Practical 4

Student Number: S4555376, Name: Trevor Atkins

9/29/2020

Part A: Reading and listening

1

```
Student <- seq(1,8)
Reading <- c(20, 40, 60, 80, 100, 120, 140, 160)
Listening <- c(65, 69, 73, 77, 80, 84, 89, 95)
Practical4a = data.frame(Student, Reading, Listening)
Practical4a$Student <- as.factor(Practical4a$Student)
```

2

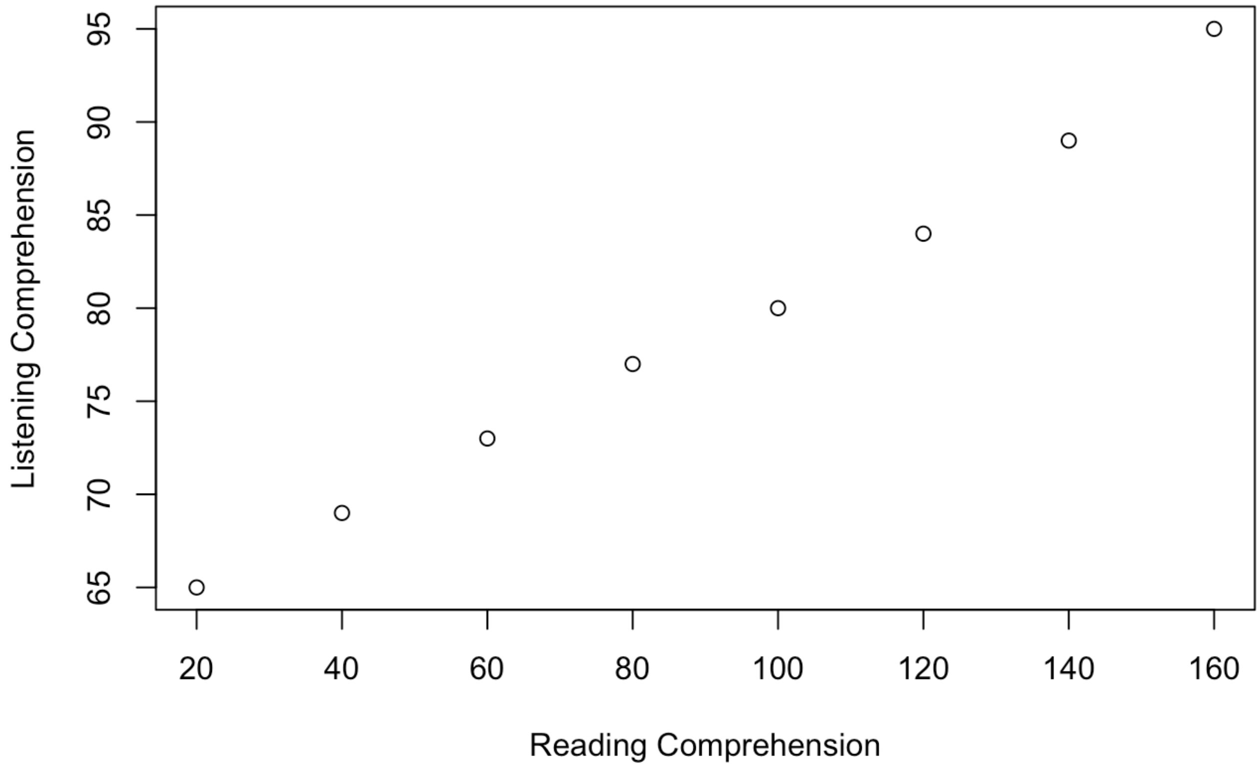
If we wanted to test the relationship between reading and listening comprehension, the H0 would be that there is no significant relationship or difference between reading and listening comprehension.

3

The mean and standard deviation was not asked to be calculated for the reading and listening comprehension variables because it is such a small sample size of 8 students.

4

```
plot(Reading,Listening, xlab="Reading Comprehension",ylab="Listening Comprehensio
n")
```



5

At face value, I think that reading and listening, as plotted in the graph, are related.

