**Quantitative Methods**

**List of Exercises N. 9**

**Selected Exercises from McClave (2014) – Chapter 9**

**9.1 Elements of a Designed Experiment**

**Exercise 1. (5). *Identifying the type of experiment*. Brief descriptions of a number of experiments are given next. Determine whether each is designed or observational and explain your reasoning.**

**a) An economist obtains the unemployment rate and gross state product for a sample of states over the past 10 years, with the objective of examining the relationship between the unemployment rate and the gross state product by census region.**

This is an observational experiment. The economist has no control over the factor levels or unemployment rates.

**b) A manager in a paper production facility installs one of three incentive programs in each of nine plants to determine the effect of each program on productivity.**

This is a designed experiment. The manager chooses only three different incentive programs to compare, and randomly assigns an incentive program to each of nine plants.

**c) A marketer of personal computers runs ads in each of four national publications for one quarter and keeps track of the number of sales that are attributable to each publication’s ad.**

This is an observational experiment. Even though the marketer chooses the publication, he has control over who responds to the ads.

**d) An electric utility engages a consultant to monitor the discharge from its smokestack on a monthly basis over a 1- year period to relate the level of sulfur dioxide in the discharge to the load on the facility’s generators.**

This is an observational experiment. The load on the facility’s generators is only observed, not controlled.

**e) Intrastate trucking rates are compared before and after governmental deregulation of prices changed, with the comparison also taking into account distance of haul, goods hauled, and the price of diesel fuel.**

This is an observational experiment. One has no control over the distance of the haul, the goods hauled, or the price of diesel fuel.

**9.2 The Completely Randomized Design: Single Factor**

**Exercise 2. (28, ADREC). *Study of recall of TV commercials*. Do TV shows with violence and sex impair memory for commercials? To answer this question, IOWA State researchers conducted a designed experiment in which 324 adults were randomly assigned to one of three viewer groups of 108 participants each (*Journal of Applied Psychology*, June 2002). One group watched a TV program with a violent content code (V) rating, the second group viewed a show with a sex content code (S) rating, and the last group watched a neutral TV program with neither a V nor a S rating. Nine commercials were embedded into each TV show. After viewing the program, each participant was scored on his or her recall of the brand names in the commercial messages, with scores ranging from 0 (no brands recalled) to 9 (all brands recalled). The data (simulated from information provided in the article) are saved in the ADREC file. The researchers compared the mean recall scores of the three viewing groups with an analysis of variance for a completely randomized design. Summary for their results are given in the table below.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **One-way ANOVA: VIOLENT, SEX, NEUTRAL** | | | | | |
| **Source** | **DF** | **SS** | **MS** | **F** | **P** |
| **Factor** | **2** | **123.27** | **61.63** | **20.45** | **0.000** |
| **Error** | **321** | **967.35** | **3.01** |  |  |
| **Total** | **323** | **1090.62** |  |  |  |
| **S = 1.736** | | **R-Sq = 11.3%** | | **R-Sq (adj)=10.75%** | |

First, we need to do as followed:

1. Clean your screen
2. Run Library(mosaic)
3. Import data
4. Create variables

For explanation how to do this look at List 1, Exercise 1.

Clean the R Script: rm(list=ls())

Library: library(mosaic)

Import data: library(readxl)

L9E2 <- read\_excel("~/SDC/Quantitative Research Methods Course/Exercises/Dataset/List9/L9E2.xlsx")

View(L9E2)

attach(L9E2)

Create variables: RATING <- RATING

RECALL\_ <- RECALL\_

FACTOR <- FACTOR

VIOLENT <- VIOLENT

SEX <- SEX

NEUTRAL <- NEUTRAL

1. **Identify the experimental units in the study.**

The experimental units are the participants in the study.

1. **Identify the dependent (response) variable in the study.**

The dependent variable is the brand recall score.

1. **Identify the factor and treatments in the study.**

There is one factor in this study – tv viewing group. Since there is only one factor the treatments correspond to the factor levels of this variable. Thus, the treatments are the same as the three levels of TV viewer group. These 3 levels are violent content code, sex content code, and neutral TV.

1. **The sample mean recall scores for the three groups were  and . Explain why one should not draw an inference about differences in the population mean recall scores on the basis of only these summary statistics.**

The means given are only sample means. If new samples were selected and sample means computed the values and order of the sample means could change. In addition, the variances are not taken into account.

1. **An ANOVA on the data yielded the results shown in the table above. Locate the test statistic and p-value on the table.**

The test statistic is F=20.45, and the p-value is p=0.000.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **One-way ANOVA: VIOLENT, SEX, NEUTRAL** | | | | | |
| **Source** | **DF** | **SS** | **MS** | **F** | **P** |
| **Factor** | **2** | **123.27** | **61.63** | **20.45** | **0.000** |
| **Error** | **321** | **967.35** | **3.01** |  |  |
| **Total** | **323** | **1090.62** |  |  |  |
| **S = 1.736** | | **R-Sq = 11.3%** | | **R-Sq (adj)=10.75%** | |

1. **Interpret the results from part e, using α = 0.01. What can the researchers conclude about the 3 groups of TV ad viewers?**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **One-way ANOVA: VIOLENT, SEX, NEUTRAL** | | | | | |
| **Source** | **DF** | **SS** | **MS** | **F** | **P** |
| **Factor** | **2** | **123.27** | **61.63** | **20.45** | **0.000** |
| **Error** | **321** | **967.35** | **3.01** |  |  |
| **Total** | **323** | **1090.62** |  |  |  |
| **S = 1.736** | | **R-Sq = 11.3%** | | **R-Sq (adj)=10.75%** | |

Since the p-value is less than α (p=0.000<0.01), H0 is rejected. There is sufficient evidence to indicate differences in the mean recall scores among the three viewing groups at α=0.01. The researchers can conclude that the content of the TV show affects the recall of imbedded commercials.

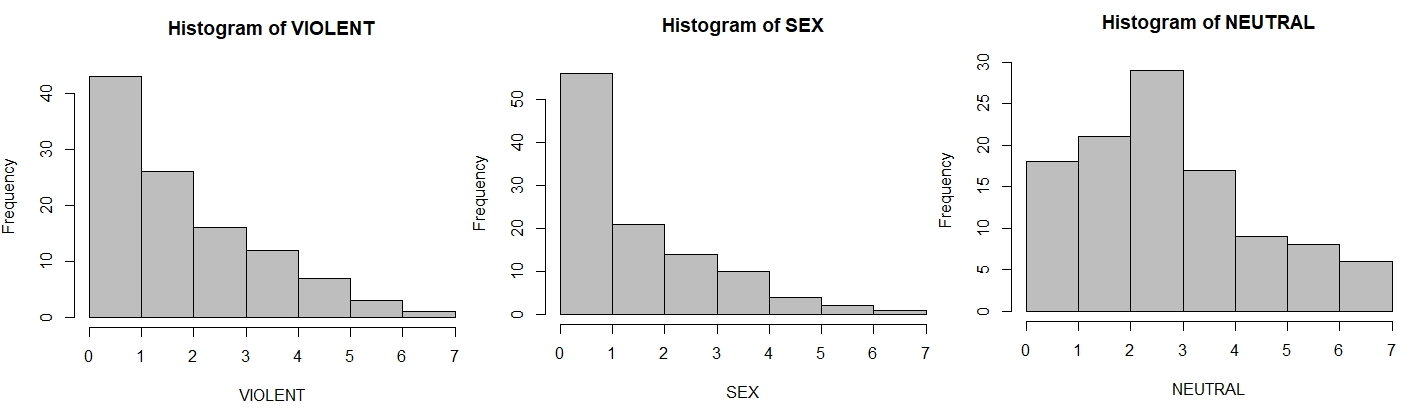
1. **Check that the ANOVA assumptions are reasonably satisfied.**

In this exercise, we need to calculate a histogram for each of the variables (Violent, Sex, and Neutral).

hist(VIOLENT,col = "grey",border = "black")

hist(SEX,col = "grey",border = "black")

hist(NEUTRAL,col = "grey",border = "black")



These histograms shows us that Violent and Sex are skewed to the right, and that only Neutral is somewhat normal distributed. However, we can also calculate the standard deviation of each variable, to see if they are the same.

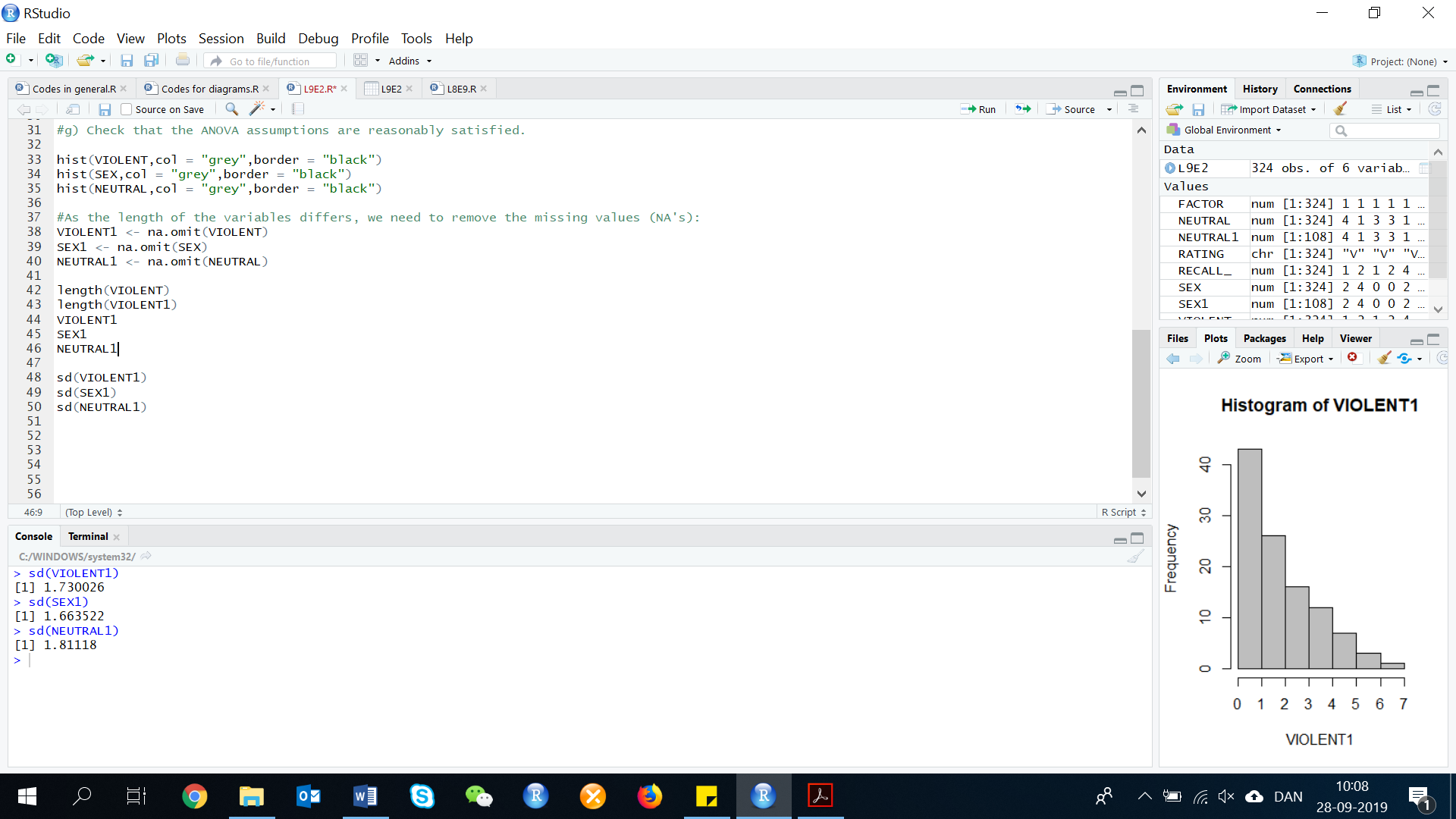
First we need to remove the missing values in each of the variables. If we look at Violent, we can see that according to R, when we import the variable, it has 328 observations. However, we can see by looking at the data, that it only has 108 observations. Therefore, we need to remove the NA’s:

VIOLENT1 <- na.omit(VIOLENT)

Now we do this for each of the variables:

SEX1 <- na.omit(SEX)

NEUTRAL1 <- na.omit(NEUTRAL)

Now we can calculate the standard deviations:

sd(VIOLENT1)

sd(SEX1)

sd(NEUTRAL1)

The assumptions for ANOVA are that the data are approximately normal and the variance of the groups are the same. From the legend above, the standard deviations are 1.730, 1.664, and 1.811. These are all very similar. From the plots, the distributions of the violent group and the neutral group are fairly normal. The distribution of the sex group is skewed to the right and may not be normal.

**Exercise 3. (30, ACCHW). *Homework assistance for accounting students*. The Journal of Accounting Education (Vol. 25, 2007) did a study of assisting accounting students with their homework. A total of 75 junior-level accounting majors who were enrolled in Intermediate Financial Accounting participated in the experiment. Students took a pretest on a topic not covered in class and then each was given a homework problem to solve on the same topic. A completely randomized design was employed, with students randomly assigned to receive one of the three different levels of assistance on the homework: (1) the completed solution, (2) check figures at various steps of the solution, and (3) no help at all. After finishing the homework, each student was given a posttest on the subject. The response variable of interest to the researchers was the knowledge gain (or, test score improvement), measured as the difference between the posttest and the pretest scores. The data (simulated from the descriptive statistics published in the article) are saved in the accompanying file.**

First, we need to do as followed:

1. Clean your screen
2. Run Library(mosaic)
3. Import data
4. Create variables

For explanation how to do this look at List 1, Exercise 1.

Clean the R Script: rm(list=ls())

Library: library(mosaic)

Import data: library(readxl)

L9E3 <- read\_excel("~/SDC/Quantitative Research Methods Course/Exercises/Dataset/List9/L9E3.xlsx")

View(L9E3)

attach(L9E3)

Create variables: NO <- NO

CHECK <- CHECK

FULL <- FULL

IMPROVE <- IMPROVE

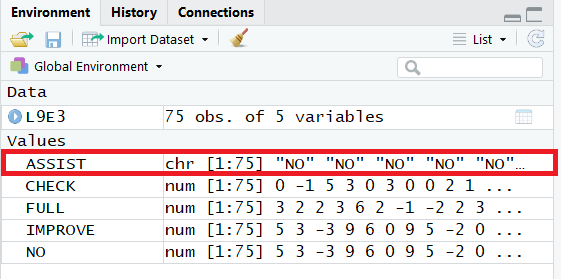
ASSIST <- ASSIST

1. **Give the null and alternative hypotheses tested in an analysis of variance of the data.**

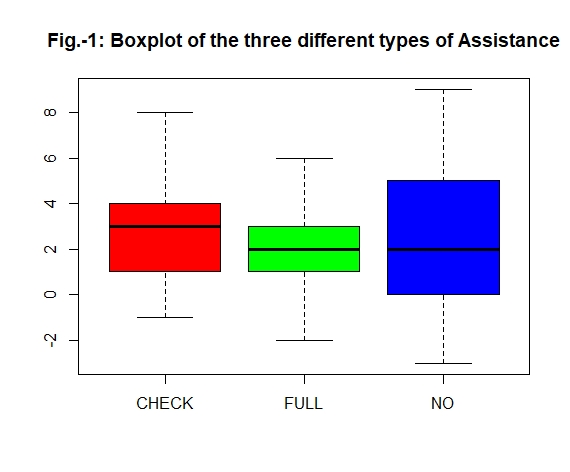
To determine if the mean knowledge gain differs among the three groups, we test:

1. **Summarize the results of the analysis in an ANOVA table.**

Before doing anything, you should check the variable type as in ANOVA, you need categorical independent variable (here the factor or treatment variable ‘Assist’). We check this, by looking at the description of the data:

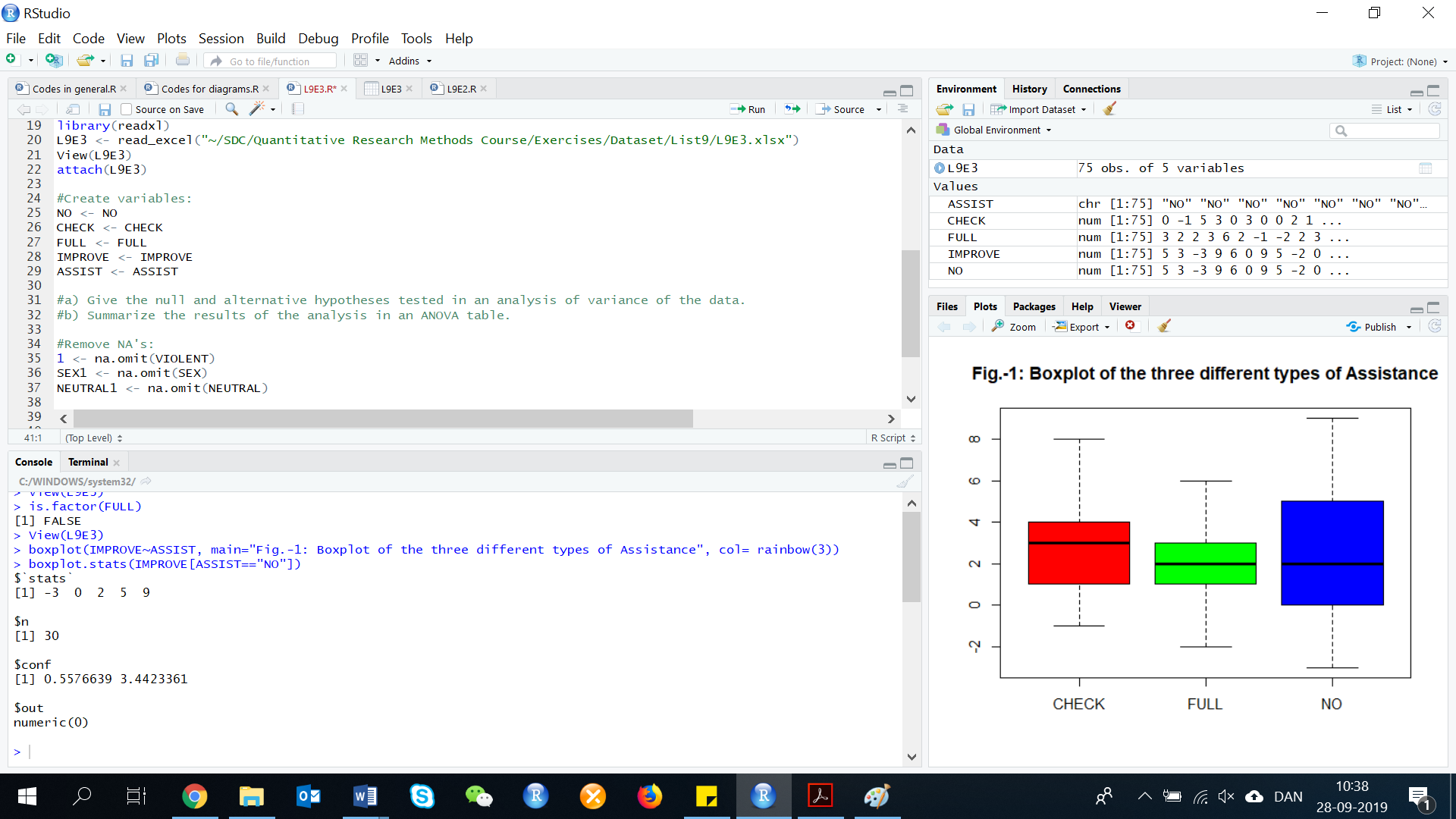


Here we can see, that the variable ‘Assist’ is in characters (chr) and not numbers (num). Now it is all set to run the ANOVA model in R. Like other linear model, in ANOVA you should check the presence of outliers. This can be detected with a boxplot. As there are three populations to study, you should use separate boxplot for each of the population. In R, it is done as:

boxplot(IMPROVE~ASSIST, main="Fig.-1: Boxplot of the three different types of Assistance", col= rainbow(3))

The above picture shows that there is no extreme observations. If there were extreme outliers, we could use the following code to detect the exact outlier. In this example, we will look at the population ‘NO’:

boxplot.stats(IMPROVE[ASSIST=="NO"])



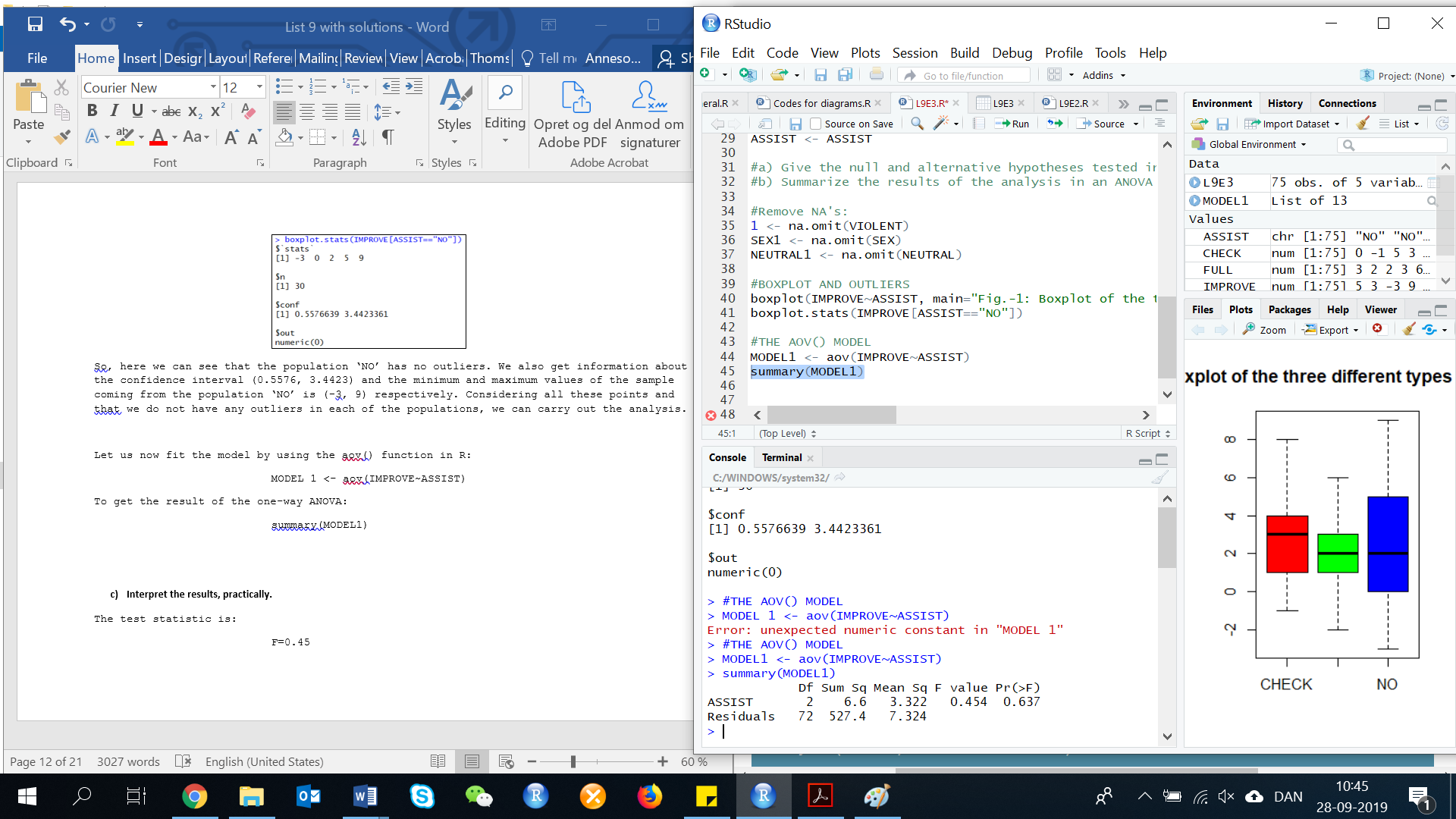
So, here we can see that the population ‘NO’ has no outliers. We also get information about the confidence interval (0.5576, 3.4423) and the minimum and maximum values of the sample coming from the population ‘NO’ is (-3, 9) respectively. Considering all these points and that we do not have any outliers in each of the populations, we can carry out the analysis.

Let us now fit the model by using the aov() function in R:

MODEL 1 <- aov(IMPROVE~ASSIST)

To get the result of the one-way ANOVA:

summary(MODEL1)



1. **Interpret the results, practically.**

The test statistic is:

F=0.454

P=0.637

Since the p-value is larger than any reasonable significance level, H0 is not rejected. There is insufficient evidence to indicate a difference in the mean knowledge gained among the three levels of assistance for any reasonable value of α. Practically speaking, there is not one type of assistance that helps students more than another.

**9.3. Multiple comparisons of means**

**Exercise 4. (46, ADREC). *Study of recall of TV commercials*. Refer to the Journal of applied psychology (June 2002) completely randomized design study to compare the mean commercial recall scores of viewers of three TV programs, presented in Exercise 28. Recall that one program had a violent content code (V) rating, one had a sex content code (S) rating, and one was a neutral TV program. Using Tukey’s method, the researchers conducted multiple comparisons of the three mean recall scores.**

First, we need to do as followed:

1. Clean your screen
2. Run Library(mosaic)
3. Import data
4. Create variables

For explanation how to do this look at List 1, Exercise 1.

