CURSUL V

MODELE STATISTICE ÎN ANALIZA VARIAŢIEI (ANOVA)

5.1. Modelul de analiză unidimensională

În analiza factorială a diverselor fenomene social-economice sunt folosite pe scară largă o serie de modele statistice de analiză variaţională, cu ajutorul cărora pot fi identificate dependenţele dintre variabile şi pot fi cuantificate intensităţile legăturilor cauzale dintre aceste fenomene.

Una dintre modalităţile practice de folosire a acestor modele este analiza varianţei (ANOVA – Analiysis of Variance). Prezentăm în continuare principalele elemente de ordin metodologic ale acestei metode.

În figura 5.1. este prezentată schema generală a elementelor de bază ale analizei varianţei pentru serii unidimensionale.

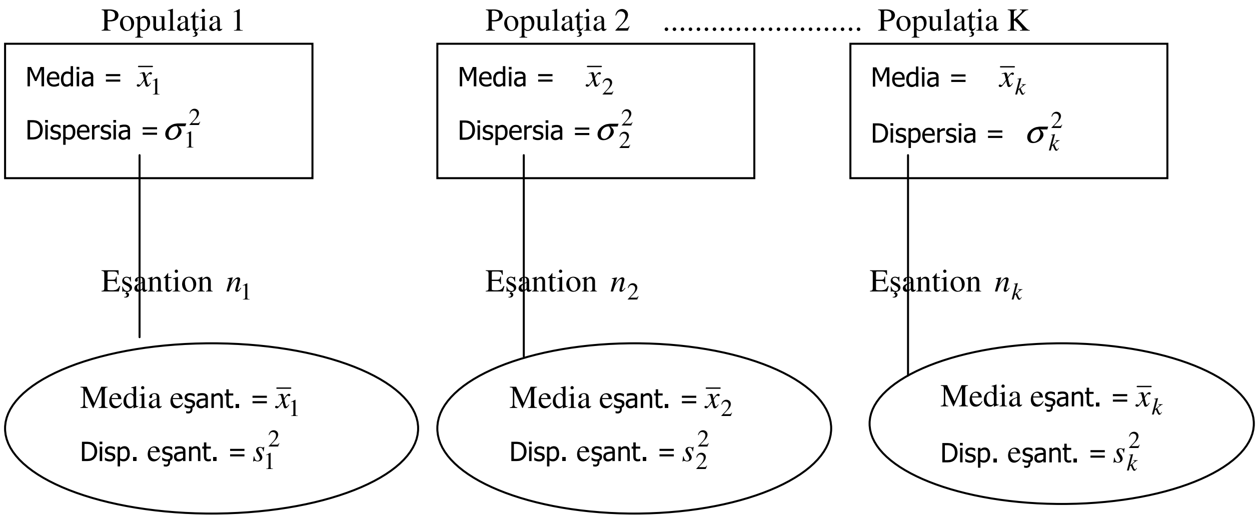

Fig.5.1. Schema pentru eşantioane independente

Din fiecare populaţie de volum N_1, N_2, ……..N_k, se extrag eşantioane de volum n_1, n_2, …..n_k. Valorile caracteristicii X, corespunzătoare fiecărei unităţi din componenţa eşantioanelor, pot fi prezentate sub formă tabelară, astfel încât prelucrarea lor în vederea obţinerii indicatorilor specifici este uşor de făcut, atât manual cât şi cu ajutorul tehnicii moderne de calcul.

Un model este prezentat în tabelul 5.1.

