

*MSc in Migration Studies
Research Methods
Hilary Term 2015
Exercises week 3*

Exercises

Note: for the exercises that require the use of Stata, please paste the commands you used in this document (preferably from your do-file with comments) as part of your answer. You can either paste the commands under each exercise that requires Stata, or paste all codes at the end of the document.

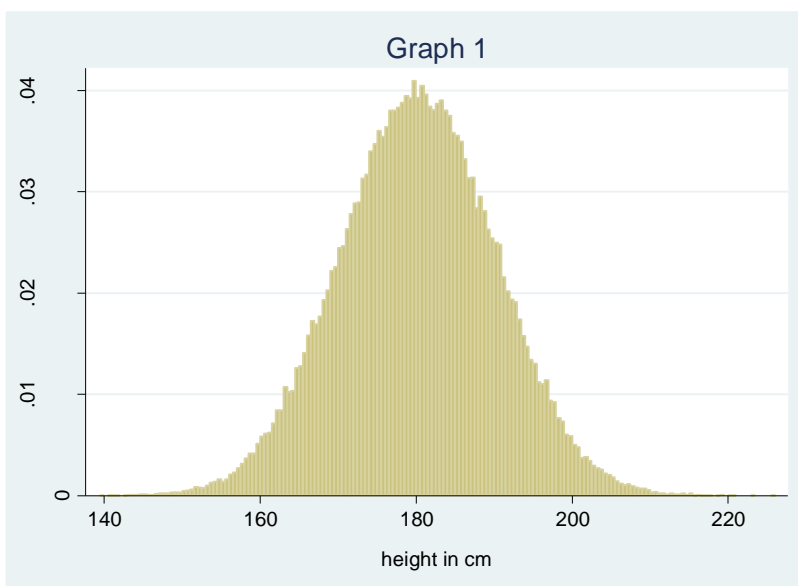
1. A girls' school in the UK needs to buy new chairs for its year-1 students (aged 6). A teacher has found a nice-looking set of chairs in an online store. They are meant for young children and the producer only guarantees the stability of the chair for users weighing less than 25kg. The weight of 6-year old girls is normally distributed, with a mean of 20 kg and a standard deviation of 3.4kg.
 - A. Use the 5 steps discussed in class to determine the probability that a 6-year old girl weighs 25 kg or more.
 - B. If you were the head of school, would you buy these chairs or encourage the teacher to look for other options? Why?

