

CRITERII DE ALEGERE A UNEI METODE DE SONDAJ ÎN CERCETAREA STATISTICĂ

DĂNUȚ-VASILE JEMNA*

Selection criteria for a sampling method in the statistical research

Abstract

In the practice of statistical researches no one method of sampling is the best. But, to plan a statistical research, demand a selecting process for a sampling method from a set of possibilities. To obtain the most appropriate method of sampling concerns to use a set of selection criteria.

The aim of this paper is to present such a set and to emphasis some issues concerning the selection process of a sampling method using the specificity of each criterion.

We start to point up that a first set of criteria is indicating by the limits of the possibilities of statistical knowledge. Another sources of limits for selecting a sampling method are the economic restrictions and the time to achieve the research.

Key words: statistical research, sampling methods, selection criteria, statistical knowledge, efficiency

1 Introducere

Realizarea unei cercetări statistice pe bază de sondaj aduce în prim plan problema alegerii unei metode de sondaj din mai multe posibile. Lucrările clasice de teoria și practica sondajelor statistice, prezintă principalele metode de sondaj prin prisma problemelor specifice legate de eșantionare, estimare, reprezentativitate precizie [Ardilly, 1994], [Cochran, 1977]. Gruparea metodelor de sondaj se realizează în funcție de principiul de extragere a eșantionului [Ardilly, 1994, p. 48]. Clasică este următoarea clasificare: sondaje aleatoare, nealeatoare (raționate sau empirice) și mixte (complexe). Din punct de vedere practic însă, alegerea uneia dintre metode ridică dificultăți și solicită cercetătorului utilizarea unui ansamblu de criterii pentru a decide ce metodă de sondaj se va utiliza într-un caz concret anume.

Obiectivul principal al acestui articol este de a prezenta un set de criterii de alegere a celei mai adecvate metode de sondaj într-o cercetare statistică. Sintagma “cea mai adecvată metodă” trebuie înțeleasă într-un sens precis, științific, care corespunde conceptului de eficiență a cercetării statistice [D.V. Jemna, 2005].

În această lucrare, criteriile de alegere a unei metode de sondaj sunt clasificate în două categorii. Un set de criterii este determinat de exigențele teoretice și metodologice ale cunoașterii statistice, iar celălalt set este impus de limitările de ordin economic și de timp specifice unui sondaj statistic.

* Lector.doctor., Catedra de Statistică, Facultatea de Economie și Administrarea Afacerilor, Universitatea “Alexandru Ioan Cuza” Iași, e-mail: djemna@uaic.ro

2 Criterii impuse de teoria și metodologia cunoașterii statistice

O primă categorie de criterii de alegere a unei metode de sondaj este sugerată de elementele specifice cunoașterii statistice.

2.1 Restricții determinate de procesul cunoașterii statistice

Cercetarea statistică vizează realități care se manifestă sub forma unor întreguri sau colectivități de unități individuale, numite populații [Al. Bărbat, 1971, p. 27]. Metoda cercetării prin sondaj presupune cunoașterea unei astfel de realități studiind o subcolectivitate (un eșantion reprezentativ) extrasă din populație după o metodă științifică cunoscută. Obiectivul cunoașterii statistice nu este atins printr-o simplă inferență inductivă de la parte la întreg, ci prin descoperirea caracteristicilor întregului într-un proces cognitiv complex. O modalitate de a înțelege filosofia cunoașterii statistice poate pleca de la soluția lui A. Dumitriu privind problema inducției [A. Dumitriu, 1997, vol. III, p. 129]. În esență, cercetarea statistică, bazată pe inducție, se fundamentează pe puterea de a reconstitui un întreg, prin integrare, cu ajutorul unei părți anume prelevată din acel întreg. Această optică de a înțelege cunoașterea statistică impune o serie de probleme specifice cu privire la: modul de definire a populației statistice ca întreg, ca totalitate, evaluarea volumului populației, identificarea unităților individuale și a caracteristicilor acestora etc. În paralel, pentru a realiza un sondaj, se evidențiază un set de probleme care privesc: evaluarea volumului eșantionului, evaluarea gradului de precizie cu care se obțin rezultatele, specificarea condițiilor de reprezentativitate etc.

