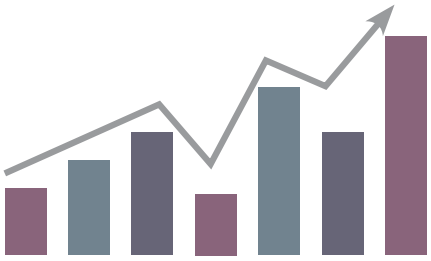
Investește în oameni!
Proiect cofinanțat din Fondul Social European prin
Programul Operațional Sectorial Dezvoltarea Resurselor Umane 2007 – 2013
Axa prioritară 1 „Educația și formarea profesională în sprijinul creșterii economice și dezvoltării societății bazate pe cunoaștere”
Domeniul major de intervenție 1.2 „Calitate în învățământul superior”
Titlul proiectului „e-Learning eficient, Individualizat și Adaptiv pentru învățământ la distanță (e-LIADA)”
Contract POSDRU/156/1.2/G/133681

Program:
Statistică descriptivă

Modul: Sistematizarea şi prezentarea datelor statistice
Tema: Populația statistică şi variabilele statistice

(60’ teorie, 30’ practică)



I. Cuprinsul secvenței – Înțelesuri majore ale comunicării

1. Ce sunt populaţia statistică şi variabilele?
2. Tipuri de variabile.

II. Obiectivele secvenței

- Însuşirea corectă a noţiunilor de populaţie statistică şi variabile;
- Identificarea corectă a tipurilor de variabile.

III. Cuvinte-cheie

Populaţie statistică, variabilă cantitativă şi calitativă, variabilă atributivă, temporală şi spaţială.

FONDUL SOCIAL EUROPEAN
Investește în
OAMENI

ROMÂNIA
UNIVERSITATEA BABEŞ-BOLYAI CLUJ-NAPOCA
CENTRUL DE FORMARE CONTINUĂ, ÎNVĂȚĂMÂNT
LA DISTANȚĂ ȘI CU FRECVENȚĂ REDUSĂ

Str Ion I.C.Brătianu, nr.20, Cluj-Napoca
Tel.: (00) 40 - 740 - 077521
Fax: (00) 40 - 364 - 815679

1. Ce este populaţia statistică ?

Citire

Populaţia statistică reprezintă mulţimea elementelor simple sau complexe, de aceeaşi natură, care au una sau mai multe însuşiri esenţiale comune, proprii elementelor cât şi populaţiei privită ca un tot unitar (Buiga, Dragoş, et al, 2015). O populaţie este finită dacă include un număr determinat de elemente, dar ea poate fi considerată drept reprezentativă pentru o populaţie teoretică infinită.

Ca urmare apare necesitatea de a delimita o populaţie în: conţinut, spaţiu şi timp. Se mai denumeşte şi populaţia univers.

Exemple de populaţii statistice: mulţimea persoanelor dintr-o anumită ţară (localitate, zonă etc.) în anul t, mulţimea gospodăriilor din România, la momentul t, mulţimea consumatorilor unui produs, mulţimea societăţilor producătoare sau concurente ale unui produs, mulţimea societăţilor distribuitoare, angajaţii unei societăţi, etc. Se notează cu majusculele de la începutul alfabetului: A, B, C etc.

Unitatea statistică constituie elementul component, al populaţiei statistice, asupra căruia se va efectua nemijlocit observarea. Unitatea statistică este purtătorul originar de informaţie sau subiectul logic al informaţiei statistice. Datorită varietăţii aspectelor sub care se poate prezenta în fapt, unitatea statistică comportă o definiţie precisă, care să excludă prin posibilitate de interpretare diferită de către observatori şi astfel orice eroare ce poate prejudicia valoarea investigaţiei.

În exemplele citate mai sus, unităţile statistice sunt: persoana, gospodăria, consumatorul, societatea producătoare sau concurentă, societatea distribuitoare, angajatul etc.

Se notează cu minusculele corespunzătoare majusculei ce simbolizează populaţia statistică, respectiv ai, bi etc..

Volumul populaţiei reprezintă numărul unităţilor statistice care alcătuiesc populaţia statistică, Acesta poate fi finit sau infinit, în funcţie de tipul populaţiei care poate fi la fel finită sau infinită.

Se notează cu N, iar pentru o populaţie A, avem:

$$A : \{a1, a2, ..., aN\}$$

Eşantion reprezintă o submulţime a unei populaţii statistice, constituită după criterii bine stabilite. În raport cu procedeul de formare a eşantionului avem eşantioane aleatoare şi eşantioane dirijate. Eşantionul aleator este format din unităţile statistice care rezultă printr-un procedeu aleator: procedeul tragerii la sorţi, tabelul cu numere întâmplătoare, procedeul extragerilor sistematice. Eşantionul dirijat este constituit pe baza unor informaţii auxiliare existente la nivelul populaţiei studiate sau lăsând liber pe anchetator să aleagă unităţile respectând doar realizarea structurii eşantionului în funcţie de criteriile stabilite. Se notează cu n. Majoritatea studiilor au ca suport datele provenite de la nivel de eşantion, de aici importanţa constituirii acestuia şi implicit, apelarea la inferenţa statistică, pentru a estima parametrii la nivelul populaţiei univers.

2. Ce sunt şi cum se clasifică variabilele statistice ?

Citire

Variabila statistică reprezintă o însuşire sau o trăsătură comună tuturor unităţilor unei populaţii. Nivelul înregistrat de o variabilă statistică la o unitate oarecare al populaţiei se numeşte realizare sau starea variabile (Buiga, Dragoş, et al, 2015). În general se notează cu majusculele de la sfârşitul alfabetului, X, Y, Z etc. Dacă se notează cu X o variabilă statistică oarecare, atunci cu x1, x2, ..., xN se vor nota stările variabilei respective.

Variabilele statistice se clasifică în raport cu natura, modul de exprimare şi modul de variaţie.

a) După natura lor variabilele statistice pot fi atributive, de timp şi de spaţiu.

- Variabila atributivă exprimă un atribut sau însuşire esenţială (alta, decât timpul sau spaţiul) unităţilor populaţiei;
- Variabila de timp ne arată timpul în care au luat fiinţă unităţile populaţiei sau perioada de timp în care au existat (exista);
- Variabila de spaţiu ne arată spaţiul în care există sau au luat naştere unităţile populaţiei.

b) După modul de exprimare a stărilor deosebim:

- Variabilă cantitativă este variabila ale cărei stări se exprimă prin valori numerice. Se mai numeşte şi variabilă metrică.
- Variabilă calitativă este variabila ale cărei stări se exprimă prin cuvinte sau coduri. Se mai numeşte variabilă nominală (stările se exprimă prin cuvinte) sau variabilă ordinală (stările se exprimă prin coduri).

c) După modul de variaţie variabila cantitativă poate fi:

- Variabilă discretă este acea variabilă care, în intervalul său de definiţie înregistrează cel mult valori raţionale, variaţia are loc în salturi.
- Variabilă continuă este acea variabilă care poate lua orice valoare reală din intervalul său de variaţie.

Exemple de variabile statistice relativ la populaţia formată din mulţimea consumatorilor unui produs:

- vârsta: variabilă atributivă, cantitativă, continuă

$$X = \{ x1 = [15\text{-}20)\ [20\text{-}30)\ \ldots \}$$

- frecvenţa de cumpărare: variabilă atributivă calitativă

$$Y = \{ y1 \text{ - foarte rar; } y2 \text{ – rar, } \ldots \}$$

- număr de sortimente cumpărate relativ la produsul analizat: variabilă atributivă, cantitativă, discretă:

$$Z = \{ z1 = 1;\ z2 = 2, \ldots \}$$

- localizarea magazinelor de unde cumpără: variabilă de spaţiu, calitativă

$$S = \{ s1 \text{ – cartierul M sau } s2 \text{ – strada P1, } \ldots \}$$

- data ultimei cumpărări a produsului analizat: variabilă de timp, cantitativă

$$T = \{ t1 = 27.01.2002;\ t2 = 24.02.2002, \ldots \}$$

Variabila aleatoare este variabila care poate lua orice valoare din valorile unei mulţimi finite sau infinite, cu o anumită probabilitate, rezultată dintr-o funcţie asociată variabilei, numită lege de probabilitate. Ca şi variabila statistică, variabila aleatoare în raport cu valorile sale poate fi discretă sau continuă. În timp ce o variabilă aleatoare înregistrează valori la întâmplare, variabila statistică constituie o însuşire certă a unităţilor statistice din populaţie. Valorile unei variabile aleatoare sunt probabile şi în strânsă legătură cu un anumit experiment. Stările unei variabile statistice nu sunt probabile, ele cuantifică o trăsătură proprie fiecărei unităţi din populaţie.

Variabila aleatoare este variabila care poate lua orice valoare din valorile unei mulţimi finite sau infinite, cu o anumită probabilitate, rezultată dintr-o funcţie asociată variabilei, numită lege de probabilitate. Ca şi variabila statistică, variabila aleatoare în raport cu valorile sale poate fi discretă sau continuă. În timp ce o variabilă aleatoare înregistrează valori la întâmplare, variabila statistică constituie o însuşire certă a unităţilor statistice din populaţie. Valorile unei variabile aleatoare sunt probabile şi în strânsă legătură cu un anumit experiment. Stările unei variabile statistice nu sunt probabile, ele cuantifică o trăsătură proprie fiecărei unităţi din populaţie.

Temă de reflecţie

- În majoritatea cazurilor, în statistică studiem populaţii statistice. Aproape tot aparatul tehnic este dezvoltat în acest scop, respectiv toată structura de organizare şi prelucrare a datelor. Uneori însă este necesar studiul statistic al unei singure unităţi, mai ales în cazurile în care ea este foarte complexă (o ţară).

- Statistica aplicată în ştiinţele sociale nu are rigiditatea unei ştiinţe perfect exacte. De exemplu vârsta unei persoane este prin natura ei o variabilă continuă, variază practic în orice moment de timp. În schimb, în orice surse de date o găsim exprimată doar în ani întregi. Ca urmare, disponibilitatea datelor ne obligă să o folosim ca o variabilă discretă şi să utilizăm metode de prelucrate adecvate.

Ştiaţi că?

Noţiunile expuse în această temă sunt relativ recente în istoria omenirii (aprox. două secole). Deşi par noţuni simple, utilizarea lor nu a fost posibilă multă vreme, datorită condiţiilor istorice care au facut mai dificilă dezvoltarea ştiinţelor sociale în raport cu ştiinţele naturii. Prin comparaţie, amintim ca logaritmii au fost descoperiţi şi utilizaţi acum două milenii.

Calculaţi

Daţi trei exemple de populaţii statistice foarte diferite între ele:

a. O populaţie formată din persoane

b. O populaţie formată din unitaţi complexe (ţări, societăţi comerciale, etc)

c. O populaţie statistică foarte diferită de cele anterioare.

Practică

Pentru fiecare din cele trei populaţii pe care le-aţi luat în considerare, daţi exemplu de cel puţin zece variabile în raport cu care populaţia ar putea fi studiată.
Clasificaţi variabilele în raport cu toate criteriile cunoscute:

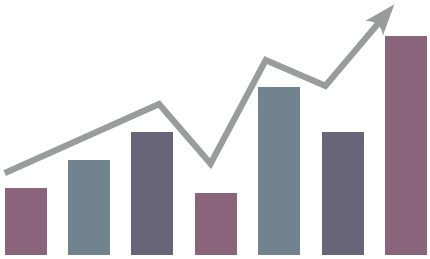
- natura lor
- modul de exprimare
- modul de variaţie

Investește în oameni!
Proiect cofinanțat din Fondul Social European prin
Programul Operațional Sectorial Dezvoltarea Resurselor Umane 2007 – 2013
Axa prioritară 1 „Educația și formarea profesională în sprijinul creșterii economice și dezvoltării societății bazate pe cunoaștere”
Domeniul major de intervenție 1.2 „Calitate în învățământul superior”
Titlul proiectului „e-Learning eficient, Individualizat și Adaptiv pentru învățământ la distanță (e-LIADA)”
Contract POSDRU/156/1.2/G/133681

Program:
Statistică descriptivă

Modul: Sistematizarea şi prezentarea datelor statistice
Tema: Seria de distribuție

(60' teorie, 30' practică)



I. Cuprinsul secvenței – Înțelesuri majore ale comunicării

1. Ce este seria de distributie ?
2. Indicatori specifici - frecvente.

II. Obiectivele secvenței

- Însuşirea corectă a noţiunilor serie statistică, serie de distribuţie, frecvenţe.
- Calculul corect al frecventelor.

III. Cuvinte-cheie

Serie statistică, serie de distribuţie, frecvenţă absolută, frecvenţă relativă.

FONDUL SOCIAL EUROPEAN
Investește în
OAMENI

ROMÂNIA
UNIVERSITATEA BABEŞ-BOLYAI CLUJ-NAPOCA
CENTRUL DE FORMARE CONTINUĂ, ÎNVĂȚĂMÂNT
LA DISTANȚĂ ȘI CU FRECVENȚĂ REDUSĂ

Str Ion I.C.Brătianu, nr.20, Cluj-Napoca
Tel.: (00) 40 - 740 - 077521
Fax: (00) 40 - 364 - 815679

Definiție

Seria statistică este o construcție care redă fie distribuția unei populații în raport cu una sau mai multe variabile, fie variația unei mărimi în timp, în spațiu sau de la o categorie la alta.

Clasificarea seriilor statistice

În raport cu numărul variabilelor

- Serii statistice unidimensionale, au la bază o singură variabilă;
- Serii statistice multidimensionale, care au la bază două sau mai multe variabile.

După natura variabilelor deosebim:

- Serii atributive, care au la bază variabile atributive;
- Serii cronologice (de timp sau istorice), care au la bază variabile de timp;
- Serii de spațiu sau teritoriale, care au la bază o variabilă de spațiu.

După modul de exprimare a stărilor variabilei deosebim:

- Serii calitative, care au la bază variabile calitative;
- Serii cantitative, care au la bază variabile cantitative și care după modul de variație a variabilei pot fi: discrete (când variabila este discretă) și continue (când variabila este continuă).

În raport cu natura indicatorului din care este alcătuită seria, avem:

- Serii de frecvență sau serii de distribuție (repartiție);
- Serii de variație.

Definiție

Seria statistică ce redă variația unei mărimi în timp, în spațiu sau de la o categorie la alta se numește serie de distribuție.

Citire

Indicatori specifici ai seriei de distributie

Conform definiţiei de mai sus, prin această serie se distribuie unităţile unei populaţii statistice în raport cu una sau mai multe variabile.

Fie o **serie statistică unidimensională** având la bază variabila X, respectiv:

$$X:\begin{pmatrix} x_1 & x_2 & \dots & x_i & \dots & x_R \\ N_1 & N_2 & \dots & N_i & \dots & N_R \end{pmatrix}$$

N_i este **frecvenţa absolută** a clasei i, $i=\overline{1,R}$ şi reprezintă numărul de unităţi ale populaţiei din clasa pentru care variabila X a înregistrat valoarea Xi

N1 + N2 + ... + NR = N.

Clasa (grupa) de unităţi în raport cu o variabilă reuneşte acele unităţi din cadrul populaţiei care înregistrează aceeaşi stare a variabilei sau stările variabilei aparţinând unui anumit interval de variaţie .

Ca urmare, în raport cu o variabilă statistică populaţia poate fi structurată într-un anumit număr d e clase. De asemenea, relativ la seria statistică unidimensională având la bază variabila X, poate fi formată cu frecvfvenţe relative, frecvenţe cumulate absolute sau relative.

Fie seria X formată cu **frecvenţe relative:**

$$X:\begin{pmatrix} x_1 & x_2 & \dots & x_i & \dots & x_R \\ f_1 & f_2 & \dots & f_i & \dots & f_R \end{pmatrix}$$

fi - ne arată ponderea unităţilor din populaţie care au înregistrat pentru variabila X starea xi:

$$f_i = \frac{N_i}{N} \qquad i=\overline{1,R}$$

Pornind de la seria iniţială se poate deduce seria formată cu **frecvenţe absolute cumulate**, respectiv:

$$X\begin{pmatrix} x_1 & x_2 & \dots & x_i & \dots & x_R \\ N(x_1) & N(x_2) & \dots & N(x_i) & \dots & N(x_R) \end{pmatrix}$$

unde: N(xi) reprezintă numărul de unităţi din populaţia studiată pentru care variabila înregistrează valori ce nu depăşesc valoarea xi.

Pornind de la seria iniţială se poate deduce seria formată cu **frecvenţe relative cumulate**, respectiv:

$$X:\begin{pmatrix} x_1 & x_2 & \dots & x_i & \dots & x_R \\ F_N(x_1) & F_N(x_2) & \dots & F_N(x_i) & \dots & F_N(x_R) \end{pmatrix}$$

unde: FN(xi) - exprimă ponderea unităţii populaţiei studiate pentru care variabila a înregistrat valori ce nu depăşesc valoarea xi.

FN(xi) = f1 + f2 + ... + fi

Seria statistică de repartiţie bidimensională este o construcţie ce redă distribuţia unei populaţii în raport cu două variabile.

Astfel, fie populaţia statistică A studiată în raport cu variabilele X şi Y, rezultatele observării se pot grupa într-un tabel de forma următoare:

Y \ X	x1	x2	...	xj	...	xj	Total
y1 y2 . . yi . . yI	N11 N21 - Ni1 - NI1	N12 N22 - Ni2 - NI2	... ... - ... - ...	N1j N2j - Nij - NIj	... ... - ... - ...	N1J N2J - NiJ - NIJ	N1. N2. - Ni. - NI.
Total	N.1	N.2	...	N.j	...	N.J	N

- Nij - reprezintă numărul de unităţi pentru care, variabila X înregistrează starea xj şi variabila Y înregistrează starea yi ;
- Ni. - numărul de unităţi pentru care Y = yi, indiferent de nivelul înregistrat de variabila X;
- N.j - numărul de unităţi pentru care X = xj, indiferent de nivelul înregistrat de variabila Y;
- N - numărul total de unităţi analizate.

Din seria bidimensională se pot extrage serii unidimensionale de forma următoare:

$$X:\begin{pmatrix} x_1 & x_2 & \dots & x_j & \dots & x_J \\ N_{.1} & N_{.2} & \dots & N_{.j} & \dots & N_{.J} \end{pmatrix}$$

$$Y:\begin{pmatrix} y_1 & y_2 & \dots & y_i & \dots & y_I \\ N_{1.} & N_{2.} & \dots & N_{i.} & \dots & N_{I.} \end{pmatrix}$$

denumite şi serii de repartiţie marginale, în raport cu X şi Y

$$Y/X = x_j:\begin{pmatrix} y_1 & y_2 & \dots & y_i & \dots & y_I \\ N_{1j} & N_{2j} & \dots & N_{ij} & \dots & N_{Ij} \end{pmatrix} \qquad \forall j = \overline{1,J}$$

denumită serie de repartiţie unidimensională în raport cu Y condiţionată de X = xj, numărul acestora fiind egal cu numărul de stări a variabilei X.

$$X/Y = y_i:\begin{pmatrix} x_1 & x_2 & \dots & x_j & \dots & x_J \\ N_{i1} & N_{i.2} & \dots & N_{ij} & \dots & N_{iJ} \end{pmatrix} \qquad \forall i = \overline{1,I}$$

denumită serie de repartiţie unidimensională în raport cu X condiţionată de Y = yi, numărul acestora fiind egal cu numărul de stări a variabilei Y.

De asemenea se poate elabora sau deduce seria de repartiţie bidimensională formată cu frecvenţe relative, unde:

$$f_{ij} = \frac{N_{ij}}{N} \qquad f_{i.} = \frac{N_{i.}}{N} \qquad f_{.j} = \frac{N_{.j}}{N} \qquad i = \overline{1,I} \qquad j = \overline{1,J}$$

Practică

Pentru fiecare din cele trei populaţii pe care le-aţi luat în considerare în aplicaţia de la secvenţa anterioară, alegeţi câte două variabile în raport cu care construiţi:

- seriile unidimensionale formate cu frecvenţele absolute
- seriile bidimensionale formate cu frecvenţe absolute
- seriile condiţionate formate cu frecvenţele absolute

Calculați

Calculul frecvențelor derivate

Pentru seriile construite în aplicaţia de mai sus, calculaţi seriile derivate cu indicatorii:

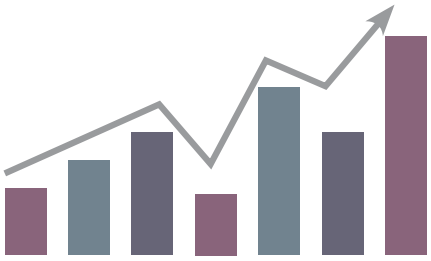
- frecvenţa absolută cumulată
- frecveţa relativă
- frecvenţa relativă cumulată

Investește în oameni!
Proiect cofinanțat din Fondul Social European prin
Programul Operațional Sectorial Dezvoltarea Resurselor Umane 2007 – 2013
Axa prioritară 1 „Educația și formarea profesională în sprijinul creșterii economice și dezvoltării societății bazate pe cunoaștere”
Domeniul major de intervenție 1.2 „Calitate în învățământul superior”
Titlul proiectului „e-Learning eficient, Individualizat și Adaptiv pentru învățământ la distanță (e-LIADA)”
Contract POSDRU/156/1.2/G/133681

Program:
Statistică descriptivă

Modul: Sistematizarea şi prezentarea datelor statistice
Tema: Seria de variație

(60’ teorie, 30’ practică)



I. Cuprinsul secvenței – Înțelesuri majore ale comunicării

1. Ce este seria de variaţie?
2. Indicatori specifici – diferenţe, indici, ritmuri.

II. Obiectivele secvenței

- Însuşirea corectă a noţiunii de serie de variaţie.
- Calculul corect al indicatorilor derivaţi.

III. Cuvinte-cheie

Serie de variaţie, indicator de nivel, indicator de intensitate, diferenţa absolută, indice, diferenţă relativă, ritm.

ROMÂNIA
UNIVERSITATEA BABEŞ-BOLYAI CLUJ-NAPOCA
CENTRUL DE FORMARE CONTINUĂ, ÎNVĂȚĂMÂNT
LA DISTANȚĂ ȘI CU FRECVENȚĂ REDUSĂ

Str Ion I.C.Brătianu, nr.20, Cluj-Napoca
Tel.: (00) 40 - 740 - 077521
Fax: (00) 40 - 364 - 815679

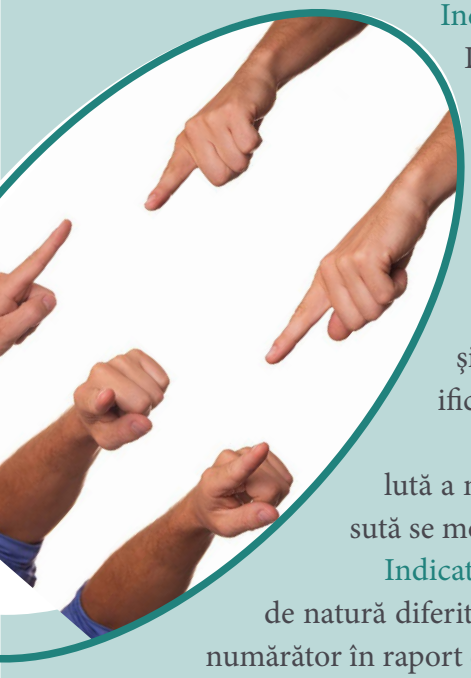
Definiție

Conform definiţiei seria de variaţie redă variaţia unei mărimi, în timp, în spaţiu sau de la o categorie la alta. Ca urmare, în continuare vom vorbi de serii cronologice (au la bază o variabilă de timp), serii de spaţiu (au la bază o variabilă de spaţiu) şi serii categoriale (au la bază variabile atributive). Cele mai des întâlnite sunt seriile cronologice şi seriile de spaţiu.

Indicatori ai seriei de variaţie

Citire

Seriile de variaţie au la bază mărimi absolute şi relative. După unii autori din cadrul mărimilor absolute fac parte indicatorul de nivel şi diferenţa absolută a unei mărimi, iar din cadrul mărimilor relative fac parte: indicatorul relativ de intensitate, indicele statistic şi diferenţa relativă a unei mărimi.



Indicatorul de nivel (Y) este o mărime ce reflectă nivelul unui fenomen analizat. De exemplu: producţia diferitelor produse, veniturile populaţiei, suprafaţa cultivată cu principalele culturi, transportul, exportul, importul etc.

Diferenţa absolută a unei mărimi (Δ_Y) exprimă diferenţa dintre nivelul cercetat şi nivelul bază de comparaţie al mărimii analizate. Se exprimă în aceeaşi unitate de măsură în care este cuantificat fenomenul analizat şi ne arată cu cât s-a modificat acesta de la un nivel la altul.

Indicele statistic al unei mărimi (I_Y) exprimă raportul dintre nivelul cercetat şi nivelul bază de comparaţie al mărimii analizate. Ne arată de câte ori se modifică acea mărime, de la un nivel la altul.

Diferenţa relativă a unei mărimi (R_Y) exprimă raportul dintre diferenţa absolută a mărimii respective şi nivelul bază de comparaţie al acesteia. Ne arată cu cât la sută se modifică mărimea de la un nivel la altul.

Indicatorul relativ de intensitate (d) se defineşte ca raport între doi indicatori de nivel de natură diferită şi arată gradul de răspândire a fenomenului cuantificat de indicatorul de la numărător în raport cu fenomenul cuantificat de indicatorul de la numitor. De exemplu: producţia diferitelor culturi / ha, densitatea populaţiei, producţia principalelor produse/ locuitor, rata şomajului etc.

Greutatea specifică (g) reflectă structura fenomenului analizat în raport cu stările variabile X, de la baza seriei.

Seriile cronologice

Seria cronologică reflectă evoluţia în timp a unei mărimi.

Valorile variabilei ca funcţie de timp pot fi fixate la un anumit moment de timp sau să se refere la un interval de timp.

Seria cronologică de momente este o serie de observaţii ordonate în timp, exprimând stocuri De exemplu, volumul populaţiei, număr de universităţi, bănci, instituţii, fonduri fixe, numărul salariaţilor, întreprinderile mici şi mijlocii din diferite domenii de activitate, unităţile de cazare turistică etc. Într-o astfel de serie însumarea mărimii analizate nu are sens din punct de vedere al conţinutului, aceasta fiind permisă din considerente de calcul, ajustări etc. Seria cronologică de intervale este o serie de observaţii ordonate în timp exprimând fluxuri. De exemplu: născuţii vii, divorţurile, decesele, producţia diferitelor culturi sau produse, venituri, cheltuieli, producţia industrială, agricolă, exportul, importul etc. Într-o astfel de serie are sens însumarea mărimii analizate.

Fie o serie cronologică de momente sau de intervale ce reflectă evoluţia în timp a nivelului unei mărimi Y,

$$Y:\begin{pmatrix} 0 & 1 & 2 & \dots & t & \dots & T \\ y_0 & y_1 & y_2 & \dots & y_t & \dots & y_T \end{pmatrix}$$

Pornind de la această serie se pot deduce seriile formate cu diferenţe absolute, indici şi diferenţe relative. În funcţie de modul de raportare a stărilor variabilei timp t, mărimile de mai sus se pot calcula cu bază fixă (t / t0) (baza de comparaţie rămâne aceeaşi) sau cu bază în lanţ (t / t-1) (baza de comparaţie se schimbă, fiind considerată cea precedentă nivelului comparat).

