Statistics 3515 Lecture **2**

Tests of Means After Experimentation

If the null hypothesis $H_0 : \mu_1 = ... = \mu_a$ is rejected then one is interested to find out which of the means are significantly different from each other. A multiple comparison procedure can be employed for this purpose. One approach is to formulate this procedure in terms of simultaneous testing of the following null hypotheses:

$$H_0^{(i,j)} : \mu_i = \mu_j,\ 1 \leq i < j \leq a.$$

The difficulty in implementing this multiple comparison procedure is to guarantee that the overall significance level for testing $\binom{a}{2}$ =a(a-1)/2 hypotheses is equal to a pre-specified value, say .05.

The following is still a popular method by many researchers that use ANOVA methods.

The LSD Method

This method is based on the following t - statistic for testing the equality of two specified means vs a two sided alternative:

$$T = \frac{\overline{Y}_{i.} - \overline{Y}_{j.}}{[MS_E(1/n_i + 1/n_j)]^{.5}}$$

The LSD method proceeds as follows.

1. Arrange the a means in order from low to high, and record the number of the observations used to compute it.

2. Get the MS_E with its degrees of freedom from the ANOVA Table.

3. Evaluate the

$$LSD(n_i,n_j) = t_{\alpha / 2; N-a} [MS_E(1 / n_i + 1 / n_j)]^{.5}, \text{ for}$$

$1 \le i < j \le a$.

Note that for a balanced design there is only one quantity

$$LSD = t_{\alpha/ 2; N-a} (2MS_E / n)^{.5} .$$

4. Compare the observed difference between each pair of averages with the corresponding LSD. If it exceeds the LSD value, we conclude that the corresponding population means are significantly different from each other.

The deficiency of this method is that for testing each pair of means the significance level (probability of type I error) is equal to α and since we testing many pairs of means the overall significance level (probability of type I error) is not known.

Bonferroni Procedure

The Bonferroni procedure is an adjusted LSD method based on the following Bonferroni inequality:

$$P(\bigcup_{i=1}^{m} A_i) \le \sum_{i=1}^{m} P(A_i) .$$

We will use this inequality for the events

$$A_{ij} = \text{reject } H_0^{(i,j)} \text{ when } H_0^{(i,j)} \text{ is true.}$$

Therefore, if we want to maintain an overall significance level of $\alpha = .05$, for example, then we have to have

$$P(\bigcup_{1 \le i < j \le a} A_{ij}) \le \alpha = .05 .$$

If we guarantee that

$$P(A_{ij}) = \frac{\alpha}{\binom{a}{2}}$$

for each i < j than the overall level of significance will be less than α.

Sidak Procedure

This procedure is based on the following product-type inequality that is valid under certain restrictions

$$P(\bigcap_{i=1}^{m} A_i^{\,c}) \geq \prod_{i=1}^{m} P(A_i^{\,c}) \text{ .}$$

Based on the above inequality one can show that

$$P(\bigcup_{1\le i<j\le a} A_{ij}) = 1 - P(\bigcap_{1\le i<j\le a} A_{ij}^{\,c}) \le 1 - \prod_{1\le i<j\le a} P(A_{ij}^{\,c}) = 1 - \prod_{1\le i<j\le a} [1 - P(A_{ij})] \text{ .}$$

Like in the Bonferroni procedure by controlling the probability of $P(A_{ij})$ we can control the overall significance level.

In this course we will use the following multiple comparison procedure.

Tukey's Multiple Comparison Procedure

This method utilizes the *Studentized range*,

$$Q = \frac{\overline{Y}_{largest} - \overline{Y}_{smallest}}{\sqrt{\frac{MS_E}{n}}}$$

where $\overline{Y}_{largest}$ and $\overline{Y}_{smallest}$ are the largest and the smallest sample means for the populations in our study, respectively. The logic behind this procedure is that, if we determine a critical value for the difference between the largest and the smallest sample means, one that implies a difference in their respective treatment means, then any other pair of sample means that differ by as much or more than this critical value would also imply a difference in corresponding treatment means.

The Tukey's multiple comparison procedure is performed as follow.

1. Evaluate

$$T_\alpha = q_\alpha(a,f)\ (MS_E / n)^{.5},$$

where

a = Number of sample means

n = number of observation in each sample

f = degrees of freedom of the MS_E

q_α (a,f) = critical value of the Studentized range from

Table V, on pages A-11 and A-12, (in this table p, instead of a, denotes the number of means), gives percentage points for the Studentized Range Statistic.

2. Rank the a sample means. Any pair of sample means differing by more than T_α will imply a difference in the corresponding population means.

Remark: If the design is not balanced one can use the Tukey-Kramer procedure that uses the following yardstick to compare each pair of means:

$$W_{ij} = q_{\alpha}\,(a,f)\,[.5\,MS_E\,(1 / n_i + 1 / n_j)]^{.5}\,.$$

Remark: Both Tukey and Tukey-Kramer multiple comparisons procedures assume that samples were randomly and independently selected from normal populations with means $\mu_1, \ldots, \mu_a$ and common variance σ^2.