

STAT 425
STAT 425
STAT 425

STAT 425 – Ass01

Daniel Girvitz, 30062205

19/01/2022

QUESTION 3.9

4.5/5

```
mix_tech <- c(1,1,1,1,
              2,2,2,2,
              3,3,3,3,
              4,4,4,4)
tensile_strength <- c(3129,3000,2865,2890,
                      3200,3300,2975,3150,
                      2800,2900,2985,3050,
                      2600,2700,2600,2765)
cement <- data.frame(mix_tech,tensile_strength)
str(cement)

## 'data.frame':   16 obs. of  2 variables:
## $ mix_tech      : num  1 1 1 1 2 2 2 2 3 3 ...
## $ tensile_strength: num  3129 3000 2865 2890 3200 ...
```

QUESTION 3.9.a

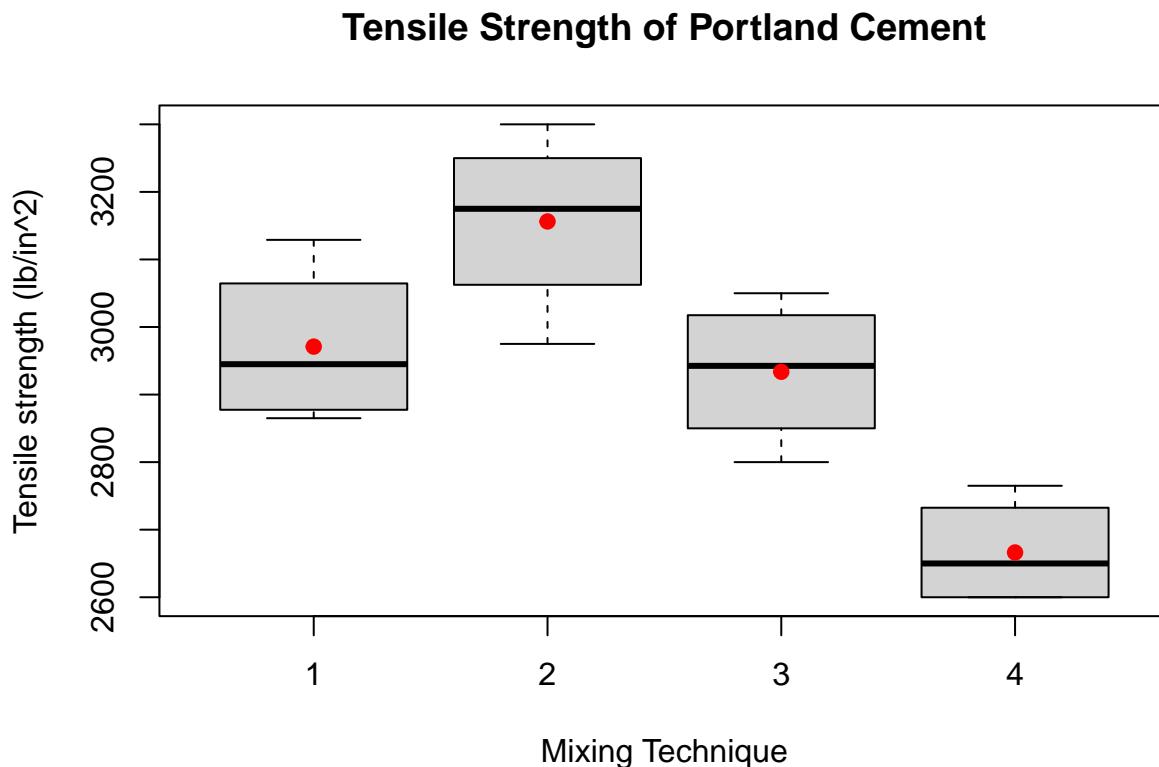
```
#With built-in R codes
cement_anova <- aov(tensile_strength~factor(mix_tech), data=cement)
summary(cement_anova)

##           Df Sum Sq Mean Sq F value    Pr(>F)
## factor(mix_tech)  3 489740 163247 12.73 0.000489 ***
## Residuals        12 153908   12826
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The chosen level of significance is alpha=0.05, which is greater than the p-value=0.000489, thus one rejects the hypothesis of equality of means/treatment effects.

QUESTION 3.9.b

```
#Boxplots
boxplot(tensile_strength~mix_tech, data=cement, main="Tensile Strength of Portland Cement",
         xlab="Mixing Technique", ylab="Tensile strength (lb/in^2)")
mean_tensile_strength <- tapply(tensile_strength, mix_tech, mean)
points(mean_tensile_strength, pch=19, col="red")
```



Clearly, one can see the difference in means among the treatment effects (mixing techniques).

QUESTION 3.9.c

```
#Refer to p.87 of the textbook
alpha <- 0.05 #significance level
N <- 16 #total number of observations
a <- 4 #total number of treatment effects / levels
n <- 4 #observations per treatment effect / level
MSE <- 12826
LSD <- qt(p=alpha, df=N-a, lower.tail=FALSE) * sqrt(2*MSE/n) #142.7277

