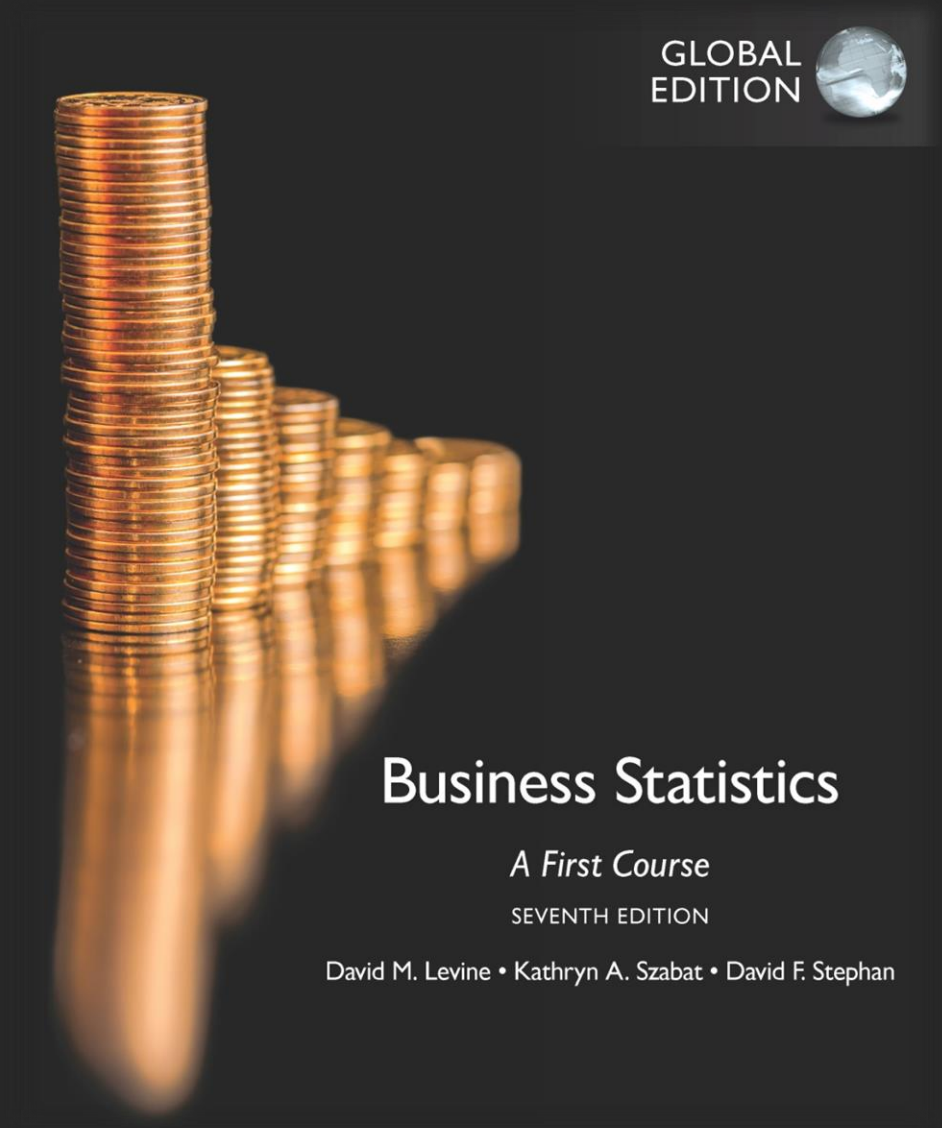


Chapter 12

Simple Linear Regression

Objectives

In this chapter, you learn:

- How to use regression analysis to predict the value of a dependent variable based on a value of an independent variable
- To understand the meaning of the regression coefficients b_0 and b_1
- To evaluate the assumptions of regression analysis and know what to do if the assumptions are violated
- To make inferences about the slope and correlation coefficient
- To estimate mean values and predict individual values

מה יכול לגרום להן להיות מופרות?