6

```
library("pastecs")
stat.desc(Practical4a$Reading, basic=FALSE, norm=TRUE)
```

```
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##   90.0000000   90.0000000   17.3205081   40.9564934 2400.0000000   48.9897949
##     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE   normtest.W
##    0.5443311    0.0000000    0.0000000   -1.6510417   -0.5574527    0.9748583
##   normtest.p
##    0.9331652
```

```
stat.desc(Practical4a$Listening, basic=FALSE, norm=TRUE)
```

```
##        median          mean       SE.mean CI.mean.0.95           var       std.dev
##    78.5000000    79.0000000     3.5807022     8.4670154   102.5714286    10.1277554
##      coef.var      skewness      skew.2SE      kurtosis      kurt.2SE    normtest.W
##     0.1281994     0.1508926     0.1003140    -1.5042864    -0.5079027     0.9835002
##    normtest.p
##     0.9781128
```

```
shapiro.test(Reading)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Reading
## W = 0.97486, p-value = 0.9332
```

```
shapiro.test(Listening)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Listening
## W = 0.9835, p-value = 0.9781
```

For samples less than 30, if the Skew.2SE and Kurt.2SE are in a range between -1 and 1 we can assume normality. For Reading, the Skew.2SE is 0.00 and Kurt.2SE is -0.56. For Listening, the Skew.2SE is 0.10 and Kurt.2SE is -0.51. If the significance value (p-value) for the Shapiro-Wilk test is above 0.05, then we can assume that the data is normally distributed. For Reading the p-value is 0.93 and for Listening the p-value is 0.98. Therefore, we can assume that the data is approximately normally distributed.

7

```
cor.test(Reading,Listening,method="pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  Reading and Listening
## t = 28.126, df = 6, p-value = 1.337e-07
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.9784280 0.9993457
## sample estimates:
##       cor
## 0.9962291
```

The value of rxy is 0.996. This is a strong correlation. The chance of incorrectly rejecting my H0 is 5 percent. I decide to reject the null hypothesis. The effect size is large.

8

A Pearson r correlation analysis showed that Reading Skills and Listening Skills were significantly positively related (r({6}) = {0.996}, p = {p <.001}, two tailed, 95% CI [{0.978},{0.999}]).

Part B: Social class

1

In this study the variables include:

low social class (independent, nominal) high social class (independent, nominal) use of haven't got (dependent, nominal) use of don't have (dependent, nominal)

2

I would formulate H0 as there is no significant relation between social class and the use of 'haven't got' versus 'don't have', whereas I would formulate Ha as there is a significant relation between social class and the use of 'haven't got' verus 'don't have'.

3

The statistic test that could be used is the Shapiro-Wilk normality test.

4

```
Table <- cbind(c(70,64),c(59,31))
rownames(Table) <- c("Haven't Got","Don't Have")
colnames(Table) <- c("Low Social Class","High Social Class")
```

5

```
library("gmodels")
CrossTable(Table, chisq=TRUE, expected=TRUE)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |              Expected N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  224
##
##
##              |
##              |  Low Social Class | High Social Class |         Row Total |
## -------------|-------------------|-------------------|-------------------|
##  Haven't Got |                70 |                59 |               129 |
##              |            77.170 |            51.830 |                   |
##              |             0.666 |             0.992 |                   |
##              |             0.543 |             0.457 |             0.576 |
##              |             0.522 |             0.656 |                   |
##              |             0.312 |             0.263 |                   |
## -------------|-------------------|-------------------|-------------------|
##    Don't Have |                64 |                31 |                95 |
##              |            56.830 |            38.170 |                   |
##              |             0.905 |             1.347 |                   |
##              |             0.674 |             0.326 |             0.424 |
##              |             0.478 |             0.344 |                   |
##              |             0.286 |             0.138 |                   |
## -------------|-------------------|-------------------|-------------------|
## Column Total |               134 |                90 |               224 |
##              |             0.598 |             0.402 |                   |
## -------------|-------------------|-------------------|-------------------|
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## ------------------------------------------------------------
## Chi^2 =  3.909115     d.f. =  1     p =  0.04802487
##
## Pearson's Chi-squared test with Yates' continuity correction
## ------------------------------------------------------------
```

```
## Chi^2 =  3.382896     d.f. =  1     p =  0.06587623
##
##
```

The assumptions concerning the expected frequencies have been met.

