Lecture 5: Comparing Treatment Means

Montgomery: Sections 3.3-5

Linear Combinations of Treatment Means

- ANOVA Model:

$$
\begin{aligned}
y_{ij} &= \mu + \tau_i + \epsilon_{ij} \quad (\tau_i\text{: treatment effect}) \\
&= \mu_i + \epsilon_{ij} \quad (\mu_i\text{: treatment mean})
\end{aligned}
$$

- Linear combination with given coefficients $c_1, c_2, \ldots, c_a$:

$$L = c_1\mu_1 + c_2\mu_2 + \ldots + c_a\mu_a = \sum_{i=1}^{a} c_i\mu_i,$$

- Want to test: $H_0 : L = \sum c_i\mu_i = L_0$
- Examples:

1. Pairwise comparison: $\mu_i - \mu_j = 0$ for all possible i and j.
2. Compare treatment vs control: $\mu_i - \mu_1 = 0$ when treatment 1 is a control and $i = 2, ..., a$ are new treatments.
3. General cases such as $\mu_1 - 2\mu_2 + \mu_3 = 0$, $\mu_1 + 3\mu_2 - 6\mu_3 = 0$, etc.

- Estimate of L:

$$\hat{L} = \sum c_i \hat{\mu}_i = \sum c_i \bar{y}_{i.}$$

$$\text{Var}(\hat{L}) = \sum c_i^2 \text{Var}(\bar{y}_{i.}) = \sigma^2 \sum \frac{c_i^2}{n_i} \left(= \frac{\sigma^2}{n} \sum c_i^2 \right)$$

- Standard Error of $\hat{L}$

$$\text{S.E.}_{\hat{L}} = \sqrt{\text{MSE} \sum \frac{c_i^2}{n_i}}$$

- Test statistic

$$t_0 = \frac{(\hat{L} - L_0)}{S.E._{\hat{L}}} \sim t(N-a) \text{ under } H_0$$

Example: Lambs Diet Experiment

- Recall there are three diets and their treatment means are denoted by μ_1, μ_2 and μ_3. Suppose one wants to consider

$$L = \mu_1 + 2\mu_2 + 3\mu_3 = 6\mu + \tau_1 + 2\tau_2 + 3\tau_3$$

and test $H_0 : L = 60$.

```
data lambs;
 input diet wtgain@@;
 cards;
  1  8 1 16 1  9 2  9 2 16 2 21
  2 11 2 18 3 15 3 10 3 17 3  6
;
proc glm;
 class diet;
 model wtgain=diet;
 means diet;
 estimate 'l1' intercept 6 diet 1 2 3;
run;
```

Example: Lambs Diet Experiment

- SAS output

```
Level of            ------------wtgain-----------
diet          N              Mean            Std Dev

1             3       11.0000000       4.35889894
2             5       15.0000000       4.94974747
3             4       12.0000000       4.96655481

Dependent Variable: wtgain
                        Standard
Parameter    Estimate      Error     t Value   Pr > |t|
l1          77.0000000  8.88506862  8.67     <.0001
```

- $t_0 = (77.0 - 60)/8.89 = 1.91$

$$P - \text{value} = P(t \leq -1.91 \text{ or } t \geq 1.91 | t(12-3)) = .088$$

- Fail to reject $H_0 : \mu_1 + 2\mu_2 + 3\mu_3 = 60$ at $\alpha = 5\%$.

Contrasts

- $\Gamma = \sum_{i=1}^{a} c_i\mu_i$ is a contrast if $\sum_{i=1}^{a} c_i = 0$.

  Equivalently, $\Gamma = \sum_{i=1}^{a} c_i\tau_i$.

- Examples

  1. $\Gamma_1 = \mu_1 - \mu_2 = \mu_1 - \mu_2 + 0\mu_3 + 0\mu_4$,
     $c_1 = 1, c_2 = -1, c_3 = 0, c_4 = 0$
     Comparing μ_1 and μ_2.

