Homework 7

Brady Lamson

Section 11.2

Problem 16

Factor A - Curing time ($I = 3$)

Factor B - Type of Mix ($J = 4$)

3 Observations for each, 36 total sample size.

$$SSA = 30763.0$$
$$SSB = 34185.6$$
$$SSE = 97436.8$$
$$SST = 205966.6$$

```r
# Sums of Squares ---
ssa <- 30763
ssb <- 34185.6
sse <- 97436.8
sst <- 205966.6
ssab <- sst - (ssa + ssb + sse)
sums <- c(ssa,ssb,sse,ssab,sst)

# Categories and samples ---
i <- 3
j <- 4
k <- 3

# Degrees of freedom ---
dfa <- i-1
```

```r
dfb <- j-1
dfe <- i*j*(k-1)
dfab <- (i-1)*(j-1)
dft <- (i*j*k)-1
dfs <- c(dfa, dfb, dfe, dfab, dft)

# Mean Square Ratios ---
mean_square_names <- c("MSA", "MSB", "MSE", "MSAB")
mean_squares <- round(sums[1:4] / dfs[1:4], 3)

# Test Statistics ---
test_stats <- (mean_squares[c(1,2,4)] / mean_squares[3]) |> round(3)
test_names <- c("fa", "fb", "fab")

# p-values ---
p_val <- pf(q = test_stats[1:3], df1 = dfs[c(1,2,4)], df2 = dfs[3], lower.tail = F)

p_val_names <- c("a", "b", "ab")
```

```
MSA = 15381.5
MSB = 11395.2
MSE = 4059.867
MSAB = 7263.533

fa = 3.789
fb = 2.807
fab = 1.789

P-Value a: 0.037
P-Value b: 0.061
P-Value ab: 0.144
```

	DF	SS	MS	f	p
Curing Time	2	30763	15381.5	3.789	0.037
Type of Mix	3	34185.6	11395.2	2.807	0.061
Interaction	6	43581.2	7263.533	1.789	0.144
Error	24	97436.6	4059.867		
Total	35	205966.6			

Utilizing all of the information collected above to generate the ANOVA table, we can draw the following conclusions:

$$H_{0AB} : \text{no interaction of factors} H_{aAB} : \text{interaction of factors} F = 1.789 p = 0.144$$

There is not statistically significant evidence to reject the null hypothesis, H_{0AB}, that there is no interaction effect between curing time and type of mix at the 0.05 significance level.

$$H_{0A} : \text{Factor A main effects are absent} H_{aA} : \text{Factor A has a main effect} F = 3.789 p = 0.037$$

There is statistically significant evidence at the 0.05 level to reject the null hypothesis that factor A, curing time, has no effect on the compression strength of hardened cement cubes. It is reasonable to suggest that curing time has a positive impact on compression strength.

$$H_{0B} : \text{Factor B main effects are absent} H_{aB} : \text{Factor B has a main effect} F = 2.807 p = 0.061$$

There is not statistically significant evidence at the 0.05 level to reject the null hypothesis that type of mix has no effect on the compression strength of hardened cement cubes.

Problem 17

Part A

$$SS_{sand} = 705 SS_{fiber} = 1278 SSE = 843 SST = 3105 SS_{AB} = 279 \alpha = 0.05$$

$$\begin{aligned}
MSA &= \frac{SSA}{df_a} = \frac{705}{2} = 352.5 \\
MSB &= \frac{1278}{2} = 639 \\
MSAB &= \frac{279}{4} = 69.75 \\
MSE &= \frac{843}{9} = 93.67 \\
F_b &= \frac{MSB}{MSE} = 6.82 \\
p_b &= 0.016
\end{aligned}$$

At the significance level 0.05 there does appear to be a statistically significant impact on wet-mold strength based on the carbon fiber addition.

