

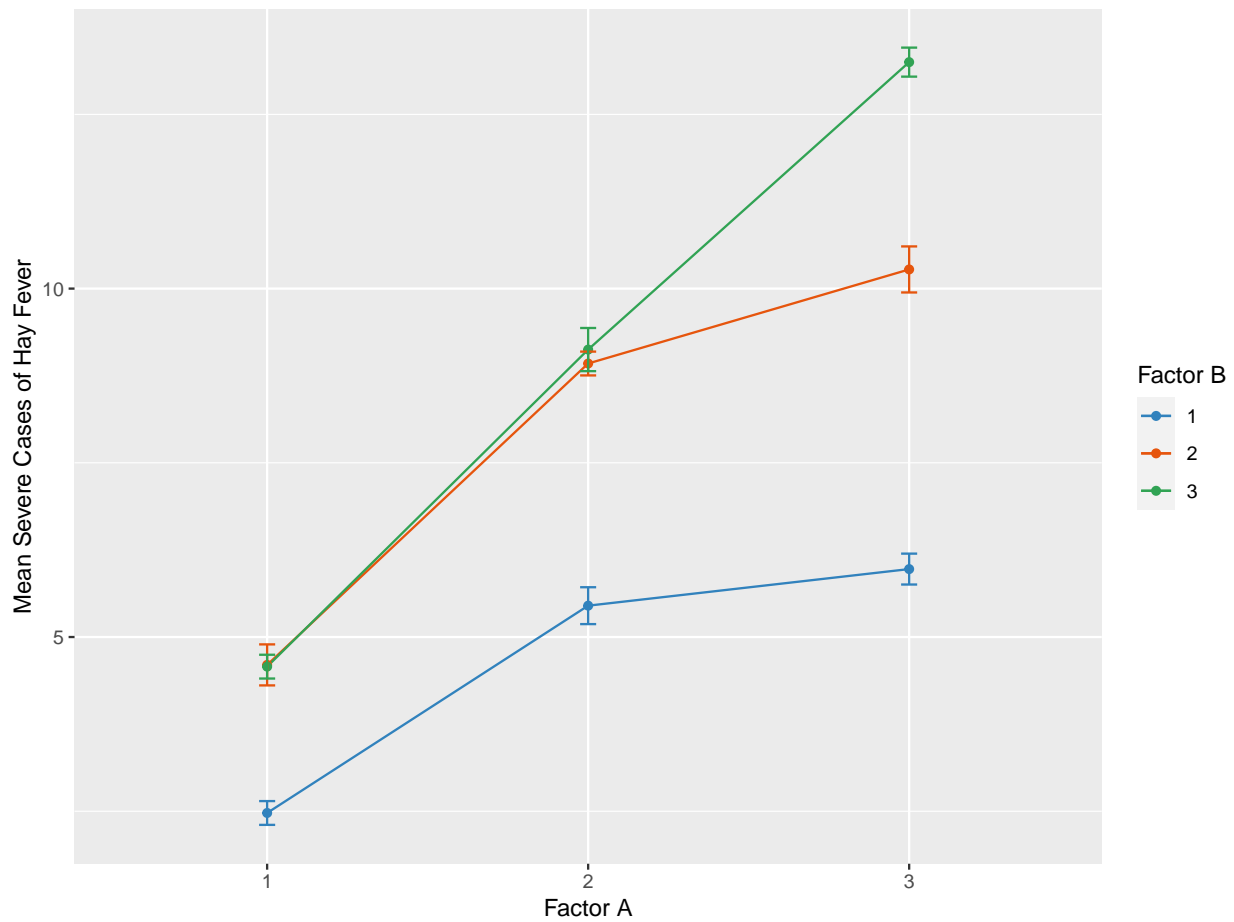
STP 531  
Applied Analysis of Variance  
Homework 5

Nathan A. Nguyen

29 March 2021

## Question 19.15

### Part A



There appears to be factor effects present. As the levels of factor A “increase”, the estimated treatment means, generally, increased though factor B (level 3) influences the degree of severe fever relief different than factor B (levels 1,2). This leads to another observation in that, there appears to be some interaction between the two factors as well e.g., the tapering off of the lines for the orange and blue curves compared to the constant green line.

## Part C

Table 1: Hay Fever ANOVA Table

	Df	Sum.Sq	Mean.Sq	F.value	Pr..F.
A	2	220.020	110.0100000	1827.8585	0
B	2	123.660	61.8300000	1027.3292	0
A:B	4	29.425	7.3562500	122.2269	0
Residuals	27	1.625	0.0601852	NA	NA

$$H_0 : (\alpha\beta)_{ij} = 0$$

$$H_a : \text{not all } (\alpha\beta)_{ij} = 0$$

$$F^* = \frac{MSAB}{MSE} = \frac{7.35}{0.06} \approx 122.5$$

$$F_{critical} = F[1 - \alpha; (a - 1)(b - 1), (n - 1)ab] = F[0.95; 4, 27] \approx 2.73$$

$$\text{if } F^* \leq F_{critical}, \text{conclude } H_0$$

$$\text{if } F^* > F_{critical}, \text{conclude } H_a$$

$$p = 0^+$$

Since  $F^* > F_{critical}$ , we reject the null hypothesis. There is sufficient evidence to suggest that there are interactions between factor A and factor B present.  $p \approx 6.7064399 \times 10^{-17}$ .

## Part D

$$H_0 : \text{all } \alpha_i = 0$$

$$H_a : \text{not all } \alpha_i = 0$$

$$F^* = \frac{MSA}{MSE} = \frac{110.010}{0.060} \approx 1833.5$$

$$F_{critical} = F[0.95; (a - 1), (n - 1)ab] = F(0.95; 2, 27) \approx 3.35$$

$$p = 0^+, F^* \approx 1833.5 > 3.35 = F_{critical}$$

$$\text{if } F^* \leq F_{critical}, \text{conclude } H_0$$

$$\text{if } F^* > F_{critical}, \text{conclude } H_a$$

Since  $F^* \approx 1833.5 > 3.35 = F_{critical}$ , we reject the null hypothesis. There is sufficient evidence to suggest that the factor A level means  $\mu_{i.}$  are all not equal. In other words, there exists main effects associated with ingredient 1 (factor A). The p-value is  $p \approx 1.4527898 \times 10^{-29}$ .

$$H_0 : \text{all } \beta_j = 0$$

$$H_a : \text{not all } \beta_j = 0$$

$$F^* = \frac{MSB}{MSE} = \frac{61.830}{0.060} \approx 1030.5$$

$$F_{critical} = F[0.95; (b - 1), (n - 1)ab] = F(0.95; 2, 27) \approx 3.35$$

$$p = 0^+, F^* \approx 1030.5 > 3.35 = F_{critical}$$

$$\text{if } F^* \leq F_{critical}, \text{conclude } H_0$$

$$\text{if } F^* > F_{critical}, \text{conclude } H_a$$

Since  $F^* \approx 1030.5 > 3.35 = F_{critical}$ , we reject the null hypothesis. There is sufficient evidence to suggest that main effects for factor B are present. The p-value is  $3.2140242 \times 10^{-26}$ .

## Question 19.32

Scheffe multiple comparison procedure with family confidence 0.90:

$$\begin{aligned}
 L_1 &= \frac{\mu_{12} + \mu_{13}}{2} - \mu_{11} \\
 L_2 &= \frac{\mu_{22} + \mu_{23}}{2} - \mu_{21} \\
 L_3 &= \frac{\mu_{32} + \mu_{33}}{2} - \mu_{31} \\
 L_4 &= L_2 - L_1 \\
 L_5 &= L_3 - L_1 \\
 L_6 &= L_3 - L_2 \\
 s^2\{\hat{L}\} &= \frac{0.060}{4} \sum (-1)^2 + (0.5)^2 + (0.5)^2 \approx 0.0225 \\
 s\{\hat{L}\} &\approx 0.15 \\
 S^2 &= (3 * 3 - 1)F[0.90; 8, 27] \approx 15.2 \\
 S &\approx 3.899 \text{ works for } i = 1, 2, 3
 \end{aligned}$$

## Part C

Table 2: 90% Scheffe Confidence for L1,L2,L3

	L1	L2	L3
Lower Limit	1.526296	2.988796	5.201295
Upper Limit	2.698704	4.161205	6.373704

Table 3: 90% Scheffe for L4, L5, L6

	L4	L5	L6
0.6334817	2.845982	1.383482	
2.2915183	4.504018	3.041518	

Table 4: Treatment Means

A	1	1	1	2	2	2	3	3	3
B	1	2	3	1	2	3	1	2	3
means	2.475	4.600	4.575	5.450	8.925	9.125	5.975	10.275	13.250



Question 23.22

23.22 a, b c

$$a.) H_0: L_i = \frac{(1)\mu_{i1} + 2\mu_{i2} + 7\mu_{i3}}{10} - \frac{\mu_{21} + 2\mu_{22} + 7\mu_{23}}{10} = 0$$

$$L_i = \frac{(1)\mu_{i1} + 2\mu_{i2} + 7\mu_{i3}}{10} - \frac{(1)\mu_{31} + 2\mu_{32} + 7\mu_{33}}{10} = 0$$

$$L_3 = \frac{(1)\mu_{11} + 2\mu_{12} + 7\mu_{13}}{10} - \frac{(1)\mu_{41} + 2\mu_{42} + 7\mu_{43}}{10} = 0$$

