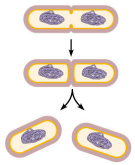
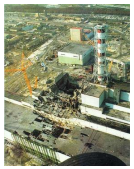
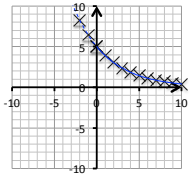
Principles of Biostatistics and Informatics



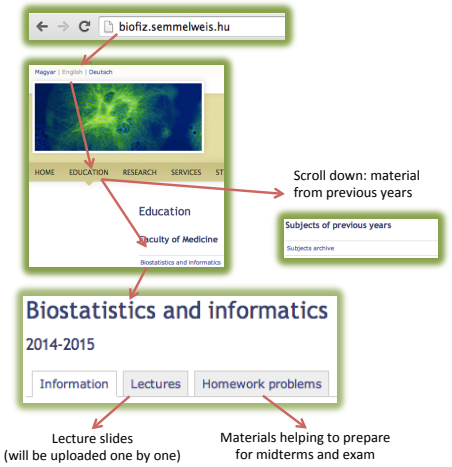
Lecture 1: An Introduction

7th September 2015

Gergely AGÓCS

How to Get Prepared?

- university = **autonomous learning**
- sources:
  - **your** notes made during lectures *(Monday 17^{40}–18^{25}; 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: 18^{00}–19^{30} and Thu: 18^{15}–19^{45}; *1st floor in the EOK building, computer labs from corridor "B")*
  - "Lab Manual of Medical Biophysics" lab practice book:
    - Biostatistics chapter *(40 page summary of theory)*
    - Problems chapter *(problems 71–77)*
  - homepage: biofiz.semmelweis.hu
    - subject requirements
    - lecture schedule and slides
    - lab schedule
    - homework problems
    - matrial from previous years



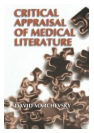
Science and Non-science

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.

Treatment of patients

Evidence based medicine (scientific medicine)



1) Decision making is based on objective evidence in all branches and at all levels of medical care. Evidences serving as base of medical care should be accessible.
2) Medical practitioners should be able to correctly judge the quality of scientific publications, as well as to critically read and understand them.
3) The development of health care requires continuous research.

Alternative or complementary medicine (non-scientific medicine, „quackery")

It is based not on ecidences but on tradition and belief, e.g.: traditional chinese medicine, acupuncture, naturopathy, homeopathy, iridology, osteopathy, cupping, bioresonance, chiropractic, etc.

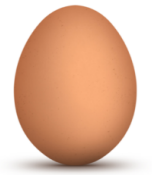

How Does Statistics Help Us?

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

Descriptive statistics → Inferential statistics

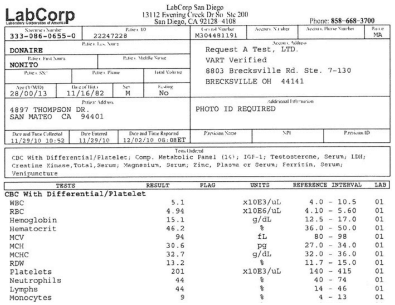
Laws → Deterministic laws → perfect prediction

Laws → Stochastic laws → uncertainty



What Type of Data do We Deal with?

Data to be processed show a high degree of variation …


LabCorp

LabCorp San Diego
13112 Evening Creek Dr So Ste 200
San Diego, CA 92128-4108
Phone: 858-668-3700

333-986-0655-0 | 22247228 | Fasting: NA

DONATER
NONITO

Request A Test, LTD
VART Verified
8803 Brecksville Rd. Ste. 7-130
BRSCKSVILLE OH 44141

28/09/13 | 11/16/82 | M | No

4897 THOMPSON DR
SAN MATEO CA 94401

PHOTO ID REQUIRED

Date and Time Collected: 11/29/10 10:52 | Date Entered: 11/29/10 | Date and Time Reported: 12/02/10 08:08ET

