Multiple Comparisons

STAT 401 - Statistical Methods for Research Workers

Jarad Niemi

Iowa State University

19 September 2013

SAS code and output for one-way ANOVA

```
DATA mice;
  INFILE 'case0501.csv' DSD FIRSTOBS=2;
  INPUT lifetime diet $;

PROC GLM DATA=mice;
  CLASS diet;
  MODEL lifetime = diet;
  LSMEANS diet / ADJUST=T;
  RUN;
```

```
                                       The GLM Procedure

Dependent Variable: lifetime

                                                  Sum of
       Source                      DF          Squares     Mean Square    F Value    Pr > F
       Model                        5      12733.94181      2546.78836      57.10    <.0001
       Error                      343      15297.41532        44.59888
       Corrected Total            348      28031.35713
```

What have we learned?

SAS code and output for one-way ANOVA

```
                    The GLM Procedure
                   Least Squares Means

                              lifetime      LSMEAN
                diet            LSMEAN      Number

                N/N85       32.6912281           1
                N/R40       45.1166667           2
                N/R50       42.2971831           3
                NP          27.4020408           4
                R/R50       42.8857143           5
                lopro       39.6857143           6


             Least Squares Means for effect diet
             Pr > |t| for H0: LSMean(i)=LSMean(j)

               Dependent Variable: lifetime

  i/j          1           2           3           4           5           6

     1                <.0001      <.0001      <.0001      <.0001      <.0001
     2    <.0001                  0.0166      <.0001      0.0731      <.0001
     3    <.0001      0.0166                  <.0001      0.6223      0.0293
     4    <.0001      <.0001      <.0001                  <.0001      <.0001
     5    <.0001      0.0731      0.6223      <.0001                  0.0117
     6    <.0001      <.0001      0.0293      <.0001      0.0117

NOTE: To ensure overall protection level, only probabilities associated with pre-planned
      comparisons should be used.
```

Multiple comparisons

How many combinations when choosing 2 items out of 6?

$$\binom{6}{2} = \frac{6!}{2!(6-2)!} = \frac{6}{2! \cdot 4!} = \frac{6 \cdot 5}{2} = 15$$

If 0.05 is our significance cutoff, what is the probability of falsely rejecting at least one true null hypothesis?

$$1 - (1 - 0.05)^{15} = 1 - (0.95)^{15} = 1 - 0.46 = 0.54$$

So there is a greater than 50% probability of falsely rejecting a true null hypothesis!

But we wanted the probability of making a mistake to be 0.05.

Multiple comparison adjustments

Confidence interval for the difference between group i and group i':

$$\overline{Y}_i - \overline{Y}_{i'} \pm M \, s_p \sqrt{\frac{1}{n_i} + \frac{1}{n_{i'}}}$$

where M is a multiplier that depends on the adjustment procedure:

Procedure	M	Use
Tukey-Kramer	$q_{\mathrm{I},n-\mathrm{I}}(1-\alpha)/\sqrt{2}$	All pairwise comparisons
Scheffé	$\sqrt{(\mathrm{I}-1)F_{(\mathrm{I}-1,n-\mathrm{I})}(1-\alpha)}$	All contrasts
Dunnett		Compare all groups to control
LSD	$t_{n-\mathrm{I}}(1-\alpha)$	After significant F-test (no adjustment)
Bonferroni	$t_{n-\mathrm{I}}(1-\alpha/2k)$ $k = \mathrm{I}(\mathrm{I}-1)/2$	k tests (most generic)

SAS code and output for one-way ANOVA

```
                    The GLM Procedure
                   Least Squares Means
     Adjustment for Multiple Comparisons: Bonferroni

                           lifetime      LSMEAN
            diet             LSMEAN      Number

            N/N85        32.6912281           1
            N/R40        45.1166667           2
            N/R50        42.2971831           3
            NP           27.4020408           4
            R/R50        42.8857143           5
            lopro        39.6857143           6


          Least Squares Means for effect diet
          Pr > |t| for H0: LSMean(i)=LSMean(j)

              Dependent Variable: lifetime

i/j             1             2             3             4             5             6

   1                     <.0001        <.0001        0.0009        <.0001        <.0001
   2       <.0001                      0.2488        <.0001        1.0000        0.0002
   3       <.0001        0.2488                      <.0001        1.0000        0.4402
   4       0.0009        <.0001        <.0001                      <.0001        <.0001
   5       <.0001        1.0000        1.0000        <.0001                      0.1751
   6       <.0001        0.0002        0.4402        <.0001        0.1751
```

SAS code and output for one-way ANOVA

```
                    The GLM Procedure
                   Least Squares Means
        Adjustment for Multiple Comparisons: Tukey-Kramer

                          lifetime      LSMEAN
            diet            LSMEAN      Number

            N/N85       32.6912281           1
            N/R40       45.1166667           2
            N/R50       42.2971831           3
            NP          27.4020408           4
            R/R50       42.8857143           5
            lopro       39.6857143           6


              Least Squares Means for effect diet
              Pr > |t| for H0: LSMean(i)=LSMean(j)

                  Dependent Variable: lifetime

i/j          1          2          3          4          5          6

   1                <.0001     <.0001     0.0008     <.0001     <.0001
   2    <.0001                 0.1565     <.0001     0.4684     0.0002
   3    <.0001     0.1565                 <.0001     0.9964     0.2460
   4    0.0008     <.0001     <.0001                 <.0001     <.0001
   5    <.0001     0.4684     0.9964     <.0001                 0.1168
   6    <.0001     0.0002     0.2460     <.0001     0.1168
```

What have we learned?

Frankly, we've learned very little.

We've learned that if $\mu_i = \mu_j$ for the following pairs, the data are highly unlikely:

- N/N85-N/R40
- N/N85-N/R50
- N/N85-NP
- N/N85-R/R50
- N/N85-lopro
- N/R40-NP
- N/R40-lopro
- N/R50-NP
- NP-R/R50
- NP-lopro

Is this information useful?

If you wanted the longest lifetime, which diet should you prefer?

Does it really make a difference?