Topic 6: Post ANOVA comparisons of means

Montgomery: chapter 3

Prof. Lingling An
University of Arizona

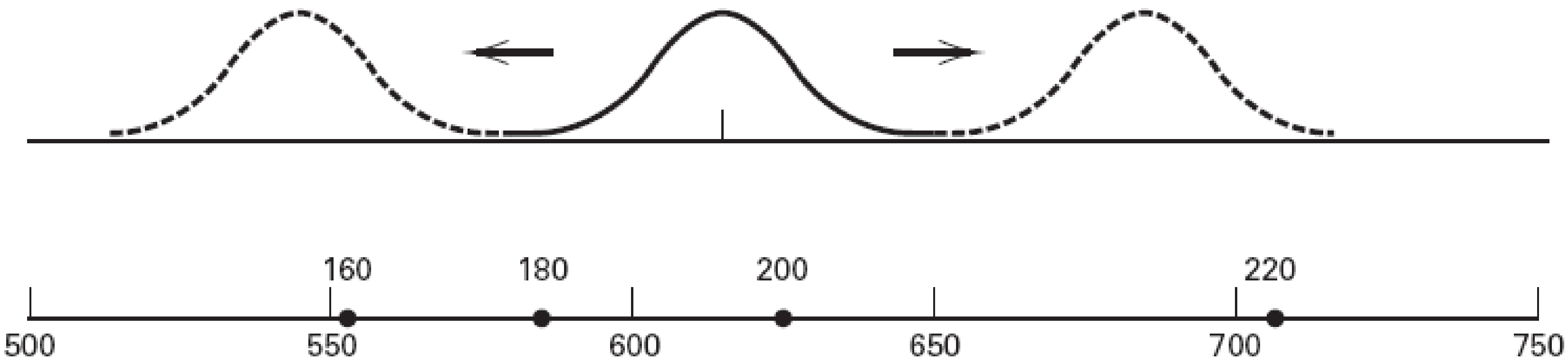
- Comparing treatment means
  - Linear combination of the treatments
  - Contrasts
  - Orthogonal contrasts
  - Simultaneous confidence intervals
- Sample size determination

Post- ANOVA Comparison of Means

- The analysis of variance tests the hypothesis of **equal treatment means**
- Assume that residual analysis is satisfactory
- If that hypothesis is rejected, we don't know **which specific means** are different
  - Determining which specific means differ following an ANOVA is called the **multiple comparisons problem**
- **How about to test:**

  H_0: $2\mu_1 + \mu_2 = \mu_3$?

Graphical comparison of means



■ FIGURE 3.11 Etch rate averages from Example 3.1 in relation to a t distribution with scale factor $\sqrt{MS_E/n} = \sqrt{330.70/5} = 8.13$

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 $i = 2, \ldots, a$ 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

- 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 $H_0 : L = 60$.