Part B

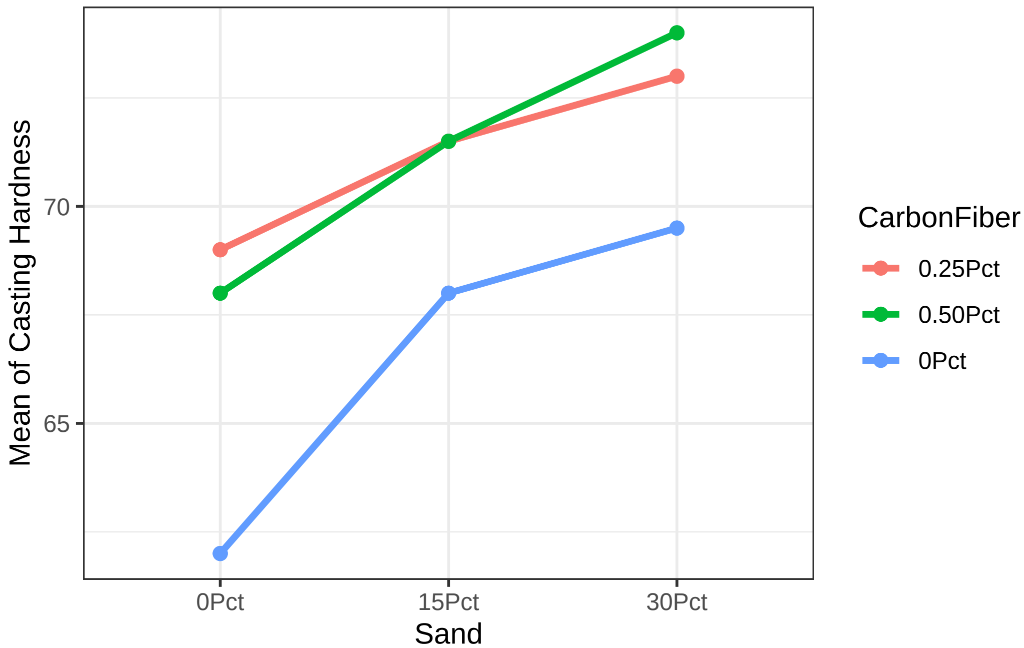
```r
df <- readr::read_table("ex_11_17.txt")
```

```
-- Column specification --------------------------------------------------------
cols(
  Sand = col_character(),
  CarbonFiber = col_character(),
  CastingHardness = col_double(),
  WetMoldStrength = col_double()
)
```

```
                 Df Sum Sq Mean Sq F value Pr(>F)
Sand              2 106.78   53.39   6.537 0.0176 *
CarbonFiber       2  87.11   43.56   5.333 0.0297 *
Sand:CarbonFiber  4   8.89    2.22   0.272 0.8887
Residuals         9  73.50    8.17
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

At the significance level 0.05, there is not a statistically significant interaction effect for sand and carbon fiber on the casting hardness. Though, at the same significance level there are main effects for both sand and carbon fiber on casting hardness. Though we fail to reject the null hypothesis for the interaction effect, the null is rejected for both main effects.

```r
df %>%
    group_by(Sand, CarbonFiber) %>%
    mutate(avg = mean(CastingHardness)) %>%
    ggplot(aes(x=Sand, y=avg)) +
    geom_line(size=1.2, aes(group=CarbonFiber, color=CarbonFiber)) +
    geom_point(size=2, aes(color=CarbonFiber)) +
    theme_bw() +
    ylab("Mean of Casting Hardness")
```



We can see that the interaction effect isn't statistically significant as they all follow roughly the same trajectory. They aren't exactly parallel but they're close.

What we can see though is that Sand has an impact because the lines aren't horizontal and carbon fiber has an impact as the lines aren't completely overlapping. it is of note here that carbon fiber likely wouldn't be statistically significant if 0Pct wasn't there. 0.25 and 0.50 are very close to overlapping.

Problem 18

```
df <- read.table("ex_11_18.txt", header = T) %>% tibble()
df %>% head(5)
```

```
# A tibble: 5 x 3
  Formulation Speed   Yield
  <chr>       <chr>   <dbl>
1 One         Sixty    190.
2 One         Sixty    189.
3 One         Sixty    190.
4 One         Seventy  185.
5 One         Seventy  179.
```

```r
my_aov_18 <-
    aov(Yield ~ Formulation + Speed + Formulation:Speed, data = df)

my_aov_18 %>% summary()
```

```
                  Df Sum Sq Mean Sq F value   Pr(>F)    
Formulation        1 2253.4  2253.4 376.271 1.99e-10 ***
Speed              2  230.8   115.4  19.270 0.000179 ***
Formulation:Speed  2   18.6     9.3   1.551 0.251639    
Residuals         12   71.9     6.0                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Part A

There does not appear to be interaction between the factors.