6

Since the p-value is less than 0.05 and the Chi-square value is larger than the critical value we can reject the null hypothesis.

7

```
library("vcd")
```

```
## Loading required package: grid
```

```
assocstats(Table)
```

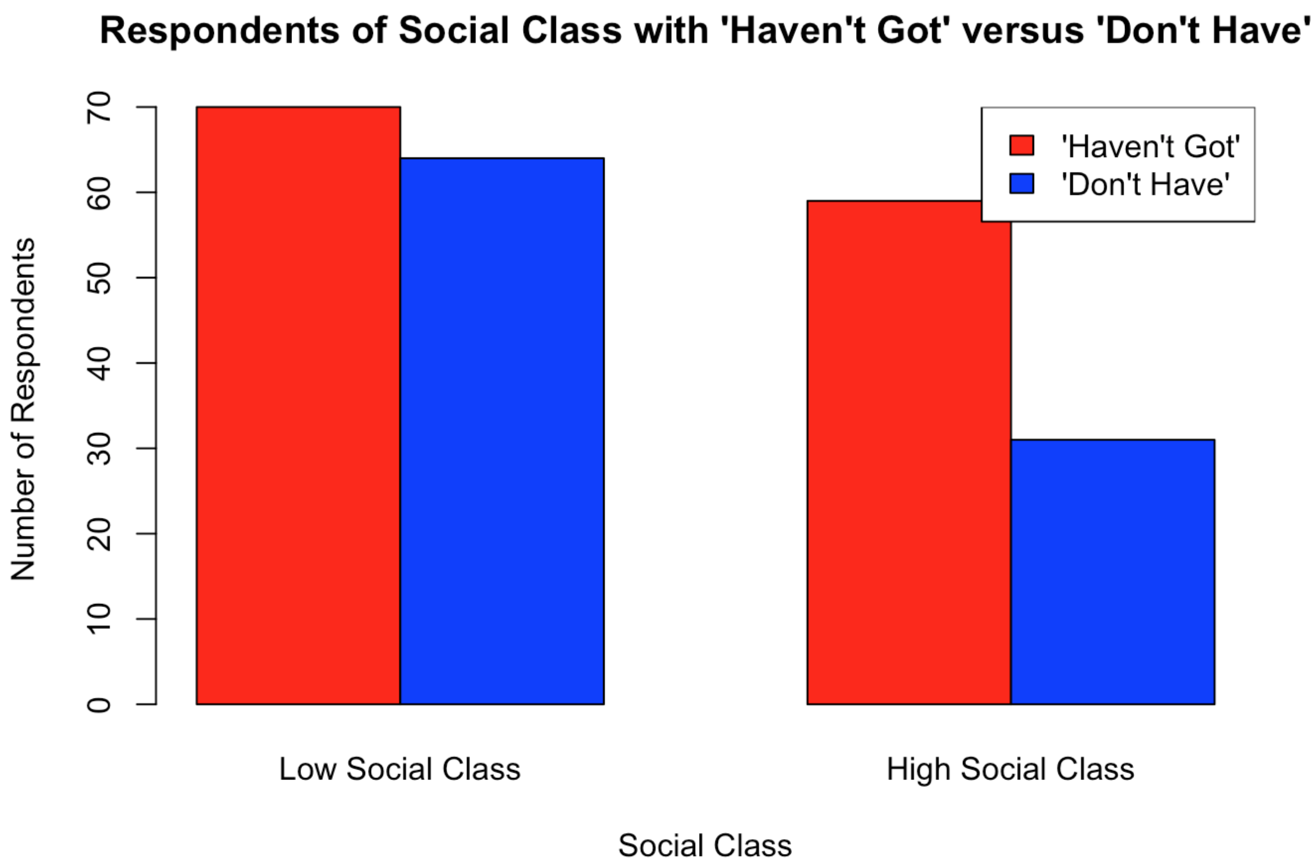
```
##                     X^2 df P(> X^2)
## Likelihood Ratio 3.9454  1 0.046999
## Pearson          3.9091  1 0.048025
##
## Phi-Coefficient   : 0.132
## Contingency Coeff.: 0.131
## Cramer's V        : 0.132
```

With the Phi-Coefficient and Cramer’s V values at 0.132, the effect size is small.

