Introducere în calculul probabilităţilor

Sorana D. BOLBOACĂ

» Experimentul aleator
» Spaţiul fundamental de evenimente
» Definiţia clasică a probabilităţii
» Independența a două evenimente
» Probabilităţi condiţionate (RR, Se, Sp, VPP, VPN, etc.)
» Variabile aleatoare

Content

Evenimentul aleator

- La aruncarea unei monede avem două rezultate posibile (cap sau pajură). Fiecare rezultate apare cu o probabilite cunoscută (0,5).
  - Testul: aruncarea monedei
  - Evenimentul: rezultatul testului
  - Evenimentul aleatoriu: rezultatul opţinut la efectuarea unui test
  - Spaţiul fundamental: {cap, pajură}

Definiţii

- Spaţiul fundamental = toate rezultatele posibile asociate unui test
  - Într-o familie cu 2 copii, spaţiul fundamental al evenimentului genul compiilor este : S = {FF, FM, MF, MM}
- Distribuţia de probabilitate: toate rezultatele posibile cu probabilităţile asociate

- Aruncarea unei monede o singură dată (cap = C, pajură = P)
  - S = {C, P}
  - P(C) = 0,5
  - P(P) = 0,5

- Aruncarea unei moneyi de două ori (cap = C, pajură = P)
  - S = {CC, CP, PC, PP}
  - P(CC) = 0,25
  - P(CP) = 0,25
  - P(PC) = 0,25
  - P(PP) = 0,25

» Evenimente complementare = două evenimente mutual exclusive a căror sumă de probabilităţi este egală cu 1

» S = {C, P}: P(C)+P(P) = 0,5+0,5 = 1

» S = {CC, CP, PC, PP}: P(CC)+P(nonCC) = 0,25+0,75 = 1

» Evenimente disjuncte: spaţiul evenimentului poate avea mai mult de 2 rezultate posibile

» Evenimente complementare: spaţiul evenimentului poate avea doar 2 rezultate posibile

Definiții

- Probabilitatea = o măsură a şansei de realizare a unui eveniment
  - P(A) $\in$ [0, 1]
  - $0 \leq Pr \leq 1$
- Fie A un eveniment:
  - P(A) = probabilitatea evenimentului A
  - Dacă evenimentul este o certitudine: P(A) = 1
  - Dacă evenimentul este imposibil de realizat: P(A) = 0
- Dacă un eveniment A se poate realiza în S probe dintr-o serie de n încercări echiprobabile, atunci probabilitatea evenimentului A este dată de numărul de cazuri favorabile raportat la numărul de cazuri posibile:

$$P(A) = (\text{nr cazuri favorabile})/(\text{nr. cazuri posibile})$$

» La aruncarea unei monede de 10 ori, obţinem de fiecare dată cap (C). Care este probabilitatea ca la următoarea aruncare să obţinem tot cap?

> 0,5

> < 0,5

> > 0,5

C, C, C, C, C, C, C, C, C, C, ?

P(cap la prima aruncare) = P(cap la a zecea aruncare) = 0,5

- Şansele sunt probabilităţi exprimate procentual
- Şansa ia valori între 0% şi 100%
  - O probabilitate de 0,75 este egală cu o şansă de 75%
- Raţia unui eveniment este probabilitatea ca un eveniment să se întâmple împărţit la probabilitatea ca acel eveniment să nu se întâmple
  - Poate lua orice valoare pozitivă
  - Fie A evenimentul de interes. Raţia de probabilitate = P(A)/[1-Pr(A)] (unde 1-P(A) = Pr(nonA))
  - Exemplu: dacă P(A) = 0,75 atunci raţia de probabilitate este de 3 la 1 (0,75/(1-0,75)=0,75/0,25=3/1)

