Korelasyon analizi

Korelasyon analizi

- Sınava hazırlanma süresi ile sınavdan alınan başarı arasında ilişki var mıdır?
  - N sayıda öğrencinin sınava hazırlanma süreleri ile sınavdan aldıkları puanlar tespit edilir.
- Reklam harcaması ile satış miktarı arasında ilişki var mıdır?
  - N sayıda firmanın reklam harcaması ile satış miktarları tespit edilir.

Korelasyon analizi

- Derslerde kimi öğrenciler çok hızlı öğrenirken kimileri başarısız oluyor?
  - “Neden?” sorusu korelasyon araştırmasına götürür.

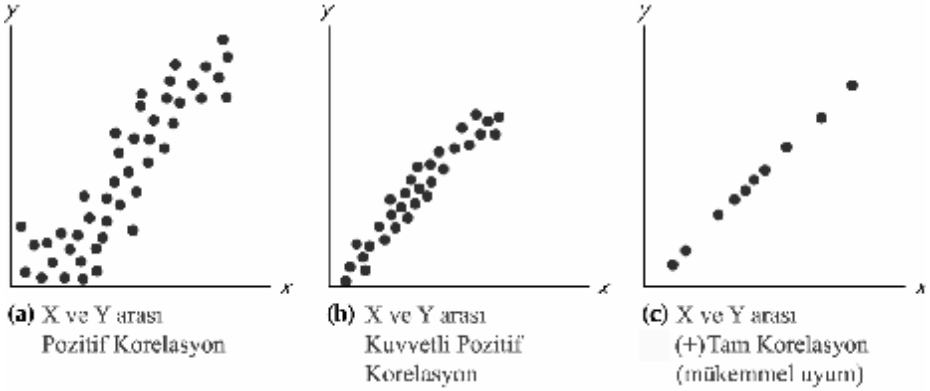
Korelasyon analizinin niteliği

- İki veya daha fazla değişkeni etkilemeye dönük bir müdahalede bulunmadan aralarındaki ilişkiyi saptama.
- Deneysel araştırmanın aksine, korelasyon araştırmasında değişkenler üzerinde bir manipülasyon olmaz.

Korelasyon analizinin niteliği

- Değişkenler arasındaki mevcut ilişkiyi betimlediği için betimleyici araştırmadır.
- Bir “korelasyon katsayısı” bularak aradaki ilişkiyi betimleriz.

Pozitif Lineer Korelasyon



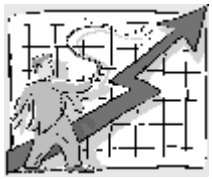
(a) X ve Y arası Pozitif Korelasyon

(b) X ve Y arası Kuvvetli Pozitif Korelasyon

(c) X ve Y arası (+)Tam Korelasyon (mükemmel uyum)

Korelasyon katsayısı

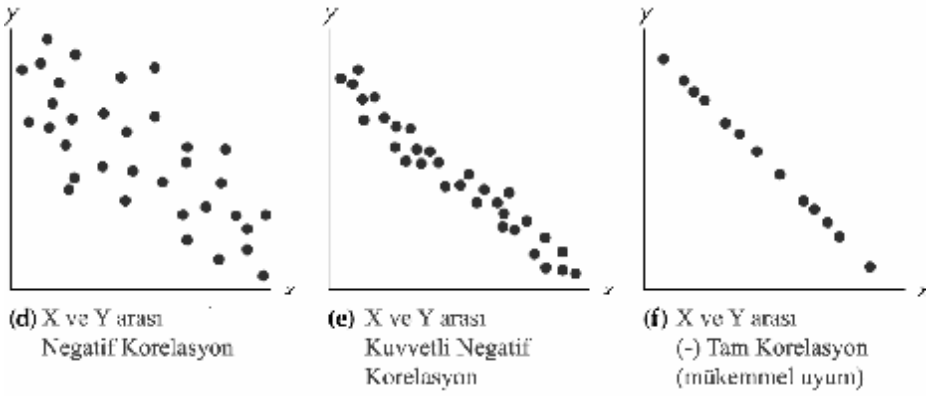
- Korelasyon katsayısının pozitif çıkması beklenen durumlar:
  - Reklam harcaması – Satış miktarı
  - Gelir – Ödenen kira
  - Dekar başına atılan gübre miktarı – Alınan verim



Korelasyon katsayısı(2)

- Korelasyon katsayısının negatif çıkması beklenen durumlar:
  - Günlük içilen sigara sayısı – koşulabilen mesafe

Negatif Lineer Korelasyon

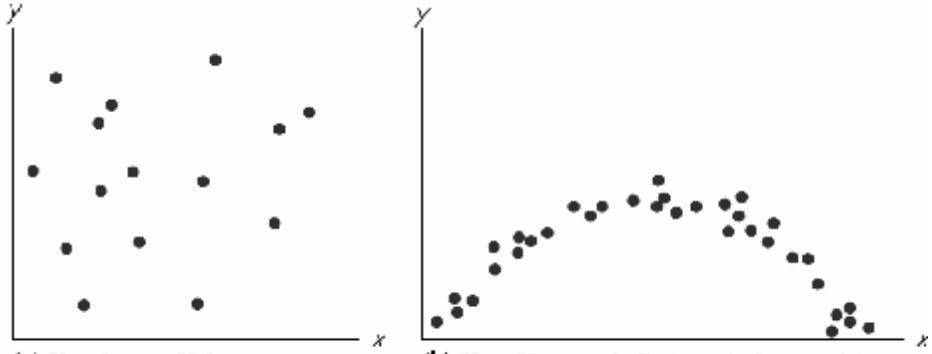


(d) X ve Y arası Negatif Korelasyon

(e) X ve Y arası Kuvvetli Negatif Korelasyon

(f) X ve Y arası (-) Tam Korelasyon (mükemmel uyum)

Doğrusal Korelasyon Yoktur



(g) Korelasyon Yok X ve Y arasında ilişki yok

(h) X ve Y arasında Doğrusal olmayan bir ilişki mevcut

Korelasyon katsayısı(3)

- Korelasyon katsayısının sıfır çıkması beklenen durumlar:
  - Öğrencinin derslerindeki başarısı ile ayakkabı numarası arasındaki korelasyon

Korelasyon analizinin amaçları

- Temel iki amacı vardır:
  1) Değişkenler arası ilişkinin açıklanmasına yardımcı olmak
  2) Benzer sonuçları önceden tahmin etmek

Açıklayıcı çalışmalar

- Değişkenler arasındaki ilişkiyi açıklamayı amaçlar.
- Sebep ve etki hakkında veri toplamaya çalışır.

Korelasyon analizinin temel aşamaları (Problem seçimi)

- Genelde 3 tip probleme yoğunlaşılır
  1. X değişkeni Y değişkeniyle ilişkili midir?
  2. X değişkeni Y değişkenini ne derece tahmin eder?
  3. Değişkenler arasındaki ilişkiler nelerdir ve bunlara dayanarak hangi kestirimlerde bulunulabilir?

Tahmin çalışmaları

- Eğer değişkenler arasında tatmin edici büyüklükte bir ilişki varsa, bilinen değişkenin değerinden bilinmeyen kestirilebilir.

Korelasyon katsayısı

- İki değişken arasındaki doğrusal ilişkinin derecesini ölçen bir değer olup -1 ile +1 arasında değişir.

