

Math 362: Mathematical Statistics II

Le Chen

le.chen@emory.edu
chenle02@gmail.com

Emory University
Atlanta, GA

Last updated on Spring 2021
Last compiled on January 15, 2023

2021 Spring

Creative Commons License
(CC By-NC-SA)

Chapter 12. The Analysis of Variance

§ 12.1 Introduction

§ 12.2 The F Test

§ 12.3 Multiple Comparisons: Turkey's Method

§ 12.4 Testing Subhypotheses with Contrasts

Chapter 12. The Analysis of Variance

§ 12.1 Introduction

§ 12.2 The F Test

§ 12.3 Multiple Comparisons: Turkey's Method

§ 12.4 Testing Subhypotheses with Contrasts



1. John Wilder Tukey (June 16, 1915 – July 26, 2000) was an American mathematician best known for development of the Fast Fourier Transform (FFT) algorithm and box plot.
2. The Tukey range test, the Tukey lambda distribution, the Tukey test of additivity, and the Teichmüller-Tukey lemma all bear his name.
3. He is also credited with coining the term 'bit'.

https://en.wikipedia.org/wiki/John_Tukey

| $N(\mu_1, \sigma^2)$ | $N(\mu_2, \sigma^2)$ | \cdots | $N(\mu_k, \sigma^2)$ |
|----------------------|----------------------|----------|----------------------|
| Y_{11} | Y_{12} | \cdots | Y_{1k} |
| Y_{21} | Y_{22} | \cdots | Y_{2k} |
| \vdots | \vdots | \vdots | \vdots |
| Y_{r1} | Y_{r2} | \cdots | Y_{rk} |

Goal For any $i \neq j$, test

$$H_0 : \mu_i = \mu_j \quad v.s. \quad H_1 : \mu_i \neq \mu_j$$

at the α level of significance defined as

$$\mathbb{P} \left(\bigcup_{j=1}^{\binom{k}{2}} E_j \right) = \alpha$$

where there are $\binom{k}{2}$ pairs, and E_j is the event of making a type I error for the j -th pair.

Goal' Simultaneous C.I.'s for $\binom{k}{2}$ pairs of means:

Given α , find I_{ij} , the C.I. for $\mu_i - \mu_j$ (with $i, j = 1, \dots, k$ and $i \neq j$), s.t.

$$\mathbb{P}(\mu_i - \mu_j \in I_{ij}, \forall i, j = 1, \dots, k, i \neq j) = 1 - \alpha.$$

??? Why not the standard pair-wise two-sample t-test?

Suppose $\mathbb{P}(E_j) = \alpha_*$. Then

$$\alpha = \mathbb{P}\left(\bigcup_{j=1}^{\binom{k}{2}} E_j\right) = 1 - \mathbb{P}\left(\bigcap_{j=1}^{\binom{k}{2}} E_j^c\right) \approx 1 - \prod_{j=1}^{\binom{k}{2}} \mathbb{P}(E_j^c) = 1 - (1 - \alpha_*)^{\binom{k}{2}}$$

Hence,

$$\alpha_* \approx 1 - (1 - \alpha)^{1/\binom{k}{2}}$$

E.g., $\alpha = 0.05$

| k | 5 | 8 | 100 |
|------------|-----------|----------|------------|
| α_* | 0.0051162 | 0.001830 | 0.00001036 |

Bonferroni's method

— A straightforward method

$$\mathbb{P}(\mu_i - \mu_j \in I_{ij}, \forall i \neq j)$$

\parallel

$$\mathbb{P}\left(\bigcap_{i \neq j} \mu_i - \mu_j \in I_{ij}\right)$$

\parallel

$$1 - \mathbb{P}\left(\bigcup_{i \neq j} \mu_i - \mu_j \notin I_{ij}\right)$$

$\vee \mid$

$$1 - \sum_{i \neq j} \mathbb{P}(\mu_i - \mu_j \notin I_{ij})$$

\parallel

$$1 - \binom{k}{2} \alpha_*$$

1. If we choose $\alpha_* = \alpha / \binom{k}{2}$,

2. let I_{ij} be the $(1 - \alpha_*)100\%$ C.I. $i \neq j$

\Downarrow

$$\mathbb{P}(\mu_i - \mu_j \in I_{ij}, \forall i \neq j)$$

$\vee \mid$

$$1 - \binom{k}{2} \alpha_*$$

\parallel

$$1 - \alpha.$$

Remark This is an approximation. The resulting C.I. are in general too wide.

