Social presence questionnaire analysis

Zilla Garama

September 2019

Contents

Introduction

This document presents inferential statistical analyses of participants' perception of the social presence of virtual characters as reported in the paper:

Creating windows to the soul: How eye gaze behaviour can make virtual characters more believable, more socially present and have mental states attributed to them

The OSF form belonging to this paper can be found here: https://osf.io/8u2kf

Authored by Zilla Garama, Ding Ding, and Willem-Paul Brinkman.

This questionnaire on the perception of the social presence of the virtual characters is from J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis titled "Equilibrium theory revisited: Mutual gaze and personal space in virtual environments". It was published in Presence: Teleoperators & Virtual Environments, pages 583–598, 2001.
Participants evaluated to what extent each of the sentences matched with the virtual character (VC), using a seven-point Likert scale (1: "Totally disagree", 7: "Totally agree") for each item. The virtual characters have been given a randomised name, as they differ, a placeholder has been used here:

- I perceive that *Name of the VC* is another person in the room with me.
  Totally disagree 1 2 3 4 5 6 7 Totally agree
- I feel that *Name of the VC* is watching me and is aware of my presence.
  Totally disagree 1 2 3 4 5 6 7 Totally agree
- The thought that *Name of the VC* is not a real person crosses my mind often.
  Totally disagree 1 2 3 4 5 6 7 Totally agree
- *Name of the VC* appears to be sentient (conscious and alive) to me.
  Totally disagree 1 2 3 4 5 6 7 Totally agree
- I perceive *Name of the VC* as being only a computerized image, not as a real person.
  Totally disagree 1 2 3 4 5 6 7 Totally agree

Abbreviations: * virtual character (VC) * Control virtual character (CVC) * Gaze aware virtual character (GAVC)

Libraries used:

```r
library(ggplot2)         # plotting & data
library(psych)           # reliability function
library(pastecs)         # plotting & data
library(lsr)             # effect size
```

Read in the questionnaire data:

```r
SPQ_data <- read.csv("SPQ_data.csv", header = TRUE, stringsAsFactors = FALSE,
                     fileEncoding="UTF-8-BOM")

#Gaze behaviour data is needed for information about the roles of the virtual characters
GB_data <- read.csv("GB_data.csv", header = TRUE, stringsAsFactors = FALSE,
                     fileEncoding="UTF-8-BOM")
GB_data<-GB_data[-c(25:28),]
```

Social presence questionnaire

Data preparation

The participants were asked to fill this questionnaire in for both virtual characters seperately. In the raw data this is represented by A and B for the physical position (left or right) of the virtual characters in the virtual environment during the experiment. However, the roles of the virtual characters (CVC or GAVC) have been randomized so A and B don't directly correspond to the roles of the virtual characters. Before further analysis can be done, information about the roles need to be added to a data table:

```r
#calculate the roles per virtual character:
CVC_position = ifelse(GB_data$GB_1_CVC %in% c("UMA_2", "UMA_3S", "UMA_4", "UMA_5S"),
                      "A", "B")
GAVC_position = ifelse(CVC_position %in% c("A"), "B", "A")

#Put the roles in a table to get a clearer overview:
GB_roles <- data.frame(Participant_number = GB_data$Participant_number,
                       CVC = GB_data$GB_1_CVC, GAVC = GB_data$GB_2_GAVC,
                       CVC_role = CVC_position, GAVC_role = GAVC_position)

print(GB_roles)
```

```
##    Participant_number    CVC   GAVC CVC_role GAVC_role
## 1                 180  UMA_4  UMA_5        A         B
## 2                 122 UMA_5S UMA_4S        A         B
## 3                 198  UMA_4  UMA_5        A         B
## 4                 182  UMA_5  UMA_4        B         A
## 5                 132 UMA_4S UMA_5S        B         A
## 6                 107  UMA_2  UMA_3        A         B
## 7                 184  UMA_3  UMA_2        B         A
## 8                 137 UMA_4S UMA_5S        B         A
## 9                 103  UMA_3  UMA_2        B         A
## 10                136  UMA_3  UMA_2        B         A
## 11                128  UMA_4  UMA_5        A         B
## 12                185 UMA_3S UMA_2S        A         B
## 13                163 UMA_2S UMA_3S        B         A
## 14                112  UMA_3  UMA_2        B         A
## 15                149 UMA_5S UMA_4S        A         B
## 16                173  UMA_3  UMA_2        B         A
```

```
## 17                195  UMA_2  UMA_3       A       B
## 18                115  UMA_4  UMA_5       A       B
## 19                169 UMA_4S UMA_5S       B       A
## 20                120 UMA_5S UMA_4S       A       B
## 21                146 UMA_5S UMA_4S       A       B
## 22                178 UMA_2S UMA_3S       B       A
## 23                119 UMA_3S UMA_2S       A       B
## 24                157 UMA_2S UMA_3S       B       A
```

