

Statistică aplicată în științele sociale

Ediție revăzută și adăugită

Cuvânt înainte de ADRIAN NETEDU

INSTITUTUL EUROPEAN
2011

Introducere / 11

Capitolul 1. INTRODUCERE ÎN STATISTICA SOCIALĂ / 13

1.1. Rolul statisticii în științele sociale / 13

1.2. Statistica ca metodologie de tip cantitativ / 16

1.3. Metodele statisticii / 19

1.3.1. Statistica descriptivă / 20

1.3.2. Statistica inferențială / 23

1.3.3. Estimări ale populației / 25

Capitolul 2. PROBLEMA MĂSURĂRII ÎN ȘTIINȚELE SOCIALE / 29

2.1. Numărare și măsurare / 29

2.2. Despre măsurare în științele sociale / 30

2.2.1. Definiții ale măsurării / 31

2.2.2. Măsurarea și cuantificarea/operationalizarea / 32

2.2.3. Elementele măsurării / 33

2.2.4. Nivele de măsurare / 33

2.3. Schema de măsurare / 60

2.4. Problema validității / 63

2.5. Problema fidelității / 70

2.6. Pregătirea datelor pentru prelucrare / 71

Capitolul 3. EȘANTIONAREA / 75

3.1. Principiile eșantionării / 75

3.2. Reprezentativitatea eșantionului / 80

3.3. Mărimea eșantionului / 86

3.4. Estimarea parametrilor la nivelul populației / 93

3.5. Aspecte practice ale eșantionării / 96

3.6. Procedee de eșantionare aleatoare / 99

3.6.1. Eșantionarea aleatoare simplă / 95

3.6.2. Eșantionare sistematică / 101

3.6.3. Eșantionarea stratificată / 102

3.6.3.1. Eșantionarea proporțională / 104

3.6.3.2. Eșantionarea stratificată neproporțională / 106

3.6.3.3. Eșantionarea stratificată optimă / 106

3.6.4. Eșantionarea multistadială / 107

3.6.5. Eșantionarea multifazică / 107

3.6.6. Eșantioane tip cluster / 109

3.6.7. Eșantioane fixe, panel / 110

- 3.7. Procedee de eşantionare nealeatoare / 111
 - 3.7.1. Eşantionarea la întâmplare / 111
 - 3.7.2. Eşantioane de voluntari / 112
 - 3.7.3. Eşantionarea dirijată / 112
 - 3.7.4. Eşantioane pe cote / 113
- 3.8. Prezentarea unor scheme de eşantionare / 116
 - 3.8.1. Anchetă de explorare (salariaţi) / 116
 - 3.8.2. Anchetă de explorare (şomeri) / 117
 - 3.8.3. Proiectarea unui eşantion / 119

Capitolul 4. UTILIZAREA SPSS ÎN ANALIZA STATISTICĂ / 123

- 4.1. Introducere în SPSS / 123
 - 4.1.1. Baza de date / 125
 - 4.1.2. Fişier de sintaxă (syntax file) / 126
 - 4.1.3. Fişier de rezultate (output file) / 128
- 4.2. Principii de bază pentru utilizarea SPSS / 129
 - 4.2.1. Definirea şi utilizarea datelor / 130
 - 4.2.2. Definirea şi utilizarea variabilelor / 130
- 4.3. Proceduri în SPSS / 132
- 4.4. Construcţia şi definirea variabilelor în SPSS / 134
 - 4.4.1. Numele variabilei / 135
 - 4.4.2. Tipul variabilei / 136
 - 4.4.3. Etichete / 139
 - 4.4.4. Valori lipsă (missing values) / 140
 - 4.4.5. Stabilirea întinderii coloanei unei variabile / 142
 - 4.4.6. Şabloane (template) / 142
- 4.5. Editarea datelor / 144
 - 4.5.1. Introducerea liniilor sau coloanelor în tabel / 145
 - 4.5.2. Copierea sau mutarea datelor în tabel / 145
 - 4.5.3. Ştergerea rândurilor (cazuri) sau coloanelor (variabile) / 146
 - 4.5.4. Selectarea în cazuri particulare / 146
 - 4.5.5. Setarea opţiunilor programului / 147
- 4.6. Unirea fişierelor de date / 149
 - 4.6.1. Situaţii în care este necesară unirea fişierelor / 150
 - 4.6.2. Adăugarea Cazurilor (Adding Cases) / 150
 - 4.6.3. Adăugarea variabilelor (add variables) / 153
 - 4.6.4. Adăugarea variabilelor cu date incomplete / 154

