

# HANDOUT # 4

## Comparison of Treatment Means

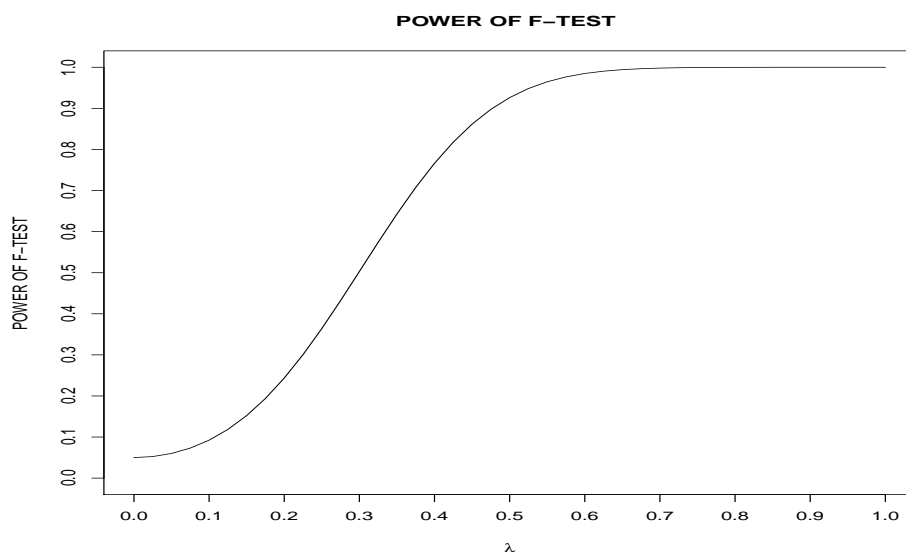
1. The Next Steps After Rejecting  $H_o : \mu_1 = \cdots = \mu_t$
2. Contrasts
3. Decomposition of Treatment Sum of Squares
4. Example: Stopping Distance
5. Multiple Testing
6. Bonferroni Simultaneous Inferences
7. Scheffé Simultaneous Inferences and Confidence Intervals
8. Multiple Comparisons Amongst All Treatment Means  
Fisher's Protected LSD, Tukey's HSD, SNK
9. Comparing Treatment Means to a Control:  
Dunnett's Procedure
10. Finding a Group of *Best* Treatments: Hsu's Procedure
11. Response Curves for Treatments with Quantitative Levels
12. Summary of the Various Procedures

## Post-hoc Analyses for Determining Differences in Treatment Means

If the F-Test fails to reject  $H_o : \mu_1 = \cdots = \mu_t$ , then we state that there is not sufficient evidence in the data to support the research hypothesis,  $H_1$  with the chance of a Type II error in the decision given by

$$\beta(\lambda) = P[\text{Type II error}] = 1 - \gamma(\lambda) \quad \text{for} \quad \lambda = \frac{\sum_{i=1}^t n_i (\mu_i - \bar{\mu})^2}{\sigma_e^2}.$$

However, in nearly all cases we do not know the value of  $\lambda$  because it is a function of the unknown treatment means:  $\mu_1, \mu_2, \cdots, \mu_t$ . By failing to reject  $H_o$ , there is not much further we can say about the treatment means except to provide the power curve of the ANOVA F-test to demonstrate possible values for Type II errors:



If the F-test rejects  $H_o : \mu_1 = \cdots = \mu_t$ , then we know that

$$P[\text{Type I error}] = \gamma(0) = \alpha.$$

Thus, we have a high degree of certainty in our decision but now our conclusion, reject  $H_o : \mu_1 = \cdots = \mu_t$  and state that there is significant evidence in the data that differences exist in the treatment means, leaves uncertainty in what type of differences may exist in the treatment means:

1. Are all  $t$  different means different?
2. Are there groups of means for which there are not significant differences?
3. If the treatment levels are quantitative, are there specific trends in the means?

These are the types of questions we will attempt to answer in this handout.

## Describing differences in Treatment Means:

For this handout we will impose the following model and then obtain results concerning specified differences in the treatment means.

Model:

$$y_{ij} = \mu_i + e_{ij}; \quad \text{for } i = 1, \dots, t \quad \text{and } j = 1, \dots, n_i,$$

where  $\mu_i = E[Y_{ij}]$  is the  $i$ th treatment mean and

$e_{ij}$  are iid  $N(0, \sigma_e^2)$  random variables.

**Definition** A **contrast** in the treatment means  $\mu_1, \dots, \mu_t$  is defined to be a linear combination

$$C = \sum_{i=1}^t k_i \mu_i \quad \text{with} \quad \sum_{i=1}^t k_i = 0$$

Examples: Suppose we have  $t = 5$  treatment means.

1. Is  $\mu_1 = \mu_3$ ?

Write Contrast  $C_1 = \mu_1 - \mu_3$  and test  $H_o : C_1 = 0$ .

2. Is the average of  $\mu_1, \mu_2$  equal to the average of  $\mu_3, \mu_4, \mu_5$ ?

Write contrast

$$C_2 = (\mu_1 + \mu_2)/2 - (\mu_3 + \mu_4 + \mu_5)/3$$

and test  $H_o : C_2 = 0$ .

Are  $C_1$  and  $C_2$  valid contrasts?

$$\text{For } C_1 : \quad \sum_{i=1}^t k_i = 1 + 0 + (-1) = 0$$

$$\text{For } C_2 : \quad \sum_{i=1}^t k_i = \frac{1}{2} + \frac{1}{2} + \frac{-1}{3} + \frac{-1}{3} + \frac{-1}{3} = 0$$

In order to make inferences about the contrasts, we need to determine point estimators, sampling distributions of the point estimators, and test statistics.

## Inferences for Contrasts

1. Point estimators: For the contrast  $C = \sum_{i=1}^t k_i \mu_i$ , the LSE of  $C$  is obtained by replacing  $\mu_i$  with its LSE:

$$\hat{C} = \sum_{i=1}^t k_i \hat{\mu}_i = \sum_{i=1}^t k_i \bar{y}_i.$$

2. Sampling Distribution of  $\hat{C}$

(a.) Unbiased estimator of  $C$ :

$$\begin{aligned} \mu_{\hat{C}} = E[\hat{C}] &= E \left[ \sum_{i=1}^t k_i \bar{y}_i \right] \\ &= \sum_{i=1}^t k_i E[\bar{y}_i] \\ &= \sum_{i=1}^t k_i \mu_i = C \end{aligned}$$

(b.)  $Var[\hat{C}]$ :

$$\begin{aligned} \sigma_{\hat{C}}^2 = Var[\hat{C}] &= Var \left[ \sum_{i=1}^t k_i \bar{y}_i \right] \\ &= \sum_{i=1}^t k_i^2 Var[\bar{y}_i] + \sum_{1 \leq i \neq \ell \leq t} k_i k_\ell Cov(\bar{y}_i, \bar{y}_\ell) \\ &= \sum_{i=1}^t k_i^2 \sigma_e^2 / n_i + 0 \\ &= \sigma_e^2 \sum_{i=1}^t k_i^2 / n_i. \end{aligned}$$

(c.) Sampling Distribution of  $\hat{C}$ :

$\hat{C}$  is a linear combination of  $\bar{y}_1, \bar{y}_2, \dots, \bar{y}_t$ , which are independent normal r.v.'s, therefore,  $\hat{C}$  is normally distributed with mean equal to  $C$  and variance equal to  $\sigma_{\hat{C}}^2$ . That is,  $\hat{C}$  is distributed  $N(C, \sigma_{\hat{C}}^2)$ .

3. A  $100(1 - \alpha)\%$  Confidence Interval for  $C$ :

$$\hat{C} \pm t_{\alpha/2, n-t} \hat{\sigma}_e \sqrt{\sum_{i=1}^t \frac{k_i^2}{n_i}},$$

where  $\hat{\sigma}_e = \sqrt{MSE}$

Justification:

$$\begin{aligned} \text{Pivot} &= \frac{\hat{C} - C}{\hat{\sigma}_e \sqrt{\sum_{i=1}^t \frac{k_i^2}{n_i}}} \\ &= \frac{(\hat{C} - C) / \left( \sigma_e \sqrt{\sum_{i=1}^t \frac{k_i^2}{n_i}} \right)}{\left( \hat{\sigma}_e \sqrt{\sum_{i=1}^t \frac{k_i^2}{n_i}} \right) / \left( \sigma_e \sqrt{\sum_{i=1}^t \frac{k_i^2}{n_i}} \right)} \\ &= \frac{(\hat{C} - C) / \left( \sigma_e \sqrt{\sum_{i=1}^t \frac{k_i^2}{n_i}} \right)}{\sqrt{\hat{\sigma}_e^2 / \sigma_e^2}} \\ &= \frac{(\hat{C} - C) / \left( \sigma_e \sqrt{\sum_{i=1}^t \frac{k_i^2}{n_i}} \right)}{\sqrt{\frac{\text{SSE} / \sigma_e^2}{(n-t)}}} \\ &\stackrel{\mathcal{D}}{=} \frac{N(0,1)}{\sqrt{\chi_{n-t}^2 / (n-t)}} \\ &\stackrel{\mathcal{D}}{=} t - \text{Distribution with df} = n - t \end{aligned}$$

4. Test Statistic for Testing  $H_o : C = 0$  versus  $H_1 : C \neq 0$

**A.** Sum of Squares Associated with the Contrast C: Let  $\hat{C} = \sum_{i=1}^t k_i \mu_i$  then

$$SS_C = \frac{(\sum_{i=1}^t k_i \bar{y}_i.)^2}{\sum_{i=1}^t \frac{k_i^2}{n_i}} = \frac{\hat{C}^2}{\sum_{i=1}^t \frac{k_i^2}{n_i}} \quad \left( SS_C = \frac{r \hat{C}^2}{\sum_{i=1}^t k_i^2} \text{ when } n_1 = \dots = n_t = r \right)$$

Derivation of the above Sum of Squares: Recall from Handout 2, that when we are testing

$$H_o : \mathbf{H}\boldsymbol{\mu} = \mathbf{h} \quad \text{vs} \quad H_1 : \mathbf{H}\boldsymbol{\mu} \neq \mathbf{h},$$

where  $\mathbf{H}$  and  $\mathbf{h}$  are specified vectors of dimension  $k \times t$  and  $k \times 1$ , respectively, then the sum of squares associated with the null hypothesis  $H_o : \mathbf{H}\boldsymbol{\mu} = \mathbf{h}$  is given by

$$SS_{\mathbf{H}} = (\mathbf{H}\hat{\boldsymbol{\mu}} - \mathbf{h})^T \left( \mathbf{H} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{H}^T \right)^{-1} (\mathbf{H}\hat{\boldsymbol{\mu}} - \mathbf{h})$$

and  $\frac{1}{\sigma_e^2} SS_{\mathbf{H}}$  has a chisquare distribution with  $df = \text{rank}(\mathbf{H})$ , under  $H_o$  and  $SS_{\mathbf{H}}$  is independent of  $MSE$ .

First write  $H_o : C = 0$  in the form of  $H_o : \mathbf{H}\boldsymbol{\mu} = \mathbf{h}$

$C = \sum_{i=1}^t k_i \mu_i = 0$  implies that  $\mathbf{H} = (k_1, \dots, k_t)$  and  $\mathbf{h} = 0$ . Therefore,

$$(\mathbf{H}\hat{\boldsymbol{\mu}} - \mathbf{h}) = \sum_{i=1}^t k_i \bar{y}_i. = \hat{C}$$

The design matrix for the cell means model is

$$\mathbf{X} = \text{Diagonal}(\mathbf{J}_{n_1}, \mathbf{J}_{n_2}, \dots, \mathbf{J}_{n_t}) \quad \text{where} \quad \mathbf{J}_{n_i},$$

is a column vector of  $n_i$  1's. Thus,

$$(\mathbf{X}^T \mathbf{X}) = \text{Diag}(n_1, n_2, \dots, n_t) \Rightarrow (\mathbf{X}^T \mathbf{X})^{-1} = \text{Diag}\left(\frac{1}{n_1}, \frac{1}{n_2}, \dots, \frac{1}{n_t}\right) \Rightarrow$$

$$\begin{aligned}
\mathbf{H}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{H}^T &= (k_1, k_2, \dots, k_t) \begin{pmatrix} \frac{1}{n_1} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{n_2} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & & 0 & \dots & \frac{1}{n_t} \end{pmatrix} \begin{pmatrix} k_1 \\ k_2 \\ \dots \\ k_t \end{pmatrix} \\
&= \left( \frac{k_1}{n_1}, \frac{k_2}{n_2}, \dots, \frac{k_t}{n_t} \right) \begin{pmatrix} k_1 \\ k_2 \\ \dots \\ k_t \end{pmatrix} \\
&= \sum_{i=1}^t k_i^2 / n_i
\end{aligned}$$

Therefore,  $\left( \mathbf{H}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{H}^T \right)^{-1} = 1 / \left( \sum_{i=1}^t k_i^2 / n_i \right)$

Finally, we obtain the sum of squares associated with contrast  $C$ ,  $SS_C$ :

$$\begin{aligned}
SS_C = SS_{\mathbf{H}} &= (\mathbf{H}\hat{\boldsymbol{\mu}} - \mathbf{h})^T \left( \mathbf{H}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{H}^T \right)^{-1} (\mathbf{H}\hat{\boldsymbol{\mu}} - \mathbf{h}) \\
&= \left( \sum_{i=1}^t k_i \bar{y}_i \right)^T \left( 1 / \left( \sum_{i=1}^t k_i^2 / n_i \right) \right) \left( \sum_{i=1}^t k_i \bar{y}_i \right) \\
&= \frac{\left( \sum_{i=1}^t k_i \bar{y}_i \right)^2}{\sum_{i=1}^t \frac{k_i^2}{n_i}} \\
&= \frac{\hat{C}^2}{\sum_{i=1}^t \frac{k_i^2}{n_i}}
\end{aligned}$$

B. From our general result we have that when  $H_o$  is true and  $C = 0$ ,

$SS_C/\sigma_e^2$  has a chisquare distribution with  $df = \text{Rank}(\mathbf{H}) = 1$  and is distributed independent of  $MSE$ . Thus, the test statistic,

$$F = \frac{SS_C}{MSE}$$

has, under  $H_o$ , a central F-distribution with  $df = 1, n - t$ .

## 5. Partition of $SS_{TRT}$ :

Next, we will write  $SS_{TRT}$  as the sum of  $t - 1$  independent sum of squares.

**Definition:** Two contrasts  $C_1 = \sum_{i=1}^t k_i \mu_i$  and  $C_2 = \sum_{i=1}^t d_i \mu_i$  are said to be **orthogonal** if  $\sum_{i=1}^t k_i d_i = 0$

**Note:** In the textbook, the author is defining contrasts in terms of  $\hat{\mu}_i$ 's and not in terms of  $\mu_i$ 's. Therefore, the author's definition of orthogonal contrasts is with respect to the  $\hat{\mu}_i = \bar{y}_i$ . and not with respect to the  $\mu_i$ 's.

There exists  $t - 1$  contrasts in the  $t$  treatment means  $\mu_1, \mu_2, \dots, \mu_t$ :

$C_1, C_2, \dots, C_{t-1}$  such that the  $t - 1$  contrasts are mutually orthogonal.

If  $n_1 = n_2 = \dots = n_t = r$ , then the  $t - 1$  contrasts satisfy:

$$SS_{TRT} = \sum_{i=1}^{t-1} SS_{C_i}$$

with the  $t - 1$  random variables  $SS_{C_1}, SS_{C_2}, \dots, SS_{C_{t-1}}$  independently distributed.

In order to have the  $t - 1$  contrasts  $SS_{C_1}, SS_{C_2}, \dots, SS_{C_{t-1}}$  independent in the case of unequal sample sizes it is necessary to impose the further restriction that

$$\sum_{i=1}^t \frac{k_i d_i}{n_i} = 0$$

The partition of sum of squares treatment into  $t - 1$  independent contrasts provides the researcher with  $t - 1$  independent bits of information on the types of differences that may exist in the  $t$  treatment means.



**Verification of the independence of the  $t - 1$  orthogonal contrasts:**

Because the  $t - 1$  contrasts,  $C_1, C_2 \dots, C_{t-1}$ , are functions of the  $t$  independent normally distributed random variables,  $\bar{y}_1, \dots, \bar{y}_t$ , we only need to show that the contrasts are pairwise uncorrelated to prove their joint independence:

$$\begin{aligned} Cov(\hat{C}_i, \hat{C}_j) &= Cov\left(\sum_{i=1}^t k_i \bar{y}_i, \sum_{l=1}^t d_l \bar{y}_l\right) \\ &= \sum_{i=1}^t \sum_{l=1}^t k_i d_l Cov(\bar{y}_i, \bar{y}_l) \\ &= \sum_{i=1}^t k_i d_i Var(\bar{y}_i) + \sum_{i \neq l} Cov(\bar{y}_i, \bar{y}_l) \\ &= \sigma_e^2 \sum_{i=1}^t \frac{k_i d_i}{n_i} + 0 \end{aligned}$$

Therefore,  $Cov(\hat{C}_i, \hat{C}_j) = 0$  if and only if,  $\sum_{i=1}^t \frac{k_i d_i}{n_i} = 0$

Note, the partition of  $SS_{TRT}$  into  $t - 1$  sum of squares associated with  $t - 1$  orthogonal contrasts yields a precise mathematical decomposition of the differences in the  $t$  treatment means into  $t - 1$  independent units of “non-overlapping” knowledge,  $C_1, C_2 \dots, C_{t-1}$ . However, the individual contrasts in the group of  $t - 1$  contrasts may not have practical interpretation for the researcher. Thus, we will often select a set of contrasts that are not orthogonal but are the contrasts which will yield insights to the researcher concerning particular types of differences in the treatment means.

