Vrije Universiteit Amsterdam	Statistical Data Analysis, Exam I
Faculty of Sciences	28 March 2019

Use of a basic calculator is allowed. Graphical calculators and mobile phones are not allowed. This exam consists of 4 questions on 2 pages (27 points).

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

You have 120 minutes to write the exam.

GOOD LUCK!

Question 1 [8 points]
Indicate for each of the following statements whether it is correct or not. Motivate/explain your answers shortly.

a. [2 points] If the skewness of a distribution is equal to 0, then the distribution is symmetric.

b. [2 points] The chi-square goodness-of-fit tests can only test whether data are χ^2_k distributed, for any degrees of freedom $k \in \mathbb{N}$.

c. [2 points] The asymptotic influence function of an estimator partially determines its asymptotic variance.

d. [2 points] In general, it is unavoidable that there are approximation errors in bootstrap procedures.

Question 2 [8 points]
Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables with unknown cumulative distribution function F.

a. [2 points] Explain the difference in the null hypotheses that can be tested with the Shapiro-Wilk and the Kolmogorov-Smirnov test.

b. [2 points] If F_0 is a continuous distribution function and $F = F_0$, discuss whether the critical values of the Kolmogorov–Smirnov test depend on the specific choice of F_0.

Reminder: the test statistic takes the form $D_n = \sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F_0(x)|$,

c. [2 points] Suppose that the distribution of D_n under the null hypothesis is unknown. Discuss whether it makes more sense to generate bootstrap samples from the distribution F_0 or to generate bootstrap samples from the empirical distribution. Motivate your answer.

d. [2 points] Describe the steps that are made in bootstrapping the Kolmogorov-Smirnov test statistic when using the bootstrap method that you preferred in part c.

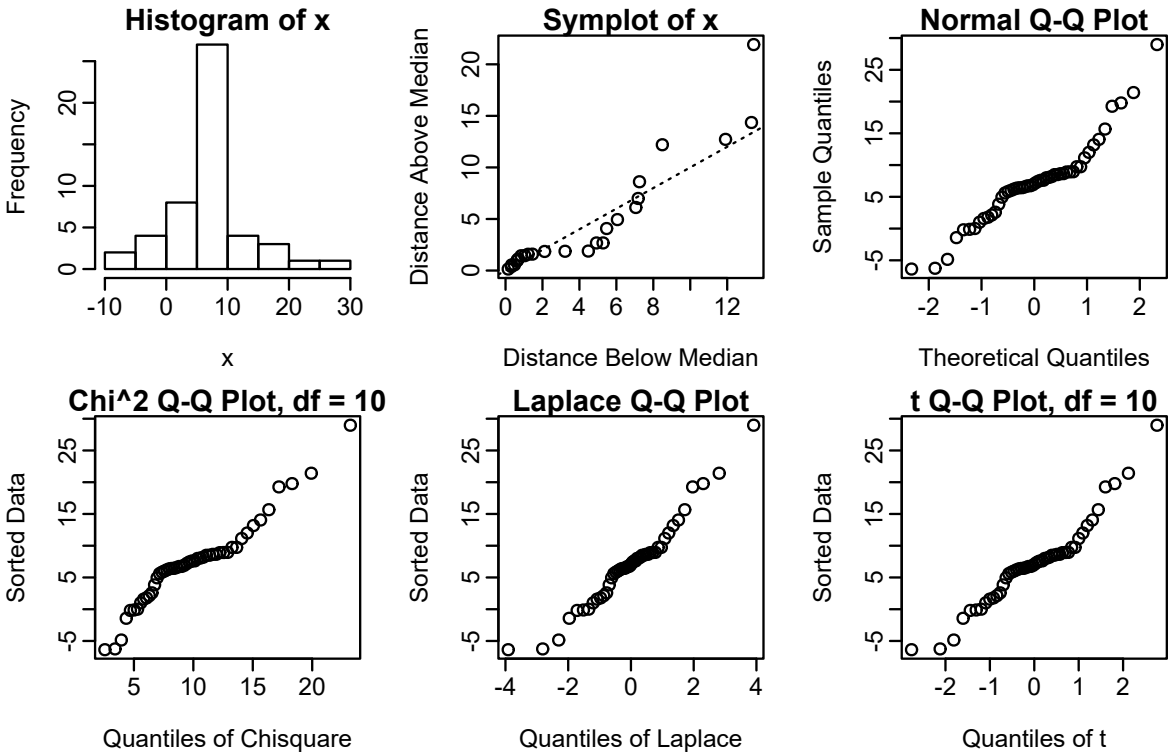

Figure 1: Histogram, symplot, and QQ-plots against indicated distributions of a sample `x` of size $n = 50$.

Question 3 [5 points]

In Figure 1 the histogram, symplot and QQ-plots with respect to the standard normal, χ^2_{10}, Laplace, and t_{10} distributions are shown for a data set `x`. It has a sample skewness of 0.64. The table to the right contains theoretical values of expectation and variance of the reference distributions.

distribution	expectation	variance
standard normal	0	1
χ^2_{10}	10	20
Laplace	0	2
t_{10}	0	1.25

a. [1 point] Based on the plots in Figure 1, which location-scale family do you think is the most appropriate for these data? Motivate your answer.

b. [2 points] With regard to the location-scale family that you chose in part a., determine the location a and the scale b. Use that the sample mean, standard deviation, and variance are $\bar{x} = 7.2$ and $\hat{\sigma} = 6.6$, and $\hat{\sigma}^2 = 43.56$ respectively.

c. [2 points] The 10% trimmed mean is one of the following numbers: 6.816 or 7.416. Indicate which number is the 10% trimmed mean, and motivate your answer.

Question 4 [6 points]

Let $X_1, X_2, \ldots, X_n$ $(n \geq 2)$ be independent random variables following a lognormal distribution with unknown parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, i.e. $\log(X_i) \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$. For lognormal distributions, $\exp(\mu)$ can be considered as a scale parameter. We are interested in an estimator of $\exp(\mu)$ and our choice is the statistic $T_n(X_1, \ldots, X_n) = \exp(\frac{1}{n}\sum_{i=1}^n \log(X_i))$.

a. [3 points] We wish to use a parametric bootstrap to determine the standard deviation of the statistic T_n. What needs to be done in the first step of the parametric bootstrap and what is required for this? Describe the remaining steps of a reasonable parametric bootstrap scheme for estimating the standard deviation of T_n.

b. [2 points] What other possibility is there to estimate the standard deviation of T_n? Do you prefer this or the method of part a.? Motivate your answer.

c. [1 points] Describe shortly which two errors are typically made in applications of bootstrap procedures.