Course: Data Analysis (task № 1)

Student’s Name and Surname Pavel Drankov

Please answer the questions below.

**Tasks to do individually**

1. What should be written in the command line to create an “age” variable with represents the age of a person in years?

generate int age = 1

2. What should be written in the command line to rename the variable “q1” to “gender”?

rename q1 gender

3. What should be written in the command line to give to the “age” variable the label “Age in years”?

label variable age "Age in years"

4. Open **nlsw88.dta** file (from Example Datasets)

sysuse nlsw88.dta

a. How many unique value labels the file contains?

17

b. Analyse **occupation** variable and answer the questions below:

- How many Managers/admins participated in the research?:264

* What in the percent of Professional/technical out of all the research participants?: 14.11

- What in the percent of Professional/technical out of the research participants that gave valid information about their occupation?: 14.17

c. Indicate the following statistical characteristics for **age** variable:

- Mode 35\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_;

- Median 39\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_;

- Mean 39.15316 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_;

- Range 34-45\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_;

- Standard deviation 3.060002\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_;

- S. E. mean .0645679 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_;

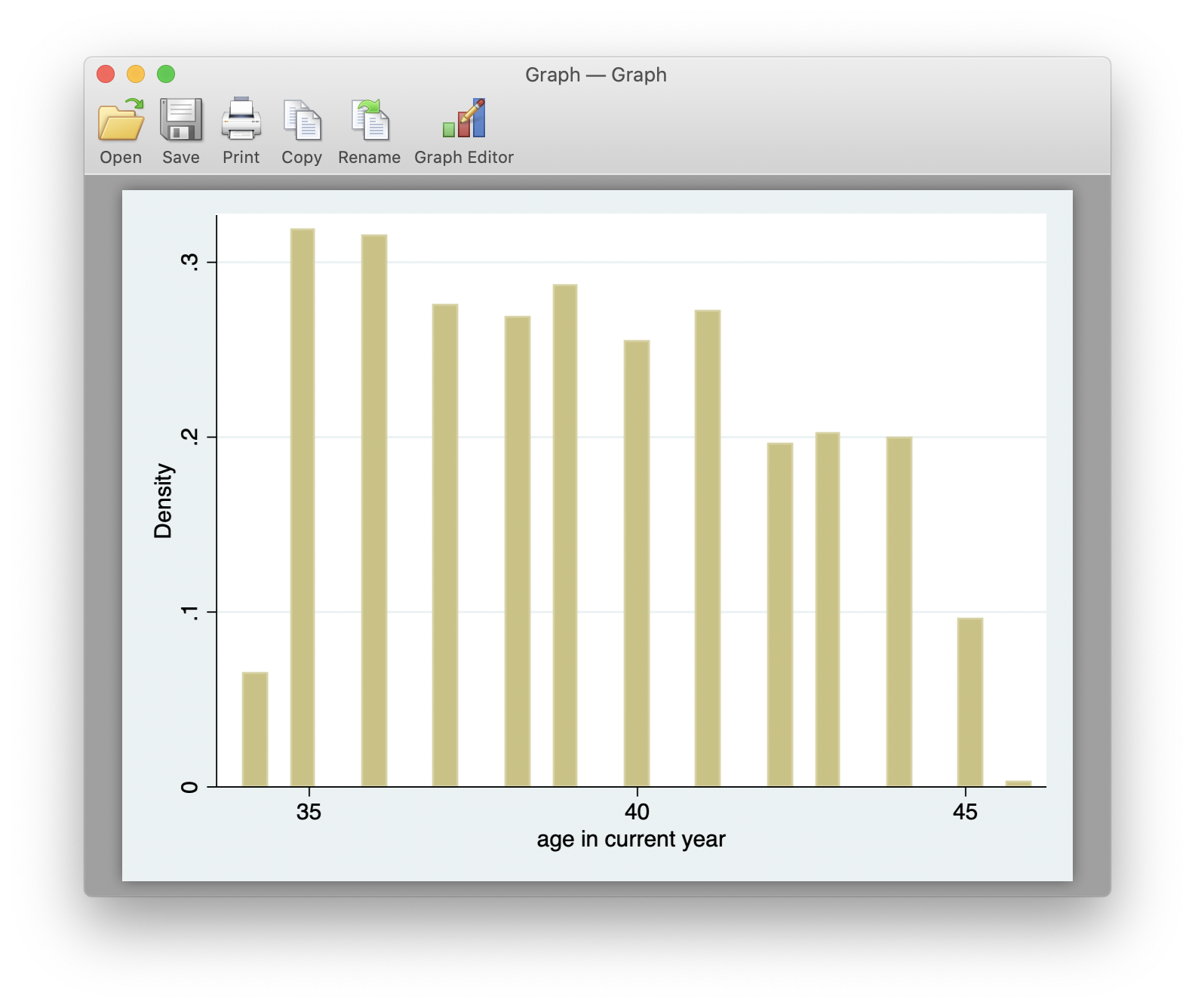
- Interquartile range 6\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_;

- Quartile deviation 3\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_;

- Decile ratio p90/p10 = 44/35 = 1.257 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

Evaluate the symmetry and pointyness of distribution of **age** variable. Indicate whether the distribution is positively or negatively skewed and what does it mean in terms of the shape of the distribution. Indicate whether the distribution is leptokurtic or platykurtic and what does it mean in terms of the shape of the distribution.

Crate a histogramfor **age** variable with normal density plot and copy it into this file.

****

d. Indicate the mean age for those who are married and not married. 1442 and 804

mean age if married == 1

e. Indicate the maximum age of those who are not married and have the tenure variable greater than 50: NO observations

f. How many college graduates are married, what is the percent of college graduates that are married? 344 and 344/532 = 0.6466...

g. How many research participants have the following characteristics:

- are not married: 804

- are not college graduates: 1714

- are not members of the union: 1630

What is the mean age and the standard deviation of age for this group of people? 39.13865