705604096_stats101b_hw3

Jade Gregory

2023-08-31

Question 1

```
compdat <- read.csv("data Readtexts paper vs ereader Article.csv")
head(compdat)
```

```
##   Studentnumber Sex School Condition Vocabulary Wordchain SumExpositoryTxt
## 1       2602142 boy      1     paper         18        53               16
## 2       2602145 boy      1     paper         18        53               17
## 3       2602158 boy      1     paper         23        53               17
## 4       2602177 boy      1     paper         10        50               13
## 5       2602261 boy      2     paper         22        52               18
## 6       2602282 boy      2     paper         16        51               17
##   SumNarrativeTxt TotalSum_Readingcompr ControlsumNarrativeTxt
## 1              15                    31                      9
## 2              11                    28                      9
## 3              15                    32                     10
## 4              10                    23                      7
## 5              13                    31                      6
## 6               7                    24                      6
##   ControlsumExpositoryTxt Controlsum Gaven2Mc Gaven3Mc Gaven4Mc Gaven5Cr
## 1                      15         24        4        1        2        1
## 2                      16         25        4        1        2        1
## 3                      15         25        4        1        2        1
## 4                       2          9        4        1        2        1
## 5                      13         19        4        1        3        1
## 6                      16         22        4        1        2        9
##   Gaven6Cr Gaven7Cr Gaven8Cr Gaven9Cr Gaven10Mc Gaven11Cr Gaven12Cr SH1Mc SH3Mc
## 1        0        2        0        1         1         1         1     3     2
## 2        0        2        0        1         3         1         0     3     2
## 3        9        2        1        0         3         0         2     3     2
## 4        9        1        0        1         2         1         0     1     4
## 5        9        2        0        0         3         9         9     3     2
## 6        9        9        9        1         3         1         9     3     4
##   SH4Mc SH5Mc SH6Mc SH10Mc SH11Mc SH12Mc SH13Mc SH15Mc SH16Mc SH17Mc SH18Mc
## 1     1     2     2      4      1      2      3      3      2      3      4
## 2     1     2     2      2      4      2      2      3      2      3      1
## 3     1     2     2      2      4      2      2      3      2      3      4
## 4     1     1     2      9      9      9      9      9      9      9      9
## 5     1     2     2      1      4      2      2      3      2      3      2
## 6     1     2     2      3      4      2      2      3      2      4      1
##   SH19Mc SH20Mc SH21Mc SH22Mc SH24Cr SH25Mc SH26Mc SH27Mc Antarktis1 Antarktis2
## 1      4      3      3      3      9      9      9      9          4          4
```

```
## 2      4      3      3      2      2      2      3      2          4          4
## 3      4      3      8      9      2      3      3      2          4          4
## 4      9      9      9      9      9      9      9      9          4          4
## 5      4      3      1      1      9      1      2      1          4          4
## 6      4      3      3      3      2      3      3      2          4          4
##   Antarktis3 Antarktis4 Antarktis5 Antarktis6 Antarktis7 Antarktis8 Antarktis9
## 1          2          4          3          3          3          1          4
## 2          2          4          3          1          3          1          4
## 3          2          4          3          1          3          1          4
## 4          3          4          3          3          3          4          4
## 5          1          4          3          1          3          1          4
## 6          1          4          3          1          3          1          4
##   Antarktis10 Antarktis11 Antarktis12 Antarktis13 Antarktis14 Antarktis15
## 1           4           1           2           3           1           2
## 2           2           1           2           3           1           2
## 3           4           1           2           3           2           2
## 4           3           1           2           3           4           2
## 5           4           1           2           3           1           2
## 6           3           1           2           3           1           2
##   Antarktis16 Antarktis17 Antarktis18 Antarktis19 Coriander1 Coriander2
## 1           1           4           1           4          3          2
## 2           3           4           1           4          3          2
## 3           3           4           1           4          3          2
## 4           4           4           1           4          3          2
## 5           3           4           1           4          3          2
## 6           3           4           1           4          3          2
##   Coriander3 Coriander4 Coriander5 Coriander6 Coriander7 Coriander8 Coriander9
## 1          2          2          3          8          1          2          3
## 2          2          2          3          8          1          4          1
## 3          2          2          3          8          1          2          3
## 4          2          2          4          8          2          2          1
## 5          4          2          3          8          1          2          3
## 6          2          1          1          8          2          1          3
##   Coriander10 Coriander11 Coriander12 Coriander13 Coriander14 Coriander15
## 1           2           1           4           1           3           4
## 2           2           4           4           2           3           4
## 3           2           1           4           1           3           4
## 4           2           1           4           1           3           2
## 5           2           1           4           1           2           4
## 6           2           2           1           3           3           2
##   Spørsmål1 Spørsmål2 Spørsmål3 Spørsmål4 Spørsmål5 Spørsmål6 Spørsmål7
## 1         1         1         0         1         1         0         9
## 2         1         1         0         1         1         0         0
## 3        NA        NA        NA        NA        NA        NA        NA
## 4         1         1         1         1         1         0         0
## 5        NA        NA        NA        NA        NA        NA        NA
## 6         1         1         0         1         0         0         9
##   pgaven2Mc pgaven3Mc pgaven4Mc pgaven5Cr pgaven6cr pgaven7Cr pgaven8Cr
## 1         1         1         1         1         0         2         0
## 2         1         1         1         1         0         2         0
## 3         1         1         1         1         0         2         1
## 4         1         1         1         1         0         1         0
## 5         1         1         0         1         0         2         0
## 6         1         1         1         0         0         0         0
```

```
##   pgaven9Cr pgaven10Mc pgaven11cr pgaven12Cr X.1Score X.2Score X.3Score
## 1         1          0          1          1        1        1        0
## 2         1          1          1          0        1        1        0
## 3         0          1          0          2        1        1        0
## 4         1          0          1          0        1        1        0
## 5         0          1          0          0        1        1        0
## 6         1          1          1          0        1        1        0
##   X.4Score X.5Score X.6Score X.7Score X.8Score X.9Score X.10Score X.11Score
## 1        1        1        0        1        1        1         1         1
## 2        1        1        1        1        1        1         0         1
## 3        1        1        1        1        1        1         1         1
## 4        1        1        0        1        0        1         0         1
## 5        1        1        1        1        1        1         1         1
## 6        1        1        1        1        1        1         0         1
##   X.12Score X.13Score X.14Score X.15Score X.16Score X.17Score X.18Score
## 1         1         1         1         1         0         1         1
## 2         1         1         1         1         1         1         1
## 3         1         1         0         1         1         1         1
## 4         1         1         0         1         0         1         1
## 5         1         1         1         1         1         1         1
## 6         1         1         1         1         1         1         1
##   X.19Score X.20Score X.21Score X.22Score X.23Score X.24Score X.25Score
## 1         1         1         1         1         1         1         1
## 2         1         1         1         1         1         1         1
## 3         1         1         1         1         1         1         1
## 4         1         1         1         1         1         0         0
## 5         1         1         1         0         1         1         1
## 6         1         1         1         1         0         0         1
##   X.26Score X.27Score X.28Score X.29Score X.30Score X.31Score X.32Score
## 1         1         1         1         1         1         1         1
## 2         1         0         0         1         0         1         0
## 3         1         1         1         1         1         1         1
## 4         0         1         0         1         1         1         1
## 5         1         1         1         1         1         1         1
## 6         0         0         1         1         0         0         0
##   X.33Score X.34Score pSH1 pSH3Mc pSH4Mc pSH5Mc pSH6Mc pSH10Mc pSH11Mc pSH12Mc
## 1         1         1    1      1      1      1      1       1       1       1
## 2         1         1    1      1      1      1      1       0       0       1
## 3         1         1    1      1      1      1      1       0       0       1
## 4         1         0    0      0      1      0      1       0       0       0
## 5         0         1    1      1      1      1      1       0       0       1
## 6         1         0    1      0      1      1      1       0       0       1
##   pSH13Mc pSH15Mc pSH16Mc pSH17Mc pSH18Mc pSH19Mc pSH20Mc pSH21Mc pSH22Mc
## 1       0       1       1       1       0       1       1       1       1
## 2       1       1       1       1       1       1       1       1       0
## 3       1       1       1       1       0       1       1       0       0
## 4       0       0       0       0       0       0       0       0       0
## 5       1       1       1       1       0       1       1       0       0
## 6       1       1       1       0       1       1       1       1       1
##   pSH24Cr pSH25Mc pSH26Mc pSH27Mc Paper_Paper Paper_Computer Computer_Computer
## 1       0       0       0       0           0              1                 0
## 2       1       0       0       1           0              1                 0
## 3       1       1       0       1           0              1                 0
## 4       0       0       0       0           0              1                 0
```

```
## 5       0       0       1       0             0                 1                 0
## 6       1       1       0       1             0                 1                 0
```

