

# Lab3

Jas Sur

2/19/2022

## Q.1

1. Problem: ( Casting of High Temperature Alloys) A metal alloy is produced by a high temperature casting process. The tensile strength of the alloy is critical for its future use. The casting process is designed produce bars of alloy with an average tensile strength above minimum requirement. An experiment was planned to isolate the variation in tensile strength due to the effects of different castings. 30 bars or alloy were produced using the 3 casting methods.

Set up the data frame. Form a factor vector “methods”. Form a vector of response “strength”. Form a data frame named “alloy”.

```
one <- c(88.0, 88.0, 94.8, 90.8, 93.0, 89.0, 86.0, 92.9, 89.0, 93.0)
two <- c(88.0, 88.0, 94.8, 90.8, 93.0, 89.0, 86.0, 92.9, 89.0, 93.0)
three <- c(94.2, 91.5, 92.0, 96.5, 95.6, 93.8, 92.5, 93.2, 96.2, 92.5)
strength <- c(one, two, three)
methods <- rep(c("one", "two", "three"), each=10)
methods <- as.factor(methods)
alloy <- data.frame(cbind(strength, methods))
str(alloy)
```

```
## 'data.frame': 30 obs. of 2 variables:
## $ strength: num 88 88 94.8 90.8 93 89 86 92.9 89 93 ...
## $ methods : num 1 1 1 1 1 1 1 1 1 1 ...
```

### a. Test the equal variance among treatments, using bartlett.test()

$H_0$  = there is no difference in conductivity  $\text{var1} = \text{var2}$   $H_a$  = there is a difference in conductivity  
 $\text{var1} <> \text{var2}$

```
bartlett.test(strength~methods, data=alloy)
```

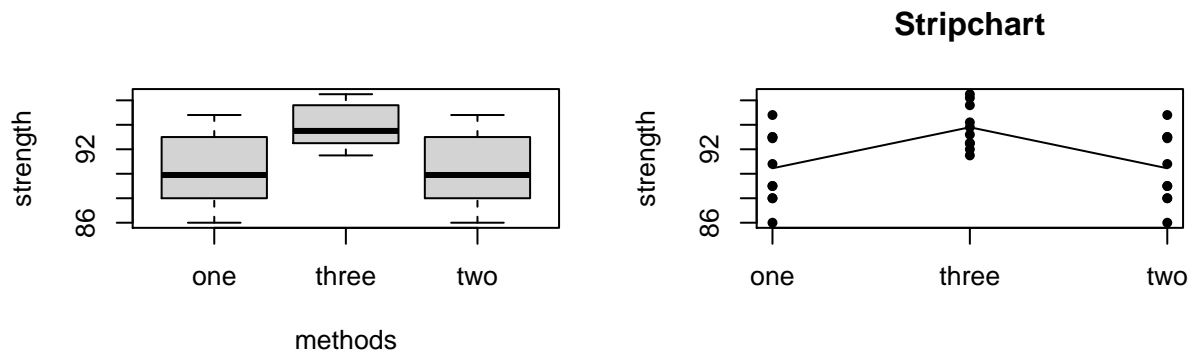
```
##
## Bartlett test of homogeneity of variances
##
## data: strength by methods
## Bartlett's K-squared = 2.2521, df = 2, p-value = 0.3243
```

