Multiple Comparisons Including Tukey's Method

Topics

- Multiple Comparisons
  - Data Set
  - Idea
  - Tukey

There are a number of other MC methods, but they will not be discussed here

Data Set

611	AlkName		923	AlkStore		445	HDName		476	HDStore
537	AlkName		794	AlkStore		490	HDName		569	HDStore
542	AlkName		827	AlkStore		384	HDName		480	HDStore
593	AlkName		898	AlkStore		413	HDName		460	HDStore

Data: Lifetime of battery per unit cost

Explained more in the R code

From Dean&Voss(1999),
Design and Analysis of Experiments

t-Test Method

- Right now, let's only consider the first 2 Types
- You have already seen this!

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij},\ \varepsilon_{ij} \sim i.i.d.N(0,\sigma^2), i = 1, 2,\ j = 1, \ldots, n$$

- For this completely randomized design, the confidence interval becomes

$$(\tau_i - \tau_{i'}) \in (\bar{y}_{i.} - \bar{y}_{i'.}) \pm w_t \sqrt{\widehat{Var}(\bar{y}_{i.} - \bar{y}_{i'.})}$$
$$(\tau_1 - \tau_2) \in (\bar{y}_{1.} - \bar{y}_{2.}) \pm w_t \sqrt{\widehat{Var}(\bar{y}_{1.} - \bar{y}_{2.})}$$

$$w_t = t_{df_E,\alpha}$$

t-Test Method with Four Types?

- Say we wanted to compare all possible pairs of the four types of batteries
- How many?
- If you construct just one 95% C.I., what is the chance that the (hypothetical) long-term mean will be captured?
- What if you construct all of them. What is the chance that the (hypothetical) long-term mean will be captured by *all* of them?

Multiple Comparisons

- Often, we are interested in:
  - Multiple hypothesis tests, e.g. for multiple contrasts
  - Multiple CI's, e.g. for multiple contrasts
- Problem
  - When several tests are made, say each at the same α level, the overall α level is greater
  - When several confidence intervals are made, the probability that they are *all* correct simultaneously is less than their individual levels
- This is called the *multiple-comparisons problem*
- The tests or CI's are called *simultaneous*

Tukey Method

- In some cases, confidence intervals are required for all possible pairwise difference contrasts
- Tukey (1953) proposed a technique to create these confidence
- When the sample sizes are equal, the overall confidence level is exactly $(1-\alpha)100\%$; otherwise it can be made to be at least $(1-\alpha)100\%$.

Tukey Method

- Used for all pairwise comparisons. Almost exclusively used when the factor has >2 levels (for 2 levels, you can just use the t-test).
- Provides an exact experimentwise error rate

Tukey Method

- For completely randomized designs, the simultaneous confidence intervals become

$$(\tau_i - \tau_{i'}) \in (\overline{y}_{i.} - \overline{y}_{i'.}) \pm w_T \sqrt{\widehat{Var}(\overline{y}_{i.} - \overline{y}_{i'.})}$$

where $i \neq i'$,

$$w_T = q_{I,df_E,\alpha} / \sqrt{2}$$

I is the number of levels of the factor, and q will be defined in the next few slides.

Studentized Range Distribution

- For equal sample sizes, the formula for Tukey's simultaneous confidence intervals is based on the distribution of the statistic

$$Q = \frac{\max(T_i) - \min(T_i)}{\sqrt{MS_E / n}}$$

where, to test the null hypothesis that all treatment means are equal

$$T_i = \bar{Y}_{i.}$$

and the max and min are taken over i=1,...,I

Studentized Range Distribution

- Under the usual assumptions, the numerator of Q is the range of I independent $N(0,\sigma^2/n)$ random variables, is standardized by an estimated standard deviation (but without the $\sqrt{2}$ term)
- It is possible to quantify this distribution. This means that we can find values $q_{I,df_E,\alpha}$ that satisfy

$$P\left(\frac{\max(T_i)-\min(T_i)}{\sqrt{MS_E/n}} \le q_{I,df_E,\alpha}\right)=1-\alpha$$

Battery Experiment

- An experiment was run to see if there are differences between 4 types of batteries comparing mean life per unit cost
- There were 4 replicates at each treatment and the MS_E was 2367.71
- The average lives per unit (in minutes/dollar) are

$$\bar{y}_{1.} = 570.75, \bar{y}_{2.} = 860.50,$$
$$\bar{y}_{3.} = 433.00, \ \bar{y}_{4.} = 496.25$$

- Find the simultaneous 95% confidence intervals for all the pairwise differences.

Results

$$\widehat{Var}\left(\overline{y}_{i.}-\overline{y}_{j.}\right)=MS_E\left(\frac{1}{n_i}+\frac{1}{n_j}\right)=2367.71(0.25+0.25)$$

$$=1183.855$$

$$q_{4,12,0.05}=4.19866$$

$$msd=\left(4.1986/\sqrt{2}\right)\sqrt{1183.855}=102.19$$

For example

$$\left(\tau_1-\tau_2\right)\in -289.75\pm 102.19=\left(-391.94,-187.56\right)$$

Remaining Contrasts

Contrast	Lower CI	Upper CI
$\tau_1 - \tau_3$	39.60	239.90
$\tau_1 - \tau_4$	-27.65	176.65
$\tau_2 - \tau_3$	325.35	529.65
$\tau_2 - \tau_4$	262.10	466.40
$\tau_3 - \tau_4$	-165.40	38.90

Conclusions

- Since the confidence intervals for the differences between battery type 1 and 4, and 3 and 4 include zero, there is not enough evidence to say that these differences are different from zero
- There is enough evidence to conclude that the differences for the other pairs are different from zero
- A hypothesis-testing framework could be used instead. For example, see if a pair of sample means differs by more than the *msd*. If so, declare the treatment means different.

Topics

- Multiple Comparisons
  - Data Set
  - Idea
  - Tukey
  - Other methods…