

Ch 17 – Analysis of Factor Level Effects (revised 1/16/09)

Motivation: ANOVA tests $H_0 : \tau_1 = \dots = \tau_r$. If the null hypothesis is rejected, what does that tell us? Not much! More informative are individual or simultaneous confidence intervals (or tests) for treatment contrasts of interest.

Treatment Contrast: $L = \sum_i c_i \tau_i$, where $\sum_i c_i = 0$, for given constants c_i .

Pairwise comparison, or pairwise difference: $D = \tau_i - \tau_{i'}$, for $i \neq i'$.

Using least squares estimates, $\hat{L} = \sum_i c_i \hat{\tau}_i = \sum_i c_i \bar{Y}_i$, $E[\hat{L}] = \sum_i c_i \tau_i$, $\sigma^2(\hat{L}) = \sum_i c_i^2 \sigma^2 / n_i$, $s^2(\hat{L}) = \sum_i c_i^2 \hat{\sigma}^2 / n_i$, and

$$t^* = \frac{\hat{L} - L}{s(\hat{L})} \sim t(n_T - r).$$

As a special case, $\hat{D} = \hat{\tau}_i - \hat{\tau}_{i'} = \bar{Y}_i - \bar{Y}_{i'}$, $E[\hat{D}] = \tau_i - \tau_{i'}$, $\sigma^2(\hat{D}) = (1/n_i + 1/n_{i'})\sigma^2$, and $s^2(\hat{D}) = (1/n_i + 1/n_{i'})\hat{\sigma}^2$.

$\hat{\sigma}^2 = MSE$

Confidence intervals and t -tests

Example: KFC, summary statistics given in Table 17.1, p. 734.

Conduct a two-tailed test that the average effect of designs 1 and 2 equals the average effect of designs 3 and 4. Determine the P-value, and interpret the results.

Construct a 95% confidence interval for each pairwise comparison.

Motivation for multiple comparison procedures: In the above example, there are $\binom{4}{2} = 6$ pairwise comparisons, each of the six confidence intervals has a 5% chance of missing, so the chance at least one misses is at most 30%, (i.e. $P(A_1 \cup A_2 \cup \dots \cup A_6) \leq P(A_1) + P(A_2) + \dots + P(A_6) = 0.30$, for A_i the event the i th interval misses). So, taking the complement of this event, we can only be at least 70% confident that all six confidence intervals work! In short, if we make many comparisons, it seems prudent to want a fairly high level of confidence that all of the inferences are correct, aka *simultaneous confidence*. Multiple comparison procedures provide this simultaneous confidence. We will consider three such procedures: the Bonferroni method, Tukey's method, and Scheffé's method.

— end of class #3, 1/13/09 —

The Bonferroni Method

The Bonferroni Method is motivated and justified by the above example and motivation. Suppose one has g pre-planned treatment contrasts of interest. Then one has at least $100(1 - \alpha)\%$ simultaneous confidence in the g corresponding confidence intervals of the form:

$$\hat{L} \pm B s(\hat{L}), \quad \text{where } B = t(1 - \alpha/2g; \nu)$$

Example: KFC, see ch17eg1.txt

Tukey's method

If Y_1, \dots, Y_r are i.i.d. $N(\mu, \sigma^2)$ (i.e. same mean, same variance), $W = \max(Y_i) - \min(Y_i)$ is the range, and S^2 is an estimate of σ^2 based on ν degrees of freedom such that $\nu S^2/\sigma^2 \sim \chi^2(\nu)$ independent of the Y_i 's, then $Q(r, \nu) = W/S$ is called the *Studentized range* and has the *Studentized Range Distribution*, which depends on the two parameters r and ν . Table B.9 (pp. 1333–1335) provides select upper- α quantiles $q(1 - \alpha; r, \nu)$.

