MEDICAL STATISTICS

Physiology

Anatomy

Chemistry

…

Yes

Statistics

?

Before modern mathematics

Medical herbs

Observation

Some of them is more effective in curing a certain illness.

The occurrence of being better is higher.

Examples

RBC: 4.5×10^{12} 1/l (3.9-5×10^{12} 1/l) $\rightarrow$ normal range?

How can we prove that a medicine decreases the fewer or not?

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

↓

Questions!

Answers

Experiment:

↓

Measure a quantitiy.

Different possible values.

↓

Variable

Readers

↓

The answer is good or not?

?

Description of a variable

- Type
- Possible values
- Occurance of the values

Type of quantities

Numerical variables

Name	Continuous	Discrete
Definition	All values are valid in a certain range	Only finite number of values
Example	Height, temperature, pressure …	No. of teeth, No. of children …

Categorial variables

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

Determination of the possible values

- Continuous : giving a possible range.
  - » e.g.: height from ~60 cm - to ~ 250 cm
- Another : listing the values, if it is possible
  - » E.g.: blood type: A, B, AB, 0

Occurance

Observation: The occurance of the value 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

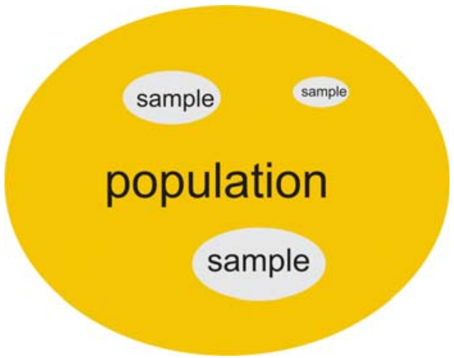
How many people?

As many as possible.

Ideal case: All of the people $\rightarrow$ ***population***

Sample

A smaller portion of the population.



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

x: the variable (quantity) testing

x_i: i-th element from the sample

Selection of the sample

Main principle: **Random sample**

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

Occurance

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

$\boldsymbol{k_i}$: no. of occurance of the i-th value in the sample.

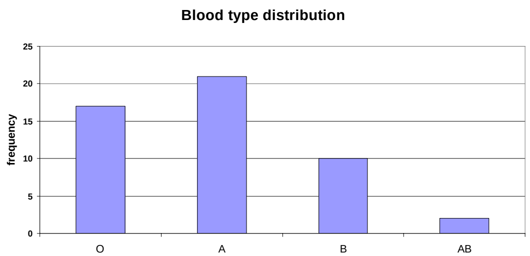
$$n = \sum_i k_i$$

Frequency distribution

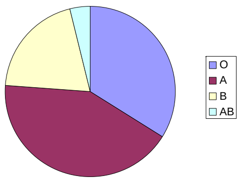
Frequency as the function of the values.

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

Presentation



Bar-chart



Pie-chart

Relative frequency, proportion

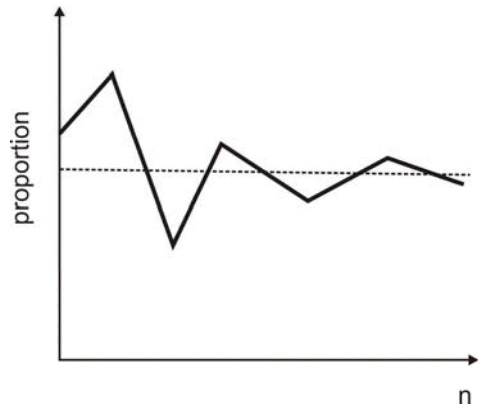
The ratio of the frequency and the total no. of the elements.