Media caracteristicii X corespunzătoare fiecărui eşantion se calculează cu ajutorul relaţiei:

$$\bar{x}_j = \frac{\sum_{j=1}^{n_j} x_{ij}}{n_j} \qquad (5.1)$$

Tab.5.1. Notaţiile pentru analiza variaţiei unei caracteristici, pe eşantioane independente

Specificaţia	Prelucrarea eşantionului numărul:			
	1	2	………j	………..k
Valorile caracteristicii X	x_{11} x_{21}	x_{12} x_{22}	x_{1j} x_{2j}	x_{1k} x_{2k}

	x_{31}	x_{32}	x_{3j}	x_{3k}
	.	. .	.	.
	.	. .	.	.
	x_{n1}	x_{n2}	x_{nj}	x_{nk}
Volumul eşant.	n_1	n_2	n_j	n_k
Media eşant.	$\bar{x}_1$	$\bar{x}$	$\bar{x}_j$	$\bar{x}_k$

Media generală a caracteristicii corespunzătoare eşantonului general n, ($n = n_1 + n_2 + \ldots..+ n_k$), se determină astfel:

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{ij}}{n} \qquad (5.2)$$

sau, ca o medie aritmetică ponderată a mediilor corespunzătoare fiecărui eşantion, folosind relaţia:

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{k} \bar{x}_j n_j}{n} \qquad (5.3)$$

Unul dintre obiectivele analizei variaţiei caracteristicii X la nivelul acestor eşantioane îl constituie verificarea ipotezei diferenţei nule, potrivit căreia mediile eşantioanelor sunt egale, adică:

$$H_0: \bar{x}_1 = \bar{x}_2 = \ldots\ldots\ldots = \bar{x}_j \ldots\ldots\ldots = \bar{x}_k$$

Pentru verificarea acestei ipoteze se poate folosi testul statistic F. Valoarea calculată a lui F se compară cu valoarea tabelară; în cazul în care valoarea calculată este mai mare, ipoteza Ho se respinge; dacă este mai mică, ipoteza H_0 se acceptă.

Pentru calculul valorii lui F se parcurg următoarele etape:

a)Determinarea sumei pătratelor abaterilor mediilor corespunzătoare fiecărui eşantion ($\bar{x}_j$) de la media generală a eşantioanelor $(\bar{\bar{x}})$, ponderate cu volumul eşantioanelor (n_j).

Formula de calcul este:

$$SST = \sum_{j=1}^{k} (\bar{x}_j - \bar{\bar{x}})^2 n_j \qquad (5.4.)$$

în care SST sunt simbolurile denumirii indicatorului respectiv în limba engleză (Sum of Squares for Treatments) adica suma pătratelor abaterilor mediilor corespunzătoare fiecărui eşantion de la media generală.

b)Determinarea sumei pătratelor abaterilor variantelor x_{ij} de la media corespunzătoare fiecărui eşantion ($\bar{x}j$). În acest scop se foloseşte relaţia:

$$SSE = \sum_{j=1}^{k}\sum_{i=1}^{nj} (x_{ij} - \bar{x}_j)^2 \qquad (5.5)$$

în care SSE – simbolurile denumirii în limba engleză (Sum of Squares for Error).

c)Determinarea mediei pătratelor abaterilor mediilor de sondaj (adică a eşantioanelor) ($\bar{x}_j$) de la media generală ($\bar{\bar{x}}$), folosind relaţia:

$$MST = \frac{SST}{k-1} \qquad (5.6.)$$

unde MST - Mean Square for Treatments.

d)Determinarea mediei pătratelor abaterilor valorilor x_{ij} de la mediile de sondaj (adică a eşantioanelor) $\bar{x}_j$, folosind relaţia:

$$MSE = \frac{SSE}{n-k} \qquad (5.7.)$$

unde MSE - Mean Square for Error

e)Calculul testului statistic de semnificaţie F cu ajutorul relaţiei:

$$F = \frac{MST}{MSE} \qquad (5.8.)$$

Observăm că metodologia prezentată oferă posibilitatea analizei variaţiei la nivelul fiecărui eşantion, prin SSE, evidenţiind influenţa factorilor care acţionează în interiorul eşantioanelor şi analiza variaţiei mediilor de sondaj ($\overline{x}_j$) faţă de media generală acestora ($\overline{\overline{x}}$), prin SST, evidenţiind diferenţierile dintre eşantioane.