Using the above mapping of roles with A/B we can seperate the questionnaire answers per role. Thus the data table of the social presence questionnaire for the CVC is:

```r
SPQ_CVC <- data.frame(SPQ_1_room=rep(NA,24), SPQ_2_presence_aware=rep(NA, 24),
                      SPQ_3_real_person=rep(NA, 24), SPQ_4_sentience=rep(NA, 24),
                      SPQ_5_computerized_image=rep(NA, 24), stringsAsFactors = FALSE)

for(i in 1:24){
  if(CVC_position[i] == "A"){
    SPQ_CVC[i,] <- SPQ_data[i, 2:6]
  }
  else{
    SPQ_CVC[i,] <- SPQ_data[i, 7:11]
  }
}

print(SPQ_CVC)
```

```
##      SPQ_1_room SPQ_2_presence_aware SPQ_3_real_person SPQ_4_sentience
## 1             5                    7                 7               1
## 2             6                    6                 3               5
## 3             5                    7                 7               2
## 4             4                    5                 4               4
## 5             6                    4                 7               5
## 6             5                    4                 3               2
## 7             4                    1                 7               2
## 8             3                    1                 2               3
## 9             6                    2                 2               6
## 10            6                    2                 2               4
## 11            1                    1                 6               5
## 12            6                    5                 5               3
## 13            1                    5                 7               1
## 14            3                    4                 4               5
## 15            7                    7                 1               5
## 16            2                    1                 4               3
## 17            6                    3                 2               5
## 18            3                    2                 3               2
## 19            4                    5                 6               1
## 20            5                    5                 7               6
## 21            5                    6                 2               3
## 22            4                    3                 5               2
## 23            1                    1                 1               3
## 24            3                    4                 3               4
##      SPQ_5_computerized_image
## 1                           7
## 2                           7
```

```
## 3                      7
## 4                      5
## 5                      3
## 6                      6
## 7                      7
## 8                      7
## 9                      4
## 10                     3
## 11                     6
## 12                     4
## 13                     7
## 14                     3
## 15                     4
## 16                     6
## 17                     3
## 18                     4
## 19                     7
## 20                     3
## 21                     3
## 22                     7
## 23                     5
## 24                     5
```

