MEDICAL STATISTICS

Physiology

Anatomy

Chemistry

…

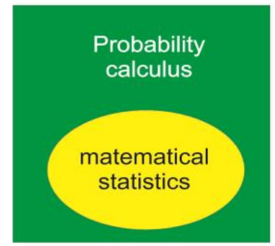
No any doubt

Statistics

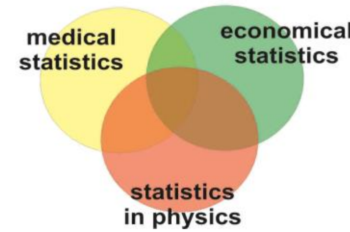
?

Medical statistics

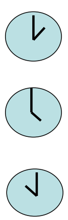
Theory: matematics



Practice: applied statistics (examples)



Example: body temperature



36.7 °C

36.9 °C

36.6 °C

36.7 °C

36.9 °C

36.5 °C

1. Inaccuracy of the measurement.
2. ***Daily fluctuation!!!***

3. Biological variability!!!

The measured value is not constant!

Measured value: 37.0 °C.

Is it healthy or not?

Another examples

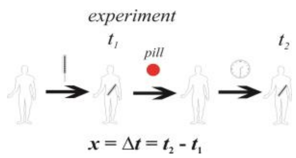
RBC: 4.5×10^{12} 1/l (<u>3.9-5 x10^{12} 1/l</u>) → normal range?

The new method in therapy is better then the old one or not?

<u>How can we prove that a medicine decreases the fewer or not?</u>

Questions!

How can we answer?



1. $\Delta t < 0$

Medicine is effective.

Medicine is not effective: fluctuation, patient will be better, etc.

?

2. $\Delta t >= 0$

Medicine is not effective.

Medicine is effective: fluctuation, doesn't increase further, etc.

Consequence:

The answer is not so simply!

(Not only the medicine influence the body temperature.)

Variables

variable	range	type	variable type	
height	~50 cm … ~250 cm	real number	numerical	continuous
no. of teeth	0 .. 32	integer		discrete
blood type	A, B, AB, 0	letters	categorical	nominal
severity of cancer	1 … 4	integer		ordinal

Descriptive statistics!

Description of a variable

- Type
- Possible values
- Occurrence of the values

Numerical variables

Name	***Continuous***	***Discrete***
Definition	Infinitely large no. of values in a certain range	Only finite number of values
Example	Height, temperature, pressure …	No. of teeth, no. of children …

Categorical variables

Name	***Nominal***	***Ordinal***
Definition	No order among the values	There is a certain order
Example	Gender, blood-type …	Severity of the illness, strength of pain …

Determination of the possible values

- Continuous : giving a possible range.
  - » e.g.: height from ~50 cm - to ~ 250 cm
- Another : listing the values, if it is possible
  - » E.g.: blood type: A, B, AB, 0

Occurence

Observation: The occurrence of the values are not the same!

↓

Trial: experiment, observation, data collection.

Deal with only the case, when the trial may be repeated!

Outcome: result of one trial. (e.g.: height of a student)

Population

How many people?

↓

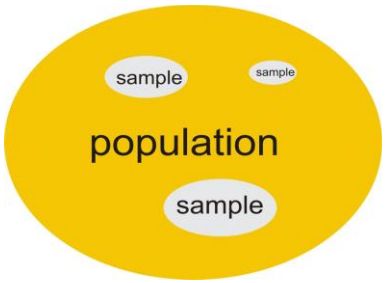
As many as possible.

↓

Ideal case: All of the people → ***population***

Sample

A smaller portion of the population.



n: no. of the elements (people) in the sample.

x: the tested variable (quantity)

x_i: i-th element from the sample

Selection of the sample

Main principle: **<u>Random sample</u>**

Medical statistics: <u>if there is no any reason to exclude,</u>
must be random!

Occurrence

Frequency (*k*): no. of occurrence in the sample.

k_i: no. of occurrence of the i-th value in the sample.

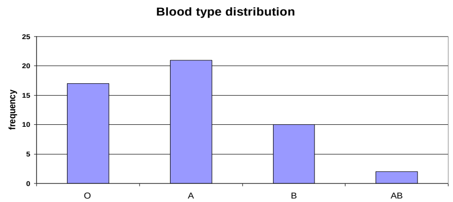
$$n = \sum_i k_i$$

Frequency distribution

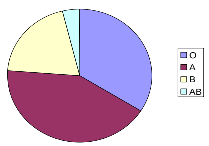
Frequency as the function of the possible values.

