Hypothesis Testing & Confidence Interval Estimation

Assoc . Prof. Marius MARUSTERI
Medical Informatics and Biostatistics Dept.
UMF Tg. Mures

Mathematical Statistics is the science which seeks to explain the "mass/global" phenomena through a relatively small number of observations.
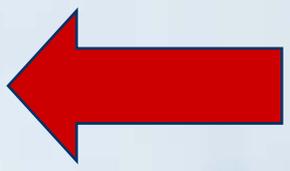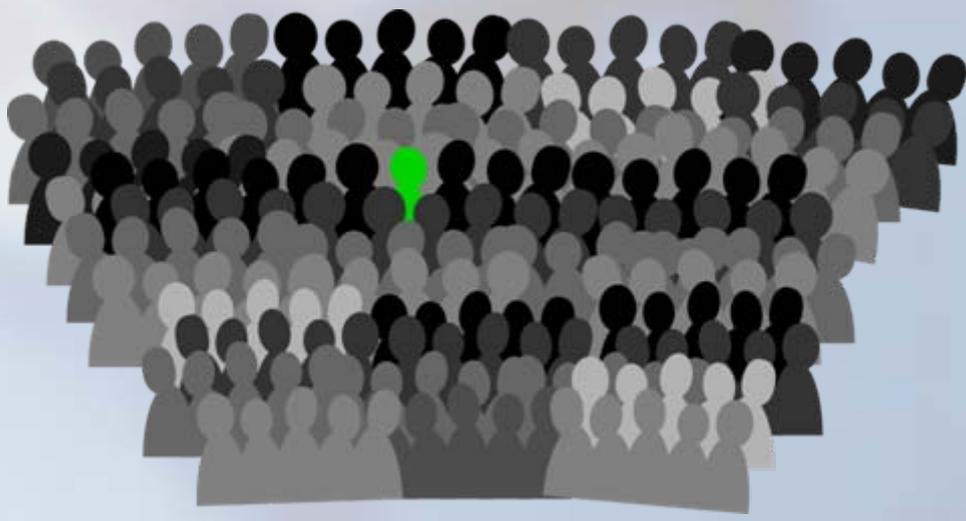
e.g. election day,
exit polls

X ± 2,5 %

Y ± 2,5 %

How is this possible?
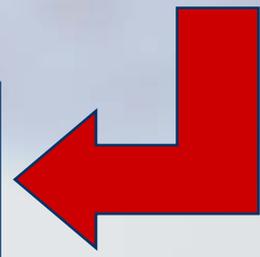
representative sample
of electors (~1400 – 21000)

All romanian electors (~ 14 mil.)

Questionnaire
(opinion poll)

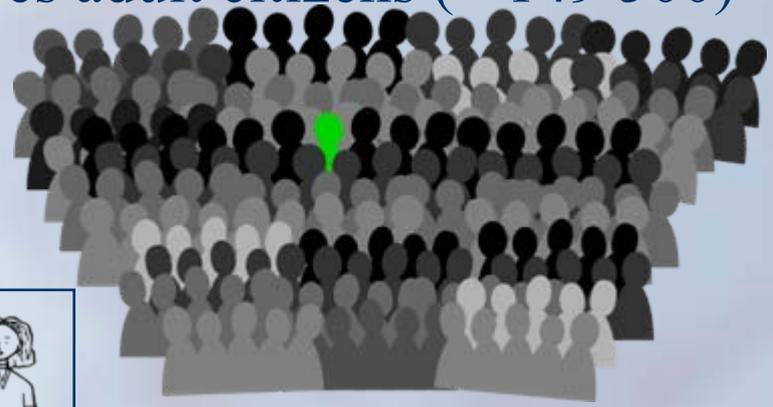# Biostatistics is a branch of statistics, specialized in investigation of biological phenomena
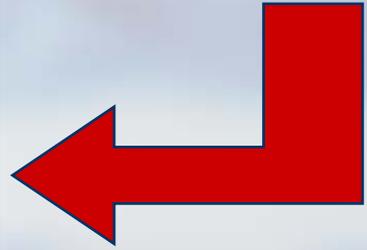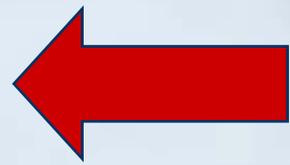
E.g. Average weight
of Tg. Mures adult citizens (~ 149 500)

Representative sample
of Tg. Mures citizens
(~1200-1400)

70,3 ± 5,7 Kg

A population is a large set of objects of a similar nature - e.g. human beings, households, readings from a measurement device - which is of interest as a whole.

Population is also used to refer to a set of potential measurements or values, **including not only cases actually observed but those that are potentially observable.**

Statistical population is a set of entities concerning which statistical inferences are to be drawn, often based for example on a random sample taken from the population.

http://en.wikipedia.org/wiki/Statistical_population

Finite population is a population that contains finite many individual (items or measurements).

Infinite population is a population that contains unlimited members.

http://www.usd.edu/~vmcbride/Stats281/Worksheet 206-1.htm

A population is not necessarily real - it may be *hypothetical* or imaginary. For example, outcomes of an experiment , that is carried out infinitely, make a hypothetical population.

http://www.statistics.com/resources/glossary/p/population.php
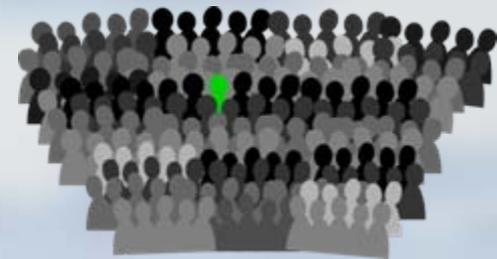
A **sample** is a subset of a population. Typically, the population is very large, making a census or a complete enumeration of all the values in the population impractical or impossible.

