Stats 101B Homework #4

Anna Piskun

5/16/2020

Question 1:

List the variables measured in this study. Response, factors, blocks, held constant-factors. What would you add to enhance this study in terms of factors and design?

Response Variable- Reading comprehension test score, word-chain test (number of word-chains solved in 4 mins), single-word-item semantic vocabulary test

Factors - digital (computer) texts, paper texts

Blocks - school (two urban primary schools in Norway were used) and gender

Held-Constant Factors - Age (all were 10th graders), Ethnicity, Socio-economic status, Environment (students' resptive classrooms), Country (Norway), type of font, type of computer screen

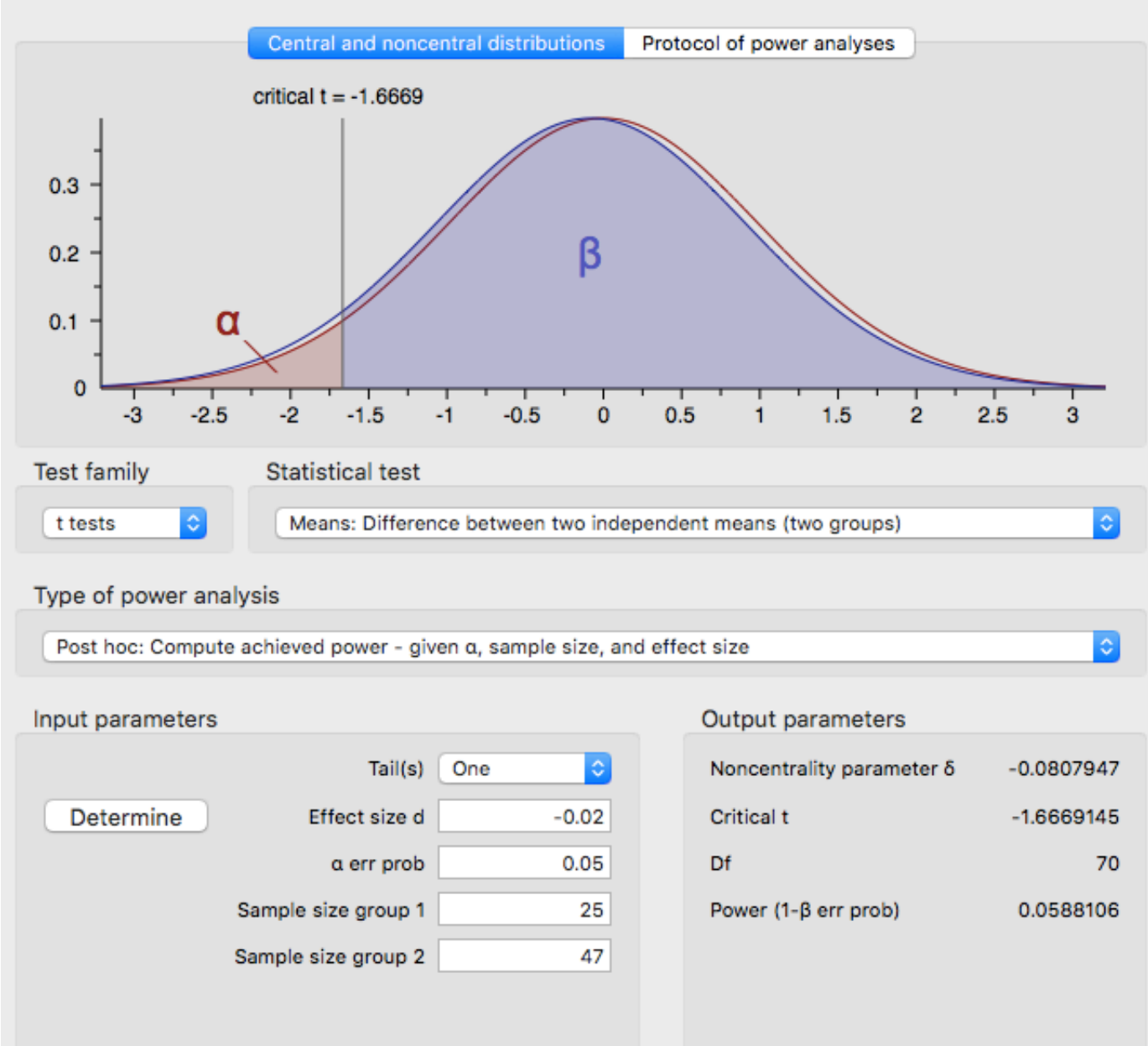
In terms of experimental design, if given a sample size of 72 and unlimited resources I would have utilized a repeated measures design to enhance the overall study. Given unlimited time and resources, 4 weeks after taking the pretest (as specified by the original design), I would have all participants read the texts on paper and then complete the following reading comprehension test. 4 weeks after that, I would have all participants read another text digitally (still consisting of a narrative and expository text produced by The Reading Centre with Cronbach's alppha >0.75). I would then compare reading comprehension vs. reading modality for each participant. Thus, each subject would be treated as a block which inherently decreases variability and increases power (without needing to increase sample size, which we later see is a significant flaw in this study).

