Populations, samples, and processes

Prof. Shou-de Lin

CSIE, NTU

sdlin@csie.ntu.edu.tw

Definitions

Population:

- a well-defined collection of objects of interest in the study. It can be a real population (e.g. all students in this class) or a virtual population (e.g. number of ideas per day).

Sample:

- a subset of population selected in some prescribed manner. (e.g. the female students in the class)
- Sampling: a process of identifying a sample

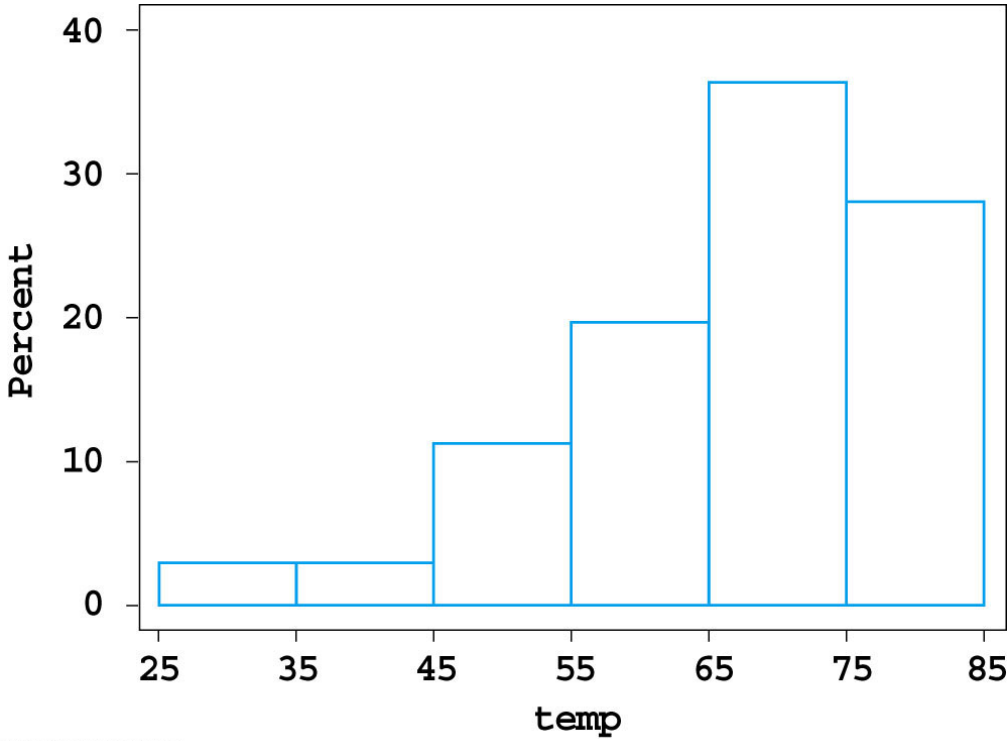
Variable:

- a characteristic of the objects. It can be a numeric measurement (e.g. the age) or a category measurement (e.g. the city you come from)
- One can make observation on data based on one or more variables. It is called
  - univariate if there is only one variable
  - bivariate if there are two variables
  - multivariate when observations are made on more than one variables. (e.g. age and height)

Example: the Challenger

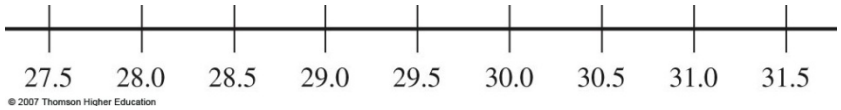
- The observation of O-ring temperature for each test firing of Challenger's engine:
  - 84, 49, 61, 40, 83…
- We can use the histogram to display the data

傅子芸



Histogram

- Def: a bar plot approximating the density of discrete or continuous numerical data.
- How to construct the histogram for discrete data?
  - First, determine the frequency and relative frequency of each x value.
  - Then mark possible x values on horizontal scale.
  - Draw a rectangle whose height is the relative frequency
- How to construct a histogram for continuous data?
  - Quantify the horizontal axis.
  - For each slot, determine the relative frequency.
  - Draw a rectangle whose height is the relative frequency



