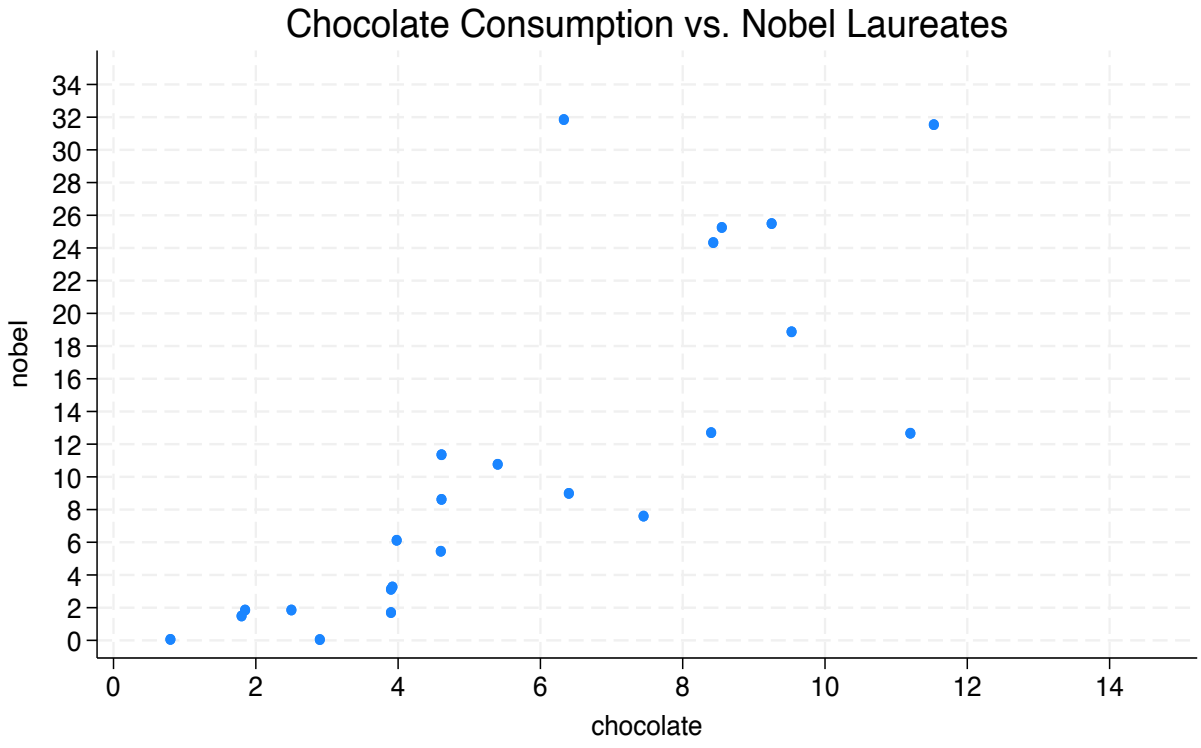
Michael Burrage II
Econometrics Assignment 1

1. Summary Statistics for Nobel and Chocolate Variables

- **Nobel Laureates (per 10 million population):**
  - Observations: 23
  - Mean: 11.08878
  - Standard Deviation: 10.21818
  - Minimum: 0.05
  - Maximum: 31.855
- **Chocolate Consumption (kg per capita/year):**
  - Observations: 23
  - Mean: 5.73217
  - Standard Deviation: 3.06525
  - Minimum: 0.8
  - Maximum: 11.53

2. Scatter Plot: Chocolate Consumption vs Nobel Laureates



3. Correlation and Linear Regression Analysis

(a) Correlation Between Chocolate Consumption and Nobel Laureates:

- The correlation coefficient between chocolate consumption and Nobel laureates is **0.7912**, which is consistent with the paper's reported **r = 0.791** ($p<0.0001$). This suggests a strong and statistically significant positive correlation.

Michael Burrage II
Econometrics Assignment 1

(b) Excluding Sweden:

- The correlation increased to **0.8623** when Sweden was excluded, closely matching the paper's recalculated value of **r = 0.862**. This exclusion strengthens the correlation by removing Sweden as a potential outlier.

(c) Linear Regression Analysis:

- The regression results indicate that chocolate consumption has a positive and significant effect on the number of Nobel laureates. The model suggests that a 1 kg increase in chocolate consumption is associated with an increase of **2.5792** Nobel laureates per 10 million population. This coefficient aligns well with the paper's estimation, which also highlights a similar dose-response relationship.

