Proposition 1 *Let $f(x)$ be the density function of absolutely continuous random variable X, $g(x): \mathbb{R} \to \mathbb{R}$ an arbitrary function, and $Y = g(X)$. Then $E(Y) = \int\limits_{-\infty}^{\infty} g(y) f_X(y)dy$.*

Remark 1 *The definition of variance and properties remain in the absolutely continuous case, i.e., $Var(X) = E\left((X - E(X))^2\right)$ etc. Also remark that for an arbitrary $k > 0$ integer, $E(X^k) = \int_{-\infty}^{\infty} z^k f(z)dz$.*

Example 1 *If X is uniformly distributed on interval $[a, b]$, then $E(X^2) = \int_a^b y^2 \frac{1}{b-a} dy = \left[\frac{y^3}{3(b-a)}\right]_{y=a}^b = \frac{a^2+ab+b^2}{3}$. Thus*

$$Var(X) = E(X^2) - E^2(X) = \frac{a^2 + ab + b^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{a^2 - 2ab + b^2}{12} = \frac{(a-b)^2}{12}.$$

Example 2 *Suppose that the distribution of random variable X is Exponential(λ). Then with integration by parts $E(X^2) = \int_0^{\infty} y^2 \lambda e^{-\lambda y} dy = \left[-y^2 e^{-\lambda y}\right]_{y=0}^{\infty} + \int_0^{\infty} 2ye^{-\lambda y} dy = \frac{2}{\lambda^2}$. Thus*

$$Var(X) = E(X^2) - E^2(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Example 3 *Let X be a standard normal random variable. Then with integration by parts*

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{\infty} x \left(x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\right) dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Thus $Var(X) = 1$. Furthermore, an arbitrary (m, σ^2) parametrized normal random variable Y has a variance of $Var(Y) = \sigma^2$, because $Var(Y) = Var(\sigma X + m) = \sigma^2 Var(X)$.

Mathematical statistics

1. **population**: we are interested in certain characteristics of the population; e.g. products, voters of Hungary, the daily changes of a currency exchange rate, effectiveness od medications; mostly it is not possible to know the 100% of the data

2. **sample**: a subset of the population, for which the values are available; since the sampling procedure is random, the sample elements are random variables; in practice we assume that these variables are independent

The result of sampling is the sample, which is a series $x_1, x_2, \ldots, x_n$ of numbers, the realization of $X_1, X_2, \ldots, X_n$ random variables.

Types of data:

1. nominal: we can only count frequencies (e.g. nationality)

2. ordinal (ordered): e.g. evaluation with words (bad, acceptable, good)

3. interval: e.g. temperature

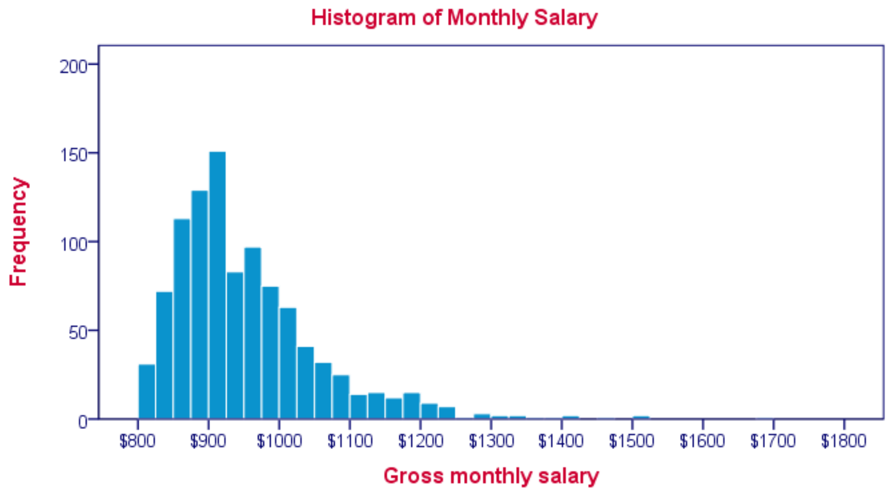
4. proportion

Graphical representations:

1. pie chart

2. histogram

3. boxplot



Histogram: We sort our data into classes, and plot the number of entries in each class.
Mean values:

1. sample average: $\overline{x} = \frac{x_1+\ldots+x_n}{n}$; if f_i denotes the frequencies of values l_i, then $\overline{x} = \frac{f_1 l_1+\ldots+f_k l_k}{n}$
2. median: the middle element of the sorted sample (if the number of elements is even, then the average of the two middle numbers)
3. quartiles: quadrant points, dividing the sorted sample into two partitions in ratio 1 : 3 (lower) and 3 : 1 (upper).

Remark 2 *The average is sensitive to the superior values, unlike the median.*