Successful statistical practice is based on focused problem definition. In sampling, this includes defining the population from which our sample is drawn.

As we already know, population can be defined as including all people or items with the characteristic one wishes to understand. Because there is very rarely enough time or money to gather information from everyone or everything in a population, the goal becomes finding a **representative sample** (or subset) of that population.

The sample represents a subset of manageable size.

Samples are collected and **statistics are calculated from the samples** so that one can make inferences or extrapolations from the sample to the population.

This process of collecting information from a sample is referred to as sampling.

http://en.wikipedia.org/wiki/Sample_(statistics)

Representative and unrepresentative samples

The best way to avoid a biased or unrepresentative sample is to select a **random sample**, also known as a probability sample.

A **random sample** is defined as a sample where the probability that any individual member from the population being selected as part of the sample is exactly the same as any other individual member of the population. Several types of random samples are simple random samples, systematic samples, stratified random samples, and cluster random samples.

A sample that is not random is called a **nonrandom sample** or a nonprobability sample.

- Nonprobability sampling include:
- **Convenience, Haphazard or Accidental sampling** - members of the population are chosen based on their relative ease of access.
- **Snowball sampling** - The first respondent refers a friend. The friend also refers a friend, etc.
- **Judgmental sampling** or **Purposive sampling** - The researcher chooses the sample based on who they think would be appropriate for the study.
- **Deviant Case**-Get cases that substantially differ from the dominant pattern (a special type of purposive sample)
- **Case study** - The research is limited to one group, often with a similar characteristic or of small size.
- **ad hoc quotas** - A quota is established (say 65% women) and researchers are free to choose any respondent they wish as long as the quota is met.

Where are we now...on this lecture...?

If we measure some numbers :

Statistical Population – including all people or items with the characteristic one wishes to understand.
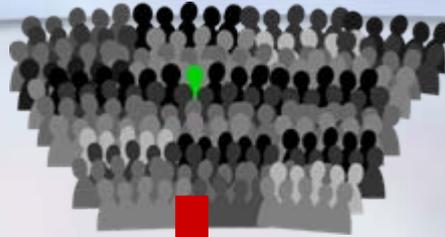
- Finite
- Infinite
- Hypothetical

Sample – a subset of a population

- Random
- Nonrandom

	Grup Test		Grup Control	
Mean		7	Mean	6.5
Standard Error		2.0155644	Standard Error	0.62361
Median			Median	6.5
Mode		#N/A	Mode	6.5
Standard Deviation		6.046693311	Standard Deviation	1.972027
Sample Variance		36.5625	Sample Variance	3.888889
Kurtosis		-0.541523434	Kurtosis	-0.38145
Skewness		0.940483152	Skewness	-0.27166
Range		16.5	Range	6.5
Minimum		1.5	Minimum	3
Maximum		18	Maximum	9.5
Sum		67.5	Sum	65
Count		9	Count	10
Largest(1)		18	Largest(1)	9.5
Smallest(1)		1.5	Smallest(1)	3
Confidence Level(95.		4.647902932	Confidence Level(9	1.410704
CV		80.622577	CV	30.33887

Descriptive statistics are used to describe the main features of a collection of data in **quantitative terms.**

Descriptive statistics are distinguished from **inferential/inductive statistics** in that they aim to **quantitatively summarize a data set**, rather than being used to support statements about the population that the data are thought to represent.

Inferential/inductive statistics

INFERENTIAL/Inductive STATISTICS

Definition(s):

Sample

Population

Statistical inference or **statistical induction** comprises the use of statistics and (random) sampling to make inferences concerning some unknown aspect of a population.

Inference is the act or process of deriving a logical consequence conclusion from premises.

Statistical (inferential) analysis on the sample

Population after statistical analysis

INFERENTIAL/Inductive STATISTICS

Definition(s):

Statistical inference - inference about a population from a random sample drawn from it or, more generally, about a random process from its observed behavior during a finite period of time. It may includes:

1. Point estimation involves the use of sample data to calculate a **single value** (known as a statistic) which is to serve as a "best guess" for an unknown (fixed or random) population parameter (e.g. RR=3,72)

2. Interval estimation is the use of sample data to calculate an interval of possible (or probable) values of an unknown population parameter, in contrast to point estimation, which is a single number(e.g. CI95% for RR is 1,57...7,92)

INFERENTIAL/Inductive STATISTICS

Questions:

Can I know it ever ??? Probably NO !!! ☺

Where is the **"true value"**

How confident should I be that other experiments will obtain "similar" results???

Interval estimation with a specific level of confidence (e.g. 95%)

± 5,7 Kg

70,3 Kg

What (average) result I obtained ?