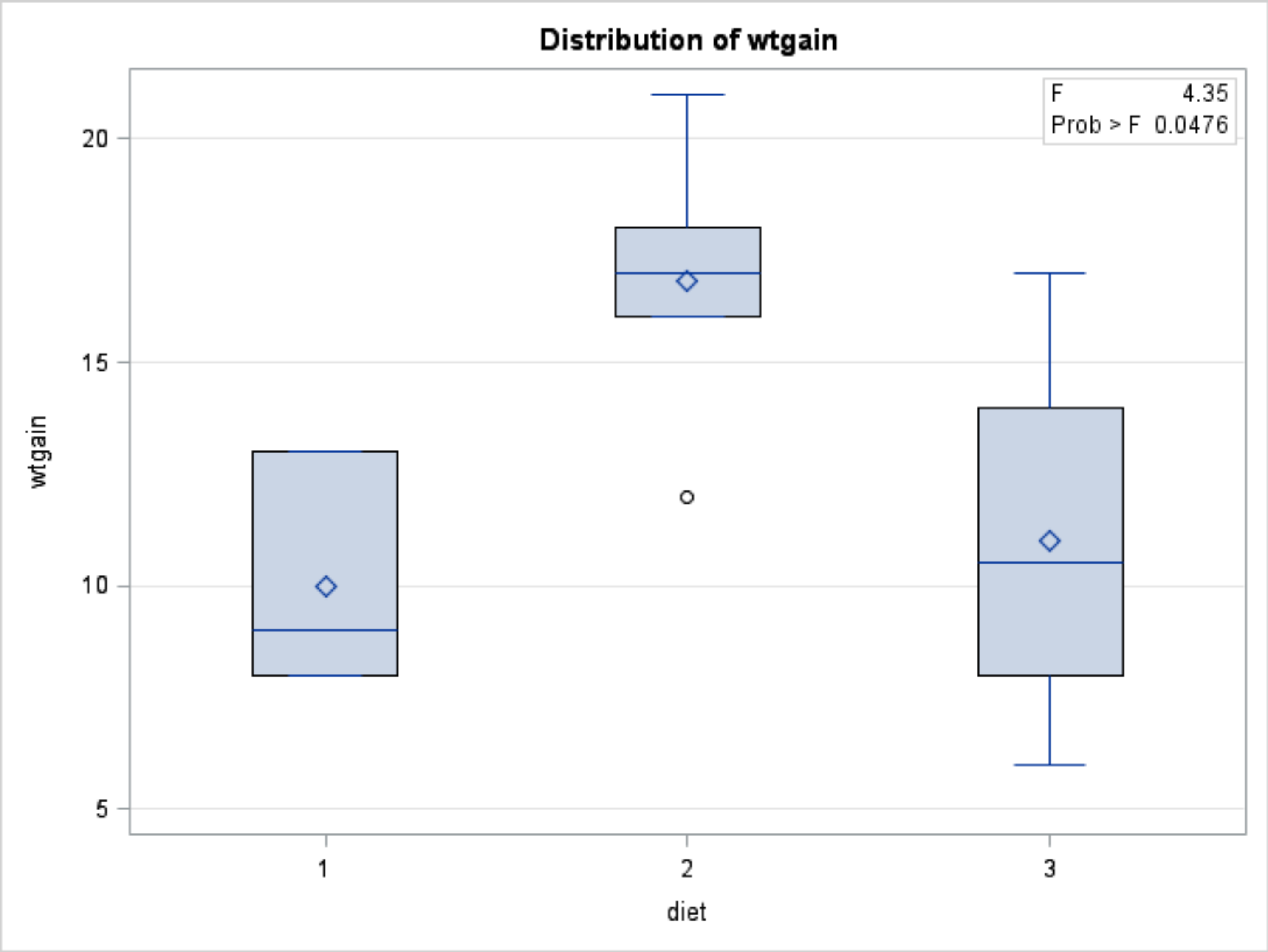
```
data lambs;
input diet wtgain@@;
datalines;
1 8 1 13 1 9 2 12 2 16 2 21 2 17 2 18 3 11 3 10 3 17 3 6
;
run;

proc glm;
class diet;
model wtgain=diet;
means diet;
estimate 'L1' intercept 6 diet 1 2 3;
run;
```

Dependent Variable: wtgain

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	114.8666667	57.4333333	4.35	0.0476
Error	9	118.8000000	13.2000000		
Corrected Total	11	233.6666667			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
diet	2	114.8666667	57.4333333	4.35	0.0476


Distribution of wtgain
F 4.35
Prob > F 0.0476
20
15
10
5
wtgain
1
2
3
diet

Level of diet	N	wtgain	
		Mean	Std Dev
1	3	10.0000000	2.64575131
2	5	16.8000000	3.27108545
3	4	11.0000000	4.54606057

Dependent Variable: wtgain

Parameter	Estimate	Standard Error	t Value	Pr > |t|
L1	76.6000000	6.68281378	11.46	<.0001

$$H_0 : L = 60$$

$$t_0 = (76.6 - 60) / 6.6828 = 2.484$$

- P-value=$P(t \le -2.484$ or $t \ge 2.484 | \; t_{(12-3)})$=0.0348
- Reject H_0 at α=0.05

Online p-value calculator:
http://graphpad.com/quickcalcs/PValue1.cfm

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: $H_0 : \Gamma = 0$

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

or

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

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

Where: Contrast Sum of Squares

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

Example: tensile data

```
data one;
 infile "/folders/myshortcuts/MySAS_folder/
tensile.dat";
 input percent strength time;
 run;
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;
run;
```

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			

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
C1	1	291.6000000	291.6000000	36.18	<.0001
C2	1	31.2500000	31.2500000	3.88	0.0630
C3	1	152.1000000	152.1000000	18.87	0.0003
C4	1	0.8100000	0.8100000	0.10	0.7545

Orthogonal contrasts

- A useful special case of the contrasts is 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.

- Generally, the method of contrasts (or orthogonal contrasts) is useful for ***preplanned comparisons***, which are specified prior to running the experiment and examining data.
  - If comparisons are selected after examining the data, most experimenters would construct tests that correspond to large observed differences in means
  - But these large differences could be the result of the real effect, or be the result of random error.
  - Read Example 3.6 on p90-91

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:

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

$$P(\text{CI not contain } L_0 | H_0) = \alpha (= \text{type I error})$$

- 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

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

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

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

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

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

- Bonferroni CIs:

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

- When m is large, Bonferroni CIs are too conservative

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: $\text{S.E.}_C = \sqrt{\text{MS}_E \sum \frac{c_i^2}{n_i}}$

- Critical value: $\sqrt{(a-1)F_{\alpha,a-1,N-a}}$

- Scheffe's simultaneous CI: $C \pm \sqrt{(a-1)F_{\alpha,a-1,N-a}}\ \text{S.E.}_C$

- 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

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

  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.

