

II.G

PRELUCRAREA STATISTICĂ A DATELOR BIOMEDICALE. PROBLEME

1. Introducere

Acest capitol include o serie de aplicații practice a căror rezolvare necesită cunoașterea noțiunilor de bază de prelucrare statistică a datelor biomedicale. Printre aceste noțiuni amintim: testele statistice, analiza corelației și regresiei liniare, analiza riscului, analiza supraviețuirii.

Problemele propuse pot fi rezolvate utilizând aplicațiile EpiInfo (<https://www.cdc.gov/epiinfo/index.html>), Microsoft Excel, OpenEpi (<http://www.openepi.com>) sau utilizând orice altă aplicație de prelucrare statistică a datelor. Pentru probleme propuse, se consideră un prag de semnificație de 0.05.

Obiectivul capitolului a fost de a pune la dispoziție studenților probleme spre rezolvare, pentru a facilita astfel procesul de însușire a noțiunilor de biostatistică.

2. Teste statistice



Problema 2.1

Într-un studiu ce viza controlul calității laboratoarelor de analize medicale, mostre din patru concentrații de enzime au fost date la trei laboratoare. S-au comparat rezultatele de la aceste laboratoare. Datele obținute în urma studiului sunt incluse în tabelul următor:

Tabelul 1. Datele obținute la cele trei laboratoare.

Concentrația de enzimă ($\mu\text{g/ml}$)					
	1	2	3	4	Total
Laborator 1	4.2	6.0	3.9	8.3	22.4
Laborator 2	3.9	7.3	4.0	7.2	22.4
Laborator 3	5.2	6.5	3.2	6.9	21.6

Cerințe:

1. Precizați ce tip de studiu este cel prezentat.
2. Precizați ce concluzie se poate trage cu privire la performanța celor trei laboratoare.



Problema 2.2

În tabelul de mai jos sunt prezentate datele cu privire la pierderea în greutate din cauza transpirației excesive în timpul unei hipoglicemii pentru 12 pacienți tratați cu propranolol și 11 pacienți în grupul de control.

	Propranolol	Control
Dimensiunea grupului	$n_1=12$	$n_2=11$
Media	$m_1 = 120 \text{ g}$	$m_2 = 70 \text{ g}$
Deviația standard	$s_1=10$	$s_2=8$

Cerințe:

1. Precizați ce tip de studiu este cel prezentat. Argumentați.
2. Testați ipoteza ca media cumulativă a greutății pierdute diferă între cele două grupuri.



Problema 2.3

Se considera un studiu în care se urmărește efectul igienei dentare în apariția ulterioară a cariilor dentare la copii. Din populația unei clinici dentare pediatrice, sunt selectați de mod aleatoriu 100 de copii. Dintre aceștia, 50 de copii au primit instrucțiuni cu privire la igiena dentară și 50 nu au fost instruiți

(aceștia reprezentând grupul de control). După 6 luni, numărul de carii dentare nou apărute au fost notate pentru fiecare copil. Rezultatele studiului sunt incluse în tabelul următor:

Tabel 2. Datele pentru grupul cu copii instruiți și grupul de control.

Numărul de carii dentare nou apărute

	0-1	2-3	4-5
Copii instruiți	30	15	5
Copii neinstruiți	20	15	15
Total	50	30	20

Cerințe:

1. Ce tip de studiu este cel prezentat?
2. Studiați dacă instrucția cu privire la igiena dentara este asociată cu numărul de noi carii dentare apărute în timpul celor 6 luni de studiu.



Problema 2.4

În tabelul 3 sunt prezentate datele unui studiu ce examinează dacă pacienții cu oboseala cronică și-au îmbunătățit starea după 6 săptămâni cu tratament cu magneziu în comparație cu grupul placebo.

Tabel 3. Datele pentru grupul cu tratament și grupul de control.

Stare

	îmbunătățită	neîmbunătățită	Total
Magneziu	12	3	15
Placebo	3	14	17
Total	15	17	32

Cerințe:

1. Precizați ce tip de studiu este cel prezentat.
2. Studiați dacă tratamentul cu magneziu a îmbunătățit starea celor cu oboseala cronică.

3. Analiza corelației și regresiei liniare



Problema 3.1

Se consideră datele hematologice de la pacienți cu anemie aplastică (sindrom al insuficienței medulare). Tabelul 4¹ prezintă procentele de reticulocite (indicator al producerii celulelor roșii în măduva spinării) și numărul de limfocite (pe mm²) de la 9 pacienți. Se dorește studierea asocierii dintre datele hematologice.