  2. $\Gamma_2 = \mu_1 - 0.5\mu_2 - 0.5\mu_3 = \mu_1 - 0.5\mu_2 - 0.5\mu_3 + 0\mu_4$
     $c_1 = 1, c_2 = -0.5, c_3 = -0.5, c_4 = 0$
     Comparing μ_1 and the average of μ_2 and μ_3.

- Estimate of Γ:

$$C = \sum_{i=1}^{a} c_i\bar{y}_{i.}$$

- Test: $H_0 : \Gamma = 0$

$$t_0 = \frac{C}{\text{S.E.}_C} \sim t(N-a)$$

$$t_0^2 = \frac{(\sum c_i \bar{y}_{i.})^2}{\text{MSE} \sum \frac{c_i^2}{n_i}} = \frac{(\sum c_i \bar{y}_{i.})^2 / \sum c_i^2/n_i}{\text{MSE}} = \frac{\text{SS}_C/1}{\text{MSE}}$$

Under H_0, $t_0^2 \sim F_{1,N-a}$.

- Contrast Sum of Squares

$$\text{SS}_C = \left(\sum c_i \overline{y}_{i.}\right)^2 / \sum (c_i^2/n_i)$$

SS_C represents the amount of variation attributable Γ.

SAS Code (cont.sas)

Tensile Strength Example

```
options ls=80;

title1 'Contrast Comparisons';

data one;
 infile 'c:\saswork\data\tensile.dat';
 input percent strength time;

proc glm data=one;
 class percent;
 model strength=percent;
 contrast 'C1' percent 0 0 0 1 -1;
 contrast 'C2' percent 1 0 1 -1 -1;
 contrast 'C3' percent 1 0 -1 0 0;
 contrast 'C4' percent 1 -4 1 1 1;
```

Dependent Variable: STRENGTH

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	475.76000	118.94000	14.76	0.0001
Error	20	161.20000	8.06000		
Corrected Total	24	636.96000			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
PERCENT	4	475.76000	118.94000	14.76	0.0001

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
C1	1	291.60000	291.60000	36.18	0.0001
C2	1	31.25000	31.25000	3.88	0.0630
C3	1	152.10000	152.10000	18.87	0.0003
C4	1	0.81000	0.81000	0.10	0.7545

Orthogonal Contrasts

- Two contrasts $\{c_i\}$ and $\{d_i\}$ are **Orthogonal** if

$$\sum_{i=1}^{a} \frac{c_i d_i}{n_i} = 0 \quad (\sum_{i=1}^{a} c_i d_i = 0 \text{ for balanced experiments})$$

- Example
  $\Gamma_1 = \mu_1 + \mu_2 - \mu_3 - \mu_4$, So $c_1 = 1, c_2 = 1, c_3 = -1, c_4 = -1$.
  $\Gamma_2 = \mu_1 - \mu_2 + \mu_3 - \mu_4$. So $d_1 = 1, d_2 = -1, d_3 = 1, d_4 = -1$
  It is easy to verify that both Γ_1 and Γ_2 are contrasts. Furthermore,

  $c_1 d_1 + c_2 d_2 + c_3 d_3 + c_4 d_4 =$
  $1 \times 1 + 1 \times (-1) + (-1) \times 1 + (-1) \times (-1) = 0$. Hence, Γ_1 and Γ_2 are orthogonal to each other.

- A **complete set** of orthogonal contrasts $\mathcal{C} = \{\Gamma_1, \Gamma_2, \ldots, \Gamma_{a-1}\}$ if contrasts are mutually orthogonal and there does not exist a contrast orthogonal outside of $\mathcal{C}$ to all the contrasts in $\mathcal{C}$.

