JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

Section F

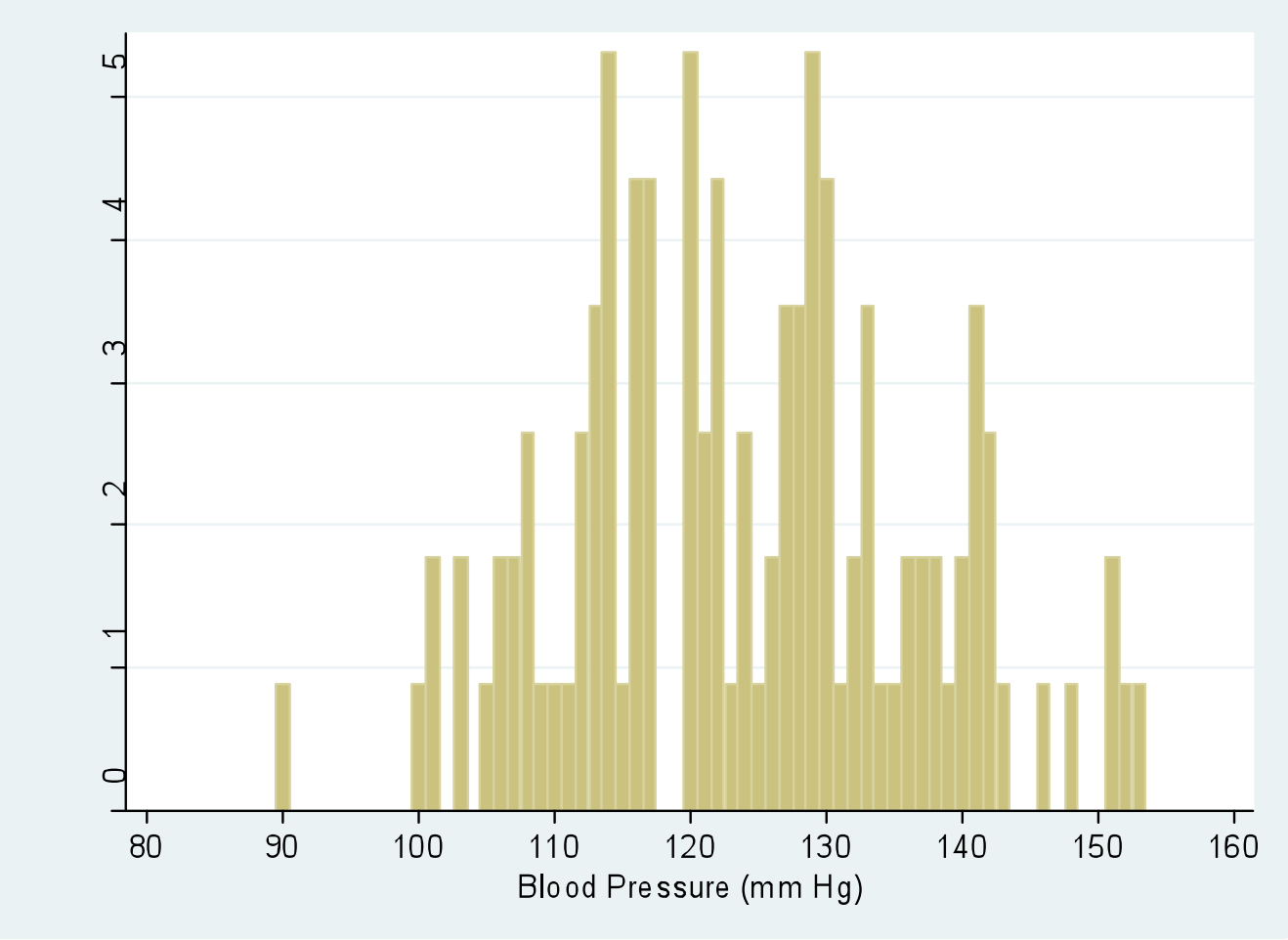
Samples versus Populations, Part 2: Sample Distribution versus Underlying "Population Distribution"

Sample Distribution

- In research, samples are taken from larger population
- If the sample is taken randomly, the sample characteristics will imperfectly mimic the population characteristics
- The characteristics include the mean, median and sd (but also the distribution of individual values)

