2 \frac{0.58 \times 0.42}{0.08^2} = 92.6 \approx 93$$

Örnek:

A ve B marka şampuanlarının içerdikleri ortalama paraben miktarları üzerine bir çalışma yapılıyor. A marka şampuandan 50 tane, B'den 40 tane ürün seçiliyor. A şampuanlarının ortalama paraben miktarı 0.5ml sapması 0.2 ve B marka şampuanların ortalama paraben miktarı 2.4ml ve sapması 1 olarak hesaplanıyor. Ortalama paraben miktarları açısından istatistiksel açıdan önemli bir fark var mıdır? (Güven aralığını %95 olasılıkla belirleyiniz)

$$(0.5 - 2.4) - 1.96\sqrt{\frac{0.2^2}{50} + \frac{1}{40}} \leq \mu_1 - \mu_2 \leq (0.5 - 2.4) + 1.96\sqrt{\frac{0.2^2}{50} + \frac{1}{40}}$$

$$-2.078 \leq \mu_1 - \mu_2 \leq -1.677$$

Aralık 0 değerini kapsamadığından farkın istatistiksel açıdan anlamlı olduğu söylenebilir.

Uygulamalar

Uygulama Soruları

Bu Bölümde Ne Öğrendik Özeti

Bu bölümde iki anakütle ortalamalarının/oranlarının farkları dikkate alınarak güven aralıkları hesaplandı ve iki anakütle parametreleri arasındaki farkın anlamlı olup olmadığı değerlendirildi.

Bölüm Soruları

Örnek büyüklüğü ve güven düzeyinin artmasının aralık tahmini üzerindeki etkisini ayrı ayrı değerlendiriniz.

2- Bir ilde yaşayan 200 üniversite öğrencisinden 142'si sanal dünyada dijital varlıklarının olduğunu söylemiştir. **a)**Buna göre bu ilde dijital varlığı olan üniversite öğrencilerinin oranını %94 güvenle tahmin ediniz. **b)** Tahmin sürecinde yapılan hata büyüklüğü nedir? Hatanın 0.04 olması için kaç örnekle çalışılmak gerekir?

3- Bir ilde yaşayan dört kişilik ailelerin günlük mutfak harcamalarını belirlemek için rastgele 4 kişi seçiliyor. 4 kişinin günlük harcamaları şöyledir: 50-20-30-20 tl. Bu ilde yaşayan 4 kişilik ailelerin ortalama mutfak harcamasını %95 güven düzeyinde tahmin ediniz.

4- A ve B bölümlerinde okuyan öğrencilerin düzenli spor yapma alışkanlıklarını değerlendirmek için sırasıyla 138 ve 124 kişi ile görüşülmüştür. A bölümünde 42 ve B bölümünde 36 öğrenci düzenli spor yaptığını söylemiştir. İki bölüm arasında spor yapma oranları arasındaki fark istatistiksel olarak anlamlı mıdır?

5-Bir bölümde okutulan X dersini 2012 yılında 124 kişi seçmiş ve ortalama başarıları 78, notlarının varyansı 144 olarak hesaplanmıştır. 2014 yılında dersi seçen öğrencilerin sayısı 88 ve bu dersin ortalama başarıları 88, notlarının varyansı 121 olarak hesaplanmıştır. Buna göre söz konusu 2 yıldaki ortalama başarı arasında bir fark var mıdır?

10. HİPOTEZ TESTLERİ

Bu Bölümde Neler Öğreneceğiz?

- 10.1.** Bu bölümde istatistiksel çıkarımanın ikinci ayağını oluşturan hipotez testleri konusu ele alacağız. Hipotez testleri ile anakütle parametresinin belli bir değere eşit olup olmadığını test edeceğiz.

Bölüm Hakkında İlgi Oluşturan Sorular

- Seçimlerde A partisinin oy oranının %47 olacağı iddia ediliyor. Görüşülen 180 kişiden 85'i A partisini destekleyeceğini söylediğine göre bu iddia desteklenebilir mi?
- A marka ampulün ortalama ömrünün 182 saatten fazla olduğu iddia ediliyor. Rastgele seçilen 28 ampulün ortalama ömrü 186 saat, standart sapması 22 saat olarak bulunduğu göre bu iddia geçerli midir

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Hipotez testleri	Anakütle parametresinin iddia edilen bir büyüklüğe eşit olup olmadığının sınanmasını öğrenmek.	Metinler, çözümlü problemler.

Anahtar Kavramlar

- Z testi
- T testi
- Hata düzeyi
- Anlamlılık düzeyi
- Ret bölgesi
- Kabul alanı
- Alternatif hipotez, sıfır hipotezi

Giriş

Bu bölümde istatistiksel çıkarsamanın diğer bileşeni olan hipotez testleri ele alınacaktır.

10.1. HİPOTEZ TESTLERİ

İstatistiksel çıkarımanın iki önemli adımı vardır bunlar tahmin ve hipotez testidir.

Önceki iki bölümde tahmin süreci ele alınmıştır. Bu bölümde ise hipotez testlerine değineceğiz. Örneklem değerleri kullanılarak hesaplanan istatistiğin değeriyle bu istatistiğin temsil ettiği anakütle parametre değerinin arasındaki farklılık olup olmadığı eğer farklılık varsa bunun istatistiksel olarak önemli olup olmadığı hipotez testi ile sınıanır.

İstatistiksel hipotez testi, bir araştırmada ilgilenilen bir ya da daha fazla parametre değeri hakkında ileri sürülen iddianın geçerliliğini bu parametreyi temsil edilen istatistik ve bu istatistiğin örneklem dağılımıyla ilgili bilgilerden yararlanarak araştıran önermedir.

Hipotez testi örneklemden yola çıkılarak anakütle parametresi hakkında ortaya atılan bir iddianın geçerliliğinin sınıanmasıdır.

Örneğin;

-İstatistik dersi final notlarının ortalaması 50'nin altındadır.

-A marka ampüllerin ortalama dayanıklılığı B marka ampüllerin ortalama dayanıklılığından daha fazladır.

Hipotez testinde hipotezin kabul edilip edilmemesi için birden fazla hipoteze ihtiyaç duyulur.

10.2. Hipotezlerin İfade Edilmesi

İstatistiksel hipotezlerin testinde, iki hipotez söz konusudur. Bunlar, “sıfır hipotezi” ve “alternatif hipotez” şeklinde isimlendirilir. Sıfır hipotezi H_0 simgesiyle gösterilir. Sıfır hipotezinde test süreci tamamlanana dek örneklem istatistiğinin değeriyle anakütle parametresi arasındaki farkın önemli olmadığı, istatistiksel olarak sıfır olduğu vurgulanır.

Sıfır hipotezinin test edilebilmesi için, bu hipotezden farklı bir hipotezin de dikkate alınması gerekir. Alternatif, H_1 , karşıt hipotez, H_0 hipotezinin belirli bir olasılıkla reddedilmesi durumunda kabul edilir ve genellikle de araştırma hipotezinin incelendiği hipotezdir. Alternatif hipotez üç farklı biçimde ifade edilir.

θ anakütle parametresi

$\hat{\theta}$ örneklem istatistiğini göstermek üzere:

$$H_0 : \theta = 0$$

$$H_1 : \theta \neq 0$$

$$H_1 : \theta < 0$$

$$H_1 : \theta > 0$$

Alternatif hipotezin yönü karar verme süreci için oldukça önemlidir.

Yukarıdaki ilk alternatif hipotez verilecek kararın anakütle parametre değerinden hem küçük hem de büyük yöndeki anlamlı farklılıklarına karşılık gelir.

İkinci alternatif hipotez anakütle parametre değerinde küçük yönde anlamlı farklılığa karşılık gelir.

Üçüncü alternatif hipotezse anakütle parametre değerinden büyük yönde anlamlı farklılığa karşılık gelir.

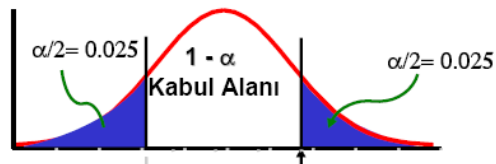
Daha önce de söylendiği gibi alternatif hipotezin yönü karar verme sürecinde önemlidir, sıfır hipotezinin reddedileceği bölgenin yönünü belirler.

Hipotez Çiftleri

Çift yönlü hipotez

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

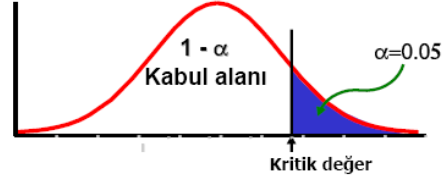


Tek yönlü hipotez

(Sağ taraf testi)

$$H_0 : \mu \leq \mu_0$$

$$H_1 : \mu > \mu_0$$

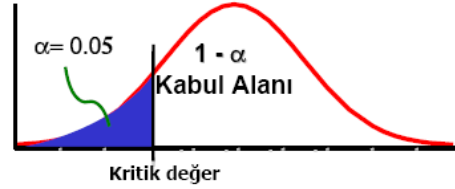


Tek taraflı hipotez

(Sol taraf testi)

$$H_0 : \mu \geq \mu_0$$

$$H_1 : \mu < \mu_0$$



Hipotezin çift yönlü olması halinde anlamlılık düzeyi iki eşit parçaya ayrılır sağ ve sol tarafta. Hipotezin tek taraflı olması halinde ise red alanları sağ veya sol tarafta yer almaktadır.

Hipotez test sürecinde en önemli adım test istatistiğinin belirlenmesidir. Eğer hesaplanan test istatistiğinin değeri belirlenen anlamlılık düzeyi/red alanı içerisinde ise sıfır hipotezi reddedilir. Eğer hesaplanan test istatistiği red alanı içinde değilse sıfır hipotezi kabul edilir.

Hipotez testinin adımları şu şekilde sıralanabilir:

1. H_0 'ı belirle.
2. H_1 'i belirle.
3. Test istatistiğini seç

4. Test istatistiğini hesapla.

5. Karar ver

10.3. I. Tür ve II. Tür Hata

Aynı anakütleden çekilen örneklemeler aynı hacimli olsalar dahi örnek istatistikleri farklı değerler alabilir. Anakütle parametresi hakkında kurulan hipotezler test edilirken örneklem istatistiğinin kullanılması halinde iki tür hata söz konusudur.

	H ₀ Gerçek	H ₀ Hatalı
H ₀ kabul	Doğru karar	II .tür hata β
H ₀ Ret	I. tür hata α	Doğru karar

Sıfır hipotezi gerçekte doğruyken yanlışlıkla hipotezin reddedilmesi durumunda I. Tip hata ile karşılaşılır. Eğer sıfır hipotezi gerçekte yanlış iken, kabul edilmişse işlenen hata II. Tip hatadır.

Hipotez testinde amaç sıfır hipotezini ya kabul ya da reddetmektir. Alternatif hipotezin kabul edilmesi veya reddedilmesi sonucunda ortaya çıkan karardır. Bu açıdan aynı hipotez testinde I. Tip ya da ikinci tip hatadan yalnızca biri söz konusudur, iki tür hatayı işlemek mümkün değildir.

Çalışmalarda hem I.tür hem de II. Tür hata riskinin küçük olması tercih edilir. α, β, n (

örneklem büyüklüğü) arasında bir ilişki vardır. I. Tür hata arttıkça ikinci tür hata olasılığı azalır ve örneklem büyüklüğü arttıkça I. Ve II. Tür hatalar küçülür.

Anlamlılık Düzeyi

I tip hata yapma düzeyine anlamlılık düzeyi denilir. Hipotez test edilirken anlamlılık düzeyi öne çıkar. Uygulamalarda genellikle anlamlılık düzeyi $\alpha = 0.01$ $\alpha = 0.05$ olarak seçilir. Yani kabul edilmesi gereken sıfır hipotezinin reddilmesi olasılıkları sırasıyla %1 ve %5,

testin güven düzeyi ise yine sırasıyla % 99 ve % 95'tir. Anlamlılık düzeyinin belirlenmesi ile reddedilecek alanın büyüklüğü de belirlenmiş olur.

Hipotez testinde test istatistiği belirlenirken örneklem büyüklüğü, test edilecek ana kütle parametresi, varyans değerinin bilinip bilinmemesi belirleyici olur.

