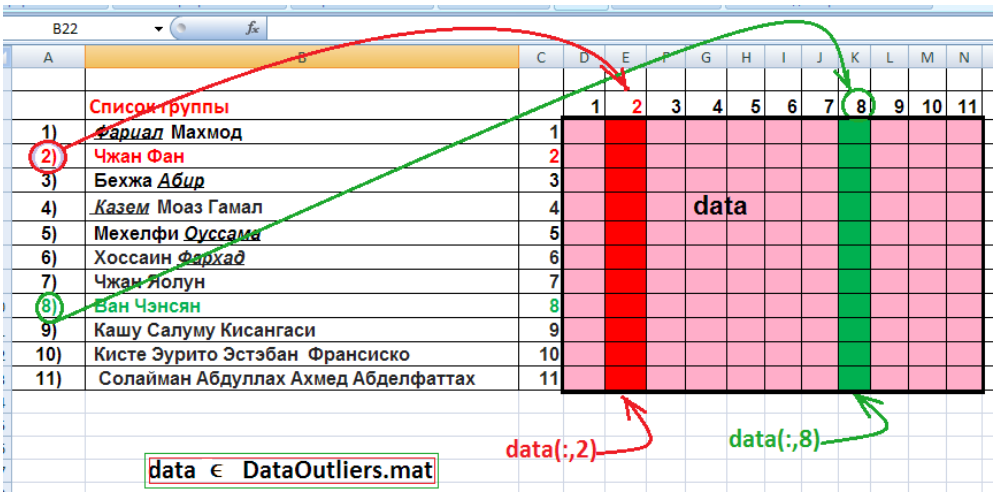
Task 4. Laboratory Work «Preliminary Analysis»

Step 1. Download the data

File DataOutliers.mat saved in Teams in the LaboratoryWork "Preliminary Analysis" folder. This file contains matrix of (100×12) size; each column – is the sample with numeric elements. Each student receives sample as a column with number coinciding with ordinal number of student in the list (see explaining table1).

Table.1.



You need to define nature of data, using statistical graphical functions:

Step 2. Data Visualization (**MatLab functions name are written by Italics**)!

Use function *boxplot* for data explanation and add values of Q_1,Q_2,Q_3 to the graph. Explain if exists data skewness, use different approaches to prove your statements.

MatLab function *boxplot* will show if there are symmetry or skewness, will help determine quartiles Q_1,Q_2,Q_3 or the same – percentiles (*prctile*), interquartiles (*iqr*), data range, and the most important – presence of outliers.

Step 3. Data Preprocessing

Reoder data and exclude outliers!

Step 4. The study of an edited experimental sample.

Control the result, using *boxplot*. Determine descriptive statistics, they are mean, median, variance, standard deviation, skewness and kurtosis.

MatLab functions: *mean, median, variance, std, skewness, kurtosis,* in addition easy to use *moment*. Draw conclusion!

Step 5. Empirical probability density function.

Using *histogram* function plot on the same axes the empirical probability density function and theoretical normal density function, depends on parameters: average and standard deviation, calculated by the sample (see ExamplePreliminaryAnalysis.mlx).

Calculate the critical point C on the axes Ox such that P(x>C)= 0.05 and fill the area under the probability density function on the graph for $x \in [C, \infty,]$ with color.

Remark! *For histogram should to choose appropriate options.*

Step 6. Empirical probability distribution function.

Using the *ecdf* function, construct an empirical probability distribution function and *normcdf* – a theoretical normal probability distribution function, depending on the parameters (the mean and the standard deviation calculated from the sample) they should be located on the same axes.