Charles Liu
304804942
Stats 101B
Disc. 3A
4-18-20

① ANOVA Table

Source	DF	SS	MS	F	P
Factor	?=4	?=987.71	246.93	?=33.10	P<0.00001 (?=)
Error	25	186.53	?=7.46		
Total	29	1174.24			

$SST = SSR + SSE$

$MS = \dfrac{SS}{DF}$

$F = \dfrac{MSR}{MSE}$

② ⓐ

Dosage	Observations				Total (M)	n	$\overline{Y}_{i \cdot}$
20g	24	28	37	30	119	4	29.75
30g	37	44	31	35	147	4	36.75
40g	42	47	52	38	179	4	44.75
					445	12	37.08

ANOVA Table

Source	DF	SS	MS	F	P
Dosage	2	453.63	226.82	7.16	P<0.05 ⊛
Error	9	285.29	31.70		
Total	11	738.92			

$(24-37.08)^2 + (28-37.08)^2 + (37-37.08)^2 + (30-37.08)^2 =$ SST 303.6656

$(37-37.08)^2 + (44-37.08)^2 + (31-37.08)^2 + (35-37.08)^2 =$ 89.1856

$(42-37.08)^2 + (47-37.08)^2 + (52-37.08)^2 + (38-37.08)^2 =$ 346.0656

738.9168

$SSE = SST - SSR = 285.29$

Answers off slightly from RStudio since I rounded up

$(4) \sum (29.75)^2 + (36.75)^2 + (44.75)^2 - (12)(37.08)^2 = (16952.75) - (16499.1168) =$

$SSR = 453.6332$

$F_0 > F_{0.05, 2, 9}$

$7.16 > 4.26$ Reject!

①

(b) "done in RStudio"

(c) Since the $F_0 > F_{0.05, 2, 99}$ we Reject the Null Hypothesis and say that there is Significant difference between Dosage & Bioactivity. (Dosage affects bioactivity)

(d) Yes, it would be appropiate to compare between pairs of means because we need to check our Power test (β) to see if we might've made a Type II Error. The F-test had the Null Rejected, too.

(e) $H_0 : \mu_i = \mu_j$ for
$H_a : \mu_i \neq \mu_j$ if at least one (i) $\Bigg\} \quad T_{ij} = \dfrac{\overline{Y_i} - \overline{Y_j}}{\sqrt{\dfrac{MSE}{n_i} + \dfrac{MSE}{n_j}}} \longrightarrow |T_{ij}| > \underbrace{\dfrac{1}{\sqrt{2}} q_\alpha (a, N-a)}_{2.79}$

"Work done in RStudio"

$T_{20g, 30g} = \left| \dfrac{(29.75 - 36.75)}{\sqrt{\dfrac{32.028}{4_2} \cdot \left(\dfrac{2}{1}\right)}} \right| = 1.75 > 2.79 \quad \otimes$

$T_{20g, 40g} = \left| \dfrac{(29.75 - 44.75)}{\sqrt{\dfrac{32.028}{2.2} \cdot \left(\dfrac{2}{1}\right)}} \right| = 3.75 > 2.79 \quad \checkmark$

$T_{30g, 40g} = \left| \dfrac{(36.75 - 44.75)}{\sqrt{\dfrac{32.028}{4_2} \cdot \left(\dfrac{2}{1}\right)}} \right| = 1.999 > 2.79 \quad \otimes$

We fail to Reject the Null, and we can say they are NOT significantly different for the means

(f) "Plots done in RStudio" We can see that the Normal Plot & Residuals Plot have their assumptions satisfied

(Next Page)

2

Coating Type	Conductivity				$Y_{i.}$	n	$\overline{Y}_{i.}$
1	143	141	150	146	580	4	145
2	152	149	137	143	581	4	145.25
3	134	136	132	127	529	4	132.25
4	129	127	132	129	517	4	129.25
Total					2207	16	137.9375

$$SST = \sum_i \sum_j Y_{ij}^2 - N \cdot \overline{Y}_{..}^2 \longrightarrow (143^2 + 141^2 + \cdots + 129^2) - (16) \cdot (137.9375^2) \Longrightarrow$$

$$(305509) - (304428.0625) = \boxed{1080.9375} = SST$$

$$SSR = n \sum_i \overline{Y}_{i.}^2 - N \cdot \overline{Y}_{..}^2 = (145^2 + \cdots + 129.25^2)(4) - (16)(137.9375^2) \Longrightarrow$$

$$(305272.75) - (304428.0625) = \boxed{844.6875} = SSR$$

$$SSE = SST - SSR \longrightarrow (1080.9375) - (844.6875) = \boxed{236.25} = SSE$$

ANOVA table

Source	DF	SS	MS	F	P
Coating Type	3	844.6875	281.5625	14.302	$P < 0.05$ ***
Error	12	236.25	19.6875		
Total	15	1080.9375			

$F_0 > F_{\alpha, a-1, N-a}$ Reject!

