


Seminar 1

Scopul acestei activități practice este însușirea tehnicilor de calcul specifice statisticii descriptive. Activitatea practică va fi efectuată folosind programul EXCEL.

Creați un nou fișier Microsoft Excel cu denumirea "**Nume_seminar 1**" și salvați-l în **directorul specific cursului**.

Folosirea programului EXCEL

Pentru a putea efectua calculele și analizele din cadrul activității practice sunt necesare cateva cunoștințe de bază a programului EXCEL. În cadrul activității practice veți folosi formule și funcții EXCEL pentru a genera statistici descriptive. De asemenea veți folosi unele reprezentări grafice pentru a obține histograme de frecvență.

O formulă EXCEL este o instrucțiune pentru programul EXCEL, care va face o anumită operație matematică într-o celulă, în funcție de conținutul altor celule. Programul EXCEL este proiectat astfel încât scrierea formulelor să fie cât mai simplă. Acest lucru este posibil prin utilizarea funcțiilor care au la bază anumite formule, și îl ajută pe utilizator să evite tastarea manuală a formulelor. Lista completă a funcțiilor valabile poate fi vizualizată prin "klik" pe icoana  din meniu.

Cei ce cunosc sintaxa formulelor, pot tasta pur și simplu formula.

De exemplu =MAX(A1:A20) ne va da maximumul valorilor aflate în celulele A1:A20.

Notă: Formulele pot fi introduse cu litere mici (lower case) sau mari (upper case), dar programul EXCEL va afișa funcția cu litere mari. **Pentru ca o formulă să fie recunoscută, ea trebuie să fie precedată de semnul "="**. De asemenea este recomandat să se folosească paranteze de câte ori este posibil.

"Referirea" la celule (cell referencing)

Modalitatea de "referire" la celule este foarte importantă și poate fi absolută, mixtă sau relativă.

Tipul referirii	Exemple
Absolut	\$A\$1
Mixt (fixarea rândului sau a coloanei)	A\$1 sau \$A1
Relativ	A1

În cazul "referirii" absolute, programul va lua valoarea din aceeași celulă indiferent unde este copiată sau mutată formula, sau dacă este umplut un rând sau o coloană. Simbolul \$ fixează rândul și/sau coloana (fixarea unei celule se poate face după selectarea ei prin apăsarea tastei **F4**).

În cazul "referirii" relative, programul schimbă celula în cazul mutării formulei. De exemplu, dacă formula "A2-1" tastată în celula B2 este mutată în celula H8 va deveni "G8-1".

"Referirea" mixtă este un amestec al tipurilor amintite mai sus. Dacă simbolul \$ este pus **în fața literei**, atunci **este fixată coloana**, prin deplasarea formulei se va modifica doar rândul, iar dacă simbolul \$ este pus **în fața cifrei**, atunci **este fixat rândul**, prin deplasarea formulei se va modifica doar coloana, rândul rămânând același.

Deplasarea unei formule se poate face după selectarea celulei respective prin copiere (copy "**Ctrl C**" și paste "**Ctrl V**") sau prin "*tragerea*" conținutului celulei respective cu ajutorul "mouse-ului" pentru a umple un rând sau o coloană (de colțul din dreapta jos, când săgeata care reprezintă mouse-ul se transformă într-o cruciuliță neagră)

Exersați aceste tipuri diferite de "referire" la diferite celule. Trebuie amintit, că rezultatul unei formule este afișat într-o celulă, în timp ce formula este afișată în *bara de funcție* de sub meniu. Tipul de referire la o celulă poate fi schimbat prin adăugarea sau stergerea simbolului \$ sau prin apăsarea tastei F4.

Reprezentări Grafice pentru o Variabilă Cantitativă

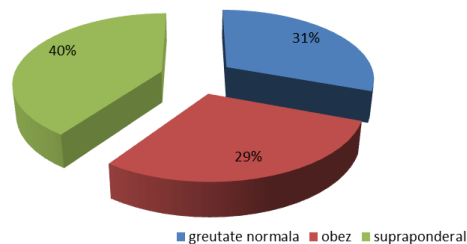
P1) Greutatea și înălțimea unui eșantion de 62 pacienți au fost măsurate și pe baza măsurătorilor s-a calculat pentru fiecare pacient indicele de masă corporală. În funcție de valorile indicelui de masă corporală subiecții au fost clasificați ca fiind cu greutate normală, supraponderal și respectiv cu obezitate. În eșantionul investigat avem 19 pacienți au avut greutate normală, 25 supraponderali și 18 obezi.

