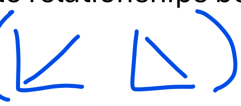
<u>Chapter 11 Theory</u>

- The independent variable is represented by X, and the dependent variable is represented by Y.

- There are 3 possible relationships between the 2 variables:
  - Linear ()
  - Non-Linear ( etc.)
  - 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)

      or error

- The smaller the <u>sum of squares error (SSE)</u> is, the better the linear regression line fit.
- The Coefficient of Correlation is notated by <u>R</u>
    - The strength of R is determined by how close R is to the extremes <u>1</u> and <u>-1</u>, as the values are <u>≥ -1 or ≤ 1</u>
    - Rough Guide: (+/-)
        - 0.01-0.19 → Very weak
        - 0.20-0.39 → Weak
        - 0.40 -0.59 → Moderate
        - 0.60-0.79 → Strong
        - 0.80-0.99 → Very Strong

0 → No correlation
&
1 → Perfect correlation

    - The <u>slope (β_1 or $\hat{\beta}_1$)</u> shares the same sign (<u>+</u>/<u>−</u>) with <u>R</u>
- If 1 of the 2 variables has a direct influence on the other, that is known as a <u>Casual relationship</u>
- The Coefficient of Determination is notated by <u>R^2</u>
    - This value is always between <u>0 & 1</u>
- A <u>residual plot</u> is a graph that pairs the variable <u>X</u> with the <u>error</u> (ϵ) for each value
    - If the linear regression is successful, these residuals should be <u>small</u> and <u>unstructured</u>
    - These plots expose <u>outliers</u> (values more than <u>3</u> standard deviations from the mean)

- The structure appears when the relationship in not linear

- A standard 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 affects 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
    - Use the function

      Stat → Tests → F: LinRegTTest → Specify lists

- LinReg(a+bx)
  - Insert data into lists
  - Use the function

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

- Formulas:
  - Error

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

  - Sum of Squares Error

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

  - 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

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

  - Coefficient of Determination

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

- Standardized Residual Plot Point

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

- 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
    - $[R^2]\%$ 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.