

II.A

PRELUCRĂRI DE DATE MEDICALE. MICROSOFT EXCEL

Obiectivul urmărit în acest material este analiza datelor medicale cu ajutorul funcțiilor statistice cu opțiunea *Insert Function* și a pachetului *Data Analysis*, disponibile în Microsoft Excel.

1. Statistică cu Insert Function (TTEST și CORREL)

(Insert Function-statistical-TTEST sau CORREL)



Problema 1

Utilizând datele de pacienți din fișierul tabel_studiu.xls (Anexă), se pune întrebarea dacă există diferențe semnificative a colesterolului recoltat comparat cu colesterolul la 6 luni, a celor 30 de pacienți investigați.



Soluție

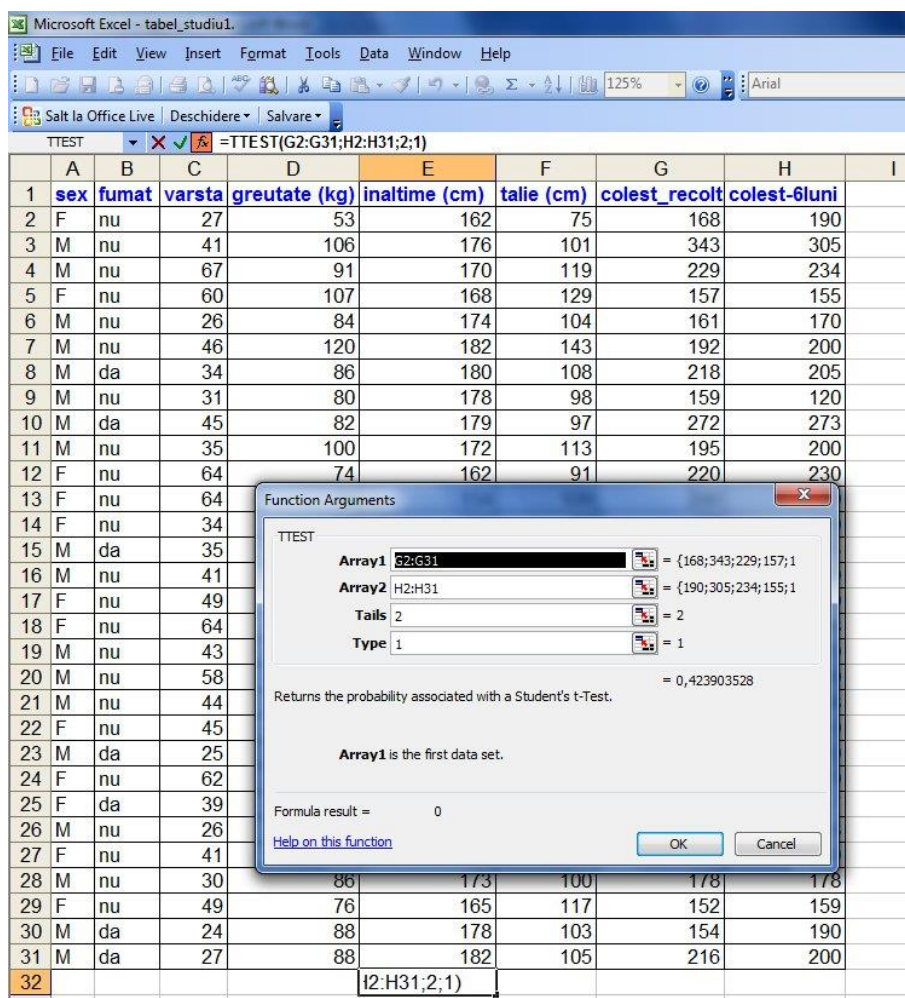


Figura 1. Pacienții și Insert function -TTEST

Observăm că $p=0.423903528>0.05$, deci *diferențele sunt nesemnificative* între colesterolul recoltat comparat cu colesterolul la 6 luni (tails =2, type =1, t pereche).



Problema 2

Utilizând aceleași date de pacienți, tabel_studiu.xls, se pune întrebarea dacă există diferențe semnificative între colesterolul recoltat dintre femei și bărbați.

**Soluție**

Microsoft Excel - tabel_studiu

File Edit View Insert Format Tools Data Window Help

Salt la Office Live Deschidere Salva

TTEST =TTEST(G2:G19;G20:G31;2;3)

	A	B	C	D	E	F	G	H	I
1		1 fumat	varsta	greutate (kg)	inaltime (cm)	talie (cm)	colest_recolt	colest-6luni	
2	M	nu	27	53	162	75	168	190	
3	M	nu	41	106	176	101	343	305	
4	M	nu	67	91	170	119	229	234	
5	M	nu	60	107	168	129	157	155	
6	M	nu	26	84	174	104	161	170	
7	M	nu	46	120	182	143	192	200	
8	M	da	34	86	180	108	218	205	
9	M	nu	31	80	178	98	159	120	
10	M	da	45	82	179	97	272	273	
11	M	nu	35	100	172	113	195	200	
12	M	nu	64	74	162	91	220	230	
13	M	nu	64	78	154	331;2;3)	246	270	
14	M	nu	34	55	152	90	147	130	
15	M	da	35	57	173	88	157	145	
16	M	nu	41						
17	M	nu	49						
18	M	nu	64						
19	M	nu	43						
20	F	nu	58						
21	F	nu	44						
22	F	nu	45						
23	F	da	25						
24	F	nu	62						
25	F	da	39						
26	F	nu	26						
27	F	nu	41						
28	F	nu	30						
29	F	nu	49						
30	F	da	24						
31	F	da	27	88	182	105	216	200	
32									

Function Arguments

TTEST

Array1 G2:G19 = {168;343;229;157;1

Array2 G20:G31 = {245;242;162;215;2

Tails 2 = 2

Type 3 = 3

= 0,887309553

Returns the probability associated with a Student's t-Test.

Type is the kind of t-test: paired = 1, two-sample equal variance (homoscedastic) = 2, two-sample unequal variance = 3.

Formula result = 1

Help on this function

OK Cancel

Figura 2. Pacienții și Insert function –TTEST

Observăm că $p=0.887309553>0.05$, deci *diferențele* sunt *nesemnificative* între colesterolul recoltat comparat între femei și bărbați (tails =2, type =3, t nepereche, varianțe diferite).

