Experimental Design
Note 3-2
Post ANOVA comparisons of means

Keunbaik Lee

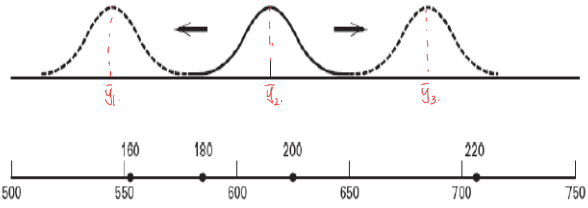
Sungkyunkwan University

Post-ANOVA Comparison of Means I

- The analysis of variance tests the hypothesis of equal treatment means $H_0: \tau_1 = \cdots = \tau_a = 0 \iff H_0: \mu_1 = \mu_2 = \cdots = \mu_a$
- Assume that residual analysis is satisfactory
- If that hypothesis is rejected, we dont 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 I



■ **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 I

- ANOVA Model:

$$\mu_i - \mu$$

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

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

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

- Want to test: $H_0 : L = \sum_i c_i\mu_i = L_0$
- Examples:
  - Pairwise comparison: $\mu_i - \mu_j = 0$ for all possible i and j.

Linear combinations of treatment means II

$i = 2, 3, \cdots, a$ control

- Compare treatment vs control: $\mu_i - \mu_1 = 0$ when treatment 1 is a control and $i = 2, \cdots, a$ are new treatments.
- General cases such as $\mu_1 - 2\mu_2 + \mu_3 = 0, \; \mu_1 + 3\mu_2 - 6\mu_3 = 0$ etc.

$\mu_2 = \frac{\mu_1 + \mu_3}{2}$

- Estimate of L:

$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$

$\bar{y}_{i\cdot} = \mu + \tau_i + \bar{\varepsilon}_{i\cdot}$

$Var(\bar{y}_{i\cdot}) = Var(\bar{\varepsilon}_{i\cdot}) = \frac{\sigma^2}{n}$

$$\hat{L} = \sum_i c_i \hat{\mu}_i = \sum_i c_i \bar{y}_{i\cdot}$$

$$var(\hat{L}) = \sum_i c_i^2 var(\bar{y}_{i\cdot}) = \sigma^2 \sum_i \frac{c_i^2}{n_i}$$

- Standard Error of $\hat{L}$

$MSE = \sigma^2$

$$SE_{\hat{L}} = \sqrt{MSE \sum_i \frac{c_i^2}{n_i}}.$$

Linear combinations of treatment means III

- Test statistic $H_0: L = L_0$

$$\frac{\hat{L}-L_0}{\sqrt{\sigma^2 \sum_{i=1}^{a} \frac{c_i^2}{n_i}}} \sim N(0,1)$$

$$\frac{SSE}{\sigma^2} \sim \chi^2_{N-a}$$

independent

$$t_0 = \frac{(\hat{L} - L_0)}{SE_{\hat{L}}} \sim t_{(N-a)} \text{ under } H_0$$

$$T = \frac{\dfrac{\hat{L}-L_0}{\sqrt{\sigma^2 \sum_{i=1}^{a} \frac{c_i^2}{n_i}}}}{\sqrt{\left(\frac{SSE}{\sigma^2}\right)/(N-a)}} = \frac{\dfrac{\hat{L}-L_0}{\sqrt{\sum_{i=1}^{a} \frac{c_i^2}{n_i}}}}{\sqrt{\underbrace{\frac{SSE}{N-a}}_{MSE}}} = \frac{\hat{L}-L_0}{\underbrace{\sqrt{MSE \sum_{i=1}^{a} \frac{c_i^2}{n_i}}}_{SE_{\hat{L}}}} \sim t_{N-a}$$

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

$$L = \overset{\mu+\tau_1}{\mu_1} + \overset{2(\mu+\tau_2)}{2\mu_2} + \overset{3(\mu+\tau_3)}{3\mu_3} = \underset{\text{intercept}}{6\mu} + \underbrace{\tau_1 + 2\tau_2 + 3\tau_3}_{\text{groups}}$$

and test: $H_0 : L = 60$.
See lambs-diet.SAS.

