

Trendline and Regression Analysis (Part 1): Simple Linear Regression

BM 4419

Business Analytics



Introduction to Regression Analysis

DCOVAA

- 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

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

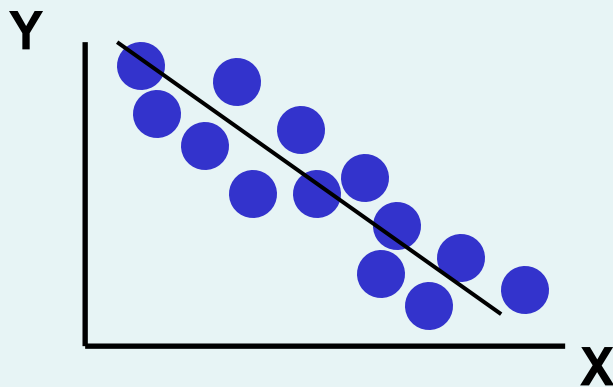
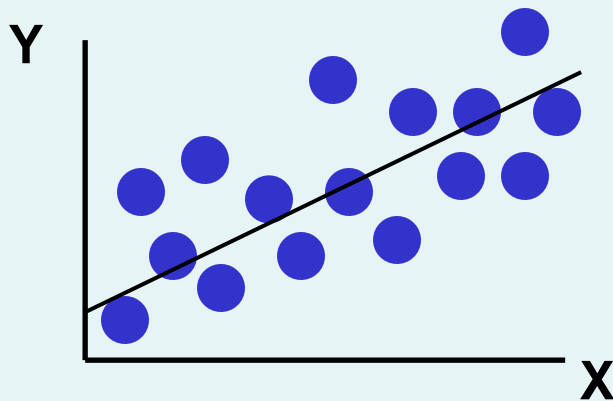
DCOVAA

- 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

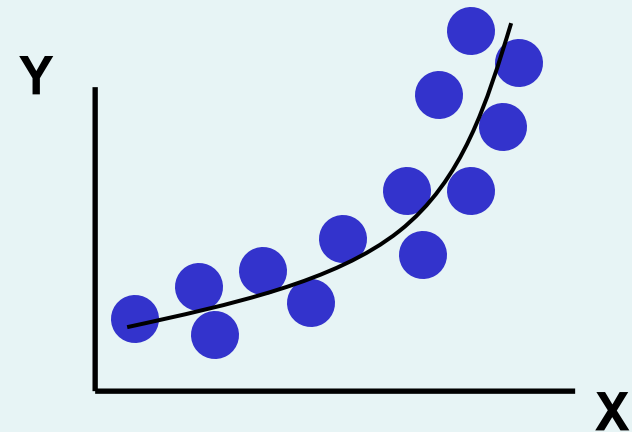
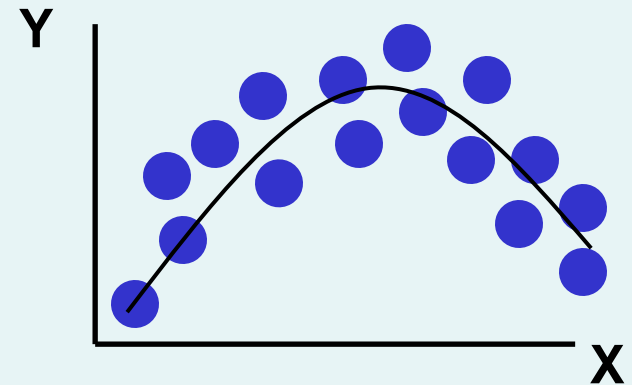
Types of Relationships

