İSTATİSTİK II

ORTAK DERS

YRD. DOÇ. DR. LEYLA İŞBİLEN YÜCEL (Bölüm 1-2)

DOÇ. DR.ÖZLEM YORULMAZ (Bölüm 3-14)

İSTANBUL ÜNİVERSİTESİ AÇIK VE UZAKTAN EĞİTİM FAKÜLTESİ

İSTANBUL ÜNİVERSİTESİ AÇIK VE UZAKTAN EĞİTİM FAKÜLTESİ

ORTAK DERS

İSTATİSTİK II

YRD. DOÇ. DR. LEYLA İŞBİLEN YÜCEL (Bölüm 1-2)

DOÇ. DR.ÖZLEM YORULMAZ (Bölüm 3-14)

Yazar Notu

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

ÖNSÖZ

Bu ders günlük hayatta, özellikle ekonomik alanda karşılaşılan çeşitli problemleri istatistiksel metotlarla ele alma, çeşitli iddiaları sınama, çıkarsama yapma, ilişkileri modelleme ve değerlendirme konularında bilgi vermeyi amaçlar. İlk iki bölümde sürekli dağılım, örnekleme kavramı gibi temel konulara değinilmiş, sonraki bölümlerde ise istatistiğin çıkarımsal konularına (aralık tahmini, hipotez testleri, regresyon analizi, ki-kare analizi, endeksler) yer verilmiştir. Bu kitap fazla sayıda örnek uygulamalarıyla, istatistiksel kavramları ve teknikleri anlamayı kolaylaştırmak hedefiyle derlenmiştir.

DOÇ. DR.ÖZLEM YORULMAZ

İÇİNDEKİLER

ÖNSÖZ.....I
KISALTMALAR V
YAZAR NOTU VI
1. SÜREKLİ DAĞILIMLAR - NORMAL DAĞILIM..... 1
1.1. Sürekli Dağılımlar 7
1.2. Sürekli Bir Fonksiyonun Olasılık Yoğunluk Fonksiyonu Olabilmesi İçin Gerekli Şartlar 8
1.3. Normal Dağılım..... 8
1.3.1. Normal Dağılımın Olasılık Yoğunluk Fonksiyonu..... 9
1.4. Z tablosunun (normal dağılım tablosu) özellikleri ve okunuşu..... 12
2. ÖRNEKLEME TEORİSİ 22
2.1. Örnekleme 28
2.2. Örneklemede Temel Kavramlar 29
2.3. Örnekleme Yapmayı Gerektiren Nedenler Nelerdir?..... 35
2.4. Basit Tesadüfi Örnekleme 36
2.5. Örnekleme Dağılımı 41
2.6. Kitle Ortalaması μ'nün Tahmini 41
2.7. Örneklemede Hata Kavramı ve Standart Hata 44
2.8. Oranların Örnekleme Dağılımı..... 45
2.9. Örneklem Hacminin Belirlenmesi..... 46
3. TAHMİN TEORİSİ VE GÜVEN ARALIĞI..... 52
3.1. İstatisitiksel Çıkarsama 58
3.2. Nokta ve Aralık Tahmini..... 59
3.3. Anakütle Varyansının Bilindiği Durumlarda Anakütle Ortalamasının Güven Aralığı:.... 59
3.4. Anakütle Varyansının Bilinmediği Durumlarda Anakütle Ortalamasının Güven Aralığı 62
3.5. Anakütle Oranının Tahmini 63
4. GÜVEN ARALIKLARI: İKİ ANAKÜTLE 74
4.1. Ortalamalar Arası Farklar İçin Güven Aralığı 80
4.2. İki Oran Farkı İçin Güven Aralığı..... 81
4.3. Örneklem Büyüklüğünün Belirlenmesi ve Hata 85
5. HİPOTEZ TESTLERİ..... 92
5.1. Hipotez Testleri 98
5.2. Hipotezlerin İfade Edilmesi 98

5.3. I. Tür ve II. Tür Hata 100
5.4. Anakütle Ortalamasına İlişkin Hipotez Testleri 101
6. HİPOTEZ TESTLERİ (İKİ ANAKÜTLE) 113
6.1. Ortalamalar Arası Farklara İlişkin Testler 119
6.2. Oranlararası Farklara İlişkin Hipotez Testleri 120
7. GENEL TEKRAR 128
8. REGRESYON ANALİZİ I 136
8.1. Regresyon Analizi 142
8.2. Regresyon Katsayılarının Tahmini 144
9. REGRESYON ANALİZİ II 157
9.1. Regresyon Doğrusunun Özellikleri ve Hata (Artık) Kavramı 163
9.2. Regresyon Denklemi Parametrelenin Hipotez Testi ve Güven Aralığı 165
10. KORELASYON KATSAYISI-DETERMİNASYON KATSAYISI 178
10.1. Regresyon Denkleminin Gücü 184
10.2. Korelasyon Katsayısı 187
10.2.1. Pearson Korelasyon Katsayısı 189
11. Kİ-KARE TESTLERİ 203
11.1. Ki-Kare Testlerini Neden Kullanırız? 208
11.2. Ki-Kare Bağımsızlık Testi 209
11.3. Ki-Kare Homojenlik Testi 212
11.4. Ki-Kare Uyum Testi 213
11.5. Kontenjans Katsayısı 214
12. VARYANS ANALİZİ TABLOSU 222
12.1. VARYANS ANALİZİ (ANOVA) 228
12.2. Tek Yönlü Anova 228
13. ENDEKS KAVRAMI 241
13.1. Endeks (İndeks) 247
13.2. Basit ve Bileşik Endeksler 251
13.3. Laspeyres Endeksi 255
13.4. Paasche Endeksi 256
13.5. Fisher Endeksi 257
14. GENEL TEKRAR 263
14.1. Çözümlü Örnekler 266

KAYNAKÇA 280

KISALTMALAR

- r.d: Rastlantı değişkeni
- N(0,1) : Ortalaması 0 ve varyansı 1 olan normal dağılım
- BTO:Basit Tesadüfi Örnekleme
- ANOVA: Varyans analizi

YAZAR NOTU

Bu kitap sosyal bilimler öğrencilerinin konuları anlamalarını zorlaştıracak teorik açıklamalara yer vermek yerine olabildiğince istatistiğin uygulama yönünü ön plana çıkarmıştır. 7. ve 14. bölümlerde önceki bölümlerde anlatılan yöntemlerin kendi içinde yeniden değerlendirmesi yapılmış ve çeşitli örnek çözümleri ile tekrarlara yer verilmiştir.

1. SÜREKLİ DAĞILIMLAR - NORMAL DAĞILIM

Bu Bölümde Neler Öğreneceğiz?

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

1) Normal dağılım neden en sık kullanılan dağılımdır?

2) 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

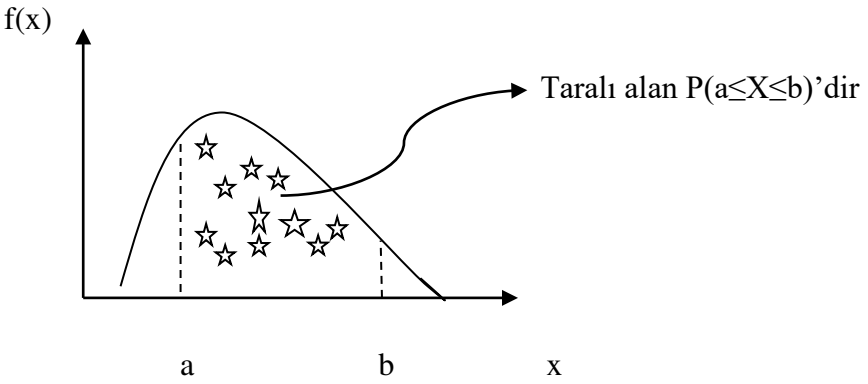
Giriş

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

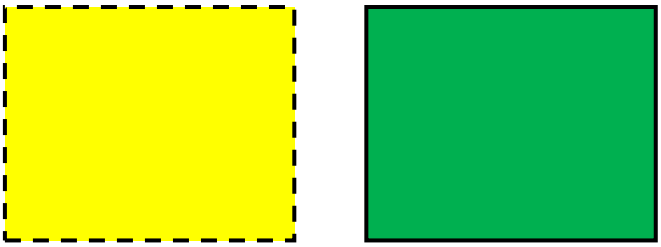
1.1. Sürekli Dağılımlar

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≤X≤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≤X≤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.

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

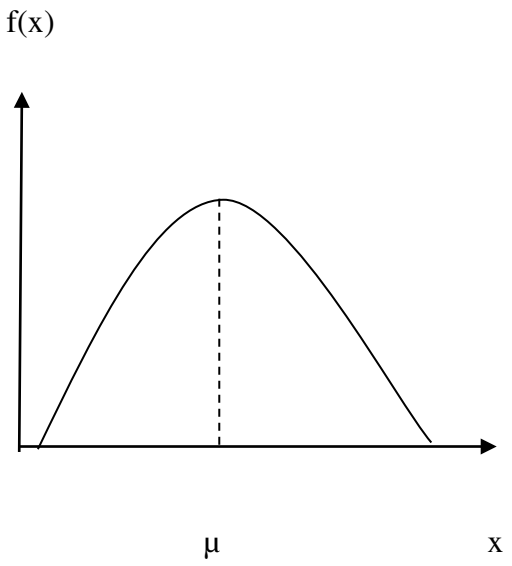
1.2. Sürekli Bir Fonksiyonun Olasılık Yoğunluk Fonksiyonu Olabilmesi İçin Gerekli Şartlar

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.

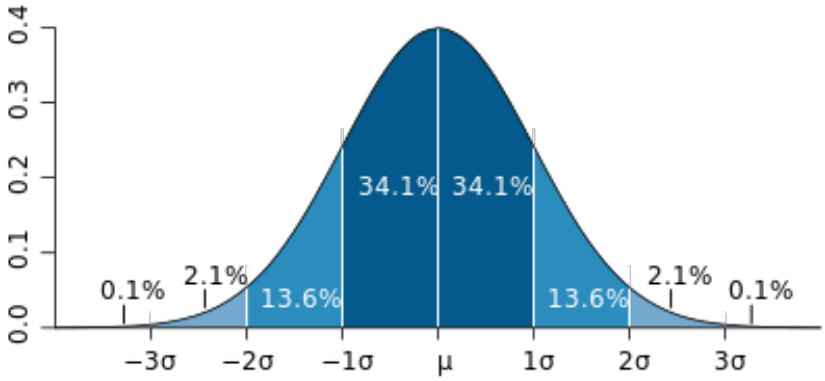
1.3. Normal Dağılım

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'nün 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:

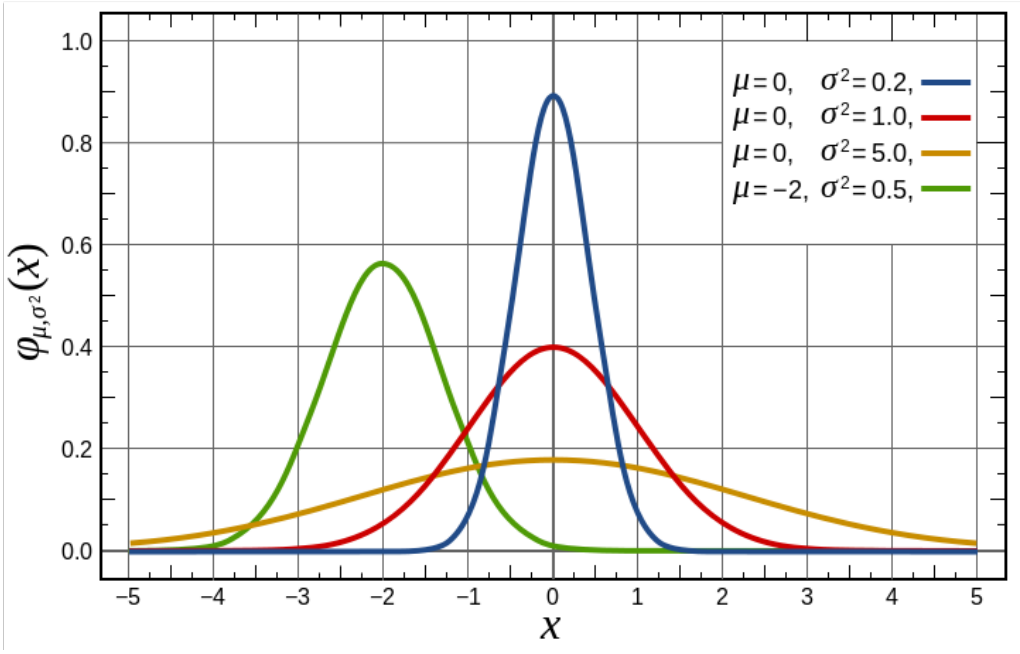


Verilerin %68.2'si (bu değer 2*34.1'den elde edilmiştir) ortalamanın 1 standart sapma dolayında bulunur. %95.4'ü (bu değer 2*47.7'den elde edilmiştir) ortalamanın 2 standart sapma dolayında bulunur. %99.6'sı (bu değer 2*49.8'den elde edilmiştir) ortalamanın 3 standart sapma dolayında bulunur.

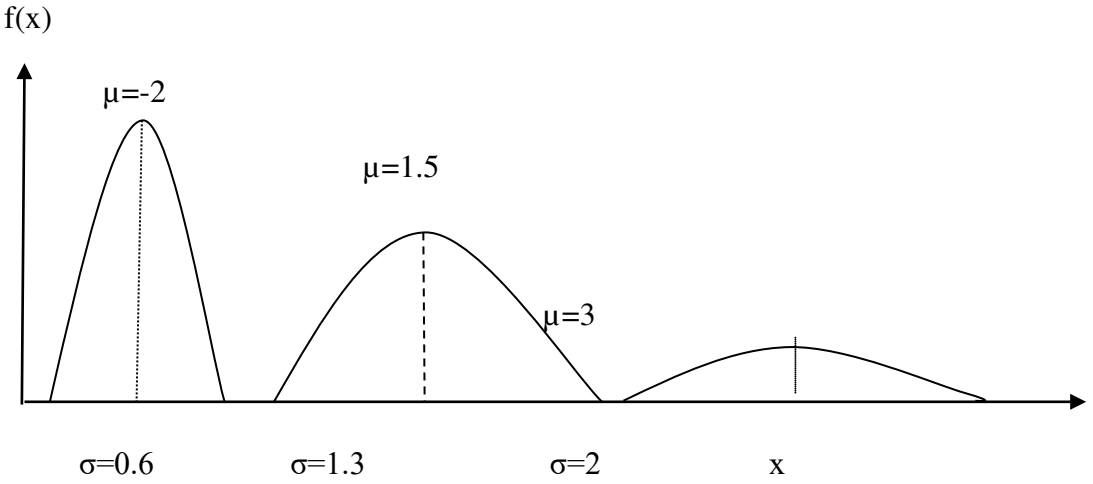
1.3.1. Normal Dağılımın Olasılık Yoğunluk Fonksiyonu

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

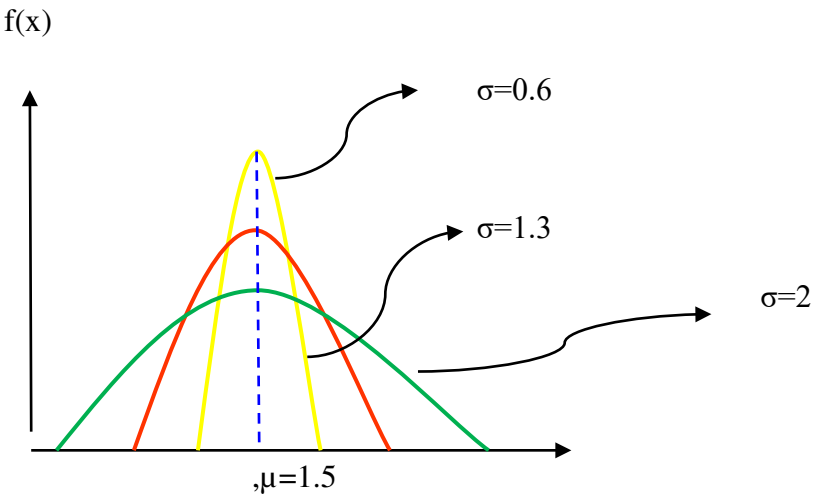
Burada e=2.718, π=3.14, σ: standart sapmadır.



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)= $\int_2^5 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx$ '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 saptırılarak standart sapmasına bölünmesiyle elde edilmektedir. Standart normal dağılım, normal dağılımla aynı özelliklere sahiptir, ancak, ortalaması sıfır, varyansı 1'dir.

X~Normal(μ,σ^2) iken; yapılan $z = \dfrac{X-\mu}{\sigma}$ dönüşümüyle elde edilen z standart değişkeni; z~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~Normal(μ,σ^2)
$f(x) = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$

$$z = \frac{X-\mu}{\sigma}$$

dönüşümü yapılarak

z~Standart Normal(0,1)
$f(z) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z)^2}$

Örnek 1.1: X~Normal(1,4) olmak üzere, P(2<X<3)=?

1.çözüm:

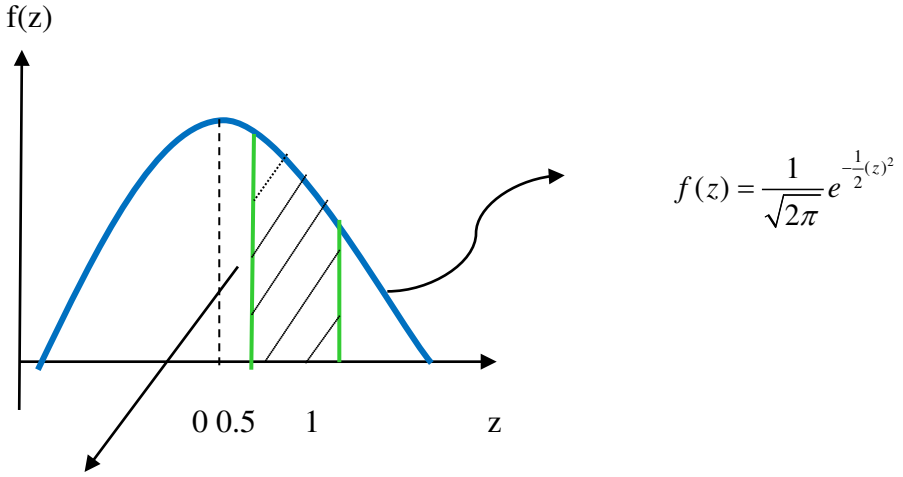
$$\int_2^3 \frac{1}{\sqrt{2\pi}\,2} e^{-\frac{1}{2}(\frac{x-1}{2})^2} dx$$

2.çözüm:

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

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

P(2<X<3)= $\int_2^3 \frac{1}{\sqrt{2\pi}\,2} e^{-\frac{1}{2}(\frac{x-1}{2})^2} dx$ =P(0.5<z<1) yani; standart normal dağılım eğrisinin altında (0.5,1) arasındaki alandır. Bu durum aşağıda görülmektedir:

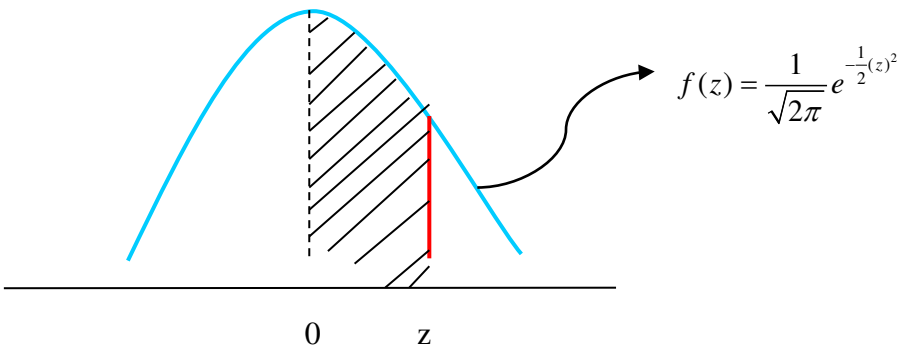


z tablosuna bakıldığında (0,0.5) arasındaki alanın büyüklüğü yani P(0<x< 0.5)=0.1915'tir. P(0<x<1)=0.3413'tür. O halde P(0.5<x<1) olasılığını bulmak için P(0<x<1)'den P(0<x< 0.5) çıkarılmalıdır.

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

1.4. 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:

z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	…	…	…	…	…	…	0.0753
0.2	0.0793	…	…	…	…	…	…	…	…	0.1141
0.3	0.1179	…	…	…	…	…	…	…	…	…
0.4	…	…	…	…	…	…	…	…	…	…
0.5	…	…	…	…	…	…	…	…	…	…
0.6	…	…	…	…	…	…	…	…	…	…
0.7	…	…	…	…	…	…	…	…	…	…
0.8	…	…	…	…	…	…	…	…	…	…
0.9	…	…	…	…	…	…	…	…	…	0.3389
1.0	…	…	…	…	…	…	…	…	…	0.3621
1.1	…	…	…	…	…	…	…	…	…	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…

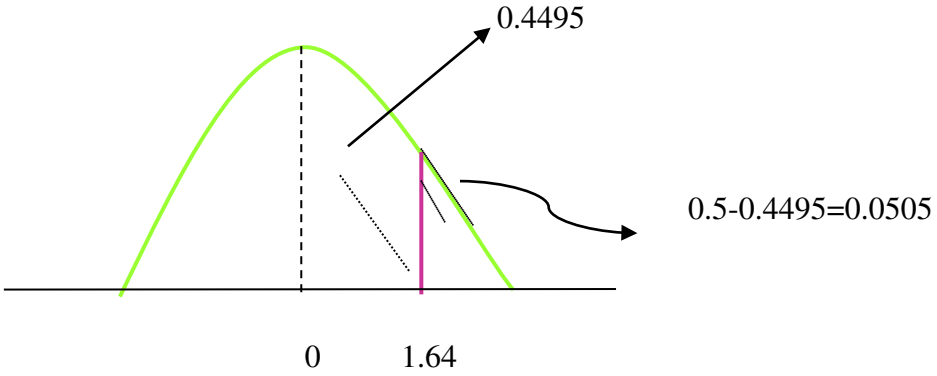
…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…
2.5	0.4938	0.4940	0.4941	…	…	…	…	…	…	0.4952
2.6	…	…	…	…	…	…	…	…	…	0.4964
2.7	…	…	…	…	…	…	…	…	…	0.4974
2.8	…	…	…	…	…	…	…	…	0.4980	0.4981
2.9	…	…	…	…	…	…	…	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

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.

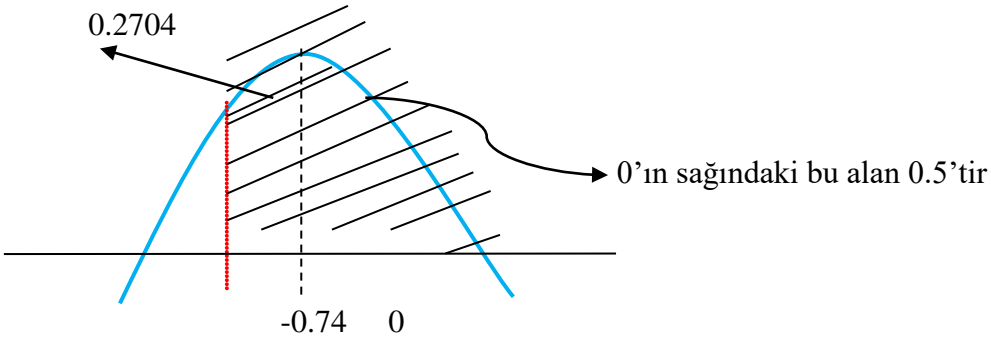
Örnek 1.2 : P(-1.33<z<1.33)=?

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



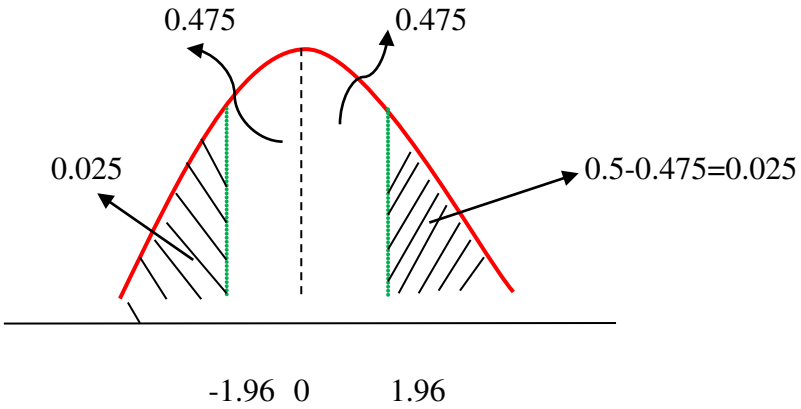
Örnek 1.4: P(z>-0.74)=?



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

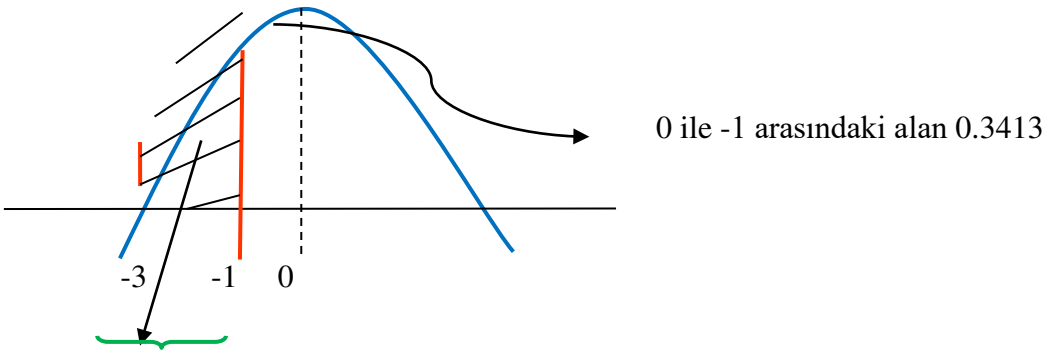
Örnek 1.5: P(│z│>1.96)=?

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



P(│z│>1.96)= P(z<-1.96)+P(z>1.96)=2x0.025=0.05'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.

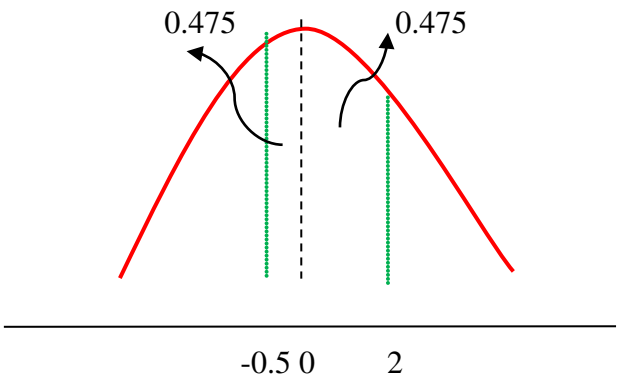
Örnek 1.7: X~Normal(12,16) veriliyor. P(10<X≤20)=?

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

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

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

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

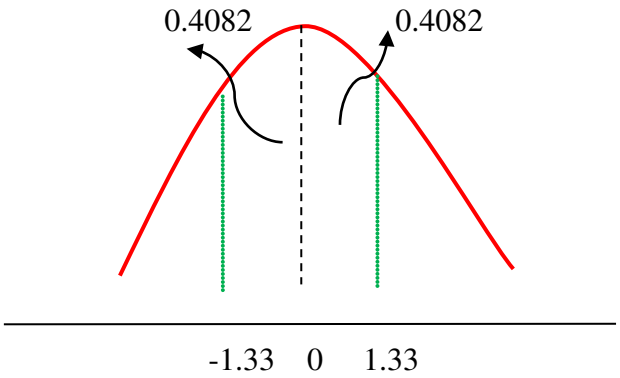


Örnek1.8: X~Normal(100,225) veriliyor. P(80<X<120)=?

μ=100

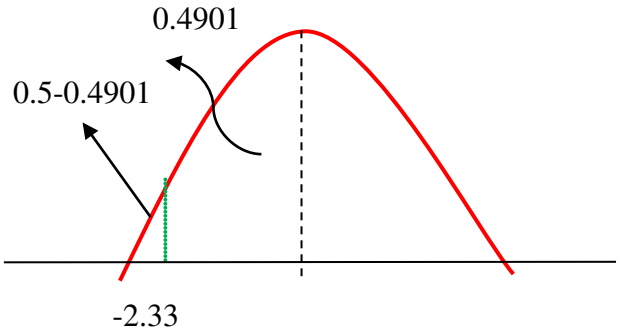
σ=15

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

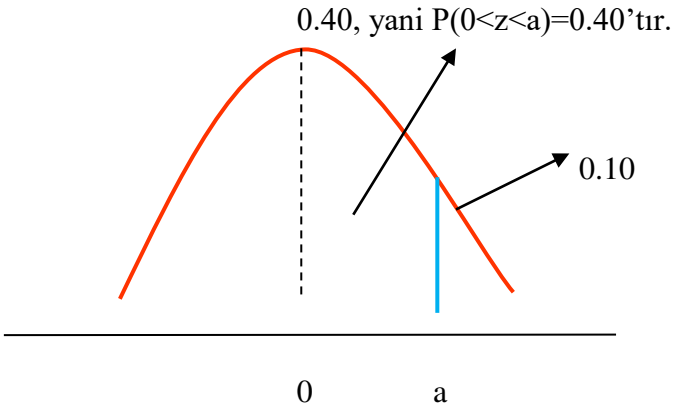


Örnek 1.9: X~Normal(27,9) veriliyor. P(X<20)=?

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



Örnek 1.10: P(z>a)=0.10 ise a=?



Z tablosuna bakarak 0.40 olasılığına denk gelen z değerini bulacağız. Tabloya bakıyoruz, fakat tamı tamına 0.40 değeri bulunmamakta. Böyle durumlarda 0.40'ı kapsayan en dar aralığı göz önüne alarak, *interpolasyon* denilen tekniği uygulamak suretiyle, 0.40'a denk gelen z tablo değerini 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'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

Bu Bölümde Ne Öğrendik Özeti

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

BÖLÜM 1 SORULARI

A) X~Normal(50,9) olmak üzere;

1) P(X<a)=0.8413 ise a değeri aşağıdakilerden hangisine eşittir?

a) -1 b)-0.5 c) 0.25 d)0.5 e)1

2) P(X>a)=0.025 ise a değeri aşağıdakilerden hangisine eşittir?

a) -1.96 b)-1.65 c) 1.65 d)1.96 e)2.58

3) P(X>a)=0.95 ise a değeri aşağıdakilerden hangisine eşittir?

a) -1.96 b)-1.65 c) 1.65 d) 1.96 e) 2.58

4) P(41<X<a)=0.7987 ise a değeri aşağıdakilerden hangisine eşittir?

a) 52.551 b)53.282 c) 58.744 d) 59.789 e) 60.121

5) P(X>a)=0.01 ise a değeri aşağıdakilerden hangisine eşittir?

a) 2.58 b) 2.33 c)1.96 d)1.65 e)hiçbiri

B) X~Normal(11,4) olmak üzere;

6) P(10<X<12) olasılık değeri aşağıdakilerden hangisine eşittir?

a) 0.08 b)0.187 c)0.383 d)0.838 e)0.883

7) P(6<X<10) olasılık değeri aşağıdakilerden hangisine eşittir?

a)0.3023 b) 0.2223 c) 0.5023 d)0.8023 e)8542

8) P(13<X<16) olasılık değeri aşağıdakilerden hangisine eşittir?

a) 0.052 b)0.1525 c)0.2525 d)0.3525 e)0.7525

9) P(7.8<X<12.6) olasılık değeri aşağıdakilerden hangisine eşittir?

a)0.7373 b)0.5353 c)0.2323 d)0.2853 e) 0.1743

10) P(X>13.24) olasılık değeri aşağıdakilerden hangisine eşittir?

a) 0.1315 b)0.1785 c)0.2461 d)0.4651 e)0.6315

Cevaplar:

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

2. ÖRNEKLEME TEORİSİ

Bu Bölümde Neler Öğreneceğiz?

Bu bölümde “Örneklemenin temel kavramları nelerdir?”, “Örnekleme yapmayı gerektiren nedenler nelerdir?”, “ Örneklemeye neden ihtiyaç duyulmuştur?”, “Başlıca örnekleme teknikleri nelerdir?” sorularına yanıt vereceğiz.

Bölüm Hakkında İlgi Oluşturan Sorular

1) Örnekleme neden gereklidir? Örnekleme tekniklerini uygulamak araştırmacıya ne tür kolaylıklar sağlar?

2) Örnekleme yapılırken hata ile karşılaşılır mı? Kaç tür hata vardır?

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Örnekleme teorisi, örnekleme yöntemleri, örnekleme dağılımları	Örnekleme yaklaşımının istatistiksel çıkarsama için gerekliliği, örnekleme yöntemleri gibi kavramlar hakkında bilgi edinilir.	Metinler, çözümlü problemler

Anahtar Kavramlar

- Örneklem
- Örnekleme
- Örnekleme Dağılımı
- Basit Tesadüfi Örnekleme
- Standart Hata

Giriş

Bu bölümde örnekleme teorisinin istatistiksel çıkarsama için önemi, örnekleme teknikleri ve örnekleme dağılımları hakkında bilgi verilecektir.

2.1. Örnekleme

Üzerinde araştırma yapılacak bir kitle hakkında bilgi derlemek için başvurulabilecek ilk ve en basit yöntem, o kitleyi oluşturan tüm gözlem birimleri üzerinde araştırma konusuyla ilgili bilgi toplamaktır. Bu durum tam sayım olarak adlandırılır. Ancak uygulamada genellikle kitlenin tüm birimlerinden bilgi derlemeye kalkışmak çeşitli nedenlerden dolayı imkânsızdır. Bunun en temel nedenleri tamsayım yönteminin genellikle pahalı ya da zaman alıcı olmasıdır.

Örneklem kitleyi en iyi şekilde temsil ettiği düşünülen küçük bir alt kümedir. Örnekleme dayanarak kitle hakkında çıkarsamalar yapılır. Örneğin Ekonometri Bölümü'nde okuyan öğrencilerin boy uzunluklarının ortalamasını hesaplamak istediğimizde, tüm öğrencilerin boyunu ölçmek yerine, her sınıftan belli bir sayıda söz gelimi 10'ar öğrencinin boyunu ölçmek suretiyle, ortalama boy uzunluğunu yaklaşık olarak hesaplayabiliriz.

Bazı durumlarda ise tam sayım yapmak, fiziksel zarara uğratabilir. Üretilen ürünün mahiyeti, örnekleme yapılmasını kaçınılmaz kılmaktadır. Örneğin bir mermi fabrikasında defolu oranını tahmin etmek istersek ve tam sayım yaparsak, sayım sonucunda elimizde boş kovanlar (patlayan yani sağlam olan mermiler) ve patlamayan (defolu) mermiler kalacaktır.

Örnekleme 4 adımda uygulanır:

1) Planlama

2) Veri toplama

3) Verilerin çözümlenmesi ve analizi

4) Sonuçların raporlanması ve sunumu

1) Planlama: Öncelikle araştırmaya konu olacak kitle belirlenir. Çerçeve oluşturulur. Gözlem tekniği belirlenir. Soru (anket) kâğıtları tasarlanır. Kullanılacak örnekleme tekniğine karar verilir. İstatistiksel tahmin ve karar alma yöntemleri belirlenir. Kesinlik, güvenilirlik ve maliyet göz önüne alınarak gerekli olan örneklem büyüklüğü (hacmi) hesaplanır.

2) Veri toplama: Örneklem birimleri seçilir ve gözlem birimlerinden ayrı ayrı bilgi toplanır.

3) Verilerin çözümlenmesi ve analizi: Örneklem istatistikleri hesaplanır. Kitle parametreleri tahmin edilir ve bunlar hakkında hipotez testleri yapılır. Genellikle ortalama, oran, varyans, v.b. şeyler hesaplanır.

