Topic 19 - Inference

STAT 525 - Fall 2013

Outline

- Inference for
  - Means
  - Differences in cell means
  - Contrasts
- Multiplicity

The Cell Means Model

- Expressed numerically

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

  where μ_i is the theoretical mean of all observations at level i (or in cell i)

- The ε_{ij} are iid $N(0, \sigma^2)$ which implies the Y_{ij} are independent $N(\mu_i, \sigma^2)$
- Parameters
  - $\mu_1, \mu_2, ..., \mu_r$
  - σ^2

Estimates

- Estimate μ_i using the sample mean of the observations at level i

$$\hat{\mu}_i = \overline{Y}_{i.}$$

- Pool the sample variances s_i^2 using weights proportional to sample size (i.e., df) to get s^2

$$\begin{aligned} s^2 &= \frac{\sum (n_i - 1)s_i^2}{\sum (n_i - 1)} \\ &= \frac{\sum (n_i - 1)s_i^2}{n_T - r} \end{aligned}$$

Confidence Intervals of μ_i's

- From model

$$\overline{Y}_{i.} \sim N(\mu_i, \sigma^2/n_i)$$

- Confidence interval

$$\overline{Y}_{i.} \pm t(1-\alpha/2; n_T - r)s/\sqrt{n_i}$$

- Degrees of freedom larger than $n_i - 1$ because pooling variance estimates across treatments (i.e., borrowing information from other groups)

SAS Commands

```
data a1;
   infile 'u:\.www\datasets525\CH15TA01.TXT';
   input cases design store;

proc means data=a1 mean std stderr clm maxdec=2;
   class design;
   var cases;

proc glm data=a1;
   class design;
   model cases=design;
   means design/t clm;

proc mixed data=a1;
   class design;
   model cases=design;
   lsmeans design / cl;
```


Output

```
The MEANS Procedure
                   Analysis Variable : cases
                              Lower 95% Upper 95%
Des   N    Mean  StdDev StdErr    CL for Mean
----------------------------------------------
1     5   14.60   2.30   1.03    11.74    17.46
2     5   13.40   3.65   1.63     8.87    17.93
3     4   19.50   2.65   1.32    15.29    23.71
4     5   27.20   3.96   1.77    22.28    32.12
```

Note: $4 \times 2.30^2 + 4 \times 3.65^2 + 3 \times 2.65^2 + 4 \times 3.96^2 = 158.24$. Except for rounding, this is equal to SSE. Also, 19-4=15 which is the df error in the ANOVA table.

There is no pooling of error (or df) when computing these confidence intervals.

Output

```
The GLM Procedure

t Confidence Intervals for cases
Alpha                            0.05
Error Degrees of Freedom           15
Error Mean Square            10.54667
Critical Value of t           2.13145

                                        95% Confidence
design          N           Mean            Limits
4               5         27.200       24.104       30.296
3               4         19.500       16.039       22.961
1               5         14.600       11.504       17.696
2               5         13.400       10.304       16.496
```

These confidence intervals are often narrower due to the increase in degrees of freedom. Results can vary if there does not appear to be a common variance.

Output

```
The Mixed Procedure

Covariance Parameter
      Estimates
Cov Parm     Estimate
Residual      10.5467

                            Least Squares Means
                  Standard
Design  Estimate     Error    DF  t Value  Pr > |t|   Lower     Upper
1        14.6000    1.4524    15    10.05    <.0001 11.5044   17.6956
2        13.4000    1.4524    15     9.23    <.0001 10.3044   16.4956
3        19.5000    1.6238    15    12.01    <.0001 16.0390   22.9610
4        27.2000    1.4524    15    18.73    <.0001 24.1044   30.2956
```

These confidence intervals are the same as the previous page. Standard errors, based on constant variance assumption, are provided.

