Assignment 3 Chi-square

Yuka Tatsumi

2021-11-27

Part 1: Personality and animal preference

Introduction

Method

Chi-square test is used to investigate if there is an association between personality and animal preference. Based on the provided data, the frequencies for each variable are calculated, using excel. Based on the calculation, Chi-square test is conducted using R.

Result

The result of Chi-square(X-squared = 13.662, df = 3, p-value=0.003403, X-squared for association) indicates that there is a statistically significant association between personality and animal preference.

Table 1 Observed Frequency (Personality and animal preference)

Personality	Bat	Rabbit	Tiger	Whale
extrovert	16	13	20	10
introvert	11	18	2	10

Table2 Expected Frequency (Personality and animal preference)

Personality	Bat	Rabbit	Tiger	Whale
extrovert	16	18	13	12
introvert	11	13	9	8

The graph below (Figure 1) indicates that more respondents who like tigers or bats are extroverts than introverts while more people who like rabbits are introverts than extroverts. A much larger portion of respondents who like tiger are extroverts than bats. Half of people who like whale are introverts and half of them are extroverts.

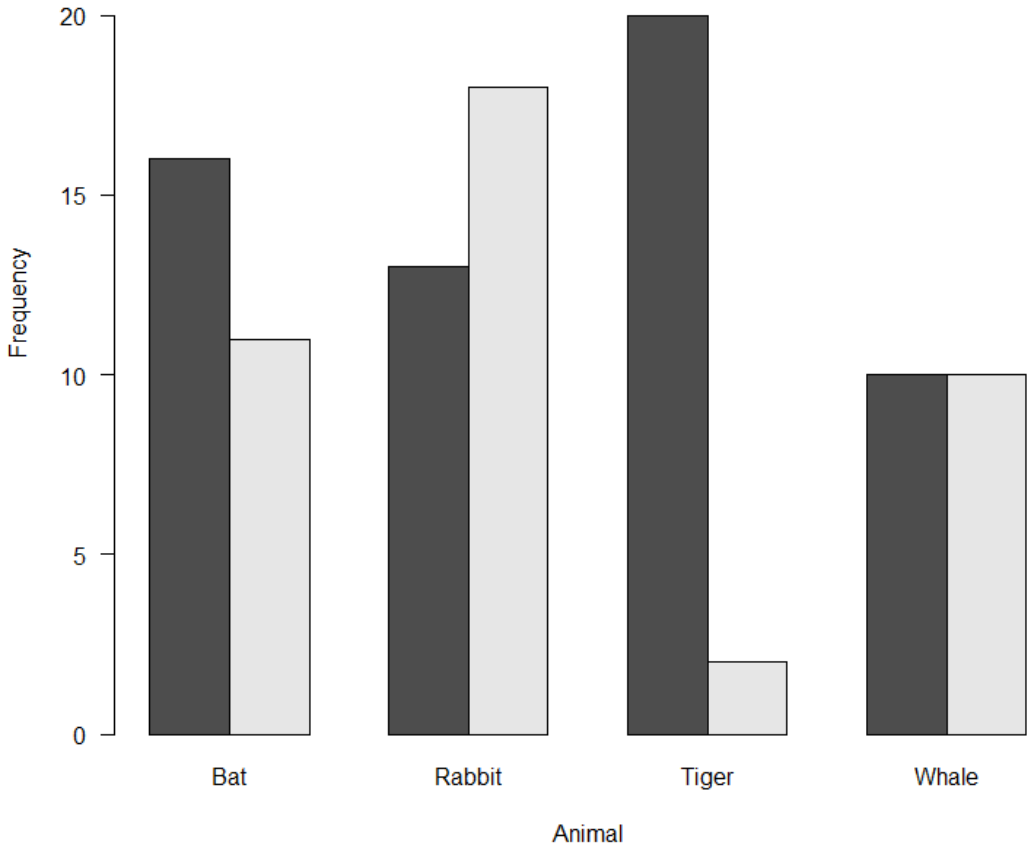

Figure 1 Bar Chart (Personality and animal preference)

Pearson's residuals(Table 3) and correlation matrix (Figure 2) suggest that tiger have a strong positive association with extroverts and a strong negative correlation with introverts. No association between bats and these personality types(extroverts/introverts) is identified. Rabbits have a relatively strong positive association with introverts and negative associations with extroverts. Whales have the same trend with a weaker association than rabbits.

