Chapter 11 Conceptual Review   Date: 5/1/25

- The independent variable is represented by X and the dependent variable is represented by Y.
- There are 3 possible relationships between the 2 variables:
  - Positive/Negative Linear
  - Non-Linear
  - No Relationship
- There are 2 models to know:
  - Population Linear Model (the truth)

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

  - Linear Regression Model (the estimate)

$$\hat{y}=\hat{\beta}_0+\hat{\beta}_1 x$$

  - For these models:
    - Slope is notated as β_1 and $\hat{\beta}_1$ depending on the model
    - Vertical intercept is notated as β_0 and $\hat{\beta}_0$ depending on the model
    - y is the true value
    - $\hat{y}$ is the estimate of the true value
    - ϵ is the noise (and is only used in the Population Linear Model)
- The smaller the sum of squares error is, the better the linear regression line fit.
- The Coefficient of Correlation is notated by R

- The strength of R is determined by how close R is to the extremes 1 and –1, as the values are greater than or equal to –1 and less than or equal to 1 → $-1 \le R \le 1$
- Rough Guide: (+ or –)
  - 1 : Perfect
  - 0.8–0.99 : Strong
  - 0.5–0.79 : Moderate
  - 0.01–0.49 : Weak
  - 0 : No Association

- If 1 of the 2 variables has a direct influence on the other, that is known as a casual relationship
- The Coefficient of Determination is notated by R^2
  - This value is always between 0 & 1
- A residual plot is a graph that pairs the variable X with the error for each value
  - If the linear regression is successful, these residuals should be small and unstructured
  - These plots expose outliers (values more than 3 standard deviantions from the mean)
  - The structure appears when the relationship in not linear
- A standardized residual plot is a residual plot where all the residuals are divided by the residual standard error.
  - The residual standard error is notated as σ_ε

- An influential point is an observation that effects the regression equation
  - These points, when removed, change the position of the regression line quite a bit.
- Relevant Calculator Methods:
  - Scatter Plots
    - Insert data into lists

      Stat → Edit → 1: Edit
    - Make your scatter plot

      2nd → Y= → Select a plot → Turn on → Specify lists → Graph
    - Zoom in to your plot

      Zoom → 9: ZoomStat
  - LinRegTTest
    - Insert data into lists

      Stat → Edit → 1: Edit
    - Use the function

      Stat → Tests → F: LinRegTTest → Specify lists
  - LinReg(a+bx)
    - Insert data into lists

      Stat → Edit → 1: Edit

- Use the function

Stat → Calc → 8:LinReg(a+bx) → Specify lists

- Formulas:
    - Error

$$\epsilon = y_i - \hat{y}_i$$

    - Sum of Squares Error

$$SS\epsilon = \sum (\epsilon)^2$$

    - Estimated Slope

$$\hat{\beta}_1 = \frac{n\sum(x_i y_i) - (\sum x_i)\cdot(\sum y_i)}{n\sum(x_i^2) - \sum(x_i)^2}$$

    - Estimated Vertical Intercept

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

    - Coefficient of Correlation

$$R = \frac{n\cdot(\sum x_i y_i) - (\sum x_i)\cdot(\sum y_i)}{\left(\sqrt{n\cdot(\sum x_i^2) - \sum(x_i)^2}\right)\cdot\left(\sqrt{n\cdot(\sum y_i^2) - \sum(x_i)^2}\right)}$$

    - Coefficient of Determination

$$R^2 = (R)^2$$

- Standardized Residual Plot Point

$$sE = \frac{\epsilon_i}{\sigma_\epsilon}$$

- Hypotheses and Interpretations
  - Both have the hypotheses where null states that $\beta_1 = 0$ while the alternative states $\beta_1 \neq 0$
  - For Correlation, the interpretation is a description of the relationship
    - Example: Strong positive linear association
  - For Determination, the interpretation is a statement in relation to the hypotheses
    - $\boxed{R^2\%}$ of the variation in y can be explained by x.
    - We have sufficient/insufficient evidence to say that the true population slope is not 0.