One-Way Analysis of Variance

Nathaniel E. Helwig

Assistant Professor of Psychology and Statistics
University of Minnesota (Twin Cities)

Updated 04-Jan-2017

Copyright

Outline of Notes

1) Categorical Predictors:

- Overview
- Dummy coding
- Effect coding

2) One-Way ANOVA:

- Model form & assumptions
- Estimation & basic inference
- Memory example (part 1)

3) Multiple Comparisons:

- Comparing means
- Bonferroni correction
- Tukey correction
- Scheffé correction
- Summary of corrections
- Memory example (part 2)

Categorical Predictors

Categorical Variables (revisited)

Suppose that $X \in \{x_1, \ldots, x_g\}$ is a categorical variable with g levels.

- Categorical variables are also called factors in ANOVA context
- Example: sex $\in \{\text{female}, \text{male}\}$ has two levels
- Example: drug $\in \{\text{A}, \text{B}, \text{C}\}$ has three levels

To code a categorical variable (with g levels) in a regression model, we need to include $g - 1$ different variables in the model.

- If we know $\mu = \frac{1}{g} \sum_{j=1}^{g} \mu_j$, where μ_j is mean for j-th factor level
- Then we know $\mu_g = g(\mu - \frac{1}{g} \sum_{j=1}^{g-1} \mu_j)$ by definition
- Only g total free parameters, so cannot estimate μ and $\{\mu_j\}_{j=1}^{g}$

Dummy Coding: Definition

Dummy coding uses $g - 1$ binary variables to code a factor:

$$x_{ij} = \begin{cases} 1 & \text{if } i\text{-th observation is in } j\text{-th level} \\ 0 & \text{otherwise} \end{cases}$$

for $i \in \{1, \ldots, n_j\}$ and $j \in \{1, \ldots, g-1\}$.

Regression model becomes

$$y_{ij} = b_0 + \sum_{j=1}^{g-1} b_j x_{ij} + e_{ij}$$

where $b_0 = \mu_g$ and $b_j = \mu_j - \mu_g$ for $j \in \{1, \ldots, g-1\}$.

Dummy Coding: Considerations

Dummy coding is useful for…

- One-way ANOVA model (unique parameter for each factor level)
- Comparing treatment groups to clearly defined “control” group

Dummy coding is less useful when…

- Have $g > 2$ levels and/or model is more complicated
- Do NOT have a clearly defined “control” or “reference” group

Dummy Coding: R Syntax

The `contrasts` function controls the coding scheme for a factor.

Use the `contr.treatment` option for dummy coding.

```
> x = factor(rep(letters[1:3], each=5))
> x
 [1] a a a a a b b b b b c c c c c
Levels: a b c
> contrasts(x) <- contr.treatment(nlevels(x))
> contrasts(x)
  2 3
a 0 0
b 1 0
c 0 1
```

Effect Coding: Definition

Effect coding also uses $g - 1$ variables to code a factor:

$$x_{ij} = \begin{cases} 1 & \text{if } i\text{-th observation is in } j\text{-th level} \\ -1 & \text{if } i\text{-th observation is in } g\text{-th level} \\ 0 & \text{otherwise} \end{cases}$$

for $i \in \{1, \ldots, n_j\}$ and $j \in \{1, \ldots, g - 1\}$.

Regression model becomes

$$y_{ij} = b_0 + \sum_{j=1}^{g-1} b_j x_{ij} + e_{ij}$$

where $b_0 = \mu$ and $b_j = \mu_j - \mu$ for $j \in \{1, \ldots, g\}$; note $b_g = -\sum_{j=1}^{g-1} b_j$.

Effect Coding: Considerations

Effect coding is useful for...

- Simple interpretation of b_0 as overall mean
- Comparing each group's effect to overall mean

Effect coding is less useful when...

- Have $g = 2$ levels for a factor
- Do have a clearly defined "control" or "reference" group

Effect Coding: R Syntax

The `contrasts` function controls the coding scheme for a factor.

Use the `contr.sum` option for effect (deviation) coding.

```
> x = factor(rep(letters[1:3],each=5))
> x
 [1] a a a a a b b b b b c c c c c
Levels: a b c
> contrasts(x) <- contr.sum(nlevels(x))
> contrasts(x)
  [,1] [,2]
a    1    0
b    0    1
c   -1   -1
```

One-Way ANOVA Model

One-Way ANOVA Model (cell means form)

The One-Way Analysis of Variance (ANOVA) model has the form

$$y_{ij} = \mu_j + e_{ij}$$

for $i \in \{1, \ldots, n_j\}$ and $j \in \{1, \ldots, g\}$ where

- $y_{ij} \in \mathbb{R}$ is real-valued response for i-th subject in j-th factor level
- $\mu_j \in \mathbb{R}$ is real-valued population mean for the j-th factor level
- $e_{ij} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$ is a Gaussian error term
- n_j is number of subjects in j-th factor level and $n = \sum_{j=1}^{g} n_j$
- g is number of factor levels

Implies that $y_{ij} \overset{\text{ind}}{\sim} \mathrm{N}(\mu_j, \sigma^2)$.

One-Way ANOVA Model (dummy coding)

Using dummy coding, the one-way ANOVA becomes

$$y_{ij} = b_0 + \sum_{j=1}^{g-1} b_j x_{ij} + e_{ij}$$

for $i \in \{1, \ldots, n_j\}$ and $j \in \{1, \ldots, g\}$ where

- $x_{ij} = \begin{cases} 1 & \text{if } i\text{-th observation is in } j\text{-th level} \\ 0 & \text{otherwise} \end{cases}$
- $b_0 = \mu_g$ is reference group mean
- $b_j = \mu_j - \mu_g$ for $j \in \{1, \ldots, g-1\}$

One-Way ANOVA Model (effect coding)

Using effect coding, the one-way ANOVA becomes

$$y_{ij} = b_0 + \sum_{j=1}^{g-1} b_j x_{ij} + e_{ij}$$

for $i \in \{1, \ldots, n_j\}$ and $j \in \{1, \ldots, g\}$ where

- $x_{ij} = \begin{cases} 1 & \text{if } i\text{-th observation is in } j\text{-th level} \\ -1 & \text{if } i\text{-th observation is in } g\text{-th level} \\ 0 & \text{otherwise} \end{cases}$
- $b_0 = \mu$ is overall mean
- $b_j = \mu_j - \mu$ for $j \in \{1, \ldots, g\}$
- Note that $b_g = -\sum_{j=1}^{g-1} b_j$ by definition

