Formulae for Basic Statistics

$$\overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$$

$$s = \sqrt{\frac{\Sigma(Y_i - \overline{Y})^2}{n-1}} \quad s = \sqrt{\frac{\Sigma\left(Y_i^2\right) - n\overline{Y}^2}{n-1}}$$

Standard error of the mean

$s / \sqrt{n}$

χ^2 test of goodness-of-fit

$$\chi^2 = \sum_i \frac{\left(O_i - E_i\right)^2}{E_i}$$

$$\ln(\hat{O}R) - Z_\alpha \; SE[\ln(\hat{O}R)] \le \ln(OR) \le \ln(\hat{O}R) + Z_\alpha \; SE[\ln(\hat{O}R)]$$

$$\hat{O}R = \frac{ad}{bc}$$

Poisson Probability Distribution

$$P[x] = \frac{\lambda^x e^{-\lambda}}{x!}$$

Binomial Probability Distribution

$$P[x] = \binom{N}{x} p^x (1-p)^{N-x}$$

Normal Probability Distribution

$$P[x] = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Confidence Interval for the mean of a normal distribution

$$\overline{Y} \pm SE_{\overline{Y}} \quad t_{\alpha(2),df}$$

Confidence Interval for the variance of a normal distribution

$$\frac{df\ s^2}{\chi^2_{\frac{\alpha}{2},df}} \le \sigma^2 \le \frac{df\ s^2}{\chi^2_{1-\frac{\alpha}{2},df}}$$

Pooled variance

$$s_p{}^2 = \frac{df_1 s_1^2 + df_2 s_2^2}{df_1 + df_2}$$

$$F = \frac{s_1^2}{s_2^2} \quad or \quad \frac{s_2^2}{s_1^2}$$

$$t = \frac{\left(\overline{Y}_1 - \overline{Y}_2\right)}{SE_{\overline{Y}_1 - \overline{Y}_2}}$$

$$\left(\overline{Y}_1 - \overline{Y}_2\right) \pm SE_{\overline{Y}_1 - \overline{Y}_2} \; t_{\alpha(2),df}$$

$$t = \frac{\overline{X} - \mu}{s / \sqrt{n}}$$

$$Z = \frac{x - \mu}{\sigma}$$

$$SE_{\overline{Y}_1 - \overline{Y}_2} = \sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$t = \frac{\overline{Y}_1 - \overline{Y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(s_2^2/n_2\right)^2}{n_2 - 1}\right)}$$

Mann-Whitney U

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

$$U' = n_1 n_2 - U$$

$$Z = \frac{2U - n_1 n_2}{\sqrt{n_1 n_2 (n_1 + n_2 + 1)/3}}$$

Formulae for regression and correlation

$$\sum(X - \overline{X})(Y - \overline{Y}) = \sum(XY) - \frac{\left(\sum X\right)\left(\sum Y\right)}{n}$$

$$b = \frac{\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum\left(X_i - \overline{X}\right)^2}$$

$$a = \overline{Y} - b\overline{X}$$

$$SS_{Total} = \sum Y_i^2 - \frac{\left(\sum Y_i\right)^2}{n}$$

$$SS_{regression} = b\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)$$

$$SS_{residual} + SS_{regression} = SS_{Total}$$

$$MS_x = \frac{SS_x}{DF_x}$$

$$r^2 = \frac{SS_{regression}}{SS_{Total}}$$

$$SE_b = \sqrt{\frac{MS_{residual}}{\sum\left(X_i - \overline{X}\right)^2}}$$

$$MS_{residual} = \frac{\sum\left(Y_i - \overline{Y}\right)^2 - b\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{n-2}$$

$$b \pm t_{\alpha[2],\nu} SE_b$$

$$\hat{Y} \pm t_{\alpha[2],\nu} SE_{\hat{Y}}$$

$$t = \frac{b - \beta_0}{SE_b}$$

$$t = \frac{\left(b_1 - b_2\right) - \left(\beta_1 - \beta_2\right)}{SE_{b1-b2}}$$

$$\left(MS_{error}\right)_p = \frac{\left(SS_{error}\right)_1 + \left(SS_{error}\right)_2}{\left(DF_{error}\right)_1 + \left(DF_{error}\right)_2}$$

$$SE_{b1-b2} = \sqrt{\frac{\left(MS_{error}\right)_p}{\left(\sum\left(X - \overline{X}\right)^2\right)_1} + \frac{\left(MS_{error}\right)_p}{\left(\sum\left(X - \overline{X}\right)^2\right)_2}}$$

$$r = \frac{\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum\left(X_i - \overline{X}\right)^2 \sum\left(Y_i - \overline{Y}\right)^2}}$$

$$SE_r = \sqrt{\frac{1 - r^2}{n-2}}$$

$$z = 0.5\ln\left(\frac{1+r}{1-r}\right)$$

$$\sigma_z = \sqrt{\frac{1}{n-3}}$$

$$r_s = 1 - \frac{6\sum d_i^2}{n^3 - n}$$

ANOVA etc.

$$F = \frac{MS_{\text{groups}}}{MS_{\text{error}}}$$

$$MS_{\text{error}} = s^2_{\text{pooled}} = \frac{\sum s_i^2 (n_i - 1)}{N - k}$$

$$MS_{\text{groups}} = \frac{\sum n_i (\overline{Y}_i - \overline{Y})^2}{k-1}$$

$$\overline{Y} = \frac{\sum n_i (\overline{Y}_i)}{N}$$

$$R^2 = \frac{SS_{\text{groups}}}{SS_{\text{total}}}$$

Kruskal-Wallis

$$H = \frac{12}{N(N+1)}\left[\sum \frac{R_i^2}{n_i}\right] - 3(N+1)$$

Tukey:

$$q = \frac{\overline{Y}_i - \overline{Y}_j}{SE} \qquad SE = \sqrt{s^2_{\text{pooled}}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}.$$

Adjusted Wald method:

$$p' = \frac{X+2}{n+4}.$$

$$\left(p' - Z\sqrt{\frac{p'(1-p')}{n+4}}\right) \le p \le \left(p' + Z\sqrt{\frac{p'(1-p')}{n+4}}\right)$$