

APLICAȚII BAZELE STATISTICII – SEMINAR 6

Tabelul 1.1. Distribuția personalului sanitar la nivelul județelor Regiunii Nord-Est, în anul 2019

	Județe Regiunea NORD-EST (y_j)						
Categorii de personal sanitar (x_i)	Bacău	Botoșani	Iasi	Neamț	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 variație ale variabilei Y sunt pe coloane – cele 6 județe ale Regiunii Nord Est (Bacău, Botoșani, Iasi, Neamț, Suceava, Vaslui).
- 🚩 Avem un total de **48179 persoane care activează în domeniul sanitar**, care pot fi împartite în 6 grupuri în funcție de județul din care face parte, sau în 5 grupuri, în funcție categoria personalului din care face parte.
- 🚩 Numarul de distribuții conditionate este: $m+p=5+6=11$ distribuții condiționate.

1. Distribuția unei serii bidimensionale 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:



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

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

b) Distribuții condiționate ($m+p$ distribuții)



- 🚩 Distribuțiile condiționate ale variabilei X în funcție de valorile variabilei Y;
- 🚩 este definită pentru fiecare valoare y_j

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

 Distribuțiile condiționate ale variabilei Y în funcție de valorile variabilei X;
 sunt definite pentru fiecare valoare x_i

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

2. Frecvențe absolute

 Frecvențe absolute marginale ($n_{i\cdot}$ și $n_{\cdot j}$).
 Frecvențele absolute marginale sunt efectivele grupurilor create de X sau Y.

3. Frecvențe relative

 Frecvențe relative marginale

$$f_{i\cdot} = \frac{n_{i\cdot}}{n_{\cdot\cdot}}; f_{\cdot j} = \frac{n_{\cdot j}}{n_{\cdot\cdot}}$$

 Frecvențe relative parțiale: $f_{ij} =$

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

 Frecvențe relative condiționate

$$f_{i|j} = \frac{n_{ij}}{n_{\cdot j}} \quad j \text{ valoare fixă, } i = 1, \dots, m$$

$$f_{j|i} = \frac{n_{ij}}{n_{i\cdot}} \quad i \text{ valoare fixă, } j = 1, \dots, p$$

4. Medii condiționate (medii de grupă)

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

5. Varianțe condiționate (varianțe de grupă)

Măsoară diferențele dintre indivizii care fac parte din același grup (intragrupă).

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


Pentru

$$Y = y_j$$

6. Media pe total


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

7. Media varianțelor de grupă (varianța intra-grupe)

 masoara, la nivel general, diferentele dintre indivizii de acelasi fel (din acelasi grup)


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

8. Varianța între grupe (varianța inter-grupe sau varianța mediilor de grupă)

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

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

9. Varianța generală


 masoara, la nivelul populatiei, diferentele dintre indivizi, negrupati

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

10. Măsurarea gradului de influență a celor două categorii de factori

 Coeficientul influenței factorului de grupare

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

 Coeficientul influenței factorilor întâmplători

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

$$k_1 + k_2 = 100\%$$

Dacă $k_1 > k_2$, atunci factorul de grupare explică mai mult din variația variabilei studiate decât factorii întâmplători.

II. Aplicații

Se considera distribuția unui esantion de 1000 de cititori ai Bibliotecii Centrale Universitare Mihai Eminescu din Iași după numărul de cărți citite într-o lună, X , și categoria de vârstă, Y . Datele sunt prezentate în tabelul de mai jos.

Tabelul 2.1. Distribuția numărului de cititori în funcție de numărul de cărți citite într-o lună și categoria de vârstă

$x_i \backslash y_j$ (cărți/lună)	Tineri (sub 26 ani)	Adulți (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

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 varianțele 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.

Rezolvare:

1. Puneți în evidență distribuția marginală a variabilei X și distribuția marginală a variabilei Y .

$x_i \backslash y_j$ (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

- 📊 O distribuție a unei variabile empirice poate fi pusă în evidență prin prezentarea valorilor variabilei și a frecvențelor lor de apariție.
- 📊 Sunt două distribuții marginale constituite din valorile variabilei și frecvențele marginale.
- 📊 In exemplul nostru, avem două distribuții marginale:
1. **Una în X:** distribuția cititorilor în funcție de numărul de cărți citite
 2. **Una în Y:** distribuția cititorilor în funcție de vârstă.

