STATISTICS

Department of Pathophysiology
Faculty of Medicine in Pilsen

Statistical sets

1. **Cardinal set** (population) contain all existing elements (individuals, measures) which fulfil conditions for involvement into the set.
2. **Sample set** involve only a subsets of the cardinal sets.

The sample set has to be large enough and representative.

Statistical variables

1. **Cardinal variables** describe the cardinal set.
2. **Sample variables** describe the sample set. They are used as an estimation of the cardinal statistical variables.

Statistical variables

1. **Size of the data set *(n)*:** number of samples involved in the data set
2. **Mean *($\bar{x}$)***

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \text{or} \qquad \bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

3. **Median:** the middle value if samples are arranged in order of magnitude (for an odd number of samples) or the average of the two middle values (for even number of observations).
4. **Modus:** the value in frequency distribution that occurs most often (two or more values with the same frequency = bimodal, resp. polymodal set)

5. **Variance *(s^2, σ^2)*:** total sum of the squared deviations from the mean, divided by number of samples (*n*), in the case of sample set divided by (*n-1*)

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

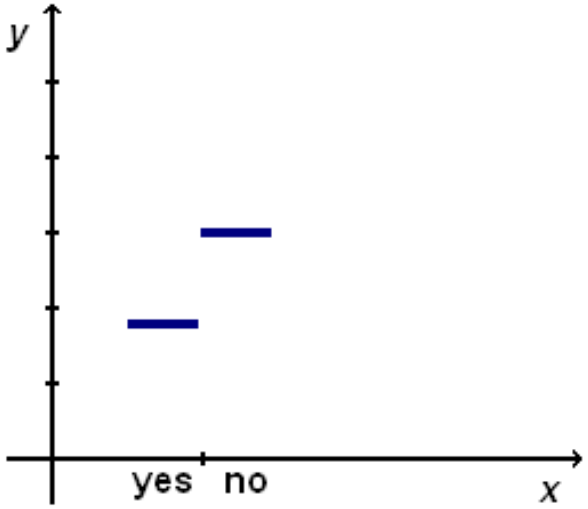
6. **Standard deviation *(s, σ)*:** square root of variance

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

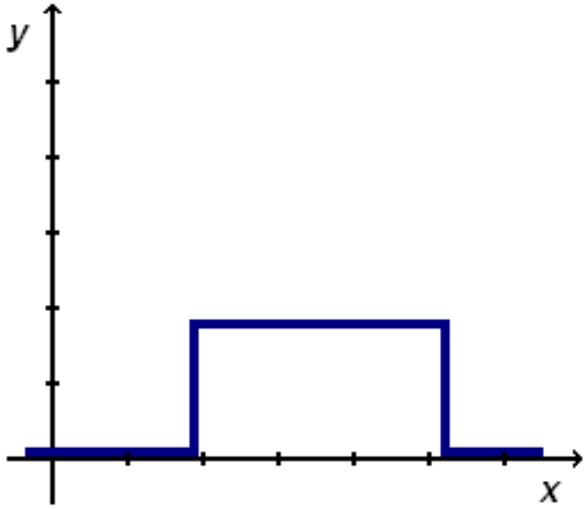
7. **Standard error of the mean:** *standard deviation divided by the square root of n.*

$$= \frac{s}{\sqrt{n}}$$

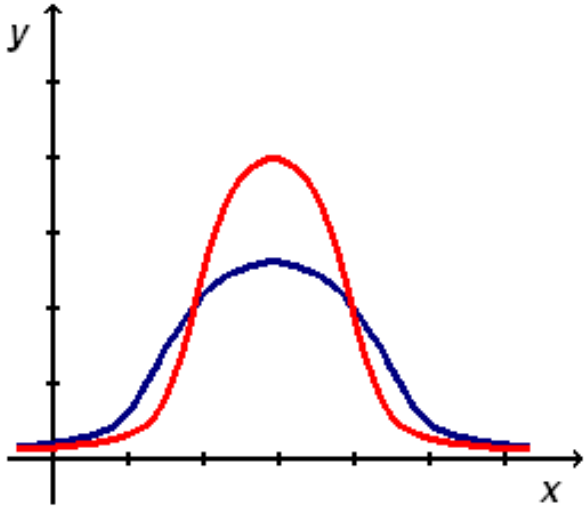
Distribution of the set



binomial



equal



Gaussian = normal

(mean = median = modus)

