

## LP 6 Rezumat

### 3.10. Teste binare

Subiect de teoria probabilităților aplicată.

Sn **test binar** un procedeu care vizează identificarea unităților dintr-o clasă  $X$  față de unitățile din clasa complementară ( $NX$ ) producând un rezultat *binar* (+ ori -).

- Produce 4 clase:  $ap$  = adevărat pozitiv     $fn$  = fals negativi  
 $fp$  = fals pozitiv     $an$  = adevărat negativi

#### 3.10.1. Caracteristicile unui test binar

1° Sensibilitatea  $Se = ap / X$

2° Specificitatea  $Sp = an / NX$

3° Observații

- $Se$ ,  $Sp$ , *proporția de fp* și cea *de fn* - cuprinse în (0, 1). Sunt probabilități.
- $Se = 1 - \text{proporția de fn în } X$ .
- $Sp = 1 - \text{proporția de fp în } NX$ .
- O înaltă  $Se$  - necesară când  $X$  este o boală infecțioasă sau o boală care trebuie tratată cât mai din timp.
- O înaltă  $Sp$  - necesară când tratamentele indicate sunt foarte nocive.

#### 3.10.2. Interpretarea rezultatelor unui test binar

Valoarea Predictivă a rezultatului Pozitiv,  $VPP = ap / +$ .

Valoarea Predictivă a rezultatului Negativ,  $VPN = an / -$ .

#### 3.10.3. Dependența valorilor predictive de caracteristicile testului și de prevalența clasei $X$

Prevalența clasei  $X$ ,  $Pre = X / (X + NX)$ .

$$VPP = \frac{Se \cdot Pre}{Se \cdot Pre + (1 - Sp) \cdot (1 - Pre)}$$

$$VPN = \frac{Sp \cdot (1 - Pre)}{Sp \cdot (1 - Pre) + (1 - Se) \cdot Pre}$$

- $VPP$  crește o dată cu (creșterea)  $Sp$ ,
- $VPN$  crește o dată cu (creșterea)  $Se$ .

Pentru același test:

- $VPP$  crește cu creșterea prevalenței clasei  $X$  (de exemplu a bolii),
- $VPN$  scade cu creșterea prevalenței clasei  $X$  (de exemplu a bolii).

În cazul unei boli rare ( $Pre < 1\%$ ) rezultatul pozitiv nu este concludent ( $VPP \ll 1$ ).

### 3.10.4. Interpretarea globală a rezultatelor a două sau mai multe teste independente

- Testul denumit **conjuncția testelor** și notat  $T_1 \cdot T_2$  (sau, mai simplu,  $T_C$ ) care este pozitiv (prin definiție)  $\Leftrightarrow$  ambele teste sunt pozitive și
- testul denumit **disjuncția testelor** și notat  $T_1 + T_2$  (sau, mai simplu,  $T_D$ ) care este negativ (prin definiție)  $\Leftrightarrow$  ambele teste sunt negative.

$$T_C \text{ are } Se_C = Se_1 \cdot Se_2 \qquad Sp_C = Sp_1 + Sp_2 - Sp_1 \cdot Sp_2$$

$$T_D \text{ are } Se_D = Se_1 + Se_2 - Se_1 \cdot Se_2 \qquad Sp_D = Sp_1 \cdot Sp_2$$

În general:

- mai multe plusuri cresc probabilitatea de a fi din clasa  $X$  și
- mai multe minusuri cresc probabilitatea de a nu fi din clasa  $X$ .

Singura cale de a reduce simultan procentele ambelor tipuri de erori este aplicarea mai multor teste. «În cazul unei afecțiuni rare ( $Pre < 1\%$ ) testul este concludent pentru rezultatul negativ, dar trebuie întotdeauna repetat pentru rezultat pozitiv.» [12]

### 3.10.5. Construcția testelor binare unidimensionale

Problemă de analiză discriminantă pe distribuții teoretice.

- $Se \uparrow \Leftrightarrow Sp \downarrow \Leftrightarrow$  dependente în mod invers.