One-Way ANOVA Model (matrix form)

In matrix form, the one-way ANOVA model is

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1(g-1)} \\ 1 & x_{21} & x_{22} & \cdots & x_{2(g-1)} \\ 1 & x_{31} & x_{32} & \cdots & x_{3(g-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{n(g-1)} \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_{g-1} \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_n \end{pmatrix}$$

where

- definition of x_{ij} and $\{b_j\}_{j=1}^{g-1}$ will depend on coding scheme
- $i \in \{1, \ldots, n\}$ and second subscript on y and e is dropped

Implies that $\mathbf{y} \sim \mathrm{N}(\mathbf{X}\mathbf{b}, \sigma^2 \mathbf{I}_n)$.

One-Way ANOVA Model (assumptions)

The fundamental assumptions of the one-way ANOVA model are:

1. x_{ij} and y_i are observed random variables (known constants)
2. $e_i \stackrel{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$ is an unobserved random variable
3. $b_0, b_1, \ldots, b_{g-1}$ are unknown constants
4. $(y_i | x_{i1}, \ldots, x_{i(g-1)}) \stackrel{\text{ind}}{\sim} \mathrm{N}(b_0 + \sum_{j=1}^{g-1} b_j x_{ij}, \sigma^2)$
   note: homogeneity of variance

Interpretation of b_j depends on coding scheme

- Dummy: b_j is difference between j-th mean and reference mean
- Effect: b_j is difference between j-th mean and overall mean

Ordinary Least Squares (cell means form)

We want to find the factor level mean estimates (i.e., $\hat{\mu}_j$ terms) that minimize the ordinary least squares criterion

$$SSE = \sum_{j=1}^{g} \sum_{i=1}^{n_j} (y_{ij} - \mu_j)^2$$

The least-squares estimates are the factor level means

$$\hat{\mu}_j = \bar{y}_j$$

where $\bar{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}$ is sample mean of Y for j-th factor level.

Ordinary Least Squares (simple proof)

Note that we want to minimize

$$(SSE)_j = \sum_{i=1}^{n_j}(y_{ij} - \mu_j)^2 = \sum_{i=1}^{n_j} y_{ij}^2 - 2\mu_j \sum_{i=1}^{n_j} y_{ij} + n_j\mu_j^2$$

separately for each $j \in \{1, \ldots, g\}$.

Taking the derivative with respect to μ_j we have

$$\frac{\mathrm{d}(SSE)_j}{\mathrm{d}\mu_j} = -2\sum_{i=1}^{n_j} y_{ij} + 2n_j\mu_j$$

and setting to zero and solving for μ_j gives $\hat{\mu}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} y_{ij} = \bar{y}_j$

Ordinary Least Squares (general case)

In general, we can use the regression approach

$$SSE = \sum_{i=1}^{n} \left(y_i - b_0 - \sum_{j=1}^{g-1} b_j x_{ij} \right)^2 = \|\mathbf{y} - \mathbf{X}\mathbf{b}\|^2$$

where $i \in \{1, \ldots, n\}$ and $n = \sum_{j=1}^{g} n_j$; note that the second subscript on Y is now dropped because there is only one summation.

The OLS solution has the form

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

which is the same as the MLR model; note that ANOVA is MLR with categorical predictors!

Fitted Values and Residuals

SCALAR FORM:

Fitted values are given by

$$\hat{y}_i = \hat{b}_0 + \sum_{j=1}^{g-1} \hat{b}_j x_{ij}$$

and residuals are given by

$$\hat{e}_i = y_i - \hat{y}_i$$

MATRIX FORM:

Fitted values are given by

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{b}}$$

and residuals are given by

$$\hat{\mathbf{e}} = \mathbf{y} - \hat{\mathbf{y}}$$

ANOVA Sums-of-Squares: Scalar Form

In one-way ANOVA model, the relevant sums-of-squares are

- Total: $SST = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{j=1}^{g}\sum_{i=1}^{n_j}(y_{ij} - \bar{y})^2$
- Between: $SSB = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = \sum_{j=1}^{g} n_j(\bar{y}_j - \bar{y})^2$
- Within: $SSW = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{j=1}^{g}\sum_{i=1}^{n_j}(y_{ij} - \bar{y}_j)^2$

The corresponding degrees of freedom are

- SST: $df_T = n - 1$
- SSB: $df_B = g - 1$
- SSW: $df_W = n - g$

ANOVA Sums-of-Squares: Matrix Form

In MLR models, the relevant sums-of-squares are

$$\begin{aligned} SST &= \sum_{i=1}^{n} (y_i - \bar{y})^2 \\ &= \mathbf{y}' \left[\mathbf{I}_n - (1/n)\mathbf{J} \right] \mathbf{y} \\ SSB &= \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \\ &= \mathbf{y}' \left[\mathbf{H} - (1/n)\mathbf{J} \right] \mathbf{y} \\ SSW &= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \\ &= \mathbf{y}' \left[\mathbf{I}_n - \mathbf{H} \right] \mathbf{y} \end{aligned}$$

Note: $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{J}$ is an $n \times n$ matrix of ones

Partitioning the Variance (same as MLR model)

We can partition the total variation in y_i as

$$\begin{aligned}
SST &= \sum_{i=1}^{n}(y_i - \bar{y})^2 \\
&= \sum_{i=1}^{n}(y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 \\
&= \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + 2\sum_{i=1}^{n}(\hat{y}_i - \bar{y})(y_i - \hat{y}_i) \\
&= SSB + SSW + 2\sum_{i=1}^{n}(\hat{y}_i - \bar{y})\hat{e}_i \\
&= SSB + SSW
\end{aligned}$$

See MLR notes for the proof.

