Multiple Comparison Procedures

Norberto E. Milla, Jr.

2025-09-13

Introduction

- Analysis of Variance uses an F-statistic that is used to test a null hypothesis of no differences between treatment means, $H_0 : \mu_1 = \mu_2 = \cdots = \mu_t$
- If this hypothesis is rejected, the investigator concludes that there are differences among the population means
- However, the test does not indicate which means differ
- It is desirable to perform follow -up analyses to examine the manner in which the treatment means differ

Introduction

- **Multiple comparison procedures** refer to making several tests for statistical significance of differences between means within a group of means
- They can be conducted in many different ways
- One way is to make all possible pairwise comparisons of the means using t-tests
- Types: *Planned*, *Unplanned*
- Types: *Pairwise*, *Group*, *Trend*

Types of multiple comparisons

- **Planned** (*priori*) pairwise comparison
  - specific pair of treatments to be compared was identified before the start of the experiment (e.g., control vs other treatments)
  - done without doing F-test
- **Unplanned** (*posteriori*) pairwise comparison
  - no specific comparison is chosen in advance
  - every possible pair of treatment means is compared to identify pairs of treatment means that are significantly different
  - done if F-test results in rejecting the null hypothesis

Pairwise comparison of means

- Compares pairs of treatment means and tests if they are significantly different
- Statistical hypotheses:
  - $H_0 : \mu_i = \mu_j$, for all pairs $i \neq j$
  - $H_1 : \mu_i \neq \mu_j$, for at least one pair of $i \neq j$
- Commonly used tests for pairwise mean comparison:
  - Least significant difference (LSD) test
  - Duncan's multiple range test (DMRT)
  - Student-Newman-Keul (SNK) test
  - Honest significant difference (HSD) test
  - Scheffe's (S) test

General Procedure

1. Sort the treatments in descending order of the means.
2. Compare the largest mean to each of the smaller means starting from the next smaller mean by computing the absolute difference (d) and comparing the value of d to the value of the test statistic.
   Decision Rule: Reject $H_0 : \mu_i = \mu_j$ if $d \geq$ the test statistic value.
3. Compare the next largest mean as in (2).
4. Repeat Steps (2) and (3) until the two smallest means have been compared.

Least significant difference (LSD) test

- Useful only if the F-test in the ANOVA is significant
- Can be used for experiments with equal or unequal replications
- Considered as the least conservative among pairwise comparison tests
- Test statistic:

$$LSD = t_{\frac{\alpha}{2},v}\sqrt{MSE\left(\frac{1}{r_i}+\frac{1}{r_j}\right)}$$

Least significant difference (LSD) test: an example

- Consider the results from an experiment to measure the effects of varying percentage of cotton on the tensile strength of fabric. There are five treatments (T1=15%, T2=20%, T3=25%, T4=30%, and T5=35%). Each treatment is replicated five times. The ANOVA table and the treatment means are given below.

```
Analysis of Variance Table

Response: TS
          Df Sum Sq Mean Sq F value    Pr(>F)    
Treatment  4 475.76  118.94  14.757 9.128e-06 ***
Residuals 20 161.20    8.06
```

```
  T1   T2   T3   T4   T5 
 9.8 15.4 17.6 21.6 10.8 
```

Least significant difference (LSD) test: an example

Treatment	Means
T4	21.6
T3	17.6
T2	15.4
T5	10.8
T1	9.8

- Test statistic:

$$
\begin{aligned}
LSD &= t_{\frac{\alpha}{2},v}\sqrt{MSE\left(\frac{1}{r_i}+\frac{1}{r_j}\right)} \\
&= t_{\frac{0.05}{2},20}\sqrt{8.06\left(\frac{1}{5}+\frac{1}{5}\right)} \\
&\approx 3.746
\end{aligned}
$$

Least significant difference (LSD) test: an example

Treatment	*Means*
T4	21.6
T3	17.6
T2	15.4
T5	10.8
T1	9.8

Pair	*d*
T4 vs T3	4.0*
T4 vs T2	6.2*
T4 vs T5	10.8*
T4 vs T1	11.8*
T3 vs T2	2.2^{ns}
T3 vs T5	6.8*
T3 vs T1	7.8*
T2 vs T5	4.6*
T2 vs T1	5.6*
T5 Vs T1	1.0^{ns}

Least significant difference (LSD) test: an example

Summary

Treatment	*Means*
T4	21.6[a]
T3	17.6[b]
T2	15.4[b]
T5	10.8[c]
T1	9.8[c]

