STAT371 DISCUSSION 11 November 17 , 2002

1. The chi-square test for the $r \times k$ contingency table

   - Null hypothesis:

     $$H_0 : \text{Row variable and column variable are independent}$$

   - Calculation of expected frequencies:

     $$E = \frac{(\text{Row total}) \times (\text{Column total})}{\text{Grand total}}$$

   - Test statistics

     $$\chi^2_s = \sum \frac{(O - E)^2}{E}$$

   - Null distribution(approximate): χ^2 distribution with

     $$df = (r-1)(k-1)$$

     where r is the number of rows and k is the number of columns in the contingency table. This approximation is adequate if $E \geq 5$ for evary cell. If r and k are large, the condition that $E \geq 5$ is less critical and the χ^2 approximatoin is adequate if the average expected frequency is at least 5, even if some of the cell counts are smaller.

2. The ANOVA Table

Source:	df	SS(Sum of Squares)	MS(Mean Square)
Between groups:	$k-1$	$\sum_{i=1}^{k} n_i(\bar{y}_i - \bar{\bar{y}})^2$	SS/df
Within groups:	$n^\star - k$	$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2$	SS/df
Total:	$n^\star - 1$	$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{\bar{y}})^2$	

where

$$\bar{y}_i = \frac{y_{i1} + y_{i2} + ... + y_{in_i}}{n_i} = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$$

$$n^\star = n_1 + n_2 + ... + n_k = \sum_{i=1}^{k} n_i$$

and

$$\bar{\bar{y}} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}}{n^\star}$$