Variance

The **variance** of a random variable X is the expected value of the squared deviation from the mean of X, $\mu = E(X)$:

$$\mathrm{Var}(X) = E((X - \mu)^2)$$

The variance of a collection of n equally likely $x_1, x_2, \ldots, x_n$ values can be written as

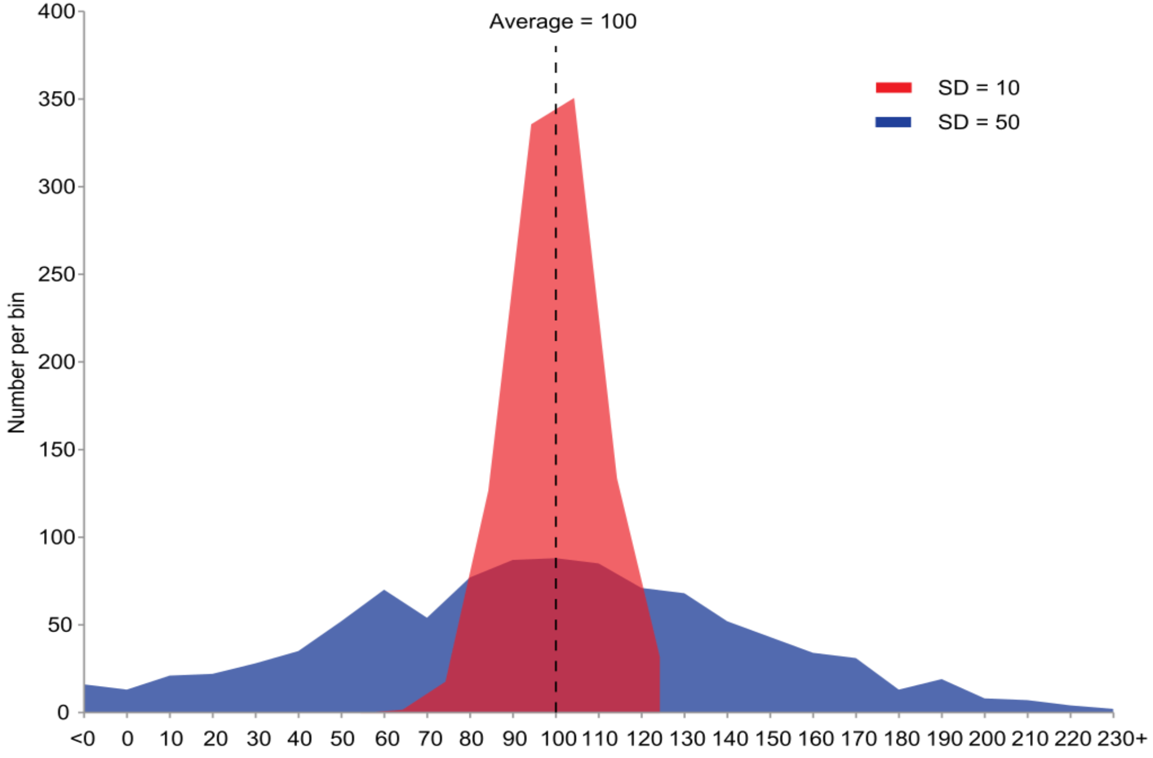
$$\mathrm{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

where μ is the **average value**, that is,

$$\mu = \frac{1}{n}\sum_{i=1}^{n}x_i$$

Variance

Variance is a measure of dispersion, meaning it is a measure of how far a set of numbers is spread out from their average value.



Variance of a discrete random variable

If the process behind a random variable *X* is discrete with probability mass function $P_X(x_1) = p_1, P_X(x_2) = p_2, \dots, P_X(x_n) = p_n,$ then

$$\mathrm{Var}(X) = \sum_{i=1}^{n} p_i \cdot (x_i - \mu)^2$$

EXAMPLE. Use R

Standard deviation

The standard deviation (SD) of a random variable is the square root of its variance:

$$SD(X) = \sqrt{\mathrm{Var}(X)}$$

Practically the standard deviation and variance measure the same thing. SD has the advantage that the standard deviation of *X* has the same unit as *X*.

Example. Use R

Normal distribution

In probability theory, a **normal distribution** (also known as **Gaussian distribution**) is a type of continuous probability distribution for a real-valued random variable. It has the density function:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

$\mu =$ mean of x

$\sigma =$ standard deviation of x

$\pi \approx 3.14159 \ldots$

$e \approx 2.71828 \ldots$

Normal distribution

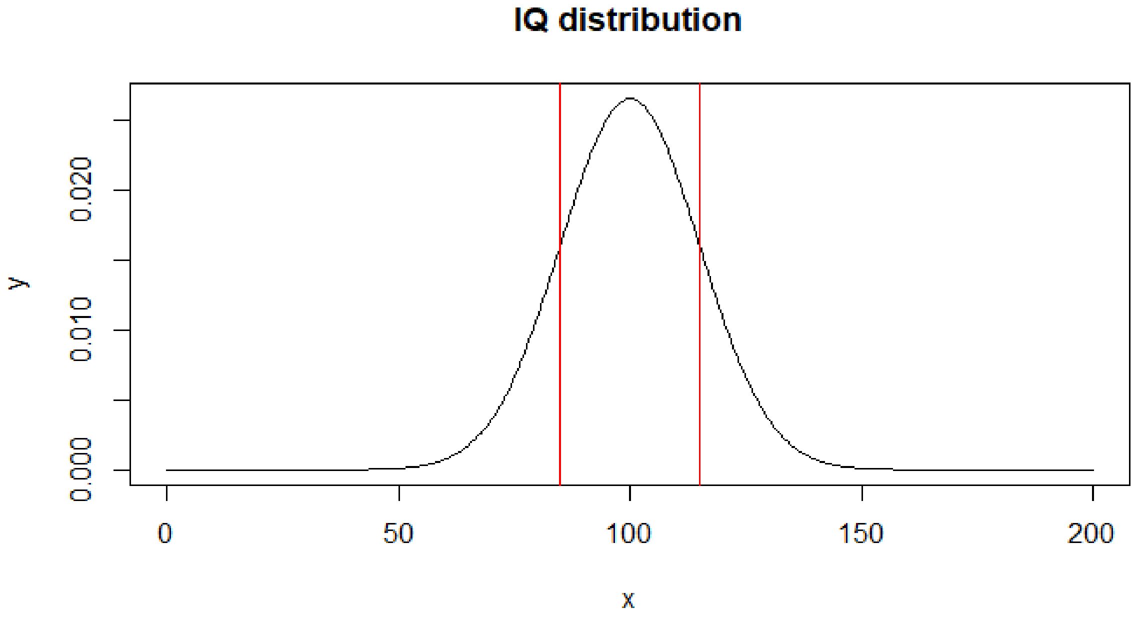
- The normal distribution then depends only on the mean and standard deviation.
- The density function can be computed in R by the function:

```
dnorm(x=15, mean=20, sd=5)
```

- Normal distribution has a bell shape, as we can seen from the following example.

Intelligence quotient IQ

- The mean, or average, IQ is 100. Standard deviation is 15 points.
- The majority of the population, 68.26%, falls within one standard deviation of the mean (IQ 85-115).



68–95–99.7 rule

In statistics, the 68–95–99.7 rule, also known as the empirical rule, is a shorthand used to remember the percentage of values that lie within an interval estimate in a normal distribution:

68%, 95%, and 99.7% of the values lie within one, two, and three standard deviations of the mean, respectively.