Simple Linear Regression
Review of least squares procedure

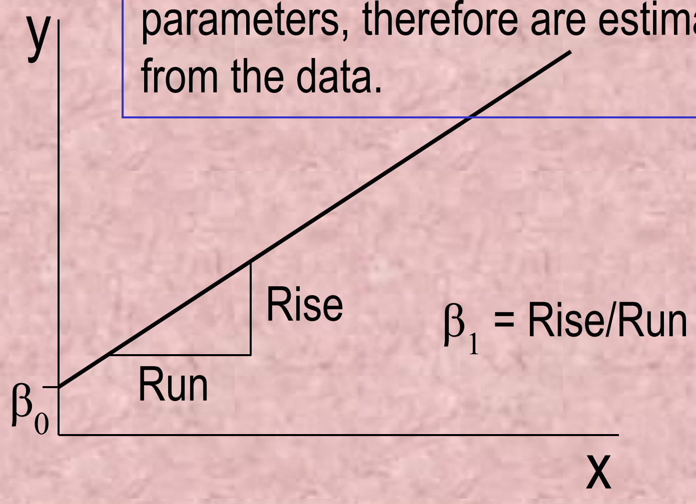
The Model

- The first order linear model

$$y = \beta_0 + \beta_1 x + \varepsilon$$

y = dependent variable
x = independent variable
β_0 = y-intercept
β_1 = slope of the line
ε = error variable

β_0 and β_1 are unknown population parameters, therefore are estimated from the data.



The Least Squares (Regression) Line

A good line is one that minimizes
the sum of squared differences between the
points and the line.

The Estimated Coefficients

To calculate the estimates of the slope and intercept of the least squares line , use the formulas:

$$b_1 = \frac{SS_{xy}}{SS_{xx}}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

$$SS_{xy} = \sum x_i y_i - \frac{\left(\sum x_i\right)\left(\sum y_i\right)}{n}$$

$$SS_{xx} = \sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n} = (n-1)s_x^2$$

Alternate formula for the slope b_1

$$b_1 = r\frac{s_y}{s_x}$$

The regression equation that estimates the equation of the first order linear model is:

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

The Simple Linear Regression Line

- Example:
  - A car dealer wants to find the relationship between the odometer reading and the selling price of used cars.
  - A random sample of 100 cars is selected, and the data recorded.
  - Find the regression line.

Car	Odometer	Price
1	37388	14636
2	44758	14122
3	45833	14016
4	30862	15590
5	31705	15568
6	34010	14718
.	Independent	Dependent
.	variable x	variable y
.	.	.

The Simple Linear Regression Line

- Solution
  - Solving by hand: Calculate a number of statistics

$\bar{x} = 36{,}009.45;$ $$SS_{xx} = \sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n} = 43{,}528{,}690$$

$\bar{y} = 14{,}822.823;$ $$SS_{xy} = \sum (x_i y_i) - \frac{\sum x_i \sum y_i}{n} = -2{,}712{,}511$$

where n = 100.

$$b_1 = \frac{SS_{xy}}{(n-1)s_x^2} = \frac{-2{,}712{,}511}{43{,}528{,}690} = -.06232$$

$$b_0 = \bar{y} - b_1\bar{x} = 14{,}822.82 - (-.06232)(36{,}009.45) = 17{,}067$$

$$\hat{y} = b_0 + b_1 x = 17{,}067 - .0623x$$