

Homework 2

Applied Multivariate Analysis

Emorie Beck

September 7, 2018

1 Workspace

1.1 Packages

```
library(car)
library(knitr)
library(kableExtra)
library(multcomp)
library(lme4)
library(plyr)
library(tidyverse)
```

1.2 data

The file, Set_2.csv, contains data for four variables and four groups.

```
wd <- "https://github.com/emoriebeck/homeworks/raw/master/multivariate/homeworks/homework2"

dat <- sprintf("%s/Set_2(1).csv", wd) %>%
  read_csv() %>%
  mutate(Group = factor(Group))

head(dat)

## # A tibble: 6 x 5
##   Group DV1 DV2 DV3 DV4
##   <fct> <int> <int> <int> <int>
## 1 1      1      2      5      3
## 2 1      2      2      1      1
## 3 1      3      2      5      3
## 4 1      2      2      2      2
## 5 1      3      1      1      1
## 6 1      4      4      1      4

long_dat <- dat %>%
  mutate(SID = 1:n()) %>%
  gather(key = DV, value = value, DV1:DV4)
```

2 Question 1

Conduct a standard ANOVA on each of the measures using `aov()`. Are the groups different on each of the measures? If so, conduct post-hoc comparisons using Holm correction, indicating the pairs of means that are significantly different.

```
Q1 <- long_dat
contrasts(Q1$Group) <- contr.treatment(4)
Q1 <- Q1 %>%
  group_by(DV) %>%
  nest() %>%
  mutate(mod = map(data, ~lm(value ~ Group, data = .)),
         aov = map(data, ~aov(value ~ Group, data = .)),
         tidy = map(aov, broom::tidy))

Q1 %>% unnest(tidy) %>% filter(term == "Group")

## # A tibble: 4 x 7
##   DV    term    df sumsq meansq statistic  p.value
##   <chr> <chr> <dbl> <dbl> <dbl>    <dbl>    <dbl>
## 1 DV1   Group     3  100.   33.4     19.4 6.40e-10
## 2 DV2   Group     3   48.6   16.2      7.14 2.22e- 4
## 3 DV3   Group     3   30.9   10.3      6.03 8.33e- 4
## 4 DV4   Group     3   21.8    7.28      3.79 1.28e- 2
```

There are group differences on all DV's.

```
compMat <- rbind(
  c( 0, 1, 0, 0), # 1 v 2
  c( 0, 0, 1, 0), # 1 v 3
  c( 0, 0, 0, 1), # 1 v 4
  c( 0, 1,-1, 0), # 2 v 3
  c( 0, 1, 0,-1), # 2 v 4
  c( 0, 0, 1,-1) # 3 v 4
)
rownames(compMat) <- c("1 v 2", "1 v 3", "1 v 4",
                      "2 v 3", "2 v 4", "3 v 4")

tab_fun <- function(x){
  x$conftint %>% data.frame %>%
    mutate(Groups = rownames(.)) %>%
    select(Groups, everything())
}

Q1 <- Q1 %>%
  mutate(comp = map(mod, ~glht(. , linfct=compMat, alternative="two.sided",rhs=0)),
         summ = map(comp, ~conftint(. , adjusted("holm"))),
         tab = map(summ, tab_fun))

Q1 %>% unnest(tab) %>%
  mutate(sign = ifelse(sign(lwr) != sign(upr), "ns", "sig")) %>%
  mutate_at(vars(Estimate:upr), funs(sprintf("%.2f", .))) %>%
  mutate(CI = sprintf("[%s, %s]", lwr, upr)) %>%
  mutate_at(vars(Estimate, CI), funs(ifelse(sign == "sig", sprintf("\\textbf{%s}", .), .))) %>%
```

Table 1: Question 1: Pairwise Comparisons

Groups	DV1		DV2		DV3		DV4	
	b	CI	b	CI	b	CI	b	CI
1 v 2	0.76	[-0.21, 1.73]	0.20	[-0.91, 1.31]	1.16	[0.19, 2.13]	1.20	[0.18, 2.22]
1 v 3	0.36	[-0.61, 1.33]	0.04	[-1.07, 1.15]	-0.28	[-1.25, 0.69]	1.08	[0.06, 2.10]
1 v 4	2.60	[1.63, 3.57]	1.68	[0.57, 2.79]	0.60	[-0.37, 1.57]	0.76	[-0.26, 1.78]
2 v 3	0.40	[-0.57, 1.37]	0.16	[-0.95, 1.27]	1.44	[0.47, 2.41]	0.12	[-0.90, 1.14]
2 v 4	-1.84	[-2.81, -0.87]	-1.48	[-2.59, -0.37]	0.56	[-0.41, 1.53]	0.44	[-0.58, 1.46]
3 v 4	-2.24	[-3.21, -1.27]	-1.64	[-2.75, -0.53]	-0.88	[-1.85, 0.09]	0.32	[-0.70, 1.34]

```
select(-lwr, -upr, -sign, b = Estimate) %>%
gather(key = est, value = value, b, CI) %>%
unite(tmp, DV, est, sep = ".") %>%
spread(tmp, value) %>%
kable(., "latex", booktabs = T, escape = F,
      col.names = c("Groups", rep(c("b", "CI"), times = 4)),
      caption = "Question 1: Pairwise Comparisons") %>%
add_header_above(c(" " = 1, "DV1" = 2, "DV2" = 2, "DV3" = 2, "DV4" = 2))
```

3 Question 2

Now combine all of the information for groups and variables into a no-intercept model using lmer. Test the following hypotheses about group differences by constructing an appropriate contrast using glht in the multcomp package.

```
Q2 <- long_dat %>%
mutate(DV = str_remove(DV, "DV"),
      G1 = mapvalues(Group, 1:4, c(1,0,0,0)),
      G2 = mapvalues(Group, 1:4, c(0,1,0,0)),
      G3 = mapvalues(Group, 1:4, c(0,0,1,0)),
      G4 = mapvalues(Group, 1:4, c(0,0,0,1)),
      DV1 = mapvalues(DV, 1:4, c(1,0,0,0)),
      DV2 = mapvalues(DV, 1:4, c(0,1,0,0)),
      DV3 = mapvalues(DV, 1:4, c(0,0,1,0)),
      DV4 = mapvalues(DV, 1:4, c(0,0,0,1))) %>%
mutate_at(vars(G1:G4, DV1:DV4), funs(as.numeric(as.character(.)))) %>%
# group_by(DV) %>%
nest() %>%
mutate(mod = map(data, ~lme4::lmer(value ~ -1 + G1:DV1 + G1:DV2 + G1:DV3 + G1:DV4 +
                                G2:DV1 + G2:DV2 + G2:DV3 + G2:DV4 +
                                G3:DV1 + G3:DV2 + G3:DV3 + G3:DV4 +
                                G4:DV1 + G4:DV2 + G4:DV3 + G4:DV4 +
                                (1|SID), data = .)),
      tidy = map(mod, broom::tidy))

Q2 %>% unnest(tidy) %>%
filter(group == "fixed") %>%
separate(term, c("DV", "Group"), sep = ":", remove = F) %>%
```

Table 2: Question 2: Model Estimated Means

Groups	DV1		DV2		DV3		DV4	
	b	SE	b	SE	b	SE	b	SE
G1	2.44	0.28	2.80	0.28	3.12	0.28	2.84	0.28
G2	3.20	0.28	3.00	0.28	4.28	0.28	4.04	0.28
G3	2.80	0.28	2.84	0.28	2.84	0.28	3.92	0.28
G4	5.04	0.28	4.48	0.28	3.72	0.28	3.60	0.28

```
mutate(DV = ifelse(DV == "G1", Group, DV),
       Group = ifelse(grepl("G1", term), "G1", Group)) %>%
mutate_at(vars(estimate:std.error), funs(sprintf("%.2f", .))) %>%
select(-statistic, -group, -term, b = estimate, se = std.error) %>%
gather(key = est, value = value, b, se) %>%
unite(tmp, DV, est, sep = ".") %>%
mutate(value = ifelse(value == "NA", NA, value),
       Group = str_replace(Group, "[_]", "\\_")) %>%
spread(tmp, value) %>%
kable(., "latex", booktabs = T, escape = F,
      col.names = c("Groups", rep(c("b", "SE"), times = 4)),
      caption = "Question 2: Model Estimated Means") %>%
add_header_above(c(" " = 1, "DV1" = 2, "DV2" = 2, "DV3" = 2, "DV4" = 2))
```

Test these hypotheses for each of the four measures: 12 contrasts in all, correct using the Holm method.