**Problema 3**

Utilizând datele de pacienți din fișierul tabel_studiu.xls, să se stabilească dacă există o legătură între variabila înălțime și variabila greutate a pacienților,

reprezentând și grafic această legătură în Excel, exportând graficul într-un fișier word.



Indicații. Soluții

În acest sens calculăm coeficientul de corelație liniară al lui Pearson, prin aplicarea funcției statistice *CORREL* și graficul care este de tip *SCATTER*.

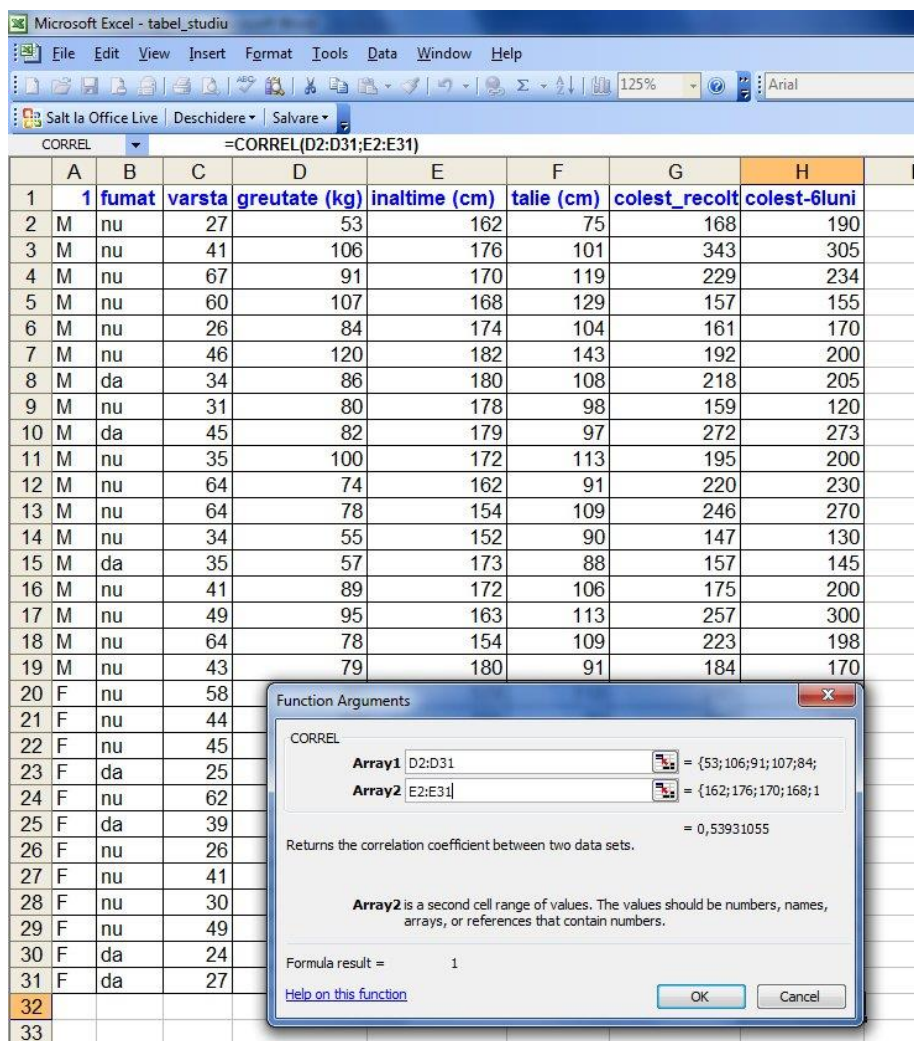


Figura 3. Pacienții și Insert Function-Correl

Observăm că se obține, $r=0.53931055$, deci există o corelație *directă*, *moderată* între înălțime și greutatea pacienților.

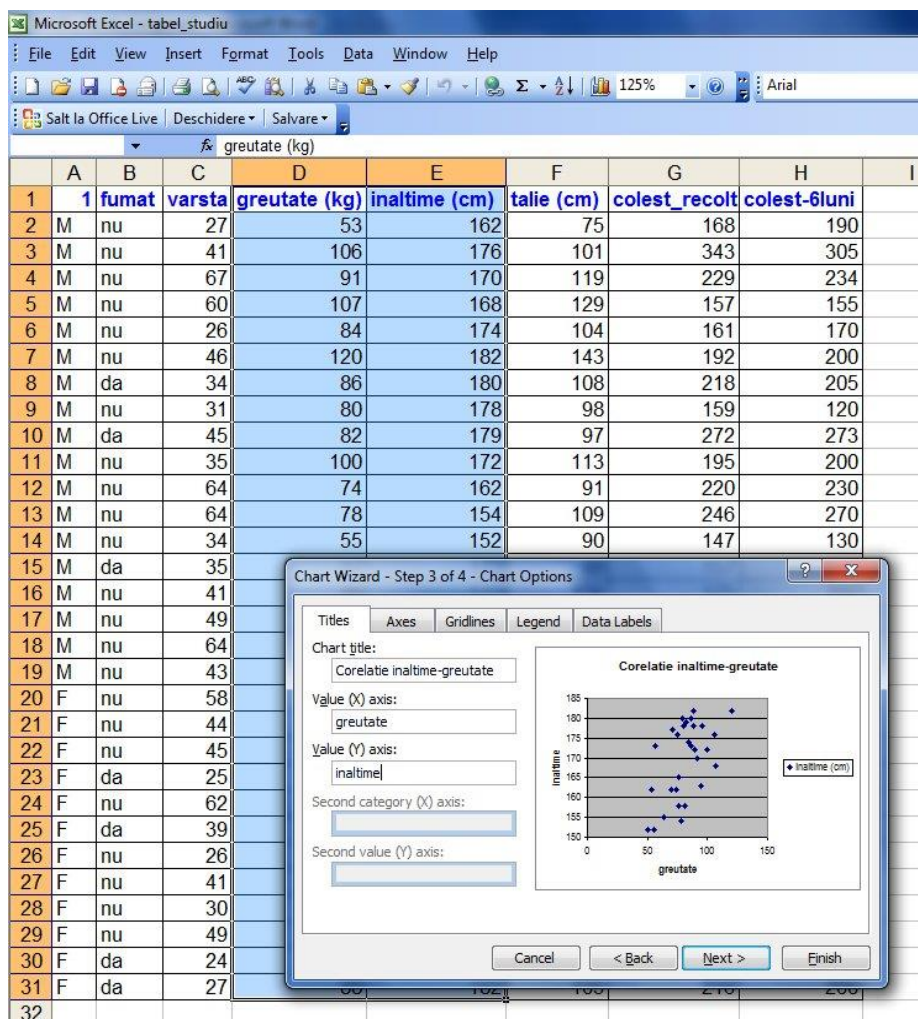


Figura 4. Pacienții și Insert Chart - Scatter

Finish, apoi clic dreapta mouse pe un punct *Add trend line* – la *Option* selectăm ecuația drepte și R^2 . Se obține fig. 5, dar în Excel, după care o trecem într-un fișier word.