a) The variables measured in this study include:

- Response Variables: Vocabulary pretest score, word chain test score, reading comprehension pretest score
- Factors: Condition (paper or computer texts)
- Blocks: School, Sex
- Held Constant: Grade level, country
- If I were to conduct my own experiment of this nature, I would increase the sample size for the study. This would provide us with a more accurate representation of our population. I would also expand the study to include participants from many other schools and possible age ranges, to account for the fact that different ages may have different reading comprehension levels. I would also include the same number of each group in my study, including the number of boys and girls surveyed as well as each number of boys and girls who participate in the electronic readings as well as the paper readings.

b)

```
nrow(compdat)
```

```
## [1] 72
```

There were 72 participants in this study. This was not enough people for this study. From the G*Power app, we find that the power associated with this sample size and the effect size -0.02 is 0.0588106. The power associated with this sample size and the effect size of -0.03 is 0.0636510. The power associated with this sample size and the effect size of 0.06 is 0.0800399. Because we have such small power values, I believe we should have more participants in this study.

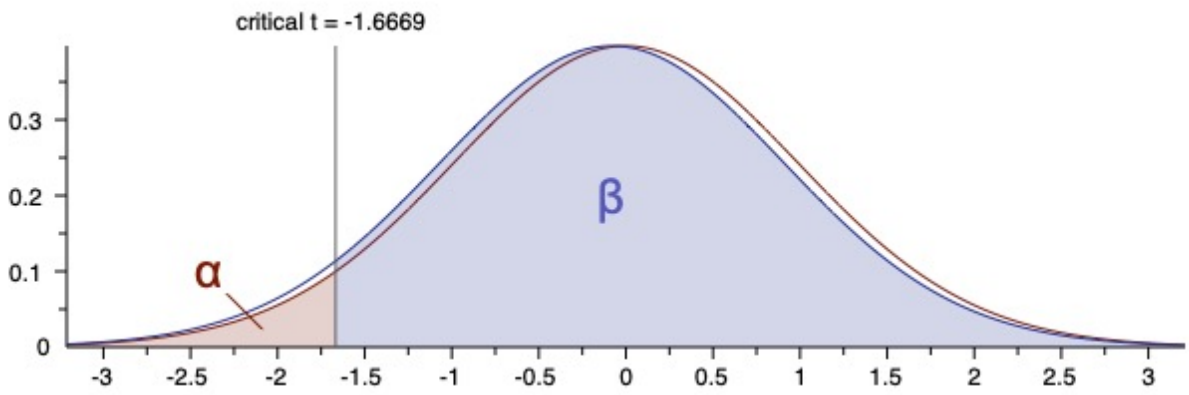


Figure 1: Distribution Plot for effective size = -0.02

c) When the author says "With respect to reliability, all texts used in this study, both for pretesting and in the main survey had Cronbach's alpha >.75" they are referring to the measure of reliability also known as internal consistency. This explains how efficiently a test is measuring what it intends to be measuring. A reliability of >0.75 is a good reliability for an experiment.

d)

```
t.test(compdat$Controlsum ~ compdat$Condition, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  compdat$Controlsum by compdat$Condition
```