## Example of an Analysis of Differences in Treatment Means

The following example will be used to illustrate the many ways in which we can evaluate the types of differences that may exist in treatment means.

### Comparison of Brands of Tires

The stopping distance at 35 mph was measured for each of six brands of automobile tires. There were four replications for each brand. The brands had the following further classification.

1. Brands S1 and S2 had an *All Terrain* construction
2. Brands B1 and B2 had a *Bias* construction
3. Brands R1 and R2 had a *Radial* construction
4. Brand S1 is the most widely used brand of tire

	Stopping Distances			
Brand	Rep 1	Rep 2	Rep 3	Rep 4
S1	22	20	25	17
S2	26	22	27	21
B1	16	20	14	18
B2	20	25	26	21
R1	28	29	23	24
R2	22	15	19	16

First test  $H_o : \mu_1 = \mu_2 = \dots = \mu_6$  vs  $H_1 : \text{ not all } \mu_i \text{ equal}$

The test rejected  $H_o$  and the researchers were then interested in testing the following seven contrasts:

	Brand							
$C_i$	$S_1$	$S_2$	$B_1$	$B_2$	$R_1$	$R_2$	Contrast	Purpose
1	2	2	-1	-1	-1	-1	$\frac{1}{2}(\mu_{S_1} + \mu_{S_2})$ vs $\frac{1}{4}(\mu_{B_1} + \mu_{B_2} + \mu_{R_1} + \mu_{R_2})$	TERRAIN vs Rest
2	1	1	-1	-1	0	0	$\frac{1}{2}(\mu_{S_1} + \mu_{S_2})$ vs $\frac{1}{2}(\mu_{B_1} + \mu_{B_2})$	TERRAIN vs BIAS
3	1	1	0	0	-1	-1	$\frac{1}{2}(\mu_{S_1} + \mu_{S_2})$ vs $\frac{1}{2}(\mu_{R_1} + \mu_{R_2})$	TERRAIN vs RADIAL
4	0	0	1	1	-1	-1	$\frac{1}{2}(\mu_{B_1} + \mu_{B_2})$ vs $\frac{1}{2}(\mu_{R_1} + \mu_{R_2})$	BIAS VS RADIAL
5	1	-1	0	0	0	0	$\mu_{S_1}$ vs $\mu_{S_2}$	WITHIN TERRAIN
6	0	0	1	-1	0	0	$\mu_{B_1}$ vs $\mu_{B_2}$	WITHIN BIAS
7	0	0	0	0	1	-1	$\mu_{R_1}$ vs $\mu_{R_2}$	WITHIN RADIAL

Since we have 6  $\mu_i$ 's (Treatment Means), there is a set of 5=6-1 contrasts which are mutually orthogonal. One such set consists of contrasts 1, 4, 5, 6, 7. There are many more such sets of 5 mutually orthogonal contrasts.

The following SAS program will be used to analyze the data:

```
* brand.sas;
ods html; ods graphics on;
option ls=80 ps=50 nocenter nodate;
title 'Stopping Distance of 6 brands of tires';
data old; array Y Y1-Y4;
input BRD $ Y1-Y4; do over Y; SD=Y; output; end;
      drop Y1-Y4;
      label BRD = 'Brand of Tire' SD = 'Stopping Distance';
cards;
S1 22 20 25 17
S2 26 22 27 21
B1 16 20 14 18
B2 20 25 26 21
R1 28 29 23 24
R2 22 15 19 16
run;
proc glm data=old order=data;
class BRD;
model SD=BRD;
contrast 'TERRAIN VS OTHERS' BRD 2 2 -1 -1 -1 -1;
contrast 'TERRAIN VS BIAS' BRD 1 1 -1 -1 0 0;
contrast 'TERRAIN VS RADIAL' BRD 1 1 0 0 -1 -1;
contrast 'BIAS VS RADIAL' BRD 0 0 1 1 -1 -1;
contrast 'WITHIN TERRAIN' BRD 1 -1 0 0 0 0;
contrast 'WITHIN BIAS' BRD 0 0 1 -1 0 0;
contrast 'WITHIN RADIAL' BRD 0 0 0 0 1 -1;
estimate 'TERRAIN VS OTHERS' BRD 2 2 -1 -1 -1 -1;
estimate 'TERRAIN VS BIAS' BRD 1 1 -1 -1 0 0;
estimate 'TERRAIN VS RADIAL' BRD 1 1 0 0 -1 -1;
estimate 'BIAS VS RADIAL' BRD 0 0 1 1 -1 -1;
estimate 'WITHIN TERRAIN' BRD 1 -1 0 0 0 0;
estimate 'WITHIN BIAS' BRD 0 0 1 -1 0 0;
estimate 'WITHIN RADIAL' BRD 0 0 0 0 1 -1;
contrast 'TRT EFFECT' BRD 2 2 -1 -1 -1 -1,
                      BRD 0 0 1 1 -1 -1,
                      BRD 1 -1 0 0 0 0,
                      BRD 0 0 1 -1 0 0,
                      BRD 0 0 0 0 1 -1;

run;
```

# OUTPUT FROM SAS

Stopping Distance of 6 brands of tires

The GLM Procedure

## Class Level Information

Class	Levels	Values
BRD	6	S1 S2 B1 B2 R1 R2

Number of Observations Read	24
Number of Observations Used	24

The GLM Procedure

Dependent Variable: SD Stopping Distance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	246.0000000	49.2000000	5.47	0.0031
Error	18	162.0000000	9.0000000		
Corrected Total	23	408.0000000			

Least Squares Means

BRD	SD LSMEAN	Standard Error	Pr >  t
S1	21.0000000	1.5000000	<.0001
S2	24.0000000	1.5000000	<.0001
B1	17.0000000	1.5000000	<.0001
B2	23.0000000	1.5000000	<.0001
R1	26.0000000	1.5000000	<.0001
R2	18.0000000	1.5000000	<.0001

Dependent Variable: SD Stopping Distance

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
C1: TERRAIN VS OTHERS	1	12.00000000	12.00000000	1.33	0.2633
C2: TERRAIN VS BIAS	1	25.00000000	25.00000000	2.78	0.1129
C3: TERRAIN VS RADIAL	1	1.00000000	1.00000000	0.11	0.7427
C4: BIAS VS RADIAL	1	16.00000000	16.00000000	1.78	0.1991
C5: WITHIN TERRAIN	1	18.00000000	18.00000000	2.00	0.1744
C6: WITHIN BIAS	1	72.00000000	72.00000000	8.00	0.0111
C7: WITHIN RADIAL	1	128.00000000	128.00000000	14.22	0.0014
TREATMENT EFFECT	5	246.0000000	49.20000000	5.47	0.0031

Parameter	Estimate	Standard Error	t Value	Pr >  t
TERRAIN VS OTHERS	6.00000000	5.19615242	1.15	0.2633
TERRAIN VS BIAS	5.00000000	3.00000000	1.67	0.1129
TERRAIN VS RADIAL	1.00000000	3.00000000	0.33	0.7427
BIAS VS RADIAL	-4.00000000	3.00000000	-1.33	0.1991
WITHIN TERRAIN	-3.00000000	2.12132034	-1.41	0.1744
WITHIN BIAS	-6.00000000	2.12132034	-2.83	0.0111
WITHIN RADIAL	8.00000000	2.12132034	3.77	0.0014

Note that

1. Sum of squares associated with each contrast:

$$SS_{C1} = \frac{\hat{C}^2}{\sum_{i=1}^6 \frac{k_i^2}{n_i}} = \frac{(6)^2}{\frac{(2)^2}{4} + \frac{(2)^2}{4} + \frac{(-1)^2}{4} + \frac{(-1)^2}{4} + \frac{(-1)^2}{4} + \frac{(-1)^2}{4}} = 12$$

2. Sum of squares associated with the difference in the 6 Brand mean stopping distances:

$$SS_{TRT} = 246.0000000$$

3. The 4 contrasts  $C_1, C_4, C_5, C_6, C_7$  are mutually orthogonal and hence their sum of squares add to the sum of squares associated with BRAND:

$$SS_{C1} + SS_{C4} + SS_{C5} + SS_{C6} + SS_{C7} = 246.0000000$$

Because of there was an equal number of replications and the 5 contrasts were orthogonal, the 5 contrasts provide a decomposition of the Sum of Squares Treatment.

The stopping distance at 35 mph was measured for each of six brands of automobile tires. However, the stopping distances for a number of the reps failed to be recorded by the measuring devices. The experiment is now unbalanced and the contrasts no longer result in a decomposition of SS Treatment. The following SAS will be used to analyze the data.

```
*brand_unbal.sas
option ls=80 ps=50 nocenter nodate;
title 'Stopping Distance of 6 brands of tires';
data old; array Y Y1-Y4;
input BRD $ Y1-Y4;
do over Y; SD=Y;
output;
end;
drop Y1-Y4;
label BRD = 'Brand of Tire' SD = 'Stopping Distance';
cards;
S1 22 20 25 17
S2 26 22 . .
B1 16 20 14 18
B2 20 25 26 .
R1 28 29 23 24
R2 22 15 . .
run;

proc glm data=old order=data;
class BRD;
model SD=BRD;
lsmeans BRD/stderr;

contrast 'TERRAIN VS OTHERS' BRD 2 2 -1 -1 -1 -1;
contrast 'TERRAIN VS BIAS' BRD 1 1 -1 -1 0 0;
contrast 'TERRAIN VS RADIAL' BRD 1 1 0 0 -1 -1;
contrast 'BIAS VS RADIAL' BRD 0 0 1 1 -1 -1;
contrast 'WITHIN TERRAIN' BRD 1 -1 0 0 0 0;
contrast 'WITHIN BIAS' BRD 0 0 1 -1 0 0;
contrast 'WITHIN RADIAL' BRD 0 0 0 0 1 -1;
run;
```

# Class Level Information

Class	Levels	Values
BRD	6	S1 S2 B1 B2 R1 R2

Number of Observations Read	24
Number of Observations Used	19

## Least Squares Means

BRD	SD LSMEAN	Standard Error	Pr >  t
S1	21.0000000	1.6002804	<.0001
S2	24.0000000	2.2631383	<.0001
B1	17.0000000	1.6002804	<.0001
B2	23.6666667	1.8478447	<.0001
R1	26.0000000	1.6002804	<.0001
R2	18.5000000	2.2631383	<.0001

Dependent Variable: SD Stopping Distance

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
TERRAIN VS OTHERS	1	5.39102564	5.39102564	0.53	0.4810
TERRAIN VS BIAS	1	14.08333333	14.08333333	1.37	0.2620
TERRAIN VS RADIAL	1	0.16666667	0.16666667	0.02	0.9005
BIAS VS RADIAL	1	11.02083333	11.02083333	1.08	0.3185
WITHIN TERRAIN	1	12.00000000	12.00000000	1.17	0.2988
WITHIN BIAS	1	76.19047619	76.19047619	7.44	0.0173
WITHIN RADIAL	1	75.00000000	75.00000000	7.32	0.0180

Parameter	Estimate	Standard Error	t Value	Pr >  t
TERRAIN VS OTHERS	4.83333333	6.66249870	0.73	0.4810
TERRAIN VS BIAS	4.33333333	3.69568933	1.17	0.2620
TERRAIN VS RADIAL	0.50000000	3.91987048	0.13	0.9005
BIAS VS RADIAL	-3.83333333	3.69568933	-1.04	0.3185
WITHIN TERRAIN	-3.00000000	2.77176700	-1.08	0.2988
WITHIN BIAS	-6.66666667	2.44446873	-2.73	0.0173
WITHIN RADIAL	7.50000000	2.77176700	2.71	0.0180

Note that  $SS_{TRT} = 206.9385965$  but

$$SS_{C1} + SS_{C4} + SS_{C5} + SS_{C6} + SS_{C7} = 179.602335160$$

Because of the unequal number of replications, the 5 contrasts no longer provide a decomposition of the Sum of Squares Treatment because the contrasts do not satisfy:

$$\sum_{i=1}^6 \frac{k_i d_i}{n_i} = 0. \text{ In fact,}$$

for  $C_1 = 2\mu_1 + 2\mu_2 - 1\mu_3 - 1\mu_4 - 1\mu_5 - 1\mu_6$  and  $C_4 = 0\mu_1 + 0\mu_2 + 1\mu_3 + 1\mu_4 - 1\mu_5 - 1\mu_6$  we have

$$\sum_{i=1}^6 \frac{k_i d_i}{n_i} = \frac{(2)(0)}{4} + \frac{(2)(0)}{2} + \frac{(-1)(1)}{4} + \frac{(-1)(1)}{3} + \frac{(-1)(-1)}{4} + \frac{(-1)(-1)}{2} = \frac{1}{6}$$

## Multiple Testing Procedures - Simultaneous Tests of Hypotheses

In examining the seven contrasts, if an  $\alpha = .05$  level was used for each of the seven contrasts then we would have declared that two of the seven contrasts had significant evidence that they were different from 0. However, seven hypotheses have thus been tested each at the .05 level. What is the probability of a Type I error taking all seven hypotheses together. This raises the question of error rates across multiple testing situations within a given experiment.

When a researcher has multiple hypotheses to be tested, each of the individual tests has Type I and Type II errors, and it is difficult to measure the overall error rate considering all tests simultaneously. A measure of this error rate is called the familywise error rate **FWER**, in comparison to the per comparison error rate **PCER**. Definitions of PCER and FWER are given as follows:

**DEFINITION:** Suppose we have  $M$  sets of hypotheses that are to be tested. The **per comparison error rate (PCER)** is the probability of a Type I error for each of the  $M$  sets of hypotheses disregarding the other  $M - 1$  sets of hypotheses (individual error rates).

We would have  $M$  probabilities to calculate:  $PC_i = \alpha_i = P[\text{Type I error}]$  in the  $i$ th experiment for  $i = 1, \dots, M$ . Tests are then developed to control the Type I error rates for the individual decisions,  $\alpha_i$  ignoring the multiplicity of tests that are to be performed. Alternatively,

**DEFINITION:** Suppose we have  $M$  sets of hypotheses that are to be tested. The **familywise error rate (FWER)** is the probability of one or more Type I errors over the  $M$  sets of hypotheses (overall error rate).

We would have one probability to calculate:

$FWER = \alpha_F = P[\text{at least one Type I errors in the } M \text{ tests}]$ . Thus, procedures are developed to control a single Type I error rate across all  $M$  tests.

In those situations where,  $M$ , the number of tests is especially large, the FWER is too strict. FWER forces too large of a value for the probability of Type II errors (low power). In these types of situations, a measure of error called the false discovery rate is used. For example, in microarray data,  $M$  may be 1000 or more.

**DEFINITION** Suppose we have  $M$  sets of hypotheses that are to be tested. The **false discovery rate (FDR)** is the expected proportion of false positive findings among all the rejected null hypotheses, that is,  $FDR = E[V/R]$ , where  $V$  is the number of Type I errors in the  $M$  tests and  $R$  is the number of rejected null hypotheses in the  $M$  tests. Multiple comparison procedures which control the FDR tend to be more liberal than procedures which control the FWER, that is, the FDR procedures find on the average more significant differences than do procedures controlling the FWER.

Two papers which discuss procedures for controlling FDR are

- Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate, a practical and powerful approach to multiple testing. *J.R. Statistical Society B*, **57**, pp. 289-300.
- Storey, J.D. (2002) A direct approach to false discovery rates. *J.R. Statistical Society B*, **64**, pp. 479-498.



## Controlling FWER

Suppose we have  $M$  sets of hypotheses to be tested. Let  $A_i$  be the event that a Type I error is committed using the  $i$ th test and let the individual error rates for each test given by

$\alpha_i = P[A_i]$  = probability of a Type I error using the  $i$ th test

$\bigcup_{i=1}^M A_i$  is the event of one or more Type I errors in the  $M$  tests with  $\alpha_F = P[\bigcup_{i=1}^M A_i]$ . We will compute and/or bound  $\alpha_F$  under various conditions on the  $M$  tests.

**Case 1** Suppose  $A_1, A_2, \dots, A_M$  are independent events with  $\alpha_{PC}$ 's given by  $\alpha_1, \alpha_2, \dots, \alpha_M$ .

