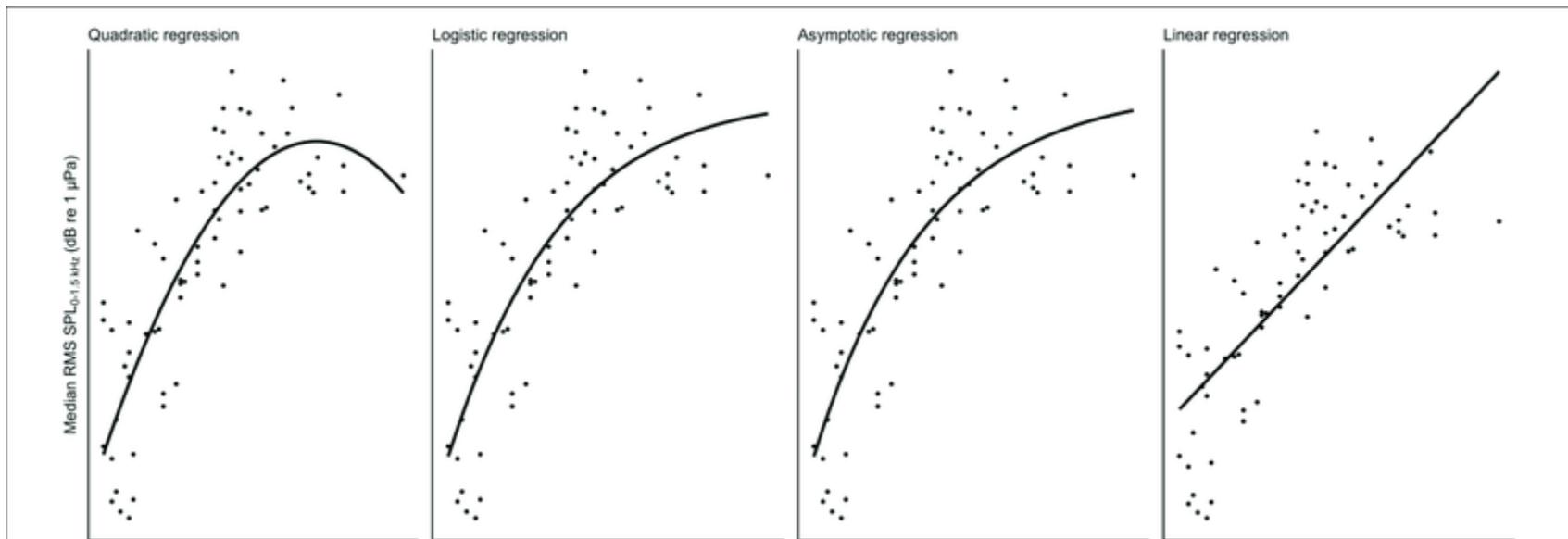


Örüntü Tanıma

Dr. Öğr. Üyesi Mehmet Zahid YILDIRIM

e-mail: m.zahidyildirim@karabuk.edu.tr

Regresyon ve Doğrusal Regresyon Analizi



Regresyon Analizi Nedir?

Regresyon analizi, aralarında neden-sonuç ilişkisi bulunan, bağımsız ve bağımlı değişkenler olarak kabul edilen iki veya daha fazla değişkenden bağımsız olan(lar) bilindiğinde bağımlı değişkenin alabileceği değeri tahmin etmek için yapılan bir analizdir.

Yani regresyon analizi ile bir veya birden çok bağımsız değişkenin, bir bağımlı değişken üzerindeki etkisi açıklanmaya çalışılmaktadır.

Ayrıca regresyon analizi ile değişkenler arasındaki ilişkiye bağlı bir matematiksel model yani denklem kurulmaya çalışılmaktadır.

$$\hat{y} = a + bx$$



Örneğin,

olarak verilen ve aralarında pozitif bir korelasyon ($+0,73$) bulunan x ve y değişkenleri için,

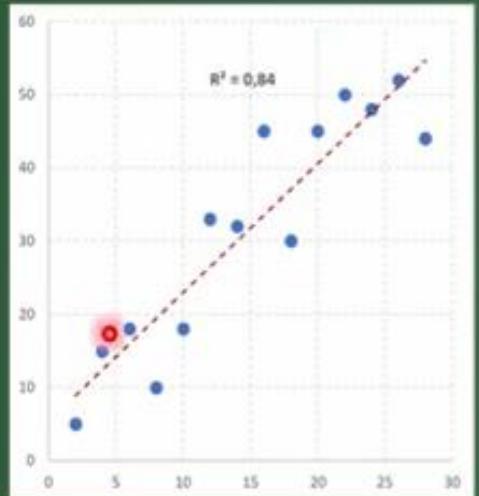
x	y
1	1
2	3
4	4
5	3

x	y
3	?
7	?

Regresyon analizinin temel amacı;

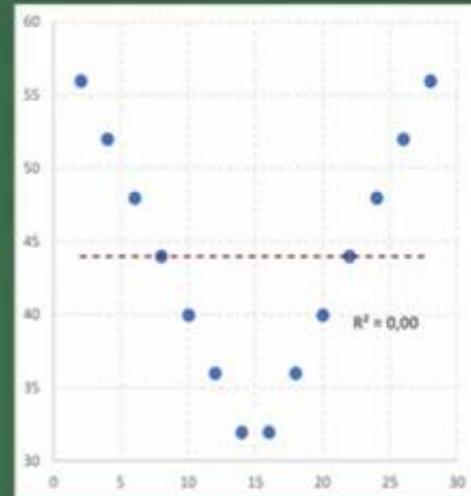
- bağımsız değişkenden bağımlı değişkeni tahmin etmek,
- bir model oluşturmak,
- ilişkiye dayalı olarak ileriye yönelik tahminde bulunmaktır.

Regresyon analizi değişkenler arasındaki ilişkinin doğrusallığına göre öncelikle doğrusal ve doğrusal olmayan olarak ikiye ayrılmaktadır.



Doğrusal (Lineer) Regresyon Analizi: Değişkenler arasındaki ilişkinin doğrusal olduğu durumlarda yapılan regresyon analizidir.

Doğrusal Olmayan (Eğrisel) Regresyon Analizi: Değişkenler arasındaki ilişkinin doğrusal olmadığı durumlarda yapılan regresyon analizidir.

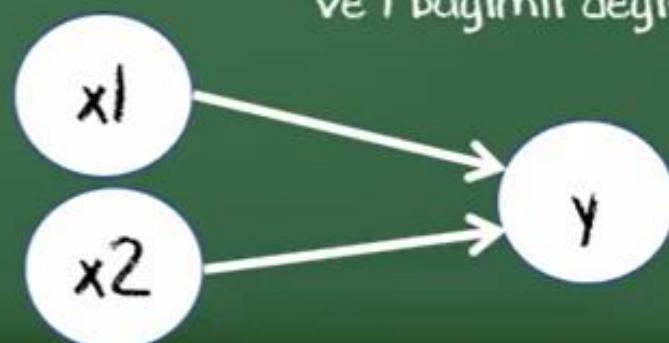


Basit Regresyon Analizi: 1 bağımsız (yordayıcı, kestirici, sebep) ve 1 bağımlı (yordanan, kestirilen, sonuç) değişken vardır.



Ayrıca, değişkenlerin sayısına göre ikiye ayrılmaktadır:

Çoklu Regresyon Analizi: Birden fazla bağımsız ve 1 bağımlı değişken vardır.



Basit Doğrusal Regresyon

1 bağımsız (yordayıcı) ve 1 bağımlı (yordanan) değişken bulunan, değişkenler arasındaki ilişkinin doğrusal olduğu regresyondur.



Az önce belirttiğimiz üzere regresyon, değişkenlerden birisi (x) bilindiğinde diğer değişkenin (y) alabileceği değerin tahmin edilmesidir.

Ancak bunun yapılabilmesi için x 'in aldığı değere göre y 'nin alacağı değeri her durumda bulmaya yarayan matematiksel denklem (eşitlik) bulunması gerekmektedir. Yani, bu matematiksel denklem bilinmelidir ki, x 'e göre y 'nin alabileceği değer tahmin edilebilsin.

$$\hat{y} = a + bx$$

Regresyon denklemi, x ve y değişkenleri arasındaki ilişkinin matematiksel olarak ifadesi yanı formülüdür ve bu formül üzerindeki değerlerin yerlerine konulması ile regresyon tahminleri yapmak mümkün olmaktadır.

İşte regresyon analizinde yapılan, değişkenler arasındaki ilişkinin matematiksel bir denklem ile açıklanmasıdır ve buna regresyon denklemi, eşitliği veya modeli adı verilmektedir.