- Creați un tabel de frecvență cu datele din problemă.

Greutate normală	
Supraponderal	
Obez	

- Creați un grafic de tip "pie". Formatați graficul astfel încât să arate ca în figură:

Indicele de masă corporală: clase de frecvență

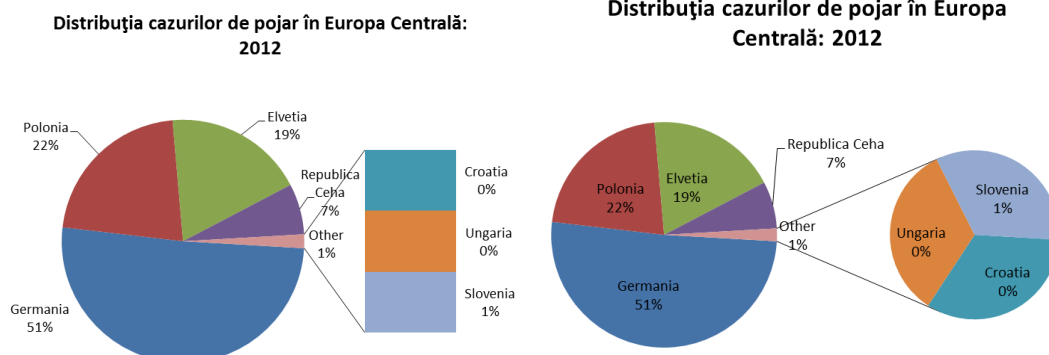


P2) S-a realizat un studiu pentru a analiza numărul de cazuri de pojar în anul 2012 în Germania, Polonia, Elveția, Republica Cehă, Croația, Ungaria, și Slovenia. Datele aferente la două au fost colectate: țara (Germania / Polonia / Elveția / Republica Cehă / Croația / Ungaria / Slovenia) și respectiv pojar (da/nu).

Datele au fost sumarizate și sunt prezentate în tabelul următor:

Țara	Număr de cazuri de pojar raportate în 2012
Germania	166
Polonia	71
Elveția	61
Republica Cehă	22
Croația	2
Ungaria	2
Slovenia	2

- Creați un grafic de tip Pie of Pie sau Pie of Bar utilizând datele cu privire la incidența pojarului în țări din Europa Centrală: Reprezentarea grafică realizată trebuie să fie similară cu una din reprezentările:



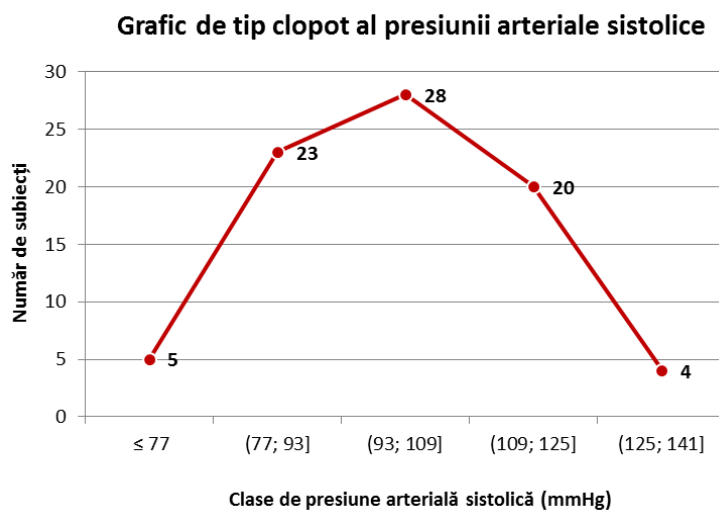
O diagramă de tip Plăcintă (Pie) este o reprezentare grafică circulară utilizată pentru a vizualiza părți ale întregului.