Estimated Error Variance (Mean Squared Error)

An unbiased estimate of the error variance σ^2 is

$$\begin{aligned} \hat{\sigma}^2 &= SSW/(n-g) \\ &= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2/(n-g) \\ &= \frac{1}{n-g}\sum_{j=1}^{g}(n_j - 1)s_j^2 \end{aligned}$$

where $s_j^2 = \frac{1}{n_j - 1}\sum_{i=1}^{n_j}(y_{ij} - \bar{y}_j)^2$ is j-th factor level's sample variance.

The estimate $\hat{\sigma}^2$ is the mean squared error (MSE) of the model, and is the pooled estimate of sample variance.

ANOVA Table and Overall F Test

We typically organize the SS information into an ANOVA table:

Source	SS	df	MS	F	p-value
SSB	$\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$	$g-1$	MSB	F^*	p^*
SSW	$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	$n-g$	MSW		
SST	$\sum_{i=1}^{n}(y_i - \bar{y})^2$	$n-1$			

$MSB = \frac{SSB}{g-1}$, $MSW = \frac{SSW}{n-g}$, $F^* = \frac{MSB}{MSW} \sim F_{g-1,n-g}$,
$p^* = P(F_{g-1,n-g} > F^*)$

F^*-statistic and p^*-value are testing $H_0 : b_1 = \cdots = b_{g-1} = 0$ versus $H_1 : b_k \neq 0$ for some $k \in \{1, \ldots, g-1\}$

- Equivalent to testing $H_0 : \mu_j = \mu \; \forall j$ versus H_1 : not all μ_j are equal

Memory Example: Data Description

Visual and auditory cues example from Hays (1994) Statistics.

- Does lack of visual/auditory synchrony affect memory?

Total of $n = 30$ college students participate in memory experiment.

- Watch video of person reciting 50 words
- Try to remember the 50 words (record number correct)

Randomly assign $n_j = 10$ subjects to one of $g = 3$ video conditions:

- `fast`: sound precedes lip movements in video
- `normal`: sound synced with lip movements in video
- `slow`: lip movements in video precede sound

Memory Example: Descriptive Statistics

Number of correctly remembered words (y_{ij}):

Subject (i)	Fast ($j = 1$)	Normal ($j = 2$)	Slow ($j = 3$)
1	23	27	23
2	22	28	24
3	18	33	21
4	15	19	25
5	29	25	19
6	30	29	24
7	23	36	22
8	16	30	17
9	19	26	20
10	17	21	23
$\sum_{i=1}^{10} y_{ij}$	212	274	218
$\sum_{i=1}^{10} y_{ij}^2$	4738	7742	4810

Memory Example: OLS Estimation (by hand)

The least-squares estimates of μ_j are the sample means:

$$\hat{\mu}_1 = \bar{y}_1 = \frac{1}{10}\sum_{i=1}^{10} y_{i1} = 212/10 = 21.2$$

$$\hat{\mu}_2 = \bar{y}_2 = \frac{1}{10}\sum_{i=1}^{10} y_{i2} = 274/10 = 27.4$$

$$\hat{\mu}_3 = \bar{y}_3 = \frac{1}{10}\sum_{i=1}^{10} y_{i3} = 218/10 = 21.8$$

Memory Example: OLS Estimation (in R: by hand)

```
# define response and factor vectors
> sync = c(23,27,23,22,28,24,18,33,21,15,
           19,25,29,25,19,30,29,24,23,36,
           22,16,30,17,19,26,20,17,21,23)
> cond = factor(rep(c("fast","normal","slow"), 10))


# sum of sync for each level of cond
> tapply(sync, cond, sum)
  fast normal   slow
   212    274    218


# sum-of-squares of sync for each level of cond
> sumsq = function(x){ sum(x^2) }
> tapply(sync, cond, sumsq)
  fast normal   slow
  4738   7742   4810
```

Memory Example: OLS Estimation (in R: dummy pt. 1)

```
> smod = lm(sync ~ cond)
> summary(smod)$coef
            Estimate Std. Error    t value     Pr(>|t|)
(Intercept)     21.2   1.408440 15.0521126 1.183308e-14
condnormal       6.2   1.991835  3.1127073 4.350803e-03
condslow         0.6   1.991835  0.3012297 7.655470e-01
```

Note that...

- $\hat{b}_0 = \bar{y}_1 = 21.2$ is mean of `fast` condition
- $\hat{b}_1 = \bar{y}_2 - \bar{y}_1 = 27.4 - 21.2 = 6.2$ is difference between means of `normal` and `fast` conditions
- $\hat{b}_2 = \bar{y}_3 - \bar{y}_1 = 21.8 - 21.2 = 0.6$ is difference between means of `slow` and `fast` conditions

Memory Example: OLS Estimation (in R: dummy pt. 2)

```
> contrasts(cond)
       normal slow
fast        0    0
normal      1    0
slow        0    1
> contrasts(cond) <- contr.treatment(3, base=2)
> contrasts(cond)
       1 3
fast   1 0
normal 0 0
slow   0 1
> smod = lm(sync ~ cond)
> summary(smod)$coef
            Estimate Std. Error   t value     Pr(>|t|)
(Intercept)     27.4   1.408440 19.454146 2.052094e-17
cond1           -6.2   1.991835 -3.112707 4.350803e-03
cond3           -5.6   1.991835 -2.811478 9.072381e-03
```

Note that...

- $\hat{b}_0 = \bar{y}_2 = 27.4$ is mean of `normal` condition
- $\hat{b}_1 = \bar{y}_1 - \bar{y}_2 = 21.2 - 27.4 = -6.2$ is difference between means of `fast` and `normal` conditions
- $\hat{b}_2 = \bar{y}_3 - \bar{y}_2 = 21.8 - 27.4 = -5.6$ is difference between means of `slow` and `normal` conditions

Memory Example: OLS Estimation (in R: effect)

```
> contrasts(cond) <- contr.sum(3)
> contrasts(cond)
       [,1] [,2]
fast      1    0
normal    0    1
slow     -1   -1
> smod = lm(sync ~ cond)
> summary(smod)$coef
             Estimate Std. Error   t value     Pr(>|t|)
(Intercept) 23.466667  0.8131633 28.858492 7.900265e-22
cond1       -2.266667  1.1499866 -1.971037 5.904989e-02
cond2        3.933333  1.1499866  3.420330 2.003601e-03
```

