U6614: Assignment 2: Assessing gender wage gaps using the Current Population Survey

Your Name (your-uni)

2021-09-27

Please submit your knitted .pdf file along with the corresponding R markdown (.rmd) via Courseworks by 11:59pm on Monday, September 27th.

Before knitting your rmd file as a pdf, you will need to install TinyTex for Latex distribution by running the following code:

```
tinytex::install_tinytex()
```

Please visit this link for more information on TinyTex installation.

1 Load and inspect CPS data:

1a) Inspect the data frame and data types for each column

- make sure to inspect the age, sex, race, college columns

1b) Use the mutate function to create new column for sex

- sex.fac = as.factor(sex),
- check if it worked by calling the str() function

1c) Include sex.fac in a new data frame called cps.temp1

- also create factors for race and college education,
- use a pipe to exclude the columns for serial, ind
- after creating cps.temp1, print the first 5 observations

1d) Inspect race.fac, sex.fac, and college.fac using the levels() function

- what package is the levels() function located in?

1e) Use filter() to only include rows only for June 2020

- store as a new object cps_2020,
- print the first 5 observations,
- confirm your data only includes observations for 2020

1f) Remove the cps.temp1 object from memory using the rm() function

2 Describe the cps_2020 data frame

2a) What is the unit of observation?

2b) How many individuals are observed? from how many households?

2c) What is the average age of individuals in the sample? Youngest and oldest person?

3 Earnings per week for different groups in June 2020

3a) Find the observation for the top weekly earnings using the summarise() function

- assign this to a new object called max_earnings

3b) Find max weekly earnings using the arrange function instead of summarise

3c) Use the filter function to subset for the observation with max weekly earnings

- don't hardcode the max earnings to filter on, refer to the max_earnings object from a),
- store in new data frame cps_max_earn,
- confirm it worked

3d) What is the age, sex, and race of the top weekly earner in the sample?

3e) List the age, sex, and race of the top 10 weekly earners in the sample

3f) How many individuals earned more than $2000 in weekly earnings?

4 Wage gaps between males and females:

4a) Use the filter function to subset observations for males

- assign to new data frame, cps_2020_male,
- sort in descending order of weekly earnings
- check if it worked

4b) Repeat part a for females and create a new data frame, cps_2020_female

4c) Use summarise to find mean, min & max for males and females, separately

- name each statistic appropriately (i.e. name each column in the 1-row table of stats)
- what is the gender gap in mean weekly earnings?

4d) What is the wage gap in weekly earnings between white males and Black females?

4e) What is the wage gap between college educated white males and college educated Black females?

NOTE: the exercises above are done using weekly earnings, but can easily be converted to hourly wages