- If there are a treatments, $\mathcal{C}$ must contain $a-1$ contrasts.
- Complete set is not unique. For example, in the tensile strength example

$$\mathcal{C}_1 \text{ : includes : } \begin{array}{rcrrrrr} \Gamma_1 & = & (0, & 0, & 0, & 1, & -1) \\ \Gamma_2 & = & (1, & 0, & 1, & -1, & -1) \\ \Gamma_3 & = & (1, & 0, & -1, & 0, & 0) \\ \Gamma_4 & = & (1, & -4, & 1, & 1, & 1) \end{array}$$

$$\mathcal{C}_2 \text{ : includes : } \begin{array}{rcrrrrr} \Gamma_1' & = & (-2, & -1, & 0, & 1, & 2) \\ \Gamma_2' & = & (2, & -1, & -2, & -1, & 2) \\ \Gamma_3' & = & (-1, & 2, & 0, & -2, & 1) \\ \Gamma_4' & = & (1, & -4, & 6, & -4, & 1) \end{array}$$

Orthogonal Contrasts

- Orthogonal contrasts (estimates) are independent with each other.
- Suppose $C_1, C_2, \ldots, C_{a-1}$ are the estimates of the contrasts in a complete set of contrasts $\{\Gamma_1, \Gamma_2, \ldots, \Gamma_{a-1}\}$, then

$$\text{SS}_{\text{Treatment}} = \text{SS}_{C_1} + \text{SS}_{C_2} + \cdots + \text{SS}_{C_{a-1}}$$

- Recall in ANOVA, $F_0 = \frac{\text{MS}_{\text{Treatment}}}{\text{MSE}}$,

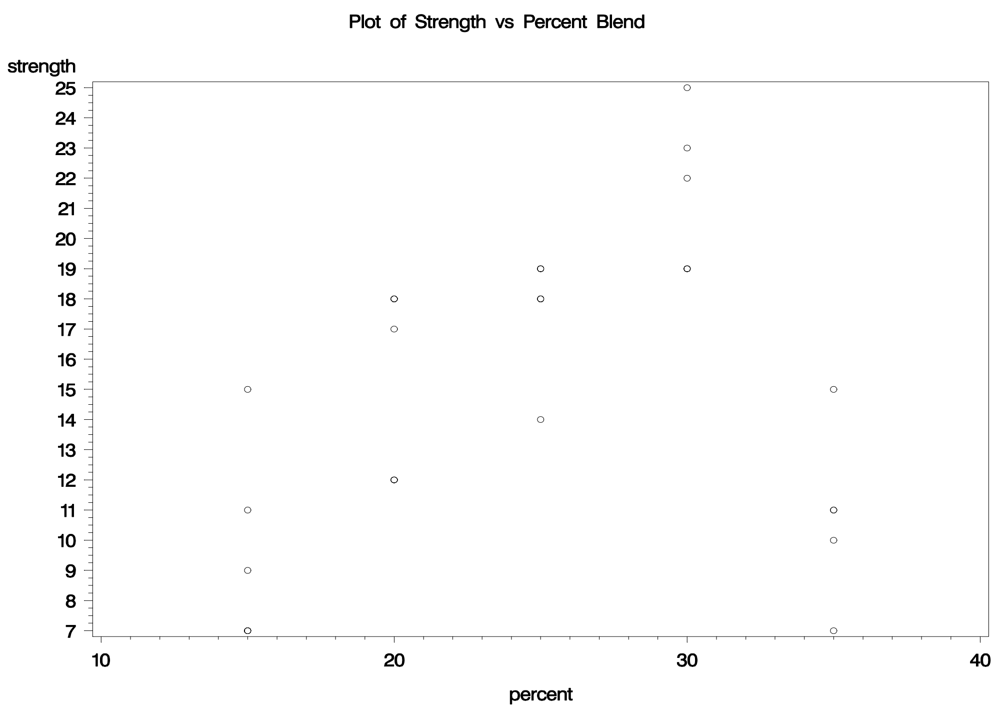
$$F_0 = \frac{\text{SS}_{C_1}/\text{MSE} + \cdots + \text{SS}_{C_{a-1}}/\text{MSE}}{a-1} = \frac{F_{10} + F_{20} + \cdots + F_{(a-1)0}}{a-1}$$

where F_{i0} is the test statistic used to test contrast Γ_i.

- Example on Slide 9

Tensile Example

Try to model mean response as a function of treatments



Orthogonal contrasts and orthogonal polynomial model

- Treatments are quantitative (assume $a = 4$)
- One can use general polynomial model to fit the trend (t: level or treatment).

$$f(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3$$

Regression can be used to get the estimates for a_1, a_2 and a_3.