Note that…

- $\hat{b}_0 = \bar{y} = 23.47$ is grand mean ($\bar{y} = \frac{212+274+218}{30} = \frac{704}{30} = 23.467$)
- $\hat{b}_1 = \bar{y}_1 - \bar{y} = 21.2 - 23.467 = -2.266667$ is difference between mean of `fast` condition and overall mean
- $\hat{b}_2 = \bar{y}_2 - \bar{y} = 27.4 - 23.467 = 3.933333$ is difference between mean of `normal` condition and overall mean
- Implicitly we have: $\hat{b}_3 = -(\hat{b}_1 + \hat{b}_2) = -(3.933333 - 2.266667) = \bar{y}_3 - \bar{y} = 21.8 - 23.467 = -1.67$ is difference between mean of `slow` condition and overall mean

Memory Example: Sums-of-Squares (by hand)

Defining $n = \sum_{j=1}^{g} n_j = 30$, the relevant sums-of-squares are

$$SST = \sum_{j=1}^{g} \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2 = \sum_{j=1}^{g} \sum_{i=1}^{n_j} y_{ij}^2 - \frac{1}{n} \left(\sum_{j=1}^{g} \sum_{i=1}^{n_j} y_{ij} \right)^2$$
$$= (4738 + 7742 + 4810) - (704^2/30) = 769.4667$$

$$SSW = \sum_{j=1}^{g} \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2 = \sum_{j=1}^{g} \sum_{i=1}^{n_j} y_{ij}^2 - \sum_{j=1}^{g} \frac{\left(\sum_{i=1}^{n_j} y_{ij} \right)^2}{n_j}$$
$$= (4738 + 7742 + 4810) - ([212^2 + 274^2 + 218^2]/10) = 535.6$$

$$SSB = SST - SSW = 769.4667 - 535.6 = 233.8667$$

Memory Example: ANOVA Table (by hand)

Putting sums-of-squares on previous slide into an ANOVA table:

Source	SS	df	MS	F	p-value
SSB	233.8667	2	116.9333	5.8947	0.0075
SSW	535.6000	27	19.8370		
SST	769.4667	29			

Note that $F^* = 5.8947 \sim F_{2,27}$ and $P(F_{2,27} > F^*) = 0.0075$.

Assuming a typical α level (e.g., $\alpha = 0.01$ or $\alpha = 0.05$), we would reject the null hypothesis $H_0 : \mu_j = \mu \; \forall j$.

We conclude that there is some mean difference on the response variable (# of remembered words) between the different conditions.