$H_a$ : Not all  $L_i = 0$ .

b.)  $a = 4, b = 3 \Rightarrow 12$  treatments  
1 trial per  $\rightarrow$

$$\beta' = \{\mu_{11}, \mu_{12}, \mu_{13}, \mu_{21}, \mu_{22}, \mu_{23}, \mu_{31}, \mu_{32}, \mu_{33}, \mu_{41}, \mu_{42}, \mu_{43}\}$$

	1	2	3	4	5	6	7	8	9	10	11	12
1	1	0	0	0	0	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0	0	0	0	0
3	0	0	1	0	0	0	0	0	0	0	0	0
4	0	0	0	1	0	0	0	0	0	0	0	0
5	0	0	0	0	1	0	0	0	0	0	0	0
6	0	0	0	0	0	1	0	0	0	0	0	0
7	0	0	0	0	0	0	1	0	0	0	0	0
8	0	0	0	0	0	0	0	1	0	0	0	0
9	0	0	0	0	0	0	0	0	1	0	0	0
10	0	0	0	0	0	0	0	0	0	1	0	0
11	0	0	0	0	0	0	0	0	0	0	1	0
12	0	0	0	0	0	0	0	0	0	0	0	1

$\neq X$

$$c) \underline{h}^T = [0, 0, 0] \quad (3 \times 2)$$

$$C = \begin{pmatrix} 0.1, 0.2, 0.7, -0.1, -0.2, -0.7, 0.0, 0.0, 0.0 \\ 0.1, 0.2, 0.7, 0.0, 0.0, -0.1, -0.2, -0.7, 0.0 \\ 0.1, 0.2, 0.7, 0.0, 0.0, 0.0, 0.0, 0.0, -0.1, -0.2, -0.7 \end{pmatrix}$$

$\underline{L}_1 = L_1$  Human vs. Social  
 $\underline{L}_2$  Human vs. Engineer  
 $\underline{L}_3$  Human vs. Management  
 across degree level.

$$H_0: C\underline{\beta} = \underline{h} \quad \underline{\beta} \text{ same as in part B.}$$

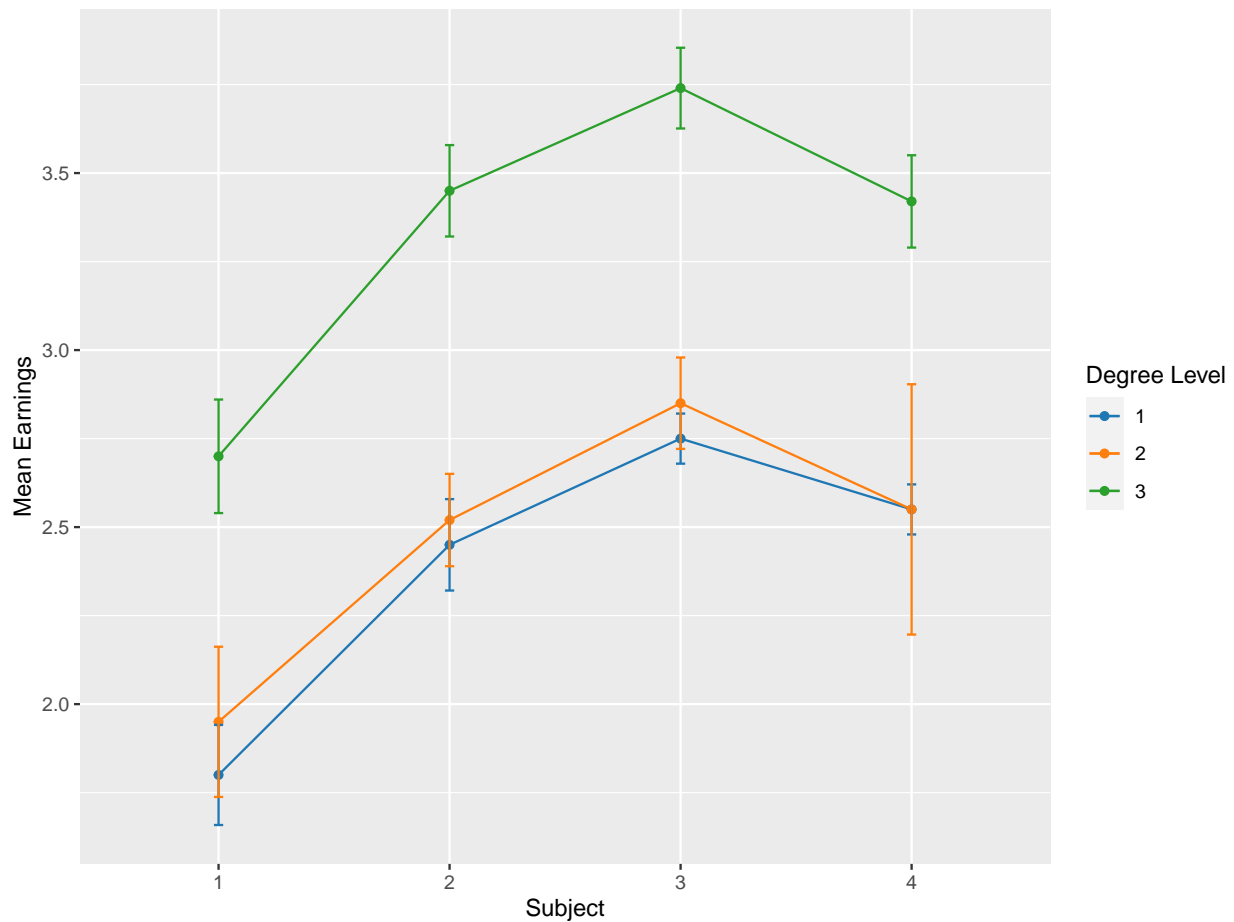
$(3 \times 12) (12 \times 1) \quad 3 \times 1$   
 $(3 \times 1)$

## Question 4 (data from 23.9)

### Part A

Table 5: Earnings ANOVA Type 3 SS

	Sum.Sq	Df	F.value	Pr..F.
(Intercept)	6.4800000	1	297.8272981	0.0000000
topic	1.0200000	3	15.6267409	0.0000017
degree	1.8375000	2	42.2266713	0.0000000
topic:degree	0.0444254	6	0.3403063	0.9104356
Residuals	0.7180000	33	NA	NA



There doesn't appear to be any serious interactions between subject matter and level of degrees, as the three curves are parallel for the most part. The gap start closes between subjects 3 and 4 (engineering and management) at degree levels 1 and 2 (bachelor's and master's). Whether or not the interaction is significant, is to be determined based on the Type III SS; however, again, the plot doesn't suggest any serious interactions between the two factors. There are main effects present.



## Part B

Table 6: Type 1 SS

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
topic	3	4.1675668	1.3891889	63.8485168	0.0000000
degree	2	8.3824522	4.1912261	192.6329541	0.0000000
topic:degree	6	0.0444254	0.0074042	0.3403063	0.9104356
Residuals	33	0.7180000	0.0217576	NA	NA

Table 7: Type 2 SS

	Sum Sq	Df	F value	Pr(>F)
topic	5.9792529	3	91.6041528	0.0000000
degree	8.3824522	2	192.6329541	0.0000000
topic:degree	0.0444254	6	0.3403063	0.9104356
Residuals	0.7180000	33	NA	NA

Table 8: Type 3 SS

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	6.4800000	1	297.8272981	0.0000000
topic	1.0200000	3	15.6267409	0.0000017
degree	1.8375000	2	42.2266713	0.0000000
topic:degree	0.0444254	6	0.3403063	0.9104356
Residuals	0.7180000	33	NA	NA

## Part C

Type 3 SS should be used over the other types of SS because in the unbalanced design, the sum of squares are no longer orthogonal. Type 3 SS tries to remove the confounding of the interaction effect as a result of the imbalance cells. A sequential approach would not be appropriate for unbalanced data.

I've read elsewhere that Type 2 can be used when there is no interaction. In this case, there is evidence of an interaction between the two factors, so would it be acceptable to use Type 2 SS?

The plot and ANOVA tables suggest that there are main effects present for factors A (subject) and B (degree level). Based on the graph, it appears that the higher level degree one has, the more one earns in an adjunct role. It also appears that the different subject matters influences earnings as well, e.g., a more STEM oriented degree (engineering) will earn more than a social science and humanities discipline.

## Part D

The least squares means should be reported because it adjusts for the differences in cell sizes. The standard mean can be misleading because it can be a confounded measure (inflated because of the unbalanced nature of the study). The standard mean may be too optimistic in the unbalanced case.

