Assignment3

Kyle Ligon

2/11/2020

$\bar{y}_{..} = 15.04$

1).

Cotton Weight Percent	Observations	
15	7 7 15 11 9	$\bar{y}_{1.} = 9.8$
20	12 17 12 18 18	$\bar{y}_{2.} = 15.4$
25	14 19 19 18 18	$\bar{y}_{3.} = 17.6$
30	19 25 22 19 23	$\bar{y}_{4.} = 21.6$
35	7 10 11 15 11	$\bar{y}_{5.} = 10.8$

a) Is there evidence to support the claim that cotton content affects the mean tensile strength? Use $\alpha = 0.05$

Source	SS	df	MS	F_0
Between Trt	475.76	4	118.94	14.75672
Error (within Trt)	161.20	20	8.06	
Total	636.96	24		

$SS_{trt} = \sum_{1}^{a}\sum_{1}^{n}(\bar{y}_{1.} - \bar{y}_{..})^2$

$= \sum_{1}^{a} n(\bar{y}_{1.} - \bar{y}_{..})^2$

$= \frac{1}{n}\sum_{1}^{5} y_{1.}^2 - \frac{Y_{..}^2}{N}$

$= \frac{1}{5}(49^2 + 77^2 + 88^2 + 108^2 + 54^2) - \frac{376^2}{25}$

$= \frac{1}{5}(30,654) - \frac{141376}{25}$

$= 6,130.8 - 5655.04$

$= 475.76$

$SS_{Tot} = \sum_{1}^{a}\sum_{1}^{n} y_{ij}^2 - \frac{(y_{..})^2}{N}$

$= 6292 - \frac{141376}{25}$

$= 6292 - 5655.04$

$= 636.96$

P-value $= 9.13 \times 10^{-6}$

With a p-value less than 0.05 there is evidence to reject the null hypothesis that all the group means are the same. There is evidence to suggest that the tensile strength is affected by cotton weights.

b) Use the Fisher LSD method to make comparisons between the pairs of means. What conclusions can you draw?

$LSD = t_{0.025, N-a}\sqrt{\frac{2MS_E}{n}} = t_{0.025, 20}\sqrt{\frac{2(8.06)}{5}} = 2.086\sqrt{\frac{2(8.06)}{5}} = 2.086(1.79555)$

$= 3.745517$

$\bar{y}_1 - \bar{y}_2 = 9.8 - 15.4 = -5.6$ *

$\bar{y}_1 - \bar{y}_3 = 9.8 - 17.6 = -7.8$ *

$\bar{y}_1 - \bar{y}_4 = 9.8 - 21.6 = -11.8$ *

$\bar{y}_1 - \bar{y}_5 = 9.8 - 10.8 = -1$

$\bar{y}_2 - \bar{y}_3 = 15.4 - 17.6 = -2.2$

$\bar{y}_2 - \bar{y}_4 = 15.4 - 21.6 = -6.2$ *

$\bar{y}_2 - \bar{y}_5 = 15.4 - 10.8 = 4.6$ *

$\bar{y}_3 - \bar{y}_4 = 17.6 - 21.6 = -4$ *

$\bar{y}_3 - \bar{y}_5 = 17.6 - 10.8 = 6.8$ *

$\bar{y}_4 - \bar{y}_5 = 21.6 - 10.8 = 10.8$ *

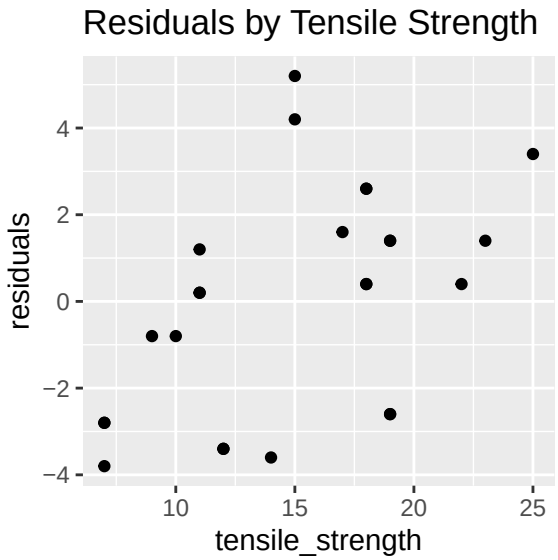
There is evidence to suggest 20, 25, and 30 are different from 15. 30 and 35 are different from 20. 30 and 35 are different from 25. 35 is different from 30.

```
library(tidyverse)
```

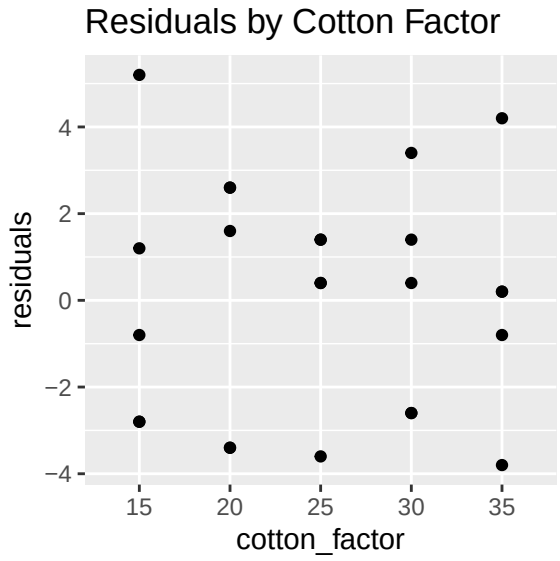
1)

C) Analyze the residuals from this experiment and comment on model adequacy. Use R.

```
cotton_data <- tibble(
  cotton_factor = gl(n = 5, k = 5, label = c("15", "20", "25", "30", "35")),
  tensile_strength = c(7, 7, 15, 11, 9, 12, 17, 12, 18, 18, 14, 19, 19, 18, 18, 19, 25, 22, 19, 23, 7,

cotton_anova <- aov(tensile_strength ~ cotton_factor, data = cotton_data)

cotton_data_clean <- cotton_data %>%
  mutate(residuals = cotton_anova$residuals)

ggplot(data = cotton_data_clean, aes(x = tensile_strength, y = residuals)) +
  geom_point() +
  labs(title = "Residuals by Tensile Strength")
```




```
ggplot(data = cotton_data_clean, aes(x = cotton_factor, y = residuals)) +
  geom_point() +
  labs(title = "Residuals by Cotton Factor")
```




```
qplot(cotton_data_clean$residuals) +
  stat_bin(bins = 5)
```

Overall, the residuals for problem 1, look fine. There's a random spattering around each of the cotton factors and the histogram of the residuals looks fine.

2) a)

Circuit Type	Response Time					
1	9	12	10	8	15	$\bar{y}_{1\cdot} = 10.8$
2	20	21	23	17	30	$\bar{y}_{2\cdot} = 22.2$
3	6	5	8	16	7	$\bar{y}_{3\cdot} = 8.4$

$\bar{y}_{\cdot\cdot} = 13.8$

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

Source	SS	df	MSE	F	p-value
Between Trt	543.61	2	271.8	16.083	0.0004
Error (within Trt)	202.80	12	16.9		
Total	746.41	14			

$SS_{Trt} = \frac{1}{n}\sum_{1}^{a} y_{i\cdot}^2 - \frac{y_{\cdot\cdot}^2}{N}$

$= \frac{1}{5}(54^2 + 111^2 + 42^2) - \frac{207^2}{15}$

$= \frac{1}{5}(17001) - 2856.6$

$= 543.61$

$SS_{Tot} = \sum_{1}^{a}\sum_{1}^{n} y_{ij}^2 - \frac{(y_{\cdot\cdot})^2}{N}$

$= 3603 - \frac{207^2}{15}$

$= 3603 - 2856.6$

$= 746.4$

$pF(16.083, 3, 12) = 0.0004$

With a p-value less than or equal to 0.01, there is evidence to suggest that one of the three means are different and we can reject the null hypothesis.

b) Use Tukey's test to compare pairs to treatment means. Note: With no stated α, I will use $\alpha = 0.05$.

$$T_{0.05} = q_{0.05}(3, 12)\sqrt{\frac{16.9}{5}} = (3.08)(1.838478) = 5.662512$$

$\bar{y}_1 - \bar{y}_2 = 10.8 - 22.2 = -11.4$ ✱

$\bar{y}_1 - \bar{y}_3 = 10.8 - 8.4 = 2.4$

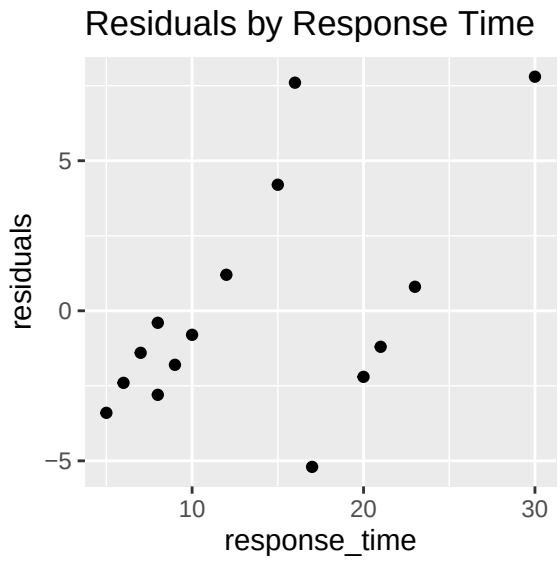
$\bar{y}_2 - \bar{y}_3 = 22.2 - 8.4 = 13.8$ ✱

With a p-value less than 0.05, there is evidence to suggest we reject the null hypothesis. There is evidence to suggest that circuit type 2 is different than 1 or 3.

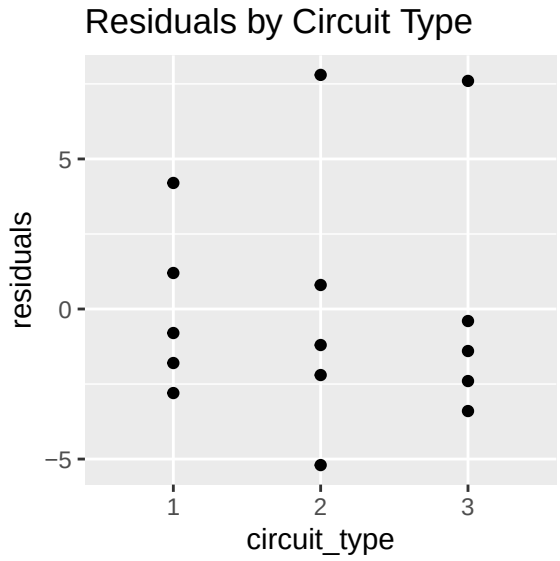
2)

C) Analyze the residuals from this experiment and comment on model adequacy. Use R.

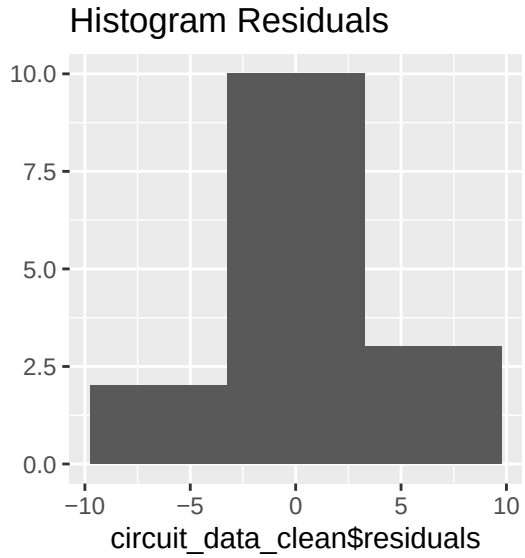
```
circuit_data <- tibble(
  circuit_type = gl(n = 3, k = 5, labels = c("1", "2", "3")),
  response_time = c(9, 12, 10, 8, 15, 20, 21, 23, 17, 30, 6, 5, 8, 16, 7)
)

circuit_anova <- aov(response_time ~ circuit_type, data = circuit_data)

circuit_data_clean <- circuit_data %>%
  mutate(residuals = circuit_anova$residuals)

ggplot(data = circuit_data_clean, aes(x = response_time, y = residuals)) +
  geom_point() +
  labs(title = "Residuals by Response Time")
```




```
ggplot(data = circuit_data_clean, aes(x = circuit_type, y = residuals)) +
  geom_point() +
  labs(title = "Residuals by Circuit Type")
```




```
qplot(circuit_data_clean$residuals) +
  stat_bin(bins = 3) +
  labs(title = "Histogram Residuals")
```



