

ANALIZA DATELOR MEDICALE. PROBLEME - Epi Info

În acest material, realizat pentru programul Epi Info sub windows, am încercat să ne familiarizăm cu secvența analizei datelor, prin prelucrări ale variabilelor calitative și cantitative, selecții, analize statistice și cu secvența de analiză epidemiologică.

1. Introducere

Epi Info versiunea 7.2.1.0 a fost produs de Centers for Disease Control and Prevention, CDC, Atlanta, Georgia în anul 2004.

Această versiune sau versiunea mai avansată poate fi descărcată de pe site-ul <https://www.cdc.gov/epiinfo/support/downloads.html>



Figura 1. Epi Info7

Epi Info cuprinde o serie de programe folosite de personalul din sănătate publică pentru investigațiile epidemiologice, folosirea bazelor de date pentru observarea în sănătatea publică și alte domenii medicale, precum și aplicații statistice.

Meniul oferă acces celorlalte programe Epi Info, dar are și un scop general, un meniu configurabil folosit în Win95, 98, NT, 2000, Vista, XP pentru a oferi meniuri și butoane de comandă care să conducă la alte programe, funcții, imagini.

Cu Epi Info și un calculator personal, medicii, medicii de sănătate publică, epidemiologii, asistentele precum și alt personal medical, pot elabora rapid un chestionar, pot introduce și analiza datele.

2. Analiza datelor

În prima parte din acest material se vor analiza statistic pacienții dintr-un fișier de date, *BPOC.xlsx* (*Anexa*), care conține date despre 60 de pacienți cu bronhopneumonie obstructivă cronică. Câteva dintre variabilele din fișierul de date sunt: NUME (numele pacientului, sub forma de inițială), SEX, AGE (vârsta), DISPN (dispnee = respirație dificilă), RAL (raluri = respirație cu zgomot ca un “huruit”), TMAX1 (presiunea arterială sistolică, la începutul studiului), FC1 (frecvența cardiacă, la începutul studiului), CVV1 (capacitatea vitală, la începutul studiului).

Pentru a analiza datele, este necesară încărcarea bazei de date în programul *ANALYZE DATA*. Fereastra programului *ANALYZE DATA* constă dintr-o porțiune care cuprinde toate comenzile disponibile (afișate în partea stânga), o porțiune centrală în care sunt afișate rezultatele și o porțiune în partea inferioară a ecranului, în care pot fi introduse comenzile ce vor fi executate ulterior.

Structura succintă a secvenței de analiza datelor (părțile componente care se utilizează mai frecvent): citirea și importul fișierelor de date, variabile, selecție variabile, comenzi statistice, statistică avansată, listări, etc.

