Practical Exam

Trevor Atkins

10/26/2020

```
install.packages(c("rmarkdown", "psych", "pastecs", "car", "effsize", "gmodels", "
vcd", "pwr", "sjstats", "ggplot2",
"visreg", "DescTools", "rcompanion", "caTools", "bitops"))
```

```
##
##   There is a binary version available but the source version is later:
##           binary source needs_compilation
## rmarkdown    2.3    2.5             FALSE
##
##
## The downloaded binary packages are in
##  /var/folders/_6/4dcygrqd30z3yrk1v612llt40000gn/T//RtmpwIRX9P/downloaded_packag
es
```

```
## installing the source package 'rmarkdown'
```

a: Learning Environment and Vocab:

1

The independent variable is the study environment that is on a nominal scale and has two levels - “library” and “cafe”. The dependent variable is the number of vocabulary learned in an hour, which is a interval scale.

2

The appropriate perspective is comparing the means of the two different groups. The appropriate statisitcal test is the t-test.

3

The null hypothesis is that the study environment does not affect the amount of vocabulary learned.

The alternative hypothesis is that the study environment does affect the amount of vocabulary learned.

4

```
dataQA <- read.csv2("StatsTest20-LocationLearning(1).csv")
dataQA$Environment <- as.factor(dataQA$Environment)
str(dataQA)
```

```
## 'data.frame':    100 obs. of  2 variables:
##  $ Environment: Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Score      : int  2 7 3 4 8 9 6 5 7 2 ...
```

Calculating the descriptives:

```
by(dataQA$Score, dataQA$Environment, mean)
```

```
## dataQA$Environment: 1
## [1] 7.173913
## ------------------------------------------------------------
## dataQA$Environment: 2
## [1] 9.166667
```

```
by(dataQA$Score, dataQA$Environment, min)
```

```
## dataQA$Environment: 1
## [1] 1
## ------------------------------------------------------------
## dataQA$Environment: 2
## [1] 3
```

```
by(dataQA$Score, dataQA$Environment, max)
```

```
## dataQA$Environment: 1
## [1] 15
## ------------------------------------------------------------
## dataQA$Environment: 2
## [1] 18
```

```
by(dataQA$Score, dataQA$Environment, sd)
```

```
## dataQA$Environment: 1
## [1] 3.573385
## ------------------------------------------------------------
## dataQA$Environment: 2
## [1] 3.254895
```

Table 1 : Descriptives for the two study environment groups

value	Library	Cafe

value	Library	Cafe
mean	7.17	9.17
minimum	1	3
maximum	15	18
sd	3.57	3.25

Since means are being compared, normality and homogeneity of varience will have to be checked. Since the sample size is larger than 30, it is better to combine skewness and kurtosis (the 2SE values) with a histogram evaluation.

```
library("pastecs")
by(dataQA$Score, dataQA$Environment, stat.desc, basic=FALSE, norm=TRUE)
```

```
## dataQA$Environment: 1
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##    7.0000000    7.1739130    0.5268669    1.0611644   12.7690821    3.5733852
##     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE   normtest.W
##    0.4981082    0.2867952    0.4095947   -0.9978168   -0.7255502    0.9607654
##   normtest.p
##    0.1227804
## ------------------------------------------------------------
## dataQA$Environment: 2
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##   9.00000000   9.16666667   0.44293507   0.88841524  10.59433962   3.25489472
##     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE   normtest.W
##   0.35507942   0.38236259   0.58905440  -0.05084620  -0.03979242   0.96945651
##   normtest.p
##   0.18242123
```

The values of Skew.2SE and Kurt.2SE are between -1 and 1 for both groups, so the data can be assumed to be normally distributed but the histogram also needs to be evaluated.

```
hist(dataQA$Score[dataQA$Environment =="1"], prob=TRUE, xlab = "difference", main
= "histogram of difference for library study environment")
curve(dnorm(x, mean=mean(dataQA$Score[dataQA$Environment =="1"]), sd=sd(dataQA$Sco
re[dataQA$Environment =="1"])), add=TRUE)
```

