Chapter 1

Descriptive Statistics

1.1 Populations, Samples, and Processes

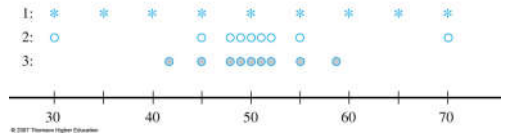
1.2 Pictorial and Tabular Methods in Descriptive Statistics

1.3 Measures of Location

1.4 Measures of Variability

28

Variability



29

The Range

is the difference between the largest and smallest sample values.

30

Example

US beer production (in millions of barrels) for different quarters during 1975 – 1982:

32, 47, 48, 34, 35, 46, 47, 33, 37, 51, 49, 33, 40, 52, 51, 39, 42, 56, 55, 43, 45, 57, 57, 43, 45, 58, 57, 44, 49, 56, 54, 50.

Find the range of this data set.

31

The (Sample) Variance

of observations $x_1, x_2, \ldots, x_n$ is given by

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

The root of the sample variance is called the sample standard deviation

32

Example

A sample of eight resistors of a certain type resulted in the following sample resistances (in ohm):

40, 43, 39, 35, 37, 43, 46, 37.

Find the sample variance.

33

A Computing Formula for the Variance

An alternative expression for s^2 is

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

34

Example

A sample of eight resistors of a certain type resulted in the following sample resistances (in ohm):

40, 43, 39, 35, 37, 43, 46, 37.

Find the sample variance using the shortcut formula.

35

Properties of the Variance

1. **Adding the same number to each element of a data set leaves the sample variance unchanged.**
2. **Multiplying each element of a data set with the same number results in a data set with new sample variance equal to the old sample variance times the square of that number.**

36

Example

A sample of eight resistors of a certain type resulted in the following sample resistances (in ohm):

40, 43, 39, 35, 37, 43, 46, 37.

Find the sample variance if each of the resistances were

(a) reduced by 35

(b) multiplied by 10.

37

Lower/Upper Fourth, Fourth Spread

1. **Order the observations from smallest to largest and separate the smallest half from the largest half (the median is included in both if the sample size is odd).**
2. **The lower fourth is the median of the smallest half.**
3. **The upper fourth is the median of the largest half.**
4. **The fourth spread is the difference between the upper fourth and the lower fourth.**

38

Boxplots

1. **Draw a horizontal measure scale.**
2. **Place a rectangle above the axis, where the left edge of the rectangle is at the lower fourth and the right edge at the upper fourth.**
3. **Place a vertical line segment inside the rectangle at the location of the median.**
4. **Draw "whiskers" out from either end of the rectangle to the smallest and largest observations.**

39

Example

US beer production (in millions of barrels) for different quarters during 1975 – 1982:

32, 47, 48, 34, 35, 46, 47, 33, 37, 51, 49, 33, 40, 52, 51, 39, 42, 56, 55, 43, 45, 57, 57, 43, 45, 58, 57, 44, 49, 56, 54, 50.

Draw a boxplot of this data set.

40

Outliers

- **Any observation farther than 1.5 times the fourth spread from the closest fourth is called an outlier.**
- **An outlier is extreme if it is more than 3 times the fourth spread from the closest fourth.**
- **An outlier is mild if it is not extreme.**

41

Example

A sample of pulse widths from slow discharges in a cylindrical cavity made of polyethylene is as follows:

5.3, 8.2, 13.8, 74.1, 85.3, 88.0, 90.2, 91.5, 92.4, 92.9, 93.6, 94.3, 94.8, 94.9, 95.5, 95.8, 95.9, 96.6, 96.7, 98.1, 99.0, 101.4, 103.7, 106.0, 113.5.

Draw a boxplot of this data set that shows and classifies the outliers.

42