Capitolul 5. OPERAŢII EXECUTATE ASUPRA DATELOR / 157

- 5.1. Ierarhizarea şi sortarea datelor / 157

- 5.1.1. Ierarhizarea scorurilor / 157
- 5.1.2. Așezarea cazurilor într-o ordine diferită: sortarea cazurilor / 159
- 5.2. Schimbarea grupării și codării datelor / 160
 - 5.2.1. Transformarea/crearea variabilelor cu ajutorul comenzii RECODE / 160
 - 5.2.2. Crearea unei variabile fictive (dummy) / 164
 - 5.2.3. Răspunsuri multiple / 165
 - 5.2.4. Crearea unor noi variabile pe baza formulelor matematice (comanda COMPUTE) / 168
- 5.3. Situații speciale / 170
 - 5.3.1. Facilitățile Functions / 170
 - 5.3.2. Valori lipsă în Compute / 171
 - 5.3.3. Cum se obțin scorurile z / 172
 - 5.3.4. Distrugerea datelor / 173
- 5.4. Selecția de subgrupuri pentru analiză / 174
 - 5.4.1. Selecția subgrupurilor bazate pe fiecare nivel al unei variabile: Split File / 174
 - 5.4.2. Compare Groups sau Organize Output by Groups în Split File / 176
 - 5.4.3. Selectări de subgrupuri folosind Select Cases / 177

Capitolul 6. ANALIZA UNIVARIATĂ A DATELOR / 183

- 6.1. Indicatori de poziție / 183
- 6.2. Mărimi multiple / 188
- 6.3. Indicatori de dispersie / 194
- 6.4. Curba distribuției normale / 205
- 6.5. Scorurile Z / 207
- 6.6. Distribuția frecvențelor / 208
- 6.7. Reprezentarea grafică a datelor / 215
 - 6.7.1. Histogramele din meniul FREQUENCIES / 215
 - 6.7.2. Grafice tip bară din meniul FREQUENCIES / 216
 - 6.7.3. Grafice din meniul Explore / 218
 - 6.7.4. Grafice din meniul GRAPHS / 219
- 6.8. Analiza datelor măsurate pe scală nominală / 222
- 6.9. Analiza datelor măsurate pe scală ordinală / 225
- 6.10. Analiza datelor măsurate pe scală de interval sau raport / 227

Capitolul 7. ANALIZA BIVARIATĂ / 231

- 7.1. Analiza variabilelor măsurate nominal sau ordinal / 231
 - 7.1.1. Grafice cu două variabile / 231
 - 7.1.2. Tabele de contingență pentru variabile nominale

și ordinale / 234

- 7.1.3. Independența într-un tabel de contingență / 238
- 7.1.4. Asocierea într-un tabel de contingență / 240
- 7.1.5. Coeficienți de asociere pentru variabile nominale / 242
 - 7.1.5.1. Șansa și raportul de șanse / 244
 - 7.1.5.2. Coeficientul ϕ / 246
 - 7.1.5.3. Coeficientul V a lui Cramer / 247
 - 7.1.5.4. Coeficientul de contingență C a lui Pearson / 247
 - 7.1.5.5. Coeficientul λ , / 248
- 7.1.6. Coeficienți de asociere pentru variabile ordinale / 250
- 7.1.7. Testul χ^2 / 254
- 7.1.8. Valoarea reziduală standardizată și ajustată / 256
- 7.2. Analiza variabilelor măsurate la nivel de interval sau raport / 257
 - 7.2.1. Grafice realizate între variabile măsurate la nivel de interval sau raport / 257
 - 7.2.2. Coeficienți de asociere între variabile măsurate la nivel de interval sau raport / 258
- 7.3. Analiza unei variabile nominale/ordinale și a unei variabile măsurată la nivel interval/raport. / 259
 - 7.3.1. Grafice utilizate pentru o variabilă nominală/ordinală și una de interval/raport / 259
 - 7.3.2. Tabele utilizate pentru o variabilă nominală/ordinală și una de interval/raport / 261
 - 7.3.3. Coeficienți de asociere utilizați pentru o variabila nominală/ordinală și una de interval/raport / 261
- 7.4. Testarea ipotezelor / 263
 - 7.4.1. Ipotezele nule / 263
 - 7.4.2. Nivelul de semnificație / 263
 - 7.4.3. Tipuri de erori / 267
 - 7.4.4. Interpretarea rezultatelor testelor de semnificație / 269
 - 7.4.5. Puterea / 270
 - 7.4.6. Semnificația statistică a diferenței dintre medii / 272
- 7.5. Testarea ipotezelor statistice / 276
 - 7.5.1. Logica testării statistice / 276
 - 7.5.2. Gestionarea erorilor / 277
 - 7.5.3. Tipuri de teste / 278
- 7.6. Teste de normalitate / 278
- 7.7. Teste parametrice / 281
 - 7.7.1. Testul t pentru un eșantion / 282
 - 7.7.2. Testul t pentru eșantioane independente / 283
 - 7.7.3. ANOVA (Analiza de varianță) / 285

