Introduction to Probability and Statistics Fourteenth Edition

William Mendenhall, III • Robert J. Beaver • Barbara M. Beaver

Introduction to Probability and Statistics Fourteenth Edition

Chapter 1
Describing Data with Graphs

Variables and Data

- A **variable** is a characteristic that changes or varies over time and/or for different individuals or objects under consideration.
- **Examples:** Hair color, white blood cell count, time to failure of a computer component.

Definitions

- An **experimental unit** is the individual or object on which a variable is measured.
- A **measurement** results when a variable is actually measured on an experimental unit.
- A set of measurements, called **data**, can be either a **sample** or a **population**.

Example

- **Variable**
  - Hair color
- **Experimental unit**
  - Person
- **Typical Measurements**
  - Brown, black, blonde, etc.

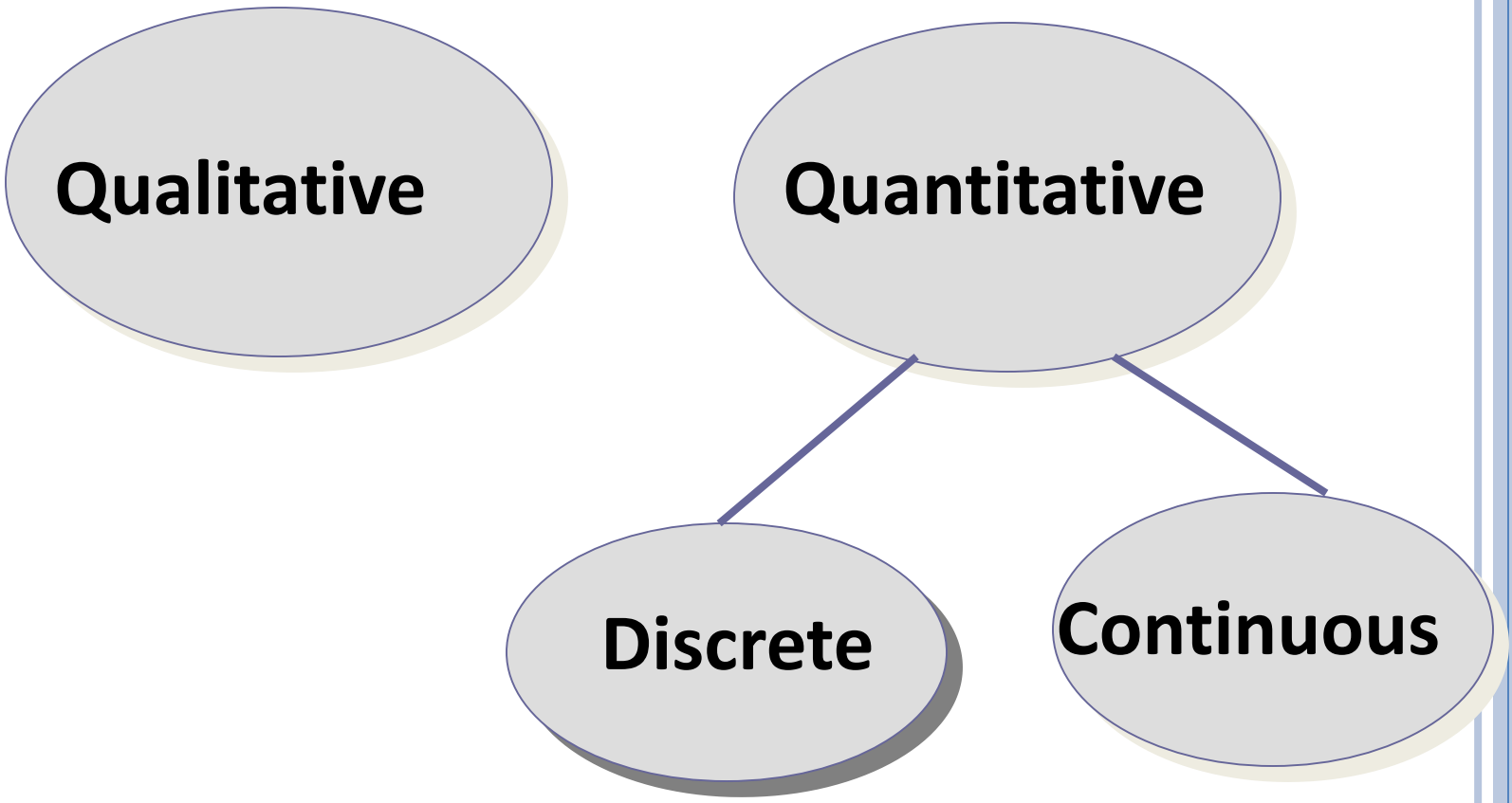
Example

- **Variable**
  - Blood Sugar Level
- **Experimental unit**
  - Individual
- **Typical Measurements**
  - 90, 100, 115 mg/dl, etc.