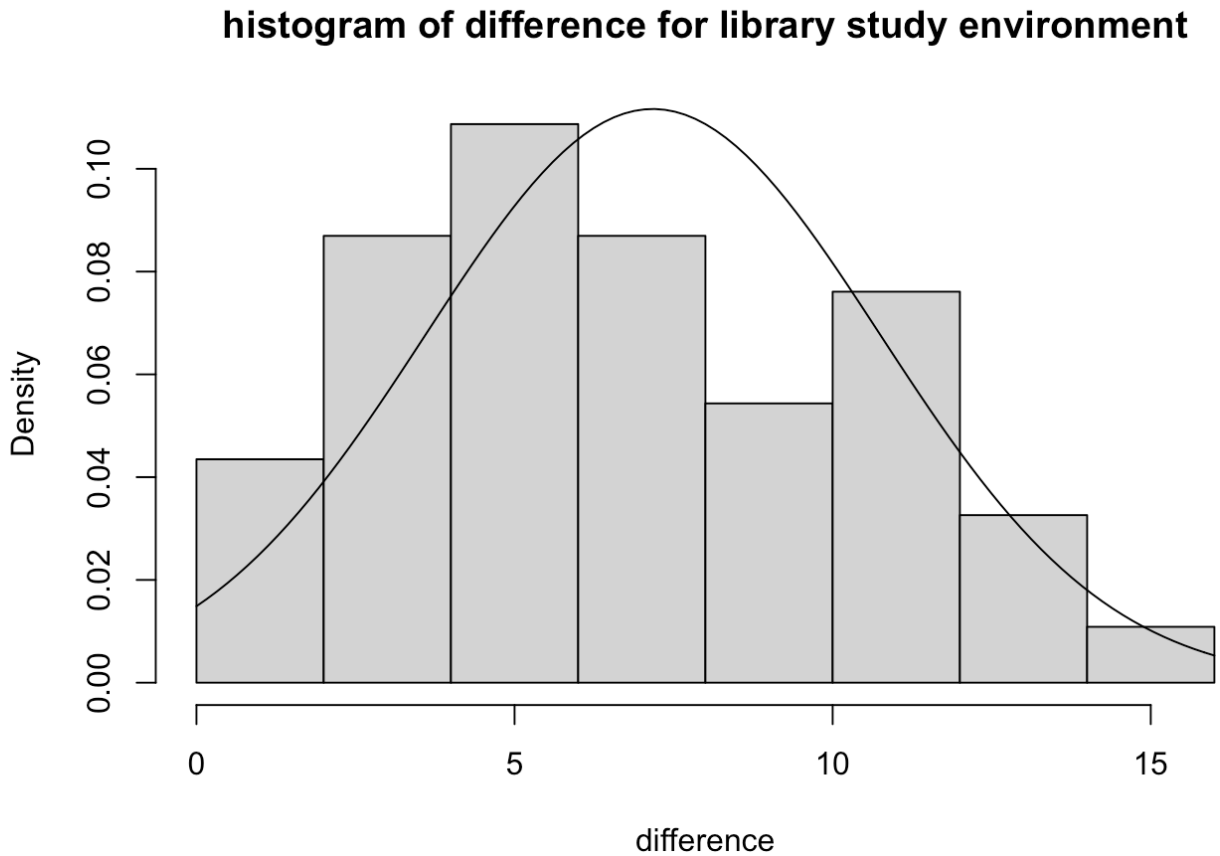

Figure 1: Probabilities of difference in vocabulary for library study environment group

```
hist(dataQA$Score[dataQA$Environment =="2"], prob=TRUE, xlab = "difference", main
= "histogram of difference for cafe study environment")
curve(dnorm(x, mean=mean(dataQA$Score[dataQA$Environment =="2"]), sd=sd(dataQA$Sco
re[dataQA$Environment=="2"])), add=TRUE)
```

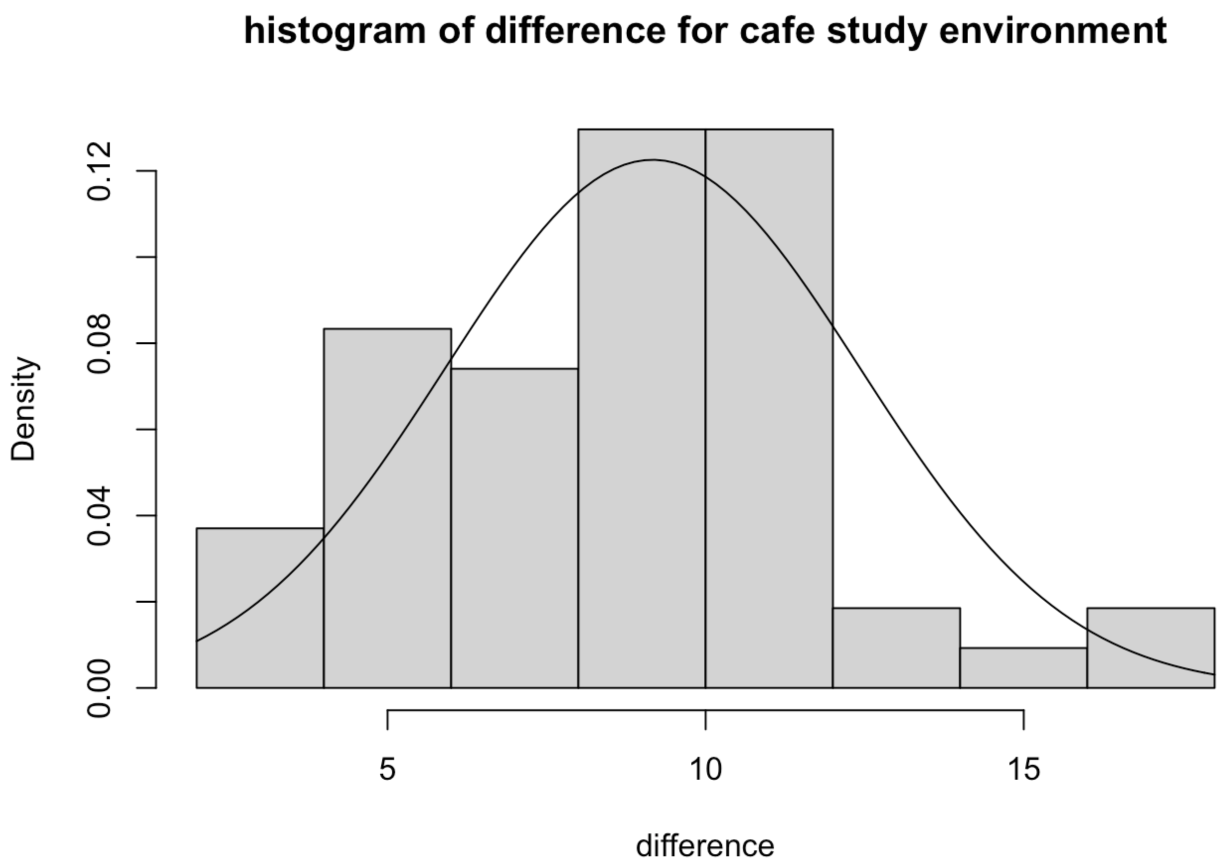

Figure 2 : Probabilities of difference in vocabulary for cafe study environment group

