Modeling Ischemic Heart Disease

Intro and Question

The Global Burden of disease attempts to map the impact of over 300 causes of death among 188 countries across a 36 year time series(1980-2015) within 20 different age groups ranging from neonatal(0-7days) to elderly (80+ years). In order to make estimates where data is incomplete or lacking across all ages, times or geographies, rates of a singe cause of death is modeled and then used to impute missing estimates. In order to contribute to improving this modeling process this project will look at the death rates of ischemic heart disease (IHD) among males.

Dataset

In order to model deaths due to ischmeic heart disease a number of data sources will be used.

- Vital registration records from 120+ central governments
- Verbal autopsy data from 30+ countires
- UN pop data on population numbers
- Word bank gdp estimates
- IHME Eduction estimates

Verbal autopsy and vital registration data will provide all cause mortality numbers and IHD death numbers. Population numbers will allow the data to scale correctly as per gompertz theory we expect data to change smoothly in log rate space across time and age within a single geography. GDP and education numbers will be used as covariates for inference.

The Spatio/Age/Temporal process

Using only education and gdp as an mechanism for inference we would expect that errors would be correlated along three separate dimensions age, time and location. Age groups that are adjacent will likely have similar rates of death due to IHD. The same can be said about estimates that are adjacent in years. We expect geographies to be smooth by similarity in development status and location (which we can operationalize by a neighborhood matrix).

Evaluation

Model development will be evaluated by holding out a portion of the data and computing out of sample RMSE as well as out of sample nll for the model.