TEMEL İSTATİSTİK II

DERS NOTLARI

PROF.DR.YÜKSEL TERZİ

ONDOKUZ MAYIS ÜNİVERSİTESİ
FEN-EDEBİYAT FAKÜLTESİ
İSTATİSTİK BÖLÜMÜ

2018

1. İSTATİSTİKSEL İLİŞKİ KATSAYILARI

İki ayrı veri setinin birbiri ile olan ilişkisi bilinmek istenebilir. Bu iki veri setinden alınan verilerin ölçme düzeylerine göre farklı test istatistiği ile ilişkiler belirlenir. Eğer veriler aralıklı ya da oranlı ölçekli elde edilmiş ise parametrik testler, eğer veriler sınıflayıcı (adlandırma) ya da sıralayıcı özellikle elde edilmiş ise parametrik olmayan testler kullanılabilir. Parametrik olmayan testlerde örnek büyüklüğü arttıkça dağılım normal dağılıma yaklaşır. Büyük örnekler için $z \sim N_z(0,1)$ standart normal test istatistikleri kullanılır.

Değişkenler arasında ilişki varsa bu ilişkinin yönü ve derecesi de önem kazanır.

İstatistikte iki değişken arasındaki ilişkinin yönü ve derecesi hakkında bilgi veren çeşitli katsayılar mevcuttur. Bunlara Bağımlılık indeksi, Bağımlılık Katsayısı, Birlikte Değişim Katsayısı, Korelasyon (Pearson, Spearman, Kendall tau vb.) örnek olarak verilebilir.

Tablo 1.Veri Tiplerine Göre İlişki Katsayıları

İlişki Katsayısı	I. Değişken	II.Değişken	Açıklama
Pearson (r)	Sürekli	Sürekli	Serpme diyagramı doğrusal
Spearman (rs)	Sürekli, kesikli, sıralı	Sürekli, kesikli, sıralı	Serpme diyagramı doğrusal değil, n az ise
Phi (ϕ)	Sırasız niteliksel	Sırasız niteliksel	2x2 tablolarda
Cramer V	Sırasız niteliksel	Sırasız niteliksel	2x2 ve rxc tablolarda
Olağanlık Katsayısı (C)	Sırasız niteliksel	Sırasız niteliksel	2x2 ve rxc tablolarda
Eta (η)	Niteliksel	Sürekli, Kesikli	Değişkenlerden biri iki veya daha fazla kategorili niteliksel, diğeri sayısal veri tipinde
Goodman, Kruskal Gamma	Sıralı	Sıralı	Değişkenler sıralı ve çapraz tablo şeklinde
Somer d	Sıralı	Sıralı	Değişkenler sıralı ve çapraz tablo şeklinde

Varieties of correlation coefficients	
Correlation coefficient	Situation in which used
Pearson product moment	When both variables are continuous
Biserial	One scale continuous, the other an artificial dichotomy of a continous scale
Point biserial	One scale continuous and the other a true dichotomy
Phi	Both scales are dichotomous
Spearman's Rho	Both scales are rank orders
Kendall's Tau	Both scales are rank orders

1.1. Bağımlılık İndeksi

Nicel iki değişkenin ölçüm sonuçlarındaki değişimlerin işaretlerine bakılarak tespit edilir.

$$B.I. = \frac{D_p - D_n}{n-1} \quad , \quad -1 \leq B.I. \leq +1$$

D_p: İşareti birbirine uyanların sayısını
D_n: İşareti birbirine uymayanların sayısını
n : Toplam gözlem sayısını

1.2. Bağımlılık Katsayısı

Nicel iki değişkenin ölçüm sonuçlarındaki değişimlerin işaretlerine ve büyüklüklerini dikkate alarak aralarındaki ilişkiyi hesaplar.

$$B.K. = \frac{\sum_{i=1}^{n} \Delta X_i \Delta Y_i}{\sum_{i=1}^{n} \left| \Delta X_i \Delta Y_i \right|} \quad , \quad -1 \leq B.K. \leq +1$$

ΔX_i: X değişkenindeki değişimin büyüklüğü (X_{i+1}-X_i)
ΔY_i: Y değişkenindeki değişimin büyüklüğü (Y_{i+1}-Y_i)

1.3. Birlikte Değişim Katsayısı

Nicel iki değişkenin ölçüm sonuçlarındaki değişimlerin işaretlerine ve büyüklüklerini dikkate alarak aralarındaki ilişkinin yönünü ve derecesini hesaplar.

$$B.D.K. = \frac{\sum_{i=1}^{n} \Delta X_i \Delta Y_i}{\sqrt{\sum_{i=1}^{n} (\Delta X_i)^2 \sum_{i=1}^{n} (\Delta Y_i)^2}} \quad , \quad -1 \leq B.D.K. \leq +1$$

ΔX_i: X değişkenindeki değişimin büyüklüğü (X_{i+1}-X_i)
ΔY_i: Y değişkenindeki değişimin büyüklüğü (Y_{i+1}-Y_i)

Örnek 1.1. Aşağıdaki veri seti için değişkenler arasındaki ilişkinin yönünü ve derecesini bağımlılık indeksi, bağımlılık katsayısı ve birlikte değişim katsayısı ile bulunuz?

X	Y	X	Y	ΔX	ΔY	ΔX*ΔY	|ΔX*ΔY|	$(\Delta X)^2$	$(\Delta Y)^2$
15	20	…	…	…	…	…	…	…	…
10	10	-	-	10-15=-5	10-20=-10	50	50	25	100
20	15	+	+	20-10=10	15-10=5	50	50	100	25
15	20	-	+	15-20=-5	20-15=5	-25	25	25	25
10	35	-	+	10-15=-5	35-20=15	-75	75	25	225
30	30	+	-	30-10=20	30-35=-5	-100	100	400	25
25	40	-	+	25-30=-5	40-30=10	-50	50	25	100
TOPLAM						-150	350	600	500

İlk gözlem referans alındığından gözlem sayısı işlem yapılan veri adar alınır. Bu örnek için n=6 olur.

$$B.I.=\frac{D_p-D_n}{n-1}=\frac{2-4}{6-1}=-0.4 \qquad B.K.=\frac{\sum_{i=1}^{n}\Delta X_i\Delta Y_i}{\sum_{i=1}^{n}\left|\Delta X_i\Delta Y_i\right|}=\frac{-150}{350}=-0.42$$

$$B.D.K.=\frac{\sum_{i=1}^{n}\Delta X_i\Delta Y_i}{\sqrt{\sum_{i=1}^{n}(\Delta X_i)^2\sum_{i=1}^{n}(\Delta Y_i)^2}}=\frac{-150}{\sqrt{600\times 500}}=-0.27$$

X ile Y arasında ters yönde düşük bir ilişki vardır.

1.4. Kİ-KARE TESTİ

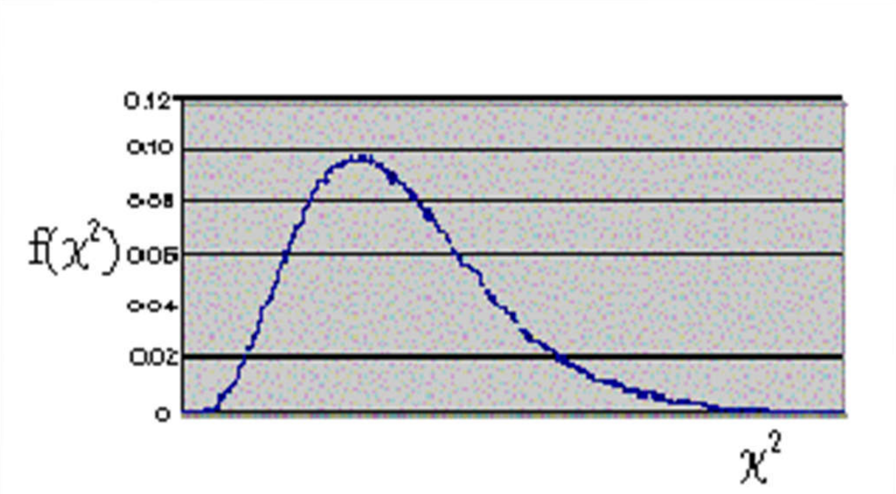
Günümüzde yapılan birçok araştırmada nicel (sayısal) değişkenlerden ziyade nitel (sayısal olmayan) değişkenlerin dikkate alındığı gözlemlenmektedir. Ayrıca bazen nicel değişkenler uygun biçimde gruplandırma ile nitel değişken durumuna getirilebilir. İşte sayısal olmayan (nitel) değişkenlere ki-kare (χ^2) testi uygulanır.

Normal dağılan bir anakütleden rasgele çekilen n hacimli örneklem için ki-kare istatistiği hesaplanır.

1. **Bir frekans dağılımının herhangi bir teorik dağılıma uyup uymadığının kontrolü için yapılan testler**
2. **İki veya daha fazla gruptaki oranların eşitliğinin testi için yapılan testler**
3. **İki özelliğin birbirinden bağımsız olup olmadığının testi**
4. **Örnek frekanslarının homojenlik kontrolü**

Yukarıdaki durumlarda χ^2 dağılışı kullanılarak hipotez testleri yapılabilir. χ^2 dağılışı sürekli ve sağa çarpık bir dağılıştır, üst kuyruğu daha uzundur.

Ki-kare dağılımı pozitif çarpık dağılıştır.



1.4.1. Kİ-KARE BAĞIMSIZLIK TESTİ

Değişkenlerin 2x2 yada rxc biçiminde ki çapraz tablolarda sınıflandırılması halinde, değişkenlerin arasında bağımlılık yada bir değişim olup olamadığını ortaya koyan testtir.

Serbestlik derecesi r satır sayısı ve c sütun sayısı olmak üzere (r-1)x(c-1) biçimindedir.

ÇAPRAZ TABLOLAR (CROSSTAB)

İki nitel değişkene ait gözlemler, rastgele n hacimli bir örnekle ele alınsın. Bir gözlemin seçimi, diğer gözlemin seçimini etkilemediği için gözlemlerin bağımsız olduğu söylenebilir. Veriler çapraz tablo biçiminde aşağıdaki gibi gösterilir.

Aralarında bağıntı bulunduğu düşünülen birinci değişkenin r şıkkı (satır), ikinci değişkenin c şıkkı (sütun) varsa rxc tablosu oluşturulur. Satır ve sütunların kesiştikleri yerlerde bulunan gözelere ilgili frekanslar kaydedilir.

rxc biçiminde oluşturulan tablolara kontenjans tablosu adı ver

rxc Kontenjans Tablosu

x \ y	1	2	…	j	…	c	Toplam
1	n_{11}	n_{12}	…	n_{1j}	…	n_{1c}	$n_{1.}$
2	n_{21}	n_{22}	…	n_{2j}	…	n_{2c}	$n_{2.}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
i	n_{i1}	n_{i2}	…	n_{ij}	…	n_{ic}	$n_{i.}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
r	n_{r1}	n_{r2}	…	n_{rj}	…	n_{rc}	$n_{r.}$
Toplam	$n_{.1}$	$n_{.2}$	…	$n_{.j}$	…	$n_{.c}$	$n_{..}$

Satır	Sütun		Toplam
	C1	C2	
S1	A A'	B B'	N1=A+B
S2	C C'	D D'	N2=C+D
Toplam	N3=A+C	N4=B+D	N

$$\chi^2 = \sum\sum (G_{ij} - T_{ij})^2 / T_{ij}$$

Burada G_{ij} gözlenen frekansları (A,B,C,D), T_{ij} (A',B', C', D') ise beklenen frekansları göstermektedir. Beklenen frekanslar satır ve sütun toplamlarının çarpımının genel toplama bölünmesi ile hesaplanır.

A'=N1×N3/N   B'=N1×N4/N   C'=N2×N3/N   D'=N2×N4/N

Örnek 1.2. **Bir okulda cinsiyet ile düzenli spor yapma arasında ilişki olduğu iddia edilmektedir. Bu amaçla rasgele seçilen 780 öğrenciden elde edilen bilgiler aşağıdaki tabloda verilmiştir.**

Cinsiyet	Düzenli Spor Yapma		Toplam
	Yapan	Yapmayan	
Bayan	246	122	368
Erkek	125	287	412
Toplam	371	409	780

Cinsiyet	Düzenli Spor Yapma		Toplam
	Evet	Hayır	
Erkek	246 175	122 193	368
Bayan	125 196	287 216	412
Toplam	371	409	780

$$\chi^2 = \sum\sum \frac{(G_{ij} - T_{ij})^2}{T_{ij}} = \frac{(246-175)^2}{175} + \frac{(122-193)^2}{193} + \frac{(125-196)^2}{196} + \frac{(287-216)^2}{216}$$

$$= 103.87$$

Örnek 1.3. **Dişlerin fırçalanması ile kişilerin eğitim durumlarının arasında bir ilişki olup olmadığı araştırılıyor. Günde diş Fırçalama sayısı (0,1,2 veya daha fazla), eğitim durumu (ilk öğretim, lise veya daha yüksek) şeklinde sınıflanıyor. Her iki değişkende sıralama ölçeğinde değişkenler olduğu için χ^2 bağımsızlık testi ile bu kontrol edilebilir.**

	0 Fırça	1 Fırça	2+ Fırça	$\sum$
İlk okul	5 (1.78)	27 (22.06)	5 (13.16)	37
Lise+	5 (8.22)	97 (101.94)	69 (60.84)	171
$\sum$	10	124	74	208

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(G_{ij} - T_{ij})^2}{T_{ij}}$$

Ki-Kare=5.833+1.107+5.063+1.262+0.24+1.095= 14.600

1.4.2. Çapraz Tablolarda İlişki Derecesi

Nitel iki değişken arasındaki ilişki derecesi bunların korelasyonunu verir. kxk tabloları için korelasyon aşağıdaki gibi hesaplanır.

$$r = \sqrt{\frac{\chi^2}{n \times (k-1)}}$$

Burada r 0 ile 1 arasında değerler alır. 0'a yakı olması ilişkinin olmadığını; 1'e yakın olması ise ilişkinin güçlü olduğunu gösterir. Formüldeki k değeri ise çapraz tablodaki satır=sütun sayısını gösterir. Eğer çapraz tablo 2x2 ise k=2 olur.

1.4.3. KONTENJANS(OLAĞANLIK) KATSAYISI

Nitel iki değişken arasındaki ilişkinin derecesini belirleyen bir katsayıdır. rxc tablolarında r>2 ve c>2 ise ki-kare değerinin gösterdiği ilişki düzeyini belirlemede kullanılır. İki değişken arasında bir ilişki bulunmuyorsa c=0 olur. İlişki yüksekse c 1'e yakın bir değer verir.

$$c = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

0≤C ≤1 olmakla beraber, kxk çapraz tablolarda C aşağıdaki değeri alır.

$$C_{max} = \sqrt{(k-1)/k}$$

2x2 tablolarda: $C_{max} = \sqrt{(2-1)/2}$ =0.707

3x3 tablolarda: $C_{max} = \sqrt{(3-1)/3}$ =0.816

4x4 tablolarda: $C_{max} = \sqrt{(4-1)/4}$ =0.866

Dikdörtgen çapraz tablolarda ise yukarıdaki gibi max bir değer bulunamaz.

Örnek 1.4. Örnek 1.2.'deki verilere göre korelasyonu ve kontenjans katsayısını bulunuz?

$$c=\sqrt{\frac{\chi^2}{\chi^2+n}}=\sqrt{\frac{103.87}{103.87+780}}=0.385$$

$$r=\sqrt{\frac{\chi^2}{n(k-1)}}=\sqrt{\frac{103.87}{780(2-1)}}=0.365$$

Çapraz tablo 2x2 olduğundan C max 0.707 değeri alabilir. C=0.385'in 1'e karşılık gelebilmesi için 1/0.707=1.414 ile çarpılması gerekir.

C=0.385x1.414=0.544 bulunur ve kuvvet açısından r gibi yorumlanır.

Orta seviyede bir ilişki vardır biçiminde yorumlanır.

Soru. Bir fakültedeki üç farklı bölümdeki öğrencilerin hafta sonları yapmış oldukları sosyal faaliyetler ile ilgili 300 kişi üzerinde yapılan bir anket çalışmasından aşağıdaki sonuçlar elde edilmiştir. Bu verileri kullanarak Ki-kare hesap değerini, korelasyon katsayısını ve kontenjans katsayısını bulunuz?

	Spor	Sinema	Tiyatro	Toplam
İstatistik	25	50	25	100
Matematik	25	50	15	90
Kimya	50	30	30	110
Toplam	100	130	70	300

Bölüm * Faaliyet Crosstabulation

			Faaliyet			Total
			Spor	Sinema	Tiyatro	
Bölüm	İstatistik	Count	25	50	25	100
		Expected Count	33,3	43,3	23,3	100,0
	Matematik	Count	25	50	15	90
		Expected Count	30,0	39,0	21,0	90,0
	Kimya	Count	50	30	30	110
		Expected Count	36,7	47,7	25,7	110,0
Total		Count	100	130	70	300
		Expected Count	100,0	130,0	70,0	300,0

Count: Gözlem değeri
Expected count: Beklenen değer

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(G_{ij} - T_{ij})^2}{T_{ij}}$$

$$= \frac{(25-33.3)^2}{33.3} + \frac{(50-43.3)^2}{43.3} + \frac{(25-23.3)^2}{23.3} + \frac{(25-30)^2}{30} + \frac{(50-39)^2}{39}$$

$$+ \frac{(15-21)^2}{21} + \frac{(60-36.7)^2}{36.7} + \frac{(30-47.7)^2}{47.7} + \frac{(30-25.7)^2}{25.7} = 21$$

$$c = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{21}{21+300}} = 0.256 \qquad r = \sqrt{\frac{\chi^2}{n(k-1)}} = \sqrt{\frac{21}{300(3-1)}} = 0.187$$

Çapraz tablo 3x3 olduğundan C max 0.816 değeri alabilir. C=0.256’nın 1’e karşılık gelebilmesi için 1/0.816=1.225 ile çarpılması gerekir.
C=0.256x1.225=0.314 bulunur ve kuvvet açısından korelasyon (r) gibi yorumlanır.

