

Recapitulare - Statistică descriptivă și corelații

Obiective:

- Recapitulare statistică descriptivă și corelații
- Funcția IF
- Funcția CountIF

Problema 1. La centul de Boli de Nutriție și Diabet au fost luați în studiu 300 de pacienți.

Pentru acești pacienți au fost înregistrați următorii parametri biologici: Vârsta, Greutate, IMC, Colesterol, Trigliceride, HDL-colesterol, TAS, TAD, Fumat, Gen, Circumferința abdominală. Datele se găsesc în fișierul Nutritie.xls.

Realizați:

- Știind că valorile normale pentru Trigliceride sunt valorile sub 135mg/dl, introduceți în coloana TG calit astfel: dacă Trigliceride < 135 atunci TG calit va avea valoarea “normale”, altfel TG calit va avea valoarea “crescute” (cu ajutorul funcției IF).
- Știind că valorile normale pentru HDL-colesterol sunt valorile peste 45mg/dl, introduceți în coloana HDL calit astfel: dacă HDL-colesterol > 45 atunci HDL calit va avea valoarea “normal”, altfel HDL calit va avea valoarea “scazut” (cu ajutorul funcției IF).
- Știind că valorile normale pentru Colesterol sunt valorile sub 200mg/dl, introduceți în coloana Colest calit astfel: dacă Colesterol < 200 atunci Colest calit va avea valoarea “normal”, altfel Colest calit va avea valoarea “crescut” (cu ajutorul funcției IF).
- Știind că o persoană are hipertensiune dacă are TAS > 140 mmHg sau dacă are TAD > 90 mmHg, introduceți în coloana Hipertensiune astfel: dacă TAS > 140 sau TAD > 90 atunci Hipertensiune va avea valoarea “Da”, altfel Hipertensiune va avea valoarea “Nu” (cu ajutorul funcției IF).
- Știind că o persoană este clasificată cu greutate normală dacă are IMC < 25 , este clasificată supraponderală dacă are IMC ≥ 25 și IMC < 30 , este clasificată obeză dacă are IMC ≥ 30 , introduceți în coloana IMC calit astfel: dacă IMC < 25 atunci IMC calit va avea valoarea “normoponderal”, dacă IMC ≥ 25 și IMC < 30 atunci IMC calit va avea valoarea “Supraponderal”, altfel IMC calit va avea valoarea “Obez” (cu ajutorul funcției IF).
- Copiați întregul tabel pe o altă pagină și ordonați pacienții după IMC calit ca primă cheie de sortare și după Colest calit ca a doua cheie de sortare.
- Reprezentați grafic distribuția pacienților după Sex

- h) Reprezentați grafic distribuția Fumatului în funcție de Sex.
- i) Calculați indicatorii statistici descriptivi (media aritmetică, mediana, abaterea standard, boltirea, asimetria, minim, maxim, amplitudinea, eroarea standard) pentru TAS, TAD și Trigliceride cu ajutorul pachetului Descriptive statistics din Data-Analysis.
- j) Calculați indicatorii statistici descriptivi (media aritmetică, cuartilele, abaterea standard, coeficientul de variație, eroarea standard, amplitudinea, valoarea centrală) pentru HDL-colesterol și pentru Colesterol cu ajutorul funcțiilor și formulelor.
- k) Reprezentați grafic IMC sub formă de Histogramă cu intervalele de frecvență sub 20, (20,25], (25,30], (30,35], (35,40], peste 40
- l) Reprezentați grafic Colesterolul sub formă de Histogramă cu intervalele de frecvență alese de dvs. în mod adecvat.
- m) Calculați matricea de corelație pentru Circumferința abdominală, TAS, TAD, Trigliceride, HDL-colesterol și Colesterol.
- n) Alcătuiți tabelul de frecvență pentru TG calit, HDL calit și Colest calit cu ajutorul funcției CounIf.
- o) Alcătuiți tabelul de frecvență pentru Hipertensiune și IMC calit cu ajutorul Pivot Table.
- p) Reprezentați grafic dependența dintre Circumferința abdominală și TAS.
- q) Reprezentați grafic dependența dintre Circumferința abdominală și TAD.
- r) Reprezentați grafic dependența dintre Varstă și Colesterol.
- s) Reprezentați grafic dependența dintre Varstă și HDL-colesterol.
- t) Reprezentați grafic dependența dintre Varstă și Trigliceride.
- u) Realizați o prezentare Power Point cu rezultatele și interpretările lor.