Clean the R Script: rm(list=ls())

Library: library(mosaic)

Import data: library(readxl)

L9E4 <- read\_excel("~/SDC/Quantitative Research Methods Course/Exercises/Dataset/List9/L9E4.xlsx")

View(L9E4)

attach(L9E4)

Create variables: RATING <- RATING

RECALL <- RECALL

FACTOR <- FACTOR

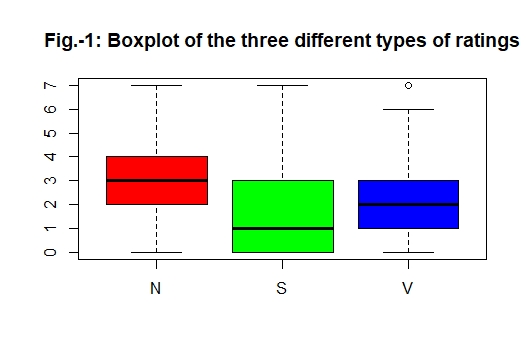
VIOLENT <- VIOLENT

SEX <- SEX

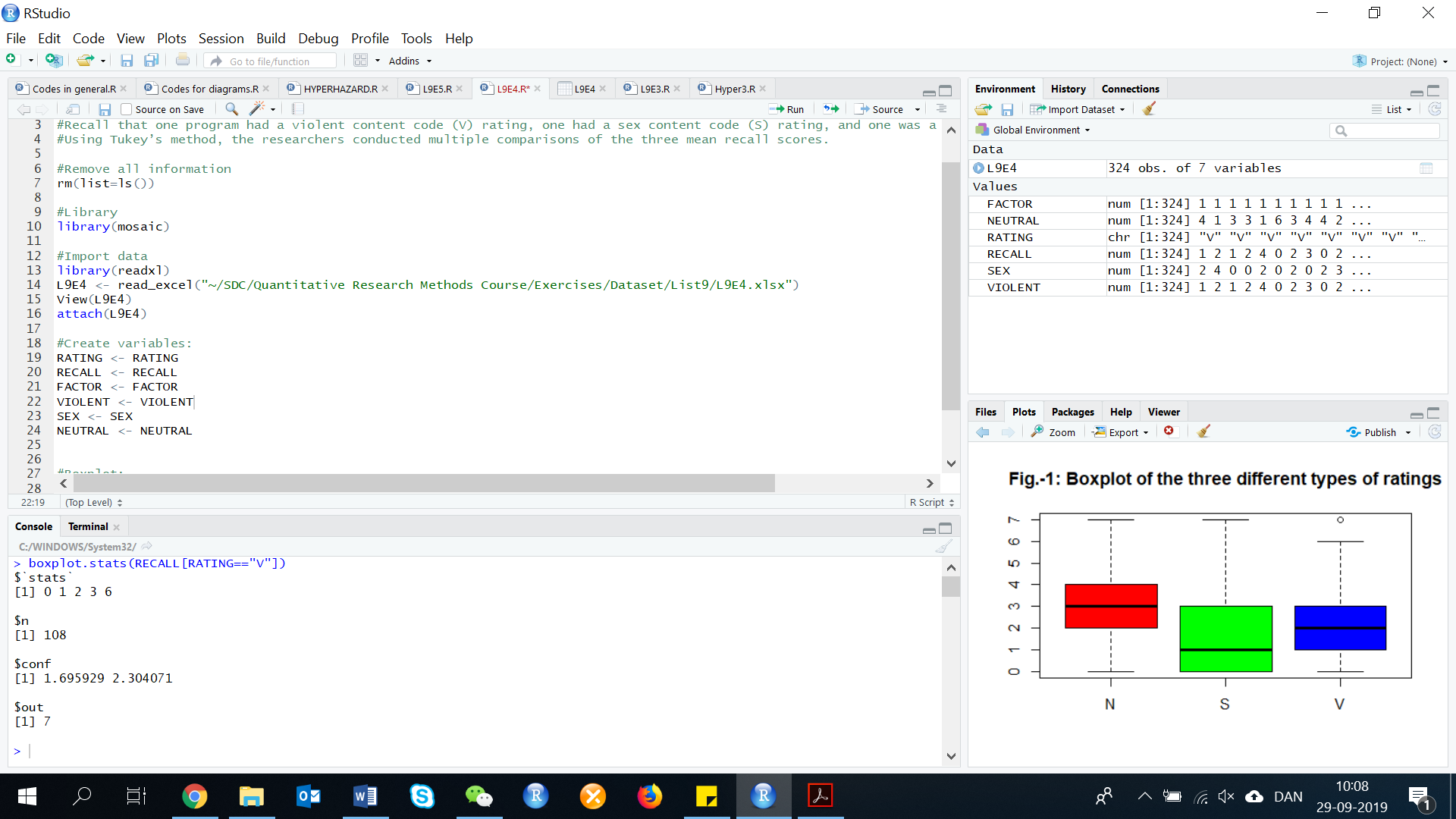
NEUTRAL <- NEUTRAL

To use the Tukey’s method, we first need to make an Anova analysis. Therefore, we start with a boxplot and detect the outliers, so we understand the data better.

boxplot(RECALL~RATING, main="Fig.-1: Boxplot of the three different types of ratings", col= rainbow(3))



The boxplot shows us that the rating type ‘V’ (Violent), has an outlier. To detect, the outlier, we use the following code:

boxplot.stats(RECALL[RATING=="V"])

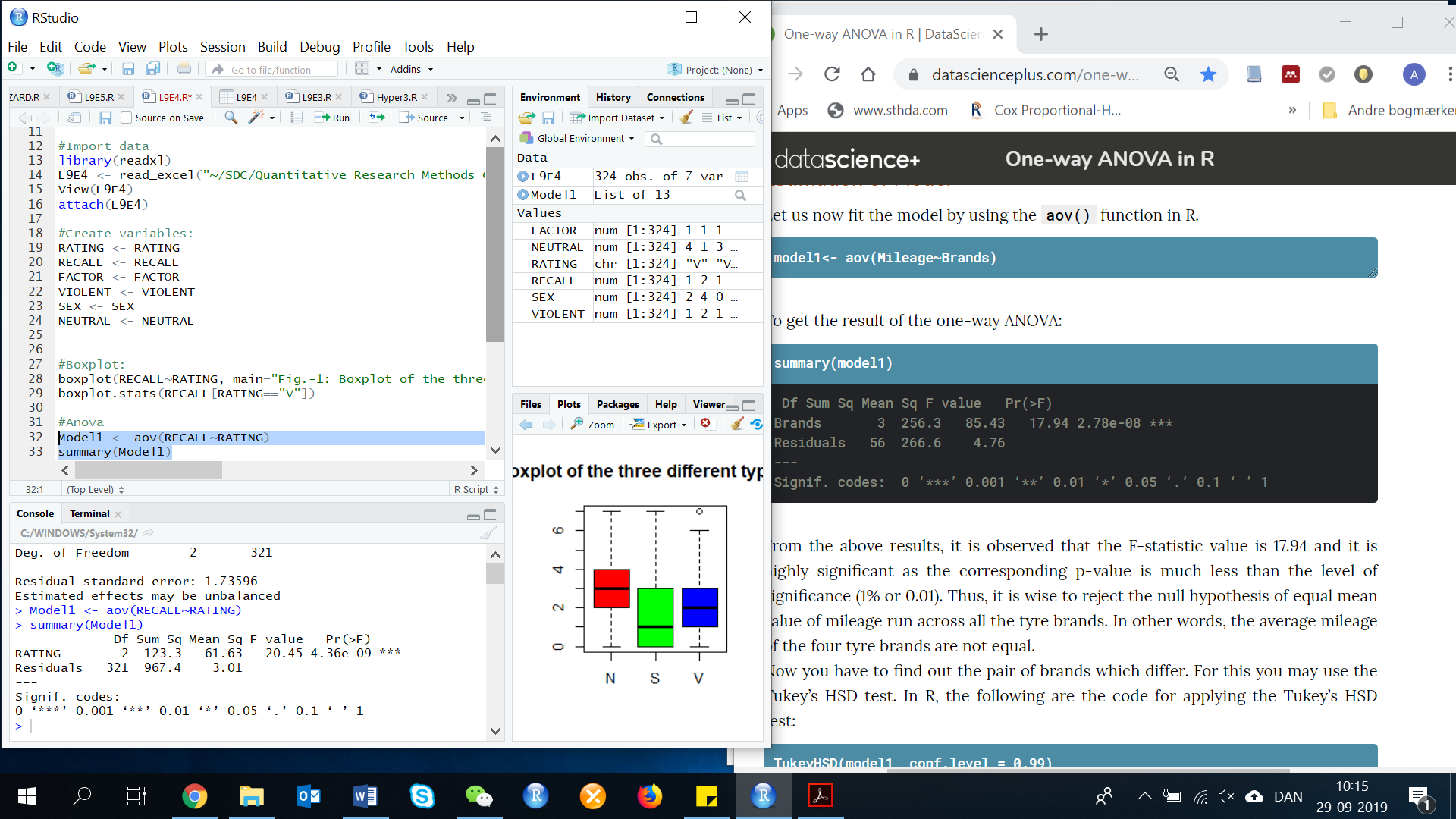
So, the outlier is the observation values ‘7’. The confidence interval is (1.695929, 2.304071) and the minimum and maximum values of the sample coming from the population ‘Violent’ is 0 and 7 respectively. Considering all these points, we ignore the outlier value ‘7’ momentarily and carry out the analysis. If at a later stage, we find that this outlier may create problems in the estimation, we will exclude it.

