Lab2

Jas Sur

2/07/2022

Q.1

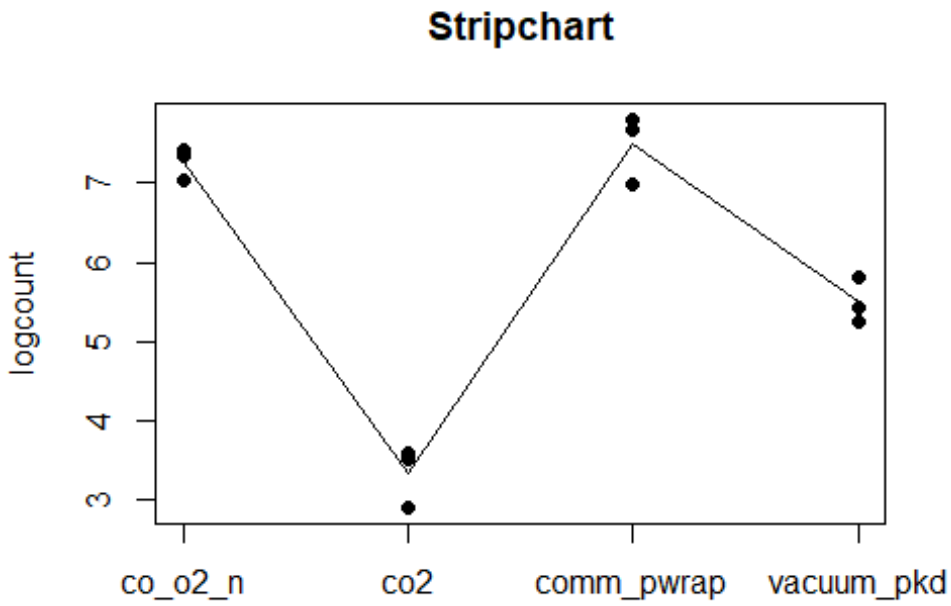
a. Set up the data frame.

```
commercial_plastic_wrap <- c(7.66, 6.98, 7.80)
vacuum_packaged <- c(5.26, 5.44, 5.80)

#1% CO,40% O2, 59% N
co_o2_n<- c(7.41, 7.33, 7.04)
co2_100_percent <-c(3.51, 2.91, 3.6)

logcount <- c(commercial_plastic_wrap, vacuum_packaged, co_o2_n,
co2_100_percent)

package <- rep(c("comm_pwrap", "vacuum_pkd", "co_o2_n", "co2"),
               each=3)
package <-factor(package)
bacteria  <-data.frame(cbind(logcount, package))
head(bacteria)
```

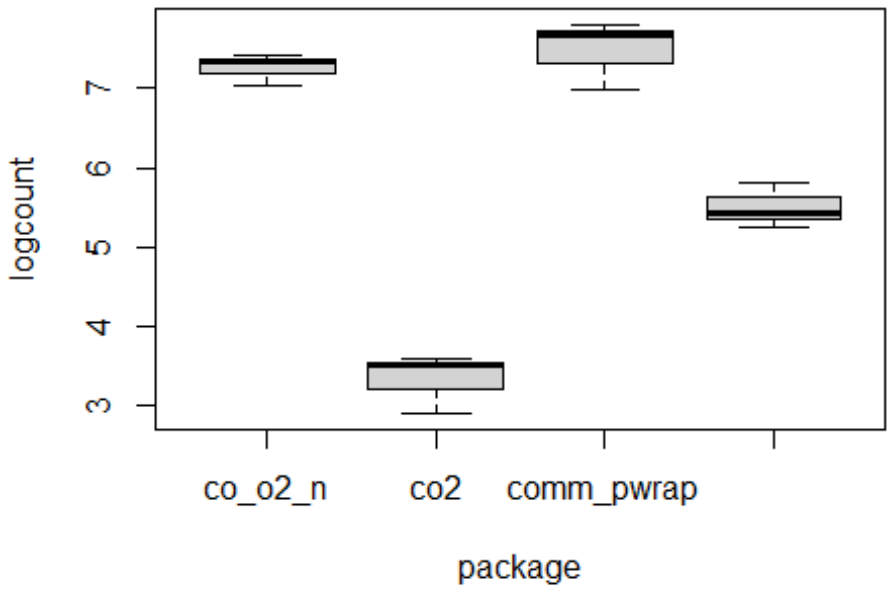
```
##   logcount package
## 1     7.66       3
## 2     6.98       3
## 3     7.80       3
## 4     5.26       4
## 5     5.44       4
## 6     5.80       4
```

b. Perform a stripchart, with line connecting means, of logcount vs package

```
stripchart(logcount~package, vertical=TRUE, pch=16, main="Stripchart")

logcount.means<-tapply(logcount, package, mean)
lines(logcount.means)
```




```
boxplot(logcount~package)
```