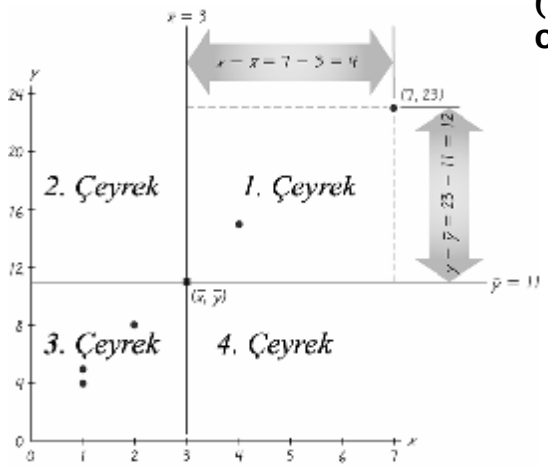
r'nin

- -1 'e yakın olması, bu değişkenler arasında çok güçlü bir negatif doğrusal ilişki olduğunu,
- +1'e yakın olması da çok güçlü pozitif doğrusal ilişki olduğunu işaret eder.
- Mutlak değer olarak r'nin 0.7 den büyük olması doğrusal ilişkinin güçlü olduğu şeklinde bir yorum yapmamıza olanak verir.

Korelasyon Katsayısı

$$r = \frac{n\mathrm{S}xy - (\mathrm{S}x)(\mathrm{S}y)}{\sqrt{n(\mathrm{S}x^2) - (\mathrm{S}x)^2}\sqrt{n(\mathrm{S}y^2) - (\mathrm{S}y)^2}}$$

Korelasyon



$(\bar{x}, \bar{y})$ **örneklerin orta noktası**

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{(n-1)\, S_x\, S_y}$$

Korelasyon Katsayısı r'nin hesaplanması

Veri				
x	1	1	3	5
y	2	8	6	4

$$r = \frac{n\mathrm{S}xy - (\mathrm{S}x)(\mathrm{S}y)}{\sqrt{n(\mathrm{S}x^2) - (\mathrm{S}x)^2}\sqrt{n(\mathrm{S}y^2) - (\mathrm{S}y)^2}}$$

$$r = \frac{4(48) - (10)(20)}{\sqrt{4(36) - (10)^2}\sqrt{4(120) - (20)^2}}$$

$$r = \frac{-8}{59.329} = -0.135$$

Basit Regresyon Analizi

Simple Linear Regression

Korelasyon vs. Regresyon

- iki değişken arasındaki ilişkiyi göstermek için Scatter plot kullanılabilir.
- Korelasyon analizi iki değişken arasındaki (lineer) ilişkinin gücünü ölçmek için kullanılır.
  - Korelasyon sadece ilişkinin gücü ile alakalıdır.
  - Korelasyon nedensellik hakkında bir şey ifade etmez

Regresyon Analizi

- Regresyon analizinde temel düşünce, bir bağımlı değişkenin bir yada birden fazla açıklayıcı değişkene istatistiksel bağımlılığını belirlemektir.
- Amacı: açıklayıcı değişkenlerin bilinen yada değişmeyen değerlerine dayanarak bağımlı değişkenin ortalama değerini tahmin etmektir.
- Başarısı: uygun verinin bulunabilmesine bağlıdır.

Regresyon ve Nedensellik

- "İstatistiksel bir ilişki, ne denli güçlü, ne denli anlamlı olursa olsun, hiçbir zaman nedensel bir ilişki olamaz; bizim nedensellik düşüncelerimiz istatistik dışından, sonuçta, şu yada bu , bir kuramdan gelmelidir".
- İstatistiksel bir ilişkinin kendi başına bir nedensellik anlamı taşımadığına dikkat edilmelidir. Bir nedensellik bulmak için kuramsal düşüncelere başvurulmalıdır.

Regresyon Analizi

- Regresyon Analizi :
  - Bağımlı değişkenin değerini en azından bir bağımsız değişkenin değerine bağlı olarak tahmin eder.
  - Bağımsız değişkende meydana gelen değişimlerin bağımlı değişkende meydana getirdiği etkileri açıklar.

Bağımlı değişken: açıklamak istediğimiz değişken

Bağımsız değişken: bağımlı değişkeni açıklamakta kullanılan değişken

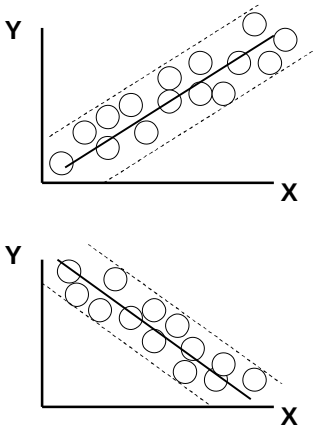
Basit Lineer Regresyon Modeli

- § Sadece bir tek bağımsız değişken, X
- § X ve Y arasındaki ilişki lineer bir fonksiyonla tanımlanmış
- § Y'deki değişime X'deki değişimin neden olduğu varsayılmakta

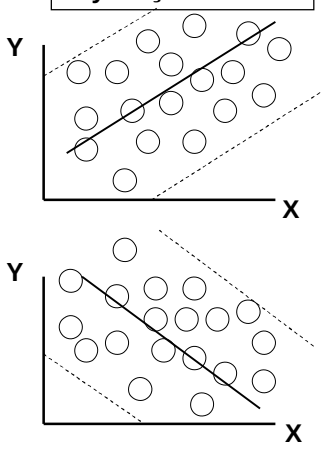
İlişki türleri

(devam)

Güçlü ilişki

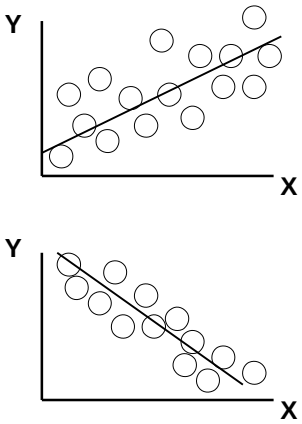


Zayıf ilişki

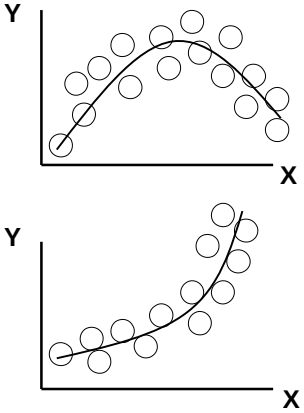


İlişki türleri

Lineer ilişki



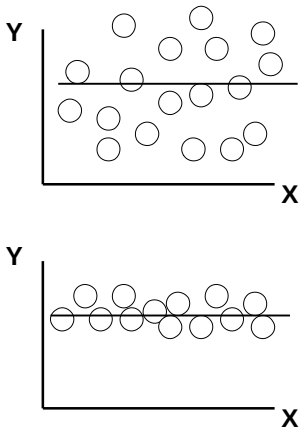
Eğrisel ilişki



İlişki türleri

(devam)

İlişki yok



Popülasyon Lineer Regresyon Eşitliği

Popülasyon regresyon modeli:

Bağımlı değişken

Popülasyon Y kesim noktası

Popülasyon eğim katsayısı

Bağımsız değişken

Rastsal hata, yada residual

$$Y_i = \underbrace{\beta_0 + \beta_1 X_i}_{\text{Lineer kısım}} + \underbrace{\varepsilon_i}_{\text{Rastsal hata kısmı}}$$

Basit Doğrusal Regresyon Modeli

§ Basit Doğrusal Regresyon Modeli

$$y = b_0 + b_1 x + e$$

§ Basit Doğrusal Regresyon Denklemi

$$E(y) = b_0 + b_1 x$$

§ Tahmini Basit Doğrusal Regresyon Denklemi

$$\hat{y} = b_0 + b_1 x$$

Örnek Lineer Regresyon Eşitliği

Basit lineer regresyon eşitliği popülasyon regresyon çizgisinin tahmin edilmesini sağlar

