

Biostatistică – Medicină Generală

Lucrarea de laborator Nr.3

Scop: la sfârșitul laboratorului veți ști:

- Să efectuați și să interpretați histogramme cu ajutorul pachetului EXCEL
- Să efectuați și să interpretați alte tipuri de grafice cu ajutorul pachetului EXCEL

1. Histograma

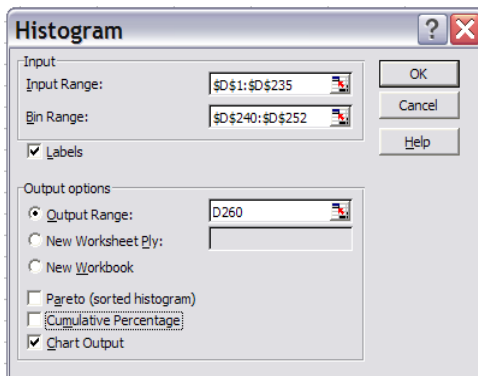
Histogramele sunt printre cele mai importante grafice în statistică, deoarece sunt bogate în informație despre distribuția valorilor pe o coloană numerică.

Să presupunem că dorim să reprezentăm printr-o histogramă distribuția vârstelor la pacienții din tabelul CIROZA. Pentru aceasta, vom proceda astfel:

Deschidem tabelul CIROZA, dacă nu este deschis.

	C	D	E	F
cu	Constantin	47	b	r
	Marin	57	b	r
	Tudora	61	f	u
	Maria	60	f	u
	Gheorghe	64	b	r
	Alexandru	61	b	r
	Stefan	62	b	r
	Joita	64	f	u
	min	26		
	max	78		
		Clasa		
		25		
		30		
		35		
		40		
		45		
		50		
		55		
		60		
		65		
		70		
		75		
		80		

- Coborâm pe coloana D, pe care se află înregistrate vârstele pacienților, până la celula D237.
- În această celulă, scrieți formula $=\min(d2:d235)$, și după apăsarea tastei ENTER, programul va calcula cea mai mică vârstă.
- În celula D238, scrieți $=\max(d2:d235)$ și după apăsarea tastei ENTER, programul va calcula cea mai mare vârstă.
- În celulele C237 și C238 scrieți Minim și, respectiv Maxim, pentru a nu uita ce reprezintă valorile calculate de program. Între 26 și 78 este o diferență de 52 de ani, deci, putem pune clasele din 5 în 5 ani (din 10 în 10 rezultă prea puține clase).
- Pe celula D240 scrieți cuvântul Clasa, iar de la D241 în jos scrieți limitele claselor (granițele dintre clase): 25, 30, 35...până la 80.
- Notați că aceste granițe ale claselor sunt situate de la celula D241 la D252, căci programul o să întrebe despre ele.



Executați secvența **Tools** → **Data Analysis**, iar din fereastra care apare, alegeți **Histogram**. Apoi apăsați butonul OK.

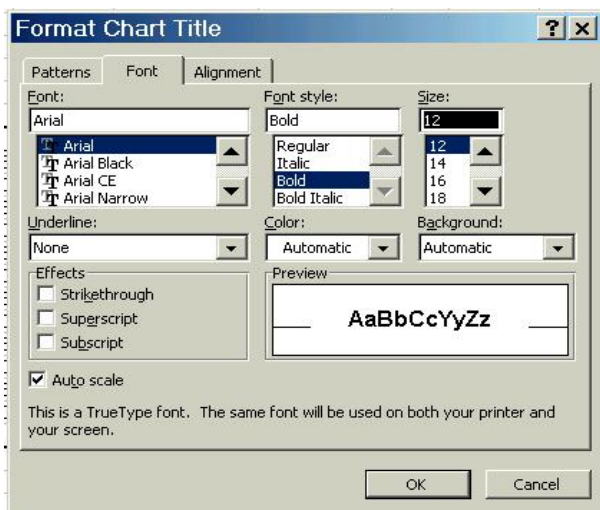
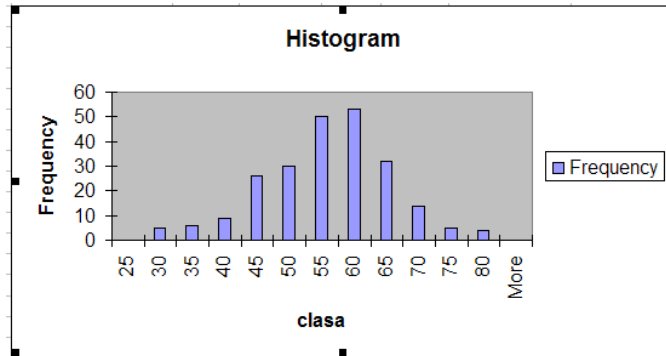
În fereastra care a apărut, completați ca în figura din stânga (fără să vă preocupați de semnele \$ pe care programul le inserează în texte):