PROBABILITATE vs. ȘANSĂ

- S-a realizat un studiu pe un eşantion de 7782 subiecţi din ţări şi s-au obţinut următoarele rezultate:
  - 36,2% din populaţia lumii au fost de acord cu următoarea propoziţie "Bărbaţii ar trebui să aibă mai mult dreptul la un loc de muncă decât femeile."
  - 13,8% din persoanele incluse în studiu aveau studii universitare
  - 3,6% din persoanele incluse în studiu îndeplineau simultan cele două criterii.

$$P(A) = 0{,}362 \text{ - } P(SU) = 0{,}138 \text{ - } P(A \ \& \ SU) = 0{,}036$$

- Evenimnetele "acord" (A) şi "studii superioare" (SU) sunt independente?
- $P(A \text{ şi } SU) = 0{,}036 \neq 0 \rightarrow$ dependente

Sursa: http://www.worldvaluessurvey.org/

P(A) = 0,362

P(UD) = 0,138

P(A & UD) = 0,036

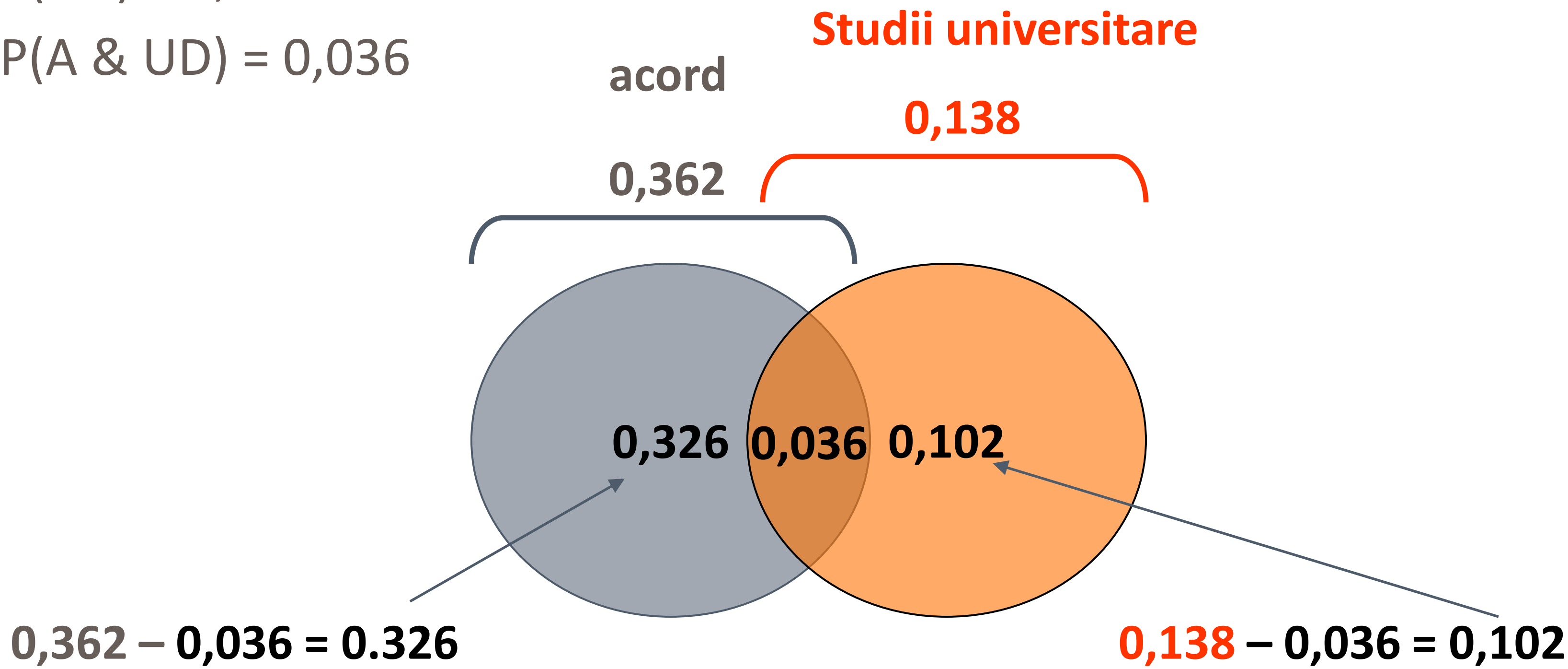


Diagrama Venn

$\cup$(reuniune) = SAU

$\cap$ (intersecţie) = ŞI