Part C

```r
df %>%
    group_by(Formulation) %>%
    summarise(avg = mean(Yield)) %>%
    mutate(effect = avg - mean(df$Yield))
```

```
# A tibble: 2 x 3
  Formulation   avg effect
  <chr>       <dbl>  <dbl>
1 One          187.   11.2
2 Two          165.  -11.2
```

The effect of Formulation One appears to be 11.19 and the effect of Formulation Two appears to be -11.19

```r
df %>%
    group_by(Speed) %>%
    summarise(avg = mean(Yield)) %>%
    mutate(effect = avg - mean(df$Yield))
```

```
# A tibble: 3 x 3
  Speed     avg effect
  <chr>   <dbl>  <dbl>
1 Eighty   179.   3.04
2 Seventy  171.  -5.03
3 Sixty    178.   1.99
```

The effects of Eighty, Seventy and Sixty speed are 3.04, -5.03 and 1.99 respectively.

Part D

```r
resids_to_check <- c(.23, -.87, .63, 4.5, -1.2, -3.3, -2.03, 1.97, 0.07, -1.1, -.3, 1.4, .

all.equal.default(
    target = resids_to_check,
    # Round the ANOVAs residuals to the same digits as the book and convert to vector
    current = as.vector(my_aov_18$residuals %>% round(2), mode = "numeric")
)
```

```
[1] TRUE
```

The residuals in the book match the residuals from the ANOVA.

Part E

The residuals do not appear to be normally distributed. They don't stay on the line well enough and follow a snaking S shape.

Problem 20

```r
df <- readr::read_table("ex_11_20.txt")
```

```
-- Column specification --------------------------------------------------------
cols(
  Adhesive = col_character(),
  Condition = col_character(),
```

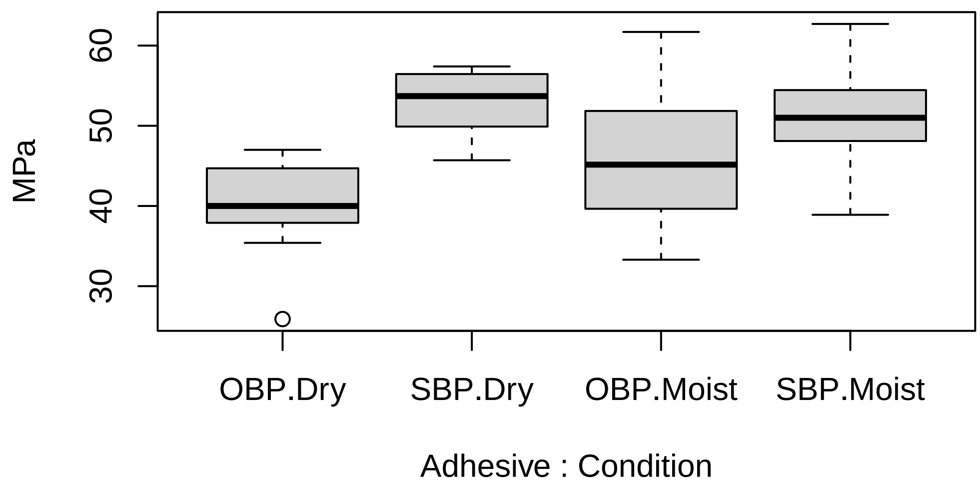
```
  MPa = col_double()
)
```

```
df %>% head(5)
```

```
# A tibble: 5 x 3
  Adhesive Condition   MPa
  <chr>    <chr>     <dbl>
1 SBP      Dry        56.7
2 SBP      Dry        57.4
3 SBP      Dry        53.4
4 SBP      Dry        54
5 SBP      Dry        49.9
```

Part A

```
boxplot(MPa ~ Adhesive + Condition + Adhesive:Condition, data = df)
```



