 Dragomirescu L. Lucrari practice de biostatistica. Editia a III-a revazuta si adaugita, 263 pp.
Editura “Agronomica”, Bucuresti, 2003.

4.7 Rezumat

Partea a III-a. ELEMENTE DE STATISTICĂ INDUCTIVĂ

Statistica inductivă, **analitică** sau **inferențială** = o cercetare selectivă a populațiilor statistice, prin **eşantioane probabiliste** (cărora li se pot calcula probabilitățile de apariție).

- Eşantioanele sunt extrase cu scopul de a *reprezenta* populațiile mamă. *Reprezentativitatea* are cel puțin două înțelesuri distincte:
    - *reprezentativitate în sens comun* şi
    - *reprezentativitate în sens probabilist.*
- Un eşantion este **reprezentativ în sens comun** (pentru populația mamă) dacă reproduce caracteristicile populației la o scară mai mică.

Asigurarea acestui tip de reprezentativitate presupune, în mod evident, cunoaşterea caracteristicilor populației, ceea ce este un nonsens atunci când scopul eşantionării este tocmai cunoaşterea acestor caracteristici. Pentru această situație trebuie introdusă accepțiunea următoare.

- Un eşantion este **reprezentativ în sens probabilist** dacă i se poate calcula probabilitatea de apariție. Probabilitatea respectivă se poate confirma experimental prin frecvența de apariție a eşantionului respectiv, în mai multe repetări ale eşantionării, dacă şi numai dacă extragerile s-au efectuat autentic aleator, adică prin “randomizare”. Efectuarea randomizării este singura cale prin care putem spera că eşantionul reprezintă caracteristicile populației aşa cum sunt ele şi nu conform « părerii » pe care o avem despre populația respectivă.

În concluzie, atunci când cunoaştem cu certitudine anumite elemente de structură ale populației, putem vorbi despre *reprezentativitate în sens comun* în legătură doar cu acele elemente, iar pentru elementele de structură necunoscute putem vorbi despre *reprezentativitate în sens probabilist* doar dacă s-au efectuat randomizările necesare.

Extrapolarea judecăților obținute pe eşantioane la populațiile statistice mamă sn **inferență** sau **inducție**, iar inferența pe baza eşantioanelor extrase prin randomizare sau, altfel spus, pe baza eşantioanelor reprezentative în sens probabilist sn **inferență statistică**. Doar inferența statistică este considerată ştiințifică, inferența nestatistică fiind acceptată doar atunci când nu se poate infera statistic.

Capitolul 4. PROBLEMA EŞANTIONAJULUI

- Când studiem o populaţie statistică (un fenomen) în mod complet, obţinem rezultate CERTE. Când o studiem incomplet, doar prin eşantioane, obţinem rezultate INCERTE, doar probabile.
- Probabilitatea unui rezultat « citit » pe un eşantion nu poate fi calculată decât dacă eşantionul a fost obţinut dintr-o urnă « bine amestecată ». Numai aşa orice eşantion de volum fixat *n* are aceeaşi şansă de a fi extras şi, astfel, i se poate calcula probabilitatea de apariţie. Acesta sn **eşantion aleator simplu**. Dacă $n \geq 2$ extragerea se poate face "cu revenire" sau "fără revenire".

4.3. Observaţii

- Dacă volumul eşantionului este 1, cele două tipuri de extragere sunt echivalente şi în cazul *eşantionului aleator simplu* orice unitate are aceeaşi probabilitate de a fi extrasă.
- Dacă populaţia (urna) este infinită sau cu un volum foarte mare şi eşantionul are un volum mult mai mic - cazul cel mai frecvent în practică - cele două moduri de extragere pot fi considerate echivalente.
- Se consideră, de regulă, drept eşantion aleator simplu, cel cu revenire (adică cel în care unele unităţi pot apărea de mai multe ori).

Două rezultate fundamentale asociate:

Rezultat de *teoria probabilităţilor*:	Rezultat de *statistică matematică:*
• **Atunci când cunoaştem conţinutul populaţiei (urnei) şi dispunem de un procedeu de a o "amesteca bine", putem să calculăm, în funcţie de tipul de extragere, probabilitatea de obţinere a oricărui eveniment** (de exemplu, extragerea unui anumit eşantion de volum fixat *n*).	• **Dacă NU cunoaştem conţinutul populaţiei (urnei), dar dispunem de un procedeu de a o "amesteca bine" şi extragem astfel un eşantion** (= doar o parte) **din ea, putem să calculăm probabilitatea de a obţine eşantionul respectiv dintr-o populaţie (urnă) de o anumită compoziţie ipotetică. Dacă probabilitatea respectivă este nulă sau aproape nulă vom respinge ipoteza asupra conţinutului urnei.**

4.3.1. Cum se asigură "buna amestecare" sau aleatorismul extragerii

- Aleatorismul este irealizabil direct de către subiecţi umani. Aceştia nu pot genera numere aleatoare, ci produc *numere într-o doară*.

- Aleatorismul autentic se poate realiza doar prin randomizare, adică apelând la *tabele de numere aleatoare* produse de fenomene aleatoare sau cel puțin la *numere pseudoaleatoare* produse de calculatoare.