$P(A) = 0,362$

$P(SU) = 0,138$

$P(A \cap SU) = 0,036$

Care este probabilitatea ca o persoană extrasă la întâmplare să aibă studii universitare sau să fie de acord?

$P(A \cup SU) = P(A) + P(SU) – Pr(A \cap SU)$

$P(A \cup SU) = 0,326 + 0,136 – 0,036 = 0,464$

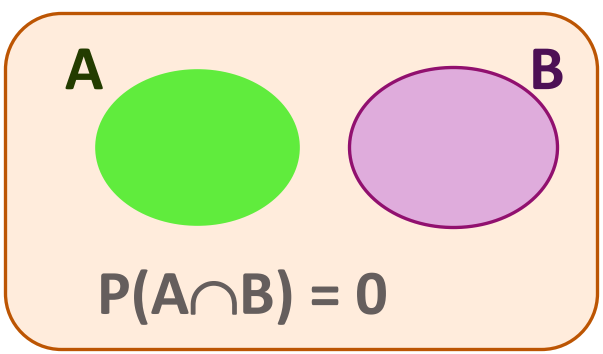
Regula generală de adunare:

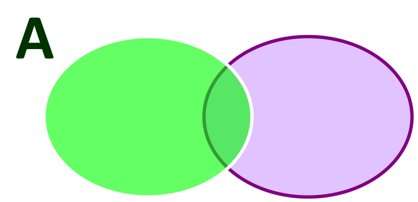
$$P(A \cup B) = P(A) + P(B) – P(A \cap B)$$

Evenimente mutual exclusive $P(A \cap B) = 0$

Evenimente mutual exclusive

- Evenimente mutual exclusive = evenimente care nu pot avea loc simultan
  - Rezultatul obţinut la aruncarea unei monede nu poate fi în acelaşi timp şi cap şi pajură
  - Un student nu poate în acelaşi timp să treacă şi să pice un examen
  - O singură carte extrasă dintr-un pachet de cărţi nu poate în acelaşi timp să fie şi 3 şi regină