$$\frac{\hat{L} - 60}{\sqrt{MSE \sum_{i=1}^{a} \frac{c_i^2}{n_i}}} = \frac{76.6 - 60}{6.683} = 2.484$$

Contrasts I

- $\Gamma = \sum_{i=1}^{a} c_i \mu_i$ is a contrast if $\sum_{i=1}^{a} c_i = 0$. Equivalently, $\Gamma = \sum_{i=1}^{a} c_i \tau_i$.
- Examples
  - $$\Gamma_1 = \mu_1 - \mu_2 = \mu_1 - \mu_2 + 0\mu_3 + 0\mu_4,$$
    $$c_1 = 1,\ c_2 = -1,\ c_3 = 0,\ c_4 = 0$$
    Comparing μ_1 and μ_2.
  - $$\Gamma_2 = \mu_1 - 0.5\mu_2 - 0.5\mu_3 = \mu_1 - 0.5\mu_2 - 0.5\mu_3 + 0\mu_4,$$
    $$d_1 = 1,\ d_2 = -0.5,\ d_3 = -0.5,\ d_4 = 0$$
    Comparing μ_1 and the average of μ_2 and μ_3.

Contrasts II

- Estimate of Γ:

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

- Test: $H_0 : \Gamma = 0$
  use $t_0 = \frac{C}{SE_C} \sim t_{(N-a)}$
  or $t_0^2 = \frac{(\sum_i c_i \bar{y}_{i\cdot})^2}{MSE \sum_i \frac{c_i^2}{n_i}} = \frac{(\sum_i c_i \bar{y}_{i\cdot})^2 / \sum_i \frac{c_i^2}{n_i}}{MSE} = \frac{SS_C/1}{MSE}$ where
  $SS_C = (\sum_i c_i \bar{y}_{i\cdot})^2 / \sum_i \frac{c_i^2}{n_i}$.
  Under H_0, $t_0^2 \sim F_{1,N-a}$.

See Tensile1.SAS.

Orthogonal contrasts I

- A useful special case of the contrasts is orthogonal contrasts.
- Two contrasts $\{c_i\}$ and $\{d_i\}$ are **orthogonal** if

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

2개의 contrast

1) $\sum_{i=1}^{a} c_i \mu_i$

2) $\sum_{i=1}^{a} d_i \mu_i$

$\rightarrow$ Let $c = [c_1 \ c_2 \ \cdots \ c_a]'$, $d = [d_1 \ d_2 \ \cdots \ d_a]'$

$c'd = \sum_{i=1}^{a} c_i d_i$

Orthogonal contrasts II

- Example

$$\Gamma_1 = \mu_1 + \mu_2 - \mu_3 - \mu_4, \text{ so } c_1 = 1,\ c_2 = 1,\ c_3 = -1,\ c_4 = -1.$$
$$\Gamma_2 = \mu_1 - \mu_2 + \mu_3 - \mu_4, \text{ so } d_1 = 1,\ d_2 = -1,\ d_3 = 1,\ d_4 = -1.$$

It is easy to verify that both Γ_1 and Γ_2 are contrasts. Furthermore,

$$\begin{aligned} & c_1 d_1 + c_2 d_2 + c_3 d_3 + c_4 d_4 \\ = \ & 1 \times 1 + 1 \times (-1) + (-1) \times 1 + (-1) \times (-1) = 0. \end{aligned}$$

Here, Γ_1 and Γ_2 are orthogonal to each other.

Orthogonal contrasts III

- 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.
- Orthogonal contrasts can be used to further partition the model sum of squares.
  - There are many sets of orthogonal contrasts and thus, many ways to partition the sum of squares.
  - The selection of particular set of orthogonal contrasts is based on

If orthogonal, $\underbrace{SS_{treatment}}_{df=a-1} = SS_{T_1} + SS_{T_2} + \cdots \underbrace{SS_{T_{a-1}}}_{df=1}$

Orthogonal contrasts IV

- Research Objective: some comparisons are more important than others.
- Experimental design.

