**HW4:Solutions**

**SDGB 7844: Statistical Methods and Computation I**

**Minxia Ji**

**Question 1**

“number of Nobel laureates per capita”supposed to measure the overall cognitive function of a given country.

I think it is **not** a reasonable measure.The number of Noble prize can only represents outstanding cognitive function, not the overall cognitive function of a given country.Instead, other measures such as the education level or IQ test results might be better than this to represent cognitive function.

**Question 2**

Yes.They are China and Brazil.

If the countries not included, it would introduce sample selection bias.

**Question 3**

They are **not** in the same scale.

The temporal scale for Nobel laureates is measured by all Nobel Prizes that were awarded through October 10, 2011.However, the temporal scale for Chocolate Consumption is 2011 for Switzerland, 2010 for 15 countries, 2004 for 5 countries, and 2002 for China.

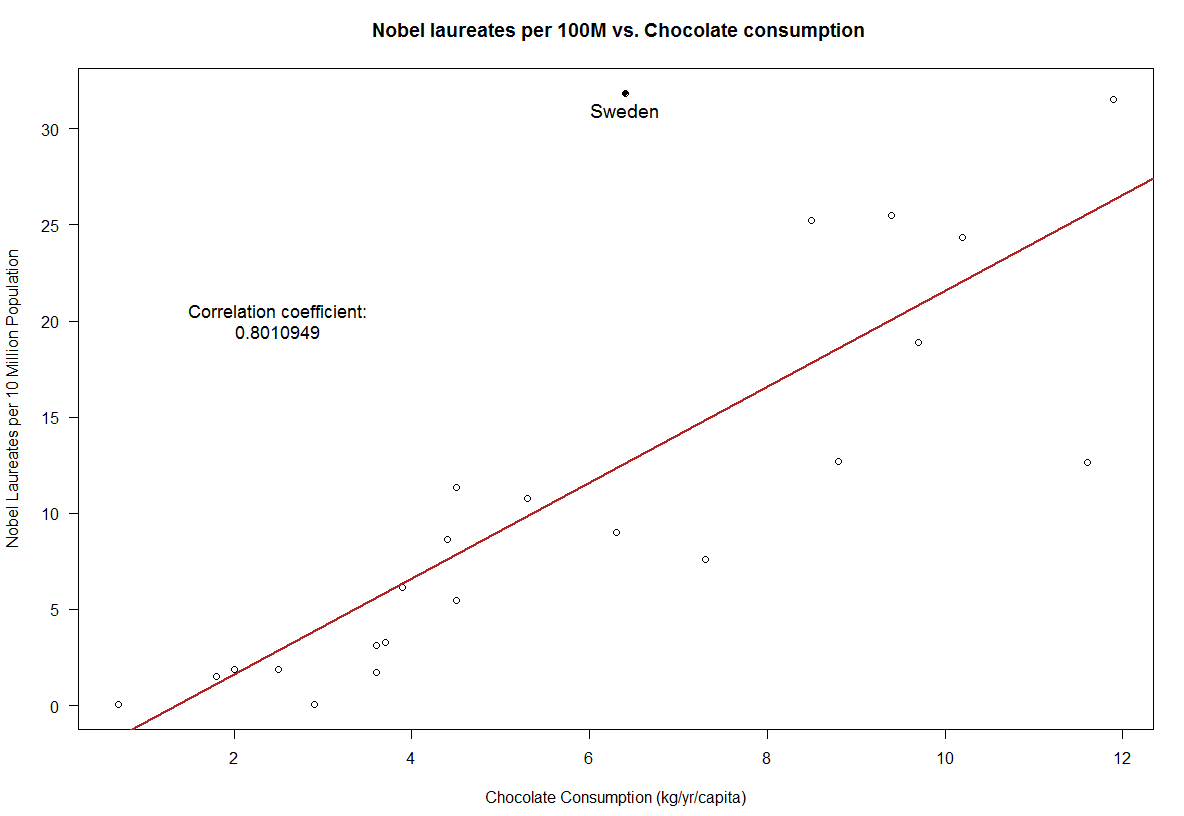
The different temporal scale for variables would **decrease the accuracy** of the analysis.The temporal scale for Nobel laureates is measured by a period of time, the temporal scale for Chocolate Consumption should also be measured by the same period of time.And the consumption of Switzerland, China and other 5 countries might change in 2011.

**Question 4**

The correlation between these two variables is 0.8010949.

The correlation coefficient shows there is a strong correlation between two variables.

It’s an **appropriate** measure, though the correlation doesn’t mean causation, but still the 0.8 shows strong relation between this two variables.



**Question 5**

Messerli’s correlation value:0.791

My correlation value is 0.801

In Messerli’s analysis, the Noble laureates is rounded to the nearest number.However, our data use the exact numbers, which are slightly differ from Messerli’s . For example, the Noble laureates for Sweden in Messerli’s analysis is 32 but in the dataset it is 31.855.