Multiplicity

- Have generated r confidence intervals
- Overall confidence level (all intervals contain its true mean) is less than $1 - \alpha$
- Many different approaches have been proposed
- Previously discussed using Bonferroni

SAS Commands

```
proc glm data=a1;
   class design;
   model cases=design;
   means design/bon clm;

proc mixed data=a1;
   class design;
   model cases=design;
   lsmeans design /alpha=0.125 cl;
run;
```


Output

```
Bonferroni t Confidence Intervals for cases

Alpha                           0.05
Error Degrees of Freedom          15
Error Mean Square           10.54667
Critical Value of t          2.83663


                                        Simultaneous 95%
design          N           Mean       Confidence Limits
4               5         27.200       23.080      31.320
3               4         19.500       14.894      24.106
1               5         14.600       10.480      18.720
2               5         13.400        9.280      17.520
```

Output

```
The Mixed Procedure

Covariance Parameter
      Estimates
Cov Parm     Estimate
Residual      10.5467

                                  Least Squares Means
               Standard
Design Estimate   Error  DF t Value  Pr > |t|  Alpha     Lower     Upper
1       14.6000  1.4524  15   10.05    <.0001  0.0125   10.4802   18.7198
2       13.4000  1.4524  15    9.23    <.0001  0.0125    9.2802   17.5198
3       19.5000  1.6238  15   12.01    <.0001  0.0125   14.8939   24.1061
4       27.2000  1.4524  15   18.73    <.0001  0.0125   23.0802   31.3198
```


Hypothesis Tests on μ_i's

- Not usually done
- SAS typically gives output for $H_0 : \mu_i = 0$ which rarely is of any interest
- If interested in $H_0 : \mu_i = c$, it is easiest to subtract of c from all observations in a data step and then test whether the new mean is equal to zero.
- Can also use CI to make decision

Differences in means

- From model

$$\overline{Y}_{i.} - \overline{Y}_{k.} \sim N\left(\mu_i - \mu_k, \sigma^2\left(\frac{1}{n_i} + \frac{1}{n_k}\right)\right)$$

- Confidence interval

$$\overline{Y}_{i.} - \overline{Y}_{k.} \pm t(1-\alpha/2; n_T - r)s\sqrt{1/n_i + 1/n_k}$$

- In this case $H_0 : \mu_i - \mu_k = 0$ is of interest
- Similar multiplicity problem
- Now have $\frac{r(r-1)}{2}$ pairwise comparisons to consider

Multiplicity Adjustment

- Approaches adjust multiplier of the SE
  - Alter α level (e.g., Bonferroni)
  - Use different distribution
- Conservative → strong control of overall Type I error - avoids false positives
- Powerful → able to pick up differences that exist - avoids false negatives
- All approaches try to strike to strike some sort of balance

Least Significant Difference

- Simply ignores multiplicity issue
- Most powerful of the procedures but also results in most false positives
- Uses $t(1-\alpha/2; n_T - r)$ to determine multiplier
- Called T or LSD in SAS

Tukey

- Based on studentized range distribution q
- Range is $\max(\overline{Y}_i) - \min(\overline{Y}_i)$ in r levels
- Accounts for any possible pair being furthest apart
- Controls overall experimentwise error rate
- Uses $q(1-\alpha; r, n_T - r)/\sqrt{2}$ to determine multiplier
- Called TUKEY in SAS

Scheffe$'$

- Based on the F distribution
- Accounts for multiplicity for <u>all</u> linear combinations of means, not just pairwise comparisons
- Protects against data snooping
- Uses $\sqrt{(r-1)F(1-\alpha; r-1, n_T - r)}$ to determine multiplier
- Called SCHEFFE in SAS

Bonferroni

- Replaces α by

$$\alpha^* = \frac{\alpha}{r(r-1)/2}$$

- Uses $t(1-\alpha^*/2; n_T - r)$ to determine multiplier
- Called BON in SAS

Holm

- Refinement of Bonferroni
- Instead of using

$$\alpha^* = \frac{\alpha}{g}$$

for all comparisons

- Rank unadjusted P-values from smallest to largest
- Continue to reject until $P_k \geq \alpha/(g-k+1)$
- Available in Proc Multtest in SAS

False Discovery Rate

- FDR defined as expected proportion of false positives in the collection of rejected null hypotheses
- Becoming more popular, especially when # of tests in the thousands or millions
- Rank P-values from smallest to largest
- Continue to reject until $P_k \geq k\alpha/g$
- Available in Proc Multtest in SAS

