

# APLICAREA ÎN DOMENIUL ONCOLOGIEI MEDICALE A METODEI VaR DE EVALUARE A RISCULUI

Gabriel-Cristian Ene

e-mail: [egabi@dbshospital.valahia.ro](mailto:egabi@dbshospital.valahia.ro)  
Spitalul Județean-Târgoviște

Rezumat: În lucrare, se prezintă unele rezultate ale cercetării științifice proprii privind EXTINDEREA metodei VaR (de evaluare a riscului) din domeniul financiar-bancar în domeniul oncologiei medicale. Studiul fundamentelor teoretice ale acestei metode conduce la ideea că metoda VaR nu depinde de domeniul aplicației bancar-financiar, medical, tehnologic-industrial etc. și că se poate aplica și în alte cazuri în care datele nu provin neapărat din mediul financiar-investițional.

Cuvinte cheie: ipoteze statistice, Value at Risk (VaR), sistem informatic, baza de date oncologie.

## 1. Introducere

În vederea achiziției și prelucrării volumului imens de date obținute în timpul supravegherii terapeutice și screening-ului în domeniul oncologiei medicale, a fost testat la un sistem informatic de achiziție și stocarea datelor din domeniul oncologiei medicale [1]. În [2], sunt prezentate rezultatele prelucrării datelor din baza de date 1999/2004 a Clinicii de Oncologie și UMF Iași. În ambele lucrări citate, sunt prezentate rezultate ale prelucrării datelor privind supraviețuirea pe zone geografice, pe decade de vârstă, pe localizări (col uterin, sân, ficat, plămân etc.) sau pe metode de tratament pre și postoperatoriu. Toate aceste evaluări statistice, făcute pe baza datelor stocate, au un caracter istoric în sensul că mulțimea de evenimente medicale, reflectate în datele respective, este caracterizată statistic din punctul de vedere al evoluției în trecut a unor asemenea mulțimi de evenimente aleatoare. În domeniul investițiilor financiare pe piața de capital, situația este oarecum similară din punctul de vedere al stocării și prelucrării datelor privind garanția obținerii unor rezultate palpabile în viitor. În asemenea stocuri de date, evenimentele de câștig ori pierdere sunt fluctuații în jurul mediei în timp, amplitudinea fluctuației pozitive este măsura succesului (câștigului), iar cele negative exprimă pierderi (insucces). Pe piața de capital, aceste fluctuații ale unor active, în final, determină randamentul financiar al investiției făcute de administratorul sau managerul unui portofoliu adică cel care a hotărât câte și de care active(acțiuni) să achiziționeze.

VaR este denumirea standard pentru un indicator statistic, obținut prin aplicarea unei noi metode denumită în extenso *Value at Risk*. Acest VaR exprimă probabilist cât de mult se poate pierde sau câștiga într-o lună într-o investiție. Pentru a putea face această evaluare, este necesară determinarea modelului matematic al randamentului activelor achiziționate de pe piața de capital. În acest scop, se folosesc datele experimentale, existente în baza de date a societății, cu privire la activele respective, pe o perioadă lungă de timp. Studiul fundamentelor teoretice ale acestei metode conduce la ideea că metoda nu depinde de domeniul aplicației bancar-financiar, medical, tehnologic-industrial etc. și că se poate aplica și în alte cazuri în care datele care nu provin neapărat din mediul financiar-investițional. Pentru aceasta, în continuare, este prezentată această metodă, mai în detaliu și este testată pe date simulate.

## 2. Modelul temporal și modelul probabilistic al fluctuațiilor

Acest exemplu ilustrează numeric faptul că *VaR este o tehnică pentru estimarea probabilității ca pierderile unui portofoliu să depășească o valoare prestabilită X*. Întrebarea la care VaR încearcă să răspundă este:

**Care este probabilitatea ca, investind în active pe o piață normală, ca în următoarele luni, pierderile să fie  $X\% \text{ €}$  din capitalul investit?**

Vom presupune că o firmă are în administrare un portofoliu care constă dintr-un singur „produs”. Randamentul acestui produs reprezintă câștigul pe care-l produce el zilnic în cadrul acestui portofoliu administrat de firmă. În condiții normale de piață, randamentul este o variabilă aleatoare cu distribuție gaussiană (adică normală). Această distribuție este caracterizată prin medie (sau speranță) și abatere standard sau volatilitate. Practic, volatilitatea exprimă amplitudinea oscilațiilor randamentului în jurul mediei.

