Important Formulas for Statistical Inference

<u>Population</u> $z = \frac{x-\mu}{\sigma}$

<u>One Sample</u>

$H_0 : \mu = \mu_0$ $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$

$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$

$H_0 : p = p_0$ $z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0 q_0}{n}}}$

<u>Two Samples</u>

$H_0 : \mu_1 - \mu_2 = \mu_0$ $z = \frac{(\bar{x}_1-\bar{x}_2)-\mu_0}{\sqrt{\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}}}$

$t = \frac{(\bar{x}_1-\bar{x}_2)-\mu_0}{\sqrt{\frac{s_p{}^2}{n_1}+\frac{s_p{}^2}{n_2}}}$

where $s_p{}^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$

$H_0 : \mu_d = \mu_{d_0}$ $t = \frac{\bar{d}-\mu_{d_0}}{s_d/\sqrt{n}}$

$H_0 : p_1 - p_2 = 0$ $z = \frac{(\hat{p}_1-\hat{p}_2)-0}{\sqrt{\frac{\overline{p}\overline{q}}{n_1}+\frac{\overline{p}\overline{q}}{n_2}}}$ where $\overline{p} = \frac{x_1+x_2}{n_1+n_2}$

a	*b*	r_1
c	*d*	r_2
c_1	c_2	*N*

$\chi_1^2 = \frac{N(ad-bc)^2}{r_1 \cdot r_2 \cdot c_1 \cdot c_2}$

$\chi_1^2 = \frac{(b-c)^2}{b+c}$

$\hat{\text{RR}} = \hat{p}_1 / \hat{p}_2$

$\text{SE } \log_e \hat{RR} = \sqrt{\frac{\hat{q}_1}{n_1\hat{p}_1}+\frac{\hat{q}_2}{n_2\hat{p}_2}}$

$\hat{\text{OR}} = \frac{\hat{p}_1/\hat{q}_1}{\hat{p}_2/\hat{q}_2}$

$\text{SE } \log_e \hat{\text{OR}} = \sqrt{\frac{1}{n_1\hat{p}_1}+\frac{1}{n_1\hat{q}_1}+\frac{1}{n_2\hat{p}_2}+\frac{1}{n_2\hat{q}_2}}$

$t = \frac{(\bar{x}_1-\bar{x}_2)-\mu_0}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$

$\hat{\text{OR}} = b/c$

$\text{SE } \log_e \hat{\text{OR}} = \sqrt{\frac{1}{b}+\frac{1}{c}}$

$n = \frac{\left[z_{\alpha/2}\sqrt{2\overline{p}\overline{q}}+z_\beta\sqrt{p_1q_1+p_2q_2}\right]^2}{\Delta^2}$

$n = \frac{\left(z_{\alpha/2}+z_\beta\right)^2\left(\sigma_1^2+\sigma_2^2\right)}{\Delta^2}$

One-Way ANOVA

$H_0 : \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_p$

Source	SS	df
Between Groups	$\sum_j \sum_i \left(\bar{y}_{.j} - \bar{y}_{..}\right)^2 = \sum_j n_j \left(\bar{y}_{.j} - \bar{y}_{..}\right)^2 = \sum_j \frac{T^2_{.j}}{n_j} - \frac{T^2_{..}}{N}$	$p-1$
Within Groups	$\sum_j \sum_i \left(y_{ij} - \bar{y}_{.j}\right)^2 = \sum_j \sum_i y_{ij}^2 - \sum_j \frac{(T_{.j})^2}{n_j} = \sum_j \left(n_j - 1\right) s_j^2$	$N-p$
Total	$\sum_j \sum_i \left(y_{ij} - \bar{y}_{..}\right)^2$ $= \sum_j \sum_i y_{ij}^2 - N(\bar{y}_{..})^2 = \sum_j \sum_i y_{ij}^2 - \frac{T^2_{..}}{N}$	$N-1$

Bonferroni $\alpha^* = \frac{\alpha}{\binom{p}{2}}$

Bonferroni $t = \frac{\bar{y}_{.i} - \bar{y}_{.j}}{\sqrt{\frac{\text{MSW}}{n_i} + \frac{\text{MSW}}{n_j}}}$

Association

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_1 - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} = \frac{(\sum_i x_i y_i) - n\bar{x}\bar{y}}{\sqrt{\left((\sum_i x_i^2) - n\bar{x}^2\right)\left((\sum_i y_i^2) - n\bar{y}^2\right)}}$$

Simple Linear Regression

$\hat{y}_i = b_0 + b_1 x_1$ where $b_0 = \bar{y} - b_1 \bar{x}$

$$b_1 = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} = \frac{(\sum_i x_i y_i) - n\bar{x}\bar{y}}{(\sum_i x_i^2) - n\bar{x}^2}$$

$$\begin{bmatrix} z_{0.01}(2-\text{sided}) = 2.58 \\ z_{0.01}(1\text{-sided}) = 2.33 \\ z_{0.05}(2-\text{sided}) = 1.96 \\ z_{0.05}(1-\text{sided}) = 1.645 \\ z_{0.20}(1-\text{sided}) = 0.84 \\ z_{0.10}(1-\text{sided}) = 1.28 \end{bmatrix}$$

ANOVA for Regression

Source	SS	df
Regression	$\sum (\hat{y}_i - \bar{y})^2$	1
Residual	$\sum \left(y_i - \hat{y}_i\right)^2$	$n-2$
Total	$\sum (y_i - \bar{y})^2$ $= \sum y_i^2 - n\bar{y}^2$	$n-1$

$r^2 = \frac{\text{SS Regression}}{\text{SST}}$