Scheffe's (S) test

- Useful if the F-test in the ANOVA is significant
- Can be used for experiments with equal or unequal replications
- Most conservative among pairwise comparison tests
- Test statistic:

$$S = \sqrt{(t-1)F_{\alpha,(v_1,v_2)}MSE\left(\frac{1}{r_i}+\frac{1}{r_j}\right)}$$

Scheffe's (S) test: an example

$$
\begin{aligned}
S &= \sqrt{(t-1)F_{\alpha,(v_1,v_2)} MSE \left(\frac{1}{r_i} + \frac{1}{r_j}\right)} \\
&= \sqrt{(5-1)F_{0.05,(4,20)} 8.06 \left(\frac{1}{5} + \frac{1}{5}\right)} \\
&\approx 6.079
\end{aligned}
$$

Scheffe's (S) test: an example

Pair	*d*
T4 vs T3	4.0[ns]
T4 vs T2	6.2*
T4 vs T5	10.8*
T4 vs T1	11.8*
T3 vs T2	2.2[ns]
T3 vs T5	6.8*
T3 vs T1	7.8*
T2 vs T5	4.6[ns]
T2 vs T1	5.6[ns]
T5 Vs T1	1.0[ns]

Treatment	*Means*
T4	21.6[a]
T3	17.6[ab]
T2	15.4[bc]
T5	10.8[c]
T1	9.8[c]

Tukey's Honest Significant Difference (HSD) test

- Used only if the F-test in the ANOVA is significant
- Used only for experiments with equal replications
- More conservative than LSD
- Test statistic:

$$HSD = q_{\alpha,t,v}\sqrt{\frac{MSE}{r}}$$

- where $q_{\alpha,t,v}$ is the upper α-quantile of the studentized-range (Tukey) distribution

Tukey-Kramer's test

- Tukey (1953) and Kramer (1956) provided a modification of Tukey's HSD test when the number of replications are unequal
- Test statistic:

$$TK = q_{\alpha,t,v}\sqrt{\frac{MSE}{2}\left(\frac{1}{r_i}+\frac{1}{r_j}\right)}$$

- If $r_i = r_j = r$, then TK reduces to HSD
- the upper α-quantile of the studentized-range (Tukey) distribution was obtained using the R command $qtukey(\alpha, t, v, lower.tail = F)$

Tukey's Honest Significant Difference (HSD) test: an example

- $HSD = q_{\alpha,t,v}\sqrt{\frac{MSE}{r}} = 4.23\sqrt{\frac{8.06}{5}} \approx 5.371$

Pair	***d***
T4 vs T3	4.0^{ns}
T4 vs T2	6.2*
T4 vs T5	10.8*
T4 vs T1	11.8*
T3 vs T2	2.2^{ns}
T3 vs T5	6.8*
T3 vs T1	7.8*
T2 vs T5	4.6^{ns}
T2 vs T1	5.6*
T5 Vs T1	1.0^{ns}

Treatment	***Means***
T4	21.6^{a}
T3	17.6^{ab}
T2	15.4^{bc}
T5	10.8^{cd}
T1	9.8^{d}

Student-Newmann-Keuls (SNK) test

- Useful if the F-test in the ANOVA is significant
- Requires equal number of replicate
- Accounts for the number of treatments in the experiment
- Test statistic:

$$W_p = q_{\alpha,p,v}\sqrt{\frac{MSE}{r}}$$

- p is the number of means in the range of means being compared

Student-Newmann-Keuls (SNK) test: an example

p	***2***	***3***	***4***	***5***
$q_{0.05,p,20}$	2.95	3.58	3.96	4.23
W_p	3.745	4.545	5.028	5.371

Pair	***d***
T4 vs T3	4.0*
T4 vs T2	6.2*
T4 vs T5	10.8*
T4 vs T1	11.8*
T3 vs T2	2.2^{ns}
T3 vs T5	6.8*
T3 vs T1	7.8*
T2 vs T5	4.6*
T2 vs T1	5.6*
T5 Vs T1	1.0^{ns}

Treatment	***Means***
T4	21.6^{a}
T3	17.6^{b}
T2	15.4^{b}
T5	10.8^{c}
T1	9.8^{c}

Duncan's multiple range test (DMRT)