### 3.10.6. Tipuri de teste binare unidimensionale

După modul de plasare a zonei de respingere:

- *teste bilaterale*
- *teste unilaterale*. După lateralitatea plasării:
  - *teste unilaterale stânga*
  - *teste unilaterale dreapta*.

## LP 6 Teste, exerciții și probleme

### TG6. Durata 200'' pe calculator.

Fie un individ care are a priori probabilitatea  $p$  de a avea boala  $B$  și un test pentru detectarea acestei boli care aplicat individului da rezultat pozitiv.

Alegeți afirmația corectă:

1. Probabilitatea devine 1.
2. Probabilitatea scade, dar nu devine 0.
3. Crește probabilitatea de a avea boala  $B$  la o valoare subunitară.

Un test binar reprezintă:

1. un proces care vizează identificarea unităților dintr-o clasă  $X$  față de unitățile din clasa complementară  $NX$ , producând un rezultat binar
2. un procedeu care vizează identificarea unităților dintr-o clasă  $X$  față de unitățile din clasa complementară  $NX$ , producând un rezultat binar
3. un proces care vizează identificarea unităților dintr-o clasă față de unitățile dintr-o altă clasă, producând un rezultat binar

Specificitatea unui test binar reprezintă:

1. proporția de adevărat negativi din totalul de negativi
2. proporția de adevărat negativi din clasa  $NX$
3. proporția de adevărat negativi din clasa  $X$

Când se vizează identificarea a aproape tuturor cazurilor din clasa  $X$ , este nevoie de:

1. o înaltă specificitate
2. o înaltă sensibilitate și o înaltă specificitate
3. o înaltă sensibilitate

Valoarea predictivă a rezultatului pozitiv este egală cu:

1. proporția de adevărat pozitivi din totalul de pozitivi
2. proporția de adevărat pozitivi din clasa  $X$
3. proporția de adevărat pozitivi din clasa  $NX$

Prevalența este o caracteristică:

1. a bolii în populație

2. a populației
3. a bolii și a testului pentru aceasta

Alegeți afirmația corectă:

1. VPP crește o dată cu sensibilitatea, iar VPN crește o dată cu specificitatea
2. VPP scade o dată cu specificitatea, iar VPN crește o dată cu sensibilitatea
3. VPP crește o dată cu specificitatea, iar VPN crește o dată cu sensibilitatea

Valorile predictive ale unui test depind de:

1. sensibilitatea și prevalența clasei X
2. sensibilitatea și specificitatea testului
3. specificitatea și prevalența clasei X
4. sensibilitatea, specificitatea și prevalența clasei X

## TC6. Durata 3'.

1. Sensibilitatea este proporția de \_\_\_\_\_ din \_\_\_\_\_.
2. Specificitatea este complementul față de 1 al proporției de \_\_\_\_\_.
3. Pentru identificarea aproape în totalitate a cazurilor din  $X$  este nevoie de o înaltă \_\_\_\_\_.
4. Pentru detectarea cancerului și evitarea iradierii oamenilor sănătoși este nevoie de un test înalt \_\_\_\_\_.
5. VPN este proporția de \_\_\_\_\_ din \_\_\_\_\_.
6. VPP \_\_\_\_\_ cu prevalența.
7. Disjuncția testelor îmbunătățește \_\_\_\_\_.

## Exerciții sau probleme rezolvate

### 1. (Probleme pentru epidemiologi și clinicieni)

În urma aplicării unui test binar pentru diagnosticarea unei boli cu prevalența în populație de 0,1 s-au obținut: proporția de fals negativi 0,15, proporția de fals pozitivi 0,05. Să se determine VPP și VPN.