Blood-type	**0**	**A**	**B**	**AB**	total
frequency	17	21	10	2	50

Presentation



Bar-chart



Pie-chart

Relative frequency, proportion

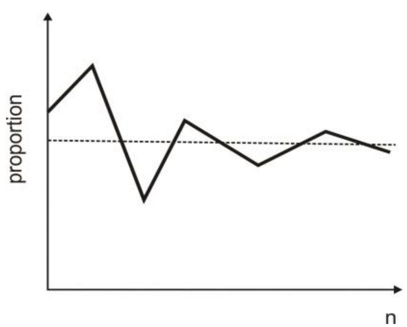
The ratio of the frequency and the total no. of the elements.

$$\sum_i \frac{k_i}{n} = \frac{1}{n}\sum_i k_i = \frac{1}{n}\times n = 1$$

Frequently it is given as percentage:

$$\frac{k_i}{n}\times 100\%$$

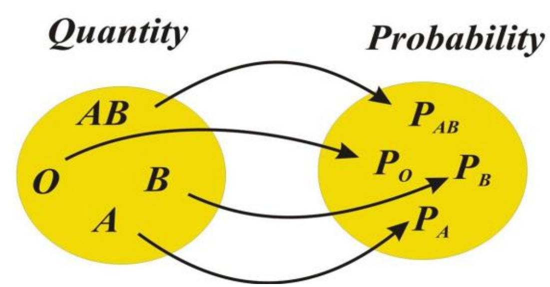
Probability (*P*)



If *n* is infinite the name of the proportion is the probability.

Probability (P): proportion in the population.

Probability distribution



Properties of the probability

$$0 \leq P \leq 1 \longrightarrow$$

P = 0 - never occur

P = 1 - always occur

Example: blood- type

$$P_A + P_B + P_{AB} + P_0 = 1 \longrightarrow \sum_i P_i = 1$$

(exclusive events)

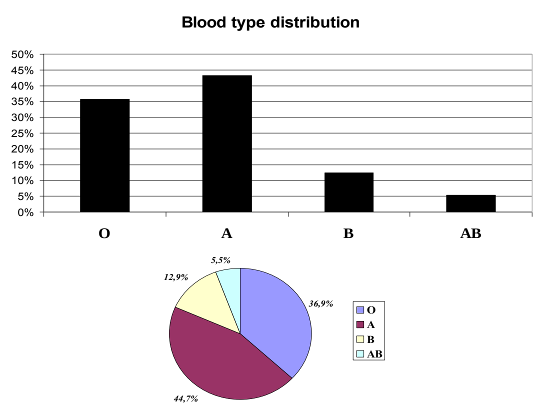
Probability and proportion

Sample	**Population**
n is finite	$n = \infty$
proportion	probability

Probability very frequently is unknown!

We usually use proportion instead of the probability.

Presentation



Continuous quantity

Infinite no. of possible values!!!

Class: a short interval in the whole range.

Class-width: the length of the class.

Frequency: no. of elements in the given class.

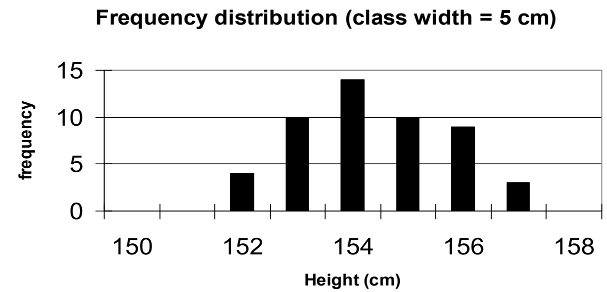
Like a discrete value!

Example

1	160 cm
2	181 cm
3	175 cm
4	163 cm
5	165 cm
6	179 cm
7	164 cm
8	185 cm
9	177 cm
10	168 cm

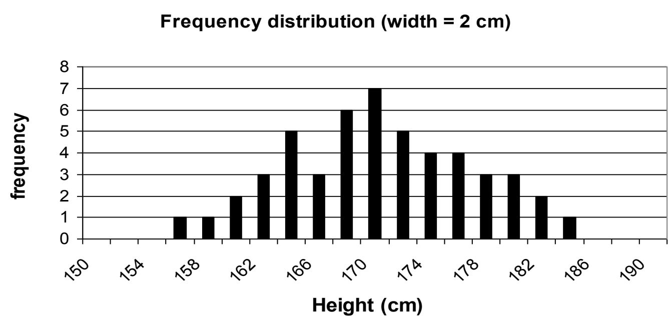
class	k_i
160-164	3
165-169	2
170-174	0
175-179	3
180-184	1
185-189	1

Presentation



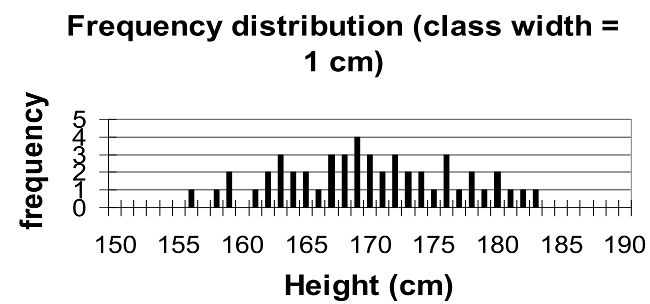
5 cm is too large!

