Yu Zhang

ACG5906
Assignment 1

Q1

Table 1: Cross-correlation table

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
nobel	23	11.09	10.22	0.0500	31.85
chocolate	23	5.732	3.065	0.800	11.53

Based on the summary statistics of the dataset, the dataset contains information on Nobel laureates and chocolate consumption from 23 countries. The data shows a wide range of Nobel Laureates per capita, with some countries producing very few Nobel Laureates (as low as 0.05) and others producing many (as high as 31.85). Similarly, there is a noticeable variation in chocolate consumption. The standard deviations for both variables support the observation that there is a considerable variation in both variables.

Q2

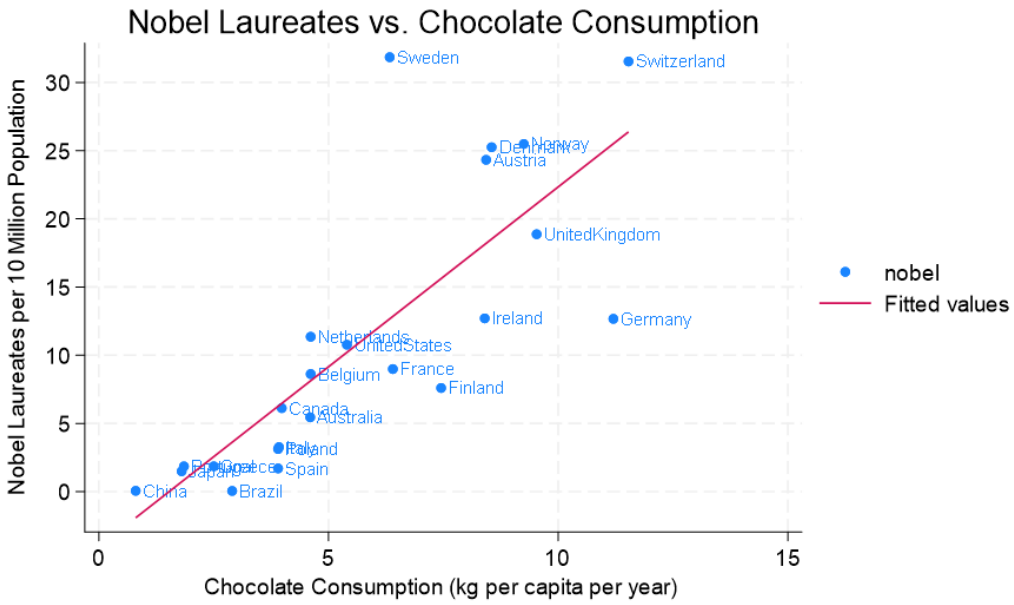

Figure 1: Nobel Laureates vs. Chocolate Consumption

The scatter plot shows a positive linear relationship between chocolate consumption and the number of Nobel laureates. Sweden appears to be an outlier, having a much higher number of Nobel laureates than its chocolate consumption might suggest compared to other countries.

Q3

(a)

Table 2: Correlation Table

Variables	nobel	chocolate
nobel	1.000	
chocolate	0.791	1.000
	(0.000)	

The correlation table shows a strong and statistically significant positive relationship between chocolate consumption and Nobel Laureates.

(b)

Table 3: Correlation Excluding Sweden

Variables	nobel	chocolate
nobel	1.000	
chocolate	0.862	1.000
	(0.000)	

After the exclusion of Sweden, there appears to be an even stronger positive correlation between the two variables.

(c)

Table 4: Linear Regression

	(1)
VARIABLES	Model 1
chocolate	2.637***
	(0.445)
Constant	-4.029
	(2.878)
Observations	23
R-squared	0.626

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The regression result shows a strong and statistically significant positive association between chocolate consumption and Nobel laureates per capita. The coefficient for chocolate is 2.637 (β), which suggests that for every additional kilogram of chocolate consumed, the number of Nobel laureates per 10 million people is predicted to increase by 2.637. In other words, the model predicts it will take about 0.4 ($1/\beta$) kg of chocolate per capita per year to increase the number of Nobel Laureates in a given country by 1.