-Ortalama ya da iki anakütle ortalama farkları için,

σ^2 biliniyorsa Z istatistiği

σ^2 bilinmiyorsa $n < 30$ ise t istatistiği

σ^2 bilinmiyorsa $n > 30$ ise Z istatistiği

-Oranlar ya da oranların farkları için Z istatistiği

10.4. Anakütle ortalamasına ilişkin Hipotez Testleri

Ortalamasının belli bir değere μ eşit olduğu iddia edilen bir anakütleden seçilen n hacimli bir örneklem aritmetik ortalamasının $\bar{x}$ ve sapmasının s olduğunu varsayalım. μ ve $\bar{x}$ arasındaki fark istatistiksel olarak önemli midir? Yoksa örnekleme hatası mıdır?

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \quad \sigma^2 \text{ biliniyorsa Z istatistiği}$$

$$z = \frac{\bar{x} - \mu}{s / \sqrt{n}} \quad \sigma^2 \text{ bilinmiyorsa } n > 30 \text{ ise Z istatistiği}$$

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} \quad \sigma^2 \text{ bilinmiyorsa } n < 30 \text{ ise t istatistiği}$$

Örnek:

Bir fabrikada üretilen margarin paketlerinin ortalama ağırlığı 100 gramdır, ve standart sapması 2 olan normal dağılım göstermektedir. Rasgele olarak seçilen 9 paketin ortalama

ağırlığı 102 gram olarak ölçülmüştür. Paketlerin ağırlığının 100 gramdan farklı olduğu söylenebilir mi? ($\alpha=0.05$)

Öncelikle hipotezler kurulur

$$H_0 : \mu = 100gr$$

$$H_1 : \mu \neq 100gr$$

$$\mu = 100$$

$$\bar{x} = 102$$

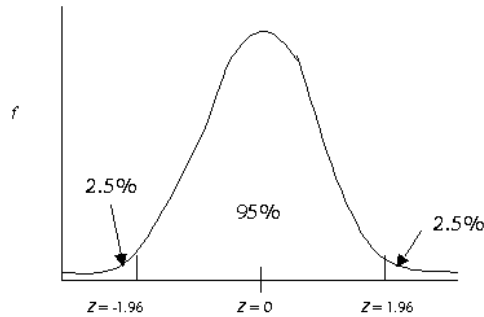
$$\sigma = 2$$

$$n = 9$$

$$Z_{hesap} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{102 - 100}{2 / \sqrt{9}} = 3$$

Hipotez çift taraflı olduğunda standart normal eğrinin hem sağ hem de sol tarafı kritik alan olarak taranır. Anlamlılık düzeyi %5 olarak verilmişti bu durumda sağ ve sol tarafta %2.5 ‘lik alan kritik/red alanı olarak belirlenecektir. Standart normal eğri tablosuna bakılarak, bu alana karşılık gelen değerler bulunur.

Toplam alan 1’dir. Sağ ve sol tarafta kalan alanların 0.5 olarak düşünülür. Tablo değeri



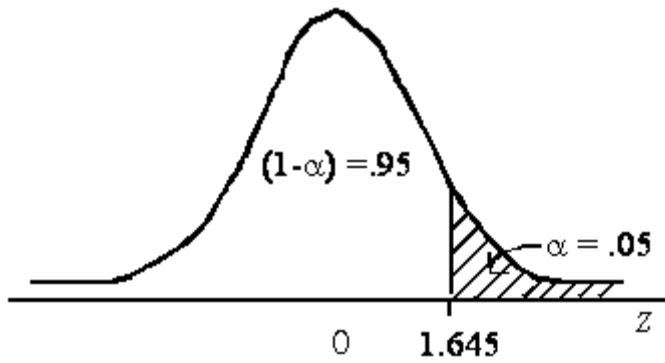
incelenirken $0.5 - 0.025 = 0.475$ değerine karşılık gelen değer alınır. Bu değer 1.96’dır.

Hesablanan Z değeri kritik değerden büyüktür. $Z_{\text{hes}} > Z_{\text{tablo}}$ ($3 > 1.96$) olduğundan H_0 hipotezi reddedilir. Paketlerin ağırlıklarının 100 gramdan farklı olduğu söylenebilir.

ÖRNEK: A marka zeytinyağı tenekelerinin ağırlığının 3680 gr'dan fazla olduğu iddia edilmektedir. Teneke ağırlıkları 150 gr sapma ile normal dağılıma uymaktadır. Rastgele seçilen 25 tenekenin ortalama ağırlığı 3725 gram olarak hesaplanmıştır. İddiayı $\alpha=0.05$ önem düzeyinde test ediniz.

$$H_0 : \mu \leq 3680$$

$$H_1 : \mu > 3680$$



Alternatif hipotezin tek taraflı olmasından dolayı tabloda kritik değere bakılırken,

$0.5 - 0.05 = 0.45$ olacaktır.

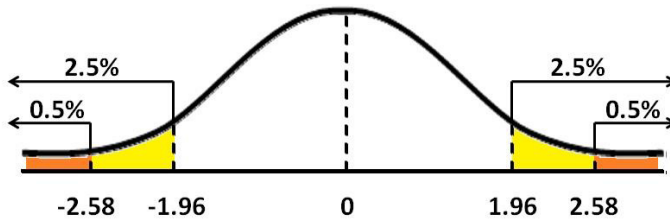
$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{3725 - 3680}{150 / \sqrt{25}} = 1.5$$

Standart Z tablosunda bulunan kritik deęer 1.645'tir. Hesaplanan Z deęeri 1.5 bu deęerin solunda yani kabul alanı içindedir. H_0 hipotezi kabul edilir yani zeytinyaęı tenekelerinin aęırlıęı 3680 gramdan fazladır iddiasını kabul için yeterli örnek yoktur.

ÖRNEK: Matematik bölümü öğrencilerinin İstatistik dersinden aldıkları notların ortalamasının 70 olduęu iddia edilmektedir. Bu amaçla seçilen 49 öğrencinin aldıkları notların ortalaması 64 ve standart sapması 14 olarak bulunmuştur. İddiayı $\alpha = 0.01$ için test ediniz.

$$H_0 : \mu = 70$$

$$H_1 : \mu \neq 70$$



Çift taraflı bir hipotez testi, bu yüzden sağ ve sol tarafta iki alan kritik olarak belirlenir.

Anlamlılık düzeyi 0.01 olduğundan, $0.5 - 0.01/2 = 0.495$ olarak bulunur, tabloda karşılık gelen deęer 2.58'dir.

$$Z = \frac{\bar{x} - \mu}{s / \sqrt{n}} = \frac{64 - 70}{14 / 7} = \frac{-6}{2} = -3$$

Hesaplanan deęer sol tarafta red alanının içindedir. $-3 < -2.58$. H_0 hipotezi reddedilir.

ÖRNEK:

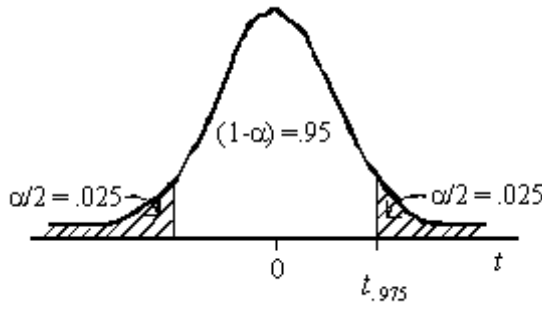
Bir marketteki margarin paketlerinin ağırlığı üzerinde brut 250 gr. yazmaktadır. Rastgele olarak seçin 14 paketin ortalama ağırlığı 235 gr ve standart sapması 14 gr olarak bulunmuştur. Paketlerin ağırlığının 250 gr olmadığı söylenebilir mi? (0.05 önem düzeyi)

$$H_0 : \mu = 250$$

$$H_1 : \mu \neq 250$$

Örnek sayısının 14 olması ve anakütle varyansının bilinmemesinden ötürü t istatistiği hesaplanmalıdır.

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} = \frac{235 - 250}{14 / \sqrt{14}} = -4$$



(n-1), 13 serbestlik dereceli t tablo değeri -2.16'dir. -4 kritik alanda olduğu için iddia reddilir paketlerin ağırlığı 250 gramdan farklıdır.

5.5 ORANLAR için hipotez testi

Çift taraflı test $H_0 : P = P_0$
 $H_1 : P \neq P_0$

$H_0 : P \geq P_0$
Tek taraflı testler $H_1 : P < P_0$

$$H_0 : P \leq P_0$$

$$H_1 : P > P_0$$

Test istatistiđi

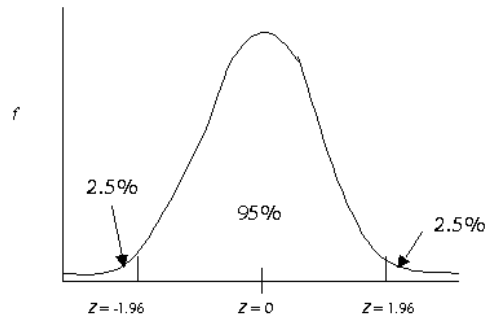
$$Z = \frac{p - P}{\sigma_p} = \frac{p - P}{\sqrt{\frac{P(1-P)}{n}}}$$

ÖRNEK: Bir dersane öğrencilerinin yabancı dil sınavındaki başarı oranlarının %95 olduğunu iddia etmektedir. Rastgele seçilen 200 öğrenciden 180 ‘i sınavdan başarılı olmuştur. İddiayı 0.05 anlam düzeyinde değerlendiriniz.

$$H_0 : P = P_0$$

$$H_1 : P \neq P_0$$

$$p = 184 / 200 = 0.92$$



$$Z = \frac{p - P}{\sigma_p} = \frac{p - P}{\sqrt{\frac{P(1-P)}{n}}} = \frac{0.8 - 0.9}{\sqrt{\frac{0.9(1-0.9)}{200}}} = -1.76$$

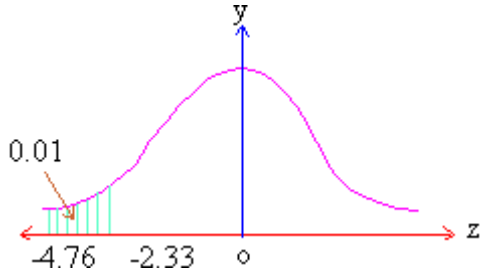
Bulunan test istatistiđi tablo deđerine oldukça yakındır kabul alanı iđerisinde görünmektedir. Eđer anlam düzeyi 0.01 olsaydı bu durumda hipotez yine kabul alanı iđerisinde olacaktı.

ÖRNEK:

A partisinin seçimlerde alacağı oy oranının %90 olacağı iddia edilmektedir. Rastgele seçilen 200 kişiye hangi partiye oy verecekleri sorulmuş, bu kişilerden 180'nin A partisine oy vereceği bilgisi elde edilmiştir. Buna göre iddiayı %1 anlamlılık düzeyinde test ediniz.

$$H_0 : P = P_0 = 0.90$$

$$H_1 : P < P_0 = 0.90$$



$$p = \frac{160}{200} = 0.8$$

$$Z = \frac{p - P}{\sigma_p} = \frac{p - P}{\sqrt{\frac{P(1-P)}{n}}} = \frac{0.8 - 0.9}{\sqrt{\frac{0.9(1-0.9)}{200}}} = -4.76$$

Tablo değeri -2.33'tür ve hesaplanan değer -4.76 'dır. H0 hipotezi reddedilir.

ÖRNEK

Bir fabrikada zeytinler 500 gr. 'lık paketler haline getirilmektedir. Paketleme işleminin kontrolü için 100 paket rastgele seçilmiş ve bu paketlerin ortalama ağırlığı 495gr. standart sapması da 20 gr. olarak hesaplanmıştır. % 5 anlamlılık düzeyinde paketlerin ağırlığının 500 gr. olduğu söylenebilir mi?

$$H_0 : \mu = 500gr$$

$$H_1 : \mu \neq 500gr$$

$$z = \frac{495 - 500}{20 / \sqrt{100}} = -2.5$$

H_0 reddedilir

ÖRNEK

Bir şirkette çalışanlara verilen günlük sigara molasının 20 dakika olduğu bilinmektedir. Rastgele seçilen 5 kişinin sigara molaları izlenmiştir. Bu değerler; 15-25-10-12-30 şeklindedir. Buna göre çalışanların kullandığı sigara molası 20 dakikadır iddiasını %5 anlamlılık düzeyinde değerlendiriniz.

