

Quantitative Methods – I

A.Y. 2021-22

Practice 4

Lorenzo Cavallo

For any clarification/meeting: cavallo@istat.it

Other resources

Beginning Statistics

(book under Creative Commons-licensed copies)

<https://2012books.lardbucket.org/>

Online Statistics Education: An Interactive Multimedia Course of Study

<https://onlinestatbook.com/mobile/index.html>

Searching on Google:

Essentials of Statistics

by Triola, Mario F.

Essentials of Statistics: Exercises

By Brink, David

(exercises from probability to regression)

THEME #1

—

Variability

Measures of dispersion

Range

It is obtained by taking the difference between the largest and the smallest values in a data set.

Range=Largest value–Smallest value

Interquartile Range

The difference between the third and the first quartiles

$IQR = Q3 - Q1$

Variance and Standard Deviation

The variance is the squared deviation of a variable from its mean.

The standard deviation is obtained by taking the positive square root of the variance.

$$\sigma^2 = \frac{\sum (x - \mu)^2}{N} \quad \text{and} \quad s^2 = \frac{\sum (x - \bar{x})^2}{n-1}$$
$$\sigma = \sqrt{\frac{\sum (x - \mu)^2}{N}} \quad \text{and} \quad s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$

Coefficient of Variation

The coefficient of variation, denoted by CV, expresses standard deviation as a percentage of the mean.

For population data : $CV = \frac{\sigma}{\mu} \times 100\%$

For sample data : $CV = \frac{s}{\bar{x}} \times 100\%$

A corporation hired 10 graduates. The starting salaries for each graduate are shown. Find the range of the starting salaries.

Starting salaries (1000s of dollars)

41 38 39 45 47 41 44 41 37 42

Solution

- Ordering the data helps to find the least and greatest salaries.

37 38 39 41 41 41 42 44 45 47

minimum maximum

- $\text{Range} = (\text{Max. salary}) - (\text{Min. salary})$
 $= 47 - 37 = 10$

The range of starting salaries is 10 or \$10,000.

Ex. 1 Find the Variance and the Standard deviation

A corporation hired 10 graduates. The starting salaries for each graduate are shown. Find the range of the starting salaries.

Starting salaries (1000s of dollars)

41 38 39 45 47 41 44 41 37 42

Solution

- First determine the mean starting salary.

$$\mu = \frac{\sum x}{N} = \frac{415}{10} = 41.5$$

- Determine the deviation for each data entry.

| Salary (\$1000s), x | Deviation (\$1000s) $x - \mu$ |
|-----------------------|----------------------------------|
| 41 | $41 - 41.5 = -0.5$ |
| 38 | $38 - 41.5 = -3.5$ |
| 39 | $39 - 41.5 = -2.5$ |
| 45 | $45 - 41.5 = 3.5$ |
| 47 | $47 - 41.5 = 5.5$ |
| 41 | $41 - 41.5 = -0.5$ |
| 44 | $44 - 41.5 = 2.5$ |
| 41 | $41 - 41.5 = -0.5$ |
| 37 | $37 - 41.5 = -4.5$ |
| 42 | $42 - 41.5 = 0.5$ |

Ex. 1 Find the Variance and the Standard deviation

A corporation hired 10 graduates. The starting salaries for each graduate are shown. Find the range of the starting salaries.

Starting salaries (1000s of dollars)

41 38 39 45 47 41 44 41 37 42

Solution

- Square each deviation

| Salary, x | Deviation: $x - \mu$ | Squares: $(x - \mu)^2$ |
|-------------|----------------------|------------------------|
| 41 | $41 - 41.5 = -0.5$ | $(-0.5)^2 = 0.25$ |
| 38 | $38 - 41.5 = -3.5$ | $(-3.5)^2 = 12.25$ |
| 39 | $39 - 41.5 = -2.5$ | $(-2.5)^2 = 6.25$ |
| 45 | $45 - 41.5 = 3.5$ | $(3.5)^2 = 12.25$ |
| 47 | $47 - 41.5 = 5.5$ | $(5.5)^2 = 30.25$ |
| 41 | $41 - 41.5 = -0.5$ | $(-0.5)^2 = 0.25$ |
| 44 | $44 - 41.5 = 2.5$ | $(2.5)^2 = 6.25$ |
| 41 | $41 - 41.5 = -0.5$ | $(-0.5)^2 = 0.25$ |
| 37 | $37 - 41.5 = -4.5$ | $(-4.5)^2 = 20.25$ |
| 42 | $42 - 41.5 = 0.5$ | $(0.5)^2 = 0.25$ |

Ex. 1 Find the Variance and the Standard deviation

A corporation hired 10 graduates. The starting salaries for each graduate are shown. Find the range of the starting salaries.

Starting salaries (1000s of dollars)

41 38 39 45 47 41 44 41 37 42

Solution

| Salary, x | Deviation: $x - \mu$ | Squares: $(x - \mu)^2$ |
|-------------|----------------------|------------------------|
| 41 | $41 - 41.5 = -0.5$ | $(-0.5)^2 = 0.25$ |
| 38 | $38 - 41.5 = -3.5$ | $(-3.5)^2 = 12.25$ |
| 39 | $39 - 41.5 = -2.5$ | $(-2.5)^2 = 6.25$ |
| 45 | $45 - 41.5 = 3.5$ | $(3.5)^2 = 12.25$ |
| 47 | $47 - 41.5 = 5.5$ | $(5.5)^2 = 30.25$ |
| 41 | $41 - 41.5 = -0.5$ | $(-0.5)^2 = 0.25$ |
| 44 | $44 - 41.5 = 2.5$ | $(2.5)^2 = 6.25$ |
| 41 | $41 - 41.5 = -0.5$ | $(-0.5)^2 = 0.25$ |
| 37 | $37 - 41.5 = -4.5$ | $(-4.5)^2 = 20.25$ |
| 42 | $42 - 41.5 = 0.5$ | $(0.5)^2 = 0.25$ |

- Find the variance

$$\sum (x_i - \mu)^2 = 88.5 \text{ and } N=10$$

$$\sigma^2 = \sum \frac{(x_i - \mu)^2}{N} = \frac{88.5}{10} = 8.85$$

- Find the standard deviation

$$\sigma = \sqrt{8.85} = 2.97$$

Ex. 1 Find the CV

A corporation hired 10 graduates. The starting salaries for each graduate are shown. Find the range of the starting salaries.

Starting salaries (1000s of dollars)

41 38 39 45 47 41 44 41 37 42

Solution

$$\mu = \frac{\sum x}{N} = \frac{415}{10} = 41.5$$

$$\sigma = \sqrt{8.85} = 2.97$$

$$CV = \frac{\sigma}{\mu} \cdot 100 = \frac{2.97}{41.5} \cdot 100 = 7.17 \%$$

Another way to calculate the variance

$$\sigma^2 = \sum \frac{(x_i - \mu)^2}{N}$$

or

$$\sigma^2 = \sum \frac{x_i^2}{N} - \mu^2$$

Different formulas same results

Ex. 1 Find the Variance and the Standard deviation

A corporation hired 10 graduates. The starting salaries for each graduate are shown. Find the range of the starting salaries.

Starting salaries (1000s of dollars)

41 38 39 45 47 41 44 41 37 42

Solution

- Find the variance

$$\sum_{i=1}^{10} x_i^2 = 41^2 + 38^2 + 39^2 + 45^2 + \dots + 42^2 = 17311$$

$$N=10$$

$$\mu = \frac{\sum x}{N} = \frac{415}{10} = 41.5$$

$$\sigma^2 = \sum \frac{x_i^2}{N} - \mu^2$$

$$\sigma^2 = \frac{17311}{10} - 41.5^2 = 1731.1 - 1722.25 = 8.85$$

Ex. 2

Calculate the mean, variance and standard deviation for the following distribution :

| Class | Frequency (f_i) |
|----------|------------------------|
| 30 – 40 | 3 |
| 40 – 50 | 7 |
| 50 – 60 | 12 |
| 60 – 70 | 15 |
| 70 – 80 | 8 |
| 80 – 90 | 3 |
| 90 – 100 | 2 |
| | $\Sigma f_i = 50$ |

