IST772 Problem Set 9

Abhijith Anil Vamadev

The homework for week 10 is based on exercises 2-4, 6, 8 and 9 on pages 209-210 but with changes as noted in this notebook (i.e., follow the problems as given in this document and not the textbook).

Attribution statement: (choose only one) 1. I did this homework by myself, with help from the book and the professor

Chapter 9, Exercise 2

Using the built in dataset CO2, use the interaction.plot() command to display a means plot of the "uptake" variable, using "Type" and "Treatment" as the factors. Interpret the results: Without knowing any of the statistics, would you guess that there is or is not an interaction between Type and Treatment? (1 pt)

```
library("HSAUR")
```

```
## Loading required package: tools
```

```
data <- CO2 #variable
interaction.plot(x.factor=CO2$Type,trace.factor=CO2$Treatment,response=CO2$uptake)
```
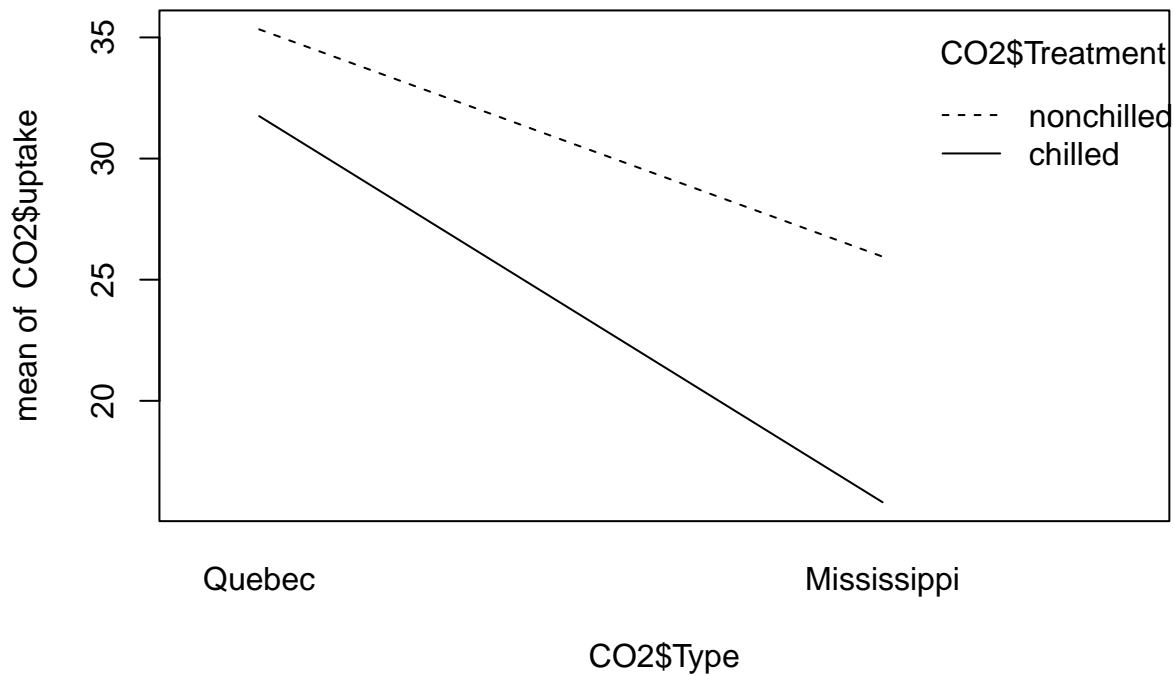
```
#interaction plot
```

- Y axis represents the uptake of Co2. The x-axis has two types of origns of the plant, Quebec and Missiippi. And there are two lines representing the two treatmnet types, one representing nonchilled and other one being chilled.
- None of the intermediate points implied by the lines really exist.
- The two lines are not entirely parallel to each other indicating there might be an interaction between the two variables.
- That suggests the possibility that the effect of CO2 Type on uptake may be dependent upon whether the treatment type is chilled or non-chilled.