The exact, and much more precise, solution is given by J.W. Turkey.

One can also construct simultaneous C.I. for all possible linear combinations of the parameters $\sum_{j=1}^k c_j \mu_j$, this can be achieved by **Scheffé's method**. A simple version is given in §12.4.

Tukey's HSD (honestly significant difference) test

Let's construct $(1 - \alpha)100\%$ C.I.'s simultaneously for all pairs.

$$\mathbb{P} \left(\left| (\bar{Y}_{.i} - \mu_i) - (\bar{Y}_{.j} - \mu_j) \right| \leq \mathcal{E}, \quad \forall i \neq j \right) = 1 - \alpha$$

\parallel

$$\mathbb{P} \left(\max_i (\bar{Y}_{.i} - \mu_i) - \min_j (\bar{Y}_{.j} - \mu_j) \leq \mathcal{E} \right)$$

\parallel

$$\mathbb{P} \left(\max_i \bar{Y}_{.i} - \min_j \bar{Y}_{.j} \leq \mathcal{E} \right)$$

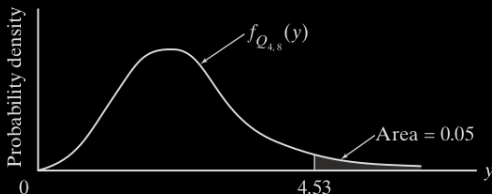
\implies Needs to study ...

Def. Let W_1, \dots, W_k be k i.i.d. r.v.'s from $N(\mu, \sigma^2)$. Let R denote their range:

$$R = \max_i W_i - \min_i W_i.$$

Let S^2 be an unbiased estimator for σ^2 independent of the W_i 's and based on ν df. Define the **Studentized range**, $Q_{k,\nu}$, to be the ratio:

$$Q_{k,\nu} := \frac{R}{S}.$$

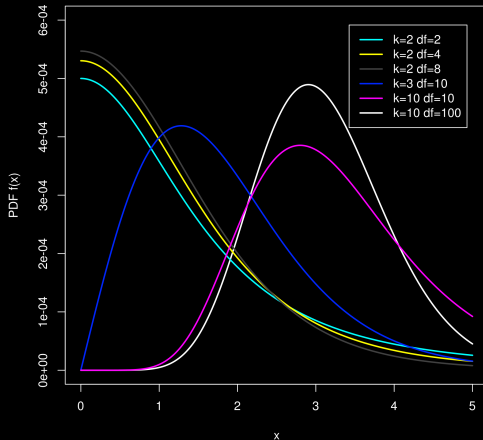


- Remark**
- 0.1 We need $R \perp S$ to mimic Student's t-distribution.
 - 0.2 In the following $\nu = n - k = rk - k = r(k - 1)$.

$Q_{k,\nu} \sim$ **Studentized range distribution** with parameters k and ν .

k : number of groups.

ν : degrees of freedom.



Let's find one example that all requirements of the $Q_{k,\nu}$ are satisfied.

$$1. \text{ Take } W_j = \bar{Y}_{.j} - \mu_j, j = 1, \dots, k \implies W_j \sim N(0, \sigma^2/r).$$

$$2. \text{ } MSE \text{ or the pooled variance } S_p^2 \quad MSE/r$$

$$\text{is an unbiased estimator for } \sigma^2 \quad \sigma^2/r$$

$$\text{is } \perp \{\bar{Y}_{.j}\}_{j=1, \dots, k}, \text{ hence } \perp \{W_j\}_{j=1, \dots, k}$$

$$3. \text{ } df \text{ of } MSE \text{ is equal to } n - k = kr - k = k(r - 1).$$

$$\implies \frac{\max_i W_i - \min_j W_j}{\sqrt{MSE/r}} \sim \text{Studentized range distribution}(k, rk - k)$$

$$\mathbb{P} \left(\frac{\max_i W_i - \min_j W_j}{\sqrt{MSE/r}} \leq Q_{\alpha,k,rk-k} \right) = 1 - \alpha$$

$$\parallel$$

$$\mathbb{P} \left(\max_i W_i - \min_j W_j \leq \frac{Q_{\alpha,k,rk-k}}{\sqrt{r}} \sqrt{MSE} \right)$$