Point estimation, the result of my experiment
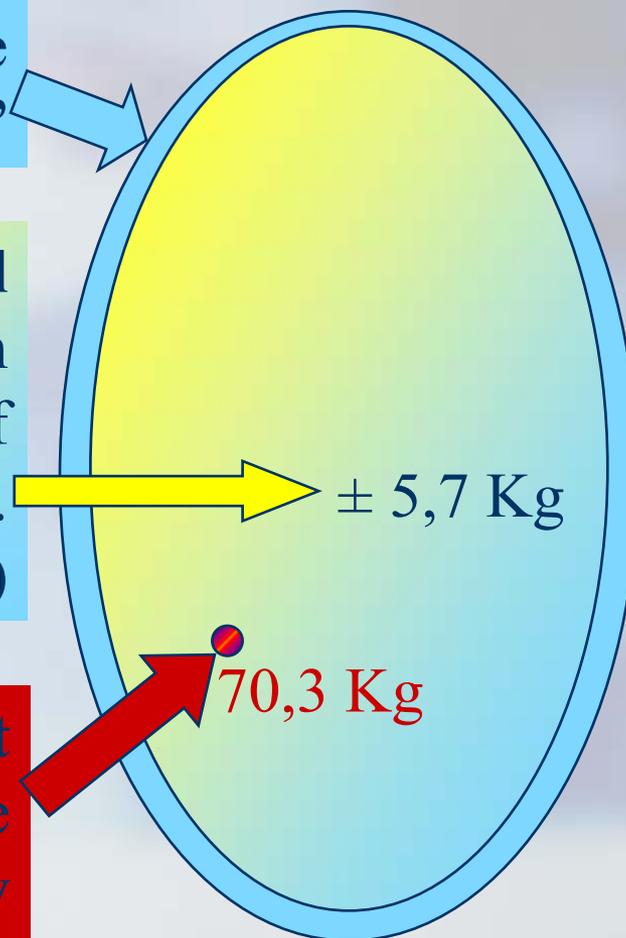
Definition(s):

3. A **statistical hypothesis test** is a method of making statistical decisions using experimental data. It is sometimes called **confirmatory data.** These decisions are almost always made using null-hypothesis tests

The **null hypothesis** (H0) formally describes some aspect of the statistical behaviour of a set of data; this description is treated as valid unless the actual behaviour of the data contradicts this assumption. Thus, the null hypothesis is contrasted against another hypothesis, so-called **"alternative hypothesis"**.

4. **Prediction/forecast.** A **prediction** is a statement or claim that a particular event will occur in the future in more certain terms than a forecast. **Forecasting** is the process of estimation in unknown situations. Prediction is a similar, but more general term. **Risk** and **uncertainty** are central to forecasting and prediction.

Why we need it?

"*A statistically significant difference*" simply means there is **statistical evidence** that **there is a difference**; it **does not mean the difference is necessarily large, important, or significant in the common meaning of the word**.

	Temp. control	Temp. test		Kg control	Kg test
	37.9	36.7		83.0	75.0
	36.1	39.2		77.0	90.0
	37.2	36.5		92.0	80.0
	36.8	39.5		68.0	89.0
	37.0	38.0		80.0	81.0
	36.6	37.0		70.0	83.0
	37.5	39.1		90.0	87.0
	36.5	36.2		86.0	100.0
	37.8	38.2		74.0	70.0
	36.2	37.6		88.0	85.0
	37.4	39.0		72.0	95.0
Average	37.0	37.9		80.0	83.0

Which means are **signifficantly different** ???

INFERENTIAL/Inductive STATISTICS

Why we need it?

We need statistical significance tests in order to conclude that something has "*statistical significance*". *No one can make statistical decisions based on "his eyes" or "his prievious experience(s)".*

	Temp. control	Temp. test		Kg control	Kg test
	37.9	36.7		83.0	75.0
	36.1	39.2		77.0	90.0
	37.2	36.5		92.0	80.0
	36.8	39.5		68.0	89.0
	37.0	38.0		80.0	81.0
	36.6	37.0		70.0	83.0
	37.5	39.1		90.0	87.0
	36.5	36.2		86.0	100.0
	37.8	38.2		74.0	70.0
	36.2	37.6		88.0	85.0
	37.4	39.0		72.0	95.0
Average	37.0	37.9		80.0	83.0
		p<0.05			p<0.05
pF	0.049	YES	pF	0.94	NO
pT	0.040	YES	pT	0.18	NO

"The Owls are not what they seem !"

Twin Peaks

INFERENTIAL/Inductive STATISTICS

What we need to know before we start the analysis ?

Data type(s):

Quantitative (numerical) data

•**Discrete data**, if there are only a **finite number of values possible** or if there is **a space on the number line between each 2 possible values (e.g old fashioned mercur based termomether).**
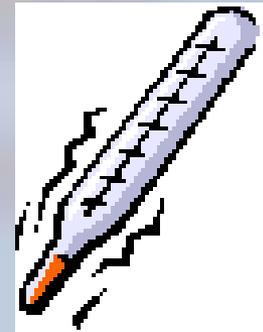
• **Continuous data** makes up the rest of numerical data. This is a type of data that is usually associated with **some sort of advanced physical measurement.**

INFERENTIAL/Inductive STATISTICS
What we need to know before we start the analysis ?

Data type(s):

Qualitative (categorical) data

• **Binary (logical) data** - a basic type of categorical data (e.g. positive/negative; present/absent etc).

• On more complex categorical data, the first (and weakest) level of data is called **nominal level data**. **Nominal** level data is made up of values that **are distinguished by name only**. There is **no standard ordering scheme** to this data (e.g. romanian, hungarian, scotish groups of people etc.).

• The second level of data is called **ordinal level data**. **Ordinal** level data is similar to nominal level data in that the data is **distinguished by name**, but **it is different than nominal level data** because there is an **ordering scheme (e.g. small, medium and high level smokers).**.

INFERENTIAL/Inductive STATISTICS
What we need to know before we start the analysis ?