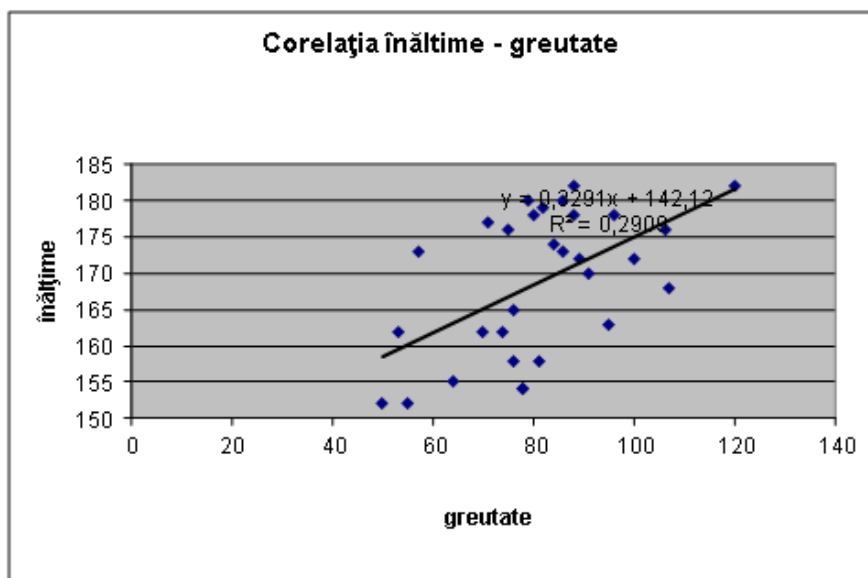


Figura 5. Pacienții și Insert Chart – Scatter-dreapta de regresie

2. Statistica cu Data Analysis (Descriptive Statistics, Histogram, t-Test, Correlation)



Problema 4 (Calcule statistice cu Data Analysis)

Realizați baza de date în MS Excel, DataAnalysis.xls:

Nr. crt.	Sex	Vârstă	Glicemie – mom. 0	Glicemie – 3 luni	Glicemie – 6 luni	Asigurat CAS
1	M	35	132	125	119	T
2	M	24	133	133	106	T
3	F	22	118	113	109	F
4	F	53	111	107	97	T
5	M	20	112	101	95	T
6	F	37	130	122	97	T
7	F	29	125	115	105	T
8	M	56	138	121	109	F
9	F	27	129	110	108	F
10	F	21	110	115	105	F
11	F	36	113	111	99	T
12	M	38	145	133	129	F

13	F	29	138	132	130	T
14	M	44	132	130	131	F
15	M	42	130	126	128	T
16	F	32	125	120	105	F
17	F	25	124	124	116	T
18	F	29	122	111	107	F
19	M	39	127	114	108	F
20	M	40	118	108	108	T
21	M	60	121	115	98	T
22	F	58	136	130	128	T
23	M	56	118	110	105	T
24	M	55	116	109	99	T
25	F	45	135	128	126	T
26	M	30	120	113	107	F
27	M	18	127	120	102	T
28	F	31	129	118	105	T
29	F	24	126	115	103	T
30	M	22	132	116	96	T
31	M	53	133	124	116	F
32	M	20	131	122	113	T
33	M	37	139	123	111	F
34	M	29	133	121	108	T
35	F	56	128	121	109	F
36	F	27	127	126	116	T
37	M	21	124	122	113	T
38	M	36	123	119	109	T
39	M	38	145	135	135	T
40	M	29	131	130	132	T
41	M	44	130	128	118	T
42	M	42	124	120	119	T
43	M	32	119	108	106	F
44	F	25	129	115	114	T
45	M	29	128	120	109	F
46	M	39	127	125	108	T
47	F	40	129	126	112	T
48	M	24	132	128	113	T
49	M	22	127	120	117	T
50	M	53	128	115	116	T

Figura 6. DataAnalysis.xls

- Calculați parametrii de statistică descriptivă pentru variabilele *Vârstă*, *Glicemie – mom. 0*, *Glicemie – 3 luni*, *Glicemie – 6 luni*.



Indicații. Soluții

Se alege din meniul superior *Data / Data Analysis / Descriptive Statistics*:

Se introduce șirul de celule pentru care dorim să calculăm parametrii de statistică descriptivă.

Se bifează atunci când prima celulă din coloana / șirul selectat conține numele variabilei pentru care se realizează analiza.

Se specifică poziția în foaia de lucru curentă unde dorim să se afișeze rezultatele.

Eventual, rezultatele se pot afișa pe o foaie de lucru separată, cu numele ales de noi, în același registru de calcul.

Se indică parametrii statistici al căror calcul se dorește:

- **Summary statistics:** parametrii standard de statistică descriptivă
- **Confidence Level for Mean:** intervalul de încredere al mediei.

Figura 7. Comanda Descriptive Statistics

Rezultatele vor fi afișate de maniera următoare:

Microsoft Excel - DataAnalysis															
File Edit View Insert Format Tools Data Window Help															
Type a question for help															
Salt la Office Live Deschidere Salva															
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
Nr.crt.	Sex	Sex	Vârsta	Glicemie - mom. 0	Glicemie - 3 luni	Glicemie - 6 luni	Asigurat CAS								
1															
2	1	M	1	35	132	125	119	T							
3	2	M	1	24	133	133	106	T							
4	3	F	2	22	118	113	109	F							
5	4	F	2	53	111	107	97	T							
6	5	M	1	20	112	101	95	T							
7	6	F	2	37	130	122	97	T							
8	7	F	2	29	125	115	105	T							
9	8	M	1	56	138	121	109	F							
10	9	F	2	27	129	110	108	F							
11	10	F	2	21	110	115	105	F							
12	11	F	2	36	113	111	99	T							
13	12	M	1	38	145	133	129	F							
14	13	F	2	29	138	132	130	T							
15	14	M	1	44	132	130	131	F							
16	15	M	1	42	130	126	128	T							
17	16	F	2	32	125	120	105	F							
18	17	F	2	25	124	124	116	T							
19	18	F	2	29	122	111	107	F							
20	19	M	1	39	127	114	108	F							
21	20	M	1	40	118	108	108	T							
22	21	M	1	60	121	115	98	T							
23	22	F	2	58	136	130	128	T							
24	23	M	1	56	118	110	105	T							

Glicemie - mom. 0				Vârsta			
Mean	127.18	Standard Error	1.110135127	Mean	35.46	Standard Error	1.67970114
Median	128	Mode	127	Median	33.5	Mode	29
Standard Deviation	7.849840763	Sample Variance	61.62	Standard Deviation	11.87728066	Sample Variance	141.0697959
Kurtosis	0.213823154	Skewness	-0.129551306	Kurtosis	-0.759353723	Skewness	0.543323705
Range	35	Minimum	110	Range	42	Minimum	18
Maximum	145	Count	50	Maximum	60	Count	50
Confidence Level(95.0%)	2.230900019	Confidence Level(95.0%)	3.375485752				

Figura 8. Rezultatele obținute pentru vârstă și glicemie mom.0

- Calculați distribuția de frecvențe și realizați histograma corespunzătoare pentru variabilele Sex, Asigurat CAS, Vârsta și Glicemie - mom. 0.



Indicații. Soluții

Se alege din meniul superior *Data / Data Analysis / Histogram*:

Deoarece trebuie să utilizăm doar valori numerice, lângă coloanele Sex, respectiv Asigurat CAS mai adăugăm două coloane cu același nume, în care codificăm aceleași valori, dar folosind numere: pentru variabila Sex: masculin = 1, feminin = 2; pentru variabila Asigurat CAS: T = 1, F = 0.

Pentru a realiza o distribuție de frecvențe, trebuie specificate:

- Șirul de valori pentru care dorim să realizăm distribuția („Input Range”);
- Coloana de valori („Bin Range”) ce conține valorile distincte, ordonate crescător, pentru care se realizează distribuția de frecvențe (în cazul variabilelor calitative, acestea reprezintă de fapt categoriile urmărite, așezate ordonat – ex. 1, 2 pentru variabila SEX, iar în cazul variabilelor cantitative reprezintă limitele inferioară și superioară ale intervalelor de clasă).

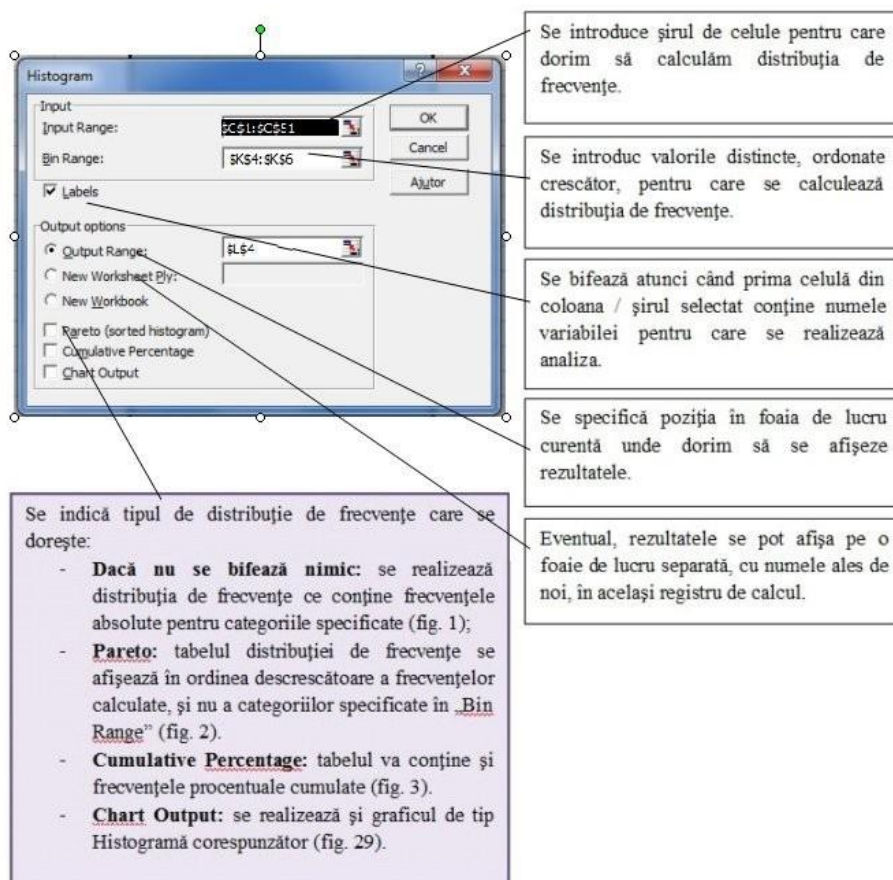


Figura 9. Comanda Histogram

Microsoft Excel - DataAnalysis														
File Edit View Insert Format Tools Data Window Help														
Salt la Office Live Deschidere Salvare														
L4 Sex														
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	Nr.crt.	Sex	Sex	Vârsta	Glicemie - mom. 0	Glicemie - 3 luni	Glicemie - 6 luni	Asigurat CAS						
1														
2	1	M	1	35	132	125	119	T						
3	2	M	1	24	133	133	106	T						
4	3	F	2	22	118	113	109	F						
5	4	F	2	53	111	107	97	T						
6	5	M	1	20	112	101	95	T						
7	6	F	2	37	130	122	97	T						
8	7	F	2	29	125	115	105	T						
9	8	M	1	56	138	121	109	F						
10	9	F	2	27	129	110	108	F						
11	10	F	2	21	110	115	105	F						
12	11	F	2	36	113	111	99	T						
13	12	M	1	38	145	133	129	F						
14	13	F	2	28	130	122	109	F						

