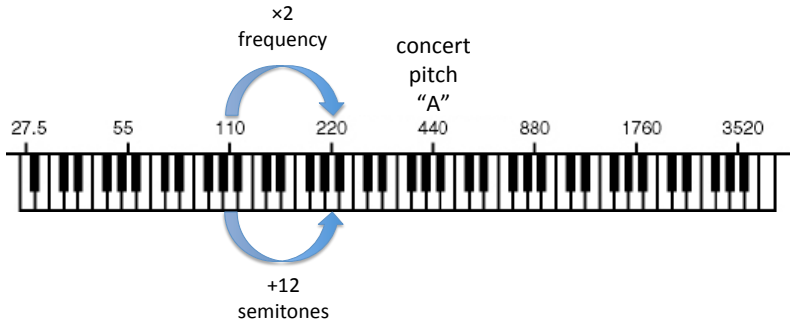
Principles of Biostatistics and Informatics

Lecture 2: Frequency Distributions

14th September 2015

Gergely AGÓCS

Logarithmic Function: Example



Logarithmic Function

INTEGRAL FORM

$$y = b \cdot \log_a(x)$$

PRACTICAL CONSIDERATIONS:

- base is 10 (sometimes *e* or 2)
- if the base is fixed this will modify the factor paramater according to the following identity:

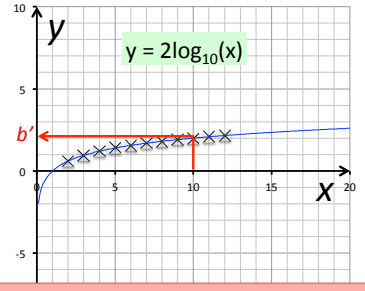
$$b \cdot \log_a(x) = b/\log_{10}(a) \cdot \log_{10}(x) = b' \cdot \log_{10}(x)$$

VARIABLES: dependent variable, independent variable

$$y = b' \cdot \log_{10}(x)$$

PARAMETERS: factor parameter

if $x = 10$ then $y = b'$



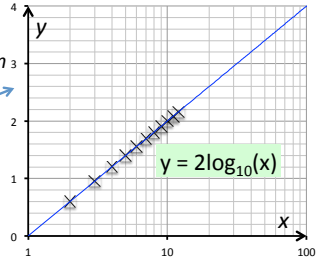
„DIFFERENTIAL" FORM

$$\Delta y \sim \Delta x/x$$

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

Logarithmic Function: Linearization

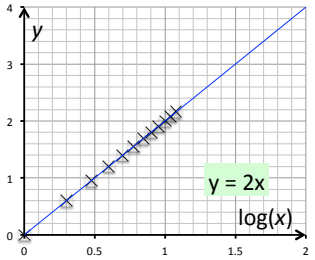
graphical linearization
plot *y* on lin and x on log scales:
*the relationship **looks** linear but it **is** still logarithmic function*



INTEGRAL FORM

$$y = b' \cdot \log_{10}(x)$$

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



Summarizing Functions

LINEAR FUNCTION

$\Delta y \propto \Delta x$

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

EXPONENTIAL FUNCTION

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

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

y vs. x

$\log y$ vs. x

Linearization

y vs. $\log x$

$\log y$ vs. $\log x$

LOGARITHMIC FUNCTION

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

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

POWER FUNCTION

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

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

Derivative and Integral: Example #1

Δ Δ

x	$y = x^2$	$y' = \Delta y/\Delta x$	$y'' = \Delta(\Delta y/\Delta x)/\Delta x$
0	0		
1	1		
2	4		
3	9		
4	16		
5	25		
6	36		
7	49		
8	64		
9	81		
10	100		

Σ Σ

Derivative and Integral: Example #1

Δ Δ

x	$y = x^2$	$y' = \Delta y/\Delta x$	$y'' = \Delta(\Delta y/\Delta x)/\Delta x$
0	0		
1	1	1	
2	4	3	2
3	9	5	2
4	16	7	2
5	25	9	2
6	36	11	2
7	49	13	2
8	64	15	2
9	81	17	2
10	100	19	2

Σ Σ

Derivative and Integral: Example #2

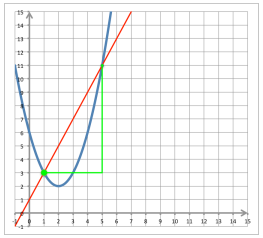
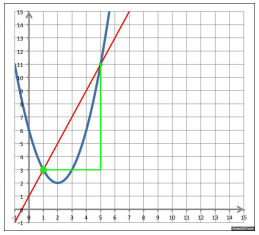
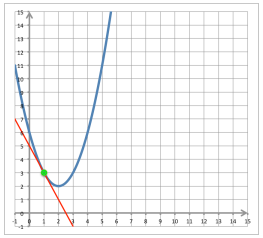
time (s)	distance (m)	speed (m/s)	acceleration (m/s^2)
0	0	0	10
1	5	10	10
2	20	20	10
3	45	30	10
4	80	40	10
5	125	50	10
6	180	60	10
7	245	70	10
8	320	80	10
9	405	90	10
10	500	100	10