Örnekleme geniş bir uygulama alanına sahiptir; kamu oyu araştırmaları, pazarlama araştırmaları, sanayi işletmelerinde kalite kontrol süreçlerinde, gelir dağılımı ölçümlerinde, eğitim konularında, demografik bilgi toplamada, seçim tahminlerinde, v.b.

2.2. Örneklemede Temel Kavramlar

Birim:

Araştırmaya konu olan olayların her birine *"birim"* adı verilir. Örneğin nüfus sayımı yapılırken çalışmanın birimi insandır. Aileler hakkında yapılan araştırmada her aile, doktorlar hakkında yapılacak bir araştırmada doktorlar birimlerdir. Buna karşılık doğum ve kazalar üzerinde bir araştırma yapılacak olsa, doğum ve kaza olaylarının her biri birer birim sayılır.

Kitle:

Hakkında bilgi edinilmek istenen ve homojen (aynı ya da benzer) birimlerden oluşan topluluktur. Kitle içerdiği birim sayısı bakımından sonlu kitle ve sonsuz kitle olarak 2 tür olabilir. Sonlu sayıda ya da sayılabilir çoklukta birim içeren evrenlere sonlu kitle denir. Örneğin bir ilçede liseye giden öğrenciler, seçmenler, öğretmenlerin kitlesi sonludur. Genellikle sonsuz kitle bir süreçle ilgilidir ve onun birimleri bu sürecin çıktılarıdır. Süreç devamlı surette aynı koşullar altında işler. Bir fabrikada üretilen ampullerin dayanma süresini belirlemek istesek, ampul üretim süreci, birimleri üretilen ampuller olan sonsuz kitleyi oluşturur.

Değişken:

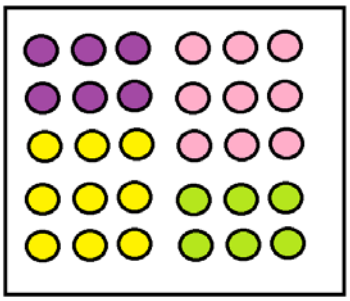
Araştırmalarda birimlerin ilgilenilen özelliklerine **değişken** adı verilir. Bu özellikler kitlede birimden birime farklı değerler alır. Ankara'daki ailelerinin, aylık gelir düzeyleri, bir araba lastiğinin kaydırmazlık özelliği, bir insanın sahip olduğu ayakkabıların sayısı, v.b. birimden birime farklılık göstermekte olan değişkendir. Değişkenler nicel ve nitel değişken şeklinde sınıflandırılırlar. Nicel değişkenler sayarak veya ölçerek ifade edilirler. Nitel değişken ise sayarak ya da ölçerek ifade edilmezler. Bazı değişkenler hem nitel hem de nicel sınıfında yer alabilir. Öğrenci başarısıyla ilgili bir araştırmada öğrenci başarı notu 10 üzerinden değerlendirmede nicel değişkendir. Bu değişken 4, 8, 10 gibi ortaya çıkar. Eğer başarı notu değişkeni geçmez, orta, iyi, pekiyi şeklinde ifade edilirse nitel değişken sınıfında yer alır. Örneğin bir ordudaki rütbeler nitel değişkendir. Göz rengi, saç rengi, memleket v.b. kavramlar nitel değişkenlerdir. Bu gibi değişkenler istatistiksel araştırmaya konu olacakları zaman sayısal olarak kodlama yapılarak çalışmaya dâhil edilirler. Örneğin; onbaşı – 1, çavuş – 2, uzman onbaşı – 3, kademeli uzman onbaşı – 4, uzman çavuş – 5, v.b.

Örneklem:

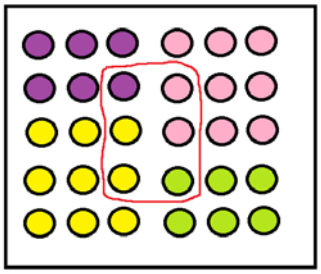
Gözlemlenmek üzere kitleden seçilen birimlerden oluşan alt kümelerin her birine örneklem denir. Örneklemin başlıca özelliği, kitleyi en iyi şekilde temsil ediyor olması gerekliliğidir. Örneklemdeki birimlerin sayısına örneklem hacmi denir ve n simgesiyle gösterilir; n< N dir. Diğer taraftan, n/N oranına örneklem oranı denir. Üzerinde çalışılan kitle hakkında yorum yapma amacıyla kitlenin seçilen bir alt parçasına örneklem adı verilir.

Şimdi, iyi bir örneklem yani kitleyi en iyi şekilde temsil eden örneklem nasıl olmalıdır aşağıdaki şekle bakarak ifade etmeye çalışalım. Kitlemiz 30 toptan oluşsun ve biz rastgele 6 top çekelim. Yani N=30, n=6 büyüklüğünde örneklem çekeceğiz. Bilindiği üzere $C_6^{30} = \dfrac{30!}{6!24!}$ sayıda farklı örneklem çekilebilir. Ama biz burada, iyi bir örneklemin ve kötü yani kitleyi temsil etmekten uzak bir örneklemin ne demek olduğunu anlatmaya çalışacağız.

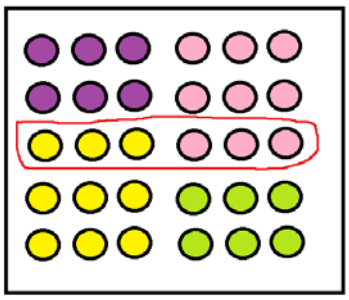
N=30 Kitlemiz



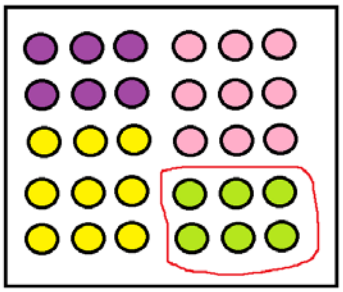
n=6 iyi (kitleyi iyi temsil eden) bir örneklem



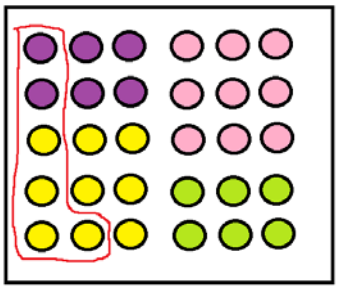
n=6 kötü (kitleyi iyi temsil edemeyen) bir örneklem



n=6 kötü (kitleyi iyi temsil edemeyen) bir örneklem



n=6 kötü (kitleyi iyi temsil edemeyen) bir örneklem



Temsili olmayan kötü bir örneklem daha…

Temsili olmayan kötü bir örneklem daha…

Örnekleme oranı: $\frac{n}{N}$ dir.

Çerçeve:

Örneklem birimlerinin listesine çerçeve denir. Örnekleme yapabilmek için örnekleme birimlerini gösteren bir listenin mevcut olması veya sıfırdan araştırmanın konusuna uygun olarak hazırlanması gerekir. Çerçevenin belli başlı örnekleri; Örneğin adres listesi, telefon rehberi, seçmen kütükleri, nüfus kayıtları, sendika kayıtları, vergi mükelleflerinin listesi, tapu ve ticaret sicilleri, dernek veya vakıfların üye kayıt listeleri vb. Bir çerçeve olmadan ne örnekleme ne de tam sayım yapılabilir. Araştırmaya başlamadan önce; amaç için uygun bir çerçeve var mı? Başka bir kaynaktan sağlanabilecek mi? Önceden düzenlenmiş bir çerçeve yoksa yeni bir çerçeve hazırlanabilir mi? Yeni çerçevenin hazırlanmasında gerekli olacak maliyet ve zamana sahip miyiz? vb. sorularının yanıtlanması gerekir. Çerçeve mümkünse kitlenin bütün birimleri ya da yapılacak içerik (kapsama) hatası sakıncalı sayılmadığı takdirde, büyük bir kısmını kapsamalıdır. Örneklemeye başlamadan önce kitlenin sınırlarını belirlemek için çerçeveyi oluşturmak gerekir. Çerçeve, kitledeki tüm birimleri bir kez içerir.

Örneklem birimi:

Örneklemde yer alan birimlere denir. Örnekleme birimi türü belirlenirken göz önünde tutulması gereken ilkeler şunlardır: Örnekleme birimi tanımlanabilir ve sınıflandırılabilir olmalıdır. Çerçeve maliyetini yükseltmemeli, kitlede birimler arası değişimi yani kitle varyansını yükseltmemelidir.

Gözlem birimi:

Örnekleme birimlerine ait ölçümlerin yapıldığı birimlere gözlem birimi denir. Aileleri örnekleme birimi olarak alan bir araştırmada, aile bireylerine ait bilgiler, örneğin, hane reisinden (anne veya babadan) elde edilebilir. Bu durumda gözlem birimi hane reisi olacaktır. Okul öncesi çocuklar üzerinde yapılacak araştırmada bilgiler her bir okul öncesi çocuktan alınacaksa her bir okul öncesi çocuk örnekleme birimi, aynı zamanda gözlem birimidir. Bu araştırma için gerekli bilgiler çocuklardan değil de onların annelerinden elde edilecekse, gözlem birimi anne olacaktır. Özetle, gözlem birimi, hakkında ayrı ayrı bilgi toplanan, kitlenin en küçük parçasıdır. Örnekleme birimi ile gözlem birimi aynı olabileceği gibi farklı da olabilirler. Örneğin haneler örnekleme birimi, hane halkı gözlem birimidir. Bazen hane halkı (veya haneler) araştırmanın mahiyetine göre hem örnekleme hem de gözlem birimi olabilir.

Parametre:

Kitleler parametre adı verilen sayısal belirleyici ölçülerle ifade edilirler. Yani parametre kitleye ilişkin ölçümler anlamına gelmektedir. Bu ölçümler çok çeşitli olabilir. Kitlenin aritmetik ortalaması (μ), kitle varyansı (σ^2), en temel kitle karakteristikleridir. Kitlenin özelliklerini tanımlayan karakteristiklere **parametre** adı verilir.

Parametre genel olarak θ simgesiyle gösterilir. θ özellikle evren aritmetik ortalaması μ, evren varyansı σ^2, kitle oranı Π, iki aritmetik ortalama arasındaki fark μ_1 - μ_2, iki oran arasındaki fark Π_1-Π_2 v.b. gibi isimleri alır. Parametre değerleri kitledeki tüm birimlerden elde edilen bilgileri kullanarak parametre adı verilen karakteristiklerle hesaplanır.

İstatistik:

Örneklemler istatistik adı verilen belirleyici ölçülerle tanımlanırlar. Bir örneklemin özelliklerini tanımlayan belirleyicilere örneklem istatistiği ya da sadece istatistik adı verilir. Bir örneklemi tanımlayan çeşitli istatistikler mevcuttur; örneklem aritmetik ortalaması $\overline{X}$, örneklem varyansı s^2 bunların başlıcalarıdır.

-Örnekleme sürecinin dört adımı vardır:

a)Araştırılacak kitleyi tanımlamak

b)Örneklem çerçevesini belirlemek

c)Örnek büyüklüğünü belirlemek

d) Uygun örnekleme tekniğini kullanarak, örneklemi seçmek

-Örnekleme yöntemleri tesadüfi olan ve tesadüfi olmayan yöntemler şeklinde iki başlık altında incelenir.

-Tesadüfi örnekleme yönteminde tüm birimlerin örnekleme girme şansları eşittir. Tesadüfi olmayan yöntemler şu şekilde incelenir:

a)Basit Tesadüfi Örnekleme

b)Tabakalı Örnekleme

c)Küme Örneklemesi

d)Çok aşamalı Örnekleme

-Tesadüfi olmayan örnekleme yöntemlerinde örnekleme girecek tüm birimlere eşit şans verilmez. Şu şekilde sıralanır:

a)Kolayda örnekleme

b)Amaçlı örnekleme

c)Kota örneklemesi

d)Kartopu örneklemesi

Burada basit tesadüfi örnekleme yöntemi ayrıntılı olarak incelenecektir. Ancak diğer örnekleme yöntemleri kısaca şöyledir:

-Basit Tesadüfi Örnekleme

Kitledeki tüm birimler listelenir ve seçimde eşit şansa sahip olan birimler bu listeden rastgele seçilir. Kitlenin çok büyük olması, örneklem çerçevesine ihtiyaç olması, birimlerin dağınık olması halinde bu yöntemi uygulamak zordur.

-Tabakalı Örnekleme

Eğer örnekleme çerçevesindeki birimler, incelenen özelliğe göre önemli farklılıklar gösterirse birimler homojen gruplara ayrılır ve bu gruplara katman adı verilir. Örnekleme sürecinde ise her birinden rastgele seçimle örnek alınır ve bu örnekler birleştirilir. Tabakalı örneklemde sabit ya da orantılı örnekleme oranı kullanılır.

-Küme Örneklemesi:

Bu yaklaşımda örneklem için tek tek birimler yerine gruplar (kümeler) birim gibi kabul edilerek seçilir. Örneklem çerçevesine ihtiyaç yoktur. Çeşitli sosyo-ekonomik-demografik özellikler dikkate alınarak kümeler belirlenir. Tanımlanan kümeler arasından rastgele seçimle kümeler belirlenir, kümelerdeki eleman sayısı çok ise her kümeden rastgele seçimle yeniden seçim yapılır.

-Kota Örneklemesi:

Tesadüfi olmayan örnekleme yöntemlerindendir. İncelenen nitelik açısından örneklem çerçevesi gruplara ayrılıyorsa tercih edilir

-Kolayda (Gelişigüzel) Örnekleme:

Belirlenen örneklem büyüklüğüne göre herhangi bir şekilde kitlenin bir parçası seçilir. Tesadüfi olmayan örnekleme yöntemidir.

-Amaçlı Örnekleme:

Araştırmanın amacı doğrultusunda kitlenin temsilci bir örneği yerine, amaçlı olarak bir kesimini örnek olarak almaktır.

-Kartopu Örnekleme:

Öncelikle kitleye ait bir birimle temas kurulur, bu birimin yardımı ile ikinci birime ve ikinci birimin yardımı ile de üçüncü birime ulaşılır. Bu şekilde sanki bir kartopunun büyümesi gibi örneklem büyüklüğü genişler.

2.3. Örnekleme Yapmayı Gerektiren Nedenler Nelerdir?

Üzerinde araştırma yapılacak olan kitle sonsuz olduğunda, onunla ilgili bilgi sadece bir örneklem üzerinden elde edilebilir. Eğer evren sonlu evren ise, N'in büyüklüğüne bağlı olarak, bilgi ya tam sayım yaparak ya da bu kitleyi temsil edecek bir örneklemden elde edilebilir. Ancak, aşağıda belirtilen nedenlerden dolayı örnekleme tam sayıma göre daha çok tercih edilir.

- **Maliyet kısıtı:** Kitle hacmi ve incelenecek özellik sayısı arttıkça tam sayım ekonomik olmaktan çıkar. Örnekleme ile çok daha az maliyetle faydalı güvenilir bilgi elde etmek mümkündür.

- **Zaman kısıtı:** Örnekleme tam sayıma göre daha kısa zamanda ve daha ayrıntılı bilgi elde etme imkânı verir. Örneklemenin bu özelliği bilgiye çok hızlı gereksinim olduğunda bilhassa önemlidir.

- **Doğru bilgiye ulaşmanın sağlanması:** Örneklemenin büyük bir avantajı vardır. Veri derleme hataları daha az sayıda birimden oluşan örneklemde daha iyi kontrol edilebildiğinden dolayı, tam sayım kadar hatta bazen daha fazla doğru bilgi elde etme imkânı sağlar. Sosyal araştırmalarda bilgiler genellikle gözlem, görüşme veya anket yöntemiyle derlenir. Çok sayıda ve istenilen özelliklere sahip, derleme hatası yapmayacak gözlemci ya da görüşmeci bulmak veya yetiştirmek zor hatta olanaksızdır. Anket yapmak başlı başına bir uzmanlık alanıdır ve bu konuda çalışacak kişilerin dikkatli, bilgili ve karşı tarafa yanlış mesajlar vererek yönlendirmede bulunmayan kimseler olması gerekmektedir. Örneğin bir

kişiye "tiyatroya ne sıklıkta gidersiniz" şeklinde bir soru sormak yerine "ayda kaç kez tiyatroya gidersiniz" diye sorarsanız, belki ayda bir bile tiyatroya gitmeyen biri bu soruya "ayda 2-3 kez giderim" vb. cevap verecektir.

- **Pratik imkânsızlık durumu:** Bir birimin (gözlem biriminin) bilgi derlemek amacıyla tahrip edilmesi gerekiyorsa, örneklemeye başvurmak zorunludur. Örneğin bir mermi fabrikasında üretilen mermilerin patlayıp patlamadığı konusunda yapılacak bir araştırma için gerekli bilgilerin derlenmesi amacıyla üretilen tüm mermilerin teste tabi tutulması olanaksızdır ve aynı zamanda anlamsızdır. Zira bu durumun sonunda elimizde sadece boş kovanlar ve defolu (patlamayan) mermiler kalacaktır.

2.4. Basit Tesadüfi Örnekleme

N hacimli bir kitleden seçilebilecek birbirinden farklı n büyüklüğünde C_n^N kadar sayıdaki örneklemlerin her birine eşit seçilme şansı tanıyan örnekleme yöntemidir.

$C_n^N = \dfrac{N!}{n!(N-n)!}$ tane farklı örnek seçilebilir. Bir örneklemin bütün olası örneklemler arasından seçilme olasılığı;

$$\frac{1}{C_n^N} = \frac{1}{\dfrac{N!}{n!(N-n)!}}$$ dir.

Basit rastgele örnekleme 2 şekilde yapılabilir:

1) Kura yöntemi

2) Rastgele sayılar tablosu kullanılarak

1) Kura Yöntemi: Kitledeki tüm birimler 1'den N'e kadar numaralandırılır. n tane birim kura ile belirlenir.

2) Rastsal sayılar tablosu kullanılarak:

Rastsal sayılar tablosu, çok çeşitli varyasyonları bulunan ve rastgele üretilmiş sayılardan oluşan bir tablodur. Basit rastgele örnekleme yöntem uygulanırken sıklıkla başvurulan bir tablodur.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1	8	0	9	4	2	5	2	5	8	2	4	7	1	3	4	7	7	4	3	3	3	6	2	0	1	8	9	7	2	1	3	4
2	3	5	6	3	2	1	9	6	8	2	1	1	9	0	4	5	2	6	1	8	2	7	5	1	2	6	2	7	1	0	9	5
3	1	3	3	0	6	3	3	1	3	7	5	3	9	6	9	3	8	7	3	8	6	6	1	5	1	5	3	8	8	5	4	3
4	3	5	6	5	0	0	1	6	2	2	4	3	6	4	3	2	4	7	9	6	6	0	9	5	5	2	8	3	1	6	2	0
5	7	8	5	0	5	9	2	6	5	5	8	8	7	3	1	1	2	1	9	2	4	6	4	5	3	5	3	0	5	5	8	9
6	4	4	9	0	5	4	1	7	9	7	2	7	6	1	5	3	5	9	0	1	4	8	7	8	9	9	8	0	9	8	7	7
7	6	6	4	6	9	1	0	4	9	3	1	8	8	8	1	9	7	5	3	7	2	7	8	5	9	3	7	3	2	4	4	5
8	3	6	2	6	5	9	9	5	1	2	1	5	9	7	5	3	9	2	2	3	5	6	5	8	2	9	4	4	2	8	9	9
9	4	6	6	5	4	8	2	0	7	5	5	4	0	6	1	2	9	6	8	3	4	2	5	1	9	1	3	8	1	7	0	9
10	6	4	9	8	7	5	1	9	0	4	7	4	7	8	1	8	6	8	3	2	9	6	8	3	9	8	7	2	4	0	9	0
11	6	7	2	2	9	8	6	9	9	3	6	1	7	8	7	5	4	8	8	3	1	3	1	5	9	6	7	9	8	8	3	4
12	9	7	4	8	5	9	3	2	5	1	1	5	2	7	2	1	0	0	3	3	9	3	0	3	9	7	1	3	4	0	1	2
13	5	6	4	1	1	4	1	7	1	4	1	9	7	4	3	4	8	1	6	5	7	3	6	8	1	2	1	8	5	0	3	9
14	7	4	4	4	9	2	0	0	8	8	4	0	5	8	8	2	4	3	9	8	3	9	0	4	9	1	9	9	9	3	3	6
15	8	2	7	9	3	0	1	9	4	6	7	2	3	7	4	3	3	9	7	9	4	6	8	9	9	0	2	1	6	9	9	0
16	0	1	6	1	7	6	1	7	1	0	2	4	2	3	8	7	2	8	9	1	6	6	7	7	1	5	8	5	2	4	8	2
17	7	3	8	8	9	7	5	9	7	5	5	5	6	6	2	4	9	9	7	7	2	0	0	8	5	5	9	6	9	7	4	0
18	7	8	3	0	4	7	1	4	3	6	9	5	2	9	1	9	1	8	0	4	4	0	4	4	1	0	3	4	2	5	9	7
19	9	8	8	7	4	2	1	6	6	5	2	6	4	5	3	5	8	4	3	0	5	2	7	0	9	6	0	5	0	7	6	8
20	1	2	6	1	2	5	1	6	8	5	6	9	2	3	1	0	3	9	3	9	8	7	0	3	9	8	4	1	0	3	5	3
21	3	9	4	7	4	9	3	7	7	6	3	4	2	5	4	3	6	2	3	9	7	4	5	5	2	0	5	5	7	7	9	5
22	4	5	5	0	8	1	0	3	1	2	5	0	2	3	0	4	1	1	3	8	9	7	8	8	9	1	4	4	4	5	2	6
23	1	3	4	4	9	6	9	7	2	3	8	3	6	9	7	6	6	2	5	1	4	2	0	1	2	0	3	8	6	5	5	2
24	8	9	7	6	5	8	2	3	8	4	8	7	0	4	5	0	3	1	0	6	9	1	6	6	2	7	1	7	7	6	0	1
25	7	7	1	0	9	9	4	3	6	9	7	8	8	2	7	3	9	7	1	4	9	7	0	0	1	5	6	6	2	8	8	9
26	6	9	5	9	6	0	0	8	8	4	4	2	2	2	8	2	1	5	2	4	2	5	1	7	5	8	1	8	0	0	8	1
27	7	9	4	1	2	3	1	2	2	4	3	1	6	7	0	2	9	9	8	4	3	4	6	9	3	0	8	5	4	7	6	2
28	2	2	8	4	0	8	9	6	9	1	0	7	5	5	4	2	7	3	1	9	3	7	8	2	1	0	6	8	9	5	7	4
29	9	5	9	4	7	4	1	6	9	3	6	5	6	0	4	5	1	1	8	3	5	9	1	6	9	5	9	9	1	1	4	3
30	4	6	1	3	8	5	4	9	6	3	6	9	3	2	0	8	5	1	0	9	9	6	8	0	1	1	6	8	6	1	3	3

Başlangıç noktası bazen size verilebilir yani mesela denebilir ki 27. Satır ile 15. Sütunun kesiştiği noktadan başlayarak N=100 gözlemli kitleden n=10 gözlemli rastgele örneklem seçiniz. Ya da kalemin ucuyla tamamen rastgele bir biçimde bir başlangıç noktası seçebilirsiniz. Ancak böylesi keyfi bir seçim pek çok istatistikçi tarafından yanlı bir seçim olarak kabul edilir. Bu nedenle bu şekilde bir başlangıç noktası seçiyorsanız, seçtiğiniz bu rakamın sağından veya solundan 2 ya da 3 basamaklı bir sayı belirleyerek bu sayılardan birini satır diğerini sütun numarası kabul ederek yansız olması umut edilen bir başlangıç seçmiş olduğunuz kabul edilir.

Kitle büyüklüğüne göre kaç basamaklı sayı alınacağı belirlenerek örneklem çekilir. Çekim işlemi herhangi bir yöne doğru yapılabilir. Seçilen birim (yani esasen birimi temsil eden rakam) bir kez daha çıkarsa, örnekleme alınmaz, seçim işlemine devam edilir.

Örneğin 100 birimlik kitlemizden rastsal sayılar tablosunu kullanarak 10 birimlik örneklem çekelim: 27. Satır ile 15. sütunun kesiştiği noktadan 10 tane rakamı belirleyelim.

Unutmamak gerekir ki 100 gözlem olduğu için 2 basamaklı rakamları göz önüne alacağız. Bahsi geçen kesişim noktası aşağıdaki şekilde görülmektedir:

	1 2 3 4	5 6 7 8	9 10 11 12	13 14 15 16	17 18 19 20	21 22 23 24	25 26 27 28	29 30 31 32
1	8 0 9 4	2 5 2 5	8 2 4 7	1 3 4 7	7 4 3 3	3 6 2 0	1 8 9 7	2 1 3 4
2	3 5 6 3	2 1 9 6	6 2 1 1	9 0 4 5	2 6 1 8	2 7 5 1	2 6 2 7	1 0 9 5
3	1 3 3 0	6 3 3 1	3 7 5 3	9 6 9 3	8 7 3 6	6 6 1 5	1 5 3 6	6 5 4 3
4	3 6 6 5	0 0 1 6	2 2 4 3	6 4 3 2	4 7 9 6	6 0 9 5	5 2 8 3	1 6 2 0
5	7 8 5 0	5 9 2 5	5 5 8 8	7 3 1 1	2 1 9 2	4 5 4 5	3 5 3 0	6 5 8 9
6	4 4 9 0	5 4 1 7	9 7 2 7	6 1 5 3	5 9 0 1	4 8 7 8	9 9 8 0	9 8 7 7
7	6 6 4 5	9 1 0 4	9 3 1 8	8 8 1 9	7 5 3 7	2 7 8 5	9 3 7 3	2 4 4 5
8	9 6 2 6	5 9 9 5	1 2 1 5	9 7 5 3	9 2 2 3	5 6 5 8	2 9 4 4	2 8 9 9
9	4 6 6 5	4 8 2 0	7 5 5 4	0 6 1 2	9 6 8 3	4 2 5 1	9 1 3 8	1 7 0 9
10	6 4 9 8	7 5 1 9	0 4 7 4	7 8 1 8	6 8 3 2	9 6 8 3	9 8 7 2	4 0 9 0
11	6 7 2 2	9 8 6 9	9 3 6 1	7 8 7 5	4 8 8 3	1 3 1 5	9 6 7 9	8 8 3 4
12	9 7 4 8	5 9 3 2	5 1 1 5	2 7 2 1	0 0 3 3	9 3 0 3	9 7 1 3	4 0 1 2
13	5 6 4 1	1 4 1 7	1 4 1 9	7 4 3 4	8 1 6 5	7 3 6 8	1 2 1 8	5 0 3 9
14	7 4 4 4	9 2 0 0	8 8 4 0	5 8 8 2	4 3 9 8	3 9 0 4	9 1 9 9	9 3 9 6
15	8 2 7 9	3 0 1 9	4 6 7 2	3 7 4 3	3 9 7 9	4 6 8 9	9 0 2 1	6 9 9 0
16	0 1 6 1	7 6 1 7	1 0 2 4	2 3 6 7	2 6 9 1	6 6 7 7	1 5 8 5	2 4 8 2
17	7 3 8 8	9 7 5 9	7 5 5 5	6 6 2 4	9 9 7 7	2 0 0 8	5 5 9 6	9 7 4 0
18	7 8 3 0	4 7 1 4	3 6 9 5	2 9 1 9	1 8 0 4	4 0 4 4	1 0 3 4	2 5 9 7
19	9 8 8 7	4 2 1 6	6 5 2 6	4 5 3 5	8 4 3 0	5 2 7 0	9 6 0 5	0 7 6 8
20	1 2 6 1	2 5 1 6	8 5 6 9	2 3 1 0	3 9 3 9	8 7 0 3	9 8 4 1	0 3 5 3
21	3 9 4 7	4 9 3 7	7 6 3 4	2 5 4 3	6 2 3 9	7 4 5 5	2 0 5 5	7 7 9 5
22	4 5 5 0	8 1 0 3	1 2 5 0	2 3 0 4	1 1 3 8	9 7 8 8	9 1 4 4	4 5 2 6
23	1 3 4 4	9 6 8 7	2 5 6 3	6 9 7 6	6 2 5 1	4 2 0 1	2 0 3 8	6 5 5 2
24	8 9 7 6	5 8 2 3	8 4 8 7	0 4 5 0	3 1 0 6	9 1 6 6	2 7 1 7	7 8 0 1
25	7 7 1 0	9 9 4 3	6 9 7 8	8 2 7 3	9 7 1 4	9 7 0 0	1 5 6 6	2 8 8 9
26	8 9 5 9	6 0 0 8	8 4 4 2	2 2 8 2	1 5 2 4	2 5 1 7	5 8 1 8	0 0 8 1
27	7 9 4 1	2 3 1 2	2 4 3 1	6 7 0 2	9 9 8 4	3 4 6 9	3 0 8 5	4 7 6 2
28	2 2 8 4	0 8 9 6	9 1 0 7	5 6 4 2	7 3 1 9	3 7 8 2	1 0 6 8	9 5 7 4
29	9 5 9 4	7 4 1 6	9 3 6 5	6 0 4 5	1 1 8 3	5 9 1 6	9 5 9 9	1 1 4 9
30	4 6 1 3	8 5 4 9	6 3 6 9	3 2 0 8	5 1 0 9	9 6 8 0	1 1 6 8	6 1 3 3

Şimdi bu 02 noktasından istediğimiz yöne doğru ilerleyerek 10 tane 2 basamaklı sayıyı tespit edelim ve bunun için yukarı doğru gidelim. 82-73-60-76-04-43-10-35-19-24. 1'den 100'e kadar sıralanmış olan kitledeki birimlerden rastgele sayılar tablosunu kullanarak 10 tane birim seçmiş olduk.

1	2	3	**4**	5	6	7	8	9	**10**
11	12	13	14	15	16	17	18	**19**	20
21	22	23	**24**	25	26	27	28	29	30
31	32	33	34	**35**	36	37	38	39	40
41	42	**43**	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	**60**
61	62	63	64	65	66	67	68	69	70
71	72	**73**	74	75	**76**	77	78	79	80
81	**82**	83	84	85	86	87	88	89	90

91	92	93	94	95	96	97	98	99	100

Şimdi bir başka örneklem seçelim:

Başlangıç noktamız 1. Satıra 1. Sütun olsun yani 80 rakamı olsun ve bu defa aşağı doğru giderek 10 tane rakamı belirleyelim:

80-35-13-35-78-44-66-38-46-64-67.

Bir gözlem sadece bir defa örnekleme alınabilir, aynı rakamları eliyoruz.

80-35-13-78-44-66-38-46-64-67.

1	2	3	4	5	6	7	8	9	10
11	12	**13**	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	**35**	36	37	**38**	39	40
41	42	43	**44**	45	**46**	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	**64**	65	**66**	**67**	68	69	70
71	72	73	74	75	76	77	**78**	79	**80**
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

Bu şekilde C_{10}^{100} kadar sayıda farklı örneklem çekilebilir.

Örnek: 50 öğrencisi olan bir dershanede 10'ar kişilik 5 sınıf bulunmaktadır. Öğrencilerin numaraları 30'dan başlamaktadır. Her sınıftan 1'er öğrenci olmak üzere 5 öğrenciyi basit rastsal örnekleme yöntemiyle seçiniz. (rastsal sayılar tablosu kullanınız)

1.sınıf	**2.sınıf**	**3.sınıf**	**4.sınıf**	**5.sınıf**

30	40	50	60	70
31	41	51	61	71
32	42	52	62	72
33	43	53	63	73
34	44	54	64	74
35	45	55	65	75
36	46	56	66	76
37	47	57	67	77
38	48	58	68	78
39	49	59	69	79

Öğrencilerin numarası 30'dan başlıyor!!!

Rastsal sayılar tablosundan herhangi bir noktadan başlayalım. N=50 iki basamaklı olduğundan, seçeceğimiz sayılar da 2 basamaklı olacak.

Rastsal Sayılarımız	**Kitledeki Birim Numarası**
09	30+9=39 1.sınıftan 1 kişi aldık
95	30+95=125 kitlede böyle bir birim yok, seçime devam
93	93+30=123 kitlede böyle bir birim yok, seçime devam
85	85+30=115 kitlede böyle bir birim yok, seçime devam
42	42+30=72 5. sınıftan bir kişi aldık
14	14+30=44 2. sınıftan bir kişi aldık
21	21+30=51 3. sınıftan bir kişi aldık
30	30+30=60 4. sınıftan bir kişi aldık

1.sınıf	**2.sınıf**	**3.sınıf**	**4.sınıf**	**5.sınıf**

Rastsal sayılar tablosunu kullanarak her sınıftan tamamen tesadüfi olarak 1'er öğrenci seçmiş olduk:

30	40	50	60	70
31	41	51	61	71
32	42	52	62	72
33	43	53	63	73
34	44	54	64	74
35	45	55	65	75
36	46	56	66	76
37	47	57	67	77
38	48	58	68	78
39	49	59	69	79

2.5. Örnekleme Dağılımı

Bir kitleye ilişkin sayısal karakteristiklere parametre dendiğini biliyoruz. Parametreler genel olarak θ ile gösterilirler. Örneklem istatistikleri ise $\hat{\theta}$ ile gösterilirler. İstatistikler tesadüfî olarak seçilen n büyüklüğündeki örneklemden elde edilen $x_1, x_2, \ldots\ldots x_n$ gözlem değerlerinin kullanılmasıyla hesaplanırlar. Olası her bir örnek için $\hat{\theta}$ istatistiği hesaplandığı varsayılırsa, $\hat{\theta}_1, \hat{\theta}_2, \ldots\ldots \hat{\theta}_{C_n^N}$ istatistiklerinin (bu istatistiklerden kasıt, örneklem ortalaması $\overline{x}$, örneklem varyansı s^2,v.b.) dağılımına örnekleme dağılımı denir.

2.6. Kitle Ortalaması μ'nün Tahmini

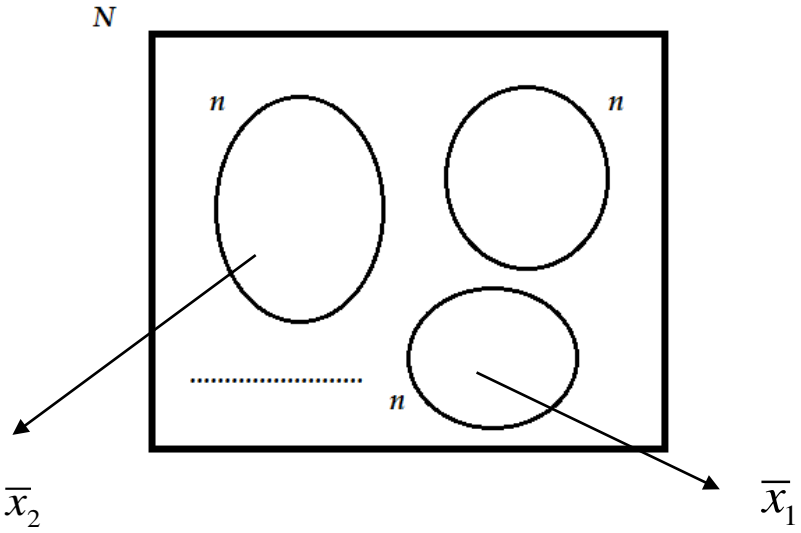
Örneklem istatistikleri kendileri birer rastlantı değişkenidirler. Çünkü örneklemden örnekleme farklı değerler alırlar.

Parametreler	Örneklem istatistikleri
μ	$\overline{x}$
σ^2	s^2

N hacimli kitleden n büyüklüğünde C_n^N kadar sayıda farklı örneklem çekilebilir ve her bir örneklemin bir ortalaması vardır; $\overline{x}_1$, $\overline{x}_2$,..... $\overline{x}_{C_n^N}$. Kitle ortalaması μ'nün tahmini, örneklem ortalamalarının beklenen değeridir.

$$E(\overline{x}_i) = \mu$$

$$E(\overline{x}_i) = \frac{\sum_{i=1}^{C_n^N} \overline{x}_i}{C_n^N}$$



Diyelim ki N=10 ve n=3 olsun. C_3^{10} tane farklı örneklem çekilebilir, yani 120 tane.

$$E(\overline{x}_i) = \mu = \frac{\sum_{i=1}^{120} \overline{x}_i}{120} = \frac{\overline{x}_1 + \overline{x}_2 + ... + \overline{x}_{120}}{120}$$

Örnek:

x	0	2	4	6
P(X=x)	¼	¼	¼	¼

n=2 büyüklüğünde örneklemler çekerek $\overline{x}$ 'nın örnekleme dağılımını bulunuz.