Let now make the model:

Model1 <- aov(RECALL~RATING)

summary(Model1)

The R console give us the following result:

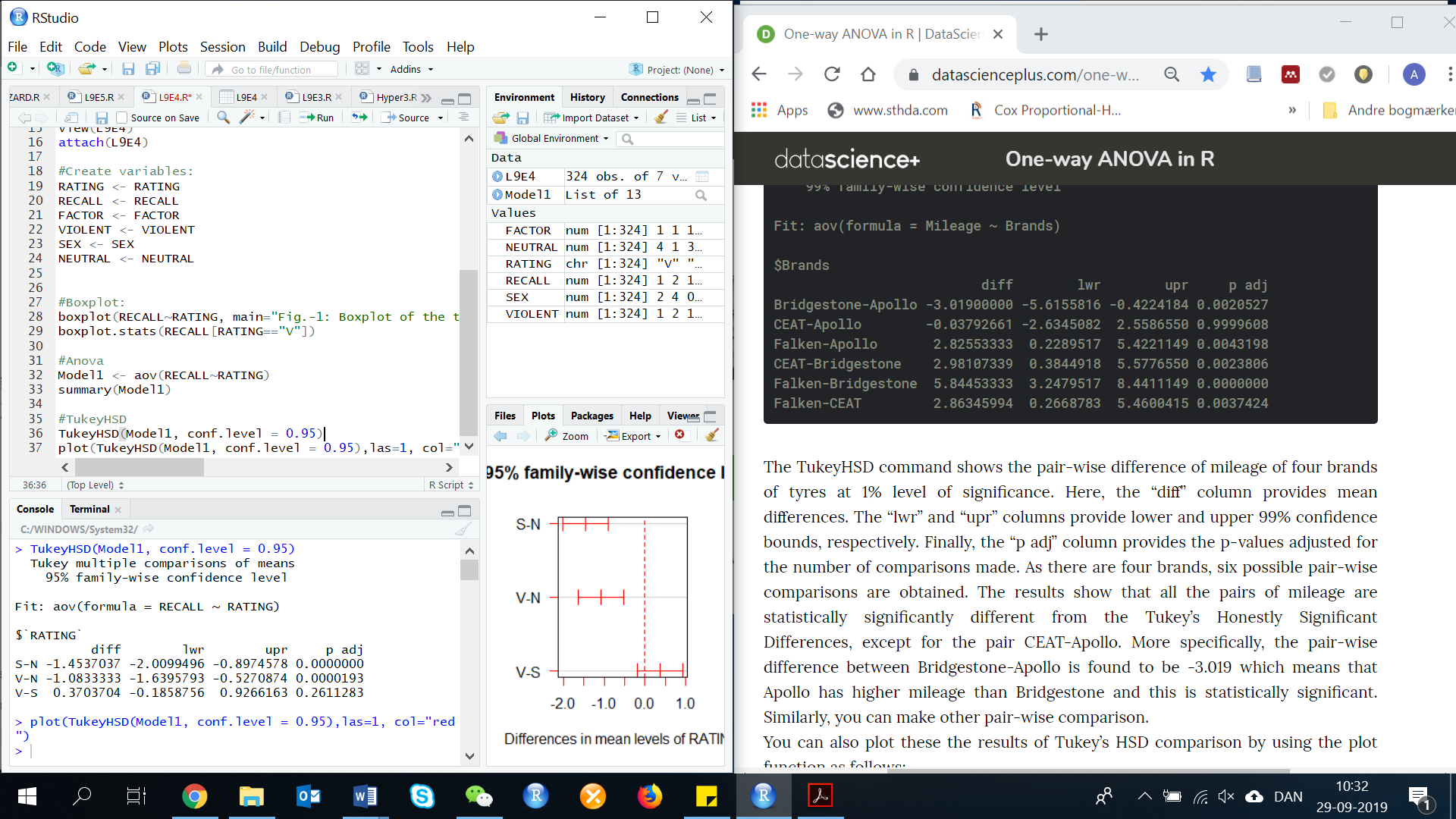


From the above results, it is observed that the F statistic value is 20.45 and it is highly significant as the corresponding p-value is much less than the level of significance (1% or 0.01).

Now we can use the Tukeys find out how the types of ratings differs.

TukeyHSD(DATA, conf.level=NUMBER)

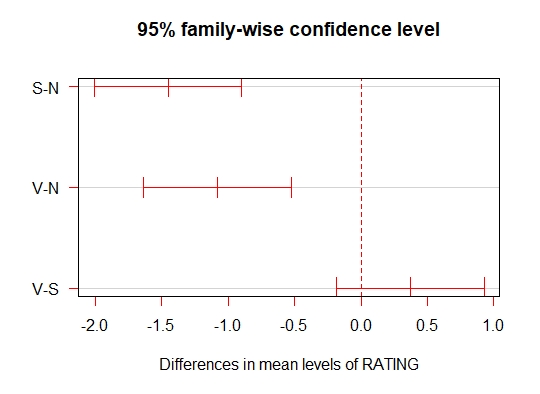
TukeyHSD(Model1, conf.level = 0.95)



The TukeyHSD command shows the pair-wise difference of recall scores of the three types of rating at 5% level of significance. Here, the ‘diff’ column provides mean differences. The ‘lwr’ and ‘upr’ columns provide lower and upper 95% confidence bounds, respectively. Finally, the ‘p adj’ column provides the p-values adjusted for the number of comparisons made. The results show that two (S-N, V-N) of the pairs of recall scores are statistical significantly different from the Tukey’s Honestly Significant Differences.

We can also plot the result of the Tukey’s HSD comparison by using the plot function as follows:

plot(TukeyHSD(Model1, conf.level = 0.99),las=1, col="red")

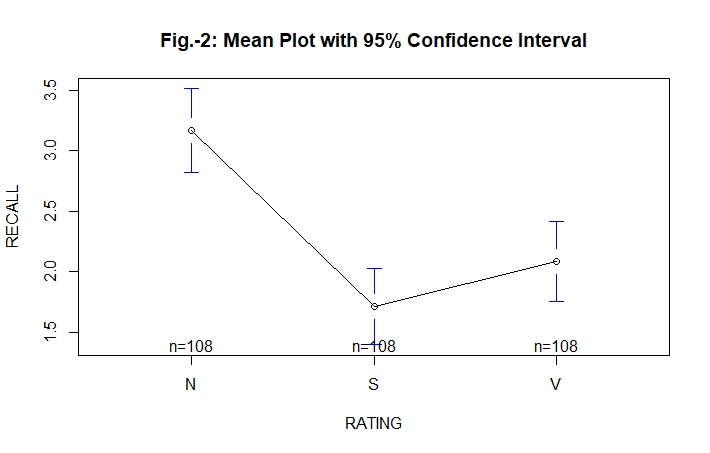


Another way of visualization the differences is, by plotting the means of the recall scores of the three types of ratings.

install.packages(gplots)

library(gplots)

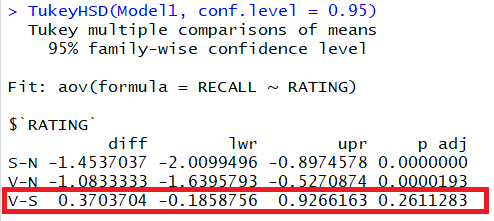
plotmeans(RECALL~RATING, main="Fig.-2: Mean Plot with 95% Confidence Interval")



**a) How many pairwise comparisons were made in this study?**

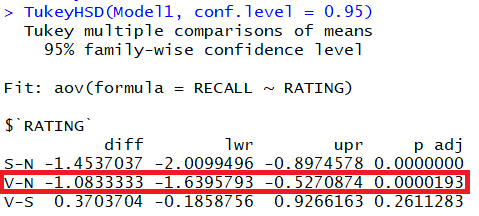
The total number of pairwise comparisons made in the Bonferroni analysis is:

**b) The multiple comparison procedure was applied to the data and the results are shown in the Minitab printout at the bottom of the page. An experimentwise error rate of .05 was used. Locate the confidence interval for the comparison of the V and S groups. Interpret this result practically.**

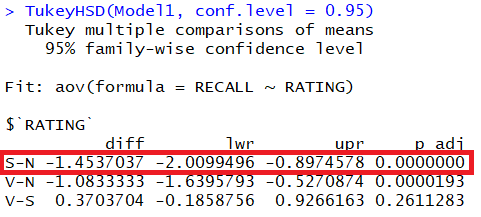
****

The confidence interval for comparing V and S groups is (-0.1858756, 0.9266163), and the pair-wise difference is found be 0.3703704. However, since it is not statistical significant and the confidence interval contains 0, there is no indication that there is a difference in mean recall between the V and S groups at .