Zayıf bir ilişki vardır biçiminde yorumlanır.

1.5. Phi Katsayısı

Phi katsayısı ikili veri (0,1 veya başarılı-başarısız gibi) yapısına sahip iki nitel değişken arasındaki ilişkileri belirlemek için kullanılır. Örneğin iki hakemin kişileri başarılı-başarısız şeklindeki değerlendirmesi arasındaki güvenilirlik Phi katsayısı ile hesaplanır.

Phi katsayısı adlandırma ölçekle elde edilmiş 2x2 lik tablolarda verilerin korelasyon katsayısını bulmak için kullanılan parametrik olmayan bir testtir.

Phi katsayısı varsayımları:

a) Veriler adlandırma ölçekle ölçülmüş olmalı
b) 2x2 lik tablolarda kullanılabilir.
c) Örnekler anakütleden rasgele seçilmiş olmalıdır (Levin ve Fox, 1991).

	X^-	X^+	Σ
Y^-	a	b	e
Y^+	c	d	f
Σ	g	h	N

Phi Katsayısı,

$$\phi = \frac{ad - bc}{\sqrt{efgh}} \text{ veya } \phi = \sqrt{\frac{\chi^2}{N}}$$

$$\chi^2 = \frac{N\left(|ad - bc| - \frac{N}{2}\right)^2}{(a+b)(c+d)(a+c)(b+d)} \sim \chi^2_{1,\alpha}$$

Phi katsayısı ki-kareye bağlı olmayan formülle bulunursa negatif çıkabilir. Bu durumda phi katsayısı işaretsiz olarak yorumlanır.

Örnek 1.5. Cinsiyet ile sigara içme arasında bir ilişki olup olmadığını belirlemek için 39 kişi üzerinde yapılan bir çalışmadan aşağıdaki sonuçlar elde edilmiştir. Cinsiyet ile sigara arasındaki ilişki katsayısı bulunuz?

Cinsiyet	Sigara içen	Sigara içmeyen	Σ
Erkek	13	7	20
Bayan	5	14	19
Σ	18	21	39

$$\varphi = \frac{ad-bc}{\sqrt{efgh}} = \frac{13\times 14 - 5\times 7}{\sqrt{18\times 21 - 20\times 19}} = 0.388$$

$$\chi^2 = \frac{N\left[|ad-bc| - N/2\right]^2}{(a+b)(c+d)(a+c)(b+d)} = \frac{N\left[|ad-bc| - N/2\right]^2}{e\times f\times g\times h}$$

$$= \frac{39\left[|13\times 14 - 7\times 5| - 39/2\right]^2}{20\times 19\times 18\times 21} = 5.867$$

$$\varphi = \sqrt{\frac{\chi^2}{N}} = \sqrt{\frac{5.867}{39}} = 0.388$$

1.6. Cramér V Testi (İlişki Katsayısı)

Cramér V testinin uygulanabilmesi için satır ve sütün sayılarının eşit olmasına gerek yoktur. Örneğin tabloların 2x3, 3x5, 4x3 şeklinde ise ilişki katsayısı Cramer V testi ile bulunabilir.

Cramér V Testi Varsayımları:

a) Veriler adlandırma ölçekle ölçülmüş olmalı

b) 2x2 ve rxc tablolarda kullanılabilir.

c) Örnekler anakütleden rasgele seçilmiş olmalıdır (Levin ve Fox, 1991).

2x2 tablolardaq Phi katsayısı ile aynı sonucu verir.

$$V = \sqrt{\frac{\chi^2}{N\times \min\{(c-1),(r-1)\}}}$$

Örnek 1.6. Eğitim düzeyi ile davranış değişikliği arasında bir ilişki olup olmadığı araştırmak için 110 kişi üzerinde yapılan bir çalışmadan aşağıdaki sonuçlar elde edilmiştir. Eğitim düzeyi ile davranış değişikliği arasındaki ilişki katsayısı bulunuz?

Eğitim Düzeyi	Davranış Değişikliği			Toplam
	Yetersiz	Orta	Çok verimli	
İlköğretim	5	14	21	40
Lise	12	18	13	43
Yüksek	17	7	3	27
Toplam	34	39	37	110

$$\chi^2 = 23.13$$

$$V = \sqrt{\frac{23.13}{110(3-1)}} = 0.324$$

Eğitim düzeyi ile davranış değişikliği arasında orta düzeyde bir ilişki vardır.

1.7. İki Serili Korelasyon Katsayısı (r_b)

İlişki derecesi incelenecek değişkenlerden biri iki veya daha çok kategorili nitelik, diğeri sürekli sayısal veri türünde ise Eta istatistiği kullanılır. Eta istatistiği ortalama (iki veya daha çok) arasındaki farkın önemlilik testi sonucunda ortalamalardaki değişime ilişkin bir ilişki katsayısıdır. Eğer nitel değişken iki kategorili ise Eta istatistiği <u>İki Serili</u> veya <u>Nokta İki Serili Korelasyon katsayısı</u> adını alır. Nitel değişken doğal 2 kategorili (erkek-bayan, başarılı-başarısız gibi) ise nokta iki serili, doğal olmayan iki kategorili ise (<50yaş, >=50 yaş gibi) iki serili korelasyon katsayısı kullanılır.

$p=n_0/n$, $q=n_1/n$

$$r_b = \frac{\overline{Y}_1 - \overline{Y}_0}{s_Y} \times \frac{p.q}{H}$$

$$P(Z_1 < Z < \infty) = P$$

H yukarıdaki eşitliği sağlayan Z1 değerine karşı gelen değerdir ve Z tablosundan bulunur. P<0.5 ise Z1 artı, p>0.5 ise eksi işaret verilir.

Elde edilen korelasyon katsayısının işareti 0 ya da 1 olarak alınan gruba göre değişir. Bunun için mutlak değere göre yorum yapılır.

Örnek 1.7. 10 kişinin yaşları bir testten aldığı skor puanları bir araştırmadan elde edilmiştir. Kişilerin yaşları >70 ve <=69 olmak üzere iki gruba ayrılmış ve aşağıdaki gibi bulunmuştur. İlişki katsayısını bulunuz.

>70	<=69
6	10
7	7
4	8
5	3
2	9

Burada nitel değişken olan yaş grubu doğal olmayan nitel değişkendir. Bu yüzden puan ile görüşme sonucu arasındaki ilişki katsayısı iki serili korelasyon katsayısı ile bulunur.

p=5/10=0,5
q=5/10=0,5

$$P(Z_1 < Z < \infty) = 0.5$$

Z tablosundan 0.5'e en yakın değere karşılık gelen Z değeri 0.00 dir. Bu Z değerine karşı gelen Ordinat tablosu değeri 0.3989 bulunur.

$$f(Z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-Z^2/2\right\} = \frac{1}{\sqrt{2\times 3.1415}} \exp\left\{-0.00^2/2\right\} = 0.3989$$

$$r_b = \frac{\overline{Y}_1 - \overline{Y}_0}{s_Y} \times \frac{p.q}{H} = \frac{7.4-4.8}{2.6} \times \frac{0.5\times 0.5}{0.3989} = 0.63$$

1.8. Nokta İki Serili Korelasyon Katsayısı

İlişki derecesi incelenecek değişkenlerden biri iki kategorili ve doğal şekilde (erkek-bayan, başarılı-başarısız gibi), diğeri sürekli sayısal veri türünde ise Nokta İki Serili Korelasyon katsayısı kullanılır.

$$r_{pb} = \frac{\overline{Y}_1 - \overline{Y}_0}{s_Y}\sqrt{p_0 \times p_1} = \frac{\overline{Y}_1 - \overline{Y}_0}{s_Y}\sqrt{\frac{n_0}{n} \times \frac{n_1}{n}}$$

Örnek 1.8. 23 erkek ve 39 bayan sporcunun oluşturduğu bir örneklemde sporcuların boy uzunlukları ve sayıları aşağıdaki gibi bulunmuştur. Cinsiyet ile boy uzunlukları arasında bir ilişki var mıdır.

Boy Uz (cm)	Erkek (p1)	Bayan (p0)	Toplam
160-170	0	7	7
170-180	1	19	20
180-190	4	13	17
190-200	10	0	10
200-210	8	0	8
Toplam	23	39	62

$$p_1 = 23/62 = 0.37 \qquad p_0 = 39/62 = 0.63$$

$$\overline{Y}_1 = \frac{\sum m_i \times f_i}{\sum f_i} = \frac{165\times0+175\times1+185\times4+195\times10+205\times8}{23} = 195.87$$

$$\overline{Y}_0 = \frac{165\times7+175\times19+185\times13+195\times0+205\times0}{39} = 176.54$$

$$\overline{Y} = \frac{165\times7+175\times20+185\times17+195\times10+205\times8}{62} = 183.71$$

$$s_Y = \sqrt{\frac{\sum f_i\left(m_i - \overline{Y}\right)^2}{\sum f_i}} = 11.98$$

$$r_{pb} = \frac{\overline{Y}_1 - \overline{Y}_0}{s_Y}\sqrt{p_0 \times p_1} = \frac{195.87-176.54}{11.98}\sqrt{0.37\times0.63} = 0.78$$

Örnek 1.9. Bir çalışmada A ve B gibi iki gruba ilişkin ölçümler aşağıdaki gibidir. Nokta iki serili ilişki katsayısını bulunuz.

No	1	2	3	4	5	6	7	8	9	10
Grup	0	1	1	0	1	0	1	0	1	0
Ölçüm	3	6	6	4	4	3	6	2	5	3

No	11	12	13	14	15	16	17	18	19	20
Grup	1	1	0	1	0	1	0	0	1	0
Ölçüm	7	4	3	9	1	5	4	0	5	2

$$p_1 = 10/20 = 0.5 \qquad p_0 = 10/20 = 0.5$$

$$\bar{Y}_0 = 2.5 \qquad \bar{Y}_1 = 5.6 \qquad s_Y = \sqrt{\frac{\sum (Y_i - \bar{Y})^2}{n-1}} = 1.96$$

$$r_{pb} = \frac{\bar{Y}_1 - \bar{Y}_0}{s_Y}\sqrt{p_0 \times p_1} = \frac{5.6 - 2.5}{1.96}\sqrt{0.5 \times 0.5} = 0.79$$

2. KORELASYON

Korelasyon (Pearson korelasyon) analizi değişkenler arasında bir ilişki bulunup bulunmadığını, eğer varsa bu ilişkinin yönünü ve gücünü belirler. Ancak korelasyon katsayısı değişkenler arasındaki nedensel ilişkiyi göstermez.

Pearson korelasyona **Pearson Çarpım Momentler Korelasyonu** olarak da bilinir.

Örneğin "insanların boy uzunlukları ile ağırlıkları", "trafik kazalarının sayısı ile mevsim özellikleri", "zeka düzeyleri ile okuldaki başarı", "eğitim düzeyi ile yıllık kazanç", "cinsiyet ile sigara içme" gibi durumlar arasında ilişki olup olmadığı korelasyon ile belirlenebilir.

Bir değişkenin değerleri, diğer değişkenin değerleri ile doğrusal biçimde ilişkili ise ve değişkenler arasındaki ilişki düz bir doğru ile açıklanıyor ise değişkenler doğrusal ilişkilidir. Eğer iki değişken birbirinden tamamen bağımsız ve birbirini etkilemiyorsa, iki değişken arasında doğrusal ilişki yoktur denir.

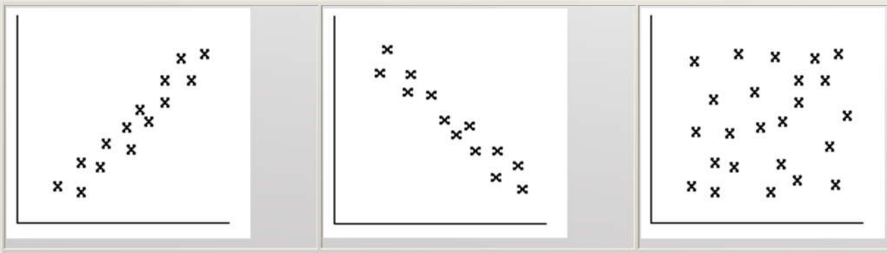
İki değişken arasındaki ilişki doğrusal olabileceği gibi eğrisel de olabilir. İkiden çok değişken arasındaki korelasyon çoklu korelasyon olarak adlandırılır.

Pearson korelasyon katsayısı, iki değişkenin de <u>sürekli</u> olmasını ve değişkenlerin <u>normal dağılım</u> göstermesini gerektirir. Eğer değişkenler normal dağılım göstermiyorsa Spearman Brown Sıra Farkları korelasyon katsayısı kullanılır.

<u>Pearson korelasyon katsayısının varsayımları:</u>

a) Değişkenlerinin her ikisi normal dağılımdan gelmiş olması gerekir.
b) Aralıklı ölçekle alınmış olması gerekir.
c) Değişkenleri her ikisi anakütleden tesadüfî olarak seçilmelidir.
d) x ve y değişkenleri arasında doğrusal bir ilişki olmalıdır.

Bir değişkenin değeri artarken diğer değişkenin değeri düzenli artıyor veya eksiliyorsa iki değişken arasındaki ilişki doğrusaldır. Bir değişkenin değerleri artarken diğer değişkenin değerleri de artıyorsa, o değişkenler pozitif ilişkilidir. Eğer bir değişkenin değerleri artarken diğer değişkenin değerleri azalıyorsa değişkenler negatif ilişkilidir.



Korelasyon=+1 | **Korelasyon=-1** | **Korelasyon=0**

2.1. Pearson Korelasyon Katsayısı

Doğrusal korelasyonun hesaplanmasında Pearson korelasyonu kullanılır. Bu formülün uygulanabilmesi için veriler en az aralıklı ölçekle toplanmalı ve süreklilik gösteren nicel bir değişken olmalıdır.

Korelasyon katsayısının değeri -1 ile +1 arasında değişir. Sonucun +1 çıkması iki değişken arasında kuvvetli olumlu ilişkinin bulunduğunu, -1 ise kuvvetli olumsuz ilişkinin bulunduğunu gösterir. Korelasyon katsayısı 0 'a yaklaştıkça ilişkinin kuvveti zayıflar, sıfır ise iki değişken arasında doğrusal bir ilişkinin olmadığını gösterir.

$$r=\frac{Cov(X,Y)}{\sqrt{V(X).V(Y)}}=\frac{\sum(X-\overline{X})(Y-\overline{Y})/n}{\sqrt{\frac{\sum(X-\overline{X})^2}{n}.\frac{\sum(Y-\overline{Y})^2}{n}}} \quad ,Cov(X,Y)=\frac{\sum(X-\overline{X})(Y-\overline{Y})}{n}$$

$$r=\frac{KT_{XY}}{\sqrt{KT_X.KT_Y}}=\frac{\sum(X-\overline{X})(Y-\overline{Y})}{\sqrt{\sum(X-\overline{X})^2\sum(Y-\overline{Y})^2}}=\frac{\sum XY-n\overline{X}.\overline{Y}}{\sqrt{[\sum X^2-n\overline{X}^2][\sum Y^2-n\overline{Y}^2]}}$$

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

Kovaryans:

$$Kov(X,Y)=\frac{\sum(X-\overline{X})(Y-\overline{Y})}{n} \qquad -\infty \text{ p } Kov\left(X,Y\right) \text{ p } +\infty$$

Ortak değişim anlamına gelen kovaryans, iki serinin aritmetik ortalamadan sapmalarının çarpımının aritmetik ortalamasıdır. Kovaryans negatif ise iki değişken arasındaki ilişki ters yönlü, pozitif ise doğru yönlü ilişki vardır denilir. Kovaryans sıfır ise ilişki yoktur.

Kovaryans serilerin ölçme biriminden etkilenir. Örneğin serilerden biri kg olarak verildiğinde bulunacak kovaryans ise aynı seriyi gr olarak verildiğinde bulunacak kovaryans farklı olacaktır. Serilerin ölçme biriminden etkilenmeyen ilişki ölçüsü olarak korelasyon kullanılması daha uygundur.

r (korelasyon değerleri)	
0.00-0.25	Çok Zayıf
0.26-0.49	Zayıf
0.50-0.69	Orta
0.70-0.89	Yüksek
0.90-1.00	Çok Yüksek

Korelasyonun karesine (r^2) determinasyon katsayısı – belirlilik katsayısı (açıklanan varyans) denir. Açıklanan varyans değişkenlerden birinde gözlenen değişkenliğin ne kadarının diğer değişken tarafından açıklandığını yorumlamada kullanılır.

ÖRNEK 2.1. **Aşağıda bir işletmede gün olarak kullanılan izin (X) ile performans puanları (Y) verilmiştir. Bu iki değişken arasında ilişki var mıdır?**

X	Y	X^2	Y^2	XY
1	14	1	196	14
2	13	4	169	26
3	12	9	144	36
3	13	9	169	39
2	11	4	121	22
1	12	1	144	12
4	12	16	144	48
5	11	25	121	55
4	14	16	196	56
3	13	9	169	39
6	12	36	144	72
5	12	25	144	60
10	10	100	100	100
9	11	81	121	99
1	14	1	196	14
8	11	64	121	88
9	10	81	100	90
7	9	49	81	63
6	12	36	144	72
7	10	49	100	70
$\sum x$ 96	$\sum y$ 236	$\sum x^2$ 616	$\sum y^2$ 2824	$\sum xy$ 1075

$$r=\frac{\sum XY-\frac{\left(\sum X\right)\left(\sum Y\right)}{n}}{\sqrt{\left(\sum X^2-\frac{\left(\sum X\right)^2}{n}\right)\left(\sum Y^2-\frac{\left(\sum Y\right)^2}{n}\right)}}=\frac{1075-\frac{96\times 236}{20}}{\sqrt{\left(616-\frac{96^2}{20}\right)\left(2824-\frac{236^2}{20}\right)}}=-0.74$$

Elde edilen sonuca göre kullanılan izin miktarı ile performans puanları arasında negatif yönlü kuvvetli ilişki vardır. Kullanılan izin miktarı arttıkça performans puanları düşmektedir.

