Statistics 101B, Dis 3B
Homework 2

Kitu Komya
UID: 404-491-375

---

1. The completed table is as follows:

Source	DF	SS	MS	F	P
Factor	3	36.15	**12.05**	**1.21**	**0.3395**
Error	**16**	**159.89**	**9.99**		
Total	19	196.04			

DF: **16** = 19 – 3
SS: **159.89** = 196.04 – 36.15
MS: **12.05** = 36.15/3
9.99 = 159.89/16
F: **1.21** = 12.05/9.99
P: **0.3395** = using online calculator

2a.

Source	DF	SS	MS	F	P
Dosage	**2**	**456**	**228**	**8.40**	**0.009**
Error	**9**	**244.3**	**27.1**		
Total	**11**	**700.3**			

DF: **2** = 3 dosage levels – 1
11 = 12 total observations – 1
9 = 11 – 2
SS: mean of 20g = 29.75, mean of 30g = 38.75, mean of 40g = 44.75, total mean = 37.75
StD of 20g = 5.44, StD of 30g = 3.86, StD of 40g = 6.08
456 = $4*(29.75 – 37.75)^2 + 4*(38.75 – 37.75)^2 + 4*(44.75 – 37.75)^2$
244.3 = $3*(5.44)^2 + 3*(3.86)^2 + 3*(6.08)^2$
700.3 = 456 + 244.3
MS: **228** = 456/2
27.1 = 244.3/9
F: **8.40** = 228/27.1
P: **0.009** = using online calculator

2b.

```
> fit <- aov(Observation~Level, data = drug1) # fit anova test
> summary(fit) # shows anova test
            Df Sum Sq Mean Sq F value Pr(>F)
Level        2  450.7  225.33   7.036 0.0145 *
Residuals    9  288.2   32.03
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The values are very similar to the ones I calculated by hand. When I did it by hand, I rounded my values, so that accounts for the differences. Both reached the same conclusions via the p-value, so the differences are negligible.

2c. At a significance level of 0.05, we see that since the p-value 0.009 is smaller than 0.05, we can conclude that there is a difference in the dosages. The null hypothesis is thus rejected and we can say that yes, dosage levels do affect bioactivity.

2d. Yes, it would be appropriate since the means of the dosages are clearly different from each other.

2e. We will use the Tukey method: where 1 is 20g, 2 is 30g, and 3 is 40g. $t = \frac{\overline{y_2} - \overline{y_1}}{\hat{\sigma}\sqrt{\frac{1}{n_2}+\frac{1}{n_1}}}$

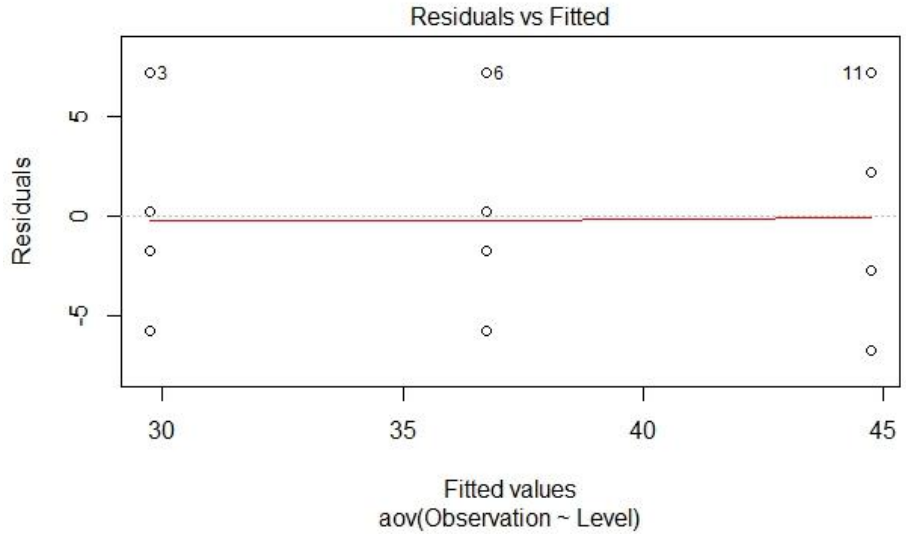
1 vs 2: $t_{12} = \frac{38.75 - 29.75}{5.21\sqrt{\frac{1}{4}+\frac{1}{4}}} = 2.44$

1 vs 3: $t_{13} = \frac{44.75 - 29.75}{5.21\sqrt{\frac{1}{4}+\frac{1}{4}}} = 4.07$

2 vs 3: $t_{23} = \frac{44.75 - 38.75}{5.21\sqrt{\frac{1}{4}+\frac{1}{4}}} = 1.63$

If these numbers are declare that $|t_{ij}| > \frac{1}{\sqrt{2}} q_{a,N-a,\alpha} = \frac{1}{\sqrt{2}} q_{4,12-4,0.05} = 3.95$ using qtukey(1-0.05, 4, 8) in R. This means that pairs 1 and 3 are different from each other. In other words, there is a difference in means between the 20g and the 40g levels.

2f. The residual plot is randomly scattered with equal amounts of observations above and below the line. Because there is no structure or pattern, the model is valid.



3a.

Source	DF	SS	MS	F	P
Dosage	**3**	**844.69**	**281.56**	**10.73**	**0.000102**
Error	**12**	**314.93**	**26.24**		
Total	**15**	**1159.62**			

DF: **3** = 4 coating type levels – 1
15 = 16 total observations – 1
12 = 15 – 3

SS: 1's mean = 145, 2's mean = 145.25, 3's mean = 132.25, 4's mean = 129.25
tot. mean = 137.93
StD of 1 = 3.92, StD of 2 = 6.65, StD of 3 = 3.86, StD of 4 = 2.06
844.69 = $4*(145 - 137.93)^2 + 4*(145.25 - 137.93)^2 + 4*(132.25 - 137.93)^2 + 4*(129.25 - 137.93)^2$
314.93 = $4*(3.92)^2 + 4*(6.65)^2 + 4*(3.86)^2 + 4*(2.06)^2$
1159.62 = 844.69 + 314.93

MS: **281.56** = 844.69/3
26.24 = 314.93/12

F: **10.73** = 281.56/26.24

P: **0.000102** = using online calculator

3b.

