GER1000 QUANTITATIVE REASONING

TUTORIAL 2

Question 1

This problem is about the height of 1,078 father-son pairs presented in Chapter 2. You will need to download the file "father-son-ordered.xlsx" that is zipped together with this tutorial question paper (not the one used in lecture) from the IVLE, and work through it using EXCEL. If your computer does not have EXCEL, you can download it for free using your NUS account, or access it via www.office.com

(a) (i) Calculate the values of the following quantities in EXCEL, to two decimal places. Use the function STDEV.P() to calculate the standard deviation (SD).

Quantity	Value	EXCEL command
Average of fathers' heights		
SD of fathers' heights		
Average of sons' heights		
SD of sons' heights		
Correlation coefficient		

(ii) Calculate the values of the quantities in the table below, but only for the part of the data corresponding to fathers of height 68 inches, i.e., fathers with height in the range 67.5--68.4 inches.

Quantity	Value	EXCEL command
Average of sons' heights		
SD of sons' heights		
Correlation coefficient		

(iii) Do the same as (ii), but for the data corresponding to fathers of height 69 inches: in the range 68.5—69.4 inches.

Quantity	Value	EXCEL command
Average of sons' heights		
SD of sons' heights		
Correlation coefficient		

(b) (i) Let h be the average height of sons whose fathers were 73 inches, i.e., with height 72.5 – 73.4 inches. We will predict the value of h in the following way. Let d be the average height of the sons in (a)(iii), minus the average height of the sons in (a)(ii). Assume that d is the amount of increase in the average height of the sons for every increase of 1 inch in the fathers' heights. What is your prediction of h?

(ii) Use EXCEL to calculate the exact value of h, like in (a)(ii) and (a)(iii). Then calculate the prediction error for (b)(i), defined as

h – (prediction of h)

(c) The regression line of son's height on father's height is given by the equation

predicted son's height = 0.514 (father's height) + 33.893

Use the regression line to obtain another prediction of h. Is it better than the prediction in (b)?

Challenge question: (i) Use the EXCEL instruction in the Appendix to construct the scatter diagram and obtain the equation of the regression line. (ii) How is the slope 0.514 related to the values obtained in (a)(i)? (iii) What does the equation give for (son's height) when substituting (father's height = 67.69).

Question 2

In October 2012, the *New England Journal of Medicine* published "Chocolate Consumption, Cognitive Function, and Nobel Laureates" by FH Messerli of Columbia University. The main findings are based on a scatterplot, reproduced below.

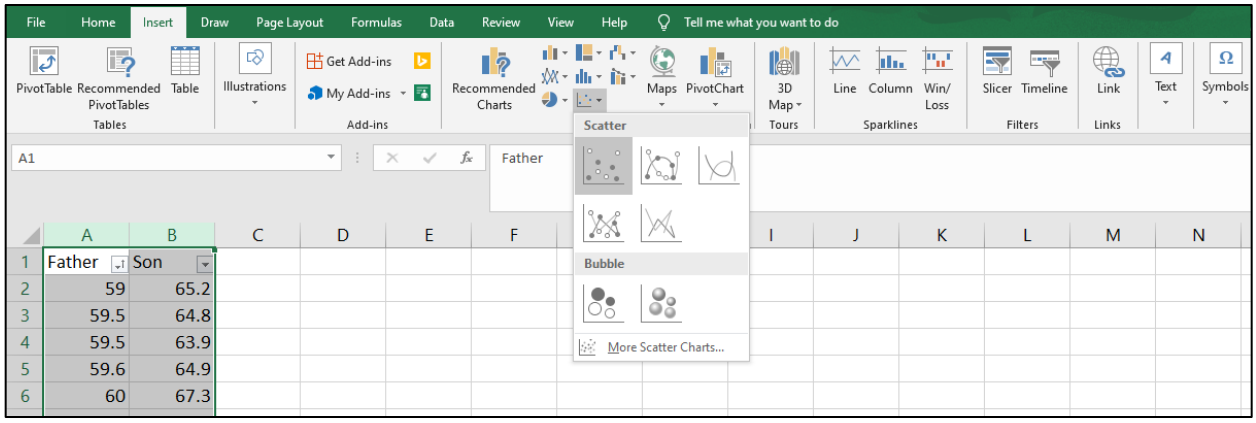
Obtain a PDF copy of Messerli's publication from the Tutorial 2 folder you have downloaded. Read the article briefly, and answer the following questions.

(a) Why did the author mention "socioeconomic status" and "geographic and climatic factors" in the Discussion section? How did he deal with the issue, and how would you?