3.1 Part A

The mean for Group 1 is different from the mean for Group 3

```
C1 <- cbind(diag(1, 4, 4), diag(0, 4, 4), diag(-1, 4, 4), diag(0, 4, 4))
rownames(C1) <- paste(paste("DV", 1:4, sep = ""), ": 1 v 3", sep = "")
```

3.2 Part B

The mean of Groups 1, 2, and 3 is different from the mean for Group 4

```
C2 <- cbind(diag(1, 4, 4), diag(1, 4, 4), diag(1, 4, 4), diag(-3, 4, 4))
rownames(C2) <- paste(paste("DV", 1:4, sep = ""), ": 1+2+3 v 4", sep = "")
```

3.3 Part C

The mean of Groups 1 and 2 is different from the mean for Groups 3 and 4

```
C3 <- cbind(diag(1, 4, 4), diag(1, 4, 4), diag(-1, 4, 4), diag(-1, 4, 4))
rownames(C3) <- paste(paste("DV", 1:4, sep = ""), ": 1+2 v 3+4", sep = "")
```

Table 3: Question 2: Group Combinations

Groups	DV1		DV2		DV3		DV4	
	b	CI	b	CI	b	CI	b	CI
1 v 3	-0.36	[-1.45, 0.73]	-0.04	[-1.13, 1.05]	0.28	[-0.81, 1.37]	-1.08	[-2.17, 0.01]
1+2 v 3+4	-2.20	[-3.74, -0.66]	-1.52	[-3.06, 0.02]	0.84	[-0.70, 2.38]	-0.64	[-2.18, 0.90]
1+2+3 v 4	-6.68	[-9.35, -4.01]	-4.80	[-7.47, -2.13]	-0.92	[-3.59, 1.75]	-0.00	[-2.67, 2.67]

```
compMat <- rbind(C1, C2, C3)

tab_fun <- function(x){
  x$confint %>% data.frame %>%
    mutate(Groups = rownames(x$confint)) %>%
    select(Groups, everything())
}

Q2 <- Q2 %>%
  mutate(comp = map(mod, ~glht(., linfct=compMat, alternative="two.sided", rhs=0)),
         summ = map(comp, ~confint(., adjusted("holm"))),
         tab = map(summ, tab_fun))

Q2 %>% unnest(tab) %>%
  mutate(sign = ifelse(sign(lwr) != sign(upr), "ns", "sig")) %>%
  mutate_at(vars(Estimate:upr), funs(sprintf("%.2f", .))) %>%
  mutate(CI = sprintf("[%s, %s]", lwr, upr)) %>%
  # mutate_at(vars(Estimate, CI), funs(ifelse(sign == "sig", sprintf("\textbf{%s}", .), .))) %>%
  select(-lwr, -upr, -sign, b = Estimate) %>%
  separate(Groups, c("DV", "Groups"), sep = ": ") %>%
  gather(key = est, value = value, b, CI) %>%
  unite(tmp, DV, est, sep = ".") %>%
  spread(tmp, value) %>%
  kable(., "latex", booktabs = T, escape = F,
        col.names = c("Groups", rep(c("b", "CI"), times = 4)),
        caption = "Question 2: Group Combinations") %>%
  add_header_above(c(" " = 1, "DV1" = 2, "DV2" = 2, "DV3" = 2, "DV4" = 2))
```

4 Question 3

3. Now construct contrasts for the following hypotheses, ignoring groups (3 contrasts, correct using the Holm method).

4.1 Part A

The mean for DV1 is different from the mean for DV2

```
C1 <- c(1,-1,0,0, 1,-1,0,0, 1,-1,0,0, 1,-1,0,0)
```

4.2 Part B

The mean for DV1 and DV2 is different from the mean for DV3 and DV4

Table 4: Question 3: Mean Differences Across Variables

Groups	b	CI
DV1 v. DV2	0.36	[-1.04, 1.76]
DV1+DV2 v. DV3+DV4	-1.76	[-3.74, 0.22]
DV1+DV2+DV3 v DV4	-2.64	[-6.07, 0.79]

```
C2 <- c(1,1,-1,-1, 1,1,-1,-1, 1,1,-1,-1, 1,1,-1,-1)
```