How many variables have you measured?

- **Univariate data:** One variable is measured on a single experimental unit.
- **Bivariate data:** Two variables are measured on a single experimental unit.
- **Multivariate data:** More than two variables are measured on a single experimental unit.

Types of Variables



Types of Variables

- **Qualitative variables** measure a quality or characteristic on each experimental unit.

- **Examples:**
  - Hair color (black, brown, blonde…)
  - Blood Group (A, B, AB, or O)
  - Gender (male, female)
  - City of birth (Amman, Irbid, Alkarak,…)

Types of Variables

- **Quantitative variables measure a numerical quantity on each experimental unit.**

  - ✓ **Discrete** if it can assume only a finite or countable number of values.

  - ✓ **Continuous** if it can assume the infinitely many values corresponding to the points on a line interval.

Examples

- **Quantitative discrete**

- For a particular day, the number of cars entering a college campus is measured.
- For a clinic in a hospital, the number of patients is recorded.

- **Quantitative continuous**

- Time under surgery.
- Patient Temperature.

Graphing Qualitative Variables

- Use a **data distribution** to describe:
  - **What values** of the variable have been measured
  - **How often** each value has occurred
- "How often" can be measured 3 ways:
  - Frequency
  - Relative frequency = Frequency/n
  - Percent = 100 x Relative frequency

