Assignment4
Saeidehsadat Mirjalili

Acupuncture for chronic headache in primary care: large, pragmatic, randomized trial

Reference: https://pubmed.ncbi.nlm.nih.gov/15023828/

Abstract

The assignment tries to uncover if acupuncture is a better option in comparison to the conventional treatments for lowering the level of tension-type or migraine headache. We use two box plots and demonstrate that a) age distribution across the groups is not different and b) there is a larger reduction of headache severity scores after one year in acupuncture group. According to the reference paper, there are 301 patients in the dataset.

Variates

Variates, or characteristics of each unit in the population, are defined as the followings:

- Age
- Sex (female/male)
- Migraine diagnosis (migraine (1)/tension-type (0))
- Chronicity (number of years of headache disorder)
- Acupuncturist (acupuncturist id code)
- Practice ID (gp practice id)
- Group (treatment group; acupuncture (1) vs. control (0))
- pk1 (headache severity score baseline)
- pk2 (headache severity score 3 month)
- pk5 (headache severity score 1 year)
- f1 (headache frequency baseline)
- f2 (headache frequency 3 month)
- f5 (headache frequency 1 year)

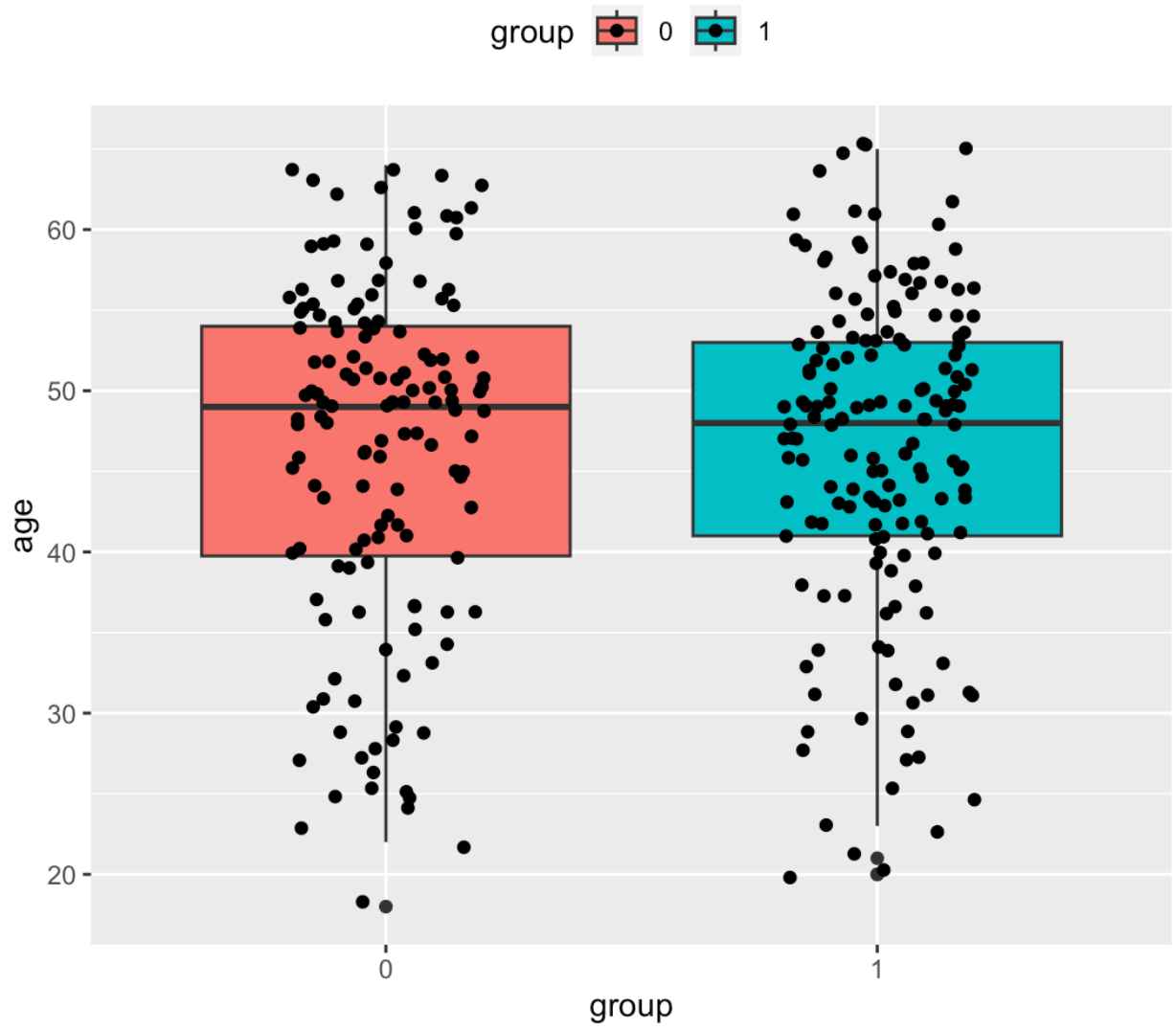

Figure 1. Distribution of age across the groups

- What software did you use to create your data visualization?
R

- Who is your intended audience?
Research community, patients with chronic headache.

- What information or message are you trying to convey with your visualization?
Whether distribution of age was different across the groups of patients. This is done to ensure age is not a confounding factor in the analysis.

- What design principles (substantive, perceptual, aesthetic) did you consider when making your visualization? How did you apply these principles? With what elements of your plots?

Substantive Principle: The substantive aspect is addressed by focusing on the relevant data that corresponds to the change in headache scores, which is the key measure of acupuncture effectiveness. Also, I ensured the same inclusion/exclusion criteria was applied and the number of observations matched the paper.

- How did you ensure that your data visualizations are reproducible? If the tool you used to make your data visualization is not reproducible, how will this impact your data visualization? I have provided the codes and the data for plotting the figures. There is no effective randomness in the procedure to be controlled with seeds. For example, jitter in the box plots does not have to be exactly reproduced.

- How did you ensure that your data visualization is accessible?
Accessibility in data visualization is ensured by providing text descriptions for the visual content and using color schemes that are colorblind friendly.

- Who are the individuals and communities who might be impacted by your visualization?
Patients with chronic headache, research community and TA of this course ☺ who needs to mark my assignment.

- How did you choose which features of your chosen dataset to include or exclude from your visualization?
I mainly followed the reference paper. Although there were multiple variables in the dataset, I focused on the headache scores at baseline and 1-year follow up. This was done to validate the main findings of the paper.

- What 'underwater labour' contributed to your final data visualization product?
The acupuncturists and GPs who provided the care, the patients who completed the study, the researchers who designed the research and those who curated the dataset.

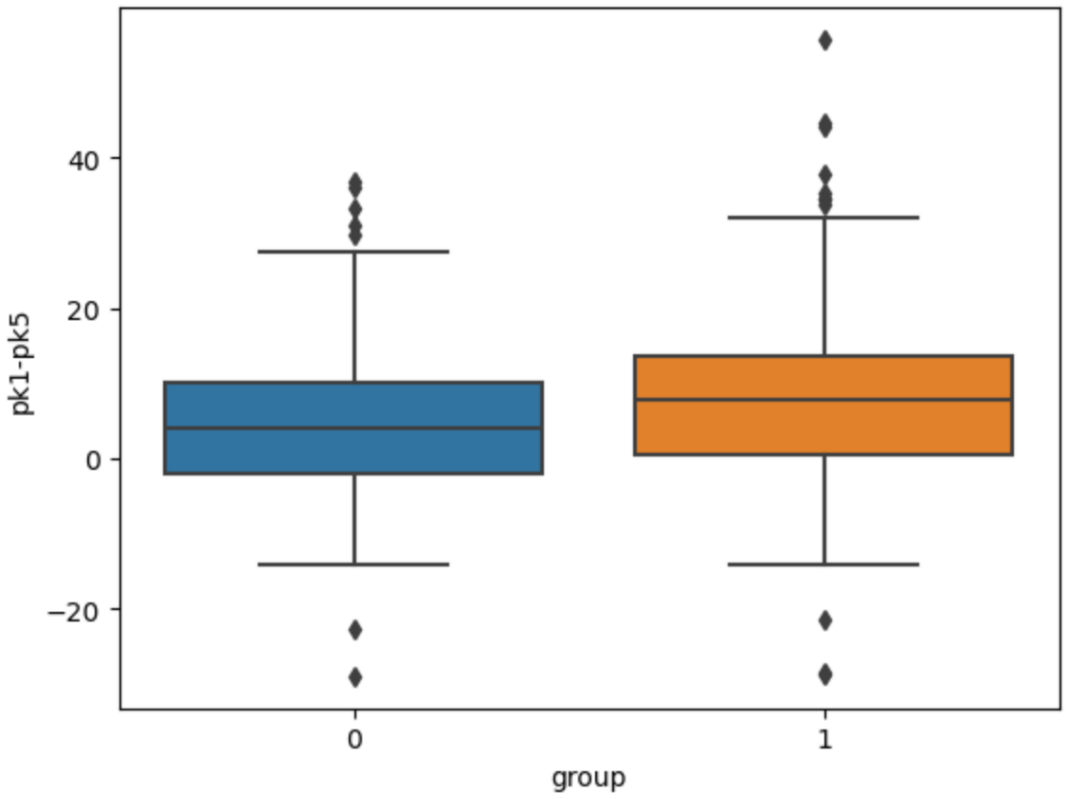

Figure 2. Headache score reduction after one year across the groups

- What software did you use to create your data visualization?
Python

- Who is your intended audience?
Same as Figure 1.

- What information or message are you trying to convey with your visualization?
Whether acupuncture combined with standard care is more effective for reducing headache scores after one year compared with standard case alone.

- What design principles (substantive, perceptual, aesthetic) did you consider when making your visualization? How did you apply these principles? With what elements of your plots?
Substantive Principle: The substantive aspect is addressed by focusing on the relevant data that corresponds to the change in headache scores, which is the key measure of acupuncture effectiveness. Also, I ensured the same inclusion/exclusion criteria was applied and the number of observations matched the paper.

Aesthetic Principle: The choice of colors (blue and orange) differentiates between the two groups without implying one is more significant or positive than the other. The colors are also likely to be distinguishable by individuals with color vision deficiencies, which is a consideration for accessibility.

- How did you ensure that your data visualizations are reproducible? If the tool you used to make your data visualization is not reproducible, how will this impact your data visualization?
Again, I have provided the codes and the data for plotting the figures, and there is no need for seeding.

- How did you ensure that your data visualization is accessible?
Same as Figure 1.

- Who are the individuals and communities who might be impacted by your visualization?
Same as Figure 1.

- How did you choose which features of your chosen dataset to include or exclude from your visualization?
I mainly followed the reference paper.

- What 'underwater labour' contributed to your final data visualization product?
Same as Figure 1.

Appendix A: Source code for generating Figure 1 in R

```
# Assignment4
# Saeidehsadat Mirjalili
setwd("/Users/saeideh/Desktop/DSI/DataVisualization/Assignment4/Code/")
acupuncture_dset <- readxl::read_excel("../Data/Data.xlsx")
pk5_filetered <- acupuncture_dset[!is.na(acupuncture_dset$pk5),] #ok_ 301
cat("Total number of patients in the test group", sum(pk5_filetered$group), '\n') # ok

# ########################Descriptive Plots############################
pk5_filetered$group <- factor(pk5_filetered$group)
age_plot <- ggplot2::ggplot(pk5_filetered, ggplot2::aes(x = group,
                                        y = age,
                                        fill = group)) +
  ggplot2::geom_boxplot() + ggplot2::geom_jitter(width = 0.2) +
  ggplot2::theme(legend.position="top")
age_plot
```

Appendix B: Source code for generating Figure 2 in Python

```
#Assignment4
#Saeidehsadat Mirjalili
import pandas as pd
# Loading the dataset
acupuncture_dset = pd.read_excel("../Data/Data.xlsx")
acupuncture_dset
acupuncture_dset.columns
acupuncture_dset.keys()
acupuncture_dset.info()
len(acupuncture_dset)
len(acupuncture_dset.columns)
acupuncture_dset.shape
pk5_filetered = acupuncture_dset[~pd.isnull(acupuncture_dset["pk5"])] #ok_ 301
print("nrows after filtering nans in pk5:", pk5_filetered.shape[0])
print("Total number of patients in the test group", sum(pk5_filetered["group"])) # ok_161
boxplot = pk5_filetered.boxplot(column=['age', 'chronicity', 'acupuncturist', 'practice_id'],
by='group', figsize=(10,20), layout=(4, 1))
import seaborn as sns
ax = sns.boxplot(x=pk5_filetered['group'], y=pk5_filetered['pk1']-pk5_filetered['pk5'])
ax.set(ylabel='pk1-pk5')
```