

STAT 523 HW 4

Md Muhtasim Billah

2/13/2020

Problem 1

From the textbook- Probability and Statistics for Engineering and Sciences by Jay L. Devore, Chapter 10, Section 10.2, Exercise 17, the contrast is given as, $\theta = 1/2(\mu_1 + \mu_2) - \mu_3$. Compute a 2-sided CI for this contrast. Use $MSE = 0.6603$, $I = 3$, $J = 10$ with 95% confidence.

Answer

Given,

Number of treatments, $I = 3$.

Number of replicates, $J = 10$.

Level of significance, $\alpha = 0.05$.

$MSE = 0.6603$.

The contrast,

$$\theta = C = 1/2(\mu_1 + \mu_2) - \mu_3$$

Since, contrasts that differ by a scalar (constant) factor are equivalent, we can rewrite the contrast as,

$$C = 2 \times C = 2 \times [1/2(\mu_1 + \mu_2) - \mu_3]$$

$$C = \mu_1 + \mu_2 - 2\mu_3$$

Since, we don't know the population (true) means μ_1, μ_2, μ_3 of the treatments, we replace them with the sample means $\bar{x}_1, \bar{x}_2, \bar{x}_3$, respectively. The sample means (from the same textbook, chapter 10, section 10.1, exercise 5, for the modulus of elasticity for lumber of three different grades) are as below.

Grade	J	\bar{x}_i
1	10	1.63
2	10	1.56
3	10	1.42

So, an unbiased estimate of the contrast, C becomes,

$$\hat{C} = \sum_{i=1}^I c_i \bar{x}_i$$

$$\hat{C} = \bar{x}_1 + \bar{x}_2 - 2\bar{x}_3$$

Where,

$$\sum_{i=1}^I c_i = 1 + 1 - 2 = 0$$

Plugging in the values, we get,

$$\hat{C} = 1.63 + 1.56 - 2 \times 1.42$$

$$\hat{C} = 0.35 > 0$$

Now, from Formula Packet 10G, the 2-sided CI for contrast C is,

$$\hat{C} \pm t_{\alpha/2, I(J-1)} \sqrt{\frac{MSE \sum_{i=1}^I c_i^2}{J}}$$

Here,

$$\sum_{i=1}^I c_i^2 = (1)^2 + (1)^2 + (-2)^2 = 6$$

Plugging in all the values,

$$0.35 \pm t_{0.05/2, 3(10-1)} \sqrt{\frac{0.6603 \times 6}{10}}$$

Here,

$$t_{0.05/2, 3(10-1)} = t_{0.025, 3(10-1)} = 2.0518$$

```
qt(1-0.025,27)
```

```
## [1] 2.051831
```

