

# Practical: Instrumental Variables

11 December 2014

A Short Course on Concepts and Methods in Causal Inference  
V Edition

## Dataset

For this practical, we will use data from a cluster-randomized community intervention trial on the effect of vitamin A supplements on one-year mortality in Indonesian children [1]. Of 450 villages, 229 were assigned to a treatment in which village children were provided two oral doses of vitamin A. Children in the 221 control communities were not provided vitamin A supplements. This design resulted in 12,094 children assigned to supplements and 11,588 to control. We will analyze this data with the 2SLS approach using the randomization indicator as an IV, ignoring clustering for simplicity.

You will find the data for this practical in the file `vitamina.txt`. You will need to copy them into an appropriate folder to be accessed from Stata (use the command `insheet` using ‘‘*path-to-file*\vitamina.txt’’). The dataset contains the following variables:

`id` sequential id number of study subjects;

`assvita` treatment assignment indicator; 1 if assigned to vit. A, 0 otherwise.

`recvita` received treatment indicator; 1 if received vit. A, 0 otherwise.

`death` one-year mortality indicator; 1 if deceased during follow-up, 0 otherwise.

If you are not familiar with Stata, remember: `help command-name` is your friend!

## Exercises

1. Perform a conventional as-treated analysis by regressing `death` on `recvita` with the command `regress death recvita, vce(robust)` (we use robust standard to account for the heteroschedasticity of the binary outcome). How do you interpret the estimated coefficient of `recvita`?
2. Compute a cross-tabulation of `assvita` by `recvita` using the `tab2` command. Do you think the monotonicity assumption could be satisfied?

3. Compute the first-stage regression of `recvita` on `assvita` and compute the predicted values `recvitapred` obtained from the model. Is `assvita` a strong IV? Use the commands:
  - `regress recvita assvita`
  - `predict recvitapred`
4. Compute the second-stage regression of `death` on `recvitapred` with the command `regress death recvitapred`. How do you interpret the estimated coefficient of `recvita`?
5. Compare the results obtained in the conventional analysis with those obtained by the 2SLS procedure. How do you explain the observed differences?
6. Implement the 2SLS procedure directly using the `ivregress` procedure with the following command:
  - `ivregress 2sls death (recvita=assvita), vce(robust)`

Compare the results with those obtained previously.

## References

- [1] Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol* 2000; 29:722-9