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

Frequently, it is given as percentage:

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

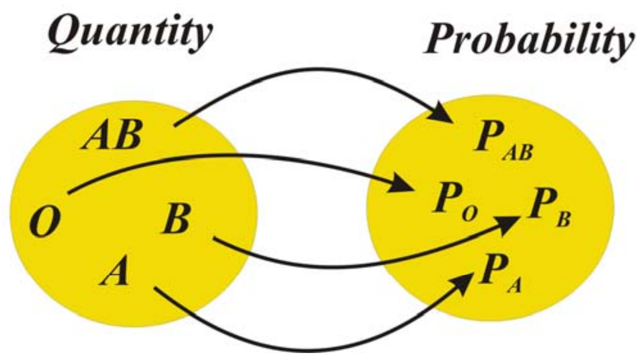
Probability (P)



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

Probability (P): proportion in the population.

Probability is a function



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)

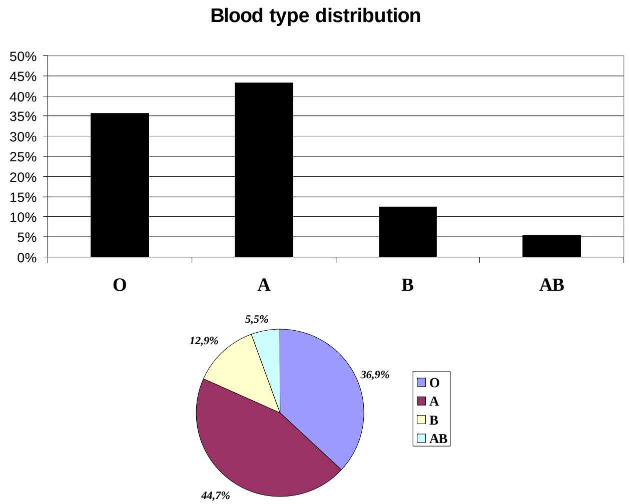
Probability and proportion

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

Probability is usually unknown!

Normally we use proportion instead of the probabilty.

Presentation



Continuous quantity

Infinite no. of possible values!!!

Class: a short interval in the whole ange.

Class-width: the length of the class.

Frequency: no. of element belonging to the given class.

↓

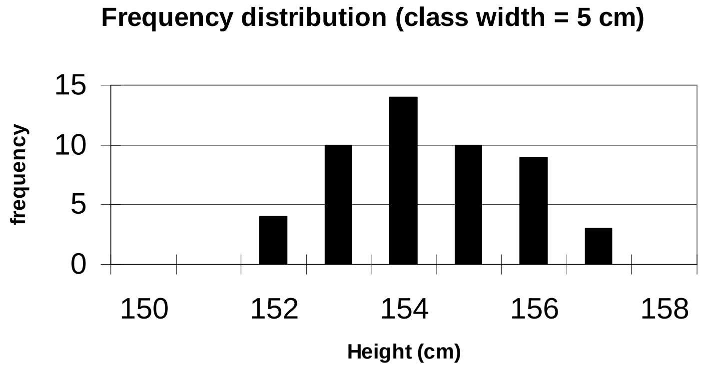
Like a discrete value!

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

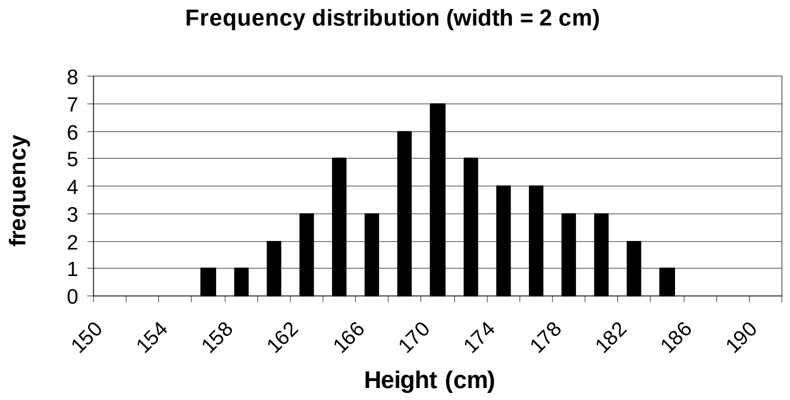
class	k_i
160-164	3
165-169	2
170-174	0
175-179	3
180-184	1
185-189	1

