Regression and Correlation

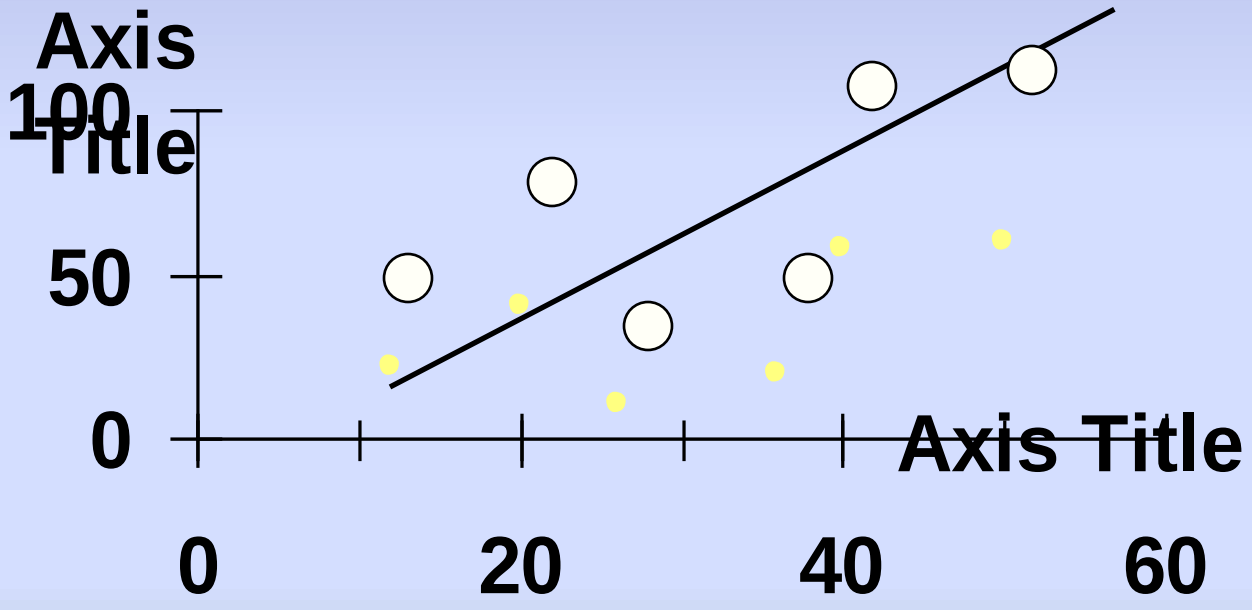
Purpose of Regression and Correlation Analysis

Regression Analysis is used primarily for **Prediction**

Correlation Analysis is used to **Measure Strength of the Association between Numerical Variables**