Örnek 2.2. **Aşağıda ailedeki fert sayısı (X) ile aylık ekmek tüketimi (Y) verilmiştir. Bu iki değişken arasında ilişki var mıdır?**

X	Y	X^2	Y^2	XY
2	42	4	1764	84
6	78	36	6084	468
4	66	16	4356	264
5	50	25	2500	250
4	60	16	3600	240
3	40	9	1600	120
$\sum X=24$	$\sum Y=336$	$\sum X^2=106$	$\sum Y^2=19904$	$\sum XY=1426$

$$r=\frac{\sum XY-\frac{\left(\sum X\right)\left(\sum Y\right)}{n}}{\sqrt{\left(\sum X^2-\frac{\left(\sum X\right)^2}{n}\right)\left(\sum Y^2-\frac{\left(\sum Y\right)^2}{n}\right)}}=\frac{1426-\frac{24\times 336}{6}}{\sqrt{\left(106-\frac{24^2}{6}\right)\left(19904-\frac{336^2}{6}\right)}}=0.79$$

Örnek 2.3. **Aşağıda gübre miktarlarına (X) göre buğday verimleri (Y) verilmiştir. Bu iki değişken arasında ilişki var mıdır?**

X	Y	X^2	Y^2	XY
30	150	900	22500	4500
20	128	400	16384	2560
42	140	1764	19600	5880
50	200	2500	40000	10000
45	180	2025	32400	8100
65	210	4225	44100	13650
$\sum X=252$	$\sum Y=1008$	$\sum X^2=11814$	$\sum Y^2=174984$	$\sum XY=44690$

$$r=\frac{\sum XY-\frac{\left(\sum X\right)\left(\sum Y\right)}{n}}{\sqrt{\left(\sum X^2-\frac{\left(\sum X\right)^2}{n}\right)\left(\sum Y^2-\frac{\left(\sum Y\right)^2}{n}\right)}}=\frac{44690-\frac{252\times1008}{6}}{\sqrt{\left(11814-\frac{252^2}{6}\right)\left(174984-\frac{1008^2}{6}\right)}}$$

$$=0.89$$

Örnek 2.4. **Aşağıda verilen üç değişken arasındaki ikili (XY, XZ ve YZ) korelasyonları bulunuz?**

	X	Y	Z	X^2	Y^2	Z^2	XY	XZ	YZ
	2	6	6	4	36	36	12	12	36
	4	8	10	16	64	100	32	40	80
	5	10	14	25	100	196	50	70	140
	6	12	15	36	144	225	72	90	180
	8	14	20	64	196	400	112	160	280
	10	20	25	100	400	625	200	250	500
Toplam	35	70	90	245	940	1582	478	622	1216

$r_{xy}=0.98$ $r_{xz}=0.99$ $r_{yz}=0.98$

2.2.KISMİ KORELASYON KATSAYISI (Partial Correlation)

İki değişken arasında korelasyon katsayısının yüksek çıkması muhakkak o değişkenler arasında bir neden-sonuç ilişkisi bulunduğu anlamına gelmez. İki değişken arasında neden-sonuç ilişkisi olmadığı halde korelasyon katsayısı eğer yüksek çıkıyorsa, bu iki değişkenin üçüncü bir değişkenden etkilenmesi söz konusu olabilir.

Kısmi korelasyon katsayısı, iki değişken arasındaki ilişkinin bir yada daha çok değişkenin kontrol edilmesiyle hesaplanan bir değerdir.

$$r_{XY,Z} = \frac{r_{XY} - r_{XZ}r_{YZ}}{\sqrt{(1-r_{XZ}^2)(1-r_{YZ}^2)}}$$

$r_{XY,Z}$: Z değişkeni sabitken (etkisi giderildiğinde), X ile Y arasındaki kısmi korelasyon
r_{XY} : X ile Y arasındaki korelasyon
r_{XZ} : X ile Z arasındaki korelasyon
r_{YZ} : Y ile Z arasındaki korelasyon

Örnek 2.5. Talep (Y), gelir (X) ve fiyat (Z) serileri arasındaki ilişki araştırılıyor. Fiyatı sabit alarak talep ile gelir arasındaki kısmi korelasyonu bulup ve doğrusal ilişki olup olmadığını araştırınız?

Yıllar	Gelir(X)	Talep (Y)	Fiyat(Z)
1995	3	8	4
1996	2	9	7
1997	5	13	11
1998	6	22	18
Σ :	16	52	40

Yıllar	X	Y	Z	X²	Y²	Z²	XY	XZ	YZ
1995	3	8	4	9	64	16	24	12	32
1996	2	9	7	4	81	49	18	14	63
1997	5	13	11	25	169	121	65	55	143
1998	6	22	18	36	484	324	132	108	396
Toplam	16	52	40	74	798	510	239	189	634

$$r_{XY}=\frac{\sum XY-\frac{\left(\sum X\right)\left(\sum Y\right)}{n}}{\sqrt{\left(\sum X^2-\frac{\left(\sum X\right)^2}{n}\right)\left(\sum Y^2-\frac{\left(\sum Y\right)^2}{n}\right)}}=\frac{239-\frac{16\times 52}{4}}{\sqrt{\left(74-\frac{16^2}{4}\right)\left(798-\frac{52^2}{4}\right)}}=0.89$$

$$r_{XZ}=\frac{\sum XZ-\frac{\left(\sum X\right)\left(\sum Z\right)}{n}}{\sqrt{\left(\sum X^2-\frac{\left(\sum X\right)^2}{n}\right)\left(\sum Z^2-\frac{\left(\sum Z\right)^2}{n}\right)}}=\frac{189-\frac{16\times 40}{4}}{\sqrt{\left(74-\frac{16^2}{4}\right)\left(510-\frac{40^2}{4}\right)}}=0.87$$

$$r_{YZ} = \frac{\sum YZ - \frac{(\sum Y)(\sum Z)}{n}}{\sqrt{\left(\sum Y^2 - \frac{(\sum Y)^2}{n}\right)\left(\sum Z^2 - \frac{(\sum Z)^2}{n}\right)}} = \frac{634 - \frac{52 \times 40}{4}}{\sqrt{\left(758 - \frac{52^2}{4}\right)\left(510 - \frac{40^2}{4}\right)}} = 0.98$$

$$r_{XY,Z} = \frac{r_{XY} - r_{XZ} r_{YZ}}{\sqrt{(1 - r_{XZ}^2)(1 - r_{YZ}^2)}}$$

$$= \frac{0.89 - 0.98 \times 0.87}{\sqrt{(1 - 0.98^2)(1 - 0.87^2)}} = 0.4$$

Z(fiyat) sabitken x (gelir) ile y(talep) arasında ki ilişki aynı yönlü fakat düşük bir ilişki söz konusudur.

Örnek 2.6. **6 kişinin aylık satış (X), yaş (Y) ve ortalama not (Z) değerleri arasındaki ilişki araştırılıyor. Yaşı sabit alarak aylık satış ile ortalama not arasındaki kısmi korelasyonu bulup, bulunan değerin anlamlı olup olmadığını-doğrusal ilişki olup olmadığını araştırınız?**

Kişiler	Satış (X)	Yaş(Y)	Ort.not(Z)
Ali	8	18	80
Ayşe	3	20	72
Hasan	12	19	50
Ahmet	9	21	75
Mehmet	5	22	60
Fatma	7	18	62

$$r_{XZ,Y} = \frac{r_{XZ} - r_{XY} r_{ZY}}{\sqrt{(1 - r_{XY}^2)(1 - r_{ZY}^2)}}$$

	X	Y	Z	X²	Y²	Z²	XY	XZ	YZ
	8	18	80	64	324	6400	144	640	1440
	3	20	72	9	400	5184	60	216	1440
	12	19	50	144	361	2500	228	600	950
	9	21	75	81	441	5625	189	675	1575
	5	22	60	25	484	3600	110	300	1320
	7	18	62	49	324	3844	126	434	1116
Toplam	44	118	399	372	2334	27153	857	2865	7841

$$r_{XY} = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{n}\right)\left(\sum Y^2 - \frac{(\sum Y)^2}{n}\right)}} = \frac{857 - \frac{44 \times 118}{6}}{\sqrt{\left(372 - \frac{44^2}{6}\right)\left(2334 - \frac{118^2}{6}\right)}} = -0.32$$

$$r_{XZ} = \frac{\sum XZ - \frac{(\sum X)(\sum Z)}{n}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{n}\right)\left(\sum Z^2 - \frac{(\sum Z)^2}{n}\right)}} = \frac{2865 - \frac{44 \times 399}{6}}{\sqrt{\left(372 - \frac{44^2}{6}\right)\left(27153 - \frac{399^2}{6}\right)}} = -0.35$$

$$r_{YZ} = \frac{\sum YZ - \frac{(\sum Y)(\sum Z)}{n}}{\sqrt{\left(\sum Y^2 - \frac{(\sum Y)^2}{n}\right)\left(\sum Z^2 - \frac{(\sum Z)^2}{n}\right)}} = \frac{7841 - \frac{118 \times 399}{6}}{\sqrt{\left(2334 - \frac{118^2}{6}\right)\left(27153 - \frac{399^2}{6}\right)}} = -0.07$$

$$r_{XZ,Y} = \frac{r_{XZ} - r_{XY} r_{ZY}}{\sqrt{(1-r_{XY}^2)(1-r_{ZY}^2)}} = \frac{-0.35-(-0.32)\times(-0.07)}{\sqrt{[1-(-0.32^2)][1-(-0.07^2)]}} = -0.39$$

Soru : A sabitken B ile C arasındaki korelasyonu bulunuz?

	A	B	C
	2	3	4
	4	5	6
	5	6	10
	6	10	12
	8	12	15
Toplam :	25	36	47

$r_{BC,A}$=?

2.3.KORELASYON KATSAYILARININ FISHER Z PUANLARINA DÖNÜŞTÜRÜLMESİ

Sıralı ölçek verisi niteliğindeki veriler varsa, normalllik varsayımı geçerli olmaz. Fisher (1915) böyle durumlarda elde edilen değerleri Z puanlarına dönüştürerek, verilerin normal dağılım özelliğine sahip olmasını sağlamıştır. Normal dağılım özelliği gösteren Fisher Z değerleri, standart Z puanlarıyla karıştırılmaması gerekir. Her iki yöntemin de hesaplama formülleri farklıdır.

Korelasyon (güvenilirlik) katsayılarının Fisher Z puanlarına dönüştürülmesi işlemi test veya ölçek sonuçlarının daha sağlıklı karşılaştırılmasına imkan sağlar. Bu şekilde güvenilirlik katsayıları yansız biçimde değerlendirilmiş olur.

Korelasyon analizine dayalı güvenilirlik katsayıları Fisher Z puanlarına aşağıdaki amaçla için dönüştürülebilir:

1. Korelasyon katsayısının hipotez testi ile ana kütle için anlamlı olup olmadığını belirlemek.
2. Aynı ölçek veya teste ait farklı güvenilirlik katsayılarını karşılaştırmak.
3. Elde edilen güvenilirlik katsayısının (korelasyon) güven aralığını belirlemek.
4. Farklı güvenilirlik katsayılarını tek bir güvenilirlik katsayısı haline getirmek ve daha sonra bu güvenilirlik katsayısının güven aralığını belirlemek.
5. Anakütle etki büyüklüğünü doğru biçimde ölçmek.

İstatistiksel test uygulamalarından elde edilen korelasyon katsayısının;

a) Ne ölçüde yüksek güvenilirliğe sahip olduğunu
b) Ne ölçüde anakütleye genellenebileceğini
belirlemek için korelasyon (r) değeri Fisher Z puanlarına dönüştürülmelidir.

$$Z = 0.5\left[\ln(1+r) - \ln(1-r)\right] \quad , \quad Z = 0,5\ln\left[\frac{1+r}{1-r}\right]$$

Örnek 2.7. 50 soruluk bir ölçek aynı kitle üzerinde iki farklı zamanda uygulanmış ve test-yeniden test yöntemi sonucunda iki ölçümün korelasyon katsayısı 0,45 bulunmuştur. Bulunan korelasyon katsayının Z skor puanını bulunuz?

$$Z = 0.5\ln\left[\frac{1+r}{1-r}\right] = 0,5\ln\left[\frac{1+0.45}{1-0.45}\right] = 0.484$$

$$SS = \sigma_Z = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{50-3}} = 0.15$$

2.4. SPEARMAN SIRA KORELASYON KATSAYISI

Doğrusal korelasyonda ilişkisi araştırılan değişkenlerin nicel ve normal olması gerekir. Bu varsayımlar sağlanmadığında Spearman Sıra Korelasyon Katsayısı kullanılır.

Sıra korelasyon katsayısının hesaplanmasında önce gözlem değerleri büyükten küçüğe veya küçükten büyüğe doğru sıralanır ve bu sıralamaya göre sıra numarası verilir.

$$r_s = 1 - \frac{6\sum D_i^2}{n(n^2-1)}$$

D_i : X ve Y'nin sıra numaraları arasındaki fark
n : Gözlem sayısı

Eğer verilerde bağlı-aynı (ties) gözlemler varsa Spearman sıra korelasyonu aşağıdaki gibi hesaplanır. Bağlı gözlem sayısı az olduğunda Spearman, çok olduğunda ise Kendall Tau-b kullanılır.

$$r_s = \frac{\sum_{i=1}^{n}\left[R(x_i)-R(\bar{x})\right]\left[R(y_i)-R(\bar{y})\right]}{(n-1)\times S_{R(x)}\times S_{R(y)}}$$

Sıra Korelasyonunun Yorumu

0.90-1.00	Çok güçlü ilişki
0.70-0.89	Güçlü
0.50-0.69	Orta seviyede
0.30-0.4.9	Düşük
0.16 -0.29	Zayıf
<0.16	Çok zayıf

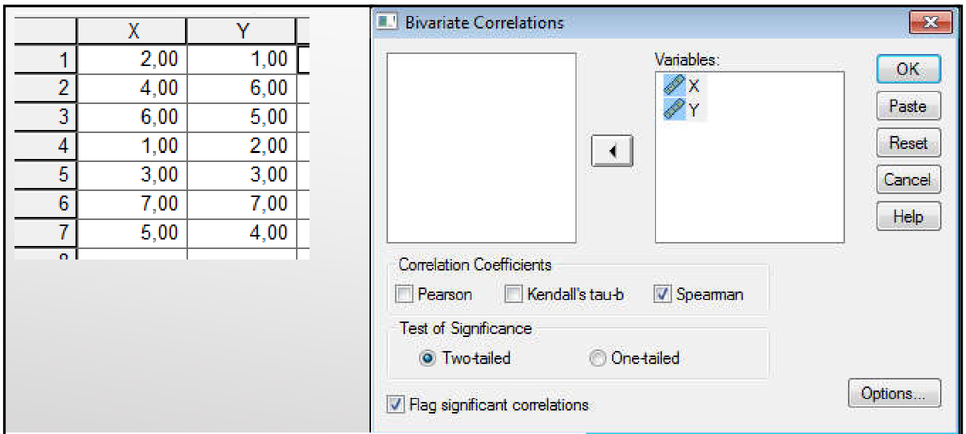
Örnek 2.8. 7 öğrencinin boy ve ağırlıklarının büyükten küçüğe doğru sıra puanları aşağıda gösterilmiştir. Veriler normal dağılım göstermediğine göre, boy ve ağırlıklar arasındaki ilişkiyi hesaplayınız?

X	Y	D	D^2
2	1	1	1
4	6	-2	4
6	5	1	1
1	2	-1	1
3	3	0	0
7	7	0	0
5	4	1	1

$$r_s = 1 - \frac{6\sum D_i^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 8}{7(7^2-1)} = 0.86$$

Öğrencilerin boy uzunlukları ile ağırlıkları arasında aynı yönde önemli bir ilişki vardır.



Correlations

			X	Y
Spearman's rho	X	Correlation Coefficient	1,000	,857*
		Sig. (2-tailed)	.	,014
		N	7	7
	Y	Correlation Coefficient	,857*	1,000
		Sig. (2-tailed)	,014	.
		N	7	7

*. Correlation is significant at the 0.05 level (2-tailed).

Örnek 2.9. Aşağıdaki verilere göre fiyat ile talep arasındaki sıra korelasyon katsayısını bulunuz?

Fiyat (X)	Talep(Y)	Rx	Ry
60	800	8	2,5
100	750	5.5	4
120	700	3	5
100	800	5.5	2,5
80	850	7	1
120	650	3	6.5
130	650	1	6.5
120	600	3	8

	x	y	Rx	Ry
1	60	800	8,0	2,5
2	100	750	5,5	4,0
3	120	700	3,0	5,0
4	100	800	5,5	2,5
5	80	850	7,0	1,0
6	120	650	3,0	6,5
7	130	650	1,0	6,5
8	120	600	3,0	8,0

Descriptive Statistics

	N	Mean	Std. Deviation
Rank of x	8	4,500	2,3755
Rank of y	8	4,500	2,4202
Valid N (listwise)	8		

$$r_s = \frac{\sum_{i=1}^{n}\left[R(x_i)-R(\bar{x})\right]\left[R(y_i)-R(\bar{y})\right]}{(n-1)\times S_{R(x)}\times S_{R(y)}}$$

$$= \frac{\sum\left[(8-4.5)(2.5-4.5)+\ldots+(3-4.5)(8-4.5)\right]}{(8-1)\times 2.375\times 2.42}$$

$$= \frac{-34.25}{40.233} = -0.851$$

SORU : **Aşağıda fiyat ve talep miktarları verilmiştir. Veriler normal dağılım göstermediğine göre, fiyat ile talep arasındaki ilişki miktarını bulunuz?**

FİYAT	TALEP
60	80
10	75
12	70
10	80
8	85
12	65
13	65
12	60

rs=-0,453

73

2.5. Kendall Tau-b

Kendall (1938) tarafından geliştirilen bu test, ikili ve sıralı ölçekli veriler arasındaki ilişkileri belirler. İki farklı gözlemcinin yaptığı değerlendirmeler puan büyüklüğü sırası içinde verilmişse veya hakem puanları büyüklük sırasına sokmuşsa Spearman korelasyonun yanında, Kendall tau b analizi de yapılır. Spearman rho büyüklük sırasına sokulmuş değerler arasındaki Pearson momentler çarpımını temsil ederken, Kendall tau olasılığı temsil eder. Yani iki değişkende aynı sırada yer alan(uyuşan) verilerin gözlenme olasılığı ile farklı sırada yer alan (uyuşmayan) verilerin gözlenme olasılığı arasındaki farktır. İki serideki veriler uyuşuyorsa tau-b +1 değerini alır. Uyuşma yoksa tau-b -1 olur.