7.7.4. Testul t pentru eșantioane dependente (Paired samples t -test) / 287

7.8. Teste neparametrice / 287

7.8.1. Teste neparametrice pentru 2 eșantioane independente / 288

7.8.1.1. Testul U propus de Mann-Whitney / 288

7.8.1.2. Testul Z propus de Kolmogorov-Smirnov / 288

7.8.1.3. Testul Wald-Wolfowitz / 289

7.8.2. Teste neparametrice pentru $k > 2$ eșantioane independente / 290

7.8.2.1. Testul Kruskal Wallis / 290

7.8.2.2. Testul medianei / 291

7.8.3. Teste neparametrice pentru 2 eșantioane dependente / 291

7.8.3.1. Testul Semnului (Sign Test) / 291

7.8.3.2. Testul Wilcoxon / 292

7.8.3.3. Testul McNemar / 293

7.8.4. Teste neparametrice pentru $k > 2$ eșantioane dependente / 293

7.8.4.1. Testul Friedman / 293

7.8.4.2. Testul W a lui Kendall / 294

7.8.4.3. Testul Q a lui Cochran / 294

Capitolul 8. ANALIZA MULTIVARIATĂ / 297

8.1. Regresia lineară / 297

8.1.1. Regresia și corelația / 297

8.1.2. Valori așteptate (prezise) și observate ale lui Y / 306

8.1.3. Coeficienții de regresie / 306

8.1.4. Intervalul de încredere pentru coeficientul de regresie / 307

8.1.5. Coeficientul de determinație / 308

8.1.6. Coeficientul de corelație / 309

8.1.7. Condiții de aplicare a regresiei / 310

8.1.8. Etape în construcția unui model de regresie / 313

8.2. Regresia logistică / 321

8.2.1. Interpretarea coeficienților regresiei logistice / 322

8.2.2. Indicatori ai modelului de regresie logistică / 323

8.2.3. Exemplu comentat de fișier de rezultate SPSS (Barometrul de opinie publică, iunie 1998) / 325

8.3. Analiza logliniară / 328

8.3.1. Șansa / 329

8.3.2. Raportul de șanse / 329

8.3.3. Șansa condiționată / 330

8.3.4. Riscul relativ / 330

8.3.5. Șansele locale (Local odds ratios) / 332

8.3.6. Modelarea log liniară / 334

8.3.7. Decompoziția unui model cu 3 variabile. Interacțiuni de ordinul 2 și 3 / 336

8.3.8. Modele saturate/ nesaturate / 338

8.3.9. Selectarea unui model potrivit / 339

8.4. Analiza factorială / 343

8.4.1. Noțiuni introductive / 343

8.4.2. Factori și variabile / 344

8.4.3. Variație, covariație și corelație / 346

8.4.4. Combinațiile liniare și derivațiile variației și covariației / 348

8.4.5. Derivația structurii covariației din structura factorială / 349

8.4.6. Obținerea soluțiilor analizei factoriale / 359

8.4.7. Obținerea analizei factoriale în SPSS / 361

Capitolul 9. CONSTRUCȚIA INDICILOR

(Coautor: lect. drd. Meseșan Schmitz Luiza) / 369

9.1. Definirea indicilor / 369

9.2. Tipuri de indici / 371

9.3. Construirea indicilor / 373

9.4. Construirea indicilor cu ajutorul programului SPSS / 375

