Lab 11: COMPAS Case Study – SOLUTIONS

PSTAT 100, Summer Session A 2025 with Ethan P. Marzban

MEMBER 1 (NetID 1) MEMBER 2 (NetID 2)
MEMBER 3 (NetID 3)

July 31, 2025

Required Packages

```
library(ottr)        # for checking test cases (i.e. autograding)
library(pander)      # for nicer-looking formatting of dataframe outputs
library(tidyverse)   # for graphs, data wrangling, etc.
```

Logistical Details

Logistical Details

- This lab is due by **11:59pm on Friday, August 1, 2025**.
- Kindly note that there will be **NO LATE SUBMISSION PERIOD** for this lab, as the quarter ends at 11:59pm on Friday August 1.
- Collaboration is allowed, and encouraged!
  - If you work in groups, list ALL of your group members' names and NetIDs (not Perm Numbers) in the appropriate spaces in the YAML header above.
  - Please delete any "MEMBER X" lines in the YAML header that are not needed.
  - No more than 3 people in a group, please.
- Ensure your Lab properly renders to a `.pdf`; non-`.pdf` submissions will not be graded and will receive a score of 0.
- Ensure all test cases pass (test cases that have passed will display a message stating `"All tests passed!"`)

Lab Overview and Objectives

Welcome to this final (and bonus) PSTAT 100 Lab! As a reminder, you are not required to submit this lab if you are satisfied with your grades on Labs 1 through 10; however, if you do submit this lab your score on this lab will replace your lowest Lab score from the quarter.

This lab is structured differently from the previous labs in this course, as we will be examining an already-conducted Case Study. My main goals for you in this lab are to:

- Read through a very well-crafted analysis report
- Become aware of ways in which bias can appear in machine learning algorithms
- Familiarize yourself with a landmark case study from statistics/data science.

Part 0: Introduction

This lab continues our discussion on the *ProPublica* report written about the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) algorithm. The main article can be accessed here.

As a reminder, the COMPAS Algorithm (developed by *Northpointe*, now called *Equivant*) is an algorithm designed to predict an individual's change of **recidivism** (committing a subsequent crime after having been arrested). Through careful analysis, *ProPublica* identified evidence of serious racial bias in the algorithm. Specifically, the authors found that the COMPAS algorithm predicted disproportionately high recidivism rates for black individuals as compared to their white counterparts; similarly, the algorithm predicted disproportionately low recidivism rates for white individuals as compared to their black counterparts.

In lecture, we didn't have time to go through much of the authors' analysis. I did mention, however, that the authors published a follow-up article detailing the statistical methodologies they used in their study. It is this follow-up report that will form the backbone of today's lab.

Part I: Understanding Decisions

> **! Action Item**
>
> Read through the report located at https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm **fully**, then answer the questions below.
>
> Note that there are "correct" answers, in the sense that answers to these questions can be found directly from the report.

Why do the authors claim they chose to analyze the COMPAS algorithm specifically?

Answers to the Question Above

Replace this line with your answers

Describe the COMPAS rating system. Specifically, what does the algorithm output? How does this tie in with the classification of individuals as having "Low", "Medium", or "High" risks of recidivism?

Answers to the Question Above

To quote the authors: 'We chose to examine the COMPAS algorithm because it is one of the most popular scores used nationwide and is increasingly being used in pretrial and sentencing, the so-called "front-end" of the criminal justice system.'

The authors point out that there are some difficulties in concretely defining "recidivism." Specifically, individuals may be arrested twice but for different crimes/reasons. How do the authors reconcile this? More broadly, how to the authors define recidivism? Be specific.

Answers to the Question Above

To quote the authors: 'To match COMPAS scores with accompanying cases, we considered cases with arrest dates or charge dates within 30 days of a COMPAS assessment being conducted. In some instances, we could not find any corresponding charges to COMPAS scores. We removed those cases from our analysis.'

The COMPAS algorithm produces a rating in each of three categories: "Risk of Recidivism", "Risk of Violent Recidivism", and "Risk of Failure to Appear". Which of these three categories was excluded from *ProPublica*'s analysis?

Answers to the Question Above

The authors chose not to analyze the "Risk of Failure to Appear" category.

In the section titled "How We Acquired the Data," the authors describe the process they used to "[match] criminal records to COMPAS records." This is likely the result of a join, like those we discussed on Lab 02. Which type (or types) of join do you think the authors used? What was the `join_by` key?

Answers to the Question Above

In this case, an **inner join** is likely the most appropriate choice - we only want to include individuals that have both a case number on file, and also COMPAS file. The authors indicate that they joined using "a person's first and last names and date of birth."

