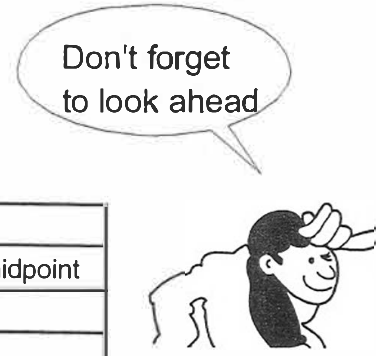
Chapter 5 Measuring Central Tendency of Grouped Data



I. Introduction

A. When actual data is unavailable or of an unmanageable volume, it may be necessary to determine parameters and statistics using a frequency distribution.

B. Important symbols:

Symbol	Definition	Symbol	Definition
$\bar{X}$	the sample mean	fx	frequency times the class midpoint
X	the midpoint of a class	Σfx	summation of fx
f	the frequency of a class	n	total frequency

II. The grouped sample mean

$$\bar{X} = \frac{\Sigma fx}{n}$$

A. Linda needs to estimate this year's tape rentals for a bank loan application. She will use the page 4 tape rentals summarized with a frequency distribution to estimate average daily rentals for the year.

B. Linda must calculate each class midpoint and then multiply it by the class frequency.

The midpoint formula is

$$X = \frac{X_1 + X_2}{2}$$

For class one

$$X = \frac{50 + 59}{2} = 54.5$$

$$\bar{X} = \frac{\Sigma fx}{n} = \frac{1{,}117.5}{15} = 74.5$$

Daily Rentals Beginning 1/2/98			
Stated Class Limits	Frequency (f)	x	fx
50 - 59	2.0	54.5	109.0
60 - 69	3.0	64.5	193.5
70 - 79	5.0	74.5	372.5
80 - 89	3.0	84.5	253.5
90 - 99	2.0	94.5	189.0
Totals	n = 15.0		Σfx = 1,117.5

Estimated yearly tape rentals would be (52)(7)(74.5) = 27,118.

III. The grouped median

A. The median is the middle number.

$$L + \frac{\frac{n}{2} - CF_b}{f}(i)$$

Symbols	Definitions
L	lower real limit of the median's class
CF_B	cumulative frequency before the median's frequency
i	class interval (width)

B. Use $\frac{n}{2}$ to determine the location of the middle frequency.

$$\frac{n}{2} = \frac{15}{2} = 7.5$$

C. Beginning at the top of the frequency distribution and counting down the frequency column reveals that the 7.5 frequency is located in the third class from the top (or bottom for that matter). The lower real limit of the median's class is 69.5 and the class is 10 wide.

Class Limits	Frequency	
50 - 59	2	Used 2 here
60 - 69	3	Used 3 here
70 - 79 (lower limit)	5	Need 2.5 from here to get to 7.5 Out of 5
80 - 89	3	
90 - 99	2	
	15	

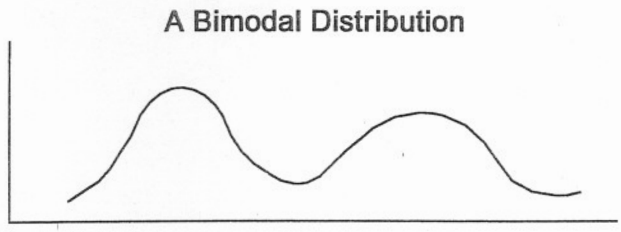
$$L + \frac{\frac{n}{2} - CF_b}{f}(i)$$

$$= 69.5 + \frac{\frac{15}{2} - 5}{5}(10)$$

$$= 69.5 + 5 = 74.5$$

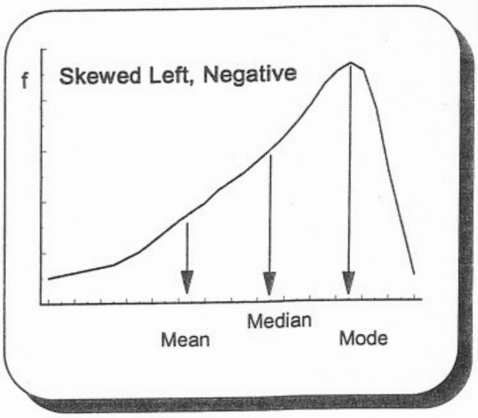
IV. The grouped mode

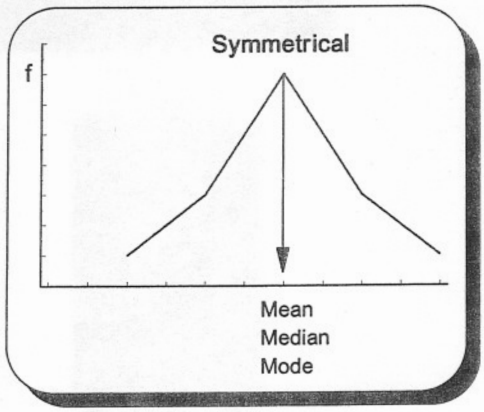
A. The grouped mode is the midpoint of the class with the highest frequency.
B. For the page 22 distribution, the mode is 74.5.
C. The mean, median, and mode are all 74.5 because this distribution is symmetrical (normal).
D. Frequency distributions with two peaks are said to be **bimodal**. More than two is **multimodal**.

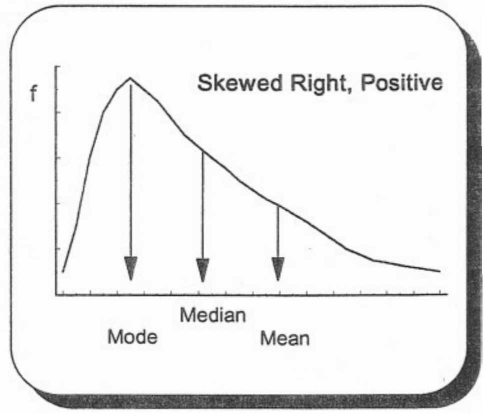


V. Nonsymmetrical distributions

A. Frequency distributions that are not symmetrical are said to be **skewed**.
   1. With negatively skewed data, the mean is the smallest of the three measures of central tendency.
   2. With positively skewed data, the mean is the largest of the three measures of central tendency.







B. **Measuring skewness**
   1. The degree to which a distribution (curve) is skewed is measured by **Pearson's coefficient of skewness.**
   2. The measure applies to both sample and population data.
   3. When data is positively skewed, the mean is larger than the median, and the measure is positive.
   4. When data is negatively skewed, the mean is smaller than the median, and the measure is negative.
   5. An increase in skewness increases the difference between the mean and the median. This causes an increase in the coefficient of skewness.
   6. Normal distributions have a zero coefficient of skewness.

$$\text{Skewness} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

$$= \frac{3(74.5 - 74.5)}{12.5} = 0$$

Note: The standard deviation of 12.5 was taken from page 28.

C. For highly skewed distributions, the median measures central tendency better than the mean because it is not as influenced by extreme values.
   1. Income is skewed right (positive) by a few people making a large amount of money.
   2. Comparing the mean and median salaries of these unionized workers yields interesting results.

$14,000
$15,000
→ **$16,000**
$17,000
$28,000

$$\mu = \frac{\sum x}{N} = \frac{\$90{,}000}{5} = \$18{,}000$$

The median is $16,000.

   3. In situations like this, a union would use the median salary to make the average look low. Management would use the mean salary to make the average look high.

Note: Suppose the top salary of $28,000 was increased to $48,000. The mean would increase from $18,000 to $22,000, but the median would remain unchanged.