#There are 4C2=6 pairs
LSD_procedure = c(
  abs(mean_tensile_strength[1]-mean_tensile_strength[2]),
  abs(mean_tensile_strength[1]-mean_tensile_strength[3]),
  abs(mean_tensile_strength[1]-mean_tensile_strength[4]),
  abs(mean_tensile_strength[2]-mean_tensile_strength[3]),
```

```

abs(mean_tensile_strength[2]-mean_tensile_strength[4]),
abs(mean_tensile_strength[3]-mean_tensile_strength[4])
)
LSD_procedure=LSD
##          1          1          1          2          2          3
##  42.52226 -105.47774 162.02226  79.77226 347.27226 124.77226

```

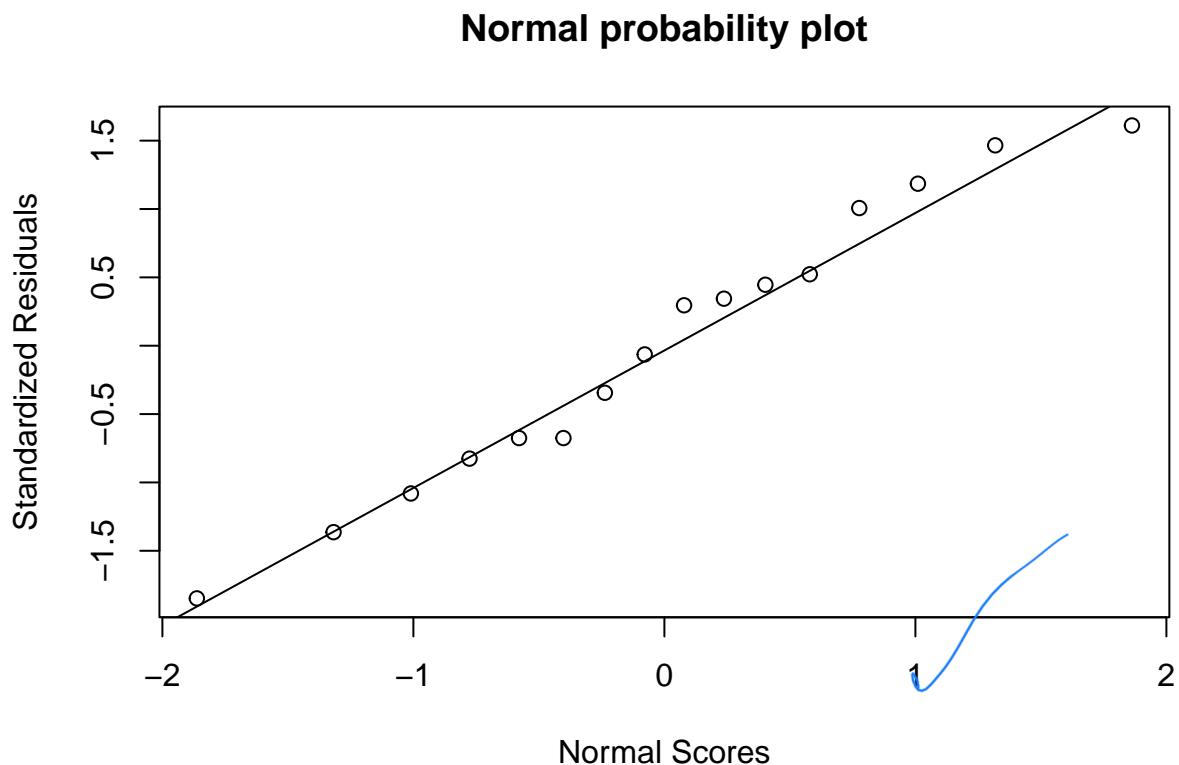
*#Any values above 0 imply the absolute value of difference in means between the corresponding pair of averages is greater than the Fisher LSD value.
#Thus only the means of the pair of y_{bar_1} . to y_{bar_3} . do not differ.*

QUESTION 3.9.d

