WirelessLab (WiSe 2016/2017)

Tutorial 2: Tools of the Trade - Statistics

Theresa Enghardt
Mirko Palmer
Apoorv Shukla

Contact & Discussions via ISIS:
Course ID 8501

http://www.inet.tu-berlin.de/menue/teaching0/ws201617/wl_1617/

Outline

- Organization
- Today's homework
- Performance Analysis
  - Visualizing performance data
  - Summarizing performance data
  - Confidence intervals and assumptions
  - Moving average

Organization: Groups

- Any changes in the groups?
- Anyone still needs a group partner?

  → Group work **mandatory** from assignment 3!
- Problems?
  - First, try to solve it within the group.
  - If necessary, talk to other students.
  - If this fails, **talk to us.**

Organization: Schedule

- Registration on QISPOS:
  By next **Tuesday, November 8th**, 10pm
- Oral exams:
  Next week **Wednesday/Thursday, Nov. 9th and 10th**
- When do you have time? (As a group)
  Please fill out the poll by **Monday, November 7th**

  → Your time slot for oral exam and debriefings

  Schedule of oral exams: On Monday.

Homework 2: Performance Analysis

- Basic concepts of performance analysis and statistics

- Due by: Wednesday, November 9th at 11.55 p.m (23:55).
- Not graded, but contents part of oral exam
- Early hand-ins possible (to get early feedback)

“There's lies, damn lies, and statistics”

Performance analysis

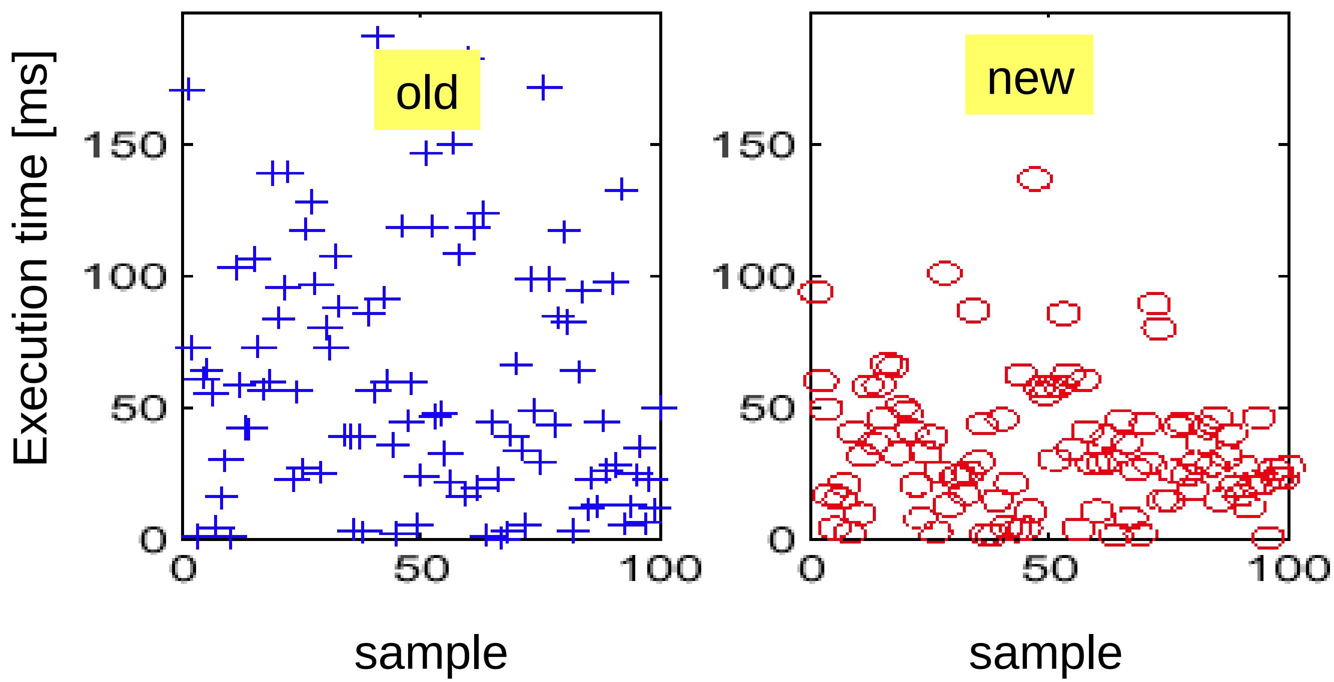
"How good is my throughput?"

"Has execution time improved?"

"Which system performs better?"