Kendall tau-b genelde kare tablolar için, Kendall-tau-c ise dikdörtgen tablolar için kullanılır. Tau-a ise uyuşan ve uyuşmayan çiftler arasındaki farkın toplam çift sayısına bölünmesiyle bulunur.

Kendall tau_b bağlı (ties) sıralığa sahip (aynı gözlemli olan değerlerin sıra puanlarının ortalaması alınır) ölçek verileri için uygundur. Bu analiz verilerin normal dağılış göstermediği ve n<20 olduğu durumlarda daha iyi sonuç vermektedir.

Kendall tau katsayıları aşağıdaki gibi yorumlanır:

>0.5	**Yüksek ilişki**
0.36-0.49	**Önemli ilişki**
0.20-0.35	**Orta derecede ilişki**
0.10-0.19	**Düşük ilişki**
<0.1	**İlişki yok**

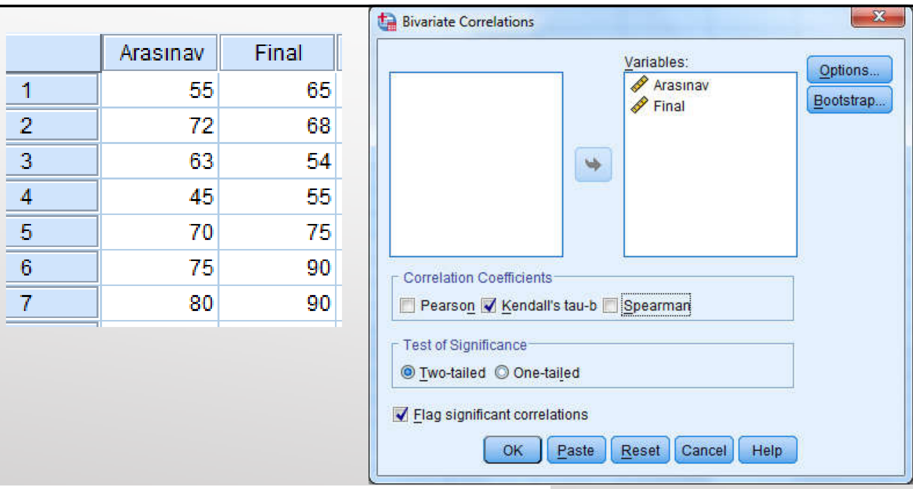
- İki seri verildiğinde birinci seri doğal sırada olacak şekilde (küçükten-büyüğe) iki seri yeniden dizilir.
-İkinci serideki her bir Yi değerine bakarak bu değeri izleyen değerlerden kaç tanesi Yi'den büyük (ai) ve kaç tanesi Yi'den küçük (bi) olduğu sayılır. Bu işlem her satır için yapılarak ai ve bi serileri oluşturulur.
- Na=Σai ve Nb= Σbi değerleri el de edilir.

$$\tau = \frac{N_a - N_b}{N(N-1)/2}$$

Örnek 2.10. 7 öğrencinin bir dersin arasınav ve final sınavı notları aşağıdaki gibidir. Verilerin normal dağılış göstermediği varsayılarak iki sınav notu arasındaki korelasyon katsayısını

Arasınav	Final	Xi	Yi	ai	bi
55	65	45	55	5	1
72	68	55	65	4	1
63	54	63	54	4	0
45	55	70	75	2	1
70	75	72	68	2	0
75	90	75	90	0.5	0.5
80	90	80	90	0	0
				17.5	3.5

$$\tau = \frac{N_a - N_b}{N(N-1)/2} = \frac{17.5 - 3.5}{7(7-1)/2} = 0.67$$



Correlations

			Arasınav	Final
Kendall's tau_b	Arasınav	Correlation Coefficient	1,000	,683*
		Sig. (2-tailed)	.	,033
		N	7	7
	Final	Correlation Coefficient	,683*	1,000
		Sig. (2-tailed)	,033	.
		N	7	7

*. Correlation is significant at the 0.05 level (2-tailed).

3. BASİT DOĞRUSAL REGRESYON ANALİZİ

Regresyon analizi yapılırken, gözlem değerlerinin ve etkilenilen olayların bir matematiksel gösterimle yani bir fonksiyon yardımıyla ifadesi gerekmektedir. Kurulan bu modele regresyon modeli denilmektedir. Regresyon analizinde bir bağımlı değişken ve birde ona etki eden bağımsız değişken veya değişkenlerden oluşan bir model sözkonusudur.

Regresyon analizi araştırma, matematik, finans, ekonomi, tıp gibi bilim alanlarında yoğun olarak kullanılmaktadır. Regresyon analizinin temelinde; gözlenen bir olayın değerlendirilirken, hangi olayların etkisi içinde olduğunun araştırılması yatmaktadır. Bu olaylar bir veya birden çok olacağı gibi dolaylı veya direkt etkileniyor da olabilirler.

Değişken, belirli bir zaman aralığı gözönüne alınıp, o zaman aralığında bir kütleyi oluşturan belli birimdeki olayları içeren örneklerdir. Sayılabilir veya ölçülebilir nitelikte olmalıdır.

Bir hissenin fiyatını bir değişken alırsak, ona dolaylı olarak veya direkt etkili bir veya birden çok değişken alabiliriz (Örneğin: Faiz oranları, enflasyon, ekonomik, politik, finansal olaylar vs.). Sadece faiz oranlarının etkisi ile ilgileniyorsak, tek değişkenli bir matematiksel model, faiz oranları ile birlikte enflasyon oranı ile de ilgileniyorsak, iki değişkenli bir matematiksel modelden söz ediyoruzdur. Faiz oranları hisse senedinin fiyatını direkt etkileyen bir unsur olmadığı halde faiz oranlarının yükseldiği durumda hisse senedinin fiyatının düşüyor olmasının gözlemlenmesi bir etkileşim olduğunun göstergesidir.

Değişkenler arasındaki ilişkilerin fonksiyonel şekillerini ararken, neden durumunda olan değişkenlere **bağımsız değişken** (independent variable), sonuç durumunda olan değişkenlere ise **bağımlı değişken** (dependent variable) adı verilir.

Örneğin satışlar ve reklam harcamaları çalışmasında, firmanın reklam harcamaları bağımsız, satışlar ise bağımlı değişkendir. Yine üretim miktarı ile maliyet arasındaki bir çalışmada üretim miktarı bağımsız değişken, maliyet ise bağımlı değişkendir. İşçilere verilen prim ile verimlilik arasındaki bir modelde ise pirimler bağımsız değişken, verimlilik ise bağımlı değişkendir.

Reklam harcamaları, üretilen miktar, pirimler gibi değişkenler daha kontrol edilebilen değişkenler olduğundan, bu değişkenler bağımsız değişkenlerdir.

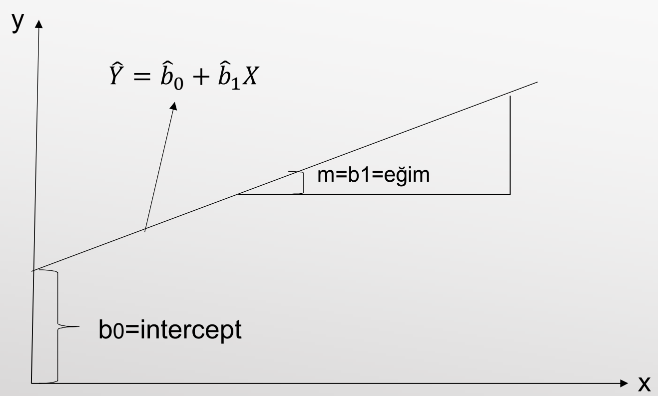
Bağımsız değişkenler kurulacak modelde bir değişkenli olarak ele alınırsa, **basit doğrusal regresyon**, birden fazla bağımsız değişkenli olarak alınırsa, **çoklu doğrusal regresyon modeli** konusunu oluşturmaktadır.

Örneğin satışlar bağımlı, reklam harcamaları ve satış personeline verilen pirim bağımsız değişken olabilir. Bir tarladan alınan verim bağımlı değişken, gübre ve yağış miktarı da bağımsız değişken olarak alınabilir.

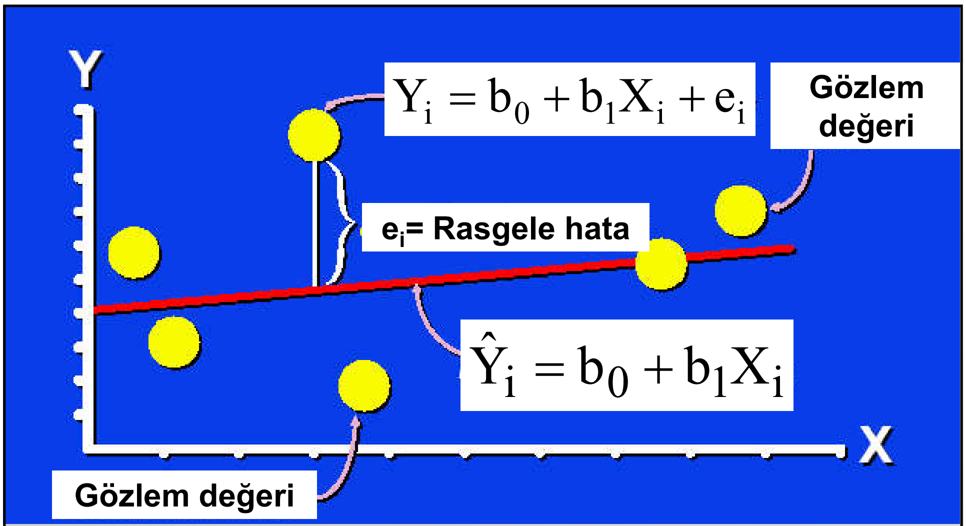
Basit Doğrusal Regresyon Modeli	$Y = b_0 + b_1X + e_i$
Çoklu Regresyon Modeli	$Y = b_0 + b_1X_1 + b_2X_2 + ... + e_i$
Y	**Bağımlı değişken**
$X_1 , X_2 , X_3 , ...$	**Bağımsız değişkenler**
$b_0,b_1, b_2, \dots$	**Sabit katsayılar**
e_i	**Hata terimi**

b_0:Regresyon doğrusunun y eksenini kestiği nokta

b_1: Regresyon doğrusunun eğimi yani X_i değişkeni bir birim değiştiği takdirde Y_i'nin ne kadar değişeceğini gösteren değer.



Eğim=m=b_1 : Doğrunun X ekseni ile pozitif yönde yaptığı açının tanjantı



3.1.Doğrusal Regresyon Modelinin Varsayımları

- Hata terimi (e_i) **rassal bir değişkendir**. X'in her bir değeri için hata terimi pozitif, negatif ve sıfır değerlerini belli olasılıklarla alabilir. Bağımlı değişkendeki ölçme hatalarından, model dışında bırakılan değişkenlerden kaynaklanan hatalardan dolayı hata terimine regresyon modelinde yer verilir.

- $E(e_i)=0$ yani hata teriminin ortalaması sıfırdır. Her bir X değeri için hata terimlerinin ortalaması sıfır olur.

- Hata teriminin varyansı X değerlerine göre değişmez yani **sabittir**. Yani tüm X değerleri için e hata terimleri kendi ortalamaları etrafında aynı değişkenliğe sahiptir. Ayrıca hata teriminin varyansı, bağımlı değişkenin varyansına eşittir.

$$V(Y)=V(e)$$

Örneğin X=10 olduğunda ortaya çıkabilecek Y'nin varyansı ile X=20 olduğunda oluşacak Y'nin varyansı aynı olacaktır.

- Hata terimi (e_i) normal dağılıma sahiptir.

  $e \sim N(0, \sigma_e^2)$

- Hata terimlerin ardışık (birbirini izleyen) değerleri birbirinden bağımsızdır. $i \neq j$ olmak üzere e_i ve e_j 'nin kovaryansı sıfıra eşittir.

  $Kov(e_i\, e_j)=0$

- Bağımsız değişkeler arasında <u>doğrusal bağımlılık</u> (ilişki) yoktur. Doğrusal bağımlılık yüksek ise bağımsız değişkenlerden her birinin etkisini diğerlerinden ayırmak zorlaşır.

3.2. Regresyon Katsayılarının (b_0 ve b_1) Tahmini

$$Y_1 = b_0 + b_1 X_1$$
$$Y_2 = b_0 + b_1 X_2$$
$$\vdots$$
$$Y_n = b_0 + b_1 X_n$$

$$\sum Y_i = n b_0 + b_1 \sum X_i$$

$$\frac{\sum Y_i}{n} = \frac{n b_0}{n} + \frac{b_1 \sum X_i}{n} \Rightarrow \overline{Y} = b_0 + b_1 \overline{X}$$

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

$$X_1Y_1 = b_0X_1 + b_1X_1^2$$
$$X_2Y_2 = b_0X_2 + b_1X_2^2$$
$$\vdots$$
$$X_nY_n = b_0X_n + b_1X_n^2$$

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

b_0 yerine değeri yazılır.

$$\hat{b}_1 = \frac{\sum X_iY_i - \frac{(\sum X_i)(\sum Y_i)}{n}}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}$$

Örnek 3.1. Aşağıda verilen verilere göre regresyon denklemini kurunuz?

Yıllar	Y_i	X_i	Y_i^2	X_i^2	$X_i Y_i$
2002	2	7	4	49	14
2003	3	9	9	81	27
2004	1	14	1	196	14
2005	6	10	36	100	60
Toplam	12	40	50	426	115

$$\overline{X} = \frac{\sum X_i}{n} = \frac{40}{4} = 10$$
$$\overline{Y} = \frac{\sum Y_i}{n} = \frac{12}{4} = 3$$

$$\sum x_i^2 = \sum X_i^2 - n\overline{X}^2 = 426 - 4\times 10^2 = 26$$
$$\sum y_i^2 = \sum Y_i^2 - n\overline{Y}^2 = 50 - 4\times 3^2 = 14$$
$$\sum x_iy_i = \sum X_iY_i - n\overline{X}\,\overline{Y} = 115 - 4\times 10\times 3 = -5$$

$$\hat{b}_1 = \frac{\sum X_i Y_i - \frac{(\sum X_i)(\sum Y_i)}{n}}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}} = \frac{-115 - \frac{40 \times 12}{4}}{426 - \frac{40^2}{4}} = -0.192$$

$$\hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X} \quad = 3 - (-0.192) \times 10 = 4.923$$

$$\hat{Y} = \hat{b}_0 + \hat{b}_1 X = 4.923 - 0.193X$$

2006 yılında X=11 olması bekleniyor ise Y değeri yukarıdaki denklemden tahmin edilebilir.

$$\hat{Y}_{2006} = 4.923 - 0.193 \times 11 = 2.81$$

Örnek 3.2. **Aşağıda öğrencilerin boy ve ağırlıkları verilmiştir. Buna göre regresyon denklemini kurunuz?**

Öğrenci	Ağırlık Y_i	Boy X_i	Y_i^2	X_i^2	$X_i Y_i$
1	65	170	4225	28900	11050
2	70	180	4900	32400	12600
3	50	155	2500	24025	7750
4	50	165	2500	27225	8250
5	80	190	6400	36100	15200
6	69	178	4761	31684	12282
Toplam	384	1038	25286	180334	67132

$$\overline{X}=\frac{\sum X_i}{n}=\frac{1038}{6}=173 \quad ; \quad \overline{Y}=\frac{\sum Y_i}{n}=\frac{384}{6}=64$$

$$\hat{b}_1=\frac{\sum X_iY_i-\frac{(\sum X_i)(\sum Y_i)}{n}}{\sum X_i^2-\frac{(\sum X_i)^2}{n}}=\frac{67132-\frac{1030\times 384}{6}}{180334-\frac{1038^2}{6}}=0.92$$

$$\hat{b}_0=\bar{Y}-\hat{b}_1\bar{X}=64-(0.92)\times 173=-95.34$$

$$\hat{Y}i=\hat{b}_0+\hat{b}_1X_i=-95{,}34+0{,}92X_i$$

160 cm boyundaki bir öğrencinin ağırlığını tahmin edelim:

$$\hat{Y}_{160}=-95{,}34+0{,}92*160=51{,}86\ kg$$

3.3. b_1 (Eğim) ile r (korelasyon) Arasındaki İlişki

- **r>0 ise, b_1>0 dır. X ile Y arasındaki birlikte değişme aynı yönde ise; r ve b_1'nin işareti pozitiftir.**
- **r<0 ise, b_1<0 dır.**
- **r=0 ise, b_1=0 dır.**
- **r=±1 ise, b_1= ±1 dır.**
- **b_1>1 ise; bağımsız değişkende gözlenecek bir birimlik değişmeye karşılık bağımlı değişkende gözlenecek değişimin miktarı bir birimlikten fazladır.**
- **b_1<1 ise; bağımsız değişkende gözlenecek bir birimlik değişmeye karşılık bağımlı değişkende gözlenecek değişimin miktarı bir birimlikten azdır.**

Eğim($m=b_1$)

Eğim, doğrunun X ekseni ile pozitif yönde yaptığı açının tanjantıdır.

Regresyon doğrusu üzerinde (X_1,Y_1) ve (X_2,Y_2) gibi herhangi iki nokta verildiğinde b_0 ve b_1 sabitleri belirlenerek, doğru denklemi elde edilebilir.