**Question 6**

Because Sweden’s per capita chocolate consumption is 6.4 kg per year, he would predict that Sweden should have produced a total of about 14 Nobel laureates, yet we observe 32, which exceeds the expected number by a factor of more than 2.

Messerli explains that it might because 1.Nobel Committee in has some inherent patriotic bias when assessing the Swedish candidates or 2.Swedes are particularly sensitive to chocolate, even minuscule amounts greatly enhance their cognition.

**Quesiton 7**

1. Nobel Laureates(per 10 Million Population) = -3.400 (per 10 Million Population)

+ 2.496(per 10 Million Population)/ (kg/yr/capita)\* Chocolate Consumption

1. Every additional kilogram increase per year per capita in chocolate consumption is associated with a 2.496 increase per 100 Million population in Nobel laureates.
2. We can conduct hypothesis tests for this regression model.
3. Linearity

No pattern is observed in the top plot.

Assumption satisfied.

2. Constant variance

Fanning residuals is observed in the top plot and the number of residuals above or below the line 0 is almost the same.

Assumption is not completely satisfied.

3. Normality

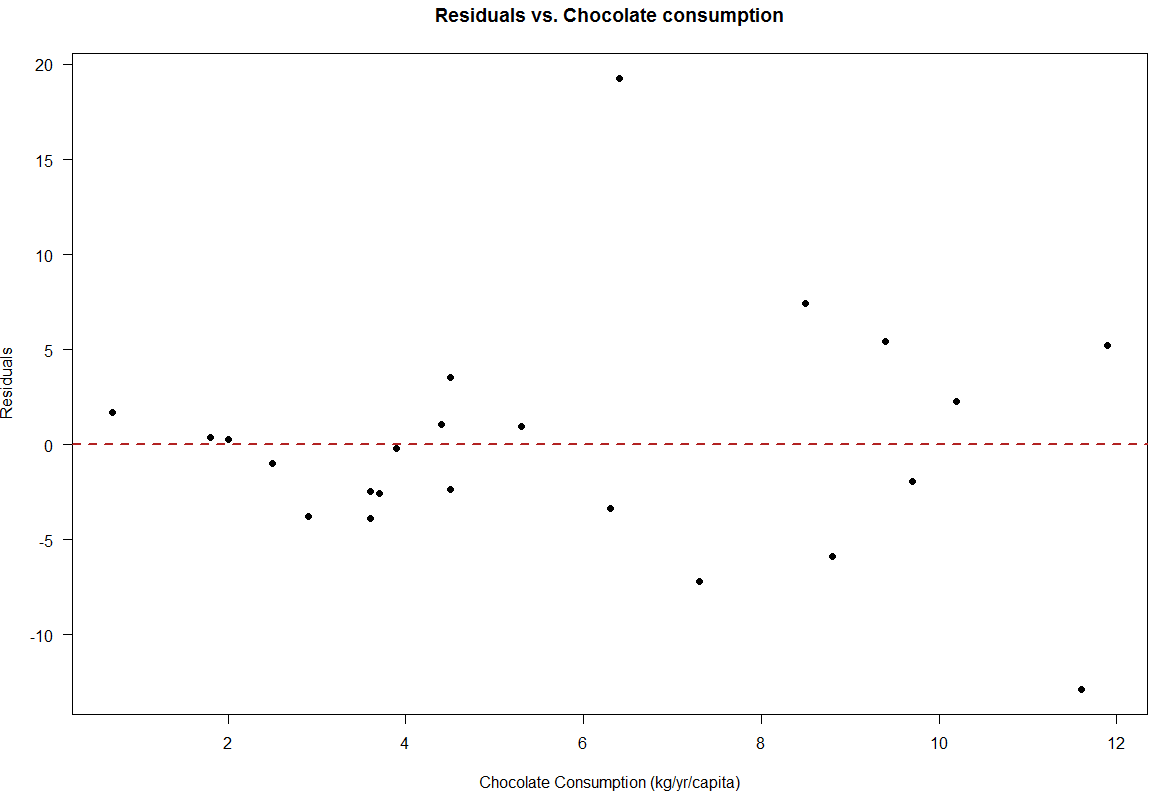
points almost form a straight line in the bottom plot.The first and the last points are a little bit like the sign of heavy tailed.