```
> fit2 <- aov(Conductivity~Coating, data = conduct) # fit anova test
> summary(fit2)
            Df Sum Sq Mean Sq F value   Pr(>F)
Coating      1  726.0   726.0   28.64 0.000102 ***
Residuals   14  354.9    25.4
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The values are very similar to the ones I calculated by hand. When I did it by hand, I rounded my values, so that accounts for the differences. Both reached the same conclusions via the p-value, so the differences are negligible.

3c. At a significance level of 0.05, we see that since the p-value 0.000102 is smaller than 0.05, we can conclude that there is a difference in the coating types. The null hypothesis is thus rejected and we can say that yes, coating types do affect conductivities.

3d. confidence interval for coating type 4: $129.25 \pm 2.179* \sqrt{\frac{19.69}{4}}$

So, the interval is (124.42, 134.08).

3e. confidence interval between types 1 and 4: $(145 - 129.25) \pm 2.179* \sqrt{\frac{19.69}{4}}$

So, the interval is (6.164, 25336).

3f. 3.6203We will use the Tukey method: where the numbers refer to coating types. $t = \frac{\overline{y_2} - \overline{y_1}}{\hat{\sigma}\sqrt{\frac{1}{n_2}+\frac{1}{n_1}}}$

1 vs 2: $t_{12} = \frac{145.25-145}{5.12\sqrt{\frac{1}{4}+\frac{1}{4}}} = 0.0691$