Let  $\text{FWER} = \alpha_F = P[\text{at least one Type I error occurs}] = P[\text{at least one } A_i \text{ occurs}]$

$$\alpha_F = P\left[\bigcup_{i=1}^M A_i\right] = 1 - P\left[\bigcap_{i=1}^M A_i^c\right] = 1 - \prod_{i=1}^M P[A_i^c] = 1 - \prod_{i=1}^M (1 - P[A_i]) = 1 - \prod_{i=1}^M (1 - \alpha_i)$$

If  $\alpha_i = \alpha_{pc}$  for all  $i$ , then  $\alpha_F = 1 - (1 - \alpha_{pc})^M$

To control  $\alpha_F$ , say we want  $\alpha_F = \alpha_o$ , then we could select  $\alpha_{pc}$  to satisfy

$$\alpha_{pc} = 1 - (1 - \alpha_o)^{1/M}$$

**EXAMPLE** Suppose  $\alpha_F = \alpha_o = .05$  is specified. The following table provides the upper bound on  $\alpha_F$ ,  $1 - (1 - \alpha_o)^M$  for  $\alpha_o = .05$  and the value of  $\alpha_{pc}$  to obtain  $\alpha_F \leq .05$  for various values of  $M$ .

$M$	Value of $\alpha_F$ $1 - (1 - \alpha_o)^M$	Required Value for $\alpha_{pc}$ $1 - (1 - \alpha_o)^{1/M}$
1	.0500	.0500
2	.0975	.0253
3	.1426	.0170
4	.1855	.0128
5	.2262	.0102
10	.4013	.0051
13	.4867	.0039
14	.5123	.0037

From the above table we can conclude that if we had  $M = 14$  contrasts to be tested then in order to have  $\alpha_F \leq .05$  then it would be necessary to have  $\alpha_{pc}$  no larger than .0037. This would greatly increase the probability of a Type II error for the individual tests and there would be a substantial reduction in the power of the tests to detect nonzero contrasts. Alternatively, if the 14 tests were run each at a  $\alpha = .05$  level then  $\alpha_F$  could be as large as 0.5123, more than a 50% chance of one or more of the 14 tests would result in a Type I error.

It is important to note that if we had  $M$  contrasts:  $C_1, \dots, C_M$  with all contrasts satisfying

$$\sum_{i=1}^t \frac{k_i d_i}{n_i} = 0$$

then  $SS_{C_1}, SS_{C_2}, \dots, SS_{C_M}$  are independent.

However, are the events  $A_1, A_2, \dots, A_M$  independent?

$$A_i = \{F_i \geq F_{\alpha_i} | H_o \text{ is true}\}$$

where the test statistic for the  $M$  tests are given by

$$F_i = \frac{SS_{C_i}}{MSE} \text{ for } i = 1, \dots, M$$

Thus,  $MSE$  is present in each of the test statistics. However, if  $n - t$  is very large then the level of dependency is very small and the above formula would be close to correct.

The following inequality holds provided the **contrasts are orthogonal**:

$$\alpha_F \leq 1 - (1 - \alpha_{pc})^M, \quad \text{where } \alpha_i = \alpha_{pc} \text{ for } i = 1, 2, \dots, M$$

The proof is in the book, *Simultaneous Statistical Inference*, by R. Miller.

To obtain  $\alpha_F \leq \alpha_o$  take

$$\alpha_{pc} = 1 - (1 - \alpha_o)^{1/M}.$$

Refer to the examples on the previous page but with the understanding that now we are only obtaining an **upper bound** on  $\alpha_F$  and not the exact value for the familywise error rate.

**Case 2** Suppose the contrasts are not orthogonal and the  $M$  events  $A_1, A_2, \dots, A_M$  are not independent. What can be concluded in this very general setting?

The **Bonferroni inequality** states  $\alpha_F \leq \sum_{i=1}^M \alpha_i$ .

Proof by Induction:

When  $M=2$ ,  $\alpha_F = P[A_1 \cup A_2] = P[A_1] + P[A_2] - P[A_1 \cap A_2] \leq P[A_1] + P[A_2] = \alpha_1 + \alpha_2$

Assume true for any integer  $M > 2$ :

$$\alpha_F = P \left[ \bigcup_{i=1}^M A_i \right] \leq \sum_{i=1}^M P[A_i] = \sum_{i=1}^M \alpha_i$$

Then, for  $M + 1$ , we have

$$\begin{aligned} \alpha_F = P \left[ \bigcup_{i=1}^{M+1} A_i \right] &= P \left[ \left( \bigcup_{i=1}^M A_i \right) \cup A_{M+1} \right] \\ &\leq P \left[ \left( \bigcup_{i=1}^M A_i \right) \right] + P[A_{M+1}] \\ &\leq \sum_{i=1}^M P[A_i] + P[A_{M+1}] = \sum_{i=1}^{M+1} \alpha_i \end{aligned}$$

Thus by Mathematical Induction we have proved the Bonferroni Inequality.

**Note:** If  $\alpha_i = \alpha_{pc}$  for all  $i = 1, \dots, M$ , then  $\alpha_F \leq M\alpha_{pc}$

Thus, to have  $\alpha_F \leq \alpha_o$ , we would specify

$$\alpha_{pc} = \frac{\alpha_o}{M}$$

**EXAMPLE** Suppose  $\alpha_F = \alpha_o = .05$  is specified. The following table provides the upper bound on  $\alpha_F$ ,  $M\alpha_o$ , for  $\alpha_o = .05$  and the value of  $\alpha_{pc}$  to obtain  $\alpha_E \leq .05$  for various values of  $M$ .

$M$	Bonferroni Bound on $\alpha_F$	Required Value for $\alpha_{pc}$
	$M\alpha_o$	$\alpha_o/M$
1	.05	.0500
2	.10	.0250
3	.15	.0167
4	.20	.0125
5	.25	.0100
10	.50	.0050
13	.65	.0038
14	.70	.0036

From the above table the values of the upper bound are considerably larger for large  $M$  than the values we had in the previous table. However, the values of  $\alpha_{pc}$  do not differ much in this general case from the values obtained in the case of orthogonal contrasts. In both cases, the use of the above values for  $\alpha_{pc}$  would result in procedures that would not have desirable properties for the power of the procedures and for the inflated probabilities for Type II errors. The application of these types of procedures are labeled as Bonferroni procedures:

## Bonferroni Simultaneous Procedures

If there are  $M$  hypotheses to be tested (contrasts) and it is specified that  $\alpha_F \leq \alpha_o$ , a given value, then the Bonferroni procedure consisting of

Running each of the  $M$  tests of hypotheses at level  $\alpha_{pc} = \alpha_o/M$

which results in  $\alpha_F \leq M\alpha_{pc} = \alpha_o$

### Bonferroni F-Test for Testing $M$ Contrasts

Suppose there are  $M$  contrasts  $C_1, \dots, C_M$ , and tests of  $H_o : C_i = 0$  versus  $H_1 : C_i \neq 0$  are to be constructed. The Bonferroni procedure is defined as follows:

Let  $F_i = \frac{SS_{C_i}}{MSE}$

State there is significant evidence ( $\alpha_F = \alpha_o$ ) that  $C_i$  is different from 0 if

$$F_i \geq F_{\frac{\alpha_o}{M}, 1, n-t} \quad \text{or if} \quad p\text{-value} \leq \frac{\alpha_o}{M}$$

where  $p\text{-value} = 1 - G(F_i)$  and  $G(\cdot)$  is the cdf of an F-distribution with  $df = 1, n - t$ .

**EXAMPLE** Consider the 7 contrasts defined for the Brand of Tires example. Suppose it is specified that  $\alpha_F \leq .05$ . Using the Bonferroni procedure, the 7 contrasts would be tested individually using  $\alpha_{pc} = \frac{.05}{7} = .0071$ . The critical value would be

$$F_{\frac{\alpha_o}{M}, 1, n-t} = F_{.0071, 1, 18} = 9.22$$

Thus, reject  $H_o : C_i = 0$  if  $F_i \geq 9.22$ .

Alternatively, reject  $H_o : C_i = 0$  if  $p\text{-value} \leq .0071$

Examining the SAS output, whereas previously there was significant evidence at the  $\alpha = .05$  level that two of the seven contrasts were different from 0, now there is only contrast  $C_7$ , the **Within Radial** contrast for which there is significant evidence that the contrast is different from 0.

**Problem** Because it can only be stated that  $\alpha_F \leq \alpha_o$ , the true size of  $\alpha_F$  is unknown in general. In some situations, it may be that  $\alpha_F$  is very much smaller than  $\alpha_o$  and hence the overall power of the Bonferroni procedure to detect  $C_i \neq 0$  would be small and hence the procedure would not be very sensitive at detecting false null hypotheses, that is, large probabilities of Type II errors.

In the case where the  $M$  contrasts are orthogonal, a slightly less conservative procedure can be obtained by using  $\alpha_{pc} = 1 - (1 - \alpha_o)^{1/M}$  in the Bonferroni procedure.

## Scheffé's Procedure for Testing Multiple Contrasts

Let  $C_h = \sum_{i=1}^t k_{ih}\mu_i$  for  $h = 1, \dots, M$

$\hat{C}_h = \sum_{i=1}^t k_{ih}\bar{y}_i$  for  $h = 1, \dots, M$

$D_h = \sqrt{\sum_{i=1}^t \frac{k_{ih}^2}{n_i}}$   $\hat{\sigma}_e^2 = MSE$ ,

$S_h = D_h \hat{\sigma}_e \sqrt{(t-1)F_{\alpha_o, t-1, \nu_2}}$

where  $t$  is the number of treatments,  $\nu_2$  is the df for MSE, and  $\alpha_o$  is the specified value for  $\alpha_F$ .

State there is significant evidence at the  $\alpha_o$  level that  $C_h$  is different from 0,

that is, reject  $H_o : C_h = 0$  if  $|\hat{C}_h| \geq S_h$

Note that only  $D_h$  in  $S_h$  changes in testing the  $M$  contrasts.

The Scheffé procedure has an exact familywise error rate,  $\alpha_F = \alpha_o$

Reference: *The Analysis of Variance*, by H. Scheffé

**EXAMPLE** In the Brand Example,  $t = 6$ ,  $n_i = 4$ ,  $\nu_2 = 24 - 6 = 18$ ,  $\hat{\sigma}_e^2 = 9$ , and take  $\alpha_o = .05$ , then  $D_h^2 = \frac{1}{4} \sum_{i=1}^t k_{ih}^2$ ,  $F_{\alpha_o, t-1, \nu_2} = F_{.05, 5, 18} = 2.773 = qf(.95, 5, 18)$

$S_h = D_h \hat{\sigma}_e \sqrt{(t-1)F_{\alpha_o, t-1, \nu_2}} = D_h 3 \sqrt{(6-1)F_{.05, 5, 18}} = (11.169)D_h$

The following table contains the values of  $D_h$  and  $S_h$  for testing the 7 contrasts:

Contrast	$D_h^2$	$S_h = 11.169D_h$	$ \hat{C}_h $	Conclusion
$C_1: 2 \ 2 \ -1 \ -1 \ -1 \ -1$	3	19.35	6	Evidence is not Significant that $C_1 \neq 0$
$C_2: 1 \ 1 \ -1 \ -1 \ 0 \ 0$	1	11.17	5	Evidence is not Significant that $C_2 \neq 0$
$C_3: 1 \ 1 \ 0 \ 0 \ -1 \ -1$	1	11.17	1	Evidence is not Significant that $C_3 \neq 0$
$C_4: 0 \ 0 \ 1 \ 1 \ -1 \ -1$	1	11.17	4	Evidence is not Significant that $C_4 \neq 0$
$C_5: 1 \ -1 \ 0 \ 0 \ 0 \ 0$	.5	7.90	3	Evidence is not Significant that $C_5 \neq 0$
$C_6: 0 \ 0 \ 1 \ -1 \ 0 \ 0$	.5	7.90	6	Evidence is not Significant that $C_6 \neq 0$
$C_7: 0 \ 0 \ 0 \ 0 \ 1 \ -1$	.5	7.90	8	Significant Evidence that $C_7 \neq 0$

Thus, in this example, the Scheffé procedure agrees with the Bonferroni procedure. Does this conclusion hold true in general?

Note in the Scheffé's  $\alpha_o = \alpha_F$ . This is an exact procedure.

In the Scheffé procedure, the hypothesis  $H_o : C_h$  is rejected if

$$|\hat{C}_h| \geq S_h \quad \text{iff} \quad \left( \frac{\hat{C}_h}{D_h \sigma_e} \right)^2 \geq (t-1)F_{\alpha_o, t-1, \nu_2}$$

In the Bonferroni procedure, the hypothesis  $H_o : C_h$  is rejected if

$$F_h = \frac{SS_{C_h}}{MSE} = \left( \frac{\hat{C}_h}{D_h \sigma_e} \right)^2 \geq F_{\frac{\alpha_o}{M}, 1, \nu_2}$$

where  $M$  is the number of contrasts being tested. Note that in the Scheffé procedure,  $M$  does not appear. The Scheffé procedure provides a specified Experimentwise error rate for all possible contrasts, not just a specified number of contrasts. To determine which of the two procedures is more conservative it is necessary to compare

Scheffé  $(t-1)F_{\alpha_o, t-1, \nu_2}$  to  $F_{\frac{\alpha_o}{M}, 1, \nu_2}$  for Bonferroni.

For the Brand of Tire example, with  $t = 6$  and  $\alpha_o = .05$ , we have that the Scheffé procedure is more conservative because  $(t-1)F_{\alpha_o, t-1, \nu_2} = 5F_{.05, 5, 18} = 13.864 > 9.202 = F_{.05/7, 1, 18} = F_{\frac{\alpha_o}{M}, 1, \nu_2}$

Thus, it takes a larger value of  $\hat{C}_h$  to reject  $H_o$  using Scheffé procedure than it does for the Bonferroni procedure. Why?

Scheffé procedure must consider “all possible” contrasts in terms of yielding the specified Experimentwise Type I error rate, whereas Bonferroni only needs to consider the  $M = 7$  specified contrasts. As the number of contrasts  $M$  increases, the critical value of the Scheffé procedure remains constant whereas the Bonferroni’s critical value will increase. For example, with  $t = 6, \alpha_o = .05$  and  $M$  contrasts, we have

M	7	10	20	30	32	33
Scheffe Critical Value	13.864	13.864	13.864	13.864	13.864	13.864
Bonferroni Critical Value	9.202	10.218	12.321	13.634	13.849	13.952

Thus, for the case of  $t = 6, \nu_2 = 18$ , and  $\alpha_o = .05$ , when  $M > 32$  contrasts are to be tested the Bonferroni procedure would be more conservative than Scheffé. The contrasts needed for the Bonferroni procedure to be more conservative than the Scheffé depends on all three factors:  $t, \nu_2$ , and  $\alpha_o$ .

The following table contains the required number,  $M$ , of contrasts needed for the critical value of the Bonferroni procedure to exceed the critical value for the Scheffé procedure. If the number of contrasts is less than or equal  $M$  than use the Bonferroni procedure. We will only consider  $\alpha_o = .05$  but with varying values for  $t$  and  $\nu_2$ :

$\nu_2 = 20$		$\nu_2 = 30$		$\nu_2 = 40$		$\nu_2 = 50$		$\nu_2 = 100$	
t	M	t	M	t	M	t	M	t	M
5	18	5	19	5	21	5	21	5	23
6	34	6	40	6	44	6	46	6	51
7	64	7	79	7	88	7	95	7	110
8	113	8	148	8	171	8	188	8	229
10	317	10	471	10	587	10	675	10	915
15	2725	15	5619	15	8583	15	11340	15	21154

If the specified number of contrasts,  $M$  is not too large, it is advisable to use Bonferroni. If the researcher is just exploring the data and does not specify a particular set of contrasts prior to running the experiment then use Scheffé procedure. However, if the researcher is constructing formal tests of whether or not the contrasts are significant, then this would not be a good procedure. Contrasts should never be constructed based on the observed data if formal tests are to be conducted. This would violate all statistical procedures.

### Simultaneous Confidence Intervals for $M$ Contrasts

**The Scheffé procedure** can be used to construct simultaneous  $100(1 - \alpha)$  confidence intervals for any number of contrasts.

Let  $\Theta$  be the set of all possible contrasts in  $t$  treatment means  $\mu_1, \dots, \mu_t$ :

$$\Theta = \left\{ C : C = \sum_{i=1}^t k_i \mu_i, \text{ with } \sum_{i=1}^t k_i = 0 \right\}$$

Then, a  $100(1 - \alpha)$  Scheffé simultaneous confidence interval for all  $C \in \Theta$  is given by

$$\hat{C} \pm \hat{\sigma}_e \left( \sqrt{\sum_{i=1}^t \frac{k_i^2}{n_i}} \right) \sqrt{(t-1)F_{\alpha, t-1, \nu_2}} \quad \text{that is} \quad \hat{C} \pm \widehat{SE}(\hat{C}) \sqrt{(t-1)F_{\alpha, t-1, \nu_2}}$$

where  $\hat{\sigma}_e = \sqrt{MSE}$  and  $\nu_2 = df$  associated with  $MSE$ .

