REGRESSION ANALYSIS

[And it's application in Business]

M.Ravishankar

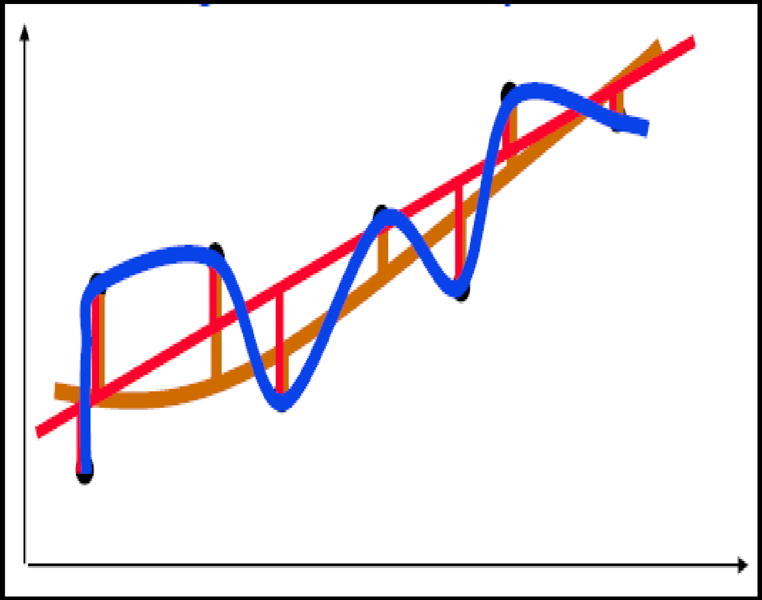
Introduction. . .

- Father of Regression Analysis
  Carl F. Gauss (1777-1855).

- contributions to physics, Mathematics & astronomy.

- The term “Regression” was first used in 1877 by Francis Galton.

Regression Analysis. . .

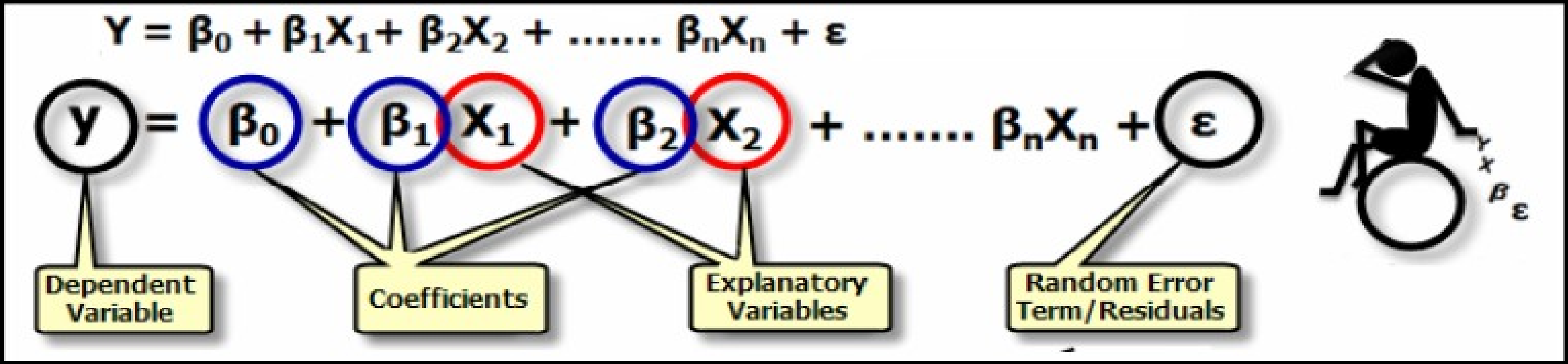


- It is the study of the relationship between variables.
- It is one of the most commonly used tools for business analysis.
- It is easy to use and applies to many situations.

Regression types. . .

- Simple Regression: single explanatory variable

- Multiple Regression: includes any number of explanatory variables.



- Dependant variable: the single variable being explained/ predicted by the regression model

- Independent variable: The explanatory variable(s) used to predict the dependant variable.

- Coefficients (β): values, computed by the regression tool, reflecting explanatory to dependent variable relationships.

- Residuals (ε): the portion of the dependent variable that isn't explained by the model; the model under and over predictions.

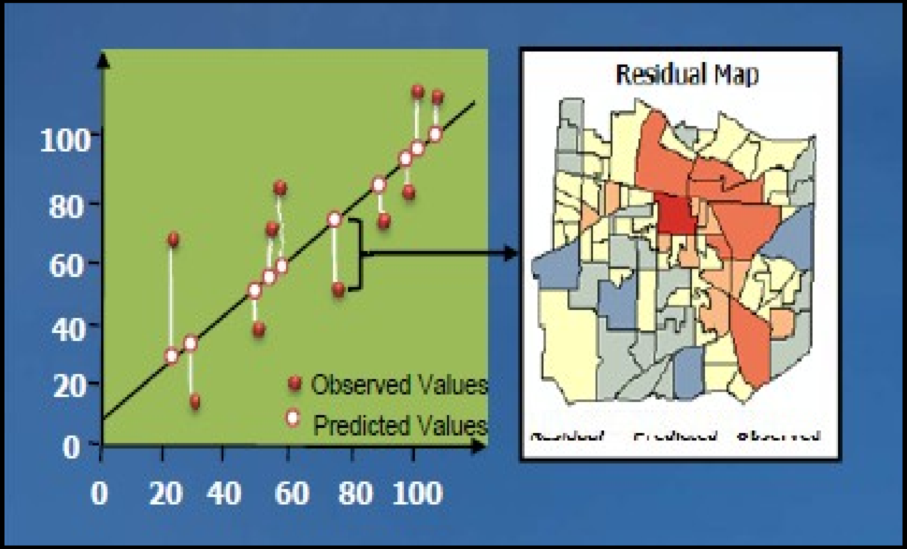
Regression Analysis. . .

- Linear Regression: straight-line relationship
  - Form: y=mx+b
- Non-linear: implies curved relationships
  - logarithmic relationships

Regression Analysis. . .

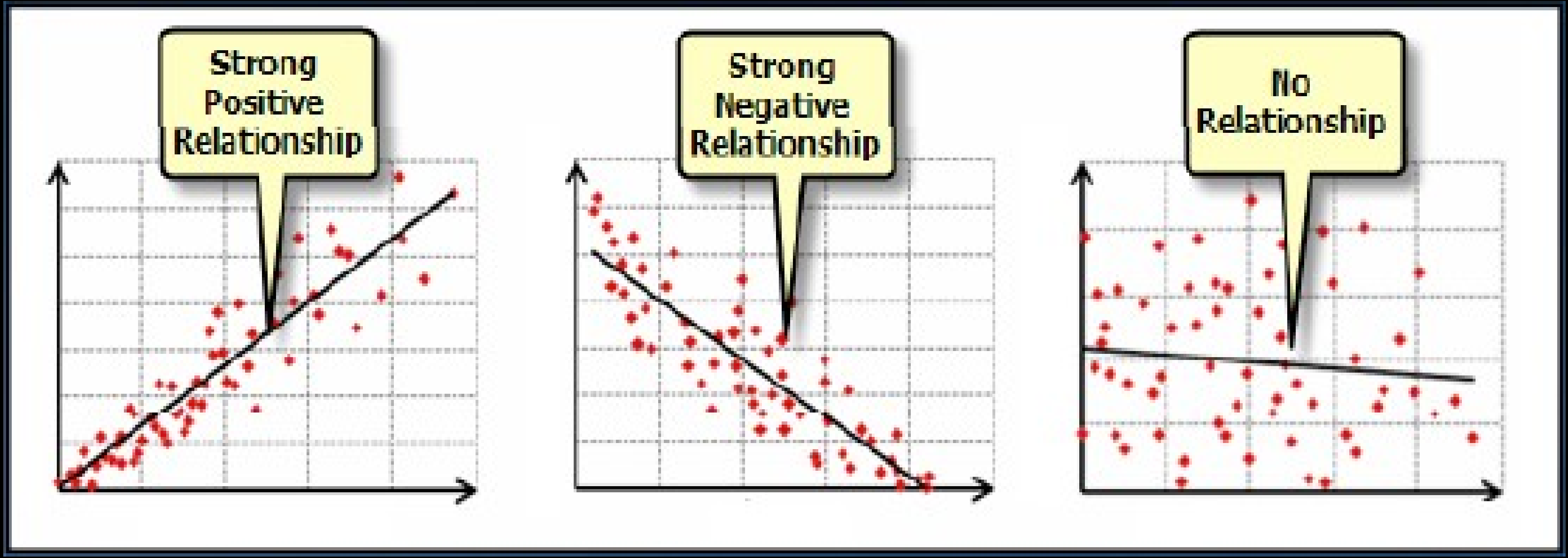
- Cross Sectional: data gathered from the same time period
- Time Series: Involves data observed over equally spaced points in time.