4.3 Part C

The mean for DV1, DV2, and DV3 is different from the mean for DV4

```
C3 <- c(1,1,1,-3, 1,1,1,-3, 1,1,1,-3, 1,1,1,-3)
```

```
compMat <- rbind(C1, C2, C3)
rownames(compMat) <- c("DV1 v. DV2", "DV1+DV2 v. DV3+DV4", "DV1+DV2+DV3 v DV4")

Q2 <- Q2 %>%
  mutate(compQ3 = map(mod, ~glht(., linfct=compMat, alternative="two.sided", rhs=0)),
         summQ3 = map(compQ3, ~confint(., adjusted("holm"))),
         tabQ3 = map(summQ3, tab_fun))

Q2 %>% unnest(tabQ3) %>%
  mutate(sign = ifelse(sign(lwr) != sign(upr), "ns", "sig")) %>%
  mutate_at(vars(Estimate:upr), funs(sprintf("%.2f", .))) %>%
  mutate(CI = sprintf("[%s, %s]", lwr, upr)) %>%
  mutate_at(vars(Estimate, CI), funs(ifelse(sign == "sig", sprintf("\\textbf{%s}", .), .))) %>%
  select(-(lwr:sign)) %>%
  kable(., "latex", escape = F, booktabs = T,
        col.names = c("Groups", "b", "CI"),
        caption = "Question 3: Mean Differences Across Variables")
```

5 Question 4

4. Finally, test each of the hypotheses from Question 3, but combine them with each of the following group questions (a total of 9 contrasts, correct using the Holm method):

5.1 Part A

Just consider Group 1 alone

```
C1 <- rbind(c( 1,-1, 0, 0, 0,0,0,0, 0,0,0,0, 0,0,0,0),
           c(1,1,-1,-1, 0,0,0,0, 0,0,0,0, 0,0,0,0),
           c(1,1, 1, -3, 0,0,0,0, 0,0,0,0, 0,0,0,0)
)
rownames(C1) <- c("DV1vDV2:G1", "DV1+2vDV3+4:G1", "DV1+2+3vDV4:G1")
```

5.2 Part B

Compare Group 2 to Group 3

```
C2 <- rbind(
  # variable 1 v 2
  # (DV1:G2 - DV1:G3 - DV2:G2 + DV2:G3)
  c(0,0,0,0, 1,-1,0,0, -1,1,0,0, 0,0,0,0),
  # variable 1+2 v. 3+4
  # ((DV1:G2 + DV2:G2) = (DV3:G2 + DV4:G2)) = ((DV1:G3 + DV2:G3) = (DV3:G3 + DV4:G3))
  # (DV1:G2 + DV2:G2 - DV3:G2 - DV4:G2 - DV1:G3 - DV2:G3 + DV3:G3 + DV4:G3) = 0
  c(0,0,0,0, 1, 1,-1,-1, -1,-1,1,1, 0,0,0,0),
  # variable 1+2+3 v 4
  # ((DV1:G2 + DV2:G2 + DV3:G2) = DV4:G2) = ((DV1:G3 + DV2:G3 + DV3:G3) + DV4:G3)
  # (DV1:G2 + DV2:G2 + DV3:G2 - DV4:G2 - DV1:G3 - DV2:G3 - DV3:G3 + DV4:G3) = 0
  c(0,0,0,0, 1,1,1,-3, -1,-1,-1,3, 0,0,0,0)
)
rownames(C2) <- c("DV1vDV2:G2vG3", "DV1+2vDV3+4:G2vG3", "DV1+2+3vDV4:G2vG3")
```