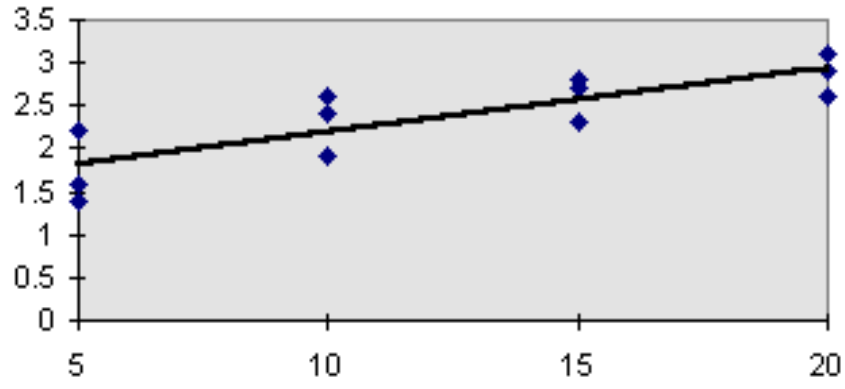
The Scatter Diagram

Plot of all (X_i , Y_i) pairs

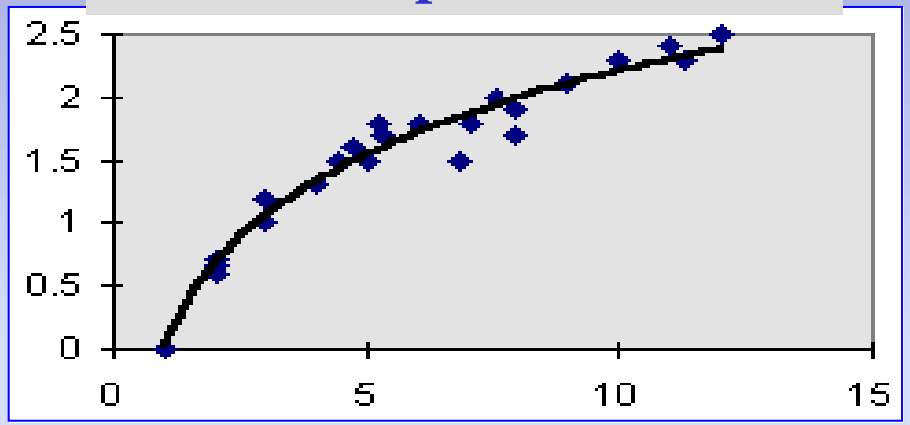


Types of Model

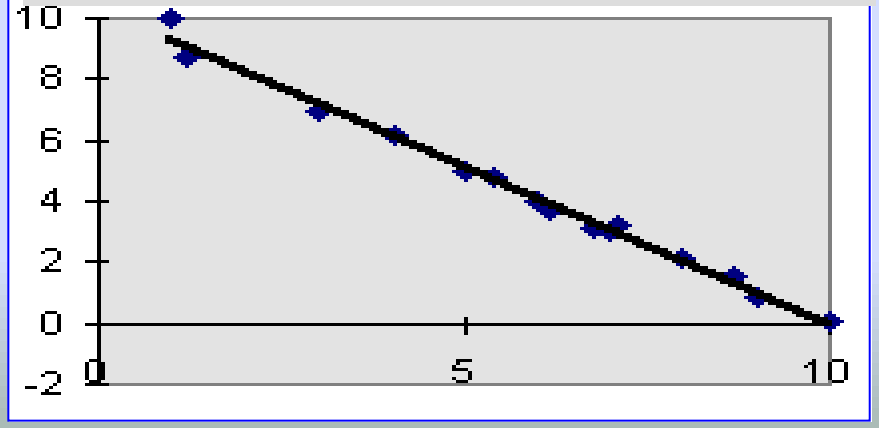
