Exam 4 Theory

- There are 3 possible relationships between variables X and Y
  - Linear or etc.
  - 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 $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$
    - This is what we are calculating/working with
- Model notations:
  - β_0 & $\hat{\beta}_0$ = Vertical Intercept
  - β_1 & $\hat{\beta}_1$ = Slope
  - y = the true y-value
  - $\hat{y}$ = the estimated y-value
  - ϵ = Error or Noise
- The smaller the sum of squares error (SSE) is, the better the linear regression line fit is.
- Coefficient of Correlation (R)
  - The strength of R is determined by how close R is to the extremes 1 & -1
  - Guide for Interpretation (Memorize)

    Think 0, 5 levels with ~0.20 in each, then 1.

    Can be + or −
    - 0 : No Relationship
    - 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
    - 1 : Perfect
  - The slope shares the same sign (+/−) with R

- Coefficient of Determination (R^2)
  - This number is always between 0 & 1
- An influential point is an observation that affects the regression equation when included versus when it is not
- Calculator Tricks
  - Those included in formula sheet
    - LinRegTTest
  - Those not included in formula sheet
    - Scatter Plots
      - Make your scatter plot

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

        Zoom → 9: ZoomStat
    - LinReg(a+bx)

      Stat → Calc → 8: LinReg(a+bx) → Specify lists
- Formulas
  - Those included in formula sheet
    - Both models from above
    - Estimated Slope
    - Estimated Vertical Intercept
    - Coefficient of Correlation
    - Error
  - Those not included in formula sheet
    - Sum of Squares Error

      $SS\epsilon = \sum(\epsilon)^2$
    - Coefficient of Determination

      $R^2 = (R)^2$

- Standardized Residual Plot Point

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

- Hypotheses and Interpretations
  - Null states $\beta_1 = 0$ & the alternative states $\beta_1 \neq 0$
  - Coefficient of Correlation interpretation is a description of the relationship
    - Example: Strong Positive Linear Relationship
  - Coefficient of Determination interpretation:
    - [R^2]% of variation in y can be explained by x.
    - sufficient/insufficient statement

- Reading the graphs
  - We look at the lines and not the dots to see if there is an influential point
  - Can only tell if a trend is linear, non-linear, or has no relationship & whether it is positive or negative without more information (like R)
  - When there is a boundary, the dots outside that boundary are outliers because they lie beyond 3 standard deviations from the trend line