MATH 422 – SPRING 2020 – HOMEWORK 5

1. Show that if X and Y are independent random variables, then $\mu_{Y|x} = E(Y)$.

2. Exercise 14.2

3. Exercise 14.3

4. Exercise 14.8 (note that the regression is linear!)

5. Investigate the relationship between CPI and GDP. CPI is the *consumer price index* which is calculated by tracking the cost of a collection of goods and services. GDP is *gross domestic product* which includes all economic activity for the year.

   (a) Find the equation of the regression line.

   (b) Plot the data and the regression line.

   (c) Find a 95% confidence interval for the slope.

   (d) Find a 95% confidence interval for the correlation coefficient.

   (e) Assume that GDP for 2020 was projected to be 19,452 million dollars (2.1% growth). Due to the coronavirus pandemic, the GDP is now projected to be 18,665 million dollars (2.9% contraction). Use your regression line to predict the CPI for both of these values. Construct a 95% confidence interval for CPI in both of these scenarios. *Note that projections for 2020 GDP vary depending on the source and change often as new information becomes available. However, overall growth will likely be significantly less than the original projection of 2.1% growth.*

6. Investigate state level COVID-19 data.

   (a) Let X be the state GDP and Y be the number of confirmed COVID-19 cases. Find the equation of the regression line for $\log(X)$ and $\log(Y)$. Plot the log-log data along with the regression line.

   (b) Find a 95% confidence interval for the correlation between $\log(X)$ and $\log(Y)$.

   (c) Think about these calculations and write a few sentences on you thoughts about the relationship between GDP and number of confirmed COVID-19 cases. Here are a few things you might consider, but feel free to discuss other things. You are not obliged to discuss these particular points.

      - Is the relationship between GDP and number of COVID-19 cases strong?
      - What is the relationship between other variables in the dataset?
      - Does GDP *cause* a high number of COVID-19 cases? What underlying causal factors could there be?