SAS Commands

```
proc glm data=a1;
   class design;
   model cases=design;
   means design/lsd tukey bon scheffe;
   means design/lines tukey;
run;

proc mixed data=a1;
   class design;
   model cases=design;
   lsmeans design / diff=all; lsmeans design / adjust=tukey;
   lsmeans design / adjust=bon; lsmeans design / adjust=scheffe;
run;

proc glimmix data=a1;
   class design;
   model cases=design;
   lsmeans design / adjust=tukey lines;
run;
```


Output

```
t Tests (LSD) for cases
NOTE: This test controls the Type I comparisonwise error rate,
not the experimentwise error rate.

Alpha                          0.05
Error Degrees of Freedom         15
Error Mean Square          10.54667
Critical Value of t         2.13145

Comparisons significant at the 0.05 level are indicated by ***.

               Difference
  design        Between      95% Confidence
Comparison       Means           Limits
4    - 3         7.700       3.057   12.343  ***
4    - 1        12.600       8.222   16.978  ***
4    - 2        13.800       9.422   18.178  ***
3    - 4        -7.700     -12.343   -3.057  ***
3    - 1         4.900       0.257    9.543  ***
3    - 2         6.100       1.457   10.743  ***
1    - 4       -12.600     -16.978   -8.222  ***
1    - 3        -4.900      -9.543   -0.257  ***
1    - 2         1.200      -3.178    5.578
2    - 4       -13.800     -18.178   -9.422  ***
2    - 3        -6.100     -10.743   -1.457  ***
2    - 1        -1.200      -5.578    3.178
```

Output

```
Tukey's Studentized Range (HSD) Test for cases
NOTE: This test controls the Type I experimentwise error rate.

Alpha                                  0.05
Error Degrees of Freedom                 15
Error Mean Square                  10.54667
Critical Value of Studentized Range  4.07597

Comparisons significant at the 0.05 level are
indicated by ***.

               Difference
  design          Between     Simultaneous 95%
Comparison          Means    Confidence Limits
4    - 3            7.700       1.421   13.979  ***
4    - 1           12.600       6.680   18.520  ***
4    - 2           13.800       7.880   19.720  ***
3    - 4           -7.700     -13.979   -1.421  ***
3    - 1            4.900      -1.379   11.179
3    - 2            6.100      -0.179   12.379
1    - 4          -12.600     -18.520   -6.680  ***
1    - 3           -4.900     -11.179    1.379
1    - 2            1.200      -4.720    7.120
2    - 4          -13.800     -19.720   -7.880  ***
2    - 3           -6.100     -12.379    0.179
2    - 1           -1.200      -7.120    4.720
```


Output

```
Bonferroni (Dunn) t Tests for cases
NOTE: This test controls the Type I experimentwise error rate, but it
generally has a higher Type II error rate than Tukey's for all
pairwise comparisons.

Alpha                        0.05
Error Degrees of Freedom       15
Error Mean Square        10.54667
Critical Value of t       3.03628

Comparisons significant at the 0.05 level are indicated by ***.

               Difference
  design          Between     Simultaneous 95%
Comparison          Means    Confidence Limits
4    - 3            7.700       1.085   14.315  ***
4    - 1           12.600       6.364   18.836  ***
4    - 2           13.800       7.564   20.036  ***
3    - 4           -7.700     -14.315   -1.085  ***
3    - 1            4.900      -1.715   11.515
3    - 2            6.100      -0.515   12.715
1    - 4          -12.600     -18.836   -6.364  ***
1    - 3           -4.900     -11.515    1.715
1    - 2            1.200      -5.036    7.436
2    - 4          -13.800     -20.036   -7.564  ***
2    - 3           -6.100     -12.715    0.515
2    - 1           -1.200      -7.436    5.036
```


Output

```
Scheffe's Test for cases
NOTE: This test controls the Type I experimentwise error rate, but it
generally has a higher Type II error rate than Tukey's for all
pairwise comparisons.

Alpha                        0.05
Error Degrees of Freedom       15
Error Mean Square        10.54667
Critical Value of F       3.28738

Comparisons significant at the 0.05 level are indicated by ***.

               Difference
  design          Between     Simultaneous 95%
Comparison          Means    Confidence Limits
4    - 3            7.700       0.859   14.541  ***
4    - 1           12.600       6.150   19.050  ***
4    - 2           13.800       7.350   20.250  ***
3    - 4           -7.700     -14.541   -0.859  ***
3    - 1            4.900      -1.941   11.741
3    - 2            6.100      -0.741   12.941
1    - 4          -12.600     -19.050   -6.150  ***
1    - 3           -4.900     -11.741    1.941
1    - 2            1.200      -5.250    7.650
2    - 4          -13.800     -20.250   -7.350  ***
2    - 3           -6.100     -12.941    0.741
2    - 1           -1.200      -7.650    5.250
```