- Least significant difference (LSD):

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

not control overall error rate

- Bonferroni method (for m pairs)

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

control overall error rate for the m comparisons.

- Tukey's method (for all possible pairs)

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

$q_\alpha(a, N-a)$ from studentized range distribution (Table VII).

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

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

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

4. Remark: control overall error rate. Read Example 3-9

For the tensile data:

```
 proc glm data=one;
class percent;
model strength=percent;
/* Construct CI for Treatment Means*/
means percent /alpha=.05 clm lsd;
means percent /alpha=.05 clm bon;
means percent /alpha=.05 clm scheffe;
/* Pairwise Comparison*/
means percent /alpha=.05 lines lsd;
means percent /alpha=.05 lines tukey;
means percent /alpha=.05 dunnett ('15');
run;
```

Individual C.I.

t Confidence Intervals for strength

percent	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

Simultaneous C. I.s (Bonferrroni)

Bonferroni t Confidence Intervals for strength

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

percent	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

Simultaneous C. I.s (Scheffe)

Scheffe's Confidence Intervals for strength

Alpha	0.05
Error Degrees of Freedom	20
Error Mean Square	8.06
Critical Value of F	2.71089
Half Width of Confidence Interval	4.674374

percent	N	Mean	Simultaneous 95% Confidence Limits	
30	5	21.600	16.926	26.274
25	5	17.600	12.926	22.274
20	5	15.400	10.726	20.074
35	5	10.800	6.126	15.474
15	5	9.800	5.126	14.474

Pairwise comparison (LSD)

t Tests (LSD) for strength

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

Pairwise comparison (LSD) - 2

Means with the same letter are not significantly different.			
t Grouping	Mean	N	percent
A	21.600	5	30
B	17.600	5	25
B			
B	15.400	5	20
C	10.800	5	35
C			
C	9.800	5	15

Pairwise comparison (Tukey)

Tukey's Studentized Range (HSD) Test for strength

Note:	**This test controls the Type I experimentwise error rate, but it generally 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

Pairwise comparison (Tukey) -2

Means with the same letter are not significantly different.				
Tukey Grouping		Mean	N	percent
	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

Compared to control (Dunnett)

Dunnett's t Tests for strength

Note:	**This test controls the Type I experimentwise error for comparisons of all treatments against a control.**

Alpha	0.05
Error Degrees of Freedom	20
Error Mean Square	8.06
Critical Value of Dunnett's t	2.65103
Minimum Significant Difference	4.7601

Compared to control (Dunnett) - 2

Comparisons significant at the 0.05 level are indicated by ***.				
percent Comparison	Difference Between Means	Simultaneous 95% 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	

Which method should I use?

- Multiple comparisons (i.e., contrasts) but not pairwise comparisons
  - If m is very small, use Bonferroni method
  - If m is very large, use Scheffe method
- Pairwise comparison
  - Tukey method
- Comparing treatment means with a control
  - Dunnett method

Determining Sample Size (OC curve)

- 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 $N-a$. Let

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

- OC curves of β vs n and Φ are included in Chart V for various α and a.
- Read Example 3.10

Example 3.10: etching rate

What we know:

four treatment means: 575 , 600 , 650 , 675

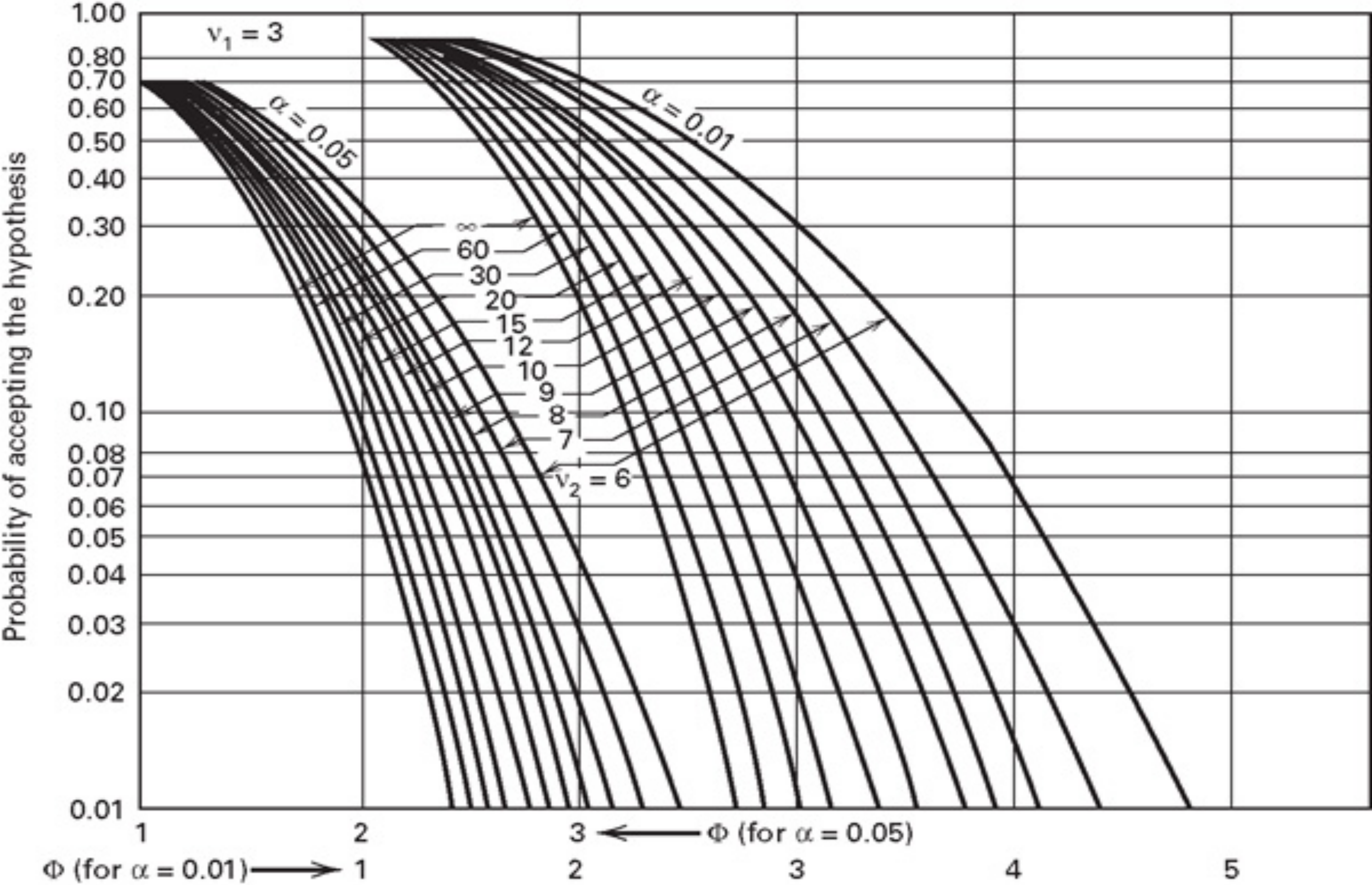
Standard deviation at each level: 25

Alpha=0.01

Power=0.9

n =?

n	Φ^2	Φ	$a(n-1)$	β	Power $(1-\beta)$
3	7.5	2.74	8	0.25	0.75
4	10.0	3.16	12	0.04	0.96
5	12.5	3.54	16	<0.01	>0.99



Probability of accepting the hypothesis
$\nu_1 = 3$
$\alpha = 0.05$
$\alpha = 0.01$
∞
60
30
20
15
12
10
9
8
7
$\nu_2 = 6$
1.00
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
1
2
3
Φ (for $\alpha = 0.05$)
Φ (for $\alpha = 0.01$)
1
2
3
4
5

SAS code: sample size calculation for one way ANOVA

```
proc power ;
  onewayanova
  groupmeans = 575 | 600 | 650 | 675
  stddev = 25
  alpha = 0.01
  npergroup = .
  power = .9;
run;
```

The POWER Procedure
Overall F Test for One-Way ANOVA

Fixed Scenario Elements	
Method	Exact
Alpha	0.01
Group Means	575 600 650 675
Standard Deviation	25
Nominal Power	0.9

Computed N Per Group	
Actual Power	N Per Group
0.962	4

Determining Sample Size (Conf. Interval. approach)

- Assume experimenter wishes to express the final results in terms of C. I. and is willing to specify in advance how wide he/she wants these intervals to be.
- So Margin of error (=half width of C.I) is assumed and solve for n
  - e.g, accuracy of the confidence interval for the difference of two treatment means:

$$\pm t_{\alpha/2, N-a}\sqrt{2\frac{MSE}{n}}$$

  - Or use simultaneous confidence interval

Last slide

- Read Sections: finish Ch3