1 vs 3: $t_{13} = \frac{132.25-145}{5.12\sqrt{\frac{1}{4}+\frac{1}{4}}} = -3.52$

1 vs 4: $t_{23} = \frac{129.25-145}{5.12\sqrt{\frac{1}{4}+\frac{1}{4}}} = -4.35$

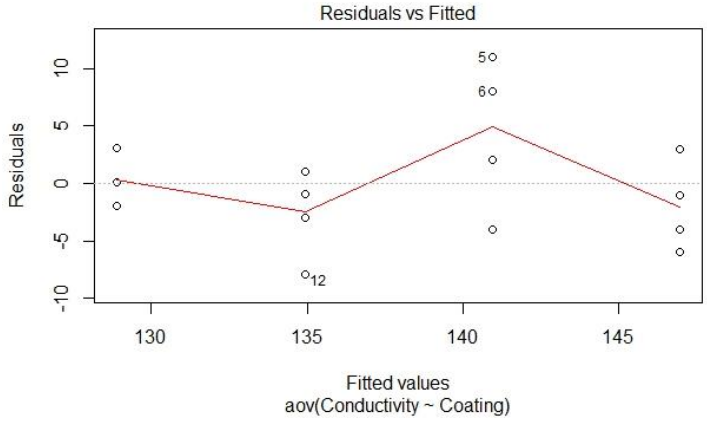
2 vs 3: $t_{23} = \frac{145.25-132.25}{5.12\sqrt{\frac{1}{4}+\frac{1}{4}}} = 3.59$

2 vs 4: $t_{23} = \frac{129.25-145.25}{5.12\sqrt{\frac{1}{4}+\frac{1}{4}}} = -4.42$

3 vs 4: $t_{23} = \frac{129.25-132.25}{5.12\sqrt{\frac{1}{4}+\frac{1}{4}}} = -0.829$

If these numbers are declare that $|t_{ij}| > \frac{1}{\sqrt{2}} q_{a,N-a,\alpha} = \frac{1}{\sqrt{2}} q_{4,16-4,0.05} = 3.19$ using qtukey(1-0.01, 4, 8) in R. This means that pairs 1 and 3 and pairs 1 and 4 and pairs 2 and 3 and pairs 2 and 4 are different from each other. In other words, there is a difference in means between the coating type 1 and coating type 3 as well as between coating type 1 and coating type 4 as well as between coating type 2 and coating type 3 as well as between coating type 2 and coating type 4.

3g. The residual plot is not as randomly scattered as I would have hoped. In fact, there is a clear heteroskedastic trend (funnel trend from left to right) which could lead us to believe that the assumption of constant variance has been violated. Thus, the model may be invalid.

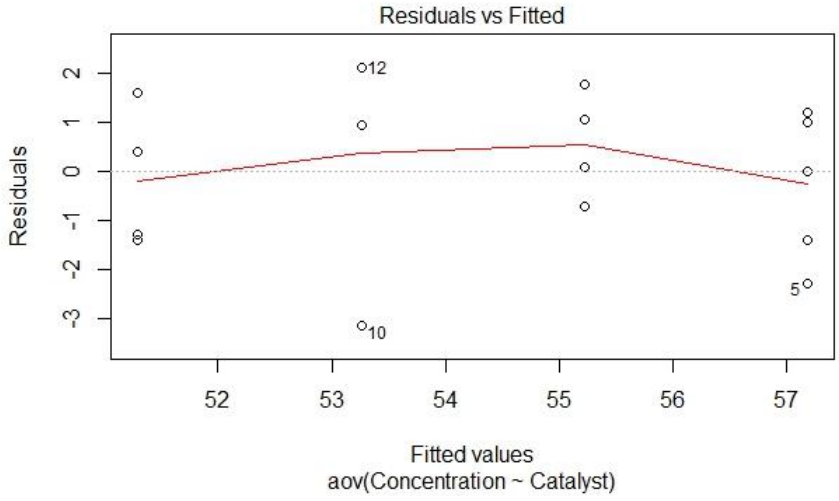


4a.

```
> fit3 <- aov(Concentration~Catalyst, data = cata) # fit anova test
> summary(fit3)
            Df Sum Sq Mean Sq F value  Pr(>F)
Catalyst     1  83.93   83.93   32.36 5.6e-05 ***
Residuals   14  36.31    2.59
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Using R, we see that at a significance level of 0.05, we see that since the p-value 0.000056 is smaller than 0.05, we can conclude that there is a difference in the means. The null hypothesis is thus rejected and we can say that yes, the type of catalyst does have an effect on concentration.

4b. The residual plot is randomly scattered with equal amounts of observations above and below the line. Because there is no structure or pattern, the model is valid.



4c. confidence interval for catalyst 1: $56.9 \pm 3.055 * \sqrt{\frac{2.88}{5}}$

Thus, the interval is (54.58, 59.22).