Presentation



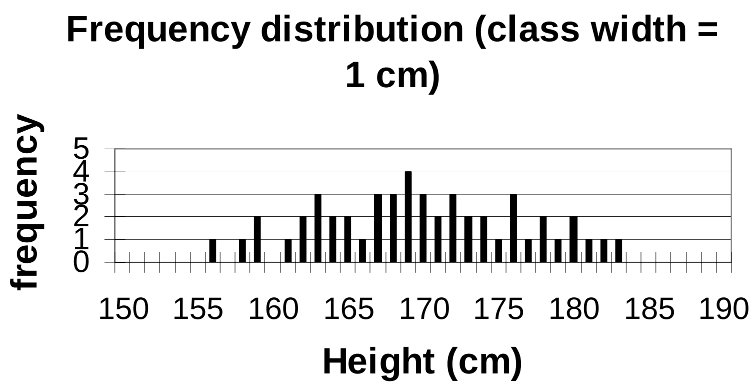
5 cm is too large!

Decrease the width



Observation: frequency decreases!

Presentation



Reason: n is too small!

Consequence

Class-width

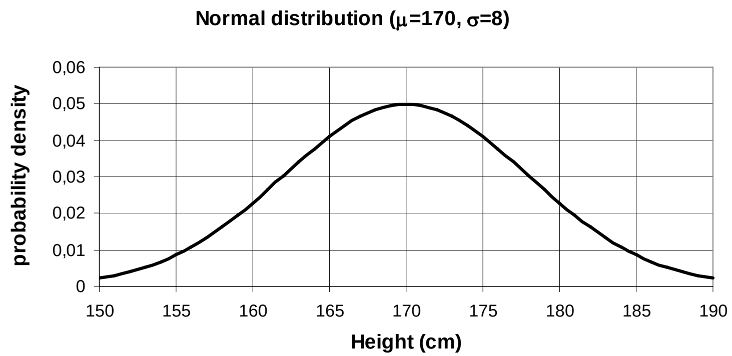
No. of classes

Frequencies

We must increase the no. of the elements!

Normal distribution

If *n* and no. of classes are infinite!



Theoretical description

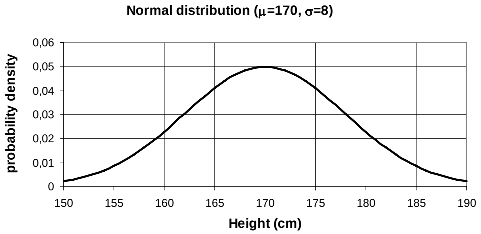
Normal or Gauss-distribution

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

Parameters:

μ – expected value or mean

σ – theoretical standard deviation

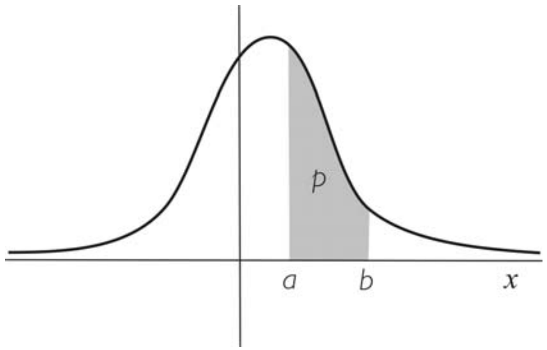


Meaning of the parameters

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



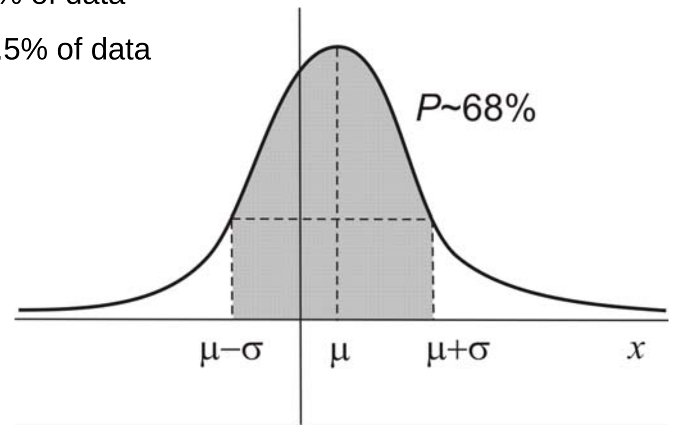
P is the probability that x is in the interval of (a,b).