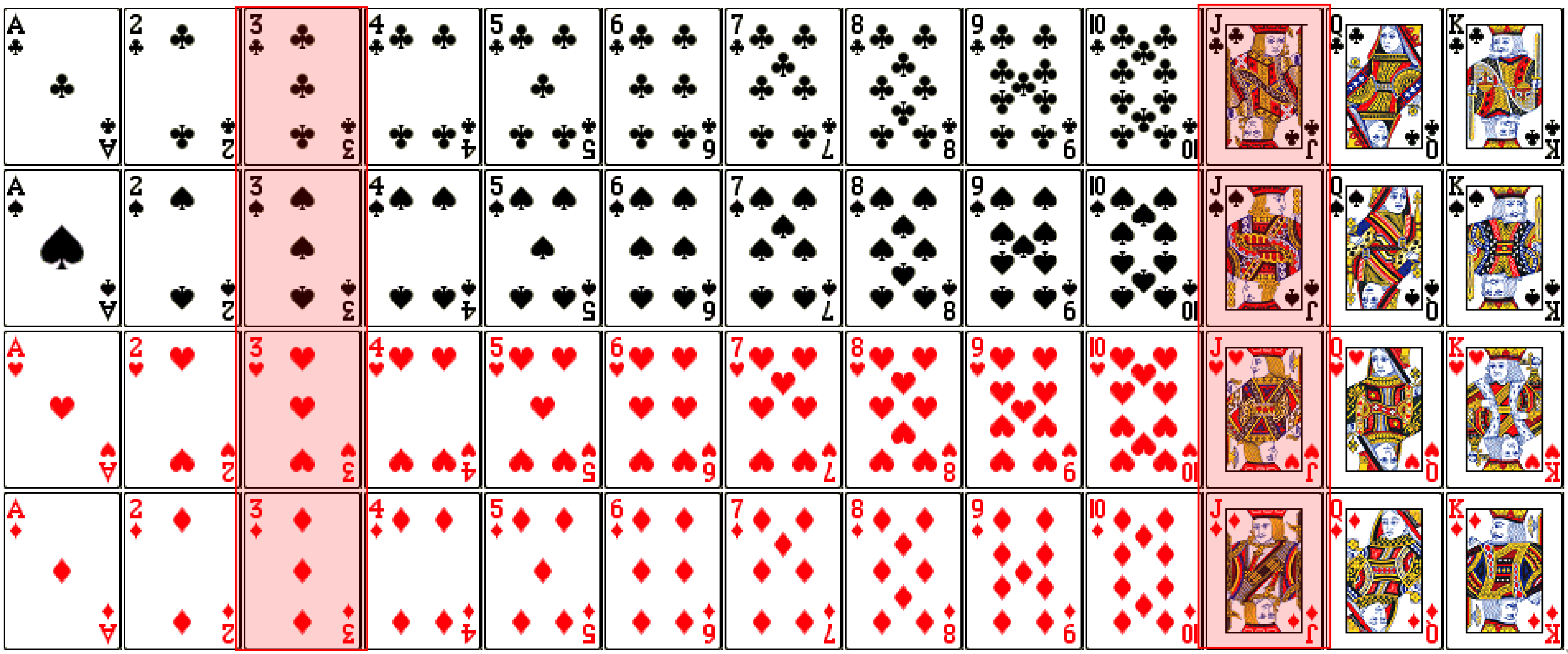


P(A∩B) ≠ 0

Care este probabilitatea de a extrage dintr-un pachet de cărţi de joc bine amestecat un J sau un 3?

Regula generală de adunare:
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
Evenimente mutual exclusive $P(A \cap B) = 0$

$P(J \cup 3) = P(J) + P(3) = 4/52 + 4/52 = 0,1538$



Evenimente mutual exclusive

$\cup$ **(reuniune) = SAU**
$\cap$ **(intersecție) = ȘI**

$P(A) = 0,362$

$P(SU) = 0,138$

$P(A \cap SU) = 0,036$

» Care este probabilitate ca o persoană aleasă aleator să nu aibă studii universitare sau să nu fie de acord cu afirmația...?

» $P(nonA \cap nonSU) = 1 - P(A \cup SU)$

» $P(nonA \cap nonSU) = 1 – 0,464 = 0,536$

Operații

$P(A) = 0,362$

$P(SU) = 0,138$

$P(A \cap SU) = 0,036$

∪ **(reuniune) = SAU**

∩ **(intersecție) = ȘI**

» Evenimentul reprezentat de existenţa studiilor superioare este independent faţă de evenimentul reprezentat de acordul că bărbaţii ar trebui să aibă mai mult dreptul la un loc de muncă decât femeile?

Produsul a două evenimente independente:
$P(A \cap B) = P(A) \times P(B)$

$P(A \cap SU) = P(A) \times P(SU)$

$0,036 = 0,362 + 0,138 \rightarrow$

$0,036 \neq 0,05 \rightarrow$ evenimentele nu sunt independente

Operații

$\cup$ **(reuniune) = SAU**
$\cap$ **(intersecție) = ȘI**

P(A) = 0,362

P(SU) = 0,138

P(A$\cap$SU) = 0,036

» Care este probabilitatea ca cel puţin unul din 5 persoane selectate la întâmplare să fie de acord cu propoziţia Bărbaţii ar trebui să aibă mai mult dreptul la un loc de muncă decât femeile?