Table 3 Pearson's Residuals (Personality and animal preference)

personality	**Bat**	**Rabbit**	**Tiger**	**Whale**
extrovert	0.0	-1.2	1.9	-0.5
introvert	-0.0	1.5	-2.3	0.6

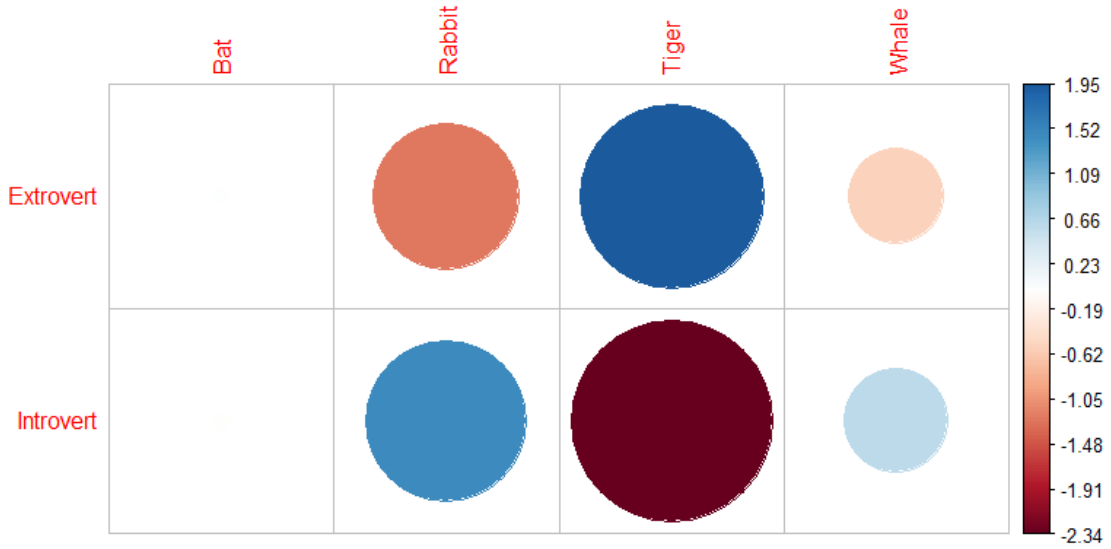

Figure 2 Correlation Plot (Residuals:Personality and animal preference)

The tables (Table 4 & 5) and correlation matrix for contribution to Chi square test score(Figure 3) indicates that tigers have the highest contribution % followed by rabbits and whales. Bats have the lowest contribution to the test score. It means that tigers and personality have the strongest association while bats and the personalities(extroverts/introverts) have the weakest association.

Table 4 Chi-square Components (Personality and animal preference)

personality	Bat	Rabbit	Tiger	Whale
extrovert	0	1.53	3.8	0.27
introvert	0	2.2	5.46	0.4

Table 5 Contribution % to Chi-square score (Personality and animal preference)

personality	Bat	Rabbit	Tiger	Whale
extrovert	0.00225	11.2	27.8	2.01
introvert	0.00324	16.1	40	2.89

