İSTATİSTİK-II

5

Amaçlarımız

Bu üniteyi tamamladıktan sonra;

- İki değişken arasındaki ilişkiyi açıklayan doğrusal modeli kurabilecek,
- İki değişken arasındaki ilişkinin derecesini belirleyebilecek

bilgi ve becerilere sahip olacaksınız.

Anahtar Kavramlar

- Regresyon Analizi
- Basit Doğrusal Regresyon Analizi
- En Küçük Kareler Tekniği
- Varyansın(σ^2) Tahmini
- Basit Doğrusal Regresyonda Aralık Tahmini
- Regresyon Katsayısının Anlamlılık Testi
- Korelasyon Analizi
- Korelasyon Katsayısı
- Belirlilik Katsayısı
- Korelasyon Katsayısının Anlamlılık Testi

İçindekiler

Regresyon ve Korelasyon Analizi

GİRİŞ

Regresyon teriminin kökeni Sör Francis Galton'a dayanır. Galton bir değişken üzerinde meydana gelen değişimin başka bir ya da daha fazla değişken tarafından açıklanıp açıklanamayacağı konusunda çalışan ilk araştırmacılardandır. 1885 yılında kalıtım konusunda çalışırken babalar ve oğulların boyları konusunda yaptığı araştırmasında, oğulların boylarının ortalamaya doğru yönlendiğini (İngilizce karşılığı "regressed") belirterek regresyon kelimesinin temelini oluşturmuştur. Günümüzde ise regresyona, değişkenler arasındaki ilişkinin modellenmesi işlemlerinin tümünü içeren geniş bir anlam yüklenmiştir.

Korelasyon ise birlikte değişim gösterme eğilimindeki değişkenler arasındaki ilişkinin derecesini temsil eden bir katsayıdır. Bu katsayı değişkenler arasındaki ilişkinin derecesinin yanı sıra ilişkinin yönünüde belirler. İki değişken arasında yüksek korelasyon olması bile, iki değişkenden birinin diğerinin nedeni olabileceğini göstermez. Korelasyon analizi iki değişken arasındaki nedensellik için kullanılamadığından, nedensellik araştırması için diğer farklı istatistik tekniklerinin kullanılması gerekir.

REGRESYON ANALİZİ

Regresyon analizinde iki veya daha fazla değişken arasındaki ilişkinin hangi matematiksel modelle ifade edileceği araştırılır. Sözü edilen iki değişken arasındaki ilişki genellikle neden-sonuç ilişkisi biçiminde ortaya çıkar: Gelir, Harcama; Yaş, Boy; Gübre miktarı, Verim miktarı; Toprak kalitesi, Verim miktarı; Çalışma süresi, Kıdem tazminatı ikilileri arasında neden-sonuç ilişkisi bulunmaktadır. Bu değişkenlerden sonuç durumundaki değişken bağımlı değişken neden durumundaki değişken ise bağımsız değişken olarak isimlendirilir. Regresyon analizinde bağımlı değişken üzerinde oluşan değişimlerin açıklanmasına çalışılır. Örneğin yukarıdaki ikililerin ilkinde, bir bireyin "harcaması" *bağımlı değişken* olarak ele alınmıştır. *Bağımsız değişken* ise neden durumunda olan "gelir" değişkenidir. Genellikle bağımlı değişken Y sembolüyle bağımsız değişken ise X'le ifade edilir.

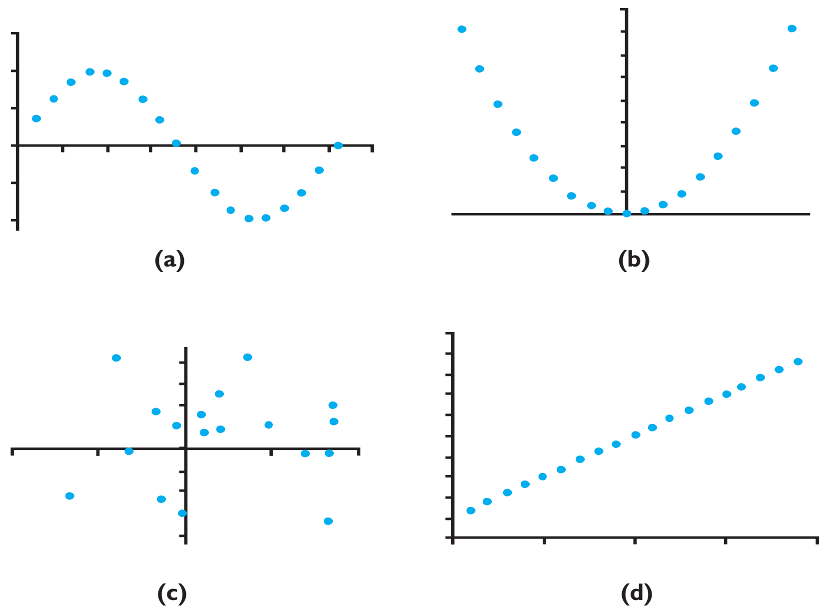
Değişkenler arasındaki ilişkiyi göstermenin ilk yolu grafik yöntemidir. X ve Y gözlem ikilileri bir grafik üzerinde birer nokta hâlinde gösterilir. İşaretlenen bu noktaların oluşturduğu şekil **serpilme diyagramı** olarak isimlendirilir. Farklı (X,Y) ikilileri için Şekil 5.1'dekine benzer görünümler elde edilebilir. Yatay eksen X; düşey eksen Y değişkeni olmak üzere (X,Y) koordinatlı noktalardan oluşan Şe-

X ve Y gözlem ikililerinin grafik üzerinde birer nokta hâlinde işaretlendiği şekil **serpilme diyagramı** olarak isimlendirilir.

kil 5.1'de verilen grafikler birer serpilme diyagramıdır. Şekil 5.1'de iki değişken arasında aldıkları değerlere göre gözlemlenebilecek 4 farklı serpilme diyagramı örneklenmiştir. Serpilme diyagramı üzerinde bağımlı ve bağımsız değişken arasındaki ilişkinin yapısı gözlenebilmektedir. Şekil 5.1(a) ve 5.1(b)'de iki değişken arasında doğrusal olmayan ilişki, 5.1(d)'de doğrusal ilişki görülmektedir. Şekil 5.1(c)'de ise iki değişken arasında ilişki olmadığı söylenebilir. Noktaların oluşturduğu şekle bakarak ilişkinin yönü ve derecesini tahmin etmek de mümkündür. Bunun için noktaların en dışta kalanları birleştirilerek bir şekil elde edilir. Söz konusu şeklin durumuna göre ilişkinin derecesi hakkında tahminde bulunulur. Eğer şekil dar bir elipse benziyorsa ilişki kuvvetlidir, elips genişledikçe ilişki zayıflar.

Şekil 5.1

Serpilme Diyagramları



Basit Doğrusal Regresyon

Regresyon analizinde bağımsız değişken sayısı bir olması durumunda *basit regresyon analizi;* birden fazla bağımsız değişken olması durumunda ise *çoklu regresyon analizi* söz konusu olmaktadır. Bu ünitede sadece basit doğrusal regresyon analizine yer verilecektir.

