

Stats 101B HW 2 Lecture 3A

Charles Liu (304804942)

4/18/2020

Loading Packages:

```
require(Hmisc)
```

```
## Loading required package: Hmisc
## Warning: package 'Hmisc' was built under R version 3.6.3
## Loading required package: lattice
## Warning: package 'lattice' was built under R version 3.6.3
## Loading required package: survival
## Loading required package: Formula
## Loading required package: ggplot2
##
## Attaching package: 'Hmisc'
## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
require(DescTools)
```

```
## Loading required package: DescTools
## Warning: package 'DescTools' was built under R version 3.6.3
##
## Attaching package: 'DescTools'
## The following objects are masked from 'package:Hmisc':
##
##     %nin%, Label, Mean, Quantile
```

Problem 2:

Creating Data for Problem 2:

```
observations <- c(24, 28, 37, 30,
                 37, 44, 31, 35,
                 42, 47, 52, 38)
dosages <- c(rep("20g", 4), rep("30g", 4), rep("40g", 4))
```

```
drug.data <- data.frame(dosages = dosages, observations = observations)
drug.data
```

```
##      dosages observations
## 1      20g             24
## 2      20g             28
## 3      20g             37
## 4      20g             30
## 5      30g             37
## 6      30g             44
## 7      30g             31
## 8      30g             35
## 9      40g             42
## 10     40g             47
## 11     40g             52
## 12     40g             38
```

2b)

```
summary(observations ~ dosages, data = drug.data)
```

```
## observations      N= 12
##
## +-----+-----+-----+
## |      |      |N|observations|
## +-----+-----+-----+
## |dosages|20g| 4|29.75000  |
## |      |30g| 4|36.75000  |
## |      |40g| 4|44.75000  |
## +-----+-----+-----+
## |Overall|  |12|37.08333  |
## +-----+-----+-----+
```

```
m1 <- lm(observations ~ dosages, data = drug.data)
summary(m1)
```

```
##
## Call:
## lm(formula = observations ~ dosages, data = drug.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.75  -3.50  -0.75   3.50   7.25
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    29.750     2.830  10.514 2.35e-06 ***
## dosages30g      7.000     4.002   1.749  0.11418
## dosages40g     15.000     4.002   3.748  0.00457 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.659 on 9 degrees of freedom
## Multiple R-squared:  0.6099, Adjusted R-squared:  0.5232
```

```
## F-statistic: 7.036 on 2 and 9 DF, p-value: 0.01446
```

```
aov(m1)
```

```
## Call:
## aov(formula = m1)
##
## Terms:
##          dosages Residuals
## Sum of Squares 450.6667 288.2500
## Deg. of Freedom      2      9
##
## Residual standard error: 5.659309
## Estimated effects may be unbalanced
```

```
anova(m1)
```

```
## Analysis of Variance Table
##
## Response: observations
##          Df Sum Sq Mean Sq F value Pr(>F)
## dosages    2 450.67  225.333   7.0356 0.01446 *
## Residuals  9 288.25   32.028
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2e)

```
MSE_2 <- 32.028
```

```
qtukey(1 - 0.05, 3, 12 - 3)/(sqrt(2))
```

```
## [1] 2.792006
```

```
tstat <- (mean(observations[dosages == "30g"]) -
          mean(observations[dosages == "20g"])) / sqrt(2 * MSE_2 / 4)
tstat #  $H_0: \mu_1 = \mu_2$ 
```

