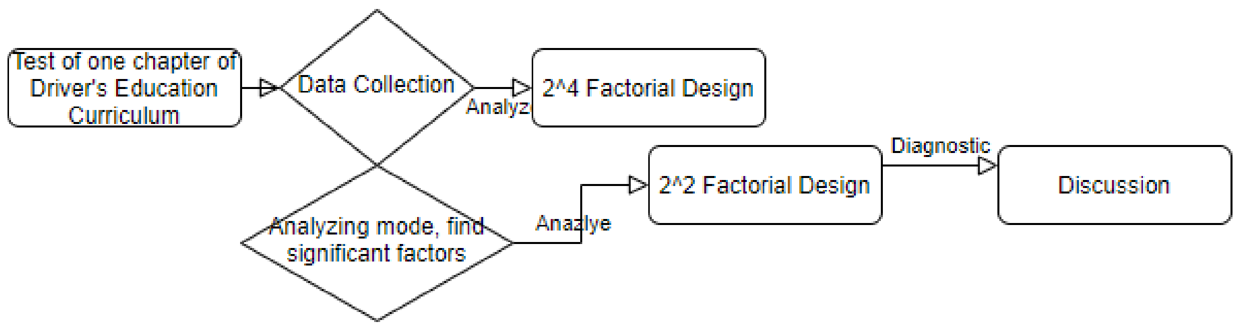
Effect of Study Mode and Strategy on Driver's License Test

Anantha Rao, Jenna Ji, Desmond Fung, Zhiyi Chen

Introduction and motivation

Experiment Overview Diagram:



The problem that we are seeking to understand is the relationship between mode and parameters of learning and test scores. In our particular experiment, we used learning materials from a driver's education curriculum and administered a driver's test to examine the amount of learning. We are interested in what kinds of learning parameters improve test scores because it can offer insight into how students learn best. Done on a much larger scale, our design might shed light on the importance of learning mode, effect of test time, amount of practice and role of prior experience on test scores.

This has been a widely studied phenomenon for decades with a variety of findings. Nist, Simpson, and Hogrebe (1985) suggest that people who used appropriate strategies performed better on content area tests when same strategies were exposed to all the participants, and, in contrast, an increase in negative behaviors was related to lower test performance. This finding indicates that positive and appropriate learning strategies will produce high test scores. More recently, an article published by The Infovore Secrets Editorial (2020) studied which learning materials would result in higher test performance. It concluded that most students had a preference for reading materials, and indeed performed better in tests than those who had video materials.

We are interested, here, in analyzing what learning strategies boost people's driver's test scores. We chose 4 explanatory variables: video/reading, short/long testing time, pre-test/no pre-test, driver license/ no driver license. To measure response, we adapted a quiz from the DMV of New York and will use participant's scores on this quiz. We realize that there are some uncontrollable factors that might affect the response variable; such as the amount of participant driving experience. However, we hope to capture some of this difference with our driver's license factor.

To recap, the objective of this study is to gain an in-depth understanding between mode of learning and test score. We will reaffirm the feasibility of our study by conducting a 2^4 factorial design on different learning strategies and evaluate their performances. Obviously due to COVID, many possible hurdles are expected. One of these difficulties includes data collection. For example, we don't have an effective way to collect a random sample as we will be mostly asking friends to fill out the quiz, thereby introducing bias into our experiment. These are all risks that we need to address as we progress.

Design and methodology

In this study, we investigated the effect of several learning strategies and learning parameters on test performance. Using learning materials provided publicly by the Department of Motor Vehicles in New York, we administered a chapter to study and a subsequent quiz to measure the amount of learning. We decided to settle on these four factors as they proved to be the most feasible variables to study: reading or video study materials (A), a testing time of 5 minutes or 10 minutes (B), a pre-test before studying or no pre-test (C), and whether or not our subjects possessed a driver’s license (D). We evaluated these participants on a content-related quiz to see how the four factors affect quiz score.

To analyze the significance of the four factors introduced, hypothesis testing was performed with the theoretical linear model defined:

$$y = \beta_0 + \beta_i x_i + \beta_{j,k} x_j x_k + \beta_{l,m,n} x_l x_m x_n + \beta_{1,2,3,4} x_1 x_2 x_3 x_4 + \varepsilon \;\; ; \;\; i, j, k, l, m, n \in \{1, 2, 3, 4\}$$

and $j < k$, and $l < m < n$.