- D1:D235, pentru **Input Range**.
- D240:D252 pentru **Bin Range**
- **Bifați Labels**
- D260, pentru **Output Range**
- Bifați caseta de validare **Chart Output**
- **Apoi apăsați OK**

Rezultatul este apariția tabelului din figura de mai jos, din stânga, precum și a histogrammei (dreapta), care este un grafic mic și destul de inestetic. Se aranjează graficul astfel încât să arate mai estetic.

Prima operație este dimensionarea. Trageți de colțuri sau de mijloacele laturilor, care sunt marcate de pătrățele mici de culoare închisă, în total 8 astfel de markeri, patru la colțuri și patru la fiecare mijloc de latură. Dacă lipsesc acești markeri, înseamnă că graficul nu este selectat. În acest caz, selectați graficul executând un clic pe suprafața lui, într-o zonă albă. Apoi trageți, pentru dimensionarea graficului. Ajustarea dimensiunii se face în pași, tragerile bruște aducând probleme. Trageți de grafic astfel încât să ocupe o suprafață cât mai mare, chiar dacă acoperă tabelul furnizat la pasul precedent. În mod normal, trebuie să ocupe aproape toată suprafața liberă de pe ecran.

Clasa	Frequency	Cumulative %
25	0	,00%
30	5	2,14%
35	6	4,70%
40	9	8,55%
45	26	19,66%
50	30	32,48%
55	50	53,85%
60	53	76,50%
65	32	90,17%
70	14	96,15%
75	5	98,29%
80	4	100,00%
More	0	100,00%



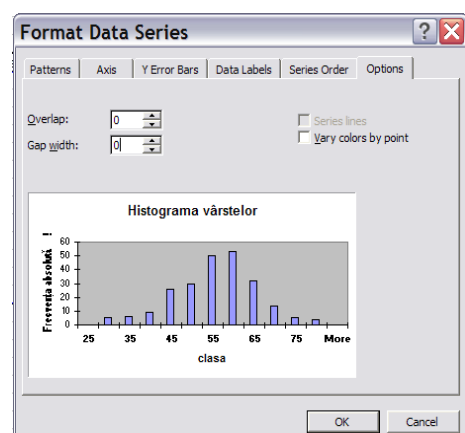
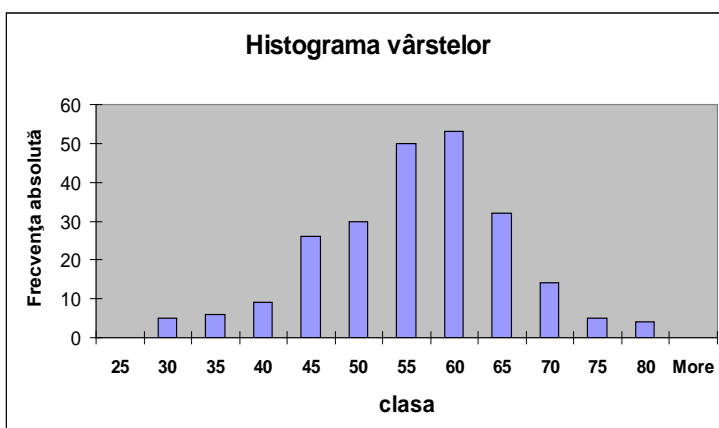
Al doilea pas este ajustarea textelor de pe suprafața graficului. Executați clic dublu pe fiecare text de pe figură, inclusiv titlul. Alegeți, pe rând, din fereastra din stânga, fontul **Arial**, stilul **Regular**, dimensiunea 10 sau 12, cu excepția titlului care este bine să aibă dimensiunea 14, sau 16.

De fiecare dată când faceți o alegere, apăsați butonul OK.