Memory Example: ANOVA Table (in R)

```
> sync = c(23,27,23,22,28,24,18,33,21,15,
           19,25,29,25,19,30,29,24,23,36,
           22,16,30,17,19,26,20,17,21,23)
> cond = factor(rep(c("fast","normal","slow"), 10))
> smod = lm(sync ~ cond)
> anova(smod)
Analysis of Variance Table

Response: sync
          Df Sum Sq Mean Sq F value   Pr(>F)
cond       2 233.87 116.933  5.8947 0.007513 **
Residuals 27 535.60  19.837
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Multiple Comparisons

Limitations of Overall F Test

If we reject $H_0 : \mu_j = \mu \; \forall j$ then we know that there are some mean differences, but we do not know where the mean differences exist

- This is assuming that $g > 2$
- Note that if $g = 2$ we have an independent samples t test

If $g > 2$ we need to perform followup tests (multiple comparisons) to determine where the mean differences are occurring in the data.

Linear Combinations of Factor Level Means

A linear combination L of the factor level means has the form

$$L = \sum_{j=1}^{g} c_j \mu_j$$

where c_j are the coefficients defining the particular linear combination.

In practice we never know μ_j so we define

$$\hat{L} = \sum_{j=1}^{g} c_j \hat{\mu}_j$$

where $\hat{\mu}_j$ is our least-squares estimate of μ_j.

Testing Linear Combinations

Remember $\hat{\mu}_j = \bar{y}_j$ which implies

$$V(\hat{\mu}_j) = V\left(\frac{1}{n_j}\sum_{i=1}^{n_j} y_{ij}\right) = \frac{1}{n_j^2}V\left(\sum_{i=1}^{n_j} y_{ij}\right) = \frac{\sum_{i=1}^{n_j} V(y_{ij})}{n_j^2} = \frac{n_j\sigma^2}{n_j^2} = \frac{\sigma^2}{n_j}$$

is the variance of $\hat{\mu}_j$, given that $V\left(y_{ij}\right) = \sigma^2$ from the assumptions.

This implies that

$$V(\hat{L}) = V\left(\sum_{j=1}^{g} c_j\hat{\mu}_j\right) = \sum_{j=1}^{g} c_j^2 V(\hat{\mu}_j) = \sigma^2\sum_{j=1}^{g} c_j^2/n_j$$

is the variance of any linear combination $\hat{L}$.

Testing Linear Combinations (continued)

To test $H_0 : L = L^*$ versus $H_1 : L \neq L^*$ use

$$t^* = \frac{\hat{L} - L^*}{\sqrt{\hat{V}(\hat{L})}} \sim t_{n-g}$$

where $\hat{V}(\hat{L}) = \hat{\sigma}^2 \sum_{j=1}^{g} c_j^2 / n_j$ uses the MSE to estimate σ^2.

If population σ^2 is known, then use

$$Z^* = \frac{\hat{L} - L^*}{\sqrt{V(\hat{L})}} \sim \mathrm{N}(0, 1)$$

Contrasts and Pairwise Comparisons

A contrast is a linear combination of the factor level means such that the coefficients sum to zero, i.e., $\sum_{j=1}^{g} c_j = 0$.

- $\mu_1 - \mu_2$ is mean difference of first two levels
- $(\mu_1 + \mu_2)/2 - \mu_3$ is mean of first two levels minus third level

A pairwise comparison is a contrast involving two factor level means:

- $\mu_1 - \mu_2$ is a pairwise comparison of first two levels
- $\mu_1 - \mu_3$ is a pairwise comparison of first and third levels
- $\mu_2 - \mu_3$ is a pairwise comparison of second and third levels

Multiple Comparison Problem

If we test multiple linear combinations of factor level means, we need to worry about the Familywise Type I Error Rate (FWER).

FWER is probability of making at least one Type I Error among all tested linear combinations.

- Single test Type I Error $= P(\text{Reject } H_0 \mid H_0 \text{ true}) = \alpha$
- For q independent tests with level α: $FWER = 1 - (1 - \alpha)^q$

Generally FWER will depend on number of tests and whether or not tests are independent of one another.

Bonferroni's Correction: Definition

Suppose we want to test f linear combinations of factor level means.

According to Boole's inequality, for f tests with level α^*

$$FWER \leq \sum_{k=1}^{f} P(\text{Reject } H_{0k} \mid H_{0k} \text{ true}) = \sum_{k=1}^{f} \alpha^* = f\alpha^*$$

regardless of whether or not the tests are independent of one another.

Bonferroni's correction sets $\alpha^* = \alpha/f$ to ensure that $FWER \leq \alpha$.

Bonferroni's Correction: Properties

Major strength: applicable to many situations (no assumptions)

Major weakness: overly conservative in some cases

Suppose we have $f = 3$ independent tests and want $FWER \leq 0.05$

- $FWER = 1 - (1 - \alpha^*)^3$
- Bonferroni: $\alpha^* = 0.05/3 = 0.0167$
  $FWER = 1 - (1 - 0.0167)^3 = 0.04917$

Suppose we have $f = 10$ independent tests and want $FWER \leq 0.05$

- $FWER = 1 - (1 - \alpha^*)^{10}$
- Bonferroni: $\alpha^* = 0.05/10 = 0.005$
  $FWER = 1 - (1 - 0.005)^{10} = 0.0489$

Detour: Studentized Range Distribution

Assume the following...

- $z_k \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2)$ for $k \in \{1, \ldots, r\}$
- $\hat{\sigma}^2$ is an estimate of σ^2 based on ν degrees-of-freedom
- $\hat{\sigma}^2$ is independent of z_k for $k \in \{1, \ldots, r\}$

The studentized range statistic is defined as

$$q_{r,\nu} = \frac{\text{range}(z_k)}{\hat{\sigma}}$$

where $\text{range}(z_k) = \max(z_k) - \min(z_k)$, and (r, ν) are the numerator and denominator degrees of freedom.

One-sided probability distribution (similar to F), so only reject if observed $q > q_{r,\nu}^{(\alpha)}$ where $P(q_{r,\nu} > q_{r,\nu}^{(\alpha)}) = \alpha$.

Testing All Possible Pairwise Comparisons

Want to test all possible pairwise comparisons between means.

- $H_0 : \mu_j - \mu_k = 0 \; \forall j, k$ versus H_1 : not all $\mu_j - \mu_k = 0$
- There are $g(g-1)/2$ unique pairwise comparisons

Note that the variance of a pairwise comparison $\hat{L} = \mu_j - \mu_k$ is

$$V(\hat{L}) = \sigma^2 \left(\frac{1}{n_j} + \frac{1}{n_k} \right)$$

where n_j and n_k are the sample sizes of factor levels j and k.

If $n_j = n_* \forall j$, simplifies to $V(\hat{L}) = \sigma^2(\frac{2}{n_*})$ for any pairwise comparison.

Tukey's Honest Significant Difference (HSD) Test

Proposed by John Tukey (1953) for balanced ANOVA, i.e., $n_j = n_* \forall j$.

Test statistic is defined as

$$q^* = \frac{\sqrt{2}\hat{L}}{\sqrt{\hat{V}(\hat{L})}} = \frac{\bar{y}_j - \bar{y}_k}{\sqrt{\hat{\sigma}^2/n_*}}$$

where $\hat{L} = \bar{y}_j - \bar{y}_k$, $\hat{V}(\hat{L}) = \hat{\sigma}^2(\frac{2}{n_*})$, and $\hat{\sigma}^2$ is MSE of model.

Considering all pairwise comparisons, $q^* \sim q_{g,n-g}$ where $q_{g,n-g}$ is studentized range distribution with $(g, n-g)$ degrees of freedom.

- Note that $\bar{y}_j \sim \mathrm{N}(\mu_j, \sigma^2/n_*)$ for all $j \in \{1, \ldots, g\}$
- Under $H_0 : \mu_j - \mu_k = 0 \; \forall j, k$, we have $\bar{y}_j \sim \mathrm{N}(\mu, \sigma^2/n_*) \; \forall j$

Tukey's HSD Test (continued)

To form a $100(1-\alpha)\%$ CI around the mean difference $\bar{y}_j - \bar{y}_k$ use

$$(\bar{y}_j - \bar{y}_k) \pm \frac{q_{g,n-g}^{(\alpha)}}{\sqrt{2}} \sqrt{\hat{V}(\hat{L})}$$

where $q_{g,n-g}^{(\alpha)}$ is critical value from studentized range distribution.

If you form all possible CIs around pairwise mean differences, you will control FWER exactly at level α using Tukey's HSD test.

- More conservative than forming all CIs using t_{n-g} critical values
- Example 95% CI: $q_{3,27}^{(0.95)}/\sqrt{2} = 2.48$ and $t_{27}^{(0.975)} = 2.05$

Tukey-Kramer Test

In unbalanced ANOVA, i.e., $n_j \neq n_k$ for at least one (j,k), use HSD extension proposed by John Tukey (1953) and Clyde Kramer (1956)

- Called the Tukey-Kramer test or Tukey-Kramer procedure

To form a $100(1-\alpha)\%$ CI around the mean difference $\bar{y}_j - \bar{y}_k$ use

$$(\bar{y}_j - \bar{y}_k) \pm \frac{q_{g,n-g}^{(\alpha)}}{\sqrt{2}} \sqrt{\hat{V}(\hat{L})}$$

where $\hat{V}(\hat{L}) = \hat{\sigma}^2(\frac{1}{n_j} + \frac{1}{n_k})$ is estimated variance of $\hat{L} = \hat{\mu}_j - \hat{\mu}_k$.

If you form all possible CIs around pairwise mean differences, you will control FWER below (but not exactly at) level α using Tukey-Kramer.

- See Hayter (1984) for formal proof of TK conservativeness

Testing All Possible Contrasts

Want to test all possible contrasts between factor level means

- $H_0 : \sum_{j=1}^{g} c_j \mu_j = 0$ for all $\mathbf{c} \in \mathcal{C}$ where
  $\mathcal{C} = \{\mathbf{c} = (c_1, \ldots, c_g) : \sum_{j=1}^{g} c_j = 0\}$ is set of all contrasts
- $H_1 : \sum_{j=1}^{g} c_j \mu_j \neq 0$ for some $\mathbf{c} \in \mathcal{C}$

Note that $\mu_j = \mu \ \forall j \iff \sum_{j=1}^{g} c_j \mu_j = 0$ for all $\mathbf{c} \in \mathcal{C}$

Proof of $\mu_j = \mu \ \forall j \implies \sum_{j=1}^{g} c_j \mu_j = 0$ for all $\mathbf{c} \in \mathcal{C}$

- If $\mu_j = \mu \ \forall j$, then $\sum_{j=1}^{g} c_j \mu_j = \mu \sum_{j=1}^{g} c_j = \mu(0) = 0$ for all $\mathbf{c} \in \mathcal{C}$

Proof of $\sum_{j=1}^{g} c_j \mu_j = 0$ for all $\mathbf{c} \in \mathcal{C} \implies \mu_j = \mu \ \forall j$

- If $\sum_{j=1}^{g} c_j \mu_j = 0$ for all $\mathbf{c} \in \mathcal{C}$, then $\mu_j - \mu_k = 0$ for all j, k

Contrasts and Overall ANOVA F Test

Remember the overall F test (associated with ANOVA table) is testing $H_0 : \mu_j = \mu \; \forall j$ versus H_1 : not all μ_j are equal

- Equivalent to testing $H_0 : \sum_{j=1}^{g} c_j \mu_j = 0$ for all $\mathbf{c} \in \mathcal{C}$ versus $H_1 : \sum_{j=1}^{g} c_j \mu_j \neq 0$ for some $\mathbf{c} \in \mathcal{C}$
- See previous slide for proof of equivalence

Point: if we want to test all possible contrasts, we can use $F_{g-1,n-g}$ distribution to control FWER at level α.

Scheffé's Method

Want to test all possible contrasts between factor level means

- $H_0 : \sum_{j=1}^{g} c_j \mu_j = 0$ for all $\mathbf{c} \in \mathcal{C}$ where $\mathcal{C} = \{\mathbf{c} = (c_1, \ldots, c_g) : \sum_{j=1}^{g} c_j = 0\}$ is set of all contrasts
- $H_1 : \sum_{j=1}^{g} c_j \mu_j \neq 0$ for some $\mathbf{c} \in \mathcal{C}$

To form a $100(1-\alpha)\%$ CI for a contrast $L = \sum_{j=1}^{g} c_j \mu_j$ use

$$\hat{L} \pm \sqrt{(g-1)F_{g-1,n-g}^{(\alpha)}}\sqrt{\hat{V}(\hat{L})}$$

where $\hat{V}(\hat{L}) = \hat{\sigma}^2 \sum_{j=1}^{g} c_j^2 / n_j$ is estimated variance of $\hat{L} = \sum_{j=1}^{g} c_j \bar{y}_j$.