DCOVA

Linear relationships



Curvilinear relationships

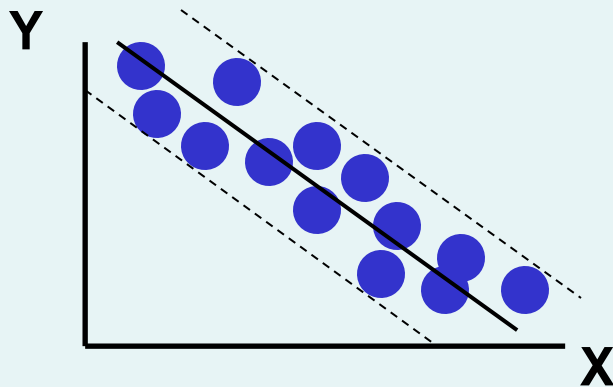
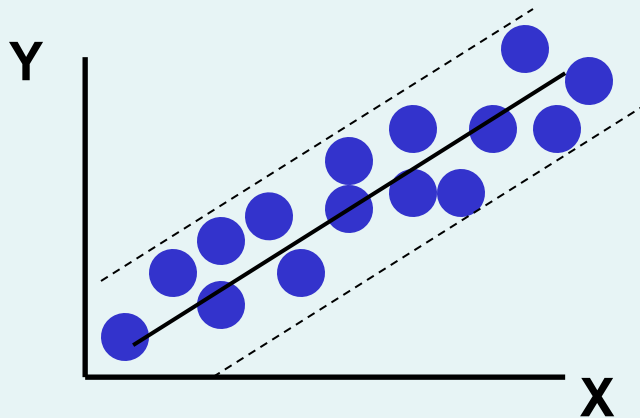


Types of Relationships

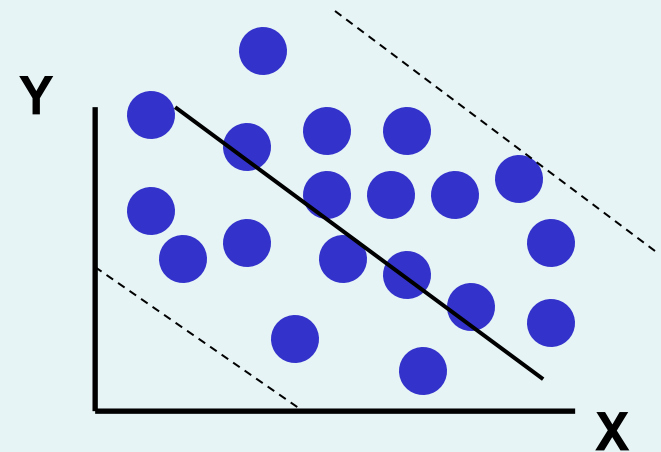
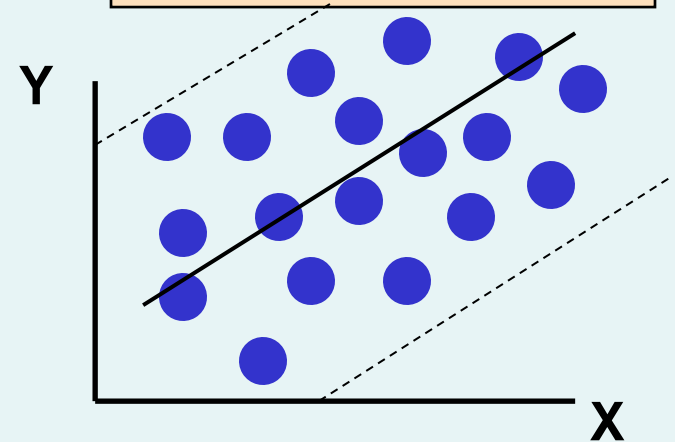
DCOVA

(continued)

Strong relationships



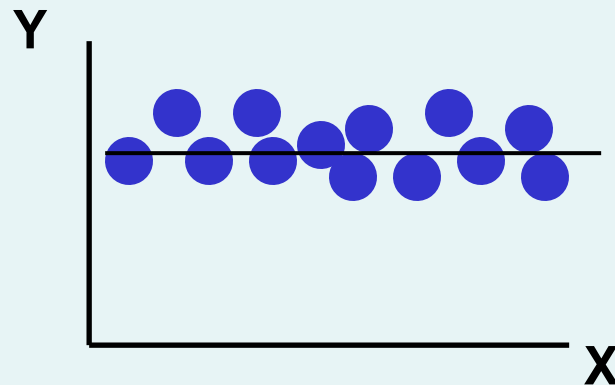
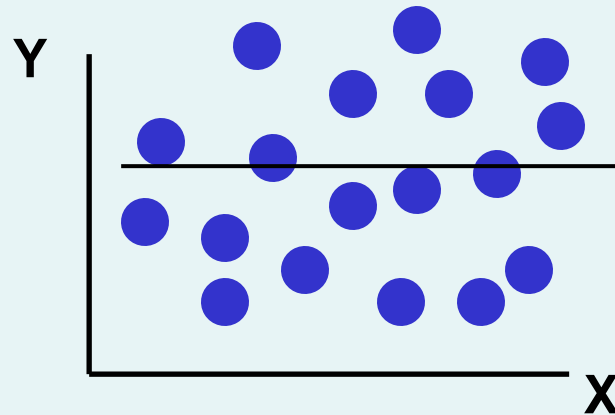
Weak relationships



