

# Homework 12

## Applied Multivariate Analysis

Emorie Beck

November 27, 2018

## 1 Workspace

### 1.1 Packages

```
library(car)
library(knitr)
library(kableExtra)
library(psych)
library(MASS)
library(car)
library(multcomp)
library(Rmisc)
library(mlogit)
library(candisc)
library(broom)
library(plyr)
library(tidyverse)
```

### 1.2 data

The file, Set\_11.csv, contains the data from an experiment investigating the effects of goal-related instructions on performance, persistence, and anxiety. Participants attempted to solve challenging anagrams (words with their letters scrambled, e.g., ULERITO = OUTLIER) during a 30-minute period. Participants could attempt as many anagrams as they wished and could give up and move on to a new one at any time. Persistence was the number of anagrams attempted. Performance was the percentage of anagrams solved correctly, of those attempted. Anxiety was the total score on a questionnaire designed to measure performance-related anxiety, given after the instructions, immediately before starting the task (scores could range from 10 to 40). Participants were given a set of 10 practice anagrams to give them an idea of their difficulty. Performance on this practice set was 51%.

Prior to beginning the task, groups were given instructions intended to influence their motivation on the task. Group 1 was told to simply "do your best." Group 2 was told, most students are able to correctly solve about 60% of the anagrams you will attempt." Given the practice test experience, these instructions were expected to make the task seem challenging, but not impossible. Group 3 was told, most students are able to correctly solve 80% of the anagrams you will attempt." These instructions were expected to make the task seem very difficult.

The researchers posed the following hypotheses:

1. The increasingly challenging instructions should lead to increasing levels of anxiety.
2. Groups 2 and 3 should exert more persistence than Group 1 but should not differ from each other because there is an upper limit to the number of anagrams that can reasonably be attempted in 30 minutes.

- Performance should be greatest in Group 2. Group 1 should have lower performance than Group 2 because of lower motivation. Group 3 should have lower performance than Group 2 because of the incapacitating effects of high anxiety.

```
wd <- "https://github.com/emoriebeck/homeworks/raw/master/multivariate/homeworks/homework13"

dat <- sprintf("%s/Set_11.csv", wd) %>% read.csv(., stringsAsFactors = F) %>% tbl_df %>%
  mutate(Group = factor(Group))

head(dat)

## # A tibble: 6 x 5
##   Perform Persist Anxiety Group   ID
##   <int>    <int>    <int> <fct> <int>
## 1     43     19     15  1     1
## 2     40     21     15  1     2
## 3     43     27     11  1     3
## 4     36     19     13  1     4
## 5     25     15     19  1     5
## 6     51     34     10  1     6
```

## 2 Question 1

Analyze each outcome measure separately and comment on the support those analyses provide for the hypotheses. Use any contrasts or follow-up comparisons that you think are important for testing each hypothesis.

```
nested1 <- dat %>%
  gather(key = outcome, value = value, Perform:Anxiety) %>%
  group_by(outcome) %>%
  nest() %>%
  mutate(lm = map(data, ~lm(value ~ Group, data = .)),
         aov = map(data, ~aov(value ~ Group, data = .)),
         tidy.lm = map(lm, broom::tidy),
         tidy.aov = map(aov, broom::tidy))

nested1 %>%
  unnest(tidy.lm) %>%
  filter(grepl("Group", term)) %>%
  mutate(p.value = ifelse(p.value < .001, "<.001", ifelse(p.value < .01,
                                                         "<.01", ifelse(p.value < .05, "<.05", ">.05")))) %>%
  kable(., "latex", booktabs = T, escape = F, digits = 2) %>%
  kable_styling(full_width = F)
```

