

STATISTICĂ

Concepte, tehnici și instrumente softwaRe

Cuprins

Despre autori	13
Introducere	15
Bookdown	17
Informații despre software	17
1 Concepte de bază în statistică	19
1.1 Ce este statistica?	19
1.2 Începutul demersului statistic	20
1.3 Mărimi relative	25
1.3.1 Mărimi relative de structură	27
1.3.2 Mărimi relative de intensitate	29
1.3.3 Mărimi relative de coordonare	30
1.3.4 Mărimi relative ale dinamicii și ale planului	31
2 Sistematizarea datelor	33
2.1 Metode și procedee de sistematizare	33
2.2 Serii de distribuție	36
2.3 Indicatorii frecvenței	43
2.4 Histograma	48
3 Vizualizarea datelor	53
3.1 Reprezentarea datelor sub formă de tabele	53
3.2 Reprezentarea grafică a datelor	54
4 Indicatorii de nivel. Tendința centrală	63
4.1 Indicatorii tendinței centrale	64
4.1.1 Media	64
4.1.2 Mediana	77
4.1.3 Modul	83
4.2 Quantilele	86

4.2.1	Quartilele	86
4.3	Aplicații - în R	87
5	Indicatorii variației	95
5.1	Indicatorii variației	95
5.1.1	Indicatorii simpli ai variației	95
5.1.2	Indicatorii sintetici ai variației	99
5.2	Indicatorii formei de repartiție	106
5.2.1	Indicatorul de asimetrie - Skewness	107
5.2.2	Indicatorul de boltire - Kurtosis	108
5.3	Indicatorii concentrării	109
5.3.1	Curba lui Lorenz	110
5.3.2	Coeficientul Gini	113
6	Distribuții de probabilitate	115
6.1	Conceptul de probabilitate	115
6.2	Distribuții de probabilitate	116
6.2.1	Distribuții discrete	116
6.2.2	Distribuții continue	129
7	Inferența statistică	145
7.1	Parametri de selecție	146
7.2	Componentele cercetării statistice selective	151
7.2.1	Selecția	151
7.2.2	Extinderea rezultatelor de selecție	154
7.2.3	Mărimea eșantionului	173
8	Indicii statistici	177
8.1	Indicii individuali	178
8.2	Indicii de grup	180
8.2.1	Indicele agregat al variabilei complexe	180
8.2.2	Indicii agregați ai factorilor	182
9	Statistici oficiale	191
9.1	Statistica oficială în România	191
9.2	Statistica oficială în Uniunea Europeană	194
9.3	Surse de date în statistica oficială	194
9.4	Diseminarea informației statistice	197
ANEXA 1 Distribuția Bernoulli		201
ANEXA 2 Distribuția Poisson		205

ANEXA 3 Distribuția Normală Standard

211

ANEXA 4 Distribuția t-Student

213

Capitolul 1

Concepte de bază în statistică

1.1 Ce este statistica?

Termenul de “statistică” este unul familiar, pe care îl auzim sau folosim în fiecare zi. Este asociat adesea cu calcule, cu un rezultat exprimat procentual sau cu un tabel de date. Cu toate acestea, termenul trebuie privit într-o accepțiune mai largă.

Poate v-ați întrebat deseori care este valoarea medie a cifrei de afaceri sau care a fost creșterea medie a acesteia într-un interval de timp, ce valoare a vânzărilor ar putea avea o firmă de telefonie mobilă care ar lansa pe piață un nou produs sau serviciu, sau care ar fi cel mai potrivit loc de muncă disponibil când veți absolvi facultatea. Cine va fi ales noul președinte al țării la următoarele alegeri prezidențiale? Sau poate vă întrebați care ar fi contribuția creșterii nivelului de educație al populației ocupate la dezvoltarea economică a țării.

Răspunsurile la astfel de întrebări, dar și la multe altele, vin din înțelegerea corectă a valorilor numerice înregistrate de o variabilă căreia îi vom spune generic “de interes”, a fluctuațiilor sau a tendinței de creștere sau de scădere a acestor valori într-o anumită perioadă de timp.