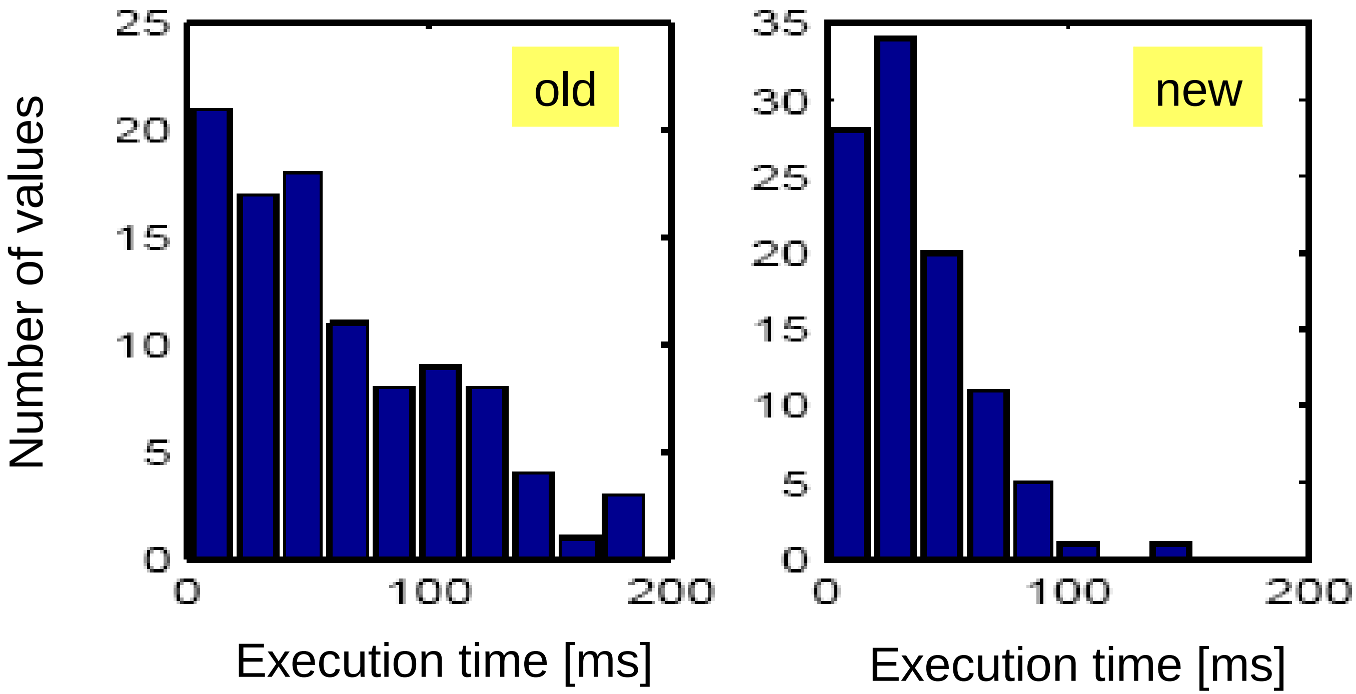
- Throughput and execution time are **performance metrics**.
  - **Measure** and/or **compute** it to get data.
  - **Analyze** the data using statistics.
  - Draw conclusions.

Performance Data



Has the performance of the system improved?

Histogram



Number of values
25
20
15
10
5
0
0
100
200
old
Execution time [ms]
35
30
25
20
15
10
5
0
0
100
200
new
Execution time [ms]

Empirical Cumulative Distribution Function (ECDF)

$$F(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{x_i \leq x\}}$$

Summarizing Performance Data

- “**How much** has performance improved?”
- Quantify:
  - Central value
  - Dispersion (Variability)

Median and p%-quantiles

- Median: value that falls in the middle of a distribution

If n is odd, the median is $x_{\left(\frac{n+1}{2}\right)}$, otherwise $\frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right)$

  - Example: median for *-3,-2,-2,1,2*

    = -2

- *p%*-quantile: *p%* of the observation below and *(100-p)%* above
  - Example: *75%*-quantile

    = *1*

Mean and standard deviation

- Mean

$$m = \frac{1}{n} \sum_{i=1}^{n} x_i$$

- Ex: mean for *-3,-2,-2,1,2*

  = *-0.8*;
- Standard deviation

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - m)^2 \text{ or } s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - m)^2$$

  – How much does the data differ from the mean?

