

**Formula Sheet for Final Exam  
POS 3713 - Spring 2007 (Dr. Hensel)**

**Chi-Square**

$$X^2 \text{ (obtained)} = \sum \frac{(f_o - f_e)^2}{f_e}$$

(where  $f_o$  = observed cell frequencies,  $f_e$  = frequencies expected if the variables were indep.)

$$f_e = (\text{Row marginal} * \text{Column marginal}) / N$$

$$df = (r - 1)(c - 1) \quad (\text{where } r = \# \text{ of rows in table, } c = \# \text{ of columns})$$

**Phi**

$$\text{Phi} = \sqrt{\frac{X^2}{N}} \quad (\text{where } X^2 \text{ is calculated as above, and } N \text{ is the } \# \text{ of cases in the table})$$

**Cramer's V**

$$V = \sqrt{\frac{X^2}{N * \min(r - 1, c - 1)}}$$

(where  $X^2$  is calculated as above,  $N$  = # cases,  $r$  = # rows in table,  $c$  = #columns in table)

**Lambda**

$$\lambda = \frac{E_1 - E_2}{E_1}$$

(where  $E_1$  = errors without knowing  $X$  = based on [N-row marginal] for each row;  
 $E_2$  = errors knowing  $X$  = based on [column marginal – largest cell] for each column)

**Gamma**

$$\text{Gamma} = \frac{N_s - N_d}{N_s + N_d}$$

(where  $N_s$  = # same order pairs in data set [cell a\*cell d in following example],  
 $N_d$  = # inverse order pairs [cell b\*cell c])

|      |        |        |
|------|--------|--------|
|      | Low    | High   |
| Low  | Cell a | Cell b |
| High | Cell c | Cell d |

## ANOVA

$$SST = \sum (X_i^2) - [N * (\bar{X})^2]$$

$$SSB = \sum [N_k * (\bar{X}_k - \bar{X})^2]$$

$$SSW = SST - SSB$$

$$dfb = k-1, dfw = N-k$$

(where  $\sum (X_i^2)$  = sum of all squared scores in the data set,  $k$ =# categories,  $N$ =sample size,  $\bar{X}$ =sample mean)

$$F(\text{obtained}) = \frac{MSB}{MSW} = \frac{[SSB/(k-1)]}{[SSW/(N-k)]}$$

## Pearson's r (Correlation)

$$\text{Covariance}(X,Y) = S_{YX} = \sum (X - \bar{X})(Y - \bar{Y}) = N \sum (XY) - \sum X \sum Y$$

$$r = \frac{\text{Covariance}(X, Y)}{\text{S.D.}(X) * \text{S.D.}(Y)} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2} * \sqrt{\sum (Y - \bar{Y})^2}} = \frac{N * \sum (XY) - \sum (X) \sum (Y)}{\sqrt{N \sum (X^2) - (\sum X)^2} * \sqrt{N \sum (Y^2) - (\sum Y)^2}}$$

## Regression

### Basic Regression Formula:

$$\text{Bivariate Regression: } Y = a + bX + e$$

$$\text{Multiple Regression: } Y = a + b_1X_1 + b_2X_2 + \dots + e$$

### Slope coefficient (b) and Intercept (a):

$$b = \frac{\text{Covariance}(X, Y)}{\text{Variance}(X)} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2} = \frac{N * \sum (XY) - \sum (X) \sum (Y)}{N \sum (X^2) - (\sum X)^2}$$

$$a = \bar{Y} - b(\bar{X})$$

(where  $\bar{Y}$  = mean of Y,  $\bar{X}$  = mean of X)

### Coefficient of Determination:

$$r^2/R^2 = \frac{\text{Explained Variation}}{\text{Total Variation}} = \frac{SSR}{SST} = \frac{\sum (\hat{Y} - \bar{Y})^2}{\sum (Y - \bar{Y})^2} = \frac{\text{Covariance}(X, Y)^2}{\text{Variance}(X) * \text{Variance}(Y)}$$

(Where  $\hat{Y}$  = predicted Y on reg. line,  $Y$  = observed Y,  $\bar{Y}$  = mean of Y)

### Statistical tables:

These tables will be provided to you during the test, along with this formula sheet:

(1) *t* distribution (pp. 511-512 of FNLG)

(2)  $X^2$  distribution (p.513)

(3) *F* distribution for  $\alpha=.05$  (p.515)