Our hypothesis is that the strategies and parameters that allow for more practice or experience will yield higher test scores than the ‘low/off’ level counterparts. We will refer to these parameters as ‘high factor settings’. We define the high factor settings as: reading materials, 10 minutes test time, pre-test, and a driver’s license. When examining our data, we ended up analyzing only the main effects of the factors and thus, have phrased our hypotheses as such. Our formal hypotheses are as follows:

$\mathbf{H_0}$: $\beta_i = 0$, for $i = 1, 2, 3, 4$. vs. $\mathbf{H_1}$: at least one of the $\beta_i \neq 0$, for $i = 1, 2, 3, 4$.

Our hypothesis is an approximation because we aren’t blocking for other variables such as IQ level or ethnicity. These factors could result in larger error variance, hence worse estimate of effects.

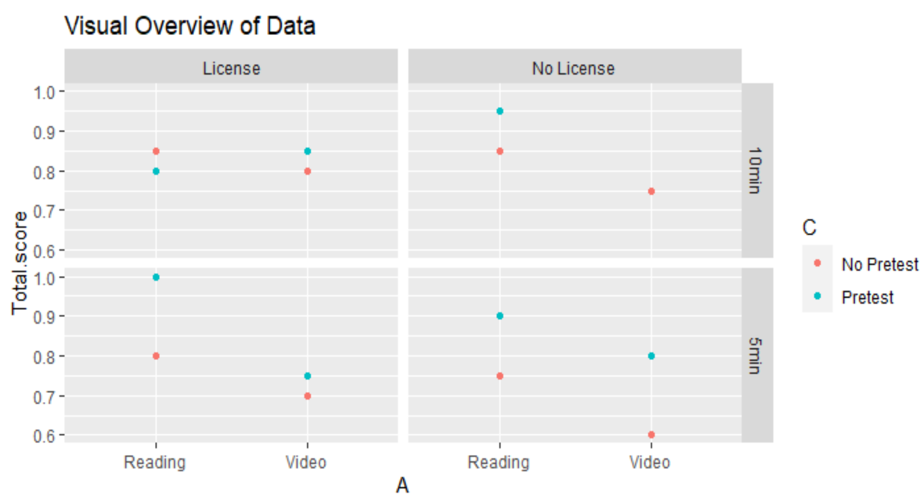
With the 4 two-level factors we mentioned before, there would be 16 participants, 8 of which had driver’s licenses and 8 who did not have a driver's license to do this experiment. Therefore, a 2^4 factorial design will be performed. We will analyze the data by looking for the presence of main effects and interactions between our factors. Normality will be checked using the normal probability plot (2.4), and diagnostics will be performed to evaluate the validity of our models.

Implementation and results

In order to learn the effect of different learning strategies on driver's test scores, we used a 2^4 full factorial model to evaluate the hypotheses. A total of 16 participants were asked to review learning materials for a driver's education course before taking a test. We then sent out a google survey test to evaluate learning. All the data are collected and stored into a table. Each participant was assigned with different combinations of the four factors (A, B, C, D). In order to ensure the fairness of the results, we used the method in which participants have no knowledge of which factor combination they and the others are receiving.

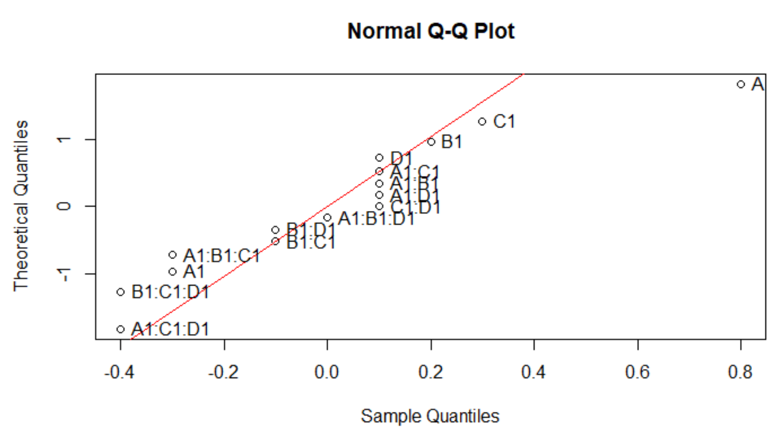
Firstly, after factoring the factors in the dataset and computing the final scores to percentage scales, we ran regressions using lm() function in programming language R. In order to determine the effects of factors, we implemented Daniel Plots, as shown below, to visualize which factors are having significant effects on the test scores.