Rezultă că variaţia totală este dată de suma celor doi indicatori, adică:

$$\text{SS(Total)} = \text{SST} + \text{SSE} \qquad (5.9)$$

sau:

$$\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(x_{ij}-\overline{\overline{x}}\right)^2 = \sum_{j=1}^{k}\left(\overline{x}_j-\overline{\overline{x}}\right)^2 n_j + \sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(x_{ij}-\overline{x}_j\right)^2 \qquad (5.10)$$

În tabelul 5.2. este prezentată o sinteză a tuturor acestor indicatori.

Tab. 2.2. Tabel ANOVA pentru analiza variaţiei unidimensionale a eşantioanelor independente.

Sursa variaţiei	*Grade de libertate*	*Suma pătratelor*	*Media pătratelor*	*F. statistic*
$\overline{x}_j - \overline{\overline{x}}$	k-1	*SST*	$MST = \frac{SST}{k-1}$	$F = \frac{MST}{MSE}$
$x_{ij} - \overline{x}_j$	n - k	*SSE*	$MSE = \frac{SSE}{n-k}$	
Total	n – 1	*SS* (Total)		

5.2. Modelul de analiză bidimensională

În principiu, metodologia utilizată în cazul analizei variaţiei bidimensionale este similară cu cea prezentată la 5.1., deosebirea constând în aceea că modularizându-se datele în funcţie de componentele unuia dintre factori, se determină setul de indicatori pentru fiecare componentă.

Forma legăturilor bifactoriale supuse analizei ANOVA este prezentată în tabelul 5.3.

Simbolurile din tabel au următoarea semnificaţie:

$\overline{x}[AB]_{ij}$ = media valorilor x_{ij} corespunzătoare fiecărui nivel de prelucrare, atât pentru factorul A cât şi pentru factorul B;

$\overline{x}[A]_i$ = media valorilor corespunzătoare fiecărui nivel i al factorului A;

$\overline{x}[B]_j$ = media valorilor corespunzătoare fiecărui nivel j al factorului B;

$\overline{\overline{x}}$ = media tuturor valorilor;

i = numărul nivelelor pentru factorul $A\ (i = 1,2,3.....a)$;

j = numărul nivelelor pentru factorul $B\ (j = 1,2,3.....b)$;

r = numărul de valori ale lui x, întâlnite în cadrul fiecărui nivel AB.

Spre exemplu:

$\bar{x}[AB]_{11}$ - este media valorilor x întâlnite în cadrul nivelului 1 al factorului A şi nivelului 1 al factorului B;

$\bar{x}[A]_1$ - este media valorilor x din nivelul 1 al factorului A;

$\bar{x}[B]_1$ - este media valorilor x din nivelul 1 al factorului B.

Suma pătratelor abaterilor este dată de următoarele relaţii:

$$SS(Total) = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}\left(x_{ijk} - \bar{\bar{x}}\right)^2 \tag{5.11}$$

$$SS(A) = rb\sum_{i=1}^{a}\left(\bar{x}[A]_i - \bar{\bar{x}}\right)^2 \tag{5.12}$$

$$SS(B) = ra\sum_{j=1}^{b}\left(\bar{x}[B]_j - \bar{\bar{x}}\right)^2 \tag{5.13}$$

$$SS(AB) = r\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{x}[AB]_{ij} - \bar{x}[A]_i - \bar{x}[B]_j + \bar{\bar{x}}\right)^2 \tag{5.14}$$

$$SSE = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}\left(x_{ijk} - \bar{x}[AB]_{ij}\right)^2 \tag{5.15}$$