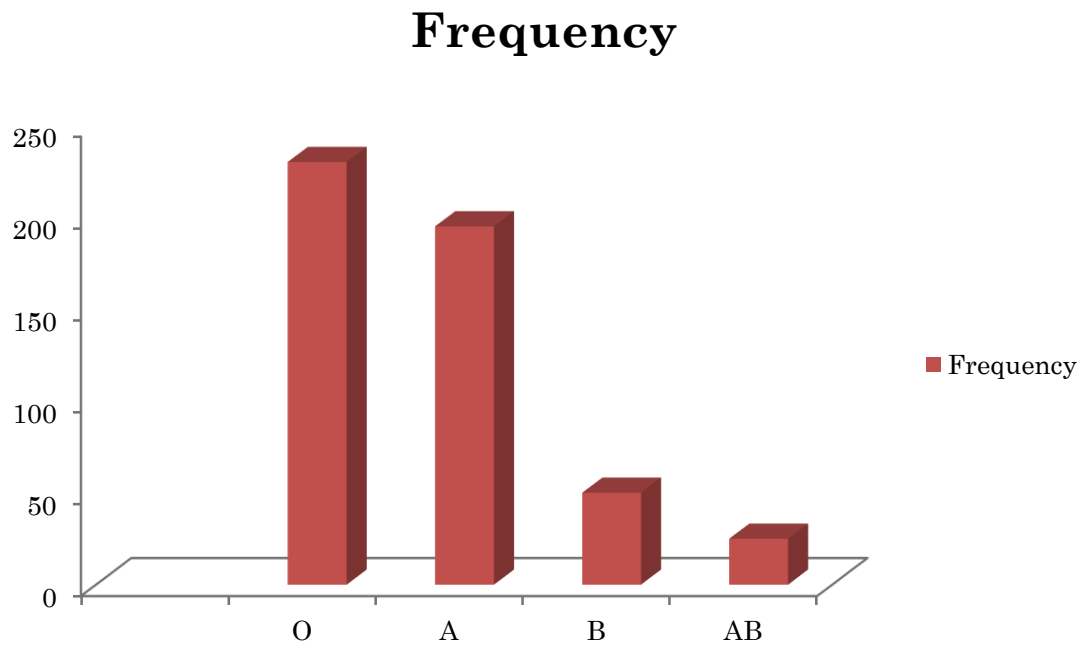
Example

- **Blood Group of 500 individuals**

- **Statistical Table:**

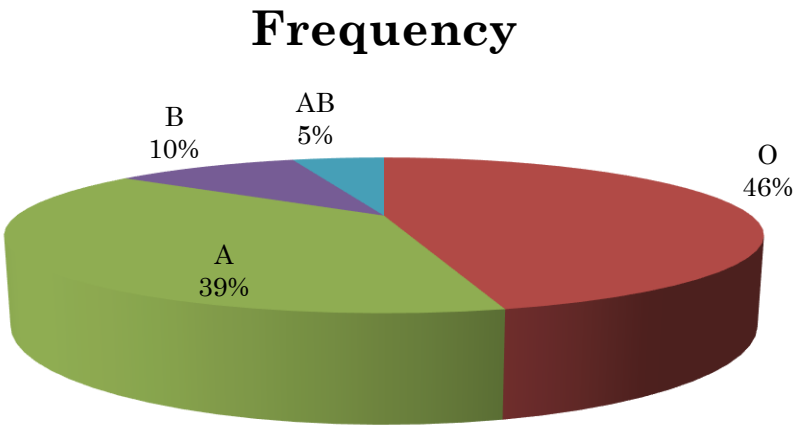
Blood Group	Frequency	Relative Frequency	Percent
O	230	0.46	46%
A	195	0.39	39%
B	50	0.10	10%
AB	25	0.05	5%
Total	500	1.00	100%

Graphs



Bar Chart

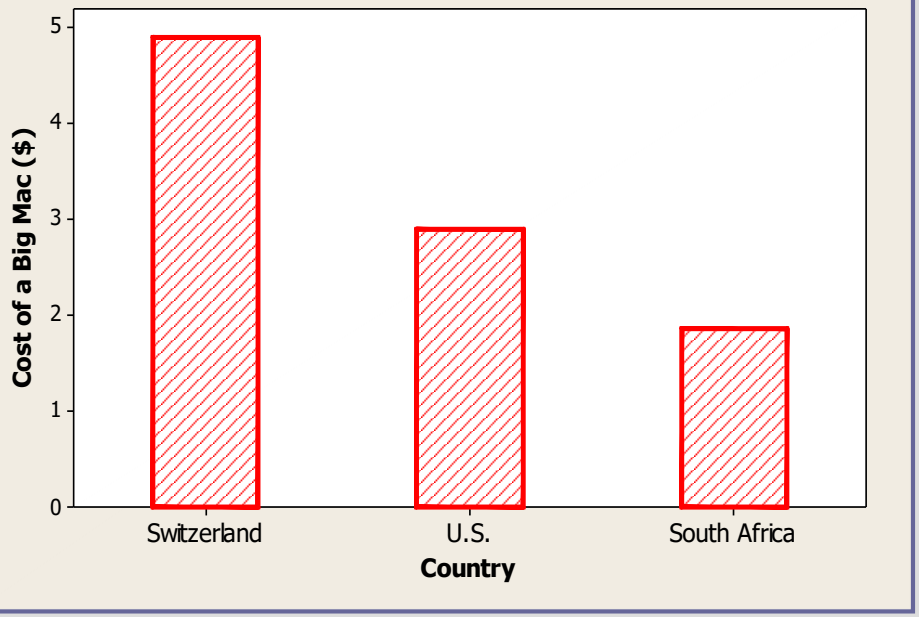
- Pie Chart



Graphing Quantitative Variables

- A single quantitative variable measured for different population segments or for different categories of classification can be graphed using a **pie** or **bar**

A Big Mac hamburger costs $4.90 in Switzerland, $2.90 in the U.S. and $1.86 in South Africa.