Data type(s)/some exercises:

Nr. crt	Varsta	Cancer	Fumator	Tip fumator	TA sistolica (mm Hg)	TA diastolica (mm H)
1	34	0	1	B	130	90
2	28	0	0	0	120	80
3	65	1	1	A	180	100
4	45	0	0	0	145	90
5	48	1	1	C	160	110
6	54	1	1	C	150	100
7	46	0	1	A	150	90
8	58	0	0	0	160	110
9	73	1	1	C	180	100
10	29	0	0	0	130	70
11	60	1	0	0	160	100
12	39	0	1	B	140	90
13	48	1	0	0	150	80
14	37	0	0	0	130	60
15	53	1	1	A	160	100
16	28	1	0	0	120	70
17	39	1	1	B	130	80
18	49	1	0	0	140	90
19	51	0	1	A	160	100

INFERENTIAL/Inductive STATISTICS

What we need to know before we start the analysis ?

Data type(s)/some exercises – "transforming data":

Nume	varsta	sex	mediu	hoehn-yahr
Balazs Ma	57	F	R	3
Markai Edi	56	M	U	2
Brustur Cri	75	F	U	1
Flueras Ar	65	F	R	3
Suteu Ada	56	M	R	3
Bucur Ron	55	M	U	3
Kacso Bel	66	M	U	4
Ungureanu	50	M	U	1
Procopie E	77	F	U	4
Andreica N	67	M	U	2
Magyari El	77	F	R	1
Suciu Ioan	48	M	U	4
Mazarini C	70	M	U	5
Bogati Eni	54	F	U	5
Ratiu Ana	79	F	R	5
Ogrean El	64	F	U	2
Miko Iulian	81	F	U	5
Rend Roza	72	F	R	4
Gheorghe .	68	F	R	4
Gherman A	72	F	R	3
Moldovan i	78	M	R	5
Teleki Eva	74	F	U	5

Nume	varsta	grupa de var	sex	mediu
Balazs Ma	57	>=50	F	R
Markai Edi	56	>=50	M	U
Brustur Cri	75	>=70	F	U
Flueras Ar	65	>=60	F	R
Suteu Ada	56	>=50	M	R
Bucur Ron	55	>=50	M	U
Kacso Bel	66	>=60	M	U
Ungureanu	50	>=50	M	U
Procopie E	77	>=70	F	U
Andreica N	67	>=60	M	U
Magyari El	77	>=70	F	R
Suciu Ioan	48	<50	M	U
Mazarini C	70	>=70	M	U
Bogati Eni	54	>=50	F	U
Ratiu Ana	79	>=70	F	R
Ogrean El	64	>=60	F	U
Miko Iulian	81	>=80	F	U
Rend Roza	72	>=70	F	R
Gheorghe .	68	>=60	F	R
Gherman A	72	>=70	F	R
Moldovan i	78	>=70	M	R
Teleki Eva	74	>=70	F	U

INFERENTIAL/Inductive STATISTICS

What we need to know before we start the analysis ?

Data type(s):

Raw data:

Qualitative (categorical) data
•**Contingency tables**
In statistics, **contingency tables** are used to record and analyse the relationship between two or more variables (e.g one independent and one dependent variable), most usually categorical variables.
It may be **2x2 or nxn** contingency tables.

Nr. crt	Varsta	Cancer	Fumator	Tip fumator	TA sistolica (mm Hg)	TA diastolica (mm H
1	34	0	1	B	130	90
2	28	0	0	0	120	80
3	65	1	1	A	180	100
4	45	0	0	0	145	90
5	48	1	1	C	160	110
6	54	1	1	C	150	100
7	46	0	1	A	150	90
8	58	0	0	0	160	110
9	73	1	1	C	180	100
10	29	0	0	0	130	70
11	60	1	0	0	160	100
12	39	0	1	B	140	90
13	48	1	0	0	150	80
14	37	0	0	0	130	60
15	53	1	1		160	100
16	28	1	0		120	70
17	39	1	1		130	80
18	49	1	0		140	90
19	51	0	1		160	100

Contingency table (2x2):
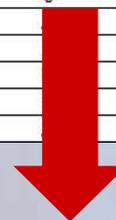
Contingency table (nxn):

	negravide	sarcini normale	disgravidii t
aldosteron	4	79	24
cortizon	15	96	37
cortizol	25	55	33

Fumator	Cancer		Total
	prezent	absent	
DA	6	4	10
NU	4	5	9
Total	10	9	19

What we need to know before we start the analysis ?

Data type(s): paired or unpaired data?

Paired data -values which fall normally into pairs and can therefore be expected to vary more between pairs than within pairs.

Are the data paired?

Choose a paired test when the experiment follows one of these designs:

• *You measure a variable before and after an intervention in each subject.*

• *You recruit subjects as pairs, matched for variables such as age, ethnic group or disease severity. One of the pair gets one treatment; the other gets an alternative treatment.*

• *You run a laboratory experiment several times, each time with a control and treated preparation handled in parallel.*

• *You measure a variable in twins, or child/parent pairs.*

More generally, you should select a paired test whenever you expect a value in one group to be closer to a particular value in the other group than to a randomly selected value in the other group.

H. Motulsky – The InStat guide to choosing and interpreting statistical tests , 2003, GraphPad Software

INFERENTIAL/Inductive STATISTICS

What we need to know before we start the analysis ?

Data type(s): paired or unpaired data?/some exercises

Ex.5	Concentratia medicamentului	
Nr. subiect	După 4 ore	După 8 ore
1	1	1
2	1.3	1.3
3	0.9	0.7
4	1	1
5	1	0.9
6	0.9	0.8
7	1.3	1.2
8	1.1	
9	1	
10	1.3	

paired

Nr. Subiec	Val uremie (mg/l) lot tratat	Val uremie (mg/l) lot martor	
1	42	48	
2	45	54	
3	48	60	
4	52	66	
5	55	72	
6	58	78	
7	60	84	
8	63	90	
9	67	96	
10	70	102	

unpaired

Subject	TA (mmHg) mother	TA (mmHg) adult child
1	160	150
2	155	160
3	170	155
4	170	150
5	170	150
6	185	155
7	190	165
8	195	165
9	205	165
10	210	170
11	210	175
12	220	175
13	220	180

paired

INFERENTIAL/Inductive STATISTICS
What we need to know before we start the analysis ?

