Components of Representativeness

- **Assessor's Population of Concern**
  - population for which assessment is being conducted
- **Surrogate Population**
  - population used by the assessor when data on the population of concern are not available
- **Population of Concern for the Surrogate Study**
  - population for which the surrogate study was designed
- **Population Sampled**
  - a sample from the population of concern of the surrogate study

Components of Representativeness

- **External Components:**

  Population of Concern vs Surrogate Population

- **Internal Components:**

  Population Sampled vs Population of Concern for the Surrogate Study

Components of Representativeness

External to Study:

How well does the surrogate population represent the population of concern?

Components of Representativeness

Internal to Study:

- How well do individuals sampled represent population of concern for the study?
- How well do actual number of respondents represent the sampled population?
- How well does the measured value represent the true value for the measured unit?

Checklists

- Assessing representativeness of population sampled vs population of concern for the surrogate study
- Assessing representativeness of surrogate population vs assessor's population of concern (individual, spatial, and temporal characteristics)

Attempting to Improve Representativeness

- Adjustments to account for differences in population characteristics
- Adjustments to account for time-unit differences

Issues

- Are the checklists useful tools?
- How does one evaluate the importance of the differences between two populations?
- How should extrapolations be made?

Discussion Topic 2

Empirical Distribution Functions vs Analytic Distribution Functions

Assessing Quality of Fit for Analytic Distributions

Empirical Distribution Function (EDF)

Given representative data, $X = \{x_1, x_2, \cdots, x_n\}$, sorted from smallest to largest, $x_1 \le x_2 \le \cdots \le x_n$, the EDF is the cumulative distribution function defined by

$$\hat{F}(x) = \frac{number\ of\ x_k \le x}{n} \qquad or \qquad \hat{F}(x) = \frac{1}{n}\sum_{k=1}^{n} H(x - x_k)$$

where H(u) is the unit step function which jumps from 0 to 1 when u $\ge$ 0. The values of the EDF are the discrete set of cumulative probabilities $(0, 1/n, 2/n, \cdots, n/n)$.
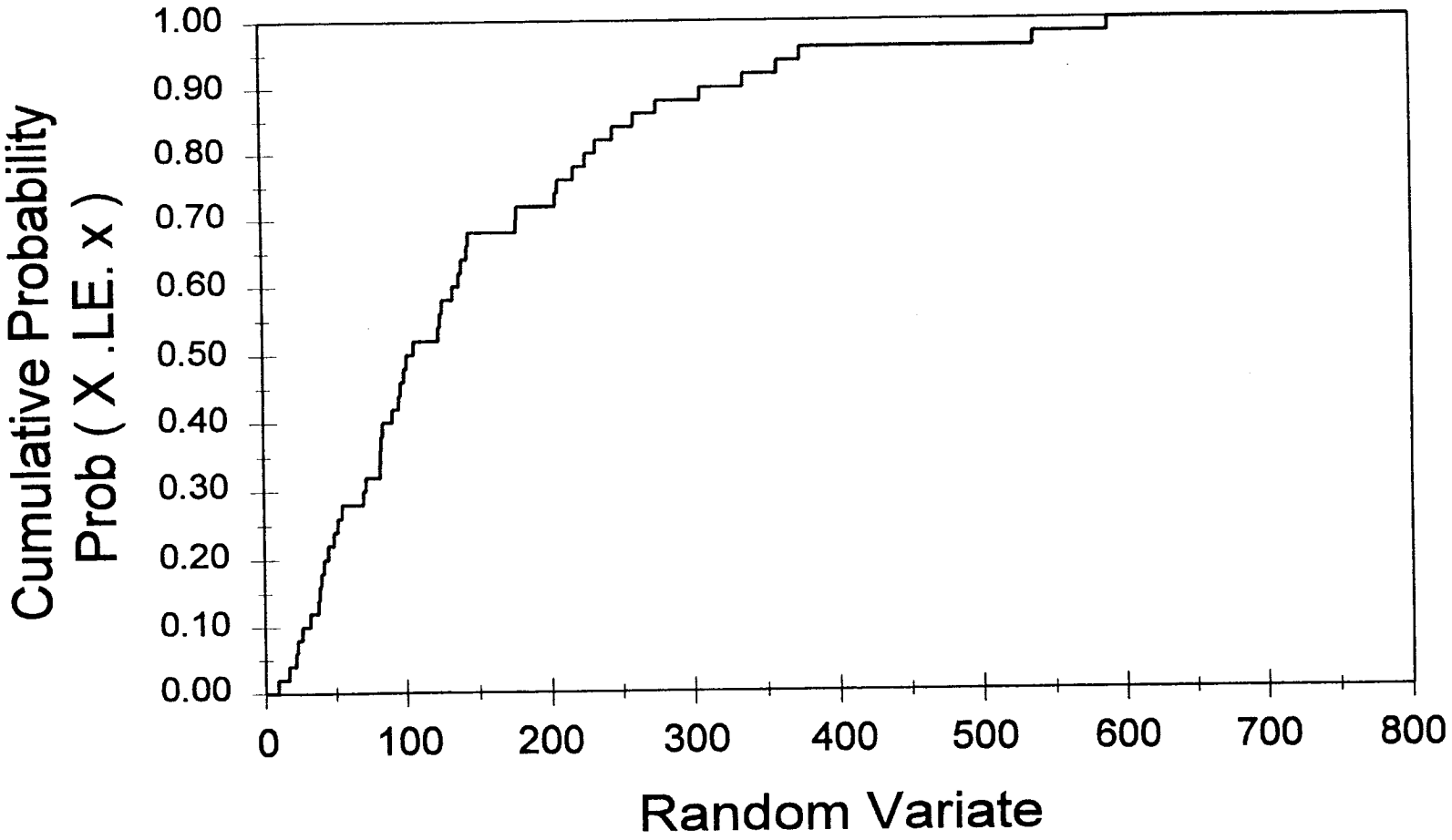
Figure 1. Example of EDF