Government Revenues and Corruption

This exercise analyzes the data from a recent paper that studies whether additional government revenues affect political corruption or the quality of politicians. The paper can be found at:

Brollo, Fernanda, et al. "The political resource curse." *The American Economic Review* 103.5 (2013): 1759-1796.

The authors argue that a "political resource curse" exists - that an increase in non-tax government revenues leads to more corruption and lowers the quality of politicians. First, with a larger budget size, incumbent politicians are more able to grab political rent without being noticed by the electorate. Second, a larger budget attracts challengers with poorer quality so that incumbents' misbehavior is punished less frequently.

The authors wish to identify the causal effect of additional federal transfers on corruption and candidate quality. Their theory states that additional non-tax revenues cause corruption, so they use transfers (the *treatment*) from the federal government to municipal governments as exogenous increases in non-tax revenues. The authors ask whether or not larger transfers lead to corruption, so the outcome is the occurence of bad administration or overt corruption. Since corruption is a somewhat vague concept, the authors use two measurements to make sure that their results do not depend on a particular definition of corruption. To avoid this, the authors use two *separate* definitions of corruption to avoid this - 'narrow' corruption includes severe irregularities in audit reports, while 'broad' corruption is a looser interpretation "which includes irregularities [in audit reports] that could also be interpreted as bad administration rather than as overt corruption" (p. 1774).

The data can found in `corruption.csv` in the `data` folder.

Name	Description
`broad`	Whether any irregularity (this might include bad administration rather than corruption) was found or not.
`narrow`	Whether any severe irregularity that is more likely to be visible to voters was found or not.
`fpm`	The FPM transfers, in $100,000 at 2000 prices.
`pop`	Population estimates.
`pop_cat`	Population category with respect to FPM cutoffs.

Question 1

The authors use a Regression Discontinuity (RD) design. What do the authors use as the *forcing variable* and outcome variable? Discuss why the authors can't simply compare all "treated" and "non-treated" villages. Then, discuss how the RD design addresses this problem. What is one weakness of the RD design?

Question 2

Read in the data below for all villages in the authors' dataset. Then, create three regressions. Regress the broad measure of corruption on:

1. the measure of federal transfers
2. the measure of federal transfers and population.
3. the measure of federal transfers, population, and the population category (as a factor).

Then, repeat this analysis for the narrow measure of corruption (so you will have six regressions in total). Interpret each of your three regressions. Can the coefficient be interpreted causally in these models? Explain why or why not.

Question 3

First, let's perform a simple RD analysis to test whether the cutoffs were properly utilized. One of the population thresholds used for FPM transfers was 10188. This means that villages with a population slightly above 10188 received different amounts of transfers to villages slightly below this population. For this analysis, we will use all villages within 500 people of this cutoff. Specifically, this means to take two separate subsets: one subset of villages with populations larger but less than 500 larger than 10188 and another subset of villages with populations smaller but less than 500 smaller than 10188.

Then, create a plot showing the relationship between population and fpm transfers for these villages. Please add a dotted vertical line to show the location of the cutoff (10188) on the x-axis. Additionally, fit two regressions and visualize them on the plot: one showing the relationship between population and FPM transfers for the subset of villages above the cutoff and another showing the relationship between population and FPM transfers for the subset of villages below the cutoff.

Question 4

Explain the plot that you just created. Why is this a useful analysis to perform? Specifically discuss how the regressions compare for villages right above and right below this cutoff: what do you notice? You will notice that the cutoff isn't "strict" - this means that entries with populations above the threshold will sometimes receive transfers lower than other villages with populations below the threshold. Sometimes villages did not receive the federal funds that they were legally entitled to. This is called "noncompliance," and often happens in real-world settings. However, our estimate at the threshold is what's called an Intention-to-Treat Effect (ITT). Even if we have noncompliance, we can calculate the ITT by using every village who was *assigned* treatment rather than those who simply complied with it. Explain why the randomization makes the ITT a valid estimate to use here. Give an example of another case in which noncompliance may lead us to estimate an ITT.

Question 5

Now, we will perform a Regression Discontuity analysis by taking observations that are close to *any* of the cutoffs. In this question we will be completing the corruption analysis and no-longer focusing on FPM transfers, but instead comparing rates of corruption for villages just above and below the threshold. First, create a subset of the data that contains observations within 500 people of the population cuttofs: $\{10188, 13584, 16980, 23772, 30564, 37356, 44148\}$. For example, we want to include all villages that have populations within 500 people above or below any of these cutoffs (the R 'OR' operator, `|`, may be useful here). Then, use `lm()` to compare the rates of broad *and* narrow corruption for villages just above and just below this threshold (hint: after you take the subset for villages within this range, create an indicator variable to show whether the remaining observation is above or below the given cutoffs. Finally, use `lm()` to compare the means for observations at the two levels of this indicator variable). For one measure of corruption (your choice!), also compare the average rates of corruption (using the function `mean()` or `tapply()`) for villages just above and below the cutoffs and show that you get the same results as with `lm()`. Explain why.

Which analysis, Question 2 or Question 5, gives a more reliable setimate of the desired causal effect? Why?