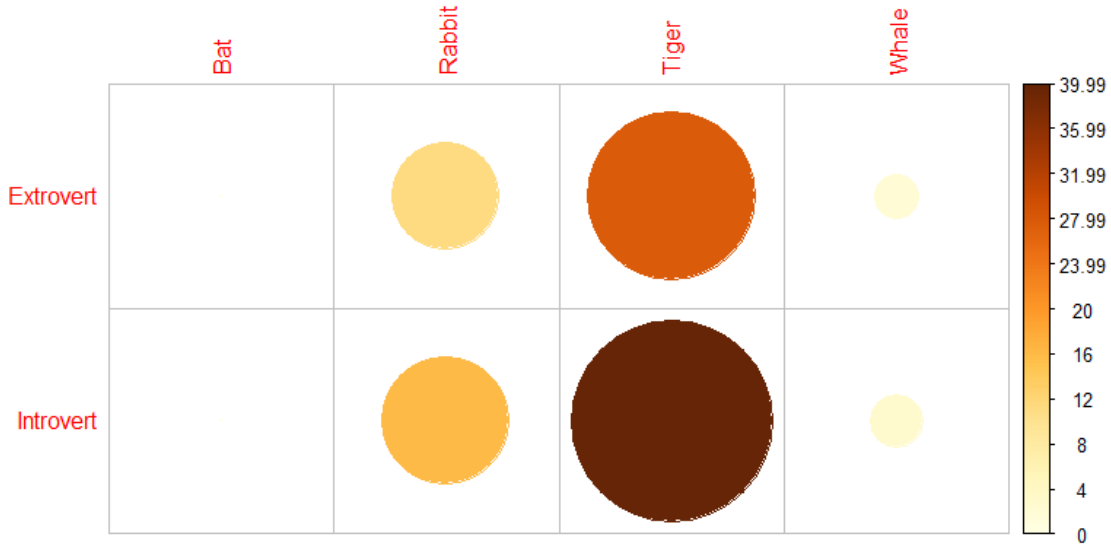

Figure 3 Correlation Plot(Personality and animal preference: Contribution % to Chi-square score

Discussion

The result of Chi-square test shows that there is a statistically significant association between personality of students and animal preference. Person who like tigers are more likely to be extroverts while person who like rabbits or whales are more likely to be introverts. Tigers have the strongest positive association with extroverts. Rabbits and whales have a positive association with introverts and rabbits have stronger association with introverts than whales. Bats have the weakest association with personalities(extroverts and introverts).

PART 2 : Stress caused by homework or exams

Introduction

It is generally said that there is a difference in attitude toward study between girls and boys and girls take study more seriously and study harder. Homework and exams is one of major activities/milestone in studies, which some students sometimes find it stressful.

The aim of the analysis in this report part 2 is to investigate if there is an association between gender and stress caused by homework or exams regarding 12-18 year old students in Sweden. Additionally, whether there is a difference between 12-15 year old girls and 16-18 year old girls regarding a feeling of stress due to exams and homework.

Method

The data in SCB website (https://www.scb.se/hitta-statistik/temaomraden/jamstalldhet/jamstalld-utbildning/trivsel-och-stress/) is used for this analysis. The data is based on the survey conducted by barns levnadsförhållanden (Barn-ULF) and SCB. The survey was targeted for 12-18 year old children in Sweden and approximately 400,000 girls and 400,000 boys participated in the survey. Chi-square test is conducted, using R in order to investigate if there is an association between gender and stress. Chi-square test without Yates' continuity correction is conducted because the number of sample is large. In order to make a contingency table, the number of boys and girls who often feel stressed due to homework or exams (frequency) are calculated using excel based on the total number of participants and the proportion who often feels stressed (%) which is available from the SCB web site mentioned earlier.

Result

The result of the Chi-squared test (X-squared = 48829, df = 1, p-value < 0.01) suggested that there is a statistically significant association between gender and feeling of stress due to exam and homework.

Table 6 Observerd Frequency (Gender and stress caused by homework and exams)

gender	Stressed	Not Stressed
boys	71010	346696
girls	152117	237927

Table 7 Expected Frequency(Gender and stress caused by homework and exams)

gender	Stressed	Not Stressed
boys	115384	302322
girls	107743	282301

The graph below (Figure 4) also indicates that a much smaller portion of boy respondents feel stressed due to exam and homework than girl respondents.

