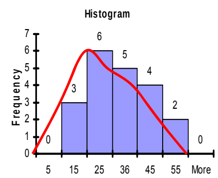
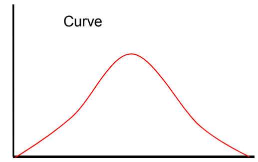
Describing Data: Numerical

Presented by:
Mahendra AN

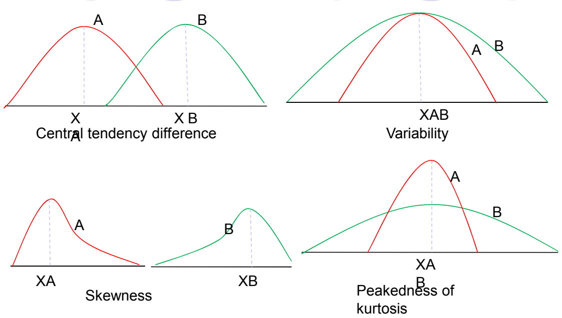
Sources:
http://business.clayton.edu/arjomand/business/stat%20presentations/busa3101.html
http://wweb.uta.edu/insyopma/baker/Courses.html
Anderson, Sweeney,wiliams, ***Statistics for Business and Economics*** , 10 e, Thomson, 2008
Djarwanto, **Statistik Sosial Ekonomi** Bagian pertama edisi 3, BPFE, 2001

Frequency distribution





Frequency Distribution Basic Characteristics



Chapter 4 (Part A)

Descriptive Statistics: Numerical Measures

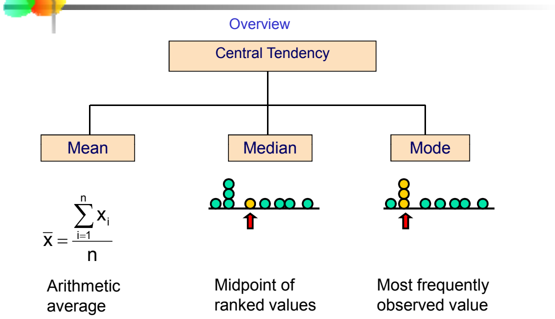
- Measures of Location
- Measures of Variability

Numerical Data Properties

- **Central Tendency**
  - Mean
  - Median
  - Mode
  - Midrange
  - Midhinge
- **Variation**
  - Range
  - Interquartile Range
  - Variance
  - Standard Deviation
  - Coeff. of Variation
- **Shape**
  - Skew
  - Kurtosis

Slide 4

Measures of Central Tendency



Statistics for Business and Economics, 6e © 2007 Pearson Education, Inc. Chap 3-5

Notation...

When referring to the number of observations in a ***population***, we use uppercase letter **N**

When referring to the number of observations in a ***sample***, we use lower case letter **n**

The arithmetic mean for a ***population*** is denoted with Greek letter "mu": μ

The arithmetic mean for a ***sample*** is denoted with an "x-bar": $\bar{x}$

Copyright © 2005 Brooks/Cole, a division of Thomson Learning, Inc. 4.6

Arithmetic Mean

- The arithmetic mean (mean) is the most common measure of central tendency
  - For a population of N values:

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

Population values; Population size

  - For a sample of size n:

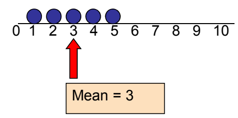
$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Observed values; Sample size

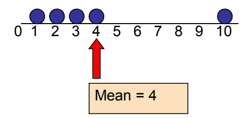
Arithmetic Mean

(continued)

- The most common measure of central tendency
- Mean = sum of values divided by the number of values
- Affected by extreme values (outliers)



$$\frac{1+2+3+4+5}{5} = \frac{15}{5} = 3$$



$$\frac{1+2+3+4+10}{5} = \frac{20}{5} = 4$$

Statistics is a pattern language...

	Population	Sample
Size	N	n
Mean	$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

Arithmetic Mean for Grouped Data

- Long method

$$\bar{X} = \frac{\sum fm}{n}$$

$\bar{X}$ = mean
U = population mean
$\sum fm$ = total freq. x mid point
n = sample freq.

- Short method

$$\bar{X} = \bar{X}o + \left(\frac{\sum fd'}{n}\right)i$$

$\bar{X}$ = mean
$\bar{X}o$ = estimated mean
$\sum fd'$ = total freq x standard devitiation
n = Total freq.
i = calss range

$$d' = \frac{m - \bar{X}o}{i}$$

m = calss mid point

Weight-Average Mean

- Giving weight for each data to make the data smoother
- This technique is used to fixing heteroscedacity problem
- Need a rationalization to decide the "weight"

$$\bar{X} = \frac{\sum XW}{\sum W}$$

W = weight

Geometric Mean

- Base on all of observation value
- Only for positive values
- GM is used if growth is counted in mean

$$G = \sqrt[n]{(X1)(X2)(X3)\ldots(Xn)}$$

Or

$$LogG = \frac{\sum logX}{n}$$

$$G = antilog \frac{\sum logX}{n}$$

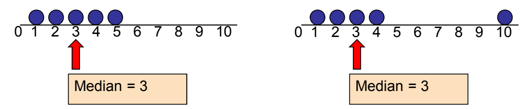
Harmonic Mean

- Base on all of observation value
- Cannot be used if one of the variables have zero value
- HM is used if ratio and the numerator have same value

$$H = \frac{n}{\sum \frac{1}{X}}$$

Median

- In an ordered list, the median is the "middle" number (50% above, 50% below)



- Not affected by extreme values

Finding the Median

- The location of the median:

Median position = $\frac{n+1}{2}$ position in the ordered data

- If the number of values is odd, the median is the middle number
- If the number of values is even, the median is the average of the two middle numbers

- Note that $\frac{n+1}{2}$ is not the *value* of the median, only the *position* of the median in the ranked data

Grouped Data Median

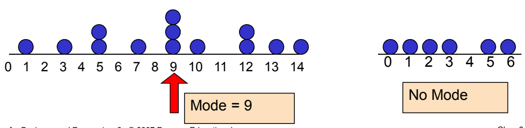
- GDMe is used for counting grouped data series

$$Md = L_{Md} + \left(\frac{\frac{n}{2} - F_{LMd}}{f_{Md}}\right)$$

Md = Median
L_{Md} = lowest limit
n = freq. number in distribution
f_{LMd} = comulative freq.
F_{Md} = class Freq.
i = class witdh

Mode

- A measure of central tendency
- Value that occurs most often
- Not affected by extreme values
- Used for either numerical or categorical data
- There may may be no mode
- There may be several modes



Grouped Data Mode

- GDMo is used for counting grouped data series
- Data present in freq. distribution

$$Mo = L_{Mo} + \left(\frac{d1}{d1 + d2}\right)$$

Mo = Mode
L_{Mo} = lowest limit of freq. that will be counted
$d1$ – highest ferq. – freq. before highest freq
$d2$ = highest ferq. – freq. after highest freq