That is, the probability that the intervals simultaneously contain their contrast is  $1 - \alpha$ :

$$P \left[ \hat{C} - \hat{\sigma}_e \sqrt{\sum_{i=1}^t \frac{k_i^2}{n_i}} \sqrt{(t-1)F_{\alpha, t-1, \nu_2}} \leq C \leq \hat{C} + \hat{\sigma}_e \sqrt{\sum_{i=1}^t \frac{k_i^2}{n_i}} \sqrt{(t-1)F_{\alpha, t-1, \nu_2}} \text{ for all } C \in \Theta \right] = 1 - \alpha$$

In order for the Scheffé C.I.'s to obtain their simultaneous coverage of all possible contrasts, the C.I.'s are considerably wider than the individual C.I.'s obtained using the pivot based C.I.'s.



**The Bonferroni procedure** can also be used to construct simultaneous  $100(1 - \alpha)$  confidence intervals for  $M$  contrasts. The Bonferroni t-based C.I. is

$$\hat{C} \pm \hat{\sigma}_e \left( \sqrt{\sum_{i=1}^t \frac{k_i^2}{n_i}} \right) t_{\alpha/(2M), \nu_2} \quad \text{that is} \quad \hat{C} \pm \widehat{SE}(\hat{C}) t_{\alpha/(2M), \nu_2}$$

Note the difference between the two intervals is the difference between

$$\sqrt{(t-1)F_{\alpha, t-1, \nu_2}} \quad \text{and} \quad t_{\alpha/(2M), \nu_2} = \sqrt{F_{\alpha/M, 1, \nu_2}}.$$

**EXAMPLE** To illustrate the difference between the simultaneous C.I.'s and the Bonferroni t-based C.I.'s, 95% C.I. on the seven contrasts from the Tire Brand Example using both methods:

From the example,

$$\hat{\sigma}_e = 3; \quad \sqrt{(t-1)F_{\alpha, t-1, \nu_2}} = \sqrt{(6-1)F_{.05, 5, 18}} = 3.723; \quad M = 7,$$

$$t_{\alpha/(2*7), n-t} = t_{.05/14, 18} = qt(1 - .05/14, 18) = 3.033631$$

$$D_h = \sqrt{\sum_{i=1}^t \frac{k_i^2}{n_i}}; \quad S_h = \hat{\sigma}_e \sqrt{(t-1)F_{\alpha, t-1, \nu_2}} D_h = (3)(3.723)D_h \Rightarrow$$

$$\text{Scheffé C.I.: } \hat{C} \pm S_h \quad \text{Bonferroni t-based C.I.: } \hat{C} \pm \hat{\sigma}_e t_{\alpha/(2M), n-t} D_h$$

$$\text{Scheffé C.I.: } \hat{C} \pm (11.17)D_h \quad \text{Bonferroni t-based C.I.: } \hat{C} \pm (9.101)D_h$$

The following table contains the two sets of SCIs for the seven contrasts:

Contrast	Scheffé C.I.		Bonferroni C.I.	
	$\hat{C}_h$	$D_h^2$	$\hat{C} \pm 11.169D_h$	$\hat{C} \pm 9.101D_h$
$C_1$	6	3	$6 \pm 19.35$	$6 \pm 15.763$
$C_2$	5	1	$5 \pm 11.17$	$5 \pm 9.10$
$C_3$	1	1	$1 \pm 11.17$	$1 \pm 9.10$
$C_4$	-4	1	$-4 \pm 11.17$	$-4 \pm 9.10$
$C_5$	-3	.5	$-3 \pm 7.90$	$-3 \pm 6.44$
$C_6$	-6	.5	$-6 \pm 7.90$	$-6 \pm 6.44$
$C_7$	8	.5	$8 \pm 7.90$	$8 \pm 6.44$

Thus, the Scheffé SCIs are wider than the Bonferron SCIs in this example. This result depends on both the number of contrasts  $M$  and  $\nu_2$ , the degrees of freedom for estimating  $\sigma_e$  as we observed previously. In general, when both  $M$  and  $\nu_2$  are relatively small, the Scheffé SCI will be wider than the Bonferroni SCI. But, as  $M$  and/or  $\nu_2$  become large, the Bonferroni SCIs will eventually become wider than the Scheffé SCIs.

## Multiple Comparison with Specific Types of Contrasts

There are three specific types of comparisons of the treatment means that researchers often make in designed experiments. For these types of comparisons, neither Bonferroni nor Scheffé is the “best” procedure.

Three specific types of Comparisons:

1. Compare all pairs: Fisher’s protected LSD, Tukey’s HSD, Student-Newman-Kuels (SNK), many others
2. Compare  $t - 1$  treatment means to a Control or Standard mean: Dunnett’s procedure
3. Determine a group of treatment means which contains the “Best” treatment: Hsu’s procedure

For comparing all pairs and comparing to a Control, a Bonferroni t-test or Scheffé’s procedure could be used. However, the Type II error may be much larger than the Type II error rate for a specialized procedure which will result in diminished power. For completeness, we will provide the details for both of these procedures, but it is **not recommended** to ever use either one of them for comparing all pairs of means or for comparing treatment means to a control.

The procedures LSD, HSD, SNK, Dunnett, and Hsu are designed to control an experimentwise error rate relative to a fixed number of contrasts:  $M = \binom{t}{2}$  contrasts for all pairwise comparisons and  $M = t - 1$  contrasts for comparing  $t - 1$  treatments to a control. The Scheffé’s procedure has an experimentwise error rate which must be maintained over all possible contrasts. Thus, we have

Let  $A_i$  be the event of a Type I error in testing  $H_o : C_i = 0$  versus  $H_1 : C_i \neq 0$  for  $i = 1, \dots, M$  contrasts

The familywise error rate for Tukey or Dunnett satisfies:

$$\alpha_F(M \text{ contrasts}) = P \left[ \bigcup_{i=1}^M A_i \right] = \alpha_o$$

The familywise error rate for Scheffé satisfies:

$$\alpha_F(\text{All possible Contrasts}) = P \left[ \bigcup_{all\ possible} A_i \right] = \alpha_o$$

All three procedures have the same Familywise error rate of  $\alpha_o$  but Scheffé’s procedure must protect for many more hypotheses than both Tukey and Dunnett.

This results in Scheffé’s procedure tending to be more conservative for testing just  $M$  contrasts which results in Scheffé procedure producing more Type II errors than Tukey or Dunnett when we are testing  $M$  specific contrasts.

**Bonferroni t-test** Suppose we have  $t$  treatment means and want to compare all  $M = \binom{t}{2} = \frac{t(t-1)}{2}$  pairs. For  $k \neq h = 1, \dots, t$ , define

$$t_{kh} = \frac{\bar{y}_{k.} - \bar{y}_{h.}}{\hat{\sigma}_e \sqrt{\frac{1}{n_k} + \frac{1}{n_h}}} \quad \text{with} \quad \hat{\sigma}_e^2 = MSE \quad \text{and} \quad df = df_{MSE}$$

State there is significant evidence at level  $\alpha_o$  that  $\mu_k$  is different from  $\mu_h$  if

$$|t_{kh}| \geq t_{\frac{\alpha_o}{2M}, df_{MSE}}$$

The Experimentwise Type I error rate,  $\alpha_F$ , satisfies,  $\alpha_F \leq \alpha_o$ .

A set of simultaneous C.I.'s on the  $M$  differences  $\mu_k - \mu_h$  for  $k \neq h = 1, \dots, t$  having simultaneous coverage probability at least  $100(1 - \alpha)\%$  are given by

$$(\bar{y}_{k.} - \bar{y}_{h.}) \pm \left( t_{\frac{\alpha_o}{2M}, df_{MSE}} \right) \hat{\sigma}_e \sqrt{\frac{1}{n_k} + \frac{1}{n_h}}$$

**Scheffé Procedure** Consider the  $M$  contrasts:

$$C_{kh} = \mu_k - \mu_h = \mu_k - \mu_h + \sum_{i \neq k, h}^t 0(\mu_i) \quad \text{for} \quad k \neq h = 1, \dots, t$$

State there is significant evidence at level  $\alpha_o$  that  $\mu_k$  is different from  $\mu_h$  if

$$|\hat{C}_{kh}| \geq \hat{\sigma}_e(D_h) \sqrt{(t-1)F_{\alpha_o, t-1, df_{MSE}}} \quad \text{where} \quad D_h = \sqrt{\frac{1}{n_k} + \frac{1}{n_h}}$$

Scheffé simultaneous C.I.'s for the  $M$  contrasts can be constructed using the procedures described earlier.

The Scheffé procedure has  $\alpha_F = \alpha_o$ . However, the Bonferroni procedure is conservative with  $\alpha_F \leq \alpha_o$ . How much less  $\alpha_F$  is than  $\alpha_o$  would vary depending on  $n_i$ 's,  $t$ , and  $\sigma_e$ .

## Tukey's Honest Significant Difference, HSD

HSD and Scheffé are two of the very few multiple comparison procedures which have an exact value for  $\alpha_F$ , that is,  $\alpha_F = \alpha_o$ . Tukey's HSD procedure requires that the data have equal sample sizes  $r = n_1 = n_2 = \dots = n_t$  and the standard AOV distributional requirements on the experimental data:

$$y_{ij} = \mu_i + e_{ij} \quad i = 1, \dots, t \quad j = 1, \dots, r \quad \text{with } e_{ij} \text{ having iid } N(0, \sigma_e^2) \text{ distributions}$$

Tukey's procedure is based on the studentized range statistic:

$$q(t, \nu_2) = \frac{\bar{y}_{max} - \bar{y}_{min}}{\hat{\sigma}_e \sqrt{\frac{1}{r}}}$$

$$\bar{y}_{min} = \min\{\bar{y}_1, \bar{y}_2, \dots, \bar{y}_t\}, \quad \bar{y}_{max} = \max\{\bar{y}_1, \bar{y}_2, \dots, \bar{y}_t\}, \quad \hat{\sigma}_e^2 = MSE.$$

Define

$$HSD(t, \nu_2) = q(\alpha_o, t, \nu_2) \sqrt{\frac{1}{2} \left( \widehat{SE}(\bar{y}_k. - \bar{y}_{h.}) \right)^2} = q(\alpha_o, t, \nu_2) \sqrt{\frac{1}{2} \left( \frac{2\sigma_e^2}{r} \right)} \Rightarrow$$

$$HSD(t, \nu_2) = q(\alpha_o, t, \nu_2) \hat{\sigma}_e \sqrt{1/r}$$

where  $\nu_2 = df_{MSE}$ ,  $t$  is number of treatment means to be compared,  $r$  is the common sample size,  $\alpha_o$  is the specified value for  $\alpha_F$ , and  $q(\alpha_o, t, \nu_2)$  is the solution to the equation:

$$P \left[ \frac{|(\bar{y}_{k.} - \mu_k) - (\bar{y}_{h.} - \mu_h)|}{\hat{\sigma}_e \sqrt{1/r}} \leq q(\alpha_o, t, \nu_2) \text{ for all } k > h \right] = 1 - \alpha_o$$

Values for  $q(\alpha_o, t, \nu_2)$  are given in Table VII in Kuehl's book. These are upper  $\alpha$  percentiles.

The R function **qtukey**( $1 - \alpha, t, df_{MSE}$ ) also yields these values.

For example, for  $\alpha = .05$ ,  $t = 5$  treatments and  $df_{MSE} = 9$ ,

**qtukey**(.95, 5, 9) = 4.755404 which is given as 4.76 in Table VII.

### Tukey HSD Procedure:

1. State there is significant evidence at level  $\alpha_o$  that the treatment means  $\mu_k$  and  $\mu_h$  are different if

$$|\bar{y}_k. - \bar{y}_h.| \geq HSD(t, \nu_2)$$

2. Simultaneous  $100(1 - \alpha_o)\%$  confidence intervals for the absolute difference in all pairs of treatment means  $\mu_k - \mu_h$  are given by

$$|\bar{y}_k. - \bar{y}_h.| \pm HSD(t, \nu_2)$$

3. Alternatively, the above simultaneous C.I.'s can be used to compare the  $M = \binom{t}{2}$  pairs of means by stating there is significant evidence at the  $\alpha_o$  level that  $\mu_k$  and  $\mu_h$  are different if 0 is not contained in the C.I.

4. When the sample sizes are unequal, a procedure called the **Tukey-Kramer procedure** replaces

$$\sqrt{1/r} \text{ with } \sqrt{\frac{1}{2} \left( \frac{1}{n_k} + \frac{1}{n_h} \right)}. \text{ This yields a multiple comparison procedure in which } \alpha_F \leq \alpha_o.$$

Tukey's HSD provided an exact value for  $\alpha_F$ . For most multiple comparison procedures only an approximate value for  $\alpha_F$  is provided. In an extensive simulation study, Carmer and Swanson(1973), "An evaluation of ten pairwise multiple comparison procedures by Monte Carlo methods," *Journal of the American Statistical Association* **68**, pp. 66-74, have been shown the procedures LSD and SNK to have  $\alpha_F$  very close to the specified value  $\alpha_o$ .

## Fisher's Protected Least Significant Difference, LSD

Most statisticians would not recommend using LSD because it does not have an exact FWER.

Fisher's protected LSD allows for unequal sample sizes  $n_1, n_2, \dots, n_t$  and the standard AOV requirements on the experimental data:

$$y_{ij} = \mu_i + e_{ij} \quad i = 1, \dots, t \quad j = 1, \dots, r \quad \text{with } e_{ij} \text{ having iid } N(0, \sigma_e^2) \text{ distributions}$$

Let  $\hat{\sigma}_e = \sqrt{MSE}$ ,  $\nu_2 = df_{MSE}$ ,  $t$  be the number of treatment means to be compared,  $\alpha_o$  be the specified value for  $\alpha_F$ , and  $t(\alpha_o/2, \nu_2)$  be the upper  $\alpha_o/2$  percentile from the t-distribution with  $df = df_{MSE}$ .

The following steps constitute the protected LSD procedure:

1. Conduct the AOV F-test of  $H_o : \mu_1 = \mu_2 = \dots = \mu_t$  at level  $\alpha_o$ . If AOV F-test fails to reject  $H_o$ , then state there is not significant evidence that  $\mu_k$  and  $\mu_h$  are different for any pair of treatment means.

2. If AOV F-test rejects  $H_o$  at level  $\alpha_o$ , then compute

$$LSD(k, h) = t_{\alpha_o/2, \nu_2} \hat{\sigma}_e \sqrt{\frac{1}{n_k} + \frac{1}{n_h}}$$

3. State there is significant evidence at level  $\alpha_o$  that the pair of treatment means  $\mu_k$  and  $\mu_h$  are different whenever

$$|\bar{y}_k - \bar{y}_h| \geq LSD(k, h)$$

4. If  $r = n_1 = \dots = n_t$ , then there is a single value of  $LSD(k, h) = t_{\alpha_o/2, \nu_2} \hat{\sigma}_e \sqrt{2 \frac{1}{r}}$

5. The most we can state is that  $\alpha_F$  is approximately  $\alpha_o$ ,  $\alpha_F$  may be somewhat larger or smaller than  $\alpha_o$ . The inability to state the value of  $\alpha_F$  is why most statisticians recommend not using Fisher's LSD.

6. The Unprotected LSD procedure does not conduct the AOV F-test. Thus, it is just running  $M = t(t-1)/2$  t-tests and  $\alpha_F$  can be much larger than  $\alpha_o$ . In fact, it could be as large as  $M\alpha_o$ .

7. Using R,  $t_{\alpha_o/2, \nu_2} = qt(1 - \alpha_o/2, \nu_2)$

For example, for  $\alpha_o = .05, \nu_2 = 10$ ,  $qt(.975, 10) = 2.228$

## Student-Newman-Keuls, SNK Procedure

Most statisticians would not recommend using SNK because it does not have an exact FWER.

SNK is a less conservative than Tukey's HSD in that it generally requires a smaller difference between the treatment sample means,  $\bar{y}_i$  and  $\bar{y}_k$ , than is required by HSD to declare a pair of treatment  $\mu_i$  and  $\mu_k$  to be different.

SNK requires equal sample sizes  $r = n_1 = n_2 = \dots = n_t$  and the standard AOV requirements on the experimental data:

$$y_{ij} = \mu_i + e_{ij} \quad i = 1, \dots, t \quad j = 1, \dots, r \quad \text{with } e_{ij} \text{ having iid } N(0, \sigma_e^2) \text{ distributions}$$

Like Tukey's procedure, SNK is based on the studentized range statistic:

$$q(m, \nu_2) = \frac{\bar{y}_{max,m} - \bar{y}_{min,m}}{\hat{\sigma}_e \sqrt{\frac{1}{r}}}$$

$\bar{y}_{min,m}$  = minimum over a set of  $m$  sample means:  $\{\bar{y}_{i_1}, \bar{y}_{i_2}, \dots, \bar{y}_{i_m}\}$ , and  $\bar{y}_{max,m}$  = maximum over a set of  $m$  sample means:  $\{\bar{y}_{i_1}, \bar{y}_{i_2}, \dots, \bar{y}_{i_m}\}$   $\hat{\sigma}_e^2 = MSE$ .

