1/25/21

Jonathan Meyer

$Y = \beta_0 + \beta_1 X + \epsilon$

12.2 #9 from Devore

Project 1: Exercise 1

$$\vec{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}_{2\times 1} \qquad \vec{\beta} = \begin{pmatrix} -45.6 \\ 1.7 \end{pmatrix}$$

$$\vec{Y} = \begin{pmatrix} 150 \\ 140 \\ 180 \\ 210 \\ 190 \\ 320 \\ 280 \\ 400 \\ 430 \\ 440 \\ 390 \\ 600 \\ 610 \\ 670 \end{pmatrix}_{14\times 1} \qquad \vec{X} = \begin{pmatrix} 1 & 100 \\ 1 & 125 \\ 1 & 125 \\ 1 & 150 \\ 1 & 150 \\ 1 & 200 \\ 1 & 200 \\ 1 & 250 \\ 1 & 250 \\ 1 & 300 \\ 1 & 300 \\ 1 & 350 \\ 1 & 400 \\ 1 & 400 \end{pmatrix}_{14\times 2} \qquad \vec{\epsilon} = \begin{pmatrix} 24.4 \\ -28.4 \\ 11.6 \\ -1.2 \\ -21.2 \\ 23.3 \\ -16.7 \\ 17.69 \\ 47.7 \\ -27.8 \\ -77.9 \\ 46.6 \\ -29 \\ 30.1 \end{pmatrix}_{14\times 1}$$

$$\vec{Y} = X\vec{\beta} + \vec{\epsilon}$$

Red

Blue

Black Black

Project 4: Exercise 2

Compare & Contrast 2 of the 4 modeling strategies indicated in the notes

All of the 4 cited modeling strategies have their place as a valid and appropriate manner to go about using statistics to either solve, predict, or understand some real world problem. I found model 2 and model 3 to be the most interesting because of the different ways they go about identifying a problem.

Compare:

- The later steps of both processes deal with assessing how well the crafted mathematical model works for understanding the real world problem
- Model 2's refinement works by fluidly moving between the short list of steps to ensure the mathematical model is a good & accurate representation of the conceptual model. Model 3 is similar in that it also wants steps repeated multiple times for the purpose of

refining the mathematical model to ensure it is an accurate conceptual reproduction of the real world model

- The obvious comparison: both models aim to construct a mathematical model that accurately represents a real world problem
- Both models aim to eliminate unwanted variables that are irrelevant to the real world problems. In other words both models seem to aim to simplify the data to be more easily readable

Contrast: - In model 2 you aim to tackle a real world problem from the onset. In model 3 you don't make assumptions with regard to the problem. You just collect and assess data and through that process the problem is revieled
- Model 3 has more structure to the overall process, with more explicit steps for what to do. Model 2 allows for more freedom in how the building process happens but doesn't have but subtle checks for determining the validity of analysis, or even the data itself.

- Model 3 advocates for making sure you do each step correctly before progressing to the next step. With model 2 you work backward & forward for several times between steps to refine the mathematical model
- Model 2 seems to prioritize be useful for seeing if a mathematical model can represent a real world problem. Model 3 is useful for gathering accurate data and discovering the most impactful variables/ interactions/ predictors from that data

Synthesize the 2 strategies:

1) RWS to derive conceptual model. gather data relevant to that model
2) check data quality & see how well it works in the CM & refine as needed. Does CM still represent RWS?
3) Determine relevant interactions & predictor variables to construct mathematical model study residuals, interactions, curvature
4) Use MM. Move backwards to CM to RWS & assess model validation

Jonothan Meyer
Project 1: Exercise 3
Linear Regression
1/29/21

Problem Title: The Opioid Crisis - 2019

1) What data collection strategies for the problem would you consider. Would you used a controlled experiment or an observational study approach?

I would use an observational study approach. There is a wealth of data available regarding doctor opiate prescriptions, incarcerations due to opiates, hospitalizations as a result of overdose, ext. I would draw from all that data to try to find interactions between opiate abuse and economic impact. I think an important first step that might be easily overlooked, is looking at the interaction between the flood of opium to China and the most significant impact it had on their economy between 1820 and 1840, prior to the Opium Wars. Obviously based on the time and place America's Crisis is not directly comparable to this event. However, I believe you would be able to derive useful interactions from that event that effectively model the same economic impact today, here in the states.

2) Identify several possible predictor variables of interest and a response variable of interest; how might you reduce this set of predictor variables for an initial analysis of a simple version of the model?

Predictor: Opiate related overdoses per capita
Response: Unemployment Rate per capita

Predictor: Opiate related arrests per capita
Response: College Level Academic Enrollment per capita

Predictor: Amount of RX's written for opiates
Response: Work related injuries or deaths

To reduce the number of predictor variables for analysis I would first look at which of these interactions can best be represented using a mathematical model such as linear regression. Some of the predictor variables could possibly be grouped together to reduced the set of predictors, as well as increase the overall sample size. An example of this would be to combine the amount of synthetic opiate drug busts with the amount of doctor opiate prescriptions. This gives a good general total of illicit and legal opiates out on the streets to compare with something like college enrollment.

3) What variable reductions should be considered to simplify the model?
Ideally I would want to find one or two response variables that had a strong correlation with opiate addiction and served as a reliable indicator of negative economic impact. Age, region, demographic, lifestyle, sex, and type of opiate preferred (heroin vs. prescription opiates) might all be reduced or combined to reduce variance in the analysis and provide simplicity. The goal, ultimately, is to find one predictor variable related to the opiate epidemic that strongly correlates with a single response variable related to negative economic impact. One possible example of this would be the number of opiate overdose hospitalizations vs. the increase in costs for insurance premiums.

4) What data or predictions would you use to validate your model?

As discussed previously, I think it would be insightful to compare The United States economic recovery as opiate availability diminishes vs. China's economic recovery after The Opium Wars. In addition, I think it would be valuable to use a region in America that has very low opiate use and use it's economic performance as a baseline to compare with my response variable. If several control regions in The US have comparable economic performance to the study despite having low levels of opiate use then I will know that the model I constructed is invalidated, and needs to be refined.