Course: Data Analysis (task № 1)

Student’s Name and Surname Iurii Aksenov

Please answer the questions below.

**Tasks to do together**

1. Open **data\_games.dta** file. Do the frequency analysis of **payment\_method** variable and answer the questions below:

- How many payments were done using a giftcard? 58

- What is the percent of payments, which were done using fb\_promotion? 0.47

- What is the percent of payments, which were done using a giftcard? 0.05

2. Open **data\_games.dta** file.

Indicate the following statistical characteristics for **crystalls\_bought** variable:

- Mode 14;

- Median 30;

- Mean 51.36338;

- Range 1695;

- Standard deviation 75.6069;

- S. E. mean .2327668

- Interquartile range 56;

- Quartile deviation (iqr /2 ) = 56/2 = 28;

- Decile ratio (p90 / p10) = 120/5.

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

****

3. Open **data\_games.sta** file.

Evaluate the symmetry and pointyness of distribution of **payment** 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.

**Answer**: Skewness 3.115122, Kurtosis 17.1964

Symmetry: Positively skewed distribution because of positive value (3.115122). It means that the frequent scores are clustered at the lower end and the tail points towards the higher or more positive scores

Kurtosis: Leptokurtic distribution because of positive value of kurtosis is >3 (17.1964). It means that the distribution is relatively thin in the tails and so look quite pointy.

**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?

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

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

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

a. How many unique value labels the file contains?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

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

- How many Managers/admins participated in the research? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

- What in the percent of Professional/technical out of all the research participants? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

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

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

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

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

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

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

- Range \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_;

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

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

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

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

- Decile ratio \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

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.

e. Indicate the maximum age of those who are not married and have the tenure variable greater than 50.

f. How many college graduates are marries, what is the percent of college graduates that are marries?

g. How many research participants have the following characteristics:

- are not marries,

- are not college graduates,

- are not members of the inion.

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

Please send this MS Word file with answers together with the do-file to amelikyan@hse.ru from your **personal** e-mail.