Documented solutions for OpenIntro labs

निरंजन*
Department of Linguistics
University of Mumbai

PhD coursework examination
Paper II: Quantitative and qualitative methods in linguistics
Term-paper submission 1
February 1, 2024

Abstract

This document contains solved exercises which are offered in the OpenIntro coursework (Diez et al., 2012). The OpenIntro coursework mainly introduces us with statistical methods and R programming language as a tool for statistical computation. The first half of the document (section 3) answers the questions found in the coursework, whereas the second half of the document (section 4) deals with the documentation of the code used to produce the respective results. This project is an amalgamation of two programming languages namely LaTeX and R.

Contents

1 Introduction

I have used the DocStrip utility of LaTeX to generate and document the required R-scripts. Hence the real source files of this project are `paper-2a.dtx` and `paper-2a.ins`. This project should be compiled with the `l3build` tool[1] developed specifically for building and unpacking big projects. All the code is bundled together in order to avoid the loss of information. Generated files are usually not version-controlled, hence the bundle doesn't have R-files. One may obtain all of them by running `latex paper-2a.ins` in the project directory. Alternatively, `l3build doc` command in the project directory will generate all of them in `build/unpacked` directory.

2 Dependencies

Some of the generated R-files depend on the files listed below. They are licensed under CC-BY-SA 3.0 Unported license which is incompatible with the license I have chosen for my code (i.e., GPLv3+), but since I haven't modified the files, I redistribute them as they are without changing the original license.

- `arbuthnot.csv`
- `cdc.csv`
- `mlb11.csv`

The figures seen in the document aren't images. They are Ti*k*Z vector codes, produced with the help of tikzDevice[2] library. I have made some minor changes to the output of the script. The respective TeX files can be seen in `support/` directory.

3 Solutions

In this section we look at the solutions of the exercises. I have attempted exercises from the following three chapters.

- Introduction to R and Rstudio
- Introduction to data
- Introduction to linear regression

3.1 Introduction to R and RStudio

3.1.1 Exercise 1

I define a variable named `arbutgirls`. This variable has the extracted data of baptised girls from the larger data-set. Simply calling this variable on R-prompt prints the desired data.

[1]Release 2023-12-15 or later
[2]https://cran.r-project.org/package=tikzDevice

3.1.2 Exercise 2

In this exercise we are supposed to describe the trend in the number of baptisation of girls over the years. For this we need to visually represent the data and then interpret it. By looking at figure 1 we can conclude the following:

1. The number of baptisation is fairly stable and increasing except one major fall.
2. Around 1640, baptisation among girls witnessed a sudden and significant fall. In a decade it reached its lowest number of the entire time-span.
3. After this fall, it again caught the momentum and quickly reinstated its earlier rate.
4. It still experienced minor falls time to time, but they weren't too steep and most of them didn't last for too long.

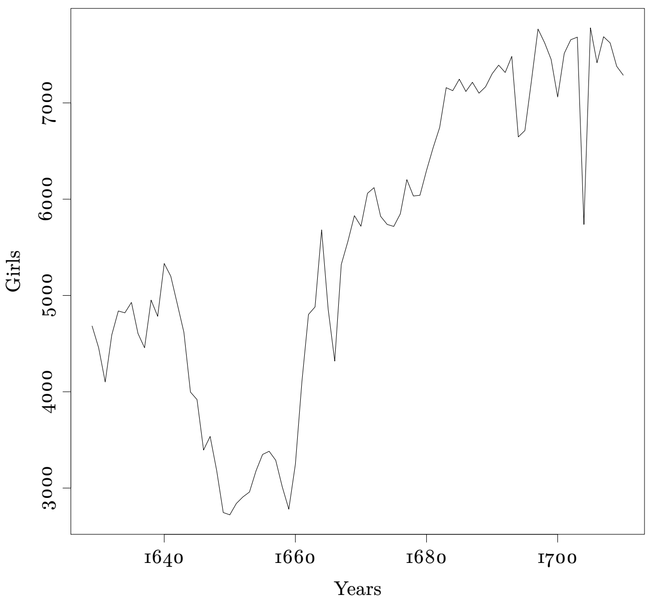

Figure 1: The graph of baptised girls

3.1.3 Exercise 3

Here We are supposed to describe the trend of baptisation of boys over the time. For that, have a look at figure 2.

We can observe fairly similar curves for both the genders and hence I infer that gender is *not* a significant variable with respect to baptisation.