Serpilme diyagramı sadece ilişki tipi üzerinde bir fikir verebilmektedir. Bu ilişkinin matematiksel bir ifadesinin geliştirilmesi gerekir. Bu denklem elde edildiğinde, bağımlı değişken Y bağımsız değişken X olmak üzere, Y'nin değerlerinin tahmini için X'in değerleri kullanılır. X, Y değişkenleri arasındaki doğrusal ilişkinin ifadesinde kullanılan eşitliğe doğrusal regresyon modeli ya da kısaca regresyon denklemi denir.

Basit doğrusal regresyon modeli,

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad i=1,2,3,...,N$$

şeklinde stokastik (olasılıklı) bir modeldir. Model'de

Y_i ; bağımlı değişkeninin i'inci gözlemlenen değerini,
X_i ; bağımsız değişkeninin i'inci gözlemlenen değerini,
ε_i ; hata terimi,
α : bilinmeyen sabit parametre,
β : bilinmeyen regresyon parametresidir.

α, doğrusal modelin sabit terimidir ve X=0 olduğunda regresyon doğrusunun dikey eksen Y'yi kestiği noktayı göstermektedir. β ise doğrusal modelin eğimini vermektedir ve regresyon analizinde, bağımsız değişken X'deki bir birimlik değişmenin, bağımlı değişken Y'de ne kadarlık bir değişmeye yol açacağını gösteren regresyon parametresidir.

$Y = \alpha + \beta X + \varepsilon$ şeklindeki

doğrusal model, X ve Y değişkenlerinin evrenleri için geçerlidir. İstatistiksel çalışmaların çoğunda olduğu gibi, regresyon analizinde de anakütleye ilişkin verilerin tümüne ulaşılamadığından bu evrenden seçilen örneklem verileriyle analiz yapılır. Örneklem verilerinden hareketle anakütle parametreleri α ve β nın tahminlerini elde edebilmek için EKK yönteminden yararlanılabilir. Bunun için öncelikle, gözlem ikililerini bir serpilme diyagramında gösterdiğimizi varsayalım. Serpilme diyagramı incelendiğinde doğrusal bir eğilim görülüyorsa, Y'nin X'e göre matematik fonksiyonunun doğrusal olduğuna (kesin olmasa da) karar verilebilir. Ancak gözlem noktaları arasından çok sayıda doğrusal fonksiyon geçirilebilir. Bu doğrusal fonksiyonlardan en uygunu Y_i gözlem değerlerine en yakın kuramsal (tahmin) $\hat{Y}_i$ değerini veren doğrusal fonksiyon olacaktır. Bir başka ifadeyle belirli bir X değeri için, elimizde iki ordinat değeri olacaktır; birincisi gözlem değeri, ikincisiyse bu noktanın doğru ya da eğri üzerinde teorik olarak hesaplanacak ordinat değeridir. İşte, $\hat{Y}_i$ kuramsal değerlerle Y_i gözlem değerleri arasındaki farklar hata terimlerini oluşturur.

$e_i = Y_i - \hat{Y}_i$ şeklinde hesaplanan hata terimleri "pozitif" ya da "negatif" ya da "sıfır" değerlerine sahip olurken, bu farkların cebirsel toplamı sıfıra eşittir.

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right) = 0$$

EKK Yönteminin esası α ve β'nın tahminleri olan a ve b'yi söz konusu farkların kareleri toplamını minimum yani

EKK Yöntemi (Y_i-$\hat{Y}_i$) artıklarının kareleri toplamını en küçük yapabilen α ve β'nın tahminlerinin elde edilmesine ilişkin bir yöntemdir.

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2 = \min$$

olacak şekilde belirlemektir.

α ve β'nın EKK tahminleri, yukarıdaki en küçük kareler yöntemi için,

$$\sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2 = \sum_{i=1}^{n} \left(Y_i - a - bX_i \right)^2$$

ifadesinin a ve b'ye göre türevleri alınıp sıfıra eşitlenerek:

$$\frac{\partial e}{\partial a} = -2\sum_{i=1}^{n}\left(Y_i - a - bX_i\right) = 0$$

$$\frac{\partial e}{\partial b} = -2\sum_{i=1}^{n}\left(Y_i - a - bX_i\right) = 0$$

bulunur, elde edilen bu iki eşitlikten;

$$\sum Y_i = an + b\sum X_i$$

$$\sum X_i Y_i = a\sum X_i + b\sum X_i^2$$

denklemleri bulunur. Bu denklemler "normal denklemler" olarak isimlendirilir. Doğru denklemleri ve katsayılarının en küçük kareler koşuluna uygun olarak hesaplanması, bu iki denklemin çözümüyle gerçekleşebilir. Normal denklemlerdeki diğer değerler n, ΣX_i, ΣY_i, ΣX_i^2 ve $\Sigma X_i Y_i$ dır ve denklemlerdeki X_i ve Y_i değerleri "sıfır orijinine" göre ifade edilmişlerdir. Bu değerleri hesapladıktan sonra basit doğrusal regresyon modelindeki α ve β nın tahminleri olan a ve b'yi normal denklemlere dayanarak kolaylıkla çözmek mümkündür. a ve b gibi iki bilinmeyenli iki denklem sisteminde bu katsayılar hesaplandığında Y'nin X'e göre tahmin edilen doğrusal regresyon denklemi,

$$\hat{Y} = a + bX$$

şeklinde ifade edilir.

Bu denklemde araştırmacı regresyon denklemini kullanarak herhangi bir X değeri için $\hat{Y}$ tahmin değerini hesaplayabilir. Bu şekilde denklemde X yerine ilgilendiği değeri yazan araştırmacı Y'nin modele göre beklenen değerini hesaplamış olur. Tersine, Y bağımlı değişkeninin istenen belirli bir değerini üretecek bir X bağımsız değişkeni için beklenen değeri de hesaplanan Y yerine ilgilendiği değeri yazan araştırmacı, denklemi çözerek X sonucunu elde edecektir.

Tahmin edilen basit doğrusal regresyon modelindeki b katsayısının sıfırdan farklı olması Y ile X arasında bir bağıntı olduğunu göstermektedir. b katsayısının işareti ise X ile Y arasındaki ilişkinin yönünü belirler. Bu işaret pozitif ise iki değişken arasında aynı yönlü negatif ise ters yönlü bir ilişkiden söz edilir.

ÖRNEK 1

Bir araştırmacı bir mağaza için rastgele seçtiği 10 günde gözlem yapmış ve günler itibarıyla gelen günlük müşteri sayısı ve elde edilen ciro (₺1 000) değerlerini gözlemlemiştir. Bu veriler Tablo 5.1'de yer almaktadır.

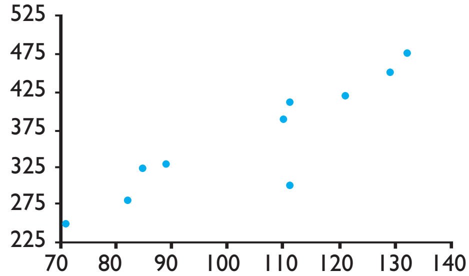
a. Serpilme diyagramını çiziniz?
b. Cironun, müşteri sayısına göre, basit doğrusal regresyon denkleminin tahminini, en küçük kareler tekniğiyle elde ediniz.

