DATA603 Statistical Modelling

Part1: Introduction to Experimental Design

What Is Experimentation?

Experimentation is part of everyday life. Will leaving 30 minutes earlier than usual in the morning make it easier to find a legal parking space at work? How about 20 minutes earlier? Or only 10 minutes earlier? Can I increase my gas mileage by using synthetic oil? Will my employees make more of an effort to be on time if I make it a practice to stop by their office to chat at the start of the day? Will a chemical reaction be faster if the amount of a specific reagent is increased threefold? How about if the temperature is increased by 10 degree celcius? Will the yield increase if an extraction is carried for 40 minutes instead of 20 minutes?

We're frequently interested to learn if and how a measure of performance is influenced by our manipulation of the factors that might affect that measure. Usually we undertake these activities in an informal manner, typically not even thinking of them as experimentation, and the stakes are such that an informal, unstructured approach is quite appropriate. Not surprisingly, as the consequences grow, if the performance improvement means a substantial increase in profitability, or the running of the experiment involves a significant expenditure of time and resources, the adoption of a more structured experimental approach becomes more important. In a research setting, **experimentation is a tool to identify the effect of a factor (with statistical significance) on a response**. On the other hand, the purpose of experimentation in an industrial context is often to obtain the maximum amount of information about different factors that can affect a process with the fewest number of observations possible.

we learned that a regression analysis of observational data has some limitations. In particular, establishing a cause-and-effect relationship between an independent variable x and the response y is difficult since the values of other relevant independent variables-both those in the model and those omitted from the model-are not controlled.

Experimental Design Terminology

The study of experimental design originated with R. A. Fisher in the early 1900s in England. During these early years, it was associated solely with agricultural experimentation. The need for experimental design in agriculture was very clear: different fertilizer blends were applied to a crop in an effort to find the particular blend that maximized crop yield. Similar motivations led to its subsequent acceptance and wide use in all fields of scientific experimentation.

The terminology associated with experimental design clearly indicates its early association with the biological sciences. We will call the process of collecting sample data **an experiment** and the (dependent) variable to be measured, the response y. The planning of the sampling procedure is called **the design of the experiment**.The object upon which the response measurement y is taken is called **an experimental unit**.

Definition 1: The process of collecting sample data is called **an experiment**.

Definition 2: The plan for collecting the sample is called **the design of the experiment.**

Definition 3: The variable measured in the experiment is called **the response variable.**

Definition 4: The object upon which the response y is measured is called **an experimental unit.**

Definition 5: The independent variables, quantitative or qualitative, that are related to a response variable y are called **factors.**

Definition 6: The intensity setting of a factor (i.e., the value assumed by a factor in an experiment) is called **a level.**

Definition 7: **A treatment** is a particular combination of levels of the factors involved in an experiment.

The three important concepts of Design of Experiment (DOE) are

I. Randomization

II. Replication

III. Blocking

The Six Steps of Experimental Design

One can frame the experimental-design process as a six-step process, as following

1. Plan the experiment.
2. Design the experiment.
3. Perform the experiment.
4. Analyze the data from the experiment.
5. Confirm the results of the experiment.
6. Evaluate the conclusions of the experiment

Completely Randomized Designs (The CRD Model)

The Statistical Model

Let Y_{ij} denote the j^{th} experimental unit for the i^{th} treatment, with $j = 1, 2, 3, \ldots, r$ and $i = 1, 2, 3, \ldots c$. We assume that the number of observations for each treatment is the same as for any other treatment. The model for a complete randomized design with a single factor can be written as

$$
\begin{aligned}
Y_{ij} &= \mu_i + \epsilon_{ij} \\
j &= 1, 2, 3, \ldots r \\
i &= 1, 2, 3, \ldots c
\end{aligned}
$$

where ϵ_{ij} is random error with mean 0 and variance σ^2. This model is known as the **means model**. The more general and useful mathematical model for CRD is

$$
\begin{aligned}
Y_{ij} &= \mu + \tau_i + \epsilon_{ij} \\
&where \\
j &= 1, 2, 3, \ldots r \\
i &= 1, 2, 3, \ldots c \\
\mu &= \text{overall average(mean)} \\
\tau_i &= \text{a reflection of the effect due to the treatment i}
\end{aligned}
$$

The CRD model in this form is known as **the effects model**. The effects model is more feasible from a practical point of view. In this form, the mean μ is a parameter common to all treatments. This parameter is also known as the baseline or control treatment. The parameter values τ_i are a reflection of the effect due to the treatment i. The error component, or the noise factor, ϵ_{ij} are noise factors associated with treatment i and experimental unit j. We will assume that the errors are iid as N(0,σ^2)

In both models each treatment receives an equal number of experimental units, that is, each treatment i receives r number of units. These kinds of models are called *balanced designs*. In practical set ups, it may not be feasible to allocate equal numbers of units, and we allow the i^{th} treatment of $r_i, i = 1, 2, \ldots, c$ number of units. In this case, the model is called **the unbalanced design**. The inferential aspects of balanced or unbalanced models do not vary drastically from each other, at least for the CRD model.

Inference for the CRD Models

Observations	Treatment(Level)					
	1	2	3	...	c	
	y_{11}	y_{21}	y_{31}		y_{c1}	
	y_{12}	y_{22}	y_{32}		y_{c2}	
	y_{13}	y_{23}	y_{33}		y_{c3}	
	. . .				. . .	
	y_{1r}	y_{2r}	y_{3r}		y_{cr}	
Totals	$y_{1.}$	$y_{2.}$	$y_{3.}$		$y_{c.}$	$y_{..}$
Average	$\bar{y}_{1.}$	$\bar{y}_{2.}$	$\bar{y}_{3.}$		$\bar{y}_{c..}$	$\bar{y}_{..}$

Typical Data for a single-factor Experiment

For example

Assume that an experiment is performed involving 120 patients. The objective of the experiment is to test a blood pressure medication against a purported placebo, but the placebo is actually garlic, suitably disguised. Sixty patients are randomly assigned to the medication, with the other 60 patients assigned to the placebo. The study is double blinded so that the investigators do not know the patient-medication/placebo assignment, and of course the patients don't know this either. The correct assignment is known only by the person who numbered the bottles, with this information later used to properly guide the computer analysis. "P" denoting the placebo and "M" denoting the medication, the results are given in **bloodpressure.csv** data file.

Observations	Treatment	
	M	P
	115	120
	100	110
	112	115
	90	90
	95	96
	.	.
	.	.
	.	.
	n1=60	n2=60

Data for blood pressure experiment

A few standard notations and the composition of the total sum of squares are as following;

The i_{th} treatment sample sum (resp. mean), denoted by $y_{i.}(resp.\ \bar{y}_{i.})$ are defined by

$$y_{i.} = \sum_{j=1}^{r} y_{ij}$$

$$\bar{y}_{i.} = \frac{y_{i.}}{r}$$

The i^{th} total sample sum (mean), denoted by $y_{..}(resp.\ \bar{y_{..}})$ are defined by

$$y_{..} = \sum_{i=1}^{c} y_{i.}$$

$$\bar{y}_{..} = \frac{y_{..}}{rc}$$

The total (corrected) sum of squares, denoted by **SST**, as

$$SST = \sum_{i=1}^{c} \sum_{j=1}^{r} (y_{ij} - \bar{y}_{..})^2.$$

The ANOVA technique partitions the SST as the sum of two components

I. the sum of squares due to treatments SSTr, and

II. the sum of squares due to error SSE

Here, SSTr and SSE are defined respectively by

$$SSTr = r \sum_{i=1}^{c} (\bar{y}_{i.} - \bar{y}_{..})^2,$$

$$SSE = \sum_{i=1}^{c} \sum_{j=1}^{r} (y_{ij} - \bar{y}_{i.})^2.$$

$$SST = SSTr + SSE$$

Define

$$S_i^2 = \frac{\sum_{j=1}^{r} (y_{ij} - \bar{y}_{i.})^2}{r-1}$$

$$i = 1, 2, \ldots, c$$

That is, S_i^2 is the sampling variance of the i-th treatment.

We can pool these c sampling variances and obtain the following:

$$MSE = \frac{(r-1)S_1^2 + (r-1)S_2^2 + (r-1)S_3^2 + \ldots + (r-1)S_c^2}{(r-1) + \ldots (r-1)} = \frac{SSE}{N-c}$$

where

MSE denotes the mean error sum of squares, and $N = rc$. Note that S_i^2 is an estimator of the variance σ^2 for the i-th treatment.

$$MSTr = \frac{SSTr}{c-1}$$

MSTr denotes the mean treatment sum of squares

The Analysis of (the) Variance (ANOVA)

Source of Variation	Df	Sum of Squares	Mean Square	F-Statistic
Between Treatment	c-1	SSTr	MSTr	MSTr/MSE
Error within Treatments	N-c	SSE	MSE	
Total	N-1	SST		

The Anova Table for CRD

Forming the F Statistic: Logic and Derivation

We are interested in testing the quality of the c treatment means; that is

$$H_o : \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_c$$
$$H_a : \text{at least one } \mu_i \text{ is different } i = 1, 2, 3, \ldots, c$$

Note that if H_0 is true, all treatments have a common mean μ. An equivalent way to write to the above hypotheses is in terms of the treatment effect τ_i, say

$$H_o : \tau_1 = \tau_2 = \tau_3 = \ldots = \tau_c = 0$$
$$H_a : \tau_i \neq 0 \text{ for at least one i } i = 1, 2, 3, \ldots, c$$

Thus, we may speak of testing the equality of treatment means or testing that the treatment effects (the $\tau_i = 0$) are zero. The appropriate procedure for testing the equality of c treatment means is the analysis of variance.

Now we investigate how a formal test of the hypothesis of no difference in treatment means can be performed. Under the assumption that the error ϵ_{ij} are normally and independently distributed with mean = 0 and variance σ^2. Therefore, if the null hypothesis of no difference in treatment means is true, the ratio

$$F = \frac{MSTr}{MSE}$$

is distributed as F (Fisher) with $c - 1$ and $N - c$ degrees of freedom. It is the test statistic for the hypothesis of no differences in treatment means.