8

A chi-square analysis revealed that the association between social class and the use of ‘haven’t got’ versus ‘don’t have’ was {significant}, χ2({1}, N={8}) = {3.9454}, p = {0.046999}. The direction of the association is that the high social class has a strong relation with using “haven’t got” over “don’t have”. The \(\phi\) coefficient value is 0.132.

```
barplot(Table, beside = TRUE, col = c("red", "blue"), main = "Respondents of Socia
l Class with 'Haven't Got' versus 'Don't Have'",
xlab = "Social Class", ylab="Number of Respondents")
legend("topright", fill = c("red", "blue"), c("'Haven't Got'", "'Don't Have'"))
```



Looking at barplot in the help menu another way to add a legend to the barplot is by adding the legend.text and args.legend in the parameters of the barplot function.

9

If the beside=TRUE is removed or changed to beside=FALSE from the code, then the columns of height are portrayed as stacked bars rather than the juxtaposed bars.

Part C: Gender and intelligence

1

The independent variable is gender (nominal measure) and the dependent variable is intelligence (interval/scale measure).

2

The independent variable has 2 levels.

3

The H0 is that there is no significant relation between gender and intelligence and the Ha is that there is a significant relation between gender and intelligence.

4

The statistical tests that could be used are the independent samples (Shapiro-Wilk) t-test or paired samples t-test.

5

```
library("psych")
Gender <- c("girl","girl","girl","girl","girl","girl","girl","girl","girl","gir
l","girl","boy","boy","boy","boy","boy","boy","boy","boy")
IntellectScore <- c(17,16,14,19,18,17,16,15,16,15,19,16,15,13,19,15,14,13,12)
Practical4b = data.frame(Gender, IntellectScore)
Practical4b$Gender <- as.factor(Practical4b$Gender)
```

6

```
Female <- subset(Practical4b, Gender == "girl")
Male <- subset(Practical4b, Gender == "boy")
aggregate(IntellectScore~Gender, Practical4b, max)
```

```
##   Gender IntellectScore
## 1    boy             19
## 2   girl             19
```

```
aggregate(IntellectScore~Gender, Practical4b, min)
```

```
##   Gender IntellectScore
## 1    boy             12
## 2   girl             14
```

```
aggregate(IntellectScore~Gender, Practical4b, mean)
```

```
##   Gender IntellectScore
## 1    boy       14.62500
## 2   girl       16.54545
```

```
aggregate(IntellectScore~Gender, Practical4b, sd)
```

```
##   Gender IntellectScore
## 1    boy       2.199838
## 2   girl       1.634848
```

```
describe(Practical4b$IntellectScore)
```

```
##    vars  n  mean   sd median trimmed  mad min max range skew kurtosis   se
## X1    1 19 15.74 2.08     16   15.76 1.48  12  19     7 0.09    -1.03 0.48
```

