



SEMINAR BAZELE STATISTICII

SEMINAR 6

Analiza unei serii bidimensionale

1. Prezentarea seriei

► O serie bidimensională prezintă variația unităților unui eșantion după două variabile de grupare în mod simultan:

- variabilele X_i cu valorile $x_i, i = \overline{1, m}$ și Y_j cu valorile $y_j, j = \overline{1, p}$

Efectivele (unitățile) eșantionului care poartă simultan valoarea x_i și valoarea y_j sunt n_{ij} .

Distribuția bivariată este definită de:

$$(x_i, y_j, n_{ij}), i = \overline{1, m}, j = \overline{1, p}$$

$x_i \backslash y_j$	$j=1$	$j=2$...	$j=p$	Total
$i=1$	n_{11}	n_{12}	n_{1p}	$n_{1.}$
$i=2$	n_{21}	n_{22}	n_{2p}	$n_{2.}$
:					:
:	n_{ij}	$n_{i.}$
:					:
$i=m$	n_{m1}	n_{m2}	...	n_{mp}	$n_{m.}$
Total	$n_{.1}$	$n_{.2}$...	$n_{.j}$	$n_{.p}$	$n_{..}$

- 🎯 Variații pe coloană – variații după i , întrucât nivelul j este constant, este același
- 🎯 Variații pe linii – variații după j , întrucât nivelul i este același
- 🎯 Variabila X are **m niveluri de variație**
- 🎯 Variabila Y are **p niveluri de variație**

Tabelul 1.1. Distributia personalului sanitar la nivelul judetelor Regiunii Nord-Est, in anul 2019

	Judete Regiunea NORD-EST (y_i)						
Categorii de personal sanitar (x_i)	Bacau	Botosani	Iasi	Neamt	Suceava	Vaslui	Total
Farmacisti	277	243	1172	315	456	183	2646
Medici (exclusiv stomatologi)	1126	651	4165	876	950	547	8315
Personal sanitar auxiliar	1888	1278	3855	1535	1838	1295	11689
Personal sanitar mediu	3785	2630	7263	3220	4225	2437	23560
Stomatologi	348	164	506	326	453	172	1969
Grand Total	7424	4966	16961	6272	7922	4634	48179

- 🚩 Nivelurile de variație ale variabilei X sunt pe randuri – cele 5 categorii ale personalului sanitar (farmacisti, medici (exclusiv stomatologi), personal sanitar auxiliar, personal sanitar mediu, stomatologi).
- 🚩 Nivelurile de variatie ale variabilei Y sunt pe coloane – cele 6 judete ale Regiunii Nord Est (Bacau, Botosani, Iasi, Neamt, Suceava, Vaslui).
- 🚩 Avem un total de **48179 persoane care activeaza in domeniul sanitar**, care pot fi impartite in 6 grupuri in functie de judetul din care face parte, sau in 5 grupuri, in functie categoria personalului din care face parte.
- 🚩 Numarul de distributii conditionate este: $m+p=5+6=11$ distribuții condiționate.

2. Tipuri de variabile

- o variabilă numerică și o variabilă nenumerică;
- ambele variabile numerice;
- ambele variabile nenumerice.

3. Distribuția după o variabilă cantitativă și o variabilă calitativă

În cadrul unei distribuții bidimensionale se disting:

a). *Două distribuții marginale*

► Distribuția marginală în X : $X : (x_i, n_{i\bullet}), i = 1, \dots, m$

$$n_{i\bullet} = \sum_{j=1}^p n_{ij}$$

- Distribuția marginală în Y :

$$Y : (y_j, n_{\bullet j}), \quad j = 1, \dots, p$$

$$n_{\bullet j} = \sum_{i=1}^m n_{ij}$$

Observație:

- Convenim să notăm variabila numerică cu X și variabila nenumerică cu Y . Prin urmare, există **m** valori ale variabilei numerice și **p** valori ale variabilei nenumerice.

b) Distribuții condiționate (m+p distribuții)

- ▶ Distribuția condiționată a variabilei X în funcție de valorile variabilei Y
 - este definită pentru fiecare valoare y_j

$(X / Y = y_j) : (x_i, n_{ij}), i = 1, \dots, m \quad \text{si} \quad j \text{ valoare fixă}$

- Distribuția condiționată a variabilei Y în funcție de valorile variabilei X
 - este definită pentru fiecare valoare \mathbf{x}_i

$$(Y / X = x_i) : (y_j, n_{ij}), j = 1, \dots, p \quad \text{și} \quad i \text{ valoare fixă}$$

4. Frecvențe absolute

- ▶ *Frecvențe absolute marginale - sunt efectivele grupurilor create de X sau Y*

$$n_{i.} \text{ și } n_{.j}$$

- ▶ *Frecvențe absolute parțiale: n_{ij}*

5. Frecvențe relative

► *Frecvențe relative marginale*

►
$$f_{i\bullet} = \frac{n_{i\bullet}}{n_{\bullet\bullet}}; \quad f_{\bullet j} = \frac{n_{\bullet j}}{n_{\bullet\bullet}}$$