Tabel 4. Datele hematologice de la pacienți cu anemie aplastică.

Pacient	Reticulocite (%)	Limfocite
1	3.6	1700
2	2.0	3078
3	0.3	1820
4	0.3	2706
5	0.2	2086
6	3.0	2299
7	0.0	676
8	1.0	2088
9	2.2	2013

Cerințe:

1. Indicați numele și tipul variabilelor implicate în analiza de corelație și regresie.
2. Formulați ipotezele de cercetare și statistice.
3. Reprezentați printr-un grafic sugestiv datele hematologice din Tabelul 4.
4. Prelucrați statistic datele (determinați coeficienții de corelație și de determinare, ecuația dreptei de regresie, testați semnificația statistică aferentă asocierii găsite).
5. Interpretați statistic și medical rezultatele obținute.

¹ The New England Journal of Medicine, 312(16), 1015–1022, 1985.



Problema 3.2

O companie farmaceutica a desfășurat un studiu pilot pentru a evalua relația dintre trei doze a unui nou agent hipnotic și timpul de somn. Datele studiului sunt prezentate în tabelul următor:

Tabel 5. Timpul de somn și doza administrată.

Pacient	Timpul de somn (ore; Y)	Doza (mM/kg; X)
1	4	3
2	6	3
3	5	3
4	9	10
5	8	10
6	7	10
7	13	15
8	11	15
9	9	15

Cerințe:

1. Indicați numele și tipul variabilelor implicate în analiza de corelație și regresie.
2. Formulați ipotezele de cercetare și statistice.
3. Reprezentați printr-un grafic sugestiv datele hematologice din Tabelul 5.
4. Prelucrați statistic datele (determinați coeficienții de corelație și de determinare, ecuația dreptei de regresie, testați semnificația statistică aferentă asocierii găsite).
5. Interpretați statistic și medical rezultatele obținute.



Problema 3.3

În tabelul următor sunt listate punctajele obținute după doi ani de pregătire în domeniul științelor de bază și rezultatele obținute la o examinare națională pentru 12 studenți la medicina.

Tabel 6. Punctajele și scorurile obținute de cei 12 studenți la medicina.

Student	Punctaj (X)	Scor (Y)
1	3.35	620
2	2.37	445
3	3.13	445
4	3.10	560
5	1.94	295
6	3.00	570
7	2.85	415
8	1.96	430
9	2.98	560
10	2.55	515
11	2.23	430
12	1.95	435

Cerințe:

1. Indicați numele și tipul variabilelor implicate în analiza de corelație și regresie.
2. Formulați ipotezele de cercetare și statistice.
3. Reprezentați printr-un grafic sugestiv datele hematologice din Tabelul 6.
4. Prelucrați statistic datele (determinați coeficienții de corelație și de determinare, ecuația dreptei de regresie, testați semnificația statistică aferentă asocierii găsite).
5. Care scorul estimat la examinarea națională pentru un student cu punctajul de 3.0?
6. Interpretați statistic și medical rezultatele obținute.

4. Analiza riscului. Aplicații. Interpretarea rezultatelor



Problema 4.1

S-a observat că accidentele vasculare cerebrale (AVC) au loc cu preponderență la tinerii ce consumă droguri recreaționale². Pentru a investiga această problemă, ei au identificat toți pacienții cu vârste între 15 și 44 de ani admiși la un anumit spital și pacienții din grupul de control pentru care consumul de droguri recreaționale nu reprezenta un factor de risc. Datele sunt prezentate în tabelul 7.

² Kaku DA, Lowenstein DH: Emergency of recreational drug abuse as a major risk factor for stroke in young adults. *Ann Intern Med* 1990;113:821–827.

Tabel 7. Datele pentru grupul cu AVC și grupul de control.

	AVC	Control
Abuz de droguri	73	18
Fără abuz de droguri	141	196
Total	214	214

Cerințe:

1. Este abuzul de droguri recreaționale un factor de risc pentru apariția AVC? Calculați OR.
2. Este factorul respectiv semnificativ statistic? Formulați ipotezele statistice. Aplicați testul statistic corespunzător.
3. Calculați intervalul de încredere pentru OR. Considerați un nivel de încredere de 95%.
4. Interpretați statistic intervalul obținut.