Positive Linear Relationship



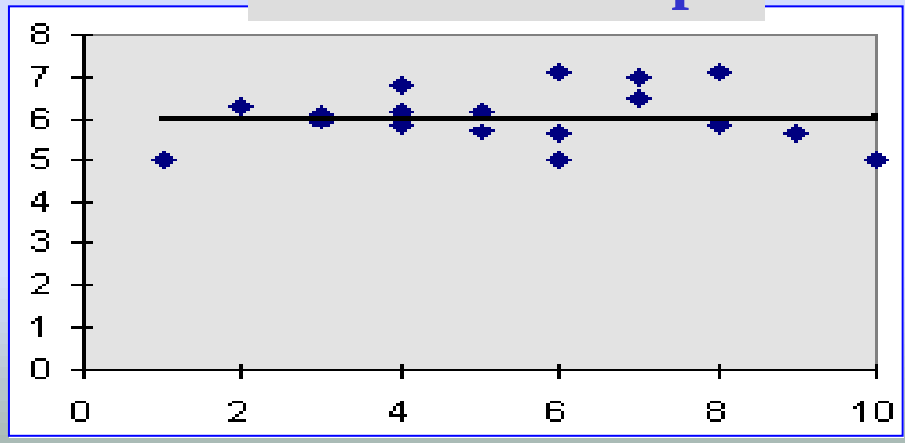
Relationship NOT Linear



Negative Linear Relationship

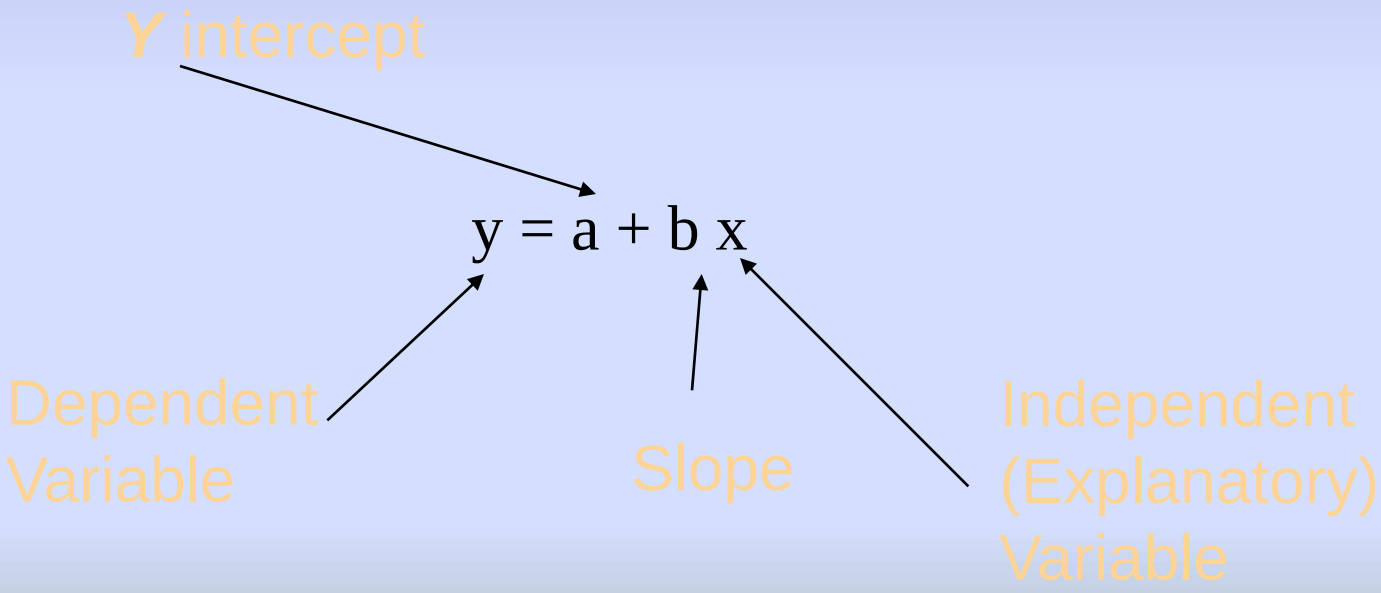


No Relationship

