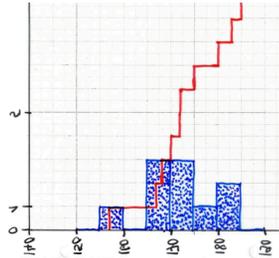
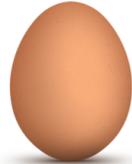


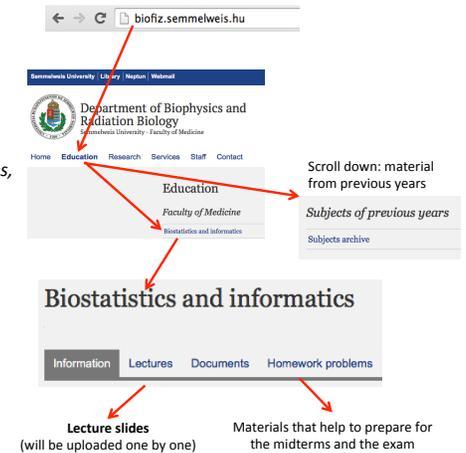
Principles of Biostatistics and Informatics



Lecture 1: An Introduction
12th September 2018
Gergely AGÓCS

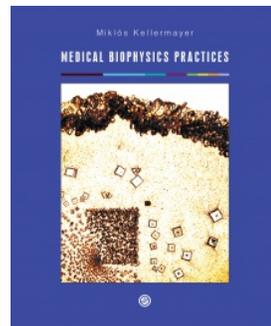
How to Get Prepared?

- university = **autonomous learning**
- sources:
 - **your notes** made during lectures (Wednesdays, 10⁰⁰–11²⁵; EOK "Szent-Györgyi Albert" lecture hall)
 - **your notes** made during computer lab classes (once a week, 90 minutes, 1st floor in the EOK building, computer labs in corridor "B")
 - **consultations** (Wed: 19⁰⁰–20³⁰; 1st floor in the EOK building, computer labs in corridor "B")
 - "Medical Biophysics Practices" lab practice **book**:
 - Biostatistics chapter (40 page summary of theory)
 - homepage: biofiz.semmelweis.hu
 - subject requirements
 - lecture schedule and slides
 - lab schedule
 - homework problems
 - material from previous years



How to Get Prepared?

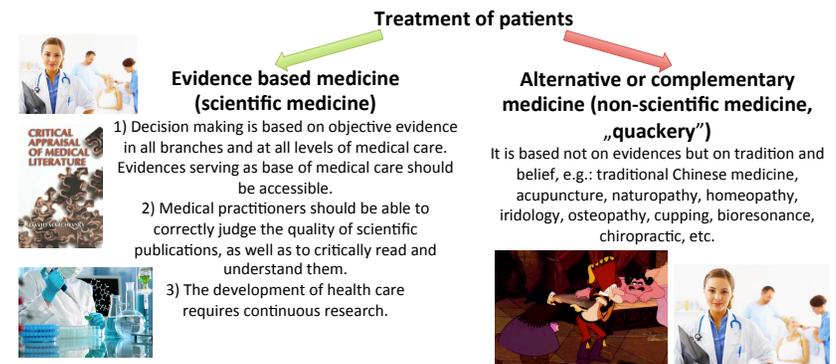
- university = **autonomous learning**
- sources:
 - **your notes** made during lectures (Wednesdays, 10⁰⁰–11²⁵; EOK "Szent-Györgyi Albert" lecture hall)
 - **your notes** made during computer lab classes (once a week, 90 minutes, 1st floor in the EOK building, computer labs in corridor "B")
 - **consultations** (Wed: 19⁰⁰–20³⁰; 1st floor in the EOK building, computer labs in corridor "B")
 - "Medical Biophysics Practices" lab practice **book**:
 - Biostatistics chapter (40 page summary of theory)



Science and Non-science

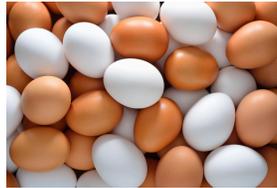
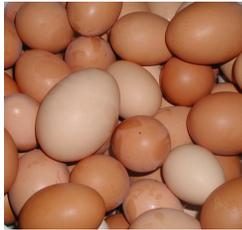
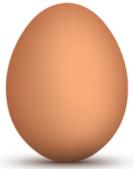
Presumption of innocence: „Everyone who has been charged shall be presumed innocent until proved guilty according to law.” CHARTER OF FUNDAMENTAL RIGHTS OF THE EUROPEAN UNION, Article 48 (1)

„Presumption of ineffectiveness”: Every treatment and remedy shall be presumed ineffective until proved effective according to scientific requirements.



How Does Statistics Help Us?

Statistics deals with the collection, organization, analysis of data, and drawing conclusions



What Type of Data do We Deal with?

Data to be processed show a high degree of variation ...

TESTS	RESULT	FLAG	UNITS	REFERENCE INTERVAL	LAB
CBC With Differential/Platelet					
WBC	5.1		x10E3/uL	4.0 - 10.5	01
RBC	4.94		x10E6/uL	4.10 - 5.40	01
Hemoglobin	15.1		g/dL	12.5 - 17.0	01
Hematocrit	46.2		%	36.0 - 50.0	01
MCV	94		fL	80 - 98	01
MCH	30.6		pg	27.0 - 34.0	01
MCHC	32.7		g/dL	32.0 - 36.0	01
RDW	13.2		%	11.7 - 15.0	01
Platelets	201		x10E3/uL	140 - 415	01
Neutrophils	44		%	40 - 74	01
Lymphs	44		%	14 - 46	01
Monocytes	9		%	4 - 13	01

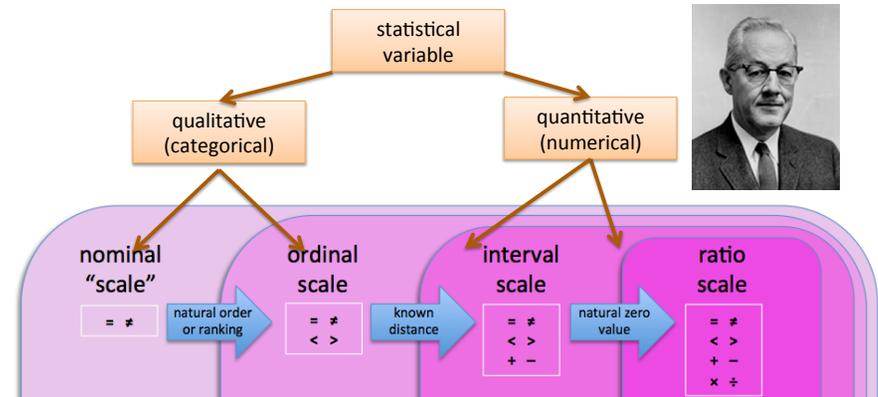
The physicist measures ...	The physician measures ...	The medical student measures ...
length	height	diameter of red blood cells (2)
frequency	heart rate	pulse frequency (22)
concentration	blood sugar level	protein conc. in blood plasma (4)
voltage	ECG-signal	ECG-signal (27)
sound intensity	hearing threshold	hearing threshold (25)
electric impedance	impedance-plethysmograph (volume)	skin impedance (24)
pressure	blood pressure	-
speed	speed of blood flow	-

What Type of Data do We Deal with?

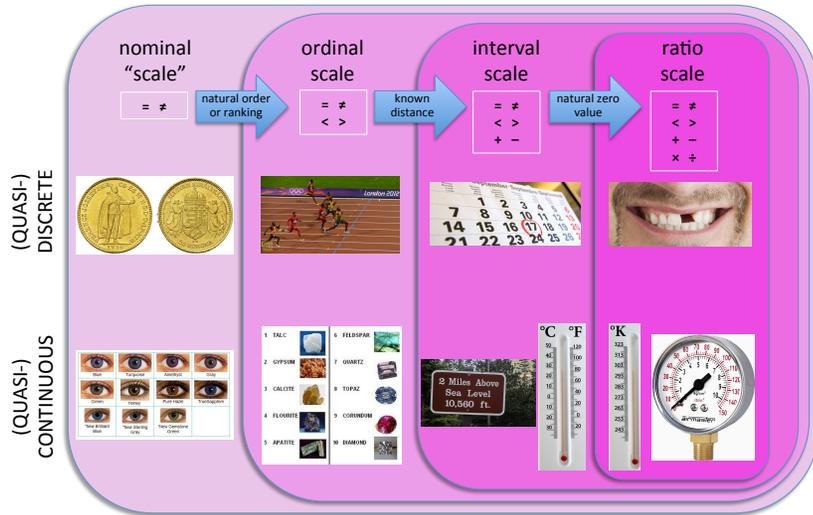
Data to be processed show a high degree of variation ...

