

Basic Concepts in Statistics

Daniel A. Menascé, Ph.D.
Dept of Computer Science
George Mason University

© 2001 D. A. Menascé. All Rights Reserved.

1

Goals in Studying Statistics

- Analyze, present, and describe numerical information properly.
- Draw conclusions about the properties of large populations from sample information (inference).
- Design experiments to learn about real-world situations.
- To forecast or predict not-measured values from a set of measurements.

© 2001 D. A. Menascé. All Rights Reserved.

2

Population and Sample

- Population (or universe): all N members of a class or group.
 - E.g., all files retrieved from a Web site since the site went into operation.
- Sample: portion of the population. Its size is denoted by n .
 - E.g., the set of files retrieved from a Web site from 10:00 AM to 2:00 PM on January 03, 2001.

© 2001 D. A. Menascé. All Rights Reserved.

3

Census, Parameter, Statistic

- Census: enumeration or count of every member of the population.
- Parameter: summary measure of the individual observations made in census of an entire population.
 - E.g., average size of all files ever retrieved from the Web site.
- Statistic: summary measure obtained from a sample.
 - E.g., average size of all files retrieved from the Web site from 10:00 AM to 2:00 PM on January 03, 2001.

© 2001 D. A. Menascé. All Rights Reserved.

4

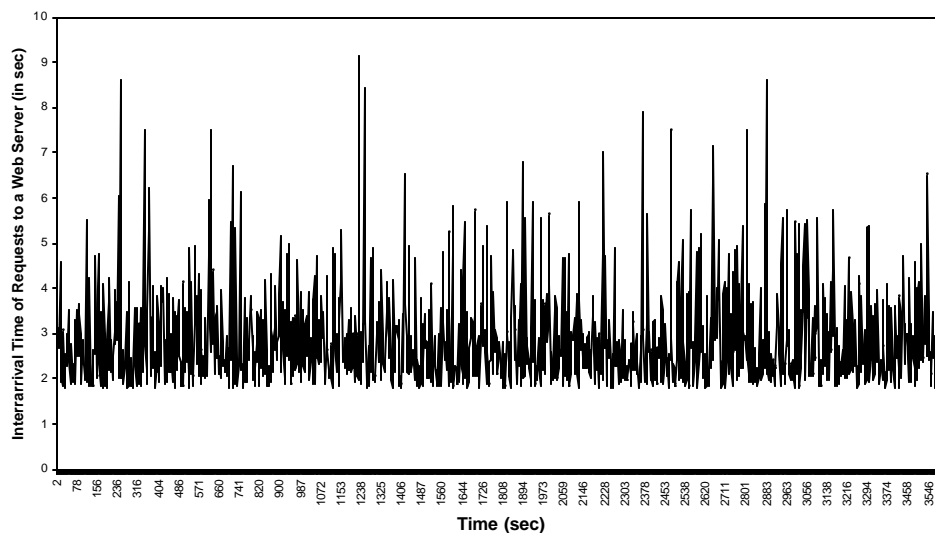
Visualizing Numerical Data

- Type of Plots:
 - Time ordered plots: the time scale is time.
 - Time-scale analysis: time is slotted into fixed time intervals. The y-axis displays a statistics over the time slot (e.g., sum, average).
 - Changing the time scale may reveal interesting properties about the variable being plotted (e.g., strong correlations between adjacent time intervals).
 - Percent frequency histograms: show the percentage of occurrences of values in a bin (range of values).
 - Cumulative frequency histograms.

© 2001 D. A. Menascé. All Rights Reserved.

5

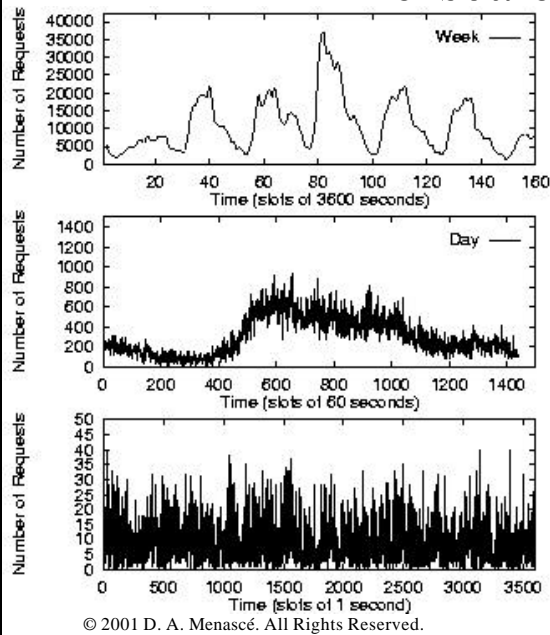
Example of a Time Plot



© 2001 D. A. Menascé. All Rights Reserved.