În exemplul real considerat, randamentul produsului administrat este normal distribuit cu o medie anuală de 10% și o abatere standard de 30% la o valoare a portofoliului zilnic de 100 mil. €. Se cere ca, la finele anului, să se răspundă la câteva simple întrebări privitoare la distribuția valorilor portofoliului:

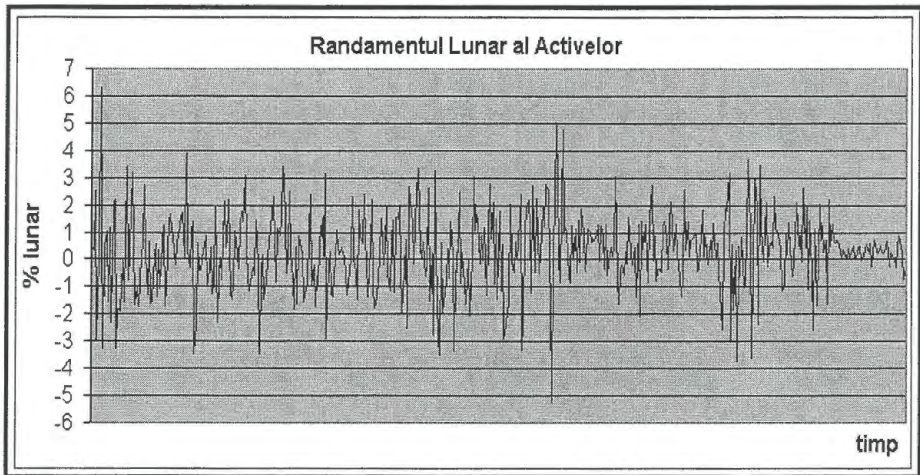


Figura 1. Realizările lunare ale randamentului activelor unui portofoliu dat [4]

1. Care este distribuția valorilor de portofoliu la finele de an?
2. Care este probabilitatea ca pierderea la finele de an să depășească 20 mil. € (cu alte cuvinte, care este probabilitatea ca pierderea la finele de an să aibă o valoare de sub 80 mil. €)?
3. Care este maximul de pierdere la finele de an, care se produce cu probabilitatea de 1%? Aceasta este VaR la 1%.

Presupunem că cele 100 mil. € sunt investite în active pe termen mediu. Cât de mult se poate pierde sau câștiga într-o lună în această investiție? Fără a răspunde la această întrebare, investitorii nu au cum să se decidă dacă randamentul pe care îl pot avea activele compensează riscul luat de investitor în momentul când a decis să facă această investiție. Pentru a putea răspunde la întrebarea de mai sus, este necesară determinarea modelului matematic al randamentului activelor. În acest scop, se folosesc datele experimentale, existente în baza de date a societății, cu privire la activele respective, pe o perioadă lungă de timp și care constituie o *serie de timp*. Pentru ca această *serie de timp* să fie utilă determinării modelului matematic al randamentului, se admite *ipoteza de staționaritate a seriei*. Această ipoteză presupune, practic, că tipul modelului și parametrii modelului (determinat pe baza acestor date experimentale) rămân aceiași și în viitor, o perioadă suficient de mare de timp.

