VU university	Statistical Data Analysis, part I
Faculty of Sciences	25 March 2013

Use of a basic calculator is allowed. Graphical calculators and mobile phones are not allowed. This exam consists of 4 questions (27 points).

Please write all answers in English. Grade = $\frac{total+3}{3}$.

GOOD LUCK!

Question 1 [6 points]
Are the following statements correct/sensible? Motivate your answer by a short argument.

a. [2 points] The Shapiro-Wilk test is for testing a composite null hypothesis.

b. [2 points] M-estimators with bounded ψ-function are robust.

c. [2 points] A two sample QQ-plot is the same as a scatter plot for paired data.

Question 2 [6 points]
Let $X_1, \ldots, X_{100}$ be independent and identically distributed random variables with unknown distribution P. Suppose we want to test $H_0 : P = P_0$, for a known P_0, using the χ^2 goodness-of-fit test.

a. [3 points] The test statistic

$$X^2 = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i}$$

has approximately a χ^2_{k-1}-distribution under H_0. Describe the rule of thumb that needs to be satisfied for this approximation to be reliable.

b. [3 points] Suppose we are given intervals $I_1, \ldots, I_k$ that do not fulfill the rule of thumb for the given sample size. In such a situation we can still use X^2 as test statistic. However, we cannot rely on its approximate χ^2-distribution. Therefore, we use a bootstrap test. Describe the steps that are made in a bootstrap test for the given null hypothesis using X^2 as test statistic.

Question 3 [7 points]
In Figure 1 a histogram, boxplot and several QQ-plots of a data set x are presented.

a. [2 point] Which of the four location scale families do you think is most appropriate for these data? Explain your answer.

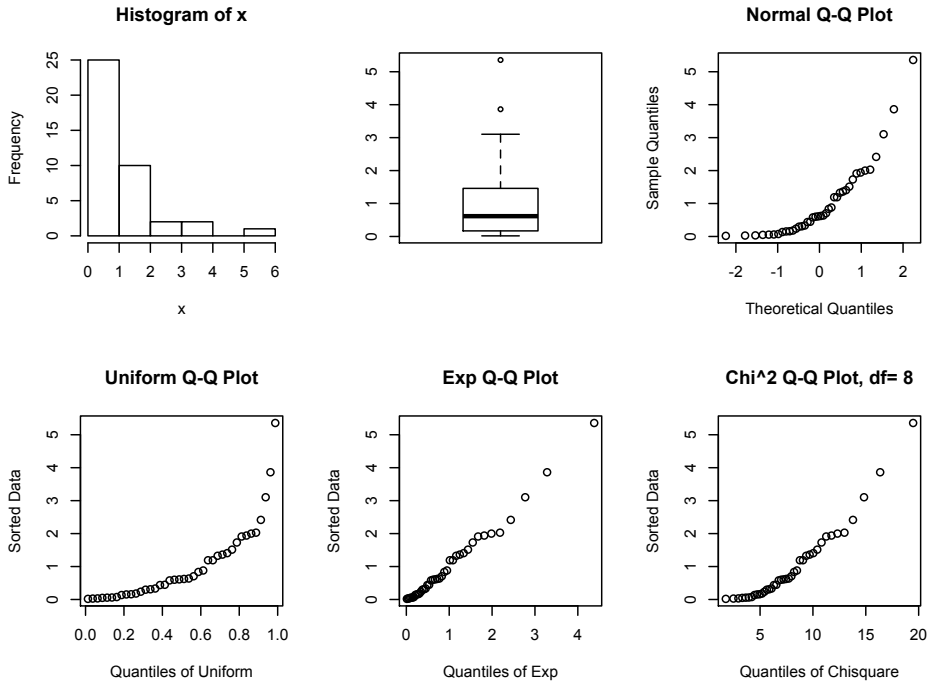

Figure 1: Histogram, boxplot and QQ-plots against the $N(0,1)$, uniform[0,1], exponential(1) and the standard χ^2_8 distributions of a data set.

b. [2 points] The $\alpha-$trimmed mean of these data was computed for $\alpha = 0, 0.1, 0.2, 0.3, 0.4, 0.5$. The values of these 6 trimmed means are, in arbitrary order: $0.69, 0.62, 0.81, 0.63, 0.74, 1.02$. Which of these values is the 0.1 trimmed mean? Motivate your answer clearly!

c. [3 points] Using the QQ-plot you have selected under part (a) determine the location a and scale b approximately. You may use that the sample variance equals 1.32. (You may use that the expectation and variance of the χ^2_k-distribution are k and $2k$, respectively.)

Question 4 [8 points]
Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with unknown distribution P. Suppose that the sample variance $T_n(X_1, \ldots, X_n) = S_X^2$ is used to estimate the variance of P. To determine the accuracy of this estimator, its standard deviation is estimated by means of the empirical bootstrap.

a. [4 points] Describe the steps of the empirical bootstrap scheme that you would use to find the bootstrap estimate of the standard deviation of T_n.

b. [2 points] Describe shortly which two errors are (necessarily) made in this bootstrap procedure.

c. [2 points] Which of the two errors in part (b) can be made arbitrarily small? What do you have to change in the procedure under (a) to make this error smaller?