Conform dicționarului publicat de institutele de cercetare din România, **statistica** este o știință care, folosind calculul probabilităților, studiază fenomenele și procesele de tip colectiv (din societate, natură etc.) din punct de vedere cantitativ în scopul descrierii acestora și a descoperirii legilor care guvernează manifestarea lor¹. Altfel definită, statistica este știința care se ocupă cu descrierea și analiza numerică a fenomenelor de masă, dezvăluind particularitățile lor de

¹Academia RPR *Dicționar Enciclopedic Român*. București: Editura Politică, 1962-1966; Academia Română, Institutul de Lingvistică, Iorgu Iordan, *Dicționarul explicativ al limbii române* (DEX), București: Editura Univers Enciclopedic, 1998

*volum, structură, dinamică, conexiune, precum și regularitățile sau legile ce le guvernează*².

În practică, metodele statistice furnizează multor domenii ale științei un puternic set de instrumente pentru analiza datelor și pentru interpretarea semnificației rezultatelor cercetărilor statistice.

Sensul termenului statistică poate avea mai multe conotații. Una dintre ele este aceea de *date* sau *colecție de date*, referitoare la un anumit fenomen sau domeniu. Într-o altă accepțiune, statistica înseamnă activitatea de *producere a datelor statistice*, de culegere, prelucrare și prezentare a informației statistice sub diverse forme (diseminare), la dispoziția utilizatorilor. Mai mult, termenul de statistică se utilizează și pentru a desemna *metodele statistice aplicate* în cunoașterea fenomenelor de masă, sau teoria statistică, implicând expunerea sistematică a conceptelor și metodelor de cercetare statistică.

Statistica are ca *obiect* studierea aspectelor cantitativ-numerice ale *fenomenelor de masă*, a dimensiunii, dinamicii și structurii acestora, a raporturilor de *interdependență* și a altor aspecte care pot fi caracterizate numeric. Fenomenele de *masă* sunt fenomene care se produc sub acțiunea comună și repetată a unui număr *mare de factori*, cu caracter sistematic sau întâmplător. Fenomenele de masă sunt caracterizate printr-o mare diversitate, formele individuale de manifestare diferind de la o unitate la alta în funcție de modul în care se combină acțiunea acestor factori. Esența acestor fenomene poate fi pusă în evidență numai prin studierea unui număr mare de cazuri. De exemplu, productivitatea muncii, la nivelul economiei naționale, poate fi cunoscută numai prin analiza informațiilor privind rezultatele activității tuturor unităților din economia națională, respectiv cantitatea de forță de muncă utilizată (dintr-o ramură de activitate sau dintr-o unitate teritorială) sau a unei părți semnificative a acestora.

Metoda statistică este constituită dintr-un ansamblu de operații, tehnici, procedee și metode de cercetare statistică. Specificitatea metodei statistice constă în faptul că se bazează pe observarea unui număr mare de cazuri și utilizează instrumente derivate din teoria probabilităților, dezvoltate în cadrul statisticii matematice.

1.2 Începutul demersului statistic

Înainte de a începe studiul statisticii este necesară o privire de ansamblu asupra termenilor cu care operează statistica. Este cunoscut faptul că, pentru organizarea unei cercetări corecte și cuprinzătoare, este necesar să se folosească un limbaj științific specific disciplinei. Acest lucru este valabil și pentru statistică, care a reușit să-și elaboreze propriile noțiuni, concepte de bază pe care să le folosească pe parcursul întregului demers statistic, de la stabilirea obiectivelor cercetării și

² *Mică enciclopedie de statistică*, Editura Științifică și Enciclopedică, București, 1985

până la analiza și interpretarea rezultatelor. În acest sens, cercetarea statistică operează cu câteva concepte specifice: **colectivitatea** sau populația statistică, unitatea statistică, caracteristicile sau **variabilele statistice**, datele statistice și indicatorii statistici.

Colectivitatea sau populația statistică reprezintă totalitatea elementelor sau a manifestărilor de aceeași natură, asupra cărora se efectuează cercetarea. În funcție de obiectivul și particularitățile cercetării, populația statistică poate fi formată din persoane sau grupuri de persoane, din instituții, din obiecte, din tranzacții economice, din evenimente etc.

