Stat 5309 Lab 4b

Tom Wilson

Feb 21, 2019

1.

An industrial engineer is conducting an experiment on eye focus time. He is interested in the effect of the distance of the object from the eye on the focus time. Four different distances are of interest. He has five subjects available for the experiment. Because there may be differences among individuals, he decides to conduct the experiment in a randomized block design. The data obtained follow. Analyze the data from this experiment (use

$$\alpha = 0.05$$

) and draw appropriate conclusions.

a

Set up the data frame named eye, using subject as a blocking factor, distance as treatment factor and time as response.

```
distances <- c("4","6","8","10")
subjects <- c("1","2","3","4","5")
eye <- expand.grid(subject = subjects, distance = distances)
eye <- cbind(eye,time=c(10,6,6,6,6,
                        7,6,6,1,6,
                        5,3,3,2,5,
                        6,4,4,2,3
                        )
             )
eye %>% kable()
```

subject	distance	time
1	4	10
2	4	6
3	4	6
4	4	6
5	4	6
1	6	7
2	6	6
3	6	6
4	6	1
5	6	6
1	8	5
2	8	3
3	8	3
4	8	2
5	8	5
1	10	6
2	10	4
3	10	4

subject	distance	time
4	10	2
5	10	3

b

Build a linear model named eye.mod.

Are the subject means significantly different?

Are the Distance means significantly different?

c

Which distances bring the longest/shortest focus time?

d

Calculate the sample size (number of treatment replicates) for power > 0.90

2

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 data that follow. Analyze the data from this experiment (use

$$\alpha = 0.05$$

) and draw conclusions.

a

set up a dataframe

```
batches <- c("b1","b2","b3","b4","b5")
days <- c("d1","d2","d3","d4","d5")
chemical <- expand.grid(day = days,
                        batch = batches
                        )
chemical <- cbind(chemical,
                  ingredient = c("A","B","D","C","E",
                                 "C","E","A","D","B",
                                 "B","A","C","E","D",
                                 "D","C","E","B","A",
                                 "E","D","B","A","C"
                                 ),
                  time = c(8,7,1,7,3,
                           11,2,7,3,8,
```

```r
                              4,9,10,1,5,
                              6,8,6,6,10,
                              4,2,3,8,8)
                     )

chemical %>% kable()
```

day	batch	ingredient	time
d1	b1	A	8
d2	b1	B	7
d3	b1	D	1
d4	b1	C	7
d5	b1	E	3
d1	b2	C	11
d2	b2	E	2
d3	b2	A	7
d4	b2	D	3
d5	b2	B	8
d1	b3	B	4
d2	b3	A	9
d3	b3	C	10
d4	b3	E	1
d5	b3	D	5
d1	b4	D	6
d2	b4	C	8
d3	b4	E	6
d4	b4	B	6
d5	b4	A	10
d1	b5	E	4
d2	b5	D	2
d3	b5	B	3
d4	b5	A	8
d5	b5	C	8

b

build a linear model using aov. Do the ingredients affect the reaction time? Day means, Batch means, Ingredient means, are significantly different? Check interaction between day and batch.

c

Find the lowest reaction time.

3

An industrial engineer is investigating the effect of four assembly methods (A,B,C,D) on the assembly time for a color television component. Four operators are selected for the study. Furthermore, the engineer knows that each assembly method produces such factigue that the time required for the last assembly may be greater than the time required for the first, regardless of the method. That is, a trend develops in the required assembly time. To account for this source of variability, the engineer uses the Latin square design shown below. Analyze the data from this experiment and draw appropriate conclusions.

a

setup the dataframe.

```
operators <- c("1","2","3","4")
assembly_order <- c("1","2","3","4")
tv_component <- expand.grid(operator = operators,
                            assembly = assembly_order
                            )
tv_component <- cbind(tv_component,
                      treatment = c("C","D","A","B",
                                    "B","C","D","A",
                                    "A","B","C","D",
                                    "D","A","B","C"
                                    ),
                      time = c(10,14,7,8,
                               7,18,11,8,
                               5,10,11,9,
                               10,10,12,14)
                      )
tv_component %>% kable()
```

operator	assembly	treatment	time
1	1	C	10
2	1	D	14
3	1	A	7
4	1	B	8
1	2	B	7
2	2	C	18
3	2	D	11
4	2	A	8
1	3	A	5
2	3	B	10
3	3	C	11
4	3	D	9
1	4	D	10
2	4	A	10
3	4	B	12
4	4	C	14

b

build a linear odel using aov. Do the treatments affect the assembly time? Operator means, assembly means, treatment means, are they significantly different?

c

Find the lowest assembly time.

1.

A chemist wishes to test the effect of four chemical agents on the strength of a particular type of cloth. Because there might be variablility from one bolt to another, the chemist decides to use a randomized block design, with the bolts of cloth considered as blocks. She selects five bolts and applies all four chemicals in random order to each bolt. The resulting tensile strengths follow. Analyze the data from this experiment (use

$$\alpha = 0.05$$

) and draw appropriate conclusions.

a

Create a vector for Blocks, named "Bold": 5 levels. Total 20. Create a vector for Treatments, named "Chemical". Total 20. Create a response vector, named "Strength". Set up the data frame named "chem".