► *Frecvențe relative parțiale:*

$$f_{ij} = \frac{n_{ij}}{n_{\bullet\bullet}}$$

► Frecvențe relative condiționate

$$f_{i/j} = \frac{n_{ij}}{n_{\bullet j}} \quad j \text{ valoare fixa}, i = 1, \dots, m$$

$$f_{j/i} = \frac{n_{ij}}{n_{i\bullet}} \quad i \text{ valoare fixa}, j = 1, \dots, p$$

6. Medii condiționate (pe grupe)

Dacă X este variabila numerică, atunci media variabilei X pe grupe este:

$$\bar{x}_j = \frac{\sum_{i=1}^m x_i \cdot n_{ij}}{n_{\bullet j}}, \text{ cu } n_{\bullet j} = \sum_{i=1}^m n_{ij}, j = \overline{1, p}$$

7. Media pe total:

$$\bar{x} = \frac{\sum_{j=1}^p \bar{x}_j \cdot n_{\bullet j}}{\sum_{j=1}^p n_{\bullet j}}.$$

8. Variația general

- masoara, la nivelul populatiei, diferentele dintre indivizi, negrupati

$$s_X^2 = s_{\bar{x}_j}^2 + \bar{s}^2$$

a) Variante între grupe (varianțe intergrupe)

- masoara, la nivel general, diferentele dintre indivizii din grupe diferite

$$s_{\bar{x}_j}^2 = \frac{\sum_{j=1}^p (\bar{x}_j - \bar{x})^2 \cdot n_{\bullet j}}{\sum_{j=1}^p n_{\bullet j}}$$

b) Media varianțelor de grupă (varianța intra-grupe)

- măsura, la nivel general, diferențele dintre indivizii de același fel (din același grup)

$$\bar{s}^2 = \frac{\sum_j s_j^2 \cdot n_{\bullet j}}{\sum_j n_{\bullet j}}$$

unde s_j^2 sunt varianțe condiționate (varianțe de grupă). Acestea măsoară variația în cadrul unei grupe (intragrupă).

Se calculează astfel:

$$s_j^2 = \frac{\sum_{i=1}^m (x_i - \bar{x}_j)^2 \cdot n_{ij}}{n_{\bullet j}}$$

pentru

$$Y = y_j$$

Măsurarea gradului de influență a factorului de grupare și a factorilor aleatori

□ Plecând de la regula de adunare a varianțelor se pot calcula:

a) Coeficientul influenței factorului de grupare

$$k_1 = \frac{s_{\bar{x}_j}^2}{s_X^2} \cdot 100$$

b) Coeficientul influenței factorilor întâmplători

$$k_2 = \frac{\bar{s}^2}{s_X^2} \cdot 100$$

Se considera distributia unui esantion de 1000 de cititori ai Bibliotecii Centrale Universitare Mihai Eminescu din Iasi dupa numarul de carti citite intr-o luna, X , si categoria de varsta, Y . Datele sunt prezentate in tabelul de mai jos.

Tabelul 2.1. Distributia numarului de cititori in functie de numarul de carti citite intr-o luna si categoria de varsta

$x_i \backslash y_j$ (carti/luna)	Tineri (sub 26 ani)	Adulti (26-55 ani)	Varstnici (peste 55 ani)	TOTAL
1	10	40	170	220
2	40	60	70	170
3	160	120	10	290
4	240	80	0	320
TOTAL	450	300	250	1000

APLICATII

Tabelul bidimensional este de forma:

$x_i \backslash y_j$	y_1	y_2	y_3	TOTAL
x_1	n_{11}	n_{12}	n_{13}	$n_{1.}$
x_2	n_{21}	n_{22}	n_{23}	$n_{2.}$
x_3	n_{31}	n_{32}	n_{33}	$n_{3.}$
x_4	n_{41}	n_{42}	n_{43}	$n_{4.}$
TOTAL	$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{..}$

► Se cere:

1. Puneți în evidență distribuția marginală a variabilei X și distribuția marginală a variabilei Y.
2. Puneți în evidență distribuțiile condiționate. Câte distribuții condiționate sunt?
3. Să se calculeze și să se interpreteze frecvențele relative parțiale.
4. Să se calculeze și să se interpreteze frecvențele relative marginale.
5. Să se calculeze și să se interpreteze frecvențele relative condiționate.
6. Să se calculeze și să se interpreteze mediile și varianțele condiționate (mediile și variantele pe grupe).
7. Să se calculeze și să se interpreteze media pe total.
8. Să se calculeze media varianțelor de grupă (varianța intra-grupe).
9. Să se calculeze varianța între grupe sau varianța mediilor de grupă (varianța intergrupe).
10. Să se calculeze varianța generală.
11. Să se calculeze coeficientul de influență a factorului de grupare și coeficientul de influență a factorilor aleatori/întâmplători.