FORMULAS

Sample statistics

Sample mean: $\overline{x} = \frac{\sum x_i}{n}$, Sample standard deviation (definition) $s = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n-1}}$

Computational formula $s = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}}$ Range=Max-Min

Interquartile Range IQR= Q_3-Q_1 , Lower Limit LL=Q1-1.5IQR , Upper Limit UL=Q3+1.5IQR

Population parameters:

Population mean: $\mu = \frac{\sum x_i}{N}$ Population standard deviation: $\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$

or $\sigma = \sqrt{\frac{\sum x_i^2}{N} - \mu^2}$

Standard score or z-score $z = \frac{x - \mu}{\sigma}$

Regression and Correlation

Linear correlation of X and Y $r = \frac{\frac{1}{n-1} \sum (x_i - \overline{x})(y_i - \overline{y})}{s_x s_y} = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$

Least-Squares Regression equation $\hat{y} = b_0 + b_1 x$, $b_1 = \frac{S_{xy}}{S_{xx}}$, $b_0 = \overline{y} - b_1 \overline{x}$

Coefficient of determination: $r^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$

$SST = \sum (y_i - \overline{y})^2 = S_{yy}$ $SSR = \sum (\hat{y}_i - \overline{y})^2 = \frac{{S_{xy}}^2}{S_{xx}}$ $SSE = \sum (y_i - \hat{y}_i)^2 = S_{yy} - \frac{{S_{xy}}^2}{S_{xx}}$

Regression Identity: $SST = SSR + SSE$

$S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$, $S_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$, $S_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$