Public Policy 529 Formula Sheet

Descriptive and Distributional Statistics

$$\bar{y} = \frac{\sum y_i}{n}$$

$$s^2 = \frac{\sum (y_i - \bar{y})^2}{n-1}$$

$$Z = \frac{y - \mu_y}{\sigma}$$

$$IQR = Q_3 - Q_1$$

$$SS = \sum (y_i - \bar{y})^2$$

$$s = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}}$$

$$\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}$$

Probability

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)}$$

$$P(A \text{ and } B) = P(A) \times P(B|A)$$

$$P(\sim A) = 1 - P(A)$$

$$P(A \text{ and } B) = P(A) \times P(B)$$

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

$$P(x) = \frac{n!}{x!(n-x)!} \pi^x (1-\pi)^{n-x}$$

Confidence Intervals and Significance Tests

$$t = \frac{\bar{y} - \mu_0}{\hat{\sigma}_{\bar{y}}}$$

$$\hat{\sigma}_{\hat{\pi}} = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$$

$$Z = \frac{\hat{\pi} - \pi_0}{\hat{\sigma}_{\pi_0}}$$

$$\text{c.i.} = \bar{y} \pm t \cdot \hat{\sigma}_{\bar{y}}$$

$$\hat{\sigma}_{\bar{y}} = \frac{s}{\sqrt{n}}$$

$$\text{c.i.} = \bar{y} \pm Z \cdot \hat{\sigma}_{\bar{y}}$$

$$\hat{\sigma}_{\pi_0} = \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

$$\text{c.i.} = \hat{\pi} \pm Z \cdot \hat{\sigma}_{\hat{\pi}}$$

$$se_{\text{diff}} = \sqrt{(se_1)^2 + (se_2)^2}$$

$$z = \frac{(\hat{\pi}_2 - \hat{\pi}_1) - H_0}{se_0}, \text{ where } se_0 = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n_1} + \frac{\hat{\pi}(1-\hat{\pi})}{n_2}} \qquad c.i. = (\hat{\pi}_2 - \hat{\pi}_1) \pm z\sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}$$

$$\hat{\pi} = \frac{\hat{\pi}_1 n_1 + \hat{\pi}_2 n_2}{n_1 + n_2}$$

$$t = \frac{(\bar{y}_2 - \bar{y}_1) - H_0}{se_{\text{diff}}} \qquad ci = (\bar{y}_2 - \bar{y}_1) \pm t \cdot se_{\text{diff}}$$

Unequal Variance:

$$se_{\text{diff}} = \hat{\sigma}_{\bar{y}_1 - \bar{y}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \qquad df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} \qquad \text{approx. } df = \min(n_1 - 1, n_2 - 1)$$

Equal Variance:

$$se_{\text{diff}} = \hat{\sigma}_{\bar{y}_1 - \bar{y}_2} = \sqrt{\frac{s_{pooled}^2}{n_1} + \frac{s_{pooled}^2}{n_2}} = s_{pooled}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \qquad s_{pooled} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$df = n_1 + n_2 - 2$$

$$t = \frac{\bar{y}_d - H_0}{\hat{\sigma}_{\bar{y}_d}} \qquad \hat{\sigma}_{\bar{y}_d} = \frac{s_d}{\sqrt{n}}, \text{ where } s_d = \sqrt{\frac{\sum(y_{di} - \bar{y}_d)^2}{n-1}}$$

Measures of Association

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}}, \quad \text{se}_r = \sqrt{\frac{1 - r^2}{n - 2}}$$

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}, \text{ where } f_e = \frac{\text{(row total)(column total)}}{n}$$