DATA 310 Formula Sheet

β_0: y-intercept of population regression line.

β_1: slope of population regression line. $\beta_1 = \frac{\mathrm{Cov}(x,y)}{\mathrm{Var}(x)}$.

$\hat{\beta}_0$: The estimated y-intercept of the regression line based on the sample data. $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

$\hat{\beta}_1$: The estimated slope of the regression line based on the sample data. $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$

$\bar{\beta}_0$ and $\bar{\beta}_1$: The population averages of β_0 & β_1 across different samples.

$E[\hat{\beta}_1] = \beta_1$, $Var[\hat{\beta}_1] = \frac{\sigma^2}{S_{xx}}$

$E[\hat{\beta}_0] = \beta_0$, $Var[\hat{\beta}_0] = \bar{X}^2 Var[\hat{\beta}_1]$

σ: The standard deviation of the error term, ϵ - this measures the spread of the errors around the regression line.

σ^2: The variance of the error term, representing the average squared deviation from the true regression line. $\sigma^2 = \frac{1}{n} \sum (y_i - \beta_0 - \beta_1 x_i)^2$

$\hat{\sigma^2}$: The estimated variance of the error term, calculated using sample data. $\hat{\sigma^2} = \frac{S_{yy}}{n-2} = \frac{\sum (y_i - \bar{y}_i)^2}{n-2}$

ϵ: The error term in the population model, the difference between real & predicted values. $\epsilon = y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$

We always assume ϵ are normally distributed with mean of 0 and variance of σ^2. All values of ϵ are independent to each other.

Simple linear regression model: $y_i = \beta_0 + \beta_1 x_i + \epsilon$

Fitted line: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

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

$E[y_i] = E[\hat{y}_i] = \beta_0 + \beta_1 x_i$, $Var[y_i] = \sigma^2$, $Var[\hat{y}_i] = 0$

$E[\bar{y}] = \beta_0 + \beta_1 \bar{x}$, $Var[\bar{y}] = \frac{\sigma^2}{n}$

Sum of squares total (SST) = $\sum (y_i - \bar{y})^2$

Sum of squares explained (SSE) = $\sum (y_i - \hat{y}_i)^2$

Sum of squares residual (SSR) = $\sum (\hat{y}_i - \bar{y})^2$

SST = SSE + SSR

$S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - n\bar{x}\bar{y}$

$S_{xx} = \sum (x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2$

$S_{yy} = \sum (y_i - \bar{y})^2 = \sum y_i^2 - n\bar{y}^2$

Sum of residuals: $\sum \epsilon_i = 0$

Sum of weighted residuals: $\sum x_i \epsilon_i = 0$

Sum of fitted value & residuals: $\sum \hat{y}_i \epsilon_i = 0$