$14.302 > 3.49$

ⓑ "done in RStudio"

ⓒ $F_0 > F_{0.05, 3, 12}$ [$14.302 > 3.49$] Therefore, we Reject the Null Hypothesis and say there is a significant difference between Coating Type & Conductivity.

ⓓ $(\mu_1 - \mu_4) \pm t_{\frac{\alpha}{2}, N-a} \sqrt{\dfrac{MSE \cdot (2)}{n}} \longrightarrow (145 - 129.25) \pm (3.055) \cdot \sqrt{\dfrac{(19.6875)(2)}{4}}$

$$= \boxed{[6.17, 25.33]}$$ Confidence Interval for 99%

ⓔ "Work in RStudio" | We fail to Reject the Null Hypothesis, and we say there is No significant difference between means.

③

④ "Plots in RStudio" We can see both Normal & Residual Plots have their assumptions satisfied.

---

④ $M_1 = 50$ $\alpha = 0.05$ $\left.\begin{array}{l}\end{array}\right\}$ $M_i = M + T_i$ for $i = 1, \ldots, 4$

$M_2 = 60$ $\sigma^2 = 25$

$M_3 = 50$ $M = \dfrac{\sum M_i}{a} = \left(\dfrac{220}{4}\right) = 55$

$M_4 = 60$ $T_i = (M_i - M)$

$T_1 = -5$ $\left.\begin{array}{l}\end{array}\right\}$ $\displaystyle\sum_{i=1}^{a=4} T_i^2 = 100$ $\phi^2 = \dfrac{n \sum T_i^2}{a \cdot \sigma^2} = \dfrac{(100)(n)}{(4)(25)} = n$

$T_2 = 5$

$T_3 = -5$ $\boxed{\phi = \sqrt{n}}$

$T_4 = 5$

n	ϕ	V_1	V_2	β	$(1-\beta)$ (Power)
4	2	3	12	0.18	0.82
5	2.24	3	16	0.08	0.92

$V_1 = 3$

$V_2 = 4(n-1)$

We would choose $n=5$ since the Power Test is 0.92 (at least 0.90).

④

Stats 101B HW 2 Lecture 3A

Charles Liu (304804942)

4/18/2020

Loading Packages:

```r
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
```

```r
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:

```r
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: M_1 = M_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)
```

Normal Q–Q Plot
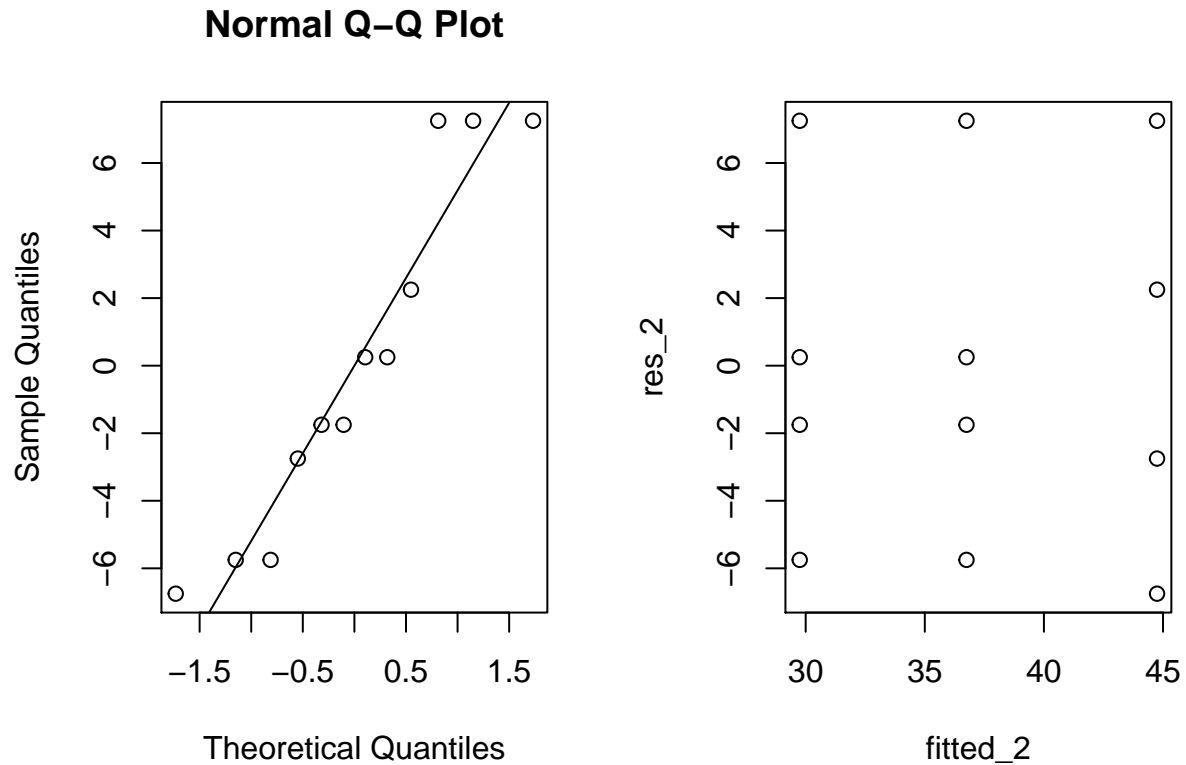
All assumptions are satisfied for all the plots.

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
```