Reprezentări Grafice pentru o Variabilă Cantitativă

P3) Presiunea arterială sistolică (mmHg) s-a măsurat la un eșantion de 80 persoane. Tabelul de frecvență obținut în urma colectării datelor este:

Clase de frecvență Presiunea arterială sistolică (mmHg)	Frecvența absolută
≤ 77	5
(77; 93]	23
(93; 109]	28
(109; 125]	20
(125; 141]	4

- Creați o reprezentare grafică ca și în figura de mai jos:



Acest tip de reprezentare grafică se utilizează pentru vizualizarea distribuției datelor. În aceeași categorie se înscrie și histograma.

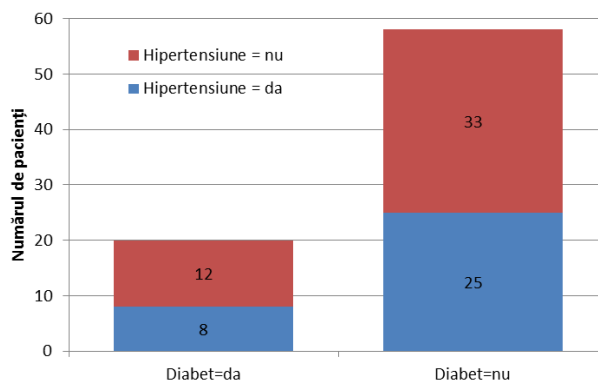
Reprezentarea Grafică a Două Variabile Calitative

P4) *Dependența dintre hipertensiune și diabet a fost investigată pe un eșantion de 78 pacienți. Pentru fiecare pacient au fost colectate următoarele date: prezența/absența hipertensiunii și prezența/absența diabetului.*

Tabelul de contingență care sumarizează datele colectate este:

	Diabet=da	Diabet=nu
Hipertensiune = da	8	25
Hipertensiune = nu	12	33

- Pe baza datelor din tabelul de contingență creați un grafic de tip Stacked Column. Graficul trebuie să fie similar cu cel din imaginea următoare:



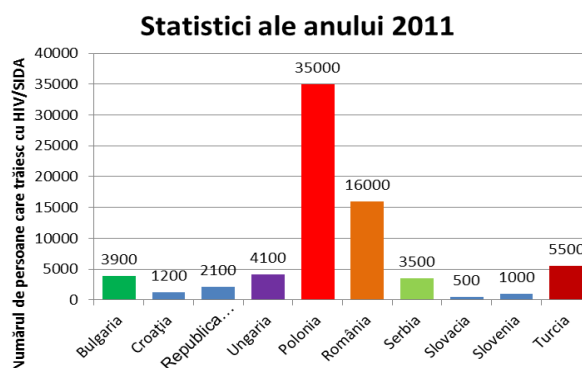
P5) *S-a realizat un studiu pentru a identifica numărul de subiecți infectați cu HIV sau diagnosticați cu SIDA în 2011 în Bulgaria, Croația, Czech Republic, Hungary, Poland, Romania, Serbia, Slovakia, Slovenia, și Turkey).*

Sumarizarea datelor colectate este prezentată în tabelul următor:

Țara	Persoane care trăiesc cu HIV/SIDA
Bulgaria	3900
Croația	1200
Republica Cehă	2100
Ungaria	4100
Polonia	35000
România	16000
Serbia	3500
Slovacia	500
Slovenia	1000
Turcia	5500

- Formatați coloana *Persoane care trăiesc cu HIV/SIDA* ca număr fără zecimale: [Format Cells ... - Number – Number without decimals].

- Creați un grafic de tip Coloane și formatați-l ca să arate ca și în figura:

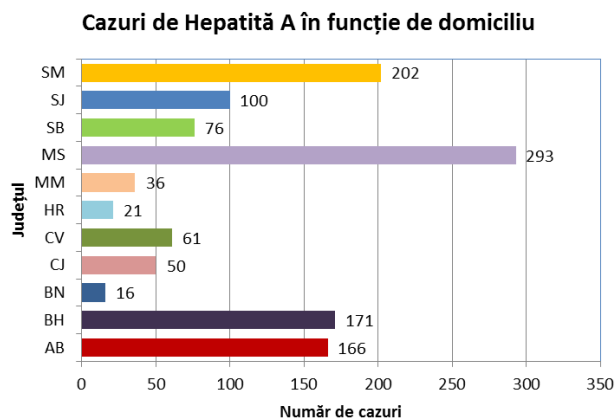


P6) S-a investigat numărul de cazuri de hepatită A în 11 județe din România (AB = Alba Iulia, BH = Bihor, BN=Bistrita Năsăud, CJ = Cluj, CV= Covasna, HR = Harghita, MM = Maramureș, MS = Targu Mures, SB = Sibiu, SJ = Sălaj și SM = Satu-Mare).