## Part E

```
## $lsmeans
## degree lsmean      SE df lower.CL upper.CL
## 1          2.39 0.0488 33      2.29      2.49
## 2          2.47 0.0444 33      2.38      2.56
## 3          3.33 0.0325 33      3.26      3.39
##
## Results are averaged over the levels of: topic
## Confidence level used: 0.95
##
## $contrasts
## contrast estimate      SE df t.ratio p.value
## 1 - 2          -0.08 0.0660 33   -1.213 0.4542
## 1 - 3          -0.94 0.0586 33  -16.042 <.0001
## 2 - 3          -0.86 0.0550 33  -15.635 <.0001
##
## Results are averaged over the levels of: topic
## P value adjustment: tukey method for comparing a family of 3 estimates

## $lsmeans
## topic lsmean      SE df lower.CL upper.CL
## 1          2.15 0.0522 33      2.04      2.26
## 2          2.81 0.0411 33      2.72      2.89
## 3          3.11 0.0479 33      3.02      3.21
## 4          2.84 0.0539 33      2.73      2.95
##
## Results are averaged over the levels of: degree
## Confidence level used: 0.95
##
## $contrasts
## contrast estimate      SE df t.ratio p.value
## 1 - 2          -0.6567 0.0664 33   -9.886 <.0001
## 1 - 3          -0.9633 0.0708 33  -13.601 <.0001
## 1 - 4          -0.6900 0.0750 33   -9.204 <.0001
## 2 - 3          -0.3067 0.0632 33   -4.856 0.0002
## 2 - 4          -0.0333 0.0678 33   -0.492 0.9604
## 3 - 4           0.2733 0.0721 33    3.791 0.0032
##
## Results are averaged over the levels of: degree
## P value adjustment: tukey method for comparing a family of 4 estimates
```

For the least squares means comparisons on degree level, it is suggest that if one has a doctorate degree, you generally make more money compared to just having a bachelor's or master's degree. There doesn't seem to be a significant difference between the amount of money a bachelor's degree and master's degree (one of the contrasts is not significant  $p = 0.45$ ), which is suggestive of the plot.

It is suggested that humanities and social sciences disciplines, in general earn less than those in engineering. There doesn't seem to be a significant difference in earnings between those in the social sciences and management subjects though ( $p = 0.9604$ ). All p-values are significant other than that particular contrast.

## Question 21.1

There is a big difference between the completely randomized design (CRD) and the randomized block design (RBD). For CRD, a random permutation is used to randomly assign treatments to the experimental units; however, this is only done in one go.

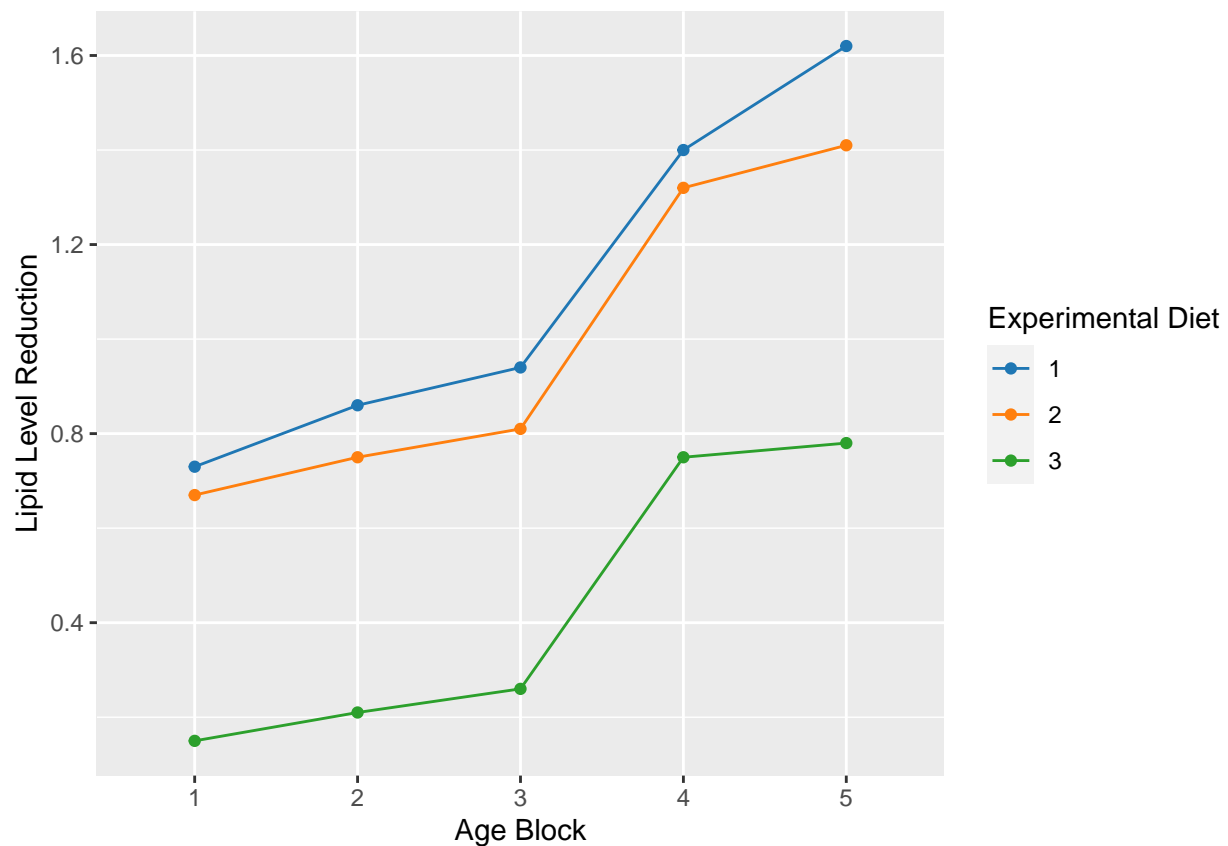
The RBD has  $n$  amount of blocks, and so there will be  $n$  independent permutations for independent assignment of treatments for the  $n$  blocks. In other words, each block is independently assigned treatments with their own randomization.

