MaCSS222, Spring 2025

Professor Bryan Graham

Problem Set 1

Due: February 10th, 2025

Problem sets are due at 11:59PM. The reader will provide instructions on how to turn in your problem set. You may work in groups, but each student should turn in their own write-up (including a "printout" of a narrated/commented and executed Jupyter Notebook if applicable). Please include a list of classmates you collaborated with when you turn in your problem set. Please also e-mail a copy of any Jupyter Notebook to the GSI (if applicable).

1 Law-of-Iterated Expectations and Analysis-of-Variance

The comma delimited text file **nlsy97ss.csv**, available on GitHub, will be used for this problem set.

The variables of interest for what follows are annual earnings (**earnings**), schooling (**yrssch**), race (**black**, **hispanic**) and sex (**female**). While not accurate, as a convenient shorthand, I will reference to non-Black, non-Hispanic respondents as 'white' (in practice this group is a mixture of respondents with European, Asian, Native American and other ancestries).

1. Let Y denote earnings and $X \in \{\text{black}, \text{hispanic}, \text{white}\}$. Explain why

$$\begin{aligned}\mathbb{E}[Y] = &\Pr(X = \text{black}) \cdot \mathbb{E}[Y \,|\, X = \text{black}] + \Pr(X = \text{hispanic}) \cdot \mathbb{E}[Y \,|\, X = \text{hispanic}] \\ &+ \Pr(X = \text{white}) \cdot \mathbb{E}[Y \,|\, X = \text{white}]\end{aligned} \tag{1}$$

2. Explain why

$$\begin{aligned}\mathbb{V}(Y) = &\Pr(X = \text{black}) \cdot \mathbb{V}(Y \,|\, X = \text{black}) + \Pr(X = \text{hispanic}) \cdot \mathbb{V}(Y \,|\, X = \text{hispanic}) \\ &+ \Pr(X = \text{white}) \cdot \mathbb{V}(Y \,|\, X = \text{white}) \\ &+ \Pr(X = \text{black}) \cdot (\mathbb{E}[Y \,|\, X = \text{black}] - \mathbb{E}[Y])^2 \\ &+ \Pr(X = \text{hispanic}) \cdot (\mathbb{E}[Y \,|\, X = \text{hispanic}] - \mathbb{E}[Y])^2 \\ &+ \Pr(X = \text{white}) \cdot (\mathbb{E}[Y \,|\, X = \text{white}] - \mathbb{E}[Y])^2,\end{aligned} \tag{2}$$

   and interpret this expression. What fraction of earnings variance is do to difference is average earnings across racial groups? What fraction reflects within-group inequality? By what percent would the variance of earnings fall if, all other things equal, average differences in earnings were eliminated across racial groups?

3. Form a dataframe consisting of all complete cases in the dataset (i.e., drop any records where one or more of our variables of interest are missing). How many records remain? Construct a table of basic summary statistics.

4. Compute the sample shares of black, hispanic, and non-black, non-hispanic respondents in your dataframe. Compute mean earnings within each group? Use these intermediate calculations to re-

cover the unconditional mean of earnings across all respondents. Compare your constructed mean with the simple sample average? Relate your analysis to (1) above.

5. Compute the conditional variance of earnings among black, hispanic, and non-black, non-hispanic respondents in your dataframe. Use your results, as well as those constructed above, to recover the unconditional variance of earnings across all respondents. Decompose earnings variance into its within-race and between-race components. Relate your analysis to (2) above.

6. Adapt your analysis thus far to decompose earnings variance into within-sex and between-sex components.

7. Discuss what you have learned about earnings inequality in the United States. See here for some basic background.

2 Experimental design

You have been hired by the Department of Labor to evaluate the effects of job training provided as part of the 2014 Workforce Innovation and Opportunity Act on earnings. Specifically to design a randomized control trial where one-half of participants are assigned to job training, whereas the balance are assigned to a control treatment (e.g., they are provided a pamphlet with information on writing a strong resume). The outcome of interest is annual earnings and the Department of Labor would like to able to detect an average treatment effect of \$1,000 with an ex ante probability of 0.90 with a test size of $\alpha = 0.05$.

1. Using the NLSY97 extra analyzed above, design an experiment meeting the specifications above. Justify your choices.

2. How would you modify your experiment to ensure a minimum detectable effect of \$1,000 for different subgroups separately (e.g., Black men, Hispanic women etc.). Design a studying that can reliable detect an effect of \$1,000 for each race-by-sex subgroup?

You may find the following references helpful: Duflo et al. (2007); Bloom (1995, 2005).

References

Bloom, H. S. (1995). Minimum detectable effects: a simple way to report the statistical power of experimental designs. *Evaluation Review*, 19(5), 547 – 556.

Bloom, H. S. (2005). *Learning More from Social Experiments: Evolving Analytic Approaches*, chapter Randomizing groups to evaluate place-based programs, (pp. 115 – 172). Russell Sage Foundation: New York.

Duflo, E., Glennerster, R., & Kremer, M. (2007). *Handbook of Development Economics*, volume 4, chapter Using randomization in development economics research: a toolkit, (pp. 3895 – 3962). North-Holland: Amsterdam.