It looks like the co-o2-n and common_package_wrap has the same log count.

c. Build a linear model, using aov() response as logcount. Do a summary.lm()

```
model<- aov(logcount ~ package)
summary.aov(model)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## package      3  33.18  11.060   99.06 1.15e-06 ***
## Residuals    8   0.89   0.112
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

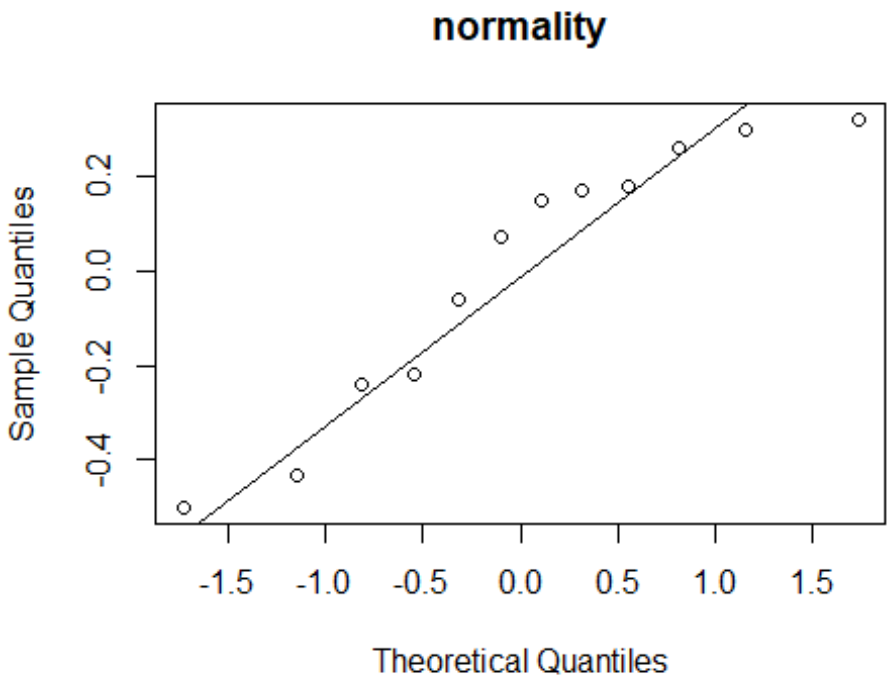
```
summary.lm(model)
```

```
##
## Call:
## aov(formula = logcount ~ package)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -0.500 -0.225  0.110  0.200  0.320
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)         7.2600     0.1929  37.633 2.73e-10 ***
## packageco2         -3.9200     0.2728 -14.368 5.38e-07 ***
## packagecomm_pwrap   0.2200     0.2728   0.806 0.443333
## packagevacuum_pkd  -1.7600     0.2728  -6.451 0.000198 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3341 on 8 degrees of freedom
## Multiple R-squared:  0.9738, Adjusted R-squared:  0.964
## F-statistic: 99.06 on 3 and 8 DF,  p-value: 1.15e-06
```

pvalue is smaller than 0.05 so it's evident that the means of some groups is not the same

d. Perform a Bartlett test of equal variances.

```
logcount.aov<- aov(model)
res <- residuals(logcount.aov)
qqnorm(res, main="normality")
qqline(res)
```



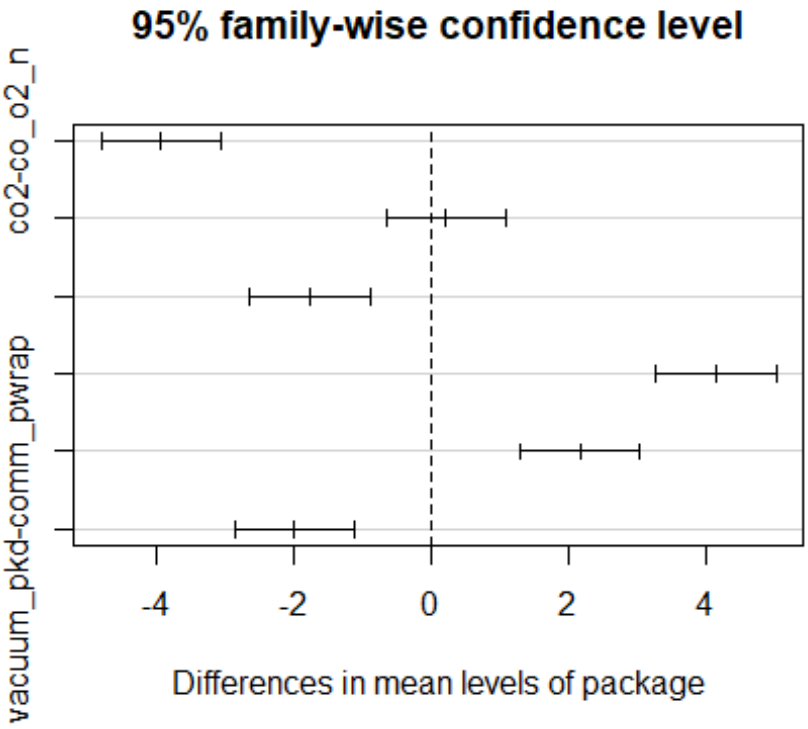