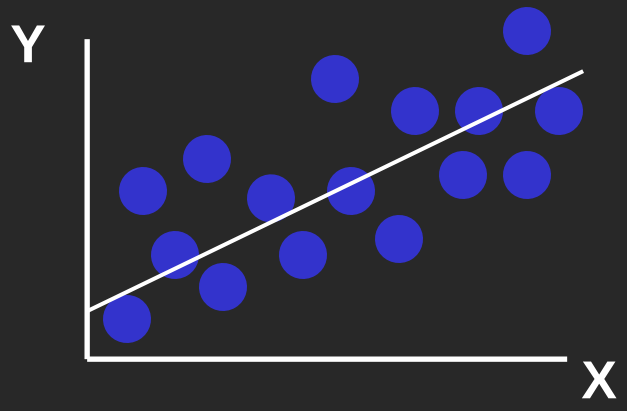
Correlation vs. Regression

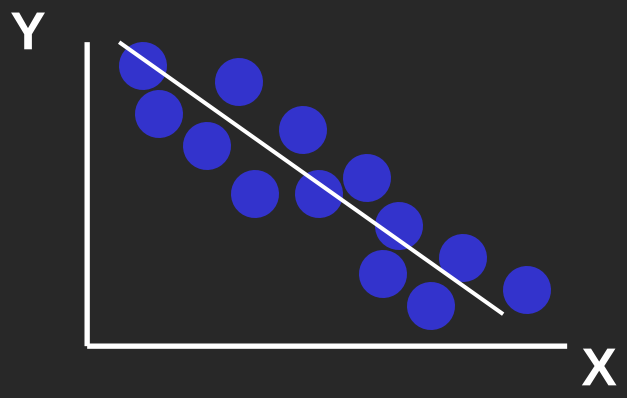
איך נגדיר קורלציה?

- A scatter plot can be used to show the relationship between two variables
- Correlation analysis is used to measure the strength of the association (linear relationship) between two variables
  - Correlation is only concerned with strength of the relationship
  - No causal effect is implied with correlation
  - Scatter plots were first presented in Ch. 2
  - Correlation was first presented in Ch. 3

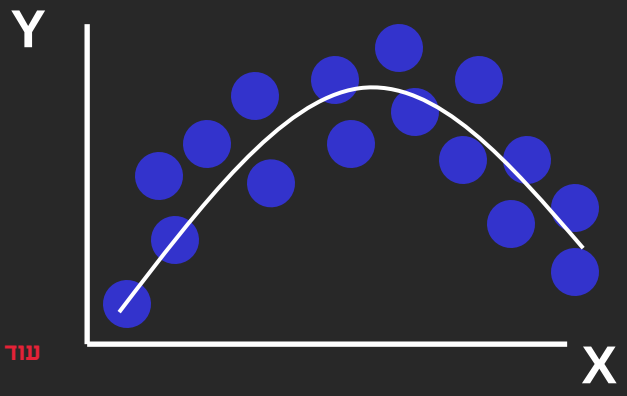
12.1 Regression Models

Linear relationships

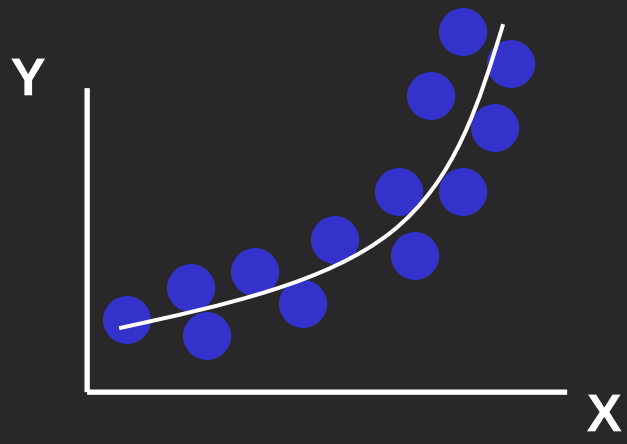




Curvilinear relationships



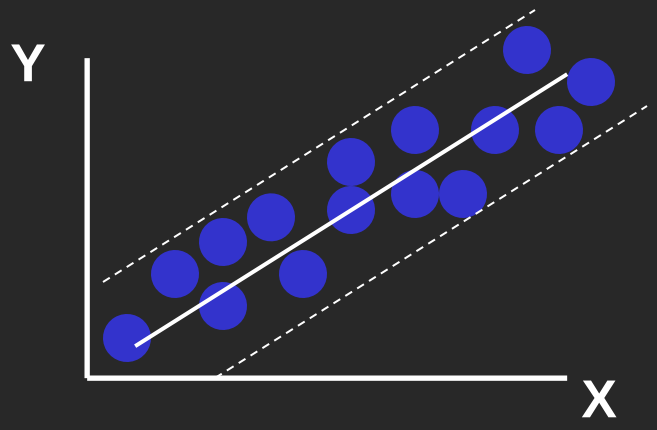
עוד דוגמאות?

