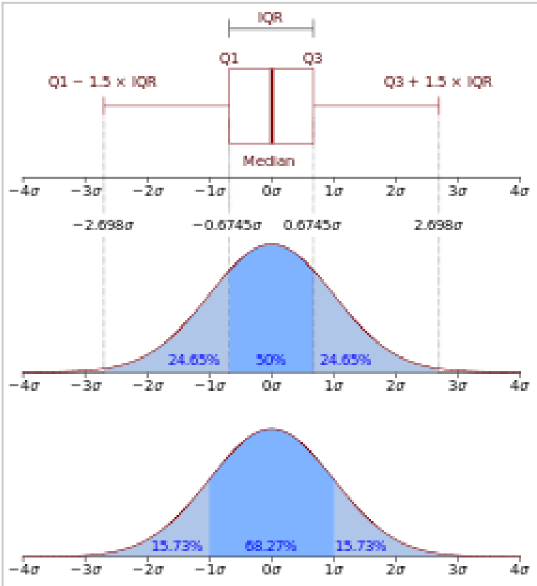
Interquartile range

From Wikipedia, the free encyclopedia

In descriptive statistics, the **interquartile range** (**IQR**), also called the **midspread** or **middle fifty**, is a measure of statistical dispersion, being equal to the difference between the upper and lower quartiles,[1][2] IQR = $Q_3 - Q_1$. In other words, the IQR is the 1st quartile subtracted from the 3rd quartile; these quartiles can be clearly seen on a box plot on the data. It is a trimmed estimator, defined as the 25% trimmed range, and is the most significant basic robust measure of scale.



Boxplot (with an interquartile range) and a probability density function (pdf) of a Normal $N(0,\sigma^2)$ Population

Contents

Use

Unlike total range, the interquartile range has a breakdown point of 50%, and is thus often preferred to the total range.

The IQR is used to build box plots, simple graphical representations of a probability distribution.

For a symmetric distribution (where the median equals the midhinge, the average of the first and third quartiles), half the IQR equals the median absolute deviation (MAD).

The median is the corresponding measure of central tendency.

Identification of outliers (see below).

Examples

Data set in a table

i	x[i]	Quartile
1	102	
2	104	
3	105	Q_1
4	107	
5	108	
6	109	Q_2 (median)
7	110	
8	112	
9	115	Q_3
10	116	
11	118	

For the data in this table the interquartile range is IQR = 115 − 105 = 10.

Data set in a plain-text box plot

```
                            +-----+-+
  o           *     |-------|     | |---|
                            +-----+-+

+---+---+---+---+---+---+---+---+---+---+---+---+   number line
0   1   2   3   4   5   6   7   8   9   10  11  12
```

For the data set in this box plot:

- lower (first) quartile $Q_1 = 7$
- median (second quartile) $Q_2 = 8.5$
- upper (third) quartile $Q_3 = 9$
- interquartile range, IQR = $Q_3 - Q_1 = 2$

Interquartile range of distributions

The interquartile range of a continuous distribution can be calculated by integrating the probability density function (which yields the cumulative distribution function — any other means of calculating the CDF will also work). The lower quartile, Q_1, is a number such that integral of the PDF from -∞ to Q_1 equals 0.25, while the upper quartile, Q_3, is such a number that the integral from -∞ to Q_3 equals 0.75; in terms of the CDF, the quartiles can be defined as follows:

$$Q_1 = \text{CDF}^{-1}(0.25),$$

$$Q_3 = \text{CDF}^{-1}(0.75),$$

where CDF^{-1} is the quantile function.

The interquartile range and median of some common distributions are shown below

Distribution	Median	IQR
Normal	μ	$2\,\Phi^{-1}(0.75)\sigma \approx 1.349\sigma \approx (27/20)\sigma$
Laplace	μ	$2b\ln(2) \approx 1.386b$
Cauchy	μ	2γ

Interquartile range test for normality of distribution

The IQR, mean, and standard deviation of a population *P* can be used in a simple test of whether or not *P* is normally distributed, or Gaussian. If *P* is normally distributed, then the standard score of the first quartile, z_1, is -0.67, and the standard score of the third quartile, z_3, is +0.67. Given *mean* = *X* and *standard deviation* = σ for *P*, if *P* is normally distributed, the first quartile

$$Q_1 = (\sigma\, z_1) + X$$

and the third quartile

$$Q_3 = (\sigma\, z_3) + X$$

If the actual values of the first or third quartiles differ substantially from the calculated values, *P* is not normally distributed.

Interquartile range and outliers

The interquartile range is often used to find outliers in data. Outliers are observations that fall below Q1 - 1.5(IQR) or above Q3 + 1.5(IQR). In a boxplot, the highest and lowest occurring value within this limit are drawn as bar of the *whiskers*, and the outliers as individual points.

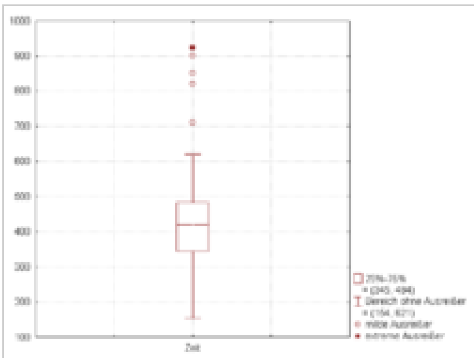


Figure 3. Box-and-whisker plot with four close and one far away extreme values, defined as outliers above Q3 + 1.5(IQR) and Q3 + 3(IQR), respectively.

See also

- Midhinge
- Interdecile range
- Robust measures of scale

References

1. Upton, Graham; Cook, Ian (1996). *Understanding Statistics*. Oxford University Press. p. 55. ISBN 0-19-914391-9.
2. Zwillinger, D., Kokoska, S. (2000) *CRC Standard Probability and Statistics Tables and Formulae*, CRC Press. ISBN 1-58488-059-7 page 18.

Categories: Scale statistics