(d) Required Chocolate for 1 Additional Nobel Laureate:

- To calculate how much additional chocolate is needed for the U.S. to produce one more Nobel laureate, the estimated coefficient suggests:

  Additional chocolate required = 1 / 2.5792 ≈ **0.3878 kg per capita/year**

  This is consistent with the paper's estimate that **0.4 kg** per capita/year would be necessary for a 1 laureate increase.

4. Discussion of the Dose-Dependent Relationship

(a) Logic Behind Chocolate and Nobel Laureates Relationship:

- The paper hypothesizes that increased chocolate intake improves cognitive function, and thus enhances the ability of a population to produce Nobel laureates. This is based on the documented cognitive benefits of flavanols found in chocolate.

(b) Do You Agree with the Author?

- While the correlation is strong, it's important to remember that correlation does not imply causation. The findings suggest an association between chocolate consumption and Nobel laureates, but they require further investigation. Chocolate consumption, as a non-essential food item, may be correlated with other socioeconomic status, education, and other lifestyle factors that contribute to cognitive development and access to resources. Simply increasing a country's chocolate consumption without improving education, access to knowledge, and resources would logically not result in producing more Nobel laureates. Thus, the relationship between chocolate and laureates is likely influenced by broader socioeconomic conditions rather than the chocolate itself.

Michael Burrage II
Econometrics Assignment 1

5. Outlier Analysis: Sweden

(a) Logic Behind Sweden's Exclusion:

- Sweden is considered an outlier in the relationship because it has a disproportionately high number of Nobel laureates relative to its chocolate consumption (6.4 kg per capita/year). The author suggests that this may be due to the Nobel Committee's patriotic bias or a higher sensitivity to chocolate.

(b) Does Sweden's Exclusion Strengthen the Correlation?

- Yes, excluding Sweden increases the correlation from **0.791** to **0.862**, indicating that Sweden may distort the overall trend between chocolate consumption and Nobel laureates.

(c) Impact on the Coefficient:

- Yes, excluding Sweden increases the regression coefficient for chocolate consumption, indicating a stronger relationship between chocolate consumption and Nobel laureates. However, the interpretation of the regression coefficient is not identical to the correlation coefficient. The regression coefficient quantifies the increase in Nobel laureates per 10 million people for each additional kilogram of chocolate consumed per capita. In contrast, the correlation coefficient simply measures the strength and direction of the linear relationship between the two variables. Therefore, while both metrics indicate a positive relationship, they capture different aspects of the relationship.

6. Reverse Causality

- The author considers the possibility of reverse causality, hypothesizing that people with higher cognitive function might be more likely to consume chocolate. However, the idea that winning a Nobel Prize could lead to increased chocolate consumption was dismissed as unlikely.

7. Omitted Variables

(a) What is the logic behind this statement according to the author?

- The author suggests that while socioeconomic status, geographic, and climatic factors may influence both chocolate consumption and the number of Nobel laureates, they do not fully account for the observed correlation. The logic is that although these factors play a role, they cannot completely explain the close relationship between chocolate consumption and cognitive outcomes (as reflected by Nobel laureates), hinting at the possibility of other unexplored variables or the cognitive benefits of chocolate itself.

(b) Why is omitting socioeconomic status a potential cause of omitted variable bias?

- Omitting socioeconomic status from the analysis could lead to omitted variable bias because countries with higher GDP, better education systems, and greater resources may both consume more chocolate and produce more Nobel laureates. By excluding socioeconomic variables, the analysis may incorrectly

attribute the relationship between chocolate and Nobel laureates to chocolate consumption alone, when in reality, socioeconomic factors might be driving both variables.

8. Extended Dataset Analysis

(a) Regression Using Additional Variables:

- A regression analysis including variables such as GDP per capita and life expectancy shows that while chocolate consumption remains significant, socioeconomic factors like GDP per capita are also significant predictors. This supports the idea that socioeconomic variables contribute to the relationship.

(b) Variable Choice Explanation:

- Including GDP and life expectancy is necessary because these variables capture both the economic capacity and health, which could influence both chocolate consumption and the ability to produce Nobel laureates. Energy and protein consumption did not appear significant in this context.