**GER1000 QUANTITATIVE REASONING**

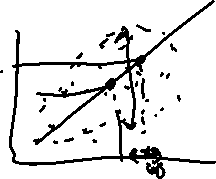
**TUTORIAL 2**

**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](http://www.office.com)

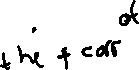
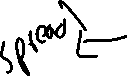
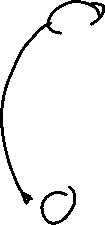
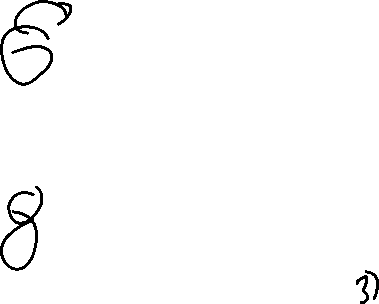
Group 4

1. (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 (Grp 4) |
| Average of fathers’ heights | 67.69inch | =AVERAGE(A2:A1079) |
| SD of fathers’ heights | 2.74 inch | =STDEV.P(A2:A1079) |
| Average of sons’ heights | 68.68 inch | =AVERAGE(B2:B1079) |
| SD of sons’ heights | 2.81inch | =STDEV.P(B2:B1079) |
| Correlation coefficient | 0.50 | =CORREL(A2:A1079, B2:B1079) |

(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 | 69.06 | =AVERAGE(B498:B650) |
| SD of sons’ heights | 2.75 | =STDEV.P(B498:B650) |
| Correlation coefficient | 0.01 | =CORREL(A498:A650, B498:B650) |



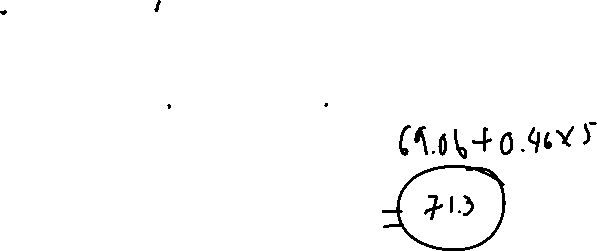
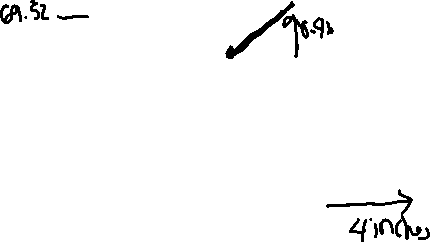
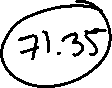
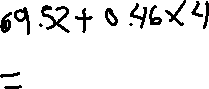
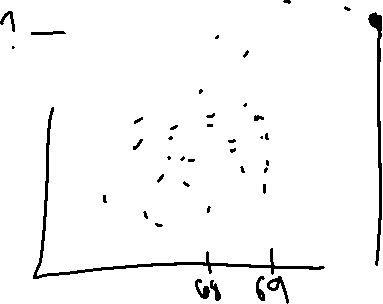
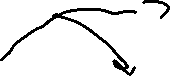
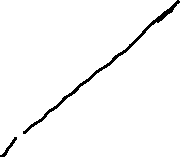
(iii) Do the same as (ii), but for the data corresponding to fathers of height 69 inches: in the range 68.5—69.4 inches.



|  |  |  |
| --- | --- | --- |
| Quantity | Value | EXCEL command |
| Average of sons’ heights | 69.52 | =AVERAGE(B651:B793) |
| SD of sons’ heights | 2.23 | =STDEV.P(B651:B793) |
| Correlation coefficient | 0.02 | =CORREL(A651:A793,B651:B793) |

Group 1

1. (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?



D=0.459



H = Average of sons’ height when dad is 69 inch + 4(d)

H=69.5195+4(0.4581)

(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)

Prediction error:71.7871-71.356=0.431

Group 5

(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)?

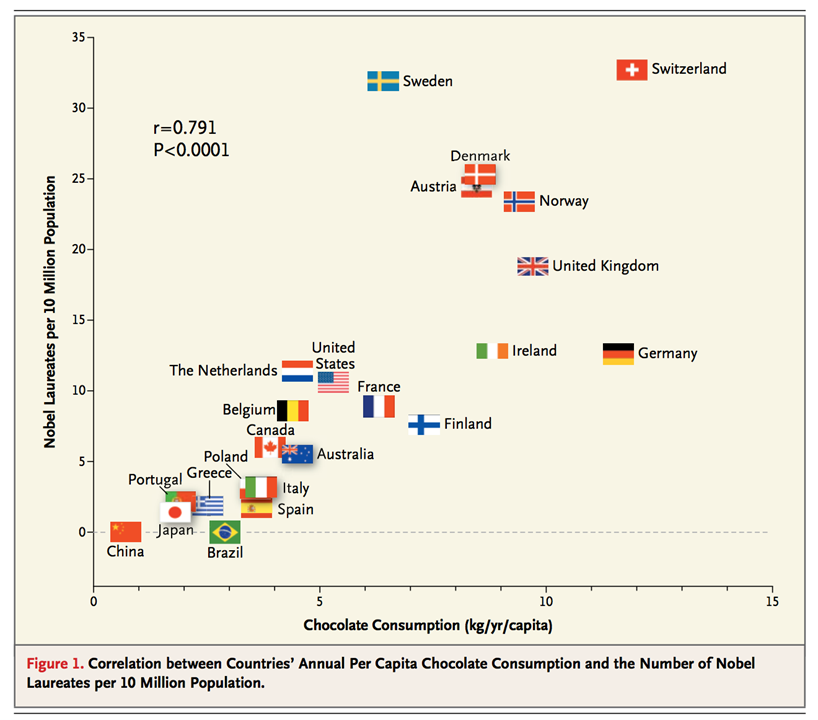
Prediction of h= 71.42, It is a better prediction because the difference is smaller.