Fie seriile cronologice formate cu:

- diferenţe absolute cu bază fixă:

$$\Delta_y^{t/t_0}:\begin{pmatrix} 0 & 1 & 2 & \dots & t & \dots & T \\ 0 & \Delta_y^{1/0} & \Delta_y^{2/0} & \dots & \Delta_y^{t/0} & \dots & \Delta_y^{T/0} \end{pmatrix}$$

$$\Delta_y^{t/0} = y(t) - y(0)$$

- diferenţe absolute cu bază în lanţ

$$\Delta_y^{t/t-1}:\begin{pmatrix} 0 & 1 & 2 & \dots & t & \dots & T \\ - & \Delta_y^{1/0} & \Delta_y^{2/1} & \dots & \Delta_y^{t/t-1} & \dots & \Delta_y^{T/T-1} \end{pmatrix}$$

$$\Delta_y^{t/t-1} = y(t) - y(t-1)$$

Între cele două tipuri de diferenţe absolute cu baza fixă şi cu bază în lanţ, există relaţii de legătură ce ne permit exprimarea unora în funcţie de celelalte. În acest context, însumând diferenţele absolute cu baza în lanţ se obţin diferenţele absolute cu baza fixă.

$$\Delta_y^{t/0} = \Delta_y^{1/0} + \Delta_y^{2/1} + \Delta_y^{3/2} + \dots + \Delta_y^{t/t-1}$$

Scăzând diferenţele succesive cu bază fixă se obţin diferenţele cu bază în lanţ.

$$\Delta_y^{t/0} - \Delta_y^{t-1/0} = y(t) - y(0) - y(t-1) + y(0) = y(t) - y(t-1) = \Delta_y^{t/t-1}$$

Diferenţa absolută ne arată cu cât se modifică mărimea analizată de la un moment la altul. Se exprimă în aceeaşi unitate de măsură în care este cuantificat fenomenul studiat. Dacă fenomenul analizat se exprimă valoric, atunci diferenţa absolută nu reflectă prea bine modificările ce intervin, impunându-se utilizarea mărimilor relative respective, indicele statistic şi diferenţa relativă.

Fie seriile cronologice formate cu:

- indici statistici cu bază fixă

$$I_y^{t/t_0}:\begin{pmatrix} 0 & 1 & 2 & \dots & t & \dots & T \\ 1 & I_y^{1/0} & I_y^{2/0} & \dots & I_y^{t/0} & \dots & I_y^{T/0} \end{pmatrix}$$

$$\Delta_y^{t/0} = y(t) - y(0)$$

- indici statistici cu bază în lanţ

$$\Delta_y^{t/t-1}:\begin{pmatrix} 0 & 1 & 2 & \dots & t & \dots & T \\ - & \Delta_y^{1/0} & \Delta_y^{2/1} & \dots & \Delta_y^{t/t-1} & \dots & \Delta_y^{T/T-1} \end{pmatrix}$$

$$\Delta_y^{t/t-1} = y(t) - y(t-1)$$

Indicatori ai seriei de variaţie

Citire

Între cele două tipuri de indici există următoarele relaţii de legătură:

- Făcând produsul indicilor cu bază în lanţ până la o anumită stare a variabilei t, se obţine indicele cu bază fixă al clasei respective.

$$I_y^{1/0}.I_y^{2/1}.\ \dots\ .I_y^{t/t-1} = \frac{y(1)}{y(0)}.\frac{y(2)}{y(1)}.\dots.\frac{y(t)}{y(t-1)} = \frac{y(t)}{y(0)} = I_y^{t/0}$$

- Împărţind doi indici succesivi cu bază fixă se obţine un indice cu bază în lanţ:

$$I_y^{t/0} : I_y^{t-1/0} = \frac{y(t)}{y(0)} : \frac{y(t-1)}{y(0)} = \frac{y(t)}{y(t-1)} = I_y^{t/t-1}$$

Indicele statistic ne arată de câte ori se modifică fenomenul analizat. Este mărimea cel mai des folosită în caracterizarea evoluţiei fenomenelor din economie.

Având ca bază de referinţă o serie cronologică se pot elabora serii formate cu:

- diferenţe relative cu bază fixă

$$R_y^{t/0} = \frac{\Delta_y^{t/0}}{y(0)} = \frac{y(t)-y(0)}{y(0)} = \frac{y(t)}{y(0)} - 1 = I_y^{t/0} - 1$$

$$R_y^{t/t_0} : \begin{pmatrix} 0 & 1 & 2 & \dots & t & \dots & T \\ 0 & R_y^{1/0} & R_y^{2/0} & \dots & R_y^{t/0} & \dots & R_y^{T/0} \end{pmatrix}$$

- diferenţe relative cu bază în lanţ

$$R_y^{t/t-1} : \begin{pmatrix} 0 & 1 & 2 & \dots & t & \dots & T \\ - & R_y^{1/0} & R_y^{2/1} & \dots & R_y^{t/t-1} & \dots & R_y^{T/T-1} \end{pmatrix}$$

$$R_y^{t/t-1} = \frac{\Delta_y^{t/t-1}}{y(t-1)} = I_y^{t/t-1} - 1 \qquad \text{sau} \qquad I_y^{t/t-1}.100 - 100$$

Această mărime la fel ca şi indicele statistice, se foloseşte frecvent în caracterizarea fenomenelor din economie.

Dacă seria cronologică analizată este de intervale, se poate deduce seria formată cu greutatea specifică:

$$g_y : \begin{pmatrix} 0 & 1 & 2 & \dots & t & \dots & T \\ g_0 & g_1 & g_2 & \dots & g_t & \dots & g_T \end{pmatrix}$$

$$g(t) = \frac{y(t)}{\sum_{t=1}^{T} y(t)}$$

Seria statistică de spaţiu (teritorială)

Seria statistică de spaţiu este o construcţie statistică ce reflectă variaţia în spaţiu a unei mărimi. Seria de spaţiu prezintă o importanţă din ce în ce mai mare, datorită dezvoltării sistemului informaţional, a necesităţii comparaţiilor internaţionale şi a comparaţiilor între regiunile unei ţări.

În cadrul Anuarului Statistic al României există capitole distincte de „Statistică teritorială” şi „Statis-

tică internaţională”. În capitolul de „Statistică teritorială” sunt cuprinse informaţii privind: populaţia, forţa de muncă, condiţii de muncă, veniturile populaţiei, cheltuielile şi consumul populaţiei, locuinţe, asistenţă socială, sănătate, învăţământ, cultură, sport, conturi naţionale, rezultate şi performanţe ale întreprinderilor, agricultură, silvicultură, industrie, transporturi, poştă, telecomunicaţii, turism, finanţe, justiţie şi starea infracţională, pe cele 7 regiuni şi Bucureşti.

La baza seriei de spaţiu se găsesc atât mărimi absolute (indicator de nivel, diferenţa absolută), cât şi mărimi relative (indicator relativ de intensitate, indicele statistic, diferenţa relativă).

Fie seria statistică Z, de forma următoare:

$$Z:\begin{pmatrix} s_0 & s_1 & s_2 & \dots & s_i & \dots & s_R \\ Z(1) & Z(2) & Z(3) & \dots & Z(i) & \dots & Z(R) \end{pmatrix}$$

si – este o stare a variabilei ce exprimă spaţiul, ;

Z(i) – exprimă o mărime (indicator de nivel sau relativ de intensitate).

Plecând de la seria de forma (1.15) se pot deduce seriile formate cu:

- diferenţe absolute cu bază fixă:

$$\Delta_Z^{s/s_0}:\begin{pmatrix} s_0 & s_1 & s_2 & \dots & s_i & \dots & s_R \\ 0 & \Delta_Z^{s_1/s_0} & \Delta_Z^{s_2/s_0} & \dots & \Delta_Z^{s_i/s_0} & \dots & \Delta_Z^{s_R/s_0} \end{pmatrix}$$

$$I_Z^{s_i/s_0} = \frac{Z(i)}{Z(0)}(100)$$

- indicii statistici cu bază fixă

$$I_Z^{s/s_0}:\begin{pmatrix} s_0 & s_1 & s_2 & \dots & s_i & \dots & s_R \\ 0 & I_Z^{s_1/s_0} & I_Z^{s_2/s_0} & \dots & I_Z^{s_i/s_0} & \dots & I_Z^{s_R/s_0} \end{pmatrix}$$

$$I_Z^{s_i/s_0} = \frac{Z(i)}{Z(0)}(100)$$

- diferenţe relative cu bază fixă

$$R_Z^{s/s_0}:\begin{pmatrix} s_0 & s_1 & s_2 & \dots & s_i & \dots & s_R \\ 0 & R_Z^{s_1/s_0} & R_Z^{s_2/s_0} & \dots & R_Z^{s_i/s_0} & \dots & R_Z^{s_R/s_0} \end{pmatrix}$$

$$I_Z^{s_i/s_0} = \frac{\Delta_Z^{s_i/s_0}}{Z(0)} = I_Z^{s_i/s_0} - 100$$

Calculaţi

- Din Anuarul Statistic sau alte surse informaţionale extrageţi două serii cronologice având la bază indicatorul de nivel, una de momente, alta de intervale şi deduceţi seriile formate cu diferenţe absolute, indici statistici, diferenţe relative, cu bază fixă şi cu bază în lanţ (interpretări).
- Din Anuarul Statistic sau alte surse informaţionale extrageţi o serie de spaţiu formată cu indicator de nivel sau indicator relativ de intensitate şi deduceţi seriile formate cu diferenţe absolute, indici şi diferenţe relative, calculate cu bază fixă. Interpretare.
- Extrageţi 2 exemple de serii de spaţiu ce conţin informaţii importante pentru domeniul economic.

Test

- Produsul intern brut (PIB) al unei ţări a crescut în 2006 faţă de 2005 cu 12% şi a scăzut în 2007 faţă de 2006 cu 12%. Cu cât s-a modificat procentual PIB în 2007 faţă de 2005?
- Produsul intern brut pe locuitor (PIB/loc) este de 32640$ în Franţa şi de 26380$ în Spania.

Determinaţi:

a) cu cât la sută este mai mare PIB/loc în Franţa faţă de Spania?

b) cu cât la sută este mai mic PIB/loc în Spania decât în Franţa?

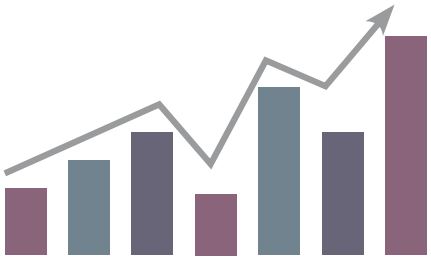
3. Salariul lui Vasile este de 1530 lei, cu 38,4% mai mic decât salariul lui Ilie. Care este salariul lui Ilie?

Investește în oameni!
Proiect cofinanțat din Fondul Social European prin
Programul Operațional Sectorial Dezvoltarea Resurselor Umane 2007 – 2013
Axa prioritară 1 „Educația și formarea profesională în sprijinul creșterii economice și dezvoltării societății bazate pe cunoaștere”
Domeniul major de intervenție 1.2 „Calitate în învățământul superior”
Titlul proiectului „e-Learning eficient, Individualizat și Adaptiv pentru învățământ la distanță (e-LIADA)”
Contract POSDRU/156/1.2/G/133681

Program:
Statistică descriptivă

Modul: Sistematizarea şi prezentarea datelor statistice
Tema: Reprezentări grafice

(60’ teorie, 30’ practică)



I. Cuprinsul secvenței – Înțelesuri majore ale comunicării

1. Ce sunt graficele statistice
2. Tipuri de grafice statistice

II. Obiectivele secvenței

- Alegerea tipului adecvat de reprezentare grafică în concordanţă cu datele dorite a fi reprezentate.
- Alegerea celei mai corecte şi sugestive figuri pentru fenomenul reprezentat.

III. Cuvinte-cheie

Histogramă, graphic de structură, nor de puncte, cronogramă, cartogramă, cartodiagramă.

FONDUL SOCIAL EUROPEAN
Investește în
OAMENI

ROMÂNIA
UNIVERSITATEA BABEŞ-BOLYAI CLUJ-NAPOCA
CENTRUL DE FORMARE CONTINUĂ, ÎNVĂȚĂMÂNT
LA DISTANȚĂ ȘI CU FRECVENȚĂ REDUSĂ

Str Ion I.C.Brătianu, nr.20, Cluj-Napoca
Tel.: (00) 40 - 740 - 077521
Fax: (00) 40 - 364 - 815679

Definiție

Reprezentarea grafică a unei serii ne dă o imagine geometrică (în plan sau spaţiu) cu privire la forma statică sau evoluţia dinamică a fenomenului cuantificat de seria respectivă. Graficul asociat unei serii constituie o imagine spaţială a fenomenului de cercetat, permiţând evidenţierea rapidă a structurii, dinamicii şi tendinţei de dezvoltare a acestuia. Reprezentările grafice sunt folosite atât în scopul cunoaşterii populaţiei în cauză, cât şi pentru popularizarea unor rezultate din diverse domenii de activitate.

Elementele unui grafic

Citire

Titlul graficului trebuie să fie scurt, clar şi semnificativ pentru conţinutul fenomenului reliefat prin seria considerată.

Scara de reprezentare reuneşte mulţimea tuturor punctelor cotate. În cazul în care variabila înregistrează valori mici, gradarea scării începe în principiu de la zero, dacă variabila înregistrează valori mari se consideră o altă origine stabilită cu aproximaţie.

Pentru a nu încărca prea mult desenul, se recomandă reprezentarea pe scară doar a valorilor dispuse la un anumit interval convenabil ales. Distanţele dintre două puncte cotate consecutive se numeşte intervalul graficului. Când intervalele sunt egale atunci avem scări uniforme, în caz contrar avem scări neuniforme. Scara aritmetică este o scară uniformă în timp ce scara logaritmică sau cea binomială constituie exemple de scări neuniforme.

Reţeaua graficului permite identificarea cu uşurinţă în plan sau în spaţiu a punctelor corespunzătoare valorilor înregistrate de variabilele în cauză. Sistemul axelor rectangulare (în plan sau spaţiu) constituie cele mai uzuale reţele în reprezentarea grafică a seriilor statistice.

Semnele convenţionale se pot materializa într-o reprezentare grafică prin inscripţii, fie printr-o legendă.

Inscripţia trebuie să fie scurtă şi semnificativă şi plasată cât mai bine în raport cu elementul din grafic pe care îl explicitează.

Legenda se foloseşte pentru a explicita folosirea semnelor, culorilor sau diverselor haşuri folosite în graficul în cauză.

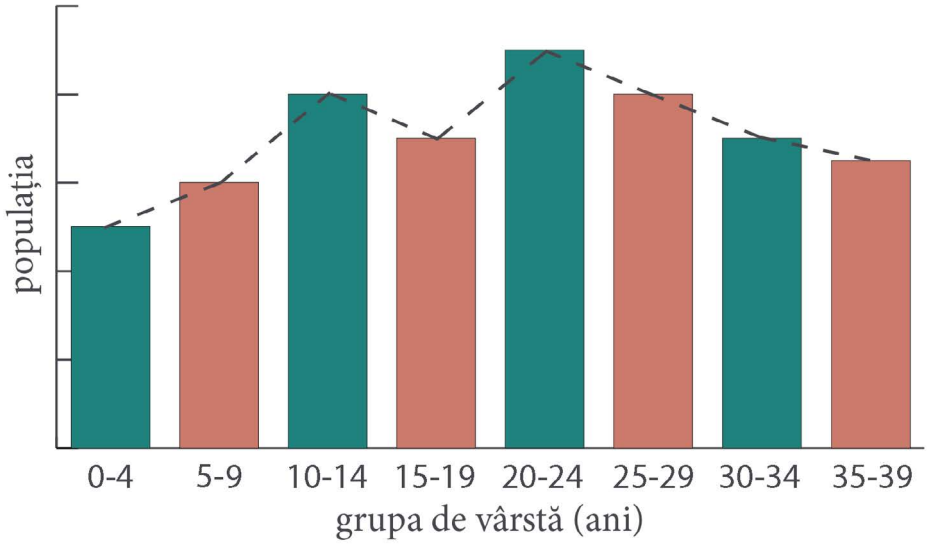
Tipuri de grafice

Citire

Histograma. Graficul specific seriilor care au la bază o variabilă continuă (de intervale) este histograma. Aceasta se construieşte într-un sistem de axe rectangulare după cum urmează: pe abscisă se trec intervalele de variaţie, iar pe ordonată se trasează scara frecvenţelor. Latura necunoscută a dreptunghiului, notată cu Li se determină din următoarea relaţie:

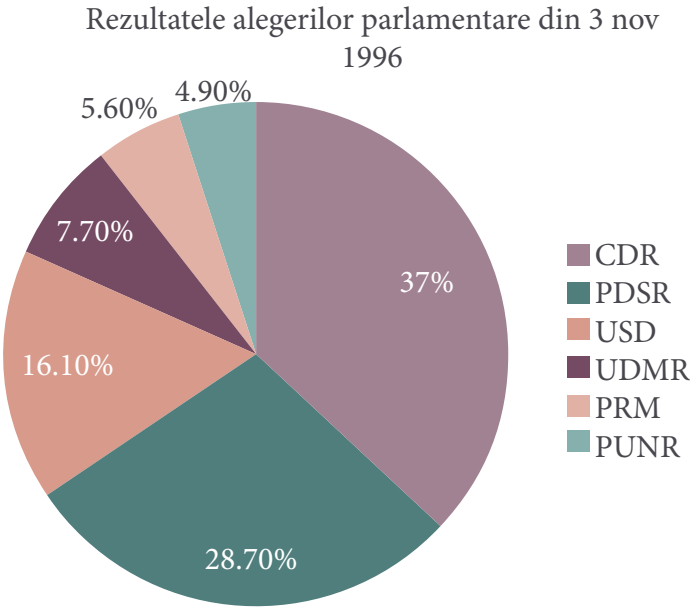
$$Li \cdot li = k \cdot Ni$$

- li = latura cunoscută a dreptunghiului corespunzător intervalului (xi-1 - xi);
- Li = latura necunoscută a dreptunghiului corespunzător intervalului (xi-1 - xi);
- Ni = frecvenţa absolută a clasei „i”;
- k = un coeficient de proporţionalitate care se alege în raport cu scara de reprezentare.



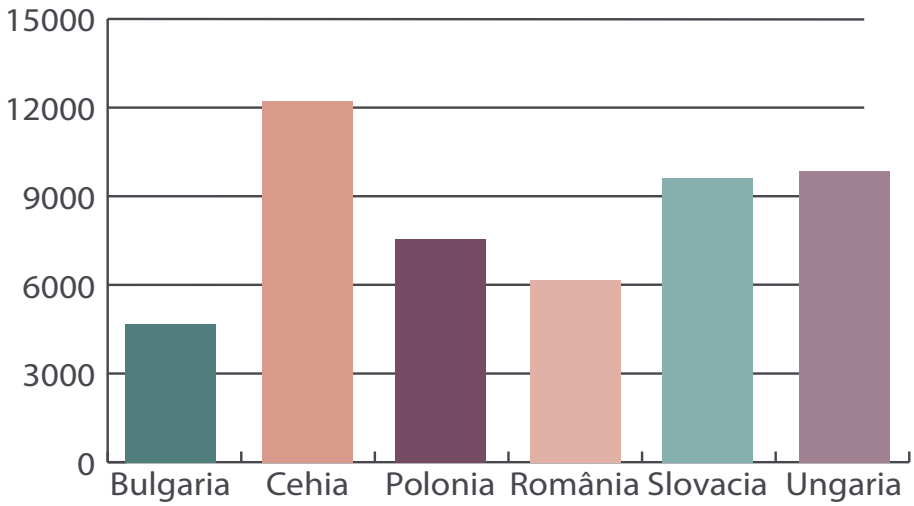
Diagramele de structură

Punerea în evidenţă sub formă grafică a structurii unei populaţii statistice este posibilă apelând la diagramele de structură: dreptunghiul, pătratul, cercul şi semicercul de structură. Aceste tipuri de grafice permit reprezentarea grafică a seriilor unidimensionale construite cu mărimi de structură (frecvenţe relative, greutate specifică). Cel mai des folosit este cercul de structură denumit şi diagrama sectorială (piechart).



Diagramele prin benzi (barchart)

Acest tip de grafic utilizează benzile (barele), pentru a reprezenta distribuţia unei populaţii în raport cu o variabilă cantitativă discretă sau calitativă. Benzile au aceeaşi lăţime (bază), iar lungimea (înălţimea) lor este direct proporţională cu frecvenţa clasei reprezentate. Numărul benzilor este egal cu numărul claselor în care este împărţită populaţia studiată. De asemenea se pot lua în considerare o variabilă sau două. În reprezentări se utilizează benzi simple sau benzi grupate. Poziţia benzilor poate fi orizontală sau verticală.

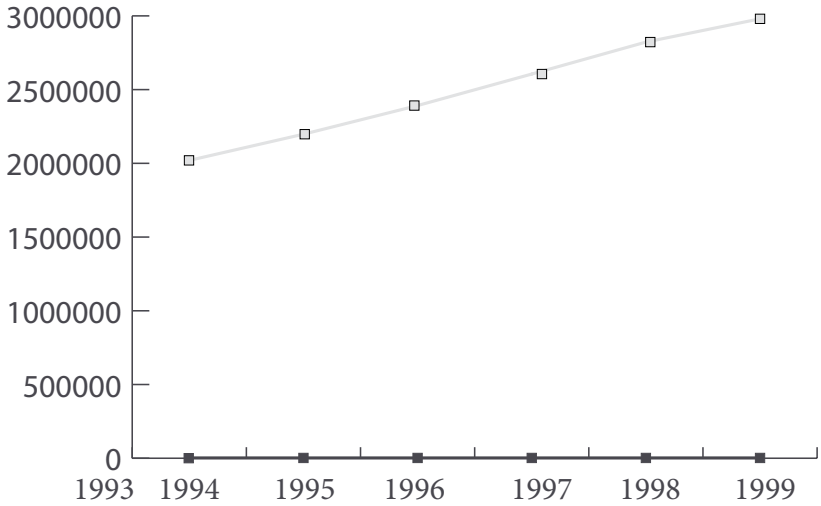


Cronograma

O categorie foarte importantă de serii o constituie seriile cronologice, a căror reprezentare grafică se realizează prin cronograme. Trasarea unei cronograme se realizează într-un sistem de axe rectangulare. Se consideră seria cronologică de forma:

$$Y:\begin{pmatrix} 0 & 1 & 2 & \dots & t & \dots & T \\ y_0 & y_1 & y_2 & \dots & y_t & \dots & y_T \end{pmatrix}$$

unde: $t=\overline{0,T}$, reprezintă momentele (sau perioadele) de timp care se reprezintă pe axa absciselor, iar mărimile yt se reprezintă pe axa ordonatelor. Fiecărei perechi de valori (t, yt), îi corespunde un punct în planul axelor rectangulare. Unind prin segmente de dreaptă punctele consecutive, astfel determinate, se obţine ceea ce se numeşte cronogramă.



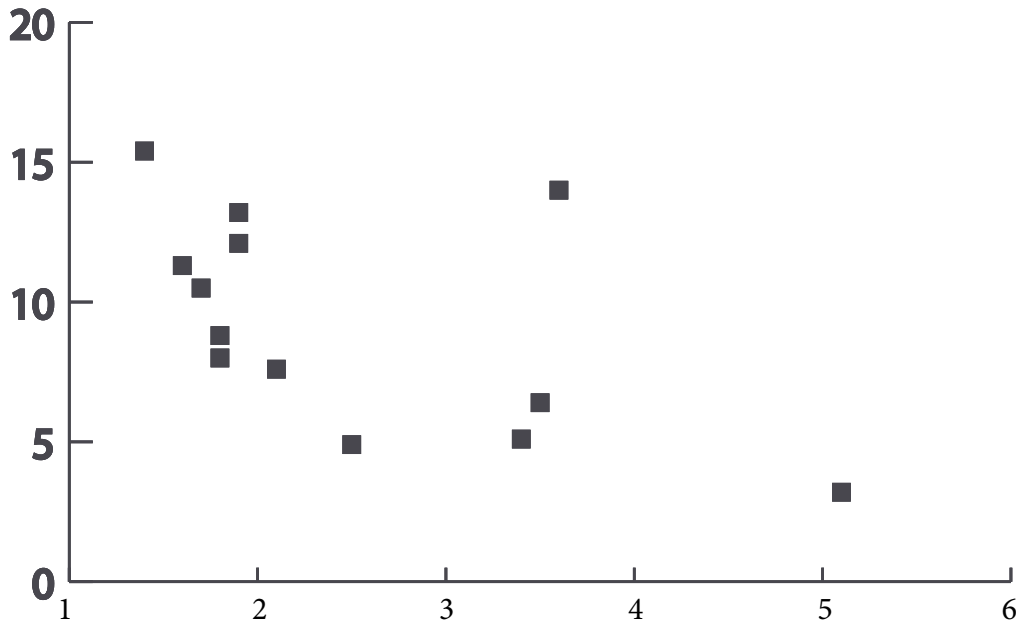
Cartograma şi cartodiagrama

Aceste tipuri de grafice se folosesc frecvent pentru reprezentarea grafică a seriilor statistice de spaţiu. Realizarea unei cartograme sau a unei cartodiagrame presupune conturarea spaţiului (sub formă de hartă) în interiorul căruia se manifestă fenomenul care este cuantificat de seria de reprezentat. În interiorul hărţii astfel realizată, prin diverse culori sau nuanţe ale aceleiaşi culori, prin haşuri sau prin diferite diagrame, este evidenţiată intensitatea dezvoltării fenomenului cercetat precum şi mărimea indicatorilor seriei. Cartodiagrama

constituie o modalitate de reprezentare grafică a seriilor de spaţiu, realizându-se ca o îmbinare între cartogramă şi diferite alte tipuri de diagrame, ca de exemplu diagrame prin benzi, cerc, pătrat, dreptunghi etc.

Norul statistic

Norul statistic constituie o modalitate de reprezentare grafică a seriilor atributive de repartiţie bidimensionale. Se consideră o serie bidimensională de repartiţie în raport cu variabilele discrete X şi Y. În sistemul de axe rectangulare xOy se marchează toate punctele de coordonate (xj, yi); $i = \overline{1, I}$; $j = \overline{1, J}$



Practică

Pentru fiecare tabel de date de mai jos, gasiţi un echivalent (alt indicator, altă populaţie studiată) în surse de date electronice.

1. Populaţia României sub 40 de ani pe grupe de vârstă

Grupa de vârstă (ani)	Populaţia
0 – 4	1147065
5 – 9	1330733
10 – 14	1737153
15 – 19	1701881
20 – 24	1978835
25 – 29	1792822
30 – 34	1698268
35 – 39	1335039

2. Distribuţia voturilor electoratului pentru Senat la alegerile din 3 noiembrie 1996:

Formaţiunea Politică	CDR	PDSR	USD	UDMR	PRM	PUNR
Voturi Obţinute (%)	37.0	28.7	16.1	7.7	5.6	4.9

3. PNB/loc în $ în România şi alte ţări est-europene, în 1998

Ţara	Bulgaria	Cehia	Polonia	România	Slovacia	Ungaria
PNB/loc ($)	4683	12197	7543	6153	9624	9832

4. Numărul total de autoturisme înscrise în circulaţie la sfârşitul anului în România în perioada 1994-1999.

Anul	1994	1995	1996	1997	1998	1999
Autoturisme înmatriculate	2020017	2197477	2391869	2605465	2822254	2980014

5. Un produs a fost lansat simultan pe 13 pieţe. Pe aceste pieţe, produsul a fost propus la preţuri diferite (P), veniturile consumatorilor (V) fiind şi ele diferite. Pentru fiecare piată s-a înregistrat un anumit nivel al cererii ©, rezultatele fiind sintetizate în tabelul următor:

Nr. Crt.	1	2	3	4	5	6	7	8	9	10	11	12	13
Cerere ©	15,4	3,2	4,9	10,5	8,0	5,1	7,6	11,3	14,0	6,4	13,2	8,8	12,1
Preţ (P)	1,4	5,1	2,5	1,7	1,8	3,4	2,1	1,6	3,6	3,5	1,9	1,8	1,9

Calculaţi

Pentru seriile extrase după modelul de mai sus, construiţi următoarele grafice:

- Histograma
- Cercul de structură
- Diagrama prin benzi
- Cronograma
- Norul de puncte

Test

Pentru fiecare din figurile de mai jos enumeraţi tipul de serie şi variabile corespunzător:

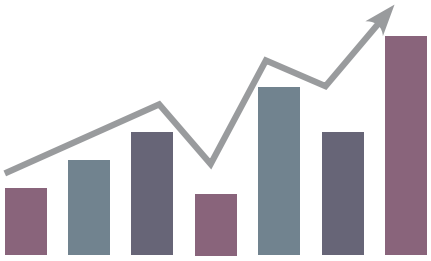
- Histograma
- Diagramele de structură
- Diagramele areale
- Norul statistic
- Cronograma
- Cartograma
- Cartodiagrama

Investește în oameni!
Proiect cofinanțat din Fondul Social European prin
Programul Operațional Sectorial Dezvoltarea Resurselor Umane 2007 – 2013
Axa prioritară 1 „Educația și formarea profesională în sprijinul creșterii economice și dezvoltării societății bazate pe cunoaștere”
Domeniul major de intervenție 1.2 „Calitate în învățământul superior”
Titlul proiectului „e-Learning eficient, Individualizat și Adaptiv pentru învățământ la distanță (e-LIADA)”
Contract POSDRU/156/1.2/G/133681

Program:
Statistică descriptivă

Modul: Sistematizarea şi prezentarea datelor statistice
Tema: Media aritmetică

(60’ teorie, 30’ practică)



I. Cuprinsul secvenței – Înțelesuri majore ale comunicării

1.Noţiunea de valoare medie
2.Media aritmetică – definire, mod de calcul, proprietăţi

II. Obiectivele secvenței

- Însuşirea corectă a noţiunilor valoare medie şi medie aritmetică.
- Calculul şi interpretarea corectă a mediei aritmetice.

