Peer Review of Lab 3 for Section 3, Group 4

Provided by Ramiro Cadavid, Katie Mo, and Errett Hobbs

General feedback

- Very nice formatting and organization (the table of contents help the reader understand the structure of the report)
- Overall theme is nice!
- Be careful about making claims of causality throughout (see examples in both the introduction and policy hypothesis sections). The instructions indicate that we should aim for causal effects, but I don't know that you can claim they exist working within your framework. For example, how do you know that areas with low crime rates don't drive lower crime rates (and not the other way around as you hypothesize)? Such a situation could exist if a county's crime rate is low because its median income is high, which eliminates a motivation to commit crime.

Section-specific feedback

1 Introduction and Research Question

- avgsen could also help answer the research question. Consider including or justifying why it was not included as a variable of interest and was instead included as a control.

- Conceptually, it may be useful to precise what is meant by tougher, since it may imply only longer sentences and a higher probability of prison. On the other hand, a higher probability of arrest and conviction may not be interpreted as a tougher system, if it usually does not lead to prison, or high sentences.

- I'd be more specific and clearly articulate what you mean by a "tougher" criminal justice system.

- Your base model doesn't address the severity of punishment. Consequently, I'd omit any discussion of severity from your introduction.

2 Data Loading and Cleaning

- It might be also useful to explore if there are unusually high concentrations of observations around certain values. This may indicate measurement errors.

- The missing counties shown on your map are not geographically diverse; they cluster to the extreme western and eastern ends of the state.

3 Model Building Process

3.1 Selection of the Outcome Variable

- In the histogram, there seems to be a high concentration of observations around certain values. It may be worth checking if there is a reason for this unusual concentration.

Peer Review of Lab 3 for Section 3, Group 4

Provided by Ramiro Cadavid, Katie Mo, and Errett Hobbs

- Your ggscatmat plots don't plot crmrte.log as a dependent variable. This also makes it difficult for me to assess the trends you observe.

3.2 Explanatory Variables, Base Model

- You have a comment in plot_grid(p1, p2, p3) as if you want to make it a 1 row, 3 columns grid. To do this you can replace the plot_grid(...) line with **par(mfr = c(1, 3))** before p1<-..
- You may want to try log transformations of the independent variables to either if the model's fit or the CLM assumptions are improved.
- Provide more explanation on why you transformed or did not transform the independent variables.

3.3 Explanatory Variables, Extended Model

- It might be useful to check for other variables as possible controls. One option is to do correlation analysis to see which variables seem to affect both crime rate and the independent variables of the base model
- Perhaps comment on how log transformation of avgsen and polpc affects the interpretability of your results
- "None of the additional variables seem to have a perfect linear relationship with the independent variable" → should be dependent variable?

4 Regression Models

4.1 Base Model

4.1.1 Coefficients - M1

- Not sure what this sentence mean, may want to consider revising it: "It is worth noting that the coefficients for prbarr and prbconv amplify the effect of increasing prbarr or prbconv. This is not true for the coefficient of prbpris."

4.1.2 Goodness of Fit - M1

4.1.3 6 CLM Assumptions - M1

- Random sampling: at a point it seems to be suggested that the sample is random, but later it seems to suggest the opposite. You may want to make this more clear.
- It would be nice to see a summary of the assumptions together, instead of having to "see below"
- Change title header "6 CLM Assumptions" to something else that doesn't start with a number; when I read the first one, I read it as another subsection (i.e., like "4.1.3.6").
- For the reader's convenience, I'd just cover CLM assumptions 4-6 in your summary as opposed to directing the him or her to the text below.

Peer Review of Lab 3 for Section 3, Group 4

Provided by Ramiro Cadavid, Katie Mo, and Errett Hobbs

4.1.4 Model Diagnostics - M1

- The findings of violations could be used to suggest a different specification of the model that could improve them.

4.2 Extended Model

- The following can be tested using scatterplots and/or correlations: "The findings of violations could be used to suggest a different specification of the model that could improve them."

4.2.1 Coefficients - M2

- log(Average Sentence) and log(Police per Capita) are not in the same order in your list of coefficients as compared to the equation you provide immediately above the code cell that calculates m2.

- the interpretation for beta4 and beta5 does not match the coefficients, perhaps they are switched

4.2.2 Goodness of Fit - M2

- To check for robustness, it may be useful to compare how the coefficients of your variables of interest change across models.

- May also want to check adjusted R2

4.2.4 Model Diagnostics - M2

- Maybe also mention that the CLT can be invoked not only because sample size is >30, but also because there is no extreme skewness in the distribution of the residuals.

- Perhaps organizing the information based on CLM assumptions rather than plots/tests would make the analysis have better flow

4.2.5 Interpretation and Conclusion - M2

- I think it'd be valuable for you to discuss how your key coefficients change from model 1 to model 2. Additionally, β_4 in model 2 is negative, so it predicts a 0.62% *decrease* (NOT increase) in the crime rate with each 1% increase in police per capita.

- I don't think your model supports a policy prescription to increase prison time in an effort to reduce crime rate. β_5 is very small; doubling the average prison sentence (i.e., raising it by 100%) is only associated with a 6.5% reduction in crime rate. Moreover, your coefficients for *prbpris* are positive in both model 1 and model 2.

4.3 Kitchen-sink Model

4.3.1 Coefficients - M3

- This section is empty. If this is intentional, it may be worth including a sentence saying so.

4.3.4 Model Diagnostics - M3

- This section seems to be incomplete, since it has much less analysis and tests than the same section for M1 and M2

Peer Review of Lab 3 for Section 3, Group 4

Provided by Ramiro Cadavid, Katie Mo, and Errett Hobbs

4.3.5 Interpretation and Conclusion - M3

- It might be useful to check for robustness, i.e. compare if the coefficients of the variables of interest changed compared to M1 and M2.

- How does the conclusion tie in with the research question?

- I think that the interpretation of what the coefficients for density and % young male indicate is not well supported by this model.

- I think you could squeeze more value from this model by elaborating on what it tells you about the coefficients for your key variables. How do their values change relative to your other models? Does this tell you anything about your base model?

4.4 Regression Table

- This would also be a good place to check for robustness (if the coefficients of the variables of interest change significantly with different model specifications.

5 Omitted Variables Analysis

5.1 Bias Table

- Need to include sign of beta1, beta2, and beta3 to determine whether bias is overestimated or underestimated

- Explanations seem more idealistic than a discussion of the trend

- I don't agree with all of your omitted variables assessment. For example, how many crimes are committed with registered guns? If a criminal is worried about using a registered gun, then can't they just get an unregistered one?

- The bias table is very wide, which makes it difficult to read.

5.2 Bias Conclusion

- "overestimating" is probably more clear than "superestimating"

- How does the analysis of bias affect your policy recommendation?

6 Conclusion

- Be careful about implicitly making causal claims in your policy prescriptions.

- Although it may very well reduce crime, I'm not sure that the American public would be supportive of aggressive video monitoring of public sites.