outcome	term	estimate	std.error	statistic	p.value
Perform	Group2	13.75	4.25	3.23	.01
Perform	Group3	3.35	4.17	0.80	.05
Persist	Group2	17.15	2.78	6.16	.001
Persist	Group3	15.74	2.73	5.77	.001
Anxiety	Group2	4.29	1.68	2.56	.05
Anxiety	Group3	8.50	1.64	5.18	.001

```

H1.1 <- c(0, 1, 0) # -1*G1 + 0*G2 + 1*G3 = 0
# -1*(I) + 0*(I + D1) + 1*(I + D2) = 0
# -I + I + D2 = 0
# D2 = 0
H2.1 <- c(1, 1, 1) # G1 = G2 + G3
# (I) = (I + D1 + I + D2)
# 0 = I + D1 + D2
H3.1 <- c(1, -1, 0) # G1 + G3 = G2
# I + I + D2 = I + D1
# I - D1 + D2 = 0
H.1 <- rbind(H1.1, H2.1, H3.1)
rownames(H.1) <- c("H1", "H2", "H3")

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

nested1 <- nested1 %>%
  mutate(comp = map(lm, ~glht(., linfct=H.1, alternative="two.sided", rhs=0)),
         summ = map(comp, ~conftint(., adjusted("holm"), calpha = univariate_calpha())),
         tab = map(summ, tab_fun))

nested1 %>% unnest(tab) %>%
  filter((outcome == "Anxiety" & Group == "H1") |
         (outcome == "Persist" & Group == "H2"))

## # A tibble: 2 x 5
##   outcome Group Estimate   lwr   upr
##   <chr>   <chr>   <dbl> <dbl> <dbl>
## 1 Persist H2      60.5  53.8  67.3
## 2 Anxiety H1       4.29  0.946  7.63

```

### 3 Question 2

Then conduct a multivariate analysis and indicate if this approach adds any additional insight or qualification to the univariate analyses.

```

dat$Group.sum <- dat$Group
contrasts(dat$Group.sum) <- contr.Sum(3)
m1 <- manova(cbind(Perform, Persist, Anxiety) ~ Group.sum, data = dat)

summary(m1)$SS

## $Group.sum
##       Perform  Persist  Anxiety
## Perform 2508.431 2464.959 353.3990
## Persist 2464.959 4518.314 1699.2928
## Anxiety  353.399 1699.293  921.8728
##
## $Residuals

```

```

##           Perform   Persist   Anxiety
## Perform 15958.289  8341.921 -4003.599
## Persist  8341.921  6837.872 -2815.093
## Anxiety -4003.599 -2815.093  2476.127

m1$coefficients

##           Perform   Persist   Anxiety
## (Intercept)  55.618632  38.605427 16.14376068
## Group.sum[S.1] -5.698632 -10.965427 -4.26376068
## Group.sum[S.2]  8.048034   6.186239  0.02290598

dat <- dat %>%
  mutate(H1 = as.numeric(as.character(mapvalues(Group, 1:3, c(-1,1,1)))),
         H2 = as.numeric(as.character(mapvalues(Group, 1:3, c(2,-1,-1)))),
         H3.1 = as.numeric(as.character(mapvalues(Group, 1:3, c(-1, 1, 0)))),
         H3.2 = as.numeric(as.character(mapvalues(Group, 1:3, c(0, 1, -1)))))

H1.2lm <- lm(cbind(Perform, Persist, Anxiety) ~ H1, data = dat)
H1.2 <- candisc(H1.2lm, term="H1", data = dat ,type="2")
H1.2$coeffs.raw

##           Can1
## Perform -0.01170095
## Persist  0.14282440
## Anxiety  0.18415089

H2.2lm <- lm(cbind(Perform, Persist, Anxiety) ~ H2, data = dat)
H2.2 <- candisc(H2.2lm, term="H2", data = dat ,type="2")
H2.2$coeffs.raw

##           Can1
## Perform -0.01170095
## Persist  0.14282440
## Anxiety  0.18415089

H3.2lm <- lm(cbind(Perform, Persist, Anxiety) ~ H3.1 + H3.2, data = dat)
H3.2 <- candisc(H3.2lm, term="H3.2", data = dat ,type="2")
H3.2$coeffs.raw

##           Can1
## Perform -0.04318753
## Persist  0.15979975
## Anxiety  0.18633789

```