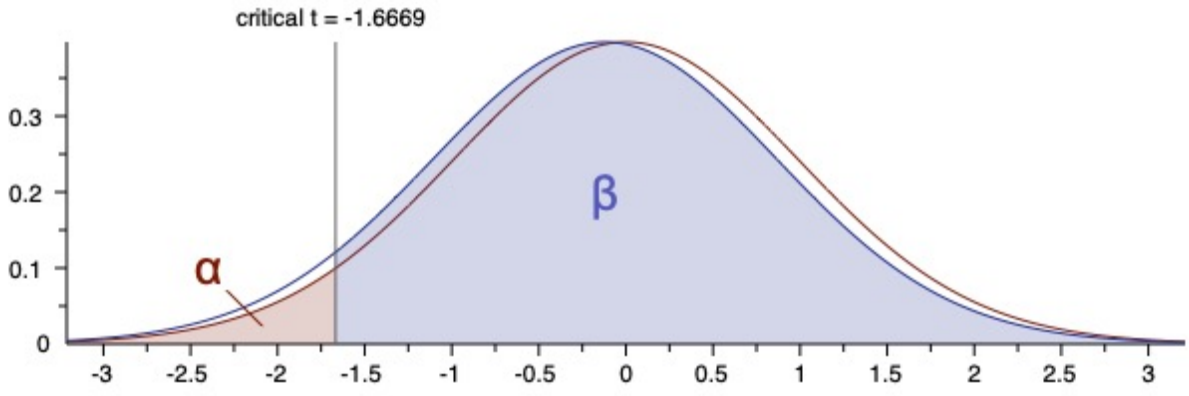


Figure 2: Distribution Plot for effective size = -0.03

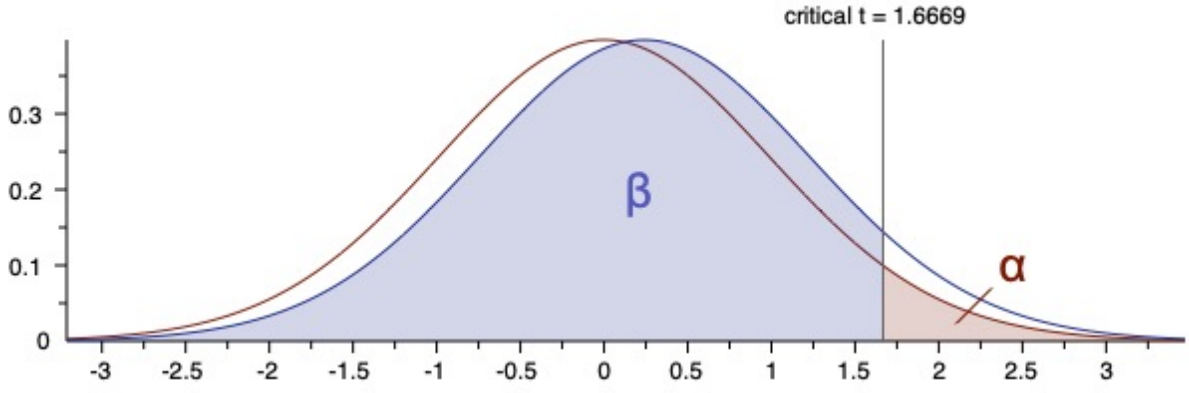


Figure 3: Distribution Plot for effective size = 0.06

```
## t = 0.086874, df = 70, p-value = 0.931
## alternative hypothesis: true difference in means between group electronic and group paper is not equa
## 95 percent confidence interval:
##  -3.587996  3.914804
## sample estimates:
## mean in group electronic      mean in group paper
##                  19.7234                  19.5600
```

This is our t test involving variables Controlsum and Condition. From our p-value of 0.931 which is greater than our significance level of 0.05, we fail to reject our null hypotehsis stating that the difference of the group means is zero.

```
t.test(compdat$Vocabulary ~ compdat$Condition, var.equal = TRUE)
```

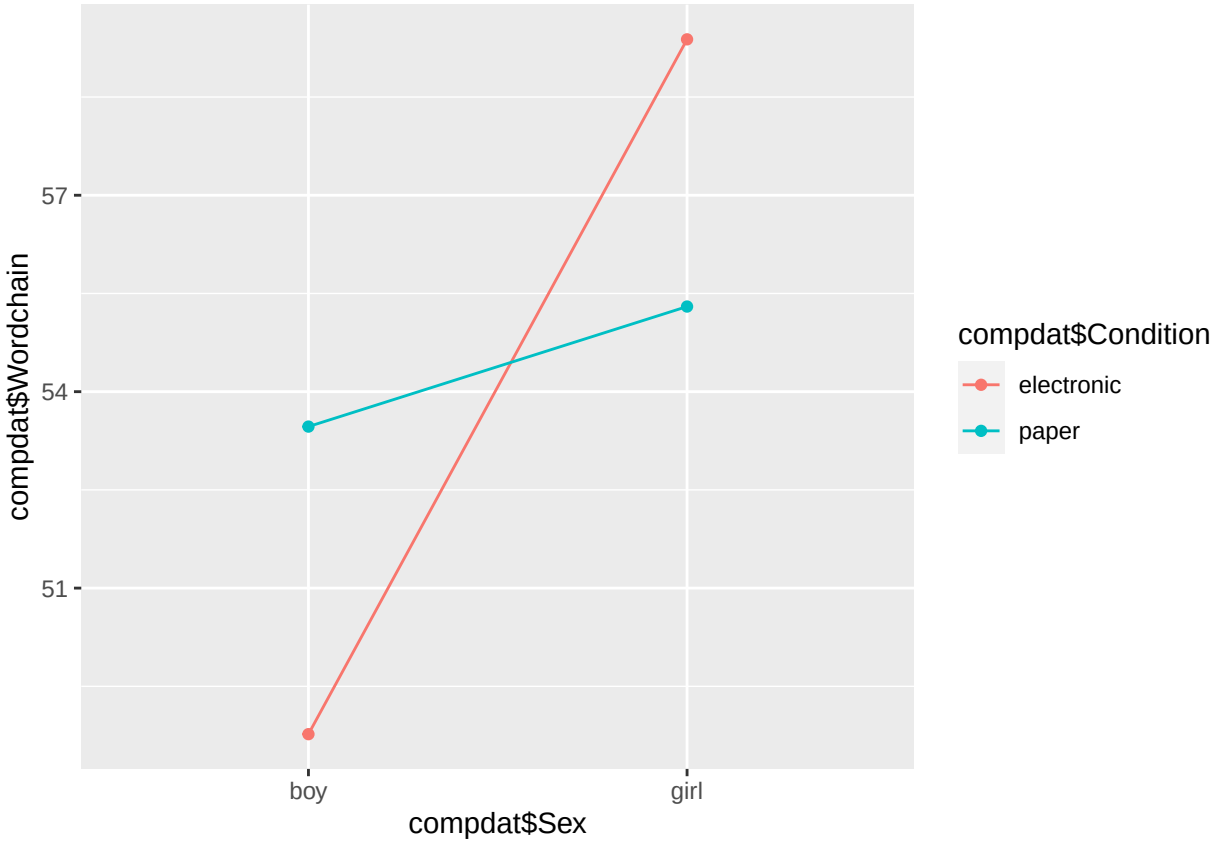
```
##
##  Two Sample t-test
##
## data:  compdat$Vocabulary by compdat$Condition
## t = 0.14123, df = 70, p-value = 0.8881
## alternative hypothesis: true difference in means between group electronic and group paper is not equa
## 95 percent confidence interval:
##  -1.485265  1.711648
## sample estimates:
## mean in group electronic      mean in group paper
##                 18.55319                 18.44000
```

This is our t test involving variables Vocabulary and Condition. From our p-value of 0.8881 which is greater than our significance level of 0.05, we fail to reject our null hypothesis stating that the difference in the group means in zero.

```
t.test(compdat$Wordchain ~ compdat$Condition, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  compdat$Wordchain by compdat$Condition
## t = -0.24983, df = 70, p-value = 0.8035
## alternative hypothesis: true difference in means between group electronic and group paper is not equa
## 95 percent confidence interval:
##  -6.192646  4.813923
## sample estimates:
## mean in group electronic      mean in group paper
##                 53.51064                 54.20000
```

```
ggplot() + aes(x = compdat$Sex, color = compdat$Condition,
               group = compdat$Condition, y = compdat$Wordchain) +
  stat_summary(fun = mean, geom = "point") +
  stat_summary(fun = mean, geom = "line")
```



This is our t test involving variables Wordchain and Condition. Since our p-value is 0.8035 which is greater than our significance level of 0.05, we fail to reject our null hypothesis stating that the difference of the two group means is zero. From our interaction plot, we can see that the two lines involving word chain comprehension and condition intersect. This would lead us to believe that there may be interaction amongst these variables, and we should investigate them further in our analysis.

e)

```
#model including Vocabulary, Word reading, and Reading comprehension pretest
m1 <- lm(TotalSum_Readingcompr ~ Vocabulary + Wordchain + Controlsum, data = compdat)
summary(m1)
```

```
##
## Call:
## lm(formula = TotalSum_Readingcompr ~ Vocabulary + Wordchain +
##     Controlsum, data = compdat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.726  -2.866   1.349   3.196   9.431
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.57604    4.28916   0.834  0.40735
## Vocabulary   0.51919    0.21615   2.402  0.01904 *
## Wordchain    0.18476    0.06334   2.917  0.00479 **
## Controlsum   0.16557    0.09754   1.697  0.09419 .
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.307 on 68 degrees of freedom
## Multiple R-squared:  0.3466, Adjusted R-squared:  0.3178
## F-statistic: 12.02 on 3 and 68 DF,  p-value: 2.088e-06
```

```
#model including Vocabulary, Word reading, Reading comprehension pretest, and Sex
m2 <- lm(TotalSum_Readingcompr ~ Vocabulary + Wordchain + Controlsum + Sex, data = compdat)
summary(m2)
```

```
##
## Call:
## lm(formula = TotalSum_Readingcompr ~ Vocabulary + Wordchain +
##     Controlsum + Sex, data = compdat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.1391  -2.8014   0.8877   3.6863   8.0322
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.12466    4.25790   0.969   0.3362
## Vocabulary   0.53998    0.21426   2.520   0.0141 *
## Wordchain    0.15193    0.06606   2.300   0.0246 *
## Controlsum   0.16200    0.09653   1.678   0.0980 .
## Sexgirl      2.09289    1.33194   1.571   0.1208
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.25 on 67 degrees of freedom
## Multiple R-squared:  0.3698, Adjusted R-squared:  0.3322
## F-statistic: 9.831 on 4 and 67 DF,  p-value: 2.561e-06
```

```
#model including Vocabulary, Word reading, Reading comprehension pretest, Sex, and Reading modality
m3 <- lm(TotalSum_Readingcompr ~ Vocabulary + Wordchain + Controlsum + Sex + Condition, data = compdat)
summary(m3)
```

```
##
## Call:
## lm(formula = TotalSum_Readingcompr ~ Vocabulary + Wordchain +
##     Controlsum + Sex + Condition, data = compdat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.0885  -1.9287   0.7578   2.9567   8.8092
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     3.22382    4.14805   0.777   0.4398
## Vocabulary      0.55082    0.20784   2.650   0.0101 *
## Wordchain       0.14356    0.06417   2.237   0.0286 *
## Controlsum      0.16550    0.09363   1.768   0.0817 .
## Sexgirl         2.27050    1.29405   1.755   0.0840 .
## Conditionpaper  2.89245    1.26390   2.289   0.0253 *
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.092 on 66 degrees of freedom
## Multiple R-squared:  0.4162, Adjusted R-squared:  0.3719
## F-statistic: 9.409 on 5 and 66 DF,  p-value: 8.342e-07
```

```
#partial f test for m1 and m2 and m3
anova(m1, m2, m3)
```