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 caused by changes in x

Types of Regression Models. . .



Estimated Regression Model. . .

The sample regression line provides an estimate of the population regression line

Estimated (or predicted) y value

Estimate of the regression intercept

Estimate of the regression slope

Independent variable

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

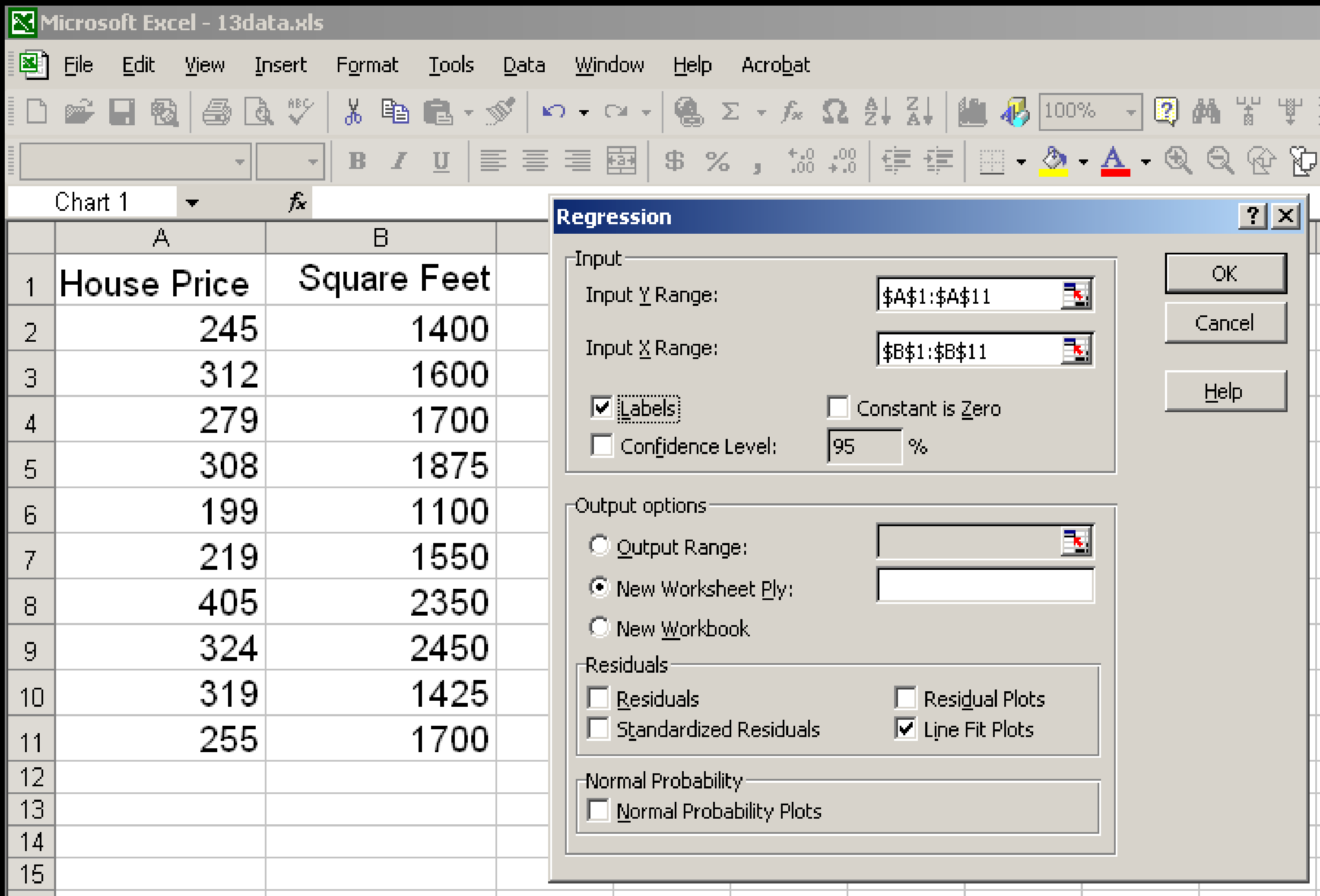
The individual random error terms e_i have a mean of zero