- We will use orthogonal polynomial model

$$f(t) = \beta_0 + \beta_1 P_1(t) + \beta_2 P_2(t) + \beta_3 P_3(t)$$

where $P_1(t)$, $P_2(t)$ and $P_3(t)$ are pre-specified polynomials of order 1, 2 and 3, respectively. $P_1(t)$ is linear, $P_2(t)$ is quadratic and $P_3(t)$ is cubic. Let t_1, t_2, ..., t_a are the treatments (equally spaced), then the polynomials correspond to the following contrasts:

t	t_1	t_2	$\cdots$	t_a	Contrasts	$\mathcal{D}$
$P_1(t)$	$P_1(t_1)$	$P_1(t_2)$	$\cdots$	$P_1(t_a)$	Γ_1	$\mathcal{D}_1$
$P_2(t)$	$P_2(t_1)$	$P_2(t_2)$	$\cdots$	$P_2(t_a)$	Γ_2	$\mathcal{D}_2$
$P_3(t)$	$P_3(t_1)$	$P_3(t_2)$	$\cdots$	$P_3(t_a)$	Γ_3	$\mathcal{D}_3$

where

$$\mathcal{D}_i = P_i(t_1)^2 + P_i(t_2)^2 + \cdots + P_i(t_a)^2$$

If Γ_1, Γ_2 and Γ_3 are orthogonal to each other, then we say $P_1(t)$, $P_2(t)$ and $P_3(t)$ are orthogonal polynomials.

- Coefficients β_i can be estimated and tested by the contrasts Γ_1, Γ_2 and Γ_3.
- Predict $f(t)$ when t is not a treatment used in the experiment.

tensile strength example: orthogonal polynomial effects

- Treatment levels t_k: 15, 20, 25, 30, 35; Median: 25; Pace: 5
- Orthogonal polynomials: let $x = (t - 25)/5$.

$$P_1(t) = x$$
$$P_2(t) = x^2 - 2$$
$$P_3(t) = 5/6[x^3 - 17x/5]$$
$$P_4(t) = 35/12[x^4 - 31x/7 + 72/35]$$

- Polynomial Contrasts and Effects

t	15	20	25	30	35	Contrast	$\mathcal{D}$	Effect (Trend)
$P_1(t)$	-2	-1	0	1	2	Γ_1	$\mathcal{D}_1 = 10$	linear
$P_2(t)$	2	-1	-2	-1	2	Γ_2	$\mathcal{D}_2 = 14$	quadratic
$P_3(t)$	-1	2	0	-2	1	Γ_3	$\mathcal{D}_3 = 10$	cubic
$P_4(t)$	1	-4	6	-4	1	Γ_4	$\mathcal{D}_4 = 70$	4th order

- The contrasts can be directly derived from Table IX or Table X.

- Want to fit the model

$$f(t) = \beta_0 + \beta_1 P_1(t) + \beta_2 P_2(t) + \beta_3 P_3(t) + \beta_4 P_4(t)$$

- Estimation and Testing
  - β_1: use Γ_1,

$$\hat{\beta}_1 = \frac{c_{11}\bar{y}_{1.} + \cdots + c_{15}\bar{y}_{5.}}{\mathcal{D}_1}$$

    Test: $H_0 : \beta_1 = 0$, $F_{10} = \frac{\text{SS}_{C_i}}{\text{MSE}} \sim F_{1,N-5}$.

  - β_2: use Γ_2,

$$\hat{\beta}_2 = \frac{c_{21}\bar{y}_{1.} + \cdots + c_{25}\bar{y}_{5.}}{\mathcal{D}_2}$$

    Test: $H_0 : \beta_2 = 0$, $F_{20} = \frac{\text{SS}_{C_i}}{\text{MSE}} \sim F_{1,N-5}$

  - Similar for β_3 and β_4

- Question: what is the estimate for β_0?