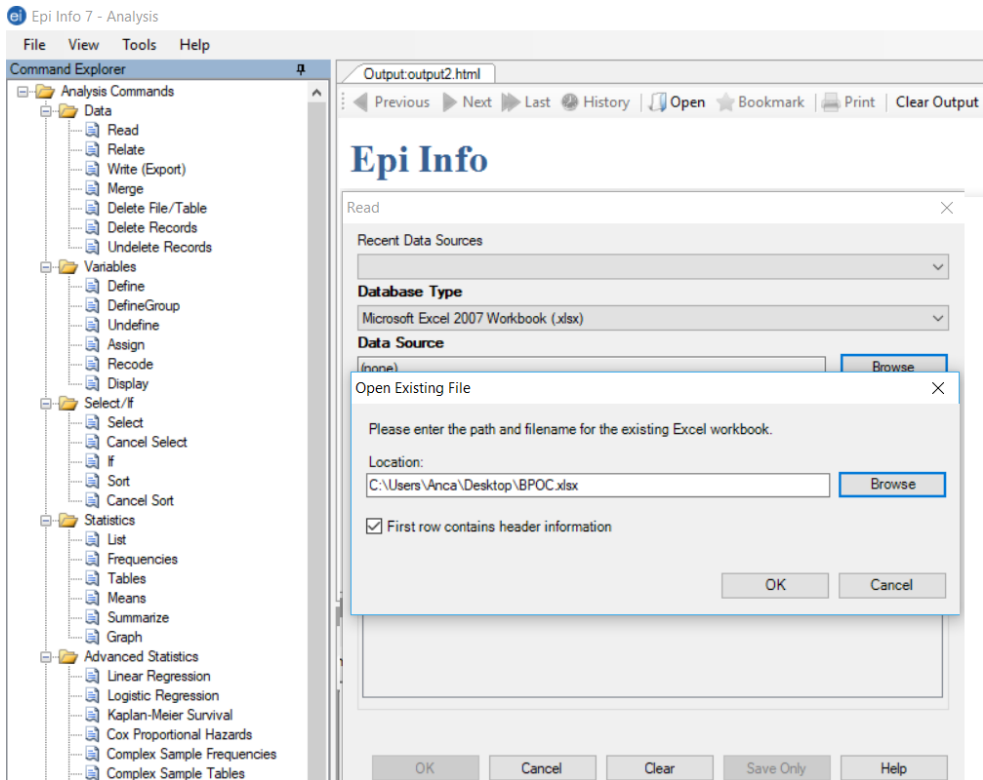


Figura 2. Secvența de analiză a datelor

Încărcați baza de date încărcată pe platformă, *BPOC.xlsx*, folosind comanda: *READ (IMPORT)*. Vizualizați setul de date (*List*). Selectați sexul masculin. (*select SEX = "M"*, utilizând doar operatorii din secvența *Select*). Calculați media de vârstă (min, max, deviația standard, mediana, moda) a lotului de pacienți (*Means AGE*). Care este frecvența pe sexe (*Frequencies SEX*)? Calculați media de vârstă pe sexe, apoi realizați graficul pentru variabila sex, vârstă (*Graph*). Succint se vor prezenta câteva comenzi din statistica mai avansată, în problemele 1 și 4.

3. Principii teoretice prezentate succint

- **Analiza corelației** (coeficientul de corelație liniară Pearson și dreapta de regresie)

Corelația reprezintă legătura dintre 2 variabile cantitative și este apreciată printr-un parametru numit coeficient de corelație Pearson.

Proprietăți ale coeficientului de corelație Pearson, r

- $r \in [-1, +1]$ și dacă $r > 0$ - corelație directă, $r < 0$ - corelație inversă
- intensitatea corelației:
 - slabă = aproape de 0
 - puternică = aproape de ± 1

Legătura dintre cele două mărimi este reprezentată și grafic prin puncte care dacă se situează pe o dreaptă, atunci corelația este liniară, iar dreapta este numită dreaptă de regresie.

- **Parametrii epidemiologici RR și OR**, obținuți cu ajutorul tabelelor de contingență 2x2)

În epidemiologia clinică se urmărește atât determinarea frecvenței de apariție a unei boli, cât și definirea unei asocieri între boală și factori cauzali sau favorizanți.

Factorii de risc pentru o anumită boală se studiază efectuând analiza riscului, sub forma unui tabel de contingență 2x2 (două linii, 2 coloane – expunerea la factorul de risc + / - și boală + / -).

Interpretare: Dacă RR, OR > 1 există risc la boală.

- **Analiza supraviețuirii – curba de supraviețuire Kaplan – Meier:** evoluția în timp a numărului de "supraviețuitori" dintr-un lot inițial.

4. Probleme aplicate teoriei amintite (Regress, Statcalc, Kaplan-Meier)

Problema 1 se bazează pe date reale dintr-un fișier cu pacienți cu bronhopneumonie obstructivă cronică. Problemele 2 și 3 sunt din: Knapp RG, Miller MC III, "Clinical Epidemiology and Biostatistics", 1992 Williams & Wilkins, Baltimore, Maryland, USA (ISBN 0-683-06206-9). Problema 4 folosește datele din fișierul *Sample.prj*, tabelul *Leukem 2*, care este dat spre utilizare împreună cu programul Epi Info 3.2.2.