Figura 10. Distribuția de frecvențe pentru variabila SEX

Microsoft Excel - DataAnalysis													
File Edit View Insert Format Tools Data Window Help													
Salt la Office Live Deschidere Salvare													
L4 Sex													
	A	B	C	D	E	F	G	H	I	J	K	L	M
	Nr.crt.	Sex	Sex	Vârsta	Glicemie - mom. 0	Glicemie - 3 luni	Glicemie - 6 luni	Asigurat CAS					
1													
2	1	M	1	35	132	125	119	T					
3	2	M	1	24	133	133	106	T					
4	3	F	2	22	118	113	109	F					
5	4	F	2	53	111	107	97	T					
6	5	M	1	20	112	101	95	T					
7	6	F	2	37	130	122	97	T					
8	7	F	2	29	125	115	105	T					
9	8	M	1	56	138	121	109	F					
10	9	F	2	27	129	110	108	F					
11	10	F	2	21	110	115	105	F					
12	11	F	2	36	113	111	99	T					
13	12	M	1	38	145	133	129	F					

Figura 11. Distribuția de frecvențe pentru variabila SEX – opțiunea „Cumulative percentage”

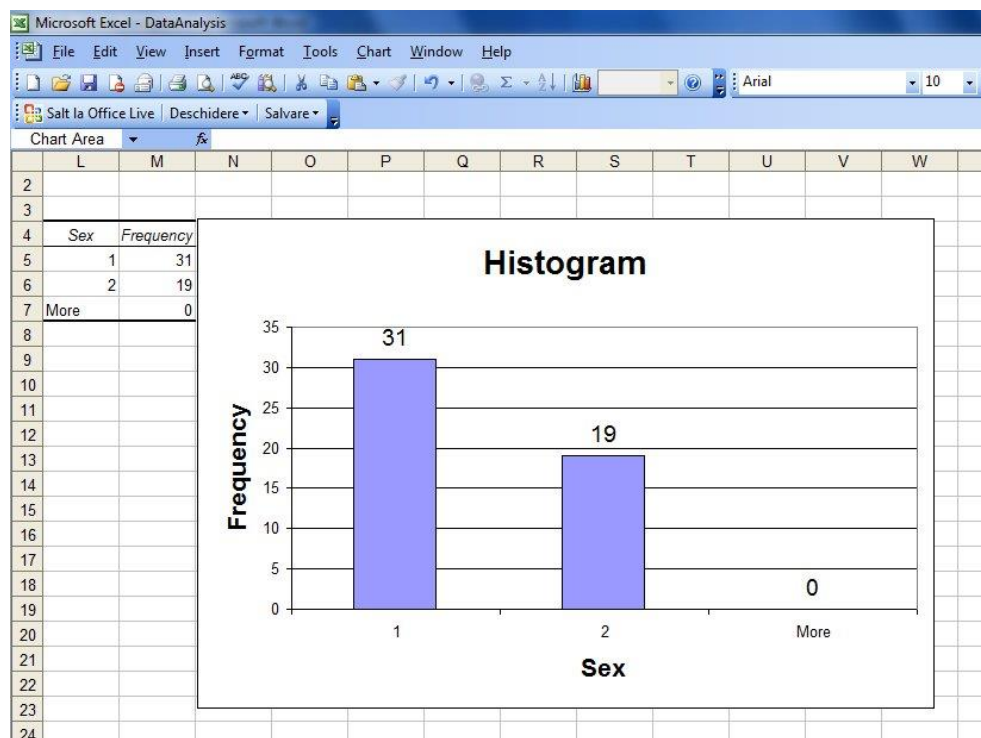


Figura 12. Histograma pentru variabila SEX – opțiunea „Chart Output”

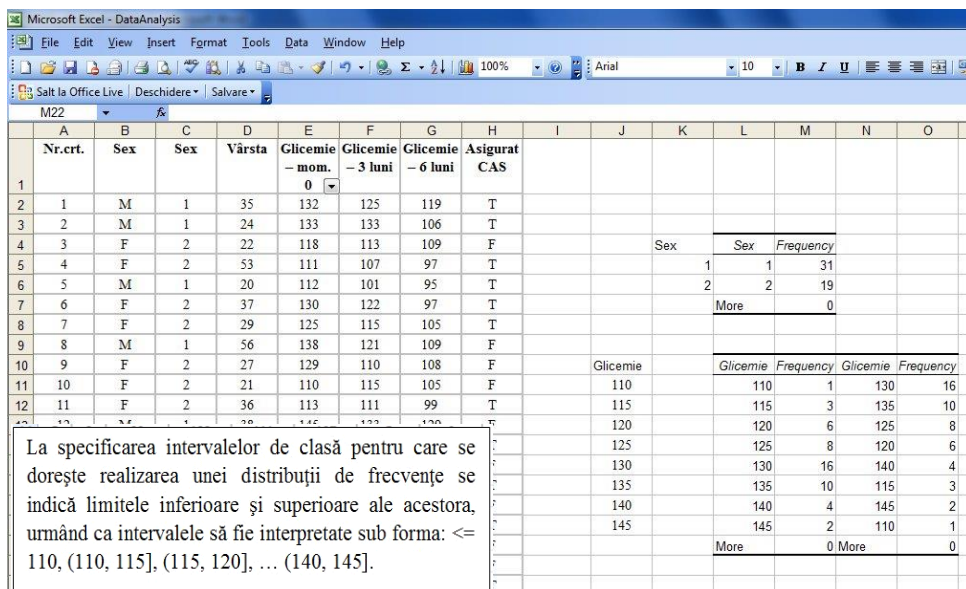


Figura 13. Pentru variabila SEX am realizat distribuția de frecvențe fără opțiuni suplimentare; Pentru variabila Glicemie – mom. 0 am realizat distribuția de frecvențe cu opțiunea „Pareto (sorted histogram)”.

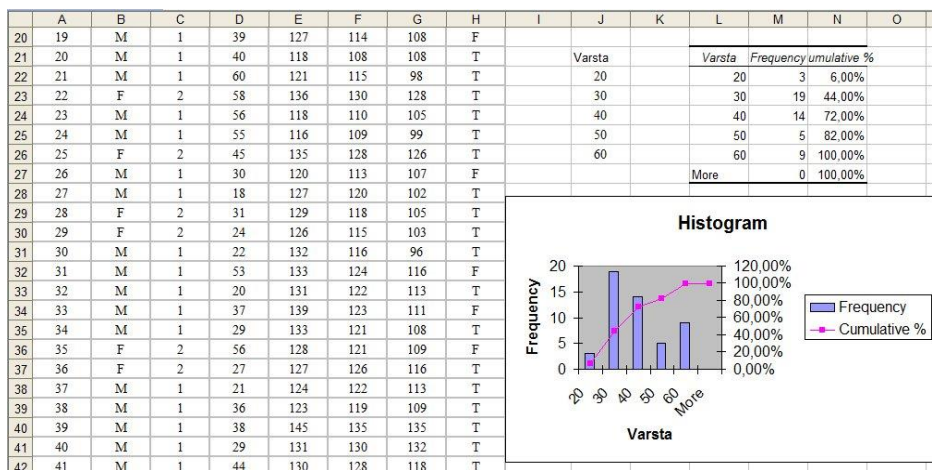


Figura 14. Distribuția de frecvențe și histograma pentru variabila Vârsta



Problema 5 (Testul t în Data Analysis)

Utilizând același set de date, se cere:

- Stabiliți dacă există diferențe semnificative statistic între valorile variabilelor *Glicemie – mom. 0* și *Glicemie – 3 luni*; *Glicemie – mom. 0* și *Glicemie – 6 luni*.