(d)

Since the model suggests that it will take about 0.4 ($1/\beta$) kg of chocolate per capita per year to increase the number of Nobel Laureates in a given country by 1, and the population of the United States is about 330 million, the model predicts that it will take around 125 ($1/\beta \times 330$) million kg additional chocolate composition per year for the United States to increase their Nobel Laureates per capita by 1.

Q4

(a)

The author suggests that dietary flavonoids found in chocolate have been shown to improve cognitive function. Then, the author reports a significant and strong correlation between national chocolate consumption rates and the number of Nobel laureates per capita. Therefore, the author speculates that the observed correlation might indicate a potential causal relationship where chocolate consumption enhances cognitive abilities, which can be proxied by Nobel laureates.

(b)

I disagree with the author because chocolate consumption and Nobel laureates could be influenced by numerous confounding factors such as socioeconomic status, education systems, investment in research, cultural factors, etc.

Q5

(a)

The author considers Sweden an outlier, which may influence the results of the statistical analyses. The author also argues that various factors may bias the observation of Sweden (e.g., the Nobel committees are based in Sweden).

(b)

Yes, we observed a stronger correlation after the exclusion of Sweden.

(c)

Table 5: Linear Regression Excluding Sweden

	(1)
VARIABLES	nobel
chocolate	2.579***
	(0.339)
Constant	-4.570*
	(2.194)
Observations	22
R-squared	0.743

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The exclusion of Sweden gives a smaller coefficient on chocolate because Sweden has a very high number of Nobel laureates compared with chocolate consumption. Therefore, including Sweden will bias the coefficient upward.
No. Correlation measures the strength and direction of a relationship between two variables, and regression estimates the effect size.

Q6

The author argues that while winning a Nobel Prize might involve celebrations, it is implausible that such events would cause a sustained increase in overall chocolate consumption. Instead, the author leans toward the hypothesis that chocolate consumption enhances cognitive function, which could contribute to higher national-level achievements, such as winning Nobel Prizes.

Q7

(a)

The author suggests that finding a single common factor, or "common denominator," could explain both high levels of chocolate consumption and the high number of Nobel laureates across different countries over many years.

(b)

Because socioeconomic status could be an important explanatory variable correlated with chocolate and Nobel laureates, omitting such variables leads to biased or inconsistent effects estimates.

Q8

(a)

Table 6: Linear Regression With Additional Variables

	(1)
VARIABLES	nobel
chocolate	1.998***
	(0.602)
gdp	0.000317**
	(0.000125)
hdi	-69.07
	(49.36)
energy	-0.00385
	(0.00645)
protein	0.217
	(0.212)
life	-0.101
	(0.738)
Constant	44.23
	(51.29)
Observations	23
R-squared	0.747

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The regression result supports the author's claim. The chocolate coefficient remains large and highly significant even after controlling for other socioeconomic and development variables. While GDP is statistically significant, its effect is minimal compared to chocolate consumption. Other controls do not significantly explain the variation in Nobel laureates.

(b)

Table 7: Cross-correlation table

Variables	nobel	chocolate	gdp	hdi	energy	protein	life
nobel	1.000						
chocolate	0.791	1.000					
	(0.000)						
gdp	0.726	0.656	1.000				
	(0.000)	(0.001)					
hdi	0.489	0.533	0.808	1.000			
	(0.018)	(0.009)	(0.000)				
energy	0.099	0.255	0.028	0.157	1.000		
	(0.655)	(0.240)	(0.900)	(0.474)			
protein	0.088	0.080	0.166	0.482	0.466	1.000	
	(0.689)	(0.716)	(0.449)	(0.020)	(0.025)		
life	0.275	0.285	0.544	0.705	0.006	0.365	1.000
	(0.204)	(0.187)	(0.007)	(0.000)	(0.978)	(0.087)	

Since energy, protein, and life have a weak correlation with both Nobel laureates and chocolate, I choose to drop them.

Table 8: Cross-correlation table

	(1)
VARIABLES	nobel
chocolate	1.846***
	(0.526)
gdp	0.000287**
	(0.000109)
hdi	-47.46
	(33.34)
Constant	29.05
	(25.72)
Observations	23
R-squared	0.730

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The results from the regression after excluding those variables do not significantly change our observations.