Chapter 9, Exercise 3

Use aov() to produce significance tests of the main effects of Type and Treatment and the interaction between Type and Treatment using uptake as the dependent variable. (1 pt) Make sure to state each null hypothesis and then use the correct language to describe the outcomes of the significance tests. Make sure to report "omnibus" statistics as well. (1 pt)

```
aovOut = aov(uptake ~ Type + Treatment + Type:Treatment, data=CO2) #aov
summary(aovOut) #summary
```

```
##               Df Sum Sq Mean Sq F value   Pr(>F)
## Type           1   3366    3366  52.509 2.38e-10 ***
```

```
## Treatment          1    988     988  15.416 0.000182 ***
## Type:Treatment  1    226     226   3.522 0.064213 .
## Residuals        80   5128      64
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- $F(1, 80) = 3.552$, for Type:treatment is not statistically significant at the conventional alpha threshold, as the p-value 0.06, we fail to reject the null hypothesis. There is no statistically significant interaction.

- $F(1, 80) = 52.509$, for the type variable,is statistically significant at the conventional alpha threshold, as $p(2.38\text{e-}10) < 0.05$. There is significant effect.

- $F(1, 80) = 15.416$, for the treatment variable,is statistically significant at the conventional alpha threshold, as $p(0.00018) < 0.05$. There is significant effect.

Chapter 9, Exercise 4

Use anovaBF() to examine the main effects and interaction as described in Exercise 3. Interpret the results in your own words. Contrast the results from the traditional ANOVA analysis with the results of the Bayes Factor ANOVA. (1 pt) Use the output object from the anovaBF() procedure to create an odds ratio that compares the complete model (with the interaction) against a main effects-only model. (1 pt)

Important note: The anovaBF() function is picky about how the data sets it analyzes are stored. Use newCO2 <- data.frame(CO2) to convert the CO2 data set to a data frame. Then conduct your analysis on newCO2.

```
library(BayesFactor)
```

```
## Loading required package: coda
```

```
## Loading required package: Matrix
```

```
## ************
## Welcome to BayesFactor 0.9.12-4.2. If you have questions, please contact Richard Morey (richarddmorey
##
## Type BFManual() to open the manual.
## ************
```

```
newCO2 <- data.frame(CO2) #data farme
aovOut3 = anovaBF(uptake ~ Type*Treatment, data=newCO2) #anova BF
aovOut3
```

```
## Bayes factor analysis
## --------------
## [1] Type                               : 2366422    ±0%
## [2] Treatment                          : 11.62652   ±0%
## [3] Type + Treatment                   : 235996150 ±1.09%
## [4] Type + Treatment + Type:Treatment : 296011205 ±1.06%
##
## Against denominator:
##   Intercept only
## ---
## Bayes factor type: BFlinearModel, JZS
```

```
aovOut3[4]/aovOut3[3] # What's the odds ratio of model 4 vs. model 3?
```

```
## Bayes factor analysis
## --------------
## [1] Type + Treatment + Type:Treatment : 1.254305 ±1.52%
##
## Against denominator:
##   uptake ~ Type + Treatment
## ---
## Bayes factor type: BFlinearModel, JZS
```

- These results confirm our findings from earlier. Type has a Bayes factor odds of 2366422 to 1 , which is heavily in favor of the alternative hypothesis, that there is an effect.

- Treatment is has a Bayes factor odds of 11.62 to 1 and is strongly in favor of the alternative, hypothesis, that there is an effect.

- The main effects only model has a Bayes factor of 238426524 to 1is also heavily in favor of the alternative hypothesis, that there is significant interaction.

- The full model containing the interaction term and the main effects, has an odds ratio of 295861285 to 1 which, is also extremely in favor of a model that combines main effects and interaction effects.

- The comparing of the interaction model versus the main effects only model, the odds show 1.27 to 1 in favor of the model that includes the interaction term. This is a very weak result, barely worth mentioning.

Chapter 9, Exercise 6

This exercise uses the USJudgeRatings dataset, contains a set of 43 ratings of the competence of state judges in the U.S. Superior Court (from the 1970s). Using similar code as that which appeared in the regression interactions section of this chapter, plot regression lines showing PREP (on the X-axis) and RTEN (on the Y-axis). Plot two regression lines, one for the half of the data set where FAMI is below its median value and one where FAMI is at or above its median value. Without knowing any of the statistics, do you think it likely that there is an interaction between FAMI and PREP? Explain why or why not in your own words. (1 pt)

