

# TEMEL İSTATİSTİK

Regresyon II

Prof. DR. Ezel Tavşancıl

# ÖRNEK II ÇÖZÜMÜ

No	Seans (X)	Depresyon Düzeyi (Y)	x <sup>2</sup>	xy
1	33	10.8	1089	356.4
2	33	9.5	1089	313.5
3	23	14.2	529	326.6
4	34	9.7	1156	329.8
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
17	27	12.8	729	345.6
18	29	11.0	841	319.0
19	24	13.5	576	324.0
20	31	10.8	961	334.8
<b>Toplam</b>	<b>581</b>	<b>236.6</b>	<b>17215</b>	<b>6761.6</b>

$$b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} = \frac{20 * 6761.6 - 581 * 236.6}{20 * 17215 - (581)^2} = -0.331$$

$$a = \bar{y} - b\bar{x} = \frac{236.6}{20} - (-0.331) \frac{581}{20} = 21.4$$

$$\hat{y} = 21.4 - 0.331x$$

Depresyon düzeyindeki değişimin %85'nin terapi seansı ile açıklanabileceği hesaplanmıştır. ( $r^2 = 0.85$ ).

## ÖRNEK III

- Bir psikolojik danışman, lise son sınıf öğrencisi olan 8 danışanına okula yönelik ilgi ve çalışma motivasyonu ölçeği uyguluyor. Uyguladığı ölçeklerden aldığı puanlar aşağıdaki tabloda yer almaktadır. Buna göre danışanlarının okula yönelik tutumları ile çalışma motivasyonları arasında anlamlı bir ilişki bulunmakta mıdır? Regresyon modelini kurarak yorumlayınız.

Çalışma Motivasyonu	Okula Yönelik Tutum	$X^2$	x.y
105	65	11025	6825
110	62	12100	6820
140	69	19600	9660
95	61	9025	5795
125	65	15625	8125
120	67	14400	8040
135	65	18225	8775
130	67	16900	8710
Toplam: 960	521	116900	62750

$r^2$   
: 0.64

- $b = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}$

$$b = 1840 / 13600 = 0.13529412$$

$$b = 0.14$$

- a:  $y - bx$

$$a = (65.13) - (0.14)120 = 48,3$$

$$a = 48.3$$

$$Y = 48.3 + 0.14X$$

r:

- $r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$

- $r = 1846 / 2306 = 0.80$

$$r = 0.80$$

$$r^2 = 0.64$$

- Çalışma motivasyonu ile okula yönelik tutum arasında pozitif yönlü bir ilişki vardır. Çalışma motivasyonu '0' olan birinin okula yönelik tutumu 48.13 olur. Çalışma motivasyonu ve okula yönelik tutum arasındaki ilişki ( $r=0.80$ ), determinasyon katsayısı 0.64'tür. İki değişken arasında pozitif ve yüksek düzeyde bir ilişki vardır. Çalışma motivasyonu, okula yönelik tutum puanlarındaki varyansın %64'ünü açıklar.

# ÖRNEK IV...

X	Y
1	7
2	8
3	9
4	8
4	9
5	11
5	10
6	13
6	14
7	13



X	Y	XY	X <sup>2</sup>
1	7	7	1
2	8	16	4
3	9	27	9
4	8	32	16
4	9	36	16
5	11	55	25
5	10	50	25
6	13	78	36
6	14	84	36
7	13	91	49
43	102	476	217



$$\hat{b} = \frac{10(476) - (43)(102)}{10(217) - (43)^2}$$

$$\hat{b} = \frac{4760 - 4386}{2170 - 1849}$$

$$\hat{b} = \frac{374}{321} = 1.165$$

## ...ÖRNEK IV

X	Y
1	7
2	8
3	9
4	8
4	9
5	11
5	10
6	13
6	14
7	13
43	102
4,3	10,2

- $a = \hat{y} - bx$
- $a = 10.2 - 1.165 (4.3)$
- $a = 5.19$

# $r^2$ Nasıl Hesaplanır?

- $r^2$  (açılanan varyans) değişkenlerin birindeki değişimin ne kadarının diğer değişkenler tarafından açıklandığını yüzde olarak verir.
- Determinasyon katsayısı olarak da isimlendirilir.

$$r^2 = \frac{\sum (\hat{Y} - \bar{Y})^2}{\sum (Y - \bar{Y})^2} \quad bx$$

# Dip Not

## Regresyon-korelasyon ilişkisi

- $r_{XY}^2 = b_{XY} * b_{YX}$
- X'in bağımlı, Y'nin bağımsız olduğu durumdaki regresyon katsayısı ile tam tersi durumdaki regresyon katsayıları çarpımı iki değişken arasındaki ilişkinin karesini verir.
- Bu durumda regresyon katsayısının işareti ile korelasyon katsayısının işareti de aynı olur.



# Tahminî Y Değeri Hesaplama

$$X_1 = 5.195 + 1.165 * 1 = 6.36$$

$$X_2 = 5.195 + 1.165 * 2 = 7.53$$

$$X_3 = 5.195 + 1.165 * 3 = 8.69$$

$$X_4 = 5.195 + 1.165 * 4 = 9.86$$

$$X_5 = 5.195 + 1.165 * 5 = 11.02$$

$$X_6 = 5.195 + 1.165 * 6 = 12.19$$

$$X_7 = 5.195 + 1.165 * 7 = 13.35$$

# Tahminin Standart Hatası

Tahminin doğruluk derecesini bulmada kullanılan ölçü

- Aralarında ilişki bulunan iki değişkenden birinden diğerini tahmin etmede, mükemmel ilişki dışında, tahminin doğruluk derecesi tartışılır **ÇÜNKÜ** yapılan kestirimler ile gerçek değerler arasında fark olacaktır. Bu farkın ortalama değerinin bilinmesi, tahminin doğruluğunu açıklamada kullanılır.
- «Bir değişkenin belli bir değeri için diğer değişkenin değeri tahmin edildiğinde, bu tahmine ne derece güvenebiliriz ya da tahminin evrende alabileceği değer hangi aralıkta olabilir?»