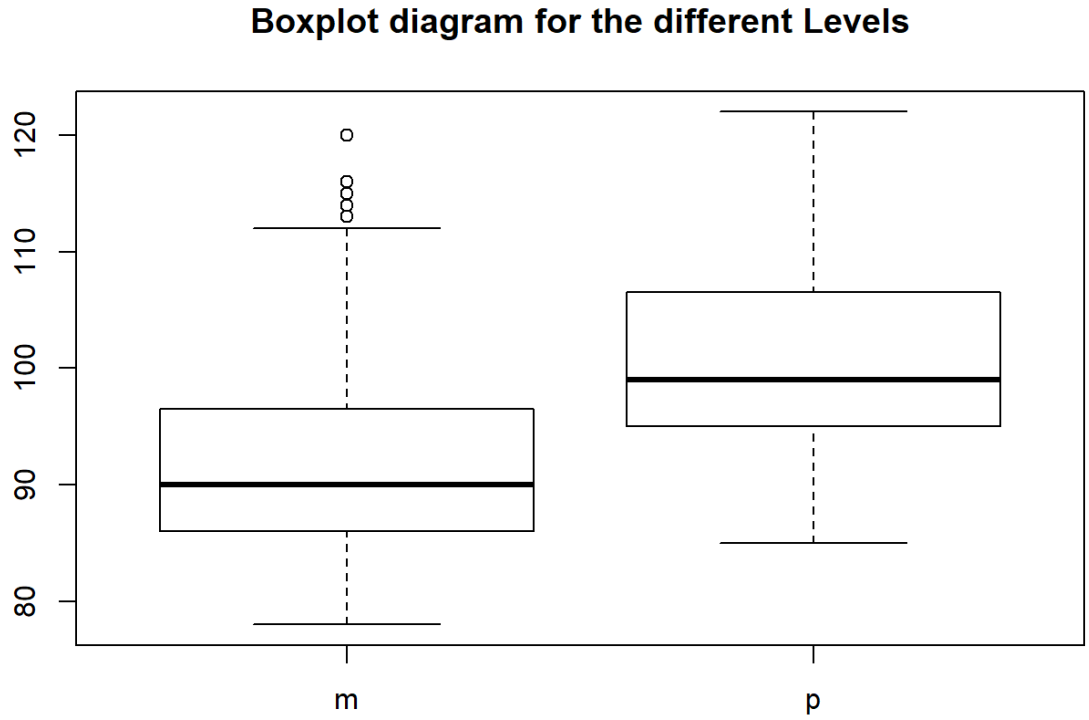
We reject H_0 and conclude that there are differences in the treatment means if $F > F_{\alpha,c-1,N-c}$ Alternatively we could use the p-value approach for decision making.

Example for one factor, Two levels: Assume that an experiment is performed involving 120 patients. The objective of the experiment is to test a blood pressure medication against a purported placebo, but the placebo is actually garlic, suitably disguised. Sixty patients are randomly assigned to the medication, with the other 60 patients assigned to the placebo. The study is double blinded so that the investigators do not know the patient-medication/placebo assignment,and of course the patients don't know this either. The correct assignment is known only by the person who numbered the bottles, with this information later used to properly guide the computer analysis. "P" denoting the placebo and "M" denoting the medication, the results are givenin **bloodpressure.csv** data file. Test if there were differences in the average blood pressure between the a blood pressure medication against a purported placebo.

```
bloodpressure=read.csv("c:/Users/thuntida.ngamkham/OneDrive - University of Calgary/datase
t603/bloodpressure.csv")
str(bloodpressure) #Read your data set and double check that dependent and independent var
iables are correctly read by R
```

```
## 'data.frame':    120 obs. of  2 variables:
##  $ bloodpressure: int  115 100 112 90 95 89 78 114 113 112 ...
##  $ treatment    : Factor w/ 2 levels "m","p": 1 1 1 1 1 1 1 1 1 1 ...
```

```
CRD<-aov(bloodpressure~treatment, data=bloodpressure) #Perform ANOVA for CRD
boxplot(bloodpressure~treatment, data=bloodpressure, main="Boxplot diagram for the differe
nt Levels") #a visual comparison of the data obtained at the different levels
```



From the blood pressure example, we test

$$H_o : \mu_1 = \mu_2$$
$$H_a : \text{at least one } \mu_i \text{ is different } i = 1, 2$$

An equivalent way to write to the above hypotheses is in term of the treatment effect τ_i, say

$$H_o : \tau_1 = \tau_2$$
$$H_a : \tau_i \neq 0 \text{ for at least one } i \; i = 1, 2$$

The output provide the Anova table as following,

Source of Variation	Df	Sum of Square	Mean Square	F	P-value
Between Treatment	1	1534	1533.7	16.19	0.000102
Within Treatment (Error/Residuals)	118	11179	94.7		
Total	119	12713			

From the Anova table, it can be seen that the Fcal=16.19 with the p-value =0.000102 < $\alpha = 0.05$, so we reject the null hypothesis. Therefore, there is sufficient evidence to indicate that the average blood pressure between a blood pressure medication against a purported placebo are different at $\alpha = 0.05$.

Note! One of the most common tests in statistics, the t-test, is used to determine whether the means of two groups are equal to each other. We also could use this test as well.

```
bloodpressure=read.csv("c:/Users/thuntida.ngamkham/OneDrive - University of Calgary/datase
t603/bloodpressure.csv", header=TRUE)
t.test(bloodpressure~treatment, data = bloodpressure)
```

```
##
##  Welch Two Sample t-test
##
## data:  bloodpressure by treatment
## t = -4.0236, df = 117.85, p-value = 0.0001016
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -10.669037  -3.630963
## sample estimates:
## mean in group m mean in group p
##           93.35          100.50
```

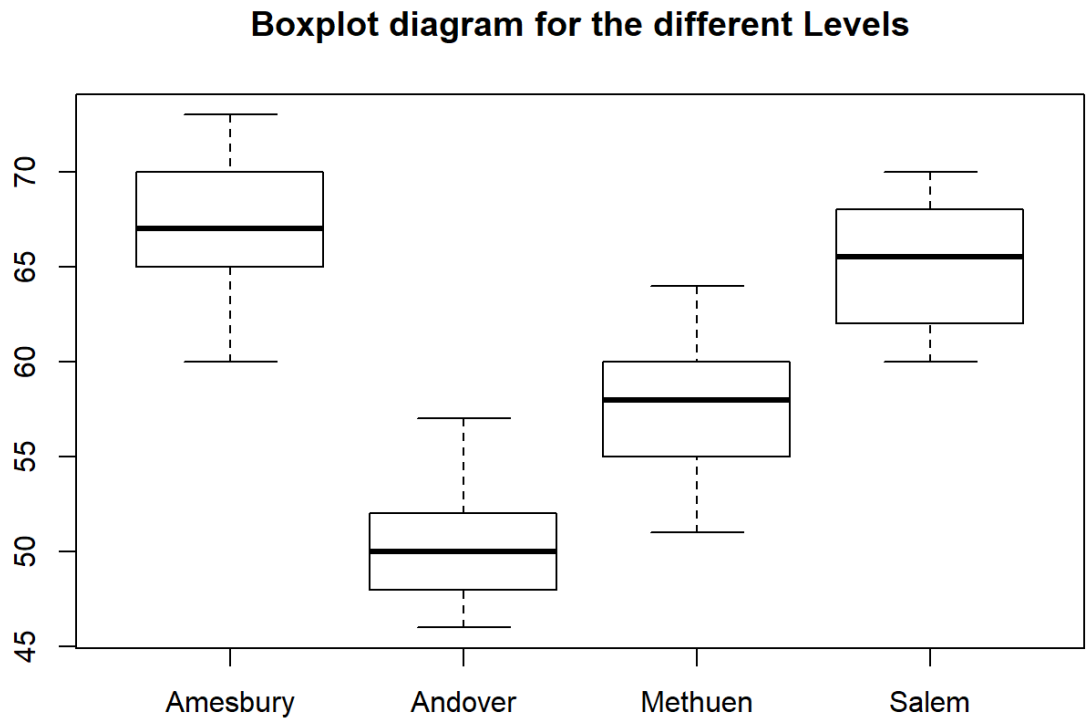
Example for one factor, more than two levels: The Merrimack Valley Pediatric Clinic (MVPC) conducted a customer satisfaction study at its four locations: Amesbury, Andover, and Methuen in Massachusetts, and Salem in southern New Hampshire. A series of questions were asked, and a respondent's "overall level of satisfaction" (using MVPC's terminology) was computed by adding together the numerical responses to the various questions. The response to each question was 1, 2, 3, 4, or 5, corresponding to, respectively, "very unsatisfied," "moderately unsatisfied," "neither unsatisfied nor satisfied," "moderately satisfied," and "very satisfied." In our discussion, we ignore the possibility that responses can be treated as an interval scale.

There were 16 questions with the possibility of a 5-rating on each, so the minimum score total was 16 and the maximum score total was 80. (For proprietary reasons, we cannot provide the specific questions.) Marion Earle, MVPC's medical director, wanted to know (among other things) if there were differences in the average level of satisfaction among customers in the four locations. Data from a random sample of 30 responders from each of the four locations are provided in **MVPC data file**

```
MVPC=read.csv("c:/Users/thuntida.ngamkham/OneDrive - University of Calgary/dataset603/MVP
C.csv", header=TRUE)
str(MVPC)#Read your data set and double check that dependent and indepent variables are co
rrectly read by R
```

```
## 'data.frame':    120 obs. of  2 variables:
##  $ Score    : int  66 66 66 67 70 64 71 66 71 67 ...
##  $ Treatment: Factor w/ 4 levels "Amesbury","Andover",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```
CRD<-aov(Score~Treatment, data=MVPC) #Perform ANOVA for CRD
boxplot(Score~Treatment, data=MVPC, main="Boxplot diagram for the different Levels") #a vi
sual comparison of the data obtained at the different levels
```




```
summary(CRD)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Treatment     3   5296  1765.5   205.3 <2e-16 ***
## Residuals   116    998     8.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As it can be seen in the output that the F value is quite large (205.3) with a p-value (called “Prob > F”) <2e-16, indicating that at any practical significance level, we reject the hypothesis that there is no difference among mean satisfaction levels for the four locations, in favor of there being differences among them.

Inclass Practice Problem

Suppose that we wish to inquire how the mean lifetime of a certain manufacturer's AA-cell battery under constant use is affected by the specific device in which it is used. It is well known that batteries of different devices have different mean lifetimes that depend on how the battery is used - constantly, intermittently with certain patterns of usage. The results of battery lifetime testing are necessary to convince a TV network to run an advertisement that claims superiority of one device over another. The testing is traditionally carried out by an independent testing agency, and the data are analyzed by an independent consultant. Suppose that we choose a production run of AA high-current-drain alkaline batteries and put to each of eight test devices; all test devices have the same nominal load impedance. Our dependent variable (yield, response, quality indicator), Y, is lifetime of the battery, measured in hours. Data **lifetime.csv file** from a random sample of 24 responders from each of the 8 brands are provided.

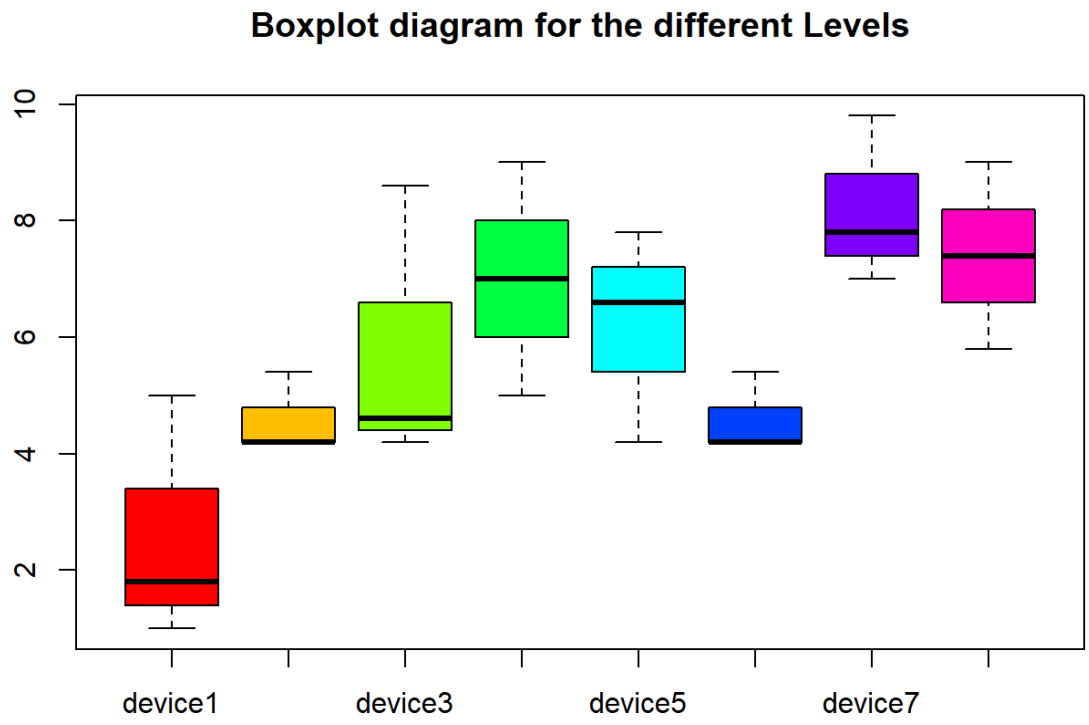
observations	Treatment							
	Device1	Device2	Device3	Device4	Device5	Device6	Device7	Device8
	1.8	4.2	8.6	.	.	.	.	.
	5	5.4	4.6	.	.	.	.	.
	1	4.2	4.2	.	.	.	.	.

Sample data for battery lifetime example