Overall, the residuals for problem 1, look fine. There's a random spattering around each of the circuit type and the histogram of the residuals looks fine.

3)

Batch 1	Batch 2	Batch 3	Batch 4	Batch 5
23.46	23.59	23.51	23.28	23.29
23.48	23.46	23.64	23.40	23.46
23.56	23.42	23.46	23.37	23.37
23.39	23.49	23.52	23.46	23.32
23.40	23.50	23.49	23.39	23.38

a) Is this a random effects model? Why?

This is a random effects model. There is no deterministic decision made by the statistician in selecting the batches. If they were to take a bunch of batches from the night shift, this would be a fixed effects model where all results can only be translated to the results of night shifts. With them being randomly selected, this is a random effects model.

b) Is there a significant variation in calcium content from batch to batch? Use $\alpha = 0.05$.

Source	SS	df	MS	F	P-value
Treatment	0.09698	4	0.02425	5.535	0.00363
Error	0.0876	20	0.00438		
Total	0.18458	24			

$$SS_{Trt} = \frac{1}{n}\sum_{i=1}^{a} y_{i.}^2 - \frac{(y_{..})^2}{N}$$
$$= \frac{1}{5}(68700.7825) - \frac{343501.4901}{25}$$
$$= 13740.1565 - 13740.05952$$
$$= \boxed{0.09698}$$

$$SS_{Tot} = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \frac{(y_{..})^2}{N}$$
$$= 13,740.2441 - 13740.05952$$
$$= 0.18458$$

With a p-value less than .005, there is evidence to reject the null hypothesis that all the means are the same. There is evidence to suggest that at least one mean is different.

c) Estimate the variance components for this model.

$$\hat{\sigma}^2 = MS_E = 0.00438$$
$$\hat{\sigma}_\tau^2 = \frac{MS_{Trt} - MS_E}{n} = \frac{0.02425 - 0.00438}{5} = \boxed{0.003974}$$
$$Var[y_{ij}] = 0.00438 + 0.02425 = \boxed{0.028605}$$

d) Find a 95% Confidence Interval for $\frac{\sigma_\tau^2}{(\sigma_\tau^2+\sigma^2)}$

$$\frac{L}{1+L} \leq \frac{\sigma_T^2}{\sigma_T^2+\sigma^2} \leq \frac{U}{1+U}$$

$$L = \frac{1}{n}\left(\frac{MS_{Trt}}{MS_\varepsilon}\left(\frac{1}{F_{0.025,4,21}}\right)-1\right), \quad U = \frac{1}{n}\left(\frac{MS_{Trt}}{MS_\varepsilon}\left(\frac{1}{F_{0.975,4,21}}\right)-1\right)$$

$$U = \frac{1}{5}\left(\frac{0.0242}{0.00238}\left(\frac{1}{0.11770}\right)-1\right), \quad L = \frac{1}{5}\left(\frac{0.0242}{0.00238}\left(\frac{1}{3.51}\right)-1\right)$$