	Gün1	Gün2	Gün3	Gün4	Gün5	Gün6	Gün7	Gün8	Gün9	Gün10
Müşteri	71	82	111	85	89	110	111	121	129	132
Ciro	250	280	301	325	328	390	410	420	450	475

Tablo 5.1
Regresyon Modeli İçin Gözlem Verisi Tablosu

a.

Şekil 5.2



Ciroya ilişkin serpilme diyagramı

Kartezyen koordinat sisteminde X ve Y'ye ait verileri işaretlediğimizde serpilme diyagramını çizmiş oluruz. Grafik incelendiğinde iki değişken arasındaki ilişkinin doğrusal olduğunu görebiliriz.

b. Basit doğrusal regresyon denklemi $\hat{Y}$ = a + bX'i oluşturabilmek için a ve b katsayılarını normal denklemler yardımıyla hesaplayalım. Normal denklemler;

$$\Sigma Y_i = an + b\Sigma X_i$$

$$\Sigma X_i Y_i = a\Sigma X_i + b\Sigma X_i^2$$

dir. Bu amaçla; ΣX_i $\Sigma Y_{i,}$ $\Sigma X_i Y_{i,}$ ΣX_i^2 , değerleri hesaplanmış ve Tablo 5.2 oluşturulmuştur. Tablonun ilk 2 sütunu günlük müşteri sayısı (X) ve günlük ciro miktarı (Y) değişkenlerinin gözlem değerlerini göstermektedir. Tablo 5.2'nin sütunlarının altında yer alan sayılar ise ilgili sütunun değerleri toplamını göstermektedir.

X	Y	XY	X^2
71	250	17750	5041
82	280	22960	6724
111	301	33411	12321
85	325	27625	7225
89	328	29192	7921
110	390	42900	12100
111	410	45510	12321
121	420	50820	14641
129	450	58050	16641
132	475	62700	17424
ΣX=1041	**ΣY=3629**	**ΣXY=390918**	**ΣX² =112359**

Tablo 5.2
Regresyon Denklemi Katsayıları Hesaplama Tablosu

$3629 = 10a + 1041b$

$390918 = 1041a + 112359b$

Bu iki denklemde yer alan a ve b katsayıları;

a = 20,175; b = 3,29

olarak hesaplanır. Buna göre tahmin edilen regresyon doğrusu denklemi;

$\hat{Y} = 20,175 + 3,29X$

şeklinde elde edilir.

DİKKAT

Ünite boyunca hesaplamalarda hassasiyet sağlamak amacıyla olabildiğince çok ondalıklı hane kullanılırken, hesaplama sonuçlarının gösteriminde yuvarlatılmış daha az sayıda ondalık haneyle yetinilmiştir.

Regresyon katsayısı 3,29 bulunduğundan, müşteri sayısı bir birim değiştiğinde ciroda 3,29 birimlik değişme gerçekleşecektir. Regresyon denklemi yardımıyla araştırmacı gözlemlediği herhangi bir X değeri için Y'nin alacağı değeri tahmin edebilir. Örneğin, mağazayı 100 kişi ziyaret ettiğinde işletmenin tahmini cirosunun bilinmesi istenebilir. Bu durumda X=100 olarak gözlemlenmiş olduğundan Y'nin tahmini, $\hat{Y} = 20,175 + 3,29X = 20,175 + 3,29x100 = 349,402$ olarak hesaplanır. Benzer biçimde, regresyon denkleminden Y'nin belirli bir değerini veren X değeri de tahmin edilebilir. 600'lük bir ciro elde etmek için kaç müşterinin mağazaya gelmesine ihtiyaç olduğu, Y=600 için X'in modele göre beklenen değeri $600 = 20,175 + 3,29X$ denkleminden X = 176,117 ≅ 177 olarak bulunur.

Normal denklemlerde X ve Y değerleri yerine bunların aritmetik ortalamalarından sapmaları olan x ve y değerlerinin konulmasıyla;

$$\sum y_i = an + b\sum x_i$$

$$\sum x_i y_i = a\sum x_i + b\sum x_i^2$$

denklemleri elde edilir. Tanım gereğince $x_i = X - \bar{X}$ ve $y_i = Y - \bar{Y}$ olduğundan aritmetik ortalamanın temel özelliklerine göre (arimetik ortalamadan cebirsel sapmaların toplamı sıfırdır.) $\sum x_i = 0$ ve $\sum y_i = 0$'dır. Böylece son iki eşitlikten

$$b = \frac{\sum x_i y_i}{\sum x_i^2}$$

ya da

$$b_{yx} = \frac{\sum x_i y_i}{\sum x_i^2}$$

elde edilir. a'nın 0'a eşit olması, ortalamalar orijinine göre, regresyon doğrusunun, değişkenlerin ortalamalarıyla tanımlanan bir noktadan geçtiğini ortaya koymaktadır. Bu durumda regresyon denklemi

$\hat{y} = bx_i$ veya $\hat{y} = b_{yx}x_i$

şeklinde yazılabilir.

Bir regresyon doğrusu ister sıfır orijinine göre yazılsın ister ortalamalar orijinine göre yazılsın eğimi değişmez. Bu nedenle her iki orijine göre hesaplanan b katsayısı aynıdır. Buna karşılık a ise;

$a = \bar{Y} - b\bar{X}$

eşitliğinden hesaplanır. Veriler için en iyi doğru denklemi

$\hat{Y} = a + bX$

şeklinde yazılabilir.

ÖRNEK 2

Örnek 1'deki veriler için Y'nin X'e göre regresyon denklemini, ortalamalar orijinine göre, en küçük kareler tekniğiyle elde edelim.

Basit doğrusal regresyon denklemini ortalamalar orijinine göre yazabilmek için, ilk olarak X ve Y serisinin aritmetik ortalamalarını ve x, y, xy ve x^2'leri elde edelim.

$$\bar{X} = \frac{\sum X_i}{n} = \frac{1041}{10} = 104,1 \qquad \bar{Y} = \frac{\sum Y_i}{n} = \frac{3629}{10} = 362,9$$

x_i	y_i	x_iy_i	x_i^2
-33,1	-112,9	3736,99	1095,61
-22,1	-82,9	1832,09	488,41
6,9	-61,9	-427,11	47,61
-19,1	-37,9	723,89	364,81
-15,1	-34,9	526,99	228,01
5,9	27,1	159,89	34,81
6,9	47,1	324,99	47,61
16,9	57,1	964,99	285,61
24,9	87,1	2168,79	620,01
27,9	112,1	3127,59	778,41
$\sum$xi=0	**$\sum$yi=0**	**$\sum$xiyi=13139,1**	**$\sum x_i^2$ =3990,9**

$a = 0$

$$b = b_{yx} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{13139,1}{3990,9} = 3,29$$

Bu durumda, ortalamalar orijinine göre, Y'nin X'e göre regresyon denklemi,

$\hat{y} = 3,29x$

şeklinde yazılır.

