



UNIVERSITATEA
DE MEDICINĂ ȘI FARMACIE
VICTOR BABEȘ | TIMIȘOARA



Corelatia si regresia liniara

Dr. Mirela Frandes

mirela.frandes@umft.ro

Studii de corelație și regresie liniară

- **Corelația liniară**

= măsură a mărimii și a sensului relației liniare dintre 2 variabile numerice

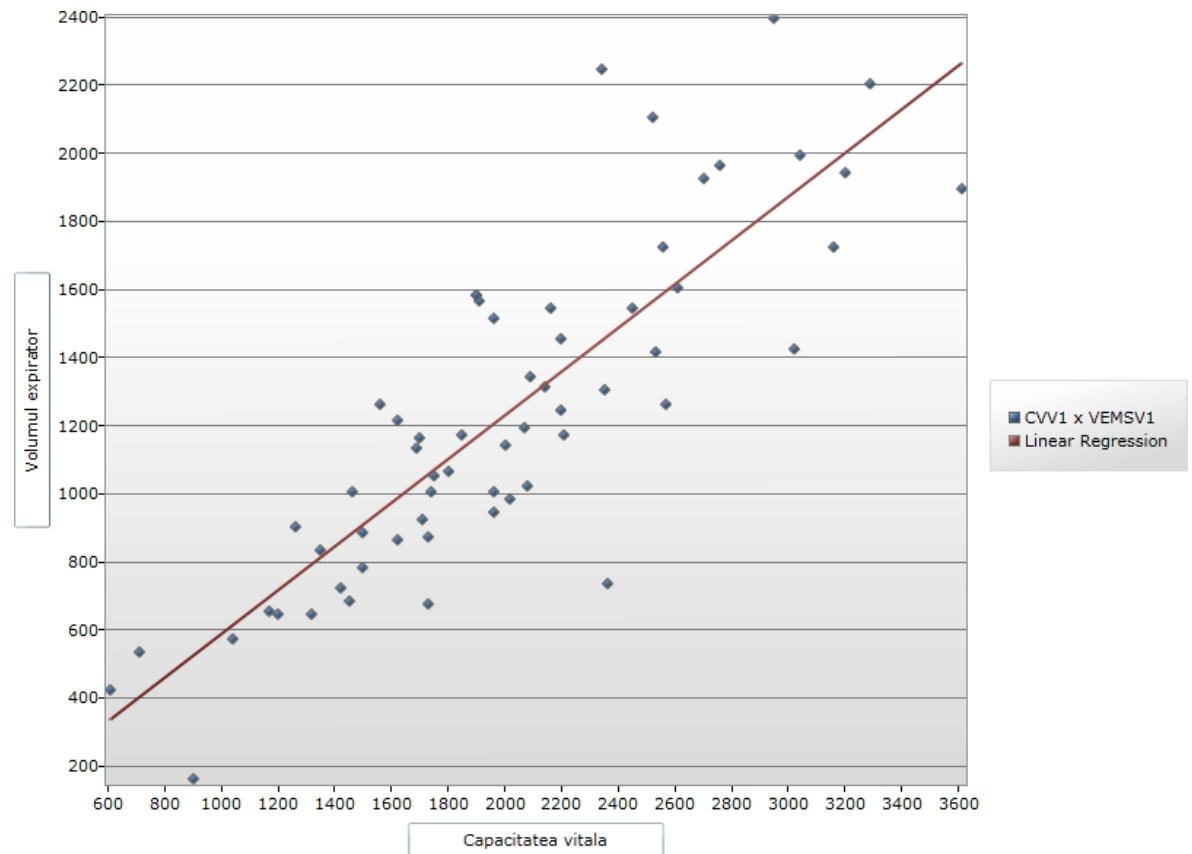
❖ Nu implică o relație cauzală!

Condiții de aplicare:

- Variabilele să fie independente
- Variabilele să fie cantitative
- Variabilele să fie măsurate pe aceiași indivizi
- Cele 2 variabile să fie distribuite normal
- Relația între cele două variabile este liniară simplă (nu exponențială, cvadratică...)

