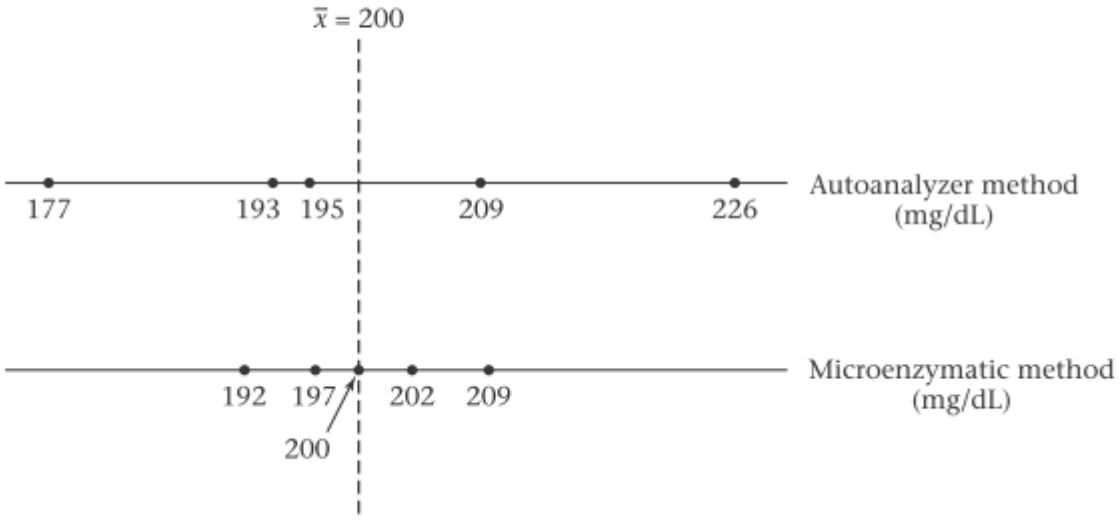
2.4 Measures of Spraed

Figure 2.4 Two samples of cholesterol measurements on a given person using the Autoanalyzer and Microenzymatic measurement methods



The mean obtained by the two methods is the same. However, the **variability** or **spread** of the Autoanalyzer method appears to be greater.

Range or variability

- Range is the difference between the largest and smallest observations in a sample.
- Once the sample is ordered, it is very easy to compute the range.
- Range is very sensitive to extreme observations or outliers.
- Larger the sample size (n), the larger the range and the more difficult the comparison between ranges from data sets of varying sizes.

EXAMPLE 2.14

The range in the sample of birthweights in Table 2.1 is

$$4146 - 2069 = 2077 \text{ g}$$

EXAMPLE 2.15

Compute the ranges for the Autoanalyzer- and Microenzymatic-method data in Figure 2.4, and compare the variability of the two methods.

Solution: The range for the Autoanalyzer method = 226 – 177 = 49 mg/dL. The range for the Microenzymatic method = 209 – 192 = 17 mg/dL. The Autoanalyzer method clearly seems more variable.

A better approach to quantifying the spread in data sets is percentiles or quantiles.

Percentiles are less sensitive to outliers and are not greatly affected by the sample size.

Quantiles or percentiles

The pth percentile is the value V_p such that p percent of the sample points are less than or equal to V_p.

Median is the 50th percentile, which is a special case of a quantile.

The ***p*th percentile** is defined by

- The $(k+1)$th largest sample point if $np/100$ is not an integer (where k is the largest integer less than $np/100$)
- The average of the $(np/100)$th and $(np/100 +1)$th largest observations if $np/100$ is an integer.

Frequently used percentiles are

- quartiles (25^{th}, 50^{th}, and 75^{th} percentiles)
- quintiles (20^{th}, 40^{th}, 60^{th}, and 80^{th} percentiles)
- deciles (10^{th}, 20^{th},..., 90^{th} percentiles)

TABLE 2.3 Sample of admission white-blood counts (× 1000) for all patients entering a hospital in Allentown, Pennsylvania, on a given day

i	x_i	i	x_i
1	7	6	3
2	35	7	10
3	5	8	12
4	9	9	8
5	8		

EXAMPLE 2.17 Compute the 20th percentile for the white-blood-count data in Table 2.3.

Solution: Because $np/100 = 9 \times .2 = 1.8$ is not an integer, the 20th percentile is defined by the (1 + 1)th largest value = second largest value = 5000.

Variance and Standard Deviation

Sample variance or variance, which is the average of the squares of the deviations from the sample mean, may be used

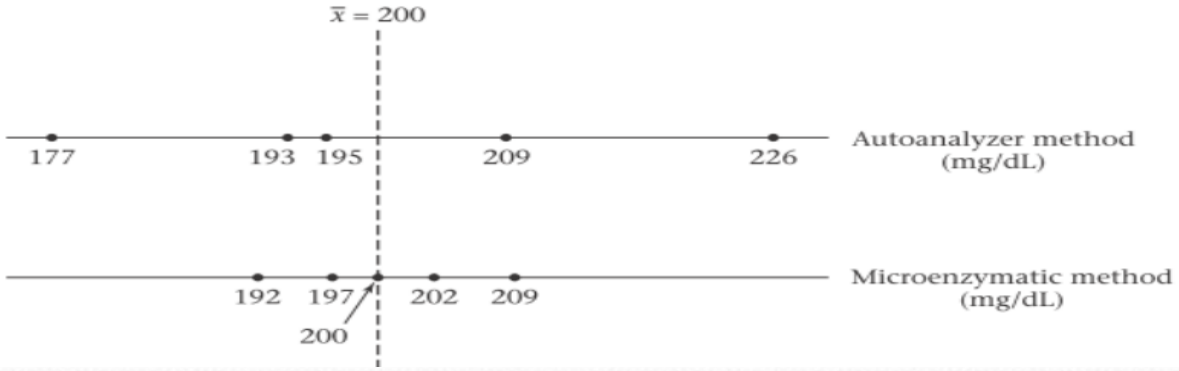
$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

standard deviation

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{\text{sample variance}}$$

EXAMPLE 2.19 Compute the variance and standard deviation for the Autoanalyzer- and Microenzymatic-method data in Figure 2.4.

Figure 2.4 Two samples of cholesterol measurements on a given person using the Autoanalyzer and Microenzymatic measurement methods



Solution: Autoanalyzer Method

$$s^2 = \left[(177-200)^2 + (193-200)^2 + (195-200)^2 + (209-200)^2 + (226-200)^2\right]/4$$
$$= (529+49+25+81+676)/4 = 1360/4 = 340$$
$$s = \sqrt{340} = 18.4$$

Microenzymatic Method

$$s^2 = \left[(192-200)^2 + (197-200)^2 + (200-200)^2 + (202-200)^2 + (209-200)^2\right]/4$$
$$= (64+9+0+4+81)/4 = 158/4 = 39.5$$
$$s = \sqrt{39.5} = 6.3$$

Thus the Autoanalyzer method has a standard deviation roughly three times as large as that of the Microenzymatic method.

Key Terms

Spread/variability: the extent to which observations differ from one another.

Range: maximum minus minimum.

Percentile: a value marking a specified percentage of observations at or below it.

Median: the 50th percentile.

Variance: a measure based on squared deviations from the mean.

Standard deviation: the square root of the variance.

Quick Review Questions

- Why is the mean alone insufficient to describe a data set?
- How is the range calculated, and what is its main limitation?
- What does the 75th percentile represent?
- Which percentile is the median?
- How are variance and standard deviation related?