Chapter 11 - Lecture 2
Multiple Comparisons in ANOVA

Yuan Huang

April 10, 2013

Ex 11.2

The lumen output was determined for each of $I = 3$ different brands of 60-watt soft-white lightbulbs, with $J = 8$ bulbs of each brand tested. The sums of squares were computed as $SSE = 4773.3$ and $SSTr = 591.2$. State the hypotheses of interest (including word definitions of parameters), and use the F test of ANOVA($\alpha = 0.05$) to decide whether there are any differences in true average lumen outputs among the three brands for this type of bulb by obtaining as much information as possible about the P-value.

Multiple Comparisons in ANOVA

When the computed value of the F statistic in single-factor ANOVA is

1. not significant, the analysis is terminated because no differences among the μs have been identified.
2. significant, then H_0 is rejected and we conclude that at least two of μs differ. In this case, the investigator will usually want to know which of the μs are different from each other. The analysis would require comparison of each pair and we need to control the family error rate (alternatively, the simultaneous confidence level). A method for carrying out this further analysis is called a multiple comparisons procedure.

Simultaneous confidence interval

For each $i < j$, the simultaneous $1 - \alpha$ confidence interval that every interval includes the corresponding value of $\mu_i - \mu_j$ is

$$\bar{x}_{i\cdot} - \bar{x}_{j\cdot} \pm Q_{\alpha, I, I(J-1)} \sqrt{\mathrm{MSR}/J}$$

What is the Q score?

Studentized range distribution

This Q score is the quantile from Studentized range distribution.

Definition

Let Z_1, Z_2, ..., Z_m be m independent standard normal random variables and W be a χ_v random variable, independent of Z_is. then the distribution of

$$Q = \frac{\max |Z_i - Z_j|}{\sqrt{W/v}} = \frac{\max(Z_1, \ldots, Z_m) - \min(Z_1, \ldots, Z_m)}{\sqrt{W/v}}$$

This distribution has two parameters :

1. m = the number of Z_is
2. v = degree freedom of the χ_v.

Usage

1. In addition to construct the Simultaneous confidence interval
2. We can apply this studentized range distribution for multiple comparisons (Tukey's procedure).

Tukey's procedure

Define $w = Q_{\alpha,I,I(J-1)}\sqrt{\text{MSR}/J}$. This w is called Tukeys honestly significantly difference (HSD).

1. List the sample means in increasing order. For example, $\bar{x}_{2\cdot} < \bar{x}_{5\cdot} < \bar{x}_{4\cdot} < \bar{x}_{1\cdot} < \bar{x}_{3\cdot}$
2. Consider first the smallest mean $\bar{x}_{2\cdot}$. If $\bar{x}_{5\cdot} - \bar{x}_{2\cdot} > w$, proceed to step 2. If $\bar{x}_{5\cdot} - \bar{x}_{2\cdot} < w$,connect these first two means with a line segment. Then if possible extend this line segment even further to the right to the largest $\bar{x}_{i\cdot}$, that differs from $\bar{x}_{2\cdot}$ by less than w.
3. Now move to $\bar{x}_{5\cdot}$, and again extend a line segment to the largest $\bar{x}_{i\cdot}$, to its right that differs from $\bar{x}_{5\cdot}$, by less than w.
4. Continue by moving to $\bar{x}_{4\cdot}$, and finally to $\bar{x}_{1\cdot}$.

Example

1. $\bar{x}_{1\cdot} = 14.5, \bar{x}_{2\cdot} = 13.8, \bar{x}_{3\cdot} = 13.3, \bar{x}_{4\cdot} = 14.3, \bar{x}_{5\cdot} = 13.1, w = 0.4$

2. $\bar{x}_{1\cdot} = 79.28, \bar{x}_{2\cdot} = 61.45, \bar{x}_{3\cdot} = 47.92, \bar{x}_{4\cdot} = 32.76, w = 17.47$