Finally,

$$0.35 \pm 2.0518 \sqrt{\frac{0.6603 \times 6}{10}}$$

$$0.35 \pm 1.29146$$

So, the two-sided confidence interval for the given contrast is,

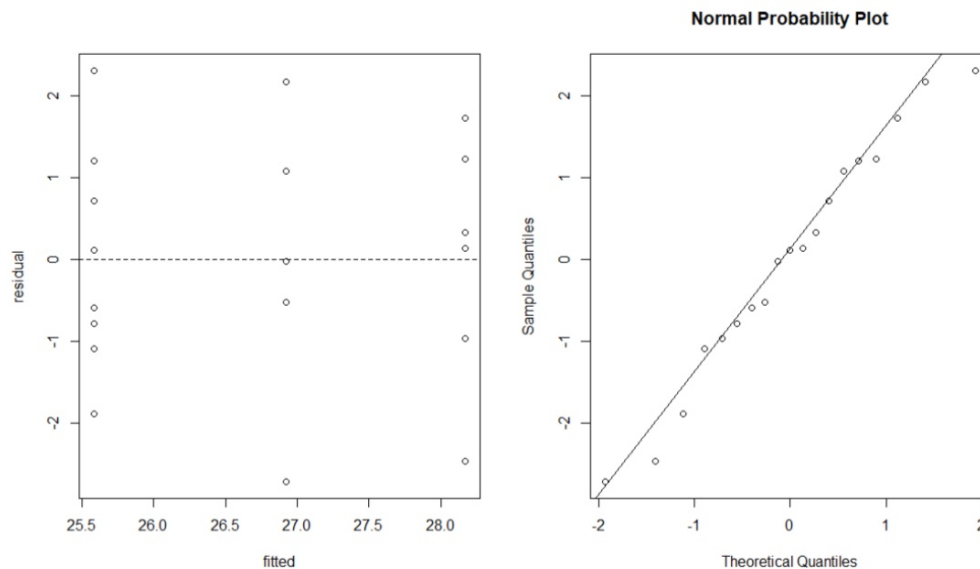
$$(-0.94146, 1.64146)$$

Since, 0 is within the CI, it can be said that the mean of grade 3 is significantly different from the mean of grade 1 and 2.

Problem 2

Consider the data in Exercise 42 in Chapter 10 Supplementary Exercises. Carry out the instructions below. Use software if possible.

- Obtain the ANOVA table for testing the hypothesis that mean CFF values are all the same.
- State/give the relevant hypotheses at $\alpha = 0.05$. Follow the 4-step hypothesis test outline. Give the reject region or the P-value.
- Do the following residual plots suggest strong evidence that the model assumptions are violated? Explain briefly.



- Perform the Tukey method of underscoring (95% confidence) to determine which means differ. Write a few sentences to summarize your findings.

Answer

a.

ANOVA table in R for testing the hypothesis that mean CFF values are all the same. Here, the treatments have unequal sample sizes.

```
#building the data frame
hw4_2=data.frame(Color=c(rep("Brown",8),rep("Green",5),rep("Blue",6)),
  CFF=c(26.8, 27.9, 23.7, 25.0, 26.3, 24.8, 25.7, 24.5,
    26.4, 24.2, 28.0, 26.9, 29.1,
    25.7, 27.2, 29.9, 28.5, 29.4, 28.3))
#carrying out ANOVA
hw4_2out=aov(CFF~Color,hw4_2)
#ANOVA table
summary(hw4_2out)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Color      2  23.00  11.499   4.802 0.0232 *
## Residuals  16  38.31   2.394
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

b.

Now, since we have the F-statistic from the ANOVA table generated in R, we can run our 4-step hypothesis testing outline.

Step1:

Null hypothesis, $H_0 : \mu_1 = \mu_2 = \mu_3$.

Alternative hypothesis, $H_a : H_0$ is false.

Level of significance, $\alpha = 0.05$

Step2:

From the ANOVA table, F-statistic,

$$F = 4.802$$

Step3:

Using built in function in R, we find (here, n=19) the p-value for unequal sample size.

$$p - value = 0.02325398$$