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

b. Do a boxplot, stripchart.

```
par(mfrow=c(2,2))
boxplot(strength~methods)
stripchart(strength~methods, vertical=TRUE, pch=16, main="Stripchart")

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



c. Build a linear model, using aov(). a summary.aov()

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

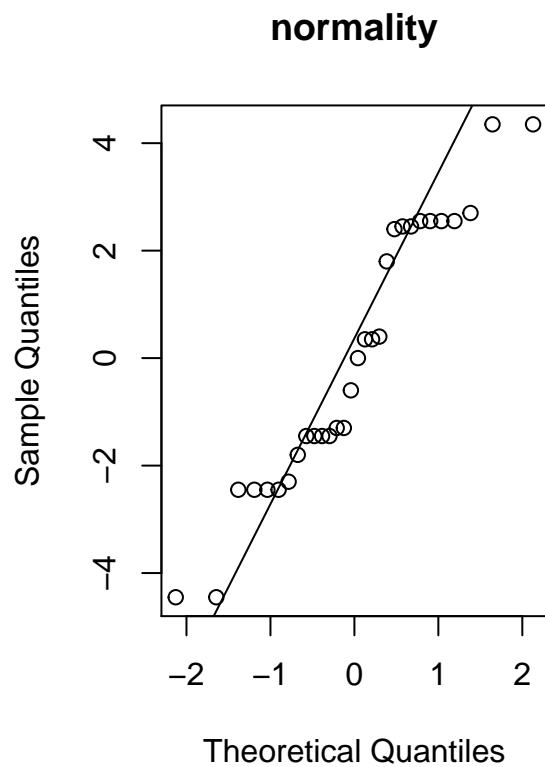
```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## methods      2  74.82   37.41    5.711 0.00854 **
## Residuals   27 176.85     6.55
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

```
##
## Call:
## aov(formula = strength ~ methods)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.450 -1.712 -0.300  2.450  4.350
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.045e+01  8.093e-01 111.760  < 2e-16 ***
## methodsthree  3.350e+00  1.145e+00   2.927  0.00687 **
## methodstwo   1.509e-14  1.145e+00   0.000  1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.559 on 27 degrees of freedom
## Multiple R-squared:  0.2973, Adjusted R-squared:  0.2452
## F-statistic: 5.711 on 2 and 27 DF,  p-value: 0.008541
```

```
par(mfrow=c(1,2))
res <- residuals(strength.model)
qqnorm(res, main="normality")
qqline(res)
shapiro.test(res)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res
## W = 0.9322, p-value = 0.05619
```



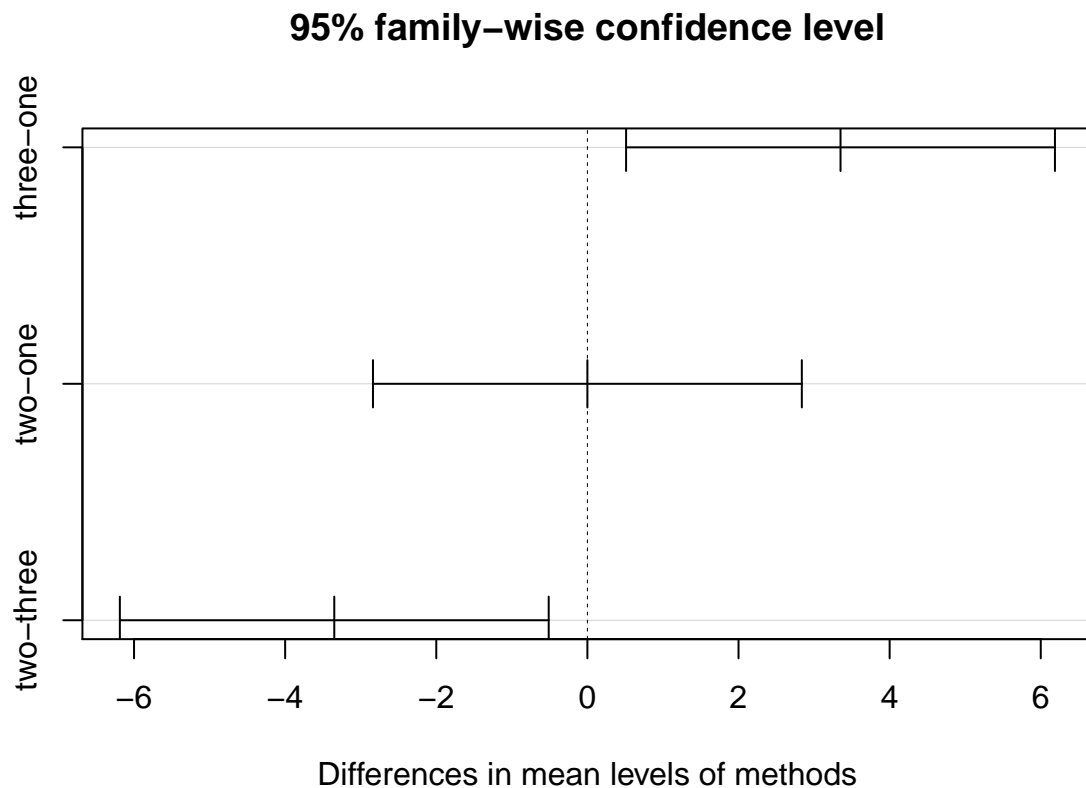
```
### d. Perform TukeyHSD().
```

```
tukey.95 <- TukeyHSD(strength.model, "methods")
```

```
tukey.95
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = strength ~ methods)
##
## $methods
##          diff          lwr          upr      p adj
## three-one 3.350000e+00 0.5121767 6.1878233 0.0182415
## two-one   1.421085e-14 -2.8378233 2.8378233 1.0000000
## two-three -3.350000e+00 -6.1878233 -0.5121767 0.0182415
```

```
plot(tukey.95)
```



