Lab Report no 1 for ST344

1632905

1. Background

The task we were given is to examine the relationship between self-perceived attractiveness and the variable which measures how much one person believes that their partner wants to meet with them again, to which we will refer as *date likelihood*. My research will be based on the results of the experiment which was conducted by researchers from Columbia, Harvard and Stanford Universities. The raw data can be found on this website (http://www.stat.columbia.edu/~gelman/arm/examples/speed.dating/).
The researchers Fisman, Iyengar, Kamenica and Simonson aimed to find how people choose their dating and/or marriage partners. After thorough consideration, they decided to pursue speed dating experiment to gather necessary data for their research. Before the events, potential participants were given an online survey to fill out which asked them to rate themselves on the following attributes: attractiveness, sincerity, intelligence, fun and ambition. They also provided various demographical information, such as their ethnicity, their field of study, the name of the university at which they studied or had been undergraduates, and their home location (ZIP code).
As the study focused only on heterosexual couples, during each event the female participant had a 4-minute conversation with a male participant, after which they filled out a post-meeting score sheet and proceeded to their next partners. Therefore, during an event, each participant took part in more than one date.

2. Introduction

We start our work by uploading the necessary data.

```
SpeedRawData <- read.csv("SpeedDatingRawData.csv")
SpeedData <- read.csv("SpeedData.csv")
```

One of our main variables will be self-perceived attractiveness as mentioned above. It will be represented by variable *attr3_1* which has a following description in the survey: *“Please rate your opinion of your own attributes, on a scale of 1-10 (be honest!): Attractive”*
We purposefully chose not to take into account variables attr3_2, atter3_3 and attr3_4. Although they also measured self-attractiveness, they were collected some time after the specific event and may be biased because of later dates or other events which can perturb our measurements.
As the variable which measures *date likelihood*, we will use *prob* which denotes the answer to the question: *How probable do you think it is that this person will say ‘yes’ for you? (1=not probable, 10=extremely probable)*

3. Data preparation

We must point out that we will use only each participant’s first speed date data, as the next ones occur in very short periods and could be not mutually independent. As part of data cleaning, we also want to erase rows without values in them. In order to perform those tasks, we prepared a function:

```
DataCleaning<-function(RawData, waves=c(1:21), attribute1="prob", attribute2="attr3_1")
{
part <- (SpeedRawData$partner == 1)
SpeedData <- RawData[part, c("iid", "gender", "prob", "attr3_1")]
SpeedData<-SpeedData[complete.cases(SpeedData),]
return(SpeedData)
}
```

