STAT 327 Formula Sheet

(to be distributed with the Final Exam)

Linear Regression:

$$Y_i = \alpha + \beta x_i + \varepsilon_i \,, \quad \varepsilon_i \sim N(0, \sigma^2)$$

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - x)^2} = \frac{\sum x_i y_i - (\sum x_i)(\sum y_i)/n}{\sum x_i^2 - (\sum x_i)^2/n}$$

$$a = \bar{y} - b\bar{x}$$

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2}\sqrt{\sum (y_i - \bar{y})^2}}$$

Estimator of σ^2: $\quad s^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n-3}$, where $\hat{y}_i = a + bx_i$, $i = 1, 2, \ldots, n$.

standard error of b: $\quad SE(b) = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}}$

$$\frac{b - \beta}{SE(b)} \sim t_{n-2}$$

(continued)

Inference about μ

Small n, Normal errors, then $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$

Large n, unknown errors, then $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ is approx. $\sim N(0,1)$

Inference about p

Large n, $p=p_0$, then $\frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ is approx $\sim N(0,1)$.

Large n, p unknown, then $\frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})/n}}$ is approx. $\sim N(0,1)$

Inference about $\mu_1-\mu_2$

Small n_1 and n_2, Normal errors with common variance,

$$\text{then } \frac{\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2},$$

$$\text{where } S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}.$$

Large n_1 and n_2, error distributions unknown (and perhaps different),

$$\text{then } \frac{\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \text{ is approx } \sim N(0,1).$$

Inference about $p_1 - p_2$

Large n_1, n_2 with p_1 and p_2 unknown,

then $$\frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$ is approx $\sim N(0,1)$.

Large n_1, n_2, p_1 and p_2 unknown but $p_1 = p_2$,

then $$\frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$ is approx $\sim N(0,1)$,

where $$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}.$$

ANOVA

$$X_{ij} = \mu_i + \varepsilon_{ij}, \quad i = 1, 2, \ldots, k$$
$$j = 1, 2, \ldots, n_i$$
$$\varepsilon_{ij} \overset{indep}{\sim} N(0, \sigma^2)$$

$$SST = SSW + SSB$$

Under $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$,

$$F = \frac{SSB/(n-k)}{SSW/(k-1)} \sim F_{k-1, n-k}, \quad \text{where } n = \sum_{i=1}^{k} n_i$$

(continued)

ANOVA (continued): Computation of Sums of Squares

Let $T_i = \sum_{j=1}^{n_i} X_{ij}$, $i=1,\ldots,k$

and $T = \sum_{i=1}^{k} T_i = \sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij}$.

Then $$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij}^2 - \frac{T^2}{n},$$

$$SSB = \sum_{i=1}^{k} \frac{T_i^2}{n_i} - \frac{T^2}{n},$$

and $SSW = SST - SSB$.