

# Bölüm 13

## Basit Doğrusal Regresyon



# Öğrenme Hedefleri

---

## **Bu bölümde aşağıdaki konulara değinilecektir:**

- Bağımsız bir değişkene dayalı olarak bağımlı bir değişkenin değerinin tahmini için regresyon analizi nasıl kullanılır
- $b_0$  ve  $b_1$  regresyon katsayılarının anlamı
- Regresyon analizinin varsayımlarının nasıl değerlendirileceği ve varsayımlar ihlal edildiğinde ne yapılması gerektiği
- Eğim ve korelasyon katsayısı ile ilgili çıkarımlar yapmak
- Ortalama değerlerin tahmin edilmesi ve bağımsız değerlerin kestirilmesi



# Korelasyon İle Regresyon İlişkisi

---

- İki değişkenin arasındaki ilişkinin gösterimi için bir **serpme diyagramı** kullanılabilir
- **Korelasyon** analizi iki değişkenin arasındaki birlikteliğin (doğrusal ilişki) gücünün ölçülmesi için kullanılır
  - Korelasyon sadece ilişkinin gücüyle ilgilenmektedir
  - Korelasyon ile hiçbir nedensel etki ifade edilmez
  - Serpme diyagramları ilk defa bölüm. 2'de gösterilmiştir
  - Korelasyon ilk defa bölüm. 3'te gösterilmiştir



# Regresyon Analizine Giriş

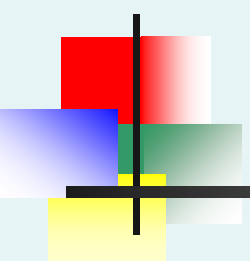
---

## ■ Regresyon analizi :

- En az bir bağımsız değişkenin değerine dayalı olarak bağımsız bir değişkenin değerini tahmin etmek,
- Bağımsız bir değişkendeki değişimin bağımsız bir değişken üzerindeki etkisini açıklamak, için kullanılır

**Bağımlı değişken:** tahmin etmek veya açıklamak istediğimiz değişken

**Bağımsız değişken:** bağımlı değişkeni tahmin etmek veya açıklamak için kullanılan değişken



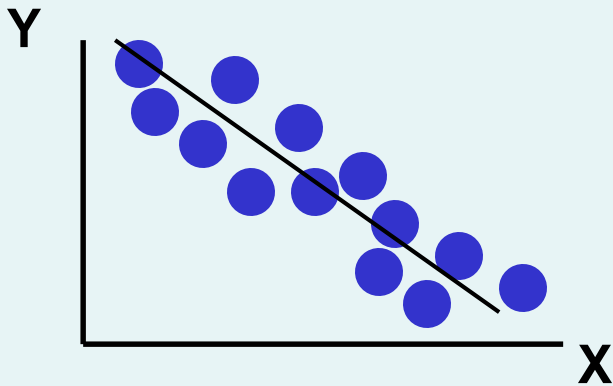
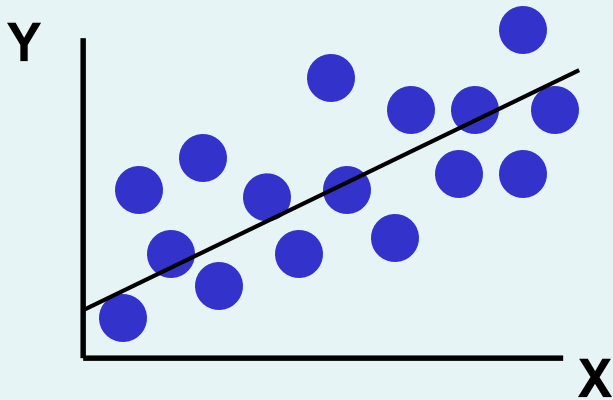
# Basit Doğrusal Regresyon Modeli

---

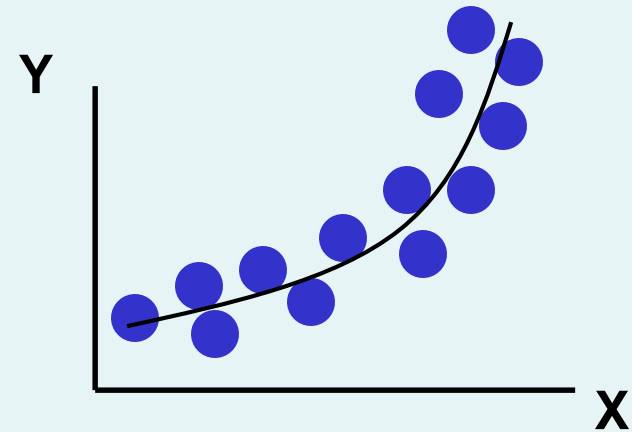
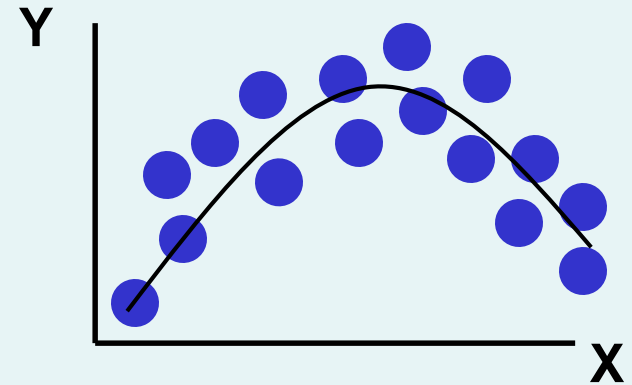
- Sadece **bir** bağımsız değişken,  $X$
- $X$  ve  $Y$  arasındaki ilişki bir doğrusal fonksiyon ile tanımlanır
- $Y$ 'deki değişimlerin  $X$ 'deki değişimler ile bağlantılı olduğu varsayımı vardır

# İlişki Tipleri

## Doğrusal İlişkiler



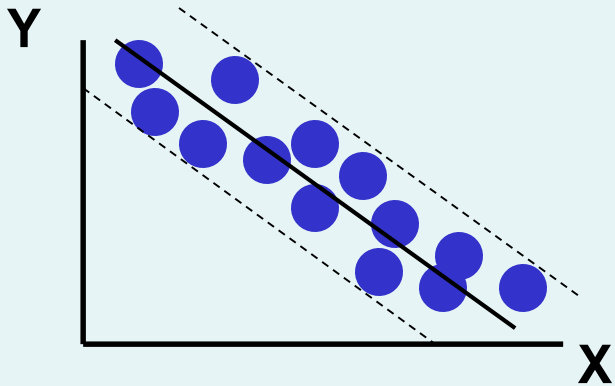
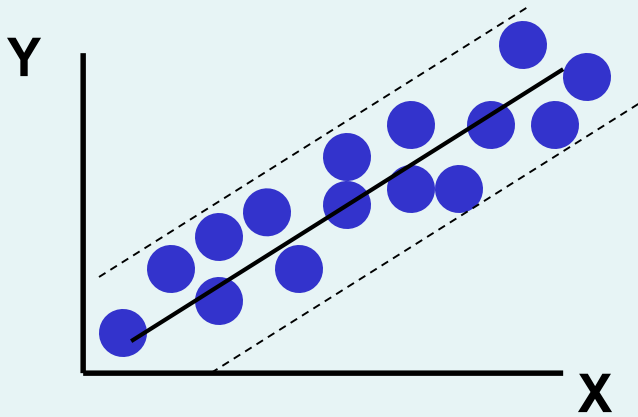
## Eğrisel İlişkiler



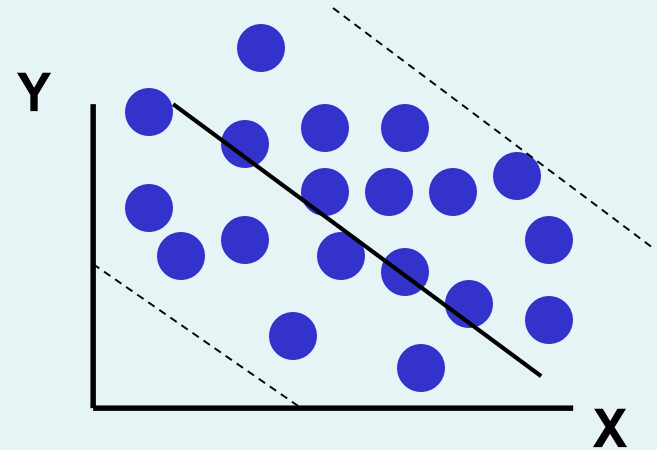
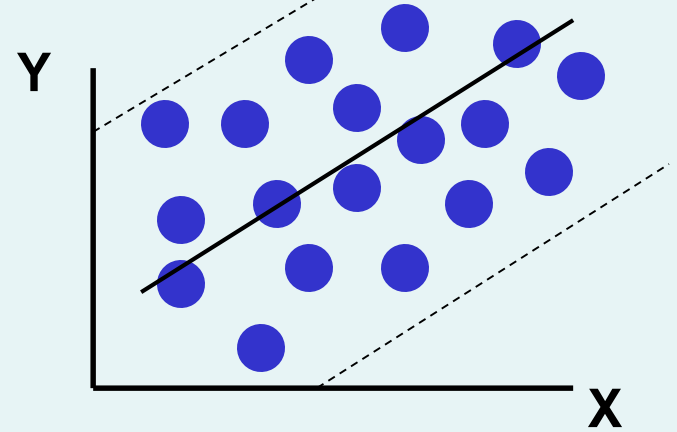
# İlişki Tipleri

(devamı)

## Kuvvetli İlişkiler



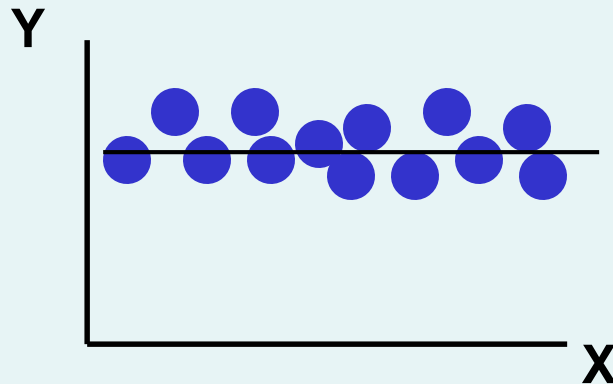
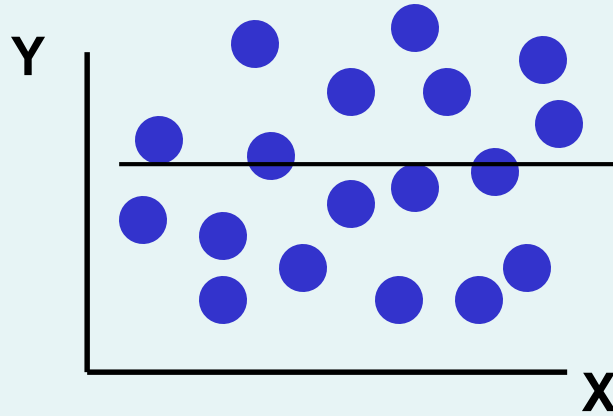
## Zayıf İlişkiler



# İlişki Tipleri

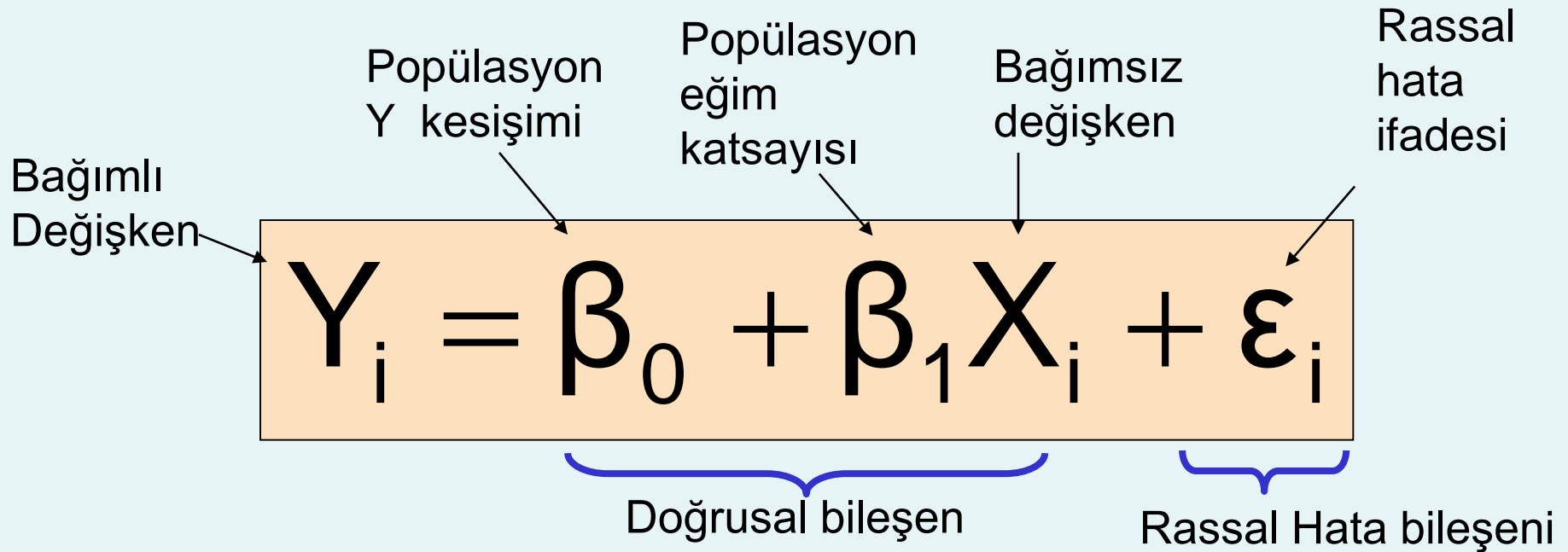
(devamı)