```
lifetime=read.csv("c:/Users/thuntida.ngamkham/OneDrive - University of Calgary/dataset603/
lifetime.csv", header=TRUE)
str(lifetime)#Read your data set and double check that dependent and independent variables
are correctly read by R
```

```
## 'data.frame':    24 obs. of  2 variables:
##  $ device: Factor w/ 8 levels "device1","device2",..: 1 1 1 2 2 2 3 3 3 4 ...
##  $ hrs   : num  1.8 5 1 4.2 5.4 4.2 8.6 4.6 4.2 7 ...
```

```
CRD<-aov(hrs~device, data=lifetime) #Perform ANOVA for CRD
boxplot(hrs~device, data=lifetime, main="Boxplot diagram for the different Levels", col= r
ainbow(8)) #a visual comparison of the data obtained at the different levels
```




```
summary(CRD)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## device       7  69.12   9.874   3.382 0.0206 *
## Residuals   16  46.72   2.920
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Estimation of the Model Parameters

We now present estimators for the parameters in the single-factor model and confidence intervals on treatment means

$$y_{ij} = \mu_i + \epsilon_{ij}$$
$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

and confidence intervals on treatment means. A point estimator of μ_i would be

$$\hat{\mu}_i = \hat{\mu} + \hat{\tau}_i = \bar{y}_{i.}$$

Therefore, a 100(1-α)% confidence interval on the i^{th} treatment mean μ_i is

$$\bar{y}_{i.} \pm t_{\alpha/2,N-c}\sqrt{MSE/r}$$

Assume that we wish to compute the confidence interval for batteries used on device6

```
lifetime=read.csv("c:/Users/thuntida.ngamkham/OneDrive - University of Calgary/dataset603/
lifetime.csv", header=TRUE)
CRD<-aov(hrs~device, data=lifetime) #Perform ANOVA for CRD
summary(CRD)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## device       7  69.12   9.874   3.382 0.0206 *
## Residuals   16  46.72   2.920
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ybar<-mean(lifetime$hrs[lifetime$device == "device6"])
tcrit<-qt(0.025,CRD$df.residual, lower.tail = F)
MSE<-sum((CRD$residuals)^2/CRD$df.residual)  #CRD$df.residual=24-8=16
r<-length(lifetime$hrs[lifetime$device == "device6"])
#construct a 95% CI
LowerCI<-ybar-tcrit*sqrt(MSE/r)
UpperCI<-ybar+tcrit*sqrt(MSE/r)
CI<-cbind(LowerCI,UpperCI)
print(CI)
```

```
##       LowerCI  UpperCI
## [1,] 2.508551 6.691449
```

Therefore, a 95% confidence interval for a mean life time of device6 is between 2.508551 hours to 6.691449 hours.

Inclass Practice Problem

From the Merrimack Valley Pediatric Clinic (MVPC), construct a 99% confidence Interval on the average customer satisfaction at Amesbury.

```
MVPC=read.csv("c:/Users/thuntida.ngamkham/OneDrive - University of Calgary/dataset603/MVP
C.csv", header=TRUE)
str(MVPC)#Read your data set and double check that dependent and indepent variables are co
rrectly read by R
```

```
## 'data.frame':    120 obs. of  2 variables:
##  $ Score    : int  66 66 66 67 70 64 71 66 71 67 ...
##  $ Treatment: Factor w/ 4 levels "Amesbury","Andover",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```
CRD<-aov(Score~Treatment, data=MVPC) #Perform ANOVA for CRD
summary(CRD)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Treatment     3   5296  1765.5   205.3 <2e-16 ***
## Residuals   116    998     8.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ybar<-mean(MVPC$Score[MVPC$Treatment == "Amesbury"])
tcrit<-qt(0.005,CRD$df.residual, lower.tail = F)
MSE<-sum((CRD$residuals)^2/CRD$df.residual)  #CRD$df.residual=24-8=16
r<-length(MVPC$Score[MVPC$Treatment == "Amesbury"])
#construct a 95% CI
LowerCI<-ybar-tcrit*sqrt(MSE/r)
UpperCI<-ybar+tcrit*sqrt(MSE/r)
CI<-cbind(LowerCI,UpperCI)
print(CI)
```

```
##       LowerCI  UpperCI
## [1,] 65.69784 68.50216
```

Inclass Practice Problem

A state securities laws was wondering whether five brokers have a difference average on stock price (hundreds dollars). The Data **brokerstudy.csv** from a random sample of 30 observations from each of the 5 brokers are provided. Based on the data, formally test a claim that average prices are the same for all brokers at α=0.05.

```
brokerstudy=read.csv("c:/Users/thuntida.ngamkham/OneDrive - University of Calgary/dataset603/brokerstudy.csv", header=TRUE)
str(brokerstudy)
```

```
## 'data.frame':    30 obs. of  2 variables:
##  $ broker: Factor w/ 5 levels "broker1","broker2",..: 1 1 1 1 1 1 2 2 2 2 ...
##  $ price : int  12 3 5 1 12 5 7 17 13 11 ...
```

```
CRD<-aov(price~broker, data=brokerstudy) #Perform ANOVA for CRD
summary(CRD)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## broker       4  622.1  155.53   7.695 0.000347 ***
## Residuals   25  505.3   20.21
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Multiple-Comparison Testing

So far, we have seen a couple of statistical tests which can indicate if a factor has an impact on the response or not, and which would make us reject or accept H_0 however, they do not show how the means differ, if, indeed, they do differ. In this section, we will discuss the multiple-comparison testing. We then present several procedures which can be used for multiple comparison of means, such as pair t test, Fisher's Least Significant Difference (LSD) test, Tukey's HSD test, the Newman-Keuls test, and Dunnett's test. Finally, we discuss the Scheffe test as a post hoc study for multiple comparisons.

Having learned that the device in which the AA-cell battery is used affects the battery's average lifetime, for example, we would likely want to know more details. Is it a case of all eight devices simply having different average battery lifetimes?

Pairwise Comparisons

Several multiple-comparison tests have as their basic procedure the comparison of all pairs of column means. Pairwise comparison tests are likely the most frequently used type of multiple-comparison tests.

For a variety of reasons, not necessarily identical for each test discussed, all of these tests should be used only when the original F-test has indicated the rejection of H_0. Indeed, one can reasonably argue that if the original F-test indicates that we cannot reject equality of all of the column means, what more is there to explore?

Pairwise tests of mean differences

The paired t-test is commonly used. It compares the means of two populations by testing if the difference between pairs is statistically different from zero.

Unadjusted Paired t tests

Example: A state securities laws was wondering whether five brokers have a difference average on stock price($hundreds). The Data **brokerstudy.csv** from a random sample of 30 observations from each of the 5 brokers are provided.

```
brokerstudy=read.csv("c:/Users/thuntida.ngamkham/OneDrive - University of Calgary/dataset603/brokerstudy.csv", header=TRUE)
str(brokerstudy)
```

```
## 'data.frame':    30 obs. of  2 variables:
##  $ broker: Factor w/ 5 levels "broker1","broker2",..: 1 1 1 1 1 1 2 2 2 2 ...
##  $ price : int  12 3 5 1 12 5 7 17 13 11 ...
```

```
CRD<-aov(price~broker, data=brokerstudy) #Perform ANOVA for CRD
summary(CRD)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## broker       4  622.1  155.53   7.695 0.000347 ***
## Residuals   25  505.3   20.21
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
mean(brokerstudy$price[brokerstudy$broker == "broker1"])
```

```
## [1] 6.333333
```

```
mean(brokerstudy$price[brokerstudy$broker == "broker2"])
```

```
## [1] 12
```

```
mean(brokerstudy$price[brokerstudy$broker == "broker3"])
```

```
## [1] 5
```

```
mean(brokerstudy$price[brokerstudy$broker == "broker4"])
```

```
## [1] 14
```

```
mean(brokerstudy$price[brokerstudy$broker == "broker5"])
```

```
## [1] 17
```

```
pairwise.t.test(brokerstudy$price,brokerstudy$broker, p.adj = "none")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  brokerstudy$price and brokerstudy$broker
##
##         broker1 broker2 broker3 broker4
## broker2 0.03863 -       -       -
## broker3 0.61200 0.01235 -       -
## broker4 0.00675 0.44823 0.00192 -
## broker5 0.00037 0.06552 9.9e-05 0.25871
##
## P value adjustment method: none
```

R function

pairwise.t.test(y,x, p.adj = “none”): is used to compare the means between two related groups of samples

Note!

p.adjust :Adjust P-values for Multiple Comparisons . it returns p-values adjusted using one of several methods.

Using p.adj = “none” in the pairwise.t.test() function makes no correction for the Type I error rate across the pairwise tests. The problem when you do a multiple comparisons you inflate this Type I error. For example, if you have 3 means (A, B and C) and you want

to make all possible piarwise comparisons i.e. A vs B, A vs C, and B vs C (i.e. 3 comparisons), then the total type I error will be 0.05*3 = 0.15. Therefore, some post-hoc tests come with a solution to this problem by making some adjustments for p-values.

adjust Pair t test

1. Bonferroni Adjustment

The Bonferroni adjustment simply divides the Type I error rate (.05) by the number of tests. For example, if we compare 3 tests, the p-value is tested at the .05/3 level. Hence, this method is often considered overly conservative. The Bonferroni adjustment can be made using p.adj = “bonferroni” in the pairwise.t.test() function.