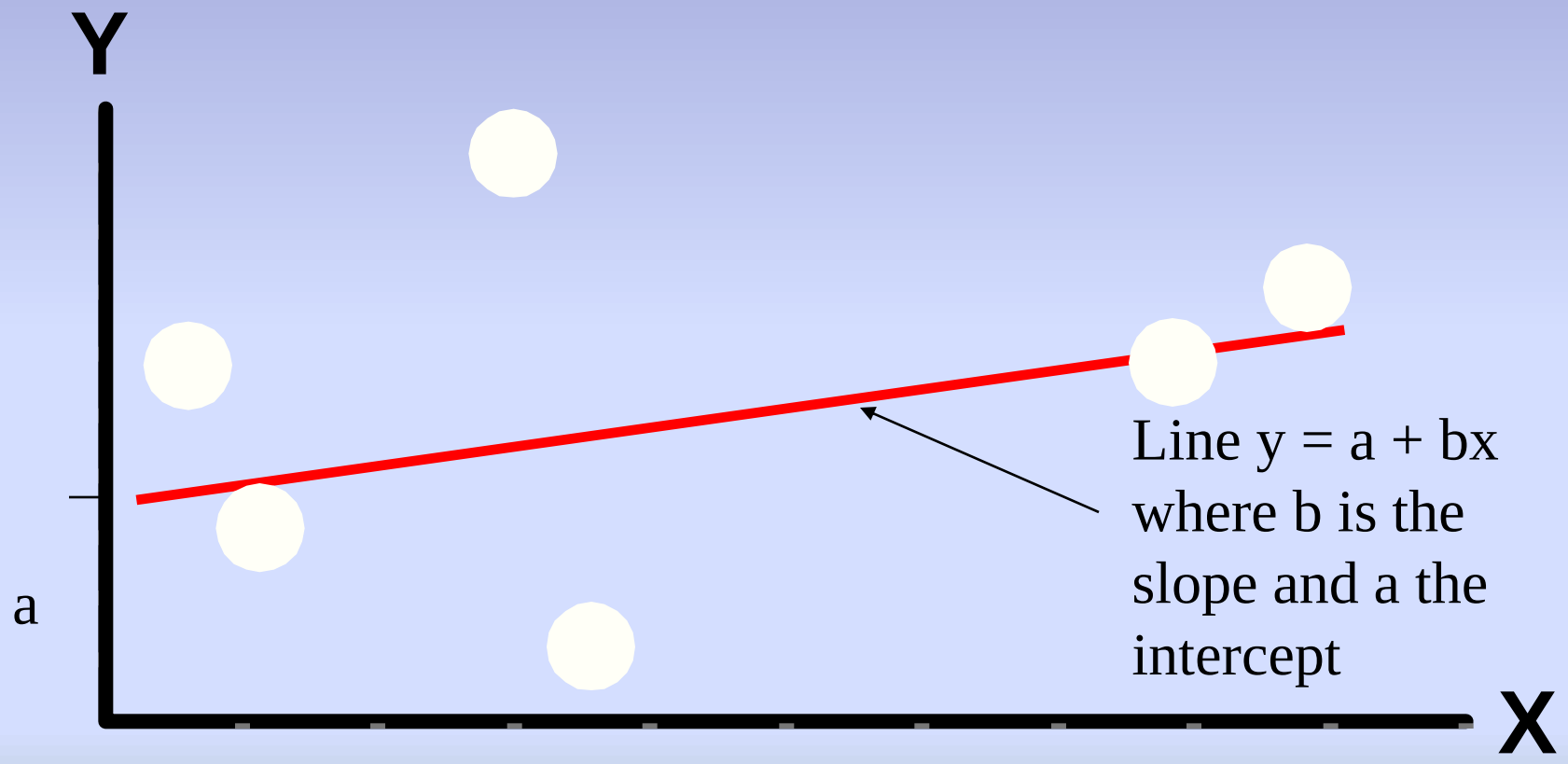


Simple Linear Regression Model

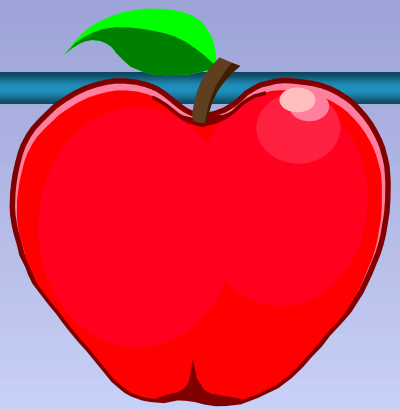
- Relationship Between Variables Is a Linear Function
- The Straight Line that Best Fit the Data