Yukarıdaki açıklamalarda Y değişkeni bağımlı değişken, X değişkeniyse bağımsız değişken kabul edilmişti. X değişkeni bağımlı Y değişkeni bağımsız değişken olduğundaysa doğrusal regresyon denklemi;

$$\hat{X} = a + bY$$

olarak ifade edilecektir. Bu durumda a ve b

$$b = \frac{\sum x_i y_i}{\sum y_i^2}$$

ya da

$$b_{xy} = \frac{\sum x_i y_i}{\sum y_i^2}$$

ve

$$a = \bar{X} - b\bar{Y}$$

formülleri yardımıyla bulunur. b_{xy} ve b_{yx}'in formülleri incelendiğinde, her ikiside daima aynı işareti taşır fakat aynı değerde değildir.

SIRA SİZDE 1

X bağımsız ve Y bağımlı değişken olmak üzere bir araştırmanın gözlemlenen değerleri izleyen tabloda verilmiştir. Verilere ilişkin serpilme diyagramını çiziniz. Serpilme diyagramı üzerindeki noktaların doğrusal eğilim içinde olup olmadıklarını gözleyiniz. b katsayısını ve a katsayısını hesaplayıp regresyon denklemini elde ediniz.

Gözlem	1	2	3	4	5	6	7	8	9	10	11	12
X	12	22	36	34	23	18	18	27	30	38	24	18
Y	10	12	14	16	18	11	17	19	12	18	14	19

Varyansın (σ^2) Tahmini

Basit doğrusal regresyon modelinde α ve β'nın tahminlerine ek olarak, aralık tahminlerinde ve hipotez testlerinde σ^2'nın tahminine gereksinim vardır.

σ^2, ε_i hata terimlerinin ortak varyansıdır. ε_i'in tahmini e_i hata terimi olduğundan e_i'lerin varyansı da σ^2'in bir tahmini olacaktır. Hataların kareler toplamı (HKT),

$$HKT = \sum e_i^2$$

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

yazılabilir. HKT'in serbestlik derecesine bölümüyle elde edilen

$$HKO = \frac{HKT}{n-k}$$

hata kareler ortalaması, (bir başka ifadeyle hataların varyansı), σ^2'in bir tahminidir. Basit doğrusal regresyon modeliyle hataların hesaplanmasında, α ve β'nın tahminleri a ve b kullanıldığından, serbestlik derecesi (n-2) olarak yazılır.

HKO'un kare kökü alındığındaysa denklemin standart hatası elde edilir ve $\hat{\sigma}$ ile gösterilir.

$$\hat{\sigma} = \sqrt{HKO}$$

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

ÖRNEK 3

Örnek 1'deki verileri kullanarak elde edilen regresyon denklemine dayanarak yapılacak tahminlerin standart hatasını hesaplayalım.

$\hat{\sigma}$'ı hesaplayabilmek için öncelikle, $\hat{Y}$ 'ları daha sonra da sırasıyla $(Y_i - \hat{Y}_i)$, $(Y_i - \hat{Y}_i)^2$ ları hesaplayalım.

$\hat{Y} = 20{,}175 + 3{,}29x$

$\hat{Y}_i$	$(Y_i - \hat{Y}_i)$	$(Y_i - \hat{Y}_i)^2$
253,93	-3,93	15,41
290,14	-10,14	102,83
385,62	-84,62	7159,94
300,02	24,98	624,12
313,19	14,81	219,44
382,32	7,68	58,92
385,62	24,38	594,56
418,54	1,46	2,13
444,88	5,12	26,24
454,75	20,25	409,90
		9213,5

buradan,

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

$$\hat{\sigma} = \sqrt{\frac{9213{,}5}{10-2}} = \sqrt{1151{,}69} = 33{,}94$$

olarak elde edilir.

Basit Doğrusal Regresyonda Aralık Tahmini

İstatistiksel çıkarsamalarda yapılan tahminlerin, gerçek değerlerle genellenmesi aralık tahminleriyle yapılır. Regresyon çözümlemesi, örneklem verileriyle yapıldığından, elde edilen a ve b'lerin anakütle parametreleri α ve β'ya ilişkin aralık tahminlerinde kullanılması söz konusudur.

α için güven aralığı;

$$P(a - t_{\alpha}.s_a \leq \alpha \leq a + t_{\alpha}.s_a) = 1 - \alpha$$

şeklinde verilir. s_a, a'nın standart hatasıdır ve

$$s_a = \hat{\sigma}\sqrt{\frac{1}{n} + \frac{\overline{X}^2}{\sum(X_i - \overline{X})^2}}$$

formülüyle hesaplanır, t_{α} ise α anlamlılık düzeyi n-2 serbestlik derecesinde t tablosundan bulunan değerdir.

β için güven aralığı;

$$P(b - t_{\alpha}.s_b \leq \beta \leq a + t_{\alpha}.s_b) = 1 - \alpha$$

$$s_b = \frac{\hat{\sigma}}{\sqrt{\sum(X_i - \overline{X})^2}}$$

şeklinde hesaplanır.

ÖRNEK 4

Örnek 1'de b=3,29 olarak hesaplanmıştır. Bu sonuca göre, β katsayısının %95 güven aralığını hesaplayalım.

b= 3,29

$$s_b = \frac{\hat{\sigma}}{\sqrt{\sum(X_i - \overline{X})^2}} = \frac{33,94}{\sqrt{\sum 3990,9}} = \frac{33,94}{63,17} = 0,54$$

ve ν=8 serbestlik derecesinde t tablo değeri 2,306'dır. Buna göre β için %95 güven aralığı

$$3,29 \pm (2,306)\ (0,54)$$
$$P(3,29 - 1,25 \leq \beta \leq 3,29 + 1,25) = 0,95$$
$$P(2,04 \leq \beta \leq 4,54) = 0,95$$

olarak hesaplanır.

Regresyon katsayısı β'nın 0,95 olasılıkla alabileceği değerler 2,04 ile 4,54 arasında olacaktır.

Regresyon Katsayısının Anlamlılık Testi

Basit doğrusal regresyon analizinde, bir bağımlı bir bağımsız değişken olması nedeniyle, test edilecek parametreler α ve β olacaktır. Daha önce açıklandığı gibi, α'nın tahmini a regresyon sabitidir ve b ise β'nın tahmini olup regresyon katsayısıdır.

Basit doğrusal regresyon modelindeki regresyon katsayısına ilişkin yapılan test, regresyon doğrusunun anlamlılığını da test etmektedir. Şöyle ki:

$H_o : \beta=0$
$H_1 : \beta\neq0$

hipotezleri

$$t_h = \frac{b}{s_b}$$

istatistiğinden yararlanılarak test edilir. α anlam düzeyinde n-2 serbestlik derecesinde t tablosundan bulunan değer hesaplanan t test istatistiğinden büyükse H_o: β=0 hipotezi kabul edilir ve regresyon doğrusu anlamlı değildir. Diğer bir ifadeyle Y'deki değişimler X'deki değişmelerden kaynaklanmamaktadır (X ve Y arasında doğrusl bir ilişki yoktur). t istatistiğinin değeri t tablo değerinden büyükse H_o reddedilir, yani regresyon doğrusu anlamlıdır. Elde edilen doğrusal regresyon modeli amaca uygun olarak kullanılabilir.