- More sensitive than the F-test in the ANOVA
- Requires equal replicate
- Test statistic:

$$R_p = r_{\alpha,p,v}\sqrt{\frac{MSE}{r}}$$

- p is the number of means in the range of means being compared

Duncan's multiple range test (DMRT): an example

p	*2*	*3*	*4*	*5*
$r_{0.05,p,20}$	2.95	3.1	3.19	3.26
R_p	3.75	3.94	4.05	4.139

Pair	*d*
T4 vs T3	4.0*
T4 vs T2	6.2*
T4 vs T5	10.8*
T4 vs T1	11.8*
T3 vs T2	2.2^{ns}
T3 vs T5	6.8*
T3 vs T1	7.8*
T2 vs T5	4.6*
T2 vs T1	5.6*
T5 Vs T1	1.0^{ns}

Treatment	*Means*
T4	21.6^{a}
T3	17.6^{b}
T2	15.4^{b}
T5	10.8^{c}
T1	9.8^{c}

Group comparison

Linear contrast: Let $\mu_1, \mu_2, \cdots, \mu_t$ be t population means. The linear function expressed as

$$\lambda = \sum_{i=1}^{t} c_i \mu_i = c_1\mu_1 + c_2\mu_2 + \cdots + c_t\mu_t$$

is a linear contrast (linear comparison) among the population means if

$$\sum_{i=1}^{t} c_i = 0$$

where the c_i's are fixed constants (coefficients).

Group comparison

Example: Suppose we have three treatments with means μ_1, μ_2, μ_3.

1. If we want to compare μ_1 and μ_2. Then

$$\lambda_1 = \mu_1 - \mu_2$$

2. If we want to compare μ_1 with the mean of μ_2 and μ_3, then

$$\lambda_2 = \mu_1 - \frac{1}{2}(\mu_2 + \mu_3)$$

or equivalently,

$$\lambda_3 = 2\mu_1 - (\mu_2 + \mu_3)$$

Group comparison

Using the sample data we estimate λ by L, where

$$L = \sum_{i=1}^{t} c_i \overline{Y}_{i.}$$

where: $\overline{Y}_{i.}$ is the mean of treatment i, and subject to the restriction $\sum_{i=1}^{t} c_i = 0$

Group comparison

Two contrasts are said to be (*pairwise*) **orthogonal** if the sum of the cross products of their coefficients is equal to zero. That is, given

$$\lambda_1 = \sum_{i=1}^{t} c_{1i}\mu_i$$

and

$$\lambda_2 = \sum_{i=1}^{t} c_{2i}\mu_i$$

we say λ_1 and λ_2 are **orthogonal** if

$$\sum_{i=1} c_{1i}c_{2i} = 0$$

Group comparison

Example: Suppose we have four treatment with means $\mu_1, \mu_2, \mu_3, \mu_4$. Let

$$\begin{aligned}
\lambda_1 &= 3\mu_1 - (\mu_2 + \mu_3 + \mu_4) \\
\lambda_2 &= \mu_2 - \mu_3 \\
\lambda_3 &= \mu_2 - \mu_4
\end{aligned}$$

Which of the following are orthogonal contrasts?

1. λ_1 and λ_2
2. λ_1 and λ_3
3. λ_2 and λ_3

Group comparison

The set of $(t-1)$ contrasts $\lambda_1, \lambda_2, \cdots, \lambda_{t-1}$, is a **set of orthogonal contrasts** if they are all *pairwise orthogonal*.

Example: Consider the four treatment with means $\mu_1, \mu_2, \mu_3, \mu_4$. Let

$$\lambda_1 = 3\mu_1 - (\mu_2 + \mu_3 + \mu_4)$$
$$\lambda_2 = \mu_2 - \mu_3$$
$$\lambda_3 = \mu_2 - \mu_4$$

Is the set of three contrasts orthogonal?

Testing the significance of a linear contrast

- $H_0 : \lambda = 0$
- $H_1 : \lambda \neq 0$
- Test statistic: $F = \frac{SSL}{MSE}$
  - where:

$$SSL = \frac{L^2}{\sum_{i=1}^{t} \frac{c_i^2}{r_i}}$$

- $F \sim F_{\alpha,(1,v)}$

Group comparison: an example

Consider the following results from an experiment conducted to determine the yield (tons/ha) response of a certain variety of rice to different kinds of fertilizer. The ANOVA table is shown below and the descriptions of the treatments are in the next slide.