Tahmin edilen Y_i

Tahmin edilen Regresyon kesim noktası

Tahmin edilen regresyon eğimi

X_i değeri

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

e_i hata terimi sıfır ortalamaya sahip

En Küçük Kareler Metodu

§ En Küçük Kareler Kriteri

$$\min \sum (y_i - \hat{y}_i)^2$$

Burada :

y_i = Bağımlı değişkenin i.ninci gözlemsel değeri

$\hat{y}_i$ = Bağımlı değişkenin i.ninci gözlemsel değeri için tahmin eğeri

Minimum Kareler Metodu

§ Tahmini Regresyon denklemi için Eğim

$$b_1 = \frac{\sum x_i y_i - (\sum x_i \sum y_i)/n}{\sum x_i^2 - (\sum x_i)^2/n}$$

§ Tahmini Regresyon denklemi için *sabit terim (Doğrunun y-eksenini kestiği noktanın başlangıç noktasına uzaklığı)*

$$b_0 = \bar{y} - b_1\bar{x}$$

x_i = *i* . Gözlem için bağımsız değişkenin değeri
$\bar{y}_i$ = *i* . Gözlem için bağımlı değişken değeri
$\bar{x}$ = bağımsız değişken için ortalama değer
y = bağımlı değişken için ortalama değer
n = gözlemlerin toplam sayısı

Basit Lineer Regresyon Örnek

§ Emlakcı evin büyüklüğü ve satış fiyatı arasındaki ilişkiyi araştırmak istemektedir

§ 10 ev rasgele seçilmiştir

§ Bağımlı değişken (Y) = ev satış fiyatı $1000

§ Bağımsız değişken (X) = evin büyüklüğü ft^2

En Küçük Karelerin Bulunması Least Squares Equation

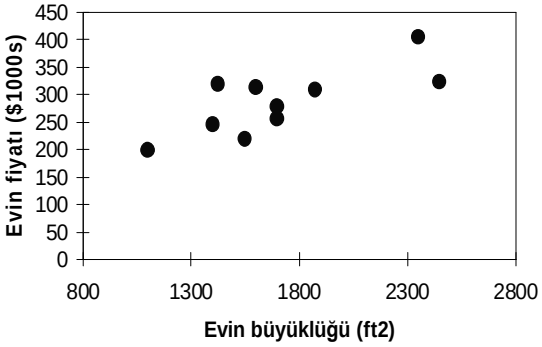
§ b_0 ve b_1 , ve diğer regresyon sonuçlarını herhangi bir istatistik programı ile kolayca hesaplamak mümkündür

Evin fiyat modeli için örnek veri

Evin fiyatı $1000s (Y)	Büyüklük (X)
245	1400
312	1600
279	1700
308	1875
199	1100
219	1550
405	2350
324	2450
319	1425
255	1700

Grafik sunumu

§ Evin fiyat modeli: scatter plot



Excel Çıktısı

Regression Statistics	
Multiple R	0.76211
R Square	0.58082
Adjusted R Square	0.52842
Standard Error	41.33032
Observations	10

Regresyon eşitliği:

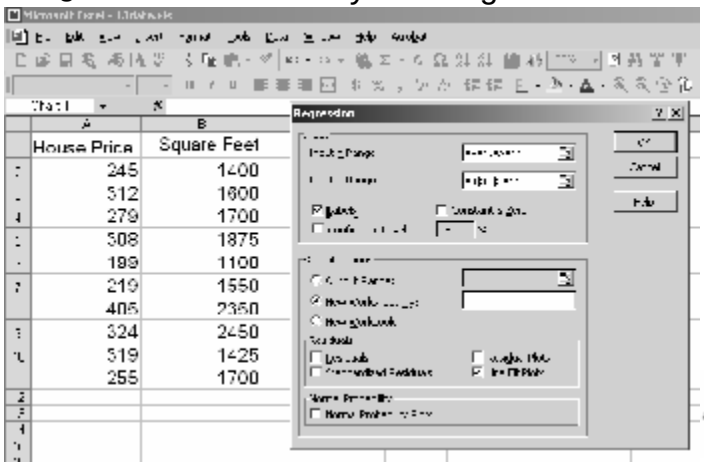
$$\widehat{\text{Evin Fiyatı}} = 98.24833 + 0.10977\ (\text{evin büyüklüğü})$$

ANOVA	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	18934.9348	18934.9348	11.0848	0.01039
Residual	8	13665.5652	1708.1957		
Total	9	32600.5000			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	98.24833	58.03348	1.69296	0.12892	-35.57720	232.07386
Square Feet	0.10977	0.03297	3.32938	0.01039	0.03374	0.18580

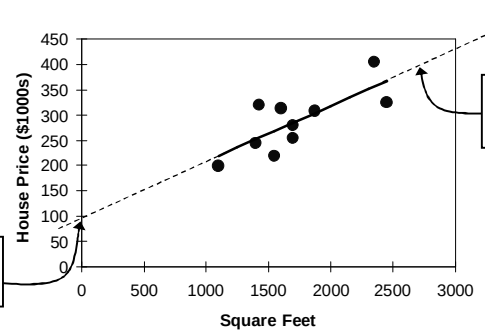
Excel'le regresyon analizi

§ Tools / Data Analysis / Regression



Grafik gösterimi

§ Evin fiyat modeli: scatter plot and regression line



Eğim = 0.10977

Kesim noktası = 98.248

$$\widehat{\text{Evin Fiyatı}} = 98.24833 + 0.10977\ (\text{evin büyüklüğü})$$

Kesim noktası, b_0'ın yorumu

$$\widehat{\text{evin fiyatı}} = \boxed{98.24833} + 0.10977\ (\text{evin büyüklüğü})$$

- b_0 X'in sıfır olduğu durumda Y'nin ortalama değerinin tahminidir. (Eğer X = 0 gözlemlenen değer aralığında ise)
  - Bu örnekte hiçbir ev 0 ft^2 olmadığı için $\boxed{b_0 = 98.24833}$ sadece gözlemlenen aralıktaki (büyüklükteki) evler için, evin büyüklüğü ile açıklanmayan kısmını belirtir

Regresyon analizi ile tahmin

2000 ft^2 bir evin fiyat tahmini:

$$\begin{aligned}\widehat{\text{evin fiyatı}} &= 98.25 + 0.1098\ (ft^2)\\ &= 98.25 + 0.1098(2000)\\ &= 317.85\end{aligned}$$

2000 ft^2 büyüklüğündeki bir evin tahmini fiyatı 317.85($1,000) = $317,850

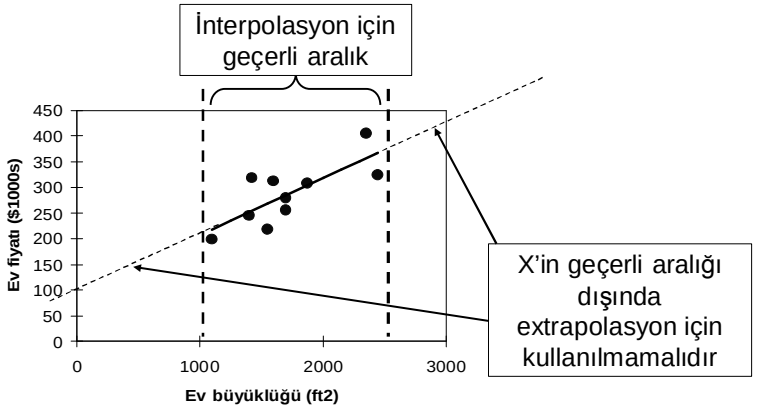
Eğim katsayısı, b_1'in yorumu

$$\widehat{\text{evin fiyatı}} = 98.24833 + \boxed{0.10977}\ (\text{evin büyüklüğü})$$

- b_1 X'in bir birim değişmesi durumunda Y'nin ortalama değerindeki değişimi ölçüsünü tahmin eder
  - Burada, $\boxed{b_1 = 0.10977}$ değeri, ev büyüklüğünün bir ft^2 artması durumunda ortalama ev fiyatında meydana gelecek artış 0.10977($1000) = $109.77 olacağını belirtir

Ara değer bulma vs. Bilinmeyene ulaşma

- Regresyon modeli tahminde kullanılacağı zaman sadece kullanılan verinin geçerli olduğu aralıkta kullanılmalıdır.



