

Fișierul BPOC.xls conține date despre 60 de pacienți cu bronhopneumonie obstructivă cronică. Aceste date au fost culese în urma unui studiu de tip cross-sectional, la diferite momente ale studiului. Fișierul conține datele culese la trei momente ale studiului: (1) la început, (2) după prima medicație și (3) după a doua medicație.

***Investigați prezența unei posibile asocieri între frecvența cardiacă (FC3) și capacitatea vitală a pacienților cu BPOC (CVV3), la momentul al III-lea al studiului.***

*In cazul în care se stabilește o asocieră semnificativă, ce capacitate vitală va avea un pacient cu frecvența cardiacă de 70 bpm?; dar pentru frecvența de 45 bpm?*

### **a. Formularea ipotezelor de cercetare**

Dorim să vedem dacă există o corelație semnificativă între frecvența cardiacă și capacitatea vitală, la momentul al III-lea al studiului.

### **b. Formularea ipotezelor statistice**

- ipoteza de nul,  $H_0$ : în populație, cele două variabile nu se corelează  
formula:  $\rho=0$
- ipoteza alternativă,  $H_a$ : în populație, cele două variabile se corelează  
formula:  $\rho \neq 0$

$\rho$  - coeficientul de corelație pentru populația din care e extras esanționul

### **c. Alegerea și aplicarea testului statistic**

#### **Analiza corelatiei**

**condiție:**

- avem 2 coloane de valori numerice continue (cazul problemei de față)

## In Excel:

1. Cu functia **CORREL** se calculeaza coeficientul r de corelatie Pearson

- *Proprietăți:*  $r \in [-1, 1]$

*sensul corelatiei:*

$r > 0 \Rightarrow$  corelație directă (pozitivă)

$r < 0 \Rightarrow$  corelație indirectă (inversa, negativă)

*taria corelatiei:*

$-0.25 < r < 0.25 \Rightarrow$  fara corelație

$-0.5 < r < -0.25$  sau  $0.25 < r < 0.5 \Rightarrow$  corelație slabă

$-0.75 < r < -0.5$  sau  $0.5 < r < 0.75 \Rightarrow$  corelație moderata

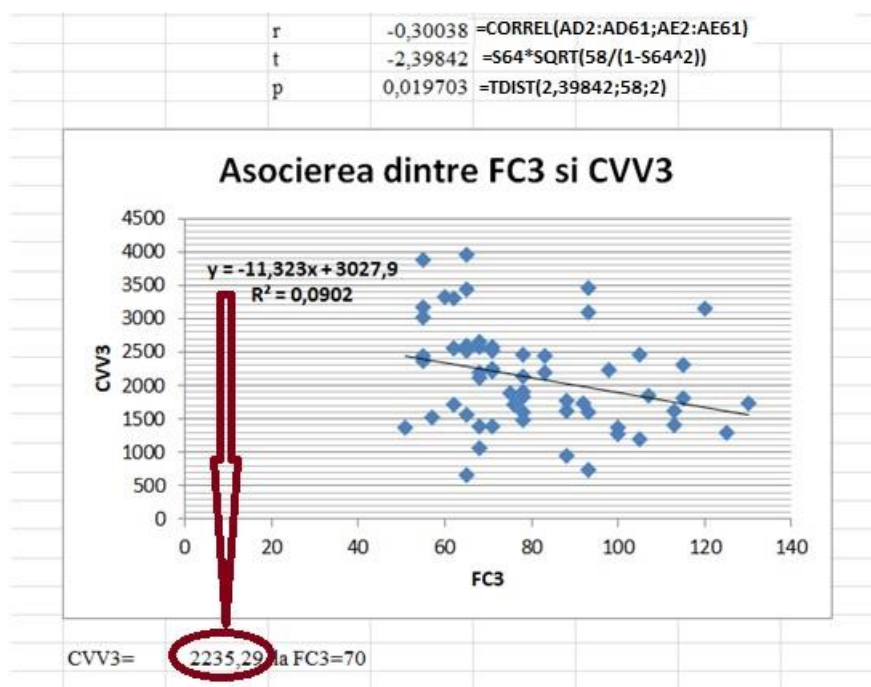
$r < -0.75$  sau  $r > 0.75 \Rightarrow$  corelație puternică

$r = -1$  sau  $r = 1 \Rightarrow$  corelație perfecta

- se calculeaza **statistica t** (premergatoare calculului probabilitatii p)

$$t = r \sqrt{\frac{n-2}{1-r^2}}$$

- se calculeaza **valoarea p**, cu functia **TDIST** (x va lua valoarea absoluta t, **n-2** = numarul de grade de libertate, **tails=2**).



2. **Data Analysis → Regression** (se selecteaza la Y coloana cu valorile dependente si la X coloana cu valorile independente).