Standard deviation

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

($\mu\pm2\sigma$) ~ 95% of data

($\mu\pm3\sigma$) ~ 99.5% of data



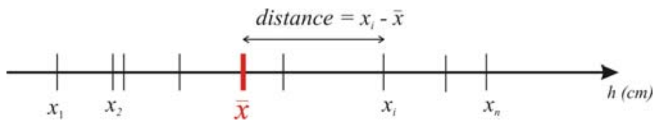
Normal distribution

Theoretical distribution! Descibre the population. In practice normally 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



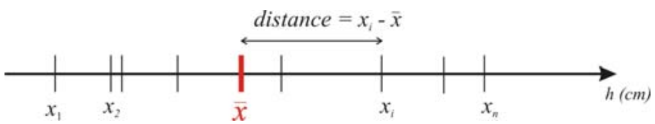
average: must be in the center of the data range.

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

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%

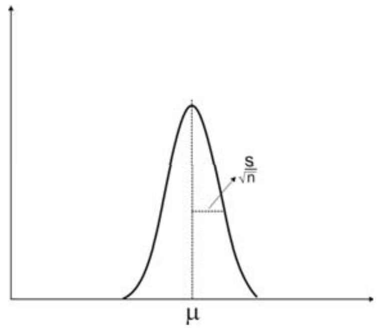
Relation of parameters

Sample	$n \to \infty$	Population
average	$\longrightarrow$	μ
s	$\longrightarrow$	σ

μ and the average

sample	average
1	170
2	168
3	166
4	173

Averages fluctuate, deviate around the μ.



Standard error

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Average deviation of the averages around the μ!

Confidence interval for μ.

$$(\bar{x} \pm s_{\bar{x}}) \sim 68\%$$

~68% is the probability that μ is in this range.

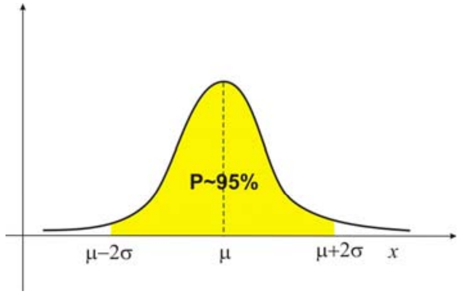
(~32% that isn't)

Information

	interval	probability	information content
$(\bar{x} \pm s_{\bar{x}}) \sim 68\%$			
$(\bar{x} \pm 2s_{\bar{x}}) \sim 95\%$			!
$(\bar{x} \pm 3s_{\bar{x}}) \sim 99.5\%$			
$(\bar{x} \pm \infty) = 100\%$	∞	P =1	= 0

Normal (reference) range

Normal distribution



Other quantites

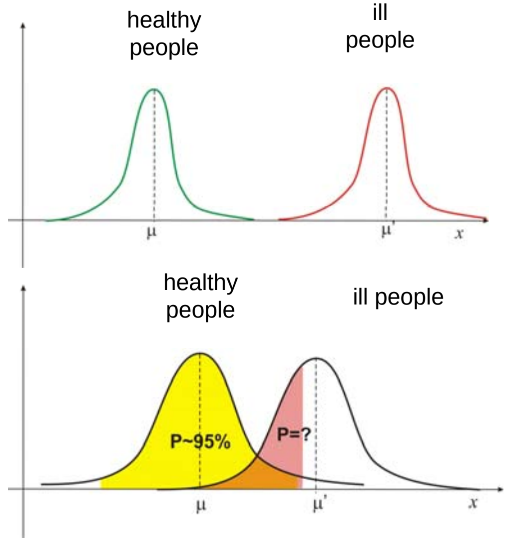
A range, that contains the 95% of the possible values.