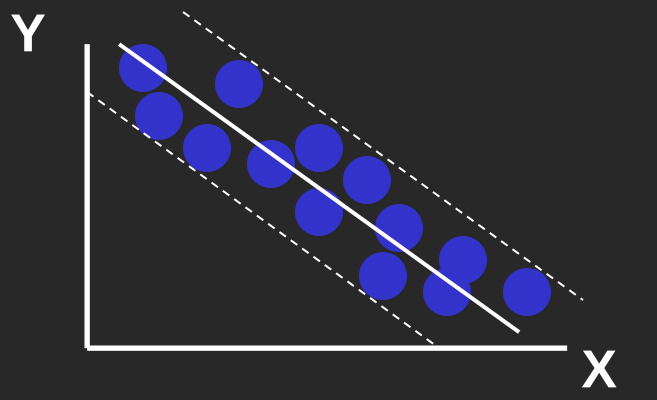


Types of Relationships

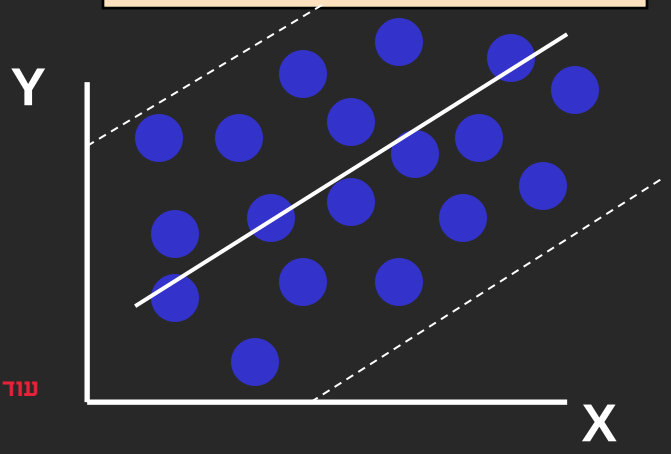
(continued)

Strong relationships

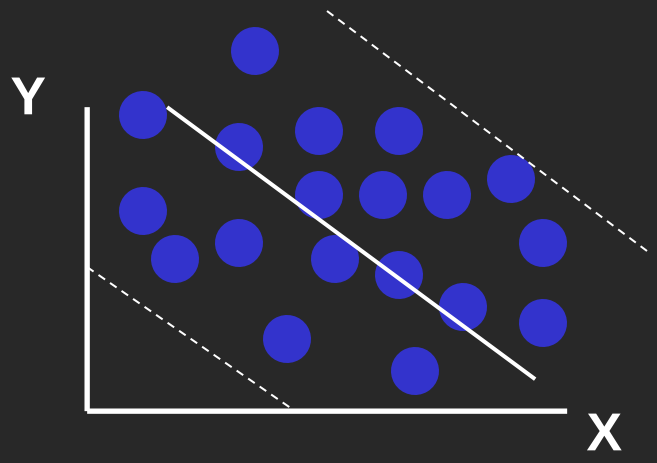




Weak relationships

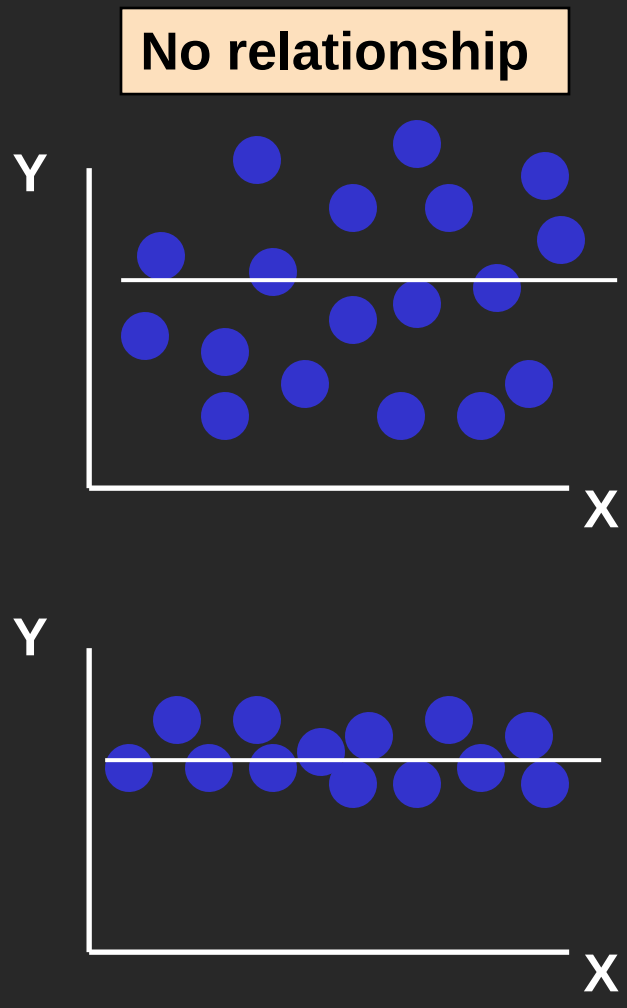


עוד דוגמאות?



Types of Relationships

(continued)



איך ייראה הגרף במקרה של קידוד 1-הוט?

12.2 Introduction to Regression Analysis

- Regression analysis is used to:
  - Predict the value of a dependent variable based on the value of at least one independent variable
  - Explain the impact of changes in an independent variable on the dependent variable

איך נגדיר שינוי?