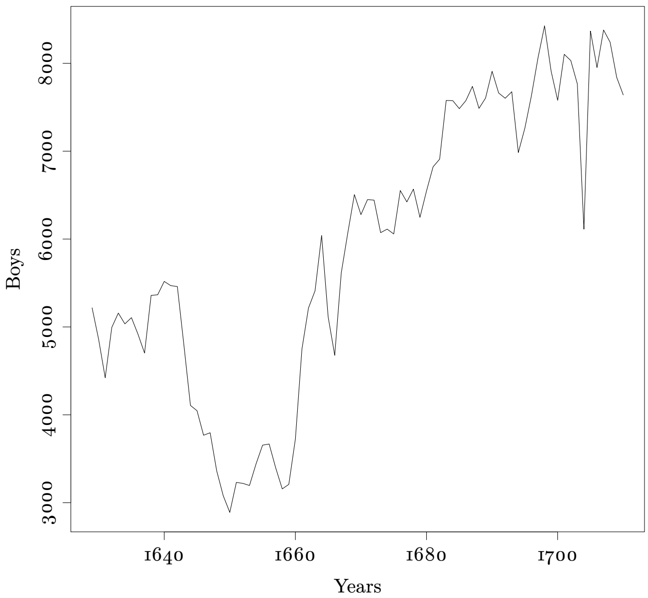

Figure 2: The graph of proportion of baptised boys

3.2 Introduction to data

3.2.1 Exercise 1

20000 is the number of cases and 9 is the number of variables in the data-set.

The following is a list of variables that we have in the data-set. The first exercise asks us to note down their types.

`genhealth`: This is an *ordinal categorical* variable, Because the values of this variable show a cline, but they are qualities and therefore they have categorical labels.

`exerany`: This is a *nominal categorical* variable. It has 0 and 1 as its values, but both of them have discrete meanings mapped to them, i.e., respectively no and yes. This isn't something that can be counted. Hence despite of using a numeric value, the variable is nominal.

`hlthplan`: The nature of this variable is identical to the `exerany` variable.

`smoke100`: This is a *discrete quantitative* variable. It counts individual events of smoking which don't have any order as such.

`height`: This is a *discrete categorical* variable, but unlike `smoke100`, the values of this variable can be arranged on a cline too for certain inferences. So it is half-way between a discrete and a continuous variable.

`weight`: The nature of this variable is identical to the `height` variable.

`wtdesire`**:** This variable is a *discrete quantitative* variable and its values can *not* be ordered. In fact, doing so will in no way help us to infer anything about its correlation with other variables.

`age`**:** This is a *continuous quantitative* variable.

`gender`**:** The way this data has been collected, gender is treated as a *discrete qualitative* variable.

3.2.2 Exercise 2

We are supposed to give summaries of height and age. R returns the following summary-table for the `height`.

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  48.00   64.00   67.00   67.18   70.00   93.00
```

On the other hand it returns the following for the `age`.

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  18.00   31.00   43.00   45.07   57.00   99.00
```

IQR (i.e., Interquartile Range) can be directly printed in R, but we can also manually subtract the figure of first quartile from the third one resulting in the IQR. The IQR for height is 6 and for age it is 26.

We get the following relative frequency distribution for gender and exercise:

```
      0    1
 f 2937 7494
 m 2149 7420
```

For finding out the number of males, we can use the frequency distribution table of gender like this:

```
    f     m
10431  9569
```

Therefore the number of males in this data-set is 9569.

For finding out the proportion of the population which reported excellent health, we will have a look at the following table.

```
excellent      fair      good      poor very good
  0.23285   0.10095   0.28375   0.03385   0.34860
```

This table means that ≈ 0.23 proportion of the total population reported excellent health.

3.2.3 Exercise 3

Figure 3 is the mosaic plot for the data of smoking and gender. It shows us that there are more women than men in this data-set who have *not* smoked 100 cigarettes in their lifetime and as both the variables are binary, the exact opposite of all the variables, i.e., more men than women in this data-sett *have* smoked more than 100 cigarettes in their lifetime, is also true.