In order to enhance the study I would significantly increase sample size (to increase the overall power of the study), sampling more schools and more 10th grade students. This also as a result introduced more blocks which will help decrease variability as well. I would also introduced balanced group design, meaning both treatment groups (digital vs. paper text) get an equal number of participants. Additionally, to further improve upon the experimental design I would add an interaction between the sex and condition variables. As seen below, there is evidence of an interaction that may help describe more of the variance in our model and as such should be included in our analysis to test for statistical signficance and reult in the best overall model. Lastly, in the final question we see how the three MLR models are not valid since all three fail to satisfy the constant variance condition. We can attempt to solve this problem and better fit the data by transforming either the response or predictor variables. Through utilizing a transformation, we will get a better model that is valid.

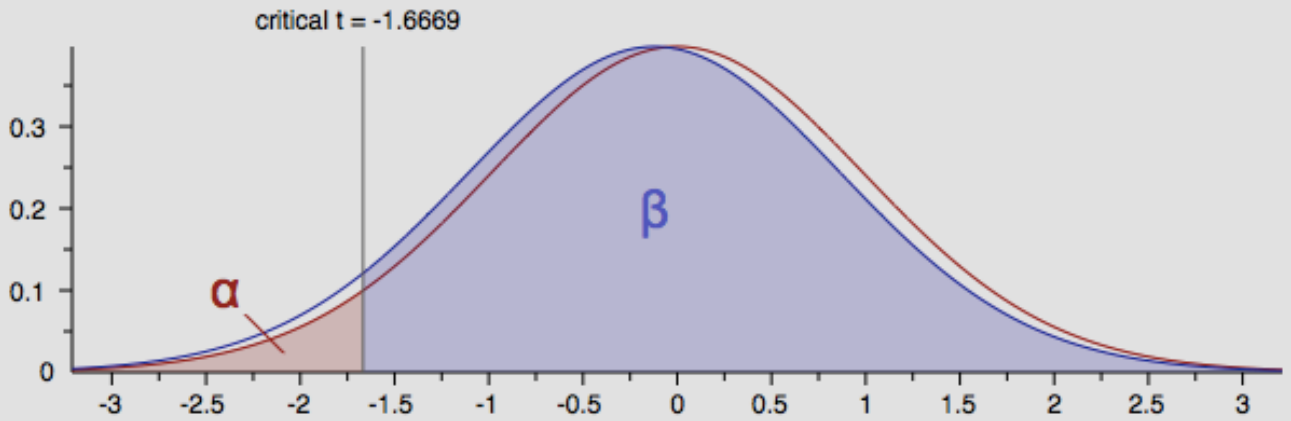
Question 2:

How many participants were there? Do you think they are enough? What is the associated power with the sample size chosen in this study?

There were 72 participants with 43% being females and 57% being males. The associated power with the sample size chosen in this study is approximately 0.058 for an effect size of -0.02, 0.063 for an effect size of -0.03, and 0.08 for an effect size of 0.06. Since power is extremely small (meaning the probability of finding a difference when a difference really does exist is low), there are not enough participants and we need to significantly increase the sample size.