```

par(mfrow=c(1,1))
cement_stdres = rstandard(cement_anova)
qqnorm(cement_stdres,
       ylab="Standardized Residuals",
       xlab="Normal Scores",
       main="Normal probability plot")
qqline(cement_stdres)

```

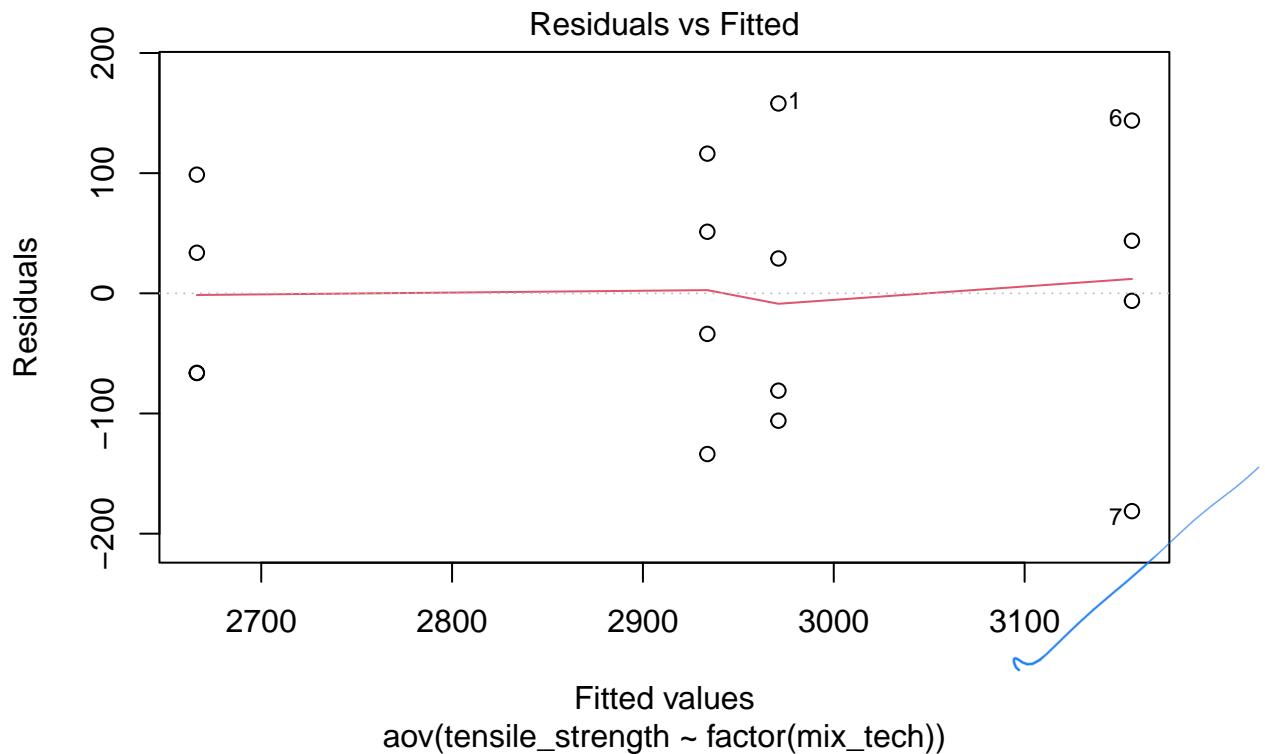


```
# NOTE: plot(cement_anova,2) achieves the same result
```

I conclude that the normality assumption is valid on account of the straight line one can draw through the residuals, and indeed on account of the relatively small amount of dispersion along the line.

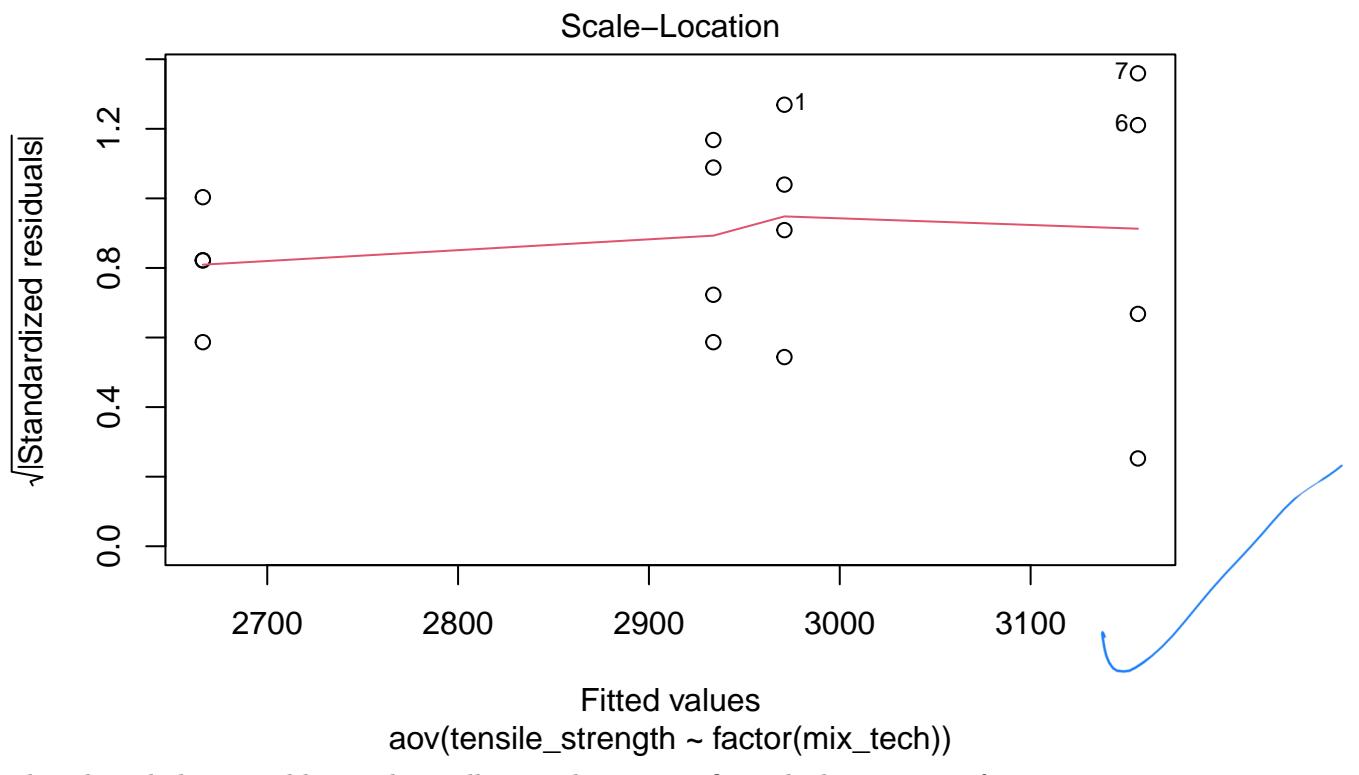
QUESTION 3.9.e