» P(A) = 0,362 → S = {0, 1, 2, 3, 4, 5} → S = {0, cel puţin 1}

» P(cel puţin 1 din 5 să fie de acord)

= 1 – P(nonA) ← P(nonA) = 1-P(acord) = 1-0,362 = 0,638

= 1 – P(nonA, nonA, nonA, nonA, nonA)

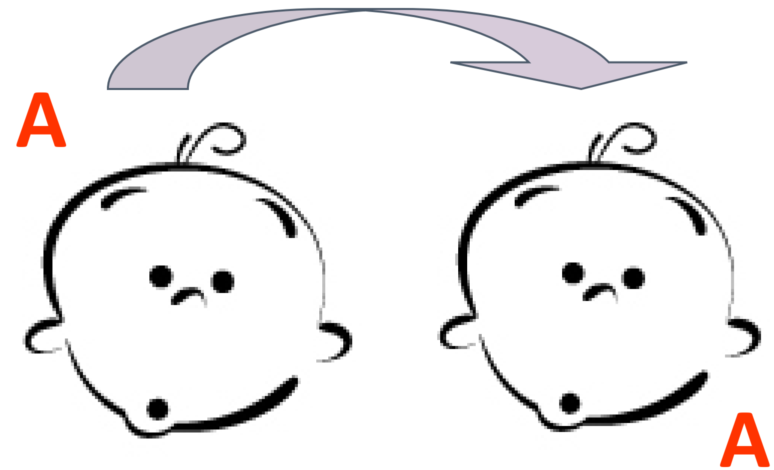
= 1 – $0,638^5$

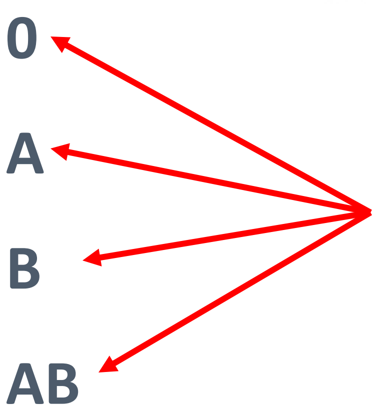
= 1 – 0,106 = 0,894

Operații

Evenimente independente

- Două evenimente sunt independente dacă cunoaşterea rezultatului unui eveniment nu aduce nici o informaţie cu privire la rezultatul celui de-al doilea eveniment
- P(A|B) = P(A)





Disjuncte

&

independente

Dependente / Independente

Sunt fraţi?

Goodman E, Amick BC, Rezendes MO, Levine S, Kagan J, Rogers WH, Tarlov AR. Adolescents' understanding of social class: a comparison of white upper middle class and working class youth. J Adolesc Health. 2000;27(2):80-3.

- S-a investigat opinia unui grup de adolescenți de 16 ani cu privire la statusul social la care consideră că aparţi (48 muncitori & 50 clasa marginală)
- Evaluarea:
  - Ocupaţia părinţilor (evaluare obiectivă)
  - Chestionar pentru copii (evaluare subiectiv[])

	Obiectiv		Total
Subiectiv	Muncitor	Marginal	
Scăzut	0	0	0
Muncitori	8	0	8
Clasa de mijloc	32	13	45
Marginal	8	37	45
Crescut	0	0	0
Total	48	50	98

Probabilităţi condiţionate

	Obiectiv		Total
Subiectiv	Muncitor	Marginal	
Scăzut	0	0	0
Muncitori	8	0	8
Clasa de mijloc	32	13	45
Marginal	8	37	45
Crescut	0	0	0
Total	48	50	98

- Care este probabilitatea clasei marginale la evaluarea obiectivă?
- P(ObiMar) = 50/98 = 0,51