```
brokerstudy=read.csv("c:/Users/thuntida.ngamkham/OneDrive - University of Calgary/dataset603/brokerstudy.csv", header=TRUE)
str(brokerstudy)
```

```
## 'data.frame':    30 obs. of  2 variables:
##  $ broker: Factor w/ 5 levels "broker1","broker2",..: 1 1 1 1 1 1 2 2 2 2 ...
##  $ price : int  12 3 5 1 12 5 7 17 13 11 ...
```

```
CRD<-aov(price~broker, data=brokerstudy) #Perform ANOVA for CRD
summary(CRD)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## broker       4  622.1  155.53   7.695 0.000347 ***
## Residuals   25  505.3   20.21
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
pairwise.t.test(brokerstudy$price,brokerstudy$broker, p.adj = "bonferroni")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  brokerstudy$price and brokerstudy$broker
##
##         broker1 broker2 broker3 broker4
## broker2 0.38630 -       -       -
## broker3 1.00000 0.12351 -       -
## broker4 0.06749 1.00000 0.01916 -
## broker5 0.00374 0.65517 0.00099 1.00000
##
## P value adjustment method: bonferroni
```

```
#pvalue (brok 1 vs brok 2)=0.03863* 10 number of comparisons=0.38630.
```

2. Holm Adjustment

The Holm adjustment sequentially compares the lowest p-value with a Type I error rate that is reduced for each consecutive test. For example, if we compare 3 tests, the ~~first~~ smallest p-value is tested at the .05/3 level (.017), the second smallest at the .05/2 level (.025), and third at the .05/1 level (.05). This method is generally considered superior to the Bonferroni adjustment and can be employed using p.adj = “holm” in the pairwise.t.test() function.

```
brokerstudy=read.csv("c:/Users/thuntida.ngamkham/OneDrive - University of Calgary/dataset6
03/brokerstudy.csv", header=TRUE)
str(brokerstudy)
```

```
## 'data.frame':    30 obs. of  2 variables:
##  $ broker: Factor w/ 5 levels "broker1","broker2",..: 1 1 1 1 1 1 2 2 2 2 ...
##  $ price : int  12 3 5 1 12 5 7 17 13 11 ...
```

```
CRD<-aov(price~broker, data=brokerstudy) #Perform ANOVA for CRD
summary(CRD)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## broker       4  622.1  155.53   7.695 0.000347 ***
## Residuals   25  505.3   20.21
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
pairwise.t.test(brokerstudy$price,brokerstudy$broker, p.adj = "holm")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  brokerstudy$price and brokerstudy$broker
##
##         broker1 broker2 broker3 broker4
## broker2 0.19315 -       -       -
## broker3 0.89645 0.07411 -       -
## broker4 0.04724 0.89645 0.01533 -
## broker5 0.00337 0.26207 0.00099 0.77613
##
## P value adjustment method: holm
```

```
#the smallest p-value (brok 3 vs brok 5) was 9.9e-05*10=0.00099
```

Inclass Practice Problem

From the MVPC experiment, compare the average level of satisfaction among customers in the four locations by using pairwise comparison t test.

```
MVPC=read.csv("c:/Users/thuntida.ngamkham/OneDrive - University of Calgary/dataset603/MVP
C.csv", header=TRUE)
str(MVPC)
```

```
## 'data.frame':    120 obs. of  2 variables:
##  $ Score    : int  66 66 66 67 70 64 71 66 71 67 ...
##  $ Treatment: Factor w/ 4 levels "Amesbury","Andover",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```
CRD<-aov(Score~Treatment, data=MVPC) #Perform ANOVA for CRD
pairwise.t.test(MVPC$Score,MVPC$Treatment, p.adj = "none")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  MVPC$Score and MVPC$Treatment
##
##         Amesbury Andover Methuen
## Andover < 2e-16  -       -
## Methuen < 2e-16  4.3e-16 -
## Salem   0.019    < 2e-16 < 2e-16
##
## P value adjustment method: none
```

```
pairwise.t.test(MVPC$Score,MVPC$Treatment, p.adj = "bonferroni")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  MVPC$Score and MVPC$Treatment
##
##         Amesbury Andover Methuen
## Andover < 2e-16  -       -
## Methuen < 2e-16  2.6e-15 -
## Salem   0.11     < 2e-16 < 2e-16
##
## P value adjustment method: bonferroni
```

```
pairwise.t.test(MVPC$Score,MVPC$Treatment, p.adj = "holm")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  MVPC$Score and MVPC$Treatment
##
##         Amesbury Andover Methuen
## Andover < 2e-16  -       -
## Methuen < 2e-16  8.5e-16 -
## Salem   0.019    < 2e-16 < 2e-16
##
## P value adjustment method: holm
```

Inclass Practice Problem

From the lifetime of AA battery experiment, compare the average life times for 8 devices by using pairwise comparison t test.

```
lifetime=read.csv("c:/Users/thuntida.ngamkham/OneDrive - University of Calgary/dataset603/lifetime.csv", header=TRUE)
str(lifetime)#Read your data set and double check that dependent and indepent variables are correctly read by R
```

```
## 'data.frame':    24 obs. of  2 variables:
##  $ device: Factor w/ 8 levels "device1","device2",..: 1 1 1 2 2 2 3 3 3 4 ...
##  $ hrs   : num  1.8 5 1 4.2 5.4 4.2 8.6 4.6 4.2 7 ...
```

```
CRD<-aov(hrs~device, data=lifetime) #Perform ANOVA for CRD
pairwise.t.test(lifetime$hrs,lifetime$device, p.adj = "none")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  lifetime$hrs and lifetime$device
##
##         device1 device2 device3 device4 device5 device6 device7
## device2 0.1710  -       -       -       -       -       -
## device3 0.0357  0.4025  -       -       -       -       -
## device4 0.0061  0.1047  0.4025  -       -       -       -
## device5 0.0201  0.2683  0.7780  0.5744  -       -       -
## device6 0.1710  1.0000  0.4025  0.1047  0.2683  -       -
## device7 0.0010  0.0201  0.1047  0.4025  0.1710  0.0201  -
## device8 0.0034  0.0620  0.2683  0.7780  0.4025  0.0620  0.5744
##
## P value adjustment method: none
```

```
pairwise.t.test(lifetime$hrs,lifetime$device, p.adj = "bonferroni")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  lifetime$hrs and lifetime$device
##
##         device1 device2 device3 device4 device5 device6 device7
## device2 1.000   -       -       -       -       -       -
## device3 1.000   1.000   -       -       -       -       -
## device4 0.172   1.000   1.000   -       -       -       -
## device5 0.564   1.000   1.000   1.000   -       -       -
## device6 1.000   1.000   1.000   1.000   1.000   -       -
## device7 0.028   0.564   1.000   1.000   1.000   0.564   -
## device8 0.094   1.000   1.000   1.000   1.000   1.000   1.000
##
## P value adjustment method: bonferroni
```

```
pairwise.t.test(lifetime$hrs,lifetime$device, p.adj = "holm")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  lifetime$hrs and lifetime$device
##
##         device1 device2 device3 device4 device5 device6 device7
## device2 1.000   -       -       -       -       -       -
## device3 0.785   1.000   -       -       -       -       -
## device4 0.160   1.000   1.000   -       -       -       -
## device5 0.503   1.000   1.000   1.000   -       -       -
## device6 1.000   1.000   1.000   1.000   1.000   -       -
## device7 0.028   0.503   1.000   1.000   1.000   0.503   -
## device8 0.091   1.000   1.000   1.000   1.000   1.000   1.000
##
## P value adjustment method: holm
```

Fisher's Least Significant Difference Test

This procedure was devised by R. A. Fisher and called Fisher's least significant difference (LSD) test, essentially involves performing a series of pairwise t-tests, each with a specified value of α.

Recall that in any hypothesis test, we establish an acceptance region for H_0 and accept H_0 if the appropriate test statistic falls within that region. If it falls in the critical region, we reject H_0. Here, for each pair of columns, i and j, the test statistic is the difference between the column means, $(\bar{y}_i - \bar{y}_j)$. If this difference is small, we conclude that the true means are equal (that is, their true difference is really zero, $\mu_i = \mu_j$, and any difference in the observed column means is just that due to statistical fluctuation). If the difference is not small, we conclude that the two levels of the factor being tested produce different true means

To find a confidence interval for the difference between two of the column means, we use

$$\bar{y}_i - \bar{y}_j \pm t_{\alpha/2,rc-c}\sqrt{MSE}\sqrt{\frac{1}{r_i}+\frac{1}{r_j}}$$

where

$t_{\alpha/2,rc-c}\sqrt{MSE}\sqrt{\frac{1}{r_i}+\frac{1}{r_j}}$ is called Fisher's least significant difference (LSD)

With the same number of data points, r, in each column, the LSD formula reduces to

$$LSD = t_{\alpha/2,rc-c}\sqrt{\frac{2MSE}{r}}$$

```
brokerstudy=read.csv("c:/Users/thuntida.ngamkham/OneDrive - University of Calgary/dataset603/brokerstudy.csv", header=TRUE)
str(brokerstudy)#Read your data set and double check that dependent and indepent variables are correctly read by R
```

```
## 'data.frame':    30 obs. of  2 variables:
##  $ broker: Factor w/ 5 levels "broker1","broker2",..: 1 1 1 1 1 1 2 2 2 2 ...
##  $ price : int  12 3 5 1 12 5 7 17 13 11 ...
```

```
CRD<-aov(price~broker, data=brokerstudy) #Perform ANOVA for CRD
summary(CRD)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## broker       4  622.1  155.53   7.695 0.000347 ***
## Residuals   25  505.3   20.21
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#example for constructing the difference in two means for broker1&3
ybar1<-mean(brokerstudy$price[brokerstudy$broker == "broker1"])
ybar3<-mean(brokerstudy$price[brokerstudy$broker == "broker3"])
tvalue<-qt(0.025,CRD$df.residual, lower.tail = F)
MSE<-sum((CRD$residuals)^2/CRD$df.residual)
r<-length(brokerstudy$price[brokerstudy$broker == "broker1"])
LSD<-tvalue*sqrt((2*MSE)/r)
LowerCI<-(ybar1-ybar3)-LSD#construct a 95% Lower CI
UpperCI<-(ybar1-ybar3)+LSD#construct a 95% Lower CI
CI<-cbind(LowerCI,UpperCI)
print(CI)
```