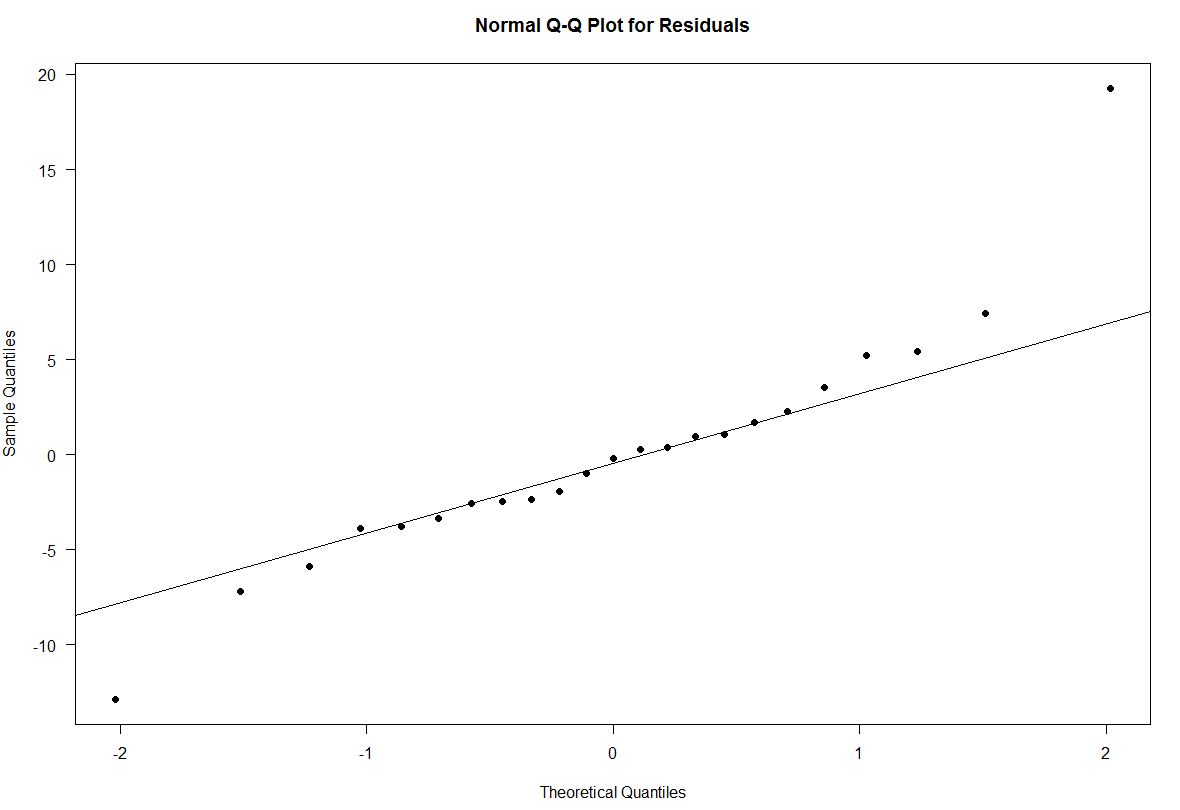
Assumption is not completely satisfied.

4. Independence

Not a time series data.Countries are probably independent of each other.

Assumption satisfied.





1. The slope is significant.

·Hypothesis:

H0 : β1 ＝ 0

H1 : β1 ≠ 0

where the β1 is the slope for chocolate consumption

·α = 0.05

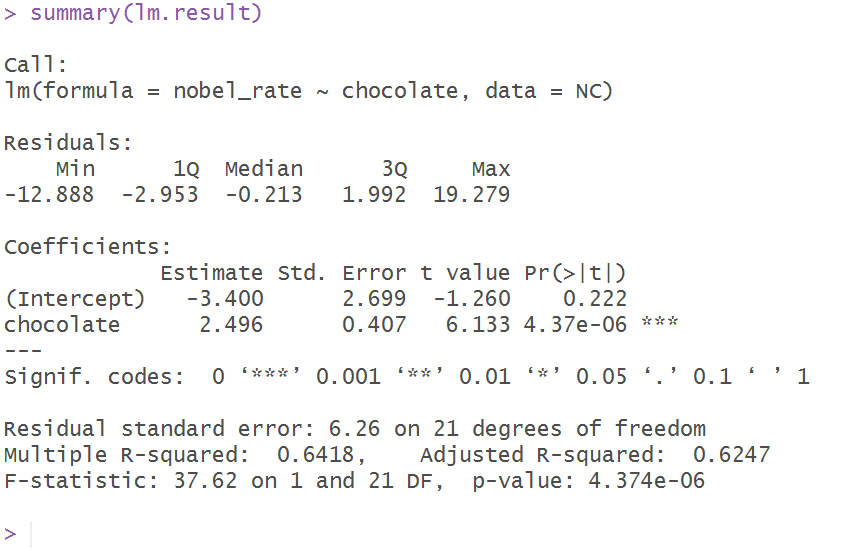
·Test statistic:

t = 6.133

which has a t distribution with 21 degrees of freedom

· P-value is 4.37e-06, much smaller than α = 0.05

· Conclusion: reject null hypothesis and conclude that the slope of the model is significantly different from 0



1. See the graph in solution in question 4.

**Question 8**

The number of Nobel laureates expected to be for Sweden is 12.57568 per 100M population.

The residual is 19.27932 per 100M population.

**Question 9**

Increasing chocolate consumption does cause an increase in the number of Nobel Laureates.

The hypothesis test reject the null hypothesis that there is no linear relation between the number of Nobel Laureates and chocolate consumption, therefore a linear relation exists between two variables.

In addition, R-square=0.6418, which means 64.18% of variation in the number of Nobel Laureates per 10 Million Population is explained by the variation in the chocolate consumption.