```
df <- USJudgeRatings #data farame
plot(df$PREP,df$RTEN) #plot
hiFAMI <- subset(df, FAMI > median(df$FAMI))#subsetting
hiLmOut <- lm(PREP ~ RTEN,data=hiFAMI) #linear regreession model
abline(hiLmOut,col="blue") #color and line

lowFAMI <- subset(df, FAMI < median(df$FAMI)) #subsetting
lowLmOut <- lm(PREP ~ RTEN,data=lowFAMI)#linear regression model
abline(lowLmOut,col="red")#color and line
```
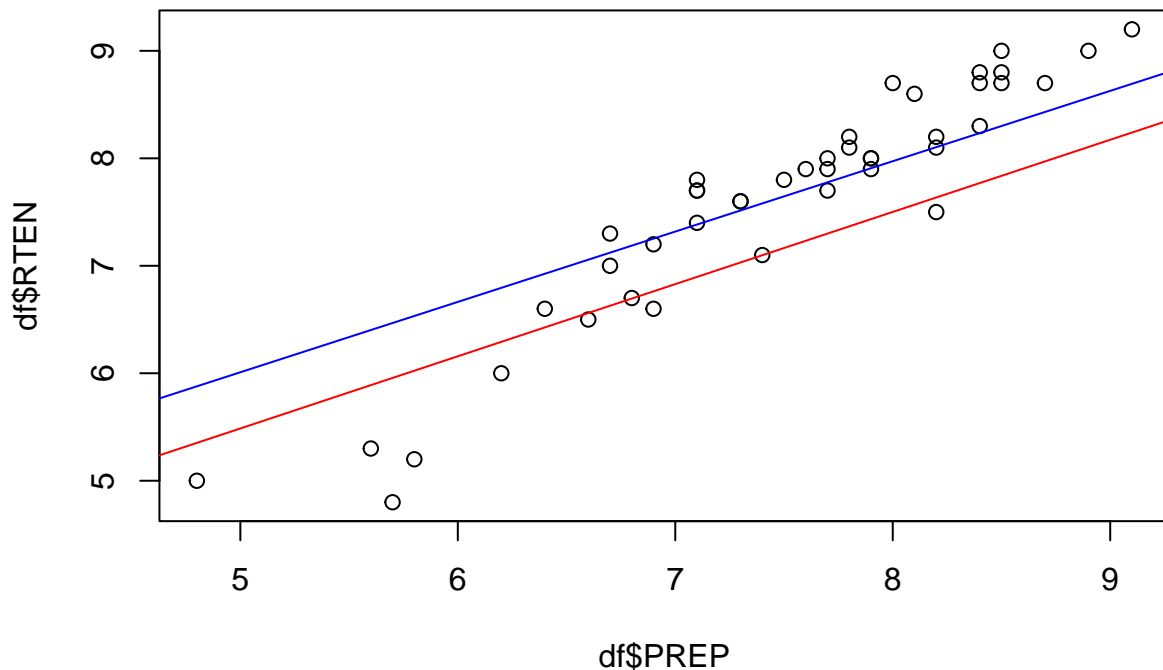
* X-axis represents PREP while Y-axis represents the RTEN, and the data points are FAMI points that are higher and lower than the median. The two lines represent at the lines is the line of best fit for the low and high FAMI (less than and greater than the median of FAMI) * It is likely there is no interaction between PREP and