*Rezolvare:*

$Pre = 0,1$ ;  $Se = ap / B = 1 - fn / B = 1 - 0,15 = 0,85$ ;  $Sp = an / NB = 1 - fp / NB = 1 - 0,05 = 0,95$ .

$$VPP = \frac{Se \cdot Pre}{Se \cdot Pre + (1 - Sp) \cdot (1 - Pre)} = \frac{0,85 \cdot 0,1}{0,85 \cdot 0,1 + (1 - 0,95) \cdot (1 - 0,1)} = 0,65.$$

$$VPN = \frac{Sp \cdot (1 - Pre)}{Sp \cdot (1 - Pre) + (1 - Se) \cdot Pre} = \frac{0,95 \cdot (1 - 0,1)}{0,95 \cdot (1 - 0,1) + (1 - 0,85) \cdot 0,1} = 0,98.$$

## 2.

În vederea diagnosticării unei boli s-au aplicat pe aceeași populație două teste diferite și s-au obținut următoarele rezultate:

| Test a | +  | -   |
|--------|----|-----|
| B      | 80 | 15  |
| NB     | 8  | 179 |

| Test b | +  | -   |
|--------|----|-----|
| B      | 87 | 8   |
| NB     | 15 | 172 |

Să se observe cum variază  $VPP$  și  $VPN$  în funcție de  $Se$  și  $Sp$ , în cazul celor două teste.

*Rezolvare:*

| Test a                          | +  | -                                 | $\Sigma$ |                                   |
|---------------------------------|----|-----------------------------------|----------|-----------------------------------|
| B                               | 80 | 15                                | 95       | $Se = ap / B = 80 / 95 = 0,84$    |
| NB                              | 8  | 179                               | 187      | $Sp = an / NB = 179 / 187 = 0,96$ |
| $\Sigma$                        | 88 | 194                               | 282      |                                   |
| $VPP = ap / + = 80 / 88 = 0,91$ |    | $VPN = an / - = 179 / 194 = 0,92$ |          |                                   |

$$Pre = B / (B+NB) = 95 / 282 = 0,34$$

| Test b                           | +   | -                                 | $\Sigma$ |                                   |
|----------------------------------|-----|-----------------------------------|----------|-----------------------------------|
| B                                | 87  | 8                                 | 95       | $Se = ap / B = 87 / 95 = 0,92$    |
| NB                               | 15  | 172                               | 187      | $Sp = an / NB = 172 / 187 = 0,92$ |
| $\Sigma$                         | 102 | 180                               | 282      |                                   |
| $VPP = ap / + = 87 / 102 = 0,85$ |     | $VPN = an / - = 172 / 180 = 0,96$ |          |                                   |

$$Pre = B / (B+NB) = 95 / 282 = 0,34$$

Se observă că în cazul aplicării a două teste aceleași populații (avem deci aceeași prevalență):

- $VPP$  a scăzut o dată cu  $Sp$ .
- $VPN$  a crescut o dată cu  $Se$ .

## 3.

În urma aplicării unui test binar pentru diagnosticarea unei boli într-o populație de 1000 de persoane s-au obținut următoarele rezultate:

| Test      | +   | -   |
|-----------|-----|-----|
| Bolnav    | 225 | 75  |
| Nonbolnav | 200 | 500 |

- Să se calculeze  $Se$ ,  $Sp$ ,  $VPP$  și  $VPN$ .  
Un cercetător a propus un alt test care ar putea fi folosit pentru diagnosticarea aceleași boli. Acest nou test este aplicat populației anterioare.
- Știind că  $fn = 100$  și  $Sp = 0,75$  să se afle  $Se$ ,  $VPP$  și  $VPN$  pentru noul test.
- Care dintre cele două teste este de preferat dacă boala poate fi tratată cu succes doar în fazele incipiente? De ce?

*Rezolvare:*

| Test              | +                             | -                             |                        |                               |
|-------------------|-------------------------------|-------------------------------|------------------------|-------------------------------|
| <b>Bolnav</b>     | 225                           | 75                            | 300                    | $Se = ap/B = 225/300 = 0,75$  |
| <b>Non-bolnav</b> | 200                           | 500                           | 700                    | $Sp = an/NB = 500/700 = 0,71$ |
|                   | 425                           | 575                           | 1000                   |                               |
|                   | $VPP = ap/+ = 225/425 = 0,53$ | $VPN = an/- = 500/575 = 0,87$ | $Pre = 300/1000 = 0,3$ |                               |

$$2. Se = ap/B = 1 - (fn/300) = 200/300 = 0,67$$

$$VPP = \frac{Se * Pr e}{Se * Pr e + (1 - Sp)(1 - Pr e)} = 0,53$$

$$VPN = \frac{Sp * (1 - Pr e)}{Sp * (1 - Pr e) + (1 - Se) Pr e} = 0,84$$

- Primul test, deoarece are  $Se$  mai mare și deci identifică mai mulți indivizi dintre cei bolnavi decât al doilea test.