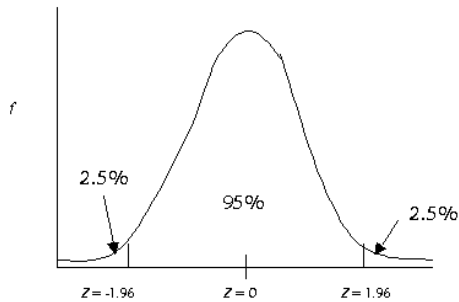
$$H_0 : \mu = 20$$

$$H_1 : \mu \neq 20$$

$$\bar{X} = \frac{15 + 25 + 10 + 12 + 30}{5} = 18.4$$

$$s = \sqrt{\frac{(15 - 18.4)^2 + \dots + (30 - 18.4)^2}{4}} = 8.67$$

$$t = \frac{18.4 - 20}{8.67 / \sqrt{5}} = -0.412$$



İddia kabul edilir, -0.412 değeri kabul bölgesi içerisindedir.

Uygulamalar

Uygulama Soruları

Bu Bölümde Ne Öğrendik Özeti

Bu bölümde anakütle parametresinin(ortalama,oran) belli bir değere eşit olup olmadığını sınamak amacıyla başvurulmuş hipotez testleri incelendi. Sıfır ve alternatif hipotezlerin nasıl oluşturulduğu, ret ve kabul alanlarının nasıl belirlendiği, test istatistiğinin nasıl düzenlendiği çeşitli örnekler üzerinde anlatıldı

Bölüm Soruları

1) Üniversitede okuyan öğrencilerden düzenli spor yapanların oranının %40'dan fazla olduğu iddiasını araştırmak amacıyla görüşülen 200 öğrenciden 88'i düzenli spor yapma alışkanlığının olduğunu söylemiştir. İddiayı $\alpha = 0,02$ sınavınız. İddia “%40 ‘dan farklı olduğu” şeklinde olursa yorumunuz değişir mi?

2) 4 yaşındaki çocukların ortalama boy uzunluklarının 92 cm olduğu iddia edilmektedir. Bir anaokulunda rastgele seçilen 4 yaşındaki 5 çocuğun boyları 90, 96, 94, 98, 92 cm ‘dir. İddiayı %2 ve % 20 anlam düzeylerinde sınavınız.

3) 3 yaşındaki çocukların ortalama boy uzunluklarının 82 cm olduğu iddia edilmektedir. Bir anaokulunda rastgele seçilen 3 yaşındaki 5 çocuğun boyları 80, 86, 84, 88, 82 cm ‘dir. İddiayı %1 ve % 18 anlam düzeylerinde sınavınız

4) Bir ilde yaşayan dört kişilik ailelerin günlük mutfak harcamalarını belirlemek için rastgele 4 kişi seçiliyor. 4 kişinin günlük harcamaları şöyledir: 50-20-30-20 tl. Bu ilde yaşayan 4 kişilik ailelerin ortalama mutfak harcamasını %95 güven düzeyinde tahmin ediniz.

11. GENEL TEKRAR

Bu Bölümde Neler Öğreneceğiz?

Bu hafta dersimizde ilk altı ders boyunca gördüğümüz konularla ilgili örnek soru çözümlerine yer vereceğiz.

Bölüm Hakkında İlgi Oluşturan Sorular

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
	.	
	.	

Anahtar Kavramlar

Giriş

11.1. Soru Çözümü

- 1- Bir işletmede üretilen vidaların çaplarının uzunluğu ortalaması 20 mm ve standart sapması da 4 mm olan normal dağılıma uymaktadır. Rastgele seçilen bir vidanın uzunluğunun 17.8 mm'den az olma olasılığı nedir?

$$X \sim N(20, 16)$$

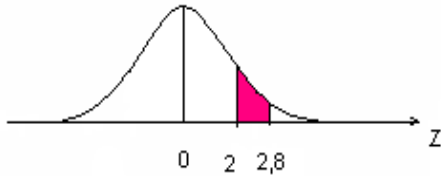
$$P(x < 17.8) = P\left(\frac{x - \mu}{\sigma} < \frac{17.8 - 20}{4}\right) = P(z < -0.55)$$

$$P(z < -0.55) = 0.5 - 0.2088 = 0.2912$$

- 2- Bir işletmede paketlenen margarin kutuların ağırlığı 100gr ve standart sapması da 2 gr. olan normal dağılıma uymaktadır. Paketlerin margarinlerden ağırlıkları 104 ile 105.6 gr arasında olanların oranı nedir?

$$Z_1 = \frac{x_1 - \mu}{\sigma} = \frac{104 - 100}{2} = 2$$

$$Z_2 = \frac{x_2 - \mu}{\sigma} = \frac{105.6 - 100}{2} = 2.8$$



$$0.4974 - 0.4773 = 0.0201 \text{ . } \% 2' \text{ dir.}$$

- 3- Bir bankanın açtığı müfettiş yardımcılığı yazılı sınavında alınan notlar 75 ortalama ve 16 standart sapma ile normal dağılıma uymaktadır. Adayların %15'i sözlü sınava katılmaya hak kazanmıştır. Sözlü sınava katılabilmek için gereken en düşük not nedir?

$$0.5 - 0.15 = 0.35 \text{ z tablosundan bakıldığında bu değer } 1.04$$

$$z = \frac{x - 75}{16} = 1.04$$

$$z = 91.64$$

- 4- Bir kurumun açtığı sınavda alınan puanların ortalaması 75 ve standart sapması da 15 olan normal dağılıma uymaktadır. Katılımcıların %10'u yeterli soru çözmediklerinden değerlendirmeye katılamamıştır. Buna göre değerlendirmeye katılabilmek için en az kaç puan almak gerekmektedir?

$$0.5 - 0.1 = 0.4, z = -1.28$$

$$z = \frac{x - 75}{16} = -1.28$$

$$z = 54.52$$

- 5- Tamsayım ve örnekleme kavramlarını açıklayınız.

Tamsayım sonlu bir ana kütlein bütün birimlerinin incelenmesi olarak isimlendirilir. Anakütleden çeşitli büyüklüklerde seçilen ve altkümeler örneklem, örneklemelerin seçilme sürecine ise örnekleme adı verilir.

- 6- Örnekleme yapmayı gerektiren nedenleri sıralayınız.

Maliyet, tamsayım bütçesi örnekleme bütçesinden daha fazladır

Zaman, örnekleme tamsayımına göre daha kısa zamanda sonuç verir.

- 7- Örnekleme dağılımı nedir?

Bir anakütleden belirli büyüklüklerde seçilen örneklemeler üzerinden hesaplanan istatistikler örneklemden örnekleme değişecektir. Hesaplanan bu istatistikler bir dağılım gösterir. Bu dağılıma örnekleme dağılımı denir.

- 8- Örnekleme Hatası ve standart hata kavramlarını açıklayınız.

Örneklemden örnekleme değişen değerler alan istatistiklerin ana kütle parametre değerlerine göre gösterdikleri sapmalara örnekleme hatası denilir. Standart hata, bir örneklem istatistiğine ait dağılımın değişkenliğinin bir ölçüsüdür, standart sapmadır.

9-Bir su dağıtım şirketi 400 abonesine ortalama su tüketimlerini sormuş aylık ortalama su tüketiminin 25 m^3 ve standart sapmanın da 5 m^3 şeklinde olduğunu

hesaplamıştır. Anakütle aylık ortalama su tüketiminin %99 güven sınırlarını hesaplayınız.

$$s = \frac{5}{\sqrt{400}} = 0.25$$

$$25 \pm 2.58(0.25)$$

$$24.333 \leq \mu \leq 25.645$$

- 9- İktisat Fakültesinde okuyan öğrencilerin düzenli spor yapma alışkanlıklarını değerlendirmek amacıyla 320 öğrenciden oluşan bir örneklem seçilmiştir. 208 öğrenci düzenli olarak spor yaptığını söylemiştir. Buna göre anakütlede düzenli olarak spor yapanların oranı için güven aralığını belirleyiniz.

$$\hat{p} = \frac{208}{320} = 0.65$$

$$s = \sqrt{\frac{0.65(1-0.65)}{320}} = 0.027$$

$$0.65 \pm 1.96(0.027)$$

$$0.59 \leq P \leq 0.70$$

- 10- İki farklı türdeki arpa tohumlarının verimlerini karşılaştırmak amacıyla araştırılma yapılmak isteniyor. Seçilen 144 tarlaya A türü arpa uygulanıyor, elde edilen ortalama ürün 300kg ve standart sapma da 70 kg olarak hesaplanıyor. Seçilen 169 tarlaya B türü arpa uygulandığında elde edilen ortalama ürün 380 kg ve standart sapma da 100kg olarak hesaplanıyor. A ve B türü arpa tohumlarının verimliliği arasındaki farkın güven sınırlarını %95 olasılıkla belirleyin.

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{70^2}{144} + \frac{100^2}{169}} = 9.65$$

$$(300 - 380) \pm 1.96(9.65)$$

$$-61 \leq \mu_1 - \mu_2 \leq -99$$

B türü arpa tohumunun verimi daha yüksektir.

11-Aynı dersi veren iki öğretim üyesinin başarıları kıyaslanmak isteniliyor. A öğretim üyesinin dersine katılan 400 öğrenciden 280'i, B öğretim üyesinin dersine katılan 440 öğrenciden 290'ı sınavdan geçtiğine göre, öğretim üyelerinin başarı oranları arasındaki farkın güven sınırlarını %99 güvenle belirleyiniz.

$$\hat{p}_1 = \frac{280}{400} = 0.7 \quad \hat{p}_2 = \frac{290}{440} = 0.66$$

$$s = \sqrt{\frac{0.7 * 0.3}{400} + \frac{0.66 * 0.34}{440}} = 0.032$$

$$(0.7 - 0.66) \pm 2.58(0.032)$$

$$-0.04 \leq P_1 - P_2 \leq 0.12$$

Öğretim üyelerinin başarı düzeyleri arasında istatistiksel açıdan önemli bir fark yoktur.

12-Bir firmanın ürettiği ürünlerin ağırlığının 25 gr. olduğu iddia edilmektedir. Rastgele seçilen 144 ürünün ortalama ağırlığı 22 gr ve standart sapma da 6gr ise %1 anlamlılık düzeyinde iddiayı değerlendiriniz.

$$H_0 : \mu = 22$$

$$H_1 : \mu \neq 22$$

$$s = \frac{6}{\sqrt{144}} = 0.5$$

$$z = \frac{22 - 25}{0.5} = -6$$

Tablo değeri 2.33 olduğu için sıfır hipotezi reddedilir.

13- Bir toplulukta sigara içenlerin oranının %40 olduğu iddia edilmektedir. Rastgele seçilen 250 kişiden 110'u sigara kullandığını söylediğine göre %5 anlamlılık düzeyinde iddiayı test ediniz.

$$H_0 : P = 0.40$$

$$H_0 : P = 0.40$$

$$\sigma = \sqrt{(0.4 * 0.6) / 250} = 0.03$$

$$z = \frac{0.44 - 0.4}{0.03} = 1.33$$

1.33 < 1.96 olduğundan sıfır hipotezi kabul edilir.

14- Bir başkanlık seçiminde A ve B bölgelerinden belli bir adayı destekleme oranlarının aynı olacağı iddia edilmektedir. A bölgesinden rastgele seçilen 50 kişiden 15'i, B bölgesinden rastgele seçilen 60 kişiden 24'ü sözkonusu başkan desteklediğine göre iddiayı %5 anlamlılık düzeyinde test ediniz.

$$H_0 : P_1 = P_2$$

$$H_0 : P_1 \neq P_2$$

$$P_1 = 15 / 50 = 0.3$$

$$P_2 = 24 / 60 = 0.4$$

$$\sigma = \sqrt{(0.3 * 0.7) / 50 + (0.4 * 0.6) / 60} = 0.091$$

$$z = \frac{0.3 - 0.4}{0.09} = -1.1$$

1.65 tablo değeri hesaplanan 1.1 değerinden büyüktür. Sıfır hipotezi kabul edilir.