If you form CIs around all possible contrasts, you will control FWER exactly at level α using Scheffé's method. Proof

Summary of Multiple Comparisons

If we want to form $100(1-\alpha)\%$ CI around all $\hat{L} = \hat{\mu}_j - \hat{\mu}_k$ use

$$\hat{L} \pm C\sqrt{\hat{V}(\hat{L})}$$

where $\hat{V}(\hat{L}) = \hat{\sigma}^2(\frac{1}{n_j} + \frac{1}{n_k})$ and C is some critical value.

Each procedure uses different critical value:

No correction: $C = t_{n-g}^{(\alpha/2)}$

Bonferroni: $C = t_{n-g}^{(\alpha^*/2)}$ with $\alpha^* = \alpha/[g(g-1)/2]$

Tukey: $C = q_{g,n-g}^{(\alpha)}/\sqrt{2}$

Scheffé: $C = \sqrt{(g-1)F_{g-1,n-g}^{(\alpha)}}$

Scheffé's critical value is ALWAYS larger than Tukey value, because set of all pairwise comparisons is subset of set of all contrasts.

Choosing Between Multiple Comparisons

You want CIs that are as narrow as possible and control FWER.

If you are interested in all pairwise comparisons, use Tukey-Kramer.

If you are interested in all possible contrasts, use Scheffé.

If you are interested in some subset of all pairwise comparisons (or all contrasts), Bonferroni may be most efficient approach.

Memory Example: Pairwise Comparison Estimates

Suppose we want to test all $g(g-1)/2 = 3$ unique pairwise comparisons between factor level means:

- $L_1 = \mu_1 - \mu_2 = \mu_{\texttt{fast}} - \mu_{\texttt{normal}}$
- $L_2 = \mu_2 - \mu_3 = \mu_{\texttt{normal}} - \mu_{\texttt{slow}}$
- $L_3 = \mu_3 - \mu_1 = \mu_{\texttt{slow}} - \mu_{\texttt{fast}}$

The estimated pairwise comparisons are given by

- $\hat{L}_1 = \hat{\mu}_1 - \hat{\mu}_2 = 21.2 - 27.4 = -6.2$
- $\hat{L}_2 = \hat{\mu}_2 - \hat{\mu}_3 = 27.4 - 21.8 = 5.6$
- $\hat{L}_3 = \hat{\mu}_3 - \hat{\mu}_1 = 21.8 - 21.2 = 0.6$

and we know that $\hat{V}(\hat{L}) = \hat{\sigma}^2(2/n_*) = (19.8370)(2/10) = 3.9674$

Memory Example: Pairwise Comparison CI Values

If we want to form 95% CI around all three $\hat{L} = \hat{\mu}_j - \hat{\mu}_k$ use

$$\hat{L} \pm C\sqrt{\hat{V}(\hat{L})} = \hat{L} \pm C\sqrt{3.9674}$$

where

- $C = t_{27}^{(.025)} = 2.0518$ with no correction
- $C = t_{27}^{(.008)} = 2.5525$ with Bonferroni correction
- $C = \frac{q_{3,27}^{(.05)}}{\sqrt{2}} = 2.4794$ with Tukey correction
- $C = \sqrt{2F_{2,27}^{(.05)}} = 2.5900$ with Scheffé correction

Note that Tukey is best (i.e., produces narrowest CIs), followed by Bonferroni, and then Scheffé.

Memory Example: Pairwise CIs (no correction)

Using no correction the CI estimates are:

$$\hat{L}_1 \pm t_{27}^{(.025)} \sqrt{\hat{V}(\hat{L}_1)} = -6.2 \pm 2.0518\sqrt{3.9674} = [-10.2869;\ -2.1131]$$

$$\hat{L}_2 \pm t_{27}^{(.025)} \sqrt{\hat{V}(\hat{L}_2)} = 5.6 \pm 2.0518\sqrt{3.9674} = [1.5131;\ 9.6869]$$

$$\hat{L}_3 \pm t_{27}^{(.025)} \sqrt{\hat{V}(\hat{L}_3)} = 0.6 \pm 2.0518\sqrt{3.9674} = [-3.4869;\ 4.6869]$$

Memory Example: Pairwise CIs (Bonferroni)

Using Bonferroni correction the CI estimates are:

$$\hat{L}_1 \pm t_{27}^{(.008)} \sqrt{\hat{V}(\hat{L}_1)} = -6.2 \pm 2.5525\sqrt{3.9674} = [-11.2841; \ -1.1159]$$

$$\hat{L}_2 \pm t_{27}^{(.008)} \sqrt{\hat{V}(\hat{L}_2)} = 5.6 \pm 2.5525\sqrt{3.9674} = [0.5159; \ 10.6841]$$

$$\hat{L}_3 \pm t_{27}^{(.008)} \sqrt{\hat{V}(\hat{L}_3)} = 0.6 \pm 2.5525\sqrt{3.9674} = [-4.4841; \ 5.6841]$$

Memory Example: Pairwise CIs (Tukey)

Using Tukey correction the CI estimates are:

$$\hat{L}_1 \pm \frac{q_{3,27}^{(.05)}}{\sqrt{2}}\sqrt{\hat{V}(\hat{L}_1)} = -6.2 \pm 2.4794\sqrt{3.9674} = [-11.1386;\ -1.2614]$$

$$\hat{L}_2 \pm \frac{q_{3,27}^{(.05)}}{\sqrt{2}}\sqrt{\hat{V}(\hat{L}_2)} = 5.6 \pm 2.4794\sqrt{3.9674} = [0.6614;\ 10.5386]$$

$$\hat{L}_3 \pm \frac{q_{3,27}^{(.05)}}{\sqrt{2}}\sqrt{\hat{V}(\hat{L}_3)} = 0.6 \pm 2.4794\sqrt{3.9674} = [-4.3386;\ 5.5386]$$

Memory Example: Pairwise CIs (Scheffé)

Using Scheffé correction the CI estimates are:

$$\hat{L}_1 \pm \sqrt{2F_{2,27}^{(.05)}}\sqrt{\hat{V}(\hat{L}_1)} = -6.2 \pm 2.59\sqrt{3.9674} = [-11.3589;\ -1.0411]$$

$$\hat{L}_2 \pm \sqrt{2F_{2,27}^{(.05)}}\sqrt{\hat{V}(\hat{L}_2)} = 5.6 \pm 2.59\sqrt{3.9674} = [0.4411;\ 10.7589]$$

$$\hat{L}_3 \pm \sqrt{2F_{2,27}^{(.05)}}\sqrt{\hat{V}(\hat{L}_3)} = 0.6 \pm 2.59\sqrt{3.9674} = [-4.5589;\ 5.7589]$$

Memory Example: Good use of Scheffé

Suppose we want to test four contrasts:

- $L_1 = \mu_1 - \mu_2 = \mu_{\texttt{fast}} - \mu_{\texttt{normal}}$
- $L_2 = \mu_2 - \mu_3 = \mu_{\texttt{normal}} - \mu_{\texttt{slow}}$
- $L_3 = \mu_3 - \mu_1 = \mu_{\texttt{slow}} - \mu_{\texttt{fast}}$
- $L_4 = \mu_2 - \frac{\mu_1+\mu_3}{2} = \mu_{\texttt{normal}} - \frac{\mu_{\texttt{slow}}+\mu_{\texttt{fast}}}{2}$

If we want to form 95% CI around all four $\hat{L}_j$ use

$$\hat{L}_j \pm C\sqrt{\hat{V}(\hat{L}_j)}$$

where

- $C = t_{27}^{(.006)} = 2.6763$ using Bonferroni ($\alpha^* = .05/4 = .0125$)
