

POLS/CSSS 510: Maximum Likelihood Methods for the Social Sciences

Problem Set 6

Professor: Christopher Adolph, Political Science and CSSS

Fall Quarter 2017

Due in class on Thursday 7 December 2017

General instructions for homeworks: Homework can be handwritten or typed. For any exercises done with R or other statistical packages, you should attach all code you have written and all (interesting) output. Materials should be stapled together in order by problem. The most readable and elegant format for homework answers incorporates student comments, code, output, and graphics into a seamless narrative, as one would see in a textbook.

Problem 1: Modeling mortality in Iraq with count models

The dataset `iraq.csv` contains survey data on mortality in Iraq from Roberts *et al*, “Mortality before and after the 2003 invasion of Iraq: a cluster sample survey,” *Lancet*, Nov. 20, 2004, 1857–1864. It also contains supplemental data added by your instructor. You should read Roberts *et al* before starting the assignment.

These data have been kindly provided by Roberts *et al* on the understanding they will NOT be shared or used in any research. Please do not circulate these data or retain them after end of the quarter.

- a. [15 points]** Fit a Poisson regression to the variable `deathpos`. You may use any covariates, transformations or interactions you wish. Be sure to justify your reasoning, and to choose the model that makes most theoretical sense to you.

NB: Not all variables in the dataset are reasonable covariate candidates. Be careful you do not control for anything that is functionally equivalent to the outcome variable.

- b. [10 points]** Explore the goodness of fit of the regression you fit in **a**. Show that the regression you fit dominates at least two plausible alternative specifications. Answering this problem will require you to choose at least two other models, fit them, and show, using tests of your choice, and/or substantive arguments, that your model from **a** is better.
- c. [10 points]** Check for outliers. You should find at least one. Remove the outlier(s), and re-estimate all three specifications. Has removing the outlier changed your judgment of which model is best?
- d. [20 points]** Having settled on a favored specification, interpret your results for a general audience. Choose two explanatory variables, and show both graphically – and explain in words – the expected effect of those variables on `deathpos`. Be sure to include confidence intervals for all quantities you report. Carry out this exercise for the models with and without the outlier(s) included. How much difference do(es) the outlier(s) make?
- e. [10 points]** Using your favored specification, fit a Negative Binomial regression to the variable `deathpos`, and test for overdispersion.
- f. [10 points]** Suppose we wanted to report the dispersion of the negative binomial model using α , where $\text{Var}(Y) = \mu + \mu^2\alpha$, as in class. Is the overdispersion parameter you obtained in **e** the same as α ? If not, transform it to be α , or write your own negative binomial code that will produce an estimate of α .
- Bonus [+10 points]:** Calculate or simulate the confidence interval of α .

Data from Roberts *et al* (2004)

Variable	Description
clustnam	The name of each cluster. An observation reports survey totals for a cluster drawn at random from Iraq. The cluster name includes the name of the governate of Iraq containing the cluster.
clustid	An id number for the observation.
deathpre	Deaths from all causes in the 14.6 months prior to the invasion.
permopre	Total number of persons times months in the cluster prior to the invasion
mortpre	Pre-invasion mortality rate: $12 \times 1000 \times \text{deathpre}/\text{permopre}$.
deathpos	Deaths from all causes in the 17.6–18 months after the invasion.
permopos	Total number of persons times months in the cluster after the invasion
mortpos	Post-invasion mortality rate: $12 \times 1000 \times \text{deathpps}/\text{permopos}$.

Data added by your instructor

Variable	Description
sunni	Whether Sunnis are a significant ethnic group in the <i>governate</i> from which the cluster is taken; based on my cursory judgment of maps from the University of Texas Perry-Castañeda Library.
kurd	Whether Kurds are a significant ethnic group in the <i>governate</i> from which the cluster is taken (see note for sunni).
shia	Whether Shias are a significant ethnic group in the <i>governate</i> from which the cluster is taken (see note for sunni).
turnout	Turnout in the <i>governate</i> from which the cluster was taken in the January 30 Iraqi national election, as reported by the Independent Electoral Commission of Iraq.
regvtrs	Registered voters in the <i>governate</i> from which the cluster was taken, as reported by the Independent Electoral Commission of Iraq (registration did not take place in some governates; these are left missing).
trrate	The rate of turnout of registered voters.
pop	The estimated January 2003 population (see Roberts <i>et al</i> Table 1).
tpctpop	The rate of turnout as a percentage of estimated total population.

Map of ethnicity:

www.lib.utexas.edu/maps/middle_east_and_asia/iraq_ethno_2003.jpg

Map of governates:

www.lib.utexas.edu/maps/atlas_middle_east/iraq_divisions.jpg