9.4.1. Construcția unui indice de numărare / 375

9.4.2. Indice sumativ sau de tip medie / 378

9.4.3. Indice ca raport de scoruri / 384

9.4.4. Indici cu formule complexe / 385

9.4.5. Construirea unui indice de tip scor factorial / 387

BIBLIOGRAFIE SELECTIVĂ / 391

*Capitolul 1***Introducere în statistica socială****1.1. Rolul statisticii în științele sociale**

Atunci când vorbim de statistică ne referim la o serie de metode subsumate conceptului general, metode care sunt utilizate pentru a descrie, a reconfigura și a analiza date. Datele trebuie să fie numerice pentru a putea fi prelucrate sau să poată fi aduse sub o formă numerică.

Sunt destule situații în care interesele științifice ale statisticienilor și sociologilor coincid. De multe ori cei care se formează ca sociologi nu au la bază o pregătire specifică de natură matematică iar acest lucru le poate crea probleme în utilizarea metodelor statistice de analiza a datelor. Chiar dacă în practică se poate constata că mulți sociologi care sunt preocupați de analiza cantitativă și implicit de analiza statistică au și o formație matematică solidă, trebuie să reținem că ne putem descurca foarte bine în statistica socială chiar dacă nu stăpânim un aparat matematic extrem de sofisticat. Iar acest minim efort trebuie făcut deoarece chiar dacă „o să ne facem un prieten” care este statistician, acesta cu greu va pricepe ce vrem de la el dacă nu are cunoștințele necesare de sociologie. Așa cum o să vedem astfel de combinații de obicei nu duc la nimic bun și se potențează apariția unor excесе ale statisticii în interpretarea fenomenelor sociale. Ceea ce face din statistica socială un demers unic și de neînlocuit este preocuparea sa pentru „reprezentarea numerică a vieții sociale, mai bine zis pentru lucruri ce pot fi numărate, măsurate, cuantificate” (N. R. Kurtz, 1982).

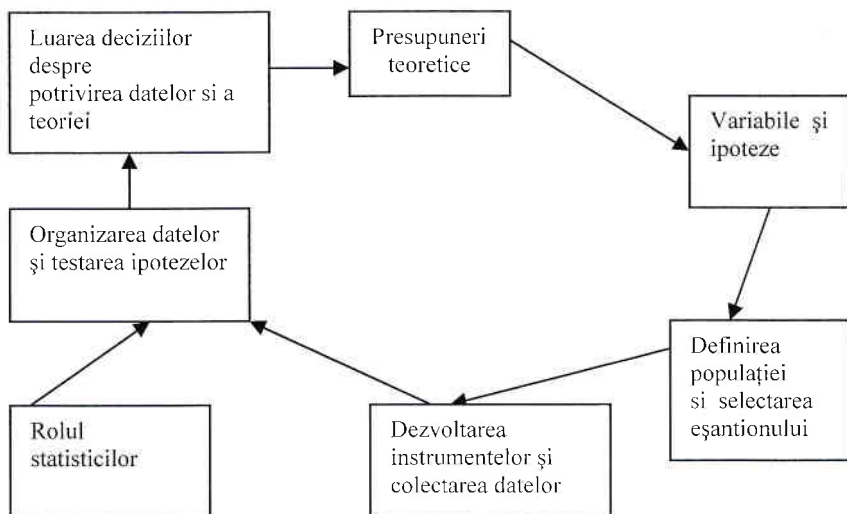
Statisticile sunt folosite în mai multe situații, în funcție de context având roluri diferite, cum ar fi:

1. rol de descriere – se poate arăta care este venitul mediu al angajaților din industria brașoveană;
2. rol de relație – se poate evidenția o legătură între mărimea venitului și sexul persoanelor angajate.

3. rol de corespondență – se poate arăta că există o corespondență între venit și atitudinea față de viitor.

Se poate spune că orice cercetare ne pune în situația de experiment. Definiția corectă a cercetării de tip cantitativ trebuie să țină seama de două aspecte: numărarea și măsurarea. (T. Rotariu, P Iluț, 1999).