Tabelul 2.2. Distribuția cititorilor în funcție de numărul de cărți citite

x_i (carti/luna)	n_i
1	220
2	170
3	290
4	320
TOTAL	1000

Tabelul 2.3. Distribuția cititorilor după categoria de vârstă

y_j	n_j
Tineri	450
Adulti	300
Varstnici	250
TOTAL	1000

2. Puneți în evidență distribuțiile condiționate. Câte distribuții condiționate sunt?

Sunt $m + p = 4 + 3 = 7$ distribuții condiționate.

Obs. 4 distribuții condiționate ale lui X în funcție de nivelurile variabilei Y , respectiv 3 distribuții condiționate ale lui Y în funcție de nivelurile variabilei X .

Tabelul 2.4. Distribuția cititorilor in functie de numarul de carti citite, condiționata incat categoria de varsta sa fie “tineri” ($X / Y = y_1$)

x_i (carti/luna)	n_{i1}
1	10
2	40
3	160
4	240
TOTAL	450

Tabelul 2.5. Distribuția cititorilor in functie de numarul de carti citite, condiționata incat categoria de varsta sa fie “adulti” ($X / Y = y_2$)

x_i (carti/luna)	n_{i2}
1	40
2	60
3	120
4	80
TOTAL	300

Tabelul 2.6. Distribuția cititorilor in functie de numarul de carti citite, condiționata incat categoria de varsta sa fie “varstnici” ($X / Y = y_3$)

x_i (carti/luna)	n_{i3}
1	170
2	70
3	10
4	0
TOTAL	250

Tabelul 2.7. Distribuția cititorilor după categoria de varsta, condiționată ca numarul de carti citite sa fie 1/luna ($Y / X = x_1$)

y_j	n_{1j}
Tineri	10
Adulti	40
Varstnici	170
TOTAL	220

Tabelul 2.8. Distribuția cititorilor după categoria de varsta, condiționată ca numărul de carti citite sa fie 2/luna ($Y/X = x_2$)

y_j	n_{2j}
Tineri	40
Adulti	60
Varstnici	70
TOTAL	170

Tabelul 2.9. Distribuția cititorilor după categoria de varsta, condiționată ca numărul de carti citite sa fie 3/luna ($Y/X = x_3$)

y_j	n_{3j}
Tineri	160
Adulti	120
Varstnici	10
TOTAL	290

Tabelul 2.10. Distribuția cititorilor după categoria de varsta, condiționată ca numărul de carti citite sa fie 4/luna ($Y/X = x_4$)

y_j	n_{4j}
Tineri	240
Adulti	80
Varstnici	0
TOTAL	320

3. Să se calculeze și să se interpreteze frecvențele relative parțiale.

$x_i \backslash y_j$ (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

$$f_{ij} = \frac{n_{ij}}{n_{..}} \cdot 100$$

$f_{11} = (10/1000) \cdot 100 = 1 \Rightarrow 1\%$ din totalul cititorilor sunt tineri si citesc 1 carte/luna.

$f_{12} = (40/1000) \cdot 100 = 4 \Rightarrow 4\%$ din totalul cititorilor sunt adulti si citesc 1 carte/luna.

$f_{13} = (170/1000) \cdot 100 = 17 \Rightarrow 17\%$ din totalul cititorilor sunt varstnici si citesc 1 carte/luna.

$f_{32} = (120/1000) \cdot 100 = 12 \Rightarrow 12\%$ din totalul cititorilor sunt adulti si citesc 3 carti/luna.

.....

4. Să se calculeze și să se interpreteze frecvențele relative marginale.

$$f_{i.} = \frac{n_{i.}}{n_{..}} \cdot 100$$

$f_{1.} = (220/1000) \cdot 100 = 22 \Rightarrow 22\%$ din totalul cititorilor citesc 1 carte/luna.

$f_{2.} = (170/1000) \cdot 100 = 17 \Rightarrow 17\%$ din totalul cititorilor citesc 2 carti/luna.

$f_{3.} = (290/1000) \cdot 100 = 29 \Rightarrow 29\%$ din totalul cititorilor citesc 3 carti/luna.