Kareler toplamı

§ SST, SSR, SSE arasındaki ilişki

(Toplam sapma) = (Açıklanan sapma) + (açıklanamayan sapma)

$$SST = SSR + SSE$$

$$\sum(y_i-\bar{y})^2=\sum(\hat{y}_i-\bar{y})^2+\sum(y_i-\hat{y}_i)^2$$

SST = Ortalamadan ayrılışların karelerinin toplamı (toplam değişkenlik ölçüsü)

SSR = regresyona dayalı kareler toplamı

SSE = hataya dayalı kareler toplamı

Excel Çıktısı

Regression Statistics	
Multiple R	0.76211
R Square	0.58082
Adjusted R Square	0.52842
Standard Error	41.33032
Observations	10

$$r^2=\frac{SSR}{SST}=\frac{18934.9348}{32600.5000}=0.58082$$

Ev fiyatındaki değişkenliğin % 58.08'i evin büyüklüğündeki değişimle açıklanmıştır.

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	18934.9348	18934.9348	11.0848	0.01039
Residual	8	13665.5652	1708.1957		
Total	9	32600.5000			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	98.24833	58.03348	1.69296	0.12892	-35.57720	232.07386
Square Feet	0.10977	0.03297	3.32938	0.01039	0.03374	0.18580

Determinasyon katsayısı, r^2

§ Determinasyon katsayısı bağımlı değişkendeki değişkenliğin bağımsız değişkenin değişkenliğiyle açıklanan kısmını ifade etmektedir

§ Determinasyon katsayısı r-squared (r^2) olarak anılmaktadır

$$r^2=\frac{SSR}{SST}=\frac{\text{regresyon kareler toplamı}}{\text{genel kareler toplamı}}$$

not: $0 \le r^2 \le 1$

Tahminin Standart Hatası

§ Regresyon doğrusu etrafındaki gözlemlerin değişkenliği (standart sapması- standart hata)

$$S_{YX}=\sqrt{\frac{SSE}{n-2}}=\sqrt{\frac{\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2}{n-2}}$$

Burada

SSE = hatanın kareler toplamı

n = örnek sayısı

Excel Çıktısı

Regression Statistics	
Multiple R	0.76211
R Square	0.58082
Adjusted R Square	0.52842
Standard Error	41.33032
Observations	10

$S_{YX} = 41.33032$

ANOVA	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	18934.9348	18934.9348	11.0848	0.01039
Residual	8	13665.5652	1708.1957		
Total	9	32600.5000			

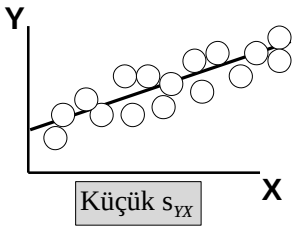
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	98.24833	58.03348	1.69296	0.12892	-35.57720	232.07386
Square Feet	0.10977	0.03297	3.32938	0.01039	0.03374	0.18580

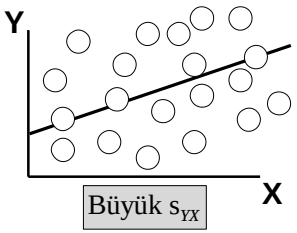
Regresyon Varsayımları

- § Hatanın normalliği
  - § Hata değerleri (ε) X'in herhangi bir verilen değeri için normal dağılım gösterir (normal olasılık dağılım çizimi ile gösterilebilinir)
- § Sabit varyans
  - § Hatanın olasılık dağılımı sabit varyans göstermelidir (scatter plot ile kolayca belirlenir)
- § Hatanın bağımsızlığı
  - § Hata değerleri istatistiksel olarak bağımsız olmalıdır (scatter plot'tan görülebilir)

Standart hatanın kıyaslaması

S_{YX} regresyon doğrusundan gözlemlenen Y değerlerindeki değişkenliğin ölçüsü

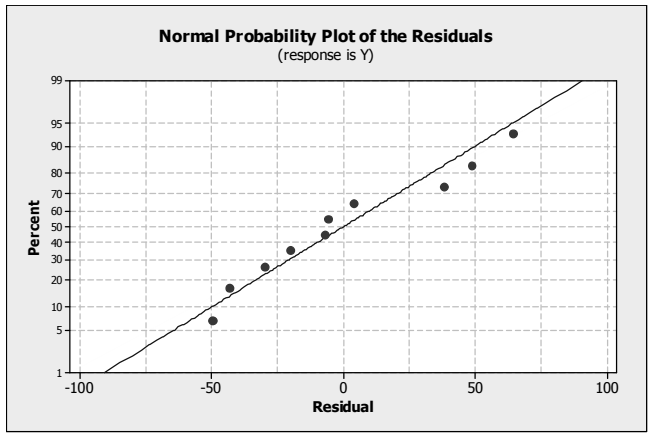




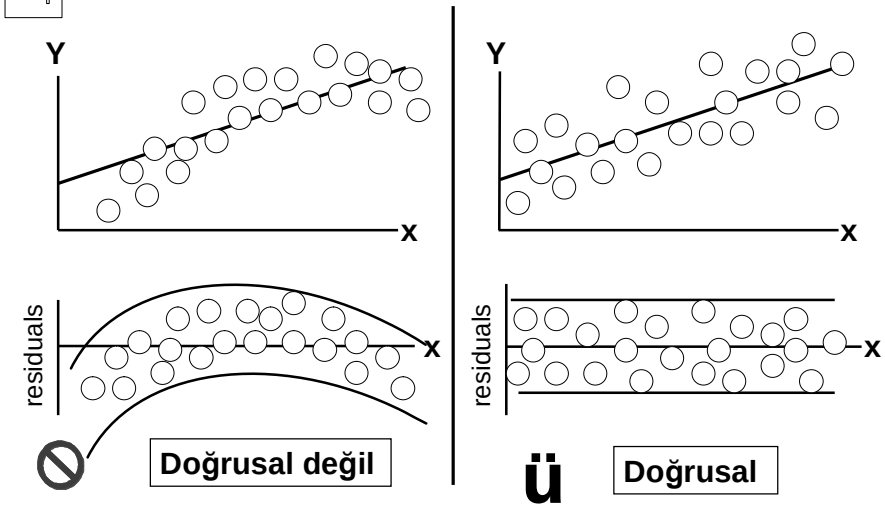
S_{YX}'nin büyüklüğü her zaman Y'nin göreceli büyüklüğü ile kıyaslanmalıdır

i.e., S_{YX} = \$41.33K evin fiyat aralığı \$200 - \$300K göre göreceli olarak küçüktür.