Figura nr. 1.1: Rolul statisticii în cercetările sociale



Sursa: Kurtz Norman R. (1982), *Introduction to social statistics*, McGraw-Hill, Boston, p. 17

Pentru a constata relația dintre sociologie și statistică trebuie să plecăm chiar de la definirea sociologiei ca știință. Astfel, sociologia se supune următoarelor reguli:

1. obiectul științei este generalul, deci și obiectul sociologiei este tot generalul;
2. acest general este totdeauna, pentru oricine același; desigur că nu se schimbă din cauza vremii sau a culorii ochilor;
3. forma legilor specifice cu ajutorul cărora poate fi interpretată realitatea socială se înscrie în niște parametri bine definiți;
4. metodele științifice sunt aceleași pentru toate disciplinele așa că este foarte important să reținem că rigoarea demersului sociologic trebuie să se înscrie în rigoarea oricărui alt demers științific;
5. știința trebuie să explice, deci să deducă efectul din cauză, ceea ce se transformă în principalul atribut al sociologiei la baza căreia stă analiza causală;
6. ideea acceptării unei realități sociale unice, la fel pentru toți. Există totdeauna o raportare între cercetător și realitatea socială. Cercetătorul este

- întotdeauna în interiorul sistemului de referință, dar trebuie să aibă grijă să nu fie influențat și să nu îl influențeze;
7. reproductibilitatea – adică metoda să poată fi aplicată pe altă unitate socială;
 8. verificabilitatea – să existe posibilitatea ca datele care sunt valide pentru o populație să fie valide și pe o subpopulație inclusă în aceasta;
 9. testabilitatea – metoda să poată fi aplicată și în alte coordonate spatio-temporale.

Plecând de la cele spuse anterior și de la schema din figura nr. 1.1. putem conchide că în cercetarea socială, analiza statistică are un rol bine determinat. La prima vedere ar părea că prelucrările statistice joacă un rol relativ minor în cadrul întregului proces de cercetare.

Acest rol este de fapt mult mai important decât pare la o privire superficială întrucât este un fapt dovedit de nenumărate cercetări că rezultatele testelor statistice arată de obicei dacă presupunerea, propoziția sugerată de speculațiile teoretice, merită atenție și munca de cercetare a respectivei probleme trebuie continuată sau din contră trebuie abandonată. În consecință, rolul analizei statistice poate fi decisiv, căci el poate determina direcția viitoare a speculațiilor teoretice cât și a cercetării.

Datorită naturii duale a realității sociale, a faptului că atunci când se discută despre ceea ce a măsurat o anchetă se face o interpretare a unei alte interpretări este normal să ne gândim că este posibil ca rezultatele prezentate să fie influențate de diferite tipuri de erori. Acestea sunt generate de modul în care încercăm să măsurăm și atunci când se vorbește despre posibilele erori se insistă în general pe erorile de eșantionare sau pe cele datorate operatorilor („Operatorul constituie cel mai important factor generator de eroare, în cazul anchetelor orale”, conform T. Rotariu și P. Iluț, 1997) sau pe erorile legate de arhitectura chestionarului.

Poate că ar trebui să ne gândim la o critică a epistemologiei în genere și în ceea ce privește utilizarea ei în analiza problemelor sociale, la o analiză și îndepărtare a ideii de obiectivitate în privința adevărului și realității. Ar fi interesant să ieșim din schemele noastre de gândire, parțial reducționiste și să admitem că sunt doar superstiții. De asemenea am putea renunța la îndepărtata relație dintre subiect și obiect, a celei dintre gândire și realitate, în general la dualismele și la modelele epistemice care universalizează structuri de limbaj și interpretare. Deci, ne-am orienta către o gândire contextuală și un relativism pragmatic, fără obsesia ultimului context, a corespondenței reprezentării cu o realitate obiectivă și ne-am referi la scopurile urmărite de actorii sociali în situații concrete. Renunțarea la ideologia comensurabilității înseamnă să acceptăm succesiunea și firescul întâmplărilor care pot fi interpretate și pot semnifica altfel o situație ce nu părea să ridice semne de întrebare. Uneori „ceea ce

trebuie” poate fi o pură întâmplare iar ceea ce socotim că este „mărunt” pentru pretențiile noastre poate fi destul de semnificativ.