$f_{4.} = (320/1000) \cdot 100 = 32 \Rightarrow 32\%$ din totalul cititorilor citesc 4 carti/luna.

$$f_{.j} = \frac{n_{.j}}{n_{..}} \cdot 100$$

$f_{.1} = (450/1000) \cdot 100 = 45 \Rightarrow 45\%$ din totalul cititorilor sunt tineri.

$f_{.2} = (300/1000) \cdot 100 = 30 \Rightarrow 30\%$ din totalul cititorilor sunt adulti.

$f_{.3} = (250/1000) \cdot 100 = 25 \Rightarrow 25\%$ din totalul cititorilor sunt varstnici.

5. Să se calculeze și să se interpreteze frecvențele relative condiționate.

$$f_{i|j} = \frac{n_{ij}}{n_{.j}} \cdot 100, j \text{ valoare fixă (categoria de varsta)}$$

Tineri

x_i (carti/luna)	n_{i1}	$f_{ij} (\%)$
1	10	$(10/450) \cdot 100 = 2.22\%$
2	40	$(40/450) \cdot 100 = 8.89\%$
3	160	$(160/450) \cdot 100 = 35.56\%$
4	240	$(240/450) \cdot 100 = 53.33\%$
TOTAL	450	100%

Interpretare:

 2.22% din totalul cititorilor tineri citesc 1 carte/luna.




 8.89% din totalul cititorilor tineri citesc 2 carti/luna.

$$f_{j/i} = \frac{n_{ij}}{n_i} \cdot 100, i \text{ valoare fixă (numarul de carti citite)}$$

Carti citite: 1carte/luna

y_j	n_{1j}	$f_{ji} (\%)$
Tineri	10	$(10/220) \cdot 100 = 4.55\%$
Adulti	40	$(40/220) \cdot 100 = 18.18\%$
Varstnici	170	$(170/220) \cdot 100 = 77.27\%$
TOTAL	220	100%

Interpretare:

-  4.55% din totalul cititorilor care citesc 1 carte/luna sunt tineri.
-  18.18% din totalul cititorilor care citesc 1 carte/luna sunt adulti.
-  77.27% din totalul cititorilor care citesc 1 carte/luna sunt sunt varstnici.

6. Să se calculeze și să se interpreteze mediile și varianțele condiționate (mediile și variantele pe grupe).

Mediile condiționate pot fi calculate pentru fiecare din distribuțiile condiționate ale variabilei cantitative, numarul de carti citite.

$$\bar{x}_j = \frac{\sum_{i=1}^m x_i n_{ij}}{\sum_{j=1}^p n_{ij}} = \frac{\sum_{i=1}^m x_i n_{ij}}{n_{1j}}$$

Tabelul 2.11. Distribuția cititorilor dupa numarul de carti citite, condiționata incat categoria de varsta sa fie “tineri” ($X / Y = y_1$)

x_i (carti/luna)	n_{i1}	$x_i n_{i1}$	$x_i - \bar{x}_1$	$(x_i - \bar{x}_1)^2$	$(x_i - \bar{x}_1)^2 n_{i1}$
1	10	10	-2.4	5.76	57.6
2	40	80	-1.4	1.96	78.4
3	160	480	-0.4	0.16	25.6
4	240	960	0.6	0.36	86.4
TOTAL	450	1530			248

$$\bar{x}_1 = \frac{1530}{450} = 3,4 \text{ carti / luna}$$

Interpretare: Numarul mediu de carti citit de tineri este de 3.4 carti/luna.

$$s_1^2 = \frac{248}{450} = 0.55$$

Tabelul 2.12. Distribuția cititorilor după numărul de cărți citite, condiționată întrucât categoria de vârstă să fie “adulti” ($X / Y = y_1$)

x_i (cărți/lună)	n_{i1}	$x_i n_i$	$x_i - \bar{x}_1$	$(x_i - \bar{x}_1)^2$	$(x_i - \bar{x}_1)^2 n_i$
1	40	40	-1.8	3.24	129.6
2	60	120	-0.8	0.64	38.4
3	120	360	0.2	0.04	4.8
4	80	320	1.2	1.44	115.2
TOTAL	300	840			288

$$\bar{x}_1 = \frac{840}{300} = 2.8 \text{ cărți/lună}$$

Interpretare: Numărul mediu de cărți citite de adulți este de 2.8 cărți/lună.

