

STA291



LECTURE 6

TUESDAY, 10 FEBRUARY 2008

Administrative

2

- Suggested problems from the textbook (not graded): 4.2, 4.3, and 4.4
- Check CengageNow for *second* online homework (section number “test”—do it, *please*.)
- Start bringing calculators (including labs—good to check skills!)

Where we've been ...

3

- Data types (scales of measurement, etc.)
- Sampling methods (~~good, bad, ugly~~ SRS, stratified, cluster versus convenience, volunteer)—why is one group good and the other bad?
- Graphical data summaries
- Order we've covered these topics are the same order we would deal with these issues in a real-world problem

Where next?

4

- **4 Numerical Descriptive Techniques**
 - Review:
 - **Parameter**
 - numerical characteristic of the **p**opulation
 - calculated using the whole population
 - **Statistic**
 - numerical characteristic of the **s**ample
 - calculated using the sample

4.1 Measures of Central Location

5

- Also called Central *Tendency*
- “What is a typical measurement in the sample/population?”
 - Mean: Arithmetic average
 - Median: Midpoint of the observations when they are arranged in increasing order
 - Mode: Most frequent value

Mean (Average)

6

- Mean (or Average): Sum of measurements divided by the number of subjects
- Example: Observations 3,8,19,12

Mean =

Mathematical Notation: Sample Mean

7

- Sample size n
- Observations x_1, x_2, \dots, x_n
- Sample Mean “ x -bar”

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Σ = SUM

Mathematical Notation: Population Mean

8

- Population size N
- Observations x_1, x_2, \dots, x_N
- Population mean μ (*mu*, read “myew”)

$$\mu = \frac{x_1 + x_2 + \dots + x_N}{N} = \frac{\sum_{i=1}^N x_i}{N}$$

Mean (Average)

9

- The mean requires numerical values. Only appropriate for quantitative data.
- It does not make sense to compute the mean for nominal variables.
- Example “Nationality” (nominal):
Germany = 1, Italy = 2, U.S. = 3, Norway = 4
Sample: Germany, Italy, Italy, U.S., and Norway
- Mean nationality = 2.4???

Mean (continued)

10

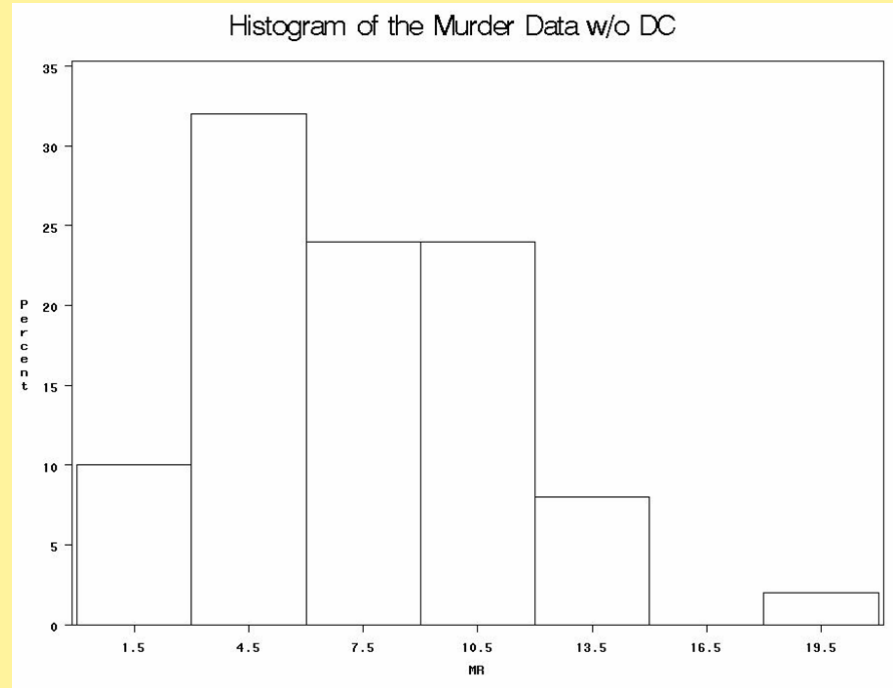
- Sometimes, the mean is calculated for ordinal variables, but this does not always make sense.
- Example “Weather” (on an ordinal scale):
Sun=1, Partly Cloudy=2, Cloudy=3,
Rain=4, Thunderstorm=5
 - Mean (average) weather=2.8
- Another example: “GPA = 3.8” is also a mean of observations measured on an ordinal scale

Mean(continued)

11

- The mean is highly influenced by outliers. That is, data points that are far from the rest of the data.