These histograms show that the data does slightly deviate from the normal distribution.

```
library(car)
```

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

```
leveneTest(Score~Environment, data=dataQA)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  1  1.2702 0.2625
##       98
```

The p-value is .2625, so homogeneity of variance can be assumed.

```
library(psych)
```

```
##
## Attaching package: 'psych'
```

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

```
by(dataQA$Score, dataQA$Environment, describe)
```

```
## dataQA$Environment: 1
##    vars  n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1    1 46 7.17 3.57      7    7.08 4.45   1  15    14 0.29       -1 0.53
## ------------------------------------------------------------
## dataQA$Environment: 2
##    vars  n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1    1 54 9.17 3.25      9    9.02 2.97   3  18    15 0.38    -0.05 0.44
```

```
boxplot(dataQA$Score~dataQA$Environment, ylab="difference in vocab", xlab="study e
nvironment", main="number of vocabulary difference by study environment group")
```

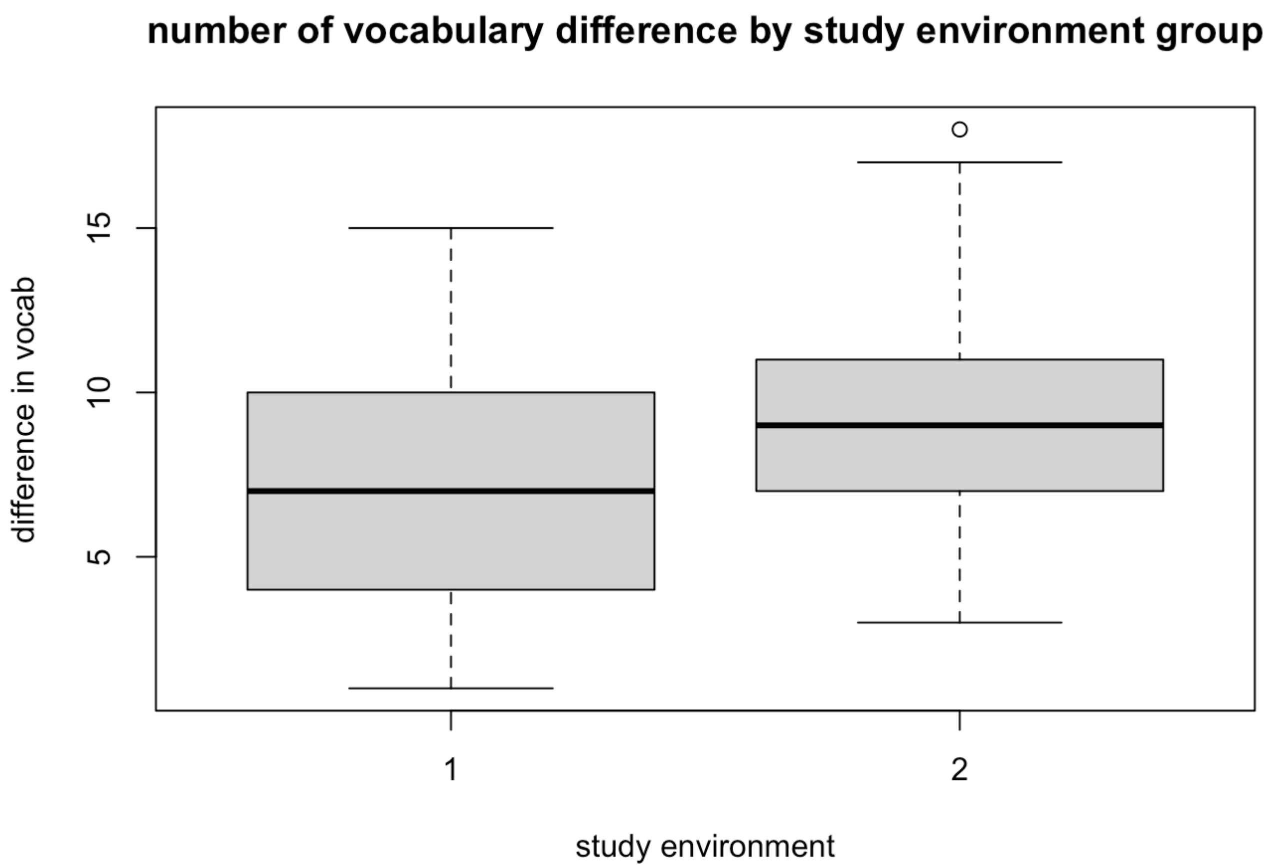

Figure 3: Boxplot showing the dispersion of scores in the library study environment group (left) and the cafe study environment group (right)

```
ttest = t.test(dataQA$Score~dataQA$Environment)
ttest
```

```
##
##  Welch Two Sample t-test
##
## data:  dataQA$Score by dataQA$Environment
## t = -2.8951, df = 92.048, p-value = 0.004733
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.3598010 -0.6257063
## sample estimates:
## mean in group 1 mean in group 2
##        7.173913        9.166667
```

```
t <- ttest$statistic[[1]]
df <- ttest$parameter[[1]]
r2 <- t^2/(t^2+df)
round(r2,3)
```

```
## [1] 0.083
```

Value of statistic: t=-2.89 Significance: p = .005 Decision Null Hypothesis: Since p = .005, we can reject null hypothesis

On average, students in a cafe study environment show greater difference in the vocabulary (M=7.17, SD=3.57) than students in a library study environment (M=9.17, SD=3.25). This difference was significant t(92)=-2.895, p = 0.004733, 95% CI [-3.3598010, -0.6257063], and the effect size was small (r2 = .083).

Meaningfulness

In terms of power we know that for a small effect size you probably need about 780 participants per group for a power of 0.8. In this case, we can say that we don't have enough partipants to reach this power of 0.8. As for the methodology of the study, we would maybe want to know more about the participants and how they were tested. This study assumes that all the students have the same L1 background before learning Polish words. It would thus be interesting to see whether a randomized sample of L1 backgrounds could reduce the chance of that variable being a factor in contributing bias to the samples.

b: Sleep, Lexical Decision Accuracy, and Gender:

1

The independent variables are the amount of sleep that is on a interval scale (with 1 level) and gender that is on a nominal scale with 2 levels (male and female). The dependent variable is the accuracy score of the lexical decision task that is a interval scale.

2

The appropriate perspective is comparing the means of the two different groups and assessing the relationship between sleep and performance. The appropriate statisitcal test is the multiple linear regression since there is both a nominal and an interval independent variable.

3

Hypothesis group 1:

The null hypothesis is that there is no effect of sleep on the lexical task performance.

The alternative hypothesis is that there is an effect of sleep on the lexical task performance.

Hypothesis group 2:

The null hypothesis is that there is no impact of gender on lexical task performance.

The alternative hypothesis is that there is an impact of gender on lexical task performance.

Hypothesis group 3:

There is no interaction between the effects of gender and sleep.

There is an interaction between the effects of gender and sleep.

4

```
dataQB <- read.csv2("StatsTest20_SleepAccuracy_BS(1).csv")
dataQB$player <- as.factor(dataQB$player)
dataQB$gender <- as.factor(dataQB$gender)
dataQB$sleep <- as.numeric(dataQB$sleep)
str(dataQB)
```

```
## 'data.frame':	56 obs. of  4 variables:
##  $ player: Factor w/ 56 levels "1","2","3","4",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ gender: Factor w/ 2 levels "f","m": 2 2 2 2 2 2 2 2 2 2 ...
##  $ sleep : num  1.7 2.5 2.9 3.1 3.5 3.59 5.06 4.34 4.45 2.2 ...
##  $ score : int  6 8 7 11 6 8 7 11 8 11 ...
```

```
by(dataQB$score, dataQB$gender, describe)
```

```
## dataQB$gender: f
##    vars  n  mean   sd median trimmed  mad min max range  skew kurtosis   se
## X1    1 28 11.96 1.97     12   11.96 2.22   9  15     6 -0.12    -1.23 0.37
## ------------------------------------------------------------
## dataQB$gender: m
##    vars  n  mean  sd median trimmed  mad min max range skew kurtosis   se
## X1    1 28 11.25 3.4     11   11.08 3.71   6  19    13 0.42    -0.51 0.64
```

```
library(ggplot2)
```

```
##
## Attaching package: 'ggplot2'
```

```
## The following objects are masked from 'package:psych':
##
##     %+%, alpha
```

```
qplot(sleep, score, colour = gender, data = dataQB)
```

