Formula Sheet — Math 222

Probability rules

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

Discrete expected value & variance

$$E(X) = \mu = \sum p_k x_k$$
$$\mathrm{Var}(X) = \sigma^2 = \sum p_k (x_k - \mu)^2$$

Expected value & variance rules

$$E(X+Y) = E(X) + E(Y)$$
$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)^*$$
$$E(cX) = cE(X)$$
$$\mathrm{Var}(cX) = c^2\,\mathrm{Var}(X)$$

* *Only if X and Y are independent.*

General form of a confidence interval

$$\text{estimate} \pm \text{margin of error, where}$$
$$\text{margin of error} = (\text{critical value}) \times SE_{\text{estimate}}$$

General form of a test statistic

$$\frac{\text{statistic} - \text{hypothesized valued}}{SE_{\text{statistic}}}$$

Least squares regression line

$$\hat{y} = b_0 + b_1 x, \text{ where } b_1 = r\frac{s_y}{s_x} \text{ and } b_0 = \bar{y} - b_1\bar{x}$$

Chi-squared formulas

$$\chi^2 = \sum \frac{(\textit{observed count} - \textit{expected count})^2}{\textit{expected count}}$$

Expected counts for 2-way tables

$$E_{ij} = \frac{\textit{row total} \times \textit{column total}}{\textit{table total}}$$

Standard errors*

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}} \qquad SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$
$$SE_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \qquad SE_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

* *When testing a null hypothesis about proportions, it is better to use the hypothesized population proportion p_0 rather than $\hat{p}$ in the formula for standard error for one-sample tests, and it is better to used the pooled proportion $\hat{p}$ rather than either individual sample proportion $\hat{p}_1$ or $\hat{p}_2$ in the formula for standard error in two-sample tests.*

Inference about regression

$$SE_{b_1} = \frac{s}{s_x\sqrt{n-1}} = \frac{r\sqrt{n-2}}{1-r^2}$$
$$SE_{\mu_y} = s\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_x^2(n-1)}} = \sqrt{\frac{s^2}{n} + (x^* - \bar{x})^2 SE_{b_1}^2}.$$
$$SE_{\hat{y}} = s\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_x^2(n-1)}} = \sqrt{s^2 + \frac{s^2}{n} + (x^* - \bar{x})^2 SE_{b_1}^2}.$$

where $s = \sqrt{MSE}$ is the standard error of the residuals. Use $dF = n - 2$ for simple linear regression. Recall that $r^2 = \frac{SSM}{SST}$ and $MST = s_y^2$. With that you can fill in the whole ANOVA table.

Inference in ANOVA Use the pooled standard deviation $s_p = \sqrt{MSE}$ as the best guess for σ. Recall also that $SSG = \sum_{i=1}^{I} N_i(\bar{x}_i - \bar{x})^2$ and $SSE = \sum_{i=1}^{I}(N_i - 1)s_i^2$.