Problem Set 5: Age-Earnings Profiles in Mexico and the US
Due 11:59 PM, Thursday, June 5

In this problem set, you will analyze data from two labor force surveys: the Current Population Survey (CPS) in the US and the National Survey of Employment (ENOE) in Mexico. The template `Rmd` script loads two datasets containing nationally representative samples of adults ages 20-64 collected in each year from 2010 to 2019. The datasets are in `rds` format, like in Problem Set 4. The dataset includes explanatory labels for the variables, so you can familiarize yourself with the variable definitions in the "Environment" window.

1. Using data from only 2019 in both datasets, compute average years of education at each age. Plot the average on the vertical axis against age on the horizontal axis. At which ages do you see patterns that would most plausibly be described as age effects? At which ages do you see patterns that would most plausibly be described as cohort effects? For the latter age range, redraw the graph with birth year rather than age on the horizontal axis. What can you say about cross-cohort trends in educational attainment in the US and Mexico?

2. Now focus on the ages you left out of the cohort figure. Using data from all years in both datasets, plot cohort age profiles of average educational attainment over this age range for birth years ending in 0 or 5. Do the age profiles have positive or negative slopes? Comparing two cohorts at a given age, does the later-born cohort have higher or lower average education than the earlier-born cohort? How do these patterns influence the cross-sectional plot for 2019 in Question 1?

3. For men and women in each country separately, draw cohort age profiles of the share employed for birth years ending in 0 or 5. Is there evidence of age effects? Cohort effects? At any given age, are men more or less likely to work than women?

4. For the rest of the problem set, restrict both samples to men who work as wage/salary workers for at least 30 hours a week (or 1560 hours a year) with positive earnings. Generate an hourly earnings variable in both datasets and call it `wage`. The table below reports the consumer price index over time in Mexico and the US.[1] Adjust the `wage` variable so it is denominated in 2010 currency units. In 2010, 10 Mexican pesos bought 1 US dollar's worth of goods and services.[2] Convert the `wage` variable in Mexico so it is denominated in dollars. What is each country's average wage in 2010 dollars?

Table 1: Consumer Price Index for Mexico and US

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Mexico	100	103	108	112	116	119	123	130	137	142
US	100	103	105	107	109	109	110	112	115	117

---

[1]E.g., in Mexico, a wage of 13 pesos per hour in 2017 would have been worth 10 pesos per hour in 2010.
[2]This is the "purchasing power parity" conversion factor, not the market exchange rate.

5. Draw the cross-sectional age profile of average wages for each country in 2019. Redraw the graph with a log scale.[3] Which version is more informative? Why?

6. In 2019, what was the ratio of the average US wage to the average Mexico wage at age 20? At age 50? Provide one age-based explanation and one cohort-based explanation for the larger ratio at age 50.

7. For each country, draw cohort age profiles of average wages for birth years ending in 0 or 5. Use a log scale. Do the cohort age profiles look similar to or different from the cross-sectional age profiles in Question 5? Do you see evidence of large cohort effects? Which of your explanations in Question 6 do you think is more likely to be correct?

8. Now zoom in on a single cohort, born in 1990. For this cohort only, compute the ratio of the average wage at each age to the average wage at age 20. Draw a graph that has this ratio on the vertical axis and age on the horizontal axis, with one plot for the US and one plot for Mexico. Do average wages rise more steeply over the life-cycle in the United States or Mexico?

9. Economists are inclined to interpret the results in Question 8 as returns to potential work experience ("potential" because we do not know whether a worker worked in every year). One critique of this interpretation is that workers of the same age can have very different amounts of potential work experience depending on when they left school. Another critique is that college graduates enter the sample after age 20, so the sample composition at age 29 may be different from the sample composition at age 20. To address these critiques, redraw the graph from Question 8 using only high school graduates. In the US, you should use workers with exactly 13 years of education (kindergarten plus 12 grades). In Mexico, you should use workers with exactly 12 years of education. Do your results suggest that the critiques are valid?

10. For high school graduates born in 1990, what is the proportional return to 10 years of potential work experience in the two countries? What do you think might drive the cross-country difference in experience returns? (The second question is open-ended, and a wide range of answers would be valid.)

---

[3]You can use `scale_y_log10()` in `ggplot()`.