And the GAVC:

```r
SPQ_GAVC <- data.frame(SPQ_1_room=rep(NA,24), SPQ_2_presence_aware=rep(NA, 24),
                       SPQ_3_real_person=rep(NA, 24), SPQ_4_sentience=rep(NA, 24),
                       SPQ_5_computerized_image=rep(NA, 24), stringsAsFactors = FALSE)

for(i in 1:24){
  if(GAVC_position[i] == "A"){
    SPQ_GAVC[i,] <- SPQ_data[i, 2:6]
  }
  else{
    SPQ_GAVC[i,] <- SPQ_data[i, 7:11]
  }
}

print(SPQ_GAVC)
```

```
##      SPQ_1_room SPQ_2_presence_aware SPQ_3_real_person SPQ_4_sentience
## 1             5                    7                 7               1
## 2             6                    7                 3               7
## 3             5                    7                 6               2
## 4             6                    7                 3               6
## 5             6                    4                 7               5
## 6             4                    3                 3               2
## 7             6                    5                 3               4
## 8             5                    7                 2               5
## 9             5                    5                 2               4
## 10            6                    6                 2               4
## 11            5                    5                 6               5
## 12            3                    3                 5               3
## 13            6                    6                 7               1
## 14            3                    1                 4               5
```

```
## 15            3                3                5                3
## 16            2                3                4                3
## 17            6                6                2                5
## 18            5                6                3                6
## 19            6                6                5                1
## 20            5                5                7                6
## 21            5                7                2                4
## 22            6                4                5                3
## 23            1                1                1                3
## 24            7                6                3                5
##     SPQ_5_computerized_image
## 1                          7
## 2                          7
## 3                          7
## 4                          4
## 5                          3
## 6                          6
## 7                          5
## 8                          6
## 9                          5
## 10                         3
## 11                         4
## 12                         5
## 13                         7
## 14                         3
## 15                         4
## 16                         6
## 17                         3
## 18                         3
## 19                         7
## 20                         3
## 21                         3
## 22                         7
## 23                         5
## 24                         4
```

Combining into a single score

According to the analysis from the source paper (referenced at the top of this document), the individual items need to be combined ito a single score:

```
SPQ_simple_sum_CVC <- SPQ_CVC
SPQ_simple_sum_CVC[, c(3,5)] <- -SPQ_simple_sum_CVC[,c(3,5)]
SPQ_simple_sum_CVC <- rowSums(SPQ_simple_sum_CVC)

SPQ_simple_sum_GAVC <- SPQ_GAVC
SPQ_simple_sum_GAVC[, c(3,5)] <- -SPQ_simple_sum_GAVC[,c(3,5)]
SPQ_simple_sum_GAVC <- rowSums(SPQ_simple_sum_GAVC)

print(SPQ_simple_sum_CVC)
```

```
##  [1] -1  7  0  4  5  2 -7 -2  8  7 -5  5 -7  5 14 -4  9  0 -3  6  9 -3 -1
## [24]  3
```

```
print(SPQ_simple_sum_GAVC)
```

```
##  [1] -1 10  1 12  5  0  7  9  7 11  5 -1 -1  2  0 -2 12 11  1  6 11  1 -1
## [24] 11
```

```
SPQ_simple_sum <- data.frame(CVC = SPQ_simple_sum_CVC, GAVC = SPQ_simple_sum_GAVC)
```

Assumption checking: normal distribution

The analysis method depends on the normality of the data distribution. This is usually done visually:

```
stem(SPQ_simple_sum$CVC)
```

```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##   -0 | 775
##   -0 | 433211
##    0 | 00234
##    0 | 555677899
##    1 | 4
```

```
stem(SPQ_simple_sum$GAVC)
```

```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##   -0 | 21111
##    0 | 001112
##    0 | 556779
##    1 | 0111122
```

```
ggplot(SPQ_simple_sum,
       aes(CVC)) + geom_histogram(aes(y=..density..), binwidth = 1,
       colour="black", fill="white") + labs(x="Density plot SPQ CVC",
       y="Density") + stat_function(fun=dnorm,
       args=list(mean=mean(SPQ_simple_sum$CVC, na.rm=TRUE),
       sd=sd(SPQ_simple_sum$CVC, na.rm=TRUE)), colour="black", size=1)
```