14) Bir dil sınavında A ve B kurslarının başarı oranlarının aynı olduğu iddia edilmektedir. Bu amaçla A kursuna devam eden 100 kişinin ortalama skoru 87.5 ve sapması 22.5 olarak hesaplanmıştır. B kursuna devam eden 125 öğrenciye notları sorulmuş ve ortalama notları 97.5 ve standart sapması da 27.5 olarak hesaplanmıştır. %1 anlamlılık düzeyinde iddiayı sınavınız.

$$H_0 : \bar{X}_1 = \bar{X}_2$$

$$H_0 : \bar{X}_1 \neq \bar{X}_2$$

$$s = \sqrt{22.5^2 / 100 + 27.5^2 / 125} = 3.33$$

$$z = \frac{87.5 - 97.5}{3.33} = 3$$

2.33 < 3 olduğundan A dershanesinin daha başarılı olduğu söylenebilir.

15) Bir GSM operatörünün kullanıcılarına sunduğu astroloji paketine talep gösteren kadınların oranının erkeklerden daha fazla olduğu iddia edilmektedir. Rassal olarak seçilen 100 kadından 60'ı ve 140 erkekten 49'u astroloji paketini almayı tercih ediyorsa, iddiayı %1 anlamlılık düzeyinde test ediniz.

$$H_0 : P_1 = P_2$$

$$H_1 : P_1 \neq P_2$$

$$p_1 = \frac{60}{100} = 0.6$$

$$p_2 = \frac{49}{140} = 0.35$$

$$s = \sqrt{(0.6 \times 0.4) / 100 + (0.35 \times 0.65) / 100} = 0.063$$

$$z = \frac{0.60 - 0.35}{0.063}$$

$$z_{tablo} = 2.33$$

$$H_0 \text{ reddedilir}$$

16- Bir ilçede oturan aileler arasından 200 aile rassal olarak seçilmiş ve bunlardan 40'nın evinde masaüstü bilgisayarı olduğu sonucuna varılmıştır. Buna göre bu ilçede yaşayanların masaüstü bilgisayara sahip olma olasılığını %99 güvenle bulunuz.

$$p = \frac{40}{200} = 0.2$$

$$s = \sqrt{\frac{(0.20 \times 0.80)}{200}} = 0.028$$

$$0.2 \pm 2.58(0.028)$$

17-Bir fabrikada üretilen pillerin ortalama ömürleri 800 saat ve standart sapmaları 72 saat olan normal dağılıma uymaktadır. Rastgele seçilen bir pilin ömrünün 764-872 saat arasında olması olasılığı nedir?

$$z_1 = \frac{764 - 800}{72} = -0.50$$

$$z_2 = \frac{872 - 800}{72} = 1$$

$$p(-0.50 < z < 1) = 0.1915 + 0.3413 = 0.5328$$

%53

Uygulamalar

Uygulama Soruları

Bu Bölümde Ne Öğrendik Özeti

Bölüm Soruları

1)

a)

b)

c)

d)

Cevaplar

1)...., 2)...., 3)...., 4)...., 5)...

12. REGRESYON ANALİZİ II

Bu Bölümde Neler Öğreneceğiz?

Bu bölümde tahmin edilen anakütle regresyon parametrelerinin istatistiksel olarak anlamlılığını sınyacağız

Bölüm Hakkında İlgi Oluşturan Sorular

- Tahmin edilmiş olan gelir tüketim denkleminde, gelir gerçekten istatistiksel olarak anlamlı mıdır?
- Tahmin edilen regresyon denkleminde gerçekleşen hata nedir?

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Regresyon parametrelerinin anlamlılığının sınanması	Regresyon denkleminde tahminin standart hatasını bulmak, parametrelerin anlamlılığını sınamak	Metinler, çözümlü problemler.

Anahtar Kavramlar

- Standart hata
- Eğim katsayısının/sabit katsayının standart hatası
- Eğim katsayısının/standart hatanın anlamlılığı

Giriş

Önceki bölümde basit doğrusal regresyon modelinde sabit ve eğim parametrelerinin nasıl belirlendiğini öğrendik bu bölümde ise söz konusu parametre tahminlerinin istatistiksel olarak anlamlılığını sınayacağız

12.1.REGRESYON DOĞRUSUNUN ÖZELLİKLERİ VE HATA (ARTIK) KAVRAMI

Regresyon doğrusu denklemi $E(Y_i) = \beta_0 + \beta_1 X_i$ olmak üzere burada β_0 , başlangıç terimidir, modelin sabit değeri olarak da bilinir. $X=0$ iken

$E(Y_i)$ 'nin değeridir. β_1 doğrunun eğimine karşılık gelir. X 'deki bir birim değişiminin $E(Y_i)$ 'deki değişimine karşılık gelir. Burada β_0 ve β_1 parametrelerdir. Bu parametrelere veri üzerinden b_0 ve b_1 tahminleriyle ulaşılır.

Y , bağımlı değişkenin elde edilen her bir gözlemi (Y_i) ana kütle ortalaması $E(Y_i)$ olan bir ana kütleden gelen rastlantı değişkeni olduğu varsayılır. Y_i gözleminin $E(Y_i)$ 'den sapması hata terimi olarak isimlendirilir ve ε ile gösterilir.

Belirlenen X değerine karşılık gelen tahmini Y değeri ya da diğer bir ifadeyle X 'in belirli değeri için, Y 'nin anakütle ortalamasının tahmini $E(Y_i)$ şöyle bulunur:

$$\hat{Y}_i = b_0 + b_1 X_i$$

$$\hat{Y}_i = \bar{Y} + b_1 (X_i - \bar{X})$$

Gözlenen Y_i değeri ile tahmini değer yani $\hat{Y}_i$ değeri karşılaştırıldığında model ile veri arasındaki uyum için bir büyüklük elde edilir buna **artık** denir.

$$e_i = Y_i - \hat{Y}_i$$

Artık, tahmin edilen modelle veri arasındaki farka karşılık gelir eğer modelde sabit terim varsa artıkların toplamı sıfırdır.

$Y_i = \hat{Y}_i + e_i$ eşitliği dikkate alındığında, $\hat{Y}_i$ Y_i gözlemini açıklayabilen fakat e_i Y_i gözlemini açıklayamayan kısımdır. Belirli varsayımlar sağlandığında artıklar tahmini hatalar olarak kabul edilir. Hatırlanacağı üzere hata,

$$\varepsilon_i = Y_i - E(Y_i) \text{ şeklindedir.}$$

12.1. Regresyon Doğrusunun Özellikleri

-Artıkların toplamı sıfırdır

$$\sum_{i=1}^n e_i = 0$$

-Artık Kareler toplamı minimumdur.

-Gözlenen Y ve tahmini Y değerlerinin toplamı birbirine eşittir.

$$\sum_{i=1}^n Y_i = \sum_{i=1}^n \hat{Y}_i$$

Dolayısıyla ortalamaları da eşittir.

- Bağımsız değişkenle artıkların çarpım toplamı sıfırdır.

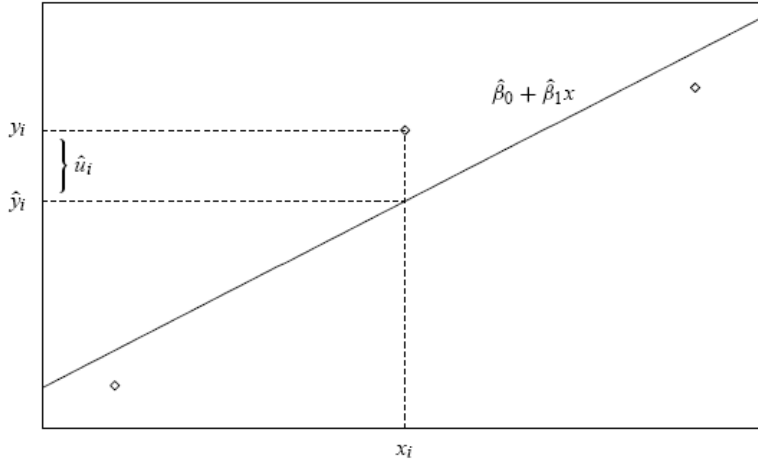
$$\sum_{i=1}^n X_i e_i = 0$$

-Tahmini Y değerleri ile artıkların çarpımlar toplamı sıfırdır.

$$\sum_{i=1}^n \hat{Y}_i e_i = 0$$

- Regresyon doğrusu daima $(\bar{X}, \bar{Y})$ noktasından geçer.

Gözlenen Y_i , tahmini $\hat{Y}_i$ ve hata arasındaki ilişki aşağıdaki temsili çizimde de görülebilmektedir.



Standart Hata

Regresyon denklemini standart hatasının tahmini hata kareler toplamının $n-2$ 'ye bölünerek karekökünün alınmasıyla bulunur. Buna tahminin standart hatası da denilir.

$$S = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n-2}} \text{ şeklinde tanımlanır.}$$

Regresyon Denkleminin Varsayımları

- Bağımlı değişken tesadüfi değişkendir, normal dağılır.
- Hatalar tesadüfidir ve normal dağılır.
- Hatalar birbirinden bağımsızdır (otokorelasyon yoktur)
- Her bağımsız değişkenin değerlerine ait olan bağımlı değişken değerlerinin alt kümelerinin varyansları birbirine eşittir (Homoskedastisite)
- Bağımsız değişkenler arasında basit doğrusal ilişki olmamalıdır.

12.2. Regresyon Denklemi Parametrelerinin Hipotez Testi ve Güven Aralığı

EKK yöntemi ile $Y = b_0 + b_1X$ regresyon denkleminin b_0 ve b_1 parametrelerinin tahmini değerleri $\hat{b}_0$ ve $\hat{b}_1$ örneklemden örnekleme değişen birer rastlantı değişkenidir. Bu rastlantı değişkenlerinin ortalama ve varyansları sırayla aşağıdaki gibidir.

$$E(\hat{b}_0) = b_0 \qquad V(\hat{b}_0) = \frac{\sum X^2}{n \sum (X - \bar{X})^2} \sigma^2$$

$$E(\hat{b}_1) = b_1 \qquad V(\hat{b}_1) = \frac{\sigma^2}{\sum (X - \bar{X})^2}$$

Regresyon analizindeki varsayımlardan ilki hatırlanacağı üzere bağımlı değişkenin Y 'nin normal dağıldığı varsayımı idi. Gerek $\hat{b}_0$ ve gerekse $\hat{b}_1$ bağımlı değişkenle doğrusal ilişki içinde olduğundan bu iki rastlantı değişkeninin de normal dağıldığı sonucuna varılır. Bu sonuca dayanarak b_0 ve b_1 için $\hat{b}_0$ ve $\hat{b}_1$ üzerinden yola çıkarak hipotez testi ve güven aralıklarını tanımlamak mümkündür. $\hat{b}_0$ ve $\hat{b}_1$ rastlantı değişkenlerinin normal dağılması halinde önceki konulardan hatırlanacağı üzere

$$z = \frac{\hat{b}_0 - b_0}{\sigma_{\hat{b}_0}} \quad \text{ve} \quad z = \frac{\hat{b}_1 - b_1}{\sigma_{\hat{b}_1}}$$

rastlantı değişkenleri de normal dağılır.