```
#1-pf(F,I-1,n-I)
1-pf(4.802,3-1,19-3)
```

```
## [1] 0.02325398
```

Step4:

It is evident that,

$$0.02325398 < 0.05$$

$$p - value < \alpha$$

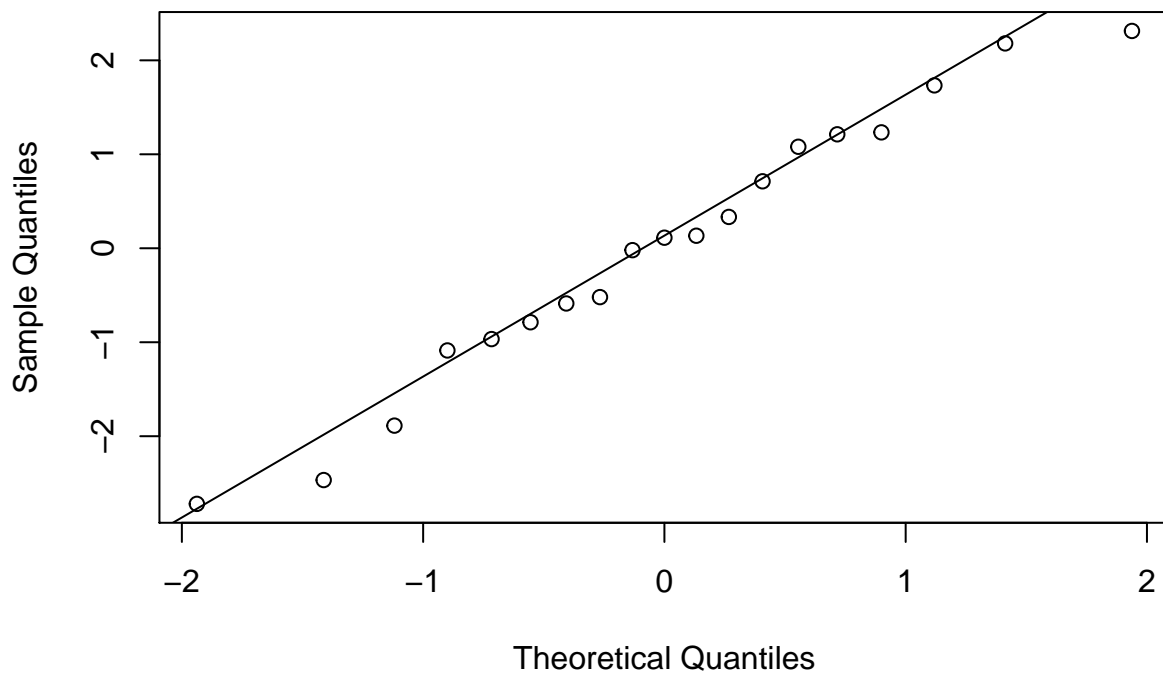
So, we can reject the null hypothesis and accept the alternate hypothesis. This indicates that there are significant differences among true mean CFF values for the three treatments.

c.

The normal probability plot for the current dataset is generated in R as below.

```
##fitted values
Fitted=hw4_2out$fitted.values
##residuals
Residuals=hw4_2out$residuals
###create the normal probability plot
qqnorm(Residuals,main="Normal Probability Plot")
##draw the straight line
qqline(Residuals)
```

Normal Probability Plot



Assumption of Normality:

From the normal probability plot, we can see a linear trend among the data points which is expected for a normally distributed data. So, it indicates that the data come from a normal distribution. It can also be verified using Shapiro-Wilk test and Anderson-Darling test for normality as below.

```
## Shapiro-Wilk normality test
shapiro.test(Residuals)
```

```
##
## Shapiro-Wilk normality test
##
## data: Residuals
## W = 0.97113, p-value = 0.7988
```

```
## Anderson-Darling normality test
library(nortest)
ad.test(Residuals)
```

```
##
```

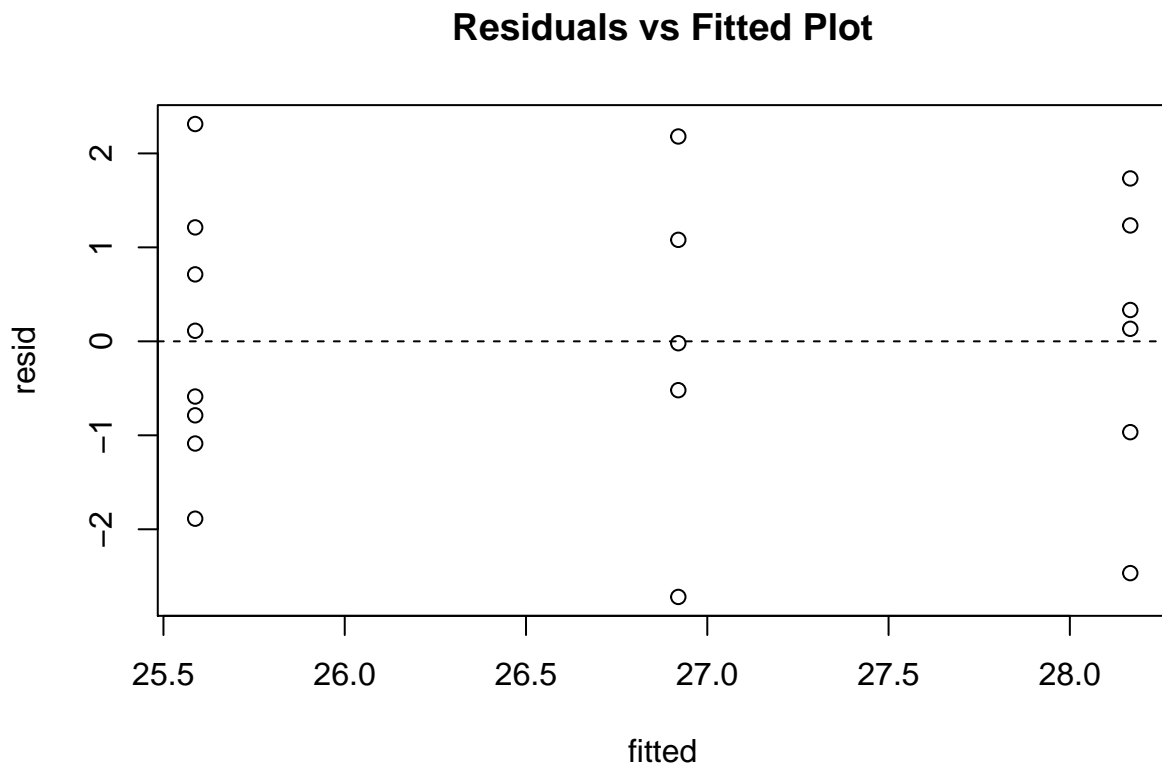
```
## Anderson-Darling normality test
##
## data:  Residuals
## A = 0.16897, p-value = 0.9219
```

Since, for both the tests, $p\text{-value} > \alpha = 0.05$, the null is retained which indicates that the data belong to a normal distribution. So, the assumption of normality, which is required for ANOVA, was not violated.

Assumption of Equal Variance:

The the residuals vs fitted plot for the current dataset is generated in R as below.