Indicații

Variabilele fiind cantitative, vom utiliza *Testul t*. Din cele 3 variante disponibile, se va folosi obligatoriu „*t-Test: Paired Two Sample For Means*”, deoarece se dorește compararea valorilor unei variabile pe același eșantion de pacienți, la momente diferite în timp.

Se selectează comanda *Data / Data Analysis / t-Test: Paired Two Sample For Means*:

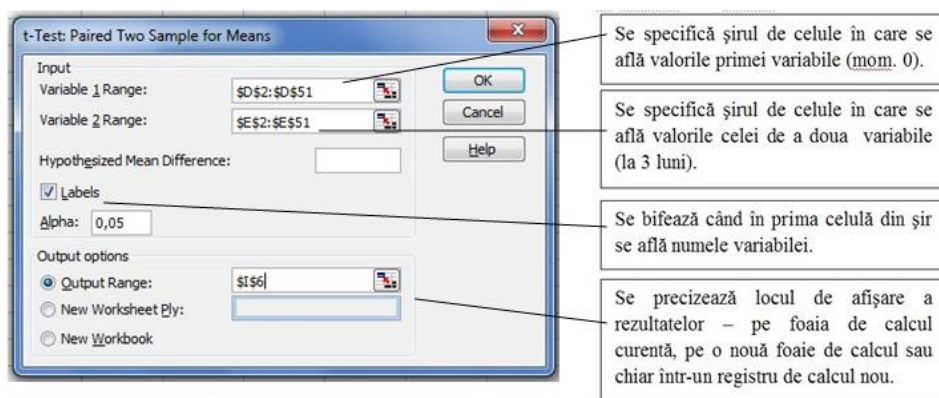


Figura 15. Comanda t-Test: Paired Two Sample For Means

Rezultatele sunt afișate după cum urmează:

Microsoft Excel - DataAnalysis23												
File Edit View Insert Format Tools Data Window Help												
Salt la Office Live Deschidere Salva												
	A	B	C	D	E	F	G	H	I	J	K	L
	Nr.crt.	Sex	Vârsta	Glicemie - mom. 0	Glicemie - 3 luni	Glicemie - 6 luni	Asigurat CAS					
1												
2	1	M	35	132	125	119	T					
3	2	M	24	133	133	106	T					
4	3	F	22	118	113	109	F					
5	4	F	53	111	107	97	T					
6	5	M	20	112	101	95	T					
7	6	F	37	130	122	97	T					
8	7	F	29	125	115	105	T					
9	8	M	56	138	121	109	F					
10	9	F	27	129	110	108	F					
11	10	F	21	110	115	105	F					
12	11	F	36	113	111	99	T					
13	12	M	38	145	133	129	F					
14	13	F	29	138	132	130	T					
15	14	M	44	132	130	131	F					
16	15	M	42	130	126	128	T					
17	16	F	32	125	120	105	F					
18	17	F	25	124	124	116	T					
19	18	F	29	122	111	107	F					
20	19	M	39	127	114	108	F					
21	20	M	40	118	108	108	T					
22	21	M	60	121	115	98	T					
23	22	M	60	121	115	98	T					
24	23	M	60	121	115	98	T					
25	24	M	60	121	115	98	T					
26	25	M	60	121	115	98	T					
27	26	M	60	121	115	98	T					
28	27	M	60	121	115	98	T					
29	28	M	60	121	115	98	T					
30	29	M	60	121	115	98	T					
31	30	M	60	121	115	98	T					
32	31	M	60	121	115	98	T					
33	32	M	60	121	115	98	T					
34	33	M	60	121	115	98	T					
35	34	M	60	121	115	98	T					
36	35	M	60	121	115	98	T					
37	36	M	60	121	115	98	T					
38	37	M	60	121	115	98	T					
39	38	M	60	121	115	98	T					
40	39	M	60	121	115	98	T					
41	40	M	60	121	115	98	T					
42	41	M	60	121	115	98	T					
43	42	M	60	121	115	98	T					
44	43	M	60	121	115	98	T					
45	44	M	60	121	115	98	T					
46	45	M	60	121	115	98	T					
47	46	M	60	121	115	98	T					
48	47	M	60	121	115	98	T					
49	48	M	60	121	115	98	T					
50	49	M	60	121	115	98	T					
51	50	M	60	121	115	98	T					
52	51	M	60	121	115	98	T					
53	52	M	60	121	115	98	T					
54	53	M	60	121	115	98	T					
55	54	M	60	121	115	98	T					
56	55	M	60	121	115	98	T					
57	56	M	60	121	115	98	T					
58	57	M	60	121	115	98	T					
59	58	M	60	121	115	98	T					
60	59	M	60	121	115	98	T					
61	60	M	60	121	115	98	T					
62	61	M	60	121	115	98	T					
63	62	M	60	121	115	98	T					
64	63	M	60	121	115	98	T					
65	64	M	60	121	115	98	T					
66	65	M	60	121	115	98	T					
67	66	M	60	121	115	98	T					
68	67	M	60	121	115	98	T					
69	68	M	60	121	115	98	T					
70	69	M	60	121	115	98	T					
71	70	M	60	121	115	98	T					
72	71	M	60	121	115	98	T					
73	72	M	60	121	115	98	T					
74	73	M	60	121	115	98	T					
75	74	M	60	121	115	98	T					
76	75	M	60	121	115	98	T					
77	76	M	60	121	115	98	T					
78	77	M	60	121	115	98	T					
79	78	M	60	121	115	98	T					
80	79	M	60	121	115	98	T					
81	80	M	60	121	115	98	T					
82	81	M	60	121	115	98	T					
83	82	M	60	121	115	98	T					
84	83	M	60	121	115	98	T					
85	84	M	60	121	115	98	T					
86	85	M	60	121	115	98	T					
87	86	M	60	121	115	98	T					
88	87	M	60	121	115	98	T					
89	88	M	60	121	115	98	T					
90	89	M	60	121	115	98	T					
91	90	M	60	121	115	98	T					
92	91	M	60	121	115	98	T					
93	92	M	60	121	115	98	T					
94	93	M	60	121	115	98	T					
95	94	M	60	121	115	98	T					
96	95	M	60	121	115	98	T					
97	96	M	60	121	115	98	T					
98	97	M	60	121	115	98	T					
99	98	M	60	121	115	98	T					
100	99	M	60	121	115	98	T					
101	100	M	60	121	115	98	T					
102	101	M	60	121	115	98	T					
103	102	M	60	121	115	98	T					
104	103	M	60	121	115	98	T					
105	104	M	60	121	115	98	T					
106	105	M	60	121	115	98	T					
107	106	M	60	121	115	98	T					
108	107	M	60	121	115	98	T					
109	108	M	60	121	115	98	T					
110	109	M	60	121	115	98	T					
111	110	M	60	121	115	98	T					
112	111	M	60	121	115	98	T					
113	112	M	60	121	115	98	T					
114	113	M	60	121	115	98	T					
115	114	M	60	121	115	98	T					
116	115	M	60	121	115	98	T					
117	116	M	60	121	115	98	T					
118	117	M	60	121	115	98	T					
119	118	M	60	121	115	98	T					
120	119	M	60	121	115	98	T					
121	120	M	60	121	115	98	T					
122	121	M	60	121	115	98	T					
123	122	M	60	121	115	98	T					
124	123	M	60	121	115	98	T					
125	124	M	60	121	115	98	T					
126	125	M	60	121	115	98	T					
127	126	M	60	121	115	98	T					
128	127	M	60	121	115	98	T					
129	128	M	60	121	115	98	T					
130	129	M	60	121	115	98	T					
131	130	M	60	121	115	98	T					
132	131	M	60	121	115	98	T					
133	132	M	60	121	115	98	T					
134	133	M	60	121	115	98	T					
135	134	M	60	121	115	98	T					
136	135	M	60	121	115	98	T					
137	136	M	60	121	115	98	T					
138	137	M	60	121	115	98	T					
139	138	M	60	121	115	98	T					
140	139	M	60	121	115	98	T					
141	140	M	60	121	115	98	T					
142	141	M	60	121	115	98	T					
143	142	M	60	121	115	98	T					
144	143	M	60	121	115	98	T					
145	144	M	60	121	115	98	T					
146	145	M	60	121	115	98	T					
147	146	M	60	121	115	98	T					
148	147	M	60	121	115	98	T					