$$\parallel$$

$$\mathbb{P} \left(|W_i - W_j| \leq \frac{Q_{\alpha,k,rk-k}}{\sqrt{r}} \sqrt{MSE}, \forall i \neq j \right)$$

$$\parallel$$

$$\mathbb{P} \left(\left| (\bar{Y}_{.i} - \bar{Y}_{.j}) - (\mu_i - \mu_j) \right| \leq \frac{Q_{\alpha,k,rk-k}}{\sqrt{r}} \sqrt{MSE}, \forall i \neq j \right)$$

$$\parallel$$

$$\mathbb{P} \left(\bar{Y}_{.i} - \bar{Y}_{.j} - \frac{Q_{\alpha,k,rk-k}}{\sqrt{r}} \sqrt{MSE} \leq \mu_i - \mu_j \leq \bar{Y}_{.i} - \bar{Y}_{.j} + \frac{Q_{\alpha,k,rk-k}}{\sqrt{r}} \sqrt{MSE}, \forall i \neq j \right)$$

Therefore, for all $i \neq j$, the $100(1 - \alpha)\%$ C.I. for $\mu_i - \mu_j$ is

$$\bar{Y}_{.i} - \bar{Y}_{.j} \pm \frac{Q_{\alpha,k,rk-k}}{\sqrt{2}} \sqrt{MSE} \sqrt{\frac{2}{r}}$$

To test $H_0 : \mu_i = \mu_j$ for specific $i \neq j$, reject H_0 in favor of $H_1 : \mu_i \neq \mu_j$ if the C.I. does NOT contain 0, at the α level of significance. \square

Note: When sample sizes are not equal, use the **Tukey-Kramer method**:

$$\bar{Y}_{.i} - \bar{Y}_{.j} \pm \frac{Q_{\alpha,k,rk-k}}{\sqrt{2}} \sqrt{MSE} \sqrt{\frac{1}{r_i} + \frac{1}{r_j}}$$

E.g. 2 A certain fraction of antibiotics injected into the bloodstream are “bound” to serum proteins. This phenomenon bears directly on the effectiveness of the medication, because the binding decreases the systemic uptake of the drug. Table below lists the binding percentages in bovine serum measured for five widely prescribed antibiotics. Which antibiotics have similar binding properties, and which are different?

| Table 12.3.1 | | | | | |
|---------------------|-----------------|-------------------|-------------------|-------------------|----------------------|
| | Penicillin G | Tetra- cycline | Strepto- mycin | Erythro- mycin | Chloram- phenicol |
| | 29.6 | 27.3 | 5.8 | 21.6 | 29.2 |
| | 24.3 | 32.6 | 6.2 | 17.4 | 32.8 |
| | 28.5 | 30.8 | 11.0 | 18.3 | 25.0 |
| | 32.0 | 34.8 | 8.3 | 19.0 | 24.2 |
| T_j | 114.4 | 125.5 | 31.3 | 76.3 | 111.2 |
| \bar{Y}_j | 28.6 | 31.4 | 7.8 | 19.1 | 27.8 |

To answer that question requires that we make all $\binom{5}{2} = 10$ pairwise comparisons of μ_i versus μ_j . First, MSE must be computed. From the entries in Table 12.3.1,

$$SSE = \sum_{j=1}^5 \sum_{i=1}^4 (Y_{ij} - \bar{Y}_{.j})^2 = 135.83$$

so $MSE = 135.83/(20 - 5) = 9.06$. Let $\alpha = 0.05$. Since $n - k = 20 - 5 = 15$, the appropriate cutoff from the studentized range distribution is $Q_{.05,5,15} = 4.37$. Therefore, $D = 4.37/\sqrt{4} = 2.185$ and $D\sqrt{MSE} = 6.58$.

