MD/PhD Course in Medical Statistics

Department of Statistics

Practical session on Minitab, descriptive statistics and the Normal distribution

The first step is to download a data file from the web. This is found on the Web page of the course, namely http://www.mas.ncl.ac.uk/~njnsm/medfac/MDPhD/notes.htm. Click on the data link under the Descriptive data and Normal distribution practical. This will give you a Minitab (version 11) save file called HEIGHT.MTW, which you should save somewhere suitable (do not work from a floppy disk).

1.
The data in HEIGHT.MTW are the heights of the 99 five-year-old boys used in the lecture. Use MINITAB (version 11) to compute the mean and SD of these data. Also find the median and quartiles and draw a Box and Whisker plot.

{n.b. the lecturer will go through this with the whole class by way of introduction to the MINITAB program}

2.
The file HEIGHT.MTW can be read only by MINITAB. If you had data that you wished to analyse you would need another method to enter the data in order to produce the .MTW file in the first place. This can be done by typing data directly into the Data Window. Try this by typing the following 15 numbers into column C2 (they are haemoglobin concentrations in g/dl).

14.2 14.4 14.6 14.8 13.4 12.9 13.2 16.6 15.8 17.8 15.8 16.0 15.9 17.0 16.4

Name the column and save the worksheet. Find the mean and SD. You are now told that some of the values are from males and some from females. This information can be entered by typing the following information (1 denotes female, 2 male) into another column.

1 1 1 1 1 1 1 2 2 2 2 2 2 2 2

Use the By variable option in the Descriptive Statistics box to obtain means and SDs separately for the sexes.

3.
A variable widely used in the assessment of photo-toxicity is distributed in a certain population with a Normal distribution having mean and SD 1.78 and 0.172 respectively (log units). What proportion of the population have a value less than 1.65 units?

{You need to compute the cumulative probability at 1.65 of a Normal distribution with the given mean and SD. To do this use the menu path Calc → Probability Distribution → Normal…. Remember to click the Cumulative probability button and enter the appropriate mean and SD: you will also need to enter 1.65 as an Input constant:.}

4.
For the variable in question 3 find the value such that 77.5% of the population is below this value. How is this related to question 3?

{Here you need the Inverse cumulative probability function: make sure you check the right boxes}.

5.
In a Normal distribution of mean 10 and SD 1, find the value such that 5% of the population have values below this threshold. Generate a sample of 100 values from this population.

{To do this use the menu path Calc → Random Data → Normal… and tell the command to Generate 100 rows of data which are then stored in a column you specify}

What proportion of values are below the threshold you calculated?

{This can be done using Sort commands under the Manip menu but a quicker way is as follows. Suppose the generated data are in the column C1, C2 is free and the numerical value of the threshold is *x*. In the Session window at the MINITAB prompt (MTB>) type the command:

MTB> let C2=(C1<*x*) (followed by return)

This will result in C2 containing 0s and 1s. The 1s will be in rows where the condition C1<*x* is true and the 0s where it is false. That is, C2 is 1 only for values of the variable below the threshold, so the proportion of 1s in C2 is the required answer. There are various ways of finding out how many 1s and 0s there are in C2. Perhaps the simplest is to use the menu path Stat → Tables → Cross tabulation…, enter C2 as the Classification variables: and click on OK.}

Repeat this exercise with a new column of 100 value and also with a new column of 10000 values. Comment on your result.