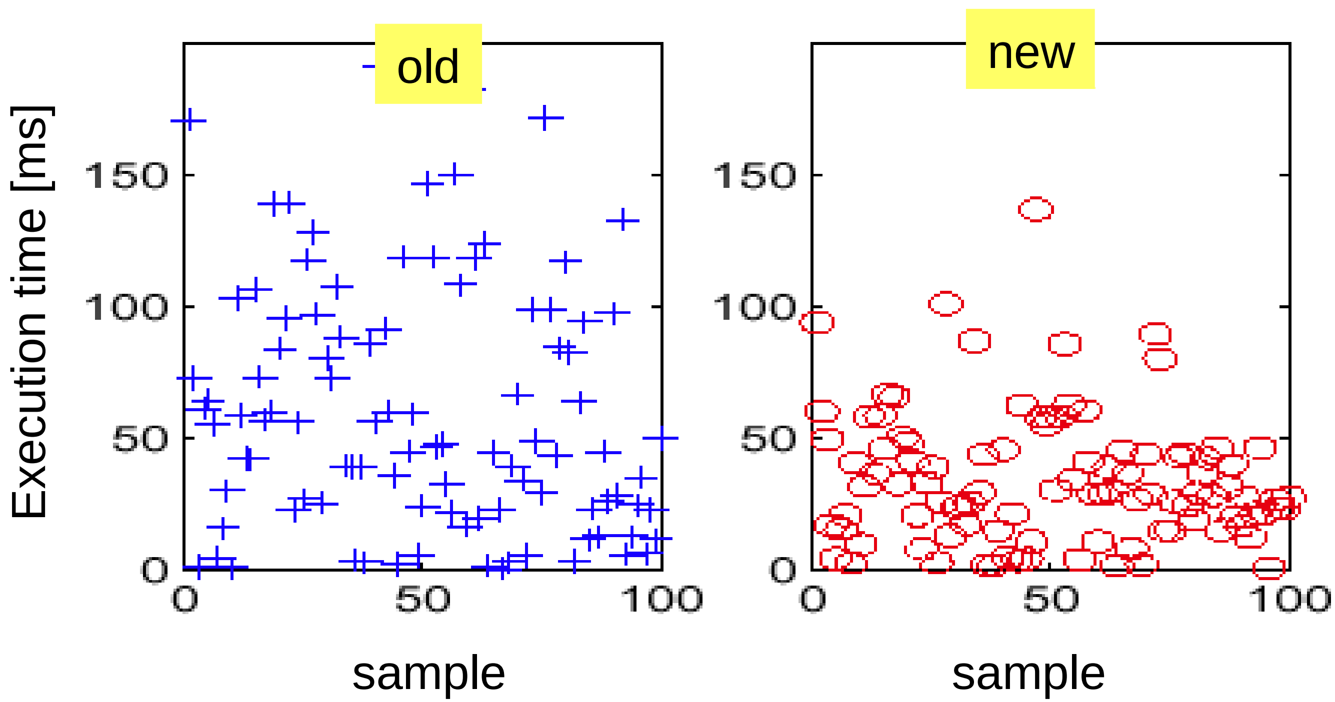
- Ex: *s* is *1.94*

Which metric to use ?

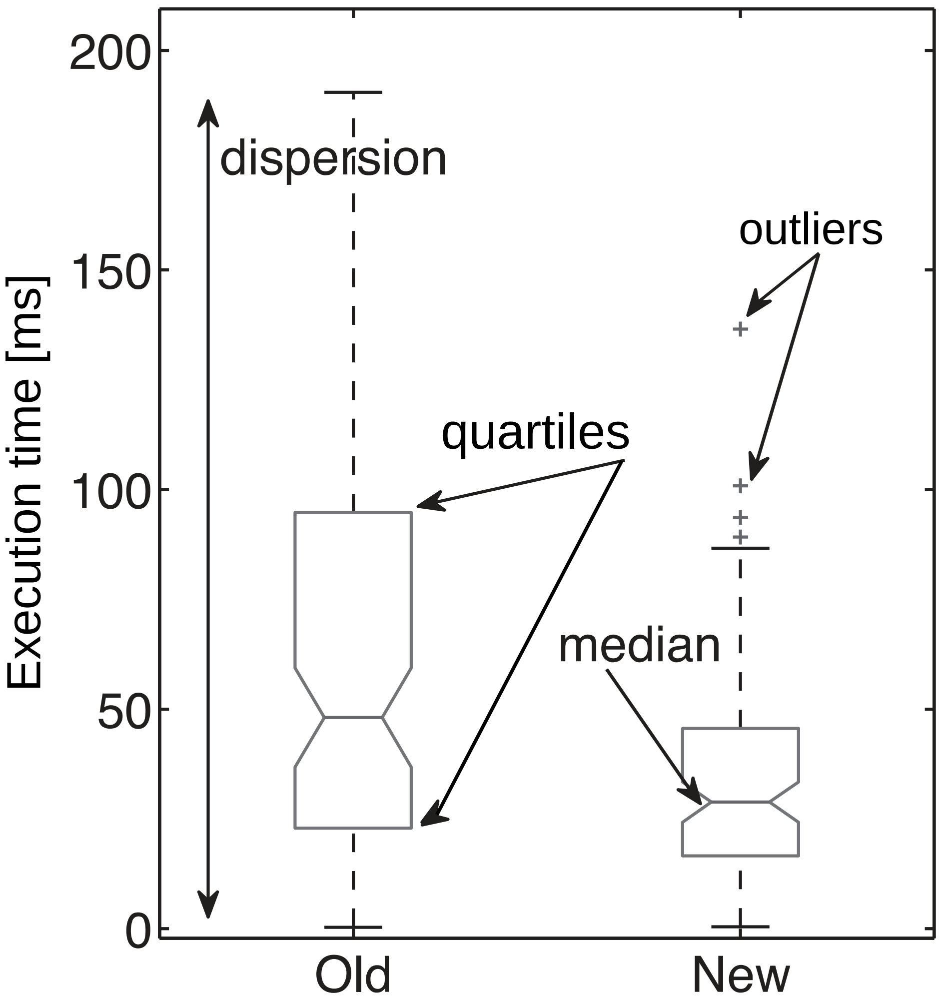
- Mean? Median? Quantiles?
  - Mean contains “more” of the data
  - Problem: Sensitive to outliers
- Example: *-3,-2,-2,1,2*
  - Median is *-2,* mean is *-0.8*
  - Now we measure a “40” (by mistake?)
  - Median is now *-0.5,* Mean is now *6*

