XMUT321 Lab 1_2 Statistics using Matlab

(Notes and Instructions to accompany Lab 1_2)

Lab 1_2 provides some background information and guides on how to use Matlab for Statistics. 2 weeks have been allocated for you to go through the materials from page 1 to page 30 in the Lab sheets (the rest of pages are for your own references).

This is a no-mark Lab, the main purpose is for you to learn how to use Matlab for Statistics. You will go through the following suggested schedule, and show the Lab instructor your **Submission** at the end of each week.

Please note: This Lab 1_2 provides a general introduction on how to use some built-in Matlab commands for Statistical computations, and it is not exhaustive. You may find the materials here useful for the subsequent Labs, but it does not necessarily mean that you must only use the commands learnt in this lab, or the commands learnt here will definitely be used in the subsequent Labs. The built-in commands, while convenience to use, behave like a "black-box" where you do not know how the codes have been written. This introductory Lab is therefore to inspire you to explore more on your own. The best way to learn is to practise using your critical thinking skills, and write your own Matlab codes.

1st Week: Descriptive Statistics (Page 1 to 11)

Here you will learn how to read the provided rain.txt file into Matlab, and use the data to derive statistical parameters, such as mean, standard deviation etc, and use the data to plot diagrams showing the distribution of the data. Refer to Chapter 1 of textbook for more details.

The file rain.txt contains the yearly maximum hourly rain storm depths in mm at Genoa University, Italy, collected from 1931 to 1988. The first column is the 'year' and the second column is the 'rain storm depth in mm'. In order to use these data, we need to read the .txt into Matlab, by using the command

[year, rain] = textread('rain.txt');

That will produce two separate columns of data (58 rows) with the title 'year' and 'rain' in the workspace of Matlab. You can then follow the rest of investigations from Page 1 to 11.

Note: on page 10 of the Lab sheets it mentions data for rainfall and runoff but where are the data??

Answer: scroll down to page 31 and you will find the table for rainfall and runoff; try create your own .txt file and follow the steps above in order to read the data into Matlab. Otherwise, a copy of the table can be downloaded from the course website.

Submission: **demonstrate to the Lab instructor the graph for boxplot(rain).**

2nd Week: Some Common Distributions (page 11 – 18)

Here you will explore Binomial, Poisson and Normal distributions. Note: we will not study these distributions until a bit later in the course, but you should have already learnt these earlier, so treat this opportunity as a revision. You can refer to textbook Chapter 4 for more details about the distributions.

a) Binomial distribution:

Here, we are conducting 20 trials (with 20 water samples) to find out if the water samples contain bacterium. We know that the probability that the water sample contains bacterium is 0.5. So out of the 20 trials, we are interested to know the probability of exactly 1 trial contains bacteria, 2 trials contain bacterium, and so on. Looking at Figure 4.1, what observations can you make?

Are you interested to know how to plot Figure 4.1? Here is how:

```
x=-5:25;                % create a vector of number of trials that we are interested
y = binopdf(x,20,0.5);  % use binopdf to create a set of probabilities (not just one)
bar(x,y,1)              % use bar to create the plot
```

Question: why do we use the command 'bar' instead of 'plot'?

b) Poisson distribution:

Follow the steps above in Binomial distribution, **see if you can plot Figure 4.4**?

c) Normal distribution:

Here is how to plot Figure 4.7:

```
x=2380:2520;            %create the range of x-axis
mu=2445;                %set the mean value
sigma=16;               %set the standard deviation
y=normpdf(x,mu,sigma);  %use normpdf to create a set of probability
plot(x,y)               %plot the probability
```

Submission: think about how to shade/colour part of the graph as shown in Figure 4.7, and show it to your Lab instructor.

Simulation using Matlab (page 19-27)

Here, you will learn how to use Matlab to simulate a set of data according to the statistical distributions that you want. These include Uniform distribution, Bernoulli distribution, Normal / Lognormal distribution etc. Refer to Chapter 4 in textbook for more details on those distributions.

Please study the Lab materials and go through the exercises.

Submission: show the Lab instructor a plot of your Figure 5.4.

If you have time, continue from page 28 to 30 on Statistical Tests.

~ end ~