STAT 525

# Chapter 17
# Analysis of Factor Level Means

Dr. Qifan Song

# The Cell Means Model

- Expressed numerically

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

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

- The $\varepsilon_{ij}$ are iid $N(0, \sigma^2)$ which implies the $Y_{ij}$ are independent $N(\mu_i, \sigma^2)$

- Parameters
  - $\mu_1, \mu_2, ..., \mu_r$
  - $\sigma^2$

# Estimates

- Estimate $\mu_i$ by the sample mean of the observations at level $i$

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

- Pool $s_i^2$ 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 $\mu_i$'s

- From model

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

- (sub-optimal) Confidence interval (one-sample $t$ CI)

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

- 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)

# Example (Page 685)

```
options nocenter;
data a1;
    infile 'u:\.www\datasets525\CH16TA01.TXT';
    input cases design store;

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

```
                Analysis Variable : cases

                                          Lower 95%    Upper 95%
Des  N    Mean  Std Dev  Std Err  CL for Mean  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
```

- $4 \times 2.30^2 + 4 \times 3.65^2 + 3 \times 2.65^2 + 4 \times 3.96^2 = 158.24$. Except for rounding, this is equal to SSE.
- 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.

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

```
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
```

- Very important for there to be a common variance.
  - These confidence intervals are often narrower due to the increase of degrees of freedom.

# Multiplicity

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

# Example (Page 685)

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

```
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
```

## Hypothesis Tests on $\mu_i$'s

- Not usually done
- SAS often 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.

## 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

# Multiplicity Adjustment

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

## Least Significant Difference

- Simply ignores multiplicity issue
- Uses $t(1-\alpha/2; n_T - r)$ to determine multiplier
- Called `T` or `LSD` in SAS

## Tukey Multiple Comparison Procedure

- Based on studentized range distribution $q$
- Range refers to $max(\overline{Y}_{i.}) - min(\overline{Y}_{i.})$ 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

# Scheffé Multiple Comparison Procedure

- Based on the $F$ distribution
- Accounts for multiplicity for all 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 Multiple Comparison Procedure

- Replaces $\alpha$ by

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

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

## Holm Multiple Comparison Procedure

- Refinement of Bonferroni
- Instead of using

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

  for all comparisons

- Rank P-values from smallest to largest

$$p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(g)}$$

- 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
- Rank P-values from smallest to largest

$$p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(g)}$$

- Continue to reject until $p_{(k)} \geq k\alpha/g$
- Available in `PROC MULTTEST` in SAS

# Others: Dunnett's and Hsu's procedure

- Dunnett's procedure: compare $r-1$ levels against one pre-determined control level (say the first level).

$$\overline{Y}_{i.} - \overline{Y}_{1.} \pm D_{2-sided,crit} s \sqrt{1/n_i + 1/n_k}$$

- SAS option `dunnett('control')`

- Hsu's MCB: Multiple Comparisons with the Best *sample mean* level

- If best means largest: all groups with group mean satisfying

$$\overline{Y}_{i.} \geq \max_j \overline{Y}_{j.} - D_{1-sided,crit} s \sqrt{1/n_i + 1/n_k}$$

  are included in the set G, and the probability that G contains the real best level is $1-\alpha$.

- SAS option `dunnettl('max level')` or `dunnettu('min level')`

- Not recommended for severely unbalanced data

## Example (Page 685)

```
/* Compare all pairs */
proc glm data=a1;
   class design;
   model cases=design;
   means design/lsd tukey bon scheffe cldiff;
run; quit;
```

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 ***.

| design Comparison | Difference Between Means | 95% Confidence 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 | |

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 ***.

| design Comparison | Difference Between Means | Simultaneous 95% 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 | |

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 ***.

| design Comparison | Difference Between Means | Simultaneous 95% 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 | |

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 ***.

| design Comparison | Difference Between Means | Simultaneous 95% 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 | |

```
/* LINES: group factor levels based on test results */
proc glm data=a1;
   class design;
   model cases=design;
   means design/lines tukey;
run;
```

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

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
```

