Homework 2

Applied Mutlivariate 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/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 %>%
  group_by(DV) %>%
  nest() %>%
  mutate(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  3.00 100    33.4      19.4  0.000000000640
## 2 DV2   Group  3.00  48.6  16.2       7.14 0.000222
## 3 DV3   Group  3.00  30.9  10.3       6.03 0.000833
## 4 DV4   Group  3.00  21.8   7.28      3.79 0.0128
```

There are group differences on all DV's.

```
compMat <- rbind(
  c(-1, 1, 0, 0), # 1 v 2
  c(-1, 0, 1, 0), # 1 v 3
  c(-1, 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$confint %>% data.frame %>%
    mutate(Groups = rownames(.)) %>%
    select(Groups, everything())
}

Q1 <- Q1 %>%
  mutate(comp = map(aov, ~glht(., linfct=compMat, alternative="two.sided",rhs=0)),
         summ = map(comp, ~confint(., adjusted("holm"), calpha = univariate_calpha())),
         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}", .), .))) %>%
  select(-lwr, -upr, -sign, b = Estimate) %>%
  gather(key = est, value = value, b, CI) %>%
  unite(tmp, DV, est, sep = ".") %>%
```

Table 1: Question 1: Pairwise Comparisons

	DV1		DV2		DV3		DV4	
Groups	b	CI	b	CI	b	CI	b	CI
1 v 2	**-1.68**	**[-2.84, -0.52]**	**-2.60**	**[-3.94, -1.26]**	**-1.96**	**[-3.12, -0.80]**	**-1.64**	**[-2.87, -0.41]**
1 v 3	**-2.08**	**[-3.24, -0.92]**	**-2.76**	**[-4.10, -1.42]**	**-3.40**	**[-4.56, -2.24]**	**-1.76**	**[-2.99, -0.53]**
1 v 4	0.16	[-1.00, 1.32]	-1.12	[-2.46, 0.22]	**-2.52**	**[-3.68, -1.36]**	**-2.08**	**[-3.31, -0.85]**
2 v 3	-0.40	[-1.14, 0.34]	-0.16	[-1.01, 0.69]	**-1.44**	**[-2.17, -0.71]**	-0.12	[-0.90, 0.66]
2 v 4	**1.84**	**[1.10, 2.58]**	**1.48**	**[0.63, 2.33]**	-0.56	[-1.29, 0.17]	-0.44	[-1.22, 0.34]
3 v 4	**2.24**	**[1.50, 2.98]**	**1.64**	**[0.79, 2.49]**	**0.88**	**[0.15, 1.61]**	-0.32	[-1.10, 0.46]

```
  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) %>%
  mutate(DV = ifelse(DV == "G1", Group, DV),
         Group = ifelse(grepl("G1", term), "G1", Group)) %>%
  mutate_at(vars(estimate:std.error), funs(sprintf("%.2f", .))) %>%
```

Table 2: Question 2: Model Estimated Means

	DV1		DV2		DV3		DV4	
Groups	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

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

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

tab_fun <- function(x){
  x$confint %>% data.frame %>%
    mutate(Groups = rownames(x$confint)) %>%
```

Table 3: Question 2: Group Combinations

	DV1		DV2		DV3		DV4	
Groups	b	CI	b	CI	b	CI	b	CI
1 v 3	-0.36	[-1.12, 0.40]	-0.04	[-0.80, 0.72]	0.28	[-0.48, 1.04]	**-1.08**	**[-1.84, -0.32]**
1+2 v 3+4	**-1.10**	**[-1.64, -0.56]**	**-0.76**	**[-1.30, -0.22]**	0.42	[-0.12, 0.96]	-0.32	[-0.86, 0.22]
1+2+3 v 4	**-2.23**	**[-2.85, -1.60]**	**-1.60**	**[-2.22, -0.98]**	-0.31	[-0.93, 0.32]	-0.00	[-0.62, 0.62]

```
    select(Groups, everything())
}

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

```
C2 <- c(.5,.5,-.5,-.5, .5,.5,-.5,-.5, .5,.5,-.5,-.5, .5,.5,-.5,-.5)
```

Table 4: Question 3: Mean Differences Across Variables

Groups	b	CI
DV1 v. DV2	0.36	[-0.80, 1.52]
DV1+DV2 v. DV3+DV4	**-0.88**	**[-1.70, -0.06]**
DV1+DV2+DV3 v DV4	**-2.23**	**[-3.16, -1.31]**

4.3 Part C

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

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

```
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"), calpha = univariate_calpha())),
         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(.5,.5,-.5,-.5, 0,0,0,0, 0,0,0,0, 0,0,0,0),
      c(.3,.3, .3, -1, 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, 1/2, 1/2,-1/2,-1/2, -1/2,-1/2,1/2,1/2, 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, 1/3,1/3,1/3,-1, -1/3,-1/3,-1/3,1, 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(1,-1,0,0, -1/3,1/3,0,0, -1/3,1/3,0,0, -1/3,1/3,0,0),

  # variable 1 + 2 v 3 + 4
  # ((.5*DV1:G1 + .5*DV2:G1) = (.5*DV3:G1 + .5*DV4:G1)) = ((DV1:G2 + DV1:G3 + DV1:4 + DV2:G2 + DV2:G3 +
  # DV1:G1 + DV2:G1 - DV3:G1 - DV4:G1 - DV1:G2 -  DV1:G3 - DV1:4 - DV2:G2 - DV2:G3 - DV2:G4 + DV3:G2 + D
  c(.5,.5,-.5,-.5, -1/6,-1/6,1/6,1/6, -1/6,-1/6,1/6,1/6, -1/6,-1/6,-1/6,1/6),