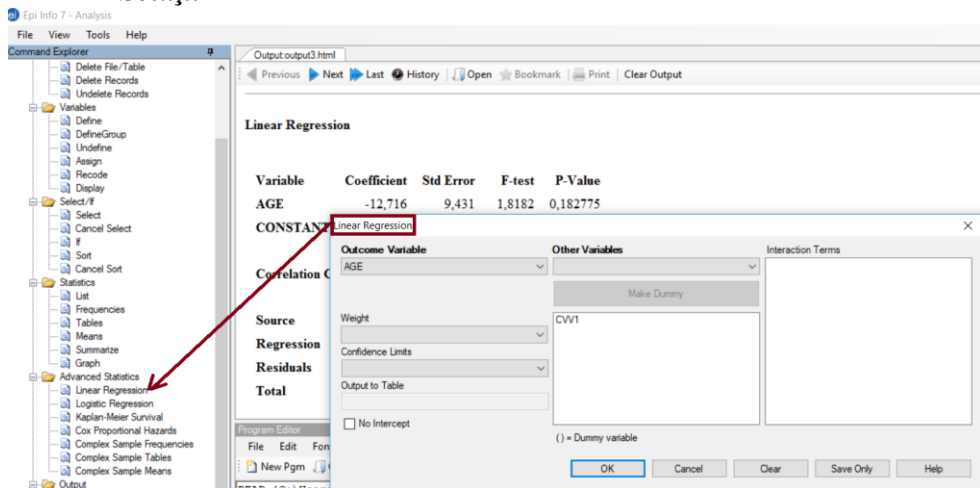
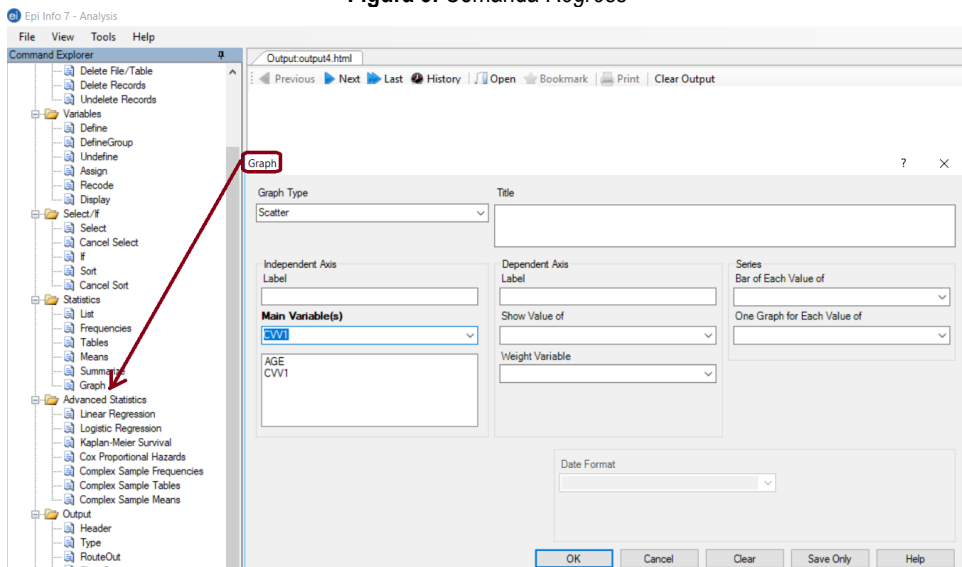


Problema 1

Utilizați ca set de date curent *BPOC.xlsx*.

Întrebări:

- a) Există corelație între *vârsta* pacienților și *CVV1*?
 - observați și interpretați valoarea lui *r*;
 - reprezentați grafic aceasta corelație și dreapta de regresie.
- b) Există corelație între *CVP1* și *VEMSP1*?
 - observați și interpretați valoarea lui *r*;
 - reprezentați grafic aceasta corelație și dreapta de regresie.

