# The Mark of a Criminal Record

## Background

In this section, we'll analyze the causal effects of a criminal record on the job prospects of white and black job applicants. This exercise is based on:

> Pager, Devah. (2003). "The Mark of a Criminal Record." *American Journal of Sociology* 108(5):937-975.[1]

The paper is an example of an "audit experiment", where researchers present two similar people that differ only according to a trait thought to be a source of discrimination. Pager hired a pair of white men and a pair of black men, and instructed each pair to apply for entry-level jobs in the city of Milwaukee, Wisconsin.

The men in each pair were matched on a number of dimensions, including physical appearance and self-presentation. The only difference between the two was that Pager randomly assigned which individual in the pair would indicate to potential employers that he had a criminal record.

To isolate the causal effect of a criminal record on job prospects for white and black respondents, she compared the difference in callback rates between applicants with and without a criminal background, and then calculated how those callback rates differed by race.

The data used for the study is available in `criminalrecord.csv`.[2] Each observation represents applicant in the experiment. The names and descriptions of variables are shown below.

| Name | Description |
|---|---|
| `jobid` | Job ID number |
| `callback` | 1 if the applicant received a callback, 0 if the applicant did not receive a callback; the experiment's **outcome variable**. |
| `black` | 1 if the applicant is black, 0 if the applicant is white. |
| `crimrec` | 1 if the applicant says he has a criminal record, 0 if the applicant says they do not have a criminal record; the experiment's **treatment variable**. |
| `interact` | 1 if applicant interacted with the employer during the job application process, 0 if applicant does not interact with the employer. |
| `city` | 1 if the job is located in the city center, 0 if the job is located in the suburbs. |
| `distance` | The job's average distance to downtown. |
| `custserv` | 1 if the job is in the costumer service sector, 0 if the job is not. |
| `manualskill` | 1 if the job requires manual skills, 0 if the job does not. |

[1] You are also welcome to watch Professor Pager discuss the design and result here.
[2] Four cases have been redacted. As a result, your findings may differ slightly from those in the paper.

## Question 1: Informative character values

First, read in the experiment's data from `data/criminalrecord.csv` into an object named `audit`, using `read_csv()`. Second, use `mutate()` and `if_else()` to replace the following variables with more informative character values, instead of their binary 0 and 1 values:

- `crimrec`: change the values of 1 and 0 to "Treated" and "Control", respectively;
- `black`: change the values of 1 and 0 to "Black" and "White", respectively.

Make sure you save these changes to the `audit` data frame object.

## Question 2: Callback rates by race

Later, we will compare the effect of a criminal record on white and black callback rates. For now, compare the overall callback rate between white and black applicants, regardless of treatment intervention.

First, use the `group_by()`, `summarize()` and `mean()` functions to create a variable `callback_rate` that reports the callback (`callback`) rate by the race of the applicant (`black`). Second, answer the following questions about these results:

- Is there a difference in callback rates between white and black applicants?
- If so, do you think this difference is small or large?

## Question 3: Average treatment effect

Let's now compare the callback rate between applicants in the treated and control groups. Because the treatment assignment is randomly assigned within applicant pairs, the difference in callback rates (i.e. "difference in means") between treated and control groups represents the *"average treatment effect"* (ATE).

First, calculate the callback rate among the treated and control groups, and assign these mean outcomes to objects named `treat_mean` and `treat_control`, respectively. This can be achieved using `filter()` to subset the data to the group of interest, and then using `summarize()` and `mean()` to calculate the mean of the outcome variable as a new variable, `callback_rate`. Second, create and print and object named `ate` which calculates the difference between the average treated and control group's callback rates.

Lastly, answer the following question: What does the resulting ATE estimate suggest about the impact a criminal record has on the prospect of receiving a callback for employment?

## Question 4: Average treatment effect, by race

Let's now compare the ATE among white and black applicants. First, similar to "Question 3", use `group_by()`, `summarize()` and `mean()` to create a variable `callback_rate` for each combination of treatment/control group (`crimrec`) and race (`black`). Second, "pivot" the preceding summarization using `pivot_wider()`, such that the names from the treatment variable `crimrec` appear as columns which contain values from the `callback_rate` variable. Third, create a variable named `ATE` which uses `mutate()` to calculate the difference between `Treated` and `Control` variables. Fourth, pipe all the previous commands into `knitr::kable()`, which rounds the cell proportions to 2 digits, in order to print a nice looking table in your .pdf output.

Lastly, answer the following questions:

- How does the average treatment effect differ by race?
- Is there anything else in the table that is worth mentioning?