```
## [1] 1.749235
```

```
1-ptukey(tstat*sqrt(2), 3, 12-3) # p-value
```

```
## [1] 0.2402996
```

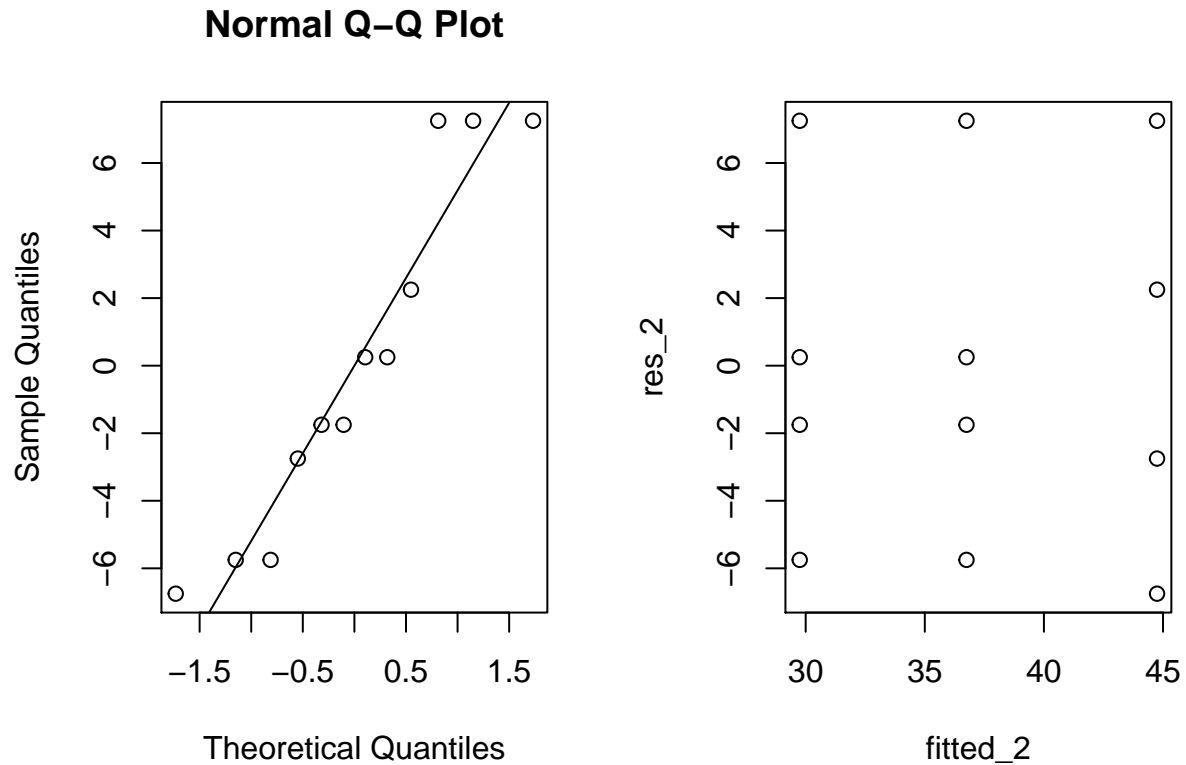
```
TukeyHSD(x= aov(m1), 'dosages', conf.level = 0.95)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = m1)
##
## $dosages
##      diff      lwr      upr    p adj
## 30g-20g    7 -4.172869 18.17287 0.2402975
## 40g-20g   15  3.827131 26.17287 0.0114434
## 40g-30g    8 -3.172869 19.17287 0.1680265
```

We fail to REJECT the NULL Hypothesis and we can say they are NOT significantly different for the means.

2f)

```
res_2 <- m1$residual
fitted_2 <- m1$fitted.values
par(mfrow = c(1, 2))
# normality check
qqnorm(res_2); qqline(res_2)
# no pattern / constant variance check
plot(fitted_2, res_2)
```



All assumptions are satisfied for all the plots.

Problem 3:

Creating Data for Problem 3:

```
conductivity <- c(143, 141, 150, 146,
                 152, 149, 137, 143,
                 134, 136, 132, 127,
                 129, 127, 132, 129)
coating_type <- c(rep("1", 4), rep("2", 4), rep("3", 4), rep("4", 4))
tv.data <- data.frame(coating_type = coating_type, conductivity = conductivity)
tv.data
```

```
##   coating_type conductivity
## 1             1          143
```

```
## 2          1          141
## 3          1          150
## 4          1          146
## 5          2          152
## 6          2          149
## 7          2          137
## 8          2          143
## 9          3          134
## 10         3          136
## 11         3          132
## 12         3          127
## 13         4          129
## 14         4          127
## 15         4          132
## 16         4          129
```

3b)

```
summary(conductivity ~ coating_type, data = tv.data)
```

```
## conductivity      N= 16
##
## +-----+-----+
## |           | N | conductivity |
## +-----+-----+
## |coating_type|1| 4|145.0000 |
## |           |2| 4|145.2500 |
## |           |3| 4|132.2500 |
## |           |4| 4|129.2500 |
## +-----+-----+
## |Overall    | |16|137.9375 |
## +-----+-----+
```

```
m2 <- lm(conductivity ~ coating_type, data = tv.data)
summary(m2)
```

```
##
## Call:
## lm(formula = conductivity ~ coating_type, data = tv.data)
##
## 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 ***
## coating_type2     0.250     3.138    0.080 0.937804
## coating_type3   -12.750     3.138   -4.064 0.001571 **
## coating_type4   -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
```

```
aov(m2)
```

```
## Call:
## aov(formula = m2)
##
## Terms:
##              coating_type Residuals
## Sum of Squares      844.6875  236.2500
## Deg. of Freedom           3        12
##
## Residual standard error: 4.43706
## Estimated effects may be unbalanced
```