Central and noncentral distributions
Protocol of power analyses
critical t = -1.6669
0.3
0.2
0.1
0
α
β
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
Test family
t tests
Statistical test
Means: Difference between two independent means (two groups)
Type of power analysis
Post hoc: Compute achieved power - given α, sample size, and effect size
Input parameters
Tail(s)
One
Determine
Effect size d
-0.02
α err prob
0.05
Sample size group 1
25
Sample size group 2
47
Output parameters
Noncentrality parameter δ
-0.0807947
Critical t
-1.6669145
Df
70
Power (1-β err prob)
0.0588106

Central and noncentral distributions | Protocol of power analyses



Test family: t tests

Statistical test: Means: Difference between two independent means (two groups)

Type of power analysis: Post hoc: Compute achieved power - given α, sample size, and effect size

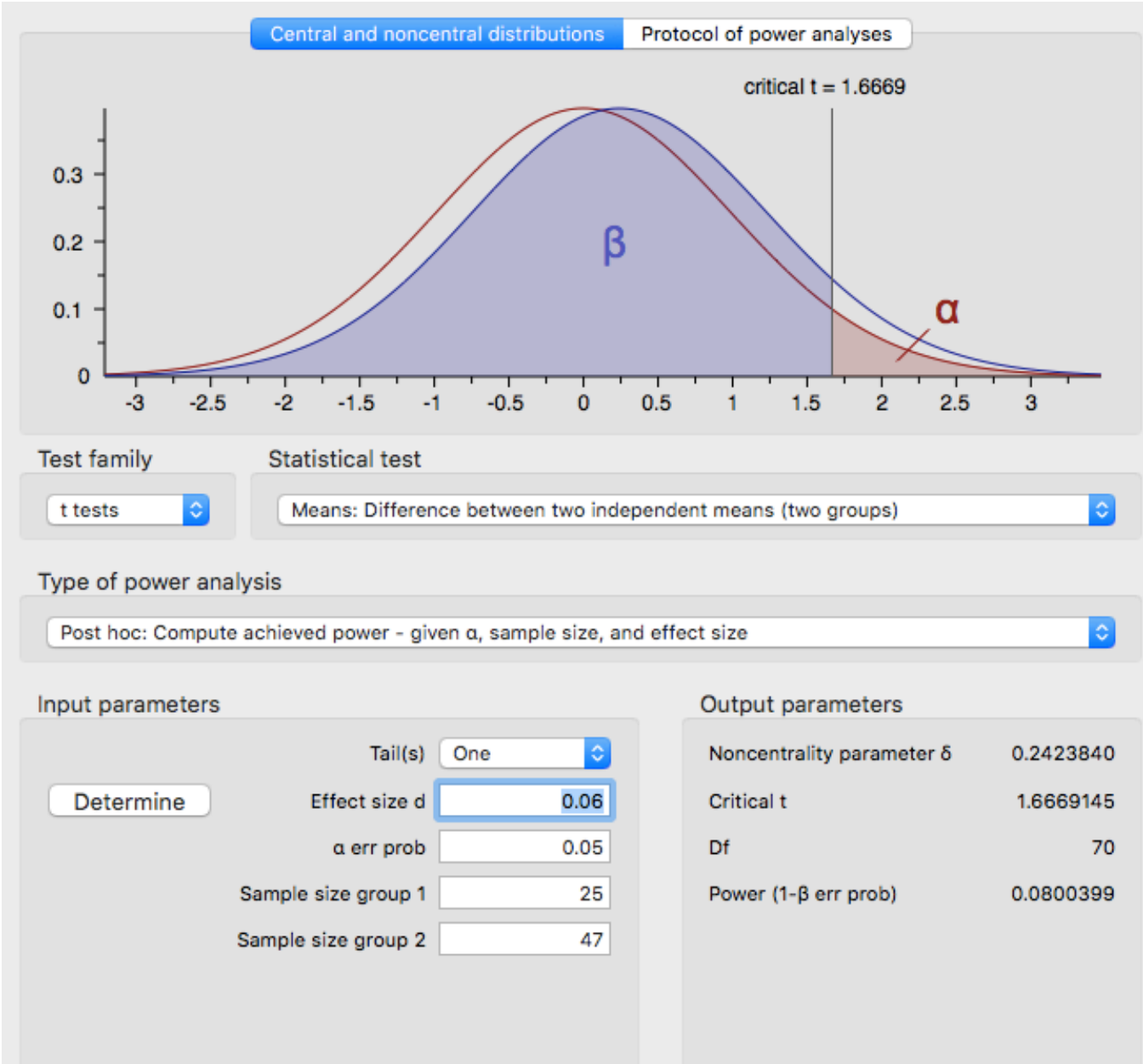
Input parameters

Tail(s)	One
Effect size d	-0.03
α err prob	0.05
Sample size group 1	25
Sample size group 2	47

Determine

Output parameters

Noncentrality parameter δ	-0.1211920
Critical t	-1.6669145
Df	70
Power (1-β err prob)	0.0636510


