

# POLI 170 Data Analysis Assignment 3

Due 3:29PM Tuesday November 9, 2021

Please submit this assignment by uploading your html (`.html`) and code (`.Rmd`) files onto Canvas before the due time, with easy-to-recognize file names (e.g., `assignment3_KirkBansak.Rmd`). Your homework will be graded based on completeness, accuracy, and readability of both code and written answers.

The point allocation in this problem set is given by:

Q1.1	Q1.2	Q1.3	Q1.4
2	2	2	4

Q2.1	Q2.2	Q2.3	Q2.4	Q2.5	Q2.6
4	4	4	4	4	4

Q3.1	Q3.2	Q3.3	Q3.4	Q3.5	Q3.6	Q3.7	Q3.8	Q3.9	Q3.10
2	4	2	3	3	4	2	6	2	4

Q4.1	Q4.2	Q4.3	Q4.4	Q4.5	Q4.6	Q4.7	Q4.8	Q4.9
2	4	2	2	2	2	2	4	4

Q5	Total
10	100

## Intro

The dataset and data dictionary you need for this assignment, which can be found on Canvas, are `broward_data.csv` and `broward_codebook.txt`. The dataset is comprised of over 7000 individuals arrested in Broward County, Florida between 2013 and 2014, for whom the commercial risk assessment algorithm COMPAS was used to predict the risk of recidivism. You should begin this assignment by reading through the data dictionary to familiarize yourself with the variables included.

## Part 1: Summary Statistics

Using the full dataset contained in `broward_data.csv`:

1. Compute and report the proportion of defendants belonging to each racial group.
2. Compute and report the two-year recidivism rate for each racial group.
3. Compute and report the mean COMPAS risk score for each racial group.
4. Comment on these summary statistics. Does there appear to be a relationship between recidivism rates and mean risk scores across the racial groups? Is this what we should expect?

## Part 2: Evaluating Fairness Metrics

Using the full dataset:

1. Compare the COMPAS risk score and COMPAS binary classification variables. What is the value of the classification threshold? And what is the difference between what the risk score and binary classification are supposed to represent?
2. Based on the COMPAS binary classification variable, compute the classification accuracy (i.e. proportion correctly classified) separately for each racial group.

3. Now, similar to the assigned ProPublica article, we will focus on the comparison between Black and White defendants. Compute and report the false positive rate and false negative rate separately for Black and White defendants.
4. Comment on the results. What definition(s) of algorithmic fairness do these results allow you to evaluate, and what are your conclusions with respect to whether COMPAS's performance meets the fairness definition(s)?
5. Now, imagine that an alternative classification threshold was applied to the algorithmic risk score. In particular, apply a threshold of 7 such that any defendant whose risk score is greater than or equal to 7 is classified as a 1 (recidivism is predicted) and otherwise classified as a 0 (recidivism not predicted). Based on this new classification, compute the false positive and false negative rates separately for Black and White defendants.
6. Comment on the results. What do you notice about the results for this new classification relative to the previous? What could be the implications if it were permissible to apply different thresholds for different racial groups?

## Part 3: Building our own Predictive Models

In this section, we will be splitting the dataset into training data and test data subsets so that we can evaluate out-of-sample predictive performance. To do so, we will use the variable `random_split`, which the teaching staff generated randomly for this section.

1. To begin, split the full dataset into two separate data frames: a training dataset containing all observations whose `random_split` variable value is "train", and a test dataset containing all observations whose `random_split` variable value is "test".
2. Using only the training data, fit a logistic regression model in which the outcome variable ( $Y$ ) is two-year recidivism and the predictors ( $X$ ) are all of the following:

- Sex
- Age
- Total number of juvenile felony criminal charges
- Total number of juvenile misdemeanor criminal charges
- Total number of non-juvenile criminal charges
- Degree of the charge

You should only include each of these predictors on their own (i.e. do not include any interaction or polynomial terms). Store the model and name it `mod1`.