Define

$$SNK(m, \nu_2) = q(\alpha_o, m, \nu_2) \hat{\sigma}_e \sqrt{1/r}, \quad \text{for } m = 2, 3, \dots, t$$

where  $\hat{\sigma}_e = \sqrt{MSE}$ ,  $\nu_2 = df_{MSE}$ ,  $t$  is number of treatment means being compared,  $r$  is the common sample size,  $\alpha_o$  is the specified value for  $\alpha_F$ , and with values for  $q(\alpha_o, m, \nu_2)$  given in Table VII in Kuehl's book, the same table we used for Tukey's HSD. Hence, we can also use the R function **qtukey**( $1 - \alpha, t, df_{MSE}$ ) to find the value of  $q(\alpha_o, m, \nu_2)$ , with  $m = t$  and  $\nu_2 = df_{MSE}$ .

1. Order the  $t$  treatment sample means:  $\bar{y}_{(1)} \leq \bar{y}_{(2)} \leq \dots \leq \bar{y}_{(t)}$  and consider the following matrix of differences (assume  $t = 6$  for illustration purposes) with  $D(k, h) = \bar{y}_{(k)} - \bar{y}_{(h)}$ :

	$\bar{y}_{(1)}$	$\bar{y}_{(2)}$	$\bar{y}_{(3)}$	$\bar{y}_{(4)}$	$\bar{y}_{(5)}$	$\bar{y}_{(6)}$	Compare to
$\bar{y}_{(1)}$		$D(2, 1)$	$D(3, 1)$	$D(4, 1)$	$D(5, 1)$	$D(6, 1)$	$SNK(6, \nu_2)$
$\bar{y}_{(2)}$			$D(3, 2)$	$D(4, 2)$	$D(5, 2)$	$D(6, 2)$	$SNK(5, \nu_2)$
$\bar{y}_{(3)}$				$D(4, 3)$	$D(5, 3)$	$D(6, 3)$	$SNK(4, \nu_2)$
$\bar{y}_{(4)}$					$D(5, 4)$	$D(6, 4)$	$SNK(3, \nu_2)$
$\bar{y}_{(5)}$						$D(6, 5)$	$SNK(2, \nu_2)$

2. Compute  $SNK(m, \nu_2)$  for  $m = 2, \dots, t$ .
3. For each of the differences in the above matrix, compare  $D(k, h)$  to  $SNK(m, \nu_2)$  for  $m = h - k + 1$ . Note, that differences along a common diagonal are compared to the same  $SNK(m, \nu_2)$  value. State there is significant evidence at the  $\alpha_o$  level that  $\mu_k$  and  $\mu_h$  are different if  $D(k, h) = |\bar{y}_{k.} - \bar{y}_{h.}| \geq SNK(m, \nu_2)$  for the appropriate value of  $m$ .
4. If  $\bar{y}_{k_1.} < \bar{y}_{k_2.} < \bar{y}_{h_2.} < \bar{y}_{h_1.}$  and it has been determined using SNK that  $\bar{y}_{k_1.}$  and  $\bar{y}_{h_1.}$  do not exceed the critical value for SNK then  $\bar{y}_{k_2.}$  and  $\bar{y}_{h_2.}$  should not be compared. It would be given that there is not significant evidence that  $\mu_{k_2}$  and  $\mu_{h_2}$  are different.
5. For SNK,  $\alpha_F$  is only approximately equal to  $\alpha_o$ , it may be larger or smaller depending on the particular values of the parameters. The inability to state the value of  $\alpha_F$  is why most statisticians recommend not using SNK.

### A few observations about HSD, LSD, and SNK:

1. If the sample sizes are unequal, most software programs use the harmonic mean in place of  $r$  when computing SNK:

$$r = \frac{t}{\sum_{i=1}^t \frac{1}{n_i}}$$

The problem is that the resulting procedure may have an inflated  $\alpha_F$  in comparison to the values given for equal sample sizes.

2. When sample sizes are equal,

$$LSD = SNK(2, \nu_2) \leq SNK(t, \nu_2) = HSD(t, \nu_2)$$

Therefore, SNK will always declare as many treatment means or more different as HSD but as many or less than LSD.

3. The Experimentwise error rate  $\alpha_F$  is exact only for HSD and Scheffé.
4. Bonferroni's t-based comparisons provide an upper bound on  $\alpha_F$ .
5. The values of  $\alpha_F$  are only approximate for LSD and SNK. The familywise error rate,  $\alpha_F$ , may be larger or smaller than the nominal value used by the experimenter. For this reason, LSD and SNK are not recommended for multiple comparisons of treatment means.



## Simultaneous Confidence Intervals on the Differences in Treatment Means

Whenever possible, a test of hypotheses should include confidence intervals on the treatment effects,  $\mu_i - \mu_k$ , in this case. Because there are many such effects,  $t(t-1)/2$ , in the case of  $t$  treatment means, it is necessary to construct simultaneous confidence intervals on the treatment mean differences. We will consider four such intervals: Separate-t intervals, Bonferroni intervals, Scheffé's intervals, Tukey intervals.

The four intervals can be obtained from the general formula:

$$\text{Simultaneous C.I. on } \mu_i - \mu_k : \quad \bar{y}_{i.} - \bar{y}_{k.} \pm C \sqrt{\frac{2}{n} \hat{\sigma}^2}$$

where  $\hat{\sigma}^2$  is MSE from the AOV table,  $n$  is the common sample size, and  $C$  is a constant specific to each method for comparing the  $t(t-1)/2$  treatment mean differences:

1. Separate-t intervals:  $C_t = t(\alpha/2; (N-t))$
2. Bonferroni intervals:  $C_B = t(\alpha/(t(t-1)); (N-t))$
3. Scheffé's intervals:  $C_S = \sqrt{(t-1)F(\alpha; (t-1), (N-t))}$
4. Tukey intervals:  $C_T = \frac{1}{\sqrt{2}}q(\alpha; t, (N-t))$

where  $t(\alpha/2; (N-t))$ ,  $t(\alpha/(t(t-1)); (N-t))$ ,  $F(\alpha; (t-1), (N-t))$  are upper percentiles from the  $t$  and  $F$  distributions and  $q(\alpha; t, (N-t))$  is the  $\alpha$  percentile from the Studentized range distribution.

The separate-t intervals are not valid simultaneous C.I.'s. They were included for just comparison purposes.

To illustrate the differences in the widths of the four C.I.'s, consider the Brand of Tires example:

$$\alpha = .05, \quad t = 6, \quad n = 4, \quad N = (6)(4) = 24, \quad \sqrt{\frac{2}{n} \hat{\sigma}^2} = \sqrt{\frac{2}{4}(9)} = 2.1213$$

$$C_t = t(.025; 18) = 2.1009, \quad C_B = t(.00167; 18) = 3.3795$$

$$C_S = \sqrt{(5)F(.05; 5, 18)} = 3.7235, \quad C_T = \frac{1}{\sqrt{2}}q(.05; 6, 18) = 3.1780$$

Based on the above calculations, the half widths of the four C.I.'s,  $C \sqrt{\frac{2}{n} \hat{\sigma}^2}$  would be

Separate-t = 4.4567, Bonferroni = 7.1708, Scheffé = 7.8987, Tukey = 6.7416

The Separate-t C.I.'s would be too narrow to provide simultaneous coverage, the Scheffé and Bonferroni C.I.'s are too wide, and Tukey's C.I.'s are just right.

## Comparison of All Treatment Means to a Control or Standard: Dunnett's Procedure

Dunnett's procedure is the most widely used multiple comparison procedure in the situation where the experiment or study has a standard treatment or a control treatment to which all other treatments are compared.

Suppose we have  $t$  treatment means:  $\mu_c, \mu_1, \mu_2, \dots, \mu_{t-1}$  with  $\mu_c$  the mean associated with a standard or control treatment.

We wish to compare  $\mu_i, i = 1, \dots, t-1$  to  $\mu_c$  in one of three possible manners:

$$\mu_i < \mu_c \quad \mu_i > \mu_c \quad \mu_i \neq \mu_c \quad \text{for } i = 1, 2, t-1.$$

Dunnett's procedure requires equal sample sizes  $r = n_1 = n_2 = \dots = n_t$  and the standard AOV requirements on the experimental data:

$$y_{ij} = \mu_i + e_{ij} \quad i = 1, \dots, t \quad j = 1, \dots, r \quad \text{with } e_{ij} \text{ having iid } N(0, \sigma_e^2) \text{ distributions}$$

Let  $d2_{\alpha_o, t-1, \nu_2}$  be the upper  $\alpha_o$  percentile of the statistic,

$$\frac{\max_{i=1, \dots, t-1} (|\bar{y}_i - \bar{y}_c|)}{\hat{\sigma}_e \sqrt{\frac{2}{r}}} \quad (\text{two-sided procedure})$$

and let  $d1_{\alpha_o, t-1, \nu_2}$  be the upper  $\alpha_o$  percentile of the statistic,

$$\frac{\max_{i=1, \dots, t-1} (\bar{y}_i - \bar{y}_c)}{\hat{\sigma}_e \sqrt{\frac{2}{r}}} \quad (\text{one-sided procedure})$$

Values of  $d1_{\alpha_o, m, \nu_2}$  and  $d2_{\alpha_o, m, \nu_2}$  (labeled 1-sided and 2-sided, respectively) and with  $k = t-1$  are given in Table VI in Kuehl's book.

Alternatively, the following R function can be used to obtain these values:

`library(mvtnorm)`

$$d1_{\alpha_o, m, \nu_2} =$$

$$qmvt(p = 1 - \alpha_o, tail = "lower.tail", df = \nu_2, corr = matrix(rep(.5, m^2), m) + diag(m) * .5) \$ quantile$$

$$d2_{\alpha_o, m, \nu_2} =$$

$$qmvt(p = 1 - \alpha_o, tail = "both.tails", df = \nu_2, corr = matrix(rep(.5, m^2), m) + diag(m) * .5) \$ quantile$$

For example, with  $\alpha_o = .01$ ,  $\nu_2 = 20$  and  $m = 4$ , we have

$$d1_{.01, 5, 20} =$$

$$qmvt(p = .99, tail = "lower.tail", df = 20, corr = matrix(rep(.5, 4^2), 4) + diag(4) * .5) \$ quantile = 3.088$$

$$d2_{.01, 5, 20} =$$

$$qmvt(p = .99, tail = "both.tails", df = 20, corr = matrix(rep(.5, 4^2), 4) + diag(4) * .5) \$ quantile = 3.386$$

The R function differs from the values in Table VI in the textbook by at most .01.

## Dunnett's Procedure

1. Compute  $D2(t-1, \nu_2) = d2_{\alpha_o, t-1, \nu_2} \hat{\sigma}_e \sqrt{\frac{2}{r}}$  or  $D1(t-1, \nu_2) = d1_{\alpha_o, t-1, \nu_2} \hat{\sigma}_e \sqrt{\frac{2}{r}}$
2. State there is significant evidence at the  $\alpha_o$  level that  $\mu_i$  is **different from**  $\mu_c$  if

$$|\bar{y}_i - \bar{y}_c| \geq D2(t-1, \nu_2)$$

- Simultaneous confidence intervals on  $\mu_i - \mu_c$  for  $i = 1, \dots, t-1$  are given by

$$\bar{y}_i - \bar{y}_c \pm D2(t-1, \nu_2)$$

3. State there is significant evidence at the  $\alpha_o$  level that  $\mu_i$  is **greater than**  $\mu_c$  if

$$\bar{y}_i - \bar{y}_c \geq D1(t-1, \nu_2)$$

- Simultaneous lower confidence bounds on  $\mu_i - \mu_c$  for  $i = 1, \dots, t-1$  are given by

$$\bar{y}_i - \bar{y}_c - D1(t-1, \nu_2)$$

4. State there is significant evidence at the  $\alpha_o$  level that  $\mu_i$  is **less than**  $\mu_c$  if

$$\bar{y}_i - \bar{y}_c \leq -D1(t-1, \nu_2)$$

- Simultaneous upper confidence bounds on  $\mu_i - \mu_c$  for  $i = 1, \dots, t-1$  are given by

$$\bar{y}_i - \bar{y}_c + D1(t-1, \nu_2)$$

5. Dunnett's procedure yields **exact** experimentwise error rates when there are equal sample sizes.

6. Table VI values are for  $n_1 = n_2 = \dots = n_{t-1} = n_c = r$ . For unequal number of replications, let  $m = \max_{i=1, \dots, t-1} (1 + .07(1 - \frac{n_i}{n_c}))$  then compute,

$$D1(t-1, \nu_2) = (m)(d1_{\alpha_o, t-1, \nu_2}) \hat{\sigma}_e \sqrt{\frac{1}{n_i} + \frac{1}{n_c}} \quad \text{or} \quad D2(t-1, \nu_2) = (m)(d2_{\alpha_o, t-1, \nu_2}) \hat{\sigma}_e \sqrt{\frac{1}{n_i} + \frac{1}{n_c}}$$

- When the sample sizes are not equal, Dunnett's procedure yields approximate experimentwise error rates.
7. Dunnett (1964), "New tables for multiple comparison with a control." *Biometrics* **20**, pp. 482-491, has tables for the critical values.
  8. Dunnett (1955), "A multiple comparison procedure for comparing several treatments with a control." *JASA* **50**, 1096-1121, showed that the relationship  $n_c = n\sqrt{t-1}$  yields the maximum coverage probability for a lower(upper) confidence bound on  $\mu_i - \mu_c$ .

## Selecting a Group of “Best” Treatments: Hsu’s Procedure

In many experimental settings, the goal of the experiment is to determine which treatments yield the **best treatment means**, either largest or smallest. In these types of experiments we can employ Hsu’s procedure to determine the group of best treatments.

Hsu’s procedure requires equal sample sizes  $n_1 = \dots = n_t = r$  and the standard AOV requirements on the experimental data:

$$y_{ij} = \mu_i + e_{ij} \quad i = 1, \dots, t \quad j = 1, \dots, r \quad \text{with } e_{ij} \text{ having iid } N(0, \sigma_e^2) \text{ distributions}$$

Let  $d_{\alpha, t-1, \nu_2}$  be the **1-sided** values given in Table VI in Kuehl’s book, that is, the 1-sided Dunnett values. The following R function can also be used to obtain  $d_{\alpha, t-1, \nu_2}$  with  $m=t-1$ :

library(mvtnorm)

`qmvtnorm(p = 1 - alpha, tail = "lower.tail", df = nu2, corr = matrix(rep(.5, m^2), m) + diag(m) * .5)$quantile`

### Hsu’s Procedure, when Best is Largest:

Hsu’s procedure produces a group  $G = \{\mu_{i_1}, \mu_{i_2}, \dots, \mu_{i_m}\}$  of  $m$  treatment means such that the probability that  $G$  contains the largest treatment mean is at least  $1 - \alpha$ .

1. For  $i = 1, \dots, t$  let  $M_i = \max_{h \neq i} \bar{y}_h$ . (largest sample mean excluding the  $i$ th sample mean.
2. Let  $K_i = M_i - d_{\alpha, t-1, \nu_2} \hat{\sigma}_e \sqrt{\frac{2}{r}}$
3. The  $i$ th treatment mean  $\mu_i$  is placed in  $G$  if  $\bar{y}_i \geq K_i$

$$\text{That is, if } M_i - \bar{y}_i \leq d_{\alpha, t-1, \nu_2} \hat{\sigma}_e \sqrt{\frac{2}{r}} \Rightarrow (\bar{y}_i \text{ is close to } M_i)$$

### Hsu’s Procedure, when Best is Smallest:

Hsu’s procedure produces a group  $G = \{\mu_{i_1}, \mu_{i_2}, \dots, \mu_{i_m}\}$  of  $m$  treatment means such that the probability that  $G$  contains the smallest treatment mean is at least  $1 - \alpha$ .

1. For  $i = 1, \dots, t$  let  $m_i = \min_{h \neq i} \bar{y}_h$ . (smallest sample mean excluding the  $i$ th sample mean.
2. Let  $k_i = m_i + d_{\alpha, t-1, \nu_2} \hat{\sigma}_e \sqrt{\frac{2}{r}}$
3. The  $i$ th treatment mean  $\mu_i$  is placed in  $G$  if  $\bar{y}_i \leq k_i$

$$\text{That is, if } \bar{y}_i - m_i \leq d_{\alpha, t-1, \nu_2} \hat{\sigma}_e \sqrt{\frac{2}{r}} \Rightarrow (\bar{y}_i \text{ is close to } m_i)$$

**Note:** For unequal replications, see Hsu(1994), *Multiple Comparisons, Theory and Methods*, pp. 89-95, Chapman & Hall.