## Question 21.7

### Part A

I believe age was used as a blocking factor because people's metabolisms change over time, generally slowing down as one gets older. It may be the case that lipid levels are different across age groups, due to changes in metabolic functions. It also may be the case that the risk of coronary heart disease is different at different age groups as well. Because of the two hypothesis, it would not be fair to directly compare the treatment effects (experimental diet) when we have this age structure going on. The results can be misleading, confounders e.g., metabolic functions etc as mentioned.

### Part C



The curves are parallel, so I'd say that the no interaction assumption is appropriate.

## Part D

$$H_0 : D = 0$$

$$H_a : D \neq 0$$

$$F^* = \frac{SSBL.TR^*}{1} \div \frac{SSRem^*}{rn_b - r - n_b}$$

$$F_{critical} = F[0.99, 1, 7] \approx 12.25$$

$$F^* \approx 6.50$$

since  $F^* < F_{critical}$ , fail to reject null hypothesis

```
##  
## Tukey's one df F test for Additivity  
##  
## data: data3$block and data3$fat_content on data3$lipd_level  
## F = 6.4453, num df = 1, denom df = 7, p-value = 0.03873  
## sample estimates:  
## D estimate  
## 0.2723121
```

The p-value is 0.0387344. Since we reject the null hypothesis at  $\alpha = 0.01$ , there is sufficient evidence for us to conclude that there are no block-treatment interactions.

## Question 21.8

### Part A

```
## Warning in kable_styling(., position = "center"): Please specify format in
## kable. kableExtra can customize either HTML or LaTeX outputs. See https://
## haozhu233.github.io/kableExtra/ for details.
```

Table 9: ANOVA Table for Age-Fat Diet

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
block	4	1.41896	0.354740	146.8903	2e-07
fat_content	2	1.32028	0.660140	273.3499	0e+00
Residuals	8	0.01932	0.002415	NA	NA

### Part C

$$H_0 : \tau_j = 0$$

$$H_a : \text{not all } \tau_j = 0$$

$$F^* = \frac{MSTR}{MSBL.TR} = \frac{0.66014}{0.00242} \approx 272.79$$

$$F_{critical} = F[0.95, 2; 8] \approx 4.46$$

since  $F^* > F_{critical}$ , reject null hypothesis

Since  $F^* > F_{critical}$  there is sufficient evidence to reject the null hypothesis. There is reason to believe that the mean lipid level reduction for the three diets differ. The p-value is  $4.3659401 \times 10^{-8}$ .

### Part D

$$B = t(1 - \frac{0.05}{4}, 8) \approx 2.75$$

$$s^2\{D\} = \frac{2MSBL.TR}{5} \approx 9.68 * 10^{-4}$$

$$s \approx 0.031$$

$$\hat{D}_1 \approx 1.11 - 0.992$$

$$\hat{D}_2 \approx 0.992 - 0.43$$

Table 10: 95 Bonferroni

	D1	D2
Lower	0.0323927	0.4763927
Upper	0.2036073	0.6476073

### Question 21.19

$$\hat{E} = \frac{(n_b - 1)MSBL + n_b(r - 1)MSBL.TR}{(n_b r - 1)MSBL.TR}$$
$$\hat{E} = \frac{(5 - 1)0.35474 + 5(3 - 1)0.00242}{(5 * 3 - 1)0.00242} \approx 43$$

If we didn't use a blocking factor, we'd need a need to increase our sample size by a factor of about 40 to achieve similar results in a completely randomized design. In other words, blocking was very efficient in this study.