Bağımlı Değişken

\hat{y} , her bir x değerine göre tahmin edilen ve bağımlı (sonuç, yordanan) değişken olan y değerini ifade etmektedir ve tahmini değer olduğu için \hat{y} simgesi ile gösterilmektedir.

Dolayısıyla, tahmini y değerinin bulunabilmesi için; x değerinin, x ve y değişkenleri arasındaki ilişkiye göre değişen belirli bir katsayı (regresyon katsayısı) ile çarpılması ve buna yine bu ilişkiye göre değişen sabit bir değer (sabit) eklenmesi gerekmektedir. Regresyon denklemi, bunun matematiksel olarak ifadesidir.



$$\hat{y} = a + bx$$

Sabit

a ise yine tahmini y değerinin bulunması için x 'e eklenmesi gereken sabit değerdir. Buna sabit adı verilmektedir. $X=0$ olduğunda y 'nin aldığı değerdir.

Bağımsız Değişken

x , bağımsız (sebep, yordayıcı) değişkendir.

Regresyon Katsayısı

b , tahmini y değerinin bulunması için x 'in çarpılması gereken katsayı değeridir. Buna regresyon katsayısı adı verilmektedir.

x 'in kendi birimi cinsinden 1 birim değişimmesine karşılık y 'de kendi birimi cinsinden meydana gelecek değişme miktarını ifade eder. Ayrıca, regresyon doğrusunun eğimidir.

x	y
1	1
2	3
4	4
5	3

Örneğin, yukarıdaki veri seti için regresyon katsayıısı yani $b = 0,5$; sabit yani $a = 1,25$ olarak bulunmuştur.

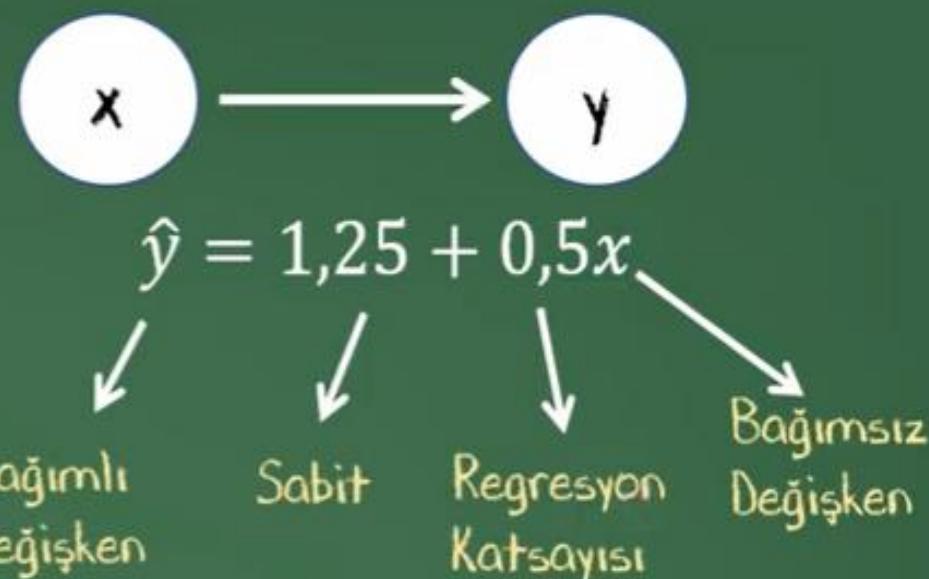
Buna göre; x değişkeninin 3 veya 7 olması halinde y'nin alabileceği değerin tahmin edilmesi için regresyon eşitliğinin kullanılması yeterlidir:

x'in 3 olması halinde

$$\hat{y} = a + bx \text{ olduğu için } \hat{y} = 1,25 + 0,5 (3) \text{ yani } \hat{y} = 2,75$$

x'in 7 olması halinde

$$\hat{y} = a + bx \text{ olduğu için } \hat{y} = 1,25 + 0,5 (7) \text{ yani } \hat{y} = 4,75$$



Ancak, regresyon katsayıısı ve sabitin bilinmediği durumlarda öncelikle bunların bulunması gerekmektedir.

Buraya kadar basit doğrusal regresyonu öğrendikten sonra
şimdi de şunu belirtmeliyiz:

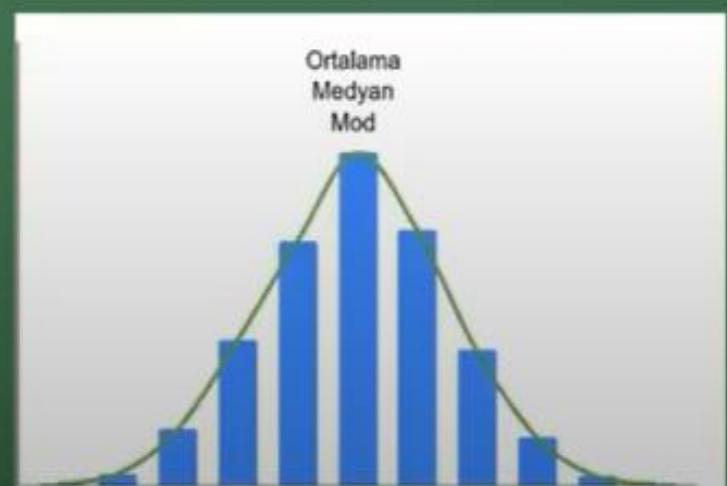
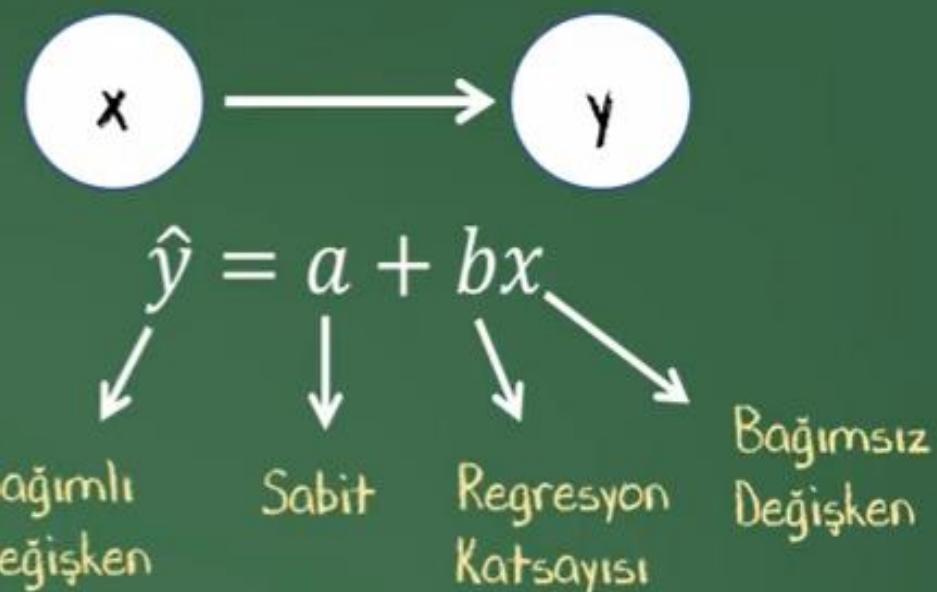
$\hat{y} = a + bx$ ile y 'nin tahmini (beklenen) değeri
bulunmaktadır. Buna **teorik** değer denilmektedir.

Gerçek y değeri ise buna, yapılan hata miktarının eklenmesi ile
bulunabilecektir. Buna **gerçek** değer adı verilmektedir.

$y = \alpha + bx + e$ veya $y = \alpha + \beta x + \varepsilon$ ile
formülleştirilmektedir.

Yani önceki formülde bulunmuş olan teorik y değerine (\hat{y}) hata
miktarı (e) eklenmektedir.

Yalnız, formüldeki e her gözlem değeri için farklıdır. Ancak
 e 'nin tüm gözlemler için ortalaması 0'dır. Ayrıca hatalar,
normal dağılıma sahiptir.



x	y
1	1
2	3
4	4
5	3

Regresyon katsayıısı yani $b = 0,5$; sabit yani $a = 1,25$ idi.