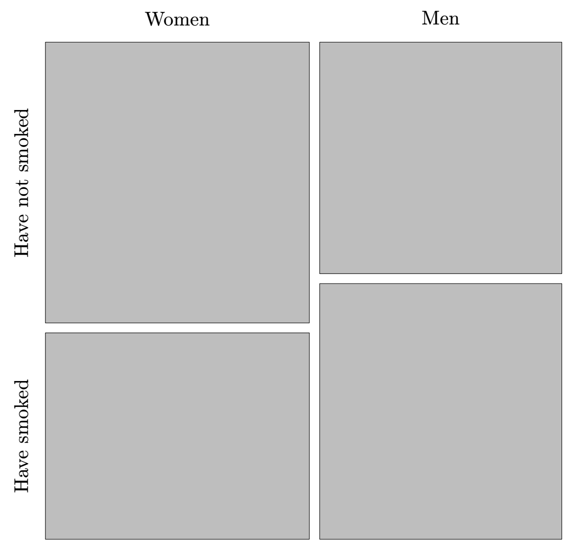

Figure 3: Gender and smoking

3.2.4 Exercise 4

I save the required data in `sdata` variable[3]. The explanation can be found in the implementation.

```
mdata <- subset(cdc,gender=="m")
adata <- subset(mdata,age=="23")
sdata <- subset(mdata,smoke100=="1")
```

3.2.5 Exercise 5

All the box-plots from figure 4 show a fairly normal distribution without any skew, but all of them have a significant number of outliers on the positive side. All the plots vary in their dispersion from which the plot representing the 'excellent' health has the lowest dispersion and the one representing 'poor' has the most dispersed data. Since this data has an introspective question to the respondent which corresponds to their BMI we can conclude that a specific BMI can't really predict what a person might feel about their health. For this the support comes from the fact that the medians and dispersions of these box-plots are not drastically different.

If we look at figure 5, one immediate observation we can make is that the data BMI for males is less dispersed than females and thus they show less variation with respect to BMIs. This is probably the reason why males seem to be showing relatively less outliers too. Females on the other hand seem to be showing a little negatively skewed dispersion

[3] I deviate from the instruction slightly, as I tend to avoid variable names with underscores. This has advantages in certain programming languages, especially TEX and friends where it is a special character.

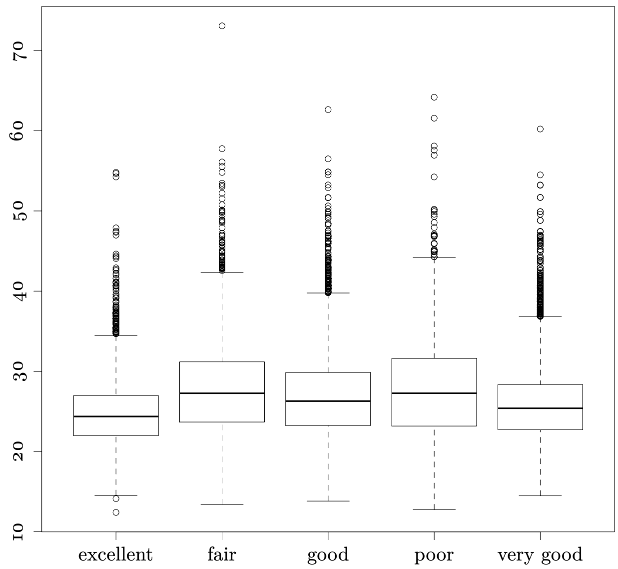

Figure 4: Box-plot of BMI to health

and have a lot of outliers on the positive end some deviating from the median way too much.

3.3 Introduction to linear regression

3.3.1 Exercise 1

To check whether a linear model will fit or not, plotting a scatter plot can help (cf. figure 6). It looks like the data is linear, because the increase in at-bats also gives a rise in the runs, but the variation in runs is still not absolutely neat and regular. There is a considerable amount of spread as seen in the scatter-plot. Considering that the correlation is mild, it would probably not be the sole factor affecting the runs and hence I would not use it for predicting runs.

3.3.2 Exercise 2

As mentioned earlier, there seems to be a linear relationship. Its strength as per the correlation-value is 0.61 which isn't a very strong correlation, but nevertheless it is a positive correlation.