Analiza corelatiei si regresiei liniare

- Asociere dintre
 - **Capacitatea vitala a pacienților cu BPOC**
CVV1
 - **volumul expirator maxim pe sec. la inceputul studiului, inainte de administrarea celor doua medicamente**
VEMSV1



Analiza corelatiei si regresiei liniare

- Coeficientul de corelatie Pearson **r**

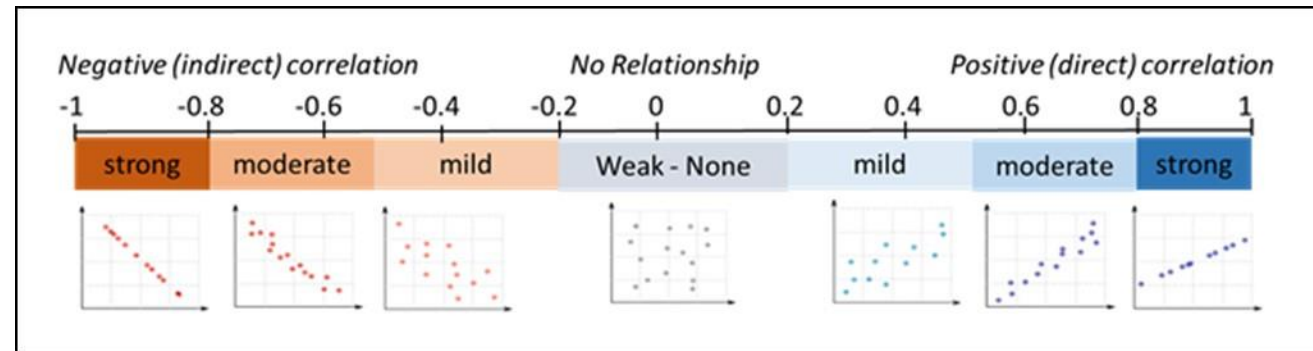
$$r \in [-1, 1]$$

- $r > 0 \Rightarrow$ corelație directă (pozitivă)
- $r < 0 \Rightarrow$ corelație indirectă (inversa, negativă)
- $r = 1 \Rightarrow$ corelație directă perfectă
- $r = -1 \Rightarrow$ corelație inversă perfectă
- $r = 0 \Rightarrow$ variabile necorelate

- Puterea relatiei

➤ Coef. de determinare

$$d = r^2$$



Analiza corelatiei si regresiei liniare

- Ipoteze statistice

H_0 : nu exista o corelatie (liniara) semnif. intre cele 2 variabile

H_a : exista o corelatie (liniara) semnif. intre cele 2 variabile

Analiza corelatiei si regresiei

Etape 1. $H_0 : \rho=0$
 $H_a : \rho \neq 0$

Etape 2 . **Statistica testului:** $t = \frac{r}{\sqrt{1-r^2}} \sqrt{n-2}$

distributia “ t ” cu “ $n-2$ ” grade de libertate

Etape 3. $\alpha = 0.05$

Etape 4. P-value $RR \in (-\infty, -t_{n-2}; t_{n-2}, \infty)$

Etape 5. Comparare P-value si α . Calcul t_c

P-value < α rejectam H_0 , acceptam H_a

Etape 6. Daca $t_c \in RR$, rejectam H_0 , acceptam H_a

Regresia liniara ca instrument de predictie

- **Regresia liniara** – masura a predictiei unei variabile in functie de valorile unei alte variabile
- **Regresia liniara *bivariata*** – relatia dintre cele 2 variabile este data de ecuatia unei drepte

$$y = ax + b$$

- Predictia lui "y" (variab. dependenta) se face in functie de valorile variabilei "x" (variab. independenta) si de coef. dreptei de regresie

a – panta dreptei

b – constanta (*intercept*)

Regresia liniara

Interpretare

- Dreapta de regresie $Y(X)$:
 - a_0 = este ordonata in origine
 - a_1 = panta dreptei de regresie.