III. Cuvinte-cheie

Valoare medie, medie aritmetică, proprietăţi ale mediei, proprietatea de adiţiune a mediei aritmetice.

FONDUL SOCIAL EUROPEAN
Investește în
OAMENI

ROMÂNIA
UNIVERSITATEA BABEŞ-BOLYAI CLUJ-NAPOCA
CENTRUL DE FORMARE CONTINUĂ, ÎNVĂȚĂMÂNT
LA DISTANȚĂ ȘI CU FRECVENȚĂ REDUSĂ

Str Ion I.C.Brătianu, nr.20, Cluj-Napoca
Tel.: (00) 40 - 740 - 077521
Fax: (00) 40 - 364 - 815679

Valoarea medie

Definiție

În vederea definirii parametrului valoarea medie se consideră o populaţie statistică studiată în raport cu variabila cantitativă X şi o funcţie G(x1,x2,…,xR) unde xi, $i=\overline{1,R}$, reprezintă stările variabilei X. Funcţia G exprimă o anumită însuşire esenţială, un atribut al populaţiei în raport cu variabila X. Această funcţie se numeşte funcţie determinantă. Prin definiţie, valoarea medie $\overline{X}$ a variabilei X este parametrul care lasă invariantă funcţia determinantă, adică:

$$G(x_1, x_2, ..., x_R) = G(\overline{X}, \overline{X}, ..., \overline{X})$$

Această egalitate se întâlneşte sub denumirea de relaţia lui BOIARSKI-KISINI. În funcţie de forma analitică a funcţiei G, din relaţie se deduce expresia analitică (indicatorul) de calcul a valorii medii $\overline{X}$.

Generalităţi privind valoarea medie

Citire

Pentru a exemplifica definiţia dată, se consideră populaţia familiilor dintr-o localitate, cercetată în raport cu numărul de copii. Datele rezultate din observare se prezintă ca o serie de repartiţie de forma:

$$X : \begin{pmatrix} x_i \\ N_i \end{pmatrix}_{i=\overline{1,R}}$$

În acest caz, funcţia determinantă are următoarea formă:

$$G(x_1, x_2, ..., x_R) = \sum_{i=1}^{R} x_i N_i$$

semnificând numărul total de copii din localitatea respectivă. Pentru a găsi numărul mediu de copii pe familie se particularizează relaţia din funcţia determinantă după cum urmează

$$\sum_{i=1}^{R} x_i \cdot N_i = \sum_{i=1}^{R} \overline{X} \cdot N_i$$

de unde rezultă:

$$\overline{X} = \frac{\sum_{i=1}^{R} x_i \cdot N_i}{\sum_{i=1}^{R} N_i}$$

Cunoaşterea “naturii” parametrului valoare medie, conduce la o definiţie mai completă şi plină de semnificaţie. Pentru a înţelege semnificaţia valorii medii $\overline{X}$, trebuie subliniat faptul că, în general, variaţia unui fenomen, de orice natură, şi în particular variaţia unei variabile X în raport cu care este cercetată o populaţie, este determinată de acţiunea simultană a două categorii de factori: factori esenţiali şi factori neesenţiali. În categoria factorilor esenţiali intră acei factori care acţionează asupra tuturor unităţilor populaţiei în mod continuu şi în acelaşi sens, determinând, în principal, nivelul de dezvoltare a variabilei pentru fiecare unitate componentă din populaţie. Factorii esenţiali se conjugă în acţiunea lor cu factorii neesenţiali, care, în general, au un caracter aleator, sunt numeroşi şi neuniform răspândiţi printre unităţile populaţiei. Fiecare din factorii consideraţi neesenţiali acţionează numai asupra unui anumit număr de unităţi din populaţie. Ca urmare, aceştia pot contribiu fie la creşterea nivelului variabilei (pentru unele unităţi din populaţie), fie la scăderea nivelului variabilei (pentru alte unităţi din populaţie). La rândul lor factor ii esenţiali nu acţionează cu aceeaşi intensitate asupra tuturor unităţilor din cadrul populaţie considerate, determinând, în acest fel, variaţia neuniformă a variabilei respective în cadrul populaţiei.

Se poate afirma că parametrul valoarea medie a unei serii statistice care are la bază variabila X, constituie acel nivel pe care l-ar putea înregistra variabila în cadrul populaţiei cercetate în condiţiile în care factorii neesenţiali nu s-ar fi manifestat, iar factorii esenţiali ar fi acţionat asupra unităţilor din populaţie cu aceeaşi intensitate. Parametrul valoarea medie, calculat pentru o serie statistică, pune în evidenţă ceea ce este comun, general şi esenţial sub aspectul nivelului de dezvoltare al variabilei, în raport cu care este studiată o populaţie. În raport cu natura variabilei ce stă la baza seriei, cât şi a formei de prezentare a indicatorilor cu care aceasta este construită, există mai multe posibilităţi de calcul a valorii medii. Funcţia determinată G, sub forma sa cea mai generală, are următoarea expresie analitică:

$$G(x_1, x_2, ..., x_R) = \left(\sum_{i=1}^{R} x_i^K \cdot f_i \right)^{\frac{1}{K}}$$

Pentru diverse valori ale lui k, în strictă concordanţă cu conţinutul şi semnificaţia funcţiei G, se întâlnesc mai multe tipuri de medii:

- media armonică (k = -1);
- media aritmetică (k = 1);
- media pătratică (k = 2);
- media cubică (k = 3);
- media de ordinul k în general.

În caz concret, valoarea medie reală este aceea care se obţine prin indicatorul (mediu) rezultat prin aplicarea criteriului relaţiei determinante.

Media aritmetică

A - Z

Definiţie

Media aritmetică este indicatorul cel mai utilizat în calculul parametrului valoarea medie a unei serii statistice, aşa cum rezultă din practica statistică.

Se consideră acum două serii statistice de repartiţie, una formată din frecvenţe absolute, iar cealaltă din frecvenţe relative:

$$X : \begin{pmatrix} x_i \\ N_i \end{pmatrix}_{i=.\overline{1,R}} \qquad X : \begin{pmatrix} x_i \\ f_i \end{pmatrix}_{i=.\overline{1,R}}$$

Media aritmetică pentru cele două serii este dată de relaţia:

$$M(X) = \overline{X} = \frac{\sum x_i N_i}{\sum N_j}; \qquad M(X) = \overline{X} = \sum x_i f_i$$

Calculul mediei aritmetice

Citire

Dacă seria este de intevale, construită cu frecvenţe absolute avem:

$$M(X) = \overline{X} = \frac{\sum x_i^{'} . N_i}{\sum N_j}$$

Fie o serie de repartiţie, care are la bază o variabilă continuă X, respectiv,

$$X : \begin{pmatrix} x_{i-1} - x_i \\ f_i \end{pmatrix}_{i=.\overline{1,R}}$$

Folosind notaţiile:

$$\frac{x_i + x_{i-1}}{2} = x_i^{'}$$

unde $x_i^{'}$ reprezintă mijlocul intervalului "i", obţinem relaţia:

$$M(X) = \overline{X} = \sum_{i=1}^{R} x_i^{'} \cdot f_i$$

Relaţia ne arată că media aritmetică a unei serii de intervale se reduce la media aritmetică a unei serii discrete în care clasele sunt reprezentate prin mijloacele intervalelor de variaţie.

Proprietăţi ale mediei aritmetice

Citire

1º Media aritmetică a unei constante este egală cu constanta respectivă, adică, dacă la baza unei serii se află o variabilă care a înregistrat o singură stare, X = C, atunci aplicând relaţia de calcul a valorii medii se obţine:

$$M(X) = M(C) = \overline{C} = C$$

2º Media produsului dintre o variabilă X şi o constantă k este egală cu produsul dintre media variabilei X şi constanta respectivă:

$$M(k \cdot X) = \sum_{i=1}^{R} k \cdot x_i \cdot f_i = k \cdot \sum_{i=1}^{R} x_i \cdot f_i = k \cdot \overline{X} = k \cdot M(X)$$

3º Media aritmetică a sumei a două sau mai multe variabile este egală cu suma mediilor acestora:

$$M(X1 + X2 + \ldots + Xr) = M(X1) + M(X2) + \ldots + M(Xr)$$

4º Media produsului a r variabile două câte două independente este egală cu produsul mediilor acestora:

$$M(X1 \cdot X2 \ldots XR) = M(X1) \cdot M(X2) \ldots M(Xr)$$

5º Proprietatea de adiţiune a mediei aritmetice:

Se presupune o populaţie care este structurată în raport cu un criteriu C cantitativ sau calitativ, în P clase, C1, C2, …, Cp. Atunci media aritmetică a variabilei X, în raport cu care este studiată populaţia, se poate obţine ca o medie a mediilor variabilei din cele P clase:

$$\overline{X} = \overline{X / C}$$

6º Media aritmetică a unei serii este cuprinsă între valoarea minimă şi valoarea maximă pe care o înregistreză variabila X, care stă la baza seriei.

$$x_{\min} \le \overline{X} \le x_{\max}$$

Această dublă inegalitate rezultă din următorul şir de inegalităţi:

$$x_{\min} = \frac{\sum_{i=1}^{R} x_{\min} \cdot N_i}{\sum_{i=1}^{R} N_i} \le \frac{\sum_{i=1}^{R} x_i \cdot N_i}{\sum_{i=1}^{R} N_i} = \overline{X} \le \frac{\sum_{i=1}^{R} x_{\max} \cdot N_i}{\sum_{i=1}^{R} N_i} = x_{\max}.$$

Calculul mediei aritmetice

Practică

Angajaţii unei societăţi comerciale se distribuie după salariul lunar cuvenit conform următoarei serii de reparaţie continuă:

$$X:\begin{pmatrix} 1{,}6-2{,}0 & 2{,}0-2{,}4 & 2{,}4-2{,}8 & 2{,}8-3{,}2 & 3{,}2-3{,}6 & 3{,}6-4{,}0 \\ 7 & 13 & 18 & 6 & 4 & 2 \end{pmatrix}$$

unde variabila X este exprimată în mii lei. Penwtru determinarea salariului mediu pe angajat, se recurge la transformarea seriei de intervale într-o serie discretă, după cum urmează:

$$X:\begin{pmatrix} 1{,}8 & 2{,}2 & 2{,}6 & 3{,}0 & 3{,}4 & 3{,}8 \\ 7 & 13 & 18 & 6 & 4 & 2 \end{pmatrix}$$

Salariul mediu se determină astfel:

$$\overline{X}=\frac{\sum_{i=1}^{6} x_i^{'}\cdot N_i}{\sum_{i=1}^{6} N_i}=\frac{1{,}8\cdot 7+2{,}2\cdot 13+2{,}6\cdot 18+3{,}0\cdot 6+3{,}4\cdot 4+3{,}8\cdot 2}{50}=\frac{12{,}72}{5}=2{,}544 \text{ mii lei}$$

Presupunem că este necesară determinarea ponderii agenţilor economici, dintr-un anumit judeţ care au înregistrat pierderi în anul calendaristic încheiat. Datele înregistrate de la instituţiile competente se aranjează într-o serie alternativă de forma:

$$X:\begin{pmatrix} 0 & 1 \\ 850 & 150 \end{pmatrix}$$

unde starea 1 pentru variabila X corespunde acelor agenţi care au înregistrat pierderi în anul considerat. Media (ponderea) agenţilor cu capital de stat care au înregistrat pierderi se determină astfel:

$$\overline{X}=\frac{\sum_{i=1}^{6} x_i^{'}\cdot N_i}{\sum_{i=1}^{6} N_i}=\frac{0\cdot 850+1\cdot 150}{1000}=\frac{150}{1000}=0{,}15 \ (15\%)$$

În judeţul considerat, 15% din agenţii economici cu capital de stat au înregistrat pierderi în anul calendaristic considerat.

Calculaţi

Pentru o serie de distribuţie extrasă şi utilizată în aplicaţiile anterioare, calculaţi:

- media aritmetică ponderată utilizând frecvenţele absolute
- media aritmetică ponderată utilizând frecvenţele absolute
- comparaţi rezultatul cu cel obţinut prin media aritmetică simplă şi comentaţi

Test

Pentru o serie de distribuţie bidimensională extrasă şi utilizată în aplicaţiile anterioare:

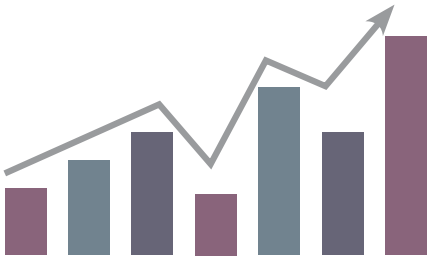
- calculaţi şi interpretaţi mediile distributiilor marginale
- calculaţi şi interpretaţi mediile distributiilor condiţionate
- demonstraţi proprietatea de adiţiune a mediei aritmetice

Investește în oameni!
Proiect cofinanțat din Fondul Social European prin
Programul Operațional Sectorial Dezvoltarea Resurselor Umane 2007 – 2013
Axa prioritară 1 „Educația și formarea profesională în sprijinul creșterii economice și dezvoltării societății bazate pe cunoaștere”
Domeniul major de intervenție 1.2 „Calitate în învățământul superior”
Titlul proiectului „e-Learning eficient, Individualizat și Adaptiv pentru învățământ la distanță (e-LIADA)”
Contract POSDRU/156/1.2/G/133681

Program:
Statistică descriptivă

Modul: Sistematizarea şi prezentarea datelor statistice
Tema: Mediana şi modala

(60’ teorie, 30’ practică)



I. Cuprinsul secvenței – Înțelesuri majore ale comunicării

1. Valoarea mediană – definire, mod de calcul, interpretare
2. Valoarea modală – definire, mod de calcul, interpretare

II. Obiectivele secvenței

- Însuşirea corectă a noţiunilor valoare mediană şi modală.
- Calculul şi interpretarea corectă a medianei şi modalei.

III. Cuvinte-cheie

Valoare mediană, valoare modală.

FONDUL SOCIAL EUROPEAN
Investește în OAMENI

ROMÂNIA
UNIVERSITATEA BABEŞ-BOLYAI CLUJ-NAPOCA
CENTRUL DE FORMARE CONTINUĂ, ÎNVĂȚĂMÂNT LA DISTANȚĂ ȘI CU FRECVENȚĂ REDUSĂ

Str Ion I.C.Brătianu, nr.20, Cluj-Napoca
Tel.: (00) 40 - 740 - 077521
Fax: (00) 40 - 364 - 815679

Valoarea mediană

Definiție

Valoarea mediană, notată cu M_e este acea valoare a variabilei cantitative X care împarte repartiția în două părți egale, respectiv:

$$F_N(M_e) = 1/2 \quad \text{sau} \quad N(M_e) = \frac{N}{2}$$

Calculul valorii mediane se face diferențiat, după cum seria are la bază o variabilă discretă sau continuă.

Pentru o repartiție discretă, calculul medianei nu implică probleme deosebite și nici un volum mare de calcule.

Se consideră o repartiție cu frecvențe absolute:

$$X : \begin{pmatrix} x_1 & x_2 & \dots & x_i & \dots & x_R \\ N_1 & N_2 & \dots & N_i & \dots & N_R \end{pmatrix}$$

În calculul valorii mediane a unei serii discrete, pot apărea două situații:

a) volumul N al populației este un număr impar;

b) volumul N al populației este un număr par.

În ambele cazuri, calculul medianei presupune, în prima fază, determinarea rangului medianei, notat cu r_{M_e}, vconform următoarei relații:

$$r_{M_e} = \frac{1}{2} \cdot \sum_{i=1}^{R} N_i = N(M_e)$$

a) Dacă volumul populației N este un număr impar, rangul medianei este un număr zecimal a cărui parte întreagă indică numărul de unități din populație pentru care variabila X a înregistrat valori mai mici ca mediana. Ca urmare, $\left[\frac{N}{2}\right]$ trebuie să fie valoarea imediat următoare celei de rang $\left[\frac{N}{2}\right]$ adică:

$$M_e = x_{\left(\left[\frac{N}{2}\right]+1\right)}$$

b) Dacă volumul populației este un număr par, rangul medianei este un număr întreg și ca urmare la mijlocul seriei nu se mai află o valoare a variabilei X cu care să coincidă mediana ci se găsesc două valori, mediana calculându-se în acest caz ca media aritmetică a acestora. Relația de calcul a medianei, în acest caz, este:

$$M_e = \frac{x_{\left[\frac{N}{2}\right]} + x_{\left(\left[\frac{N}{2}\right]+1\right)}}{2}$$

Pentru o repartiție continuă, calculul valorii mediane presupune verificarea egalității (2.12) și ca urmare, trebuie cunoscută densitatea de repartiție f(x). Determinarea funcției f(x) implică un volum mare de calcule și deci, din acest motiv, în activitatea practică f(x) este aproximat. Acest lucru va conduce la o expresie aproximativă de calcul a valorii mediane, care necesită un volum redus de calcule.

Pentru acesta se consideră o repartiție continuă în raport cu variabila X, și anume:

$$X : \begin{pmatrix} x_0 - x_1 & x_1 - x_2 & \dots & x_{i-1} - x_i & \dots & x_{R-1} - x_R \\ N_1 & N_2 & \dots & N_i & \dots & N_R \end{pmatrix}$$

unde intervalele xi-1-xi, pot fi de lungime egală sau neegală. Calcularea rangului medianei va permite stabilirea intervalului în care se află valoarea mediană, interval numit și interval median. Se cumulează frecvențele absolute din aproape în aproape până ce este îndeplinită inegalitatea:

$$N_1 + N_2 + \ldots + N_i \geq \frac{1}{2} N$$

Ultima frecvenţă Ni cumulată, ne permite să indicăm intervalul median.

$$[x_{i-1} - x_i)$$

Formula aproximativă de calcul a medianei:

$$M_e = x_{i-1} + \frac{N(M_e) - N(x_{i-1})}{N_i} \cdot (x_i - x_{i-1})$$

- $x_{i-1} = x_{M_e}$ - limita inferioară a intervalului median;
- $N_i = N_{M_e}$ - frecvenţa absolută a intervalului median;
- $x_i - x_{i-1} = l_{M_e}$ - lungimea intervalului median

Valoarea modală

Definiție

Valoarea modală Mo(X) a unei repartiţii reprezintă aceea valoare a variabilei X căreia îi corespunde frecvenţa cea mai mare.

Acest parametru se mai numeşte modul, valoare dominantă, sau modă se notează cu Mo.

Calculul modalei

Citire

Pentru o serie de repartiţie discretă, dată sub forma:

$$X : \begin{pmatrix} x_1 & x_2 & \ldots & x_i & \ldots & x_R \\ f_1 & f_2 & \ldots & f_i & \ldots & f_R \end{pmatrix}.$$

valoarea modală se citeşte direct din serie, nefiind nevoie de nici o tehnică sau formulă de calcul. În cazul acestui tip de serie, valoarea modală va fi acea valoare a variabilei X pentru care frecvenţa este cea mai mare.

Pentru serii de repartiţie continue, respectiv:

$$X : \begin{pmatrix} x_0 - x_1 & x_1 - x_2 & \ldots & x_{i-2} - x_{i-1} & x_{i-1} - x_i & x_i - x_{i+1} & \ldots & x_{R-1} - x_R \\ f_1 & f_2 & \ldots & f_{i-1} & f_i & f_{i+1} & \ldots & f_R \end{pmatrix}$$

Modala nu poate fi determinată direct. Intervalul căruia îi corespunde frecvenţa cea mai mare, se numeşte intervalul modal şi va conţine modala. Să presupunem că intervalul modal este xi-1-xi. Formula de calcul a modalei:

$$M_o(x) = x_{M_o} + \frac{\Delta_{-1}}{\Delta_{-1} + \Delta_1} \cdot l_{M_o}$$

- Mo - reprezintă valoarea modală;
- xMo - reprezintă limita inferioară a intervalului modal;
- Δ_{-1} - reprezintă diferenţa dintre frecvenţa intervalului modal şi frecvenţa intervalului precedent;
- Δ_1 - reprezintă diferenţa dintre frecvenţa intervalului modal şi frecvenţa intervalului următor;
- lMo - reprezintă lungimea intervalului modal.

O serie poate avea o singură valoare modală, caz în care seria se numeşte unimodală. Dacă o serie are mai multe valori modale, atunci se numeşte plurimodală. O serie plurimodală evidenţiază faptul că populaţia în cauză este neomogenă. Calculul valorii modale, în asemenea cazuri, presupune o delimitare mai riguroasă a obiectului

observării cât şi a populaţiei care urmează să fie studiată. O altă cale, care poate duce la eliminarea unui asemenea neajuns, o constituie comasarea a două câte două sau trei câte trei intervale etc., până se ajunge la o serie unimodală.

Calculul medianei

Practică

Următorul exemplu, privind calculul medianei unei serii, va permite evidenţierea tehnicii de calcul a acestui parametru cât şi semnificaţia lui. Repartiţia angajaţilor într-o societate comercială, în raport cu salariul lunar, este caracterizată prin următoarea serie:

$$X:\begin{pmatrix} 160-200 & 200-240 & 240-280 & 280-320 & 320-360 & 360-400 \\ 7 & 13 & 18 & 6 & 4 & 2 \end{pmatrix}$$

unde valorile variabilei X sunt exprimate în dolari.

Prima etapă în calculul medianei îl constituie calculul rangului, după cum urmează:

$$r_{M_e} = \frac{1}{2}\cdot\sum_{i=1}^{R} N_i = \frac{1}{2}(7+13+18+6+4+2) = \frac{50}{2} = 25$$

Însumând frecvenţele succesiv până ce valoarea obţinută depăşeşte rangul, rezultă că intervalul care va conţine mediana este [240-280). Fiind vorba de o serie care are la bază o variabilă continuă, mediana se determină conform relaţiei:

$$M_e = x_{M_e} + \frac{N(M_e) - N(x_{M_e})}{N_{M_e}}\cdot l_{M_e} = 240 + \frac{25-20}{18}\cdot 40 = 240 + \frac{100}{9} = 251{,}111\$$$

Ca urmare, jumătate din angajaţii societăţii respective câştigă cel mult 251,111$ şi cealaltă jumătate câştigă cel puţin 251,111$ în perioada în care s-a făcut observarea. Relevantă, în acest sens, pare a fi următoarea serie construită, având în vedere mediana:

$$\begin{pmatrix} 160{,}000-251{,}111 & 521{,}111-400{,}000 \\ 50\% & 50\% \end{pmatrix}$$

care constituie repartiţia celor două jumătăţi din populaţia angajaţilor societăţii respective.

Calculul modalei

Practică

Reconsiderăm exemplul privind distribuţia angajaţilor unei societăţi comerciale în raport cu salariul lunar:

$$X:\begin{pmatrix} 160-200 & 200-240 & 240-280 & 280-320 & 320-360 & 360-400 \\ 7 & 13 & 18 & 6 & 4 & 2 \end{pmatrix}$$

unde variabila X este exprimată în dolari. Calculul valorii modale presupune în primul rând găsirea intervalului modal.

Frecvenţa cea mai mare este 18 şi ca urmare, intervalul modal este 240-280. Intervalul modal fiind găsit, se poate aplica relaţia de calcul a valorii modale:

$$M_o = x_{M_o} + \frac{\Delta_{-1}}{\Delta_{-1}+\Delta_1}\cdot l_{M_o} = 240 + \frac{5}{5+12}\cdot 40 = 240 + \frac{200}{17} = 251{,}764\$$$

În concluzie, putem afirma că cei mai mulţi anagajaţi ai societăţii comerciale respective au un salar lunar în jur de 251,764 $.

Calculul medianei şi modalei

Calculați

Având în vedere populaţia societăţilor comerciale supusă observării concretizată în tabelul următor, calculaţia mediana şi modala pentru fiecare din variabilele:

- X1 – capital social;
- X2 – cifra de afaceri;
- X3 – profitul;

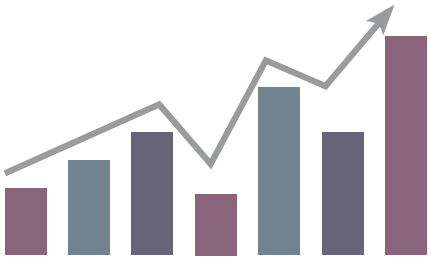
Nr	Capital social	Cifra de afaceri	Profitul
1.	8.083.365	2.309.871	0
2.	9.611.744	10.761.787	419.175
3.	10.116.125	7.694.483	467.496
4.	8.053.025	11.561.823	969
5.	2.121.925	2.317.951	80.880
6.	4.678.625	271.610	0
7.	888.490	159.351	0
8.	1.566.623	5.747.524	1.097
9.	1.076.050	5.377.189	17.363
10.	2.152.325	4.400.044	8.681
11.	2.990.126	871.025	0
12.	510.285	773.886	38.935
13.	507.277	330.839	0
14.	4.845.214	5.359.855	12.436
15.	5.963.750	9.931.327	90.223
16.	3.410.000	5.387.876	13.238
17.	6.270.418	541.334	0
18.	2.007.911	91.896	976
19.	3.354.625	6.855.127	145.649
20.	7.546.425	9.345.803	225.865
21.	4.116.425	3.720.863	375.052
22.	4.629.025	8.000.906	400.575
23.	6.848.825	176.728	1.960
24.	9.709.000	3.864.087	1.087
25.	520.686	542.748	0
26.	2.086.534	37.793	0
27.	1.891.958	988.769	62.668
28.	3.114.550	2.812.349	78.300
29.	5.389.225	12.456.122	390.659
30.	5.874.950	6.240.445	23.583

Investește în oameni!
Proiect cofinanțat din Fondul Social European prin
Programul Operațional Sectorial Dezvoltarea Resurselor Umane 2007 – 2013
Axa prioritară 1 „Educația și formarea profesională în sprijinul creșterii economice și dezvoltării societății bazate pe cunoaștere”
Domeniul major de intervenție 1.2 „Calitate în învățământul superior”
Titlul proiectului „e-Learning eficient, Individualizat și Adaptiv pentru învățământ la distanță (e-LIADA)”
Contract POSDRU/156/1.2/G/133681

Program:
Statistică descriptivă

Modul: Sistematizarea şi prezentarea datelor statistice
Tema: Parametrii variației

(60’ teorie, 30’ practică)



I. Cuprinsul secvenței – Înțelesuri majore ale comunicării

1. Definirea parametrilor variației
2. Modul de calcul al parametrilor variației
3. Proprietăţi ale varianţei

II. Obiectivele secvenței

- Însuşirea corectă a noţiunilor de varianţă şi abatere medie patratică
- Calculul şi utilizarea corectă a parametrilor variaţiei

III. Cuvinte-cheie

Variaţie, varianţă, abatere medie pătratică, coefficient de variaţie Pearson

FONDUL SOCIAL EUROPEAN
Investește în OAMENI

ROMÂNIA
UNIVERSITATEA BABEŞ-BOLYAI CLUJ-NAPOCA
CENTRUL DE FORMARE CONTINUĂ, ÎNVĂȚĂMÂNT
LA DISTANȚĂ ȘI CU FRECVENȚĂ REDUSĂ

Str Ion I.C.Brătianu, nr.20, Cluj-Napoca
Tel.: (00) 40 - 740 - 077521
Fax: (00) 40 - 364 - 815679

Rolul parametrilor variaţiei

Citire

Intensitatea diferită cu care se pot manifesta factorii esenţiali cât şi sensul contrar cu care pot acţiona factorii neesenţiali în raport cu fiecare unitate a unei populaţii statistice provoacă nivele diferite înregistrate de variabile în raport cu care este studiată populaţia. Problema măsurării variaţiei unei variabile cantitative este importantă pentru a vedea în ce măsură valoarea medie a acesteia poate reprezenta întrega populaţie.