Density plot SPQ CVC
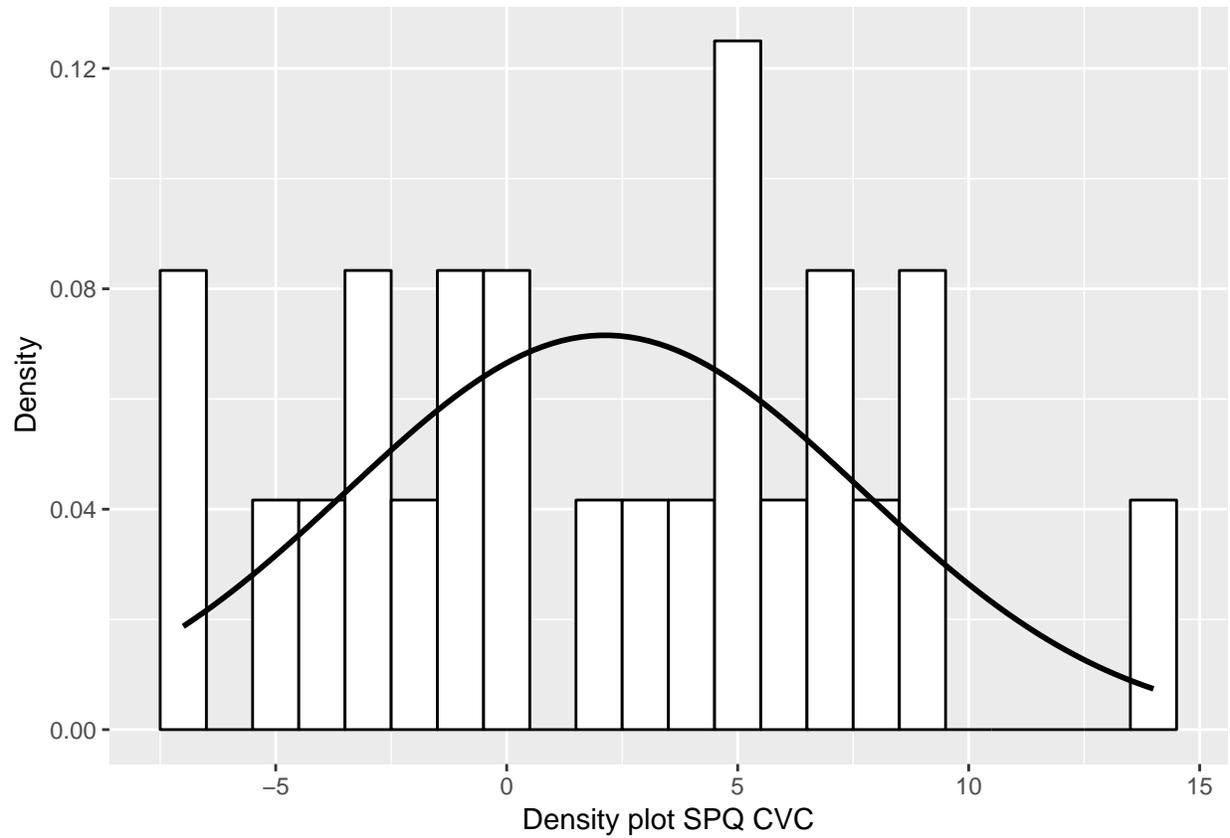
```r
ggplot(SPQ_simple_sum,
    aes(GAVC)) + geom_histogram(aes(y=..density..), binwidth = 1,
    colour="black", fill="white") + labs(x="Density plot SPQ GAVC",
    y="Density") + stat_function(fun=dnorm,
    args=list(mean=mean(SPQ_simple_sum$GAVC, na.rm=TRUE),
    sd=sd(SPQ_simple_sum$GAVC, na.rm=TRUE)), colour="black", size=1)
```