Çözüm:

Olası örneklem sayısı $C_2^4 = \dfrac{4!}{2!2!} = 6$ tanedir.

Örneklemler	$\overline{x}_i$	Olasılıklar
0,2	1	1/6
0,4	2	1/6
0,6	3	1/6
2,4	3	1/6
2,6	4	1/6
4,6	5	1/6

$\overline{x}$	1	2	3	4	5
P($\overline{x}$)	1/6	1/6	2/6	1/6	1/6

$$E(\overline{x}) = \mu = (1 \times \frac{1}{6}) + (2 \times \frac{1}{6}) + (3 \times \frac{2}{6}) + (4 \times \frac{1}{6}) + (5 \times \frac{1}{6}) = 3$$

Örnek:

x	0	2	4	6
P(X=x)	¼	¼	¼	¼

n=3 büyüklüğünde örneklemler çekerek $\overline{x}$ 'nın örnekleme dağılımını bulunuz.

Çözüm:

Olası örneklem sayısı $C_3^4 = \dfrac{4!}{3!1!} = 4$ tanedir.

Örneklemler	$\overline{x}_i$	Olasılıklar
0,2,4	2	1/4
0,2,6	2.66	1/4
2,4,6	4	1/4
0,4,6	3.33	1/4

$\overline{x}$	2	2.66	4	3.33
P($\overline{x}$)	1/4	1/4	1/4	1/4

$$E(\overline{x}) = \mu = (2 \times \frac{1}{4}) + (2.66 \times \frac{1}{4}) + (3.33 \times \frac{1}{4}) + (4 \times \frac{1}{4}) = 2.99$$

2.7. Örneklemede Hata Kavramı ve Standart Hata

Örneklemden örnekleme değişken değerler alan istatistiklerin kitle parametre değerlerine göre gösterdiği sapmalara yani $(\theta - \hat{\theta})$ 'lara örnekleme hatası denir. Sapmaların sıfır olması durumunda, yapılacak tahmin sapmasızdır (yansızdır). Örnekleme hatalarının kareli ortalamasına örneklem istatistiğinin standart hatası denir. Hesaplanan standart hata değeri, istatistiğin kitle parametresinden ortalama olarak ne kadar saptığını gösterir. Sapmanın sıfıra yakın olması istatistiklerden yararlanılarak üretilen bilgilerin güvenilir olduğunu yani yapılan çıkarsamaların güvenilir olduğunu gösterir.

Ortalamanın Standart Hatası

$$E(\overline{x}_i - \mu)^2 = \sigma_{\overline{x}}^2$$

İadesizde $\frac{n}{N} \geq 0.05 \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$

İadelide $\frac{n}{N} < 0.05 \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

$\frac{n}{N} < 0.05$ olduğunda yani n çok küçüldüğünde ya da kitle çok büyüdüğünde ($N \to \infty$), $N - n \cong N - 1$ olacaktır ya da öyle olduğu kabul edilir. Bu nedenle sonsuz kitlede düzeltme faktörü kullanmaya gerek duyulmaz.

2.8. Oranların Örnekleme Dağılımı

N hacimli bir kitlede ilgilenilen türden birimlerin sayısı A olmak üzere;

$\pi = \frac{A}{N}$ bu birimlerin kitledeki oranıdır.

Örneklemde ise;

$p = \frac{a}{n}$'dir.

İadesizde $\frac{n}{N} \geq 0.05$ iken;

$$\sigma_p^2 = \frac{\pi(1-\pi)}{n} \times \frac{N-n}{N-1}$$

İadelide $\frac{n}{N} < 0.05$ iken;

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

Π genellikle bilinmez, bunun yerine p kullanılır.

$$s_p^2 = \frac{p(1-p)}{n} \times \frac{N-n}{N-1}$$

$$s_p^2 = \frac{p(1-p)}{n}$$

2.9. Örneklem Hacminin Belirlenmesi

Örnekleme sürecinin en önemli konularından birisi örnekleme girecek birim sayısının, örnek hacminin belirlenmesidir. Çünkü örneklem hacmi arttıkça örneklem istatistiklerinin hesaplanan değeri parametre değerlerine yaklaşır, mümkün örneklem istatistiklerinin değişim aralığı dağılır, standart hata küçülür ve dolayısıyla doğruluk artar. Bir araştırmada kitlenin bilinmeyen karakteristiklerinin (parametrelerinin) doğruluğu yüksek tahminini elde etmek örneklemenin temel amacıdır. Ancak bu amacın istenen şekilde gerçekleşmesi mümkün değildir. Çünkü örneklem hacmi arttıkça örneklem için gerekli harcama ve zaman artar. Bu nedenlerle örneklem hacmi belirlenirken yalnız standart hatanın büyüklüğü değil, örnekleme bütçesi ve diğer kaynaklar göz önünde bulundurulmalıdır. Kitlenin homojen veya heterojen oluşu örneklem hacminin belirlenmesinde etkilidir. Bir homojen kitlenin (tipik olaylar kitlesi, bkz. Tipik olay, hep aynı şekilde tezahür eden olaylar, bir bebeğin bir burnu olması, bir arabanın 4 tekerleği olması vb. olasılıksız durumlar) karakteristiklerini doğru tahmin etmek için örneklem hacminin bir olması yeterlidir. Üzerinde çalışılacak kitlenin heterojenliği arttıkça, doğru sonuçlara varabilmek için, örnekleme hacmini de arttırmak gerekir.

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ ise, $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$, dir yani buradan; $n = \frac{\sigma^2}{\sigma_{\bar{x}}^2}$ elde edilir.

Örnek: N=500 birimlik bir kitlenin σ=85 cm'dir ortalamanın standart hatasının 15 cm'den fazla olmaması için örneklem kaç birim olmalıdır?

Çözüm: İadelide; $n = \frac{\sigma^2}{\sigma_{\bar{x}}^2} = \frac{85^2}{225} = 32.11 \cong 33$

İadesizde;

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \times \frac{N-n}{N-1}$$ buradan n'i çekersek;

$$n\left(\frac{N-n}{N-1}\right) = \frac{\sigma^2}{\sigma_{\bar{x}}^2}$$

$$n(\frac{500-n}{500-1}) = \frac{85^2}{15^2}$$

Buradan n=30.23≈31 birim hesaplanır.

Örnek: π=0.45 olduğu bilinmektedir. Rastgele 36 birim seçiliyor. Söz konusu oranın varyansını ve standart hatasını bulunuz.

Çözüm: π=0.45 n=36

$$\sigma_p^2 = \frac{\sigma^2}{n} = \frac{\pi(1-\pi)}{n} = \frac{0.2475}{36} = 0.007$$

$$\sigma_p = \sqrt{0.007} = 0.08$$

KAYNAKÇA

1. Örnekleme Kuramı, Hülya Çıngı, Hacettepe Üniversitesi, 1990, Ankara.

2. Uygulamalı İstatistik 1 ve 2, Özer Serper.

3. İktisatçılar ve İşletmeciler İçin İstatistik, Onur Özsoy, 2010.

4. Olasılık ve İstatistik, Fikri Akdeniz, 2013.

5. İstatistik Yöntemleri, Murat Karagöz, 2011.**5.**

6. http://w2.anadolu.edu.tr/aos/kitap/IOLTP/2294/unite03.pdf (*Doç.Dr. Ahmet ÖZMEN)*

İlk iki bölüm ***yukarıda belirtilen kaynaklardan kısmen ve(ya) aynen alıntılanmak suretiyle derlenmiştir.***

Bölüm Soruları

1- Örnekleme ve Tamsayım arasındaki fark nedir?

2- İstatistik ve parametre kavramlarını açıklayınız

3- Basit Tesadüfi Örnekleme metodunu kısaca açıklayınız

4- Örneklem büyüklüğü belirlenirken nelere dikkat edilmelidir?

5- İyi seçilmiş bir örneklem hangi özellikleri taşımalıdır?

6- Örneklem ve Örnekleme kavramları arasındaki fark nedir?

7- Standart hata, örneklem hatası kavramlarını açıklayınız.

BÖLÜM 2 SORULARI

1-3, 5, 7 ve 9 değerlerinden oluşan bir serinin örneklem standart sapması aşağıdakilerden hangisine eşittir?

a) 2.58 b)5.29 c) 6.18 d) 6.59 e)7.18

2-Kitledeki tüm birimlerin sayılması işlemi ……. olarak isimlendirilir. Boşluğa yazılması gereken kelime aşağıdakilerden hangisidir?

a)Örnekleme b)Örneklem c) Kısmi sayım d)Tamsayım e) Hiçbiri

3-Aşağıdakilerden hangisi örnekleme yapmayı gerektiren nedenlerden birisi değildir?

a)Maliyet b)Zaman c) Kitlenin yeterli büyüklükte olmaması d) Örneğe giren birimlerin fiziksel olarak zarara uğraması e)Kitlenin sonsuz olması

4-Aşağıdaki örnekleme yöntemlerinden hangisi örneklem çerçevesinin bilinmesini gerektirmez?

a)Basit Tesadüfi Örnekleme b)Tabakalı Örnekleme c)Küme örneklemesi d)Sistematik örnekleme

5- Aşağıdakilerden hangisi örneklem büyüklüğünü etkileyen faktörlerden birisi değildir?

a) Araştırma bütçesi b)Sabit maliyet c) Tolere edilebilir hata düzeyi d) varyans e)örneklem oranı

6-Aşağıdaki örnekleme yöntemlerinden hangisi olasılıklı olmayan örnekleme yöntemidir?

a)Kota örneklemesi b)Tabakalı örnekleme c)Küme örneklemesi d)Sistematik örnekleme e)Basit Tesadüfi Örnekleme

7- Aşağıdakilerden hangisi bir veri toplama aracı değildir?

a) Gözlem b)Tamsayım c)Anket d)Deney e)Mülakat

8- Beklenen kitle değerine eşit olan bir tahminci nasıl bir tahmincidir?

a)Tutarlı b)Yeterli c)Tarafsız d)Etkin e)yanlı

9- Tesadüfi olmayan kota örnekleme yöntemi aşağıdaki yöntemlerden hangisi ile benzerlik gösterir?

a)Sistematik b)Basit Tesadüfi c)Tabakalı d)Küme örneklemesi e) Çok aşamalı

10-Örneklem çerçevesine hangi tesadüfi örnekleme yönteminde ihtiyaç yoktur?

a)Sistematik b)Basit Tesadüfi c)Tabakalı d)Küme örneklemesi e) Kartopu

Cevaplar

1) A 2) D 3)C 4)C 5)E 6)A 7)B 8)C 9)C 10)D

Bu Bölümde Ne Öğrendik Özeti

Bu bölümde istatistiksel çıkarsama yöntemlerinin bel kemiğini oluşturan örnekleme konusu, temel kavramlar, örnekleme dağılımı ve yöntemleri dâhilinde ele alınmış, çeşitli sayısal ve görsel örneklerle desteklenmiştir.

3. TAHMİN TEORİSİ VE GÜVEN ARALIĞI

Bu Bölümde Neler Öğreneceğiz?

Bu bölümde tahmin teorisi, anakütle ortalama ve oranlarının parametreleri için güven aralığı konularını ele alacağız.

Bölüm Hakkında İlgi Oluşturan Sorular

1) Bir üniversitede okuyan 30 öğrenci ile görüşülmüş, 18'i düzenli spor yaptığını söylemiştir. Bu üniversitede okuyan ve düzenli spor yapan öğrencilerin oranı nedir?

2) Kadıköy'de yaşayan 5 farklı dört kişilik aile mensubu haftalık mutfak harcamalarının 200, 280, 400,340,440,320 TL olduğunu söylemiştir. Bu ilçede yaşayanların dört kişilil ailelerin haftalık ortalama mutfak harcamaları için ne söylenebilir?

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Tahmin teorisi ve güven aralıkları	Anakütle parametrelerinin çeşitli güven düzeylerinde tahmini yapılır.	Metinler, çözümlü problemler

Anahtar Kavramlar

- Ortalama tahmini
- Oran tahmini
- Güven aralığı
- Güven düzeyi
- Oran farklarının tahmini
- Ortalama farklarının tahmini
- İstatistiksel anlamlılık

Giriş

Örneklemden yola çıkarak anakütle hakkında çeşitli sonuçlara ulaşma süreci istatistiksel çıkarsama olarak adlandırılır. Bu bölümde istatistiksel çıkarsama hakkında bilgi edineceğiz.

3.1. İstatisitiksel Çıkarsama

İstatistiksel çıkarsama sürecinin iki temel ayağı tahmin ve hipotez testidir. Bu ve sonraki bölümde tahmin, dört ve beşinci bölümlerde ise hipotez testleri konularını ele alacağız. Tahmin anakütle parametresinin alabileceği değerlerin belirlenmesine ilişkin sürece karşılık gelirken, hipotez testi ise anakütle parametrelerine ilişkin iddiayı değerlendirme sürecidir.

İstatistiksel tahmin iki şekilde yapılabilir:

-Nokta Tahmini

-Aralık Tahmini

Bilinmeyen bir kütle parametresini tahmin etmek için kullanılan örnek istatistiğine **tahminci** denir. Bir tahmincinin bir tek değerle ifade edilmesine **nokta tahmini** adı verilir. Kütle ortalaması μ nün tahmincisi $\overline{X}$, Kütle varyansının tahmincisi s^2 σ^2 birer nokta tahminidir.

Literatürde tahmincilerin taşıması gereken 4 özellik vurgulanır:

-Sapmasızlık

-Tutarlılık

-Etkinlik

- Yeterlilik

Yukarıda vurgulanan tahmincinin taşıması gereken özellikler kısaca şöyle açıklanabilir:

1. Sapmasızlık: Örnek istatistiğinin beklenen değeri tahmin edilmek istenen ana kütle parametresine eşit ise, söz konusu istatistik ana kitle parametresinin "sapmasız" bir tahminidir

$$E(\overline{X}) = \mu, \quad E(\hat{p}) = p \quad ise \ \overline{X} \ ve \ \hat{p} \ sirasiyla \ \mu \ ve \ p'nin \quad tahmincisi$$

2. Tutarlılık: Örnek hacmi artarken tahmin edici tahmin edilmek istenen kütle parametresine yaklaşıyorsa bu tahmin ediciye tutarlı tahmin edici adı verilir.

3. Etkinlik: Tutarlı tahmin ediciler arasından birinin seçilmesi istendiğinde seçim etkinlik kriterine göre yapılır. Tahmin edicilerin etkinliği varyanslarına dayanır. Hangi örnek istatistiğinin dağılımı daha küçük varyansa sahipse, tahmin olarak o istatistik tercih edilir.

4.Yeterlilik: Bir tahmin edici tahmin edilmek istenen parametre hakkında örnekte bulunan bütün veriyi kullanıyorsa bu tahmin edicinin yeterli olduğu söylenir.

Eğer anakütle parametresi örneklemden yola çıkılarak tek bir sayısal değerle ifade edilirse bu değere nokta tahmini denir. Örneğin İktisat fakültesine kayıtlı bir öğrencinin matematik dersi not ortalamasını tahmin etmek istediğimizde bu tahmini tek bir değerle gösterilirse 75 gibi bu nokta tahmine örnektir.

Anakütle parametresinin tahminini tek bir değerle yapmak güvenilir bir çıkarsama olarak kabul edilmez. Çıkarsamanın güven düzeyini artırmak için bir tahminin belli bir hata payı içerdiğini düşünerek bir aralıkta değer vermektir. Örneğin İktisat fakültesine kayıtlı bir öğrencinin not ortalamasını tahmin etmek istediğimizde not ortalamasını içine alan bir aralık olarak70-80 arası şeklinde ifade edilebilir.

3.2. Nokta ve Aralık Tahmini

Nokta tahmini, bu tahminin bilinmeyen gerçek anakütle parametresine ne kadar yakın olabileceğine yani doğru parametre değerine hangi olasılıkla ve ne kadar yakın olduğuna ilişkin bir bilgi içermez. Oysa aralık tahmini bilinmeyen parametreye ilişkin belirsizligi dikkate alır.

Bir anakütle parametresinin aralık tahmincisi, örneklem bilgisinden faydalanarak o anakütle parametresinin içinde bulunduğu bir aralığı belirler. Buna karşılık gelen tahmine de aralık tahmini denir. Anakütle parametresinin tahmini tek bir değer yerine bir değerler aralığı içinde ifade edilir. Belli bir güvenle bu aralığın tahmin edilmek istenen parametreyi içerdiği söylenebilir. 1-α olarak ifade edilen güven düzeyi parametrenin gerçek değerinin 1-α olasılıkla belirlenen aralık içinde olduğunu ifade eder. Ya da α olasılıkla parametreyi içermez.

Belirli bir olasılık dâhilinde parametre değerlerinin şu aralıkta yer aldığı düşünülür:

$a < \mu < b$

α olasılık değeri dikkate alınarak yukarıdaki eştisizlik şöyle düzenlenir:

$p(a \leq \mu \leq b) = 1 - \alpha$, a ve b arasındaki uzaklık güven aralığını verir. Anakütle parametresi μ, 1- α olasılıkla a ve b arasındadır. Söz konusu olan anakütle oranının tahmini ise P parametresinin yaklaşımı da değişmeyecektir.

Ortalamaların dağılımı anakütle varyansının bilinip bilinmemesine göre farklılık göstermektedir. Güven aralıklarının oluşturulmasında da bu durum dikkate alınır.

3.3. Anakütle Varyansının Bilindiği Durumlarda Anakütle Ortalamasının Güven Aralığı:

Önceki bölümden hatırlanacağı üzere anakütle ortalaması μ ve varyansı σ^2 olan normal dağılıma uyarsa örnek ortalamaları da ortalaması μ ve varyansı σ^2/n olan normal dağılıma uyar. Öte yandan örnek hacmi (n) büyük olduğu zaman ortalamaların örnekleme dağılımı normal dağılıma yaklaşmaktadır.

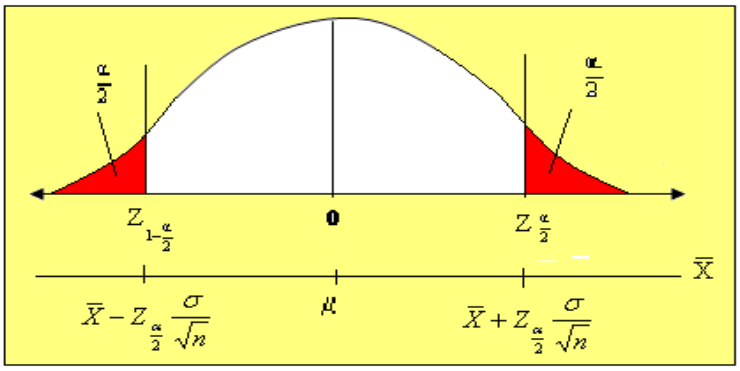
%100(1-α) güven derecesi ile μ için güven aralığı şöyle olur:

$$\overline{X} - Z_{\frac{\alpha}{2}} . \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \text{ olacaktir.}$$

Yukarıdaki eşitsizliği tahmin edilecek diğer anakütle parametreleri için de aslında şu şekilde gösterebiliriz:

$$istatistik \mp tablo\, değeri * s\tan dart\ hata$$

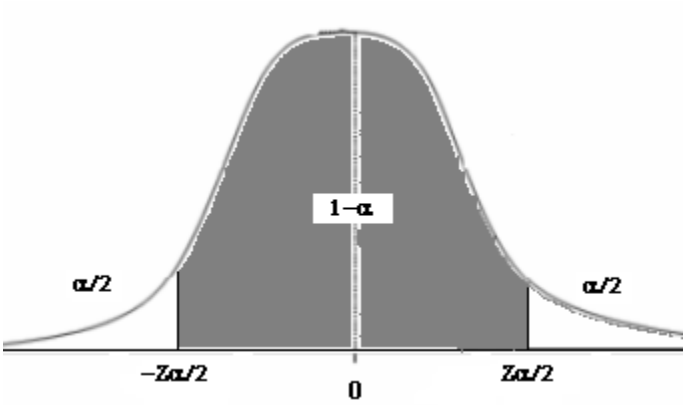
Anakütle varyans değerinin bilindiği ve örnek sayısının 30'dan fazla olduğu durumlarda normal dağılım tablosu (Z tablosu) kullanılır.



Grafikte görülen kırmızı alanların toplamı α olasılık değerine karşılık gelecektir. Beyaz alan ise anakütle parametresinin yer aldığı alandır. Şimdi çeşitli örnekler üzerinde bu kavramları değerlendirelim.

ÖRNEK:

Piyasaya yeni sürülen bir ürünün uzunluğunun stansdart sapması 2cm'dir. Rastgele seçilen 16 ürünün ortalama uzunluğu 4 cm olarak hesaplanmıştır. %95 güvenle anakütle ortalamasını tahmin ediniz.



% 95 güvenle yapılacak tahmin $1-\alpha = 0.95$ ve $\alpha = 0.05$ şeklinde yorumlanacaktır.

Tablo değerini tespit edebilmek için çizimde yer alan alanın tek bir tarafı (sağ ya da sol) için düşünerek hareket etmek tablo okumayı kolaylaştıracaktır.

$\alpha/2 = 0.025$, 0.5-0.025=0.475 ya da 0.95/2=0.475 griye boyalı alanın yarısına karşılık gelecektir. Bölüm sonunda yer alan tablodan da görülebileceği gibi 0.475 değerine karşılık gelen tablo değeri 1.96 'dır.

Z tablosundan elde edilen değer $Z_{\alpha/2} = 1.96$.

$$\overline{X} - Z_{\frac{\alpha}{2}} . \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$4 - 1.96 . \frac{2}{\sqrt{16}} \leq \mu \leq 4 + 1.96 \frac{2}{\sqrt{16}}$$

$$3.02 \leq \mu \leq 4.98$$

16 örnekten elde edilen bilgiyle söylenebilir ki anakütle ortalaması % 95 güvenle 3,02 ile 4,98 arasındadır.

Eğer anakütle varyansına ilişkin bilgi soruda verilmişse örnek sayısı 30'dan az olsa da normal dağılım tablosunu kullanıyoruz.

Örnek: Bir fabrikada üretilen margarin paketlerinin ağırlığının varyansı 100 gr dır. Rastgele seçilen 25 paketin ağırlığının ortalaması 120 gr.'dir. Ana kütle ortalamasını %90 ve %99 güvenle tahmin ediniz.

-%90 güvenle anakütle ortalama değeri

α=0.10, α/2=0.05 olacaktır. Tablo değeri Z_α=1.65 olur

$$120 - 1.65 \frac{10}{\sqrt{25}} \leq \mu \leq 120 + 1.65 \frac{10}{\sqrt{25}}$$

$$116.71 \leq \mu \leq 123.29$$

Aynı koşullar altında alınan 100 örnekten bulunan güven aralıklarının 90 tanesi parametre değerini içinde bulundurması beklenir.

-%99 güvenle anakütle ortalama değeri

α=0.01, α/2=0.005 olacaktır. Z_α=2.58 olur

$$120 - 2.58\frac{10}{\sqrt{25}} \le \mu \le 120 + 2.58\frac{10}{\sqrt{25}}$$

$$114.76 \le \mu \le 125.24$$

3.4. Anakütle Varyansının Bilinmediği Durumlarda Anakütle Ortalamasının Güven Aralığı

Anakütle varyansı çoğu zaman bilinmez. Anakütle varyansı σ^2 bilinmediği zaman anakütlenin dağılımı normal olmak koşuluyla μ'nün (1-α) güven aralığı t dağılımı ile belirlenir.

Normal dağılan bir anakütleden çekilen n büyüklüğündeki rassal bir örnek için $\frac{\bar{X} - \mu}{S/\sqrt{n}}$ nin dağılımı n-1 serbestlik dereceli t dağılımına uyar.

Buna göre anakütle varyansı bilinmeyen bir anakütleden çekilen n birimlik bir örneğe dayanarak anakütle ortalamasının (1-α) güven aralığı şöyle yazılır.

$$\overline{X} - t_{\frac{\alpha}{2},n-1} \cdot \frac{S}{\sqrt{n}} \le \mu \le \overline{X} + t_{\frac{\alpha}{2},n-1} \cdot \frac{S}{\sqrt{n}}$$

Anakütle varyansı bilinmediğinde ve örnek hacmi n<30 olduğunda t dağılımı tercih edilir, t tablosu bölüm sonunda yer almaktadır. Tabloda yer alan serbestlik derecesi kolonuna n-1 dikkate alınarak ulaşılır, tablo okuma yaklaşımı normal dağılım tablosunda olduğu gibidir.

Örnek:

ABC dersanesinin öğrencilerinin bir sınavdaki ortalama başarısını ölçmek için 16 öğrenci seçiliyor, öğrencilerin not ortalamasının 90 ve notların standart sapmasının da 12 olduğu görülüyor. %95 güvenle bu ABC dershanesinin başarı ortalamasını tamin ediniz.,

Anakütle varyansı verilmemiş, örnek sayısı da n<30 olduğundan t dağılımı tercih edilir.

$$n = 16,\ \ \overline{X} = 90,\ \ S = 12,\ \ 1 - \alpha = 0,95 \quad t_{\alpha/2,sd} = t_{0,025,15} = 2,13$$

Bölüm sonunda t tablosu yer almaktadır.

$$\overline{X} - t_{\frac{\alpha}{2},n-1} \cdot \frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\frac{\alpha}{2},n-1} \frac{S}{\sqrt{n}}$$

$$90 - 2,13 \cdot \frac{12}{\sqrt{16}} < \mu < 90 + 2,13 \cdot \frac{12}{\sqrt{16}}$$

%95 güvenle anakütle ortalaması 83.61 ile 96.39 arasındadır.

Örnek:

Bir parkuru tamamlamak isteyen 10 koşucunun dakika cinsinden süreleri şöyledir: 15, 8.5, 10.2, 11, 12.2, 11, 7.3, 10.1, 9.6, 14.5 Anakütle ortalaması için %95 güvenle sınırları oluşturunuz.

Anakütle varyansı bilinmiyor ve örnek sayısı 10 yine t dağılımı kullanılmalıdır.

$$(1-\alpha)=0.95 \quad \alpha/2=0.025 \quad t_{10-1,0.025}=2.262$$

$$s^2=\frac{\sum(x_i-\overline{x})^2}{n-1}=5.89$$

$$10.94-2.262\frac{\sqrt{5.89}}{\sqrt{10}}\le\mu\le 10.94+2.262\frac{\sqrt{5.89}}{\sqrt{10}}$$

$$9.2\le\mu\le 12.68$$

3.5. Anakütle Oranının Tahmini

Anakütle oranı p'nin tahmin edilmesi için örneklem oranı $\hat{p}$ 'dan faydalanılır.

p' nin 1-α güven aralığı normal dağılım varsayımı ile şöyle yazılır.

$$\hat{p}-z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\le p\le\hat{p}+z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Yaklaşım yine aynıdır, örnekten hesaplanan değere "tablo değeri* standart hata" eklenir ve çıkarılır. Kullanacağımız tablo z tablosudur.

$$\hat{p}\mp Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Örnek:

Bir bölgede sigara içenlerin oranını tahmin etmek amacıyla rastgele 200 kişiye sorulmuş ve 24 kişi sigara içtiğini bildirmiştir. Bu verilere göre bu bölgedeki sigara içme oranını %99 güvenle tahmin ediniz.

$$\hat{p} = \frac{24}{200} \Rightarrow \hat{p} = 0,12 \quad \hat{q} = 0,88 \quad n = 200 \quad 1-\alpha = 0,99 \quad \alpha = 0,01 \quad Z_{\alpha/2} = Z_{0,005} = 2,58$$

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0,12 - 2,58 \cdot \sqrt{\frac{0,12 \cdot 0,88}{200}} < p < 0,12 + 2,58 \cdot \sqrt{\frac{0,12 \cdot 0,88}{200}}$$

$$0.061 \leq p \leq 0.179$$

Örnek:

Bir doktora gelen 150 hastanın 50'sinde astım rahatsızlığına rastlanmıştır. Bu bölgede yaşayan astım hastalarının oranını %95 güvenle tahmin ediniz.

$$Z_{\alpha/2} = 1.96$$

$$\hat{p} = 50/150 = 0,3$$

$$0.3 - 1.96\sqrt{\frac{0.3x0.7}{150}} \leq p \leq 0.3 + 1.96\sqrt{\frac{0.3x0.7}{150}}$$

%95 güvenle bu aralık şöyledir:

$$0.226 < p < 0.373$$

Örnek:

Bir sınıfta test sınavına verilen doğru yanıtların ortalamasının 28, standart sapmasının da 2 olduğu hesaplanmıştır. Rastgele seçilen 81 öğrencinin vereceği doğru yanıt sayısını %95 güven düzeyinde belirleyiniz.

$$\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$28 - 1.96\frac{2}{9} \le \mu \le 28 + 1.96\frac{2}{9}$$

$$27.565 \le \mu \le 28.435$$

Örnek: Bir sınıftaki 25 öğrencinin test sınavına verdikleri doğru yanıtların ortalamasının 28, standart sapmasının da 2 olduğu hesaplanmıştır. Öğrencilerin verdikleri doğru yanıt sayısını %95 güven düzeyinde belirleyiniz.

$$\bar{X} - t_{\alpha/2}\frac{s}{\sqrt{n}} \le \mu \le \bar{X} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

$$28 - 2.064\frac{2}{5} \le \mu \le 28 + 2.064\frac{2}{5}$$

$$27.565 \le \mu \le 28.435$$

$$27.17 \le \mu \le 28.83$$

Örnek: Üniversite öğrencilerinin oluşturduğu bir kitleden rastgele seçilen 49 öğrencinin haftalık spor yapma sıklıkları incelenmiş ve haftalık spor süresinin 20 saat ortalamaya ve 5 saat standart sapmaya sahip olduğu görülmüştür. Üniversite öğrencilerinin haftalık ortalama spor yapma süresini %95 güven düzeyinde tahmin ediniz.

$$20 \pm 1.96\frac{5}{\sqrt{49}}$$

$$18.6 \le \mu \le 21.6$$

Örnek: Bir üretim sürecinden rastgele seçilen 25 ürünün ortalama ağırlığı 1000 gr ve varyansı da 625 gr olarak hesaplanmıştır. %95 güvenle üretilen ürünlerin ortalama ağırlığını tahmin ediniz.

Kitle sapması bilinmediğinden ve örnek sayısı 30'dan küçük olduğunda t dağılımı tercih edilir.

$$1000 \pm 2.064 \frac{25}{\sqrt{25}}$$

$$989.68 \leq \mu \leq 1010.32$$

Örnek:

Bir bölgede yaşayan ve sabahları işe kendi araçları ile giden 250 kişi ile görüşülmüş bu kişilerin 185'i sabahları yolda radyo dinlediklerini söylemiştir. Buna göre sabahları radyo dinleyenlerin oranı %99 güvenle nedir?

$$\hat{p} = \frac{185}{250}$$

$$\hat{q} = 0.26$$

$$0.74 - 2.58\sqrt{\frac{0.74 * 0.26}{250}} \leq p \leq 0.74 + 2.58\sqrt{\frac{0.74 * 0.26}{250}}$$

$$0.72 \leq p \leq 0.76$$

Örnek:

Rastgele seçilen 40 sosyal bilimler alanında okuyan üniversite öğrencisinden 12'si düzenli olarak spor yaptığını söylemiştir. Buna göre sosyal bilimler alanında okuyan üniversite öğrencilerinden düzenli spor yapanların oranını %95 güvenle tahmin ediniz. Bu örneklemden kaynaklanan standart hatanın büyüklüğünü belirleyiniz.

$$\hat{p} = \frac{12}{40} = 0.3$$

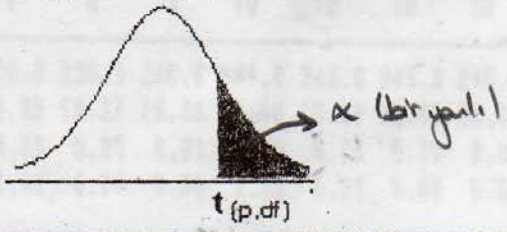
$$\sigma_{\hat{p}} = \sqrt{\frac{0.3 * 0.7}{40}} = 0.072$$

$$0.30 \pm 1.96(0.072)$$

$$0.16 \leq p \leq 0.44$$

Student'in t Tablosu

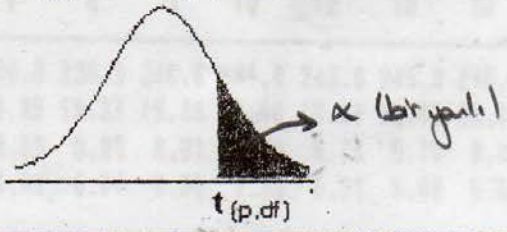
t tablosu sağ yan değerleri olasılıkları



2 yanlı alfa	*0.80*	*0.50*	*0.20*	*0.10*	*0.05*	*0.02*	*0.010*	*0.001*
Sd\1yanlı alfa	**0.40**	**0.25**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.0005**
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869
11	0.259556	0.697445	1.363430	1.795885	2.20099	2.71808	3.10581	4.4370
12	0.259033	0.695483	1.356217	1.782288	2.17881	2.68100	3.05454	4.3178
13	0.258591	0.693829	1.350171	1.770933	2.16037	2.65031	3.01228	4.2208
14	0.258213	0.692417	1.345030	1.761310	2.14479	2.62449	2.97684	4.1405
15	0.257885	0.691197	1.340606	1.753050	2.13145	2.60248	2.94671	4.0728
16	0.257599	0.690132	1.336757	1.745884	2.11991	2.58349	2.92078	4.0150
17	0.257347	0.689195	1.333379	1.739607	2.10982	2.56693	2.89823	3.9651
18	0.257123	0.688364	1.330391	1.734064	2.10092	2.55238	2.87844	3.9216
19	0.256923	0.687621	1.327728	1.729133	2.09302	2.53948	2.86093	3.8834
20	0.256743	0.686954	1.325341	1.724718	2.08596	2.52798	2.84534	3.8495
21	0.256580	0.686352	1.323188	1.720743	2.07961	2.51765	2.83136	3.8193
22	0.256432	0.685805	1.321237	1.717144	2.07387	2.50832	2.81876	3.7921
23	0.256297	0.685306	1.319460	1.713872	2.06866	2.49987	2.80734	3.7676
24	0.256173	0.684850	1.317836	1.710882	2.06390	2.49216	2.79694	3.7454
25	0.256060	0.684430	1.316345	1.708141	2.05954	2.48511	2.78744	3.7251
26	0.255955	0.684043	1.314972	1.705618	2.05553	2.47863	2.77871	3.7066
27	0.255858	0.683685	1.313703	1.703288	2.05183	2.47266	2.77068	3.6896
28	0.255768	0.683353	1.312527	1.701131	2.04841	2.46714	2.76326	3.6739
29	0.255684	0.683044	1.311434	1.699127	2.04523	2.46202	2.75639	3.6594
30	0.255605	0.682756	1.310415	1.697261	2.04227	2.45726	2.75000	3.6460
Sns.	0.253347	0.674490	1.281552	1.644854	1.95996	2.32635	2.57583	3.2905

Student'in t Tablosu

t tablosu sağ yan değerleri olasılıkları



2 yanlı alfa	*0.80*	*0.50*	*0.20*	*0.10*	*0.05*	*0.02*	*0.010*	*0.001*
Sd\1yanlı alfa	**0.40**	**0.25**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.0005**
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869
11	0.259556	0.697445	1.363430	1.795885	2.20099	2.71808	3.10581	4.4370
12	0.259033	0.695483	1.356217	1.782288	2.17881	2.68100	3.05454	4.3178
13	0.258591	0.693829	1.350171	1.770933	2.16037	2.65031	3.01228	4.2208
14	0.258213	0.692417	1.345030	1.761310	2.14479	2.62449	2.97684	4.1405
15	0.257885	0.691197	1.340606	1.753050	2.13145	2.60248	2.94671	4.0728
16	0.257599	0.690132	1.336757	1.745884	2.11991	2.58349	2.92078	4.0150
17	0.257347	0.689195	1.333379	1.739607	2.10982	2.56693	2.89823	3.9651
18	0.257123	0.688364	1.330391	1.734064	2.10092	2.55238	2.87844	3.9216
19	0.256923	0.687621	1.327728	1.729133	2.09302	2.53948	2.86093	3.8834
20	0.256743	0.686954	1.325341	1.724718	2.08596	2.52798	2.84534	3.8495
21	0.256580	0.686352	1.323188	1.720743	2.07961	2.51765	2.83136	3.8193
22	0.256432	0.685805	1.321237	1.717144	2.07387	2.50832	2.81876	3.7921
23	0.256297	0.685306	1.319460	1.713872	2.06866	2.49987	2.80734	3.7676
24	0.256173	0.684850	1.317836	1.710882	2.06390	2.49216	2.79694	3.7454
25	0.256060	0.684430	1.316345	1.708141	2.05954	2.48511	2.78744	3.7251
26	0.255955	0.684043	1.314972	1.705618	2.05553	2.47863	2.77871	3.7066
27	0.255858	0.683685	1.313703	1.703288	2.05183	2.47266	2.77068	3.6896
28	0.255768	0.683353	1.312527	1.701131	2.04841	2.46714	2.76326	3.6739
29	0.255684	0.683044	1.311434	1.699127	2.04523	2.46202	2.75639	3.6594
30	0.255605	0.682756	1.310415	1.697261	2.04227	2.45726	2.75000	3.6460
Sns.	0.253347	0.674490	1.281552	1.644854	1.95996	2.32635	2.57583	3.2905

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936

2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

Bölüm Soruları

1) Bir markette rastgele seçilen 6 şampuan kutusunun ağırlığı 330-308-298-301-295-289 ‘dur. Şampuan kutularının ortalam ağırlığını %99 güvenle tahmin ediniz.