Interpretarea pantei a_1

$> 0 \Rightarrow$ tendinta crescatoare – relatie direct proportionala

$< 0 \Rightarrow$ tendinta descrescatoare – relatie invers proportionala

$\cong 0 \Rightarrow$ nicio tendinta

- cu cat a_1 este mai mare, cu atat relatia este mai puternica
- cu cat a_1 este mai aproape 0, cu atat relatia este mai slaba

Relation entre coefficient de corrélation et droite de régression

Varianta totala de la variabila dependenta = Varianta data de dreapta + Varianta care nu este data de dreapta.

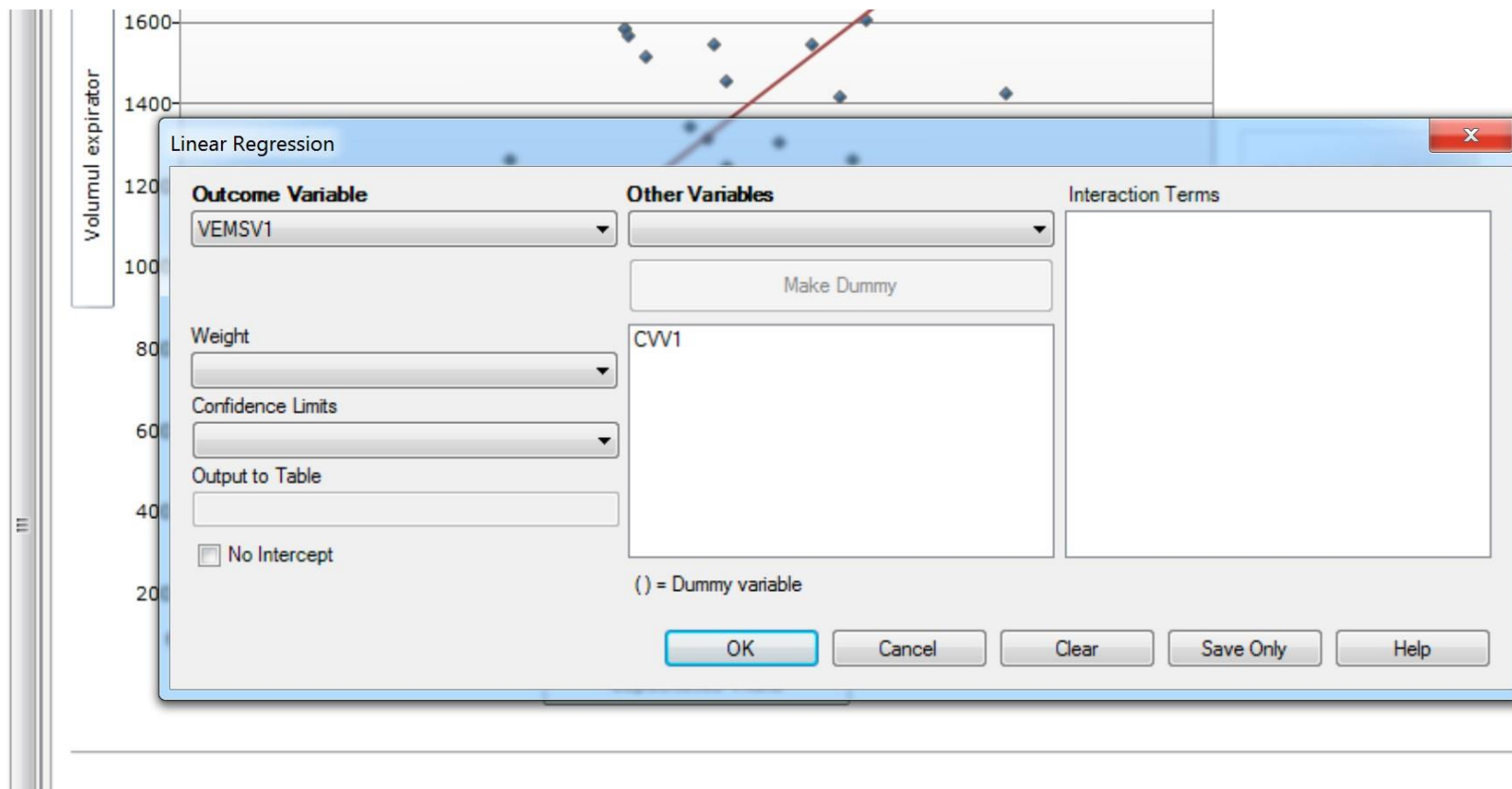
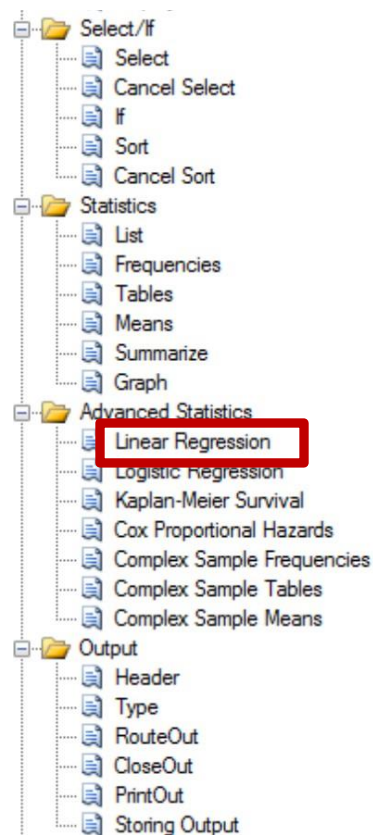
$$r^2 = \frac{\text{Variatia regresiei}}{\text{Variatia totala}} \quad \text{coefficient de détermination}$$