We compacted all our data in a facet grid. It seems like people without a pre-test generally perform worse than those with a pre-test. Another trend we noticed is that people who were given reading material tend to have a higher test score than those given a video.



Further analysis was made to statistically evaluate these observations below.

In the Daniel Plot, we again see that there may be some interesting main effects with effect A and C (video



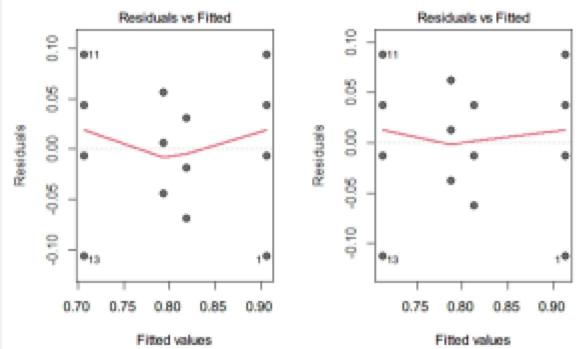
and pre-test). We also see that there appears to be some notable interactions, however we will forego all of these because some of the main effects that comprise them are themselves close to this red normal line. Thus, we invoke the sparsity of effects principle to comfortably dismiss high order effects, and we invoke heritability to dismiss those interactions whose main effects aren't important. We continue analysis with factors A and C.

After projecting out the insignificant factors, B and D, we're left with a 2^2 factorial design where each combination has 4 replicates. We settled on the regression model of $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$, where x_1 represents the study material (reading or video), x_2 represents pre-tests or not, and the β's are the regression coefficients. After summary() and anova(), we got $y = 0.82 - 0.11x_1 + 0.09x_2$. The p values for factors reading/video and pretest/no-pretest are 0.005 and 0.0213. Therefore, at the level of α=0.05, there is significant evidence to conclude that factors reading/video and pretest/no-pretest are significant.

People who were given reading materials tended to have higher test scores than those given a video. We tentatively theorize that this may have to do with people being more concentrated when reading than watching a video. This result seems to be in agreement with other recent work on the matter. Pre-test seems to improve scores less so, but still, those given a pre-test generally had higher scores than those without pre-tests. This may be because, the pretest primed test takers about the content, or that people who take the tests with pre-tests are more familiar with the form of the tests and thus more prepared. Again, these possible explanations should be examined to see which ones hold up in future work.

Discussion

In conclusion, the result of the study rejects the null hypothesis at the level of 0.05 and supports the idea that study mode and strategy can affect the drivers' license test results. More specifically, the study finds that both reading/video and pretest/no-pretest have significant effects. People who received reading materials tended to perform better on the test than people who received video materials, and people who were given pretests as practice also tended to have higher test scores. Therefore, since the alternative hypothesis is supported by the analysis, the results, in general, support the assumption that study mode and strategies affect test results.



The left residual plot is based on the linear model $y = 0.82 - 0.11x_1 + 0.09x_2$ (lm(y~A+C)). Based on the red contour line, the variances of the error terms are equal. The residuals vary a little around the horizontal band. In the future, we may consider adapting our original model to improve the model's assumption of linearity. We transformed the model by adding $(x_1 + x_2)^2$ as another predictor to (lm(y~(A+C)+(A+C)^2)), and the right residual plot based on the fixed model forms a straighter line; conforming more to the assumption of linearity.

Overall, our results are not representative with our limited sample size, but still, the results are fair and have promising effects. Examining our study, it is important to have more fine tune control over the learning materials and tests that we use. While, using publicly available materials makes it easier to attain ready-made stimuli, we have less control- in particular- on the modes of learning factor(reading vs video). This includes a different script for each, and no choices about how the information is displayed. Future research regarding mode of study, ought to leverage more control on the production of these materials to examine in more detail how different kinds of videos and readings can differ in their effectiveness.

