# Determining Distributions

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





# Problem

- Given a data set, determine the best distribution that fits the data.
- Useful for simulation and analytic studies.
- Start by plotting a histogram to narrow down type of distribution.
- Then, use a quantile-quantile plot.

# Quantile-Quantile plot

$y_i$ : observed qi-th quantile

$x_i$ : theoretical qi-th quantile

If the observations come from the same distribution, the quantile-quantile plot should be linear.

To determine the theoretical quantile, we need to be able to invert the CDF F(x).

$$q_i = F(x_i)$$

$$x_i = F^{-1}(q_i)$$





# Examples of CDFs and Their Inverse Functions

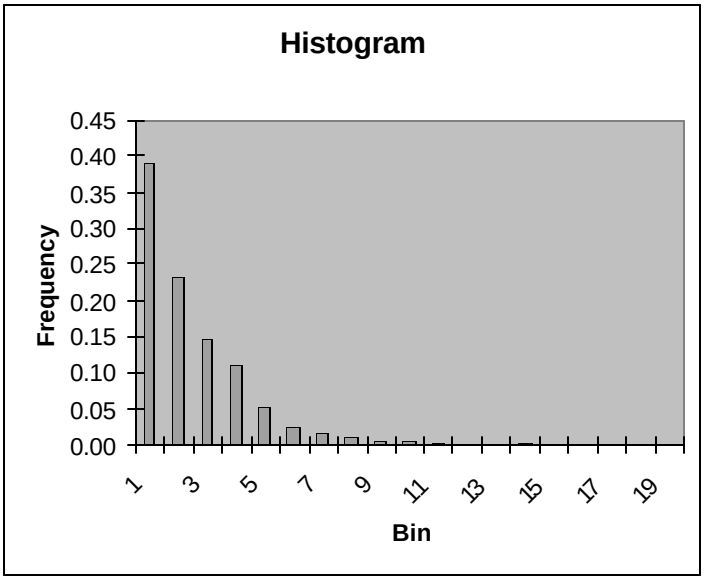
| | | |
|---|---|---|
| Exponential | $F(x) = 1 - e^{-x/a}$ | $-a\text{Ln}(1-u)$ |
| Pareto | $F(x) = 1 - x^{-a}$ | $\frac{1}{(1-u)^{1/a}}$ |
| Geometric | $F(x) = 1 - (1-p)^x$ | $\left\lceil \frac{\text{Ln}(u)}{\text{Ln}(1-p)} \right\rceil$ |

# Example of a Quantile-Quantile Plot

- One thousand values are suspected of coming from an exponential distribution (see histogram in the next slide). The quantile-quantile plot is pretty much linear, which confirms the conjecture.



