

Lecture 8/Chapter 7

Part 2. Summarizing Data

Ch.7: Measurement Data

- Summaries
- Displaying with Stemplots
- Displaying with Histograms

Course Divided into Four Parts (Review)

1. Finding Data in Life (completed): scrutinizing origin of data
2. Finding Life in Data: summarizing data yourself or assessing another's summary
3. Understanding Uncertainty in Life: probability theory
4. Making Judgments from Surveys and Experiments: statistical inference

Definitions (Review)

- **Variable:** a characteristic that varies from one individual to another
- **Statistics:** the science of principles and procedures for gaining and processing **data** (info about variables' values for a sample) and using the info to draw general conclusions
- **Statistics:** summaries of data (such as a sample average or sample proportion)

Definitions

- Summarize values of a quantitative (measurement) variable by telling **center**, **spread**, **shape**.
- **Center:** measure of what is typical in the distribution of a quantitative variable
 - **Spread:** measure of how much the distribution's values vary
 - **Shape:** tells which values tend to be more or less common

Definitions

Measures of Center

■ **mean** = average = $\frac{\text{sum of values}}{\text{number of values}}$

■ **median**:

- *the* middle for **odd** number of values
- average of middle two for **even** number of values

■ **mode**: most common value

Measures of Spread

- **Range**: difference between highest & lowest
- **Standard deviation** (discussed later)

Example: Basic Summaries

□ **Background**: Cigarettes smoked in a day for

22 smoking students:

1 2 4 5 7 10 10 10 10 12 15
15 15 20 20 20 20 20 25 30

□ **Question**: How can we summarize the data?

□ **Response**:

1. center

- mean (average) =
- median = middle:
- mode (most common) =

Example: Basic Summaries

□ **Background**: Cigarettes smoked in a day for

22 smoking students:

1 2 4 5 7 10 10 10 10 12 15
15 15 20 20 20 20 20 25 30

□ **Question**: How can we summarize the data?

□ **Response**:

2. spread (variability): range is

3. shape:

Definitions for Shape

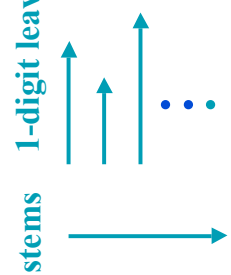
- **Symmetric distribution**: balanced on either side of center
- **Skewed distribution**: unbalanced (lopsided)
- **Skewed left**: has a few relatively low values
- **Skewed right**: has a few relatively high values
- **Outliers**: values noticeably far from the rest
- **Unimodal**: single-peaked
- **Normal**: a particular symmetric bell-shape

Displays of a Quantitative Variable

Displays help us see the shape of the distribution.

- **Stemplot**
 - Advantage: most detail
 - Disadvantage: impractical for large data sets
- **Histogram**
 - Advantage: works well for any size data set
 - Disadvantage: some detail lost
- **Boxplot**
 - Advantage: shows outliers, makes comparisons
 - Disadvantage: much detail lost

Definition

- **Stemplot:** vertical list of stems, each followed by horizontal list of one-digit leaves

- **Split stems:** If plot has too few stems, split into 2 (1st stem gets leaves 0-4, 2nd gets 5-9) or 5 (1st stem gets leaves 0-1, etc.) or 10.

Example: Basic Stemplot

- **Background:** *Cigarettes smoked in a day for 22 smoking students:*
1 2 4 5 7 10 10 10 10 12 15
15 15 20 20 20 20 20 20 25 30
- **Question:** Construct stemplot, describe shape?
- **Response:**

Example: Splitting Stems

- **Background:** *Earnings of 29 male students:*
0 2 2 3 3 3 3 4 4 5 5 5 5 5
6 6 6 6 7 8 8 10 10 10 12 15 20 25 42
- **Question:** Construct stemplot, describe shape?
- **Response:** start with 0 to 4 as stems:
0 0 2 2 etc. Almost all the values would appear in the first line, resulting in a poor display.