Linear Regression Model



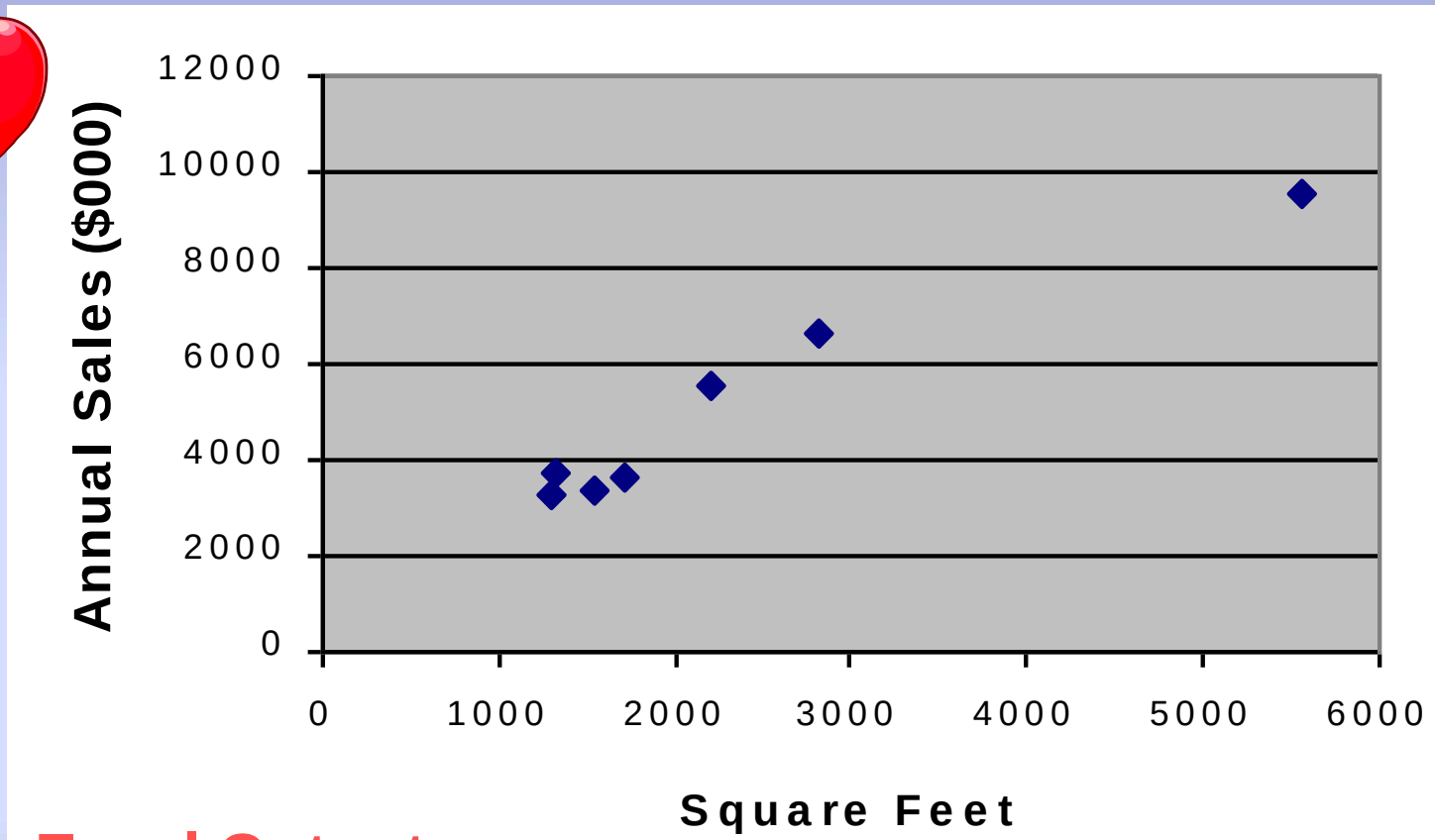
Simple Linear Regression Equation: Example



You wish to examine the relationship between the square footage of produce stores and its annual sales. Sample data for 7 stores were obtained. **Find the equation of the straight line that fits the data best**

Store	Square Feet	Annual Sales ($000)
1	1,726	3,681
2	1,542	3,395
3	2,816	6,653
4	5,555	9,543
5	1,292	3,318
6	2,208	5,563
7	1,313	3,760

