Statistics 1
Chapter 4

Describing Data 3

Chapter Four

Describing Data: Measures of Dispersion

GOALS

When you have completed this chapter, you will be able to:

ONE

Compute and interpret the range, the mean deviation, the variance, and the standard deviation from raw data.

TWO

Compute and interpret the range, the variance, and the standard deviation from grouped data.

THREE

Explain the characteristics, uses, advantages, and disadvantages of each measure of dispersion.

Chapter Four *continued*

Describing Data: Measures of Dispersion

GOALS

When you have completed this chapter, you will be able to:

FOUR

Understand Chebyshev’s theorem and the Normal, or Empirical Rule, as they relate to a set of observations.

FIVE

Compute and interpret quartiles and the interquartile range.

SIX

Construct and interpret box plots

SEVEN

Compute and understand the coefficient of variation and the coefficient of skewness.

Mean Deviation

- Mean Deviation: The arithmetic mean of the absolute values of the deviations from the arithmetic mean.
-

$$MD = \frac{\Sigma \left| X - \bar{X} \right|}{n}$$

EXAMPLE 1

- The weights of a sample of crates containing books for the bookstore are (in lbs.) 103, 97, 101, 106, 103.
- $\bar{X} = 510/5 = 102$ lbs.
- $\Sigma = 1+5+1+4+1=12$
- MD = 12/5 = 2.4
- Typically, the weights of the crates are 2.4 lbs. from the mean weight of 102 lbs.

Population Variance

- The population variance for ungrouped data is the arithmetic mean of the squared deviations from the population mean.
- $$\sigma^2 = \frac{\Sigma(X-\mu)^2}{N}$$

EXAMPLE 2

- The ages of the Dunn family are 2, 18, 34, and 42 years. What is the population variance?

$$\mu = \Sigma X / N = 96 / 4 = 24$$

$$\sigma^2 = \Sigma(X - \mu)^2 / N = 944 / 4 = 236$$

Population Variance *continued*

- An alternative formula for the population variance is:

$$\sigma^2 = \frac{\Sigma X^2}{N} - \left(\frac{\Sigma X}{N}\right)^2$$

The Population Standard Deviation

- The population standard devia[redacted]on (   ) is the square root of the population variance.
- For EXAMPLE 2, the population standard deviation is 15.19 (square root of 230.81).

Sample Variance

- The sample variance estimates the population variance.

$$\text{Conceptual Formula} = S^2 = \frac{\Sigma(X - X)^2}{n - 1}$$

$$\text{Computational Formula} = \mathrm{S}^2 = \frac{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}{n - 1}$$

EXAMPLE 3

- A sample of five hourly wages for various jobs on campus is: $7, $5, $11, $8, $6. Find the variance.
- $\overline{X} = 37/5 = 7.40$
- $s^2 = 21.2/(5-1) = 5.3$

Sample Standard Deviation

- The sample standard deviation is the square root of the sample variance.
- In EXAMPLE 3, the sample standard deviation = 2.30

Measures of Dispersion: Ungrouped Data

- For ungrouped data, the range is the difference between the highest and lowest values in a set of data.
- RANGE = Highest Value - Lowest Value
- EXAMPLE 4: A sample of five accounting graduates revealed the following starting salaries: $22,000, $28,000, $31,000, $23,000, $24,000. The range is $31,000 - $22,000 = $9,000.

Sample Variance For Grouped Data

- The formula for the sample variance for grouped data used as an estimator of the population variance is:

$$S^2 = \frac{\Sigma fX^2 - \frac{(\Sigma fX)^2}{n}}{n-1}$$

- where f is class frequency and X is class midpoint.