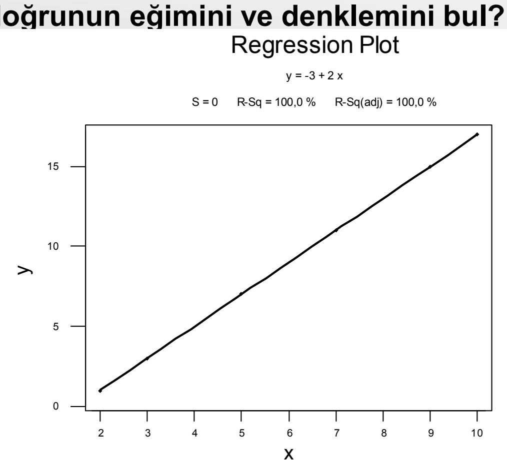
$$Y - Y_1 = \frac{Y_2 - Y_1}{X_2 - X_1}(X - X_1) \Rightarrow Y - Y_1 = m(X - X_1)$$

Burada m eğim olup, Y'deki değişmenin X'deki değişmeye bölünmesiyle elde edilir.

Örnek 3.3.

X:	**2**	**3**	**5**	**7**	**9**	**10**
Y:	**1**	**3**	**7**	**11**	**15**	**17**

a) Doğrunun grafiğini b) (2,1) ve (3,3) noktalarından geçen doğrunun eğimini ve denklemini bul?



$$Y - Y_1 = \frac{Y_2 - Y_1}{X_2 - X_1}(X - X_1)$$

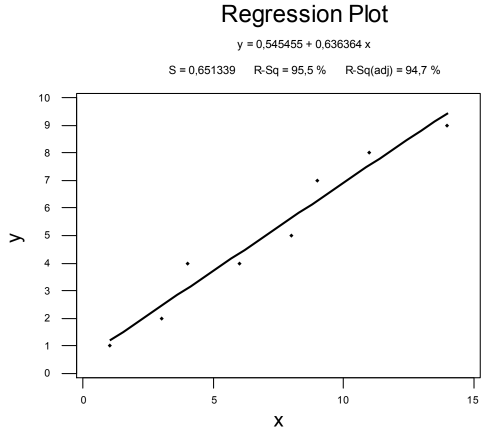
$$Y - 1 = \frac{3-1}{3-2}(X - 2)$$

$$Y = -3 + 2X$$

Örnek 3.4.

X:	1	3	4	6	8	9	11	14
Y:	1	2	4	4	5	7	8	9

a) Doğrunun grafiğini b) (0,1) ve (12,7.5) noktalarından geçen doğrunun eğimini ve denklemini bul?



$$Y - Y_1 = \frac{Y_2 - Y_1}{X_2 - X_1}(X - X_1)$$

$$Y - 1 = \frac{7.5 - 1}{12 - 0}(X - 0)$$

$$Y = 1 + 0.54X$$

3.4. EN KÜÇÜK KARELER (EKK) METODU

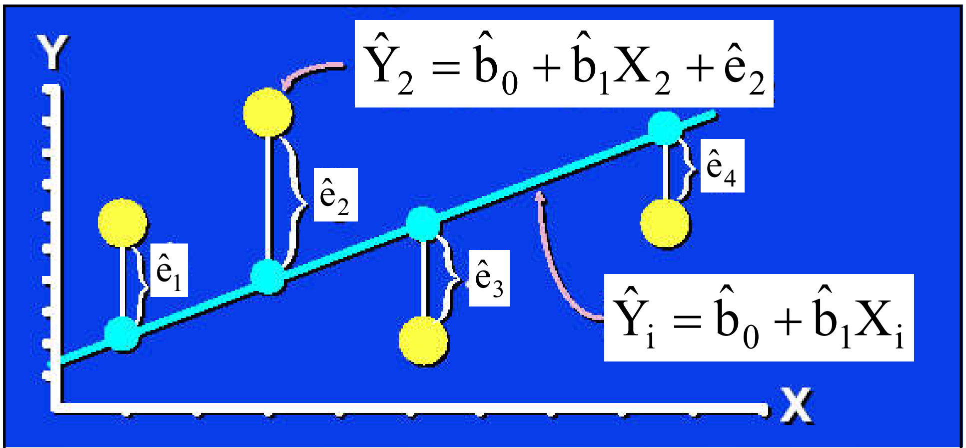
Bu metodun esası gerçek değerlerin regresyon doğrusundan (eğrisinden) uzaklaşmalarını minimum yapan denklemin bulunmasıdır. Gerçek değerler Y_i , tahmin değerleri $\hat{Y}_i$ ve bu değerler arasındaki fark hata terimini (e_i) oluşturur.

$$\sum e_i = \sum (Y_i - \hat{Y}_i) = 0$$

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

(X_1,Y_1), (X_2,Y_2),..., (X_n,Y_n) noktaları arasından sonsuz sayıda doğru geçebilir. Her doğru için Y_i değerleri ile $\hat{Y}_i$ değerleri arasında değişik farklar çıkacaktır. Bunlar içerisinde herhangi bir doğru için, <u>bu farkların kareler toplamı minimum ise o doğru, dağılımı en iyi temsil eden doğrudur.</u>

$$\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_i$$



$$\sum_{i=1}^{n} \hat{e}_i^2 = \hat{e}_1^2 + \hat{e}_2^2 + \hat{e}_3^2 + \hat{e}_4^2$$

$$\sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i\right)^2 = \sum_{i=1}^{n} \left(Y_i - b_0 - b_1 X_i\right)^2$$

ifadesinin b_0 ve b_1 'e göre türevleri alınıp sıfıra eşitlenerek:

$$\frac{\partial e}{\partial b_0} = -2\sum \left(Y_i - b_0 - b_1 X_i\right) = 0$$

$$\frac{\partial e}{\partial b_1} = -2\sum X_i\left(Y_i - b_0 - b_1 X_i\right) = 0$$

bulunur, elde edilen bu iki eşitlikten,

$$\left.\begin{aligned} \sum Y_i &= b_0 n + b_1 \sum X_i \\ \sum X_i Y_i &= b_0 \sum X_i + b_1 \sum X_i^2 \end{aligned}\right\} \textbf{Normal denklemler}$$

Örnek 3.5.

Reklam harcaması (X) (Milyon TL)	Satışlar (Y) (Adet)
10	3
20	4
30	6
40	7
50	10

$$\sum Y = b_0 n + b_1 \sum X$$

$$\sum XY = b_0 \sum X + b_1 \sum X^2$$

'dir. Bu durumda, X, Y, X^2 ve XY değerlerinin hesaplanması gerekir.

X	Y	XY	X^2
10	3	30	100
20	4	80	400
30	6	180	900
40	7	280	1600
50	10	500	2500
$\sum X = 150$	$\sum Y = 30$	$\sum XY = 1070$	$\sum X^2 = 5500$

$$30 = 5b_0 + 150b_1$$
$$1070 = 150b_0 + 5500b_1$$

Bu iki denklemin b_0 ve b_1 katsayıları;

$b_0 = 0.9$; $b_1 = 0.17$

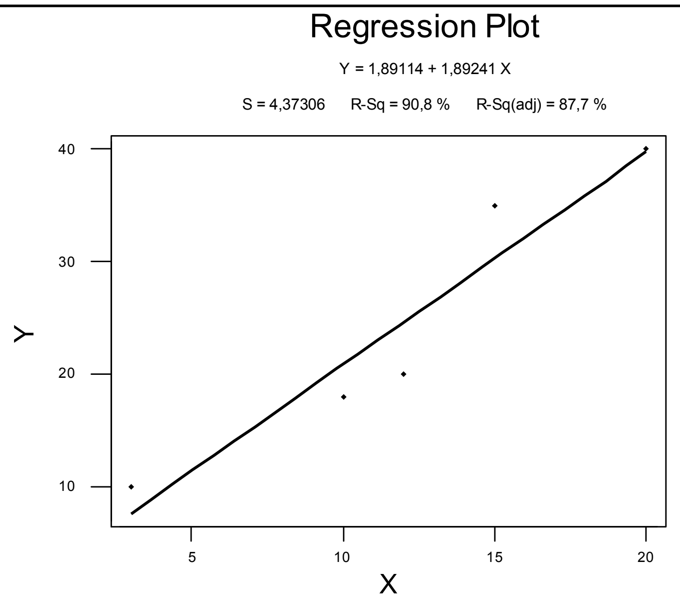
SORU. Aşağıdaki verilere göre basit doğrusal regresyon modelini tahmin ediniz.

X_i	$\mathbf{Y}X_i$	$X^2_{\mathbf{16}i}$	$\boldsymbol{Y}_i^2$	$X_i * \boldsymbol{Y}_i$
4	3	16	9	12
6	5,5	36	30,25	33
10	6,5	100	42,25	65
12	9	144	81	108

$\hat{Y} = \hat{b}_0 + \hat{b}_1 X_i = 0{,}80 + 0{,}65 X_i$

Örnek 3.6. Aşağıda bir ürünün talep ve fiyat miktarları verilmiştir. Buna göre regresyon denklemini kurunuz?

	Talep Y_i	Fiyat X_i	Y_i^2	X_i^2	$X_i Y_i$
	10	3	100	9	30
	18	10	324	100	180
	20	12	400	144	240
	35	15	1225	225	525
	40	20	1600	400	800
Toplam	123	60	3649	878	1775



3.5. TAHMİNİN STANDART HATASI

Regresyon denklemi yardımıyla tahmin edilen bağımlı değişken değerleri kesin olmayıp, birer tahmin değerleridir. Bu nedenle her tahminin bir hatası (sapması) olabilir. İşte bu hataya tahminin standart hatası denir.

$$S_{YX} = \sqrt{\frac{\sum (Y_i - \hat{Y}_i)^2}{n - m}}$$

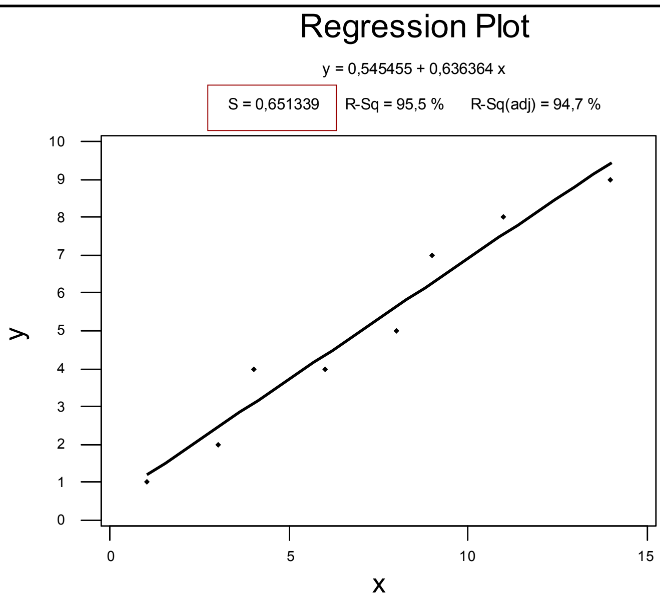
Y_i : Gözlenen değerler,
$\hat{Y}_i$: Tahmin edilen değerler
n : Gözlem sayısı
m : parametre sayısı (b0, b1)

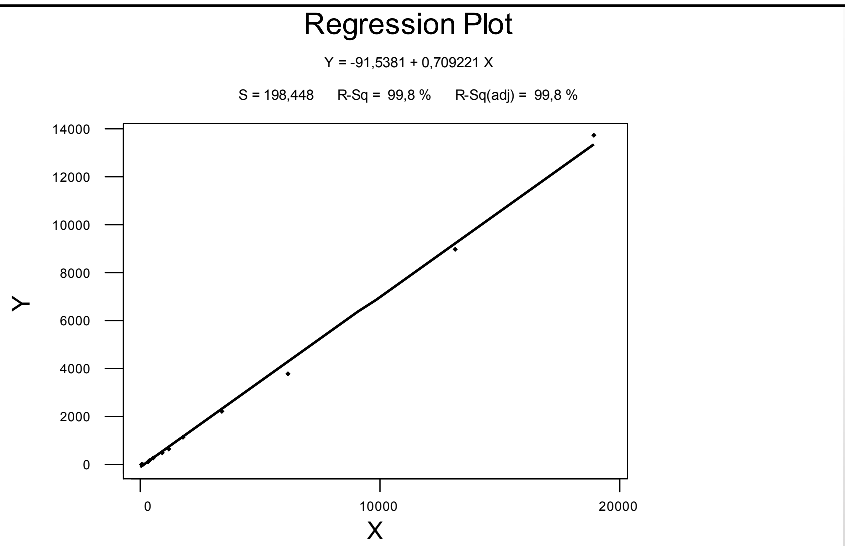
Örnek 3.7.

X	Y	$\hat{Y}_i$	$Y_i - \hat{Y}_i$	$(Y_i - \hat{Y}_i)^2$
1	1	1.18	-0.18	0.0324
3	2	2.46	-0.46	0.2116
4	4	3.1	0.90	0.8100
6	4	4.38	-0.38	0.1444
8	5	5.66	-0.66	0.4356
9	7	6.3	0.70	0.4900
11	8	7.58	0.42	0.1764
14	9	9.5	-0.50	0.2500
			TOPLAM	2.5510

$$\hat{Y}_i = b_0 + b_1 X_i = 0{,}54 + 0{,}64 X_i$$

$$S_{YX} = \sqrt{\frac{\sum (Y_i - \hat{Y}_i)^2}{n-m}} = \sqrt{\frac{2.551}{8-2}} = 0.65$$





$$S_{YX} = \sqrt{\frac{\sum (Y_i - \hat{Y}_i)^2}{n-m}} = \sqrt{\frac{768395.3}{10-2}} = 309.9$$

Tahminin standart hatası regresyon denklemi yardımıyla yapılacak tahminlerin doğruluğunu gösterir. Hata sıfıra ne kadar yakınsa, yapılacak tahminler o kadar güvenilir olur.

$$\hat{Y}_i \pm 1S_{YX} \quad , \quad \%68.27$$

$$\hat{Y}_i \pm 2S_{YX} \quad , \quad \%95.45$$

$$\hat{Y}_i \pm 3S_{YX} \quad , \quad \%99.73$$

$$\hat{Y}_i = -91.54 + 0.71 \times 1000 = 618.46$$

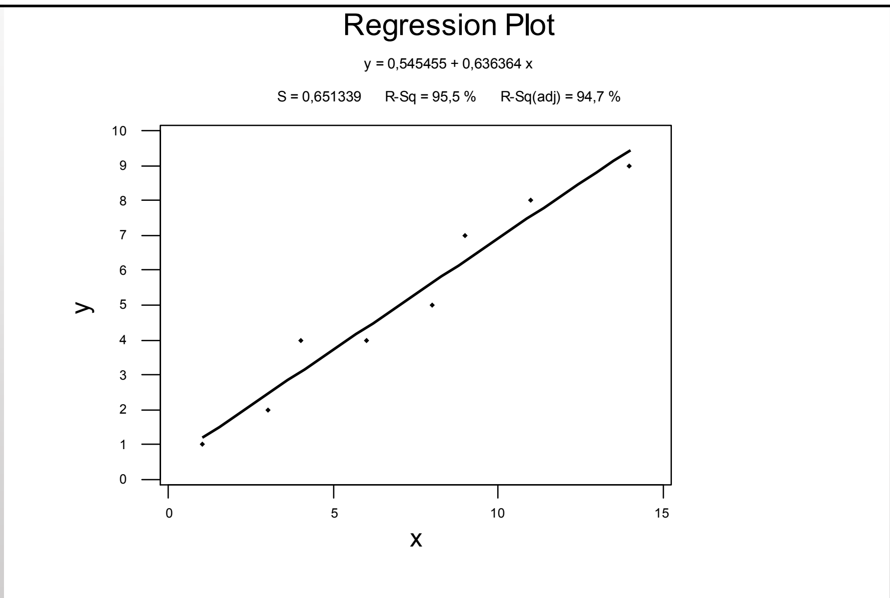
$$\hat{Y}_i \pm 2S_{YX} \quad , \quad \%95.45$$

$$618.46 \pm 2 \times 309.9 \Rightarrow (-1.43 \ ; 1238.26)$$

Örnek 3.8.

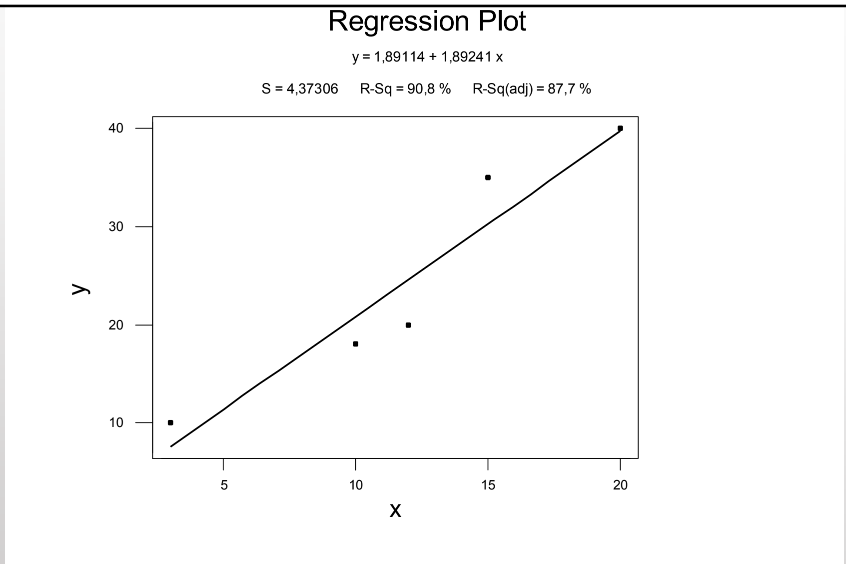
X	Y	$\hat{Y}_i$	$Y_i - \hat{Y}_i$	$(Y_i - \hat{Y}_i)^2$
1	1	1,18	-0,18	0,0324
3	2	2,46	-0,46	0,2116
4	4	3,1	0,9	0,81
6	4	4,38	-0,38	0,1444
8	5	5,66	-0,66	0,4356
9	7	6,3	0,7	0,49
11	8	7,58	0,42	0,1764
14	9	9,5	-0,5	0,25
			TOPLAM	2,551

$$\hat{Y}_i = b_0 + b_1 X_i$$
$$= 0.54 + 0.64 X_i$$



Örnek 3.9. Aşağıda bir ürünün talep ve fiyat miktarları verilmiştir. Buna göre regresyon denklemini,katsayıların anlamlığını ve güven aralıklarını bulunuz?

	Talep Y_i	Fiyat X_i	Y_i^2	X_i^2	$X_i Y_i$
	10	3	100	9	30
	18	10	324	100	180
	20	12	400	144	240
	35	15	1225	225	525
	40	20	1600	400	800
Toplam	123	60	3649	878	1775



3.6. DETERMİNASYON KATSAYISI-BELİRTME KATSAYISI (r^2)

Bağımlı değişkenin (Y_i) bağımsız değişkene (X_i) hangi oranda bağlı olduğunu gösteren katsayıya determinasyon(belirlilik) katsayısı denir. Korelasyon katsayısının karesine de yine determinasyon katsayısı adı verilir.

Determinasyon katsayısı [0,1] arasındadır.

Doğrusal korelasyonda bağımlı değişken, bağımsız değişken dışında bazı değişkenlere de bağlı bulunabilir. Bu değişkenlerin tamamına **diğer etkenler (k^2)** denir.

$r^2+k^2=1$

$$r^2 = \frac{\text{Açıklanan değişim}}{\text{Toplam değişim}}$$

$$= \frac{\sum_{i=1}^{n}(Y_i-\bar{Y})^2 - \sum_{i=1}^{n}(Y_i-\hat{Y})^2}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}$$

$$r^2 = \left(\frac{Cov(X,Y)}{S_x\, S_y}\right)^2$$

PROBLEMLER

SORU 1. 10 öğrencinin haftalık çalışma saati ile akademik ortalama değerleri aşağıda verilmiştir.

Bu değerlerden yararlanarak çalışma saati ile akademik ortalama arasındaki doğrusal ilişkinin matematiksel modelini kurunuz?

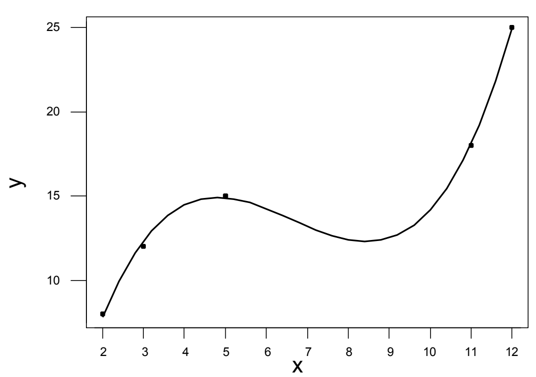
Çalışma Saati (hafta)	:15	8	9	18	14	20	7	2	12	10
Akademik Ortalama	:2.5	1.5	1.8	3	2.9	3.6	1.5	1	1.8	2

SORU 2. Buğdayın verimi ile protein yönünden zenginliği arasında doğrusal bir ilişki bulunduğu ileri sürülmektedir. Bir tarladan elde edilen buğday ağırlıkları ile protein miktarları aşağıda verilmiştir. %5 anlamlılık seviyesine göre;
Bu değerlerden yararlanarak buğday ağırlığı ile protein miktarı arasındaki doğrusal ilişkinin matematiksel modelini kurunuz?

Buğday Ağırlığı (gr)	:150	400	300	300	350
Protein Miktarı (%)	:16	13	15	13	13

EĞRİSEL REGRESYON

Bazen bağımlı değişken ile bağımsız değişken arasındaki ilişki doğrusal olmayabilir. Çeşitli eğrisel modeller vardır.



Eğrisel Modeller

(1) Linear	$E(Y_t) = \beta_0 + \beta_1 t$
(2) Logarithmic	$E(Y_t) = \beta_0 + \beta_1 \ln(t)$
(3) Inverse	$E(Y_t) = \beta_0 + \beta_1 / t$
(4) Quadratic	$E(Y_t) = \beta_0 + \beta_1 t + \beta_2 t^2$
(5) Cubic	$E(Y_t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3$
(6) Compound	$E(Y_t) = \beta_0 \beta_1^t$
(7) Power	$E(Y_t) = \beta_0 t^{\beta_1}$
(8) S	$E(Y_t) = \exp(\beta_0 + \beta_1 / t)$
(9) Growth	$E(Y_t) = \exp(\beta_0 + \beta_1 t)$
(10) Exponential	$E(Y_t) = \beta_0 e^{\beta_1 t}$
(11) Logistic	$E(Y_t) = \left(\frac{1}{u} + \beta_0 \beta_1^t\right)^{-1}$

Model	Dependent Variable	Independent Variables	Coefficients
(1)	Y	t	β_0, β_1
(2)	Y	$\ln(t)$	β_0, β_1
(3)	Y	$1/t$	β_0, β_1
(4)	Y	t, t^2	$\beta_0, \beta_1, \beta_2$
(5)	Y	t, t^2, t^3	$\beta_0, \beta_1, \beta_2, \beta_3$
(6)	$\ln(Y)$	t	β_0^*, β_1^*
(7)	$\ln(Y)$	$\ln(t)$	β_0^*, β_1
(8)	$\ln(Y)$	$1/t$	β_0, β_1
(9)	$\ln(Y)$	t	β_0, β_1
(10)	$\ln(Y)$	t	β_0^*, β_1
(11)	$\ln\left(\frac{1}{Y} - \frac{1}{u}\right)$	t	β_0^*, β_1^*

where $\beta_0^* = \ln(\beta_0)$ and $\beta_1^* = \ln(\beta_1)$.

(1) $\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 t$

(2) $\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 \ln(t)$

(3) $\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 / t$

(4) $\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 t^2$

(5) $\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 t^2 + \hat{\beta}_3 t^3$

(6) $\hat{Y}^*_t = \hat{\beta}^*_0 + \hat{\beta}^*_1 t$

(7) $\hat{Y}^*_t = \hat{\beta}^*_0 + \hat{\beta}_1 \ln(t)$

(8) $\hat{Y}^*_t = \hat{\beta}_0 + \hat{\beta}_1 / t$

(9) $\hat{Y}^*_t = \hat{\beta}_0 + \hat{\beta}_1 t$

(10) $\hat{Y}^*_t = \hat{\beta}^*_0 + \hat{\beta}_1 t$

(11) $\hat{Y}^*_t = \hat{\beta}^*_0 + \hat{\beta}^*_1 t$

1. PARABOL -KARELİ EĞRİ: $Y = b_0 + b_1 X + b_2 X^2$

2. HİPERBOL EĞRİ: $Y = 1/(b_0 + b_1 X)$

3. GEOMETRİK EĞRİ (Power): $Y = b_0 X^{b_1}$

4. ÜSTEL EĞRİ (Compound): $Y = b_0 b_1^{X}$

5. KÜBİK EĞRİ: $Y = b_0 + b_1 X + b_2 X^2 + b_3 X^3$

1. Kuadratik (Kareli) Regresyon Modeli

$$\hat{Y} = b_0 + b_1 X + b_2 X^2$$

$$\sum Y = nb_0 + b_1 \sum X + b_2 \sum X^2$$

$$\sum XY = b_0 \sum X + b_1 \sum X^2 + b_2 \sum X^3$$

$$\sum X^2 Y = b_0 \sum X^2 + b_1 \sum X^3 + b_2 \sum X^4$$

Denklemlerinden b_0, b_1 ve b_2 parametreleri bulunur.

Örnek :

	Y	X	X^2	X^3	X^4	XY	X^2Y
	99	-3	9	-27	81	-297	891
	96	-2	4	-8	16	-192	384
	91	-1	1	-1	1	-91	91
	84	0	0	0	0	0	0
	75	1	1	1	1	75	75
	64	2	4	8	16	128	256
	51	3	9	27	81	153	459
TOP	560	0	28	0	196	-224	2156

$$\sum Y = nb_0 + b_1 \sum X + b_2 \sum X^2$$

$$\sum XY = b_0 \sum X + b_1 \sum X^2 + b_2 \sum X^3$$

$$\sum X^2 Y = b_0 \sum X^2 + b_1 \sum X^3 + b_2 \sum X^4$$

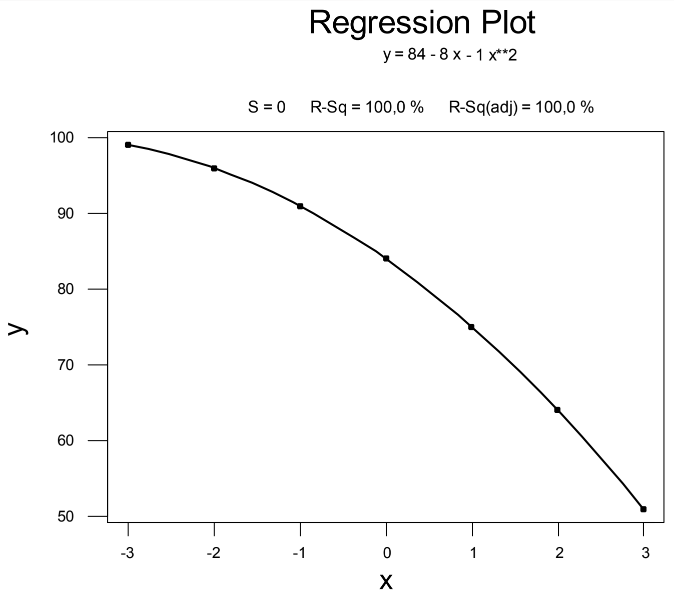
$560 = 7b_0 + 0b_1 + 28b_2$

$-224 = 0b_0 + 28b_1 + 0b_2$

$2156 = 28b_0 + 0b_1 + 196b_2$

$b_0 = 84$, $b_1 = -8$, $b_2 = -1$

$$\hat{Y} = 84 - 8X - X^2$$



Örnek :

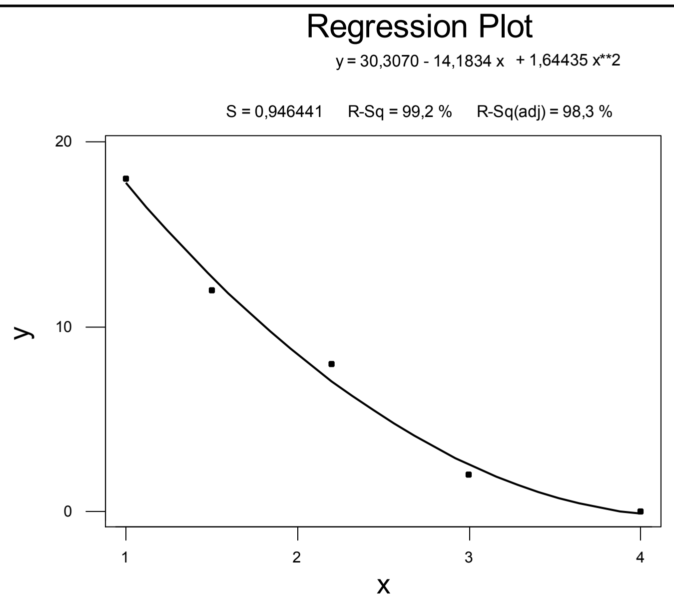
	Y	X	X^2	X^3	X^4	XY	X^2Y
	18	1,0	1,00	1,00	1,00	18,0	18,00
	12	1,5	1,25	3,37	5,06	18,0	27,00
	8	2,2	4,84	10,65	23,42	17,6	38,72
	2	3,0	9,00	27,00	81,00	6,0	18,00
	0	4,0	16,00	64,00	256,00	0,0	0,00
TOP	40	11,7	33,09	106,02	366,48	59,6	101,72

$$\sum Y = nb_0 + b_1\sum X + b_2\sum X^2$$
$$\sum XY = b_0\sum X + b_1\sum X^2 + b_2\sum X^3$$
$$\sum X^2Y = b_0\sum X^2 + b_1\sum X^3 + b_2\sum X^4$$

$40 = 5b_0 + 11{,}7b_1 + 33{,}09b_2$

$59{,}6 = 11{,}7b_0 + 33{,}09b_1 + 106{,}03b_2$

$101{,}72 = 33{,}09b_0 + 106{,}02b_1 + 366{,}48b_2$



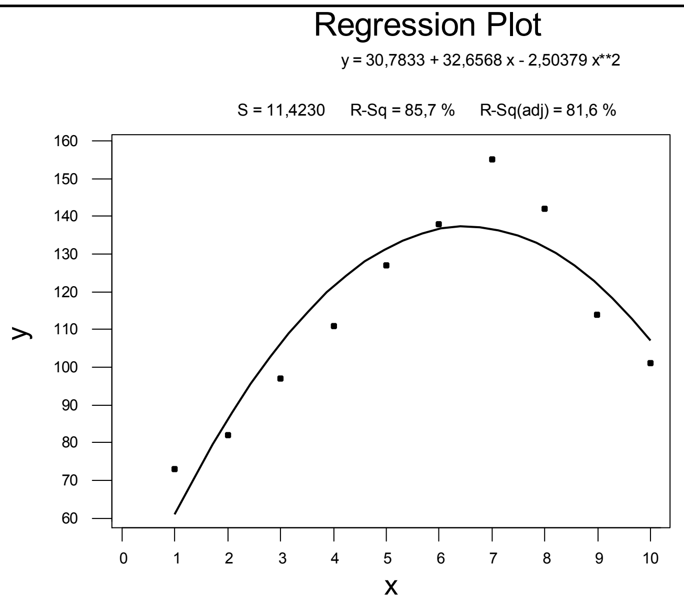
Soru :

	Y	X	X^2	X^3	X^4	XY	X^2Y
	73	1	1	1	1	73	73
	82	2	4	8	16	164	328
	97	3	9	27	81	291	873
	111	4	16			444	1776
	127	5	25			635	3175
	138	6	36			828	4968
	155	7	49			1085	7595
	142	8	64			1136	9088
	114	9	81			1026	9234
	101	10	100			1010	10100
TOP	1140	55	385	3025	25333	6692	47210

$$\sum Y = nb_0 + b_1\sum X + b_2\sum X^2$$
$$\sum XY = b_0\sum X + b_1\sum X^2 + b_2\sum X^3$$
$$\sum X^2Y = b_0\sum X^2 + b_1\sum X^3 + b_2\sum X^4$$

$1140 = 10b_0 + 55b_1 + 385b_2$
$6692 = 55b_0 + 385b_1 + 3025b_2$
$47210 = 385b_0 + 3025b_1 + 25333b_2$

$b_0 = 30,9$ $b_1 = 32,6$ $b_2 = -2,5$



2. Hiperbol Eğri $\hat{Y} = \dfrac{1}{b_0 + b_1 X}$

$\dfrac{1}{\hat{Y}} = \hat{F}$ olsun.

$$\frac{1}{\hat{Y}} = b_0 + b_1 X \quad = \hat{F}$$

$$\sum F = nb_0 + b_1 \sum X$$

$$\sum XF = b_0 \sum X + b_1 \sum X^2$$

Örnek : Aşağıdaki verilere göre $\hat{Y} = 1/(b_0 + b_1 X)$ regresyon denklemini tahmin ediniz.

	Y	X	F=1/Y	XF	X^2
	100	5	0,01	0,05	25
	33	10	0,03	0,30	100
	20	15	0,05	0,75	225
	25	20	0,04	0,80	400
	17	25	0,06	1,50	625
TOP		75	0,19	3,40	1375

$$\sum F = nb_0 + b_1 \sum X$$

$$\sum XF = b_0 \sum X + b_1 \sum X^2$$

$0,19 = 5b_0 + 75b_1$

$3,40 = 75b_0 + 1375b_1$

$b_0 = 0,005$, $b_1 = 0,0022$

$$\hat{Y} = \frac{1}{0,005 + 0,0022X}$$

3. Geometrik (Power) Eğri

$$\hat{Y} = b_0 X^{b_1}$$

$$\log \hat{Y} = \log b_0 + b_1 \log X$$

$$\hat{F} = k + b_1 G$$

$$\sum F = nk + b_1 \sum G$$

$$\sum GF = k \sum G + b_1 \sum G^2$$

Örnek : Aşağıdaki verilere göre $\hat{Y} = b_0 X^{b_1}$ power regresyon denklemini tahmin ediniz.

	Y	X	F=logY	G=logX	GF	G^2
	25	2	1,398	0,301	0,421	0,091
	38	4	1,580	0,602	0,951	0,362
	58	8	1,763	0,903	1,592	0,815
	87	16	1,940	1,204	2,336	1,450
	132	32	2,121	1,505	3,192	2,265
	200	64	2,301	1,806	4,156	3,262
	300	128	2,477	2,107	5,219	4,439
TOP			13,580	8,428	17,867	12,684

$$\sum F = nk + b_1 \sum G$$

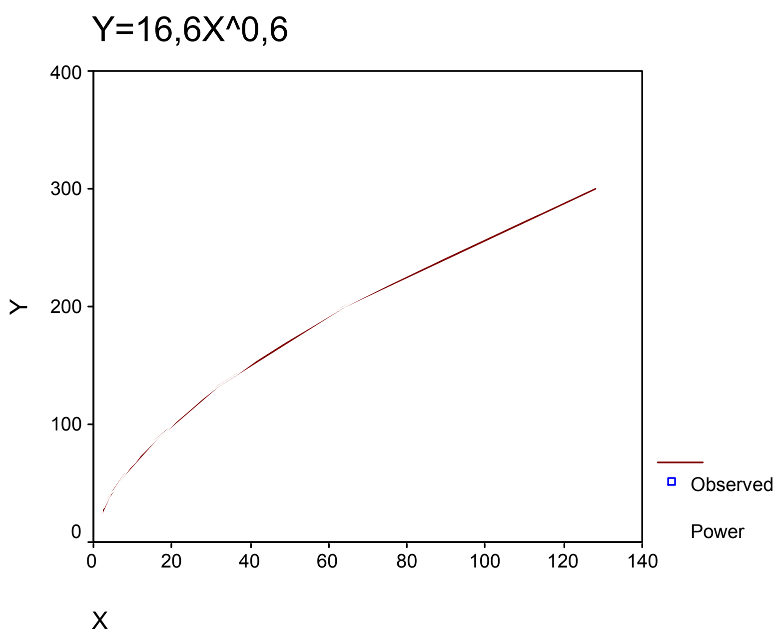
$$\sum GF = k\sum G + b_1 \sum G^2$$

$13,580 = 7k + 8,428 b_1$

$17,867 = 8,428k + 12,684 b_1$

$k = 1,22 = \log b_0$, $b_0 = 16,596$ $b_1 = 0,598$

$$\hat{Y} = 16,596 X^{0,598}$$



Soru : Aşağıdaki verilere göre power (geometrik) regresyon denklemini tahmin ediniz.