6

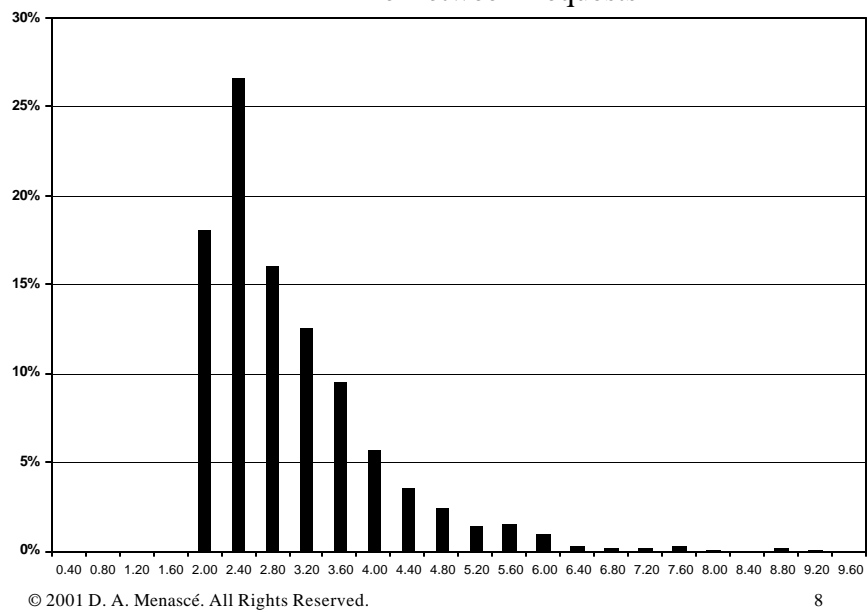
Time-scale Analysis



From "In Search of Invariants in E-commerce Workloads," Menascé et al, Proc. ACM Conference on E-commerce, Minneapolis, MN, October 17-20, 2000.

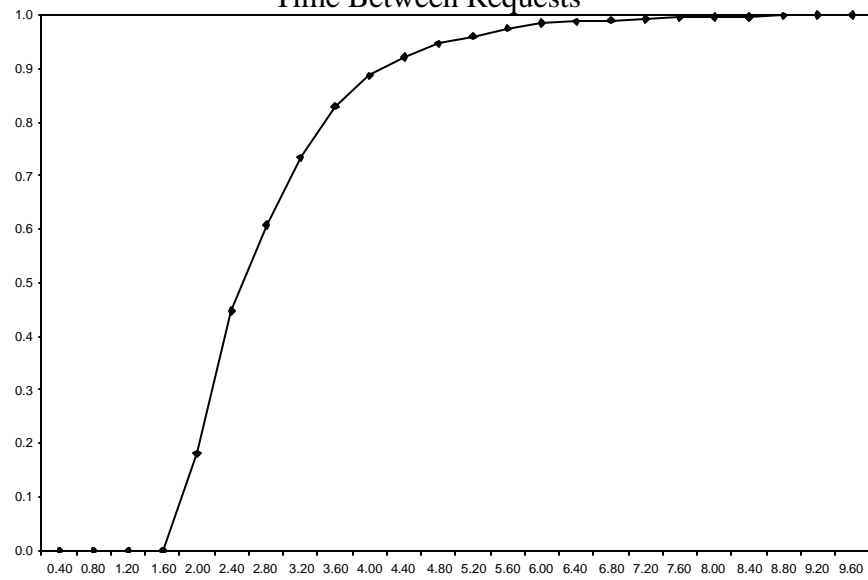
7

Example of a Percentage Frequency Histogram for Interarrival Time Between Requests



8

Example of a Cumulative Percentage Frequency Plot for Interarrival Time Between Requests



© 2001 D. A. Menascé. All Rights Reserved.