Central and noncentral distributions
Protocol of power analyses
critical t = 1.6669
0.3
0.2
0.1
0
β
α
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
Test family
t tests
Statistical test
Means: Difference between two independent means (two groups)
Type of power analysis
Post hoc: Compute achieved power - given α, sample size, and effect size
Input parameters
Tail(s)
One
Determine
Effect size d
0.06
α err prob
0.05
Sample size group 1
25
Sample size group 2
47
Output parameters
Noncentrality parameter δ
0.2423840
Critical t
1.6669145
Df
70
Power (1-β err prob)
0.0800399

Question 3:

What does the author mean by "With respect to reliability, all texts used in this study, both for pretesting and in the main survey had Cronbach's alpha>.75."

Generally speaking, Cronbach's alpha is a measure of "how related a set of items are as a group" (according to UCLA IDRE Statistical Consulting Center). It is used to measure within group consistency and is often used as a measure of scale reliability. Essentially, the cronbach's alpha sees how well a test measures what it should. A rule of thumb most statisticians follow is that a Cronbach's alpha of 0.70 and above is satisfactory (above .80 is better, and above .90 is best). Within the context of the author's research, this means that the scale formed by the reading comprehension questions in the pretest and actual test is reliable. Thus, the questions used in the pretest and actual test measured similar concepts indicating similar and reliable results.

Question 4:

Using R, download the posted data and perform the t-tests posted on page 64 (section 3 Results). Then create an interaction plot using type of gender and condition to get "Word chain" responses. Comment on your plot.

```
data <- read.csv("HW 4 data Readtexts Article.csv")

#Vocabulary T-Test
t.test(Vocabulary~Condition, data = data)
```

```
##
##  Welch Two Sample t-test
##
## data:  Vocabulary by Condition
## t = 0.13784, df = 45.837, p-value = 0.891
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.539919  1.766302
## sample estimates:
## mean in group electronic      mean in group paper
##                 18.55319                 18.44000
```

```
#Word-chain T-test
t.test(Wordchain~Condition, data = data)
```

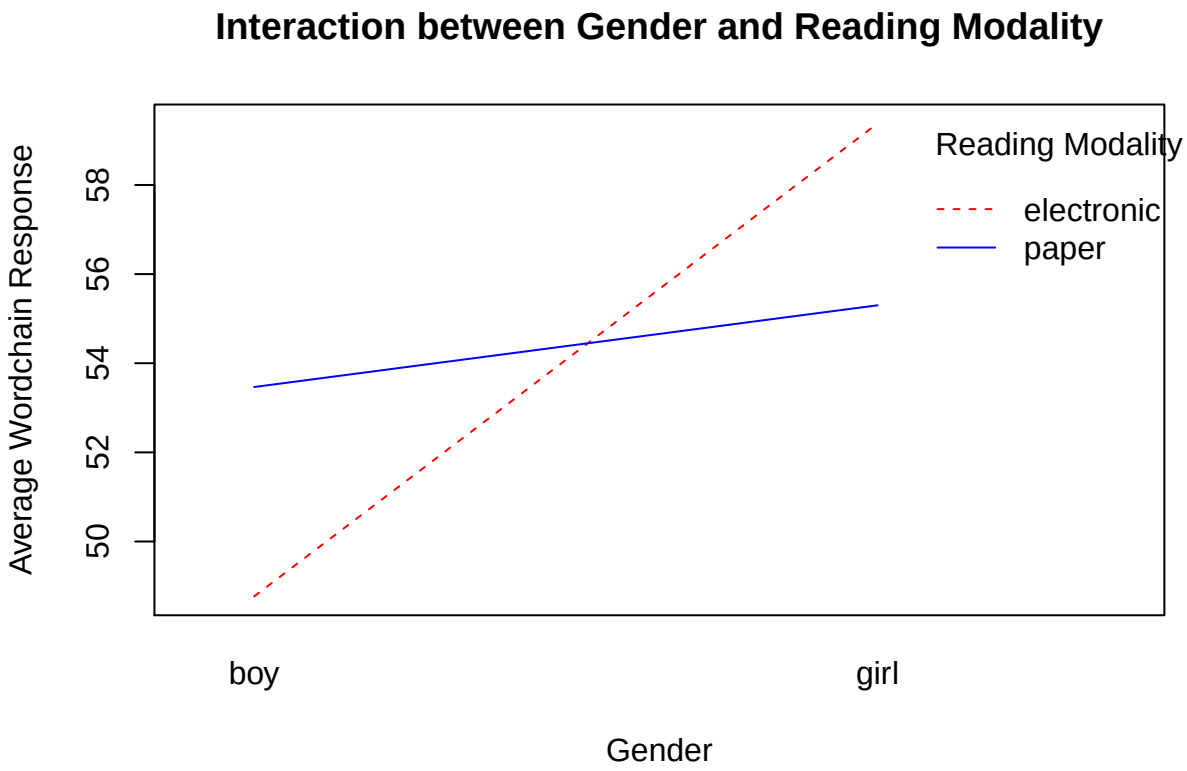
```
##
##  Welch Two Sample t-test
##
## data:  Wordchain by Condition
## t = -0.29371, df = 69.713, p-value = 0.7699
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.370857  3.992133
## sample estimates:
## mean in group electronic      mean in group paper
##                 53.51064                 54.20000
```

```
#Reading Comprehension Pre Test T-test
t.test(Controlsum~Condition, data = data)
```

```
##
##  Welch Two Sample t-test
##
```

```
## data:  Controlsum by Condition
## t = 0.09413, df = 60.972, p-value = 0.9253
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.307851  3.634660
## sample estimates:
## mean in group electronic      mean in group paper
##                  19.7234                  19.5600
```

```
#Interaction Plot
interaction.plot(x.factor = data$Sex, trace.factor = data$Condition,
                 response = data$Wordchain, fun = base::mean, trace.label = "Reading Modality",
                 col = c("red", "blue"),
                 xlab = "Gender", ylab = "Average Wordchain Response",
                 main = "Interaction between Gender and Reading Modality")
```



