Multiple Comparisons

Applied Regression and Other Multivariable Methods
Sections 17-7 - 17-9

Linear Combination of Means

- Consider fixed effects model

$$\begin{aligned} Y_{ij} &= \mu + \alpha_i + E_{ij} \\ &= \mu_i + E_{ij} \end{aligned}$$

- Often hypotheses different than H_0 : all $\alpha_i = 0$
- Would like to test $H_0 : L = \sum c_i \mu_i = L_0$
  - Pairwise comparisons
  - Treatments vs control
  - Comparing combinations of trts
  - (Curvi)linear relationship between means and trts (of interest when treatment levels are ordered - like regression)
- Hypotheses may be planned or "after the fact"

Can use statistical model to construct t-test (F-test)

$$\widehat{L} = \sum c_i \overline{Y}_i \qquad \text{Var}(\widehat{L}) = \text{Var}(\sum c_i \overline{Y}_i) = \sum c_i^2 \text{Var}(\overline{Y}_i) = \text{MSE} \sum (c_i^2/n_i)$$

$$t_o = \frac{\widehat{L} - L_0}{\sqrt{\text{Var}(\widehat{L})}}$$

Under H_0: $t_o \frown t_{N-k}$

Linear Combination of Means

- Why t-test instead of overall F test?
  - T-test specifically addresses hypothesis of interest
  - Overall F-test jointly tests all possible contrasts
  - Why include contrasts of no interest in a test?

    If overall error rate equal to α

    Error rate for single comparison $< \alpha$

    F test will reduce power ("water down" test)

    May find individual test significant while F not
- Problems with linear combinations
  - Multiple tests inflate overall error rate
  - Not all combinations independent
- Will discuss procedures that control error rates

Pairwise Comparison of Means

- Can be expressed as a linear combination
- If comparing trt j and trt j'

$$c_i = \begin{cases} 1 & \text{if } i = j \\ -1 & \text{if } i = j' \\ 0 & \text{otherwise} \end{cases}$$

- Trade-off between power (probability of finding true differences) and prob(Type I error)
- No one "best" procedure
- Comparison methods vary in protection
  - Control experimentwise error (overall Type I)
    Lower power but lower prob(Type I error)
  - Control comparisonwise error (indiv Type I)
    High prob(Type I error) but high power

Pairwise Comparison Methods

- Least Significant Difference (LSD)
  - Use MSE and Error df from ANOVA table
  - Perform usual t-test using α

$$\frac{\overline{Y}_i - \overline{Y}_j}{\sqrt{\text{MSE}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}} \sim t_{N-k}$$

  - Does not control experimentwise error
  - High power but high overall P(Type I)
- Bonferroni's Method
  - Use MSE and Error df from ANOVA table
  - Adjust individual α level for # of tests
  - Will perform $m = k(k-1)/2$ tests
  - Perform usual t-test with $\alpha^* = \alpha/m$
  - Controls experimentwise error (no greater than α)
  - Low power that gets lower the larger number of trts

Pairwise Comparison Methods

- Tukey's (Tukey-Kramer) Method
  - Uses studentized range distribution q instead of t
  - Distribution based on the range $\overline{Y}_{\text{MAX}} - \overline{Y}_{\text{MIN}}$ out of k trts
  - Controls overall experimentwise error rate α
  - Reject if (q available in Table A-6)

$$|\overline{Y}_i - \overline{Y}_j| > \frac{q_{\alpha,k,N-k}}{\sqrt{2}}\sqrt{\text{MSE}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

- Student-Newman-Keuls Test
  - Performs Tukey's test for varying number of trts
  - Start with $p = k$, then $p = k-1$, etc

$$K_p = q_\alpha(p, N-k)\sqrt{\text{MSE}/n}$$

  - Stop whenever difference not found significant
  - Controls experimentwise error for all tests with m means
  - When unequal sample size $n = k/\sum 1/n_i$

Pairwise Comparison Methods

- Dunnett's Test
  - Specifically designed for trts vs control situation
  - Distribution based on MAX($\overline{Y}_{\text{Control}} - \overline{y}_{\text{Trt}}$) for $k-1$ trts
  - Similar in approach to Tukey's test
  - Controls experimentwise error rate

All pairwise procedures of the form

$$|\overline{Y}_i - \overline{Y}_j| > C\sqrt{\text{MSE}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