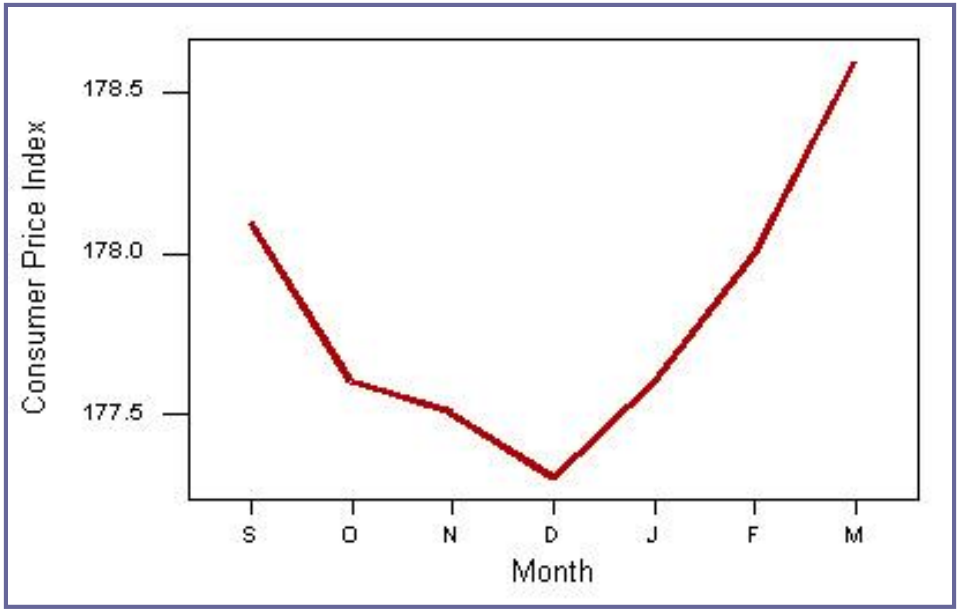


- A single quantitative variable measured over time is called a **time series.** It can be graphed using a **line** or **bar chart.**

Patients attending a health centre

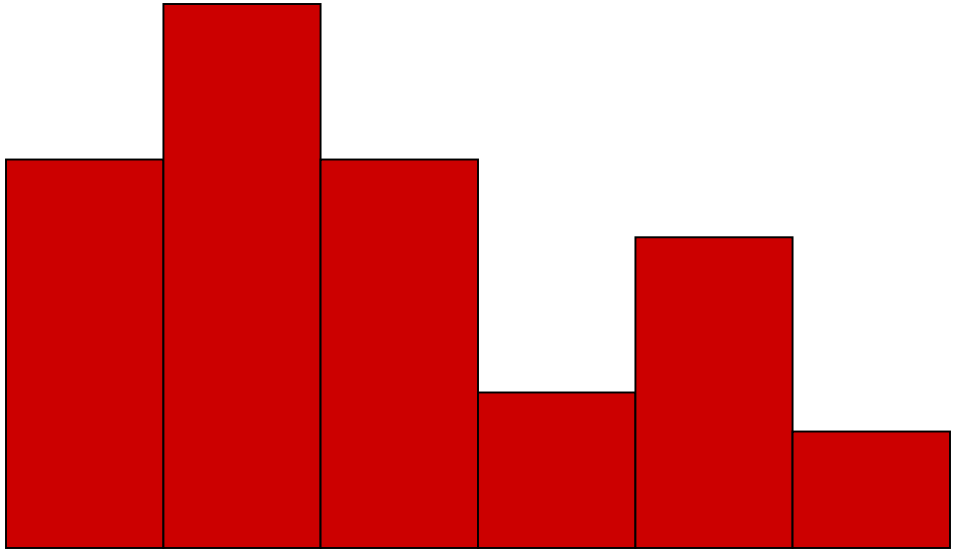
Jan	Feb	March	Apr	May	Jun	Jul
٨٥٠٠	٧٣٢٠	٦٥٠٠	٤٣٠٠	٣٢٠٠	٢٣٥٠	١٨٨٠

BUREAU OF LABOR STATISTICS



Relative Frequency Histograms

- A **relative frequency histogram** for a quantitative data set is a bar graph in which the height of the bar shows "how often" (measured as a proportion or relative frequency) measurements fall in a particular class or subinterval.



Relative Frequency Histograms

- Divide the range of the data into **5-12 subintervals** of equal length.
- Calculate the **approximate width** of the subinterval as Range/number of subintervals.
- Round the approximate width up to a convenient value.
- Use the method of **left inclusion** including the left endpoint, but not the right in your tally.
- Create a **statistical table** including the subintervals, their frequencies and relative frequencies.

Relative Frequency Histograms

- Draw the **relative frequency histogram** plotting the subintervals on the horizontal axis and the relative frequencies on the vertical axis.
- The height of the bar represents
  - The **proportion** of measurements falling in that class or subinterval.
  - The **probability** that a single measurement, drawn at random from the set, will belong to that class or subinterval.