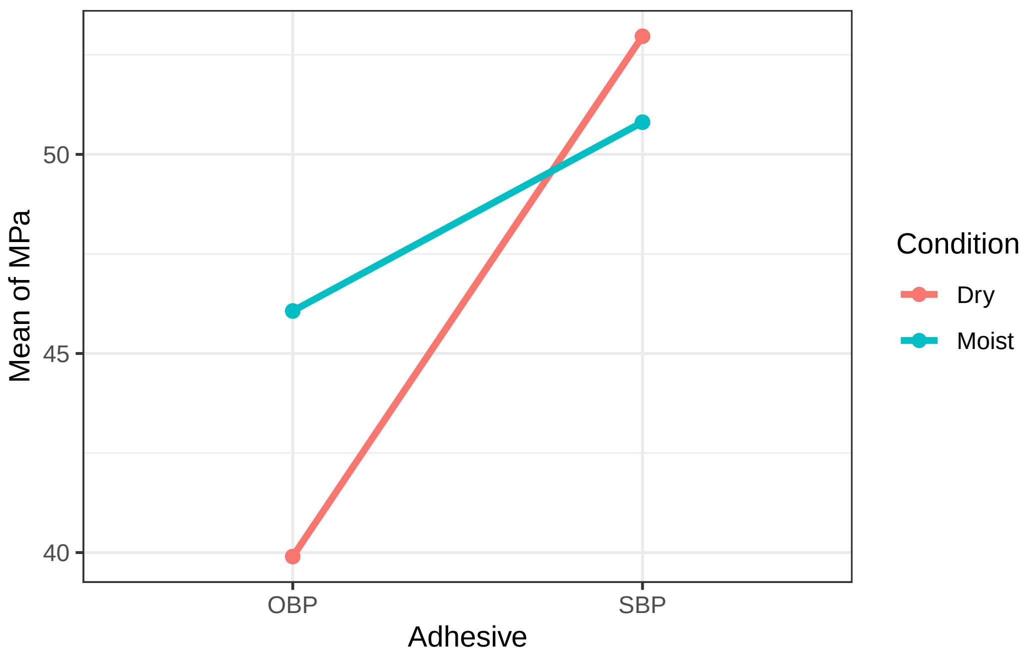
Part B

```
my_aov <-
    aov(MPa ~ Adhesive + Condition + Adhesive:Condition, data = df)

my_aov %>%
```

```
    summary()
```

```
                   Df Sum Sq Mean Sq F value   Pr(>F)    
Adhesive            1  951.4   951.4  22.846 1.99e-05 ***
Condition           1   48.2    48.2   1.157   0.2879    
Adhesive:Condition  1  207.9   207.9   4.993   0.0306 *  
Residuals          44 1832.3    41.6                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



At the 0.05 significance level there isn't substantial evidence to reject the null hypothesis that there is no interaction effect between adhesive and condition on MPa.

Part C

```
df <-
    df %>%
    mutate(
        Treatment = paste(Adhesive, Condition)
    )
```

Significant differences are found between:

- SBP Dry - OBP Dry
- SBP Moist - OBP Dry

Section 11.3

Problem 1

```
df <- readr::read_table("rice.txt")
```

```
-- Column specification --------------------------------------------------------
cols(
  Sunshine = col_character(),
  Variety = col_character(),
  Fertilizer = col_character(),
  Yield = col_double()
)
```

```
df %>% head(5)
```

```
# A tibble: 5 x 4
  Sunshine Variety  Fertilizer Yield
  <chr>    <chr>    <chr>      <dbl>
1 Low      Variety1 Low           86
2 Low      Variety1 Low          115
3 Low      Variety1 Low           83
4 Low      Variety1 Low           70
5 Low      Variety1 High         109
```

$$X_{ijkl} = \mu + \alpha_i + \beta_k + \delta_k + \gamma_{ij}^{AB} + \gamma_{ik}^{AC} + \gamma_{jk}^{BC} + \gamma_{ijk}^{ABC} + \epsilon_{ijkl}$$

Where Factor A, B and C represent sunshine, variety and fertilizer respectively. It is assumed that the ϵ_{ijkl}'s are iid $N(0, \sigma)$.

$$H_{0A} : \alpha_i = 0 \ \forall i$$
$$H_{aA} : \text{Not all } \alpha_i = 0$$

$$H_{0B} : \beta_j = 0 \ \forall j$$
$$H_{aB} : \text{Not all } \beta_j = 0$$

$$H_{0C} : \delta_k = 0 \ \forall k$$
$$H_{aC} : \text{Not all } \delta_k = 0$$

$$H_{0AB} : \gamma_{ij}^{AB} = 0 \ \forall i, j$$
$$H_{aAB} : \text{Not all } \gamma_{ij}^{AB} = 0$$

$$H_{0AC} : \gamma_{ik}^{AC} = 0 \ \forall i, k$$
$$H_{aAC} : \text{Not all } \gamma_{ik}^{AC} = 0$$

$$H_{0BC} : \gamma_{jk}^{BC} = 0 \ \forall j, k$$
$$H_{aBC} : \text{Not all } \gamma_{jk}^{BC} = 0$$

$$H_{0ABC} : \gamma_{ijk}^{ABC} = 0 \ \forall i, j$$
$$H_{aABC} : \text{Not all } \gamma_{ijk}^{ABC} = 0$$

```
my_aov_1 <- aov(Yield ~ Sunshine*Variety*Fertilizer, data = df)

my_aov_1 %>%
    summary()
```

```
                            Df Sum Sq Mean Sq F value   Pr(>F)    
Sunshine                     1   1313    1313   5.301 0.027209 *  
Variety                      2     71      35   0.143 0.867146    
Fertilizer                   1   3834    3834  15.484 0.000364 ***
Sunshine:Variety             2   4083    2041   8.244 0.001128 ** 
Sunshine:Fertilizer          1     20      20   0.081 0.777773    
Variety:Fertilizer           2      8       4   0.016 0.984231    
Sunshine:Variety:Fertilizer  2    736     368   1.486 0.239884    
Residuals                   36   8914     248                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Part Vi

The three factor interaction is not statistically significant at any reasonable significance level. $p = 0.240$, $F = 1.486$.

As the three factor interaction is not significant we may proceed to the two factor interactions.

Fertilizer is not involved in any significant interaction effects. $p = \{0.984, 0.777\}$, $F = \{0.016, 0.081\}$. Notation: first element in the set corresponds to fertilizers relationship with variety of rice, the second with sunshine.

Sunshine does have a significant interaction effect with the variety of rice. $p = 0.001128$, $F = 8.244$.

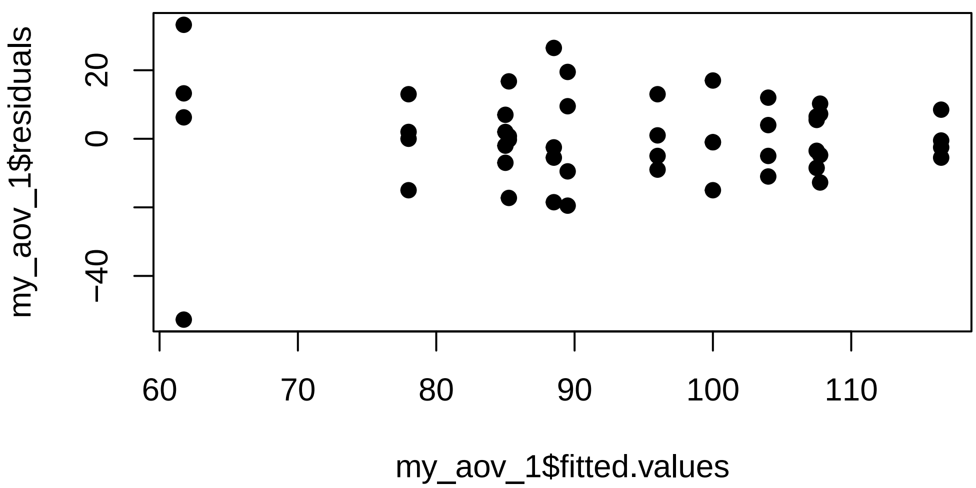
Based on the above results, it is only appropriate to examine the main effect of Fertilizer as it was not involved in any interaction effects.

Fertilizer does have a significant impact on the yield as a main effect. $p = 0.000364$, $F = 15.484$.

Variety and sunshine have an effect on the yield as they were involved in an interaction effect together. We need not examine their main effects.

Part Vii

Based on the plots, mostly the the normal probability plot, the residuals do not appear to follow a normal distribution. It's really close though and I'm unsure if I'm being too picky.



The residuals do not appear to follow a constant standard deviation. They appear to get smaller as we proceed along the fitted value axis.

Problem 27

Part A

- Factor A: Temperature profile ($I = 3$)
- Factor B: Type of Plastic ($J = 3$)
- Factor C: Speed ($K = 3$)
- Two observations per ($L = 2$)

$$
\begin{aligned}
SSA &= 14144.44 \\
SSB &= 5511.27 \\
SSC &= 224696.39 \\
SSAB &= 1069.62 \\
SSAC &= 62.67 \\
SSBC &= 331.67 \\
SSE &= 3127.50 \\
SST &= 270024.33
\end{aligned}
$$

I will be using code to construct this table as I do not have any desire to do this manually.

```
# Sums of Squares ---
SSA = 14144.44
SSB = 5511.27
SSC = 244696.39
SSAB = 1069.62
SSAC = 62.67
SSBC = 331.67
SSE = 3127.50
SST = 270024.33
SSABC <- SST - sum(SSA,SSB,SSC,SSAB,SSAC,SSBC,SSE)
sums <- c(SSA,SSB,SSC,SSAB,SSAC,SSBC,SSABC,SSE,SST)

# Categories and samples ---
i <- 3
j <- 3
k <- 3
l <- 2

# Degrees of freedom ---
```

```r
dfa <- i-1
dfb <- j-1
dfc <- k-1
dfab <- dfa * dfb
dfac <- dfa * dfc
dfbc <- dfb * dfc
dfabc <- dfa * dfb * dfc
dfe <- i*j*k*(l-1)
dft <- prod(i,j,k,l) - 1
dfs <- c(dfa, dfb, dfc, dfab, dfac, dfbc, dfabc, dfe, dft)

# Mean Square Ratios ---
mean_square_names <- c("msa", "msb", "msc", "msab", "msac", "msbc", "msabc", "mse")
mean_squares <- round(sums[1:8] / dfs[1:8], 3)

# Test Statistics ---
test_stats <- (mean_squares[1:7] / mean_squares[8]) |> round(3)
test_names <- c("fa", "fb", "fc", "fab", "fac", "fbc", "fabc")

# p-values ---
p_val <- pf(q = test_stats, df1 = dfs[1:7], df2 = dfs[8], lower.tail = F)

p_val_names <- c("a", "b", "c", "ab", "ac", "bc", "abc")
```

```
---[Mean Squares]---

MSA = 7072.22
MSB = 2755.635
MSC = 122348.195
MSAB = 267.405
MSAC = 15.668
MSBC = 82.918
MSABC = 135.096
MSE = 115.833

---[Test Statistics]---

fa = 61.055
fb = 23.79
fc = 1056.246
```

```
fab = 2.309
fac = 0.135
fbc = 0.716
fabc = 1.166

---[P-Values]---

P-Value a: 0
P-Value b: 0
P-Value c: 0
P-Value ab: 0.084
P-Value ac: 0.968
P-Value bc: 0.588
P-Value abc: 0.355
```

For the sake of my own sanity I will not be writing this out into an entire table. The values themselves should suffice. If they don't, oh well I did my best.

Part B

Examing the F statistics and p-values for the two and three factor interactions show us that, at the 0.05 significance level, none of them are significant. The f statistics are $F = \{2.31, 0.135, 0.716, 1.166\}$ and the p-values are $p = \{0.084, 0.968, 0.588, 0.355\}$. As we can see, none of those are below 0.05.

Part C

Factors A, B and C all have p-values that are essentially 0 (they show zero because they're rounded). As such, they are all lower than 0.05 and have a significant effect.

