The effect of a conversational agent on individuals' motivation to perform a cognitive restructuring exercise

Mohammed Al Owayyed, Franziska Burger, Willem-Paul Brinkman

23/8/2021

Contents

1 Introduction

This document presents inferential statistical analyses of participants' perveived usefulness and self-efficacy. This analysis was reported in:

"The effect of a conversational agent on individuals' motivation to perform a cognitive restructuring exercise"

The OSF form belonging to this report can be found here: https://osf.io/v6tkq

Libraries used:

```
library(foreign) #open various data files
library(tidyr)  # for wide to long format transformation of the data
library(ggplot2) # plotting & data
library(pander) # for pander tables
library(ez) #for ezANOVA
library(psych) # reliability function
library(stringr) #find how many repeated aruguments
library(pastecs) # plotting & data
library(lsr) # effect size
library(tidyverse) # visualize data
library(nlme) # for multilevel
library(lme4) # Non-linear multilevel
library(ggpubr) # plotting
library(rstatix) # for calculating effect size
library(psych) # reliability function
library(stringr) #find how many repeated aruguments
library(fitdistrplus) # to fit distribution

library(effectsize)
```

2 Data file

To read the data from the file:

```
P_data <- read_excel('Dataset.xlsx', sheet = 'Sheet1')
```

Description of the data presented in P_data:

Field	Description
ParticipantID	The participant Unique ID
Group	Which condition the participant followed
Scenario	The scenario presented to rate the negative and positive thoughts from
Usefulness1	The post measure for the first usefulness question from 1 (strongly disagree) to 5 (strongly agree)
Usefulness2	The post measure for the second usefulness question from 1 (strongly disagree) to 5 (strongly agree)
Usefulness3	The post measure for the third usefulness question from 1 (strongly disagree) to 5 (strongly agree)
Usefulness4	The post measure for the fourth usefulness question from 1 (strongly disagree) to 5 (strongly agree)
Usefulness5	The post measure for the fifth usefulness question from 1 (strongly disagree) to 5 (strongly agree)
Usefulness6	The post measure for the sixth usefulness question from 1 (strongly disagree) to 5 (strongly agree)
Self-Efficacy1	The post measure for the first Self-Efficacy question from 0 (highly certain cannot do) to 10 (highly certain can do)
Self-Efficacy2	The post measure for the second Self-Efficacy question from 0 (highly certain cannot do) to 10 (highly certain can do)
Self-Efficacy3	The post measure for the third Self-Efficacy question from 0 (highly certain cannot do) to 10 (highly certain can do)

Field	Description
Self-Efficacy4	The post measure for the fourth Self-Efficacy question from 0 (highly certain cannot do) to 10 (highly certain can do)
Self-Efficacy5	The post measure for the fifth Self-Efficacy question from 0 (highly certain cannot do) to 10 (highly certain can do)
Pre-Usefulness1	The pre measure for the first usefulness question from 1 (strongly disagree) to 5 (strongly agree)
Pre-Usefulness2	The pre measure for the second usefulness question from 1 (strongly disagree) to 5 (strongly agree)
Pre-Usefulness3	The pre measure for the third usefulness question from 1 (strongly disagree) to 5 (strongly agree)
Pre-Usefulness-4	The pre measure for the fourth usefulness question from 1 (strongly disagree) to 5 (strongly agree)
Pre-Usefulness5	The pre measure for the fifth usefulness question from 1 (strongly disagree) to 5 (strongly agree)
Pre-Usefulness6	The pre measure for the sixth usefulness question from 1 (strongly disagree) to 5 (strongly agree)
Pre-Self-Efficacy1	The pre measure for the first Self-Efficacy question from 0 (highly certain cannot do) to 10 (highly certain can do)
Pre-Self-Efficacy-2	The pre measure for the second Self-Efficacy question from 0 (highly certain cannot do) to 10 (highly certain can do)
Pre-Self-Efficacy-3	The pre measure for the third Self-Efficacy question from 0 (highly certain cannot do) to 10 (highly certain can do)
Pre-Self-Efficacy-4	The pre measure for the fourth Self-Efficacy question from 0 (highly certain cannot do) to 10 (highly certain can do)
Pre-Self-Efficacy-5	The pre measure for the fifth Self-Efficacy question from 0 (highly certain cannot do) to 10 (highly certain can do)

2.1 Missing data

Overall, the study was completed 225 times, and the data of 33 participants were removed from the data analysis. The reasons for exclusion were (1) performing the experiment more than once (n = 11), for which only the first evaluation completed by the participants was included in the analysis; (2) no possibility of intervention effect, because “old” thoughts were not perceived as believable (rated as 0) (n = 5); (3) writing nonsensical answers to the exercise questions (n = 15); and (4) having the same answers to the open-ended questions as other participants had (n = 1 pair).

3 Perceived usefulness analysis

3.1 Reliability check

The participant in all three conditions were asked to fill a perceived usefulness questionnaire of before and after doing the exercise. The questionnaire includes 6 questions which they were asked to rate from 1 (Strongly Disagree) to 5 (Strongly agree). Reliability analysis of usefulness questions shows an acceptable reliability level (alpha > 0.7)