```
##        LowerCI  UpperCI
## [1,] -4.012658 6.679325
```

```
#example for constructing the difference in two means for broker1&2
ybar1<-mean(brokerstudy$price[brokerstudy$broker == "broker1"])
ybar2<-mean(brokerstudy$price[brokerstudy$broker == "broker2"])
tvalue<-qt(0.025,CRD$df.residual, lower.tail = F)
MSE<-sum((CRD$residuals)^2/CRD$df.residual)
r<-length(brokerstudy$price[brokerstudy$broker == "broker1"])
LSD<-tvalue*sqrt((2*MSE)/r)
LowerCI<-(ybar1-ybar2)-LSD#construct a 95% Lower CI
UpperCI<-(ybar1-ybar2)+LSD#construct a 95% Lower CI
CI<-cbind(LowerCI,UpperCI)
print(CI)
```

```
##        LowerCI    UpperCI
## [1,] -11.01266 -0.3206751
```

```
#example for constructing the difference in two means for broker2&5
ybar2<-mean(brokerstudy$price[brokerstudy$broker == "broker2"])
ybar5<-mean(brokerstudy$price[brokerstudy$broker == "broker5"])
tvalue<-qt(0.025,CRD$df.residual, lower.tail = F)
MSE<-sum((CRD$residuals)^2/CRD$df.residual)
r<-length(brokerstudy$price[brokerstudy$broker == "broker2"])
LSD<-tvalue*sqrt((2*MSE)/r)
LowerCI<-(ybar2-ybar5)-LSD#construct a 95% Lower CI
UpperCI<-(ybar2-ybar5)+LSD#construct a 95% Lower CI
CI<-cbind(LowerCI,UpperCI)
print(CI)
```

```
##        LowerCI   UpperCI
## [1,] -10.34599 0.3459915
```

We can use the r package *agricolae* to compute LSD with the function *LSD.test*.

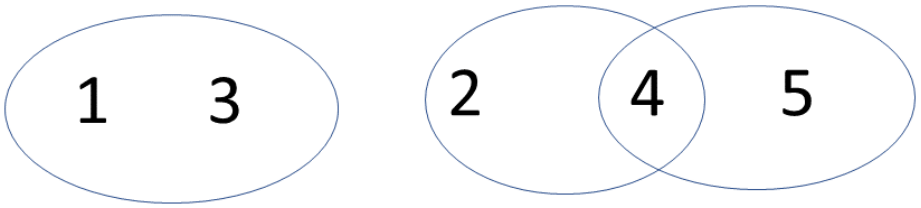
```
library(agricolae)
LS=LSD.test(CRD,trt="broker")
LS
```

```
## $statistics
##    MSerror Df     Mean       CV  t.value      LSD
##   20.21333 25 10.86667 41.37353 2.059539 5.345992
##
## $parameters
##         test p.ajusted name.t ntr alpha
##   Fisher-LSD      none broker   5  0.05
##
## $means
##             price      std r       LCL       UCL Min Max   Q25  Q50   Q75
## broker1  6.333333 4.633213 6  2.553146 10.113520   1  12  3.50  5.0 10.25
## broker2 12.000000 4.516636 6  8.219813 15.780187   7  17  8.00 12.0 16.00
## broker3  5.000000 2.756810 6  1.219813  8.780187   1   8  3.25  5.5  7.00
## broker4 14.000000 5.830952 6 10.219813 17.780187   6  21 10.50 13.5 18.75
## broker5 17.000000 4.195235 6 13.219813 20.780187  13  24 14.00 16.0 18.75
##
## $comparison
## NULL
##
## $groups
##             price groups
## broker5 17.000000      a
## broker4 14.000000      a
## broker2 12.000000      a
## broker1  6.333333      b
## broker3  5.000000      b
##
## attr(,"class")
## [1] "group"
```

```
#LSD.test(brokerstudy$price,brokerstudy$broker,DFerror=CRD$df.residual,MSerror=MSE)
```

conclusions: We would say by using LSD approach, brokers 1 and 3 are equivalent (their difference is not statistically significant, and, hence, we cannot reject their equality) with respect to buying price index, Y; similarly for brokers in the second subset

(brokers 2, 4, and 5). However, the two subsets of columns (brokers) are concluded to be different. We can show this diagrammatically



Tukey's Honestly Significant Difference Test

The honestly significant difference (HSD) test, devised by J. W. Tukey, focuses on the experimentwise error rate, a. It is a post hoc test used when **there are equal numbers of subjects contained in each group** for which pairwise comparisons of the data are being made. Post hoc tests, like this one, literally mean after the fact. They are used to determine whether any group or set of treatment conditions significantly differs from one or more others. The Tukey HSD test is more likely to identify statistically significant differences than other post hoc tests. This entry discusses the utility of the Tukey HSD post hoc test, gives a thorough developmental overview, and then provides further elaboration.

To construct a confidence interval for the difference between two of the column means, we use

$$\bar{y}_i - \bar{y}_j \pm q(c, df)_{\alpha/2}\sqrt{\frac{MSE}{r}}$$

where

$$HSD = q(c, df)_{\alpha/2}\sqrt{\frac{MSE}{r}}$$

Error *df*	a	Number of treatment means 2	3	4	5	6	7	8	9	10	11
5	.05	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17
	.01	5.70	6.98	7.80	8.42	8.91	9.32	9.67	9.97	10.24	10.48
6	.05	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49	6.65
	.01	5.24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10	9.30
7	.05	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16	6.30
	.01	4.95	5.92	6.54	7.01	7.37	7.68	7.94	8.17	8.37	8.55
8	.05	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05
	.01	4.75	5.64	6.20	6.62	6.96	7.24	7.47	7.68	7.86	8.03
9	.05	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74	5.87
	.01	4.60	5.43	5.96	6.35	6.66	6.91	7.13	7.33	7.49	7.65
10	.05	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.72
	.01	4.48	5.27	5.77	6.14	6.43	6.67	6.87	7.05	7.21	7.36
11	.05	3.11	3.82	4.26	4.57	4.82	5.03	5.30	5.35	5.49	5.61
	.01	4.39	5.15	5.62	5.97	6.25	6.48	6.67	6.84	6.99	7.13
12	.05	3.08	3.77	4.20	4.52	4.75	4.95	5.12	5.27	5.39	5.51
	.01	4.32	5.05	5.50	5.84	6.10	6.32	6.51	6.67	6.81	6.94
13	.05	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32	5.43
	.01	4.26	4.96	5.40	5.73	5.98	6.19	6.37	6.53	6.67	6.79
14	.05	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36
	.01	4.21	4.89	5.32	5.63	5.88	6.08	6.26	6.41	6.54	6.66
15	.05	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.31
	.01	4.17	4.84	5.25	5.56	5.80	5.99	6.16	6.31	6.44	6.55
16	.05	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15	5.26
	.01	4.13	4.79	5.19	5.49	5.72	5.92	6.08	6.22	6.35	6.46
17	.05	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11	5.21
	.01	4.10	4.74	5.14	5.43	5.66	5.85	6.01	6.15	6.27	6.38
18	.05	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07	5.17
	.01	4.07	4.70	5.09	5.38	5.60	5.79	5.94	6.08	6.20	6.31
19	.05	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04	5.14
	.01	4.05	4.67	5.05	5.33	5.55	5.73	5.89	6.02	6.14	6.25
20	.05	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01	5.11
	.01	4.02	4.64	5.02	5.29	5.51	5.69	5.84	5.97	6.09	6.19
24	.05	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	3.92	5.01
	.01	3.96	4.55	4.91	5.17	5.37	5.54	5.69	5.81	5.92	6.02
30	.05	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82	4.92
	.01	3.89	4.45	4.80	5.05	5.24	5.40	5.54	5.65	5.76	5.85
40	.05	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73	4.82
	.01	3.82	4.37	4.70	4.93	5.11	5.26	5.39	5.50	5.60	5.69
60	.05	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	4.73
	.01	3.76	4.28	4.59	4.82	4.99	5.13	5.25	5.36	5.45	5.53
120	.05	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56	4.64
	.01	3.70	4.20	4.50	4.71	4.87	5.01	5.12	5.21	5.30	5.37
α	.05	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47	4.55
	.01	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16	5.23

Source: E. S. Pearson and H. O. Hartley (1966), *Biometrika Tables for Statisticians*, vol. 1, 3rd ed (Reprinted with permission of Oxford University Press)

Figure

```
brokerstudy=read.csv("c:/Users/thuntida.ngamkham/OneDrive - University of Calgary/dataset603/brokerstudy.csv", header=TRUE)
str(brokerstudy)#Read your data set and double check that dependent and indepent variables are correctly read by R
```

