Exercise - Racial Bias in the Labor Market

In this question you'll partially replicate a well-known paper on racial bias in the labor market: "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination" by Marianne Bertrand and Sendhil Mullainathan. The paper, which I'll refer to as BM for short, appears in Volume 94, Issue #4 of the *American Economic Review*. You will need to consult this paper to complete this problem.

For convenience, I've posted a copy of the dataset from this paper on my website at `https://ditraglia.com/data/lakisha_aer.csv`. Each row of the dataset corresponds to a single fictitious job applicant. After loading the `tidyverse` library, you can read the data into a tibble called `bm` using the `read_csv()` function as follows:

```
library(tidyverse)
bm <- read_csv('https://ditraglia.com/data/lakisha_aer.csv')
```

1. Read the introduction and conclusion of BM. Then write a short paragraph answering the following:
   a. What research question do BM try to answer?
   b. What data and methodology do they use to address the question?
   c. What do the authors consider to be their key findings?
2. Now that you have a rough idea of what the paper is about, it's time to examine the dataset `bm`. Carry out the following steps:
   a. Display the tibble `bm`. How many rows and columns does it have?
   b. Display only the columns `sex`, `race` and `firstname` of `bm`. What information do these columns contain? How are `sex` and `race` encoded?
   c. Add two new columns to `bm`: `female` should take the value `TRUE` if `sex` is female, and `black` should take value `TRUE` if `race` is black.
3. Read parts A-D of section II in BM. Then write a short paragraph answering the following:
   a. How did the experimenters create their bank of resumes for the experiment?
   b. The experimenters classified the resumes into two groups. What were they and how did they make the classification?
   c. How did the experimenters generate identities for their fictitious job applicants?
4. Randomized controlled trials are all about *balance*: when the treatment is randomly assigned, the characteristics of the treatment and control groups will be the same on average. To answer the following parts you'll need a few additional pieces of information. First, the variable `computerskills` takes on the value `1` if a given resume says that the applicant has computer skills. Second, the variables `education` and `yearsexp` indicate level of education and years experience, while `ofjobs` indicates the number of previous jobs listed on the resume.
   a. Is sex balanced across race? Use `dplyr` to answer this question. Hint: what happens if you apply the function `sum` to a vector of `TRUE` and `FALSE` values?
   b. Are computer skills balanced across race? Hint: the summary statistic you'll want to use is the *proportion* of individuals in each group with computer skills. If you have a vector of ones and zeros, there is a very easy way to compute this.
   c. Are `education` and `ofjobs` balanced across race?
   d. Compute the mean and standard deviation of `yearsexp` by race. Comment on your findings.
   e. Why do we care if `sex`, `education`, `ofjobs`, `computerskills`, and `yearsexp` are balanced across race?
   f. Is `computerskills` balanced across `sex`? What about `education`? What's going on here? Is it a problem? Hint: re-read section II C of the paper.
5. The outcome of interest in `bm` is `call` which takes on the value `1` if the corresponding resume elicts an email or telephone callback for an interview. Check your answers to the following against Table 1 of the paper:
   a. Calculate the average callback rate for all resumes in `bm`.
   b. Calculate the average callback rates separately for resumes with "white-sounding" and "black-sounding" names. What do your results suggest?
   c. Repeat part 2, but calculate the average rates for each combination of race and sex. What do your results suggest?
6. Read the help files for the `dplyr` function `pull()` and the base R function `t.test()`. Then test the null hypothesis that there is no difference in callback rates across black and white-sounding names against the two-sided alternative. Comment on your results.