Problem 29

For this problem I utilized the same code at Problem 27, if there was any confusion as to how the values were acquired.

Part A

- Factor A: Treatment ($I = 3$)
- Factor B: Fabric Type ($J = 2$)
- Factor C: Number of Cycles ($K = 3$)
- Five observations per ($L = 5$)

$$\begin{aligned} SSA &= 1043.27 \\ SSB &= 112148.10 \\ SSC &= 3020.97 \\ SSAB &= 373.52 \\ SSAC &= 392.71 \\ SSBC &= 145.95 \\ SSABC &= 54.13 \\ SSE &= 339.30 \end{aligned}$$

```
---[Mean Squares]---

MSA = 521.635
MSB = 112148.1
MSC = 1510.485
MSAB = 186.76
MSAC = 98.177
MSBC = 72.975
MSABC = 13.533
MSE = 4.713

---[Test Statistics]---

fa = 110.68
fb = 23795.481
fc = 320.493
fab = 39.627
fac = 20.831
fbc = 15.484
fabc = 2.871

---[P-Values]---
```

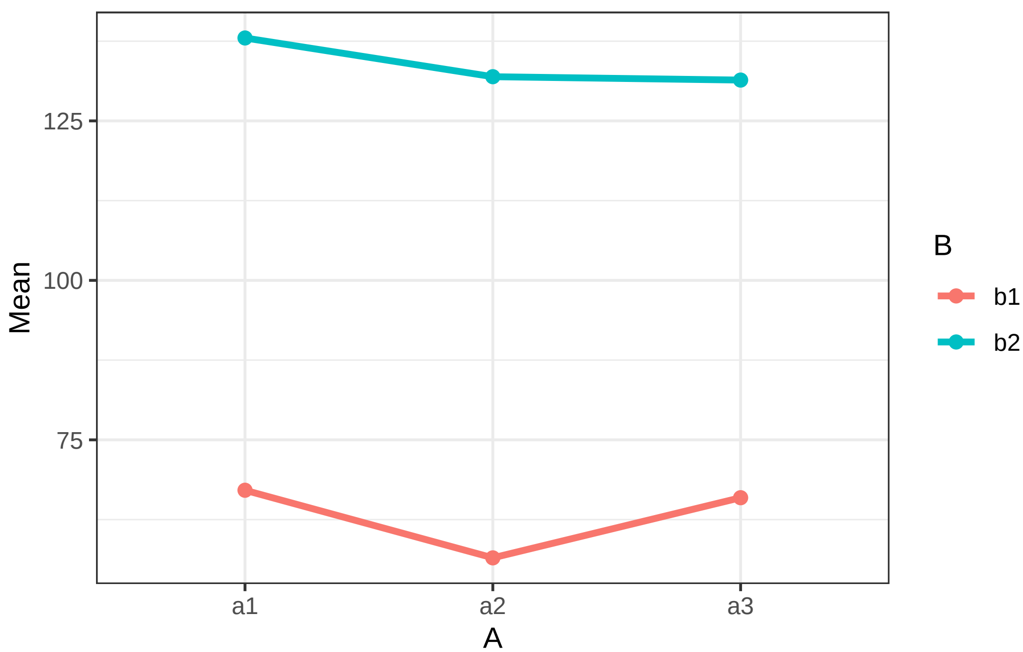
```
P-Value a: 0
P-Value b: 0
P-Value c: 0
P-Value ab: 0
P-Value ac: 0
P-Value bc: 0
P-Value abc: 0.029
```