General formulas for orthogonal polynomial of degrees 1-4

One factor of a levels $l_1, l_2, \ldots, l_a$, equally spaced. Let m be the median, δ be the difference between two consecutive levels:

$$P_1(t) = \lambda_1(\frac{t-m}{\delta})$$

$$P_2(t) = \lambda_2[(\frac{t-m}{\delta})^2 - \frac{a^2-1}{12}]$$

$$P_3(t) = \lambda_3[(\frac{t-m}{\delta})^3 - (\frac{t-m}{\delta})(\frac{3a^2-7}{20})]$$

$$P_4(t) = \lambda_4[(\frac{t-m}{\delta})^4 - (\frac{t-m}{\delta})^2(\frac{3a^2-13}{14}) + \frac{3(a^2-1)(a^2-9)}{560}]$$

(λ_i) are constants to make the polynomials have integer values at the treatment levels, they are available from Table IX or Table X.

Tensile Strength Example: m=25, $\delta = 5$, (λ_i)=(1, 1, 5/6, 35/12)

SAS

tensile strength example

```
data one;
 infile ’c:\saswork\data\tensile.dat’;
 input percent strength time;

proc glm data=one;
 class percent;
 model strength=percent;
estimate ’C1’ percent -2 -1 0 1 2;
 estimate ’C2’ percent 2 -1 -2 -1 2;
 estimate ’C3’ percent -1 2 0 -2 1;
 estimate ’C4’ percent 1 -4 6 -4 1;
contrast ’C1’ percent -2 -1 0 1 2;
 contrast ’C2’ percent 2 -1 -2 -1 2;
 contrast ’C3’ percent -1 2 0 -2 1;
 contrast ’C4’ percent 1 -4 6 -4 1;
run;
```

Output

Dependent Variable: strength

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	475.7600000	118.9400000	14.76	<.0001
Error	20	161.2000000	8.0600000		
Corrected Total	24	636.9600000			

Parameter	Estimate	Error	t Value	Pr > |t|
C1	8.2000000	4.0149720	2.04	0.0545
C2	-31.0000000	4.7505789	-6.53	<.0001
C3	-11.4000000	4.0149720	-2.84	0.0101
C4	-21.8000000	10.6226174	-2.05	0.0535

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
C1	1	33.6200000	33.6200000	4.17	0.0545
C2	1	343.2142857	343.2142857	42.58	<.0001
C3	1	64.9800000	64.9800000	8.06	0.0101
C4	1	33.9457143	33.9457143	4.21	0.0535

Estimates

Hence,

$$\hat{\beta}_1 = 8.20/10 = .82; \ \hat{\beta}_2 = -31/14 = -2.214$$

$$\hat{\beta}_3 = -11.4/10 = -1.14; \ \hat{\beta}_4 = -21.8/70 = -0.311$$

So the fitted functional relationship between tensile strength y and cotton percent (t) is

$$y = \hat{\beta}_0 + .82P_1(t) - 2.214P_2(t) - 1.14P_3(t) - 0.311P_4(t),$$

where $P_1(t), \ldots, P_4(t)$ are defined on Slide 16.

Testing Multiple Contrasts (Multiple Comparisons) Using Confidence Intervals

- One contrast:

$$H_0 : \Gamma = \sum c_i \mu_i = \Gamma_0 \text{ vs } H_1 : \Gamma \neq \Gamma_0 \text{ at } \alpha$$

100(1-α) Confidence Interval (CI) for Γ:

$$\mathrm{CI} : \sum c_i \bar{y}_{i.} \pm t_{\alpha/2, N-a} \sqrt{MS_E \sum \frac{c_i^2}{n_i}}$$

$$P(\text{CI not contain } \Gamma_0 | H_0) = \alpha (= \text{type I error})$$

- Decision Rule: Reject H_0 if CI does not contain Γ_0.

- Multiple contrasts

$$H_0 : \Gamma^1 = \Gamma_0^1, \ldots \Gamma^m = \Gamma_0^m \text{ vs } H_1 : \text{at least one does not hold}$$

If we construct CI_1, CI_2,..., CI_m, each with 100(1-α) level, then for each CI_i,

$$P(\text{CI}_i \text{ not contain} \Gamma_0^i \mid H_0) = \alpha, \text{ for } i = 1, \ldots, m$$

- But the **overall error rate** (probability of type I error for H_0 vs H_1) is inflated and much larger than α, that is,

$$P(\text{at least one CI}_i \text{ not contain } \Gamma_0^i \mid H_0) >> \alpha$$

- One way to achieve small overall error rate, we require much smaller error rate (α') of each individual CI_i.