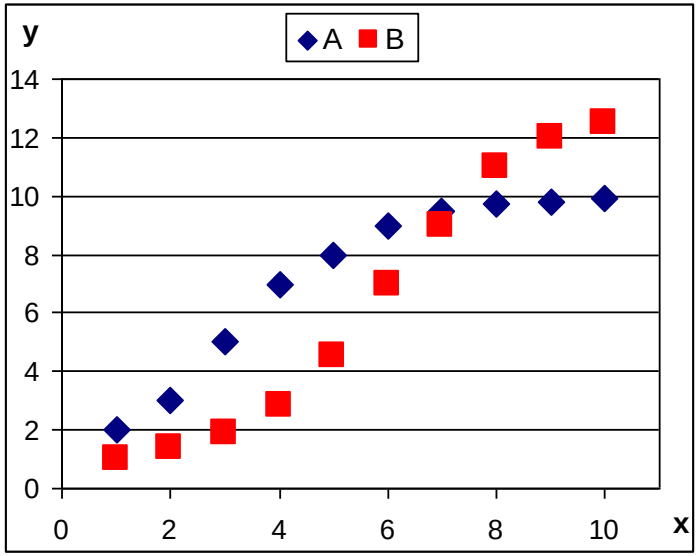
The interval $\bar{x} \pm s$ contains 2/3 of all elements, the interval $\bar{x} \pm 3s$ contains almost all elements – useful for elimination of distant values.

x … element value

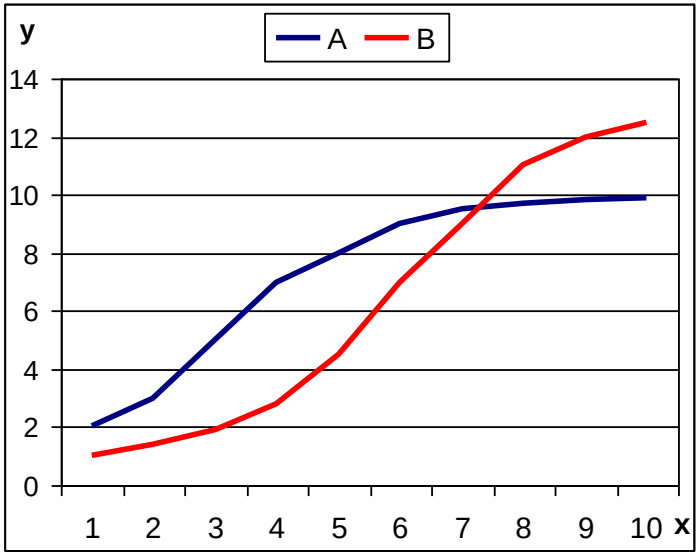
y … frequency of elements with given value

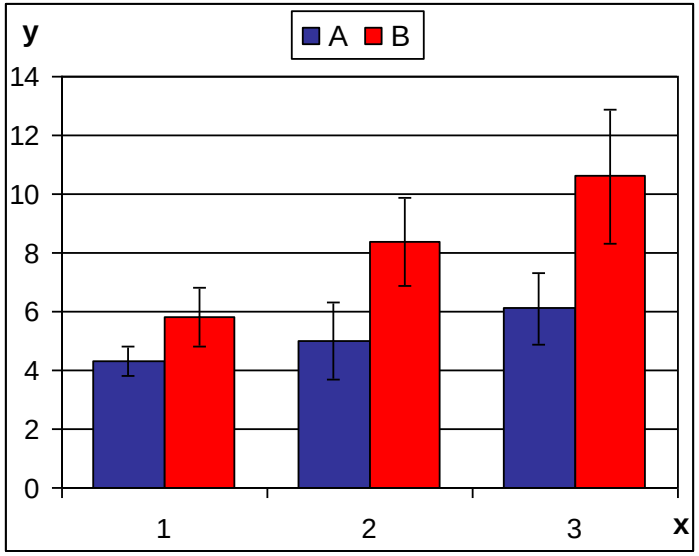
Graphs

point

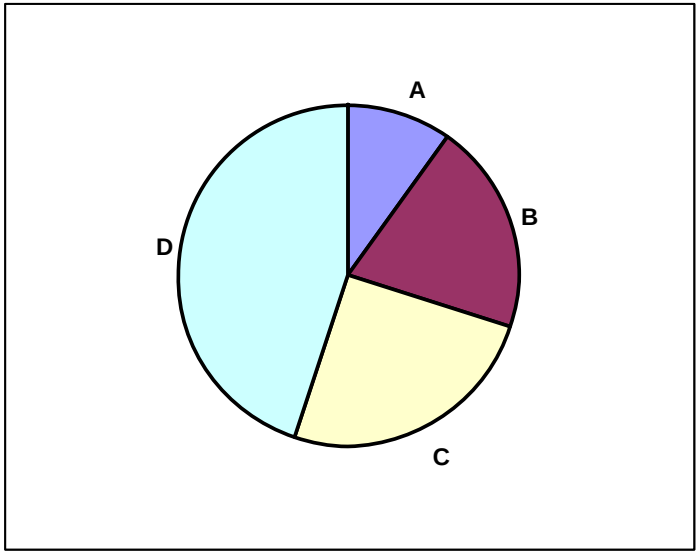


line diagram





column graph = histogram



pie graph

Hypothesis testing

Parametric tests: basic condition for their use is normal distribution of both compared sets (Gaussian distribution curve).
e.g.: t-test, U-test

Nonparametric tests: normal distribution is not required.
e.g.: Mann-Whitney test, sign test

- Zero hypothesis formulation – usually negative (the sets are not different, the differences are not statistically significant).
- With the tests we find the level of significance (*p*), it means the probability that zero hypothesis rejection is incorrect. Lower *p* value indicates higher statistical significance of the differences between the groups
- Statistically significant differences if $p < 0,05$
- Statistically highly significant differences if $p < 0,01$

Pairs of data comparison: paired tests (paired t-test, sign test)

t-test

condition for use: parametric test
(Similar is U-test, which has an additional condition of minimal set size 30 elements.)