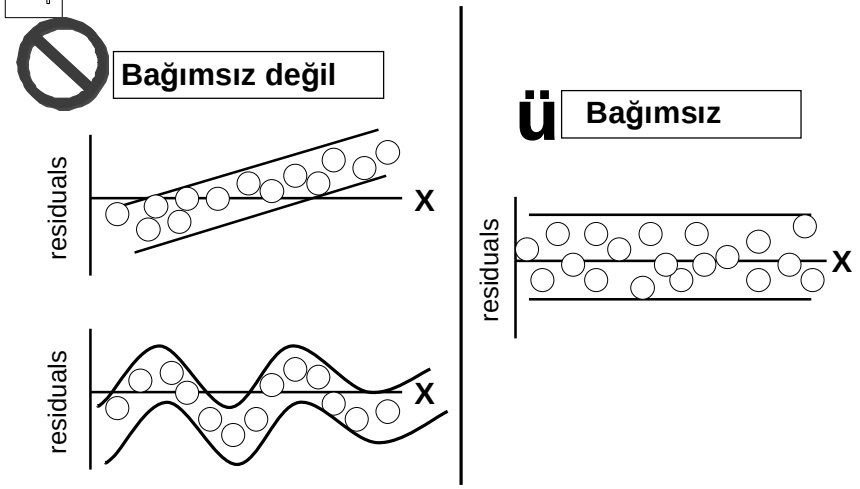
Hatanın Normalliği



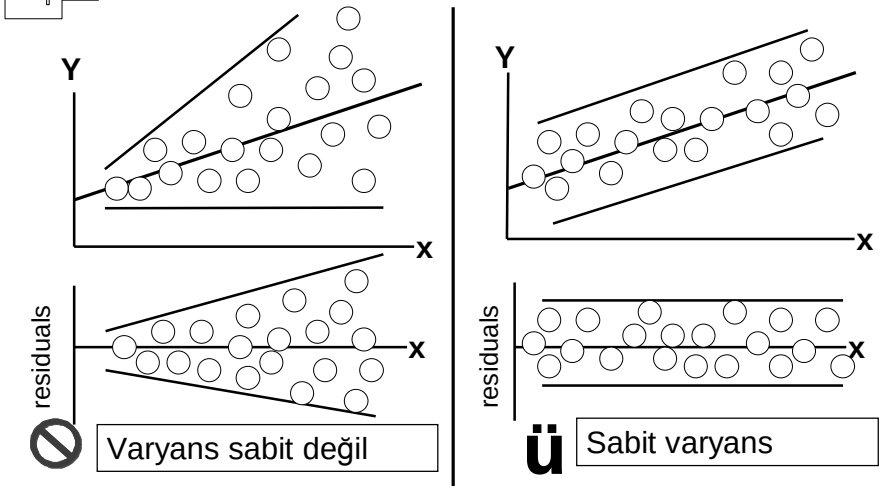
Doğrusallığın analizi



Bağımsızlık için Residual Analizi

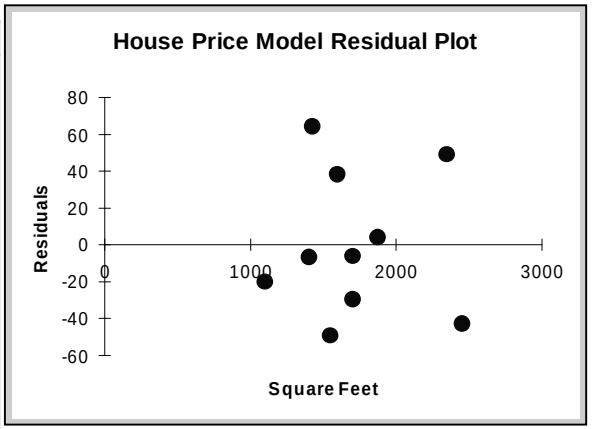


Sabit varyans için Residual analizi (Homoscedasticity)



Excel Residual Çıktısı

RESIDUAL OUTPUT		
	Tahmin edilen ev fiyatı	*Residuals*
1	251.92316	-6.923162
2	273.87671	38.12329
3	284.85348	-5.853484
4	304.06284	3.937162
5	218.99284	-19.99284
6	268.38832	-49.38832
7	356.20251	48.79749
8	367.17929	-43.17929
9	254.6674	64.33264
10	284.85348	-29.85348



Regresyon varsayımlarını ihlal eden bir durum gözükmemektedir

Oto-korelasyonun Ölçümü (Autocorrelation): Durbin-Watson istatistiği

§ Veri zaman serisi şeklinde olduğunda oto-korelasyonunun mevcudiyetini belirlemek için kullanılır.

§ Eğer her hangi bir zaman dilimindeki hata diğer zaman dilimindeki bir hata ile ilişkili ise Oto-korelasyon mevcuttur.

Durbin-Watson Testi

§ Oto-korelasyon için Durbin-Watson testi kullanılır

H_0: hatalar arasında korelasyon yoktur
H_1: oto-korelasyon mevcut

$$D = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}$$

§ alabileceği aralık is $0 \le D \le 4$

§eğer H_0 doğru ise D 2'ye yakın olmalı

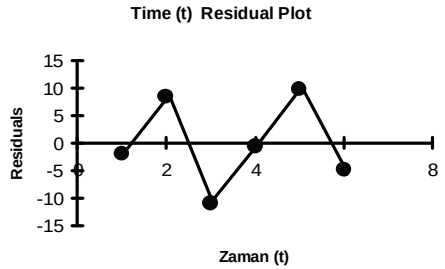
§ $D < 2$ pozitif oto-korelasyonu işaret eder,

§ $D > 2$ negatif oto-korelasyonu işaret eder

Oto-korelasyon - Autocorrelation

§ Oto-korelasyon zaman içerisinde hatalar (residuals) arası korelasyondur.

§ Burada, residuals periyodik bir yapı göstermektedir ve rastsal değildir.



§ Hatanın rastsal ve bağımsız olduğu regresyon varsayımını ihlal etmektedir

Pozitif oto-korelasyon testi

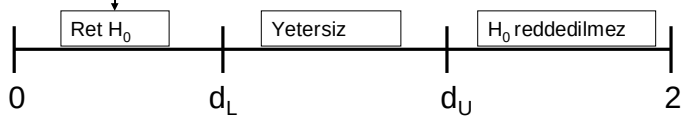
H_0: pozitif oto-korelasyon mevcut değil

H_1: pozitif oto-korelasyon mevcut

§Durbin-Watson test değerini hesapla = D
(Durbin-Watson regresyon opsiyonunda seçilerek hesaplattırılabilir.