İlişki yoktur





# Basit Doğrusal Regresyon Modeli



The diagram illustrates the Simple Linear Regression Model equation,  $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ , with labels for each component:

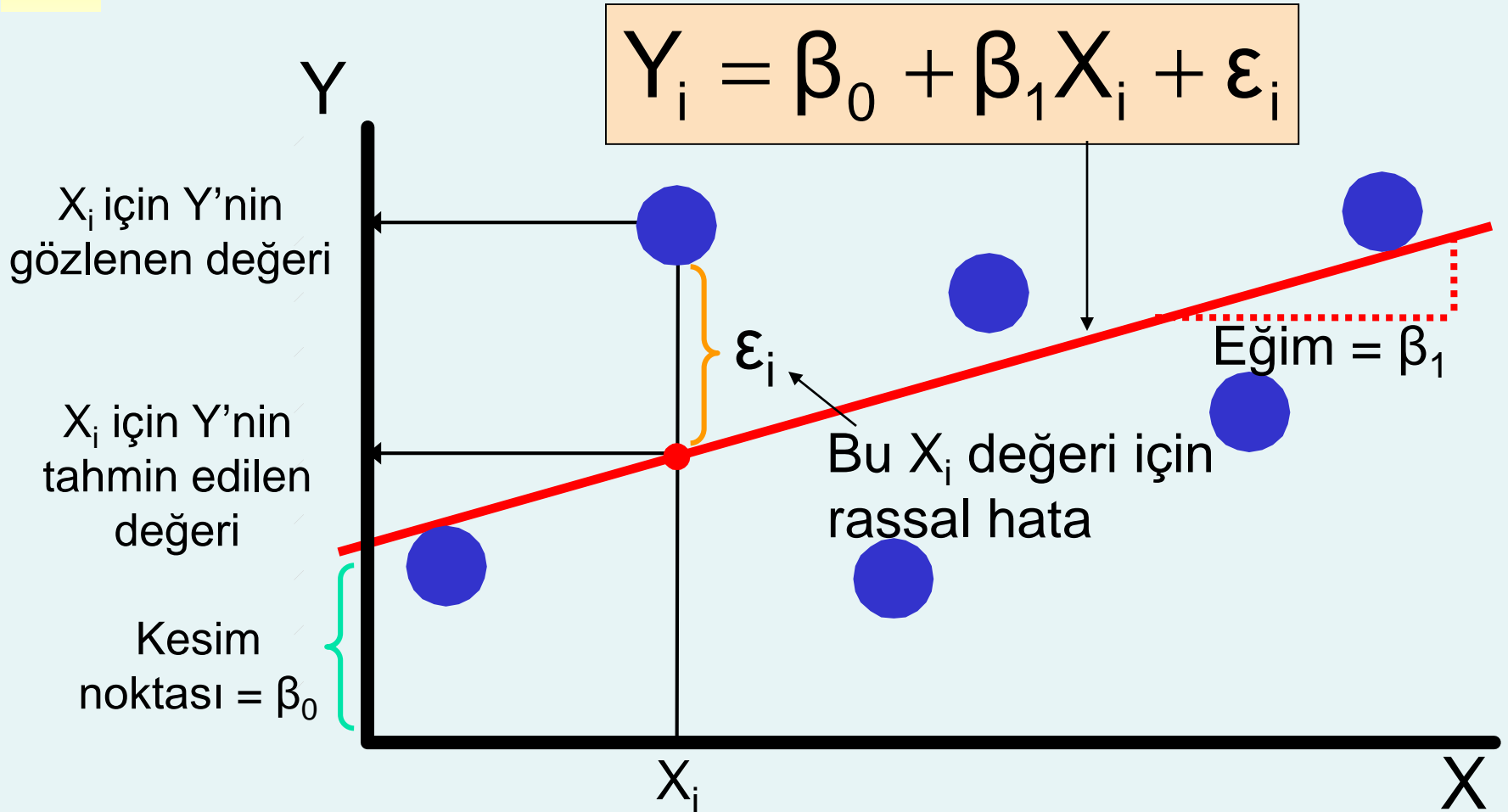
- Bağımlı Değişken** (Dependent Variable) points to  $Y_i$ .
- Popülasyon Y kesişimi** (Population Y intercept) points to  $\beta_0$ .
- Popülasyon eğim katsayısı** (Population slope coefficient) points to  $\beta_1$ .
- Bağımsız değişken** (Independent Variable) points to  $X_i$ .
- Rassal hata ifadesi** (Random error expression) points to  $\epsilon_i$ .

Below the equation, two blue brackets group the terms:

- Doğrusal bileşen** (Linear component) groups  $\beta_0 + \beta_1 X_i$ .
- Rassal Hata bileşeni** (Random error component) groups  $\epsilon_i$ .

# Basit Doğrusal Regresyon Modeli

(devamı)



# Basit Doğrusal Regresyon Denklemi (Tahmin Çizgisi)

Basit doğrusal regresyon denklemi popülasyon regresyon çizgisinin bir **tahmininin** yapılmasını sağlar

$i$  gözlemi için  
tahmin  
edilmiş  $Y$   
değeri

Regresyonun  
 $Y$  eksenini  
kesme  
noktasının  
tahmini

Regresyon  
eğiminin tahmini

$i$  gözlemi için  
 $X$  değeri

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



# En Küçük Kareler Yöntemi

---

$b_0$  ve  $b_1$ ,  $Y$  ve  $\hat{Y}$  arasındaki fark karelerinin toplamını en küçük yapan değerlerin bulunmasıyla elde edilir :

$$\min \sum (Y_i - \hat{Y}_i)^2 = \min \sum (Y_i - (b_0 + b_1 X_i))^2$$



# En Küçük Kareler Denkleminin Bulunması

---

- $b_0$  ve  $b_1$  katsayıları ve bu bölümdeki diğer regresyon sonuçları Excel kullanılarak bulunacaktır

İlgi duyanlar için formüller kitaptaki bölümde verilmiştir.



# Eğim ve Kesme noktasının Çıkarımı

---

- $b_0$  ,  $X$ 'in değeri sıfır olduğunda  $Y$ 'nin beklenen ortalama değeridir
- $b_1$  ,  $X$ 'in değerindeki bir birimlik artış sonucu olarak  $Y$ 'nin ortalama değerindeki beklenen değişim miktarıdır

# Basit Doğrusal Regresyon Örnek

- Bir emlak acentesi bir evin satış fiyatı ile boyutu (feet-kare olarak ölçülmüştür) arasındaki ilişkiyi ölçmek istemektedir.
- Rassal olarak 10 evlik bir örnek seçilmiştir
  - Bağımlı değişken( $Y$ ) = \$1000 cinsinden ev fiyatı
  - Bağımsız değişken( $X$ ) = feet-kare (ev boyutu)



# Basit Doğrusal Regresyon

## Örnek: Veri

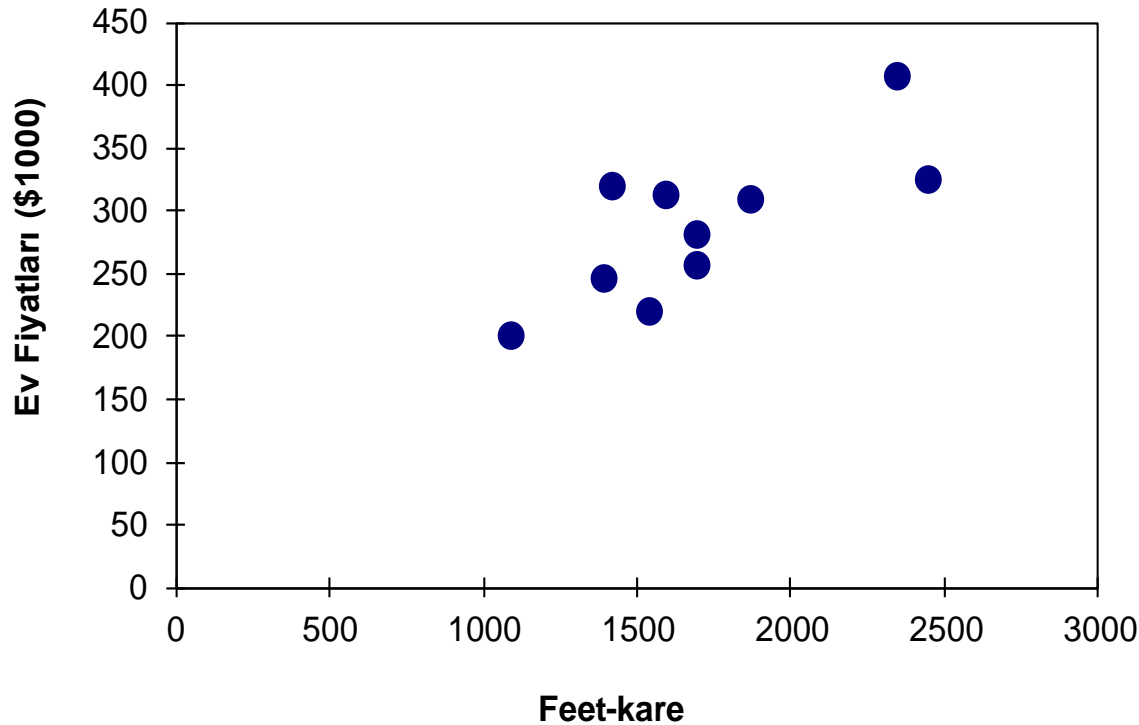
<b>\$1000 cinsinden ev fiyatları (Y)</b>	<b>Feet-kare (ev boyutu) (X)</b>
245	1400
312	1600
279	1700
308	1875
199	1100
219	1550
405	2350
324	2450
319	1425
255	1700





# Basit Doğrusal Regresyon Örneği: Serpme Diyagramı

## Ev fiyatı Modeli: Serpme Diyagramı



# Basit Doğrusal Regresyon Örnek: Excel Veri Çözümleyici Fonksiyonu Kullanılarak

1. Veri'yi Seç

2. Veri çözümleyici'yi seç

3. Regresyonu Seç

The screenshot shows the Microsoft Excel interface with the 'Data' tab selected. The 'Data Analysis' button is highlighted in the 'Data Tools' group. A red arrow points from the 'Data Analysis' button to the 'Data Analysis' dialog box. The dialog box is open, and the 'Regression' option is selected in the list of analysis tools. A red arrow points from the 'Regression' option to the 'OK' button. The background spreadsheet shows data for 'House Price' and 'Square Feet'.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	House Price	Square Feet																
2	245	1400																
3	312	1600																
4	279	1700																
5	308	1875																
6	199	1100																
7	219	1550																
8	405	2350																
9	324	2450																
10	319	1425																
11	255	1700																
12																		



# Basit Doğrusal Regresyon Örnek: Excel Veri Çözümleyici Fonksiyonu Kullanılarak

(devamı)

Y aralığını ve X aralığını ve gerekli seçenekleri girin

	A	B	C	D	E	F	G	H	I
1	House Price	Square Feet							
2	245	1400							
3	312	1600							
4	279	1700							
5	308	1875							
6	199	1100							
7	219	1550							
8	405	2350							
9	324	2450							
10	319	1425							
11	255	1700							
12									
13									
14									
15									
16									
17									
18									
19									
20									

**Regression**

Input

Input Y Range: \$A\$2:\$A\$11

Input X Range: \$B\$2:\$B\$11

☐ Labels ☐ Constant is Zero

☐ Confidence Level: 95 %

Output options

☒ Output Range: \$D\$1

☐ New Worksheet Ply:

☐ New Workbook

Residuals

☐ Residuals ☐ Residual Plots

☐ Standardized Residuals ☐ Line Fit Plots

Normal Probability

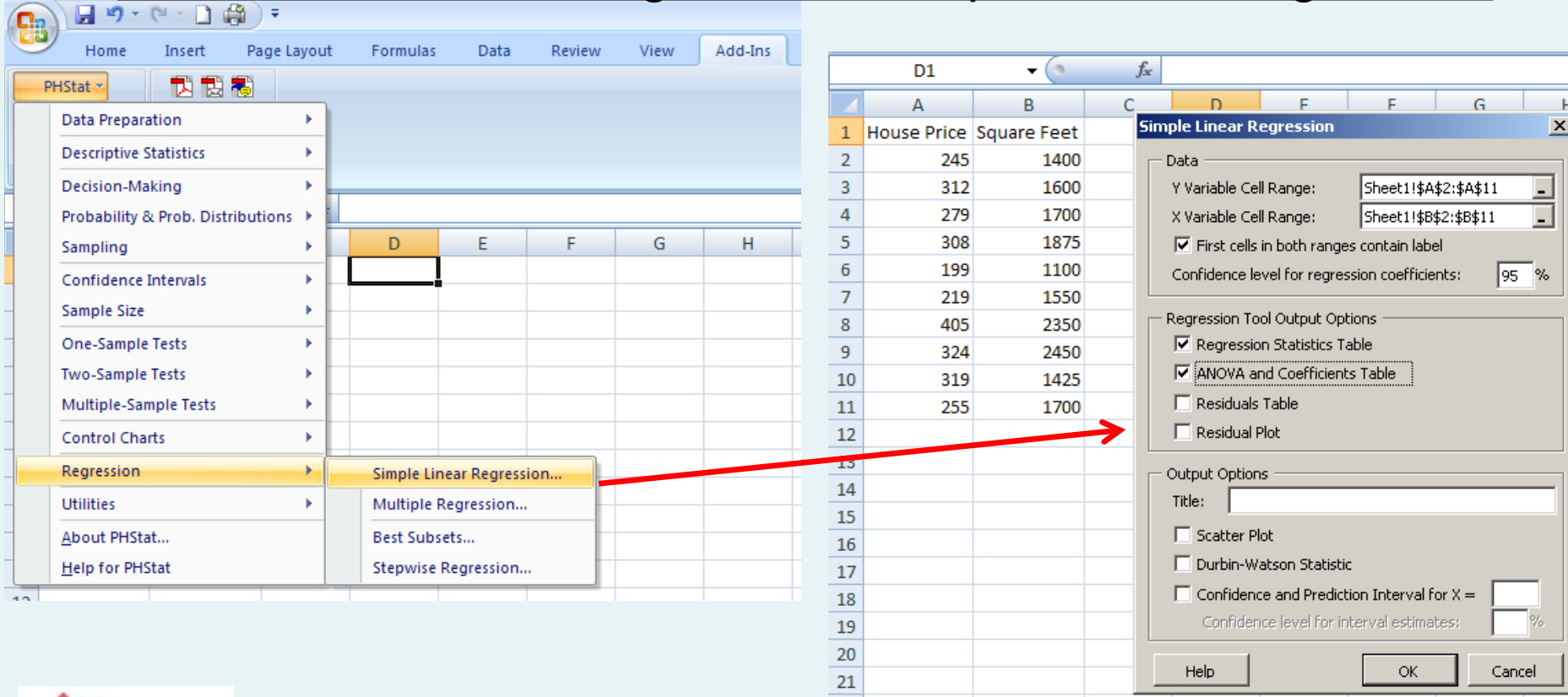
☐ Normal Probability Plots

OK Cancel Help



# Basit Doğrusal Regresyon Örnek: PHStat Kullanarak

Eklentiler: PHStat: Regression: Simple Linear Regression



The screenshot shows the PHStat add-in menu in Microsoft Excel. The 'Regression' option is selected, and the 'Simple Linear Regression...' dialog box is open. The dialog box is configured with the following settings:

- Data:**
  - Y Variable Cell Range: Sheet1!\$A\$2:\$A\$11
  - X Variable Cell Range: Sheet1!\$B\$2:\$B\$11
  - ☒ First cells in both ranges contain label
  - Confidence level for regression coefficients: 95 %
- Regression Tool Output Options:**
  - ☒ Regression Statistics Table
  - ☒ ANOVA and Coefficients Table
  - ☐ Residuals Table
  - ☐ Residual Plot
- Output Options:**
  - Title: (empty field)
  - ☐ Scatter Plot
  - ☐ Durbin-Watson Statistic
  - ☐ Confidence and Prediction Interval for X = (empty field) %
  - Confidence level for interval estimates: (empty field) %

A red arrow points from the 'Simple Linear Regression...' option in the PHStat menu to the dialog box.

	A	B	C	D	F	G	H
1	House Price	Square Feet					
2	245	1400					
3	312	1600					
4	279	1700					
5	308	1875					
6	199	1100					
7	219	1550					
8	405	2350					
9	324	2450					
10	319	1425					
11	255	1700					
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							



# Basit Doğrusal Regresyon Örnek: Excel Çıktısı

## Regression Statistics

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

Regresyon denklemi:

$$\text{ev fiyatı} = 98.24833 + 0.10977 (\text{feet} - \text{kare})$$

## ANOVA

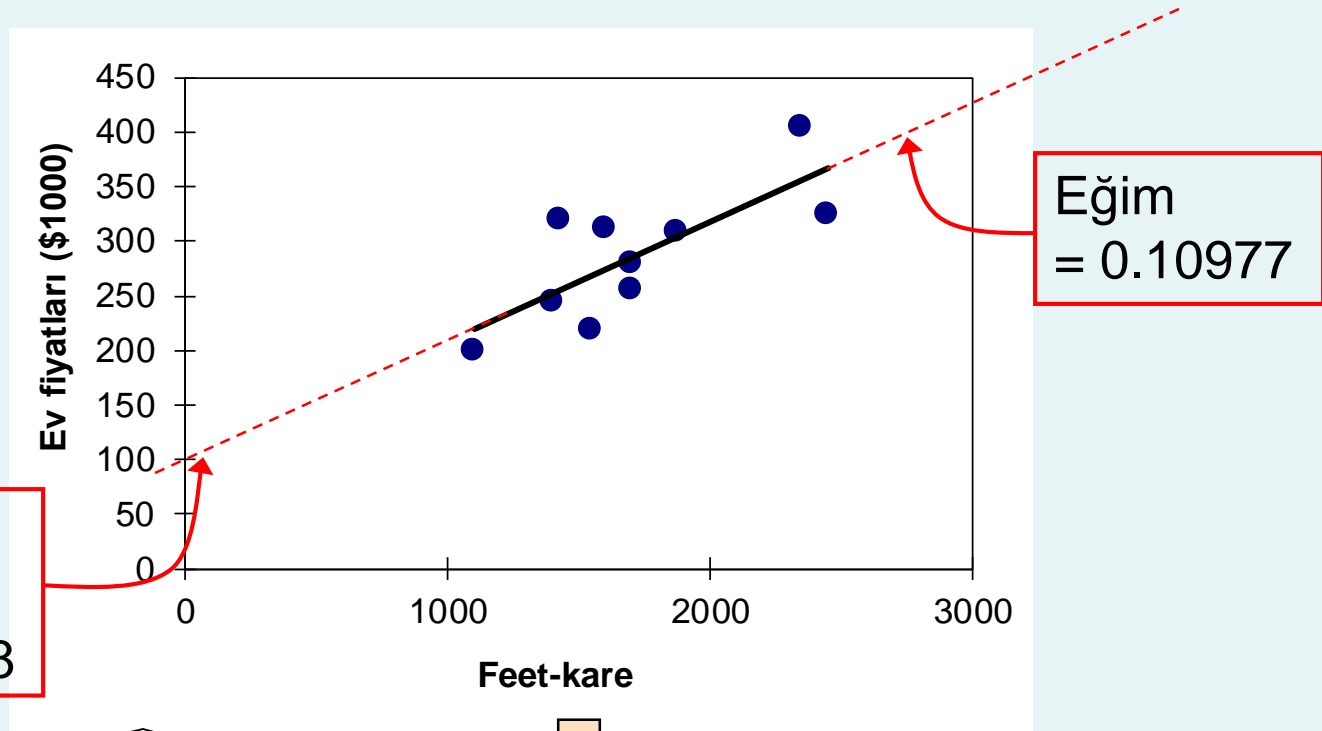
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	18934.9348	18934.9348	11.0848	0.01039
Residual	8	13665.5652	1708.1957		
Total	9	32600.5000			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
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



# Basit doğrusal Regresyon Örnek: Grafik Gösterim

Ev fiyatı modeli: Serpme diyagramı ve tahmin çizgisi



$$\text{ev fiyatı} = 98.24833 + 0.10977 (\text{feet - kare})$$

# Basit Doğrusal Regresyon Örnek: $b_0$ 'ın yorumlanması

$$\widehat{\text{ev fiyatı}} = 98.24833 + 0.10977 (\text{feet - kare})$$

- $b_0$  ,  $X$ 'in değeri sıfır olduğunda  $Y$ 'nin almış olduğu ortalama değerdir (Eğer  $X = 0$   $X$ 'in alabileceği değer aralığında ise)
- Bir evin alanı sıfır olamayacağı için,  $b_0$  'ın pratiğe dönük bir uygulaması yoktur.



# Basit Doğrusal Regresyon Örnek: $b_1$ 'in yorumlanması

$$\widehat{\text{Ev fiyatı}} = 98.24833 + 0.10977 (\text{feet - kare})$$

- $b_1$  , X'deki bir birimlik artış sonucunda Y'nin ortalama değerindeki değişimi tahmin eder.
  - Burada,  $b_1 = 0.10977$ , bize ev alanındaki bir foot-karelik artışın, bir evin ortalama fiyatının ortalama olarak  $.10977(\$1000) = \$109.77$  artacağını söylemektedir.





# Basit Doğrusal Regresyon

## Örnek: Çıkarımların Yapılması

2000 feet-kare olan bir evin fiyatı hakkında çıkarım yapalım:

$$\begin{aligned}\text{ev fiyatı} &= 98.25 + 0.1098 (\text{feet} - \text{kare}) \\ &= 98.25 + 0.1098(2000) \\ &= 317.85\end{aligned}$$

2000 feet-kare alana sahip bir evin tahmin edilen fiyatı  $317.85(\$1,000\text{s}) = \$317,850$ 'dır

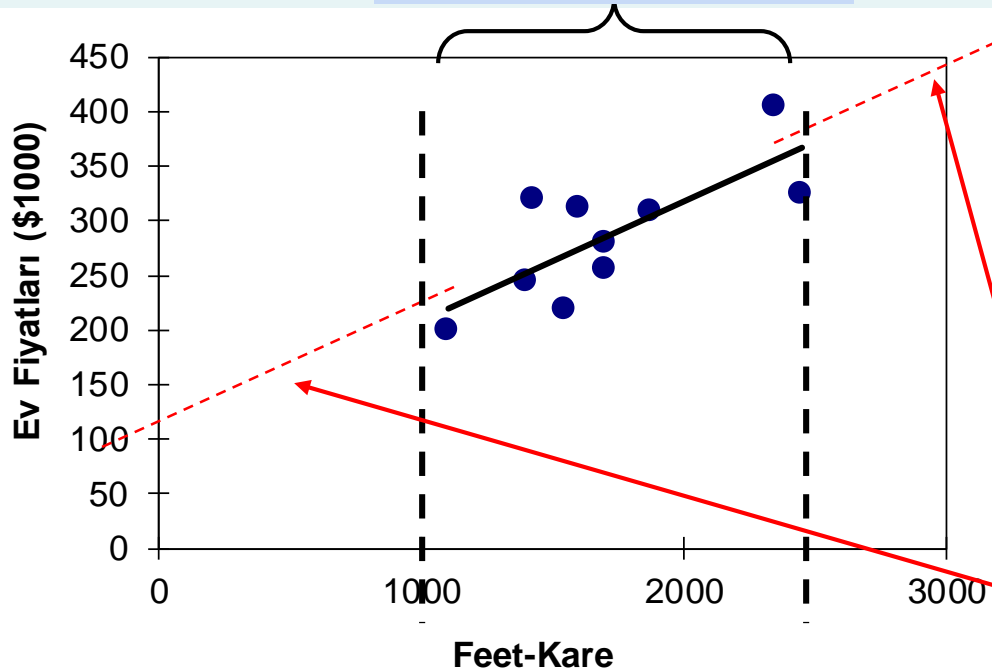


# Basit Doğrusal Regresyon

## Örnek: Çıkarımların Yapılması

- Bir regresyon modeli çıkarım yapmak için kullanılırken, sadece verinin ilgili aralığı içerisinde kullanılmalıdır

Interpolasyon için  
ilgili aralık



X'in gözlemlenen  
aralığı dışında  
ekstrapolasyon  
yapılmaya  
çalışılmamalıdır