2) Bir ilçede yalnız yaşayan 49 kişi ile görüşülmüş, görüşülen 49 kişinin aylık mutfak masraflarının ortalamasının 480 ytl ve varyansı da 2500 olarak hesaplanmıştır. Bu ilçede aylık mutfak harcalamalarını %96 güvenle tahmin ediniz.

3) Üniversitede okuyan 182 öğrenci ile görüşülmüş, görüşülen öğrencilerin 82’si düzenli spor yaptuğını söylemiştir. Üniversitede düzenli spor yapan öğrencilerin oranını %96 güven düzeyinde tahmin ediniz.

4) Nokta ve Aralık tahminleri arasındaki fark nedir?

5) Sizce güven düzeyi arttıkça aralık tahmini nasıl değişir?

BÖLÜM 3 ÇOKTAN SEÇMELİ SORULARI

1-Aşağıdakilerden hangisi tahmincilerin taşıması gereken özelliklerinden birisi değildir?

a)Sapmasızlık b)Tutarlılık c)Yeterlilik d)Etkinlik e) Sağlamlık

2-Aşağıdaki ifadelerden hangisi doğru değildir?

a) Rassal hata ele alınan örneklem nedeniyle açığa çıkar

b)Rassal hata istatistik ve parametre arasındaki farka karşılık gelir

c) Rassal hata kitle büyüdükçe azalır

d)Sistematik hata ölçüm aracının ayarsız olmasından kaynaklanır

e)Örneklem büyüklüğü arttıkça standart hata azalır

3- Bir fabrikada üretilen ürünlerin ortalama ağırlıkları araştırılmaktadır. Bu amaçla 100 birimlik örnek seçilmiş ve örnek ortalaması 20 gr olarak bulunmuştur. Kitle dağılımı normalse ve varyansı 16 gr ise, kitle ortalaması % 95 güvenle hangi aralıkta yer alır?

a) 19.22-20.79 b) 18.23-21.19 c)19.53- 22.03 d) 18.47- 20.98 e) 17.47-21.59

4- 5000 birimden oluşan bir kitlenin varyansı 1600'dür. Bu kitleden çekilen 400 birimin ortalaması 80 ise, %90 olasılıkla kitle ortalaması hangi aralıkta değerler alır?

a)70.84-87.68 b)76.84-83.17 c) 77.98-84.99 d) 78.69-86.66 e) 79.02- 87.48

5- Normal dağıldığı bilinen bir kitleden seçilen 16 birimin ağırlığı 30kg ve varyansı ise 12.96 kg olarak hesaplanmıştır. Buna göre kitle ortalaması %95 olasılıkla hangi aralıkta değerler alır?

a) 26.04-30.48 b) 27.47-31.56 c) 27.98- 32.99 d)28.08-31.92 e)29.01- 32.29

6- 10 hastanın A marka ilacı kullanmaları sonucunda iyileşme süreleri saat olarak şu şekilde 3,3,4,4,7,6,8,5,7,3 gözlenmiştir. Buna göre A marka ilacı kullananların ortalama iyileşme süresine ilişkin kitle ortalaması %99 olasılıkla hangi aralıkta değerler alır?

a)3.06-6.94 b) 2.57-7.02 c)2.47-7.35 d)2.36-7.96 e) 2.28-8.02

7- Bir markette yapılan araştırmaya göre 100 kişiden 60 'ı A marka şampuanları tercih etmektedir. Şampuan alan müşteri kitlesinde A markasını tercih edenlerin oranının % 95 olasılıkla tahmini aşağıdakilerden hangsine karşılık gelir?

a) 0.504-0.696 b)0.514-0.711 c)0.524-0.723 d)0.531-0.745 e)0.591-0.789

8)Bir ilçedeki 400 seçmenden 190'ı A partisini destekleyeceklerini söylemişlerdir. Buna göre A partisini destekleyenlerin oranı %99 olasılıkla aşağıdakilerden hangisine karşılık gelir?

a)0.38-0.52 b)0.41-0.54 c) 0.43-0.57 d)0.44-0.59 e)0.45-0.61

9) Standart sapması 6 olan normal dağılımdan seçilmiş rassal olarak seçilmiş 16 gözlemlin örnek ortalaması 25'dir. Kitle ortalaması μ için %90 güven aralığının üst sınırı aşağıdakilerden hangisidir?

a)27.47 b)28.04 c)28.96 d)29.02 e)29.87

10)Yukarıdaki soruda elde edilen alt sınır aşağıdakilerden hangisine eşittir?

a)22.53 b)23.04 c)23.68 d)24. 04 e)24.68

Yanıtlar

1- E 2- C 3- A 4-B 5-D 6-A 7-A 8-B 9-A 10-A

Bu Bölümde Ne Öğrendik Özeti

Bu bölümde anakütle parametrelerinin tahmin edilmesi sürecini değerlendirdik. Nokta tahmin ile aralık tahmini arasındaki farka değindik. Anakütle ortalamasının aralık tahmininde anakütle varyansının bilinip bilinmemesi ve örnek büyüklüğünün istatistik tablolarının seçiminde belirleyici olduğunu öğrendik.

4. 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

1) 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?

2) 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

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
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

- Ortalamalarası 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.

$$\textit{Örnek istatistikleri farkı} \mp \textit{tablo değeri} * \textit{s}\tan\textit{dart hata}$$

4.1. Ortalamalar Arası Farklar İçin Güven Aralığı

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

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

Eğer anakütle varyansları bilinmiyorsa bu durumda;

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

Her bir anakütleye 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.

$$\bar{X}_1=86 \qquad S_1=12 \qquad n_1=40$$
$$\bar{X}_2=72 \qquad S_2=14 \qquad n_2=35$$

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

$$P\left(6.18\le\mu_1-\mu_2\le21.82\right)=0.99$$

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.

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

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

$$8\pm5.62$$

$$2.40\le\mu_1-\mu_2\le13.62$$

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

4.2. İki Oran Farkı İçin Güven Aralığı

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

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

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.

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

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

$$= 0.018$$

$$\left(0.82 - 0.760\right) - 1.96 \times 0.018 \le \mathrm{P}_1 - \mathrm{P}_2 \le \left(0.82 - 0.760\right) + 1.96 \times 0.018$$

$$0.029 \le \mathrm{P}_1 - \mathrm{P}_2 \le 0.10$$

%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?

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

$$0.175 - 1.645 \cdot \sqrt{\frac{.175 \cdot (.825)}{200}} \le p \le 0.175 + 1.645 \cdot \sqrt{\frac{.175 \cdot (.825)}{200}}$$

$$0.1308 \le p \le 0.2192$$

Örnek:

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

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

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

$$0.053 \le p \le 0.107$$

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?

$$P\left(\overline{X} - z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}} \le \mu_X \le \overline{X} + z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}}\right) = 1-\alpha$$

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

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?

$$P(\overline{X} - Z_{\alpha/2}\frac{S_x}{\sqrt{n}} \le \mu \le \overline{X} + Z_{\alpha/2}\frac{S_x}{\sqrt{n}}) = 1-\alpha$$

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

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 ilşkin 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)

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

$$0.906 \le \mu_1-\mu_2 \le 1.093$$

Aralık sıfır değerini kapsamamaktadır, iki ilacın tedavi süreleri açısısından 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 ver erkeklerin oranı arasında istatistiksel açıdan önemli bir fark var mıdır?

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

$$0.2632 \le p_1-p_2 \le 0.4363$$

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 yarışı 3 dakikanın altında bir sürede tamamlamıştır. Kitle oran farkı için %70 güven düzeyinde aralığı oluştururak, iki cins at arasında fark var mıdır değerlendiriniz.

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

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

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

$$0.1009 \le p_1-p_2 \le 0.1103$$

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

Örnek:

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

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

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

$$0.0249 \le p_1 - p_2 \le 0.1951$$

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 erkekten 60'ı düzenli kitap okumaktadır. Kitle oran farkları için güven sınırlarını %86 olasılıkla belirleyiniz.

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

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

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

$$0.3841 \le p_1 - p_2 \le 0.4345$$

4.3. Örneklem Büyüklüğünün Belirlenmesi ve Hata

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

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

$e = tablo\ değeri * s\tan dart\ hata$

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

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

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

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

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

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 ihitiyaç duyulurdu?

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

$$0.58 - 1.56\sqrt{\frac{0.58x0.42}{100}} \le p \le 0.58 + 1.56\sqrt{\frac{0.58x0.42}{100}}$$

$$0.58 - 0.07644 \le p \le 0.58 + 0.07644$$

$$\square \ hata \qquad \square \ hata$$

$$0.50356 \le p \le 0.6564$$

$$hata = 0.08 \quad olursa$$

$$n = z_{\alpha/2}^2 \frac{pq}{hata^2} = 1.56^2 \frac{0.58x0.42}{0.08^2} = 92.6 \square 93$$

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)

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

$$-2.078 \le \mu_1-\mu_2 \le -1.677$$

Aralık 0 değerini kapsamadığından farkın istatistiksel açıdan anlamlı olduğu söylenebilir.

Örnek: İki ayrı atölyede üretilen ürünlerin ortalama ağırlıkları arasında fark olup olmadığının sınanması için A atölyesinden 36 ve B atölyesinden 49 ürün seçilmiştir. A atölyesinde üretilen ürünlerin ortalama ağırlıkları 3.5kg ve sapması 1.5kg; B atölyesinde üretilen ürünlerin ağırlıkları 2.5 kg ve sapması 1kg'dır. % 95 güven düzeyinde kitle ortalamaları arasındaki fark için güven aralığını oluşturunuz. Ortalama ağırlıklar açısından istatistiksel olarak anlamlı bir fark var mıdır?

Fark yoktur,

$$(3.5-2.5)\pm 1.96\sqrt{\frac{1.5^2}{36}+\frac{1}{49}}$$

$$0.44 \le \mu_1-\mu_2 \le 1.56$$

Örnek:

Rastgele seçilen 100 erkek lise öğrencisinden 40'ı ve 80 kız lise öğrencisinden 40'ı düzenli olarak roman okuduğunu söylemiştir. Kız ve erkek öğrencilerin roman okuma oranları arasında istatistiksel açıdan önemli bir fark var mıdır (%95 güven düzeyinde değerlendiriniz)?

$$(0.5-0.4)\pm 1.96\sqrt{\frac{0.4*0.6}{100}+\frac{0.5*0.5}{80}}$$

$$0.09 \le p_1-p_2 \le 0.11$$

Bölüm Soruları

1-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 başarısı 88, notlarının varyansı 121 olarak hesaplanmıştır. Buna göre söz konusu 2 yıldaki ortalama başarı arasında bir fark var mıdır?

BÖLÜM 4 ÇOKTAN SEÇMELİ SORULARI

1- A ve B okullarında okuyan öğrencilerin günlük harçlıklarının dağılımının normal olduğu bilinmektedir. Öğrencilerin harçlıkları şöyledir:
A: 15,15,17,18,20,20,22 B:7,9,9,10,12,12,15,16
A ve B okullarında okuyan öğrencilerin ortalama harçlıkları arasındaki fark için %95 olasılıkla oluşturulacak güven aralığı hangisine eşittir?

a) 2.7-9.3 b)2.8-9.4 c)2.8- 9.5 d) 2.9- 9.8 e)3.1-10.2

2-Yukarıdaki sorudan hareketle nasıl bir yorum yapılır?

A ve B okullarında okuyan öğrencilerin ortalama harçlıkları arasındaki fark istatistiksel olarak anlamlıdır, A okullarında okuyan öğrencilerin ortalama harçlığı B'de okuyanlardan daha fazladır.

3- A ürününden 20 ve B ürününden 25 tane örnek seçilmiştir. A ürünlerinin ortalaması 30 birim ve varyansı 18 birim, B ürünlerininse ortalaması 25 birim ve varyansı 12 birim olarak

hesaplanmıştır. A ve B ürünlerinin ortalamaları arasındaki fark için %95olasılıkla güven sınırları aşağıdakilerden hangisine eşittir?

a) 2.5- 6.9 b)2.7- 7.3 c)2.8- 7.4 d)2.9- 7.5 e)3.1-7.8

4-Yukarıdaki sorudan hareketle nasıl bir yorum yapılır?

A ve B ürünlerinin ortalamaları arasındaki fark istatistiksel olarak anlamlıdır, A ürünün ortalaması daha yüksektir.

5- A ürününden 75 ve B ürününden 128 tane örnek seçilmiştir. A ürünlerinin ortalaması 82 birim ve standart sapması 22 birim, B ürünlerininse ortalaması 74 birim ve standart sapması 35 birim olarak hesaplanmıştır. A ve B ürünlerinin ortalamaları arasındaki fark için %95 güven sınırları aşağıdakilerden hangisine eşittir?

a)-0.30-18.02 b)-0.25 -16.12 c) -0.20- 16.21 d) -0.15- 15.89 e) -0.15- 16.82

6-Yukarıdaki sorudan hareketle nasıl bir yorum yapılır?

A ve B ürünlerinin ortalamaları arasında istatistiksel açıdan fark yoktur.

7- A ilacını kullanan hastalardan tesadüfi olarak seçilen 100 kişide 60'ı ve B ilacını kullanan hastalardan tesadüfi olarak seçilen 100 kişiden 80'i iyileşmiştir. Aşağıdakilerden hangisi %95 olasılıkla A ve B ilaçlarının iyileşme oranları arasındaki farka karşılık gelir?

a) -0.47-(-0.20) b)-0.45-(-0.15) c)-0.32-(-0.08) d)-0.20-0.12 e)-0.18-0.20

8-Yukarıdaki sorudan hareketle nasıl bir yorum yapılır?

A ve B ilacını kullanarak iyileşenlerin oranları arasında istatistiksel açıdan fark vardır. B ilacının etkinliğinin daha fazla olduğu söylenebilir.

9-A tekniğini kullanarak yabancı dil öğrenen 200 kişiden %40 'ı ve B tekniğini kullanarak yabancı dil öğrenen 400 kişiden %15'i girdikleri sınavda başarılı olmuştur. A ve B tekniklerinin başarı oranları arasındaki fark için %99 güvenle tahmin aşağıdakilerden hangisine karşılık gelir?

a) 0.15-0.35 b) 0.20-0.40 c)0.25-0.45 d)0.30-0.45 e)0.32-0.64

10- Yukarıdaki sorudan hareketle nasıl bir yorum yapılır?

Her iki teknik arasındaki fark istatistiksel açıdan anlamlıdır. A tekniğinin başarı oranı daha yüksektir.

Yanıtlar

1)A 3)B 5)A 7)C 9)A

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.

5. HİPOTEZ TESTLERİ

Bu Bölümde Neler Öğreneceğiz?

Bu bölümde istatistiksel çıkarsamanın 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

1) Seçimlerde A partisinin oy oranının %47 olacağı iddia ediliyor. Görüşülen 180 kişiden 85'i A partisini destekleyeceğini söylediğine göre bu iddia desteklenebilir mi?

2) A marka ampulun ortalama ömrünün 182 saatten fazla olduğu iddia ediliyor. Rastgele seçilen 28 ampulun ortalama ömrü 186 saat,standart sapması 22 saat olarak bulunduğuna 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
- Anlamlılılık düzeyi
- Ret bölgesi
- Kabul alanı
- Alternatif hipotez, sıfır hipotezi

Giriş

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

5.1. Hipotez Testleri

İstatiksel çıkarsamanın 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 sınanır.

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 örnekleme dağılımıyla ilgili bilgilerden yararlanarak araştıran önermedir.

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

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.

5.2. 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.

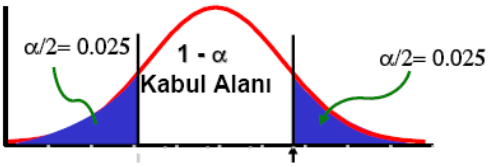
Üçüncü alternatif hipotezse anakütle parametre değerinden büyük yönde anlamlı farklılığa karşılık gelir.

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

Hipotez Çiftleri

Çift yönlü hipotez

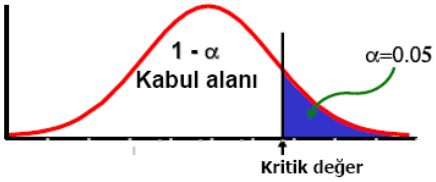
$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$



Tek yönlü hipotez

(Sağ taraf testi)

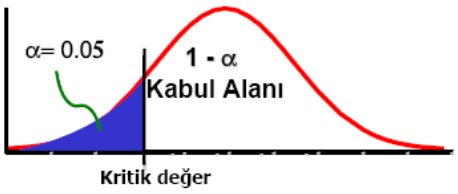
$$H_0 : \mu \leq \mu_0$$
$$H_1 : \mu > \mu_0$$



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

5.3. I. Tür ve II. Tür Hata

Aynı anakütleden çekilen örneklemler 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.

	H0 Gerçek	H0 Hatalı
H0 kabul	Doğru karar	II .tür hata β
H0 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

5.4. Anakütle Ortalamasına İlişkin Hipotez Testleri

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

$$z = \frac{\overline{x} - \mu}{\sigma / \sqrt{n}}$$ σ^2 biliniyorsa Z istatistiği

$$z = \frac{\overline{x} - \mu}{s / \sqrt{n}}$$ σ^2 bilinmiyorsa n>30 ise Z istatistiği

$$t = \frac{\overline{x} - \mu}{s / \sqrt{n}}$$ σ^2 bilinmiyorsa n< 30 ise t istatistiği

Örnek:

Bir fabrikada üretilen margarin paketlerinin ortalama ağırlığı 100 gramdır, ve standart sapması 2 olan normal dağılım göstermektedir. Ratgele olarak seçilen 9 paketin ortalama ağırlığı 102 gram olarak ölçülmüştür. Paketlerin ağrılığının 100 gramdan farklı olduğu söylenebilir mi?(α=0.05)

-Öncelikle hipotezler kurulur

$$H_0 : \mu = 100gr$$
$$H_1 : \mu \neq 100gr$$

-

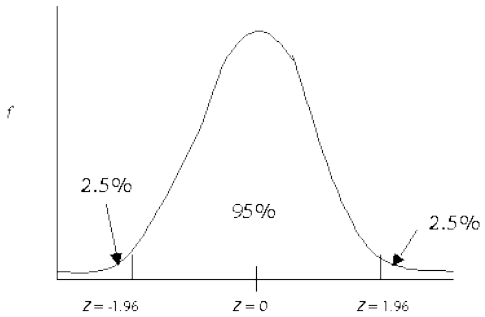
$$\mu = 100$$
$$\overline{x} = 102$$
$$\sigma = 2$$
$$n = 9$$

-

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

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.

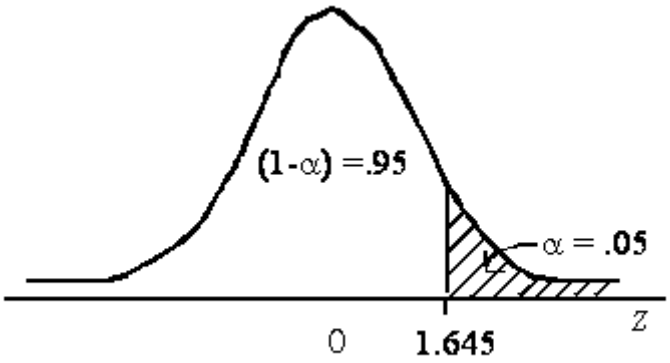
Toplam alan 1'dir. Sağ ve sol tarafta kalan alanların 0.5 olarak düşünülür. Tablo değeri incelenirken 0.5 -0.025=0.475 değerine karşılık gelen değer alınır. Bu değer 1.96'dır.



Hesaplanan Z değeri kritik değerden büyüktür. $Z_{hes} > Z_{tablo}$ (3>1.96) olduğundan H0 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 α=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,

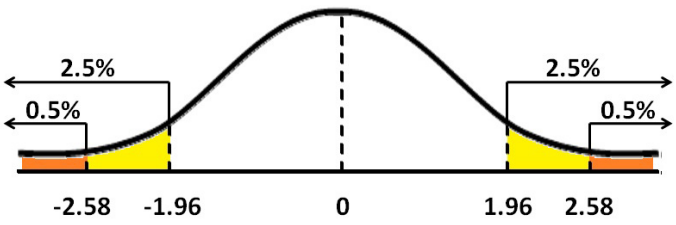
0.5-0.05=0.45 olacaktır.

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

Standart Z tablosunda bulunan kritik değer 1.645'tir. Hesaplanan Z değeri 1.5 bu değerin solunda yani kabul alanı içindedir. H0 hipotezi kabul edilir yani zeytinyağı tenekelerinin ağırlığı 3680 gramdan fazladır iddiasını kabul için yeterli örnek yoktur.

Örnek: Matematik bölümü öğrencilerinin İstatistik dersinden aldıkları notların ortalamasının 70 olduğu 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 karşılk gellen değer 2.58'dir.

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

Hesaplanan değer sol tarafta red alanının içindedir. -3< -2.58. H0 hipotezi reddedilir.

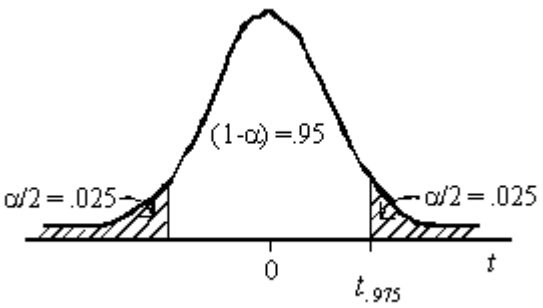
Örnek:

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 'dan farklı olduğu söylenebilir mi? (0.05 önem düzeyi)

$$H_0 : \mu = 250$$
$$H_1 : \mu \neq 250$$

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

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



(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 İçin Hipotez Testi

Çift taraflı test

$$H_0 : P = P_0$$
$$H_1 : P \neq P_0$$

Tek taraflı testler

$$H_0 : P \geq P_0$$
$$H_1 : P < P_0$$

$$H_0 : P \leq P_0$$
$$H_1 : P > P_0$$

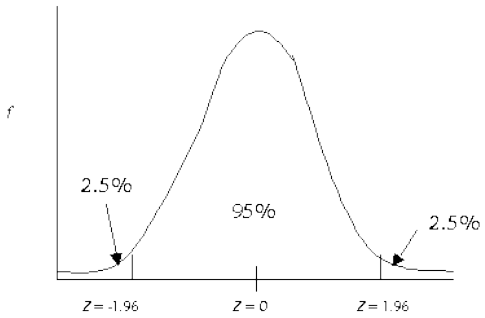
Test istatistiği

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

Örnek: Bir dershane 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.

$$H_0 : P = P_0$$
$$H_1 : P \neq P_0$$

$$p = 184 / 200 = 0.92$$



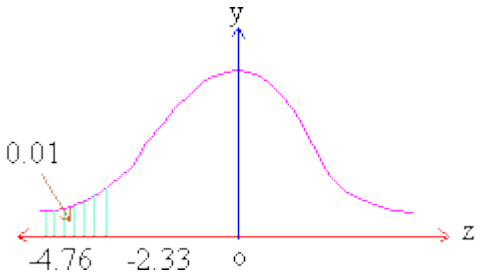
$$Z = \frac{p-P}{\sigma_p} = \frac{p-P}{\sqrt{\frac{P(1-P)}{n}}} = \frac{0.8-0.9}{\sqrt{\frac{0.9(1-0.9)}{200}}} = -1.76$$

Bulunan test istatistiği tablo değerine oldukça yakındır kabul alanı içerisinde görünmektedir. Eğer anlam düzeyi 0.01 olsaydı bu durumda hipotez yine kabul alanı içerisinde olacaktı.

Örnek:

A partisinin seçimlerde alacağı oy oranının %90'as olacağı iddia edilmektedir. Rastgele seçilen 200 kişiye hangi partiye oy verecekleri sorulmuş, bu kişilerden 180'nın 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. farklı olduğu söylenebilir mi?

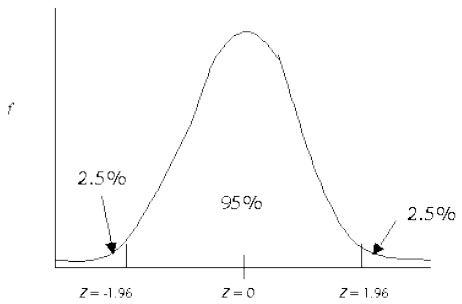
$$H_0 : \mu = 500gr$$
$$H_1 : \mu \neq 500gr$$

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

$$H_0 \quad reddedilir$$

Örnek

Bir şirkette çalışanlara verilen günlü 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.



$$H_0 : \mu = 20$$
$$H_1 : \mu \neq 20$$
$$\overline{X} = \frac{15 + 25 + 10 + 12 + 30}{5} = 18.4$$
$$s = \sqrt{\frac{(15 - 18.4)^2 + ... + (30 - 18.4)^2}{4}} = 8.67$$

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

İddia kabul edilir, -0.412 değeri kabul bölgesi içerisindedir.

Bölüm Soruları

1) Üniversitede okuyan öğrencilerden düzenli spor yapanların oranın %40'dan fazla olduğu iddiasını araştırmak amacıyla görüşülen 200 öğrenciden 88'i düzenli spor yapma alışkanlığının olduğunu söylemiştir. İddiayı α =0,02 sınayınız. 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 sınayınız.

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 sınayınız

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.

BÖLÜM 5 ÇOKTAN SEÇMELİ SORULARI

1-Aşağıdaki ifadelerden hangisi doğrudur?

I) İstatistiksel hipotez testinde örneklem verilerinden hareketle hesaplanan istatistikle, kitlenin bilinen ya da varsayımsal değeri arasındaki farkın istatistiksel olarak anlamlı olup olmadığı sınanır.
II) İstatistiksel hipotez testinde örneklem istatistiğinin belirli bir değere eşit olup olmadığı sınanır
III) Hipotez testlerinde test edilen alternatif hipotezin doğruluğudur.
a)Yalnız I b) Yalnız II c)Yalnız III d) I ve II e) I ve III

2- Hipotez test sürecinin aşamaları için aşağıdaki sıralamalardan hangisi doğrudur?

I) Sıfır ve alternatif hipotezlerin ifade edilmesi
II) Örneklem istatistiğinin standart hale dönüştürülmesi
III) Anlamlılık düzeyinin belirlenmesi
IV) Kitleden örneklemin çekilmesi
V) İstatistik kararın verilmesi
a)I –II-III-IV-V b)I-III-IV-II-V c) III-II-I-IV-V d)IV-I-III-II-V

3-Gerçekte doğru olan sıfır hipotezin yanlışlıkla reddedilmesi durumunda işlenen hataya ne denir?

a) Standart hata b)I . Tip hata c)II. Tip hata d) III. Tip hata e)IV. Tip hata

4-Gerçekte yanlış olan sıfır hipotezinin yanlışlıkla kabul edilmesi durumunda işlenen hataya ne denir?

a) Standart hata b)I . Tip hata c)II. Tip hata d) III. Tip hata e)IV. Tip hata

5-Örneklem istatistiğinin normal dağıldığı tek yönlü üst kuyruk testinde, %1 önem düzeyinde yapılacak testte kabul bölgesinin oransal büyüklüğü ne olur?

a) 0.01 b) 0.05 c) 0.90 d) 0.95 e)0.99

6- Aşağıdaki adımlardan hangisi test sürecinin adımlarından biri olarak kabul edilemez?

a) Hipotezlerin yazılması b)Anlamlılık düzeyinin belirlenmesi c)Örneklemin seçilmesi d)İstatistiksel kararın verilmesi e)Tamsayım yapılması

7-Normal dağılıma sahip bir kitleden rassal olarak seçilen 100 birimlik örneklemin ortalamasının standart hatası 0.25 ise bu dağılımın standart sapması aşağıdakilerden hangisine eşittir?

A) 2.25 b) 2.5 c) 4.5 d)5.5 e)10.5

8- Normal dağılan bir kitlenin ortalamasının 55 olup olmadığının sınanmak için seçilen örneğin ortalaması 58, standart hata ise 1.3 olarak bulunmuştur. Standart z değerinin sağında kalan bölgenin alanı kaçtır?

a)0.02 b) 0.01 c)0.30 d)0.49 e)0.99

9- Kitle ortalamasının 100 birim olup olmadığının %95 güvenle test etmek için seçilen 25 birimlik örneklemin ortalaması 85 ve standart sapması 50 olarak bulunmuştur. Örnek ortalamasını temsil esen test istatistiğinin değeri aşağıdakilerden hangisidir?

a)-3 b) -2 c)-1.5 d)-1 e)1

10- Kitle ortalamasının 100 birim olup olmadığının %95 güvenle test etmek için seçilen 25 birimlik örneklemin ortalaması 120 ve varyansı 2500 olarak bulunmuştur. Örnek ortalamasını temsil esen test istatistiğinin değeri aşağıdakilerden hangisidir?

a)3 b) 2 c)1.5 d)1 e)-1

Yanıtlar

1) A 2) B 3)B 4)C 5) E 6)E 7)B 8)D 9)C 10)B

Bu Bölümde Ne Öğrendik Özeti

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

6. HİPOTEZ TESTLERİ (İKİ ANAKÜTLE)

Bu Bölümde Neler Öğreneceğiz?

Bu bölümde iki anakütle parametreleri arasındaki farkı dikkate alarak hipotez testlerini yeniden gözden geçireceğiz. Yanısıra tek örneklemde küçük örneklem testlerini de ele alacağız.

Bölüm Hakkında İlgi Oluşturan Sorular

1) Bir toplulukta sigara içen kadınların oranının sigara içen erkeklerin oranından daha az olduğu iddiasını sınamak.

2) İşletme ve İktisat bölümlerinde okutulan matematik dersin işletme öğrencilerinin not ortalamasının daha yüksek olduğu iddiasını sınamak

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	**Kazanım**	**Kazanımın nasıl elde edileceği veya geliştirileceği**
İki anakütle olması halinde anakütle parametrelerinin farklarına ilişkin hipotez testleri ve tek anakütle olması halinde hipotez testleri	Farklı anakütle parametrelerinin farklarına ilişkin iddiaları sınama	Metinler, çözümlü problemler.

Anahtar Kavramlar

- Sıfır ve alternatif hipotezler
- T istatistiği
- Ortalama farkları
- Oran farkları

Giriş

Bu bölümde iki anakütle olması halinde ilgilenilen parametre farklarına ilişkin hipotez testlerinden ve küçük örnek testlerinden söz edilecektir.

6.1. Ortalamalar Arası Farklara İlişkin Testler

Eğer iki anakütleye ilişkin hipotezler test edilecekse hipotezler şu şekilde olacaktır:

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

Çift taraflı hipotez iki anakütle ortalamalarının eşitliği iddiası

$$H_0 : \mu_1 \geq \mu_2$$
$$H_1 : \mu_1 < \mu_2$$

Tek taraflı hipotez testleri

$$H_0 : \mu_1 \leq \mu_2$$
$$H_1 : \mu_1 > \mu_2$$

Kullanılacak test istatistiği de aşağıdaki gibidir.

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}} = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Örnek hacimlerinin toplanı n1+n2< 30 ise

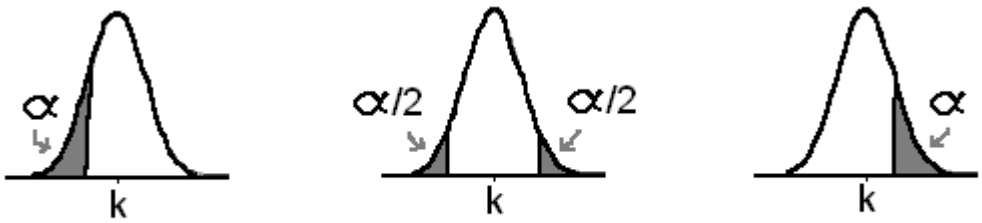
$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Örnek:

Konya ve İstanbul'da yaşayan 300'er ev hanımı ile görüşülerek haftalık mutfak harcamaları soruluyor. Konya'da yaşayan ev hanımlarının aylık mutfak harcamalarının ortalamasını 600 ytl ve sapması da 100 ytl olarak hesaplanmıştır. İstanbul'da yaşayan ev hanımlarının aylık ortalama mutfak harcamalarının 700ytl ve standart sapmaları da 150 ytl olarak hesaplanmıştır. Konya ve İstanbul'da yaşayan ev hanımlarının ortalama mutfak harcamaları arasında fark var mıdır? (0.01 anlamlılık düzeyi)

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

Çift yönlü hipotez testi anlamlalık düzeyinin 0.01 olmasından ötürü sağ ve sol kısımdaki ret alanları 0.005 olacaktır. 0.5-0.005 değerine karşılık gelen tablo değeri 2.58 olacaktır. Eğer alternatif hipotez tek yönlü olsaydı tablo değeri 2.33 olacaktı. Aşağıdaki gösterim, standart normal dağılım tablosundan değerleri bularak değerlendirmekte yardımcı olacaktır.



$H_0: \mu = k$ $H_1: \mu < k$

α	z critical
0.10	-1.28
0.05	-1.65
0.01	-2.33

$H_0: \mu = k$ $H_1: \mu \neq k$

α	z critical
0.10	± 1.65
0.05	± 1.96
0.01	± 2.58

$H_0: \mu = k$ $H_1: \mu > k$

α	z critical
0.10	1.28
0.05	1.65
0.01	2.33

$$\sigma_{\bar{x}1-\bar{x}2} = \sqrt{\frac{100^2}{300} + \frac{150^2}{300}} = 10.04$$

$$z = \frac{700-600}{10.04} = 10$$

Hesaplanan z değeri tablo değerinden büyüktür, sıfır hipotezi reddedilir. 10>2.58

İstanbul ve Konya'da ailelerin aylık mutfak harcamaları farklıdır.

6.2. Oranlararası Farklara İlişkin Hipotez Testleri

İki farklı anakütleden n1 ve n2 büyüklüğünde çekilen örneklemlerin oranları sırasıyla p1 ve p2 olsun. Belirlenen anlamlılık düzeyinde söz konusu örneklemlerin aynı anakütleden gelip gelmediğinin araştırılmasında yine hipotez testlerine başvurulur.