```
bartlett.test(res~package)

##
##  Bartlett test of homogeneity of variances
##
## data:  res by package
## Bartlett's K-squared = 1.1501, df = 3, p-value = 0.765
```

The variance is same for different groups

e. Perform a multiple comparison of treatment mean, using TukeyHSD()

```
tuskey <-TukeyHSD(model, "package")
tuskey

##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = logcount ~ package)
##
## $package
##                          diff        lwr        upr     p adj
## co2-co_o2_n             -3.92 -4.7936808 -3.0463192 0.0000026
## comm_pwrap-co_o2_n       0.22 -0.6536808  1.0936808 0.8497833
## vacuum_pkd-co_o2_n      -1.76 -2.6336808 -0.8863192 0.0008974
## comm_pwrap-co2           4.14  3.2663192  5.0136808 0.0000017
## vacuum_pkd-co2           2.16  1.2863192  3.0336808 0.0002167
## vacuum_pkd-comm_pwrap   -1.98 -2.8536808 -1.1063192 0.0004001
```

```
plot(tuskey)
```



Q.2. Data: Tensile strength of Portland Cement

a. Set up a data frame , with varibles: mixing (factor) and strength (response)

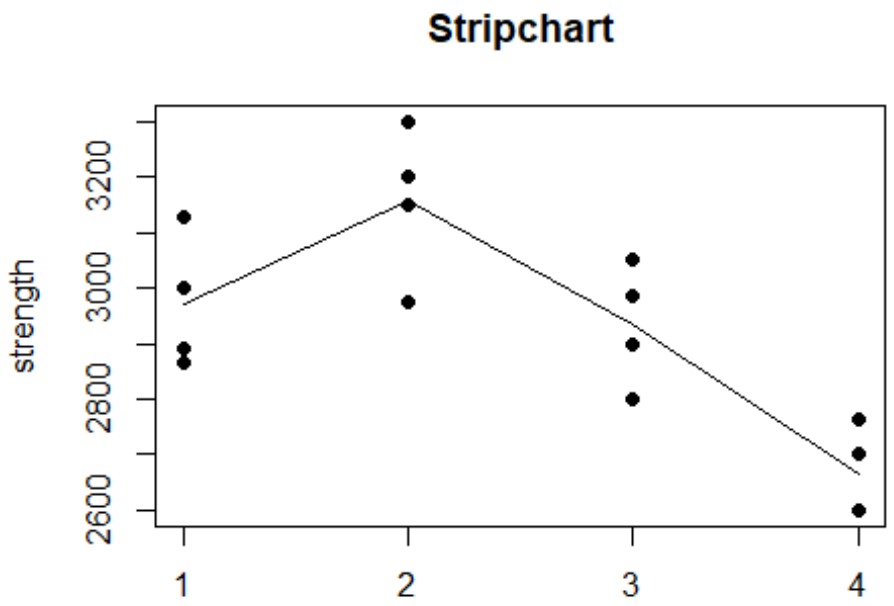
```
one <- c(3129, 3000, 2865, 2890)
two <- c(3200, 3300, 2975, 3150)
three <- c(2800, 2900, 2985, 3050)
four <- c(2600, 2700, 2600, 2765)
strength <- c(one, two, three, four)

material <- rep(c("1", "2", "3", "4"), each=4)
material <-factor(material)
cement  <-data.frame(cbind(strength, material))
head(cement)
```

```
##   strength material
## 1     3129        1
## 2     3000        1
## 3     2865        1
## 4     2890        1
## 5     3200        2
## 6     3300        2
```

b. Perform a stripchart. Perform a Box plot.