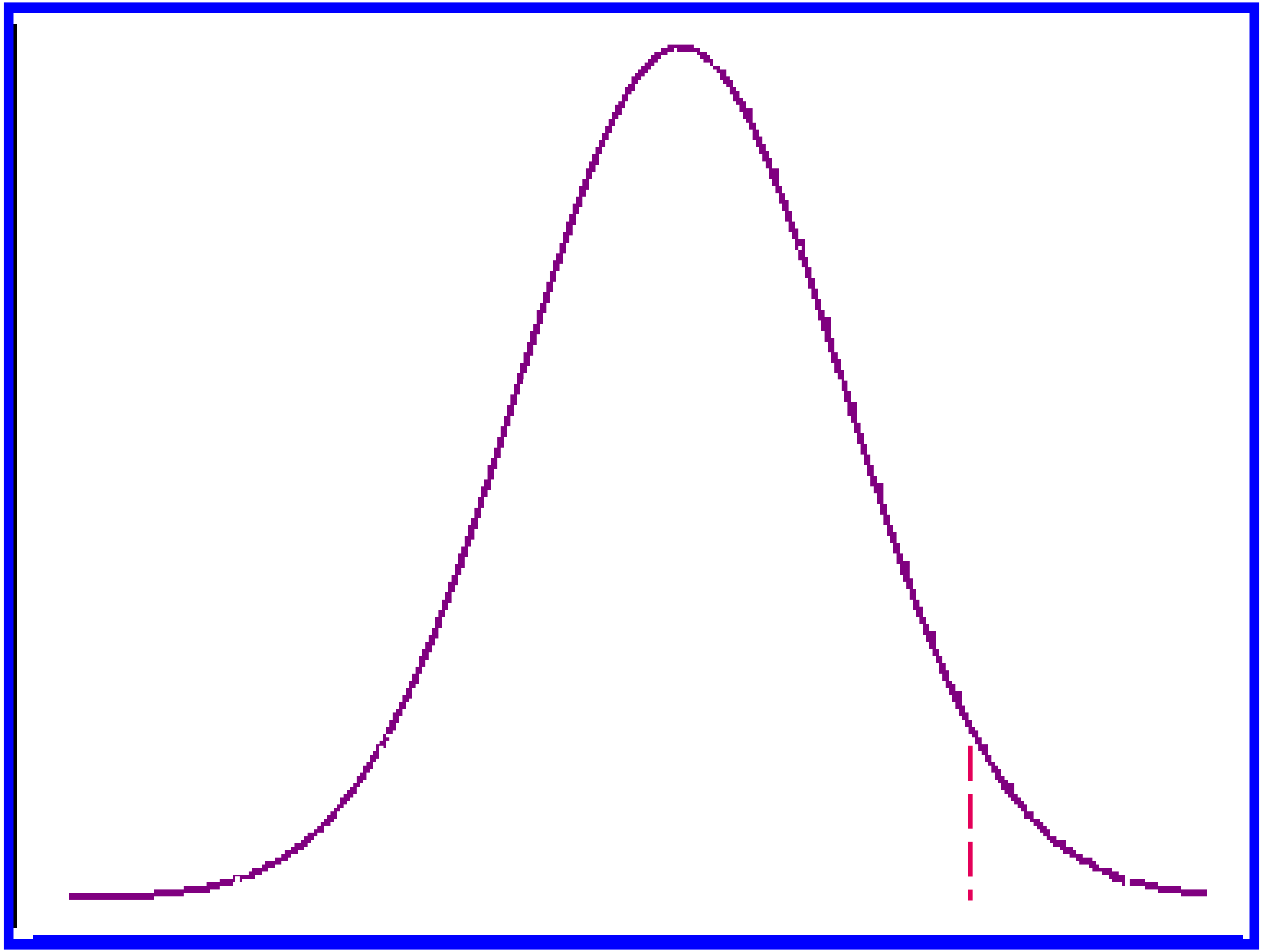
Interpretation and Uses of the Standard Deviation

- Chebyshev's theorem: For any set of observations, the minimum proportion of the values that lie within *k* standard deviations of the mean is at least 1 - 1/k , where k^2 is any constant greater than *1*.

Interpretation and Uses of the Standard Deviation

- Empirical Rule: For any symmetrical, bell-shaped distribution, approximately 68% of the observations will lie within $\pm 1\sigma$ of the mean (μ); approximately 95% of the observations will lie within $\pm 2\sigma$ of the mean (μ) approximately 99.7% within $\pm 3\sigma$ of the mean (μ).

Bell - Shaped Curve showing the relationship between σ and μ.



Relative Dispersion

- The coefficient of variation is the ratio of the standard deviation to the arithmetic mean, expressed as a percentage:

$$CV = \frac{s}{\overline{X}}(100\%)$$

Skewness

- Skewness is the measurement of the lack of symmetry of the distribution.
- The coefficient of skewness is computed from the following formula: Sk = 3(Mean - Median) / (Standard deviation)

Interquartile Range

- The Interquartile range is the distance between the third quartile Q_3 and the first quartile Q_1.
- Interquartile range = third quartile - first quartile = $Q_3 - Q_1$

First Quartile

- The First Quartile is the value corresponding to the point below which 25% of the observations lie in an ordered data set.

$$Q_1 = L + \frac{\frac{n}{4} - CF}{f}(i)$$

- where L=lower limit of the class containing Q1, CF= cumulative frequency preceding class containing Q1, f= frequency of class containing Q1, i= size of class containing Q1.

Third Quartile

- The Third Quartile is the value corresponding to the point below which 75% of the observations lie in an ordered data set:

$$Q_3 = L + \frac{\frac{3n}{4} - CF}{f}(i)$$

  - where L=lower limit of the class containing Q3, CF= cumulative frequency preceding class containing Q3, f= frequency of class containing Q3, i= size of class containing Q3.

Quartile Deviation

- The Quartile deviation is half the distance between the third quartile, Q_3, and the first quartile, Q_1.
- $QD = [Q_3 - Q_1]/2$

EXAMPLE 5

- If the third quartile = 24 and the first quartile = 10, what is the quartile deviation? The interquartile range is 24 - 10 = 14; thus the quartile deviation is 14/2 = 7.

Percentile Range

- Each data set has 99 percentiles, thus dividing the set into 100 equal parts.
- The percentile range is the distance between two stated percentiles. The 10-to-90 *percentile range* is the distance between the 10th and 90th percentiles.

Formula For Percentiles

$$Lp = (n + 1)\frac{p}{100}$$

Box Plots

- A box plot is a graphical display, based on quartiles, that helps to picture a set of data.
- Five pieces of data are needed to construct a box plot: the *Minimum Value*, the *First Quartile*, the *Median*, the *Third Quartile*, and the *Maximum Value*.

EXAMPLE 6

- Based on a sample of 20 deliveries, Marco's Pizza determined the following information: minimum value = 13 minutes, Q_1 = 15 minutes, median = 18 minutes, Q_3 = 22 minutes, maximum value = 30 minutes. Develop a box plot for the delivery times.

EXAMPLE 6 *continued*