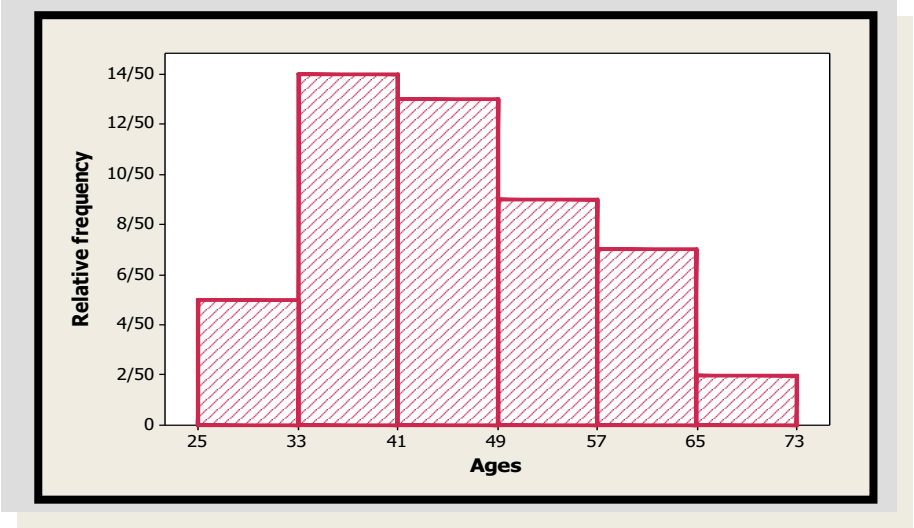
EXAMPLE

The ages of 50 tenured faculty at a university.

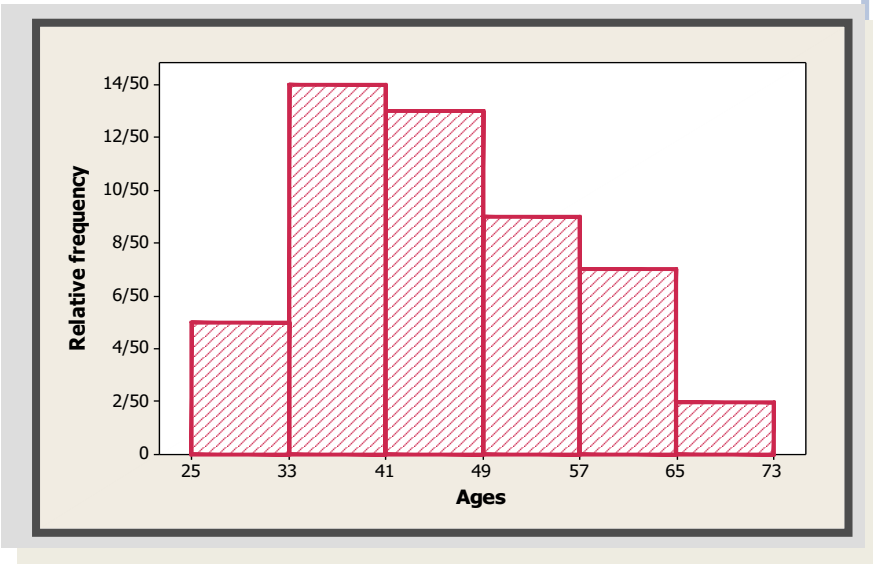
34	48	**70**	63	52	52	35	50	37	43	53	43	52	44
42	31	36	48	43	**26**	58	62	49	34	48	53	39	45
34	59	34	66	40	59	36	41	35	36	62	34	38	28
43	50	30	43	32	44	58	53						

- We choose to use **6** intervals.
- Minimum class width **= (70 – 26)/6 = 7.33**
- Convenient class width **= 8**
- Use **6** classes of length **8**, starting at **25.**

Age	Tally	Frequency	Relative Frequency	Percent
25 to < 33	1111	5	5/50 = .10	10%
33 to < 41	1111 1111 1111	14	14/50 = .28	28%
41 to < 49	1111 1111 111	13	13/50 = .26	26%
49 to < 57	1111 1111	9	9/50 = .18	18%
57 to < 65	1111 11	7	7/50 = .14	14%
65 to < 73	11	2	2/50 = .04	4%



Describing the Distribution



Shape? Skewed right

Outliers? No.

What proportion of the tenured faculty are younger than 41?

(14 + 5)/50 = 19/50 = .38

What is the probability that a randomly selected faculty member is 49 or older?

(8 + 7 + 2)/50 = 17/50 = .34

Key Concepts

I. How Data Are Generated

1. Experimental units, variables, measurements
2. Samples and populations
3. Univariate, bivariate, and multivariate data

II. Types of Variables

1. Qualitative or categorical
2. Quantitative
    a. Discrete
    b. Continuous

III. Graphs for Univariate Data Distributions

1. Qualitative or categorical data
    a. Pie charts
    b. Bar charts

Key Concepts

2. Quantitative data
    a. Pie and bar charts
    b. Line charts
    c. Relative frequency histograms
3. Describing data distributions
    a. Shapes—symmetric, skewed left, skewed right, unimodal, bimodal
    b. Proportion of measurements in certain intervals
    c. Outliers