

Homework 4: MANOVA

S&DS 5360 | Multivariate Statistics

Brandon Tran (brandon.tran@yale.edu)

February 27, 2026

Setup & Data Overview

We will study the following variables, some of which were manually encoded after data cleaning, and other of which were part of the original study:

- **Continuous Response:** Depression Score (`deprawsc`) and Flourishing Score (`flourish`)
- **Categorical Predictors:** Therapy Yes/No (`ther_any`) and Sexual Orientation (`queer`)
- **Continuous Predictor:** Average Weeknight Hours of Sleep (`sleep_wknight`)

Per feedback from the grader, we omit the code to preprocess the data.

```
tibble [30,323 x 5] (S3: tbl_df/tbl/data.frame)
$ deprawsc      : num [1:30323] 24 8 6 1 1 22 7 21 1 2 ...
$ flourish       : num [1:30323] 26 47 48 45 47 16 46 15 45 56 ...
$ ther_any       : Factor w/ 2 levels "0","1": 2 2 1 1 1 1 1 1 1 1 ...
$ queer          : Factor w/ 2 levels "0","1": 1 2 1 1 1 1 1 1 1 1 ...
$ sleep_wknight: num [1:30323] 3 6 7 9 6 6 6 6 6 7 ...
- attr(*, "na.action")= 'omit' Named int [1:51623] 2 3 5 7 8 9 10 11 12 15 ...
..- attr(*, "names")= chr [1:51623] "2" "3" "5" "7" ...

# A tibble: 6 x 5
  deprawsc flourish ther_any queer sleep_wknight
    <dbl>     <dbl> <fct>   <fct>      <dbl>
1      24       26 1     0           3
2      8        47 1     1           6
3      6        48 0     0           7
4      1        45 0     0           9
5      1        47 0     0           6
6     22       16 0     0           6
```

Part 1 | Visualizing Our Categorical Response Variables

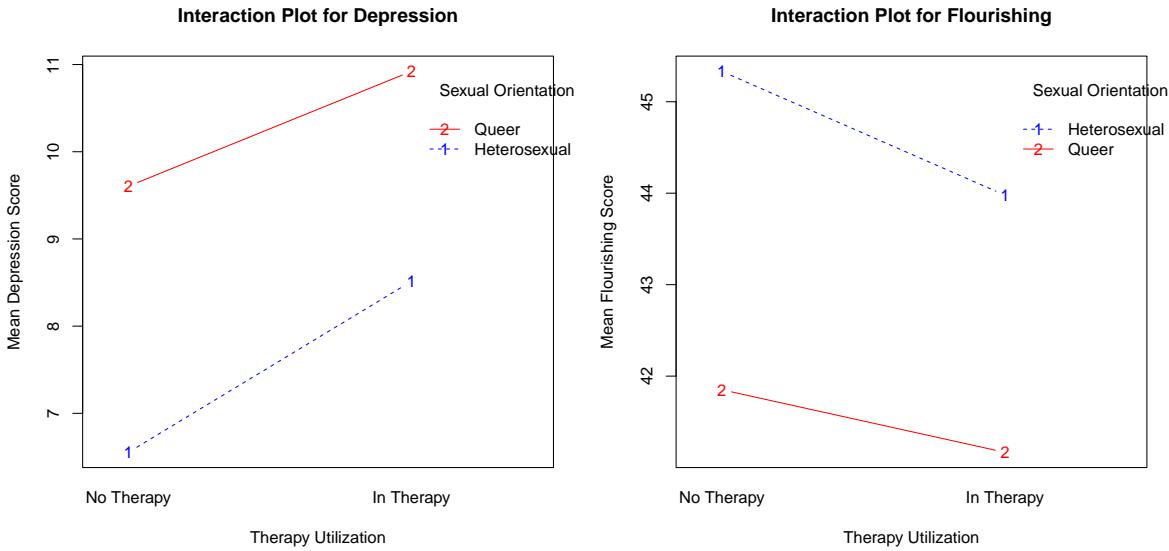
We present interaction plots for each of our response variables, `deprawsc` and `flourish`.

```
df_plot <- df %>%
  mutate(
    Therapy_Status = factor(ther_any, levels = c(0, 1), labels = c("No Therapy", "In Therapy"),
    Sexual_Orientation = factor(queer, levels = c(0, 1), labels = c("Heterosexual", "Queer"))
  )

par(mfrow = c(1, 2))

interaction.plot(
  x.factor = df_plot$Therapy_Status,
  trace.factor = df_plot$Sexual_Orientation,
  trace.label = 'Sexual Orientation',
  response = df$deprawsc,
  fun = mean,
  type = 'b',
  legend = TRUE,
  col = c('blue', 'red'),
  main = 'Interaction Plot for Depression',
  xlab = 'Therapy Utilization',
  ylab = 'Mean Depression Score'
)

interaction.plot(
  x.factor = df_plot$Therapy_Status,
  trace.factor = df_plot$Sexual_Orientation,
  trace.label = 'Sexual Orientation',
  response = df$flourish,
  fun = mean,
  type = 'b',
  legend = TRUE,
  col = c('blue', 'red'),
  main = 'Interaction Plot for Flourishing',
  xlab = 'Therapy Utilization',
  ylab = 'Mean Flourishing Score'
)
```



We observe that across both continuous response variables, we see roughly parallel lines. This indicates there is likely no interaction; that is, the effect of being in therapy or not seems to be equivalent for both queer and heterosexual students.

The positive slope of both lines for depression indicates that therapy utilization is associated with higher levels of depression. By comparison, the negative slope of both lines for flourishing indicates that therapy utilization is associated with lower levels of flourishing, as expected.

Finally, observing the gap between lines in both plots, we determine the main effects. Queer students seem to have lower flourishing scores than their heterosexual counterparts, regardless of therapy utilization. Conversely, heterosexual students yield higher flourishing scores, regardless of therapy utilization.

Part 2 | Two-Way MANOVA

With an understanding of how our variables interact in hand, we proceed to run two-way multivariate analysis of variance (MANOVA) to further examine the effect of therapy and sexual orientation on depression and flourishing. We perform a Type III (Marginal) MANOVA test because it is the most conservative approach for our Likert-scaled survey data with unequal group sizes. It does not assume there is no interaction (Type II), nor does it assume a perfectly balanced experiment (Type I). We test the following hypotheses:

Test 1: Does the effect of therapy depend on sexual orientation?

- H_0 : No multivariate interaction between therapy utilization and sexual orientation; effect on mental health response variables is the same for queer and heterosexual students.

- H_1 : There exists a multivariate interaction; the effect of therapy on mental health response variables differs based on sexual orientation.

Test 2: Do queer and heterosexual students have different mental health profiles?

- $H_0: \mu_{\text{heterosexual}} = \mu_{\text{queer}}$
- $H_1: \mu_{\text{heterosexual}} \neq \mu_{\text{queer}}$

Test 3: Do therapy users and non-users have different mental health profiles?

- $H_0: \mu_{\text{therapy}} = \mu_{\text{no therapy}}$
- $H_1: \mu_{\text{therapy}} \neq \mu_{\text{no therapy}}$

```
manova_fit <- lm(cbind(deprawsc, flourish) ~ ther_any * queer, data = df)
Manova(manova_fit, type = 'III')
```

```
Type III MANOVA Tests: Pillai test statistic
          Df test stat approx F num Df den Df Pr(>F)
(Intercept)  1  0.95441   317320      2 30318 < 2e-16 ***
ther_any     1  0.01623      250       2 30318 < 2e-16 ***
queer        1  0.03272      513       2 30318 < 2e-16 ***
ther_any:queer 1  0.00065      10       2 30318 5.1e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Note that all