Dacă abaterile de la valoarea medie sunt neesenţiale atunci se poate afirma că populaţia este omogenă şi că acest parametru poate reprezenta tendinţa centrală, iar dacă aceste abateri sunt mari atunci populaţia este eterogenă şi valoarea medie nu are capacitatea de a reprezenta populaţia. Parametrii variaţiei se pot calcula atât sub formă absolută cât şi relativă, şi măsoară împrăştierea valorilor unei variabile cantitative faţă de valoarea medie sau valoarea mediană.

Definirea abaterii medii pătratice şi a varianţei

Definiţie

Abaterea medie pătratică, notată cu σx , se defineşte ca fiind media pătratică a abaterilor valorilor variabilei X, de la valoarea medie $\overline{X}$, adică:

$$\sigma_x = \sqrt{M(X-\overline{X})^2}$$

Un calcul intermediar în aflarea acestui parametru, îl constituie calcularea pătratului abaterii medii pătratice, care se numeşte varianţă şi are următoarea expresie de calcul:

$$\sigma_x^2 = M(X-\overline{X})^2 = V(X)$$

Modul de calcul al abaterii medii pătratice şi al varianţei

Citire

Varianţa fiind un calcul intermediar în aflarea abaterii medii pătratice, în cele ce urmează se va prezenta modul de calcul al acesteia.

Relaţia de calcul a varianţei se particularizează în raport cu tipul seriei. În cazul unei serii care are la bază o variabilă X discretă, conform definiţiei, varianţa are expresia:

$$\sigma_x^2 = \frac{\sum_{i=1}^{R}(x_i-\overline{X})^2 \cdot N_i}{\sum_{i=1}^{R} N_i}$$

În cazul unei serii care are la bază o variabilă X continuă, varianţa se calculează conform următoarei relaţii:

$$\sigma_x^2 = \int_{x_1}^{x_R}\left(x-\overline{X}\right)^2 \cdot f(x)\cdot dx$$

a cărei aplicare presupune cunoaşterea densităţii de repartiţie f(x).

Pentru o serie dată, varianţa calculată nu are interpretare, dar dacă se extrage rădăcina pătrată din acesta se obţine un număr care se exprimă în aceleaşi unităţi de măsură ca şi variabila de la baza seriei. Acest număr (valoare) reprezintă abaterea medie pătratică, simbolizând cu cât se abate în medie în plus sau minus orice valoare xi a variabilei X de la valoarea medie $\overline{X}$.

Parametrul abaterea medie pătratică se poate exprima şi sub formă relativă, caz în care se numeşte coeficientul de variaţie Pearson, şi se notează cu Vx. Expresia de calcul este:

$$V_x = \frac{\sigma_x}{\overline{X}} \cdot 100\%$$

şi reprezintă abaterea medie a orcărei valori a variabilei X de la valoarea medie, considerată egală cu 1 sau 100%.

Coeficientul de variaţie a lui Pearson calculat pentru două sau mai multe serii, poate fi folosit în aprecieri comparative privind gradul de reprezentativitate a valorii medii calculate. Deoarece gradul de reprezentativitate a valorii medii este în raport invers cu mărimea coeficientului de variaţie a lui Pearson, se poate afirma, în cazul mai multor serii, că este mai reprezentativă valoarea medie a acelei serii pentru care Vx este mai mic. În concluzie, trebuie reţinut că parametrul abaterea medie pătratică sub formă absolută σx şi sub formă relativă Vx sunt indicatori fundamentali utilizaţi în măsurarea variaţiei unei variabile.

Proprietăţi ale varianţei

Citire

1. Varianţa unei constante este egală cu zero. Luându-se X = C şi aplicând relaţia de calcul a varianţei se obţine:

$$V(X) = V(C) = M[(C-C)2] = M(0) = 0$$

2. Varianţa unei variabile X este independentă de schimbarea originii. Ţinând cont de definiţia varianţei avem:

$$V(Y) = M(Y-\overline{Y})^2 = M[\ X - x_0) - M(X - x_0]^{\ 2} =$$
$$= M(X - x_0 - \overline{X} + x_0)^2 = M(X-\overline{X})^2 = V(X)$$

Dacă între două variabile Y şi X există următoarea relaţie liniară:

$$Y = a \cdot X + b$$

atunci are loc următoarea egalitate:

$$V(Y) = a2 \cdot V(X)$$

Conform definiţiei varianţei, se poate scrie:

$$V(Y) = M(Y-\overline{Y})^2$$

Valoarea medie a varibilei Y este:

$$M(Y) = \overline{Y} = M(a \cdot X + b) = a \cdot M(X) + b$$

care înlocuită ne conduce la:

$$V(Y) = M\left(Y-\overline{Y}\right)^2 = M\left(a \cdot X + b - a \cdot M(x) - b\right)^2 =$$
$$= M(a(x - M(x)^{\ 2} = a^2 M(X-\overline{X})^2 = a^2 \cdot V(X)$$

3. Varianţa unei variabile, se poate exprima cu ajutorul diferenţei dintre momentul de ordinul doi şi pătratul momentului de ordinul întâi al variabilei, adică:

$$V(X) = M\left(X^2\right) - \left(M(X)\right)^2$$

Următorul calcul simplu ne conduce la egalitatea de mai sus:

$$V(X) = M\left(X-\overline{X}\right)^2 = M(X^2 - 2X \cdot \overline{X} + \overline{X}^2) = M(X^2) - 2 \cdot \overline{X} \cdot M(X) + \overline{X}^2$$
$$= M(X^2) - 2\overline{X}^2 + \overline{X}^2 = M(X^2) - \overline{X}^2 = M(X^2) - (M(X)^{\ 2}$$

4. Varianţa sumei a două variabile dependente este egală cu suma varianţelor celor două variabile plus covarianţa dintre cele două variabile:

$$V(X+Y) = V(X) + V(Y) + 2M\left(\left(X-\overline{X}\right)\left(Y-\overline{Y}\right)\right)$$

5. Regula de adunare a varianţelor. Se consideră o serie de repartiţie, care are la bază variabila X, iar populaţia respectivă de volum N este împărţită în R clase în raport cu un criteriu C. Se demonstrează că varinţa în întrega populaţie, care măsoară variaţia totală în raport cu X, este egală cu suma a două varianţe din care una măsoară variaţia în cadrul grupelor, iar cealaltă variaţia dintre grupe. Variaţia lui X în cadrul populaţiei cercetate se compune din variaţia în cadrul celor R grupe datorată acţiunii factorilor neesenţiali şi variaţia de la o grupă la alta datorată acţiunii factorilor esenţiali care nu acţionează cu aceeaşi intensitate asupra unităţilor din populaţie. Notând cu:

- $\sigma^2_{x_i}$ - varianţa variabilei X în grupa "i" care măsoară împrăştierea în această grupă provocată de factori neesenţiali;
- $\sigma^2_{\overline{X_i}}$ - varianţa dintre grupe, care măsoară împrăştierea de la o grupă la alta;
- $\overline{\sigma^2_{x_i}}$ - media varianţelor grupelor, care măsoară variaţia medie în cadrul celor R grupe;
- $\overline{X_i}$ - media din grupa "i";

proprietatea enunţată mai sus se rezumă la următoarea egalitate:

$$\sigma^2_x = \sigma^2_{\overline{X_i}} + \overline{\sigma^2_{X_i}}$$

Se consideră populaţia agenţilor economici care îşi desfăşoară activitatea într-o anumită zonă geografică. Seria de distribuţie a celor 1160 de agenţi economici în raport cu profitul realizat este:

$$X:\begin{pmatrix} 0-50 & 50-100 & 100-150 & 150-200 & 200-250 \\ 170 & 250 & 350 & 260 & 130 \end{pmatrix}$$

unde variabila profit X are valorile exprimate în milioane lei.

Profitul mediu este:

$$\overline{X} = \frac{25\cdot 17 + 75\cdot 25 + 125\cdot 35 + 175\cdot 26 + 225\cdot 13}{116} = \frac{14{,}100}{116} = 121{,}55 \quad mil.lei$$

Varianţa calculată conform definiţiei este:

$$\sigma^2_X = \frac{\sum_{i=1}^{5}\left(x_i' - \overline{X}\right)^2 \cdot N_i}{\sum_{i=1}^{5} N_i} =$$

$$= \frac{(25-121{,}55)^2\cdot 17 + (75-121{,}55)^2\cdot 25 + (125-121{,}55)^2\cdot 35}{116} +$$

$$+ \frac{(175-121.55)^2.26 + (225-121.55)^2.13}{116} =$$

$$= \frac{(96{,}55)^2\cdot 17 + (46{,}55)^2\cdot 25 + (3{,}45)^2\cdot 35 + (53{,}45)^2\cdot 26 + (103{,}45)^2\cdot 13}{116} =$$

$$= \frac{426.465{,}68}{116} = 3676{,}428$$

de unde:

$$\sigma_X = \sqrt{\sigma^2_X} = \sqrt{3676{,}428} = 60{,}63 \quad mil.lei$$

Valoarea obţinută pentru σx de 60,63 mil. lei reprezintă cu cât se abate în medie profitul fiecărui agent economic de la profitul mediu considerat de 121,55 mil. Lei.

Abaterea medie pătratică exprimată sub formă relativă, prin intermediul coeficientului de variaţie Pearson este:

$$V_X = \frac{\sigma_X}{\overline{X}}\cdot 100\% = \frac{60{,}61}{121{,}55}\cdot 100\% \approx 49{,}86\%$$

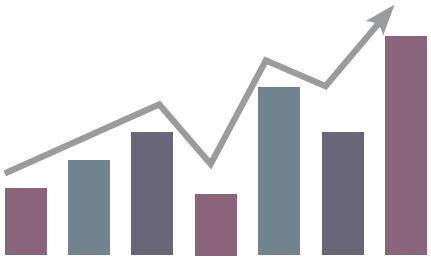
reprezentând abaterea medie în procente a profiturilor înregistrate de cei 1160 de agenţi economici de la profitul mediu.

Investește în oameni!
Proiect cofinanțat din Fondul Social European prin
Programul Operațional Sectorial Dezvoltarea Resurselor Umane 2007 – 2013
Axa prioritară 1 „Educația și formarea profesională în sprijinul creșterii economice și dezvoltării societății bazate pe cunoaștere”
Domeniul major de intervenție 1.2 „Calitate în învățământul superior”
Titlul proiectului „e-Learning eficient, Individualizat și Adaptiv pentru învățământ la distanță (e-LIADA)”
Contract POSDRU/156/1.2/G/133681

Program:
Statistică descriptivă

Modul: Sistematizarea şi prezentarea datelor statistice
Tema: Aspecte generale privind legăturile dintre variabile

(60’ teorie, 30’ practică)



I. Cuprinsul secvenței – Înțelesuri majore ale comunicării

1. Tipologia legăturilor dintre variabile
2. Aspecte punctuale privind utilizarea legăturilor de tip probabilist între variabile
3. Etapele tehnice parcurse în modelarea legăturilor

II. Obiectivele secvenței

- Înţelegerea corectă a scopului studiului legăturilor dintre variabile
- Învăţarea etapelor necesare prelucrărilor statistice în cazul legăturilor dintre variabile

III. Cuvinte-cheie

Legătură deterministă, legătură probabilistă, variabilă endogenă şi exogenă, variabilă reziduală, corelaţie, regresie

FONDUL SOCIAL EUROPEAN
Investește în
OAMENI

ROMÂNIA
UNIVERSITATEA BABEŞ-BOLYAI CLUJ-NAPOCA
CENTRUL DE FORMARE CONTINUĂ, ÎNVĂȚĂMÂNT
LA DISTANȚĂ ȘI CU FRECVENȚĂ REDUSĂ

Str Ion I.C.Brătianu, nr.20, Cluj-Napoca
Tel.: (00) 40 - 740 - 077521
Fax: (00) 40 - 364 - 815679

Tipologia legăturilor dintre variabile

Citire

În capitolul de faţă ne propunem abordarea unor metode statistice caracteristice studiului seriilor multidimensionale. Scopul acestora este de a identifica şi utiliza eventualele legături care se pot manifesta între două sau mai multe variabile. Ne pot interesa: existenţa legăturii, intensitatea acesteia, forma funcţională a legăturii, parametrii şi reprezentativitatea ei privind fenomenul cercetat.

Problematica legăturilor dintre variabile este foarte curent întâlnită în economie. Spunem că salariul unui angajat este în funcţie de productivitatea muncii sale, vechimea în muncă, responsabilitatea activităţii sale, etc; sau cererea dintr-un produs este în funcţie de preţul produsului, venitul consumatorilor, etc. De fiecare dată, atât în teoria economică, cât şi în aplicaţii se întâlneşte expresia "fie funcţia cererii…". În realitatea economică însă, această funcţie nu se dă, nu se cunoaşte, ci trebuie estimată pornind de la o bază de date. Această problemă de estimare a unei funcţii şi alte probleme colaterale ei fac obiectul acestui capitol. Pentru a putea aborda studiul legăturilor dintre variabile trebuie să ştim în primul rând dacă există sau nu o legătură între variabilele studiate (sau între fenomenele pe care acestea le reprezintă) şi care este natura acestora.

Definirea legăturii nule

Definiţie

Legătura nulă. Semnifică lipsa oricărei legături între două sau mai multe fenomene sau variabile care cuantifică fenomenele. De exemplu, o legătură nulă se manifestă între înălţimea unui angajat şi salariul acestuia sau între produsul intern brut al unei ţări şi vârsta primului ministru. Din punct de vedere statistic, spunem că între două variabile X şi Y există o legătură nulă, sau nu există legătură, dacă $\mathrm{cov}(x, y) = 0$.

Definirea legăturii deterministe

Definiţie

2. Legătura deterministă. Spunem că între variabilele X şi Y există o legătură deterministă dacă unei valori a lui X îi corespunde o singură valoare a lui Y. Astfel de legături se întâlnesc în special în fizică, unde de exemplu viteza este egală cu distanţa împărţită la timp: $v = d / t$, sau forţa este egală cu masa înmulţită cu acceleraţia: $F = m \cdot a$. Astfel de exemple există şi în economie, unde rata profitului este egală cu profitul împărţit la cifra de afaceri: $r_\pi = \pi / C.A. \cdot 100\%$. Legătura este deterministă pentru că variabila este perfect determinată de celelalte două: π şi $C.A.$ Adică pentru o anumită valoare a profitului şi o anumită valoare a cifrei de afaceri nu putem avea decât o singură valoare a ratei profitului.

Definirea legăturii statistice

Definiţie

3. Legătura statistică. Se mai numeşte şi stocastică sau probabilistă. Este tipul de legătură cel mai des întâlnită în ştiinţele sociale, deci şi în economie. Fiecărei valori a variabilei X îi corespunde o distribuţie de valori ale variabilei Y. Matematic, o astfel de legătură se exprimă sub forma $y = f(x) + \varepsilon$, unde am notat p rin ε componenta aleatoare reziduală, datorată acţiunii asupra lui Y a celorlalţi factori decât X.

Deşi s-ar putea spune că prin luarea în considerare a tuturor factorilor care influenţează variabila Y, legătura este intrinsec deterministă, în ştiinţele economice vom întâlni aproape întotdeauna un număr

foarte mare de factori, care nu pot fi identificaţi şi cuantificaţi în totalitatea lor. Asfel, funcţia care îl explicitează pe Y are două componente: una determnistă, $f(x_1, x_2, ..., x_n)$, cuprinzând variabilele cuantificabile de care depinde Y, şi una aleatoare, ε, cuprinzând variabilele ce nu au putut fi cuantificate.

Aspecte punctuale privind legăturile statistice

Citire

Studiul legăturilor dintre variabile s-a dezvoltat într-o disciplină aparte, numită econometrie. În capitolul de faţă nu ne propunem deci decât o introducere în această problematică, fără a aborda elemente de inferenţă statistică specifice acestor legături. În cele ce urmează vom prezenta câteva aspecte legate de variabile şi fenomenele reprezentate de acestea, probleme atât de natura aparatului statistic utilizat, cât şi de aplicabilitatea lui în contextul economic.

1. Natura probabilistă a legăturilor. Această natură are ca urmare un anume grad de nedeterminare ceea ce duce la o dificilă identificare a corelaţiilor dintre variabile. Nedeterminarea parţială se datorează multitudinii factorilor care determină un anumit fenomen. Mecanismul proceselor economice este adeseori deosebit de complex, presupunând existe nţa unui număr foarte mare de factori de luat în considerare. Struc turile cauzale se pot dovedi a fi complicate, cauza unui fenomen depinzând la rândul ei de o altă cauză. De asemenea legăturile pot fi atât univoce cât şi biunivoce, două variabile putând fi în acelaşi timp cauză şi efect. De multe ori chiar teoria economică poate clarifica situaţia, mecanismul formării cererii de exemplu fiind explicat prin preţ şi venit. În alte cazuri însă, din lipsa unei teorii pertinente într-un anumit domeniu se poate începe prin instrumentul statistic, iar dacă anumite ipoteze formulate se adeveresc prin efectuarea observaţiilor, se încearcă o justificare economică a rezultatului găsit. Această din urmă cale poate fi totuşi periculoasă dacă explicaţia economică nu este găsită, deoarece din punt de vedere statistic se poate identifica o corelaţie puternică între două variabile aparent fără nici o legătură între ele, deoarece se realizează prin intermediul unei a treia variabile, care poate fi cauza comună a celorlalte două.

Toată această argumentaţie mai sus expusă nu este făcută cu scopul de a încurca şi mai mult cititorul, ci pentru a atrage atenţia atât asupra coplexităţii problematicii acestui capitol, cât şi a modului de utilizare a aparatului statistic ce urmează a fi expus, nu într-un mod mecanic, după o anumită reţetă, ci în strânsă legătură cu teoria economică.

2. Variabilele endogene, exogene şi reziduale. În studiul legăturilor dintre fenomene intervin mai multe tipuri de variabile, după rolul jucat de ele în funcţia care explicitează legătura. Variabila efect, cea pe care dorim să o explicăm prin intermediul altor factori se numeşte variabilă endogenă (sau explicată sau dependentă) şi obişnium să o notăm cu Y. Factorii de influenţă (cei cuantificabili), variabilele prin care explicităm fenomenul, se numesc variabile exogene (sau explicative sau factoriale).

Obişnuim să le notăm cu $X_1, X_2, ..., X_n$. Variabilele care au influenţă asupra fenomenului cercetata, a variabilei endogene, dar care nu pot fi identificate sau cuantificate, se grupează sub forma unei variabile reziduale, notată cu ε.

Identificarea variabilelor explicative în scopul studierii influenţei lor asupra celei dependente se bazează pe cunoştinţele şi experienţa specialistului în domeniul cercetat, dar şi pe metode statistice specifice. În economie majoritatea legăturilor între variabile sunt cunoscute pornind de la teoria economică. Totuşi, fiecare situaţie economică concretă poate presupune o altă intensitatea a legăturii, o altă influenţă a fiecărui factor asupra variabilei enndogene. În alte cazuri însă, dependenţele teoretice nu se cunosc, ci trebuie identificate pe baza observării fenomenelor studiate. Aceste fenomene vor trebui însă urmărite în timp pentru a verifica dacă legătura este una conjuncturală sau ţine de un anumit mecanism economic.

În funcţie de situaţia concretă în fiecare domeniu sau caz, în alegerea variabilelor explicative se poate proceda în două moduri:

a) Introducerea succesivă a variabilelor în model (funcţie). Pentru început se aleg variabilele presupuse a fi cele mai semnificative. Dacă fenomenul cercetat nu este suficient de bine explicat de acestea, se mai introduc variabile până la obţinerea unei funcţii satisfăcătoare.

b) Eliminarea succesivă a variabilelor din model. Dacă funcţia care conţine toate variabilele explicative (sau potenţial explicative) este prea "încărcată", se poate renunţa la o parte din ele, care se dovedesc neesenţiale, adică explică o prea mică parte din fenomenul cercetat.

Alegerea uneia sau alteia din metode poate depinde şi de scopul cercetării. Scopul studiului trebuie să ne indice precis variabila de explicat şi ne poate sugera unele variabile explicative. Alegerea acestora însă, poate să fie condiţionată de bugetul alocat care să nu permită o extindere prea mare a numărului lor. În acelaşi timp un număr mare de variabile explicative explică mai bine un fenomen, dar face modelul mai complicat, mai dificil de înţeles şi utilizat. Scopul este deci identificarea unui model cât mai simplu, dar care să cuantifice cât mai mult variaţia explicită.

3. Multicoliniaritatea. Aşa cum am mai precizat, un model cuprinde o variabilă de explicat Y şi câteva variabile explicative $X_1, X_2, ..., X_n$. Acestea din urmă trebuie să fie independente între ele, dar corelate cu variabila dependentă. În cazul în care condiţia nu este îndeplinită, apare fenomenul de multicoliniaritate. Neluarea în considerare a acestuia duce la erori de estimare şi dificultăţi de utilizare a modelului. De exemplu, nu putem şti cum se modifică Y la o modificare a unei variabile factoriale oarecare X_i, în condiţiile în care celelalte rămân constante. Corelaţia dintre ele nu permite modificarea uneia fără modificarea celorlalte. În acest fel, nu putem simula prin funcţie o anumită situaţie prin modificarea valorilor variabilelor factoriale şi studiul impactului acestor modificări asupra celei endogene. Eliminarea multicoliniarităţii se face prin renunţarea la unele variabile între care există corelaţii. Putem proceda astfel, deoarece nu se pierde toată informaţia conţinută în ele, ea existând de asemenea în mare parte în variabilele rămase. Nu trebuie exagerată totuşi eliminarea, deoarece duce la o mai proastă specificare a varianţei explicate.

4. Măsurarea variabilelor. Variabilele din model trebuie exprimate în aşa fel încât să permită o cât mai uşoară şi corectă prelucrare, estimare şi utilizare. Este vorba atât de unitatea de măsură, cât şi de indicatorul în care sunt exprimate. Nu sunt rare situaţiile în care este preferată utilizarea diferenţelor relative şi absolute, a indicilor, logaritmilor, exponenţilor, etc.

5. Modelul matematic. Este o etapă care urmează în mod firesc după stabilirea variabilelor din model. Si aici în majoritatea cazurilor teoria economică este cea care ne furnizează şi forma legăturii. În multe situaţii ni se pare normal ca relaţia dintre factori trebuie să fie aditivă, multiplicativă, exponenţială, etc, sau după o funcţie compusă. Enumerăm aici cu titlu de exemplu funcţiile Cob-Douglas, Tornquist, Philips. În cazul în care tipul relaţiei matematice nu este precizat, acest rol revine instrumentului statistic. Cea mai simplă metodă o constituie utilizarea unui grafic adecvat punerii în evidenţă a formei legăturii. Totuşi, concluziile obţinute pe baza metodelor statistico-matematice trebuie corelate cu teoria economocă, pentru că putem trage adeseori concluzii pripite, datorate unor valori conjuncturale ale variabilelor, cu un rol foarte mare al factorilor reziduali.

6. Calitatea datelor. Datele ce se află la dispoziţia utilizatorului au de asemenea o mare importanţă. Calitatea proastă a acestora poate deteriora rezultatele modelului, chiar în urma parcurgerii reuşite a etapelor construirii acestuia.

Un aspect al calităţii datelor îl reprezintă numărul de observaţii, acurateţea şi corectitudinea culegerii şi înregistrării lor. O funcţie va fi cu atât mai bună cu cât se bazează pe o bază de date mai importantă. Ne putem uşor imagina ce concluzii eronate am putea să trgem dintr-o funcţie care explică intenţiile de vot pentru un candidat sau partid în funcţie de diverse variabile proprii electorului ca vârsta, profesia, venitul, etc, dacă studiul are loc pe folosind date cu privire la trei persoane. Un alt aspect îl reprezintă omogenitatea

plicate pe baza aceloraşi variabile. Pentru populaţii foarte eterogene se poate dovedi benefică o împărţire a acestora în clase şi construirea unui model în fiecare clasă. În cazul (şi este cel mai des întâlnit) în care datele provin dintr-un eşantion este necesară urmărirea reprezentativităţii eşantionului pentru populaţia din care a fost extras. Altfel, rezultatele modelului nu pot fi extinse pentru întreaga populaţie. Luând acelaşi exemplu de mai sus, ne putem imagina ce concluzii am putea trage la nivelul întregului electorat dacă datele disponibile s-ar referi numai la un singur segment de vârstă sau venit.

Etapele parcurse în modelarea legăturilor

Citire

Analiza legăturii dintre variabilele unei repartiţii multidimensionale presupune abordarea următoarelor probleme, care se pot constitui şi în etape ce trebuie parcurse în demersul statistic necesar: Aceste etape pot fi parcurse integral sau parţial, în funcţie de natura variabilelor. Pentru variabilele calitative nu vor fi parcurse (în statistica descriptivă) decât primele trei, deoarece posibilităţile de prelucrare sunt mai reduse. În schimb, toate cele şase etape pot fi parcurse în cazul variabilelor cantitative.

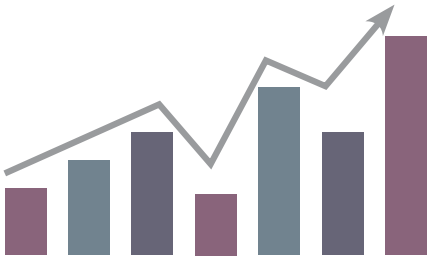
1. Organizarea rezultatelor observării populaţiei sau eşantionului în raport cu variabilele cercetate
2. Analiza statistică a existenţei legăturii
3. Analiza statistică a intensităţii legăturii sau a gradului de asociere dintre variabilele observate
4. Formularea unor ipoteze cu privire la forma matematică a legăturii
5. Estimarea parametrilor funcţiei de regresie
6. Analiza reprezentativităţii funcţiei de regresie

Investește în oameni!
Proiect cofinanțat din Fondul Social European prin
Programul Operațional Sectorial Dezvoltarea Resurselor Umane 2007 – 2013
Axa prioritară 1 „Educația și formarea profesională în sprijinul creșterii economice și dezvoltării societății bazate pe cunoaștere”
Domeniul major de intervenție 1.2 „Calitate în învățământul superior”
Titlul proiectului „e-Learning eficient, Individualizat și Adaptiv pentru învățământ la distanță (e-LIADA)”
Contract POSDRU/156/1.2/G/133681

Program:
Statistică descriptivă

Modul: Sistematizarea şi prezentarea datelor statistice
Tema: Legăturile dintre variabile calitative

(60’ teorie, 30’ practică)



I. Cuprinsul secvenței – Înțelesuri majore ale comunicării

1. Roblematica legăturilor dintre variabile calitative
2. Evidenţierea existenţei legăturii dintre variabile calitative
3.Măsurarea intensităţii legăturii dintre variabile calitative

II. Obiectivele secvenței

- Învăţarea tehnicilor necesare privind legăturile dintre variabilele calitative
- Interpretarea corectă a rezultatelor provenite din tehnicile aplicate

III. Cuvinte-cheie

Corelaţie statistică, testul χ^2, coeficientul de asociere Pearson

FONDUL SOCIAL EUROPEAN
Investește în
OAMENI

ROMÂNIA
UNIVERSITATEA BABEŞ-BOLYAI CLUJ-NAPOCA
CENTRUL DE FORMARE CONTINUĂ, ÎNVĂȚĂMÂNT
LA DISTANȚĂ ȘI CU FRECVENȚĂ REDUSĂ

Str Ion I.C.Brătianu, nr.20, Cluj-Napoca
Tel.: (00) 40 - 740 - 077521
Fax: (00) 40 - 364 - 815679

Tipologia legăturilor dintre variabile

Citire

Un prim instrument ce ne stă la îndemână este tabelul de corelaţie, un tabel cu două intrări, reprezentând o repartiţie bidimensională. Modul de construcţie al unui astfel de tabel se cunoaşte de la seriile statistice. O primă ipoteză asupra existenţei sau inexistenţei legăturii ne este furnizată de metode pur descriptive, ca studiul tabelului de corelaţie sau a norului de puncte asociat.