Reference

Nist, S. L., Simpson, M. L., & Hogrebe, M. C. (1985). The Relationship between the Use of Study Strategies and Test Performance. *Journal of Reading Behavior, 17*(1), 15-28. doi:10.1080/10862968509547528

Reading vs Watching Videos: What Science Says. (2020, July 13). Retrieved December 12, 2020, from https://infovoresecrets.com/reading-vs-watching-videos-what-science-says/

Appendix

```
library(tidyverse)
```

```
## -- Attaching packages ------------------------------------------------------------------------

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

## -- Conflicts ---------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
df <- read.csv("C:/Users/q8357/OneDrive/Desktop/FINAL_DATA.csv")
df
```

```
##    ï..Name_of_test_taker Video Test_Time Pre_Test. Drivers_License. Total.score
## 1                   Jack    -1         1         1                1  16.00 / 20
## 2                   Lisa     1        -1         1               -1  16.00 / 20
## 3                   Joel    -1        -1         1               -1  18.00 / 20
## 4                 Daniel     1        -1         1                1  15.00 / 20
## 5                   Jose     1         1         1                1  17.00 / 20
## 6                 Parker    -1        -1         1                1  20.00 / 20
## 7                   Andy     1         1         1               -1  15.00 / 20
## 8                Jessica    -1         1         1               -1  19.00 / 20
## 9                 Josiah    -1        -1        -1                1  16.00 / 20
## 10                 Lynda     1        -1        -1                1  14.00 / 20
## 11                Summit     1         1        -1                1  16.00 / 20
## 12               Zackery    -1         1        -1                1  17.00 / 20
## 13                 Quinn     1        -1        -1               -1   12.00 /20
## 14                  Alex    -1        -1        -1               -1  15.00 / 20
## 15                  Will    -1         1        -1               -1  17.00 / 20
## 16                 Lucas     1         1        -1               -1  15.00 / 20
```