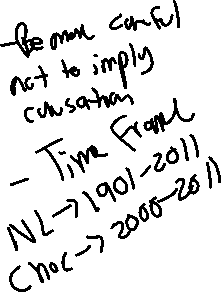
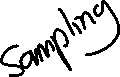
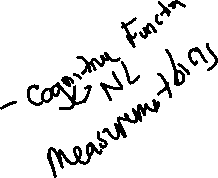
Error = 71.79-41.42 = 0.36

*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).*

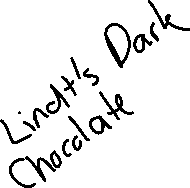
**Question 2**

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.



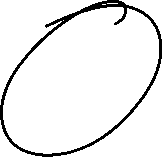


Obtain a PDF copy of Messerli’s publication from the NUS Libraries website. One way is to search by the title, then click on the DOI link. Read the article briefly, and answer the following questions.

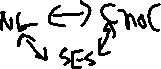


Group 3

1. 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?



The author is suggesting that there might be other confounders in this study that can and will affect the correlation between chocolate consumption and the number of Nobel laureates over the years.



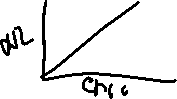
SES is associated to NL = give some reasons

SES is associated to choc = give some reasons

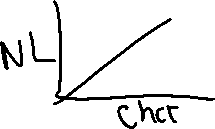
He is suggesting that a difference in the socioeconomic statuses and geographic and climactic factors can also differ from country to country which in turn will affect the correlation, however they cannot fully explain the correlation observed either. There are many other factors hindering the conclusion that “chocolate intake provides the abundant fertile ground needed for sprouting of Nobel laureates”

By slicing the factors such as by socioeconomic status and geographic and climatic factors and analyze the trends based on new data groups to overcome confounders.

Low SES (N=12)



High SES (N=11)



Group 2

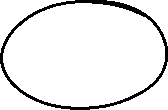
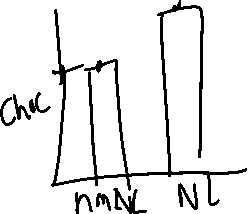
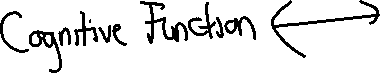
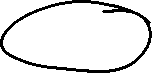
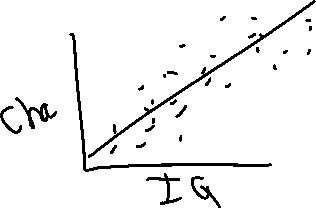
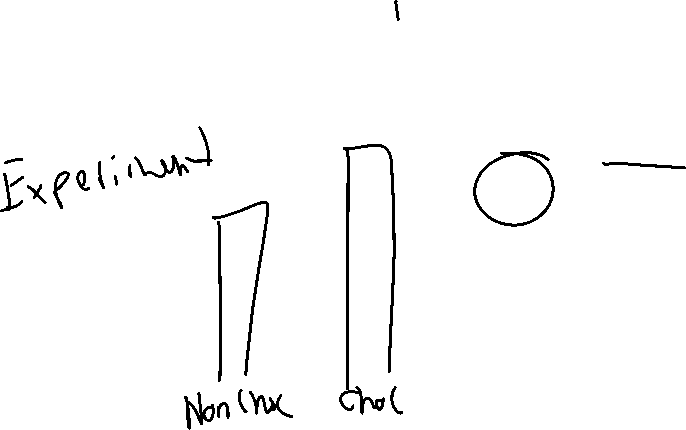
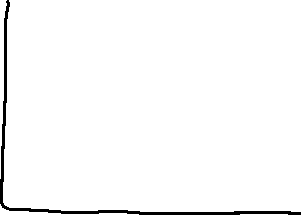
1. What issue was brought up in the first sentence in the Study Limitations section? How would you deal with it, if you were to replicate the study on Nobel prizes and chocolate consumption?

The present data are based on country averages, and the specific chocolate intake of individual Nobel laureates of the past and present remains unknown. The cumulative dose of chocolate that is needed to sufficiently increase the odds of be- ing asked to travel to Stockholm is uncertain. This research is evolving, since both the number of Nobel laureates and chocolate consumption are time-dependent variables and change from year to year.

Issue brought up is that the data is based on country averages, which indicates that this study is looking at Ecological Correlation, in which linear association strength tends to be “overstated” based on aggregated data. This could possibly give rise to the Ecological fallacy, whereby inferences deduced on correlation about individuals based on aggregated data may not hold if they are deduced from individual data instead.

If I were to replicate the study, I would obtain individual data on specific chocolate consumption of individuals (including Nobel laureates), so as to eliminate the Ecological fallacy and be able to observe the strength of linear association between individual chocolate consumption and Nobel prizes.

Alternatively, we can not draw a conclusion based on the individual, we can conclude that the average chocolate consumption is positively associated with the average Nobel Prizes.



**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.

A screenshot of a computer

Description automatically generated

Step 2: EXCEL would take entries in columns A and B as the values of and 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.

A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated

A screenshot of a cell phone

Description automatically generated

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.

A close up of a map

Description automatically generated

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.