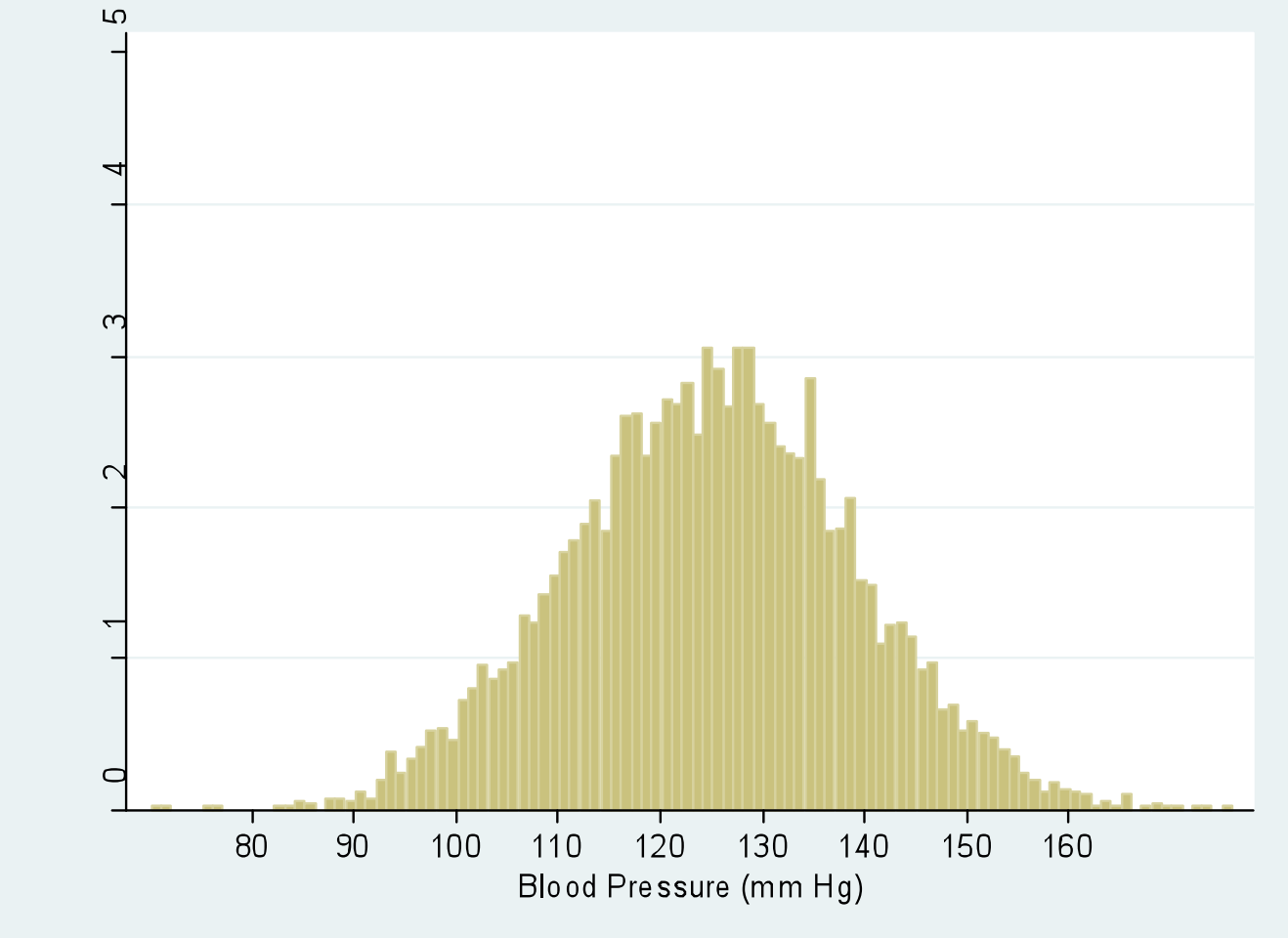
Example 1: Blood Pressure in Males

- Histogram of BP values for random sample of 113 men



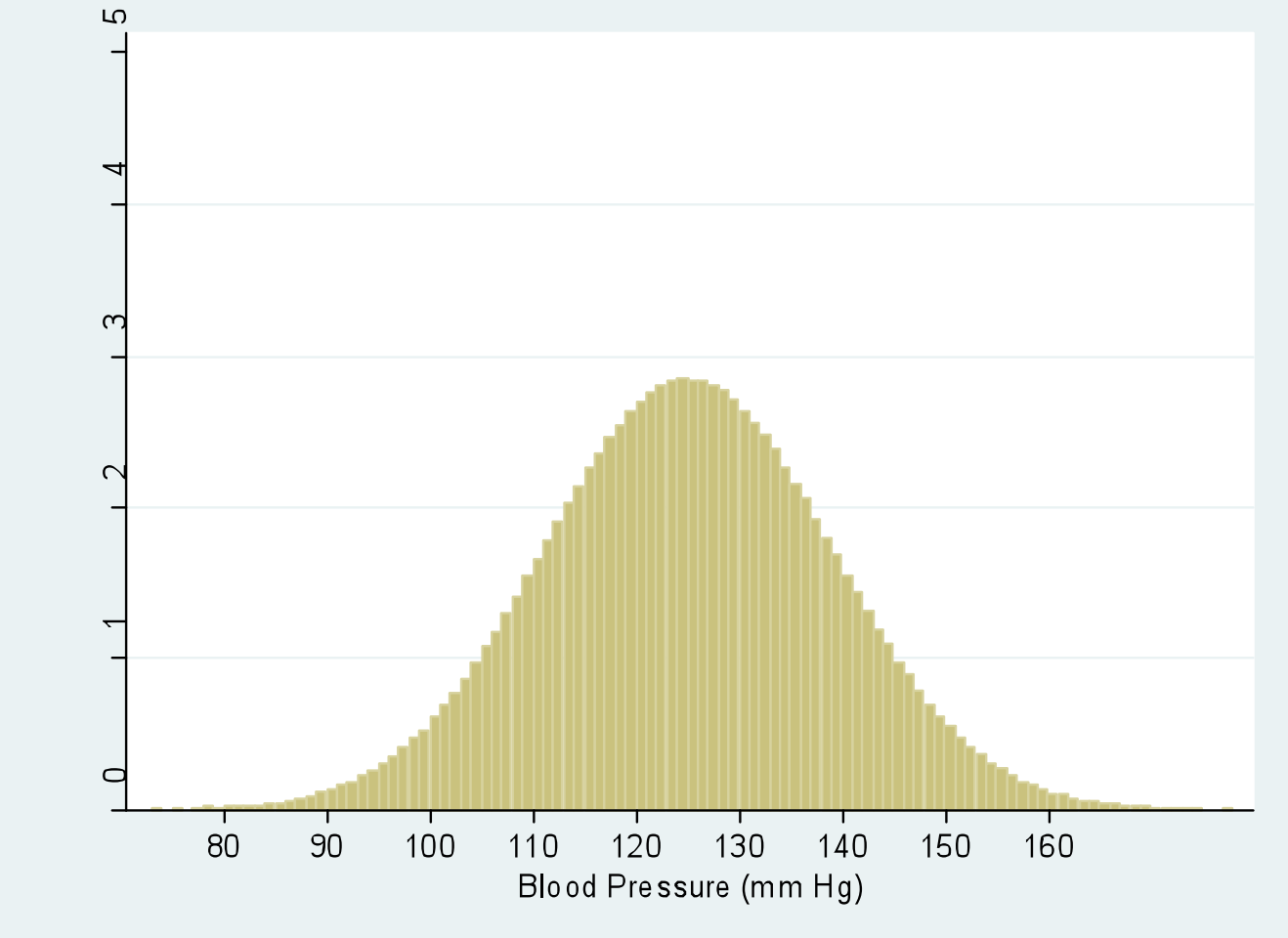
Example 1: Blood Pressure in Males

- Histogram of BP values for random sample of 500 men



Example 1: Blood Pressure in Males