$\hat{b}_0$ ve $\hat{b}_1$ tahminlerinin standart hataları şöyledir:

$$S(\hat{b}_0) = \sqrt{\frac{\sum X^2}{n \sum (X - \bar{X})^2}} S$$

$$S(\hat{b}_1) = \frac{S}{\sqrt{\sum (X - \bar{X})^2}}$$

Varyans bilinmediği zaman z yerine t eşitlikleri kullanılır:

$$t = \frac{\hat{b}_0 - b_0}{S_{\hat{b}_0}}$$

$$t = \frac{\hat{b}_1 - b_1}{S_{\hat{b}_1}}$$

Bu istatistikler $n-2$ serbestlik dereceli t dağılır. Bu eşitliklerden yola çıkarak sırasıyla sabit ve eğim parametreleri için güven aralıkları şöyledir:

$$P[-t_{n-2, \alpha/2} < \frac{\hat{b}_0 - b_0}{S_{\hat{b}_0}} < t_{n-2, \alpha/2}] = 1 - \alpha$$

$$P[\hat{b}_0 - t_{n-2, \alpha/2} S_{\hat{b}_0} \leq b_0 \leq \hat{b}_0 + t_{n-2, \alpha/2} S_{\hat{b}_0}] = 1 - \alpha$$

$$\hat{b}_0 - t_{n-2, \alpha/2} S_{\hat{b}_0} \leq b_0 \leq \hat{b}_0 + t_{n-2, \alpha/2} S_{\hat{b}_0}$$

Sabit parametre için yukarıdaki eşitsizliklerin düzenlenmesiyle bulunan yaklaşım eğim parametresi için de yapıldığında bulunan güven aralığı şöyledir:

$$\hat{b}_1 - t_{n-2, \alpha/2} S_{\hat{b}_1} \leq b_1 \leq \hat{b}_1 + t_{n-2, \alpha/2} S_{\hat{b}_1}$$

İlgilinenilen hipotez testi sabit katsayı için aşağıdaki gibi düzenlendiğinde,

$$H_0 : b_0 = b$$

$$H_1 : b_0 \neq b$$

aşağıdaki test istatistiği kullanılır:

$$t_0 = \frac{\hat{b}_0 - b}{S_{\hat{b}_0}}$$

Alternatif hipotez yukarıda çift yönlü düzenlenmiş olmakla beraber aşağıdaki gibi tek yönlü de düzenlenebilir:

$$H_1 : b_0 < b$$

$$H_1 : b_0 > b$$

Alternatif hipotezin çift yönlü olması halinde eğer bulunan kritik değer ve tablo değeri arasında aşağıdaki gibi bir ilişki varsa H_0 reddedilir.

$$t_0 > t_{n-2, \alpha/2} \quad \text{veya} \quad t_0 < -t_{n-2, \alpha/2}$$

Alternatif hipotez eğer aşağıdaki gibi tek yönlü ise,

$$H_1 : b_0 < b$$

$t_0 < -t_{n-2, \alpha}$ halinde H_0 hipotezi reddedilir.

Ve son olarak da alternatif hipotez

$$H_1 : b_0 > b \text{ şeklindeyse}$$

$t_0 > t_{n-2, \alpha}$ olmalı halinde H_0 hipotezi reddedilir.

Eğim parametresinin hipotez testinde genellikle aşağıdaki şekilde düzenlenir. Katsayının genellikle sıfıra eşitliği sınanır. Çünkü katsayının sıfıra eşitliği bağımlı değişkenin bağımsız değişkene doğrusal bağımlı olmadığı anlamına gelir. Sıfır hipotezi ve olası alternatif hipotezler aşağıdaki gibi düzenlenir;

$$H_0 : b_1 = 0$$

$$H_1 : b_1 \neq 0$$

$$H_1 : b_1 > 0$$

$$H_1 : b_1 < 0$$

Burada kullanılan istatistik aşağıdaki gibidir.

$$t_0 = \frac{\hat{b}_1}{S_{\hat{b}_1}}$$

Tablodan bulunan kritik değerle yukarıdaki istatistikten hesaplanan değer karşılaştırılması sabit parametre için hipotez testinde değinildiği gibidir, aynıdır.

ÖRNEK: Önceki örnekte bir sınıftaki öğrencilerin matematik ve muhasebe derslerindeki başarıları arasındaki ilişki incelenmişti. Regresyon denkleminin parametrelerinin %99 ve %95 güvenle bulundukları aralıkları belirleyiniz ve katsayıların anlamlılığını aynı güven düzeyinde sınavınız.

Öncelikle tahminin standart hatasını bulalım;

$$\hat{y} = -0.625 + 0.9375(2) = 1.25$$

$$\hat{y} = -0.625 + 0.9375(3) = 2.188$$

$$\hat{y} = -0.625 + 0.9375(5) = 4.063$$

$$\hat{y} = -0.625 + 0.9375(6) = 5$$

$$\hat{y} = -0.625 + 0.9375(7) = 5.938$$

$$\hat{y} = -0.625 + 0.9375(10) = 8.75$$

$$\hat{y} = -0.625 + 0.9375(7) = 5.938$$

$$\hat{y} = -0.625 + 0.9375(8) = 6.875$$

Y	$\hat{Y}$	$(Y - \hat{Y})^2$
1	1,25	0,0625
2	2,188	0,035344
3	4,063	1,129969
5	5	0
6	5,938	0,003844
7	8,75	3,0625
8	5,938	4,251844
8	6,875	1,265625
40	40,002	9,811626

$$S = \sqrt{\frac{9.81}{8-2}} = 1.278$$

$$0.486 \leq b_1 \leq 1.389$$

$$S_{\hat{b}_1} = \frac{S}{\sqrt{\sum (X - \bar{X})^2}} = \frac{1.278}{\sqrt{48}} = 0.1846 \text{ olacağından } \%99 \text{ ve } \%95 \text{ güvenle parametrenin}$$

yer alacağı aralıklar şöyledir:

$$0.9375 - (0.1846)(3.707) \leq b_1 \leq 0.9375 + (0.1846)(3.707)$$

$$0.2532 \leq b_1 \leq 1.622$$

$$(t_{0.01;6} = 3.707)$$

$$0.9375 - (0.1846)(2.447) \leq b_1 \leq 0.9375 + (0.1846)(2.447)$$

$$0.486 \leq b_1 \leq 1.389$$

$$(t_{0.05;6} = 2.447)$$

Sabit parametrenin standart hatası, güven aralığı aşağıdaki gibidir:

$$S_{\hat{b}_0} = \sqrt{\frac{\sum X^2}{n \sum (X - \bar{X})^2}} S = \sqrt{\frac{336}{8(48)}} 1.278 = 1.195$$

$$-0.625 - (1.195)(3.707) \leq b_0 \leq -0.625 + (1.195)(3.707)$$

$$-5.055 \leq b_0 \leq 3.805$$

$$-0.625 - (1.195)(2.447) \leq b_0 \leq -0.625 + (1.195)(2.447)$$

$$-3.549 \leq b_0 \leq 2.299$$

Eğim parametresinin anlamlılığı için kurulan hipotezler ve testi şöyledir:

$$H_0 : b_1 = 0$$

$$H_1 : b_1 \neq 0$$

$$t_0 = \frac{0.9375}{0.1846} = 5.079$$

elde edilen değer her iki anlamlılık düzeyindeki kritik değerlerle sınıldığında sıfır hipotezi reddedilir. Yani, Y bağımlı değişkeni X'e bağımlıdır.

Sabit katsayının testi benzeri yaklaşımla incelendiğinde,

$$H_0 : b_0 = 0$$

$$H_1 : b_0 \neq 0$$

$$t_0 = \frac{-0.625}{1.195} = -0.523$$

Her iki anlamlılık düzeyinde sıfır hipotezi kabul edilir.

-Bağımlı değişkenin ortalama değeri için güven aralığı bulunmak istendiğinde, belirli bir x değerinde bağımlı değişkenin ortalamasının standart hatası aşağıdaki biçimde bulunur:

$$S_{y_i} = \sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x - \bar{x})^2}} S$$

- Eğer belirli bir x değerine karşılık gelen tek bir nokta için güven aralığı oluşturmak istenirse kullanılacak standart hata şöyledir:

$$S_{Y_d} = \sqrt{1 + \frac{1}{n} + \frac{(x_d - \bar{x})^2}{\sum (x - \bar{x})^2}} S$$

ÖRNEK: Yukarıdaki örneğe ait verileri kullanarak matematik dersinden 9 alan öğrencilerin muhasebe dersinden alacakları ortalama not için %95 güven aralığı oluşturunuz.

$$S = \sqrt{\frac{1}{8} + \frac{(9-6)^2}{48}} 1.278 = 0.559(1.278) = 0.714$$

$$\hat{Y} = -0.625 + 0.9375(9) = 7.8125$$

Aralık tahmini için:

$$7.8125 - 2.447(0.714) \leq Y_9 \leq 7.8125 + 2.447(0.714)$$

$$6.065 \leq Y_9 \leq 9.56$$

ÖRNEK: Yukarıdaki örneğe ait verileri kullanarak matematik dersinden 9 alan bir öğrencinin muhasebe dersinden alacağı not için %95 güven aralığı oluşturunuz.

$$S = \sqrt{1 + \frac{1}{8} + \frac{(9-6)^2}{48}} 1.278 = 1.146(1.278) = 1.465$$

$$\hat{Y} = -0.625 + 0.9375(9) = 7.8125$$

Aralık tahmini:

$$7.8125 - 2.447(1.465) \leq Y_9 \leq 7.8125 + 2.447(1.465)$$

$$4.23 \leq Y_9 \leq 11.40$$

ÖRNEK:

Önceki bölümde ele alınan suç ve işsizlik oranlarına ilişkin örnekteki parametrelerin anlamlılığını % 5 önem düzeyinde sınamak istersek;

X		Y			
işsizlik oranı	suç oranı	Ytahmin	(Y- Ytah)	(Y- Ytah)^2	
0,8	3	2,592	0,408	0,166464	
1,4	6	5,211	0,789	0,622521	
2,3	7	9,1395	2,1395	4,57746	
3,5	15	14,3775	0,6225	0,387506	
4,5	19	18,7425	0,2575	0,066306	
12,5	50	50,0625	0,0625	5,820258	

$$\sum Y = 50 \quad \sum X = 12.5 \quad \sum XY = 164.9$$

$$\sum X^2 = 40.39 \quad \sum Y^2 = 680$$

$$\sum (X - \bar{X})(Y - \bar{Y}) = 39.9$$

$$\sum (X - \bar{X})^2 = 9.14 \quad \sum (Y - \bar{Y})^2 = 180$$

$$Y = -0.9 + 4.365X$$

$$S = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n-2}} = \sqrt{\frac{5.82}{5-3}} = 1.39$$

$$S(\hat{b}_1) = \frac{S}{\sqrt{\sum (X - \bar{X})^2}} = \frac{1.39}{\sqrt{9.14}} = 0.455$$

$$S_{\hat{b}_0} = \sqrt{\frac{\sum X^2}{n \sum (X - \bar{X})^2}} S = \sqrt{\frac{40.39}{5(9.14)}} 1.39 = 1.306$$

$$4.365 - 3.82 * 0.455 \leq b_1 \leq 4.365 + 3.82 * 0.455$$

$$2.63 \leq b_1 \leq 6.1$$

$$t = \frac{4.365}{0.455} = 9.6$$

Eğim parametresi anlamlıdır, güven aralığı sıfırı kapsamadığı gibi hesaplanan t istatistik değeri de tablo değerinden büyüktür ($9.6 > 3.82$)

$$-0.9 - 3.82 * 1.306 \leq b_0 \leq -0.9 + 3.82 * 1.306$$

$$-5.88 \leq b_0 \leq 4$$

$$t = \frac{-0.9}{1.306} = -0.69$$

Sabit parametresi ise anlamsızdır. Aralık sıfırı kapsar ve test istatistiği sonucu kritik değerden küçük çıkmıştır.

Uygulamalar

Uygulama Soruları

Bu Bölümde Ne Öğrendik Özeti

Bölüm Soruları

Aşağıda matematik dersine çalışma süresi ve alınan notlara ilişkin bilgi mevcuttur. Regresyon denklemini kurarak eğim ve sabit parametrelerin anlamlılığını sınavınız.

çalışma saati	alınan not
3	30
5	45
6	55
8	65
9	80

-Tahmin edilen notların toplamı yaklaşık olarak alınan notlara eşit mi?

-Hataların toplamı yaklaşık olarak sıfır mı?

2- Uygulanan ilaç dozu ve iyileşme sürelerine ilişkin regresyon denkleminin parametrelerinin anlamlılığını sınavınız.

Doz	İyileşme süresi
1	8
2	8
3	6
4	4
5	2

13. Kİ-KARE TESTLERİ

Bu Bölümde Neler Öğreneceğiz?

13.1. Ki kare testleri ile sayısal olmayan değişkenler arasındaki ilişkinin varlığı, iki ya da daha fazla sayıda örneklemin aynı anakütleden gelip gelmediği, bir seriye ait frekans dağılımının belli bir dağılıma uygun olup olmadığı sınamalarını yapmayı öğreneceğiz.