BANDA	Nr. de evenimente	BANDA	Nr. de evenimente
A1<-5 mil €	1 din 503	0,5<A13<1 mil €	65 din 503
-5<A2<-4,5 mil €	0 din 503	1<A14<1,5 mil €	40 din 503
-4,5<A3<-4 mil €	0 din 503	1,5<A15<2 mil €	30 din 503
-4<A4<-3,5 mil €	10 din 503	2<A16<2,5 mil €	15 din 503
-3,5<A5<-3 mil €	9 din 503	2,5<A17<3 mil €	10 din 503
-3<A6<-2,5 mil €	10 din 503	3<A18<3,5 mil €	3 din 503
-2,5<A7<-2 mil €	15 din 503	3,5<A19<4 mil €	2 din 503
-2<A8<-1,5 mil €	35 din 503	4<A20<4,5 mil €	0 din 503
-1,5<A9<-1 mil €	30 din 503	4,5<A21<5 mil €	1 din 503
-1<A10<-0,5 mil €	55 din 503	5<A22<5,5 mil €	0 din 503
-0,5<A11<0 mil €	85 din 503	5,5<A23<6 mil €	0 din 503
0<A12<0,5 mil €	95 din 503	6<A24<6,5 mil €	1 din 503

Figura 2. Frecvențele de apariție ca evenimente discrete (A1...A12) a valorilor seriei din figura 1

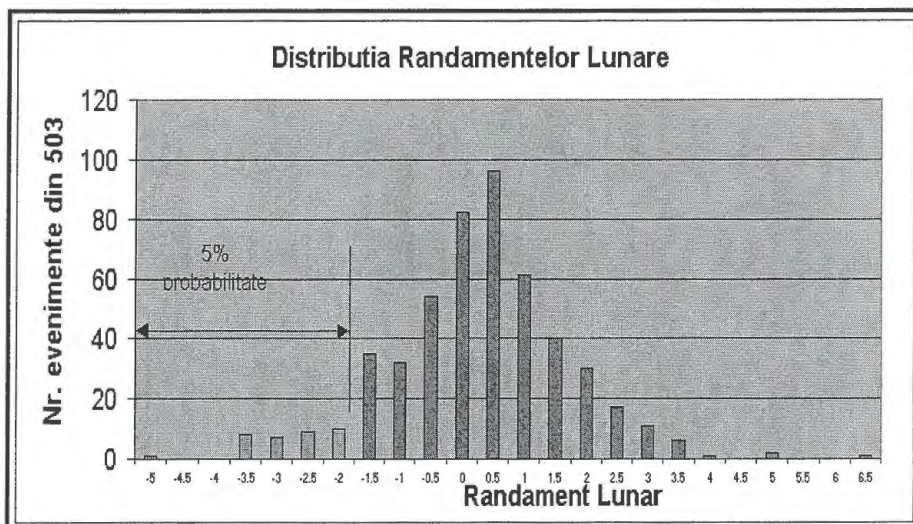


Figura 3 Histograma frecvențelor de apariție a valorilor seriei de timp din figura 2

Modelele matematice care pot caracteriza și descrie aceste date experimentale disponibile sunt de două categorii: *temporale și probabilistice*. Una din caracteristicile unui portofoliu este *volatilitatea* acestuia, care depinde de fluctuațiile randamentului, respectiv, de dispersia acestuia. Deci, valorile lunare ale randamentului nu sunt constante, ele reflectă fluctuațiile pieții de capital și, prin urmare, sunt evenimente discrete aleatoare. În figura 1, sunt prezentate valorile lunare ale randamentului activelor pe parcursul a 503 luni. Aceste date experimentale sunt disponibile pentru calculul parametrilor modelului matematic. Există două variante de model pentru reprezentarea datelor: modelul temporal și modelul probabilistic. Aceste modele au semnificații și reprezentări diferite. Modelul probabilistic al seriei temporale din figură exprimă probabilitățile (frecvențele)  $P_k$  cu care se produce evenimentul: Aceasta exprimă probabilitatea ca valorile,  $y(t)$ ,  $t = 1, 2, \dots, 503$ , ale randamentului să fie cuprinse în intervalul  $A_k$  din mulțimea dată de intervale:

$$-\infty < A_1 < -5, -5 < A_2 < -4,5, -4,5 < A_3 < -4, \dots, -1 < A_{10} < -0,5, \dots, 6 < A_{24} < 6,5$$

