STATISTICS FORMULAS

DESCRIPTIVE STATISTICS:

MEAN: \( \bar{x} = \frac{1}{n} \sum x_i \)

VARIANCE: \( s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 \)

STANDARD DEVIATION: \( s = \sqrt{s^2} \)

STANDARD ERROR: \( SE_{\bar{x}} = \frac{s}{\sqrt{n}} \)

Z-SCORE: \( Z = \frac{x - \mu}{\sigma} \)

REGRESSION LINES:

For a data set \((x_i, y_i), \) where \((\bar{x}, \bar{y})\) are the centroids (means) of the data set, and \(r\) is the correlation coefficient:

LEAST-SQUARES REGRESSION LINE: \( \hat{y} = b_0 + b_1 x \)

RESIDUALS: \( e_i = y_i - \hat{y} \)

\( \sum (y_i - \hat{y}_i)^2 \)

SSM = \( \frac{SSM}{SST} = SSM + SSE \)

COEFFICIENT OF DETERMINATION: \( r^2 = \frac{SSM}{SST} \)

CORRELATION COEFFICIENT: \( r = \sqrt{r^2} \)

SLOPE: \( b_1 = r \frac{s_x}{s_y} \)

INTERCEPT: \( b_0 = \bar{y} - b_1 \bar{x} \)

VARIANCE: \( MSE = s^2 = \frac{s^2}{n-2} \)

ST DEV: \( s = \sqrt{s^2} \)

STANDARD ERROR b1: \( SE_{b1} = \frac{s}{\sqrt{\Sigma(x_i-\bar{x})^2}} \)

STANDARD ERROR b0: \( SE_{b0} = s \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\Sigma(x_i-\bar{x})^2}} \)

CONFIDENCE LEVEL FOR THE INTERCEPT \( \beta_0: \) \( b_0 \pm t \cdot SE_{b0} \)

CONFIDENCE LEVEL FOR THE SLOPE: \( \beta_1: \) \( b_1 \pm t \cdot SE_{b1} \)

PREDICTION INTERVAL: \( \hat{y} \pm t \cdot SE \)

HYPOTHESIS TESTING – MEANS:

STANDARD ERROR: \( \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \)

MARGIN OF ERROR: \( m = z \cdot \frac{\sigma}{\sqrt{n}} \) or \( m = t \cdot \frac{\sigma}{\sqrt{n}} \)

CONFIDENCE INTERVAL: \( C.I. = \bar{x} \pm m \)

SAMPLE SIZE FOR A GIVEN \( m: \) \( n = \left( \frac{z \cdot \sigma}{m} \right)^2 \)

ONE SAMPLE Z-TEST: \( z = \frac{x - \mu_0}{\sigma} \)

T-TEST: \( t = \frac{x - \mu_0}{\frac{s}{\sqrt{n}}} \)

TWO SAMPLE Z-TEST: \( z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \)

TWO SAMPLE T-TEST: \( t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \)

PROPORTION: \( \hat{p} = \frac{X}{n} \) where \( X = \) number of successes

STANDARD ERROR: \( SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \)

MARGIN OF ERROR, 2-SAMP PROP: \( m = z \cdot SE_{\hat{p}} \)

Z-TEST, ONE-SAMPLE PROPORTION: \( z = \frac{\hat{p} - p_0}{\frac{SE_{\hat{p}}}{n}} \)

STD ERR, 2-SAMP PROP: \( SE_{\hat{p}} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \)

MARGIN OF ERR, 2-SAMP PROP: \( m = z \cdot SE_{\hat{p}} \)

PLUS FOUR PROPORTIONS: \( \hat{p}_i = \frac{X_i + 1}{n_i + 2} \)

EST DIFF BTWN PROPS: \( \bar{\hat{p}} = \frac{\hat{p}_1 + \hat{p}_2}{n_1 + n_2} \)

STD DEV: \( \sigma_{\bar{\hat{p}}} = \sqrt{\frac{p_1(1-p_1)}{n_1 + 2} + \frac{p_2(1-p_2)}{n_2 + 2}} \)

POOLED PROPORTION: \( \hat{p} = \frac{X_1 + X_2}{n_1 + n_2} \)

POOLED STD ERR: \( SE_{\hat{p}} = \sqrt{\left( \frac{\hat{p}(1-\hat{p})}{n_1} \right) + \left( \frac{1}{n_2} \right)} \)

TWO SAMPLE Z-SCORE: \( z = \frac{\hat{p}_1 + \hat{p}_2}{SE_{\hat{p}}} \)

DESCRIPTIONS OF STATISTICS FORMULAS

MEAN: The mean, symbolized by x-bar, equals one divided by the number of samples multiplied by the sum of all data points, symbolized by x-sub-i.