Median and mean

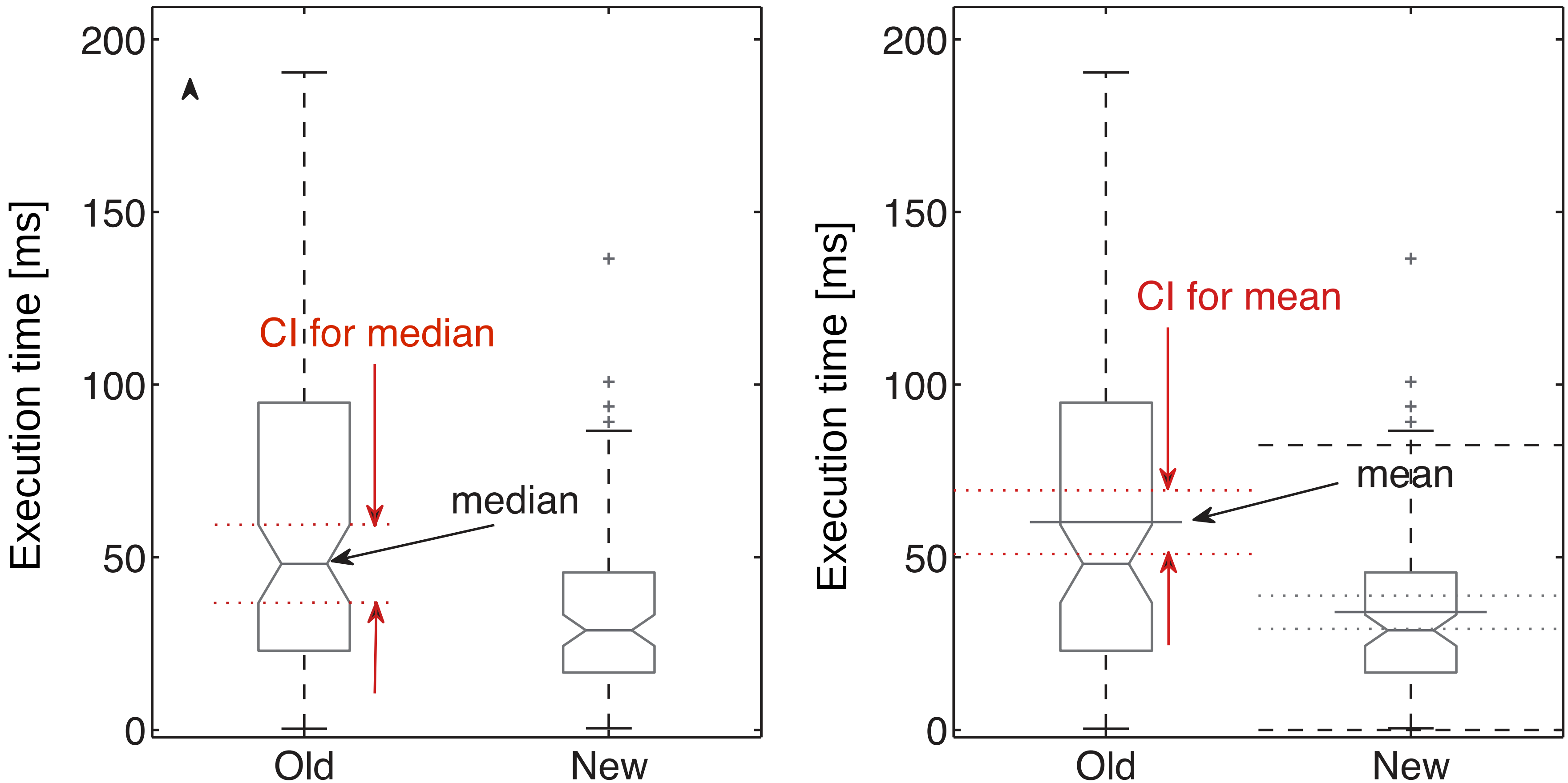
- Let‘s look at the first example again...



Boxplots of median and mean



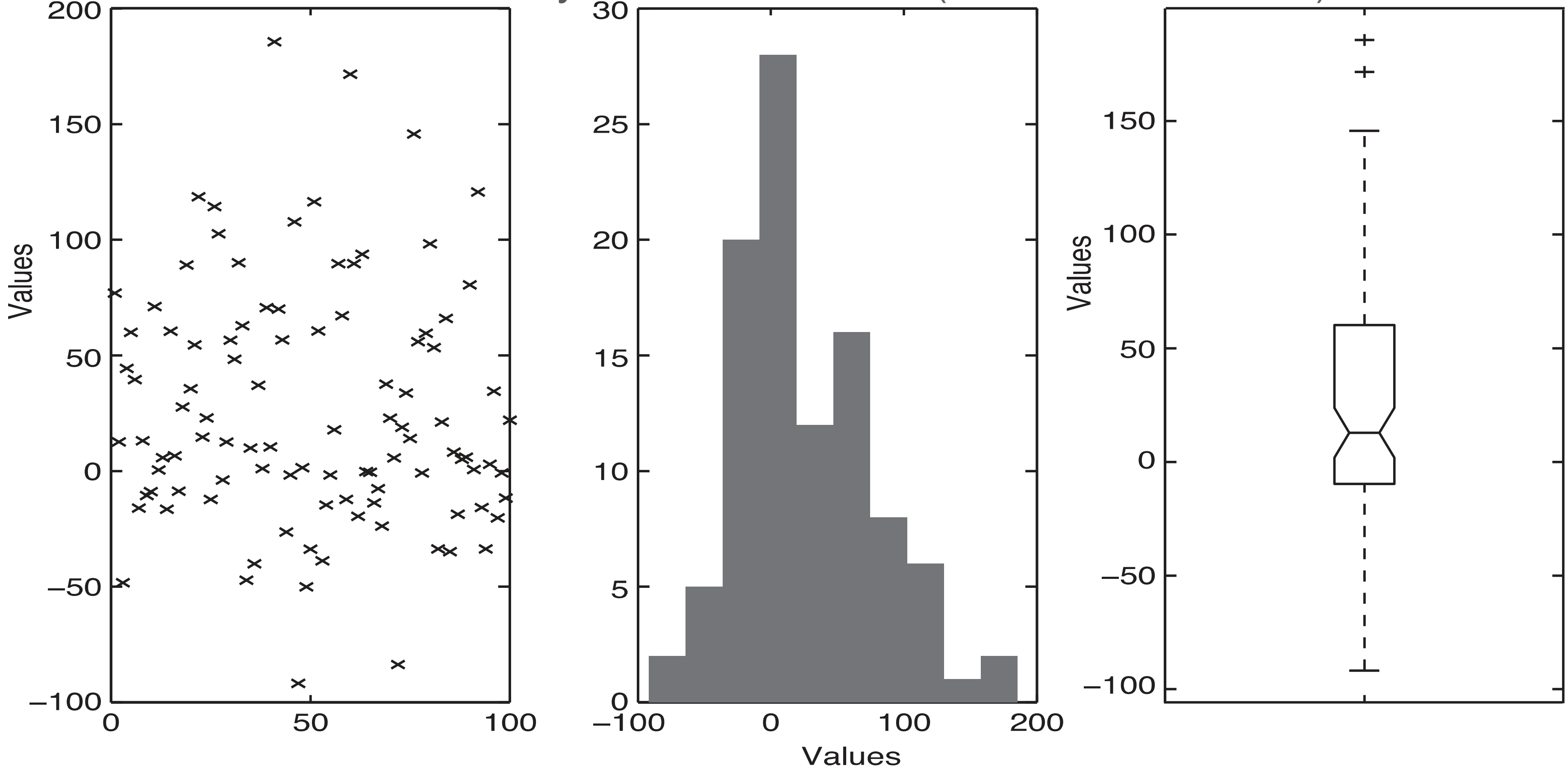
Confidence Interval (CI)



Quantifies *uncertainty* of an estimation...
(Based on our measured sample, where do we estimate the “true” median/mean of the system's execution time?)