ÖRNEK 5

Örnek 1'de elde edilen regresyon katsayısının α=0,05 anlam düzeyinde anlamlılık testini yapalım.

H_o: β=0

H_1: $\beta \neq 0$

Hipotezlerini test edelim $t_h = \frac{b}{s_b}$ örneklem küçük olduğu için benimsenen t, ν= n-2 = 8 serbestlik derecesinde t dağılır.

$$t_h = \frac{b}{s_b} = \frac{3,29}{0,54} = 6,09$$

α=0,05 ve 8 serbestlik derecesi ile $t_{0,05}$= 2,306 olduğundan $t_h > t_{0,05}$'dir ve H_0 reddedilecektir, b katsayısı istatistiksel olarak anlamlıdır.

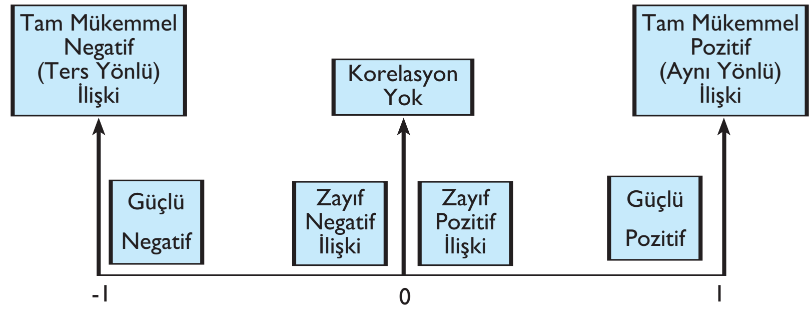
KORELASYON ANALİZİ

Şu ana kadar çalışmalarımız regresyon analizi başlığı altında biri açıklayıcı (bağımsız) diğeri açıklanan (bağımlı) olmak üzere iki değişken arasındaki doğrusal ilişkiye ait denklem üzerinde yoğunlaşmıştır. Bundan sonraki kesimde, iki değişken arasındaki doğrusal ilişkinin derecesi incelenecektir.

Basit Doğrusal Korelasyon Katsayısı

İki nicel değişken arasındaki ilişkinin derecesini ve yönünü belirlemek için r ile gösterilen *Pearson korelasyon katsayısı* veya *basit korelasyon katsayısı* hesaplanır. Bu katsayı, $-1 \leq r \leq 1$ aralığında değer alır. Pozitif korelasyon katsayısı değişkenlerden birinin değeri arttığında diğerinin de değerinin arttığını; negatif korelasyon katsayısı ise değişkenlerden birinin değeri artarken diğerinin değerinin azaldığını belirtir. $r = \pm 1$ olduğunda, söz konusu iki değişken mükemmel/tam ilişki içindedir. Buna karşılık $r = 0$ olması iki değişkenin hiçbir ilişki içinde olmadıklarını gösterir. r -0,50 ya da +0,50 etrafında bir değer aldığında ise değişkenler arasında orta düzeyli bir ilişkinin varlığından söz edilir. Korelasyon en az iki değişken için tanımlansa da, bir değişkenin kendisi ile korelasyonu +1'dir. Korelasyon katsayısı r'nin belirttiği ilişkinin derece ve yönü Şekil 5.3'te özetlenmektedir.

Şekil 5.3



Pearson Korelasyon Katsayısının İlişki Yön ve Dereceleri

Basit doğrusal korelasyon katsayısı iki değişken arasındaki doğrusal ilişkinin derecesini ve yönünü belirler.

X ve Y gibi iki değişken arasındaki **basit doğrusal korelasyon katsayısı** farklı formüller yardımıyla hesaplanabilir. Bunlardan birincisi S_x, S_y sırasıyla *X* ve *Y* değişkenlerinin standart sapması olmak üzere, *X* ve *Y* değişkenleri arasındaki. Pearson

korelasyon katsayısının $r = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{(n-1)s_x s_y}$ olarak hesaplanmasıdır. İkinci olarak

doğrudan iki değişkene ilişkin verilerden hareketle $r = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2 \sum_{i=1}^{n}(Y_i - \overline{Y})^2}}$

formülüyle hesaplanabilir.

Korelasyon katsayısı hesaplarken yararlanabileceğimiz bir diğer formül ise şu şekildedir:

$$r = \pm\sqrt{b_{yx} b_{xy}}$$

Formüldeki;

b_{yx} = Y'nin X'e göre regresyon katsayısı

b_{xy} = X'in Y'ye göre regresyon katsayısıdır.

Burada dikkat edilmesi gereken nokta regresyon katsayıları pozitifse korelasyon katsayısı pozitif, eğer her iki regresyon katsayısının işareti negatifse korelasyon katsayısı da negatif olacaktır.

ÖRNEK 6

Örnek 1'deki X ve Y değişkenlerine ilişkin Pearson korelasyon katsayısını standart sapma değerlerini kullanarak bulunuz.

X ve Y değişkenlerinin standart sapmalarını kullanarak korelasyon katsayısının hesaplanabilmesi için izleyen tabloda verildiği gibi sırasıyla $(X_i - \overline{X})$, $(Y_i - \overline{Y})$ ve $(X_i - \overline{X})(Y_i - \overline{Y})$'nın hesaplanması gerekir.

Müşteri	Ciro
X_i	Y_i
71	250
82	280
111	301
85	325
89	328
110	390
111	410
121	420
129	450
132	475
21,058	76,3551

$X_i-\bar{X}$	$Y_i-\bar{Y}$	$(X_i-\bar{X})(Y_i-\bar{Y})$
-33,1	-112,9	3736,99
-22,1	-82,9	1832,09
6,9	-61,9	-427,11
-19,1	-37,9	723,89
-15,1	-34,9	526,99
5,9	27,1	159,89
6,9	47,1	324,99
16,9	57,1	964,99
24,9	87,1	2168,79
27,9	112,1	3127,59
		13139,10

$n = 10$,

$\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y}) = 13139{,}10$ olarak hesaplanmıştır. Korelasyon katsayısı için gerekli olan X ve Y değişkenlerinin standart sapmaları $s_x = 21{,}058$ ve $s_y = 76{,}355$'dır. Bu değerler formülde yerine konularak bulunur. Hesaplanan bu korelasyon katsayısına göre; iki değişken arasında güçlü aynı yönlü (pozitif) ilişki olduğu söylenebilir.

SIRA SİZDE 2

Örnek 1'deki X ve Y değişkenlerine ilişkin Pearson korelasyon katsayısını aşağıdaki eşitliği kullanarak hesaplayınız.

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

Belirlilik Katsayısı

Değişkenler arasındaki ilişkinin derecesi korelasyon katsayısı yardımıyla belirlenirken bu katsayının değerine bağlı olarak zayıf, orta ve güçlü ilişki gibi nitelemeler söz konusu olur. Bağımlı değişkende meydana gelen değişmelerin ne kadarının bağımsız değişkendeki değişmelerle açıklanabileceğini belirlemek amacıyla **belirlilik katsayısı** kulanılır. Belirlilik katsayısı Pearson korelasyon katsayısı r'nin karesinin alınması ile hesaplanır ve 0 ile 1 arasında değerler alır, r^2 sembolü ile gösterilir.