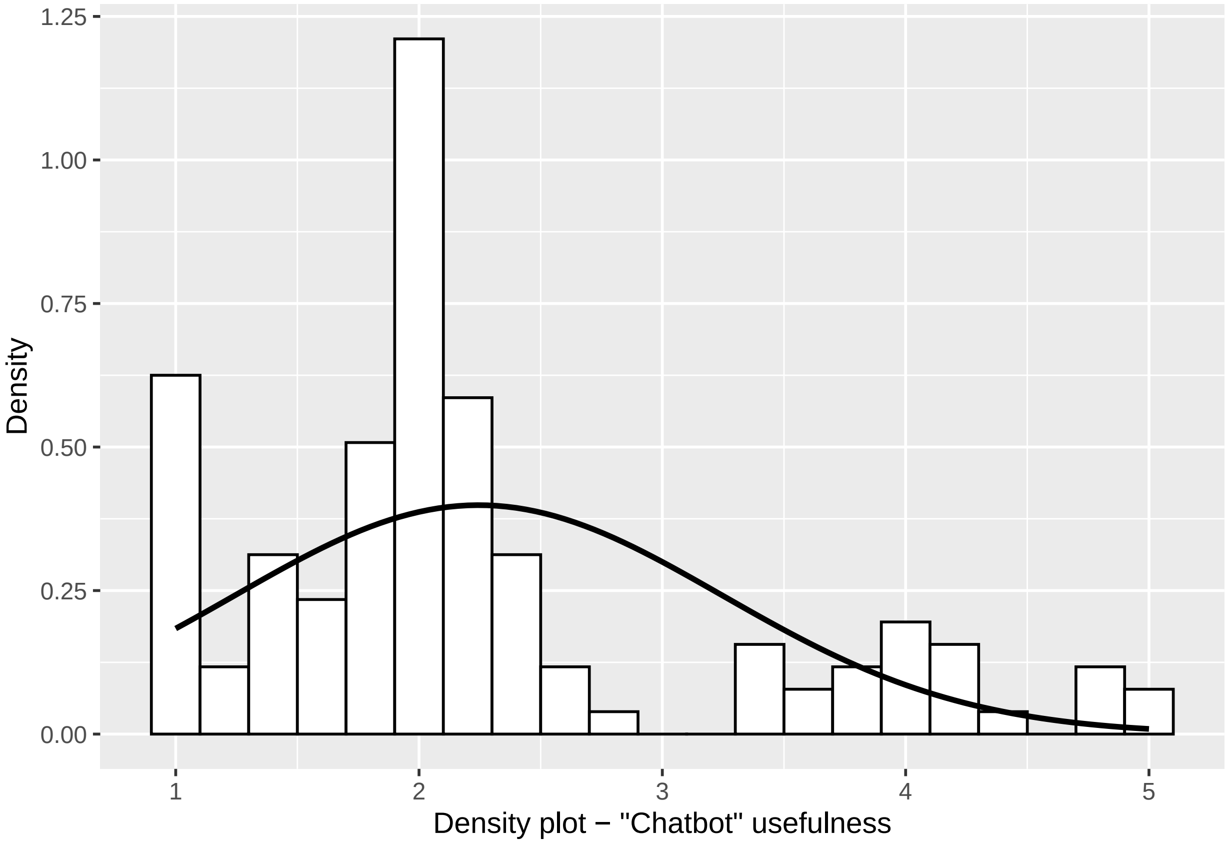
3.2 Data preparation

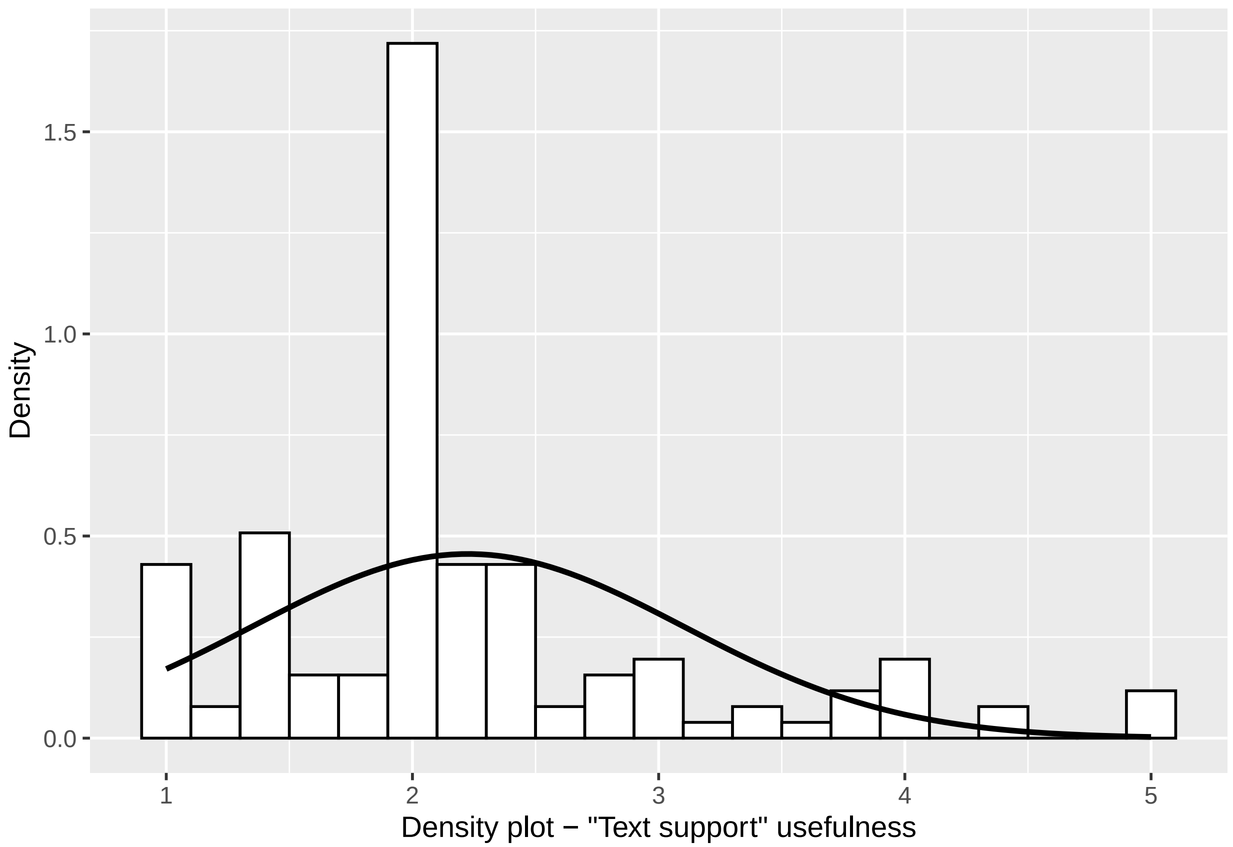
Since the reliability level was acceptable, we continued with getting a unified score for usefulness. First, we calculated the average of the pre and post questionnaire. Then, we transfer the data into another structure