Decrease the width!



Observation: frequency decreases!

Presentation



Reason: n is too small!

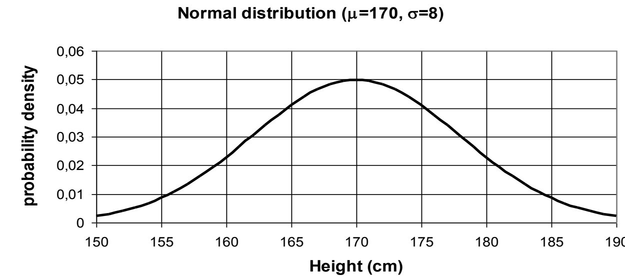
Consequence

Class-width | No. of classes | Frequencies

We must increase the no. of the elements!

Normal distribution

If n and no. of classes are infinite!



Theoretical description

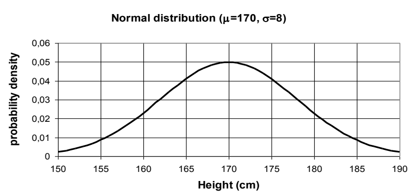
Normal or Gauss-distribution

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

Parameters:

μ – expected value or mean

σ – theoretical standard deviation

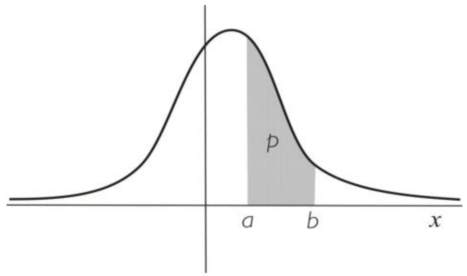


Meaning of the parameters

μ **(mean):**
the value belonging to the maximum of the curve.

σ **(theoretical standard deviation):**
the average deviation of the data from the μ.

Probability



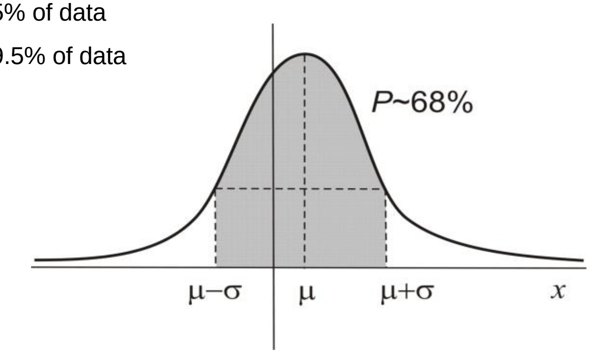
***P* is the probability that x is in the (a,b) interval.**

Standard deviation

($\mu \pm \sigma$) ~ 68% of data

($\mu \pm 2\sigma$) ~ 95% of data

($\mu \pm 3\sigma$) ~ 99.5% of data



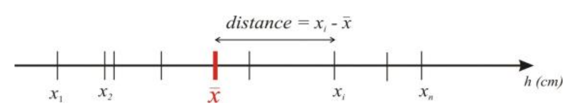
Normal distribution

Theoretical distribution that describes the population. In practice usually we don't know the parameters of this.

We usually have a ***random sample*** from the population.

We must estimate the parameters!

Estimation of the μ



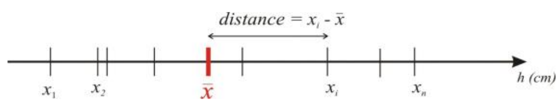
average: must be in the center of the data range.

$$\sum_i (x_i - \overline{x}) = 0 \longrightarrow \overline{x} = \frac{\sum_i x_i}{n}$$

Estimation of the σ

σ = average deviation of the data from the μ.

s (standard deviation) = average deviation of the elements from the average.



$$Q_x = \sum_i (x_i - \bar{x})^2 \geq 0$$

Standard deviation

$$s = \sqrt{\frac{Q_x}{n-1}}$$

s: the average deviation of the elements from the average.

($\bar{x}$±s) ~ 68%

($\bar{x}$±2s) ~ 95%

($\bar{x}$±3s) ~ 99.5%

Relation of parameters

Sample	n → ∞	Population
average	→	μ
s	→	σ

Question of the week!

How can we estimate the μ and the σ?