**4ST630 Homework 2**

**(sampling from the multivariate distributions and some nonparametric statistics)**

**Evaluation: 0 – 10 points**

**Deadline: see InSIS coursework submission**

Use data in diabetes.csv

* Source: <https://www.kaggle.com/uciml/pima-indians-diabetes-database>
* Original file contains the data about women older 21 years coming from the North-American Indian tribe Pima, known historically for its predisposition to “diseases of civilization”, especially type II diabetes.
* It contains 768 observations of 9 variables:
  + Number of pregnancies, blood glucose, blood pressure, skin thickness, blood insulin, BMI, hereditary score and age.
  + Plus variable “outcome” – whether a woman has been diagnosed with diabetes (1) or not(0).
* First delete all the observations that contain at least one 0 (which means NA in the used coding) in its variables except for the variable Outcome (in which 0 means “without diagnosed diabetes”).
* Find out:
  + Vector of means
  + Wishart matrix
  + Sample covariance matrix
  + Sample correlation matrix
  + Eigenvalues of sample correlation matrix. Are they all positive?
* Suppose the sample comes from multivariate normal distribution with the vector of expected values:
  + Find out a value of quadratic form Q, its distribution and the probability, that it will reach the value or higher value.

Choose three of the variables (but not the variable Outcome) and:

* Find out medians, trimeans, trimmed means, windsorized means (use 5 % observations to be trimmed).
* Find out IQRs, MADs and Gini indices.
* Use sample means and sample standard deviations as estimates of μ and σ of normal distributions.
* Draw a sample quantile functions, empirical distribution functions and kernel density estimates with histograms and compare them (graphically) to the theoretical functions of estimated normal distributions (from previous point).
* Use the Kolmogorov-Smirnov and χ2 GOF test in order to find out whether the distributions are drawn from the estimated normal distributions.