Homework 1

Applied Mutlivariate Analysis

Emorie Beck

September 7, 2018

1 Workspace

1.1 Packages

```
library(car)
library(knitr)
library(kableExtra)
library(multcomp)

## Loading required package:  mvtnorm
## Loading required package:  survival
## Loading required package:  TH.data
## Loading required package:  MASS
##
## Attaching package:  'TH.data'
## The following object is masked from 'package:MASS':
##
##    geyser

library(lme4)

## Loading required package:  Matrix

library(plyr)
library(tidyverse)

## -- Attaching packages ---------------------------------------------------------------------------
tidyverse 1.2.1 --
##  ggplot2 2.2.1     purrr  0.2.4
##  tibble  1.4.2     dplyr  0.7.5
##  tidyr  0.8.0     stringr 1.3.0
##  readr  1.1.1     forcats 0.3.0
## -- Conflicts ------------------------------------------------------------------------------------
tidyverse_conflicts() --
## x dplyr::arrange()  masks plyr::arrange()
## x purrr::compact()  masks plyr::compact()
## x dplyr::count()    masks plyr::count()
## x tidyr::expand()    masks Matrix::expand()
## x dplyr::failwith()  masks plyr::failwith()
## x dplyr::filter()    masks stats::filter()
## x dplyr::id()        masks plyr::id()
## x dplyr::lag()      masks stats::lag()
```

```
## x dplyr::mutate()    masks plyr::mutate()
## x dplyr::recode()    masks car::recode()
## x dplyr::rename()    masks plyr::rename()
## x dplyr::select()    masks MASS::select()
## x purrr::some()      masks car::some()
## x dplyr::summarise() masks plyr::summarise()
## x dplyr::summarize() masks plyr::summarize()
```

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))

## Parsed with column specification:
## cols(
##  Group = col_integer(),
##  DV1 = col_integer(),
##  DV2 = col_integer(),
##  DV3 = col_integer(),
##  DV4 = col_integer()
## )

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)
```

```
## # A tibble: 8 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 DV1   Residuals 96.0  165     1.72     NA    NA
## 3 DV2   Group      3.00  48.6  16.2       7.14  0.000222
## 4 DV2   Residuals 96.0  218     2.27     NA    NA
## 5 DV3   Group      3.00  30.9  10.3       6.03  0.000833
## 6 DV3   Residuals 96.0  164     1.71     NA    NA
## 7 DV4   Group      3.00  21.8   7.28      3.79  0.0128
## 8 DV4   Residuals 96.0  184     1.92     NA    NA
```

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 = ".") %>%
  spread(tmp, value) %>%
  kable(., "latex", booktabs = T, escape = F,
        col.names = c("Groups", rep(c("b", "CI"), times = 4))) %>%
  add_header_above(c(" " = 1, "DV1" = 2,  "DV2" = 2,  "DV3" = 2,  "DV4" = 2))
```

	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]**

```
# Q1 <- Q1 %>%
#   mutate(ref.grid = map(aov, emmeans::ref_grid),
#          emmeans = map(ref.grid, emmeans(., "Group")))
```

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.

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

3.2 Part B

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

3.3 Part C

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

4 Question 2

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

4.2 Part B

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

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

5.2 Part B

Compare Group 2 to Group 3

5.3 Part C

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