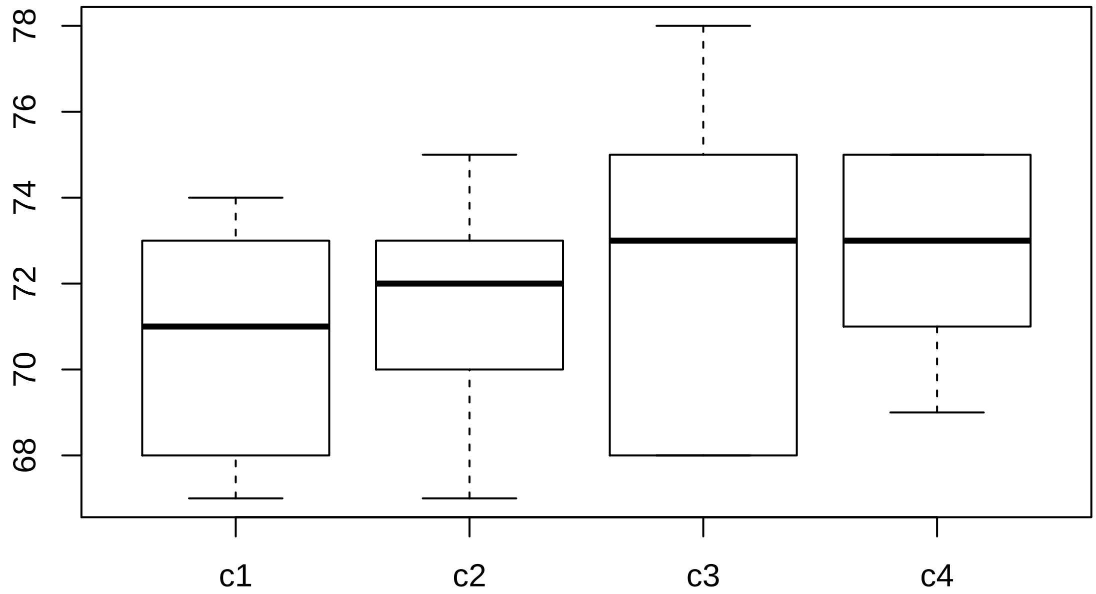
```
bolts <- c("b1","b2","b3","b4","b5")
chemicals <- c("c1","c2","c3","c4")
chem_data <- expand.grid(bolt = bolts,chemical = chemicals)
chem_data <- cbind(chem_data,strength = c(73,68,74,71,67,
                                          73,67,75,72,70,
                                          75,68,78,73,68,
                                          73,71,75,75,69))
chem_data %>% kable()
```

bolt	chemical	strength
b1	c1	73
b2	c1	68
b3	c1	74
b4	c1	71
b5	c1	67
b1	c2	73
b2	c2	67
b3	c2	75
b4	c2	72
b5	c2	70
b1	c3	75
b2	c3	68
b3	c3	78
b4	c3	73
b5	c3	68
b1	c4	73
b2	c4	71
b3	c4	75
b4	c4	75
b5	c4	69

b

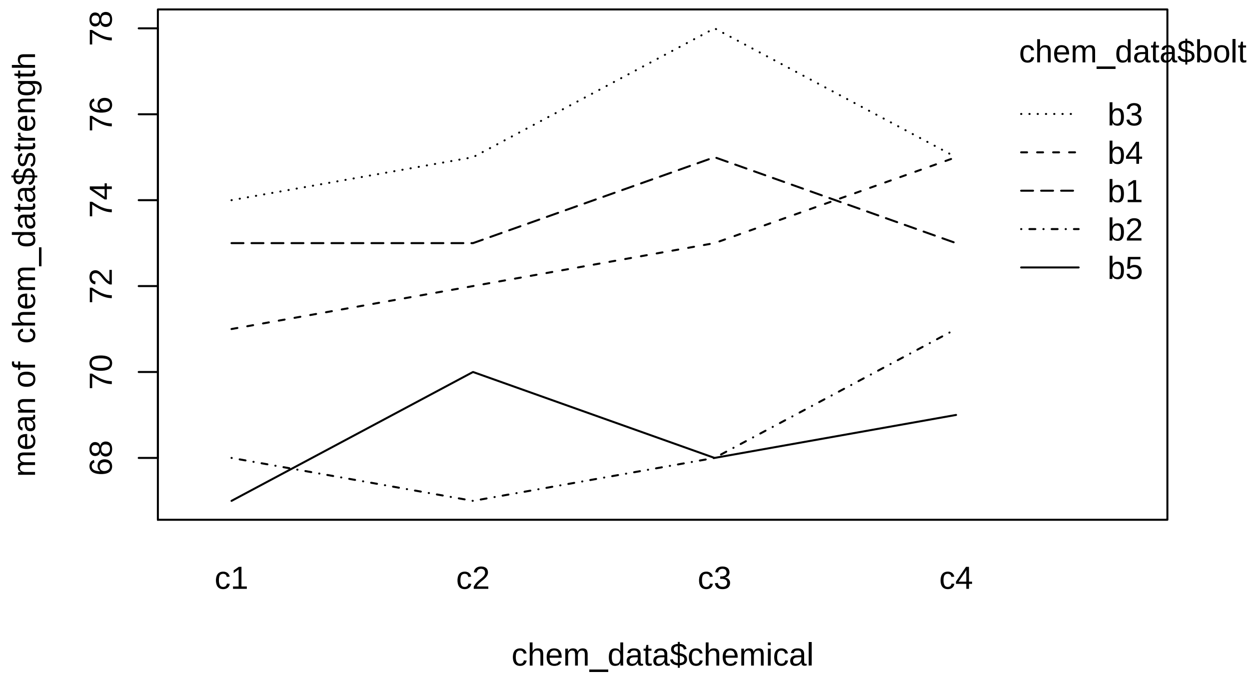
Any evidence that the Chemical affect Strength?

```
boxplot(strength~chemical,data=chem_data)
```



