Proposal Outline

Alyssa Forber

October 16, 2017

Background

Opioid information

-explain issue of opioid addiction

-rare outcome

-whether or not to prescribe to patients if likely to become addicted

Data

Describe dataset
-show a table1 of the data
-number of subjects
-outcome
-variables in dataset

Aims?

Describe the aim of the thesis

-accurate predicting, better sens and spec for unbalanced outcome

-using and comparing methods of cutpoints and sampling

Methods

describe analysis

Model

-temporal split of data to get train set
-using logistic regression for the easier interpretation
-clinical colleagues understand this
-stepwise selection to reduce variables in data

First approach

Approach 1:
-predict and change cutoff with train set- no sampling
-explore cutoff probability to get the “best”

ROC methods

ROC:
-explain ROC curves and youden
-show an example ROC curve
-could also be chosen if a certain sens or spec in mind instead of youden

Second approach (sampling)

Approach 2:
-predict and use 0.5 cutoff with sampled train set that is balanced
-explain up, down, and smote (explain smote)

Preliminary Results

Explain prelim results (sens, spec, accuracy, AUC, npv, ppv):
Table of–
-non-sampled with 0.5 cutoff
-non-sample with youden cutoff
-up sample 0.5
-down sample 0.5
-smote sample 0.5

Discussion of Results

Depening on situation the clinician may like different sens/spec
Some may want to be more conservative, others may not (cancer patients in a lot of pain who need opioids)

Moving Forward

-simulation of different outcome percents, see when you could run model without sampling or changing cutoffs
-trying different sampling other than just the defaults for each method
-different ways to do stepwise selection

Questions

Should I show code in this?
Am I giving the talk as if for an audience of statisticians or clinicians or mixed?
What more should be expanded on?
Any figures or tables that would be helpful
Should I go into detail about what logistic regression and predicting is doing behind the scenes?
Do our prediction percents need to be back transformed?? Because it's a logit model? (relates to project for adv data)
Is it ok to show all that I've done or will they think I've already done too much?
Other ideas of things to be done moving forward?
Work on Github?