**c) Repeat part b for the remaining comparisons. Which of the groups has the largest mean recall score?**



The confidence interval for comparing the V and N groups is (-1.6395793, -0.5270874), and the pair-wise difference is found to be -1.0833333. Since the confidence interval does not contain 0 and that the p-value (0.0000193) is much less than the level of significance (1% or 0.01), there is evidence to indicate there is a difference in mean recall between V and N groups at . The evidence indicates that the Neutral group is significantly higher than the mean recall of the Violent group.



The confidence interval for comparing the S and N groups is (-2.0099496, -0.8974578), and the pairwise difference is found to be -1.4537037. Since the confidence interval does not contain 0 and the p-value is statistical significant, there is evidence to indicate there is a difference in mean recall between S and N groups at . This means that Neutral has a higher recall score than Sex, and this is statistical significant.

**d) In the journal article, the researchers concluded that “memory for (television) commercials is impaired after watching violent or sexual programming”. Do you agree?**

Yes. When compared to the Neutral group, the mean recalls for the V and S groups are significantly lower than the mean recall for the Neutral group.

**9.4 The randomized block design**

**Exercise 5. (60, PLANTS). *Reducing on-the-job stress*. Plant therapist believe that plants can reduce on-the-job stress. A Kansas State University study was conducted to investigate this phenomenon. Two weeks prior final exams, 10 undergraduate students took part in an experiment to determine what effect the presence of a live plant, a photo of a plant, or absence of a plant has on a student’s ability to relax while isolated in a dimly lit room. Each student participated in three sessions - one with a live plant, one with a plant photo, and one with no plant (control). During each session, finger temperature was measured at 1-minute intervals for 20 minutes. Because increasing finger temperature indicates an increased level of relaxation, the minimum temperature (in degrees) was used as the response variable. For example, one student’s finger measured 95.6 in the “live plant” condition, 92.6 in the “plant photo” condition, and 96.6 in the “no plant” condition (those are information for Student 1). The temperatures under the three conditions for the other nine students follow:**

|  |  |
| --- | --- |
| **Student 2** | **95.6 94.8 96.0** |
| **Student 3** | **96.0 97.2 96.2** |
| **Student 4** | **95.2 94.6 95.7** |
| **Student 5** | **96.7 95.5 94.8** |
| **Student 6** | **96.0 96.6 93.5** |
| **Student 7** | **93.7 96.2 96.7** |
| **Student 8** | **97.0 95.8 95.4** |
| **Student 9** | **94.9 96.6 90.5** |
| **Student 10** | **91.4 93.5 96.6** |

**These data are saved in the accompanying file. Conduct an ANOVA and make the proper inferences at α = .10.**

First, we need to do as followed:

1. Clean your screen
2. Run Library(mosaic)
3. Import data
4. Create variables

For explanation how to do this look at List 1, Exercise 1.

Clean the R Script: rm(list=ls())

Library: library(mosaic)

Import data: library(readxl)

L9E5 <- read\_excel("~/SDC/Quantitative Research Methods Course/Exercises/Dataset/List9/L9E5.xlsx")

View(L9E5)

attach(L9E5B)

Create variables: STUDENT <- STUDENT

MEASURE <- MEASURE

PLANT <- PLANT

To determine if there are differences among the mean temperatures among the three treatments, we test:

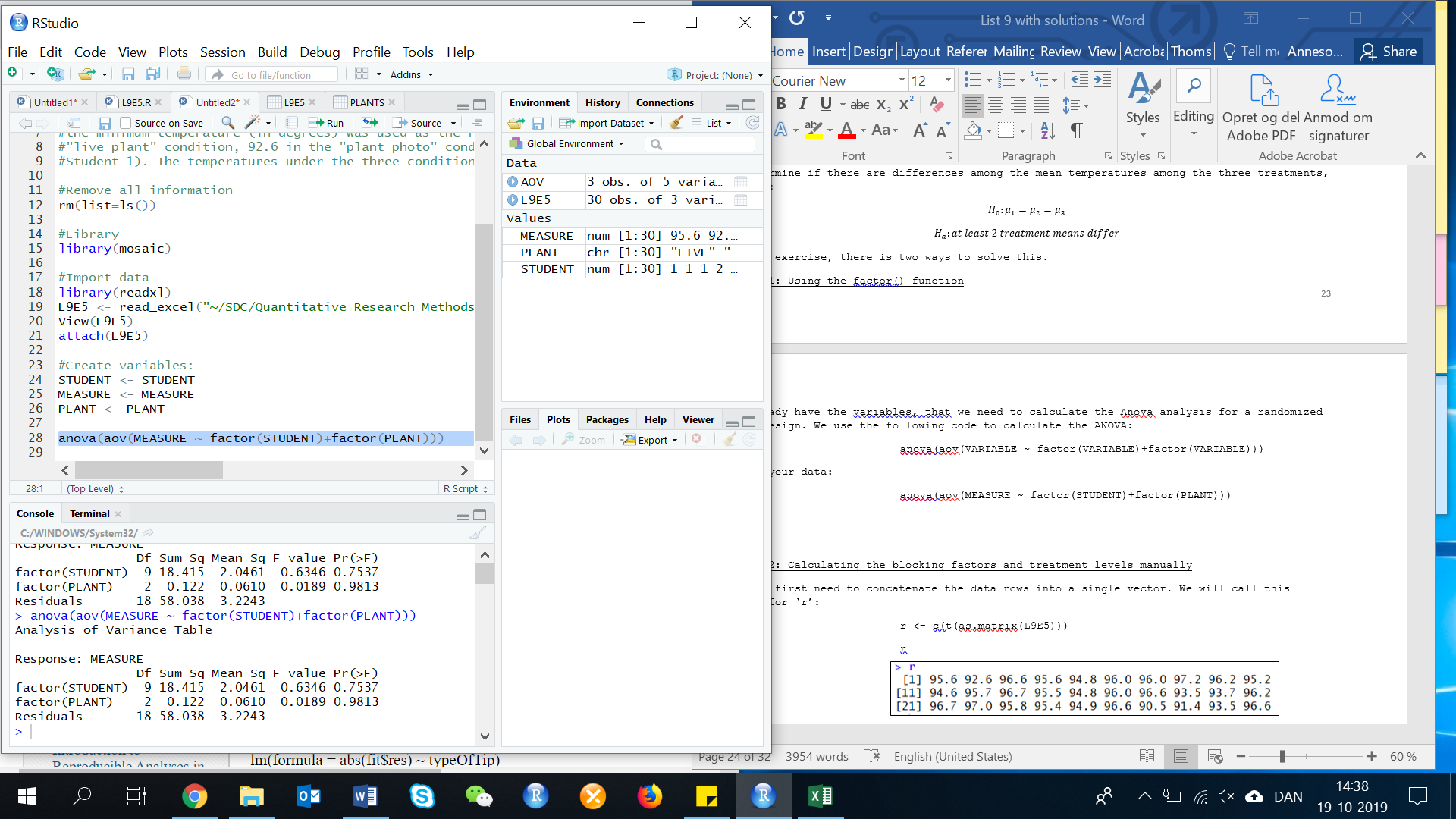
In this exercise, there is two ways to solve this.

We already have the variables, that we need to calculate the Anova analysis for a randomized block design. We use the following code to calculate the ANOVA:

anova(aov(VARIABLE ~ factor(VARIABLE)+factor(VARIABLE)))

Insert your data:

anova(aov(MEASURE ~ factor(STUDENT)+factor(PLANT)))

Result:

The test statistic is F=0.0189. The associated p-value is p=0.9813. Since the p-value is very large, there is no evidence of a difference in mean temperature among the three treatments for any reasonable value of α. Since there is no difference, we do not need to compare the means. It appears that the presence of plants or pictures of plants does not reduce stress.

