Tasks for exam "Applied Statistics in R"

1. Descriptive statistics.

   (a) For a give sample calculate summary statistics (numerical characteristics including mean, variance, standard deviation, median, 0.25-quantile, 0.5-quantile, 0.75-quantile, mode, minimal value, maximal value, skewness, asymmetry).

   (b) Plot the graph of cumulative distribution function.

   (c) Plot three histograms with different number of intervals.

   (d) Plot "box plot" and give interpretation of the result.

2. Statistical goodness-of-fit tests. For a given sample test a hypothesis about normality with confidence level 0.9 using following tests (some of them may need to use "nortest" package):

   (a) Kolmogorov-Smirnov test,

   (b) Chi-squared test,

   (c) Cramer-von Mises test,

   (d) Anderson-Darling test.

   Compare results and give the interpretation.

3. Comparing two or more samples.

   For Samples 1 and 2 test the following hypotheses (with conf. level 0.95):

   (a) Test if two means are equal using Student's t-test (note it is needed to verify that the distribution of both samples is normal).

   (b) Test if two variances are equal using Fisher's F-test (note it is needed to verify that the distribution of both samples is normal).

   (c) Test if the distributions of two random variables are equal using Kolmogorov-Smirnov test.

   (d) Test if the distributions of two random variables are equal using Wilcoxon test.

   Compare results and give the interpretation.

4. Regression analysis.

   (a) For Data 1 construct a model of linear and ridge (or quantile) regression ($Y \sim X_1 + \ldots + X_m$). Test if any parameter of the model and the model in general are significant and give the interpretation of the results.

   (b) For Data 2 construct a model of binary regression (logit and probit). Test if any parameter of the model is significant and give interpretation of the results. Test the fitness of the model using Wald's statistics. Give interpretation of the results.

5. Cluster analysis.

   (a) Find an appropriate number of clusters with Sample "Task_5" using k-means. Give interpretation of the results.

   (b) Find an appropriate number of clusters with Sample "Task_5" using hierarchical clusterization method. Give interpretation of the results.