Belirlilik katsayısı, Bağımlı değişkende meydana gelen değişmelerin ne kadarının bağımsız değişkendeki değişmelerle açıklanabileceğini gösterir.

$0 \leq r^2 \leq 1$

$r^2 = 1$ ise Y'deki değişimin %100'ünün X bağımsız değişkeni tarafından açıklanabildiği kabul edilir.

$r^2 = 0$ ise X bağımsız değişkeni, Y bağımlı değişkenini hiç açıklayamıyor demektir.

ÖRNEK 7

Örnek 1'de verilen günlük elde edilen ciro değişkenin ne kadarlık bir bölümünün günlük müşteri sayısı ile açıklanabileceğini bulunuz.

Bu sorunun cevaplanması için korelasyon katsayı değerine ihtiyaç bulunmaktadır. Bu değerde önceki örnekten r=0,908 olarak bulunmuştu. Bu sayının karesi belirlilik katsayısı olacağından $(0{,}908)^2=0{,}82$ olarak hesaplanır. Söz konusu mağazanın günlük cirosuna, günlük müşteri sayısının etkisinin %82 olduğu söylenir. Geriye kalan %18 ise müşteri sayısına bağlı olamayan kısımdır.

Korelasyon Katsayısı Anlamlılık Testi

Evren korelasyon katsayısı ρ (ro) ile sembolize edilir. Evren korelasyon katsayısı $\rho=0$ olan bir ana kütleden seçilen örneklemlerin r katsayıları normal dağılıma sahiptir. Bu nedenle,

H_0: $\rho = 0$

H_1: $\rho \neq 0$

hipotezleri test edilir.

Hipotezlerin kurulmasının ardından örneklem korelasyon katsayısı yardımıyla $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ istatistiği hesaplanır.

Daha sonra belirlenen anlam düzeyi ve n-2 serbestlik derecesine göre t tablosundan kritik değer belirlenir. Belirlenen bu kritik değer yardımıyla (-kritik değer, +kritik değer) aralığı tanımlanır. Hesaplanan t değeri, söz konusu bu aralıkta yer aldığında H_0 hipotezi kabul, bu aralıkta yer almadığında ise reddedilir. H_0'ın reddedilmesinin anlamı hesaplanmış korelasyon katsayısının istatistiksel olarak anlamlı olduğudur.

ÖRNEK 8

Örnek 1'de verilen probleme ilişkin olarak ana kütle korelasyon katsayısının değeri 0'a eşit midir? %5 anlam düzeyine göre test ediniz.

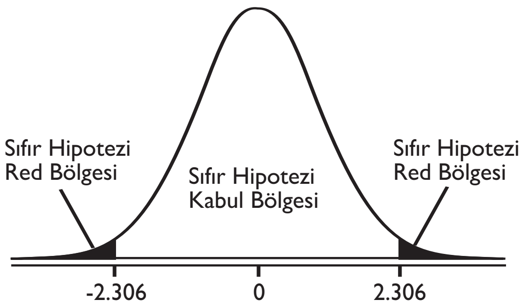
İlk olarak

H_0: $\rho = 0$ (Evren korelasyonu 0'dır.)

H_1: $\rho \neq 0$ (Evren korelasyonu 0 değildir.)

hipotezleri yazılır.

Daha sonra tablodan gerekli kritik değerler elde edilir. t dağılımı tablosu yardımıyla %5 anlam düzeyi ve n-2=10-2=8 serbestlik derecesi için elde edilen kritik değerler ile belirlenen H_0 hipotezi red bölgeleri Şekil 5.4'teki gibidir.



Şekil 5.4

%5 Anlam Düzeyi ve 8 Serbestlik Derecesi ile Korelasyon Katsayısı Anlamlılık Testinde H_0'ın Kabul ve Red Bölgeleri

Sonunda test için gerekli olan gözlemlenen t istatistiği değeri hesaplanır.

Bu değer $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \frac{0,908\sqrt{8}}{\sqrt{1-0,908^2}} = 6,129$ olarak bulunur. Hesaplanan bu 6,129 değeri Şekil 5.4'te gösterilen kritik değerler aralığında, yani H_0'ın kabul bölgesinde yer almadığından "İlgilenilen değişkenler arasındaki korelasyon sıfırdır." hipotezi reddedilir. Değişkenler arasında ana kütle düzeyinde de ilişki bulunmaktadır.

SIRA SİZDE 3

10 kişilik bir sınıftaki öğrencilerin istatistik ve matematik derslerindeki başarı puanları tabloda verilmiştir. Bu sınıftaki öğrencilerin istatistik ve matematik derslerindeki başarı seviyeleri arasında bir ilişki olabileceği düşünülmektedir. Bu ilişkinin belirlenmesi için, öğrencilerin izleyen tabloda verilen istatistik ve matematik sınavları başarı puanları incelenecektir:

Öğrenci	1	2	3	4	5	6	7	8	9	10
İstatistik Başarı Puanı	75	80	95	60	40	70	65	25	90	85
Matematik Başarı Puanı	85	80	90	50	40	55	50	20	95	80

a. İstatistik ve matematik dersleri başarı puanlarıyla ilgili korelasyon katsayısını hesaplayınız.
b. İstatistik dersi başarı puanının ne kadarlık kısmının matematik dersi başarı puanı tarafından açıklanabileceğini hesaplayınız.
c. Korelasyon katsayısının istatistiksel olarak anlamlı olup olmadığını %5 anlamlılık düzeyine göre test ediniz.

Özet

AMAÇ 1

İki değişken arasındaki ilişkiyi açıklayan doğrusal modeli kurmak

Regresyon analizi ilişki içinde bulunan değişkenler arasındaki ilişkinin doğasını belirlemek ve bu ilişkiyi kullanarak o konu ile ilgili tahminler (estimation) ya da kestirimler (prediction) yapabilmek amacıyla kullanılan istatistiksel bir tekniktir. Regresyon analizinde bağımlı değişken üzerinde oluşan değişimlerin açıklanmasına çalışılır. Regresyon analizinin yalnızca bir açıklayıcı değişkenle yapılması düşünülemez. Gerçekten de regresyon analizi birden fazla bağımsız değişken üzerinde de yapılabilir. Bir bağımsız değişken olması durumunda *basit doğrusal regresyon analizi;* birden fazla bağımsız değişken olması durumunda *çoklu doğrusal regresyon analizi* söz konusu olmaktadır. X, Y değişkenleri arasındaki doğrusal ilişkinin ifadesinde kullanılan eşitliğe doğrusal regresyon modeli ya da kısaca regresyon denklemi denir. Doğrusal regresyon denkleminin tahmini için kullanılabilen farklı teknikler arasında en çok bilineni en küçük kareler tekniğidir. Basit doğrusal regresyon modeli,

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

olarak yazılır. En küçük kareler tekniğinde, basit doğrusal regresyon modelinde yer alan ana kütle parametreleri α ve β için eldeki veri yardımıyla

$$a = \frac{\sum Y_i - b\sum X_i}{n} \quad \text{ve} \quad b = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - \left(\sum X_i\right)^2}$$

tahminleri hesaplanır. Böylelikle bulunan a ve b parametre tahminleri doğru denkleminde yerlerine konarak basit doğrusal regresyon denklemi

$\hat{Y} = a + bX$ olarak yazılır.

