Cursul 2

Șl. Dr. Lavinia Moleriu

Noţiuni introductive de Biostatistică

Prelucrarea şi descrierea datelor biomedicale

Rolul biostatisticii...

Prelucrarea şi descrierea datelor biomedicale

BIOSTATISTICA

– De ce este importantă statistica?

– De ce este esenţială în lumea medicală?

– Ce trebuie să ştim pentru a înţelege biostatistica?

Prelucrarea şi descrierea datelor biomedicale

BIOSTATISTICA

- Statistica elementară:
  - Inferenta statistica
  - Statistica descriptiva - Parametrii statistici
  - Estimarea statistica
  - Teste statistice
  - Corelatia si regresia

Prelucrarea şi descrierea datelor biomedicale

BIOSTATISTICA

- *Aplicaţii speciale:*
  - **Epidemiologie (Analiza riscului)**
  - **Aplicaţii în demografie (Analiza supravietuirii)**
  - **Sănătate publică (Statistica sanitară)**
- *Prelucrări avansate*
  - **Analiza clasificării etc**

BIOSTATISTICA

Concepte generale

- Statistica = gruparea, analiza şi interpretarea datelor referitoare la fenomenele de masa

- Biostatistica = ramura a statisticii, care studiază, interpretează şi analizează datele din domeniul biologic, inclusiv cel medical

- **Principala caracteristică a proceselor biologice :**
  - *Variabilitatea -> presupune un anumit grad de incertitudine*

- => **Statistica** ne permite să stabilim „reguli” cu care să ţinem seama de incertitudine

BIOSTATISTICA

Concepte generale

- **Biostatistica** = ştiinţa estimării caracteristicilor unei populatii şi/sau compararea populatiilor
- **Populaţie** = totalitatea (multimea) persoanelor la care se referă studiul; anasamblul indivizilor integraţi în studiu, având cel putin o proprietate comună
- **Individ** = unitate statistică, se poate identifica concret şi i se poate atribui un moment de timp şi o locaţie ; un element din populație
  - În cadrul unei populaţii indivizii işi pierd individualitatea
- **Volumul populaţiei** = numărul indivizilor din populaţie

BIOSTATISTICA

Metode de studiu – evaluarea populaţiilor

- **Recensamânt**
  - **metoda exactă**
  - Situare bine definită în timp şi spaţiu, dar foarte costisitoare
- **Screening**
  - populaţii cu abateri deosebite ale unui parametru
  - nu este necesara localizarea în timp şi este relativ costisitoare
- **Selectie (esantionare)**
  - alegerea unei submultini din populatie – esantion
  - măsuratorile se realizează doar la nivel de eşantion şi costuri sunt reduse

BIOSTATISTICA

Concepte generale

- **Eşantion** = o submulţime din populaţie
- **Eşantion reprezentativ** = eșantion care cuprinde toate straturile populației în proporții similare și are toate caracteristicile populației; eșantionul care conţine proporţional indivizi reprezentând toate caracteristicile populaţiei
- **Criterii de selectie în cadrul unui eșantion reprezentativ:**
  - echiprobabilitate
  - independență

BIOSTATISTICA

- **Metode de selectie în cadrul unui eşantion reprezentativ**
  - sondajul simplu
  - sondajul dirijat
  - sondajul mixt

BIOSTATISTICA

Concepte generale

– **BIAS** = orice condiție care influențează procesul de selecție (influențează reprezentativitatea eșantionului**)**

- **eşantioane neselective** (“unbiased”) - care respectă echiprobabilitatea

- **eşantioane selective** (“biased”) în care, cu sau fără ştirea noastră, un factor a influenţat componenţa lotului.

OBS! În marea majoritate a cazurilor eşantioanele selective sunt evitate; excepţie fac studiile din “analiza riscului” unde aceşti factori sunt chiar căutaţi pentru eventuala definire a unor “straturi”.

BIOSTATISTICA

Concepte generale

- **Statistica descriptivă**
  - **Descrierea variabilelor (calcularea indicatorilor statistici descriptivi, reprezentări grafice, tabele de frecvente, etc.)**
- **Statistica inferentială**
  - **Operația de generalizare, la nivelul întregii populații a caracteristicilor obținute pe un eșantion**
    - ***estimarea parametrilor distributiei unei populatii***
    - ***testarea ipotezelor statistice***

BIOSTATISTICA

- ***VARIABILĂ=*** o caracteristica a populatiei care este studiata si masurata la toți indivizii din esantion
- ***CLASIFICARE:***
  - *NUMERICE*
    - măsurate cu un instrument
    - se precizează unitatea de măsură
  - *ORDINALE (rang)*
    - scară convenţională, fără unităţi
    - exprimă ordinea (ranguri)
  - *NOMINALE (calitative, "count data")*
    - nu există un număr de ordine între variabile (nu se pot compara)

Indicatorii Statistici

Indicatorii Tendinței Centrale

Indicatori de Dispersie

Indicatorii tendinţei centrale:

- **Media**
- **Mediana**
- **Moda**
- Alţi indicator: media geometrică, media armonică

Media

- **MEDIA ARITMETICA (MEAN):**

$$\overline{X} = \left(\sum_{i=1}^{N} X_i \right) / N$$

Mediana

- ***MEDIANA (MEDIAN)*** **= valoarea care împarte lotul de valori ordonateîn două subintervale egale ca număr de valori**
  - ***Ex: pentru un numar par sau impar de elemente***

- **Recomandată pentru variabilele ordinale**

Moda

- ***MODA (MODE)*** **= valoarea cea mai frecventă**
  - **distribuţii uni~, bi~, multimodale**
- **Recomandate pentru variable nominale**