The difference is how C is determined

SNK method allows C to vary

General rule: Bigger $C \rightarrow$ lower power

Contrasts

- Linear combination with coeff's that sum to zero
- Contains pairwise comparisons but allows others
- $L = \sum c_i\mu_i = 0$ with $\sum c_i = 0$

  To compare Trt1 and Trt2 : $c_i = \{1, -1, 0, \ldots, 0\}$

  To compare Trt1 vs .5(Trt2+Trt3) : $c_i = \{1, -.5, -.5, 0, \ldots, 0\}$
- Scheffé's Method

  Set up $1-\alpha$ simultaneous CI for **all contrasts**

  Protects for unplanned comparisons

  Overall error rate at most $\alpha \leftrightarrow$ low power

  Uses $\sqrt{(k-1)F_{k-1,N-k}}$ instead of t distribution

  For pairwise comparisons, reject if

$$|\overline{Y}_i - \overline{Y}_j| > \sqrt{(k-1)F_{k-1,N-k,\alpha}}\sqrt{\text{MSE}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

Text Example

```
options nocenter ls=75;

data potency;
 input sub y @@;
 cards;
 1 29 1 28 1 23 1 26 1 26
 1 19 1 25 1 29 1 26 1 28
 2 17 2 25 2 24 2 19 2 28
 2 21 2 20 2 25 2 19 2 24
 3 17 3 16 3 21 3 22 3 23
 3 18 3 20 3 17 3 25 3 21
 4 18 4 20 4 25 4 24 4 16
 4 20 4 20 4 17 4 19 4 17
 ;

proc glm;
 class sub;
 model y = sub;
 means sub / lsd bon tukey snk scheffe;
run;
quit;
```



```
The GLM Procedure
                                   Sum of
Source            DF      Squares   Mean Square  F Value  Pr > F
Model              3   249.8750000   83.2916667     8.55  0.0002
Error             36   350.9000000    9.7472222
Corrected Total   39   600.7750000

R-Square     Coeff Var      Root MSE        y Mean
0.415921      14.23970      3.122054      21.92500

Source            DF    Type I SS   Mean Square  F Value  Pr > F
sub                3   249.8750000   83.2916667     8.55  0.0002

Source            DF  Type III SS   Mean Square  F Value  Pr > F
sub                3   249.8750000   83.2916667     8.55  0.0002

t Tests (LSD) for y
NOTE: This test controls the Type I comparisonwise error
rate, not the experimentwise error rate.
Alpha                              0.05
Error Degrees of Freedom             36
Error Mean Square              9.747222
Critical Value of t             2.02809
Least Significant Difference     2.8317

Means with the same letter are not significantly different.

           Mean      N    sub
A        25.900     10    1

B        22.200     10    2
B
B        20.000     10    3
B
B        19.600     10    4
```



```
Bonferroni (Dunn) t Tests for y
NOTE: This test controls the Type I experimentwise error
rate, but it generally has a higher Type II error rate.
Alpha                              0.05
Error Degrees of Freedom             36
Error Mean Square              9.747222
Critical Value of t             2.79197
Minimum Significant Difference   3.8982

              Mean      N    sub
     A      25.900     10    1
     A
B    A      22.200     10    2
B
B           20.000     10    3
B
B           19.600     10    4

Tukey's Studentized Range (HSD) Test for y
NOTE: This test controls the Type I experimentwise error
rate, but it generally has a higher Type II error rate.
Alpha                                   0.05
Error Degrees of Freedom                  36
Error Mean Square                   9.747222
Critical Value of Studentized Range  3.80880
Minimum Significant Difference        3.7604

              Mean      N    sub
     A      25.900     10    1
     A
B    A      22.200     10    2
B
B           20.000     10    3
B
B           19.600     10    4
```



```
Student-Newman-Keuls Test for y
NOTE: This test controls the Type I experimentwise error
rate under the complete null hypothesis but not under
partial null hypotheses.
Alpha                          0.05
Error Degrees of Freedom         36
Error Mean Square          9.747222
Number of Means               2            3            4
Critical Range         2.831702    3.4127901     3.760353

           Mean      N    sub
A        25.900     10    1

B        22.200     10    2
B
B        20.000     10    3
B
B        19.600     10    4

Scheffe's Test for y
NOTE: This test controls the Type I experimentwise error rate.
Alpha                              0.05
Error Degrees of Freedom             36
Error Mean Square              9.747222
Critical Value of F             2.86627
Minimum Significant Difference   4.0942

              Mean      N    sub
     A      25.900     10    1
     A
B    A      22.200     10    2
B
B           20.000     10    3
B
B           19.600     10    4
```