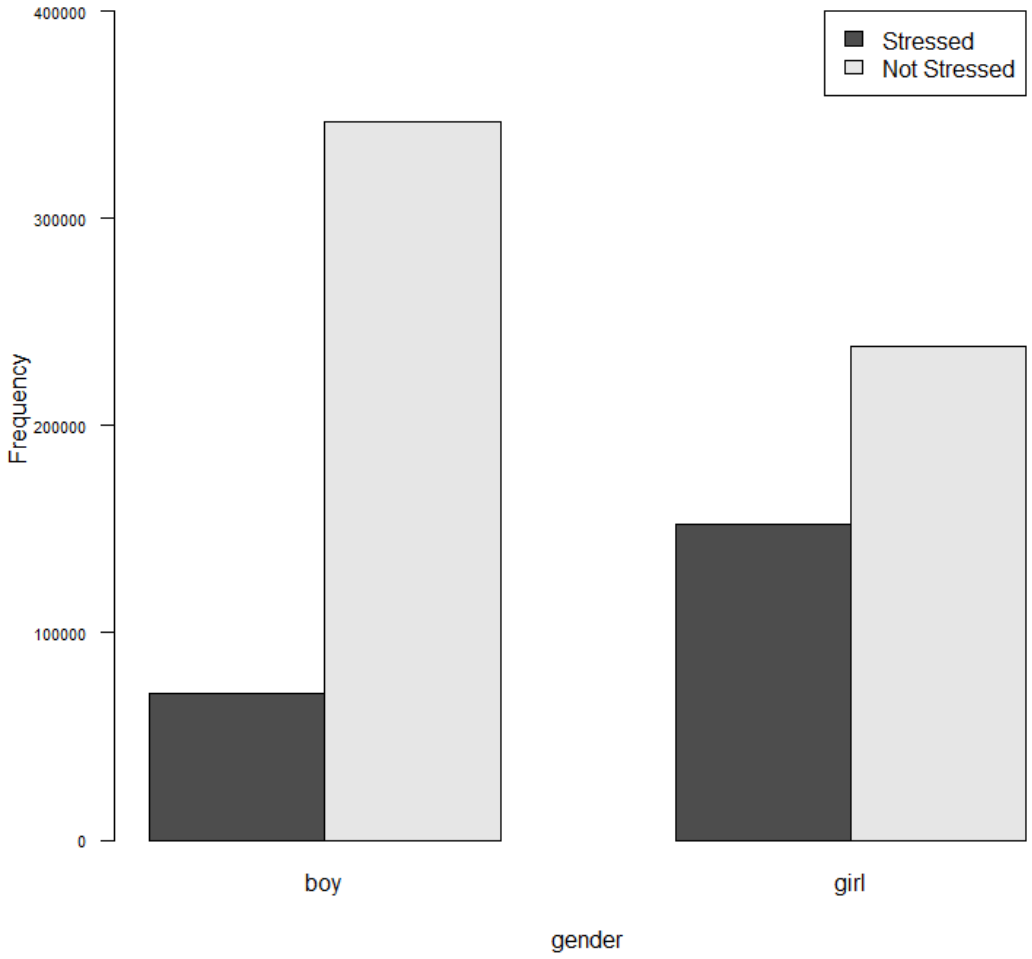

Figure 4 Bar chart (Gender and stress caused by homework and exams)

Pearson's residuals (Table 8) and correlation matrix (Figure 5) suggest that boys have a strong negative association with stress du to homework and tests while girls have a strong positive association with stress by homework and exams. This implies that girls are more likely to feel stress due to homework and exams than boys.

Table 8 Pearson's residuals (Gender and stress caused by homework and exams)