TESTS	RESULT	FLAG	UNITS	REFERENCE INTERVAL	LAB
CBC With Differential/Platelet					
WBC	5.1		x10E3/uL	4.0 - 10.5	01
RBC	4.94		x10E6/uL	4.10 - 5.60	01
Hemoglobin	15.1		g/dL	12.5 - 17.0	01
Hematocrit	46.2		%	36.0 - 50.0	01
MCV	94		fL	80 - 98	01
MCH	30.6		pg	27.0 - 34.0	01
MCHC	32.7		g/dL	32.0 - 36.0	01
RDW	13.2		%	11.7 - 15.0	01
Platelets	201		x10E3/uL	140 - 415	01
Neutrophils	44		%	40 - 74	01
Lymphs	44		%	14 - 46	01
Monocytes	9		%	4 - 13	01

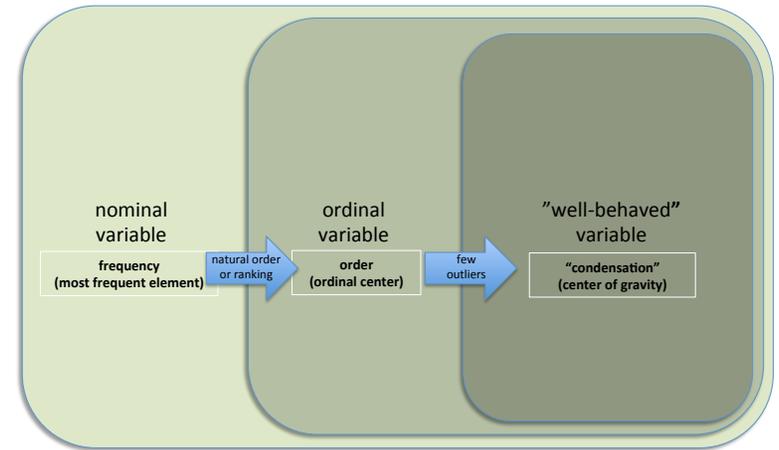
Types of Statistical Variables (I): Levels of Measurement (by S. S. Stevens)



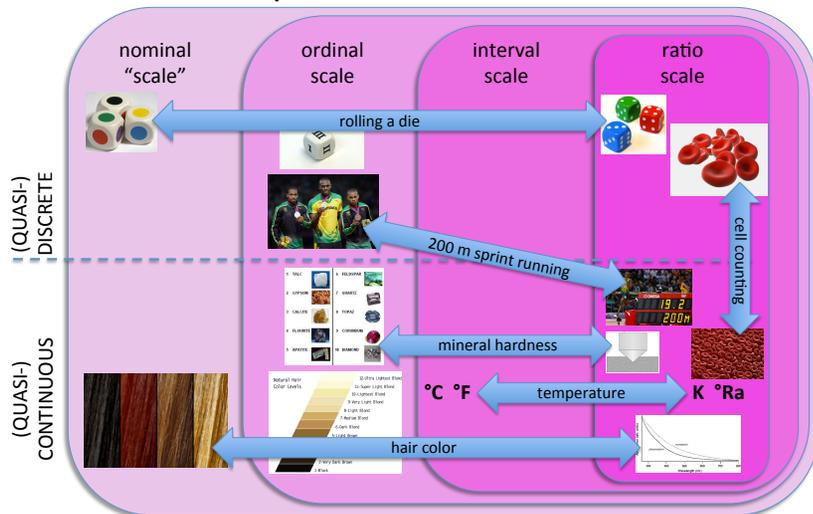
Types of Statistical Variables (I): Levels of Measurement (by S. S. Stevens)



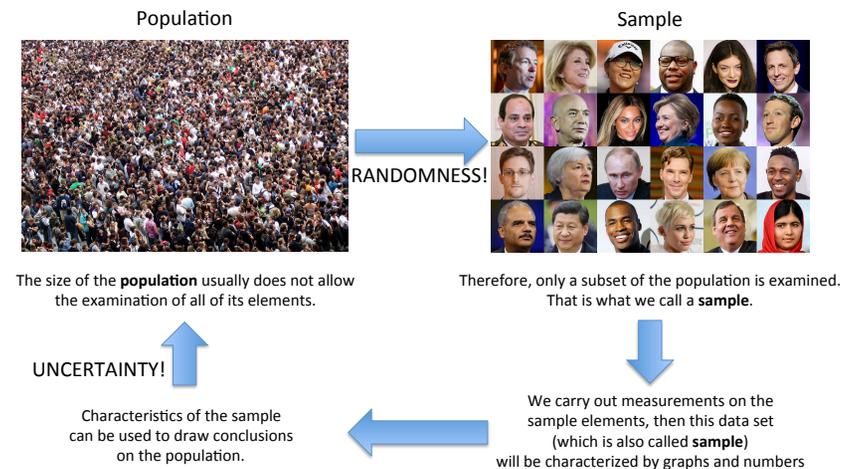
Statistical Variable Types (II): The Basis for Statistical Comparison



Types of Statistical Variables (I): The Importance of the Context



Population and Sample



Representation of Data (I)

- 1) A simple list of data
- 2) Summary of frequencies in tables
 - absolute frequency (Δn) and relative frequency ($\Delta n/n$)
 - categories are evident for qualitative variables [Excel: =COUNTIF() function]
 - categories (bins) are created arbitrarily for quantitative variables [Excel: =FREQUENCY() function can also be used]
 - frequency density ($\Delta n/\Delta x$) and relative frequency density ($[\Delta n/n]/\Delta x$)

list:
an enumeration of results of all experiments

(absolute) frequency:
number of experiments with the given outcome

relative frequency:
the proportion of the given outcome within the sample

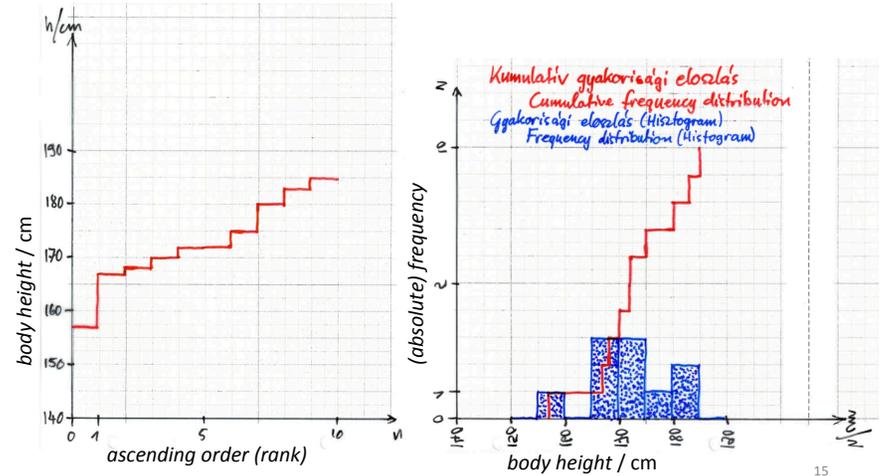
patient №	blood group (ABO)	cholesterol level (mg/dL)
1	B	148
2	AB	169
3	B	159
4	B	159
5	B	150
6	B	167
7	A	147
8	B	15
9	AB	177
10	B	150
11	B	141

blood group	(absolute) frequency	relative frequency
A	85	0.425
B	28	0.14
AB	10	0.05
O	77	0.385
Σ	200	1

gyakorlati elosztások (differenciálszámítások függvények)					
osztályok	osztályok feloldó (szé) határa	(abszolút) gyakoriság (GYAKORISÁG)	(abszolút) gyakoriság (DARABTEL)	relatív gyakoriság	relatív gyakoriság: sűrűség
$x \leq 100$	110	0	0	0	0
$100 < x \leq 110$	120	2	2	0.01	0.2
$110 < x \leq 120$	130	5	5	0.025	0.5
$120 < x \leq 130$	140	22	22	0.11	2.2
$130 < x \leq 140$	150	31	31	0.155	3.1
$140 < x \leq 150$	160	48	48	0.24	4.8
$150 < x \leq 160$	170	40	40	0.2	4
$160 < x \leq 170$	180	32	32	0.16	3.2
$170 < x \leq 180$	190	10	10	0.05	1
$180 < x \leq 190$	200	9	9	0.045	0.9
$190 < x \leq 200$	210	11	11	0.055	1.1
$200 < x \leq 210$					
$210 < x$					
összeg		200	200	1	0.0005