An approximation to Hsu's procedure is obtained by replacing

$$d_{\alpha, t-1, \nu_2} \hat{\sigma}_e \sqrt{\frac{2}{r}} \quad \text{with} \quad t_{\alpha/(t-1), \nu_2} \hat{SE}(\hat{\mu}_i - \hat{\mu}_j)$$

where  $\hat{SE}(\hat{\mu}_i - \hat{\mu}_j) = \sigma_e \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$  for the case of unequal sample sizes in a CRD design.

### Tire Brand example:

In the Brand Example the “Best” treatment is the treatment having **Smallest**  $\mu_i$

$d(5, 0.05, 18) = 2.41$ , which yields  $k_i = m_i + (2.41)(3)\sqrt{2/4} = m_i + 5.11$

$d(5, 0.01, 18) = 3.21$ , which yields  $k_i = m_i + (3.21)(3)\sqrt{2/4} = m_i + 6.81$

Brand	$\bar{y}_i$	$m_i$	0.05	0.01
			$k_i$	$k_i$
$S_1$	21	17	22.11	23.81
$S_2$	24	17	22.11	23.81
$B_1$	17	18	23.11	24.81
$B_2$	23	17	22.11	23.81
$R_1$	26	17	22.11	23.81
$R_2$	18	17	22.11	23.81

For  $\alpha = 0.05$ , the Collection,  $G_1$ , containing the treatment with the “Best” mean is

$$G_1 = \{S_1, B_1, R_2\}$$

That is, we are 95% confident that  $G_1$  contains the treatment with the “Best” mean.

For 99% confidence, i.e., for  $\alpha = 0.01$ , the group of treatments becomes  $G_2 = \{S_1, B_1, B_2, R_2\}$

Notice to increase the level of confidence from 95% to 99% for  $G$  to contain the treatment with the “Best” mean it was necessary to add another treatment to the group.

A 95% C.I. for the difference between the  $i$ th treatment mean and the best treatment mean  $\mu^*$ :  $\mu_i - \mu^*$  is given by  $(0, D_i)$ , where  $D_i = \max(0, \bar{y}_i - m_i + 5.11)$ .

Brand	95% C.I. on $\mu_i - \mu^*$
$S_1$	(0, 9.11)
$S_2$	(0, 12.11)
$B_1$	(0, 4.11)
$B_2$	(0, 11.11)
$R_1$	(0, 14.11)
$R_2$	(0, 6.11)

## SAS code for Dunnett, LSD, Tukey, SNL:

```
ods html; ods graphics on;
option ls=80 ps=50 nocenter nodate;
title 'Stopping Distance of 6 brands of tires';
data old; array Y Y1-Y4;
input BRD $ Y1-Y4; do over Y; SD=Y; output; end;
      drop Y1-Y4;
      label BRD = 'Brand of Tire' SD = 'Stopping Distance';
cards;
S1 22 20 25 17
S2 26 22 27 21
B1 16 20 14 18
B2 20 25 26 21
R1 28 29 23 24
R2 22 15 19 16
run;
proc glm data=old order=data;
class BRD;
model SD=BRD;
run;
lsmeans BRD/cl pdiff alpha=.05 adjust=tukey;
lsmeans BRD/cl pdiff=controll('S1') adjust=DUNNETT alpha=.05;
run;
means BRD/hovtest=bf;
output out=ASSUMP r=RESID p=MEANS;
proc univariate def=5 plot normal; var RESID;
run;
proc gplot;
plot SD*BRD='*';
run;
ods graphics off;ods html close;
```

1. **pdiff=control('S1') adjust=DUNNETT** provides a 2-sided Dunnett comparison with the treatment **S1** designated as the control and tests  $H_1 : \mu_i \neq \mu_{S1}$
2. **pdiff=controll('S1') adjust=DUNNETT** provides a 1-sided Dunnett comparison with the treatment **S1** designated as the control and tests  $H_1 : \mu_i < \mu_{S1}$
3. **pdiff=controlu('S1') adjust=DUNNETT** provides a 1-sided Dunnett comparison with the treatment **S1** designated as the control and tests  $H_1 : \mu_i > \mu_{S1}$

The GLM Procedure

Class Level Information

Class	Levels	Values
BRD	6	B1 B2 R1 R2 S1 S2

Number of Observations Read	24
Number of Observations Used	24

Dependent Variable: SD Stopping Distance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	246.0000000	49.2000000	5.47	0.0031
Error	18	162.0000000	9.0000000		
Corrected Total	23	408.0000000			

Least Squares Means

Adjustment for Multiple Comparisons: Tukey

BRD	SD LSMEAN	LSMEAN Number
S1	21.0000000	1
S2	24.0000000	2
B1	17.0000000	3
B2	23.0000000	4
R1	26.0000000	5
R2	18.0000000	6

Least Squares Means for effect BRD  
Pr > |t| for H0: LSMean(i)=LSMean(j)  
Dependent Variable: SD

i/j	1	2	3	4	5	6
1		0.7185	0.4412	0.9298	0.2229	0.7185
2	0.7185		0.0392	0.9966	0.9298	0.0979
3	0.4412	0.0392		0.0979	0.0055	0.9966
4	0.9298	0.9966	0.0979		0.7185	0.2229
5	0.2229	0.9298	0.0055	0.7185		0.0149
6	0.7185	0.0979	0.9966	0.2229	0.0149	

BRD	SD LSMEAN	95% Confidence Limits
S1	21.000000	17.848617, 24.151383
S2	24.000000	20.848617, 27.151383
B1	17.000000	13.848617, 20.151383
B2	23.000000	19.848617, 26.151383
R1	26.000000	22.848617, 29.151383
R2	18.000000	14.848617, 21.151383

Least Squares Means for Effect BRD

i j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)
-----	--------------------------	--

1 2	-3.000000	-9.741630, 3.741630
1 3	4.000000	-2.741630, 10.741630
1 4	-2.000000	-8.741630, 4.741630
1 5	-5.000000	-11.741630, 1.741630
1 6	3.000000	-3.741630, 9.741630
2 3	7.000000	0.258370, 13.741630
2 4	1.000000	-5.741630, 7.741630
2 5	-2.000000	-8.741630, 4.741630
2 6	6.000000	-0.741630, 12.741630
3 4	-6.000000	-12.741630, 0.741630
3 5	-9.000000	-15.741630, -2.258370
3 6	-1.000000	-7.741630, 5.741630
4 5	-3.000000	-9.741630, 3.741630
4 6	5.000000	-1.741630, 11.741630
5 6	8.000000	1.258370, 14.741630

The GLM Procedure

Least Squares Means

Adjustment for Multiple Comparisons: Dunnett

BRD	SD LSMEAN	H0:LSMean=Control Pr < t
S1	21.0000000	
S2	24.0000000	0.9947
B1	17.0000000	0.1263
B2	23.0000000	0.9799
R1	26.0000000	0.9997
R2	18.0000000	0.2566

Least Squares Means for Effect BRD

i j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)
-----	--------------------------	--

2 1	3.000000	-Infinity, 8.106229
3 1	-4.000000	-Infinity, 1.106229
4 1	2.000000	-Infinity, 7.106229
5 1	5.000000	-Infinity, 10.106229
6 1	-3.000000	-Infinity, 2.106229



## SUMMARY OF TREATMENT COMPARISONS - BRAND EXAMPLE

### I. Comparison of LSD, Tukey, SNK Procedures:

Technique	Groups of Brands with Similar Means	Pairs of Brands with Significantly Different Means
HSD	G1 = $[R_1, S_2, B_2, S_1]$ G2 = $[S_1, S_2, B_2, R_2]$ G3 = $[S_1, R_2, B_1, B_2]$	$[R_1, R_2]$ , $[B_1, R_1]$ $[S_2, B_1]$
SNK	G1 = $[R_1, S_2, B_2, S_1]$ G2 = $[S_1, B_2, R_2]$ G3 = $[S_1, R_2, B_1]$	$[R_1, R_2]$ , $[B_1, R_1]$ $[S_2, B_1]$ , $[S_2, R_2]$ $[B_2, B_1]$
LSD	G1 = $[R_1, S_2, B_2]$ G2 = $[S_1, S_2, B_2]$ G3 = $[S_1, R_2, B_1]$	$[S_1, R_1]$ , $[R_1, R_2]$ $[B_1, R_1]$ , $[S_2, R_2]$ $[S_2, B_1]$ , $[B_2, R_2]$ , $[B_2, B_1]$

Note that LSD declares the most pairs of treatments to be different, then SNK, and finally HSD declares the fewest pairs of treatments to be different.

### II. Comparison of All Brands to Brand $S_1$ (Standard Brand) Dunnett's Procedure:

This is a 1-sided test  $H_1 : \mu_i < \mu_{S_1}$  (shorter mean stopping distances). Based on the Dunnett procedure our conclusion is that there is not significant evidence that any of the brands have a shorter mean stopping distance than Brand  $S_1$ .

### III. Hsu's Procedure for finding the "Best" Treatment In the Brand Example the "Best" treatment is the treatment having **Smallest** $\mu_i$

With 95% confidence ( $\alpha = 0.05$ ), the Collection containing the treatment having the "Best" mean consists of

$\{S_1, B_1, R_2\}$ .

With 99% confidence ( $\alpha = 0.01$ ), the Collection containing the treatment having the "Best" mean consists of

$\{S_1, B_1, B_2, R_2\}$ .

## Analysis of Treatment Means

### When Treatment is Levels of Quantitative Variable

The following example will be used to illustrate the many ways in which we can evaluate the types of differences that may exist in treatment means when the levels of the treatments are quantitative.

#### EXAMPLE

An experiment is run to examine the relationship between the temperature of the reaction and the percent yield of reaction. The treatment in this example is the temperature of a chemical reaction. There are 5 equally spaced values for the temperatures with four replications at each temperature. Thus, 4 orthogonal polynomial contrasts will be used to examine the effects of temperature on the response variable.

	PerCent Yield				
Temperature	Rep 1	Rep 2	Rep 3	Rep 4	$\bar{y}_i$
550	6	4	5	5	5
600	32	26	24	22	26
650	45	45	44	34	42
700	63	62	44	39	52
750	87	85	72	80	81

A plot of the data is given on the next page. There appears to be an increasing trend in the data for increasing reaction temperature. Does this observed trend hold in the treatment means or is it just an artifact of the observed data?

When the treatments consist of  $t$  equally spaced values of a quantitative variable and there are an equal number of replications per treatment level, it is possible to construct  $t - 1$  contrasts which reflect the reduction in  $SSE$  due to fitting an increasing higher order polynomial model relating the response variable to the treatment variable.

In the above example, the response variable is  $y_{ij}$  the % yield of the  $j$ th replication of a chemical reaction run using temperature level  $T_i$  for  $i = 1, 2, 3, 4, 5$  and  $j = 1, 2, 3, 4$ . The treatment levels are 550, 600, 650, 700, and 750 which are equally spaced and there are an equal number of replications per treatment:  $n_i = 4$  for all  $i$ . It is thus possible to fit a polynomial of order  $t - 1 = 4$  relating % yield to temperature. The following notation will describe the general setting of such an experiment.

1. Let  $X$  be a treatment with  $t$  equally spaced quantitative levels used in the experiment:  $X_1, X_2, \dots, X_t$ :  $X_i = X_1 + (i - 1)d$ , where  $d$  is the common spacing between treatment levels.

With  $t=5$ ,  $X_1 = 550$ ,  $d = 50 \Rightarrow X_i = 500 + (i - 1)50$  for  $i = 2, 3, 4, 5$

2. Suppose  $r$  independent experiments are conducted at each level, yielding average responses:  $\bar{y}_1, \bar{y}_2, \dots, \bar{y}_t$ .
3. Using the contrasts given in Table XI,  $t - 1$  mutually orthogonal contrasts can be constructed which reflect the gain in model fit by fitting a higher order polynomial to the experimental data.
4. This set of contrasts provide a decomposition of  $SS_{TRT}$  into  $t - 1$  independent components.

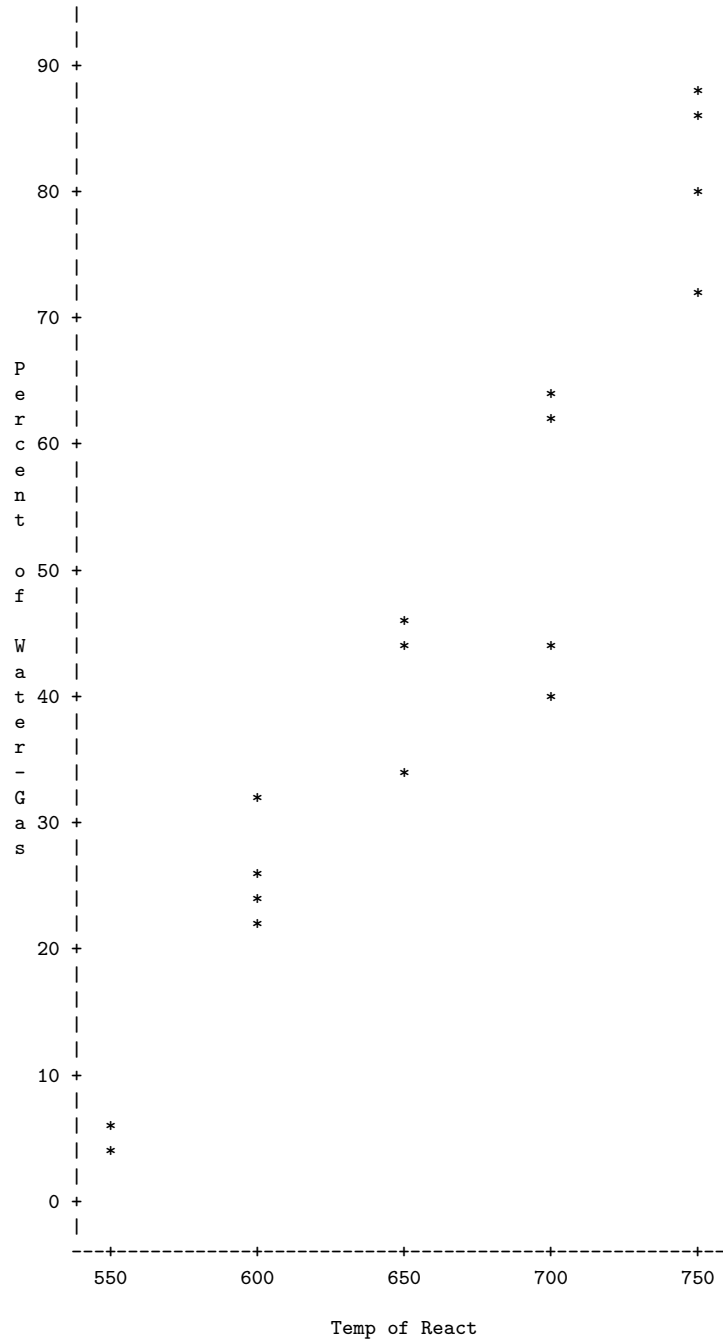
**Table XI** Orthogonal polynomials *From Kuehl, "Design of Experiments"*

	$t = 3$			$t = 4$			$t = 5$				$t = 6$					$t = 7$					
	$X_j$	$P_1$	$P_2$	$P_1$	$P_2$	$P_3$	$P_1$	$P_2$	$P_3$	$P_4$	$P_1$	$P_2$	$P_3$	$P_4$	$P_5$	$P_1$	$P_2$	$P_3$	$P_4$	$P_5$	$P_6$
1	-1	1	-3	1	-1	-2	2	-1	1	-5	5	-5	1	-1	-3	5	-1	3	-1	1	
2	0	-2	-1	-1	3	-1	-1	2	-4	-3	-1	7	-3	5	-2	0	1	-7	4	-6	
3	1	1	1	-1	-3	0	-2	0	6	-1	-4	4	2	-10	-1	-3	1	1	-5	15	
4			3	1	1	1	-1	-2	-4	1	-4	-4	2	10	0	-4	0	6	0	-20	
5						2	2	1	1	3	-1	-7	-3	-5	1	-3	-1	1	5	15	
6										5	5	5	1	1	2	0	-1	-7	-4	-6	
7															3	5	1	3	1	1	
$\sum_{j=1}^t \{P_i(X_j)\}^2$		2	6	20	4	20	10	14	10	70	70	84	180	28	252	28	84	6	154	84	924
$\lambda$		1	3	2	1	$\frac{10}{3}$	1	1	$\frac{5}{6}$	$\frac{35}{12}$	2	$\frac{3}{2}$	$\frac{5}{3}$	$\frac{7}{12}$	$\frac{21}{10}$	1	1	$\frac{1}{6}$	$\frac{7}{12}$	$\frac{7}{20}$	$\frac{77}{60}$

	$t = 8$						$t = 9$						$t = 10$						
	$X_j$	$P_1$	$P_2$	$P_3$	$P_4$	$P_5$	$P_6$	$P_1$	$P_2$	$P_3$	$P_4$	$P_5$	$P_6$	$P_1$	$P_2$	$P_3$	$P_4$	$P_5$	$P_6$
1	-7	7	-7	7	-7	1	-4	28	-14	14	-4	4	-9	6	-42	18	-6	3	
2	-5	1	5	-13	23	-5	-3	7	7	-21	11	-17	-7	2	14	-22	14	-11	
3	-3	-3	7	-3	-17	9	-2	-8	13	-11	-4	22	-5	-1	35	-17	-1	10	
4	-1	-5	3	9	-15	-5	-1	-17	9	9	-9	1	-3	-3	31	3	-11	6	
5	1	-5	-3	9	15	-5	0	-20	0	18	0	-20	-1	-4	12	18	-6	-8	
6	3	-3	-7	-3	17	9	1	-17	-9	9	9	1	1	-4	-12	18	6	-8	
7	5	1	-5	-13	-23	-5	2	-8	-13	-11	4	22	3	-3	-31	3	11	6	
8	7	7	7	7	7	1	3	7	-7	-21	-11	-17	5	-1	-35	-17	1	10	
9							4	28	14	14	4	4	7	2	-14	-22	-14	-11	
10													9	6	42	18	6	3	
$\sum_{j=1}^t \{P_i(X_j)\}^2$		168	168	264	616	2184	264	60	2772	990	2002	468	1980	330	132	8580	2860	780	660
$\lambda$		2	1	$\frac{2}{3}$	$\frac{7}{12}$	$\frac{7}{10}$	$\frac{11}{60}$	1	3	$\frac{5}{6}$	$\frac{7}{12}$	$\frac{3}{20}$	$\frac{11}{60}$	2	$\frac{1}{2}$	$\frac{5}{3}$	$\frac{5}{12}$	$\frac{1}{10}$	$\frac{11}{240}$

Adapted from *Biometrika Tables for Statisticians*, Vol. 1, 1966, edited by E. S. Pearson and H. O. Hartley, with permission of the Biometrika Trustees.