```
par(mfrow=c(1,1))
plot(cement_anova,1)
```



The plot is unstructured, and variance appears to be constant. Note homogeneity of variance can be checked with the “scale-location plot” or “spread-location plot”, which I will include below.

```
par(mfrow=c(1,1))
plot(cement_anova,3)
```

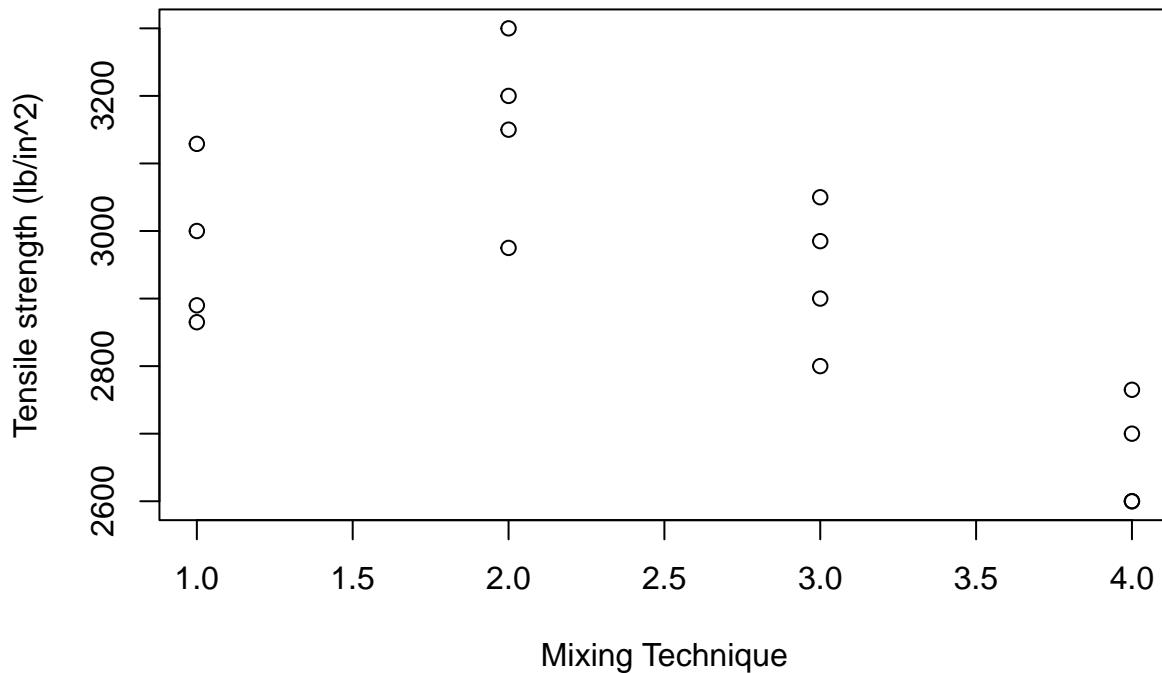


The relatively-horizontal line with equally spread points confirms the homogeneity of variance.

QUESTION 3.9.f *Comments?* - O.S

```
#Scatterplot
par(mfrow=c(1,1))
plot(tensile_strength~mix_tech, data=cement,
      main="Tensile Strength of Portland Cement",
      xlab="Mixing Technique", ylab="Tensile strength (lb/in^2)")
```

Tensile Strength of Portland Cement



QUESTION 3.10.a

2/2

Note Tukey's Test is on p.85 of the textbook.

```

alpha=0.05
MSE=12826
n=4
studentized_range_statistic=(max(mean_tensile_strength)-min(mean_tensile_strength)) / sqrt(MSE/n) #8.65
q_alpha_a_f=4.20 # Appendix Table V gives q_alpha=0.05(a=4,f=DFE=12)=4.20
T_a=q_alpha_a_f*sqrt(MSE/n)

#There are 4C2=6 pairs
Tukey_procedure = c(
  abs(mean_tensile_strength[1]-mean_tensile_strength[2]),
  abs(mean_tensile_strength[1]-mean_tensile_strength[3]),
  abs(mean_tensile_strength[1]-mean_tensile_strength[4]),
  abs(mean_tensile_strength[2]-mean_tensile_strength[3]),
  abs(mean_tensile_strength[2]-mean_tensile_strength[4]),
  abs(mean_tensile_strength[3]-mean_tensile_strength[4]))
)
Tukey_procedure-T_a

##          1         1         1         2         2         3
## -52.57906 -200.57906  66.92094 -15.32906 252.17094 29.67094

```

#Any values above 0 imply the two means compared are significantly different.

Compared to the Fisher LSD method, it appears that Tukey's Test is less sensitive, as it only noticed differences in means between treatment effects 1 to 4, 2 to 4, and 3 to 4. It missed the differences in means between treatment effects 1 to 2, and 2 to 3, which are also graphically obvious.

QUESTION 3.10.b

Whereas the Tukey test does not control the “overall error rate”, the Fisher LSD method controls the “error rate alpha for each individual pairwise comparison”.

QUESTION 3.11

25/25

Let us first construct 95% confidence intervals on the mean tensile strengths for each mixing technique. Procedure taken from textbook section 3.3.3 (Estimation of Model Parameters) on p.66-67.