§ Durbin-Watson tablosundan d_L değerini bul
(**n:** örnek sayısı ve **k:** bağımsız değişken)

Karar: H_0 ret eğer $D < d_L$



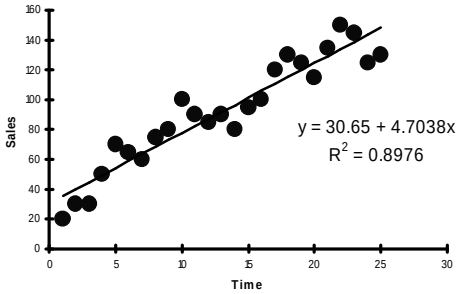
Pozitif oto-korelasyon testi

(devam)

§ Örnek boyutu n = 25:

Excel/PHStat output:

Durbin-Watson Calculations	
Sum of Squared Difference of Residuals	3296.18
Sum of Squared Residuals	3279.98
Durbin-Watson Statistic	**1.00494**



$$D = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2} = \frac{3296.18}{3279.98} = 1.00494$$

Pozitif oto-korelasyon testi

(continued)

§ Burada, n = 25 örnek ve k = 1 bağımsız değişken var

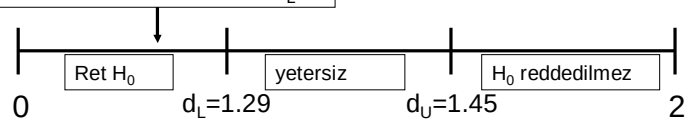
§ Durbin-Watson tablosundan, $d_L = 1.29$

§ $D = 1.00494 < d_L = 1.29$, H_0 reddedilir önemli derecede pozitif oto-korelasyonun mevcuttur

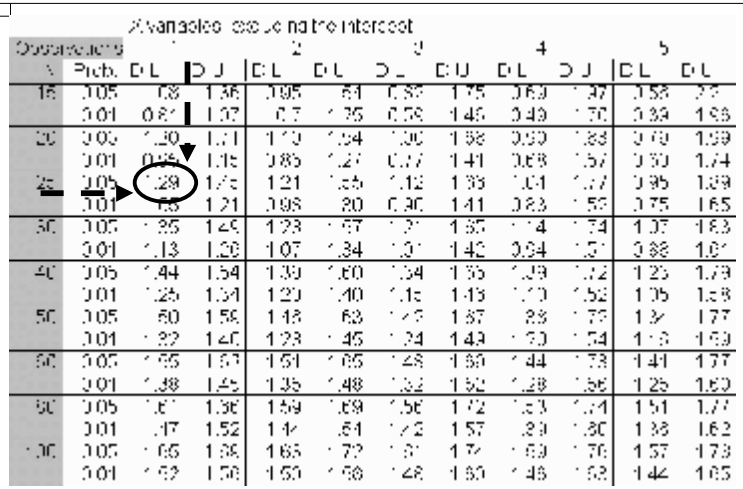
§ Bu nedenle doğrusal satış tahminde kullanmak için uygun bir model değildir

Karar: **ret H_0**

$D = 1.00494 < d_L$



Durbin-Watson tablosu



Eğim hakkında çıkarımlar: Anlamlılık testi: t Testi

§ Popülasyon eğimi için t testi

§ X ve Y arasında doğrusal bir ilişki var mı?

§ Boş ve alternatif hipotez

H_0:	$\beta_1 = 0$	(doğrusal bir ilişki yok)
H_1:	$\beta_1 \neq 0$	(doğrusal ilişki mevcut)

§ Test istatistigi

$$t = \frac{b_1 - \beta_1}{S_{b_1}}$$

Burada:

b_1 = regresyon eğim katsayısı

β_1 = eğim hipotezi

S_{b1} = eğimin standart hatası

$$d.f. = n - 2$$

Eğim hakkında çıkarımlar: t Testi

(devam)

House Price in $1000s (y)	Square Feet (x)
245	1400
312	1600
279	1700
308	1875
199	1100
219	1550
405	2350
324	2450
319	1425
255	1700

Tahmin edilen regresyon eşitliği:

$$\widehat{\text{ev fiyatı}} = 98.2483 + 0.1098\,(\text{sq.ft.})$$

Modeldeki eğim =0.1098

Evin büyüklüğü satış fiyatını etkiler mi?

Inferences about the Slope: t Test Example

(devam)

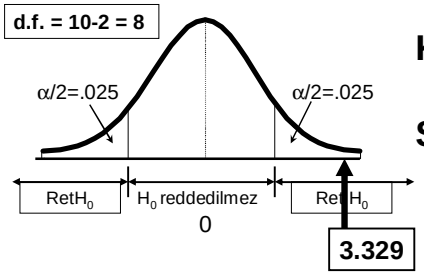
Test istatistiği: **t = 3.329**

H_0: $\beta_1 = 0$
H_1: $\beta_1 \neq 0$

Excel çıktısı:

b_1 p

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	**98.24833**	**58.03348**	**1.69296**	**0.12892**
Square Feet	**0.10977**	**0.03297**	**3.32938**	**0.01039**



Karar:

Reject H_0 at α=.05

Sonuç:

Evin büyüklüğünün satış fiyatını etkilediği konusunda yeterince bulgu mevcuttur.

Eğim hakkında çıkarımlar: t Testi örneği

H_0: $\beta_1 = 0$
H_1: $\beta_1 \neq 0$

Excel çıktısından:

b_1

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	**98.24833**	**58.03348**	**1.69296**	**0.12892**
Square Feet	**0.10977**	**0.03297**	**3.32938**	**0.01039**

$$t = \frac{b_1 - \beta_1}{S_{b_1}} = \frac{0.10977 - 0}{0.03297} = 3.32938$$

$$df = n - 2 = 8 \qquad 2 \text{ tails}$$

$$\text{TDIST}(3.32938, 8, 2) = .010393985$$

Anlamlılık Testi: *F* Testi

§ Hipotez

H_0: $b_1 = 0$
H_a: $b_1 \neq 0$

§ Test İstatistiği

F = MSR/MSE

§ Reddetme Kuralı

H_0 reddet eğer $F > F_a$

F_a , F 'nin serbestlik dereceleri,

pay için 1, payda için ise (n-2) 'dir $F_{1;\,n-2}$.

F Dağılımı

n-2

Den. df	α	Numerator df 1	2	3	4	5	6	7	8	9
8	.50	0.499	0.757	0.860	0.915	0.948	0.971	0.988	1.00	1.01
	.90	[illegible]	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
	.95	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
	.975	[illegible]	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36
	.99	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
	.995	14.7	11.0	9.60	8.81	8.30	7.95	7.69	7.50	7.34
	.999	25.4	18.5	15.8	14.4	13.5	12.9	12.4	12.0	11.8
9	.50	0.494	0.749	0.852	0.906	0.939	0.962	0.978	0.990	1.00
	.90	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
	.95	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
	.975	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03
	.99	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
	.995	13.6	10.1	8.72	7.96	7.47	7.13	6.88	6.69	6.54
	.999	22.9	16.4	13.9	12.6	11.7	11.1	10.7	10.4	10.1
10	.50	0.490	0.743	0.845	0.899	0.932	0.954	0.971	0.983	0.992
	.90	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
	.95	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
	.975	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78
	.99	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
	.995	12.8	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97
	.999	21.0	14.9	12.6	11.3	10.5	9.93	9.52	9.20	8.96
17	.50	0.484	0.735	0.835	0.888	0.921	0.943	0.959	0.972	0.981
	.90	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
	.95	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
	.975	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44
	.99	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
	.995	11.8	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20
	.999	18.6	13.0	10.8	9.63	8.89	8.38	8.00	7.71	7.48

F Testi: Ev fiyatı Örneği

- Hipotez H_0: $b_1 = 0$
  H_a: $b_1 \neq 0$
- Reddetme Kuralı
  α = 0.05 ve d.f. = 1; 8 için: $F_{.05} = 5.32$
  H_0 reddet eğer F > 5.32.
- Test İstatistiği
  F = MSR/MSE = 18935/1708 = 11.09
- Sonuç
  H_0'yu reddedebiliriz.