visually, there is a small difference in strength based on chemical.

```
interaction.plot(chem_data$chemical,chem_data$bolt, chem_data$strength)
```



On closer inspection, the effect of chemical is consistent across different bolts (not much interaction).

c

Perform a TukeyHSD to compare the the treatment means. Which chemical is the preferred (bring the highest strengh)

```
strength_model <- aov(strength~chemical+bolt,data = chem_data)
TukeyHSD(strength_model, conf.level=0.95)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = strength ~ chemical + bolt, data = chem_data)
##
## $chemical
##       diff        lwr      upr     p adj
## c2-c1  0.8 -1.7308322 3.330832 0.7852734
## c3-c1  1.8 -0.7308322 4.330832 0.2042593
## c4-c1  2.0 -0.5308322 4.530832 0.1417326
## c3-c2  1.0 -1.5308322 3.530832 0.6540138
## c4-c2  1.2 -1.3308322 3.730832 0.5182726
## c4-c3  0.2 -2.3308322 2.730832 0.9952030
##
## $bolt
##        diff        lwr        upr     p adj
```

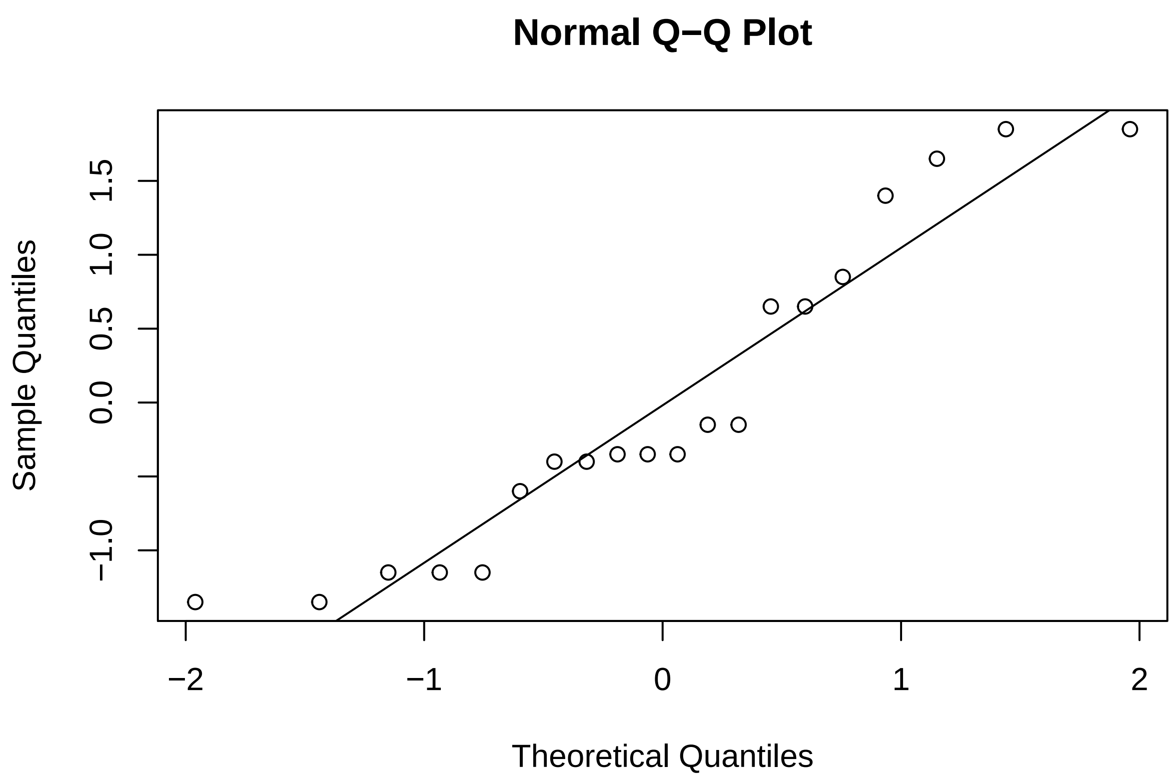
```
## b2-b1 -5.00  -8.037831 -1.9621691 0.0015656
## b3-b1  2.00  -1.037831  5.0378309 0.2814173
## b4-b1 -0.75  -3.787831  2.2878309 0.9295872
## b5-b1 -5.00  -8.037831 -1.9621691 0.0015656
## b3-b2  7.00   3.962169 10.0378309 0.0000717
## b4-b2  4.25   1.212169  7.2878309 0.0056966
## b5-b2  0.00  -3.037831  3.0378309 1.0000000
## b4-b3 -2.75  -5.787831  0.2878309 0.0830636
## b5-b3 -7.00 -10.037831 -3.9621691 0.0000717
## b5-b4 -4.25  -7.287831 -1.2121691 0.0056966
```