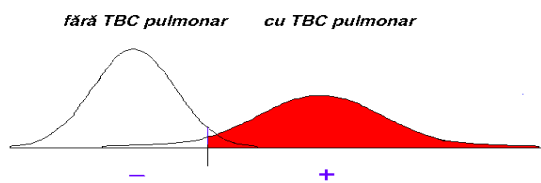
#### 4.

În vederea construirii unui test de depistare a tuberculozei bronhopulmonare s-au construit prin randomizare, cu ajutorul unui biostatistician, două loturi foarte mari. Primul lot conținea pacienți infectați cu bacilul lui Koch, dar care nu au făcut boala, iar al doilea pacienți cu boala confirmată prin examen radiologic. Tuturor pacienților li s-a făcut IDR și li s-a măsurat în mm, la 72 de ore, diametrul maxim.

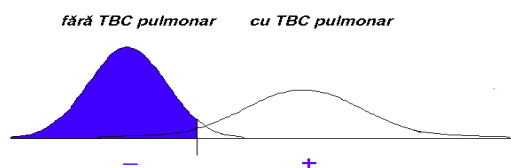
S-au calculat apoi mediile și abaterile standard pentru fiecare lot, obținându-se următoarele valori:

|                   | <b>Medii</b> | <b>Abateri standard</b> |
|-------------------|--------------|-------------------------|
| Pacienți fără TBC | 5,12         | 1,9                     |
| Pacienți cu TBC   | 12,67        | 1,6                     |

- a) Să se explice pe un desen adecvat cum se va calcula limita de discriminare dacă dorim ca testul să aibă sensibilitatea 0,99.



- b) Să se calculeze efectiv limita respectivă de discriminare.  
**L = 9 mm. (z = -2,29)**  
 c) Fiind cunoscută limita de discriminare să se explice, repetând desenul de la punctul a, cum se poate determina specificitatea testului.



- d) Să se determine efectiv specificitatea. pentru limita calculată la punctul b.  
**Sp = 0,98. (z(9)=2,04)**  
 e) Dacă TBC-ul este o boală rară în populația din care provin pacienții noștri ce putem spune despre un rezultat pozitiv ?  
**Este neconcludent.**  
 f) Ce trebuie făcut în acest caz ?  
**Trebuie repetat testul în condiții de independență.**

## 5. (Problemă pentru clinician)

Pentru diagnoza unui pacient suspect de boala X cu prevalența 0,005 în populația din care face parte s-a utilizat un test cu sensibilitatea 0,95 și specificitatea 0,97.

1. Ce veți face dacă rezultatul testului este pozitiv ?
2. Dar dacă este negativ ?
3. Cât de probabil are boala X pacientul dacă rezultatul a fost pozitiv ?
4. Cât de probabil are boala X pacientul dacă rezultatul a fost pozitiv și la repetarea testului.
5. Cât de probabil nu are boala X pacientul dacă rezultatul a fost negativ ?

*Rezolvare:*

1. Se repetă testul, fiind vorba de o boală rară ( $Pre = 0,005 < 0,01$ )
2. Pentru rezultat negativ testul este concludent, fiind boală rară. Se înlătură suspiciunea bolii X.
3.  $VPP = 0,14$
4. Calculăm  $Se$  și  $Sp$  pentru conjuncție:  $Se=0,9$ ;  $Sp = 1$ , apoi  $VPP=1$ .
5.  $VPN = 1$ .

## Exerciții sau probleme propuse

### 6.

*Variația VPP și VPN cu Pre.*

În vederea diagnosticării unei boli s-a aplicat același test pe două populații diferite și s-au obținut următoarele rezultate:

| Test a    | +  | -   |
|-----------|----|-----|
| <i>B</i>  | 80 | 15  |
| <i>NB</i> | 8  | 179 |

| Test b    | +  | -   |
|-----------|----|-----|
| <i>B</i>  | 60 | 11  |
| <i>NB</i> | 8  | 203 |

Să se observe cum variază  $VPP$  și  $VPN$  în funcție de prevalența ( $Pre$  a) bolii în populație.