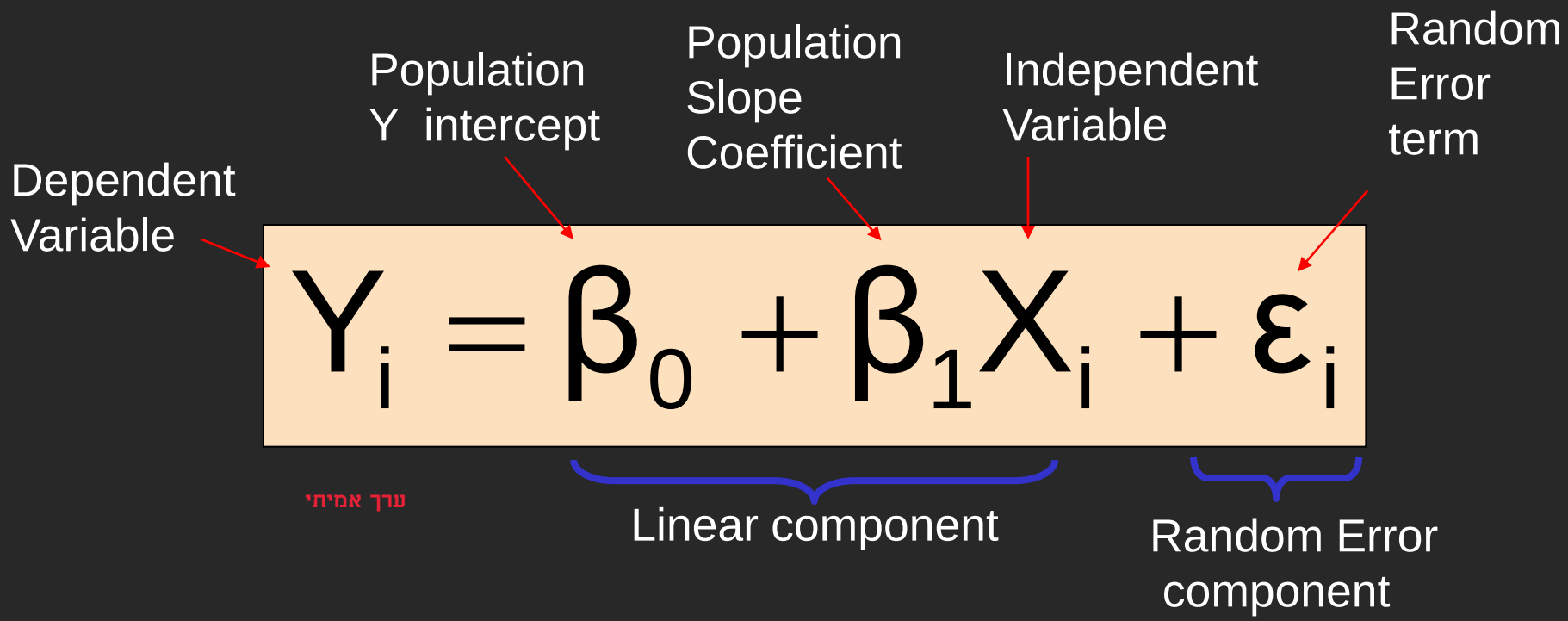
- Dependent variable: the variable we wish to predict or explain
- Independent variable: the variable used to predict or explain the dependent variable

Simple Linear Regression Model

- Only **one** independent variable, X
- Relationship between X and Y is described by a linear function
- Changes in Y are assumed to be related to changes in X

מה יקרה כאשר תהיה לנו סדרה של משתנים מסבירים?