```
stripchart(strength~material, vertical=TRUE, pch=16, main="Stripchart")
```

```
strength.means<-tapply(strength, material, mean)
lines(strength.means)
```



c. Test the hypothesis that mixing techniques affect the strength of the cement. Use α=0.05

What test do use. Perform the test. Conclusion.

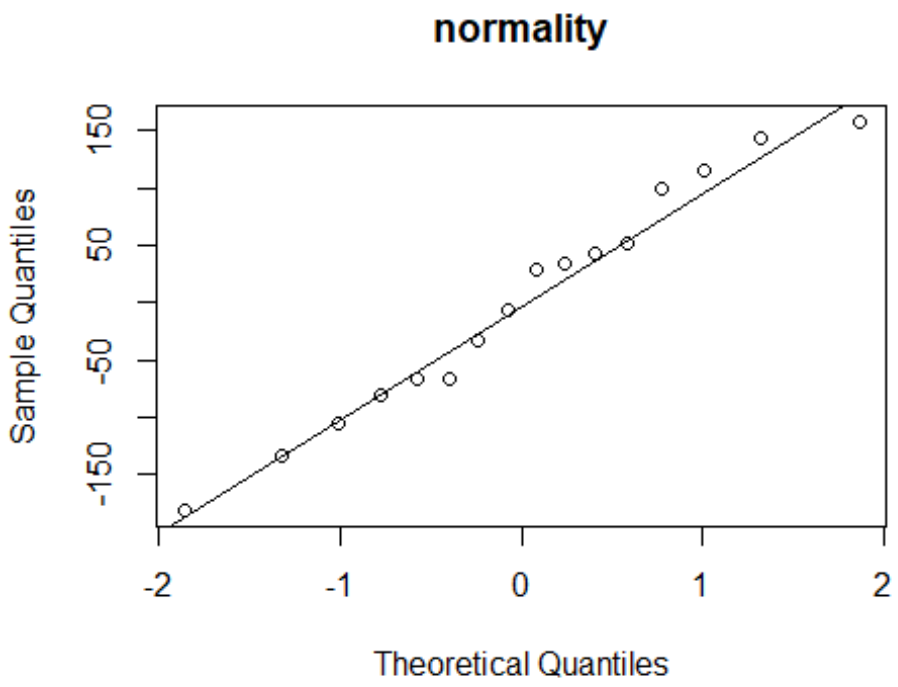
H0 = mixing techniques doesn't affect the strength of material Ha = mixing techniques affect the strength of material

assumptions > data is normal

```
strength.aov.model<- aov(strength ~ material)
summary.aov(strength.aov.model)

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

summary.lm(strength.aov.model)

##
## Call:
## aov(formula = strength ~ material)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -181.25  -69.94   11.38   63.12  158.00
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2971.00      56.63  52.468 1.51e-15 ***
## material2     185.25      80.08   2.313   0.0392 *
## material3     -37.25      80.08  -0.465   0.6501
## material4    -304.75      80.08  -3.806   0.0025 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 113.3 on 12 degrees of freedom
## Multiple R-squared:  0.7609, Adjusted R-squared:  0.7011
## F-statistic: 12.73 on 3 and 12 DF,  p-value: 0.0004887
```

```
res <- residuals(strength.aov.model)
qqnorm(res, main="normality")
qqline(res)
```




```
shapiro.test(res)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res
## W = 0.97046, p-value = 0.846
```

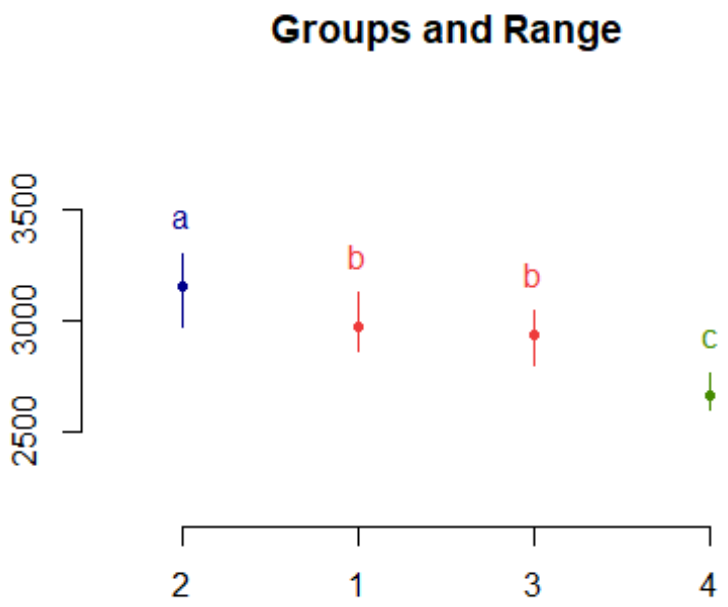
```
#check variance
bartlett.test(res~material)

##
##  Bartlett test of homogeneity of variances
##
## data:  res by material
## Bartlett's K-squared = 0.71158, df = 3, p-value = 0.8705
```

pvalue is greater than alpha (0.05), reject null hypothesis, accept alternative hypothesis that material does affect the strength

d. Use the Fisher LSD (Least Significant Difference) α = 0.05 to make comparison