- A ~~specail~~ special set of orthogonal contrasts that are used when the levels of a factor can be assigned values on a metric scale are called orthognal polynomials.
- Thus for $t =$ the number of treatments, the following table can be used to obtain the contrast coefficients:

Orthogonal contrasts V

Table: Orthogonal polynomial contrasts

$t=3$			$t=4$			$t=5$		
L	Q	C	L	Q	C	L	Q	C
-1	1		-3	1	-1	-2	2	-1
0	-2		-1	-1	3	-1	-1	2
1	1		1	-1	-3	0	-2	0
			3	1	1	1	-1	-2
						2	2	1

$t=6$			$t=7$			$t=8$		
L	Q	C	L	Q	C	L	Q	C
-5	5	-5	-3	5	-1	-7	7	-7
-3	-1	7	-2	0	1	-5	1	5
-1	-4	4	-1	-3	1	-3	-3	7
1	-4	-4	0	-4	0	-1	-5	3
3	-1	-7	1	-3	-1	1	-5	-3
5	5	5	2	0	-1	3	-3	-7
			3	5	1	5	1	-5
						7	7	7

Orthogonal contrasts VI

- In $t = 3$, linear and quadratic contrasts for assessing trends in mean response across factor:

$$\Gamma_{Linear} = (-1)\mu_1 + (0)\mu_2 + (1)\mu_3,$$
$$\Gamma_{quadratic} = (1)\mu_1 + (-2)\mu_2 + (1)\mu_3.$$

Testing multiple contrasts (multiple comparisons) using Confidence Intervals I

- One contrast:

$$H_0 : \Gamma = \sum_{i=1}^{a} c_i \mu_i = \Gamma_0 \quad \text{vs } H_1 : \Gamma \neq \Gamma_0$$

$$\frac{\hat{\Gamma} - \Gamma_0}{\sqrt{MSE \sum_{i=1}^{a} \frac{c_i^2}{n_i}}}$$

$100(1-\alpha)\%$ confidence interval (CI) for Γ:

$$CI : \sum_{i=1}^{a} c_i \bar{y}_{i\cdot} \pm t_{\alpha/2, N-a} \sqrt{MSE \sum_{i=1}^{a} c_i^2 / n_i},$$

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

Testing multiple contrasts (multiple comparisons) using Confidence Intervals II

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

Previously ; contrast $H_0 : \Gamma_1 = 0$, $H_a : \Gamma_1 \neq 0$

⋮

$H_0 : \Gamma_a = 0$, $H_a : \Gamma_a \neq 0$

$H_0 : \Gamma^1 = \Gamma_0^1, \cdots, \Gamma^m = \Gamma_0^m$ vs $H_1 :$ at least one does not hold

If we construct $CI_1, \ CI_2, \ \cdots, \ CI_m$, each with $100(1-\alpha)\%$ level, then for each CI_i,

$$P(CI_i \text{ not contain } \Gamma_0^i | H_0) = \alpha, \text{ for } i = 1, \cdots, m.$$

Testing multiple contrasts (multiple comparisons) using Confidence Intervals III

- 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,

$$P(\text{at least one } CI_i \text{ not contain } \Gamma_0^i | H_0) >> \alpha.$$

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

Let $A_i = \{CI : \text{contain } \Gamma_0^i\}$

$$\Rightarrow P(A_1^c \cup A_2^c \cup \cdots \cup A_m^c | H_0) = P[(A_1 \cap A_2 \cap \cdots \cap A_m)^c | H_0] = 1 - P[(A_1 \cap A_2 \cap \cdots \cap A_m) | H_0]$$

$$= 1 - \underbrace{P(A_1|H_0)}_{1-\alpha} P(A_2|H_0) \cdots \underbrace{P(A_m|H_0)}_{1-\alpha}$$

$$= 1 - (1-\alpha)^m$$

rate

rate

rate

Bonferroni Method for Testing Multiple Contrasts

- Bonferroni Inequality

$$
\begin{aligned}
& P(\text{at least one } CI_i \text{ not contain} \Gamma_0^i | H_0) \\
= \; & P(CI_1 \text{not contain or} \cdots \text{ or } CI_m \text{ not contain} | H_0) \\
\neq \; & P(CI_1 \text{ not} | H_0) + \cdots + P(CI_m \text{ not} | H_0) = m\alpha^{'}
\end{aligned}
$$