Regresyon denklemi ise şu şekilde idi: $\hat{y} = 1,25 + 0,5(x)$

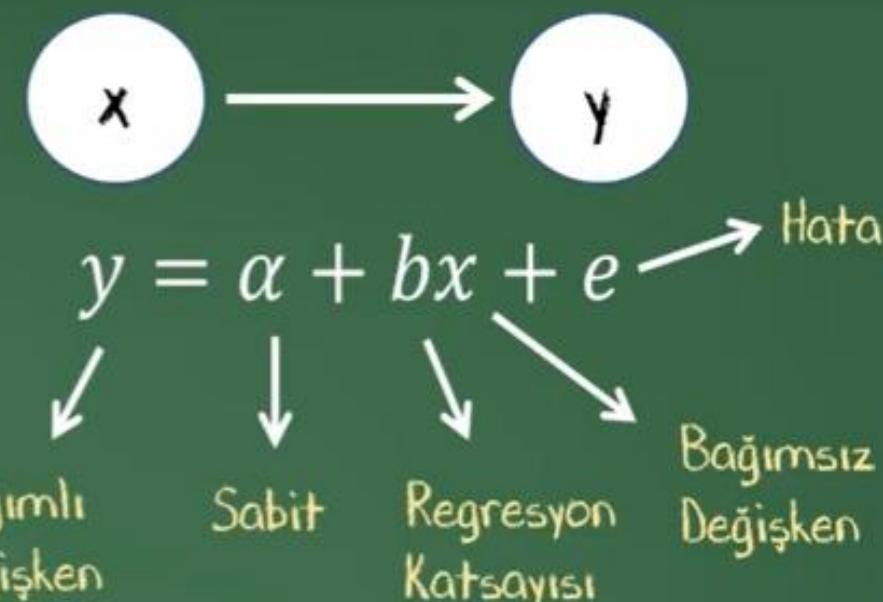
Bu denkleme göre gözlenen değerler arasında yer alan 4'ün y tahminini bulmak istersek;

$$\hat{y} = 1,25 + 0,5(4)$$

$\hat{y} = 3,25$ olacaktır.

Ancak gördüğümüz üzere $x=4$ olduğunda $y=4$ 'tür.

Yani teorik değer 3,25'tir ama gerçek değer 4'tür. 0 zaman 0,75 hata bulunmaktadır.



Bu şekilde bir hata çıkışının nedeni iki değişken arasındaki ilişkinin gücüne bağlıdır. $r=1,00$ olduğunda mükemmel tahmin yapılmaktadır.



x	y
1	1
2	3
4	4
5	3

Regresyon katsayıısı yani $b = 0,5$; sabit yani $a = 1,25$ idi.

Regresyon denklemi ise şu şekilde idi: $\hat{y} = 1,25 + 0,5(x)$

Bu denkleme göre gözlenen değerler arasında yer alan

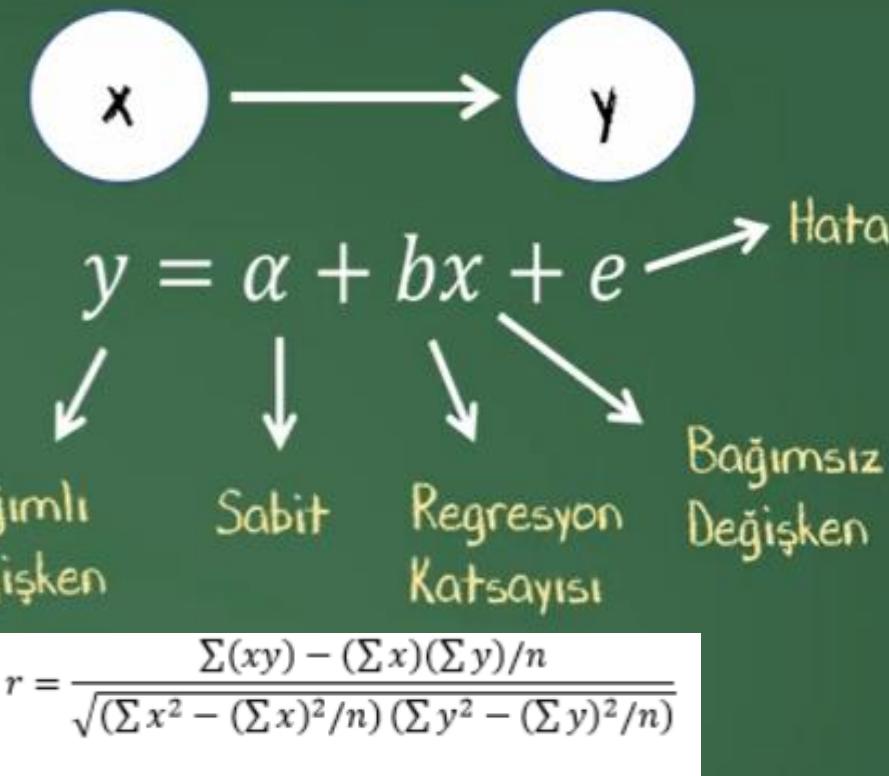
4'ün y tahminini bulmak istersek;

$$\hat{y} = 1,25 + 0,5(4)$$

$\hat{y} = 3,25$ olacaktır.

Ancak gördüğümüz üzere $x=4$ olduğunda $y=4$ 'tür.

Yani teorik değer 3,25'tir ama gerçek değer 4'tür. O zaman 0,75 hata bulunmaktadır.



Bu şekilde bir hata çıkışının nedeni iki değişken arasındaki ilişkinin gücüne bağlıdır. $r=1,00$ olduğunda mükemmel tahmin yapılmaktadır.

Ancak ilişkinin gücü azaldığında tahminin doğruluğu düşmekte, hata artmaktadır. Burada $r=0,73$ 'tir. Bundan dolayı bu hata ortaya çıkmıştır.

Ayrıca, burada çok önemli bir notu belirtmeliyiz:

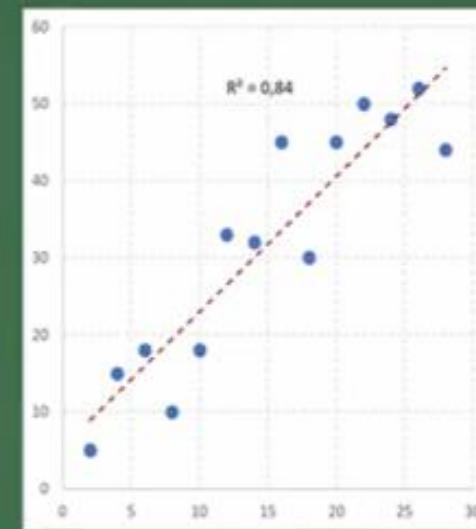


Regresyon analizinin yapılabilmesi için,

(I) değişkenler arasında nötr olmayan bir korelasyon bulunmalıdır. Aksi halde, yani aralarında hiçbir korelasyon bulunmayan değişkenler arasında, temeli tahmin yapma olan regresyon analizinin yapılması anlamsız olacaktır.

(II) Değişkenler arasında doğrusal bir ilişki bulunmalıdır.

(III) Bağımlı değişken eşit aralıklı veya eşit orantılı ölçme düzeyinde ve sürekli olmalıdır. Bağımsız değişkenin de bu özelliklere sahip olması beklenir ancak kategorik olmaları halinde de regresyon yapılabilir.



(IV) Her iki değişken de normal dağılım göstermelidir. Çünkü regresyon analizi aşırı uç değerlerden ve dağılım özelliklerinden ciddi olarak etkilenir.



Regresyon Katsayıısı $\hat{y} = a + bx$

$$1,5 = a + 1,5(1)$$
$$3 = a + 1,5(2)$$

$$6 = a + 1,5(4)$$

Regresyon katsayıısı (b), tahmini y değerinin bulunması için x 'in çarpılması gereken katsayıı değeridir.

Bağımsız değişkende bir birimlik değişme (artma veya azalma) olduğunda, bağımlı değişkende bunun kaç katı artma veya azalma olacağını gösterir.

Aşağıdaki formüllerden birisi kullanılarak bulunabilmektedir:

$$b = \frac{\sum xy - n \cdot \bar{x} \bar{y}}{\sum x^2 - n \cdot \bar{x}^2}$$

$$b = r \frac{s_y}{s_x}$$

Buna göre; regresyon katsayıısı, I. formülde x ve y değerleri kullanılarak, II. formülde ise korelasyon katsayıısı (r) ve standart sapma (s) değerleri arasındaki ilişkiye göre bulunabilmektedir.

x	y
1	1
2	3
4	4
5	3

Örneğin, verilen değerler için regresyon katsayıısı her iki formül kullanılarak şu şekilde hesaplanabilmektedir:

Regresyon Katsayısı $\hat{y} = a + bx$

$$b = \frac{\sum xy - n \cdot \bar{x} \bar{y}}{\sum x^2 - n \cdot \bar{x}^2}$$
 kullanırsak

x	y	xy	x	y	x^2
1	1	1	3	2,75	1
2	3	6	3	2,75	4
4	4	16	3	2,75	16
5	3	15	3	2,75	25
=12	=11	=38			=46

$$b = \frac{38 - (4) \cdot (3) \cdot (2,75)}{46 - 4 \cdot (3)^2} \rightarrow \frac{38 - 33}{46 - 36} \rightarrow \frac{5}{10} \rightarrow b = 0,5$$

Sabit $\hat{y} = a + bx$

Sabit (a), tahmini y değerinin bulunması için x 'e eklenmesi gereken sabit değerdir.

Aşağıdaki formül yardımıyla bulunabilmektedir:

$$a = \bar{y} - b\bar{x}$$

x	y
1	1
2	3
4	4
5	3

Örneğin, verilen değerler için şu şekilde hesaplanabilmektedir:

$a = \bar{y} - b(\bar{x})$ formülünde y ve x değişkenlerinin ortalamalarına ve b 'ye yani regresyon katsayısına ihtiyaç duyulmaktadır.

Bunlar $\bar{y} = 2,75$, $\bar{x} = 3$ ve $b=0,5$ 'dir.

Bu değerler formüldeki yerlerine koyulduğunda

$$a = 2,75 - 0,5(3) \rightarrow a = 2,75 - 1,50 \text{ yani}$$

$$a = 1,25 \text{ olacaktır.}$$

Dikkat edilirse, 1,25 sonucu buraya kadar belirttiğimiz denklem ile aynı çıkmıştır.



$$\hat{y} = 1,25 + 0,5(x)$$

x	y
2	8
6	5
7	7
9	4
8	6

Yukarıda verilen x değişkeni bağımsız,
y değişkeni ise bağımlı değişkendir.
Değişkenler arasındaki korelasyon
katsayısı ise -0,76'dır.

1. regresyon katsayısını,
2. regresyondaki sabit değeri,
3. regresyon denklemini ve bunun
ifade ettiği anlamı,
4. x 'in 7 olması halinde y 'nin
alabileceği tahmini değeri bulunuz.

x	y
2	8
6	5
7	7
9	4
8	6

Yukarıda verilen x değişkeni bağımsız, y değişkeni ise bağımlı değişkendir. Değişkenler arasındaki korelasyon katsayısı ise -0,76'dır.

1. regresyon katsayısını,
2. regresyondaki sabit değeri,
3. regresyon denklemini ve bunun ifade ettiği anlamı,
4. x'in 7 olması halinde y'nin alabileceği tahmini değeri bulunuz.

Regresyon katsayısını hesaplamak için $b = \frac{\sum xy - N \cdot \bar{x} \bar{y}}{\sum x^2 - N \cdot \bar{x}^2}$ formülünü kullanmak gerekmektedir.

Bunun için ise aşağıdaki tablo hazırlanmalıdır:

x	y	xy	\bar{x}	\bar{y}	x^2
2	8	16	6,4	6	4
6	5	30	6,4	6	36
7	7	49	6,4	6	49
9	4	36	6,4	6	81
8	6	48	6,4	6	64
=32	=30	=179			=234

Buna göre, değerler formül üzerine yerleştirildiğinde aşağıdaki gibi olacaktır:

$$b = \frac{179 - (5) \cdot (6,4) \cdot (6)}{234 - 5 \cdot (6,4)^2} \rightarrow \frac{179 - 192}{234 - 204,8} \rightarrow \frac{-13}{29,2} \rightarrow b = -0,4452$$

x	y
2	8
6	5
7	7
9	4
8	6

Yukarıda verilen x değişkeni bağımsız, y değişkeni ise bağımlı değişkendir.
Değişkenler arasındaki korelasyon katsayısı ise -0,76'dır.

1. regresyon katsayısını, $b = -0,4452$

2. regresyondaki sabit değeri,

3. regresyon denklemini ve bunun ifade ettiği anlamı,

4. x'in 7 olması halinde y'nin alabileceği tahmini değeri bulunuz.

Sabit değeri hesaplamak için $a = \bar{y} - b(\bar{x})$ formülünde y ve x değişkenlerinin ortalamalarına ve b'ye yani regresyon katsayısına ihtiyaç duyulmaktadır.

x	y	xy	\bar{x}	\bar{y}	x^2
2	8	16	6,4	6	4
6	5	30	6,4	6	36
7	7	49	6,4	6	49
9	4	36	6,4	6	81
8	6	48	6,4	6	64
=32	=30	=179			=234

Bunlar $\bar{y} = 6$, $\bar{x} = 6,4$ ve $b = -0,4452$ 'dir.

Bu değerler formüldeki yerlerine koymduğunda

$$a = 6 - (-0,4452(6,4))$$

$$a = 6 - (-2,8493) \text{ yani } a = 8,8493 \text{ olacaktır.}$$

x	y
2	8
6	5
7	7
9	4
8	6

Yukarıda verilen x değişkeni bağımsız, y değişkeni ise bağımlı değişkendir. Değişkenler arasındaki korelasyon katsayısı ise -0,76'dır.

1. regresyon katsayısını, $b = -0,4452$

2. regresyondaki sabit değeri, $a = 8,8493$

3. regresyon denklemini ve bunun ifade ettiği anlamı, $\hat{y} = 8,8493 - 0,4452(x)$

4. x 'in 7 olması halinde y 'nin alabileceği tahmini değeri bulunuz.

Regresyon denkleminin formülü $\hat{y} = a + bx$ 'tir.

Bu formülün, verilen değişkenler arasındaki ilişkiyi ifade eden hali ise formüldeki değerleri yerlerine koyarak bulunabilecektir:

$$\hat{y} = 8,8493 - 0,4452(x)$$

Bunun anlamı; tahmini y değerlerinin, x değerlerinin $-0,4452$ katının $8,8493$ fazlasına eşit olduğunu ifade etmektedir.

x	y
2	8
6	5
7	7
9	4
8	6

Yukarıda verilen x değişkeni bağımsız, y değişkeni ise bağımlı değişkendir. Değişkenler arasındaki korelasyon katsayısı ise -0,76'dır.

1. regresyon katsayısını, $b = -0,4452$

2. regresyondaki sabit değeri, $a = 8,8493$

3. regresyon denklemini ve bunun ifade ettiği anlamı, $\hat{y} = 8,8493 - 0,4452(x)$

4. x'in 7 olması halinde y'nin alabileceği tahmini değeri bulunuz. $\hat{y} = 5,73$

x'in 7 olması halinde y'nin alabileceği tahmini değeri, önceki adımda bulduğumuz regresyon eşitliğini kullanarak kolayca bulmak mümkündür.

$\hat{y} = 8,8493 - 0,4452(x)$ olduğu için x'in 7 olması halinde

$\hat{y} = 8,8493 - 0,4452(7) \rightarrow \hat{y} = 5,73$ olacaktır.

Standart Hata

$$y = \alpha + bx + e$$

$$y = \alpha + \beta x + \varepsilon$$

$e = y - \hat{y}$ 'dır.

Peki standart hatayı nasıl hesaplayabiliriz?

$$n \geq 30 \text{ ise } s_e = \sqrt{\frac{\sum(y-\hat{y})^2}{n}}$$

$$n < 30 \text{ ise } s_e = \sqrt{\frac{\sum(y-\hat{y})^2}{n-2}}$$

Ya da

$$s_e = s_y \sqrt{\frac{n-1}{n-2}(1-r^2)}$$

x	y
2	8
6	5
7	7
9	4
8	6

Örneğin, yine aynı veri seti için hesaplayalım.

$\hat{y} = 8,8493 - 0,4452(x)$ idi:

y	\hat{y}	$y - \hat{y}$	$(y - \hat{y})^2$
8	7,96	0,04	0,002
5	6,18	-1,18	1,388
7	5,73	1,27	1,606
4	4,84	-0,84	0,710
6	5,29	0,71	0,507
$\Sigma = 30$		$\Sigma = 0$	$\Sigma = 4,21$

$$s_e = \sqrt{\frac{\sum(y-\hat{y})^2}{n-2}} \rightarrow s_e = \sqrt{\frac{4,21}{3}} \rightarrow s_e = 1,18$$

$$y = \alpha + bx + e \text{ idi.}$$

0 zaman gerçek y değerinin denklemi şu olacaktır:

$$y = 8,8493x - 0,4452(x) + 1,18$$

* Teorik değerlerin toplamı ve gerçek değerlerin toplamı birbirine eşittir.