**Problema 4.2**

Un studiu *cross-sectional* își propune să analizeze dacă există o relație între fumat și anxietate. Pentru aceasta, un eșantion de 1000 de persoane au fost simultan clasificate în fumători și nefumători, precum și în funcție de nivelul de anxietate. Astfel s-a observat că 300 de persoane au fost identificate ca având un nivel ridicat de anxietate, 500 identificați drept fumători și 200 de persoane au fost identificate drept fumători și cu nivel ridicat de anxietate.

Tabel 8. Datele cu privire la studiul legat de relația dintre fumat și anxietate**Nivel de anxietate**

	Ridicat	Scăzut
Fumător	200	300
Nefumător	100	400
Total	300	700

Cerințe:

1. Care este incidența în grupul de studiu a nivelelor ridicate de anxietate.
2. Care este RR pentru un nivel ridicat de anxietate pentru fumători în comparație cu cei nefumători.
3. Care este valoarea lui OR. Interpretați rezultatul.

5. Analiza supraviețuirii. Aplicații. Interpretarea rezultatelor



Problema 5.1

S-au comparat două grupuri de pacienți cu transplant de rinichi³. Un grup a fost tratat post-transplant cu *azathioprine* pentru minimizarea riscului de rejectare a organului. Celălalt grup a fost tratat cu *cyclosporine*, o substanță imunomodulară. Datele celor două grupuri sunt în fișierul *Birtch.xls* (Anexa).

Cerințe:

1. Realizați curbele de supraviețuire Kaplan-Meier.
2. Aplicați testul *logrank*. Interpretați rezultatele.



Sugestie de rezolvare:

În *EpiInfo v.7* → *Analyze Data (Classic)*, se citește fișierul *Birtch.xls*, ce include informațiile din tabelele 1 și 2.

Pentru realizarea curbelor Kaplan-Meier, se alege comanda *Advanced Statistics* → *Kaplan-Meier Survival*.

☺ Explicați semnificația fiecărei variabile ce intervine în comanda *Kaplan-Meier Survival*.

³ Datele sunt de la studiul inițiat de Dr. Alan Birtch (*Southern Illinois University School of Medicine, Department of Surgery*)

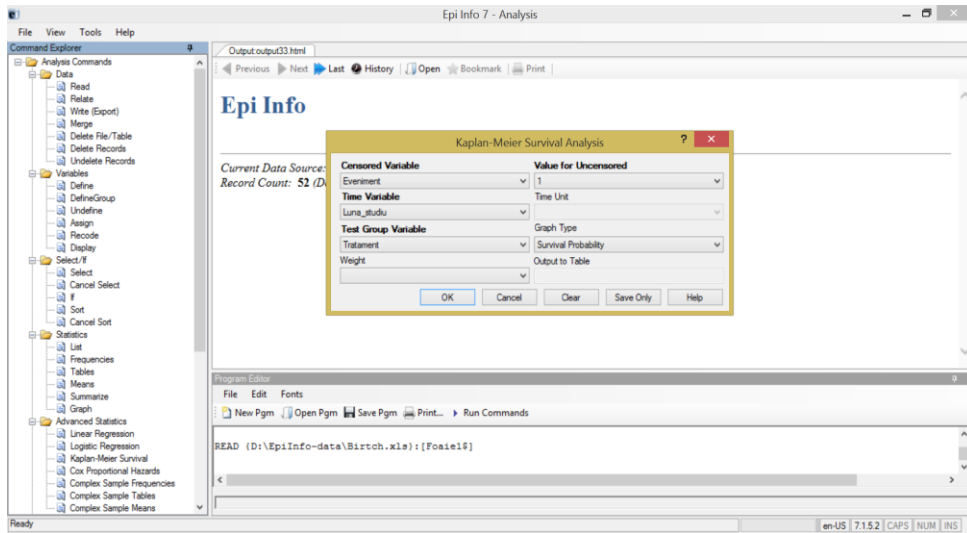


Figura 1. Captură de ecran – comanda *Kaplan-Meier Survival* în *EpiInfo* v. 7.