Simple Linear Regression Model

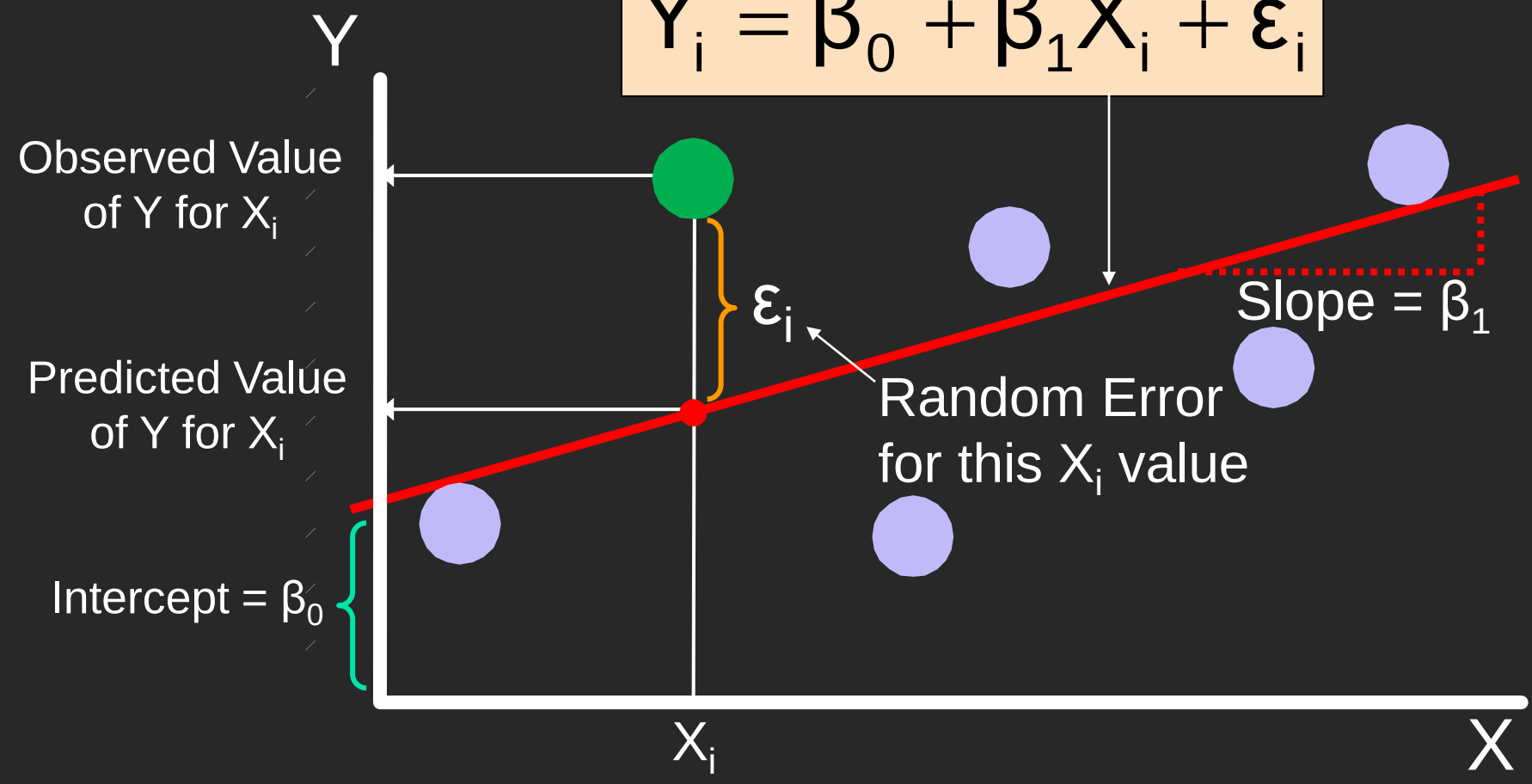


Simple Linear Regression Model

(continued)

ערך אמיתי

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$



Simple Linear Regression Equation (Prediction Line)

The simple linear regression equation provides an estimate of the population regression line

Estimated (or predicted) Y value for observation i

Estimate of the regression intercept

Estimate of the regression slope

Value of X for observation i

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

ערך חזוי

The Least Squares Method

b_0 and b_1 are obtained by finding the values that minimize the sum of the squared differences between Y and $\hat{Y}$:

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

אמיתי - חזוי

MSE = 1/n(sigma(Yi-yi)^2)

Finding the Least Squares Equation

- The coefficients b_0 and b_1, and other regression results in this chapter, will be found using Excel or Minitab

או בכל דרך אחרת לחיישוב נומרי

Interpretation of the Slope and the Intercept

נקודת חיתוך עם ציר אנכי

- b_0 is the estimated mean value of Y when the value of X is zero

מקדם הרגרסיה

- b_1 is the estimated change in the mean value of Y as a result of a one-unit increase in X

Simple Linear Regression Example

- A real estate agent wishes to examine the relationship between the selling price of a home and its size (measured in square feet)

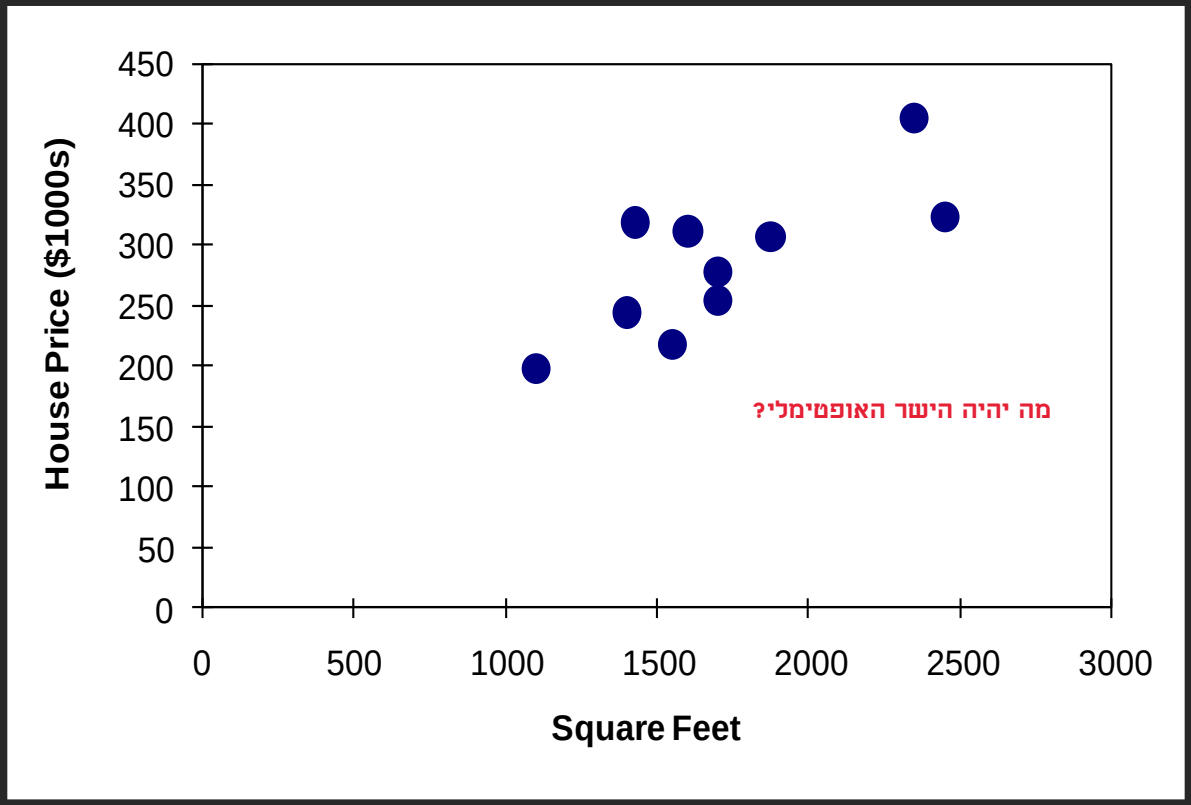
- A random sample of 10 houses is selected
  - Dependent variable (Y) = house price in $1000s
  - Independent variable (X) = square feet