Density plot SPQ GAVC
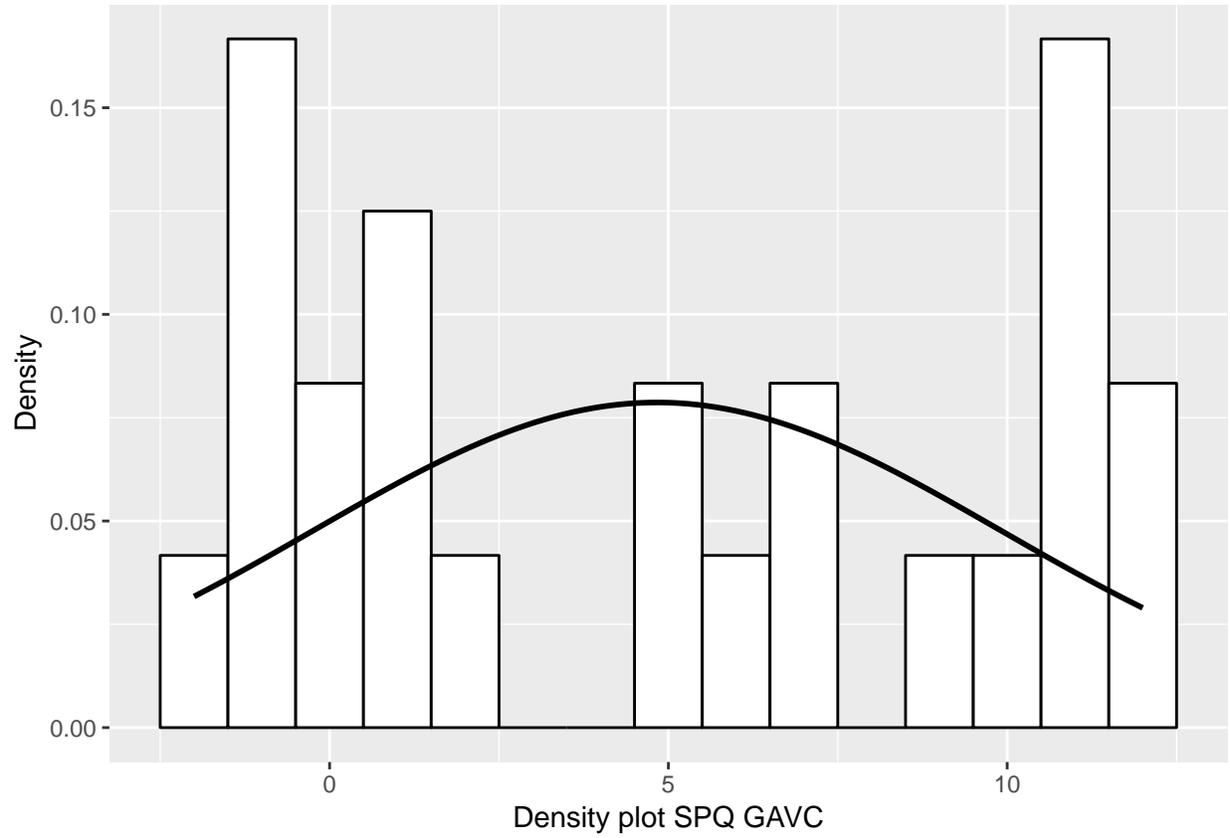
```
qplot(sample=SPQ_simple_sum$CVC, stat="qq") + labs(x="Q-Q plot SPQ CVC")
```

```
## Warning: `stat` is deprecated
```
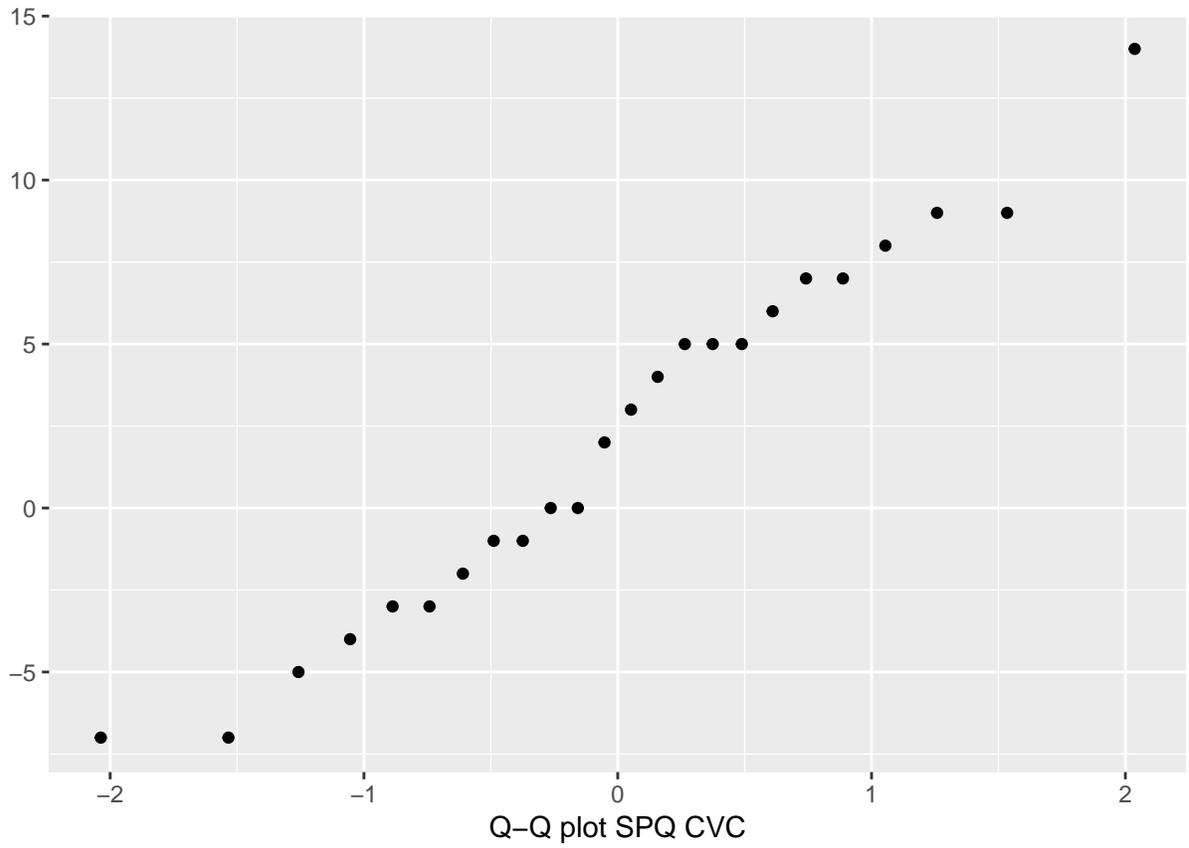
Q–Q plot SPQ CVC

```r
qplot(sample=SPQ_simple_sum$GAVC, stat="qq") + labs(x="Q-Q plot SPQ GAVC")
```

```
## Warning: `stat` is deprecated
```