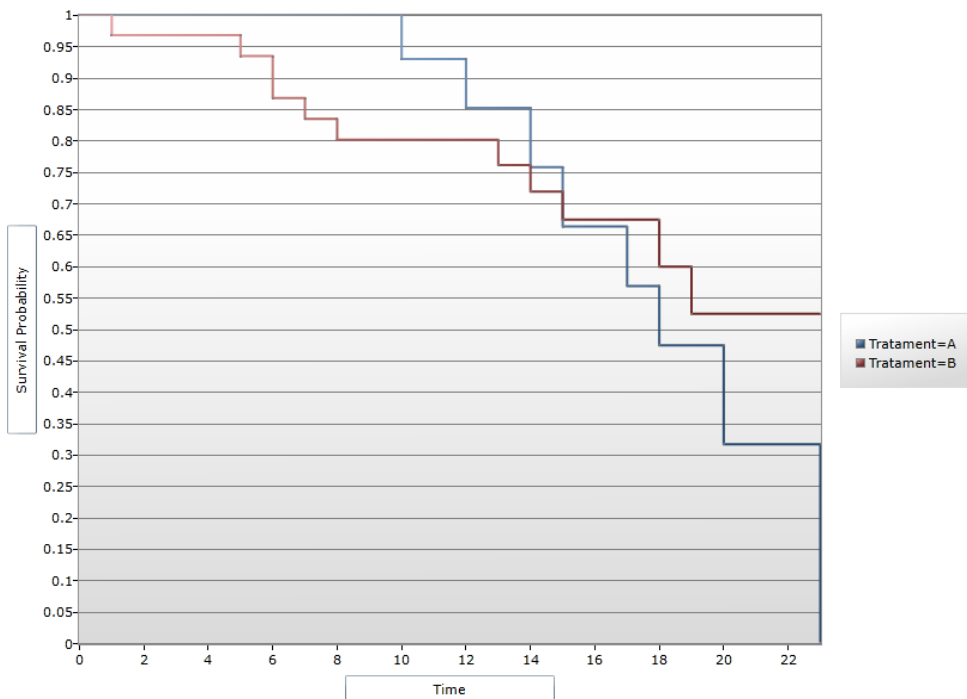


Figura 2. Curbele Kaplan-Meier în *EpiInfo* v.7.



Problema 5.2

Se consideră studiul clinic *Anderson*⁴ ce și-a propus să compare efectul unui steroid (6-mercaptopurine) cu efectul placebo la 42 de pacienți cu leucemie. S-a urmărit durata perioadei fără recidivă (sau până la apariția unei recidive).

1. Realizați curbele de supraviețuire Kaplan-Meier.
2. Aplicați testul *logrank*. Interpretați rezultatele.



Sugestie de rezolvare:

În *EpiInfo* v.7 → *Analyze Data (Classic)*, se citește proiectul **Sample.prj** și se alege fișierul **Anderson**. În prealabil, se bifează căsuța de la *Show* → *Tables*.

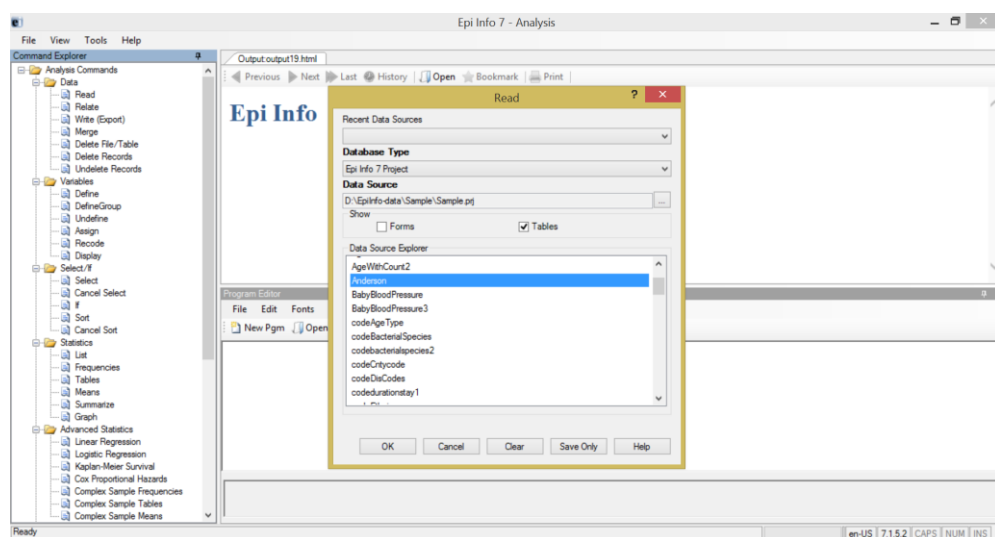


Figura 3. Captură de ecran – citirea fișierului de date *Anderson* în *EpiInfo* v.7.

Pentru realizarea curbelor Kaplan-Meier, se alege comanda *Advanced Statistics* → *Kaplan-Meier Survival*.

⁴ Freireich et al. The effect of 6-mercaptopurine on the duration of steroid-induced remissions in acute leukemia. *Blood* 21: 699-716, 1963.

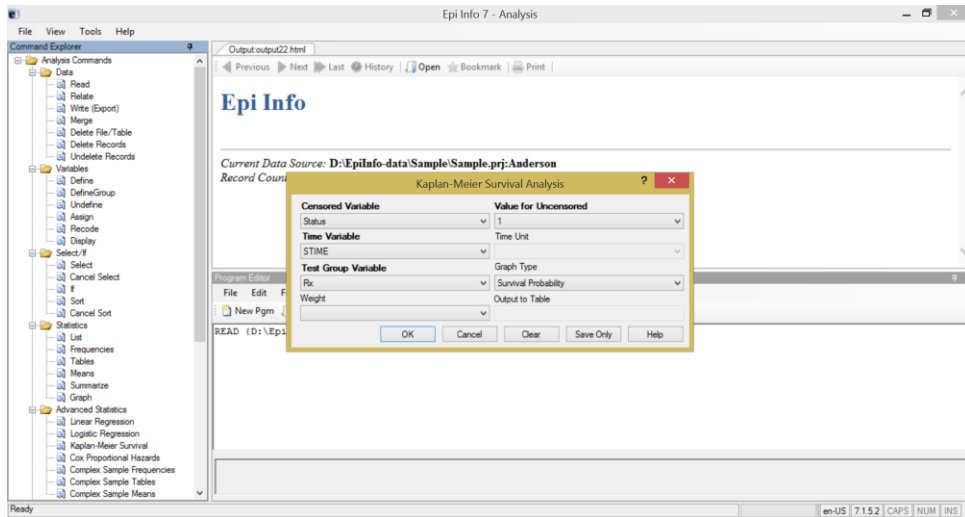


Figura 4. Captură de ecran – comanda *Kaplan-Meier Survival* în *EpiInfo* v.7.

☺ Explicați semnificația fiecărei variabile din comanda *Kaplan-Meier Survival*. Timpul de supraviețuire în săptămâni până la recidivă – **Stime**; variabila care indică dacă s-a întâmplat un eveniment – **Status**; variabila care indică dacă pacientul a luat un tratament sau este doar sub efect placebo – **Rx (0=tratament, 1=placebo)**.

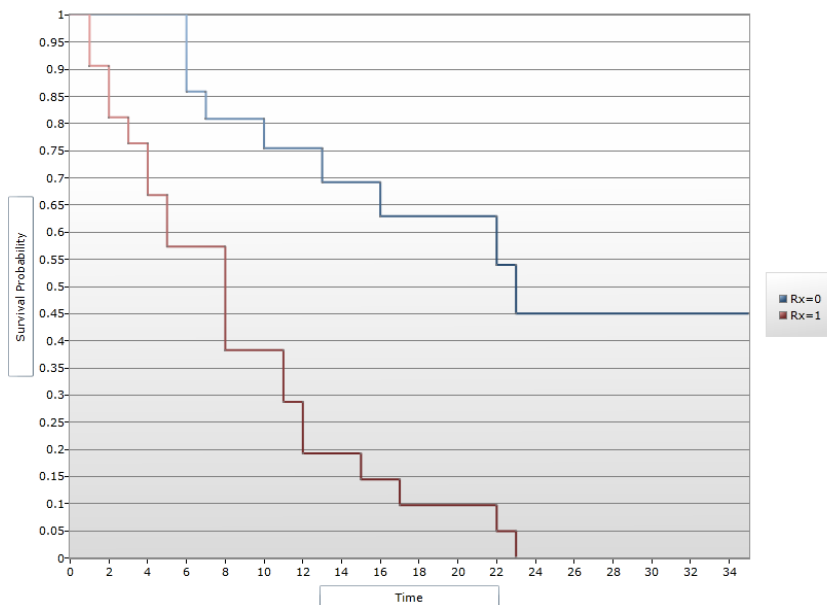


Figura 5. Curbele Kaplan-Meier în *EpiInfo* v.7. Albastru – pacienții cu tratament, roșu – pacienții sub efect placebo.

Figura 5 include comparativ curbele de supraviețuire după un anumit număr de zile de urmărire pentru cele două grupuri de pacienți. Probabilitatea de supraviețuire la 12 zile este de aprox. 20%(0,2) pentru grupul pacienților fără tratament, comparativ cu aprox. 75%(0,75) pentru grupul pacienților cu tratament. După 34 de zile de urmărire, probabilitatea de supraviețuire este de 0% pentru grupul pacienților fără tratament, comparativ cu 45% (0,45) pentru grupul pacienților cu tratament.

Aplicând testul *logrank*, se observă că există diferențe semnificative între cele două curbe – între cele două grupe de pacienți ($p < 0.05$). Grupul pacienților cu tratament au o probabilitate de supraviețuire mai mare.



Problema 5.3

Se consideră datele extrase dintr-un studiu cu pacienți HIV pozitiv, ce și-a propus să examineze dacă există diferențe semnificative între probabilitățile de supraviețuire ale pacienților care au folosit substanțe (intravenos) și cei care nu au folosit.

1. Realizați curbele de supraviețuire Kaplan-Meier.
2. Aplicați testul *logrank*. Interpretați rezultatele.



Sugestie de rezolvare:

În *EpiInfo v.7* → *Analyze Data (Classic)*, se citește proiectul **Sample.prj** și se alege fișierul **hmohiv**.

Setul de date conține variabilele: ID pacient, timpul de supraviețuire, data de intrare în studiu, ultima data înregistrată/data ieșirii din studiu, vârsta, utilizare substanță și variabila eveniment (*censor*).

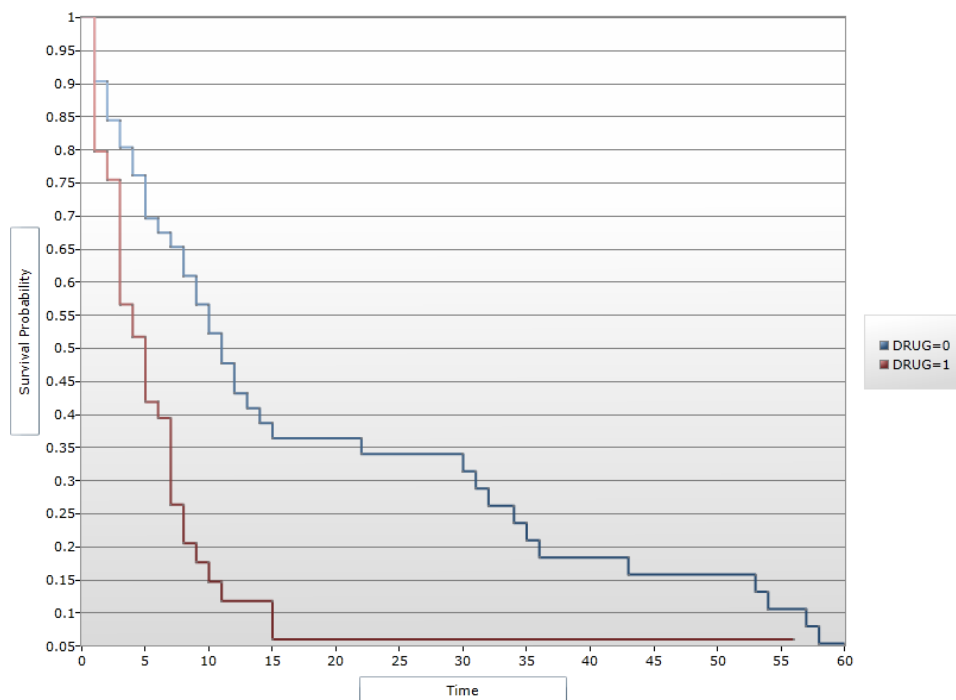


Figura 6. Curbele Kaplan-Meier în EpiInfo v.7.

6. Probleme recapitulative

Fișier de lucru: **BPOC.xlsx** (Anexa)

Descrierea fișierului:

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-over*, la diferite momente ale studiului. Fișierul conține datele culese la două momente ale studiului: (1) la început, (2) după medicație.

