STAT371 DISCUSSION 10 November 10 , 2002

1. **Comparison of Paired Samples**
   In a paired design, the observations (Y_1, Y_2) occur in pairs (not independent). Instead of considering Y_1 and Y_2 independently, we consider the DIFFERENCE d, defined as $d = Y_1 - Y_2$.

   - Relationship between sample means and population means:
   $$\bar{d} = \bar{y_1} - \bar{y_2}$$
   $$\mu_d = \mu_1 - \mu_2$$
   - Standard error of $\bar{d}$ is
   $$\mathrm{SE}_{\bar{d}} = \frac{s_d}{\sqrt{n_d}}$$
   - t Test
   $$H_0 : \mu_d = 0$$
   $$t_s = \frac{\bar{d} - 0}{\mathrm{SE}_{\bar{d}}}$$
   - $(1 - \alpha)$ Confidence Interval for μ_d is
   $$\bar{d} \pm t_{\frac{\alpha}{2}} \mathrm{SE}_{\bar{d}}$$
   e.g., If $\alpha = 10\%$, then 90% confidence interval of μ_d is $\bar{d} \pm t_{0.05} \mathrm{SE}_{\bar{d}}$

2. **Analysis of Categorical Data**

   - The chi-square goodness-of-fit test
   $$H_0 : \Pr\{\text{categorical } 1\} = p_1, \Pr\{\text{categorical } 2\} = p_2, ...$$
   H_A: At least one of the probabilities specified in H_0 is incorrect
   Chi-square Statistics is
   $$\chi^2_s = \sum \frac{(O - E)^2}{E}$$

where the summation is over all the categories. O represents the observed frequency of the category and E represents the expected frequency. For categorial i, $E = n \times p_i$, where n is the total number of observations: $n = \sum O$.

Under H_0, if the sample size is large enough, the distribution of χ_s^2 can be approximated by χ^2 distribution with degree of freedom as

$$\text{df} = (\text{number of categories}) - 1$$

- The chi-square test for the 2×2 contingency table

$$H_0 : p_1 = p_2$$

The Chi-squate statistics is

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

where the sum is taken over all four cells in the contingency table. O represents the observed frequency and E represents the correspondint expected frequency according to H_0, and

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

The degree of freedom is

$$df = 1$$