2.2 Criterii de alegere a metodei de sondaj

Din condițiile teoretice și metodologice ale cunoașterii statistice, se desprinde un prim grup de criterii care permit alegerea unei metode de sondaj. În această categorie pot fi cuprinse: natura fenomenului sau a populației statistice, informațiile disponibile despre populația de referință, natura și volumul eșantionului, precizia rezultatelor cercetării prin sondaj. Pentru fiecare criteriu se vor prezenta câteva aspecte care necesită atenție din partea specialiștilor în organizarea unei cercetări prin sondaj.

a. Natura populației de referință și posibilitățile de definire a acesteia

Populațiile statistice sunt populații reale, adică sunt finite și permit calculul volumului acestora și sunt alcătuite din unități individuale identificabile. În majoritatea cercetărilor pe bază de sondaj există posibilitatea stabilirii naturii, a caracteristicilor intrinseci, a spațiului și a timpului de manifestare a fenomenului real studiat, deci a tuturor caracteristicilor de definire a populației statistice. Există însă și situații când definirea populației ridică probleme de tipul: evaluării volumului, delimitării unităților individuale, stabilirii caracteristicilor de definire a populației sub cele trei aspecte: calitativ, în spațiu și în timp. Menționăm câteva situații în acest sens:

- În controlul calității produselor, de exemplu, populația de referință este colectivitatea produselor obținute pe fluxul de fabricație, timp de o zi sau de o lună, însă *volumul acestei populații este necunoscut a priori*.
- În studiul unor fenomene reale, cum ar fi fenomenele naturii, de exemplu, *delimitarea evenimentelor individuale* nu se poate realiza decât în momentul observării statistice.

Aceste restricții legate de definirea populației reprezintă tot atâtea restricții în posibilitatea de a alege o metodă de sondaj. Probleme ridică îndeosebi sondajul aleator, care

impune identificare unităților din populația de referință și realizarea unei liste, a unei baze de sondaj cu toate unitățile populației. În consecință, în astfel de cazuri, se optează pentru sondaje empirice sau mixte.

În practică, problemele prezentate mai sus conduc deseori la realizarea unor sondaje “pseudo-aleatoare” sau “oportunistice” [R. Cassady, 1945, p. 319]. În aceste cazuri, realizatorii sondajelor “eficientizează” munca prin renunțarea la rigoarea științifică și obțin eșantioanele alegând “la întâmplare”, fără un criteriu sau principiu de selecție, unități din populația de referință.

b. Informația disponibilă

Despre populația de referință se pot obține informații de diverse tipuri, prin intermediul mai multor surse: studii anterioare, documente, cercetări preliminare anchetei etc. Aceste informații sunt disponibile fie la nivel de unitate individuală, fie la nivelul întregii populații și pot fi utilizate în faza de obținere a eșantionului (a priori) sau după realizarea sondajului (a posteriori).

Informația disponibilă la nivel individual presupune existența bazei de sondaj și permite realizarea unor sondaje aleatoare sau mixte. La nivel global, informația suplimentară provine din rezultate anterioare sau din baze de sondaj care nu sunt accesibile în totalitate pentru organizarea anchetei. În acest caz, se pot construi sondaje empirice, cum ar fi sondajul pe cote, sau sondaje mixte, cum ar fi sondajul stratificat. Dacă se cunoaște structura populației după una sau mai multe variabile puternic corelate cu tema cercetării, eșantionul poate apărea ca o miniatură a populației totale, prin utilizarea sondajului stratificat sau a unui sondaj pe cote.

O informație suplimentară se poate utiliza și după ce a fost extras eșantionul, pentru a ameliora calitatea rezultatelor [Ardilly, 1994, p. 205]. Ameliorarea calității rezultatelor este posibilă doar în cazul sondajelor aleatoare. Decizia de a utiliza informația suplimentară în construirea unui plan de sondaj sau în faza de ameliorare a rezultatelor se ia cumulând condițiile care parvin de la toate criteriile.

c. Natura și volumul eșantionului

În realizarea unui sondaj, după definirea populației, o problemă esențială este definirea eșantionului. Această operație vizează în mod deosebit culegerea datelor, adică posibilitățile practice de a realiza observarea unităților statistice individuale. Metoda de sondaj se alege în funcție de modul în care pot fi identificate unitățile individuale ca surse de date statistice. Aceste unități pot fi simple (ființe, obiecte) sau complexe (firme, familii) și pot fi identificate prin individualizare sau prin apartenența la un grup (o serie, un cluster).

În general, pentru studiul populației nu există baze de sondaj cu unități individuale [J.M. Grosbras, 1987, p. 31]. Ca atare, obținerea eșantionului presupune, în prealabil, extragerea unor unități complexe și apoi identificarea unităților de observare, utilizând un plan de sondaj mai complex.

Stabilirea volumului eșantionului reprezintă o altă problemă delicată în realizarea unui sondaj. Se întâlnesc două situații: când volumul eșantionului este cunoscut sau fixat a priori și situația când trebuie calculat sau estimat. Volumul eșantionului este fixat a priori din rațiuni metodologice, din restricții economice sau este cunoscut din studii anterioare similare.

În practică, în cazul sondajelor empirice, nu există nici o posibilitate de calcul a volumului eșantionului. Ca urmare, acesta este fixat a priori. De exemplu, în cazul unei anchete de opinie, dacă se utilizează sondajul pe cote, volumul eșantionului este aproximat la 1000 persoane [J.M. Grosbras, 1987, p. 245].

Volumul eșantionului poate fi fixat și prin limitele impuse de buget. Dacă prin bugetul anchetei s-a stabilit un volum total al cheltuielilor (C) și un cost unitar (c) pentru realizarea anchetei la nivelul unei unități individuale, volumul eșantionului se poate calcula prin relația: $n = \frac{C}{c}$. Acest volum al eșantionului, odată stabilit, restricționează posibilitățile de alegere a unei metode de sondaj.

Calculul volumului eșantionului este posibil în cazul sondajelor aleatoare sau mixte. Dacă nu există constrângeri de cost, volumul eșantionului se determină în funcție de gradul de precizie a rezultatelor. De exemplu, pentru o variabilă de interes, în cazul în care se dorește să se estimeze o proporție sau un procent, se poate calcula volumul eșantionului care estimează parametrul cu o eroare de plus - minus 3%, pentru un nivel de încredere de 95% și în ipoteza de maximă eterogenitate a populației. În cazul unui sondaj aleator simplu, este necesar un eșantion de 1067 persoane pentru a asigura condițiile de estimare precizate. Deși pare de prisos, reluăm mențiunea că cele 1067 de persoane nu se aleg "la întâmplare", ci în condițiile riguroase ale extragerii aleatoare, utilizând baza de sondaj și un algoritm de extragere construit cu ajutorul numerelor aleatoare.

d. Precizia rezultatelor

În statistică, termenul de precizie este utilizat pentru a desemna măsura în care un rezultat statistic se apropie de realitatea comensurată. De regulă, precizia este apreciată cu ajutorul mărimii erorilor statistice. În fiecare etapă a demersului statistic se obțin erori care determină gradul de precizie a rezultatelor.

Posibilitatea măsurării erorilor reprezintă o problemă fundamentală în cunoașterea statistică. Există erori care pot fi măsurate, iar altele scapă oricărei posibilități de evaluare cantitativă. În cazul sondajelor, un rol important îl au erorile de eșantionare, care apreciază diferența dintre rezultatele obținute la nivelul eșantionului și realitatea de la nivelul populației totale. Nivelul preciziei este de interes atât pentru cei care produc datele, dar mai ales pentru cei care le utilizează. Astfel, este necesar ca toate rezultatele statistice să fie însoțite de explicații metodologice cu privire la modul de determinare a datelor, la interpretarea și posibilitățile de utilizare a acestora, dar mai ales cu privire la erorile cu care acestea au fost determinate.

În funcție de cerințele beneficiarului rezultatelor sondajului sau de condițiile științifice impuse unei anumite cercetări, apar și posibilitățile de alegere a unei metode de sondaj. Precizia rezultatelor unui sondaj poate fi apreciată distinct pentru fiecare metodă în parte. Sondajele aleatoare permit calcularea limitelor de precizie a rezultatelor, pe baza erorilor de eșantionare. Sondajele empirice permit doar aprecieri calitative asupra gradului de precizie a rezultatelor, comparativ cu alte metode sau pe marginea procedeele utilizate în realizarea anchetei.

Afirmația: "rezultatele sunt garantate cu o eroare de plus-minus 3%" este valabilă doar în cazul unui sondaj aleator sau mixt. Dacă cineva dorește un rezultat cu o precizie de acest tip nu poate apela la un sondaj empiric.

3 Criterii impuse de limitele economice și de timp

În practica sondajelor statistice, criteriul economic și cel de timp sunt esențiale, dar nu determinante. Ca orice activitate umană, cercetarea prin sondaj statistic depinde de o serie de limite stabilite într-un buget și într-un calendar de timp. Este de preferat însă ca aceste limite să fie flexibile și să se coreleze cu cele impuse de rigoarea științifică, a căror coeficient de importanță trebuie să fie mai ridicat.