Bonferroni Method for Testing Multiple Contrasts

- Bonferroni Inequality

$$P(\text{ at least one CI}_i \text{ not contain } \Gamma_0^i \mid H_0)$$

$$= P(\text{CI}_1 \text{ not contain..or ....or CI}_m \text{ not contain } \mid H_0)$$

$$\leq P(\text{ CI}_1 \text{ not } \mid H_0) + \cdots + P(\text{ CI}_m \text{ not } \mid H_0) = m\alpha'$$

- In order to control overall error rate (or, overall confidence level), let

$$m\alpha' = \alpha, \text{we have, } \alpha' = \alpha/m$$

- Bonferroni CIs:

$$\text{CI}_i : \sum c_{ij}\bar{y}_{j.} \pm t_{\alpha/2m}(N-a)\sqrt{\text{MS}_E \sum \frac{c_{ij}^2}{n_j}}$$

- When m is large, Bonferroni CIs are too conservative (overall type II error too large).

Scheffe's Method for Testing All Contrasts

- Consider all possible contrasts: $\Gamma = \sum c_i \mu_i$

  Estimate: $C = \sum c_i \bar{y}_{i.}$, St. Error: $\mathrm{S.E.}_C = \sqrt{\mathrm{MS}_E \sum \frac{c_i^2}{n_i}}$
- Critical value: $\sqrt{(a-1)F_{\alpha,a-1,N-a}}$
- Scheffe's simultaneous CI: $C \pm \sqrt{(a-1)F_{\alpha,a-1,N-a}}\ \mathrm{S.E.}_C$
- Overall confidence level and error rate for m contrasts

$$P(\text{CIs contain true parameter for any contrast}) \geq 1-\alpha$$

$$P(\text{at least one CI does not contain true parameter}) \leq \alpha$$

Remark: Scheffe's method is also conservative, too conservative when m is small

Methods for Pairwise Comparisons

- There are $a(a-1)/2$ possible pairs: $\mu_i - \mu_j$ (contrast for comparing μ_i and μ_j). We may be interested in m pairs or all pairs.
- Standard Procedure:
  1. Estimation: $\bar{y}_{i.} - \bar{y}_{j.}$
  2. Compute a **Critical Difference (CD)** (based on the method employed)
  3. If

  $$|\ \bar{y}_{i.} - \bar{y}_{j.}\ | > \mathrm{CD}$$

  or equivalently if the interval

  $$(\bar{y}_{i.} - \bar{y}_{j.} - \mathrm{CD},\ \bar{y}_{i.} - \bar{y}_{j.} + \mathrm{CD})$$

  does not contain zero, declare $\mu_i - \mu_j$ significant.

Methods for Calculating CD.

- Least significant difference (LSD):

$$\text{CD} = t_{\alpha/2,N-a}\sqrt{\text{MS}_E(1/n_i + 1/n_j)}$$

  not control overall error rate

- Bonferroni method (for m pairs)

$$\text{CD} = t_{\alpha/2m,N-a}\sqrt{\text{MS}_E(1/n_i + 1/n_j)}$$

  control overall error rate for the m comparisons.

- Tukey's method (for all possible pairs)

$$\text{CD} = \frac{q_\alpha(a, N-a)}{\sqrt{2}}\sqrt{\text{MS}_E(1/n_i + 1/n_j)}$$

  $q_\alpha(a, N-a)$ from studentized range distribution (Table VII or Table VIII).
  Control overall error rate (exact for balanced experiments). (Example 3.7).

Comparing treatments with control (Dunnett's method)

1. Assume μ_1 is a control, and $\mu_2, \ldots, \mu_a$ are (new) treatments
2. Only interested in $a-1$ pairs: $\mu_2 - \mu_1, \ldots, \mu_a - \mu_1$
3. Compare $\mid \bar{y}_{i.} - \bar{y}_{1.} \mid$ to

$$\text{CD} = d_\alpha(a-1, N-a)\sqrt{\text{MS}_E(1/n_i + 1/n_1)}$$

where $d_\alpha(p, f)$ from Table IX or Table VIII: critical values for Dunnett's test.