- Gerçekteki Karşılığı: **gözlenen ve tahmin edilen değerlerin arasındaki fark puanlarının standart sapmasıdır.**
- Pratikte Karşılığı: Fark Puanlarının kareleri toplamı (hata kareleri toplamı), serbestlik derecesine bölünerek (a ve b katsayıları için  $sd=n-2$ ) karekökü alınır.

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

- Fark puanlarına dayalı standart hata tan sayılı küçük gruplar için kolay olmakla beraber büyük gruplar için korelasyona dayalı şu formül kullanılır.

$$S_{YX} = S_y \sqrt{\frac{n-1}{n-2} (1-r^2)}$$

$S_y$ : Y puanlarına ait standart sapma

$\frac{n-1}{n-2}$  değeri büyük örneklerde etkisini yitirir.

n=100 için  $\frac{n-1}{n-2}=1.01$

n=30 için  $\frac{n-1}{n-2}=1.035$

n= 15 için  $\frac{n-1}{n-2}=1.076$

n=10 için  $\frac{n-1}{n-2}=1.25$

Bu nedenle büyük veri setlerinde kesir yok sayılarak sadece korelasyona dayalı değer hesaplanır.

$$S_{YX} = S_y \sqrt{1-r^2}$$

# Tahminin Standart Hatasını Hesaplamada İki Formülün Uygulanması

Y	$\hat{Y}$	$(Y-\hat{Y})$	$(Y-\hat{Y})^2$
45	42,01	2,99	8,9401
40	30,69	9,31	86,6761
28	36,35	-8,35	69,7225
55	42,01	12,99	168,74
62	58,99	3,01	9,0601
45	47,67	-2,67	7,1289
40	58,99	-18,99	360,62
80	75,97	4,03	16,2409
90	87,29	2,71	7,3441
20	25,03	-5,03	25,3009
			759,774

$$S_{YX} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n - 2}} \quad S_{YX} = \sqrt{\frac{759,774}{10 - 2}}$$

$$S_{YX} = \sqrt{\frac{759,774}{10 - 2}} \quad S_{YX} = \sqrt{94,97}$$

$$S_{YX} = 9,745$$

Bağımlı değişkenin regresyon eğrisi etrafındaki değişkenliğinin ölçüsü

X	Y	$\hat{Y}$	$Y-\hat{Y}$	$(Y-\hat{Y})^2$
1	7	6,36	0,64	0,4096
2	8	7,53	0,47	0,2209
3	9	8,69	0,31	0,0961
4	8	9,86	-1,86	3,4596
4	9	9,86	-0,86	0,7396
5	11	11,02	-0,02	0,0004
5	10	11,02	-1,02	1,0404
6	13	12,19	0,81	0,6561
6	14	12,19	1,81	3,2761
7	13	13,35	-0,35	0,1225
				10,0213

$$S_{YX} = S_y \sqrt{1-r^2}$$

$$S_{YX} = 2,44\sqrt{1-0,81}$$

$$S_{YX} = 2,44\sqrt{0,19}$$

$$S_{YX} = 2,44(0,435) = 1,061$$

- Tahminin standart hatası, iki formülde farklı sonuçlar verdi. Örneklemin küçük olduğu dikkate alınarak hesaplamada örneklem büyüklüğünü işleme katarak 9.75 değerini veren eşitlik yorumlanmalı.
- Bu değer, bir birimlik standart hataya karşılık ortalama sapma miktarıdır.
- Sapma miktarı, tahmin edilen değer her iki tarafı için de geçerlidir (hem sağ hem de solu). Bu nedenle, nokta tahmini olarak regresyon eşitliği bulunacaksa bunu belli güven düzeyleri ya da hata olasılıkları için oluşturabilir ev sonuçlarını normal dağılımda alan ilişkilerinden yararlanarak yorumlanabilir.

Örneğimizde,  $X=4$  için  $\widehat{Y}=3.6$  bulunmuştu. Aynı örnek için  $X$ 'den  $Y$ 'nin tahmin edilmesinde standart hata (tahminin birim standart sapması) 1.48 bulundu. Normal dağılımda  $\mp$  bir birim sapma aralığında kalan alan toplam alanın %68.26'sına karşılıktır. Buna göre %68.26 güvenle (olasılıkla) çalışma saati  $X=4$  olan birinin başarı puanının,  $3.6 \mp 1.48$  aralığında kalacağı söylenebilir. Buna göre  $X=4$  için tahmin edilen  $Y$  değerinin gerçekte %68.26 olasılıkla 2.12 ile 5.08 arasında olacağı ( $2.12 < \widehat{Y} < 5.08$ ) ya da bir başka anlatımla 2.12'den küçük ve 5.08'den büyük bir değer alamayacağı ifade edilebilir. Güven düzeyi %95 alındığında tahminin güven aralığını bulmak için ilk olarak tahminin standart hatası 1.96 ile çarpılarak tahmine ilişkin ortalama sapma miktarı bulunur. Bulunan sapma miktarı tahmin edilen değere eklenip çıkartılarak anılan güven aralığının alt ve üst sınır değerleri elde edilir.



# KAYNAKLAR

Büyüköztürk, Ş., Çokluk, Ö. ve Köklü N. (2013). *Sosyal bilimler için istatistik*. Ankara: Pegem Akademi.

Tan, Ş. (2016). *SPSS ve excel uygulamalı temel istatistik I*. Ankara: Pegem Akademi.