Formula Sheet for Quiz 1

STAT 011

For a sample of data

If $\{x_1, x_2, \ldots, x_n\}$ is a data set of n observational units, we have the following:

Sample mean

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Sample variance

$$Var(x_1, \ldots, x_n) = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Sample standard deviation

$$sd(x_1, \ldots, x_n) = s = \sqrt{s^2}$$

If we want to standardize the data set X, to create a new standardized data set $Z = \{z_1, z_2, \ldots, z_n\}$ we preform

$$z_i = \frac{x_i - \bar{x}}{sd(x_1, \ldots, x_n)}, \text{ for } i = 1, \ldots, n.$$

Simple linear regression notation

The fitted/estimated regression model is $\hat{y}_i = b_0 + b_1 x_i$ where $b_0 = \bar{y} - b_1\bar{x}$ and $b_1 = \frac{s_{xy}}{\sqrt{s_x s_y}} \cdot \frac{s_y}{s_x}$.

$$\text{Residual} = e = y - \hat{y} = \text{observed value} - \text{predicted value}$$

Standard error of the residuals: $s_e = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-2}}$

Sum of squares terms

$$s_x = \sum_{i=1}^{n}(x_i - \bar{x})^2, \quad s_y = \sum_{i=1}^{n}(y_i - \bar{y})^2, \quad s_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Correlation coefficient

$$r = \frac{s_{xy}}{\sqrt{s_x s_y}}$$

Coefficient of determination/R-squared

$$R = \left(\frac{s_{xy}}{\sqrt{s_x s_y}}\right)^2$$