3.3.3 Exercise 4

The summary for the `homeruns` variable can be seen below:

```
Call:
lm(formula = runs ~ homeruns, data = mlb)
```

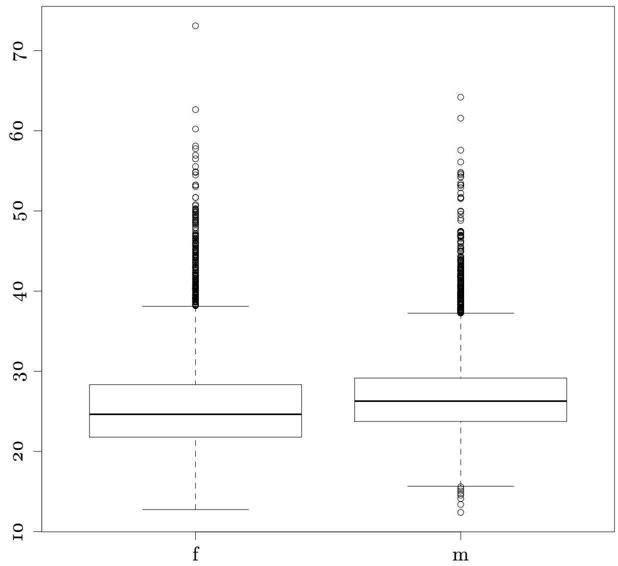

Figure 5: Box-plot of BMI to gender

```
Residuals:
    Min      1Q  Median      3Q     Max 
-91.615 -33.410   3.231  24.292 104.631 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 415.2389    41.6779   9.963 1.04e-10 ***
homeruns      1.8345     0.2677   6.854 1.90e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 51.29 on 28 degrees of freedom
Multiple R-squared:  0.6266,	Adjusted R-squared:  0.6132 
F-statistic: 46.98 on 1 and 28 DF,  p-value: 1.9e-07
```

Based on this, the equation for the regression line can be written as:

$$\hat{y} = 415.24 + 1.83 \times homeruns$$

The slope for home runs tells us that for each home-run we see 1.83 times runs, which looks like a very neat and stable correlation as opposed to the `at_bats` one. Using this variable for predicting score is much more reliable. Thus success-rate of a team can also be predicted with this.

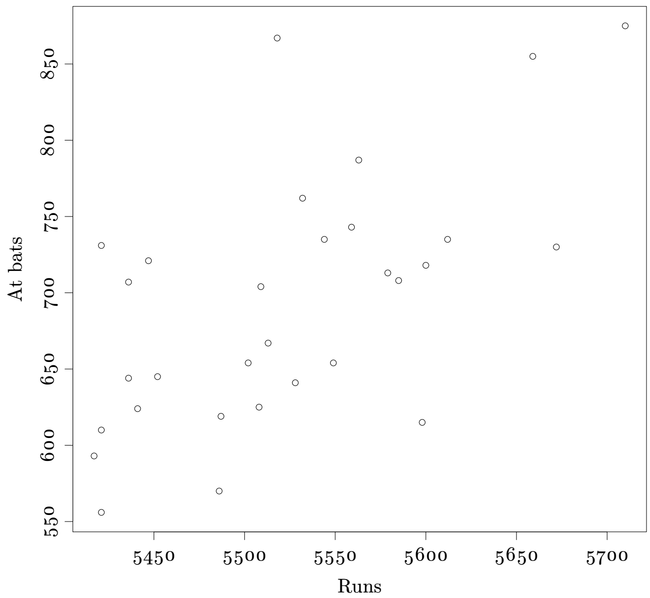

Figure 6: Major League Baseball

3.3.4 Exercise 5

If the team manager just looks at the regression line, they would predict around 725 runs, but the actual data says that they are near 700. So this is an overestimate and the residual is ≈ 25.

3.3.5 Exercise 6

There is no apparent pattern in the plot of residuals (cf. figure 7). This tells us that there is *constant variance.*

3.3.6 Exercise 7

Looking at figure 8 and figure 9 it seems that residuals are normally distributed. The representation we have got is unimodal. It is fairly symmetric, so yes, the *nearly normal distribution condition* does seem to meet here.

3.3.7 Exercise 8

The variability looks pretty less except in some cases. Thus the constant variability condition does not seem to be met.