at a confidence of 95%, there are no significant differences in strength between chemicals.

d

Check the assumption of the residuals.

```r
qqnorm(strength_model$residuals)
qqline(strength_model$residuals)
```



The quantiles of the residual closely match the quantiles of a normal distribution.

2.

Three different washing solutions are being compared to study their effectiveness in retarding bacteria growth in 5-gallon milk containers. The analysis is done in a laboratory, and only three trials can be run on any day. Because days could represent a potential source of variability, the experimenter decides to use a randomized block design. Observations are taken for four days, and the data are shown here. Analyze the data from this experiment (use

$$\alpha = 0.05$$

) and draw conclusions.

e

Create a vector for blocks named days 4 levels total 12. Create a vector for treatments named solutions total 12. create a response vector named Growth set up the data frame.

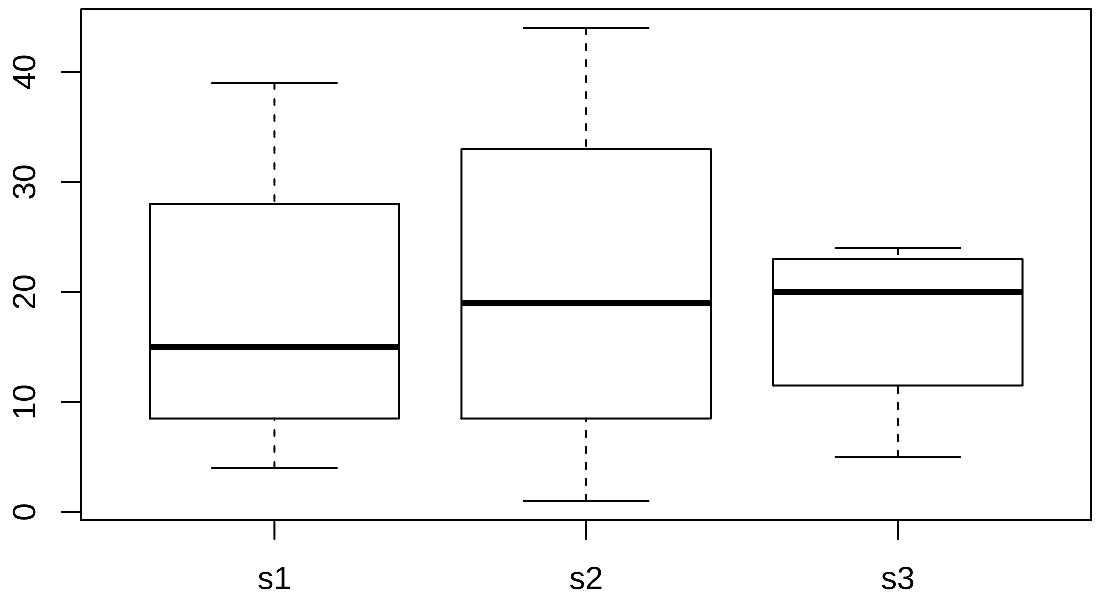
```
days <- c("d1","d2","d3","d4")
solutions <- c("s1","s2","s3")
bacteria_data <- expand.grid(solution = solutions, day = days)
bacteria_data <- cbind(bacteria_data, growth = c(13,22,18,39,
                                                  16,24,17,44,
                                                  5,4,1,22))
bacteria_data %>% kable()
```

solution	day	growth
s1	d1	13
s2	d1	22
s3	d1	18
s1	d2	39
s2	d2	16
s3	d2	24
s1	d3	17
s2	d3	44
s3	d3	5
s1	d4	4
s2	d4	1
s3	d4	22