Eğim için güven aralığı hesabı

Eğim için güven aralığı hesabı :

$$b_1 = \mathrm{b}_1 \pm t_{n-2} S_{b_1}$$

d.f. = n - 2

Ev fiyatı için Excel Çıktısı:

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	**98.24833**	**58.03348**	**1.69296**	**0.12892**	**-35.57720**	**232.07386**
Square Feet	**0.10977**	**0.03297**	**3.32938**	**0.01039**	**0.03374**	**0.18580**

95% güven düzeyinde, eğim için güven aralığı
B: (0.0337, 0.1858)

Eğim için güven aralığı hesabı

(devam)

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	**98.24833**	**58.03348**	**1.69296**	**0.12892**	**-35.57720**	**232.07386**
Square Feet	**0.10977**	**0.03297**	**3.32938**	**0.01039**	**0.03374**	**0.18580**

Evin satış fiyat birimi $1000 olduğundan, 95% güvenle evin birim büyüklüğünün satış fiyatı üzerine ortalama etkisinin $33.70 ve $185.80 arasında olduğunu söyleyebiliriz

95% güven aralığı 0'ı içermemektedir.

Sonuç: =.05 anlamlılık düzeyinde ev fiyatı ve büyüklüğü arasında önemli bir ilişki mevcuttur.

Ortalama Y için Güven aralığı (X'in ortalaması için)

Verilen bir $\overline{X}$ değeri için **Y'nin ortalama değerinin** güven aralığı

Güven aralığı $\mu_{Y|X=X_i}$:

$$\hat{Y} \pm t_{n-2} S_{YX} \sqrt{h_i}$$

Aralığın boyutu X'in ortalaması ile olan mesafesine göre değişir

$$h_i = \frac{1}{n} + \frac{(X_i - \overline{X})^2}{SSX} = \frac{1}{n} + \frac{(X_i - \overline{X})^2}{\sum (X_i - \overline{X})^2}$$

Ortalama Y için Güven aralığı (X, herhangi bir değer için)

Verilen bir X_i değeri için **Y değerinin** güven aralığı

güven aralığı $Y_{X=X_i}$:

$$\hat{Y} \pm t_{n-2} S_{YX} \sqrt{1 + h_i}$$

Ortalama için güven aralığı hesabı Örnek

Güven aralığı $\mu_{Y|X=X_i}$

2,000 ft² evin ortalama fiyatı için 95% güven aralığını hesapla

Tahmini fiyat $\hat{Y}_i$ = 317.85 ($1,000s)

$$\hat{Y} \pm t_{n-2} S_{YX} \sqrt{\frac{1}{n} + \frac{(X_i - \overline{X})^2}{\sum (X_i - \overline{X})^2}} = 317.85 \pm 37.12$$

Güven aralığı 280.66 -- 354.90,

yada $280,660 -- $354,900

Herhangi bir değer için güven aralığı hesabı : Örnek

Güven Aralığı tahmini $Y_{X=X_i}$

2,000 ft² bir ev için 95% güven aralığı nedir?

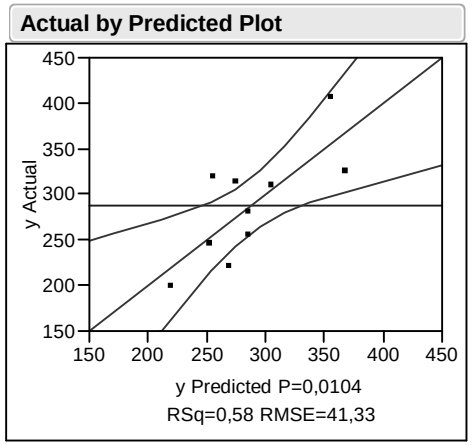
Tahmin fiyat $\hat{Y}_i$ = 317.85 ($1,000s)

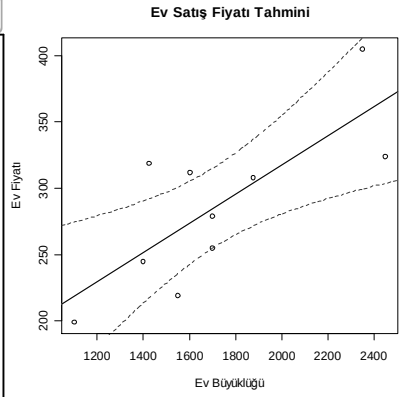
$$\hat{Y} \pm t_{n-1} S_{YX} \sqrt{1 + \frac{1}{n} + \frac{(X_i - \overline{X})^2}{\sum (X_i - \overline{X})^2}} = 317.85 \pm 102.28$$

Güven aralığı 215.50 -- 420.07, yada

$215,500 -- $420,070

Güven Aralığı





SPSS Çıktısı

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95% Confidence Interval for B Lower Bound	95% Confidence Interval for B Upper Bound
1	(Constant)	98,248	58,033		1,693	,129	-35,577	232,074
	X	,110	,033	,762	3,329	,010	,034	,186

a. Dependent Variable: Y

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	18934,935	1	18934,935	11,085	,010[a]
	Residual	13665,565	8	1708,196		
	Total	32600,500	9			

a. Predictors: (Constant), X

b. Dependent Variable: Y

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,762[a]	,581	,528	41,330	3,222

a. Predictors: (Constant), X

b. Dependent Variable: Y

Minitab Çıktısı

Regression Analysis: Y versus X
The regression equation is
Y = 98,2 + 0,110 X

Predictor	Coef	SE Coef	T	P
Constant	98,25	58,03	1,69	0,129
X	0,10977	0,03297	3,33	0,010

S = 41,3303 R-Sq = 58,1% R-Sq(adj) = 52,8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	18935	18935	11,08	0,010
Residual Error	8	13666	1708		
Total	9	32600			

Durbin-Watson statistic = 3,22166

Rgui-Lineer Regresyon analizi

Veri girişi

```
y<-c(245,312,279,308,199,219,405,324,319,255)
x<-c(1400,1600,1700,1875,1100,1550,2350,2450,1425,1700)
Ev<-data.frame(cbind(y, x))
Ev
attach(Ev)
```

Lineer regresyon

```
mod1<-lm(y~x)
summary(mod1)
anova(mod1)
```

Grafik

```
plot(x, y, xlab="Ev Büyüklüğü", ylab="Ev Fiyatı")
title("Ev Satış Fiyatı Tahmini")
abline(lm(y~x), col="green", lwd=2)
```