Plot of PE\*T. Symbol used is '\*'.



NOTE: 3 obs hidden.

To illustrate the above idea, consider the simplest possible case, fitting a first order polynomial to the  $t$  data points:  $(X_1, \bar{y}_1.), (X_2, \bar{y}_2.), \dots, (X_t, \bar{y}_t.) \Rightarrow \bar{y}_i = \beta_o + \beta_1 X_i + e_i$

Let  $\bar{X} = \frac{1}{t} \sum_{i=1}^t X_i$ ,  $SS_X = \sum_{i=1}^t (X_i - \bar{X})^2$ ,  $k_i = \frac{X_i - \bar{X}}{SS_X}$ ,  $i = 1, \dots, t$

Define the contrast:  $C = \sum_{i=1}^t k_i \mu_i$  where  $\mu_i = E[\bar{y}_i]$  the  $i$ th Treatment Mean

What does this contrast measure?

$$\begin{aligned} \hat{C} &= \sum_{i=1}^t k_i \bar{y}_i. \\ &= \sum_{i=1}^t \frac{X_i - \bar{X}}{SS_X} \bar{y}_i. \\ &= \frac{\sum_{i=1}^t (X_i - \bar{X}) \bar{y}_i.}{SS_X} \\ &= \frac{\sum_{i=1}^t (X_i - \bar{X}) (\bar{y}_i. - \bar{y}.)}{\sum_{i=1}^t (X_i - \bar{X})^2} \\ &= \hat{\beta}_1 \quad \text{the LSE of the slope of the line relating } \mu_{Y|X} \text{ to } X \end{aligned}$$

Thus,  $\hat{C} = 0$  if and only if  $\hat{\beta}_1 = 0$ . The test of  $H_o : C = 0$  is equivalent to the test of  $H_o : \beta_1 = 0$ . Thus,  $C$  measures whether or not there is a 1st order polynomial relationship between the mean response and the explanatory variable  $X$ . The decomposition of  $SS_{TRT}$  into the  $t - 1$  components extends the 1st order model to a  $t - 1$  order polynomial.

**Goal: Model  $\mu_{y|X}$  as a polynomial function of  $X$ :**

$$y_i = \beta_o + \beta_1 X_i + \beta_2 X_i^2 + \dots + \beta_{t-1} X_i^{t-1} + e_i,$$

where the values of  $X$  are equally spaced:

$$X_i = X_1 + (i - 1)d \text{ for } i = 1, \dots, t \text{ and } e_i \text{ are iid r.v.'s with distribution } N(0, \sigma_e^2).$$

Observe  $r$  values of  $y$  for each value of  $X$  yielding:

$$(X_1, \bar{y}_1.), (X_2, \bar{y}_2.), \dots, (X_t, \bar{y}_t.)$$

Determine the reduction in the value of  $SSE$  for each added term in the polynomial:

$$\text{Model 0: } y_i = \beta_o + e_i$$

$$\text{Model 1: } y_i = \beta_o + \beta_1 X_i + e_i$$

$$\text{Model 2: } y_i = \beta_o + \beta_1 X_i + \beta_2 X_i^2 + e_i$$

$\vdots$

$$\text{Model } t - 1: y_i = \beta_o + \beta_1 X_i + \beta_2 X_i^2 + \dots + \beta_{t-1} X_i^{t-1} + e_i$$

Sequentially compare the above models:

Let  $SSE_i$  be the sum of squares error from Model  $i$  and

let  $SSE_{ch,i} = SSE_{i-1} - SSE_i$  ; for  $i = 1, 2, \dots, t-2$

be the gain in the fit of the model due to adding another term to the current model.

Equation 1 is given by

**Equation 1:**  $\bar{y}_i = \beta_o + \beta_1 X_i + \beta_2 X_i^2 + \dots + \beta_{t-1} X_i^{t-1} + e_i$ ,

is written in matrix form as:

$\bar{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ , where

$$\bar{\mathbf{y}} = \begin{bmatrix} \bar{y}_1. \\ \bar{y}_2. \\ \vdots \\ \bar{y}_t. \end{bmatrix}; \quad \mathbf{X} = \begin{bmatrix} 1 & X_1 & X_1^2 & \dots & X_1^{t-1} \\ 1 & X_2 & X_2^2 & \dots & X_2^{t-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_t & X_t^2 & \dots & X_t^{t-1} \end{bmatrix}; \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_o \\ \beta_1 \\ \vdots \\ \beta_{t-1} \end{bmatrix}; \quad \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_t \end{bmatrix}$$

To obtain the orthogonal polynomials, first it is necessary to standardize the  $X$  variable:

$$X_i = X_1 + (i-1)d \Rightarrow \bar{X}_i = X_1 + \frac{d}{2}(t-1) \Rightarrow$$

$$Z_i = \frac{X_i - \bar{X}_i}{d} = i - \frac{t+1}{2} \text{ for } i = 1, 2, \dots, t$$

Rewrite the model in terms of the  $Z_i$ 's:

**Equation 2:**  $\bar{y}_i = \gamma_o + \gamma_1 Z_i + \gamma_2 Z_i^2 + \dots + \gamma_{t-1} Z_i^{t-1} + e_i$

Define the  $t-1$  polynomials in  $Z_i$  as:

$$P_o(Z_i) = 1; \quad P_1(Z_i) = Z_i; \quad P_2(Z_i) = Z_i^2 - 2$$

$$P_{l+1} = Z_i P_l(Z_i) - \frac{l^2(t^2-l^2)}{4(4l^2-1)} P_{l-1}(Z_i) \text{ for } l = 2, \dots, t-2$$

The above polynomials, after converting them to integers, yield the coefficients in Table XI in Kuehl's book (see page 44 of this handout).

The regression equation can now be written in terms of these polynomials:

**Equation 3:**  $\bar{y}_i = \alpha_o + \alpha_1 P_1(Z_i) + \alpha_2 P_2(Z_i) + \dots + \alpha_{t-1} P_{t-1}(Z_i) + e_i$

Express the above model in matrix form as  $\bar{\mathbf{y}} = \mathbf{P}\boldsymbol{\alpha} + \mathbf{e}$ , where

$$\bar{\mathbf{y}} = \begin{bmatrix} \bar{y}_1. \\ \bar{y}_2. \\ \vdots \\ \bar{y}_t. \end{bmatrix}; \quad \mathbf{P} = \begin{bmatrix} P_o(Z_1) & P_1(Z_1) & \cdots & P_{t-1}(Z_1) \\ P_o(Z_2) & P_1(Z_2) & \cdots & P_{t-1}(Z_2) \\ \vdots & \vdots & \ddots & \vdots \\ P_o(Z_t) & P_1(Z_t) & \cdots & P_{t-1}(Z_t) \end{bmatrix}; \quad \boldsymbol{\alpha} = \begin{bmatrix} \alpha_o \\ \alpha_1 \\ \vdots \\ \alpha_{t-1} \end{bmatrix}; \quad \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

The LSE of  $\boldsymbol{\alpha}$  is obtained as  $\hat{\boldsymbol{\alpha}} = (\mathbf{P}'\mathbf{P})^{-1}(\mathbf{P}'\bar{\mathbf{y}})$

Because the columns of  $\mathbf{P}$  are orthogonal and the sum of each column is 0, we obtain:

$$\mathbf{P}'\mathbf{P} = \text{Diag} \left( t, \sum_{i=1}^t P_1^2(Z_i), \sum_{i=1}^t P_2^2(Z_i), \dots, \sum_{i=1}^t P_{t-1}^2(Z_i) \right).$$

Therefore,

$$\hat{\alpha}_l = \frac{\sum_{i=1}^t P_l(Z_i) \bar{y}_{i.}}{\sum_{i=1}^t P_l^2(Z_i)} \quad \text{and} \quad SS_{P_l} = \frac{\left( \sum_{i=1}^t P_l(Z_i) \bar{y}_{i.} \right)^2}{\frac{1}{r} \sum_{i=1}^t P_l^2(Z_i)}; \quad l = 1, 2, \dots, t-1$$

Note that  $\hat{\alpha}_l = \frac{\hat{C}_l}{\sum_{i=1}^t k_{li}^2}$ , where  $k_{li} = P_l(Z_i)$ ,  $\hat{C}_l = \sum_{i=1}^t k_{li} \bar{y}_{i.}$  and  $SS_{P_l} = SS_{C_l} = \frac{\hat{C}_l^2}{\frac{1}{n} \sum_{i=1}^t K_{li}^2}$ .

The sum of squares  $SS_{P_l}$  is the resulting decrease in  $SSE$  by including the term  $X^l$  in a model containing the terms  $X, X^2, \dots, X^{l-1}$ :

$$SS_{P_1} = SS_{TOT} - SSE_1$$

$$SS_{P_2} = SSE_1 - SSE_2$$

$$\vdots$$

$$SS_{P_{t-1}} = SSE_{t-2} - SSE_{t-1}$$

The LSE of the coefficients in the original model,  $\hat{\beta}_i$  can be obtained from the  $\hat{\alpha}_i$ 's:

$$\text{In the model } \hat{y} = \hat{\alpha}_o + \hat{\alpha}_1 P_1(Z) + \hat{\alpha}_2 P_2(Z) + \cdots + \hat{\alpha}_{t-1} P_{t-1}(Z)$$

Replace  $Z$  with  $\frac{X-\bar{X}}{d}$  and expand the polynomial to obtain:

$$\hat{y} = \hat{\beta}_o + \hat{\beta}_1 X + \hat{\beta}_2 X^2 + \cdots + \hat{\beta}_{t-1} X^{t-1}$$

The above ideas will be illustrated using the Chemical Reaction Experiment.

There are  $t = 5$  equally spaced levels of the treatment factor, temperature of reaction, with  $r = 4$  independent experiments at each of the five temperature levels. The  $t - 1 = 4$  mutually orthogonal contrasts are obtained as follows: (See Table IX):

$$Z_i = i - \frac{t+1}{2} \text{ for } i = 1, 2, 3, 4, 5 \text{ yielding } Z = -2, -1, 0, 1, 2$$

$$P_o(Z) = 1 \text{ for all } Z$$

$$P_1(Z) = Z$$

$$P_{l+1} = Z_i P_l(Z_i) - \frac{l^2(t^2-l^2)}{4(4l^2-1)} P_{l-1}(Z_i) \text{ for } l = 2, \dots, t-2 \Rightarrow$$

$$P_2(Z) = Z P_1(Z) - \frac{(t^2-1)}{4(4-1)} P_o(Z) = Z^2 - 2$$

$$P_3(Z) = Z P_2(Z) - \frac{(2)^2(t^2-(2)^2)}{4(4(2)^2-1)} P_1(Z) = Z(Z^2 - 2) - \frac{7}{5} Z = Z^3 - \frac{17}{5} Z$$

$$P_4(Z) = Z P_3(Z) - \frac{(3)^2(t^2-(3)^2)}{4(4(3)^2-1)} P_2(Z) = Z(Z^3 - \frac{17}{5}) - \frac{36}{35}(Z^2 - 2) = Z^4 - \frac{31}{7} Z^2 + \frac{72}{35}$$

The values given for the 4 contrasts are obtained by evaluating the above for  $Z = -2, -1, 0, 1, 2$  and multiplying by a common denominator to obtain integer values for the five coefficients. We can compare the values obtained from the values given in Table IX:

For example,  $P_3(Z) = \frac{-6}{5}$  for  $Z = -2$ ;  $\frac{12}{5}$  for  $Z = -1$ ;  $0$  for  $Z = 0$ ;  $\frac{-12}{5}$  for  $Z = 1$ ;  $\frac{6}{5}$  for  $Z = 2$ ;

Multiplying the five values by  $\frac{5}{6}$ , yields the coefficients given in Table IX:  $-1, 2, 0, -2, 1$

		Polynomial				
		Mean	Linear	Quadratic	Cubic	Quartic
Temp( $X_i$ )	$\bar{y}_i$	$P_{l0}$	$P_{l1}$	$P_{l2}$	$P_{l3}$	$P_{l4}$
550	5	1	-2	2	-1	1
600	26	1	-1	-1	2	-4
650	42	1	0	-2	0	6
700	52	1	1	-1	-2	-4
750	81	1	2	2	1	1
Multiplier	$\lambda_l$	1	1	1	5/6	35/12
$\hat{C}_l = \sum_{i=1}^t P_{li} \bar{y}_i$		206	178	10	24	26
$D_l = \sum_{i=1}^t P_{li}^2$		5	10	14	10	70
$SS_{P_l} = \frac{\hat{C}_l^2}{\frac{1}{r} D_l}$			12673.6	28.571	230.4	38.6286
$\hat{\alpha}_l = \frac{\hat{C}_l}{D_l}$		41.2	17.8	5/7	2.4	13/35

From the above table, we obtain the following AOV for Regression:



SV	DF	SS	MS	F	p-value
Model	4	12971.2	3242.8	66.45	$2 \times 10^{-9}$
Linear	1	12673.6	12673.6	259.7	$7 \times 10^{-11}$
Quadratic	1	28.6	28.6	0.59	0.4560
Cubic	1	230.4	230.4	4.72	0.0462
Quartic	1	38.6	38.6	0.79	0.3877
Error	15	732	48.80		

The predictive model can also be obtained:

$$\hat{y} = \lambda_o \hat{\alpha}_o + \lambda_1 \hat{\alpha}_1 P_1(Z) + \lambda_2 \hat{\alpha}_2 P_2(Z) + \lambda_3 \hat{\alpha}_3 P_3(Z) + \lambda_4 \hat{\alpha}_4 P_4(Z)$$

It is necessary to multiply the coefficients  $\alpha_l$  by the multipliers  $\lambda_l$  because we had converted the actual coefficients in the orthogonal contrasts into integers.

$$\hat{y} = (1)(41.2) + (1)(17.8)P_1(Z) + (1)\left(\frac{5}{7}\right)P_2(Z) + \left(\frac{5}{6}\right)(2.4)P_3(Z) + \left(\frac{35}{12}\right)\left(\frac{13}{35}\right)P_4(Z)$$

$$\hat{y} = 41.2 + (17.8)P_1(Z) + \left(\frac{5}{7}\right)P_2(Z) + (2)P_3(Z) + \left(\frac{13}{12}\right)P_4(Z)$$

The equation can be written in terms of  $X$  by substituting  $Z = \frac{X - \bar{X}}{d} = \frac{X - 650}{50}$  into the above equation and collecting terms:

$$\begin{aligned} \hat{y} = & 41.2 + (17.8) \left( \frac{X - 650}{50} \right) + \left( \frac{5}{7} \right) \left( \left( \frac{X - 650}{50} \right)^2 - 2 \right) + (2) \left( \left( \frac{X - 650}{50} \right)^3 - \frac{17}{5} \left( \frac{X - 650}{50} \right) \right) \\ & + \frac{13}{12} \left( \left( \frac{X - 650}{50} \right)^4 - \frac{31}{7} \left( \frac{X - 650}{50} \right)^2 + \frac{72}{35} \right) \end{aligned}$$

which simplifies to

$$\hat{y} = 25756 - 167.783X + 0.40657X^2 - 0.0004347X^3 + 0.000000173X^4$$

Examine the following SAS program to compare the results obtained above to the results from directly running a regression analysis.