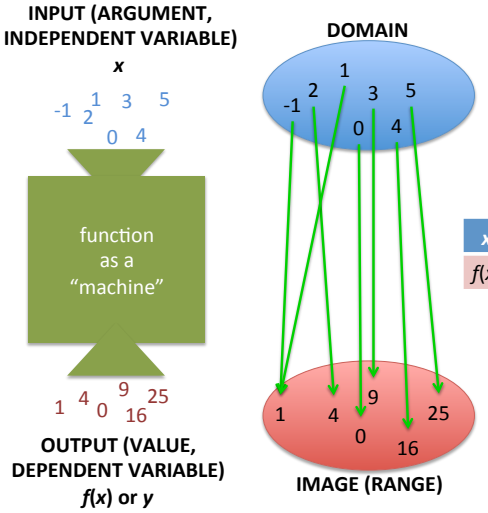
Test Ordered
CBC With Differential/Platelet; Comp. Metabolic Panel (14); IGF-1; Testosterone, Serum, LDH; Creatine Kinase,Total,Serum; Magnesium, Serum; Zinc, Plasma or Serum; Ferritin, Serum; Venipuncture

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	12.6 - 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 is a Function?

Unambiguous assignment of one set of values to an other set of values



$x \mapsto f(x)$ or $y = f(x)$

x	–1	0	1	2	3	4	5
$f(x)$	1	0	1	4	9	16	25

$x \mapsto f(x)$ or $y = f(x)$

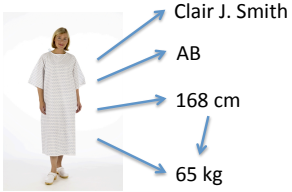
f is the function defining the relationship between *x* and *f*(*x*)

Types of Functions

Unambiguous assignment of one set of values to an other set of values

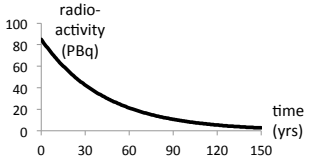
patient ↦ NAME(*patient*)
patient ↦ ABO BLOOD GROUP(*patient*)
patient ↦ BODY HEIGHT(*patient*)
patient ↦ BODY MASS(*patient*)

BODY HEIGHT(*patient*) ↦ BODY MASS(*patient*)