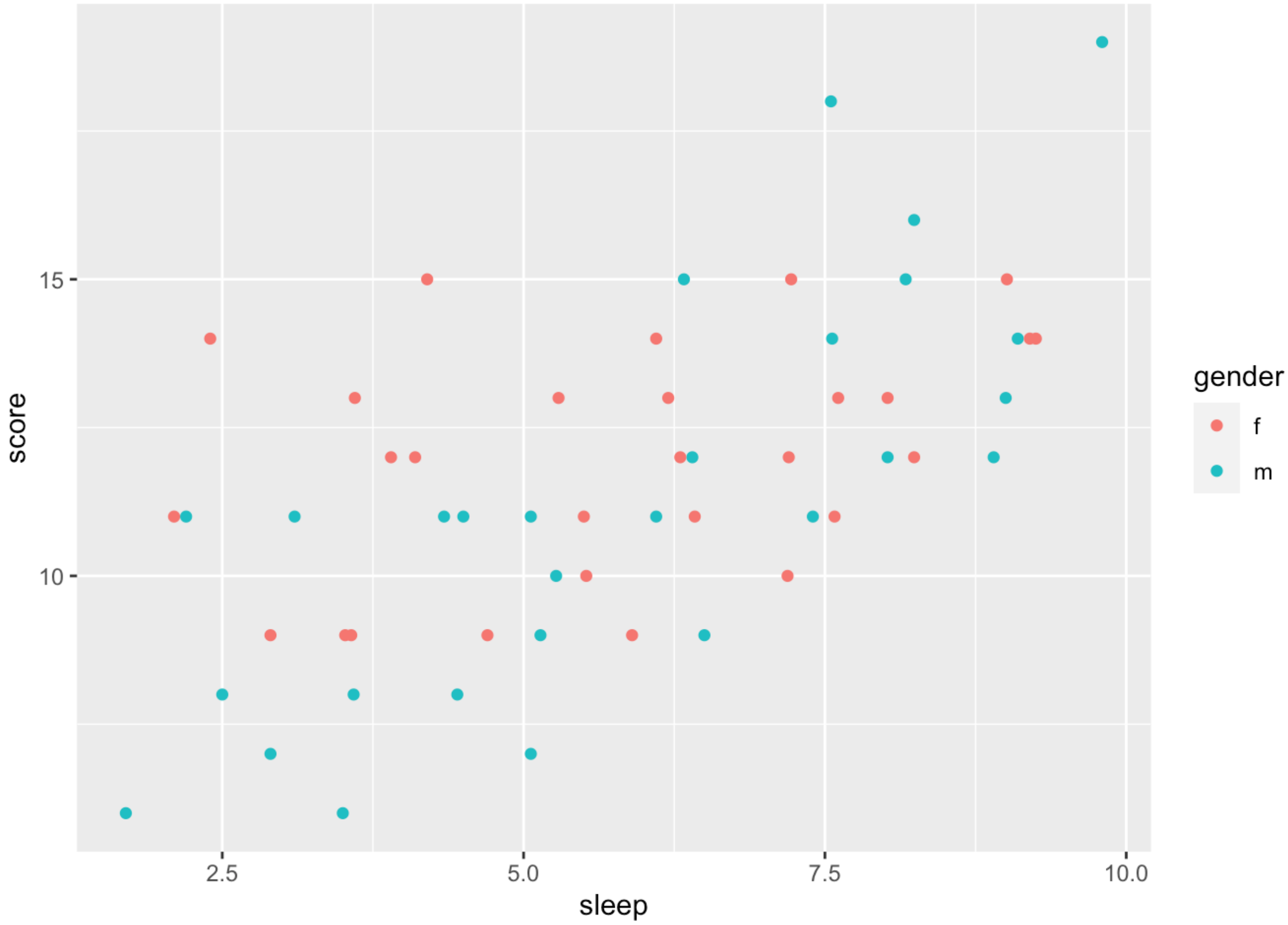

Figure 4: Scatterplot showing the relationship between sleep (x-axis) and score (y-axis) with separated colours for people who are male (blue) and those who are female (orange).

The pattern in Figure 4 reveals that there seems to be a linear positive relationship between sleep and score. The people who are female additionally seem to score higher overall than those who are male.

```
m1 = lm(score~sleep + gender, data = dataQB)
summary(m1)
```

```
##
## Call:
## lm(formula = score ~ sleep + gender, data = dataQB)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.674 -1.563 -0.624  1.463  5.386
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.4348     0.8879   8.373 2.85e-11 ***
## sleep         0.7793     0.1351   5.771 4.19e-07 ***
## genderm      -0.7043     0.5869  -1.200    0.236
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.196 on 53 degrees of freedom
## Multiple R-squared:  0.3962, Adjusted R-squared:  0.3734
## F-statistic: 17.39 on 2 and 53 DF,  p-value: 1.56e-06
```

Negative effect of gender and positive effect of sleep, but is there also an interaction?

```
m2 = lm(score~sleep*gender, data = dataQB)
summary(m2)
```

```
##
## Call:
## lm(formula = score ~ sleep * gender, data = dataQB)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4305 -1.7224 -0.0294  1.1995  4.8092
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     9.8363     1.1695   8.411 2.87e-11 ***
## sleep           0.3661     0.1898   1.929  0.05919 .
## genderm        -5.0151     1.5760  -3.182  0.00247 **
## sleep:genderm   0.7424     0.2544   2.918  0.00519 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.055 on 52 degrees of freedom
## Multiple R-squared:  0.4812, Adjusted R-squared:  0.4513
## F-statistic: 16.08 on 3 and 52 DF,  p-value: 1.607e-07
```

There is no significant interaction between sleep and gender. To be sure, we can check this in a visualization using the visreg package.

```
library("visreg")
visreg(m2,xvar="sleep", by="gender")
```

