Sumi Choudhury HW #4

SDGB 7844 November 16, 2016

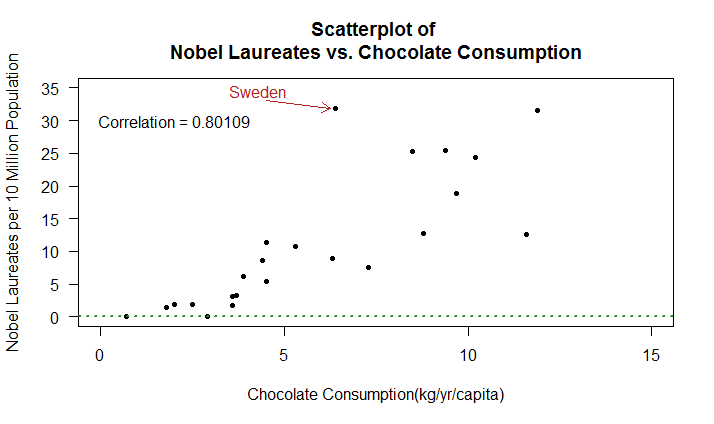
Solutions HW 4

1) According to Messerli, the number of Nobel laureates per capita in this model measures the number of Nobel laureates per 10 million persons in a total of 23 countries. It is the response variable. The per capita measure does not take into account that there may be more socioeconomically enriched areas within a given country, and therefore certain areas of these countries may be heavily influenced by other variables that produce high cognitive individuals. Also this variable is measured against chocolate consumption per capita for a given country. However, the number of Nobel laureates per country can be produced or influenced by a number of other variables or predictors which were not analyzed in this model. For example, proportion of GDP spend on education can affect the number of Nobel laureates in a country and when running a regression with multiple variables, the partial derivative of the slope for the chocolate consumption may not be as significant as it appears in this current example. Also, there can be underlying factors such as other variables which influence both these X and Y variables positively at the same time. Examples include differences in socioeconomic status from country to country and geographic and climatic factors.

2) Countries without Nobel Prize recipients are not included in this model. This would then leave out persons with high cognitive functioning from other regions of the world, who may or may not have benefited from this proposed correlation between chocolate consumption and improved cognitive function. We are restricting our assumption to the point that greater number of Nobel laurates in a country equates to higher levels of cognitive ability. Yet, there may be other variables that are contributing to the observations of chocolate consumption and Nobel laureates having a high correlation, such as the aforementioned underlying factors. Also, there are countries which due to socioeconomic reasons, or cultural differences, may not have a high chocolate consumption rate, yet may produce intelligent individuals with high cognitive abilities.

3) The number of Nobel laureates per capita and chocolate consumption per capita are not measured on the same temporal scale. The number of Nobel laureates are taken from as far back as the beginning of the 1900s, while the chocolate consumption per capita is taken for only four years with the earliest being 2002. The sample spaces are so vastly different that this cannot provide us with a reasonable relationship with which to compare the two variables. Even the countries that have the lowest turnout of Nobel laureates, may have produced a Nobel laureate during a time that large amounts of chocolate consumption was not so possible as it is today.

4)



There is a linear relationship between the two variables, so correlation is an appropriate measure. However, this does not mean that it is necessarily a useful measure. There is a bias in the data that is selected. In this case we have the predictor variable chocolate consumption, which is on a temporal scale that is much smaller than the variable Nobel laureates. So the correlation does not indicate a useful linear relationship although a linear relationship exists.

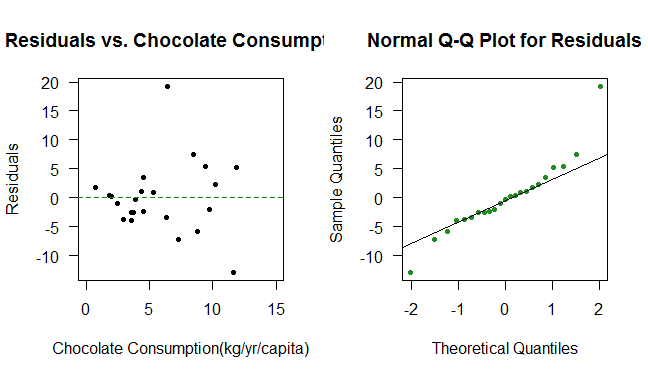
5) Messerli’s correlation (which includes Sweden) is 0.791.The reason that my output correlation of 0.801 is slightly different is because Messerli uses whole numbers for the calculations, while my data set is using the decimal values.