gender	Stressed	Not Stressed
boys	-131	81
girls	135	-84

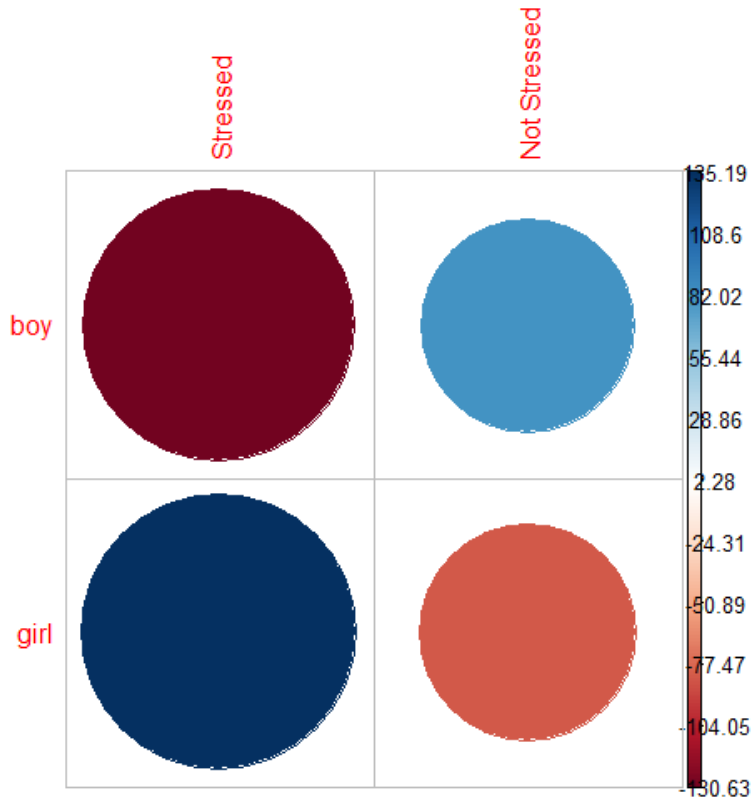

Figure 5 Correlation Plot (Residual:Gender and stress caused by homework and exams)

Regarding the association between age and stress due to homework and exams for girl, the result of the Chi-squared test (X-squared = 10211, df = 1, p-value < 0.01, X-squared test for association) suggests that there is a statistically significant association between the age (12-15 years old girls and 16-18 years old) and feeling stressed due to exam and homework

Table 9 Observed Frequency (Age and stress due to homework and exams for girls)

Age	Stressed	Not Stressed
12-15 years old	73307	155776
16-18 years old	77261	83700

Table 10 Expected Frequency (Age and stress due to homework and exams for girls)

Age	Stressed	Not Stressed
12-15 years old	88433	140650
16-18 years old	62135	98826

The graph below (Figure 6) indicates that a larger portion of 16-18 years old girl respondents feel stress than 12-15 years old girls. Almost half of the 16-18 years old girls feel stress while almost 30% of 12-15 year old girl respondents feel stress due to homework and exams.