Derivative: slope of tangent line

difference quotient:
$\Delta y/\Delta x$
slope of **secant** line

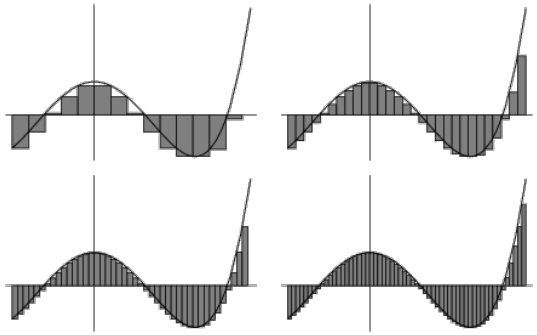
$\Delta \longrightarrow d$

derivative:
dy/dx
slope of **tangent** line

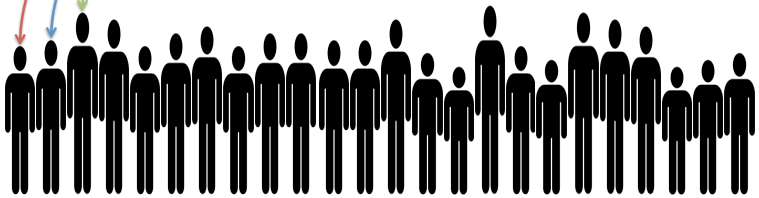
Integral: Area Under the Curve (AUC)

$\Sigma \longrightarrow \int$



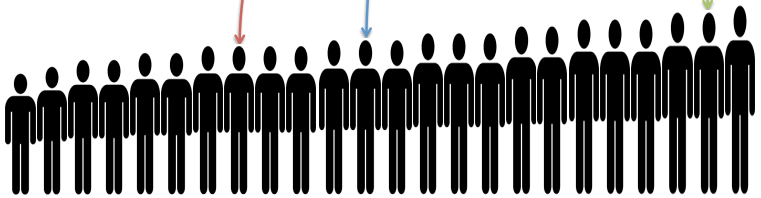
Composition of the Data Set

Sample

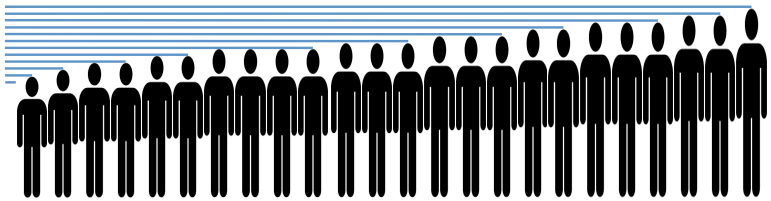


Composition of the Data Set

Sample

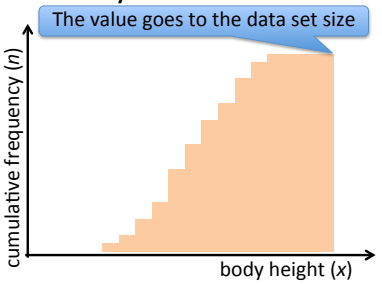


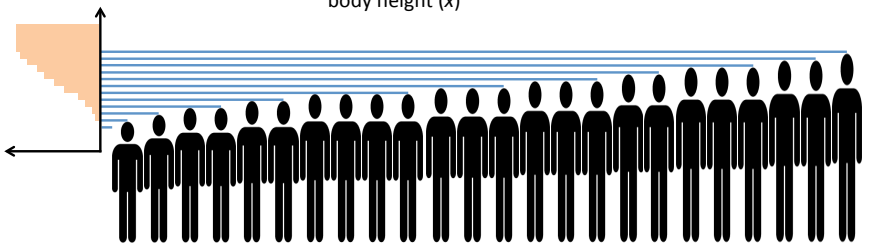
Composition of the Data Set



Cumulative Distribution Function

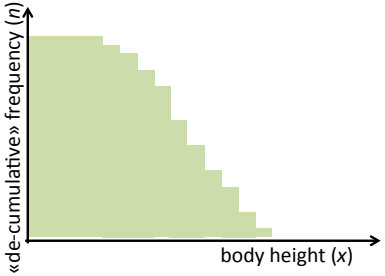
How many elements are **less** than a given *x* value?

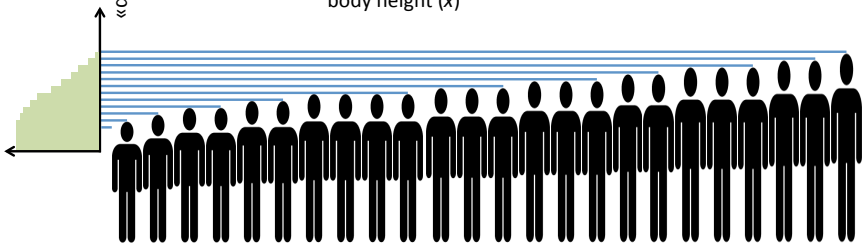




Integral Discrimination Function

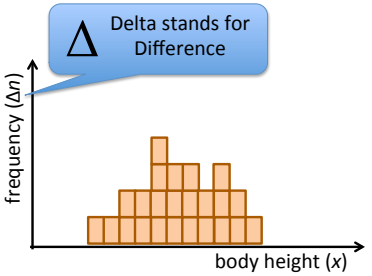
How many elements are **greater** than a given *x* value?

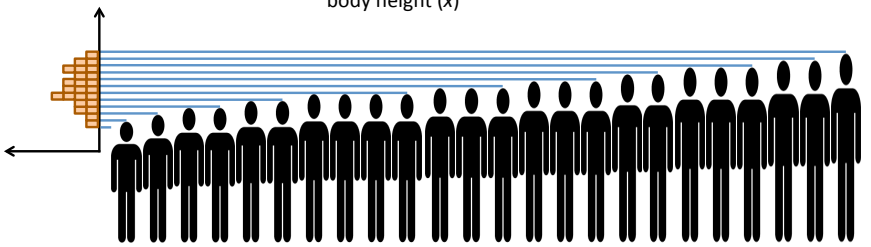




Frequency Distribution Function

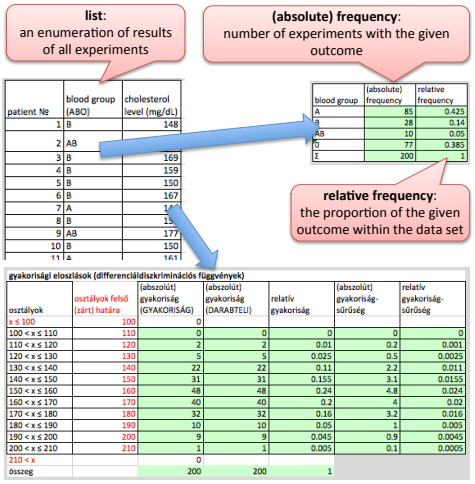
How many elements fall **within** a bin of Δ*x* width?





Representation of Data Sets (I)

1) *A simple list of the 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 ($[\Delta n/n]/\Delta x$)



patient №	blood group (ABO)	cholesterol level (mg/dL)
1	B	148
2	AB	
3	B	169
4	B	159
5	B	150
6	B	167
7	A	14
8	B	15
9	AB	177
10	B	150
11	A	161

blood group	(absolute) frequency	relative frequency
A	85	0.425
B	28	0.14
AB	10	0.05
0	77	0.385
Σ	200	1

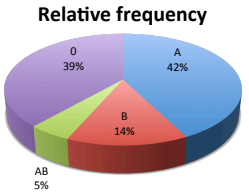
gyakorisági eloszlások (differenciáldiszkriminációs függvények)

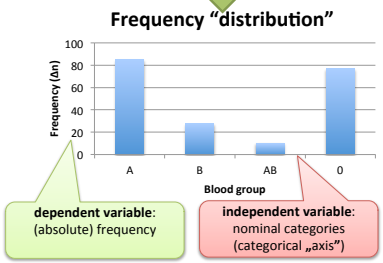
osztályok	osztályok felső (zárt) határa	(abszolút) gyakoriság (GYAKORISÁG)	(abszolút) gyakoriság (DARABTEL)	relatív gyakoriság	(abszolút) gyakoriság-sűrűség	relatív gyakoriság-sűrűség
x ≤ 100	100	0				
100 < x ≤ 110	110	0	0	0	0	0
110 < x ≤ 120	120	2	2	0.01	0.2	0.001
120 < x ≤ 130	130	5	5	0.025	0.5	0.0025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

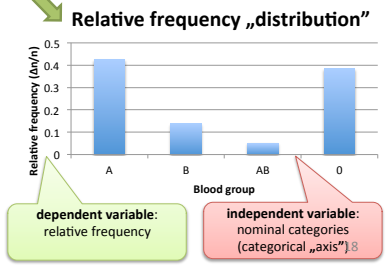
Representation of Data Sets (II)

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
0	77	0.385
Σ	200	1







Representation of Data Sets (III)

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