# SAS Program for the Orthogonal Polynomial Example

```
* orthopoly.sas;
ods html;ods graphics on;
option ls=72 ps=60 nocenter nodate;
title 'Orthogonal Polynomials';

data poly; array Y Y1-Y4;
input T Y1-Y4; do over Y; PE=Y; output; end;
      drop Y1-Y4;
      label T = 'Temp of React' PE = 'Percent of Water-Gas';
cards;
550 6 4 5 5
600 32 26 24 22
650 45 45 44 34
700 63 62 44 39
750 87 85 72 80
run;
proc plot;
plot PE*T='*';
run;
proc glm;
class T;
model PE = T/ss3;
contrast 'LINEAR' T -2 -1 0 1 2;
contrast 'QUADRATIC' T 2 -1 -2 -1 2;
contrast 'CUBIC' T -1 2 0 -2 1;
contrast 'QUARTIC' T 1 -4 6 -4 1;

*simultaneous test of all 4 contrasts;
contrast '4 TREND CONTRASTS' T -2 -1 0 1 2,
                             T 2 -1 -2 -1 2,
                             T -1 2 0 -2 1,
                             T 1 -4 6 -4 1;
run; ods graphics off; ods html close;
```

OUTPUT FROM SAS PROGRAM: orthopoly.sas:

Dependent Variable: PE Percent of Water-Gas

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	12971.20000	3242.80000	66.45	<.0001
Error	15	732.00000	48.80000		
Corrected Total	19	13703.20000			

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
LINEAR	1	12673.60000	12673.60000	259.70	<.0001
QUADRATIC	1	28.57143	28.57143	0.59	0.4560
CUBIC	1	230.40000	230.40000	4.72	0.0462
QUARTIC	1	38.62857	38.62857	0.79	0.3877
4 TREND CONTRASTS	4	12971.20000	3242.80000	66.45	<.0001

## Matrix Calculations for Combined Test of 4 Contrasts:

Recall from page 6 of this handout, the Sum of Squares needed to test the significance of the 4 Contrasts simultaneously is obtained as follows:

The test of  $H_o : C_1 = 0, C_2 = 0, C_3 = 0, C_4 = 0$  versus  $H_1 : \text{At least one } C_i \neq 0$

is obtained by formulating the Hypothesis matrix:

Test  $H_o : \mathbf{H}\boldsymbol{\mu} = 0$  vs  $H_o : \mathbf{H}\boldsymbol{\mu} \neq 0$  with

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \end{pmatrix}; \quad \mathbf{H} = \begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ 2 & -1 & -2 & -1 & 2 \\ -1 & 2 & 0 & -2 & 1 \\ 1 & -4 & 6 & -4 & 1 \end{pmatrix}; \quad \hat{\boldsymbol{\mu}} = \begin{pmatrix} \bar{y}_{1.} \\ \bar{y}_{2.} \\ \bar{y}_{3.} \\ \bar{y}_{4.} \\ \bar{y}_{5.} \end{pmatrix} = \begin{pmatrix} 5 \\ 26 \\ 42 \\ 52 \\ 81 \end{pmatrix};$$

$$(\mathbf{X}^T \mathbf{X}) = \text{Diag}(n_1, n_2, n_3, n_4, n_5) = \begin{pmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix}$$

The sum of squares associated  $\mathbf{H}$  is given by

$$SS_{\mathbf{H}} = (\mathbf{H}\hat{\boldsymbol{\mu}} - \mathbf{0})^T (\mathbf{H}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{H}^T)^{-1} (\mathbf{H}\hat{\boldsymbol{\mu}} - \mathbf{0}) = 12971.2$$

The following R code yields the above matrix calculations:

```
library(MASS)
H = matrix(c(-2,-1,0,1,2,2,-1,-2,-1,2,-1,2,0,-2,1,1,-4,6,-4,1),nrow=4, byrow=T)
muhat = matrix(c(5,26,42,52,81),nrow=5)
h = matrix(c(0,0,0,0),nrow=4)
x=rep(4,5)
D = diag(x,5,5)
A = H%*%muhat - h
Dinv = solve(D)
Cinv = solve(H%*%Dinv%*%t(H))
SSH = t(A)%*%Cinv%*% A
SSH
12971.2
```

The test statistic for testing  $H_o : \mathbf{H}\boldsymbol{\mu} = 0$  vs  $H_o : \mathbf{H}\boldsymbol{\mu} \neq 0$  is given by

$$F = \frac{SS_H/k}{MSE} = \frac{12971.2/4}{48.80} = 66.45 \Rightarrow \text{p-value} = P[F_{4,15} \geq 66.45] = 2.3 \times 10^{-9}$$

Next we will compare the analysis of the orthogonal polynomials to fitting regression models:

```
*orthoreg.sas
option ls=70 ps=55 nocenter nodate;
data poly;
array Y Y1-Y4;
input T Y1-Y4;
T2=T**2; T3=T**3; T4=T**4;
do over Y; PE=Y; output; end;
    drop Y1-Y4;
    label T = 'Temp of React' PE = 'Percent Yield of Reaction';
cards;
550 6 4 5 5
600 32 26 24 22
650 45 45 44 34
700 63 62 44 39
750 87 85 72 80
run;
proc reg data=poly ;
model PE= T T2 T3 T4/ ss1 ss2;
run;
proc reg data=poly ;
model PE= T T2 T3 / ss1 ss2;
run;
proc reg data=poly ;
model PE= T T2 / ss1 ss2;
run;
proc reg data=poly ;
model PE= T / ss1 ss2;
run;
```

OUTPUT FROM SAS PROGRAM:  
 MODEL WITH 4TH ORDER POLYNOMIAL:

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	12971.20008	3242.80002	66.451	0.0001
Error	15	731.99992	48.79999		
C Total	19	13703.20000			

Root MSE	6.98570	R-square	0.9466
Dep Mean	41.20000	Adj R-sq	0.9323

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob >  T
INTERCEP	1	25756	33926.424847	0.759	0.4595
T	1	-167.783738	211.41191517	-0.794	0.4398
T2	1	0.406568	0.49192672	0.826	0.4215
T3	1	-0.000435	0.00050659	-0.858	0.4044
T4	1	0.000000173	0.00000019	0.890	0.3877

Variable	DF	Type I SS	Type II SS
INTERCEP	1	33949	28.125652
T	1	12674	30.736937
T2	1	28.571429	33.333775
T3	1	230.400000	35.926892
T4	1	38.628651	38.628651

-----  
 MODEL WITH 3RD ORDER POLYNOMIAL:

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	12932.57143	4310.85714	89.503	0.0001
Error	16	770.62857	48.16429		
C Total	19	13703.20000			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob >  T
INTERCEP	1	-4376.514287	1974.7238678	-2.216	0.0415
T	1	20.128571	9.22280000	2.182	0.0443
T2	1	-0.030914	0.01426996	-2.166	0.0457
T3	1	0.000016000	0.00000732	2.187	0.0439

Variable	DF	Type I SS	Type II SS
INTERCEP	1	33949	236.575123
T	1	12674	229.416922
T2	1	28.571429	226.046678
T3	1	230.400000	230.400000

-----  
MODEL WITH 2ND ORDER POLYNOMIAL:

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	12702.17143	6351.08571	107.858	0.0001
Error	17	1001.02857	58.88403		
C Total	19	13703.20000			

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob >  T
INTERCEP	1	-70.914286	171.97973568	-0.412	0.6852
T	1	-0.015429	0.53377392	-0.029	0.9773
T2	1	0.000286	0.00041017	0.697	0.4955

Variable	DF	Type I SS	Type II SS
INTERCEP	1	33949	10.011761
T	1	12674	0.049197
T2	1	28.571429	28.571429

-----  
MODEL WITH 1ST ORDER POLYNOMIAL:

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	12673.60000	12673.60000	221.566	0.0001
Error	18	1029.60000	57.20000		
C Total	19	13703.20000			

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob >  T
INTERCEP	1	-190.200000	15.63745504	-12.163	0.0001
T	1	0.356000	0.02391652	14.885	0.0001

Variable	DF	Type I SS	Type II SS
INTERCEP	1	33949	8462.231579
T	1	12674	12674

What exactly are the tests of  $H_o : C_i = 0$  versus  $H_1 : C_i \neq 0$ ,  $i = 1, \dots, t - 1$  evaluating where  $C_i$  is the contrast associated with the  $i$ th order polynomial?

We will explain their meaning through the chemical yield example.

We have  $t = 5$  levels of the variable  $X$  temperature of the reaction with five potential models:

Model 0:  $Y_i = \beta_o + e_i$  with  $SSE_0 = SS_{Total} = 13703.20000$

Model 1:  $Y_i = \beta_o + \beta_1 X_i + e_i$  with  $SSE_1 = 1029.60000$

Model 2:  $Y_i = \beta_o + \beta_1 X_i + \beta_2 X_i^2 + e_i$  with  $SSE_2 = 1001.02857$

Model 3:  $Y_i = \beta_o + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + e_i$  with  $SSE_3 = 770.62587$

Model 4:  $Y_i = \beta_o + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \beta_4 X_i^4 + e_i$  with  $SSE_4 = 731.99992$

Using the SSEs from the five models we have the following reductions in SSE due to adding another term to the regression model:

$SSE_{Red1} = SSE_0 - SSE_1 = 13703.20000 - 1029.60000 = 12673.6$  This Sum of Squares measures the gain in fitting a straight line with arbitrary slope to the data over fitting a horizontal line.

$SSE_{Red2} = SSE_1 - SSE_2 = 1029.60000 - 1001.02857 = 28.57143$  This Sum of Squares measures the gain in fitting a quadratic equation to the data over fitting a straight line.

$SSE_{Red3} = SSE_2 - SSE_3 = 1001.02857 - 770.62587 = 230.4027$  This Sum of Squares measures the gain in fitting a cubic equation to the data over fitting a quadratic equation.

$SSE_{Red4} = SSE_3 - SSE_4 = 770.62587 - 731.99992 = 38.62595$  This Sum of Squares measures the gain in fitting a quartic equation to the data over fitting a cubic equation.

The sum of squares from the Four contrasts are given below.

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
LINEAR	1	12673.600	12673.600	259.70	0.0001
QUADRATIC	1	28.571	28.571	0.59	0.4560
CUBIC	1	230.400	230.400	4.72	0.0462
QUARTIC	1	38.629	38.629	0.79	0.3877

Thus, we can now see that the Contrast SS are exactly what we obtained from fitting the four models to the data.

The p-value for the Linear Contrast is testing  $H_o : \beta_1 = 0$  vs  $H_o : \beta_1 \neq 0$  in the model:  $Y_i = \beta_o + \beta_1 X_i + e_i$

The p-value for the Quadratic Contrast is testing  $H_o : \beta_2 = 0$  vs  $H_o : \beta_2 \neq 0$  in the model:  $Y_i = \beta_o + \beta_1 X_i + \beta_2 X_i^2 + e_i$

The p-value for the Cubic Contrast is testing  $H_o : \beta_3 = 0$  vs  $H_o : \beta_3 \neq 0$  in the model:  $Y_i = \beta_o + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + e_i$

The p-value for the Quartic Contrast is testing  $H_o : \beta_4 = 0$  vs  $H_o : \beta_4 \neq 0$  in the model:  $Y_i = \beta_o + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \beta_4 X_i^4 + e_i$

When the selected levels of the treatment variable are unequally spaced, the following R code can be used to obtain the coefficients. It is not necessary to convert the coefficients to integers.

Suppose we have a variable X from which the levels of the treatment will be selected. The researcher wants to run experiment with a control (X=0) and the following 4 levels from X: 5, 15, 30, 50. The levels are unequally spaced so the table in the textbook cannot be used to select the coefficients for determining if the relationship between the response Y and the treatment variable X follow a  $t-1 = 5-1 = 4$ th degree polynomial.

```
con = contr.poly(5,scores=c(0,5,15,30,50))
con_std = matrix(0,5,4)
m = c(rep(0,4))
for (i in 1:4) {
m[i] =min(abs(con[,i]))
con_std[,i] = con[,i]/m[i]
}
```

con\_std

	[,1]	[,2]	[,3]	[,4]
[1,]	-4	6.000000	-2.666667	14
[2,]	-3	1.000000	2.000000	-28
[3,]	-1	-5.666667	3.333333	20
[4,]	2	-7.333333	-3.666667	-7
[5,]	6	6.000000	1.000000	1

```
con_std[,2] = 3*con_std[,2]
con_std[,3] = 3*con_std[,3]
```

con\_std

	[,1]	[,2]	[,3]	[,4]
[1,]	-4	18	-8	14
[2,]	-3	3	6	-28
[3,]	-1	-17	10	20
[4,]	2	-22	-11	-7
[5,]	6	18	3	1

The four contrasts would thus be given by

$$C_{linear} = -4\mu_1 + -3\mu_2 - \mu_3 + 2\mu_4 + 6\mu_5$$

$$C_{quadratic} = 18\mu_1 + 3\mu_2 - 17\mu_3 + -22\mu_4 + 18\mu_5$$

$$C_{cubic} = -8\mu_1 + 6\mu_2 + 10\mu_3 + -11\mu_4 + 3\mu_5$$

$$C_{quartic} = 14\mu_1 + -28\mu_2 + 20\mu_3 + -7\mu_4 + \mu_5$$



## SUMMARY of Procedures

1. Testing a Single Contrast: Use F-test with  $\alpha_C = \alpha_F$ , FamilyWiseErrorRate (FWER). Yields exact result for  
 $\Pr[\text{Type I Error}]$ .
2. Testing a fixed number,  $M$ , of contrasts - Selected prior to running the experiment.
  - (a) If  $M$  Contrasts are mutually orthogonal: Use F-test with  $\alpha_C = 1 - (1 - \alpha_F)^{1/M}$ . Yields an upper bound of size  $\alpha_F$  on  $\Pr[\text{Experiment-wise Type I Error}]$ .
  - (b) If  $M$  Contrasts are NOT mutually orthogonal: Use F-test with  $\alpha_C = \alpha_F/M$  (*Bonferroni F-test*) Yields an upper bound of size  $\alpha_F$  on  $\Pr[\text{Experiment-wise Type I Error}]$ .
3. Although it is never a good idea to test contrasts selected after the experiment has been run, the Scheffé Procedure would be the procedure to be used in the unusual circumstances when this is necessary. Scheffé's Procedure handles an unspecified number of contrasts. Scheffé's Procedure yields Exact Result with  
 $\Pr[\text{Experiment-wise Type I Error}] = \alpha_F$ .
4. For comparing ALL possible Pairs of Treatment Means,  $\mu'_i$ s:
  - (a) Strongly recommend against using Fisher's Protected LSD and SNK due to the lack of an exact value for FWER
  - (b) In nearly all cases use Tukey's HSD when sample sizes are equal.
  - (c) When the  $t$  sample sizes are unequal, use the Tukey-Kramer in place of Tukey.
5. When comparing a Control or Standard Treatment to a Group of Treatments: Use Dunnett's Procedure.
6. When selecting a subset of Treatments containing the "BEST" Treatment: Use Hsu's Procedure.
7. When the treatments consist of equally spaced levels of an ordinal scaled variable, use Orthogonal Polynomial with  $\alpha_C = 1 - (1 - \alpha_F)^{1/M}$  to evaluate trends in the Treatment Means  $\mu'_i$ s. Yields an upper bound of size  $\alpha_F$  on  $\Pr[\text{Experiment-wise Type I Error}]$ .
8. Procedures yielding exact experimentwise error rates are Scheffé, Tukey's HSD, and Dunnett's procedures.
9. Procedures yielding results having a bounded experimentwise error rate are Bonferroni and Hsu's procedures.

10. Fisher's protected LSD and SNK have experimentwise error rates which are neither exact nor bounded by their nominal values.

11. References:

- (a) Miller, R.(1981), *Simultaneous Statistical Inferences, 2nd Edition*
- (b) Carmen and Swanson, *JASA, Vol. 68, 1973, pp. 66-74, Evaluation of 10 Pairwise Comparison Procedures by Monte Carlo Methods*
- (c) Benjamini and Hochberg(1995), **Controlling the false discovery rate, a practical and powerful approach to multiple testing**, *J.R.S.S. B*, Vol. 57, pp. 289-300.
- (d) Storey(2002), **A direct approach to false discovery rates**, *J.R.S.S. B*, Vol. 64, pp. 479-498.
- (e) Westfall, Tobias, Rom, Wolfinger, Hochberg (1999), **Multiple Comparisons and Multiple Tests**, SAS Institute, Inc.
- (f) Hsu, Jason.(1996), *Multiple Comparisons, Theory and Methods*