Output

```
Tukey's Studentized Range (HSD) Test for cases

Alpha                                  0.05
Error Degrees of Freedom                 15
Error Mean Square                  10.54667
Critical Value of Studentized Range  4.07597
Minimum Significant Difference       6.1019
Harmonic Mean of Cell Sizes        4.705882

NOTE: Cell sizes are not equal.

Means with the same letter are not significantly different.

          Mean      N    design

A       27.200      5    4

B       19.500      4    3
B
B       14.600      5    1
B
B       13.400      5    2
```

Mixed Output

Effect	design	_design	Estimate	Standard Error	DF	t Value	Pr > |t|	Adjustment	Adj P
design	1	2	1.2000	2.0539	15	0.58	0.5677		.
design	1	3	-4.9000	2.1785	15	-2.25	0.0399		.
design	1	4	-12.6000	2.0539	15	-6.13	<.0001		.
design	2	3	-6.1000	2.1785	15	-2.80	0.0135		.
design	2	4	-13.8000	2.0539	15	-6.72	<.0001		.
design	3	4	-7.7000	2.1785	15	-3.53	0.0030		.
design	1	2	1.2000	2.0539	15	0.58	0.5677	Tukey-Kramer	0.9353
design	1	3	-4.9000	2.1785	15	-2.25	0.0399	Tukey-Kramer	0.1549
design	1	4	-12.6000	2.0539	15	-6.13	<.0001	Tukey-Kramer	0.0001
design	2	3	-6.1000	2.1785	15	-2.80	0.0135	Tukey-Kramer	0.0583
design	2	4	-13.8000	2.0539	15	-6.72	<.0001	Tukey-Kramer	<.0001
design	3	4	-7.7000	2.1785	15	-3.53	0.0030	Tukey-Kramer	0.0142
design	1	2	1.2000	2.0539	15	0.58	0.5677	Bonferroni	1.0000
design	1	3	-4.9000	2.1785	15	-2.25	0.0399	Bonferroni	0.2397
design	1	4	-12.6000	2.0539	15	-6.13	<.0001	Bonferroni	0.0001
design	2	3	-6.1000	2.1785	15	-2.80	0.0135	Bonferroni	0.0808
design	2	4	-13.8000	2.0539	15	-6.72	<.0001	Bonferroni	<.0001
design	3	4	-7.7000	2.1785	15	-3.53	0.0030	Bonferroni	0.0180
design	1	2	1.2000	2.0539	15	0.58	0.5677	Scheffe	0.9507
design	1	3	-4.9000	2.1785	15	-2.25	0.0399	Scheffe	0.2125
design	1	4	-12.6000	2.0539	15	-6.13	<.0001	Scheffe	0.0002
design	2	3	-6.1000	2.1785	15	-2.80	0.0135	Scheffe	0.0895
design	2	4	-13.8000	2.0539	15	-6.72	<.0001	Scheffe	<.0001
design	3	4	-7.7000	2.1785	15	-3.53	0.0030	Scheffe	0.0248

Glimmix Output

Differences of design Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

design	_design	Estimate	Standard Error	DF	t Value	Pr > |t|	Adj P
1	2	1.2000	2.0539	15	0.58	0.5677	0.9353
1	3	-4.9000	2.1785	15	-2.25	0.0399	0.1549
1	4	-12.6000	2.0539	15	-6.13	<.0001	0.0001
2	3	-6.1000	2.1785	15	-2.80	0.0135	0.0583
2	4	-13.8000	2.0539	15	-6.72	<.0001	<.0001
3	4	-7.7000	2.1785	15	-3.53	0.0030	0.0142

Tukey-Kramer Grouping for design Least Squares Means (Alpha=0.05)
LS-means with the same letter are not significantly different.