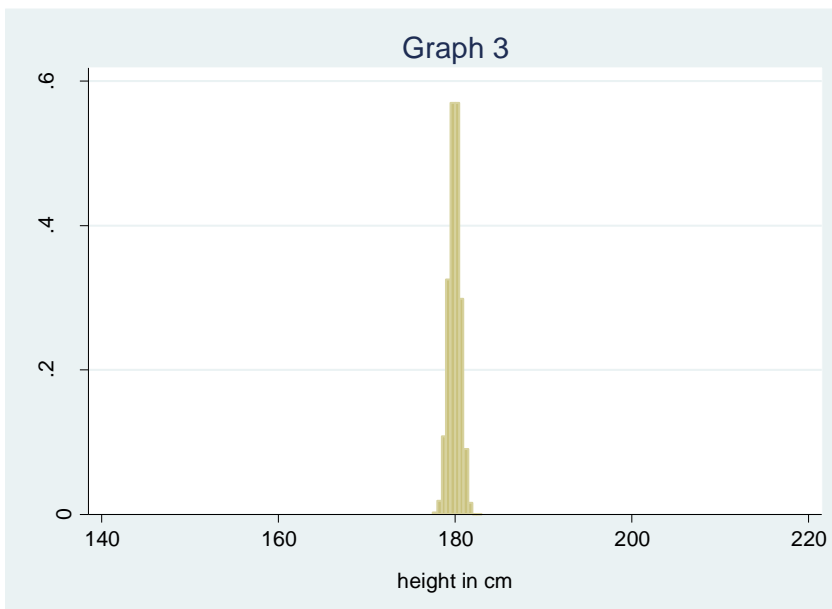
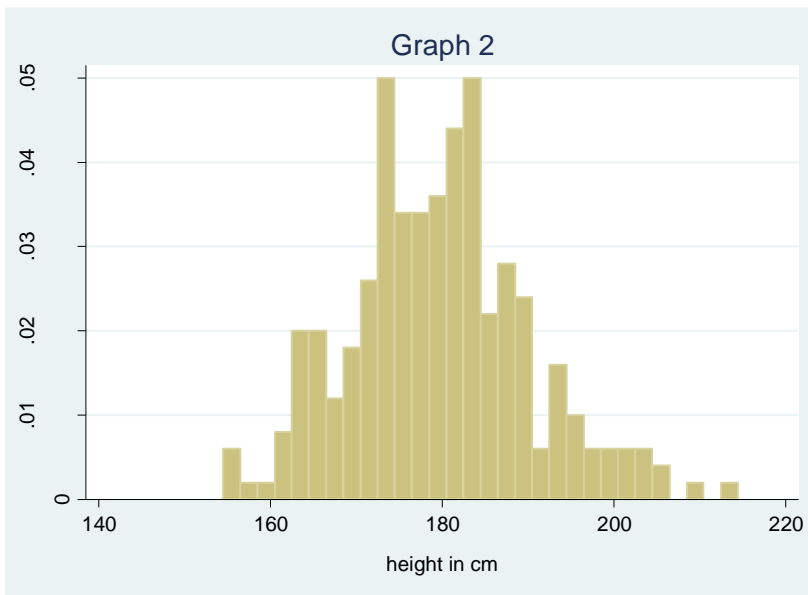
2. Family services is monitoring the weight of young children. They will intervene in the case of obese or underweight children. They define this as children who belong to the 2.5% heaviest children or the 2.5% lightest children. The weight of 6-year old girls is normally distributed with a mean of 20 kg and a standard deviation of 3.4kg.
 - A. Let's first look at the obese girls. What is the weight above which family services will intervene?
 - B. Now let's look at underweight girls. What is the weight under which family services will intervene?

3. You would like to do a research project on discrimination in the justice system. You would like to know if people perceive the justice system as colour-blind or biased. You decide to explore your research question with the ESS subset because of its large sample and good reputation for data quality. Looking at the codebook, you decide that question D30 best approaches what you are interested in.
 - A. In Stata, make a frequency table of variable D30. What are your preliminary findings?
 - B. You wonder whether people who consider themselves part of an ethnic minority have a different opinion about discrimination in the justice system than people who belong to the dominant group. You decide to use variable C32 as a proxy for belonging to an ethnic minority. First inspect variable C32 by running a frequency table and give a short description of the composition of the ESS sample in terms of ethnic minority status.
 - C. Recode the missing values in variables C32 and D30 to variables that Stata recognises as missing values (as we did in the Stata workshop. See the Stata guide for examples).

- D. Then make a crosstab (a two-way table) of variables C32 and D30 and request Stata to display percentages in a way that allows you to determine whether people who consider themselves part of an ethnic minority have a different opinion about discrimination in the justice system than people who belong to the dominant group.
- E. What do the data tell you about the differences in perceptions of the justice system between people who do and do not belong to an ethnic minority?
- F. For the week 1 exercises you considered the sampling bias in ESS for immigrants. Might this bias have affected the results of the relation between ethnic minority status and perception of discrimination in the justice system? If so, in what way?
- G. Think of another factor that might influence people's perception of discrimination in the justice system. Look for a variable in the ESS subset that measures this. (HINT: You can browse through the codebook of the subset that is posted on WebLearn.) Once you have found an appropriate variable in ESS, inspect that variable in a frequency tab to see if there are no strange values, and then make a cross-tab between that variable and D30. What does this tell you?