VARIANCE: Variance, symbolized by s squared, equals 1 divided by the number of samples minus one, multiplied by the sum of each data point subtracted by the mean then squared.

STANDARD DEVIATION: Standard deviation, symbolized by s, equals the square root of the variance s-squared.

STANDARD ERROR: The standard error of the mean equals the standard deviation divided by the square root of the number of samples.

Z-SCORE: Z equals the test data minus the population mean, then divided by the population standard deviation.

REGRESSION LINES:

LEAST-SQUARES REGRESSION LINE: The predicted value, symbolized by y-hat, equals the intercept, symbolized by b-sub-o, plus the slope, symbolized by b-sub-1, times the data point x.

RESIDUALS: The residual, symbolized by e-sub-i, equals the data point y, symbolized by y-sub-i, minus the predicted value from the least-squares regression line, symbolized y-hat.

SSM, SSE, SST: Sum of square means equals the sum of the centroid, symbolized by y-bar, minus the predicted value of each x data point, symbolized by y-hat sub I. Sum of square errors equal the sum of each y data point, symbolized by y-sub-I, minus the predicted value of each data point, symbolized by y-hat-sub-I, then squared. The Sum of Square Total = Sum of Square Means plus Sum of Square Errors.

COEFFICIENT OF DETERMINATION: The coefficient of determination, symbolized r-squared, equals the sum of square means divided by the sum of squares total.

CORRELATION COEFFICIENT: The correlation coefficient r equals the square root of the coefficient of determination, symbolized by r-squared.

SLOPE: Slope, symbolized b-sub-one, equals the correlation coefficient r multiplied by the ratio of the standard deviation of the x data points to the standard deviation of the y data points.

INTERCEPT: Intercept, symbolized by b-sub-zero, equals the mean of the y data points, symbolized by y-bar, minus the slope, symbolized by b-sub-one multiplied by the mean of the x data points, symbolized by x-bar.

VARIANCE: Mean of Square Errors, symbolized s-squared or MSE, is equal to the sum of the residuals, symbolized by e-sub-I, squared then divided by the number of data points subtracted by two. STANDARD DEVIATION, symbolized by s, equals the square root of variance.

STANDARD ERROR: The standard error of the slope, symbolized by SE-sub-b1, equals the standard deviation, symbolized by s, divided by the square root of the sum of each data point, symbolized by x-sub-I, subtracted from the mean of all x data points, symbolized by s-bar, then squared.

The STANDARD ERROR of the intercept, symbolized by SE-sub-bo, equals the standard deviation, symbolized by s, multiplied by the square root of one divided by the number of data points plus the mean of all x’s squared, symbolized by x-bar squared, divided by the sum of all x data points, symbolized by x-sub-I minus the mean of all x data points, symbolized by x-bar, squared.
CONFIDENCE LEVEL FOR THE INTERCEPT: The confidence level for the intercept, symbolized $\beta_{-0}$, equals the sample intercept, symbolized by $b_{-0}$, plus or minus the $t$-score for the interval, symbolized by $t$, multiplied by the standard error of the intercept.

CONFIDENCE LEVEL FOR THE SLOPE: The confidence level for the slope, symbolized by $\beta_{-1}$, equals the sample slope, symbolized by $b_{-1}$, plus or minus the $t$-score for the interval, symbolized by $t$, multiplied by the standard error of the slope.

PREDICTION INTERVAL: The prediction interval equals the predicted value of $y$, symbolized by $y$-hat, plus or minus the $t$-score for the interval, symbolized by $t$, multiplied by the standard error.