RTEN, so there two linear regression models, are parallel. * That suggests the possibility that the effect of PREP on RTEN may not be dependent upon whether there is high Familiarity with law or low Familiarity with law.
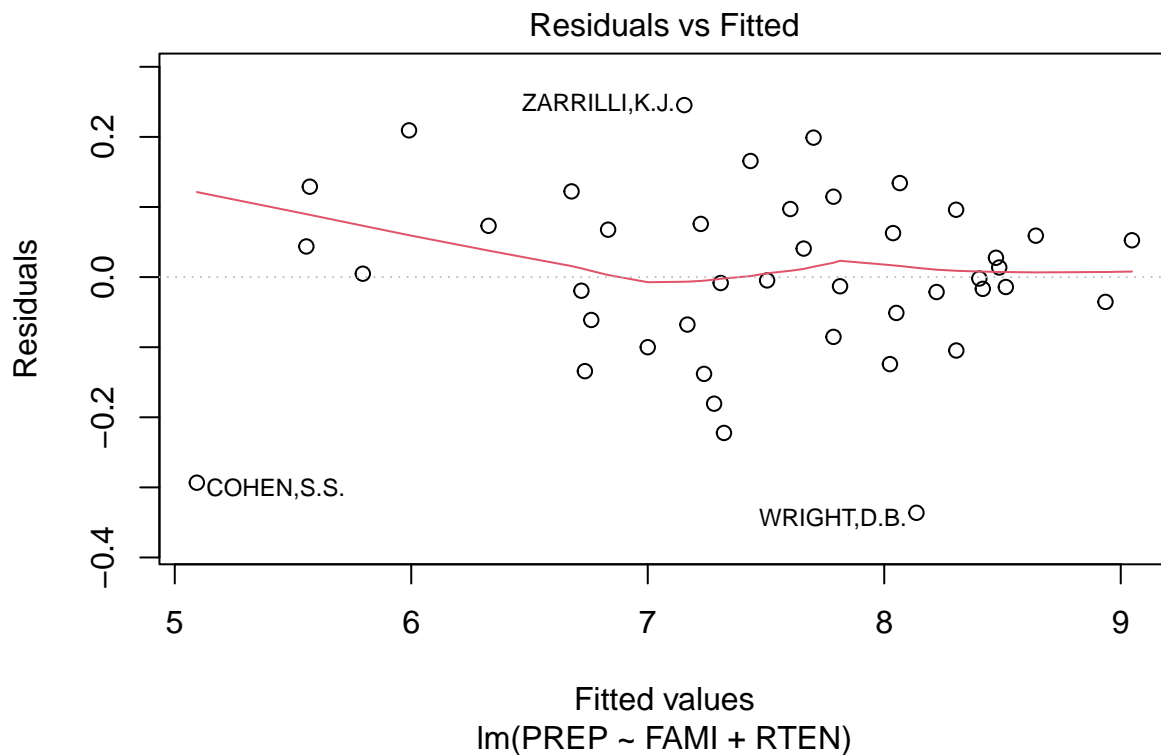
Chapter 9, Exercise 8

Conduct a regression analysis of the linear and interactive effects of PREP and FAMI on RTEN using the lm() function. Be sure to centre the variables first. (1 pt) Interpret the results in your own words, making sure to report the outcomes of the significance tests. (1 pt)
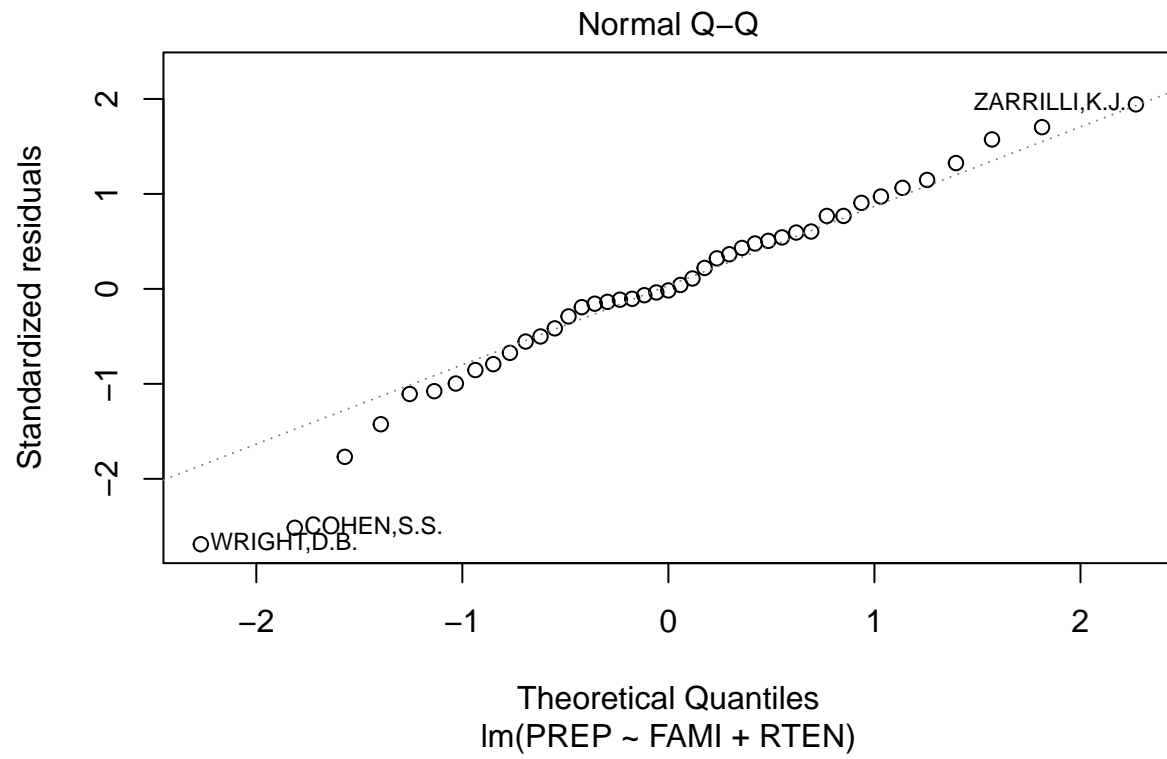
```r
reg <- lm(PREP~FAMI+RTEN, data = df) #linear regression
summary(reg)#summary
```
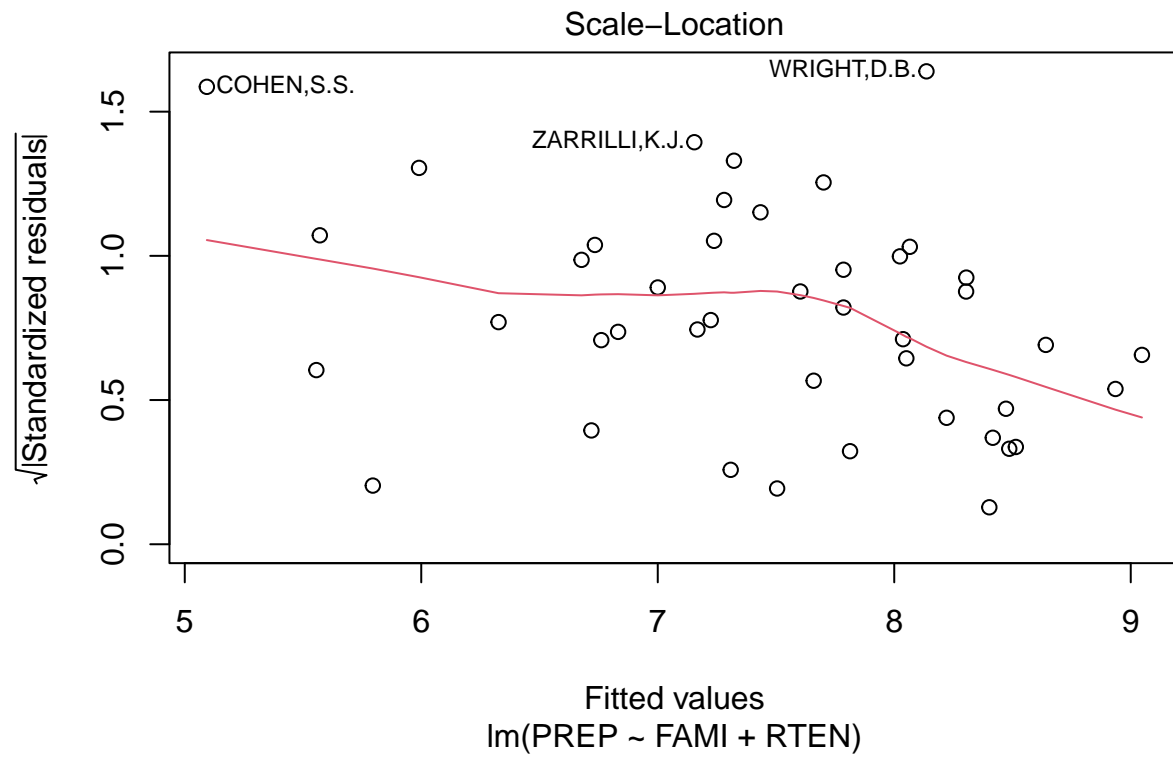
```