- Histogram of BP values for male population

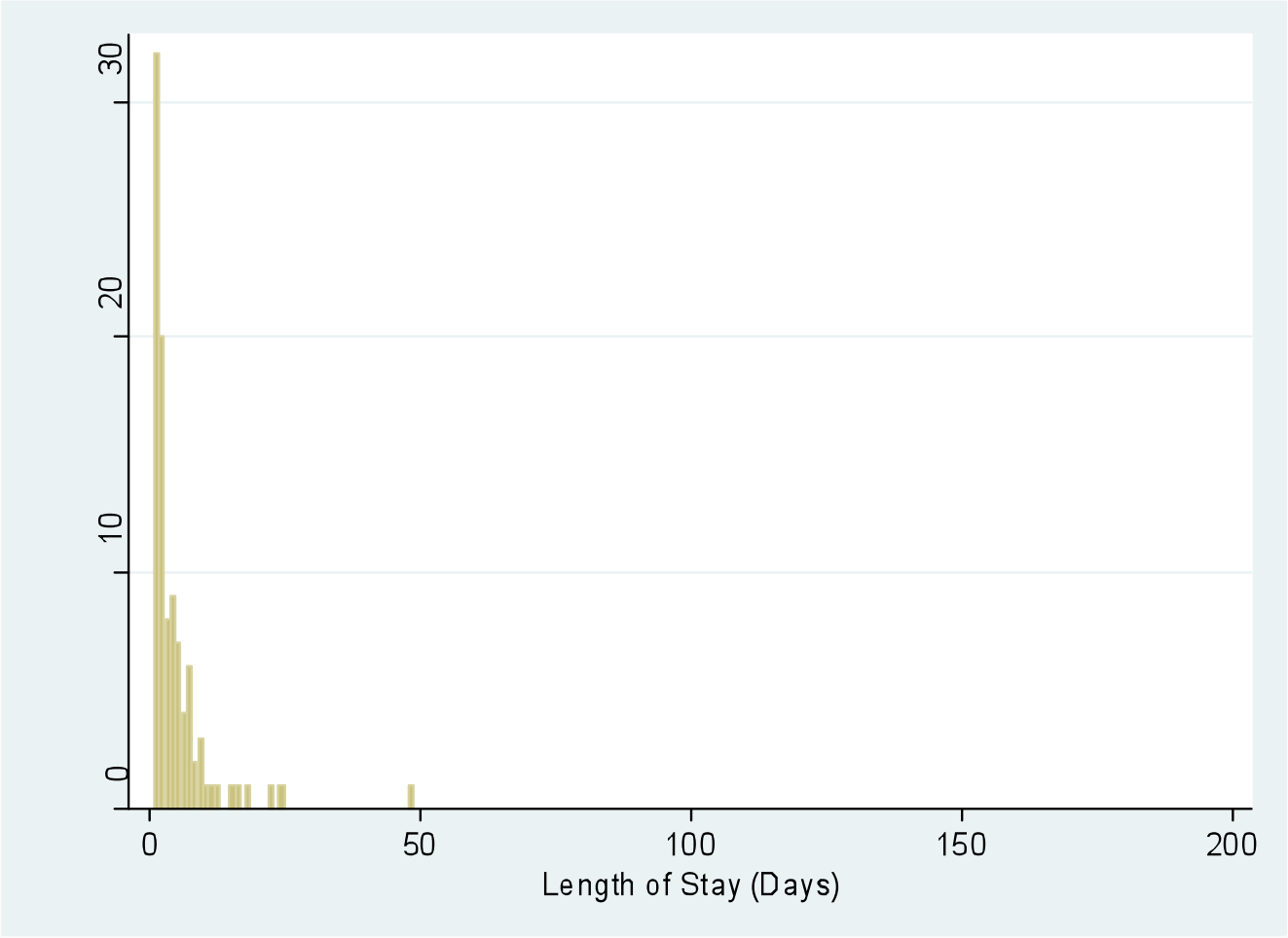


The Histogram and the Probability Density

- The *probability density* is a smooth idealized curve that shows the shape of the distribution in the population
- This is generally a theoretical distribution that we can never see: we can only estimate it from the distribution presented by a representative (random) sample from the population
- Areas in an interval under the curve represent the percentage of the population in the interval
- The distributions shown are indicative of a symmetric, bell shaped distribution for blood pressure measurements in men