Urmărim dacă frecvenţele absolute N_{ij} iau valori apropiate în tot tabelul, caz în care nu există legătură între variabilele X şi Y, iar norul de puncte asociat prezintă o repartizare omogenă a punctelor pe domeniul de valori al variabilelor. Dacă frecvenţele se repartizează după una din diagonale sau o curbă oarecare sugerând existenţa unei legături, norul de puncte asociat prezintă o grupare a punctelor după o anumită formă (o dreaptă sau o curbă). Cu cât norul de puncte este mai grupat, cu atât ne putem aştepta mai mult la existenţa unei legături.

Metodele descriptive amintite anterior sunt imediate şi uşor de utilizat, dar nu pot stabili clar existenţa sau inexistenţa unei legături. Pentru aceasta, ne vom folosi de metode cantitative, din care cea mai utilizată este χ^2. După cum vom vedea, aceasta utilizează doar frecvenţele ca informaţie numerică. Evident, pentru variabilele cantitative, existenţa stărilor numerice permite folosirea unor metode care să ţină cont de acestea.

X / Y	x_1	x_2	…	x_i	…	x_I	Total
y_J	N_{1J}	N_{2J}	.	N_{iJ}	.	N_{IJ}	$N_{\bullet J}$
…	.	.	.	.	.	.	.
y_j	N_{1j}	N_{2j}	.	N_{ij}	.	N_{Ij}	$N_{\bullet j}$
…	.	.	.	.	.	.	.
y_2	N_{12}	N_{22}	.	N_{i2}	.	N_{I2}	$N_{\bullet 2}$
y_1	N_{11}	N_{21}	.	N_{i1}	.	N_{I1}	$N_{\bullet 1}$
Total	$N_{1\bullet}$	$N_{2\bullet}$	.	$N_{i\bullet}$	.	$N_{I\bullet}$	N

Metoda constă în esenţă în a compara frecvenţele absolute observate cu cele teoretice corespunzătoare cazului în care nu există legătură. Frecvenţele teoretice se vor calcula făcând apel la teoria probabilităţilor. Asociem tabelului o urnă cu N bile, reproducând structura populaţiei în raport cu variabilele X şi Y. Considerăm N_{ij} bile cu inscripţia (i, j). Dacă evenimentele $X = x_i$ şi $Y = y_i$ sunt independente, putem scrie:

$$P(X = x_i; Y = y_i) = P(X = x_i) \cdot P(Y = y_i) = \frac{N_{i\bullet}}{N} \cdot \frac{N_{\bullet j}}{N} = P_{ij}^*$$

notând cu P_{ij}^* produsul probabilităţilor celor două evenimente. Putem astfel calcula toate frecvenţele teoretice. Suma lor fiind egală cu 1, ele vor fi comparabile cu frecvenţele relative empirice. Pentru a le putea compara cu frecvenţele absolute, le vom înmulţi cu N. Astfel:

$$N_{ij}^* = P_{ij}^* \cdot N = \frac{N_{i\bullet}}{N} \cdot \frac{N_{\bullet j}}{N} \cdot N = \frac{N_{i\bullet} \cdot N_{\bullet j}}{N}$$

unde sunt frecvenţe absolute teoretice. Obţinem deci două tabele, unul cu frecvenţele absolute reale, observate N_{ij}^* şi altul cu frecvenţele absolute teoretice corespunzătoare cazului în care nu există legătură între variabile. Dacă între acestea nu există nici o diferenţă, concluzionăm că nu există legătură între variabile. Va trebui însă să ne bazăm în aprecieri pe o măsură globală a distanţei dintre cele două tabele, respectiv ale frecvenţelor teoretice şi empirice.

Definirea testului χ^2

Definiție

Definim testul ca o măsură a distanţei dintre cele două tabele, astfel:

$$\chi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(N_{ij} - N_{ij}^*)^2}{N_{ij}^*}$$

Observăm că dacă $N_{ij} = N_{ij}^*, \forall\, i = 1,\ldots, I\ j = 1,\ldots, J$ nu există legătură, iar $\chi^2 = 0$. Dimpotrivă, cu cât frecvenţele empirice diferă mai mult de cele teoretice, cu atât χ^2 ia valori mai mari.

Ca o concluzie, distingem cele două cazuri:

1) Dacă $\chi^2 = 0$ nu există legătură între variabile

2) Dacă $\chi^2 > 0$ există legătură între variabile

În statistica descriptivă ne limităm doar la a judeca egalitatea sau nonegalitatea cu zero. În realitate însă, pentru a garanta cu o anumită probabilitate existenţa legăturii, χ^2 trebuie să depăşească un anumit prag. Acesta ne este dat de statistica inferenţială prin legea de probabilitate χ^2. Astfel,

$$\chi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(N_{ij} - N_{ij}^*)^2}{N_{ij}^*}$$

este o variabilă χ^2 cu $v = (I-1)(J-1)$ grade de libertate. Vom compara deci valoarea calculată a lui χ^2 cu valorile tabelate. Dacă $\chi^2 > \chi^2_{tab(v,p)}$ putem afirma cu un risc de eroare p că există legătură între variabilele X şi Y. Valoarea tabelată o privim ca pe un prag peste care trebuie să treacă valoarea calculată a lui χ^2 pentru a accepta cu o anumită probabilitate existenţa legăturii. Cu cât riscul de eroare dorit este mai mic, cu atât "pragul" va fi mai exigent, adică $\chi^2_{tab(v,p)}$ va avea o valoare mai mare.

Definirea coeficientului de asociere Pearson

Definiție

Coeficientul de asociere Pearson se bazează pe metoda utilizată la existenţa legăturii, folosind frecvenţele absolute din tabelul de corelaţie. Se poate utiliza ca o metodă rapidă (această facilitate există chiar şi în soft-urile de birotică cum ar fi Microsoft Excel) şi pentru variabilele cantitative, dar se pierde o parte din informaţie prin neutilizarea stărilor numerice ale acestora.

Relaţia de calcul a coeficientului este:

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}}$$

unde N este volumul populaţiei. Ne interesează care sunt limitele acestui coeficient, pentru a-i putea aprecia valorile numerice pe care le ia. Reamintim că χ^2 este o măsură globală a distanţei dintre două repartiţii: cea observată şi cea teoretică corespunzătoare cazului în care nu există legătură între variabile.

- dacă $\chi^2 = 0$ atunci $C = 0$
- dacă $\chi^2 \to \infty$ atunci $C \to 1$

Ca urmare, $\chi^2 \in [0;1)$

Interpretarea gradului de asociere (intensităţii legăturii) pe baza coeficientului este următoarea:

- dacă $C = 0$ legătura este nulă (lipsa legăturii)
- dacă $C \in (0;0{,}3)$ legătura este de intensitate slabă

- dacă $C \in [0,3\,;0,7)$ legătura este de intensitate medie
- dacă legătura este de intensitate puternică

Desigur aceste limite nu trebuie interpretate într-o manieră foarte rigidă. O legătură este tot medie atât la 0,35 cât şi la 0,68. Prezintă mai mult interes compararea intensităţii legăturii dintre aceleaşi două variabile pentru aceeaşi populaţie în momente de timp diferite sau pentru populaţii similare.

Calculul medianei

Definiție

Următorul exemplu, privind calculul medianei unei serii, va permite evidenţierea tehnicii de calcul a acestui parametru cât şi semnificaţia lui. Repartiţia angajaţilor într-o societate comercială, în raport cu salariul lunar, este caracterizată prin următoarea serie:

$$X:\begin{pmatrix} 160-200 & 200-240 & 240-280 & 280-320 & 320-360 & 360-400 \\ 7 & 13 & 18 & 6 & 4 & 2 \end{pmatrix}$$

unde valorile variabilei X sunt exprimate în dolari.

Prima etapă în calculul medianei îl constituie calculul rangului, după cum urmează:

$$r_{M_e} = \frac{1}{2}\cdot\sum_{i=1}^{R} N_i = \frac{1}{2}(7+13+18+6+4+2) = \frac{50}{2} = 25$$

Însumând frecvenţele succesiv până ce valoarea obţinută depăşeşte rangul, rezultă că intervalul care va conţine mediana este [240-280). Fiind vorba de o serie care are la bază o variabilă continuă, mediana se determină conform relaţiei:

$$M_e = x_{M_e} + \frac{N(M_e)-N(x_{M_e})}{N_{M_e}}\cdot l_{M_e} = 240 + \frac{25-20}{18}\cdot 40 = 240 + \frac{100}{9} = 251{,}111\$$$

Ca urmare, jumătate din angajaţii societăţii respective câştigă cel mult 251,111$ şi cealaltă jumătate câştigă cel puţin 251,111$ în perioada în care s-a făcut observarea. Relevantă, în acest sens, pare a fi următoarea serie construită, având în vedere mediana:

$$\begin{pmatrix} 160{,}000-251{,}111 & 521{,}111-400{,}000 \\ 50\% & 50\% \end{pmatrix}$$

care constituie repartiţia celor două jumătăţi din populaţia angajaţilor societăţii respective.

Calculul existenţei şi intensităţii legăturii

Definiție

Dintre societăţile comerciale cotate la Bursa de Balori Bucureşti am extras un eşantion de 34 societăţi. Să se studieze existenţa şi intensitatea legăturii dintre variabilele în raport cu care s-a făcut observarea: X – domeniul de activitate al firmei şi Y – riscul acţiunilor societăţii.

X / Y	Industrie	Financiar-Bancar	Alte Domenii	Total
Mic	3	1	7	11
Mediu	10	2	4	16
Mare	2	5	-	7
Total	15	8	11	34

Avem astfel distribuţia empirică (observată) a societăţilor în raport cu cele două variabile. Indicatorul de la baza seriei este frecvenţa absolută:

$$N_{11} = 3 \qquad N_{12} = 1 \qquad N_{13} = 7$$

$$N_{21} = 10 \qquad N_{22} = 2 \qquad N_{23} = 4$$

$$N_{31} = 2 \qquad N_{32} = 5 \qquad N_{33} = 0$$

Deducem frecvenţele absolute teoretice, corespunzătoare cazului în care nu există legătură între variabile:

$$N_{11}^* = \frac{11 \cdot 15}{34} = 4{,}85 \qquad N_{12}^* = \frac{11 \cdot 8}{34} = 2{,}59 \qquad N_{13}^* = \frac{11 \cdot 11}{34} = 3{,}56$$

$$N_{21}^* = \frac{16 \cdot 15}{34} = 7{,}06 \qquad N_{22}^* = \frac{16 \cdot 8}{34} = 3{,}76 \qquad N_{23}^* = \frac{16 \cdot 11}{34} = 5{,}18$$

$$N_{31}^* = \frac{7 \cdot 15}{34} = 3{,}09 \qquad N_{32}^* = \frac{7 \cdot 8}{34} = 1{,}54 \qquad N_{33}^* = \frac{7 \cdot 11}{34} 2{,}26$$

Scriem şi sub formă tabelară această repartiţie teoretică:

X / Y	Industrie	Financiar-Bancar	Alte Domenii	Total
Mic	4,85	2,59	3,56	10
Mediu	7,06	3,76	5,18	16
Mare	3,091	1,65	2,26	7
Total	15	8	11	34

Calculăm distanţa dintre cele două tabele:

$$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(N_{ij} - N_{ij}^*)^2}{N_{ij}^*}$$

$$\chi^2 = \frac{(3-4{,}85)^2}{4{,}85} + \frac{(1-2{,}59)^2}{2{,}59} + \frac{(7-3{,}56)^2}{3{,}56} +$$

$$+ \frac{(10-7{,}06)^2}{7{,}06} + \frac{(2-3{,}76)^2}{3{,}76} + \frac{(4-5{,}18)^2}{5{,}18} +$$

$$+ \frac{(2-3{,}09)^2}{3{,}09} + \frac{(5-1{,}65)^2}{1{,}65} + \frac{(0-2{,}26)^2}{2{,}26} = 16{,}77$$

$\chi^2 > 0$ deci există legătură între variabile la nivel de eşantion

Pentru a verifica dacă legătura există şi la nivel de populaţie, facem apel la valorile tabelate ale distribuţiei de probabilitate χ^2.

I = 3 , J = 3 , $v = (3-1)(3-1) = 4$

- pentru $p = 5\%$, $\chi^2_{tab} = 9{,}49$, $\chi^2 > \chi^2_{tab}$ deci cu un risc de eroare de 5% acceptăm existenţa legăturii între cele două variabile
- pentru $p = 1\%$, $\chi^2_{tab} = 13{,}28$, $\chi^2 > \chi^2_{tab}$ deci cu un risc de eroare de 1% acceptăm existenţa legăturii între cele două variabile

Intensitatea legăturii:

- coeficientul de asociere al lui Pearson: $C = \sqrt{\frac{\chi^2}{N + \chi^2}}$ $\quad C = \sqrt{\frac{16{,}77}{34 + 16{,}77}} = 0{,}57$

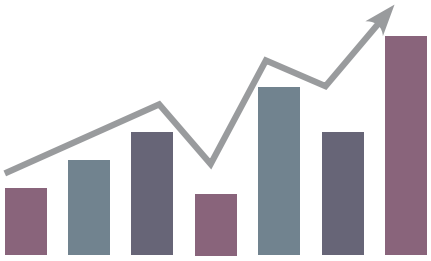
legătura dintre cele două variabile este de intensitate medie.

Investește în oameni!
Proiect cofinanțat din Fondul Social European prin
Programul Operațional Sectorial Dezvoltarea Resurselor Umane 2007 – 2013
Axa prioritară 1 „Educația și formarea profesională în sprijinul creșterii economice și dezvoltării societății bazate pe cunoaștere”
Domeniul major de intervenție 1.2 „Calitate în învățământul superior”
Titlul proiectului „e-Learning eficient, Individualizat și Adaptiv pentru învățământ la distanță (e-LIADA)”
Contract POSDRU/156/1.2/G/133681

Program:
Statistică descriptivă

Modul: Sistematizarea şi prezentarea datelor statistice
Tema: Legăturile dintre variabilele ordinale

(60’ teorie, 30’ practică)



I. Cuprinsul secvenței – Înțelesuri majore ale comunicării

1. Problematica legăturilor dintre variabile ordinale
2. Măsurarea intensităţii legăturii dintre variabile ordinale

II. Obiectivele secvenței

- Învăţarea tehnicilor necesare privind legăturile dintre variabilele ordinale
- Interpretarea corectă a rezultatelor provenite din tehnicile aplicate

III. Cuvinte-cheie

Corelaţie statistică, coeficient Kendall, coeficient Spearmann

FONDUL SOCIAL EUROPEAN
Investește în OAMENI

ROMÂNIA
UNIVERSITATEA BABEŞ-BOLYAI CLUJ-NAPOCA
CENTRUL DE FORMARE CONTINUĂ, ÎNVĂȚĂMÂNT
LA DISTANȚĂ ȘI CU FRECVENȚĂ REDUSĂ

Str Ion I.C.Brătianu, nr.20, Cluj-Napoca
Tel.: (00) 40 - 740 - 077521
Fax: (00) 40 - 364 - 815679

Specificul legăturilor dintre variabilele ordinale

Citire

Variabilele ordinale sunt tot variabile calitative, dar ale căror stări sunt ierarhizabile. Ca exemplu putem da inteligenţa sau performanţele fizice ale indivizilor, atractivitatea unei firme sau a unei ţări din punctul de vedere al investitorilor, renumele firmei, etc. Un grup de firme se poate ordona în raport cu renumele, dar pentru această variabilă nu există o unitate de măsură foarte obiectivă, astfel încât să putem spune că o firmă este de 2,7 ori mai renumită decât alta.

Ordonarea unităţilor populaţiei în raport cu stările variabilei aduce un plus de informaţie faţă de variabilele calitative neordinale, informaţie ce poate fi valorificată prin construirea unor coeficienţi specifici lor care să aprecieze intensitatea legăturii.

Coeficient de corelaţie a rangurilor Kendall

Definiție

Presupunem o populaţie de volum n observată în raport cu m variabile. Unităţilor statistice li se atribuie ranguri în raport cu fiecare variabilă după următoarea schemă redată în tabelul următor. Vom numi rang al unităţii statistice în raport cu variabila numărul natural care indică locul ocupat de valoarea variabilei înregistrată la unitatea în şirul ordonat al valorilor variabilei înregistrate la cele n unităţi considerate. Astfel, pentru fiecare variabilă, rangurile vor fi primele n numere naturale.

Coeficientul de corelaţie construit de Kendall va lua valoarea 1 pentru legătura directă de intensitate maximă şi –1 pentru legătura inversă de intensitate maximă. Pentru a putea construi coeficientul, vom defini mai întâi indicatorul de concordanţă (P) şi respectiv indicatorul de discordanţă (Q). În raport cu variabila X se ordonează crescător rangurile unităţilor, iar în raport cu variabila Y păstrăm ordinea unităţilor şi deci o succesiune oarecare (rezultată din ordonarea în raport cu X) a rangurilor.

	A_1	A_2	...	A_i	...	A_n
X	1	2	...	I	...	N
Y	r_1	r_2	...	r_i	...	r_n

unde r_i, $i = 1,...,n$ reprezintă unul şi numai unul din numerele naturale de la 1 la n.

Pentru fiecare rang r_i, $i = 1,...,n$, se determină numărul rangurilor mai mari decât r_i situate la dreapta, număr pe care îl notăm cu P_i. Însumând toate numerele P_i, $i = 1,...,n$ se obţine un număr notat cu P pe care îl numim indicator de concordanţă. În mod asemănător, pentru fiecare rang r_i, $i = 1,...,n$, se determină numărul rangurilor mai mici decât r_i situate la dreapta, număr pe care îl notăm cu Q_i. Însumând toate numerele Q_i, $i = 1,...,n$ se obţine un număr notat cu Q pe care îl numim indicator de discordanţă. Relativ la aceşti doi indicatori se verifică relaţia:

$$P + Q = \frac{n(n-1)}{2}$$

Pe baza indicatorilor de concordanţă şi discordanţă construim coeficientul de corelaţie simplă a rangurilor al lui Kendall, definit astfel:

$$\tau = \frac{P-Q}{P+Q} = \frac{P-Q}{\frac{n(n-1)}{2}}$$

În cazul unei legături directe de intensitate maximă, P va lua valoare sa maximă, iar Q pe cea minimă, adică: $P = \frac{n(n-1)}{2}$ iar $Q = 0$, deci $\tau = 1$. În cazul unei legături inverse de intensitate maximă, P va lua valoare sa mi nimă, iar Q pe cea maximă, adică: $P = 0$ iar, $Q = \frac{n(n-1)}{2}$ deci $\tau = -1$. În cazul lipsei legăturii,

$P = Q$, iar $\tau = 0$. Putem determina astfel intervalul în care va fi cuprins τ, respectiv $\tau \in [-1\ ;1]$. Interpretarea intensităţii legăturii pe baza acestui coeficient se va face astfel:

- dacă $\tau > 0$ egătura este directă
- dacă $\tau = 0$ legătura este nulă
- dacă $\tau < 0$ legătura este inversă
- dacă $|\tau| \in [0\,;0{,}3)$ legătura este de intensitate slabă
- dacă $|\tau| \in [0{,}3\,;0{,}7)$ legătura este de intensitate medie
- dacă $\tau \in [-1\,;1]$ legătura este de intensitate puternică

Calculul coeficientului Kendall

Definiţie

Vom lua ca exemplu 10 firme dintr-un anumit domeniu în raport cu variabilele X = renumele firmei (cât este de cunoscută în rândul consumatorilor) şi Y = încrederea în produsele firmei. Rezultă după observare următorul tabel al rangurilor.

	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
X	1	2	3	4	5	6	7	8	9	10
Y	2	3	1	5	6	4	7	9	8	10

Calculăm indicatorii de concordanţă şi discordanţă:

$$P = 8+7+7+5+4+4+3+1+1 = 40$$

$$Q = 1+1+0+1+1+0+0+1+0 = 5$$

De unde rezultă coeficientul de corelaţie a rangurilor al lui Kendall:

$$\tau = \frac{P-Q}{P+Q} = \frac{40-5}{40+5} = 0{,}78$$

deci între renumele firmei şi încrederea în produsele sale există o legătură irect foarte puternică.

Coeficient de corelaţie multiplă Kendall

Definiţie

În cazul mai general, cu m variabile, pentru o populaţie de volum n, legătura va fi de intensitate maximă dacă dacă fiecare unitate a populaţiei are acelaşi rang în raport cu toate variabilele. Deoarece între două variabile poate exista o legătură directă, între altele două o legătură inversă, nu putem aprecia pe ansamblu un sens al legăturii. De aceea, coeficientul va fi unul de cuncordanţă şi nu de corelaţie. El se defineşte tot pe baza tabelului rangurilor, considerând următorul sistem de numere:

$$S = \{S_1, S_2, \ldots, S_k, \ldots, S_n\}$$

unde S_k, $k = 1,\ldots,n$ este suma rangurilor pentru cele m variabile în raport cu unitatea A_k. Dacă legătura este directă de intensitate maximă, avem:

$$S = \{m\,, 2m\,, \ldots, m\ \}$$

Dacă nu există legătura (rangurile sunt perfect amestecate) avem:

$$S_1 = S_2 = \ldots = S_k = \ldots = S_n$$

iar dispersia lor va fi nulă.

Calculăm dispersia în cazul legăturii de intensitate maximă:

$$M(S) = \frac{m + 2m + \ldots + nm}{n} = \frac{m(n+1)}{2}$$

$$M(S^2) = \frac{m^2 + (2m)^2 + \ldots + (nm)^2}{n} = \frac{m^2(n+1)(2n+1)}{6}$$

$$\sigma_S^2 = M(S^2) - [M(S)]^2 = \frac{m^2(n^2-1)}{12}$$

Considerând că dispersia efectivă a sistemului S este proporţională cu concordanţa dintre cele m variabile, coeficientul de concordanţă multiplă a rangurilor al lui Kendall se defineşte astfel:

$$k = \frac{\sigma_S^2(\text{efectiva})}{\sigma_S^2(\text{maxima})} = \frac{\sigma_S^2(efectiva)}{\frac{m^2(n^2-1)}{12}}$$

Deoarece ne arată doar intensitatea, nu şi sensul legăturii, . Interpretarea intensităţii legăturii pe baza acestui coeficient se va face astfel:

- dacă $k = 0$ legătura este nulă
- dacă $k \in [0\,;0{,}3)$ legătura este de intensitate slabă
- dacă $k \in [0{,}3\,;0{,}7)$ legătura este de intensitate medie
- dacă $k \in [0{,}7\,;1]$ legătura este de intensitate puternică

Calculul coeficientului multiplu Kendall

Practică

Ca exemplu, luăm aceleaşi 10 fime, dar pe lângă variabilele X = renumele firmei (cât este de cunoscută în rândul consumatorilor) şi Y = încrederea în produsele firmei vom lua în calcul şi Z = performanţele economice (poate fi cifra de afaceri, profitul, mărimea dividendelor, etc). Distribuţia rangurilor în raport cu variabilele X, Y şi Z :

	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
X	1	2	3	4	5	6	7	8	9	10
Y	2	3	1	5	6	4	7	9	8	10
Z	2	4	1	5	3	7	8	6	10	9
S	5	9	5	14	14	17	22	23	27	29

$$M(S) = \frac{5+9+5+14+14+17+22+23+27+29}{10} = 16{,}5$$

$$M(S^2) = \frac{5^2+9^2+5^2+14^2+14^2+17^2+22^2+23^2+27^2+29^2}{10} = 339{,}5$$

$$\sigma_S^2(\text{efectiva}) = M(S^2) - [M(S)]^2 = 339{,}5 - 16{,}5^2 = 67{,}25$$

$$\sigma_S^2(\text{maxima}) = \frac{m^2(n^2-1)}{12} = \frac{3^2(10^2-1)}{12} = 74{,}25$$

$$k = \frac{\sigma_S^2(\text{efectiva})}{\sigma_S^2(\text{maxima})} = \frac{67{,}25}{74{,}25} = 0{,}91$$

Există deci o legătură de intensitate foarte mare între renumele firmei, încrederea în produsele ei şi performanţele sale economice.

Coeficient de corelaţie Spearman

Definiţie

Ca şi coeficientul similar propus de Kendall, şi acesta se calculează pornind de la tabelul de concordanţă a rangurilor. Ne vom folosi de diferenţele dintre ranguri pentru aceeaşi unitate a populaţiei relativ la cele două variabile. Coeficientul are următoarea expresie:

$$\eta = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

Calculul coeficientului Spearman

Practică

Vom lua ca exemplu numeric aceeaşi populaţie şi variabile ca şi în cazul coeficientului similar al lui Kendall. Limitele celor doi coeficienţi sunt aceleaşi, la fel şi interpretările valorilor numerice.

Distribiţia rangurilor şi a diferenţelor dintre ranguri:

	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
X	1	2	3	4	5	6	7	8	9	10
Y	2	3	1	5	6	4	7	9	8	10
d_i	-1	-1	2	-1	-1	2	0	-1	1	0

$$\eta = 1 - \frac{6(1+1+4+1+1+4+0+1+1+0)}{10(10^2-1)} = 0{,}92$$

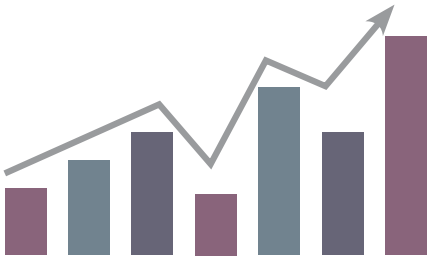
Există deci o legătură puternică între cele două variabile. Valoarea numerică este diferită faţă de cea a coeficientului lui Kendall. Nu trebuie privite însă într-un mod foarte rigid valorile obţinute. Nu este importantă alegerea unuia sau altuia din coeficienţi, ci folosirea aceluiaşi coeficient pentru aceeaşi populaţie la momente diferite de timp sau populaţii perechi.