Our first t-test for vocabulary gives us a pvalue of 0.891 meaning we fail to reject the null hypothesis, indicating there is not a difference in the mean score on the vocabulary test between electronic and paper texts. Same goes for the t-tests for Wordchain and Reading comprehension which found pvalues of 0.7699 and 0.9253 respectively. Therefore, we come to the same conclusion that both indicate there is no difference in the mean amount of wordchain responses and reading comprehension score between participants who used an electronic vs. paper text.

Looking at the above interaction plot, we see some indication of interaction between sex and condition because the lines are not parallel (in fact they cross each other), meaning they do not have equal rates of change.Thus, we must do further analysis to determine whether these interactions are significant.

Question 5:

Using R, redo the MLR and the t-tests listed in tables 1 and 2 on page 65. What would be your final MLR. Summarize your results in an ANOVA table for each MLR in table 2. Check the assumptions.

```r
#Table 2 MLR's
model <- lm(TotalSum_Readingcompr~Vocabulary+Wordchain+Controlsum, data = data)
summary(model)
```

```
##
## Call:
## lm(formula = TotalSum_Readingcompr ~ Vocabulary + Wordchain +
##     Controlsum, data = data)
##
## 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
```

```r
model.2 <- lm(TotalSum_Readingcompr~Vocabulary+Wordchain+Controlsum + Sex, data = data)
summary(model.2)
```

```
##
## Call:
## lm(formula = TotalSum_Readingcompr ~ Vocabulary + Wordchain +
##     Controlsum + Sex, data = data)
##
## 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
```

```r
#Final MLR
model.3 <- lm(TotalSum_Readingcompr~Vocabulary+Wordchain+Controlsum+Sex+Condition, data = data)
summary(model.3)
```

```
##
## Call:
## lm(formula = TotalSum_Readingcompr ~ Vocabulary + Wordchain +
##     Controlsum + Sex + Condition, data = data)
##
## 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
```

```
anova(model, model.2, model.3)
```

```
## 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
```

```
anova(model, model.3)
```