Representation of Data (II)

4) Representation of frequencies in case of **quantitative** variables



Representation of Data (II)

3) Representation of frequencies in case of **qualitative** variables

blood group	(absolute) frequency	relative frequency
A	85	0.425
B	28	0.14
AB	10	0.05
O	77	0.385
Σ	200	1

Relative frequency

Frequency "distribution"

Relative frequency „distribution“

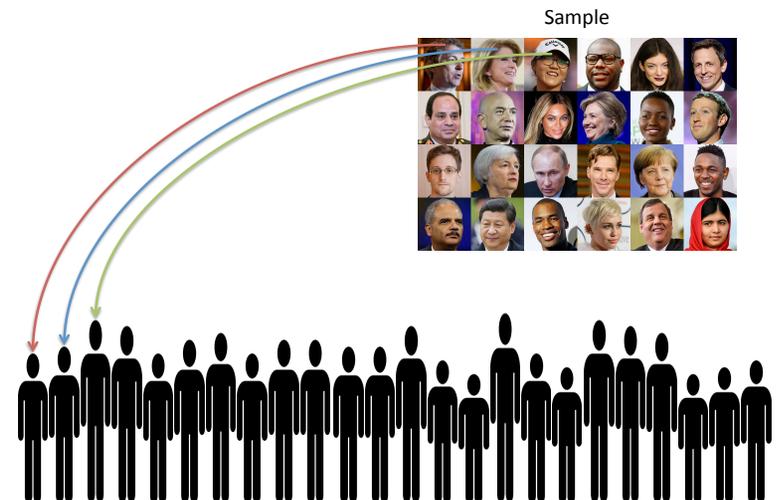
dependent variable: (absolute) frequency

independent variable: nominal categories (categorical „axis“)

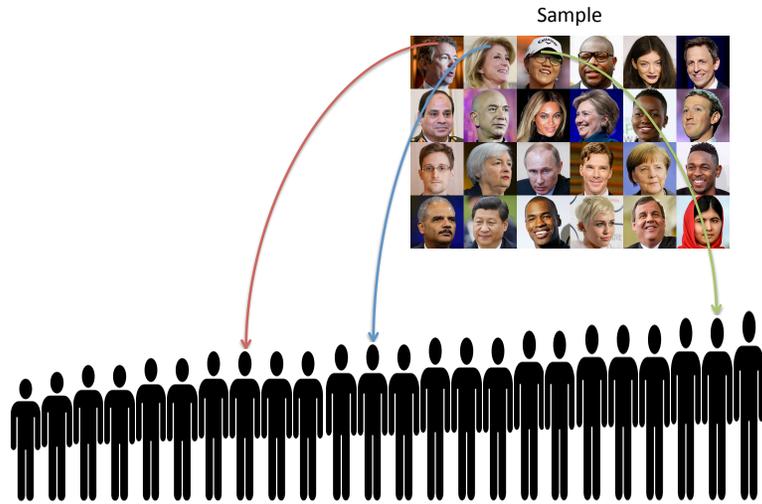
dependent variable: relative frequency

independent variable: nominal categories (categorical „axis“)

Composition of the Data

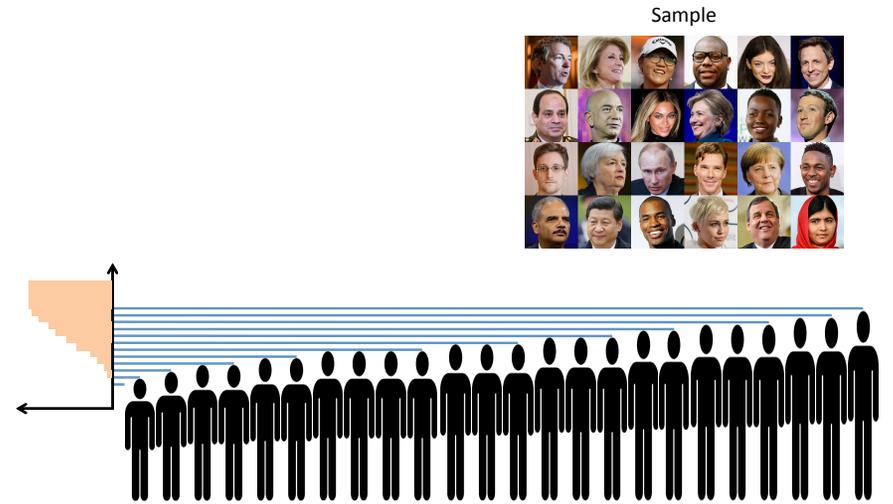


Composition of the Sample

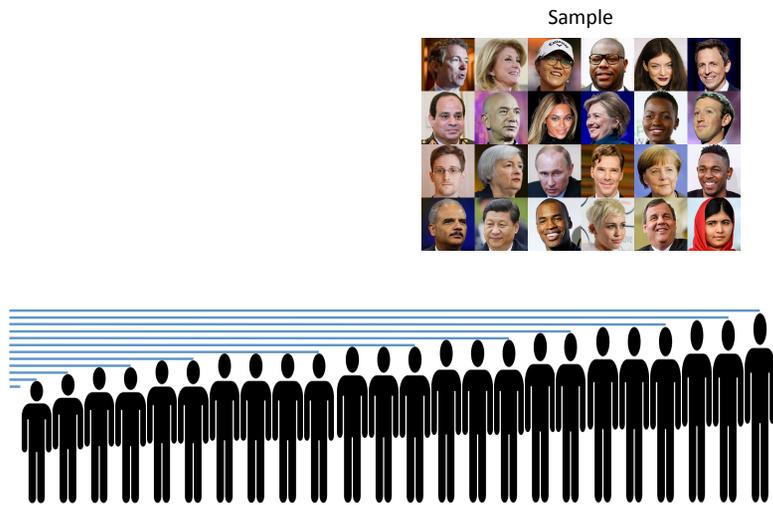


Cumulative Distribution Function

How many elements are less than a given x value?

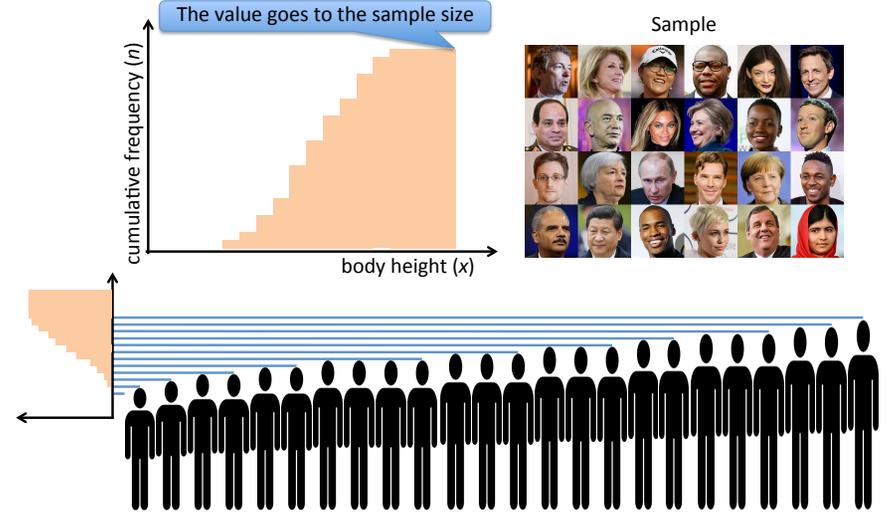


Composition of the Sample



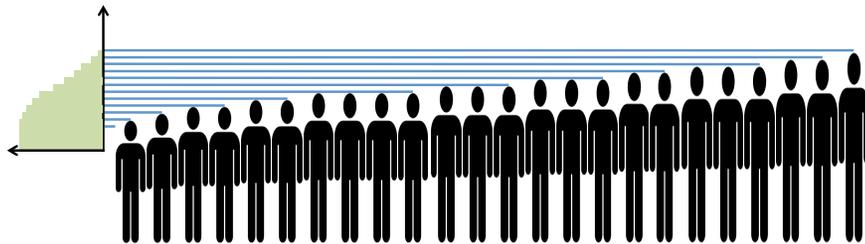
Cumulative Distribution Function

How many elements are less than a given x value?



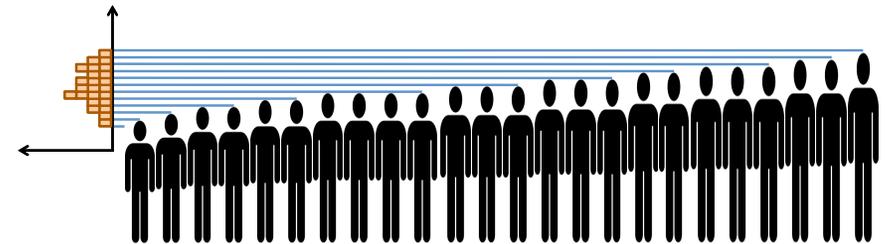
Integral Discrimination Function

How many elements are **greater** than a given x value?



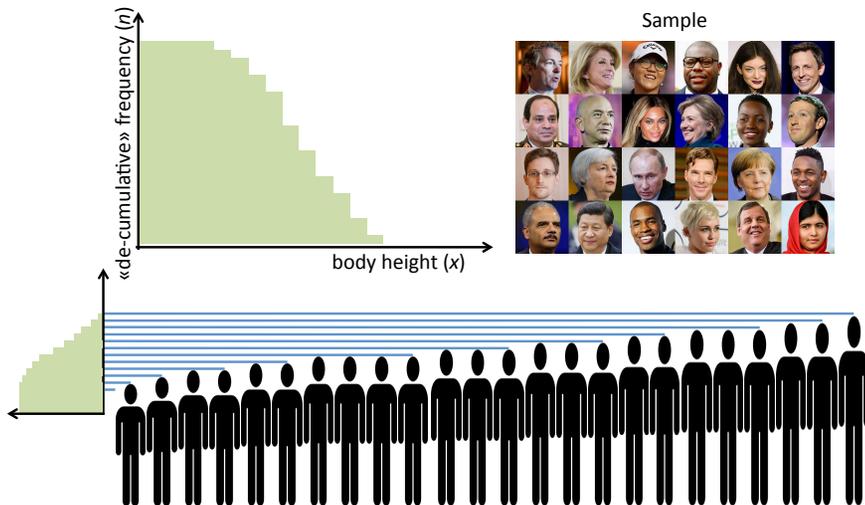
Frequency Distribution Function

How many elements fall **within** a bin of Δx width?



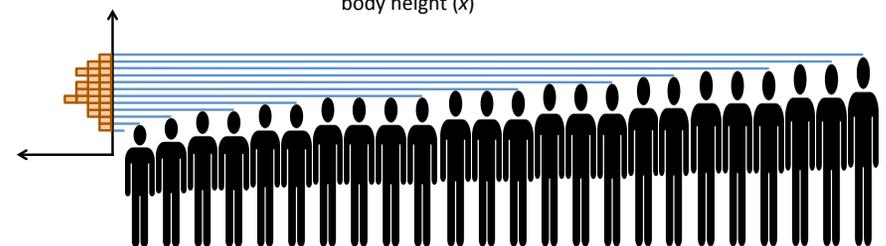
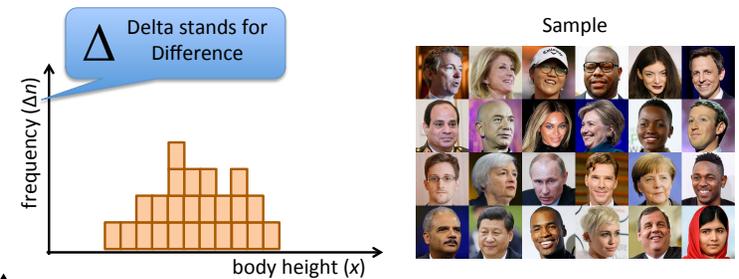
Integral Discrimination Function

How many elements are **greater** than a given x value?