Investește în oameni!
Proiect cofinanțat din Fondul Social European prin
Programul Operațional Sectorial Dezvoltarea Resurselor Umane 2007 – 2013
Axa prioritară 1 „Educația și formarea profesională în sprijinul creșterii economice și dezvoltării societății bazate pe cunoaștere”
Domeniul major de intervenție 1.2 „Calitate în învățământul superior”
Titlul proiectului „e-Learning eficient, Individualizat și Adaptiv pentru învățământ la distanță (e-LIADA)”
Contract POSDRU/156/1.2/G/133681

Program:
Statistică descriptivă

Modul: Sistematizarea şi prezentarea datelor statistice
Tema: Regresia liniară

(60’ teorie, 30’ practică)



I. Cuprinsul secvenței – Înțelesuri majore ale comunicării

1. Specificarea matematică a regresiei liniare
2. Calculul coeficientilor regresiei liniare
3. Reprezentativitatea funcţiei de regresie

II. Obiectivele secvenței

- Învăţarea tehnicilor necesare pentru calculul coeficienţilor regresiei liniare
- Utilizarea corectă a regresiei liniare pentru modelarea fenomenelor economice

III. Cuvinte-cheie

Regresie liniară, coefficient de regresie, coefficient de corelaţie liniară Pearson

FONDUL SOCIAL EUROPEAN
Investește în
OAMENI

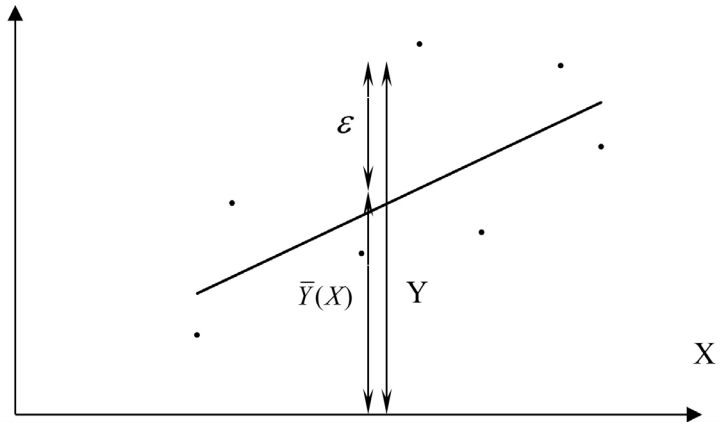
ROMÂNIA
UNIVERSITATEA BABEŞ-BOLYAI CLUJ-NAPOCA
CENTRUL DE FORMARE CONTINUĂ, ÎNVĂȚĂMÂNT
LA DISTANȚĂ ȘI CU FRECVENȚĂ REDUSĂ

Str Ion I.C.Brătianu, nr.20, Cluj-Napoca
Tel.: (00) 40 - 740 - 077521
Fax: (00) 40 - 364 - 815679

Estimarea coeficienţilor regresiei liniare

Citire

Această etapă succede alegerii formei funcţiei. În estimarea coeficienţilor va trebui să ţinem cont de abaterea punctelor norului faţă de modelul matematic ales $\hat{Y}$, datorat altor factori decât $X_1, X_2, ..., X_n$, consideraţi neesenţiali, cuantificaţi prin variabila reziduală ε. Principiul de la care se porneşte în estimarea parametrilor este cel al pătratelor minime. Minimizăm suma pătratelor abaterilor valorilor observate ale lui Y de la nivelul calculat prin $\hat{Y}$.



Condiţia de minim a sumei este echivalentă cu condiţia de minim a mediei:

$$M\left[Y - \hat{Y}\right]^2 = M(\varepsilon^2)$$

Ecuaţia $\hat{Y}(X_1, X_2, ..., X_n)$ care descrie legătura dintre Y şi factorii de influenţă $X_1, X_2, ..., X_n$ se numeşte ecuaţia de regresie. Metoda regresiei constă în modelarea legăturilor statistice prin ecuaţia de regresie. Deoarece problema de minim se poate rezolva doar cunoscând forma particulară a funcţiei, vom aborda estimarea coeficienţilor seprat, pe tipuri de funcţii.

Estimarea generală a coeficienţilor prin metoda pătratelor minime

Citire

Sub formă generală, pentru observaţia i, regresia liniară multiplă se scrie:

$$y_i = a_0 + a_1 x_{1i} + a_2 x_{2i} + ... + a_n x_{ni} + \varepsilon_i$$

- $i = \overline{1, I}$ indexează observaţiile (unităţile statistice)
- I numărul de observaţii
- y_i valoarea variabilei endogene Y pentru observaţia i
- x_{1i} valoarea variabilei explicative X1 pentru observaţia i
- x_{2i} valoarea variabilei explicative X2 pentru observaţia i

  .....
- x_{ni} valoarea variabilei explicative Xn pentru observaţia i
- ε_i valoarea variabilei reziduale pentru observaţia i
- $a_0, a_1, a_2, ..., a_k$ coeficienţii de regresie care trebuie estimaţi.

Dacă rescriem modelul observaţie cu observaţie, obţinem:

$$
\begin{aligned}
&y_1 = a_0 + a_1x_{11} + a_2x_{21} + \ldots + a_nx_{n1} + \varepsilon_1\\
&y_2 = a_0 + a_1x_{12} + a_2x_{22} + \ldots + a_nx_{n2} + \varepsilon_2\\
&\ldots\ldots\\
&y_i = a_0 + a_1x_{1i} + a_2x_{2i} + \ldots + a_nx_{ni} + \varepsilon_i\\
&\ldots\ldots\\
&y_I = a_0 + a_1x_{1I} + a_2x_{2I} + \ldots + a_nx_{nI} + \varepsilon_I
\end{aligned}
$$

ceea ce sub formă matricială (între paranteze numărul de linii şi respectiv de coloane) devine:

$$
\underset{(T,1)}{Y} = \underset{(T,k+1)}{X} \cdot \underset{(k+1,1)}{a} + \underset{(T,1)}{\varepsilon}
$$

unde:

$$
Y = \begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_i \\ \ldots \\ y_I \end{pmatrix} \qquad X = \begin{pmatrix} 1 & x_{11} & x_{21} & \ldots & x_{n1} \\ 1 & x_{12} & x_{21} & \ldots & x_{n2} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{1i} & x_{2i} & \ldots & x_{ni} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{1I} & x_{2I} & \ldots & x_{nI} \end{pmatrix}
$$

$$
a = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \ldots \\ a_n \end{pmatrix} \qquad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \ldots \\ \varepsilon_i \\ \ldots \\ \varepsilon_I \end{pmatrix}
$$

Prima coloană a matricei X conţine doar valoarea 1, care corespunde coeficientului a_0. Astfel matricea X are I linii şi n+1 coloane (n variabile explicative plus constanta).

Considerăm modelul sub formă matricială, cu n variabile explicative şi I observaţii:

$$
Y = X \cdot a + \varepsilon
$$

Pentru estimarea vectorului a, având ca şi componente coeficienţii $a_0, a_1, a_2, \ldots, a_n$ aplicăm ca şi la modelul liniar simplu metoda patratelor minime (MPM), minimizând suma patratelor erorilor.

Notăm: $S = \sum_{i=1}^{I} \varepsilon_i^2$

$$
\begin{aligned}
\min \sum_{i=1}^{T} \varepsilon_i^2 &= \min(\varepsilon'\varepsilon)\\
&= \min(Y - Xa)'(Y - Xa)\\
&= \min(Y'Y - Y'Xa - a'X'Y + a'X'Xa)\\
&= \min(Y'Y - 2a'X'Y + a'X'Xa)
\end{aligned}
$$

unde :

- ε' este vectorul transpus al lui ε;
- a' este vectorul transpus al lui a;
- Y' este vectorul transpus al lui Y;
- X' este transpusa matricii X.

Pentru a minimiza expresia, derivăm în raport cu a :

$$\frac{\partial S}{\partial a} = -2X'Y + 2X'X\,a = 0$$

$$a = (X'X)^{-1}X'Y$$

Această soluţie este realizabilă dacă matricea patratică este inversabilă. Matricea este este neinversabilă doar în caz de coliniaritate perfectă între oricare două variabile explicative.

Acest mod de calcul al coeficienţilor de regresie este mai dificil de efectuat manual. Principalul obstacol este inversarea matricii . Desigur, tehnica de calcul actuală rezolvă simplu această dificultate. Există soft-ware de specialitate care calculeaza prin proceduri simple coeficienţii de regresie. Câteva exemple de astfel de soft-uri de statistică şi econometrie: SPSS, SAS, STATA, Eviews, LIMDEP, GAUSS.

Estimarea coeficienţilor regresiei liniare

Citire

Particularizarea calculului coeficienţilor de regresie liniară pentru un model liniar simplu

În ipoteza în care legătura dintre Y şi factorii săi de influenţă $X_1, X_2, ..., X_n$ este liniară, ecuaţia de regresie va fi de forma:

$$\hat{Y} = a_0 + a_1X_1 + a_2X_2 + ... + a_nX_n$$

Coeficienţii $a_0, a_1, a_2, ..., a_n$ modelului vor rezulta din minimizarea următoarei funcţii cu (n+1) necunoscute:

$$G(a_0, a_1, ..., a_n) = M\left[Y - (a_0 + a_1X_1 + ... + a_nX_n)\right]^2$$

Condiţiile de minim constau în anularea celor (n+1) derivate parţiale ale funcţiei $G(a_0, a_1, .., a_n)$ în raport cu necunoscutele $a_0, a_1, ..., a_n$, ceea ce conduce la următorul sistem de ecuaţii:

$$\begin{cases} \dfrac{\partial G(a_0, a_1, ..., a_n)}{\partial a_0} = -2M\left[Y - (a_0 + a_1X_1 + ... + a_nX_n)\right] = 0 \\ \dfrac{\partial G(a_0, a_1, ..., a_n)}{\partial a_j} = -2M\left[Y - (a_0 + a_1X_1 + ... + a_nX_n)\right] \cdot X_j = 0 \end{cases} \quad \forall j = \overline{1,n}$$

sau într-o formă echivalentă:

$$\begin{cases} M\left(a_0 + a_1X_1 + ... + a_nX_n\right) = M(Y) \\ M\left(a_0 + a_1X_1 + ... + a_nX_n\right) \cdot X_n = M(Y) \cdot X_j \end{cases} \quad \forall j = \overline{1,n}$$

de unde rezultă:

$$\begin{cases} a_0 + a_1M(X_1) + ... + a_nM(X_n) = M(Y) \\ a_0 + a_1M(X_1X_j) + ... + a_nM(X_nX_j) = M(YX_j) \end{cases} \quad \forall j = \overline{1,n}$$

Prin rezolvarea acestui sistem liniar de ecuaţii în raport cu necunoscutele $a_0, a_1, ..., a_n$, se obţin valorile coeficienţilor ecuaţiei de regresie. Astfel, legătura statistică dintre Y şi $X_1, X_2, ..., X_n$ este modelată prin aproximare cu o legătură funcţională. Pentru cazul unui singur factor X_1, ecuaţia de regresie se scrie:

$$\hat{Y} = a_0 + a_1X_1$$

iar sistemul de ecuaţii devine:

$$\begin{cases} a_0 + a_1M(X_1) = M(Y) \\ a_0M(X_1) + a_1M(X_1^2) = M(YX_1) \end{cases}$$

Sau sub formă echivalentă:

$$\begin{cases} a_1 = \dfrac{M(YX_1) - M(Y)M(X_1)}{M(X_1^2) - [M(X_1)]^2} \\ a_0 = M(Y) - a_1 M(X_1) \end{cases}$$

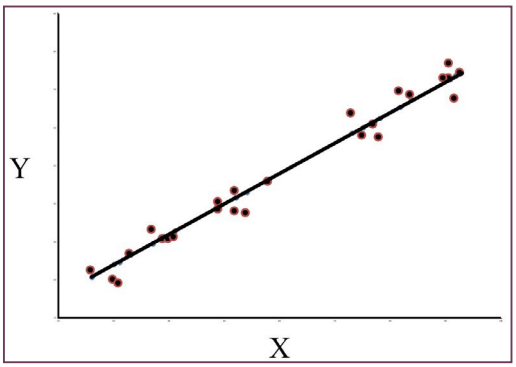
Analiza reprezentativităţii regresiei liniare

Citire

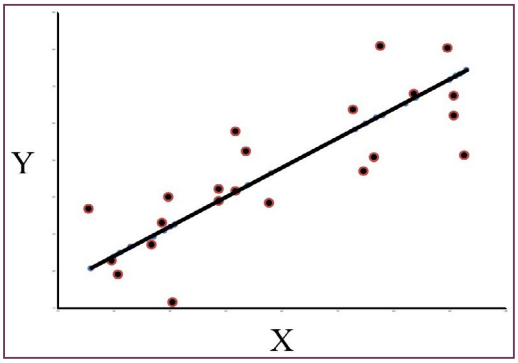
Raport de determinare şi coeficient de corelaţie

Funcţia de regresie obţinută prin metoda pătratelor minime poate fi o ajustare mai mult sau mai puţin performantă a distribuţiei norului de puncte. Figurile de mai jos ilustrează două situaţii diferite din acest punct de vedere.

Ajustare bună prin funcţia de regresie



Ajustare mai slabă prin funcţia de regresie



Cu cât punctele norului sunt mai apropiate de linia de regresie, cu atât ajustarea este mai bună. Desigur, graficul ne arată doar o imagine a ajustării. Pentru a o măsura numeric facem apel la analiza varianţei (ANOVA – Analysis Of Variance). Se compară varianţa "explicată", datorată regresiei cu varianţa totală a variabilei endogene (Y).

Pentru orice formă a funcţiei de regresie şi orice număr de variabile explicative, folosim notaţiile:

$$Y = f(X_1, X_2, ..., X_n) + \varepsilon$$

$$\hat{Y} = f(X_1, X_2, ..., X_n)$$

Descompunerea varianţei totale a lui Y în raport cu regresia este:

$V_{totală} = V_{explicată} + V_{reziduală}$

unde:

- Varianţa totală: $V_{tot} = \dfrac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}$
- Varianţa explicată: $V_{exp} = \dfrac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{n}$
- Varianţa reziduală: $V_{rez} = \dfrac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} = \dfrac{\sum_{i=1}^{n}\varepsilon_i^2}{n}$

Folosind aceste componente ale varianţei definim raportul de determinare definit astfel:

$$R^2 = \frac{V_{\exp}}{V_{tot}} = 1 - \frac{V_{rez}}{V_{tot}}$$

$$0 \le R^2 \le 1$$

Cu cât R2 este mai mare, cu atât calitatea ajustării este mai bună. Această remarcă corespunde şi imaginii vizuale pe care ne-o facem cu privire la ajustare. O valoare mare a lui R2 corespunde unei valori mici a lui V_{rez} respectiv unor valori ε mici.

Coeficientul de corelaţie liniară Pearson

În cazul particular al unei regresii liniare simple de ecuaţie $Y = a + bX + \varepsilon$, rădăcina pătrată a raportului de determinare $R = \sqrt{R^2}$ se numeşte coeficient de corelaţie liniară Pearson.

$$R = \sqrt{\frac{V_{\exp}}{V_{\text{tot}}}} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad \text{unde } \hat{Y} = a + bX,$$

deci $\hat{y}_i = a + bx_i$.

Înlocuind valorile coeficientilor a şi b cu expresiile lor deduse din metoda celor mai mici pătrate obţinem o formulă echivalentă pentru R:

$$\lg \hat{Y} = \lg a + X \cdot \lg b$$

Dacă R>0, legătura dintre variabilele X şi Y este directă, adică ele variază în acelaşi sens: Y creşte când X creşte. Dacă R<0, legătura dintre variabilele X şi Y este inversă, adică ele variază în acelaşi opus: Y scade când X creşte. Aceste interpretări sunt coerente cu panta dreptei:

$$b = \frac{\text{cov}(X,Y)}{\sigma_X^2}$$

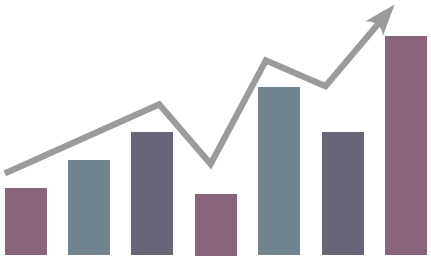
Diferenţa dintre expresiile lui R şi b este doar la numitor, care oricum este un termen pozitiv. Ca urmare, R şi b au acelaşi semn. Deci dacă există o legătură directă între X şi Y, aceasta va fi ajustată printr-o dreaptă de pantă pozitivă. Dimpotrivă, dacă există o legătură inversă între X şi Y, aceasta va fi ajustată printr-o dreaptă de pantă pozitivă.

Investește în oameni!
Proiect cofinanțat din Fondul Social European prin
Programul Operațional Sectorial Dezvoltarea Resurselor Umane 2007 – 2013
Axa prioritară 1 „Educația și formarea profesională în sprijinul creșterii economice și dezvoltării societății bazate pe cunoaștere”
Domeniul major de intervenție 1.2 „Calitate în învățământul superior”
Titlul proiectului „e-Learning eficient, Individualizat și Adaptiv pentru învățământ la distanță (e-LIADA)”
Contract POSDRU/156/1.2/G/133681

Program:
Statistică descriptivă

Modul: Sistematizarea şi prezentarea datelor statistice
Tema: Regresii neliniare şi aplicații ale regresiei

(60’ teorie, 30’ practică)



I. Cuprinsul secvenței – Înțelesuri majore ale comunicării

1. Specificarea matematică a regresiilor neliniare
2. Calculul coeficientilor regresiilor neliniare
3. Aplicaţii ale regresiei

II. Obiectivele secvenței

- Învăţarea tehnicilor necesare pentru calculul coeficienţilor regresiei neliniare
- Utilizarea corectă a regresiilor liniare şi neliniare pentru modelarea fenomenelor economice

III. Cuvinte-cheie

Regresie parabolică, regresie hiperbolică, regresie exponenţială

FONDUL SOCIAL EUROPEAN
Investește în
OAMENI

ROMÂNIA
UNIVERSITATEA BABEŞ-BOLYAI CLUJ-NAPOCA
CENTRUL DE FORMARE CONTINUĂ, ÎNVĂŢĂMÂNT
LA DISTANŢĂ ŞI CU FRECVENŢĂ REDUSĂ

Str Ion I.C.Brătianu, nr.20, Cluj-Napoca
Tel.: (00) 40 - 740 - 077521
Fax: (00) 40 - 364 - 815679

Estimarea coeficienţilor regresiilor neliniare

Citire

Regresia parabolică

În economie sunt numeroase exemplele în care legătura dintre fenomene şi deci variabilele care le cuantifică nu este liniară. Dacă Y reprezintă recolta la hectar dintr-un produs agricol, iar X cantitatea de îngrăşăminte, ne vom da seama chiar şi intuitiv că o anumită creştere a lui X nu provoacă aceeaşi creştere a lui Y pe tot intervalul de variaţie al celor două variabile.

La valori mari ale cantităţii de îngrăşăminte, acestea provoacă saturaţie sau chiar nocivitate, ducând la o stagnare, respectiv diminuare a producţiei. Alte exemple pot fi: legătura dintre vechimea în muncă şi mărimea salariului, dintre cheltuielile cu publicitatea şi volumul vânzărilor, dintre venit şi cererea pentru un produs, etc.

Având modelul parabolic de ecuaţie:

$$\hat{Y} = a_0 + a_1 X + a_2 X^2$$

facem substituţiile:

$$X = X_1$$

$$X^2 = X_2$$

după care ecuaţia devine:

$$\hat{Y} = a_0 + a_1 X_1 + a_2 X_2$$

care reprezintă un model liniar cu doi factori.

Regresia exponenţială

Dacă ecuaţia de regresie are formă exponenţială:

$$\hat{Y} = a \cdot b^X$$

se încearcă aducerea la forma liniară. Mai întâi se logaritmează ecuaţia:

$$\lg \hat{Y} = \lg a + X \cdot \lg b$$

iar apoi se fac substitiţiile:

$$\hat{Z} = \lg \hat{Y}$$

$$a_0 = \lg a$$

$$a_1 = \lg b$$

Rezultă astfel modelul liniar simplu:

$$\hat{Z} = a_0 + a_1 X$$

Regresia hiperbolică

Dacă ecuaţia de regresie are formă hiperbolică:

$$\hat{Y} = a + b \cdot \frac{1}{X}$$

se face substituţia:

$$X_1 = \frac{1}{X}$$

de unde rezultă modelul liniar:

$$\hat{Y} = a + bX_1$$

Alte tipuri de regresie

În practica economică se întâlnesc frecvent şi alte tipuri de funcţii (unele chiar funcţii compuse). Principiul de lucru pentru estimarea coeficienţilor va rămâne însă întotdeauna acelaşi: încercarea de a aduce funcţia la o formă liniară.

Foarte des întâlnite sunt funcţiile de producţie. Forma generală a acestora este:

$$\hat{Y} = a \cdot X_1^{m_1} \cdot X_2^{m_2} \cdot \ldots \cdot X_n^{m_n}$$

Printr-o astfel de funcţie se defineşte o legătură între nivelul producţiei Y şi factorii de care aceasta depinde: productivitatea muncii, calificarea forţei de muncă, gradul de înzestrare cu capital fix, etc. Calculul coeficienţilor se face prin reducere la cazul liniar prin logaritmare:

$$\lg \hat{Y} = \lg a + m_1 \cdot \lg X_1 + \ldots + m_n \cdot \lg X_n$$

Dacă în această nouă ecuaţie facem substituţiile:

$$\hat{Z} = \lg \hat{Y}$$

$$X_i' = \lg X_i, \quad i = \overline{1,n}$$

reducem ecuaţia la una liniară multiplă.

Practică

Estimarea funcţiei de regresie

Un produs a fost lansat simultan pe 13 pieţe. Pe aceste pieţe, produsul a fost propus la preţuri diferite (P), veniturile consumatorilor (V) fiind şi ele diferite. Pentru fiecare piată s-a înregistrat un anumit nivel al cererii (C), rezultatele fiind sintetizate în tabelul următor:

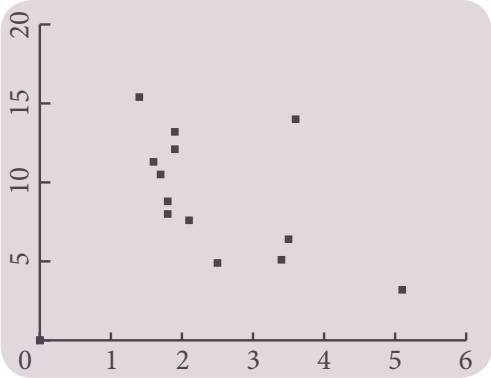
Nr. crt.	Cerere (C)	Preţ (P)	Venit (V)
1	15,4	1,4	620
2	3,2	5,1	530
3	4,9	2,5	490
4	10,5	1,7	800
5	8,0	1,8	630
6	5,1	3,4	410
7	7,6	2,1	670
8	11,3	1,6	920
9	14,0	3,6	990
10	6,4	3,5	320
11	13,2	1,9	520
12	8,8	1,8	700
13	12,1	1,9	730

1) Să se formuleze ipoteze cu privire la forma legăturii dintre cerere (C) şi preţ (P). Pentru formele funcţiilor de regresie reţinute ca fiind posibile, să se calculeze coeficienţii funcţiilor şi reprezentativitatea acestora.

2) Similar pentru legătura dintre cerere şi venit.

3) Să se calculeze coeficienţii funcţiei care modelează legătura liniară multiplă dintre cerere şi factorii săi de influenţă. Analizaţi reprezentativitatea acestei funcţii în raport cu reprezentativitatea funcţiilor de regresie simple. Care va fi valoarea estimată a cererii pe o piaţă unde preţul de vânzare va fi 3,2 iar venitul mediu al consumatorilor de 550 ?

1) Construim norul de puncte în raport cu P şi C:



Reţinem ca forme posibile ale funcţiei de regresie:

- dreapta: $\overline{C}(P) = a_1 + b_1 \cdot P$
- hiperbola: $\overline{C}(P) = a_2 + b_2 \cdot \dfrac{1}{P}$

a) Pentru primul caz, cel al dreptei de regresie, avem:

$$M(P) = \frac{1{,}4 + 5{,}1 + 2{,}5 + \ldots\ldots + 1{,}9}{13} = 2{,}485$$

$$M(P^2) = \frac{1{,}4^2 + 5{,}1^2 + 2{,}5^2 + \ldots\ldots + 1{,}9^2}{13} = 7{,}273$$

$$M(C) = \frac{15{,}4 + 3{,}2 + 4{,}9 + \ldots\ldots + 12{,}1}{13} = 9{,}269$$

$$M(Z) = \frac{0{,}714 + 0{,}196 + 0{,}400 + \ldots\ldots + 0{,}526}{13} = 0{,}462$$

$$M(Z^2) = \frac{0{,}714^2 + 0{,}196^2 + 0{,}400^2 + \ldots\ldots + 0{,}526^2}{13} = 0{,}238$$

$$M(CZ) = \frac{15{,}4 \cdot 0{,}714 + 3{,}2 \cdot 0{,}196 + \ldots\ldots + 12{,}1 \cdot 0{,}526}{13} = 4{,}639$$

$$b_2 = \frac{M(CZ) - M(C)M(Z)}{M(Z^2) - [M(Z)]^2} = \frac{4{,}639 - 9{,}269 \cdot 0{,}462}{0{,}238 - 0{,}462^2} = 14{,}527$$

$$a_2 = M(C) - b_2 \cdot M(Z) = 9{,}269 - 14{,}527 \cdot 0{,}462 = 2{,}56$$

Dreapta de regresie se scrie:

$$\hat{C}(P) = 14{,}3 - 2{,}025 \cdot P$$

Calculăm reprezentativitatea dreptei de regresie:

$$\sigma_P^2 = \frac{(1{,}4 - 2{,}458)^2 + (5{,}1 - 2{,}458)^2 + \ldots\ldots + (1{,}9 - 2{,}458)^2}{13} = 1{,}1$$

$$\sigma_P = 1{,}049$$

$$\sigma_C^2 = \frac{(15{,}4 - 9{,}269)^2 + (3{,}2 - 9{,}269)^2 + \ldots\ldots + (4{,}9 - 9{,}269)^2}{13} = 13{,}479$$

$$\sigma_C = 3{,}671$$

$$R_{C,P} = \frac{M(CP) - M(C)M(P)}{\sigma_C \cdot \sigma_P} = \frac{20{,}81 - 9{,}269 \cdot 2{,}485}{3{,}671 \cdot 1{,}049} = -0{,}58$$

Drepta de regresie are o reprezentativitate medie.

b) Pentru al doilea caz, cel al hiperbolei de regresie, facem schimbarea de variabilă: $Z = \frac{1}{P}$

Datele cu privire la cerere şi noua variabilă Z:

Nr.	C	Z		Nr.	C	Z		Nr.	C	Z
1	15,4	0,714		6	5,1	0,294		11	13,2	0,526
2	3,2	0,196		7	7,6	0,476		12	8,8	0,556
3	4,9	0,400		8	11,3	0,625		13	12,1	0,526
4	10,5	0,588		9	14,0	0,278				
5	8,0	0,556		10	6,4	0,286				

$$M(Z) = \frac{0{,}714 + 0{,}196 + 0{,}400 + \ldots.. + 0{,}526}{13} = 0{,}462$$

$$M(Z^2) = \frac{0{,}714^2 + 0{,}196^2 + 0{,}400^2 + \ldots.. + 0{,}526^2}{13} = 0{,}238$$

$$M(CZ) = \frac{15{,}4 \cdot 0{,}714 + 3{,}2 \cdot 0{,}196 + \ldots.. + 12{,}1 \cdot 0{,}526}{13} = 4{,}639$$

$$b_2 = \frac{M(CZ) - M(C)M(Z)}{M(Z^2) - [M(Z)]^2} = \frac{4{,}639 - 9{,}269 \cdot 0{,}462}{0{,}238 - 0{,}462^2} = 14{,}527$$

$$a_2 = M(C) - b_2 \cdot M(Z) = 9{,}269 - 14{,}527 \cdot 0{,}462 = 2{,}56$$

Hiperbola de regresie se scrie:

$$\hat{C}(P) = 2{,}56 + 14{,}527 \cdot \frac{1}{P}$$

Calculăm reprezentativitatea hiperbolei de regresie:

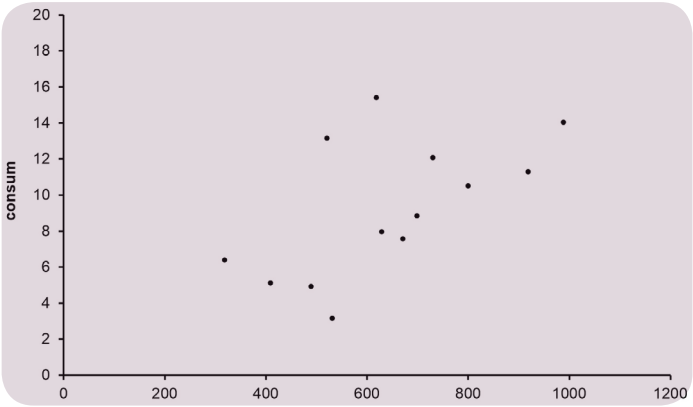
$$\sigma_Z = 0{,}1519$$

$$\sigma_C = 3{,}671$$

$$R_{C,Z} = \frac{M(CZ) - M(C)M(Z)}{\sigma_C \cdot \sigma_Z} = \frac{4{,}639 - 9{,}269 \cdot 0{,}462}{3{,}671 \cdot 0{,}1519} = 0{,}64$$

Drepta de regresie are o reprezentativitate medie, dar este mai reprezentativă decât dreapta.