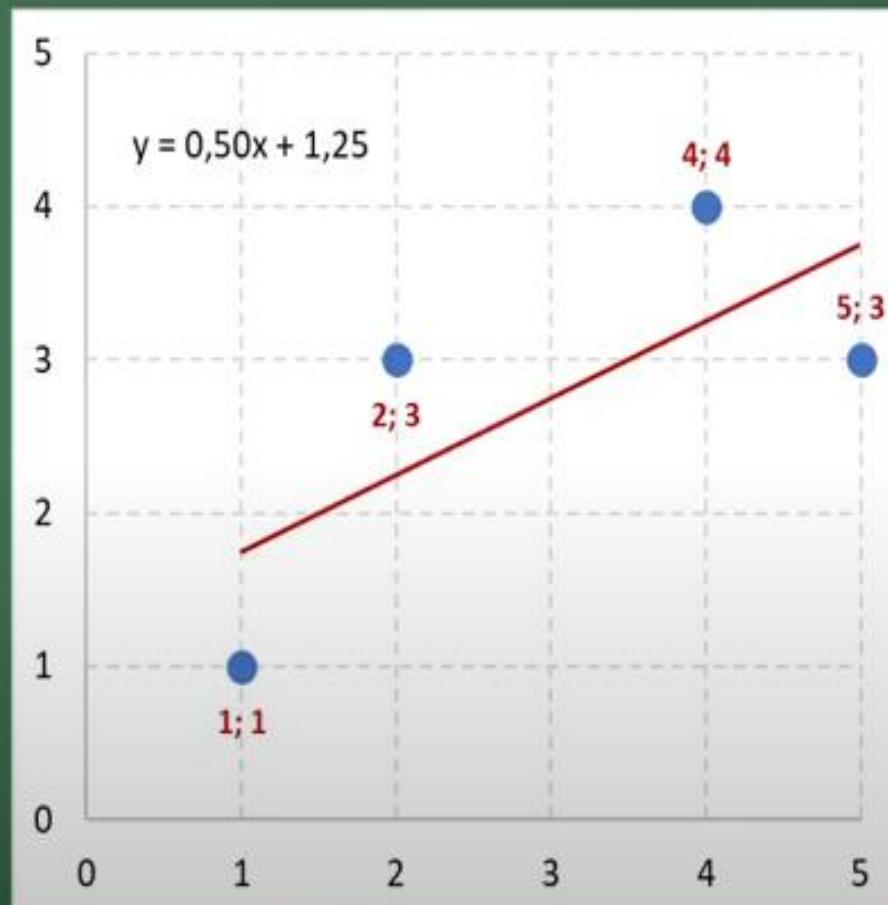
* Teorik değerler ile gerçek değerlerin farklarının toplamı her zaman 0'dır.

Dağılım Grafiği (Serpilme Diyagramı) ve Regresyon Doğrusu



Regresyon ile korelasyon birbiri ile yakından ilişkili konular olduğu için tipki korelasyonda olduğu gibi regresyonu da görselleştirmek ve analiz etmek için dağılım grafiği kullanmak son derece kullanışlı bir yöntemdir. Yalnız bu defa grafik üzerine regresyon doğrusu adı verilen çizgi eklenmektedir.

x	y
1	1
2	3
4	4
5	3



Bu adımları
bir örnek
üzerinde
uygulayalım:

x	y
1	1
2	3
4	4
5	3

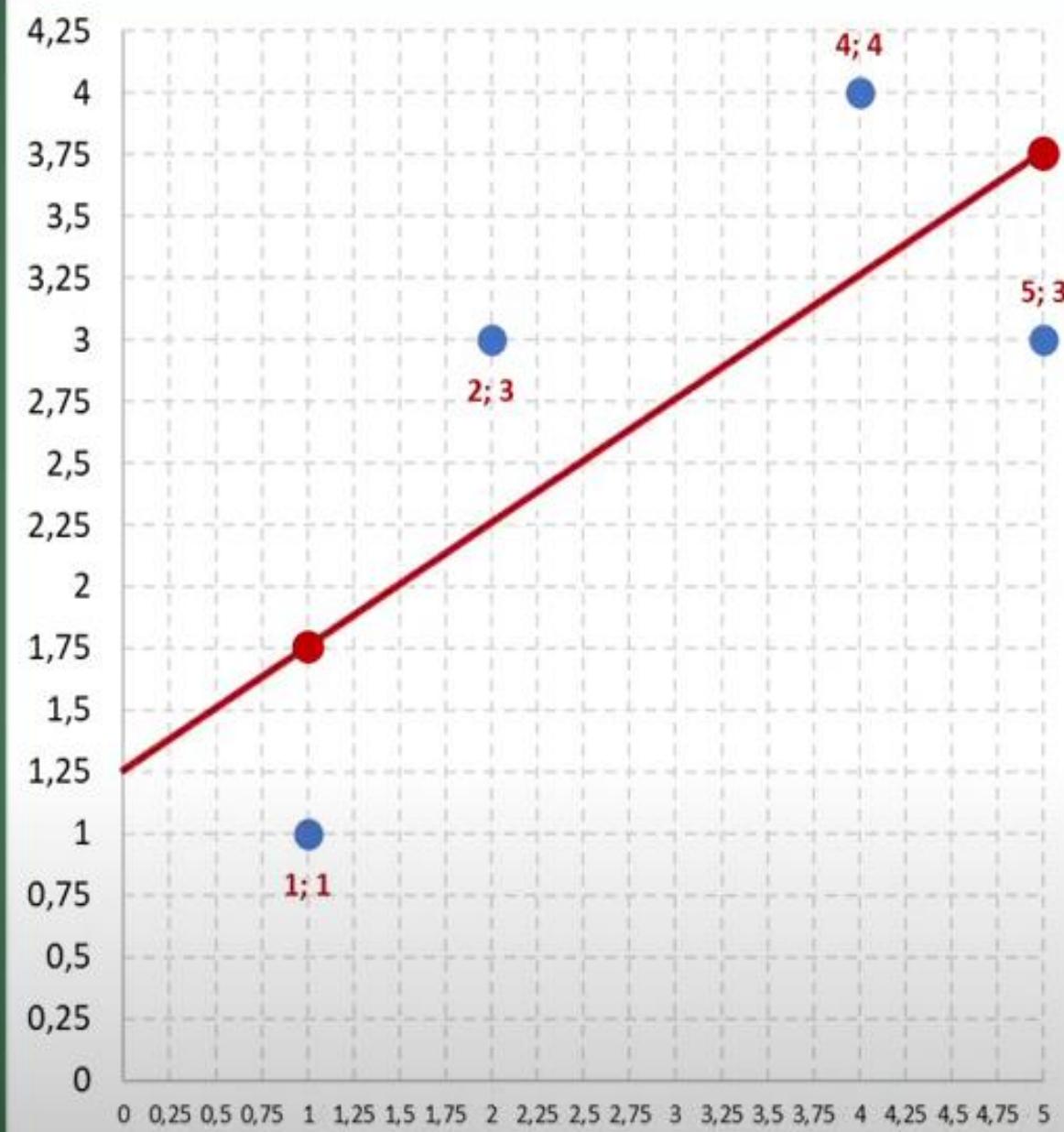
1. Bu verilere ait regresyon eşitliği hesaplanmalıdır. $\hat{y} = 1,25 + 0,5(x)$

2. İstenen iki farklı x değerinin tahmini y değerleri bulunmalıdır.

3. x ve y gözlenen değerleri grafik üzerine yerleştirilmelidir.

4. Tahmini y değerleri ve bunların gözlenen x karşılıkları grafik üzerine işaretlenmelidir. Ardından bu noktalar düz bir çizgi ile birleştirilerek regresyon doğrusu çizilmelidir.