Cu atât mai mult atunci când se încearcă măsurarea unor comportamente sau atitudini ale unor persoane (noi îi vom numi de aici înainte indivizi statistici) dificultățile practice devin și mai mari. În primul rând pentru că nu se poate măsura direct ci doar estima. Apoi, nu trebuie să uităm, că în multe cazuri măsurarea repetată nu aduce un plus semnificativ de informație datorită volatilității opiniilor subiecților ceea ce face puțin probabil realizarea unui model de predicție. Experiența practică a demonstrat că de multe ori ajustările și ponderările s-au făcut în funcție de intuiția cercetătorului renunțându-se la estimările obținute pe baza modelului. În sfârșit, dificultatea analizei crește în cazul sondajelor de opinie politică care reprezintă un important atribut al comunicării politice, iar în momentul în care principala strategie adoptată la nivelul întregului spectru politic a avut serioase conotații persuasive, ca să mă exprim elegant și să nu spun manipulatorii, este greu să se facă distincția între erori intenționate și erori neintenționate. Într-un asemenea context, indiferent de tipul de eroare, cred că ar trebui să redimensionăm importanța acordată analizei acestora pentru că pot afecta calitatea datelor, respectiv validitatea și fidelitatea măsurării.

1.2. Statistica ca metodologie de tip cantitativ

Din punctul de vedere al dualismului metodologic, care face o distincție clară între abordările calitative și cele cantitative, statistica se leagă de partea cantitativă a demersurilor metodologice din sociologie.

Statistica este așadar o metodologie de tip cantitativ. Există în cadrul statisticii noțiuni importante precum: indicatori statistici, analiză univariată, analiză multivariată și eșantionare.

Statistica ca disciplină științifică se ocupă de fabricarea unor instrumente de producere a datelor care se extind pe un număr de entități ce formează așa numita *populație statistică*.

Metodele statistice au ca punct de plecare diversitatea indivizilor, a entităților cu care lucrează. Aceste entități se manifestă diferit și ele trebuie evaluate în funcție de variabilele care le caracterizează. Se lucrează cu:

- populații statistice;
- individ statistic;
- variabile statistice (caracteristici ale indivizilor).

Metodele statistice au avantajul că reușesc să capteze manifestările individuale ale indivizilor și diferențele individuale se regăsesc pe plan general.

Din punct de vedere matematic comportamentul indivizilor se explică printr-o funcție $f(x)$.

Statistica are un corp comun de tehnici și procedee care sunt valabile indiferent de natura variabilelor cu care lucrează. Statistica nu are un obiect empiric de cercetat, ea este o știință exclusiv metodologică pentru că prescrie niște modalități de lucru cu anumite fenomene statistice și nu se leagă de conținut.

Conținutul și interpretarea indicatorilor statistici nu țin de domeniul statisticienilor ci de cel al sociologilor. Statistica s-a dezvoltat în corelație strânsă cu științele sociale. Termenul de statistică a fost elaborat în germană și desemna știința chemată să ajute la conducerea treburilor publice ale statului.

Fără a intra în amănunte trebuie să reținem că în a doua jumătate a secolului al XIX-lea apar problemele corelației și asocierii multivariate care implică dezvoltarea modelelor de corelație și regresie ca și problema a cercetărilor selective.

După 1960 s-au dezvoltat programe și pentru analiza variabilelor de tip calitativ. Utilizarea programelor de calculator a permis aplicare unor metode pe variabile calitative, precum: analiza de corespondență, analiza loglineară, analiza discriminativă multiplă.

În prezent există diferite programe, cum ar fi SPSS, care pot fi folosite pentru a realiza funcțiile matematice cerute pentru analiza statistică. Din fericire, softul respectiv ne scutește de calcule complicate și minuțioase, dar cunoștințele de bază de matematică sunt totuși indispensabile pentru înțelegerea scopului și rolului procedurilor statistice.