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

Sample 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



Microsoft Excel - 13data.xls
File Edit View Insert Format Tools Data Window Help Acrobat
100%
Chart 1
A
B
House Price
Square Feet
245 1400
312 1600
279 1700
308 1875
199 1100
219 1550
405 2350
324 2450
319 1425
255 1700
Regression
Input
Input Y Range:
$A$1:$A$11
Input X Range:
$B$1:$B$11
Labels
Constant is Zero
Confidence Level:
95
%
OK
Cancel
Help
Output options
Output Range:
New Worksheet Ply:
New Workbook
Residuals
Residuals
Residual Plots
Standardized Residuals
Line Fit Plots
Normal Probability
Normal Probability Plots

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:

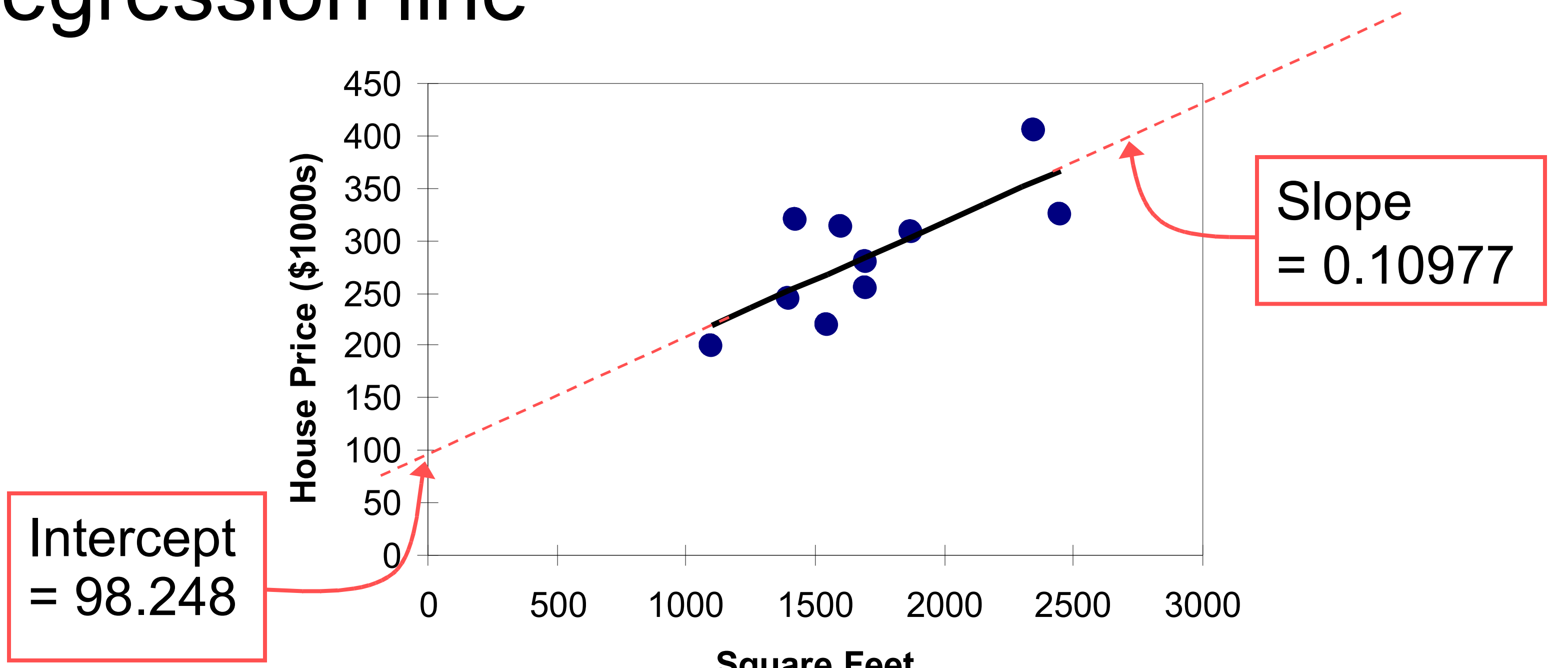
$$\widehat{\text{house price}} = 98.24833 + 0.10977\ (\text{square feet})$$