## Appendix

### Question 19.15

#### Part A

```
getdata <- function(...){
  e = new.env()
  name = data(..., envir = e)[1]
  e[[name]]
}

data <- getdata("HayFeverRelief")
# two-factor, 3x3 = 9 treatments
data$A <- as.factor(data$A)
data$B <- as.factor(data$B)

treatment_means <- data %>%
  group_by(A, B) %>%
  summarise(means = mean(y),
            sd = sd(y))

p1 <- treatment_means %>%
  ggplot(aes(x = A, y = means, color = B)) +
  geom_line(aes(group = B)) + geom_point() +
  geom_errorbar(aes(ymin = means-sd, ymax = means+sd), width = 0.05) +
  scale_color_d3(palette = "category20c")

p1
#
# p2 <- ggline(data, x = "A", y = "y", color = "B",
#              add = c("mean_se"))
```

#### Part C

```
# test for interaction (n = 4 (number of sample sizes per treatment))
mod.1 <- lm(y ~ A*B, data = data)
anova.table <- data.frame(anova(mod.1))
kable(anova.table[, ], caption = "Hay Fever ANOVA Table", format = "markdown") %>%
  kable_styling(position = "center")
p <- 1-pf(7.356/0.060,4,27) # or
p2 <- pf(7.356/0.060,4,27, lower.tail = FALSE)
```

#### Part D

```
# Factor A
F.star <- 110.010/0.060
F.crit <- qf(0.95,2,27)
p <- pf(1833.5,2,27,lower.tail = FALSE)

# Factor B
F.star <- 61.830/0.060
F.crit <- qf(0.95,2,27)
p <- pf(1030.5,2,27,lower.tail = F)
```



## Question 19.32

### Part C

```
results <- data %>%
  group_by(A, B) %>%
  summarise(means = mean(y))
temp2 <- data.frame(t(results))
temp2 <- temp2 %>%
  mutate_if(is.character, as.numeric)

L1 <- (temp2[3,2]+temp2[3,3])/2 - temp2[3,1]
L2 <- (temp2[3,5]+temp2[3,6])/2 - temp2[3,4]
L3 <- (temp2[3,8]+temp2[3,9])/2 - temp2[3,7]

MSE <- 0.060
a <- 3
b <- 3
n <- 4
est.variance <- (0.060/4) * ((-1)^2 +(0.5)^2 + (0.5)^2)
est.std <- sqrt(est.variance)
# sigma2.L <- sqrt(0.060/(b*n) * ((-1)^2 +(1/2)^2 +(1/2)^2))
S2 = (3*3-1)*qf(0.90,8,27)
S = sqrt(S2)

L1.LL <- L1 - S*est.std; L1.UL <- L1 + S*est.std;
L2.LL <- L2 - S*est.std; L2.UL <- L2 + S*est.std;
L3.LL <- L3 - S*est.std; L3.UL <- L3 + S*est.std;

corgi <- data.frame(
  L1 = c(L1.LL, L1.UL),
  L2 = c(L2.LL, L2.UL),
  L3 = c(L3.LL, L3.UL)
)
row.names(corgi) <- c("Lower Limit", "Upper Limit")
kable(corgi[, ], caption = "90% Scheffe Confidence for L1,L2,L3", format = "markdown") %>% kable_styling(

EST.VAR <- (0.060/4) * (2*(-1)^2+ 2*(0.5)^2+ 2*(0.5)^2)
EST.STD <- sqrt(EST.VAR)
L4 <- L2 - L1
L5 <- L3 - L1
L6 <- L3 - L2
L4.LL <- L4 - S*EST.STD; L4.UL <- L4 + S*EST.STD;
L5.LL <- L5 - S*EST.STD; L5.UL <- L5 + S*EST.STD;
L6.LL <- L6 - S*EST.STD; L6.UL <- L6 + S*EST.STD

poodle <- data.frame(
  L4 = c(L4.LL, L4.UL),
  L5 = c(L5.LL, L5.UL),
  L6 = c(L6.LL, L6.UL)
)
kable(poodle[, ], caption = "90% Scheffe for L4, L5, L6", format = "markdown") %>%
  kable_styling(position = "center")
```

```
# kable(anova.table[], caption = "Hay Fever ANOVA Table", format = "markdown") %>%  
#   kable_styling(position = "center")  
temp3 <- t(results)  
kable(temp3[], caption = "Treatment Means", format = "markdown") %>%  
  kable_styling(position = "center")
```

## Question 4 (data from 23.9)

### Part A

```
data2 <- read.table("https://people.stat.sc.edu/Hitchcock/adjunctprofessors.txt",
                    header = FALSE, col.names = c("earnings", "topic", "degree", "observation"))
data2$topic <- as.factor(data2$topic)
data2$degree <- as.factor(data2$degree)

# library(car)
earnings.mod <- lm(earnings ~ topic + degree + topic:degree, data = data2)
corgi <- data.frame(Anova(earnings.mod, type = "III"))
kable(corgi[], caption = "Earnings ANOVA Type 3 SS", format = "markdown") %>%
  kable_styling(position = "center")

# plotMeans(data2$earnings, data2$topic, data2$degree)

alt.means <- data2 %>%
  group_by(topic, degree) %>%
  summarise(means = mean(earnings),
            sds = sd(earnings))

p2 <- alt.means %>%
  ggplot(aes(x = topic, y = means, color = degree)) +
  geom_line(aes(group = degree)) + geom_point() +
  geom_errorbar(aes(ymin = means-sds, ymax = means+sds), width = 0.04) +
  scale_color_d3() +
  labs(x = "Subject",
       y = "Mean Earnings",
       color = "Degree Level")

p2
```