Data type(s): How many samples ?

One sample

Ex. 8	
Nr. Subiec	Val uremie (mg/l) lot tratat
1	42
2	45
3	48
4	52
5	55
6	58
7	60
8	63
9	67
10	70

Two samples

Ex.5	Concentratia medicamen	
Nr. subiect	După 4 ore	După 8 ore
1	1	1
2	1.3	1.3
3	0.9	0.7
4	1	1
5		
6		
7		
8		
9		
10		

Three or more samples

	M_1	M_2	M_3	M_4	M_5	M_6
Experimentul 1	25.1	22.8	25.5	24.5	25.5	24.7
Experimentul 2	27	23.8	27.9	25.2	28.7	27.1
Experimentul 3	29.6	27.1	28.8	27.7	26.2	26
Experimentul 4	26.6	22.7	26.9	26.9	25.7	26.2
Experimentul 5	25.2	22.8	25.4	27.1	27.2	25.7
Experimentul 6	28.3	27.4	30	30.6	27.9	29.2
Experimentul 7	24.7	22.2	29.6	26.4	25.6	28
Experimentul 8	25.1	25.1	23.5	26.6	28.5	24

INFERENTIAL/Inductive STATISTICS

What we need to know before we start the analysis ?

Data type(s): how to collect data for various statistical software

"indexed data" - there is a *"grouping variable":* SPSS, SAS, EpiInfo,OpenStat

"raw data" -
there is *NO "grouping variable":*
Excel,OOO Statistics, GraphPad
Instat, Prism etc

Ex9.	Tymol	
Bonav	Înainte	După ratie
1	10	8
2	8	8
3	16	10
4	5	5
5	6	4
6	12	7
7	9	8
8	10	14
9	14	10
10	10	6

Steps required to apply a statistical test

1. **The statement** of relevant **null** and **alternative hypotheses** to be tested.

2. **Choosing significance level** (represented by the Greek symbol α (alpha). Popular levels of significance are 5%, 1% and 0.1%

3. **Compute the relevant test statistic**

4. **Compare the test-statistic (S) to the relevant critical values (CV)** (obtained from tables in standard cases). Here we may obtain so-called **"p value"**.

5. **Decide to either fail to reject the null hypothesis or reject it in favor of the alternative**. The decision rule is to reject the null hypothesis (H0) if S > CV and vice versa. Practically if p ≤ α, we will reject the null hypothesis (H0), otherwise we will accept it.

* The Steps labeled in red are now usually done by a statistical software

Steps required to apply a statistical test

1. The statement of relevant null and alternative hypotheses

A **hypothesis** is a **statement of belief** about the values of population parameters.

In hypothesis testing, we usually consider two hypotheses: **the null and alternative hypotheses.**

The **null hypothesis**, denoted by H_0, is usually a hypothesis of **no difference**. Initially, we will consider a type of H_0, that is a claim that **there is no difference between the population parameter and its hypothesized value or set of values.**

The hypothesized values chosen for the null hypothesis are usually chosen

to be uninteresting values. An example might be that in a trial comparing two diabetes drugs, the mean values for fasting plasma glucose are the same for the two treatment groups.

In general, the experimenter is interested in rejecting the null hypothesis.

The **alternative hypothesis**, denoted by H_1, is a claim that the null hypothesis is false; e.g. the population parameter takes on a value different from the value or values specified by the null hypothesis.

The alternative hypothesis is usually the scientifically interesting hypothesis that we would like to confirm.

Steps required to apply a statistical test
1. The statement of relevant null and alternative hypotheses

Examples:
Comparing means

	Temp. control	Temp. test	
	37.9	36.7	
	36.1	39.2	
	37.2	36.5	
	36.8	39.5	
	37.0	38.0	
	36.6	37.0	
	37.5	39.1	
	36.5	36.2	
	37.8	38.2	
	36.2	37.6	
	37.4	39.0	
Average	37.0	37.9	

Analyzing contingency tables

Reacție locală		intensă	normală
tip vaccin	A	165	12
	B	142	29

The **null hypothesis** – **nothing has happened** ☺
"Empiric": These means are NOT significantly different !
"State of the art": The mean of the differences between Temp control and Temp. test DO NOT differ significantly from zero!
The **alternative hypothesis**
Vice-versa

The **null hypothesis** – **nothing has happened** ☺
"Empiric": There is NO statistically significant association between independent variable and dependent variable
"State of the art": There is NO statistically significant association between rows and columns
The **alternative hypothesis**
Vice-versa

Steps required to apply a statistical test

2. Choosing significance level (α - alpha).

Types of error

Statisticians speak of two significant sorts of statistical error:

Type I error

Also known as an **"error of the first kind",** an **α error**, or a **"false positive": the error of rejecting a null hypothesis when it is actually true**. Plainly speaking, it occurs when we are observing a difference when in truth there is none. An example of this would be if a test shows that a woman is pregnant when in reality she is not. Type I error can be viewed as the error of **excessive credulity.**

Type II error

Also known as an **"error of the second kind",** a **β error**, or a **"false negative":** the error of failing to reject a null hypothesis when it is in fact not true. In other words, this is the error of failing to observe a difference when in truth there is one. An example of this would be if a test shows that a woman is not pregnant when in reality she is. Type II error can be viewed as the error of **excessive skepticism.**

If we consider **C(onfidence)=1(00)- α(alpha)**

In simple cases, **significance level (α - alpha),** it is defined as **the probability of making a decision to reject the null hypothesis when the null hypothesis is actually true** (a decision known as a **Type I error**, or "false positive determination"). Popular levels of significance are 5%, 1% and 0.1%, empirically coresponding to a "confidence level" of 95%, 99% and 99.9%.