There is no statistical difference between two and one, however the difference of mean between three-one and two-three is statistically difference cause mean-diff > tukey's value. It's further established by the fact that the p-value < 0.05 (alpha)

e. Perform a power analysis.

```
strength.means<-tapply(strength, methods, mean)
#strength.sd<-tapply(strength, methods, sd)

#library(dplyr)

#find mean and standard deviation of weight loss for each treatment group

#summarize.data <- alloy %>%
#           group_by(methods) %>%
#           summarize(mean = mean(strength), sd = sd(strength))

summary.aov(strength.model)
```

```
##           Df Sum Sq Mean Sq F value  Pr(>F)
## methods      2  74.82   37.41    5.711 0.00854 **
## Residuals    27 176.85    6.55
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

MSE <- 6.55
n = 2
power.anova.test(groups=3, n=n, between.var=var(strength.means),
                  within.var= MSE,
                  sig.level=.05, power=NULL)

##
##      Balanced one-way analysis of variance power calculation
##
##      groups = 3
##      n = 2
##      between.var = 3.740833
##      within.var = 6.55
##      sig.level = 0.05
##      power = 0.1264235
##
## NOTE: n is number in each group

```

## Q.2

Setup data response time (ms) for three type of circuits that could be used in automatic valve shutoff

```

one <- c(9,12,10,8,15)
two <- c(20,21,23,17,30)
three <- c(6,5,8,16,7)

response.time<- c(one, two, three)
circuit.type<-rep(c("one","two","three"), each =5)
circuit.type<-as.factor(circuit.type)
valves <-data.frame(cbind(response.time, circuit.type))
str(valves)

```

```

## 'data.frame':   15 obs. of  2 variables:
## $ response.time: num  9 12 10 8 15 20 21 23 17 30 ...
## $ circuit.type : num  1 1 1 1 1 3 3 3 3 3 ...

```

a. Test the hypothesis that the three circuit types have the same response time. Use alpha = 0.01

$H_0$  = three circuit have same response time  $H_a$  = atleast two of the three circuits have different response time

assumptions > data is normal

```
attach(valves)
```

```

## The following objects are masked _by_ .GlobalEnv:
##
##      circuit.type, response.time

```

```
valves.model<- aov(response.time ~circuit.type)
summary.aov(valves.model)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## circuit.type  2   543.6    271.8    16.08 0.000402 ***
## Residuals    12   202.8     16.9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

```
##
## Call:
## aov(formula = response.time ~ circuit.type)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
##    -5.2    -2.3    -1.2     1.0     7.8
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)     10.800      1.838   5.874 7.55e-05 ***
## circuit.typethree  -2.400      2.600  -0.923 0.374155
## circuit.typetwo    11.400      2.600   4.385 0.000889 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.111 on 12 degrees of freedom
## Multiple R-squared:  0.7283, Adjusted R-squared:  0.683
## F-statistic: 16.08 on 2 and 12 DF,  p-value: 0.0004023
```

```
res <- residuals(valves.model)
```

```
#check variance
bartlett.test(res~circuit.type)
```

```
##
## Bartlett test of homogeneity of variances
##
## data:  res by circuit.type
## Bartlett's K-squared = 1.1345, df = 2, p-value = 0.5671
```

p values (0.007446) is less than 0.01, so null hypothesis that the circuit type affect doesn't affect response time is rejected and the alternative hypothesis that atleast two circuit type have different response time is accepted.

In-order to get which circuit type have different response time, we have check the pair-wise comparison using Tukeys test