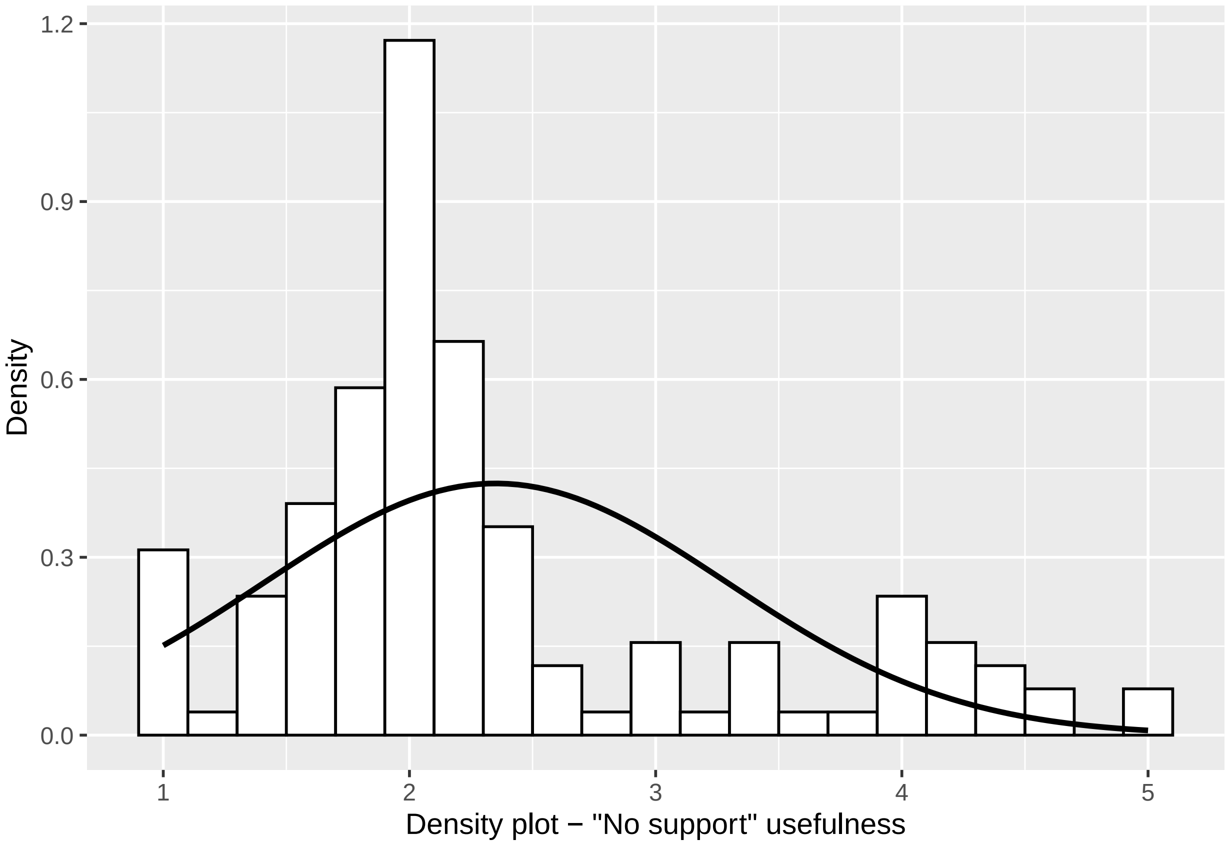
(Id, Group, Session, Score). The new structure can be used to fit a generalized model. After that, we subset the data to the three groups (i.e., chatbot, text support, no support).

3.3 Assumption checking

Before analysing the data, we checked for distribution normality. This was done visually for the 3 conditions:



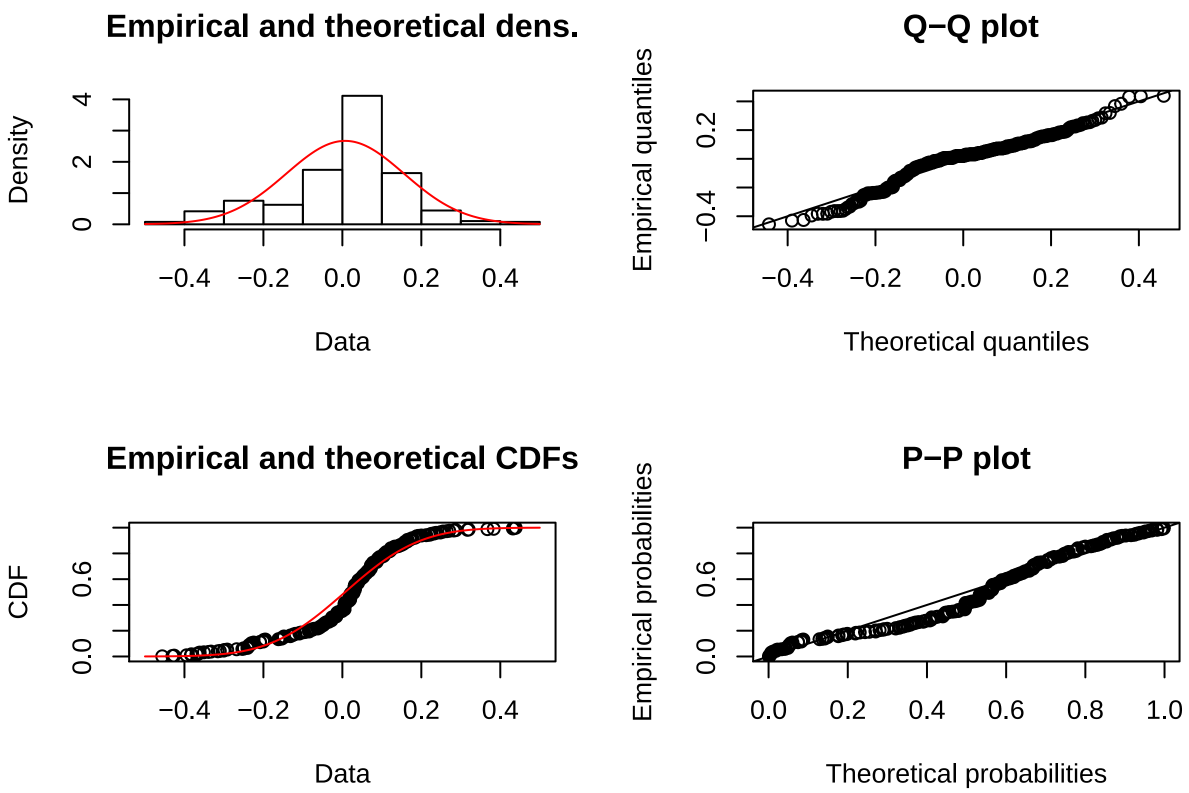

Density
1.5
1.0
0.5
0.0
1
2
3
4
5
Density plot − "Text support" usefulness



The data in the histograms shows a clear deviation from normal distribution.

3.4 Analysis of data

A generalized multilevel mixed effect model was fitted, wherein as a random effect, we used participant, and as fixed effects, we used the pre and post sessions. The model has a random intercept and a fixed slope, as we are assuming that all participants have the same direction but with various starting points. First, we checked if the residuals fits the distribution in case of using Gamma distribution.



The plots looks reasonable. We continued analysing the data using the same model.

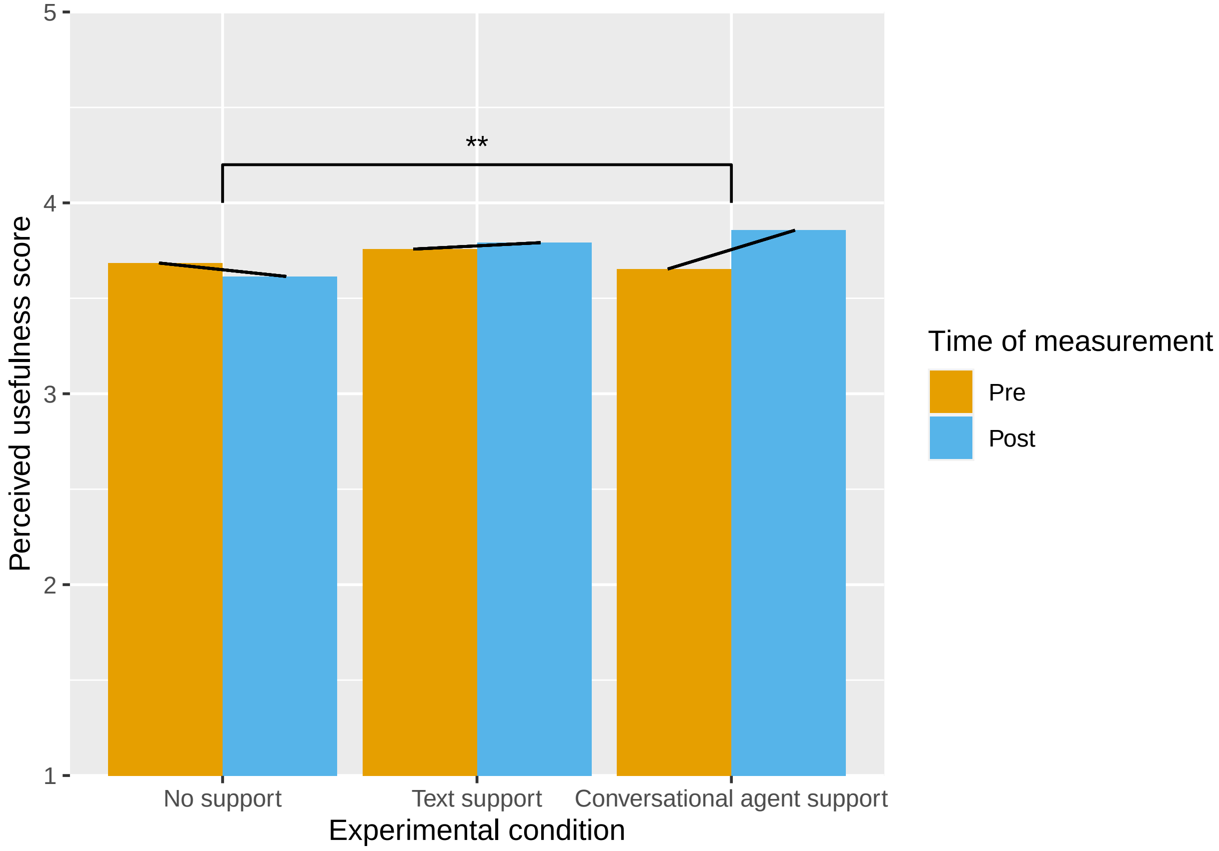
Table 2: Analysis of Deviance Table (Type II Wald chisquare tests)