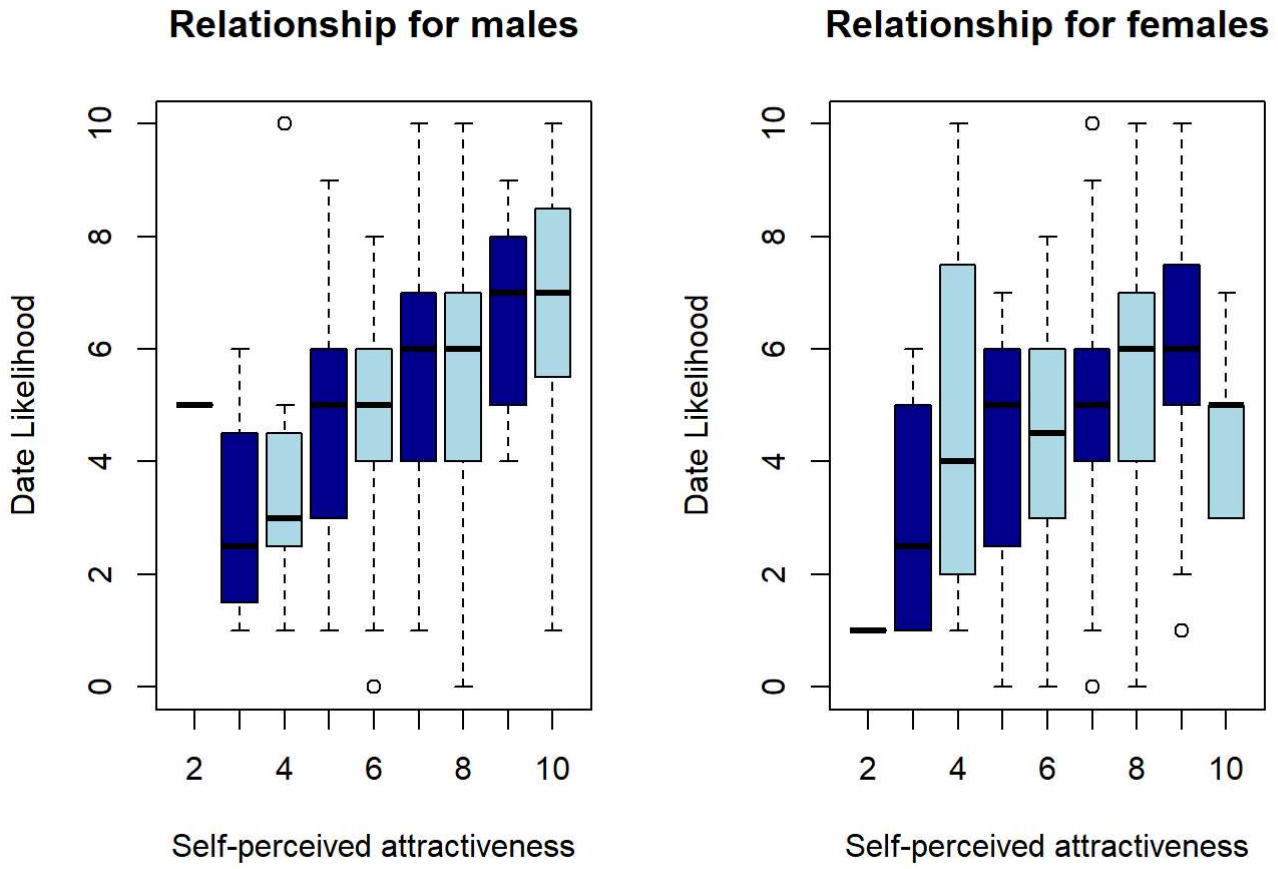
4. Analysis

For the purpose of this analysis, we will construct the function which will create boxplots for our data.
We will plot data for males and females separately to examine if there are differences between them. In our dataset males are denoted as *1* and females as *0*.

```
PlotData <- function(inputfile, waves = c(1:21),
                     attribute1 = "prob", attribute2 = "attr3_1")
{
  Data <- DataCleaning(inputfile)
  par(mfrow=c(1,2))
  boxplot(Data[,3] ~ Data[,4], data = Data, subset = (gender==1),
          main = "Relationship for males", col = (c("lightblue", "darkblue")),
          xlab = "Self-perceived attractiveness", ylab = "Date Likelihood")
  boxplot(Data[,3] ~ Data[,4], data = Data, subset = (gender==0),
          main = "Relationship for females", col = (c("lightblue", "darkblue")),
          xlab = "Self-perceived attractiveness", ylab = "Date Likelihood")
}
```

We use this function to create box plots for our data.

```
PlotData(SpeedRawData)
```



In the first boxplot, we see that there is a strong positive correlation between self-perceived attractiveness and date likelihood for males. The mean of date likelihood is higher for males with higher self-perceived attractiveness. However, this relationship is not that clear for females. The trend is mostly positive but the plot is much more flat in this case.

Furthermore, we see from the plots that some people with very high self-perceived attractiveness rate their dating chances very low as the plot shows that there are people who rated their attractiveness in the range 8-10 and still thinks that their Date Likelihood is in the range 0-2. Also, we can find out that the median of Date Likelihood for

very attractive females with self-perceived attractiveness of 8 or 9, is only 6.

From both plots, we see that almost nobody rate themselves as low as 1 or 2. In fact, there is no person which rated their attractiveness as 1 alnd only 2 people who assigned themselves the score of 2.

6. Conclusions

We found out that typically people with higher self-perceived attractiveness have stronger beliefs that the other person will date them again. These findings seem to be in line with common sense. The correlation is stronger among males than females. In the latter case, the relation is more flat, as even females with very high self-perceived attractiveness are not sure if the other person is interested in them.

7. References

Fisman, R., S. S. Iyengar, E. Kamenica, and I. Simonson. 2006. “Gender Differences in Mate Selection: Evidence from a Speed Dating Experiment.” The Quarterly Journal of Economics 121 (2). MIT Press: 673–97.