x	y	\hat{y}
1	1	1,25
5	3	3,75

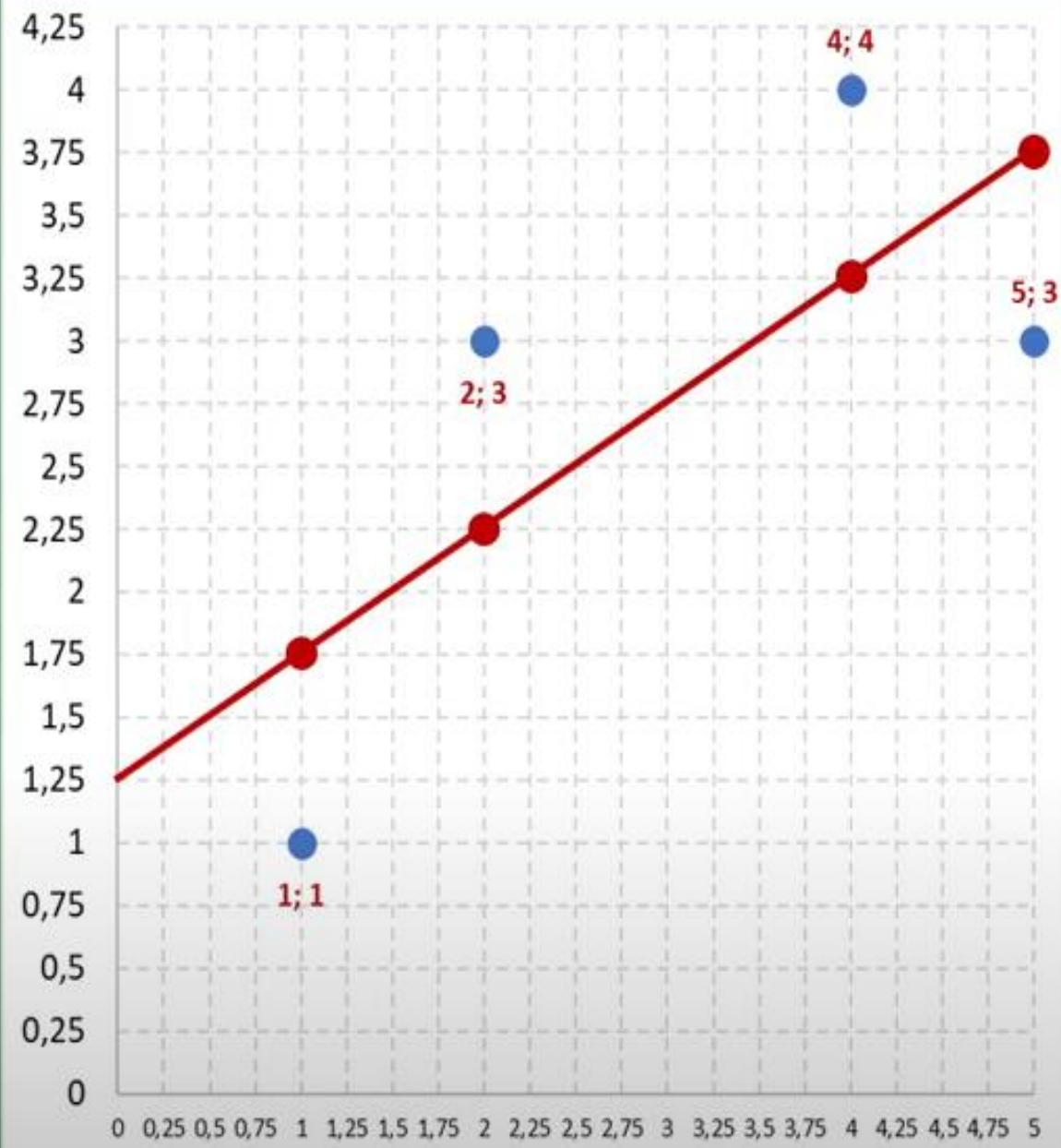


1. Regresyon doğrusu, her bir x değerine karşılık olarak tahmin edilen y değerlerinin birleştirilmesi ile ortaya çıkan bir doğrudur.

Yani regresyon doğrusu $\hat{y} = a + bx$ formülü yardımıyla hesaplanan \hat{y} 'leri birleştiren doğrudur.

x	y	\hat{y}
1	1	1,75
2	3	2,25
4	4	3,25
5	3	3,75

2. Regresyon doğrusu, tahmini değerlerin birleştirilmesi sonucu çizildiği için gerçek (gözlenen) değerler her zaman doğru üzerinde olmayı beklemektedir.

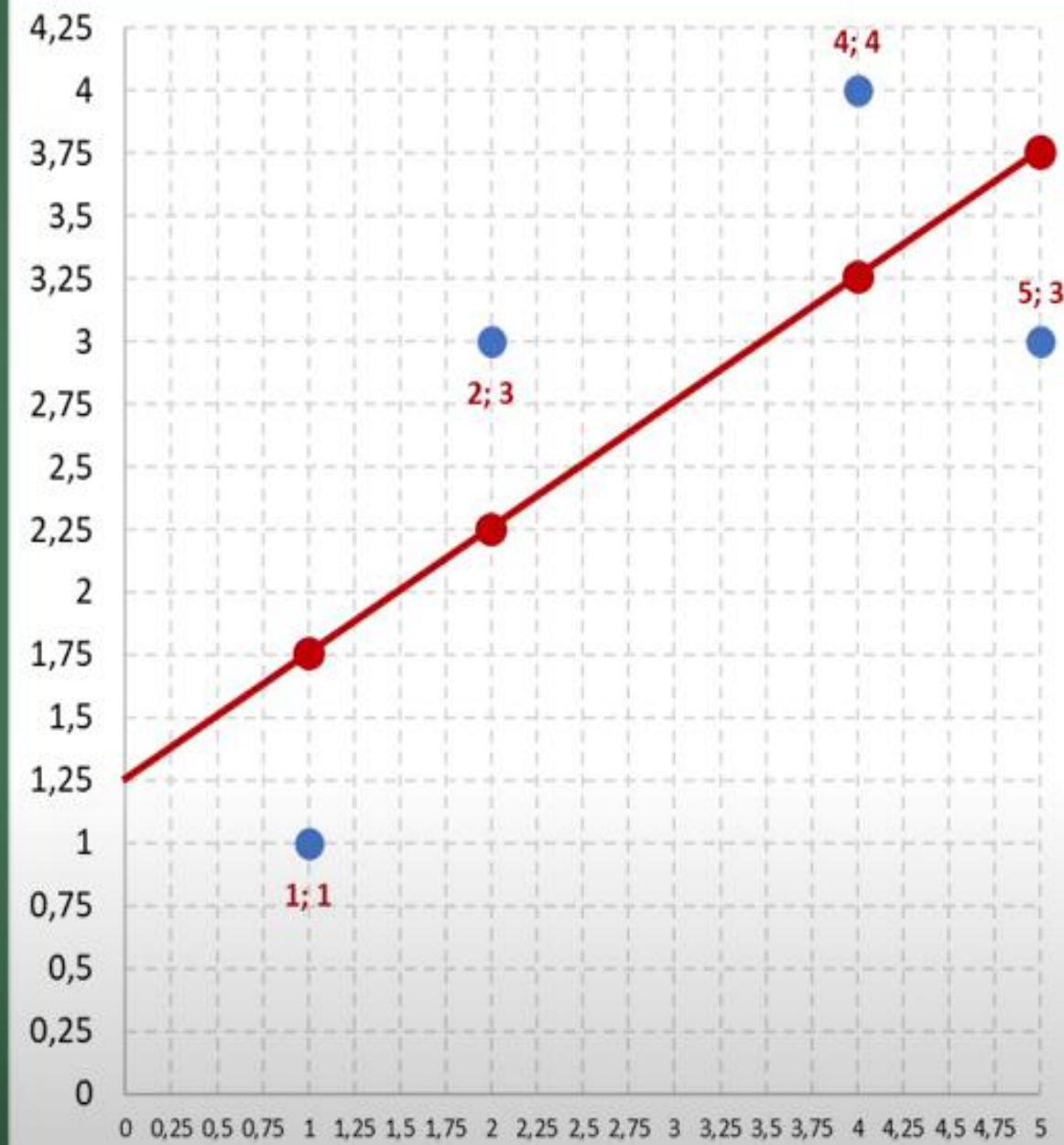


3. Gözlenen değerler, sadece x ve y değişkenleri arasındaki korelasyonun mükemmel olduğu (yani korelasyon katsayısının +1,00 veya -1,00 olduğu) durumlarda regresyon çizgisi üzerinde yer almaktadır.

Bunun nedeni, değişkenlerin ancak mükemmel korelasyon olması halinde bir çizgi halinde uzanmasıdır.

Bu durumda korelasyon katsayısının karesine eşit olan determinasyon katsayıısı (r^2) yani tahminin gücü %100 olduğu için tahmini y değerleri ile gözlenen y değerleri aynı olmaktadır.

4. Korelasyonun mükemmel olduğu durumlarda, gözlenen ve tahmini değerler arasındaki farkı ifade eden ve "sapma" veya "hata" olarak adlandırılan değer sıfır olmaktadır. Çünkü bu durumda gözlenen ve tahmini değerler arasında fark yoktur.