Confidence Interval (CI)

… or accuracy of the difference (“new” - “old” value)



→ Could the “true” difference be zero? (= No improvement)

Computing Confidence Intervals

- We assume that the data comes from an *iid* stochastic model

  Independent Identically Distributed

- We treat data like samples of random variables
  - with the same stochastical distribution
  - where previous samples do not influence next sample

Independence assumption

How do I know if this is true ?

- In controlled experiment:
  Avoid unintended influence of factors
  (Draw factors randomly with replacement)
- In measurement on a running system:
  Randomize the measurements

Confidence Interval for Median

- Based on binomial distribution shown in table on the right

- Order the data and take two values as indexed by j and k

- p = probability / confidence level

n	j	k	p
$n \leq 5$: no confidence interval possible.			
6	1	6	0.969
7	1	7	0.984
8	1	7	0.961
9	2	8	0.961
10	2	9	0.979
11	2	10	0.988
12	3	10	0.961
13	3	11	0.978
14	3	11	0.965
15	4	12	0.965
16	4	12	0.951
17	5	13	0.951
18	5	14	0.969
19	5	15	0.981
20	6	15	0.959
21	6	16	0.973
22	6	16	0.965
23	7	17	0.965
24	7	17	0.957
25	8	18	0.957
26	8	19	0.971
27	8	20	0.981
28	9	20	0.964
29	9	21	0.976
30	10	21	0.957
31	10	22	0.971
32	10	22	0.965

Confidence Intervals for mean

- Calculate based on standard deviation:
  e.g: $1.96 \times s/\sqrt{n}$ for 95% CI
- Assumptions:
  - iid
  - Large number of samples
  - Common distribution has finite variance (e.g. normal distribution)

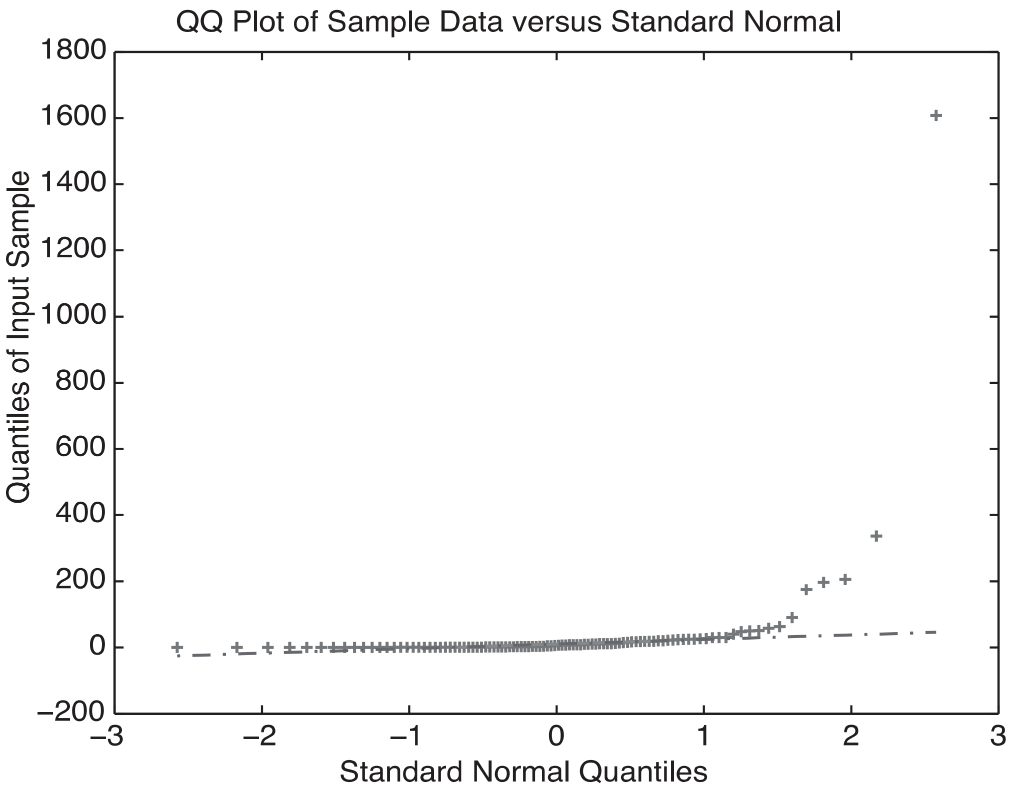
CI for mean, assumptions

- When is the distribution "right"?
  - n>30
    - If central limit theorem holds (in practice: *n* is large and distribution is not "wild")
    - Close to normal, or not heavy tailed
  - n<30
    - Data must come from an *iid* + *normal* distribution

