STA291
Fall 2008

LECTURE 7

THURSDAY, 11 FEBRUARY

Chap. 4: Numerical Descriptive Techniques

4.1 Measures of Central Location (last time)

4.2 Measures of Variability (this time)

4.3 Measures of Relative Standing and Box Plots (next time?)

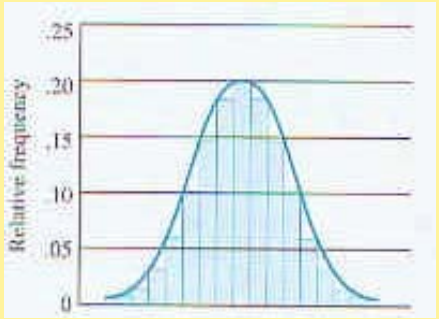
Homework and Suggested Study Material

- [10 points] Due Saturday, 27 September, 11pm

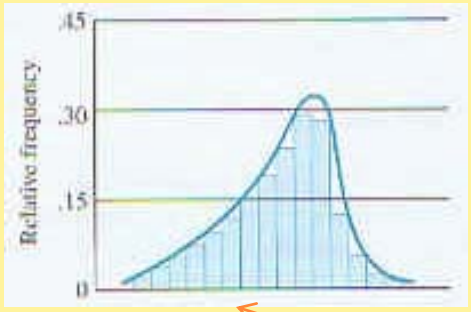
  Assignment HW3 on CengageNOW.
- Use the Study Tools at Thomson Now, click on our Courseware Book, and work through "Chapter 4 – Numerical Descriptive Techniques". (Pre-test, study plan, and post-test)
- Suggested problems from the textbook:

  4.20, 4.23, 4.24, 4.25, 4.40
- If you are interested in Finance, please read p.96 and try problem 4.8. Feel free to ask about it in lab/office hour.

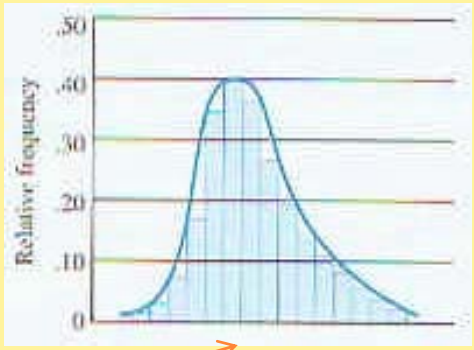
Review: Shapes of Distributions



Symmetric Distribution



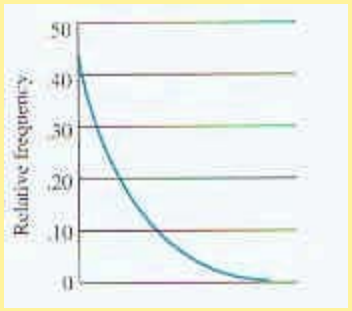
Skewed to the left



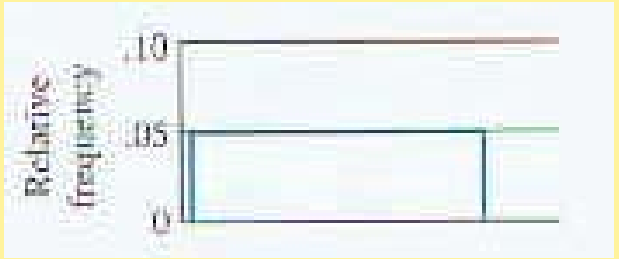
Skewed to the right

Examples

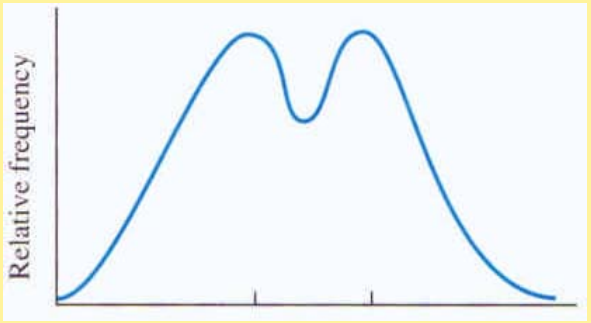
Exponential Distribution



Uniform Distribution



Bimodal Distribution



Summarizing Data Numerically

- Center of the data
  - Mean
  - Median
  - Mode
- Dispersion of the data
  - Variance, Standard deviation
  - Interquartile range
  - Range

