LS 123 Data, Prediction, and Law

Meeting 10

**Racial Bias in ARAIs, COMPAS’ Problems, ML Primer and Logit**

1. **DIP proposal and PSET 1: questions**

**PSET 1** geopandas problems?

project proposal—the question and the data you will use; you can get into how you will model your attempt to answer your question later

1. **Skeem & Lowenkamp 2016**—selecting into (an incarcerated) population
   1. first some questions to start us discussing
      1. What risk assessment instrument are Skeem & Lowenkamp evaluating? What does it predict?  Is there anything about the variable that they’re predicting that should worry us? Why?
      2. What is the main claim of the Skeem & Lowenkamp paper? what evidence do they use to support their claim?
      3. How do the authors explain the “disparate impact” of the predictive instrument? Why is the association between mean scores on the instrument and a person’s race not a problem, in their view?
      4. What tool do the authors use to assess the accuracy of the predictions their instrument yields up? Explain it briefly (and yes, go ahead and use the Internetz!).
      5. How does the argument for using previous criminal history connect to Harcourt’s argument about the ratchet effect on certain sub-groups?
      6. What factors contribute to arrests? Are arrests random? What do you need to arrest someone?
   2. First, let’s not worry too much about the psychometric summary statistics and measures of difference (or measures of association)
      1. Psychologists have their own language for measuring association and for measuring difference
      2. For now we will stick with the ideas and the authors’ presentation of the evidence
   3. connect to Harcourt ch. 1 (and Harcourt 2015)
      1. Harcourt: use of actuarial tools to decide on criminal justice supervision oversamples the highest risk part of a population
      2. By focusing on criminal history, predictive tools actually subject the initially identified population that was oversampled to even greater surveillance
      3. Focusing on *selective incapacitation* seems smart in utilitarian terms, but reliance on criminal history as predictor means that you will create a bigger footprint of criminal history in the targeted group
      4. Thus you will have continued hyperincarceration of African Americans
   4. Skeem & Lowenkamp: wait, not so fast—predictive tools like **P**ost **C**onviction **R**isk **A**ssessment work
      1. Separate what they claim to be retributive thrust of the punitive turn in criminal justice from what they say are the utilitarian aims of prediction—get the amount of punishment (and the money spent on it) “correct”
      2. Is this still “selective incapacitation”? Skeem and Lowenkamp say that they are only talking about prediction in terms of deciding on community supervision (i.e. probation), rather than “front end” sentencing decisions—what will probation consist of rather than the decision itself
         1. Other caveats? They are talking about federal offender population that has already been convicted, mainly for drug offenses
         2. So the PCRA might not be valid for the vast bulk of incarcerated population (who are convicted under state law, which is most criminal law)
         3. Also, we need to decide first as a society what we want to do with people who are convicted, and the predictive instrument isn’t for sentencing decisions other than alternative sanctions
      3. But they are still talking about recidivism (**measured as re-arrest**, not conviction), not “just desserts” or rehabilitation—so they are interested in getting the most bang for the buck: selective incapacitation in another form?
         1. what leads to arrest? is it a random process? what do you need to arrest someone? (a warrant, or probable cause to believe that the person has committed or is committing a crime); lower standard than conviction
         2. arrest is endogenous: the arresting person has knowledge of your criminal history in making the decision, at least a lot of the time
         3. surveillance itself is required for an arrest (and so if you watch parolees, they are more likely than others to be arrested)
         4. What we want to do is prevent recidivism (although maybe by providing services rather than deciding to continue incarceration)
         5. Also, the idea of just desserts is congruent with the most important predictor of reoffending: criminal record
      4. Key point? There is not an important independent effect of race (independent variable) on the outcome (dependent variable) they care about—recidivism (**arrest** for a new nonviolent or violent crime)
         1. That does not mean there are not group differences between Black and white offenders on the prediction instrument
         2. But, the difference is mediated by criminal record and by the other clusters of variables (remember those? there were five)
            1. Criminal history (prior **arrests**, age, probation/parole violations, offending pattern, institutional adjustment, violent offense)
            2. Employment and education (years education, recent work history, unemployment)
            3. Social networks
            4. Substance abuse
            5. Attitude (motivation to change)
         3. Some of these are fixed (criminal record) but some can change over period of supervision (job or education record, substance abuse, motivation)
      5. So for the limited purpose of predicting how “risky” offenders on community supervision will be, the PCRA is fine, right?
         1. The logistic regression model, which predicts the log of the odds of the dependent variable being true, shows that we cannot reject the null hypothesis—that race has no effect on re-arrest—when we account for PCRA score
         2. Of course, the model does not account for much of the variation in the probability of arrest
         3. The other problem is that there is a selection process into the initial pool of people who have criminal records—see Harcourt
         4. But the key idea is that process the initially selects people into the higher-risk pool is biased based on race: Black men especially are more likely to be selected into it
         5. So the key predictor, criminal history, is not independent of the bias we are worried about but is actually caused by that bias
         6. so yes, criminal history mediates (or is an “intervening variable”) that ties race to outcome, but does that really mean that taking criminal history into account means that the instrument itself is unbiased?
         7. bonus: how good of a predictor is the PCRA?
            1. race variable alone does not meet the standard to reject the null, but that’s all we can say
            2. the fit of the model is not that great (pseudo-R2)—down there with other social science models
            3. AUC-ROC—how good of a classifier

