SOME USEFUL FORMULAS

Univariate Data

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad and \qquad s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$\bar{x} \pm t_{\alpha/2;n-1} \times (s_x/\sqrt{n}) \qquad and \qquad t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$$

$$\text{ttail}(df, t) = \Pr[T > t] \text{ where } T \sim t(df)$$

$t_{\alpha/2}$ such that $\Pr[|T| > t_{\alpha/2}] = \alpha$ is calculated using $\text{invttail}(df, \alpha/2)$.

Bivariate Data

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \times \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{s_{xy}}{\sqrt{s_{xx} \times s_{yy}}} \quad [\text{Here } s_{xx} = s_x^2 \text{ and } s_{yy} = s_y^2].$$

$$\hat{y} = b_1 + b_2 x_i \qquad b_2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad b_1 = \bar{y} - b_2\bar{x}$$

$$\text{TSS} = \sum_{i=1}^{n}(y_i - \bar{y}_i)^2 \qquad \text{ErrorSS} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad \text{RegSS} = \text{TSS - ErrorSS}$$

$$R^2 = 1 - \text{ErrorSS/TSS}$$

$$b_2 \pm t_{\alpha/2;n-2} \times s_{b_2}$$

$$t = \frac{b_2 - \beta_{20}}{s_{b_2}} \qquad s_{b_2}^2 = \frac{s_e^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad s_e^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$y|x = x^* \in b_1 + b_2 x^* \pm t_{\alpha/2;n-2} \times s_e \times \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} + 1}$$

$$\text{E}[y|x = x^*] \in b_1 + b_2 x^* \pm t_{\alpha/2;n-2} \times s_e \times \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}}$$

Multivariate Data

$$\hat{y} = b_1 + b_2 x_{2i} + \cdots + b_k x_{ki}$$

$$R^2 = 1 - \text{ErrorSS/TSS} \qquad \bar{R}^2 = R^2 - \frac{k-1}{n-k}(1 - R^2)$$

$$b_j \pm t_{\alpha/2;n-k} \times s_{bj} \qquad and \qquad t = \frac{b_j - \beta_{j0}}{s_{b_j}}$$

$$F = \frac{R^2/(k-1)}{(1 - R^2)/(n - k)} \qquad and \qquad F = \frac{(SSE_r - SSE_u)/(k - g)}{SSE_u/(n - k)}$$

$$\text{Ftail}(df1, df2, f) = \Pr[F > f] \text{ where } F \text{ is } F(df1, df2) \text{ distributed.}$$

F_α such that $\Pr[F > f_\alpha] = \alpha$ is calculated using $\text{invFtail}(df1, df2, \alpha)$.