individ_intro_background_EDA

Trevor Kwan

20/04/2021

1. Introduction

Background The study investigates changes in subjects' mental distress over time in an intervention group and a control group.

Methods Subjects answered questionnaires at baseline (month = 0), 3, 6, 18, and 60 months. These questionnaires measured a number of variables, including mental distress (GSI), gender, and education.

Measurement (Variable Descriptions) The independent variable was group status. All subjects were randomly assigned into one of two groups: a treatment/intervention group and a control group. Other predictor variables were measured in the questionnaires, including gender and education.

The dependent variable was the Global Severity Index (GSI), which is used as a measure of mental distress in subjects. A higher GSI score implies a higher level of mental distress.

Objectives The main objectives of the study are to compare the treatment effect with the control effect on reducing mental distress, to study how mental distress changes over time in all groups, and to investigate whether changes in mental distress can be explained by the treatment variable or other predictor variables.

Statistical Questions of Interest There was interest in investigating whether or not subjects' mental distress in both of the treatment groups decrease significantly over time. There was also interest in testing the effectiveness of the treatment. If there is a significant difference in subjects' mental distress between the treatment and the control group, this suggests the treatment to be effective.

Why is the problem important? If a treatment is found to be effective in reducing mental distress in subjects, this intervention can be recommended to people with overwhelming levels of mental distress.

delete later Objectives: - compare treatment with control on reducing GSI - see how subjects' GSI changes over time - see if changes in GSI over time can be partially explained by treatment variable, gender variable, or education variable

Statistical Questions: - Do subjects' GSI decrease significantly in treat group? In control group? (month is predictor, mixed effects regression) - Is there a sig difference in GSI between treat and control? (pre-post phase example)

"classes" == "gender" or "education"

2. Exploratory Data Analysis (EDA)

A large number of observations seem to be missing (see Table 1). The rate of missing observations per month was found to be greater for later months in comparison to earlier months (see Table 2). Exploratory data analysis was done using only data without missing observations. The data were examined for trends corresponding to the research questions: trajectory of mental distress over time and differences in mental distress between treatment and control groups.

```
## Rows: 1,355
## Columns: 6
## $ SN      <int> 161, 161, 161, 161, 161, 49, 49, 49, 49, 49, 383, 383, 383, 38~
## $ NEW_GRP <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
## $ month   <int> 0, 3, 6, 18, 60, 0, 3, 6, 18, 60, 0, 3, 6, 18, 60, 0, 3, 6, 18~
## $ DDS1    <int> 1, 1, 1, 1, 1, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ DDS4    <int> 11, 11, 11, 11, 11, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
## $ GSI     <dbl> NA, NA, NA, NA, NA, NA, NA, NA, 0.5094340, NA, NA, 2.0943396, ~
```

Table 1: Descriptive Summary Statistics

Statistic	Value
Number of Observations (Rows)	1355.0
Number of Observations with Missing Data (Rows)	317.0
Number of Complete Observations	1038.0
Number of Unique Subjects	271.0
Number of Unique Education Levels	14.0
Mean GSI	0.8
Standard Deviation GSI	0.4

Table 2: Rate of Missing Observations Per Month Interval

Month	Missing Rate
0	0.03
3	0.14
6	0.17
18	0.34
60	0.32

Large variations in GSI per month were found between individuals, suggesting the use of mixed-effects models in formal analysis (see Figure 1). Overall trends seemed to indicate that GSI scores were decreasing over time in both treatment and control groups (see Figure 1 and Figure 2). Grouping individuals by treatment status showed the median GSI of the treatment group to be slightly higher than the median GSI of the control group (see Figure 3). An independent sample t-test was conducted to explore the possibility of a significant difference in GSI scores across treatment groups, but the difference was not found to be significant (see Figure 3). A Wilcoxon test was also conducted since t-tests are sensitive to outliers, and this difference was found to be signficiant (see Figure 3). The figures in this section reflect only the data for which there were no missing observations.