Types of Relationships

DCOVA
(continued)

No relationship



Simple Linear Regression Model

DCOVA

Dependent Variable

Population Y intercept

Population Slope Coefficient

Independent Variable

Random Error term

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

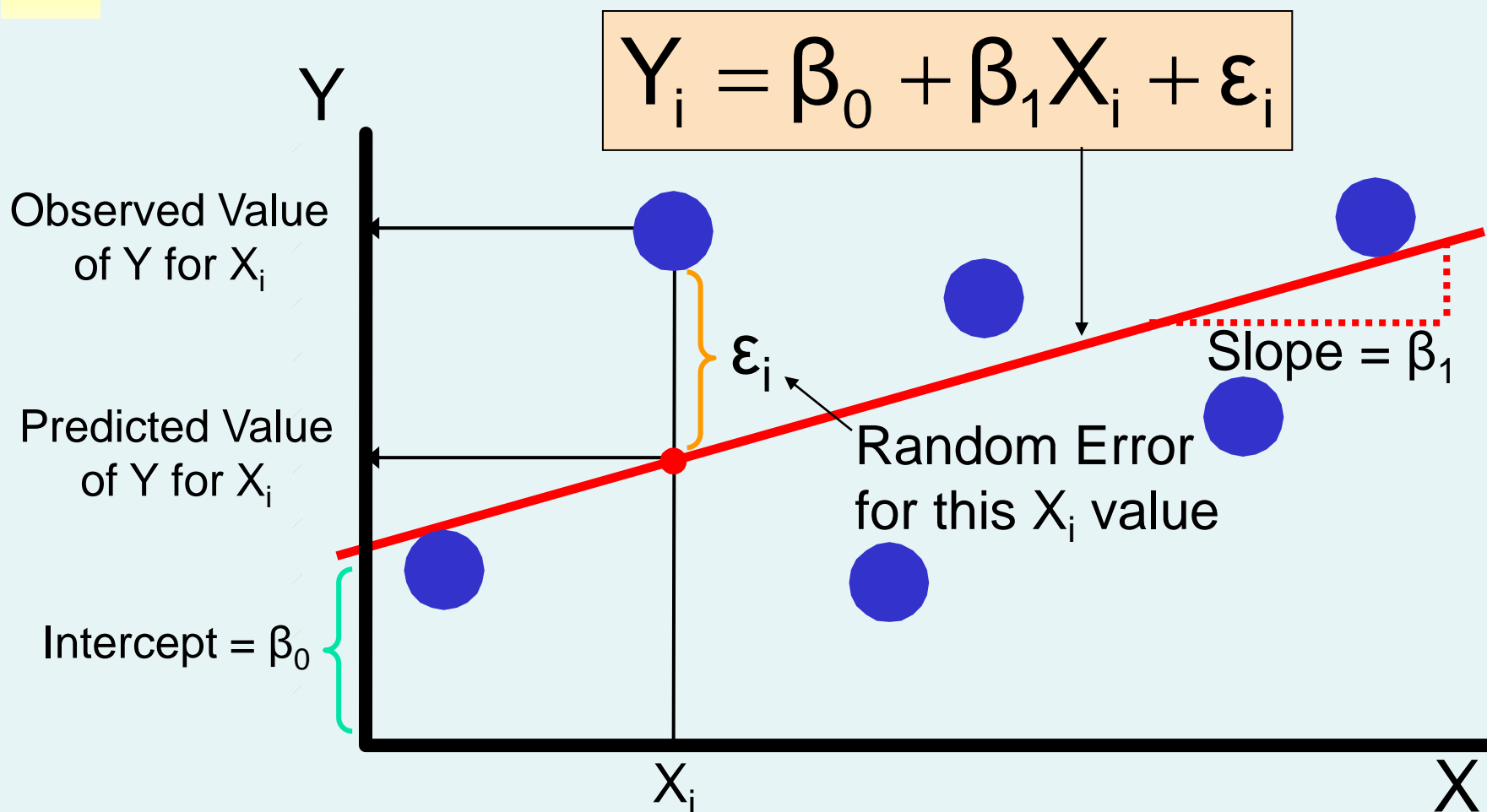
Linear component

Random Error component

The diagram illustrates the Simple Linear Regression Model equation: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$. The equation is enclosed in a light orange rectangular box. Labels with arrows point to specific parts of the equation: 'Dependent Variable' points to Y_i ; 'Population Y intercept' points to β_0 ; 'Population Slope Coefficient' points to β_1 ; 'Independent Variable' points to X_i ; and 'Random Error term' points to ε_i . Below the box, two blue curly braces group the terms: the first brace under $\beta_0 + \beta_1 X_i$ is labeled 'Linear component', and the second brace under ε_i is labeled 'Random Error component'.

Simple Linear Regression Model

DCOVA_A
(continued)



Simple Linear Regression Equation (Prediction Line)

DCOVA

