

SDGB/DGGB 7844 HW 4: Chocolate & Nobel Prizes

Instructor: Prof. Nagaraja

Due: 11/12 in class

Submit your homework as follows: (a) hardcopy of assignment only in class; (b) e-mail assignment and code to fordhamRcomputing@gmail.com by the start of class. In your e-mail, use the subject heading: “HW 4: [Your name]” and include HW 4 and your name the file names you attach. Comment your code for full credit. Complete your work individually.

Read the *New England Journal of Medicine* article, “Chocolate Consumption, Cognitive Function, and Nobel Laureates” (Messerli, F.H., Vol. 367(16), 1562-1564; 2012) which is posted on Blackboard. We will be using a reconstruction of Messerli’s data along with GDP per capita¹. The variables in the data set you will use are (file: “nobel_chocolate.txt” on Blackboard) are “country”, “nobel_rate”, and “chocolate”.

The information gathered in the data set you will be using is from several different sources. The number of Nobel prize winners is from Wikipedia and includes winners through November 2012, population information (used to compute the “nobel_rate” variable) is from the World Bank, and chocolate market size is from the Euromonitor International’s Passport Database.

Goal: In this assignment, you will be replicating Messerli’s analysis.

¹Source: McClintock, S., Stangl, D., and Çetinkaya-Rundel, M. (2014). The real secret to genius? Reading between the lines. *CHANCE*. **27**(1).

1. According to Messerli, what is the variable “number of Nobel laureates per capita” supposed to measure? Do you think it is a reasonable measure? Justify your answer.
2. Are countries without Nobel prize recipients included in Messerli’s study? If not, what types of bias(es) would that introduce?
3. Are the number of Nobel laureates per capita and chocolate consumption per capita measured on the same temporal scale? If not, how could this affect the analysis?
4. Construct a scatterplot of Nobel laureates vs. chocolate consumption. Label Sweden on your plot (on the computer, not by hand). Compute the correlation between these two variables and add it to the scatterplot. How would you describe this relationship? Is correlation an appropriate measure? Why or why not?
5. Why is your correlation different from Messerli’s?
6. Why does Messerli consider Sweden an outlier? How does he explain it?
7. Regress Nobel laureates against chocolate consumption (include Sweden):
 - (a) What is the regression equation? (Include units of measurement.)
 - (b) Interpret the slope.
 - (c) Is the slope significant (conduct a hypothesis test and include your regression output in your answer)? Test at the $\alpha = 0.05$ level.
 - (d) Add the regression line to your scatterplot.
 - (e) Conduct a residual analysis to check the regression assumptions. Make all plots within one figure.
8. Using your model, what is the number of Nobel laureates expected to be for Sweden? What is the residual? (Remember to include units of measurement.)
9. Does increasing chocolate consumption cause an increase in the number of Nobel Laureates? Justify your answer.