MODELE DE PRELUCRARE STATISTICĂ PRIMARĂ A DATELOR ÎN CERCETAREA SOCIOLOGICĂ

Lect.univ.dr. MONICA ANGELA BARA,
Universitatea "1 Decembrie 1918" Alba Iulia

ABSTRACT: Working models of primary statistics data in Sociology. *Statistics provides indicators whoo summarizing the main characteristics of collections of cases and data interpretation techniques, assessing their significance and comparing. The paper presents the main primary analysis statistical methods easy to use all those working in education and make teaching research.*

Keywords: *statistical estimation, mean, standard deviation.*

Metodele statistice oferă nu numai indici care sintetizează caracteristicile principale ale unei colecții de cazuri, dar şi tehnici de interpretare a datelor, de estimare a semnificației acestora, de comparare între diferite frecvențe şi între diferiți indici. În cercetarea concretă, determinarea unor indici cum sunt: media, mediana, dispersia, coeficienții de corelație etc., nu este suficientă pentru formularea unor concluzii cu un grad ridicat de certitudine.

Spre exemplu, media notelor obținute de membrii unui grup experimental la un test de cunoştințe, după ce s-a folosit o metodă specială de învățare, nu este suficientă pentru a evalua eficiența metodei experimentale, întrucât trebuie cunoscut şi modul în care se raportează această medie la media obținută la aceeaşi probă de un grup de control cu care s-a lucrat în alt mod, cu altă metodă de învățare. În acelaşi fel se pune şi problema determinării semnificației unei frecvențe sau a unui coeficient de corelație.

Problema interpretării statistice apare, în mod deosebit, în cazul *cercetărilor selective,* respectiv în acele investigații efectuate pe eşantioane, dar care vizează formularea unor concluzii generale, valabile pentru întreaga populație din care eşantionul a fost extras şi pe care o reprezintă.

Într-adevăr, cercetările psiho-sociale se efectuează rareori pe colectivități totale, întrucât acest lucru este practic imposibil de înfăptuit. De regulă, cercetările se efectuează pe eşantioane extrase (sau considerate a fi extrase} dintr-o *colectivitate generală.* Extrapolarea datelor de la eşantion asupra populației totale nu se face însă fără riscul de a greşi, într-o măsură mai mare sau mai mică. Tehnicile de estimare nu fac altceva decât să "estimeze" acest risc atunci când se formulează concluzii generale pe bază de date parțiale.

1. Probleme de estimare statistică

În analiza şi interpretarea datelor unei cercetări nu este suficientă doar calcularea câtorva valori caracteristice (frecvențe, medii, coeficienți), fiind necesară şi stabilirea semnificației acestor valori pentru populația din care a fost extras eşantionul pe care s-a făcut cercetarea, fapt care poate fi ilustrat prin câteva exemple:

➤ în cazul în care se studiază răspândirea tulburărilor de vorbire la copii de vârstă şcolară mică pe un eşantion de circa 3000 de copii, se pune întrebarea: *în ce limite se poate afirma că frecvența tulburărilor stabilite pe eşantion este valabilă şi în cazul populației totale a şcolarilor mici dintr-un oraş, dintr-o zonă sau chiar dintr-o țară ?;*

- în cazul unui studiu referitor la volumul mediu al vocabularului la copii de 7 ani, efectuat pe un eşantion de circa 1000 copii, se pune întrebarea: *în ce limite volumul mediu obținut (4500 de cuvinte) este reprezentativ pentru întreaga populație a copiilor de 7 ani?*

Datele obținute pe unul sau mai multe eşantioane se apropie, mai mult sau mai puțin, de cele ale colectivității totale, dar rareori concordă pe deplin. Altfel spus, *indicii eşantionului reprezintă estimări ale indicilor colectivității*.

Aceste estimări, mai bine spus *şansa de eroare* a acestora depinde de trei factori, respectiv: *volumul eşantionului, tehnica de eşantionare* şi *variabilitatea colectivității* din care a fost extras eşantionul. Cum acest din urmă factor poate fi atenuat printr-o tehnică adecvată de eşantionare, de regulă rămân doi factori principali, volumul eşantionului (*N*) şi tehnica eşantionării.

JEAN STOETZEL, calculând tabelul probabilităților de eroare, a arătat că şansa de eroare nu depinde de raportul dintre mărimea eşantionului şi universul anchetei (populația totală) ci numai *de volumul absolut al eşantionului* şi de *tehnica de eşantionare*. Rezultă că nu interesează ponderea eşantionului în cadrul populației totale, ci doar mărimea lui absolută, aşa încât faptul că un eşantion deține o pondere mare în colectivitatea din care a fost extras, nu este un argument pentru reprezentativitatea lui.