Orthogonal Contrasts

- Suppose you have two contrasts $\{c_{Ai}\}$ and $\{c_{Bi}\}$
- **Orthogonal** if $\sum c_{Ai}c_{Bi}/n_i = 0$

- Can divide SS(Trt) into $k-1$ orthogonal contrast SS
- Each contrast has 1 and $N-k$ df
- Pairwise comparisons are not all orthogonal

- Can also use orthogonal contrasts to study trend
- Only interesting if treatments quantitative (ordered)
- If equally spaced trts and $n_i = n$, c_i in Table A-7
- Results in breakdown of polynomial regression

Text Example 17-10

The dissolved oxygen content was measured at points 0, 10, 20, 30 miles downstream from a plant's wastewater discharge location. Three measurements were taken at each point.

Is there a difference in the dissolved oxygen content?

Is there a linear trend with distance?

```
data oxygen;
 input loc y @@;
 if loc=1 then dist=0;  if loc=2 then dist=10;
 if loc=3 then dist=20; if loc=4 then dist=30;
 cards;
 1 4 1 5 1 6
 2 6 2 6 2 6
 3 7 3 8 3 9
 4 8 4 9 4 10
 ;

 proc glm;
  class loc;
  model y = loc;
  contrast 'linear' loc -3 -1 1 3;
  contrast 'quadratic' loc 1 -1 -1 1;
  contrast 'linear' loc -1 3 -3 1;

 proc reg;
  model y = dist;
 run;
 quit;
```



```
The GLM Procedure
Dependent Variable: y
                              Sum of
Source             DF        Squares   Mean Square  F Value  Pr > F
Model               3    30.00000000   10.00000000    13.33  0.0018
Error               8     6.00000000    0.75000000
Corrected Total    11    36.00000000

R-Square     Coeff Var      Root MSE        y Mean
0.833333      12.37179      0.866025      7.000000

Source           DF     Type I SS   Mean Square  F Value  Pr > F
loc               3   30.00000000   10.00000000    13.33  0.0018
Source           DF   Type III SS   Mean Square  F Value  Pr > F
loc               3   30.00000000   10.00000000    13.33  0.0018

Contrast         DF   Contrast SS   Mean Square  F Value  Pr > F
linear            1   29.40000000   29.40000000    39.20  0.0002
quadratic         1    0.00000000    0.00000000     0.00  1.0000
cubic             1    0.60000000    0.60000000     0.80  0.3972

The REG Procedure
                   Analysis of Variance
                             Sum of          Mean
Source             DF       Squares        Square   F Value    Pr > F
Model               1      29.40000      29.40000     44.55    <.0001
Error              10       6.60000       0.66000
Corrected Total    11      36.00000

Root MSE              0.81240    R-Square     0.8167
Dependent Mean        7.00000    Adj R-Sq     0.7983
Coeff Var            11.60577

                        Parameter Estimates
                       Parameter       Standard
Variable      DF        Estimate          Error    t Value    Pr > |t|
Intercept      1         4.90000        0.39243      12.49      <.0001
dist           1         0.14000        0.02098       6.67      <.0001
```


a From the GLM output, the overall $F = 13.33$ is significant at the $\alpha = .05$ level (P-value=.0018). This implies at least one of the distance means is different than the others.

b The contrast statement was used in SAS. The P-value is .0002. There appears to be a linear relationship among the means. Looking at the estimated means, the dissolved oxygen increases as the distance increases.

c. From the REG output, the P-value is $< .0001$. This again suggests a linear relationship.

d. A regression model describes the relationship among means. If $\beta_1 = 0$, all the means would be the same. If we plug in $\mu_i = \mu$ into the contrast, we see the sum equals zero.

e. The remaining contrast SS (quadratic and cubic) have been pooled into SSE. If these SS and 2 df are removed from the error, the tests are identical.

f. SST-SS($\hat{L}$) is simply the SS for the cubic and quadratic contrasts. The F statistic would then be $F = (0.6/2)/0.75 = 0.4$. This is smaller than 1 so it will not be significant.

e. Lack of fit test was described on pages 290-292. The F=$\text{MS}_{\text{LOF}}/\text{MS}_{\text{PE}}$. In this case, that would be $F = (0.6/2)/(6.0/8) = 0.4$. This is the same as above.