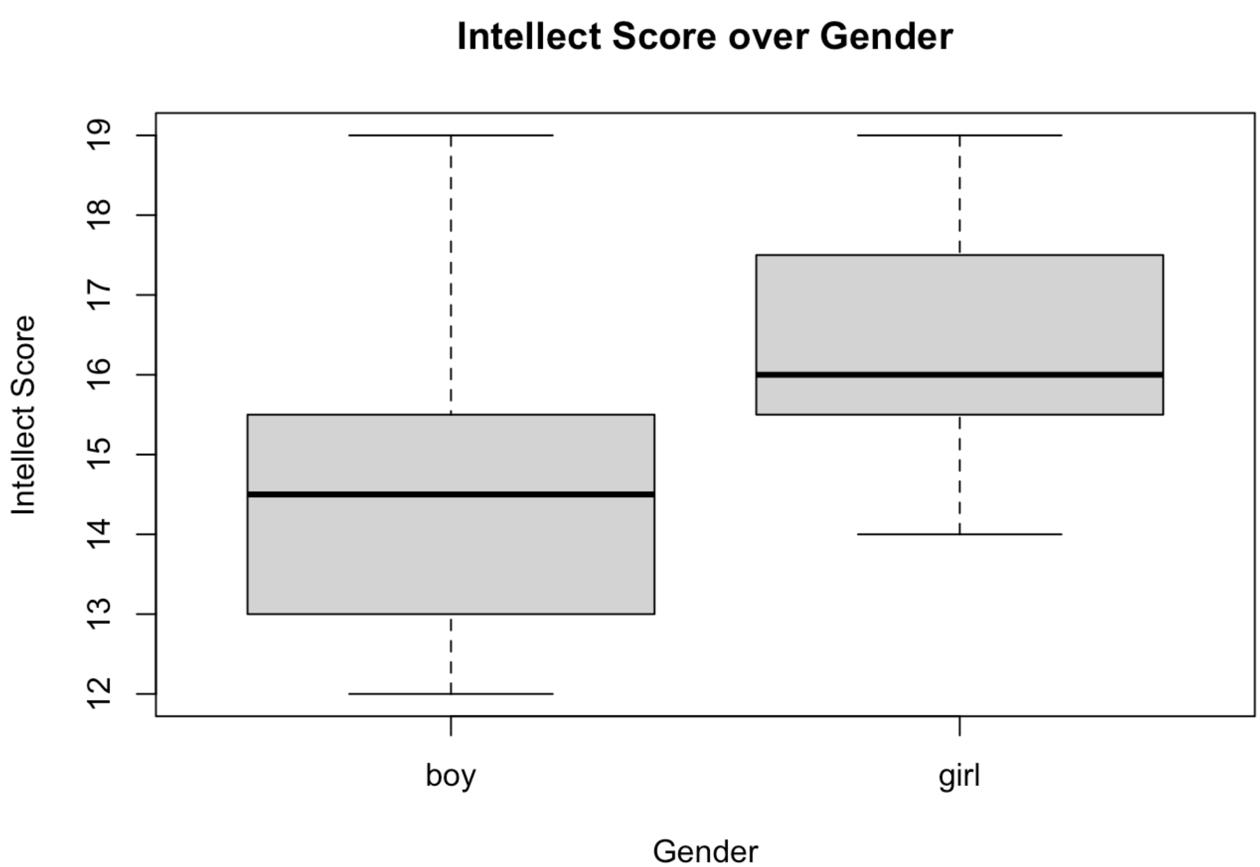
value	Overall	Female	Male
mean	15.74	16.55	14.63
max	19	19	19
min	12	14	12
sd	2.08	1.63	2.20

7

My first impressions about the difference between the boys and girls are that the girls have higher intelligence scores.