Source of Variation	*SS*	*df*	*MS*	*F*	$F_{0.05}$	*p-value*
Treatment	59.45	5	11.89	49.54	2.77	<0.001
Error	4.30	18	0.24			
Total	63.75	23				

Group comparison: an example

TRMT	*FERTILIZER AND APPLICATION*	*r*	*TOTAL*	*MEAN*
T1	Control (no fertilizer application)	4	15.32	3.83
T2	Complete fertilizer (14-14-14)	4	31.20	7.80
T3	Organic Fertilizer A , single application	4	26.84	6.71
T4	Organic Fertilizer A , split application	4	29.20	7.30
T5	Organic Fertilizer B , single application	4	16.12	4.03
T6	Organic Fertilizer B , split application	4	19.68	4.92

Group comparison: an example

Comparisons of interest:

1. Mean of the control vs the means of all treatments: $\lambda_1 = 5\mu_1 - \mu_2 - \mu_3 - \mu_4 - \mu_5 - \mu_6$
2. Mean of complete fertilizer vs the means of the organic fertilizer: $\lambda_2 = 4\mu_2 - \mu_3 - \mu_4 - \mu_5 - \mu_6$
3. Means of organic fertilizer A vs means of organic fertilizer B: $\lambda_3 = \mu_3 + \mu_4 - \mu_5 - \mu_6$
4. Mean of organic fertilizer A single vs A split: $\lambda_4 = \mu_3 - \mu_4$
5. Mean of organic fertilizer B single vs B split: $\lambda_5 = \mu_5 - \mu_6$

Group comparison: an example

- $H_0 : \lambda_1 = 0$ versus $H_1 : \lambda_1 \neq 0$, or
- $H_0 : 5\mu_1 - \mu_2 - \mu_3 - \mu_4 - \mu_5 - \mu_6 = 0$ versus
  $H_1 : 5\mu_1 - \mu_2 - \mu_3 - \mu_4 - \mu_5 - \mu_6 \neq 0$, or
- $H_0 : 5\mu_1 = \mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6$ versus
  $H_1 : 5\mu_1 \neq \mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6$
- Test statistic:

$$
\begin{aligned}
L_1 &= 5\overline{Y}_{1.} - \overline{Y}_{2.} - \overline{Y}_{3.} - \overline{Y}_{4.} - \overline{Y}_{5.} - \overline{Y}_{6.} \\
&= 5(3.83) - 7.80 - 6.71 - 7.30 - 4.03 - 4.92 \\
&= -11.61
\end{aligned}
$$

Group comparison: an example

- Test statistic:

$$SSL_1 = \frac{L_1^2}{\sum_{i=1}^{t} \frac{c_i^2}{r_i}}$$

$$= \frac{(-11.61)^2}{\frac{5^2}{4} + \frac{(-1)^2}{4} + \frac{(-1)^2}{4} + \frac{(-1)^2}{4} + \frac{(-1)^2}{4} + \frac{(-1)^2}{4}}$$

$$\approx 17.97$$

$$\rightarrow F = \frac{SSL_1}{MSE} = \frac{17.97}{0.24} \approx 74.88$$

- Critical value: 4.41 ($qf(0.05, 1, 18, lower.tail = F)$)
- p-value: 0.00 ($pf(74.88, 1, 18, lower.tail = F)$)

Group comparison: an example

Source of Variation	*SS*	*df*	*MS*	*F*	$F_{0.05}$
Treatment	59.45	5	11.89	49.54	2.77
Control vs Treated	*17.97*	*1*	*17.97*	*74.88*	*4.41*
Complete vs Organic	*13.58*	*1*	*13.58*	*56.58*	
Organic A vs Organic B	*26.50*	*1*	*26.50*	*110.42*	
A single vs A split	*0.70*	*1*	*0.70*	*2.92*	
B single vs B split	*1.58*	*1*	*1.58*	*6.58*	
Error	4.30	18	0.24		
Total	63.75	23			

Group comparison: another example

1. Control vs Treated
2. Complete vs Organic
3. Single vs Split
4. A single vs B single
5. A split vs B split

Trend comparison

- Experiments are often designed to study the effect of increasing levels of a factor, e.g., increments of a fertilizer, planting dates, doses of a chemical, concentrations of a feed additive, etc.
- How does the yield of rice respond to increasing levels of N fertilizer?
- Used to determine the functional relationship between the treatment and the response variable
- We consider only the case where the quantitative levels of the treatment are equally spaced

Trend comparison

No. of Treatments	Degree of Polynomial	Coefficients						Sum of squared coefficients
		T1	T2	T3	T4	T5	T6	
3	Linear	1	0	-1				2
	Quadratic	1	-2	1				6
4	Linear	-3	-1	1	3			20
	Quadratic	1	-1	-1	1			4
	Cubic	-1	3	-3	1			20
5	Linear	-2	-1	0	1	2		10
	Quadratic	2	-1	-2	-1	2		14
	Cubic	-1	2	0	-2	1		10
	Quartic	1	-4	6	-4	1		70
6	Linear	-5	-3	-1	1	3	5	70
	Quadratic	5	-1	-4	-4	-1	5	84
	Cubic	-5	7	4	-4	-7	5	180
	Quartic	1	-3	2	2	-3	1	28
	Quintic	-1	5	-10	10	-5	1	252

