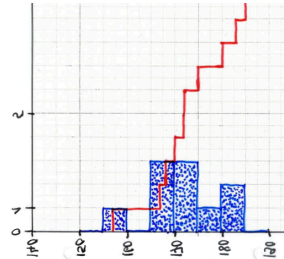


Principles of Biostatistics and Informatics



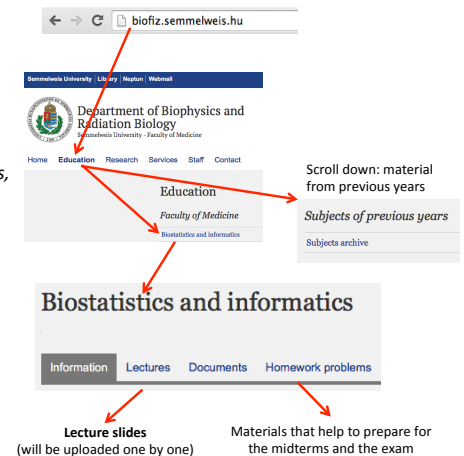
Lecture 1: An Introduction

12th September 2018

Gergely AGÓCS

How to Get Prepared?

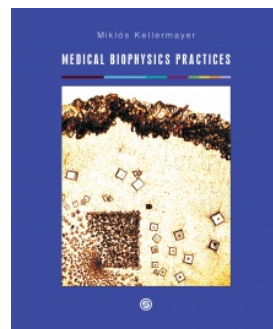
- university = **autonomous learning**
- sources:
 - your notes** made during lectures (Wednesdays, 10⁰⁰–11²⁵; EOK "Szent-Györgyi Albert" lecture hall)
 - your notes** made during computer lab classes (once a week, 90 minutes, 1st floor in the EOK building, computer labs in corridor "B")
 - consultations** (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



3

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)

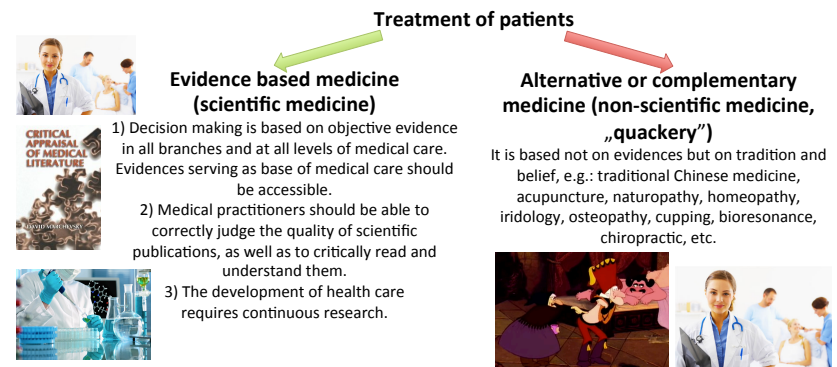


2

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.



4

How Does Statistics Help Us?

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

Descriptive statistics \longleftrightarrow Inferential statistics



5

What Type of Data do We Deal with?

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

LabCorp		13112 Fanning Creek Dr. Ste. 300 San Diego, CA 92128-4108		Phone: 858-668-3700	
333-086-0655-0	22247228	22247228	22247228	22247228	22247228
DONATRE		Request A Test, LTD.		VART Verified	
NONITO		8803 Brecksville Rd. Ste. 7-130		BRECKSVILLE OH 44141	
4897 THOMPSON DR.		PHOTO ID REQUIRED		Additional Information	
SAN MATEO CA 94401		11/28/10		12/02/10 08:08ET	
CBC With Differential/Platelet		Temp. Unit: Celsius		Temp. Unit: Fahrenheit	
WBC		5.1		x10E3/uL	
RBC		4.94		x10E6/uL	
Hemoglobin		15.1		g/dL	
Hematocrit		46.2		%	
MCV		94		fL	
MCH		30.6		pg	
MCHC		32.7		g/dL	
RDW		13.2		%	
Platelets		201		x10E3/uL	
Neutrophils		44		%	
Lymphs		44		%	
Monocytes		9		%	

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	-

7

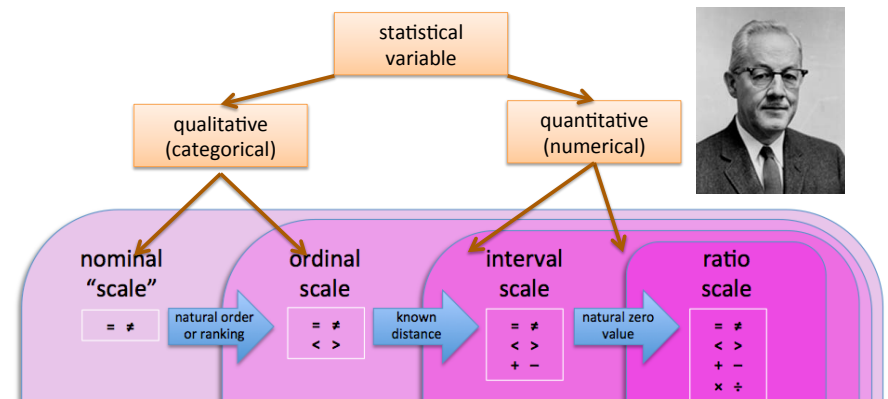
What Type of Data do We Deal with?

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

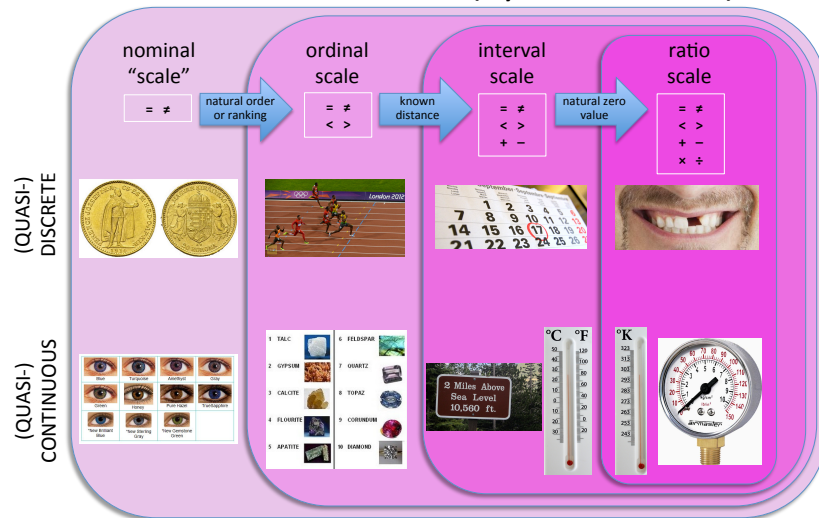
LabCorp		13112 Fanning Creek Dr. Ste. 300 San Diego, CA 92128-4108		Phone: 858-668-3700	
333-086-0655-0	22247228	22247228	22247228	22247228	22247228
DONATRE		Request A Test, LTD.		VART Verified	
NONITO		8803 Brecksville Rd. Ste. 7-130		BRECKSVILLE OH 44141	
4897 THOMPSON DR.		PHOTO ID REQUIRED		Additional Information	
SAN MATEO CA 94401		11/28/10		12/02/10 08:08ET	
CBC With Differential/Platelet		Temp. Unit: Celsius		Temp. Unit: Fahrenheit	
WBC		5.1		x10E3/uL	
RBC		4.94		x10E6/uL	
Hemoglobin		15.1		g/dL	
Hematocrit		46.2		%	
MCV		94		fL	
MCH		30.6		pg	
MCHC		32.7		g/dL	
RDW		13.2		%	
Platelets		201		x10E3/uL	
Neutrophils		44		%	
Lymphs		44		%	
Monocytes		9		%	

6

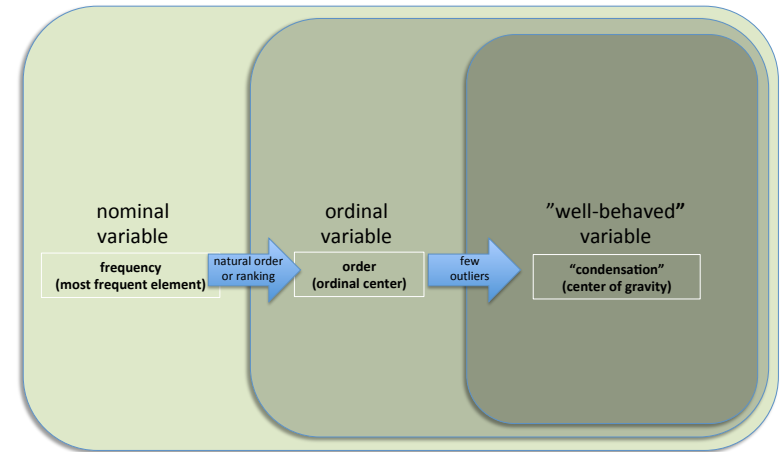
Types of Statistical Variables (I): Levels of Measurement (by S. S. Stevens)



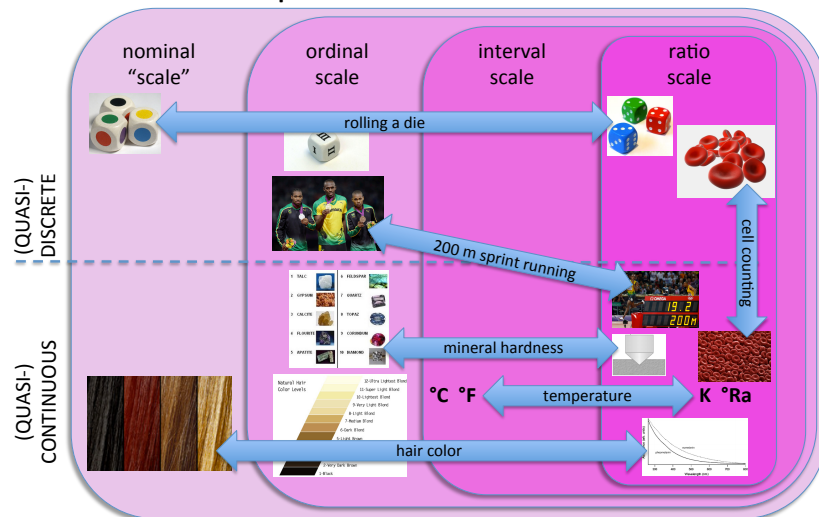
Types of Statistical Variables (I): Levels of Measurement (by S. S. Stevens)



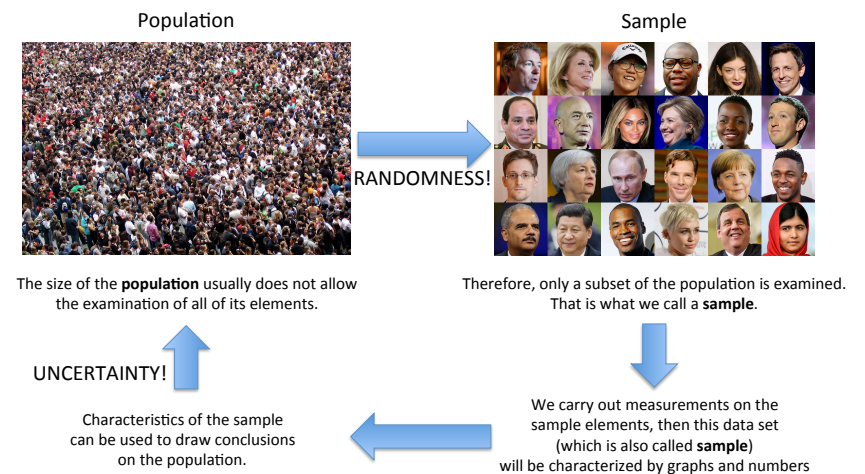
Statistical Variable Types (II): The Basis for Statistical Comparison



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

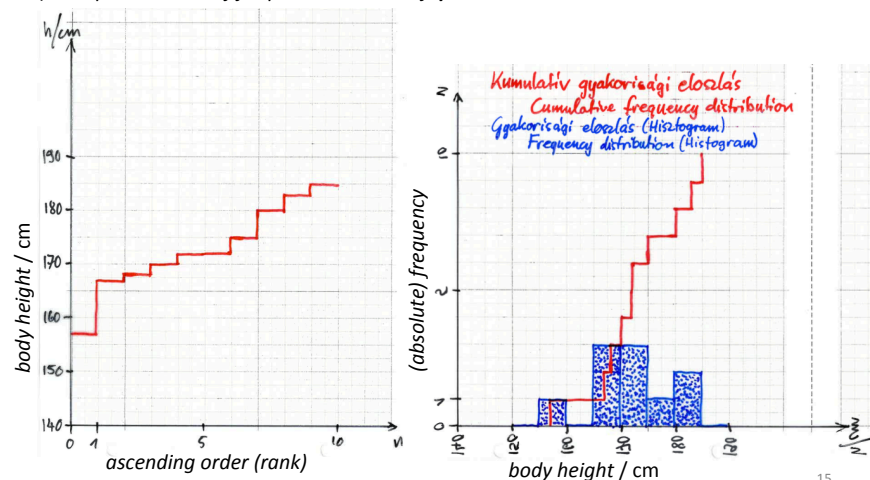
gyakorlati eloszlások (differenciálszámítások függvények)

osztályok	abszolút gyakoriság (GYAKORISÁG)	relatív gyakoriság	abszolút gyakoriság-sűrűség	relatív gyakoriság-sűrűség
$x \leq 100$	0	0	0	0
$100 < x \leq 110$	2	0.01	0.2	0.001
$110 < x \leq 120$	5	0.025	0.5	0.0025
$120 < x \leq 130$	22	0.11	2.2	0.011
$130 < x \leq 140$	31	0.155	3.1	0.0155
$140 < x \leq 150$	48	0.24	4.8	0.024
$150 < x \leq 160$	40	0.2	4	0.02
$160 < x \leq 170$	32	0.16	3.2	0.016
$170 < x \leq 180$	10	0.05	1	0.005
$180 < x \leq 190$	9	0.045	0.9	0.0045
$190 < x \leq 200$	1	0.005	0.1	0.0005
$200 < x \leq 210$	0	0	0	0
$210 < x$	0	0	0	0
összeg	200	1		

13

Representation of Data (II)

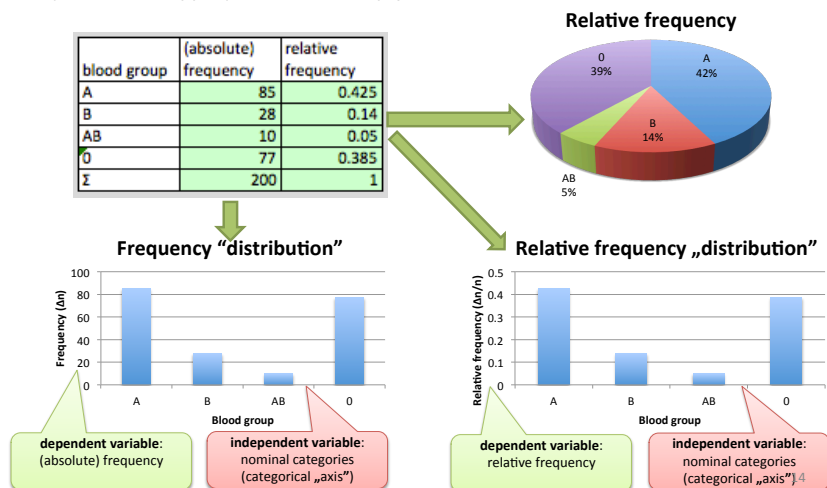
4) Representation of frequencies in case of **quantitative** variables



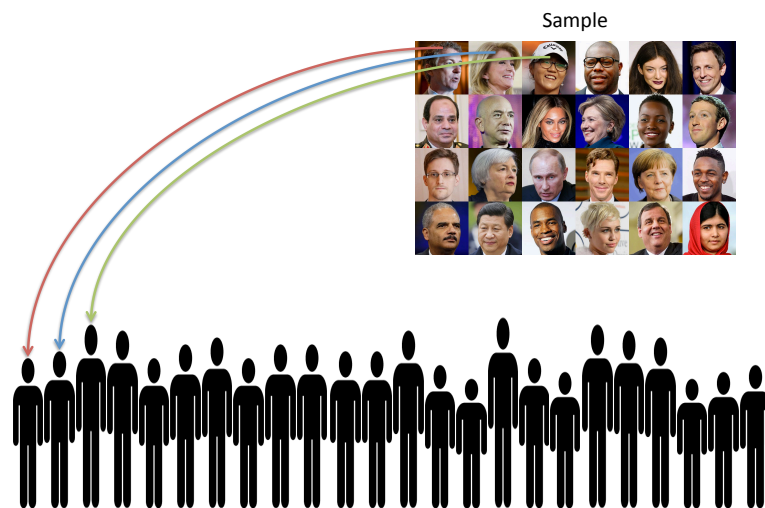
15

Representation of Data (II)

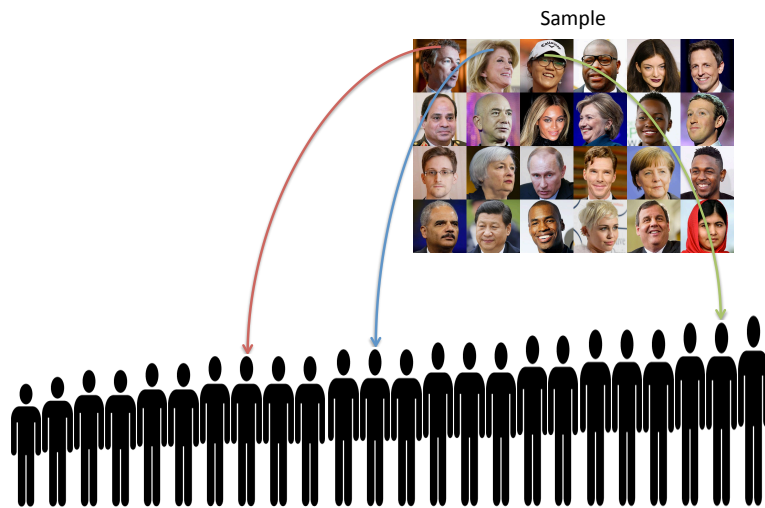
3) Representation of frequencies in case of **qualitative** variables



Composition of the Data

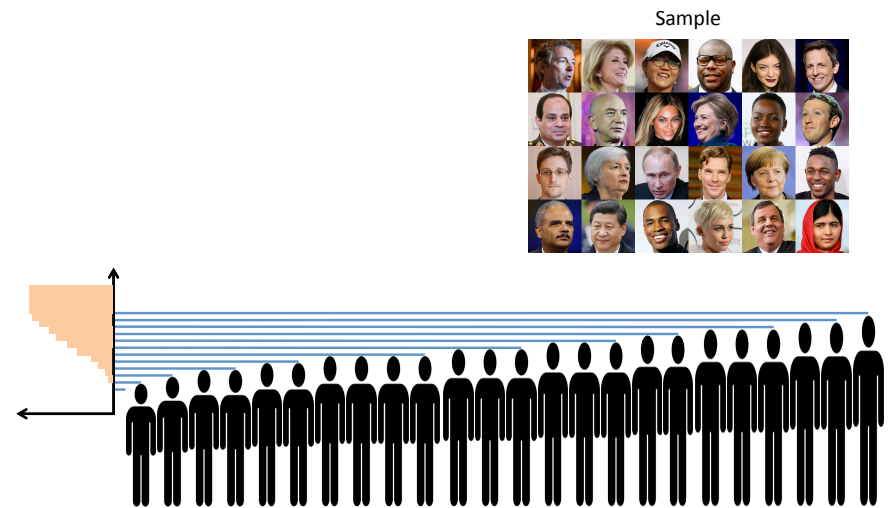


Composition of the Sample

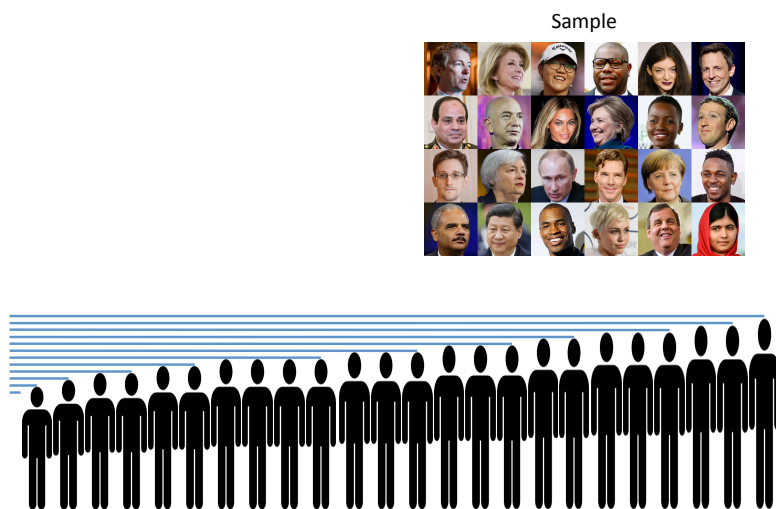


Cumulative Distribution Function

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

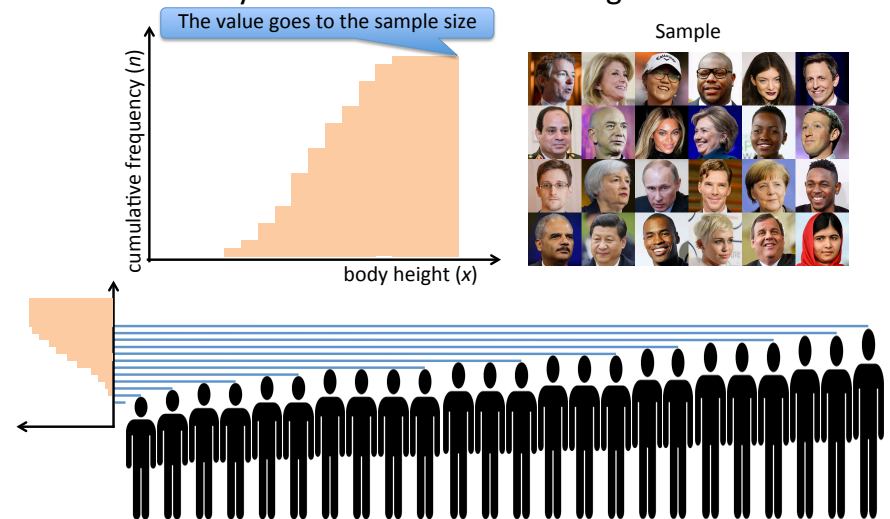


Composition of the Sample



Cumulative Distribution Function

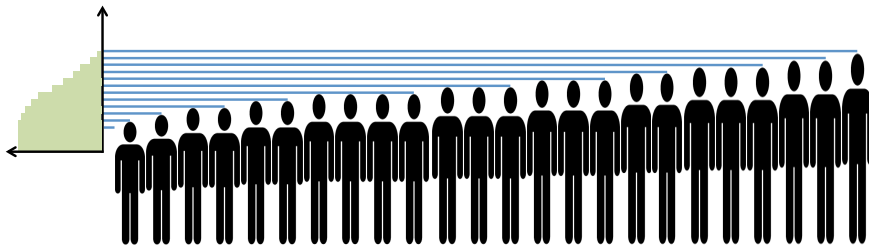
How many elements are **less** than a given x value?



Integral Discrimination Function

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

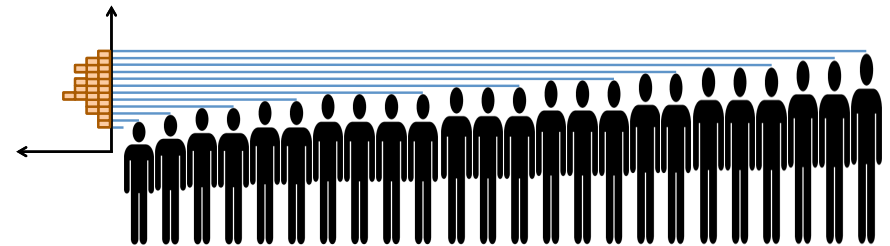
Sample



Frequency Distribution Function

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

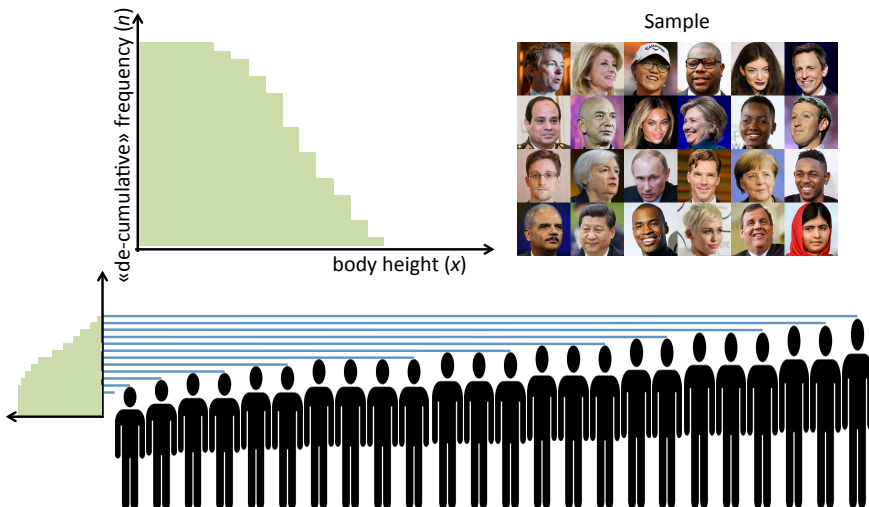
Sample



Integral Discrimination Function

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

Sample

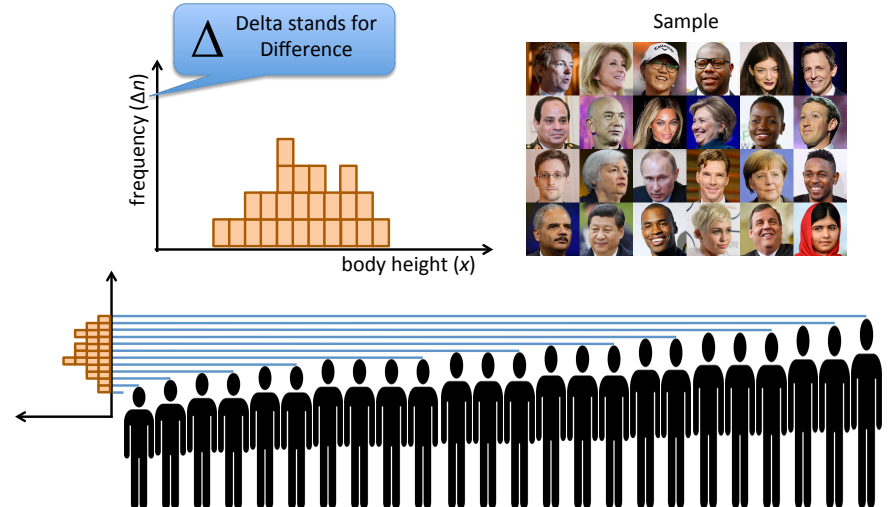


Frequency Distribution Function

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

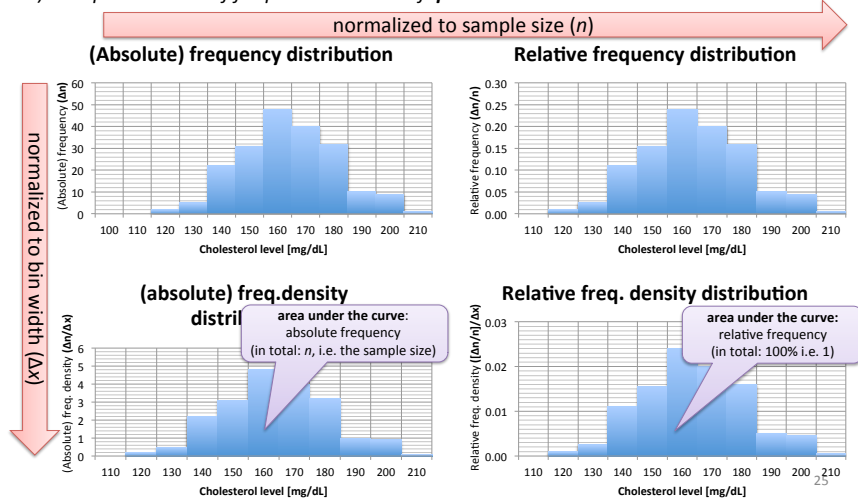
Δ Delta stands for Difference

Sample



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