# Exploring Correlation Between Program of Study and DERS Responses

Andy Chen

2025-11-15

## Introduction

Research Question: Is there a correlation between the Program of Study and the responses to the Difficulties in Emotional Regulation Scale (DERS) Questionnaire?

The DERS Questionnaire asks respondents to rank the frequency that they experience difficulties in emotional regulation. The scale is interpreted as follows:

1 = Almost never (0% - 10% of the time)

2 = Sometimes (11% - 35% of the time)

3 = About half the time (36% - 65% of the time)

4 = Most of the time (66% - 90% of the time)

5 = Almost always (91% - 100% of the time)

The questions are phrased such that a higher response indicates greater frequency of difficulties in emotional regulation.

### Install and Load Libraries

```
options(repos = c(CRAN = "https://cloud.r-project.org"))
install.packages(c("survey"))
```

```
## package 'survey' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##   C:\Users\toadcode\AppData\Local\Temp\RtmpC8nF2B\downloaded_packages
```

```
library(mosaic)
library(tidyverse)
library(survey)
library(knitr)
```

### Import Data

```
Data <- read.csv("student_mental_health.csv")[1:1193, ]
attach(Data)
```

**Filter Data**

```r
# Remove observations where participants failed the catch question
passed_catch = filter(Data, Data$Catch_question != "NA")

# Isolate the appropriate columns
this_subset <- passed_catch[, c("Program", "DERS_1", "DERS_2", "DERS_3", "DERS_4",
                                "DERS_5", "DERS_6", "DERS_7", "DERS_8",
                                "DERS_9", "DERS_10", "DERS_11", "DERS_12",
                                "DERS_13", "DERS_14", "DERS_15","DERS_16")]

# Add new column to data
this_subset['DERS_mean'] <- 0

for (i in 1:length(this_subset$Program)){
  # Populate the new columns with the mean of DERS responses from each row
  this_subset$DERS_mean[i] = rowMeans(this_subset[i, c(2:17)])

}
```

**Boxplots of the Mean Response to DERS Questions Per Program**

```r
this_subset$Program <- as.factor(this_subset$Program)

ggplot(this_subset,
       aes(x=this_subset$Program,
           y=this_subset$DERS_mean,
           fill=this_subset$Program)) +

# Label the scale
scale_fill_discrete('Programs',
                    labels=c('Business', 'Education', 'Engineering',
                    'Fine Arts', 'Human & Social Development',
                    'Humanities', 'Sciences',
                    'Social Sciences')) +

geom_boxplot() +

labs(title = "Distribution of Average Response Per Program",
         x = "Programs of Study",
         y = "Response") +

theme_minimal()+

theme(plot.title = element_text(hjust = 0.5), # center the title
      axis.text.x = element_blank())
```
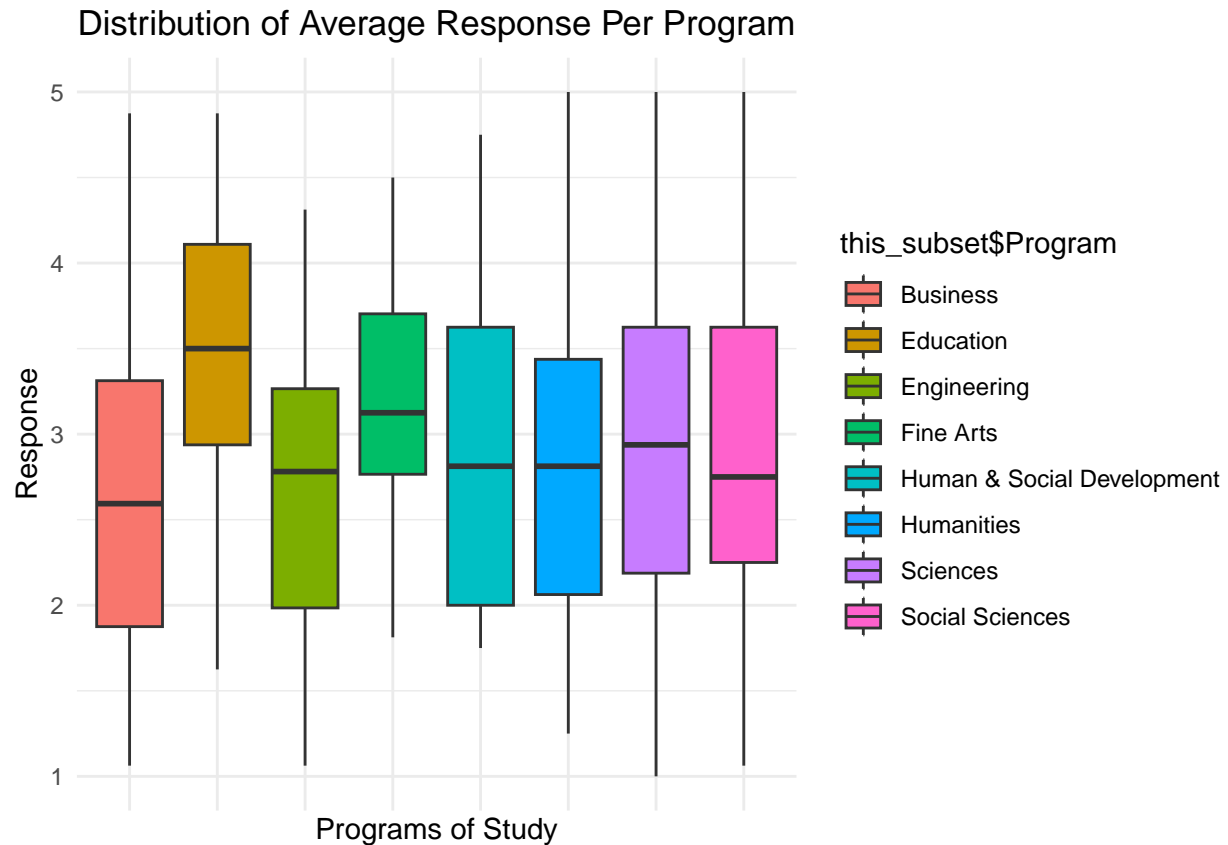
## Distribution of Average Response Per Program



The box plots produced suggests that students studying education, on average, have an average response higher than students not studying education.

We will proceed with Analysis of Variance (AoV) to compare the means and determine if there is significant differences.

Our null hypothesis is: "There is no significant differences in mean DERS responses between different programs of study" with alpha = 0.05.

**AoV and F Test**

```r
results <- aov(this_subset$DERS_mean ~ this_subset$Program, data = this_subset)

aov_summary <- summary(results)

aov_table <- aov_summary[[1]]

# F Test
f_value <- aov_table[["F value"]][1]

df_1 <- aov_table[["Df"]][1] # degrees of freedom
df_2 <- aov_table[["Df"]][2] # degrees of freedom

f_crit <- qf(0.95, df_1, df_2)

if (f_value > f_crit) {
  print("Reject the null hypothesis.")
```

```
} else {
  print("Fail to reject the null hypothesis.")

}
```

```
## [1] "Reject the null hypothesis."
```

The Anova and F test concluded that there is evidence to suggest there exists significant differences between the mean of the DERS responses between the Programs of study.

**Condifence Intervals**

```
Program <- 1:8
Mean <- 0
Lower_Bound <- 0
Upper_Bound <- 0

ci_summary = data.frame(Program, Mean, Lower_Bound, Upper_Bound)


for (i in 1:8){
  this_data <- subset(this_subset$DERS_mean, this_subset$Program == i)

  t_result <- t.test(this_data)

  this_ci <- t_result$conf.int

  ci_summary$Mean[i] = t_result$estimate
  ci_summary$Lower_Bound[i] = this_ci[1]
  ci_summary$Upper_Bound[i] = this_ci[2]
}


means <- ci_summary$Mean
lower <- ci_summary$Lower_Bound
upper <- ci_summary$Upper_Bound

labels <- c('Business', 'Education', 'Engineering',
                 'Fine Arts', 'Human & Social Development',
                 'Humanities', 'Sciences',
                 'Social Sciences')

ggplot(ci_summary, aes(labels, means)) +
# Label the scale
scale_fill_discrete('Programs',
                 labels=c('Business', 'Education', 'Engineering',
                 'Fine Arts', 'Human & Social Development',
                 'Humanities', 'Sciences',
                 'Social Sciences')) +

geom_point() +
geom_errorbar(aes(ymin = lower, ymax = upper)) +
labs(title = "Confidence Interval of Mean of DERS Responses",
      x = "Programs",
```

```
        y = "Average Response to DERS") +

theme_minimal()+
theme(plot.title = element_text(hjust = 0.5)) + # center the title
coord_flip() +
scale_y_continuous(limits = c(2, 4), breaks = seq(2, 4, by = 0.25))
```

## Confidence Interval of Mean of DERS Responses