Alternatif hipotezin yönüne göre üç farklı durum söz konusudur.

$$H_0 : P_1 = P_2$$
$$H_1 : P_1 \neq P_2$$

$$H_0 : P_1 \geq P_2$$
$$H_1 : P_1 < P_2$$

$$H_0 : P_1 \leq P_2$$
$$H_1 : P_1 > P_2$$

Test istatistiği ise şu şekilde hesaplanır:

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

Örnek: Bir işletmede çalışan kadın ve erkeklerin sigara içme oranlarını değerlendirmek için 60 erkekle ve 40 kadın ile görüşülüyor. Görüşme sonucunda 51 erkek ve 20 kadın sigara kullandığı bilgisine ulaşılıyor. Erkeklerin sigara içme oranının kadınlardan sigara içme oranından yüksek olduğu söylenebilir mi? (%5 önem düzeyinde değerlendiriniz)

$$H_0 : P_1 \leq P_2$$
$$H_1 : P_1 > P_2$$

erkeklerde sigara içme oranı $p_1 = \frac{51}{60} = 0.85$

kadınlarda sigara içme oranı $p_2 = \frac{20}{40} = 0.50$

%5 anlamlılık düzeyinde tek taraflı hipotez testi için tablo değeri Z_{tab}= 1.65 olacaktır.

$$Z_h = \frac{0.85 - 050}{\sqrt{\frac{0.85x0.15}{60} + \frac{0.5x0.5}{40}}} = 3.78$$

Sıfır hipotezi reddedilir. Erkeklerin sigara içme oranı kadınlarınkinden fazladır.

Örnek: Ampül üreten iki firmanın ürettikleri ampüllerin ortalama dayanma süresi araştırılmak isteniliyor. A fabrikasından seçilen 80 ampülün ortalama dayanma süresi 135 gün ve standart sapması 15 gün; B fabrikasında seçilen 90 ampülün ortalama dayanma süresi 130 gün ve standart sapması 18 gündür. %1 anlamlılık düzeyinde A fabrikasında üretilen ampüllerin daha dayanıklı olduğu söylenilebilir mi?

$$H_0 : \mu_1 \equiv \mu_2$$
$$H_1 : \mu_1 > \mu_2$$

%1 anlamlılık düzeyi, tek yönlü hipotez testi z tablo değeri 2.33.

$$Z_h = \frac{(\overline{X}_1 - \overline{X}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(135 - 130)}{\sqrt{\frac{15^2}{80} + \frac{18^2}{90}}} = 1.95$$

Tablo değeri hesap değerinden büyüktür sıfır hipotezi kabul edilir. Ampüllerin dayanıklılık süreleri arasında fark yoktur.

Küçük Örneklerle çalışılması halinde z yerine t testi kullanılmalıdır.

t Dağılımı

- Küçük örneklerden (n<30) elde edilen istatistiklerin dağılımı Student t dağılımına uyar.

- Küçük örnek istatistiklerinin gösterdiği dağılım normal eğri gibi simetriktir.Normal eğriye göre daha basık ve yaygın bir şekil alır. Böylece eğrinin kuyruklarında daha büyük bir alan oluşur.

- Küçük örnekler için z cetveli yerine, çeşitli örnek büyüklükleri ve olasılık seviyeleri için ayrı ayrı hesaplanmış t cetvelleri kullanılır.

Ortalamalarla ilgili hipotez testinde n<30 iken t dağılımı kullanılmalıdır.

$$t_h = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$$

Örnek:

Bir markette satılan margarinlerin paketlerinin üzerinde 455 gr yazmaktadır. Rastgele seçilen 17 paketin ortalama ağırlığı 450 gr ve standart sapması 13 gr bulunmuştur. %5 anlamlılık düzeyinde paket ağırlığının 455 gr olduğu iddiası kabul edilebilir mi?

$$H_0 : \mu = 455$$
$$H_1 : \mu \neq 455$$

$$n = 17$$
$$\overline{X} = 450 gr.$$
$$s = 13 gr.$$

Tablodan kritik değer bulunurken n-1 serbestlik derecesi dikkate alınmalıdır. 16 serbestlik derecesiyle 0.05 anlamlılık düzeyinde t tablo değeri 2.12'dir.

$$t_h = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}} = \frac{450 - 455}{\frac{13}{\sqrt{17}}} \cong 1.5$$

Hesaplanan t istatistiği tablo değerinden küçüktür. Sıfır hipotezi kabul edilir, yani paketlerin ağırlığının 455 gr. olduğu iddiası geçerlidir.

Örnek:

A üniversitesi, İktisat Bölümü öğrencilerinin istatistik dersi ortalama başarısı 2010 yılında 65 puandır. 2011 yılında mezun olan 26 öğrenci restgele seçilmiş ve ortalama başarı puanlarının 70, standart sapmasının 10 puan olduğu hesaplanmıştır. Buna göre öğrencilerin başarıları arasında her iki dönem dikkatte alındığında fark var mıdır %1 anlamlılık düzeyinde test edin.

$$H_0 : \mu = 65$$
$$H_1 : \mu \neq 65$$

26-1=25 serbestlik derecesi %1 anlam düzeyinde t tablo değeri 2.787 'dir.

$$t = \frac{70-65}{10/\sqrt{26}} = 2.5$$

Hesaplanan t istatistği tablo değerinden küçüktür. 2010 ve 2011 yıllarındaki ortalama başarı oranları arasında anlamlı bir fark yoktur.

Örnek: Uyumadan önce diş ipi kullananların oranın %30 olduğu iddia edilmektedir. Rastgele seçilen 100 kişiden 28'i diş ipi kullandığını söylemiştir. Buna göre söz konusu iddiayı %1 anlamlılık düzeyinde sınayınız.

$$H_0 : p = 0.3$$
$$H_0 : p \neq 0.3$$
$$\hat{p} = 28/100 = 0.28$$
$$z = \frac{0.28-0.3}{\sqrt{\frac{0.3*0.7}{100}}} = \frac{-0.02}{0.0458} = -0.44$$

İddia kabul edilir.

Örnek: Bir bölgede yaşayan dört kişilik ailelerin aylık mutfak harcamalarının en az 700 tl olduğu iddia edilmektedir. Rastgele seçilen 6 ailenin harcamalarının şöyle olduğu belirlenmiştir. 620, 920, 750, 680, 830, 950. Buna göre 0.05 önem düzeyinde iddia kabul edilebilir mi?

5 s.d. t tablo değeri 2.01 olduğundan iddia kabul edilemez

$$Ha : \mu > 700$$
$$\bar{X} = 791.8$$
$$s = 13.2$$
$$t = \frac{791.8-700}{13.2/\sqrt{5}} = 1.7$$

Örnek:Yurtta kalan üniversite öğrencilerinin aylık harcamalarının en az 1200 tl olduğu iddiasını sınamak amacıyla yurtta kalan öğrencilerden rassal olarak 10 kişi seçiliyor. Elde edilen kayıtlara göre öğrencilerin aylık harcamaları şöyledir: 1390, 1340, 1240, 1080,1590,1630,1150,1380,1060, 1240. Buna göre 0.01 önem düzeyinde iddia doğrulanır mı?

Tek taraflı tablo değeri 2.82 olduğundan 1200'den fazla olduğunu söylemek için yeterli kanıt yoktur.

$$Ha : \mu > 1200$$
$$\bar{X} = 1310$$
$$s = 19.5$$
$$t = \frac{1310-700}{19.5/\sqrt{10}} = 1.8$$

Örnek: Bir eğitim programının işletme ve iktisat bölümü öğrencileri üzerindeki etkisini değerlendirmek amacıyla rastgele seçilen 100 işletme öğrencisinin notlarının ortalaması 53 ve varyansı 9 olarak hesaplanmıştır. Benzer şekilde seçilen 50 iktisat öğrencisinin notlarının ortalaması 55 ve varyansı 25 olarak bulunmuştur. %1 önem düzeyinde iktisat öğrencilerinin daha başarılı olduğu söylenebilir mi?

Kritik tablo değeri 2.33<3 olduğu için iktisat öğrencilerinin daha başarılı olduğu söylenebilir, sıfır hipotezi reddedilir.

$$Ha: \mu_{iktisat} > \mu_{işletme}$$

$$z = \frac{55-53}{\sqrt{\frac{16}{50}+\frac{9}{100}}} = 3$$

BÖLÜM 6 SORULARI

A) Aşağıdaki soru metninden hareketle ilk 3 soruyu yanıtlayınız.
Bir fabrikada paketlenen ürünlerin ortalama ağırlığı 80 gr ve standart sapması 20gr'dır. Rassal olarak seçilen 100 paketin ortalama ağırlığı 60gr olarak bulunmuştur. Ürünlerin ortalamasının 80gr'dan az olduğu iddiasını % 5 önem düzeyinde araştırınız.

1) Örnekleme dağılımının red bölgesi aşağıdakilerden hangisidir?
   a) Sol uçta %2.5 'lik alan
   b) Sol uçta %5'lik alan
   c) Sol uçta %10'luk alan
   d) Sağ uçta %2.5'luk alan
   e) Sağ uçta %5'lik alan

2) Bu sınamada alternatif hipotez nedir?
   $a) \mu > 80$
   $b) \mu < 80$
   $c) \mu = 80$
   $d) \overline{X} < 80$
   $e) \overline{X} > 80$
3) Örneklem ortalamasını temsil eden test skoru nedir?
   a) -2.5 b) -2 c)-1 d)0 e)1

B)4-5-6 numaralı soruları aşağıdaki soru metninden hareketle çözünüz.

Bir fabrikada üretilen margarin paketlerinin ağırlığı 250gramdır. Normal dağılan bu kitleden çekilen örneklemin büyüklüğü 25 birimdir. Örneklem ortalaması 220 gram ve standart

sapması 15 olarak hesaplanmıştır.% 95 güven düzeyinde kitle ortalaması ile örneklem ortalaması arasındaki farkın anlamlı olup olmadığını sınanacaktır.

4)Örnekleme dağılımının ret bölgesi aşağıdakilerden hangisidir?

a) Sağ ve sol uçta %5'lik alan

b)Sağ ve sol uçta % 1'lik alan

c)Sağ ve sol uçta %2.5'luk alan

d)Sağ uçta %5 lik alan

e)Sol uçta %5'lik alan

5) Bu sınamada alternatif hipotez aşağıdakilerden hangisidir?

$a)\mu > 250$
$b)\mu < 250$
$c)\mu \neq 250$
$d)\overline{X} < 250$
$e)\overline{X} > 250$

6) Örneklem ortalamasını temsil eden skor aşağıdakilerden hangisidir?

a) -2.5 b) -2 c)-1 d)0 e)1

C)7-8-9-1 0 numaralı soruları aşağıdaki metinden hareketle yanıtlayınız.

Bir bölgedeki seçmenlerin % 90'nının A partisini desteklediği bilinmektedir. Rassal olarak seçilen 25 kişinin %80 'ninin A partisini desteklediği sonucuna ulaşılmıştır. %1 önem düzeyinde bu bölgedeki seçmenlerin %90'nından daha azının A partisini desteklediği iddiası desteklenebilir mi araştırılacaktır.

7- Örnekleme dağılımının ret bölgesi aşağıdakilerden hangisidir?

a) Sağ ve sol uçta %5'lik alan

b)Sağ ve sol uçta % 1'lik alan

c)Sağ ve sol uçta %2.5'luk alan

d)Sağ uçta %10 luk alan

e)Sol uçta %10'lik alan

8) Bu sınamada alternatif hipotez aşağıdakilerden hangisidir?

$a)\pi > 0.90$
$b)\pi < 0.80$
$c)\pi < 0.90$
$d)\, p < 0.90$
$e)\, p > 0.80$

9)Örneklem oranını temsil eden skor aşağıdakilerden hangisidir?

a) -1.67 b) -1.5 c) -1 d) 1.67 e)2

10) İddia geçerli midir?

%1 önem düzeyinde -1.67 > -2.33 olduğundan iddia geçerli değildir.

Yanıtlar

1)B 2)B 3)C 4)C 5)C 6)C 7)D 8)C 9)A

Bu Bölümde Ne Öğrendik Özeti

Bu bölümde iki anakütle olması halinde anakütle parametre farklarına ilişkin iddialer test edildi. Yanı sıra küçük örnekler de gözden geçirildi.

7. GENEL TEKRAR

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

ÇÖZÜMLÜ ÖRNEKLER

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~ N(20,16)

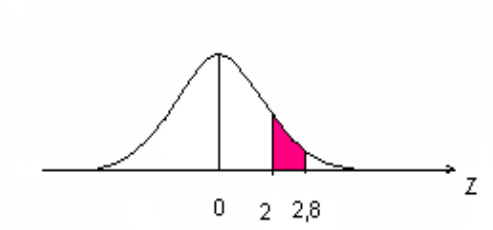
$$P(x < 17.8) = P\left(\frac{x-\mu}{\sigma} < \frac{17.8-20}{4}\right) = P(z < -0.55)$$

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

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 . % 2'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 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 an ez 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 kütlenin bütün birimlerinin incelenmesi olarak isimlendirilir. Anakütleden çeşitli büyüklüklerde seçilen ve altkümeler örneklem, örneklemlerin seçilme sürecine ise örnekleme adı verilir.

6- Örnekleme yapmayı gerektiren nedenleri sıralayınız.

Maliyet, tamsayım bütçesi örnekleme bütçesinden daha fazladır

Zaman, örnekleme tamsayıma göre daha kısa zamanda sonuç verir.

7- Örnekleme dağılımı nedir?

Bir anakütleden belirli büyüklüklerde seçilen örneklemler 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 eğerlerine 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 5m^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.

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

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

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

$$-0.04 \le P_1 - P_2 \le 0.12$$

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.

$$H_0 : \mu = 22$$
$$H_1 : \mu \neq 22$$
$$s = \frac{6}{\sqrt{144}} = 0.5$$

$$z = \frac{22 - 25}{0.5} = 8$$

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 başkanı desteklediğine göre iddiay %5 anlamlılık düzeyinde test ediniz.

$$H_0 : P_1 = P_2$$
$$H_0 : P_1 \neq P_2$$

$$P_1 = 15 / 50 = 0.75$$
$$P_2 = 24 / 60 = 0.40$$
$$\sigma = \sqrt{(0.3 * 0.7) / 50 + (0.4 * 0.6) / 60} = 0.091$$

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

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 amacla A kursuna devam eden 100 kişinin ortalama skoru 87.5 ve sapması 22.5 olarak hesaplanmıştır. B kursuna devam eden125 öğrenciye notları sorulmuş ve ortalama notları 97.5 ve standart sapması da 27.5 olarak hesaplanmıştır. %1 anlamlılık düzeyinde iddiayı sınayınız.

$$H_0 : \overline{X}_1 = \overline{X}_2$$
$$H_0 : \overline{X}_1 \neq \overline{X}_2$$

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

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

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.

$$H_0 : P_1 = P_2$$
$$H_1 : P_1 \neq P_2$$
$$p_1 = \frac{60}{100} = 0.6$$
$$p_2 = \frac{49}{140} = 0.35$$

$$s = \sqrt{(0.6x0.4) / 100 + (0.35x0.65) / 100} = 0.063$$

$$z = \frac{0.60 - 0.35}{0.063}$$
$$z_{tablo} = 2.33$$
$$H_0 \quad reddedilir$$

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.

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

$$s = \sqrt{\frac{(0.20x0.80)}{200}} = 0.028$$

$$0.2 \pm 2.58(0.028)$$

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?

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

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

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

$$\%53$$

18- İki farklı atölye üretilen aynı marka ürünlerin ortalama ağırlıkları arasında fark olup olmadığı değerlendirilecektir. Bu amaçla I. Atölyeden seçilen 15 birimin ortalaması 242gr ve varyansı 10 olarak hesaplanmıştır. II. Atölyeden seçilen 12 birimin ortalaması 239 gr ve varyansı 20 olarak hesaplanmıştır. Buna göre %5 önem düzeyinde ortalamalar arasında fark var mıdır?
Bu soru güven aralıkları yardımıyla çözülebileceği gibi hipotez testleri ile çözülebilir.

$$H_A : \mu \neq 368$$

$$t = \frac{242 - 239}{\sqrt{\frac{10}{15} + \frac{20}{10}}} = 0.19$$

$$(242 - 239) - 1.75\sqrt{\frac{10}{15} + \frac{20}{10}} \leq \mu_1 - \mu_2 \leq (242 - 239) + 1.75\sqrt{\frac{10}{15} + \frac{20}{10}}$$

8. REGRESYON ANALİZİ I

Bu Bölümde Neler Öğreneceğiz?

Bu bölümde iki değişken arasındaki ilişkiyi modellemeyi öğreneceğiz. Regresyon analizi iki değişken arasındaki nedensel ilişkinin fonksiyonel yapısını verir.

Bölüm Hakkında İlgi Oluşturan Sorular

1) Bir topluluktaki kişilerin gelir ve tüketimleri üzerine araştırılma yapılmak istenmiştir. Aynı gelire sahip kişilerin tüketimlerinin farklı olabildiği görülmüştür. Topluluktaki tüm bireyler gözetilerek gelir ve tüketim arasındaki ilişki nasıl modellenebilir?

2) Suç oranı ve işsizlik arasındaki beklenen ilişki nasıl modellenebilir?

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	**Kazanım**	**Kazanımın nasıl elde edileceği veya geliştirileceği**
Regresyon analizi, sabit ve eğim parametrelerinin belirlenmesi	İki değişken arasındaki ilişki, değişkenlerinin bağımlı ve bağımsız yapıda olmaları dikkate alınarak modellenir.	Metinler, çözümlü problemler.

Anahtar Kavramlar

- Regresyon Analizi
- Bağımlı değişken
- Bağımsız değişken
- Sabit ve eğim katsayısı

Giriş

Bu bölümde regresyon analizini ele alarak iki değişken arasındaki ilişkiyi modelleyeceğiz. Değişkenlerden hangisinin etkileyen, hangisinin etkilenen değişken olduğunu belirleyerek bir modelleme yapacağız.

8.1. Regresyon Analizi

Regresyon analizi değişkenler arasındaki ilişkinin araştırılmasında kullanılan istatistiksel bir araçtır. Bu araçla, bir değişkenin diğer değişken üzerindeki nedensel ilişkisi araştırılır. İncelenen ilişkideki değişkenler aralarındaki ilişki göz önüne alınarak değişkenler bağımlı ve bağımsız olarak isimlendirilir. Önceki derslerde de değinildiği gibi istatistiğin öncelikli ilgi alanını rastlantı değişkeninin davranışını bir modelle tahmin etmek oluşturur. Davranışı tahmin edilecek olan rastlantı değişkeni bir diğer değişken(ler)in fonksiyonu olarak gösterilebilir ve bu değişken bağımlı olarak isimlendirilir ve Y ile gösterilir. Bağımlı değişkeni etkileyen değişken ise X ile gösterilir ve bağımsız değişken olarak isimlendirilir. Yani, bağımlı değişken, bağımsız değişken(ler) tarafından açıklanmaya çalışılır ve açıklayıcı değişkenlerin modelde bilinen sabitler olduğu varsayılır.

Model şu şekilde gösterilir; $Y_i = b_0 + b_1 X_i + \varepsilon_i$

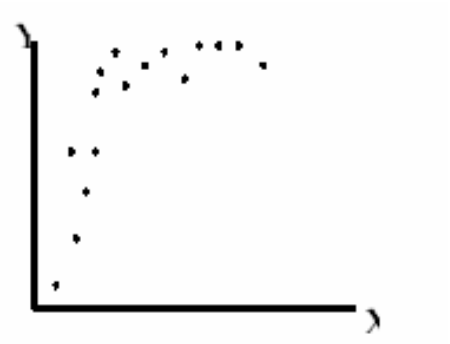
burada b_0 sabit katsayıyı gösterir buna başlangıç parametresi de denir, b_1 ise eğim parametresidir. X'deki 1 birimlik değişmenin Y üzerinde nasıl bir değişim yaptığını gösterir. Denklemdeki ε_i ise daha sonra da açıklanacağı gibi hata terimine karşılık gelir.

Örneğin, tüketim ve gelir üzerine yapılan bir çalışmada bağımsız değişken gelir, bağımlı değişkense tüketimdir ya da bir hastaya uygulanan ilacın dozu ve hastanın iyileşme süreci çalışmasında bağımsız değişken ilacın dozu ve bağımlı değişkense hastanın iyileşme süreci olur.

Regresyon analizi, bilinen gerçekleşen olaylar sonucunda elde edilen bulgulardan yola çıkarak gelecekteki olaylarla ilgili tahmin yani öngörü yapılmasını sağlar. Regresyon modelinde amaç, koşullar değiştiğinde bağımlı değişkenin ortalamasının $E(Y_i)$ 'nin nasıl değiştiğini tanımlamaktır.

Değişkenler arasındaki ilişki deterministik (kesin) ya da olasılıksal (istatistiksel) olarak isimlendirilir. Deterministik ilişkileri açıklamada kullanılan matematiksel fonksiyondan farklı olarak, regresyon analizi olasılıksal ilişkileri açıklar. Arz-talep, gelir-tüketim gibi ilişkilerin modellenmesinde deterministik ilişkiye sahip değişkenler yerine istatistiksel ilişkiye sahip değişkenler kullanılır.

Söz konusu ilişkide bağımsız değişken sayısının bir tane olması basit regresyonla, birden fazla olması ise çoklu regresyonla açıklanır. Regresyon modeli, doğrusal yapıda olabileceği gibi parabolik, logaritmik, üstel biçimli de olabilir. Modelde bir bağımlı ve bir bağımsız değişken söz konusu olduğunda, yani basit regresyon söz konusu iken serpilme diyagramı kullanılarak uygun model seçimi yapılabilir. Serpilme diyagramı, i. gözlemin bağımlı değeri y_i ve bağımsız değeri x_i olmak üzere tüm gözlem çiftleri üzerinden, her ikili yani Y ve X değişkenlerinin aldığı tüm değerler birer nokta ile temsil edilecek şekilde çizilir. Diyagramdaki dağılıma bakılarak uygun model belirlenir.



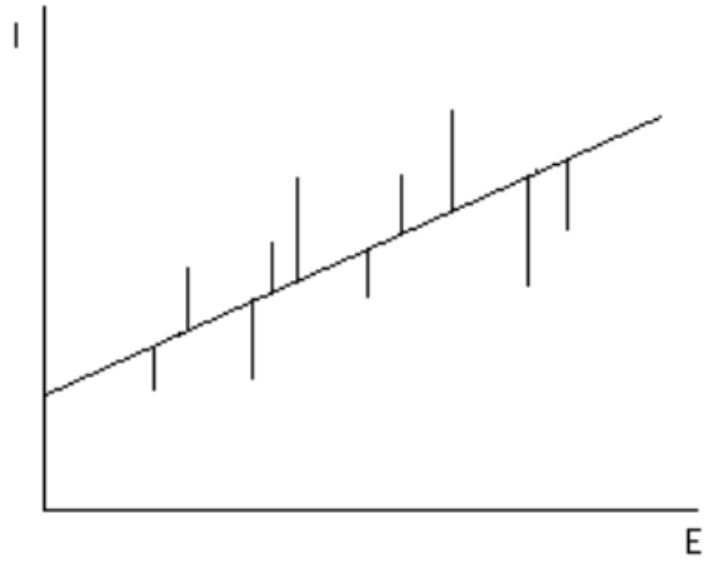
Yukarıdaki serpilme diyagramlarında noktaların ortasından geçecek olan eğri dikkate alınır ve bu eğri incelenen ilişki biçimi hakkında bilgi verir. Buna göre, ilk çizimde noktaların bir doğru etrafında toplandığı söylenebilir ve değişkenler arasında aynı yönlü doğrusal bir ilişkinin varlığından söz edilebilir. İkinci çizimse ters yönlü doğrusal bir ilişkinin varlığını gösterir. Üçüncü çizimde doğrusal olmayan bir ilişkinin varlığı söz konusudur. Son çizim

dikkate alındığında ise bir eğri oluşturmak mümkün görünmemektedir, değişkenler arasında bir ilişkinin olmadığı sonucuna varılır.

Serpilme diyagramı çizimi sonrasında uygun modele karar verilir ve modeldeki parametreler tahmin edilir. Tahmin sürecinde çeşitli kriterler doğrultusunda kullanılan yöntemler olmakla beraber, burada bu yöntemlerden sadece En Küçük Kareler Yöntemine (EKK) değinilecektir.

8.2. Regresyon Katsayılarının Tahmini

EKK ile bulunacak eğrinin her (xi,yi) gözlem çiftine karşılık gelen nokta ile bu noktanın EKK ile elde edilecek eğri üzerindeki dik izdüşümü arasındaki farklar toplamı sıfır olmalıdır. Bu farklar, yani Y_i değerlerinin regresyon doğrusuna olan uzaklığı, daha sonrada bahsedileceği gibi 'hata' olarak isimlendirilir. Y_i değerlerinin regresyon doğrusu üzerindeki görüntüsü $\hat{Y}_i$ (tahmini Y_i) ile arasındaki fark hataya karşılık gelir. İdeal regresyon doğrusu, bu farkların karelerinin toplamını $\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$, minimum verenle elde edilir.



Hata kareler toplamının minimum olabilmesi için, sabit ve eğim parametrelerine göre türevleri alınarak sıfıra eşitlenir:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2$$

$$\frac{\partial(\sum e_i^2)}{\partial b_0} = 2\sum (Y - b_0 - b_1 X)(-1) = 0$$

$$\frac{\partial(\sum e_i^2)}{\partial b_1} = 2\sum (Y - b_0 - b_1 X)(-x) = 0$$

Bu işlem sonrasında elde edilen denklemler Normal Denklemler olarak isimlendirilir:

$$\sum Y_i = nb_0 + b_1 \sum X_i$$

$$\sum X_i Y_i = b_0 \sum X_i + b_1 \sum X_i^2$$

Cramer yöntemine göre bu denklemler çözüldüğünde;

$$b_0 = \frac{\begin{vmatrix} \sum Y & \sum X \\ \sum YX & \sum X^2 \end{vmatrix}}{\begin{vmatrix} n & \sum X \\ \sum X & \sum X^2 \end{vmatrix}} = \frac{\sum Y \sum X^2 - \sum X \sum XY}{n\sum X^2 - (\sum X)^2}$$

$$b_1 = \frac{\begin{vmatrix} n & \sum Y \\ \sum X & \sum XY \end{vmatrix}}{\begin{vmatrix} n & \sum X \\ \sum X & \sum X^2 \end{vmatrix}} = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2}$$

elde edilir. Ancak eğim ve sabit parametre tahminlerini ortalamadan sapmalar üzerinden giderekte belirlemek mümkündür:

$$x = X - \bar{X}$$

$$y = Y - \bar{Y}$$

Eğim parametresi bu x ve y üzerinden gidilerek belirlenir.

$$\frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2} = \frac{n\sum (x+\bar{X})(y+\bar{Y}) - [\sum (x+\bar{X})(y+\bar{Y})]}{n\sum (x+\bar{X})^2 - (\sum (x+\bar{X}))}$$

$$= \frac{n\sum (xy + x\bar{Y} + \bar{X}y + \bar{X}\bar{Y}) - (\sum x + n\bar{X})(\sum y + n\bar{Y})}{n\sum (x^2 + 2x\bar{X} + \bar{X}^2) - (\sum x + n\bar{X})^2}$$

$$= \frac{n\sum xy + n\bar{Y}\sum x + n\bar{X}\sum y + n^2\bar{X}\bar{Y} - (\sum x + n\bar{X})(\sum y + n\bar{Y})}{n\sum x^2 + 2n\bar{X}\sum x + n^2\bar{X}^2 - (\sum x + n\bar{X})^2}$$

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

$$\sum y = \sum (Y - \bar{Y}) = 0$$

olduğundan eşitlikler düzenlenirse;

$$b_1 = \frac{n\sum xy + n^2\bar{X}\bar{Y} - n^2\bar{X}\bar{Y}}{n\sum x^2 + n^2\bar{X}^2 - n^2\bar{X}^2} = \frac{\sum xy}{\sum x^2}$$

$$b_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

Sabit parametre için,

$\sum Y_i = nb_0 + b_1 \sum X_i$ olduğundan, eşitliğin her iki tarafı n'e bölünürse,

$$\bar{Y} = b_0 + b_1\bar{X}$$

$$b_0 = \bar{Y} - b_1\bar{X}$$

elde edilir.

Örnek: Aşağıda bir sınıftaki öğrencilerin muhasebe ve matematik derslerine ait veri bulunmaktadır. Muhasebe dersinden başarının matematik dersinden başarıya bağımlı olup olmadığını sınamak için regresyon denklemini oluşturunuz.

Muhasebe	**Matematik**
1	2
2	3
3	5
5	6
6	7
7	10
8	7
8	8

Y	**X**	**Y^2**	**X^2**	**YX**	**X-Xort=x**	**Y-Yort=y**	**xy**	**x^2**
1	2	1	4	2	-4	-4	16	16
2	3	4	9	6	-3	-3	9	9
3	5	9	25	15	-1	-2	2	1
5	6	25	36	30	0	0	0	0
6	7	36	49	42	1	1	1	1
7	10	49	100	70	4	2	8	16
8	7	64	49	56	1	3	3	1
8	8	64	64	64	2	3	6	4
40	**48**	**252**	**336**	**285**	**0**	**0**	**45**	**48**

Yort=5

Xort=6

Normal denklemlerden elde edilen eşitlikler kullanılarak katsayılar şöyle tahmin edilmiştir:

$$b_0 = \frac{\sum Y \sum X^2 - \sum X \sum XY}{n\sum X^2 - (\sum X)^2}$$

$$b_0 = \frac{40(336) - 48(285)}{8(336) - 48^2} = -0.625$$

$$b_1 = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2}$$

$$b_1 = \frac{8(285) - 48(40)}{8(336) - 48^2} = 0.9375$$

Ortalamadan sapmalarla,

$$b_1 = \frac{\sum xy}{\sum x^2} = \frac{45}{48} = 0.9375$$

$$b_0 = \overline{Y} - b_1\overline{X} = 5 - 0.9375(6) = -0.625$$

yine aynı donuca ulaşılmıştır. Tahmin edilen regresyon denklemi şöyledir:

$$\hat{y} = -0.625 + 0.9375$$

Örnek:

Aşağıda bir eyaletteki suç ve işsizlik oranlarına ilişkin veri mevcuttur.

işşizlik oranı	**suç oranı**
0,8	3
1,4	6
2,3	7
3,5	15
4,5	19

İşsizlik ve suç işleme oranları arasındaki ilişkiyi gösteren regresyon denklemini oluşturunuz.

Sorunun çözümünün ilk aşamasında bağımlı ve bağımsız değişkenleri belirleyelim. İşssizlik oranı bağımsız değişkendir (x), suç oranını etkiler ki bu da bağımlı değişken (Y) olarak isimlendirilir. İlgili kolonların toplam ve çarpımlarının toplamına ait bilgiler aşağıdaki gibi özetlenmiştir.

$$\sum Y = 50 \quad \sum X = 12.5 \quad \sum XY = 164.9$$
$$\sum X^2 = 40.39 \quad \sum Y^2 = 680$$
$$\sum (X - \overline{X})(Y - \overline{Y}) = 39.9$$
$$\sum (X - \overline{X})^2 = 9.14 \quad \sum (Y - \overline{Y})^2 = 180$$

$$b1 \equiv \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2} \qquad b_0 = \overline{Y} - b_1 \overline{X}$$

$$b_1 = \frac{39.9}{9.14} = 4.365$$

$$b_0 = \frac{50}{5} - 4.365(\frac{12.5}{5}) = -0.9$$

$$Y = -0.9 + 4.365X$$

İşsizlik oranı 1 birim arttığında suç oranı 4.365 birim artar.

Ortalamadan sapmalar serisi yerine orijinal seriden hareket edilirse eğim katsayısı şöyle bulunacaktır:

$$b_1 = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2} = \frac{5(164.9) - (12.5)(50)}{5(40.39) - 12.5^2} = 4.365$$

Örnek:

Bir firmanın reklam harcamaları ve satış rakamlarına ilişkin veri mevcuttur. Regresyon denklemini oluşturunuz

Reklam Harca.	**Satışlar**
1.6	6
2.8	12
4.6	14
7	30
9	38

Bağımlı değişken satışlardır., Reklam harcamaları ise bağımsız değişkendir.

$$\sum Y = 100 \quad \sum X = 25 \quad \sum XY = 659.6$$
$$\sum X^2 = 161.56 \quad \sum Y^2 = 2720$$
$$\sum (X - \bar{X})(Y - \bar{Y}) = 159.6$$
$$\sum (X - \bar{X})^2 = 36.56 \quad \sum (Y - \bar{Y})^2 = 720$$

$$b_1 = \frac{159.9}{36.56} = 4.37$$
$$b_0 = \frac{100}{5} - 4.37(\frac{25}{5}) = -1.8$$
$$Y = -1.8 + 4.37X$$

Reklam harcamaları 1 birim artarsa satış 4.37 birim artar.

Diğer yaklaşımla:

$$b_1 = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2} = \frac{5*659.6 - 25*100}{5*161.56 - 25^2} = 4.37$$

$$b_0 = \frac{\sum Y \sum X^2 - \sum X \sum XY}{n\sum X^2 - (\sum X)^2} = -1.8$$

Örnek: Aşağıdaki x,y serilerinden hareketle regresyon denklemini bularak yorumlayınız.

x	y
10.2	7
8.4	5
6.2	4
4.2	1
11	8

$$n = 5, \quad \sum X = 85, \quad \sum Y = 45$$
$$\sum XY = 789, \quad \sum X^2 = 1475, \quad \sum Y^2 = 425$$

$$b_1 = \frac{789 - 5(85/5)(45/5)}{1475 - 5(85/5)^2} = 0.8$$

$$b_0 = 45/5 - (0.8)(85/5) = -4.6$$

Örnek: Aşağıda verilen x,y ikilisinden hareketle regresyon denklemini bulunuz.

x	y
20	12
19	10
17	9
16	8
13	6

$$n = 5, \quad \sum X = 40, \quad \sum Y = 25$$

$$\sum XY = 230.4, \quad \sum X^2 = 351.68, \quad \sum Y^2 = 155$$

$$b_1 = \frac{230.4 - 5(40/5)(25/5)}{351.68 - 5(40/5)^2} = 0.96$$

$$b_0 = 25/5 - (0.96)(40/5) = -2.677$$

$$y = -2.677 + 0.96x$$

Bölüm Soruları

1) X={2,3,4,5}, Y={6,8,8,10} X ve Y sırasıyla bağımsız ve bağımlı değişkenler ise regresyon denklemi nasıl ifade edilir?,

2) Uygulanan ilacın dozu ve ilacın etki süresi aşağıdaki gibidir. Regresyon denklemini bularak yorumlayınız.

Doz	**İyileşme süresi**
1	8
2	8
3	6
4	4
5	2

BÖLÜM 8 ÇOKTAN SEÇMELİ SORULARI

1- Aşağıdaki grafiklerden hangisi kullanılarak değişkenler arasındaki ilişki hakkında fikir sahibi olunabilir?

a) Pasta Grafiği
b) Frekans Poligonu
c) Serpilme diyagramı
d) Histogram
e) Dal-yaprak grafiği

2-Aşağıdaki ifadelerden hangisi ya da hangileri doğrudur?

I)EKK Yöntemi, artıkların karelerinin toplamını en küçük yapabilen regresyon denklemindeki sabit ve eğim parametrelerinin elde edilmesini sağlayan bir yöntemdir.
II)Korelasyon analizi ile değişkenler arasındaki ilişkinin yönü, büyüklüğü ve matematiksel modeli belirlenir.
III) Regresyon analizinde bağımlı değişken ve gözlem değerlerinden elde edilen tahmini Y değerleri arasındaki farka hata terimi adı verilir.
a)I ve III b) Yalnız I c) Yalnız II d) Yalnız III e) I ve II

3- Aşağıdaki ifadelerden hangisi ya da hangileri doğrudur?