În această categorie de criterii se includ costul anchetei - cu tipurile de costuri specifice: costul total și costul unitar pentru o unitate studiată – și timpul de realizare a sondajului.

a. Costul

În practică, există tendința de a realiza sondajele în condiții de cost fixate a priori. Odată fixat costul sau bugetul anchetei, se poate analiza ce tip de metodă de sondaj se poate aplica în respectivele condiții economice. Totuși, această perspectivă este una îngustă. Pot exista situații în care nici măcar condițiile științifice minimale nu pot fi acoperite de un buget fixat. În consecință, realizarea unui sondaj nu înseamnă a comanda un produs în limitele unui cost fix, ci impune un minim dialog între beneficiar și specialistul care realizează ancheta. Acesta din urmă poate aprecia tipul de sondaj cel mai potrivit pentru fenomenul considerat și poate realiza un cost estimativ.

Costul diferă de la un tip de sondaj la altul. În cazul sondajului aleator, costul este ridicat, deoarece implică un efort financiar în plus pentru: baza de sondaj, informația suplimentară, culegerea și prelucrarea datelor. Dacă se realizează un sondaj pe cote, de exemplu, costurile se pot reduce cu 20-25% sau chiar pot fi mai mici cu 50% față de cazul sondajului aleator [J. Desabie, 1966, p. 54].

Într-un sondaj, costul și precizia sunt într-o relație de interdependență directă: o precizie mai mare a rezultatelor presupune un cost mai ridicat. De regulă, în practică, trebuie rezolvată întotdeauna o dilemă: costul sau precizia? Alegerea unei metode de sondaj presupune rezolvarea acestei dileme. Uneori se stabilește un grad de precizie și o limită de cost, în funcție de care se determină un volum al eșantionului. Alteori se stabilește volumul eșantionului, un cost și apoi se determină nivelul de precizie care poate fi atins.

J. Neyman (1894-1981) a dat o soluție optimă acestei dileme prin metoda *alocării optimale* într-un sondaj stratificat [P. Ardilly, p. 86]. Această metodă presupune obținerea unui eșantion care minimizează erorile în condiții de cost date sau pentru un volum al

eșantionului fixat. Se obțin relațiile:
$$\frac{n_h}{N_h \sigma_h} = \frac{n}{\sum_{h=1}^H N_h \sigma_h} = \text{const.},$$
 unde n_h este eșantionul

extras din stratul h , N_h este volumul stratului h , care are o abatere standard σ_h , iar n este

volumul eșantionului, iar $\sum_{h=1}^H N_h = N$ este volumul populației totale.

În mod analog, dacă se consideră costul anchetei fixat dinainte, se poate realiza o alocare în funcție de costul unitar de sondaj la nivelul fiecărui strat (c_h):

$$\frac{\frac{n_h}{N_h \sigma_h}}{\sqrt{c_h}} = \frac{C}{\sum_{h=1}^H N_h \sigma_h \sqrt{c_h}} = \text{const.}, \text{ unde } \sum_{h=1}^H c_h n_h = C$$

reprezintă costul total al anchetei și este fixat.

Din relațiile de mai sus, rezultă că din fiecare strat se va extrage un sub-eșantion cu atât mai mare cu cât: volumul stratului este mai mare, dispersia stratului este mai mare, costul unitar de anchetă din fiecare strat este mai mic. Limita acestor alocări optimale este determinată de cunoașterea dispersiilor din fiecare strat, precum și a costurilor unitare din fiecare strat.

Soluția la dilema cost-precizie conduce deseori la alegerea unei metode de sondaj empirice, care presupune costuri mai reduse și un grad de precizie considerat a fi satisfăcător.

b. Timpul de realizare

Rezultatele unui sondaj trebuie obținute într-un timp cât mai scurt, pentru că aceste rezultate, de regulă, fundamentează decizii practice cu privire la fenomenul studiat.

Fiecare activitate din programul de cercetare are un loc și un timp de realizare. Cercetarea prin sondaj este un șir sau lanț de operațiuni care necesită respectarea unui calendar precis pentru fiecare operațiune și etapă în parte. În general, în practică, timpul estimat pentru un sondaj este depășit cu un procent de până la 150% [D.P. Warwick, C.A. Lininger, 1975, p.35].

În cazul unui sondaj empiric, timpul de realizare a anchetei este mai mic decât cel pentru un sondaj aleator. În cel de-al doilea caz, timpul necesar crește cu până la 10% sau 15%, datorită observării datelor, corectării non-răspunsurilor și a pregătirii anchetei. În plus, anchetatorii au un randament aproape dublu în cazul sondajului pe cote, de exemplu, comparativ cu sondajul aleator [J. Desabie, 1966, p. 54].