0.74 means it is better than flipping a coin, but is it good enough?

they don’t give details (like a confusion matrix) on where the classification errors are

what if the predictor tends to generate false negatives for white offenders and false positives for black offenders? also, are false negatives worse (from a justice standpoint) than false positives?

* + 1. So, what should we do? We want to be efficient, but we don’t want to be unjust (and deciding who to punish based on race would be unjust)
       1. Harcourt: be less punitive
       2. Skeem & Lowenkamp: that is a political decision, but we still will need to use our resources wisely and so we will still need predictive tool
    2. their measure to assess the PCRA? the area under the Receiver Operating Characteristic curve
       1. from assessment of accuracy of radar operators during WWII
       2. represents the total amount of signal as opposed to noise, that is, the tradeoff between specificity and sensitivity; how well does a test discriminate between alternatives (like “has the disease” vs “does not have the disease”)

1. **COMPAS prediction instrument and its problems (20 min)**
   1. today we read a critique of COMPAS as a prediction instrument: it is biased, and in a way that goes against the norms of the United States (see 14th Amendment)
   2. we will also read a rebuttal (the Flores piece for next time)—you critics are just wrong about the bias
   3. note that the **data are available at the link in the article** in the form of a Jupyter notebook (written in R)
      1. the problem is that it is a survival analysis type of model: what is the accuracy of prediction after someone has been out for two years, i.e., were they arrested for another crime
      2. the key idea is that the calculation of risk shows that all else equal, it matters whether the person being evaluated is Black (compared to the base case of White)
   4. main findings
      1. all else equal, race matters (as does gender) in getting a high score on the COMPAS instrument (there was an effect for being Black even controlling for actual recidivism in the two year period
      2. the COMPAS instrument does predict two year recidivism (Cox survival model) but not as well as Northpointe claims it does, and seems to underestimate probabilities for White offenders and overestimate probabilities for Black offenders
      3. we will read a defense of COMPAS for next time, and saw a court challenge in Wisconsin v. Loomis; next time we will talk about a replication study by Dressler and Farid (Prof. Hany Farid himself from the I-School
      4. once we get to machine learning algorithms in a few weeks you might want to check out the Broward county data that the ProPublica team left for us in their Git repository, if people are interested (it has the COMPAS score and the variables the ProPublica team generated)
      5. I didn’t run with the Broward County data because the modeling is a bit fancier than what I wanted to get into, but we will touch upon logistic regression models and prediction in a few weeks
   5. logistic model
      1. PPT slide: converting something discontinuous into a continuous estimator
         1. remember we are estimating whether or not someone gets rearrested in logistic model
         2. graphically in two dimensions (so predictor and outcome)—draw/project on board
         3. that’s a yes or no outcome, which violates the assumption that the dependent variable or prediction is a continuous measure
         4. thus we need to convert the yes or no outcome into a continuous variable
         5. we will talk more about logistic regression models and prediction later
      2. ln(Pr(event)/1-Pr(event)) = intercept + vector of predictors + error
      3. the coefficients are not as easy to interpret as OLS 🡪 you have to plug numbers back in to see what the difference is between black and white offenders
      4. note that the base case in the model is white, and that all the other race variables are dichotomous (1 if you are in category, zero if you are not)🡪
   6. Gelman blog: the all else equal problem (a regression model isolates the effect of a single variable and holds other variables constant; to interpret logit, you have to select the values to plug into the model, and some combinations of values represent very small proportions of the overall sample)
   7. Cox proportional hazards model: do people make it two years without rearrest?
      1. another species of logistic regression model
      2. here the question is not just yes or no, but what is the probability of survival for two years
      3. I’m no expert in survival analysis (or event history analysis) but it is widely used in the social sciences and in medicine
      4. race maybe mattered in how the score predicts recidivism over two years, but not for prediction of rearrest for violent crime
      5. there were more false positive “high risks” for Black offenders and more false negatives for White offenders