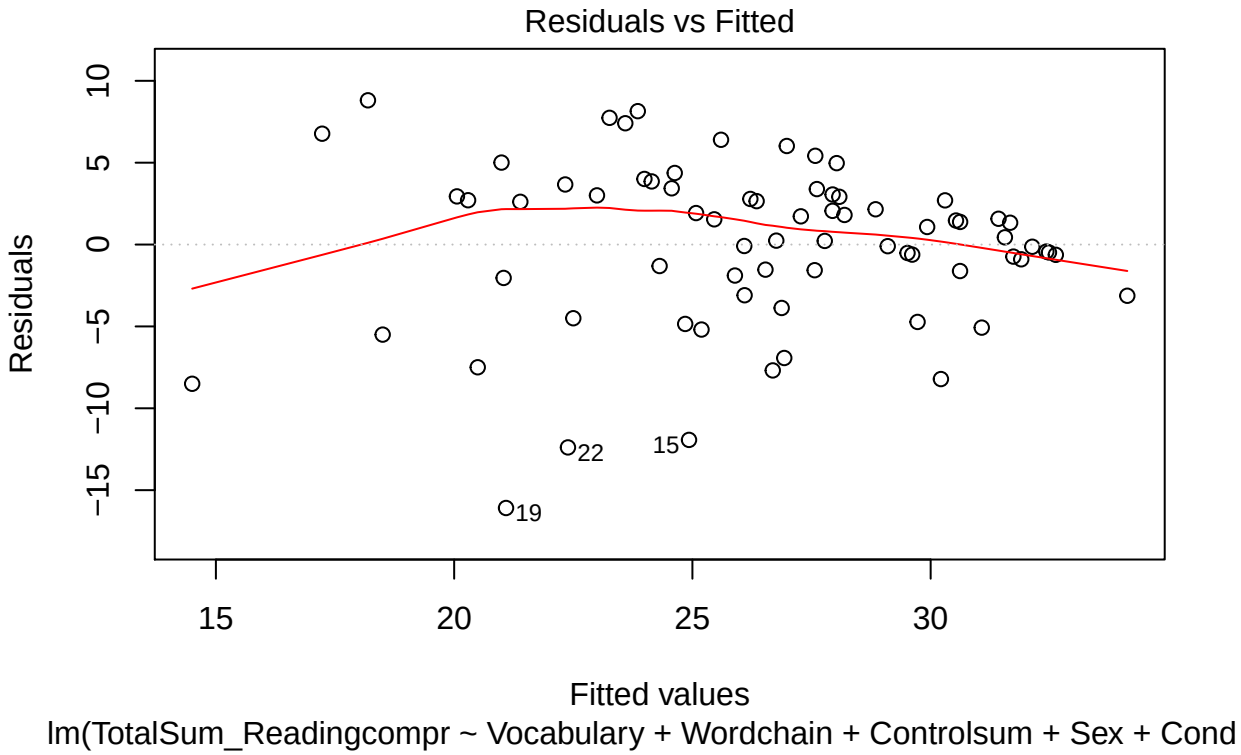
```
## Analysis of Variance Table
##
## Model 1: TotalSum_Readingcompr ~ Vocabulary + Wordchain + Controlsum
## Model 2: TotalSum_Readingcompr ~ Vocabulary + Wordchain + Controlsum +
##     Sex + Condition
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     68 1915.0
## 2     66 1711.1  2    203.84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

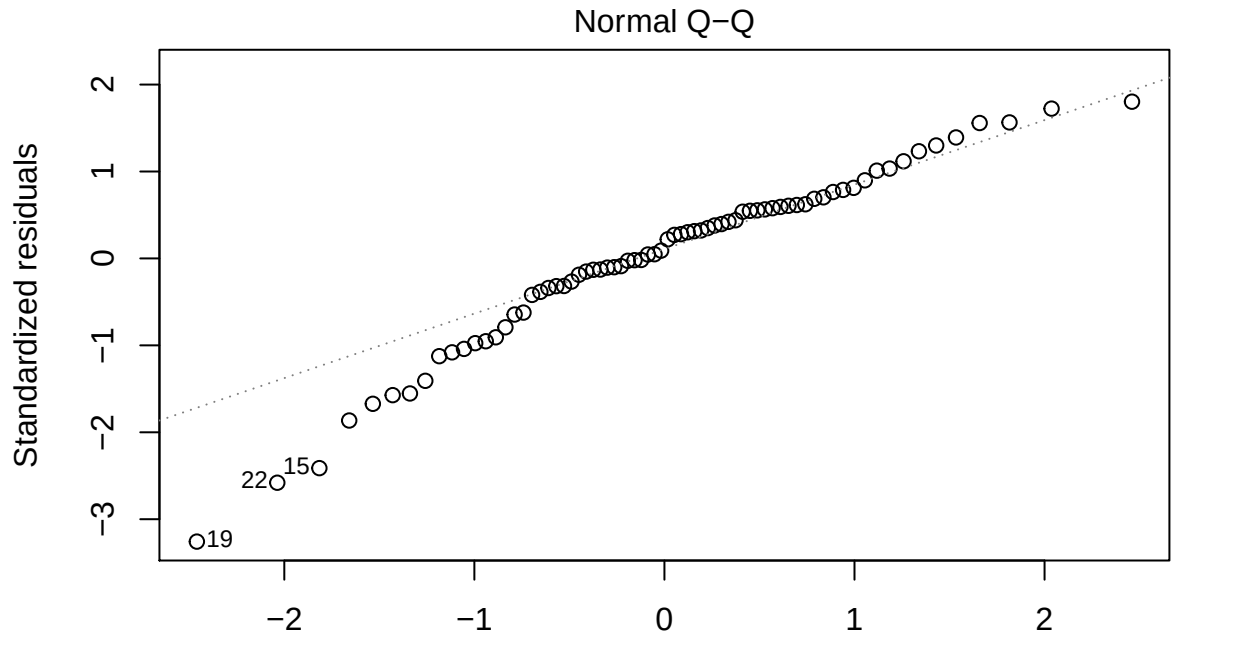
```
anova(model.3)
```

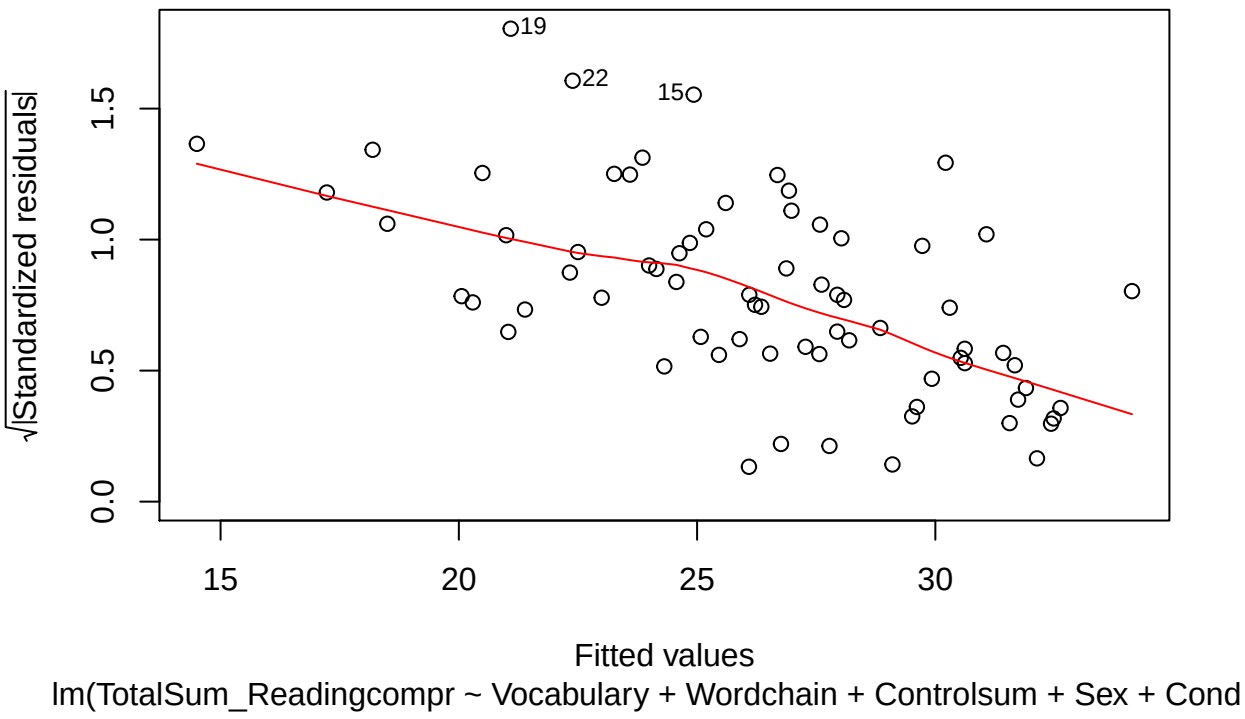
```
## Analysis of Variance Table
```

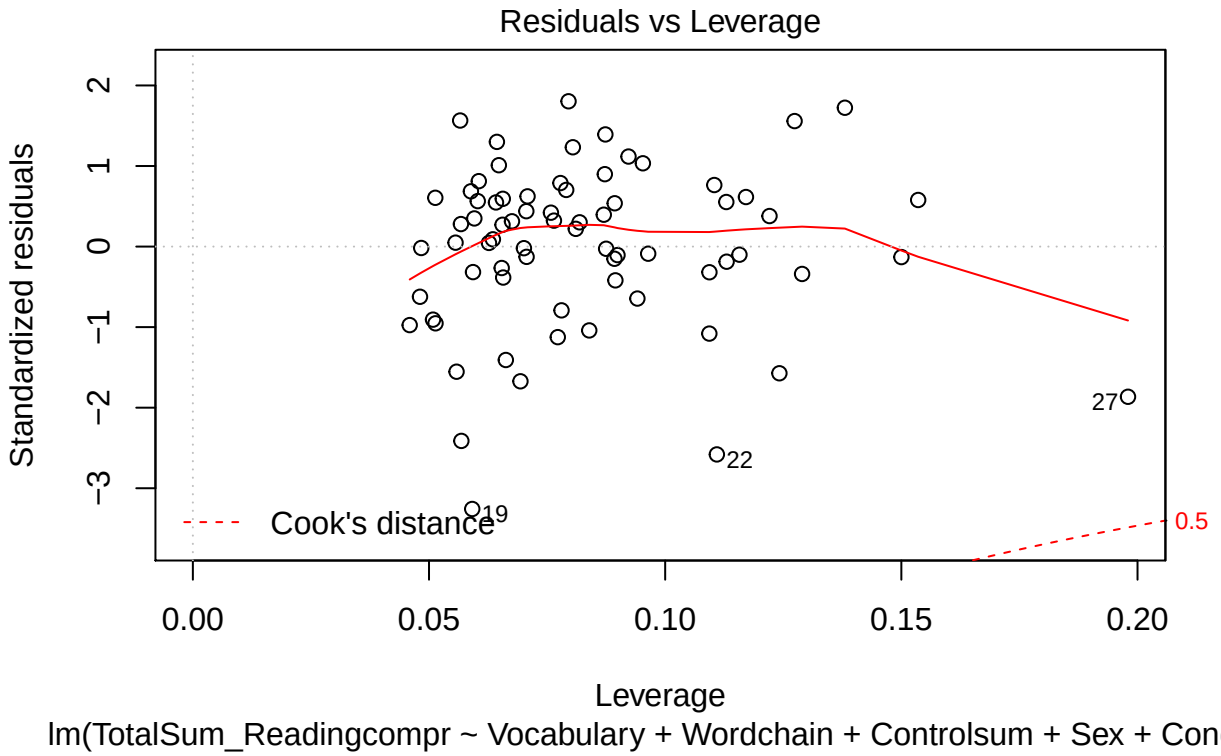
```
##
## Response: TotalSum_Readingcompr
##             Df  Sum Sq Mean Sq F value    Pr(>F)
## Vocabulary   1  537.53  537.53 20.7331 2.332e-05 ***
## Wordchain    1  397.22  397.22 15.3211 0.0002173 ***
## Controlsum   1   81.14   81.14  3.1297 0.0814968 .
## Sex          1   68.06   68.06  2.6251 0.1099520
## Condition    1  135.78  135.78  5.2373 0.0253172 *
## Residuals   66 1711.14   25.93
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(model.3)
