U.S. News & World Report

Stata Coding Exercise

If you need a trial copy of Stata, StataCorp has generously (and rapidly) provided candidates with copies before for the purposes of completing this code test. You can contact them here or at service@stata.com to request this.

This exercise deals with simulated administrative claims (data file attached). After working through the exercise, please provide us with your code, your responses to #1, #6, #8, and #9, your graph for #6, and your log and output files. Please write clearly and comment your code, with the expectation that it will be reviewed by someone who has never seen it before.

The provided data file (USNtest\_Stata.csv) is structured as follows:

| **Variable Name** |
| --- |
| patientId |
| age |
| admitDate |
| systolic |
| procedure1 |
| procedure2 |
| procedure3 |
| procedure4 |
| procedure5 |
| diagnosis1 |
| diagnosis2 |

**Type Description** int Unique patient id int Age in years str9 Date of admission int Systolic BP, mm hg int ICD-9 procedure code int “

int “

int “

int “

int ICD-9 diagnosis code int “

diagnosis3 int “

diagnosis4 int “

diagnosis5 int “

aha\_id byte Unique hospital id The provided data file (hospital\_Stata.csv) is structured as follows:

| **Variable Name** |
| --- |
| aha\_id |
| bdtot |

**Type Description** byte Unique hospital id int Bed count

1) Create descriptive statistics for the above variables, and run any other code you would normally use to evaluate quality of newly imported data. Describe any concerns you may have about the suitability of this dataset for further analysis.

2) Subset the data to include only those patients age 65 or older (>=). Save this new working file and use it for the rest of the exercise.

3) Create a dummy variable indicating whether each admission involves a readmission, defined as a subsequent hospitalization for the same patientId within 30 days of the index admission.

4) Create a dummy variable indicating whether each admission involved coronary artery bypass graft (CABG) surgery:

● Inclusion criteria are any procedure code in the following group: 3610, 3611, 3612, 3613, 3614, 3615, 3616.

● Exclusion criteria are any procedure code where the first three characters are 350 or 351.

5) Install the community-contributed Stata module “elixhauser” (ssc install elixhauser). Run the module, using the ICD-9 enhanced option, creating binary variables to flag each Elixhauser comorbidity and a count of comorbidities.

6) Merge the hospital\_Stata.csv file to your data. Create a bar graph of average age by bed count for CABG and non-CABG patients and briefly describe your observations in a few sentences.

7) Specify and run a regression model that estimates how likely a patient is to be readmitted among patients undergoing CABG surgery, controlling for systolic blood pressure and the number of Elixhauser comorbidities in the admission record. Update the dataset with a predicted probability of readmission for each patient undergoing CABG surgery.

8) Interpret the model output and explain your model choice over the alternatives. Discuss the assumptions that must hold in order to obtain unbiased estimates from your choice of model.

9) In a few sentences, interpret the graph below.

