Solution manual for excercises in Statistics the art and science of learning from data

Martin Summer

Jan 07, 2024

CONTENTS

This is a manual with worked solutions for all the exercises and project steps in the course "Statistics: The art and science of learning from data". The manual is for the use of on-site learning facilitators to support students in their efforts to work on exercises and assignments.

Note that there are usually several ways to solve a problem, in particular how to implement certain tasks in R. So the solutions worked out here should be seen as a suggested solution.

I have taken great efforts to assure that all exercises and problems can be solved and when numerical answers are asked for that they are correct. Should you despite of all this detect a mistake, please report the issue to martin.summer@chello.at

CHAPTER

ONE

EXERCISES: UNIT 1, CATEGORICAL DATA AND PROPORTIONS

1.1 Pen and paper exercises

1.1.1 Exercise 1: Introduce yourself

My name is Martin Summer. I live in Vienna, the capital of Austria in Europe with my wife Gabriele and my two sons Paul, 17 and Peter, 15. I am an economist by training and profession and I am currently heading a research department at Oesterreichische Nationalbank, which is the central bank of Austria and together with the European Central Bank part of the Europan system of central banks. I came to statistics and data analysis through my work, where I use these tools to do research, advise policy and more generally trying to understand what is going on in the world more broadly.

I have admired Jesuit Worldwide Learning as a project and institution for many years. This at some stage led to a cooperation when JWL asked and entrusted me to develop a statistics course for their program. The course you are taking now is the first result in this attempt and I hope you will enjy it and find it interesting and useful. I love teaching and I love statistics and this is a key reason why I decided to take on this honorable assignemt. Of course I also want to support you in your desire to learn by giving the best I can from my experience. I also want to support my Jesuit friends who founded and maintain this wonderful project, which I truely admire and love.

Statistics impacts my life in many ways: As a source of wonder and a well of knowledge about the world, as an amazing field of insight and inspiration, to which many of the greatest minds in human history, both theoretically and practically have contributed. It is just wonderful to partcipate in this human project of knowledge. Statistics also impacts my professional life, where it is one tool of the trade in economic analysis and policy advice. At the moment I am together with colleagues at OeNB are engaged in a research project where we uses statistical methods to find out whether people in Austria would be willing to adopt a digital form of the Euro, our money which we currently just as banknotes or through accounts at commercial banks. Isn't it almost like magic that statistics allows us to glimpse not only into an unknown future but also into potential behavior of people in using a payment method - the digital Euro - that does not yet exist at all but as an idea?

1.1.2 Exercise 2: Computing the percentage decrease in infant mortality rates of 8 European countries between 1860 and 2020

Here are the numbers we discussed in the text in a table. I named the variables for the infant mortality rate in this exercise MR-1860 for the infant mortality rate in 1860 and MR-2020 for the corresponding mortality rate in 2020.

Country	Continent	MR-1860	MR-2020
Austria	Europe	0.237	0.0030
Belgium	Europe	0.139	0.0034
Denmark	Europe	0.136	0.0031
France	Europe	0.150	0.0034
Germany	Europe	0.260	0.0031
Norway	Europe	0.102	0.0018
Spain	Europe	0.174	0.0027
Sweden	Europe	0.124	0.0021

In case you wonder how I did this table in the Jupyter notebook, look at the cell input and study the code. For textcells in Jupyter you can use a markup language called Markdown. Mardown will allow you to format text by using certain markup symbols that the Jupyter notebook can interpret and process. Here is a link to a cheat sheet summarizing the most important markup instructions: https://www.markdownguide.org/cheat-sheet/

I encourage you as facilitators to familiarize yourself with markup and teach it to the students you work with when working with the Jupyter Notebooks.

The infant mortality rate in Austria was 0.237 in 1860 and was 0.0030 160 years later in 2020. The percentage decrease over this period is computed as: \begin{equation*} \frac{0.237 - 0.003}{0.237} \end{equation*}

you can compute this by hand or by calculator or by R. Let me use R for the sake of the example

```
((0.237 - 0.003)/0.237)*100
```

$$98.7341772151899$$

This is a reduction of roughly 99 % percent. One could also say that the mortality rate decreased by a factor of 78. If you express the change as a relative change in percent you see how enormous the reduction was over a period of 160 years. This is next to a miracle, after centuries of extremely high rates.

In case you are uncertain how to compute percentage changes, let us discuss some examples:

Example 1: Suppose you have worked in a Supermarket for a wage of 10 (in some monetary unit) and your pay is increased to 12. What is your pay increase in percent? The way to compute the percentage increase is: \begin{equation*} \frac{(12-10)}{10} \end{equation*}

which is 0.20. Take this times 100 and you have the increase in percent, which is 20 %.

Example 2: The staff of a company decreased from 40 to 29. What was the percentage decrease in its labor force? \begin{equation*} \frac{(29 - 40)}{29} \end{equation*}

which is 0.275. Take this by 100 and you have 27.5 %.

Did you see the general pattern. When computing a percentage change, the formula you apply is

\begin{equation*} \frac{(final - initial)}{initial} \times 100 \end{equation*}

This is the formula for a percentage change. When this change is positive, we say naturally that we have an increase, when it is negative we say we have a decrease and fold the minus sign into the workd decrease. So in the second example, we would say that the labor force of the company decreased by27.5 %.

When the percentage change is 100 it means that something doubled, when it is 1000 it means that something has increased 10 fold, or as in the case of infant mortality in Austria between 1860 and 2020 has decreased 78 fold.

We can also do the computations with R based on what we have learned in the course so far. Let's first store the values in two objects, which we call 'mr_1860' and 'mr_2020'.