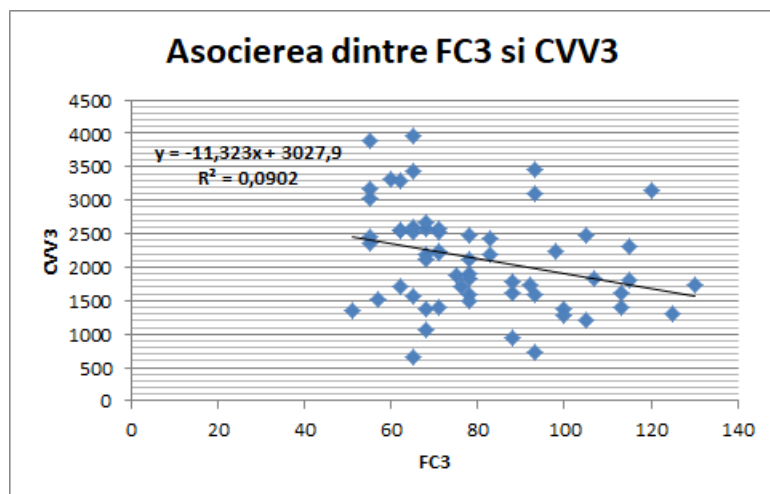
SUMMARY OUTPUT									
Regression Statistics									
Multiple R	0,300384								
R Square	0,090231								
Adjusted R	0,074545								
Standard Error	727,6983								
Observations	60								
ANOVA									
	df	SS	MS	F	Significance F				
Regression	1	3046169	3046169	5,752429	0,019703				
Residual	58	30713596	529544,8						
Total	59	33759765							
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95,0%	Upper 95,0%	
Intercept	3027,874	389,4064	7,775615	1,46E-10	2248,392	3807,355	2248,392	3807,355	
FC3	-11,3226	4,720852	-2,39842	0,019703	-20,7724	-1,87278	-20,7724	-1,87278	

$$y = -11.3236 x + 3027,874$$

$$CVV3 = -11.3236 FC3 + 3027,874 = 2235.292$$

## 3. Reprezentarea grafica:

Dependenta dintre cele 2 variabile se vizualizeaza cu ajutorul graficului **Scatter** (se selecteaza intai coloana cu valorile independente - care vor fi afisate pe axa OX si apoi coloana cu valorile dependente - care vor fi afisate pe axa Oy).



In EpiInfo:

Advanced Statistics → Linear Regression

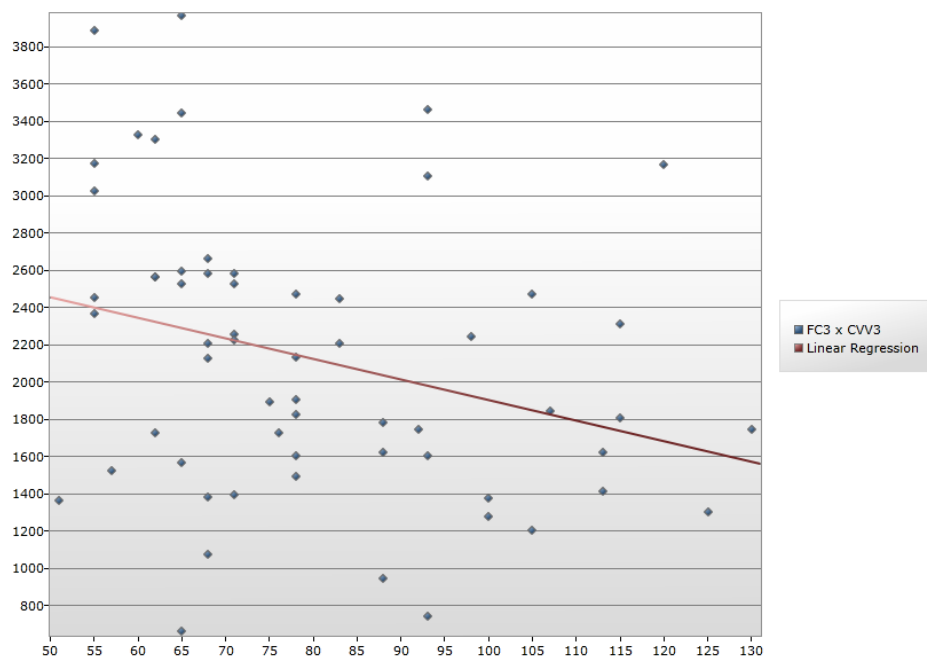
Rezultate:

## Linear Regression

Variable	Coefficient	Std Error	F-test	P-Value
FC3	-11,323	4,721	5,7524	0,019703
CONSTANT	3027,874	389,406	60,4602	0,000000

Correlation Coefficient,  $r^2 = 0,09$

Source	df	Sum of Squares	Mean Square	F-statistic
Regression	1	3046168,895	3046168,895	5,752
Residuals	58	30713596,105	529544,760	
Total	59	33759765,000		



În urma aplicării oricărui test statistic vom obține valoarea probabilității ca asocierile să fie întâmplătoare ( $p$ ), după care luăm decizia statistică astfel:

$p \geq 0,05 \rightarrow H_0$  se accepta, diferentele sunt neseemnificative

$p < 0,05 \rightarrow H_0$  se respinge, diferentele sunt semnificative

$p < 0,01 \rightarrow H_0$  se respinge, diferentele sunt foarte semnificative

$p < 0,001 \rightarrow H_0$  se respinge, diferentele sunt extrem de semnificative

### **d. Decizia statistică**

$p = 0.0197 < 0.05 \rightarrow H_0$  se respinge; exista corelatie semnificativa

### **e. Interpretarea rezultatelor**

CVV3 se afla intr-o corelatie semnificativa, inversa si slaba cu FC3.