```




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


Scale−Location
√|Standardized residuals|
Fitted values
lm(TotalSum_Readingcompr ~ Vocabulary + Wordchain + Controlsum + Sex + Cond
19
22
15



Our final MLR would be the one that included Vocabulary, Word reading, Reading comprehension pretest, Sex, and reading modality as after controlling for the variance associated with with vocabulary, word-reading, reading comprehension, and gender, the reading modality variable accounted for additional variance and was therefore statistically significant. Looking at the ANOVA table for each MLR, we see that the third model is statistically significant when compared to the second model and likewise when compared to model 1, model 3 is still statistically significant. Therefore, as seen by analysis of our anova table, model 3 is our final MLR model with vocabulary, wordchain, being statistically significant variables and condition also being statistically significant given the reading comprehension pretest and sex variables are accounted for in our model. Our final model with the added variables helps to explain more of the variability in the data.

Looking at the diagnostic plots for this model, however, there is a slight backwards fanshape pattern and overall curvature in the residual plot indicating non-constant variance. The normal probability plot is relatively straight, however there is some significant deviation at the tails, indicating that the errors are not normally distributed. The scale-location plot shows a clear decreasing trend, which again confirms that the constant variance assumption is not satisfied. While there are no high leverage or influential points, since there is evidence of non-constant variance and the normality condition is not satisfied, our final model is not valid.