### Part B

```
blah1 <- anova(earnings.mod)
blah2 <- Anova(earnings.mod, type = "2")
blah3 <- Anova(earnings.mod, type = "3")
kable(blah1[], caption = "Type 1 SS", format = "markdown") %>%
  kable_styling(position = "center")
kable(blah2[], caption = "Type 2 SS", format = "markdown") %>%
  kable_styling(position = "center")
kable(blah3[], caption = "Type 3 SS", format = "markdown") %>%
  kable_styling(position = "center")
```

### Part C

### Part D

### Part E

```
library(lsmeans)
earnings.mod <- lm(earnings ~ topic + degree + topic:degree, data = data2)

a = lsmeans(earnings.mod,
```

```

    pairwise ~ degree,
    adjust = "tukey")

b = lsmeans(earnings.mod,
    pairwise ~ topic,
    adjust = "tukey")

kable(a[], caption = "lsmeans degree", format = "markdown") %>%
  kable_styling(position = "center")

kable(b[], caption = "lsmeans topic", format = "markdown") %>%
  kable_styling(position = "center")

```

### Question 21.1

## Question 21.7

### Part A

### Part B

```
p3 <- data3 %>%
  ggplot(aes(x = block, y = lipid_level, color = fat_content)) +
  geom_line(aes(group = fat_content)) + geom_point() +
  labs(x = "Age Block",
       y = "Lipid Level Reduction",
       color = "Experimental Diet")
  scale_color_d3()
p3
```

### Part C

```
# original author of the function: Jim Robison-Cox
# citation: http://fisher.utstat.utoronto.ca/~mahinda/stac52/tukey.txt
tukeys.add.test <- function(y,A,B){
  ## Y is the response vector
  ## A and B are factors used to predict the mean of y
  ## Note the ORDER of arguments: Y first, then A and B
  dname <- paste(deparse(substitute(A)), "and", deparse(substitute(B)),
                 "on",deparse(substitute(y)) )
  A <- factor(A); B <- factor(B)
  ybar.. <- mean(y)
  ybari. <- tapply(y,A,mean)
  ybar.j <- tapply(y,B,mean)
  len.means <- c(length(levels(A)), length(levels(B)))
  SSAB <- sum( rep(ybari. - ybar.., len.means[2]) *
              rep(ybar.j - ybar.., rep(len.means[1], len.means[2])) *
              tapply(y, interaction(A,B), mean))^2 /
            ( sum((ybari. - ybar..)^2) * sum((ybar.j - ybar..)^2))
  aovm <- anova(lm(y ~ A+B))
  SSrem <- aovm[3,2] - SSAB
  dfdenom <- aovm[3,1] - 1
  STATISTIC <- SSAB/SSrem*dfdenom
  names(STATISTIC) <- "F"
  PARAMETER <- c(1, dfdenom)
  names(PARAMETER) <- c("num df", "denom df")
  D <- sqrt(SSAB/ ( sum((ybari. - ybar..)^2) * sum((ybar.j - ybar..)^2)))
  names(D) <- "D estimate"
  RVAL <- list(statistic = STATISTIC, parameter = PARAMETER,
              p.value = 1 - pf(STATISTIC, 1,dfdenom), estimate = D,
              method = "Tukey's one df F test for Additivity",
              data.name = dname)
  attr(RVAL, "class") <- "htest"
  return(RVAL)
}

tukey.results <- tukeys.add.test(y = data3$lipid_level, A = data3$block, data3$fat_content)
tukey.results
```

## Question 21.8

Part A

Part C

Part D

```
age_means <- data3 %>%
  group_by(block, fat_content) %>%
  summarise(means = mean(lipd_level))

woof <- data3 %>%
  group_by(fat_content) %>%
  summarise(means = mean(lipd_level))

D1 = 1.11 - 0.992
D2 = 0.992 - 0.43
S2 <- (2*0.00242)/5
s <- sqrt(S2)
B <- qt(0.9875,8)
D1.LL <- D1 - B*s; D1.UL <- D1 + B*s
D2.LL <- D2 - B*s; D2.UL <- D2 + B*s

blah <- data.frame(
  D1 = c(D1.LL, D1.UL),
  D2 = c(D2.LL, D2.UL)
)
row.names(blah) <- c("Lower", "Upper")
kable(blah[, caption = "95 Bonferroni", format = "markdown"]) %>%
  kable_styling(position = "center")
```

### Question 21.19

```
num <- (5-1)*0.35474 + 5*(3-1)*0.00242  
denom <- (5*3-1)*0.00242  
E.prime <- num/denom
```