



Prefață

Analiza datelor cu Microsoft Excel vă propune o incursiune introductivă în lumea instrumentelor **Microsoft Excel** dedicate analizei statistice a datelor experimentale. Deși **MS Excel** nu este un software dedicat prelucrării statistice a datelor, răspândirea acestui software și mai ales faptul că acesta conține un **add-ins** puternic pentru analiza datelor, **Data Analysis**, îl recomandă pentru un curs introductiv în Statistică. Această carte presupune că utilizatorul are deja cunoștințele de bază privind lucrul cu **MS Excel** și este la început de drum privind **Statistica**. Odată stăpânite conceptele de bază ale statisticii, trecerea la produse profesionale dedicate prelucrării statistice a datelor, ca de exemplu SPSS¹, PSPP² etc. devine foarte simplă și ușoară.

Prezenta lucrare face o introducere graduală în problematica prelucrării datelor. În acest context:

- **Capitolul 1** introduce principalele concepte de statistică descriptivă și statistică distributivă, prezentând noțiunile de variabilă, frecvențe, distribuții, histogramme, medii, mediane și abateri, precum și modul în care **MS Excel** ne ajută să calculăm aceste valori.
- **Capitolul al 2-lea** tratează distribuția probabilităților, prezentând conceptele de probabilitate, distribuție a acestora, variabilă aleatoare și detaliază cea mai utilizată distribuție, distribuția normală.
- **Capitolul al 3-lea** stabilește cum pot fi determinate în **MS Excel** intervalele de încredere, cum pot fi testate ipotezele de lucru și descrie o altă distribuție, distribuția t (Student).

¹ *** <https://www.ibm.com/products/spss-statistics> (accesat la 25.11.2022).

² *** <https://www.gnu.org/software/pspp/> (accesat la 25.11.2022).

- **Capitolul al 4-lea** este dedicat tabelelor pivot, tabelelor cu două căi și testului Pearson χ^2 și locului pe care acestea le ocupă în analiza statistică.
- **Capitolul al 5-lea** ne învață să efectuăm regresia liniară simplă, să verificăm modelul regresional, să determinăm coeficientul de corelație și să construim matricea de corelație.
- **Capitolul al 6-lea** este dedicat regresiei multiple și cum pot fi testate presupunerile.
- **Capitolul al 7-lea** tratează analiza varianței cu un singur factor cât și a celei cu doi factori.
- **Capitolul al 8-lea** introduce seriile de timp. Este doar o introducere în problematica previzionării, prezentând câteva dintre principalele metode de previzionare, cum sunt media mobilă și netezirea exponențială cu unul și cu doi parametri.
- **Capitolul al 9-lea** arată cum poate fi utilizat **MS Excel** în controlul statistic al calității, cum pot fi generate diagramele de control și cum poate fi construită diagrama χ .
- **Anexa 1** conține o listă a funcțiilor **MS Excel** prezentate în această carte. Nu este o listă completă a tuturor funcțiilor de care dispune **MS Excel** și de aceea vă invită să folosiți sistemul de ajutor din **MS Excel** precum și sursele bibliografice pentru a identifica și multe alte funcții care v-ar putea fi utile în analiza statistică a datelor.
- **Anexa 2** vă prezintă add-ins-ul **Data Analysis** și instrumentele care vă ajută să efectuați analize statistice complexe în **Excel**.
- **Anexa 3** cuprinde un mic dicționar de termeni statistici utilizați în lucrare.

Desigur că pot fi adăugate și multe alte elemente dar lăsăm și în seama cititorului sarcina de a explora mulțimea de funcții și metode de care dispune **MS Excel** pentru analiza statistică a datelor.

Prezenta lucrare nu este un curs de statistică, deși prezintă în mod succint principalele elemente de statistică la care se face referință în fiecare capitol. Mai mult chiar, presupune că utilizatorul are cunoștințe de bază în statistică și chiar încurajează, prin bibliografia suplimentară de

la finalul fiecărui capitol, utilizatorii să consulte diversele tutoriale și materiale sugerate.

Prezenta lucrare nu este nici un curs introductiv în **MS Excel**. Ea pleacă de la presupunerea că utilizatorul este deja familiarizat cu mediul de lucru **Excel** și știe să-i folosească funcțiile de bază.