Simple Linear Regression Example: Data

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

Simple Linear Regression Example: Scatter Plot

House price model: Scatter Plot



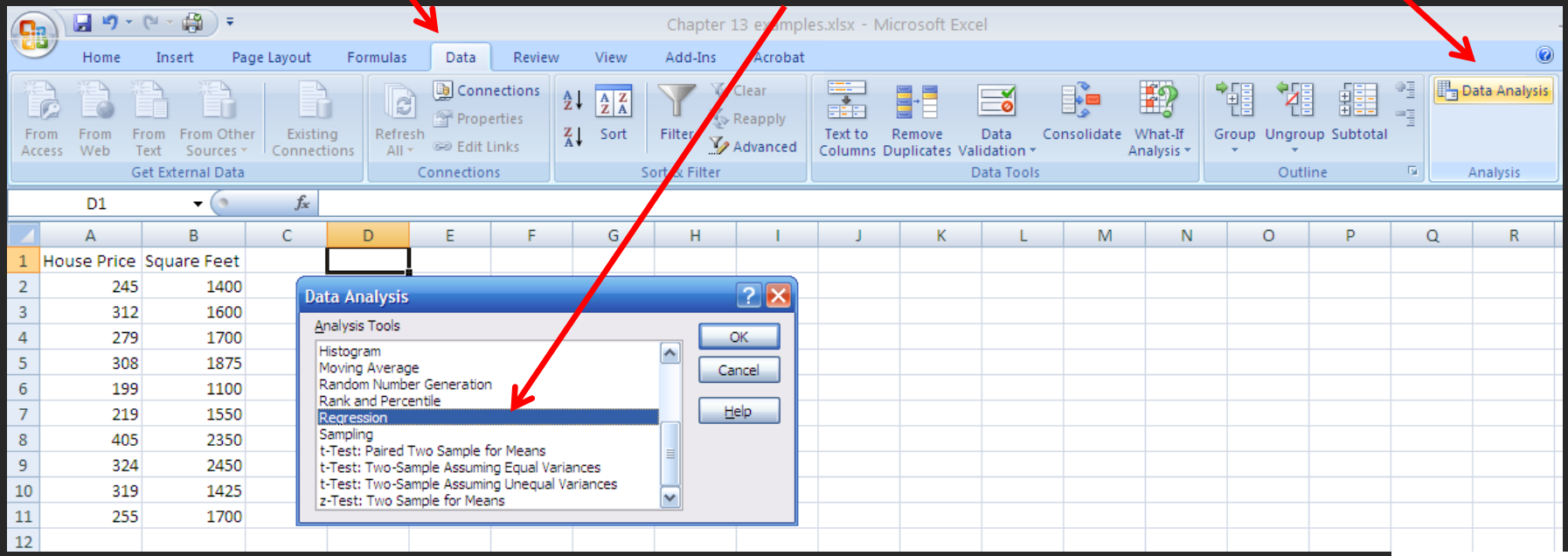
Simple Linear Regression Example: Using Excel Data Analysis Function