f

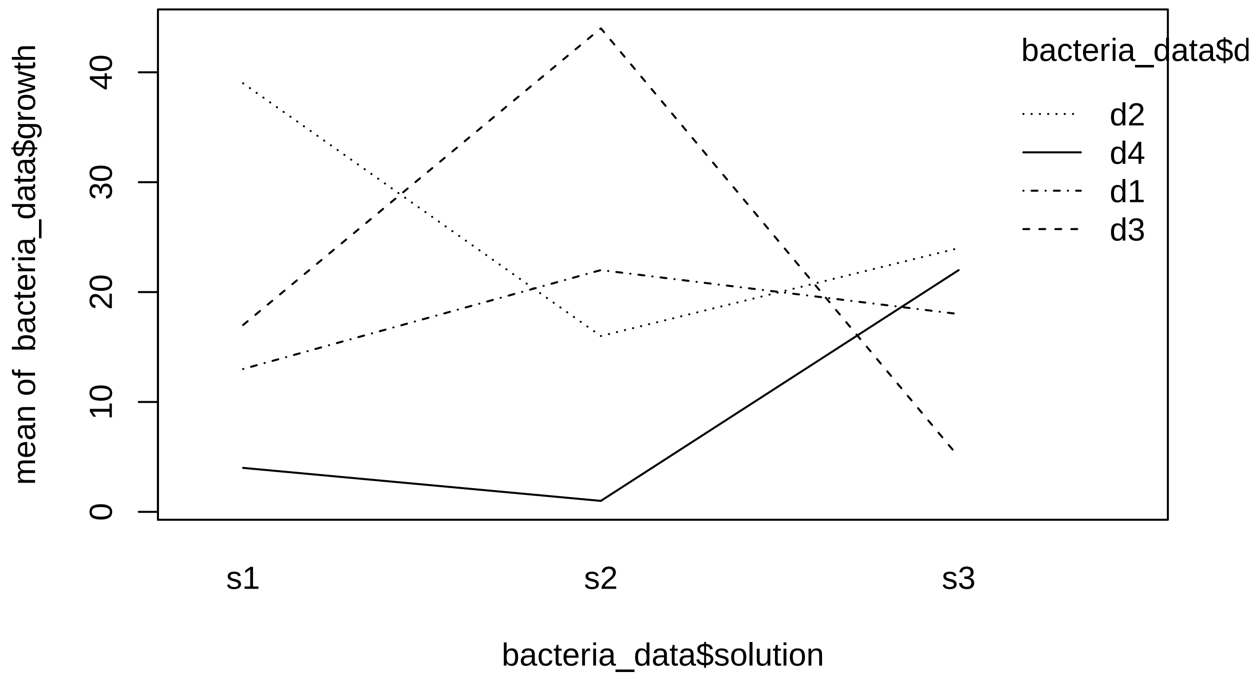
Any evidence that the solutions affect bacteria growth?

```
boxplot(growth~solution,data=bacteria_data)
```



Visually, there is not much difference in growth based on solution.

```
interaction.plot(bacteria_data$solution,bacteria_data$day, bacteria_data$growth)
```



On closer inspection, there could be some significant interactions between solution and day.

g

Perform a TukeyHSD to compare the the treatment means. Which chemical is the preferred (brings the lowest bacterial growth)

```
growth_model <- aov(growth~solution+day,data = bacteria_data)
TukeyHSD(growth_model, conf.level=0.95)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = growth ~ solution + day, data = bacteria_data)
##
## $solution
##       diff       lwr      upr     p adj
## s2-s1  2.5 -29.96351 34.96351 0.9698442
## s3-s1 -1.0 -33.46351 31.46351 0.9950912
## s3-s2 -3.5 -35.96351 28.96351 0.9420329
##
## $day
##             diff       lwr      upr     p adj
## d2-d1   8.666667 -33.62565 50.95899 0.8899174
## d3-d1   4.333333 -37.95899 46.62565 0.9832670
## d4-d1  -8.666667 -50.95899 33.62565 0.8899174
## d3-d2  -4.333333 -46.62565 37.95899 0.9832670
```