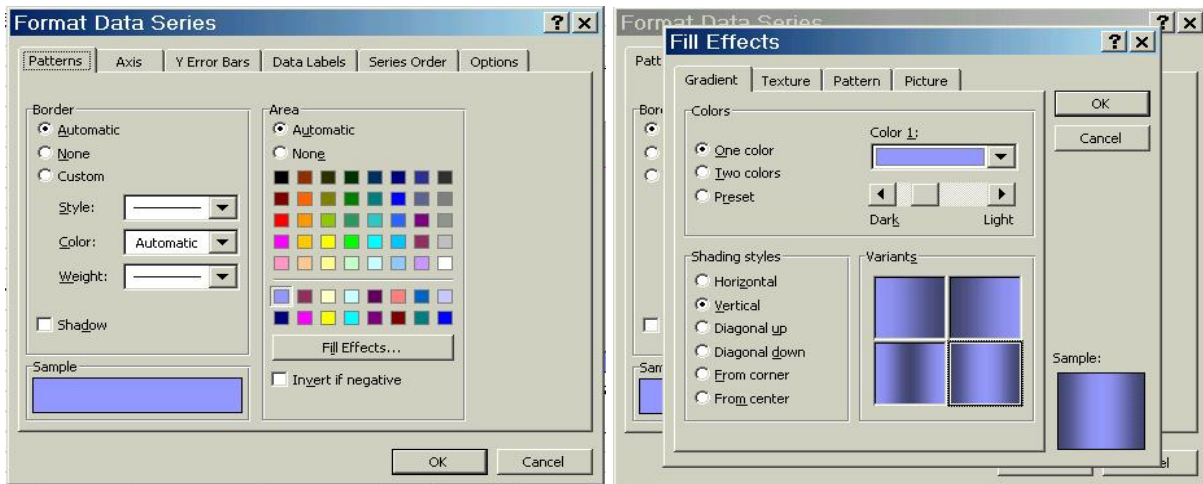
Dacă ați reușit să executați aceste operații, graficul ar trebui să arate în acest moment ca cel din figura de mai jos, stânga.

Al treilea pas este schimbarea textelor. Executați un singur clic pe cuvântul **Histogram** și apoi, cu ajutorul cursorului pus automat de program, scrieți **Histograma**

vârstelor. La fel faceți cu cuvântul **Frequency**, și scrieți **Frecvența absolută**. Apoi, executați clic pe legendă și o stergeți prin apăsarea tastei **Delete**.



Al patrulea pas este ajustarea grosimii barelor. Executați clic pe suprafața unei bare, și din fereastra **Format Data Series**, care apare, apăsați în meniu ultima opțiune din dreapta, **Options**, și alegeți în caseta **Gap width**, o valoare mai mică decât cea afișată de program (vezi figura de mai sus, dreapta). Recomandabil este să alegeți 0 (așa se obișnuiește la histogramme). Apoi apăsați butonul OK