# Değişim Ölçüleri

- Toplam Değişim iki bölümden oluşmaktadır:

$$TKT = RKT + HKT$$

Toplam Kareler  
Toplamı

Regresyon  
Kareler Toplamı

Hata Kareler  
toplamı

$$TKT = \sum (Y_i - \bar{Y})^2$$

$$RKT = \sum (\hat{Y}_i - \bar{Y})^2$$

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

$\bar{Y}$  = Bağımlı değişkenin ortalama değeri

$Y_i$  = Bağımlı değişkenin gözlenen değeri

$\hat{Y}_i$  = Verilen  $X_i$  değerine göre  $Y$ 'nin tahmin edilen değeri

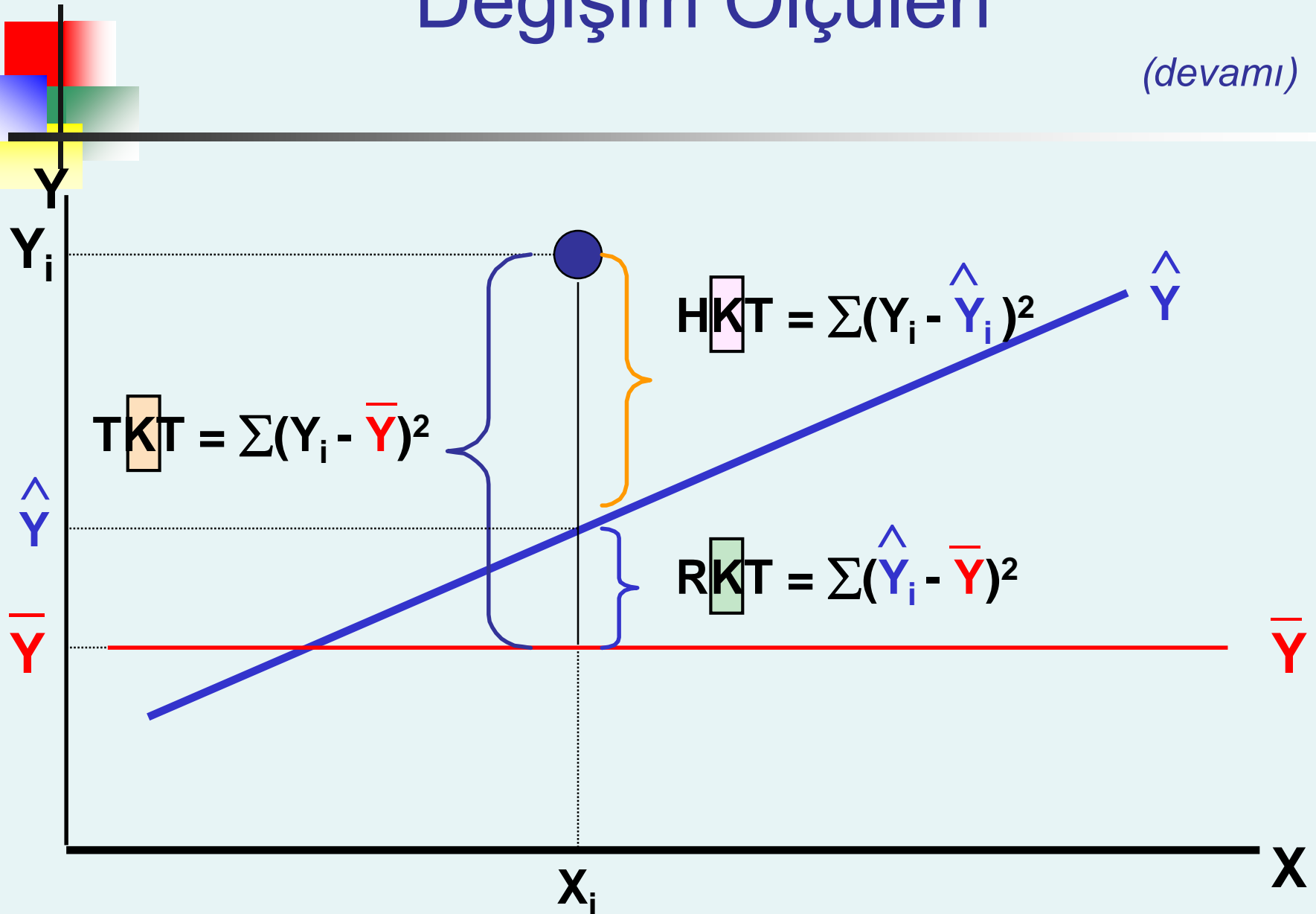
# Değişim Ölçüleri

(devamı)

- TKT = toplam kareler toplamı (Toplam değişim)
  - $Y_i$  değerlerinin, ortalama değerleri olan  $\bar{Y}$  etrafındaki değişim miktarını ölçer
- RKT = regresyon kareler toplamı (Açıklanmış değişim)
  - X ve Y arasındaki ilişkiye atfedilen değişim
- HKT = hata kareler toplamı (Açıklanmamış değişim)
  - X'den başka diğer faktörlere atfedilen Y'deki değişim

# Değişim Ölçüleri

(devamı)





# Determinasyon Katsayısı, $r^2$

- **Determinasyon katsayısı**, bağımlı değişkende toplam değişimin bağımsız değişkendeki değişim tarafından açıklanan kısmıdır
- Determinasyon katsayısı **r-kare** olarak da adlandırılır ve  **$r^2$**  olarak gösterilir.

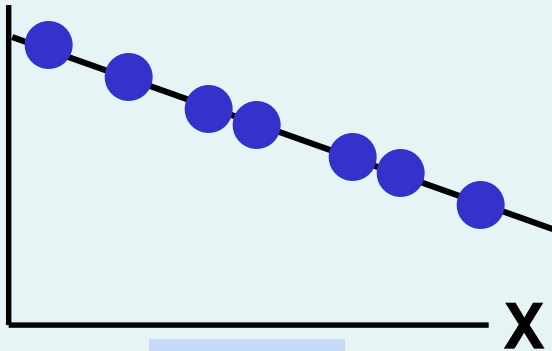
$$r^2 = \frac{RKT}{TKT} = \frac{\text{regresyon kareler toplamı}}{\text{toplam kareler toplamı}}$$

not:

$$0 \leq r^2 \leq 1$$

# Yaklaşık $r^2$ Değerlerine Örnekler

Y

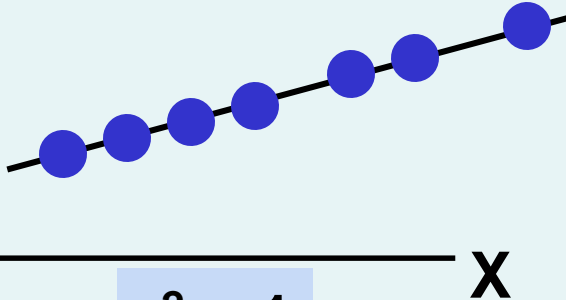


$$r^2 = 1$$

$$r^2 = 1$$

**X ve Y arasında kusursuz bir ilişki mevcuttur:**

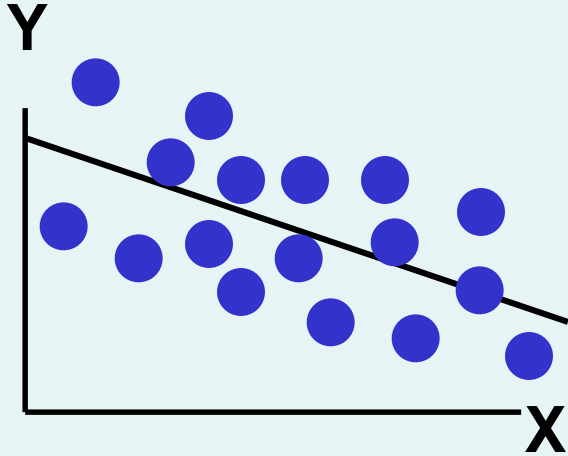
Y



$$r^2 = 1$$

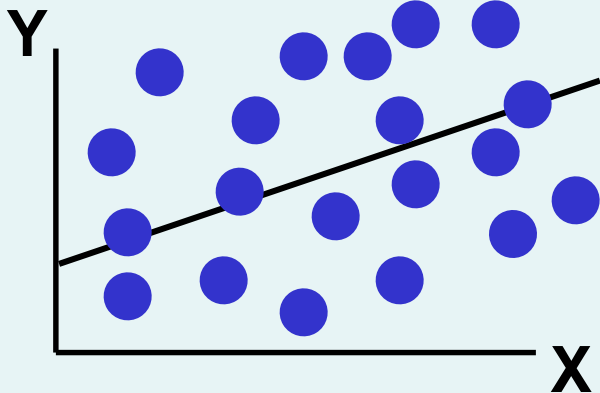
**Y'deki değişimin %100'ü, X'deki değişim ile açıklanmaktadır**

# Yaklaşık $r^2$ Değerlerine Örnekler



$$0 < r^2 < 1$$

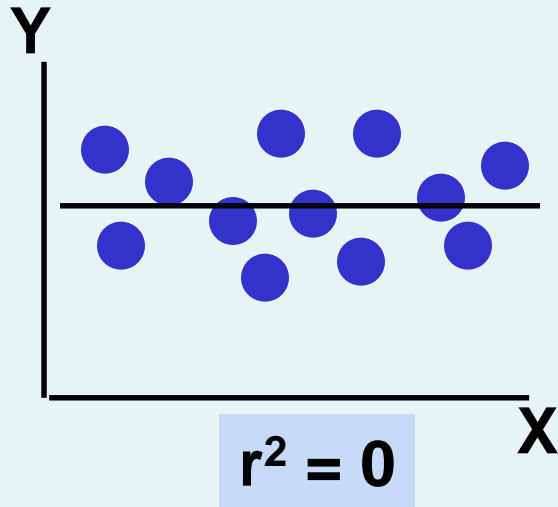
**X ve Y arasında daha zayıf ilişkiler :**



**Y'deki bir kısım değişimler X'deki değişimlerle açıklanır fakat tüm değişimleri içermez**



# Yaklaşık $r^2$ Değerlerine Örnekler



$$r^2 = 0$$

**X ve Y arasında doğrusal bir ilişki yoktur:**

**Y'nin değeri X'e bağlı olarak açıklanamaz.  
(Y'deki değişimlerin hiçbiri X'deki değişimlerle açıklanamaz)**

# Basit Doğrusal Regresyon Örnek: Excel'de Determinasyon Katsayısı, $r^2$

## Regression Statistics

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

$$r^2 = \frac{RKT}{TKT} = \frac{18934.9348}{32600.5000} = 0.58082$$

Ev fiyatlarındaki değişimin %58.08'i ev alanındaki (feet-kare) değişim ile açıklanmaktadır

## ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	18934.9348	18934.9348	11.0848	0.01039
Residual	8	13665.5652	1708.1957		
Total	9	32600.5000			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
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





# Tahminin Standart Hatası

- Regresyon çizgisi etrafındaki gözlemlerin değişimlerinin standart sapması şöyle tahmin edilir:

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

HKT = Hata kareleri toplamı

n = örnek büyüklüğü

# Basit Doğrusal Regresyon Örnek: Excel'de Standart Hata Tahmini

## 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

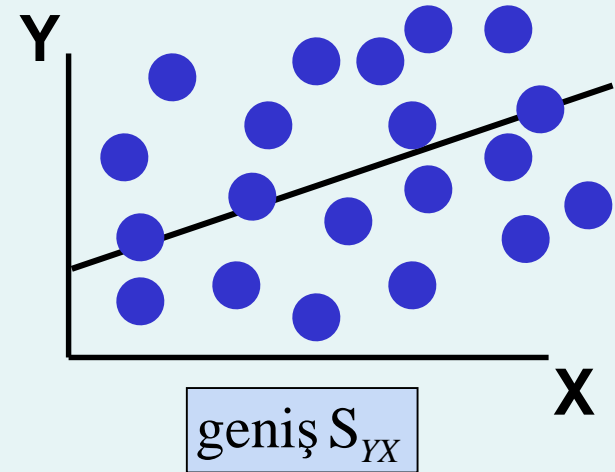
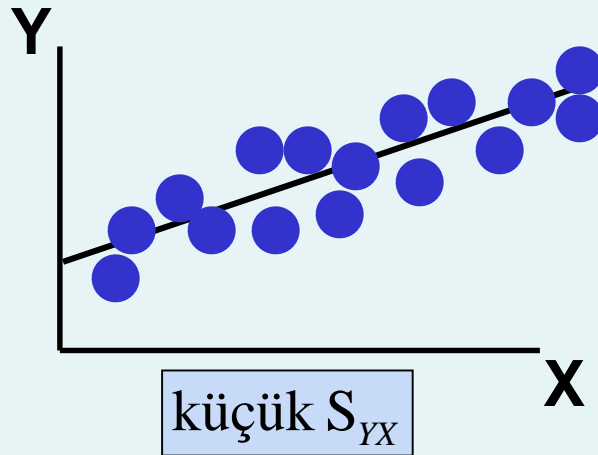
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	18934.9348	18934.9348	11.0848	0.01039
Residual	8	13665.5652	1708.1957		
Total	9	32600.5000			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
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



# Standart Hataların Karşılaştırılması

$S_{YX}$  gözlemlenen Y değerlerinin regresyon çizgisinden sapmasının bir ölçüsüdür



$S_{YX}$ 'in büyüklüğü (şiddeti) örnek verisindeki Y değerlerinin boyutuyla ilişkili olarak değerlendirilmelidir

yani, \$200K - \$400K aralığındaki ev fiyatlarıyla bağlantılı olarak  $S_{YX} = \$41.33K$  kısmen küçük bir değerdir.