5

```
## F-statistic:  14.3 on 3 and 12 DF,  p-value: 0.0002881
```

```r
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
```

```r
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)

```r
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
```

```r
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)

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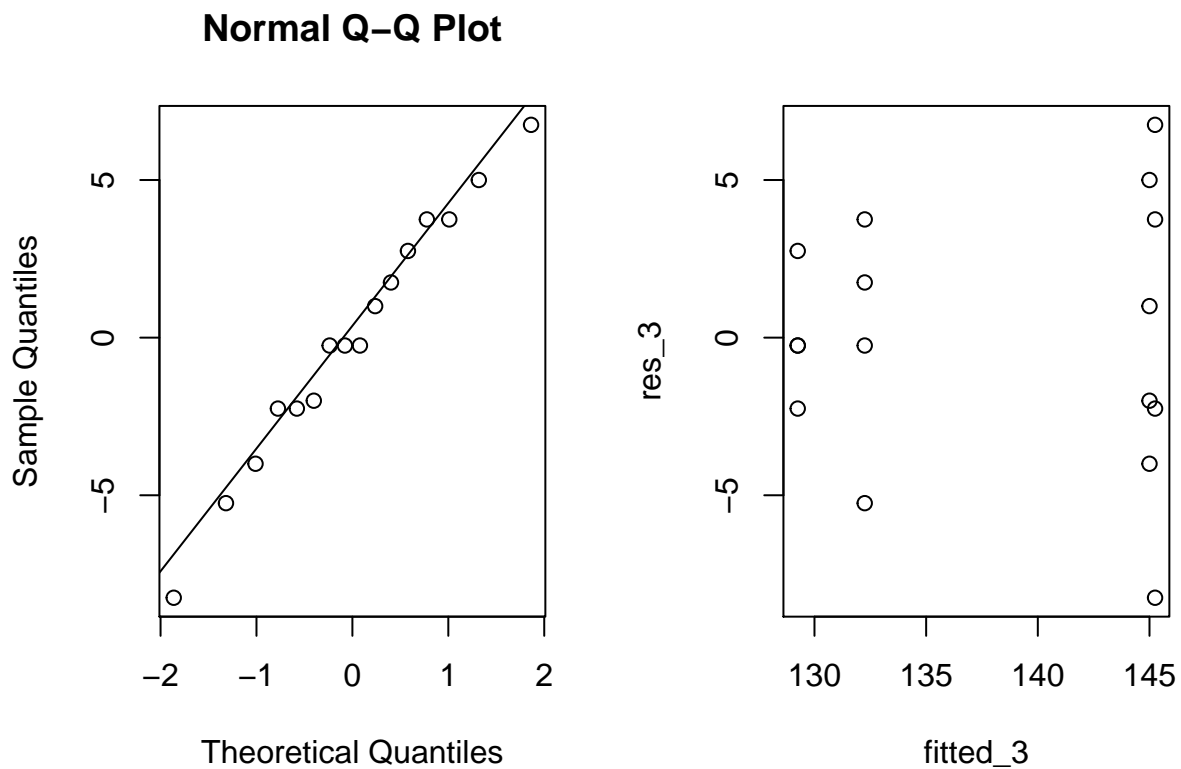
```
## 4-2 -16.00 -25.314853 -6.685147 0.0012913
## 4-3  -3.00 -12.314853  6.314853 0.7759360
```

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

3f)

```r
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.