4 Implementation

In this section, we will discuss how we have coded our R-files. Since we have used DocStrip For generating the R-scripts, the explanation of the code will be a fusion of explaining a

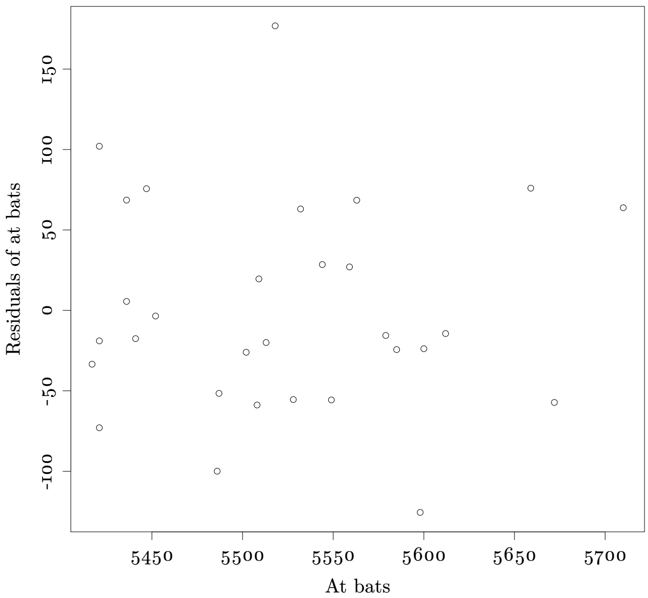

Figure 7: Residuals plot

few bits of DocStrip and a few bits of R.

To begin with, the DocStrip utility works on something called as '*tags*'. Any kind of source code is added in those tags. E.g., if I want to create a tag named `foo`, I will first issue `%<*foo>` add some code which is supposed to go inside this tag, end the tag with `%</foo>`. In the `.ins` file then, I can direct DocStrip to create a file by extracting the code seen in tag `foo`. A file can have code from multiple tags. I have segregated several layers of code and created various tags for this precise purpose. Our source code for R starts with a tag called `tikz` as seen below.

```
⟨*tikz⟩
library(tikzDevice)
⟨/tikz⟩
```

This R-code merely loads the library `tikzDevice` which is used to produce the Ti*k*Z code of the plots. Since this isn't required by all the Rscripts, I have created a tag (i.e., `tikz`) and only used it in the `.ins` for the files which require this library.

The first chapter relies on the arbuthnot.R data-set. We are advised to access the data-set from the web with the following R-command:

```
source("http://www.openintro.org/stat/data/arbuthnot.R")
```

Even though this looks like a convenient way, I prefer saving the data-set offline and then processing it. The aforementioned website collects data using cookies without users' knowledge and consent. Users come to know about it only (if/)when they read the 'Terms of use'. I save the required .csv data-sets in my working directory which not only saves

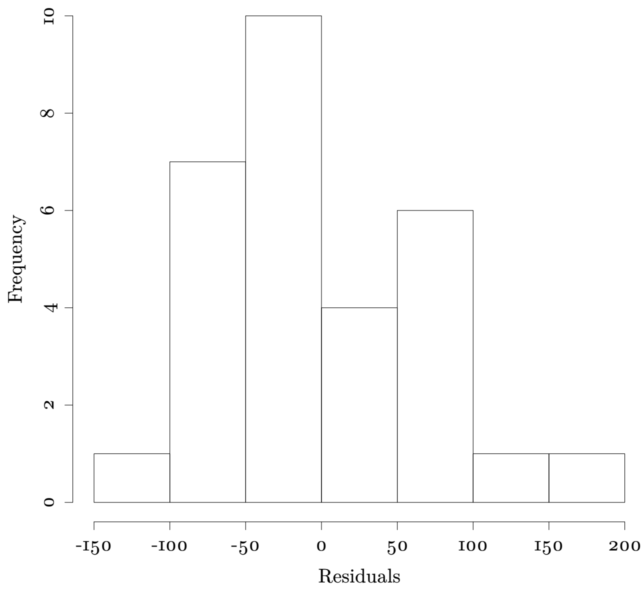

Figure 8: Histogram of residuals

the users from the trackers on the website, but makes the script executable without an internet connection too. For the first exercise the required data-file is `arbuthnot.csv`[4].

After loading the data-set, we need to retrieve the data of baptised girls. The data-set used by the tutorial is an `.R` file. As it isn't a `.csv` file, it requires special `R`-internal methods for saving the data-set which aren't commonly known. As users are advised to load the data-set with the `source()` function, it is highly likely that they don't even know the input of the loaded data-set. The problem is, often, we don't get such hand-tailored `.R` files. Instead what we usually get are the `.csv` data-files. For reading a plain `.csv` R has a function named `read.csv()`. I use it to save the data under a custom variable.

