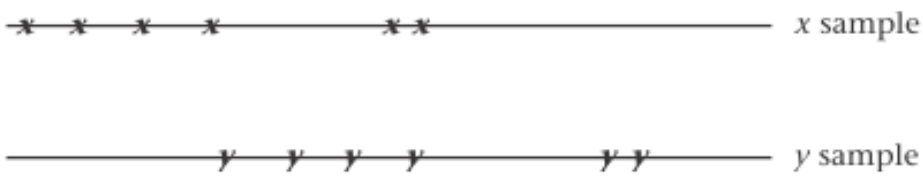
2.5 Properties of Variance and Standard Deviation

Figure 2.5 Comparison of the variances of two samples, where one sample has an origin shifted relative to the other



Samples $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ where $y_i = x_i + c$ $\quad i = 1, \ldots, n$

if respective sample variances are s_x^2 and s_y^2 then $s_y^2 = s_x^2$

Samples $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ where $y_i = cx_i$ $\quad i = 1, \ldots, n$ *and* $c>0$

then $s_y^2 = c^2 s_x^2$ which is $s_y = cs_x$

$$s_y^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1} = \frac{\sum_{i=1}^{n}(cx_i - c\bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n}[c(x_i - \bar{x})]^2}{n-1} = \frac{\sum_{i=1}^{n}c^2(x_i - \bar{x})^2}{n-1}$$

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

$$s_y = \sqrt{c^2 s_x^2} = cs_x$$

EXAMPLE 2.22

Compute the variance and standard deviation of the birthweight data in Table 2.1 in both grams and ounces.

Solution: The original data are given in grams, so first compute the variance and standard deviation in these units.

$$s^2 = \frac{(3265-3166.9)^2 + \cdots + (2834-3166.9)^2}{19}$$

$$= 3{,}768{,}147.8/19 = 198{,}323.6 \text{ g}^2$$

$$s = 445.3 \text{ g}$$

To compute the variance and standard deviation in ounces, note that

$$1 \text{ oz} = 28.35 \text{ g} \quad \text{or} \quad y_i = \frac{1}{28.35} x_i$$

Thus, $s^2(\text{oz}) = \frac{1}{28.35^2} s^2(\text{g}) = 246.8 \text{ oz}^2$

$$s(\text{oz}) = \frac{1}{28.35} s(\text{g}) = 15.7 \text{ oz}$$

2.6 Coefficient of Variation (CV)

Defined as $100\% \times (s/\bar{x})$

Table 2.6 Reproducibility of cardiovascular risk factors in children, Bogalusa Heart Study, 1978–1979

	n	Mean	*sd*	*CV* (%)
Height (cm)	364	142.6	0.31	0.2
Weight (kg)	365	39.5	0.77	1.9
Triceps skin fold (mm)	362	15.2	0.51	3.4
Systolic blood pressure (mm Hg)	337	104.0	4.97	4.8
Diastolic blood pressure (mm Hg)	337	64.0	4.57	7.1
Total cholesterol (mg/dL)	395	160.4	3.44	2.1
HDL cholesterol (mg/dL)	349	56.9	5.89	10.4

The coefficient of variation is independent of the measurement units used. It is useful for comparing the relative variability of different samples, especially when their means differ, and for comparing the reproducibility of different variables. A larger CV indicates greater variability relative to the mean.

2.8 Graphic Methods

Graphic methods of displaying data give a quick overall impression of data. The following are some graphic methods.

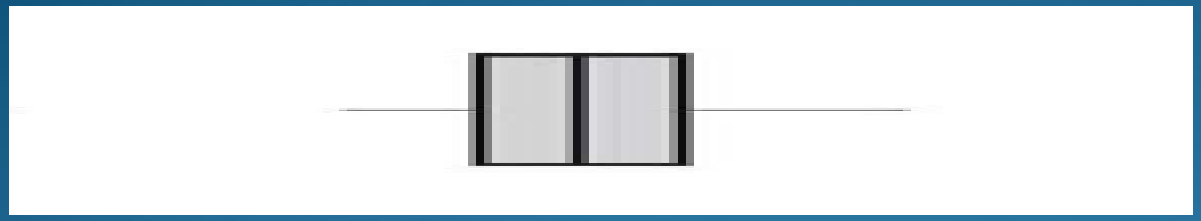
- Bar graph
- Histogram
- Boxplot

Box plot

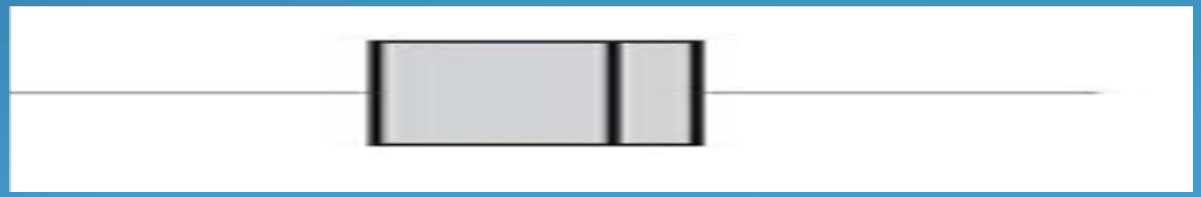
- If the distribution is symmetric, then upper and lower quartiles should be approximately equally spaced from the median

- If the upper quartile is farther from the median than the lower quartile, then the distribution is positively skewed

- If the lower quartile is farther from the median than the upper quartile, then the distribution is negatively skewed







Key Terms

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

Standard deviation: the square root of the variance.

Coefficient of variation: standard deviation expressed as a percentage of the mean.

Quartiles: values that divide ordered data into four parts.

Skewness: asymmetry in a distribution.

Quick Review Questions

- What happens to the variance when the same constant is added to every observation?
- If every observation is multiplied by 4, how does the standard deviation change?
- Write the formula for the coefficient of variation.
- Which graphical display uses quartiles and the median?
- How can quartile distances help identify positive or negative skewness?
- Why should a data set be described using both location and spread?