Table 12.3.2

| Pairwise Difference | $\bar{Y}_{.i} - \bar{Y}_{.j}$ | Tukey Interval | Conclusion |
|---------------------|-------------------------------|------------------|------------|
| $\mu_1 - \mu_2$ | -2.8 | (-9.38, 3.78) | NS |
| $\mu_1 - \mu_3$ | 20.8 | (14.22, 27.38) | Reject |
| $\mu_1 - \mu_4$ | 9.5 | (2.92, 16.08) | Reject |
| $\mu_1 - \mu_5$ | 0.8 | (-5.78, 7.38) | NS |
| $\mu_2 - \mu_3$ | 23.6 | (17.02, 30.18) | Reject |
| $\mu_2 - \mu_4$ | 12.3 | (5.72, 18.88) | Reject |
| $\mu_2 - \mu_5$ | 3.6 | (-2.98, 10.18) | NS |
| $\mu_3 - \mu_4$ | -11.3 | (-17.88, -4.72) | Reject |
| $\mu_3 - \mu_5$ | -20.0 | (-26.58, -13.42) | Reject |
| $\mu_4 - \mu_5$ | -8.7 | (-15.28, -2.12) | Reject |

```

1 > # Case Study 12.3.1
2 > # Input data first
3 > Input <- c("
4 + rates group
5 + 29.6 M1
6 + 24.3 M1
7 + 28.5 M1
8 + 32.0 M1
9 + 27.3 M2
10 + 32.6 M2
11 + 30.8 M2
12 + 34.8 M2
13 + 5.8 M3
14 + 6.2 M3
15 + 11.0 M3
16 + 8.3 M3
17 + 21.6 M4
18 + 17.4 M4
19 + 18.3 M4
20 + 19.0 M4
21 + 29.2 M5
22 + 32.8 M5
23 + 25.0 M5
24 + 24.2 M5
25 + ")
26 > Data = read.table(
27   textConnection(Input),
28   + header=TRUE)

```

```

1 > # Compute one-way ANOVA test
2 > res.aov <- aov(rates ~ group, data = Data)
3 > # Summary of the analysis
4 > summary(res.aov)
5           Df Sum Sq Mean Sq F value Pr(>F)
6 group         4 1480.8   370.2   40.88 6.74e-08 ***
7 Residuals    15  135.8     9.1
8 ---
9 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

1 > # Tukey multiple pairwise-comparisons
2 > TukeyHSD(res.aov)
3   Tukey multiple comparisons of means
4     95% family-wise confidence level
5
6 Fit : aov(formula = rates ~ group, data = Data)
7
8 $group
9      diff      lwr      upr    p adj
10 M2-M1 2.775 -3.795401  9.345401 0.6928357
11 M3-M1 -20.775 -27.345401 -14.204599 0.0000006
12 M4-M1 -9.525 -16.095401 -2.954599 0.0034588
13 M5-M1 -0.800 -7.370401  5.770401 0.9952758
14 M3-M2 -23.550 -30.120401 -16.979599 0.0000001
15 M4-M2 -12.300 -18.870401 -5.729599 0.0003007
16 M5-M2 -3.575 -10.145401  2.995401 0.4737713
17 M4-M3 11.250  4.679599 17.820401 0.0007429
18 M5-M3 19.975 13.404599 26.545401 0.0000010
19 M5-M4 8.725  2.154599 15.295401 0.0071611

```

```

1 > round(TukeyHSD(res.aov)$group,2)
2           diff      lwr      upr    p adj
3 M2-M1      2.78     -3.80      9.35    0.69
4 M3-M1    -20.77    -27.35    -14.20    0.00
5 M4-M1     -9.52    -16.10     -2.95    0.00
6 M5-M1     -0.80     -7.37      5.77    1.00
7 M3-M2    -23.55    -30.12    -16.98    0.00
8 M4-M2    -12.30    -18.87     -5.73    0.00
9 M5-M2     -3.58    -10.15      3.00    0.47
10 M4-M3     11.25      4.68     17.82    0.00
11 M5-M3     19.97     13.40     26.55    0.00
12 M5-M4      8.73      2.15     15.30    0.01
13 ---
14 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
15 (Adjusted p values reported — single-step method)