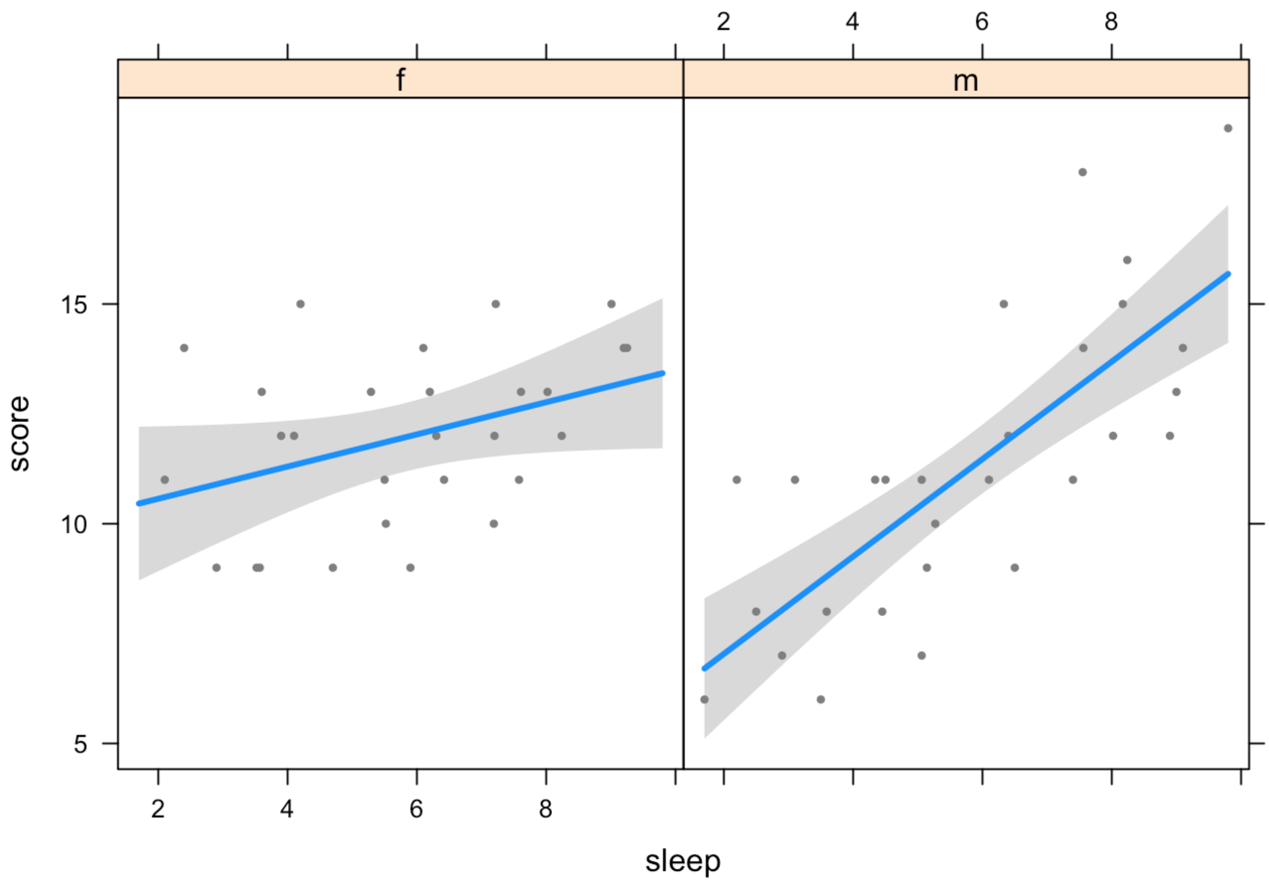

Figure 5: Plot of the slopes of the effect of sleep for people who are female (left) and those who are male (right).

The effects look very much the same, confirming the absence of an interaction.

```
anova(m1, m2)
```

```
## Analysis of Variance Table
##
## Model 1: score ~ sleep + gender
## Model 2: score ~ sleep * gender
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     53 255.61
## 2     52 219.64  1    35.972 8.5164 0.005189 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The ANOVA reveals that model 2 (with the interaction) is not better at explaining the variance than model 1. So, we'll stick with the simpler model 1 with the main effects only.

We do still need to check our assumptions:

linearity; heteroskedasticity; normally distributed residuals; collinearity.

```
plot(fitted(m1), residuals(m1))
```

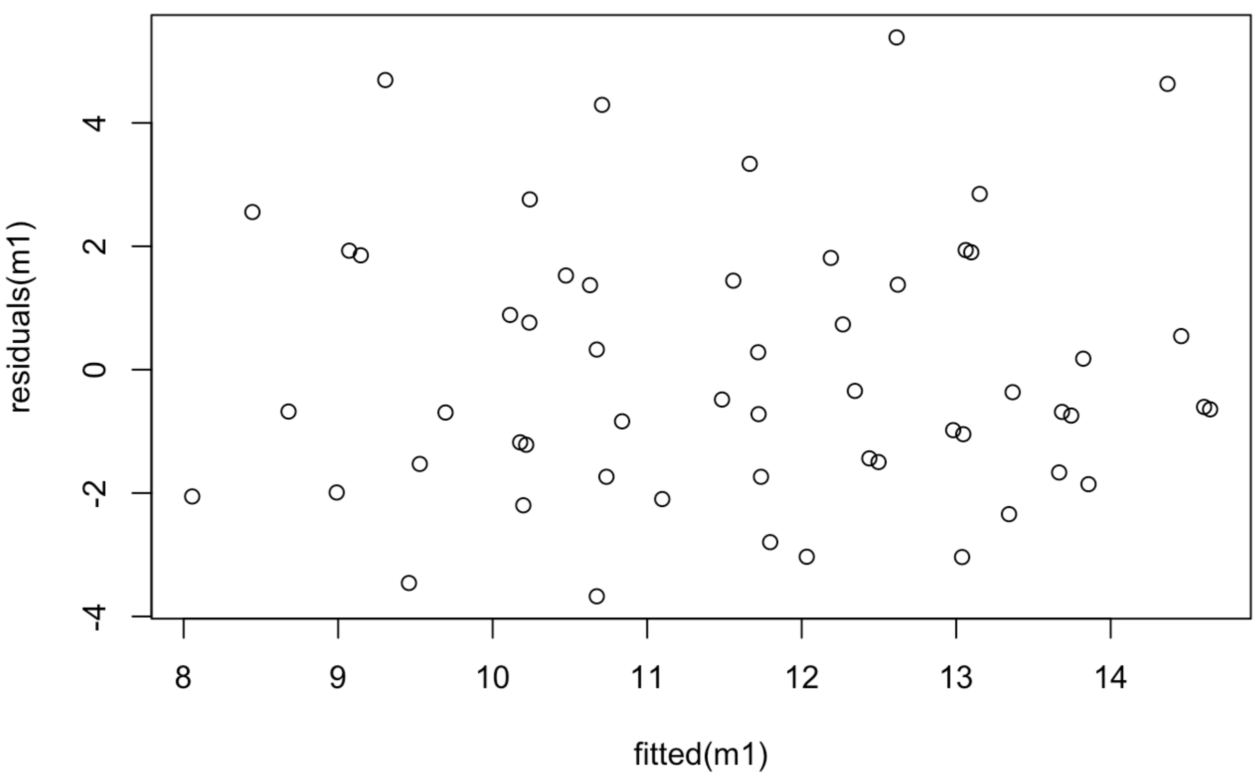

Figure 6: Scatterplot showing the residuals and their deviations from the fitted values.

The relationship is linear (also see Figure 4) and the residuals plot in Figure 6 does not reveal any signs of heteroskedasticity. To be sure about the second assumption, we can do the non Constant Variance error test.

```
ncvTest(m1)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 0.2167011, Df = 1, p = 0.64157
```

The ncvTest is non-significant, so we can assume homoscedasticity.

Remember that the residuals also have to be normally distributed, so we'll have to check that as well.

```
qqnorm(residuals(m1))
```

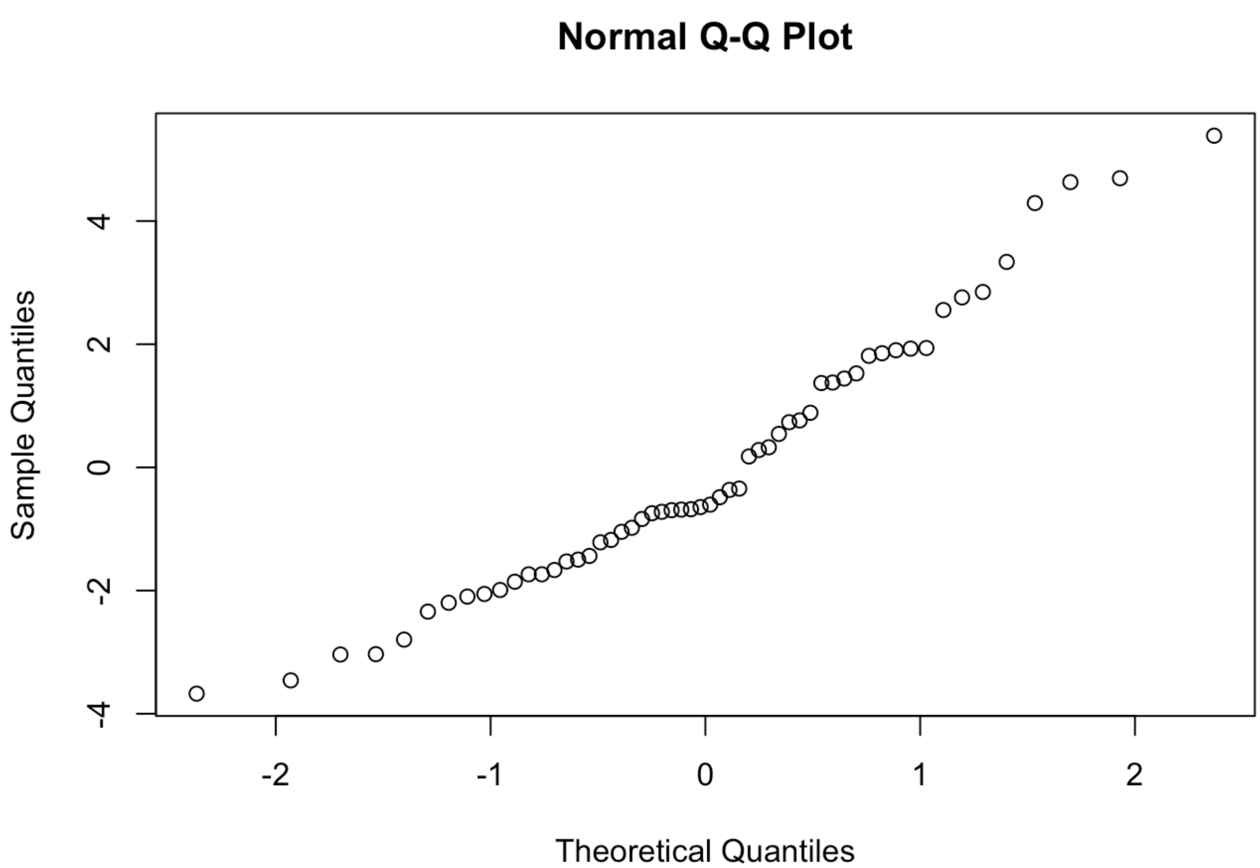

Figure 7: qqplot of the residuals.

The qqplot indicates that the data follow a straight line, so it approximates the normal distribution. Note that this could also have been checked by plotting a histogram of the residuals or even by performing a Shapiro-Wilk on the residuals.

```
shapiro.test(residuals(m1))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(m1)
## W = 0.95927, p-value = 0.05611
```

The Shapiro-Wilk confirms that the residuals of our model are normally distributed.

Based on common sense, we would not expect any correlations between sleep and accuracy on the lexical decision test, but we can check using the VIF.

```
car::vif(m1)
```

```
##    sleep   gender
## 1.000009 1.000009
```

As the values are well below 5, we can assume that there is no collinearity issue.

```
summary(m1)
```

```
##
## Call:
## lm(formula = score ~ sleep + gender, data = dataQB)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.674 -1.563 -0.624  1.463  5.386
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.4348     0.8879   8.373 2.85e-11 ***
## sleep         0.7793     0.1351   5.771 4.19e-07 ***
## genderm      -0.7043     0.5869  -1.200    0.236
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.196 on 53 degrees of freedom
## Multiple R-squared:  0.3962, Adjusted R-squared:  0.3734
## F-statistic: 17.39 on 2 and 53 DF,  p-value: 1.56e-06
```

We constructed a linear model of proficiency score as a function of sleep and gender. This model was significant ($F(2,53)=17.39$, $p< .001$) and explained 37.34% of the variance in the data (adjusted R-squared). Regression coefficients are shown in Table 2. The negative coefficient for age of acquisition reveals that, as sleep increases the accuracy increases significantly. The negative coefficient for gender indicates a significantly lower proficiency score for those who are male as compared to those who are female. This pattern can also be seen in Figure 8 below.

Table 2: Regression coefficients for the linear model of score as a function of sleep and gender.

-	Estimate	SE	t-value	p-value
Intercept	7.4348	0.8879	8.373	2.85e-11
Sleep	0.7793	0.1351	5.771	4.19e-07
Gender=male	-0.7043	0.5869	-1.200	0.236

```
boxplot(dataQB$score~dataQB$gender, ylab="Accuracy of Lexical Decision Task", main
="Effect of gender on scores")
```

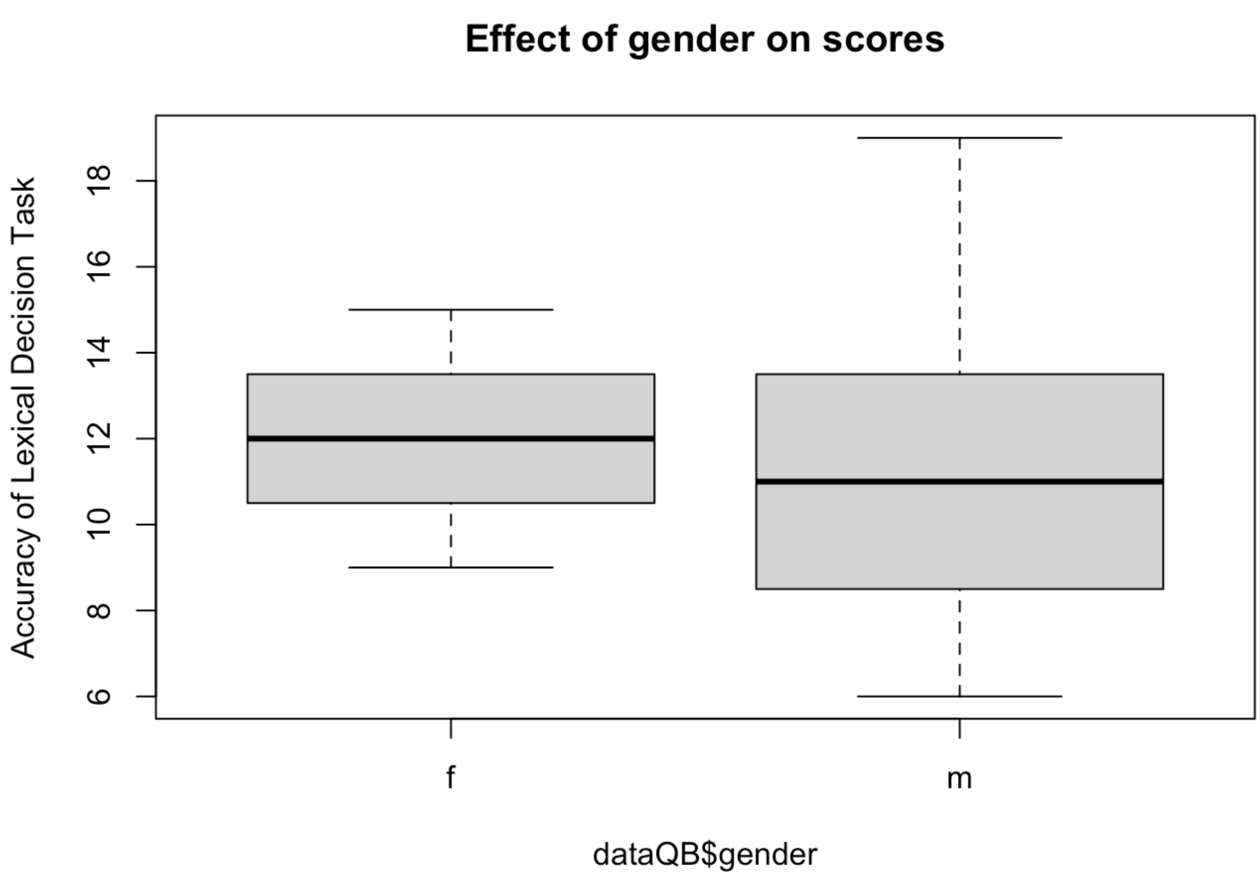

Figure 8: Box plot of the dispersion in scores for the people who are female (left) and those who are male (right).

Meaningfulness

```
0.3734/(1-0.3734)
```

```
## [1] 0.5959145
```

```
library(pwr)
pwr.f2.test(u=2, v=53, f2=0.5959, sig.level=1.56e-06) # where the sig.level is the
p-value of the model.
```

```
##
##      Multiple regression power calculation
##
##               u = 2
##               v = 53
##              f2 = 0.5959
##       sig.level = 1.56e-06
##           power = 0.498524
```

The power is 0.499, which is not the 0.8 that we want. Therefore, we can say that we don't have enough partipants to reach this power of 0.8. As for the methodology of the study, we would maybe want to know more about the lexical decision task and further description on the educational background of the participants.