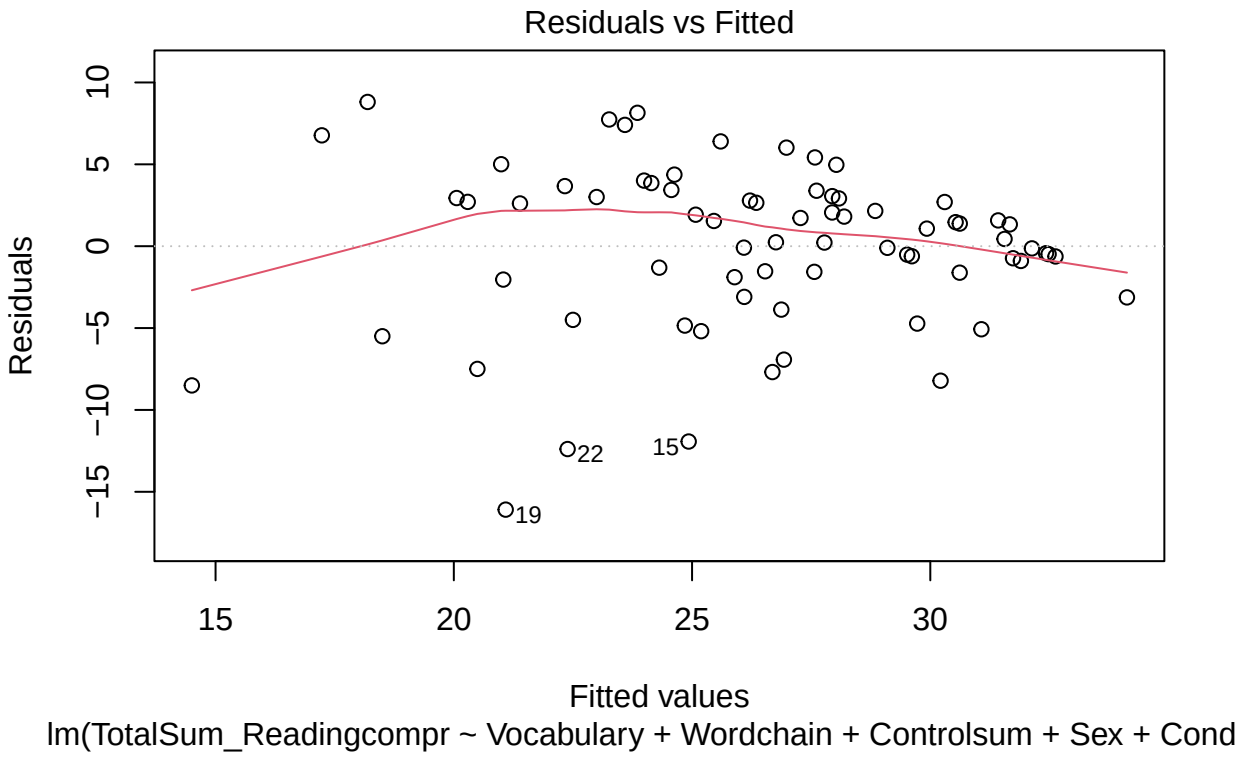
```
## Analysis of Variance Table
##
## Model 1: TotalSum_Readingcompr ~ Vocabulary + Wordchain + Controlsum
## Model 2: TotalSum_Readingcompr ~ Vocabulary + Wordchain + Controlsum +
##     Sex
## Model 3: TotalSum_Readingcompr ~ Vocabulary + Wordchain + Controlsum +
##     Sex + Condition
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     68 1915.0
## 2     67 1846.9  1     68.06 2.6251 0.10995
## 3     66 1711.1  1    135.78 5.2373 0.02532 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

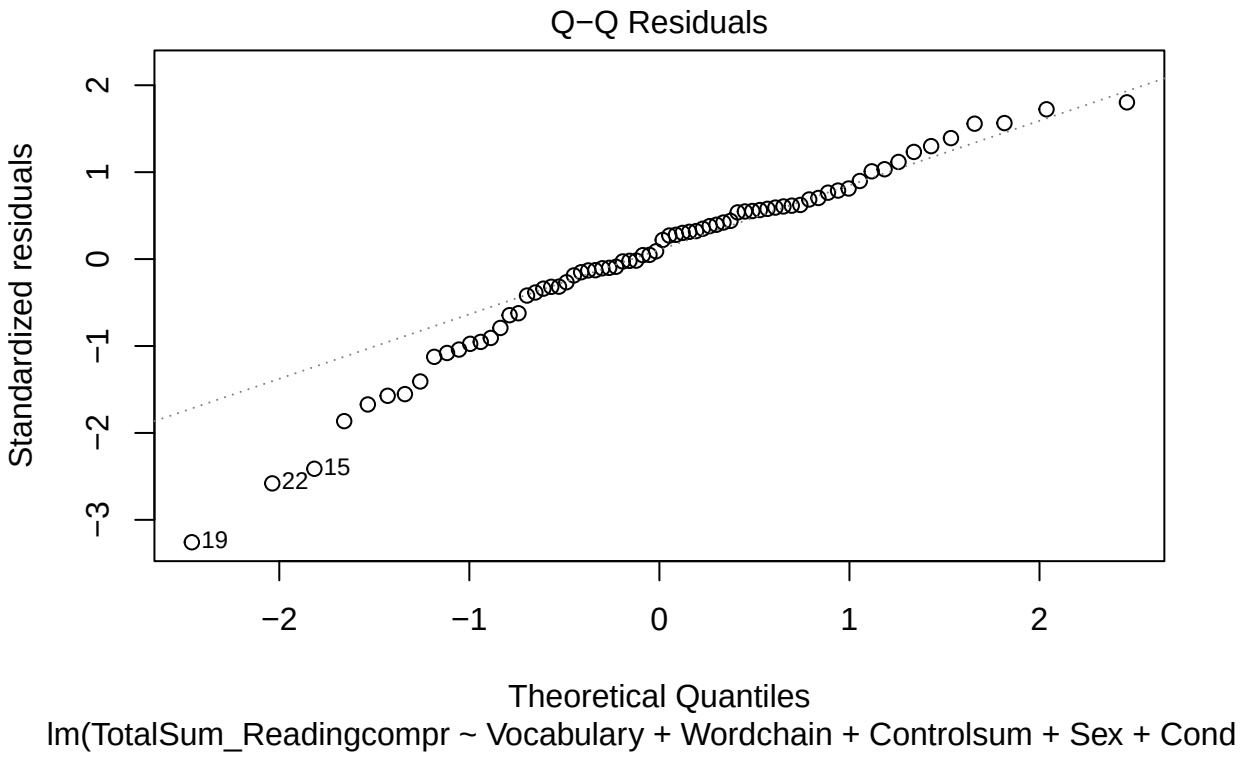
```
#partial f test comparing m1 and m3
anova(m1, m3)
```

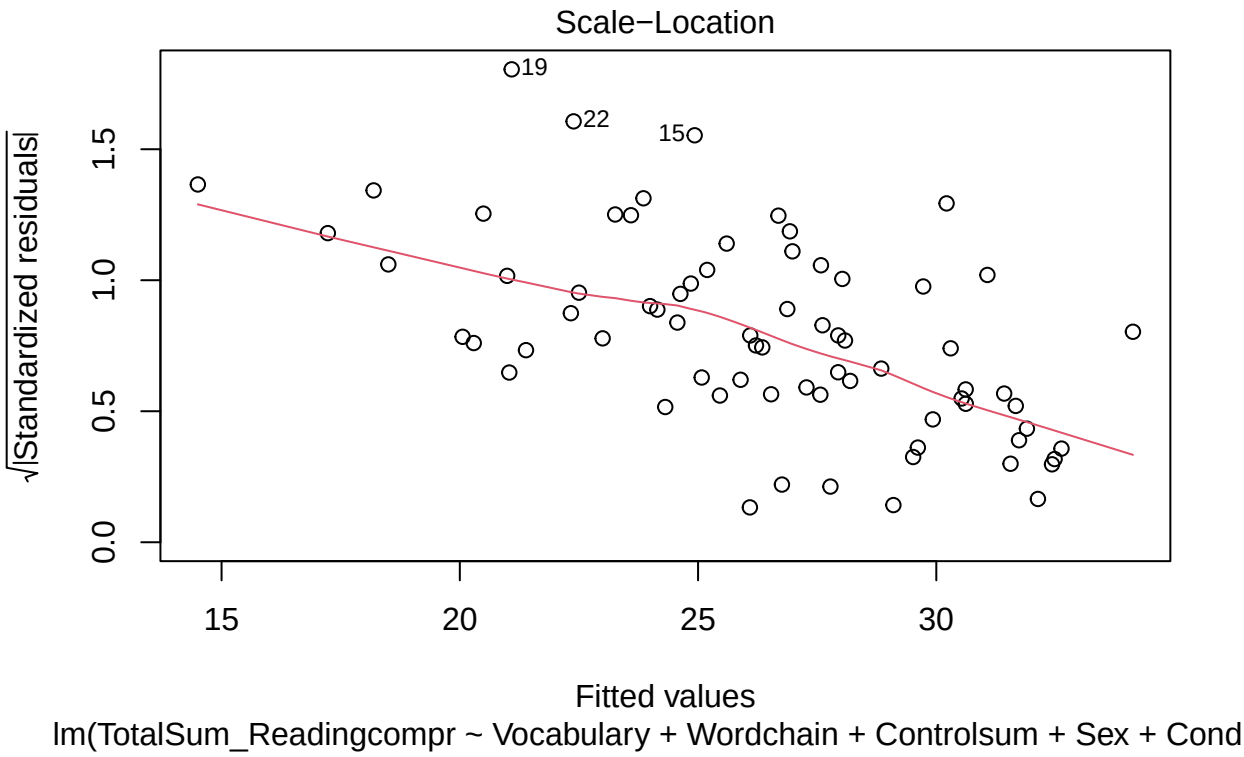
```
## Analysis of Variance Table
##
## Model 1: TotalSum_Readingcompr ~ Vocabulary + Wordchain + Controlsum
## Model 2: TotalSum_Readingcompr ~ Vocabulary + Wordchain + Controlsum +
##     Sex + Condition
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     68 1915.0
## 2     66 1711.1  2    203.84 3.9312 0.02438 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