Indicații. Soluții

Variabilele fiind cantitative, vom utiliza *Testul t*.

De această dată se lucrează evident cu eșantioane diferite pentru care se înregistrează valorile Glicemiei (bărbații și femeile din tabel), deci se va folosi testul t: Two-Sample; trebuie însă determinat care dintre cele 2 variante ale acestuia se va folosi: varianta „t-Test: Two-Sample Assuming Equal Variances” sau varianta „t-Test: Two-Sample Assuming Unequal Variances”.

Pentru a putea decide ce variantă de test t se folosește în situația noastră, vom verifica mai întâi dacă cele două eșantioane au varianțele egale sau nu. Astfel, se va folosi un alt test statistic, și anume *testul F* de egalitate a varianțelor.

Se selectează comanda *Data / Data Analysis / F-Test Two Sample for Variances*:

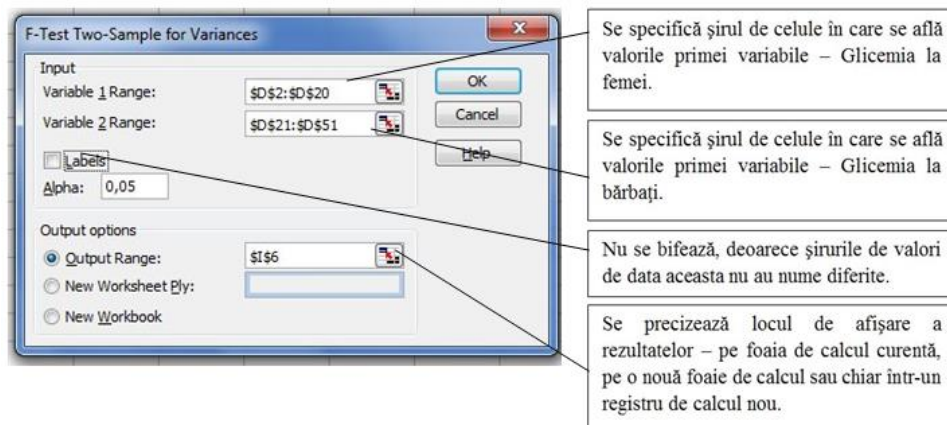


Figura 17. Comanda F-Test Two Sample for Variances

Rezultatele sunt afișate după cum urmează:

Microsoft Excel - DataAnalysis23												
File Edit View Insert Format Tools Data Window Help												
Salt la Office Live Deschidere Salvare												
	A	B	C	D	E	F	G	H	I	J	K	L
	Nr.crt.	Sex	Vârsta	Glicemie - mom. 0	Glicemie - 3 luni	Glicemie - 6 luni	Asigurat CAS					
1												
2	1	F	35	132	125	119	T					
3	2	F	24	133	133	106	T					
4	3	F	22	118	113	109	F					
5	4	F	53	111	107	97	T					
6	5	F	20	112	101	95	T					
7	6	F	37	130	122	97	T					
8	7	F	29	125	115	105	T					
9	8	F	56	138	121	109	F					
10	9	F	27	129	110	108	F					
11	10	F	21	110	115	105	F					
12	11	F	36	113	111	99	T					
13	12	F	38	145	133	129	F					
14	13	F	29	138	132	130	T					
15	14	F	44	132	130	131	F					
16	15	F	42	130	126	128	T					
17	16	F	32	125	120	105	F					

Figura 18. Rezultate testului F

Sunt calculați următorii indicatori:

- Media, varianța și numărul de observații pentru cele două șiruri de valori;
- Valoarea statisticii calculate $F = 2.357949$;
- $P(F \leq f)$ one-tail: $0.018270 > 0.05$, diferențele sunt *semnificative*, deci varianțele sunt inegale.

