HIV Documentation

This data comes from the study: Thornton, Rebecca L. 2008. ‘The Demand for, and Impact of, Learning Hiv Status.’ American Economic Review 98 (5): 1829–63. Online [here](https://www.aeaweb.org/articles?id=10.1257/aer.98.5.1829). The data originally came from the [casualdata](https://cran.r-project.org/web/packages/causaldata/index.html) package. This version of the dataset excludes all observations with missing values of any of the variables.

# Description of the data collection, from the paper:

“The Malawi Diffusion and Ideational Change Project (MDICP) is conducted in Malawi, a land-locked country located in southern Africa (Figure 1). This collaborative project between the University of Pennsylvania and the Malawi College of Medicine is an ongoing study of men and women randomly selected from approximately 120 villages in the districts of Rumphi, Mchinji, and Balaka, located in the Northern, Central, and Southern regions respectively. Approximately 25 percent of all households in each village were randomly selected to participate in 1998, and ever-married women and their husbands from these households were interviewed in 1998, 2001, and 2004. During data collection in 2004, an additional sample of young adults (ages 15–24) residing in the original villages was added to the sample.”

“Between May and August of 2004, nurses from outside each area offered respondents free tests in their homes for HIV and three other sexually transmitted infections (STIs): gonorrhea, chlamydia, and trichomoniasis. At the time that tests were offered, respondents were given pretest counseling about HIV prevention strategies…After taking the HIV test samples, nurses gave each respondent vouchers redeemable upon obtaining either HIV or STI results. Voucher amounts were randomized by letting each respondent draw a token out of a bag indicating a monetary amount [which included 0]…Vouchers ranged between zero and three dollars, with an average total voucher amount (including zeros) of 1.01 dollars, worth approximately a day’s wage.”

“Two to four months after sample collection, test results became available, and temporary test results (VCT) centers, consisting of small portable tents, were placed randomly throughout the districts. Based on their geospatial (GPS) coordinates, respondents’ households in villages were grouped into zones, and a location within each zone was randomly selected to place a tent. The average distance to a center was two kilometers and over 95 percent of those tested lived within five kilometers. Distance to the VCT center is calculated as a straight line and does not account for roads or paths. In most cases, tents were placed in the exact randomly selected location and paths were created for easy accessibility.”

# Variables

villnum - ID of the village the participant lived in

`results`` - Whether the participant recieved their HIV results

distvct - distance from the test result center in kilometers

amount - total incentive in USD

incentive - whether the participant recieved any incentive

age - age of the participant in years

hiv2004 - HIV results, coded 1 for positive test, 0 for negative test, and -1 for can’t be determined/missing (a little unsure about the coding of -1, this is a best guess)