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

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



Interpretation of the Slope and the Intercept

DCOVAA

- b_0 is the estimated average value of Y when the value of X is zero
- b_1 is the estimated change in the average value of Y as a result of a one-unit increase in X

Simple Linear Regression Example

DCOVA A

- 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

DCOVA

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

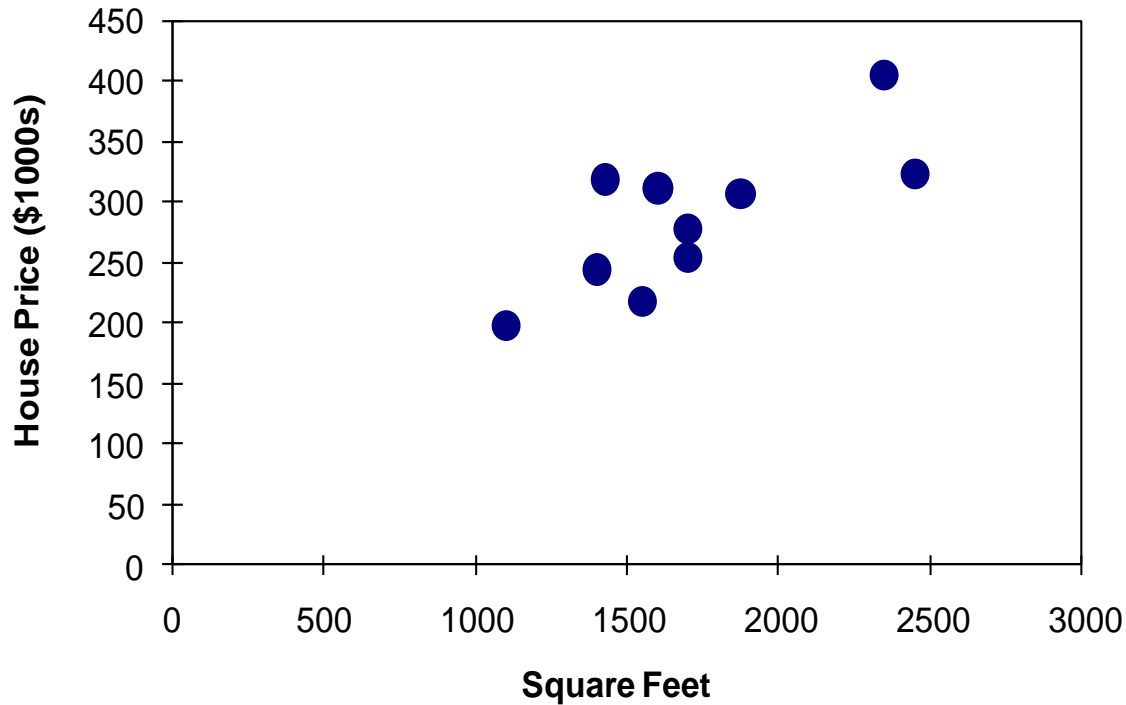


Simple Linear Regression

Example: Scatter Plot

DCOVA

House price model: Scatter Plot



Simple Linear Regression Example: Using Excel Data Analysis Function

DCOVA

1. Choose Data

2. Choose Data Analysis

3. Choose Regression

Chapter 13 examples.xlsx - Microsoft Excel

Home Insert Page Layout Formulas Data Review View Add-Ins Acrobat

From Access From Web From Text From Other Sources Get External Data

Existing Connections Refresh All Properties Edit Links Connections

Sort Filter Sort & Filter Clear Reapply Advanced

Text to Columns Remove Duplicates Data Validation Data Tools Consolidate What-If Analysis

Group Ungroup Subtotal Outline Analysis

Data Analysis

Analysis Tools

- Histogram
- Moving Average
- Random Number Generation
- Rank and Percentile
- Regression
- Sampling
- t-Test: Paired Two Sample for Means
- t-Test: Two-Sample Assuming Equal Variances
- t-Test: Two-Sample Assuming Unequal Variances
- z-Test: Two Sample for Means

	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																		



Simple Linear Regression Example: Using Excel Data Analysis Function

(continued)

Enter Y's and X's and desired options

DCOVA_A

	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:

Input X Range:

☐ Labels ☐ Constant is Zero

☐ Confidence Level: %

Output options

☒ Output Range:

☐ New Worksheet Ply:

☐ New Workbook

Residuals

☐ Residuals ☐ Residual Plots

☐ Standardized Residuals ☐ Line Fit Plots

Normal Probability

☐ Normal Probability Plots

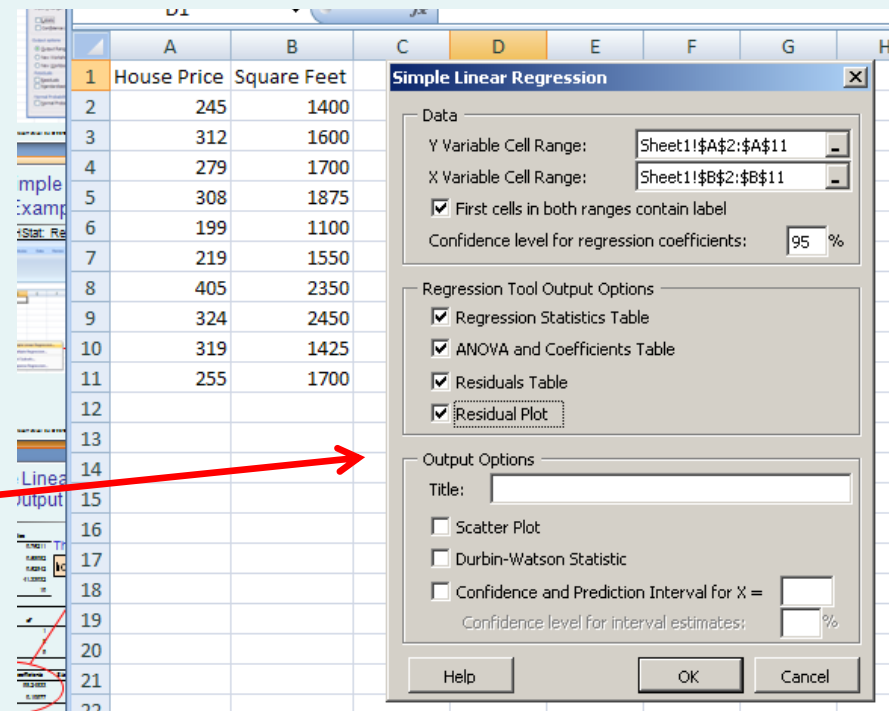
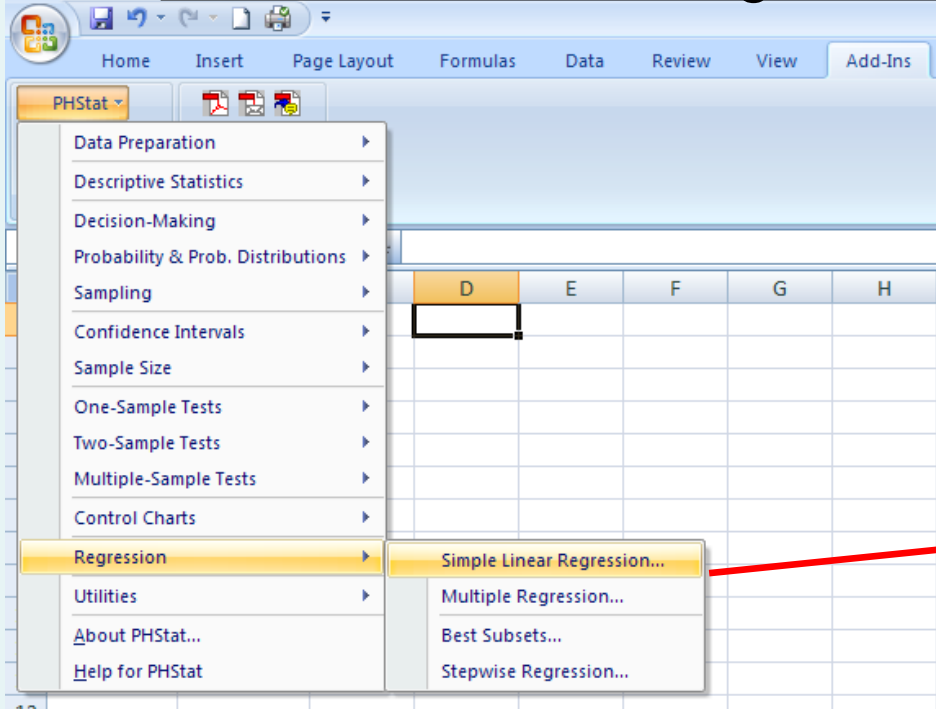
OK Cancel Help