```
alpha=0.05
N=16
a=4
MSE=12826

#For mixing technique 1
mean_tensile_strength[1]+c(-1,1)*qt(p=alpha/2, df=N-a, lower.tail=FALSE) * sqrt(MSE/n)

## [1] 2847.623 3094.377

#For mixing technique 2
mean_tensile_strength[2]+c(-1,1)*qt(p=alpha/2, df=N-a, lower.tail=FALSE) * sqrt(MSE/n)

## [1] 3032.873 3279.627

#For mixing technique 3
mean_tensile_strength[3]+c(-1,1)*qt(p=alpha/2, df=N-a, lower.tail=FALSE) * sqrt(MSE/n)

## [1] 2810.373 3057.127

#For mixing technique 4
mean_tensile_strength[4]+c(-1,1)*qt(p=alpha/2, df=N-a, lower.tail=FALSE) * sqrt(MSE/n)

## [1] 2542.873 2789.627
```

Let us second construct a 95% confidence interval on the difference in means for techniques 1 and 3 using the procedure from textbook section 3.3.3 (Estimation of Model Parameters) on p.66-67.

```
alpha=0.05
N=16
a=4
MSE=12826
mean_tensile_strength[1]-mean_tensile_strength[3]+c(-1,1)*qt(p=alpha/2, df=N-a, lower.tail=FALSE) * sqrt(MSE/n)
```

```
## [1] -137.232 211.732
```

Let us again construct a 95% confidence interval on the difference in means for techniques 1 and 3, but now we employ the results from 3.10.a according to textbook p.85 (3.5.7 Comparing Pairs of Treatment Means).

```
(mean_tensile_strength[1]-mean_tensile_strength[3])+c(-1,1)*T_a
```

```
## [1] -200.5791 275.0791
```

These two confidence intervals in which we compare the difference in means for techniques 1 and 3 do indeed aid in the interpretation of results: Since both cross zero, it can be concluded there is no statistically significant difference between the means corresponding to the two techniques.

QUESTION 3.24

3/3

```
fluid_type <- c(  
  seq(1,1,length.out=6),  
  seq(2,2,length.out=6),  
  seq(3,3,length.out=6),  
  seq(4,4,length.out=6)  
)  
  
life_at_35kV_load <- c(  
  17.6,18.9,16.3,17.4,20.1,21.6,  
  16.9,15.3,18.6,17.1,19.5,20.3,  
  21.4,23.6,19.4,18.5,20.5,22.3,  
  19.3,21.1,16.9,17.5,18.3,19.8  
)  
  
fluids <- data.frame(fluid_type,life_at_35kV_load)  
str(fluids)  
  
## 'data.frame': 24 obs. of 2 variables:  
## $ fluid_type : num 1 1 1 1 1 1 2 2 2 2 ...  
## $ life_at_35kV_load: num 17.6 18.9 16.3 17.4 20.1 21.6 16.9 15.3 18.6 17.1 ...
```

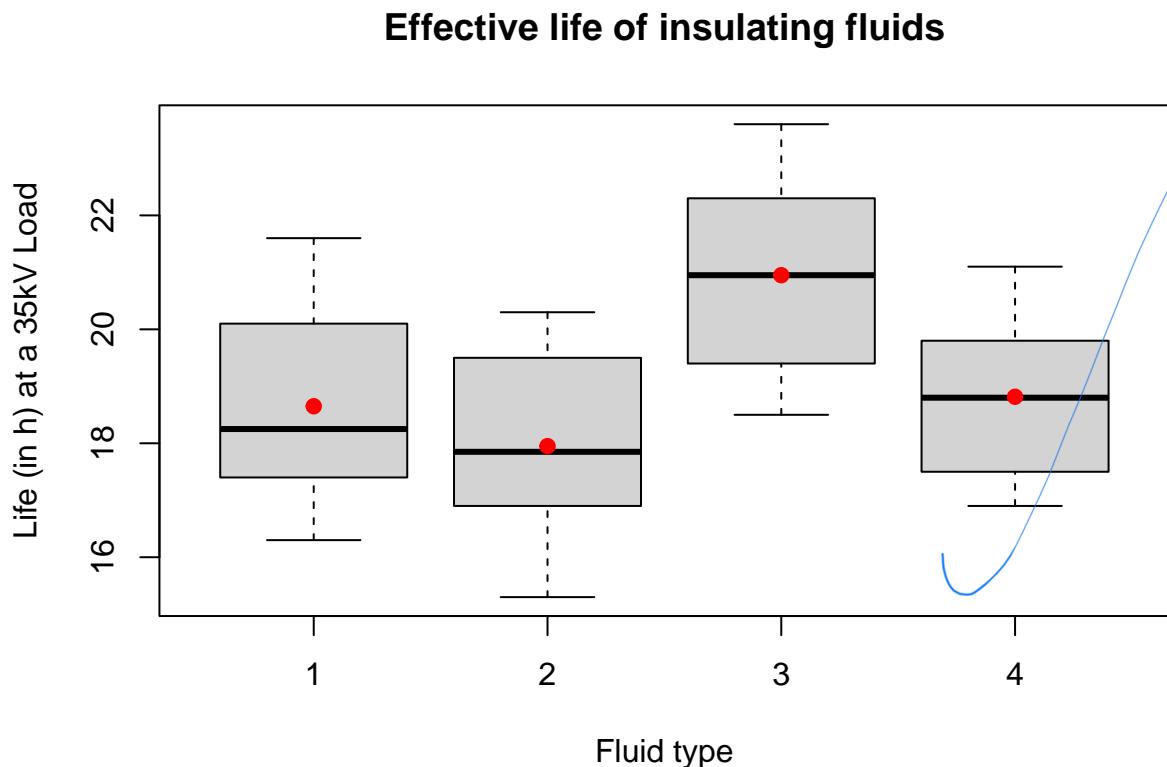
QUESTION 3.24.a

