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 σ^2 would be: (please circle your answer)

25.0   25.5   35.0   37.5   None of the above

$$\frac{(500+250)}{4\times 6-4}=\frac{750}{20}=37.5$$

Ans: 37.5

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	**252.64**	29.84	**3.545e-08**	
Block	**5**	**323.82**	64.765	?	?	(needed it?)
Error	20	169.33	**8.467**			
Total	29	1503.71				

a) Fill in all the blanks.

$MS_{treatment}=1010.56/4=252.64; SS_{Blocks}=1503.71-1010.56-169.33=323.82$
$df_{blocks}=323.82/64.765=5, MS_E=169.33/20=8.467;$

```
> pf(29.84,4,20,lower.tail=FALSE)
[1] 3.544848e-08
```

Ans: see bold numbers on ANOVA table

b) How many blocks were used in this experiment?

5 + 1 = 6

Ans: 6

c) What conclusions can you draw?

Ans: The treatments seem to differ significantly (*p*-value very very small)

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 Rate	**Furnace 1**	**2**	**3**	**4**
5	8	4	5	6
10	14	5	6	9
15	14	6	9	2
20	17	9	3	6

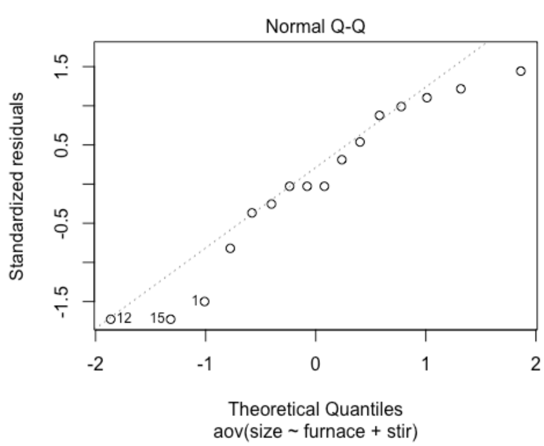
```
> stir=factor(rep(1:4, each=4))
> furnace=factor(rep(1:4,times=4))
> size=c(8,4,5,6,14,5,6,9,14,6,9,2,17,9,3,6)
> gsize=data.frame(size,stir,furnace)
> gsize
   size stir furnace
1     8    1       1
2     4    1       2
3     5    1       3
4     6    1       4
5    14    2       1
6     5    2       2
7     6    2       3
8     9    2       4
9    14    3       1
10    6    3       2
11    9    3       3
12    2    3       4
13   17    4       1
14    9    4       2
15    3    4       3
16    6    4       4
```

```
> str(gsize)
'data.frame':	16 obs. of  3 variables:
 $ size   : num  8 4 5 6 14 5 6 9 14 6 ...
 $ stir   : Factor w/ 4 levels "1","2","3","4": 1 1 1 1 2 2 2 2 3 3 ...
 $ furnace: Factor w/ 4 levels "1","2","3","4": 1 2 3 4 1 2 3 4 1 2 ...
> 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
```

a) Is there any evidence that stirring rate affects grain size?

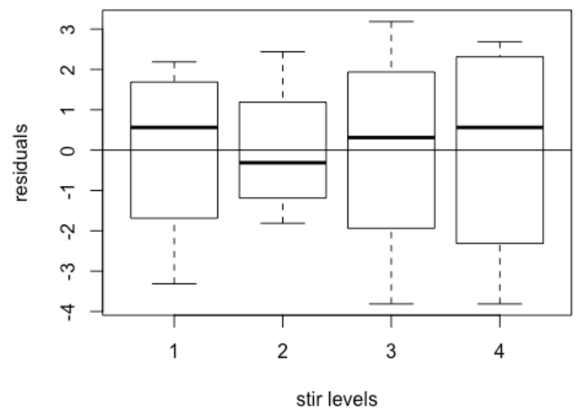
Ans: No (p-value = 0.4995 > 0.05)

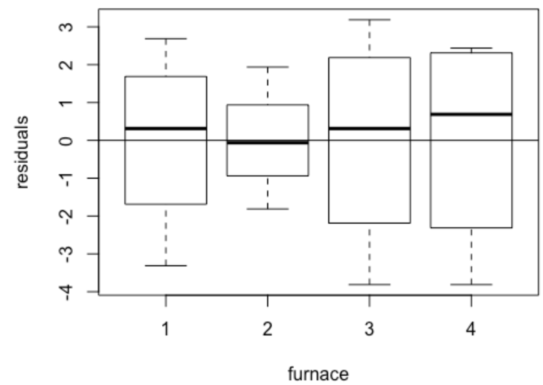
b) Graph the residuals from this experiment on a normal probability plot. Interpret this plot.



Ans: Although the residuals do fall close to a normal probability plot, there appears to be a S-shaped pattern

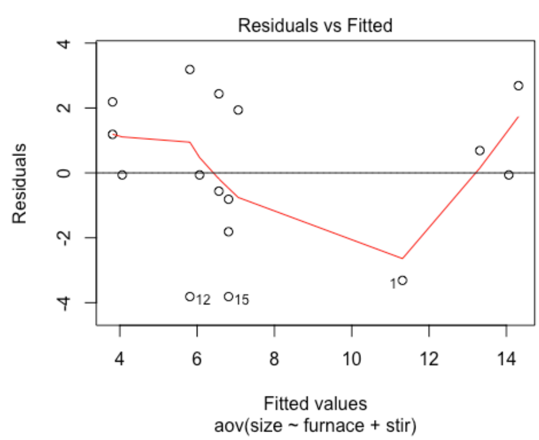
c) Plot the residuals versus furnace and stirring rate. Does this plot convey any useful information?





Ans: Not much useful information (both plots show that residuals fall within ±1 residuals)

d)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?



Ans: a. Rate 5, furnace 1; b. Rate 15, Furnace 4; c. Rate 20, Furnace 3

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.

(a) 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

```
> 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
   row col trt  y
1    1   1   1  8
2    1   2   2  7
3    1   3   4  1
4    1   4   3  7
5    1   5   5  3
6    2   1   3 11
7    2   2   5  2
8    2   3   1  7
9    2   4   4  3
10   2   5   2  8
11   3   1   2  4
12   3   2   1  9
13   3   3   3 10
14   3   4   5  1
15   3   5   4  5
16   4   1   4  6
17   4   2   3  8
18   4   3   5  6
19   4   4   2  6
20   4   5   1 10
21   5   1   5  4
22   5   2   4  2
23   5   3   2  3
24   5   4   1  8
25   5   5   3  8
```

```
> str(ls5by5)
'data.frame':	25 obs. of  4 variables:
 $ row: Factor w/ 5 levels "1","2","3","4",..: 1 1 1 1 1 2 2 2 2 2 ...
 $ col: Factor w/ 5 levels "1","2","3","4",..: 1 2 3 4 5 1 2 3 4 5 ...
 $ trt: Factor w/ 5 levels "1","2","3","4",..: 1 2 4 3 5 3 5 1 4 2 ...
 $ y  : num  8 7 1 7 3 11 2 7 3 8 ...
> attach(ls5by5)
The following objects are masked _by_ .GlobalEnv:

    col, row, trt, y

> trt.means=tapply(y,trt,mean)
> trt.means
  1   2   3   4   5 
8.4 5.6 8.8 3.4 3.2 
```

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

Ans: Significance difference is observed (*p*-value = 0.0004877 < 0.05)

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

```
> bartlett.test(y~trt,data=ls5by5)

	Bartlett test of homogeneity of variances

data:  y by trt
Bartlett's K-squared = 1.5544, df = 4, p-value = 0.817
```

Ans: Normality and homogeneity of variance is verified (*p*-value = 0.817 > 0.05)

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 $\boldsymbol{\Sigma\tau_i = 0}$ *we have* $\boldsymbol{\tau_i = \bar{Y}_{i..} - \bar{Y}_{...}}$ *with* $\boldsymbol{\mu = \bar{Y}_{...}}$

But R use the restriction $\boldsymbol{\tau_1 = 0.}$ *Therefore, we have* $\boldsymbol{\tau_i = \bar{Y}_{i..} - \bar{Y}_{...}}$ *for i = 2, 3, 4, 5 and* $\boldsymbol{\tau_1 = 0.}$

We know that the cell means are the estimators of $\boldsymbol{\mu_1 = 8.4}$ $\boldsymbol{\mu_2 = 5.6}$ $\boldsymbol{\mu_3 = 8.8}$ $\boldsymbol{\mu_4 = 3.4}$ $\boldsymbol{\mu_5 = 3.2.}$ ***and*** $\boldsymbol{\mu = 5.88}$ ***(mean of the*** $\boldsymbol{\mu_i}$ ***values)*** *We can always estimate these cell means. They are of the form* $\boldsymbol{\mu_i = \mu + \tau_i,}$ *but we cannot estimate in unique form the* $\boldsymbol{\tau_i}$ *parameters neither* $\boldsymbol{\mu}$ *separately. To estimate these parameters, we need restrictions to solve the Normal Equations, and these restrictions are not unique.*