The Mark of a Criminal Record Revisited

The dataset is called `exam3.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.

```
exam3 <- read.csv("/Users/clayparham/Documents/Bush631-/Exam 3/exam3-1.csv")
```

The problem will give you practice with:

- 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

In the data, the number of cases is 696 cases. Black is a dummy variable, showing that the percentage of respondents that are black is 56.9% while the number of respondents that are white is 43.1%. That makes the number of black respondents 396 and the number of white respondents 300.

```
summary(exam3)
```

```
##      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
```

```
#number of respondents
NROW(exam3)
```

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

```
#number of black respondents
NROW(exam3$black[exam3$black == "1"])
```

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

```
#number of white respondents
NROW(exam3$black[exam3$black == "0"])
```

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

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.

Answer 2

We see in the data that the percentage of white with a criminal record who receive a callback (16%) is lower than white without a criminal record (34%), but still higher than both black with (5%) and without (14%) a criminal record. Black with a criminal record is very lower while without a criminal record is still lower than any white respondent.

```
white <- subset(exam3, black == "0")
black <- subset(exam3, black == "1")

#percentage of callback white with a criminal record
mean(white$callback[white$crimrec == "1"])
```

```
## [1] 0.1666667
```

```
#percentage of callback white without a criminal record
mean(white$callback[white$crimrec == "0"])
```

```
## [1] 0.34
```

```
#percentage of callback black with a criminal record
mean(black$callback[black$crimrec == "1"])
```

```
## [1] 0.05076142
```

```
#percentage of callback black without a criminal record
mean(black$callback[black$crimrec == "0"])
```

```
## [1] 0.1407035
```

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.

Answer 3

Using a t-test, we find that the confident interval for white with a criminal record at 95% confidence is .11 to .23. At a 99% confidence interval, that range increases to .09 to .25.

```
t.test(white$callback[white$crimrec == "1"])
```

```
##
##  One Sample t-test
##
## data:  white$callback[white$crimrec == "1"]
## t = 5.4589, df = 149, p-value = 1.956e-07
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.1063371 0.2269963
## sample estimates:
## mean of x
## 0.1666667
```

```
t.test(white$callback[white$crimrec == "1"], conf.level = .99)
```

```
##
##  One Sample t-test
##
## data:  white$callback[white$crimrec == "1"]
## t = 5.4589, df = 149, p-value = 1.956e-07
## alternative hypothesis: true mean is not equal to 0
## 99 percent confidence interval:
##  0.0870044 0.2463289
## sample estimates:
## mean of x
## 0.1666667
```

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.

Answer 4

We find that the estimated effect is around a 17% decrease in the amount of callbacks for white respondents with a criminal degree versus those who do not. Running a t-test, we find that we are 95% confident the true

impact is between 7% and 27% for those with criminal degrees. This is shown to be significant by the very low p-value (.0005)

```
mean(white$callback[white$crimrec == "0"]) - mean(white$callback[white$crimrec == "1"])
```

```
## [1] 0.1733333
```

```
t.test(white$callback[white$crimrec == "0"],white$callback[white$crimrec == "1"])
```

```
##
##  Welch Two Sample t-test
##
## data:  white$callback[white$crimrec == "0"] and white$callback[white$crimrec == "1"]
## t = 3.5103, df = 282.36, p-value = 0.0005207
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.07613775 0.27052892
## sample estimates:
## mean of x mean of y
## 0.3400000 0.1666667
```

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?

Answer 5

Looking at the data, the t-value is greater than zero, meaning that we can (probably), reject the null hypothesis. The distribution value is 282 which coupled with the t-value of 3.5, gives us the probability that the data could be just caused by random chance to be well within the allowed standard (since $.0005 < .05$). This means that is is very unlikely the the null hypothesis might be correct.

```
t.test(white$callback[white$crimrec == "0"],white$callback[white$crimrec == "1"])
```

```
##
##  Welch Two Sample t-test
##
## data:  white$callback[white$crimrec == "0"] and white$callback[white$crimrec == "1"]
## t = 3.5103, df = 282.36, p-value = 0.0005207
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.07613775 0.27052892
## sample estimates:
## mean of x mean of y
## 0.3400000 0.1666667
```

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?

Answer 6

In the context of this problem, a type I error is one where there is a "false positive". That is, the null hypothesis (where the data that having a criminal record does not impact your chances of a callback) is actually correct and what we see as a difference is actually just meaningless data.

A type II error is the opposite, where we get a "false negative", or where we falsely say that the null hypothesis is true while it is actually false. In the context of this problem, it would be saying that a criminal record does not impact the chances of a callback when it actually does.

Setting the alpha at .05 for a two-tailed test is testing whether the data could be found in the .025 of the left side of the graph or the .025 for the right side. This would be testing for a type I error.