Sumarizarea datelor colectate este prezentată în tabelul următor:

	AB	BH	BN	CJ	CV	HR	MM	MS	SB	SJ	SM
Hepatita A	166	171	16	50	61	21	36	293	76	100	202

- Formatați rândul Hepatită A ca număr fără zecimale.
- Creați un grafic de tip bare și formatați-l astfel încât să arate ca în imaginea de mai jos:

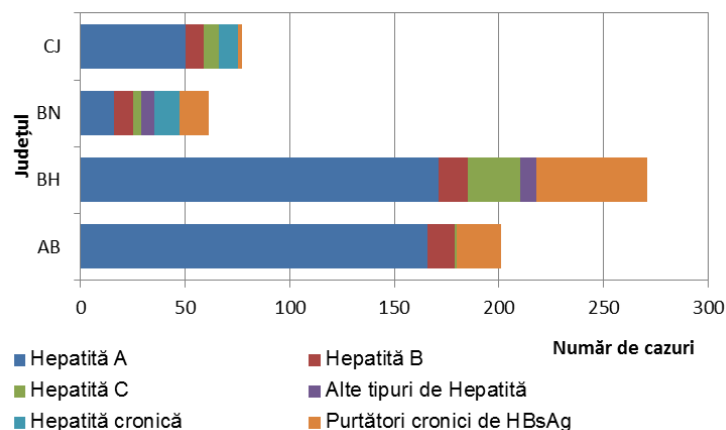


P7) Tipul de hepatită (A, B, C, alte tipuri, hepatită cronică și respectiv purtători cronici de HbsAg) a fost investigat în 4 județe din România.

Sumarizarea datelor colectate este prezentată în tabelul următor:

	AB	BH	BN	CJ
Hepatită A	166	171	16	50
Hepatită B	13	14	9	9
Hepatită C	1	25	4	7
Alte tipuri de hepatită	0	8	6	0
Hepatită cronică	0	0	12	9
Purtători cronici de HBsAg	21	53	14	2

- Creați un grafic de tip Stacked Bar:



Graficul de tip coloane este compus din coloane discrete, fiecare coloană reprezentând o categorie diferită. Înălțimea coloanei este egală cu cantitatea din categoria dată. Similar cu graficul de tip Bare, graficul de tip coloane se folosește pentru a compara valorile diferitelor categorii.

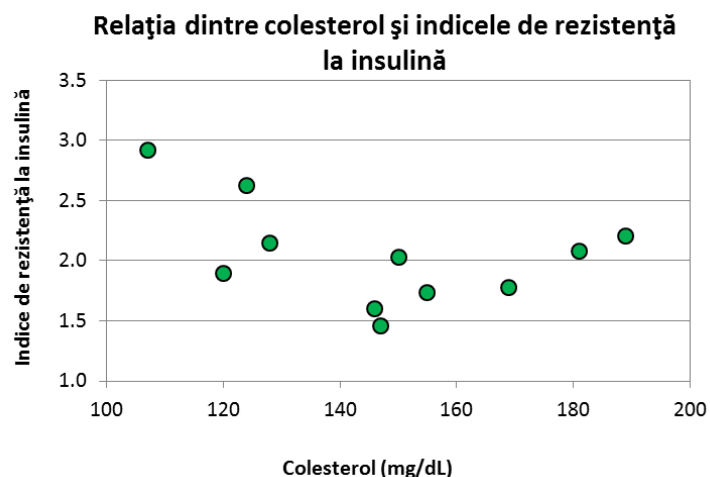
Reprezentarea Grafică a Două Variabile Cantitative

P8) S-a realizat un studiu pe un eșantion de 11 pacienți pentru a analiza relația dintre colesterolul total și indicele de rezistență la insulină.

Tabele colectate sunt prezentate în tabelul de mai jos:

Colesterol (mg/dL)	Indice de rezistență la insulină
181	2.08
146	1.60
155	1.73
107	2.92
128	2.14
120	1.90
150	2.03
169	1.77
147	1.46
189	2.21
124	2.62

- Copiați tabelul anterior în fisierul personal.
- Formatați coloana *Colesterol* ca număr fără zecimale.
- Formatați coloana *Indice de rezistență la insulină* ca număr cu două zecimale.
- Creați o reprezentare grafică de tip "scatter" și formatați-o ca și în figura următoare:



Graficul de tip Scatter permite vizualizarea relației dintre două variabile cantitative dependente. Datele sunt prezentate ca o colecție de puncte, fiecărui punct corespunzându-i valoarea primei variabile pe axa OX și respectiv valoarea celei de-a doua variabile pe axa OY.

Alte Tipuri de Reprezentări Grafice

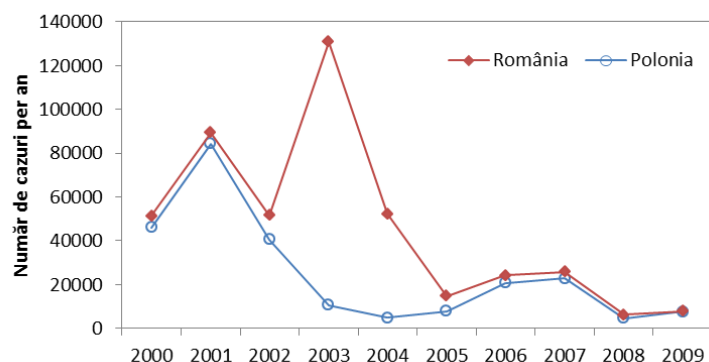
P9) S-a investigat numărul de cazuri de rubeolă din Polonia și România în perioada 2000-2009.

Sumarizarea datelor colectate este prezentată în tabelul următor:

Țara	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Polonia	46181	84419	40518	10588	4857	7946	20668	22890	4598	7586
România	50125	85076	51079	120377	47444	6801	3563	2958	1746	343

- Copiați datele din tabelul anterior într-un nou "worksheet" denumit *Linie*.
- Formatați rândurile ca număr fără zecimale.
- Creați un grafic de tip linie și formatați-l ca și în imaginea de mai jos:

Trendul cazurilor de rubeolă



Graficul de tip linie permite vizualizarea relației dintre două variabile și este utilizat pentru ilustrarea modificărilor în timp.

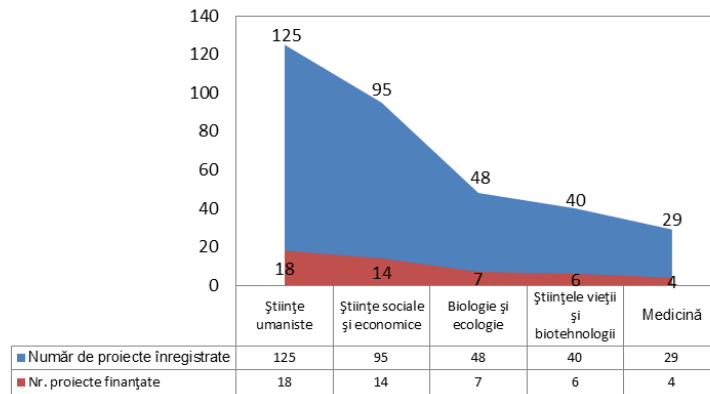
P10) S-a analizat numărul de proiecte submise și respectiv finanțate pentru proramul de burse individuale postdoctorale în România (2012) pentru cinci arii de cercetare (Științe umaniste, Științe sociale și economice, Biologie și ecologie, Biotehnologii și Medicină).

Sumarizarea datelor colectate este prezentată în tabelul de mai jos:

Domeniul de cercetare	Număr de proiecte înregistrate	Nr. proiecte finanțate
Științe umaniste	125	18
Științe sociale și economice	95	14
Biologie și ecologie	48	7
Biotehnologii	40	6
Medicină	29	4

- Copiați tabelul anterior într-un nou "worksheet" denumit **Arie**.
- Formatați coloanele *Nr. proiecte înregistrate* și *Nr. proiecte finanțate* ca număr fără zecimale.
- Creați un grafic de tip Arie și formatați-l ca și în imaginea de mai jos:

Proiecte de cercetare postdoctorală competiția 2012



Graficul de tip arie permite vizualizarea sumarizărilor cantitative și arată importanța relativă a valorilor fiind utilizat pentru a compara doua sau mai multe cantități

2. În foaia de calcul **Buline** creați reprezentarea grafică cerută mai jos.

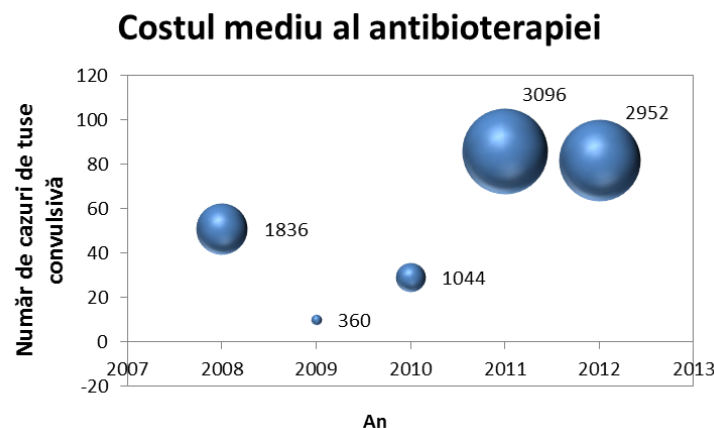
P11) S-a realizat un studiu pentru a evalua costul antibioterapiei în tratamentul tusei convulsive pe o durată de 5 ani (2008-2012).

Sumarizarea datelor colectate este prezentată în tabelul de mai jos:

Anul	Nr. cazuri tuse convulsivă	Costul mediu al antibioterapiei
2012	82	2952
2011	86	3096
2010	29	1044
2009	10	360
2008	51	1836

- Copiați tabelul de mai sus într-un "worksheet" denumit *Buline*.

- Creați reprezentarea grafică de tip "Bubble" și formatați-o astfel încât să arate ca cea din figura următoare:



Graficul de tip "buble" permite reprezentarea grafică simultană a 3 dimensiuni. Mărimea bulinei indică valoarea celei de a treia dimensiuni în timp ce prima și cea de-a doua dimensiune sunt reprezentate pe axa OX și respectiv OY.

Tema: Pentru fiecare din datele urmatoare realizați reprezentarea grafică potrivită. Redenumiți foile de calcul cu tipul reprezentării grafice. Salvați documentul (Nume_Tema_grafice) și trimiteți fișierul prin e-mail (atașat)

a. Glicemia versus nivelul total al colesterolului sanguin

Glicemie (mg/dL)	75	92	77	128	81	138	88	72	71	80	91	94	90	96
Colesterol total (mg/dL)	168	343	229	157	161	192	218	159	272	195	220	246	147	175

b. Costul mediu al tratamentului cu antibiotice pentru tusea convulsivă în asociere cu numărul de cazuri raportate

Anul	Costul mediu al tratamentului antibiotic	Nr. cazuri tuse convulsivă
2000	17748	493
1999	2952	82
1998	3492	97
1997	9468	263
1996	33372	927

c. Numărul de cazuri de tuse convulsivă și pojar în România în intervalul 2000-2012

Anul	2012	2011	2010	2009	2008
Tuse convulsivă	82	86	29	10	51
Pojar	7450	4189	193	8	12

d. Populația din județul Cluj ca funcție a ariei de rezidență (rural/urban, 2010)

Grupa de vârstă (ani)	Total	Urban	Rural
≤4	244589	128735	115854
5 - 9	235719	114571	121148
10-14	246079	114909	131170
15-19	268288	136903	131385
20-24	374411	211521	162890
25-34	747264	430203	317061
35-44	721838	406603	315235
45-54	586172	350814	235358
55-64	553880	311276	242604
65-74	387927	182821	205106
75-84	227548	95072	132476
> 84	47135	20169	26966

e. Scorul Qol în relație cu clasele simptomatice:

	Pur obstructiv	Pur irigativ	Mixt
Qol<4	31	0	3
Qol≥4	102	4	72

f. Scorul clinic la internare: pacienți cu insuficiență respiratorie acută:

Scor clinic	Nr. pacienți
4	1
5	1
6	5
7	5
8	3
9	12
10	3