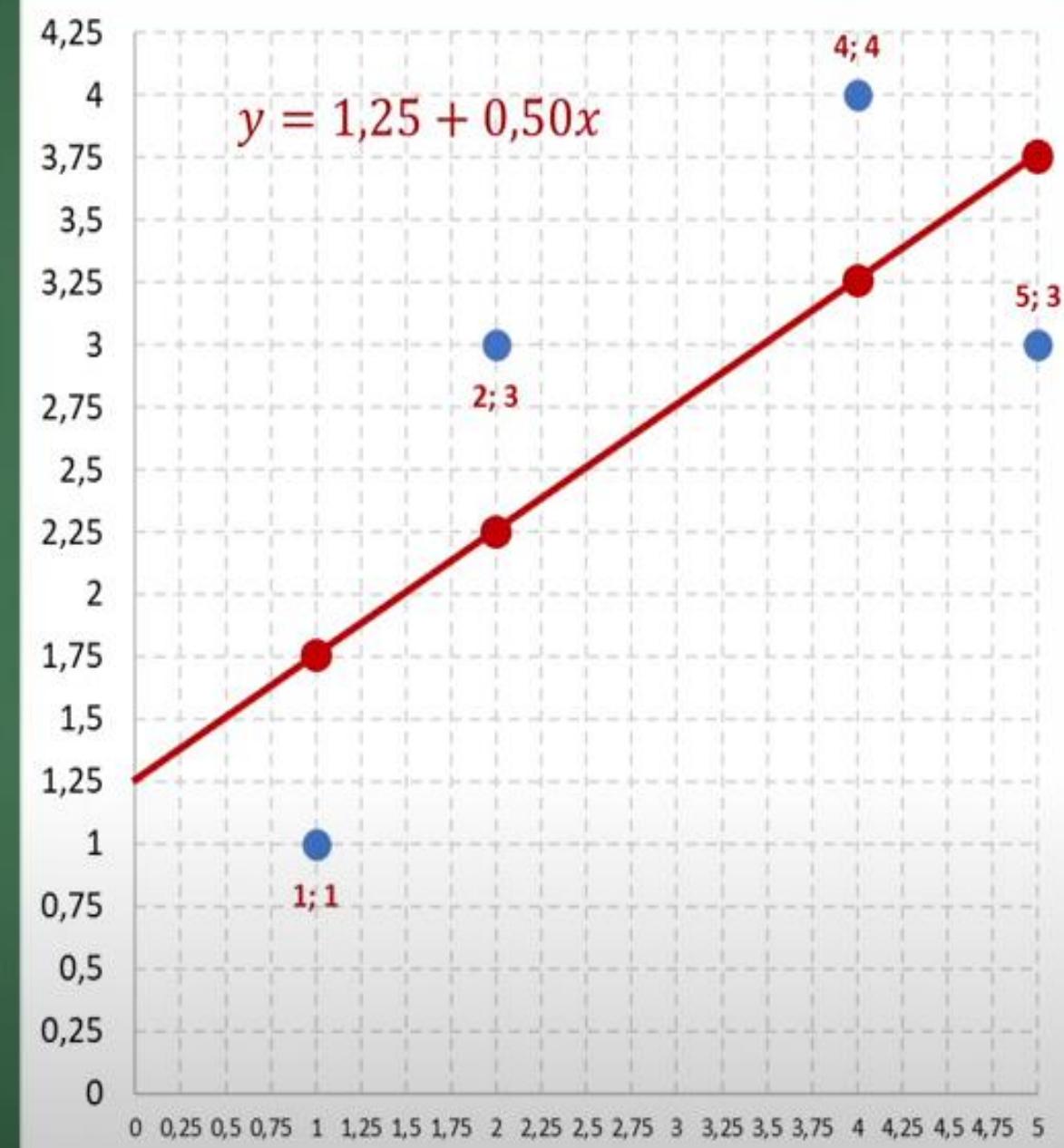
6. Değişim grafiği üzerinde çizginin en soldaki başlangıç noktası her zaman sabit değeridir, yani çizgi sabitten başlar ve aşağıya veya yukarıya doğru uzanır.

Bunu x ekseni 0'dan başlayacak şekilde düzenlendiyse görebilirsiniz.

Şekilde sabit değer 1,25'tir ve $x = 0$ olması hâlinde $y = 1,25$ olacağı için regresyon çizgisi buradan başlayacaktır.

7. Regresyon katsayısı, regresyon çizgisinin aşağıya veya yukarıya doğru eğimini vermektedir. Katsayının değerinin negatif (-) olması hâlinde çizgi aşağıya doğru; pozitif (+) olması ise yukarıya doğru uzanır.

Şekilde katsayı 0,50 olduğu için çizgi yukarıya doğru uzanmıştır.

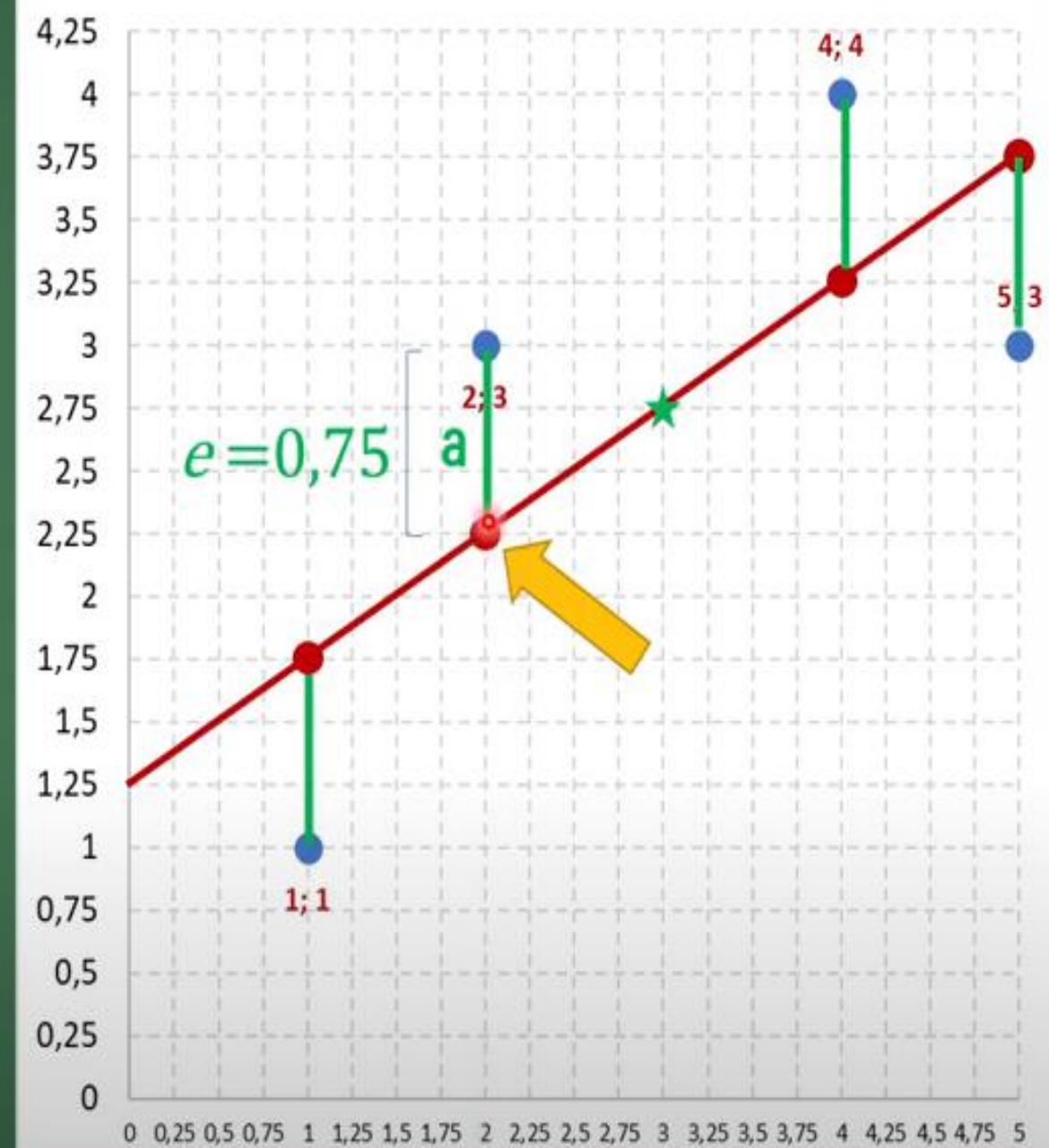


8. Regresyon doğrusu, beklenen y değerleri yanında her zaman x ve y değişkenlerinin aritmetik ortalamalarının kesişiminden geçer.

Şekilde bu kesişim noktası $x=3$ ve $y=2,75$ noktasıdır ve çizgi buradan geçmiştir.

