

# Introduction to the Practice of Statistics

## List 2

### Laboratory

1. Referring to the data set `grades.txt` decide if the distribution of the results of IQ tests is normal:
  - (a) by computing the number of observations in intervals  $[\bar{x} - s, \bar{x} + s]$ ,  $[\bar{x} - 2s, \bar{x} + 2s]$ ,  $[\bar{x} - 3s, \bar{x} + 3s]$ , where  $\bar{x}$  is the mean and  $s$  is the standard deviation;
  - (b) by using QQ-plots.
2. Referring to the data set `income.txt` decide (by using QQ-plots) if the distribution of incomes is normal. If not, check if for some nice transformation (e.g.  $\sqrt{X}$ ,  $\ln X$ ) we obtain a normal distribution.
3. Repeat ten times the following experiment using the data set `income.txt`:
  - (a) Draw 1000-element sample from the data set. Draw trajectories of the following statistics: mean, median, standard deviation, IQR and the fraction of people with at least bachelor's degree, computed for the first  $t$  elements of the sample ( $t \in [1, 1000]$ ). Compare with the values of those statistics for the whole data set.
  - (b) Draw ten 100-element groups from the data set, using stratified sampling with respect to the job class (i.e. the number of people drawn from each sector is equal to the percentage of people from that sector in the whole data set). Then draw trajectories of the appropriate statistics, gradually adding the groups. Compare with the values of those statistics for the whole data set.
4. Simulate ten times 1000 coin tosses. For each of the ten experiments draw trajectories of the fraction of heads in the first  $k$  tosses ( $k \in [1, 1000]$ ).