Solution

| Class | Frequency (f_i) | m_i | $m_i f_i$ |
|----------|------------------------|-------|-------------------------|
| 30 – 40 | 3 | 35 | $35 \times 3 = 105$ |
| 40 – 50 | 7 | 45 | $45 \times 7 = 315$ |
| 50 – 60 | 12 | 55 | $55 \times 12 = 660$ |
| 60 – 70 | 15 | 65 | $65 \times 15 = 975$ |
| 70 – 80 | 8 | 75 | $75 \times 8 = 600$ |
| 80 – 90 | 3 | 85 | $85 \times 3 = 255$ |
| 90 – 100 | 2 | 95 | $95 \times 2 = 190$ |
| | $\Sigma f_i = 50$ | | $\Sigma m_i f_i = 3100$ |

The mean of the distribution is:

$$\mu = \sum \frac{m_i f_i}{N} = \frac{3100}{50} = 62$$

Ex. 2

Finding Variance and Standard Deviation

| Class | Frequency (f_i) | m_i | $(m_i - \mu)^2$ | $(m_i - \mu)^2 f_i$ |
|----------|------------------------|-------|----------------------|------------------------|
| 30 – 40 | 3 | 35 | $(35 - 62)^2 = 729$ | $3 \times 729 = 2187$ |
| 40 – 50 | 7 | 45 | $(45 - 62)^2 = 289$ | $7 \times 289 = 2023$ |
| 50 – 60 | 12 | 55 | $(55 - 62)^2 = 49$ | $12 \times 49 = 588$ |
| 60 – 70 | 15 | 65 | $(65 - 62)^2 = 9$ | $15 \times 9 = 135$ |
| 70 – 80 | 8 | 75 | $(75 - 62)^2 = 169$ | $8 \times 169 = 1352$ |
| 80 – 90 | 3 | 85 | $(85 - 62)^2 = 529$ | $3 \times 529 = 1589$ |
| 90 – 100 | 2 | 95 | $(95 - 62)^2 = 1089$ | $2 \times 1089 = 2187$ |
| | $\Sigma f_i = 50$ | | | Sum = 10050 |

$$\sigma^2 = \sum \frac{(m_i - \mu)^2 f_i}{N} = \frac{10050}{50} = 201$$

$$\sigma = \sqrt{201} = 14.17$$

...and the CV?