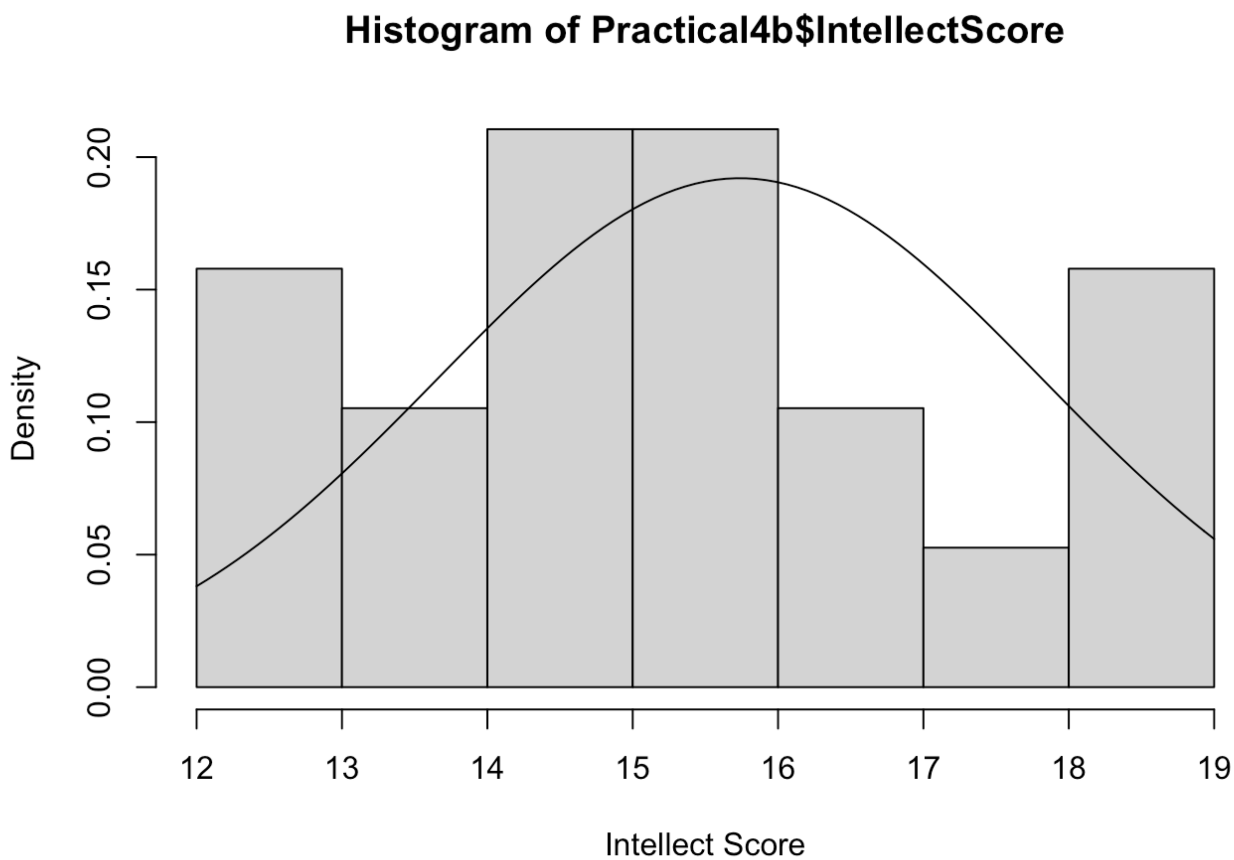
8

```
boxplot(IntellectScore~Gender, data=Practical4b, main="Intellect Score over Gende
r", xlab="Gender", ylab="Intellect Score")
```



9

```
hist(Practical4b$IntellectScore, prob=TRUE, xlab="Intellect Score")
curve(dnorm(x, mean=mean(Practical4b$IntellectScore),sd=sd(Practical4b$IntellectSc
ore)), add=TRUE)
```




```
stat.desc(Practical4b$IntellectScore, basic=FALSE, norm=TRUE)
```

```
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##  16.00000000  15.73684211   0.47659922   1.00129780   4.31578947   2.07744783
##     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE   normtest.W
##   0.13201173   0.08615109   0.08224186  -1.02744047  -0.50649266   0.95088138
##   normtest.p
##   0.40902520
```

```
by(Practical4b$IntellectScore, Practical4b$Gender, stat.desc,
basic=FALSE, norm=TRUE)
```

```
## Practical4b$Gender: boy
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##   14.5000000   14.6250000    0.7777601    1.8391103    4.8392857    2.1998377
##     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE   normtest.W
##    0.1504163    0.6990096    0.4647043   -0.6965155   -0.2351694    0.9227958
##   normtest.p
##    0.4529579
## ------------------------------------------------------------
## Practical4b$Gender: girl
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##  16.00000000  16.54545455   0.49292515   1.09830568   2.67272727   1.63484778
##     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE   normtest.W
##   0.09880948   0.17641574   0.13350922  -1.33523222  -0.52181325   0.94182426
##   normtest.p
##   0.54217663
```

```
library('car')
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:psych':
##
##     logit
```

```
leveneTest(IntellectScore~Gender, data=Practical4b)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  1    0.39 0.5406
##       17
```

```
test <- t.test(Practical4b$IntellectScore ~ Practical4b$Gender, var.equal = FALSE)
test
```

```
##
##  Welch Two Sample t-test
##
## data:  Practical4b$IntellectScore by Practical4b$Gender
## t = -2.0856, df = 12.357, p-value = 0.05838
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.92030824  0.07939915
## sample estimates:
##  mean in group boy mean in group girl
##           14.62500           16.54545
```

The value of t is -2.0856. The degree of freedom is 12.357. The level of significance or p-value is 0.05838. The 95 percent confidence interval is -3.92030824 and 0.07939915. Since the level of significance is larger than 0.05, H0 cannot be rejected. As the two outer ends of the spectrum are both positive, then the chance of the difference being positive is relatively large.