### 7.

*Determinare VPP și VPN.*

În urma aplicării unui test binar pentru diagnosticarea unei boli cu prevalența în populație de 0,15 s-au obținut: proporția de fals negativi 0,14, proporția de fals pozitivi 0,06. Să se determine  $VPP$  și  $VPN$ .

### 8.

*Combinarea testelor binare.*

Se dau două teste caracterizate de:

$$Se_1 = 0,84; Sp_1 = 0,96$$

$$Se_2 = 0,92; Sp_2 = 0,92$$

Să se compare  $Se$  și  $Sp$  ale testelor conjuncție, respectiv disjuncție cu ale testelor inițiale.

## 9.

Pentru construirea unui test biomedical de identificare a glaucomului s-a măsurat tensiunea intraoculară la două loturi de pacienți. Primul lot a fost format din pacienți cu glaucom, iar al doilea, pacienți fără glaucom. Mediile și abaterile standard ale celor două loturi au fost:

| Lot          | Medii (M) în mm Hg | Abateri standard (S) în mm Hg |
|--------------|--------------------|-------------------------------|
| Cu glaucom   | $M_1 = 29$         | $S_1 = 3$                     |
| Fără glaucom | $M_2 = 20$         | $S_2 = 2$                     |

1. Presupunând că tensiunea intraoculară se distribuie gaussian la cele două loturi cu mediile și abaterile standard specificate, să se figureze pe același desen cele două distribuții.
2. Fixându-se limita de discriminare între cele două categorii la valoarea  $L = 23,5$  mm Hg, să se calculeze care va fi procentul de glaucoame nedepistate de acest test, precum și cel de indivizi fără glaucom care sunt considerați în mod greșit cu glaucom.
3. Care dintre limitele 23 și 24,5 va produce un procent mai scăzut de glaucoame nedepistate?

## 10. (Colocviu-iunie 2002b)

În vederea construirii unui test de depistare a tuberculozei bronhopulmonare s-au construit prin randomizare, cu ajutorul unui biostatistician, două loturi foarte mari. Primul lot conținea pacienți cu boala confirmată prin examen radiologic, iar al doilea pacienți infectați cu bacilul lui Koch, dar care nu au făcut boala. Tuturor pacienților li s-a făcut IDR și li s-a măsurat în mm, la 72 de ore, diametrul maxim.

Pentru simplificarea calculului, presupunem că lotul de bolnavi a fost format din 6 bolnavi iar diametrele măsurate au fost 10, 12, 15, 13, 14, 12, iar lotul de infectați nonbolnavi a fost format din 8 pacienți cu diametrele 6, 2, 3, 4, 8, 6, 5, 7.

Studiindu-se distribuțiile teoretice „din spatele” acestor date experimentale, s-a stabilit, ca limită de discriminare pentru construcția testului, valoarea 9,5 mm.

1. Să se calculeze parametri necesari pentru cele două serii.
2. Să se figureze, în același desen, cele două distribuții teoretice „din spatele” datelor și limita de discriminare propusă pentru test și să marcheze zona de rezultate pozitive, cea de rezultate negative și să se hașureze diferit zonele de fals.
3. Să se calculeze procentele de erori de diagnostic.
4. Să se calculeze caracteristicile testului.
5. Ce trebuie făcut pentru a mări sensibilitatea?

6. Considerând că prevalența TBC pulmonar în populația investigată este de 1%, să se calculeze care este probabilitatea ca un individ cu rezultat pozitiv să aibă boala.
7. Ce trebuie făcut pentru ca să fim mai siguri în cazul anterior?
8. Calculați probabilitatea de a fi bolnav dacă aplicăm soluția propusă la punctul anterior.
9. Care ar fi fost procentele de fals dacă în locul populațiilor teoretice am fi considerat doar loturile noastre?