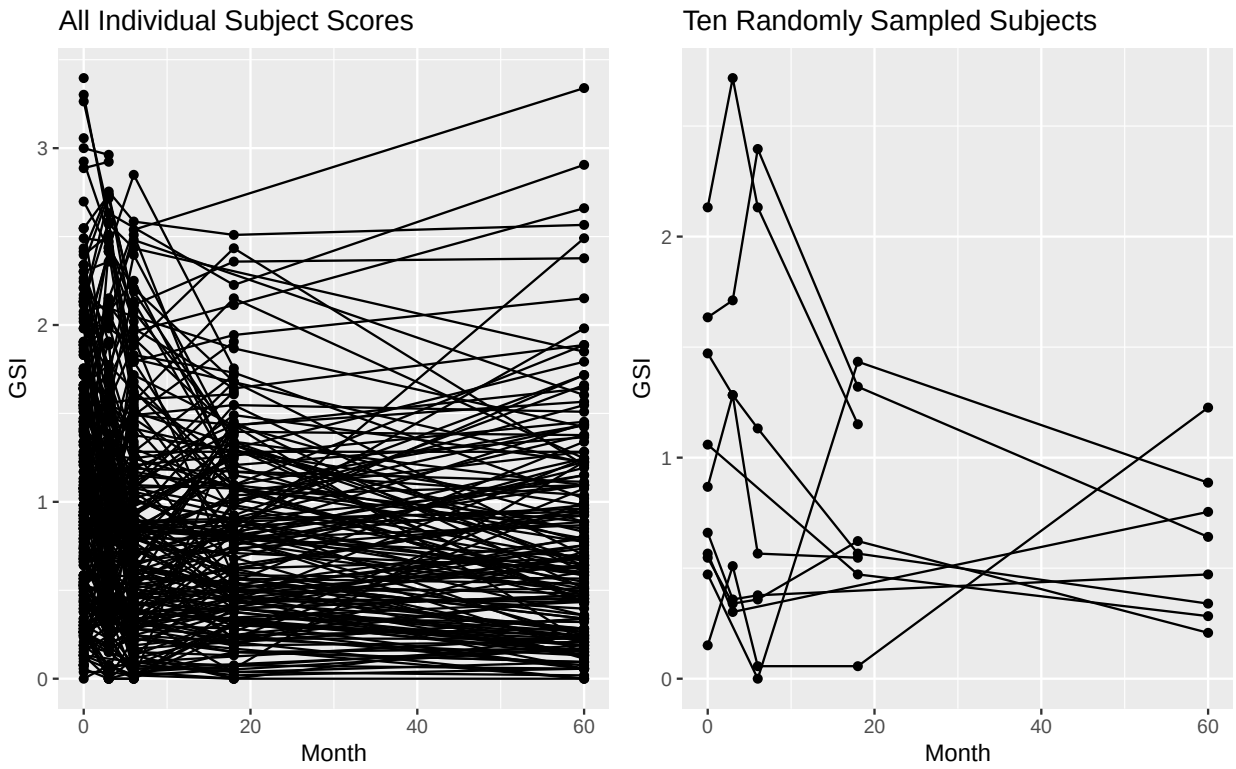

Figure 1: Individual GSI Scores Across Months

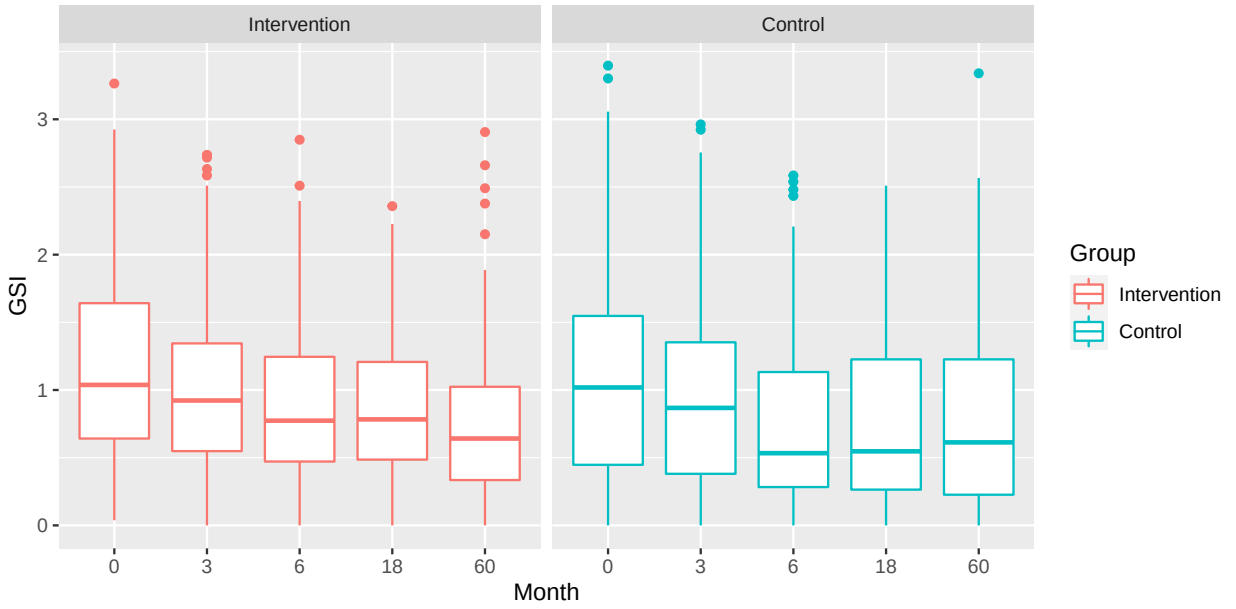

Figure 2: Total Writing Scores Across Intervention Weeks Grouped By Class

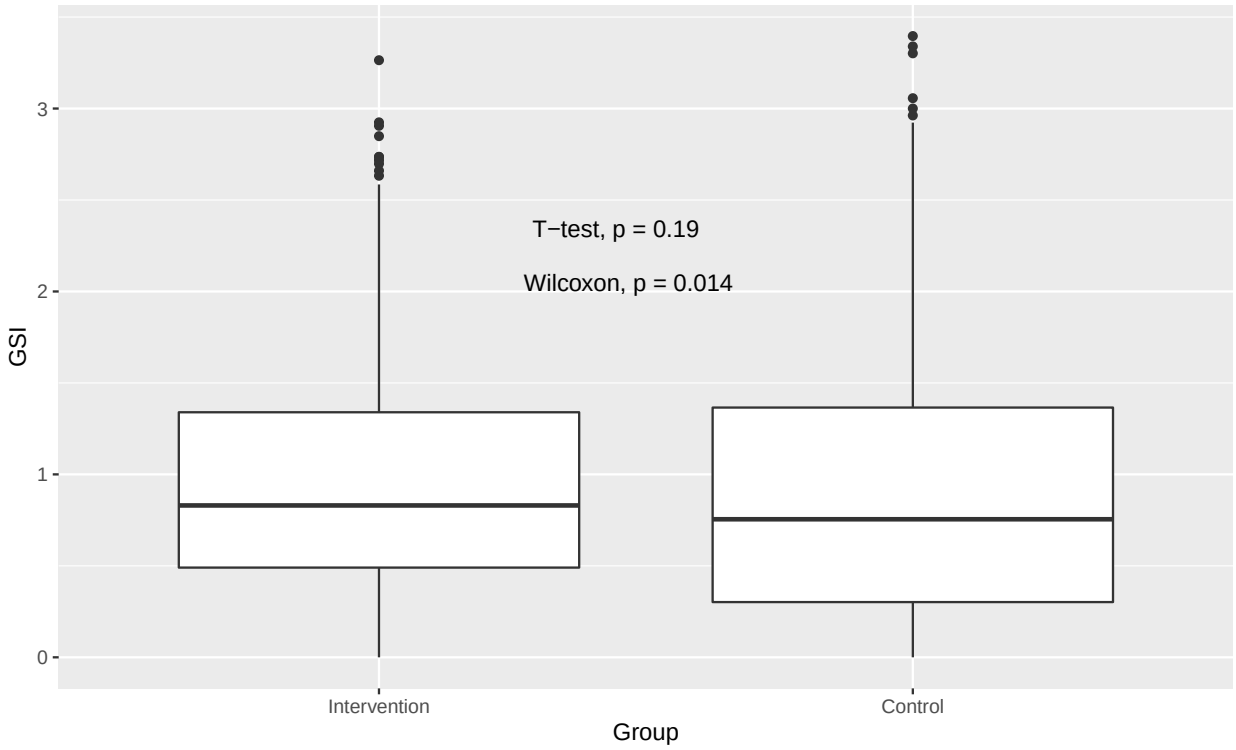

Figure 3: GSI Scores Across Treatment Groups