5





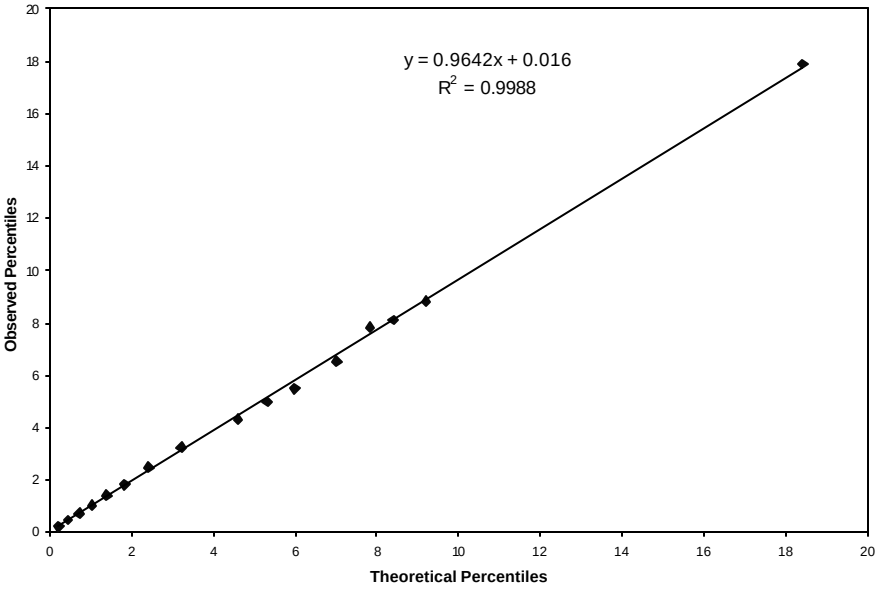

6

# Data for Quantile-Quantile Plot

| qi | yi | xi |
|---|---|---|
| 0.100 | 0.22 | 0.21 |
| 0.200 | 0.49 | 0.45 |
| 0.300 | 0.74 | 0.71 |
| 0.400 | 1.03 | 1.02 |
| 0.500 | 1.41 | 1.39 |
| 0.600 | 1.84 | 1.83 |
| 0.700 | 2.49 | 2.41 |
| 0.800 | 3.26 | 3.22 |
| 0.900 | 4.31 | 4.61 |
| 0.930 | 4.98 | 5.32 |
| 0.950 | 5.49 | 5.99 |
| 0.970 | 6.53 | 7.01 |
| 0.980 | 7.84 | 7.82 |
| 0.985 | 8.12 | 8.40 |
| 0.990 | 8.82 | 9.21 |
| 1.000 | 17.91 | 18.42 |

7






8

# What if the Inverse of the CDF Cannot be Found?

- Use tables and interpolate.
- Approximation for N(0,1):

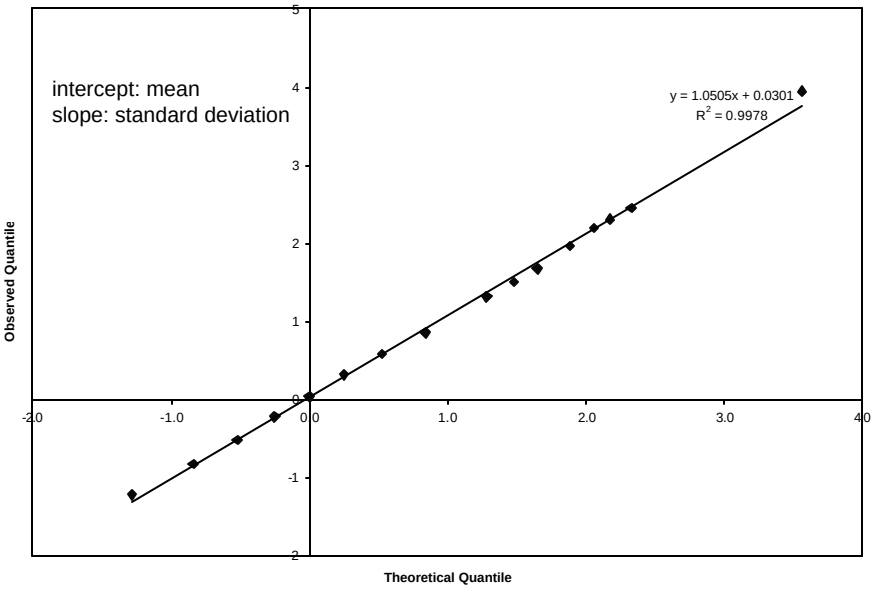
$$x_i = 4.91[q_i^{0.14} - (1 - q_i)^{0.14}]$$

- For N(μ,σ) the $x_i$ values are scaled as

$m + s\,x_i$ before plotting.