```
## 'data.frame':    30 obs. of  2 variables:
##  $ broker: Factor w/ 5 levels "broker1","broker2",..: 1 1 1 1 1 1 2 2 2 2 ...
##  $ price : int  12 3 5 1 12 5 7 17 13 11 ...
```

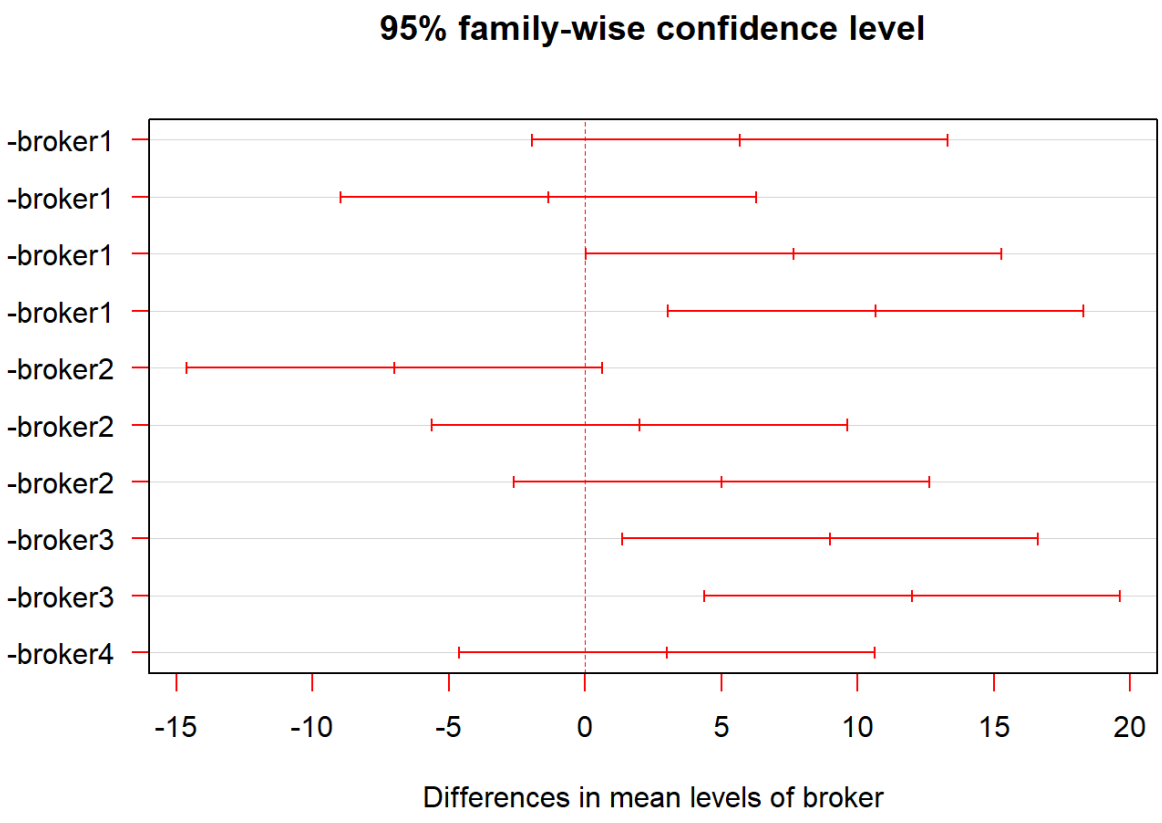
```
CRD<-aov(price~broker, data=brokerstudy) #Perform ANOVA for CRD
summary(CRD)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## broker       4  622.1  155.53   7.695 0.000347 ***
## Residuals   25  505.3   20.21
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(CRD, conf.level = 0.95)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = price ~ broker, data = brokerstudy)
##
## $broker
##                       diff           lwr        upr     p adj
## broker2-broker1  5.666667  -1.95663742 13.2899707 0.2186942
## broker3-broker1 -1.333333  -8.95663742  6.2899707 0.9851976
## broker4-broker1  7.666667   0.04336258 15.2899707 0.0482197
## broker5-broker1 10.666667   3.04336258 18.2899707 0.0031451
## broker3-broker2 -7.000000 -14.62330408  0.6233041 0.0829559
## broker4-broker2  2.000000  -5.62330408  9.6233041 0.9367210
## broker5-broker2  5.000000  -2.62330408 12.6233041 0.3301346
## broker4-broker3  9.000000   1.37669592 16.6233041 0.0149948
## broker5-broker3 12.000000   4.37669592 19.6233041 0.0008643
## broker5-broker4  3.000000  -4.62330408 10.6233041 0.7756233
```

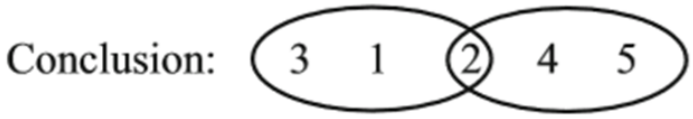
```
plot(TukeyHSD(CRD, conf.level = 0.95),las=1, col = "red")
```



Thus, we need to check each of the 10 differences in column means, as shown in the output. What does the story tell? If we ignore the column two mean for the moment, we conclude that column means three and one are the same, but different from (smaller than) column means four and five, which are themselves the same; that is,

Figure

However, column mean two is “the same as column means three and one,” but also “the same as column means four and five.” (As pointed out in footnote 5, this type of phrasing is more concise than it is elegant.) We have an inconsistency of the type expressed earlier. In a consulting capacity, we would tell a client that “column means three and one cannot be said to be different, column means four and five cannot be said to be different, but the former two can be said to be different from (smaller than) the latter two; we cannot determine the role of column mean two.” We would express this thought diagrammatically as



Figure

It should be noted that one potential reason for obtaining different results from those of the Fisher's LSD analysis is that the per-comparison error rates are very different. In the immediately preceding Tukey's HSD analysis, with 10 comparisons and alpha= .05, each comparison error rate is,relatively speaking, very small; if they were independent comparisons, which they're not, each would be only .005. The per comparison error rate for the Fisher's LSD analysis was α=.05, about ten times larger.

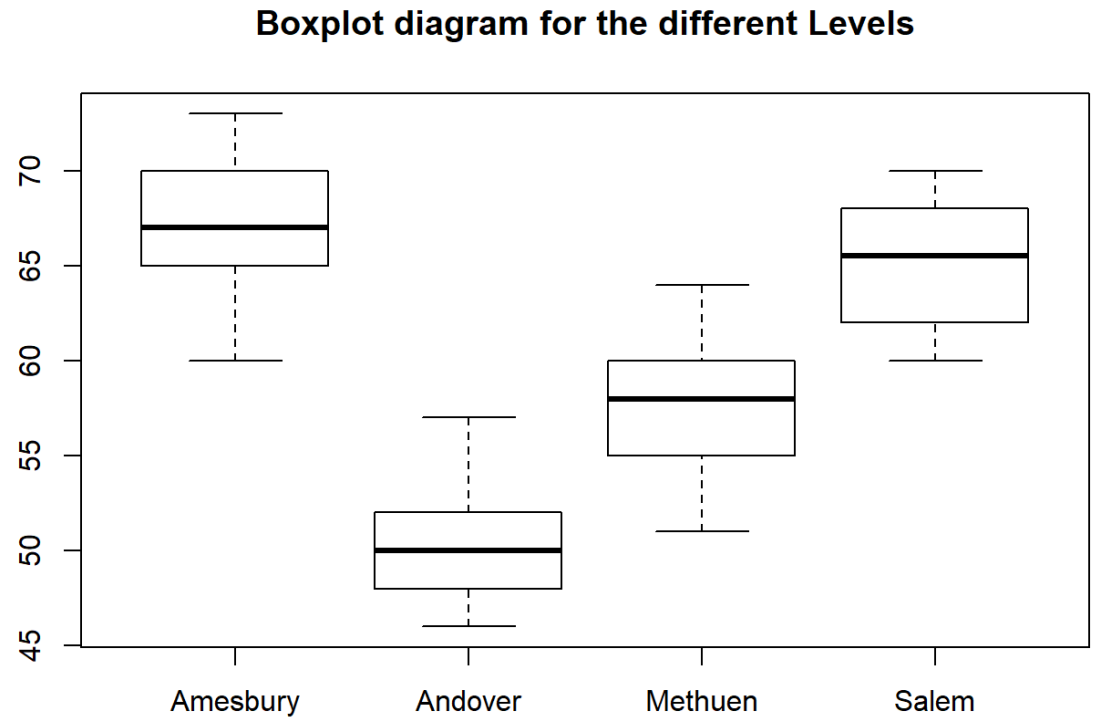
Inclass Practice Problem

From the MVPC experiment, compare the average level of satisfaction among customers in the four locations by Tukey's HSD. Compare your results with pair t test.

```
library(agricolae)
MVPC=read.csv("c:/Users/thuntida.ngamkham/OneDrive - University of Calgary/dataset603/MVP
C.csv", header=TRUE)
str(MVPC)#Read your data set and double check that dependent and indepent variables are co
rrectly read by R
```

```
## 'data.frame':    120 obs. of  2 variables:
##  $ Score    : int  66 66 66 67 70 64 71 66 71 67 ...
##  $ Treatment: Factor w/ 4 levels "Amesbury","Andover",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```
CRD<-aov(Score~Treatment, data=MVPC) #Perform ANOVA for CRD
boxplot(Score~Treatment, data=MVPC, main="Boxplot diagram for the different Levels") #a visual comparison of the data obtained at the different levels
```



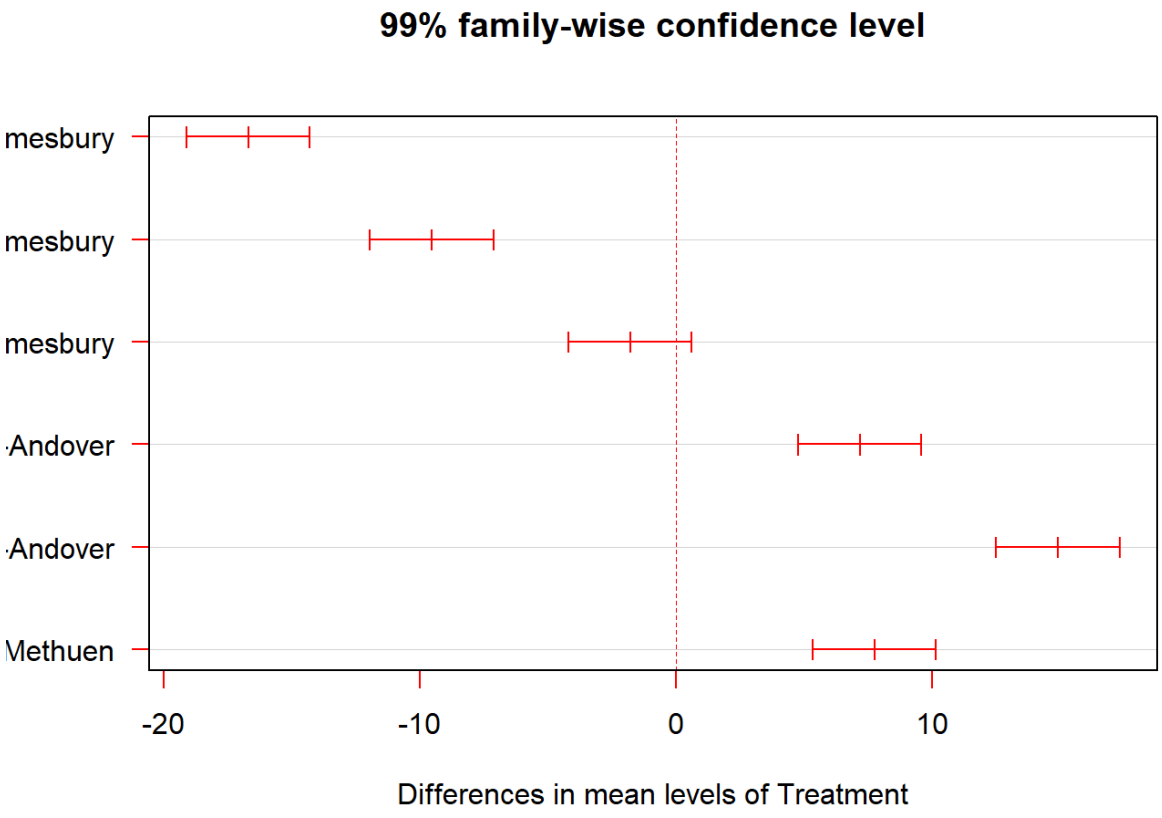

```
summary(CRD)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Treatment     3   5296  1765.5   205.3 <2e-16 ***
## Residuals   116    998     8.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(CRD, conf.level = 0.99)
```

