Homework 3: Models for Exploration

A. Short answer
B. Empirical exercise
Submission
Start Over

B. Empirical exercise

In this exercise, we will continue the gender pay gap analysis started Homework 1, using the same March 2009 CPS extract. This week we extended that analysis to career earnings comparisons. We'll pick things up there.

✓ Gender earnings gap review

First, let's review what we learned in class from the 23-62 year-olds in the March 2009 CPS.

Question 1:

Average earnings for men were $______ (round to the nearest integer), while average earnings for women were roughly $______ (round to the nearest 1,000) less.

64190, 19000

Correct!

Question 2:

This average dollar difference translates into roughly a _____ (round to the nearest integer) percent earnings gap.

43

Correct!

Question 3:

Based on Table 2, you would say male earnings increases are (more/less) ______ variable than female earnings.

more

Correct!

Question 4:

Based on Figures 11 and 12, you would say male earnings increase (more/less) ______ rapidly than female earnings early in a career.

more

Correct!

Question 5:

Based on Figure 12, in the first year of a career, male earnings increase _____ % on average while female increase by only _____ % (round to the nearest integer for both answers).

5, 3

Correct!

Continue

Percentage earnings gap

First, we'll replicate the estimated CEFs for women and men we showed in class, but using actual earnings instead of log earnings. Then, we'll evaluate the percentage gap in earnings between women and men.

Make sure that you click Submit Answer on each coding exercise where it appears.

As in Homework 1, we start by loading the data, refer to prior HW if you need to be reminded of the file name.

R Code Start Over Run Code Submit Answer

```r
library(readxl)
cps_mar <- read_xlsx("./data/cps09mar.xlsx")

```

Super job! Correct!

Picking up where we left off in class, we filter down to workers who are between 23-62 years old using `filter` and recreate the gender variable, using the `mutate()` and `case_when()` functions. We'll do it in one chunk.

R Code Start Over Run Code

```r
cps_mar_2362 <- cps_mar %>%
  filter(age >= 23,
         age <= 62) %>%
  mutate(gender = case_when(female == 1 ~ "Female",
                            female == 0 ~ "Male"))
```

Now we are ready to replicate the estimated CEFs for women and men using actual earnings. Remember, to talk in terms of career years, we "center" `age` on 23.

R Code Start Over Run Code

```r
cef_fvm <- cps_mar_2362 %>%
  mutate(age = age-23) %>%   # Center on age=23
  group_by(age, gender) %>%
  summarise (
    earnings = mean(earnings)
  )
```

```
`summarise()` has grouped output by 'age'. You can override us
ing the `.groups`
argument.
```

Now, plot the estimated CEFs just like in Figure 11, except the vertical axis should show actual dollar values.

R Code Start Over Run Code

```r
options(scipen=999)
ggplot(cef_fvm, aes(age, earnings, color=gender)) +
  geom_point() +
  geom_line() +
  ylab("Average earnings by age") +
  labs(title="CEFs of earnings by gender")
```

Question 6:

The CEF plots indicate that the gender earnings gap (grows/shrinks) _____ over a typical career.

grows

Correct!

Because our interest is in the earnings gap, it would probably be more informative just to examine the ratio of the two CEFs. So, let's do it. We'll use `filter` to separate the male and female CEF estimates. Then, we'll compute the `ratio`, put it in a new data frame with `age`, and list its values. This is a newer skill, so we will check your code here.

R Code Start Over Run Code

```r
males   <- filter(cef_fvm, gender == "Male")
females <- filter(cef_fvm, gender == "Female")
df_ratio <- data.frame(age = males$age, ratio = males$earnings/fem
df_ratio
```

age <dbl>	ratio <dbl>
0	1.185431
1	1.111521
2	1.168452
3	1.187692
4	1.139309
5	1.169614
6	1.282535
7	1.292464
8	1.195199
9	1.250615

1-10 of 40 rows Previous 1 2 3 4 Next

Question 7:

The ratio begins at _____, rises to _____ in year 23 (age 46), and then tapers off until the last year. (Round to two decimal places for both.)

1.19, 1.52

Correct!

Question 8:

A ratio of 1.69 implies an earnings gap of _____ %.

69

Correct!

Now let's plot the ratio.

R Code Start Over Run Code

```r
ggplot(df_ratio, aes(x = _____, y = _____)) +
  geom_line() +
  geom_point() +
  ylab("Male-female earnings ratio") +
  xlab("Age") +
  labs(title="Male-female earnings ratio by age")
```

Continue

✓ Gender earnings gaps by educational attainment

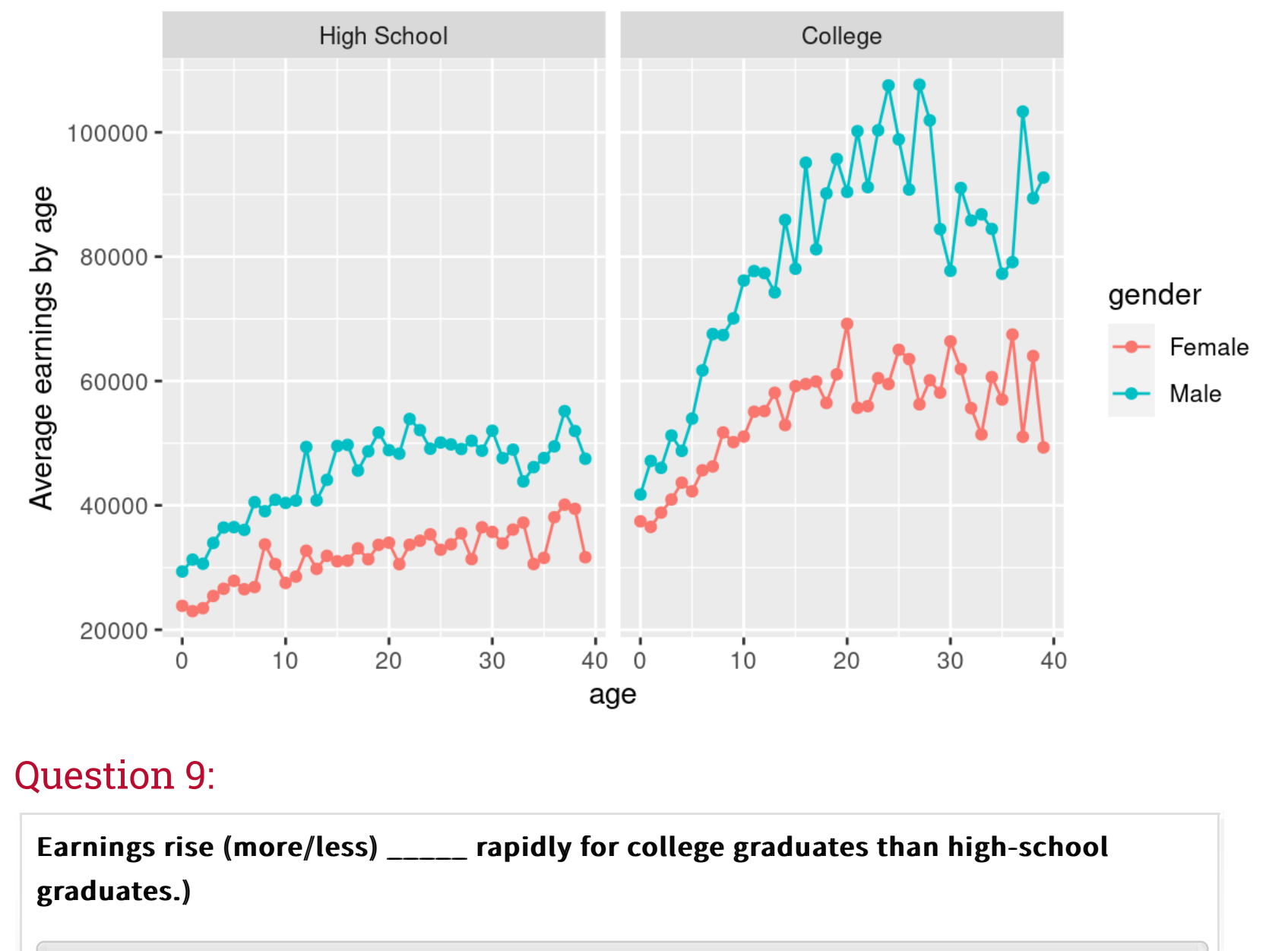
The size and pattern of the gender earnings gap begs for an explanation. Let's start down this path by comparing earnings gaps by educational attainment. We'll focus on high-school and college graduates and add those `education` categories to `group_by`. The plot code will use `facet_wrap` to arrange the high school and college plots side-by-side. Use the plot to answer a couple of questions to finish things out.

R Code Start Over Run Code

```r
cef_fvm_edu <- cps_mar_2362 %>%
  filter(education == 12 | education == 16) %>%
  mutate(age = age - 23) %>%
  group_by(age, gender, education) %>%
  summarise(
    earnings = mean(earnings)
  )

ggplot(cef_fvm_edu, aes(x = age, y = earnings, color = gender)) +
  geom_point() +
  geom_line() +
  facet_wrap(~ education,
             labeller = as_labeller(c(`12` = "High School",
                                      `16` = "College"))) +
  ylab("Average earnings by age") +
  labs(title = "CEFs of earnings by gender and education level")
```

```
`summarise()` has grouped output by 'age', 'gender'. You can o
verride using the
`.groups` argument.
```



Question 9:

Earnings rise (more/less) _____ rapidly for college graduates than high-school graduates.)

more

Correct!

Question 10:

By mid-career, the earnings gap among college graduates appears to (increase/decrease) ______ relative to the gap among high-school graduates.

increase

Correct!

Continue

Previous Topic Next Topic