Autorii doresc să mulțumească dlui lector dr. Remus Butănescu-Volanin, dnei prof. dr. ing. Carmen Simion precum și dlui prof. univ. dr. ing. Florin Lungu pentru recenzia atentă efectuată asupra acestui material și pentru observațiile pertinente care au condus la îmbunătățirea acestuia. Mulțumim de asemenea studenților noștri fără de care această carte nu ar fi apărut. Le mulțumim și pentru răbdarea de a studia materialul și dorința de a ne împărtăși din problemele și dificultățile întâmpinate la înțelegerea materialului.

În speranța că acest material se va dovedi util tuturor celor care vor să învețe acest produs și conștient că acesta poate fi încă îmbunătățit, autorii mulțumesc totodată și tuturor celor care vor face observații, vor aduce completări sau vor veni cu sugestii de îmbunătățire.

Sibiu,
10.01.2023

Autorii

1

Statistică descriptivă

În acest capitol veți învăța:

- Ce sunt variabilele și de câte feluri sunt acestea.
- Ce se înțelege prin statistică descriptivă și statistică inferențială.
- Cum să creați tabele de frecvențe, frecvențe cumulative, procentaje și procentaje cumulative.
- Cum să creați histograme și să grupați histogramele în grupuri.
- Cum să folosiți Excel pentru a face diverse calcule din cadrul statisticii descriptive.

1.1. Variabile și statistică descriptivă

Statistica descriptivă folosește o serie de instrumente matematice pentru a sintetiza valorile dintr-un set de date, ca de exemplu atunci când se utilizează datele privind profiturile medii ale unor companii dintr-o anumită ramură pentru a se observa care sunt companiile care fac profit maxim și care nu. Statistica descriptivă nu se folosește pentru a trage concluzii generale și nici pentru a face previziuni. Pentru acestea din urmă se utilizează **statistica inferențială**. Scopul statisticii descriptive este acela de a descrie conținutul unui anumit set de date. Acest tip de statistică poate implica o singură variabilă sau mai multe variabile. **Variabila** reprezintă o caracteristică singulară a unui obiect sau eveniment, ca de exemplu vârsta studenților dintr-o grupă. Variabilele pot fi:

- **Cantitative**, care implică valori care apar sub forma unor numere care nu sunt arbitrare, ce de exemplu vârsta, greutatea, venitul lunar etc.. Acestea, la rândul lor pot fi:
 - **Discrete**, care presupun valori numerice dintr-o listă definită, ca de exemplu valorile unui zar sau numărul studenților dintr-o grupă.
 - **Continue**, care pot avea valori dintr-o gamă foarte largă de numere, cum sunt numerele aparținând mulțimii numerelor reale, ca de exemplu valorile greutateii indivizilor: 52,3 kg, 76,3 kg etc.

- **Calitative**, valori care se încadrează într-o anumită categorie, indicând proprietatea unui obiect, ca de exemplu genul, apartenența etnică, numele produsului etc. Variabilele calitative pot fi:
 - **Ordinale**, variabile cărora le pot fi asociate o ordine, ca de exemplu ordinea preferințelor.
 - **Nominale**, valori care nu pot fi clasificate într-o anumită ordine, ca de exemplu rasa, județul sau genul.

Figura 1.1 exemplifică variabile cantitative discrete (e.g. coloana Vârsta) și calitative nominale (e.g.: coloanele Genul și Continentul).

ID Student	Vârsta	Genul (0 = Barbat 1 = Femeie)	Continent (0 = Europa 1 = Africa 2 = America 3 = Asia)
1000	18	1	0
1001	19	0	1
1002	23	1	2
1003	21	1	0
1004	22	0	0
1005	18	0	0
1006	18	0	1
1007	17	1	1
1008	19	0	1
1009	20	1	2
1010	17	1	3
1011	18	0	0
1012	18	0	0

Figura 1.1. *Exemple de valori cantitative și calitative*

1.2. Distribuții folosind tabele de frecvențe

Distribuția valorilor se referă la modul în care datele sunt repartizate într-un domeniu de valori. De exemplu, pentru datele din figura 1.1 poate fi semnificativă vârsta medie a studenților. Sau câți studenți provin din Asia?