Probabilitatea marginală

	Obiectiv		Total
Subiectiv	Muncitor	Marginal	
Scăzut	0	0	0
Muncitori	8	0	8
Clasa de mijloc	32	13	45
Marginal	8	37	45
Crescut	0	0	0
Total	48	50	98

- Care e probabilitatea clasei marginale (Mar) atât la clasificarea subiectivă (Sub) cât și la cea obiectivă (Obi)?
- P(SubMae) = 45/98
- P(ObjMar) = 50/98
- $P(SubMar \cap ObjMar) = 37/98 = 0,38$

Probabilități condiționate

Sorana D. BOLBOACĂ

	Obiectiv		Total
Subiectiv	Muncitor	Marginal	
Scăzut	0	0	0
Muncitori	8	0	8
Clasa de mijloc	32	13	45
Marginal	8	37	45
Crescut	0	0	0
Total	48	50	98

- Care este probabilitatea ca un adolescent din clasa muncitor la evaluarea obiectivă să fie clasificat în clasa marginal la evaluarea subietivă?
- P(SubMar) = 45/98
- P(ObjMun) = 48/98
- P(SubMar|ObjMun) = 8/48 = 0,17

Probabilități condiționate

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

» Fie A şi B două evenimente

» Prin probabilitatea condiţionată a lui A de către B (simbol: P(A|B)) se înţelege probabilitatea de a se realiza evenimentul A dacă în prealabil s-a realizat evenimentul B

» Exemplu:

> P(Test pozitiv tuberculină|TBC) este probabilitatea de a obţine un test pozitiv la tuberculină la un pacient care are TBC.

» P(B|A) nu este acelaşi lucru cu P(A|B)

» **P(B|A) ≠ P(A|B)**

Probabilități condiționate

Riscul Relativ (RR)

» Raportul dintre P(B|A) și P(B|nonA)

unde B = boala, A = factorul de risc

» Pentru două evenimente independente:

  > $P(B|A) = P(B) = P(B|nonA)$ → RR = 1

» Pentru două evenimente dependente:

  » $P(B|A) \neq P(B) \neq P(B|nonA)$ → RR ≠ 1

Riscul Relativ (RR)

- Mamografia în depistarea cancerului de sân
  - Din 10 femei cu mamografie pozitivă 1 va dezvoltat cancer de sân la 2 ani după mamografie
  - Din 100000 femei cu mamografie negativă 20 vor dezvoltat cancer de sân la 2 ani după mamografie
- Care este RR de a dezvolta cancer de sân la 2 ani după o mamografie pozitivă?
- B = cancer de sân; A = mamografie pozitivă
- $RR = P(B|A)/P(B|nonA)$

Rata șansei (OR)

PMC full text: Int J Cancer. Author manuscript; available in PMC 2013 Sep 1.
Published in final edited form as:
Int J Cancer. 2012 Sep 1; 131(5): 1210–1219.
Published online 2011 Dec 14. doi: 10.1002/ijc.27339
Copyright/License ► Request permission to reuse

Table 2

Relative risk of lung cancer in current smokers by duration, average intensity, and pack-years of smoking cigarettes.