The first exercise of the first chapter asks us to find out a command for extracting the count of baptised girls in the data-set. I use `arbut` as a variable which reads the data with `read.csv()` function from the `.csv` file. As the data has a header-line, I have used `header=TRUE` option. The internal tag we have used here is `intro`, as the code inside it will be used by all the exercises of this chapter.

```
⟨*intro⟩
arbut <- read.csv(
  "arbuthnot.csv", ## name of the csv file to be read.
  header=TRUE      ## if the first row is a header.
)
⟨/intro⟩
```

After reading the data-set, I define a variable `arbutgirls` which stores the data of baptised girls. This is a variable which has the extracted data of all the baptised girls. For this, we use the variable of the data-set, i.e., `arbut` followed by a `$` sign which stands

[4] https://openintro.info/stat/data/arbuthnot.csv

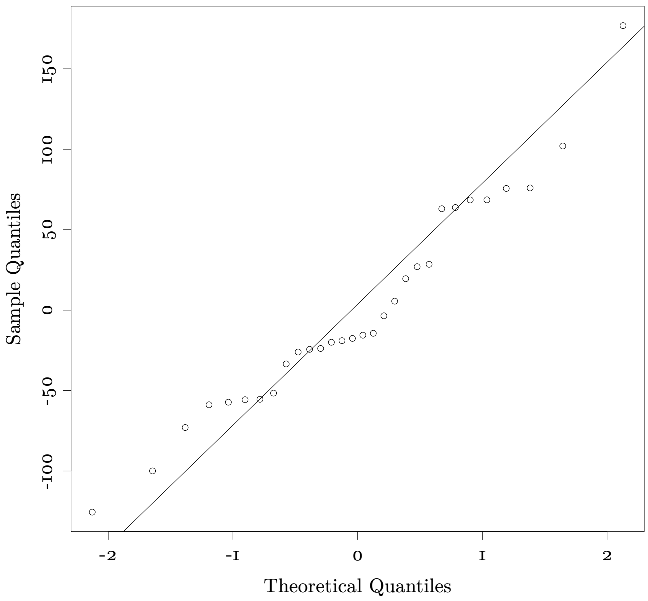

Figure 9: Normal probability plot of residuals

for a column. Then we write the column-name we are interested in (which is `girls` in this case). Since this code is only used for storing and printing the data of baptised girls; I have used the internal tag `girls`. This code is found in the `intro.R` file.

```
⟨*girls⟩
arbutgirls <- arbut$girls
arbutgirls
⟨/girls⟩
```

Then we name the file with the `tikz()` function for the output Ti*k*Z code. This line is enclosed in the `gplot` tag, so that we can omit it in other files which don't require it.

```
⟨*gplot⟩
tikz('girlsplot.tex')
⟨/gplot⟩
```

The following code is used to plot the graph for baptised girls. It can be seen in figure 1.

```
⟨*plot1⟩
plot(
  x=arbut$year,   ## the column to be plotted on the x-axis.
  y=arbut$girls,  ## the column to be plotted on the x-axis.
  type="l",       ## to get a line-plot instead of dots-plot
  xlab=("Years"), ## label to be given to the x-axis.
  ylab=("Girls")  ## label to be given to the y-axis.
)
⟨/plot1⟩
```

We use a similar approach for writing the source code for the plot with the data for boys. We use `bplot` and `plot2` tags for this.

```
⟨*bplot⟩
tikz('boysplot.tex')
⟨/bplot⟩
⟨*plot2⟩
plot(
  x=arbut$year,   ## the column to be plotted on the x-axis.
  y=arbut$boys,   ## the column to be plotted on the x-axis.
  type="l",       ## to get a line-plot instead of dots-plot
  xlab=("Years"), ## label to be given to the x-axis.
  ylab=("Girls")  ## label to be given to the y-axis.
)
⟨/plot2⟩
```

We need to read another data-set named `cdc.csv`[5] for the exercises found in the second chapter. For that the following code is used with `idata` as the internal tag-name.

```
⟨*idata⟩
cdc <- read.csv(
  "cdc.csv",      ## name of the csv file to be read.
  header=TRUE     ## if the first row is a header.
)
⟨/idata⟩
```

After this we are supposed to extract the number of cases and number of variables which we do with the following commands. We use the `stats` tag for this. It has all the code for the exercises asking for simple numerical statistics without any graphs or visuals. I have clubbed them all together. We use the `message()` function of R for printing text on the prompt. Running `idata.R` can demonstrate the effects. The `nrow()` and `ncol()` functions of R return the number of rows and columns respectively, from which most typically the former is the number of cases and the latter is the number of variables.

```
⟨*stats⟩
message(
  "--------------------------\n",
  "**Chapter 2: Exercise 1**\n",
  "--------------------------"
)
cases <- nrow(cdc)
variables <- ncol(cdc)
message(
  "Cases:     ",cases
)
message(
  "Variables: ",variables
)
```

