

ECON 7020
Philip Shaw
Problem Set 1
Due date: Feb. 8, 2024

Problem 1. For this problem you should use the file `wagesub.csv` which contains data on logged wage, years of education, year of birth, and other controls.

- a. Describe the type of estimator we could use to estimate the population CDF for years of education. What condition would have to be true for us to know this estimator is a consistent estimator for a general population CDF?
- b. Calculate the mean values for logged wage and years of education. Calculate the variance and standard deviation of each series.
- c. Plot a histogram for years of education and logged wage across the observations in the data.
- d. Using the script file `cdffun.R`, estimate and plot the cdf across the support of values for both years of education and logged wage.¹²
- e. Given your estimates from part d., do either logged wage or years of education appear to be normally distributed? How do you know?
- d. Approximate the PDF $f(x) = \frac{dF(x)}{dx}$ with the following function:

$$\hat{f}(x) = \frac{F_n(x+h) - F_n(x-h)}{2h} \quad (1)$$

Using the equation above and $h = .01$, estimate $f(x)$ at the same points in the support of years of education as in part d. Plot this function.

¹Try using `edgrid=seq(min(Educ), max(Educ), by=1)`. Also consult the `cdfsims.R` code to compute the values of the CDF across this grid of values. You will need to do a for loop for this to work.

²Use the following command to plot functions in R `plot(x,y, type = "l", lty = 1)`.