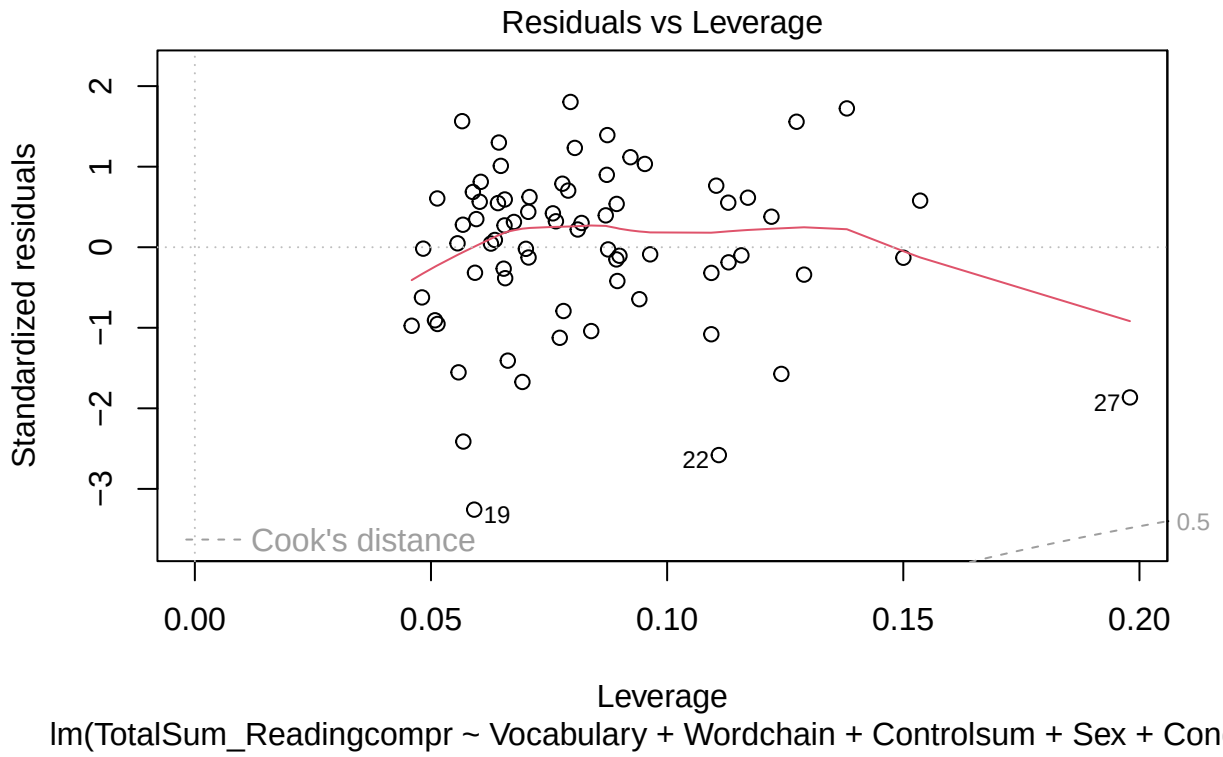
From our anova of all three models, we can see that m3 has a p-value of 0.02532 which is less than our significance level of 0.05, making this model statistically significant compared to our m2 which has a p-value of 0.10995 which is greater than our significance level of 0.05. Because of this, we can then compare m3 directly to m1 to see how it holds. Again, the m3 model is statistically significant with a p-value of 0.02438 being less than the significance level of 0.05. Now we can conclude that our final model from our MLR is m3 which includes the variables TotalSum_Readingcompr ~ Vocabulary + Wordchain + Controlsum + Sex + Condition.

```
plot(m3)
```


Residuals vs Fitted
Residuals
10
5
0
−5
−10
−15
15
20
25
30
22
15
19
Fitted values
lm(TotalSum_Readingcompr ~ Vocabulary + Wordchain + Controlsum + Sex + Cond


Q−Q Residuals
Standardized residuals
Theoretical Quantiles
lm(TotalSum_Readingcompr ~ Vocabulary + Wordchain + Controlsum + Sex + Cond
19
22
15


Scale−Location
19
22
15
√|Standardized residuals|
0.0
0.5
1.0
1.5
15
20
25
30
Fitted values
lm(TotalSum_Readingcompr ~ Vocabulary + Wordchain + Controlsum + Sex + Cond



In our residuals vs fitted plot for the m3 model, we can see that the data points are plotted horizontally across the graph. We do notice a fan pattern emerge from the graph which can indicate that the constant variance assumption is not held by our model. In the QQ norm plot for the m3 model we can see that the data points follow the dashed line tightly and stray at the ends of the graph, indicating that our normality assumption is held by our model. Our scale location plot has data points plotted horizontally across the graph, but there is a noticeable decreasing trend amongst them further supporting the idea that the constant variance assumption is not held by this model. In our residuals vs leverage plot we can see that the data points are plotted equally and horizontally across the graph with no noticeable points residing in the Cook's distance portion of the graph. We can further assume that there are not many outliers or influential points in this data. Overall, we can conclude that the model most likely holds the normality assumption but the constant variance assumption must be analyzed further.

Question 2

a)

```
baldness <- read.delim2("Ch 8 HeartDiseaseBaldness.txt")
head(baldness)
```

```
##   Heart_Disease Baldness
## 1            no     none
## 2            no     none
## 3            no     none
## 4            no     none
## 5            no     none
## 6           yes     none
```

b)

```
dim(baldness)
```

```
## [1] 1435    2
```

The baldness data frame has the dimensions of 2 columns and 1435 rows.

c) The names of the variables in the data include Heart_Disease and Baldness.

d) The null hypothesis states that the two variables do not have a significant relationship to one another and the alternative hypothesis states that the two variables do have a significant relationship to one another.