- $C = \sqrt{2F_{2,27}^{(.05)}} = 2.59$ using Scheffé

Memory Example: Good use of Scheffé (continued)

Note that $\hat{L}_4 = 27.4 - \frac{21.8+21.2}{2} = 5.9$ and

$$\hat{V}(\hat{L}_4) = \hat{\sigma}^2 \sum_{j=1}^{3} \frac{c_j^2}{n_j} = (19.8370)\left(\frac{1}{10} + \frac{(-1/2)^2}{10} + \frac{(-1/2)^2}{10}\right) = 2.9756$$

Using Scheffé correction the CI estimates are:

$$\hat{L}_1 \pm \sqrt{2F_{2,27}^{(.05)}}\sqrt{\hat{V}(\hat{L}_1)} = -6.2 \pm 2.59\sqrt{3.9674} = [-11.3589;\ -1.0411]$$

$$\hat{L}_2 \pm \sqrt{2F_{2,27}^{(.05)}}\sqrt{\hat{V}(\hat{L}_2)} = 5.6 \pm 2.59\sqrt{3.9674} = [0.4411;\ 10.7589]$$

$$\hat{L}_3 \pm \sqrt{2F_{2,27}^{(.05)}}\sqrt{\hat{V}(\hat{L}_3)} = 0.6 \pm 2.59\sqrt{3.9674} = [-4.5589;\ 5.7589]$$

$$\hat{L}_4 \pm \sqrt{2F_{2,27}^{(.05)}}\sqrt{\hat{V}(\hat{L}_4)} = 5.9 \pm 2.59\sqrt{2.9756} = [1.4322;\ 10.3678]$$

Appendix

Scheffé's Method: Logic

Remember that the overall ANOVA F test has the form

$$F^* = \frac{MSB}{MSW} = \frac{\frac{1}{g-1}\sum_{j=1}^{g} n_j(\bar{y}_j - \bar{y})^2}{\hat{\sigma}^2} \sim F_{g-1,n-g} \quad \text{under } H_0$$

which implies that

$$S^2 = \frac{SSB}{MSW} = \frac{\sum_{j=1}^{g} n_j(\bar{y}_j - \bar{y})^2}{\hat{\sigma}^2} \sim (g-1)F_{g-1,n-g} \quad \text{under } H_0$$

Defining the test of a single contrast as $T_{\mathbf{c}} = \frac{\hat{L}}{\sqrt{\hat{V}(\hat{L})}}$, note that

$$\sup_{\mathbf{c}\in\mathcal{C}} T_{\mathbf{c}}^2 = S^2$$

where sup denotes the supremum (i.e., least upper-bound).

Scheffé's Method: Proof (part 1)

To prove the claim $\sup_{\mathbf{c}\in\mathcal{C}} T_{\mathbf{c}}^2 = S^2$, define the $n \times 1$ vector

$$\mathbf{a}' = \begin{pmatrix} \frac{c_1}{n_1}\mathbf{1}'_{n_1} & \frac{c_2}{n_2}\mathbf{1}'_{n_2} & \cdots & \frac{c_g}{n_g}\mathbf{1}'_{n_g} \end{pmatrix}$$

where c_j are contrast coefficients and $\mathbf{1}_{n_j}$ is an $n_j \times 1$ vector of ones.

Define $\mathbf{y}' = (\mathbf{y}'_1, \ldots, \mathbf{y}'_g)$ where $\mathbf{y}'_j = (y_{1j}, \ldots, y_{n_j j})$ and note that

$$\mathbf{a}'\mathbf{y} = \sum_{j=1}^{g} \frac{c_j}{n_j}\mathbf{1}'_{n_j}\mathbf{y}_j = \sum_{j=1}^{g} c_j \bar{y}_j = \hat{L}$$

$$\|\mathbf{a}\|^2 = \mathbf{a}'\mathbf{a} = \sum_{j=1}^{g} \left(\frac{c_j}{n_j}\right)^2 \mathbf{1}'_{n_j}\mathbf{1}_{n_j} = \sum_{j=1}^{g} c_j^2/n_j$$

which implies that

$$T_{\mathbf{c}}^2 = \frac{(\mathbf{a}'\mathbf{y})^2}{\hat{\sigma}^2\|\mathbf{a}\|^2}$$

for any contrast $\mathbf{c} \in \mathcal{C}$.

Scheffé's Method: Proof (part 2)

Now note that $\mathbf{a} = \mathbf{X}\tilde{\mathbf{b}}$ where

$$\mathbf{X} = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \cdots & \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \cdots & \mathbf{0}_{n_2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{n_g} & \mathbf{0}_{n_g} & \cdots & \mathbf{1}_{n_g} \end{pmatrix} \qquad \tilde{\mathbf{b}} = \begin{pmatrix} c_1/n_1 \\ c_2/n_2 \\ \vdots \\ c_g/n_g \end{pmatrix}$$

which implies that $\mathbf{a} = \mathbf{H}\mathbf{a}$ where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is hat matrix.

Also note that $\mathbf{a}'\mathbf{1}_n = \sum_{j=1}^{g} \frac{c_j}{n_j}\mathbf{1}_{n_j}'\mathbf{1}_{n_j} = \sum_{j=1}^{g} c_j = 0$, which implies that

$$\mathbf{a} = (\mathbf{H} - \mathbf{H}_0)\mathbf{a}$$

where $\mathbf{H}_0 = \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'$ is projection matrix for constant space.

Scheffé's Method: Proof (part 3)

Putting things together, we can write the single contrast test as

$$T_{\mathbf{c}}^2 = \frac{(\mathbf{a}'\mathbf{y})^2}{\hat{\sigma}^2 \|\mathbf{a}\|^2} = \frac{[\mathbf{a}'(\mathbf{H} - \mathbf{H}_0)\mathbf{y}]^2}{\hat{\sigma}^2 \|\mathbf{a}\|^2}$$

for any contrast $\mathbf{c} \in \mathcal{C}$, given that $\mathbf{a}' = \mathbf{a}'(\mathbf{H} - \mathbf{H}_0)$.

By the Cauchy-Schwarz inequality, we know that

$$(\mathbf{u}'\mathbf{v})^2 \leq \|\mathbf{u}\|^2 \|\mathbf{v}\|^2$$

with equality holding only when $\mathbf{u} = w\mathbf{v}$ for some $w \neq 0$.

Scheffé's Method: Proof (part 4)

Letting $\mathbf{a} = \mathbf{u}$ and $\mathbf{v} = (\mathbf{H} - \mathbf{H}_0)\mathbf{y}$, we have

$$T_{\mathbf{c}}^2 = \frac{[\mathbf{a}'(\mathbf{H} - \mathbf{H}_0)\mathbf{y}]^2}{\hat{\sigma}^2\|\mathbf{a}\|^2}$$
$$\leq \frac{\|\mathbf{a}\|^2\|(\mathbf{H} - \mathbf{H}_0)\mathbf{y}\|^2}{\hat{\sigma}^2\|\mathbf{a}\|^2} = \frac{\|(\mathbf{H} - \mathbf{H}_0)\mathbf{y}\|^2}{\hat{\sigma}^2} = S^2$$

If we define $\mathbf{a} = (\mathbf{H} - \mathbf{H}_0)\mathbf{y}$, then $T_{\mathbf{c}}^2$ reaches its upper bound of S^2.

To control FWER at level α note that

$$P(T_{\mathbf{c}}^2 \leq S^2 \; \forall \mathbf{c} \in \mathcal{C}) = P(\sup_{\mathbf{c} \in \mathcal{C}} T_{\mathbf{c}}^2 \leq S^2)$$

and we know that $S^2 \sim (g-1)F_{g-1,n-g}$ under H_0.

- Use $S = \sqrt{(g-1)F_{g-1,n-g}^{(\alpha)}}$ to form $100(1-\alpha)\%$ CI for $T_{\mathbf{c}}$

Return