3. Extract the **in-sample** predictions of recidivism from `mod1`, in the form of predicted probabilities, and plot and display a histogram of these predicted probabilities.
4. Compute and report the **in-sample** Brier score for `mod1`.
5. Now, using the in-sample predicted probabilities, construct a binary classification variable using 0.5 as the threshold (i.e. greater than or equal to 0.5). Based on this, compute the **in-sample** classification accuracy separately for Black and White defendants.
6. Now, apply your already fitted `mod1` to the test data, and generate predicted probabilities of recidivism for the test data. Based on this, compute the **out-of-sample** Brier score overall, as well as the **out-of-sample** classification accuracy separately for Black and White defendants (again using the 0.5 threshold).
7. Compare the in-sample performance to the out-of-sample performance. Is this what you would have expected and why?
8. Now, using the **training data** again, fit a new logistic regression model that is the same as `mod1` except that it also includes race as a predictor. Store this new model and name it `mod2`. (Note that in real-world deployments of risk assessment algorithms in criminal justice, race is typically prohibited from being included. We may be interested to know, however, what the effect would be if it were hypothetically included.) Then, apply `mod2` to the **test data** and follow the same

process as above to compute the **out-of-sample** performance metrics (overall Brier score, as well as classification accuracy separately for Black and White defendants) for `mod2`.

9. Using a scatterplot, plot the **out-of-sample** predicted probabilities of `mod1` vs. `mod2` (with `mod1` on the  $x$ -axis).
10. Comment on the results of the previous questions. Does the explicit inclusion of race in the predictive model appear to have a significant influence on model performance?

## Part 4: Exploring Predictions

1. Now, copy the following into your R Markdown file:

```
hypo_def <-  
  data.frame(sex = c(1,0), age = c(18,41),  
             juv_fel_count = c(0,0), juv_misd_count = c(1,0),  
             priors_count = c(0,1), charge_degree = c(0,1),  
             race = c("black","white")  
  )
```

This will create two hypothetical defendants. After you have inspected the `hypo_def` data frame, describe and compare the characteristics of each of the two defendants in words.

2. Using `mod2` that you already fit in the previous section, compute and report the predicted probabilities that each of these defendants contained in `hypo_def` recidivates. Based on these results alone, can you tell which characteristics of the defendants account for higher or lower recidivism probabilities?
3. Now we will modify the characteristics of the second defendant to gradually match the first defendant. First, modify the `hypo_def` data frame to change the second defendant's race to match that of the first. Re-compute the predicted probabilities of recidivism for both defendants, and comment on the results. How much did that change the results for the second defendant?

4. Keeping the modification you just made, now also change the second defendant's sex to match the first. Again recompute the predicted probabilities of recidivism, and comment on the results. How much did that change the results for the second defendant?
5. Keeping both modifications you already made, now also change the second defendant's priors count to match the first. Again recompute the predicted probabilities of recidivism, and comment on the results. How much did that change the results for the second defendant?
6. Keeping all modifications you already made, now also change the second defendant's age to match the first. Again recompute the predicted probabilities of recidivism, and comment on the results. How much did that change the results for the second defendant?
7. Finally, keeping all modifications you already made, now also change the second defendant's charge degree to match the first. Again recompute the predicted probabilities of recidivism, and comment on the results. How much did that change the results for the second defendant?
8. Comment on these results as a whole. What characteristics seem to be most important in predicting the probability of recidivism.
9. In the full dataset, how different are Black and White defendants on average with respect to those most predictive characteristics? You should use whatever functions or statistics you deem appropriate/useful to answer this question.

## Part 5: Calibration Plot

To the best of your ability, recreate the COMPAS calibration plot shown on slide 31 of the Criminal Justice Case Study lecture slide deck from class (October 29). To do this, you should use the full dataset. Given the data you have access to, your plot will not be an exact match, but it should be very close. You do not need to include the light grey shading around the lines.