2) Construim norul de puncte în raport cu V şi C:



Reţinem ca formă posibilă a funcţiei de regresie:

- dreapta: $\hat{C}(V) = a_3 + b_3 \cdot V$

$$M(V) = \frac{620 + 530 + ... + 730}{13} = 640{,}8$$

$$M(V^2) = \frac{620^2 + 530^2 + ... + 730^2}{13} = 444731$$

$$M(C) = 9{,}269$$

$$M(CV) = \frac{15{,}4 \cdot 620 + 3{,}2 \cdot 530 + ... + 12{,}1 \cdot 730}{13} = 6340{,}7$$

$$b_3 = \frac{M(CV) - M(C)M(V)}{M(V^2) - [M(V)]^2} = \frac{6340{,}7 - 9{,}269 \cdot 640{,}8}{444731 - 640{,}8^2} = 0{,}01176$$

$$a_3 = M(C) - b_3 \cdot M(V) = 9{,}269 - 0{,}01176 \cdot 640{,}8 = 1{,}73$$

Dreapta de regresie se scrie:

$$\overline{C}(V) = 1{,}73 + 0{,}01176 \cdot V$$

Calculăm reprezentativitatea dreptei de regresie:

$$\sigma_V^2 = 34146$$

$$\sigma_V = 184{,}8$$

$$\sigma_C = 3{,}671$$

$$R_{C,V} = \frac{M(CV) - M(C)M(V)}{\sigma_C \cdot \sigma_V} = \frac{6340{,}7 - 9{,}269 \cdot 640{,}8}{3{,}671 \cdot 184{,}8} = 0{,}59$$

Drepta de regresie are o reprezentativitate medie.

3) Pentru funcţia de regresie liniară multiplă $C = a_4 + b_4 \cdot P + c_4 \cdot V + \varepsilon$, estimăm coeficienţii prin metoda celor mai mici pătrate:

$$a = (X'X)^{-1} X'Y$$

În cazul aplicaţiei noastre avem:

Nr. crt.	Cerere (C)	Preţ (P)	Venit (V)
1	15,4	1,4	620
2	3,2	5,1	530
3	4,9	2,5	490
4	10,5	1,7	800
5	8,0	1,8	630
6	5,1	3,4	410
7	7,6	2,1	670
8	11,3	1,6	920
9	14,0	3,6	990
10	6,4	3,5	320
11	13,2	1,9	520
12	8,8	1,8	700
13	12,1	1,9	730

$$Y = \begin{pmatrix} c_1 \\ c_2 \\ \dots \\ c_{13} \end{pmatrix} = \begin{pmatrix} 15{,}4 \\ 3{,}2 \\ \dots \\ 12{,}1 \end{pmatrix}$$

$$a = \left[\begin{pmatrix} 1 & 1 & \dots & 1 \\ 1{,}4 & 5{,}1 & \dots & 1{,}9 \\ 620 & 530 & \dots & 730 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1{,}4 & 620 \\ 1 & 5{,}1 & 530 \\ \dots & \dots & \dots \\ 1 & 1{,}9 & 730 \end{pmatrix} \right]^{-1} \cdot \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1{,}4 & 5{,}1 & \dots & 1{,}9 \\ 620 & 530 & \dots & 730 \end{pmatrix} \cdot \begin{pmatrix} 15{,}4 \\ 3{,}2 \\ \dots \\ 12{,}1 \end{pmatrix}$$

$$a = \begin{pmatrix} 7{,}2733 \\ -1{,}5116 \\ 0{,}00898 \end{pmatrix}$$

Coeficienţii sunt deci:

$$a_4 = 7{,}2733$$
$$b_4 = -1{,}5116$$
$$c_4 = 0{,}00898$$

Putem scrie deci funcţia de regresie liniară multiplă:

$$C = 7{,}2733 - 1{,}5116 \cdot P + 0{,}00898 \cdot V + \varepsilon$$

sau

$$\hat{C}(P,V) = 7{,}2733 - 1{,}5116 \cdot P + 0{,}00898 \cdot V$$

Apreciem reprezentativitatea funcţiei prin raportul de determinare R2 şi respectiv coeficientul de corelaţie liniară multiplă R.

Nr. crt.	Cerere (C)	Preţ (P)	Venit (V)	$\hat{C} = a_4 + b_4 \cdot P + c_4 \cdot V$	$\varepsilon = C - \hat{C}$
1	15,4	1,4	620	10,72	4,68
2	3,2	5,1	530	4,32	-1,12
3	4,9	2,5	490	7,89	-2,99
4	10,5	1,7	800	11,89	-1,39
5	8,0	1,8	630	10,21	-2,21
6	5,1	3,4	410	5,82	-0,72
7	7,6	2,1	670	10,12	-2,52
8	11,3	1,6	920	13,12	-1,82
9	14,0	3,6	990	10,72	3,28
10	6,4	3,5	320	4,86	1,54
11	13,2	1,9	520	9,07	4,13
12	8,8	1,8	700	10,84	-2,04
13	12,1	1,9	730	10,96	1,14

Calculăm raportul de determinare R2:

$$R^2 = \frac{V_{\exp}}{V_{tot}} = 1 - \frac{V_{rez}}{V_{tot}} = 1 - \frac{\sigma_\varepsilon^2}{\sigma_C^2}$$

$$R^2 = 1 - \frac{6{,}519}{13{,}479} = 0{,}516$$

$$R_{CPV} = 0{,}719$$

Funcţia liniară multiplă este foarte reprezentativă pentru modelarea legăturii dintre consum (C) şi factorii săi de influenţă: preţul produsului (P) şi venitul consumatorilor. Observăm că funcţia multiplă are un coeficient de corelaţie mai mare decât ceilalţi coeficienţi, ai funcţiilor de regresie cu câte un singur factor. Se confirmă astfel teoria conform căreia o variabilă este cu atât mai bine explicată, cu cât numărul de variabile explicative este mai mare.

$$\hat{C}(P,V) = 7{,}2733 - 1{,}5116 \cdot P + 0{,}00898 \cdot V$$

$$\hat{C}(3{,}2\,;550) = 7{,}2733 - 1{,}5116 \cdot 3{,}2 + 0{,}00898 \cdot 550 = 7{,}38$$

Pe o piaţă unde produsul se va vinde la un preţ de 3,2 iar venitul mediu al consumatorilor este 550, ne aşteptăm la o valoare cea mai probabilă a consumului de 7,38.

Practică

Din diverse surse de informaţii de statistică internaţională, cunoaştem cu privire la 16 state ale lumii valorile variabilelor X = P.I.B./locuitor (mii $) şi Y = cheltuielile pentru activităţi de turism pe locuitor (mii $) :

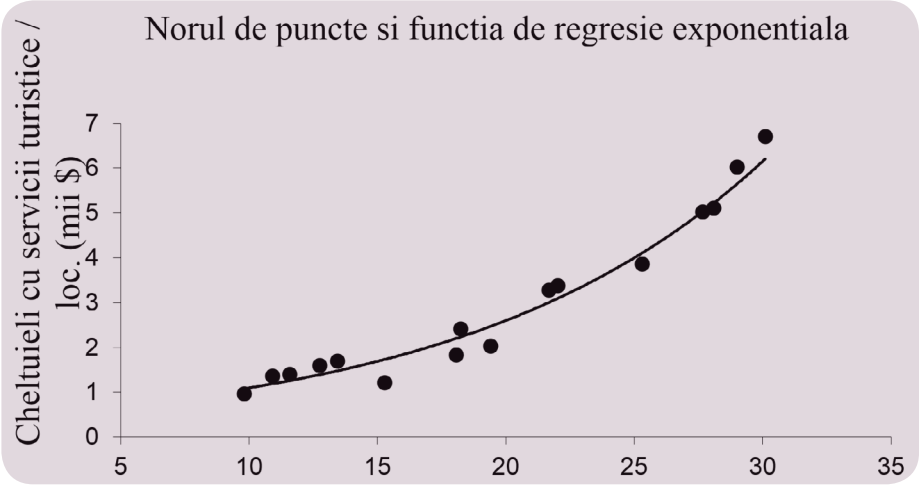
Nr. crt.	X (mii $)	Y (mii $)
1	9.8	0.97
2	15.3	1.23
3	12.8	1.59
4	30.1	6.69
5	22.0	3.4
6	18.1	1.82
7	13.4	1.71
8	28.1	5.11
9	29.0	6.02
10	11.6	1.42
11	25.3	3.84
12	10.9	1.36
13	19.4	2.01
14	21.7	3.29
15	18.2	2.42
16	27.7	5.03

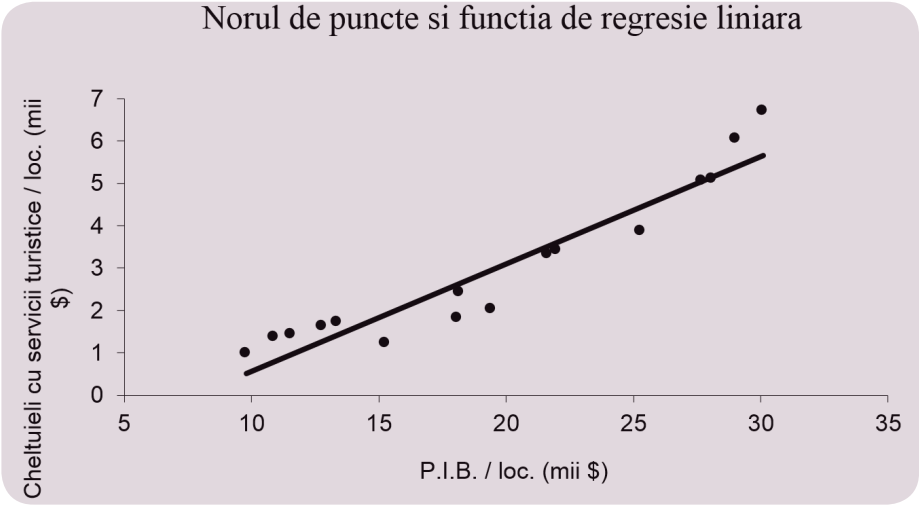
1) Folosind un grafic adecvat şi elemente din teoria economică, să se formuleze ipoteze cu privire la forma posibilă a funcţiei de regresie.

2) Pentru una din formele funcţionale selectate, să se estimeze parametrii şi să se scrie funcţia.

3) Decideţi care funcţie este mai bună, atât pe baza coeficienţilor de corelaţie, cât şi pe baza valorilor previzionate pe baza funcţiilor.

1) Construim norul de puncte în raport cu cele două variabile:





La o primă vedere, ambele funcţii par adaptate pentru modelarea legăturii, punctele norului nefiind prea îndepartate de ele. Dacă judecăm puţin din punct de vedere economic, cererea de servicii turistice este crescătoare în raport cu P.I.B./loc., ceea ce este valabil pentru ambele funcţii. Trebuie să ţinem cont însă că serviciile turistice sunt într-o măsură importantă un bun de lux, deci cu o elasticitate supraunitară a cererii în raport cu venitul. Mai potrivită pare deci funcţia exponenţială.

2) Funcţia exponenţială dintre cele două variabile, sub forma ei deterministă se scrie:

$$Y = a \cdot b^X$$

care prin logaritmare devine:

$$\ln Y = \ln a + \ln b \cdot X$$

iar sub forma aleatoare:

$$\ln Y = \ln a + \ln b \cdot X + \varepsilon$$

Facem următoarele schimbări de variabilă:

$$\ln Y = Z$$

$$\ln a = A$$

$$\ln b = B$$

Funcţia devine :

$$Z = A + B \cdot X + \varepsilon$$

care este o funcţie liniară. Datele numerice pentru estimarea parametrilor le prezentăm în tabelul următor:

Nr. crt.	X (mii $)	Z=lnY
1	9.8	-0.030
2	15.3	0.207
3	12.8	0.464
4	30.1	1.901
5	22.0	1.224
6	18.1	0.599
7	13.4	0.536
8	28.1	1.631
9	29.0	1.795
10	11.6	0.351
11	25.3	1.345
12	10.9	0.307
13	19.4	0.698
14	21.7	1.191
15	18.2	0.884
16	27.7	1.615

$$\begin{cases} A + B \cdot M(X) = M(Z) \\ A \cdot M(X) + B \cdot M(X^2) = M(XZ) \end{cases}$$

$$M(X) = 19{,}59$$

$$M(X^2) = 428{,}81$$

$$M(Z) = 0{,}9199$$

$$M(XZ) = 21{,}91$$

$$B = \frac{M(XZ) - M(X) \cdot M(Z)}{M(X^2) - M(X)^2} = \frac{21{,}91 - 19{,}59 \cdot 0{,}9199}{428{,}81 - 19{,}59^2} = 0{,}08634$$

$$A = M(Z) - B \cdot M(X) = 0{,}9199 - 0{,}08634 \cdot 19{,}59 = -0{,}7715$$

De unde:

$$b = e^{0{,}08634} = 1{,}0902$$

$$a = e^{-0{,}7715} = 0{,}4623$$

Putem scrie deci funcţia de regresie:

$$Y = 0{,}4623 \cdot 1{,}0902^X + \varepsilon$$

3) Calculăm coeficienţii de corelaţie pentru funcţia exponenţială, respectiv cea liniară. Cum şi cea exponenţială a fost liniarizată, avem:

$$R_{ZX} = \frac{M(XZ) - M(X) \cdot M(Z)}{\sigma_X \cdot \sigma_Z} = 0{,}97$$

$$R_{YX} = \frac{M(XY) - M(X) \cdot M(Y)}{\sigma_X \cdot \sigma_Y} = 0{,}95$$

Funcţia exponenţială este deci mai reprezentativă.

Pentru a compara valorile prezise prin cele două funcţii, trebuie să calculăm şi parametrii funcţiei liniare. Fără a mai detalia calculele, scriem funcţia:

$$Y = -2{,}328 + 0{,}2723 \cdot X + \varepsilon$$

Pentru o valoare a venitului de 7000 $ (X=7) obţinem:

- pentru funcţia exponenţială $\overline{Y}(7) = 0{,}4623 \cdot 1{,}0902^{7} = 0{,}85$
- pentru funcţia liniară $\overline{Y}(7) = -2{,}328 + 0{,}2723 \cdot 7 = -0{,}42$ rezultat evident absurd

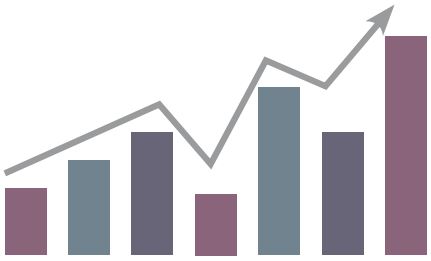
Ca urmare, şi valorile prezise de cele două funcţii o indică pe cea exponenţială ca fiind cea adecvată modelării legăturii dintre PIB/loc şi cheltuielile cu servicii turistice.

Investește în oameni!
Proiect cofinanțat din Fondul Social European prin
Programul Operațional Sectorial Dezvoltarea Resurselor Umane 2007 – 2013
Axa prioritară 1 „Educația și formarea profesională în sprijinul creșterii economice și dezvoltării societății bazate pe cunoaștere”
Domeniul major de intervenție 1.2 „Calitate în învățământul superior”
Titlul proiectului „e-Learning eficient, Individualizat și Adaptiv pentru învățământ la distanță (e-LIADA)”
Contract POSDRU/156/1.2/G/133681

Program:
Statistică descriptivă

Modul: Sistematizarea şi prezentarea datelor statistice
Tema: Indici statistici

(60’ teorie, 30’ practică)



I. Cuprinsul secvenței – Înțelesuri majore ale comunicării

1. Definirea indicilor statistici
2. Tipologia indicilor statistici
3. Mod de calcul şi aplicaţii

II. Obiectivele secvenței

- Însuşirea corectă a noţiunilor de variaţie integrală şi factorială
- Calculul şi utilizarea corectă a indicilor variaţiei factoriale

III. Cuvinte-cheie

Indice al variaţiei integrale, indice al variaţiei factoriale, indice Laspeyres, indice Paasche, indice Fisher

FONDUL SOCIAL EUROPEAN
Investește în OAMENI

ROMÂNIA
UNIVERSITATEA BABEŞ-BOLYAI CLUJ-NAPOCA
CENTRUL DE FORMARE CONTINUĂ, ÎNVĂȚĂMÂNT
LA DISTANȚĂ ȘI CU FRECVENȚĂ REDUSĂ

Str Ion I.C.Brătianu, nr.20, Cluj-Napoca
Tel.: (00) 40 - 740 - 077521
Fax: (00) 40 - 364 - 815679

Aspecte generale asupra indicilor statistici

Citire

Pentru studierea fenomenelor economice şi sociale este adesea nevoie să descriem şi să studiem variaţia unor mărimi. Indicele statistic este o funcţie ale cărei valori numerice exprimă variaţia relativă totală sau parţială a unei mărimi între două situaţii (două perioade, două spaţii diferite, două categorii). Fie Z o mărime a cărei variaţie constituie obiectivul de studia şi λ un parametru de care depinde mărimea Z, respectiv Z = Z(λ). Relativ la parametrul λ, acesta poate exprima timpul, spaţiul sau anumite categorii economice sau sociale.

Considerând două stări oarecare ale parametrului λ, respectiv λ = j cu Z(λ) = Z(j) şi λ = k cu Z(λ) = Z(k), atunci, în sens clasic, indicele mărimii Z este:

$$I_Y^{k/j} = \frac{Z(k)}{Z(j)}$$

După numărul de unităţi la care se referă mărimea Z, indicii sunt:

- indici elementari sau individuali, ce măsoară variaţia unei mărimi relativ la o singură unitate statistică a populaţiei. De exemplu, pentru mărimea Z = x.y, avem:

$$I_x^{k/j} = \frac{x(k)}{x(j)}, \quad I_y^{k/j} = \frac{y(k)}{y(j)} \quad \text{şi} \quad \frac{Z(k)}{Z(j)} = \frac{x(k)\cdot y(k)}{x(j)\cdot y(j)}$$

- indici sintetici sau de grup, ce măsoară variaţia unei mărimi relativ la două sau mai multe unităţi ale populaţiei.

De exemplu $Z = \sum_{i=1}^{m} x_i y_j$, unde „i" reprezintă unitatea statistică, x şi z fiind factorii ce se măsoară la nivelul tuturor celor „m" unităţi. Indicele sintetic este:

$$I_Y^{k/j} = \frac{Z(k)}{Z(j)} = \frac{\sum_i x_i(k) y_i(k)}{\sum_i x_i(j) y_i(j)}$$

Frecvent, în domeniul economic, prezintă interes studiul unei mărimi de tip valoare , unde pi sunt preţuri iar qi sunt cantităţi.

După natura parametrului λ, indicii sunt:

- indici ai dinamicii, unde λ exprimă timpul;
- indici teritoriali, unde λ exprimă spaţiul;
- indici ai unor categorii, unde λ exprimă o anumită categorie economică sau socială.

O abordare sistemică a indicilor unei mărimi îi împarte în:

- indici ai variaţiei integrale;
- indici ai variaţiilor factoriale, denumiţi şi indici factoriali.

Indicele variaţiei integrale

Definiţie

Fie Z = f(x1, x2, ..., xm) modelul ce exprimă dependenţa mărimii Z de factorii x1, x2 , ..., xm. Variaţia mărimii Z rezultată prin acţiunea tuturor factorilor de influenţă, se numeşte variaţie integrală. Indicele variaţiei integrale aferent mărimii Z este:

$$I_Z^{k/j} = \frac{Z(k)}{Z(j)}$$

Proprietăţi ale indicelui variaţiei integrale:

i) Identitatea (reflexivitatea)

$$I_z^{k/k} = 1$$

respectiv:

$$I_z^{k/k} = \frac{Z(k)}{Z(k)} = 1$$

ii) Reversibilitatea (simetria)

$$I_Z^{k/j} \cdot I_Z^{j/k} = 1 \quad \text{sau} \quad I_Z^{k/j} = \frac{1}{I_Z^{j/k}}$$

adică:

$$I_Z^{k/j} = \frac{Z(k)}{Z(j)} = \frac{1}{\frac{Z(j)}{Z(k)}} = \frac{1}{I_Z^{j/k}}$$

iii) Circularitatea (tranzactivitatea). Fie j, h şi k trei stări a parametrului λ, h fiind starea intermediară atunci:

$$I_z^{h/j} \cdot I_z^{k/h} = \frac{Z(h)}{Z(j)} \cdot \frac{Z(k)}{Z(h)} = \frac{Z(k)}{Z(j)} = I_z^{k/j}$$

deoarece

$$I_z^{h/j} \cdot I_z^{k/h} = I_z^{k/j}$$

Indicii variaţiei factoriale

Definiție

Indicii exprimând cote părţi din variaţia mărimii Y determinate, respectiv de factorii x1, x2, ..., xm constituie sistemul de indici factoriali ai mărimii Y. Lipsa unei definiţii efective a condus cercetarea spre găsirea unor algoritmi de calcul ai indicilor factoriali. Cei mai cunoscuţi algoritmi au ca trăsătură comună menţinerea la un nivel constant a factorilor nestudiaţi, fiind numiţi şi sisteme de ponderare.

Principalele tipuri de indici factoriali (sisteme de ponderare) utilizaţi în practică sunt:

- indicii de tip Laspeyres;
- indicii de tip Paasche;
- indicii de tip Fisher.

Se cunosc o serie de alte expresii de calcul a indicilor factoriali, destinate în principal pentru calculul indicelui preţurilor.

Indicii factoriali de tip Laspeyres

Definiție

Cunoscut iniţial sub denumirea de indicele preţurilor, indicele de tip Laspeyres a fost definit având în vedere o mărime exprimând valoarea:

$$V = \sum_{i=1}^{r} p_i q_i$$

unde pi sunt preţuri iar qi cantităţi. Indicele factorial al preţurilor, de tip Laspeyres se determină conform relaţiei:

$$I_{V/p}^{k/j}(\cdot L) = \frac{\sum_{j=1}^{r} p_i(k) q_i(j)}{\sum_{i=1}^{r} p_i(j) q_i(j)}$$

unde:

pi(j) – preţul din perioada de bază;

pi(k) – preţurile din perioada curentă;

qi(j) – cantităţile din perioada de bază;

qi(k) – cantităţile din perioada curentă.

Asemănător se poate scrie expresia indicelui volumului fizic (a cantităţii):

$$I_{V/q}^{k/j}(L\cdot) = \frac{\sum_{j=1}^{r} p_i(j)q_i(k)}{\sum_{i=1}^{r} p_i(j)q_i(j)}$$

Având în vedere expresiile celor doi indici, se deduce o regulă generală pentru elaborarea indicilor factoriali de tip „Laspeyres": partea din variaţia unei mărimi determinată de un factor se obţine menţinând ceilalţi factori la nivelul bază de comparaţie. Astfel, în cazul mărimii Z depinzând de m factori, adică Z = f(X1, X2, ..., Xm), avem:

$$I_{Z/x_1}^{k/j}(\cdot L...L...L) = \frac{f(x_1(k),x_2(j),...,x_i(j),...,x_m(j))}{f(x_1(j),x_2(j),...,x_i(j),...,x_m(j))}$$

$$I_{Z/x_i}^{k/j}(L\,L...\cdot...L) = \frac{f(x_1(j),x_2(j),...,x_i(k),...,x_m(j))}{f(x_1(j),x_2(j),...,x_i(j),...,x_m(j))}$$

$$I_{Z/x_m}^{k/j}(L\,L...L...\cdot) = \frac{f(x_1(j),x_2(j),...,x_i(j),...,x_m(k))}{f(x_1(j),x_2(j),...,x_i(j),...,x_m(j))}$$

Practică

Calculul indicilor de tip Laspeyres

Relativ la sortimentele importate, dintr-o anumită grupă de mărfuri, dispunem de următoarele informaţii:

Se cere:

- a) indicele valorii importului pentru această grupă de mărfuri;
- b) indicele preţului de import respectiv indicele cantităţii importate (a volumului importului), ambii de tip Laspeyres.

Sortiment	an j		an k	
	Cantitate importată	Preţ unitar de import	Cantitate importată	Preţ unitar de import
S1	18	10	15	18
S2	15	6	25	12
S3	25	14	30	20

a) Valoarea importului, din această grupă de mărfuri (formată din trei sortimente), se determină astfel:

unde: $V = \sum_{i=1}^{3} p_i q_i$

pi – preţul unitar de import al sortimentului i;

qi – cantitatea importată din sortimentul i.

Indicele exprimând variaţia integrală a valorii importului este:

$$I_V^{k/j} = \frac{V(j)}{V(k)} = \frac{15\cdot 18 + 25\cdot 12 + 30\cdot 20}{18\cdot 10 + 15\cdot 6 + 25\cdot 14} = \frac{1170}{620} = 1{,}890$$

$I_V^{k/j} = 1.890$ indică o creştere a valorii importului, în anul k faţă de anul j, de 1.89 ori.

b) Indicii factoriali, de tip Laspeyres

$$I_{V/p}^{k/j}(L) = \frac{\sum_{i=1}^{3} p_i(k)q_i(j)}{\sum_{i=1}^{3} p_i(j)q_i(j)} = \frac{18\cdot 18 + 12\cdot 15 + 20\cdot 25}{18\cdot 10 + 15\cdot 6 + 25\cdot 14} = \frac{1004}{620} = 1{,}619$$

$$I_{V/q}^{k/j}(L) = \frac{\sum_{i=1}^{3} p_i(j)q_i(k)}{\sum_{i=1}^{3} p_i(j)q_i(j)} = \frac{10\cdot 15 + 6\cdot 25 + 14\cdot 30}{18\cdot 10 + 15\cdot 6 + 25\cdot 14} = \frac{720}{620} = 1{,}162$$

Urmare a creşterii preţurilor unitare, valoarea importului a crescut, în anul k faţă de anul j, de 1,619 ori. Privind acest indice factorial ca un indice sintetic al preţurilor, rezultatul obţinut arată căci per ansamblu celor trei sortimente preţurile unitare au crescut de 1,619 ori. Modificarea cantităţilor importate a determinat o creştere a valorii importului, în anul k faţă de anul j, de 1,162 ori, sau cantităţile importate au crescut per ansamblu de 1,162 ori.