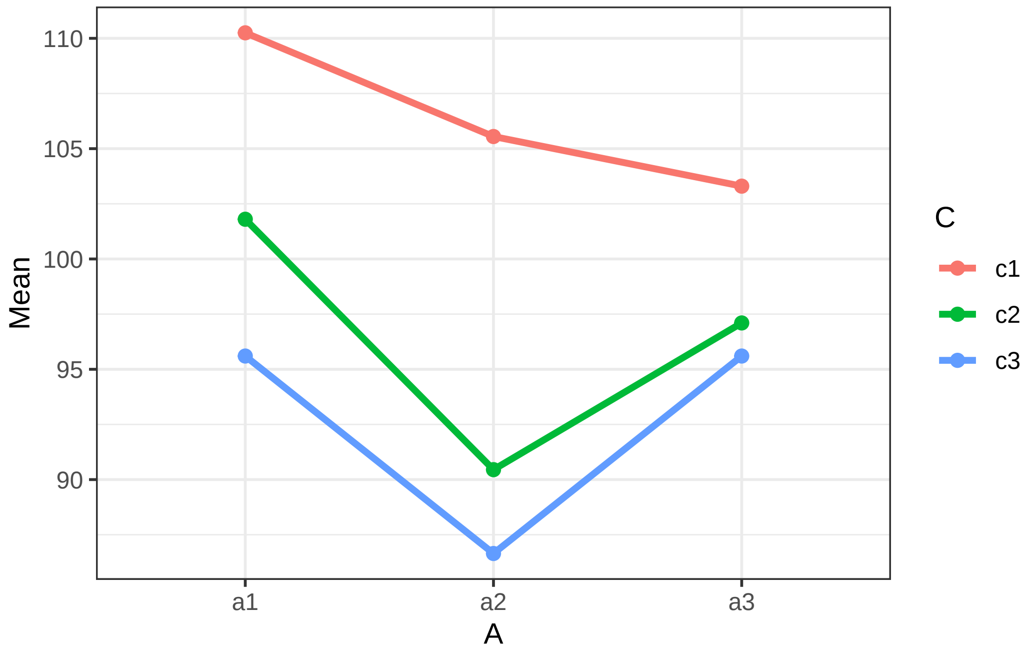
Yet again, I will not be converting this into an ANOVA table. If that docks me points I totally understand.

As for the tests, only the triple interaction is not significant ($F = 2.871, p = 0.029$) at the 0.01 level. All of the double interactions were highly significant (rounded to 0 at 3 digits).

Part B

```
# A tibble: 5 x 5
  interaction A     B     means C    
  <chr>       <chr> <chr> <dbl> <chr>
1 a1 b1       a1    "b1"   67.1 ""   
2 a1 b2       a1    "b2"  138   ""   
3 a1 c1       a1    ""    110.  "c1" 
4 a1 c2       a1    ""    102.  "c2" 
5 a1 c3       a1    ""     95.6 "c3" 
```





Honestly a lot of these lines appear fairly close to parallel. I'm surprised to see such high significance levels for pretty similar looking plots. With AB it does make sense though, as they're very far apart on the mean axis. AC also probably skates by on the vertical difference.

Problem 31

```
df <- read.table("ex_11_31.txt", header = T) %>% tibble()
df %>% head(5)
```

```
# A tibble: 5 x 4
  Power      Speed       PasteThickness NickelWt
  <chr>      <chr>       <chr>             <dbl>
1 SixHundred SixHundred  TwoTenths          38.6
2 SixHundred SixHundred  ThreeTenths        35.1
3 SixHundred SixHundred  FourTenths         19.2
4 SixHundred NineHundred TwoTenths          38.2
5 SixHundred NineHundred ThreeTenths        34.2
```

```
my_aov_31 <- aov(NickelWt ~ Power*Speed*PasteThickness - Power:Speed:PasteThickness, data=

my_aov_31 %>% summary()
```

```
                     Df Sum Sq Mean Sq F value Pr(>F)
Power                 2  124.6   62.30   4.849 0.0417 *
Speed                 2   20.6   10.30   0.802 0.4815
PasteThickness        2  356.9  178.47  13.892 0.0025 **
Power:Speed           4   57.5   14.37   1.119 0.4118
Power:PasteThickness  4   61.4   15.35   1.195 0.3834
Speed:PasteThickness  4   11.1    2.76   0.215 0.9226
Residuals             8  102.8   12.85
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Part B

The p-values for all of the double interactions are well above 0.05, $\{0.9, 0.38, 0.4\}$. As such, none of them are significant.

Part C

Paste Thickness and Power are both significant at the 0.05 level $\{0.0025, 0.0417\}$.

Part D

ThreeTenths and FourTenths have a significant difference with a p-value of 0.025. TwoTenths and FourTenths have a significant difference with a p-value of 0.002.