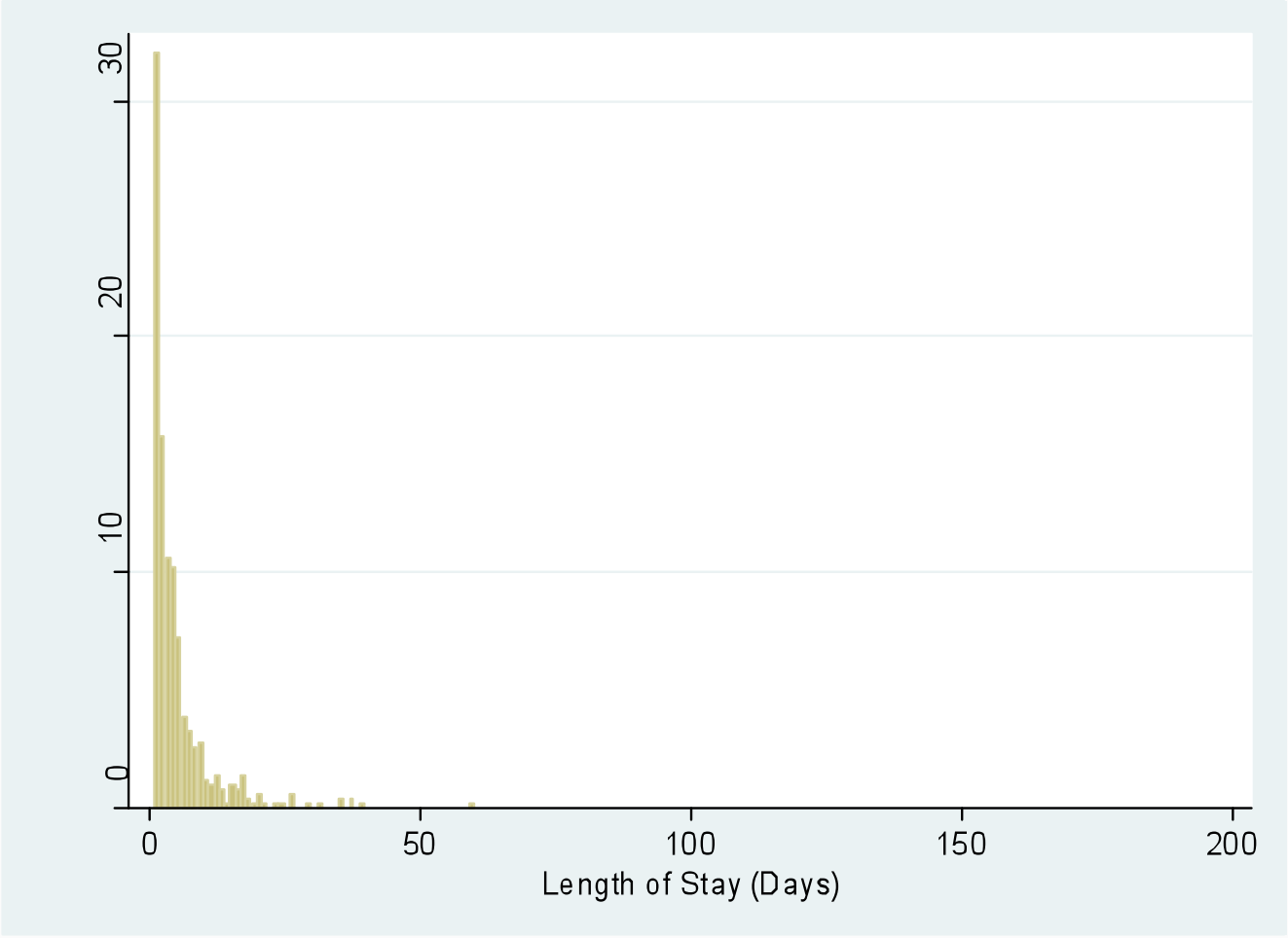
Example 2: Hospital Length of Stay

- Histogram of LOS values for 100 patients



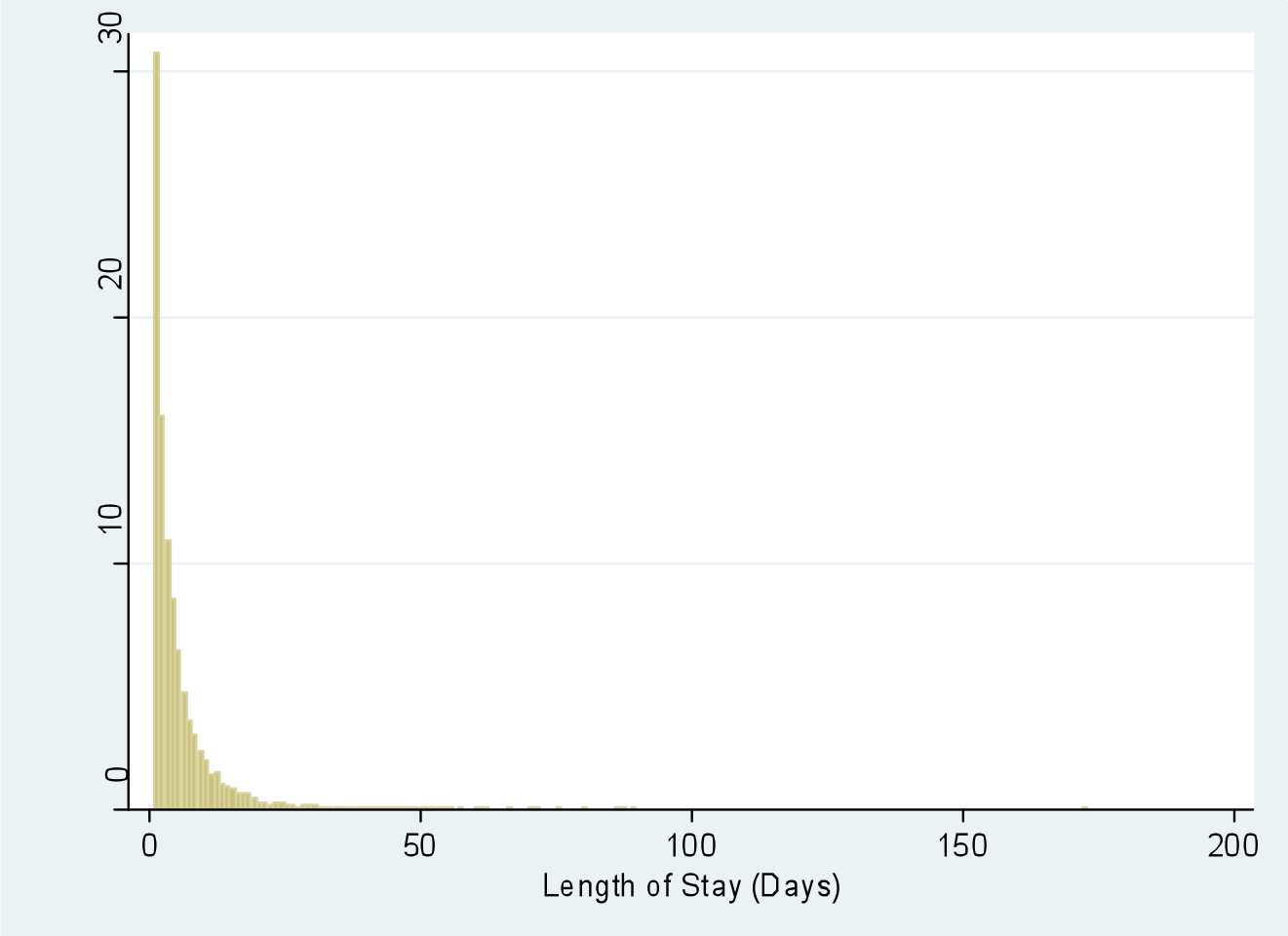
Example 2: Hospital Length of Stay