$$\sigma = 14.17$$

$$\mu = \sum \frac{m_i f_i}{N} = 62$$

$$CV = \frac{\sigma}{\mu} \cdot 100$$

$$CV = \frac{14.17}{62} \cdot 100$$

$$CV = 22.85 \%$$

THEME #2



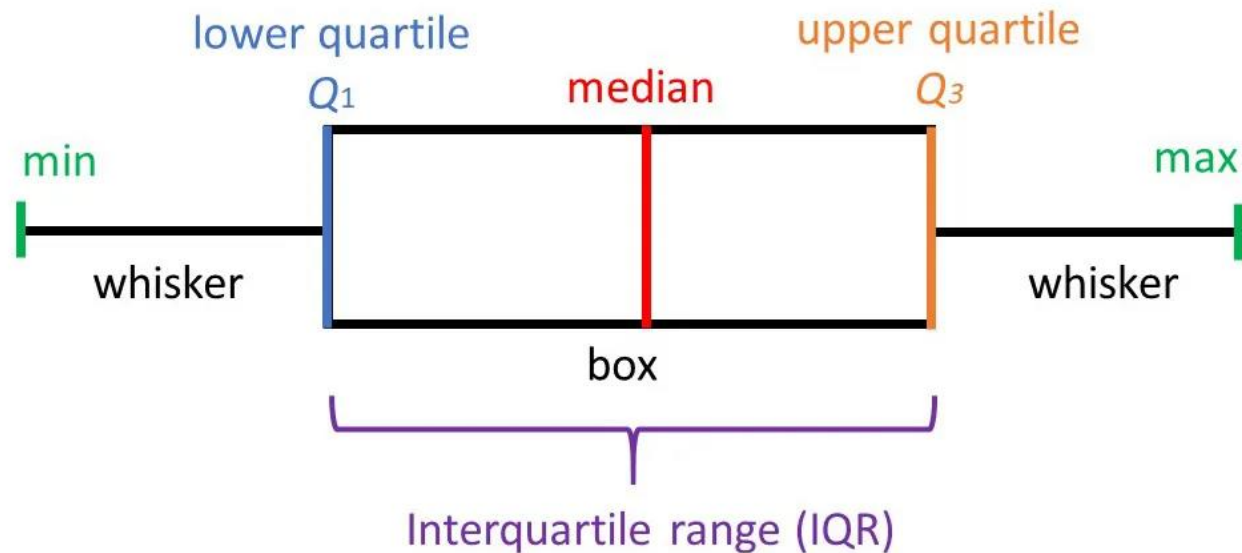
Box-Plot

Box-Whiskers Plot (or Box-Plot)

Comprehensive graphical representation of a distribution.

Indeed, it provides info on:

- position: median, Q_1 , and Q_3 (lines)
- dispersion: interquartile range (box)
- shape of the distribution (whiskers)
- extreme values: outliers



bp.1 An employee of a computer store recorded the number of sales he made each month. In the past 12 months, he sold the following numbers of computers:

51, 20, 25, 39, 7, 44, 92, 41, 22, 6, 42, 18.

Make the box and whisker plots.

Solution

First, put the data in ascending order. Then find the median.

6, 7, 18, 20, 22, 25, 39, 41, 42, 44, 51, 92

$N=12$

Median position = $\frac{N+1}{2} = \frac{12+1}{2} = 6.5\text{th value}$

Median = $\frac{\text{sixth} + \text{seventh obs.}}{2} = \frac{25 + 39}{2} = 32$

There are six numbers below the median, namely: 6, 7, 18, 20, 22, 25.

Q_1 position = the median of these six items = $\frac{N+1}{2} = \frac{6+1}{2} = 3.5\text{th value}$

$Q_1 = (\text{third} + \text{fourth observations}) / 2 = (18 + 20) / 2 = 19$

There are six numbers above the median, namely: 39, 41, 42, 44, 51, 92.

Q_3 position = the median of these six items = $\frac{N+1}{2} = \frac{6+1}{2} = 3.5\text{th value}$

$Q_3 = (\text{third} + \text{fourth observations}) / 2 = (42+44) / 2 = 43$

bp.1 An employee of a computer store recorded the number of sales he made each month. In the past 12 months, he sold the following numbers of computers:

51, 20, 25, 39, 7, 44, 92, 41, 22, 6, 42, 18.

Make the box and whisker plots.