Q–Q plot SPQ GAVC
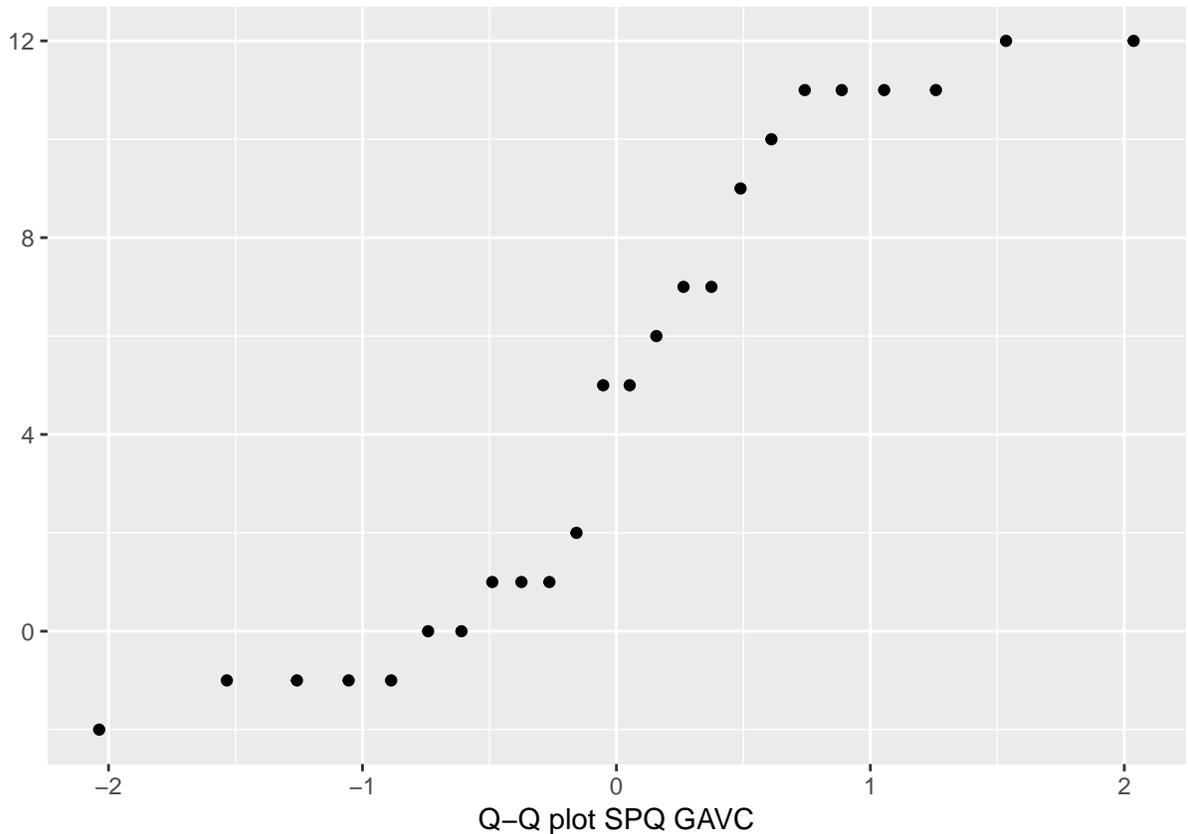
By visually inspecting the histogram and qqplot it seems that the data is not normally distributed. However, the sample size is rather small (n<30) so it is better to quantify the shape of the distribution:

```
round(stat.desc(SPQ_simple_sum[,c("CVC", "GAVC")], basic = FALSE, norm = TRUE), digits = 3)
```

```
##                  CVC    GAVC
## median         2.500   5.000
## mean           2.125   4.833
## SE.mean        1.138   1.035
## CI.mean.0.95   2.354   2.141
## var           31.071  25.710
## std.dev        5.574   5.071
## coef.var       2.623   1.049
## skewness       0.088   0.120
## skew.2SE       0.093   0.127
## kurtosis      -0.977  -1.693
## kurt.2SE      -0.532  -0.922
## normtest.W     0.970   0.875
## normtest.p     0.664   0.007
```