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

Graphical Presentation . . .

- House price model: scatter plot and regression line



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

Interpretation of the Intercept, b_0

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

- 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)
  - Here, no houses had 0 square feet, so b_0 = 98.24833 just indicates that, for houses within the range of sizes observed, $98,248.33 is the portion of the house price not explained by square feet

Interpretation of the Slope Coefficient, b_1

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

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

Example: House Prices

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

Estimated Regression Equation:

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

Predict the price for a house with 2000 square feet

Example: House Prices

Predict the price for a house with 2000 square feet:

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

$$= 98.25 + 0.1098(2000)$$

$$= 317.85$$

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

Coefficient of Determination, R^2

Coefficient of determination

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

Note: In the single independent variable case, the coefficient of determination is

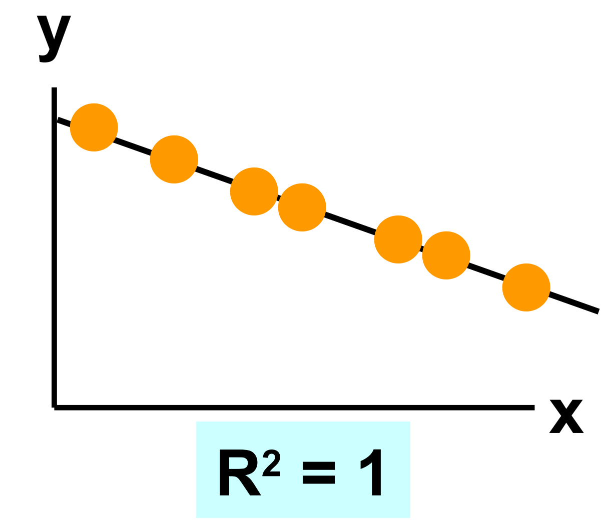
$$R^2 = r^2$$

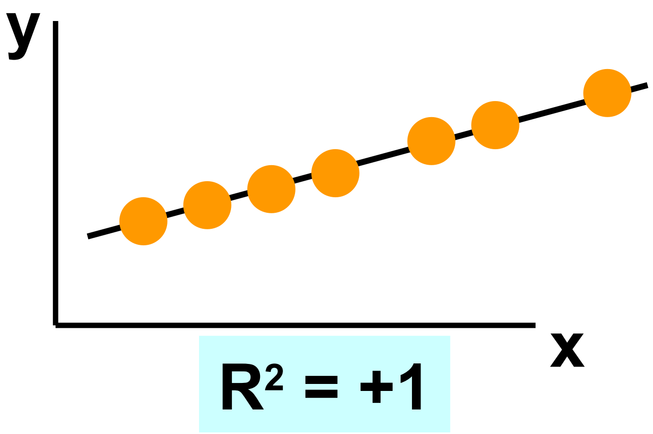
where:

R^2 = Coefficient of determination

r = Simple correlation coefficient

Examples of Approximate R^2 Values



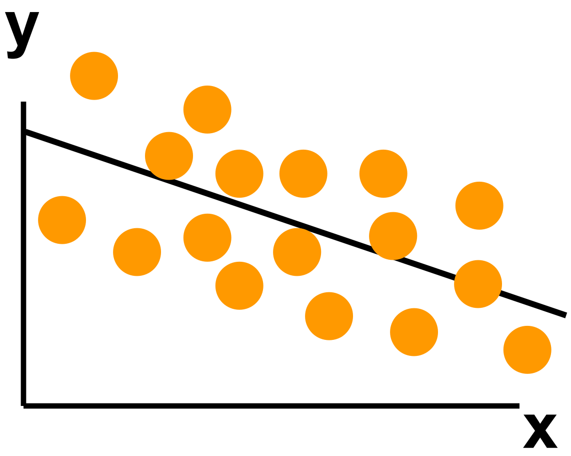


$R^2 = 1$

Perfect linear relationship between x and y:

100% of the variation in y is explained by variation in x

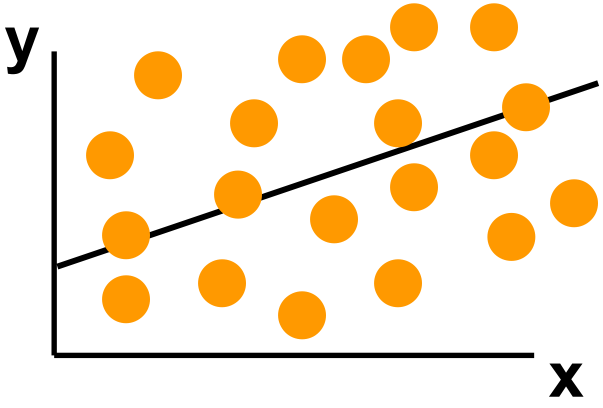
Examples of Approximate R^2 Values



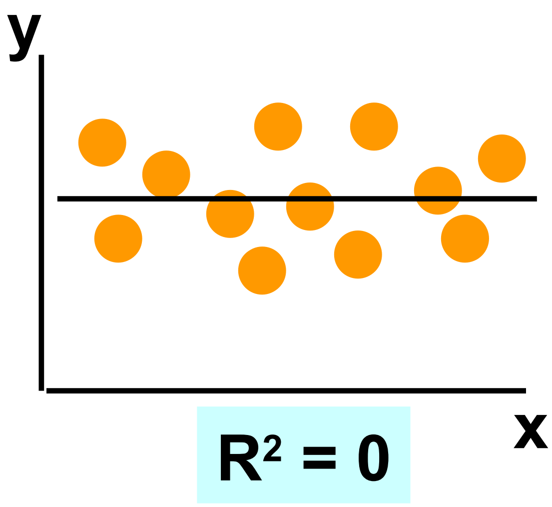
$0 < R^2 < 1$

Weaker linear relationship 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



$R^2 = 0$

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)

Output. . .

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	*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

Standard Error of Estimate. . .

- The standard deviation of the variation of observations around the regression line is estimated by

$$s_\varepsilon = \sqrt{\frac{SSE}{n-k-1}}$$

Where

SSE = Sum of squares error

n = Sample size

k = number of independent variables in the model

The Standard Deviation of the Regression Slope

- The standard error of the regression slope coefficient (b_1) is estimated by

$$s_{b_1} = \frac{s_\varepsilon}{\sqrt{\sum(x-\bar{x})^2}} = \frac{s_\varepsilon}{\sqrt{\sum x^2 - \frac{(\sum x)^2}{n}}}$$

where:

s_{b_1} = Estimate of the standard error of the least squares slope

$s_\varepsilon = \sqrt{\frac{SSE}{n-2}}$ = Sample standard error of the estimate

Output. . .

$$s_\varepsilon = 41.33032$$

$$s_{b_1} = 0.03297$$

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

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

Reference. . .

- Business statistics by S.P.Gupta & M.P.Gupta

Sources retrieved from Internet…

- www.humboldt.edu
- www.cs.usask.ca
- www.cab.latech.edu
- www.quickmba.com
- www.wikipedia.com
- www.youtube.com

M.RAVISHANKAR
MBA(AB) 2008-2010 Batch
NIFTTEA KNITWEAR FASHION INSTITUTE
TIRUPUR

OXYGEN024@GMAIL.COM

QUESTIONS?


TO
WIN
THE GAME,
JUST REMAIN
IN THE
GAME
FAILURE IS A PATH TO SUCCESS; SO DON'T GIVE UP, JUST MOVE ON WITH FIRM DETERMINATION
COPYRIGHTS SANTABANTA.COM
COPYRIGHTS SANTABANTA.COM
COPYRIGHTS SANTABANTA.COM
COPYRIGHTS SANTABANTA.COM