```
##   Tukey multiple comparisons of means
##     99% family-wise confidence level
##
## Fit: aov(formula = Score ~ Treatment, data = MVPC)
##
## $Treatment
##                          diff        lwr         upr     p adj
## Andover-Amesbury -16.700000 -19.109498 -14.2905018 0.0000000
## Methuen-Amesbury  -9.533333 -11.942832  -7.1238351 0.0000000
## Salem-Amesbury    -1.800000  -4.209498   0.6094982 0.0872085
## Methuen-Andover    7.166667   4.757168   9.5761649 0.0000000
## Salem-Andover     14.900000  12.490502  17.3094982 0.0000000
## Salem-Methuen      7.733333   5.323835  10.1428315 0.0000000
```

```
plot(TukeyHSD(CRD, conf.level = 0.99),las=1, col = "red")
```



Inclass Practice Problem

From the lifetime of AA battery experiment, compare the average life times for 8 devices by Tukey's HSD. Compare your result with pair t test

```
lifetime=read.csv("c:/Users/thuntida.ngamkham/OneDrive - University of Calgary/dataset603/
lifetime.csv", header=TRUE)
str(lifetime)#Read your data set and double check that dependent and indepent variables ar
e correctly read by R
```

```
## 'data.frame':    24 obs. of  2 variables:
##  $ device: Factor w/ 8 levels "device1","device2",..: 1 1 1 2 2 2 3 3 3 4 ...
##  $ hrs   : num  1.8 5 1 4.2 5.4 4.2 8.6 4.6 4.2 7 ...
```

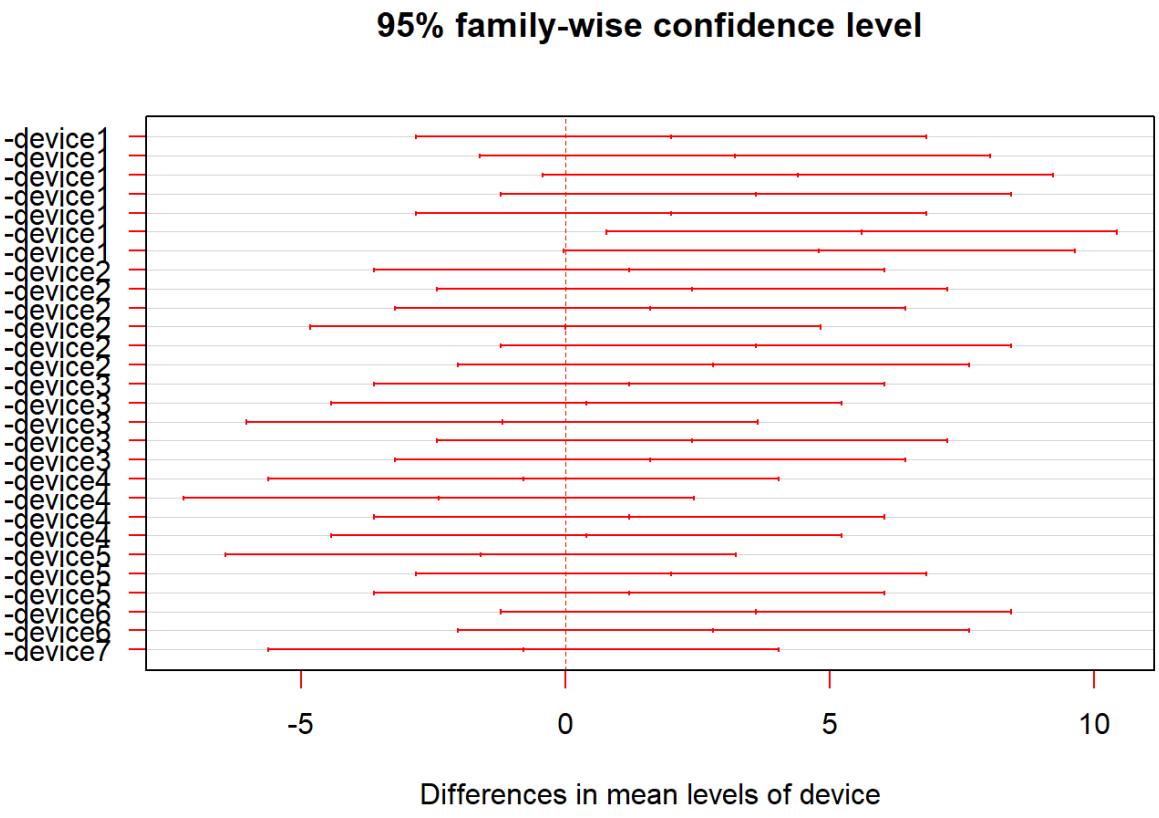
```
CRD<-aov(hrs~device, data=lifetime) #Perform ANOVA for CRD
summary(CRD)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## device       7  69.12   9.874   3.382 0.0206 *
## Residuals   16  46.72   2.920
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(CRD, conf.level = 0.95)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = hrs ~ device, data = lifetime)
##
## $device
##                           diff        lwr       upr     p adj
## device2-device1  2.000000e+00 -2.8304964  6.830496 0.8293865
## device3-device1  3.200000e+00 -1.6304964  8.030496 0.3532560
## device4-device1  4.400000e+00 -0.4304964  9.230496 0.0881336
## device5-device1  3.600000e+00 -1.2304964  8.430496 0.2320154
## device6-device1  2.000000e+00 -2.8304964  6.830496 0.8293865
## device7-device1  5.600000e+00  0.7695036 10.430496 0.0174163
## device8-device1  4.800000e+00 -0.0304964  9.630496 0.0520860
## device3-device2  1.200000e+00 -3.6304964  6.030496 0.9858452
## device4-device2  2.400000e+00 -2.4304964  7.230496 0.6758100
## device5-device2  1.600000e+00 -3.2304964  6.430496 0.9358676
## device6-device2 -1.776357e-15 -4.8304964  4.830496 1.0000000
## device7-device2  3.600000e+00 -1.2304964  8.430496 0.2320154
## device8-device2  2.800000e+00 -2.0304964  7.630496 0.5067527
## device4-device3  1.200000e+00 -3.6304964  6.030496 0.9858452
## device5-device3  4.000000e-01 -4.4304964  5.230496 0.9999877
## device6-device3 -1.200000e+00 -6.0304964  3.630496 0.9858452
## device7-device3  2.400000e+00 -2.4304964  7.230496 0.6758100
## device8-device3  1.600000e+00 -3.2304964  6.430496 0.9358676
## device5-device4 -8.000000e-01 -5.6304964  4.030496 0.9987759
## device6-device4 -2.400000e+00 -7.2304964  2.430496 0.6758100
## device7-device4  1.200000e+00 -3.6304964  6.030496 0.9858452
## device8-device4  4.000000e-01 -4.4304964  5.230496 0.9999877
## device6-device5 -1.600000e+00 -6.4304964  3.230496 0.9358676
## device7-device5  2.000000e+00 -2.8304964  6.830496 0.8293865
## device8-device5  1.200000e+00 -3.6304964  6.030496 0.9858452
## device7-device6  3.600000e+00 -1.2304964  8.430496 0.2320154
## device8-device6  2.800000e+00 -2.0304964  7.630496 0.5067527
## device8-device7 -8.000000e-01 -5.6304964  4.030496 0.9987759
```

```
plot(TukeyHSD(CRD, conf.level = 0.95),las=1, col = "red")
```



Newman-Keuls Test

The Newman-Keuls and Tukey HSD work with different distributions (Newman-Keuls with the studentized range) but that doesn't necessarily mean that one is better than the other. In fact, there's no consensus on how to choose between the two. Although the N-K test was designed to have more power than Tukey's HSD, the probability of making a Type I error can't be calculated for the N-K test, nor is it possible to calculate confidence intervals around difference between means.

The Newman-Keuls test is an alternative to Tukey's HSD test. It is similar to the HSD test

```
library(agricolae)# SNK.test() is avalible in the agricolae package for Newman-Keuls (SNK)
brokerstudy=read.csv("c:/Users/thuntida.ngamkham/OneDrive - University of Calgary/dataset603/brokerstudy.csv", header=TRUE)
print(brokerstudy)
```

```
##     broker price
## 1  broker1    12
## 2  broker1     3
## 3  broker1     5
## 4  broker1     1
## 5  broker1    12
## 6  broker1     5
## 7  broker2     7
## 8  broker2    17
## 9  broker2    13
## 10 broker2    11
## 11 broker2     7
## 12 broker2    17
## 13 broker3     8
## 14 broker3     1
## 15 broker3     7
## 16 broker3     4
## 17 broker3     3
## 18 broker3     7
## 19 broker4    21
## 20 broker4    10
## 21 broker4    15
## 22 broker4    12
## 23 broker4    20
## 24 broker4     6
## 25 broker5    24
## 26 broker5    13
## 27 broker5    14
## 28 broker5    18
## 29 broker5    14
## 30 broker5    19
```

```
str(brokerstudy)#Read your data set and double check that dependent and indepent variables
are correctly read by R
```

```
## 'data.frame':    30 obs. of  2 variables:
##  $ broker: Factor w/ 5 levels "broker1","broker2",..: 1 1 1 1 1 1 2 2 2 2 ...
##  $ price : int  12 3 5 1 12 5 7 17 13 11 ...
```

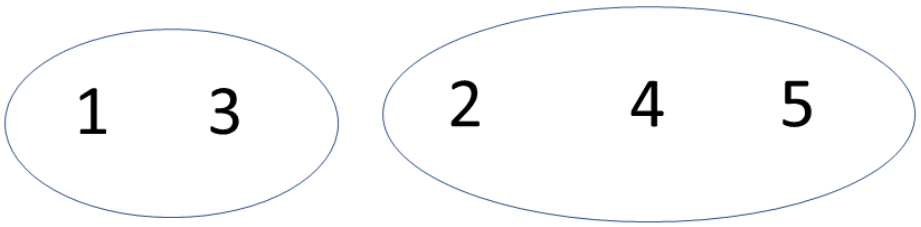
```
CRD<-aov(price~broker, data=brokerstudy) #Perform ANOVA for CRD
summary(CRD)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## broker       4  622.1  155.53   7.695 0.000347 ***
## Residuals   25  505.3   20.21
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
print(SNK.test(CRD,"broker",group=TRUE))#SNK.test() function can be used for Newman-Keuls
 test in R,
```

```
## $statistics
##    MSerror Df     Mean       CV
##   20.21333 25 10.86667 41.37353
##
## $parameters
##   test name.t ntr alpha
##    SNK broker   5  0.05
##
## $snk
##      Table CriticalRange
## 2 2.912627      5.345992
## 3 3.522566      6.465505
## 4 3.889997      7.139908
## 5 4.153363      7.623304
##
## $means
##             price      std r Min Max   Q25  Q50   Q75
## broker1  6.333333 4.633213 6   1  12  3.50  5.0 10.25
## broker2 12.000000 4.516636 6   7  17  8.00 12.0 16.00
## broker3  5.000000 2.756810 6   1   8  3.25  5.5  7.00
## broker4 14.000000 5.830952 6   6  21 10.50 13.5 18.75
## broker5 17.000000 4.195235 6  13  24 14.00 16.0 18.75
##
## $comparison
## NULL
##
## $groups
##             price groups
## broker5 17.000000      a
## broker4 14.000000      a
## broker2 12.000000      a
## broker1  6.333333      b
## broker3  5.000000      b
##
## attr(,"class")
## [1] "group"
```

We observe that the Newman-Keuls (N-K) test finds two results that are different from those of the HSD test: column means three and two, as well as column means one and two, are now concluded to be different, whereas the HSD test did not conclude that they were different.



Figure

The Scheffe Test

The Scheffe Test (also called Scheffe's procedure or Scheffe's method) is a post-hoc test used in Analysis of Variance. It is named for the American statistician Henry Scheffe. After you have run ANOVA and got a significant F-statistic (i.e. you have rejected the null hypothesis that the means are the same), then you run Sheffe's test to find out which pairs of means are significant.

For pair-wise comparisons, Scheffe's can be computed as follows:

$$S = \sqrt{(k-1)F_{\alpha,df_1,df_2}}$$

$$\bar{y_i} - \bar{y_j} \pm S\sqrt{MSE(\frac{1}{n_i} + \frac{1}{n_j})}$$

In this class, we use R software to calculate all multiple comparison.