				All histologies		Squamous cell cancer		Small cell lung cancer		Adenocarcinoma	
			Controls	Cases	OR[a] (95% CI)	Cases	OR[a] (95% CI)	Cases	OR[a] (95% CI)	Cases	OR[a] (95% CI)
Duration (years)	Men	1-<20	190	139	7.4 (5.5-9.8)	49	12.5 (7.9-19.6)	43	22.0 (12.3-39.2)	25	2.5 (1.5-4.2)
		20-<30	552	739	13.0 (10.6-15.9)	286	23.2 (16.5-32.5)	163	26.6 (16.4-43.1)	159	5.5 (4.0-7.5)
		30-<40	1106	1959	21.4 (18.1-25.3)	886	41.0 (30.3-55.5)	391	43.1 (27.6-67.3)	380	9.1 (7.1-11.7)
		40-<50	1341	2662	29.9 (25.5-35.2)	1212	57.0 (42.3-76.8)	463	61.9 (39.7-96.7)	570	15.0 (11.8-19.0)
		50-<60	602	1213	32.8 (27.2-39.7)	572	67.0 (48.5-92.5)	179	61.5 (37.9-100.0)	248	15.4 (11.5-20.5)
		≥60	34	71	26.9 (16.9-43.0)	33	57.6 (31.6-105.1)	9	49.9 (19.3-129.1)	16	14.7 (7.4-29.3)
	Women	1-<20	85	79	2.8 (1.9-3.9)	16	4.6 (2.4-8.5)	20	7.5 (4.0-14.0)	26	1.5 (0.9-2.5)
		20-<30	181	285	4.9 (3.8-6.2)	51	7.2 (4.7-11.1)	86	16.0 (10.2-25.2)	93	2.8 (2.0-3.8)
		30-<40	215	473	8.6 (7.0-10.5)	94	11.1 (7.8-15.8)	126	25.7 (17.2-38.4)	155	4.6 (3.5-5.9)
		40-<50	144	409	11.8 (9.4-14.9)	126	22.6 (15.8-32.4)	75	29.6 (18.8-46.7)	136	6.3 (4.8-8.4)
		50-<60	63	165	12.2 (8.7-17.1)	52	27.7 (16.7-45.8)	23	33.6 (17.2-65.6)	51	6.2 (4.0-9.4)
		≥60	3	10	20.2 (5.1-80.2)	9	37.3 (6.7-207.1)	1	49.0 (2.4-995.0)	1	4.2 (0.4-45.4)

Author Manuscript

	Boală+	Boală-	Total
Test+	AP	FP	= AP+FP
Test-	FN	AN	= FN+AN
Total	= AP+FN	=FP+AN	= n

Denumire parametru	Probabilitate	Formula
Rata falşilor pozitivi		=FP/(FP+AN)
Rata falşilor negativi		=FN/(FN+AP)
Sensibilitatea	=P(T+|B+)	=AP/(AP+FN)
Specificitatea	=P(T-|B-)	=AN/(AN+FP)
Acurateţea		=(AP+AN)/n
Valoarea predictivă pozitivă	=P(B+|T+)	=AP/(AP+FP)
Valoarea predictivă negativă	=P(B-|T-)	=AN/(AN+FN)
Riscul relativ		=AP(FP+AN)/FN(AP+FP)
Rata şansei		=(AP·AN)/(FN·FP)
Riscul atribuabil		=AP/(AP+FP)-FN/(FN+AN)

Probabilități în tabelul de contingență 2×2

Sorana D. BOLBOACĂ

Testul diagnostic

	Probabilitate	Formula
Se	=Pr(T+|B+)	=AP/(AP+FN)
Sp	=Pr(T-|B-)	=AN/(AN+FP)
VPP	=Pr(B+|T+)	=AP/(AP+FP)
VPN	=Pr(B-|T-)	=AN/(AN+FN)

	Boală+	Boală-	Total
Test+	AP	FP	= AP+FP
Test-	FN	AN	= FN+AN
Total	= AP+FN	=FP+AN	= n