design	Estimate	
4	27.2000	A
3	19.5000	B
		B
1	14.6000	B
		B
2	13.4000	B

SAS Commands

```
proc multtest data=a1 holm fdr out=new noprint;
 class design;
 contrast '12' 1 -1 0 0;
 contrast '13' 1 0 -1 0;
 contrast '14' 1 0 0 -1;
 contrast '23' 0 1 -1 0;
 contrast '24' 0 1 0 -1;
 contrast '34' 0 0 1 -1;
 test mean(cases);
run;

proc print data=new;
run;
```


Output

Obs	_test_	_var_	_contrast_	_value_	_se_	_nval_	raw_p	stpbon_p	fdr_p
1	MEAN	cases	12	22.8	39.0248	15	0.56774	0.56774	0.56774
2	MEAN	cases	13	-93.1	41.3921	15	0.03995	0.07990	0.04794
3	MEAN	cases	14	-239.4	39.0248	15	0.00002	0.00010	0.00006
4	MEAN	cases	23	-115.9	41.3921	15	0.01346	0.04038	0.02019
5	MEAN	cases	24	-262.2	39.0248	15	0.00001	0.00004	0.00004
6	MEAN	cases	34	-146.3	41.3921	15	0.00300	0.01201	0.00601

Instead of comparing each raw P-value to a different α level, the P-values are adjusted based on the procedure.

This approach works for experiments with a small number of levels. Can also input a set of P-values and perform the analysis.

Linear Combination of Means

- Would like to test $H_0 : L = \sum c_i\mu_i = L_0$
- Hypotheses usually planned but can be "after the fact"

Can use statistical model to construct t-test

$$\widehat{L} = \sum c_i\overline{Y}_{i.} \qquad \begin{aligned} \mathrm{Var}(\widehat{L}) &= \mathrm{Var}(\textstyle\sum c_i\overline{Y}_{i.}) \\ &= \textstyle\sum c_i^2\mathrm{Var}(\overline{\mathrm{Y}}_{\mathrm{i.}}) \\ &= \mathrm{MSE}\textstyle\sum(\mathrm{c_i^2/n_i}) \end{aligned}$$

$$t^* = \frac{\widehat{L}-L_0}{\sqrt{\mathrm{Var}(\widehat{\mathrm{L}})}}$$

Under H_0: $t^* \sim t_{n_T-r}$

Contrasts

- Special case of linear combination
- Requires $\sum c_i = 0$
- Example 1: $\mu_1 - \mu_2 = 0$
- Example 2: $\mu_1 - (\mu_2 + \mu_3)/2 = 0$
- Example 3: $(\mu_1 + \mu_2) - (\mu_3 + \mu_4) = 0$

SAS Commands

```
proc glm data=a1;
   class design;
   model cases=design;
   contrast '1&2 v 3&4' design .5 .5 -.5 -.5;
   estimate '1&2 v 3&4' design .5 .5 -.5 -.5;
run;

*Joint test of several contrasts;
proc glm data=a1;
   class design;
   model cases=design;
   contrast '1 v 2&3&4' design 1 -.3333 -.3333 -.3333;
   estimate '1 v 2&3&4' design 3 -1 -1 -1 /divisor=3;
   contrast '2 v 3 v 4' design 0 1 -1  0,
                        design 0 0  1 -1;
```


Output

```
Contrast    DF  Contrast SS   Mean Square  F Value  Pr > F
1&2 v 3&4    1  411.4000000   411.4000000    39.01  <.0001

                             Standard
Parameter      Estimate         Error  t Value    Pr > |t|
1&2 v 3&4   -9.35000000  1.49705266    -6.25      <.0001
```

Contrast does an F test while Estimate does a t-test and gives an estimate of the linear combination. Contrast allows you to simultaneously test a collection of contrasts.

```
Contrast    DF   Contrast SS   Mean Square  F Value  Pr > F
1 v 2&3&4    1   108.4739502  108.4739502     10.29  0.0059
2 v 3 v 4    2   477.9285714  238.9642857     22.66  <.0001
```

Background Reading

- KNNL Sections 17.1-17.8
- knnl738.sas
- KNNL Chapter 18