6) Messerli considers Sweden an outlier. This is because Sweden does not represent a typical data point in this data set. Its actual output of Nobel laureates strays far from the estimated Y hat value of 14 Nobel laureates given its per capita chocolate consumption of 6.4 kg per year. Its observed output is 32 Nobel laureates.

7) (a) The regression equation is Ŷ = -3.4 Nobel laureates per 10 million pop. + 2.496X kilograms of chocolate consumption per capita per year.

(b) The slope is 2.496. This means that for every one kilogram increase, on average, of chocolate consumption per capita per year for a given country it is associated with an additional 2.496 Nobel laureates for the country on average.

(c)



Based on the first plot, the residual plot, we can see that the linearity assumption is met based on the steady green line; there is no apparent trend.

Based on the first plot, we can also see that there is a fairly constant variance, meaning that as the X values become larger, the variations in Y, or the errors, remain fairly constant without too much fanning.

Based on the second plot, the Normal Q-Q plot, we can see that if our errors or residuals are normally distributed, the fitted line would generally be a diagonal line; this assumption is met.

Independence – this data is not a time series, so chocolate consumption rates are probably independent of one another and not necessarily contributing to the residuals getting larger over time; this assumption is met.

(d) The slope is significant in this model. As shown in the below regression output, we can refer to the t-value for Chocolate Consumption or the P-value associated with its slope. Here the hypothesis we are testing is:

H**0**: βchocolate consumption = 0 Ha: βchocolate consumption ≠ 0

The t-value of 6.133 is quite large, which indicates that it will be far right of the t-critical value. Also, we can see by the P-value of 4.37e-06 that this is quite small and 2 times the P-value is less than α = 0.05, so we reject the null hypothesis. We can conclude that the slope is significantly different from zero.

lm(formula = x$Nobel\_Laureates ~ x$Chocolate\_Consumption)

Residuals:

Min 1Q Median 3Q Max

-12.888 -2.953 -0.213 1.992 19.279

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -3.400 2.699 -1.260 0.222

x$Chocolate\_Consumption 2.496 0.407 6.133 4.37e-06 \*\*\*

---

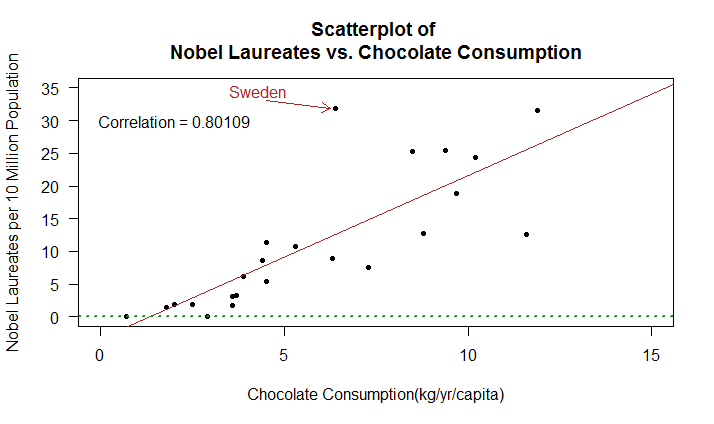
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.26 on 21 degrees of freedom

Multiple R-squared: 0.6418, Adjusted R-squared: 0.6247

F-statistic: 37.62 on 1 and 21 DF, p-value: 4.374e-06

(e)



8) The number of Nobel laureates for Sweden is expected to be 12.57568 Nobel laureates per 10 million population and the residual is 19.27932 Nobel laureates per 10 million population.

9) Chocolate consumption does not cause an increase in the number of Nobel laureates. As part of my answer to question 1, where I explained that the response variable may not be best indicated by the highly correlated predictor which is chocolate consumption, we must be careful to note that correlation does not equate to causation. First, the number of Nobel laureates per country can be produced or influenced by a number of other variables or predictors which were not analyzed in this model. For example, proportion of GDP spend on education can affect the number of Nobel laureates in a country and when running a regression with multiple variables, the partial derivative of the slope for the chocolate consumption may not be as significant as it appears in this current example. Another consideration is that there can be underlying factors such as other variables which influence both these X and Y variables positively at the same time. Examples include differences in socioeconomic status from country to country and geographic and climatic factors.

In some cases also, the Y variable may be influencing the X variable, although it does not seem to be the case in this example.

Yet another factor to think about is that other food items containing flavonoids (and other nutrients) can just as easily be hypothesized and tested in the regression model. Perhaps these countries with high chocolate consumption also consume more green tea, or wine. There are still a multitude of variables to consider in order to make the model for meaningful.