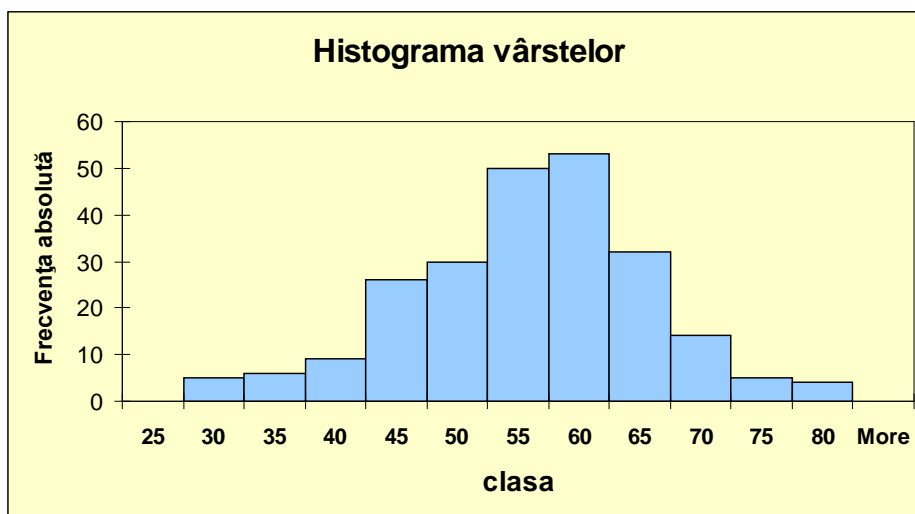


Al cincilea pas este colorarea barelor. Executați clic dublu pe suprafața uneia dintre barele graficului, și din fereastra **Format Data Series**, alegeți, dacă nu este aleasă, opțiunea **Patterns** (prima din stânga, vezi figura de mai sus, stânga). Alegeți una dintre culorile pe care vi le propune programul prin apăsarea ei cu un clic de mouse (preferabil o culoare mai vie, albastru, roșu, portocaliu, etc). Apoi, apăsați butonul **Fill Effects...** Din fereastra care apare (vezi figura de mai sus, dreapta), alegeți ca în figură, **One Color** și **Vertical**. Apoi apăsați butonul OK. Veți observa că barele graficului sunt colorate cu un degradeu (gradient de culoare), având la bază culoarea aleasă. Puteți repeta acest pas, pentru a exersa și pentru a alege cea mai frumoasă combinație. Atunci când repetați puteți schimba butoanele pe care apăsați în fereastra **Fill Effects**.

Al șaselea și ultimul pas este colorarea suprafețelor graficului. Dreptunghiul în care se află barele graficului și care are ca delimitatori axele de coordonate, o numim **zona grafic**. Marginile dincolo de acest dreptunghi, le numim **zona text**. Atât **zona grafic** cât și **zona text**, se colorează identic cu modul în care ați colorat barele. Procedați asemănător și cu aceste suprafețe care au rămas necolorate. Este indicat să alegeți culori pastel, astfel ca suprafața întregului grafic să nu aibă un aspect prea închis.

Dacă o histogramă este destinată tipăririi alb-negru, este bine să alegeți opțiuni de hașurare și nu culori cu nuanțe fine. Exersați și acest mod de umplere, de exemplu cu un clic dublu pe barele graficului și, după ce apăsați pe butonul **Fill effects**, alegeți opțiunea **Pattern**.

Pentru afișarea pe ecran însă, se pot folosi ca opțiuni de umplere atât gradientul sau texturile cât și fotografiile de fond. Astfel, cu clic dublu pe zona grafic, în fereastra **Fill Effects**, puteți alege **Texture**, sau **Picture**. Repetați, pentru a învăța aceste opțiuni. **Texture** oferă posibilitatea umplerii cu texturi de diverse materiale ca marmură, piatră, lemn, etc. Opțiunea **Picture**, oferă posibilitatea ca fondul graficului să fie o fotografie. Pentru aceasta este bine să știți unde sunt păstrate fotografiile implicite ale programelor **Microsoft Office**. De obicei, locul unde trebuie căutate este folderul firmei Microsoft din **C:\Program Files**.



Cel mai important lucru după executarea unui grafic este însă interpretarea lui, care de obicei este o problemă deosebit de simplă, căci nu se cere decât să observăm dacă barele graficului se dispun simetric față

de centru (distribuție simetrică), sau există o tendință la o asimetrie mai mult sau mai puțin accentuată (asimetrie stânga sau asimetrie dreapta). Este posibil uneori să avem histogramme cu două vârfuri, în acest caz spunem că avem o distribuție bimodală. În cazul histogrammei de mai sus, observăm că ea are un aspect de relativă simetrie, deci pacienții au vârstele repartizate asemănător unei curbe Gauss.

2. Graficul liniar de evoluție

Este unul din cele mai simple grafice atât ca realizare cât și ca interpretare. Este folosit atunci când avem un tabel cum este cel de mai jos, în care se înregistrează pe o perioadă de timp, zilnic sau la alte intervale, o mărime cum ar fi temperatura la unul sau mai mulți bolnavi, și avem nevoie de un grafic care să ne arate evoluția în timp.

În tabelul de mai jos, pe fiecare linie, sunt înregistrate temperaturile câte unui pacient, măsurate timp de două săptămâni, notate cu T1, T2,...până la T14. Pe linia 2, sunt înregistrate temperaturile pacientului B1, pe linia 3 cele ale pacientului B2, iar pe linia 4, cele ale pacientului B3. Se pot reprezenta toate cele trei linii de evoluție în același timp, chiar puteam reprezenta mai mulți pacienți, dar nu este recomandat să punem mai mult de 2-4.

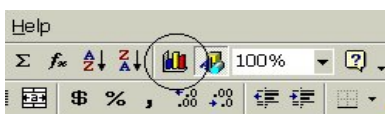
Pentru a executa un grafic liniar, primul pas este introducerea datelor.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	
2	B1	37.1	37.2	37	36.9	37	37.1	37.2	37.2	37.1	37.1	37	37	36.9	37	
3	B2	37.4	37.5	37.4	37.3	37.4	37.5	37.3	37.2	37.3	37.5	37.4	37.6	37.5	37.6	
4	B3	38.2	38.6	38.5	38.4	38.5	38.4	38.6	38.1	37.6	37.4	37.1	37.3	37	37.2	
5																
6																
7																

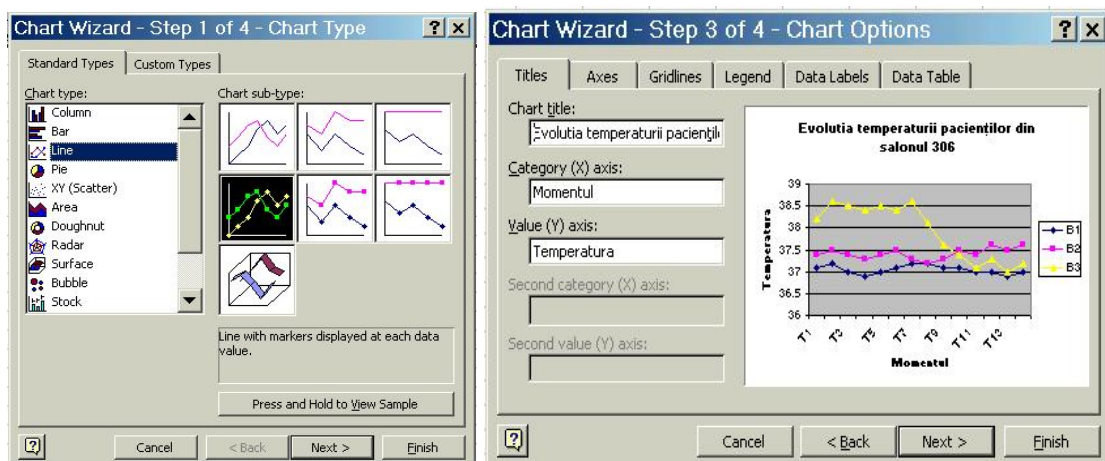
Introduceți trei linii de câte 14 temperaturi, ca în figura din stânga, simulând astfel că ați luat temperatura a trei pacienți pe parcursul a 14 zile. Nu trebuie să introduceți temperaturile din imagine, dar este bine să fie plauzibile.

Atenție, dacă programul aliniază valorile introduse pe stânga celulelor, introducerea nu este corectă. Probabil că ați inversat punctul zecimal cu virgula zecimală. **Folosiți punctul zecimal!**

Selectați începând din celula A1 și până în celula O4. Atenție, țineți minte că celula de început rămâne albă, totuși ea este selectată.



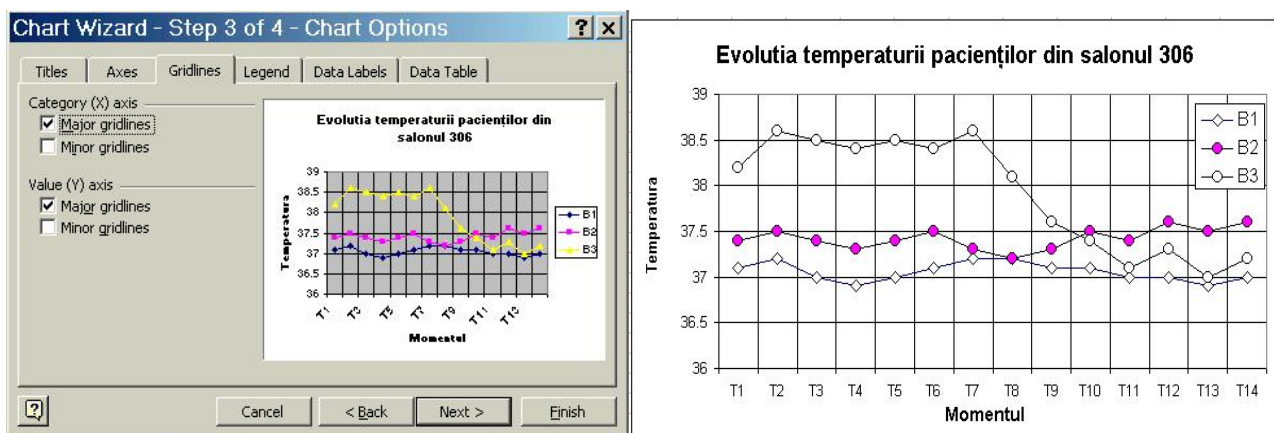
După ce ați selectat celulele cum am descris mai sus, trebuie să cereți ajutorul unui mic program care are pictograma în partea de sus a ecranului, așa cum vedeți în figura din stânga. Executați, deci, clic pe pictograma încercuită în imagine (pe ecran nu este încercuită).



Va apărea fereastra din figura de mai sus, stânga, care este primul pas în colectarea de informație despre graficul pe care dorim să îl executăm. Observați că în partea de sus a ferestrei, este scris că este primul pas din cei patru pași necesari de parcurs (Step 1 of 4).

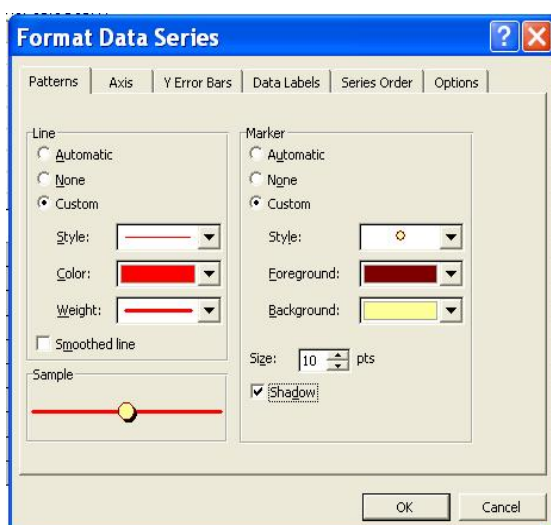
- Alegeți graficul de tip **Line** (cel selectat în figura din stânga), și apoi selectați subtipul care se vede marcat pe fond negru.
- În final apăsați pe butonul **Next**, pentru a trece la pasul al doilea. Din fereastra de la pasul al doilea nu trebuie nimic selectat sau marcat, așa încât, apăsați doar butonul **Next**, pentru a trece la pasul al treilea.
- La pasul al treilea sunt mai multe informații pe care trebuie să le transmitem programului pentru ca acesta să execute graficul. Acestea sunt împărțite pe secțiuni.
- În figura de mai sus, dreapta vedeți fereastra pentru pasul al treilea, secțiunea **Titluri (Titles)**. Completați cu textele pe care le vedeți în figură.

Atenție: La caseta de dialog **Chart title** (titlul graficului), textul este trunchiat deoarece nu încapă în casetă. Scrieți textul **“Evoluția temperaturii pacienților din salonul 306”**, sau ceva asemănător, chiar dacă textul este mai lung decât caseta, programul va reține întregul text corect. Apăsați pe secțiunea **Gridlines**, din aceeași fereastră, și lăsați validate casetele de validare din figura de mai jos, stânga.



- Apoi, apăsați butonul **Next**, pentru a trece la pasul al patrulea.
- La pasul al patrulea nu faceți nici o alegere ci numai apăsați pe butonul **Finish**.
- Efectul este apariția graficului din figura din stânga (de fapt unul asemănător, care trebuie prelucrat și adus la acest aspect sau unul similar).

Pentru prelucrare, procedați așa cum ați procedat la histogramă. Ajustați dimensiunea graficului astfel încât să fie cât mai mare, dar să nu depășească limitele ecranului, micșorați textele, în afară de titlu, care este bine să rămână așa cum este, micșorați cifrele cu care este făcută numerotarea de pe axe, schimbați fondurile, etc.



Un aspect deosebit de plăcut al acestui grafic, rezultă dacă urmați următorii pași:

Puneți vârful pointerului de mouse, chiar pe una din cele trei linii, și executați dublu clic. Va apare fereastra din stânga, pe care alegeți astfel: La caseta de dialog **Color**, alegeți roșu aprins, sau o altă culoare tare. La caseta **Weight**, alegeți cel mai gros model de linie. În dreapta, la **Style**, alegeți cerculețul. La caseta **Foreground**, alegeți culoarea maro (pentru a afi potrivită cu roșu), iar la caseta **Background**, alegeți galben, care se potrivește cu roșu și maro. La caseta **Size**, alegeți 9, sau 10 sau chiar 11, iar la **Shadow**, lăsați bifat. Apoi apăsați OK. Graficul are acum una din linii mult schimbată ca aspect. Procedați analog cu celelalte linii, alegând evident alte culori sau markeri.

3. Graficul cu bare

Este cel mai indicat grafic în cazul tabelelor de incidență. Pentru a face un exercițiu, procedați astfel: introduceți date ca cele din figură, care clasifică pacienții după sex și grupa de vârstă.

The screenshot shows the Microsoft Excel interface with a data table and two dialog boxes for creating a chart.

	< 30	30 - 40	40 - 50	50 - 60	> 60
Femei	14	47	86	64	26
Bărbați	25	64	124	98	36

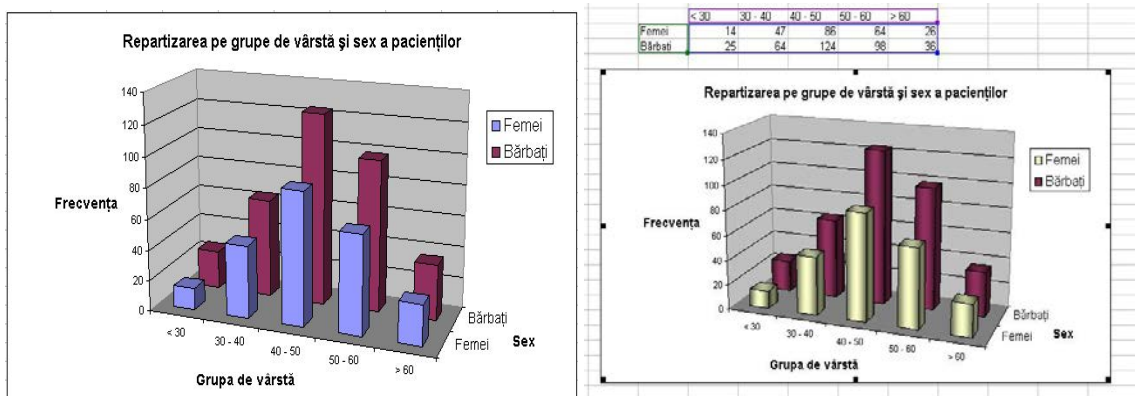
The **Chart Wizard - Step 1 of 4 - Chart Type** dialog box shows the **Column** chart type selected. The **Chart Wizard - Step 3 of 4 - Chart Options** dialog box shows the chart title "Repartizarea pe grupe de vârstă și sex a pacienților", the category (X) axis "Grupa de vârstă", the series (Y) axis "Sex", and the value (Z) axis "Frecvența".

Acest tip de grafic reprezintă prin bare verticale frecvențele absolute din celulele unui tabel de incidență (numărul de pacienți înscris în fiecare celulă). Are exact atâtea rânduri de linii și coloane de bare, câte linii și coloane are tabelul.

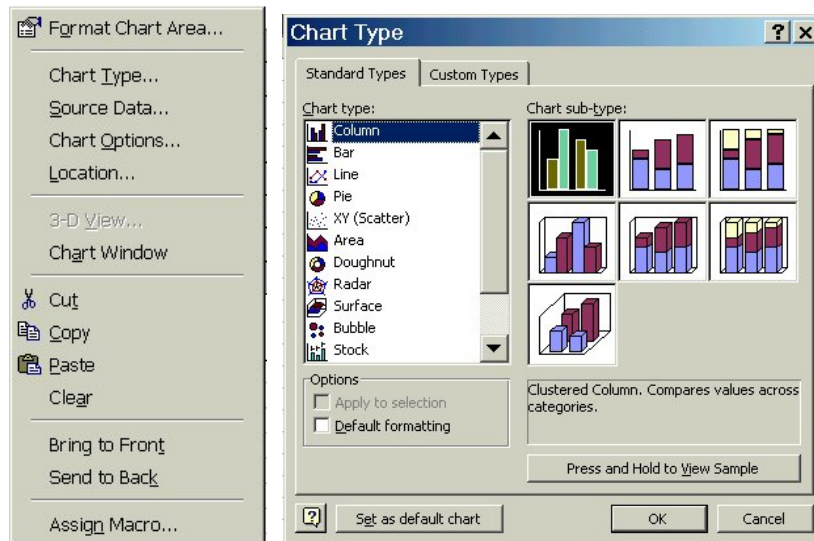
Selectați tabelul așa cum vedeți în figură. (Atenție, este foarte important să faceți **selecția exact ca în figură**, adică să nu selectați nici o linie sau coloană în plus sau în minus. În figură este selectată zona **de la C2 la H4**). Apoi, executați clic pe pictograma asistentului pentru grafice, care afișază fereastra de mai sus, din stânga.

La primul pas, alegeți tipul de grafic **Column** (grafic cu coloane), și subtipul care se vede că este marcat pe fond negru. Apoi, apăsați butonul **Next**. În fereastra de la pasul doi nu faceți nici o opțiune ci doar apăsați butonul **Next**. Efectul, este apariția ferestrei pentru pasul al treilea, care se vede în figura de mai sus, din dreapta. Scrieți titlurile pe care le vedeți în figură (Atenție, titlul graficului este mai lung decât cuprinde caseta de dialog, dar îl scrieți pe cel din graficul pus ca probă). Apoi, apăsați butonul **Next**.

În fereastra de la pasul al patrulea, nu faceți nici o opțiune ci doar apăsați butonul **Finish**. Rezultatul este apariția unui grafic pe care îl mai ajustați ca mărime trăgând de marginile lui, micșorați textele, pentru a arăta ca în una din figurile de mai jos.



În cazul în care tipul graficului nu vă satisface, el poate fi schimbat și după ce a fost realizat.



Executați un clic dreapta de mouse pe suprafața liberă a graficului, adică acolo unde este culoare albă sau culoarea aleasă pentru fond. Din lista care apare (vezi mai sus, în stânga), alegeți opțiunea **Chart Type**, și din fereastra pe care ați mai folosit-o, optați așa cum vedeți în figură, apoi apăsați butonul OK.

Acum graficul are atâtea grupe de coloane câte coloane a avut tabelul și arată cu totul altfel. Procedați ca la histogramă pentru a realiza un grafic estetic și în acest caz.

4. Exerciții și chestiuni de examen

2. Executați o histogramă pentru valorile de pe coloana BRT (bilirubina totală). Când calculați minimul și maximul, veți constata că trebuie să luați lungimea unei clase de aproximativ 1 sau cel mult 2. Recomandabil este 1. De aceea, limitele claselor le veți pune din 1 în 1 (sau din 2 în 2 dacă ați optat pentru lungimea 2). Veți începe cu 0, apoi 1, apoi 2 și așa mai departe până la o valoare imediat superioară maximului. Astfel ați cerut programului să calculeze câți pacienți avem cu valoarea BRT între 0 și 1, câți avem cu valoarea BRT între 1 și 2, și așa mai departe până la ultima clasă. Urmați pașii de la exemplul de la începutul acestui material.

3. Executați o histogramă cu valorile Hemoglobinei (coloana Hb). Și de data aceasta, este recomandabil ca lungimea unei clase să fie 1, deoarece valoarea minimă este în jurul lui 7, iar valoarea maximă în jurul lui 16. Pe o foaie de hârtie, precizați următoarele: valoarea minimă, valoarea maximă, lungimea unei clase, precum și aspectul histogramei, adică dacă distribuția este simetrică sau asimetrică.

4. Deschideți tabelul SHR. În acest tabel sunt înregistrați 240 de pacienți cu hepatite, sindrom hepato-renal, ascită refractară și alte afecțiuni (vezi pe coloana LOT). Executați o histogramă cu valorile greutateilor celor 240 de pacienți (coloana **Greut1**). Apreciați dacă histograma are un aspect simetric sau avem un caz de distribuție asimetrică.

5. Ca și la exercițiul 3, executați histograme pentru coloanele pe care sunt înregistrate valorile taliei (Talie), tensiunea arterială sistolică (TAM1), hemoglobina (HbSeric1), calcemia, (CaSeric1), potasemia (KSeric1), sodiul seric (NaSeric1), creatinina (Creat1).

6. Deschideți tabelul ASTM. În acest tabel sunt înregistrați 240 de pacienți cu astm bronșic. Executați histograme cu valorile greutateilor celor 240 de pacienți (coloana **Greut**). Apreciați dacă histograma are un aspect simetric sau avem un caz de distribuție. La fel pentru talie, vârstă, scorul Borg, scorul 6MWD, (coloanele **Inălțime**, **Varsta**, **Borgi**, **6MWDi**).

7. Folosind o histogramă, apreciați dacă duratele de supraviețuire la pacienții din tabelul LIMFOM sunt distribuite simetric.