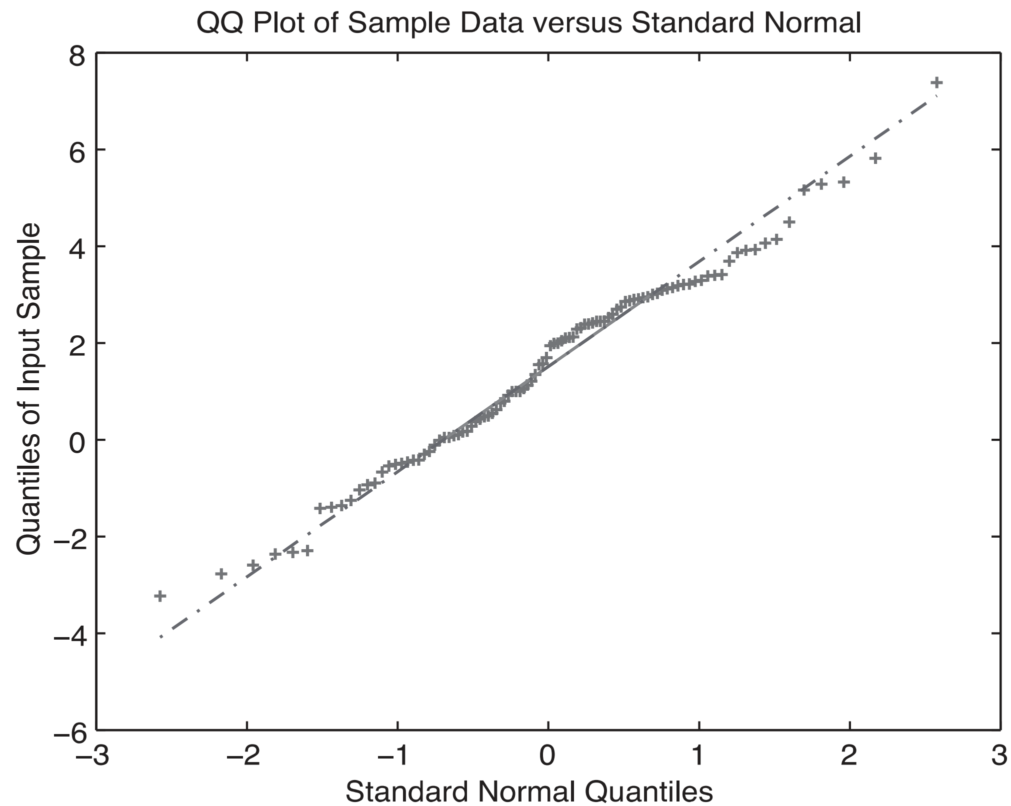
Normal distribution assumption

- Normal Qqplot
  - X-axis: standard normal quantiles
  - Y-axis: Ordered statistic of sample
- If data comes from a normal distribution, qqplot is close to a straight line (except for end points)
  - Visual inspection is often enough

QQPlots



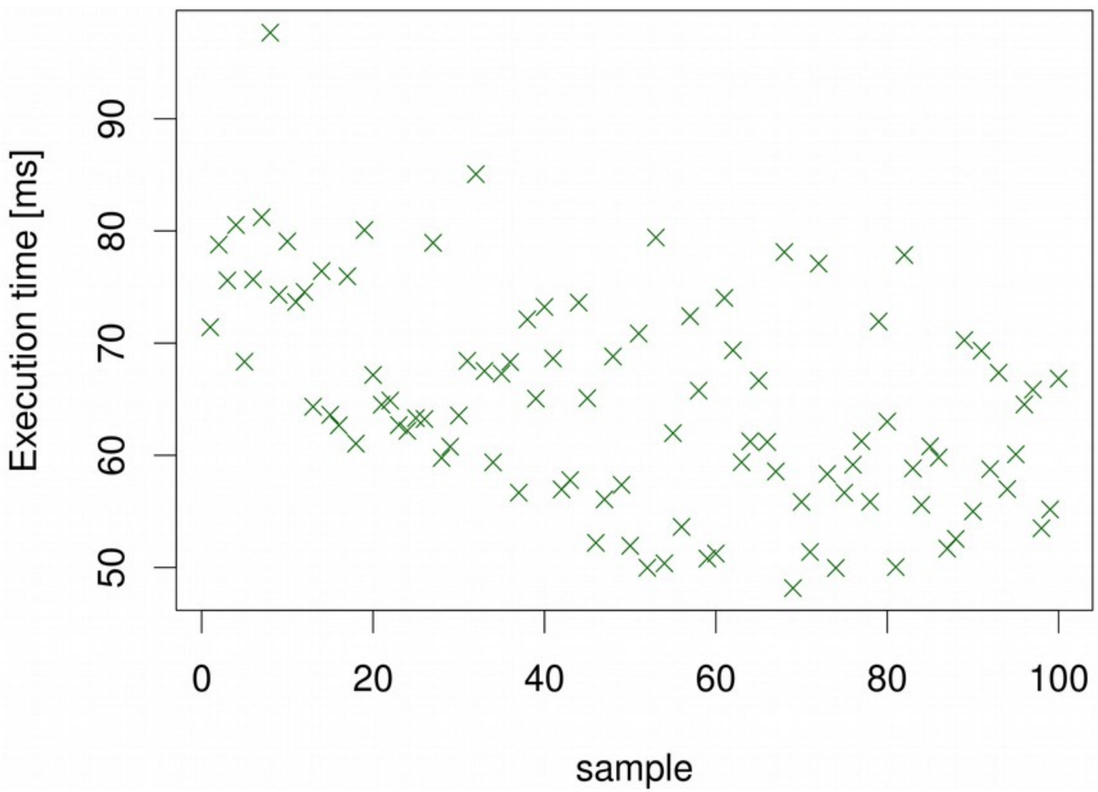
(a) (qq-plot)