Solutions

Solution to Part 2

```
library(tidyverse)
bm
```

```
# A tibble: 4,870 × 65
   id    ad    education ofjobs yearsexp honors volunteer military empholes
   <chr> <chr>     <dbl>  <dbl>    <dbl>  <dbl>     <dbl>    <dbl>    <dbl>
 1 b     1             4      2        6      0         0        0        1
 2 b     1             3      3        6      0         1        1        0
 3 b     1             4      1        6      0         0        0        0
 4 b     1             3      4        6      0         1        0        1
 5 b     1             3      3       22      0         0        0        0
 6 b     1             4      2        6      1         0        0        0
 7 b     1             4      2        5      0         1        0        0
 8 b     1             3      4       21      0         1        0        1
 9 b     1             4      3        3      0         0        0        0
10 b     1             4      2        6      0         1        0        0
# ℹ 4,860 more rows
# ℹ 56 more variables: occupspecific <dbl>, occupbroad <dbl>,
#   workinschool <dbl>, email <dbl>, computerskills <dbl>, specialskills <dbl>,
#   firstname <chr>, sex <chr>, race <chr>, h <dbl>, l <dbl>, call <dbl>,
#   city <chr>, kind <chr>, adid <dbl>, fracblack <dbl>, fracwhite <dbl>,
#   lmedhhinc <dbl>, fracdropout <dbl>, fraccolp <dbl>, linc <dbl>, col <dbl>,
#   expminreq <chr>, schoolreq <chr>, eoe <dbl>, parent_sales <dbl>, …
```

```
bm <- bm |>
  mutate(female = (sex == 'f'),
         black = (race == 'b'))
```

Solution to Part 4

a. Yes sex is balanced across race:

```r
bm |>
  group_by(black) |>
  summarize(n_female = sum(female))
```

```
# A tibble: 2 × 2
  black n_female
  <lgl>    <int>
1 FALSE     1860
2 TRUE      1886
```

b. Yes, computer skills are balanced across race:

```r
bm |>
  group_by(black) |>
  summarize(avg_computerskills = mean(computerskills))
```

```
# A tibble: 2 × 2
  black avg_computerskills
  <lgl>              <dbl>
1 FALSE              0.809
2 TRUE               0.832
```

c. Yes, both are balanced across race:

```r
bm |>
  group_by(black) |>
  summarize(avg_numjobs = mean(ofjobs), avg_educ = mean(education))
```

```
# A tibble: 2 × 3
  black avg_numjobs avg_educ
  <lgl>       <dbl>    <dbl>
1 FALSE        3.66     3.62
2 TRUE         3.66     3.62
```

d. The mean and standard deviation are about the same across race, as we'd expect given randomization.

```r
bm |>
  group_by(black) |>
  summarize(avg_exp = mean(yearsexp), sd_exp = sd(yearsexp))
```

```
# A tibble: 2 × 3
  black avg_exp sd_exp
  <lgl>   <dbl>  <dbl>
1 FALSE    7.86   5.08
2 TRUE     7.83   5.01
```

e. We care about balance because we want to know that the perception of *race* is responsible for any difference in callback rates, not some other factor.

f. These *aren't* balanced across sex. As the authors write "we use nearly exclusively female names for administrative and clerical jobs to increase callback rates."

```r
bm |>
  group_by(female) |>
  summarize(avg_computerskills = mean(computerskills),
            avg_educ = mean(education))
```

```
# A tibble: 2 × 3
  female avg_computerskills avg_educ
  <lgl>               <dbl>    <dbl>
1 FALSE               0.662     3.73
2 TRUE                0.868     3.58
```

Solution to Part 5

```r
# (a)
bm |>
  summarize(avg_callback = mean(call))
```

```
# A tibble: 1 × 1
  avg_callback
         <dbl>
1       0.0805
```

```r
# (b)
bm |>
  group_by(black) |>
  summarize(avg_callback = mean(call))
```

```
# A tibble: 2 × 2
  black avg_callback
  <lgl>        <dbl>
1 FALSE       0.0965
2 TRUE        0.0645
```

```r
# (c)
bm |>
  group_by(female, black) |>
  summarize(avg_callback = mean(call))
```

```
# A tibble: 4 × 3
# Groups:   female [2]
  female black avg_callback
  <lgl>  <lgl>        <dbl>
1 FALSE  FALSE       0.0887
2 FALSE  TRUE        0.0583
3 TRUE   FALSE       0.0989
4 TRUE   TRUE        0.0663
```

Solution to Part 6

```r
call_black <- bm |>
  filter(race == 'b') |>
```

```r
  pull(call)
call_white <- bm |>
  filter(race == 'w') |>
  pull(call)
t.test(call_black, call_white)
```

```
	Welch Two Sample t-test

data:  call_black and call_white
t = -4.1147, df = 4711.6, p-value = 3.943e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.04729503 -0.01677067
sample estimates:
 mean of x  mean of y 
0.06447639 0.09650924 
```