

Analiza datelor aplicata în tratamentul radiologic al cancerelor pulmonare

Ing. Carmen DOGARU
ORANGE România

The article presents a study of data analysis performed on a lot of patients with lung cancer treated by radiotherapy in the Clinic of Radiology-Oncology from Coltea Hospital Bucharest from 1997 till 2002. The results are obtained by using a software applying the Ward method on the above mentioned lot of patients are presented and commented along with the Principal Component Analysis applied.

Keywords: data analasys, radiology.

Analiza datelor s-a dezvoltat din statistica matematica preluându-i metodele si amplificându-le si devenind astfel specifice acestei noi discipline. Analiza datelor foloseste algoritmi matematici si prelucreaza volume mari de date, ceea ce presupune un volum imens de calcule, evolutia ei ca disciplina a luat un avânt deosebit în paralel cu evolutia tehnicii de calcul. În practica medicala analiza de date a devenit tot mai mult un ajutor în analiza volumului mare de informatii si în oferirea unor interpretari ale fenomenelor medicale, care sa asiste medicul în luarea deciziilor optime de tratament.

Astfel, exista numeroase încercari de grupare obiectiva a pacientilor în functie de comportarea lor la diversi stimuli. Pe baza acestor clasificari se pot construi scoruri de evaluare a sanatatii unor colectivitati.

Prezentam un studiu care evalueaza subclase pentru un lot de pacienti suferind de cancer pulmonare, care au fost tratati în Clinica de Radiologie-Oncologie a Spitalului clinic Coltea, în perioada 1997 – 2002.

Sex	Numar	%
Barbati	870	96,03
Femei	36	3,97
Total	906	100

Din aceste date se poate constata o nota predominanta a sexului masculin, fapt cunoscut, ce se regaseste în toate studiile mondiale legate de cancerul pulmonar. Ne atrage atentia procentul de aproape 4% a neoplaziilor aparute la femei, ceea ce este în continua crestere, pentru ca o analiza similara facuta în urma

cu 5 ani depistase sub 2% cazuri de cancer pulmonare la femei. Acest fapt poate fi explicat în mare masura prin tabagism, care a devenit o practica tot mai raspândita în rândul femeilor, în plus consumul de alcool si expunerea la noxe profesionale cu potential cancerigen contribuie si ele la aceasta crestere semnalata de cazuri.

Mediu	Numar	%
Urban	441	51,33
Rural	405	48,67
Total	906	100

Se observa ca din mediul urban provin 441 de bolnavi fata de 405 de la sate, aceasta se explica prin poluarea marilor orase, cu toate implicatiile ei in carcinogeneza. Totusi diferentele sat si oras sunt minime si asta se explica numai prin influenta majora pe care fumatul si consumul de bauturi alcoolice îl are asupra organismului uman.

Vârsta (ani)	Numar	%
< 30	6	0,66
31 – 40	48	5,30
41 – 50	153	16,89
51 – 60	357	39,40
61 – 70	270	29,80
> 70	72	7,95
Total	906	100

Repartitia statistica în functie de vârsta permite o serie de consideratii interesante. Astfel, în categoria sub 30 de ani se observa 2 pacienti, ceea ce este o raritate la acest tip de localizare a cancerului. Si în grupa de vârsta 31–40 de ani se constata un numar de bolnavi care procentual depaseste datele existente în

literatura, privind incidenta neoplaziilor pulmonare pe plan mondial. La celelalte grupe de vârstă incidenta acestui tip de cancer este

în limite normale și similare studiilor din alte țări.

Tipologie	Numar	%
Mari fumatori (> 20 țigarete / zi)	459	50,66
Fumatori de 10 – 20 țigarete / zi	219	23,18
Fumatori ocazionali (< 10 țigarete / zi)	138	15,23
Nefumatori	99	10,93
Total	906	100

Datele din tabelul de mai sus confirmă încă odată rolul tabagismului ca factor principal în apariția cancerului pulmonar și accentuează faptul că numărul de țigarete fumate zilnic

are o mare importanță, ca și numărul de ani cât a fumat fiecare din pacienții lotului examinat.

Tipologie	Numar	%
Consumatori ocazionali de alcool	501	55,30
Mari consumatori de alcool	405	44,70
Total	906	100

În cazul consumului de alcool se constată că acesta intervine mai puțin pregnant în etiologia cancerului pulmonar. Totuși s-a remarcat

că el reprezintă un factor care potențează apariția acestui tip de boală în combinație cu fumatul.

Tipul histopatologic	Numar	%
Carcinom spinocelular bine diferentiat	270	29,80
Carcinom spinocelular mediu diferentiat	123	13,58
Carcinom spinocelular puțin diferentiat	390	43,05
Carcinom epidermoid nediferentiat	99	10,93
Carcinom 'in situ'	15	1,65
Alte forme	9	0,99
Total	906	100

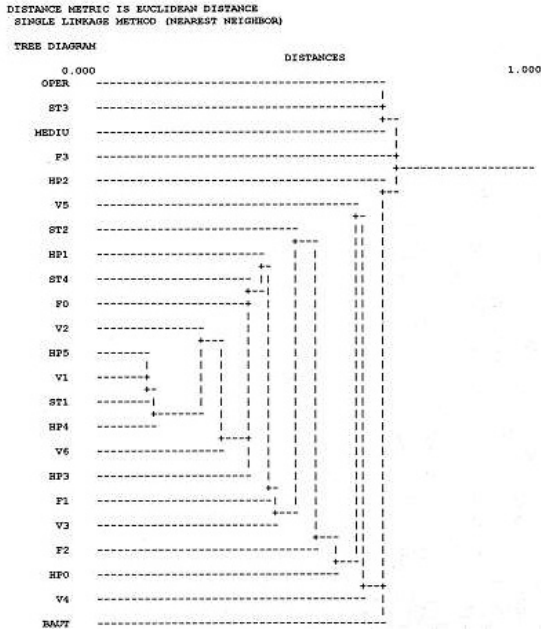
Se observă că forma histopatologică dominantă la acest tip de neoplazie malignă este carcinomul spinocelular. Numărul redus de carcinoame "in situ" și alte forme histopatologice se înscrie în procentul de 2–3 % menționat de toate studiile de specialitate.

În cazul cancerelor pulmonare tratamentul este complex și include intervenția chirurgicală, radioterapia și chimioterapia. Locul fiecărui tip de tratament menționat în protocolul terapeutic este stabilit de la caz la caz de o comisie de medici alcătuită dintr-un chirurg O.R.L., un radioterapeut și un chimioterapeut. Se observă că stadiul III este cel mai bine reprezentat numeric și atitudinea care se adoptă în aceste situații este radioterapia și chimioterapia, urmate eventual de chirurgie.

Pentru evaluarea de subclase pe diverse criterii în lotul de pacienți studiat s-a încercat o împărțire în clustere a acestora după criteriile

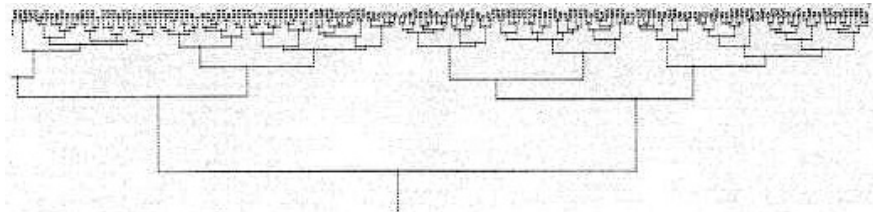
principale ale bazei de date: sex, mediu de viață, vârstă, număr de țigarete fumate, consum de alcool, tipul histopatologic al bolii. Se observă că grupul se divide net în două subgrupe după sex.

În continuare am alcătuit două baturi și am analizat clasificarea celui de bărbați, care este cel mai bine reprezentat din cele două loturi. Facând gruparea pe coloane a bărbaților suferind de cancer pulmonar după metoda Ward se observă că anumite caracteristici ale lotului cum ar fi VI (vârsta sub 30 de ani) sunt apropiate de HP4 (examen histopatologic pentru carcinomul "in situ"), HP5 (alte tipuri de tumori benigne) și stadiul I de boală. Aceste date au puțini reprezentanți în lotul studiat ceea ce corespunde cu datele din literatura internațională de specialitate, de aceea se decide eliminarea lor din analiza ulterioară pentru aflarea caracteristicilor principale.



Restul caracteristicilor sunt suficient de departate pentru a fi considerate independente. Împartirea în clustere evidentiaza doua grupe distincte si una dintre ele se divide în alte doua. Exista deci trei grupe de bolnavi care vor fi folosite pentru analiza în componente principale.

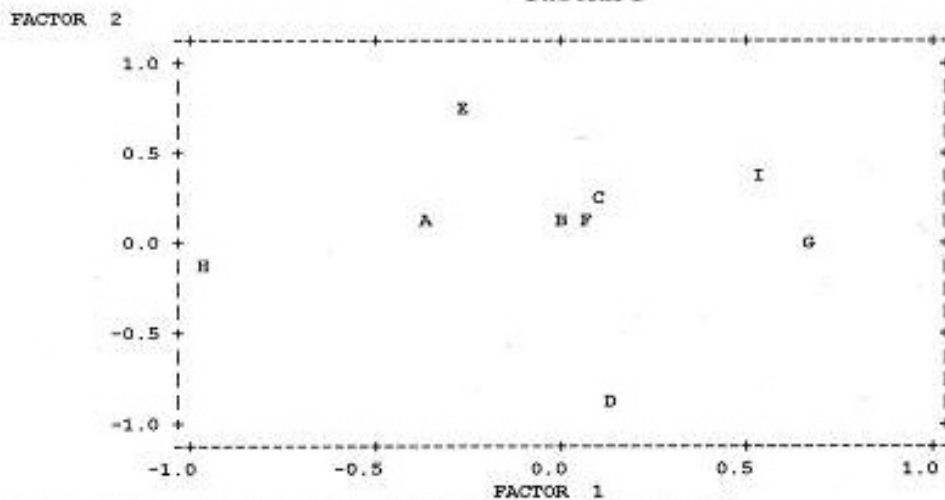
În urma analizei în componente principale a pacientilor barbati suferind de cancer pulmonare tratati în Clinica de Radiologie-Oncologie Coltea în perioada 1997-2002 observam ca "Factorul 1" este cel principal si determina împartirea lotului în proportie de 21%. Asta se poate interpreta astfel: factorul mentionat este puternic influentat negativ de caracteristica "fumat mai mult de 20 țigarete zilnic" si puternic influentat pozitiv de caracteristica "nefumatori".



Deci aceste doua caracteristici sunt în opozitie (tabelul 1) si factorul 1 este un factor

principal reprezentând aproape o patrime din valori.

Tabelul 1



unde H este "fumat mai mult de 20 de țigarete zilnic"

G este "nefumători"

E este vârsta pacientului este cuprinsă în intervalul 51-60 de ani

D este vârsta pacientului este cuprinsă în intervalul 61-70 de ani

PERCENT OF TOTAL VARIANCE EXPLAINED		
1	2	3
21.130	17.130	15.470

Un al doilea factor principal are o proportie de 17% si se explica prin opozitia celor doua grupe de vârsta care marcheaza pragul dupa care cancerule pulmonare sunt foarte frecvente. Este vorba de decadele 51-60 ani si 61-70 de ani, care sunt cele mai bine reprezentate în lotul analizat. Se poate deci trage concluzia din analiza în componente principale ca 40%

din cazurile de cancerule pulmonare la barbati se datoreaza fumatului a mai mult de un pachet de tigari zilnic, dar si trecerii pragului de vârsta de 51-60 ani. În mod normal se continua analiza cu calculul de discriminant pentru componentele principale depistate. Raspândirea pacientilor cu cancerule pulmonare din lotul analizat este prezentat în figura 1.

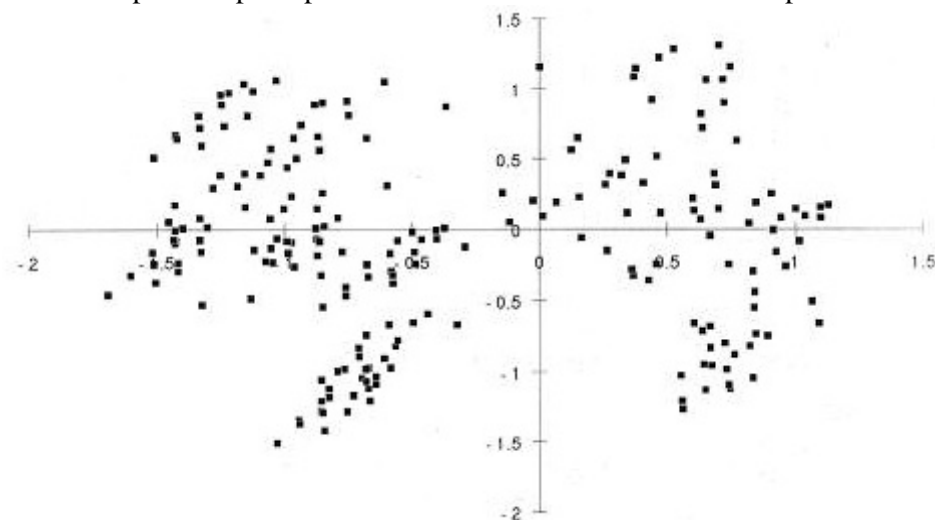


Fig. 1. Raspândirea pacientilor cu cancerule pulmonare

Studiul de fata a fost realizat prin utilizarea unui program care foloseste algoritmul Ward si rezultatele obtinute au fost validate în S.P.S.S.

Bibliografie

1. R.K.Brock & W. Krisher, 'The Data Analysis', Ed. European Physical Society, Geneva, 2003

2. J. Hand, 'Discrimination and Classification', Ed. John Wiley&Sons, London, 1990

3. 'Cancer', Ed. Springer-Verlag Heide l-berg, 2003

4. G. Saporta & V. Stefanescu, 'Analiza Datelor & Informatica', Ed. Economica, Buc u-resti, 1996