Simple Linear Regression

Example: Using PHStat

Add-Ins: PHStat: Regression: Simple Linear Regression



Simple Linear Regression Example: Excel Output

DCOVA

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:

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

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			

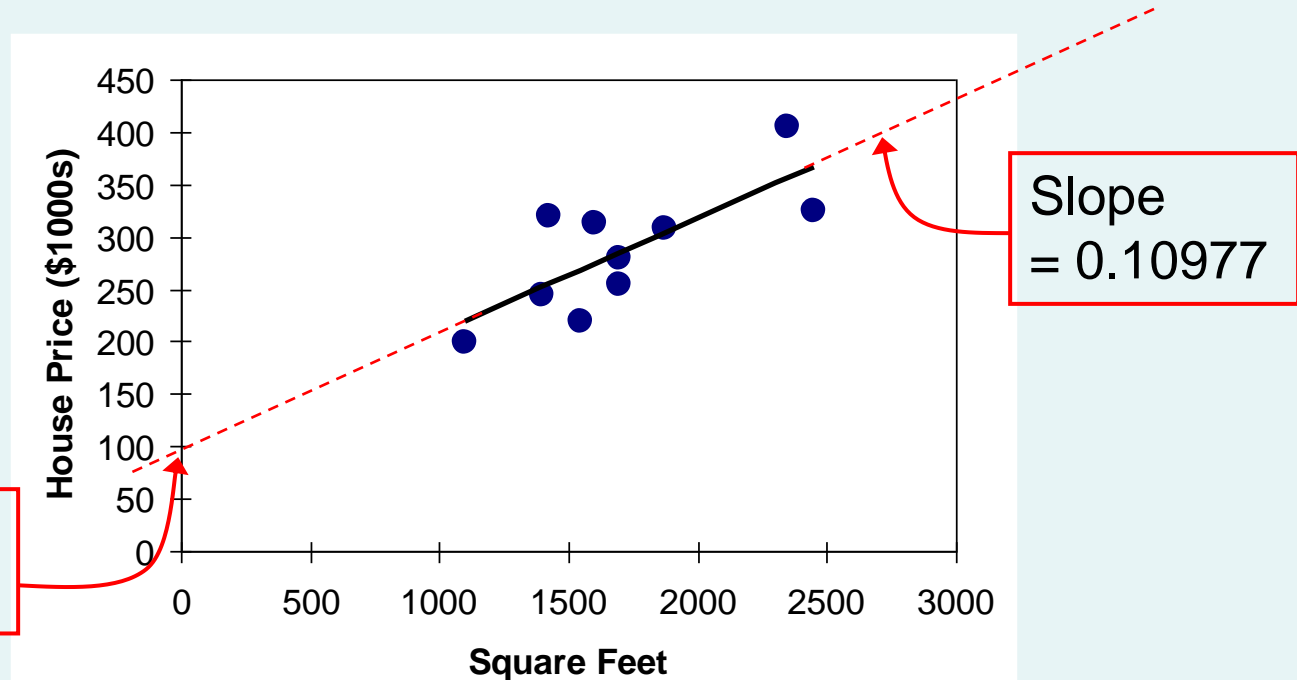
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	98.24833	58.03348	1.69296	0.12892	-35.57720	232.07386
Square Feet	0.10977	0.03297	3.32938	0.01039	0.03374	0.18580