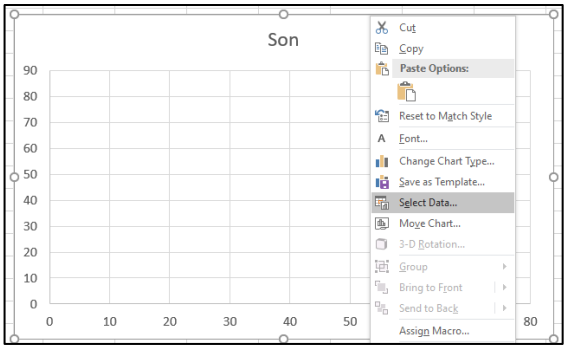
(b) What issue was brought up in the first sentence in the Study Limitations section? How would you deal with it, if you were to replicate the study on Nobel prizes and chocolate consumption?

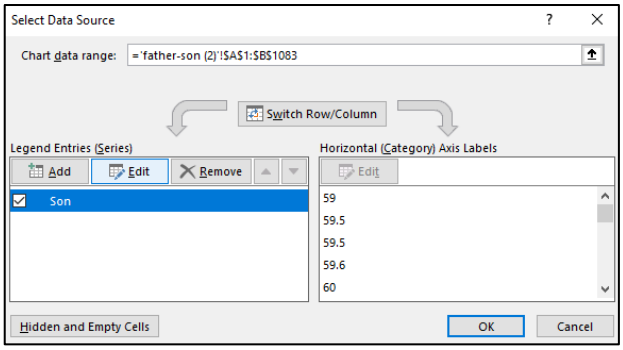
Appendix: How to plot a scatter diagram and show the regression line on Microsoft EXCEL (only relevant to the challenge question)

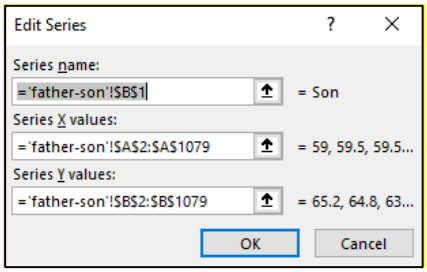
Step 1: Select the columns A (fathers' heights) and B (son's heights) and press the Scatter plot (shown in grey in the image below) under the *Insert* tab. You can centralise the data, by clicking on the axes, then in Axis Options, changing the values in Bounds.



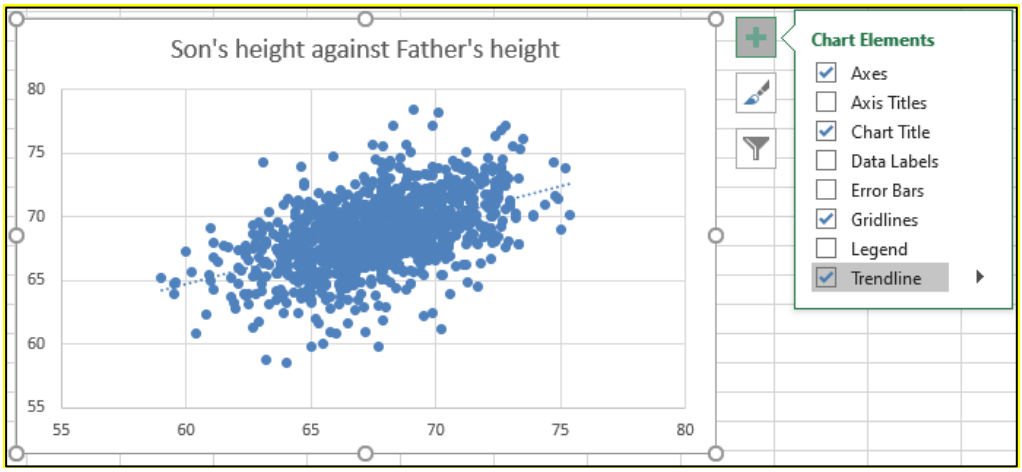
Step 2: EXCEL would take entries in columns A and B as the values of x and y respectively. However, you can always check it by right-clicking the scatter plot and press *Select Data*. Subsequently, click on the series and press *Edit*. You should see that Series X values are from column A and Series Y values are from column B. Note that the values should start from row 2 and end at row 1079, since there are 1078 pairs of values.







Step 3: Back to the scatter plot, insert the trendline by selecting the chart and clicking on the "+" and check the "Trendline" box. Note that the axes are adjusted in the image below so that it is easier to see the trendline.



Step 4: To show the equation of the line, click on the triangle beside it, select "More Options" and check the "Display Equation on chart". You would see the equation of the line on your scatter plot.