4.3.2. Cum se consultă o tabelă de numere aleatoare

- Se numerotează indivizii populației de studiat cu valori de la 1 la *N*. Se stabileşte numărul de cifre zecimale, *p*, ale numărului *N*, şi volumul *n* al eşantionului ce se va extrage prin randomizare.
- Se aleg într-o doară: un număr în tabelă, o direcție şi un sens de parcurgere.
- Se urmează direcția aleasă în sensul stabilit şi se rețin *n* numere formate din *p* cifre care sunt mai mici sau egale cu *N*.
  - Dacă dorim un eşantion cu revenire vom accepta între cele *n* numere şi numerele care s-au repetat.
  - În caz contrar, vom reține doar numere care nu se repetă.
- Includem în eşantion exact unitățile cu numerele din lista astfel produsă.

4.3.3. Concluzii practice

- Este <u>obligatorie</u> folosirea numerelor aleatoare,
- Utilizarea numerelor pseudoaleatoare este acceptabilă dar
- este <u>strict interzisă</u> folosirea numerelor într-o doară, deoarece numai randomizarea poate permite calcularea probabilităților evenimentelor de interes, probabilități care se numesc **grad de încredere** în cazul estimării, respectiv, **nivel de semnificație** în cazul testării statistice.

Capitolul 5. PROBLEMA ESTIMĂRII

5.1. Mic dicționar de termeni

A estima = a aproxima. **Estimare** = aproximare. **Estimație** = aproximație.
Estimație punctuală = o valoare care aproximează un parametru, fără a se cunoaşte precizia (respectiv, marja de eroare a) aproximării. **Estimație prin interval de încredere** = două numere (diferența lor indicând direct “marja de eroare” sau, în mod invers, “precizia aproximării”) şi o probabilitate care exprimă **gradul de încredere** ataşat intervalului.
Estimată = valoarea (necunoscută) aproximată.
Estimator = persoană şi/sau instrument care aproximează.
Parametru (în sens restrâns) = indicator al unei populații. **Statistică** = indicator al unui eşantion. Prin statistici estimăm parametri.

5.2. Estimarea mediei unei populații statistice prin media unui eşantion aleator simplu. Studii de caz

Media de eşantionaj (sau **distribuția mediei de eşantionaj** sau **distribuția de eşantionaj a mediei**) = distribuția mediilor tuturor eşantioanelor aleatoare simple de volum fixat *n* care se pot extrage dintr-o populație dată. Se notează *m*.
Dispersia mediei de eşantionaj sn **dispersia standard a mediei** (de eşantionaj). Se notează $S^2(m)$.
Abaterea standard a mediei de eşantionaj sn **eroarea standard a mediei** (de eşantionaj). Se notează $S(m)$.

- $M(m)$ = (media mediei de eşantionaj) = μ (media populației).
- $S(m) = \dfrac{\sigma}{\sqrt{n}}$
- Experimentele numerice sugerează ideea că indiferent de forma distribuției caracterului în populație, media de eşantionaj tinde să devină o distribuție normală dacă volumul eşantionului creşte. (Fapt demonstrat printr-o teoremă limită, numită *teorema limită centrală*.)
  Dacă distribuția caracterului în populație este gaussiană atunci media de eşantionaj se distribuie gaussian pentru orice volum *n* fixat. (Teoremă exactă.)

- Consecință aplicativă:
  Teoremele ne asigură, de exemplu, de următorul "pariu" probabilist:
  $\mu - 1,28 \cdot S(m) < m < \mu + 1,28 \cdot S(m)$, cu probabilitatea 0,8 (80%),
  ceea ce ne conduce la inferența statistică:
  $(m - 1,28 \cdot S(m);\ m + 1,28 \cdot S(m)) \ni \mu$ cu gradul de încredere 80%,
  (citeşte "intervalul... conține media μ cu gradul de încredere 80%)
- Acest interval aleator sn **intervalul de 80% încredere** ceea ce înseamnă că dacă repetăm de multe ori extragerea unor eşantioane aleatoare simple de volum mare ($n > 30$, pentru a asigura condiția teoremei limită centrală) în circa 80% din cazuri intervalele de mai sus vor acoperi media populației, μ..

- În practică se extrage un singur eşantion aleator simplu de medie m_1. Spunem despre intervalul fix **$(m_1 - 1,28 \cdot S(m);\ m_1 + 1,28 \cdot S(m))$** că este **UN interval de 80% încredere**. Despre acest interval:
  - nu ştim dacă acoperă sau nu media μ,
  - ştim însă că dacă repetăm extragerea aleatoare, intervalele de forma generală de mai sus, **$(m - 1,28 \cdot S(m);\ m + 1,28 \cdot S(m))$**, obținute vor acoperi pe μ în circa 80% din situații şi că,

- în consecință, este mai natural să sperăm că intervalul unic obținut de noi, ***(m_1 - 1,28·S(m); m_1 + 1,28·S(m))***, face parte din cele 80% care acoperă media decât din cele 20% care nu o acoperă.

De aici rezultă că gradul de încredere trebuie să fie mai mare de 50%.