Problema timpului de realizare a anchetei este întemeiată, mai ales în condițiile societății actuale aflate într-o permanentă schimbare. Problematika este dublă. Pe de o parte, fără o încadrare în timp, este posibil ca la finalizarea sondajului rezultatele să nu mai aibă efectul scontat. Pe de altă parte, poate apare tentația de a alege o metodă care necesită un timp de realizare mai scurt sau de a grăbi etapele de realizare a sondajului în detrimentul calității.

4 Concluzii

Cercetarea pe bază de sondaj nu este doar o activitate științifică, ci și o activitate economică. Dintre mai multe planuri de sondaj cu un grad de precizie apropiată, de obicei, se optează pentru un plan de sondaj mai economic, dacă eforturile economice sunt semnificativ mai reduse. Însă, dacă reducerea costurilor presupune și o reducere importantă a gradului de precizie, preferabil este să se aleagă metoda mai scumpă dar cu rezultate de calitate.

Cu cât volumul eșantionului este mai mare, cu atât precizia sondajului este mai ridicată. Acest câștig de precizie prin mărirea volumului eșantionului se realizează până la un anumit nivel, după care orice mărire nu aduce contribuții substanțiale. Volumul eșantionului este limitat nu doar din rațiuni legate de precizia rezultatelor, ci și din rațiuni economice și de natura fenomenului studiat.

Costurile și rigorile științifice ale sondajelor aleatoare sunt uneori semnificative, ceea ce determină opțiunea pentru sondaje empirice, a căror rezultate nu pot fi apreciate sub aspectul preciziei decât într-o expresie calitativă. Însă, opțiunea pentru un sondaj empiric poate avea argumente bine întemeiate, precum: lipsa bazei de sondaj, lipsa informațiilor despre populație etc.

Pentru practica sondajelor statistice, se impune utilizarea criteriilor de alegere a unei metode de sondaj într-o perspectivă integrată, unitară. Deși, există înclinația de a balansa în favoarea criteriilor economice și de timp, așa cum tot practica o demonstrează, soluția este analiza tuturor criteriilor într-un sistem care să permită validarea celei mai adecvate metode. De asemenea, analiza trebuie să arate gradul în care au fost atinse exigențele cerute de fiecare criteriu în parte pentru metoda aleasă. Ansamblul acestor criterii permite un echilibru între condiția de reprezentativitate, impusă de rigoarea științifică, și eficiența economică a activității de cercetare.

Bibliografie

- Ardilly, P., *Les techniques de sondage*, Editions Technip, Paris, 1994.
- Bărbat, Al., *Teoria statisticii sociale*, Universitatea "Al.I. Cuza" Iași, 1971.
- Bobko, P., Miller, S., Tusing, R., "A Sample of Sampling Definitions", *Teaching of Psychology*, American Psychological Association, vol. 7, nr. 3, October 1980.
- Cassady, R. Jr., "Statistical sampling techniques and marketing research", *The Journal of Marketing*, American Marketing Association, vol.IX, nr.4, april 1945.
- Cochran, W.G., *Sampling Techniques, 3rd Edition*, Wiley, New York, 1977.
- Davidson, J.H., "Accuracy in statistical sampling", *The Accounting Review*, American Accounting Association, Vol. 34, No. 3 (Jul., 1959).
- Desabie, J., *Theorie et pratique des sondages*, Dunod, Paris, 1966.
- Dumitriu, Anton, *Istoria logicii*, vol. 1-3, Editura Tehnică, București, 1997.
- Grosbras, J.M., *Methodes statistiques des sondages*, Economica, Paris, 1987.
- Hitzig, Neal B., "Elements of Sampling: The Population, the Frame, and the Sampling Unit", *The CPA Journal*, New York State of CAPs, November 2004, Vol. 74 Issue 11.
- Jaba, E., *Statistica*, ediția a treia, Editura Economică, București, 2002.
- Jemna, D.V., *Eficiența sondajului statistic*, Editura Sedcom Libris, Iași, 2005.
- Morgenstern, O., *L'illusion statistique. Precision et incertitude des donnees economiques*, Dunod, Paris, 1972.
- Ruggiero M.A., "Out of sample, out of mind", *Futures Magazine*, Apr. 2005, Vol. 34 Issue 5.
- Warwick, D.P., Lininger, C.A., *The Sample Survey: Theory and Practice*, McGraw-Hill, New York, 1975.