	Chisq	Df	Pr(>Chisq)
Group	1.253	2	0.5344
Session	2.211	1	0.137
Group:Session	9.167	2	0.01022

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: Gamma  ( inverse )
## Formula: ScoreRverse ~ Group + Session + Group:Session + (1 | ParticipantID)
##    Data: Usf
##
##      AIC      BIC   logLik deviance df.resid
##    531.8    563.4   -257.9    515.8      376
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.7800 -0.2097  0.1299  0.4148  2.3083
##
## Random effects:
##  Groups        Name        Variance Std.Dev.
##  ParticipantID (Intercept) 0.01992  0.1411
##  Residual                  0.04793  0.2189
## Number of obs: 384, groups:  ParticipantID, 192
##
## Fixed effects:
##                               Estimate Std. Error t value Pr(>|z|)
## (Intercept)                    0.56460    0.02894  19.511  < 2e-16 ***
## GroupNo Support               -0.02230    0.04051  -0.550  0.58199
## GroupText Support             -0.01299    0.04009  -0.324  0.74594
## SessionPost                    0.03434    0.01088   3.156  0.00160 **
## GroupNo Support:SessionPost   -0.04547    0.01512  -3.007  0.00264 **
## GroupText Support:SessionPost -0.02828    0.01559  -1.814  0.06967 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) GrpNSp GrpTxS SssnPs GNS:SP
## GroupNSpprt -0.687
## GrpTxtSpprt -0.692  0.491
## SessionPost -0.173  0.123  0.124
## GrpNSppr:SP  0.125 -0.181 -0.089 -0.720
## GrpTSppr:SP  0.121 -0.086 -0.185 -0.698  0.502
```

the interaction between the groups and the sessions shows a significance $p < 0.05$. Therefore, there is a difference between the groups. Also, the usefulness shows a significant p-value between the the chatbot and no support (p<0.05)

The following bar chart show the difference between the pre and post questionnaire means for the 3 conditions.



4 Self-Efficacy analysis

4.1 Reliability check

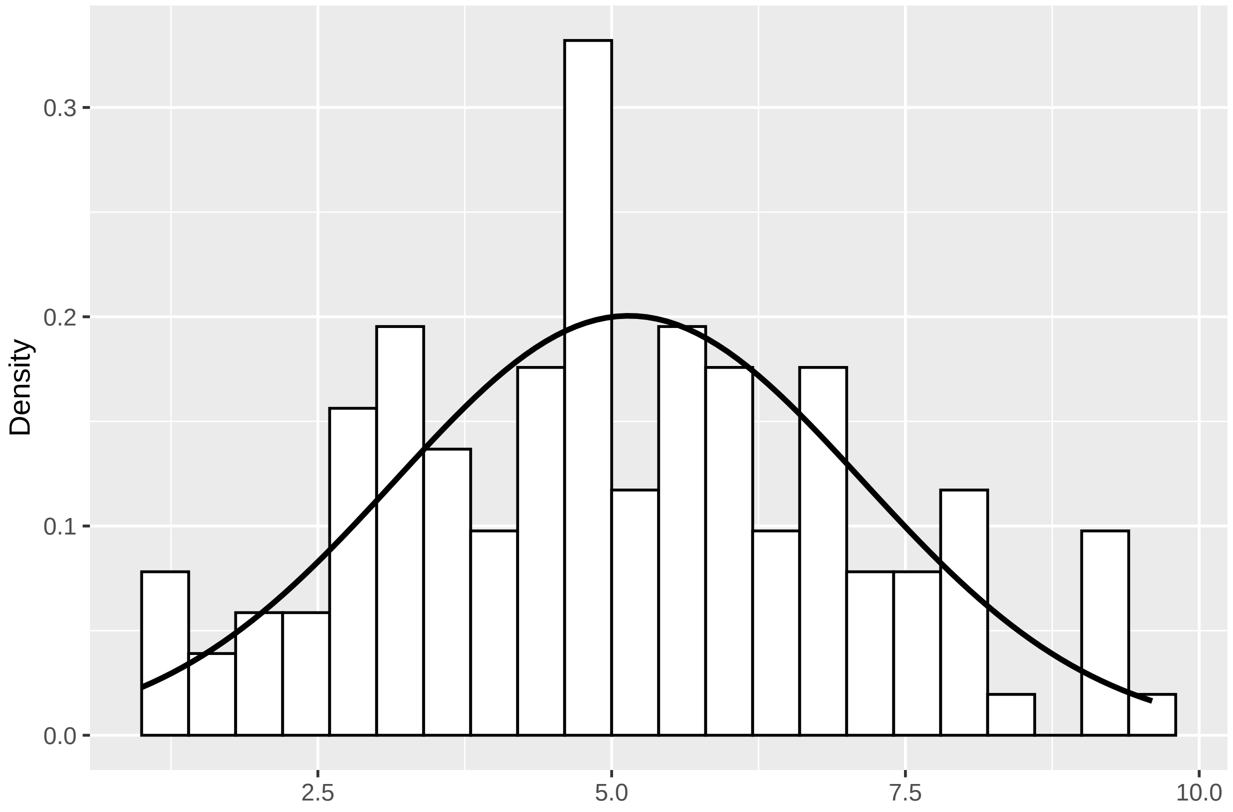
The participant in all three conditions were asked to fill a self-efficacy questionnaire of before and after doing the exercise. The questionnaire includes 5 questions which they were asked to rate from 0 (highly certain cannot do) to 10 (highly certain can do). Reliability analysis of self-efficacy questions shows an acceptable reliability level (alpha > 0.7)

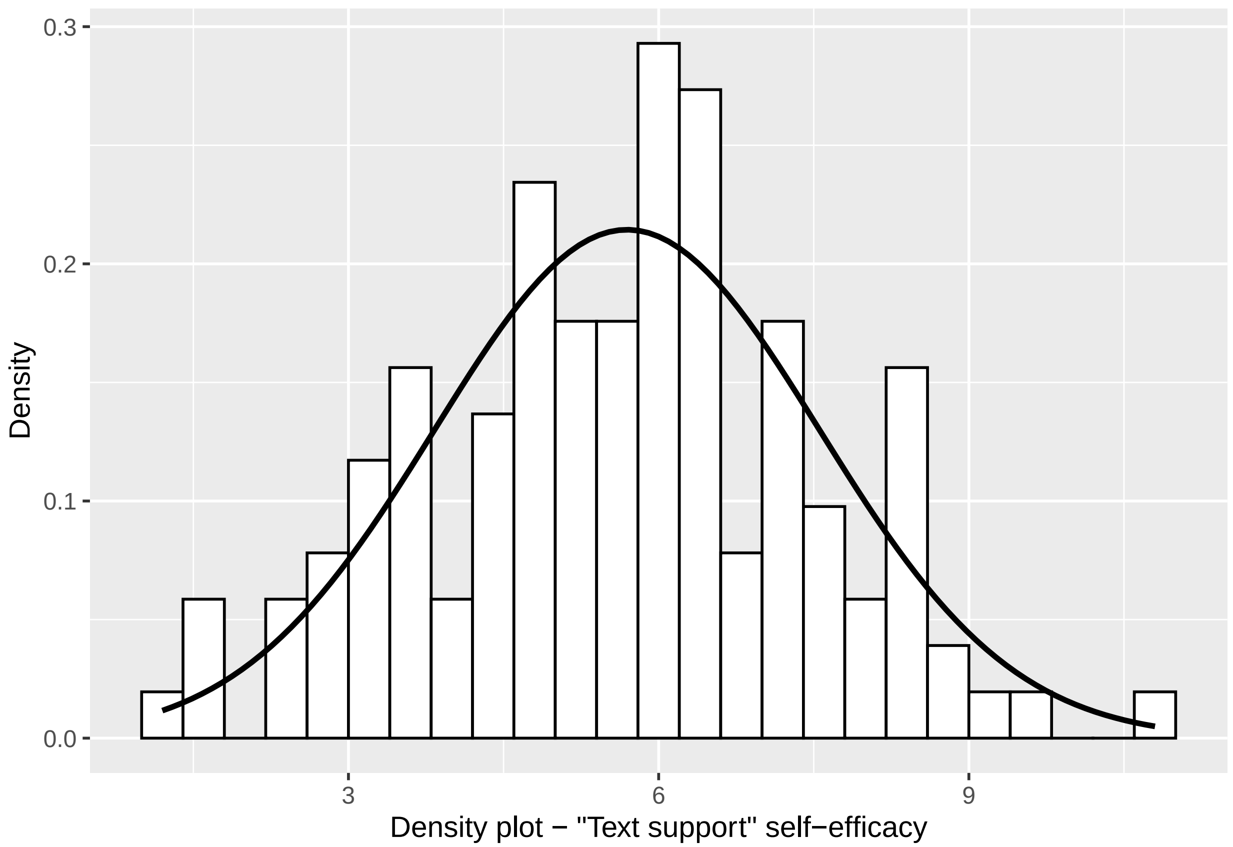
4.2 Data preparation

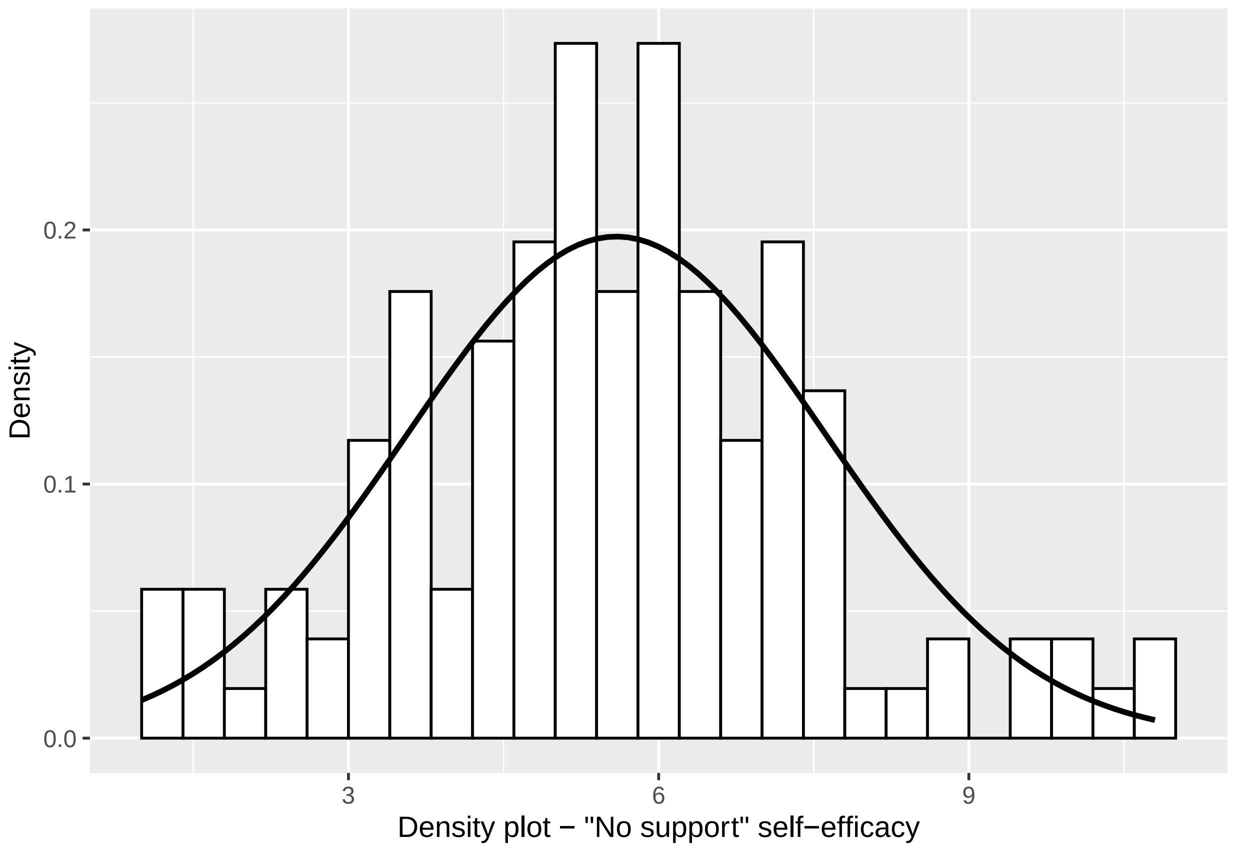
Since the reliability level was acceptable, we continued with getting a unified score for self-efficacy First, we calculated the average of the pre and post questionnaire. Then, we transfer the data into another structure (Id, Group, Session, Score). The new structure can be used to fit a generalized model. After that, we subset the data to the three groups (i.e., chatbot, text support, no support).

4.3 Assumption checking

Before analysing the data, we checked for distribution normality. This was done visually for the 3 conditions:



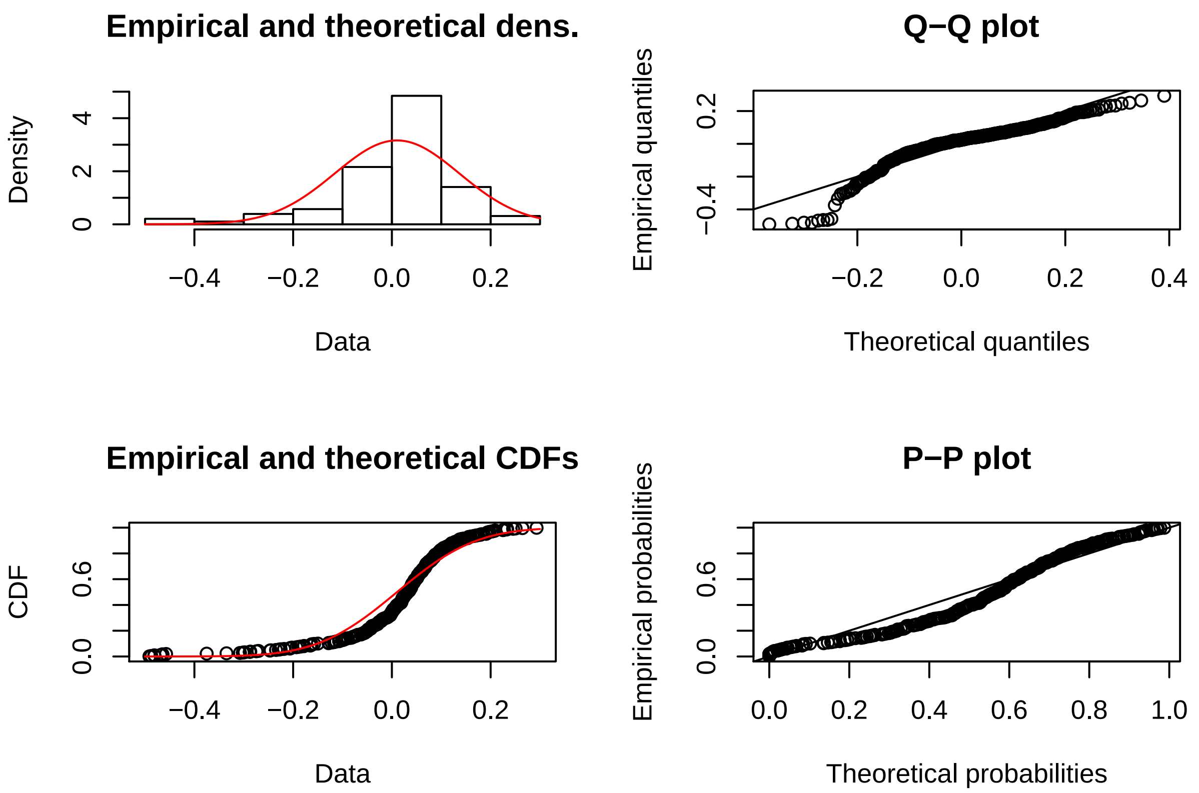

0.3
0.2
0.1
0.0
Density
3
6
9
Density plot − "Text support" self−efficacy



The data in the histograms shows a clear deviation from normal distribution.

4.4 Analysis of data

A generalized multilevel mixed effect model was fitted, wherein as a random effect, we used participant, and as fixed effects, we used the pre and post sessions. The model has a random intercept and a fixed slope, as we are assuming that all participants have the same direction but with various starting points. First, we checked if the residuals fits the distribution in case of using Gamma distribution.


Empirical and theoretical dens.

Q−Q plot

Empirical and theoretical CDFs

P−P plot

The plots looks reasonable. We continued analysing the data using the same model.

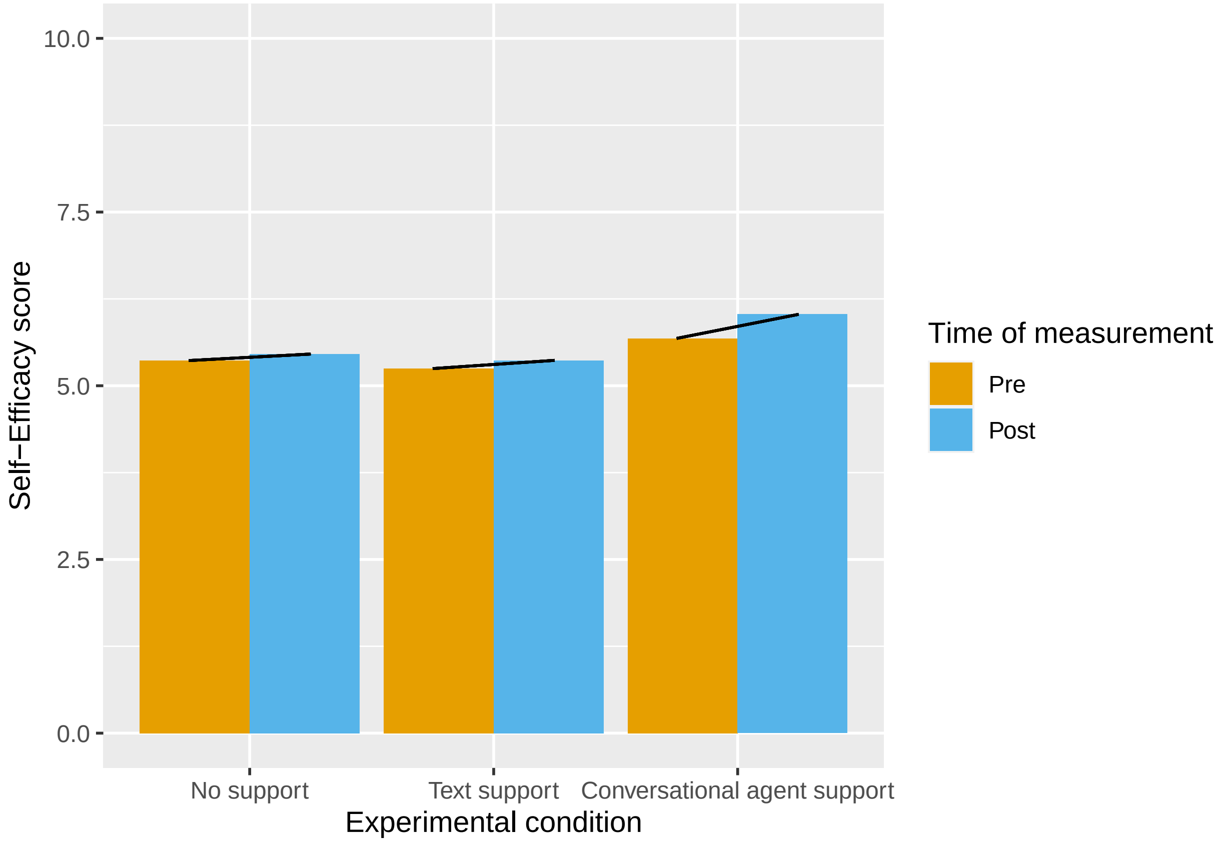
```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: Gamma  ( inverse )
## Formula: ScoreRverse ~ Group + Session + Group:Session + (1 | ParticipantID)
##    Data: SeEf
##
##      AIC      BIC   logLik deviance df.resid
##   1257.2   1288.8   -620.6   1241.2      376
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.9841 -0.1405  0.1551  0.3941  1.5445