- In order to control overall error rate (or, overall confidence level), let $m\alpha^{'}$, we have $\alpha^{'} = \alpha/m$.
- Bonferroni CIs:

$$
CI_i : \sum_{j=1}^{a} c_{ij}\bar{y}_{j\cdot} \pm t_{\alpha/2m, N-a} \sqrt{MSE \sum_{j=1}^{a} \frac{c_{ij}^2}{n_j}}
$$

- When m is large, Bonferroni CIs are too conservative.

Scheffe's Method for Testing All Contrasts

- Consider all possible contrasts: $\Gamma = \sum_{i=1}^{a} c_i \mu_i$.

  Estimate: $C = \sum_{i=1}^{a} c_i \bar{y}_{i\cdot}$, St. Error: $SE_C = \sqrt{MSE \sum_{i=1}^{a} \frac{c_i^2}{n_i}}$
- Critical value: $\sqrt{(a-1)F_{\alpha,a-1,N-a}}$
- Scheffe's simultaneous CI: $C \pm \sqrt{(a-1)F_{\alpha,a-1,N-a}} SE_C$
- 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 I

- 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:
  - Estimate $\bar{y}_{i\cdot} - \bar{y}_{j\cdot}$
  - Compute a **Critical Difference** (CD) (based on the method employed)
  - If

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

    or equivalently if the interval

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

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

Methods for Pairwise Comparisons II

- Least significant difference (LSD):

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

  not control overall error rate.

- Bonferroni method (for m pairs)

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

Methods for Pairwise Comparisons III

- Tukey's method (for all possible pairs)
  Tukey's method makes use of the distribution of the studentized range statistic $q = \frac{\bar{y}_{max} - \bar{y}_{min}}{\sqrt{MSE/n}}$ where $\bar{y}_{max}$ and $\bar{y}_{min}$ are the largest and smallest sample means, respectively, out of a group of a means.

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

  where $q_\alpha(a, N-a)$ from studentized range distribution (Table VII).
  Control overall error rate (exact for balanced experiments) (Examples 3.7 and 3.8).

Methods for Pairwise Comparisons IV

Tukey's method makes use of the distribution of the studentized range statistic

$$q = \frac{\bar{y}_{max} - \bar{y}_{min}}{\sqrt{MSE/n}}$$

where $\bar{y}_{max}$ and $\bar{y}_{min}$ are the largest and smallest sample means, respectively out of a group of sample means

Methods for Pairwise Comparisons V

- SNK (Student-Newman-Keuls) method Similar to Tukey's method except calculation of CD:

$$CD = q_\alpha(p, N-a)\sqrt{\frac{MSE}{n}}.$$

where p is the number of means ranging the two comparing means.
For example, $\bar{Y}_2 < \bar{Y}_5 < \bar{Y}_1 < \bar{Y}_3 < \bar{Y}_4$.
1) To compare μ_2 and μ_4, $p = 5$
2) To compare μ_5 and μ_3, $p = 3$

Comparing treatments with control (Dunnetts method)

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

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

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

- Remark: control overall error rate. Read example 3.9.

See Tensile-Comparison.SAS.

Which method should I use?

- Multiple comparisons (i.e., contrasts) but not pairwise comparisons
  - If m is very small, use Bonferroni method conservative 해서
  - If m is very large, use Scheffe method
- Pairwise comparison
  - Tukey method
  - Tukey and SNK (Student-Newman-Keuls) are commonly used
  - Duncan is too liberal (not recommended) too much rejections
  - LSD is not recommended
- 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

$$\begin{aligned} \beta &= P(\text{Accept}\, H_0 | H_0 \text{ is false}) \\ &= P(F_0 < F_{\alpha, a-1, N-a} | H_1 \text{ is correct}) \end{aligned}$$

- 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
Then $n =$?

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

Thus, 4 or 5 replicates are sufficient to obtain a test with the required power. See Sample-size.SAS

Determining Sample Size (Confidence 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