

QUEEN MARY, UNIVERSITY OF LONDON

MTH 4106

Introduction to Statistics

Practical 7

28 February 2012

Today's practical does two different things:

- plotting the probability density function of some normal distributions for you to insert in your lecture notes (this part continues from Practical 6);
- plotting three or four variables (this part continues from Practical 3, and will not be covered in lectures).

1 (Plotting the standard normal probability density function) Use

Graph $\rightarrow$ Probability Distribution Plot...,

click on View Single and then choose the appropriate options for the standard normal distribution.

2 (Plotting another normal probability density function) Now suppose that $X \sim N(4, 25)$. What interval on the x -axis should lie under $99\frac{3}{4}\%$ of the area under the curve?

Input the appropriate parameters for Minitab to plot the probability density function of X . Does this graph agree with your prediction above? If not, why not?

3 (Printing the graphs) Print off those two graphs, either directly or via the **Report Pad**. Then Exit from Minitab and start again.

4 (The pulse data) We are going to do some more plots using the pulse data. There are two ways to obtain the data. One is to open the project that you saved at the end of Practical 3: see Practical 2 for how to do this. The other method is to simply open the worksheet **Pulse.MTW** again, just as you did at the start of Practical 3. In this case you will need to recalculate the variable **Diff**. Whichever method you choose, it would be sensible to reread your notes on Practical 3 to remind yourself about the data.

5 (Three qualitative variables) Is a person's activity level affected by whether or not they smoke? If so, is the effect the same for men and women? Cross-tabulate the three relevant variables, using the two two-level variables **Smokes** and **Sex** to label the rows and the single three- (or four?)-level variable **Activity** to label the columns. You should find that there is one row for each combination of levels of **Smokes** and **Sex**.

Use this table to find out what percentage of the non-smoking men report the highest level of activity.

--

6 (One quantitative variable and two qualitative variables) Here we do a similar trick, drawing a separate boxplot for each combination of the levels of the two qualitative factors. Do this for **Diff**, **Activity** and **Sex** by choosing **One Y, With Groups** in the dialogue box for boxplots and then entering both **Activity** and **Sex** as the **categorical variables**.

What difference does it make if you enter the 'categorical variables' in the opposite order?

--

7 (Two quantitative variables and one qualitative variable) In Practical 3 we made a scatterplot of **Weight** against **Height**. Which of the points correspond to smokers, and which to non-smokers? To see this, redo the scatterplot, clicking **With Groups** and entering **Smokes** as the **categorical variable**. Smokers and non-smokers are then plotted with different symbols. Which is which? Does smoking make any difference to the relationship between weight and height?



8 (Three quantitative variables) For this, we really need a three-dimensional scatterplot. Of course, this is impossible on a piece of paper or a flat screen. But Minitab can pretend to plot the variables in three dimensions and just give you a two-dimensional view of it. Use

Graph → 3D Scatterplot...

and choose **Simple**. Enter **Pulse1** as the **Z** variable, **Weight** as the **Y** variable and **Height** as the **X** variable. Give the scatterplot a suitable title.

I find this diagram quite hard to interpret. But we can get more information by changing our viewing angle. Do this by using the **3D Graph Tools**.

Find an angle that you find helpful, and save that version of the graph in your **Report Pad**.

9 (Three quantitative variables and one qualitative variable) We can now add the qualitative variable **Sex** to the previous scatterplot by doing it again and choosing **With Groups**, so that men and woman are plotted with different symbols. Again, the display is not very easy to interpret, but it is clear that the men are generally taller than the women.

10 (Two quantitative variables and two qualitative variables) Now we plot the two quantitative variables on a scatterplot, and use the previous trick of doing something different (in this case, a different plotting symbol) for each combination of the levels of the two qualitative factors. Try this for the scatterplot of **Weight** versus **Height**, including the variables **Smokes** and **Sex**.

After appending this graph to your **Report Pad**, write a few lines after the graph to explain what each of the coloured symbols represents.

11 (That's enough variables!) As you can see, the display (and its interpretation) get quite difficult when there are four or more variables. There are some more advanced statistical techniques that let us reduce the dimension, but we won't cover them in this module.