- Histogram of LOS values for 500 patients



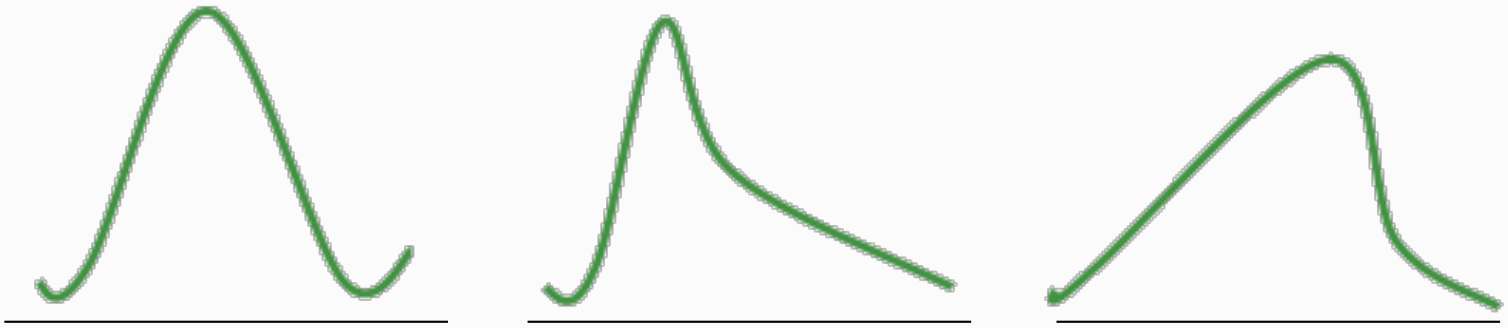
Example 2: Hospital Length of Stay

- Histogram of LOS values for all patients