##
## Call:
## lm(formula = PREP ~ FAMI + RTEN, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.33621 -0.06444 -0.00205  0.07451  0.24537
##
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.10115    0.16045   0.630   0.5320
## FAMI         0.84254    0.06198  13.594   <2e-16 ***
## RTEN         0.13904    0.05342   2.603   0.0129 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1283 on 40 degrees of freedom
## Multiple R-squared:  0.9828, Adjusted R-squared:  0.9819
## F-statistic:  1139 on 2 and 40 DF,  p-value: < 2.2e-16
```
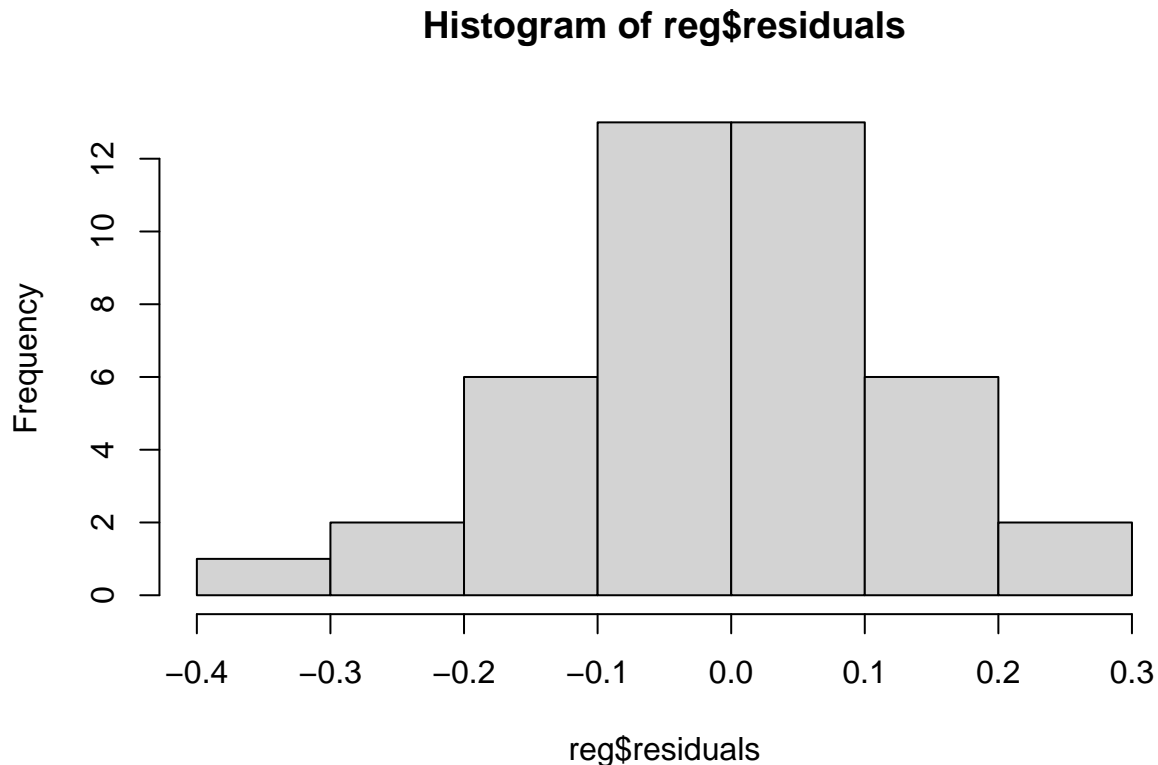
```
plot(reg, which = 1:3) #plots
```

Normal Q–Q

Standardized residuals

ZARRILLI,K.J.

COHEN,S.S.
WRIGHT,D.B.

Theoretical Quantiles
lm(PREP ~ FAMI + RTEN)

7

Scale–Location

Fitted values
lm(PREP ~ FAMI + RTEN)