Median = 32

$Q_1 = 19$

$Q_3 = 43$

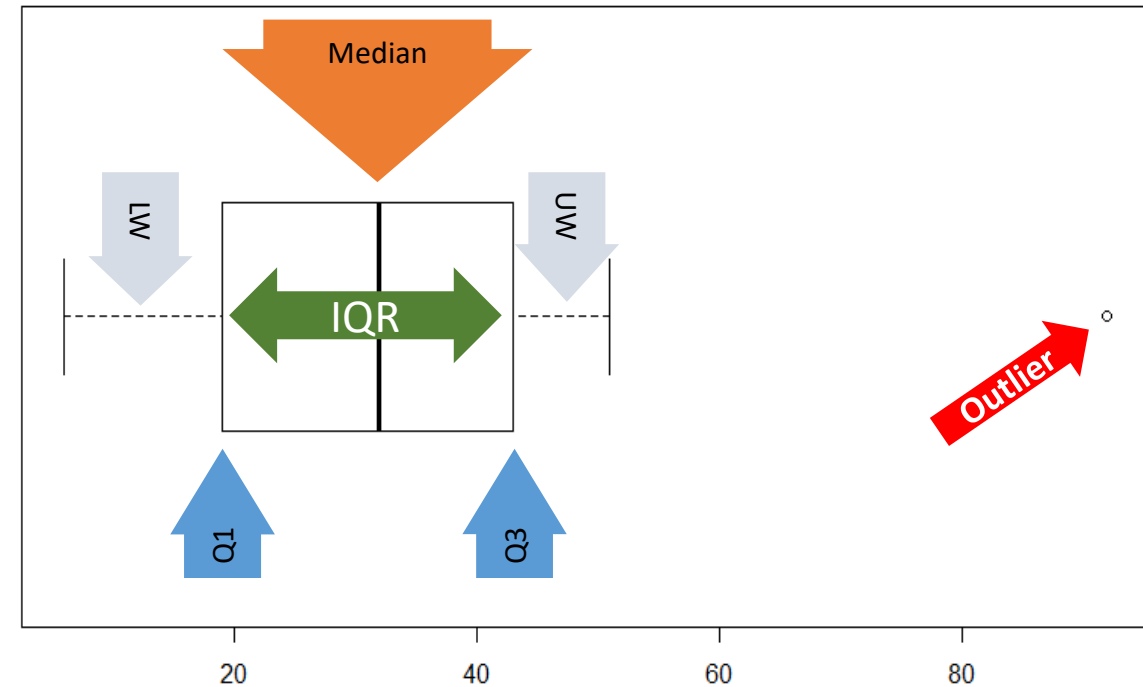
$IQR = Q_3 - Q_1 = 43 - 19 = 24$

Whiskers:

Upper = $Q_3 + 1.5 IQR = 43 + 1.5 \cdot 24 = 43 + 36 = 79$

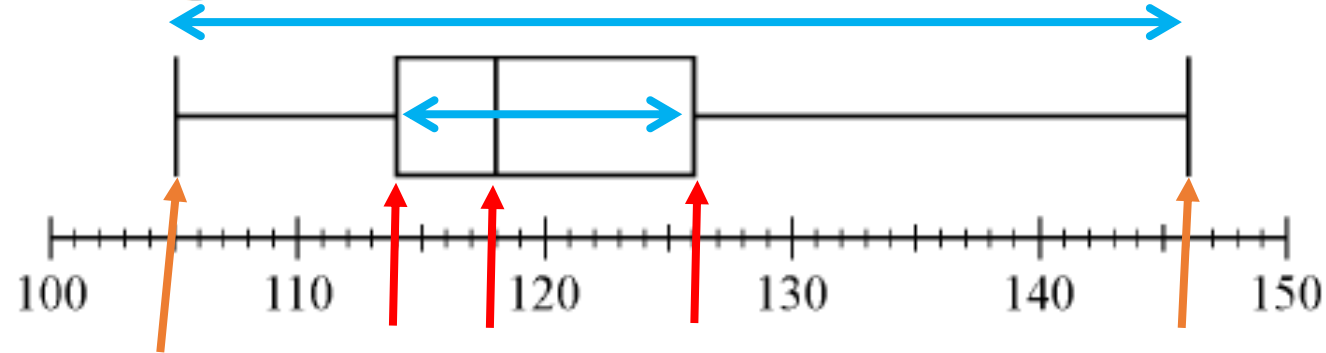
Lower = $Q_1 - 1.5 IQR = 19 - 36 = -17$
(smaller than the minimum value)

1 Upper outlier (92)



bp.2

Eight hundred insects were weighed, and the resulting measurements, in milligrams, are summarized in the boxplot below.



(a) What are the range, the three quartiles, and the interquartile range of the measurements?

Solution

We have 800 data ($n=800$).

From the box-plot we can conclude that:

The minimum value is 105 and maximum value is 146

The **range** it is maximum minus smallest or $146 - 105 = 41$

The **median** is 118.

First quartile is 114 and **third** is 126.

The **IQR (Interquartile Range)** is $126 - 114 = 12$