# Regresyon Çizgisinin Varsayımları

- Doğrusallık
  - X ve Y arasındaki ilişki doğrusaldır
- Hataların bağımsızlığı
  - Hata değerleri istatistiksel olarak bağımsızdır
- Hatanın normalitesi
  - Hata değerleri verilen bir X değeri için normal dağılmıştır
- Eşit Sapma (homosedastisite olarak da adlandırılır)
  - Hata değerlerinin olasılık dağılımı sabit varyansa sahiptir

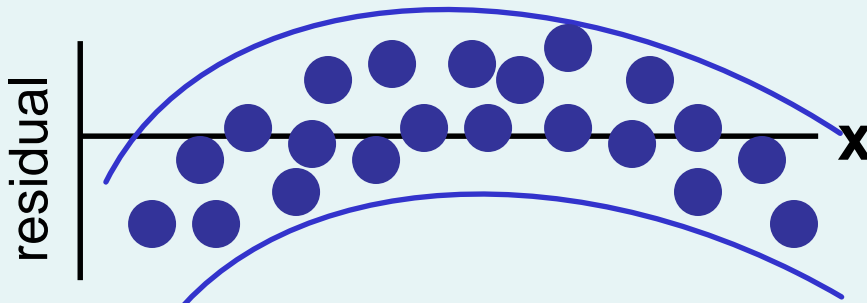
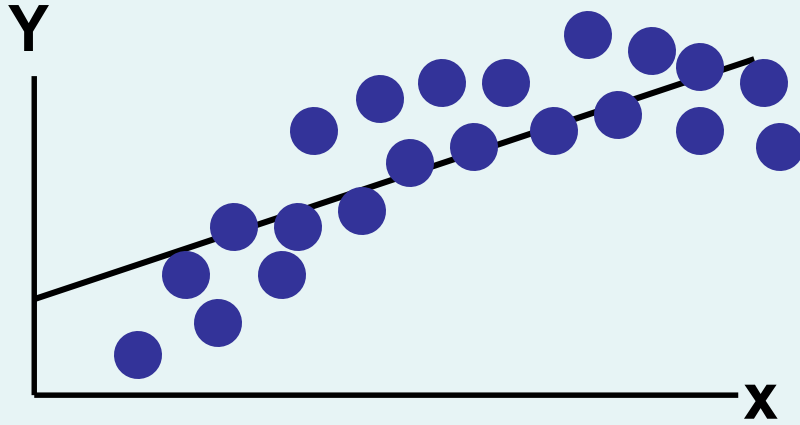


# Artık Değer (Residual) Analizi

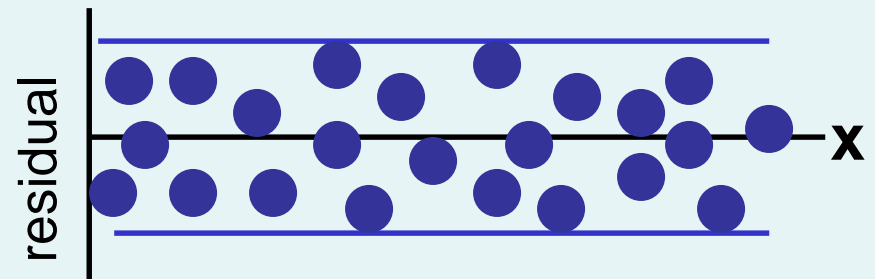
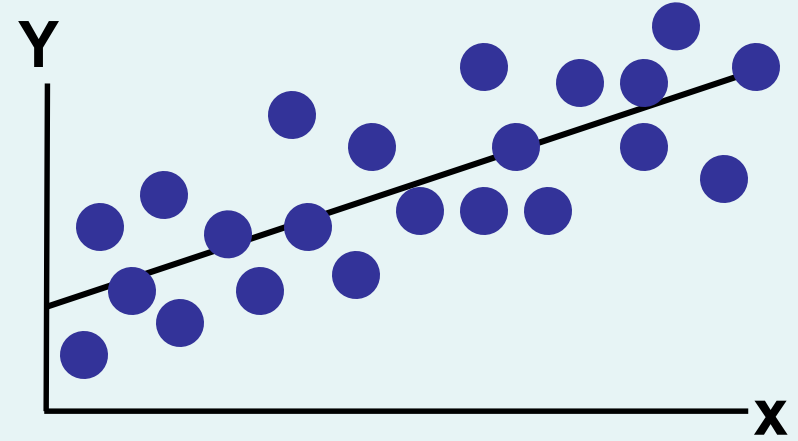
$$e_i = Y_i - \hat{Y}_i$$

- $i$  gözlemi için artık değeri,  $e_i$ , gözlenen değer ile tahmin değeri arasındaki farktır.
- Artık değerlerini test ederek regresyon varsayımlarını kontrol edin
  - Doğrusal varsayımını test edin
  - Bağımsızlık varsayımını değerlendirin
  - Normal dağılım varsayımını değerlendirin
  - $X$ 'in tüm seviyeleri için sabit değişimi (homosedastisite) sorgulayın
- Artık Değerlerin Grafik Analizi
  - $X$ 'e karşılık artık değerlerin çizimi

# Doğrusallık İçin Artık Değer Analizi



**Doğrusal  
değil**



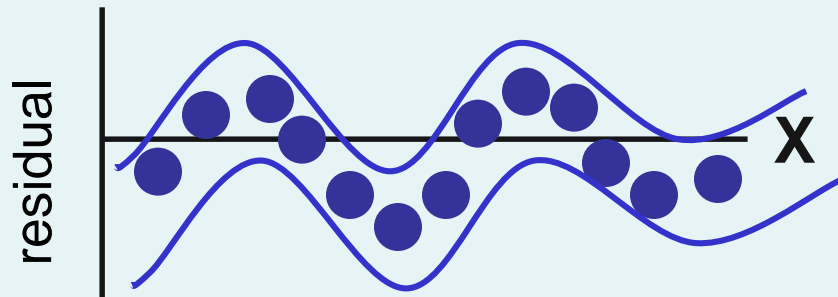
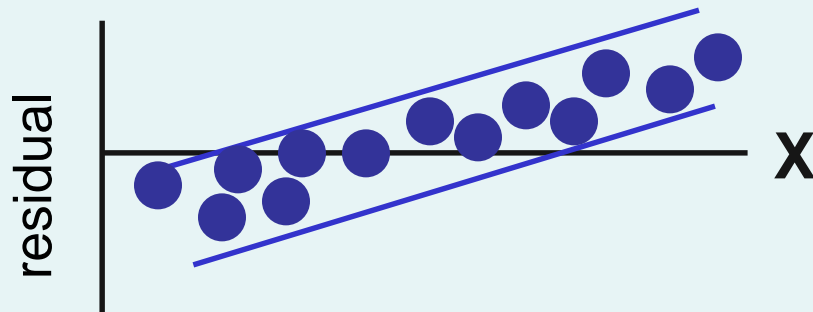
**Doğrusal**



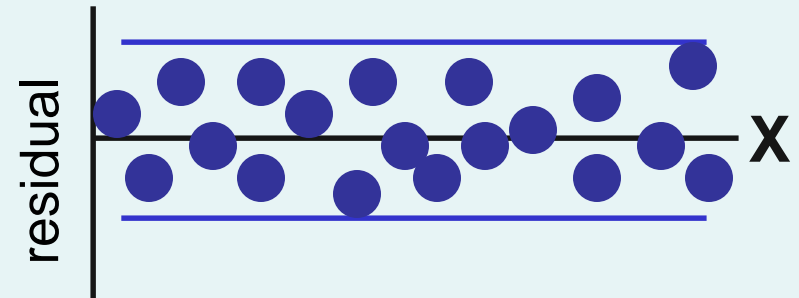
# Bağımsızlık İçin Artık Değer Analizi



Bağımsız değil



Bağımsız





# Normalitenin Kontrolü

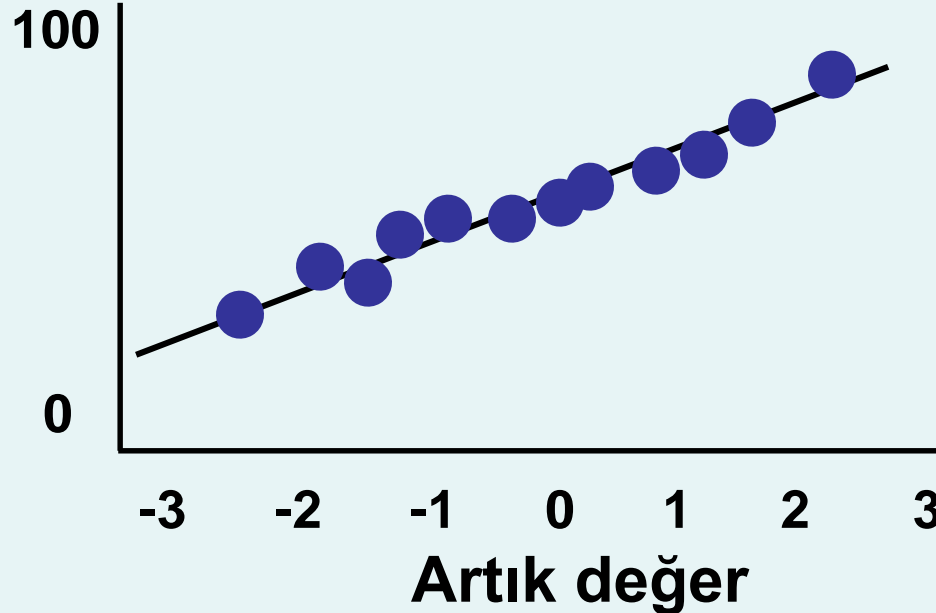
---

- Artık değerler için Kök ve Yaprak gösteriminin incelenmesi
- Artık değerlerin kutu diyagramının incelenmesi
- Artık değerlerin histogramının incelenmesi
- Artık değerlerin bir normal olasılık diyagramının oluşturulması

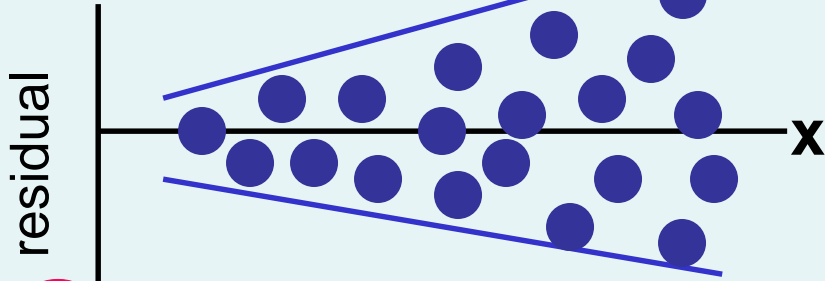
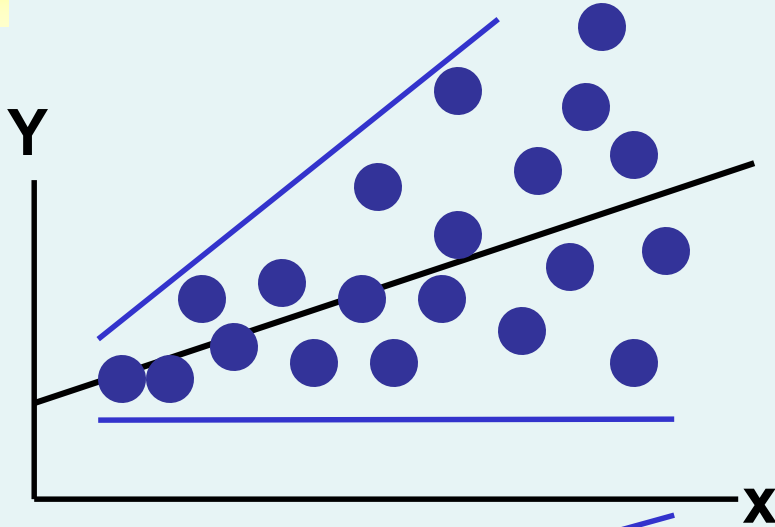
# Normalite için Artık Değer Analizi

Normal olasılık grafiği kullanılırken, normal hatalar yaklaşık olarak doğru bir çizgi üzerinde olacaktır

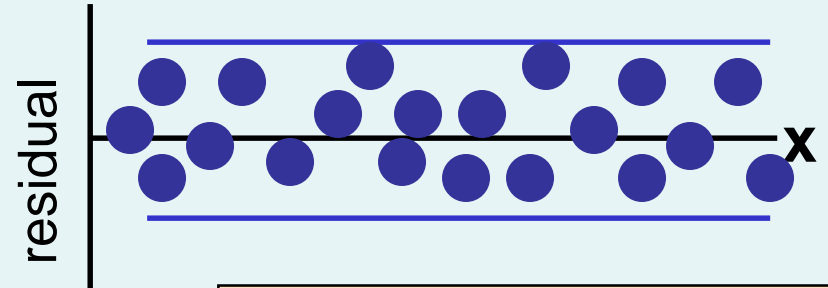
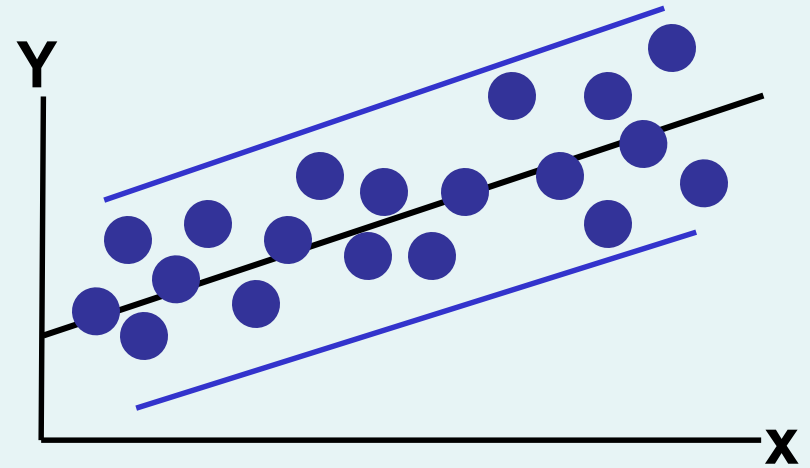
**Yüzde**



# Eşit Değişim İçin Artık Değer Analizi



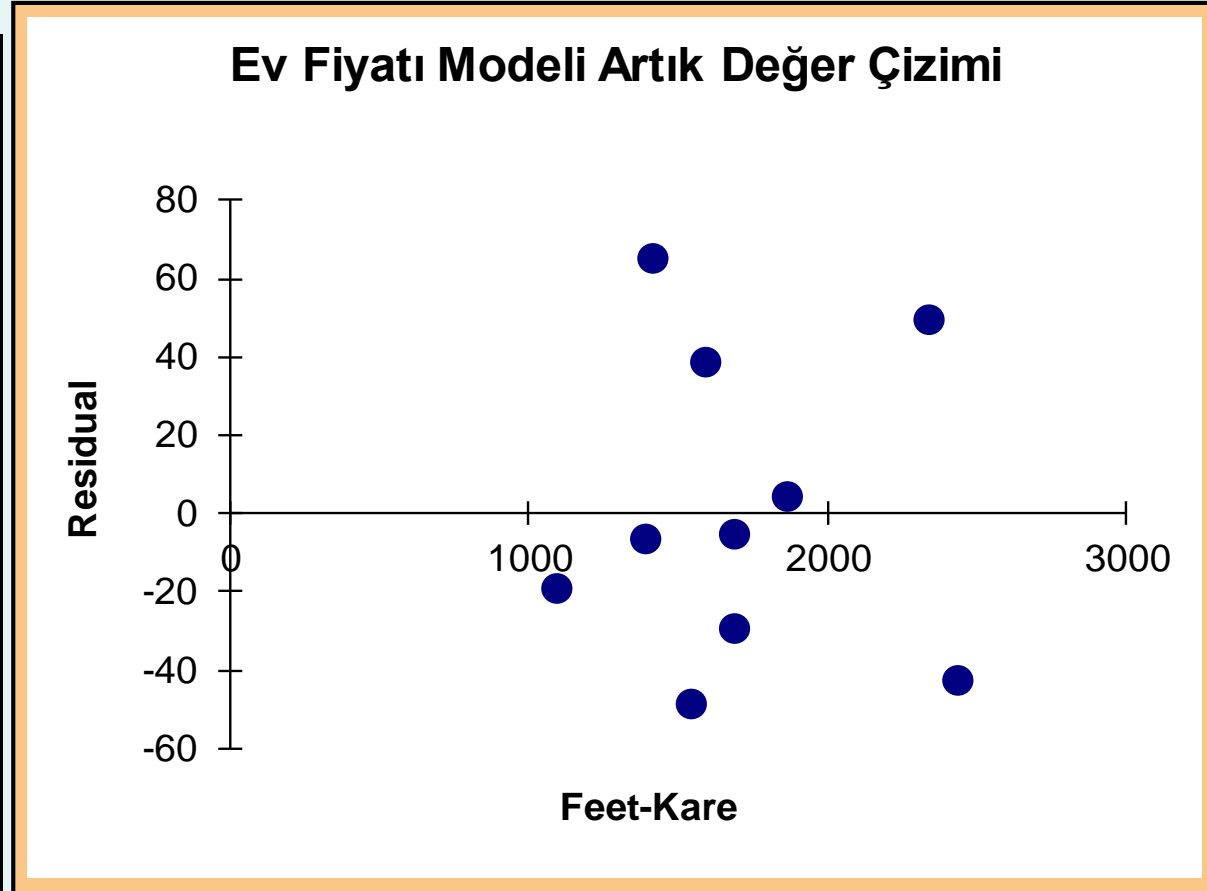
Sabit olmayan değişim



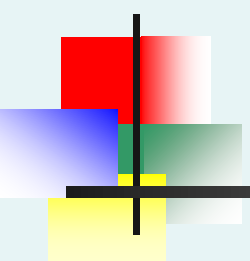
Sabit değişim

# Basit Doğrusal Regresyon Örnek: Excel Artık değer Çıktısı

ARTIK DEĞER ÇIKTISI		
	<i>Tahmini Ev Fiyatı</i>	<i>Artık Değerler</i>
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



Herhangi bir regresyon varsayımını ihlal ettiği görülmemektedir



# Otokorelasyonun Ölçümü: Durbin-Watson İstatistiği

---

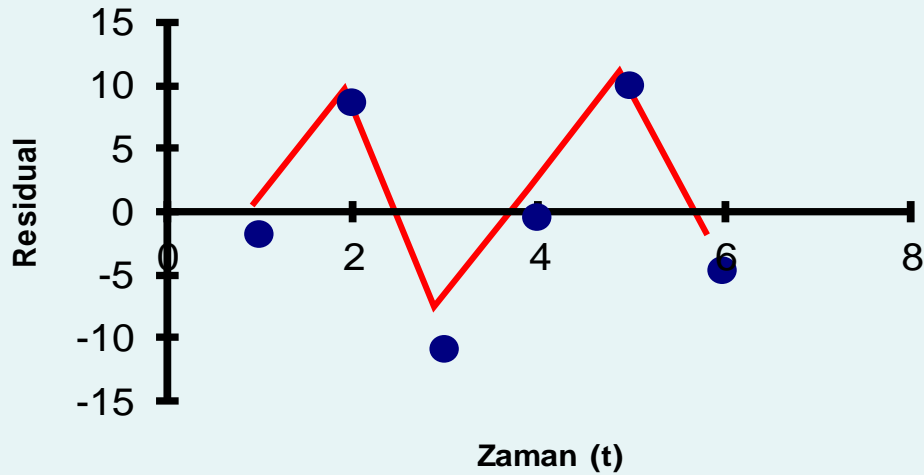
- Veriler zamanla toplandığında otokorelasyon olup olmadığını tespit etmek için kullanılır
- Bir zaman periyodundaki artık değerler başka bir periyottaki artık değerlerle ilgili ise, otokorelasyon mevcuttur

# Otokorelasyon

- Otokorelasyon, zamana göre hataların (artık değerlerin) korelasyonudur

Zaman (t) Artık değer çizimi

- Burada artık değerler, rassal değil, döngüsel bir model gösterirler. Çevrimsel modeller pozitif bir otokorelasyonun işaretidir



- Artık değerlerin rassal ve bağımsız olduğu regresyon varsayımını ihlal eder



# Durbin-Watson İstatistiği

- Durbin-Watson istatistiği, otokorelasyonu test etmek için kullanılır

$H_0$ : artık değerler ilişkili değildir

$H_1$ : pozitif otokorelasyon mevcuttur

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

- Mümkün aralık  $0 \leq D \leq 4$
- $H_0$  doğru ise  $D$  2'ye yakın olmalıdır
- 2'den küçük  $D$  değeri pozitif otokorelasyonun işaretidir, 2'den büyük  $D$  değeri negatif otokorelasyonun işaretidir.



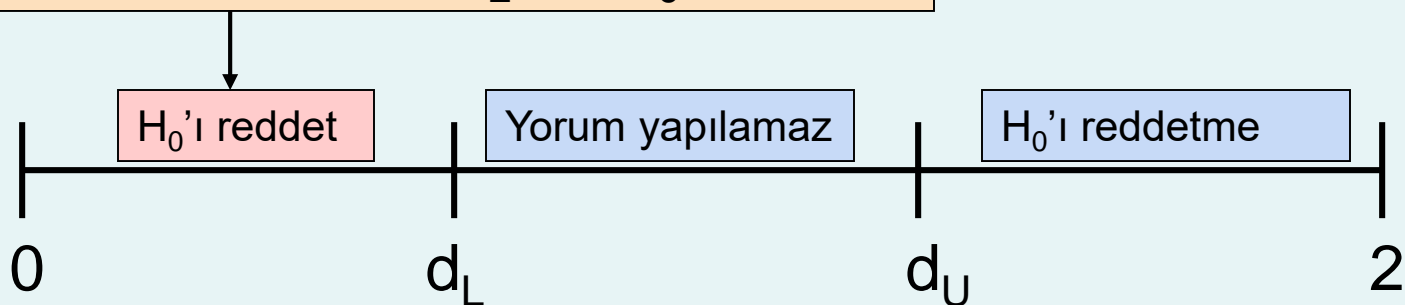
# Pozitif Otokorelasyon İçin Test Etme

$H_0$ : pozitif otokorelasyon yoktur

$H_1$ : pozitif otokorelasyon mevcuttur

- Durbin-Watson test istatistiğini =  $D$  hesapla  
(Durbin-Watson İstatistiği Excel veya Minitab kullanılarak hesaplanabilir)
- Durbin-Watson tablosundan  $d_L$  ve  $d_U$  değerlerini bulun  
( $n$  örnek boyutu için ve  $k$  bağımsız değişken sayısı için)

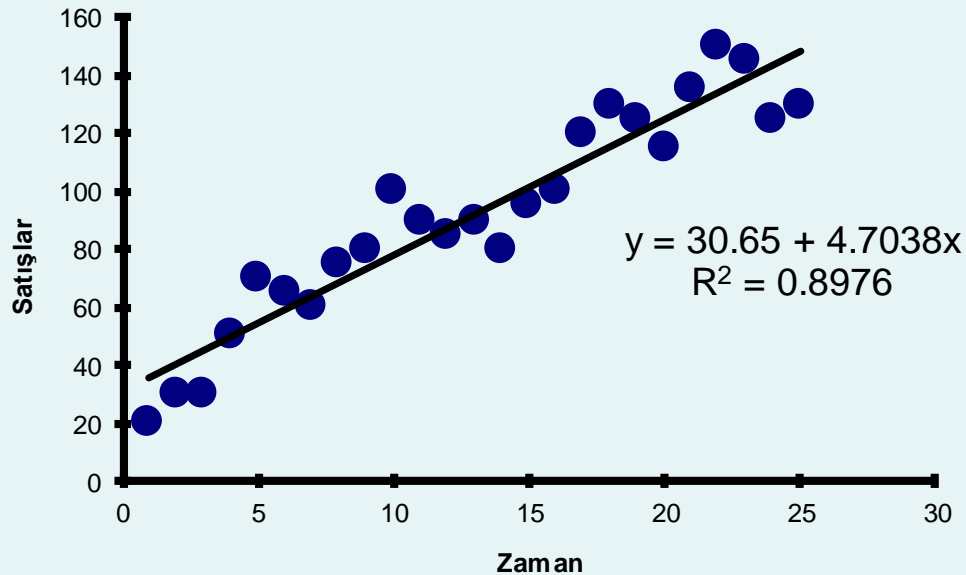
Karar Kuralı: Eğer  $D < d_L$  ise  $H_0$ 'ı reddet



# Pozitif Otokorelasyon İçin Test Etme

(devamı)

- Elimizde şu zaman seri tablosu olduğunu düşünelim:



- Otokorelasyon var mıdır?

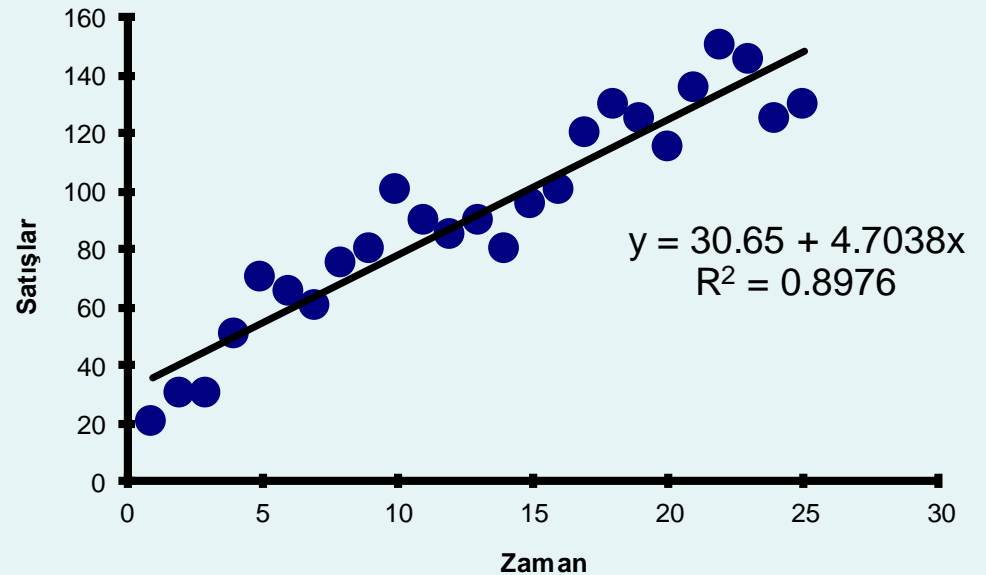
# Pozitif Otokorelasyon İçin Test Etme

(devamı)

- $n = 25$  büyüklüklü örnek:

Excel/PHStat çıktısı:

Durbin-Watson Calculations	
Sum of Squared Difference of Residuals	3296.18
Sum of Squared Residuals	3279.98
<b>Durbin-Watson Statistic</b>	<b>1.00494</b>



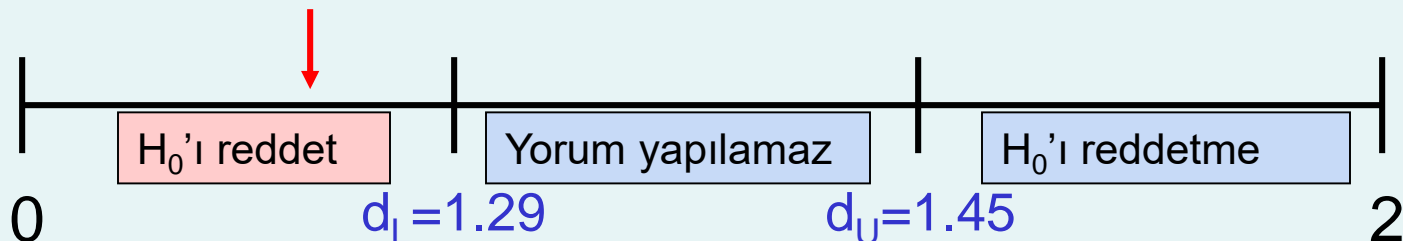
$$D = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} = \frac{3296.18}{3279.98} = 1.00494$$

# Pozitif Otokorelasyon İçin Test Etme

(devamı)

- Burada,  $n = 25$  ve  $k = 1$  olmak üzere bir bağımsız değişken vardır
- Durbin-Watson tablosunu kullanarak,  $d_L = 1.29$  ve  $d_U = 1.45$  elde edilir
- $D = 1.00494 < d_L = 1.29$ ,  $H_0$  reddedilir ve önemli bir pozitif otokorelasyon olduğu yorumu yapılır

**Karar:**  $D = 1.00494 < d_L$   
olduğundan  $H_0$  reddedilir





# Eğimle İlgili Çıkarımlar

- Regresyon eğimi katsayısı ( $b_1$ ) standart hata tahmini şu formülle yapılır:

$$S_{b_1} = \frac{S_{YX}}{\sqrt{XKT}} = \frac{S_{YX}}{\sqrt{\sum (X_i - \bar{X})^2}}$$

$S_{b_1}$  = Eğimin standart hata tahmini

$S_{YX} = \sqrt{\frac{HKT}{n-2}}$  = Tahminin standart hatası

# Eğimle İlgili Çıkarımla: t Testi

- Bir popülasyon eğimi için t testi
  - X ve Y arasında doğrusal bir ilişki var mıdır?
- Sıfır ve alternatif hipotezler
  - $H_0: \beta_1 = 0$  (doğrusal ilişki yok)
  - $H_1: \beta_1 \neq 0$  (doğrusal bir ilişki mevcut)
- Test istatistiği

$$t_{\text{STAT}} = \frac{b_1 - \beta_1}{S_{b_1}}$$

$$s.d. = n - 2$$

$b_1$  = regresyon eğim  
katsayısı

$\beta_1$  = hipotezle belirlenmiş  
eğim

$S_{b_1}$  = eğimin standart hatası

# Eğimle İlgili Çıkarımlar: t Testi Örneği

Ev Fiyatları \$1000 (y)	Feet-Kare (x)
245	1400
312	1600
279	1700
308	1875
199	1100
219	1550
405	2350
324	2450
319	1425
255	1700

## Beklenen Regresyon Denklemi:

$$\text{ev fiyatı} = 98.25 + 0.1098 (\text{feet} - \text{kare})$$

Bu modelin eğimi 0.1098'dir

Evin alanı (feet-kare) ile satış fiyatı arasında doğrusal bir ilişki var mıdır?

# Eğim İle İlgili Çıkarımlar: t Testi Örneği

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

**Excel Çıktısından:**

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	98.24833	58.03348	1.69296	0.12892
Square Feet	0.10977	0.03297	3.32938	0.01039

$b_1$

$S_{b_1}$

$$t_{\text{STAT}} = \frac{b_1 - \beta_1}{S_{b_1}} = \frac{0.10977 - 0}{0.03297} = 3.32938$$



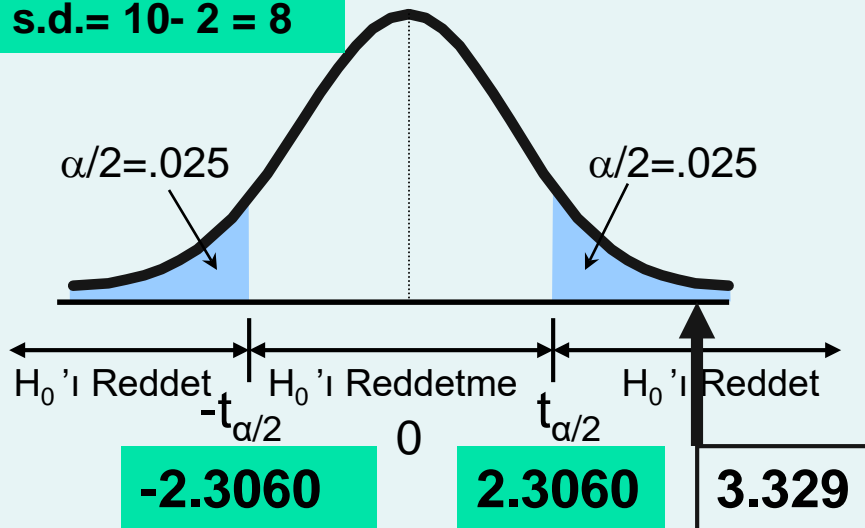
# Eğimle İlgili Çıkarımlar: t Testi Örneği

Test İstatistiği:  $t_{\text{STAT}} = 3.329$

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

s.d. =  $10 - 2 = 8$



Karar:  $H_0$ 'ı Reddet

Ev alanının (feet-kare) ev fiyatlarını etkilediği yönünde yeterli kanıt mevcuttur

# Eğimle İlgili Çıkarımlar: t Testi Örneği

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

**Excel Çıktısından:**

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	98.24833	58.03348	1.69296	0.12892
Square Feet	0.10977	0.03297	3.32938	0.01039

p-değeri

Karar: p-değeri  $< \alpha$  olduğundan  $H_0$ 'ı reddet

Ev alanının (feet-kare) ev fiyatlarını etkilediği yönünde yeterli kanıt mevcuttur.



# Anlamlılık İçin F Testi

- F Test istatistiği:

$$F_{STAT} = \frac{MSR}{MSE}$$

$$MSR = \frac{RKT}{k}$$

$$MSH = \frac{HKT}{n - k - 1}$$

$F_{STAT}$  ,  $k$  pay ve  $(n - k - 1)$  payda serbestlik derecesine sahip F dağılımına uymaktadır

( $k$  = regresyon modelindeki bağımsız değişkenlerin sayısı)

# Anlamlılık İçin F Testi

## Excel Çıktısı

### Regression Statistics

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

$$F_{\text{STAT}} = \frac{\text{MSR}}{\text{MSH}} = \frac{18934.9348}{1708.1957} = 11.0848$$

1 ve 8 serbestlik  
dereceleriyle

F testi için  
p-değeri

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			

# Anlamlılık İçin F Testi

(devamı)

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

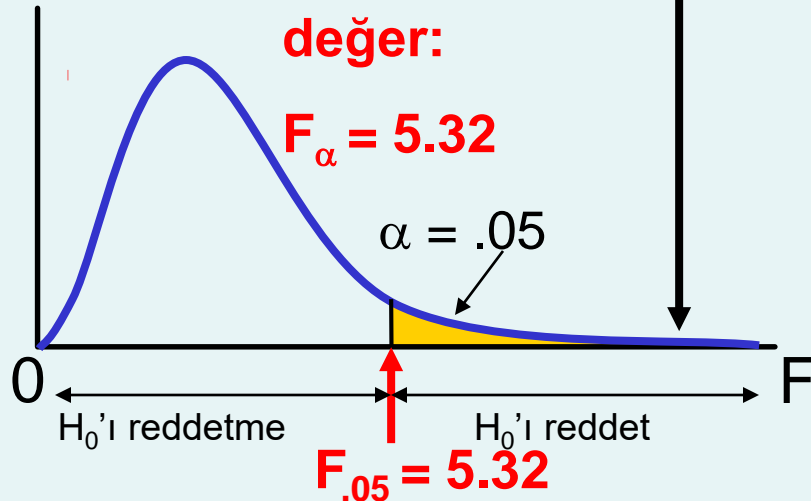
$$\alpha = .05$$

$$df_1 = 1 \quad df_2 = 8$$

**Kritik  
değer:**

$$F_{\alpha} = 5.32$$

$$\alpha = .05$$



**Test İstatistiği:**

$$F_{\text{STAT}} = \frac{MSR}{MSH} = 11.08$$

**Karar:**

$\alpha = 0.05$  düzeyinde  $H_0$  reddedilir

**Çıkarım:**

Ev boyutlarının satış fiyatlarını etkilediği yönünde yeterli kanıt mevcuttur.

# Eğim İçin Güven Aralığı Tahmini

Eğimin Güven Aralığı Tahmini:

$$b_1 \pm t_{\alpha/2} S_{b_1}$$

$$\text{s.d.} = n - 2$$

Ev fiyatları İçin Excel Çıktısı:

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
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 (0.0337, 0.1858) olacaktır

# Eğim İçin Güven Aralığı Tahmini

(devamı)

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
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

Ev fiyatı değişkeninin birimi \$1000 olduğundan, satış fiyatları üzerindeki ortalama etkinin ev boyutunun her bir foot-karesi için \$33.74 ve \$185.80 arasında olduğunu %95 güven düzeyinde söyleyebiliriz

Bu %95 güven aralığı 0 içermeyecektir.

**Çıkarım:** .05 önem seviyesinde ev fiyatları ile ev boyutu arasında önemli bir ilişki mevcuttur.



# Korelasyon Katsayısı için t Testi

- Hipotezler

$H_0: \rho = 0$  (X ve Y arasında korelasyon yoktur)

$H_1: \rho \neq 0$  (korelasyon mevcuttur)

- Test istatistiği

$$t_{\text{STAT}} = \frac{r - \rho}{\sqrt{\frac{1 - r^2}{n - 2}}}$$

(n – 2 serbestlik derecesiyle)

$$r = +\sqrt{r^2} \text{ if } b_1 > 0$$

$$r = -\sqrt{r^2} \text{ if } b_1 < 0$$



# Korelasyon Katsayısı için t Testi

(devamı)

.05 önem seviyesinde ev boyutu (feet-kare) ile satış fiyatı arasında bir doğrusal ilişki olduğuna dair bir kanıt var mıdır?

$H_0: \rho = 0$  (korelasyon yok)

$H_1: \rho \neq 0$  (korelasyon mevcut)

$\alpha = .05$  ,  $sd = 10 - 2 = 8$

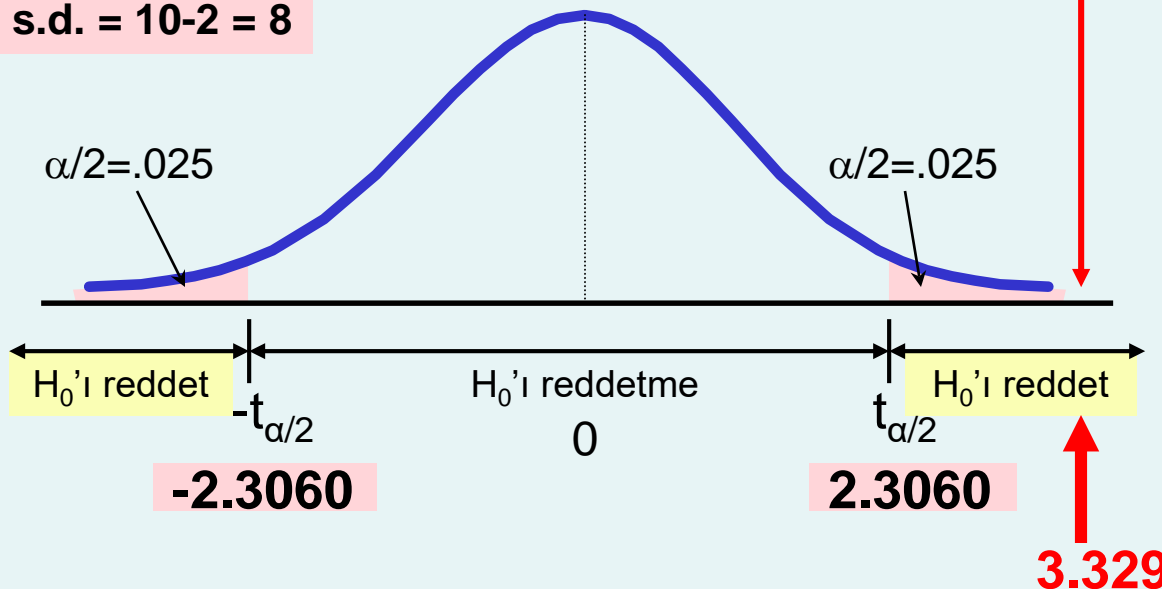
$$t_{STAT} = \frac{r - \rho}{\sqrt{\frac{1 - r^2}{n - 2}}} = \frac{.762 - 0}{\sqrt{\frac{1 - .762^2}{10 - 2}}} = 3.329$$

# Korelasyon Katsayısı için t Testi

(devamı)

$$t_{\text{STAT}} = \frac{r - \rho}{\sqrt{\frac{1 - r^2}{n - 2}}} = \frac{.762 - 0}{\sqrt{\frac{1 - .762^2}{10 - 2}}} = 3.329$$

$$\text{s.d.} = 10 - 2 = 8$$



**Karar:**  
H<sub>0</sub>'ı reddet

**Conclusion:**  
%5 önem  
seviyesinde bir  
doğrusal  
etkileşim  
olduğuna dair  
**kanıt vardır**

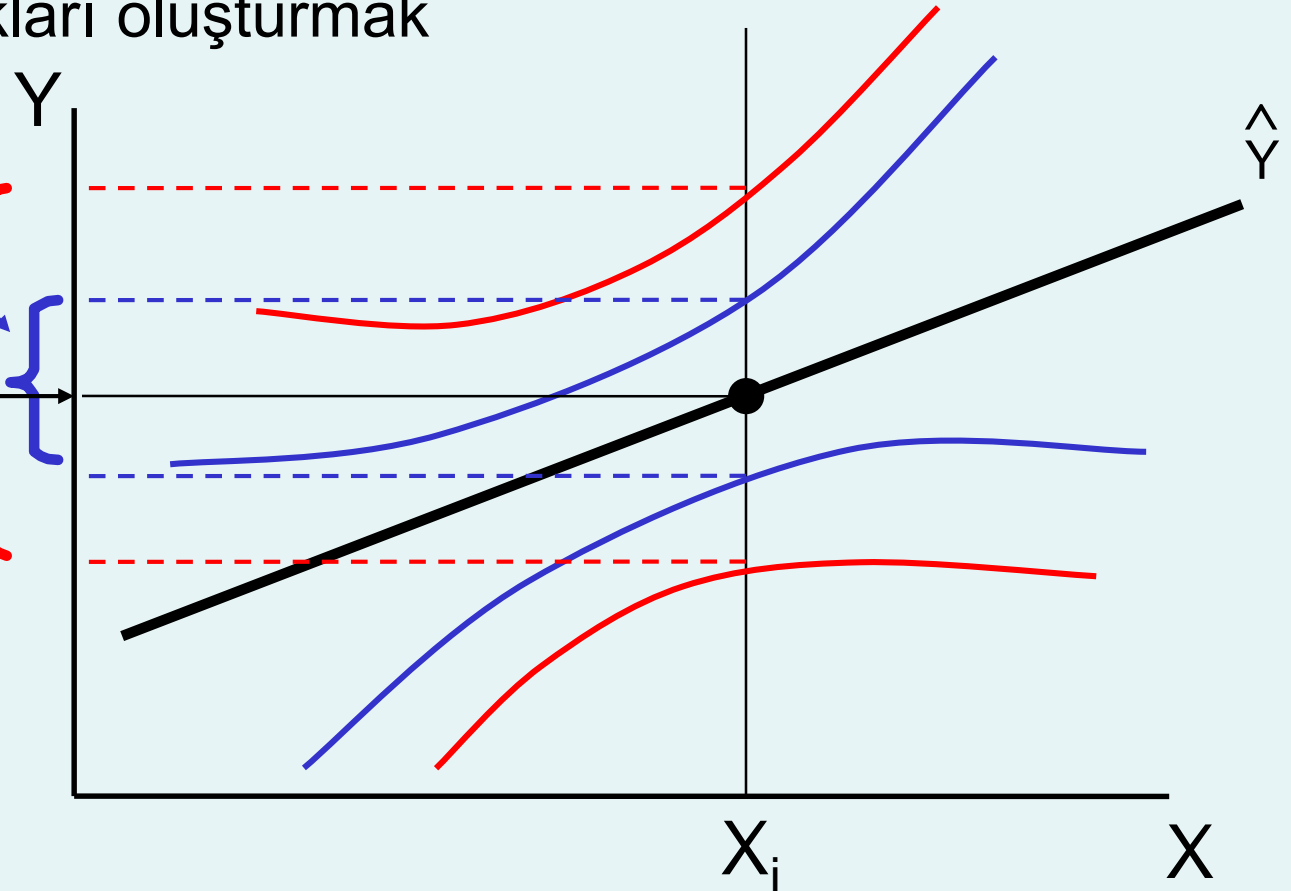
# Ortalama Değerlerinin Tahmini ve Bağımsız Değerlerin Çıkarımı

Amaç: Verilen bir  $X_i$  için  $Y$ 'nin değeriyle ilgili belirsizliğin aktarılması için  $Y$  etrafındaki aralıkları oluşturmak

$X_i$  verildiğinde  $Y$ 'nin ortalaması için Güven aralığı

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

$X_i$  verildiğinde, bağımsız bir  $Y$  değerinin tahmin aralığı



# X Verildiğinde Ortalama Y Değeri İçin Güven Aralığı

Belirli bir  $X_i$  verildiğinde **Y'nin ortalama değeri** için güven aralığı tahmini

$\mu_{Y|X=X_i}$  için Güven aralığı :

$$\hat{Y} \pm t_{\alpha/2} S_{YX} \sqrt{h_i}$$

Aralığın boyutu Ortalamadan  $\bar{X}$ , olan uzaklığa göre değişir

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

# X Verildiğinde Bağımsız Y Değeri İçin Güven Aralığı

Belirli bir  $X_i$  verildiğinde **Y'nin bağımsız bir değeri** için güven aralığı tahmini

$Y_{X=X_i}$  için guven araligi :

$$\hat{Y} \pm t_{\alpha/2} S_{YX} \sqrt{1 + h_i}$$

Bu fazladan terim, aralık genişliğine, tek bir bağımsız değer için eklenmiş belirsizliği yansıtacak şekilde eklenir



# Ortalama Değerlerin Tahmini: Örnek

$\mu_{Y|X=X_i}$  için Güven aralığı

2,000 feet-kare evlerin ortalama fiyatlarının %95 güven aralığında tahmini değerini bulunuz

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

$$\hat{Y} \pm t_{0.025} S_{YX} \sqrt{\frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum (X_i - \bar{X})^2}} = 317.85 \pm 37.12$$

Güven aralığı uç noktaları( Excel'den) 280.66 ve 354.90'dır, veya \$280,660'dan \$354,900'a değer alır

# Bağımsız tek Değerlerin Tahmini: Örnek

$Y_{X=X_i}$  için kestirim aralığı tahmini

2,000 feet-kare olan bir bağımsız evin %95 güven düzeyinde kestirim aralığını bulunuz

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

$$\hat{Y} \pm t_{0.025} S_{YX} \sqrt{1 + \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum (X_i - \bar{X})^2}} = 317.85 \pm 102.28$$

Excel'den elde edilen kestirim aralığı uç noktaları 215.50 ve 420.07, veya \$215,500'dan \$420,070'a olacaktır



# Excel'de Güven ve Kestirim Aralıklarının Bulunması

---

- Excel'de

PHStat | regression | simple linear regression ...  
sıralamasını kullan

- “X için güven ve kestirim aralığı =”  
“confidence and prediction interval for X=”  
kutucuğunu işaretle  
ve X-değerini ve istenen güven seviyesini gir



# Excel'de Güven ve Kestirim Aralıklarının Bulunması

(devamı)

	A	B
1	Confidence Interval Estimate	
2		
3	Data	
4	X Value	2000
5	Confidence Level	95%
6		
7	Intermediate Calculations	
8	Sample Size	10
9	Degrees of Freedom	8
10	t Value	2.306006
11	Sample Mean	1715
12	Sum of Squared Difference	1571500
13	Standard Error of the Estimate	41.33032
14	h Statistic	0.151686
15	Average Predicted Y (YHat)	317.7838
16		
17	For Average Predicted Y (YHat)	
18	Interval Half Width	37.11952
19	Confidence Interval Lower Limit	280.6643
20	Confidence Interval Upper Limit	354.9033
21		
22	For Individual Response Y	
23	Interval Half Width	102.2813
24	Prediction Interval Lower Limit	215.5025
25	Prediction Interval Upper Limit	420.0651

Girdi değerleri

$\hat{Y}$

$\mu_{Y|X=X_i}$  için güven aralığı tahmini

$Y_{X=X_i}$  için kestirim aralığı tahmini

# Regresyon Analizinde Düşülebilecek Hatalar

- En küçük kareler regresyonunun altında yatan varsayımların farkında olmamak
- Varsayımların nasıl değerlendireceğini bilmemek
- Belli bir varsayım ihlal edilirse, en küçük kareler regresyonunun alternatiflerini bilmemek
- Konu bilgisi olmaksızın bir regresyon modeli kullanılması
- İlgili alanın dışında ekstrapolasyon yapmak



# Regresyonda Karşılaşılabilecek Hataların Önlenmesi İçin Stratejiler

- Olası ilişkinin gözlemlenmesi için X ve Y'nin bir serpmeye diyagramı ile başlanır
- Varsayımları kontrol etmek için artık değer analizi yapılmalıdır
  - Homoskedastisite gibi varsayımların ihlali için X ve Artık değerlerin diyagramı çizilir
  - Olası normal olmayanlığı ortaya çıkarmak için artık değerlerin histogramı, kök ve yaprak diyagramı, kutu diyagramı veya artık değerlerin normal olasılık grafiği kullanılır



# Regresyonda Karşılaşılabilecek Hataların Önlenmesi İçin Stratejiler

(devamı)

- Herhangi bir varsayımın ihlali mevcutsa, alternatif metot veya modeller kullanılır
- Varsayım ihlali ile ilgili kanıt yoksa, regresyon katsayılarının önem derecesi için test edilir ve güven aralığı ve kestirim aralıkları oluşturulur
- Kestirim yapılması önlenir veya ilgili aralık dışında tahmin yapılır



# Bölüm Özeti

---

Bu bölümde aşağıdaki konulara değinilmiştir;

- Regresyon modellerinin tipleri
- Regresyon ve korelasyonun varsayımları
- Basit doğrusal regresyonun elde edilmesi
- Değişim ölçüleri
- Artık değer (Residual) analizi
- Otokorelasyonun ölçümü



# Bölüm Özeti

(devamı)

- Eğim ile ilgili çıkarımlar yapmak
- Korelasyon – birlikteliğin gücünü ölçmek
- Ortalama değerlerin tahmini ve bağımsız değerlerin kestirimi
- Regresyondaki olası tuzaklar ve bunların önlenmesi için önerilen stratejiler