Common Shapes of the Distribution

- Some shapes of data distributions



A

Symmetrical and bell shaped

B

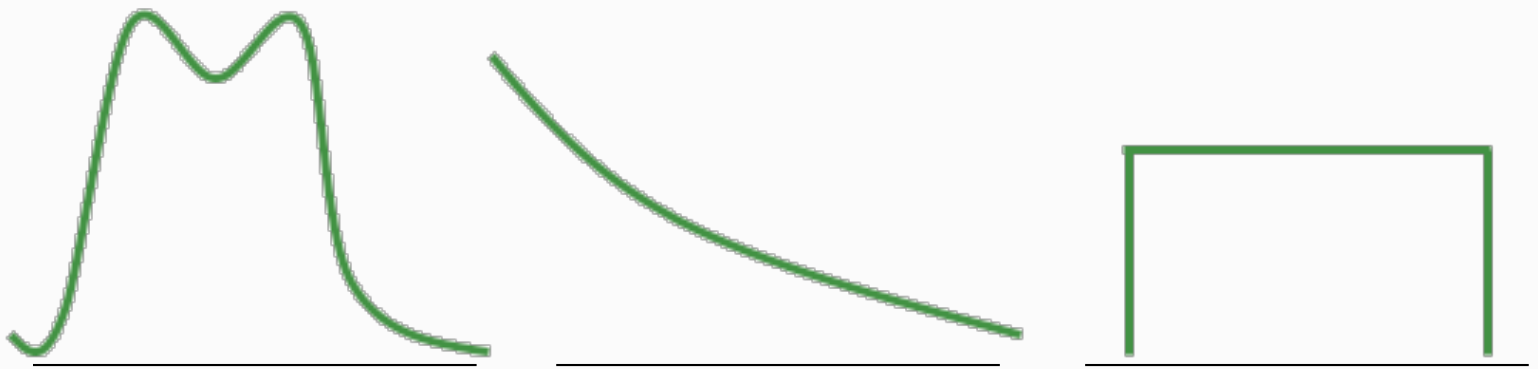
Positively skewed or skewed to the right