```
fluids_anova <- aov(life_at_35kV_load~factor(fluid_type), data=fluids)  
summary(fluids_anova)  
  
## Df Sum Sq Mean Sq F value Pr(>F)  
## factor(fluid_type) 3 30.17 10.05 3.047 0.0525 .  
## Residuals 20 65.99 3.30  
## ---  
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ',' 1
```

At a level of significance of alpha=0.05, according to the ANOVA table, there is no difference the fluids differ in regard to their life (in h) at a 35kV load.

QUESTION 3.24.b

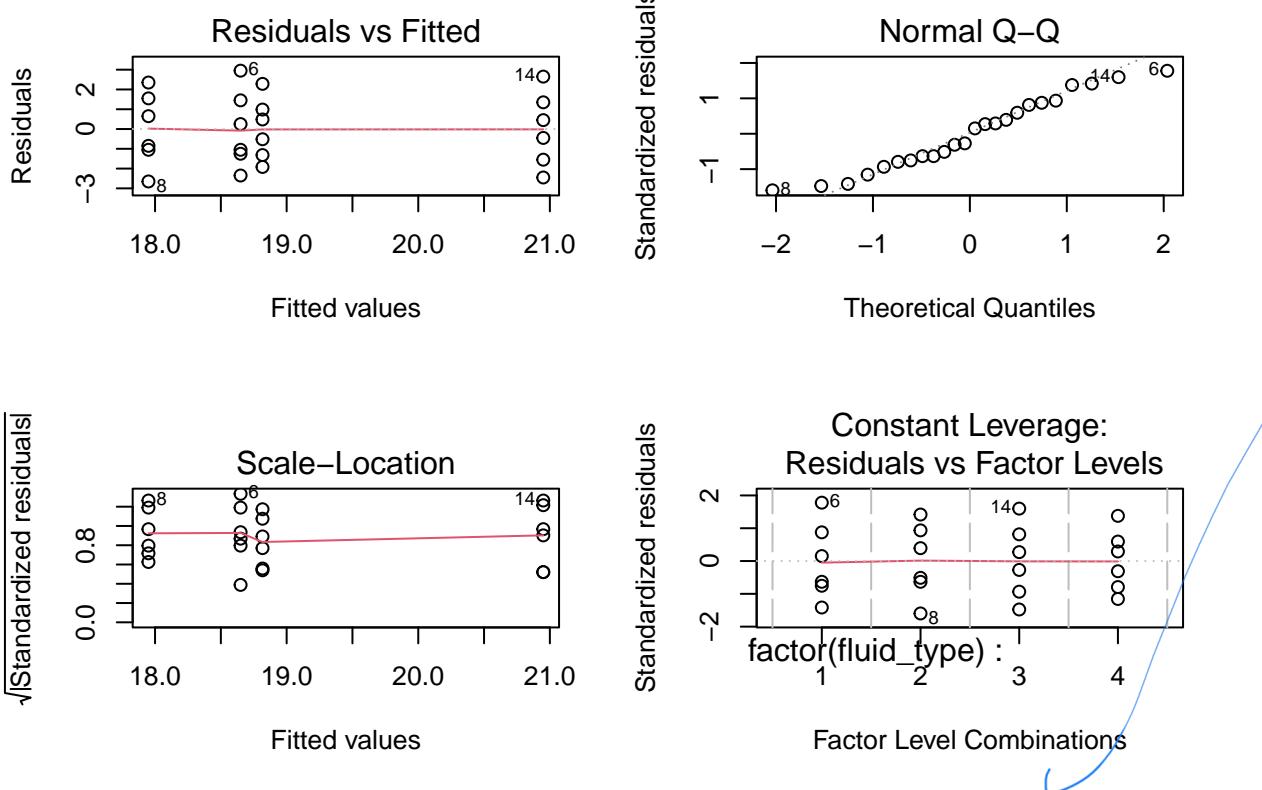
```
#Boxplots
boxplot(life_at_35kV_load~fluid_type, data=fluids,
        main="Effective life of insulating fluids",
        xlab="Fluid type", ylab="Life (in h) at a 35kV Load")
mean_fluid_life <- tapply(life_at_35kV_load, fluid_type, mean)
points(mean_fluid_life, pch=19, col="red")
```



Although the interpretation of the ANOVA table, taken at a significance level of alpha=0.05, refuted the difference in means of the fluid types, the boxplots suggest fluid type #3 may last the longest.

QUESTION 3.24.c

```
par(mfrow=c(2,2))
plot(flights_anova)
```



Anova assumptions: "Errors are iid $\sim \text{Normal}(\text{mean}=0, \text{variance}=\sigma^2)$ ".

What the diagnostic plots tell me: 1. Normal Q-Q: residuals are normally distributed, which I know because they more or less follow the diagonal line, with more attention to residuals in the middle and less attention to residuals at the ends. 2. Scale-Location: Residuals residuals follow the horizontal line are are equally spread out (homogeneity of variance), which also is a strong argument for homoscedasticity.

QUESTION 3.52

The Kruskal-Wallis test is located on p. 113 (3.11.1) of the textbook. It test for H_0 : equality of treatment means against H_a : inequality of treatment means for at least one $i \neq j$. It is used when normality cannot be assumed.

```
kruskal.test(life_at_35kV_load~factor(fluid_type), data=fluids)
```

```
##  
##  Kruskal-Wallis rank sum test  
##  
##  data:  life_at_35kV_load by factor(fluid_type)  
##  Kruskal-Wallis chi-squared = 6.2177, df = 3, p-value = 0.1015
```

At a level of significance of $\alpha=0.05$, according to the test, with $p=0.1015$, there is no difference in that the fluids differ in regard to their life (in h) at a 35kV load, which matches the conclusion found with the ANOVA table.

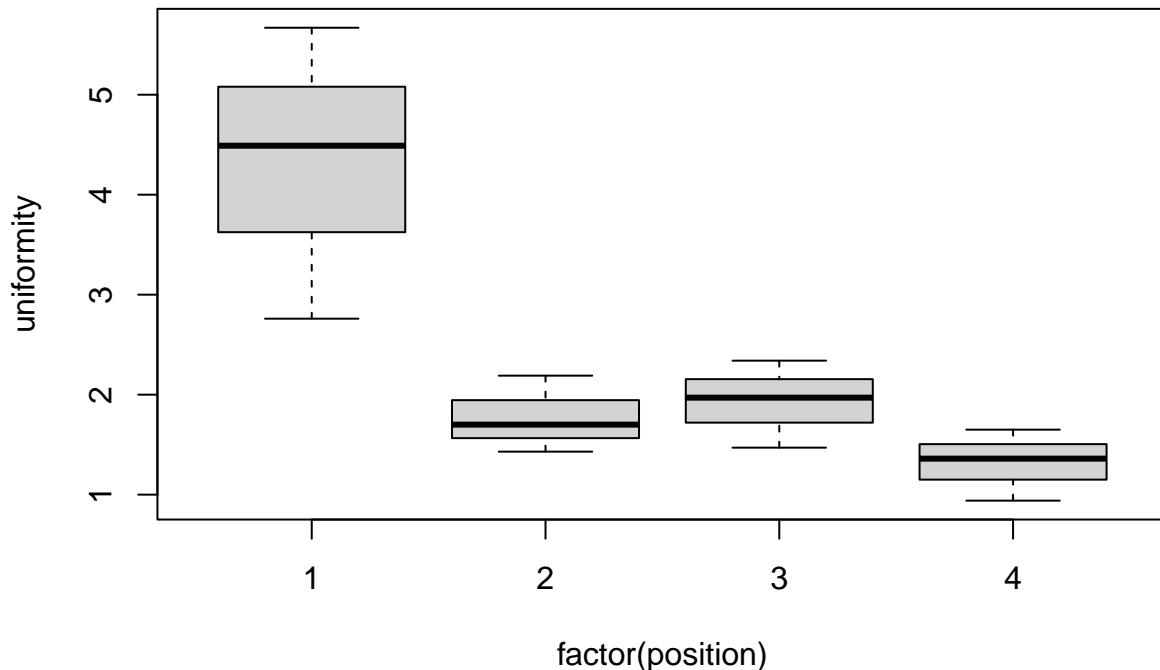
QUESTION 3.31

```
position = c(1,1,1,
           2,2,2,
           3,3,3,
           4,4,4)

uniformity = c(2.76, 5.67, 4.49,
              1.43, 1.70, 2.19,
              2.34, 1.97, 1.47,
              0.94, 1.36, 1.65)

wafers = data.frame(position, uniformity)

boxplot(uniformity~factor(position), data=wafers)
```



QUESTION 3.31.a

```
wafers_anova=aov(uniformity~factor(position), data=wafers)
summary(wafers_anova)
```

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

```

## factor(position) 3 16.220 5.407     8.29 0.00775 **
## Residuals       8  5.217  0.652
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

At a significance of alpha=0.05, the p-value from the ANOVA table indicates rejection of the equality of treatment means, thus there is a difference in wafer positions between at least two levels.