Given equal samples sizes n for r levels of a single treatment factor, under our factor effects model, consider the treatment sample means \bar{Y}_i , ($i = 1, \dots, r$). The quantities $\bar{Y}_i - \tau_i$ are i.i.d. $N(\mu, \sigma^2/n)$, and $S^2/n = MSE/n$ provides an appropriate estimate of σ^2/n with $\nu = n_T - r = r(n - 1)$ degrees of freedom. It follows that

$$Q = \frac{\max(\bar{Y}_i - \tau_i) - \min(\bar{Y}_i - \tau_i)}{S(\bar{Y}_i)} \sim Q(r, \nu),$$

for $S^2(\bar{Y}_i) = \sigma^2/n$ and $\nu = r(n - 1)$. Hence,

$$\begin{aligned} 1 - \alpha &= P \left[\frac{\max(\bar{Y}_i - \tau_i) - \min(\bar{Y}_i - \tau_i)}{S/\sqrt{n}} \leq q(1 - \alpha; r, \nu) \right] \\ &= P \left[\frac{|\bar{Y}_i - \tau_i - (\bar{Y}_{i'} - \tau_{i'})|}{S/\sqrt{n}} \leq q(1 - \alpha; r, \nu) \text{ for all } i \neq i' \right] \\ &= P \left[\tau_i - \tau_{i'} \in \bar{Y}_i - \bar{Y}_{i'} \pm q(1 - \alpha; r, \nu)(S/\sqrt{n}) \text{ for all } i \neq i' \right] \\ &= P \left[\tau_i - \tau_{i'} \in \bar{Y}_i - \bar{Y}_{i'} \pm [q(1 - \alpha; r, \nu)/\sqrt{2}]s(\hat{D}) \text{ for all } i \neq i' \right] \end{aligned}$$

Taking $\bar{y}_i - \bar{y}_{i'} \pm q(1 - \alpha; r, \nu)(s/\sqrt{n})$, or equivalently, $\bar{y}_i - \bar{y}_{i'} \pm [q(1 - \alpha; r, \nu)/\sqrt{2}]s(\hat{D})$, as the confidence interval for $\tau_i - \tau_{i'}$, this formulae provides simultaneous $100(1 - \alpha)\%$ confidence intervals for all pairwise comparisons of the treatment effects.

The quantity $q(1 - \alpha; r, \nu)(s/\sqrt{n}) = [q(1 - \alpha; r, \nu)/\sqrt{2}]s(\hat{D}) = \text{msd}$, say, is called the *minimum significant difference*. Why? The first form, $\text{msd} = q(1 - \alpha; r, \nu)(s/\sqrt{n})$, is the way it was presented by Tukey, but this requires equal sample sizes. The second form, $\text{msd} = [q(1 - \alpha; r, \nu)/\sqrt{2}]s(\hat{D})$, in terms of the standard error of the pairwise difference, was recommended by Kramer (to allow) for the case of unequal sample sizes, in which case the method is called the *Tukey-Kramer method* and $s^2(\hat{D}) = s^2(1/n_i + 1/n_{i'})$. Using this approach with unequal sample sizes has since been shown to be conservative. (What does this mean?)

General formulae for equal or unequal sample sizes:

$$\hat{D} \pm T s(\hat{D}), \quad \text{where } T = q(1 - \alpha; r, \nu)/\sqrt{2}$$

Example: KFC, see ch17eg1.txt

Example: ch17eg2.txt, SAS graphical method for displaying results if equal sample sizes.

Scheffé's method

Scheffé's method provides simultaneous $100(1 - \alpha)\%$ confidence intervals for all treatment contrasts and corresponds directly to the ANOVA F -test of equality of treatment effects.

The ANOVA test for equal treatment means *fails* to reject the null hypothesis at significance level α if $F^* \leq F(1 - \alpha; r - 1, \nu)$, for $\nu = n_T - r$, where $F^* = [\text{SSTR}/[(r - 1)\text{MSE}]]$. One can show, for any set of $r - 1$ orthogonal (i.e. independently estimated) treatment contrasts L_1, \dots, L_{r-1} , that $\text{SSTR} = \text{SS}(L_1) + \dots + \text{SS}(L_{r-1})$. In other words, SSTR is equal to the sum of the $r - 1$ sums of squares associated with any set of $r - 1$ orthogonal treatment contrasts. What is the sum of squares for a treatment contrast? For any treatment contrast L , if $V(\hat{L}) = a^2\sigma^2$ for some constant a , then under our model $\hat{L}/a \sim N(L/a, \sigma^2)$ and $\text{SS}(L) = \text{SS}(\hat{L}/a) = (\hat{L}/a)^2$, with $E[\text{SS}(L)] = \sigma^2 + (L/a)^2$.