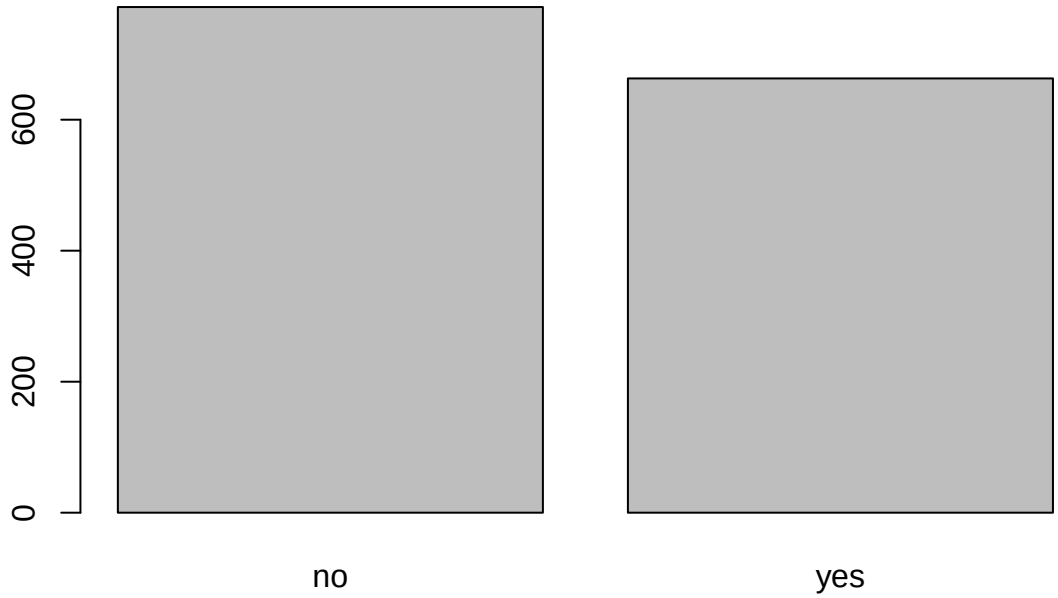
e)

```
table(baldness)
```

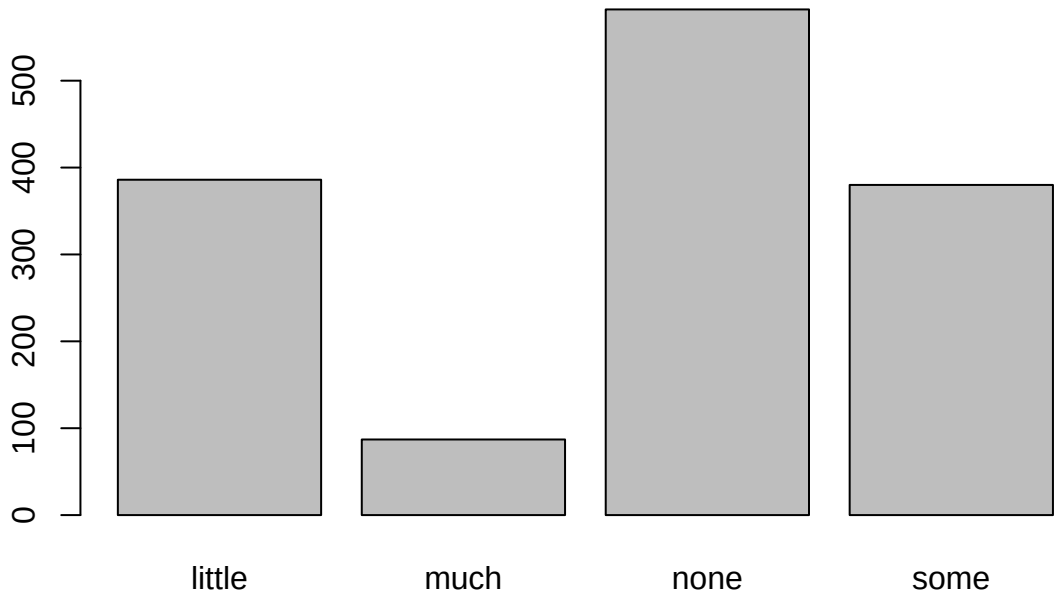
```
##               Baldness
## Heart_Disease little much none some
##           no     221   35  331  185
##           yes    165   52  251  195
```

f)

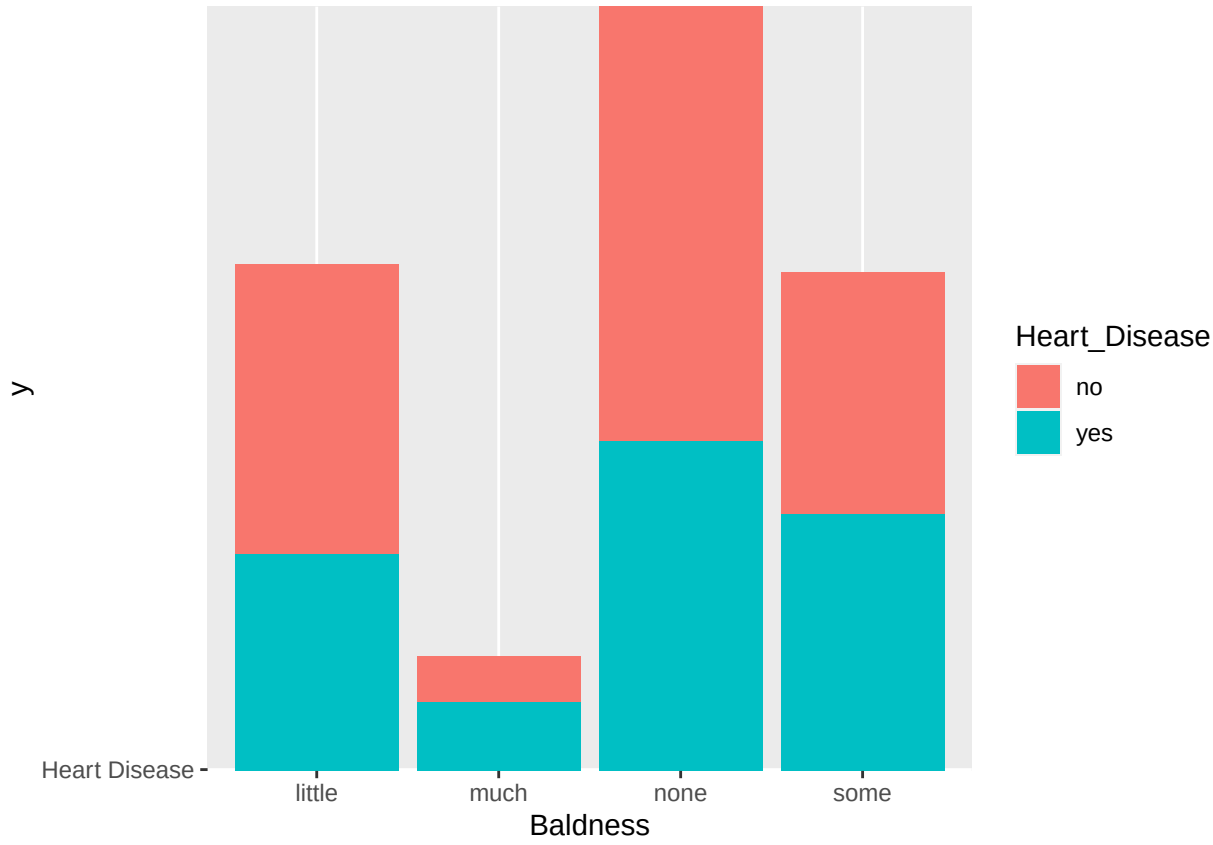
```
barplot(table(baldness$Heart_Disease))
```




```
barplot(table(baldness$Baldness))
```




```
ggplot(baldness, aes(fill = Heart_Disease, y = "Heart Disease", x = Baldness)) +
  geom_bar(position = "stack", stat = "identity")
```



g)

```
chisq.test(table(baldness))
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(baldness)
## X-squared = 14.51, df = 3, p-value = 0.002287
```

Since our p-value of 0.002287 is less than our significance level of 0.05, it is significant and we reject our null hypothesis and can conclude that there is evidence to support the idea that the two variables, Heart_Disease and Baldness are related to one another.