**[1] What implicit claim about causality does Obama's "cycle of crime" theory assert?**

* Causality, i.e. X 🡪 Y
  + Going to jail (particularly for the first time) 🡪 Higher probability of performing criminal activity

**[2] Your friend has an ingenious idea. He/she has detailed case data about criminal sentencing in a large jurisdiction for everyone charged with a felony. The data includes the length of the prison sentence (in days), and whether the person was convicted of a *second* crime after he/she was out ("recidivism"). This seems to be what the "cycle of crime" theory is talking about.   
  
The proposed research design is: Run a regression whose outcome is recidivism and whose main explanatory variable is the length of the prison sentence. React (to) your friend's research design.**

* [1] Correlation =/= causation
  + I believe that this type of regression need not provide a causal understanding. In particular, confounding variables may exist that influence both the independent and dependent variables heterogeneously (for example, having a family member in prison may increase both length of prison stay and recidivism)
  + Moreover, I believe that in order to substantiate any claims as to the above, we would want to see both those who ultimately went to jail in addition to those who did not go to jail, to make a proper comparison – not only those charged with a felony (which I assume to mean those who went to jail)
* [2] Additionally, any effect as described by the above “cycle of crime” may be threshold-based vs. continuous (i.e., length of prison sentence may or may not be the best independent variable)
* [3] Finally, questions of generality/external validity, i.e., to how large a set of people/how general a claim do we desire and can this dataset deliver (e.g., any prison sentence vs. certain types, any geography vs. certain locations, etc.)

**[3] This study**

* Z (instrumental variable) 🡪 X (independent variable) 🡪 Y (dependent variable)
  + (random) Judge assignment 🡪 Going to jail (particularly for the first time) 🡪 Higher probability of performing criminal activity

**[4] Perform a balance test. Does the judge's party really seem to be randomly assigned?**

* (taken from the balance table note) It appears as though average severity of crime is approximately equal in the republican and democratic judge groups, whereas months in jail and recidivates is a bit uneven (though I believe this to be ok, given we need only the attributes of assignment to be random, not the downstream effects such as months in jail and recidivates).

**[5] Describe in words the ``first stage'' of the IV design. Then, create a publication-quality table for the first stage only.**

* The first stage of the IV design will regress the instrumental variable (Z) against the independent variable (X).
* In this case, Z is judge assignment and X is number of months spent in jail (we could alternatively build a threshold based model where X is a binary event – whether or not the individual goes to jail or not)
  + I may be understanding this incorrectly, but it seems that all the data we have except for severity of crime is downstream from Z (i.e., both months in jail and recidivates are functions of Z)

**[6] Interpret the coefficient on your instrument from the first stage.**

* (from the table note) Controls for severity of crime. The effect of political affiliation of judge assignment on months in jail is statistically significant at the 0.01 level, and being assigned a republican judge corresponds to approximately 3 more months in jail than being assigned a non-republican (democratic?) judge.

**[8] Calculate the ratio of the reduced form**

* (Z\_coefficientreduced\_form)/(Z\_coefficientfirst\_stage)
* 🡪 (0.143)/(3.094) = 0.04621848739495798

(I think the intended meaning may be that each additional month in jail corresponds to an increase in probability of recidivism by 4.6%)

**[9] Now complete the IV regression and make a publication quality table of the second stage. Use the setup below.**

* Publication quality table in separate doc, below terminal output is for my records.

A screenshot of text

Description automatically generated

**[10] State the F-stat in your writeup. It does not need to go into your table (although, in an actual publication it would). Is it above the conventional threshold?**

* F-stat: 2397.4; p-value: 0.0000
* I’m not totally sure, I thought that the conventional threshold is an f-stat of about 10, given this looks like it’s over 2000 I feel like something may be off, but the low p-value gives me confidence. Would appreciate your thoughts in review – thank you!

**[11] Compare your answer to question #8 (above) to the IV coefficient in #9.**

* Crazy, almost as if they’re one in the same.

**[12] Complete these sentences.**

* + In the research design above (using randomized judges), the **always-takers** are the defendants who are always experiencing recidivism/go to jail again no matter how much time they go to jail for (or if they go at all) the first/prior time.
  + The **never-takers** are the defendants who are always never experiencing recidivism/go to jail again no matter how much time they go to jail for (or if they go at all) the first/prior time.
  + The **compliers** are the defendants who are less likely to experience recidivism/go to jail only if they don’t go to jail or go to jail for a short period the first/prior time.
  + The **defiers** are the defendants who are more likely to experience recidivism/go to jail only if they don’t go to jail or go to jail for a short period the first/prior time.

**[13] Comment on the monotonicity assumption and the possibility of "defiers" in this setting.**

* Deterministically (sharp?) seems to be violated, might be imperfect or use fuzzy/probabilistic. Conservatively, dataset looks like it has 212 instances where defendants get 0 months in jail but do go to jail later (“experience recidivism”, if that is the proper term for this series of events) – so if we count months in jail vs. a binary, this is likely even worse.
* It looks like months in jail, not controlling for anything (controlling doesn’t appear to change much), has a very low coefficient when correlated to recidivism. So I’m a bit skeptical of the monotonicity assumption, but am even more skeptical of my understanding here – I’d greatly appreciate your thoughts/input.
  + A screenshot of text

    Description automatically generated

**[14] In your dataset, what types of defendants are compliers?**

* Compliers are defendants for whom amount of time spent in jail actually changes their probability of going back to (or to for the first time) jail. E.g., a defendant who if they don’t go to jail, may never go to jail, but in another reality, if they do go to jail, they may then get out only to go back to jail.

**[15] Does the cycle of crime hypothesis appear to be true for the compliers?**

* Many grains of salt. If my understanding of 2SLS and instrumental variable is correct, then the effect/coefficient we uncover only applies to compliers. So, yes – it does appear significant (as per questions [8], [9] and [11].