“””

The Cox model is expressed by the *hazard function* denoted by h(t). Briefly, the hazard function can be interpreted as the risk of dying at time t. It can be estimated as follow:

*h*(*t*)=*h*0(*t*)×*exp*(*b*1*x*1+*b*2*x*2+...+*bpxp*)

where,

* *t* represents the survival time
* *h*(*t*) is the hazard function determined by a set of p covariates (*x*1,*x*2,...,*xp*)
* the coefficients (*b*1,*b*2,...,*bp*) measure the impact (i.e., the effect size) of covariates.
* the term *h*0 is called the baseline hazard. It corresponds to the value of the hazard if all the *xi*are equal to zero (the quantity exp(0) equals 1). The ‘t’ in h(t) reminds us that the hazard may vary over time.

The Cox model can be written as a multiple linear regression of the logarithm of the hazard on the variables *xi*

, with the baseline hazard being an ‘intercept’ term that varies with time.

The quantities *exp*(*bi*)

are called hazard ratios (HR). A value of *bi* greater than zero, or equivalently a hazard ratio greater than one, indicates that as the value of the *ith*

covariate increases, the event hazard increases and thus the length of survival decreases.

Put another way, a hazard ratio above 1 indicates a covariate that is positively associated with the event probability, and thus negatively associated with the length of survival.

In summary,

* HR = 1: No effect
* HR < 1: Reduction in the hazard
* HR > 1: Increase in Hazard

“””

1. **Lab 11—Math in SciPy**
   1. please see the explainer on ML and Optimization in bCourses Sp22 Slides (maybe include a bit)
   2. This is important stuff; this will help you understand the ML algorithms we start talking about next week
   3. This will all be familiar to you if you are a computer science or data science or applied math major, but bear with us (or at least with me—I was an Asian Studies major)
   4. Common kinds of problems: fitting a line using Ordinary Least Squares—this time represented in the language of linear algebra, optimization, summation (integration),
2. Skeem & Lowenkamp take-aways
   1. ARAIs are useful in assessing how to release someone into the community
   2. Incapacitation of people likely to re-offend is still the goal here; policy should be set by overall goals of system of punishment, not merely reducing risk of reoffense
   3. Reoffense, in their main model, is measured in two ways: arrest, and arrest for a crime of violence
   4. Race has no independent effect in the PCRAs assessment of risk
   5. Furthermore, based on its ability to identify likely recidivists (as measured by area under the AUC-ROC curve), it is a useful predictive tool
   6. If there’s a problem, though, maybe it is not in the predictive instrument so much as it is in the selection of people who get classified as “arrested”—they attempt to minimize the problem by using arrest for crimes of violence as their outcome
   7. AUC-ROC curve: a representation of how much signal there is relative to noise; e.g., across the levels of sensitivity of the classifier, what is the ratio of true positives to false positives
      1. Maybe the radar analogy is pretty good: what proportion of the blips are actual airplanes and not a ghost—what we are aiming for is 1 (all the blips are airplanes, and all the negative readings are not airplanes—i.e. It classifies every signal accurately)
      2. The explainer from Medium does a good job of putting into machine learning terms—you are plotting sensitivity against 1-specificity
      3. This comes up as a measure of classifier accuracy fairly often, and as you will see is a source of contention as to whether it is the one true measure of model quality
      4. The AUC of the PCRA measure was about 0.74—there is a 74% probability that the model will classify someone correctly as a recidivist or non-recidivist; is that good, or not good?