```
anova(m2)
```

```
## Analysis of Variance Table
##
## Response: conductivity
##              Df Sum Sq Mean Sq F value    Pr(>F)
## coating_type  3 844.69  281.562   14.302 0.0002881 ***
## Residuals    12 236.25   19.688
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3d)

```
MSE_3 <- 19.6875 # from the ANOVA table
mean(conductivity[coating_type == "1"]) -
  mean(conductivity[coating_type == "4"]) +
  qt(1 - 0.005, 16 - 4)*sqrt(2*MSE_3/4)
```

```
## [1] 25.33354
```

```
mean(conductivity[coating_type == "1"]) -
  mean(conductivity[coating_type == "4"]) -
  qt(1 - 0.005, 16 - 4)*sqrt(2*MSE_3/4)
```

```
## [1] 6.166458
```

3e)

```
TukeyHSD(x= aov(m2), 'coating_type', conf.level = 0.95)
```

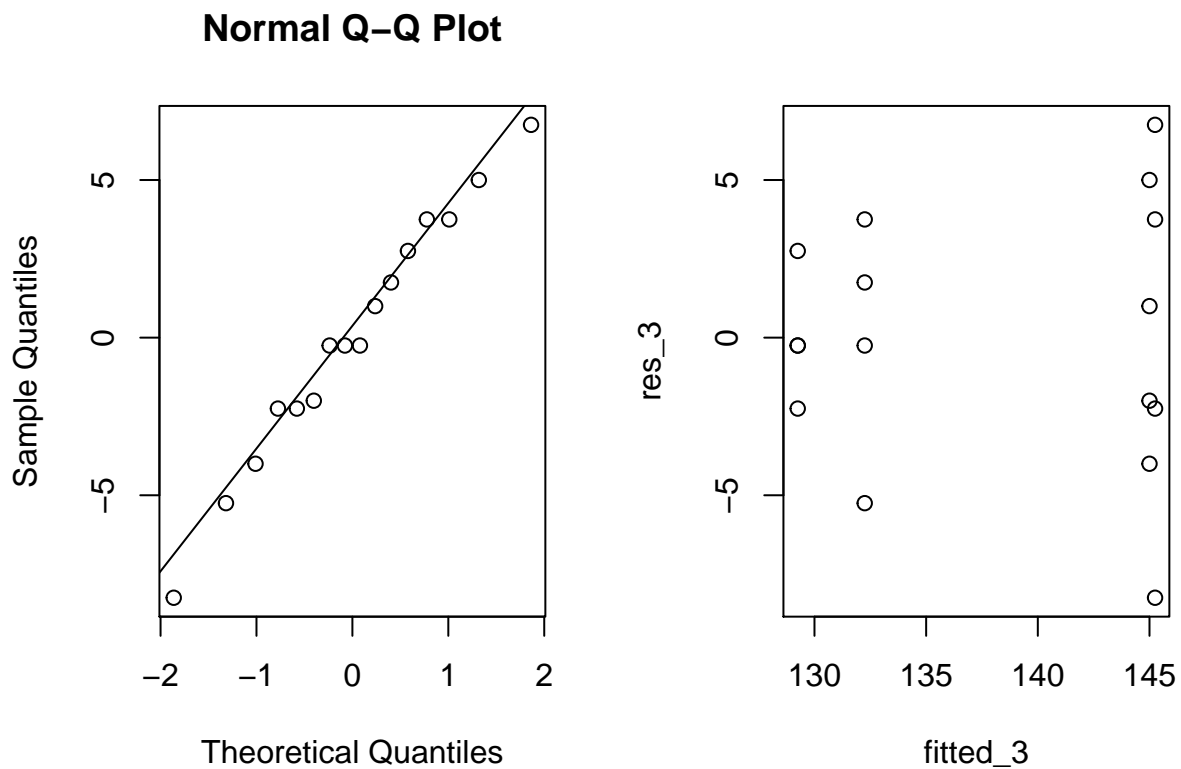
```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = m2)
##
## $coating_type
##      diff      lwr      upr    p adj
## 2-1   0.25 -9.064853  9.564853 0.9998078
## 3-1 -12.75 -22.064853 -3.435147 0.0073964
## 4-1 -15.75 -25.064853 -6.435147 0.0014707
## 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
```

We fail to REJECT the NULL Hypothesis and we can say they are NOT significantly different for the means.

3f)

```
res_3 <- m2$residual
fitted_3 <- m2$fitted.values
par(mfrow = c(1, 2))
# normality check
qqnorm(res_3); qqline(res_3)
# no pattern / constant variance check
plot(fitted_3, res_3)
```



All assumptions are satisfied for all the plots.