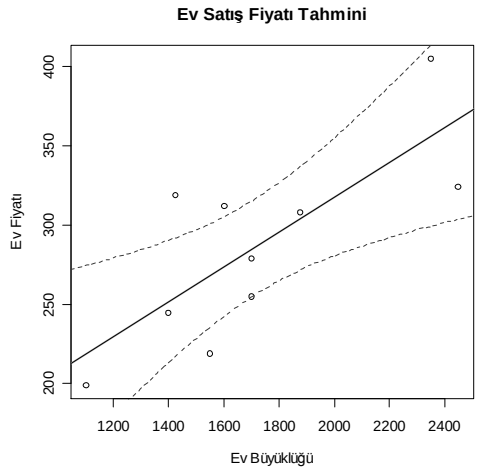
Residuals

Hata analizi

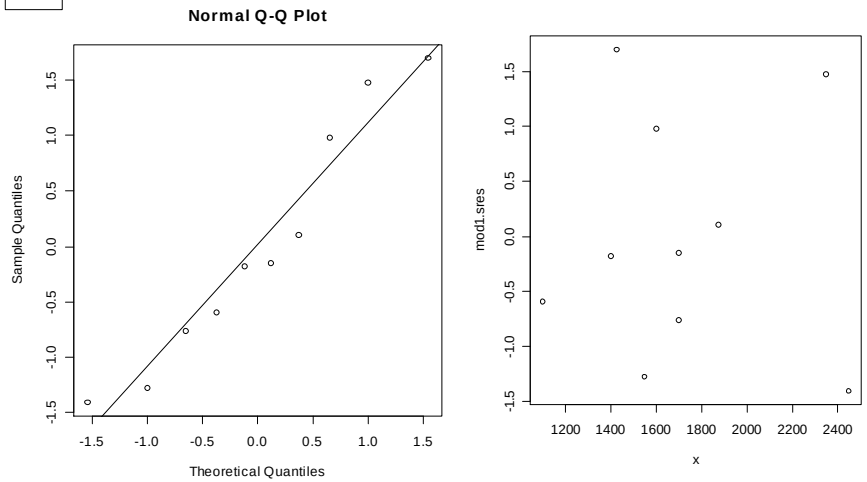
```
mod1.hat<-lm.influence(mod1)$hat #hatayı hesaplar.
mod1.res<- mod1$residuals #residuals.
mod1.sres<- mod1.res/sqrt(deviance(mod1)*(1- mod1.hat)/ mod1$df)
plot(x, mod1.sres) #standart residuals
summary(mod1.sres)
qqnorm(mod1.sres) #normallik analizi
qqline(mod1.sres) #normallik analizi
abline(0, 0, col="green", lwd=2)
plot(mod1$fitted, mod1.sres)
```

Rgui-güven aralığı çizimi

```
library(sfsmisc)
plot(x, y, xlab="Ev Büyüklüğü", ylab="Ev Fiyatı")
title("Ev Satış Fiyatı Tahmini")
abline(lm(y~x), col="green", lwd=2)
linesHyperb.lm(mod1, c.prob=0.95, conf=TRUE,
    col="red")
```



Residuals-Regresyon Hatası



Rgui Çıktısı

```
> summary(mod1)
Call:
lm(formula = y ~ x)

Residuals:
   Min      1Q     Median      3Q     Max
-49.388 -27.388  -6.388   29.577  64.333

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        98.24833   58.03348   1.693   0.1289
x                   0.10977    0.03297   3.329   0.0104 *
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 41.33 on 8 degrees of freedom
Multiple R-Squared: 0.5808,     Adjusted R-squared: 0.5284
F-statistic: 11.08 on 1 and 8 DF,  p-value: 0.01039
```

Regresyon Analizinde Görünmez Tehlikeler

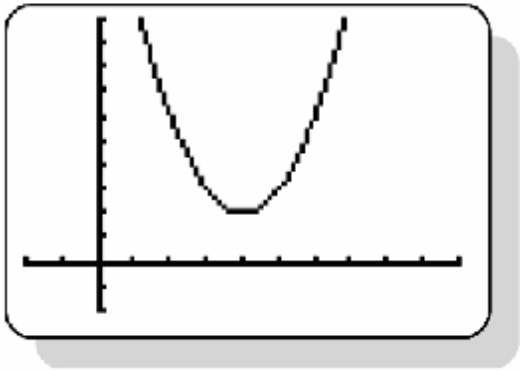
- § EKK (OLS) regresyon analizinin temel varsayımlarının farkında olunmaması
- § Varsayımların nasıl değerlendirileceğinin bilinmemesi
- § Her hangi bir varsayımın ihlal edilmesi durumunda regresyon alternatifinin bilinmemesi
- § Regresyon yöntemi ile modellenecek konu hakkında bilgi sahibi olunmaması
- § Uygun aralık dışında extrapolasyon (ileriye dönük tahmin) için kullanılması

Fonksiyon Formları

- § **Lineer:** $y = a + bx$
- § **İkinci derece:** $y = ax^2 + bx + c$
- § **Logaritmik:** $y = a + b \ln x$
- § **Eksponansiyel:** $y = ab^x$
- § **Üstel:** $y = ax^b$
- § **Lojistik:** $y = \dfrac{c}{1 + ae^{-bx}}$

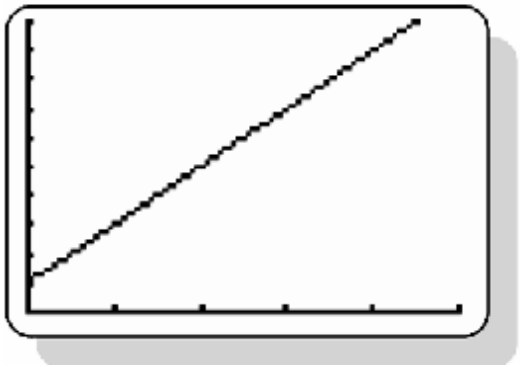
Fonksiyon Formları

Quadratic: $y = 2x^2 - 8x + 9$



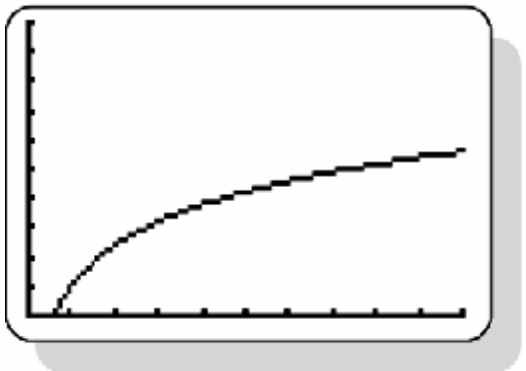
Fonksiyon Formları

Linear: $y = 1 + 2x$



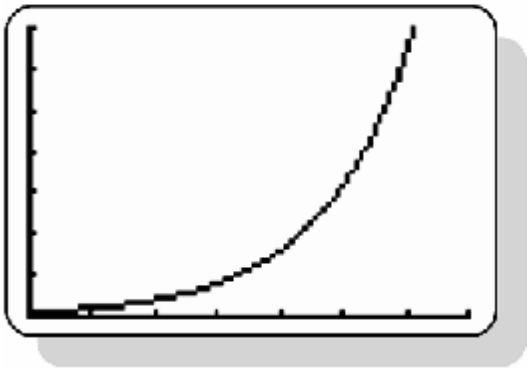
Fonksiyon Formları

Logarithmic: $y = 1 + 2\ln x$



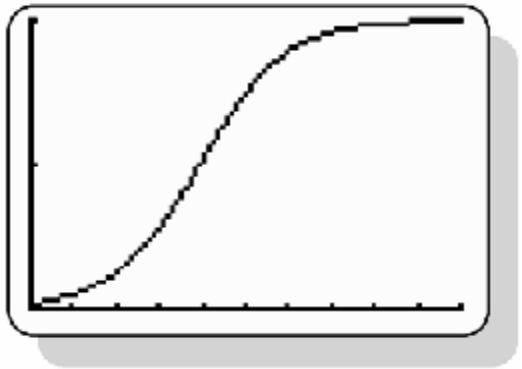
Fonksiyon Formları

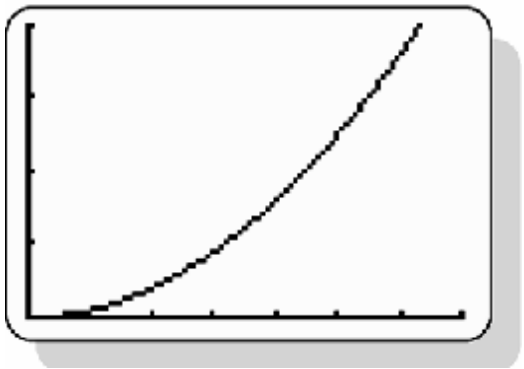
Exponential: $y = 2^x$



Fonksiyon Formları

Logistic: $y = \dfrac{2}{1 + 50e^{-x}}$



Fonksiyon Formları

Power: $y = x^2$