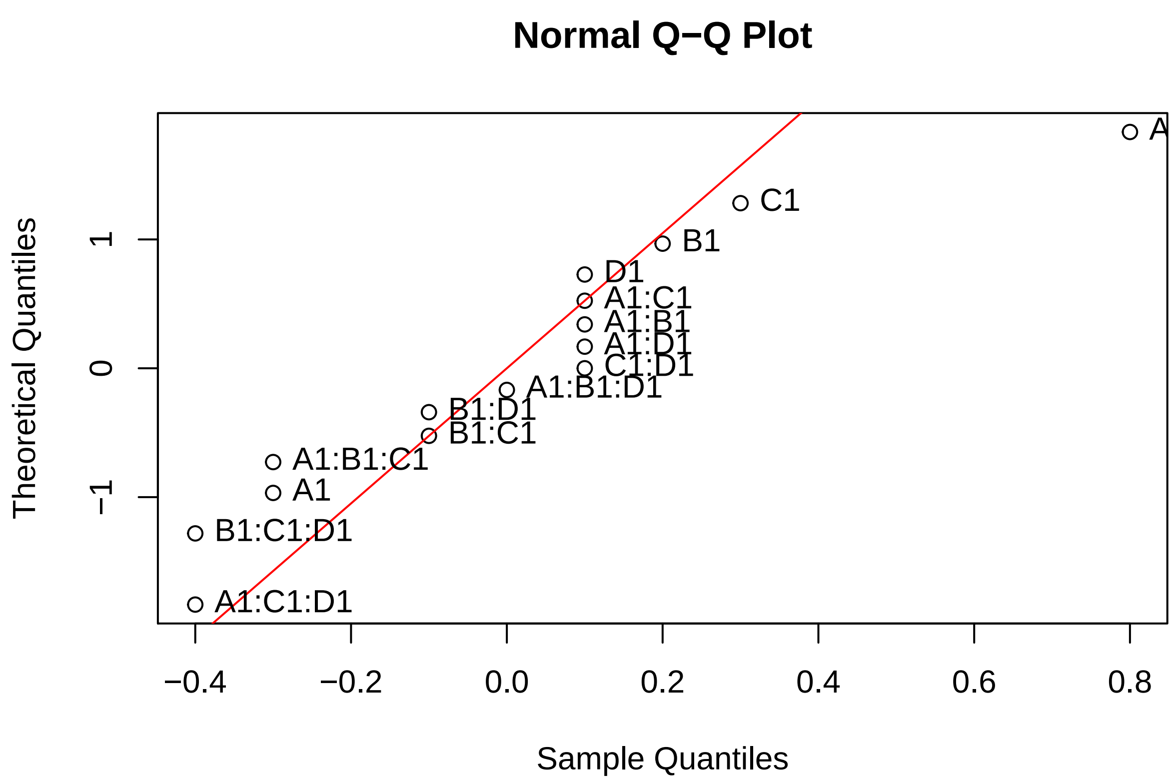
```
daniel_plot <- function(effects) {
  qq <- qqnorm(effects, datax = TRUE)
  qqline(effects, col = "red", probs = c(0.3, 0.7), datax = TRUE)
  text(qq$x, qq$y, names(effects), pos=4)
}

# building data frame
df$Total.score <- as.numeric(str_sub(df$Total.score,1,2)) / 20
df<- df%>% rename(A=Video, B=Test_Time, C=Pre_Test., D=Drivers_License.)
df$A <- as.factor(df$A)
```

```
df$B <- as.factor(df$B)
df$C <- as.factor(df$C)
df$D <- as.factor(df$D)

# building model (2^4)
fit_coded <- lm(Total.score~A*B*C*D, data=df)
effects <- 2 * coef(fit_coded)[-1]

# daniel plot + test
daniel_plot(effects)
```




```
summary(fit_coded)
```

```
##
## Call:
## lm(formula = Total.score ~ A * B * C * D, data = df)
##
## Residuals:
## ALL 16 residuals are 0: no residual degrees of freedom!
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.500e-01         NA      NA       NA
## A1          -1.500e-01         NA      NA       NA
## B1           1.000e-01         NA      NA       NA
```

```
## C1           1.500e-01         NA      NA       NA
## D1           5.000e-02         NA      NA       NA
## A1:B1        5.000e-02         NA      NA       NA
## A1:C1        5.000e-02         NA      NA       NA
## B1:C1       -5.000e-02         NA      NA       NA
## A1:D1        5.000e-02         NA      NA       NA
## B1:D1       -5.000e-02         NA      NA       NA
## C1:D1        5.000e-02         NA      NA       NA
## A1:B1:C1    -1.500e-01         NA      NA       NA
## A1:B1:D1     4.663e-16         NA      NA       NA
## A1:C1:D1    -2.000e-01         NA      NA       NA
## B1:C1:D1    -2.000e-01         NA      NA       NA
## A1:B1:C1:D1  4.000e-01         NA      NA       NA
##
## Residual standard error: NaN on 0 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:    NaN
## F-statistic:   NaN on 15 and 0 DF,  p-value: NA
```

```r
anova(fit_coded)
```

```
## Warning in anova.lm(fit_coded): ANOVA F-tests on an essentially perfect fit are
## unreliable
```

```
## Analysis of Variance Table
##
## Response: Total.score
##           Df   Sum Sq  Mean Sq F value Pr(>F)
## A          1 0.050625 0.050625
## B          1 0.005625 0.005625
## C          1 0.030625 0.030625
## D          1 0.002500 0.002500
## A:B        1 0.005625 0.005625
## A:C        1 0.000625 0.000625
## B:C        1 0.015625 0.015625
## A:D        1 0.002500 0.002500
## B:D        1 0.002500 0.002500
## C:D        1 0.002500 0.002500
## A:B:C      1 0.000625 0.000625
## A:B:D      1 0.010000 0.010000
## A:C:D      1 0.000000 0.000000
## B:C:D      1 0.000000 0.000000
## A:B:C:D    1 0.010000 0.010000
## Residuals  0 0.000000
```

```r
# final model (2^2)
projected_fit <- lm(Total.score~A+C, data=df)
summary(projected_fit)
```

```
##
## Call:
## lm(formula = Total.score ~ A + C, data = df)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.10625 -0.04375  0.00000  0.04375  0.09375
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.81875    0.02895  28.277 4.63e-13 ***
## A1          -0.11250    0.03343  -3.365  0.00507 **
## C1           0.08750    0.03343   2.617  0.02130 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06687 on 13 degrees of freedom
## Multiple R-squared:  0.583,  Adjusted R-squared:  0.5188
## F-statistic: 9.086 on 2 and 13 DF,  p-value: 0.003397
```