```
tukey.95 <- TukeyHSD(valves.model, which="circuit.type")
tukey.95
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = response.time ~ circuit.type)
##
## $circuit.type
##      diff      lwr      upr      p adj
## three-one -2.4 -9.336445  4.536445 0.6367043
## two-one   11.4  4.463555 18.336445 0.0023656
## two-three 13.8  6.863555 20.736445 0.0005042
```

The circuit.type three-one is not significant but two-one and two-three is significant

### c. Use the graphical procedure to compare treatment means

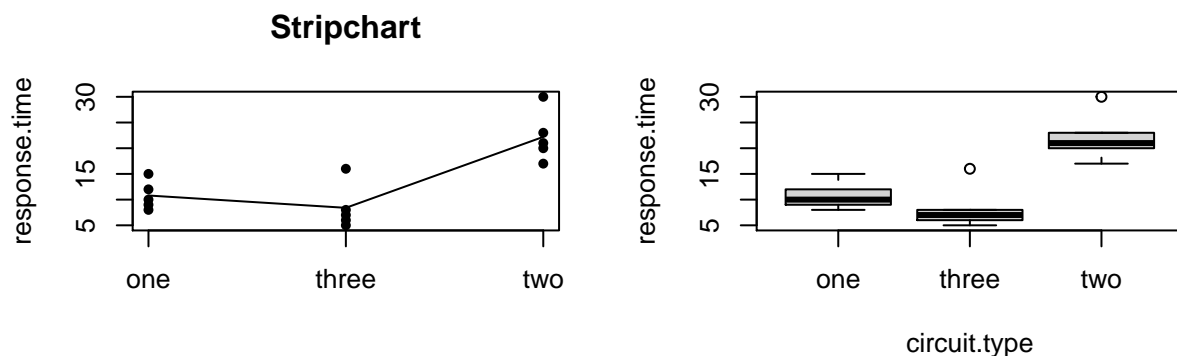
What conclusions can be drawn

```
par(mfrow=c(2,2))
stripchart(response.time~circuit.type, vertical=TRUE, pch=16, main="Stripchart")

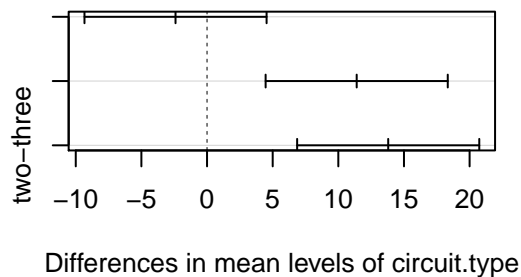
response.means<-tapply(response.time, circuit.type, mean)
lines(response.means)

boxplot(response.time~circuit.type)

plot(tukey(.95))
```



### 95% family-wise confidence level





How do they compare with the conclusions from part b. > Comparing with tukey's table, three-one doesn't have significant difference of means, which is substantiated by the tukey's plot. > two-one and two-three have significant difference in mean because the confidence interval doesn't have any zeros in it. ### d. Construct a set of orthogonal contrasts

```
C1 <-c(1,0,-1)
C2 <-c(-2,3,-1)
C <- rbind(C1,C2)
rownames(C) <- c("Tr1 ==Tr3", "Tr2 = average of Tr3, Tr4")

library(gmodels)
```

```
## Warning: package 'gmodels' was built under R version 4.0.5
```

```
fit.contrast(valves.model, circuit.type, C, conf.int=0.95)
```

```
##                                Estimate Std. Error   t value
## circuit.typeTr1 ==Tr3           -11.4    2.600000  -4.384615
## circuit.typeTr2 = average of Tr3, Tr4  -18.6    6.878953  -2.703900
##                                Pr(>|t|) lower CI  upper CI
## circuit.typeTr1 ==Tr3           0.0008890695 -17.06491 -5.735087
## circuit.typeTr2 = average of Tr3, Tr4 0.0191705620 -33.58795 -3.612048
## attr(,"class")
## [1] "fit_contrast"
```

e. If you were the design engineer and you wished to minimize the response time, which circuit type would you select?

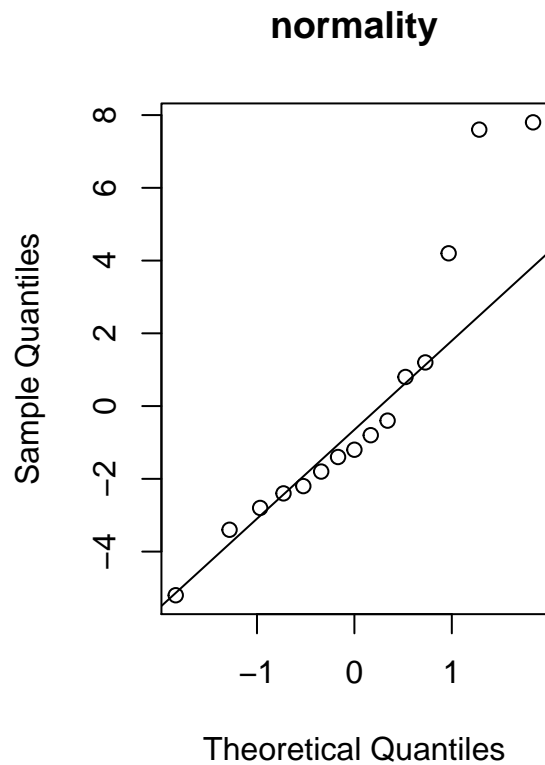
If I have to select a circuit to minimize the response time, I would select circuit type three, because this one has the lowest response time.

f. Analyze the residuals from this experiment. Are the basic analysis of variance assumptions satisfied.

```
par(mfrow=c(1,2))
qqnorm(res, main="normality")
qqline(res)
shapiro.test(res)
```

```
##
## Shapiro-Wilk normality test
##
## data:  res
## W = 0.87352, p-value = 0.03802
```

```
detach()
```



The residuals doesn't appear to be normal. The shapiro test also has p-value smaller than 0.05, which substantiate it. So the assumptions are not satisfied.

### Q.3

Build the data set

```
one <- c(58.2,57.2,58.4,55.8,54.9)
two <- c(56.3,54.5,57.0,55.3)
three <- c(50.1,54.2,55.4)
four <- c(52.9,49.9,50.0,51.7)

concentration <-c(one, two, three, four)
catalyst <-c(rep(c("one"), each=5), rep(c("two"), each=4), rep(c("three"), each=3), rep(c("four"), each=4))
catalyst <- as.factor(catalyst)
mixture <- data.frame(cbind(concentration, catalyst))
str(mixture)

## 'data.frame':  16 obs. of  2 variables:
## $ concentration: num  58.2 57.2 58.4 55.8 54.9 56.3 54.5 57 55.3 50.1 ...
## $ catalyst      : num  2 2 2 2 2 4 4 4 4 3 ...
```

a. Do the four catalyst have the same effect on the concentration?

```
attach(mixture)

## The following objects are masked _by_ .GlobalEnv:
##
##      catalyst, concentration

mixture.model <- aov(concentration~catalyst)
summary.aov(mixture.model)

##              Df Sum Sq Mean Sq F value    Pr(>F)
## catalyst      3  85.68   28.56    9.916 0.00144 **
## Residuals    12  34.56    2.88
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary.lm(mixture.model)

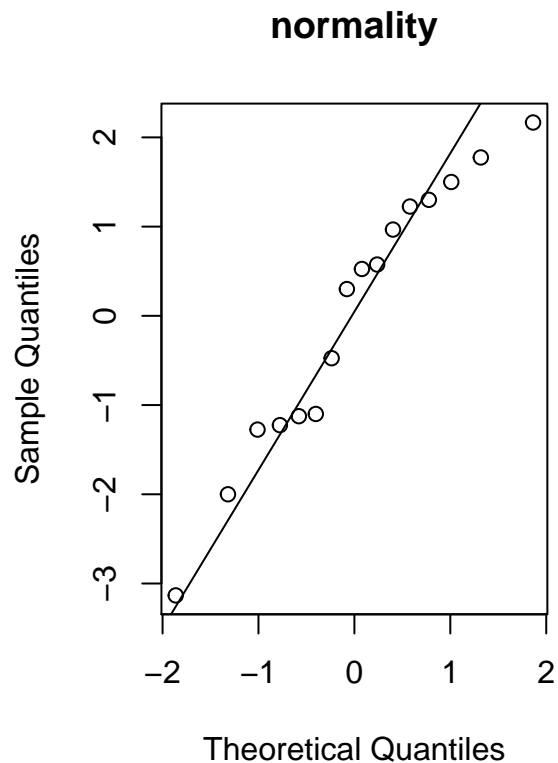
##
## Call:
## aov(formula = concentration ~ catalyst)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.1333 -1.1500  0.4125  1.2437  2.1667
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    51.1250     0.8485  60.250 2.89e-16 ***
## catalystone     5.7750     1.1384   5.073 0.000274 ***
## catalystthree   2.1083     1.2962   1.627 0.129783
## catalysttwo     4.6500     1.2000   3.875 0.002208 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.697 on 12 degrees of freedom
## Multiple R-squared:  0.7126, Adjusted R-squared:  0.6407
## F-statistic: 9.916 on 3 and 12 DF,  p-value: 0.001436
```

pvalue is small, so null hypothesis that the means of concentration for different catalyst is same is rejected and it's accepted that the catalyst have different effect on concentration.

b. Analyse the residuals from this experiment

```
par(mfrow=c(1,2))
res <- residuals(mixture.model)
qqnorm(res, main="normality")
qqline(res)
shapiro.test(res)
```