$$s_2^2 = \frac{288}{300} = 0.96$$

Tabelul 2.13. Distribuția cititorilor după numărul de cărți citite, condiționată întrucât categoria de vârstă să fie “varstnici” ($X / Y = y_1$)

x_i (cărți/lună)	n_{i1}	$x_i n_i$	$x_i - \bar{x}_1$	$(x_i - \bar{x}_1)^2$	$(x_i - \bar{x}_1)^2 n_i$
1	170	170	-0.36	0.1296	22.032
2	70	140	0.64	0.4096	28.672
3	10	30	1.64	2.6896	26.896
4	0	0	2.64	6.9696	0
TOTAL	250	340			77.6

$$\bar{x}_1 = \frac{340}{250} = 1.36 \text{ cărți/lună}$$

Interpretare: Numărul mediu de cărți citite de varstnici este de 1.36 cărți/lună.

$$s_3^2 = \frac{77.6}{250} = 0.31$$

Tabelul de sinteza este:

	Tineri	Adulti	Varstnici
Media de grupa	$\bar{x}_1 = 3.4$	$\bar{x}_2 = 2.8$	$\bar{x}_3 = 1.36$
Varianța de grupa	$s_1^2 = 0.55$	$s_2^2 = 0.96$	$s_3^2 = 0.31$
Volumul grupei	n.1 = 450	n.2 = 300	n.3 = 250

7. Să se calculeze și să se interpreteze media pe total.

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

$$\bar{x} = \frac{3.4 \cdot 450 + 2.8 \cdot 300 + 1.36 \cdot 250}{1000} = 2.71 \text{ carti/luna}$$

Interpretare: Numarul mediu de carti citit de cititorii BCU este de 2.71 carti/luna.

8. Să se calculeze media varianțelor de grupă (varianța intra-grupe).

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

Obs. Măsoară, la nivel general, diferențele dintre indivizii de același fel (din aceeași grupă).

$$\bar{s}^2 = \frac{0.55 \cdot 450 + 0.96 \cdot 300 + 0.31 \cdot 250}{1000} = 0.61$$

Interpretare: Valoare de 0.61 este expresia cantitativa a diferentelor de carti citite dintre cititorii din aceeași categorie de varsta.

9. Să se calculeze varianța între grupe sau varianța mediilor de grupă (varianța intergrupe).

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

Obs. Masoara, la nivel general, diferentele dintre indivizii din grupe diferite.

\bar{x}_j	n_j	$\bar{x}_j - \bar{x}$	$(\bar{x}_j - \bar{x})^2$	$(\bar{x}_j - \bar{x})^2 n_{r_j}$
3.4	450	0.69	0.4761	214.245
2.8	300	0.09	0.0081	2.43
1.36	250	-1.35	1.8225	455.625
TOTAL	1000			672.3

$\bar{x} = 2.71$ (am calculat anterior)

$$s_{x_j}^2 = \frac{672.3}{1000} = 0.67$$

Interpretare: Aceasta valoare de 0.67 este expresia numerica a diferentelor de carti citite de catre cititorii din categorii diferite de varsta.

10. Să se calculeze varianța generala.

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

$$s_X^2 = 0.61 + 0.67 = 1.28$$

11. Să se calculeze coeficientul de influență a factorului de grupare și coeficientul de influență a factorilor aleatori/întâmplători.

$$k_1 = \frac{s_{x_j}^2}{s_X^2} \cdot 100 = \frac{0.67}{1.28} \cdot 100 = 52.34\%$$

$$k_2 = \frac{\bar{s}^2}{s_X^2} \cdot 100 = \frac{0.61}{1.28} \cdot 100 = 47.66\%$$

Interpretare: Deoarece coeficientul factorului de grupare k_1 este mai mare decât coeficientul influenței factorilor întâmplători k_2 , factorul de grupare, categoria de varsta are influență asupra numărului de carti citite.

Obs.

Daca $k_1 < k_2$, atunci factorul de grupare nu are o influență semnificativă asupra variației variabilei numerice; în acest caz, varianța din interiorul grupelor este mai mare decât varianța dintre grupe. Există diferențe mai mari de comportament între unitățile statistice din aceeași grupă decât între unitățile statistice din grupuri diferite.