But: 5% is the chance being out!!!

Normal range

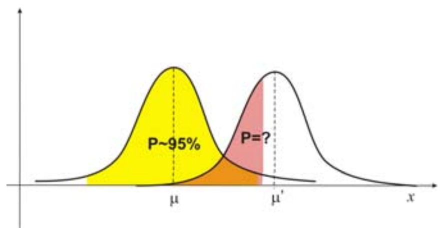
ideal case

in practice



Why is the normal range?

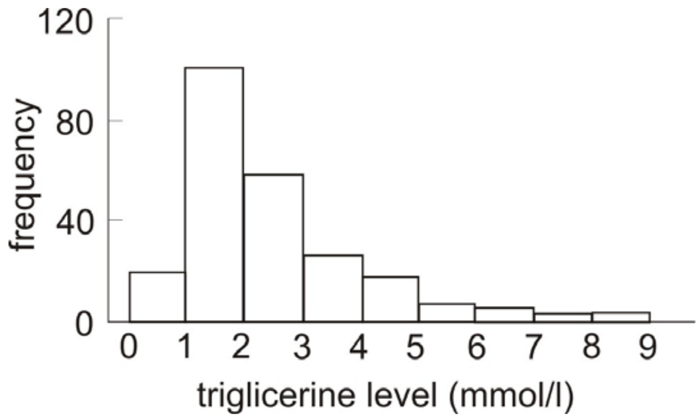
How much is the probability being ill?



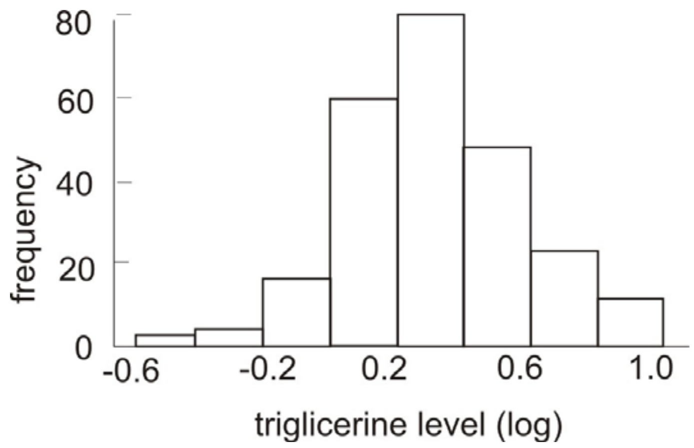
Healthy people:
n is enough large.
Estimation is good.

Ill people:
n is usually small.
Estimation is poor.

lognormal distribution

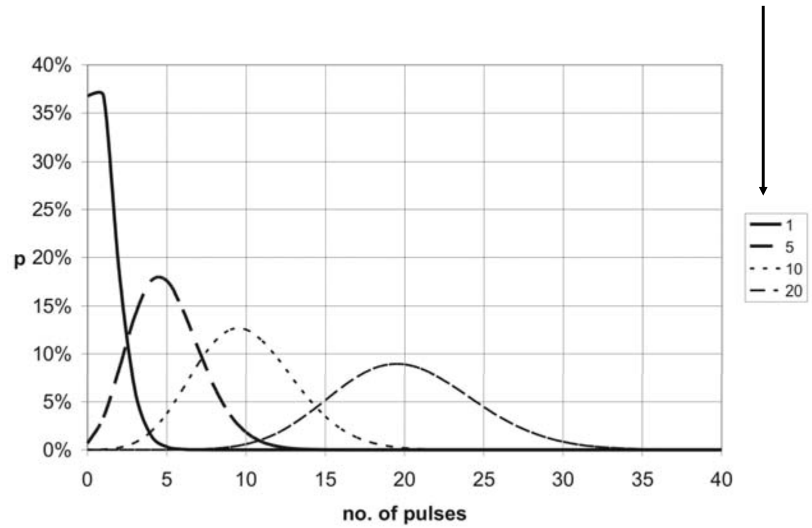


Transformation



Poisson distribution

mean:



Similarity