AMAÇ 2

İki değişken arasındaki ilişkinin derecesini belirlemek

İki ya da daha fazla ve en az aralıklı ölçeğe uygun şekilde ölçümlenmiş değişkenler arasındaki ilişkinin derecesini belirlemek için r ile gösterilen Pearson korelasyon katsayısı ile hesaplanır. Korelasyon katsayısı, $-1 \leq r \leq 1$ aralığında değer alır. Pozitif korelasyon katsayısı değişkenlerden birinin değeri arttığında diğerinin de değerinin arttığını; negatif korelasyon katsayısı ise değişkenlerden birinin değeri artarken diğerinin değerinin azaldığını belirtir. $r = \pm 1$ olduğunda, söz konusu iki değişken mükemmel/tam ilişki içindedir. Buna karşılık $r = 0$ olması iki değişkenin hiçbir ilişki içinde olmadıklarını gösterir.

Kendimizi Sınayalım

1. Aşağıdakilerden hangisi değişkenler arasındaki ilişkinin yapısı hakkında genel bir bakışa imkan sağlar?

a. Histogram
b. Serpilme diyagramı
c. Frekans poligonu
d. Normal eğri
e. Pasta grafik

2. Hangi yaklaşımda bağımlı ve bağımsız değişkenler arasındaki doğrusal ilişkinin matematiksel fonksiyonu ortaya konmaya çalışılır?

a. Hipotez testi
b. Aralık Tahmini
c. Regresyon analizi
d. Korelasyon analizi
e. Varyans analizi

3. Bağımlı değişkenle (Y_i) gözlem değerleriyle regresyon denkleminden hesaplanan ($\hat{Y}_i$) değeri arasındaki farka ne ad verilir?

a. Hata Terimi
b. Varyans
c. Korelasyon
d. Katsayı
e. Parametre

4. $\hat{Y} = 2 + 5X$ Regresyon denkleminde (b) regresyon katsayısı aşağıdakilerden hangisinde doğru olarak ifade edilmiştir?

a. X'deki bir birimlik değişim Y'de 5 birimlik değişime neden olur.
b. X'deki bir birimlik artış Y'de 1 birimlik artışa neden olur.
c. X'de 1 birimlik azalma Y'de 1 birimlik azalışa neden olur.
d. X'de bir birimlik değişim Y'de 2 birimlik değişime neden olur.
e. X'in değerindeki değişimin Y'nin değerine etkisi yoktur.

Aşağıdaki tabloyu inceleyiniz. 5-6 ve 7. Sorular aşağıdaki bilgilere göre cevaplandırılacaktır.

X_i	Y_i	$\hat{Y}_i = a + bX_i$
3	12	12
5	16	16
11	28	28
7	20	20
1	8	8

5. Yukarıdaki tablodaki verilerden oluşturulan regresyon modelinde a = 6 olduğunda, regresyon doğrusunun eğimi aşağıdakilerden hangisidir?

a. -1
b. 0
c. 1
d. 2
e. 4

6. Tablo verilerine göre HKT değeri nedir?

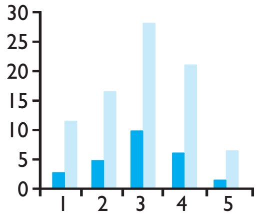
a. 4,6
b. 16,8
c. 0
d. -16,8
e. -4,6

7. Tablo verilerine göre tahminlerin standart hatası(σ) nedir?

a. 3,2
b. 6,4
c. 3
d. 6
e. 0

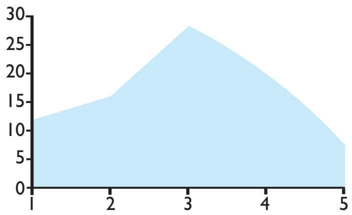
8. Aşağıdakilerden hangisi serpilme diyagramı örneğidir?

a.

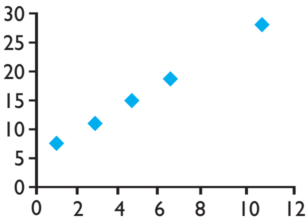


b.

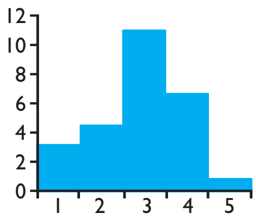
c.



d.



e.



9. Bir araştırmada X ve Y değişkenleri için gözlem değerleri aşağıdaki tabloda verilmiştir.

X	Y
1	2
3	4
4	6
7	8
9	10
12	12
45	14
50	16
80	18
50	20

X ve Y değişkenleri arasındaki korelasyon katsayısı nedir?

a. -0,92
b. -0,45
c. 0,02
d. 0,55
e. 0,88

10. 9. Sorudaki gözlem verilerinden hareketle hesaplanacak belirlilik katsayısı nedir?

a. 0,81
b. 0,21
c. 0.12
d. 0,42
e. 0,77

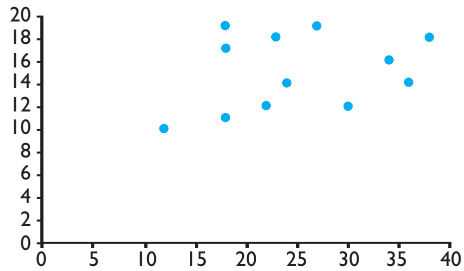
Kendimizi Sınayalım Yanıt Anahtarı

1. b	Yanıtınız yanlış ise "Regresyon Analizi" konusunu yeniden gözden geçiriniz.
2. c	Yanıtınız yanlış ise "Regresyon Analizi" konusunu yeniden gözden geçiriniz.
3. a	Yanıtınız yanlış ise "Basit Doğrusal Regresyon Analizi" konusunu yeniden gözden geçiriniz.
4. a	Yanıtınız yanlış ise "Basit Doğrusal Regresyon Analizi" konusunu yeniden gözden geçiriniz.
5. d	Yanıtınız yanlış ise "Basit Doğrusal Regresyon Analizi" konusunu yeniden gözden geçiriniz.
6. c	Yanıtınız yanlış ise "Varyansın (σ_2) Tahmini" konusunu yeniden gözden geçiriniz.
7. e	Yanıtınız yanlış ise "Basit Doğrusal Regresyon Analizi" konusunu yeniden gözden geçiriniz.
8. d	Yanıtınız yanlış ise "Basit Doğrusal Regresyon Analizi" konusunu yeniden gözden geçiriniz.
9. e	Yanıtınız yanlış ise "Korelasyon Analizi" konusunu yeniden gözden geçiriniz.
10. e	Yanıtınız yanlış ise "Korelasyon Analizi" konusunu yeniden gözden geçiriniz.