```
#install.packages("agricolae")
library(agricolae)

## Warning: package 'agricolae' was built under R version 4.0.5

MSerror <- 12826
Fisher<- LSD.test(strength.aov.model, "material", MSerror, console=T)

##
## Study: strength.aov.model ~ "material"
##
## LSD t Test for strength
##
## Mean Square Error:  12825.69
##
## material,  means and individual ( 95 %) CI
##
##   strength       std r      LCL      UCL  Min  Max
## 1  2971.00 120.55704 4 2847.624 3094.376 2865 3129
## 2  3156.25 135.97641 4 3032.874 3279.626 2975 3300
## 3  2933.75 108.27242 4 2810.374 3057.126 2800 3050
## 4  2666.25  80.97067 4 2542.874 2789.626 2600 2765
##
## Alpha: 0.05 ; DF Error: 12
## Critical Value of t: 2.178813
##
## least Significant Difference: 174.4798
##
## Treatments with the same letter are not significantly different.
##
##   strength groups
## 2  3156.25      a
## 1  2971.00      b
## 3  2933.75      b
## 4  2666.25      c

Fisher
```

```
## $statistics
##    MSerror Df     Mean       CV  t.value      LSD
##   12825.69 12 2931.812 3.862817 2.178813 174.4798
##
## $parameters
##         test p.ajusted   name.t ntr alpha
##   Fisher-LSD      none material   4  0.05
##
## $means
##   strength       std r      LCL      UCL  Min  Max     Q25    Q50     Q75
## 1  2971.00 120.55704 4 2847.624 3094.376 2865 3129 2883.75 2945.0 3032.25
## 2  3156.25 135.97641 4 3032.874 3279.626 2975 3300 3106.25 3175.0 3225.00
## 3  2933.75 108.27242 4 2810.374 3057.126 2800 3050 2875.00 2942.5 3001.25
## 4  2666.25  80.97067 4 2542.874 2789.626 2600 2765 2600.00 2650.0 2716.25
##
## $comparison
## NULL
##
## $groups
##   strength groups
## 2  3156.25      a
## 1  2971.00      b
## 3  2933.75      b
## 4  2666.25      c
##
## attr(,"class")
## [1] "group"

plot(Fisher)
```



Only one group has pvalue > 0.05 and rest of the groups has pvalue < 0.05, so the means for each of other groups is not the same comm_pwrap-co_o2_n 0.22 -0.6536808 1.0936808 0.8497833

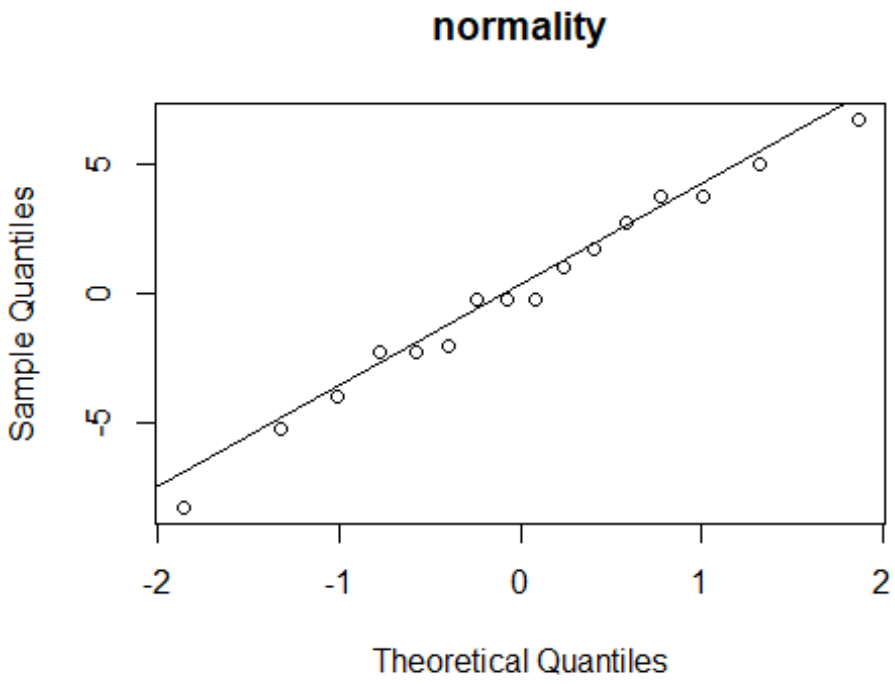
Q 3

```
one <- c(143, 141, 150, 146)
two <- c(152, 149, 137, 143)
three <- c(134, 136, 132, 127)
four <- c(129, 127, 132, 129)

conductivity <- c(one, two, three, four)

coating <- rep(c("1", "2", "3", "4"), each=4)
coating <-factor(coating)
television  <- data.frame(cbind(conductivity, coating))
head(television)

##   conductivity coating
## 1          143       1
## 2          141       1
## 3          150       1
## 4          146       1
## 5          152       2
## 6          149       2
```

a. Is there a difference in conductivity due to coating type?

alpha = 0.05

```
cond.model<- aov(conductivity ~ coating)
summary.aov(cond.model)

##             Df Sum Sq Mean Sq F value   Pr(>F)
## coating      3  844.7  281.56    14.3 0.000288 ***
## Residuals   12  236.3   19.69
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary.lm(cond.model)

##
## Call:
## aov(formula = conductivity ~ coating)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -8.25  -2.25  -0.25   3.00   6.75
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  145.000      2.219  65.359  < 2e-16 ***
## coating2       0.250      3.138   0.080 0.937804
## coating3     -12.750      3.138  -4.064 0.001571 **
## coating4     -15.750      3.138  -5.020 0.000299 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.437 on 12 degrees of freedom
## Multiple R-squared:  0.7814, Adjusted R-squared:  0.7268
## F-statistic:  14.3 on 3 and 12 DF,  p-value: 0.0002881

res <- residuals(cond.model)
qqnorm(res, main="normality")
qqline(res)
```




```
shapiro.test(res)

##
##  Shapiro-Wilk normality test
##
## data:  res
## W = 0.98422, p-value = 0.9882
```

pvalue is greater than 0.05,so it's ascertained that the data is normal

compare the variance > H0 = there is no difference in conductivity var1 = var2 > Ha = there is a difference in conductivity var1 <> var2

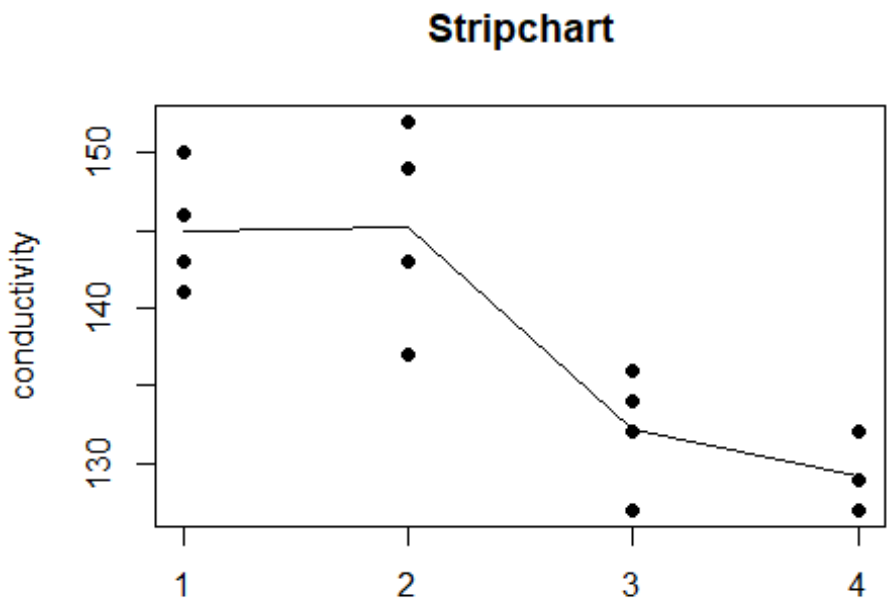
```
bartlett.test(res~coating)

##
##  Bartlett test of homogeneity of variances
##
## data:  res by coating
## Bartlett's K-squared = 3.2944, df = 3, p-value = 0.3484
```

pvalue is greater than 0.05, so null hypothesis is accepted and it's ascertained the variance is same

Since the data is normal and variance is equal, it may be concluded that there is no difference in conductivity due to coating type

b. Estimate the mean and the treatment effects view the model output

```
stripchart(conductivity~coating, vertical=TRUE, pch=16, main="Stripchart")

cond.means<-tapply(conductivity, coating, mean)
lines(cond.means)
```



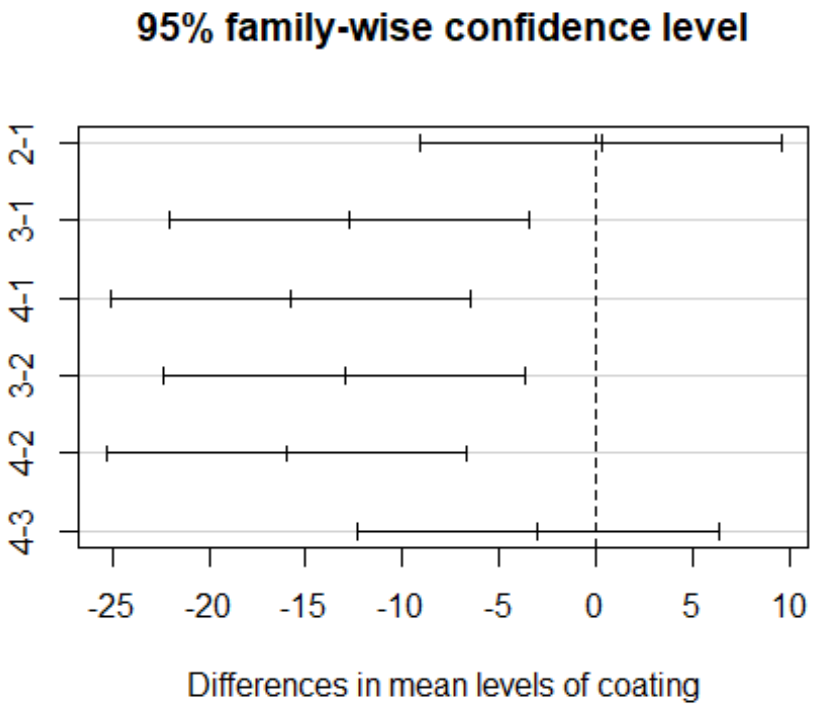
It can be seen from the stripchart that coating type of 1,2 have the same conductivity while 3, 4 have another group of same conductivity, but doesn't match 1,2

The pvalue from aov test is smaller than 0.05, so we have sufficient evidence to say that one of the means is different from the others.

```
tukey.95 <- TukeyHSD(cond.model, "coating")

tukey.95
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = conductivity ~ coating)
##
## $coating
##       diff        lwr       upr     p adj
## 2-1   0.25  -9.064853  9.564853 0.9998078
## 3-1 -12.75 -22.064853 -3.435147 0.0073964
## 4-1 -15.75 -25.064853 -6.435147 0.0014707
## 3-2 -13.00 -22.314853 -3.685147 0.0064441
```

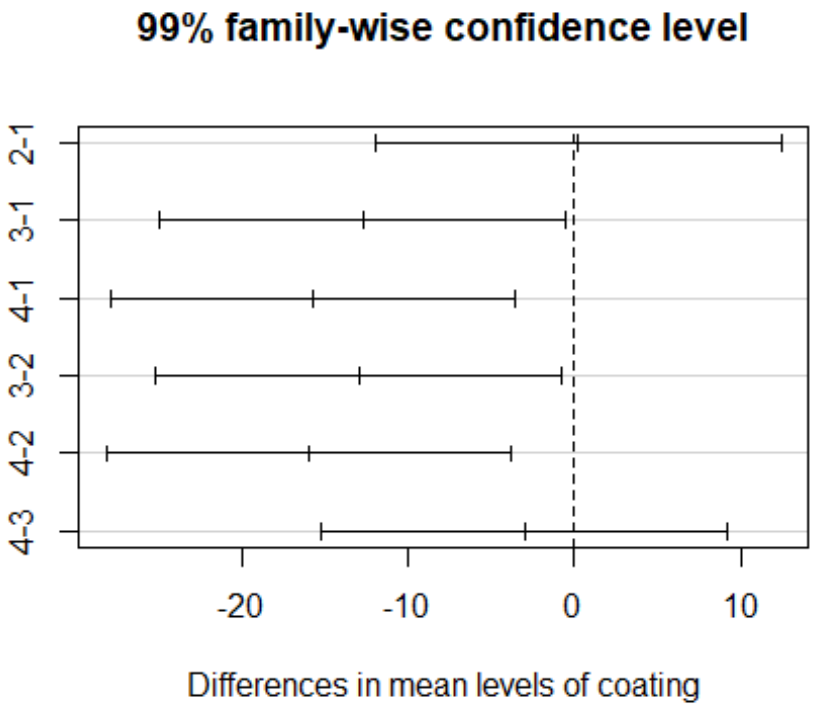
```
## 4-2 -16.00 -25.314853 -6.685147 0.0012913
## 4-3  -3.00 -12.314853  6.314853 0.7759360

plot(tukey.95)
```



For coating type 4, the confidence intervals for the mean value between groups 4-2 and 4-1 contain the value zero, which indicates that there is a statistically significant difference in mean loss between the two groups. This is consistent with the fact that two of these groups of the p-values from our hypothesis tests are below 0.05.