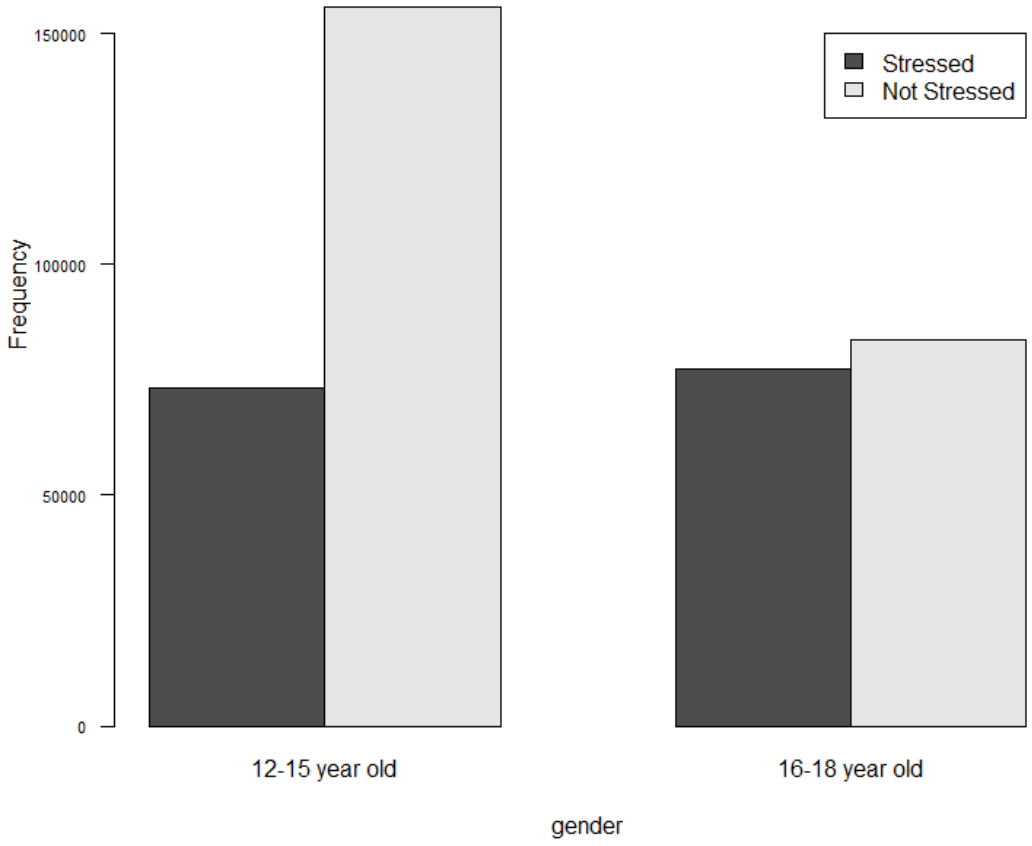

Figure 6 Bar chart (Age and stress due to homework and exams for girls)

Pearson's residuals (Table 11) and correlation matrix (Figure 7) suggest that 16-18 years old girls have a positive association with stress by homework and tests while 12-15 years old girls have a negative association with stress by homework and tests. It means that 16-18 years old girls are more likely to feel stress by homework and tests than 12-15 years old girls.

Table 11 Pearson's residuals (Age and stress due to homework and exams for girl)