Count the t-value

$$t = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \cdot \frac{n_1+n_2}{n_1 \cdot n_2}}}$$

x_1, x_2 … means of both sets
n_1, n_2 … size of both sets
s_1, s_2 … standard deviations of both sets

Degrees of freedom = n_1+n_2-2

Counted *t* compare with values in the table (*t tab*) in the line for appropriate degree of freedom. If *t>ttab*, the differences are statistically significant on appropriate level of significance. It means, the probability of incorrect rejection of zero hypothesis is lower than *p*.

Degrees of freedom	Statistical significance level (*p*)				
	0.5	**0.1**	**0.05**	**0.02**	**0.01**
1	**1.000**	**6.314**	**12.706**	**31.821**	**63.657**
2	**0.816**	**2.920**	**4.308**	**6.965**	**9.925**
3	**0.765**	**2.353**	**3.182**	**4.541**	**5.841**
4	**0.741**	**2.132**	**2.776**	**3.747**	**4.604**
5	**0.727**	**2.015**	**2.571**	**3.365**	**4.032**
6	**0.718**	**1.943**	**2.447**	**3.143**	**3.707**
7	**0.711**	**1.895**	**2.365**	**2.998**	**3.409**
8	**0.706**	**1.860**	**2.306**	**2.896**	**3.335**
9	**0.703**	**1.833**	**2.262**	**2.821**	**3.250**
10	**0.700**	**1.812**	**2.228**	**2.764**	**3.169**
11	**0.679**	**1.796**	**2.201**	**2.716**	**3.106**
12	**0.695**	**1.782**	**2.177**	**2.681**	**3.055**
13	**0.694**	**1.781**	**2.160**	**2.650**	**3.012**
14	**0.692**	**1.761**	**2.145**	**2.624**	**2.977**
15	**0.691**	**1.753**	**2.131**	**2.602**	**2.947**
20	**0.687**	**1.725**	**2.086**	**2.528**	**2.845**
25	**0.684**	**1.708**	**2.060**	**2.485**	**2.787**
30	**0.683**	**1.697**	**2.042**	**2.457**	**2.750**

Statistical studies

1. **Retrospective studies** - problematic quality of data
   - less time consuming

2. **Prospective studies** - better data collection
   - time consuming and more expensive
   - problematic in rare diseases

	♂	♀
n		
$\bar{x}$		
median		
modus		
σ^2		
σ		
standard error of the mean		

♂	body weight (g)	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
rat 1			
rat 2			
⋮			
rat n			

♀	body weight (g)	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
rat 1			
rat 2			
⋮			
rat n			

Variance

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

Standard deviation

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

Standard error of the mean $= \frac{s}{\sqrt{n}}$

t-test

$$t = \frac{\left| \bar{x}_1 - \bar{x}_2 \right|}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \cdot \frac{n_1 + n_2}{n_1 \cdot n_2}}}$$

Degrees of freedom	Statistical significance level (*p*)				
	0.5	**0.1**	**0.05**	**0.02**	**0.01**
1	**1.000**	**6.314**	**12.706**	**31.821**	**63.657**
2	**0.816**	**2.920**	**4.308**	**6.965**	**9.925**
3	**0.765**	**2.353**	**3.182**	**4.541**	**5.841**
4	**0.741**	**2.132**	**2.776**	**3.747**	**4.604**
5	**0.727**	**2.015**	**2.571**	**3.365**	**4.032**
6	**0.718**	**1.943**	**2.447**	**3.143**	**3.707**
7	**0.711**	**1.895**	**2.365**	**2.998**	**3.409**
8	**0.706**	**1.860**	**2.306**	**2.896**	**3.335**
9	**0.703**	**1.833**	**2.262**	**2.821**	**3.250**
10	**0.700**	**1.812**	**2.228**	**2.764**	**3.169**
11	**0.679**	**1.796**	**2.201**	**2.716**	**3.106**
12	**0.695**	**1.782**	**2.177**	**2.681**	**3.055**
13	**0.694**	**1.781**	**2.160**	**2.650**	**3.012**
14	**0.692**	**1.761**	**2.145**	**2.624**	**2.977**
15	**0.691**	**1.753**	**2.131**	**2.602**	**2.947**
20	**0.687**	**1.725**	**2.086**	**2.528**	**2.845**
25	**0.684**	**1.708**	**2.060**	**2.485**	**2.787**
30	**0.683**	**1.697**	**2.042**	**2.457**	**2.750**