The skew.2SE and kurt.2SE are smaller than 0.98 (ignoring the plus or minus sign), which means the skew and kurtosis are not significant (at $p < 0.05$). The p-values (indicated by normtest.p) obtained by the Shapiro-Wilk test show that the CVC is normally distributed. The GAVC, however, has a p-value < 0.05 and thus deviates from normality.

Paired difference test

The results were analysed using either a parametric or non-parametric paired difference test depending on the normality of the distribution. In this case the GAVC data is not normally distributed and thus a Wilcoxon signed-rank test was used:

```
wilcox_result <- wilcox.test(SPQ_simple_sum$CVC, SPQ_simple_sum$GAVC, paired = TRUE, exact = TRUE, corr
```

```
## Warning in wilcox.test.default(SPQ_simple_sum$CVC, SPQ_simple_sum$GAVC, :
## cannot compute exact p-value with ties
```

```
## Warning in wilcox.test.default(SPQ_simple_sum$CVC, SPQ_simple_sum$GAVC, :
## cannot compute exact p-value with zeroes
```

```
wilcox_result
```

```
##
##  Wilcoxon signed rank test
##
## data:  SPQ_simple_sum$CVC and SPQ_simple_sum$GAVC
## V = 44.5, p-value = 0.0237
## alternative hypothesis: true location shift is not equal to 0
```

The p-value is 0.02 which is < 0.05 which means that we can reject the null hypothesis and the result is statistically significant.

```
Zstat <- qnorm(wilcox_result$p.value/2)
Zstat
```

```
## [1] -2.261979
```

The Z-value is -2.26, which will be used to calculate the effect size.

Effect size

The effect size is calculated using Pearson's r for a non-parametric test:

```
abs(Zstat)/sqrt(24)
```

```
## [1] 0.4617245
```

The effect size r is 0.46.