Age	Stressed	Not Stressed

12-15 years old	-51	40
16-18 years old	61	-48

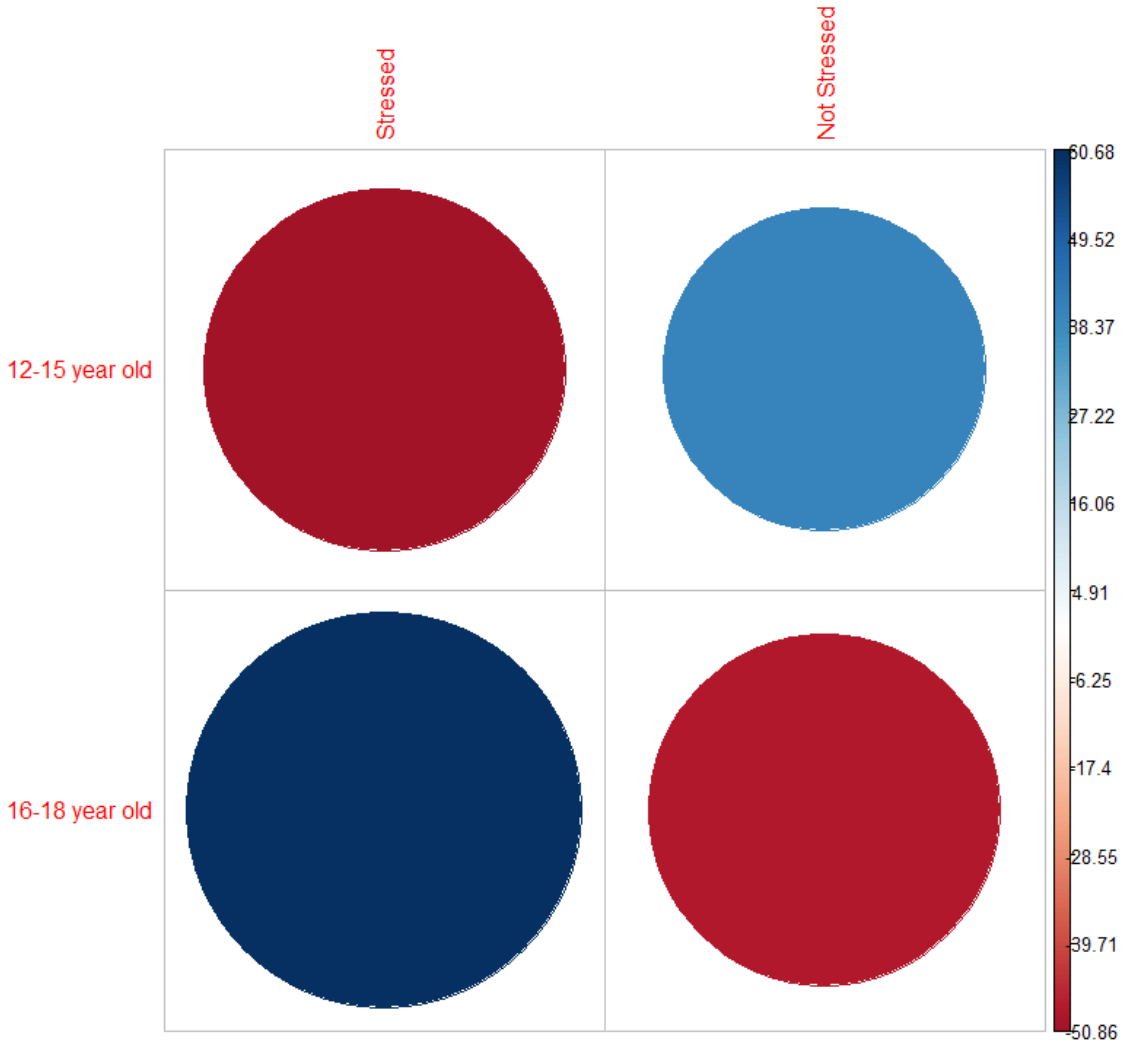

Figure 7 Correlation Plot (residual) for Age and stress due to homework and exams for girls

Discussion

The result of Chi-squared test indicates that there is an association between gender and feeling of stress due to exam and homework. Much larger portion of girls feel stressed due to homework and exams than boys. Whether and how this trend will change in the future is interesting to see as gender equality is making progress.

Additionally, there is a statistically significant difference between 12-15 year old girls and 16-18 year old girls regarding feeling stressed due to exam and homework. 16-18 year old girls feel more stressed than 12-15 year old. There can be different reasons for this

difference. The reason might be increased amount of study, increased level of difficulty in study, pressures/workload before applying for university etc.

References

Appendix

Part 1

```
counts <- matrix(c(16,13,20,10,11,18,2,10), 2, 4, byrow=TRUE)
dimnames(counts) <- list("2"=c("Extrovert", "Introvert"), "4"=c("Bat",
"Rabbit", "Tiger", "Whale"))
counts
chisq_result1 <- chisq.test(counts, correct=FALSE)
```

Table 1

Observed Frequency (Personality and animal preference)

```
observed1<-chisq_result1$observed
observed1 %>%
  as_hux() %>%
  add_colnames() %>%
  set_number_format(0)%>%
  set_outer_padding(5) %>%
  set_align("center")%>%
  insert_column(c("personality","extrovert","introvert"))%>%
  theme_basic()
```

Chisquare test (Personality and animal preference)

```
result_table<- chisq_result1
result_table
```

Table 2

Expected Frequency (Personality and animal preference)

```
chisq_result1$expected %>%
  as_hux() %>%
  add_colnames() %>%
  set_number_format(0)%>%
  set_outer_padding(5) %>%
  set_align("center")%>%
  insert_column(c("personality","extrovert","introvert"))%>%
  theme_basic()
```

Figure 1

Bar Chart (Personality and animal preference)

```
barplot(counts, beside=T, xlab="Animal", ylab="Frequency",
legend.text=T, args.legend = list(x=3, y=1500),
ylim=c(0,1.2*max(counts)),col=gray.colors(2),las=1)
```

Table 3

Pearson's Residuals (Personality and animal preference)

```
chisq_result1$residuals %>%
  as_hux() %>%
  add_colnames() %>%
  set_number_format(1)%>%
  set_outer_padding(5) %>%
  set_align("center")%>%
  insert_column(c("personality","extrovert","introvert"))%>%
  theme_basic()
```

Figure 2

Correlation Plot (Residuals: Personality and animal preference)

```
corrplot(chisq_result1$residuals, is.cor = FALSE)
```

Table 4

Chi-square Components (Personality and animal preference)

```
result_res2<-round(chisq_result1$residuals^2, 2) # Chi-square Components
result_res2 %>%
  as_hux() %>%
  add_colnames() %>%
  set_outer_padding(4) %>%
  set_align("center")%>%
  insert_column(c("personality","extrovert","introvert"))%>%
  theme_basic()
```

Table 5

Contribution % to Chi-square score(Personality and animal preference)

```
contribution1<-100*chisq_result1$residuals^2/chisq_result1$statistic

contribution1 %>%
  as_hux()%>%
  add_colnames() %>%
  set_align("center")%>%
  insert_column(c("personality","extrovert","introvert"))%>%
  theme_basic()
```