(b) (qq-plot of log of data)

Time series data

- Time dependencies → data is *not* iid!
- But: Observe how performance changes over time

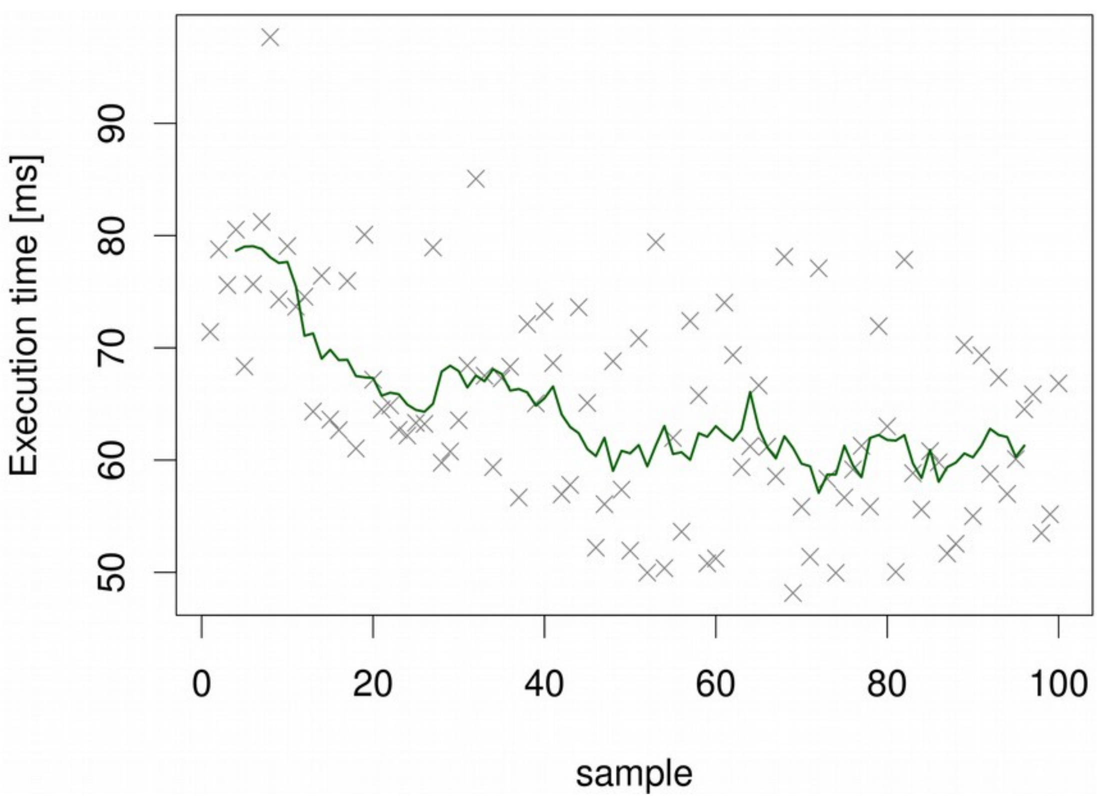


Moving average

- Get a “smoother” plot of the trend
- **Simple Moving Average (SMA)** with n = 8 (window size 8):

$$SMA = \frac{p_M + p_{M-1} + \cdots + p_{M-(n-1)}}{n}$$

$$= \frac{1}{n}\sum_{i=0}^{n-1} p_{M-i}$$



Factors

- So far: “Old” and “New” performance data
- In practice
  - e.g. different server
  - e.g. different times of day
- These influence **factors**
  - System load
  - Network traffic from others (cross traffic)
  - Network configuration (e.g. socket buffer size)

$\rightarrow$ Is the performance change really due to the factor I wanted?

Reading material

- "Performance Evaluation of Computer and Communication Systems"
  by Jean-Yves Le Boudec.
- https://infoscience.epfl.ch/record/146812/files/perf
  PublisherVersion_1.pdf