The Mark of a Criminal Record Revisited

In one of the additional exercises for Chapter 2, we analyzed data from an important field experiment by Devah Pager about the the effect of race and criminal record on employment:

"The Mark of a Criminal Record". *American Journal of Sociology* 108(5):937-975. Look here to watch Professor Pager discuss the design and result.

This is a follow-up exercise using the same data set. Last time you encountered the paper, you described the different callback rates between groups. Now we are going to use what we've learned about statistical inference to better understand those patterns. You are welcome—and even encouraged—to reuse code from that exercise. In fact, in practice you often have to work with the same dataset many times, and writing good code the first time helps you reuse the code in future projects.

The dataset is called `criminalrecord.csv`. You may not need to use all of these variables for this activity. We've kept these unnecessary variables in the dataset because it is common to receive a dataset with much more information than you need.

Name	Description
`jobid`	Job ID number
`callback`	`1` if tester received a callback, `0` if the tester did not receive a callback.
`black`	`1` if the tester is black, `0` if the tester is white.
`crimrec`	`1` if the tester has a criminal record, `0` if the tester does not.
`interact`	`1` if tester interacted with employer during the job application, `0` if tester doesn't interact with employer.
`city`	`1` is job is located in the city center, `0` if job is located in the suburbs.
`distance`	Job's average distance to downtown.
`custserv`	`1` if job is in the costumer service sector, `0` if it is not.
`manualskill`	`1` if job requires manual skills, `0` if it does not.

The problem will give you practice with:

- re-using old code (optional)
- constructing confidence intervals
- difference-of-means tests
- p-values
- type I and type II errors

Question 1

Begin by loading the data into R and explore the data. How many cases are there in the data? Run `summary()` to get a sense of things. In how many cases is the tester black? In how many cases is he white?

Answer

```r
audit <- read.csv("data/criminalrecord.csv")

## (1) Number of observations
dim(audit)
```

```
## [1] 696   9
```

```r
## (2) quick summary
summary(audit)
```

```
##      jobid            callback          black          crimrec
##  Min.   :   1.00   Min.   :0.0000   Min.   :0.000   Min.   :0.0000
##  1st Qu.:  87.75   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.0000
##  Median :1024.50   Median :0.0000   Median :1.000   Median :0.0000
##  Mean   : 658.57   Mean   :0.1638   Mean   :0.569   Mean   :0.4986
##  3rd Qu.:1112.25   3rd Qu.:0.0000   3rd Qu.:1.000   3rd Qu.:1.0000
##  Max.   :1200.00   Max.   :1.0000   Max.   :1.000   Max.   :1.0000
##
##     interact           city           distance        custserv
##  Min.   :0.0000   Min.   :0.0000   Min.   : 0.00   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.: 8.00   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :12.00   Median :1.0000
##  Mean   :0.2428   Mean   :0.3919   Mean   :11.96   Mean   :0.6282
##  3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:16.00   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :25.00   Max.   :1.0000
##                   NA's   :2        NA's   :2       NA's   :2
##   manualskill
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.4813
##  3rd Qu.:1.0000
##  Max.   :1.0000
##  NA's   :2
```

```r
## (3) White and black
length(audit$jobid[audit$black == 1])
```

```
## [1] 396
```

```r
length(audit$jobid[audit$black == 0])
```

```
## [1] 300
```

There are 696 observations. There are 396 cases with black applicants and 300 cases with white applicants.

Question 2

Now we examine the central question of the study. Calculate the proportion of callbacks for white applicants with a criminal record, white applicants without a criminal record, black applicants with a criminal record, and black applicants without a criminal record.

Question 3

Now consider the callback rate for white applicants with a criminal record. Construct a 95% confidence interval around this estimate. Also, construct a 99% confidence interval around this estimate.

Question 4

Calculate the estimated effect of a criminal record for white applicants by comparing the callback rate in the treatment condition and the callback rate in the control condition. Create a 95% confidence interval around this estimate. Next, describe the estimate and confidence interval in a way that could be understood by a general audience.

Question 5

Assuming a null hypothesis that there is no difference in callback rates between white people with a criminal record and white people without a criminal record, what is the probability that we would observe a difference as large or larger than the one that we observed in a sample of this size?

Question 6

Imagine that we set up a hypothesis test where the null hypothesis is that there is no difference in callback rates between whites with and without a criminal record. In the context of this problem, what would it mean to commit a type I error? In the context of this problem, what would it mean to commit a type II error? If we set $\alpha = 0.05$ for a two-tailed test are we specifying the probability of type I error or type II error?