Testing the significance of a trend

- For $t = 3$, the coefficients of a linear trend are: $c_1 = 1, c_2 = 0, c_3 = -1$, hence, $\lambda_L = \mu_1 - \mu_3$
- $H_0 : \lambda_L = 0$ (*Trend is not linear*) vs $H_1 : \lambda_L \neq 0$ (*Trend is linear*)
- Test statistic: $F = \frac{SSL_L}{MSE}$
  - where:

$$SSL_L = \frac{L_L^2}{\sum_{i=1}^{t} \frac{c_i^2}{r_i}}$$

- $F \sim F_{\alpha,(1,v)}$

Trend comparison: an example

Consider the results from an experiment conducted in CRD to determine the yield response of corn to a newly formulated organic fertilizer.

Fertilizer (kg/ha)	Observations				TOTAL	MEAN
0	4.89	4.79	4.65	4.47	18.80	4.70
50	5.08	5.19	4.89	4.92	20.08	5.02
100	5.25	5.18	5.26	5.23	20.92	5.23
150	5.38	5.37	5.46	5.31	21.52	5.38
200	5.55	5.34	5.30	5.25	21.44	5.36
150	5.29	5.22	5.40	5.21	21.12	5.28
					123.88	5.16

SV	df	SS	MS	F_c	P-value
Treatments	5	1.356	0.271	19.57**	<0.01
Error	18	0.249	0.014		
TOTAL	23	1.605			

Trend comparison: an example

Determining if the linear trend is significant

- Here we have $t = 6$ treatments, so the coefficients of a linear trend are: $c_1 = -5, c_2 = -3, c_3 = -1, c_4 = 1, c_5 = 3, c_6 = 5$
- we estimate λ_L with L_L, where

$$\begin{aligned} L_L &= \sum_{i=1}^{t} c_i \overline{Y}_{i.} \\ &= -5 \times 4.70 - 3 \times 5.02 - 1 \times 5.23 + 1 \times 5.38 + 3 \times 5.36 + 5 \times 5.28 \\ &= 4.07 \end{aligned}$$

- Verify that $SSL_L = 0.947$!
- Compute the SS Residual as follows:

$$\begin{aligned} SSResidual &= SSTr - SSL_L \\ &= 1.356 - 0.947 = 0.409 \end{aligned}$$

Trend comparison: an example

Determining if the linear trend is significant

- Revise the ANOVA Table as follows:

SV	df	SS	MS	F_c	P-value
Treatments	5	1.356	0.271	19.57**	<0.0001
Linear	1	0.947	0.947	67.64**	<0.0001
Residual	4	0.409	0.102	7.38**	0.0011
Error	18	0.249	0.014		
TOTAL	23	1.605			

- Since the residual is significant we proceed to testing the significance of the next higher-order polynomial
- If the residual is NOT significant, then there exists a linear relationship between the treatments and the response

Trend comparison: an example

Determining if the quadratic trend is significant

- For $t = 6$ treatments the coefficients of a quadratic trend are: $c_1 = 5, c_2 = -1, c_3 = -4, c_4 = -4, c_5 = -1, c_6 = 5$
- We have

$$\begin{aligned} L_Q &= 5 \times 4.70 - 1 \times 5.02 - 4 \times 5.23 - 4 \times 5.38 - 1 \times 5.36 + 5 \times 5.28 \\ &= -2.92 \end{aligned}$$

- Verify that $SSL_Q = 0.406$
- Hence,

$$\begin{aligned} SSResidual &= SSTr - SSL_L - SSL_Q \\ &= 1.356 - 0.947 - 0.406 \\ &= 0.003 \end{aligned}$$

Trend comparison: an example

SV	df	SS	MS	F_c	P-value
Treatments	5	1.356	0.271	19.57**	<0.0001
Linear	1	0.947	0.947	67.64**	<0.0001
Quadratic	1	0.406	0.406	29.00**	<0.0001
Residual	3	0.003	0.001	0.07^{ns}	0.981
Error	18	0.249	0.014		
TOTAL	23	1.605			

- Therefore, the yield response of corn to this fertilizer formulation is quadratic.