Example: Murder rates



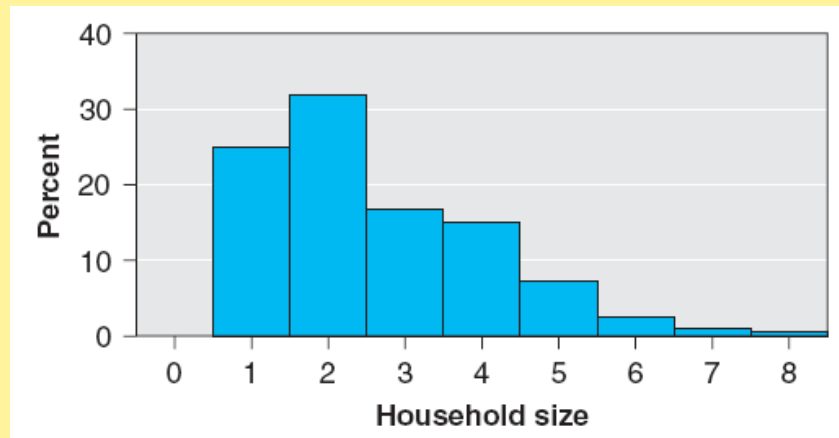
Mean (continued)

12

- Example: Murder Rate Data

Mean incl. DC: 8.73

Mean w/o DC: 7.33



- Any right-skewed distribution: the mean is “pulled” to the right

Central Location

13

- If the distribution is highly skewed, then the mean is not representative of a typical observation

- Example:

Monthly income for five persons

1,000 2,000 3,000 4,000 100,000

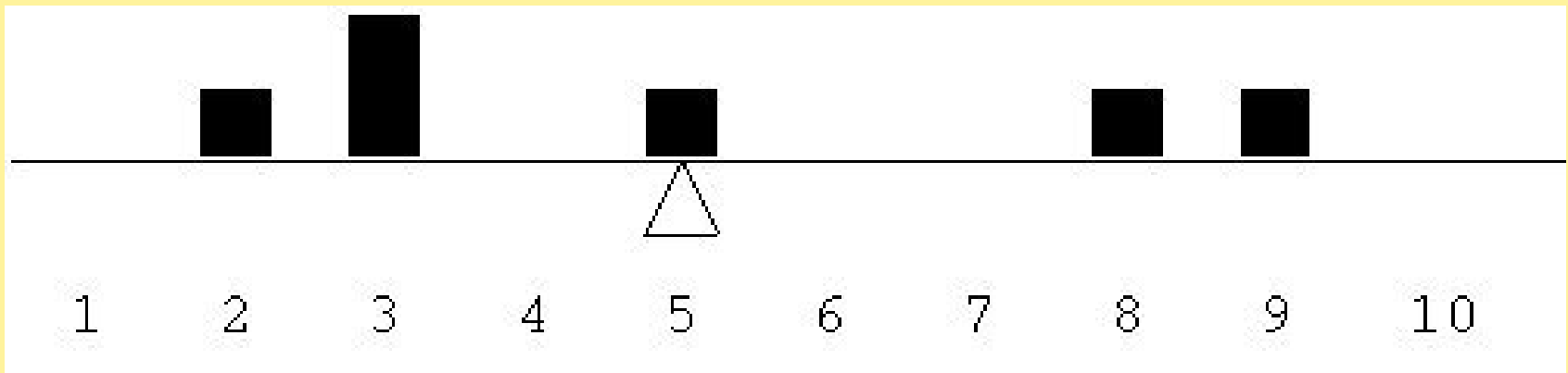
Average monthly income:

- Not representative of a typical observation.