10

A paired sample t-test compares means from the same group at different times and a one sample t-test tests the mean of a single group against a known mean, so using an independent samples t-test that compares the means for two groups makes the most appropriate sense in the context of the data provided.

11

```
t <- test$statistic[[1]]
df <- test$parameter[[1]]
r2 <- t^2/(t^2+df)
r2
```

```
## [1] 0.2603588
```

```
library('effsize')
```

```
##
## Attaching package: 'effsize'
```

```
## The following object is masked from 'package:psych':
##
##     cohen.d
```

```
cohen.d(Practical4b$IntellectScore,Practical4b$Gender)
```

```
##
## Cohen's d
##
## d estimate: -1.017148 (large)
## 95 percent confidence interval:
##       lower       upper
## -2.05747110  0.02317563
```

We found a large effect size. What this would mean for the number of participants I need to get enough power is small, although there is a fair chance that it is mistaken since there is only 19 participants in total as big effects can be detected with small sample sizes whereas to detect smaller effects larger samples are needed.

12

An independent samples t-test revealed that, on average, the {girls} showed a higher level of intelligence (M={16.5455}, SD= {1.6348}) than the {boys} (M={14.6250}, SD= {2.1998}). This difference was {not significant}, t({12.357})= {-2.0856}, p = {0.05838)}, 95% CI [{ -3.92030824}, {0.07939915}]. This effect was of a {large} size, r2/d = {0.2603588}.

13

The meaningfulness of this outcome is that we cannot yet assume that there is a significant relation between gender and intellegince.

14

Additional information I would like to have about this study include: How well each gender scored on each question on the test.

Part D: Alpha - This assignment is slightly more

1

```
Practical3a <- read.csv(file="/Users/trekkatkins/Documents/R Practicals/Data-Pract
ical3a.csv",sep=";")
str(Practical3a)
```

```
## 'data.frame':    130 obs. of  22 variables:
##  $ Student: int  1 2 3 4 5 6 7 8 9 10 ...
##  $ teacher: chr  "A" "A" "A" "A" ...
##  $ group  : chr  "1A" "1A" "1A" "1A" ...
##  $ Q1     : int  10 12 10 18 20 16 10 7 20 11 ...
##  $ Q2     : int  5 5 4 5 5 5 3 4 4 5 ...
##  $ Q3     : int  5 4 5 6 6 6 4 6 6 5 ...
##  $ Q4     : int  7 8 6 8 7 8 8 6 6 7 ...
##  $ Q5     : int  4 4 2 5 5 6 5 3 6 6 ...
##  $ Q6     : int  2 4 3 3 4 3 3 2 3 3 ...
##  $ Q7     : int  4 5 0 4 4 1 3 0 4 2 ...
##  $ Q8     : int  14 18 8 15 19 19 16 14 17 17 ...
##  $ Q9     : int  5 5 0 5 5 0 0 0 5 5 ...
##  $ Q10    : int  5 0 5 5 5 0 5 0 5 5 ...
##  $ Q11    : int  5 5 5 5 5 5 5 5 5 5 ...
##  $ Q12    : int  0 0 0 0 0 0 5 0 0 0 ...
##  $ Q13    : int  5 0 5 5 5 5 5 0 5 0 ...
##  $ Q14    : int  0 5 0 0 0 0 0 0 0 0 ...
##  $ Q15    : int  0 0 0 5 0 0 0 0 5 0 ...
##  $ Q16    : num  4 4 4 4 5 4 2 5 4 4 ...
##  $ Q17    : int  12 17 11 12 13 11 11 2 8 12 ...
##  $ TOTAL  : num  87 96 68 105 108 89 85 54 103 87 ...
##  $ Grade  : int  6 7 5 8 8 6 6 4 7 6 ...
```