```

Table 12.3.2

| Pairwise Difference | $\bar{Y}_i - \bar{Y}_j$ | Tukey Interval | Conclusion |
|---------------------|-------------------------|------------------|------------|
| $\mu_1 - \mu_2$ | -2.8 | (-9.38, 3.78) | NS |
| $\mu_1 - \mu_3$ | 20.8 | (14.22, 27.38) | Reject |
| $\mu_1 - \mu_4$ | 9.5 | (2.92, 16.08) | Reject |
| $\mu_1 - \mu_5$ | 0.8 | (-5.78, 7.38) | NS |
| $\mu_2 - \mu_3$ | 23.6 | (17.02, 30.18) | Reject |
| $\mu_2 - \mu_4$ | 12.3 | (5.72, 18.88) | Reject |
| $\mu_2 - \mu_5$ | 3.6 | (-2.98, 10.18) | NS |
| $\mu_3 - \mu_4$ | -11.3 | (-17.88, -4.72) | Reject |
| $\mu_3 - \mu_5$ | -20.0 | (-26.58, -13.42) | Reject |
| $\mu_4 - \mu_5$ | -8.7 | (-15.28, -2.12) | Reject |

```

1 > # Or one may use multcomp package or multiple comparisons
2 > library (multcomp)
3 > summary(glht(res.aov, linfct = mcp(group = "Tukey"))))
4
5 Simultaneous Tests for General Linear Hypotheses
6
7 Multiple Comparisons of Means: Tukey Contrasts
8
9
10 Fit : aov(formula = rates ~ group, data = Data)
11
12 Linear Hypotheses:
13           Estimate Std. Error t value Pr(>|t|)
14 M2 - M1 == 0  2.775      2.128   1.304 0.69283
15 M3 - M1 == 0 -20.775     2.128  -9.764 < 0.001 ***
16 M4 - M1 == 0  -9.525     2.128  -4.477 0.00348 **
17 M5 - M1 == 0  -0.800     2.128  -0.376 0.99528
18 M3 - M2 == 0 -23.550     2.128 -11.068 < 0.001 ***
19 M4 - M2 == 0 -12.300     2.128  -5.781 < 0.001 ***
20 M5 - M2 == 0  -3.575     2.128  -1.680 0.47374
21 M4 - M3 == 0  11.250     2.128   5.287 < 0.001 ***
22 M5 - M3 == 0  19.975     2.128   9.388 < 0.001 ***
23 M5 - M4 == 0   8.725     2.128   4.101 0.00717 **
24 ---
25 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
26 (Adjusted p values reported -- single-step method)

```

| | | Estimate | Std. Error | t value | Pr(> t) |
|----|--------------|----------|------------|---------|-------------|
| 2 | M2 – M1 == 0 | 2.775 | 2.128 | 1.304 | 0.69283 |
| 3 | M3 – M1 == 0 | -20.775 | 2.128 | -9.764 | < 0.001 *** |
| 4 | M4 – M1 == 0 | -9.525 | 2.128 | -4.477 | 0.00348 ** |
| 5 | M5 – M1 == 0 | -0.800 | 2.128 | -0.376 | 0.99527 |
| 6 | M3 – M2 == 0 | -23.550 | 2.128 | -11.068 | < 0.001 *** |
| 7 | M4 – M2 == 0 | -12.300 | 2.128 | -5.781 | < 0.001 *** |
| 8 | M5 – M2 == 0 | -3.575 | 2.128 | -1.680 | 0.47371 |
| 9 | M4 – M3 == 0 | 11.250 | 2.128 | 5.287 | < 0.001 *** |
| 10 | M5 – M3 == 0 | 19.975 | 2.128 | 9.388 | < 0.001 *** |
| 11 | M5 – M4 == 0 | 8.725 | 2.128 | 4.101 | 0.00719 ** |

Table 12.3.2

| Pairwise Difference | $\bar{Y}_i - \bar{Y}_j$ | Tukey Interval | Conclusion |
|---------------------|-------------------------|------------------|------------|
| $\mu_1 - \mu_2$ | -2.8 | (-9.38, 3.78) | NS |
| $\mu_1 - \mu_3$ | 20.8 | (14.22, 27.38) | Reject |
| $\mu_1 - \mu_4$ | 9.5 | (2.92, 16.08) | Reject |
| $\mu_1 - \mu_5$ | 0.8 | (-5.78, 7.38) | NS |
| $\mu_2 - \mu_3$ | 23.6 | (17.02, 30.18) | Reject |
| $\mu_2 - \mu_4$ | 12.3 | (5.72, 18.88) | Reject |
| $\mu_2 - \mu_5$ | 3.6 | (-2.98, 10.18) | NS |
| $\mu_3 - \mu_4$ | -11.3 | (-17.88, -4.72) | Reject |
| $\mu_3 - \mu_5$ | -20.0 | (-26.58, -13.42) | Reject |
| $\mu_4 - \mu_5$ | -8.7 | (-15.28, -2.12) | Reject |

Two more examples of ANOVA using R

E.g. 1 <http://www.sthda.com/english/wiki/one-way-anova-test-in-r>

E.g. 2 <https://datascienceplus.com/one-way-anova-in-r/>