1. Choose Data

2. Choose Data Analysis

In Python: SKLEARN

3. Choose Regression

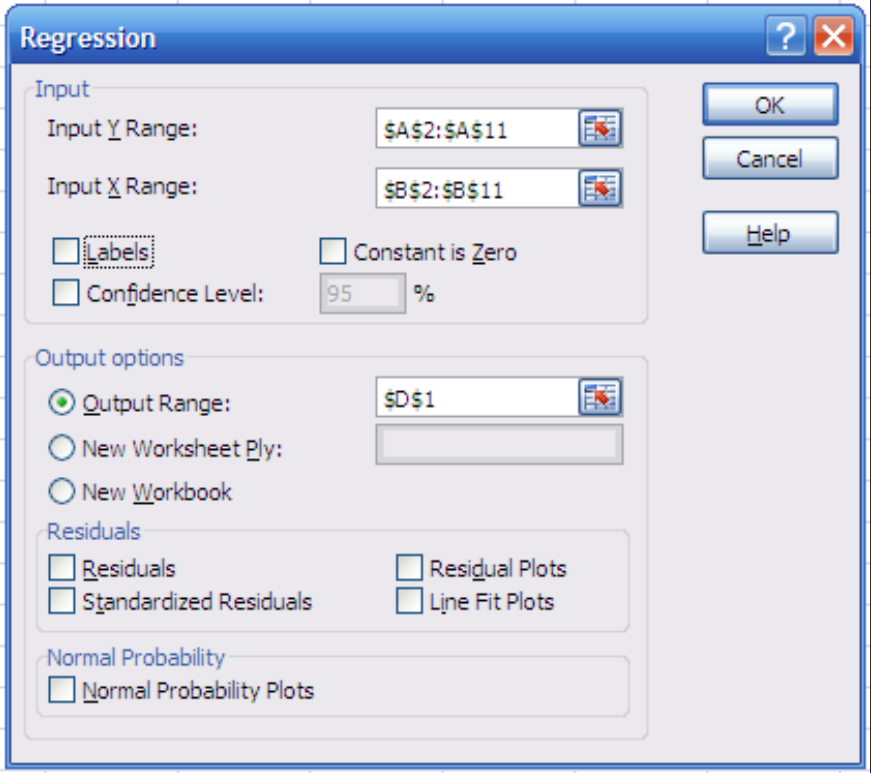


Simple Linear Regression Example: Using Excel Data Analysis Function

(continued)

Enter Y range and X range and desired options

	A	B
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



Simple Linear Regression Example: Excel Output

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

The regression equation is:

$$\widehat{\text{house price}} = 98.24833 + 0.10977\,(\text{square feet})$$

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

Simple Linear Regression Example: Minitab Output

The regression equation is

Price = 98.2 + 0.110 Square Feet

Predictor	Coef	SE Coef	T	P
Constant	98.25	58.03	1.69	0.129
Square Feet	0.10977	0.03297	3.33	0.010

S = 41.3303 R-Sq = 58.1% R-Sq(adj) = 52.8%

Analysis of Variance

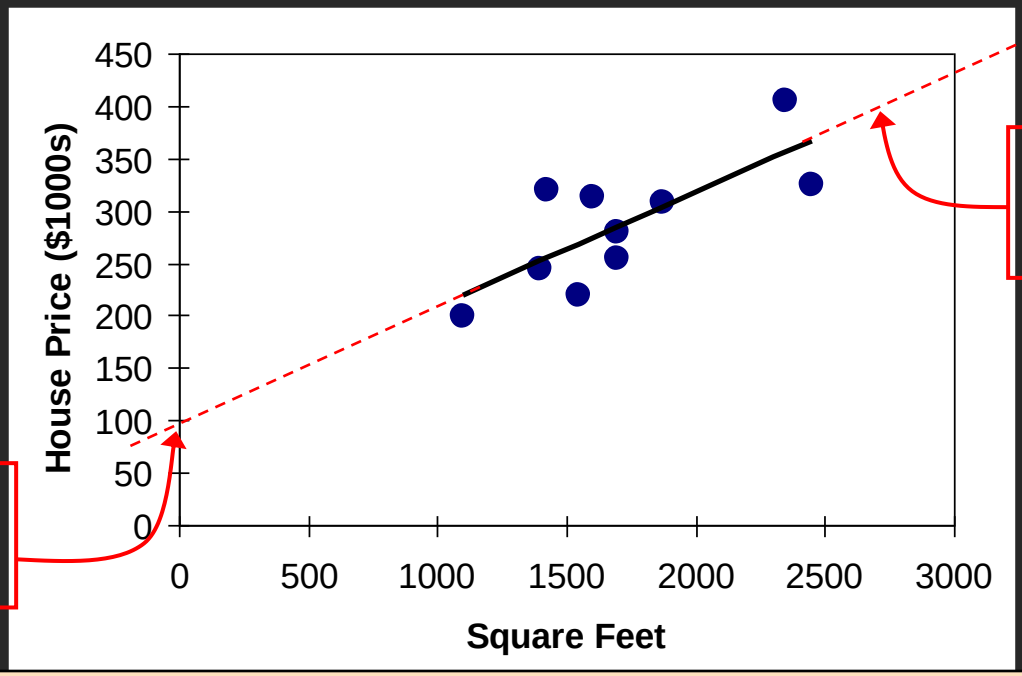
Source	DF	SS	MS	F	P
Regression	1	18935	18935	11.08	0.010
Residual Error	8	13666	1708		
Total	9	32600			