- NUME (numele pacientului, sub formă de inițiale)
- SEX (genul)
- AGE (vârsta)
- DISPN (dispnee = respirație dificilă) 0 = Nu; 1 = Da
- RAL (raluri = respirație cu zgomot ca un “hurut”) 0 = Nu; 1 = Da

La începutul studiului (înainte de tratament):

- TMAX1 (tensiunea arterială sistolică)
- TMIN1 (tensiunea arterială diastolică)

- FC1 (frecvența cardiacă)
- CVV1 (capacitatea vitală) - valoare absolută
- CVP1 (capacitatea vitală) - procent din valoarea estimată ca fiind normală pentru pacient
- DEP1 (debitul expirator) - procent din valoarea estimată ca fiind normală pentru pacient
- VEMSV1 (volumul expirator maxim pe secundă) - valoare absolută
- VEMSP1 (volumul expirator maxim pe secundă) - procent din valoarea estimată ca fiind normală pentru pacient
- DEM1 (debitul expirator maxim)
- IPB1 (indicele de permeabilitate bronșică, indicele Tiffeneau)

La momentul al II-lea al studiului (după tratament):

- TMAX2, TMIN2, FC2, CVV2, CVP2, DEP2, VEMSV2, VEMSP2, DEM2, IPB2
- MEDIC (medicație primită de pacient)



Pentru fiecare dintre problemele următoare, răspundeți la cerințele:

- a) Definiți ipoteza de cercetare
- b) Definiți ipotezele statistice
- c) Alegeți și aplicați testul potrivit
- d) Interpretați rezultatul statistic și medical.

Probleme

- P1. Studiați existența unei corelații între vârsta și capacitatea vitală la începutul studiului (CVV1). Realizați un grafic sugestiv care să conțină și dreapta de regresie.
- P2. Studiați existența unei corelații între variabila CVP1 și volumul expirator maxim pe secundă la începutul studiului (VEMSP1). Realizați un grafic sugestiv care să conțină și dreapta de regresie.
- P3. Estimați media vârstei pentru populația din care s-a extras acest eșantion, cu un nivel de încredere de 68%.
- P4. În ce interval se regăsesc 95% din valorile tensiunii sistolice după tratament?
- P5. Realizați tabelul de frecvență după variabila RAL și realizați un grafic sugestiv.
- P6. Există asociere între dispnee și sexe? Realizați un grafic sugestiv.
- P7. Frecvența cardiacă se modifică semnificativ după tratament?

- P8. Exista diferențe semnificative între indicele Tiffeneau (IPB2) după sexe?
- P9. Exista diferențe semnificative între valorile capacității vitale după tratament (CVV2) după tipul medicației?
- P10. Realizați histograma valorilor TMIN2 cu pasul h calculat.
- P11. Vârsta de peste 50 de ani reprezintă factor de risc pentru creșterea tensiunii (TMAX1) peste 140?
- P12. Volumul expirator maxim pe secundă în valoare absolută se modifica semnificativ după tratament?

7. Concluzii

În acest material ne-am însușit noțiunile legate de:

- ✚ Identificarea testului statistic potrivit în funcție de tipul de studiu/problema;
- ✚ Formularea ipotezelor de cercetare și statistice;
- ✚ Aplicarea testelor statistice și formularea deciziei statistice;
- ✚ Identificarea problemelor de analiză a corelației și regresiei liniare;
- ✚ Reprezentarea grafică a datelor într-o problemă de analiza corelației și regresiei;
- ✚ Calcularea coeficientului de corelație și a coeficientului de determinare;
- ✚ Interpretarea dreptei de regresie;
- ✚ Identificarea unei probleme de analiza riscului;
- ✚ Formularea ipotezelor de cercetare și statistice pentru o problemă de analiză a riscului;
- ✚ Calcularea și interpretarea indicatorilor RR și OR;
- ✚ Calcularea și interpretarea intervalelor de încredere pentru RR și OR;
- ✚ Reprezentarea curbelor de supraviețuire Kaplan-Meier în EpiInfo v.7;
- ✚ Interpretarea curbelor Kaplan-Meier;
- ✚ Aplicarea și interpretarea testului *logrank*.

Referințe

- Mihalas GI, Lungeanu D, Informatica medicală și biostatistică, Editura „Victor Babes”, 2009: 1-225.
- Rosner B. Fundamentals of biostatistics (5th Edition). Pacific Grove: Duxbury. Thomson Learning, 2000.
- Knapp RG, Clinton Miller M: Clinical epidemiology and biostatistics. Baltimore: Williams & Wilkins, 1992.
- Machin D, Campbell MJ, Walters SJ, Medical Statistics Fourth Edition, John Wiley & Sons, 2007.