

# Introduction to the Practice of Statistics

## List 1

### Laboratory

1. Data set `grades.txt` contains data on 78 seventh-grade students. The data include each student's grade point average (GPA) (where numerical values applied to grades: F, D–, D, D+, C–, C, C+, B–, B, B+, A–, A, are 0 to 11), score on a standard IQ test, gender and a score on the Piers-Harris Children's Self-Concept Scale (a psychological test).
  - (a) Draw histograms of all quantitative variables. Describe the shape, center and spread of their distributions. Give values of corresponding statistics (minimum, maximum, median, quartiles, mean, standard deviation, variance, coefficient of variation).
  - (b) Repeat point (a) separately for boys and girls. Compare the results.
2. For one chosen quantitative variable from the data set `grades.txt` draw histograms for a few different choices of the number of classes (including one for a very small number of classes and one for a very large number of classes). Describe the effect of too few classes and the effect of too many classes.
3. Data set `income.dat` contains data on 55 899 people, from a survey of the Bureau of Labor Statistics, USA. The data set contains age (in years), education (1 = did not reach high school, 2 = some high school, without diploma, 3 = high school diploma, 4 = some college, no bachelor's degree, 5 = bachelor's degree, 6 = postgraduate degree), gender (1 = male, 2 = female), income (in US dollars) and job class (5 = private sector, 6 = public sector, 7 = self-employed).
  - (a) Draw histograms of incomes. Describe the shape, center and spread of their distributions. Give values of corresponding statistics (minimum, maximum, median, quartiles, mean, standard deviation, variance, coefficient of variation).
  - (b) Repeat point (a) separately for men and women. Compare the results.
  - (c) Identify outliers at point (a).