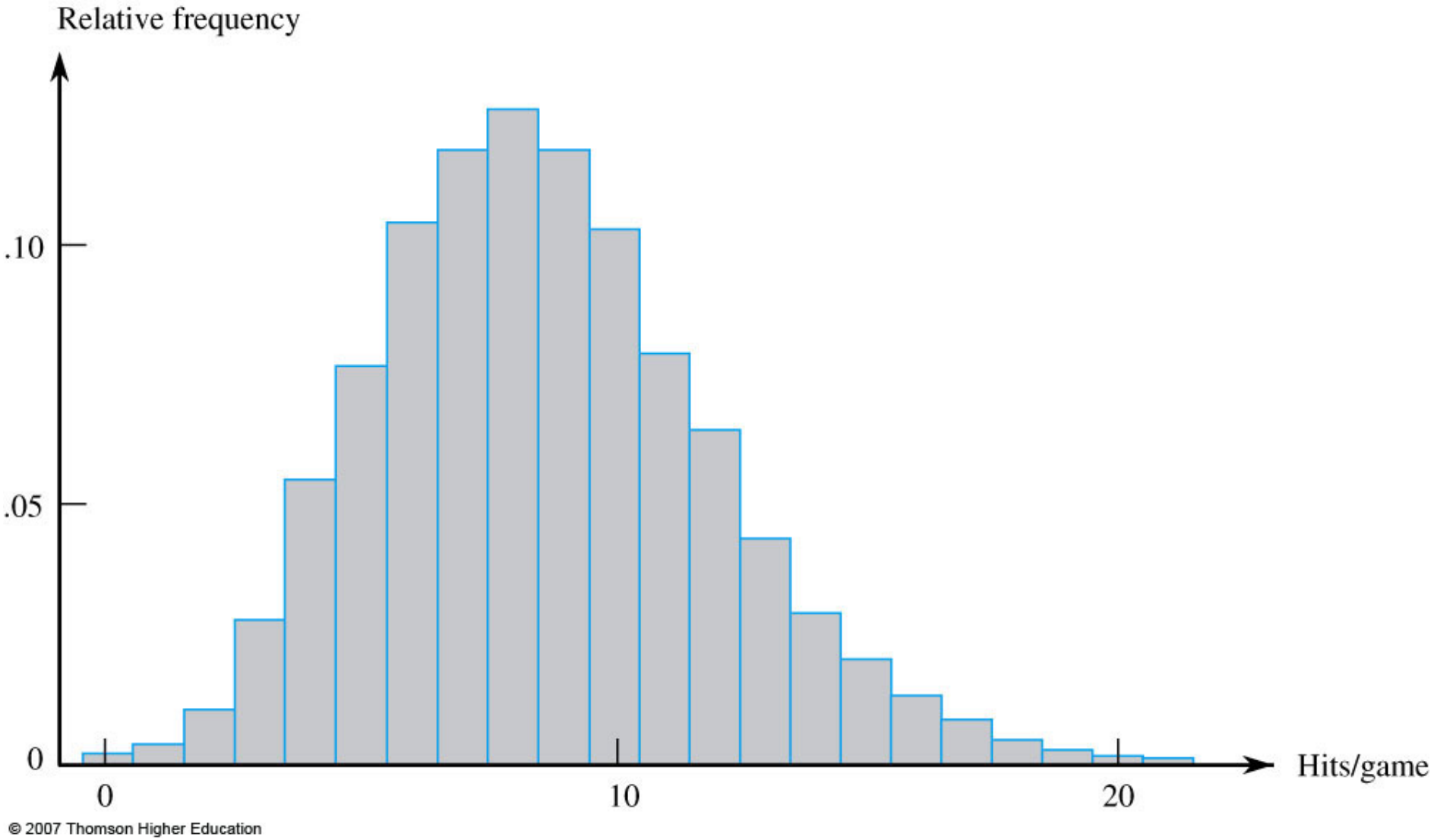
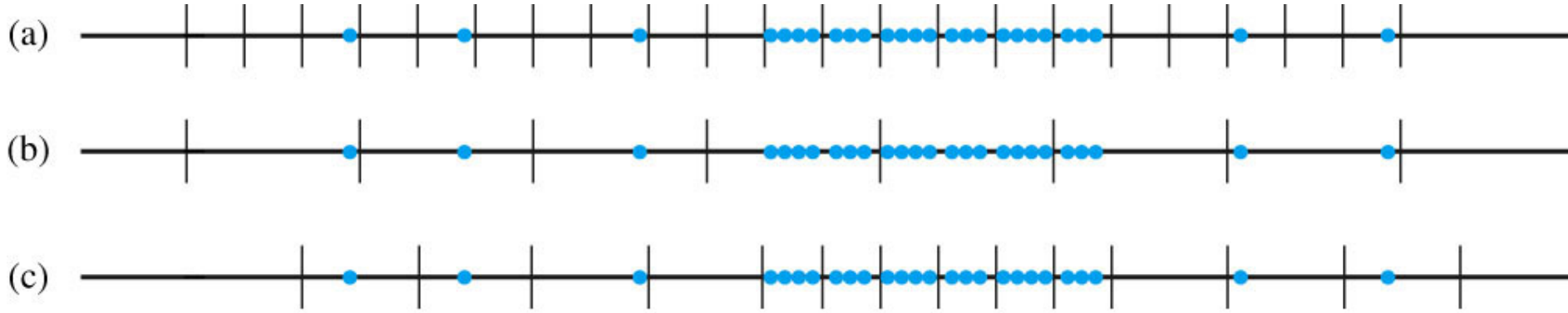
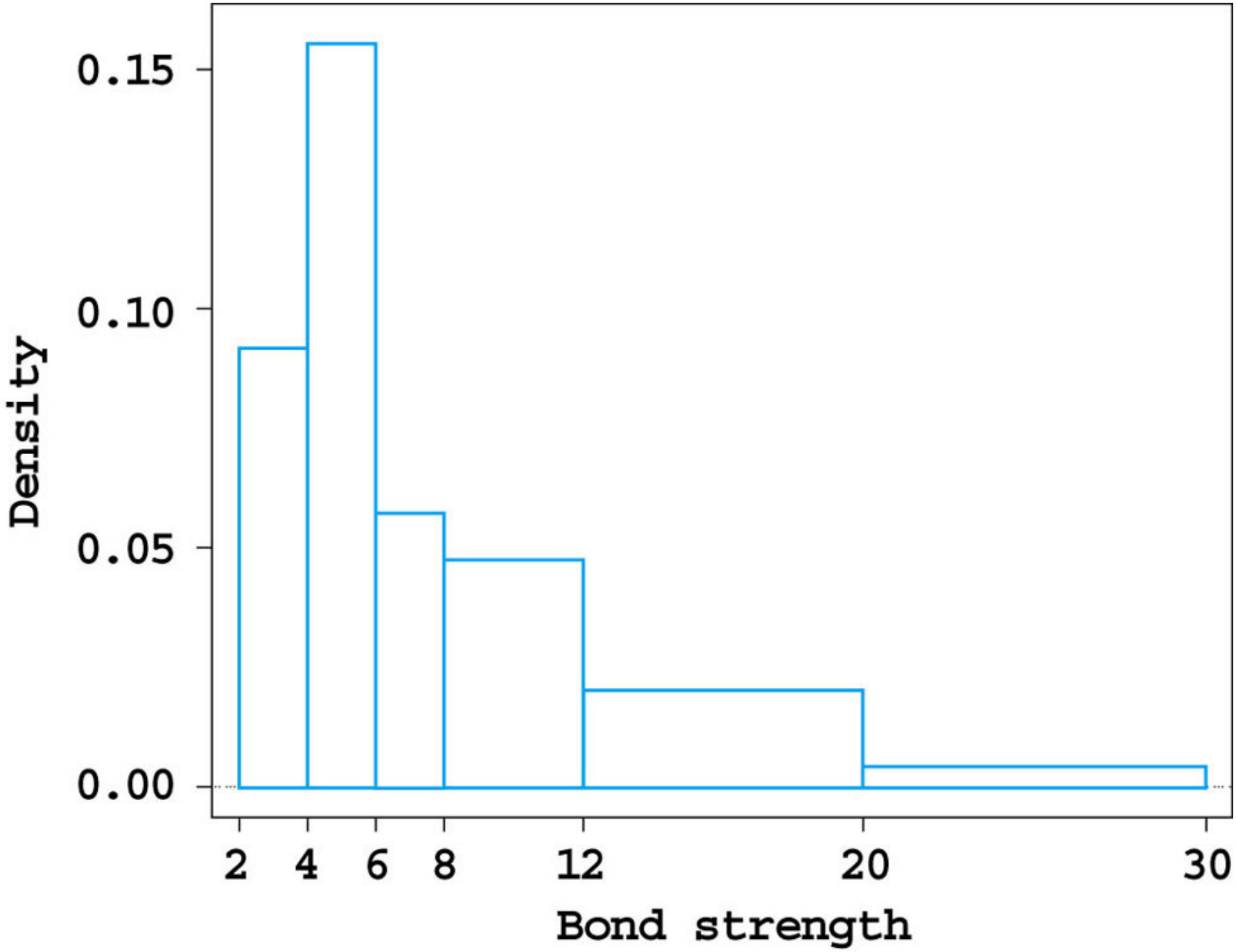


Fig. 1-7, p. 15

Constructing Histogram for Continuous Data: Unequal class Widths

- Equal-width classes may not be a sensible choice if a data set “stretches out” to one side or the other.
- It is better to use unequal class widths
  - Rectangle height=relative frequency/width





Histogram Shapes

- (a) Symmetric unimodal
- (b) bimodal
- (c) positively skewed
- (d) negatively skewed