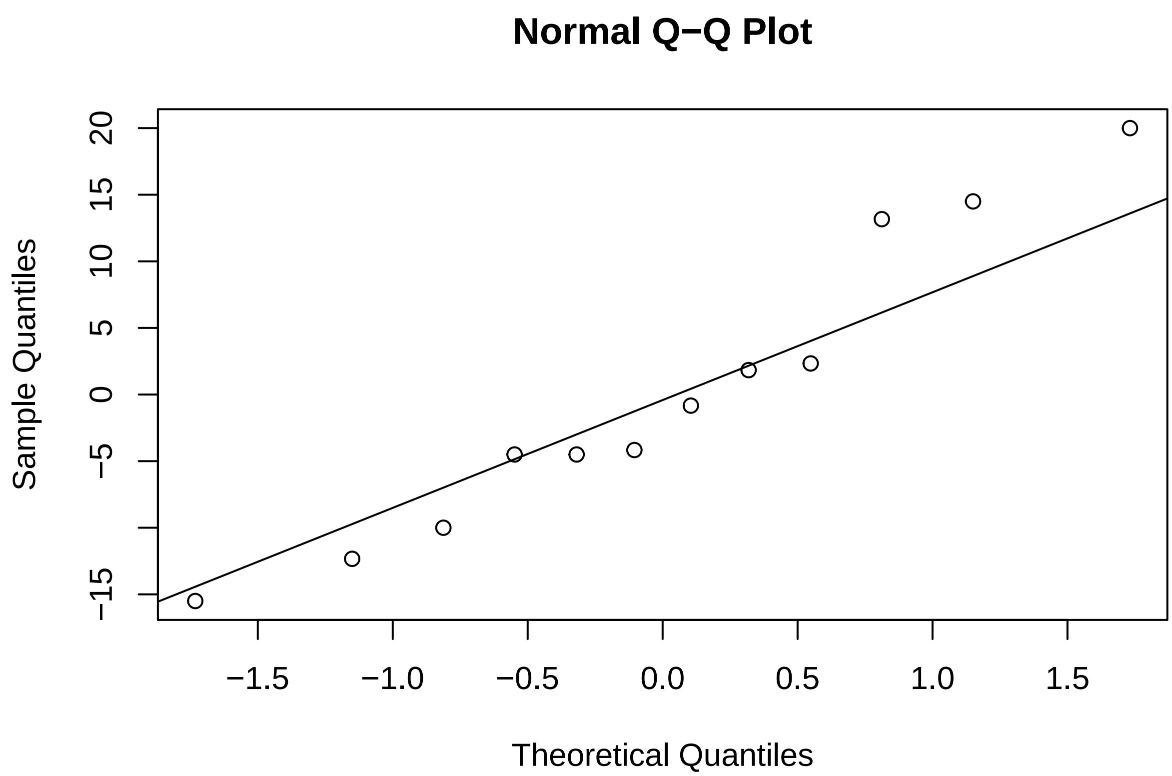
```
## d4-d2 -17.333333 -59.62565 24.95899 0.5327553
## d4-d3 -13.000000 -55.29232 29.29232 0.7217433
```

Comparing pair-wise, there is no significant difference growth between solutions or days.

h

Check the assumption of the residuals.

```
qqnorm(growth_model$residuals)
qqline(growth_model$residuals)
```



The residuals are very close to normally distributed.

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 characterisitcs, 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 affects the grain size of the product. Each furnace can be run at four different stirring rates. A randomized block design is run fro a particular refiner and the resulting grain size data is shown below.

```
stir_rates <- c("5","10","15","20")
furnaces <- c("f1","f2","f3","f4")
```

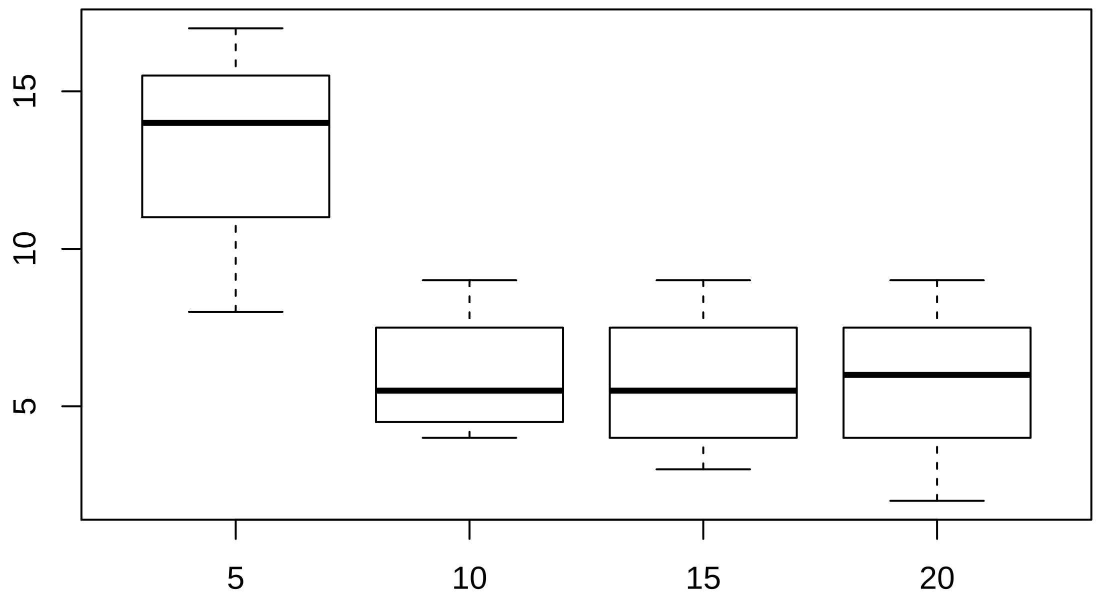
```r
aluminum_data <- expand.grid(stir_rate = stir_rates,furnace = furnaces)
aluminum_data <- cbind(aluminum_data, grain_size = c(8,4,5,6,
                                                      14,5,6,9,
                                                      14,6,9,2,
                                                      17,9,3,6))
aluminum_data %>% kable()
```

stir_rate	furnace	grain_size
5	f1	8
10	f1	4
15	f1	5
20	f1	6
5	f2	14
10	f2	5
15	f2	6
20	f2	9
5	f3	14
10	f3	6
15	f3	9
20	f3	2
5	f4	17
10	f4	9
15	f4	3
20	f4	6

a

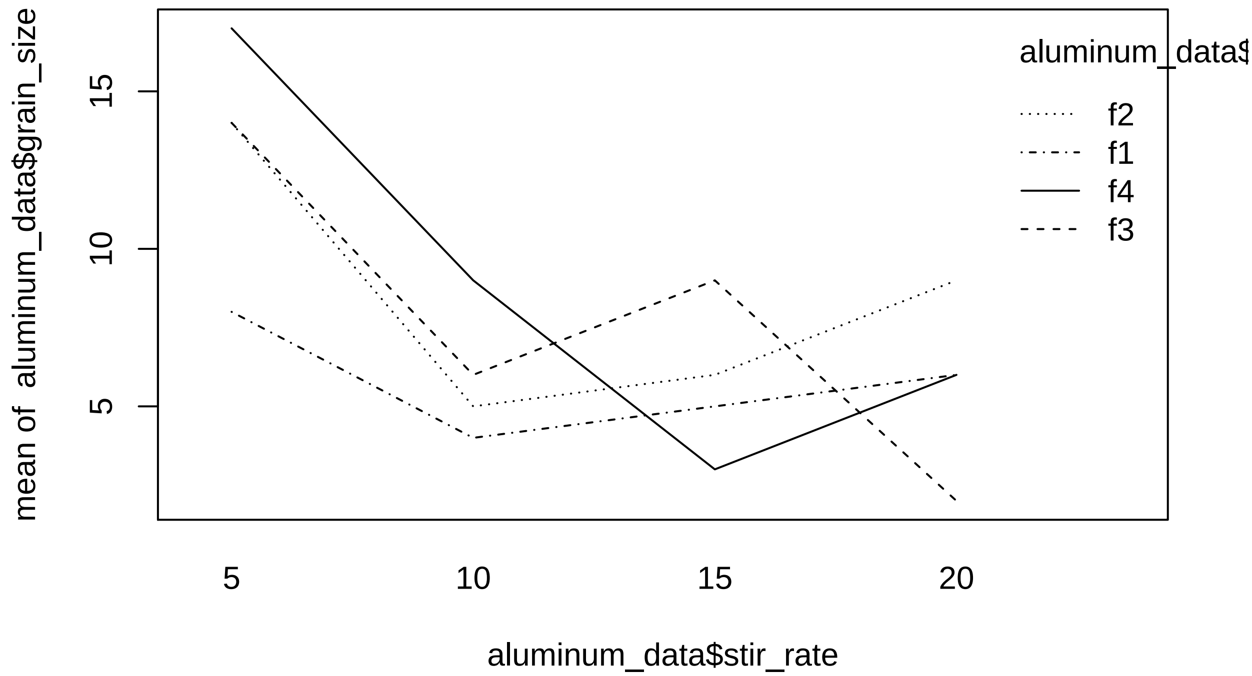
Is there any evidence that stirring rate affects grain size?

```r
boxplot(grain_size~stir_rate,data=aluminum_data)
```



Visually, the slowest stir rate (5 rpm) has a dramatically larger grain size.

```
interaction.plot(aluminum_data$stir_rate,aluminum_data$furnace, aluminum_data$grain_size)
```