- Notatie: $d = r^2$ (patratul coeficientul de corelatie Pearson)
- Numar de la 0 la 1. O transformam in procent pentru a-l interpreta ($0 \rightarrow 100\%$)

Interpretare: procentul de la variatia unei variabile explica dependenta de cealalta variabila

- Cu cat **d** este mai mare, cu atat relatia este mai puternica
- Cu cat **d** este aproape de 0, cu atat relatia este mai slaba

Regresia liniara ca instrument de predictie



Regresia liniara ca instrument de predictie

REGRESS VEMSV1 = CVV1

Linear Regression

Variable	Coefficient	Std Error	F-test	P-Value
CVV1	0.642	0.054	141.1617	0.000000
CONSTANT	-59.025	113.646	0.2698	0.605474

Correlation Coefficient $r^2 = 0.71$

$$r = \sqrt{0.71} = 0.84$$

$$P = 0.000 < 0.001$$



Se respinge H_0 ,
se accepta H_a

Regresia liniara ca instrument de predictie

REGRESS VEMSV1 = CVV1

Linear Regression

Variable	Coefficient	Std Error	F-test	P-Value
CVV1	0.642	0.054	141.1617	0.000000
CONSTANT	-59.025	113.646	0.2698	0.605474

Correlation Coefficient: $r^2 = 0.71$

Source	df	Sum of Squares	Mean Square	F-statistic
Regression	1	10105756.195	10105756.195	141.162
Residuals	58	4152217.138	71589.951	
Total	59	14257973.333		

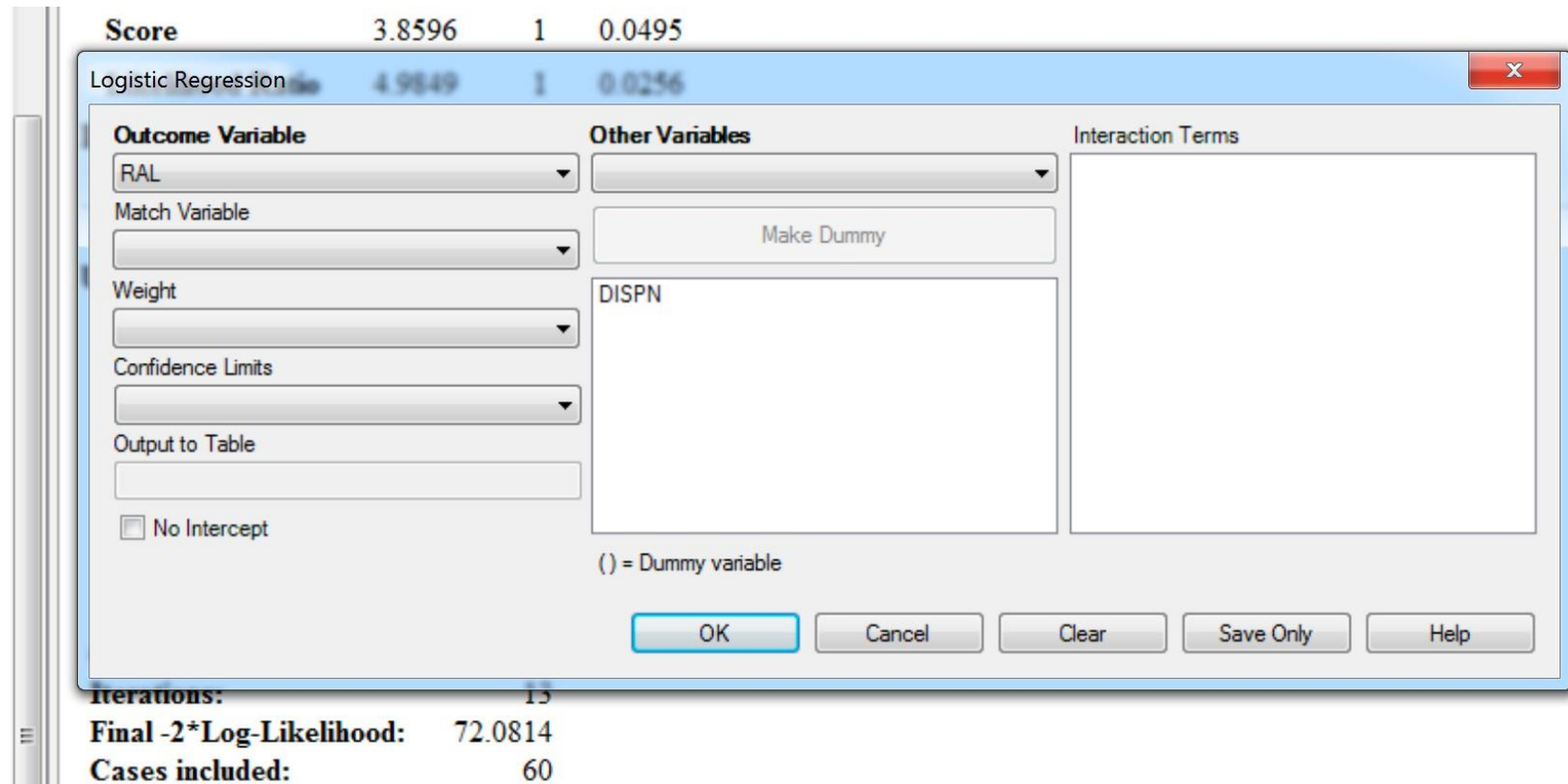
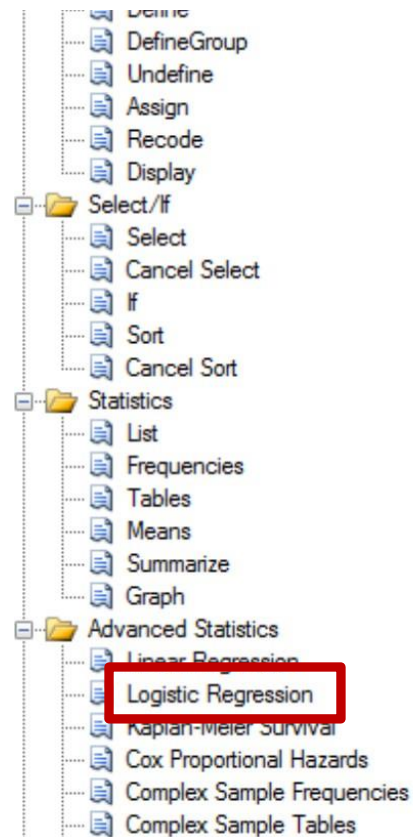
- Predictia **volumului expirator maximal** in functie de (cunoscand) **capacitatea vitala**
- Se verifica daca modelul este semnif. statistic
F-statistic > 1
- Se verifica daca contributia predictorului este semnif. ($P < 0.001$)
- Se identifica **ecuatia de predictie**

$$VEMSV1 = 0.642 CVV1 - 59.025$$

Regresia logistica

- Predictia variabilelor dihotomice (2 categorii)
- Predictorii pot fi:
 - Variabile continue
 - Variabile de tip categorie
 - Variabile dihotomice
- Conditii de aplicare:
 - Lipsa corelatiei puternice intre predictorii
 - Lipsa *outliers*

Regresia logistica



Regresia logistica – Model I

LOGISTIC RAL = FC1

Verificam daca modelul este valid

Unconditional Logistic Regression

Term	Odds Ratio	95%	C.I.	Coefficient	S. E.	Z-Statistic	P-Value
FC1	<u>1.0384</u>	<u>1.0037</u>	<u>1.0744</u>	0.0377	0.0174	2.1711	<u>0.0299</u>
CONSTANT	*	*	*	-2.7522	1.3728	-2.0048	<u>0.0450</u>

Convergence: Converged
Iterations: 4
Final -2*Log-Likelihood: 77.2398
Cases included: 60

Test	Statistic	D.F.	P-Value
Score	5.0828	1	<u>0.0242</u>
Likelihood Ratio	5.3368	1	0.0209

Valid daca P-value < 0.05

P = 0.0242



Frecventa cardiaca (**FC1**) este un predictor
Pentru prezenta ralurilor (**RAL**)

Regresia logistica – Model II

LOGISTIC RAL = FC1 IPB1

Unconditional Logistic Regression

Term	Odds Ratio	95%	C.I.	Coefficient	S. E.	Z-Statistic	P-Value
FC1	<u>1.0424</u>	<u>1.0027</u>	<u>1.0837</u>	0.0415	0.0198	2.0976	<u>0.0359</u>
IPB1	<u>0.9088</u>	<u>0.8540</u>	<u>0.9670</u>	-0.0957	0.0317	-3.0181	<u>0.0025</u>
CONSTANT	*	*	*	2.8320	2.1948	1.2903	0.1969

Convergence: Converged
Iterations: 5
Final -2*Log-Likelihood: 64.5580
Cases included: 60

Test	Statistic	D.F.	P-Value
Score	14.4334	2	<u>0.0007</u>
Likelihood Ratio	18.0187	2	0.0001

Modelul este valid

P= 0.0007 < 0.001



Frecventa cardiaca si indicele de permeabilitate bronsica sunt predictorii pentru prezenta ralurilor

Regresia logistica – Model II

LOGISTIC RAL = FC1 IPB1

Unconditional Logistic Regression

Term	Odds Ratio	95%	C.I.	Coefficient	S. E.	Z-Statistic	P-Value
FC1	<u>1.0424</u>	<u>1.0027</u>	<u>1.0837</u>	0.0415	0.0198	2.0976	<u>0.0359</u>
IPB1	<u>0.9088</u>	<u>0.8540</u>	<u>0.9670</u>	-0.0957	0.0317	-3.0181	<u>0.0025</u>
CONSTANT	*	*	*	2.8320	2.1948	1.2903	0.1969

Convergence: Converged
Iterations: 5
Final -2*Log-Likelihood: 64.5580
Cases included: 60

Test	Statistic	D.F.	P-Value
Score	14.4334	2	0.0007
Likelihood Ratio	18.0187	2	0.0001

Contributia fiecarui predictor este semnificativa

- FC1 P = 0.0356
- IPB1 P = 0.0025
- FC1 OR = 1.0424
- IPB OR = 0.9088

Pacientii cu o FC crescuta au riscul de 1.04 ori mai mare de a prezenta raluri.

Indicele de permeabilitate bronsica este un factor de protectie contra riscului de a prezenta raluri.