Figura 3. Comanda *Regress*Figura 4. Comanda *Scatter* pentru reprezentarea grafică a corelației dintre AGE, CVV 1

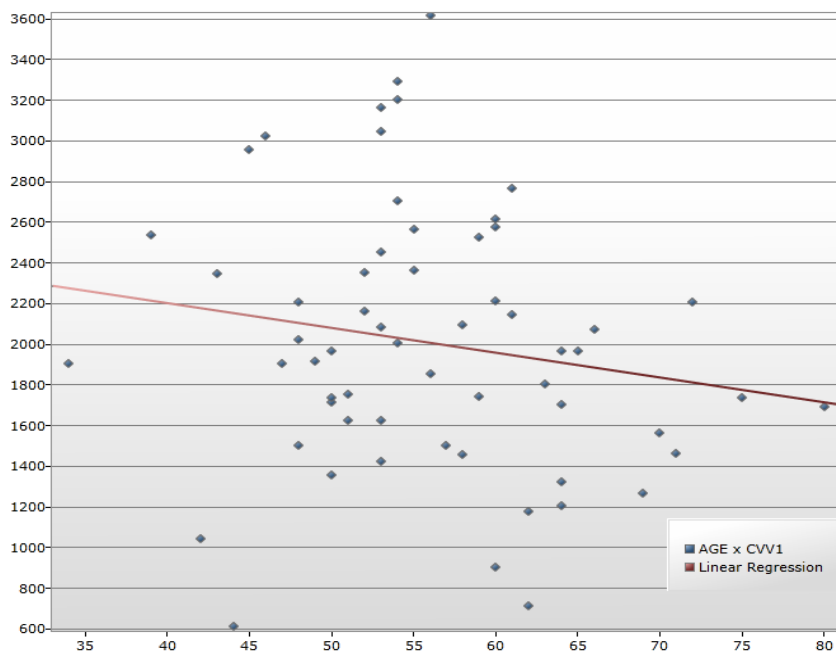


Figura 5. Reprezentarea grafică a corelației dintre AGE, CVV 1

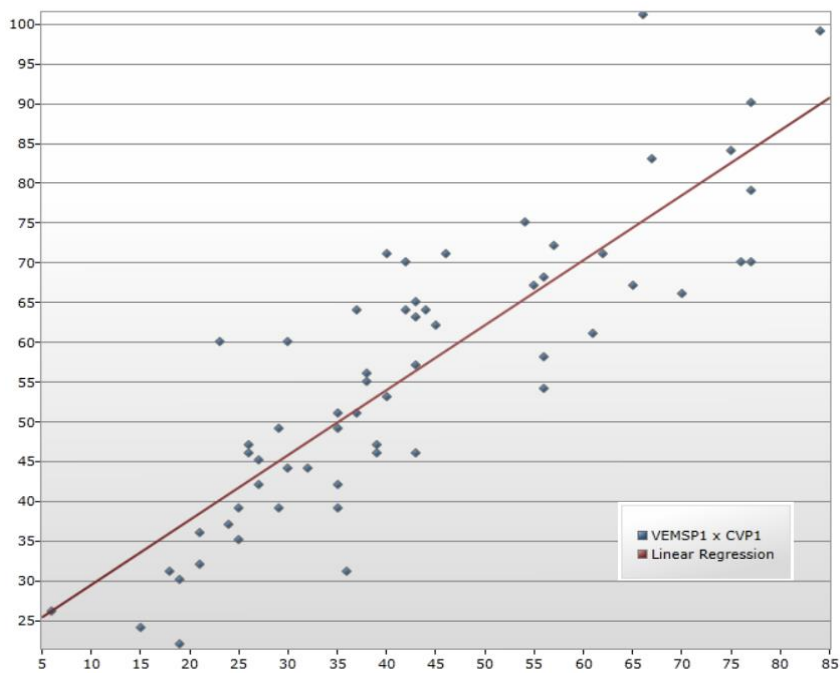


Figura 6. Reprezentarea grafică a corelației dintre VEMSP1, CVP1



Problema 2

Într-un studiu cross-sectional s-a cercetat legătura dintre fumat și anxietate a 1000 de persoane care au fost clasificate simultan, în funcție de fumat (fumător sau nefumător) și de nivelul actual de anxietate (crescut sau scăzut). Așa cum este prezentat în tabelul de mai jos, 300 dintre aceste persoane au fost găsite a avea un nivel crescut de anxietate, 500 au fost identificate ca fumători și 200 ca fumători care au anxietate crescută.

	Nivel de anxietate		Total
	Crescut	Scazut	
Fumători	200	300	500
Nefumători	100	400	500
Total	300	700	1000

- Care este RR al anxietății crescute pentru fumători comparativ cu cei nefumători?
- Care este OR?



Indicații

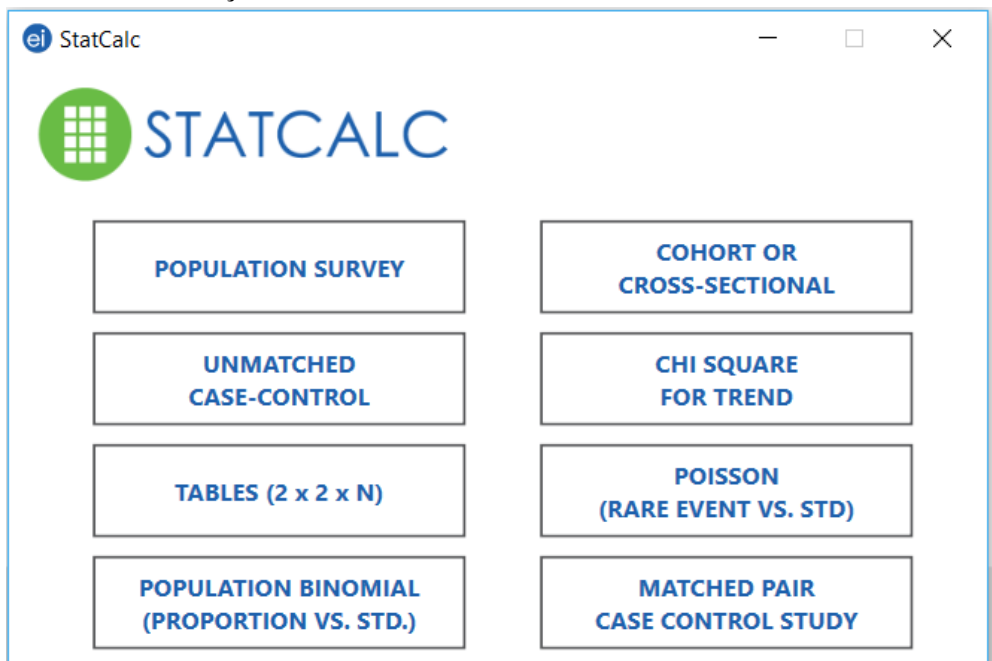


Figura 7. Secvența StatCalc

StatCalc - 2x2 Tables

Strata 1	Strata 2	Strata 3	Strata 4	Strata 5	Strata 6	Strata 7	Strata 8	Strata 9
----------	----------	----------	----------	----------	----------	----------	----------	----------

		Outcome		
		Yes	No	Total
Exposure	Yes	200	300	500
	Row %	40,00 %	60,00 %	100,00 %
	Col %	66,67 %	42,86 %	50,00 %
	No	100	400	500
	Row %	20,00 %	80,00 %	100,00 %
	Col %	33,33 %	57,14 %	50,00 %
Total	300	700	1000	
Row %	30,00 %	70,00 %	100,00 %	
Col %	100,00 %	100,00 %	100,00 %	

Figura 8. Tabelele de contingență (2 x 2, 2 x n)