De exemplu:

- O cercetare privind veniturile populației se realizează pe o colectivitate formată din gospodării;
- O cercetare referitoare la nivelul de educație a populației – se realizează pe o colectivitate formată din persoane;
- Colectivitatea care face obiectul unei investigații privind producția și eficiența economică este alcătuită din unități economice;
- Colectivitatea cercetată în cadrul unui studiu privind activitatea sistemului de sănătate cuprinde unități sanitare (toate spitalele, policlinicile, centrele de tratament etc.);
- Studiul statistic al calității producției se realizează pe colectivități formate din obiecte (produse finite, semifabricate etc.);
- Studiul statistic al natalității sau mortalității se realizează pe colectivități compuse din evenimente (nașteri sau decese).

Colectivitatea supusă cercetării poate fi totală (cercetare exhaustivă), atunci când se fac înregistrări referitoare la toate elementele care formează obiectul studiului, sau parțială (eșantion), atunci când se fac înregistrări referitoare numai la o parte din această colectivitate.

Unitățile statistice reprezintă mulțimea numărabilă de elemente care compun colectivitatea statistică. De exemplu, unitatea statistică a unei cercetări referitoare la veniturile populației poate fi gospodăria sau persoana. Unitatea statistică a unei cercetări privind calitatea producției este produsul finit, semifabricatul sau piesa a căreia i se testează caracteristicile.

Unitățile statistice pot fi simple sau complexe. Acestea din urmă sunt formate din mai multe unități simple. O astfel de unitate este gospodăria³. Unitățile pe care

³Prin **gospodărie** se înțelege grupul de două sau mai multe persoane care locuiesc împreună în mod obișnuit, având, în general, legături de rudenie și care se gospodăresc (fac menajul) în comun, participând în totalitate sau parțial la formarea veniturilor și la cheltuirea lor. Persoana care nu aparține de o gospodărie și care declară că locuiește și se gospodărește singură se consideră gospodărie formată dintr-o singură persoană. Se consideră membri ai gospodăriei și persoanele plecate din localitate pentru o perioadă mai mare de 6 luni, care se află în țară sau străinătate, dacă acestea păstrează legături familiale cu gospodăria

se realizează cercetarea salariilor sunt, de asemenea unități complexe: unitățile economice la care se face înregistrarea datelor. Fiecare element al colectivității este purtătorul unei *caracteristici* supuse observării statistice. Caracteristica statistică reprezintă acea proprietate/însușire care este comună tuturor unităților unei colectivități statistice cercetate. Formele sau nivelurile concrete ale acestora, denumite *variante* sau *valori*, diferă de la o unitate la alta (sau în timp, în cazul aceleiași unități) sub influența unui complex de factori. Numărul unităților la care se înregistrează aceeași variantă sau valoare poartă denumirea de *frecvență* a variantei/valorii respective.

Caracteristicile statistice poartă numele și de variabile. Acestea pot fi calitative sau cantitative.

Iată câteva exemple de variabile calitative în tabelele de mai jos:

Tabel 1.1: Exemple de variabile calitative

<i>Variabilă</i>	<i>Categorie</i>
Gen	masculin, feminin
Mediu de rezidență	urban, rural
Formă juridică	societate comercială pe acțiuni, societate cu răspundere limitată, regie autonomă
Activitate economică	industrie, agricultură, comerț
Statut ocupațional	salariat, agricultor, șomer, pensionar
Nivel de educație	scăzut, mediu, superior

```
date_educatie <- read.csv("date/date_educatie.csv")
head(date_educatie)
```

```
##   id varsta      gen mediu_rezidenta nivel_educatie
## 1  1     63 masculin          urban          mediu
## 2  2     55 masculin          rural          scazut
## 3  3     31 masculin          rural          scazut
## 4  4     56 masculin          urban          mediu
```

```
## 5 5 31 feminin urban mediu
## 6 6 63 feminin rural superior
summary(date_educatie[2:4])
```

```
##      varsta      gen      mediu_rezidenta
## Min.   :16.00  feminin :39      rural:41
## 1st Qu.:33.00  masculin:31    urban:29
## Median :44.50
## Mean   :43.79
## 3rd Qu.:55.00
## Max.   :65.00
```

Variabilele cantitative sunt măsurabile și se exprimă numeric. Acestea pot fi discrete sau continue.