Indicatorii de dispersie:

– **Domeniul de valori (R)**

– **Deviaţia standard(s)**

– **Varianţa, dispersia (s^2)**

– **Eroarea standard (sx)**

Definiții

- **Domeniul de valori (R)=** Xmax-Xmin
- **Deviaţia standard(s) =** gradul de împrăștiere al valorilor individuale în jurul mediei eșantionului
- **Varianța, dispersia (s^2)=** pătratul deviației standard
- **Coeficientul de variaţie (cv)=** raportul dintre deviația standard si media aritmetica

Definiții

- **Abatere centrală=** distanța unei valori individuale fața de valoarea medie
- **Eroarea standard (a mediei)=** arată gradul de împrăștiere a mediilor eșantioanelor în jurul mediei populației
- **Alți indicatori de dispersie: cuantile=** valorile care împart eșantionul in n subclase echinumerice
  - ***Ex:*** *4 clase = cuartile*
    - *10 clase = decile, ... Centile, promile*

Formule

$$s = \sqrt{\frac{\sum \left(x_i - \bar{x}\right)^2}{n-1}}$$

$$cv = s / \bar{x}$$

Formule

$$s^2 = \frac{\sum_{i=1}^{N} (X_i - \overline{X})^2}{N}$$

$$S_{\overline{X}} = \frac{S}{\sqrt{N}}$$

Distribuţia unei variabile

- **Descrierea distributiilor**
  - Tabelara
  - Grafica

- **Caracterizarea unei distributii**
  - Tendinta centrala
  - Imprastierea
  - Forma distributiei (simetria - asimetrie, aplatizare)

Functia de distributie

- Functia ce exprima probabilitatea de aparitie a valorii unei variabile independente intr-o populatie

Distribuţii uzuale

- Distributii continue
  - Distributia normala Gauss
  - Distributia t Student
  - Distributia chi-square
  - Distributia F

- Distributii discrete
  - Distributia binomiala (variabile nominale)
  - Distributia Poisson
  - Distributia hypergeometrica

Distribuţia normală

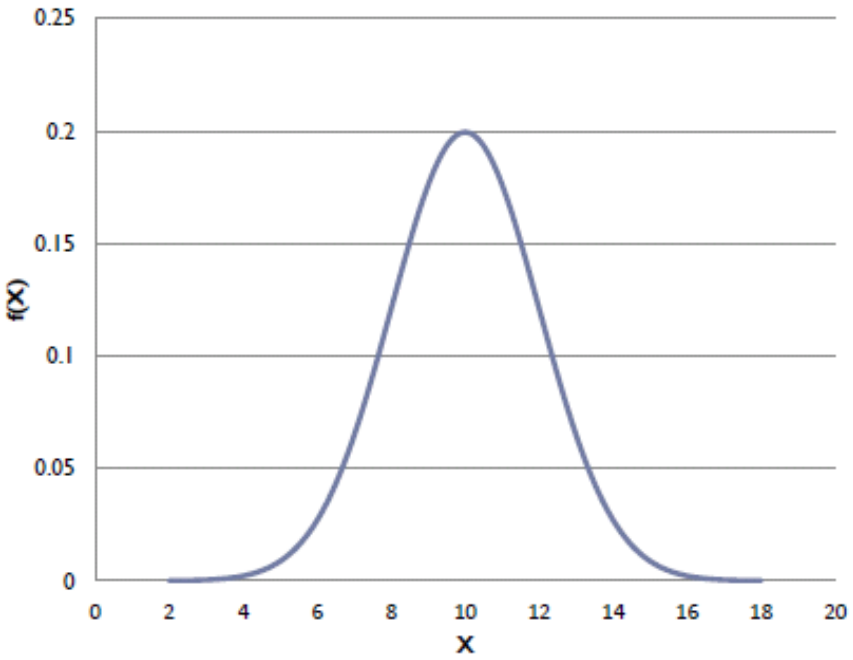
- Multe marini biologice se pot reprezenta printr-o curba simetrica fata de medie, avand forma unui “clopot” – distributia normala Gauss
- Modeleaza numeroase variabile numerice masurate pe populatii de volum mare
- Constituie fundamentul inferentei statistice, întrucât reprezinta distributia estimarilor parametrului unei populatii, estimarile fiind obtinute din toate esantioanele posibile

Distribuţia normală Gauss

- Functia de distributie

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Depinde de:
  - Medie
  - Deviatia standard



Media = 10, Dev. Std = 2

- Aria totala de sub curba este egala cu 1
- Este simetrica in jurul mediei
- "Cozile" se intind la infinit, nu ating niciodata planul orizontal

Intervale de încredere pentru eșantion

$$x_i \in (\bar{x} - s, \bar{x} + s);\ p = 68\%$$

$$x_i \in (\bar{x} - 2s, \bar{x} + 2s);\ p = 95\%$$

$$x_i \in (\bar{x} - 3s, \bar{x} + 3s);\ p = 99.7\%$$

- p = proporţia indivizilor care au mărimea în intervalul respectiv
- p = probabilitatea ca un individ să aiba mărimea în intervalul respectiv

Intervale de încredere pentru populaţii

- **Eroarea standard a mediei: s_x = s /√ n**

$$\mu \in (\bar{x} - s_x, \bar{x} + s_x);\ p = 68\%$$

$$\mu \in (\bar{x} - 2s_x, \bar{x} + 2s_x);\ p = 95\%$$

$$\mu \in (\bar{x} - 3s_x, \bar{x} + 3s_x);\ p = 99.7\%$$

- p = probabilitatea ca media populatiei (μ) să fie situată în intervalul respectiv