Odds-based Parameters				Statistical Tests		
	Estimate	Lower	Upper		X ²	2 Tailed P
Odds Ratio	2,6667	2,0096	3,5386	Uncorrected	47,6190	0,00000000
MLE Odds Ratio (Mid-P)	2,6640	2,0103	3,5433	Mantel-Haenszel	47,5714	0,00000000
Fisher-Exact		1,9914	3,5787	Corrected	46,6714	0,00000000

Risk-based Parameters					
	Estimate	Lower	Upper		
Risk Ratio	2,0000	1,6284	2,4564	Mid-P Exact	0,00000000
Risk Difference	20,0000	14,4563	25,5437	Fisher Exact	0,00000000

Figura 9. OR, RR al anxietății pentru fumători comparativ cu cei nefumători

Interpretarea se face conform teoriei.



Problema 3

Pentru a studia posibila asociere între uzul contraceptivelor orale și apariția artritei reumatoide, un cercetător a selectat ca și subiecți 100 de femei cu un diagnostic confirmat de artrită reumatoidă și 200 femei supuse unui tratament în aceeași instituție medicală pentru alte boli musculoscheletale ca și subiecți.

Foile de observație ale pacienților au fost cercetate pentru a se documenta de folosirea contraceptivelor orale înainte de data primei diagnosticări. Rezultatele sunt prezentate în următorul tabel.

Artrita reumatoidă			
Contraceptive orale	Prezentă	Absentă	Total
Utilizare	40	120	160
Fără utilizare	60	80	140
Total	100	200	300

- Care este tipul de studiu epidemiologic utilizat?
- Care este OR la utilizarea contraceptivelor orale în artrită reumatoidă?
- Care este RR?
- Calculați intervalul de încredere 95% pentru OR și RR?
- Există o corelație statistic semnificativă între utilizarea contraceptivelor orale și apariția artritei reumatoide?



Problema 4

Analiza supraviețuirii – Curba de supraviețuire Kaplan – Meier

Folosim datele din fișierul *Sample.prj*, tabelul *Leukem2* (din Epi Info 7.2.1.0), în care avem supraviețuirea în luni a 42 pacienți diagnosticați cu leucemie.

Current Data Source: C:\Users\Anca\Desktop\Sample\Sample.prj:LEUKEM2

Record Count: 42

Line List								
	UniqueKey	RECSTATUS	OldUniqueKey	GlobalRecordId	ID	SURVTIME	CENSORED	TREATMENT
▶	29	1	29	09f53459-cdb3-4...	29	5	0	Placebo
	5	1	5	0dbfed12-ae0c-4...	5	10	0	6-MP
	41	1	41	0df7cc90-c977-4...	41	22	0	Placebo
	4	1	4	10daa887-5d20-...	4	7	0	6-MP
	16	1	16	123d59b1-e2e1-...	16	20	1	6-MP
	2	1	2	157a7e25-62b8-...	2	6	0	6-MP
	10	1	10	17d05bba-b129-...	10	6	1	6-MP
	19	1	19	2324d6e3-a6c4-...	19	32	1	6-MP
	20	1	20	26805dac-7bd0-...	20	34	1	6-MP
	24	1	24	357315e6-2200-...	24	2	0	Placebo

Figura 10. Fișierul de date LEUKEM2

Kaplan-Meier Survival Analysis

Censored Variable	Value for Uncensored
CENSORED	0
Time Variable	Time Unit
SURVTIME	
Test Group Variable	Graph Type
TREATMENT	Survival Probability
Weight	Output to Table

