Economics 172: Problem Set #3

Due on November 17, 2025 at 11:00pm

Zachary Brandt
zbrandt@berkeley.edu

Problem 1: Weather and Witch Killing

This question builds on the econometric analysis in *Miguel 2005* about Tanzania poverty and witch-killing. You will carry out some econometrics analysis related to that paper and also related to the *Miguel, Satyanath and Sergenti (2004)* article. You may write up your answers using a word processor, include copies/screenshots of your regression tables, and attach a copy of your R script at the end. Alternatively, you may choose to produce an RMarkdown file that integrates your code, your written responses, and tables displaying your regression results into a single document (please "knit to PDF" and turn in the resulting PDF file; do not simply turn in your `.rmd` file).

In either case, your submission must include: Your entire R code/script; Your written answers; Your regression output. Please merge all documents into a single PDF before submitting.

Please download "pset3-2025-killing.csv" from the bCourses page.

Use the `read.csv` command to open it in R (or RStudio), either on your local computer, or on UC Berkeley's DataHub, found at https://r.datahub.berkeley.edu/.

This dataset is a partial extract of the data from *Miguel (2005)*, organized such that each observation (row) contains data for a particular village (denoted by the variable vid) in a particular year (denoted year) in Meatu district, Tanzania. In other words, this is panel data.

Variables include:

- `witch_murders`: number of witch murders in a given year village-year
- `oth_murders`: number of non-witch murders
- `any_rain`: indicator (1/0) for whether a drought or flood occurred
- `any_disease`: indicator for whether a disease outbreak (measles, cholera, etc.) occurred
- `famine`: indicator for whether there was an extreme food shortage
- `educat`: average years of schooling in the village
- `trad_relig`: proportion of households practicing tradiational religions

a) Construct a new variable for the total number of murders in a village-year (witch + non-witch murders).

b) Create a table of summary statistics for all variables in the dataset, including the mean, standard deviation, minimum, maximum, and number of observations, using `stargazer`, `summary` or `describe` commands in R. Discuss any noteworthy patterns. Pay particular attention to the murder and rainfall variables.

c) Now consider the effect of extreme weather on murders in the village.

   (i) Install "miceadds" and "sandwich." Using the `lm.cluster` command, regress total murders (in a village in a particular year) on the indicator for whether a drought or flood occurred in that year. Make sure that error terms should be allowed to be correlated ("clustered") across years for the same village (use `vid`). Simply use `summary` to report the results in this question. [Note: Results estimated by `lm.cluster` could not be exported directly with `stargazer` so we use `summary` for simplicity. In the section we will teach how to export clustered regression results in a neater way.]

   (ii) In a second regression, add average years of schooling and proportion of households practicing traditional religions as additional explanatory variables.

   (iii) Interpret both regressions carefully.

Problem 1 continued on next page...

d) Finally, consider a possible instrumental variables (IV) approach. Economic theory suggests that extreme economic hardship—such as a famine—may be associated with more violence, including murders. Famine may be caused by extreme rainfall (which would be the instrumental variable).

(i) Write out the first stage regression, the second stage regression, and the reduced form regression.

(ii) Evaluate whether this is a valid IV approach by discussing the plausibility of the three key IV conditions: Relevance; Exclusion restriction; Exogeneity. What are some specific ways in which each of these assumptions might be appropriate or might fail in this context?

Problem 2: The Primary School Deworming Project (PSDP)