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.

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

Quantity	Value	EXCEL command
Average of father's heights	67.69 in	AVERAGE(A2:A1079)
SD of father's heights	2.74 in	STDEV.P(A2:A1079)
Average of son's heights	68.68 in	AVERAGE(B2:B1079)
SD of son's heights	2.81 in	STDEV.P(B2:B1079)
Correlation	0.50	CORREL(A2:A1079,B2:B1079)

(b) Repeat (a) (iii), (iv) and (v), 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. Do the same for the data corresponding to fathers of height 69 inches: 68.5—69.4 inches.

Fathers of height 68 inches are in rows 498 to 650 inclusive.

Quantity	Value	EXCEL command
Average of son's heights	69.06 in	AVERAGE(B498:B650)
SD of son's heights	2.75 in	STDEV.P(B498:B650)
Correlation	0.01	CORREL(A498:A650,B498:B650)

Fathers of height 69 inches are in rows 651 to 793 inclusive.

Quantity	Value	EXCEL command
Average of son's heights	69.52 in	AVERAGE(B651:B793)
SD of son's heights	2.23 in	STDEV.P(B651:B793)
Correlation	0.02	CORREL(A651:A793,B651:B793)

(c) How are the numbers in (b) different from those in (a)? Can you explain why?

The average son's height of 68-in fathers is more than the overall average son's height. That of 69-in fathers is even higher. This reflects the positive correlation.

The two SD's of son's height are smaller than the overall SD. This is because the spread in the vertical direction is less than the overall spread.

The two correlations are very small compared to the overall correlation. This is attenuation due to range restriction.

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.

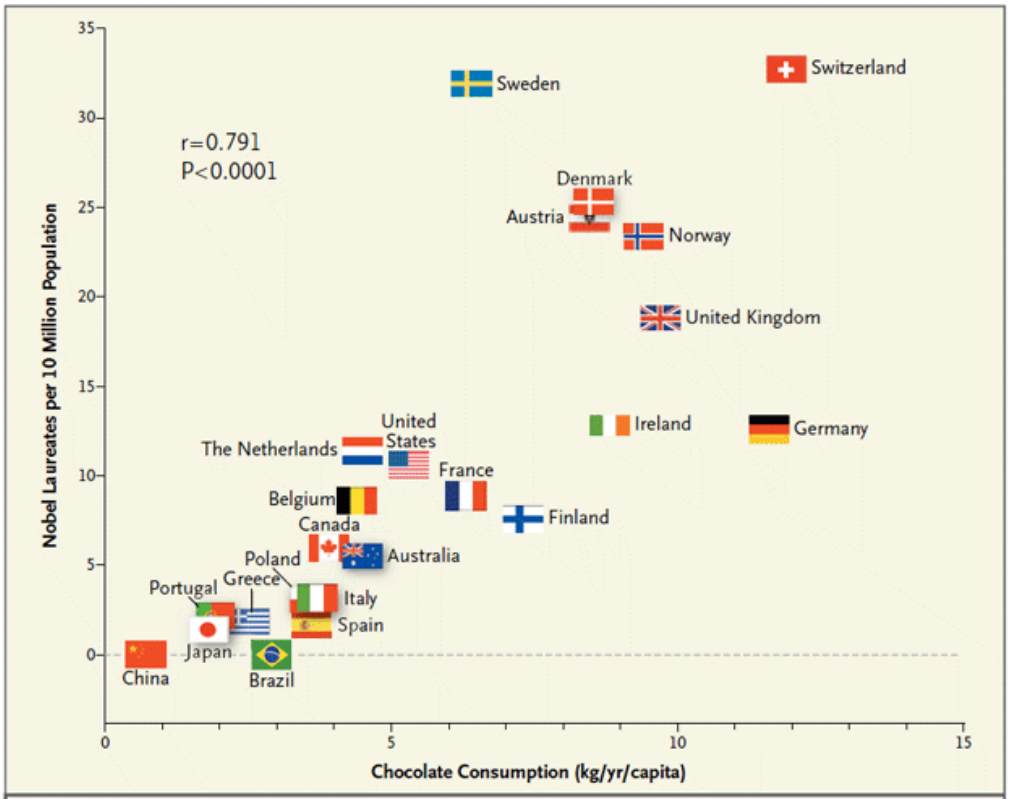

Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

NEW ENGLAND JOURNAL OF MEDICINE

(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? This concerns which step(s) of the QR framework (Frame, Specify, Collect, Analyse and Communicate)?

These are potential confounders. Socioeconomic status of a country is likely associated with chocolate consumption (chocolate is expensive), and also associated with Nobel prizes (education is expensive too). Messerli claims they "fall short of fully explaining the close correlation observed", without further details. We can collect data on the socioeconomic status of countries, then control for it by studying the association separately in say more well-off countries and less well-off countries. This is the slicing method used for controlling confounders. Analyse, Collect.

(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? This concerns which step(s) of the QR framework?

The data involve averages: number of Nobel laureates per 10 million people, and chocolate consumption rate, so the correlation is ecological. The association at the individual level is likely to be less strong. As Messerli admitted, we don't know how much chocolate was consumed by individual Nobel laureates. We should try to find out how much chocolate was eaten by some Nobel laureates, and compare with some non-Nobel laureates. Even then, we have to watch out for confounders. Analyse, Collect.

Two other problematic issues relate to the time frame and choice of countries. While all Nobel Prizes from 1901 to 2011 were included, the chocolate consumption data only came from after 2000. If the consumption habits differ substantially between these recent years and the previous 100 years, as one suspects it would, it weakens the study's conclusion quite substantially. There are many countries that have no Nobel laureates, but nevertheless consume copious amount of chocolate, which are not included in this study of only 22 countries. We would expect the scatter diagram to look quite different if all the almost 200 countries were included.