Steps required to apply a statistical test

2. Choosing significance level (α - alpha).

Examples:

Comparing means

	Temp. control	Temp. test
	37.9	36.7
	36.1	39.2
	37.2	36.5
	36.8	39.5
	37.0	38.0
	36.6	37.0
	37.5	39.1
	36.5	36.2
	37.8	38.2
	36.2	37.6
	37.4	39.0
Average	37.0	37.9

α(alpha) =0.05 or
α(alpha) =5%
C(onfidence)=1-0.05=0.95 or
C(onfidence)=100-5= 95 (%)

Analyzing contingency tables

Reacție locală		intensă	normală
tip vaccin	A	165	12
	B	142	29

α(alpha) =0.05 or
α(alpha) =5%
C(onfidence)=1-0.05=0.95 or
C(onfidence)=100-5= 95 (%)

Steps required to apply a statistical test

3. Compute the relevant test statistic.

In statistical hypothesis testing, a **test statistic is a numerical summary of a set of data that reduces the data to one** or a small number of values that **can be used to perform a hypothesis test.**

Given a **null hypothesis and a test statistic T**, we can specify a **"null value" T0** such that **values of T close to T0 present the strongest evidence in favor of the null hypothesis**, whereas **values of T far from T0 present the strongest evidence against the null hypothesis.**

An important property of a test statistic is that **we must be able to determine its sampling (normal?) distribution under the null hypothesis**, which allows us **to calculate p-values.**

Steps required to apply a statistical test

3. Compute the relevant test statistic (a computing intensive task).

Examples:

Comparing means

	Temp. control	Temp. test
	37.9	36.7
	36.1	39.2
	37.2	36.5
	36.8	39.5
	37.0	38.0
	36.6	37.0
	37.5	39.1
	36.5	36.2
	37.8	38.2
	36.2	37.6
	37.4	39.0
Average	37.0	37.9

Analyzing contingency tables

Reacție locală		intensă	normală
tip vaccin	A	165	12
	B	142	29

Every test has a mathematical formula. E.g. :

t (Student) statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sum d_1^2 + \sum d_2^2}{n_1 + n_2 - 2}}} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$$

Chi square statistic

$$\chi^2 = \frac{e_e - e_t}{e_t}$$

* This step is now usually done by a statistical software. For explanation of the terms in the equations, go to "Marius Marusteri – Fondamentals in Biostatistics – lecture notes", at
http://www.umftgm.ro/info/Curs_Notiuni_fundamentale.pdf

Steps required to apply a statistical test

4. Compare the test-statistic (S) to the relevant critical values (CV) (obtained from tables in standard cases).

Examples:

Comparing means

	Temp. control	Temp. test
	37.9	36.7
	36.1	39.2
	37.2	36.5
	36.8	39.5
	37.0	38.0
	36.6	37.0
	37.5	39.1
	36.5	36.2
	37.8	38.2
	36.2	37.6
	37.4	39.0
Average	37.0	37.9

The results :

t (Student) statistic

t (Student) statistic = 2.25
Degrees of freedom = 11+11-2=20

Analyzing contingency tables

Reacție locală		intensă	normală
tip vaccin	A	165	12
	B	142	29

Chi square statistic

Chi square statistic = 8.67
Degrees of freedom=1

* This step is now usually done by a statistical software.

Steps required to apply a statistical test

4. Compare the test-statistic (S) to the relevant critical values (CV) (obtained from tables in standard cases). Here we may obtain so-called "p value".

Examples:

Comparing means

	Temp. control	Temp. test
	37.9	36.7
	36.1	39.2
	37.2	36.5
	36.8	39.5
	37.0	38.0
	36.6	37.0
	37.5	39.1
	36.5	36.2
	37.8	38.2
	36.2	37.6
	37.4	39.0
Average	37.0	37.9

t (Student) statistic = 2.25
Degrees of freedom = 20

Nivel de semnificație pentru testul bilateral

n \ a	0,50	0,20	0,10	0,05	0,02	0,01	0,002	0,001	0,0001
1	1,000	3,078	6,314	12,70	31,821	63,657	318,309	636,618	6366,198
2	0,816	1,886	2,290	4,303	6,965	9,925	22,327	31,598	99,992
3	0,765	1,638	2,353	3,182	4,541	5,841	10,214	12,924	28,000
4	0,741	1,533	2,132	2,77	3,747	4,604	7,173	8,610	15,544
5	0,727	1,476	2,015	2,571	3,365	4,032	5,893	6,869	11,178
6	0,718	1,440	1,943	2,44	3,143	3,707	5,208	5,959	9,082
7	0,711	1,415	1,895	2,365	2,998	3,499	4,785	5,408	7,885
8	0.706	1,397	1,860	2,30	2,896	3,355	4,501	5,041	7,120
9	0,703	1,383	1,833	2,262	2,821	3,250	4,297	4,781	6,594
10	0,700	1,372	1,812	2,228	2,764	3,169	4,144	4,587	6,211
11	0,697	1,363	1,796	2,201	2,718	3,106	3,025	4,437	5,921
12	0,695	1,356	1,782	2,17	2,681	3,102	3,930	4,318	5,694
13	0,694	1,350	1,771	2,16	2,650	3,055	3,852	4,221	5,513
14	0,692	1,345	1,761	2,145	2,624	2,977	3,787	4,140	5,363
15	0,691	1,341	1,753	2,131	2,602	2,947	3,733	4,073	5,239
16	0,690	1,337	1,746	2,12	2,583	2,921	3,686	4,015	5,134
17	0,689	1,333	1,740	2,11	2,567	2,898	3,646	3,965	5,014
18	0,688	1,330	1,734	2,101	2,552	2,878	3,610	3,922	4,966
19	0,688	1,328	1,729	2,093	2,539	2,861	3,579	3,883	4,897
20	0,687	1,325	1,725	2,086	2,528	2,845	3,552	3,850	4,837