Instrucțiuni

- a) Știind că valorile normale pentru Trigliceride sunt valorile sub 135mg/dl, introduceți în coloana TG calit astfel: dacă Trigliceride < 135 atunci TG calit va avea valoarea “normale”, altfel TG calit va avea valoarea “crescute” (cu ajutorul funcției IF).
 1. Selectați celula M2
 2. Introduceți funcția IF știind că:

Funcția IF (Intoarce o valoare dacă condiția specificată este adevărată și o altă valoare dacă condiția specificată este falsă.)

=IF(conditie;valoarea_daca_adevarat;valoarea_daca_fals)

sau

=IF(conditie,valoarea_daca_adevarat,valoarea_daca_fals)

conditie orice valoare sau expresie care poate fi evaluată cu adevărat sau fals. De exemplu $I2 < 135$. Expresia este adevărată dacă în I2 avem o valoare mai mică decât 135, altfel este falsă.

valoarea_daca_adevarat este valoarea (rezultatul) întoarsă de funcția IF dacă condiția este adevărată. De exemplu “normale” sau 1.

valoarea_daca_fals este valoarea (rezultatul) întoarsă de funcția IF dacă condiția este falsă. De exemplu “crescute” sau 0.

D	E	F	G	H	I	J	K	L	M	N	O	
Greutate	IMC	Circumferinta Abdominala	TAS	TAD	Trigliceride	Hdl-colesterol	Colesterol	Fumat	TG calit	HDL calit	Colest calit	Hi
58	22	95	120	70	124	64	225	Nu	=IF(I2<135;"normale";"crescute")			
76	31	99	99	66	380	31	225	Nu				

sau

D	E	F	G	H	I	J	K	L	M	N	O	
Greutate	IMC	Circumferinta Abdominala	TAS	TAD	Trigliceride	Hdl-colesterol	Colesterol	Fumat	TG calit	HDL calit	Colest calit	Hi
58	22	95	120	70	124	64	225	Nu	=IF(I2<135;"normale";"crescute")			
76	31	99	99	66	380	31	225	Nu	=IF(logical_test; [value_if_true]; [value_if_false])			
104	34	119	129	85	314	48	284	Nu				

b) Știind că valorile normale pentru HDL-colesterol sunt valorile peste 45mg/dl, introduceți în coloana HDL calit astfel: dacă HDL-colesterol > 45 atunci HDL calit va avea valoarea “normal”, altfel HDL calit va avea valoarea “scazut” (cu ajutorul funcției IF).

Instrucțiuni: analog punctul a

c) Știind că valorile normale pentru Colesterol sunt valorile sub 200mg/dl, introduceți în coloana Colest calit astfel: dacă Colesterol < 200 atunci Colest calit va avea valoarea “normal”, altfel Colest calit va avea valoarea “crescut” (cu ajutorul funcției IF).

Instrucțiuni: analog punctul a

d) Știind că o persoană are hipertensiune dacă are TAS >140 mmHg sau dacă are TAD >90 mmHg, introduceți în coloana Hipertensiune astfel: dacă TAS > 140 sau TAD >90 atunci Hipertensiune va avea valoarea “Da”, altfel Hipertensiune va avea valoarea “Nu” (cu ajutorul funcției IF).

Instrucțiuni:

Funcția OR (întoarce fals dacă toate condițiile sunt evaluate cu fals, în celelalte cazuri funcția întoarce valoarea adevărat):

`OR(conditie1;conditie2; ...)`

sau

`OR(conditie1,conditie2, ...)`

`conditie1,conditie2, ...` sunt de la 1 la 30 de condiții care pot fi evaluate fiecare cu adevărat sau fals.

Funcția AND (întoarce adevărat dacă toate condițiile sunt evaluate cu adevărat, în celelalte cazuri funcția întoarce valoarea fals):

`AND(conditie1;conditie2; ...)`

sau

`AND(conditie1,conditie2, ...)`

`conditie1,conditie2, ...` sunt de la 1 la 30 de condiții care pot fi evaluate fiecare cu adevărat sau fals.

Cholesterol	Fumat	TG calit	HDL calit	Colest calit	Hipertensiune	IMC calit
225	Nu	normale	normal	crescut	=IF(OR(G2>140;H2>90);"Da";"Nu")	
190	Nu	crescute	crescut	normal		

- e) Știind că o persoană este clasificată cu greutate normală dacă are $IMC < 25$, este clasificată supraponderală dacă are $IMC \geq 25$ și $IMC < 30$, este clasificată obeză dacă are $IMC \geq 30$, introduceți în coloana IMC calit astfel: dacă $IMC < 25$ atunci IMC calit va avea valoarea “normoponderal”, dacă $IMC \geq 25$ și $IMC < 30$ atunci IMC calit va avea valoarea “Supraponderal”, altfel IMC calit va avea valoarea “Obez” (cu ajutorul funcției IF).

Instrucțiuni:

Datorită faptului că avem 3 criterii vom folosi două funcții IF imbricate:

Colest calit	Hipertensiune	IMC calit
crescut	Nu	=IF(E2<25;"normoponderal";IF(E2<30;"supraponderal";"obez"))
normal	Nu	

f) Copiați întregul tabel pe o altă pagină și ordonați pacienții după IMC ca primă cheie de sortare și după Colest ca a doua cheie de sortare.

Instrucțiuni: Copiați întregul tabel pe o altă pagină în Excel. Din meniul Data alegeți opțiunea Sort. Alegeți IMC ca primă și Colest ca a doua cheie de sortare:

	A	B	C	D	E	F	G	H	
1	ID	Varsta	Sex	Greutate	IMC	Circumferinta Abdominala	TAS	TAD	Trigl
2	1	55	f	58	22	95	120	70	
3	2	79	f	76	31	99	99	66	
4	3	51	m	104	34	119	129	85	
5	4	54	f						93
6	5	53	r						65
7	6	52	f						77
8	7	65	f						90
9	8	56	r						01
10	9	64	r						75
11	10	56	f						55
12	11	68	r						79
13	12	76	f						92
14	13	68	r						76
15	14	54	r						86
16	15	63	r						96
17	16	42	r						82
18	17	54	r						88
19	18	51	f						78
20	19	67	f						88
21	20	63	r						16
22	21	74	f						70

Se poate sorta crescător sau descrescător (opțiunile Ascending sau Descending).

Reveniți în Sheet 1.

g) Reprezentați grafic distribuția pacienților după Sex

Instrucțiuni: în laborator 4 punctul a) grafic PIE

h) Reprezentați grafic distribuția Fumatului în funcție de Sex.

Instrucțiuni: în laborator 4 punctul b) grafic BAR în procente

i) Calculați indicatorii statistici descriptivi (media aritmetică, mediana, abaterea standard, boltirea, asimetria, minim, maxim, amplitudinea, eroarea standard) pentru TAS, TAD și Trigliceride cu ajutorul pachetului Descriptive statistics din Data-Analysis.

Instrucțiuni: în laborator 5 punctul 4

j) Calculați indicatorii statistici descriptivi (media aritmetică, quartilele, abaterea standard, coeficientul de variație, eroarea standard, amplitudinea, valoarea centrală) pentru HDL-colesterol și pentru Colesterol cu ajutorul funcțiilor și formulelor.

Instrucțiuni: în laborator 5 punctul 3 pentru calculul mediei, quartilelor (toate cinci: minim, quartila 1, mediana, quartila 3, maxim) și abaterii standard

Coeficientul de variație se calculează cu formula: $CV = \frac{s}{\bar{X}}$, unde s este abaterea standard, iar \bar{X} este media aritmetică

Eroarea standard se calculează cu formula: $E_s = \frac{s}{\sqrt{n}}$

Amplitudinea se calculează cu formula: Amplitudinea = $X_{\max} - X_{\min}$

Valoarea centrală se calculează cu formula: Valoarea centrală = $\frac{X_{\min} + X_{\max}}{2}$, unde X_{\min} este minimul, iar X_{\max} este maximul.

k) Reprezentați grafic IMC sub formă de Histogramă cu intervalele de frecvență sub 20, (20,25], (25,30], (30,35], (35,40], peste 40

Instrucțiuni: în Laborator 4 punctul d) grafic Histogramă. Pentru acest caz Binary range va fi șirul: 20, 25, 30, 35, 40

l) Reprezentați grafic Colesterolul sub formă de Histogramă cu intervalele de frecvență alese de dvs. în mod adecvat.

Instrucțiuni:

La punctul i) ati calculat minumul 110, maximul 520 și amplitudinea 410 pentru colesterol.

Lungimea intervalului de clasă pentru histogramă se calculează cu formula:

$$\text{Amplitudinea}/(\text{Nr.de intervale}-1)$$

Numărul de intervale se alege între 4 și 10, în funcție de talia eșantionului sau de preferințe.

In cazul nostru dacă alegem 6 intervale, lungimea intervalului de clasă va fi 82.

Introducem Intervalele de clasă astfel: Minim pentru prima valoare,

Minim+Lungimea Intervalului pentru a doua valoare. Apoi adaugăm Lungimea intervalului la valoarea precedentă de mai multe ori cât timp valoarea este < maxim. (Se poate executa Fill-Down).

	N	O	P	Q	R	S	T	U	V
1	HDL calit	Colest calit	Hipertensiune	IMC calit					
2	normal	crescut	Nu	normoponderal					
3	crescut	normal	Nu	obez					
4	normal	crescut	Nu	obez			minim	110	
5	normal	normal	Da	supraponderal			maxim	520	
6	crescut	normal	Nu	supraponderal			amplitudinea	410	
7	crescut	crescut	Nu	normoponderal			Lungimea intervalului	82	
8	normal	crescut	Da	normoponderal					
9	normal	crescut	Da	normoponderal					
10	crescut	crescut	Nu	supraponderal			Intervalele de clasa		
11	crescut	crescut	Da	normoponderal				110	
12	crescut	normal	Nu	obez			=U4+U7		
13	normal	normal	Da	supraponderal				274	
14	crescut	normal	Nu	normoponderal				356	
15	crescut	normal	Da	normoponderal					

	N	O	P	Q	R	S	T	U	V
1	HDL calit	Colest calit	Hipertensiune	IMC calit					
2	normal	crescut	Nu	normoponderal					
3	crescut	normal	Nu	obez					
4	normal	crescut	Nu	obez			minim	110	
5	normal	normal	Da	supraponderal			maxim	520	
6	crescut	normal	Nu	supraponderal			amplitudinea	410	
7	crescut	crescut	Nu	normoponderal			Lungimea intervalului	82	
8	normal	crescut	Da	normoponderal					
9	normal	crescut	Da	normoponderal					
10	crescut	crescut	Nu	supraponderal			Intervalele de clasa		
11	crescut	crescut	Da	normoponderal				110	
12	crescut	normal	Nu	obez				192	
13	normal	normal	Da	supraponderal			=T12+\$U\$7		
14	crescut	normal	Nu	normoponderal				356	
15	crescut	normal	Da	normoponderal					

Se continuă cu executarea graficului Histogramă analog cu punctul anterior.

m) Calculați matricea de corelație pentru Circumferința abdominală, TAS, TAD, Trigliceride, HDL-colesterol și Colesterol.

Instrucțiuni: în Laborator 6 punctul d)

n) Alcătuiți tabelul de frecvență pentru TG calit, HDL calit și Colest calit cu ajutorul funcției CounIf.

Instrucțiuni:

Executați tabelul următor:

Trigliceride	Frecventa
Normale	
Crescute	
Total	

In celula unde ar trebui să fie numărul de pacienți cu Trigliceride cu valori normale construiți sintaxa funcției CountIF:

Funcția CountIF (numără câte celule din domeniul dat îndeplinesc un criteriu):

=COUNTIF(domeniu;criteriu)

sau

=COUNTIF(domeniu,criteriu)

Înlocuiți **domeniu** cu domeniul pe care ați calculat TG calit (fără titlul variabilei, numai valorile) ex: M2:M20

Criteriu este o condiție sub forma unui număr, expresie, referință, sau text între ghilimele, de exemplu 135, >=135, B4 sau "normal". În cazul nostru înlocuiți criteriu cu "normale".

fx =COUNTIF(M2:M301;"normale")							
P	Q	R	S	T	U	V	W
	obez						
	supraponderal						
	obez			Trigliceride	Frecventa		
	supraponderal			Normale	=COUNTIF(M2:M301;"normale")		
	supraponderal			Crescute	COUNTIF(range; criteria)		
	supraponderal			Total			

sau

fx =COUNTIF(M2:M301;"normale")							
P	Q	R	S	T	U	V	W
	obez						
	supraponderal						
	obez			Trigliceride	Frecventa		
	supraponderal			Normale	=COUNTIF(M2:M301;"normale")		
	supraponderal			Crescute	COUNTIF(range; criteria)		
	supraponderal			Total			

Pentru a număra valorile crescute în sintaxa de mai sus a funcției IF se înlocuiește "normale" cu "crescute".

În dreptul Totalului se calculează totalul cu funcția SUM. Numărul total trebuie să fie egal cu numărul de pacienți.

În mod analog se calculează tabelul de frecvență pentru HDL calit și Colest calit.

o) Alcătuiți tabelul de frecvență pentru Hipertensiune și IMC calit cu ajutorul Pivot Table.

Instrucțiuni: în Laborator 4 punctul a)

p) Reprezentați grafic dependența dintre Circumferința abdominală și TAS.

Instrucțiuni: în Laborator 4 punctul c)

q) Reprezentați grafic dependența dintre Circumferința abdominală și TAD.

Instrucțiuni: analog punctul o)

r) Reprezentați grafic dependența dintre Varstă și Colesterol.

Instrucțiuni: analog punctul o)

s) Reprezentați grafic dependența dintre Varstă și HDL-colesterol.

Instrucțiuni: analog punctul o)

t) Reprezentați grafic dependența dintre Varstă și Trigliceride.

Instrucțiuni: analog punctul o)

u) Realizați o prezentare Power Point cu rezultatele obținute la subpunctele de mai sus și interpretările lor.