I) Regresyon analizinde değişkenlerden biri artarken diğeri azalıyorsa eğim katsayısı negatif değerlidir.
II) Basit regresyon analizinde serpilme diyagramına bakılarak ilişkinin yönünü belirlemek mümkündür.
III) Regresyon analizinde hata terimlerinin toplamı sıfırdır.
a) Yalnız I b) Yalnız II c)Yalnız III d)I ve II e) I, II ve III

4-X bağımsız ve Y ise bağımlı değişkeni göstermek üzere, $\sum_{i=1}^{8} Y_i = 360$ $\sum_{i=1}^{8} X_i = 184$ $\sum_{i=1}^{8} X_i^2 = 4834$ $\sum_{i=1}^{8} X_i Y_i = 8566$ regresyon denkleminin eğim parametresi aşağıdakilerden hangisine eşit olur?

a)1.58 b) 1.64 c) 2.11 d)-1.58 e)-1.64

5-Yukarıdaki bilgiden hareketle regresyon denkleminin sabit parametresi aşağıdakilerden hangisine eşit olur?

a)8.66 b) 7.66 c) -12.15 d) -8.66 e)-2.23

6-Y=8.66+1.58X regresyon denklemi için aşağıdakilerden hangisi söylenemez?

a)X artarken Y artar

b)Y artarken X artar
c)X ve Y arasında aynı yönlü ilişki vardır
d)X de 1 birimlik artıs Y'de ortalama 1.58 birimlik artmaya neden olur.
e) X ile Y arasında doğrusal artan bir ilişki vardır

7-Regresyon modelinde gözlenen değerler ve tahmin arasındaki farkın kareli ortalamasına ne denir?

a) Değişim katsayısı b)Standart sapma c)Korelasyon katsayısı d)Eğim katsayısı e)Tahminin standart hatası

8- X bağımsız ve Y ise bağımlı değişkeni göstermek üzere,

$\sum_{i=1}^{7} Y_i = 140$ $\sum_{i=1}^{7} X_i = 280$ $\sum_{i=1}^{7} X_i^2 = 6574$ $\sum_{i=1}^{7} X_i Y_i = 5870$ regresyon denkleminin eğim parametresi aşağıdakilerden hangisine eşit olur?

a) -2.3 b) -1.5 c)0.8 d)1.5 e)2.3

9) Yukarıdaki bilgiden hareketle regresyon denkleminin sabit parametresi aşağıdakilerden hangisine eşit olur?

a) 60 b) 40 c) 20 d) -20 e)-40

10-Y=-40+1.5X regresyon denklemi için aşağıdakilerden hangisi söylenemez?

a)X artarken Y artar
b)Y artarken X azalır
c)X ve Y arasında aynı yönlü ilişki vardır
d)X de 1 birimlik artıs Y'de ortalama 1.5 birimlik artmaya neden olur.
e) X ile Y arasında doğrusal artan bir ilişki vardır

Yanıtlar

1)C 2)A 3)E 4)A 5)A 6)B 7)E 8)D 9)E 10) B

Bu Bölümde Ne Öğrendik Özeti

Çeşitli olayları, olguları açıklamakta kullandığımız değişkenler ve bunlar arasındaki ilişkinin anlaşılmasında kullanılan regresyon analizini, basit doğrusal regreyon denkleminde yer alan sabit ve eğim parametrelerinin elde edilmesini, yorumlanmasını öğrendik.

9. REGRESYON ANALİZİ II

Bu Bölümde Neler Öğreneceğiz?

Bu bölümde tahmin edilen anakütle regresyon parametrelerinin istatistiksel olarak anlamlılığını sınayacağız.

Bölüm Hakkında İlgi Oluşturan Sorular

1) Tahmin edilmiş olan gelir tüketim denkleminde, gelir gerçekten istatistiksel olarak anlamlı mıdır?

2) 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.

9.1. Regresyon Doğrusunun Özellikleri ve Hata (Artık) Kavramı

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 değişmenin $E(Y_i)$ 'deki değişimine karşılık gelir. Burada β_0 ve β_1 parametrelerdir. Bu parametlere 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. Yi 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:

$$\hat{Y}_i = b_0 + b_1 X_i$$

$$\hat{Y}_i = \overline{Y} + b_1 (X_i - \overline{X})$$

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,

$\varepsilon_i = Y_i - E(Y_i)$ şeklindedir.

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 toplamları sıfırdır.

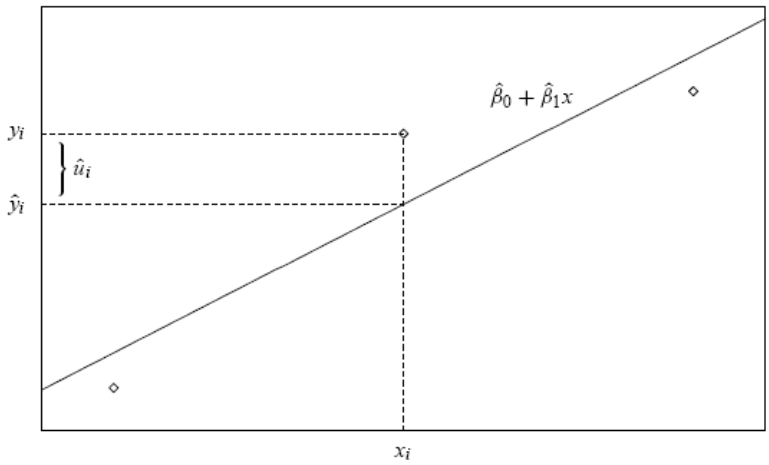
$$\sum_{i=1}^{n} X_i e_i = 0$$

-Tahmini Y değerleri ile artıkların çarpımlar toplamı sıfırdır.

$$\sum_{i=1}^{n} \hat{Y}_i e_i = 0$$

- Regresyon doğrusu daima $(\overline{X}, \overline{Y})$ noktasından geçer.

Gözlenen Y_i, tahmini 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.

$$S = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n-2}}$$ 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.

9.2. Regresyon Denklemi Parametrelenin Hipotez Testi ve Güven Aralığı

EKK yöntemi ile $Y = b_0 + b_1 X$ 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 değişkenin 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:

$$t = \frac{\hat{b}_0 - b_0}{s_{\hat{b}_0}} \qquad\qquad t = \frac{\hat{b}_1 - b_1}{s_{\hat{b}_1}}$$

Bu istatistikler n-2 serbestlik dereceli t dağılır. Bu eşitliklerden yola çıkarak sırasıyla sabit ve eğim paramereleri için güven aralıkları şöyledir:

$$P[-t_{n-2,\alpha/2} < \frac{\hat{b}_0 - b_0}{S_{\hat{b}_0}} < t_{n-2,\alpha/2}] = 1-\alpha$$

$$P[\hat{b}_0 - t_{n-2,\alpha/2}S_{\hat{b}_0} \le b_0 \le \hat{b}_0 + t_{n-2,\alpha/2}S_{\hat{b}_0}] = 1-\alpha$$

$$\hat{b}_0 - t_{n-2,\alpha/2}S_{\hat{b}_0} \le b_0 \le \hat{b}_0 + t_{n-2,\alpha/2}S_{\hat{b}_0}$$

Sabit parametre için yukarıdaki eşitsizliklerin düzenlenmesiyle bulunan yaklaşım eğim parametresi için de yapıldığında bulunan güven aralığı şöyledir:

$$\hat{b}_1 - t_{n-2,\alpha/2}S_{\hat{b}_1} \le b_1 \le \hat{b}_1 + t_{n-2,\alpha/2}S_{\hat{b}_1}$$

İlgilenilen hipotez testi sabit katsayı için aşağıdaki gibi düzenlendiğinde, aşağıdaki test istatistiği kullanılır:

$$H_0 : b_0 = b$$
$$H_1 : b_0 \neq b$$

$$t_0 = \frac{\hat{b}_0 - b}{S_{\hat{b}_0}}$$

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.

$t_0 > t_{n-2,\alpha/2}$ veya $t_0 < -t_{n-2,\alpha/2}$

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

$H_1 : b_0 > b$ 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 değerin kıyaslanması sabit parametre için hipotez testinde değinildiği gibidir, aynıdır.

Örnek: Önceki örnekte bir sınıftaki öğrencilerin matematik ve muhasebe derslerindeki başarıları arasındaki ilişki incelenmişti. Regresyon denkleminin paremetrelerinin %99 ve %95 güvenle bulundukları aralıkları belirleyiniz ve katsayların anlamlılığını aynı güven düzeyinde sınayınız.

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**

$$S=\sqrt{\frac{9.81}{8-2}}=1.278$$

$$0.486 \le b_1 \le 1.389$$

$S_{\hat{b}_1}=\dfrac{S}{\sqrt{\sum(X-\bar{X})^2}}=\dfrac{1.278}{\sqrt{48}}=0.1846$ olacağından %99 ve %95 güvenle parametrenin yer alacağı aralıklar şöyledir:

$$0.9375-(0.1846)(3.707) \le b_1 \le 0.9375+(0.1846)(3.707)$$

$$0.2532 \le b_1 \le 1.622$$

($t_{0.01;6}$ = 3.707)

$$0.9375-(0.1846)(2.447) \le b_1 \le 0.9375+(0.1846)(2.447)$$

$$0.486 \le b_1 \le 1.389$$

($t_{0.05;6}$ = 2.447)

Sabit parametrenin standart hatası, güven aralığı aşağıdaki gibidir:

$$S_{\hat{b}_0} = \sqrt{\frac{\sum X^2}{n\sum (X-\bar{X})^2}}S = \sqrt{\frac{336}{8(48)}}1.278 = 1.195$$

$$-0.625-(1.195)(3.707) \le b_0 \le -0.625+(1.195)(3.707)$$

$$-5.055 \le b_0 \le 3.805$$

$$-0.625-(1.195)(2.447) \le b_0 \le -0.625+(1.195)(2.447)$$

$$-3.549 \le b_0 \le 2.299$$

Eğim parametresinin anlamlılığı için kurulan hipotezler ve testi şöyledir:

$H_0:\ b_1 = 0$
$H_1:\ b_1 \ne 0$

$$t_0 = \frac{0.9375}{0.1846} = 5.079$$

elde edilen değer her iki anlamlılık düzeyindeki kritik değerlerle sınandığında sıfır hipotezi reddedilir. Yani, Y bağımlı değişkeni X'e bağımlıdır.

Sabit katsayının testi benzeri yaklaşımla incelendiğinde,

$H_0:\ b_0 = 0$
$H_1:\ b_0 \ne 0$

$$t_0 = \frac{-0.625}{1.195} = -0.523$$

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:

$$S_{Y_i} = \sqrt{\frac{1}{n} + \frac{(x_i - \overline{x})^2}{\sum (x - \overline{x})^2}} S$$

- Eğer belirli bir x değerine karşılık gelen tek bir nokta için güven aralığı oluşturmak istenirse kullanılacak standart hata şöyledir:

$$S_{Y_d} = \sqrt{1 + \frac{1}{n} + \frac{(x_d - \overline{x})^2}{\sum (x - \overline{x})^2}} S$$

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:

$$7.8125 - 2.447(0.714) \le Y_9 \le 7.8125 + 2.447(0.714)$$

$$6.065 \le Y_9 \le 9.56$$

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:

$$7.8125 - 2.447(1.465) \le Y_9 \le 7.8125 + 2.447(1.465)$$

$$4.23 \le Y_9 \le 11.40$$

Örnek:

Önceki bölümde ele alınan suç ve işsizlik oranlarına ilişkin örnekteki parametrelerin anlamlılığını % 5 önem düzeyinde sınamak istersek;

X Y

işşizlik 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

$$\sum Y = 50 \quad \sum X = 12.5 \quad \sum XY = 164.9$$
$$\sum X^2 = 40.39 \quad \sum Y^2 = 680$$
$$\sum (X - \bar{X})(Y - \bar{Y}) = 39.9$$
$$\sum (X - \bar{X})^2 = 9.14 \quad \sum (Y - \bar{Y})^2 = 180$$

$$Y = -0.9 + 4.365X$$

$$S = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n-2}} \sqrt{\frac{5.82}{5-3}} = 1.39$$

$$S(\hat{b}_1) = \frac{S}{\sqrt{\sum (X - \bar{X})^2}} = \frac{1.39}{\sqrt{9.14}} = 0.455$$

$$S_{\hat{b}_0} = \sqrt{\frac{\sum X^2}{n \sum (X - \bar{X})^2}} S = \sqrt{\frac{40.39}{5(9.14)}} 1.39 = 1.306$$

$$4.365 - 3.82 * 0.455 \le b_1 \le 4.365 + 3.82 * 0.455$$

$$2.63 \le b_1 \le 6.1$$

$$t = \frac{4.365}{0.455} = 9.6$$

Eğim parametresi anlamlıdır, güven aralığı sıfırı kapsamadığı gibi hesaplanan t istatistik değeri de tablo değerinden büyüktür (9.6> 3.82)

$$-0.9 - 3.82 * 1.306 \le b_0 \le -0.9 + 3.82 * 1.306$$

$$-5.88 \le b_0 \le 4$$

$$t = \frac{-0.9}{1.306} = -0.69$$

Sabit parametresi ise anlamsızdır. Aralık sıfırı kapsar ve test istatistiği sonucu kritik değerden küçük çıkmıştır.

Örnek: Aşağıdaki veriden hareketle regresyon denklemi eğim katsayısının anlamlılığını değerlendiriniz.

x	y
20	6
19	8
17	9
16	10
13	12

$$n = 5, \quad \sum X = 85, \quad \sum Y = 45$$

$$\sum XY = 741, \quad \sum X^2 = 1475, \quad \sum Y^2 = 425$$

$$b_1 = \frac{741 - 5(85/5)(45/5)}{1475 - 5(85/5)^2} = -0.8$$

$$b_0 = 45/5 - (-0.8)(85/5) = 22.6$$

x	y	ytah	e	e^2
20	6	6.6	-0.6	0.36
19	8	7.4	0.6	0.36
17	9	9	0	0
16	10	9.8	0.2	0.04
13	12	12.2	-0.2	0.04
85	45	45	-3.6E-15	0.8

$$s = \sqrt{\frac{0.8}{5-2}} = 0.516$$

$$s_{b1} = \frac{0.516}{\sqrt{30}} = 0.094$$

$$t = \frac{-0.8}{0.094} = -8.9$$

3 sd li t tablo değeri -3.82 olduğundan eğim parametresinin anlamsız olduğunu ifade eden sıfır hipotezi reddedilir.

Bölüm Soruları

1-Aşağıda matematik dersine çalışma süresi ve alınan notlara ilişkin bilgi mevcuttur. Regresyon denklemini kurarak eğim ve sabit parametrelerin anlamlılığını sınayınız.

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 sınayınız.

Doz	**İyileşme süresi**
1	8
2	8
3	6
4	4
5	2

BÖLÜM 9 ÇOKTAN SEÇMELİ SORULARI

A) Aşağıdaki 1-2-3 ve 4 numaralı soruları şu soru metninden hareketle çözünüz:
X bağımsız ve Y bağımlı değişkene karşılık gelmek üzere,

$$\sum_{i=1}^{9} Y_i = 70 \quad \sum_{i=1}^{9} X_i = 44 \quad \sum_{i=1}^{9} X_i^2 = 298 \quad \sum_{i=1}^{9} X_i Y_i = 462$$

1) Regresyon denkleminin eğim katsayısı aşağıdakilerden hangisine eşittir?
a) 1.45 b)0.69 c) -0.80 d)-1.29 e)-1.45

2) Regresyon denkleminde tahminin standart hatası nedir?
a) 0.896 b)0.799 c)0.639 d)0.489 e)0.398

3) Kitle için eğim katsayısının yer alacağı sınırlar % 95 olasılıkla aşağıdakilerden hangisine eşittir?
a)-0.72- 0.80 b)-0.52-1.90 c)0.40-1.96 d)0.58-2.16 e)0.70-3.28

4) Bir önceki soruda elde edilen güven aralığından hareketle aşağıdakilerden hangisi söylenemez?
I) X değişkeninin Y değişkeni üzerinde etkisi yoktur.
II) Eğim katsayısı %5 önem düzeyinde anlamlı değildir.
III) Y değişkenin X üzerinde etkisi vardır
a) Yalnız I b) Yalnız II c) I ve II d) Yalnız III e) II ve III

B) 5-6-7-8-9-10 numaralı soruları aşağıdaki metinden hareketle yanıtlayınız. X bağımsız ve Y bağımlı değişkene karşılık gelmek üzere;

$$\sum_{i=1}^{10} Y_i = 160 \quad \sum_{i=1}^{10} X_i = 50 \quad \sum_{i=1}^{10} X_i^2 = 304 \quad \sum_{i=1}^{10} X_i Y_i = 908$$

5- Sabit ve eğim katsayıları sırasıyla aşağıdakilerden hangisine eşit olur?

a)(2, 6) b) (1, 3) c) (6, 2) d) (4,5) e)(2,5)

6- Regresyon denkleminde hata karelerinin toplamı aşağıdakilerden hangisine eşittir?

a) 22 b) 19 c) 17 d)12 e)11

7-Regresyon denkleminde eğim katsayısının standart hatası aşağıdakilerden hangisine eşittir?

a)0.91 b)0.78 c) 0.67 d)0.17 e)0.14

8- Regresyon denkleminde sabit katsayının standart hatası aşağıdakilerden hangisine eşit olur?

a)0.91 b)0.78 c) 0.67 d)0.57 e)0.4

9- Eğim katsayısının anlamlılığını sınamak için yapılan t testi skoru aşağıdakilerden hangisine eşittir?

a) 16.23 b)14.78 c)15.47 d)14.06 e) 11.98

10- Eğim katsayısının anlamlılığı için kullanılan t testi skoru aşağıdakilerden hangisine eşittir (alfa=0.05)?

a) 2.899 b) 2.306 c) 1.95 d) 1.65 e)1.28

Yanıtlar

1)A 2)B 3)B 4)D 5)C 6)D 7)D 8)A 9)E 10)B

Bu Bölümde Ne Öğrendik Özeti

Bu bölümde regresyon parametrelerinin anlamlı olup olmadığını sınadık.

10. KORELASYON KATSAYISI-DETERMİNASYON KATSAYISI

Bu Bölümde Neler Öğreneceğiz?

Bu bölümde regresyon denkleminde bağımlı değişkenin ne oranda bağımsız değişken tarafından açıklandığını determinasyon (Belirlilik) katsayısı ile değerlendireceğiz. Ayrıca iki değişken arasındaki ilişkinin yönünü ve büyüklüğünü korelasyon katsayısı ile inceleyeceğiz.

Bölüm Hakkında İlgi Oluşturan Sorular

1) Enflasyon ve faiz oranı arasında nasıl bir ilişki var? Aynı yönlü mü? Ters yönlü mü? Güçlü mü? Zayıf mı?

2) Gelir ve tüketim arasındaki regresyon denkleminde, tüketimin gelir tarafından açıklanma yüzdesi nedir?

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	**Kazanım**	**Kazanımın nasıl elde edileceği veya geliştirileceği**
Determinasyon katsayısı, Korelasyon katsayısı	İki değişken arasındaki ilişkinin yönünün ve büyüklüğünün belirlenmesi, regresyon denkleminin gücünün belirlenmesi.	Metinler, çözümlü problemler

Anahtar Kavramlar

- Determinasyon katsayısı
- Korelasyon katsayısı
- Regresyon denkleminin gücü

Giriş

Bu bölümde regresyon denkleminde yer alan bağımlı değişkenin bağımsız değişken tarafından açıklanma yüzdesini determinasyon katsayısı ile değerlendireceğiz. Determinasyon katsayısı, regresyon denkleminin gücünü veren bir büyüklüktür. Konularımıza dahil ettiğimiz korelasyon katsayısı ise değişkenler arasındaki ilişkinin nedenselliği ile değil de ilişkinin sadece yönü ve büyüklüğü hakkında bilgi veren bir istatistiktir.

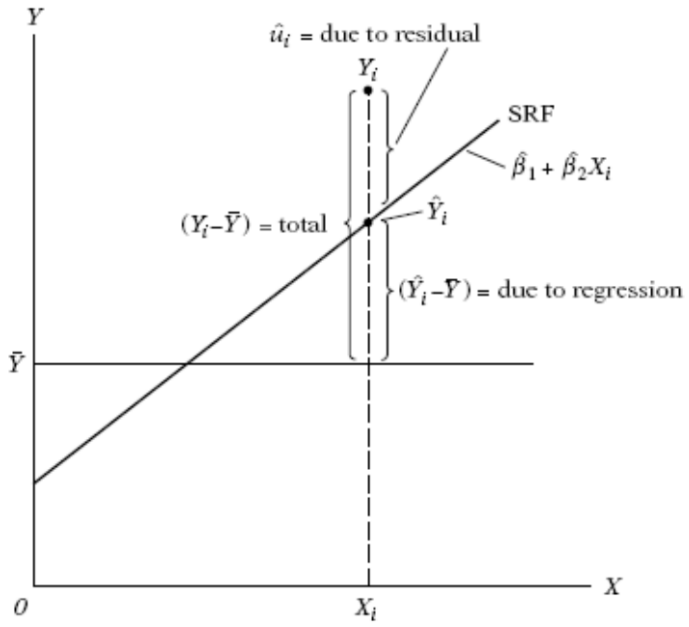
10.1. Regresyon Denkleminin Gücü

Regresyon analiziyle bağımsız değişken(ler)le bağımlı değişkendeki değişme açıklanır. Değişimi açıklamanın tam olması beklenemez, örnekleme ve regresyon denklemin yapısına bağlı olarak farklı oranlarda açıklanır ancak istenen yüksek bir açıklanma oranıdır. Bu orana regresyon denkleminin gücü- determinasyon katsayısı- belirlilik katsayısı denilir ve R^2 ile gösterilir.

Determinasyon katsayısı bağımlı değişkendeki değişimin ne kadarının bağımsız değişken(ler) tarafından açıklandığını gösterir ve $0 \leq R^2 \leq 1$ arasındadır. Katsayının 1'e yaklaşması açıklama gücünün yüksekliğini gösterir.

Aşağıdaki çizimden de görülebileceği gibi bir gözlem için şu eşitlik yazılabilir:

$$Y_i - \overline{Y} = (\hat{Y}_i - \overline{Y}) + (Y_i - \hat{Y}_i)$$



bu eşitlikte şöyle bir düzenleme yapılır.

$$Y_i - \hat{Y}_i = (Y_i - \hat{Y}_i) - (\hat{Y}_i - \overline{Y})$$

ve her iki tarafın karesi alınırsa

$$\sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - \overline{Y})^2 + \sum (\hat{Y}_i - \overline{Y})^2 - 2\sum (Y_i - \overline{Y})(\hat{Y}_i - \overline{Y})$$

elde edilir. Bu eşitlikte bir takım değişiklikler yapmak için normal denklemler hatırlansın.

$$\sum Y = n\hat{b}_0 + \hat{b}_1 \sum X$$

denkleminde her iki taraf n'e bölünürek yukarıdaki karesel eşitliğin sağ tarafındaki ifade yeniden yazılır:

$$\bar{Y} = \hat{b}_0 + \hat{b}_1 \bar{X}$$

$$\hat{Y} = \hat{b}_0 + \hat{b}_1 X$$

$$\hat{Y} = \bar{Y} + \hat{b}_1 (X - \bar{X})$$

$$-2\sum (Y_i - \bar{Y})(\hat{Y}_i - \bar{Y}) = -2\sum (Y_i - \bar{Y})\hat{b}_1 (X_i - \bar{X})$$

$$= -2\hat{b}_1 \sum (Y_i - \bar{Y})(X_i - \bar{X})$$

$$= -2\hat{b}_1^2 \sum (X_i - \bar{X})^2$$

$$= -2\sum (\hat{Y}_i - \bar{Y})^2$$

Karesel eşitlik yeniden düzenlenerek şu hale getirilir:

$$\sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - \bar{Y})^2 - \sum (\hat{Y}_i - \bar{Y})^2$$

$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum (Y_i - \hat{Y}_i)^2$$

$$\sum (Y_i - \bar{Y})^2 = \hat{b}_1^2 \sum (X_i - \bar{X})^2 + \sum (Y_i - \hat{Y}_i)^2$$

Burada, sol taraf regresyon denklemindeki toplam değişmeye karşılık gelir, eşitliğin sağ tarafının ilk kısmı yani regresyon parametresinin yer aldığı kısım açıklanabilen değişimdir. Kalan diğer kısımsa açıklanamayan değişim olarak isimlendirilir.

Toplam Değişme= Açıklanabilen Değişme + Açıklanamayan Değişme

Kısaca şu notasyonla da gösterilebilir:

SST=SSR+SSE

İşte determinasyon katsayısı burada, toplam değişim içerisindeki açıklanabilen değişim yüzdesidir:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

Örnek: Önceki örnekten yola çıkarak regresyon denkleminin gücünü belirleyiniz.

Y	$\hat{Y}$	$(Y-\hat{Y})^2$	$(Y-\overline{Y})^2$
1	1,25	0,0625	16
2	2,188	0,035344	9
3	4,063	1,129969	4
5	5	0	0
6	5,938	0,003844	1
7	8,75	3,0625	4
8	5,938	4,251844	9
8	6,875	1,265625	9
40	**40**	**9,811626**	**52**

$$R^2 = 1 - \frac{9.81}{52} = 0.811$$

Örnek:

Önceki iki bölümde ele alınan işsizlik ve suç oranı istatistikleri verisinden hareketle determinasyon katsayısını bulunuz.

X Y

işşizlik 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	4,5774603
3,5	15	14,3775	0,6225	0,3875063
4,5	19	18,7425	0,2575	0,0663062
12,5	50	50,0625	-0,0625	5,8202578

$$\sum Y = 50 \quad \sum X = 12.5 \quad \sum XY = 164.9$$
$$\sum X^2 = 40.39 \quad \sum Y^2 = 680$$
$$\sum (X - \bar{X})(Y - \bar{Y}) = 39.9$$
$$\sum (X - \bar{X})^2 = 9.14 \quad \sum (Y - \bar{Y})^2 = 180$$

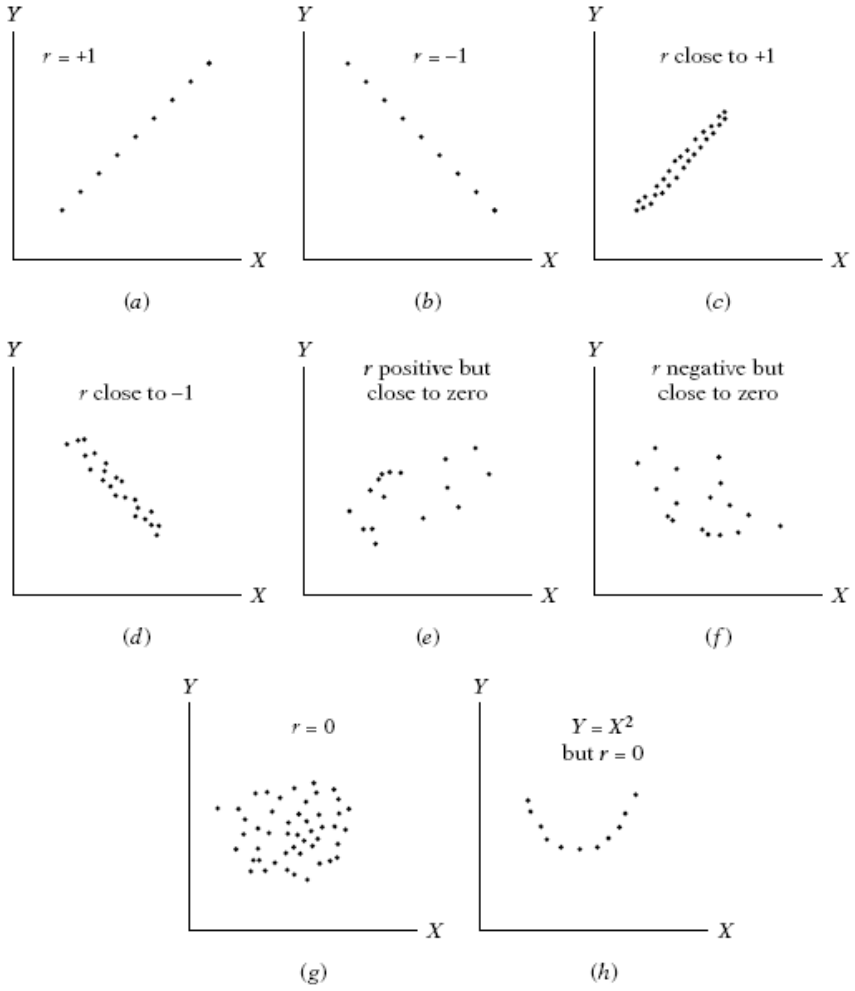
$$R^2 = 1 - \frac{\sum (Y_i - \hat{Y})^2}{\sum (Y_i - \bar{Y})^2} = 1 - \frac{5.82}{180} = 0.967$$

Suç oranı istatistiklerinin açıklanmasında işsizlik oranı oldukça önemli bir yer tutar, %97 oranında açıklar.

10.2. Korelasyon Katsayısı

Regresyon analizi ile bağımlı ve bağımsız değişken(ler) arasında ilişkinin varlığı, bu ilişkinin gücü sorgulandı ve değişkenler arasındaki ilişki matematiksel bir modelle açıklandı. Bu yeni konuda, iki değişken arasındaki ilişkinin ölçüsünün, büyüklüğü ve yönü incelenecek. İki değişken arasındaki ilişkinin ölçüsü değerlendirilirken Korelasyon Katsayısı (r) kullanılır. Korelasyon analizinde regresyon analizindeki gibi serpilme diyagramı çizilerek değişkenler arasındaki ilişki için genel bir bilgi edinilir. Korelasyon katsayısı –1 ve +1 arasında değiişm gösterir.

Aşağıda farklı korelasyon ilişkisini gösteren çizimlere yer verilmiştir (Gujarati, Temel Ekonometri 1998).



0-0.5 arası ilişki zayıf ilişki olarak değerlendirilir. Korelasyon katsayısının "-" olması değişkenler arasındaki ters yönlü ilişkiye işaret ederken, bu katsayının "+" olması aynı yönlü ilişkiye işaret eder.

Çizim (a) da aynı yönlü tam doğrusal bir ilişki söz konsudur.

Çizim (b)'de ise ters yönlü tam doğrusal bir ilişki söz konusudur.

Çizim (c) ve (d) tama yakın sırayla aynı yönlü ve ters yönlü ilişki söz konusudur.

(e) ve (f) çizimlerinde ise korelasyon katsayısı sıfıra yakındır. Çizim (g)'de ilişki yoktur, korelasyon katsayısı sıfırdır. Son çizim (h) de de korelasyon katsayısı sıfırdır ama bu durum değişkenler arasında ilişki olmamasından değil ilişkinin doğrusal olmamasından kaynaklanır.

Regresyon analizinde değişkenler arasındaki ilişkinin matematiksel ifadesi öne çıkar, değişkenler bağımlı ve bağımsız değişkenler olarak ayrılır, bağımlı değişken, rastlantı değişkenidir bağımsız değişkense kontrol edilebilen değişkendir. Korelasyon analizinde ise değişkenlerde bağımlı ve bağımsız ayırımı yoktur tüm değişkenler rastlantı değişkenidir.

Regresyon analizinin korelasyon analizinden bir diğer farklı yaklaşımıysa, regresyon analizinin değişkenler arasındaki neden sonuç önbilgisini sınaması fakat korelasyon katsayısının her zaman nedensel bir ilişkinin göstergesi olmamasıdır.

Korelasyon katsayılar ilgilenilen değişkenlerin türlerine göre sınıflanabilir. Buna göre eğer ilgilenilen değişkenler nitelse kullanılan korelasyon katsayılarından bazıları şöyledir:

-Phi Katsayısı

-Cramer Katsayısı

-Lambda Katsayısı

-Gamma Katsayısı

-Kendall'ın tau Katsayısı

İlgilenilen değişkenler nicelse kullanılan korelasyon katsayıları şöyle sıralanabilir:

-Pearson Korelasyon Katsayısı

-Spearman Korelasyon Katsayısı

Burada sıralanan korelasyon katsayılarına daha sonra üzerinde detaylı bir biçimde durulacak olan Kısmi Korelasyon katsayısını da eklemek mümkündür. Kısmi korelasyon katsayısı iki değişken arasındaki ilişkiyi gösterirken diğer değişkenlerin etkilerini dikkate almaz.

10.2.1. Pearson Korelasyon Katsayısı

İki değişken arasındaki doğrusal ilişkiyi gösteren bir diğer ölçütte hatırlanacağı gibi Kovaryans katsayısıdır. Ancak kovaryans katsayısı ölçü birimlerine bağlı bir büyüklüktür.

$$Kov(X,Y) = \sigma_{xy} = \frac{\sum (X-\bar{X})(Y-\bar{Y})}{n-1}$$

Korelasyon katsayısı ise kovaryans katsayısının standart sapma değerlerine oranlanmasıyla elde edilir ve ölçü birimlerinden bağımsızdır. Korelasyon katsayısı şöyle ifade edilir:

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

Anakütleye ait korelasyon katsayısı yukarıdaki gibi olmakla beraber örneğe ait olan korelasyon katsayısı ise şöyle gösterilir:

$$r = \frac{S_{xy}}{S_x S_y} = \frac{\dfrac{\sum (X-\bar{X})(Y-\bar{Y})}{n-1}}{\sqrt{\dfrac{\sum (X-\bar{X})^2}{n-1}}\sqrt{\dfrac{\sum (Y-\bar{Y})^2}{n-1}}} = \frac{\sum (X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum (X-\bar{X})^2}\sqrt{\sum (Y-\bar{Y})^2}}$$

$$r = \frac{N\sum XY - \sum X \sum Y}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}}$$

$$x = X - \bar{X} \qquad y = Y - \bar{Y}$$

Korelasyon katsayısının regresyon denkleminin gücünü ifade eden determinasyon katsayısıyla ve eğim katsayısı β_1 ile arasında ilişki vardır. Şöyle ki;

Eğim katsayısı, ortalamadan sapmalarla $\hat{\beta}_1 = \dfrac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2}$ şeklinde gösterilmek üzere,

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2}\sqrt{\sum (y-\bar{y})^2}}$$

eğim katsayısı ve korelasyon katsayısı birbirine oranlanırsa,

$$\frac{\hat{\beta}_1}{r} = \frac{\sqrt{\sum (y-\bar{y})^2}}{\sqrt{\sum (x-\bar{x})^2}}$$

ve sonrasında eşitliğin sağ tarafı $\sqrt{n-1}$ 'e bölünürse,

$\dfrac{\hat{\beta}_1}{r} = \dfrac{s_y}{s_x}$ elde edilir.

Eğim katsayısı ve korelasyon arasındaki ilişki şöyle gösterilebilir: $\hat{\beta}_1 = r\dfrac{s_y}{s_x}$

Anlaşılacağı üzere korelasyon katsayısı ve regresyon katsayısının işaretleri aynı olacaktır.

Hatırlanacağı üzere regresyonun temel eşitliğinde, toplam değişme açıklanabilen ve açıklanamayan değişmeye eşittir.

$$\sum (y_i - \overline{y})^2 = \hat{\beta}_1^2 \sum (x - \overline{x})^2 + \sum (y_i - \hat{y}_i)^2$$

$$\sum (y_i - \overline{y})^2 = r^2 \frac{s_y^2}{s_x^2} \sum (x - \overline{x})^2 + \sum (y_i - \hat{y}_i)^2$$

$$r^2 = \frac{\sum (y_i - \overline{y})^2 - \sum (y_i - \hat{y}_i)}{\sum (y_i - \overline{y})^2}$$

$$r^2 = \frac{\sum (\hat{y}_i - \overline{y})^2}{\sum (y_i - \overline{y})^2}$$

Buradan açıkça basit regresyondaki determinasyon katsayısının korelasyon katsayısının karesine yani $(r)^2$ eşit olduğu görülebilir.

Örnek:

Muhasebe dersinden başarının matematik dersinden başarıya bağlı olup olmadığını araştırmak için seçilen öğrencilere ait notlar aşağıdaki gibidir. Korelasyon katsayısını hesaplayınız.

Muhasebe	**Matematik**
1	2
2	3
3	5
5	6
6	7
7	10
8	7
8	8

Korelasyon katsayısı ve regresyon denklemi arasındaki ilişkiyi değerlendirmek için öncelikle regresyon denklemi bulunmak istenirse;

y	x	y^2	x^2	xy
1	2	1	4	2
2	3	4	9	6
3	5	9	25	15
5	6	25	36	30
6	7	36	49	42
7	10	49	100	70
8	7	64	49	56
8	8	64	64	64
40	**48**	**252**	**336**	**285**

Normal denklemler:

$40=8b_0+48b_1$

$285=48b_0+336b_1$

$b_1=0.9375$

$b_0=-0.625$

$$r=\frac{N\sum XY-\sum X\sum Y}{\sqrt{[N\sum X^2-(\sum X)^2][N\sum Y^2-(\sum Y)^2]}}$$

$$r=\frac{8(285)-(48)(40)}{\sqrt{[8(336-48^2)][8(252-40^2]}}=0.90$$

Regresyon denkleminin eğimini veren katsayı 0.9375'tir. Bu katsayının pozitif olması aynı yönlü ilişki olduğunu gösterir. Nitekim bulunun korelasyon katsayısı da pozitifit. Korelasyon katsayısı, r, oldukça yüksektir. Buradan muhasebe ve matemetik dersleri arasında güçlü doğrusal bir ilişki olduğu sonucuna varılır.

Tahmini y değerleri üzerinden bulunacak determinasyon katsayısının karekökü de aynı sonucu verecektir. Şöyle ki;

y	**ytah**	**(y-ytah)^2**	**(y-yort)^2**
1	1,25	0,0625	16
2	2,188	0,035344	9
3	4,063	1,129969	4
5	5	0	0
6	5,938	0,003844	1
7	8,75	3,0625	4
8	5,938	4,251844	9
8	6,875	1,265625	9
40	**40**	**9,811626**	**52**

$$R^2 = 1 - \frac{9.81}{52} = 0.811$$

$$r = \sqrt{1 - \frac{9.81}{52}} = 0.9$$

Örnek:

Önceki iki bölümde ele alınan işsizlik ve suç oranı istatistikleri verisinden hareketle korelasyon katsayısını bulunuz.

X Y

işşizlik oranı	**suç oranı**
0,8	3
1,4	6
2,3	7
3,5	15
4,5	19
12,5	50

$$\sum Y = 50 \quad \sum X = 12.5 \quad \sum XY = 164.9$$
$$\sum X^2 = 40.39 \quad \sum Y^2 = 680$$
$$\sum (X - \bar{X})(Y - \bar{Y}) = 39.9$$
$$\sum (X - \bar{X})^2 = 9.14 \quad \sum (Y - \bar{Y})^2 = 180$$

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2}\sqrt{\sum (y - \bar{y})^2}} = \frac{39.9}{\sqrt{9.14 * 180}} = 0.983$$
$$(r)^2 = R^2$$

Örnek:

Aşağıda bir firmanın reklam harcamalarına ve satışlarına ait bilgi mevcuttur. Satış rakamlarının reklam harcamalarına göre regresyon denklemini bularak eğim katsayısının anlamlılığını sınayın ve regresyon denkleminin gücünü değerlendirin.

Reklam Harcamaları(X)	Satışlar(Y)
10	3
20	4
30	6
40	7
50	10

Y	X	Y^2	X^2	YX	Y-Yort=y	X-Xort=x	xy	x^2
3	10	9	100	30	-3	-20	60	400
4	20	16	400	80	-2	-10	20	100
6	30	36	900	180	0	0	0	0
7	40	49	1600	280	1	10	10	100
10	50	100	2500	500	4	20	80	400
30	**150**	**210**	**5500**	**1070**	**0**	**0**	**170**	**1000**

$$b_0 = \frac{30(5500) - 150(1070)}{5(5500) - 150^2} = 0.9$$

$$b_1 = \frac{5(1070) - 30(150)}{5(5500) - 150^2} = 0.17$$

$$b_1 = \frac{170}{1000} = 0.17$$

$b_0 = 6 - (0.17)30 = 0.9$

Reklam harcamalarındaki 1 birimlik değişme satıişlarda 0.17 birimlik değişime neden olur.

Y	$\hat{Y}$	$(Y-\hat{Y})^2$	$(Y-\overline{Y})^2$
3	2,6	0,16	9
4	4,3	0,09	4
6	6	0	0
7	7,7	0,49	1
10	9,4	0,36	16
30	30	1,1	30

$$R^2 = 1 - \frac{1.1}{30} = 0.963$$

Reklam harcamaları satışları %96 oranında açıklamaktadır.

Eğim parametresi için güven aralığı şöyledir:

$$0.17 - (3.181)(0.02) \le b_1 \le 0.17 + (3.181)(0.02)$$
$$0.106 \le b_1 \le 0.234$$

$$S = \sqrt{\frac{1.1}{5-2}} = 0.605$$

$$S_{\hat{b}_1} = \frac{0.605}{\sqrt{1000}} = 0.02$$

Eğim parametresinin testinde bulunan değer tablo kritik değeri 3.18'den büyük olduğu için bağımsız değişkenin ağımlı değişkeni açıklama da yetersiz olduğunu savunan sıfır hipotezi reddedilir.

$$H_0 : b_1 = 0$$
$$H_1 : b_1 \neq 0$$

$$t = \frac{0.17}{0.02} = 8,5$$

Örnek: Bir ilacın dozu ve iyileşme süresine ait seri aşağıdaki gibidir. Korelasyon katsayısını ve regresyon denklemini bulunuz.

İlaç Dozu(x)	İyileşme süresi
0	90
1	70
4	70
6	40
7	50
12	10

$$n = 6, \quad \sum X = 30, \quad \sum Y = 330$$
$$\sum X^2 = 246, \quad \sum Y^2 = 22100, \quad \sum XY = 1060$$

$$r = \frac{\sum XY - n\bar{X}\bar{Y}}{\sqrt{\sum (X^2 - n\bar{X})(\sum (Y^2 - n\bar{Y})}}$$
$$= \frac{1060 - 6(5)(55)}{\sqrt{(22100 - 6(55^2))(246 - 6(5^2)}} = -0.95812$$

$$b_1 = \frac{1060 - 6(55)(5)}{246 - 6(5^2)} = -6.146$$

$$b_0 = 55 - (-6.146 * 5) = 85.729$$

$$Y = 85.729 - 6.146X$$

Örnek: Aşağıda yer alan x,y ikilisinden hareketle regresyon denklemini oluşturunuz, eğim katsayısının anlamlılığını değerlendiriniz, determinasyon ve korelasyon katsayısını bulunuz.

X	Y
2	18
3	22
5	25
6	26
8	30
9	32
12	40

$$n = 7, \quad \sum X = 45, \quad \sum Y = 193$$
$$\sum X^2 = 363, \quad \sum Y^2 = 5633, \quad \sum XY = 1391$$

$$r = \frac{1391 - 7(45/7)(193/7)}{\sqrt{(363 - 7((45/7)^2))(5633 - 7((193/7)^2)}} = 0.9912$$

$$b_1 = \frac{1391 - 7(45/7)(193/7)}{363 - 7(45/7)^2} = 2.039$$

$$b_0 = (193/7) - 2.039(45/7) = 14.465$$

X	Y	ytah	y-ytah=e	hata kare
2	18	18.543	-0.543	0.294849
3	22	20.582	1.418	2.010724
5	25	24.66	0.34	0.1156
6	26	26.699	-0.699	0.488601
8	30	30.777	-0.777	0.603729
9	32	32.816	-0.816	0.665856
12	40	38.933	1.067	1.138489
45	193		-0.01	5.317848

$$s = \sqrt{\frac{5.32}{7-2}} = 1.032$$

$$s_{b1} = \frac{1.032}{\sqrt{73.72}} = 0.12$$

$$t = \frac{2.039}{0.12} = 16.99$$

$$R^2 = 1 - \frac{5.32}{2993} = 0.998$$

Bölüm Soruları

1-Aşağıda matematik dersine çalışma süresi ve alınan notlara ilişkin bilgi mevcuttur. Determinasyon ve korelasyon katsayılarını hesaplayarak yorumlayınız.

Çalışma saati	**Alınan not**
3	30
5	45
6	55
8	65
9	80

2- Uygulanan ilaç dozu ve iyileşme sürelerine ilişkin bilgi mevcuttur. Determinasyon ve korelasyon katsayılarını bularak yorumlayınız.

Doz	**İyileşme süresi**
1	8
2	8
3	6
4	4
5	2

BÖLÜM 10 ÇOKTAN SEÇMELİ SORULARI

1-Aşağıdaki ifadelerden hangisi ya da hangileri doğrudur?

I) Korelasyon katsayısı 2 değişken arasındaki ilişkinin yönünün ve büyüklüğünün belirlenmesinde kullanılır.
II) Korelasyon katsayısının mutlak değerce 1'e yaklaşması güçlü bir ilişki göstergesidir.
III) Determinasyon katsayısı korelasyon katsayısının karesine eşittir.
a)Yalnız I b) Yalnız II c)Yalnız III d)I ve II e)Hepsi

2- X ve Y iki değişkene karşılık gelmek üzere, aşağıdaki bilgiden hareketle elde edilecek olan pearson korelasyon katsayısı kaçtır?

$$\sum_{i=1}^{7} Y_i = 70 \quad \sum_{i=1}^{7} X_i = 140 \quad \sum_{i=1}^{7} X_i^2 = 3596 \quad \sum_{i=1}^{7} X_i Y_i = 1726 \quad \sum_{i=1}^{7} Y_i^2 = 854$$

a)0.93 b) 0.98 c)-0.87 d)-0.92 e)-0.98

3- 8 'er gözlemden oluşan X ve Y değişkenleri arasındaki korelasyon katsayısı aşağıdakilerden hangisine eşittir?

$$\sum_{i=1}^{8} Y_i = 400 \quad \sum_{i=1}^{8} X_i = 80 \quad \sum_{i=1}^{8} (X_i - \overline{X})^2 = 368 \quad \sum_{i=1}^{8} (X_i - \overline{X})(Y_i - \overline{Y}) = 1345 \quad \sum_{i=1}^{8} (Y_i - \overline{Y})^2 = 5750$$

a)0.72 b)0.82 c)0.92 d)-0.72 e)-0.92

A) Aşağıdaki soruları burada verilen bilgiden hareketle yanıtlayınız:
6 adet gözlemden oluşan X bağımsız ve Y bağımlı değişkenlerine ait değerler şöyledir,

$$\sum_{i=1}^{6} Y_i = 385 \quad \sum_{i=1}^{6} X_i = 41 \quad \sum_{i=1}^{6} X_i^2 = 297 \quad \sum_{i=1} X_i Y_i = 2795$$

4- Regresyon denkleminin eğim katsayısı aşağıdakilerden hangisine eşittir?

a) -12.01 b) -9.76 c) -6.26 d)9.66 e)12.23

5- Regresyon denkleminde bağımlı değişken ve tahmin edilen bağımlı değişken arasındaki farkı gösteren hata terimlerinin toplamı aşağıdakilerden hangisine eşittir?

a)0 b)2 c)12 d)14 e)20

6)Regresyon denkleminde hata kareler toplamı aşağıdakilerden hangisine eşittir?

a) 0 b) 200 c)890 c)1280 d)1571 e)1600

7) Regresyon denkleminde determinasyon katsayısı kaçtır?

a) 0.78 b) 0.86 c) 0.91 d)0.93 e)0.98

8)Regresyon denkleminde X değişkeni Y değişkenin % kaçını açıklar?

a) 0.78 b) 0.86 c) 0.91 d)0.93 e)0.98

9-X ve Y değişkenleri arasındaki korelasyon katsayısı kaçtır?

a)0.95 b)0.90 c)0.88 d) 0.83 e)0.80

10- Regresyon denkleminde sabit katsayı kaçtır?

a)-3.18 b)-1.81 c)-0.78 d) 0.79 e)1.81

Yanıtlar

1) E 2) A 3)C 4)D 5)A 6)D 7)C 8)C 9)A 10)A

Bu Bölümde Ne Öğrendik Özeti

Determinasyon katsayısının ve korelasyon katsayısının hangi amaçla kullanıldığı, nasıl hesaplandığı ve yorumlandığı konuları ele alındı.

11. Kİ-KARE TESTLERİ

Bu Bölümde Neler Öğreneceğiz?

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

1) Eğitim düzeyi ile izlenen televizyon programları arasında ilişki var mıdır?

2) Kadınların makyaj yapma sıklığı ile yaşları arasında bir ilişki var mıdır?

3) 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

11.1. Ki-Kare Testlerini Neden Kullanırız?

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

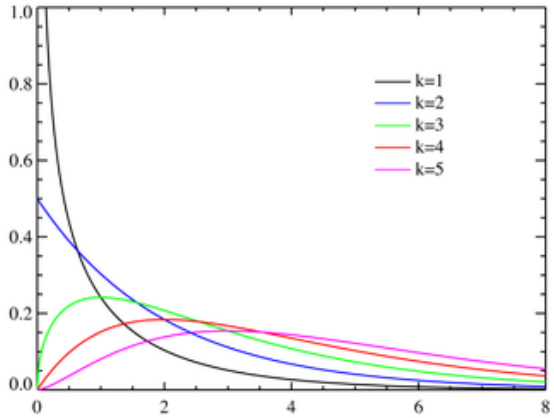
Ki kare testlerinde χ^2 dağılımı kullanılır. Kısaca bahsetmek gerekirse ki-kare dağılımı, X tesadüfi değişken dağılımından bir değer seçilip standart hale dönüştürülüp, karesinin alınmasıyla bulunur.

$$Z_i = \frac{X_1 - \mu}{\sigma}$$

Standardize edilen X tesadüfi değişkeni, standart normal değişken Z_i'nin aşağıdaki şekilde karesi alındığında ki-kare dağılımına dönüşür.

$$Z_i^2 = (\frac{X_i - \mu}{\sigma})^2$$

Ki-kare dağılımlarının serbestlik derecesi n ise şu şekilde gösterilir: χ_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.

11.2. Ki-Kare Bağımsızlık Testi

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

Partition A Class	B_1	B_2	...	B_C	Sums
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 sütün toplamlarının 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.

$$\chi^2_{test} = \sum \frac{(G_i - B_i)^2}{B_i} \qquad \chi^2_{tablo} = \chi^2_{(k-1)(r-1),\alpha}$$

$$\chi^2_{test} < \chi^2_{tablo} \rightarrow H_0 \quad reddedilemez$$

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	5
17-22 yaş	12	19	5	6
22+	6	5	5	6

H_0:Evlenme usulü yaşa bağımlı değildir.

H_1: bağımlıdır.

Gözlenen	**Görücü**	**İsteyerek**	**Zorla**	**Akraba**	**Satır toplamı**
13-16yaş	15	20	10	5	50
17-22 yaş	12	19	5	6	42
22+	6	5	5	6	22
Sütun toplamı	33	44	20	17	114

B_{11}=(29*46)/114= 13.3 … B_{34}=(17*22)/114=..

Tüm beklenen değerler yukarıdaki gibi hesaplanırsa şu tabloya ulaşılır;

	Görücü	**İsteyerek**	**Zorla**	**Akraba**
13-16yaş	13,3	20,2	9,2	5,2
17-22 yaş	11,3	17,2	7,8	5,7
22+	5,4	6,6	6	5,1

$\chi^2 = \frac{(15-13.3)^2}{13.3} + \frac{(20-20.2)^2}{20.2} + ... + \frac{(3-1.1)^2}{1.1} = 9.69$ hesaplanan bu değer tablo değeri ile kıyalanırsa,

$$\chi^2_{3x2,0.05} = 12.59$$

(Ki-kare tablosu bölüm sonunda yer almaktadır)

H_0 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 gözeye 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}= (140x180)/260=75.38 B_{12}=(120x140)/260=64.62

B_{21}= (140x120)/260=64.62 B_{22}=(120x120)/260=55.38

$$\chi^2 = \frac{(80-75.38)^2}{75.38} + \frac{(60-64.62)^2}{64.62} + ... + \frac{(60-55.38)^2}{55.38} = 1.429$$

Tablo değeri $\chi^2_{1x1,0.05} = 3.841$ olduğundan sigara içme alışkanlığı ve cinsiyet arasında ilişki yoktur hipotezi kabul edilir.

11.3. Ki-Kare Homojenlik Testi

Bu test ile iki ya da daha fazla bağımsız örneklemin aynı anakütleden seçilip seçilmediğinin araştırılmasında kullanılır. Bağımsızlık testinden farklı olarak, sıfır hipotezi örneklemlerin aynı ana kütleden seçildiği şeklinde kurulur.

H_0: Örneklemler aynı ana kütleden seçilmiştir.

H_1: Örneklemler farklı ana kütlelerden seçilmiştir.

Örnek: Bir süpermarket 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 örneklemler 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}= (50x70)/125=28 . . . B_{23}= (15x75)/125=9

Şampuan	Beğenmeyen	Kararsız	Beğenen
A	28	16	6
B	42	24	9

H_0: İki örneklem aynı ana kütleden seçilmiştir.

H_1: Örneklemler farklı ana kütlelerden seçilmiştir.

$$\chi^2 = \frac{(30-28)^2}{28} + ... + \frac{(10-9)^2}{9} = 0.631$$

Tablo değeri $\chi^2_{1x2,0.05} = 5.99$ olduğundan örneklemler aynı anakütleden seçilmiştir.

11.4. Ki-Kare Uyum Testi

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 $\overline{x} = ((32x0) + (1x16) + (2x8) + (3x4))/60 = 0.734$

$$p(x=0) = \frac{e^{-0.734}0.734^{0}}{0!} = 0.48$$

$$p(x=1) = \frac{e^{-0.734}0.734^{1}}{1!} = 0.35$$

$$p(x=2) = \frac{e^{-0.734}0.734^{2}}{2!} = 0.13$$

$$p(x=3) = \frac{e^{-0.734}0.734^{3}}{3!} = 0.04$$

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.

0.45x60=27…

$$\chi^{2} = \frac{(32-27)^{2}}{27} + \ldots + \frac{(4-2.4)^{2}}{2.4} = 3.21$$

$\chi^{2}_{2,0.05} = 5.99$ (Burada serbestlik derecesi k-m-1 den, 4-1-1 şeklinde bulunmuştur.

Buna göre dağılım poisson'a uymaktadır.

11.5. Kontenjans Katsayısı

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	Lise	Üniversite	Yüksek 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.

$$\chi^2 = \frac{(50-36)^2}{36} + \ldots + \frac{(30-15)^2}{15} = 42.93$$

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.

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? (α =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?

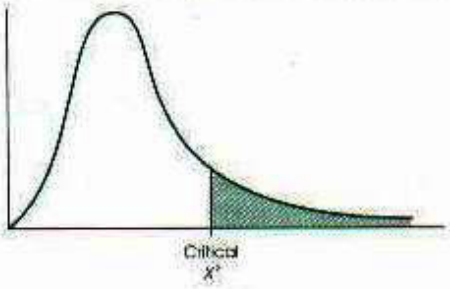
	Akciğer Kanseri	
Sigara içme	Var	Yok
Kullanan	20	80
Kullanmayan	5	95

3- Tablo değerlerini $\alpha = 0.01$ ve $\alpha = 0.05$ alarak değerlendirin

4- Kontenjans katsayısını bularak yorumlayın

TABLE B.7 THE CHI-SQUARE DISTRIBUTION*

*The table entries are critical values of χ^2.

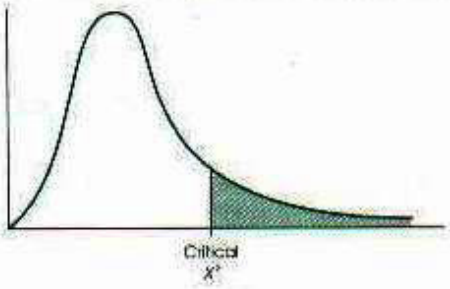


df	PROPORTION IN CRITICAL REGION 0.10	0.05	0.025	0.01	0.005
1	2.71	3.84	5.02	6.63	7.88
2	4.61	5.99	7.38	9.21	10.60
3	6.25	7.81	9.35	11.34	12.84
4	7.78	9.49	11.14	13.28	14.86
5	9.24	11.07	12.83	15.09	16.75
6	10.64	12.59	14.45	16.81	18.55
7	12.02	14.07	16.01	18.48	20.28
8	13.36	15.51	17.53	20.09	21.96
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.

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.

BÖLÜM 11 SORULARI

1-Aşağıdaki ifadelerden hangisi ya da hangileri doğrudur?

I) İki değişken arasında ilişki yoksa bu iki değişkenin bağımsız olduğu söylenir
II) Sınıflama düzeyinde ölçülmüş iki değişkenin bağımsızlığı kikare analizi ile araştırılır.
III) Kikare bağımsızlık testinde serbestlik derecesi (satır sayısı-1)x(sütun sayısı-1) ile bulunur
a)Yalnız I b)Yalnız II c)Yalnız III d)I veII e) Hepsi

A-Aşağıdaki soruları tablodan hareketle çözünüz.

	sigara kullanan	sigara kullanmayan
Kadın	5	115
Erkek	20	160

2-Yukarıdaki tablodan hareketle uygun sıfır hipotezi aşağıdakilerden hangisidir?

a) Kadınlarda sigara içme alışkanlığı daha azdır
b) Erkeklerde sigara içme alışkanlığı daha fazladır
c) Sigara içme alışkanlığı ile cinsiyet arasında ilişki yoktur
d) Sigara içme alışkanlığı ile cinsiyet arasında ilişki vardır
e) Hiçbiri

3-Tablodaki serbestlik derecesi kaçtır?

a) 5 b) 4 c) 3 d)2 e)1

4-% 5 önem düzeyinde tablodan bulunacak kritik değer nedir?

a)1.96 b) 3.84 c)5.52 d)9.22 e)13.82

5-Tablodan hesaplanacak kikare değeri aşağıdakilerden hangisine eşittir?

a)1.04 b)2.21 c)3.18 d)4.57 e)8.41

B)Tablodan hareketle aşağıdaki soruları yanıtlayınız

Dört farklı yabancı dil eğitimine katılan öğrencilerin başarı ve başarısızlık durumları tabloda özetlenmiştir.

	A	B	C	D
Başarılı	23	30	40	25
Başarısız	27	20	10	25

6)Tablodaki bilgi ile uyumlu olan alternatif hipotez aşağıdakilerden hangisidir?

a) Eğitim türü ve öğrencinin başarısı arasında ilişki vardır

b)Eğitim türü ve öğrencinin başarısı arasında bir ilişki yoktur

c)C türü eğitim en iyi sonucu verir

d)C türü eğitim diğer tekniklerden daha üstündür

e)Hiçbiri

7) Tablodaki serbestlik derecesi kaçtır?

b) 5 b) 4 c) 3 d)2 e)1

8) % 5 önem düzeyinde tablodan bulunacak kritik değer nedir?

a)1.96 b) 3.84 c)5.52 d)7.81 e)13.82

9) Tablodan hesaplanacak kikare değeri aşağıdakilerden hangisine eşittir?

a)11.04 b)12.21 c)13.18 d)14.37 e)18.41

10) Kikare bağımsızlık testinin uygulanacağı 60 gözlemden oluşan 3x3 kontenjans tablosu için uygun serbestlik derecesi aşağıdakilerden hangisidir?

a)59 b) 57 c)9 d) 4 e)3

Yanıtlar:

1)E 2) C 3) E 4) B 5) A 6) A 7) C 8) D 9)D 10) D

Bu Bölümde Ne Öğrendik Özeti

Önceki bölümlerde nicel değişkenler arasındaki ilişkilerin analizi değerlendirilmiştir. Bu bölümde ise nitel değişkenler arasında bir ilişkinin olup olmadığı sorusuna kikare testleri ile yanıt verilmiştir. Test süreci hipotez testleri başlığı altında incelediğimiz süreçten farklı değildir.

12. VARYANS ANALİZİ TABLOSU

Bu Bölümde Neler Öğreneceğiz?

Bu bölümün konusunu Varyans Analizi (ANOVA) oluşturmaktadır. İkiden fazla anakütle ortalamalarını karşılaştırmak amacıyla ANOVA kullanılmaktadır.

Bölüm Hakkında İlgi Oluşturan Sorular

1) Aynı öğretim üyesinin farklı dört bölüme verdiği başarı oranları arasında fark var mıdır?

2) Bir tarlaya uygulanan 3 farklı gübreden elden edilen sonuçlara bakarak verim farklılığı olduğunu söyleyebilir misiniz?

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	**Kazanım**	**Kazanımın nasıl elde edileceği veya geliştirileceği**
ANOVA	Birden fazla anakütle ortalamalarının birbirine eşit olup olmadığının sınanabilmesi	Metinler, çözümlü problemler.

Anahtar Kavramlar

- ANOVA
- F testi

Giriş

Bu haftaki dersimizin konusunu Varyans Analizi (ANOVA) oluşturmaktadır. İkiden fazla örnekleme ilişkin test yapılmak istenildiğinde ANOVA kullanılmaktadır.

12.1. VARYANS ANALİZİ (ANOVA)

Varyans analizi ikiden fazla örneklem grubunun ortalamalarının birbirine eşit olup olmadığının sınanmasında kullanılır. Önceki derslerden hatırlanacağı üzere iki örneklem ortalamasının testi için örneklem büyüklüğüne bağlı olarak z ve t testleri kullanılmıştı.

Bu iki testi ikiden fazla örneklem için kullanılması halinde I. Tip hata artacak yani doğru olduğu halde yanlışlıkla sıfır hipotezi reddedilebilecektir. Örneğin, 3 örneklem için t testi ikili gruplar 1-2, 1-3, 2-3 halinde ayrı ayrı uygulanacaktır ve her testin I.tip hata yapmama olasılığı 0,95 olduğunda üç ayrı t testinde I.tür hata yapmama olasılığı 0,857'ye düşecektir (0,95 * 0,95 * 0,95 = 0,857). Yani I.tür hata yapma olasılığı 0,05'ten 0,143'e yükselir (1 - 0,857 = 0,143). Bu sonuç kabul edilemeyeceğinden ANOVA testine başvurulur.

ANOVA testinin uygulanabilmesi için çeşitli varsayımların sağlanması beklenir:

- örneklemlerin elde edildiği anakütlelerin dağılımı normal olmalı

- bağımsız örneklemler olmalı

-anakütle varyansları eşit olmalı

ANOVA tek yönlü ve çift yönlü şeklinde sınıflandırılabilir

12.2. Tek Yönlü Anova

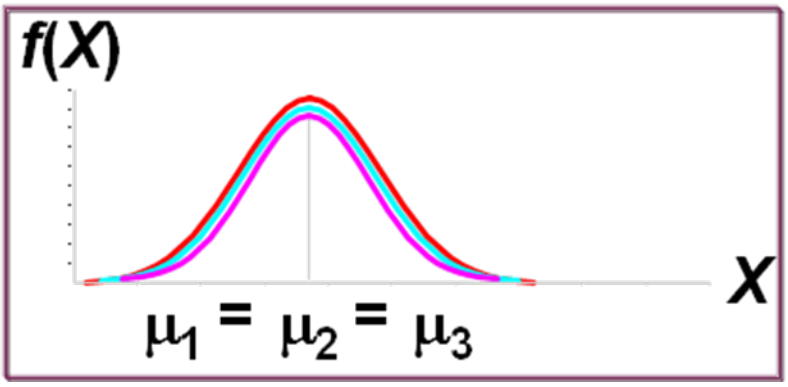
ANOVA ile amaç k tane anakütleden n hacimli bağımsız tesadüfi örnekler seçildiğinde, bu örneklerin ortalamalarından hareketle anakütle ortalamalarının birbirinden farklı olup olmadığı test etmektir.

Hipotezler şöyle ifade edilir:

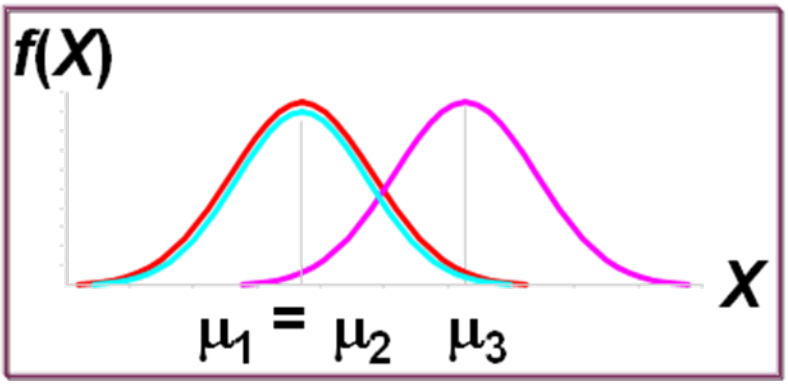
$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$$

$$H_1 : En \text{ az iki anakütle ortalaması birbirine eşit değildir}$$

Sıfır hipotezinde tüm anakütle ortalamalarının eşit olduğu varsayılır:



Alternatif hipotez de ise en az bir anakütlesi ortalaması diğerlerinden farklıdır.



	İşlemler
	1 2 … i … k X_{11} X_{21} … X_{i1} … X_{k1} X_{12} X_{22} … X_{i2} … X_{k2} . . X_{1n} X_{2n} … X_{in} … X_{kn}
Toplam	T_1 T_2 T_i T_k T
Ortalama	$\overline{X}_1$ $\overline{X}_k$ $\overline{X}$

Anakütle ortalamalarını karşılaştırmak için varyans analizi uygulanırken, herbir gruptaki gözlemlerin, X_{ij}, grup ortalaması olan $\overline{X}_i$ değerlerinden farklarına dayalı olan grup içi değişim ve gruplara ait olan ortalama değerler $\overline{X}_i$ genel ortalama değeri olan $\overline{X}$ dan farklarının kareler toplamına dayalı olan gruplar arası değişim iki önemli ölçüttür. Grup içi değişim ve gruplar arası değişim toplamı, Toplam Değişkenliği verir.

ANOVA ile gruplar (örnekler) arasında uygunluk olup olmadığı yani incelenen grupların aynı anakütleye ait olup olmadıkları sorusuna cevap aranır.

$$\sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij}-\overline{X})^2$$

Toplam Değişkenlik (Genel Kareler Toplamı-GKT)

Gruplar arası Değişkenlik (Gruplar arası Kareler Toplamı-GAKT) $n\sum_{i=1}^{k}(X_i - \overline{X})^2$

Grup ortalamalarının genel ortalamadan gösterdiği sapmalar hesaplanır.

Gruplar içi Değişkenlik (Gruplar içi Kareler Toplamı- GİKT) $\sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij} - \overline{X}_i)^2$

Her bir gözlemin dâhil olduğu grubun ortalmasından gösterdiği sapmalar hesaplanır.

Toplam değişimi oluşturan gruplar içi ve gruplar arası değişim aslında varyans analizinde test sürecinin temelini oluşturur.

$$\sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij} - \overline{X})^2 = n\sum_{i=1}^{k}(X_i - \overline{X})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij} - \overline{X_i})^2$$

Test sürecinin temel varsayımı anakütle varyanslarının eşit olduğudur. Eğer gruplar birbirine eşitse yani sıfır hipotezi doğru ise GİKT anakütle varyansının tahmininde kullanılır, yansız tahminciyi bulmak için "n-k" ya bölünür. Bulunan bu değer grup içi kareler ortalaması olarak isimlendirilir. Benzer şekilde gruplar arası kareler ortalama değeri de "k-1" serbestlik derecesine oranlanır.

Eğer sıfır hipotezi doğru ise gruplar arası kareler ortalama değeri gruplar içi kareler ortalama değerine oranlanır.

Değişim Kaynağı	**Kareler Toplamı**	**Serbestlik Derecesi**	**Kareler Ortalaması**	**Test İstatistiği**
işlem	GAKT	**v_1=k-1**	$s_1^2 = \frac{GAKT}{k-1}$	
Hata	GİKT	**v_2= n-k**	$s_2^2 = \frac{GİKT}{n-k}$	
Toplam	GKT	n-1		$F = \frac{s_1^2}{s_2^2}$

Çeşitli önem seviyeleri ve örnek büyüklükleri için **s_1^2 / s_2^2** nin hangi noktaya kadar rastlantısallıktan, hangi noktadan sonra istatistiksel olarak önemli bir farktan yani grupların farklı anakütlelere ait olmasından kaynaklandığı F tablosu ile değerlendirilir.

Hesaplanan F değeri, F tablosundan elde edilen kritik değerden küçükse grup ortalamaları arasındaki farklılık tesadüfidir ve söz konusu gruplar aynı anakütleye aittir sonucuna varılır.

Eğer hesaplanan test istatistiği, kritik değerden büyükse örnek ortalamaları arasındaki fark istatistiksel olarak önemlidir ve bu grupların farklı anakütlelere ait olduklarına karar verilir.

F değeri, iki varyansın birbirine bölümü olduğu için negatif değer almaz.

Bu yüzden F dağılımı sağa çarpıktır. H_0 hipotezinin red bölgesi eğrinin sağ ucunda yer alır.

Aşağıda F tablosuna yer verilmiştir, görüleceği üzere, iki serbestlik derecesi vardır.

TABLE B.5 THE STUDENTIZED RANGE STATISTIC (q)*

*The critical values for q corresponding to $\alpha = .05$ (lightface type) and $\alpha = .01$ (boldface type).

df FOR ERROR TERM	k = NUMBER OF TREATMENTS										
	2	3	4	5	6	7	8	9	10	11	12
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17	7.32
	5.70	**6.98**	**7.80**	**8.42**	**8.91**	**9.32**	**9.67**	**9.97**	**10.24**	**10.48**	**10.70**
6	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49	6.65	6.79
	5.24	**6.33**	**7.03**	**7.56**	**7.97**	**8.32**	**8.61**	**8.87**	**9.10**	**9.30**	**9.48**
7	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16	6.30	6.43
	4.95	**5.92**	**6.54**	**7.01**	**7.37**	**7.68**	**7.94**	**8.17**	**8.37**	**8.55**	**8.71**
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05	6.18
	4.75	**5.64**	**6.20**	**6.62**	**6.96**	**7.24**	**7.47**	**7.68**	**7.86**	**8.03**	**8.18**
9	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74	5.87	5.98
	4.60	**5.43**	**5.96**	**6.35**	**6.66**	**6.91**	**7.13**	**7.33**	**7.49**	**7.65**	**7.78**
10	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.72	5.83
	4.48	**5.27**	**5.77**	**6.14**	**6.43**	**6.67**	**6.87**	**7.05**	**7.21**	**7.36**	**7.49**
11	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49	5.61	5.71
	4.39	**5.15**	**5.62**	**5.97**	**6.25**	**6.48**	**6.67**	**6.84**	**6.99**	**7.13**	**7.25**
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39	5.51	5.61
	4.32	**5.05**	**5.50**	**5.84**	**6.10**	**6.32**	**6.51**	**6.67**	**6.81**	**6.94**	**7.06**
13	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32	5.43	5.53
	4.26	**4.96**	**5.40**	**5.73**	**5.98**	**6.19**	**6.37**	**6.53**	**6.67**	**6.79**	**6.90**
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36	5.46
	4.21	**4.89**	**5.32**	**5.63**	**5.88**	**6.08**	**6.26**	**6.41**	**6.54**	**6.66**	**6.77**
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.31	5.40
	4.17	**4.84**	**5.25**	**5.56**	**5.80**	**5.99**	**6.16**	**6.31**	**6.44**	**6.55**	**6.66**
16	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15	5.26	5.35
	4.13	**4.79**	**5.19**	**5.49**	**5.72**	**5.92**	**6.08**	**6.22**	**6.35**	**6.46**	**6.56**
17	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11	5.21	5.31
	4.10	**4.74**	**5.14**	**5.43**	**5.66**	**5.85**	**6.01**	**6.15**	**6.27**	**6.38**	**6.48**
18	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07	5.17	5.27
	4.07	**4.70**	**5.09**	**5.38**	**5.60**	**5.79**	**5.94**	**6.08**	**6.20**	**6.31**	**6.41**
19	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04	5.14	5.23
	4.05	**4.67**	**5.05**	**5.33**	**5.55**	**5.73**	**5.89**	**6.02**	**6.14**	**6.25**	**6.34**
20	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01	5.11	5.20
	4.02	**4.64**	**5.02**	**5.29**	**5.51**	**5.69**	**5.84**	**5.97**	**6.09**	**6.19**	**6.28**
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.01	5.10
	3.96	**4.55**	**4.91**	**5.17**	**5.37**	**5.54**	**5.69**	**5.81**	**5.92**	**6.02**	**6.11**
30	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82	4.92	5.00
	3.89	**4.45**	**4.80**	**5.05**	**5.24**	**5.40**	**5.54**	**5.65**	**5.76**	**5.85**	**5.93**
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73	4.82	4.90
	3.82	**4.37**	**4.70**	**4.93**	**5.11**	**5.26**	**5.39**	**5.50**	**5.60**	**5.69**	**5.76**
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	4.73	4.81
	3.76	**4.28**	**4.59**	**4.82**	**4.99**	**5.13**	**5.25**	**5.36**	**5.45**	**5.53**	**5.60**
120	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56	4.64	4.71
	3.70	**4.20**	**4.50**	**4.71**	**4.87**	**5.01**	**5.12**	**5.21**	**5.30**	**5.37**	**5.44**
∞	2.77	3.31	3.63	3.86	4.03	4.17	4.28	4.39	4.47	4.55	4.62
	3.64	**4.12**	**4.40**	**4.60**	**4.76**	**4.88**	**4.99**	**5.08**	**5.16**	**5.23**	**5.29**

Table 29 of E. Pearson and H. Hartley, *Biometrika Tables for Statisticians*, 3rd ed. New York: Cambridge University Press, 1966. Adapted and reprinted with permission of the Biometrika trustees.

Örnek:

Fakülteye alınan 4 farklı markadaki 20 fotokopi makinesinin 1 sene içerisnde çıkardığı arıza sayıları aşağıdaki gibidir. Ortalama servis dışı kalma süreleri arasında fark olduğu söylenebilir mi %5 anlamlılık düzeyinde değerlendiriniz.

	I	**II**	**III**	**IV**
1	10	11	16	12
2	10	10	13	10
3	11	10	15	14
4	12	9	16	13
5	12	10	15	11
T_i	**55**	**50**	**75**	**60**

$$\overline{X}_I = \frac{55}{5} = 11 \qquad \overline{X}_{II} = \frac{50}{5} = 10 \qquad \overline{X}_{III} = \frac{75}{5} = 15 \qquad \overline{X}_{IV} = \frac{60}{5} = 12$$

$$GİKT = (10-11)^2 + (10-11)^2 + ... + (13-12)^2 + (11-12)^2 = 22$$

Genel ortalama= (55+50+75+60)/20=12

$$GAKT = 5(11-12)^2 + 5(10-12)^2 + 5(15-12)^2 + 5(12-12) = 70$$

Değişim Kaynağı	**Kareler Toplamı**	**Serbestlik Derecesi**	**Kareler Ortalaması**	**Test İstatistiği**
işlem (GAKT)	70	v_1=4-1	$s_1^2 = \frac{70}{3} = 23.333$	
Hata (GİKT)	22	v_2= 20-4	$s_2^2 = \frac{22}{16} = 1.375$	
Toplam (GKT)	92	20-1		$F = \frac{23.333}{1.375} = 16.97$

Tablo değerine bakılırsa $\alpha = 0.05$, v1=3, v2=16 serbestlik derecesi ile F=3.24

Test istatistiği $F = 16.97 >$ F_{tab}= 3.24

H_0 hipotezini reddedilir, en az iki grup ortalamasının birbirinden farklı olduğuna karar verilir.

ÖRNEK

X	Z	W	Q
32	36	42	16
35	30	28	18
38	30	23	22
26	20	15	14
46	24	26	20

Yukarıda verilen dört birimin ortalamaları arasında fark var mıdır %5 anlamlılık düzeyinde değerlendiriniz.

$$Ho: \mu_1 = \mu_2 = \mu_3$$
$$H_1: En \quad az\ biri\ farklıdır$$

$$Xort = 541/20 = 27.05$$
$$X_{1ort} = 35.4$$
$$X_{2ort} = 28$$
$$X_{3ort} = 26.8$$
$$X_{4ort} = 18$$
$$GAKO = [5(18-27.05)^2 + ... + 5(26.8-27.05)^2]/3 = 254.31$$
$$GIKO = ((32-35.4)^2 + ... + (20-18)^2)/(20-4) = 49.875$$
$$F = 254.31/49.875 = 5$$
$$F_{(3,16),0.05} = 3.24$$

Söz konusu ortalamalar farklıdır.

Örnek: Bir işletmenin 4 atölyesinde üretilen aynı ürünün ortalama ağırlıkları arasında 0.05 önem düzeyinde anlamlı bir farklılık var mıdır?

A	B	C	D
61	81	65	92
90	72	80	79
78	75	60	97
84	73	69	84
67	69		93
	80		

$H_0: \mu_1 = \mu_2 = ... = \mu_k$
$H_1: En$ az iki anakütle ortalaması birbirine eşit değildir

$$\overline{Y}_A = 76 \quad \overline{Y}_B = 75 \quad \overline{Y}_C = 68.5 \quad \overline{Y}_D = 89$$
$$\overline{Y} = 77.5$$
$$GAKT_{ort} = (5(76-77.5)^2 + 6(75-77.5)^2 + 4(68.5-77.5)^2 + 5(89-77.5)^2)/3 = 344.65$$
$$GIKT_{ort} = ((61-76)^2 + ... + (93-89)^2)/16 = 69.44$$
$$F = 4.96$$

F kritik değeri 3.24 olduğundan H0 reddedilir.

Örnek: Bir ürün A, B ve C atölyelerinde üretilmektedir. Bu üç atölyede üretilen ürünlerinin ortalama ağırlıkları arasında anlamlı bir fark olup olamadığını 0.05 önem düzeyinde değerlendiriniz.

A	B	C
8	7	6
10	6	8
9	8	9
13		5
		7

$$\overline{Y}_A = 10 \quad \overline{Y}_B = 7 \quad \overline{Y}_C = 7$$

$$\overline{Y} = \frac{40+21+35}{12} = 8$$

$$GAKT = 4(10-8)^2 + 3(7-8)^2 + 5(7-8)^2 = 24$$

$$GIKT = (8-10)^2 + ... + (7-7)^2 = 26$$

$$GAKT_{ort} = \frac{24}{3-1} = 12$$

$$GIKT_{ort} = \frac{26}{12-3} = 2.89$$

$$F = \frac{12}{2.89} = 4.15$$

F tablo değeri 4.26 olduğundan fark yoktur.

Bölüm Soruları

1-Üç farklı eğitim düzeyine sahip grubun haftalık gelirlerine ait veri aşağıdaki gibidir. Bu iç grubun ortalama gelirleri arasında fark var mıdır? (alfa=0.05)

Lise Mezunu	**Üniversite Mezunu**	**Yüksek Lisans**
300	520	530
320	540	538
400	490	600
410	470	650
500	600	

BÖLÜM 12 SORULAR

1- Birden fazla kitle ortalamasının testi için aşağıdakilerden hangisi kullanılır?
   a) İki örneklem t testi b) Kikare testi c)Korelasyon d)Regresyon e)ANOVA

A) Aşağıdaki soruları tablodan hareketle yanıtlayınız

3farklı yabancı dil eğitim tekniğinin uygulandığı bir toplulukta öğrencilerin aldığı notlar aşağıdaki gibidir. A, B ,C teknikleri arasında bir fark olup olmadığı araştırılacaktır.

A	B	C
8	8	8
10	6	7
9	7	6
13	9	
	5	

2- Sıfır hipotezi nasıl kurulmalıdır?
   a) Her üç tekniğin ortalama başarısı eşittir
   b) Tekniklerin ortalama başarıları birbirinden farklıdır
   c) En az bir tekniğin ortalama başarısı diğerlerinden farklıdır
   d) A tekniğinin başarısı diğerlerinden daha fazladır
   e) Hiçbiri

3-Gruplar arası kareler toplamı aşağıdakilerden hangisine eşittir?

a) 34 b) 28 c) 24 d) 26 e)16

4-Gruplar içi kareler toplamı aşağıdakilerden hangisine eşittir?
a)34 b) 28 c) 24 d) 26 e)16

5- Gruplar arası kareler ortalaması nedir?

a)12 b)13 c)14 d)15 e)16

6- %5 önem düzeyinde F tablosu kritik değeri nedir?

a)4.26 b) 5.26 c) 6.18 d)7.14 e)9.12

7-Gruplar içi kareler ortalaması nedir?

a)7.14 b)6.17 c)5.87 d)2.89 e)1.13

8- Hesaplanan F istatistik değeri kaçtır?

a)5.17 b)4.15 c)3.89 d)2.79 e)1.49

9) Aşağıdaki ifadelerden hangisi ya da hangileri doğrudur?

I) Sıfır hipotezi reddedilir

II)Sıfır hipotezi reddedilemez

III)Teknikler arasında önemli bir fark yoktur

a)Yalnız I b)Yalnız II c)Yalnız III d) I ve III e) II ve III

Yanıtlar

1)E 2)A 3)C 4)D 5)A 6)A 7)D 8)B 9)E

Bu Bölümde Ne Öğrendik Özeti

Bu bölümde ikiden fazla anakütleden seçilen grupların ortalamaları arasındaki farkların anlamlı olup olmadığı varyans analizi ile incelendi.

13. ENDEKS KAVRAMI

Bu Bölümde Neler Öğreneceğiz?

Bu bölümde günlük hayatta çok sık kullandığımız endekslere yer verilecektir.

Bölüm Hakkında İlgi Oluşturan Sorular

1) Bir sepetteki 5 malın 2005 yılındaki fiyatı 2006 yılına göre nasıl bir değişim göstermiştir?

2) Enflasyon hesapları neden farklı enstitüler tarafından farklı sonuçlar verir?

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	**Kazanım**	**Kazanımın nasıl 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 işittimiğimiz bir kavramdır. Endeks bir orandır, fiyat değişimleri, ücret değişimleri hakkında bize bilgi verir.

13.1. 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.

$$I = \frac{x_i}{x_0} x100$$

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)x100=100
2001	275	(275/350)x100=78.57
2002	200	(200/350)x100=57.14
2003	375	(375/350)x100=107.14
2004	390	(390/350)x100=111.43
2005	170	(390/350)x100=48.57
2006	280	(280/350)x100=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.

Endeklser basit ve bileşik, zaman ve mekân şeklinde farklı başlıklar altında incelenebilir.

Mekân Endeksi

Eğer endeksler bir mekân serisinden yola çıkılarak hesaplanıyorsa bu endekslere mekân endeksi denilir. İlgilenilen çeşitli değerlerin şehirler, ilçeler, ülkeler içindeki oransal değimine mekân endeksi ile ulaşılır. Mekân endeksinde temel alınan değer serideki tüm değerlerin aritmetik ortalamasından yola çıkılarak bulunur.

$$I = \frac{x_i}{\overline{x}} x100$$

Örnek: Aşağıda 5 farklı 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.

Xort=(20+15+12+10+8)/5= 3

Endeks
(20/13)x100 =153.85
(15/13)x100 =115.38
(12/13)x100 =92.31
(10/13)x100 =76.92
(8/13)x100= 61.54

Buna göre söz konusu 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.

$$I = \frac{x_i}{x_0} x100$$

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)**	**SEI**
2001	1280	(1280/1280)x100=100
2002	1300	(1300/1280)x100=101.5
2003	1360	(1360/1280)x100=106.3
2004	1100	(1100/1280)x100=85.9
2005	1400	(1400/1280)x100=109.4
2006	1490	(1490/1280)x100=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.

$$I = \frac{x_i}{x_{i-1}} x100$$

Örnek:

Önceki soruda yer verilen veriden hareketle Değişir Esaslı Endens 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.

13.2. Basit ve Bileşik 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 alınıyorda 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:

$$I = \frac{p_i}{p_0} x100$$

$$I = \frac{q_i}{q_0} x100$$

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)x100=% 126.5 artış oranı %26.5'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)x100= % 91. %9 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 tartlı aritmetik ortalaması

Basit Toplam Endeks

Endekse kapsamındaki maddelerin cari devre fiyatları (miktarları) toplanarak temel devre fiyatlar (miktarlar) toplamına oranlanır.

$$I = \frac{\sum p_1}{\sum p_0} x100$$

Örnek:

Aşağıda 4 maddeye ilişkin 2004 ve 2005 fiyatları verilmiştir. 2005 yılı basit toplam indeksi hesaplayınız.

Ürünler	2004 kg fiyatı	2005 kg fiyatı
Kaşar Peyniri	11000	12000
Zeytin	12000	14000
Beyaz Peynir	23000	24000
Tereyağ	18000	18500

Toplam fiyatlar dikkate alınırsa:

Ürünler	2004 kg fiyatı	2005 kg fiyatı
Kaşar Peyniri	11000	12000
Zeytin	12000	14000
Beyaz Peynir	23000	24000
Tereyağ	18000	18500
Toplam	64000	68500

I=(68500/64000)x100=107 Söz konusu 4 ürün ortalama olarak % 7 oranında artmıştır.

Basit Endekslerin Tartısız Aritmetik Ortalaması

Cari yıldaki fiyatlar temel devre fiyatlarına bölünür ve daha sonra ortalama alınır.

$$I = \frac{\sum(\frac{p_1}{p_0})}{n} x100$$

Örnek: Bir önceki soruda sunulan veriyi kullanarak 4 madde için Basit Endekslerin Tartısız Aritmetik Ortalamasını hesaplayın

Ürünler	2004 kg fiyatı	2005 kg fiyatı
Kaşar Peyniri	11000	12000
Zeytin	12000	14000
Beyaz Peynir	23000	24000
Tereyağ	18000	18500

Öncelikle her bir dönem için endeksler hesaplanır.

Ürünler	2004 p0	2005 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)x100=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.

$$I = \frac{\sum (\frac{p_1}{p_0}) xt}{\sum t} x100$$

Tartı endeksin çeşidine göre farklılık gösterir. Genel yaklaşıma göre tartlılar endeks kapsamındaki maddelerin miktar ve fiyatlarının çarpılmasıyla elde edilir. Tartılar değerlendirilirken temek 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.

13.3. Laspeyres Endeksi

Tartı hesaplanırken temel devre miktarı alınır.

$t=p0 \times q0$

Yukarıda belirtilen formülde yerine yerleştirilirse;

$$I = \frac{\sum (\frac{p_1}{p_0}) x(p_0 x q_0)}{\sum (p_0 x q_0)} x100$$

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.

	2004		2005	
Ürünler	**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

	2004		2005			
Ürünler	**Po**	**qo**	**p1**	**q1**	**p0q0**	**p1q0**
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)x100=130.8

İncelenen 4 tarım ürününün 2005 yılında fiyatları 2004 yılına göre %31 oranında artmıştır.

13.4. Paasche Endeksi

Endeks devresi miktarının tartı olarak kullanılmasıyla elde edilir.

$t=p_0xq_1$

Bu tartı formüle yerleştirilirse;

$$\frac{\sum(\frac{p_1}{p_0})x(p_0xq_1)}{\sum(p_0xq_1)}x100=\frac{\sum(p_1xq_1)}{\sum(p_0xq_1)}$$

Örnek:

Önceki örnekte ele alınan veriden hareketle Paasche endeksini hesaplayarak yorumlayın

	2004		2005			
Ürünler	**Po**	**Qo**	**p1**	**q1**	**p0q1**	**p1q1**
Kivi	420000	60	500000	70	29400000	35000000
Avakado	670000	70	690000	74	49580000	51060000
Brokoli	580000	80	750000	90	52200000	67500000
Muz	900000	100	990000	120	108000000	118800000
					239180000	272360000

I=(239180000/208500000)x100=113.8

İncelenen 4 tarım ürününün 2005 yılında fiyatları 2004 yılına göre %13.8 oranında artmıştır.

13.5. Fisher Endeksi

Laspeyres ve Paasche endeksleri hesaplanırken farklı ağırlıklandırmalar yüzünden farklı sonuçlara ulaşılur.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.

$$I = \sqrt{\frac{\sum p_1 x q_0}{\sum p_0 x q_0} x \frac{\sum p_1 x q_1}{\sum p_0 x q_1}} x100$$

Örnek:

Önceki örnekte ele alınan veri için Fisher Endeksini hesaplayarak yorumlayınız.

	2004		**2005**	
Ürünler	**Po**	**qo**	**p1**	**q1**
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,

p0q0	**p0q1**	**p1q1**	**p1q0**	
25200000	29400000	35000000	30000000	
46900000	49580000	51060000	48300000	
46400000	52200000	67500000	60000000	
90000000	108000000	118800000	99000000	
208500000	**239180000**	**272360000**	**237300000**	Toplam

$$I = \sqrt{\frac{\sum p_1 x q_0}{\sum p_0 x q_0} x \frac{\sum p_1 x q_1}{\sum p_0 x q_1}} x100 = \sqrt{\frac{237300000}{208500000} x \frac{272360000}{239180000}} x100 = 106.3$$

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.

- **Örnek:**

İki ürünün 2004-2007 yılları arasındaki fiyat ve miktarlarına ait bilgi aşağıdaki gibidir.2005 yılını baz yılı olmak üzere fiyat indekslerini Laspeyres formülüne göre hesaplayınız

Yıllar	Fiyat		Miktar	
	A-ürünü	B-ürünü	A-ürünü	B-ürünü
2004	255	190	1475	450
2005	273	209	1650	490
2006	292	248	1665	475
2007	372	374	1670	475

$$I_{Laspeyres} = \frac{\sum_{i=1}^{n} p_{1i} q_{0i}}{\sum_{i=1}^{n} p_{0i} q_{0i}}$$

$$I_{2004} = \frac{255(1650) + 190(490)}{273(1650) + 209(490)} 100 = 92.9$$

$$I_{2006} = \frac{295(1650) + 248(490)}{273(1650) + 209(490)} 100 = 109.13$$

Örnek:

İki ürünün 2004-2007 yılları arasındaki fiyat ve miktarlarına ait bilgi aşağıdaki gibidir. 2005 yılını baz yılı olmak üzere fiyat indekslerini Paasche formülüne göre hesaplayınız

Yıllar	Fiyat		Miktar	
	A-ürünü	B-ürünü	A-ürünü	B-ürünü
2004	255	190	1475	450
2005	273	209	1650	490
2006	292	248	1665	475
2007	372	374	1670	475

$$I_{Paasche} = \frac{\sum_{i=1}^{n} p_{1i} q_{1i}}{\sum_{i=1}^{n} p_{0i} q_{1i}}$$

$$I_{2004} = \frac{255(1475) + 190(450)}{273(1475) + 209(450)} 100 = 92.9$$

$$I_{2006} = \frac{292(1665) + 248(475)}{273(1665) + 209(475)} 100 = 109.1$$

Örnek: 6 ürüne ait 2005 ve 2010 yıllarına ait fiyatlar tablodaki gibidir. Geometrik ortalamaya göre fiyat indeksini hesaplayınız (2005=100)

Ürünler	2005 fiyat	2010 fiyat
A	100	200
B	150	300
C	125	250
D	400	800
E	75	150
F	100	200

$$I_{geometrik} = 100\sqrt[n]{\frac{p_{11}}{p_{01}}\frac{p_{12}}{p_{02}}..\frac{p_{1n}}{p_{0n}}} = 100\sqrt[6]{\frac{200}{100}\frac{300}{150}..\frac{200}{100}}$$

Bölüm Soruları

1)

Ürünler	**po**	**Qo**	**p1**	**q1**
portakal	60 0	12 0	70 0	13 0
mandalina	70 0	13 0	80 0	13 0
armut	50 0	15 0	50 0	12 0
elma	40 0	20 0	45 0	22 0

Yukarıdaki 4 madde için fisher fiyat endeksini hesaplayınız.

2) 2004 yılı baz yılı ise 2006 ve 2008 yılları için sabit esaslı endeksi hesaplayınız.

BÖLÜM 13 SORULAR

1- Aşağıdaki ifadelerden hangisi ya da hangileri doğrudur?
I) İndeks hesaplanırken Temel Devre dikkate alınır
II) Temel devre ile karşılaştırılan devreye cari devre denilir
III) Sabit indekste her cari devre,aynı temel devre ile karşılaştırılır
a)Hepsi b) Yalnız I c)Yalnız II d) Yalnız III e) I ve III

2- Aşağıdaki ifadelerden hangisi ya da hangileri doğrudur?

I)Bileşik indekste dikkate alınan değişkenlerin önem düzeyleri farklı olduğundan tartılar kullanılır

II)Mekan indeksi hesaplanırken temel devre olarak seriyi oluşturan değerlerin ortalaması alınır III)Eğer indeks hesaplanırken her değer, bir önceki devre değeriyle kıyaslanıyorsa değişken esaslı indeks değerine ulaşılır.

a) Yalnız I b) Yalnız II c) Yalnız III d) I ve III e) Hepsi

3-5 madde için Fisher indeksi %23.16 ve Paasche indeksi %21.61 hesaplanmış ise Laspeyres indeksi aşağıdakilerden hangisine eşittir?

a)24.82 b) 24.79 c) 23.16 d)22.89 e) 22.54

4-2005 yılını baz yılı olmak üzere fiyat indekslerini Paasche, Laspeyres ve Fisher formülüne göre hesaplayınız

Yıllar	Fiyat		Miktar	
	A-ürünü	B-ürünü	A-ürünü	B-ürünü
2004	250	210	1475	450
2005	279	229	1650	490
2006	300	250	1665	475
2007	372	378	1670	475

Bu Bölümde Ne Öğrendik Özeti

Bu bölümde endeks kavramı, endeks türleri, hesaplanması ve yorumlanması konularına değinilmiştir. Bir serideki fiyat/miktar değişimini yorumlamak istediğimizde kullanacağımız farklı endeksler farklı sonuçları verecektir, bu konuda örneklerle detaylı bilgi verilmiştir.

14. GENEL TEKRAR

Bu Bölümde Neler Öğreneceğiz?

Bu bölümde soru cevap şeklinde önceki bölümde ele aldığımız konuları gözden geçireceğiz.

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	**Kazanım**	**Kazanımın nasıl elde edileceği veya geliştirileceği**
Genel Tekrar	Farklı örnekler üzerinden şimdiye kadar öğrenilmiş konuların tekrarı	Çözümlü problemler

14.1. Çözümlü Örnekler

1- Regresyon ve korelasyon analizleri arasındaki farklar nelerdir?

-Regresyon analizi iki değişken arasındaki ilişkinin nedenselliğini inceler. Söz konusu ilişkide etkilyen ve etkilenen değişkenler belirlenir.

-Korelasyon katsayısı ise iki değişken arasındaki ilişkinin yönü ve büyüklüğüne odaklanır.

2- Aşağıdaki ifadelerin tek doğuruluğunu tartışın.

I) Regresyon denkleminden elde edilen hataların toplamı sıfırdır.

II) Regresyon denkleminden elde edilen Y değerleri (tahmini Y), toplamı sıfırdır

III) Determinasyon karesinin karekökü korelasyon katsayısını verir

IV) Korelasyon katsayısı -1 ve +1 ‘e yaklaştıkça güçlü bir ilişkiyi gösterir, bu sınırlar determinasyon katsayısının yorumunda da değişmez

3-Regresyon denkleminde hata kareler toplamı ile standart hata arasında nasıl bir ilişki vardır? 28 gözlemli bir regresyon denkleminde hata kareler toplamı 180 ise regresyon denkleminin standart hatası kaç olur?

Standart Hata= $(180/ (28-2))^{1/2}$= 2.63

5- İkiden fazla anakütle ortalamasının karşılaştırılmasında t/ z testi kullanmayı tercih eder misiniz? Neden

İkili ikili t /z testlerinin kullanımı testin gücünü düşüreceğinden ANOVA’ya başvurulur.

6-

Bir ülkenin2006-2011 tarihleri arasında gerçekleşen satış rakamları aşağıdaki gibidir. 2006 yılını temel devre olarak seçerek sabit esaslı endeksleri hesaplayın.

Yıllar	Satış	Sabit Esaslı
2006	2158	100
2007	2055	95.2
2008	2318	107.4
2009	2131	98.7
2010	2620	121.4
2011	2660	123.3

7- Öğretim düzeyi ile izlenen programlar arasında bir ilişki var mıdır?

		Öğrenim Düzeyi		
Programlar	**İlk**	**Orta**	**Yüksek**	**Toplam**
Magazin	50	20	10	80
Film	20	32	8	60
Belgesel	14	11	60	85
Eğlence	6	37	32	75
Toplam	90	100	110	300

		Öğrenim Düzeyi		
Programlar	**İlk**	**Orta**	**Yüksek**	**Toplam**
Magazin	50 (24)	20(26.7)	10(29.3)	80
Film	20(18)	32(20)	8(22)	60
Belgesel	14(25.5)	11(28.3)	60(31.2)	85
Eğlence	6(22.5)	37(25)	32(27.5)	75
Toplam	90	100	110	300

Sıfır hipotezi izlenen programlar ve eğitim düzeyi arasında ilişki yoktur şeklinde oluşturulur.

$$\chi^2 = \frac{(50-24)^2}{24} + \frac{(20-18)^2}{18} + \ldots + \frac{(32-27.5)^2}{27.5} = 119.82$$

Serbestlik derecesi (3-1)(4-1)=6

Tablo değeri 12.5 < 119.8.

Sıfır hipotezi reddedilir. İzlenen programlar ve eğitim düzeyi arasında ilişki vardır.

8)

Sigara içme ve cinsiyet arasında ilişki var mıdır?

sigara	**kadın**	**erkek**
kullanan	20	80
kullanmayan	5	95

sigara	**kadın**	**erkek**	
kullanan	20(12.5)	80(87.5)	100
kullanmayan	5(12.5)	95(87.5	100
	25	175	200

H0:Sigara içme ve cinsiyet arasında ilişki vardır.

Tablo değeri 3.841 (sd=(2-1)(2-1)=3.841

Sigara içme ve cinsiyet arasında ilişki vardır hipotezi reddedilir.

$$\chi^2 = \frac{(20-12.5)^2}{12.5} + \frac{(80-87.5)^2}{87.5} + \frac{(5-12.5)^2}{12.5} + \frac{(95-87.5)^2}{87.5}) = 10.28$$

9)

Bir ilköğretim okulundaki öğrencilerin ailelerinin eğitim düzeyi ile öğrencilerin başarı düzeyleri arasında bir ilişki olup olmadığını araştırmak için bir çalışma yapılıyor. Çalışma sonuçları aşağıdaki gibidir. Buna göre %5 anlamlılık düzeyinde bir ilişki olduğu söylenebilir mi?

Eğitim Düzeyi	**Başarılı**	**Başarısız**
İlköğretim	140	110
Lise	70	80
Üniversite	90	10

H_0: Öğrencinin başarısı ile ailesinin eğitim düzeyi arasında ilişki yoktur

H_1: Öğrencinin başarısı ile ailesinin eğitim düzeyi arasında ilişki vardır

Eğitim Düzeyi	**Başarılı**	**Başarısız**	**Toplam**
İlköğretim	140 (150)	110(100)	250
Lise	70(90)	80(60)	150
Üniversite	90(60)	10(40)	100
Toplam	300	200	500

Parantez içerisinde yer alan beklenen frekanslar şöyle bulunur:

İlk göze: (250*300)/500=150

Son göze: (200*100)/500=40

$$\chi^2 = \frac{(140-150)^2}{150} + \ldots + \frac{(10-40)^2}{40} = 50.278$$

$$\chi^2_{tablo} = 5.99$$

s.d=(3-1)(2-1)=2

Sıfır hipotezi reddedilir.

10)

Aşağıda reklam harcamaları ve satış rakamları verilmiştir. Satış rakamları ve reklam harcamaları arasındaki korelasyon katsayısını hesaplayınız.

Reklam harcamaları	**Satış rakamları**
14	28
35	66
22	38
29	70
6	22
15	27
17	28
20	47
12	14
29	68

x	y	x^2	y^2	xy
14	28	196	784	392
35	66	1225	4356	2310
22	38	484	1444	836
29	70	841	4900	2030
6	22	36	484	132
15	27	225	729	405
17	28	289	784	476
20	47	400	2209	940
12	14	144	196	168
29	68	841	4624	1972
199	**408**	**4681**	**20510**	**9661**

$$r = \frac{10x9661 - 199x408}{\sqrt{(10(4681) - 199^2)(10(20510 - 408^2))}} = 0.924$$

11)

X:{3,4,5,8,10} ve Y:{2,6,4,9,10} değişkenleri arasındaki basit doğrusal regesyon denklemini hesaplayınız. (Xbağımsız değişken)

X	**y**	**x^2**	**y^2**	**xy**
3	2	9	4	6
4	6	16	36	24
5	4	25	16	20
8	9	64	81	72
10	14	100	196	140
30	**35**	**214**	**333**	**262**

Xort=6

Yort=7

$$b_0 = \frac{35*214-30.262}{5*214-900} = -2.176$$

$$b_1 = \frac{5*262-35*30}{5*214-900} = 1.522$$

$$y = -2.176 + 1.522x$$

12)

X {80,75,93,73,84}, Y={180,184,193,176,182} bu iki değişken arasındaki **korelasyon** katsayısını hesaplayınız.

x	y	x-xort	y-yort	(x-xort)^2	(y-yort)^2	(x-xort)(y-yort)
80	180	-1	-3	1	9	3
75	184	-6	1	36	1	-6
93	193	12	10	144	100	120
73	176	-8	-7	64	49	56
84	182	3	-1	9	81	-3
405	**915**	**0**	**0**	**254**	**240**	**170**

$$r = \frac{170}{\sqrt{254}\sqrt{170}} = 0.689$$

Aynı yönlü ilişki.

13)

Faiz oranı X ={0.6,0.7,0.85,0.9,1.05,1.2} Talep edilen kredi sayısı Y={70,52,48,35,19,7}. Regresyon denklemini oluşturunuz.

x	y	xy	x^2
0.6	70	42	0.36
0.7	52	36.4	0.49
0.85	48	40.8	0.7225
0.9	35	31.5	0.81
1.05	19	19.95	1.1025
1.2	7	8.4	1.44
5.3	**231**	**179.05**	**4.925**

$$b_0 = \frac{231*4.92 - 5.3*179.05}{6*4.92 - 28.09} = \frac{187.56}{1.43} = 131.16$$

$$b_1 = \frac{6*179.05 - 5.3 - 231}{6*4.92 - 28.09} = -\frac{150}{1.43} = -104.9$$

$$y = 131.16 - 104.9x$$

14)

İlköğretim öğrencilerinin akıllı telefona sahip olma oranlarına ilişkin yapılan bir çalışma için görüşülen 500 öğrenciden 313'ü telefon sahibi olduğunu söylemiştir. Buna göre %98 güven düzeyinde anakütledeki telefon kullanım oranı için güven aralığını oluşturunuz.

$$\hat{p} = \frac{313}{500} = 0.626$$

$$0.626 - 2.33\sqrt{\frac{0.626*0.374}{500}} \le P \le 0.626 + 2.33\sqrt{\frac{0.626*0.374}{500}}$$

$$0.576 \le P \le 0.676$$

15)

Bir şirkette çalışan kadın ve erkeklerin memnuniyet oranı üzerine yapılan araştırmada görüşülen 80 kadından 72'si ve 150 erkekten 146'sı işyerinden memnun olduklarını söylemişlerdir. %90 güven düzeyini dikkate alarak anakütle oran farkları için güven aralığını luşturunuz. Memnuniyet açısından cinsiyete gmre fark var mıdır?

$$\hat{p}_1 = \frac{72}{80} = 0.9$$

$$\hat{p}_2 = \frac{143}{150} = 0.95$$

$$(0.9 - 0.95) - 1.65\sqrt{\frac{0.9*0.1}{80} + \frac{0.95*0.05}{150}} \le P_1 - P_2 \le (0.9 - 0.95) + 1.65\sqrt{\frac{0.9*0.1}{80} + \frac{0.95*0.05}{150}}$$

$$-0.11 \le P_1 - P_2 \le 0.01$$

Aralık sıfır değerini kapsıyor, cinsiyet memnuniyet oranları arasında fark yoktur.

16)

Bir ilacın etkinliğinin araştırılması için seçilen 8 hastanın ilaç kullanımı sonrası iyileşme süreleri gün olarak şöyledir: 8,5,6,5,7,8,9,6. Anakütle varyansının 4 olduğu biliniyorsa ilacın iyileştime süresinin 7 gün olduğu iddiası desteklenebilir mi?

$$Ho : \mu = 7$$
$$Ho : \mu < 7$$
$$\overline{x} = \frac{5+8+6+5+7+8+9+6}{8} = 6.75$$

$$z = \frac{6.75-7}{2/\sqrt{8}} = -0.35$$

%99 güven düzeyinde -2.58 tablo değeri olduğundan sıfır hipotezi kabul edilir.

17)

Mıknatıslı cam sileceği üreten bir firma ürünlerinin dayanıklıık süresinin 8 ay olduğunu iddia etmektedir. Bu ürünlerden rastgele olarak 49 tane seçilmiş ve ortalaması 7.3 ay, standart sapması 1.27 ay olarak bulunmuştur. %99 güven düzeyinde iddiayı test edeniz.

$$Ho : \mu = 8$$
$$Ho : \mu \neq 8$$

$$z = \frac{7.3-8}{1.27/\sqrt{49}} = -4.4$$

Sıfır hipotezi reddedilir!

18)

X	Y
50	20
60	15
70	30
90	35
100	50

X bağımsız Y bağmlı değişkenleri için basit doğrusal regrsyon denklemini hesaplayın.

X	Y	Xi-74	Yi-30	xiyi	xi^2
50	20	-24	-10	240	576
60	15	-14	-15	210	196
70	30	-4	0	0	16
90	35	16	5	80	256
100	50	26	20	520	676
				1050	1720

Xort=74 Yort=30

$$b_1 = \frac{1050}{1720} = 0.61$$

$$b_0 = 30 - 74(0.61) = -15.17$$

$$Y = -15.17 - 0.61X$$

Y tahmin	**hata=ei**	**e^2**	**y^2**
15.33	4.67	21.8	100
21.43	-6.43	41.34	125
27.53	2.47	6.1	0
39.73	-4.73	22.37	25
45.83	4.17	17.38	400
		108.99	**650**

$$\sigma = \sqrt{\frac{108.99}{5-2}} = 6.02$$

$$S_{b1} = \frac{6.02}{\sqrt{1720}} = 0.055$$

$$0.61 - 3.182 * 0.055 \leq \beta_1 \leq 0.61 + 3.182 * 0.055$$

$$0.434 \leq \beta_1 \leq 0.785$$

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

$$t = \frac{0.61}{0.055} = 11.09 \quad sıfırhipotezi \quad red$$

Eğim katsayısı anlamlıdır.

$$R^2 = 1 - \frac{108.99}{650} = 0.832$$

X, Y'yi %83 oranında açıklar

19)

Bir fabrikada üretilen margarinlerin ağırlıklarının ortalaması 200 gr. ve sapması 4 olan normal dağılıma uymaktadır. Margarinlerin ağırlığı 180gr ve210 gr dışına çıkarsa satışa sunulmamaktadır. Bu verilere göre satış sunulmayan margarinlerin yüzdesi nedir?

$$z_1 = \frac{\overline{x} - \mu}{\sigma} = \frac{210 - 200}{4} = 2.5$$

$$z_2 = \frac{\overline{x} - \mu}{\sigma} = \frac{180 - 200}{4} = -5$$

Tablo değerleri sırasıyla

$0.4938 \quad ve \quad 0.4999$

$$(0.5 - 0.4999) + (0.5 - 0.4938) = 0.0063$$

$\%0.63$

KAYNAKÇA

Esin A., Ekni M. ve Gamgam H., “İstatistik”, Gazi Kitabevi ,2006

Serper, O. (2002). Uygulamalı İstatsistik, Ezgi Kitabevi

Sınıksaran, E. (2011). *İstatistiksel Yöntemler*. İstanbul: Türkmen, 3. Baskı.

Anadolu Üniversitesi, İstatistik II Ders Notları