t stat > 2,086 (p=0.05) and t stat < 2.528 (p=0.02)

* This step is now usually done by a statistical software **(real p=0.04).**

Steps required to apply a statistical test

4. Compare the test-statistic (S) to the relevant critical values (CV) (obtained from tables in standard cases). Here we may obtain so-called "p value".

Examples:

Analyzing contingency tables

Reacție locală		intensă	normală
tip vaccin	A	165	12
	B	142	29

n α	0,990	0,975	0,950	0,900	0,100	0,050	9,025	0,010	0,001
1	0,0002	0,0010	0,0039	0,0198	2,71	3,84	5,02	6,63	10,83
2	0,02	0,05	0,10	0,21	4,61	5,99	7,38	9,21	13,82

Chi square statistic = 8.67
Degrees of freedom=1

Chi stat > 6.63 (p=0.01) and t stat < 10.83 (p=0.001)

* This step is now usually done by a statistical software **(real p=0.0032).**
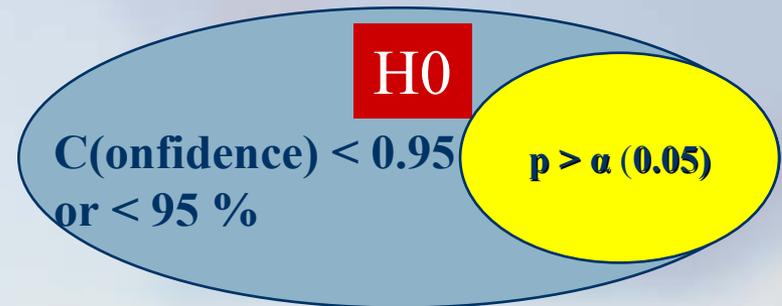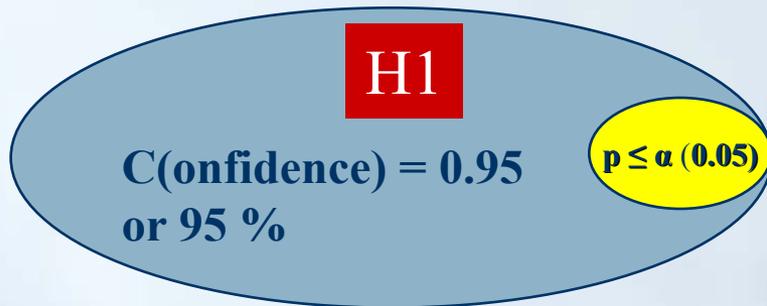
Steps required to apply a statistical test

5. Decide to either fail to reject the null hypothesis or reject it in favor of the alternative hypothesis.

The decision rule is to reject the null hypothesis (H0) if S > CV and vice versa, for a given alpha.

Practically if **p ≤ α, we will reject the null hypothesis (H0),** otherwise **(p > α)** we **will accept it**.

Some "empirical" explanations:

H1

C(onfidence) = 0.95
or 95 %

p ≤ α (0.05)

H0

C(onfidence) < 0.95
or < 95 %

p > α (0.05)

Steps required to apply a statistical test

5. Decide to either fail to reject the null hypothesis or reject it in favor of the alternative hypothesis.

Examples:

Comparing means

	Temp. control	Temp. test	
	37.9	36.7	
	36.1	39.2	
	37.2	36.5	
	36.8	39.5	
	37.0	38.0	
	36.6	37.0	
	37.5	39.1	
	36.5	36.2	
	37.8	38.2	
	36.2	37.6	
	37.4	39.0	
Average	37.0	37.9	

α(alpha) =0.05 or
α(alpha) =5%
C(onfidence)=1-0.05=0.95 or
C(onfidence)=100-5= 95 (%)

p=0.04 => we accept H1,
These **means** are **significantly different** !
OR
The **mean of the differences** between Temp control
and Temp. test **differ significantly from zero!**

Analyzing contingency tables

Reacție locală		intensă	normală
tip vaccin	A	165	12
	B	142	29

α(alpha) =0.05 or
α(alpha) =5%
C(onfidence)=1-0.05=0.95 or
C(onfidence)=100-5= 95 (%)

p=0.0032 => we accept H1,
There is **statistically significant association** between
independent variable and dependent variable **OR**
There is **statistically significant association** between rows
and columns

Steps required to apply a statistical test

5a. VERY IMPORTANT !!! – every STATISTICAL CONCLUSION must be followed by a SCIENTIFIC (Biomedical) ONE !