As we have seen in the documentation, we can compute the IQR manually, but R has a function precisely with this name and unlike other common functions this is an all-caps-function, named `IQR()`. Thus to find and print the two IQRs, I have used the following code.

```
hiqr <- IQR(cdc$height)
```

[5] https://www.openintro.org/book/statdata/cdc.csv

```
aiqr <- IQR(cdc$age)
message(
  "\nIQR of the height variable is ",hiqr,"\n"
)
message(
  "IQR of the age variable is ",aiqr,"\n"
)
```

The following code gives us the summaries of heights and ages of the participants. The argument of the `summary()` function here is the name of the data-set, i.e. `cdc` here in both the cases followed by a `$` sign and the variable we are interested in.

```
message("\nSummary of height:\n")
summary(cdc$height)
message("\nSummary of age:\n")
summary(cdc$age)
```

The following code gives us the tables for gender alone and gender combined with the exercising variable. For the former we just pass one argument to the command and for the latter we pass two arguments which are to be compared. When two arguments are used, the first one forms the rows and the second one forms the columns.

```
table(cdc$gender)
table(cdc$gender,cdc$exerany)
⟨/stats⟩
```

We use the following code for the mosaic plot. We also add the parameter `santize=TRUE` for treating LaTeX special characters differently.

```
⟨*mplot⟩
tikz('mosaicplot.tex',sanitize=TRUE)
⟨/mplot⟩
```

The following command plots a mosaic plot for visualising the correlation between gender and smoking.

```
⟨*plot3⟩
mosaicplot(table(cdc$gender,cdc$smoke100))
⟨/plot3⟩
```

For extracting under 23 male smokers from the data we use the following code. It has three subsets. First we store the male participants in `mdata`, then from `mdata` we extract out under 23 participants and at last we take out smokers from them. The syntax for all the three commands is exactly the same. We have used the `subset()` function. Its first argument is the data-set which we want to refine and the second argument is the variable with the expected categorical value which we want to extract. Since the question is only about the command, I haven't included it in any of the R-scripts.

For calculating the BMI of the given data-set, we have used the following command. It will be used in every file because of the common `bmi` tag.

```
⟨*bmi⟩
bmi <- (cdc$weight/cdc$height^2)*703
⟨/bmi⟩
```

The following are the commands defining the filenames of two plots followed by the commands for actually plotting them.

```
⟨*hplot⟩
tikz('plot4.tex')
```

```
⟨/hplot⟩
⟨*splot⟩
tikz('plot5.tex')
⟨/splot⟩
⟨*plot4⟩
boxplot(bmi~cdc$genhlth)
⟨/plot4⟩
⟨*plot5⟩
boxplot(bmi~cdc$gender)
⟨/plot5⟩
```

OpenIntro recommends to download a data-set named `mlb11.Rdata` with the following command.

```
download.file(
  "http://www.openintro.org/stat/data/mlb11.RData",
  destfile = "mlb11.RData"
)
```

I have not followed this method for two reasons. One is again, as mentioned before, the website uses trackers. The second issue is the nature of the `.Rdata` file itself. It is a binary file which is opaque. Plain text files can be read with any text editors. They don't require special software for interpretation. `.Rdata` file stores the state of an R-prompt at any desired point. We get no information regarding how the coder managed to reach that stage. In my opinion it is harmful to the learning as well as software-wise distributing opaque files restricts the 'free' nature of the project. Hence I have supported the project with another `.csv`-file (`mlb11.csv`)[6] and entered all the required commands to reach the state it reaches with `mlb11.Rdata`. This also gives me more freedom for having my own variable-names.

```
⟨*rdata⟩
mlb <- read.csv(
  "mlb11.csv",      ## name of the csv file to be read.
  header=TRUE       ## if the first row is a header.
)
⟨/rdata⟩
```

We use `cor()` function for finding correlation. It takes two arguments which are the concerned variables.

