NAME: Himesh Buch RU ID: 190000724

HOMEWORK SET # 3. Due Date: Mar. 5, 2020

**Problem 1.**

Suppose that a single-factor experiment with four levels of the factor has been conducted. There are six replicates and the experiment has been conducted in blocks. The error sum of squares is 500 and the block sum of squares is 250. If the experiment had been conducted as a completely randomized design the estimate of the experimental error variance σ2 would be: (please circle your answer)

25.0 25.5 35.0 37.5 None of the above

**Answer 1.**

K = number of levels of the factor = 4

N = number of replicates = 6

Ntotal = N \* K = 24

Error = N – K = 20

SSE = 500

MSE = SSE / Error = 500 / 20 = **25**

**Problem 2.**

The ANOVA from a randomized complete block experiment output is shown below.

SoV DF SS MS F P-value

Treatment 4 1010.56 ? 29.84 ?

Block ? ? 64.765 ? ? (needed it?)

Error 20 169.33 ?

Total 29 1503.71

1. Fill in all the blanks.

**Sol a):** Here’s the new table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SoV | DF | SS | MS | F | P-value |
| Treatment | 4 | 1010.56 | **252.64** | 29.84 | **3.545E-08** |
| Block | **5** | **323.825** | 64.765 | **7.649** | **0.00037** |
| Error | 20 | 169.33 | **8.467** |  |  |
| Total | 29 | 1503.71 |  |  |  |

MS for treatment = 1010.56 / 4 = **252.64**

MS for error = 169.33 / 20 = **8.467**

SS for blocks = 64.765 \* 5 = **323.825**

F for blocks = 64.765 / 8.467 = **7.649**

Using excel’s ability to find P-Values:

(Using the fdist command)

= fdist(29.84, 4, 20) = 3.545E-08 = **0.00000003545**

P value for block: = fdist(7.649, 5, 20) = **0.00037**

1. How many blocks were used in this experiment?

**Sol b):** Degrees of freedom for blocks = b-1 = 29-24 = **6**

Hence, there are 6 boxes thar were used in this experiment

1. What conclusions can you draw?

**Sol c):** Looking at the P-Value of the treatment, we can say that,

P-Value of treatment < 0.05 (which is the value of \alpha)

Hence, we can **reject** the null hypothesis

**Problem 3**

An aluminum master alloy manufacturer produces grain refiners in ingot form. The company produces the product in four **furnaces**. Each furnace is known to have its own unique operating characteristics, so any experiment run in the foundry that involves more than one furnace will consider furnaces as a nuisance variable. The process engineers suspect that **stirring rate (rpm)** affects the grain size of the product. Each furnace can be run at four different stirring rates. A randomized block design is run for a particular refiner, and the resulting grain size data is as follows:

**Stirring** **Furnace**

**Rate** **1 2 3 4**

**5 |**    8 4 5 6

**10 |** 14 5 6 9

**15 |** 14 6 9 2

**20 |** 17 9 3 6

1. Is there any evidence that stirring rate affects grain size?

**Sol a):** Here is the output generated by the given code:

> ## ANOVA analysis: notice we include the block effect first in the model equation

> afit <- aov(size ~ furnace + stir, data = gsize)

> summary(afit)

Df Sum Sq Mean Sq F value Pr(>F)

furnace 3 165.19 55.06 6.348 0.0133 \*

stir 3 22.19 7.40 0.853 0.4995

Residuals 9 78.06 8.67

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

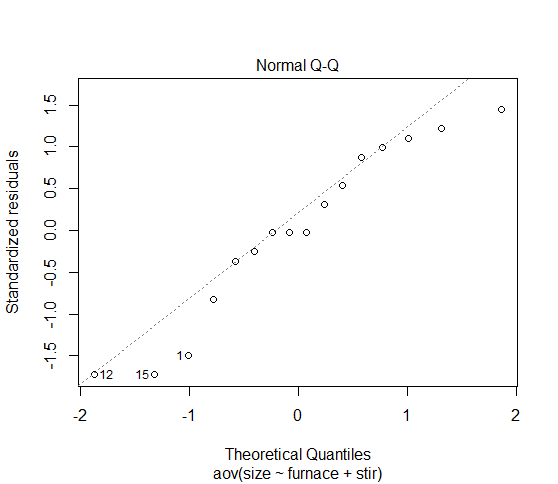
>

The above ANOVA output indicates that there is no difference in mean grain size, hence, it proves that stirring rate has not affected the grain size

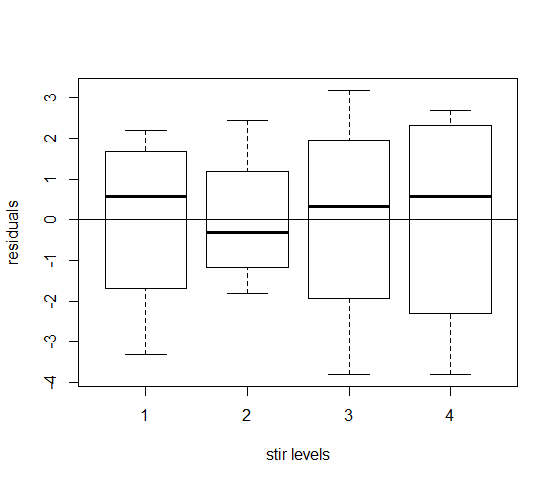
1. Graph the residuals from this experiment on a normal probability plot. Interpret this plot.

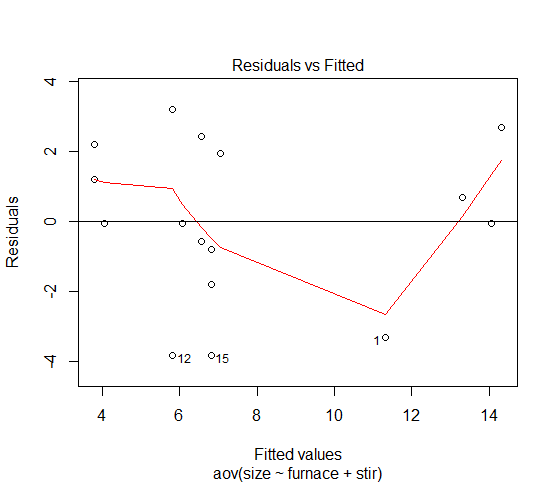
**Sol b):** Here is the graph generated by the given R code:

It indicates that the normality assumption is valid



1. Plot the residuals versus furnace and stirring rate. Does this plot convey any useful information?





Looking at the above graphs, we can say that, the variance is consistent at different stirring rates. It also identifies that different stirring rates do not affect variance.

1. What should the process engineers recommend concerning the choice of stirring rate and furnace for this particular grain refiner if small grain size is desirable?

**Sol d):** There is no effect due to the stirring rate

**Problem 4**

The effect of five different ingredients (*A, B, C, D, E*) on the reaction time of a chemical process is being studied. Each batch of new material is only large enough to permit five runs to be made. Furthermore, each run requires approximately 1.5 hours, so only five runs can be made in one day. The experimenter decides to run the experiment as a Latin square so that day and batch effects may be systematically controlled. She obtains the data that follow.

1. Analyze the data from this experiment (use α=0.5) and draw conclusions.

|  | **Day** | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Batch** | **1** | **2** | **3** | **4** | **5** |  |
| **1** | A=8 | B=7 | D=1 | C=7 | E=3 |  |
| **2** | C=11 | E=2 | A=7 | D=3 | B=8 |  |
| **3** | B=4 | A=9 | C=10 | E=1 | D=5 |  |
| **4** | D=6 | C=8 | E=6 | B=6 | A=10 |  |
| **5** | E=4 | D=2 | B=3 | A=8 | C=8 |  |

**Sol a):** Here’s the ANOVA output which was generated by running the given code:

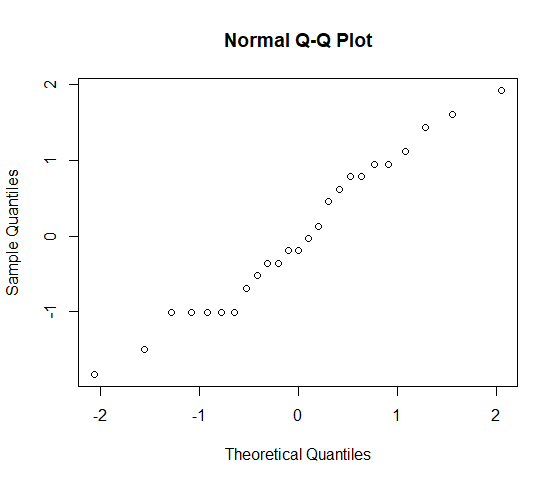
|  |
| --- |
| > anova(ls.fit)  Analysis of Variance Table  Response: y  Df Sum Sq Mean Sq F value Pr(>F)  row 4 15.44 3.860 1.2345 0.3476182  col 4 12.24 3.060 0.9787 0.4550143  trt 4 141.44 35.360 11.3092 0.0004877 \*\*\*  Residuals 12 37.52 3.127  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1 |
|  |
| |  | | --- | | > | |

By looking at the output, we can say that the variable “trt” (the treatment factor) has a significant effect because its p-value is very close to 0. We can also say that the effect of the other two variables “row” and “col” are negligible by looking at the p-values.

**(b)** Check Normality and Homogeneity of Variance assumptions. (can use plots or tests)

**Sol b):** The normal probability plot and the residual graphs below do not give us any reason to question our normality or identical independent assumptions. This was generated by using the R command qqnorm. Here is the command that was used,

*qqnorm(rstandard(ls.fit))*



R CODE FOR PROBLEM # 3

## Create factor levels and enter the data

stir <- **factor**(**rep**(1:4, each = 4)) ## you can use the **=** sign

furnace <- **factor**(**rep**(1:4, times = 4)) ## instead of the **<-** arrow sign

size <- **c**(8, 4, 5, 6,

14, 5, 6, 9,

14, 6, 9, 2,

17, 9, 3, 6)

gsize <- **data.frame**(size, stir, furnace)

gsize # Check the data

str(gsize) # This command shows the structure of object called gsize

## ANOVA analysis: notice we include the block effect first in the model equation

afit <- **aov**(size ~ furnace + stir, data = gsize)

**summary**(afit)

## The plot to visualize normality is given by the command below

plot(afit,2)

## The plot to visualize residuals versus stir is given by the command below

plot(stir, afit$residual, xlab='stir levels', ylab='residuals'); abline(0,0)

## plot(stir, afit$residual, xlab=’stir levels’, ylab=’residuals’); abline(0,0)

## Notice the difference in the quotes symbol. The first is done using the R editor.

## The second was done using MS Word, and it does not work in R. MS Word adds metadata.

## You can use a .txt editor such as MS Notepad to generate a flat file, with no metadata

## Next plot shows residuals versus predicted values

plot(afit,1); abline(0,0)

R CODE FOR PROBLEM # 4

## Create the two-blocking factor and treatment levels and enter the data

row = **factor**(**rep**(1:5, each = 5))

col = **factor**(**rep**(1:5, times = 5))

trt = **factor**(**c(1,2,4,3,5,3,5,1,4,2,2,1,3,5,4,4,3,5,2,1,5,4,2,1,3**))

y = **c**(8, 7, 1, 7, 3,

11,2, 7, 3, 8,

4, 9, 10, 1, 5,

6, 8, 6, 6, 10,

4, 2, 3, 8, 8)

ls5by5 = **data.frame**(row, col, trt, y)

ls5by5 # Check if the data is correct

str(ls5by5)

attach(ls5by5)

trt.means = tapply(y,trt,mean) # these are the means for each treatment level

trt.means

# Now we fit the latin square model

ls.fit = lm(y ~ row + col + trt)

anova(ls.fit)

*Notice that R does not use the same restrictions that we have in the book to solve the Normal Equations. With the usual restriction* ***Στi = 0*** *we have* ***τi = Ῡi .. - Ῡ...*** *with* ***µ =******Ῡ...***

*But R use the restriction* ***τ1 = 0.*** *Therefore, we have* ***τi = Ῡi .. - Ῡ...*** *for i = 2, 3, 4, 5 and* ***τ1 = 0.***

*We know that the cell means are the estimators of* ***µ1 = 8.4 µ2 = 5.6 µ3 = 8.8 µ4 = 3.4 µ5 = 3.2. and µ = 5.88 (mean of the µi values)*** *We can always estimate these cell means. They are of the form* ***µi = µ + τi,*** *but we cannot estimate in unique form the* ***τi*** *parameters neither* ***µ*** *separately. To estimate these parameters, we need restrictions to solve the Normal Equations, and these restrictions are not unique.*