Sıra Sizde Yanıt Anahtarı

Sıra Sizde 1

Verilere ilişkin serpilme diyagramı çizildiğinde aşağıda verilen grafik elde edilir.



Serpilme diyagramındaki noktaların doğrusal eğilim içinde oldukları gözlenmektedir. Önerilen eşitlik yardımıyla b katsayısının hesaplanabilmesi için öncelikle x ve y değişkenlerinin ortalamaları bulunmalıdır.

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{12+22+\ldots+18}{12} = \frac{300}{12} = 25$$

$$\bar{Y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{10+12+\ldots+19}{12} = \frac{180}{12} = 15$$

Ardından ara hesaplamalar için gereken aşağıdaki tablo oluşturulur. Tablonun altındaki sayıların ilgili sütunların toplamları olduğu unutulmamalıdır. Tablodan elde edilen değerlerle

$$b = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{85}{730} = 0{,}116$$

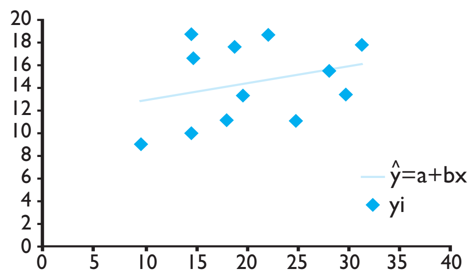
olarak bulunur. Daha sonra bu b değeri kullanılarak

$$a = \frac{\sum y_i - b\sum x_i}{n} = \frac{180-(0{,}116).(300)}{12} = 12{,}089$$

olarak elde edilir.

X_i	Y_i	$X_i-\bar{X}$	$Y_i-\bar{Y}$	$(X_i-\bar{X})(Y_i-\bar{Y})$	$(X_i-\bar{X})^2$
12	10	-13	-170	2210	169
22	12	-3	-168	504	9
36	14	11	-166	-1826	121
34	16	9	-164	-1476	81
23	18	-2	-162	324	4
18	11	-7	-169	1183	49
18	17	-7	-163	1141	49
27	19	2	-161	-322	4
30	12	5	-168	-840	25
38	18	13	-162	-2106	169
24	14	-1	-166	166	1
18	19	-7	-161	1127	49
300	180			85	730

$\hat{Y} = a + bX = 12{,}089 + 0{,}116X$ biçiminde regresyon denklemi elde edilir. Bu doğru serpilme diyagramı üzerinde gösterildiğinde ise izleyen grafik elde edilir.



Sıra Sizde 2

Eşitliğin sağ tarafındaki değerlerin elde edilmesi amacıyla izleyen tablo oluşturulur.

X_i	Y_i	$X_i-\bar{X}$	$Y_i-\bar{Y}$	$(X_i-\bar{X})(Y_i-\bar{Y})$	$(X_i-\bar{X})$	$(Y_i-\bar{Y})^2$
71	250	-33,1	-112,9	3736,99	1095,61	12746,41
82	280	-22,1	-82,9	1832,09	488,41	6872,41
111	301	6,9	-61,9	-427,11	47,61	3831,61
85	325	-19,1	-37,9	723,89	364,81	1436,41
89	328	-15,1	-34,9	526,99	228,01	1218,01
110	390	5,9	27,1	159,89	34,81	734,41
111	410	6,9	47,1	324,99	47,61	2218,41
121	420	16,9	57,1	964,99	285,61	3260,41
129	450	24,9	87,1	2168,79	620,01	7586,41
132	475	27,9	112,1	3127,59	778,41	12566,41
				13139,10	3990,90	52470,90

İlgili değerler eşitlikte yerlerine konarak

$$r = \frac{13139,10}{\sqrt{(3990,90).(52470,90)}} = 0,908$$

olarak bulunur.

Sıra Sizde 3

a. Pearson korelasyon katsayısının hesaplanması için kullanılacak

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

formülüne uygun olarak, gerekli ara değerleri içeren bir tablo oluşturulur.

X_i	Y_i	$X_i-\bar{X}$	$Y_i-\bar{Y}$	$(X_i-\bar{X})(Y_i-\bar{Y})$	$(X_i-\bar{X})^2$	$(Y_i-\bar{Y})^2$
75	85	6,5	20,5	133,25	42,25	420,25
80	80	11,5	15,5	178,25	132,25	240,25
95	90	26,5	25,5	675,75	702,25	650,25
60	50	-8,5	-14,5	123,25	72,25	210,25
40	40	-28,5	-24,5	698,25	812,25	600,25
70	55	1,5	-9,5	-14,25	2,25	90,25
65	50	-3,5	-14,5	50,75	12,25	210,25
25	20	-43,5	-44,5	1935,75	1892,25	1980,25
90	95	21,5	30,5	655,75	462,25	930,25
85	80	16,5	15,5	255,75	272,25	240,25
				4692,5	4402,5	5572,5

$$r = \frac{4692,5}{\sqrt{4402,5x5572,5}} = \frac{46925}{4953,073} = 0,947$$

Korelasyon değerine göre; iki değişken arasında aynı yönlü güçlü korelasyon olduğu söylenebilir.

b. Belirlilik katsayısı korelasyon katsayısı değerinin karesi olan (0,947)2=0,897 olacaktır. Dolayısıyla bu sınıftaki öğrencilerin istatistik dersinden elde ettikleri başarı puanlarına matematik dersinden elde ettikleri başarı puanlarının etkisinin %89,7 olduğu söylenebilir. Bu, oldukça büyük bir değer olduğundan, bu sınıftaki 10 öğrencinin matematik dersindeki başarılarının istatistik dersindeki başarılarına olumlu (pozitif) yönde güçlü bir etkisi olduğu söylenebilir.

c. H_0: $\rho = 0$

H_1: $\rho \neq 0$

hipotezleri kurulur.

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0,947\sqrt{10-2}}{\sqrt{1-0,897}} = \frac{2,6785}{0,3209} = 8,3468$$

bulunur.

8,3468 olarak hesaplanan t değeri Şekil 5.5'teki sıfır hipotezi red bölgesinde kaldığından, "İlgilenilen değişkenler arasındaki korelasyon sıfırdır" hipotezi reddedilir. Başka bir ifadeyle iki değişken arasında istatistiksel olarak anlamlı bir ilişki söz konusudur.

Yararlanılan ve Başvurulabilecek Kaynaklar

Cornillon, Pierre-André ve Matzner-Løber, Éric (2007), **Regression Theorie et Applications,** Springer-Verlag France, Paris.

Çömlekçi, Necla (1994), **Temel İstatistik İlke ve Teknikleri,** Bilim Teknik Yayınevi Yayını, Eskişehir.

Durucasu, Hasan (2003), **Excel ile Matris Uygulamaları,** Birlik Ofset Yayını, Eskişehir.

Şıklar, Emel (2000), **Regresyon Analizine Giriş,** Anadolu Üniversitesi Yayını, Eskişehir.

Yüzer, Ali Fuat (1996), **Olasılık ve İstatistik,** Anadolu Üniversitesi Yayını, Eskişehir.