4. Remark: control overall error rate. Read Example 3-9 (or 3-10)

For pairwise comparison, which method should be preferred? LSD, Bonferroni, Tukey, Dunnett or others?

SAS Code

```
data one;
 infile 'c:\saswork\data\tensile.dat';
 input percent strength time;

proc glm data=one;
 class percent;
 model strength=precent;

/* Construct CI for Treatment Means*/
means percent /alpha=.05 lsd clm;
means percent / alpha=.05 bon clm;

/* Pairwise Comparison*/
means percent /alpha=.05 lines lsd;
means percent /alpha=.05 lines bon;
means percent /alpha=.05 lines scheffe;
means percent /alpha=.05 lines tukey;
means percent /dunnett;
run;
```

The GLM Procedure

t Confidence Intervals for y

Alpha	0.05
Error Degrees of Freedom	20
Error Mean Square	8.06
Critical Value of t	2.08596
Half Width of Confidence Interval	2.648434

trt	N	Mean	95% Confidence Limits	
30	5	21.600	18.952	24.248
25	5	17.600	14.952	20.248
20	5	15.400	12.752	18.048
35	5	10.800	8.152	13.448
15	5	9.800	7.152	12.448

The GLM Procedure

Bonferroni t Confidence Intervals for y

Alpha	0.05
Error Degrees of Freedom	20
Error Mean Square	8.06
Critical Value of t	2.84534
Half Width of Confidence Interval	3.612573

trt	N	Mean	Simultaneous 95% Confidence Limits	
30	5	21.600	17.987	25.213
25	5	17.600	13.987	21.213
20	5	15.400	11.787	19.013
35	5	10.800	7.187	14.413
15	5	9.800	6.187	13.413

```
                     t Tests (LSD) for y
NOTE: This test controls the Type I comparisonwise error rate, not the
                     experimentwise error rate.
                 Alpha                                0.05
                 Error Degrees of Freedom               20
                 Error Mean Square                    8.06
                 Critical Value of t               2.08596
                 Least Significant Difference       3.7455
         Means with the same letter are not significantly different.

              t Grouping          Mean        N     trt
                       A        21.600        5     30

                       B        17.600        5     25
                       B        15.400        5     20

                       C        10.800        5     35
                       C         9.800        5     15
```

```
                    Bonferroni (Dunn) t Tests for y
  This test controls the Type I experimentwise error rate, but it generall
              has a higher Type II error rate than REGWQ.
                Alpha                                     0.05
                Error Degrees of Freedom                    20
                Error Mean Square                         8.06
                Critical Value of t                    3.15340
                Minimum Significant Difference          5.6621
          Means with the same letter are not significantly different.
             Bon Grouping          Mean       N    trt
                        A        21.600       5    30
                   B    A        17.600       5    25
                   B    C        15.400       5    20
                        C        10.800       5    35
                        C         9.800       5    15
```

Scheffe's Test for y

NOTE: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	20
Error Mean Square	8.06
Critical Value of F	2.86608
Minimum Significant Difference	6.0796

Means with the same letter are not significantly different.