  # 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:
  # (DV1:G1 + DV2:G1 + DV3:G1 - DV4:G1 - DV1:G2 - DV2:G2 - DV3:G2 - DV1:G3 - DV2:G3 - DV3:G3 - DV1:G4 -
  c(1/3,1/3,1/3,-1/3, -1/9,-1/9,-1/9,1/9, -1/9,-1/9,-1/9,1/9, -1/9,-1/9,-1/9,1/9)
)
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]
## DV1vDV2:G1             1.0000000 -1.0000000  0.0000000  0.0000000
## DV1+2vDV3+4:G1         0.5000000  0.5000000 -0.5000000 -0.5000000
## DV1+2+3vDV4:G1         0.3000000  0.3000000  0.3000000 -1.0000000
## DV1vDV2:G2vG3          0.0000000  0.0000000  0.0000000  0.0000000
## DV1+2vDV3+4:G2vG3      0.0000000  0.0000000  0.0000000  0.0000000
## DV1+2+3vDV4:G2vG3      0.0000000  0.0000000  0.0000000  0.0000000
## DV1vDV2:G1vG2+3+4      1.0000000 -1.0000000  0.0000000  0.0000000
## DV1+2vDV3+4:G1vG2+3+4  0.5000000  0.5000000 -0.5000000 -0.5000000
## DV1+2+3vDV4:G1vG2+3+4  0.3333333  0.3333333  0.3333333 -0.3333333
##                              [,5]       [,6]       [,7]       [,8]
## DV1vDV2:G1             0.0000000  0.0000000  0.0000000  0.0000000
## DV1+2vDV3+4:G1         0.0000000  0.0000000  0.0000000  0.0000000
```

```
## DV1+2+3vDV4:G1          0.0000000  0.0000000  0.0000000  0.0000000
## DV1vDV2:G2vG3           1.0000000 -1.0000000  0.0000000  0.0000000
## DV1+2vDV3+4:G2vG3       0.5000000  0.5000000 -0.5000000 -0.5000000
## DV1+2+3vDV4:G2vG3       0.3333333  0.3333333  0.3333333 -1.0000000
## DV1vDV2:G1vG2+3+4      -0.3333333  0.3333333  0.0000000  0.0000000
## DV1+2vDV3+4:G1vG2+3+4 -0.1666667 -0.1666667  0.1666667  0.1666667
## DV1+2+3vDV4:G1vG2+3+4 -0.1111111 -0.1111111 -0.1111111  0.1111111
##                             [,9]      [,10]      [,11]     [,12]
## DV1vDV2:G1              0.0000000  0.0000000  0.0000000 0.0000000
## DV1+2vDV3+4:G1          0.0000000  0.0000000  0.0000000 0.0000000
## DV1+2+3vDV4:G1          0.0000000  0.0000000  0.0000000 0.0000000
## DV1vDV2:G2vG3          -1.0000000  1.0000000  0.0000000 0.0000000
## DV1+2vDV3+4:G2vG3      -0.5000000 -0.5000000  0.5000000 0.5000000
## DV1+2+3vDV4:G2vG3      -0.3333333 -0.3333333 -0.3333333 1.0000000
## DV1vDV2:G1vG2+3+4      -0.3333333  0.3333333  0.0000000 0.0000000
## DV1+2vDV3+4:G1vG2+3+4 -0.1666667 -0.1666667  0.1666667 0.1666667
## DV1+2+3vDV4:G1vG2+3+4 -0.1111111 -0.1111111 -0.1111111 0.1111111
##                            [,13]      [,14]      [,15]     [,16]
## DV1vDV2:G1              0.0000000  0.0000000  0.0000000 0.0000000
## DV1+2vDV3+4:G1          0.0000000  0.0000000  0.0000000 0.0000000
## DV1+2+3vDV4:G1          0.0000000  0.0000000  0.0000000 0.0000000
## DV1vDV2:G2vG3           0.0000000  0.0000000  0.0000000 0.0000000
## DV1+2vDV3+4:G2vG3       0.0000000  0.0000000  0.0000000 0.0000000
## DV1+2+3vDV4:G2vG3       0.0000000  0.0000000  0.0000000 0.0000000
## DV1vDV2:G1vG2+3+4      -0.3333333  0.3333333  0.0000000 0.0000000
## DV1+2vDV3+4:G1vG2+3+4 -0.1666667 -0.1666667 -0.1666667 0.1666667
## DV1+2+3vDV4:G1vG2+3+4 -0.1111111 -0.1111111 -0.1111111 0.1111111
```

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

Table 5: Question 4: Group x DV Combinations

	DV1+2+3 v DV4		V1+2 v DV3+4		DV1 v DV2	
Groups	b	CI	b	CI	b	CI
G1	-0.33	[-0.79, 0.13]	-0.36	[-0.77, 0.05]	-0.36	[-0.94, 0.22]
G1vG2+3+4	**-0.45**	**[-0.87, -0.04]**	**-1.43**	**[-1.91, -0.94]**	-0.60	[-1.27, 0.07]
G2vG3	0.55	[-0.12, 1.22]	-0.50	[-1.08, 0.08]	0.24	[-0.58, 1.06]