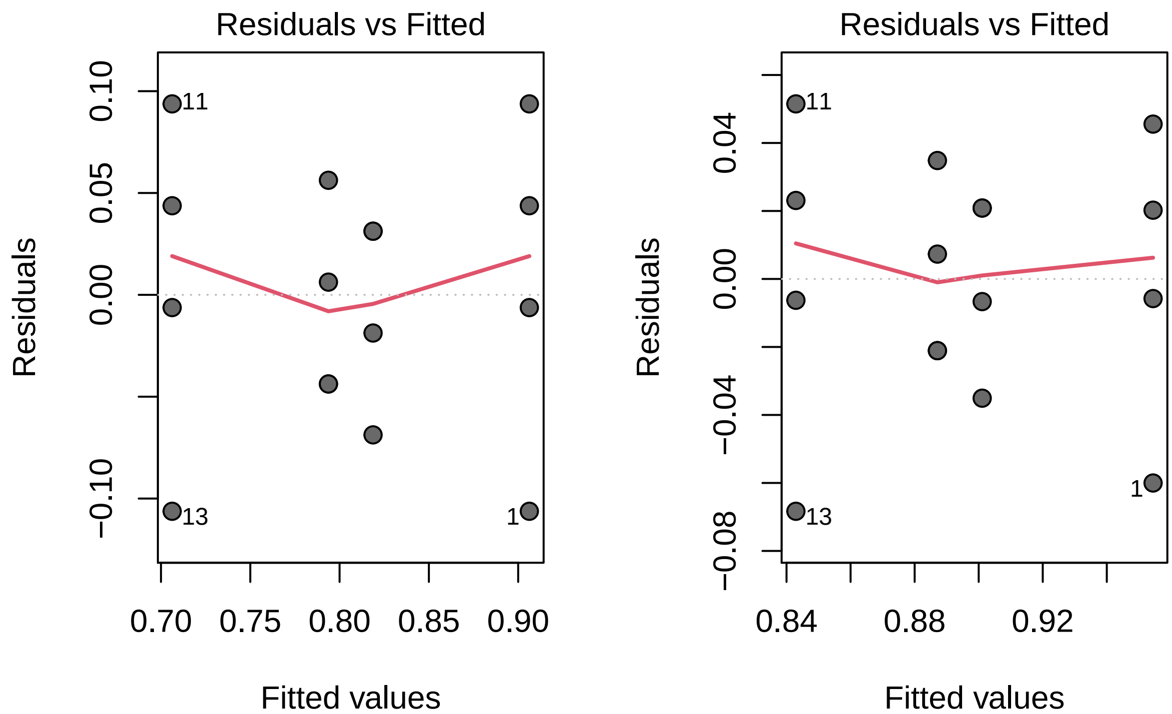
```r
anova(projected_fit)
```

```
## Analysis of Variance Table
##
## Response: Total.score
##           Df   Sum Sq  Mean Sq F value   Pr(>F)
## A          1 0.050625 0.050625 11.3226 0.005073 **
## C          1 0.030625 0.030625  6.8495 0.021301 *
## Residuals 13 0.058125 0.004471
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# residual plot + revised residual plot
par(mfrow=c(1,2))
plot(projected_fit, which=1, pch=21, bg="dimgray", cex=1.2, lwd=2)
fit_coded_0 <- lm(sqrt(Total.score)~(A+C)+(A+C)^2, data=df)
anova(fit_coded_0)
```

```
## Analysis of Variance Table
##
## Response: sqrt(Total.score)
##           Df    Sum Sq   Mean Sq F value   Pr(>F)
## A          1 0.0157509 0.0157509 10.3197 0.007459 **
## C          1 0.0095136 0.0095136  6.2332 0.028088 *
## A:C        1 0.0000842 0.0000842  0.0552 0.818260
## Residuals 12 0.0183155 0.0015263
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
plot(fit_coded_0, which=1, pch=21, bg="dimgray", cex=1.2, lwd=2)
```


Residuals vs Fitted
Residuals
Fitted values
0.10
0.05
0.00
−0.10
0.70
0.75
0.80
0.85
0.90
11
13
1
Residuals vs Fitted
Residuals
Fitted values
0.04
0.00
−0.04
−0.08
0.84
0.88
0.92
11
13
1