QUESTION 3.31.b

The estimation of the variability due to wafer positions is MSE=0.652.

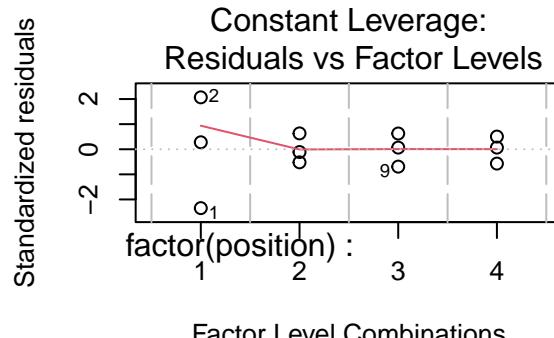
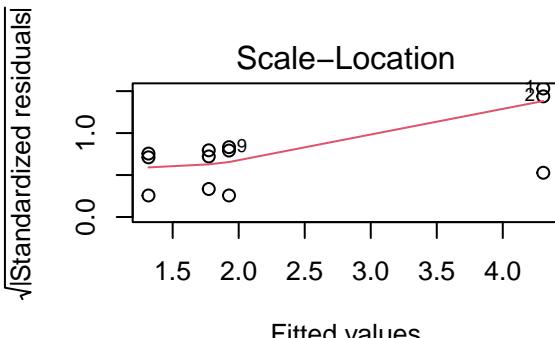
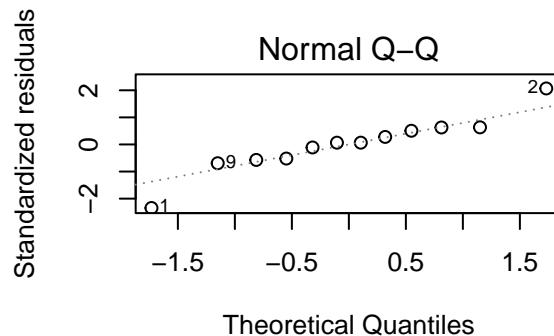
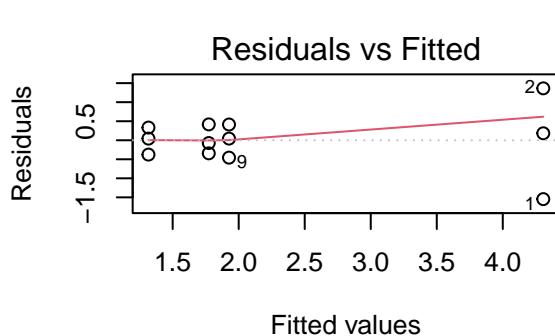
QUESTION 3.31.c

QUESTION 3.31.d

```

par(mfrow=c(2,2))
plot(wafers_anova)

```



1. Residuals vs fitted is roughly linear, thus data is roughly linear.
2. Normal Q-Q plot: data follow the diagonal line except for the two points at opposite extremes.
3. Scale-Location plot: line obviously is not horizontal, thus cannot assume homogeneity of variance (homoscedasticity).

QUESTION 3.32.a

```
Q  
SSTreatment = 16.220  
SSError = 5.217  
SSTotal = SSTreatment + SSError  
X  
DFtotal = 3+8 #11  
  
MSTotal = SSTotal/DFtotal  
MSTotal
```

```
## [1] 1.948818
```

Estimate in the total variability of the uniformity response is MSTotal = 1.948818

QUESTION 3.32.b

QUESTION 3.32.c

QUESTION 3.33.a

```
V  
#Clear variables  
rm(list = ls())  
a=4  
N=3*4 #12  
SSTotal=330.56  
SSTreatment=250.65  
SSError=(SSTotal-SSTreatment)  
MSE=SSError/(N-a)  
MSE
```

```
## [1] 9.98875
```

The MSE (9.98875) is the estimate of the error variance.

QUESTION 3.33.b

```
R_squared = SSTreatment/SSTotal  
X  
R_squared
```

```
## [1] 0.7582587
```

About 75.8% of the variability in the response variable is explained by the treatment effect.

QUESTION 3.35

```

# see p106 of the textbook

alpha=0.05

a = 4 #four levels
n = 3 #three replicates

N = 12

SSTreatment = 250.65
SSTotal = 330.56

SSError = SSTotal-SSTreatment #79.91

DFtreatment = a-1 #3
DFerror = N-a #8
DFtotal = N-1 #11

MStreatment = SSTreatment/DFtreatment
MSerror = SSError/DFerror

F_lower = qf(1-alpha/2, a-1, N-a, lower.tail = TRUE)
L = (1/n)*((MStreatment/MSerror)*(1/F_lower)-1)

F_upper = qf(alpha/2, a-1, N-a, lower.tail = TRUE)
U = (1/n)*((MStreatment/MSerror)*(1/F_upper)-1)

lower_limit = L/(1+L)
upper_limit = U/(1+U)

lower_limit

## [1] 0.1535943

upper_limit

## [1] 0.9757316

```