Bölüm Hakkında İlgi Oluşturan Sorular

- Eğitim düzeyi ile izlenen televizyon programları arasında ilişki var mıdır?
- Kadınların makyaj yapma sıklığı ile yaşları arasında bir ilişki var mıdır?
- Annenin eğitim düzeyi çocuğun okuldaki başarısını etkiler mi?

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Kikare Testleri	İki Nitel değişken arasında ilişkinin varlığını değerlendirmek	Metinler, çözümlü problemler

Anahtar Kavramlar

- Ki kare Uyum testi
- Ki kare Bağımsızlık testi

Giriş

13.1. Kİ-KARE TESTLERİNİ NEDEN KULLANIRIZ?

İstatistik I derslerinden hatırlanacağı üzere istatistikte değişkenler sadece nicel(sayısal) olanlarla sınırlı değildir, nitel değişkenler de istatistiğin konusunu oluşturur. Önceki derslerde nicel değişkenler tahmin teknikleri ve hipotez testlerinin konusunu oluşturdu. Zaman zaman analizlerde sayısal olmayan yani nitel değişkenler de dikkate alınabilir. Örneğin insanların eğitim düzeyi ve izledikleri programlar arasındaki ilişkinin varlığı araştırılmak istenebilir. Böyle bir durumda söz konusu 2 değişken de niteldir ve şıkları sırasıyla şu şekilde ifade edilebilir; “ilkokul-lise-üniversite”, “dizi-eğlence-yarışma”.

Ki kare testleri ile sayısal olmayan değişkenler arasındaki ilişkinin varlığı (ilişkinin varolmadığı sıfır hipotezi varsayımı ile), iki ya da daha fazla sayıda örneklemin aynı anakütleden gelip gelmediği, bir seriye ait frekans dağılımının belli bir dağılıma uygun olup olmadığı sınamaları yapılabilir. Ki-kare testleri üç başlık altında incelenir:

-Ki kare bağımsızlık

-Ki-kare homojenlik

-Ki-kare uyum

Ki kare testlerinde χ^2 dağılımı kullanılır. Kısaca bahsetmek gerekirse ki-kare dağılımı, X tesadüfi değişken dağılımından bir değer seçilip standart hale dönüştürülüp, karesinin alınmasıyla bulunur.

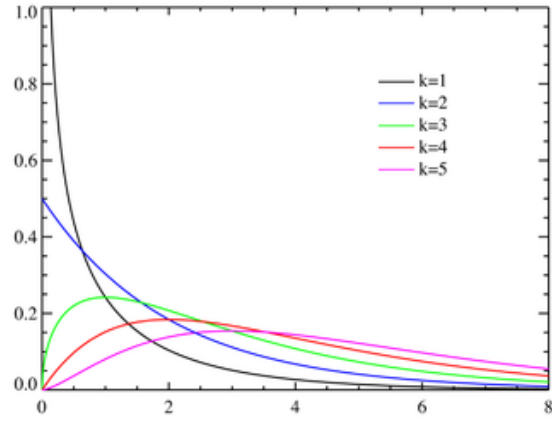
$$Z_i = \frac{X_i - \mu}{\sigma}$$

Standardize edilen X tesadüfi değişkeni, standart normal değişken Z_i 'nin aşağıdaki şekilde karesi alındığında ki-kare dağılımına dönüşür.

$$Z_i^2 = \left(\frac{X_i - \mu}{\sigma}\right)^2$$

Ki-kare dağılımlarının serbestlik derecesi n ise şu şekilde gösterilir:

$$\chi_n^2$$



Ki-kare dağılımı sağa çarpıktır ve normalden daha diktir. N büyüdükçe merkezi limit teoreminin bir sonucu olarak diklik ve asimetri azalır ve dağılım normale yaklaşır.

Ki-kare değerleri 0 ile ∞ arasında değişir. Yukarıdaki tabloda yer alan k değerleri serbestlik derecesine karşılık gelmektedir.

13.2. Kİ-KARE BAĞIMSIZLIK TESTİ

İncelenen birimlerin iki ya da daha fazla özelliğinin birbirlerinden bağımsız olup olmadıklarını sınamak amacıyla kullanılan bir testtir. Örneğin sigara içme alışkanlığı ve cinsiyet gibi, eğitim düzeyi ve alınan ücret düzeyi (düşük/orta/yüksek) gibi...

Bu tür araştırmalarda test yapılırken kontenjans tablosundan faydalanılır. Tabloda incelenen iki değişkene ilişkin şıklar yatay ve dikey gözelerde yer alır.

		Partition B				
Class		B_1	B_2	...	B_C	Sums
Partition A	A_1	n_{11}	n_{12}	...	n_{1C}	$n_{1\bullet}$
	A_2	n_{21}	n_{22}	...	n_{2C}	$n_{2\bullet}$
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	A_C	n_{R1}	n_{R2}	...	n_{RC}	$n_{R\bullet}$
	Sums	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet C}$	N

Hipotezler şu şekilde kurulur:

H_0 : A ve B özellikleri bağımsızdır

H_A : A ve B özellikleri bağımlıdır

Yukarıdaki kontenjans tablosunda satır sayısı R ve kolon sayısı C ile gösterilmiştir buna göre serbestlik derecesi, $sd = (c-1)(n-1)$ 'dir.

Bu hipotez testini uygulamak için öncelikle her bir “gözlenen” göze için “beklenen” değerlerin bulunması gerekir. Beklenen değerler, satır ve sütun toplamalarının çarpımlarının toplam gözlem sayısına oranlanması ile bulunur. Örneğin n_{11} gözlenen değerdir. Bu gözeye ilişkin beklenen değerse $((n_{1.}) * (n_{.1})) / N$ ile bulunur. Gözlenen ve beklenen arasındaki farkların büyüklüğüne bakmak gerekir. Aradaki fark mutlak değerce ne kadar büyükse, sıfır hipotezi hakkında o ölçüde şüphe duyulur. Her bir göze için beklenen ve gözlenen değerler gözönüne alınarak aşağıdaki test istatistiği hesaplanır.

$$\chi^2_{test} = \sum \frac{(G_i - B_i)^2}{B_i} \quad \chi^2_{tablo} = \chi^2_{(k-1)(r-1), \alpha}$$

$$\chi^2_{test} < \chi^2_{tablo} \rightarrow H_0 \text{ reddedilemez}$$

Örnek

Türkiye’de (Doğu ve Güneydoğu Anadolu B.) yaşayan kadınların evlenme şekilleri ve yaşları arasında bir ilişki var mıdır?

	Görücü	İsteyerek	Zorla	Akraba
13-16yaş	15	20	10	1
17-22 yaş	12	19	5	3
22+	2	5	5	3

H_0 : Evlenme usulu yaşa bağımlı değildir.

H_1 : bağımlıdır.

Gözlenen	Görücü	İsteyerek	Zorla	Akraba	satır toplamı
13-16yaş	15	20	10	1	46
17-22 yaş	12	19	5	3	39
22+	2	5	5	3	15
sütun toplamı	29	44	20	7	100

$$B_{11}=(29*46)/100= 13.3 \quad \dots$$

$$B_{34}=(7*15)/100=1.1$$

Tüm beklenen değerler yukarıdaki gibi hesaplanırsa şu tabloya ulaşılır;

	Görücü	İsteyerek	Zorla	Akraba
13-16yaş	13,3	20,2	9,2	3,2
17-22 yaş	11,3	17,2	7,8	2,7
22+	4,4	6,6	3	1,1

$$\chi^2 = \frac{(15-13.3)^2}{13.3} + \frac{(20-20.2)^2}{20.2} + \dots + \frac{(3-1.1)^2}{1.1} = 9.69$$

hesaplanan bu değer tablo değeri ile kıyılanırsa,

$$\chi^2_{3 \times 2, 0.05} = 12.59$$

(Ki-kare tablosu bölüm sonunda yer almaktadır)

H₀ hipotezi kabul edilir yani, evlenme usulu yaşa bağımlı değildir.

Örnek

Sigara kullanma alışkanlığına ilişkin bir araştırma için 120 kadın ve 140 erkekle görüşülüyor. Sigara kullanımının her iki cinsiyet için de aynı olduğu iddiası desteklenmeli midir ($\alpha = 0.05$) ?

	Kullanıyor	Kullanmıyor
Erkek	80	60
Kadın	60	60

Tabloda gözlenen değerler yer almaktadır. Her gözeyle ilişkin beklenen değerlerse şu şekilde bulunur:

	Kullanıyor	Kullanmıyor	
Erkek	80	60	140
Kadın	60	60	120
	140	120	260

H_0 : Sigara içme alışkanlığı cinsiyete bağımlı değildir.

H_1 : Sigara içme alışkanlığı cinsiyete bağımlıdır.

$$B_{11} = (140 \times 180) / 260 = 75.38 \quad B_{12} = (120 \times 140) / 260 = 64.62$$

$$B_{21} = (140 \times 120) / 260 = 64.62 \quad B_{22} = (120 \times 120) / 260 = 55.38$$

$$\chi^2 = \frac{(80 - 75.38)^2}{75.38} + \frac{(60 - 64.62)^2}{64.62} + \dots + \frac{(60 - 55.38)^2}{55.38} = 1.429$$

Tablo değeri $\chi^2_{1 \times 1, 0.05} = 3.841$ olduğundan sigara içme alışkanlığı ve cinsiyet arasında ilişki yoktur hipotezi kabul edilir.

11.3. Kİ-KARE HOMOJENLİK TESTİ

Bu test ile iki ya da daha fazla bağımsız örneklemin aynı anakütleden seçilip seçilmediğinin araştırılmasında kullanılır. Bağımsızlık testinden farklı olarak, sıfır hipotezi örneklemelerin aynı ana kütleden seçildiği şeklinde kurulur.

H_0 : Örneklemeler aynı ana kütleden seçilmiştir.

H_1 : Örneklemeler farklı ana kütlelerden seçilmiştir.

Örnek: Bir supermarket aynı işletmenin piyasaya sürdüğü 2 farklı şampuan markasına ilişkin tüketici ile memnuniyet anketi yapmıştır. Bu amaçla ilk grupta 50 ikinci grupta ise 75 kişiyle görüşülmüştür. Buna göre seçilen örneklemeler aynı anakütleden mi gelmektedir?

Şampuan	Begenmeyen	Kararsız	Begenen	Toplam
A	30	16	5	50
B	40	25	10	75
Toplam	70	40	15	125

Yukarıdaki örneklerde hesaplanıldığı gibi burada da her bir göze için tek tek beklenen frekanslar bulunur. Buna göre elde edilen değerler aşağıdaki gibi olacaktır.

$$B_{11} = (50 \times 70) / 125 = 28 \dots$$

$$B_{23} = (15 \times 75) / 125 = 9$$

Şampuan	Begenmeyen	Kararsız	Begenen
A	28	16	6
B	42	24	9

H_0 : İki örneklem aynı ana kütleden seçilmiştir.

H_1 : Örneklemeler farklı ana kütlelerden seçilmiştir.

$$\chi^2 = \frac{(30-28)^2}{28} + \dots + \frac{(10-9)^2}{9} = 0.631$$

Tablo değeri $\chi^2_{1 \times 2, 0.05} = 5.99$ olduğundan örneklemir aynı anakütleden seçilmiştir.

11.4. Kİ-KARE UYUM TESTİ

Bir örneklemin ilgili olduğu ana kütleyi temsil edip etmediğini belirlemek amacıyla kullanılan bir testtir. Testte beklenen frekanslar ilgili olasılık dağılımından elde edilmektedir.

Örnek:

Bir fotokopi makinesinin verdiği arıza sayılarının Poisson dağıldığı iddia edilmektedir. Aşağıda arıza haftaları ve sayılarının sıklığına ilişkin veri mevcuttur. Buna göre arıza sayılarının poisson dağılımına uyduğu iddiası desteklenmeli midir?

Arıza sayısı	0	1	2	3
Hafta	32	16	8	4

H_0 : Arıza sayıları poisson dağılımına uyar

H_1 : Arıza sayıları poisson dağılımına uymaz

Yukarıdaki tablodan görülebileceği gibi 60 haftanın 32 sinde hiç arıza oluşmazken, 4'ün de 3 arıza oluşmaktadır.