```
hist(reg$residuals) #residual histogram
```

Histogram of reg$residuals

reg$residuals

* Looking at the graph and data, the residuals appear to be distributed without any skewness and looks to be normally distrusted. There is no Heteroscedasticity either, looking at the plots, as it isn't in a cone shape. * The intercept has an estimate of 0.1011 with a t-value of 0.630 and an associated p-value of 0.532 which is higher than the p-value, making it fail to reject the null hypothesis. * The FAMI has an estimate of 0.842 with a t-value of 13.594 and a p-value of 2e-16 which is much smaller than the alpha, we reject the null hypothesis, that each estimated coefficient is equal to zero. * The RTEN has an estimate of 0.139 with a t-value of 2.63 and a p-value of 0.0129 which issmaller than the alpha, we reject the null hypothesis, that each estimated coefficient is equal to zero. * The r-squared is 0.982 and the adjusted r-squared being 0.981 which is a very high indication that the variable PREP can be well predicted by FAMI and RTEN. * The f-statistic is 1139 with df of 2 and 40. The p-value associated with it is 2.2e-16 which is less than the alpha so we reject that the null hypothesis that R-squared is equal to 0, and shows that the test is statistically significant.

Chapter 9, Exercise 9

Repeat Exercise 8 using lmBF() to conduct one regression analysis that includes the interaction term and one that does not. (1 pt) Interpret the results. Make sure you report an odds ratio for the comparison of a model with linear effects versus a model with linear and interaction effects. Interpret the results in your own words. (1 pt)