Scheffe Grouping		Mean	N	trt
	A	21.600	5	30
	A			
B	A	17.600	5	25
B				
B	C	15.400	5	20
	C			
	C	10.800	5	35
	C			
	C	9.800	5	15

```
                 Tukey's Studentized Range (HSD) Test for y
      This test controls the Type I experimentwise error rate, but it general
                 has a higher Type II error rate than REGWQ.
                 Alpha                                        0.05
                 Error Degrees of Freedom                       20
                 Error Mean Square                            8.06
                 Critical Value of Studentized Range       4.23186
                 Minimum Significant Difference              5.373
           Means with the same letter are not significantly different.
               Tukey Grouping           Mean       N    trt
                        A             21.600       5    30
                        A
                   B    A             17.600       5    25
                   B
                   B    C             15.400       5    20
                        C
                   D    C             10.800       5    35
                   D
                   D                   9.800       5    15
```

```
                    Dunnett's t Tests for y
This test controls the Type I experimentwise error for comparisons of a
                  treatments against a control.
              Alpha                                   0.05
              Error Degrees of Freedom                  20
              Error Mean Square                       8.06
              Critical Value of Dunnett's t        2.65112
              Minimum Significant Difference        4.7602
    Comparisons significant at the 0.05 level are indicated by ***.

                            Difference
           trt                 Between     Simultaneous 95%
        Comparison               Means    Confidence Limits

        30    - 15              11.800       7.040   16.560  ***
        25    - 15               7.800       3.040   12.560  ***
        20    - 15               5.600       0.840   10.360  ***
        35    - 15               1.000      -3.760    5.760
```

Determining Sample Size

- More replicates required to detect small treatment effects
- Operating Characteristic Curves for F tests
- Probability of type II error

$$\beta = P(\text{ accept } H_0 \mid H_0 \text{ is false})$$

$$= P(F_0 < F_{\alpha,a-1,N-a} \mid H_1 \text{ is correct })$$

- Under H_1, F_0 follows a **noncentral** F distribution with noncentrality λ and degrees of freedom, $a-1$ and $N-a$, where

$$\lambda = a \times \Phi^2, \quad \Phi^2 = \frac{n\sum_{i=1}^{a}\tau_i^2}{a\sigma^2}$$

- OC curves of β vs n and Φ are included in Chart V for various α and a.

Example 3-10: Experiment Involving 4 Treatments

- Suppose want to detect (at $\alpha = 0.01$) $\mu_1 = 575$, $\mu_2 = 600$, $\mu_3 = 650$, $\mu_4 = 675$, and can assume $\sigma = 25$.
- How many replicates per treatment is needed such that $\beta < 0.10$?
- We have $\tau_1 = -50$, $\tau_2 = -25$, $\tau_3 = 25$, $\tau_4 = 50$, and

$$\Phi^2 = \frac{n \sum_{i=1}^{a} \tau_i^2}{a\sigma^2} = \frac{6250n}{4(25)^2} = 2.5n,$$

- $\nu_1 = a - 1 = 3$, $\nu_2 = N - a = 4(n-1)$, and $\beta = f(\alpha, \nu_1, \nu_2, n, a * \Phi^2)$:

n	Φ^2	Φ	ν_2	β	Power
3	7.5	2.74	8	0.25	0.75
4	10.0	3.16	12	0.04	0.04
5	12.5	3.54	16	$<$0.01	$>$0.99

SAS Code

Determining Sample Size

```
data new;
a=4; alpha=.01;
do n=3 to 5;
df = a*(n-1);
Phi2 = 2.5 * n;
fcut = finv(1-alpha,a-1,df);
beta = probf(fcut,a-1,df,a*Phi2);
power = 1-beta;
output;
end;
proc print;
var n Phi2 df beta power; run;
```

Obs	n	Phi2	df	beta	power
1	3	7.5	8	0.25	0.75
2	4	10.0	12	0.04	0.96
3	5	12.5	16	0.003	0.996

Another Approach

- Suppose want to guarantee $\beta < 0.10$ when there is at least one pair of treatments that differ by D(e.g. $\mu_1 - \mu_2 \geq D$).
- The smallest Φ^2 is

$$\Phi^2 = \frac{nD^2}{2a\sigma^2}$$

- In Example 3.10, consider $D = 75$ and assume $\sigma^2 = 25$,

$$\Phi^2 = \frac{n(75)^2}{2(4)(25^2)} = 1.125n$$

- $\nu_1 = a - 1 = 3$, $\nu_2 = N - a = 4(n-1)$, and $\beta = f(\alpha, \nu_1, \nu_2, n, a\Phi^2)$:

n	Φ^2	Φ	ν_2	β	Power
4	4.5	2.12	12	0.39	0.61
5	5.625	2.37	16	0.20	0.80
6	6.75	2.60	20	$<$0.10	$>$0.90