Comparing means

	Temp. control	Temp. test
	37.9	36.7
	36.1	39.2
	37.2	36.5
	36.8	39.5
	37.0	38.0
	36.6	37.0
	37.5	39.1
	36.5	36.2
	37.8	38.2
	36.2	37.6
	37.4	39.0
Average	37.0	37.9

Analyzing contingency tables

Reacție locală		intensă	normală
tip vaccin	A	165	12
	B	142	29

Examples:

The temperature of the test group (animals which received pyrogenic agents) is significantly different by the control (placebo) group !
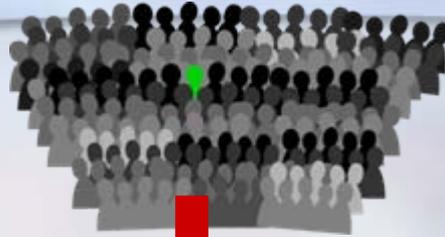
There is an association between vaccin types (A or B) and the local reaction !

If we measure some numbers :

Statistical Population – including all people or items with the characteristic one wishes to understand.

Sample – a subset of a population

Descriptive statistics are used to describe the main features of a collection of data in **quantitative terms.**

Grup Test		Grup Control	
Mean	7	Mean	6.5
Standard Error	2.0155644	Standard Error	0.62361
Median		Median	6.5
Mode	#N/A	Mode	6.5
Standard Deviation	6.046693311	Standard Deviation	1.972027
Sample Variance	36.5625	Sample Variance	3.888889
Kurtosis	-0.541523434	Kurtosis	-0.38145
Skewness	0.940483152	Skewness	-0.27166
Range	16.5	Range	6.5
Minimum	1.5	Minimum	3
Maximum	18	Maximum	9.5
Sum	67.5	Sum	65
Count	9	Count	10
Largest(1)	18	Largest(1)	9.5
Smallest(1)	1.5	Smallest(1)	3
Confidence Level(95.0)	4.647902932	Confidence Level(9	1.410704
CV	80.62257748	CV	30.33887

**Optional tests
e.g. test for outliers**

**Mandatory tests
e.g. goodness-of-fit (normality) tests**

Inferential/inductive statistics

Test for outliers (optional tests):

In statistics, an **outlier** is an observation that is **numerically distant from the rest of the data**.

They can **occur by chance** in any distribution, but they are often **indicative** either of **measurement error** or that the population has a **heavy-tailed distribution**.

Outliers tests:

Grubbs Test (common use)

Chauvenet elimination criteria (e.g. in Romanian Pharmacopoeia)

Basic principles:

• We **have to apply an outlier test** when we meet high dispersion of data (high values of variance OR standard deviation, compared with the mean)

•**Outliers are eliminated on by one, on each sample**.

• If we are dealing with **paired data**, we have to **eliminate the pair also**.

Test for outliers (optional tests)- Grubbs Test:

1. **The statement of relevant null and alternative hypotheses** to be tested
(H0: The values are NOT signifficantly different from average values.)
(H1: The values are signifficantly different from average values.)
2. **Choosing significance level** (represented by the Greek symbol α (alpha), in this case 0,05.
3. **Compute the relevant test statistic** $Z = \dfrac{|mean - value|}{SD}$
4. **Compare the test-statistic to the relevant critical values** (obtained from tables).
5. **Decide to either accept the null hypothesis or reject it in favor of the alternative.**
Practically if p ≤ α, we will reject the null hypothesis (H0), otherwise we will accept it.

And some examples ....

Test for outliers (optional tests):

Raw unpaired data

First sample, first outlier

First sample, no more outliers

Second sample, first outlier

Subiect	TA Sănăto	TA Bolnav
1	80	150
2	155	160
3	170	155
4	170	150
5	170	50
6	185	155
7	190	
8	195	165
9	205	165
10	210	90
11	210	175
12	220	175
13	220	180

Grubb's test for outliers.

Instructions:
Enter a list of values into the white column, then press CALCULATE

Clear   Calculate

If nothing happens when you press Calculate, press the RETURN key first. You cannot press buttons while you are entering or editing data.

Data: P<0.05??
80   Yes
155
170
170
170
185
190
195
205
220
220

Descriptive statistics
Mean= 183.0769
SD= 37.39087
N= 13

Results:
Grubb's test calculates a P value only for the value furthest from the rest. If the test determines that the value is an outlier, you'll see "Yes" next to that value. Otherwise, you'll see the word "No".

Reference:
Outliers in Statistical Data (3rd edition), V. Barnett and T. Lewis, Wiley, 1994.

Final unpaired data, no outliers

Subiect	TA Sănăto	TA Bolnav
1		150
2	155	160
3	170	155
4	170	150
5	170	
6	185	
7	190	165
8	195	165
9	205	165
10	210	
11	210	175
12	220	175
13	220	180

Second sample, no more outliers

Second sample, second outlier

Test for outliers (optional tests)- Grubbs Test:

Outliers make statistical analyses difficult.

When you encounter an outlier, **you may be tempted to delete it from the analyses**. First, ask yourself these questions:

   * **Was the value entered into the computer correctly?** If there was an error in data entry, fix it.
   * **Were there any experimental problems with that value?** For example, if you noted that one tube looked funny, you have justification to exclude the value resulting from that tube without needing to perform any calculations.
   * **Is the outlier caused by biological diversity?** If each value comes from a different person or animal, the outlier may be a correct value. It is an outlier not because of an experimental mistake, but rather because that individual may be different from the others. This may be the most exciting finding in your data!

After answering no to those three questions, you have to decide what to do with the outlier. There are two possibilities.

   * **One possibility is that the outlier was due to chance.** In this case, you should keep the value in your analyses. The value came from the same population as the other values, so should be included.
   * **The other possibility is that the outlier was due to a mistake** - bad pipetting, voltage spike, holes in filters, etc. Since including an erroneous value in your analyses will give invalid results, you should remove it. In other words, the value comes from a different population than the other and is misleading.