```
lmOutBayes1 <- lmBF(PREP~FAMI+RTEN,data=df) #linear regression BF
lmOutBayes2 <- lmBF(PREP~FAMI+RTEN + FAMI:RTEN,data=df)  #linear regresion BF
lmOutBayes1
```

```
## Bayes factor analysis
```

```
## --------------
## [1] FAMI + RTEN : 2.311798e+32 ±0%
##
## Against denominator:
##   Intercept only
## ---
## Bayes factor type: BFlinearModel, JZS
```

lmOutBayes2

```
## Bayes factor analysis
## --------------
## [1] FAMI + RTEN + FAMI:RTEN : 8.214559e+30 ±0%
##
## Against denominator:
##   Intercept only
## ---
## Bayes factor type: BFlinearModel, JZS
```

- The Bayes factor analsysis for the data, odds ratio of 2.3117e+32 to 1, This shows that the odds are overwhelmingly in favor of the alternative hypothesis, in the sense that the model containing FAMI and RTEN as predictors is hugely favored over a model means that only contains the y-intercept.
- The bayes factor analysis for the model containing the interaction term, has an odds ratio of 8.21e+32 to 1, which is heavily in favor of the alternative hypothesis.

lmOutBayes2/lmOutBayes1 #ratio

```
## Bayes factor analysis
## --------------
## [1] FAMI + RTEN + FAMI:RTEN : 0.0355332 ±0%
##
## Against denominator:
##   PREP ~ FAMI + RTEN
## ---
## Bayes factor type: BFlinearModel, JZS
```

- An odds ratio for the comparison of a model with linear effects versus a model with linear and interaction effects, has an odds ratio of 0.03 to 1, which is very low, not in favor of the alternative hypothesis, and should not be mentioned.