	Y	X	F=logY	G=logX	GF	G^2
	4,00	1,00	0,60	0,00	0,00	0,00
	6,00	2,00	0,78	0,30	0,23	0,09
	8,00	4,00	0,90	0,60	0,54	0,36
	10,00	5,00	1,00	0,70	0,70	0,49
	15,00	6,00	1,18	0,78	0,92	0,61
	20,00	10,00	1,30	1,00	1,30	1,00
TOP	63,00	28,00	5,76	3,38	3,69	2,55

$$\sum F = nk + b_1 \sum G$$

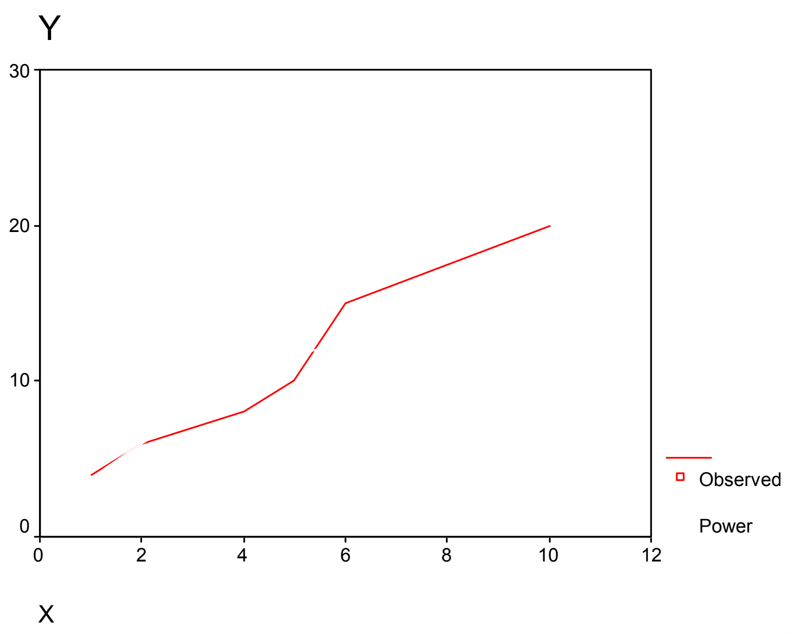
$$\sum GF = k \sum G + b_1 \sum G^2$$

$5{,}76=6k+3{,}38b_1$

$3{,}69=3{,}38k+2{,}55b_1$

$b_0=3{,}69$ $\quad$ $b_1=0{,}69$

$$\hat{Y} = 3{,}69X^{0{,}69}$$



4. Üstel (Compound) Model

$$\hat{Y} = b_0 * b_1^{X}$$

$$\log \hat{Y} = \log b_0 + (\log b_1)X$$

$$\hat{F} = k + mX$$

$$\sum F = nk + m\sum X$$

$$\sum XF = k\sum X + m\sum X^2$$

Örnek : Aşağıdaki verilere göre compound regresyon denklemini tahmin ediniz?

	Y	X	F=logY	XF	X^2
	105,6	1	2,024	2,024	1
	111,8	2	2,048	4,096	4
	118,0	3	2,072	6,216	9
	124,9	4	2,097	8,388	16
	131,7	5	2,120	10,600	25
TOP		15	10,361	31,324	55

$$\sum F = nk + m\sum X$$

$$\sum XF = k\sum X + m\sum X^2$$

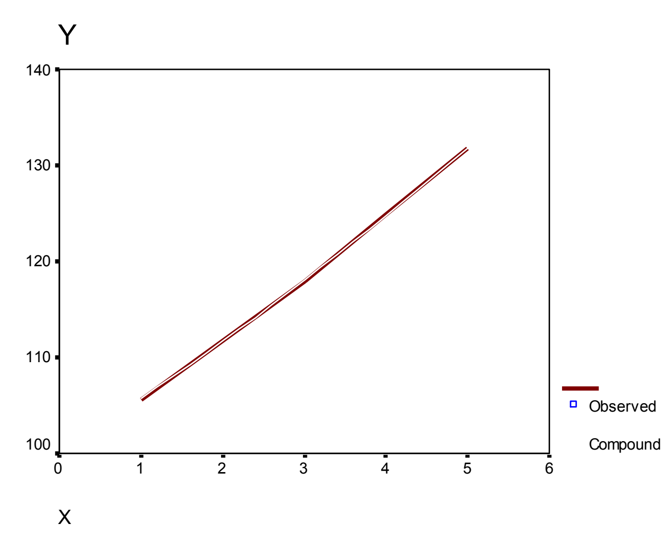
10,361=5k+15m

31,324=15k+55m

k=2=logb_0, b_0=100 m=0,024=logb_1 , b_1=1,057

$$\hat{Y} = 100(1{,}057)^{X}$$

Dependent	Mth	Rsq	d.f.	F	Sigf	b0	b1
Y	COM	1,000	3	21428,6	,000	100,009	1,0568



SORU: Aşağıdaki verilere göre compound regresyon denklemini tahmin ediniz?

	Y	X	F=logY	XF	X^2
	10	1	1,00	1,00	1,00
	12	2	1,08	2,16	4,00
	15	4	1,18	4,70	16,00
	20	5	1,30	6,51	25,00
	25	6	1,40	8,39	36,00
TOP	82,00	18,00	5,95	22,76	82,00

$$\sum F = nk + m\sum X$$

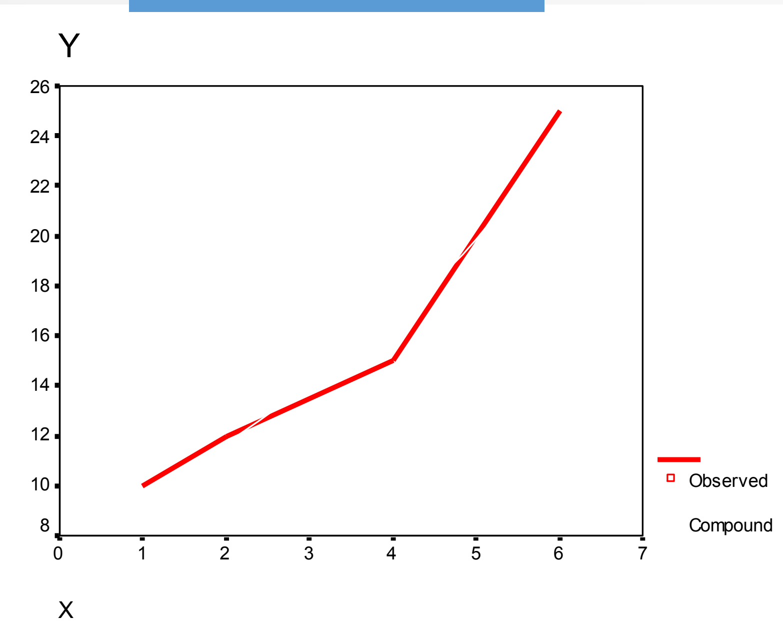
$$\sum XF = k\sum X + m\sum X^2$$

5,95 =5k+18m

22,76=18k+82m

$$\hat{Y} = 8,21(1,19)^X$$

$$\hat{Y} = 8{,}21(1{,}19)^X$$



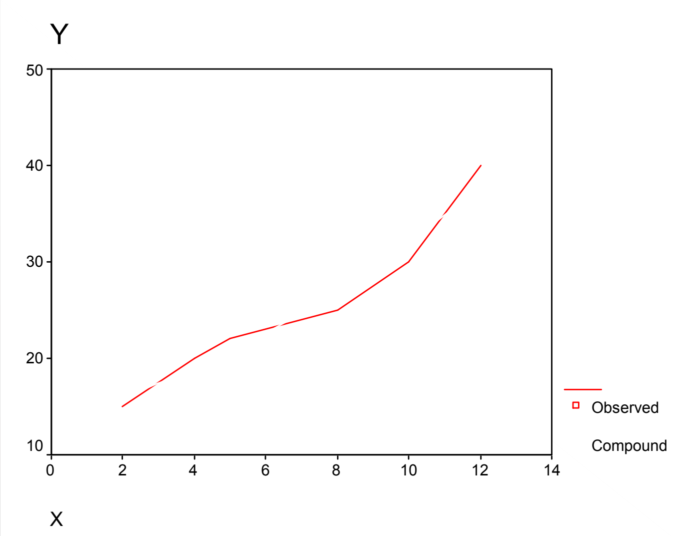
SORU: Aşağıdaki verilere göre compound regresyon denklemini tahmin ediniz?

Y	X	F=logY	XF	X^2
15	2			
20	4			
22	5			
25	8			
30	10			
40	12			

$$\sum F = nk + m\sum X$$

$$\sum XF = k\sum X + m\sum X^2$$

Dependent	Mth	Rsq	d.f.	F	Sigf	b0	b1
Y	COM	,962	4	101,31	,001	13,31	1,09



PROBLEMLER

SORU 1. 10 öğrencinin haftalık çalışma saati ile akademik ortalama değerleri aşağıda verilmiştir.

a) Bu değerlerden yararlanarak çalışma saati ile akademik ortalama arasındaki doğrusal ilişkinin matematiksel modelini kurunuz?

b) Kurulan modelin anlamlı olup olmadığını ANOVA tablosu ile araştırınız?

c) Modelin katsayılarının önemli olup olmadığını test ediniz?

d) Belirtme katsayısını bulunuz ve anlamlı olup olmadığını test ediniz?

T tablo=3,83 F Tablo=5,12

Çalışma Saati (hafta)	:15	8	9	18	14	20	7	2	12	10
Akademik Ortalama	:2,5	1,5	1,8	3	2,9	3,6	1,5	1	1,8	2

SORU 2. Buğdayın verimi ile protein yönünden zenginliği arasında doğrusal bir ilişki bulunduğu ileri sürülmektedir. Bir tarladan elde edilen buğday ağırlıkları ile protein miktarları aşağıda verilmiştir. %5 anlamlılık seviyesine göre;

a) Bu değerlerden yararlanarak buğday ağırlığı ile protein miktarı arasındaki doğrusal ilişkinin matematiksel modelini kurunuz?

b) Kurulan modelin anlamlı olup olmadığını ANOVA tablosu ile araştırınız?

c) Modelin katsayılarının önemli olup olmadığını test ediniz?

d) Belirtme katsayısını bulunuz ve anlamlı olup olmadığını test ediniz?

T tablo=7,45 F tablo=7,71

Buğday Ağırlığı (gr)	:150	400	300	300	350
Protein Miktarı (%)	:16	13	15	13	13

SORU 3. Bir fabrikada çalışan işçi sayısı ile bir yıl içinde dokunan kumaş uzunluğu için 5 yıl verileri aşağıda verilmiştir. %5 anlamlılık seviyesine göre;

a) Kuadratik (kareli) regresyon modeline göre matematiksel modeli kurunuz?

b) Compound regresyon modeline göre matematiksel modeli kurunuz?

c) Power regresyon modeline göre matematiksel modeli kurunuz?

d) Hiperbol regresyon modeline göre matematiksel modeli kurunuz?

İşçi sayısı	:1	2	3	4	5
Dokunan kumaş (m^2)	:4	6	10	15	20

SORU 4. Bir tarladaki soya fasulyelerinin haftalara göre boy uzunluğunun değişip değişmediği araştırılmak isteniyor. Buna göre %5 anlamlılık seviyesinde;

a) Kuadratik (kareli) regresyon modeline göre matematiksel modeli kurunuz?
b) Compound regresyon modeline göre matematiksel modeli kurunuz?
c) Power regresyon modeline göre matematiksel modeli kurunuz?
d) Hiperbol regresyon modeline göre matematiksel modeli kurunuz?

Boy uzunluğu (cm)	:5	13	16	23	33	38	40
Haftalara göre yaş	:1	2	3	4	5	6	7

4. ENDEKSLER

Endeks bir olaya ait sayısal verilerde zaman veya mekan boyutunda meydana gelen oransal değişimin göstergesidir.

Endeks değer değişmelerinin oransal bir ölçüsüdür. Bu oransal değişmeler zaman içinde meydana gelebileceği gibi mekanlar arasında da ortaya çıkabilir.

Günümüzde endeksler, çok yaygın bir kullanım alanına sahiptir. Özellikle üretim, tüketim, dış ticaret, para, fiyat gibi durumlar için çeşitli endeksler düzenlenmektedir.

Endekste biri kıyaslanan, diğeri temel olmak üzere iki değer vardır. Kıyaslanan değer paya, temel değer paydaya yazılır. Oransal kıyaslamayı kolaylaştırmak içinde *100 alınır. Böylece temel değer 100 kabul edilerek, diğerlerinin buna göre kaç olacağı belirlenmiş olur.

X_i : Kıyaslanan değer
X_0:Temel değer

$$I_{i/0} = \frac{X_i}{X_0} \times 100$$

ENDEKS ÇEŞİTLERİ

Endeks hesabında, endeksin cinsine ve kapsamına uygun çeşitli yöntemler kullanılır. Bu nedenle, endeks hesaplama yöntemlerini endeks çeşitleri ile birlikte incelemek kolaylık sağlar. Endeksler şu şekilde sınıflandırılmaktadır.

1. Yer ve zaman endeksleri
2. Sabit ve değişken esaslı endeksler
3. Basit ve bileşik endeksler

4.1.YER(MEKAN) VE ZAMAN ENDEKSLERİ

4.1.1. Mekan Endeksleri

Nüfus, üretim ve fiyat gibi herhangi bir istatistiki değişkene ait değerlerin bölgeler, iller vb. gibi mekanlar itibariyle gösterdiği oransal değişimin ölçüsüne mekan (yer) endeksi denir. Mekan endeksinde serinin aritmetik ortalaması kullanılır.

$$I_m = \frac{X_i}{\overline{X}} \times 100$$

Örnek 4.1. 1991 yılında 10 şehrimizde ortalama koyun eti fiyatlarının (TL/Kg) aşağıdaki şekilde bulunmuştur. Bu şehirlerimize ait mekan endekslerini bulunuz?

Şehirler	Fiyat
İstanbul	21.210
Ankara	18.560
İzmir	22.270
Adana	18.580
Konya	19.830
Samsun	20.830
Sivas	19.250
Erzurum	18.670
Gaziantep	19.330
Antalya	20.420
Toplam	198.950

$$\overline{X} = 198.950 / 10 = 19.895$$

$$I_{ist.} = \frac{21.210}{19.895} \times 100 = 106.6$$

$$I_{ist.} = \frac{18.560}{19.895} \times 100 = 93.3$$

Şehirler	Endeks (%)
İstanbul	106.6
Ankara	93.3
İzmir	111.9
Adana	93.4
Konya	99.7
Samsun	104.7
Sivas	96.8
Erzurum	93.8
Gaziantep	97.2
Antalya	102.6
Toplam	1000.0
Ortalama	100.0

Yorum : Koyun eti fiyatı ortalamaya göre İstanbul'da %6.6, İzmir'de %11.9, Samsun'da %4.7 ve Antalya'da %2.6 oranında yüksektir. Buna karşılık koyun eti fiyatı ortalamaya göre Ankara'da %6.7, Adana'da %6.6, Konya'da %0.3, Sivas'ta %3.2, Erzurum'da %6.2 ve Gaziantep'te %2.8 oranında düşüktür.

4.1.2. Zaman Endeksi

Nüfus, üretim ve fiyat gibi istatistiki değişkenlere ait değerlerin zaman itibariyle gösterdiği oransal değişimlerin ölçüsüne zaman endeksi denir. Bu endeksler bir zaman serisine dayanırlar.

$$I_z = \frac{X_i}{X_0} \times 100$$

Örnek 4.2. 1985-1992 yılları arasında İstanbul'daki ortalama zeytinyağı fiyatları (TL/Kg) aşağıdaki gibi bulunmuştur. Bu şehrimize ait zaman endekslerini hesaplayınız?

Yıllar	Fiyat
1985	6.630
1986	7.270
1987	7.260
1988	8.680
1989	8.350
1990	10.960
1991	12.270
1992	16.540

$$I_{1/0} = \frac{X_1}{X_0} \times 100 = \frac{7.270}{6.630} \times 100 = \%109.6$$

$$I_{7/0} = \frac{X_7}{X_0} \times 100 = \frac{16.540}{6.630} \times 100 = \%249.5$$

Yıllar	Endeks (%)
1985 - 0	100.0
1986 - 1	109.6
1987 - 2	109.5
1988 - 3	130.9
1989 - 4	125.9
1990 - 5	165.3
1991 - 6	185.1
1992 - 7	249.5

Yorum : Tüm yıllarda zeytinyağı fiyatı esas devre (1985) fiyatına göre artmıştır. Fiyat artışları esas devreye göre 1986'da %9.6, 1987'de %9.5,…,1992'de ise %149.5 oranında olmuştur.

Soru : 1985-1991 yılları arasında Ankara'daki simit fiyatları (TL/Adet) aşağıdaki gibi bulunmuştur. Bu şehrimize ait zaman endekslerini hesaplayınız?

Yıllar	Fiyat
1985	380
1986	320
1987	350
1988	330
1989	360
1990	450
1991	520

Yıllar	Endeks(%)
1985	100.0
1986	84.2
1987	92.1
1988	86.8
1989	94.7
1990	118.4
1991	138.8

4.2. SABİT VE DEĞİŞKEN ESASLI ENDEKSLER

4.2.1. Sabit Esaslı Endeksler

Endekslerin sabit ve değişken esaslı olması genellikle zaman serileri için söz konusudur. Belirli bir dönemi veya belirli bazı dönemlerin ortalaması esas kabul edilip, serinin bütün değerlerini bunun yüzdesi olarak göstermesiyle elde edilen endekse sabit esaslı endeks denir.

Esas devrenin enflasyon (paranın değerini kaybetmesi-tüketicinin satın alma gücünü kaybetmesi), devalüasyon (Ulusal paranın, yabancı para birimleri karşısında değerinin isteyerek belli bir amaca yönelik olarak düşürülmesi) ve deflasyon (fiyat seviyesindeki genel düşüş) gibi iktisadi olayların aşırılık kazanmadığı yani istikrarlı bir yıl olması şarttır.

$$I_{SEE} = I_{i/0} = \frac{X_i}{X_0} \times 100$$

Örnek 4.3. **Bir ülkenin 1990-1998 yılları arasında gerçekleşen ihracatı aşağıdaki gibidir. 1990 yılını esas devre kabul edip, sabit esaslı endeksleri hesaplayınız?**

Yıllar	İhracat (bin Ton)
1990	2.158
1991	2.055
1992	2.318
1993	1.131
1994	2.620
1995	2.661
1996	2.564
1997	2.023
1998	2.182

$$I_{4/0} = \frac{X_4}{X_0} \times 100 = \frac{2.620}{2.158} \times 100 = \%121.4$$

$$I_{7/0} = \frac{X_7}{X_0} \times 100 = \frac{2.023}{2.158} \times 100 = \%93.7$$

Yıllar	Endeks (%)
1990 – 0	100.0
1991 – 1	95.2
1992 – 2	107.4
1993 – 3	98.7
1994 – 4	121.4
1995 – 5	123.3
1996 – 6	118.8
1997 – 7	93.7
1998 - 8	101.1

Yorum : 1990 yılına göre ihracat 1992'de %7.4 , 1994'te %21.4 , 1995'te %23.3 , 1996'da %18.8 ve 1998'de %1.3 oranında artmıştır. Buna karşılık 1991'de %4.8 , 1993'te %1.3 ve %1997'de %6.3 oranında azalmıştır.

4.2.2. Değişken Esaslı Endeksler

Esas alınan dönem değişken olduğunda, yani her değer bir önceki dönemin değeri ile karşılaştırıldığında buna değişken esaslı endeks denir.

Kesrin paydası sabit kaldığında sabit esaslı endeks, değiştiğinde ise değişken esaslı endeksten söz edilir.

$$I_{DEE} = I_{i/(i-1)} = \frac{X_i}{X_{i-1}} \times 100$$

Örnek 4.4. Bir ülkenin 1990-1998 yılları arasında gerçekleşen ihracatı aşağıdaki gibidir. Değişken esaslı endeksleri hesaplayınız?

Yıllar	İhracat (bin Ton)
1990	2.158
1991	2.055
1992	2.318
1993	1.131
1994	2.620
1995	2.661
1996	2.564
1997	2.023
1998	2.182

$$I_{5/4} = \frac{X_5}{X_4} \times 100 = \frac{2.661}{2.620} \times 100 = \%101.6$$

$$I_{8/7} = \frac{X_8}{X_7} \times 100 = \frac{2.182}{2.023} \times 100 = \%107.9$$

Yıllar	Endeks (%)
1990 – 0	100.0
1991 – 1	95.2
1992 – 2	112.8
1993 – 3	91.9
1994 – 4	123.0
1995 – 5	101.6
1996 – 6	96.4
1997 – 7	78.9
1998 - 8	107.9

Yorum : Bir önceki döneme göre 1992'de %12.8 , 1994'te %23, 1995'te %1.6 ve 1998'te %7.9 oranında artış olmuştur. 1991'de %4.8 , 1993'de %8.1, 1996'da %3.6 ve 1997'de %21.1 oranında bir önceki döneme göre azalma olmuştur.

4.3. BASİT VE BİLEŞİK ENDEKSLER

Basit endeksler tek maddeyi kapsayan endekslerdir. Bileşik endeksler ise iki veya daha çok maddeyi kapsamaktadır.

4.3.1. Basit Endeksler

Basit endeks tek bir maddenin fiyatında veya miktarında zaman içinde meydana gelen oransal değişmeleri belirlemek için kullanılır. Aksi belirtilmedikçe endeks denildiğinde basit endeks anlaşılır.

Sabit Esaslı Fiyat Endeksi

$$I_{SEFE} = \frac{p_i}{p_0} \times 100$$

Sabit Esaslı Miktar Endeksi

$$I_{SEME} = \frac{q_i}{q_0} \times 100$$

Değişken Esaslı Fiyat Endeksi

$$I_{DEFE} = \frac{p_i}{p_{i-1}} \times 100$$

Değişken Esaslı Miktar Endeksi

$$I_{DEME} = \frac{q_i}{q_{i-1}} \times 100$$

p_0 : Esas devredeki fiyat
q_0: Esas devredeki miktar
p_i : i devresindeki fiyat
q_i : i devresindeki miktar

Örnek 4.4. **A malının çeşitli yıllardaki satış fiyatı ve miktarları aşağıdaki gibidir. Sabit ve değişken esaslı fiyat ve miktar endekslerini bulunuz?**

Yıllar	Fiyat (p)	Miktar (q)
1994	200	50
1995	250	40
1996	280	48
1997	420	36

$$I_{SEFE} = \frac{p_1}{p_0} \times 100 = \frac{250}{200} \times 100 = \%125$$

$$I_{SEFE} = \frac{p_2}{p_0} \times 100 = \frac{280}{200} \times 100 = \%140$$

$$I_{SEFE} = \frac{p_3}{p_0} \times 100 = \frac{420}{200} \times 100 = \%210$$

$$I_{SEME} = \frac{q_1}{q_0} \times 100 = \frac{40}{50} \times 100 = \%80$$

$$I_{SEME} = \frac{q_2}{q_0} \times 100 = \frac{48}{50} \times 100 = \%96$$

$$I_{SEME} = \frac{q_3}{q_0} \times 100 = \frac{36}{50} \times 100 = \%72$$

$$I_{DEFE} = \frac{p_1}{p_0} \times 100 = \frac{250}{200} \times 100 = \%125$$

$$I_{DEFE} = \frac{p_2}{p_1} \times 100 = \frac{280}{250} \times 100 = \%112$$

$$I_{DEFE} = \frac{p_3}{p_2} \times 100 = \frac{420}{280} \times 100 = \%150$$

$$I_{DEME} = \frac{q_1}{q_0} \times 100 = \frac{40}{50} \times 100 = \%80$$

$$I_{DEME} = \frac{q_2}{q_1} \times 100 = \frac{48}{40} \times 100 = \%120$$

$$I_{DEME} = \frac{q_3}{q_2} \times 100 = \frac{36}{48} \times 100 = \%75$$

4.3.2. Bileşik Endeksler

Birbiriyle ilgili iki veya daha fazla maddenin fiyatlarında veya miktarlarında zaman içinde meydana gelen oransal değişmelerin belirlenmesinde bileşik endeksler kullanılır.

Bileşik endekse dahil olacak madde sayısı ne çok az ne de çok fazla olmalıdır. Çünkü az sayıda maddeye dayanan endeks temsili olmayacağı gibi, çok sayıdaki madde de hesaplamayı zorlaştırır.

Uygulamada bileşik endeks Laspeyres-Paasche Endeksleri ve Fisher Endeksi yaklaşımları ile hesaplanır.

4.3.2.1. Laspeyres ve Paasche Endeksleri

Laspeyres ve Paasche endeksleri maddelerin önem farklılıklarını dikkate almamızı sağlayan endekslerdir. Bunlar hem fiyat hem de miktar serileri için hesaplanır.

Laspeyres Fiyat Endeksi

$$I_{LFE} = \frac{\sum p_i q_0}{\sum p_0 q_0} \times 100$$

Paasche Fiyat Endeksi

$$I_{PEE} = \frac{\sum p_i q_i}{\sum p_0 q_i} \times 100$$

Laspeyres Miktar Endeksi

$$I_{LME} = \frac{\sum p_0 q_i}{\sum p_0 q_0} \times 100$$

Paasche Miktar Endeksi

$$I_{PME} = \frac{\sum p_i q_i}{\sum p_i q_0} \times 100$$

i=0 ise yukarıdaki formüllerin sonucu %100 olur. Yani esas devrenin Laspeyres ve Paasch fiyat ve miktar endeksleri daima %100 dür.

p_0 : Esas devredeki fiyat
q_0: Esas devredeki miktar
p_i : i devresindeki fiyat
q_i : i devresindeki miktar

$\Sigma p_0 q_0$: Esas devrede satın alınan belirli miktarlardaki mallar için yapılan harcamalar toplamı
$\Sigma p_i q_0$: Esas devrede satın alınan belirli miktarlardaki malları i devresinde satın alabilmek için yapılması gerekli olan harcamalar toplamı
$\Sigma p_i q_i$: i devresinde satın alınan belirli miktarlardaki mallar için yapılan harcamalar toplamı
$\Sigma p_0 q_i$: i devresinde satın alınan belirli miktarlardaki malları esas devrede satın alabilmek için yapılması gerekli olan harcamalar toplamı

Not : Laspeyres endekslerinde tartılar değişmediği için, bu endeksler kıyaslanabilir. Paasch endekslerinde ise her devre için değişik tartılar kullanıldığından, kıyaslama yapmak mümkün değildir. Bu yüzden Paasch endeksleri pek tercih edilmez.

2003 Temel Yıllı Tüketici Fiyatları ve Üretici Fiyatları Endeksi Laspeyres formülü kullanılarak hesaplanmaktadır.

Örnek 4.5. 3 maddenin bazı yıllardaki satış fiyat ve miktarları aşağıdaki gibidir. Laspeyres ve Paasche fiyat ve miktar endekslerini bulunuz?

Yıllar	A Maddesi		B Maddesi		C Maddesi	
	Fiyat-p	Miktar-q	Fiyat-p	Miktar-q	Fiyat-p	Miktar-q
1996	20	15	30	30	15	20
1997	16	18	18	20	20	22
1998	24	16	22	18	23	16

Laspeyres Fiyat Endeksleri

$$I_{LFE} = \frac{\sum p_1 q_0}{\sum p_0 q_0} \times 100 = \frac{(16\times 15)+(18\times 30)+(20\times 20)}{(20\times 15)+(30\times 30)+(15\times 20)} \times 100 = \%78.7$$

$$I_{LFE} = \frac{\sum p_2 q_0}{\sum p_0 q_0} \times 100 = \frac{(24\times 15)+(22\times 30)+(23\times 20)}{(20\times 15)+(30\times 30)+(15\times 20)} \times 100 = \%98.7$$

3 maddenin fiyatlarında esas devre olan 1996'ya göre ortalama 1997'de %21.3 ve 1998'de %1.3 oranında bir azalış olmuştur.

Paasche Fiyat Endeksleri

$$I_{PFE} = \frac{\sum p_1 q_1}{\sum p_0 q_1} \times 100 = \frac{(16\times 18)+(18\times 20)+(20\times 22)}{(20\times 18)+(30\times 20)+(15\times 22)} \times 100 = \%84.3$$

$$I_{PFE} = \frac{\sum p_2 q_2}{\sum p_0 q_2} \times 100 = \frac{(24\times 16)+(22\times 18)+(23\times 16)}{(20\times 16)+(30\times 18)+(15\times 16)} \times 100 = \%104.4$$

Hesaplamada her bir yıl için farklı tartılar kullanıldığından kıyaslama yapılmaz.

Laspeyres Miktar Endeksleri

$$I_{LME} = \frac{\sum p_0 q_1}{\sum p_0 q_0} \times 100 = \frac{(20\times18)+(30\times20)+(15\times22)}{(20\times15)+(30\times30)+(15\times20)} \times 100 = \%86.0$$

$$I_{LME} = \frac{\sum p_0 q_2}{\sum p_0 q_0} \times 100 = \frac{(20\times16)+(30\times18)+(15\times16)}{(20\times15)+(30\times30)+(15\times20)} \times 100 = \%73.3$$

3 maddenin miktarlarında esas devre olan 1996'ya göre ortalama 1997'de %14 ve 1998'de %26.7 oranında bir azalış olmuştur.

Paasche Miktar Endeksleri

$$I_{PME} = \frac{\sum p_1 q_1}{\sum p_1 q_0} \times 100 = \frac{(16\times18)+(18\times20)+(20\times22)}{(16\times15)+(18\times30)+(20\times20)} \times 100 = \%92.2$$

$$I_{PME} = \frac{\sum p_2 q_2}{\sum p_2 q_0} \times 100 = \frac{(24\times16)+(22\times18)+(23\times16)}{(24\times15)+(22\times30)+(23\times20)} \times 100 = \%77.6$$

Farklı tartılarla hesaplama yapıldığından bu endeksler kıyaslanamaz.

4.3.2.2. Fisher Endeksi

Fisher endeksi Laspeyres ve Paasche endekslerinin geometrik ortalamasıdır.

Fisher Fiyat Endeksi

$$I_{FisherFE} = \sqrt{LFE \times PFE} = \sqrt{\left(\frac{\sum p_i q_0}{\sum p_0 q_0} \times 100\right) \times \left(\frac{\sum p_i q_i}{\sum p_0 q_i} \times 100\right)}$$

Fisher Miktar Endeksi

$$I_{FisherME} = \sqrt{LME \times PME} = \sqrt{\left(\frac{\sum p_0 q_i}{\sum p_0 q_0} \times 100\right) \times \left(\frac{\sum p_i q_i}{\sum p_i q_0} \times 100\right)}$$

Örnek 4.6. **Örnek 4.5'e göre Fisher fiyat ve miktar endekslerini bulunuz?**

Fisher Fiyat Endeksi

$$I_{FisherFE1997} = \sqrt{\left(\frac{\sum p_i q_0}{\sum p_0 q_0} \times 100\right) \times \left(\frac{\sum p_i q_i}{\sum p_0 q_i} \times 100\right)}$$

$$= \sqrt{(78.7) \times (84.3)} = \%81.5$$

$$I_{FisherME1998} = \sqrt{(98.7) \times (104.4)} = \%101.5$$

Fisher Miktar Endeksi

$$I_{FisherFE1997} = \sqrt{\left(\frac{\sum p_0 q_i}{\sum p_0 q_0} \times 100\right) \times \left(\frac{\sum p_i q_i}{\sum p_i q_0} \times 100\right)}$$

$$= \sqrt{(86) \times (92.2)} \;\; = \%89$$

$$I_{FisherFE1998} = \sqrt{(73.3) \times (77.6)} \;\; = \%75.4$$

4.4. ENFLASYON ORANI

Enflasyon fiyatlarda meydana gelen şişkinliktir. Bir zaman endeksinde ardışık dönemler arasında meydana gelen büyüme oranına enflasyon oranı adı verilir. Buna göre t dönemindeki enflasyon oranı aşağıdaki formülle hesaplanabilir.

$$P_{t:t-1} = \frac{I_t - I_{t-1}}{I_{t-1}} \times 100$$

Enflasyon değerleri hesaplanırken amaca göre uygun endeks türü seçilmelidir. Örneğin sendikaların ücret zammı görüşmelerinde TÜFE, Kamuya iş yapmış müteahhit alacaklarının güncellenmesinde ÜFE'deki enflasyon dikkate alınmalıdır.

Enflasyon ardışık yıllar yerine herhangi bir s yılına göre de bulunabilir.

$$P_{t:t-1} = \frac{I_t - I_{t-s}}{I_{t-s}} \times 100$$

Örnek 4.7. Aşağıdaki tabloda verilen endeks değerlerine göre enflasyon oranlarını bulunuz.

Yıllar	Fisher Endeks
1990	100.0
1991	112.3
1992	148.3
1993	167.3

$$P_{1991} = \frac{I_t - I_{t-1}}{I_{t-1}} \times 100 = \frac{112.3 - 100.0}{100.0} \times 100 = 12.3$$

$$P_{1992} = \frac{148.3 - 112.3}{112.3} \times 100 = 32.0$$

$$P_{1993} = \frac{167.3 - 148.3}{148.3} \times 100 = 12.8$$

5. KAYNAKLAR

- Serper Ö (2000). Uygulamalı İstatistik II. Ezgi Kitabevi, Bursa.
- Aytaç, M (1999). Matematiksel İstatistik, Ezgi Kitabevi, Bursa.
- Kartal, M. (2006). Bilimsel Araştırmalarda Hipotez Testleri, Nobel Yayın Dağıtım, Ankara.
- Orhunbilge, N. (2000). Örnekleme Yöntemleri ve Hipotez Testleri, Avcıol Basım Yayın, İstanbul.
- Karagöz, M. (2009). İstatistik Yöntemleri, Ekin Basın Yayın Dağıtım, Bursa. Akdeniz F.(1984). Olasılık ve İstatistik, Ankara Ünv. Basımevi, Ankara.
- Canküyer, E. ve Aşan, Z. (2001). Parametrik Olmayan İstatistiksel Teknikler, Anadolu Ünv. Yayınları, No:1266, Eskişehir.
- Gamgam, H.(1998). Parametrik Olmayan İstatistiksel Teknikler, Gazi Ün. Yayın No:140.Ankara
- Şenol, Ş. (2008). Çıkarsamalı İstatistik, Nobel Yayın Dağıtım, Ankara.
- Üçkardeş, F(2006). İstatistik Testler Üzerine Bir Çalışma, Sütçü İmam Ünv. Fen Bil. Ens. Zootekni ABD, Y.Lisans Tezi, K.Maraş.
- İstatistik (2010). Murat Açık Öğretim Yayınları.
- Esin, A. Çelebioğlu, S. (1999). İstatistik, Schaum's Outlines, Nobel Yayın Dağıtım.
- Çil, B. (2004). İstatistik, Detay Yayıncılık.
- Alpar, R. (2010). Uygulamalı İstatistik ve Geçerlik-Güvenilirlik, Detay Yayıncılık, Ankara.