TECHNISCHE UNIVERSITÄT BERLIN
Fakultät IV – Elektrotechnik und Informatik
Fachgebiet Internet Network Architectures

WirelessLab WS 2016/17
Homework 2: Tools of the Trade - Statistics

The goal of this assignment is to keep on teaching you the necessary tools to perform and analyze measurements. The objective is to make sure you understand a few basic statistical concepts that will allow you to analyze data later in the semester.
Software tools such as Matlab, Octave, R, S-plus, or Matplotlib provide commands that implement median, histograms or boxplots. We want you to know what lies behind these commands.
Download "Performance Evaluation Lecture Notes" from ISIS. Read Chapter 1 up to and including section 1.4. In Chapter 2, read the following sections: 2.1.1, 2.1.2, 2.2.1 to 2.2.3, 2.3, 2.4.3, 2.5, and 2.6. You don't have to understand all the details of what you read. If you actually decide to build your understanding from another source, this is perfectly fine too. You are free to use any other textbook or even Wikipedia. *But* make sure you understand the following concepts:

- Metrics, factors
- Histograms and ECDF
- Median, quantiles, and confidence intervals for the median, as well as the underlying assumptions
- Mean and standard deviation, and confidence intervals
- Boxplot
- Simple Moving Average

The goal of the following questions is to help you check your understanding.

Question 1: (60 Points) *Computing variables and checking assumptions*

(a) What is the ECDF $F(x)$ of the data set

$$-10.1, -1.2, -9.5, -1, -1, -1, 0.1, 5, 7, 7, 7, 7, 2, 2, 2 \tag{1}$$

for $x = -100$, $x = -10$, $x = 0$, $x = 7$ and $x = 10$?

(b) Let $x_i < x_j$ if $i < j$. What is the median of

$$x_{-2}, x_1, x_0$$

and the median of

$$x_{-2}, x_1, x_0, x_2, x_4, x_{-3}.$$

What is the median of the data set in (a)?

(c) What is the main necessary assumption for the calculation of a confidence interval?

(d) Compute the confidence interval for the median of the data set in (a).

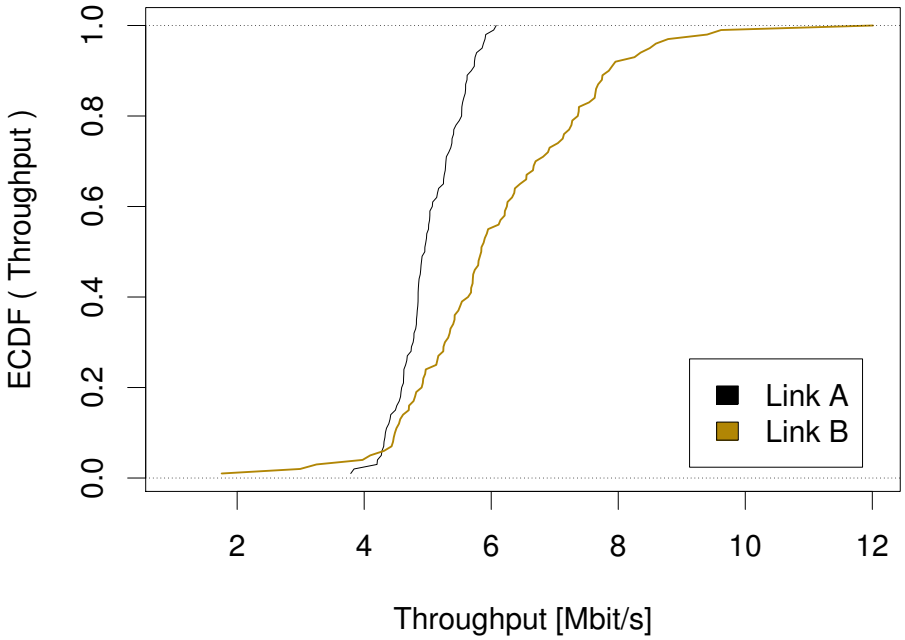
(e) Does it make sense to compute a confidence interval for the mean on the data set in (a)? Why (not)?

(f) If the confidence interval of the difference between the means of two data sets includes 0, what does it imply?

(g) What is the simple moving average of the data set in (a) with $n = 3$?

Question 2: (40 Points) *Reading a plot*

The following plot shows the result of a performance test over two different wireless links:



(a) What is the *metric* here?

(b) What are the median, the quartiles and the 95% quantiles for both links?
(Please give approximated numbers.)

(c) Which link do you think has the higher mean? Which one has the higher standard deviation? Why?
(Do not give numbers, just a tendency.)

(d) Based on this data, which link do you think had the better performance? Why?

(e) What possible *factors* could have influenced the performance? (You can guess.)

Submission

`https://isis.tu-berlin.de/course/view.php?id=8501`

Please submit a PDF document containing *a cover page* with your names and group ID, and *having your group number in its file name*.

If you want to submit multiple files, make an archive (.tar.gz, .zip) containing *a directory* with all of your files and *having your group number in its file name*. All files that belong to a specific question must have the question/subquestion in their filenames. Please try not to clutter your submission with temporary files.

Due Date: Wednesday, November 9th at 11.55 p.m. (23:55).

(Early submissions are possible in order to get feedback before the oral exam.)