9. Dağılım grafiği üzerinde; gözlenen herhangi bir x noktasından, bu noktanın regresyon doğrusu üzerinde yer alan tahmini y karşılığına doğru çizilen çizginin uzunluğu hata değerine eşittir.

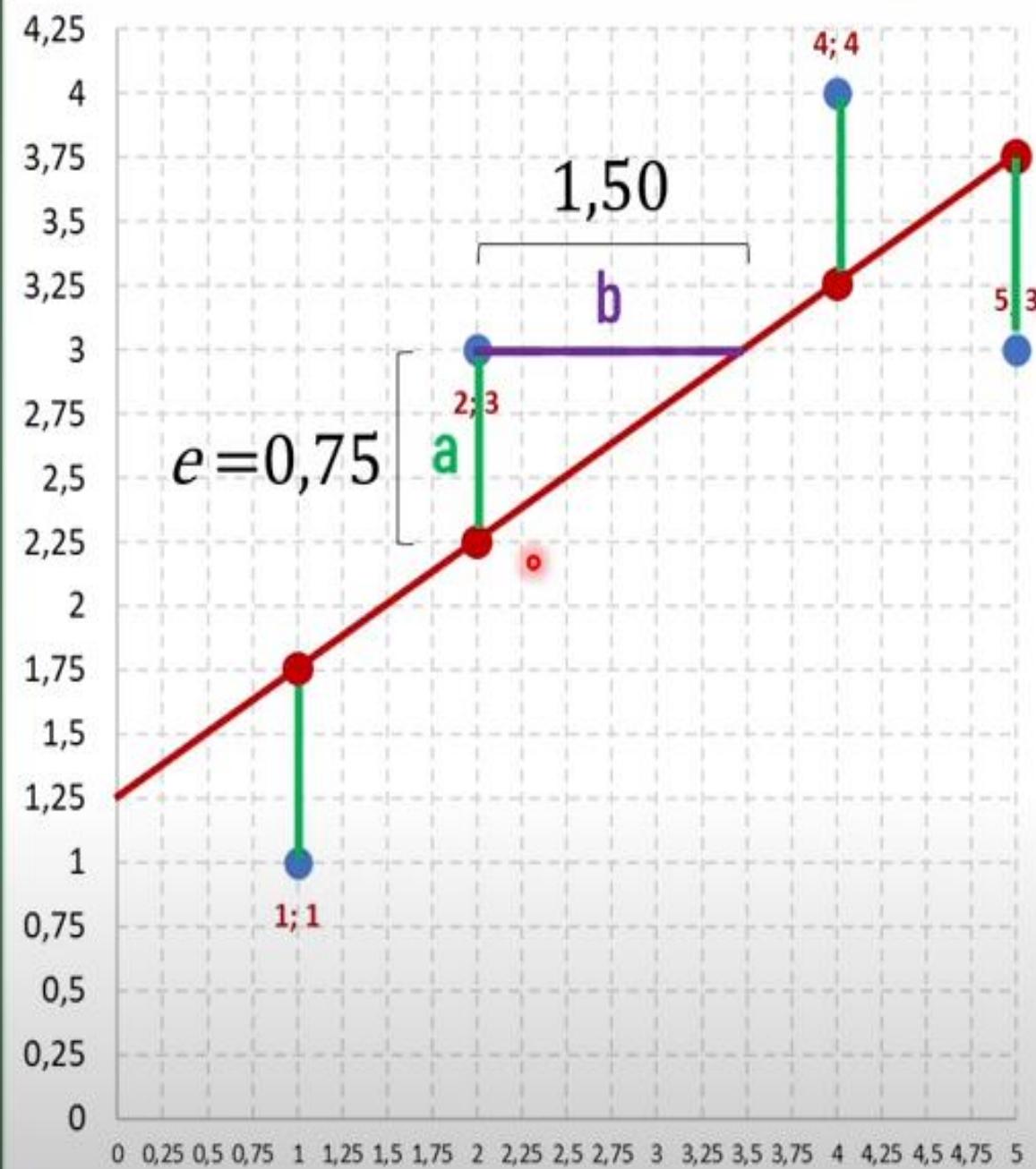
Şekilde $x=2$ 'nin tahmini y karşılığı 2,25'tir. Bu nedenle regresyon doğrusu bu noktadan geçmiştir. Ancak, gözlenen y değeri 3'tür. Dolayısıyla sapma yani hata miktarı 0,75'tir ve bu değer a ile gösterilen çizginin uzunluğuna eşittir.



10. Gözlenen x değerinden regresyon doğrusuna birleşecek yatay bir çizgi çizilmesi halinde bu kenarın uzunluğu da bulunabilir.

Bunun için çizilen çizginin doğru ile keşşim noktasının x değeri (hesaplama veya görsel olarak) bulunmalıdır.

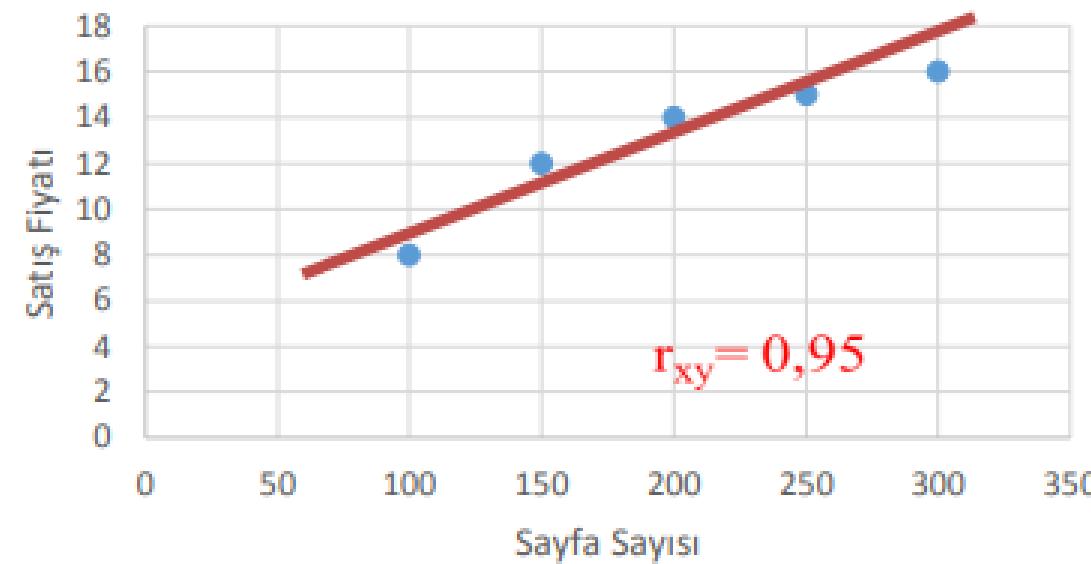
Örneğin, şekilde bu noktanın $x=3,5$ olduğu görülmektedir. Bu durumda şekilde b ile gösterilen kenarın uzunluğu $3,5 - 2$ 'den $1,5$ olacaktır.



Örnek

Aşağıdaki tabloda kitapların fiyatı ve sayfa sayıları verilmiştir. Sayfa sayısını kullanarak kitabı tahmin edebileceğimiz regresyon doğrusunu (Tahmin Denklemi) bulunuz.

Sayfa Sayısı (X)	Satış Fiyatı (Y)
100	8
150	12
200	14
250	15
300	16



Çözüm

Sayfa Sayısı bağımsız değişken (X)

Fiyat bağımlı değişken (Tahmin edilen Y) olacaktır. Buna göre;

Sayfa Sayısı (X)	Satış Fiyatı (Y)	X*Y	X ²
100	8	800	10000
150	12	1800	22500
200	14	2800	40000
250	15	3750	62500
300	16	4800	90000
1.000	65	13.950	225.000

$$\bar{x} = \frac{1000}{5} = 200$$

$$\bar{y} = \frac{65}{5} = 13$$

$$b_{yx} = \frac{\sum_{i=1}^n x_i * y_i - n * \bar{x} * \bar{y}}{\left(\sum_{i=1}^n x_i^2 - n * (\bar{x})^2 \right)} = \frac{13950 - 5 * 200 * 13}{(225000 - 5 * 200^2)} = \frac{13950 - 13000}{(225000 - 200000)}$$

$$b_{yx} = \frac{950}{25000} \quad b_{yx} = 0,038 \text{ TL}$$

Sayfa Sayısı (X) 1 adet arttığında kitabıın Fiyatı (Y) ortalama 0,038 TL artmaktadır.

$$a = \bar{Y} - b_{yx} * \bar{X} \quad a = 13 - 0,038 * 200 \quad a = 5,4$$

Tahmin denklemi ise $Y = 5,4 + 0,038 * X$