**9.5 Factorial experiments: Two factors**

**Exercise 6. (76, EGGS). *Commercial eggs produced from different housing systems*. Refer to the Food Chemistry (Vol. 106, 2008) study of four different types of egg housing systems, Exercise 33. Recall that the four housing systems were cage, barn, free range, and organic. In addition to housing system, the researchers also determined the weight class (medium or large) for each sampled egg. The data on whipping capacity (percent overrun) for the 28 sampled eggs are shown in the accompanying table. The researchers want to investigate the effect of both housing system and weight class on the mean whipping capacity of the eggs. In particular, they want to know whether the difference between the mean whipping capacity of medium and large eggs depends on the housing system.**

|  |  |  |
| --- | --- | --- |
| **Housing** | **Wtclass** | **Overrun (%)** |
| **Cage** | **M**  **L** | **495, 462, 488, 471, 471**  **502, 472, 474, 492, 479** |
| **Free** | **M**  **L** | **513, 510, 510**  **520, 531, 521** |
| **Barn** | **M**  **L** | **515, 516, 514**  **526, 501, 508** |
| **Organic** | **M**  **L** | **532, 511, 527**  **530, 544, 531** |

First, we need to do as followed:

1. Clean your screen
2. Run Library(mosaic)
3. Import data
4. Create variables

For explanation how to do this look at List 1, Exercise 1.

Clean the R Script: rm(list=ls())

Library: library(mosaic)

Import data: library(readxl)

L9E6 <- read\_excel("~/SDC/Quantitative Research Methods Course/Exercises/Dataset/List9/L9E6.xlsx")

View(L9E6)

attach(L9E6)

Create variables: HOUSING <- HOUSING

THICKNESS <- THICKNESS

OVERRUN <- OVERRUN

STRENGTH <- STRENGTH

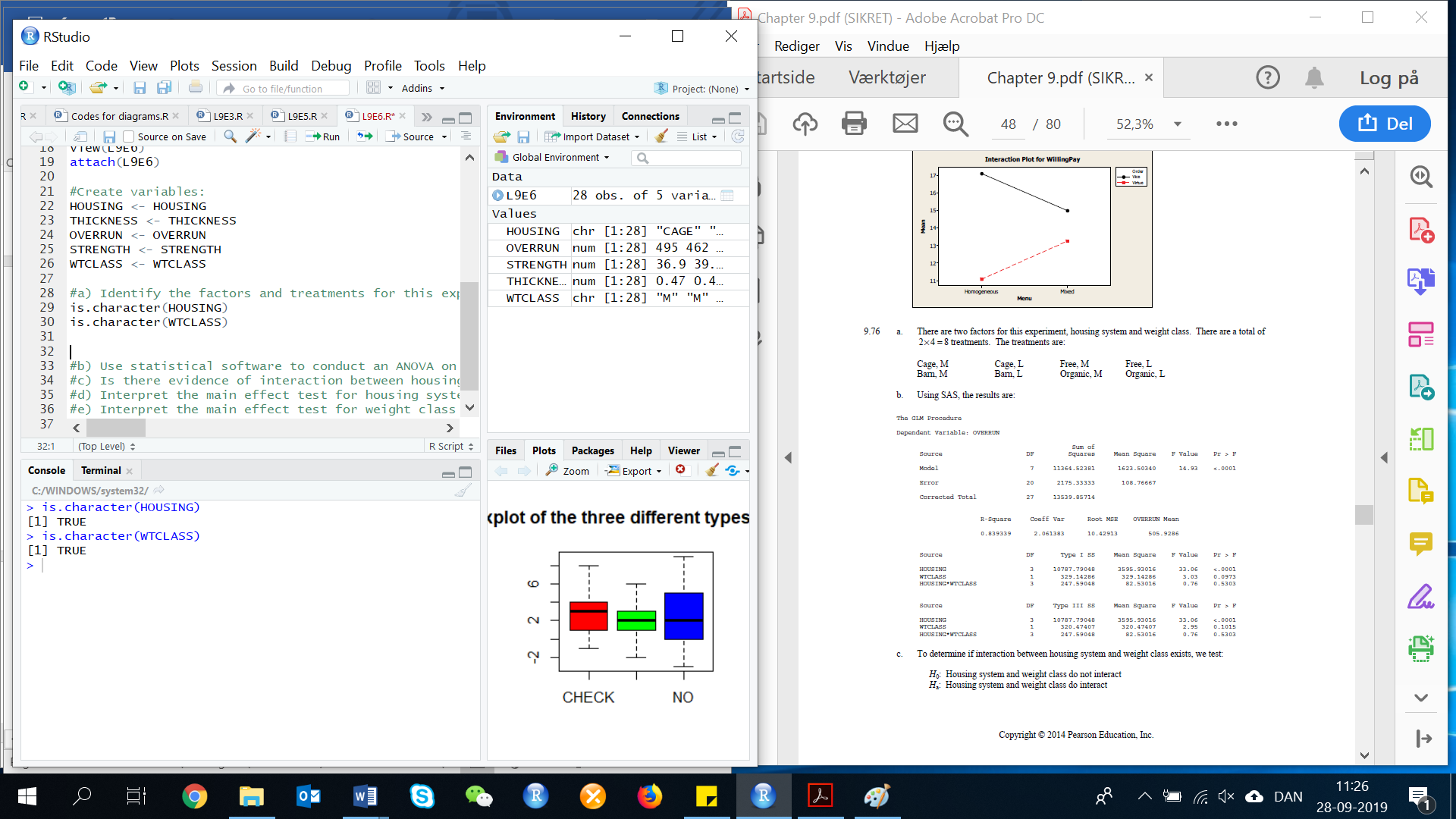
WTCLASS <- WTCLASS

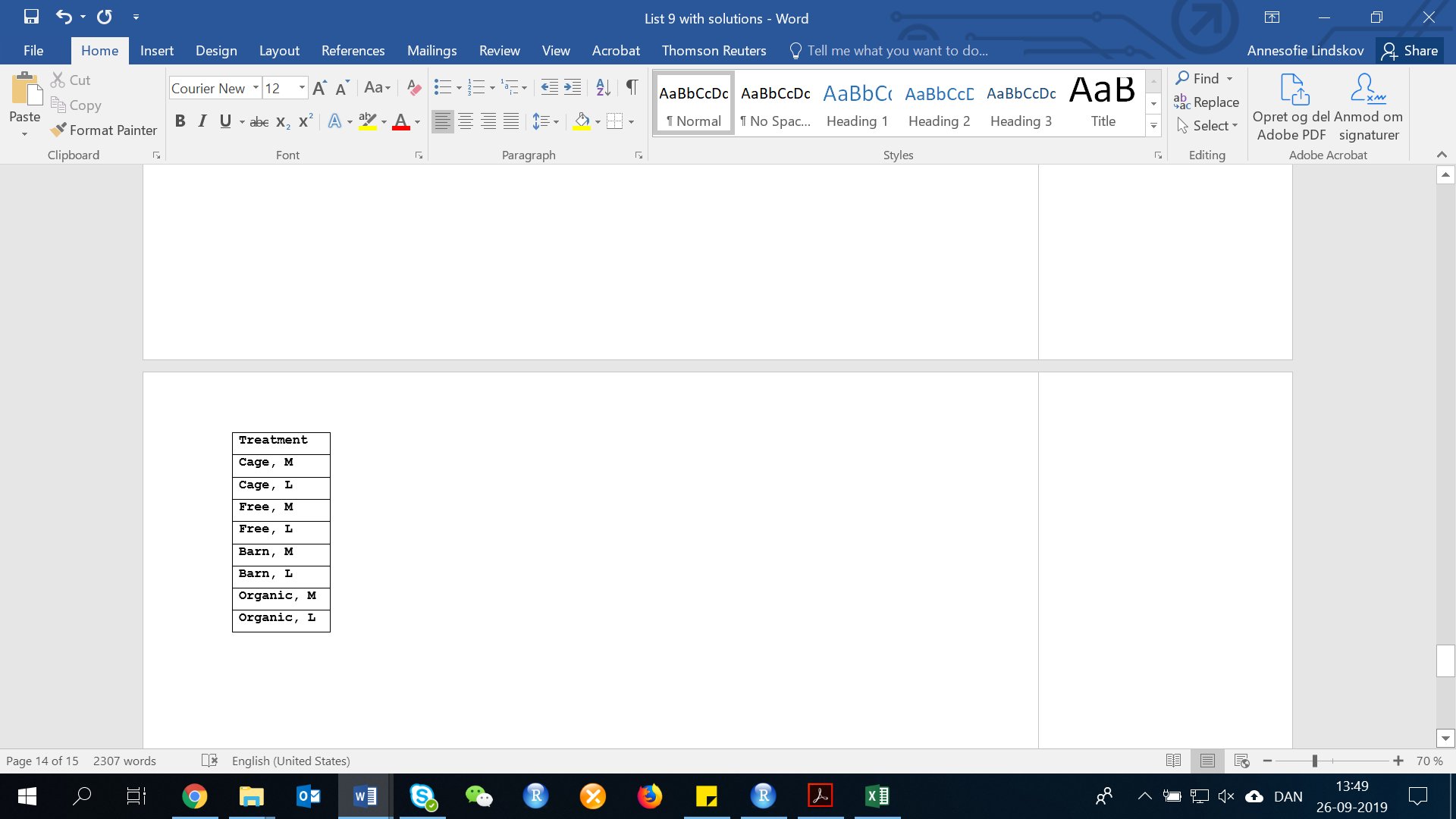
1. **Identify the factors and treatments for this experiment.**