```
library(agricolae)
brokerstudy=read.csv("c:/Users/thuntida.ngamkham/OneDrive - University of Calgary/dataset603/brokerstudy.csv", header=TRUE)
str(brokerstudy)#Read your data set and double check that dependent and indepent variables are correctly read by R
```

```
## 'data.frame':    30 obs. of  2 variables:
##  $ broker: Factor w/ 5 levels "broker1","broker2",..: 1 1 1 1 1 1 2 2 2 2 ...
##  $ price : int  12 3 5 1 12 5 7 17 13 11 ...
```

```
CRD<-aov(price~broker, data=brokerstudy) #Perform ANOVA for CRD
scheffe.test(CRD,"broker", group=TRUE,console=TRUE)
```

```
##
## Study: CRD ~ "broker"
##
## Scheffe Test for price
##
## Mean Square Error  : 20.21333
##
## broker,  means
##
##             price      std r Min Max
## broker1  6.333333 4.633213 6   1  12
## broker2 12.000000 4.516636 6   7  17
## broker3  5.000000 2.756810 6   1   8
## broker4 14.000000 5.830952 6   6  21
## broker5 17.000000 4.195235 6  13  24
##
## Alpha: 0.05 ; DF Error: 25
## Critical Value of F: 2.75871
##
## Minimum Significant Difference: 8.622663
##
## Means with the same letter are not significantly different.
##
##             price groups
## broker5 17.000000      a
## broker4 14.000000     ab
## broker2 12.000000    abc
## broker1  6.333333     bc
## broker3  5.000000      c
```

```
library(agricolae)
MVPC=read.csv("c:/Users/thuntida.ngamkham/OneDrive - University of Calgary/dataset603/MVPC.csv", header=TRUE)
str(MVPC)#Read your data set and double check that dependent and indepent variables are correctly read by R
```

```
## 'data.frame':    120 obs. of  2 variables:
##  $ Score    : int  66 66 66 67 70 64 71 66 71 67 ...
##  $ Treatment: Factor w/ 4 levels "Amesbury","Andover",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```
CRD<-aov(Score~Treatment, data=MVPC) #Perform ANOVA for CRD
scheffe.test(CRD,"Treatment", group=TRUE,console=TRUE)
```

```
##
## Study: CRD ~ "Treatment"
##
## Scheffe Test for Score
##
## Mean Square Error  : 8.599713
##
## Treatment,  means
##
##             Score      std  r Min Max
## Amesbury 67.10000 2.916658 30  60  73
## Andover  50.40000 2.835733 30  46  57
## Methuen  57.56667 3.013571 30  51  64
## Salem    65.30000 2.961244 30  60  70
##
## Alpha: 0.05 ; DF Error: 116
## Critical Value of F: 2.682809
##
## Minimum Significant Difference: 2.148087
##
## Means with the same letter are not significantly different.
##
##             Score groups
## Amesbury 67.10000      a
## Salem    65.30000      a
## Methuen  57.56667      b
## Andover  50.40000      c
```

```
lifetime=read.csv("c:/Users/thuntida.ngamkham/OneDrive - University of Calgary/dataset603/
lifetime.csv", header=TRUE)
str(lifetime)#Read your data set and double check that dependent and indepent variables ar
e correctly read by R
```

```
## 'data.frame':    24 obs. of  2 variables:
##  $ device: Factor w/ 8 levels "device1","device2",..: 1 1 1 2 2 2 3 3 3 4 ...
##  $ hrs   : num  1.8 5 1 4.2 5.4 4.2 8.6 4.6 4.2 7 ...
```

```
CRD<-aov(hrs~device, data=lifetime) #Perform ANOVA for CRD
summary(CRD)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## device       7  69.12   9.874   3.382 0.0206 *
## Residuals   16  46.72   2.920
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
scheffe.test(CRD,"device", group=TRUE,console=TRUE)
```

```
##
## Study: CRD ~ "device"
##
## Scheffe Test for hrs
##
## Mean Square Error  : 2.92
##
## device,  means
##
##         hrs       std r Min Max
## device1 2.6 2.1166010 3 1.0 5.0
## device2 4.6 0.6928203 3 4.2 5.4
## device3 5.8 2.4331050 3 4.2 8.6
## device4 7.0 2.0000000 3 5.0 9.0
## device5 6.2 1.8330303 3 4.2 7.8
## device6 4.6 0.6928203 3 4.2 5.4
## device7 8.2 1.4422205 3 7.0 9.8
## device8 7.4 1.6000000 3 5.8 9.0
##
## Alpha: 0.05 ; DF Error: 16
## Critical Value of F: 2.657197
##
## Minimum Significant Difference: 6.017369
##
## Means with the same letter are not significantly different.
##
##         hrs groups
## device7 8.2      a
## device8 7.4      a
## device4 7.0      a
## device5 6.2      a
## device3 5.8      a
## device2 4.6      a
## device6 4.6      a
## device1 2.6      a
```

That is, column means three and one are the same, and are different from (smaller than) column means two,four,and five,the latter three column means being the same.

Inclass Practice Problem

From the MVPC experiment, compare the average level of satisfaction among customers in the four locations by Newman-Keuls Test.Compare your result with pair t test and Tukey's HSD.

```
library(agricolae)
MVPC=read.csv("c:/Users/thuntida.ngamkham/OneDrive - University of Calgary/dataset603/MVPC.csv", header=TRUE)
str(MVPC)#Read your data set and double check that dependent and indepent variables are correctly read by R
```

```
## 'data.frame':    120 obs. of  2 variables:
##  $ Score    : int  66 66 66 67 70 64 71 66 71 67 ...
##  $ Treatment: Factor w/ 4 levels "Amesbury","Andover",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```
CRD<-aov(Score~Treatment, data=MVPC) #Perform ANOVA for CRD
print(SNK.test(CRD,"Treatment",alpha=0.01, group=TRUE))#SNK.test() function can be used for Newman-Keuls test in R,
```

```
## $statistics
##    MSerror  Df     Mean       CV
##   8.599713 116 60.09167 4.880089
##
## $parameters
##   test    name.t ntr alpha
##    SNK Treatment   4  0.01
##
## $snk
##      Table CriticalRange
## 2 3.703652      1.982949
## 3 4.202736      2.250160
## 4 4.500339      2.409498
##
## $means
##             Score      std  r Min Max   Q25  Q50   Q75
## Amesbury 67.10000 2.916658 30  60  73 65.25 67.0 69.75
## Andover  50.40000 2.835733 30  46  57 48.00 50.0 52.00
## Methuen  57.56667 3.013571 30  51  64 55.25 58.0 59.75
## Salem    65.30000 2.961244 30  60  70 62.25 65.5 68.00
##
## $comparison
## NULL
##
## $groups
##             Score groups
## Amesbury 67.10000      a
## Salem    65.30000      a
## Methuen  57.56667      b
## Andover  50.40000      c
##
## attr(,"class")
## [1] "group"
```

Inclass Practice Problem

From the lifetime of AA battery experiment, compare the average life times for 8 devices by Tukey's HSD. Compare your result with pair t test and Tukey's HSD.

```
lifetime=read.csv("c:/Users/thuntida.ngamkham/OneDrive - University of Calgary/dataset603/
lifetime.csv", header=TRUE)
str(lifetime)#Read your data set and double check that dependent and indepent variables ar
e correctly read by R
```

```
## 'data.frame':    24 obs. of  2 variables:
##  $ device: Factor w/ 8 levels "device1","device2",..: 1 1 1 2 2 2 3 3 3 4 ...
##  $ hrs   : num  1.8 5 1 4.2 5.4 4.2 8.6 4.6 4.2 7 ...
```

```
CRD<-aov(hrs~device, data=lifetime) #Perform ANOVA for CRD
print(SNK.test(CRD,"device",alpha=0.05, group=TRUE))#SNK.test() function can be used for N
ewman-Keuls test in R,
```

```
## $statistics
##   MSerror Df Mean       CV
##      2.92 16  5.8 29.46208
##
## $parameters
##   test name.t ntr alpha
##    SNK device   8  0.05
##
## $snk
##      Table CriticalRange
## 2 2.997999      2.957755
## 3 3.649139      3.600155
## 4 4.046093      3.991781
## 5 4.332688      4.274528
## 6 4.556809      4.495641
## 7 4.740611      4.676976
## 8 4.896220      4.830496
##
## $means
##         hrs       std r Min Max Q25 Q50 Q75
## device1 2.6 2.1166010 3 1.0 5.0 1.4 1.8 3.4
## device2 4.6 0.6928203 3 4.2 5.4 4.2 4.2 4.8
## device3 5.8 2.4331050 3 4.2 8.6 4.4 4.6 6.6
## device4 7.0 2.0000000 3 5.0 9.0 6.0 7.0 8.0
## device5 6.2 1.8330303 3 4.2 7.8 5.4 6.6 7.2
## device6 4.6 0.6928203 3 4.2 5.4 4.2 4.2 4.8
## device7 8.2 1.4422205 3 7.0 9.8 7.4 7.8 8.8
## device8 7.4 1.6000000 3 5.8 9.0 6.6 7.4 8.2
##
## $comparison
## NULL
##
## $groups
##         hrs groups
## device7 8.2      a
## device8 7.4      a
## device4 7.0     ab
## device5 6.2     ab
## device3 5.8     ab
## device2 4.6     ab
## device6 4.6     ab
## device1 2.6      b
##
## attr(,"class")
## [1] "group"
```