Indicii factoriali de tip Paasche

Definiție

Paasche, spre deosebire de Laspeyres, pentru a defini indicele preţurilor, ponderează preţurile bunurilor cu cantităţile din perioada curentă. Prin urmare, rezultă:

$$I_{V/p}^{k/j}(P) = \frac{\sum_{j=1}^{r} p_i(k)q_i(k)}{\sum_{i=1}^{r} p_i(j)q_i(k)}$$

Asemănător se poate deduce expresia de calcul a indicelui volumului fizic:

$$I_{V/q}^{k/j}(P) = \frac{\sum_{j=1}^{r} p_i(k)q_i(k)}{\sum_{i=1}^{r} p_i(k)q_i(j)}$$

Şi în acest caz se poate deduce o regulă generală pentru elaborarea indicilor factoriali de tip „Paasche": partea din variaţia unei mărimi determinată de un factor se obţine menţinând ceilalţi factori la nivelul curent. Pentru mărimea Z, depinzând cei m factori, indicii factoriali de tip „Paasche" sunt:

$$I_{Z/x_1}^{k/j}(\cdot P...P...P) = \frac{f(X_1(k), X_2(k), ..., X_i(k), ..., X_m(k))}{f(X_1(j), X_2(k), ..., X_i(j), ..., X_m(k))}$$

$$I_{Z/x_i}^{k/j}(PP...\cdot...P) = \frac{f(X_1(k), X_2(k), ..., X_i(k), ..., X_m(k))}{f(X_1(k), X_2(k), ..., X_i(j), ..., X_m(k))}$$

$$I_{Z/x_m}^{k/j}(PP...P...\cdot) = \frac{f(X_1(j), X_2(j), ..., X_i(j), ..., X_m(k))}{f(X_1(k), X_2(k), ..., X_i(k), ..., X_m(j))}$$

Indicii factoriali de tip Fisher

Definiție

Irving Fisher a stabilit o nouă expresie de calcul a indicelui prețurilor pornind de la limitele indicilor Laspeyres respectiv Paasche. Astfel, indicii Fisher se exprimă ca o medie geometrică a indicilor Laspeyres și Paasche:

$$I_{V/p}^{k/j}(F)=\sqrt{I_{V/p}^{k/j}(L)\cdot I_{V/p}^{k/j}(P)}$$

(indicele prețurilor)

$$I_{V/q}^{k/j}(F)=\sqrt{I_{V/q}^{k/j}(L)\cdot I_{V/q}^{k/j}(P)}$$

(indicele volumului fizic)

Expresii de calcul ale indicilor Fisher

Citire

Se poate extinde acest principiu de calcul pentru o mărimi Z depinzând după o lege oarecare de m factori. Astfel, indicii factoriali obținuți în urma aplicării principiului Fisher sunt definiți ca o medie geometrică a tuturor indicilor factoriali de tip Laspeyres și Paasche, relativ la același factor, determinați prin toate succesiunile posibile de analiză a factorilor.

Spre exemplu în cazul în care mărimea Y depinde de doi factori, respectiv Z = f(X1, X2), indicii factoriali se exprimă astfel:

$$I_{z/x_1}^{k/j}(F)=\sqrt{I_{z/x_1}^{k/j}(\cdot L)\cdot I_{z/x_1}^{k/j}(\cdot P)}=\sqrt{\frac{f(X_1(k),X_2(j))}{f(X_1(j),X_2(j))}\cdot\frac{f(X_1(k),X_2(k))}{f(X_1(j),X_2(k))}}$$

$$I_{z/x_2}^{k/j}(F)=\sqrt{I_{z/x_2}^{k/j}(L\cdot)\cdot I_{z/x_2}^{k/j}(P\cdot)}=\sqrt{\frac{f(X_1(j),X_2(k))}{f(X_1(j),X_2(j))}\cdot\frac{f(X_1(k),X_2(k))}{f(X_1(k),X_2(j))}}$$

Dacă mărimea Z depinde de 3 factori, respectiv Z = f(X1, X2, X3), având 6 succesiuni de analiză a factorilor, indicii factoriali au următoarele expresii:

$$I_{z/x_1}^{k/j}(F)=\sqrt[6]{\left[I_{z/x_1}^{k/j}(\cdot LL)\right]^2\cdot I_{z/x_1}^{k/j}(\cdot LP)\cdot I_{z/x_1}^{k/j}(\cdot PL)\cdot\left[I_{z/x_1}^{k/j}(\cdot PP)\right]^2}$$

$$I_{z/x_2}^{k/j}(F)=\sqrt[6]{\left[I_{z/x_2}^{k/j}(L\cdot L)\right]^2\cdot I_{z/x_2}^{k/j}(L\cdot P)\cdot I_{z/x_2}^{k/j}(P\cdot L)\cdot\left[I_{z/x_2}^{k/j}(P\cdot P)\right]^2}$$

$$I_{z/x_3}^{k/j}(F)=\sqrt[6]{\left[I_{z/x_3}^{k/j}(L\cdot L)\right]^2\cdot I_{z/x_3}^{k/j}(LP\cdot)\cdot I_{z/x_3}^{k/j}(PL\cdot)\cdot\left[I_{z/x_1}^{k/j}(PP\cdot)\right]^2}$$

Justificăm în continuare aceste expresii. Dacă mărimea Z depinde de 3 factori se înregistrează 6 succesiuni de studiere a influenței factorilor asupra variației mărimii. O primă succesiune poate fi următoarea (X1, X2, X3), iar indicii factoriali vor înregistra următoarele expresii:

$$I_{z/x_1}^{k/j}(\cdot LL)=\frac{f(X_1(k),X_2(j),X_3(j))}{f(X_1(j),X_2(j),X_3(j))}$$

$$I_{z/x_2}^{k/j}(P\cdot L)=\frac{f(X_1(k),X_2(k),X_3(j))}{f(X_1(k),X_2(j),X_3(j))}$$

$$I_{z/x_3}^{k/j}(PP\cdot)=\frac{f(X_1(k),X_2(k),X_3(k))}{f(X_1(k),X_2(k),X_3(j))}$$

Astfel, factorii a căror influenţă a fost măsurată sunt consideraţi la nivelul lor din perioada curentă, iar cei. În notaţia indicilor au intervenit simbolurile „L”, „P” şi „ · ”, unde „L” semnifică faptul că factorul se menţine la nivelul bază de comparaţie, „P” factorul se menţine la nivelul curent, iar „ · ” acel factor este analizat. Indiferent de ordinea în care sunt analizaţi factorii, poziţia lor în model nu se schimbă. Pentru toate succesiunile vor fi prezentaţi indicii factoriali, fără a fi redate expresiile lor.

Succesiunea	X1	X2	X3
(X1, X3, X2)	$I_{z/x_1}(\cdot LL)$	$I_{z/x_2}(P\cdot P)$	$I_{z/x_2}(L\cdot P)$
(X1, X2 X3)	$I_{z/x_1}(\cdot LL)$	$I_{z/x_2}(P\cdot L)$	$I_{z/x_3}(PP\cdot)$
(X2, X1, X3)	$I_{z/x_1}(\cdot PL)$	$I_{z/x_2}(L\cdot L)$	$I_{z/x_3}(PP\cdot)$
(X2 X3, X1)	$I_{z/x_1}(\cdot PP)$	$I_{z/x_2}(L\cdot L)$	$I_{z/x_3}(LP\cdot)$
(X3, X1, X2)	$I_{z/x_1}(\cdot LP)$	$I_{z/x_2}(P\cdot P)$	$I_{z/x_3}(LL\cdot)$
(X3, X2, X1)	$I_{z/x_1}(\cdot PP)$	$I_{z/x_2}(L\cdot P)$	$I_{z/x_3}(LL\cdot)$

Dacă mărimea Z depinde de m factori, indicii factoriali se determină ca o medie geometrică a m! indici obţinuţi prin cele m! succesiuni.

Practică

Considerăm mărimea Z=C reprezentând cheltuielile privind consumul a două produse. Datele referitoare la factorii săi de influenţă, adică preţul mediu de cumpărare şi cantitatea cumpărată, pentru două luni consecutive sunt prezentate în tabel.

- a) indicele variaţiei integrale a mărimii C;
- b) indicii factoriali ai mărimii C, de tip Laspeyres, Paasche respectiv Fisher.
- c) exprimarea indicilor factoriali sintetici în funcţie de indicii elementari corespunzători.

Produs	Aprilie 2004		Mai 2004	
	Preţ mediu	Cantitate	Preţ mediu	cantitate
Produs A	3400	2 (kg)	3550	3 (kg)
Produs B	890	10 (buc)	1100	8 (buc)

a) Mărimea C se exprimă în funcţie de preţ şi cantitate astfel:

$$C = \sum_{i=1}^{n} p_i q_i$$

pi – preţul mediu unitar al produsului i;

qi – cantitatea cumpărată din produsul i;

În cazul analizat avem n=2.

Notăm cu: 0 - luna aprilie

1 - luna mai.

Indicele exprimând variaţia totală a mărimii C se determină astfel:

$$I_C^{1/0} = \frac{C(1)}{C(0)} = \frac{19450}{15700} = 1{,}2388$$

$$C(1) = \sum_{i=1}^{2} p_i(1) q_i(1) = 3550 \cdot 3 + 1100 \cdot 8 = 19450$$

$$C(0)=\sum_{i=1}^{2} p_i(0)q_i(0)=3400\cdot 2+890\cdot 10=15700$$

$I_C^{1/0}=1{,}2388$ arată o creştere a cheltuielilor cu consumul celor două produse, în luna aprilie faţă de luna mai, de 1,238 ori.

b) Indicii factoriali

b1) Indicii factoriali Laspeyres

$$I_{C/p}^{1/0}(\cdot L)=\frac{\sum_{i=1}^{2} p_i(1)q_i(0)}{\sum_{i=1}^{2} p_i(0)q_i(0)}=\frac{3550\cdot 2+1100\cdot 10}{C(0)}=\frac{18100}{15700}=1{,}1528$$

$$I_{C/q}^{1/0}(L\cdot)=\frac{\sum_{i=1}^{2} p_i(0)q_i(1)}{\sum_{i=1}^{2} p_i(0)q_i(0)}=\frac{3400\cdot 3+890\cdot 8}{C(0)}=\frac{17320}{15700}=1{,}1032$$

- 1,1528 ne arată de câte ori ar fi crescut cheltuielile C, dată s-ar fi modificat numai preţurile, cantităţile rămânând constante la nivelul lunii aprilie.
- 1,1032 arată de câte ori ar fi crescut cheltuielile C dacă s-ar fi modificat numai cantităţile cumpărate, iar preţurile ar fi rămas constante la nivelul lunii aprilie.

$$I_{C/p}^{1/0}(\cdot P)=\frac{\sum_i p_i(1)q_i(1)}{\sum_i p_i(0)q_i(1)}=\frac{C_1}{3400\cdot 3+890\cdot 8}=\frac{19450}{17320}=1{,}12298$$

$$I_{C/q}^{1/0}(P\cdot)=\frac{\sum_i p_i(1)q_i(1)}{\sum_i p_i(1)q_i(1)}=\frac{19450}{18100}=1{,}0746$$

b2) Indicii factoriali Paasche

- 1,12298 ne arată de câte ori ar fi crescut cheltuielile C dacă s-ar fi modificat numai preţurile, cantităţile rămânând constante la nivelul lunii mai.
- 1,0746 ne arată de câte ori ar fi crescut cheltuielile C dacă s-ar fi modificat numai cantităţile cumpărate, preţurile rămânând constante la nivelul lunii mai.

b3) Indicii factoriali Fisher

$$I_{C/p}^{1/0}(F)=\sqrt{I_{C/p}^{1/0}(\cdot L)I_{C/p}^{1/0}(P\cdot)}=\sqrt{1{,}1528\cdot 1{,}12298}=1{,}1378$$

$$I_{C/q}^{1/0}(F)=\sqrt{I_{C/q}^{1/0}(L\cdot)I_{C/q}^{1/0}(P\cdot)}=\sqrt{1{,}1032\cdot 1{,}0746}=1{,}0888$$

1,1378 ne arată de câte ori cresc cheltuielile C pe seama modifvicării preţurilor de cumpărare.

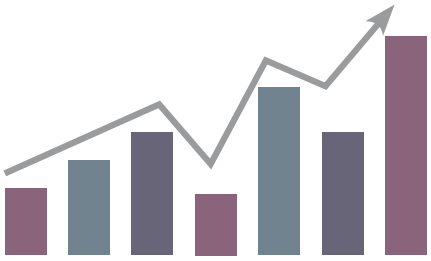
1,0888 ne arată de câte ori cresc cheltuielile C pe seama modificării cantităţilor cumpărate.

Investește în oameni!
Proiect cofinanțat din Fondul Social European prin
Programul Operațional Sectorial Dezvoltarea Resurselor Umane 2007 – 2013
Axa prioritară 1 „Educația și formarea profesională în sprijinul creșterii economice și dezvoltării societății bazate pe cunoaștere”
Domeniul major de intervenție 1.2 „Calitate în învățământul superior”
Titlul proiectului „e-Learning eficient, Individualizat și Adaptiv pentru învățământ la distanță (e-LIADA)”
Contract POSDRU/156/1.2/G/133681

Program:
Statistică descriptivă

Modul: Sistematizarea şi prezentarea datelor statistice
Tema: Serii cronologice

(60’ teorie, 30’ practică)



I. Cuprinsul secvenței – Înțelesuri majore ale comunicării

1. Definirea indicatorilor medii ai seriilor cronologice
2. Mod de calcul al indicatorilor medii
3. Aplicații

II. Obiectivele secvenței

- Însuşirea corectă a noţiunilor de indicatori medii ai seriilor cronologice
- Calculul şi utilizarea corectă a nivelului mediu, diferenţei absolute medii, indicelui mediu şi ritmului mediu

III. Cuvinte-cheie

Nivelul mediu, diferența absolută medie, indicele mediu, ritmul mediu

FONDUL SOCIAL EUROPEAN
Investește în OAMENI

ROMÂNIA
UNIVERSITATEA BABEŞ-BOLYAI CLUJ-NAPOCA
CENTRUL DE FORMARE CONTINUĂ, ÎNVĂȚĂMÂNT LA DISTANȚĂ ȘI CU FRECVENȚĂ REDUSĂ

Str Ion I.C.Brătianu, nr.20, Cluj-Napoca
Tel.: (00) 40 - 740 - 077521
Fax: (00) 40 - 364 - 815679

Definirea seriei cronologice

Definiție

Anticiparea, previziunea evoluţiei viitoare a fenomenelor economice presupune în primul rând cunoaşterea istoriei acestora, punerea în evidenţă a unor legităţi privind comportamentul lor trecut. Baza de date pe care se fundamentează analiza evoluţiei fenomenelor în timp este constituită din serii cronologice.

O serie cronologică este o secvenţă de observaţii asupra unei variabile, ordonate după parametrul timp. Frecvent, măsurătorile asupra variabilei sunt efectuate la intervale egale de timp, seria cronologică fiind prezentată sub forma:

$$Y:\begin{pmatrix} 1 & 2 & \dots & t & \dots & n \\ y_1 & y_2 & \dots & y_t & \dots & y_n \end{pmatrix}$$

Din punct de vedere statistic, în contextul modelării şi previziunii, seria cronologică cu datele observate $y_1, y_2, \dots, y_n$ constituie o realizare a secvenţei de variabile aleatoare $Y_1, Y_2, \dots, Y_n$, adică a unui proces stochastic de tip discret.

Aspecte generale privind seriile cronologice

Citire

În derularea activităţii lor, frecvent agenţii economici sunt puşi în situaţia de a anticipa viitorul, iar apoi de a lua decizii în consecinţă. Oamenii de afaceri sunt nevoiţi să previzioneze anual cifra de afaceri şi alte elemente necesare întocmirii unui plan de afaceri, investitorii sunt interesaţi de profitul viitor degajat de investiţie, respectiv guvernele de previziunea consumului sau a cheltuielilor guvernamentale etc.. Obţinerea rapidă de previziuni utilizând modele cantitative de previziune este la îndemâna analiştilor, urmare şi a softurile de statistică accesibile şi uşor de exploatat.

Anticiparea, previziunea evoluţiei viitoare a fenomenelor economice presupune în primul rând cunoaşterea istoriei acestora, punerea în evidenţă a unor legităţi privind comportamentul lor trecut. Baza de date pe care se fundamentează analiza evoluţiei fenomenelor în timp este constituită din serii cronologice.

În analiza seriilor cronologice este necesar ca lungimea perioadei observate să fie suficient de lungă pentru a face posibilă estimarea unui model adecvat calitativ, care să surprindă mecanismul real de generare al fenomenului, respectiv să permită identificarea unor componente ale evoluţiei pe termen lung. De regulă se impune utilizarea unor serii cronologice cu cel puţin 15 termeni, respectiv pentru serii sezoniere este de dorit ca perioada observată să acopere cel puţin cinci cicluri sezoniere.

În acelaşi timp datele trebuie să rămână comparabile în timp. Condiţiile în care evoluează fenomenul este necesar să rămână în esenţă aceleaşi. Astfel, nu este indicat a se utiliza în elaborarea de modele, serii cronologice ce acoperă perioade de schimbări economice sau politice majore, război, sau alte evenimente excepţionale; în analiza evoluţiei majorităţii indicatorilor economici pentru ţara noastră este indicat ca datele să înceapă după 1989. Înainte de aplicarea tehnicilor specifice de analiză şi previziune, dacă este necesar, unii indicatori vor fi exprimaţi în preţuri comparabile. Când se analizează spre exemplu evoluţia cifrei de afaceri sau a indicatorilor macroeconomici de rezultate şi ne interesează evoluţia datelor neafectate de schimbările de preţ, este indicat a se exprima datele în preţurile anului bază de comparaţie, prin împărţirea acestora la un indice adecvat al preţurilor. De asemenea, creşterea în timp a unor variabile din economie se datorează în principal creşterii populaţiei, astfel că în aceste situaţii este mai util a se analiza evoluţia variabilei per cap de locuitor.

Atunci când cronograma indică prezenţa unor valori aberante, corespunzătoare unor greve, calamităţi naturale sau altor evenimente punctuale, acestea vor fi înlocuite cu valorile medii ce ar fi fost înregistrate în circumstanţe normale. De asemenea, dacă graficul indică puncte de schimbare bruscă a sensului evoluţiei,

de tipul o tendinţă crescătoare ce se schimbă brusc într-una descrescătoare, se vor elabora modele diferite pentru cele doua părţi ale seriei.

Frecvenţa măsurătorilor este condiţionată şi de practică. Spre exemplu, vânzările unui magazin pot fi înregistrate zilnic, profitul poate fi observat lunar şi / sau anual respectiv indicele bursier la încheierea zilei de cotaţie. În general, acolo unde sunt disponibile, poate fi utilă utilizarea unor date cât mai frecvente. Datele anuale nu fac posibilă observarea caracterului sezonier specific anumitor indicatori respectiv modelarea unor dependenţe în care timpul de reacţie al variabilei efect este scurt.

Scopul final al tehnicilor de analiză a seriilor cronologice este previziunea. Informaţiile obţinute respectiv modelele elaborate sunt valabile, strict vorbind, doar pentru perioada observată. Atunci când se fac previziuni pornim de la presupunerea că fenomenul va continua să aibă acelaşi comportament ca şi în trecut. Este important ca analistul să se întrebe în ce măsură această presupunere este realistă, respectiv să ţină seama de aşteptările sale. Se spune pe bună dreptate că „previziunea rămâne în acelaşi timp ştiinţă şi artă". Previziunea fenomenelor economice este o sarcină relativ dificilă, urmare a complexităţii mediului economic. Abordările tradiţionale sunt uneori subiective şi prea simplificatoare, în timp ce metodele moderne sunt mai riguros fundamentate teoretic dar sunt şi mai complexe, necesitând experienţă şi o intervenţie activă a analistului.

Definirea diferenţei absolute medii

Definiţie

În definirea şi calculul diferenţei medii întâlnim mai multe abordări.

a) Diferenţa medie satisface relaţia autoregresivă:

$$y_t = \overline{\Delta}_y + y_{t-1} + \varepsilon_t, \quad t = 2, 3, \ldots, n$$

Metoda celor mai mici pătrate, respectiv condiţia: $\sum_{t=1}^{n} \varepsilon_t^2 \to \min$, conduce la următoarea expresie de calcul a diferenţei medii absolute:

$$\overline{\Delta}_y = \frac{\sum_{t=2}^{n}(y_t - y_{t-1})}{n-1} = \frac{\sum_{t=2}^{n}\Delta_y^{t/t-1}}{n-1}$$

sau echivalent:

$$\overline{\Delta}_y = \frac{y_n - y_1}{n-1}$$

Acest mod de calcul conduce la rezultate satisfăcătoare în condiţiile în care indicatorul evoluează, pe perioada considerată, aproximativ liniar.

b) Diferenţa medie se defineşte prin raportare la tendinţă:

$$y_t = \overline{\Delta}_y + T_{t-1} + \varepsilon_t$$

expresia de calcul pentru obţinută prin metoda celor mai mici pătrate fiind:

$$\overline{\Delta}_y = \frac{\sum_{t=2}^{n}(y_t - T_{t-1})}{n-1}$$

Calculul diferenţei medii are sens doar pe perioade sau subperioade de timp în care indicatorul are o tendinţă monotonă. Aşa cum sugerează şi relaţia de calcul, diferenţă medie $\overline{\Delta}_y$ măsoară cu cât s-a modificat în medie indicatorul Y de la o unitate de timp la alta, de-a lungul perioadei analizate.

Calculul diferenţei medii

Practică

Emisiile de CO2 pe locuitor (în tone metrice) au înregistrat în Camerun următoarea evoluţie:

Anul	2005	2006	2007	2008	2009	2010
Emisii	0,204	0,206	0,305	0,283	0,332	0,351

Diferenţa medie absolută, calculată din formulă:

$$\overline{\Delta}_y = \frac{y_n - y_1}{n-1} = \frac{0{,}351 - 0{,}204}{5} = 0{,}0294 \text{ mii}$$

Emisiile de CO2 pe locuitor au crescut în medie de la un an la altul cu 0,0294 mii tone metrice.

Definirea indicelui mediu

Definiţie

Indicele mediu indică de câte ori ($\overline{I}_y$) sau proporţia în care ($\overline{I}_y$ x100), s-a modificat în medie variabila Y de la o perioadă de timp la alta. Pentru calculul indicelui mediu se pot utiliza metoda autoregresivă sau metoda exponenţială. De asemenea, menţionăm că baza exponentului din tendinţa exponenţială, studiată în paragraful 5.4, poate fi interpretată ca o estimaţie a indicelui mediu.

a)Metoda autoregresivă

Presupunem a termenii seriei se obţin din valoarea precedentă, prin adăugarea unei valori nepredictibile $\varepsilon_t : y_t = a \cdot y_{t-1} + \varepsilon_t, \quad t = 2, 3, ..., T$. Indicele mediu, notat cu , este coeficientul a din ecuaţia anterioară: Utilizând metoda celor mai mici pătrate pentru estimarea coficientului $\overline{I}_y$, din condiţia:

$$y_t = \overline{I}_y \cdot y_{t-1} + \varepsilon_t, \quad t = 2, 3, ..., T$$

unde

$$S(\overline{I}_y) \to \min \quad \Leftrightarrow \quad \frac{\partial S}{\partial \overline{I}_y} = 0$$

se obţine următoarea expresie de calcul a indicelui mediu:

$$S(\overline{I}_y) = \sum_{t=2}^{T} \varepsilon_t^2 = \sum_{t=2}^{T} (y_t - \overline{I}_y \cdot y_{t-1})^2$$

b)Metoda exponenţială

$$\overline{I}_y = \frac{\sum_{t=2}^{T} y_{t-1} \cdot y_t}{\sum_{t=2}^{T} y_{t-1}^2}$$

Presupunem că indicii cu bază în lanţ sunt aproximativ egali:

valoarea comună acestora fiind indicele mediu. Având în vedere că $I_y^{k/k-1} = y_k / y_{k-1}$ se obţin următoarele relaţii:

$$I_y^{2/1} \approx I_y^{3/2} \approx ... \approx I_y^{T/T-1} = \overline{I}_y$$

Din ultima relaţie rezultă următoarea relaţie de calcul a indicelui mediu:

$$y_2 = y_1 \cdot \overline{I}_y$$

$$y_3 = y_2 \cdot \overline{I}_y = y_1 \cdot (\overline{I}_y)^2$$

$$y_T = y_{T-1} \cdot \overline{I}_y = y_1 \cdot (\overline{I}_y)^{T-1}$$

Indicii cu bază în lanţ sunt aproximativ egali atunci când seria are o tendinţă exponenţială, prin ur-

$$\bar{I}_y = \sqrt[T-1]{\frac{y_T}{y_1}}$$

mare această relaţie de calcul este adecvată dacă seria are o tendinţă exponenţială.

Calculul indicelui mediu

Practică

Emisiile de CO2 pe locuitor (în tone metrice) au înregistrat în Camerun următoarea evoluţie:

Anul	2005	2006	2007	2008	2009	2010
Emisii	0,204	0,206	0,305	0,283	0,332	0,351

a)Calculul indicelui mediu prin metoda autoregresivă:

$$\bar{I}_y = \frac{\sum_{t=2005}^{2010} y_{t-1} \cdot y_t}{\sum_{t=2005}^{2010} y_{t-1}^2} = \frac{0{,}204 \times 0{,}206 + \ldots.. + 0{,}332 \times 0{,}351}{0{,}204^2 + \ldots.. + 0{,}332^2} \cong 1{,}093$$

Emisiile de CO2 pe locuitor au crescut în medie de la un an la altul de 1,093 ori.

b)Calculul indicelui mediu prin metoda exponenţială:

$$\bar{I}_y = \sqrt[T-1]{\frac{y_T}{y_1}} = \sqrt[5]{\frac{y_{2010}}{y_{2005}}} = \sqrt[5]{\frac{0{,}351}{0{,}204}} \cong 1{,}115$$

Emisiile de CO2 pe locuitor au crescut în medie de la un an la altul de 1,093 ori.

Definirea ritmului mediu

Definiţie

Ritmul mediu se determină pornind de la indicele mediu:

$$\bar{R}_y = \bar{I}_y - 1$$

sau

$$\bar{R}_y(\%) = (\bar{I}_y - 1) \cdot 100\%$$

Forma procentuală $\bar{R}_y(\%)$, utilizată frecvent, indică cu câte % s-a modificat în medie Y de la o perioadă de timp la alta.

Calculul indicelui mediu

Practică

Emisiile de CO2 pe locuitor (în tone metrice) au înregistrat în Camerun următoarea evoluţie:

Anul	2005	2006	2007	2008	2009	2010
Emisii	0,204	0,206	0,305	0,283	0,332	0,351

a)Calculul indicelui mediu prin metoda autoregresivă:

$$\bar{R}_y(\%) = (\bar{I}_y - 1) \cdot 100\% = (1{,}093 - 1) \cdot 100\% = 9{,}3\%$$

Emisiile de CO2 pe locuitor au crescut în medie de la un an la altul de 1,093 ori.

b)Calculul indicelui mediu prin metoda exponenţială:

$$\bar{R}_y(\%) = (\bar{I}_y - 1) \cdot 100\% = (1{,}115 - 1) \cdot 100\% = \mathbf{1\ \ ,5\%}$$

Emisiile de CO2 pe locuitor au crescut în medie de la un an la altul de 1,093 ori.

Program:
Statistică descriptivă

Conceperea şi pregatirea materialelor video (1,5 ore)

La conceperea materialelor video s-a avut în vedere complementaritatea cu materialul scris. Astfel, înregistrările video conţin mai degrabă:

- Aspecte privind necesitatea şi interesul studiului secvenţei respective de învăţare
- Elemente de originalitate şi inedit pentru încurajarea abordării secvenţei
- Explicaţii colaterale

Acest mod de a concepe materialul filmat are rolul de a nu se suprapune peste materialul scris, care pune accentul pe:

- Abordare axiomatică, riguros-ştiinţifică a disciplinei "Statistică Descriptivă"
- Aplicaţii practice orientate spre înţelegerea materialului teoretic
- Aplicaţii practice orientate spre lărgirea orizontului de aplicaţii posibile ale statisticii

Înregistrarea materialelor video (1,5 ore)

Înregistrarea efectivă pe suport electronic a materialelor video s-a relizat în mod profesional la CF-CIDFR (Centrul pentru Formare Continuă, Învăţământ la Distanţă şi cu Frecvenţă Redusă) al Universităţii Babeş-Bolyai din Cluj Napoca. Pentru unele secvenţe a fost necesară repetarea filmării pentru a obţine un material de calitate cât mai bună.