Öncelikle varsayılan poisson dağılımına ilişkin olasılıklar ve beklenen frekanslar bulunur sonrasında ki-kare sınaması yapılır.

$$p(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

Ortalama arızalanma sayısı $\bar{x} = ((32 \times 0) + (1 \times 16) + (2 \times 8) + (3 \times 4)) / 60 = 0.734$

$$p(x=0) = \frac{e^{-0.734} 0.734^0}{0!} = 0.48$$

$$p(x=1) = \frac{e^{-0.734} 0.734^1}{1!} = 0.35$$

$$p(x=2) = \frac{e^{-0.734} 0.734^2}{2!} = 0.13$$

$$p(x=3) = \frac{e^{-0.734} 0.734^3}{3!} = 0.04$$

Gözlenen frekans	32	16	8	4
Beklenen frekans	27	21	8,4	2,4

Beklenen frekans değerleri olasılık değerleri ve toplam frekans değeriyle çarpılarak bulunmuştur.

$$0.45 \times 60 = 27 \dots$$

$$\chi^2 = \frac{(32-27)^2}{27} + \dots + \frac{(4-2.4)^2}{2.4} = 3.21$$

$$\chi_{2,0.05}^2 = 5.99 \quad (\text{Burada serbestlik derecesi } k-m-1 \text{ den, } 4-1-1 \text{ şeklinde bulunmuştur.})$$

Buna göre dağılım poisson'a uymaktadır.

11.5. KONTENJANS KATSAYISI

İki değişken arasındaki ilişkinin büyüklüğü hakkında bilgi verir. Kontenjans tablolarından ($r > 2$, $c > 2$) hesaplanır, “c” notasyonu kontenjans katsayısını göstermek üzere $c = 0$ ise ilişki yoktur eğer $c = 1$ ise oldukça güçlü ilişki varlığını gösterir.

$$c = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

İle hesaplanır.

Örnek:

Eğitim Düzeyi ve Gelir arasındaki ilişki var mıdır? Bu ilişkinin derecesi nedir?

Gelir Düzeyi/ Eğitim Düzeyi	Yüksek		
	Lise	Üniversite	Lisans
Düşük	50	20	10
Orta	20	30	10
Yüksek	20	10	30

Soruda verilen gözlenen frekanslarla beraber beklenen frekans değerleri de dikkate alınmalıdır. Her bir gözeye ilişkin hesaplanan beklenen frekanslardan sonra ki-kare değeri hesaplanır.

$$\chi^2 = \frac{(50-36)^2}{36} + \dots + \frac{(30-15)^2}{15} = 42.93$$

Tablo değeri $\chi^2_{4,0.01} = 13.28$ olduğundan, gelir düzeyi ve eğitim düzeyi arasındaki ilişki varlığı kabul edilir.

Kontenjans katsayısı ise,

$$c = \sqrt{\frac{42.93}{42.93 + 200}} = 0.42$$

Orta düzeyde bir ilişki varlığına işaret eder.

Uygulamalar

Uygulama Soruları

Bu Bölümde Ne Öğrendik Özeti

Bölüm Soruları

- 1- Evli kadınların maruz kaldığı şiddet ve eşlerinin eğitim düzeyleri arasında ilişkiyi değerlendirmek için 150 evli kadınla görüşülmüştür, çalışma sonuçları aşağıdaki gibidir. Erkeğin eğitim düzeyi kadına yönelik şiddet eğilimi üzerinde etkili midir? ($\alpha = 0,01$) (15 puan)

	Sözlü şiddet	Fiziksel şiddet	Şiddet yok
İlk+Orta	30	10	20
Yüksek	40	10	40

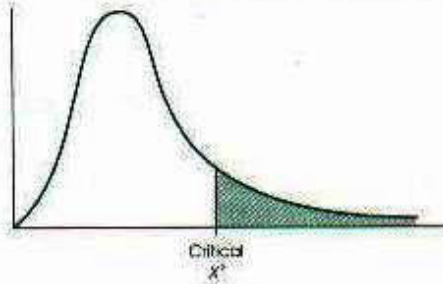
- 2- Sigara içme ve akciğer kanseri arasında bir ilişki varlığının sınanması için hangi test seçilmelidir?

	Akciğer Kanseri	
Sigara içme	Var	Yok
Kullanan	20	80
Kullanmayan	5	95

- 3- Tablo değerlerini $\alpha = 0.01$ ve $\alpha = 0.05$ olarak değerlendirin
- 4- Kontenjans katsayısını kullanarak yorumlayın

TABLE B.7 THE CHI-SQUARE DISTRIBUTION*

*The table entries are critical values of χ^2 .



df	PROPORTION IN CRITICAL REGION				
	0.10	0.05	0.025	0.01	0.005
1	2.71	3.84	5.02	6.63	7.88
2	4.61	5.99	7.38	9.21	10.60
3	6.25	7.81	9.35	11.34	12.84
4	7.78	9.49	11.14	13.28	14.86
5	9.24	11.07	12.83	15.09	16.75
6	10.64	12.59	14.45	16.81	18.55
7	12.02	14.07	16.01	18.48	20.28
8	13.36	15.51	17.53	20.09	21.96
9	14.68	16.92	19.02	21.67	23.59
10	15.99	18.31	20.48	23.21	25.19
11	17.28	19.68	21.92	24.72	26.76
12	18.55	21.03	23.34	26.22	28.30
13	19.81	22.36	24.74	27.69	29.82
14	21.06	23.68	26.12	29.14	31.32
15	22.31	25.00	27.49	30.58	32.80
16	23.54	26.30	28.85	32.00	34.27
17	24.77	27.59	30.19	33.41	35.72
18	25.99	28.87	31.53	34.81	37.16
19	27.20	30.14	32.85	36.19	38.58
20	28.41	31.41	34.17	37.57	40.00
21	29.62	32.67	35.48	38.93	41.40
22	30.81	33.92	36.78	40.29	42.80
23	32.01	35.17	38.08	41.64	44.18
24	33.20	36.42	39.36	42.98	45.56
25	34.38	37.65	40.65	44.31	46.93
26	35.56	38.89	41.92	45.64	48.29
27	36.74	40.11	43.19	46.96	49.64
28	37.92	41.34	44.46	48.28	50.99
29	39.09	42.56	45.72	49.59	52.34
30	40.26	43.77	46.98	50.89	53.67
40	51.81	55.76	59.34	63.69	66.77
50	63.17	67.50	71.42	76.15	79.49
60	74.40	79.08	83.30	88.38	91.95
70	85.53	90.53	95.02	100.42	104.22
80	96.58	101.88	106.63	112.33	116.32
90	107.56	113.14	118.14	124.12	128.30
100	118.50	124.34	129.56	135.81	140.17

Table 8 of E. Pearson and H. Hartley, *Biometrika Tables for Statisticians*, 3d ed, New York: Cambridge University Press, 1966. Adapted and reprinted with permission of the Biometrika trustees.

14. ENDEKS KAVRAMI

Bu Bölümde Neler Öğreneceğiz?

Bu bölümde günlük hayatta çok sık kullandığımız endekslere yer verilecektir.

Bölüm Hakkında İlgi Oluşturan Sorular

- Bir sepetteki 5 malın 2005 yılındaki fiyatı 2006 yılına göre nasıl bir değişim göstermiştir?
- Enflasyon hesapları neden farklı enstitüler tarafından farklı sonuçlar verir?

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın elde edileceği veya geliştirileceği
Endeksler	Endekslerin hesaplanması, yorumu, aralarındaki farklar	Metinler, çözümlü problemler.

Anahtar Kavramlar

- Zaman Endeksi
- Mekan Endeksi
- Laspeyres Endeksi
- Paasche Endeksi
- Fisher Endeksi

Giriş

Endeksler günlük hayatta özellikle ekonomi haberlerinde sıkça işittiğimiz bir kavramdır. Endeks bir orandır, fiyat değişimleri, ücret değişimleri hakkında bize bilgi verir.

14.1. ENDEKS (İNDEKS)

Günlük hayatta sıklıkla ülke ekonomisine ilişkin yorum yaparken endeks sözcüğü kullanılır. Endeks, belirli bir istatistik olaya ait değerlerin zaman ve yer itibariyle gösterdiği oransal değişmelerin ölçüsüdür. Tanım gereği endekslerde “temel” ve karşılaştırılmanın yapıldığı “cari” yıl dikkate alınır. Karşılaştırma yapılacak değer (cari) pay kısmına yazılırken temel değerse paydaya yazılır ve 100 ile çarpılır.

Cari yıl değerinin (x_i) temel yıl değerine (x_0) oranıdır.

$$I = \frac{x_i}{x_0} \times 100$$

Örnek: Aşağıda bir şirketin yıllara göre ihracat rakamlarına ilişkin bilgi mevcuttur. 2000 yılını temel (baz) yıl kabul ederek endeks değerlerini hesaplayınız.

Yıllar	İhracat(1000 \$)
2000	350
2001	275
2002	200
2003	375
2004	390
2005	170
2006	280

Yıllar	İhracat(1000 \$)	Endeks (1990=100)
2000	350	$(350/350) \times 100 = 100$
2001	275	$(275/350) \times 100 = 78.57$
2002	200	$(200/350) \times 100 = 57.14$
2003	375	$(375/350) \times 100 = 107.14$
2004	390	$(390/350) \times 100 = 111.43$
2005	170	$(170/350) \times 100 = 48.57$
2006	280	$(280/350) \times 100 = 80$

Soruda temel yıl 2000 senesi olarak belirtildiğinden 2000=100 şeklinde yazılmıştır.

2001 yılında ihracat rakamı 2000 yılına göre % 21 oranında, 2002 yılında ise % 43 oranında azalmıştır. 2003 yılında 2000 yılına göre % 7 oranında artmıştır. 2004 yılının artışı ise % 11 civarındadır. 2006 semesinde ise % 20 oranında azalmıştır.

Endeksler basit ve bileşik, zaman ve mekan şeklinde farklı başlıklar altında incelenebilir.

Mekan Endeksi

Eğer endeksler bir mekan serisinden yola çıkılarak hesaplanıyorsa bu endekslere mekan endeksi denilir. İlgilenilen çeşitli değerlerin şehirler, ilçeler, ülkeler içindeki oransal değişimine mekan endeksi ile ulaşılır. Mekan endeksinde temel alınan değer serideki tüm değerlerin aritmetik ortalamasından yola çıkılarak bulunur.

$$I = \frac{x_i}{\bar{x}} \times 100$$

Örnek: Aşağıda 5 farklı il ile ait sinema bileti satış fiyatları verilmiştir. Mekan endeksini hesaplayarak yorumlayınız.

İller	Fiyatlar
İstanbul	20
Ankara	15
Bursa	12
Kayseri	10
Niğde	8

Serinin aritmetik ortalaması hesaplanır.

$$X_{ort}=(20+15+12+10+8)/5=13$$

Endeks
$(20/13) \times 100 = 153.85$
$(15/13) \times 100 = 115.38$
$(12/13) \times 100 = 92.31$
$(10/13) \times 100 = 76.92$
$(8/13) \times 100 = 61.54$

Buna göre sözkonusu 5 ildeki ortalama bilet fiyatı dikkate alındığında, İstanbul ve Ankara'da ortalama üzerindedir. En ucuz biletse Niğde ilindedir.

ZAMAN ENDEKSİ

İlgilenilen değerlerin yıl, ay, gün gibi zaman içindeki oransal değişimleri zaman endeksini verir. Sabit ve Değişken Esaslı olmak üzere ikiye ayrılır.

Sabit Esaslı Endeks (SEİ)

Sabit esaslı endeksler hesaplanırken, devrelerden birisi temel olarak kabul edilir ve kalan diğer tüm devrelerin temel devreye göre oranı dikkate alınır.

$$I = \frac{x_i}{x_0} \times 100$$

Örnek:

Bir tatil beldesinde Ağustos ayındaki nüfus yıllara göre aşağıdaki gibidir. Sabit Esaslı Endeksi hesaplayarak yorumlayın(2001 yılını temel devre olarak alın).

Yıllar	Nüfus (bin kişi)
2001	1280
2002	1300
2003	1360
2004	1100
2005	1400
2006	1490

Temel devre 2001=100 olarak alınır. Sabit esaslı indeks (SEİ) aşağıdaki gibi hesaplanır.