##
## Random effects:
##  Groups        Name        Variance Std.Dev.
##  ParticipantID (Intercept) 0.009455 0.09724
##  Residual                  0.043574 0.20874
## Number of obs: 384, groups:  ParticipantID, 192
##
## Fixed effects:
##                                 Estimate Std. Error t value Pr(>|z|)
## (Intercept)                     0.301648   0.023957  12.591  < 2e-16 ***
## GroupNo Support                -0.020947   0.032982  -0.635  0.52536
## GroupText Support              -0.046626   0.033624  -1.387  0.16553
## SessionPost                     0.011501   0.004076   2.822  0.00478 **
## GroupNo Support:SessionPost   -0.008841   0.005550  -1.593  0.11115
## GroupText Support:SessionPost -0.008178   0.005532  -1.478  0.13930
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) GrpNSp GrpTxS SssnPs GNS:SP
## GroupNSpprt -0.656
## GrpTxtSpprt -0.668  0.464
## SessionPost -0.080  0.057  0.056
## GrpNSppr:SP  0.058 -0.080 -0.041 -0.734
## GrpTSppr:SP  0.059 -0.042 -0.078 -0.737  0.541
```

Table 3: Analysis of Deviance Table (Type II Wald chisquare tests)

	Chisq	Df	Pr(>Chisq)
Group	2.274	2	0.3208
Session	6.176	1	0.01295
Group:Session	3.075	2	0.2149

The self-efficacy between the groupd does not show a significant difference (p>0.05). However, there is a significant increase between the pre and post measurements (p<0.05).

The following bar chart shows the difference between the pre and post questionnaire means for the 3 conditions.


Self−Efficacy score
10.0
7.5
5.0
2.5
0.0
No support
Text support
Conversational agent support
Experimental condition
Time of measurement
Pre
Post