# 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

$$
\begin{aligned}
& \widehat{L} = \sum c_i \overline{Y}_{i.} \\
& Var(\widehat{L}) = Var(\sum c_i \overline{Y}_{i.}) = \sum c_i^2 \text{Var}(\overline{Y}_{i.}) \\
\Longrightarrow \quad & \widehat{Var}(\widehat{L}) = \text{MSE} \sum (c_i^2/n_i) \\
\Longrightarrow \quad & t^* = \frac{\widehat{L} - L_0}{\sqrt{\widehat{Var}(\widehat{L})}}
\end{aligned}
$$

- 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: $(\mu_1 + \mu_2) - (\mu_3 + \mu_4) = 0$

# Example (Page 685)

```
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; quit;
```

```
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` actually performs a general linear $F$ test, hence allows you to simultaneously test a collection of contrast

```
/* 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;
run; quit;
```

| 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 |

| Parameter | Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|
| 1 v 2&3&4 | -5.43333333 | 1.69441348 | -3.21 | 0.0059 |

```
/* HOLM & FDR */
proc multtest data=a1 holm fdr out=out1 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);

proc print data=out1; quit;
```

| _contrast_ | _value_ | _se_ | _nval_ | raw_p | stpbon_p | fdr_p |
|---|---|---|---|---|---|---|
| 12 | 22.8 | 39.0248 | 15 | 0.56774 | 0.56774 | 0.56774 |
| 13 | -93.1 | 41.3921 | 15 | 0.03995 | 0.07990 | 0.04794 |
| 14 | -239.4 | 39.0248 | 15 | 0.00002 | 0.00010 | 0.00006 |
| 23 | -115.9 | 41.3921 | 15 | 0.01346 | 0.04038 | 0.02019 |
| 24 | -262.2 | 39.0248 | 15 | 0.00001 | 0.00004 | 0.00004 |
| 34 | -146.3 | 41.3921 | 15 | 0.00300 | 0.01201 | 0.00601 |

- Instead of comparing each raw $p$-value to a different $\alpha$ level, the $p$-values are adjusted based on the procedure.

```
/* HOLM & FDR */
proc multtest data=a1 holm fdr out=out2 noprint;
    class design;
    contrast '12vs34' 1 1 -1 -1;
    contrast '13vs24' 1 -1 1 -1;
    contrast '12' 1 -1 0 0;
    contrast '13' 1 0 -1 0;
    contrast '24' 0 1 0 -1;
    contrast '34' 0 0 1 -1;
    test mean(cases);

proc print data=out2; run; quit;
```

| _contrast_ | _value_ | _se_ | _nval_ | raw_p | stpbon_p | fdr_p |
|---|---|---|---|---|---|---|
| 12vs34 | -355.3 | 56.8880 | 15 | 0.00002 | 0.00008 | 0.00005 |
| 13vs24 | -123.5 | 56.8880 | 15 | 0.04639 | 0.11984 | 0.05567 |
| 12 | 22.8 | 39.0248 | 15 | 0.56774 | 0.56774 | 0.56774 |
| 13 | -93.1 | 41.3921 | 15 | 0.03995 | 0.11984 | 0.05567 |
| 24 | -262.2 | 39.0248 | 15 | 0.00001 | 0.00004 | 0.00004 |
| 34 | -146.3 | 41.3921 | 15 | 0.00300 | 0.01201 | 0.00601 |

- Instead of comparing each raw $p$-value to a different $\alpha$ level, the $p$-values are adjusted based on the procedure.
  - stpbon_p$_{(k)}$ = raw_p$_{(k)}$ $\times (g - k + 1)$
  - fdr_p$_{(k)}$ = raw_p$_{(k)}$ $\times g/k$

# Chapter Review

- Inference for
  - Means
  - Differences in cell means
  - Contrasts
- Multiplicity
  - The Tukey procedure can be modified to handle general contrasts of factor level means
  - If only pairwise comparisons are to be made, the Tukey procedure gives narrower confidence limits and is therefore the preferred method than the Sheffé procedure
  - For general contrasts of factor level means, the Sheffé procedure is preferred than the Tukey procedure
  - Bonferroni, Sheffé and Tukey procedures are of the form "estimator $\pm$ multiplier $\times$ SE" with different multipliers. For any given problem, one may select the one with the smallest multiplier