```
##
## Shapiro-Wilk normality test
##
## data:  res
## W = 0.9486, p-value = 0.4679
```



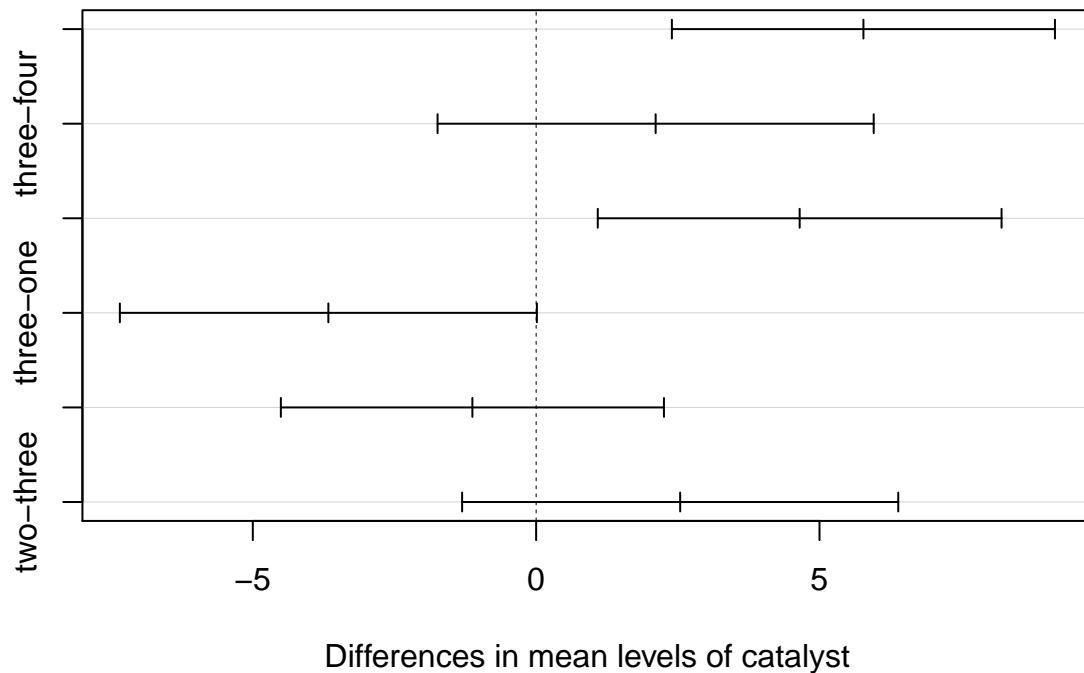
The qqnorm plot demonstrate that residuals doesn't appear to have normal values, but the shapiro test has pvalue  $> 0.05$ , so it can be ascertained that the residuals are normal.

```
tukey <- TukeyHSD(mixture.model, "catalyst")
tukey
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = concentration ~ catalyst)
##
## $catalyst
##          diff          lwr          upr      p adj
## one-four    5.775000    2.395062    9.15493778 0.0013492
## three-four   2.108333   -1.739895    5.95656179 0.4010616
## two-four     4.650000    1.087233    8.21276725 0.0102709
## three-one   -3.666667   -7.346277    0.01294353 0.0509093
## two-one     -1.125000   -4.504938    2.25493778 0.7587514
## two-three    2.541667   -1.306562    6.38989512 0.2552535
```

```
plot(tukey)
```

### 95% family-wise confidence level



c. Construct a 99 percent interval estimate of the mean response for catalyst 1

Calculate SST, SS(Treatment), SS(Error)

```
mixture.means<-tapply(concentration, catalyst, mean)

a <- concentration[catalyst=="one"]
b <- concentration[catalyst=="two"]
c <- concentration[catalyst=="three"]
d <- concentration[catalyst=="four"]

a.mean <-mean(a)
b.mean <-mean(b)
c.mean <-mean(c)
d.mean <-mean(d)

grand.mean <- mean(mixture.means)

ss.error <-sum((a-a.mean)^2)+ sum((b-b.mean)^2)+sum((c-c.mean)^2)+sum((d-d.mean)^2)
ss.error
```

```
## [1] 34.56167
```

```
ss.total <-sum((concentration-grand.mean)^2)
ss.total
```

```
## [1] 121.0778
```

```
ss.treatment <-ss.total-ss.error
ss.treatment
```

```
## [1] 86.51611
```

The ss.treatment and ss.error values are matching with those calculated in anova table.