THE UNIVERSITY OF SYDNEY

STAT2012 STATISTICAL TESTS

Semester 2	**Tutorial Week 7**	2015

Summary of week 6

- Multiple comparisons using the *Bonferroni* method: Suppose that $Y_{i1}, Y_{i2}, \cdots, Y_{in_i}$ are independent random samples from $N(\mu_i, \sigma^2)$.

  If the null hypothesis of the equality of means for the g groups is rejected, the test statistic that tests $H_0: \ \mu_l = \mu_m$ is
  $$t_{lm} = \frac{\bar{x}_{l.} - \bar{x}_{m.}}{s_p\sqrt{\frac{1}{n_l} + \frac{1}{n_m}}},$$
  where $s_p^2 = \frac{SSR}{N-g}$ and N is the total sample size. We reject H_0 if
  $$p_{lm} \leq \alpha/r$$
  where p_{lm} is the p-value and $r = g(g-1)/2$.

- Kruskal Wallis test: To test $H_0: \ \mu_1 = \mu_2 = \cdots = \mu_g$ against H_1 : Not all the μ_i's are equal, the test statistic is
  $$k_0 = \frac{SST}{MST_0} = \frac{\sum_{i=1}^{g} n_i \bar{r}_{i.}^2 - N\bar{r}^2}{\left[\sum_{i=1}^{g}\sum_{j=1}^{n_i} r_{ij}^2 - N\bar{r}^2\right]/(N-1)} \overset{\text{no ties}}{=} \frac{12}{N(N+1)}\sum_{i=1}^{g} n_i\left(\bar{r}_i\right)^2 - 3(N+1)$$
  under H_0 where r_{ij} is the rank of y_{ij} in the combined sample.

Tutorial Questions

1. Water samples were taken at three different locations in a river to determine whether the quantity of dissolved oxygen, a measure of water pollution, varied from one location to another. Location one was selected above an industrial plant; location two was adjacent to the industrial water discharge for the plant; and location three was slightly downriver in midstream. Five water specimens were randomly selected at each location. The data are shown in the accompanying table (The greater the pollution, the lower will be the dissolved oxygen readings).

Location 1	Location 2	Location 3
5.9	4.8	6.0
6.1	5.0	6.1
6.3	4.3	5.8
6.1	4.7	5.6
6.0	5.1	5.7

$$\sum_{i=1}^{g}\sum_{j=1}^{n_i} y_{ij} = 83.5, \qquad \sum_{i=1}^{g}\sum_{j=1}^{n_i} y_{ij}^2 = 470.25.$$

(a) Test whether the data provide sufficient evidence to indicate a difference in mean dissolved content for the three locations by using the Kruskal-Wallis test (Hint: $\sum_{i=1}^{g}\sum_{j=1}^{n_i} r_{ij}^2 = 1237.5$).

(b) Test whether there is a difference in mean dissolved content across the three locations using the Bonferroni test.

2. Suppose that you wish to compare the means for two populations X and Y and that $\sigma_X^2 = 9$ and $\sigma_Y^2 = 25$.

   (a) What allocation of $n = 90$ to the two samples will result in the maximum amount of information about $\mu_X - \mu_Y$? (minimize the $var(\bar{X} - \bar{Y})$)

   (b) Suppose that you allocate $n_1 = n_2$ observations to each sample. How large must n_1 and n_2 be in order to obtain the same amount of information as that implied by the solution to (a)?

3. In the case of no ties, show that the Kruskal-Wallis test statistic

$$k_0 = \frac{12}{N(N+1)} \sum_{j=1}^{g} n_j \left(\bar{r}_{\cdot j}\right)^2 - 3(N+1).$$

   Hint: note that if there are no ties in the data,

$$\sum_{j=1}^{g}\sum_{i=1}^{n_j} r_{ij}^2 = \sum_{i=1}^{N} i^2 = \frac{1}{6}N(N+1)(2N+1).$$

4. Assuming $g = 2$, show that the Kruskal-Wallis test statistic $K = \widetilde{W}^2$, where the $\widetilde{W}$ is the standardised Wilcoxon test statistic defined by

$$\widetilde{W} = \frac{n_1\bar{r}_1 - \dfrac{n_1(N+1)}{2}}{\sqrt{\dfrac{n_1 n_2 (N+1)}{12}}}$$

   so $K \sim \chi^2_{g-1}$.

   Hint: $n_1 + n_2 = N$, $\bar{r} = (n_1\bar{r}_{\cdot 1} + n_2\bar{r}_{\cdot 2})/N = (N+1)/2$.

Extra Practice Problems

1. In a diet test, each of four diet programs is applied to a sample of people. At the end of three weeks, the amount of pounds people lost are shown below.

Diet Program			
1	2	3	4
12	19	16	28
6	10	20	17
18	13	26	22
23	20	19	16
	25		20

Test to determine if there is enough evidence at the 5% significance level to infer that at lease two population locations differ.