```
###create the residuals vs fitted plot
plot(Fitted, main="Residuals vs Fitted Plot", Residuals,xlab="fitted",ylab="resid")
abline(h=0,lty=2)
```



From the plot, we can see the residual values for each treatment plotted along the vertical direction. The horizontal axis denotes the mean values for different treatments. From looking at the vertical spread of the variance of the replicates, it can be said that the four treatments have same variance. So, the second assumption required for ANOVA test was not violated either.

d.

The Tukey method of underscoring (95% confidence) to determine which means differ.

```
#building the data frame
hw4_2=data.frame(Color=c(rep("Brown",8),rep("Green",5),rep("Blue",6)),
  CFF=c(26.8, 27.9, 23.7, 25.0, 26.3, 24.8, 25.7, 24.5,
        26.4, 24.2, 28.0, 26.9, 29.1,
```

```

25.7, 27.2, 29.9, 28.5, 29.4, 28.3))
#carrying out ANOVA
hw4_2out=aov(CFF~Color,hw4_2)
TukeyHSD(hw4_2out,conf.level=0.95)

```

```

## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = CFF ~ Color, data = hw4_2)
##
## $Color
##          diff          lwr          upr      p adj
## Brown-Blue -2.579167 -4.7354973 -0.422836 0.0183579
## Green-Blue -1.246667 -3.6643959  1.171063 0.3994319
## Green-Brown  1.332500 -0.9437168  3.608717 0.3124225

```

```

source(url("http://math.wsu.edu/math/faculty/jpascual/stat423/R/one-way-T-method.R"))
oneway.t.method(hw4_2out,conf.level=0.95,ndigits=2) # T Method

```

```

##          [,1]      [,2]      [,3]
## [1,]      NA 2.156331 2.417729
## [2,] 2.156331      NA 2.276217
## [3,] 2.417729 2.276217      NA

```

T Method (95% Confidence)

	Treatment 1 = Blue Treatment 2 = Brown Treatment 3 = Green		
Treatment:	2	3	1
Mean:	25.59	26.92	28.17
	<hr/> <hr/>		

```

## [1] "Trt 2 mean +wij = 27.86"
## [1] "2 3 Not different."
## [1] "Trt 2 mean +wij = 27.74"
## [1] "2 1 Different."
## [1] "Trt 3 mean +wij = 29.34"
## [1] "3 1 Not different."

```

We see from the underscore plot that the treatment 2 and 3 are not significantly different. Same goes for the treatment 3 and 1. But treatment 2 and 1 are significantly different.