Part II: Analysis

In this section, we'll dive a little deeper into *ProPublica*'s analysis. Specifically, we are going to focus on the first **logistic regression** fit in the section titled "Analysis".

ProPublica has, very helpfully, provided public access to their data. I am including a `.csv` file in our `data/` subfolder, however I encourage you to read through the source GitHub page, at https://github.com/propublica/compas-analysis.

Action Item 1

Load in the data, and assign it to a data frame called `compas`.

```
## replace this line with your code
compas <- read.csv("data/compas-scores.csv")
```

Action Item 2

Based on the categorization of `age` in the dataset and the regression table provided in the report (titled "Risk of General Recidivism Logistic Model"), which `age_cat` level is being used as the baseline? (**Yes, there is a correct answer; yes, you can tell from the regression table!**)

```
## replace this line with your code
compas$age_cat %>% unique()
```

```
[1] "Greater than 45" "25 - 45"         "Less than 25"
```

Replace this line with your answers

There are three categories, but only two appear in the regression table. This means the missing category is the baseline level - that is, '25-45'.

! Action Item 3

Based on the categorization of `age` in the dataset and the regression table provided in the report (titled "Risk of General Recidivism Logistic Model"), which `age_cat` level is being used as the baseline? (**Yes, there is a correct answer; yes, you can tell from the regression table!**)

```
## replace this line with your code
```

The authors have also very helpfully provided *all of their source code* on the aforementioned GitHub repository. Below is a direct copy of their code (with some minor formatting changes; all rights remain with the authors):

```
df <- mutate(df, crime_factor = factor(c_charge_degree)) %>%
      mutate(age_factor = as.factor(age_cat)) %>%
      within(age_factor <- relevel(age_factor, ref = 1)) %>%
      mutate(race_factor = factor(race,
                                  labels = c("African-American",
                                             "Asian",
                                             "Caucasian",
                                             "Hispanic",
                                             "Native American",
                                             "Other"))) %>%
      within(race_factor <- relevel(race_factor, ref = 3)) %>%
      mutate(gender_factor = factor(sex, labels= c("Female","Male"))) %>%
      within(gender_factor <- relevel(gender_factor, ref = 2)) %>%
      mutate(score_factor = factor(v_score_text != "Low",
                                   labels = c("LowScore","HighScore")))
model <- glm(score_factor ~ gender_factor + age_factor + race_factor +
                            priors_count + crime_factor + two_year_recid,
             family="binomial", data=df)
summary(model)
```

One thing to note is that the object `df` has been assigned the full dataset (what we are calling `compas` in this lab).

! Action Item 4

Provide a line-by-line description of the code above. That is, using **words**, describe what each line of code is doing. The main moral here is that, from our accrued PSTAT 100 knowledge, we actually have enough know-how to understand every line!

1) Factorize `crime_factor`
2) Factorize `age_factor`
3) Relevel the baseline within the factored ages
4) Factorize `race`
5) Factorize `race`
6) Factorize `race`

7) Factorize `race`
8) Factorize `race`
9) Factorize `race`
10) Factorize `race`
11) Relevel the baseline within the factored races
12) Factorize gender
13) Relevel the baseline within gender
14) Factorize recidivism score
15) Factorize recidivism score
16) Fit a GLM of recidivism score on the other desired covariates
17) Fit a GLM of recidivism score on the other desired covariates
18) Fit a GLM of recidivism score on the other desired covariates
19) Produce a regression table

! Action Item 5

You may notice that each recidivism score is classified as one of *three* categories ("Low", "Medium", "High"), yet we fit a *logistic* regression (which assumes a binary output). This is because two levels were combined into a single level - which two are these, and why did the authors choose to combine these particular levels?

Replace this line with your answers

Medium and High scores were combined to simply read as 'high'. To quote the authors: 'We used those factors to model the odds of getting a higher COMPAS score. According to Northpointe's practitioners guide, COMPAS "scores in the medium and high range garner more interest from supervision agencies than low scores, as a low score would suggest thereis little risk of general recidivism," so we considered scores any higher than "low" to indicate a risk of recidivism.'

Submission Details

Congrats on finishing this PSTAT 100 lab! Please carry out the following steps:

i Submission Details

1) Check that all of your tables, plots, and code outputs are rendering correctly in your final `.pdf`.
2) Submit **ONLY** your `.pdf` to Gradescope. Make sure to **match ALL pages to the ONE question on Gradescope**; failure to do so will incur a penalty of 0.1 points.