Pentru început vom încerca să clarificăm noțiunea de variabilă, ca un prim pas în inițierea privind lucrul cu softurile specifice analizei statistice. Variabilele, adică orice mărime care se modifică, se pot clasifica în două moduri de bază diferite. Prima clasificare, așa cum am văzut, în urma distincției pe care am făcut-o, împarte variabilele în variabile **calitative** sau **cantitative** (dacă spunem mama este foarte grasă este vorba de o variabilă calitativă, iar dacă spunem mama are 1m și 60 de cm. este vorba de o variabilă cantitativă).¹

A doua clasificare se face între variabile **discontinue** și variabile **continue**. O variabilă discontinuă este o variabilă care există numai în clase sau valori distincte. De exemplu, dacă dorim să ne cumpărăm o pereche de pantofi, vom sesiza că avem un număr limitat de opțiuni cum ar fi 36, 37, 45 și nu putem lua o pereche de pantofi care are mărimea 36,3 sau 36,4. Cu alte cuvinte acest tip de variabilă ia doar anumite valori. În schimb o variabilă continuă există într-un număr infinit de valori pe un anumit interval (de exemplu vârsta unei

¹ Pentru cei interesați amănunte suplimentare despre distincția cantitativ-calitativ și despre numărare și măsurare se pot găsi în T. Rotariu, P. Iluț, *Ancheta sociologică și sondajul de opinie*, Editura Polirom, 1997.

persoane trece printr-un număr infinit de valori diferite de la naștere până la moarte, de la 0 la 104 ani spre exemplu).

Ținând cont de cele două clasificări putem grupa variabilele în patru categorii. În tabelul nr. 1.1. vom prezenta această clasificare mixtă.

Tabelul nr. 1.1: *Tipuri de variabile*

Variable	Discontinue	Continue
Calitative	Aspectul unei mașini Numele șoferului mașinii	Culoarea mașinii
Cantitative	Numărul roților mașinii Numărul locurilor mașinii Numărul ușilor mașinii	Lungimea mașinii Viteza mașinii Vechimea mașinii

Statistica s-a constituit pe două direcții:

- parametrică (analiza variabilele cantitative);
- nonparametrică (analiza variabilele calitative).

Analiza statistică nu poate din păcate să ne răspundă la toate întrebările. De multe ori cercetătorii pot cădea în capcana „relațiilor statistice” uitând că este vorba de un simplu instrument care poate stabili asocieri între date, fără ca în realitate să existe vreo relație între cele două fenomene și astfel pot să apară aberații ca în celebra corelație dintre numărul de berze dintr-o localitate și numărul de nou născuți din respectiva localitate. S-a stabilit statistic o corelație dintre numărul de berze dintr-o localitate și numărul de copii născuți. Conform corelației ar reieși că berzele ar putea fi responsabile de nașterea copiilor dintr-o localitate.

Un alt exces al statisticii este încercarea de a explica toate variațiile care au loc între diferite fenomene cu ajutorul indicatorilor statistici.

Alte excese statistice sunt:

- încercarea de a măsura aproximativ;
- suprasimplificarea fenomenelor;
- prezentarea trunchiată a fenomenelor, scoasă din context;
- abuzul de formule și de teste statistice;
- încercarea de a măsura ceea ce nu se poate măsura;
- exagerarea intensității unei corelații (de exemplu o corelație de 0,1 deși semnificativă statistic, relevă că doar o sutime din variația unei variabile este comun împărțită de cealaltă variabilă).

Din observațiile de până acum deducem că înțelegerea realității se poate obține îmbinând cunoașterea sociologică cu cea statistică. Nu trebuie uitat că statistica este doar un instrument care ne ajută în analiza fenomenelor sociale.

1.3. Metodele statisticii

Metodele statisticii pot fi împărțite în două categorii: statistica descriptivă și statistica deductivă sau inferențială. Dacă ne raportăm însă la faza în care vorbim de analiză statistică în economia unei cercetări sociologice putem să facem o discriminare mai fină.

Când ajungem la etapele finale ale anchetei, datele obținute de la/sau despre persoanele anchetate au fost redactate și codificate, iar primii pași spre sintetizare, care constau de obicei din construirea tabelelor de bază, au fost făcuți. Ceea ce urmează este, din mai multe puncte de vedere, sarcina care pretinde cea mai mare calificare: analiza și interpretarea rezultatelor. Desigur, este o sarcină care solicită raționamentul și priceperea cercetătorului și nu trebuie lăsată pe seama unui personal care nu este calificat. Activitatea curentă de analiză poate să nu fie dificilă, dar conducerea ei corectă și interpretarea care o însoțește pretind o cunoaștere adâncită a fundamentării anchetei și a tuturor stadiilor ei pe care o are probabil numai conducătorul anchetei. Este încă un argument în favoarea ideii ca sociologii să fie calificați în ceea ce privește toate etapele anchetei, inclusiv introducerea datelor în calculator și prelucrarea acestora.