Simple Linear Regression Example: Graphical Representation

DCOVA

House price model: Scatter Plot and Prediction Line



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

Simple Linear Regression

Example: Interpretation of b_0

DCOVA

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

- b_0 is the estimated average value of Y when the value of X is zero (if $X = 0$ is in the range of observed X values)
- Because a house cannot have a square footage of 0, b_0 has no practical application



Simple Linear Regression

Example: Interpreting b_1

DCOVA

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

- b_1 estimates the change in the average 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

DCOVA

Predict the price for a house
with 2000 square feet:

$$\begin{aligned}\widehat{\text{house price}} &= 98.25 + 0.1098 \text{ (sq.ft.)} \\ &= 98.25 + 0.1098(2000) \\ &= 317.85\end{aligned}$$

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



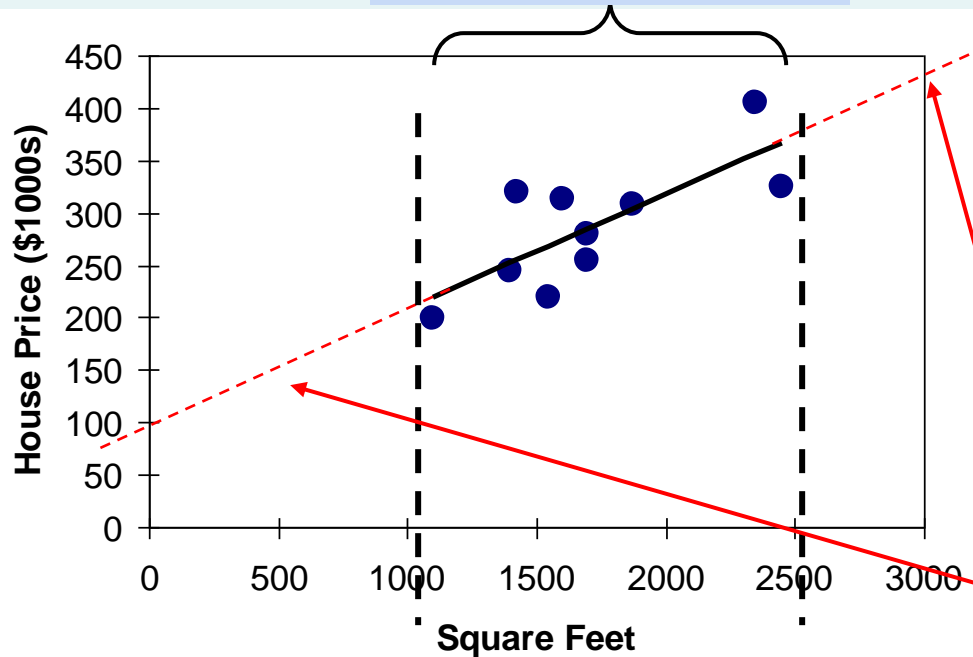
Simple Linear Regression

Example: Making Predictions

DCOVA

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

Relevant range for interpolation



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



Coefficient of Determination, r^2

DCOVA

- The **coefficient of determination** is the portion of the total variation in the dependent variable that is explained by variation in the independent variable
- The coefficient of determination is also called **r-squared** and is denoted as r^2

$$r^2 = \frac{SSR}{SST} = \frac{\text{regression sum of squares}}{\text{total sum of squares}}$$

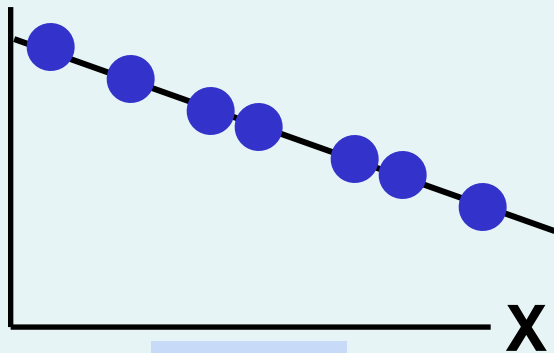
note:

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

Examples of Approximate r^2 Values

DCOVAA

Y

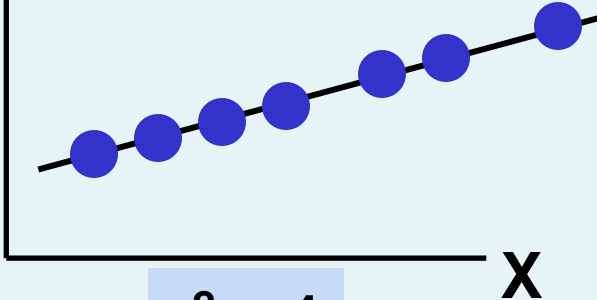


$$r^2 = 1$$

$$r^2 = 1$$

**Perfect linear relationship
between X and Y:**

Y

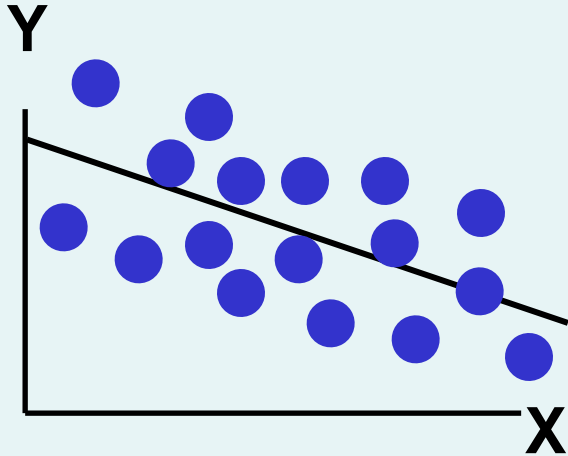


$$r^2 = 1$$

**100% of the variation in Y is
explained by variation in X**

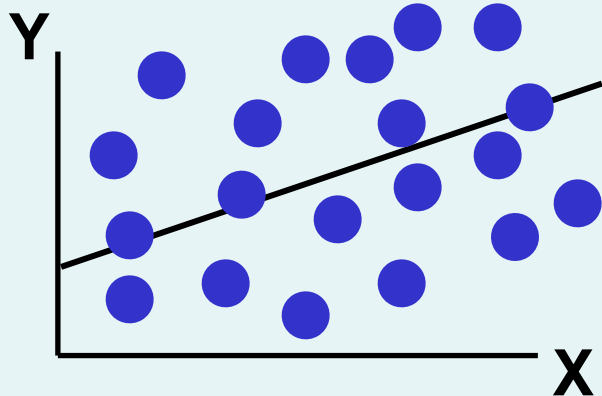
Examples of Approximate r^2 Values

DCOVAA



$$0 < r^2 < 1$$

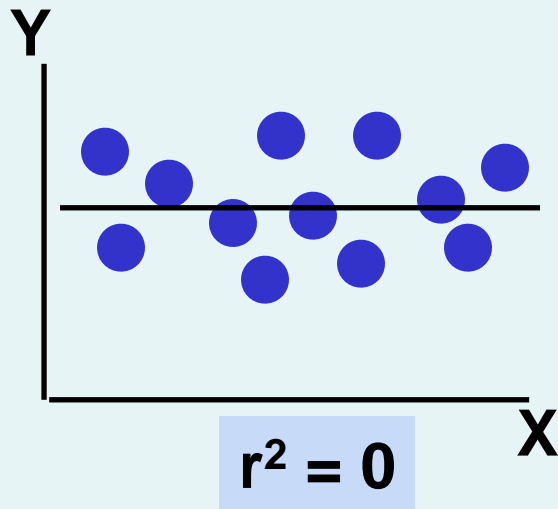
**Weaker linear relationships
between X and Y:**



**Some but not all of the
variation in Y is explained
by variation in X**

Examples of Approximate r^2 Values

DCOVA



$$r^2 = 0$$

**No linear relationship
between X and Y:**

**The value of Y does not
depend on X. (None of the
variation in Y is explained
by variation in X)**

Simple Linear Regression Example: Coefficient of Determination, r^2 in Excel

DCOVA

Regression Statistics

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

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

58.08% of the variation in house prices is explained by variation in square feet

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



EXAMPLE:

Eight students revealed the number of hours per day spent studying for the final exams and their relevant marks. These are tabulated in a table.

- Find the least square regression equation.
- Interpret the value of a and b .
- Predict the mark of final exam, if the student spent 9 hours and 15 minutes for studying.

Number of hours	2	6	8	1	10	7	6	3
Marks	40	50	80	20	60	80	90	40

SOLUTION:

$$b = \frac{SS_{xy}}{SS_{xx}} = 5.7090$$

$$a = \bar{y} - b\bar{x} = \frac{\sum y}{n} - b \frac{\sum x}{n} = 26.81$$

$$\hat{y} = 26.81 + 5.7090x$$