$U = 9.2435$, $L = 0.11756$

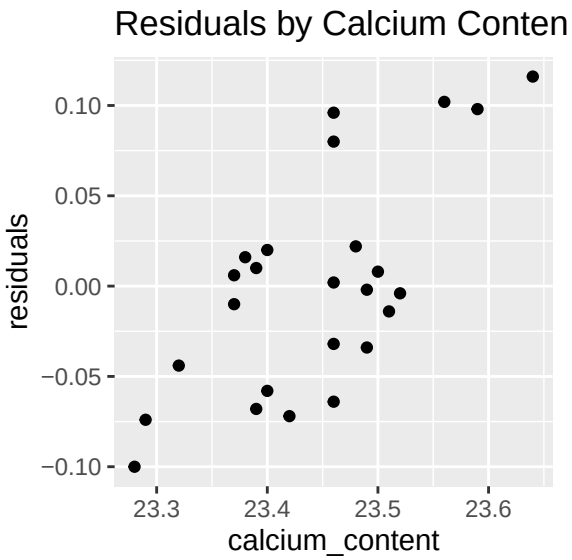
$$U = \frac{9.2435}{10.2435} = 0.902377 \qquad L = 0.1148$$

95% Confidence Interval for ICC is between 0.1148 & 0.90238

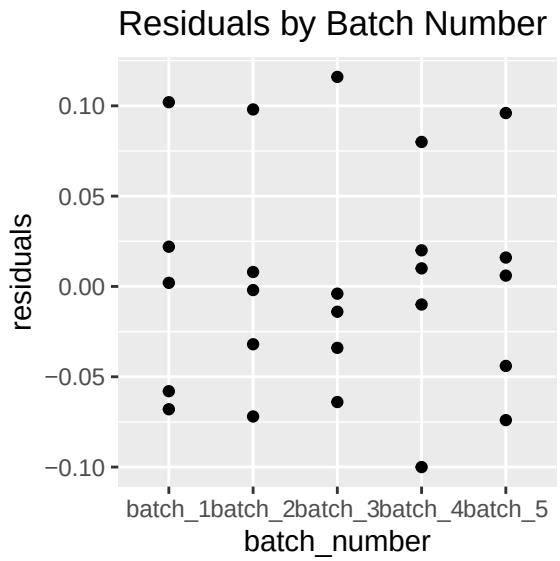
3)

E) Analyze the residuals from this experiment and comment on model adequacy. Use R.

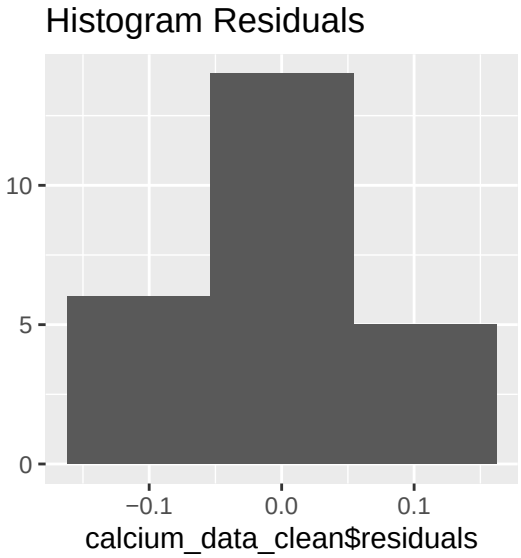
```
calcium_data <- tibble(
  batch_number = gl(n = 5, k = 5, labels = c("batch_1", "batch_2", "batch_3",
                                             "batch_4", "batch_5")),
  calcium_content = c(23.46, 23.48, 23.56, 23.39, 23.40, 23.59, 23.46, 23.42, 23.49, 23.50, 23.51, 23.64
)

calcium_anova <- aov(calcium_content ~ batch_number, data = calcium_data)

calcium_data_clean <- calcium_data %>%
  mutate(residuals = calcium_anova$residuals)

ggplot(data = calcium_data_clean, aes(x = calcium_content, y = residuals)) +
  geom_point() +
  labs(title = "Residuals by Calcium Content")
```




```
ggplot(data = calcium_data_clean, aes(x = batch_number, y = residuals)) +
  geom_point() +
  labs(title = "Residuals by Batch Number")
```




```
qplot(calcium_data_clean$residuals) +
  stat_bin(bins = 3) +
  labs(title = "Histogram Residuals")
```



These residuals are the best out of the three problems. Their spattering by batch number and nice histogram make for an awesome model and we could proceed to post-hoc testing bearing a check on the p-value.