C

Negatively skewed or skewed to the left

Shapes of the Distribution

- Some possible shapes for frequency distributions



A

Bimodal

B

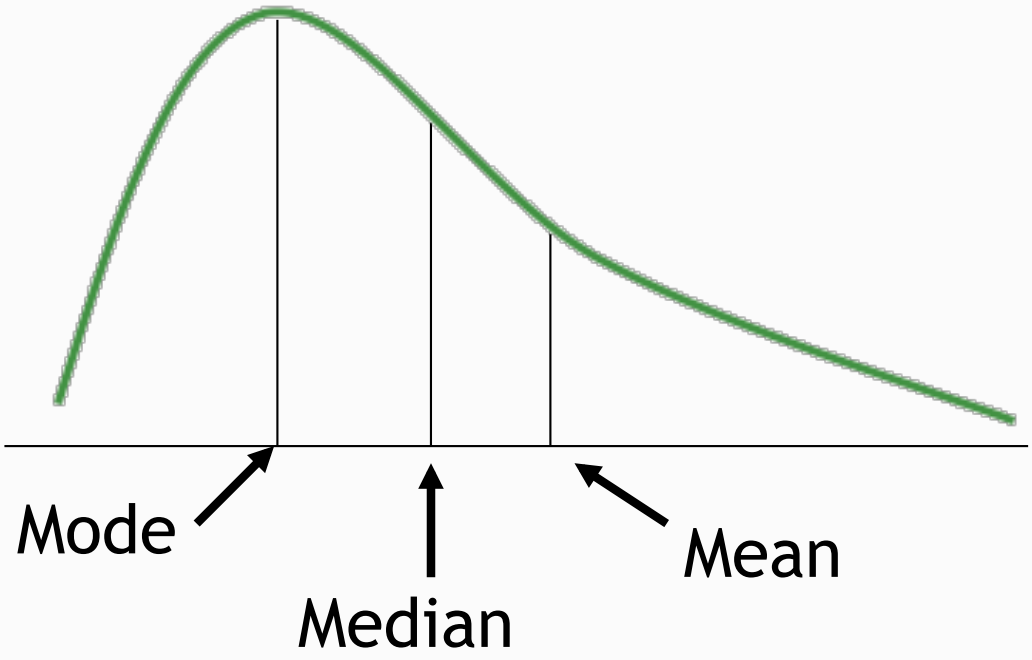
Reverse
J-shaped

C

Uniform

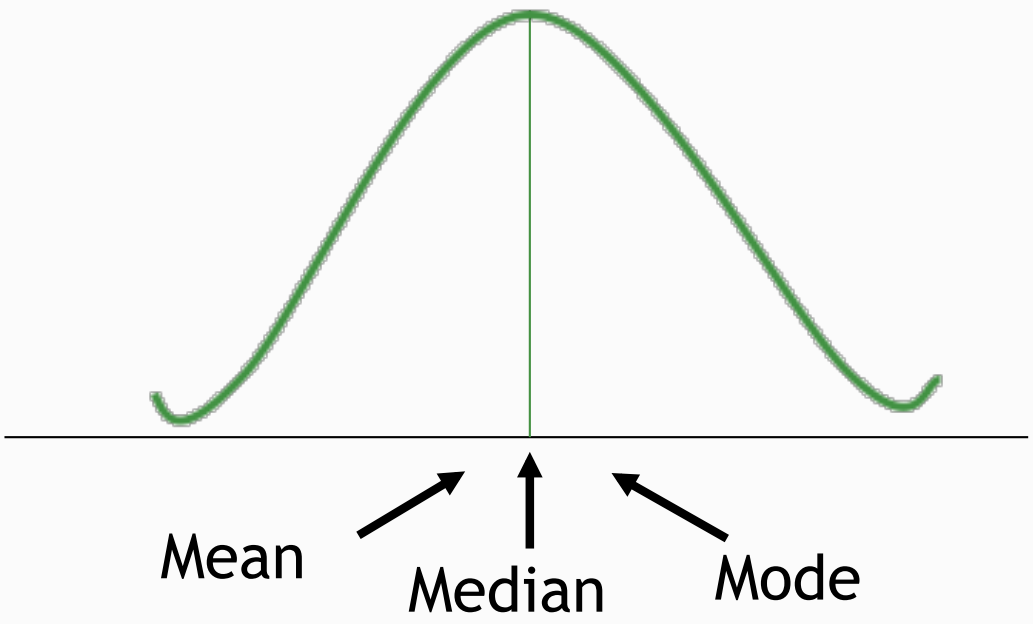
Distribution Characteristics

- Mode: Peak(s)
- Median: Equal areas point
- Mean: Balancing point



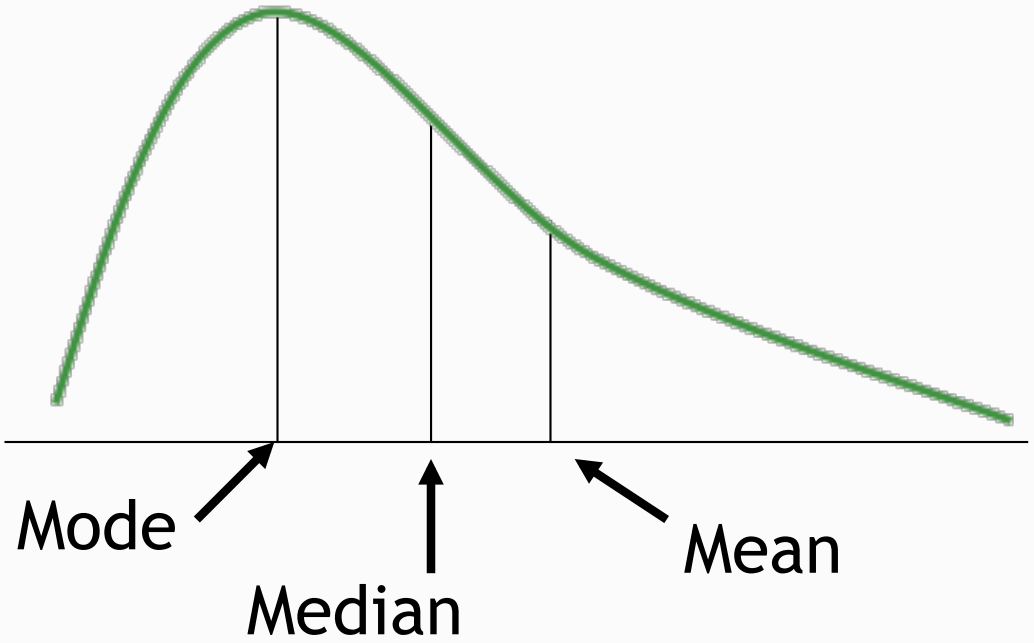
Shapes of Distributions

- *Symmetric* (right and left sides are mirror images)
  - Left tail looks like right tail
  - Mean = Median = Mode



Shapes of Distributions

- *Right skewed* (positively skewed)
  - Long right tail
  - Mean > Median



Shapes of Distributions

- *Left skewed* (negatively skewed)
  - Long left tail
  - Mean < Median