VPP

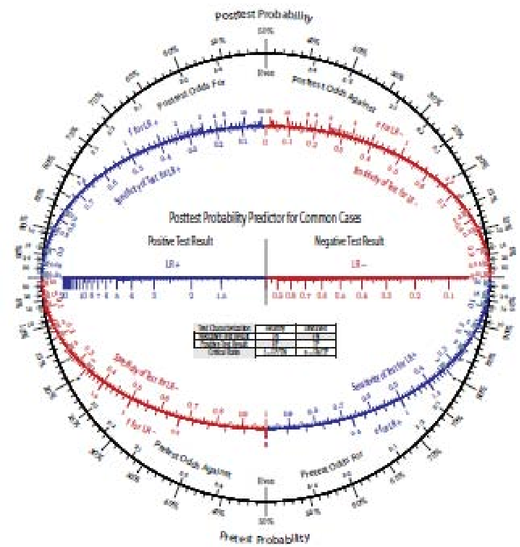
VPN

Se

Sp

Diagnostic

Este vorba de probabilitatea



Regula Bayer

Probabilități

P_{old} + Test = P_{new}

Nomograma

Tehnică grafică de aplicare a teoretiei probabilităților

Teorema lui BAYES

- Un test diagnostic nu este un eveniment
- Testul detectează ceea ce nu există (rezultate fals pozitive) și poate să nu detecteze ceea ce există (rezultate fals negative)
- Un test diagnostic ne dă o probabilitate a testului care nu este identică cu probabilitatea reală
- Persoanele preferă frecvențe absolute nu relative (100 din 10.000 NU 1%)

	Cancer mamar+ (2%)	Cancer mamar- (98%)
Mammo +	70	10
Mammo-	30	90

- 2% din femei au cancer mamar
- 70% din mamografii detectează prezența cancerului mamar atunci când acesta este prezent → 30% dă rezultate fals negative
- 10% din mamografii detectează prezența cancerului atunci când acesta nu există (rezultate fals pozitive) → 90% din mamografii returnează un diagnostic negativ corect

	Cancer mamar+ (2%)	Cancer mamar- (98%)
Mammo +	70	10
Mammo-	30	90

- 2% din femei au cancer de sân
- O femeie cu cancer de sân are șansa de 70% de a avea un rezultat mamografic pozitiv și 30% șansa de a avea un rezultat mamografic negativ
- O femeie care nu prezintă cancer de sân are o șansă de 10% de a avea un rezultat mamografic pozitiv și 90% de a avea un rezultat mamografic negativ

- Aveți o pacientă cu rezultat mamografic pozitiv. Care este șansa acesteia de a avea cancer mamar?

	Cancer mamar+ (2%)	Cancer mamar - (98%)
Mammo +	70	10
Mammo-	30	90

» Probabilitatea adevărat pozitivului = probabilitatea de a avea cancer mamar × probabilitatea de a avea un rezultat mamografic pozitiv = 0,02 × 0,7 = 0,014 → șansa de 1,4%

» Probabilitatea fals pozitivului = probabilitatea de a nu prezenta cancer mamar × probabilitatea unui rezultat mamografic pozitiv = 0,98 × 0,10 = 0,098 → șansa de 9,8%

Anatomia testului diagnostic

	Cancer mamar+ (2%)	Cancer mamar - (98%)
Mammo +	Adevărat pozitiv 0,02*0,70 = 0,014	Faps pozitiv 0,10*0,98=0,098
Mammo-	Fals negativ 0,02*30 = 0,600	Adevărat negativ 0,90*0,98 = 0,882

- Care este șansa ca pacienta să prezinte cancer mamar dacă are un rezultat mamografic pozitiv?
  - Probabilitatea = (evenimentul de interes)/(toate posibilitățile)
  - = 0,014/(0,014+0,098) = 0,125 → șansa = 12,5%

Anatomia testului diagnostic

- → un rezultat mamografic pozitiv ne spune doar că pacienta are șansa de a avea cancer mamar de 12,5% (comparat cu valoarea expectată de 70%)
- → în 10% din cazuri mamografia dă rezultate fals pozitive → vor exista rezultate pozitive în orice populație
- → această problemă poate fi rezolvată cu ajutorul teoremei Bayes

Adunare:

» $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

» $P(A \cup B) = P(A) + P(B)$: **evenimente mutual exclusive**

Înmulțire:

» $P(A \cap B) = P(A) \cdot P(B|A)$

» $P(A \cap B) = P(A) \cdot P(B)$: **evenimente independente**

RECALL!