Yıllar	Nüfus (bin kişi)	SEİ
2001	1280	$(1280/1280) \times 100 = 100$
2002	1300	$(1300/1280) \times 100 = 101.5$
2003	1360	$(1360/1280) \times 100 = 106.3$
2004	1100	$(1100/1280) \times 100 = 85.9$
2005	1400	$(1400/1280) \times 100 = 109.4$
2006	1490	$(1490/1280) \times 100 = 116.4$

Buna göre 2003 yılında 2001 yılına göre % 6.3 'lük bir artış olmuştur. Nüfustaki artış en fazla 2006 tarihindedir %16.4. 2004 yılında 2001 yılına göre yaklaşık olarak % 16 oranında azalma olmuştur.

Değişir Esaslı İndeks (DEİ)

Değişir Esaslı Endeks hesaplanırken, her bir devreye ait değer bir önceki devre değerine oranlanır. Zincirleme endeks olarak da bilinir.

$$I = \frac{x_i}{x_{i-1}} \times 100$$

Örnek:

Önceki soruda yer verilen veriden hareketle Değişir Esaslı Endeks değerlerini hesaplayarak yorumlayın.

Yıllar	Nüfus (bin kişi)	DEİ
2001	1280	(1280/1280)x100=100
2002	1300	(1300/1280)x100=101.5
2003	1360	(1360/1300)x100=104.6
2004	1100	(1100/1360)x100=80.88
2005	1400	(1400/1100)x100=127.3
2006	1490	(1490/1400)x100=106.4

2006 yılındaki nüfus 2005 yılına göre %6.4 oranında artmıştır, 2005 yılında ise 2004 yılına göre %27.3 oranında artış göstermiştir. 2004 yılında ise 2003 yılına göre yaklaşık % 19 oranında azalma söz konusudur.

14.2. BASİT VE BİLEŞİK ENDEKSLER

Endeksler kapsadıkları maddelere göre basit ve bileşik olarak iki sınıfa ayrılırlar.

Hesap esnasında tek bir madde dikkate alınıyorsa Basit endeks, iki ya da daha fazla madde dikkate alınıyorsa Bileşik endeks olarak adlandırılır.

Basit Endeks:

İlgilenilen maddenin fiyatındaki oransal değişim hesaplanacaksa basit fiyat endeksi, miktarındaki değişim hesaplanacaksa basit miktar endeksi adını alır.

“p” fiyata “q” miktara karşılık gelmek üzere Basit miktar ve fiyat endeksleri şöyle hesaplanır:

$$I = \frac{P_i}{P_0} \times 100$$

$$I = \frac{q_i}{q_0} \times 100$$

Örnek:

Bir ürünün fiyatı 2006 yılı fiyatı 132 TL iken 2007 yılındaki fiyatı 167 TL olmuştur. 2007 yılındaki fiyat artış oranını hesaplayınız.

$$I = (167/132) \times 100 = \% 126.5 \text{ artış oranı } \%26.5 \text{’tir.}$$

ÖRNEK:

Bir zeytin üreticisinin 2004 yılı zeytin hasadı 2000 ton, 2005 yılı zeytin hasadı ise 1820 ton’dur. 2004 yılına göre 2005 yılındaki zeytin üretimindeki düşüş oranı nedir?

$$I = (1820/2000) \times 100 = \% 91. \%9 \text{ oranında düşüş yaşanmıştır.}$$

BİLEŞİK ENDEKSLER

İndeksin kapsadığı maddelere için fiyat ya da miktarların zaman içindeki oransal değişimleri dikkate alınır. Üç başlık altında incelenebilir. Bunlar:

-Basit Toplam Endeks

-Basit Endekslerin tartısız aritmetik ortalaması

-Basit Endekslerin tartılı aritmetik ortalaması

Basit Toplam Endeks

Endekse kapsamındaki maddelerin cari devre fiyatları (miktarları) toplanarak temel devre fiyatlar (miktarlar) toplamına oranlanır.

$$I = \frac{\sum p_1}{\sum p_0} \times 100$$

Örnek:

Aşağıda 4 maddeye ilişkin 2004 ve 2005 fiyatları verilmiştir. 2005 yılı basit toplam indeksi hesaplayınız.

	2004	2005
Ürünler	kg fiyatı	kg fiyatı
Kaşar Peyniri	11000	12000
Zeytin	12000	14000
Beyaz Peynir	23000	24000
Tereyağ	18000	18500

Toplam fiyatlar dikkate alınırsa:

	2004	2005
Ürünler	kg fiyatı	kg fiyatı
Kaşar Peyniri	11000	12000
Zeytin	12000	14000
Beyaz Peynir	23000	24000

Tereyağ	18000	18500
Toplam	64000	68500

Basit Endekslerin Tartısız Aritmetik Ortalaması

Cari yıldaki fiyatlar temel devre fiyatlarına bölünür ve daha sonra ortalama alınır.

$$I = (68500/64000) \times 100 = 107$$

Söz konusu 4 ürün ortalama olarak % 7 oranında artmıştır.

$$I = \frac{\sum \left(\frac{p_1}{p_0} \right)}{n} \times 100$$

Örnek: Bir önceki soruda sunulan veriyi kullanarak 4 madde için Basit Endekslerin Tartısız Aritmetik Ortalamasını hesaplayın

	2004	2005
Ürünler	kg fiyatı	kg fiyatı
Kaşar Peyniri	11000	12000
Zeytin	12000	14000
Beyaz Peynir	23000	24000
Tereyağ	18000	18500

Öncelikle her bir dönem için endeksler hesaplanır.

	2004	2005	
Ürünler	p0	p1	p1/p0
Kaşar Peyniri	11000	12000	1,090909
Zeytin	12000	14000	1,166667
Beyaz Peynir	23000	24000	1,043478
Tereyağ	18000	18500	1,027778
Toplam	64000	68500	4,328832

$$I=(4.33/4) \times 100=108$$

Ortalama olarak fiyatlar % 8 oranında artmıştır.

Basit Endekslerin Tartılı Aritmetik Ortalaması

Eğer endeks kapsamındaki maddelerin fiyatlarının bir önemi varsa, endeks hesaplanırken bu farklılıklar dikkate alınarak ağırlıklandırılır. Aşağıdaki formülde tartıları “t” temsil etmektedir.

$$I = \frac{\sum (\frac{p_1}{p_0})_{xt}}{\sum t} \times 100$$

Tartı endeksin çeşidine göre farklılık gösterir. Genel yaklaşıma göre tartılar endeks kapsamındaki maddelerin miktar ve fiyatlarının çarpılmasıyla elde edilir. Tartılar değerlendirilirken temel devre fiyatı baz alınır.

$$t = p_0 \times q$$

Tartı hesaplanırken yukarıda yer alan eşitlikte miktarın (q) temel devre ya da cari değeri alınabilir. Temel devre ve endeks devresi değerlerinin tartıda yer almasına göre iki farklı endeks türü karşımıza çıkar. Bunlar Laspeyres ve Paasche'dir.

14.3. LASPEYRES ENDEKSİ

Tartı hesaplanırken temel devre miktarı alınır.

$$t = p_0 \times q_0$$

Yukarıda belirtilen formülde yerine yerleştirilirse;

$$I = \frac{\sum \left(\frac{p_1}{p_0} \right) x (p_0 x q_0)}{\sum (p_0 x q_0)} \times 100$$

Örnek:

Aşağıda 2004-2005 yıllarına ait üretilen tarım ürünleri ve hasad miktarları miktarları mevcuttur. Temel devreyi 2004 alarak 2005 yılı için Laspeyres fiyat endeksini hesaplayınız.

Ürünler	2004		2005	
	Fiyat TL	Miktar TON	Fiyat TL	Miktar TON
Kivi	420000	60	500000	70
Avakado	670000	70	690000	74
Brokoli	580000	80	750000	90
Muz	900000	100	990000	120

Ürünler	2004		2005		2005	
	P ₀	q ₀	P ₁	q ₁	P ₀ q ₀	P ₁ q ₁
Kivi	42	60	50	70	2520	3000
Avakado	67	70	69	74	4690	4830
Brokoli	58	80	75	90	4640	6000
Muz	90	100	99	120	9000	9900
					20850	23730

$$I = (23730 / 20850) \times 100 = 113.8$$

İncelenen 4 tarım ürününün 2005 yılında fiyatları 2004 yılına göre %31 oranında artmıştır.

14.4. PAASCHE ENDEKSİ

Endeks devresi miktarının tartı olarak kullanılmasıyla elde edilir.

$$t=p_0xq_1$$

Bu tartı formüle yerleştirilirse;

$$\frac{\sum (\frac{p_1}{p_0})x(p_0xq_1)}{\sum (p_0xq_1)} \times 100 = \frac{\sum (p_1xq_1)}{\sum (p_0xq_1)}$$

ÖRNEK:

Önceki örnekte ele alınan veriden hareketle Paasche endeksini hesaplayarak yorumlayın

Ürünler	2004		2005		q	p ₀ q ₁	p ₁ q ₁
	P ₀	Q ₀	P ₁	Q ₁			
Kivi	420000	60	500000	70	70	29400000	35000000
Avakado	670000	70	690000	44	44	49580000	51060000
Brokoli	580000	80	750000	90	90	52200000	67500000
Muz	900000	100	990000	120	120	108000000	118800000
						239180000	272360000

$$I = (239180000 / 208500000) \times 100 = 113.8$$

İncelenen 4 tarım ürününün 2005 yılında fiyatları 2004 yılına göre %13.8 oranında artmıştır.

14.5. FİŞHER ENDEKSİ

Laspeyres ve Paasche endeksleri hesaplanırken farklı ağırlıklandırmalar yüzünden farklı sonuçlara ulaşılır. Paasche endeksi fiyat artışlarını olduğundan az gösterirken, Laspeyres endeksi de fiyat artışlarını olduğundan fazla gösterir. Fisher Endeksi ile Laspeyres ve Paasche endekslerinin geometrik ortalaması alınır, bu yolla gerçeğe daha yakın bir sonuca ulaşılabileceği düşünülür.

$$I = \sqrt{\frac{\sum p_1 x q_0}{\sum p_0 x q_0} \times \frac{\sum p_1 x q_1}{\sum p_0 x q_1}} \times 100$$

Örnek:

Önceki örnekte ele alınan veri için Fisher Endeksini hesaplayarak yorumlayınız.

Ürünler	2004	q ₀	2005	q ₁
	P ₀		p ₁	
Kivi	420000	60	500000	70
Avakado	670000	70	690000	74
Brokoli	580000	80	750000	90
Muz	900000	100	990000	120

Formül için gereken seriler tek tek hesaplanırsa,

p ₀ q ₀	p ₀ q ₁	p ₁ q ₁	p ₁ q ₀	Topla
25200000	29400000	35000000	30000000	
46900000	49580000	51060000	48300000	
46400000	52200000	67500000	60000000	
	10800000	11880000		
90000000	0	0	99000000	
20850000	23918000	27236000	23730000	
0	0	0	0	m

$$I = \sqrt{\frac{\sum p_1 x q_0}{\sum p_0 x q_0} \times \frac{\sum p_1 x q_1}{\sum p_0 x q_1}} \times 100 = \sqrt{\frac{237300000}{208500000} \times \frac{272360000}{239180000}} \times 100 = 106.3$$

2005 yılında fiyatlar 2004 yılına göre %6.3 oranında artmıştır.

Endeksler hesaplanırken temel devrenin seçimi keyfi olmamalıdır. Temel devre seçilirken ekonomi açısından görece durağan/ hareketli olmayan bir dönem temel devre seçimi için uygundur. Bu açıdan, savaş dönemleri, ekonomik kriz dönemleri seçilmemelidir.

Uygulamalar

Uygulama Soruları

Bu Bölümde Ne Öğrendik Özeti

Bölüm Soruları

1)

<i>Ürünler</i>	p₀	Q₀	p₁	q₁
<i>portakal</i>	600	120	700	130
<i>mandalina</i>	700	130	800	130
<i>armut</i>	500	150	500	120
<i>elma</i>	400	200	450	220

Yukarıdaki 4 madde için fisher fiyat endeksini hesaplayınız

2) 2004 yılı baz yılı ise 2006 ve 2008 yılları için sabit esaslı endeksi hesaplayınız.

KAYNAKLAR