5.3. Eroarea standard a unui estimator, eroarea limită a unei estimații și intervalul de încredere aferent

5.3.1. Eroarea standard a unui estimator

Estimatori ai mediei μ a unei populații statistice:	**Erorile standard ale estimatorilor:**	**Denumirile erorilor standard:**
Media de eşantionaj, *m*	$S(m) = \dfrac{\sigma}{\sqrt{n}}$	Eroarea standard a mediei (de eşantionaj)
Mediana de eşantionaj, *med*	$S(med) = \dfrac{\sigma}{\sqrt{n}} \cdot \sqrt{\dfrac{\pi}{2}}$, cu π = 3,1415...	Eroarea standard a medianei (de eşantionaj)

Aberații din literatură: eroarea standard a eşantionului, eroarea standard a mediei estimate.

	Notații, în general:	**Exemple:**
Parametru estimat:	θ	μ (media populației)
Estimator al lui θ	$\hat{\theta}$	
Exemple:	$\hat{\theta}_1$ $\hat{\theta}_2$	*m*, media de eşantionaj *med*, mediana de eşantionaj

5.3.2. Eroarea (absolută) limită a unei estimații sau MARJA de EROARE și intervalul de încredere corespunzător

Eroarea standard Eroarea (absolută) limită cu gradul de încredere 1-α Intervalul de încredere 1-α	atribut al:	unui estimator unei estimații unei estimații

Dacă folosim un eşantion cu volum $n > 30$:

eroarea (absolută) limită (a *estimației - lui* μ *- cu gradul de încredere* $1-\alpha$), realizată prin estimatorul *m*, este produsul:

$$\Delta_{\alpha/2} = z_{\alpha/2} \cdot S(m)$$

în care:

- $z_{\alpha/2}$ este α-cuantila bilaterală superioară a distribuției normale standard şi
- $S(m)$ este eroarea standard a mediei (de eşantionaj).

În general:

eroarea (absolută) limită (a *estimației - lui* θ *- cu gradul de încredere* $1-\alpha$), realizată prin estimatorul $\hat{\theta}$, este produsul:

$$\Delta_{\alpha/2} = y_{\alpha/2} \cdot S(\hat{\theta})$$

în care:

- $y_{\alpha/2}$ este α-cuantila bilaterală superioară a distribuției de eşantionaj corespunzătoare, şi
- $S(\hat{\theta})$ este eroarea standard a estimatorului $\hat{\theta}$.

Sn **interval de încredere** (sau de **confidență**) $\mathbf{1-\alpha}$ (sau $\mathbf{(1-\alpha)\cdot 100\%}$) intervalul centrat în estimația punctuală calculată pe baza eşantionul extras aleator şi de lungime $2 \cdot \Delta_{\alpha/2}$. Definiție este valabilă doar în cazul distribuțiilor de eşantionaj simetrice față de origine (ca normala standard şi *t*-urile, care intervin în cazul estimării mediei).

- Eroarea limită (marja de eroare) şi intervalul de încredere indică în mod invers precizia estimației.
- Eroarea limită (marja de eroare) a unei estimații şi intervalul de încredere cresc o dată cu gradul de încredere $p = 1-\alpha$.
- În consecință, precizia unei estimații scade o dată cu creşterea gradului de încredere.

5.4. Proprietăți ale estimatorilor

5.4.1. Justețe - deplasare (proprietate binară)

1°

Un estimator $\hat{\theta}$ sn **just** (sau **nedeplasat**) dacă media sa coincide cu parametrul estimat θ: $M(\hat{\theta}) = \theta$. În caz contrar, se numeşte **deplasat** (sau **injust**): $M(\hat{\theta}) \neq \theta$.

2°

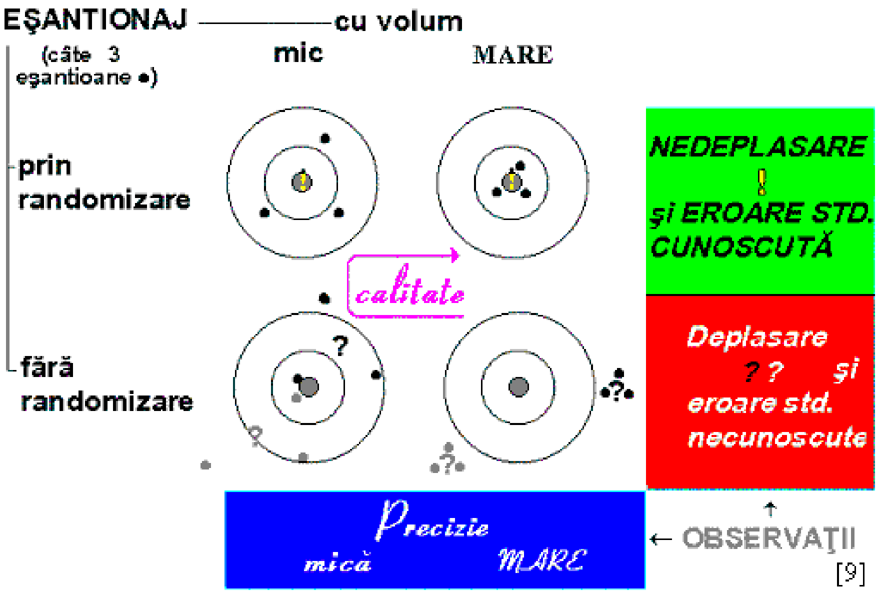
- Media de eşantionaj, m, este un estimator just al mediei populației statistice, μ:
  $M(m) = \mu$.
- Dispersia de eşantionaj, s^2, este un estimator deplasat al dispersiei populației, σ^2, mai precis: $M(s^2) < \sigma^2$.
- Un estimator just pentru σ^2 este dispersia corectată de eşantionaj:
  $s_c^2 = \dfrac{\sum (x - m)^2}{n-1} = s^2 \cdot \dfrac{n}{n-1}$. Deci $s_c = s \cdot \sqrt{\dfrac{n}{n-1}}$.

5.4.2. Eficacitate mai mare sau mai mică (proprietate graduală), eficiență

Un estimator just $\hat{\theta}_1$ este **mai eficace decât un alt estimator** just $\hat{\theta}_2$ dacă are o eroare standard mai mică: $S(\hat{\theta}_1) < S(\hat{\theta}_2)$.
Cel mai eficace estimator - al aceluiaşi parametru estimat - se numeşte estimator **eficient**.

- Media de eşantionaj este un estimator mai eficace decât mediana de eşantionaj.
- Media de eşantionaj este chiar un estimator eficient al mediei populației.
- Putem obține estimări la fel de precise (pentru un anumit grad de încredere 1-α fixat):
  - fie calculând media pe un eşantion de 64 subiecți,
  - fie calculând mediana pe un eşantion de 100 subiecți.

5.5. Problema estimării – rezumat



- Etapele realizării unei estimații corecte și cu efort minim:
  - stabilirea estimatorului eficient în condițiile date,
  - calcularea volumului (eşantionului) necesar pentru obținerea preciziei dorite, la gradul de încredere fixat şi
  - efectuarea randomizării.
- Erori grave, foarte frecvente: (1) aplicarea acestei teorii la eşantioane obținute fără randomizare, caz în care toate conceptele anterioare, mai puțin cel de estimație punctuală, nu au sens ; (2) "Volum mare = eşantion reprezentativ". Se gândeşte în termenii conceptului de « reprezentativitate în sens comun », care este un nonsens în statistica inductivă. Aici acționează reprezentativitatea în sens probabilist. Aceasta este asigurată de justețea estimatorului şi de randomizare, nu de un volum mare; (3 = 1+2) utilizarea de eşantioane obținute fără randomizare şi cu volume mari. Volumul mare creşte precizia estimării, dar fără randomizare estimațiile sunt deplasate necontrolabil şi toate conceptele anterioare nu au sens, inclusiv estimațiile punctuale. Cele mai frecvente valori ale acestora sunt cele mai aberante estimații.

5.6. Aplicația 1: estimarea mediei μ prin media de eşantionaj aleator simplu, *m*, în diverse ipoteze

	1° volumul eşantionului $n > 30$	2° volumul eşantionului $n \leq 30$
5.6.1. **Cazul 1:** **σ cunoscut**	*Cazul 1.1:* Indiferent de forma distribuției variabilei în populație $m \in N\left(\mu, \sigma/\sqrt{n}\right)$ sau, altfel scris, $\frac{m-\mu}{\sigma/\sqrt{n}} \in N(0{,}1)$ (conform teoremei limită centrală).	*Cazul 1.2:* a) Dacă variabila $x \in N(\mu, \sigma)$ în populația statistică, atunci $\frac{m-\mu}{\sigma/\sqrt{n}} \in N(0{,}1)$ (conform teoremei exacte). b) Dacă variabila $x \notin N(\mu, \sigma)$ se caută o transformare de cvasigaussianizare. Dacă se găseşte, teorema se aplică datelor astfel transformate.
5.6.2. **Cazul 2: σ estimat** prin $s_c = s \cdot \sqrt{\frac{n}{n-1}}$ $(\Rightarrow \boxed{\frac{s_c}{\sqrt{n}} = \frac{s}{\sqrt{n-1}}})$	*Caz 2.1*: Indiferent de forma distribuției variabilei în populație: $m \in N\left(\mu, s\ /\sqrt{n-1}\right)$ sau, altfel scris, $\frac{m-\mu}{s/\sqrt{n-1}} \in N(0{,}1)$ (conform teoremei limită centrală).	*Caz 2.2:* a) Dacă variabila $x \in N(\mu, \sigma)$ în populația statistică, atunci $\frac{m-\mu}{s/\sqrt{n-1}} \in t_{n-1}$ b) Dacă variabila $x \notin N(\mu, \sigma)$, se caută o transformare de cvasigaussianizare. Dacă se găseşte, teorema se aplică datelor astfel transformate.

- Intervalul de (1-α) · 100 % încredere pentru media μ este deci:

$\mu = m \pm \Delta_{\alpha/2}$, unde valorile $\Delta_{\alpha/2}$ (de forma $y_{\alpha/2} \cdot S(m)$)

sunt înscrise în celulele din chenarul dublu:

<table>
<tr><td rowspan="2">Abaterea standard în populație σ:</td><td colspan="2">Volumul eşantionului:</td></tr>
<tr><td>1° mare (n > 30)</td><td>2° mic (n ≤ 30) şi x ∈ N (μ, σ)</td></tr>
<tr><td>Cazul 1: cunoscută
$\Rightarrow S(m) = \sigma/\sqrt{n}$</td><td>$z_{\alpha/2} . \sigma/\sqrt{n}$</td><td>$z_{\alpha/2} . \sigma/\sqrt{n}$</td></tr>
</table>

Cazul 2: estimată prin s (sau s_c) $\Rightarrow$ S (m) = $s/\sqrt{n-1}$ = $(s_c/\sqrt{n})$	$z_{\alpha/2} \cdot s/\sqrt{n-1}$	$t_{n-1,\,\alpha/2} \cdot s/\sqrt{n-1}$

5.6.3. Observații practice

La publicarea rezultatelor unei cercetări se indică, pe baza eşantionului de volum n analizat, una din următoarele patru variante, în care prin $\bar{x}$ sau m_1 se notează media eşantionului utilizat, iar e.a.s. înseamnă eşantion aleator simplu:

	Exemple:	*Observații:*
1. $\bar{x} \pm s$ Eşantion de volum n	$\bar{x} \pm s$ = 160,5 ± 10 n = 100	Prezentarea mediei şi abaterii standard a eşantionului fără intenția estimării statistice a acestora în populația mamă.
2. $\bar{x} \pm s_c$ Eşantion de volum n	$\bar{x} \pm s_c$= 160,5 ± 10,05, cu volumul n = 100	Prezentarea mediei şi abaterii standard corectate şi volumului e.a.s. pentru estimarea statistică punctuală a mediei şi abaterii standard în populația mama.
3. $\mu = \bar{x} \pm S(m)$, unde $S(m)$ = $s/\sqrt{n-1}$ =$(s_c/\sqrt{n})$ Eşantion de volum n	μ = 160,5 ± 1, cu volumul n = 100	Prezentarea mediei şi volumului e.a.s şi a erorii standard a estimatorului, pentru estimarea mediei din populația mamă. (Este forma cea mai uzitată.)
4. $\mu = \bar{x} \pm \Delta_{\alpha/2}$, cu α specificat (Eşantion de volum n)	μ=160,5 ± 1,64, cu α = 0,1 (n = 100)	Prezentarea mediei e.a.s, a gradului de încredere şi a erorii limită a estimației asociate acestuia. (Este forma cea mai direct sugestivă şi pentru care nu este obligatorie, deşi este recomandată, specificarea volumului.)

Erori foarte frecvente: (1) Se notează cu s atât abaterea standard, cât şi abaterea standard corectată (s_c). (2) Se scrie doar «160,5 ± 12» fără a se preciza care dintre cele patru statistici reprezintă valoarea 12.

RECOMANDARE: Scrieți «160,5 ± 12, unde 12 este abaterea standard / abaterea standard corectată / eroarea standard a mediei / eroarea limită (marja de eroare de estimare) cu α = 0,1», în primele trei cazuri adăugând obligatoriu şi n = 100. În locul denumirilor se pot pune formulele, dar nu notațiile, acestea nefiind standardizate în mod unic.

OBSERVAȚIE: Prima variantă este valabilă pentru orice eşantion indiferent de modul de obținere, iar ultimele trei variante au sens doar dacă eşantionul a fost obținut prin randomizare şi este e.a.s.. (Există şi eşantioane care se obțin tot prin randomizare, dar mai « complicat » diferind de cele aleatoare simple. Şi prin acestea se poate face estimare statistică dar cu alte formule.)

5.7. Aplicația 2: Estimarea unei proporții *p* dintr-o populație statistică, prin frecvența relativă *f*, dintr-un eşantion aleator simplu.

Propoziție:

Frecvența de eşantionaj f se distribuie:

1) gaussian, cu

2) $M(f) = p$, p fiind proporția din populația mamă şi

3) $S(f) \approx \sqrt{\dfrac{f \cdot (1-f)}{n}}$, în care n este volumul fixat al eşantioanelor aleatoare simple,

dacă:

a) $n > 30$ şi

b) $n \cdot f$, respectiv $n \cdot (1 - f) \geq 10$ sau, cel puțin, 5.

$\Delta_{\alpha/2} = z_{\alpha/2} \cdot S(f)$ sn **precizia absolută** *a estimării cu gradul de încredere 1-*α*.*

Greşeală tipică:
Uneori, mai ales unii sociologi, folosesc exprimarea: "Estimăm că procentul în populație este în jurul valorii 65% cu o marjă de eroare de ±13%", exprimare care este doar aparent ştiințifică, dar care nu are nici o valoare, dacă nu se specifică totodată şi gradul de încredere al estimării, căci putem afirma o marjă de eroare de estimare (precizie) oricât de bună (5%, la 3%, la 1% etc.), reducând însă gradul de încredere al estimării, în mod corespunzător (poate chiar sub 50%, ceea ce ar reprezenta o aberație). Publicul larg nu reacționează pentru că se confundă bine cunoscuta eroarea de măsurare cu eroarea de estimare. Confuzia produce manipulare, considerându-se că valoarea estimată este cu certitudine în intervalul respectiv, ca la o simplă valoare măsurată.

47 Teste, exerciții şi probleme

TG7. Durata 170'' pe calculator.

Dintre calitatile unui estimator cea mai importanta este:
1. precizia.
2. justetea.

Alegeti afirmatia corecta:
1. Studiul nostru este reprezentativ deoarece a fost efectuat pe un esantion extras prin randomizare si are un volum suficient de mare pentru a asigura precizia necesara.
2. Studiul nostru este reprezentativ pentru lacurile din Delta Dunarii, deoarece a fost efectuat pe un esantion cu un volum suficient de mare pentru a asigura precizia necesara.
3. Studiul nostru este reprezentativ pentru lacurile din Delta Dunarii, deoarece a fost efectuat pe un esantion extras prin randomizare din toate lacurile din delta si are un volum suficient de mare pentru a asigura precizia necesara.

Alegeti comentariul corect la urmatoarea afirmatie din literatura: 'Reprezentativitatea unui esantion este data, in esenta, de eroarea standard care ii corespunde.'
1. Eroarea standard nu corespunde unui esantion ci unui estimator iar reprezentativitatea unui esantion se asigura mai intai prin randomizare si apoi prin volum cat mai mare.
2. Eroarea standard nu corespunde unui esantion ci unui estimator iar reprezentativitatea unui esantion se asigura prin volum cat mai mare.

Exprimarea 'intervalul de incredere 95% (20,30) acopera media populatiei statistice' inseamna:
1. ca daca repetam de multe ori extragerea a cate unui esantion aleator, in cca 95% din cazuri intervalul (20, 30) va acoperi media populatiei statistice;
2. ca 95% din unitatile statistice ale populatiei au valorile cuprinse in intervalul respectiv;

3. ca daca repetam de multe ori extragerea a cate unui esantion aleator, in cca 95% din cazuri intervalul general de forma (m-delta, m+delta) va acoperi media populatiei statistice.

Daca dorim sa micsoram un interval de confidenta trebuie:
1. fie sa marim volumul esantionului, fie sa acceptam un grad de incredere mai redus;
2. fie sa micsoram volumul esantionului, fie sa acceptam un grad de incredere mai redus;
3. fie sa marim volumul esantionului, fie sa lucram cu un grad de incredere mai ridicat;

Daca marim volumul esantionului, intervalul de incredere se va:
1. mentine acelasi
2. micsora
3. mari

Estimarea dispersiei unei populatii se face prin:
1. dispersie, deoarece este un estimator eficient.
2. dispersia corectata, deoarece este mai eficace decit dispersia.
3. dispersia corectata, deoarece este un estimator nedeplasat.

TC7. Durata 4'.

1. Teoria estimaţiei oferă instrumente pentru estimarea ________________ prin _____________.
2. Un eşantion ____________________ este orice eşantion de volum *n* fixat care are aceeaşi şansă de a fi extras.
3. Abaterea standard a mediei de eşantionaj se numeşte ___________________.
4. Teorema exactă pentru distribuţia mediei de eşantionaj se referă la distribuţia ________ a caracterului în populaţie.
5. Corectaţi afirmaţia următoare:

Eroarea standard a eşantionului este un atribut al estimatorului.

8. Intervalul centrat în estimaţia punctuală şi de lungime $2\cdot\Delta_{\alpha/2}$ se numeşte ____.
9. Eroarea limită a unei estimaţii __________ cu gradul de încredere $p = 1- \alpha$.
10. Un bun estimator trebuie să fie ______________ şi apoi ___________.
11. Comparaţia între eficacitatea a 2 estimatori nedeplasaţi se face pe baza _____________________.

13. Un volum mai mare al eşantionului pentru realizarea unei estimări are sens doar dacă ________________________________.

14. Când σ este cunoscut şi n > 30, distribuţia mediei de eşantionaj poate fi considerată_________________________.

Exerciţii sau probleme rezolvate

1.

Să se estimeze media în populaţie, printr-un interval de încredere 90%, pe baza următorului eşantion aleator simplu alcătuit din următoarele 32 elemente:
1 1 1 1 2 2 2 2 2 2 3 3 3 3 3 3 4 4 4 4 4 4 5 5 5 5 5 5 6 6 6 6.
Nu se cunoaşte distribuţia caracterului în populaţie dar se ştie că abaterea standard în populaţie are valoarea 10.

Rezolvare:
Deoarece se cunoaşte abaterea standard în populaţie şi volumul eşantionului n = 32 este mare (n > 30), suntem în cazul 1.1 (vezi paragraful 5. 6 din rezumat). Calculăm media eşantionului, după ce grupăm datele în distribuţia de frecvenţe din primele două coloane ale tabelului următor:

x_j	n_j	$x_j \cdot n_j$
1	4	4
2	6	12
3	6	18
4	6	24
5	6	30
6	4	24
	$n = \Sigma N_j = 32$	$T_1 = 112$

$m_1 = T_1 / n = 112 / 32 = 3{,}5.$

Eroarea standard a mediei de eşantionaj:
$S(m) = \sigma / \sqrt{n} = 10 / \sqrt{32} \approx 10 / 5{,}66 \approx 1{,}77.$
$1 - \alpha = 90\% = 0{,}9 \Rightarrow \alpha = 0{,}1.$
$z_{0,1\,/\,2} = z_{0,05} = 1{,}64$ (din tabela distribuţiei normale standard. Vezi Anexa 2.) Eroarea limită cu gradul de încredere 90% = (lungimea semiintervalului de încredere) = $\Delta_{\alpha/2} = z_{\alpha/2} \cdot S(m) = 1{,}64 \cdot 1{,}77 \approx 2{,}9$. Deci $\mu = m_1 \pm \Delta_{\alpha/2} = 3{,}5 \pm 2{,}9$ *cu gradul de încredere 90%* sau, altfel scris, *intervalul (0,6; 6,4)* ∍ μ *cu gradul de încredere 90%.* (*Observaţie:* Semnul “∍” se citeşte “conţine” sau “acoperă”.)

2.

Să se estimeze media în populaţie, printr-un interval de încredere 80%, pe baza următorului eşantion extras aleator şi alcătuit din următoarele 9 elemente:
100 20 30 40 60 30 40 50 20.

Se ştie că distribuţia caracterului în populaţie este gaussiană şi că abaterea standard în populaţie are valoarea 90.

Rezolvare:
Deoarece se cunoaşte abaterea standard în populaţie, volumul eşantionului $n = 9$ este mic ($n \leq 30$) şi se cunoaşte că distribuţia caracterului în populaţie este gaussiană, suntem în cazul 1.2a (vezi paragraful 5. 6. din rezumat). Media eşantionului este $m_1 = 43{,}33$. Eroarea standard a mediei de eşantionaj $S(m) = \sigma / \sqrt{n} = 90 / 3 = 30$. $\alpha = 0{,}2$. $z_{0,2/2} = z_{0,1} = 1{,}28$. Eroarea limită cu gradul de încredere 90% = (lungimea semiintervalului de încredere) = $\Delta_{\alpha/2} = z_{\alpha/2} \cdot S(m) = 1{,}28 \cdot 30 = 38{,}4$. Deci $\mu = m_1 \pm \Delta_{\alpha/2} = 43{,}33 \pm 38{,}4$ *cu gradul de încredere 80%* sau, altfel scris, *intervalul (4,93; 81,73)* $\ni \mu$ *cu gradul de încredere 80%.*

3.

Să se estimeze media în populaţie, printr-un interval de încredere 90%, pe baza următorului eşantion extras aleator şi alcătuit din următoarele 34 elemente:
1 1 1 1 1 2 2 2 2 2 2 3 3 3 3 3 3 4 4 4 4 4 4 5 5 5 5 5 5 6 6 6 6 6.
Nu se cunoaşte distribuţia caracterului în populaţie şi nu se cunoaşte abaterea standard în populaţie.

Rezolvare:
Deoarece nu se cunoaşte abaterea standard în populaţie şi volumul eşantionului $n = 32$ este mare ($n > 30$), suntem în cazul 2.1 (vezi paragraful 5. 6 din rezumat). Calculăm media şi abaterea standard corectată ale eşantionului, după ce grupăm datele în distribuţia de frecvenţe din primele două coloane ale tabelei următoare:

x_j	n_j	$x_j \cdot n_j$	x_j^2	$n_j \cdot x_j^2$
1	5	5	1	5
2	6	12	4	24
3	6	18	9	54
4	6	24	16	96
5	6	30	25	150
6	5	30	36	180
	$n = \Sigma n_j = 34$	$T_1 = 119$		$T_2 = 509$

Media este $m = T_1 / n = 119 / 34 = 3{,}5$, iar dispersia corectată

$$s^2 = \frac{T_2}{n} - m^2 = \frac{509}{34} - 3{,}5^2 \approx 14{,}97 - 12{,}25 = 2{,}72.$$

Abaterea standard este $s \cong 1{,}65$.

Eroarea standard a mediei de eşantionaj $S(m)$ = $s/\sqrt{n-1} \cong$ 1,65 / 5,74 = 0,2875. α = 0,1. $z_{0,1/2}$ = $z_{0,05}$ = 1,645. Eroarea limită cu gradul de încredere 90% = (lungimea semiintervalului de încredere 90%). $\Delta_{\alpha/2}$ = $z_{\alpha/2} \cdot S(m)$ = 1,645 · 0,2875 = 0,4729375 ≅ 0,47. Deci μ = $m_1 \pm \Delta_{\alpha/2}$ = *3,5* $\pm$ *0,47 cu gradul de încredere 90%* sau, altfel scris, *intervalul (3,03; 3,97) ∍* μ *cu gradul de încredere 90%.*

4.

Să se estimeze media în populaţie, printr-un interval de încredere 90%, pe baza următorului eşantion extras aleator şi alcătuit din următoarele 9 elemente:
79 91 57 35 28 37 33 90 10.
Caracterul este distribuit gaussian în populaţie dar nu se cunoaşte abaterea standard în populaţie.

Rezolvare:
Deoarece nu se cunoaşte abaterea standard în populaţie, volumul eşantionului n = 9 este mic ($n \leq 30$) şi se cunoaşte că distribuţia caracterului în populaţie este gaussiană, suntem în cazul 2. 2a.
Calculăm media, dispersia şi abaterea standard ale eşantionului: m = 51,11 şi s^2 ≅ 769,65 $\Rightarrow$ s ≅ 27,74 Eroarea standard a mediei de eşantionaj $S(m)$ = $s/\sqrt{n-1}$ = 27,74/2,83 ≅ 9,8. α = 0,1 $\Rightarrow$ $t_{n-1;\ \alpha/2}$ = $t_{8;\ 0,1/2}$ = 1,86 (din Anexa 3). Eroarea limită cu gradul de încredere 90% = (lungimea semiintervalului de încredere) $\Delta_{\alpha/2}$ = $t_{n-1;\ \alpha/2} \cdot S(m)$ = 1,86 · 9,8 = 18,24. Deci μ = $m_1 \pm \Delta_{\alpha/2}$ = *51,11* $\pm$ *18,24 cu gradul de încredere 90%* sau, altfel scris, *intervalul (32,87; 69,35) ∍* μ *cu gradul de încredere 90%.*

5.

Să se estimeze media într-o populaţie, printr-un interval de încredere 95%, pe baza următorului eşantion extras aleator şi alcătuit din următoarele 28 elemente:
4 4 4 4 4 5 5 5 5 5 5 6 6 6 6 6 6 7 7 7 7 7 7 8 8 8 8 8.
Se cunosc următoarele informaţii: caracterul nu este distribuit gaussian în populaţie şi nu se cunoaşte abaterea standard în populaţie.

Rezolvare:

Cazul 2. 2. b.

Deoarece nu se cunoaşte distribuţia caracterului în populaţie, deci nici o transformare care să poată fi aplicată datelor pentru cvasigaussianizare, iar volumul eşantionului este mai mic decât 30, problema nu se poate rezolva cu datele existente. Pentru rezolvare trebuie fie să găsim distribuţia caracterului în populaţie, fie să prelevăm un eşantion cu volum mai mare de 30.

6.

Din 200 exemplare extrase randomizat dintr-o populaţie de crapi, 20 prezentau paraziţi la nivelul branhiilor. Să se estimeze cu gradul de încredere 90% frecvenţa relativă a indivizilor parazitaţi din populaţie. Rezultatele finale se vor rotunji la două zecimale.

Rezolvare:

$n = 200 > 30$. $f = 20 / 200 = 0{,}1$. $n \cdot f = 200 \cdot 0{,}1 = 20 \geq 10$. $n \cdot (1 - f) = 200 \cdot 0{,}9 = 180 \geq 10$.

$S(f) \approx \sqrt{\dfrac{f \cdot (1 - f)}{n}} = \sqrt{\dfrac{0{,}1 \cdot 0{,}9}{200}} \approx 0{,}0212$ = eroarea standard a frecvenţei de eşantionaj. $\alpha = 0{,}1 \Rightarrow z_{0{,}1/2} = z_{0{,}05} = 1{,}64 \Rightarrow \Delta_{\alpha/2} = z_{\alpha/2} \cdot S(f) = 1{,}64 \cdot 0{,}0212 \approx 0{,}03$ = lungimea semiintervalului de încredere = eroarea limită cu gradul de încredere 90%.

$F = f \pm \Delta_{\alpha/2} = 0{,}1 \pm 0{,}03$ cu gradul de încredere 90%
sau, altfel scris,
intervalul (0,07; 0,13) ∍ F cu gradul de încredere 90%.

(*Observaţie:* Semnul ∍ se citeşte "conţine" sau "acoperă".)

Exerciții sau probleme propuse

7.

Să se estimeze media în populație, printr-un interval de încredere 90%, pe baza următorului eşantion extras aleator şi alcătuit din următoarele 28 elemente:
7 7 7 7 8 8 8 8 8 9 9 9 9 9 10 10 10 10 10 11 11 11 11 11 12 12 12 12.
Nu se cunoaşte distribuția caracterului în populație, se cunoaşte abaterea standard în populație, ea având valoarea 10.

8.

Să se estimeze media într-o populație, printr-un interval de încredere 80%, pe baza următorului eşantion obținut prin extragerea fiecărui al 100-lea element din populație şi alcătuit din următoarele 11 elemente:
21 29 35 41 47 25 38 43 30 27 33.
Se cunosc următoarele informații: caracterul este distribuit gaussian în populație şi nu se cunoaşte abaterea standard în populație.

9.

Să se estimeze media în populație, printr-un interval de încredere 90%, pe baza următorului eşantion extras aleator şi alcătuit din următoarele 33 elemente:
7 7 7 7 8 8 8 8 8 9 9 9 9 9 10 10 10 10 10 11 11 11 11 11 12 12 12 12 12 13 13 13 13.
Nu se cunoaşte distribuția caracterului în populație, se cunoaşte abaterea standard în populație, ea având valoarea 10.

10.

Să se estimeze media într-o populație, printr-un interval de încredere 80%, pe baza următorului eşantion extras aleator şi alcătuit din următoarele 8 elemente:
21 40 35 57 29 41 53 38.
Se cunosc următoarele informații: caracterul este distribuit gaussian în populație şi se cunoaşte abaterea standard în populație, aceasta având valoarea 18.

11.

Să se estimeze media într-o populaţie, printr-un interval de încredere 95%, pe baza următorului eşantion extras aleator şi alcătuit din următoarele 34 elemente:
3 3 3 3 3 4 4 4 4 4 4 5 5 5 5 5 5 6 6 6 6 6 6 7 7 7 7 7 7 8 8 8 8 8.
Se cunosc următoarele informaţii: caracterul nu este distribuit gaussian în populaţie şi nu se cunoaşte abaterea standard în populaţie.

12.

Să se estimeze media într-o populaţie, printr-un interval de încredere 80%, pe baza următorului eşantion extras aleator şi alcătuit din următoarele 11 elemente:
21 29 35 41 47 25 38 43 30 27 33.
Se cunosc următoarele informaţii: caracterul este distribuit gaussian în populaţie şi nu se cunoaşte abaterea standard în populaţie.

13.

Din 200 exemplare extrase randomizat dintr-o populaţie de bibani, 2 prezentau paraziţi la nivelul branhiilor. Să se estimeze cu gradul de încredere 90% frecvenţa relativă a indivizilor parazitaţi din populaţie.