Pentru a oferi un răspuns, un prim pas îl presupune construcția unui tabel cu frecvențele vârstelor sau numărul studenților grupat pe continente. Un **tabel de frecvențe** este un tabel în care sunt numărate aparițiile unei anumite variabile.

Pentru a afla câți studenți au 17, 18, 19 etc. ani, se poate proceda astfel (Fig. 1.2):

1. Selectați coloana pentru care doriți să calculați frecvențele, de exemplu \$B1:\$B14.
2. Dați clic pe pictograma **Analyze Data** de sub eticheta **Home**.
3. Se va deschide fereastra **Analyze Data** care vă va sugera o listă de răspunsuri. Observați cum valorile pentru vârstă sunt aranjate în ordinea descrescătoare a frecvențelor și, alăturat lor, numărul de apariții.
4. Dați clic pe butonul **+ Insert PivotTable** pentru a insera aceste date în foaia de calcul. Acest nou tabel va fi inserat într-o nouă foaie de calcul botezată implicit **Suggestion**. Modificați denumirile câmpurilor și al foii de calcul cu nume semnificative pentru Dvs., de exemplu **Frecvența Vârste** (Fig. 1.3).

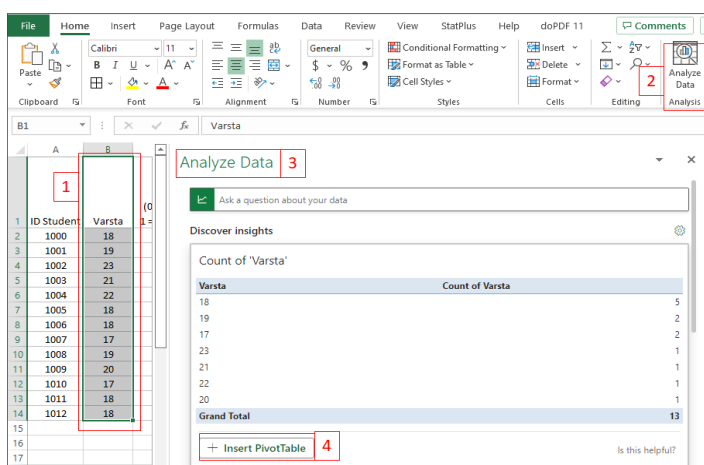


Figura 1.2. Crearea tabelului frecvențelor pentru coloana Vârsta

Varsta	Frecvența Varsta
18	5
17	2
19	2
23	1
22	1
21	1
20	1
Grand Total	13

Figura 1.3. Tabelul frecvențelor pentru coloana Vârsta

Dacă doriți să grupați datele pentru care vreți să aflați frecvențele, atunci puteți utiliza **containerele** (bins). De exemplu, dacă ne interesează grupele de vârstă < 20 de ani, 20..21 și peste 22 de ani, putem crea 3 containere și folosi funcția **FREQUENCY** din Excel (Fig. 1.4), astfel:

	A	B	C	D	E	F
			Genul (0 = Barbat 1 = Femeie)	1 = Africa 2 = America 3 = Asia)		
1	ID Student	Varsta				
2	1000	18	1	0		
3	1001	19	0	1		
4	1002	23	1	2		
5	1003	21	1	0		
6	1004	22	0	0		
7	1005	18	0	0		
8	1006	18	0	1		
9	1007	17	1	1		
10	1008	19	0	1		
11	1009	20	1	2		
12	1010	17	1	3		
13	1011	18	0	0		
14	1012	18	0	0		
15						
16		Containere		Rezultat		
17		19	1	=FREQUENCY(Sub 20 de ani	
18		22			20 .. 22 de ani	
19					Peste 22 de ani	
20						

Figura 1.4. Calculul frecvențelor folosind containerele

1. Introduceți în celula \$D17, formula: =FREQUENCY(B2:B14,B17:B18)
2. Excel va completa pe liniile următoare valorile din container, așa cum se poate observa în figura 1.5. Adăugați manual explicații la rezultatele obținute.

Containere	Rezultat	
19	9	Sub 20 de ani
22	3	20 .. 22 de ani
	1	Peste 22 de ani

Figura 1.5. Rezultatele obținute cu funcția **FREQUENCY**