GEORGE GALLUP a arătat, de exemplu, că probabilitatea de eroare a unui eşantion național pentru S.U.A, în mărime de 100 de persoane este de 15%, un eşantion reprezentativ de 900 de persoane dă o probabilitate de eroare de 5% iar un eşantion de 10.000 de persoane prezintă un grad de eroare de 1,5%, limită dincolo de care o creştere a volumului eşantionului nu mai aduce un spor semnificativ de reprezentativitate.

După P.R. HOFFSTÄTTER *probablitatea de eroare (O_p)* se poate calcula conform relației:

$$O_p = \sqrt{\frac{p(100 - p)}{N}} = \sqrt{\frac{pq}{N}}$$

unde: *p = procentul celor care exprimă o anumită opinie;*
q = 100 - p;
N = volumul (efectivul) eşantionului.

Exemplu: dacă dintr-un eşantion de 2400 persoane, 960 (respectiv 40%) exprimă opinia A, probabilitatea ca 40% din totalul populției din care a fost extras eşantionul să exprime aceeaşi opinie este:

$$O_p = \sqrt{\frac{40(100 - 40)}{2400}} = \sqrt{\frac{40 \cdot 60}{2400}} = \sqrt{\frac{2400}{2400}} = \pm 1\%$$

ceea ce însemnă că probabilitatea de eroare este de ±1%, respectiv, ponderea celor care vor exprima aceeaşi opinie A din totalul populției poate oscila între 39-41%.

2. Semnificația unei medii

Semnificația sau fidelitatea unei medii depinde de *volumul eşantionului studiat (N)* şi de *variabilitatea colectivității (σ)* din care a fost selectat eşantionul, în sensul că dacă volumul eşantionului creşte, media devine mai stabilă şi mai reprezentativă.

În cazul în care dintr-o populație se selectează mai multe eşantioane având fiecare acelaşi volum *(N)*, mediile calculate pentru fiecare eşantion *(m_1, m_2,...,m_n)* notate cu *(m')* vor prezenta o distribuție normală. Aceasta înseamnă că mediile calculate vor oscila în jurul mediei reale *(m)*, *abaterea standard a mediilor (σ_m)* având ca expresie:

$$\sigma_m = \frac{\sigma}{\sqrt{N}} = E$$

valoarea abaterii fiind denumită *eroare standard (E).*

Cu ajutorul erorii standard se pot determina *limitele intervalului de încredere*, respectiv *amplitudinea intervalului de*

încredere, şi anume:
-intervalul de încredere pentru *P = 0,005*:
lim.inf. = m'- 1,96E; lim.sup. = m' + 1,96E
-intervalul de încredere pentru *P = 0,001:*
lim.inf. = m'- 2,58E; lim.sup. = m' + 2,58E

Exemplu: în cazul unui eşantion extras dintr-o populaţie având:

N = 51, m' = 13.17, σ = 4,74

dacă se pune problema determinareii intervalului de încredere pentru pragul de semnificaţie *P = 0,005,* demersul este următorul:

-se calculează eroarea standard:

$$E = \frac{\sigma}{\sqrt{N}} = \frac{4{,}74}{\sqrt{51}} = \frac{4{,}74}{7{,}14} = 0{,}66$$

-se determină limitele intervalului de încredere:

lim.inf = 13,17 - 1,96 0,66 = 13,17 - 1,29 = 11,88
lim.sup = 13,17 + 1.96 0,66 = 13,17 + 1,29 = 14,46

Media reală a caracteristicii studiate pentru întreaga populaţie va fi cuprinsă în intervalul 11,88-14,46, însă există probabilitatea ca 5% din mediile eşantioanelor extrase din aceeaşi populaţie să nu se înscrie în acest interval, având în vedere pragul de semnificaţie pentru care au fost calculate limitele intevalului de încredere *P* = 0,005.

Acelaşi calcul se poate efectua şi dacă se alege un alt prag de probabilitate, spre exemplu: *P = 0,001,* caz în care limitele în care se înscrie media populaţiei, cu riscul ca 1% din mediile eşantioanelor să nu intre în acest interval, este:

lim.inf = 13,17 - 2,58 0,66 = 13,17 - 1,50 = 11,67
lim.inf = 13,17 + 2,58 0,66 = 13,17 + 1,50 + 14,67

În acest caz intervalul de încredere este: 11,67-14,67.

3. Semnificaţia relaţiei dintre medie şi abaterea standard

În cele ce urmează se va insista mai mult asupra proprietăţilor mediei şi abaterii standard, precum şi asupra relaţiilor dintre aceşti indicatori, întrucât deseori cunoşterea celor doi indicatori sintetici este suficientă pentru caracterizarea unui eşantion sau colecţii de cazuri, ca şi pentru plasarea unui subiect sau grup de subiecţi în distribuţia colectivităţii (eşantionului) din care fac parte.

În acest sens, este necesară prezentarea semnificaţiei abaterii standard şi a mediei aritmetice pentru cazul unei distribuţii normale. Sub acest aspect, s-a demonstrat că în cazul în care distribuţia este normală (simetrică) sau apropiată de acestă formă, atunci $m \pm 2\sigma$ (mai exact m ± 1,98σ) acoperă 95% din cazurile (subiecţii, elementele) ce alcătuiesc efectivul total *(N),* iar $m \pm 2{,}58\sigma$ acoperă 99% din cazuri. Lucrând cu multiplii întregi ai lui *σ,* relaţiiile dintre medie şi abaterea standard au semnificaţia din graficul prezentat în fig. 1.

Pentru a pune în evidenţă importanţa relaţiilor din acest grafic, se poate apela la un exemplu concret. Se presupune că la o clasă de elevi se cunoaşte media aritmetică şi abaterea standard ale notelor şcolare, acestea fiind:

m = 7,80 şi σ = 0,70

Dacă distribuţia clasei în funcţie de notele şcolare este normală (sau apropiată), atunci se pot determina, pe baza mediei, abaterii standard şi relaţiilor dintre ele, următoarele date despre clasa respectivă:

- *amplitudinea intervalului de variţie (A),* care se situează aproximativ între $m - 3\sigma$ şi $m + 3\sigma$ (deoarece $A=6\sigma$), adică în cazul de faţă, între notele 5,70 şi 9,90 întrucât:
  m - 3σ = 7,80 - 3 · 0,70 = 5,70;
  m + 3σ = 7,80 + 3 · 0,70 = 9,90.
- dacă efectivul clasei este de 36 de elevi,atunci se pot aproxima *tranşele*

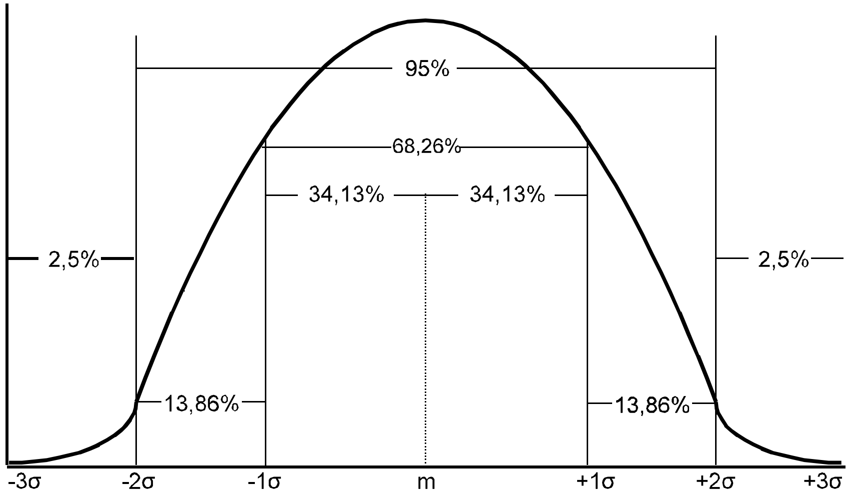

Fig. 1. Interpretarea abaterii standard

(efectivele) de elevi situați între diferitele intervale ale notelor şcolare, în modul următor:

-între 5,40 - 6,40 ≈ 1 elev (2,5%)
-între 6,40 - 7,10 ≈ 5 elevi (13,86%)
-între 7,10 - 7,80 ≈ 12 elevi (34%)
-între 7,80 - 8,50 ≈ 12 elevi (34%)
-între 8,50 - 9,20 ≈ 5 elevi (13,86%)
-între 9,20 - 9,20 ≈ 1 elev (2,5%)

Fireşte că pe baza datelor de mai sus, cunoaşterea mediei şcolare pe care un anumit elev o are, permite plasarea lui în colectivul clasei, sub aspectul nivelului la învățătură.

De pildă, despre un elev cu media 8,10 se poate spune că se situează aproximativ între al şaptelea şi al optsprezecelea elev ai clasei.

Desigur, diferențierea este mai slabă la mijlocul intervalului, dar mai puternică la extreme.

Astfel, despre un elev cu media sub 6,40 se poate afirma cu certitudine că este între ultimii 2-3 elevi din clasă.

Semnificația abaterii standard şi relațiile ei cu media aritmetică are o mare utilitate în analiza şi interpretarea datelor rezultate din aplicarea unor probe psihologice pe eşantioane reprezentative.

BIBLIOGRAFIE

1. Luduşan, N.; Voiculescu, F., *Măsurarea şi analiza statistică în ştiințele educației*, Ed. Imago, Sibiu, 1997.
2. Luduşan, N.; Bara, Monica Angela, *Statistică socială. Curs universitar*, Ed. Univ. Alba Iulia, Seria Didactica, 2006.
3. Miftode, V., *Metodologia sociologică, metode şi tehnici de cerectare sociologică*, Ed. Porto-Franco, Galați, 1995.