Figure 3

Correlation Plot(Personality and animal preference:Contribution % to Chi-square score)

```
corrplot(100*chisq_result1$residuals^2/chisq_result1$statistic, is.cor=FALSE)
```

PART 2

Table 6

Observerd Frequency (Gender and stress caused by homework and exams)

```
count_stress = matrix(c(71010,152117,346696,237927),nrow=2)

dimnames(count_stress)<-list(gender=c("boy", "girl"),stress=c("Stressed","Not Stressed"))

count_stress %>%
  as_hux() %>%
  add_colnames() %>%
  set_number_format(0)%>%
  #set_caption("Observed frequency") %>%
  set_outer_padding(5) %>%
  set_align("center")%>%
  insert_column(c("gender","boys","girls"))%>%
  theme_basic()
```

Chisquare test (Gender and stress caused by homework and exams)

```
result_stress<-chisq.test(count_stress,correct=FALSE)
result_stress
```

Figure 4

Bar chart (Gender and stress caused by homework and exams)

```
options(scipen = 999)
barplot(t(count_stress), beside=T, xlab="gender", ylab="Frequency",
        legend.text=T, args.legend = list(x=6, y=400000),
        ylim=c(0,1.2*max(count_stress)),col=gray.colors(2),las=1)
```

Table 7

Expected Frequency (Gender and stress caused by homework and exams)

```
result_stress$expected %>%
  as_hux() %>%
  add_colnames() %>%
  set_number_format(0)%>%
  set_outer_padding(5) %>%
  set_align("center")%>%
  insert_column(c("gender","boys","girls"))%>%
  theme_basic()
```

Table 8

Pearson's Residuals(Gender and stress caused by homework and exams)

```
result_stress$residuals %>%
  as_hux() %>%
  add_colnames() %>%
```

```
  set_number_format(0)%>%
  set_outer_padding(5) %>%
  set_align("center")%>%
  insert_column(c("gender","boys","girls"))%>%
  theme_basic()
```

Figure 5

Correlation Plot(residual:Gender and stress caused by homework and exams)

```
corrplot(result_stress$residuals, is.cor = FALSE)
```

Table 9

Observed Frequency (Age and stress due to homework and exams for girls)

```
count_stress2 = matrix(c(73307,77261,155776,83700),nrow=2)

dimnames(count_stress2)<-list(gender=c("12-15 year old", "16-18 year
old"),stress=c("Stressed","Not Stressed"))

count_stress2 %>%
  as_hux() %>%
  add_colnames() %>%
  set_number_format(0)%>%
  set_outer_padding(5) %>%
  set_align("center")%>%
  insert_column(c("Age","12-15 years old","16-18 years old"))%>%
  theme_basic()
```

Chi square test (Age and stress due to homework and exams for girls)

```
result_stress2<-chisq.test(count_stress2,correct=FALSE)
result_stress2
```

Table 10

Expected Frequency (Age and stress due to homework and exams for girls)

```
result_stress2$expected %>%
  as_hux() %>%
  add_colnames() %>%
  set_number_format(0)%>%
  set_outer_padding(5) %>%
  set_align("center")%>%
  insert_column(c("Age","12-15 years old","16-18 years old"))%>%
  theme_basic()
```

Figure 6

Bar chart (Age and stress due to homework and exams for girls)

```
options(scipen = 999)
barplot(t(count_stress2), beside=T, xlab="gender", ylab="Frequency",
        legend.text=T, args.legend = list(x=6, y=150000),
```

```
        ylim=c(0,1.2*max(count_stress2)),col=gray.colors(2),las=1,
cex.axis=0.7)
```

Table 11

Pearson's Residuals (Age and stress due to homework and exams for girl)

```
result_stress2$residuals %>%
  as_hux() %>%
  add_colnames() %>%
  set_number_format(0)%>%
  set_outer_padding(5) %>%
  set_align("center")%>%
  insert_column(c("Age","12-15 years old","16-18 years old"))%>%
  theme_basic()
```

Figure 7

Correlation Plot (residual) for Age and stress due to homework and exams for girls

```
corrplot(result_stress2$residuals, is.cor = FALSE)
```