Section 8.2: Pairwise Comparisons and Inference After ANOVA

Eric D. Nordmoe

Math 261
Biostatistics
Kalamazoo College

Outline

- ANOVA Review
- Follow-up Tests
  - Fisher's LSD Method (no multiple comparison adjustment)
  - Bonferroni Method (k comparisons)
  - Tukey's Method (all comparisons)

Required Conditions for ANOVA

- Independent random samples from
- Either ...
  - Each of k populations OR
  - A randomized comparative experiment with k treatments
- The response variable in each population has a Normal distribution
- The standard deviations are equal

$$\sigma_1 = \sigma_2 = \cdots = \sigma_k = \sigma$$

A Problem and a Solution

- Problem: Multiple comparisons
  - Using separate two-sample t confidence intervals to compare multiple pairs of means renders the nominal confidence levels of the pairwise comparisons invalid.
  - If all pairwise intervals are computed at the 95% level, the chance that all intervals contain the corresponding parameters will be (much) less than 95%.
- Solution: ANOVA with follow-up procedures
  - Carry out ANOVA to test the hypothesis that all population means are the same using the F statistic

$$F = \frac{\text{variation among the sample means}}{\text{variation within individuals in the same sample}}$$

Multiple Comparisons

Follow-up to Global *F* Test

- If the global F-test rejects equality of means, use multiple comparison (post hoc or follow-up) methods to calculate "simultaneous" confidence intervals.
- Three methods are described in what follows:
  - Fisher's Least Significant Difference (no adjustment for multiple comparisons):
  - Bonferroni method
  - Tukey's Pairwise Multiple Comparisons Method

Note: All can be done using results from RStudio.

Fisher's Least Significant Difference Method

- The Fisher's LSD method is an extension of the 2-sample t procedures which makes no adjustment for multiple comparisons .
- It is rarely advisable since it yields faulty intervals and significance tests.
- For a one-way design with k groups, a Fisher's LSD confidence interval for $\mu_1 - \mu_2$ is:

$$(\bar{x}_1 - \bar{x}_2) \pm t^* s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where t^* is the ordinary (unadjusted) critical value for the t(df) density curve with area $1 - \alpha$ between $-t^*$ and t^* and $s_p = \sqrt{\text{MSE}}$.

Bonferroni Method

- The Bonferroni method is a general method for multiple comparisons.
  - Can be applied whenever multiple significance tests are being carried out.
  - Very conservative.
- Basic idea:
  - Goal: Obtain *c* confidence intervals with *simulataneous* confidence level $100(1-\alpha)\%$ confidence.
  - Procedure: Calculate each confidence interval using $100(1-\alpha)\%$ confidence.
- For a one-way ANOVA with *k* groups:
  - Total possible comparison: $c = k(k-1)/2$.
  - Test each at α/c level of significance.
- For confidence intervals, use multiplier t^* with area $1-\alpha/c$ between $-t^*$ and t^* to obtain overall $100(1-\alpha)\%$ confidence.

Bonferroni Example

Mice Diet Restriction Study

- In the Mice diet restriction study, there are $k = 6$ groups and $6(6-1)/2 = 15$ possible comparisons but ...
  - Only 5 comparisons are of interest so let $c = 5$.
- Desired overall level of significance is $\alpha = 0.05$.
- Set pairwise error rate equal to $.05/5 = .01$.
- Test $H_0 : \mu_1 = \mu_2$ by test statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where

$$t \sim t(\text{df}_E) = \text{ and } s_p = \sqrt{\text{MSE}}$$

and $\text{df}_E = n - k$.

Bonferroni Example

Confidence Intervals

- Bonferroni confidence interval:

$$(\bar{x}_1 - \bar{x}_2) \pm t^* s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

  where c is the number of intervals and t^* is the critical value for the t(df) density curve with area $1 - \alpha/c$ between $-t^*$ and t^*.

  - Each interval has confidence level $100(1 - \alpha/c)\%$.
  - Overall confidence level is $100(1 - \alpha)\%$.

Bonferroni Example

R Output for Diet Restriction Study

Consult R code examples

Tukey's Pairwise Multiple Comparisons Method

- The Tukey method takes the number of comparisons into account by replacing the critical value t^* by another critical value based on the distribution of the difference between the largest and smallest of a set of k sample means.
  - The Studentized Range Distribution
    - See simulation

Tukey's Pairwise Multiple Comparisons Method

- Tukey Simultaneous Confidence Intervals for all pairwise differences $\mu_i - \mu_j$ among the population means have the form

$$(\bar{x}_i - \bar{x}_j) \pm \frac{q*}{\sqrt{2}} s_p \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

  where $s_p = \sqrt{\text{MSE}}$ and q^* is the upper α critical value from the Studentized Range distribution with parameters k equal to the number of group means being compared and $n - k$, the error degrees of freedom.
  - If all group sample sizes n_i are equal, the overall level of confidence C is correct.
  - If sample sizes differ across groups, the true confidence level is at least C so the test is conservative.
  - R can calculate values of q^* using the qtukey function:
    - `qtukey(1-alpha,k,n-k)`
    - `qtukey(.95,6,343)`

Note: R computes all intervals upon request using `TukeyHSD` function.

Tukey's Pairwise Multiple Comparisons Method

- Tukey Pairwise tests of significance to carry out by hand simultaneous tests of the hypotheses

$$H_0 : \mu_i = \mu_j$$
$$H_a : \mu_i \neq \mu_j$$

  for all pairs of population means, reject H_0 for any pair whose confidence interval does not contain 0. These tests have an overall significance level no less than $1 - C$.

Note: R calculates a test statistic based on the Studentized Range distribution and obtains an adjusted p-value for each comparison. The TukeyHSD output also includes pairwise confidence intervals.

Tukey's Pairwise Multiple Comparisons Method

R Output

See R script and output.

Summary of Multiple Comparison Methods

- Fisher's LSD method is most liberal and is generally not advisable.
- Bonferroni's method is most conservative but versatile
  - Can be used in many contexts
  - Can be used for all pairwise comparisons or a subset (number being compared determines c)
- Tukey's method is best when all pairwise comparisons are of interest.
  - First, do the overall F test.