De exemplu, nota la statistică, este o variabilă cantitativă, discretă.

Tabel 1.2: Exemple de variabile cantitative discrete

<i>Variabilă</i>	<i>Valoare</i>
	10
	9
Nota la statistică	8
	7
	6
	5

Un exemplu de variabilă cantitativă continuă este venitul net al salariaților unei firme.

Considerăm că se cunosc veniturile tuturor celor $n=200$ salariați.

Tabel 1.3: Exemple de variabile cantitative continue

<i>Variabilă</i>	<i>Valoare</i>
	3500
	1800
Venit net	...
	6300
	5450

Statistica operează, în principal, cu frecvențele diferitelor variante ale caracteristicilor calitative, iar analiza statistică a acestor caracteristici se realizează prin

metode ce diferă de cele utilizate în studiul variabilelor cantitative.

În cazul unei cercetări referitoare la activitatea economică, de exemplu, se vor înregistra informații privind volumul producției, cheltuielile de producție și profitul, numărul de salariați și dotarea cu active corporale etc. din fiecare unitate economică.

Evaluarea situației existente pe piața forței de muncă, evoluția fenomenelor de ocupare, șomaj și inactivitate va înregistra informații referitoare la populația în vârstă de muncă, populația activă, populația ocupată, numărul șomerilor etc. Un alt exemplu este cel al unei cercetări referitoare la venituri, care va cuprinde, în principal, informații numerice despre nivelul veniturilor de care beneficiază gospodăriile (din activități salariale, din agricultură, din activități pe cont propriu, din prestații sociale etc.), despre impozitele și contribuțiile de asigurări sociale plătite de gospodării, despre mărimea gospodăriei, vârsta membrilor acesteia etc. Prin natura lor, variabilele cantitative se grupează în:

- variabile discrete - pot lua numai anumite valori (exprimate prin numere întregi);
- variabile continue - pot lua orice valori într-un interval finit sau infinit.

Numărul salariaților unei firme sau numărul de persoane care compun o gospodărie sunt variabile discrete, în timp ce valoarea producției, nivelul salariilor sau nivelul veniturilor sunt variabile continue (chiar dacă nivelurile înregistrate ale acestora apar ca valori ale unor variabile discrete). În practică, variabilele discrete sunt tratate ca fiind continue, dacă valorile lor posibile sunt numeroase (de ex., numărul salariaților unităților economice).

Datele statistice exprimă valori ale unor caracteristici cantitative ale unităților statistice, precum și ale unor grupe ale colectivității sau ale colectivității în ansamblul ei. Informațiile numerice sunt date statistice numai dacă sunt definite sub aspectul conținutului, al unității de măsură, al identificării unității/grupului/colectivității la care se referă, precum și al localizării în spațiu și în timp. Datele se obțin fie direct prin observare, fie sunt rezultate ale prelucrării datelor primare. De exemplu, date statistice sunt nivelurile cifrei de afaceri înregistrate de firmele X și Y din București, în luna septembrie 2014.

Indicatorii statistici reprezintă expresia numerică ce exprimă manifestările unor fenomene economice sau sociale, caracterizând nivelul și structura, modificările în timp și variația spațială, precum și interdependențele dintre ele.

Indicatorii statistici sunt rezultate ale prelucrării informațiilor primare înregistrate în cursul unei observări statistice sau a unor informații cuprinse în publicații statistice sau baze de date. De exemplu, modificarea cifrei de afaceri a firmei X în luna septembrie 2014 față de aceeași lună a anului precedent este un indicator statistic.

Pentru cunoașterea fenomenelor de masă, indicatorii statistici îndeplinesc mai multe funcții și anume: de măsurare, de comparare, de analiză, de sinteză; de