Hence,

$$\begin{aligned}
 1 - \alpha &= P(F^* \leq F(1 - \alpha; r - 1, \nu) \mid H_0 \text{ true}), \quad \text{for } F^* = \text{SSTR}/[(r - 1)\text{MSE}] \\
 &= P\left(\sum_i [(\hat{L}_i - L_i)/a_i]^2 / [(r - 1)\text{MSE}] \leq F(1 - \alpha; r - 1, \nu) \text{ for any orthogonal decomposition}\right) \\
 &\quad \text{where } V(\hat{L}) = a_i^2\sigma^2, \\
 &\leq P([\hat{L} - L]/a]^2 / [(r - 1)\text{MSE}] \leq F(1 - \alpha; r - 1, \nu) \text{ for all treatment contrasts } L) \\
 &= P([\hat{L} - L]^2 \leq (r - 1) F(1 - \alpha; r - 1, \nu) [a^2\text{MSE}] \text{ for all treatment contrasts } L) \\
 &= P([\hat{L} - L]^2 \leq (r - 1) F(1 - \alpha; r - 1, \nu) s^2(\hat{L}) \text{ for all treatment contrasts } L) \\
 &= P([\hat{L} - L]^2 \leq S^2 s^2(\hat{L}) \text{ for all treatment contrasts } L), \quad \text{for } S^2 = (r - 1) F(1 - \alpha; r - 1, \nu) \\
 &= P(|\hat{L} - L| \leq S s(\hat{L}) \text{ for all treatment contrasts } L) \\
 &= P(L \in \hat{L} \pm S s(\hat{L}) \text{ for all treatment contrasts } L)
 \end{aligned}$$

Note: One can show that the one inequality is in fact an equality—namely, there exists a treatment contrast L such that $\text{SS}(L) = \text{SSTR}$, (i.e. take $c_i = n_i(\bar{y}_i - \bar{y}..)$). So, the F -test rejects equality of treatment effects at significance level α if and only if there exists a treatment contrast that is significantly non-zero based on $100(1 - \alpha)\%$ confidence intervals for all treatment contrasts by Scheffé's method.

General formulae for Scheffé's method for $100(1 - \alpha)\%$ simultaneous confidence for all treatment contrasts L :

$$L \in \hat{L} \pm S s(\hat{L}), \quad \text{where } S^2 = (r - 1)F(1 - \alpha; r - 1, \nu)$$

Example: KFC, see ch17eg1.txt

Comments:

- Recall when each method of multiple comparisons is applicable: the Bonferroni method requires a fixed number of pre-planned contrasts, Tukey's method is for all pairwise comparisons, and Scheffé's method is for all treatment contrasts.
- Given a choice between applicable methods, choose the one that gives the smallest confidence intervals—namely, the one with the smallest msd , or equivalently, the one corresponding to the smallest value of B , T or S .
- Tukey's method is always best for all pairwise comparisons.
- The Bonferroni method is generally best for a relatively small number of pre-planned comparisons.
- Scheffé's method gives one carte blanche to look at all treatment contrasts. Furthermore, it is equivalent to the F -test for equality of treatment effects in the following sense: the F -test rejects equality of treatment effects at significance level α if and only if there exists a treatment contrast that is significantly non-zero based on $100(1 - \alpha)\%$ confidence intervals for all treatment contrasts by Scheffé's method.
- The Bonferroni, Tukey and Scheffé methods can be conducted as simultaneous (or multiple) tests rather than confidence intervals.
- One can combine methods. Example:
- Other methods exist: Dunnett's method (MCC), Hsu's method (MCB), etc.

Homework: 17.10b–f, 17.11b–f (ignoring the last sentence of 17.11e), 17.16.

— end of class #4, 1/15/09 —

17.8 Planning of Sample Sizes

Example: Consider a completely randomized design to compare five treatments using Tukey's method and simultaneous 95% confidence. What sample size is required to have the minimum significant difference be at most four if one anticipates having $s = \sqrt{mse} \leq 4$?

Homework: 17.23 (assume mse as obtained in exercise 16.8).

Example: ch16eg1.txt — discuss SAS statements: `lsmeans`, `estimate`, `contrast`

Summary of multiple comparisons formulae:

- Bonferroni: $\hat{L} \pm B s(\hat{L})$, where $B = t(1 - \alpha/2g; \nu)$, for g preplanned treatment contrasts
- Tukey: $\hat{D} \pm T s(\hat{D})$, where $T = q(1 - \alpha; r, \nu)/\sqrt{2}$, for all pairwise comparisons
- Scheffé: $\hat{L} \pm S s(\hat{L})$, where $S^2 = (r - 1)F(1 - \alpha; r - 1, \nu)$, for all treatment contrasts