```
tukey.99 <- TukeyHSD(cond.model, "coating", conf.level = 0.99)
tukey.99

##   Tukey multiple comparisons of means
##     99% family-wise confidence level
##
## Fit: aov(formula = conductivity ~ coating)
##
## $coating
##       diff       lwr        upr     p adj
## 2-1   0.25 -11.95552 12.4555225 0.9998078
## 3-1 -12.75 -24.95552 -0.5444775 0.0073964
## 4-1 -15.75 -27.95552 -3.5444775 0.0014707
## 3-2 -13.00 -25.20552 -0.7944775 0.0064441
## 4-2 -16.00 -28.20552 -3.7944775 0.0012913
## 4-3  -3.00 -15.20552  9.2055225 0.7759360
```

```
plot(tukey.99)
```



For coating type 4, the confidence intervals for the mean value between groups 4-2 and 4-1 contain the value zero, which indicates that there is a statistically significant difference in mean loss between the two groups. This is consistent with the fact that two of these groups of the p-values from our hypothesis tests are below 0.05.

There is not much statistical difference between .95 and .99

d. Test all pairs of mean using Fisher LSD

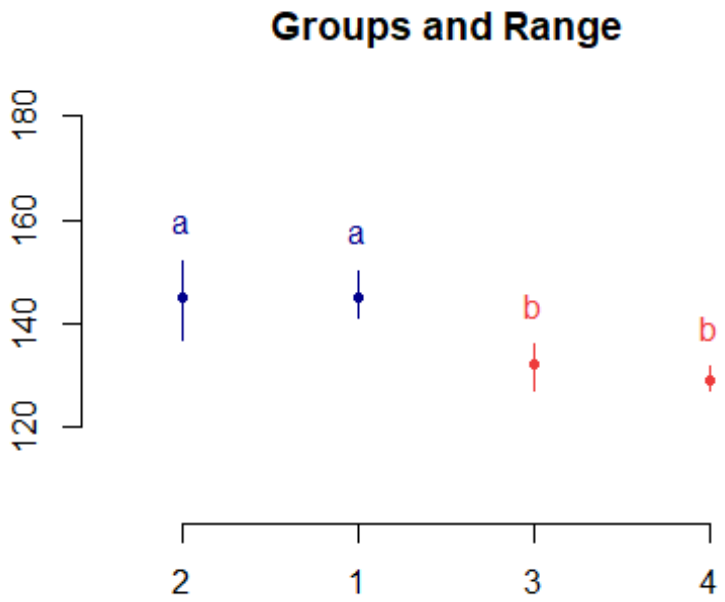
```
MSerror <-19.69
Fisher<- LSD.test(cond.model, "coating", MSerror, console=T)

##
## Study: cond.model ~ "coating"
##
## LSD t Test for conductivity
##
## Mean Square Error:  19.6875
##
## coating,  means and individual ( 95 %) CI
##
##   conductivity      std r      LCL      UCL Min Max
## 1       145.00 3.915780 4 140.1662 149.8338 141 150
## 2       145.25 6.652067 4 140.4162 150.0838 137 152
## 3       132.25 3.862210 4 127.4162 137.0838 127 136
## 4       129.25 2.061553 4 124.4162 134.0838 127 132
```

```
##
## Alpha: 0.05 ; DF Error: 12
## Critical Value of t: 2.178813
##
## least Significant Difference: 6.835971
##
## Treatments with the same letter are not significantly different.
##
##   conductivity groups
## 2       145.25      a
## 1       145.00      a
## 3       132.25      b
## 4       129.25      b
```

Fisher

```
## $statistics
##   MSerror Df     Mean       CV  t.value      LSD
##   19.6875 12 137.9375 3.216718 2.178813 6.835971
##
## $parameters
##         test p.ajusted  name.t ntr alpha
##   Fisher-LSD      none coating   4  0.05
##
## $means
##   conductivity      std r      LCL      UCL Min Max    Q25   Q50    Q75
## 1       145.00 3.915780 4 140.1662 149.8338 141 150 142.50 144.5 147.00
## 2       145.25 6.652067 4 140.4162 150.0838 137 152 141.50 146.0 149.75
## 3       132.25 3.862210 4 127.4162 137.0838 127 136 130.75 133.0 134.50
## 4       129.25 2.061553 4 124.4162 134.0838 127 132 128.50 129.0 129.75
##
## $comparison
## NULL
##
## $groups
##   conductivity groups
## 2       145.25      a
## 1       145.00      a
## 3       132.25      b
## 4       129.25      b
##
## attr(,"class")
## [1] "group"
```

```
plot(Fisher)
```



2,1 (group a) and 3,4 (group b) have same mean

e. Assuming that coating type 4 is currently in use, what are your recommendations to the manufacturer? We wish to minimize conductivity.

Since group 3 and 4 have the same conductivity, coating# 4 can be used interchangebly with coating# 3.