Physical Interpretation of the Mean

14

- Assume that each measurement has the same “weight”



- Then, the mean is the center of gravity for the set of observations
- This is because the sum of the distances to the mean is the same for the observations above the mean as for the observations below the mean

Median

15

- The median is the measurement that falls in the middle of the ordered sample
- When the sample size n is odd, there is a middle value
- It has the ordered index $(n+1)/2$
- Example: 1.1, 2.3, 4.6, 7.9, 8.1

$$n=5, \quad (n+1)/2=6/2=3, \quad \text{Index} = 3$$

So,

Median = 3rd smallest observation = 4.6

Median

16

- When the sample size, n , is even, *average* the two middle values

- Example: 3, 4, 7, 10, 13, 19

$$n=6, \quad (n+1)/2=7/2=3.5, \quad \text{Index} = 3.5$$

Median = midpoint between 3rd and 4th
smallest observations = $(7+10)/2 = 8.5$

Mean and Median

17

- For skewed distributions, the median is often a more appropriate measure of central tendency than the mean
- The median usually better describes a “typical value” when the sample distribution is highly skewed

- Example:

Monthly income for five persons ($n = 5$)

1,000 2,000 3,000 4,000 100,000

- Median monthly income: 3000

Mean and Median

18

- **Example: Murder Rate Data**
- **Mean including DC: 8.73**
Mean without DC: 7.33
- **Median including DC: 6.8**
Median without DC: 6.7

Mean and Median

19

• Example: Keeneland Sales

Fillies Rule on Tuesday at Keeneland

September 16, 2008

[email this article](#)

[bookmark this article](#)

Fillies by Indian Charlie and Empire Maker topped Tuesday's session of Keeneland's September Yearling Sale.

John Brocklebank, as agent, went to \$250,000 to purchase a filly by Indian Charlie out of the stakes-winning Dehere mare Her She Kisses. Consigned by Mill Ridge Sales, agent, the filly is from the family of graded stakes winners Crafty Shaw, Shawklit Mint, and Mr. Shawklit.

A filly by Empire Maker out of Grade 3 Violet Handicap winner Changing World, by Spinning World, brought a final bid of \$230,000 from Ken and Sarah Ramsey. The filly was consigned by The Acorn LLC, agent for White Oaks (Mr. and Mrs. Samuel H. Rogers Jr.).

Gross receipts for Tuesday totaled \$14,116,400, down 11.6 percent from the \$15,969,400 posted last year. The session average of \$52,283 was down 6.4 percent from \$55,837 recorded in 2007, while the median of \$40,000 remained the same.

Cumulative gross sales for the eight days totaled \$295,453,300, down 13.1 percent from \$340,060,600 in 2007. Average was down 12.4 percent from \$171,488 to \$150,205, while the median price of \$95,000 was down 5 percent from last year's \$100,000.

Mean and Median

20

- Is there a compromise between the median and the mean? Yes!
- Trimmed mean:
 1. Order the data from smallest to largest
 2. Delete a selected number of values from each end of the ordered list
 3. Find the mean of the remaining values
- The trimming percentage is the percentage of values that have been deleted from each end of the ordered list.

Mode

21

- Mode of a data set is the most frequently occurring value
- Can speak of a data set being *unimodal*, *bimodal*, or *multimodal*
- Can be calculated on nominal (!) data
- On a histogram, where would the mode be?

Summary: Measures of Location

22

Mean- Arithmetic Average

{ Mean of a Sample - \bar{x}
Mean of a Population - μ

Can be calculated only on quantitative data

Notation: Subscripted variables
 n = # of units in the sample
 N = # of units in the population
 x = Variable to be measured
 x_i = Measurement of the i th unit

Median – Midpoint of the observations when they are arranged in increasing order

Can be calculated on *quantitative or ordinal* data

Mode- Most frequent value.

Can be calculated on quantitative, ordinal, or nominal data!

Attendance Survey Question #6



- On an index card
 - Please write down your name and section number
 - Today's Questions: