

BIostatISTICS AND INFORMATICS

INTRODUCTION
MIKLÓS KELLERMAYER

FAITH - CERTAINTY - KNOWLEDGE

Is faith in our knowledge strong enough to bet our life on it?



Murillo: Adoration of the magi



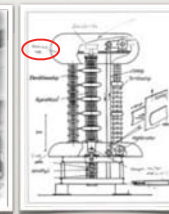
Károly Simonyi (1916-2001)



Accelerator today (ELTE)



Van de Graaff particle accelerator (Sopron, 1951)



Statistics is inherently manifested in human thinking...

POLIOMYELITIS

Poliomyelitis anterior acuta, Heine-Medin disease, infantile paralysis



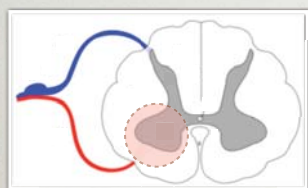
Jakob Heine,
1840



Oskar Medin,
1890



Flaccid paralysis of limb muscles, muscle atrophy, deformation of limbs.



Poliovirus preferably attacks the motor neurons of the anterior horn in the spinal chord.



"Iron lung" (negative-pressure ventilator)

In severe cases (in bulbospinal polio), respiratory failure occurs, requiring life-long respiratory assistance.

POLIO VACCINE



Jonas Salk, 1955
IPV: Intravenous Polio Vaccine



Albert Sabin, 1962
OPV: Oral Polio Vaccine ("Sabin drops")

Is the polio vaccine effective?

Consideration	Problems
The vaccine is simply provided.	Intensity of epidemic varies by itself (solution: comparative study).
Establishment of a Control group	Ethical concerns (reassurance: treatment also carries risks)
Comparison	Different size of control and treated groups (solution: calculate ratios)
Group selection	Lurking variable (e.g., financial background, hygiene) (solution: similar groups - randomization)
Selection of administration method	Effect of subconscious factors (solution: placebo)
Diagnostics	Driven diagnosis (solution: double blind experiment)

	Group size	Incidence
Treated group	200 000	28
Control group	200 000	71

Statistics: science that aids decision making.

FOUNDATIONS OF BIOSTATISTICS

- * Medical decision making
- * Properties of data, variables
- * Random variation
- * Relationship between variables

MEDICAL ACTIVITY

Series of decisions!

The logic of a research scientist and a physician are similar:

Observation	Symptoms
Consideration, hypothesis	preliminary (target) diagnosis
Experiment	Tests (laboratory, imaging)
Theory	Diagnosis



Therapy

LOGIC, ARGUMENT, INFERENCE

Deductive logic:

Specific inference drawn from general premises

1. Every Greek is mortal.
2. Socrates is Greek.
3. Thus, Socrates is mortal.



Diagnostics:

1. Pneumonia is accompanied by fever.
2. This patient has fever.
3. Thus, the patient has pneumonia. (!?)

In medical logic statistical inference is manifested, because it is either not possible or we are unable to consider every circumstance.

Inductive logic:

General inference drawn from specific facts

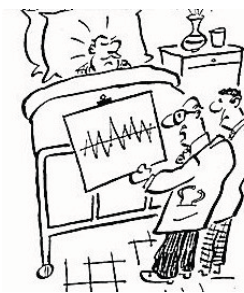
1. This ice cube is cold.
2. Every ice cube is cold.



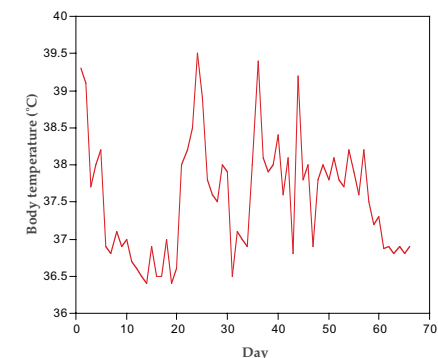
Identification, discovery of new syndromes.

DATA ARE EVERYWHERE

Data (*datum*, *sing.*): qualitative or quantitative characteristic



Fever chart

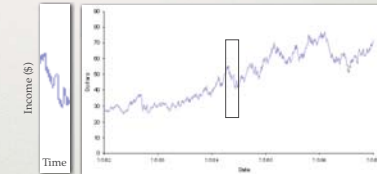


BIOSTATISTICAL THINKING I.

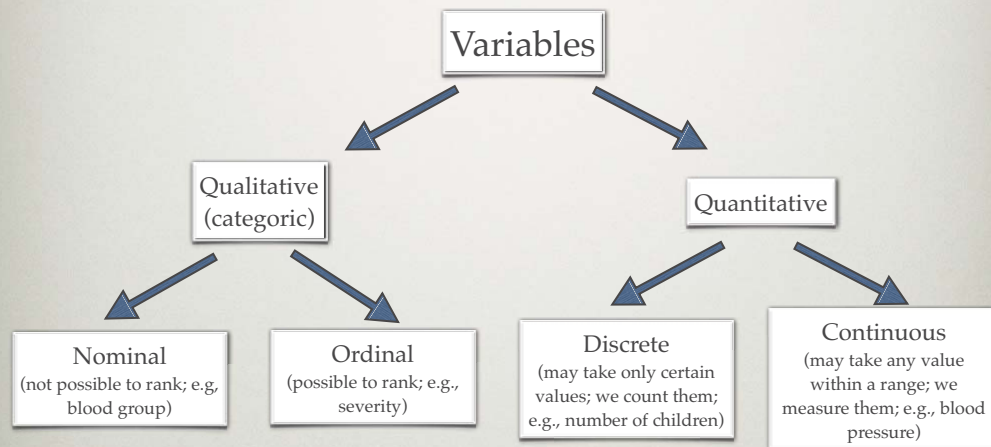
- Data are not “only numbers” (They are numbers with context)
3850 gram infant
- Data beat anecdotes
5 year, \$5 million National Cancer Institute study vs.
TV interview with the mother of a leukemic child
- Beware of the lurking variable
Students playing music excel in other studies...(?)

BIOSTATISTICAL THINKING II.

- Source of data is important
Ann Landers, columnist:
“If you had it to do again, would you have children?”
- Variation is everywhere
CEO: “Shall I fire the sales reps because the income is down?”
- Conclusions are not certain!
Mammograms reduce the risks of death in women aged 50 to 64 years by 26%...



DATA: VALUES OF VARIABLES



There is variation in the values of the variable.

RELATIONSHIP BETWEEN VARIABLES I. DEFINITION - PLOT

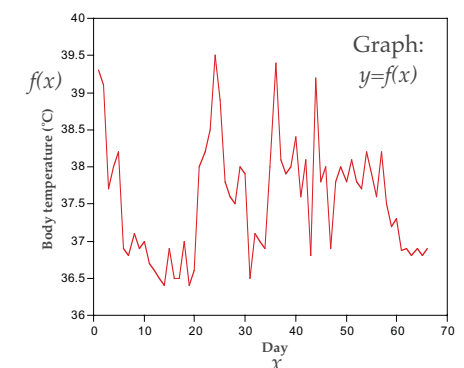
Function: assignment
(between sets;
set: collection of well-defined objects)

x Day	f	$f(x)$ Body temperature (°C)
1.0000	→	39.300
2.0000	→	39.100
3.0000	→	37.700
4.0000	→	38.000
5.0000	→	38.200
6.0000	→	36.900
7.0000	→	36.800
8.0000	→	37.100
9.0000	→	36.900
10.0000	→	37.000
11.0000	→	36.700
12.0000	→	36.600
13.0000	→	36.500
14.0000	→	36.400
15.0000	→	36.900

Range
Independent variable

Set
Dependent variable

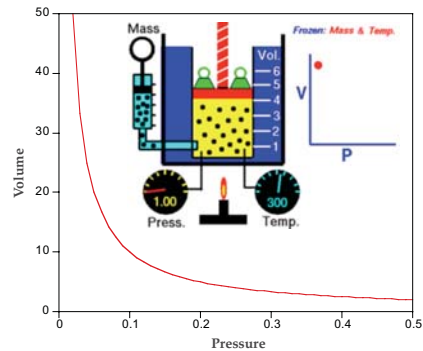
Plotting the function: in xy coordinate system



RELATIONSHIP BETWEEN VARIABLES II.

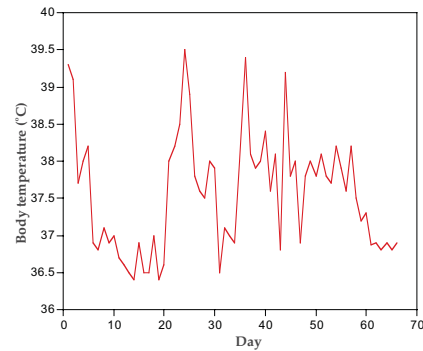
Deterministic

Physical law, e.g., Boyle's law
 $pV = \text{constant}$



Stochastic

Random variations play an important role



SUMMARY

- ***Medical activity:** series of decisions
- ***Biostatistics:** important tool in the quantification of medicine; aids medical decisions
- ***First step of medical activity:** data collection.
- ***Data:** numbers with context; qualitative or quantitative characteristics; values of variables
- ***Functions:** describe relationship between variables.
- ***Random variation:** always present.