Discutarea fazelor finale ale unei anchete în termeni generali este extrem de dificilă. Nu există două anchete asemănătoare prin problemele lor de analiză și interpretare, iar singurele reguli care se pot stabili cu o oarecare certitudine se referă, în sens strict, la metodele statistice.

Din acest punct de vedere, putem distinge patru aspecte diferite ale acestor metode:

(1) Analiza materialului anchetei nu trebuie să aibă neapărat un caracter statistic. În măsura în care interesul se concentrează asupra cazurilor individuale mai mult decât asupra caracteristicilor ansamblului, metodele calitative de analiză și evaluare sunt preferabile.

(2) O mare parte din ceea ce se numește de obicei analiză depinde de prelucrarea distribuțiilor statistice, de construirea graficelor și de calcularea indicatorilor simpli, cum sunt mediile, dispersiile, procentajele, coeficienții de corelație etc. Pentru că termenul de analiză este prea pretențios pentru a numi această etapă se consideră că „descriere statistică ar fi un termen mai potrivit, deoarece tot ceea ce se face este să se descrie caracteristicile agregatului anchetat” (C. A. Moser, 1967).

(3) Descrierea statistică, în acest sens, este însă numai o parte a analizei anchetei. Inferența, în sens larg, este cealaltă parte. De fiecare dată când datele anchetei se bazează pe un eșantion din populație despre care urmează să fie trase concluzii se recurge la un anumit tip de inferență. Cercetătorului i se pune

problema să estimeze caracteristicile populației din datele eșantionului precum și să estimeze erorile de eșantionare.

(4) Probleme de inferență apar, de asemenea, într-un context diferit. Una este să extragi o măsură a corelației dintre două variabile cuprinse de anchetă și cu totul altceva este să cauți prin metode analitice o explicație a modului cum ia naștere această relație demonstrată. Utilizarea de tehnici statistice complicate în anchete provine adesea din dorința de a stabili și a interpreta relații care variază după mai multe caracteristici.

1.3.1. Statistica descriptivă

Statistica descriptivă înseamnă exact ceea ce sugerează numele. Este vorba de metode pentru descrierea fenomenelor sociale care furnizează descrieri numerice. Aceste descrieri nu trebuie făcute la voia întâmplării. Ne aflăm într-o situație similară cu cea de la observația participativă. Dacă nu urmăm un set de reguli și proceduri nu putem spune că am făcut o descriere adecvată. Aceasta trebuie subscrisă obiectivelor pe care le urmărim.

Să presupunem că trebuie să descriem grupa din care face parte o persoană în funcție de trei caracteristici de structurare: sex, vârstă și venit. Scopul acestei descrieri ar fi acela de a înțelege mai bine fenomenele care au loc în grupul respectiv și care ar putea fi influențate de discriminări de gen sau de diferențe foarte mari de vârstă între membrii grupului. De asemenea venitul ar putea constitui un factor important de segmentare dacă am constata că acesta nu este distribuit omogen la nivelul grupului ceea ce desigur ar genera așteptări și expectanțe total diferite ale membrilor acestuia. Fiecare dintre cele trei caracteristici este de fapt în limbaj sociologic o variabilă, adică fiecare reprezintă o entitate care poate lua valori diferite pentru oricare dintre persoane. Așa cum o să vedem, sexul sau genul este o variabilă care poate lua o valoare din două posibile: masculin sau feminin. Vârsta poate lua valori începând de la o zi până la mai mult de o sută de ani. De asemenea, venitul poate lua valori pe o scară foarte largă de la venituri mici sau lipsa venitului până la venituri foarte mari.

Pornind de la clasificarea variabilelor realizată în tabelul nr 1.1, variabilele sex, vârstă și venit se încadrează astfel:

Tabelul nr. 1.2: *Tipuri de variabile folosite în statistica descriptivă*

Variabile	Discontinue	Continue
Calitative	Sex	
Cantitative		Vârsta Venit