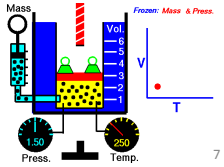


height ↦ FREQUENCY(*height*) pl. 164 cm ↦ 4 people

time ↦ *RADIOACTIVITY(time)*



temperature ↦ *VOLUME(temperature)*



Linear Function

INTEGRAL FORM

VARIABLES: dependent variable, independent variable

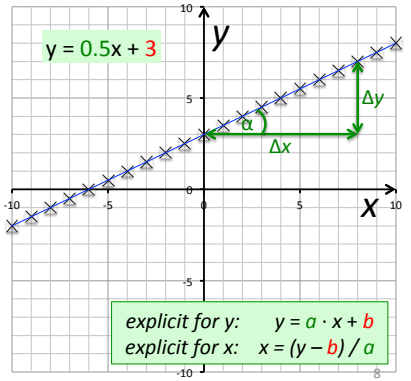
$y = a \cdot x + b$

PARAMETERS: slope (gradient, increment), y-axis intercept

if $x = 0$ then $y = b$

if $\Delta x = 1$ then $\Delta y = a$

$a = \Delta y / \Delta x = \tan\alpha$



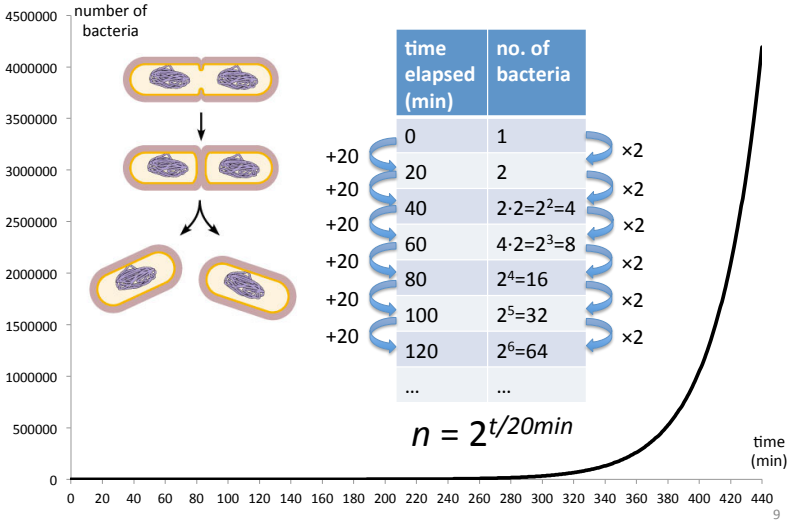
explicit for y: $y = a \cdot x + b$
explicit for x: $x = (y - b) / a$

"DIFFERENTIAL" FORM

$\Delta y \propto \Delta x$

The **change** of the dependent variable is proportional to the **change** of the independent variable

Exponential Function: Example #1

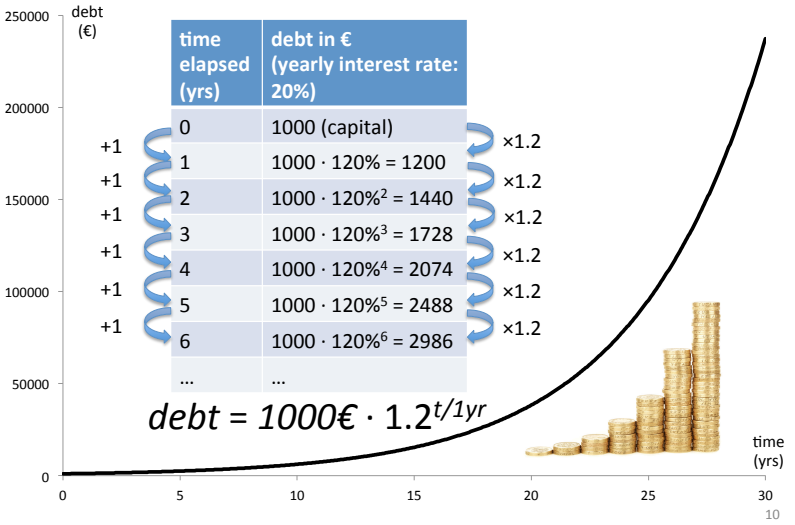


time elapsed (min)	no. of bacteria
0	1
20	2
40	$2 \cdot 2 = 2^2 = 4$
60	$4 \cdot 2 = 2^3 = 8$
80	$2^4 = 16$
100	$2^5 = 32$
120	$2^6 = 64$
...	...

+20 ×2

$$n = 2^{t/20min}$$

Exponential Function: Example #2

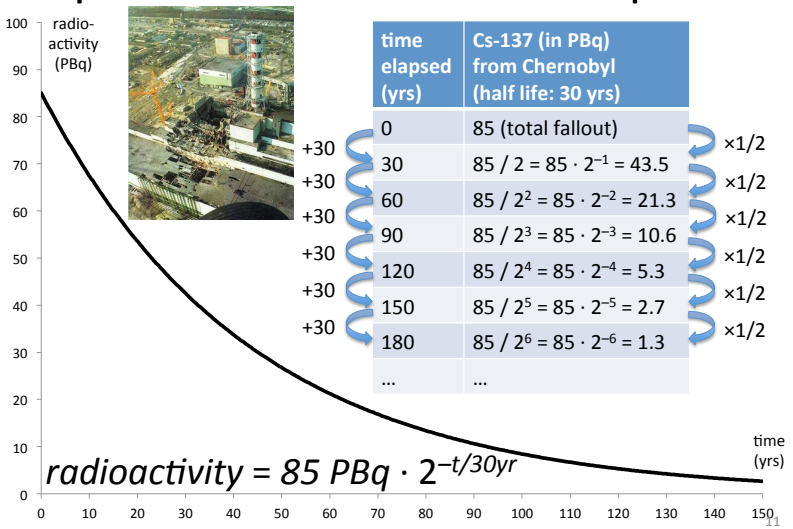


time elapsed (yrs)	debt in € (yearly interest rate: 20%)
0	1000 (capital)
1	$1000 \cdot 120\% = 1200$
2	$1000 \cdot 120\%^2 = 1440$
3	$1000 \cdot 120\%^3 = 1728$
4	$1000 \cdot 120\%^4 = 2074$
5	$1000 \cdot 120\%^5 = 2488$
6	$1000 \cdot 120\%^6 = 2986$
...	...

+1 ×1.2

$$debt = 1000€ \cdot 1.2^{t/1yr}$$

Exponential Function: Example #3



time elapsed (yrs)	Cs-137 (in PBq) from Chernobyl (half life: 30 yrs)
0	85 (total fallout)
30	$85 / 2 = 85 \cdot 2^{-1} = 43.5$
60	$85 / 2^2 = 85 \cdot 2^{-2} = 21.3$
90	$85 / 2^3 = 85 \cdot 2^{-3} = 10.6$
120	$85 / 2^4 = 85 \cdot 2^{-4} = 5.3$
150	$85 / 2^5 = 85 \cdot 2^{-5} = 2.7$
180	$85 / 2^6 = 85 \cdot 2^{-6} = 1.3$
...	...

+30 ×1/2

$$radioactivity = 85\ PBq \cdot 2^{-t/30yr}$$

Exponential Function

INTEGRAL FORM

$$y = b \cdot a^x$$

PRACTICAL MODIFICATIONS:

- the base number is preferred to be *e*
- a new factor parameter *p* (or 1/*k*) is necessary in the exponent
- use a negative sign in the exponent
- *b* is rather denoted by y_0

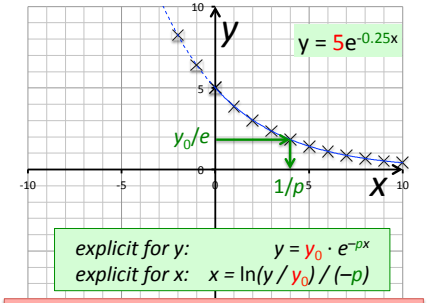
VARIABLES: dependent variable, independent variable

$$y = y_0 \cdot e^{-px} = y_0 \cdot e^{-x/k}$$

PARAMETERS: pre-exponential coefficient, exponential coefficient

if $x = 0$ then $y = y_0$

if $y = y_0/e$ then $x = 1/p = k$



explicit for y: $y = y_0 \cdot e^{-px}$

explicit for x: $x = \ln(y / y_0) / (-p)$

"DIFFERENTIAL" FORM

$$\Delta y/y \propto \Delta x$$

The **relative change** of the dependent variable is proportional to the **change** of the independent variable

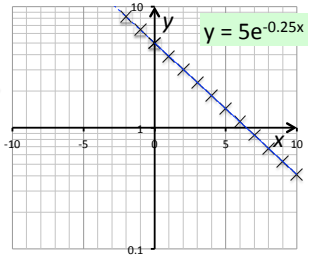
Exponential Function: Linearization

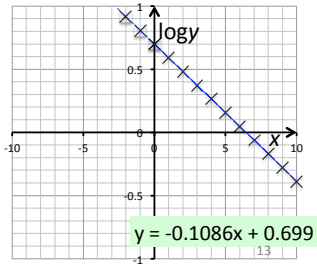
graphical linearization
plot *y* on a log scale as a function of x:
*the relationship **looks** linear but it **is** still exponential*

INTEGRAL FORM

$$y = y_0 \cdot e^{-px}$$

$$\log y = \log\left(y_0 \cdot e^{-p \cdot x}\right)$$
$$\log y = \log y_0 + \log\left(e^{-p \cdot x}\right)$$
$$\log y = \log y_0 - p \cdot x \cdot \log e$$
$$\underbrace{\log y}_{y} = \underbrace{-p \cdot \log e}_{a} \cdot x + \underbrace{\log y_0}_{b}$$



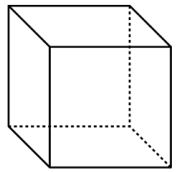


intercept = logy_0
log5 = 0.699
slope = $-p \cdot \log e$
$-0.25 \cdot \log e = -0.1086$

arithmetical linearization
plot log*y* as a function of *x*:
*the relationship **is** linear*

Power Function: Example

mass ∝ volume ∝ [body]length^3
surface area ∝ [body]length^2



Power Function

INTEGRAL FORM

VARIABLES: dependent variable, independent variable

$$y = b \cdot x^a$$

PARAMETERS: pre-exponential coefficient, exponent

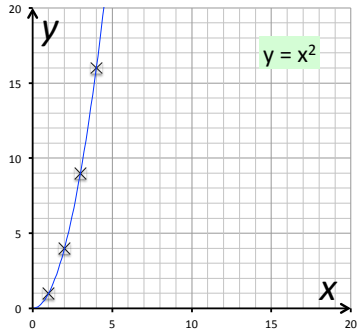
explicit for y: $y = b \cdot x^a$
explicit for x: $x = (y / b)^{1/a}$

"DIFFERENTIAL" FORM

$$\Delta y/y \propto \Delta x/x$$

The **relative change** of the dependent variable is proportional to the **relative change** of the independent variable

if *x* = 1
then *y* = *b*



inverse proportionality and **square root** functions are also power functions

$$y = \frac{b}{x} = b \cdot x^{-1}$$

$$y = \sqrt{x} = x^{\frac{1}{2}}$$

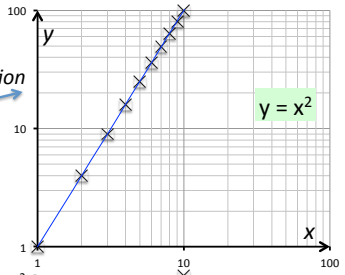
Power Function: Linearization

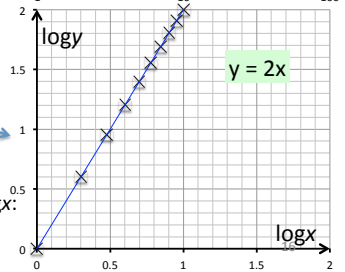
graphical linearization
plot both *y* and x on log scales:
*the relationship **looks** linear but it **is** still power function*

INTEGRAL FORM

$$y = b \cdot x^a$$

$$\log y = \log\left(b \cdot x^a\right)$$
$$\log y = \log b + \log\left(x^a\right)$$
$$\log y = \log b + a \cdot \log x$$
$$\underbrace{\log y}_{y} = a \cdot \underbrace{\log x}_{x} + \underbrace{\log b}_{b}$$





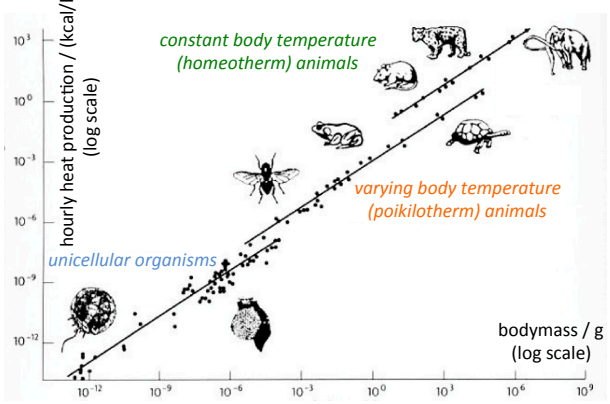
intercept = log*b*
log1 = 0
slope = *a*
a = 2

arithmetical linearization
plot log*y* as a function of log*x*:
*the relationship **is** linear*

Power Function: Example

Allometric scaling
(e.g. Kleiber's law)

hourly heat production ∝ body mass$^{3/4}$



mass ∝ volume ∝ [body]length^3
surface area ∝ [body]length^2