The regression equation is:

$\widehat{\text{house price}} = 98.24833 + 0.10977 \text{ (square feet)}$

Simple Linear Regression Example: Graphical Representation

House price model: Scatter Plot and Prediction Line



Slope
= 0.10977

Intercept
= 98.248

$$\widehat{\text{house price}} = 98.24833 + 0.10977\,(\text{square feet})$$

Simple Linear Regression Example: Interpretation of b_o

$$\widehat{\text{house price}} = 98.24833 + 0.10977\,(\text{square feet})$$

- b_0 is the estimated mean value of Y when the value of X is zero (if X = 0 is in the range of observed X values) האם ערך כזה בכלל אפשרי?
- Because a house cannot have a square footage of 0, b_0 has no practical application

Simple Linear Regression Example: Interpreting b_1

ממוצע השינוי לאורך הגדילה

$$\widehat{\text{house price}} = 98.24833 + 0.10977(\text{square feet})$$

- b_1 estimates the change in the mean value of Y as a result of a one-unit increase in X
  - Here, $b_1 = 0.10977$ tells us that the mean value of a house increases by .10977($1000) = $109.77, on average, for each additional one square foot of size

Simple Linear Regression Example: Making Predictions

Predict the price for a house with 2000 square feet:

פשוט נציב, לא?

$$\widehat{\text{house price}} = 98.25 + 0.1098\ (\text{sq.ft.})$$

$$= 98.25 + 0.1098(2000)$$

$$= 317.85$$

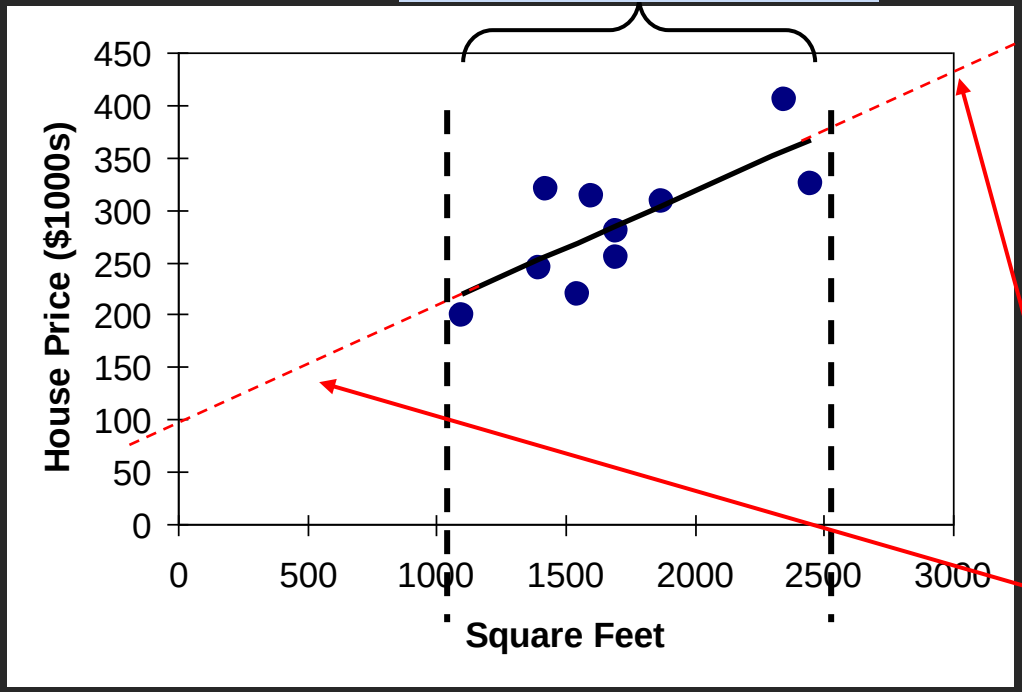
The predicted price for a house with 2000 square feet is 317.85($1,000s) = $317,850

Simple Linear Regression Example: Making Predictions

- When using a regression model for prediction, only predict within the relevant range of data

האם כדאי להשתמש ברגרסיה בטווח לא מוגדר?

Relevant range for interpolation



Do not try to extrapolate beyond the range of observed X's

12.4 Assumptions of Regression L.I.N.E

- Linearity
  - The relationship between X and Y is linear
- Independence of Errors
  - Error values are statistically independent
  - Particularly important when data are collected over a period of time
- Normality of Error
  - Error values are normally distributed for any given value of X
- Equal Variance (also called homoscedasticity)
  - The probability distribution of the errors has constant variance

Chapter Summary

In this chapter we discussed:

- How to use regression analysis to predict the value of a dependent variable based on a value of an independent variable
- To understand the meaning of the regression coefficients b_0 and b_1
- To evaluate the assumptions of regression analysis and know what to do if the assumptions are violated
- To make inferences about the slope and correlation coefficient
- To estimate mean values and predict individual values