Exercise 1-Solutions
TMA4255
Applied Statistics

January 16, 2017

Intro

0.1 Start MINITAB

Start MINITAB on your laptop, or remote desktop to cauchy.math.ntnu.no and log in with

win-ntnu-no\yourusername
(notice that its not /) and password.

When you start MINITAB you are presented with an empty set-up (project) with a session window on top and a worksheet below. In the session window the results from our statistical analyses are presented (and we may also write commands). **In the worksheet we enter or read in data, and some results may also be written here**. We may save our worksheet(s) and plots and session window as a project.

0.2 New project

First, create a folder called TMA4255. Let us now start a project, called Ex1.PRJ, and put it in this folder.

Commands: *File − > Save Project As − > Ex1.PRJ*

0.3 General information:

- The default decimal separator in MINTAB is comma. That is, 15/2==7**,**5 NOT 7.5
- In this course we only need to use the menus to execute commands, but if you set Editor-Enable Commands the commands used will be echoed in the session window even though you used the menus. However, if you need to do programming - you should instead go for R (not MINITAB).

- We go through the menus together - orally.

1 Problem 1

1.1 Reading data

Read in the data for Exercise 1 from the file data1.MTW. Go to the TMA4255 WWW-page at the Exercises tab, and save the data1.MTW file to your new TMA4255 folder.

Alternatives:

1. **Commands:** *File − > Open Worksheet − > Go where data1.MTW is saved*

2. Or, you may just click on the file and it opens in a new MINITAB window (then you need to do the new project step again).

NOTE: A file with extension .MTW is a Minitab Worksheet. If you want to open other files you need to choose *File − > Other Files − > Import special text*. The open-icon will open project files (.PRJ extension). Now you have a

project with data in the worksheet. Open the Project Manager, and see that you now have a data1 worksheet and many columns with names and values. Notice that there is also a folder for constants and for matrices.

1.2 Analyzing the data (1a)

Run the following commands:

Commands: *Stat − > Basic statistics − > Display Descriptive statistics*

Double click on the variable you want to analyze on the left panel. It will automatically jump to the "variables" box. On the option "Statistics" , we can choose which statistic you want to calculate. Accordingly, on the option "Graphs" you can check the boxes for the graphs you want to generate. The graphs can also be generated by the following command:

Commands: *Graph − > choose your graph*

NOTE: Look at *Assistant − > Graphical Analysis* for more information about the graphs.

1.3 The normal plot (1b)

To check if the data in normally distributed run the following command:

Commands: *Graph − > Probability plot*

Choose "single", then choose your variable (Co). You can choose the distribution you want by clicking "Distribution". If you choose Normal distribution, and then click OK, you will get a plot whose dots are outside the confidence intervals, this is an indication of a bad fit. That is, your data is not normally distributed. Try choosing "lognormal", what do you see?

May also look for trend in data with:
Commands: *Graph − > Time Series Plot*

1.4 Transformations with the Calc tab (1c)

For any transformation of the data do:

Commands: *Clac − > Calculator*

First, choose where you want to save your results, "store results in variable" and give a name of an empty column, like C3 for example. Then find your transformation from the "Functions" panel (note: for ln, choose log base 10). Double click on the function and then from the left panel you choose the variable to be transformed, double click on it. Finally, in the "expression" panel you should have something like LOGTEN("C1") for example.

For the three transformations save the results into columns C2, C3 and C4. Repeat the steps above to produce normal plots. You may produce multiple plots at once, but deselect graph variables form groups.

2 Problem 2

2.1 Simulate normal data (2a)

When you simulate normal data you don't need an already existing data! To simulate normal data with mean 2 and sd 4 do:

Commands: *Clac − > Random Data − > Normal*

Fill in the mean and the standard deviation. In "Number of rows of data to generate" fill in 50. And in "Store in columns" write C5-C6. Present descriptive stats and plots as in exercise 1, for each column.
NOTE: You have generated 2 data, one in column C5 and one in C6.

2.2 Look up probabilities or critical values (2b)

Here we will use the command:

Commands: *Clac − > Probability Distributions − > your distribution*

For I-III you will choose the "Input constant" panel, there you will put 32, or 28 accordingly. "Input column" needs a column input, you don't have it. Furthermore, you fill in the characteristics of the distribution, like "number of trials" or "event probability" for the binomial distribution. Finally, if you want to calculate P(X=x) you choose "Probability" and if you want P(X$\leq$ x) you choose "cumulative probability".

NOTE: For point 2, you need to calculate $P(X \leq 32) - P(X \leq 27)$. Calculate each one of them and then subtract the one from the other.

NOTE: For the third point, you have to approximate the Normal distribution, that is, find the mean and the standard deviation. This is approximated by $N(np, np(1-p))$. Then do points 1 and 2, but choose Normal and fill in the mean and the sd you have found.

2.3 Plotting distributions (extra)

Some times it is nice to look at the probability distributions we work with. For example we may plot a T distribution with 8 df, and show the right 5% tail.

Commands: *Graph − > Probability Distribution Plots − > your distribution*

Then choose "single". Choose your distribution and fill in its characteristics.

3 Problem 3

3.1 Generating

To generate the data here, do as before but: (1) in the "number of rows of data to generate" fill in 500, (2) in the "store in columns" write "C11-C20". Remember the mean and the sd.

3.2 3a

For each row we want to calculate the maximum value:

Commands: *Calc − > Row statistics*

Then choose "maximum", in "input variables" write C11-C20 and in "Store results in" write C21. Now we have the maximum for each row. Then do:

Commands: *Stat* $->$ *Basic statistics* $->$ *Display Descriptive statistics*

As before, and the result in the "mean" is the estimate of the mean.

3.3 3b

Finally, in 3b, we may solve this theoretically (see solutions) or by sorting the data in C21 and using the number of observations smaller than or equal to 205 divided by 500 as an estimate for this probability. For sorting the data do:

Commands: *Data* $->$ *sort*

Both in "Sort columns" and in the first "By column" fill in C21. Click OK, then count the total number of elements that are ≤ 205.

Making a report (extra)

To make a report you may either use the ReportPad or copy and paste into you favorite software for writing reports (Word? or Latex?).

If you do not want to use ReportPad:
In MINITAB: Graph window selected
Edit- Copy Graph
In writing software: paste

Or, if you want a specific format for the graph choose File-Save Graph as and choose .PNG or what you want, and import into your writing software.
With ReportPad:
Or, you right-click on a graph that you want to be included in your report using ReportPad and choose Append graph to report, you may also add sections from the session window by right-click and Append Section to Report. By clicking on the ReportPad in the ProjectManager you may choose Save Report As or Print Report. The ReportPad can be saved as txt or html. Save your report as a .html file and open it in your WWW-browser.

Saving and retrieving projects

Remember to save your project before you quit MINITAB. Maybe you can have one PRJ-file for each exercise in TMA4255?