

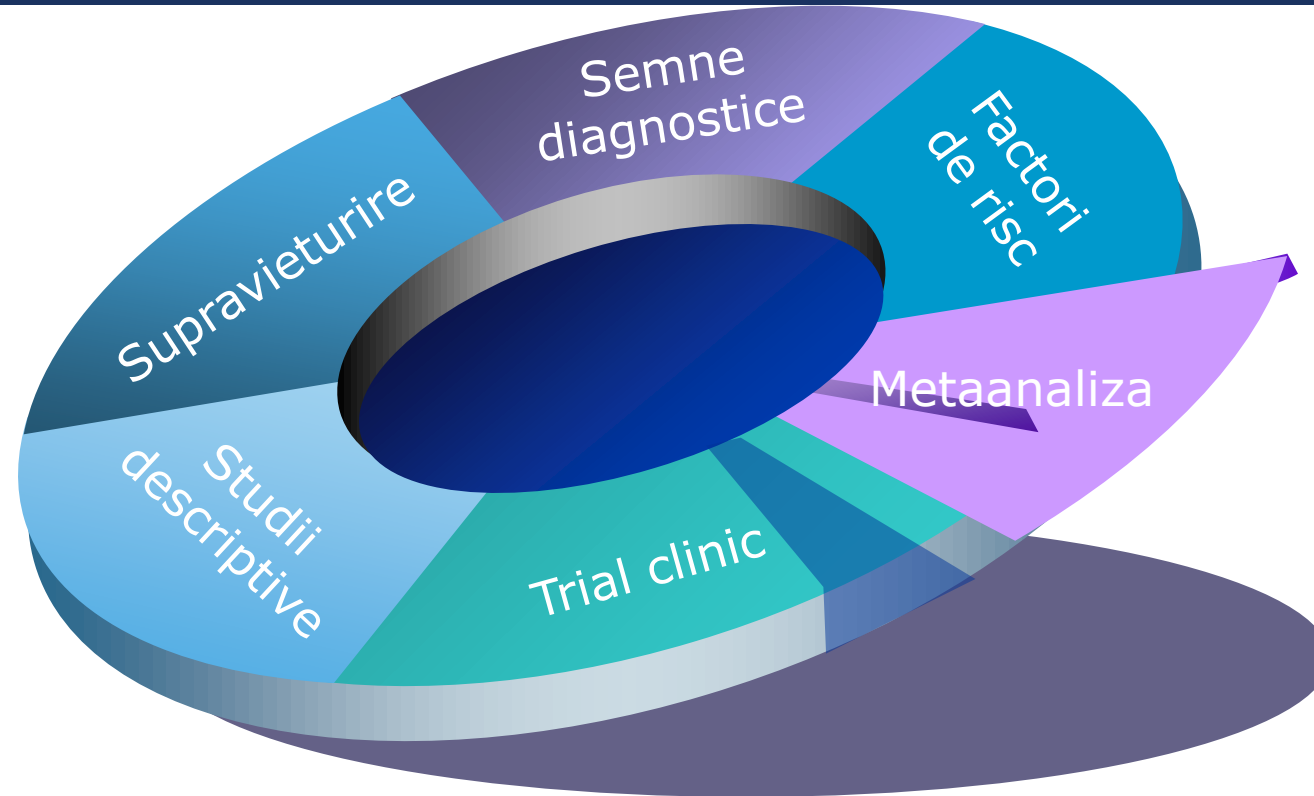


STUDII DESCRIPTIVE

MODELAREA FENOMENELOR MEDICALE

TUDOR DRUGAN 2023

STUDII, STUDII, STUDII



STUDII DESCRIPTIVE

Definiție

Descriptiv vs Analitic

De ce?

MODELAREA FENOMENELOR MEDICALE

Corelații

Regresii

Exemple

STUDII DESCRIPTIVE

Prezent

Se ameliorează?

Funcționează?

Situația actuală

Intervenția

Situația ameliorată

STUDIILE ANALITICE

Prezent
cauze

De ce cauzele persistă?

Funcționează? De ce?

Situația actuală

Intervenția

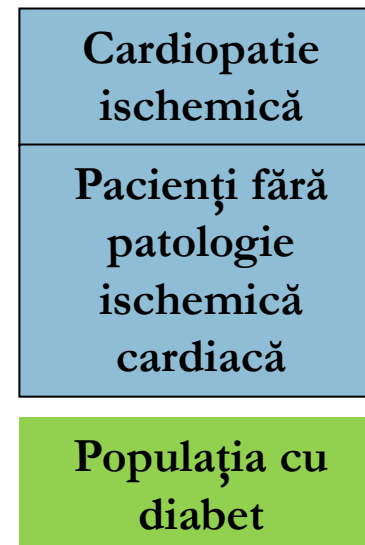
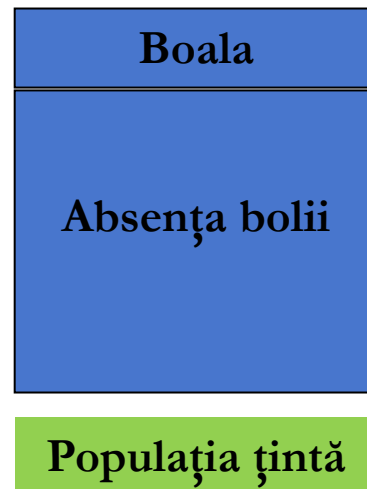
Situația ameliorată

STUDII DESCRIPTIVE

- ✖ Studiile descriptive nu trebuie privite numai ca și complement sau supliment al cercetărilor analitice.
- ✖ Se pot folosi și pentru a genera ipoteze în ceea ce privește soluțiile unor probleme, ipoteze care ulterior pot fi testate prin metode cantitative bazate pe rezultatele studiului descriptive sau o combinație de metode descriptive și analitice.

STUDII DE PREVALENȚĂ

- × Studiu descriptiv
- × Se determină prevalența unei boli la o populație.





CORELAȚII ȘI REGRESII

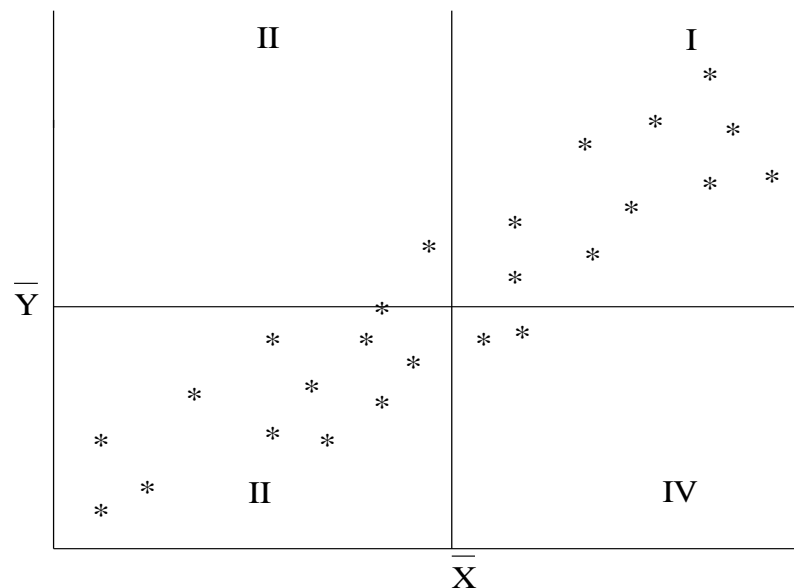
CUM SE TRANSFORMĂ STUDIILE DESCRIPTIVE ÎN ANALITICE 2 ?

STATISTICI DESCRIPTIVE ÎN DOUĂ DIMENSIUNI

- Varsta X : X_1, X_2, \dots, X_n
- TAS Y : Y_1, Y_2, \dots, Y_n .
- Să se stabilească dacă există o legătură între variabilele X și Y și să se determine o modalitate de a măsura intensitatea acestei legături.
 - Coeficientul de corelație
- Să se stabilească legătura dintre cele două variabile
 - Funcția de regresie

DIAGRAMA DE DISPERSIE

- Diagrama de dispersie asociată unei tabel de date bidimensional:
 - $X: X_1, X_2, \dots, X_n$
 - $Y: Y_1, Y_2, \dots, Y_n$
- se obține reprezentând grafic punctele de coordonate (X_i, Y_i) $i=1, 2, \dots, n$.



RELAȚIA DE CAUZALITATE

- Întotdeauna $X \Rightarrow Y$
- Atenție la factorii cofondatori sau la variabilele derivate
- Pentru a obține un indicator independent și de unitățile de măsură ale celor două variabile se utilizează coeficientul de corelație sau coeficientul Bravais-Pearson:

$$r = \frac{COV(X,Y)}{S_x \cdot S_y} = \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{[n \sum X^2 - (\sum X^2)][n \sum Y^2 - (\sum Y^2)]}}$$

REGULILE EMPIRICE COLTON (1974)

- un coeficient de corelație de la -0.25 la 0,25 înseamnă **o corelație slabă sau nulă**,
- un coeficient de corelație de la 0.25 la 0.50 (sau de la -0.25 la -0.50) înseamnă **un grad de asociere acceptabil**
- un coeficient de corelație de la 0.5 la 0.75 (sau de la -0.5 la -0.75) **înseamnă o corelație moderată spre bună**
- un coeficient de corelație mai mare decât 0.75 (sau mai mic decât -0.75) înseamnă **o foarte bună asociere sau corelație**

TESTUL DE SEMNIFICAȚIE PENTRU COEFICIENTUL DE CORELAȚIE PEARSON

- **Semnificația coeficientului de corelație Pearson poate fi evaluată dacă valoarea observată a apărut datorită întâmplării**
 - dacă testul este semnificativ probabilitatea este mai mică de 0,05
 - Interpretarea este dată de valorile lui r
- **Dacă p este mai mare de 0,05 interpretarea este că datele experimentale nu ne permit enunțarea existenței unei relații între variabilele luate în calcul**
 - $r = 0$

<https://www.xltoolbox.net/>



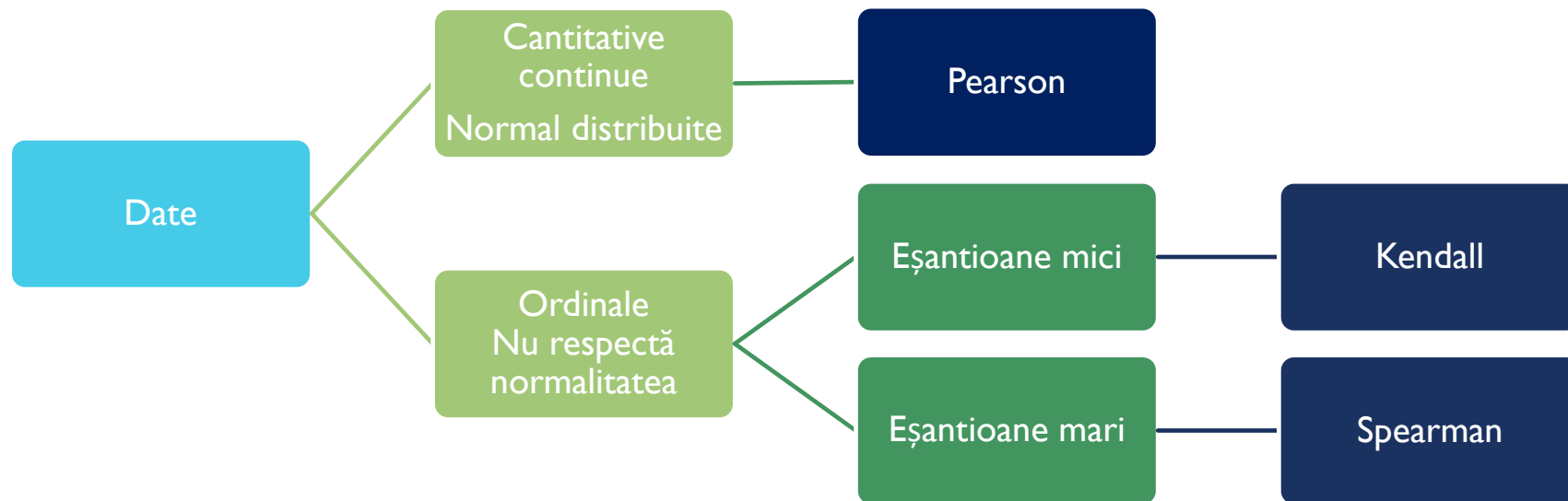
INTERPRETAREA R,P

Valoarea r	$p > 0,05$	$p < 0,05$
-0.25 la 0,25	corelație slabă sau nulă	corelație slabă sau nulă
0.25 la 0.50 (-0.25 la -0.50)	Nu are semnificatie statistica	un grad de asociere acceptabil
0.5 la 0.75 (-0.5 la -0.75)	Nu are semnificatie statistica	o corelație moderată spre bună
0.75 (sau mai mic decât -0.75)	Nu are semnificatie statistica	o foarte bună asociere sau corelație

INTERPRETAREA R,P

- **Atentie:**
 - $p > 0,05$ coeficientul descrie doar esantionul din care a fost extras
 - Variabilele corelate trebuie sa fie **cantitative continue**
 - **Datele trebuie să fie normal distribuite**
 - Intre variabilele pentru care se calculează corelație trebuie sa **existe o relație de cauzalitate**
 - <https://www.graphpad.com/quickcalcs/linear1/>

CE COEFICIENT SĂ ALEG?



CORELAREA RANGURILOR: COEFICIENTUL DE CORELAȚIE SPEARMAN

- Coeficientul de corelație Spearman, notat r_s , este analogul nonparametric al coeficientul de corelație Pearson, calculat pentru a fi utilizat cu date ordinale sau date care nu respectă distribuția normală.
- Semnificația coeficientului de corelație Spearman poate fi evaluată dacă valoarea observată a apărut datorită întâmplării (dacă este mai mic de 0,05 regresia este și semnificativă statistic).
- Coeficientul de corelație Spearman măsoară gradul de asociere dintre două variabile

CORELAREA RANGURILOR: COEFICIENTUL DE CORELAȚIE KENDALL

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)}$$

- Coeficientul de corelație Kendal tau, notat τ puterea asocierii dintre două variabile ordinale

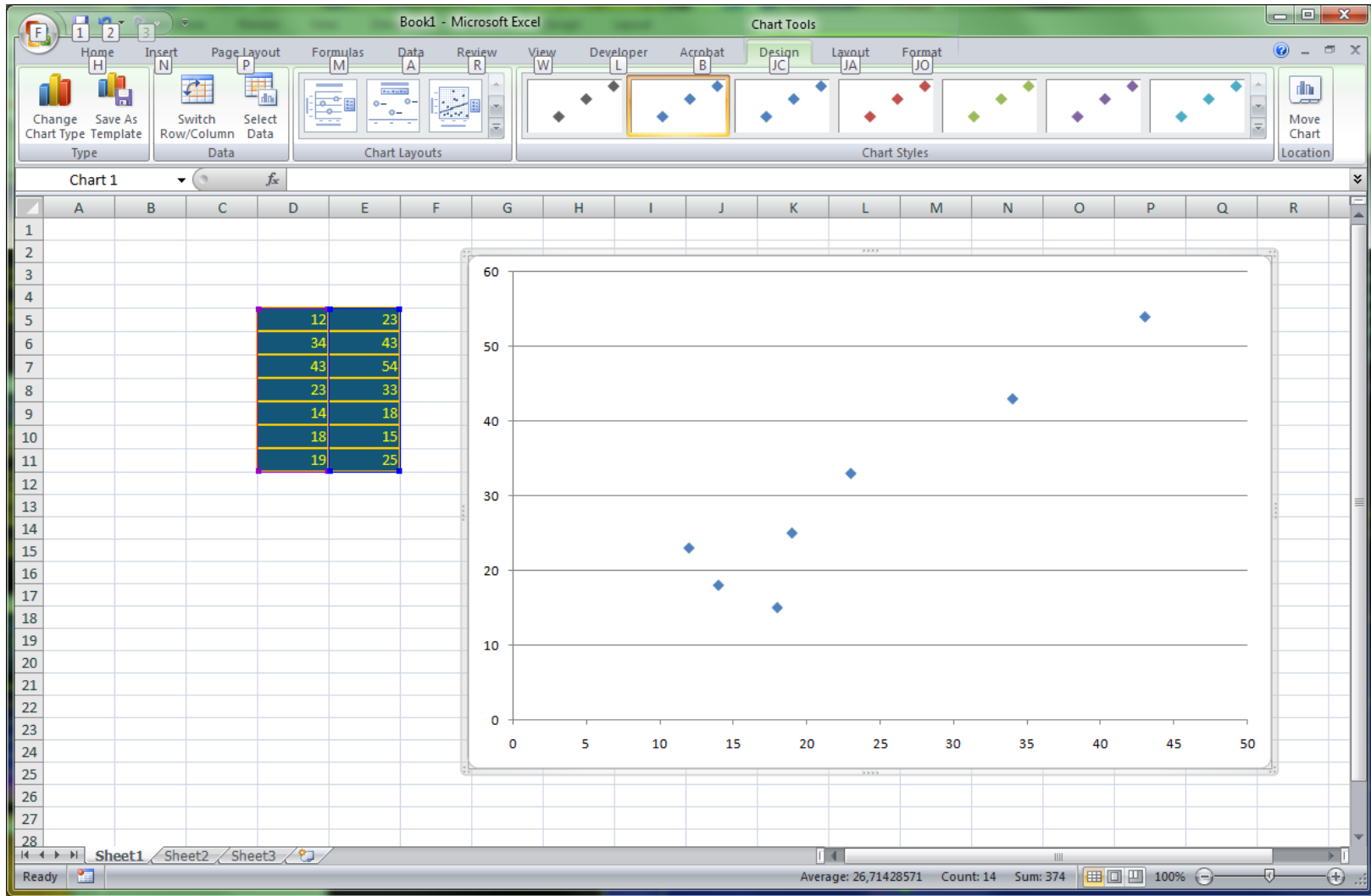
KENDALL VS SPEARMAN

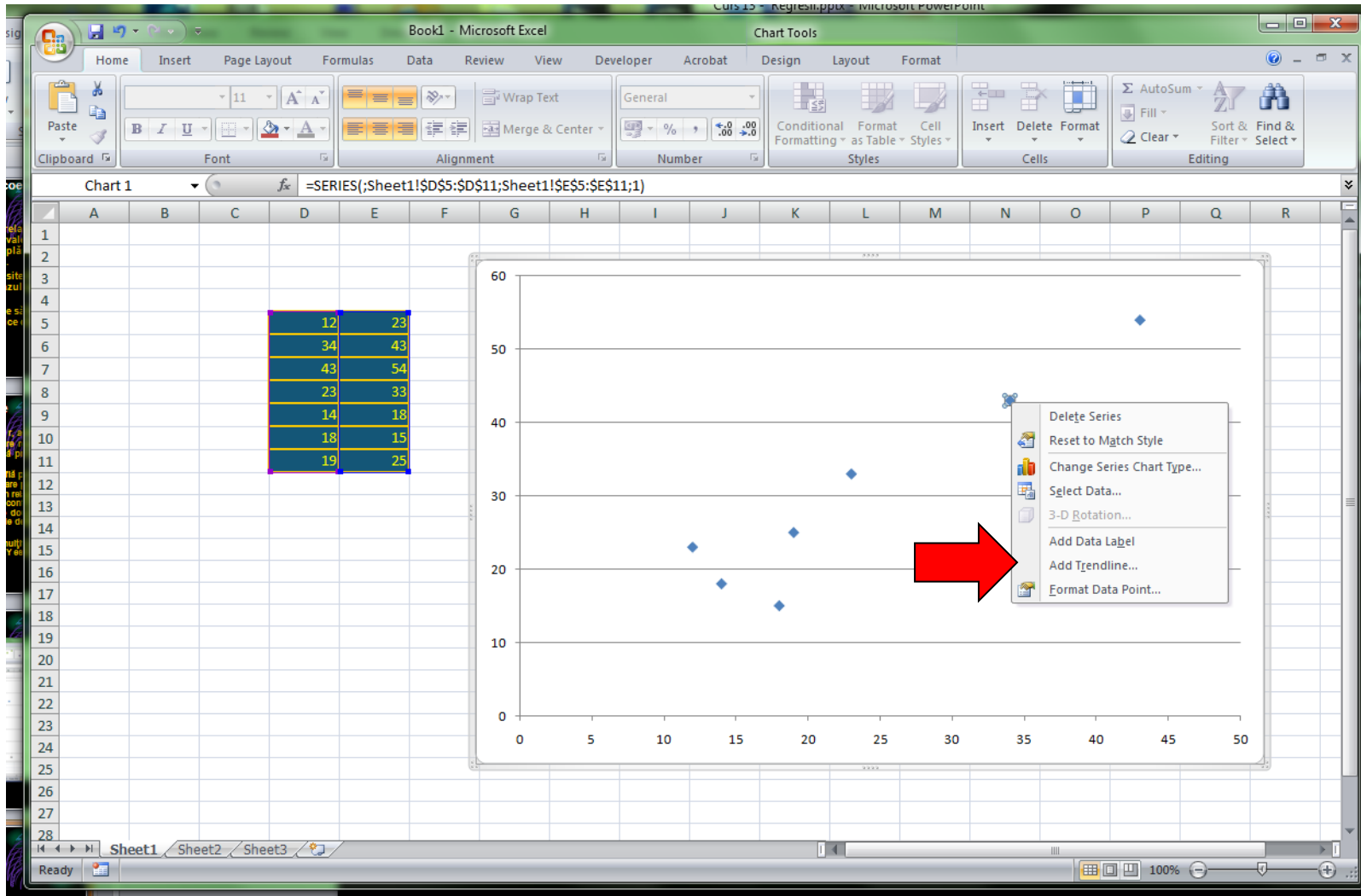
- În mod normal, corelația Kendall este mai eficientă decât cea Spearman și este preferată dacă avem eșantioane mici sau valori aberante (outliers).
- Spearman's rho este mai mare de obicei decât Kendall's tau.

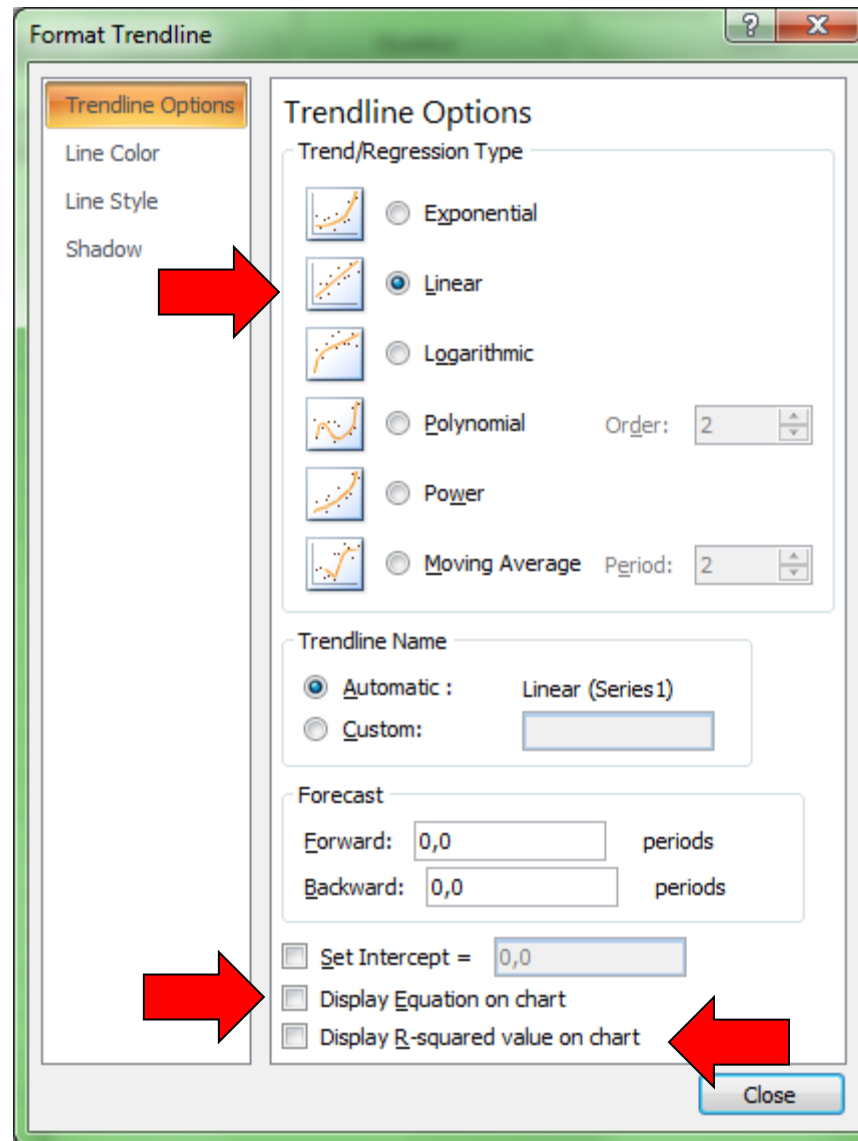


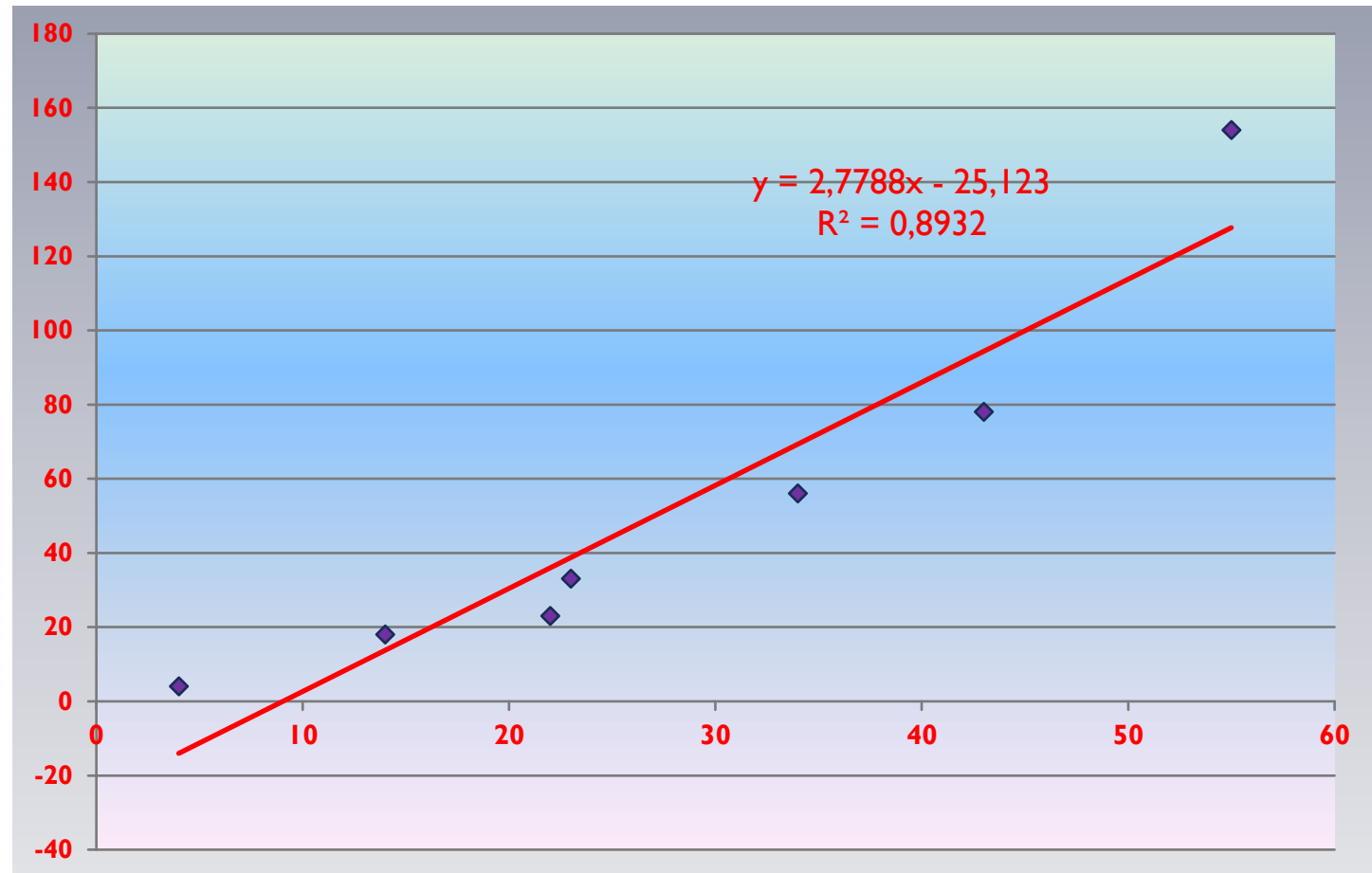
REGRESIA











EXEM

Final General Concepts 128

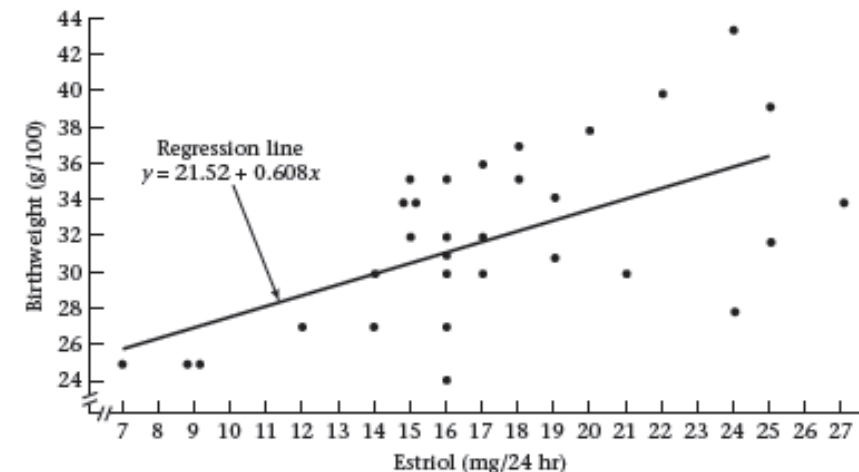
Sample data from the Greene-Touchstone study relating birthweight and estriol level in pregnant women near term

i	Estriol (mg/24 hr) x_i	Birthweight (g/100) y_i	i	Estriol (mg/24 hr) x_i	Birthweight (g/100) y_i
1	7	25	17	17	32
2	9	25	18	25	32
3	9	25	19	27	34
4	12	27	20	15	34
5	14	27	21	15	34
6	16	27	22	15	35
7	16	24	23	16	35
8	14	30	24	19	34
9	16	30	25	18	35
10	16	31	26	17	36
11	17	30	27	18	37
12	19	31	28	20	38
13	21	30	29	22	40
14	24	28	30	25	39
15	15	32	31	24	43
16	16	32			

Source: Reprinted with permission of the American Journal of Obstetrics and Gynecology, 85(1), 1-9, 1963.

$$E(y|x) = \alpha + \beta x$$

Data from the Greene-Touchstone study relating birthweight and estriol level in pregnant women near term



Source: Reprinted with permission of the American Journal of Obstetrics and Gynecology, 85(1), 1-9, 1963.

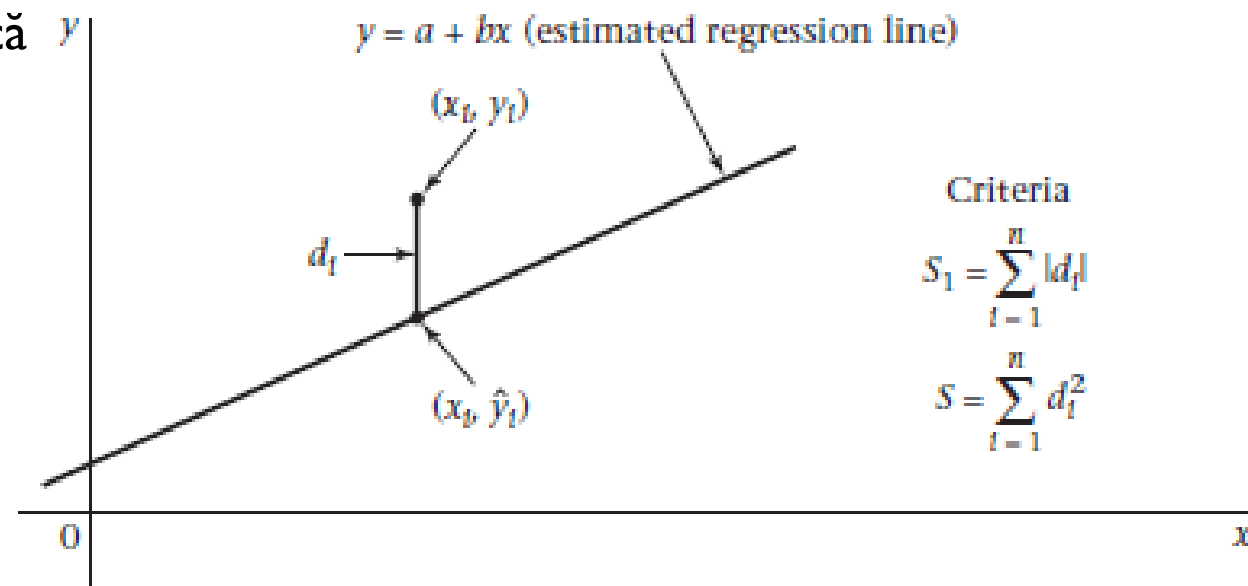
OBȚINEREA DREPTEI DE REGRESIE

Dreapta de regresie este acea funcție matematică pentru care

$$S = \sum_{i=1}^n d_i^2$$

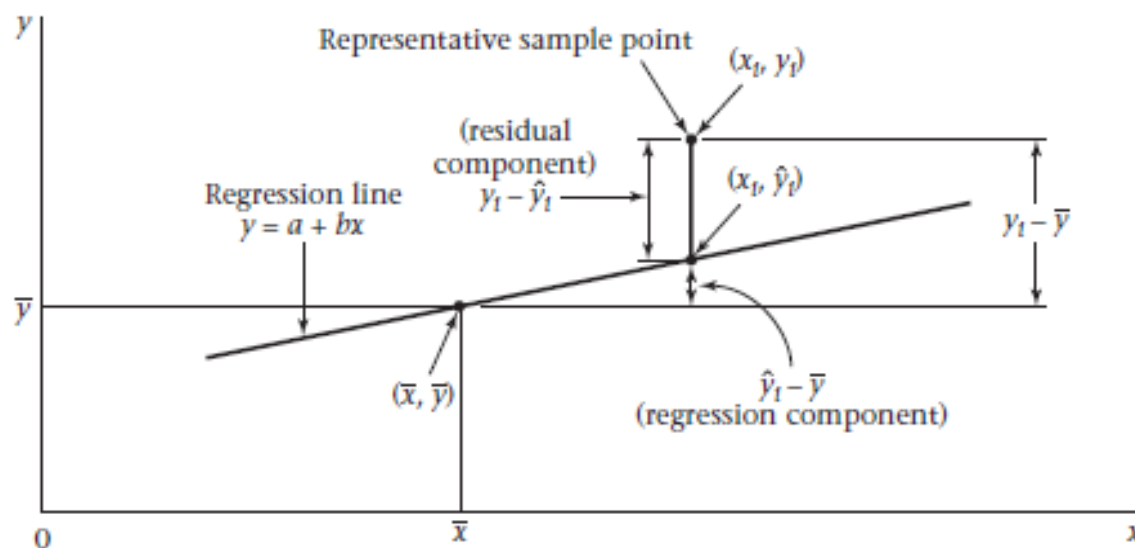
are o valoare minimă (metoda celor mai mici pătrate).

! Corolar pentru regresia liniară coeficienții dreptei a și b se deduc din valorile seriei, pentru alte forme de funcții de cele mai multe ori softurile de regresie folosesc soluții iterative.



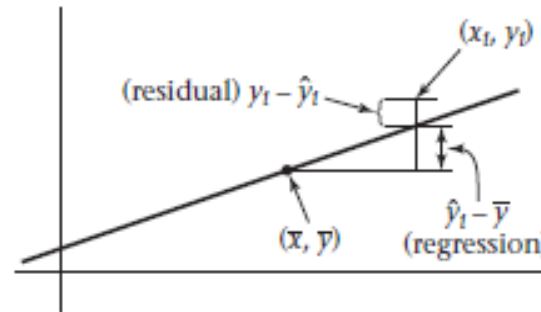
REZIDUALE...

- Orice punct care nu se află pe dreapta de regresie este la o distanță de aceasta numită reziduală

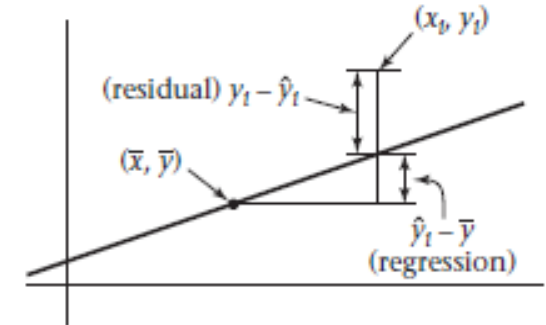


COMPONENTA DE REGRESIE ȘI CEA REZIDUALĂ

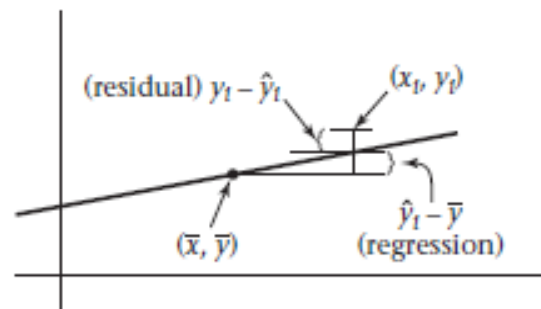
- O regresie “bună” are o componentă de regresie în general mai mare decât cea reziduală
- Metoda de apreciere a calității regresiei



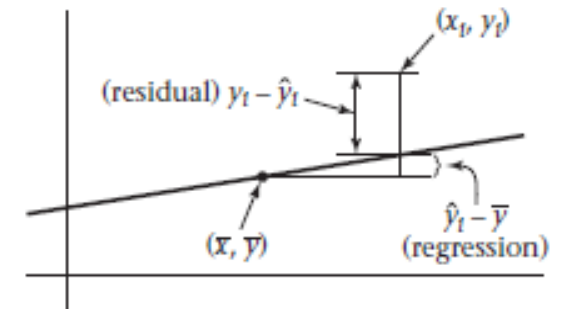
(a) Large regression, small residual components



(b) Large regression, large residual components



(c) Small regression, small residual components



(d) Small regression, large residual components

EXEMPLU SIGMAPLOT

Equation: Polynomial; Linear

$f = y_0 + a \cdot x$

R	Rsqr	Adj Rsqr	Standard Error of Estimate
0,9444	0,8918	0,8702	5,0881

	Coefficient	Std. Error	t	P
y0	2,6087	4,6996	0,5551	0,6027
a	1,1824	0,1842	6,4211	0,0014

Analysis of Variance:

	DF	SS	MS
Regression	2	7427,5552	3713,7776
Residual	5	129,4448	25,8890
Total	7	7557,0000	1079,5714

Corrected for the mean of the observations:

	DF	SS	MS	F	P
Regression	1	1067,4123	1067,4123	41,2304	0,0014
Residual	5	129,4448	25,8890		
Total	6	1196,8571	199,4762		

Statistical Tests:

Normality Test (Shapiro-Wilk)	Passed	(P = 0,3520)
-------------------------------	--------	--------------

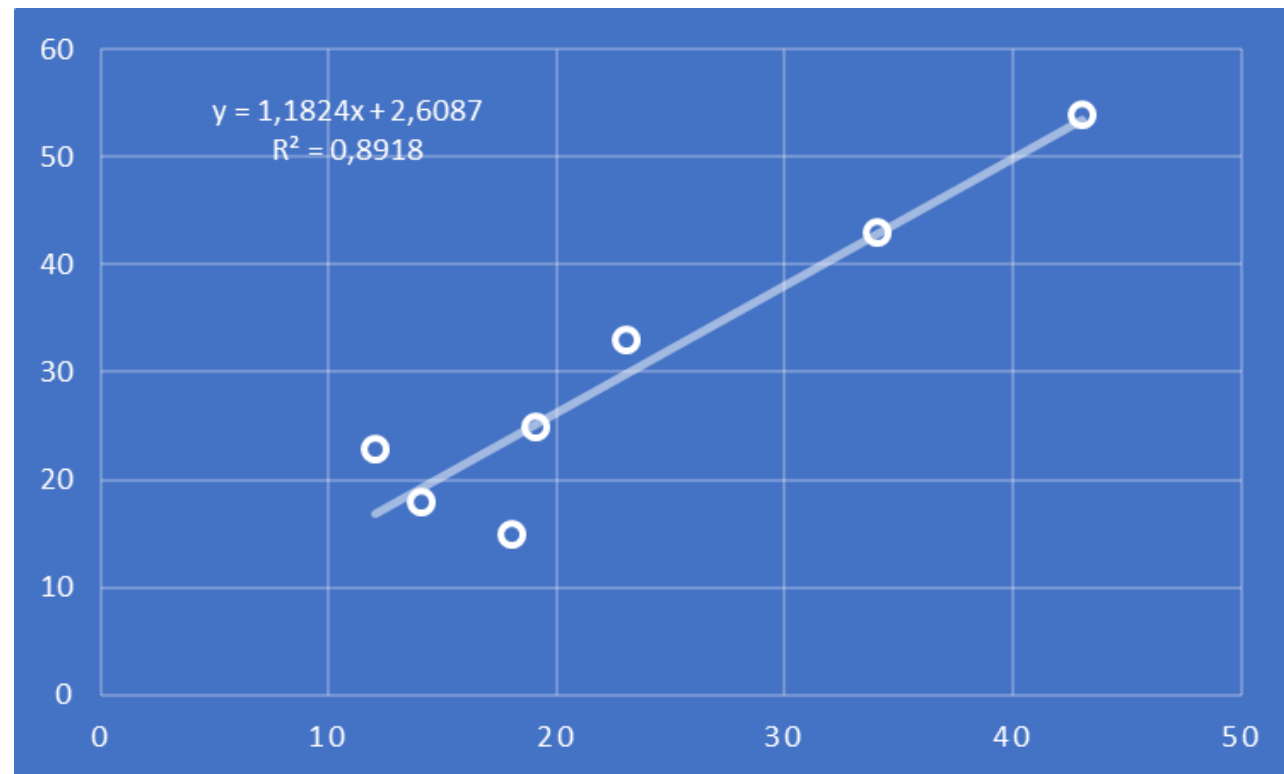
$$f = y_0 + a \cdot x$$

$$y_0 = 2,6087$$

$$a = 1,1824$$

$$f = 1,1824x + 2,6087$$

R	Rsqr	Adj Rsqr	Standard Error of Estimate
0,9444	0,8918	0,8702	5,0881





REGRESIA MULTIPLĂ



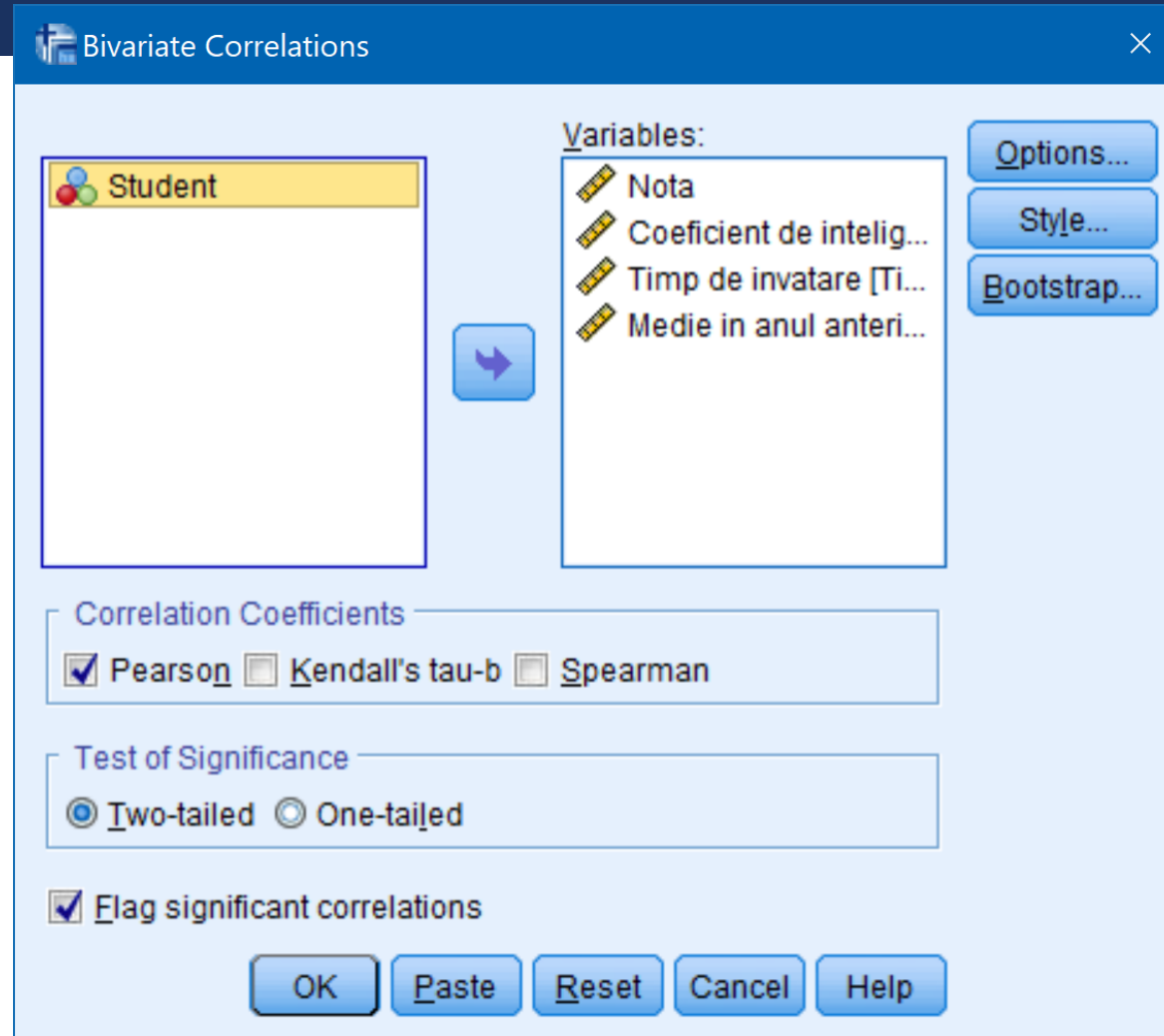
INTERPRETAREA ÎN SPSS A REGRESIEI MULTIPLE

SPSS Statistics Data Editor window showing a dataset named "Untitled2 [DataSet1]". The window displays the Variable View tab, showing the structure of the dataset with 29 rows and 12 columns.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Student	Numeric	8	0		None	None	8	Right	Nominal	Input
2	Nota	Numeric	8	0		None	None	8	Right	Scale	Input
3	IQ	Numeric	8	0	Coeficient de in...	None	None	8	Right	Scale	Input
4	Timp	Numeric	8	2	Timp de invatare	None	None	8	Right	Scale	Input
5	Medie	Numeric	8	2	Medie in anul a...	None	None	8	Right	Scale	Input
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											
25											
26											
27											
28											
29											

The bottom of the window shows the "Data View" and "Variable View" tabs, with "Variable View" currently selected.

INTERPRETAREA ÎN SPSS A REGRESIEI MULTIPLE



INTERPRETAREA ÎN SPSS A REGRESIEI MULTIPLE

Correlations

Correlations

		Nota	Coefficient de inteligenta	Timp de invatare	Medie in anul anterior
Nota	Pearson Correlation	1	,582 [*]	,577 [*]	,574 [*]
	Sig. (2-tailed)		,023	,024	,025
	N	15	15	15	15
Coefficient de inteligenta	Pearson Correlation	,582 [*]	1	,704 ^{**}	,363
	Sig. (2-tailed)	,023		,003	,183
	N	15	15	15	15
Timp de invatare	Pearson Correlation	,577 [*]	,704 ^{**}	1	,147
	Sig. (2-tailed)	,024	,003		,602
	N	15	15	15	15
Medie in anul anterior	Pearson Correlation	,574 [*]	,363	,147	1
	Sig. (2-tailed)	,025	,183	,602	
	N	15	15	15	15

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

INTERPRETAREA ÎN SPSS A REGRESIEI MULTIPLE

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,763 ^a	,583	,469	1,563

a. Predictors: (Constant), Medie in anul anterior, Timp de invatare, Coeficient de inteligenta

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-6,960	5,790		-1,202	,255
	Coeficient de inteligenta	,025	,070	,105	,355	,729
	Timp de invatare	,272	,174	,434	1,560	,147
	Medie in anul anterior	1,098	,493	,472	2,228	,048

a. Dependent Variable: Nota

CORRELATIONS

$$\text{Nota} = -6,96 + 0,25 * \text{IQ} + 0,272 * \text{timp} + 0,493 * \text{medie}$$

REGRESIA LOGISTICĂ

CE ESTE REGRESIA LOGISTICĂ?

- Tip de regresie care permite predicția unei variabile discrete printr-un mix de predictor, variabile atât continue cât și discrete
- Răspunde la aceeași întrebare: există o funcție de discriminare dar nu este necesară existența unor condiții pentru predictor: normalitate, relație liniară...
- Extrem de utilizată când nu se găsește o funcție de predicție:
 - Este dificil să prezici o problemă la un pacient dacă TA îi crește cu 10mmHg dacă acesta nu are HTA în schimb la un hipertensiv este un semn al unei probleme

LIMITARE

- Clasificarea poate să fie doar discretă

DEFINIȚII

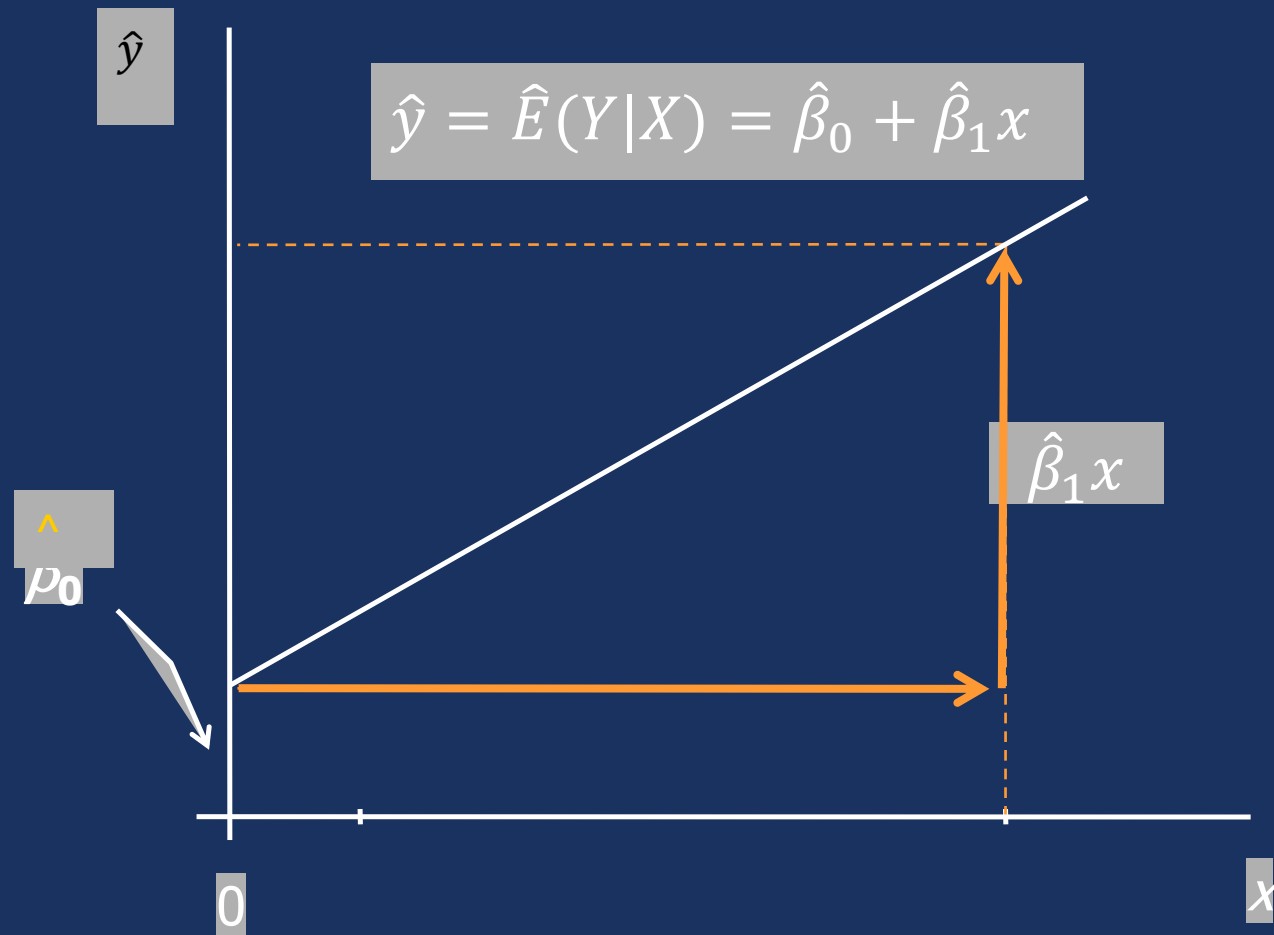
Termeni

- **Probabilitate**
 - PROBABILITY
- **Șansă = $p/(1-p)$**
 - ODDS
- **Raportul șanselor**
 - ODDS RATIO

Exemplu- obținerea fețelor pare la un zar

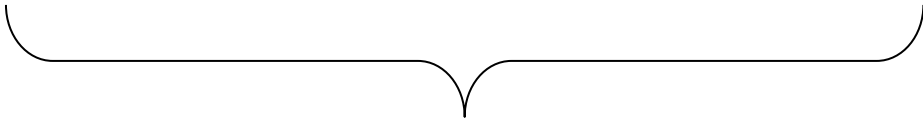
- **ZAR**
 - Probabilitate = $3/6 = 0,50$
 - Șansa = $p/(1-p) = 0,5/0,5=1$

REGRESIA LINIARĂ



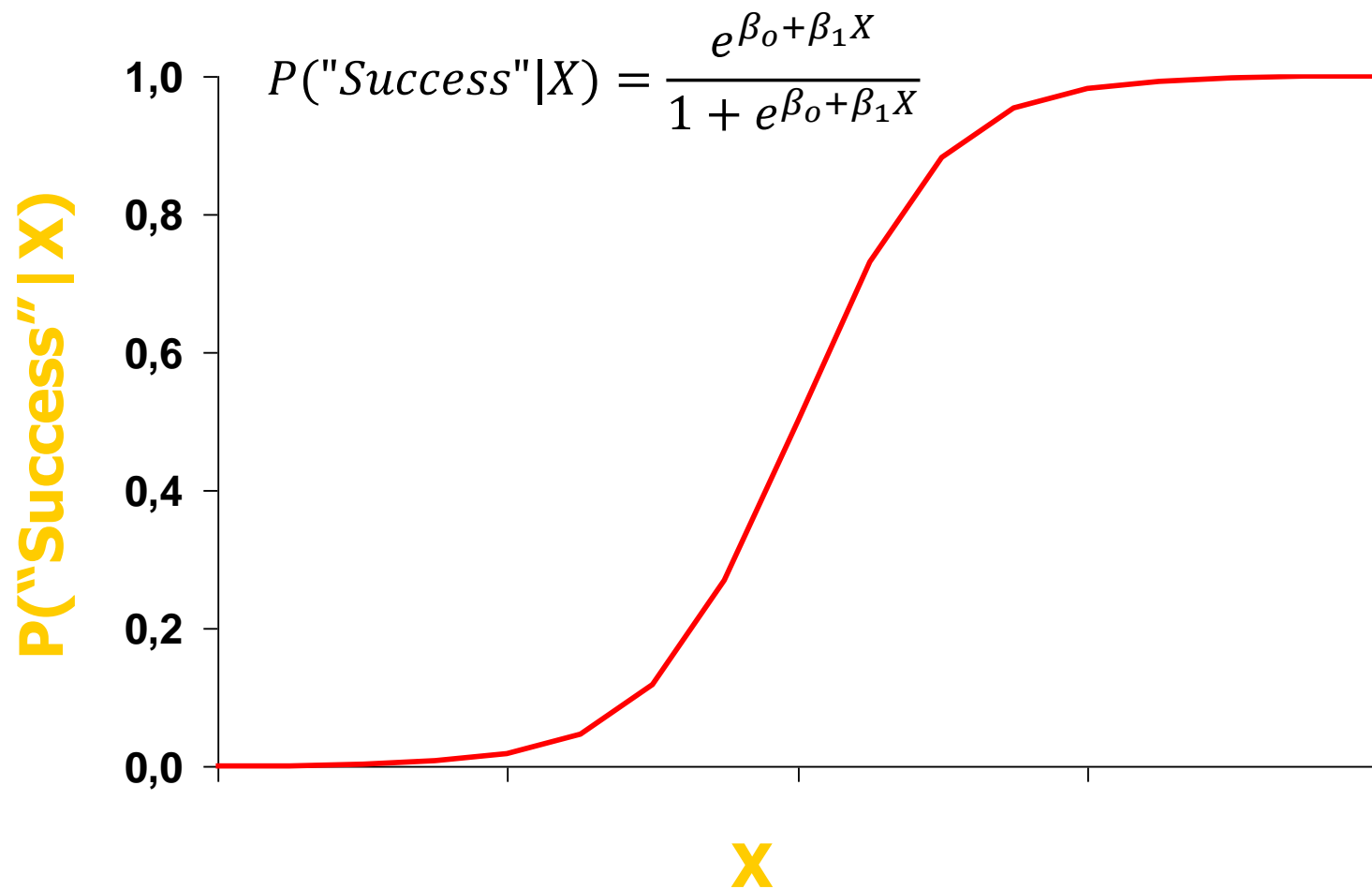
TRANSFORMAREA LOGIT

- **Clasificarea poate să fie doar discretă**

$$\ln(\text{șansă}) = \ln\left(\frac{p}{1-p}\right) = \beta_o + \beta_1 X$$


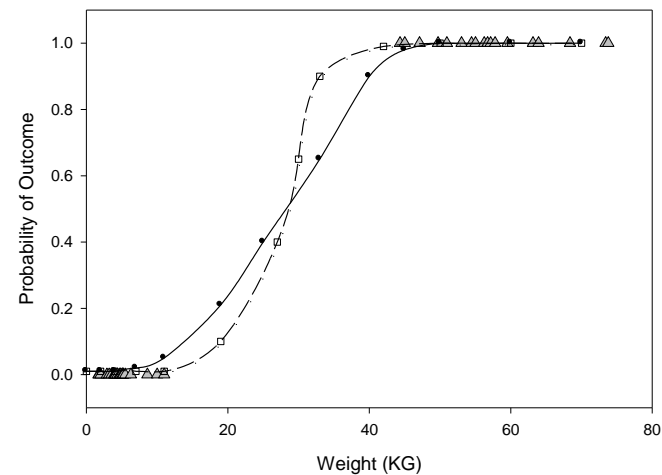
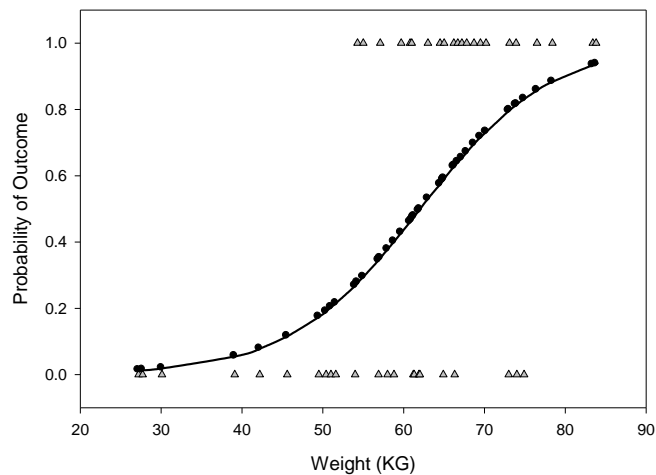
$$p = \frac{e^{\beta_o + \beta_1 X}}{1 + e^{\beta_o + \beta_1 X}}$$

FUNCȚIA LOGISTICĂ



REGRESIA LOGISTICĂ

- Dacă variabila dependentă este de formă binară
 - Estimarea probabilității și a șansei clasificării corecte





APLICAȚII



ORIGINAL ARTICLE



Modelling long-term evolution of chitotriosidase in non-neuronopathic Gaucher disease

Cristina Drugan^a , Tudor C. Drugan^b , Paula Grigorescu-Sido^c and Ioana Naşcu^c

^aDepartment of Medical Biochemistry, “Iuliu Hațieganu” University of Medicine and Pharmacy, Cluj-Napoca, Romania; ^bDepartment of Medical Informatics and Biostatistics, “Iuliu Hațieganu” University of Medicine and Pharmacy, Cluj-Napoca, Romania; ^cDepartment of Paediatrics, Paediatric Clinic I, “Iuliu Hațieganu” University of Medicine and Pharmacy, Cluj-Napoca, Romania

ABSTRACT

Chitotriosidase, an enzyme secreted by activated macrophages, is widely used as a biomarker for therapeutic monitoring and patient follow-up in Gaucher disease (GD), a lysosomal disorder caused by an inherited deficiency of glucocerebrosidase. We analyzed the long-term evolution of chitotriosidase aiming to establish an accurate model that describes the influence of enzyme replacement therapy (ERT) and the impact of several covariates. A total of 55 patients with non-neuronopathic (type 1) GD were followed for almost 17 years (during a maximum of 7.57 and 8.96 years, before and after the onset of ERT, respectively). Plasma chitotriosidase activity, measured yearly before the onset of ERT and at 6-month intervals after the initiation of ERT was analyzed as a function of several covariates (age at

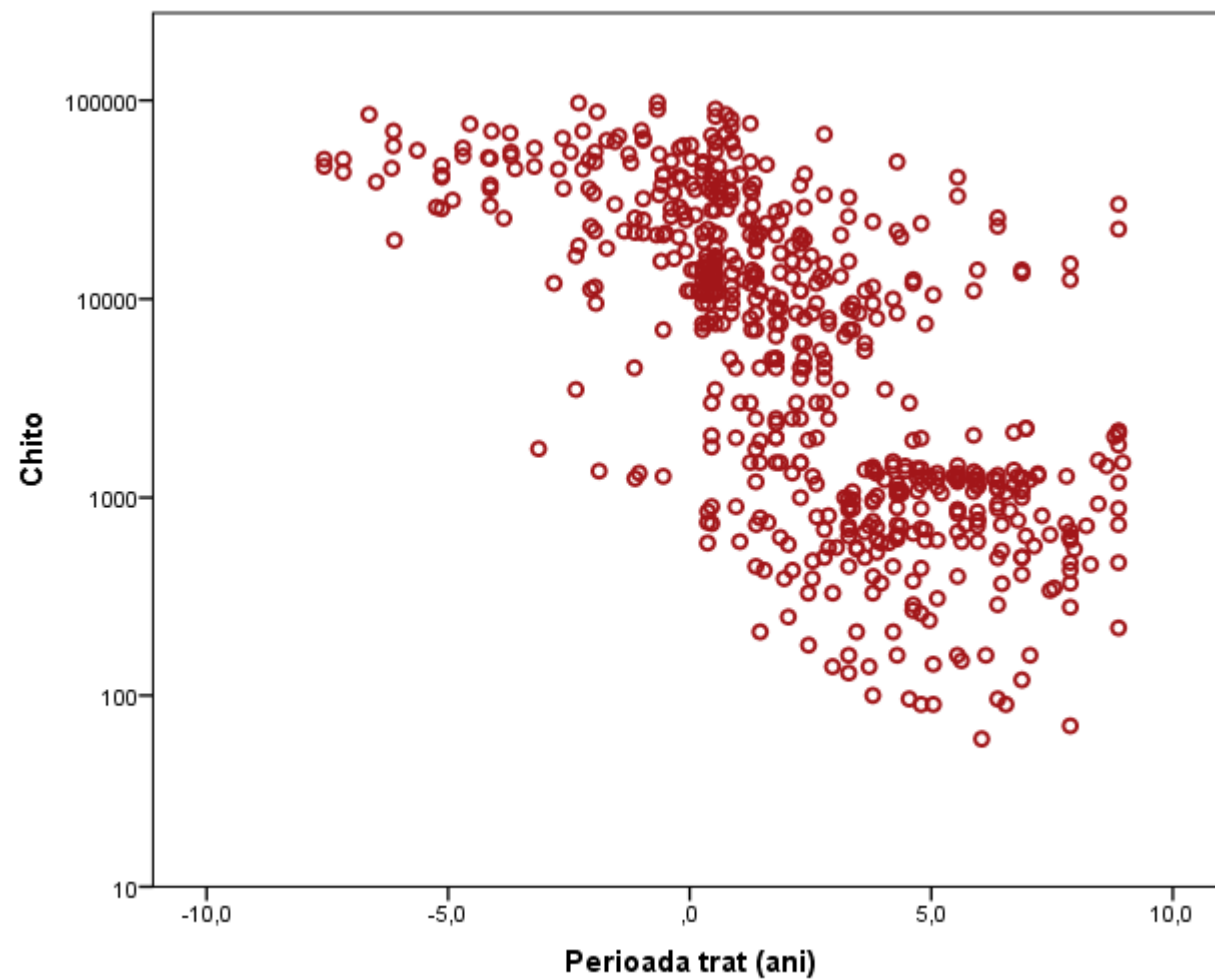
ARTICLE HISTORY

Received 8 September 2016
Revised 26 December 2016
Accepted 2 March 2017

KEYWORDS

Biomarkers; enzyme replacement therapy; lysosomal storage disorders

MODELAREA EFECTELOR TRATAMENTULUI



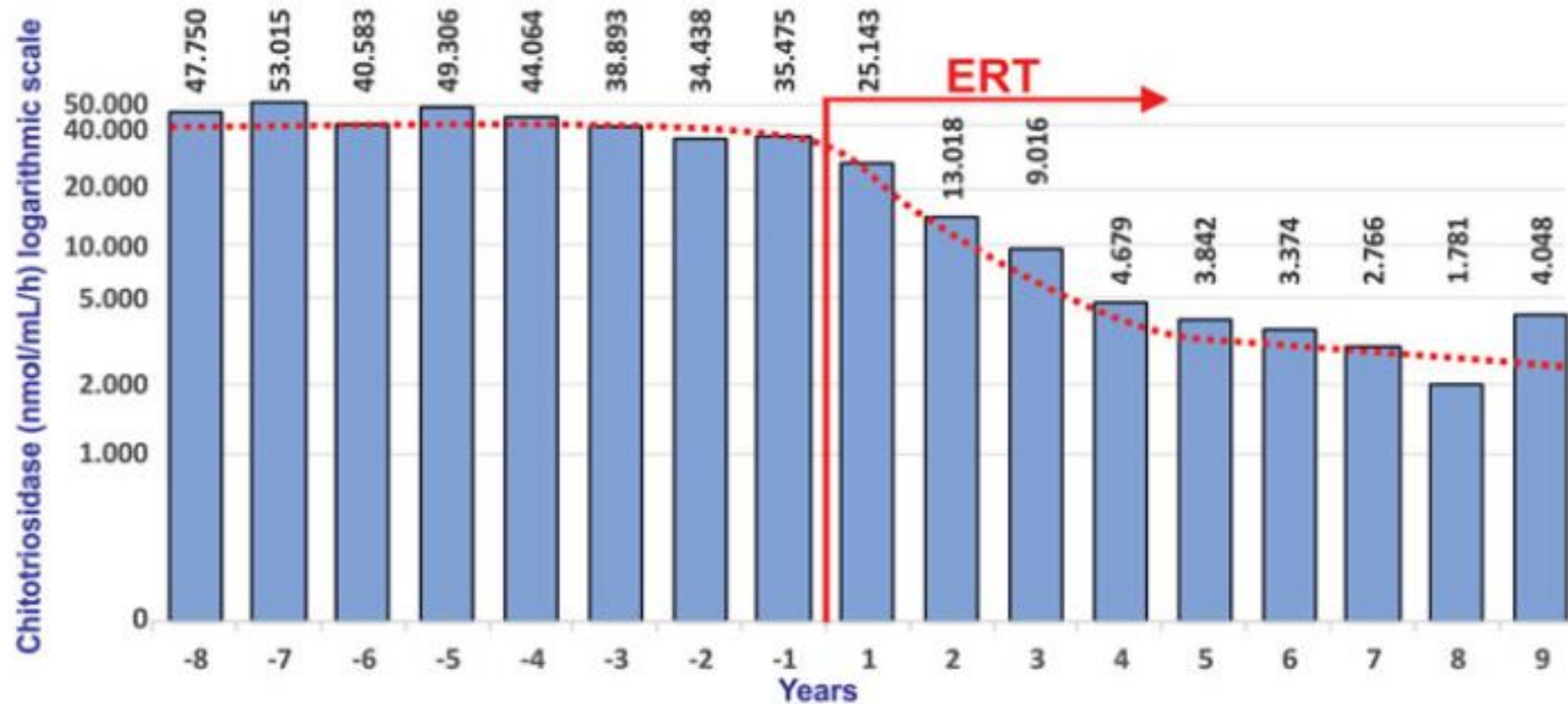
MODELAREA EFECTELOR TRATAMENTULUI

Table 2. (A) Overall evolution of mean chitotriosidase activity in all patients (global evolution) and covariate impact on this evolution. (B) Parameters of the sigmoidal regression that describes the overall evolution of chitotriosidase (ChT) in all patients (global evolution) and covariate impact on these parameters

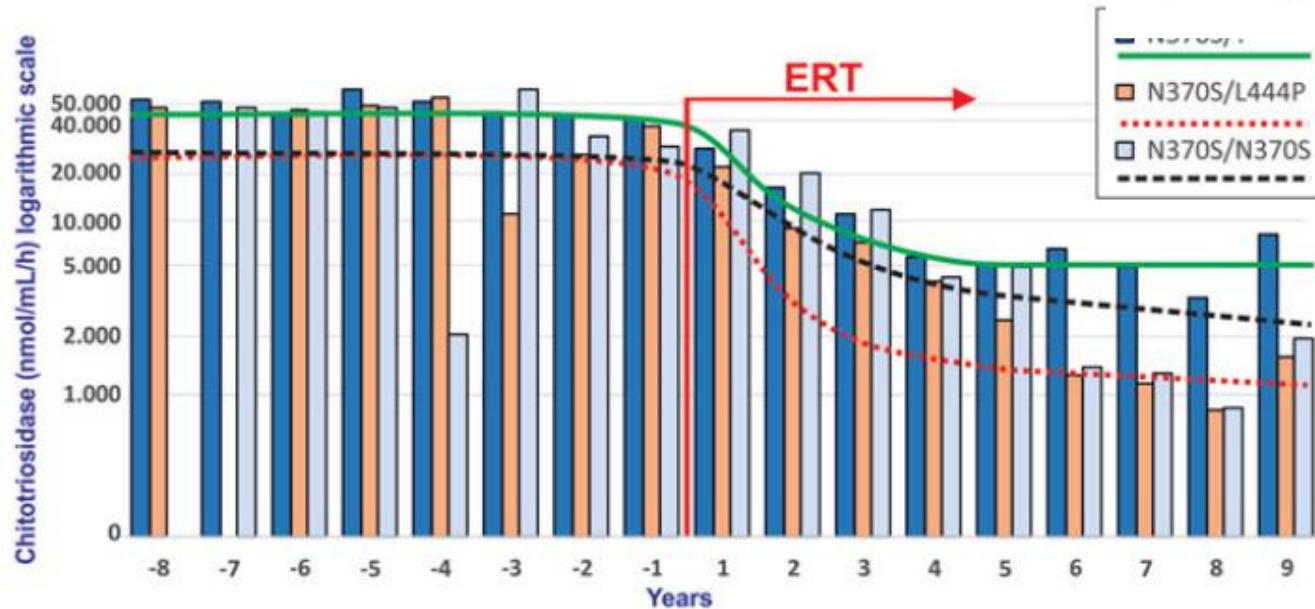
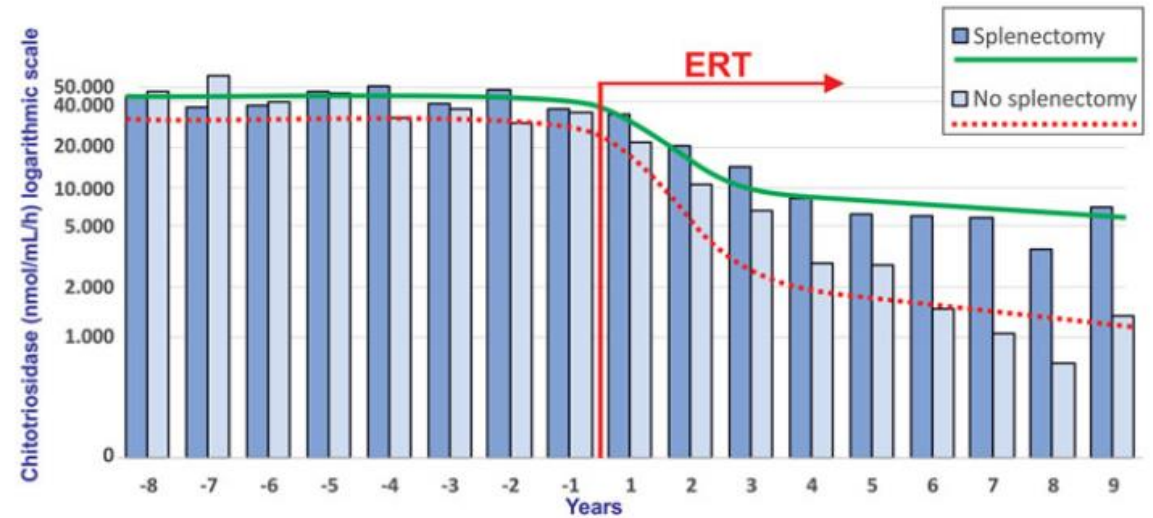
Time intervals	(A) Mean chitotriosidase values (nmol/mL/h)				(B) Sigmoidal regression				
	Before ERT	1st year of ERT	2nd–3rd years of ERT	4–9 years of ERT	Minimal ChT value ^b	Maximal ChT value ^b	Mean response time (years)	Slope	R
Global evolution of chitotriosidase ^c	39442	25143	9224	3248	2289	43489	0.497	−1.165	.668
Impact of covariates on the evolution of chitotriosidase									
Splenectomy									
Yes	44029	34677	13990	5822	5313	40179	1.288	−0.892	.676
No	36548	21224	7106	1539	90	47535	0.091	−1.348	.688
Mann-Whitney test	0.025 ^a	0.043 ^a	0.000 ^a	0.000 ^a					
Bone complication									
Yes	33788	41054	13445	7282	9184	30447	1.268	−0.004	.633
No	39966	20568	7769	1444	577	47418	0.176	−1.226	.732
Mann-Whitney test	0.418	0.002 ^a	0.002 ^a	0.000 ^a					
Genotype									
N370S/?	43603	25113	10264	5171	4582	44769	0.353	−1.007	.675
N370S/L444P	34444	19313	6372	1296	252	43394	0.015	−1.410	.690
N370S/N370S	35678	32367	11502	2345	4847	29341	1.323	−0.781	.662
Kruskal-Wallis test	0.140	0.051	0.127	0.179					
Age at diagnosis									
Over 15	40394	26752	10042	2752	1439	45337	0.652	−1.231	.679
Under 15	33667	20122	6151	4956	5511	35831	0.022	−1.056	.617
Mann-Whitney test	0.283	0.080	0.074	0.621					
Age at treatment onset									
Over 15	40288	25350	9532	2554	1556	44993	0.533	−1.245	.682
Under 15	33385	24091	7460	7491	3591	38418	0.280	−0.959	.574
Mann-Whitney test	0.298	0.367	0.414	0.020 ^a					

$$ChT_{predicted} = ChT_{min} + \frac{ChT_{max} - ChT_{min}}{1 + e^{\left(\frac{mean\ response\ time - time}{slope}\right)}}$$

MODELAREA EFECTELOR TRATAMENTULUI



MODELAREA TRATAMENTULUI





Article

An Improved Score for the Evaluation of Mucosal Healing in Inflammatory Bowel Disease—A Pilot Study

Lidia Neamți ^{1,2}, Tudor Drugan ^{3,*}, Cristina Drugan ¹, Ciprian Silaghi ¹, Lidia Ciobanu ²
and Alexandra Crăciun ¹

¹ Department of Medical Biochemistry, “Iuliu Hațieganu” University of Medicine and Pharmacy, 400012 Cluj-Napoca, Romania

² Gastroenterology Department, Regional Institute of Gastroenterology and Hepatology “O. Fodor”, 400000 Cluj-Napoca, Romania

³ Department of Medical Informatics and Biostatistics, “Iuliu Hațieganu” University of Medicine and Pharmacy, 400012 Cluj-Napoca, Romania

* Correspondence: tdrugan@umfcluj.ro

Abstract: Inflammatory bowel diseases are chronic conditions characterized by periods of remission, alternating with episodes of exacerbation, in which the primary therapeutic target is mucosal healing. Although colonoscopy is currently considered the gold standard for assessing disease activity, it presents a significant number of disadvantages. Over time, various inflammatory biomarkers have been proposed to detect disease activation, but current biomarkers have many limitations. Our study aimed to analyze the most commonly used biomarkers for patient monitoring and follow-up both independently and taken together as a group, in order to propose an improved activity score that more accurately reflects the changes occurring at the intestinal level, in order to limit the number of

AN IMPROVED SCORE FOR THE EVALUATION OF MUCOSAL HEALING IN INFLAMMATORY BOWEL DISEASE

		FC	CRP	ESR 1 h	ESR 2 h	Leukocytes
Crohn's disease	CDAI					
Se	54 (48–55)	78 (71–80)	59 (53–59)	87 (82–92)	94 (91–98)	28 (22–33)
Sp	92 (64–99)	84 (56–97)	100 (74–100)	36 (13–62)	15 (2–31)	61 (34–84)
PPV	97 (86–99)	95 (88–99)	100 (90–100)	87 (82–92)	82 (80–86)	77 (61–90)
NPV	30 (21–33)	45 (30–52)	35 (26–35)	36 (13–62)	40 (7–82)	16 (9–21)
Ulcerative colitis	Mayo score					
Se	81 (78–83)	79 (76–79)	40 (37–40)	82 (80–85)	92 (90–95)	20 (17–21)
Sp	75 (44–93)	100 (71–100)	100 (71–100)	41 (17–69)	41 (17–67)	83 (53–97)
PPV	97 (93–99)	100 (96–100)	100 (93–100)	92 (89–96)	93 (90–96)	92 (79–98)
NPV	28 (16–34)	31 (22–31)	14 (9–14)	20 (8–34)	38 (16–61)	9 (5–10)

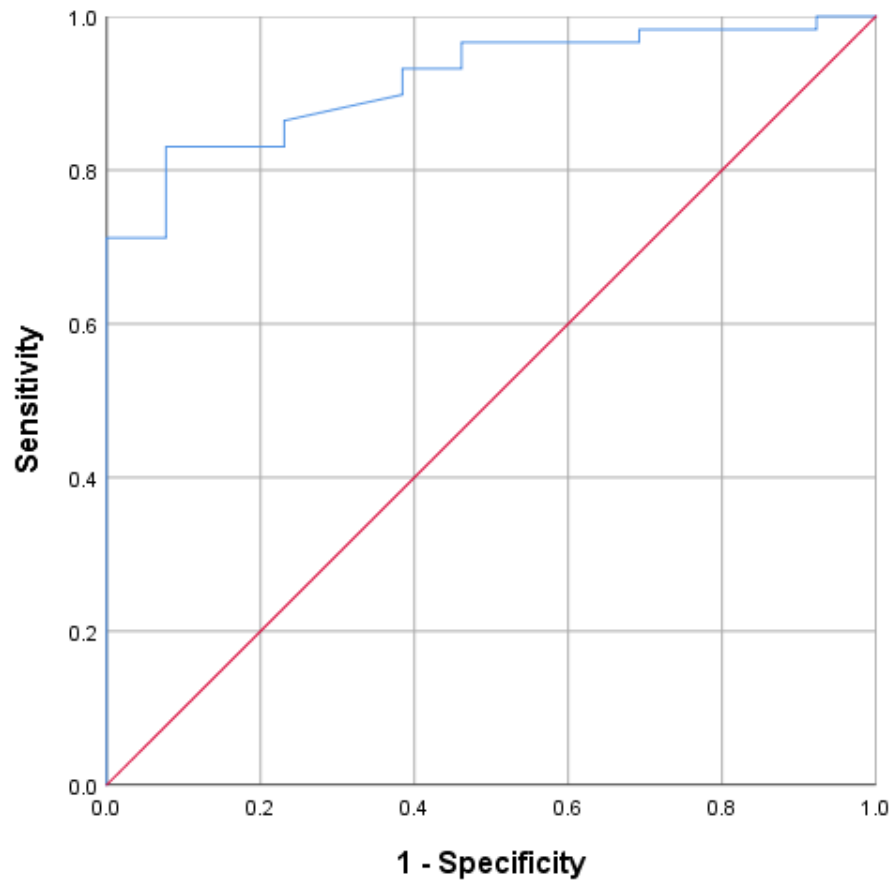
AN IMPROVED SCORE FOR THE EVALUATION OF MUCOSAL HEALING IN INFLAMMATORY BOWEL DISEASE

Crohn's Disease			Ulcerative Colitis		
	Regression Coefficient	Proposed Coefficient		Regression Coefficient	Proposed Coefficient
CDAI	0.001	1/1000	Mayo score	0.264	1/4
CRP (mg/dL)	0.718	1.000	CRP (mg/dL)	0.265	1/4
FC (μg/g)	0.002	1/500	FC (μg/g)	0.009	1/100

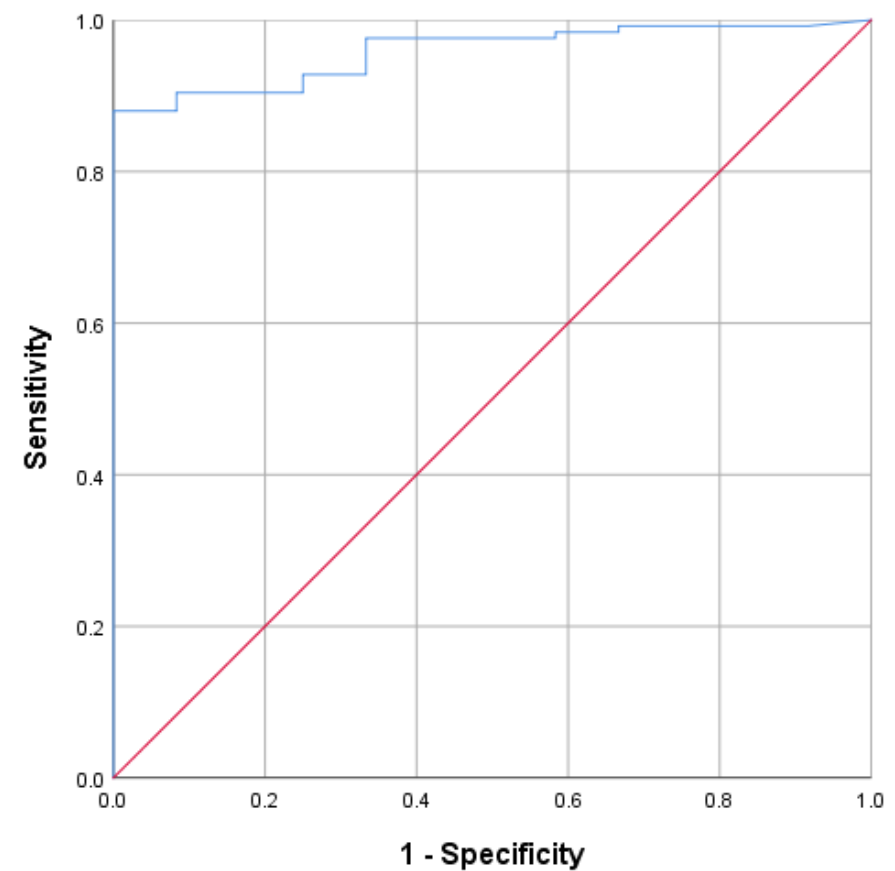
$$\text{CD Score} = \frac{\text{CDAI}}{1000} + \frac{\text{FC}}{500} + \text{CRP}$$

$$\text{UC Score} = \frac{\text{Mayo score}}{4} + \frac{\text{FC}}{100} + \frac{\text{CRP}}{4}$$

AN IMPROVED SCORE FOR THE EVALUATION OF MUCOSAL HEALING IN INFLAMMATORY BOWEL DISEASE



Diagonal segments are produced by ties.



Diagonal segments are produced by ties.

AN IMPROVED SCORE FOR THE EVALUATION OF MUCOSAL HEALING IN INFLAMMATORY BOWEL DISEASE

- The ROC curves corresponding to the activity scores developed for CD and UC are presented in Figure 1. For CD, the validation of the score on the studied patients produced an AUC of 0.917 [0.853–0.983], whereas, for UC, the AUC was 0.958 [0.924–0.992]. Using the Youden coefficient, we set the threshold level at 1 for the CD score and at 2 for the UC score. For CD, this corresponds to a sensitivity of 0.831 and a specificity of 0.923 ($p < 0.001$), while for UC, it implies a sensitivity of 0.904 and a specificity of 0.917 ($p < 0.001$).

PRINTRE RÂNDURI...

Stockflare

If you torture the data long enough
it will confess to anything

RONALD COASE