We store the name of the Ti*k*z file in the `scatter` tag and the code for plotting it in the `plot6` tag.

```
⟨*scatter⟩
tikz('plot6.tex',sanitize=TRUE)
⟨/scatter⟩
⟨*plot6⟩
plot(
  mlb$at_bats,
  mlb$runs,
  xlab="Runs",
  ylab="At bats"
)
⟨/plot6⟩
```

[6] https://www.openintro.org/book/statdata/mlb11.csv

The following is the code for plotting figure 7.

```
⟨*resid⟩
tikz('plot7.tex',sanitize=TRUE)
⟨/resid⟩
⟨*mlbdata⟩
mlbatbat <- lm(runs~at_bats,data=mlb)
mlbdata <- lm(runs~at_bats,data=mlb)
⟨/mlbdata⟩
⟨*plot7⟩
plot(
  mlbatbat$residuals~mlb$at_bats,
  xlab="At bats",
  ylab="Residuals of at bats"
)
⟨/plot7⟩
```

The following is the code for getting two summaries from the `mlbdata` variable.

```
⟨*rstat⟩
mlbdata <- lm(runs~homeruns,data=mlb)
summary(mlbdata)
⟨/rstat⟩
```

The following is the code for the histogram. We use the `hist` tag for including it.

```
⟨*hist⟩
tikz('plot8.tex',sanitize=TRUE)
⟨/hist⟩
⟨*plot8⟩
hist(
  mlbdata$residuals,
  main="",
  xlab="Residuals",
)
⟨/plot8⟩
```

We use the following code for getting the normal probability plot. It is enclosed in the `qqnorm` tag.

```
⟨*qqnorm⟩
tikz('plot9.tex',sanitize=TRUE)
⟨/qqnorm⟩
⟨*plot9⟩
qqnorm(
  mlbdata$residuals,
  main=""
)
qqline(mlbdata$residuals)
⟨/plot9⟩
```

Every file having the Ti*k*Z-conversion of R-plots, should have a line closing off the device, hence we have used the following line with the generic tag `tikz` and loaded it at the end of all the graphic files with the `.ins`.

```
⟨*tikz⟩
dev.off()
⟨/tikz⟩
```

References

Diez, D. M., Barr, C. D., & Cetinkaya-Rundel, M. (2012). *Openintro statistics.* OpenIntro.

GNU Free Documentation License

Preamble

The purpose of this License is to make a manual, textbook, or other functional and useful document 'free' in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of 'copyleft', which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The '**Document**', below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as '**you**'. You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A '**Modified Version**' of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A '**Secondary Section**' is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The '**Invariant Sections**' are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The '**Cover Texts**' are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A '**Transparent**' copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not 'Transparent' is called '**Opaque**'.

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The '**Title Page**' means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, 'Title Page' means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

The '**publisher**' means any person or entity that distributes copies of the Document to the public.

A section '**Entitled XYZ**' means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as '**Acknowledgements**', '**Dedications**', '**Endorsements**', or '**History**'.) To '**Preserve the Title**' of such a section when you modify the Document means that it remains a section '**Entitled XYZ**' according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

C. State on the Title page the name of the publisher of the Modified Version, as the publisher.

D. Preserve all the copyright notices of the Document.

E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

H. Include an unaltered copy of this License.

I. Preserve the section Entitled 'History', Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled 'History' in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the 'History' section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

K. For any section Entitled 'Acknowledgements' or 'Dedications', Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

M. Delete any section Entitled 'Endorsements'. Such a section may not be included in the Modified Version.

N. Do not retitle any existing section to be Entitled 'Endorsements' or to conflict in title with any Invariant Section.

O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled 'Endorsements', provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled 'History' in the various original documents, forming one section Entitled 'History'; likewise combine any sections Entitled 'Acknowledgements', and any sections Entitled 'Dedications'. You must delete all sections Entitled 'Endorsements'.

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an 'aggregate' if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled 'Acknowledgements', 'Dedications', or 'History', the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See https://www.gnu.org/licenses/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License 'or any later version' applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Document.

11. RELICENSING

'Massive Multiauthor Collaboration Site' (or 'MMC Site') means any World Wide Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A 'Massive Multiauthor Collaboration' (or 'MMC') contained in the site means any set of copyrightable works thus published on the MMC site.

'CC-BY-SA' means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.

'Incorporate' means to publish or republish a Document, in whole or in part, as part of another Document.

An MMC is 'eligible for relicensing' if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

> Copyright © YEAR YOUR NAME.
>
> Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled 'GNU Free Documentation License'.

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the 'with … Texts.' line with this:

> with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.