Scatter Diagram Example



Excel Output

Equation for the Best Straight Line

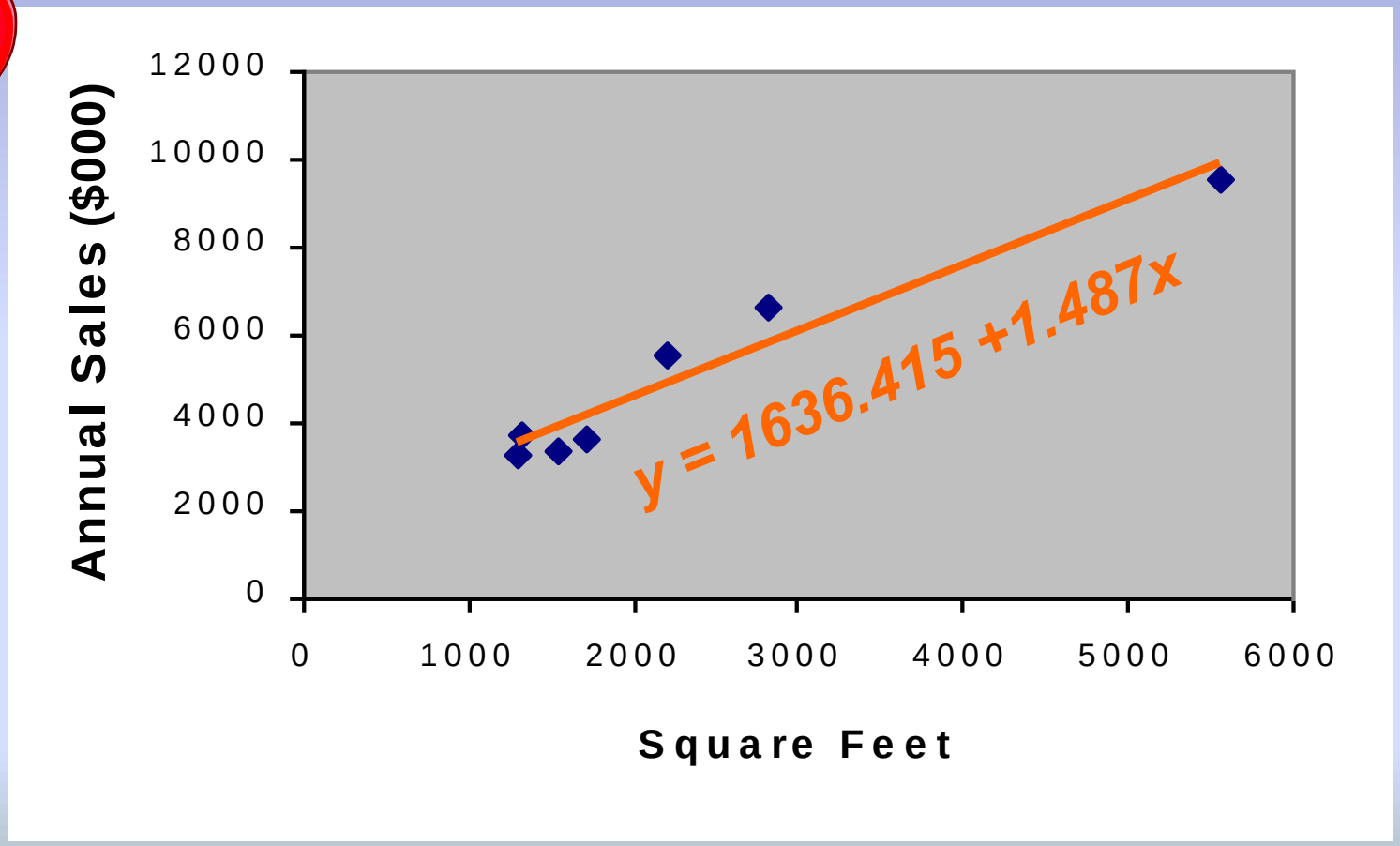
$$y = a + bx$$

$$= 1636.415 + 1.487x$$

From Excel Printout:

	Coefficients
Intercept	1636.414726
X Variable	1.486633657

Graph of the Best Straight Line



Interpreting the Results

$$y = 1636.415 + 1.487x$$

The slope of 1.487 means for each increase of one unit in X, the Y is estimated to increase 1.487units.

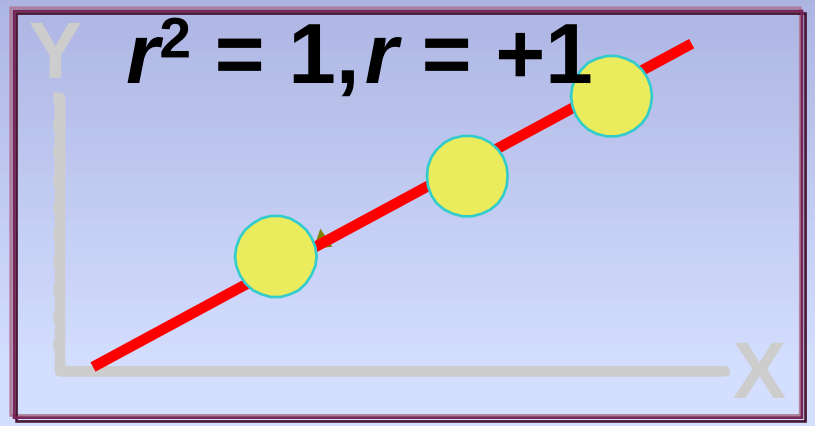
For each increase of 1 square foot in the size of the store, the model predicts that the expected annual sales are estimated to increase by $1487.

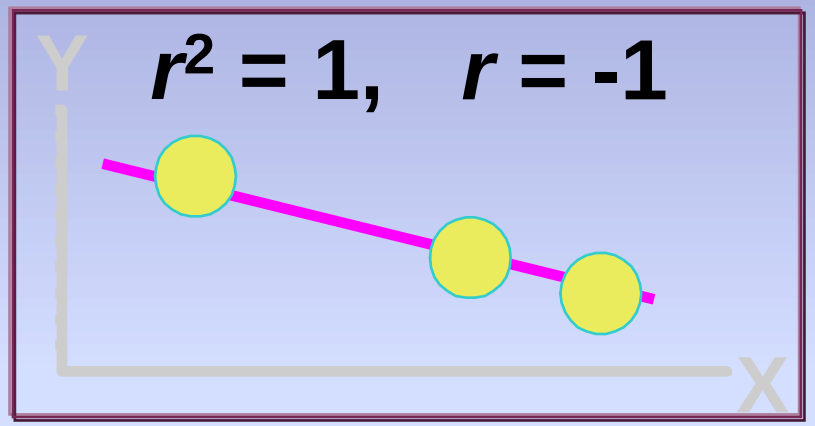
The Coefficient of Determination

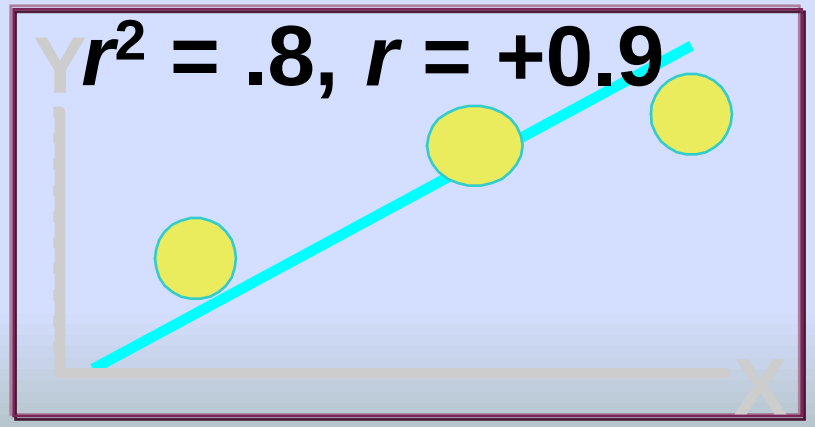
$$r^2 = \frac{SSR}{SST} = \frac{\text{regression sum of squares}}{\text{total sum of squares}}$$

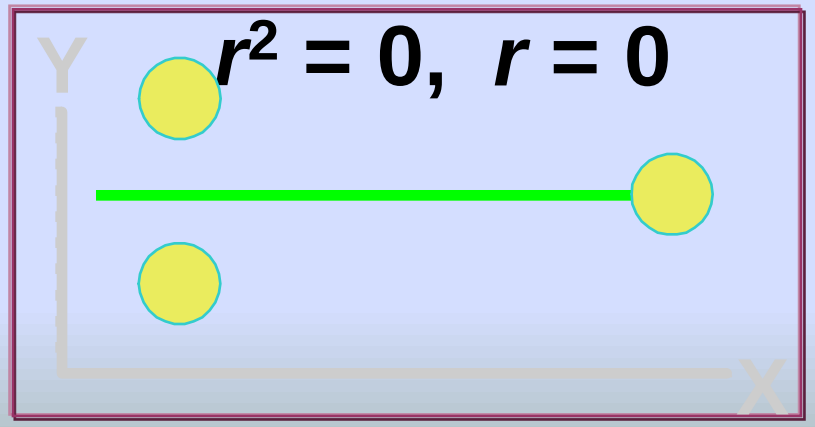
Measures the proportion of variation that is explained by the independent variable ***X*** in the regression model

Coefficients of Determination (r^2) and Correlation (r)









Measures of Variation: Example

Excel Output for Produce Stores

Regression Statistics	
Multiple R	0.9705572
R Square	0.94198129
Adjusted R Square	0.93037754
Standard Error	611.751517
Observations	7

$r^2 = .94$

94% of the variation in annual sales can be explained by the variability in the size of the store as measured by square footage

Correlation: Measuring the Strength of Association

- **Answer** ‘How Strong **Is the Linear Relationship Between 2 Variables?’**
- **Coefficient of Correlation Used**
  - **Values range from -1 to +1 (two variables can be perfectly correlated, partly correlated, uncorrelated**
  - **Measures degree of association**
- **Is the Square Root of the Coefficient of Determination**