

## Seminar no. 2

In the dataset *Household\_marriage.RData*, there are 1346 respondent answers on two questions:

question 1: „How important is that the youth do not live in the same household as their parents?“

Question 2: „How important is to get married?“

The answers were on 5 value scale: Very important(1) - Quite important(2) - To some extent important(3) - Not so important(4) - Absolutely not important(5).

Using suitable descriptive statistics characterize both variables and compare the distribution of the two questions.

(General Social Survey, USA, 2002)

### Příklad 1

For the *Household\_marriage.RData* dataset depict the plot of the *ownhh* and *getmar* variables. Since the repeated values of the pairs are located the same, slightly misrepresent their location.

### Příklad 2

In the *Household\_marriage.RData* dataset state and compare the associations between the variables "opinion on own household importance" and "opinion on marriage importance".

By the Spearman correlation coefficient test the independence of the two variables. Regarding the value of  $r_S$ , evaluate the significance and practical application relevance of the test. Compare the  $p$  value of this test with the  $p$  value of the  $\chi^2$  test of independence. What is the change in the  $p$  values related to the data ordinality?

### Příklad 3

Do the opinions about own household and marriage importance differ? Discover in the *Household\_marriage.RData*.

### Příklad 4

In the dataset *Criminality.RData*, there are particular states of USA with values of 7 variables: ratio of violators on 100k inhabitants, ratio of murders,..., percentage of households under the poverty line, percentage of incomplete families. Characterize graphically the multidimensional data a) using pairs of variables, b) using charts suitable for multidimensional visualization.

**Homework:** The same tasks as in the problem 4 for the *Countries.RData* dataset.