STATISTICS FORMULAS

Campus Academic Resource Program

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: $\hat{y} = b_o + b_1 x$

RESIDUALS: $e_i = y_i - \hat{y}$

SSM $= \sum(\bar{y} - \hat{y}_i)^2$ SSE $= \sum(y_i - \hat{y}_i)^2$ 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: $MSE = s^2 = \frac{\sum e_i^2}{n-2}$ ST DEV: $s = \sqrt{s^2}$

STANDARD ERROR b_1: $SE_{b1} = \frac{s}{\sqrt{\sum(X_I - \bar{x})^2}}$

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

CONFIDENCE LEVEL FOR THE INTERCEPT β_0: $b_0 \pm t^*SE_{b0}$

CONFIDENCE LEVEL FOR THE SLOPE: β_1: $b_1 \pm t^*SE_{b1}$

PREDICTION INTERVAL: $\hat{y} \pm t * SE$

HYPOTHESIS TESTING – MEANS:

STANDARD ERROR: $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

MARGIN OF ERROR: m = $z * \frac{\sigma}{\sqrt{n}}$ or m = $t * \frac{\sigma}{\sqrt{n}}$

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

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

ONE SAMPLE Z-TEST: $z = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$ T-TEST: $t = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}$

TWO SAMPLE Z-TEST: $z = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$

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: m = $z * SE_{\hat{p}}$

Z-TEST, ONE-SAMPLE PROPORTION: $z = \frac{\hat{p}-p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}}$

STD ERR, 2-SAMP PROP: $SE_D = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

MARGIN OF ERR, 2-SAMP PROP: m = $z * SE_D$

PLUS FOUR PROPORTIONS: $\tilde{p}_i = \frac{X_i+1}{n_i+2}$

EST DIFF BTWN PROPS: $\tilde{D} = \tilde{p}_1 - \tilde{p}_2$

STD DEV: $\sigma_{\tilde{D}} = \sqrt{\frac{P_1(1-p_1)}{n_1+2} + \frac{p_2(1-p_2)}{n_2+2}}$

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

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

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

Reference: Moore DS, McCabe GP & Craig BA. *Introduction to the Basic Practice of Statistics*. New York: W.H. Freeman & Co, 5th edition.

DESCRIPTIONS OF STATISTICS FORMULAS

Campus Academic Resource Program

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 centriod, 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.

Reference: Moore DS, McCabe GP & Craig BA. *Introduction to the Basic Practice of Statistics*. Ne w York: W.H. Freeman & Co, 5th edition.

CONFIDENCE LEVEL FOR THE INTERCEPT: The confidence level for the intercept, symbolized beta-sub-zero, equals the sample intercept, symbolized by b-sub-zero, 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-sub-one, equals the sample slope, symbolized by b-sub-one, 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.

:

Reference: Moore DS, McCabe GP & Craig BA. *Introduction to the Basic Practice of Statistics*. Ne w York: W.H. Freeman & Co, 5th edition.