Randamentul oscilează de la un minim de  $-5.27\%$  (ceea ce corespunde unei pierderi de  $5,27$  mil. € la o sută de  $100$  mil. € investiție) până la un maxim de  $+6.31\%$  (care corespunde unui câștig de  $6,31$  mil. € la  $100$  mil. € investiție). Putem construi și analiză histogramă a acestor valori de la cea mai mică la cea mai mare (figura 3). Observăm că există o singură apariție peste  $6\%$ , încă două între  $5,5\%$  și  $4,5\%$  ș.a.m.d. Putem construi o distribuție a randamentelor care numără câte evenimente aleatoare aparțin fiecărui subinterval. Frecvențele care exprimă de câte ori apar valori în fiecare din aceste intervale, sunt date în tabelul din figura 2. Putem alege un nivel de încredere, să spunem  $95\%$ . Pentru acest nivel de încredere, putem găsi pe grafic un punct pentru care există o probabilitate de  $100-95=5\%$  și să găsim un randament mai mic. Numărul căutat este  $-1,75\%$ , deoarece situațiile în care randamentul este mai mic decât  $-1,75\%$  reprezintă  $5\%$  din numărul total de luni ( $25$  din cele  $503$  luni).

### 3. Modelul fluctuațiilor unor date de natură oncologică [1]

Rezultatul de mai sus putea fi obținut și pornind de la expresia analitică  $f(m, \sigma)$  a unei distribuții normale date cu media  $m$  și varianța  $\sigma$  (deviația standard). Spre exemplu, pentru datele statistice  $Y$  privind afecțiunile de *COL UTERIN*, reprezentate în figura 4, s-a aplicat metoda de verificare a normalității datelor. Testul statistic a confirmat caracterul gaussian al datelor din figura 4. Deoarece funcția densității de repartiție normală (Gauss) are doar doi parametri, s-au calculat aceștia (media  $m$  și abaterea standard  $\sigma$  *sigmaCOL*). Valorile calculate ale acestor parametri sunt afișate pe graficul din figura de mai jos.

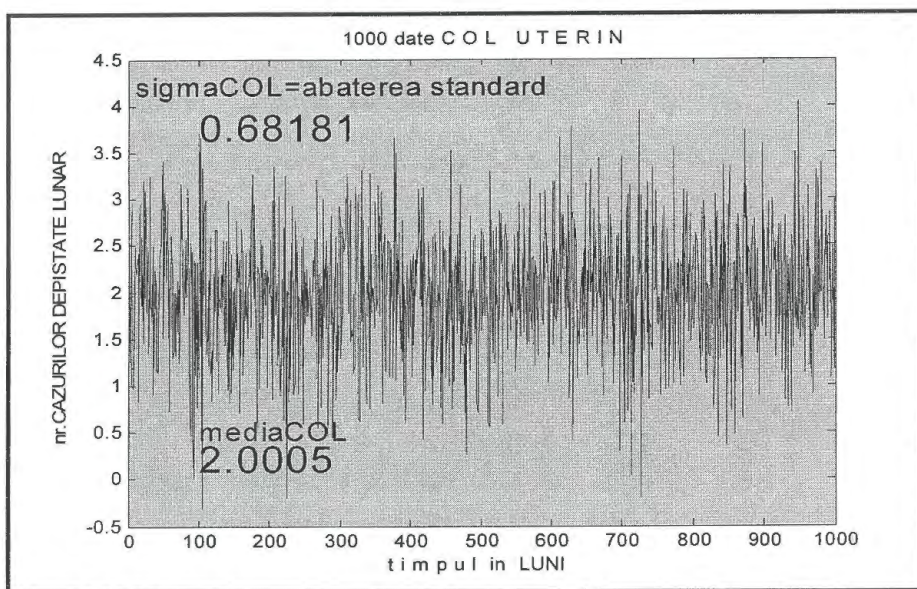


Figura 4. Seria de timp privind evoluția numărului de cazuri analizate lunar privind „col uterin” [2]

Pentru a aplica VaR, în acest caz, nu este necesar tot ansamblul de date, ci este suficientă funcția distribuției densității de probabilitate fie în forma analitică [5], fie în forma grafică.