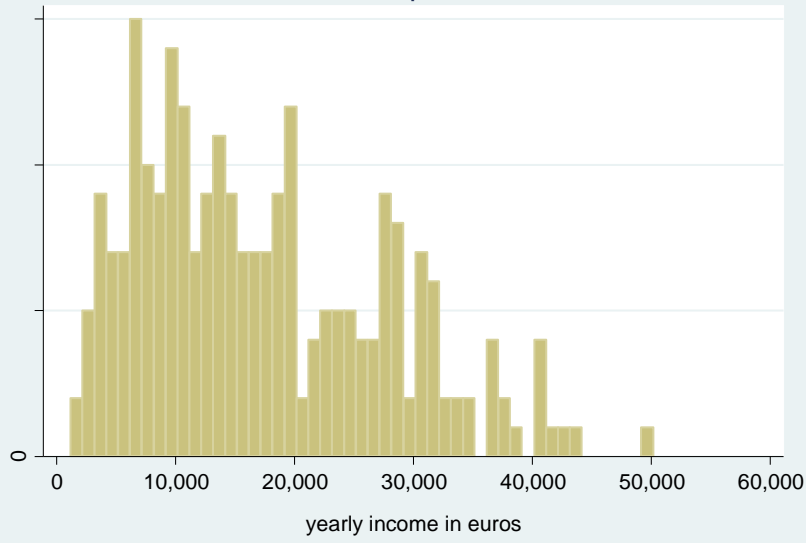
4. Below are graphs of three distributions of height of male conscripts for the national army in 2010. Height is normally distributed. Which of the graphs represents:
- A. The population distribution: i.e. the distribution of the height of all 100,000 conscripts
 - B. The sampling distribution of the mean height of a series of random samples of 250 conscripts
 - C. A sample data distribution of height in a random sample of 250 conscripts
- For each explain your reasoning; why do you think that graph represents that distribution



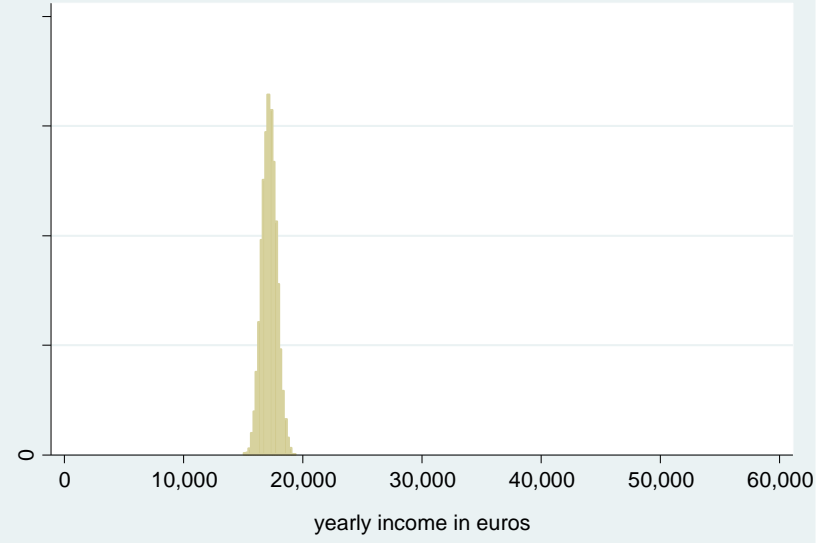


5. Below are graphs of three distributions of yearly income (in Euros) of a cohort of university graduates. There are 100,000 members in the cohort. Which of the graphs represents
- The population distribution: i.e. the income of all 100,000 graduates
 - The sampling distribution of the mean income of a series of random samples of 250 graduates
 - A sample data distribution of income in a random sample of 250 graduates
- For each explain your reasoning; why do you think that graph represents that distribution

Graph 1



Graph 2



Graph 3