OK Cancel Clear Save Only Help

Figura 11. Comanda pentru Kaplan-Meier

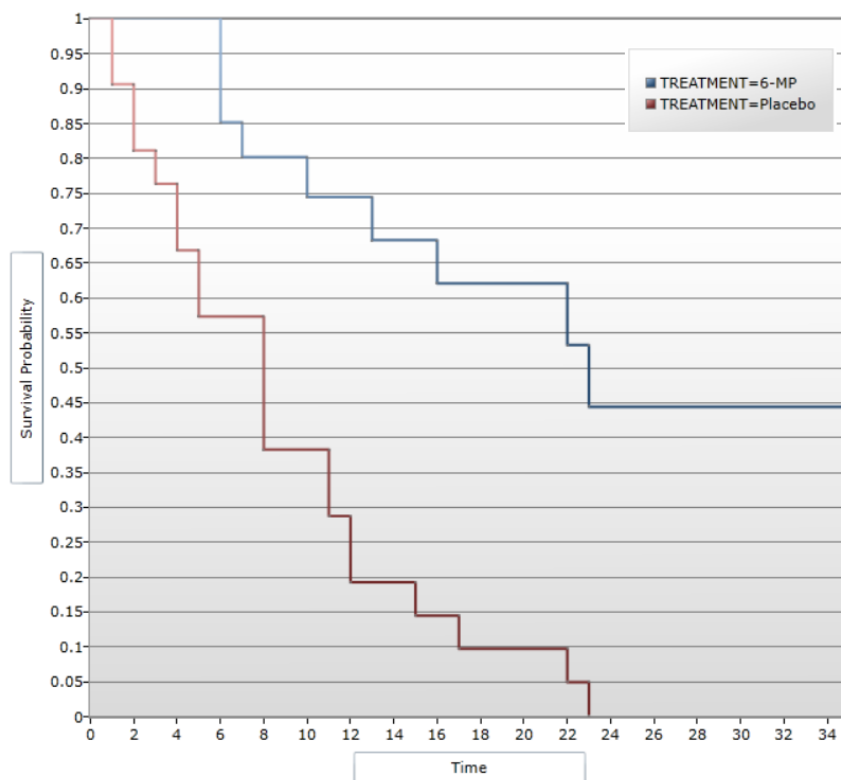


Figura 12. Curba de supraviețuire Kaplan-Meier

Graficul ne permite să analizăm comparativ supraviețuirea după un anumit număr de luni de urmărire pentru cele două tratamente. În acest exemplu, probabilitatea de supraviețuire la 8 luni este de circa 50% (0,5) pentru pacienții

tratați cu placebo, comparativ cu circa 80% (0,8) pentru pacienții tratați cu 6-MP. După 23 de luni de urmărire, probabilitatea de supraviețuire este de 0% pentru placebo, comparativ cu 45% (0,45) pentru 6-MP.

5. Concluzii

În acest material am reușit să ne familiarizăm cu:

- + secvența de analiza datelor, realizând prelucrări descriptive, analize statistice și analize epidemiologice;
- + analize avansate, prin calcularea coeficientului de corelație, realizarea graficului care interpretează o corelație, obținerea curbei de supraviețuire Kaplan-Meier;
- + analize epidemiologice prin calcularea parametrilor epidemiologici OR, RR.

Referințe

- Vernic CV, Apostol SA, Frandes M, Mada L, Lungeanu D. APLICATII PRACTICE DE INFORMATICA SI BIOSTATISTICA MEDICALA IN NURSING. Editura Eurostampa, ISBN 978-606-32-0487-6, Colectia Derzelas, 2017:1-216.
- Vernic CV, Timar B, Mada L, Apostol SA. NOȚIUNI DE INFORMATICĂ MEDICALĂ ȘI ABORDĂRI DE BIOSTATISTICĂ ÎN NURSING. Editura Eurostampa, ISBN 978-606-32-0152-3, Colectia Derzelas, 2015:1-282.
- Vernic CV, Timar B, Mada L, Apostol SA. *Informatică medicală și metode de biostatistică aplicate în nursing*. Editura Eurostampa, 2014.
- Vernic CV, Mada L, Lungeanu D, Muntean C, Apostol SA, Catu CO, Ursoniu S. *Aplicații practice de Informatică Medicală și Biostatistică*. Editura Victor Babes, 2010.
- Knapp RG, Clinton Miller M. *Clinical epidemiology and biostatistics*. Baltimore: Williams & Wilkins, 1992.
<http://www.cdc.gov/Epiinfo/>