Measuring Central Tendency (review)

- "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 vs. Median vs. Mode

- The mean is sensitive to outliers, median and mode are not
- In general, the median is more appropriate for skewed data than the mean
- In some situations, the median may be too insensitive to changes in the data
- The mode may not be unique

Mean vs. Median vs. Mode

- Mean: Interval data with an approximately symmetric distribution

- Median: Interval or ordinal data

- Mode: All types of data

Mean and Median

- Example: For towns with population size 2500 to 4599 in the U.S. Northeast in 1994, the mean salary of chiefs of police was $37,527, and the median was $30,500.

- Does this suggest that the distribution of salary was skewed to the left, symmetric, or skewed to the right?

Mean, Median, Mode—Another Example

Response	Frequency	Relative Frequency
every day	969	
a few times a week	452	
once a week	261	
less than once a week	196	
Never	76	
TOTAL		

- Identify the mode
- Identify the median response
- Mean?

Percentiles

- The p^{th} *percentile* is a number such that *p%* of the observations take values below it, and (100-*p)%* take values above it
- 50th percentile = median
- 25th percentile = lower quartile = Q_1
- 75th percentile = upper quartile = Q_3

In general,

- L_p = (n+1)p/100th is the spot in the ordered list of observations to find the p^{th} percentile

Quartiles

- 25th percentile

  = lower quartile

  = *approximately* median of the observations below the median

- 75th percentile

  = upper quartile

  = *approximately* median of the observations above the median

Median and Quartiles can be found from a stem and leaf plot

- Example: Murder Rate Data (w/o DC—key: 20|3 = 20.3)

Stem	Leaf	#
20	3	1
19		
18		
17		
16		
15		
14		
13	135	3
12	7	1
11	334469	6
10	2234	4
9	08	2
8	03469	5
7	5	1
6	034689	6
5	0238	4
4	46	2
3	0144468999	10
2	039	3
1	67	2

A quarter of the states has murder rate above...

The median murder rate is...

A quarter of the states has murder rate below...

Five-Number Summary

- Maximum, Upper Quartile, Median, Lower Quartile, Minimum
- Statistical Software SAS output (Murder Rate Data)

Quantile	Estimate
100% Max	20.30
75% Q3	10.30
50% Median	6.70
25% Q1	3.90
0% Min	1.60

Note the distance from the median to the maximum compared to the median to the minimum.

Interquartile Range

- The Interquartile Range (IQR) is the difference between upper and lower quartile
- IQR = $Q_3 - Q_1$
- IQR= Range of values that contains the middle 50% of the data
- IQR increases as variability increases

Box Plot (AKA Box-and-Whiskers Plot)

- A box plot is basically a graphical version of the five-number summary (unless there are outliers)
- It consists of a ***box*** *that contains the central 50%*of the distribution (from lower quartile to upper quartile),
- A ***line*** *within the box that marks the median*,
- And ***whiskers*** *that extend to the maximum and* minimum values, unless there are outliers

Outliers

- An observation is an outlier if it falls
  - more than 1.5 IQR above the upper quartile or
  - more than 1.5 IQR below the lower quartile
- Example: Murder Rate Data w/o DC
  - upper quartile Q3 = 10.3
  - IQR = 6.4
  - Q3 + 1.5 IQR = _____
  - Any outliers?

Five-Number Summary/Box Plot

- On right-skewed distributions, minimum, Q_1, and median will be "bunched up", while Q_3 and the maximum will be farther away.
- For left-skewed distributions, the "mirror" is true: the maximum, Q_3, and the median will be relatively close compared to the corresponding distances to Q_1 and the minimum.
- Guess on symmetric distributions?

Attendance Survey Question 7

- ***On a your index card:***
  - Please write down your name and section number
  - Today's Question: