# Lesson 19 - Linear Regression

Two Quantitative Variables

Association between two Quantitative variables
examined through scatterplot.

- Direction - Positive or negative association
- Form - The shape. Linear or not.
- Strength - concentrated in a shape or scattered

Correlation Coefficient ($r$)

$-1 \leq r \leq 1$ measure of the linear association

Linear Regression - when association is linear, a line is used as a math model to summarize relationship and predict values of response var from explanatory var.

least squares regression line - Slope and intercept combo that minimizes the sum of squared error

$$\hat{y} = \beta_0 + \beta_1 x_1$$

* Slope and intercept interpretation

Residuals $(y - \hat{y})$ - The error.

Coefficient of Determination

$R^2 \rightarrow$ The variation explained by $x$

Association between 2 Quant variables - Theory Based

* Use $\beta$ (Slope) or $\rho$ ($r \rightarrow$ correlation)

$H_0: \beta = 0$ There is no association between [explanatory] and [response]

$H_a: \beta \neq 0$ There is an association .. .. .. ..

$$t = \frac{\beta_1 - 0}{SE(\beta_1)} \quad \text{or} \quad \frac{r}{\sqrt{\frac{1 - r^2}{n-2}}}$$

references the slope
↓

95% Confidence Interval → 95% confident that a one unit increase in explanatory variable is associated w/ a # to # increase in response variable in the population

Validity Conditions

Linearity - predicted vs. residuals plot is liner

Independence - each response is independent of the other observations

Normality - histogram of residuals is normally distributed

Equal Variance - predicted vs. residuals plot show equal variance

Interpretation of R output for regression