By looking at our data, we can see (by detecting it with is.character()) that we have two characters (categorical data) for this experiment, ‘HOUSING’ and ‘WTCLASS’.

is.character(HOUSING)

is.character(WTCLASS)



These two characters have respectively 28 observations. However, Housing is has four populations (CAGE, FREE, BARN, ORGANIC) and WTCLASS has two populations (M, L). Therefore, we have in total 2\*4=8 treatments.

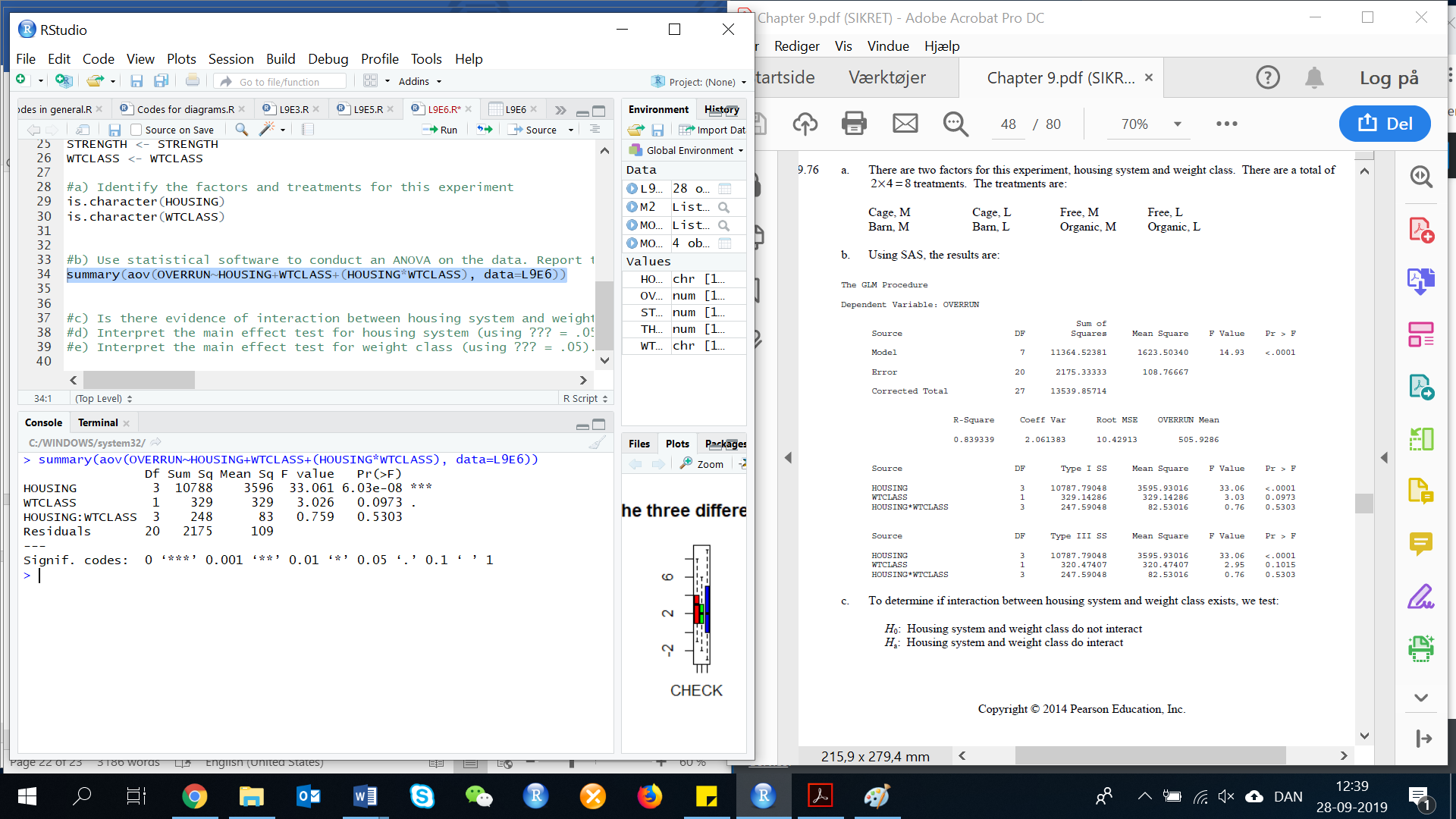
|  |  |
| --- | --- |
| **Housing** | **Wtclass** |
| **Cage** | **M**  **L** |
| **Free** | **M**  **L** |
| **Barn** | **M**  **L** |
| **Organic** | **M**  **L** |

**b) Use statistical software to conduct an ANOVA on the data. Report the results in an ANOVA table.**

In this exercise, we need to make new ANOVA analysis. The Dependent variable is ‘OVERRUN’, and the three independent variables are ‘HOUSING’, ‘WTCLASS’ and the interaction term ‘HOUSING\*WTCLASS’. To make this ANOVA analysis, we use the following code:

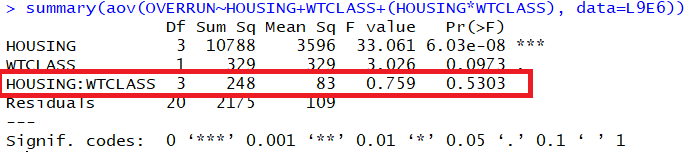
summary(aov(VARIABLE~VARIABLE+VARIABLE+(INTERACTION.TERM), data=DATA))

summary(aov(OVERRUN~HOUSING+WTCLASS+(HOUSING\*WTCLASS), data=L9E6))

Then the R console gives us:

**c) Is there evidence of interaction between housing system and weight class? Test using α = .05 (Hint: Due to an unbalanced design, you will need to analysis the data using the general linear model procedure of your statistical software). What does this imply, practically?**

To determine if interaction between housing system and weight class exists, we test:



By looking at this ANOVA table, we can see that the test statistic is:

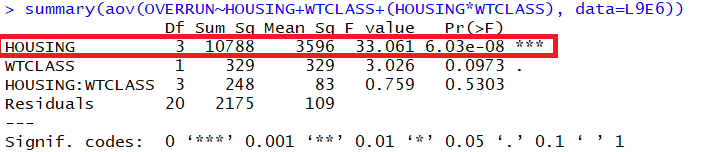
F=0.759

P=0.5303

Since the p-value is not less than α(p=0.5303 differs from 0.05), H0 is not rejected. There is insufficient evidence to indicate that housing system and weight class interact at α=0.05.

**d) Interpret the main effect test for housing system (using α = .05). What does this imply, practically?**

To determine if there is a difference in mean whipping capacity among the 4 housing systems, we test:



By looking at this ANOVA table, we can see that the test statistic is:

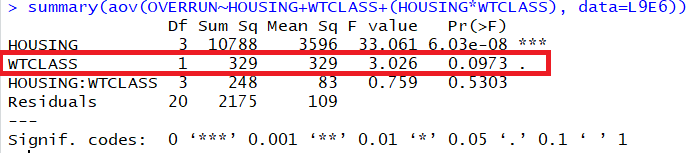
F=33.06

P=0.0001

Since the p-value is less than α(p<0.001<0.05), H0 is rejected. There is sufficient evidence to indicate a difference in mean whipping capacity among the 4 housing systems at α=0.05.

**e) Interpret the main effect test for weight class (using α = .05). What does this imply, practically?**

To determine if there is a difference in mean whipping capacity between the 2 weight classes, we test:



By looking at the data, we can see that the test statistic is:

F=2.95

P=0.1015

Since the p-value is not less than α(p<0.1015 differs 0.05), H0 is not rejected. There is insufficient evidence to indicate a difference in mean whipping capacity between the 2 weight classes at α=0.05.