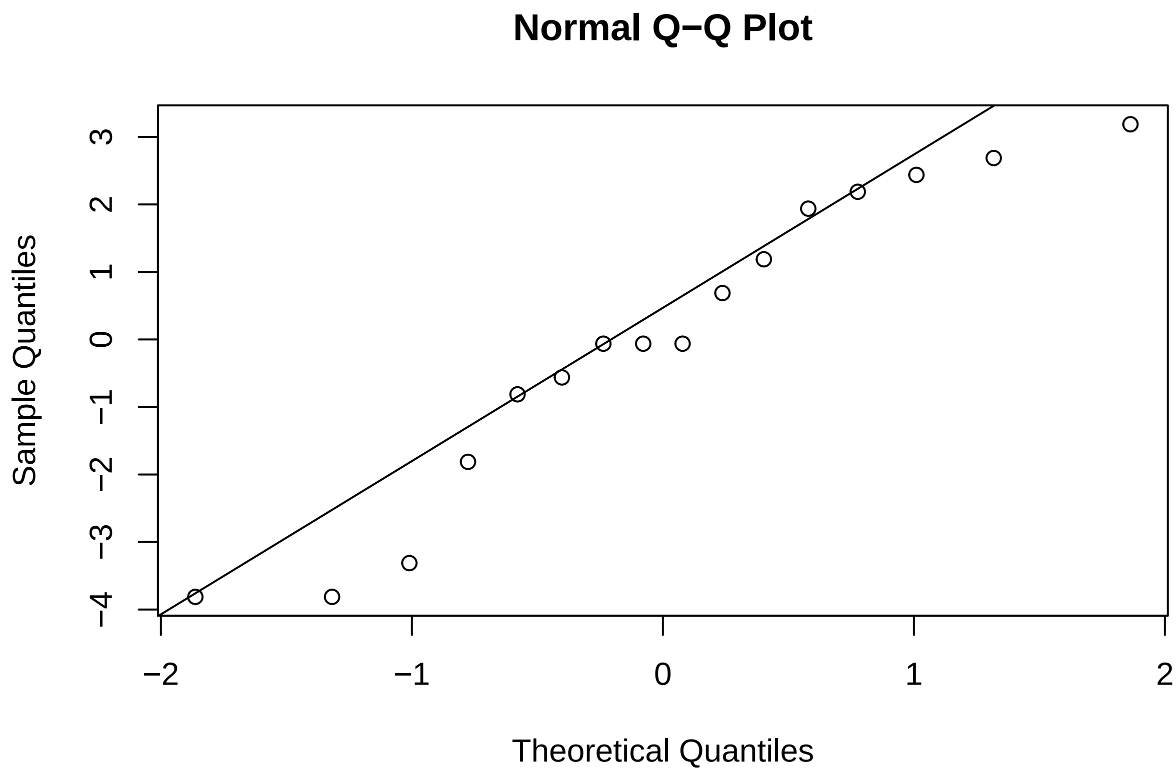


On closer inspection, the effect of stir rate on grain size is consistent across different furnaces.

b

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

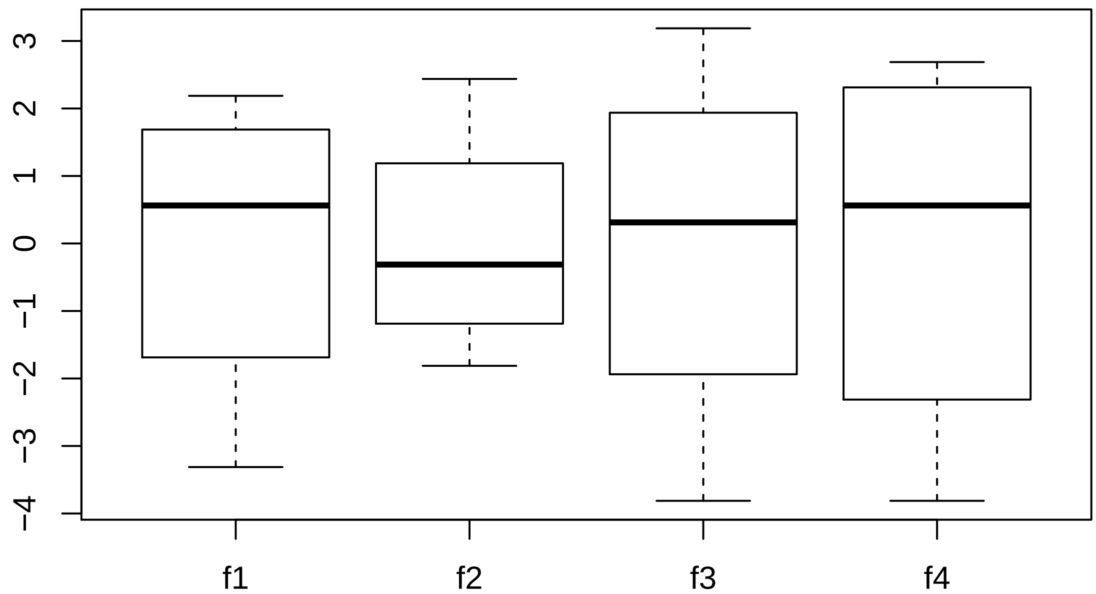
```
grain_model <- aov(grain_size~stir_rate+furnace,data=aluminum_data)
qqnorm(grain_model$residuals)
qqline(grain_model$residuals)
```



c

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

```
plot(x=aluminum_data$furnace,y=grain_model$residuals)
```



This plot shows unequal variance among furnaces included in the model.

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?

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

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = grain_size ~ stir_rate + furnace, data = aluminum_data)
##
## $stir_rate
##        diff        lwr        upr     p adj
## 10-5  -7.25 -13.751145 -0.7488546 0.0291472
## 15-5  -7.50 -14.001145 -0.9988546 0.0243746
## 20-5  -7.50 -14.001145 -0.9988546 0.0243746
## 15-10 -0.25  -6.751145  6.2511454 0.9993332
## 20-10 -0.25  -6.751145  6.2511454 0.9993332
## 20-15  0.00  -6.501145  6.5011454 1.0000000
##
## $furnace
##        diff       lwr      upr     p adj
## f2-f1  2.75 -3.751145 9.251145 0.5736833
## f3-f1  2.00 -4.501145 8.501145 0.7743492
```

```
## f4-f1  3.00 -3.501145 9.501145 0.5074914
## f3-f2 -0.75 -7.251145 5.751145 0.9829916
## f4-f2  0.25 -6.251145 6.751145 0.9993332
## f4-f3  1.00 -5.501145 7.501145 0.9615966
```

Stir rate of 5 produces the largest grain size. Choice of furnace is not significant.