9



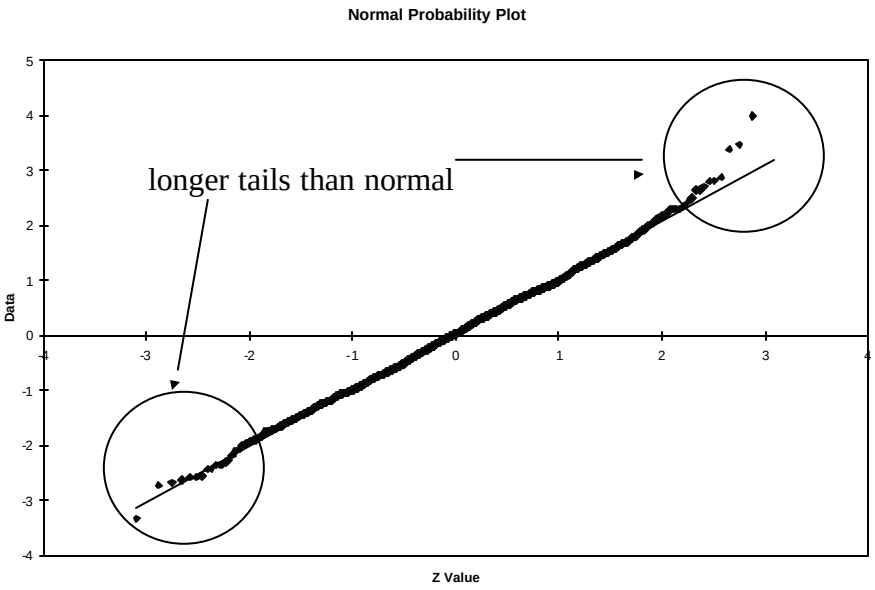


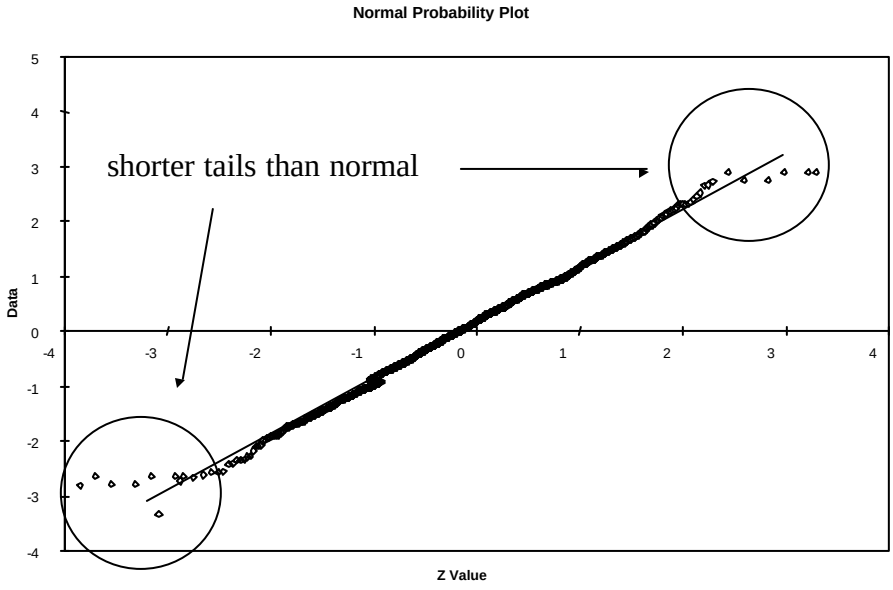

10

# Using PhStat

- The normality assumption can be checked with PhStat by using the Normal Probability Plot option from the Probability Distribution option. (example in next slide).

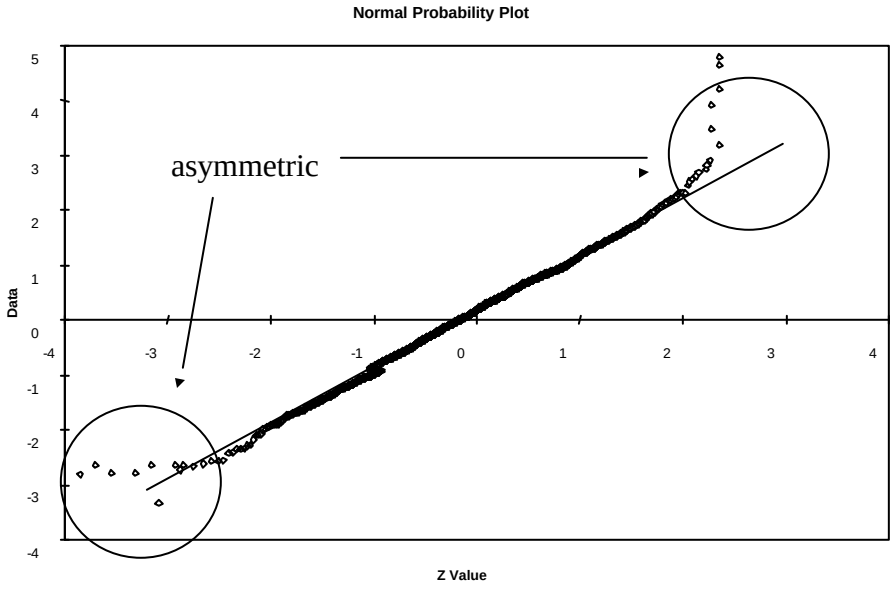





13






14