#### INSTRUCIUNILE MATLAB PENTRU CALCULUL DISTRIBUTIEI GAUSS

```
media=mean(y)
sigma=std(y)
a=1/(sigma*(sqrt(2*pi)))
b=2*sigma2
for k=-.5:1:4.5
c=a*exp(-(k-media)2/b)
end
```

Figura 5. Program Matlab pentru calculul mediei ,dispersiei și repartiției probabilităților

În figura 5, este prezentat programul Matlab pentru calculul funcției de distribuție, pe baza celor doi parametri calculați mai sus (figura 4). Primele două linii din această secvență calculează media și abaterea standard dată pentru seria de timp  $Y$ .

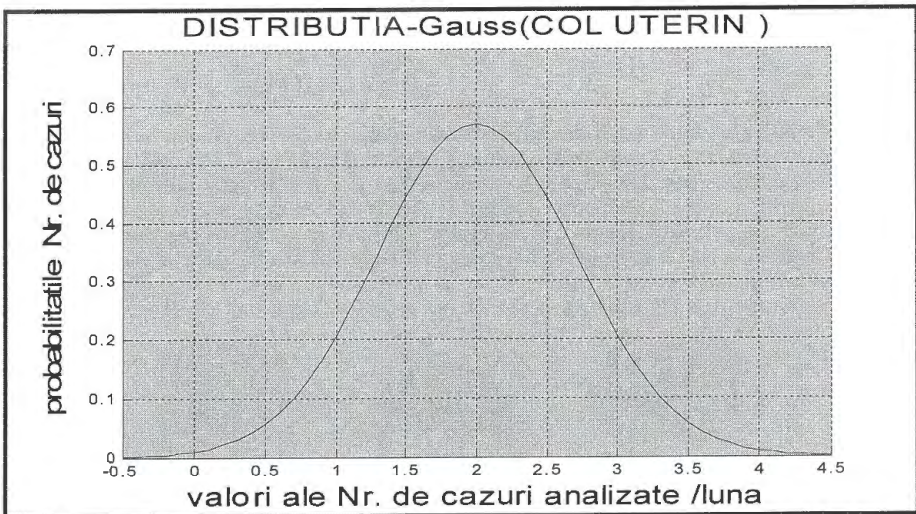


Figura 6. Reprezentarea grafică a repartiției probabilităților pentru datele din figura 4

Graficul funcției normale de distribuție, astfel calculat, este reprezentat în figura 6.

Din acest grafic, se poate determina, de exemplu, care este probabilitatea (riscul), ca numărul lunar de cazuri să crească peste nr.=3. Din grafic rezultă că, în luna curentă, probabilitatea aceasta este 0,2.

## 4. Concluzii

Scopul acestei cercetări a constat în verificarea metodei VaR prin simulare și în detalierea aspectelor de aplicare, cu caracter general, care să faciliteze extinderea aplicațiilor metodei și în alte domenii inclusiv cel medical. Lucrarea prezintă o cale de extindere a unei metode de evaluare a riscului din domeniul financiar în domeniul medical. Este indicată o cale prin care în locul masivelor de date concrete să se utilizeze modelul acestor date. În cazul concret prezentat, se folosește ca model funcția distribuției normale a densității de probabilitate. Pentru a putea aplica această reprezentare, se impune efectuarea prealabilă a testului statistic de normalitate a datelor inițiale.

## Bibliografie

1. **ENE G. C., FIROIU E., OGERU M., LUPĂȘCU:** Sistem informatic de procesare evaluare și monitorizare a datelor screening-ului în cancerul colului uterin. A treia Conferință Națională de oncologie medicală. Ed. Medicală Universitară, Craiova, 2004.
2. **CIORNEA D., E. BILD:** Analiza cancerului de col uterin din baza de date ONCOSOFT pe perioada 1999-2004. A treia Conferință Națională de oncologie medicală. Ed. Medicală Universitară, Craiova, 2004.
3. **GHINEA M.:** Matlab, Ed. Teora, București, 2006.
4. **TOMIȚĂ T.:** Valoarea la Risc, Value at Risk (VaR) - 2006. t.tomita@brm.ro
5. **CRAIU V.:** Verificarea Ipotezelor Statistice. Editura Didactică și Pedagogică, București, 2006.