Ch 12.1

- Least-squares regression fits a straight line of the form $\hat{y} = a + bx$ to data to predict a response variable y from an explanatory variable x. Inference in this setting uses the **sample regression line** to estimate or test a claim about the **population (true) regression line**.
- The conditions for regression inference are
  - **Linear:** The actual relationship between x and y is linear. For any fixed value of x, the mean response μ_y falls on the population (true) regression line $\mu_y = \alpha + \beta x$.
  - **Independent:** Individual observations are independent. When sampling is done without replacement, check the *10% condition*.
  - **Normal:** For any fixed value of x, the response y varies according to a Normal distribution.
  - **Equal SD:** The standard deviation of y (call it σ) is the same for all values of x.
  - **Random:** The data are produced from a well-designed random sample or randomized experiment.
- The slope b and intercept a of the sample regression line estimate the slope β and intercept α of the population (true) regression line. Use the standard deviation of the residuals, s, to estimate σ.
- Confidence intervals and significance tests for the slope β of the population regression line are based on a t distribution with $n - 2$ degrees of freedom.
- The ***t* interval for the slope** β has the form $b \pm t^* SE_b$, where the **standard error of the slope** is $SE_b = \dfrac{s}{s_x\sqrt{n-1}}$.
- To test the null hypothesis $H_0: \beta = \beta_0$, carry out a ***t* test for the slope.** This test uses the statistic $t = \dfrac{b - \beta_0}{SE_b}$. The most common null hypothesis is $H_0: \beta = 0$, which says that there is no linear relationship between x and y in the population.

DO NOT FIND THE TEST STATISTIC AND P-VALUE BY HAND. USE YOUR CALCULATOR OR MINITAB!!!

C12.2

- Curved relationships between two quantitative variables can sometimes be changed into linear relationships by **transforming** one or both of the variables. Once we transform the data to achieve linearity, we can fit a least-squares regression line to the transformed data and use this linear model to make predictions.
- When theory or experience suggests that the relationship between two variables follows a **power model** of the form $y = ax^p$, there are two transformations involving powers and roots that can linearize a curved pattern in a scatterplot. Option 1: Raise the values of the explanatory variable x to the power p, then look at a graph of (x^p, y). Option 2: Take the pth root of the values of the response variable y, then look at a graph of $(x, \sqrt[p]{y})$
- Another useful strategy for straightening a curved pattern in a scatterplot is to take the **logarithm** of one or both variables. When a power model describes the relationship between two variables, a plot of log y (ln y) versus log x (ln x) should be linear.
- In a linear model of the form $y = a + bx$, the values of the response variable are predicted to increase by a constant amount b for each increase of 1 unit in the explanatory variable. For an **exponential model** of the form $y = ab^x$, the predicted values of the response variable are multiplied by a factor of b for each increase of 1 unit in the explanatory variable. When an exponential model describes the relationship between two variables, a plot of log y (ln y) versus x should be linear.

- Ignore the transformations that don't involve logs