5.3 Part C

Compare Group 1 to the combination of Groups 2, 3, and 4.

```
C3 <- rbind(
  # variable 1 v 2
  # (DV1:G1 = (DV1:G2 + DV1:G3 + DV1:G4)) = (DV2:G1 = (DV2:G2 + DV2:G3 + DV2:G4))
  # (DV1:G1 - DV1:G2 - DV1:G3 - DV1:G4 - DV2:G1 + DV2:G2 + DV2:G3 + DV2:G4) = 0
  c(3,-3,0,0, -1,1,0,0, -1,1,0,0, -1,1,0,0),

  # variable 1 + 2 v 3 + 4
  # ((DV1:G1 + DV2:G1 - DV3:G1 - DV4:G1)) - DV1:G2 - DV1:G3 - DV1:G4 - DV2:G2 - DV2:G3 - DV2:G4 + DV3:G3 + DV4:G4 = 0
  c(3,3,-3,-3, -1,-1,1,1, -1,-1,1,1, -1,-1,1,1),

  # variable 1 + 2 + 3 v 4
  # ((DV1:G1 + DV2:G1 + DV3:G1) = DV4:G1) = ((DV1:G2 + DV2:G2 + DV3:G2 + DV1:G3 + DV2:G3 + DV3:G3 + DV1:G4 + DV2:G4 + DV3:G4) = DV4:G4)
  # DV1:G1 + DV2:G1 + DV3:G1 - DV4:G1 - DV1:G2 - DV2:G2 - DV3:G2 - DV1:G3 - DV2:G3 - DV3:G3 - DV1:G4 - DV2:G4 - DV3:G4 + DV4:G4 = 0
  c(3,3,3,-9, -1,-1,-1,3, -1,-1,-1,3, -1,-1,-1,3)
)
rownames(C3) <- c("DV1vDV2:G1vG2+3+4", "DV1+2vDV3+4:G1vG2+3+4", "DV1+2+3vDV4:G1vG2+3+4")
```

```
(compMat <- rbind(C1,C2,C3))
```

```
##           [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
## DV1vDV2:G1      1  -1   0   0   0   0   0   0   0   0
## DV1+2vDV3+4:G1  1   1  -1  -1   0   0   0   0   0   0
## DV1+2+3vDV4:G1  1   1   1  -3   0   0   0   0   0   0
## DV1vDV2:G2vG3   0   0   0   0   1  -1   0   0  -1   1
## DV1+2vDV3+4:G2vG3  0   0   0   0   1   1  -1  -1  -1  -1
## DV1+2+3vDV4:G2vG3  0   0   0   0   1   1   1  -3  -1  -1
## DV1vDV2:G1vG2+3+4  3  -3   0   0  -1   1   0   0  -1   1
## DV1+2vDV3+4:G1vG2+3+4  3   3  -3  -3  -1  -1   1   1  -1  -1
## DV1+2+3vDV4:G1vG2+3+4  3   3   3  -9  -1  -1  -1   3  -1  -1
##           [,11] [,12] [,13] [,14] [,15] [,16]
```

Table 5: Question 4: Group x DV Combinations

Groups	DV1+2+3 v DV4		V1+2 v DV3+4		DV1 v DV2	
	b	CI	b	CI	b	CI
G1	-0.16	[-2.11, 1.79]	-0.72	[-1.85, 0.41]	-0.36	[-1.16, 0.44]
G1vG2+3+4	2.00	[-4.77, 8.77]	-1.12	[-5.03, 2.79]	-1.80	[-4.56, 0.96]
G2vG3	1.64	[-1.12, 4.40]	-1.00	[-2.60, 0.60]	0.24	[-0.89, 1.37]

```
## DV1vDV2:G1          0    0    0    0    0    0
## DV1+2vDV3+4:G1      0    0    0    0    0    0
## DV1+2+3vDV4:G1      0    0    0    0    0    0
## DV1vDV2:G2vG3       0    0    0    0    0    0
## DV1+2vDV3+4:G2vG3   1    1    0    0    0    0
## DV1+2+3vDV4:G2vG3  -1    3    0    0    0    0
## DV1vDV2:G1vG2+3+4   0    0   -1    1    0    0
## DV1+2vDV3+4:G1vG2+3+4 1    1   -1   -1    1    1
## DV1+2+3vDV4:G1vG2+3+4 -1    3   -1   -1   -1    3
```

```
Q2 <- Q2 %>%
  mutate(compQ4 = map(mod, ~glht(., linfct=compMat, alternative="two.sided", rhs=0)),
         summQ4 = map(compQ4, ~confint(., adjusted("holm"))),
         tabQ4 = map(summQ4, tab_fun))
```

```
Q2 %>% unnest(tabQ4) %>%
  mutate(sign = ifelse(sign(lwr) != sign(upr), "ns", "sig")) %>%
  mutate_at(vars(Estimate:upr), funs(sprintf("%.2f", .))) %>%
  mutate(CI = sprintf("[%s, %s]", lwr, upr)) %>%
  mutate_at(vars(Estimate, CI), funs(ifelse(sign == "sig", sprintf("\\textbf{%s}", .), .))) %>%
  select(-lwr, -upr, -sign, b = Estimate) %>%
  separate(Groups, c("DV", "Groups"), sep = ":") %>%
  gather(key = est, value = value, b, CI) %>%
  unite(tmp, DV, est, sep = ".") %>%
  spread(tmp, value) %>%
  kable(., "latex", booktabs = T, escape = F,
        col.names = c("Groups", rep(c("b", "CI"), times = 3)),
        caption = "Question 4: Group x DV Combinations") %>%
  add_header_above(c(" " = 1, "DV1+2+3 v DV4" = 2, "V1+2 v DV3+4" = 2, "DV1 v DV2" = 2))
```