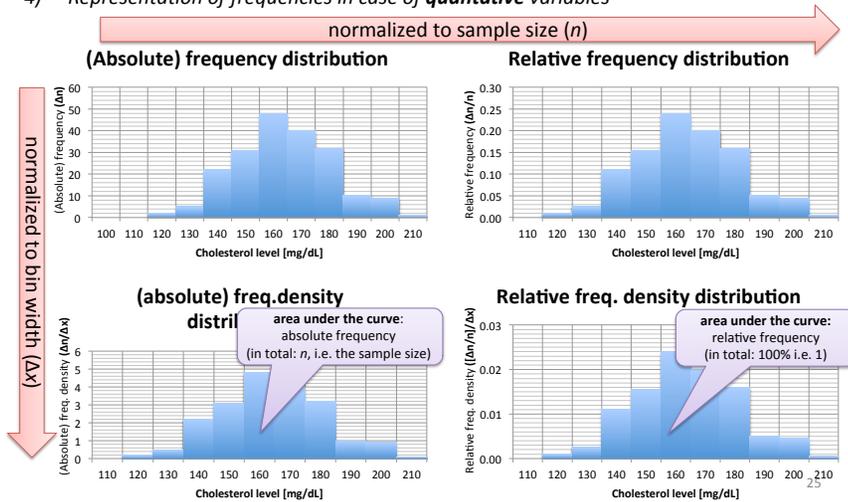
Frequency Distribution Function

How many elements fall **within** a bin of Δx width?



Representation of Data (III)

4) Representation of frequencies in case of **quantitative** variables



Test questions #2

- Give examples for ordinal scale.
- What is the substantial difference between a nominal and an ordinal scale?
- Give example for interval scale.
- What is the substantial difference between an ordinal and an interval scale?
- Give examples for ratio scale.
- What is the substantial difference between an interval and a ratio scale?
- Why is it important to define a statistical variable properly?
- Show the importance of the context of the statistical variable with examples.
- How can the statistical variables be grouped from the point of view of comparison?
- What is the relationship between Stevens' levels of measurements and the hierarchy of variable comparison?
- What does it mean, that a variable is "well-behaved"?
- How can the different levels defined in the hierarchy of variable comparison be characterized?
- Is a quantity measured on (Stevens') interval scale always "well-behaved" from the aspect of statistical comparison?
- What is a population?
- What is a sample?
- How do we take a sample from the population?
- Why does the sample not perfectly represent the population?
- What is the reason of the uncertainty burden on conclusions on the populations drawn from the sample?
- What are the ways of characterization of the sample?
- What is the meaning of frequency and absolute frequency?

27

Test questions #1

- What are the most important characteristics of science?
- What are the properties of scientific thinking?
- How does scientific medicine differ from quackery?
- How can we make sure that a given medical method is based on scientific evidence?
- Give examples for non-scientific medical methods!
- Who has to carry the burden of proof: Who states that a new method is efficient or who says it is not?*
- What is statistics?
- What is the aim of statistics?
- Which branches of mathematics does statistics rely on?
- What is the aim of model making?
- What is the relationship between a model and reality?
- What should be proven about a medicine or treatment: whether it is efficient or it is inefficient?
- What is the problem with the following statement: "No one has proven that this method is inefficient, therefore it would be unjust to limit its use."
- How can the variables acquired during the examination of a patient be grouped? Name some nominal variables. Name some ordinal variables. Name some discrete numerical variables. Name some continuous numerical variables. Name some "well-behaved" statistical variables. Name some non-"well-behaved" statistical variables. Give examples for nominal "scale". Why is the word "scale" in the term "nominal scale" a misnomer?

26

Test questions #3

- If we just say "frequency" what does it refer to: absolute or relative frequency?
- How can the frequency values of a set of data be summarized in a table?
- What do we need to pay attention to during the tabular summary of frequencies of numerical variables?
- How can we make frequencies directly comparable, if samples differ in size?
- How can we make frequencies directly comparable, if bins differ in width?
- Why is relative frequency "relative"?
- What does the word "density" refer to in the term frequency density?
- What are the practical ways of graphical representation for qualitative variables?
- What is represented on the horizontal axis of a column chart constructed from qualitative data?
- What is the meaning of the "categorical axis" of a column chart constructed from qualitative data?
- What and how can be read out from the graph of a frequency distribution?
- What and how can be read out from the graph of a relative frequency distribution?
- What and how can be read out from the graph of a frequency density distribution?
- What and how can be read out from the graph of a relative frequency density distribution?
- What is the total area under the curve of a frequency and a relative frequency density distribution?

28