2

```
Questions <- Practical3a[, c(4:20)]
psych::alpha(Questions)
```

```
##
## Reliability analysis
## Call: psych::alpha(x = Questions)
##
##   raw_alpha std.alpha G6(smc) average_r S/N   ase mean  sd median_r
##        0.8      0.83    0.85      0.22 4.9 0.021  4.3 1.4     0.19
##
##  lower alpha upper     95% confidence boundaries
## 0.75 0.8 0.84
##
##  Reliability if an item is dropped:
##     raw_alpha std.alpha G6(smc) average_r S/N alpha se var.r med.r
## Q1       0.77      0.81    0.83      0.21 4.2    0.026 0.026  0.18
## Q2       0.79      0.82    0.84      0.22 4.5    0.022 0.028  0.18
## Q3       0.78      0.81    0.83      0.21 4.2    0.023 0.025  0.18
## Q4       0.78      0.82    0.84      0.22 4.4    0.023 0.027  0.18
## Q5       0.78      0.81    0.83      0.21 4.3    0.023 0.025  0.18
## Q6       0.78      0.81    0.84      0.21 4.4    0.022 0.027  0.18
## Q7       0.78      0.82    0.84      0.22 4.5    0.022 0.030  0.18
## Q8       0.78      0.82    0.84      0.22 4.6    0.022 0.029  0.18
## Q9       0.79      0.83    0.85      0.24 5.0    0.021 0.029  0.21
## Q10      0.79      0.83    0.85      0.23 4.8    0.021 0.030  0.18
## Q11      0.80      0.84    0.85      0.24 5.1    0.021 0.028  0.22
## Q12      0.80      0.84    0.86      0.24 5.1    0.021 0.028  0.22
## Q13      0.79      0.83    0.85      0.23 4.7    0.022 0.030  0.19
## Q14      0.80      0.84    0.86      0.25 5.4    0.020 0.024  0.22
## Q15      0.80      0.84    0.85      0.24 5.1    0.021 0.029  0.22
## Q16      0.78      0.81    0.83      0.21 4.3    0.022 0.027  0.18
## Q17      0.76      0.81    0.83      0.21 4.3    0.026 0.027  0.18
##
##  Item statistics
##       n raw.r std.r r.cor r.drop mean  sd
## Q1  130  0.81  0.73  0.73  0.670 10.6 6.4
## Q2  130  0.54  0.60  0.58  0.505  4.0 1.1
## Q3  130  0.68  0.73  0.73  0.640  4.3 1.8
## Q4  130  0.62  0.63  0.62  0.560  6.6 2.1
## Q5  130  0.65  0.70  0.71  0.607  3.9 1.8
## Q6  130  0.62  0.66  0.64  0.585  2.1 1.2
## Q7  130  0.54  0.58  0.55  0.482  2.1 1.6
## Q8  130  0.60  0.56  0.52  0.470 13.2 4.1
## Q9  130  0.35  0.33  0.26  0.253  2.8 2.5
## Q10 130  0.44  0.44  0.37  0.346  2.6 2.5
## Q11 130  0.27  0.30  0.21  0.172  3.6 2.3
## Q12 130  0.29  0.28  0.19  0.194  1.3 2.2
## Q13 130  0.48  0.47  0.41  0.392  2.3 2.5
## Q14 130  0.10  0.14  0.03  0.015  1.0 2.0
## Q15 130  0.28  0.31  0.23  0.196  1.1 2.1
```

```
## Q16 130  0.64  0.68  0.67  0.600  3.5 1.3
## Q17 130  0.77  0.71  0.70  0.665  8.3 4.7
```

3

A general accepted rule is that a Cronbach alpha that is 0.6 to 0.8 is considered an acceptable level of reliability.

4

Removing any of the items would not substantially improve the reliability of the test since none of the values are greater than the overall alpha value.

5

The items that should be removed because of problematic r.drop correlations include: Q9 with a r.drop value below 0.3 with a value of 0.253, Q11 with a value of 0.172, Q12 with a value of 0.194, Q14 with a value of 0.015, and Q15 with a value of 0.196.