Tab. 5.3 Notaţiile pentru modelul ANOVA bifactorial

Factorul B	Factorul A				
	1	2	....	a	
1	$\left.\begin{matrix} x_{111}\\ x_{112}\\ .\\ .\\ x_{11r}\end{matrix}\right] \bar{x}[AB]_1$	$\left.\begin{matrix} x_{211}\\ x_{212}\\ .\\ .\\ x_{21r}\end{matrix}\right] \bar{x}[AB]_{21}$	....	$\left.\begin{matrix} x_{a11}\\ x_{a12}\\ .\\ .\\ x_{a1r}\end{matrix}\right] \bar{x}[AB]_{a1}$	$\bar{x}[B]_1$
2	$\left.\begin{matrix} x_{121}\\ x_{122}\\ .\\ .\\ x_{12r}\end{matrix}\right] \bar{x}[AB]_1$	$\left.\begin{matrix} x_{221}\\ x_{222}\\ .\\ .\\ x_{22r}\end{matrix}\right] \bar{x}[AB]_{22}$	....	$\left.\begin{matrix} x_{a21}\\ x_{a22}\\ .\\ .\\ x_{a2r}\end{matrix}\right] \bar{x}[AB]_{a2}$	$\bar{x}[B]_2$
:	:	:	:	:	:
b	$\left.\begin{matrix} x_{1b1}\\ x_{1b2}\\ .\\ .\\ x_{1br}\end{matrix}\right] \bar{x}[AB]_1$	$\left.\begin{matrix} x_{2b1}\\ x_{2b2}\\ .\\ .\\ x_{2br}\end{matrix}\right] \bar{x}[AB]_{2b}$	....	$\left.\begin{matrix} x_{ab1}\\ x_{ab2}\\ .\\ .\\ x_{abr}\end{matrix}\right] \bar{x}[AB]_{ab}$	$\bar{x}[B]_b$

	$\bar{x}[A]_1$	$\bar{x}[A]_2$	....	$\bar{x}[A]_a$	$\bar{\bar{x}}$

Testul statistic F pentru verificarea ipotezei diferenţei nule H_0 se determină astfel:

a) Pentru diferenţa dintre mediile corespunzătoare nivelelor factorului A:

$H_0 : \bar{x}[A]_1 = \bar{x}[A]_2 = ... = \bar{x}[A]_a$

$H_1 :$ cel puţin două medii diferă.

Testul statistic:

$$F = \frac{MS(A)}{MSE} \qquad (5.16)$$

b) Pentru diferenţa dintre mediile corespunzătoare nivelelor factorului B:

$H_0 : \bar{x}[B]_1 = \bar{x}[B]_2 = ... = \bar{x}[B]_b$

H_1 :cel puţin două medii sunt diferite.

Testul statistic:

$$F = \frac{MS(B)}{MSE} \qquad (5.17)$$

c)Pentru interacţiunea între factorii A şi B:

H_0 : Factorii A şi B nu interacţionează pentru a afecta nivelul mediilor.

H_1 : Interacţiunea celor doi factori afectează nivelul mediilor.

Testul statistic:

$$F = \frac{MS(AB)}{MSE} \qquad (5.18)$$

Sinteza metodologiei de calcul pentru aceşti indicatori este prezentată în tabelul 5.4.

Tab. 5.4. Tabel ANOVA pentru analiza variaţiei bidimensionale a eşantioanelor independente.

Sursa variaţiei	*Grade de libertate*	*Suma pătratelor*	*Media pătratelor*	*F - statistic*
Factorul A	a – 1	*SS(A)*	$MS(A) = \frac{SS(A)}{a-1}$	$F = \frac{MS(A)}{MSE}$
Factorul B	b – 1	*SS(B)*	$MS(B) = \frac{SS(B)}{b-1}$	$F = \frac{MS(B)}{MSE}$
Interacţiune AB	(a-1)(b-1)	*SS(AB)*	$MS(AB) = \frac{SS(AB)}{(a-1)(b-1)}$	$F = \frac{MS(AB)}{MSE}$
Abateri Individuale $(x_{ijk} - \bar{x}[AB])_{ij}$	n – ab	*SSE*	$MSE = \frac{SSE}{n-ab}$	
Total	n – 1	SS(Total)		