În continuare este clar că vom folosi testul t pentru varianțe diferite:

Se selectează comanda *Data / Data Analysis / t-Test Two-Sample Assuming Unequal Variances*:

t-Test: Two-Sample Assuming Unequal Variances

Input

Variable 1 Range:

Variable 2 Range:

Hypothesized Mean Difference:

☐ Labels

Alpha:

Output options

☒ Output Range:

☐ New Worksheet Ply:

☐ New Workbook

OK Cancel Help

Se specifică șirul de celule în care se află valorile primei variabile – Glicemia la femei.

Se specifică șirul de celule în care se află valorile primei variabile – Glicemia la bărbați.

Nu se bifează, deoarece șirurile de valori de data aceasta nu au nume

Se precizează locul de afișare a rezultatelor – pe foaia de calcul curentă, pe o nouă foaie de calcul sau chiar într-un registru de calcul nou.

Figura 19. Comanda t-Test Two-Sample Assuming Unequal Variances

Rezultatele sunt afișate după cum urmează:

Microsoft Excel - DataAnalysis3

File Edit View Insert Format Tools Data Window Help

Salt la Office Live | Deschidere | Salvare

J47

	A	B	C	D	E	F	G	H	I	J	K	L	M
	Nr.crt.	Sex	Vârsta	Glicemie - mom. 0	Glicemie - 3 luni	Glicemie - 6 luni	Asigurat CAS						
1													
2	1	F	35	132	125	119	T						
3	2	F	24	133	133	106	T						
4	3	F	22	118	113	109	F						
5	4	F	53	111	107	97	T						
6	5	F	20	112	101	95	T						
7	6	F	37	130	122	97	T						
8	7	F	29	125	115	105	T						
9	8	F	56	138	121	109	F						
10	9	F	27	129	110	108	F						
11	10	F	21	110	115	105	F						
12	11	F	36	113	111	99	T						
13	12	F	38	145	133	129	F						
14	13	F	29	138	132	130	T						
15	14	F	44	132	130	131	F						
16	15	F	42	130	126	128	T						
17	16	F	32	125	120	105	F						
18	17	F	25	124	124	116	T						
19	18	F	29	122	111	107	F						
20	19	F	39	127	114	108	F						
21	20	M	40	118	108	108	T						

t-Test: Two-Sample Assuming Unequal Variances

	Variable 1	Variable 2
Mean	126	127.9032258
Variance	96.8888889	41.09032258
Observations	19	31
Hypothesized Mean Difference	0	
df	27	
t Stat	-0.750856269	
P(T<=t) one-tail	0.22961682	
t Critical one-tail	1.703288423	
P(T<=t) two-tail	0.45923364	
t Critical two-tail	2.051830493	

Figura 20. Rezultate testului t

Interpretarea lor este similară cu cea de la testul t pentru eșantioane perechi.

Observăm că $t \text{ Stat} = -0,750856$, $P(T \leq t) \text{ one-tail} = 0.229616 > 0.05$, diferențele între valorile medii ale glicemiei la mom.0 la femei și bărbați sunt *ne semnificative* statistic.



Problema 6 (Corelații în Data Analysis)

Utilizând datele din următorul tabel realizați o bază de date, DataAnalysis3.xls.

Nr.crt.	Sex	Vârsta	Greutate	Glicemie	Glicemie	Glicemie	Asigurat
1	M	35	65	132	125	119	T
2	M	24	70	133	133	106	T
3	F	22	75	118	113	109	F
4	F	53	80	111	107	97	T
5	M	20	90	112	101	95	T
6	F	37	100	130	122	97	T
7	F	29	75	125	115	105	T
8	M	56	98	138	121	109	F
9	F	27	59	129	110	108	F

10	F	21	55	110	115	105	F
11	F	36	65	113	111	99	T
12	M	38	87	145	133	129	F
13	F	29	82	138	132	130	T
14	M	44	94	132	130	131	F
15	M	42	115	130	126	128	T
16	F	32	97	125	120	105	F
17	F	25	98	124	124	116	T
18	F	29	100	122	111	107	F
19	M	39	112	127	114	108	F
20	M	40	85	118	108	108	T
21	M	60	102	121	115	98	T
22	F	58	101	136	130	128	T
23	M	56	92	118	110	105	T
24	M	55	91	116	109	99	T
25	F	45	89	135	128	126	T
26	M	30	88	120	113	107	F
27	M	18	97	127	120	102	T
28	F	31	75	129	118	105	T
29	F	24	74	126	115	103	T
30	M	22	76	132	116	96	T
31	M	53	80	133	124	116	F
32	M	20	65	131	122	113	T
33	M	37	70	139	123	111	F
34	M	29	80	133	121	108	T
35	F	56	60	128	121	109	F
36	F	27	58	127	126	116	T
37	M	21	68	124	122	113	T
38	M	36	85	123	119	109	T
39	M	38	93	145	135	135	T
40	M	29	94	131	130	132	T
41	M	44	103	130	128	118	T
42	M	42	111	124	120	119	T
43	M	32	97	119	108	106	F
44	F	25	58	129	115	114	T
45	M	29	69	128	120	109	F
46	M	39	76	127	125	108	T
47	F	40	77	129	126	112	T
48	M	24	78	132	128	113	T
49	M	22	84	127	120	117	T
50	M	53	96	128	115	116	T

Figura 21. DataAnalysis3.xls

- Determinați coeficienții de corelație între *Vârstă*, *Greutate*, *Glicemie la mom. 0, 3 luni și 6 luni*.



Indicații. Soluții

Se selectează comanda *Data / Data Analysis / Correlation*:

Se specifică șirul de celule în care se află valorile variabilelor pentru care dorim să calculăm coeficienții de corelație.

Se specifică modul de grupare a valorilor – în coloane în cazul de față.

Se bifează când în prima celulă din fiecare coloană se află numele variabilei.

Se precizează locul de afișare a rezultatelor – pe foaia de calcul curentă, pe o nouă foaie de calcul sau chiar într-un registru de calcul nou.

Figura 22. Comanda Correlation

Rezultatele sunt afișate după cum urmează:

Nr.crt.	Sex	Vârsta	Greutate	Glicemie - mom. 0	Glicemie - 3 luni	Glicemie - 6 luni	Asigurat CAS
1							
2	1	M	35	65	132	125	119
3	2	M	24	70	133	133	106
4	3	F	22	75	118	113	109
5	4	F	53	80	111	107	97
6	5	M	20	90	112	101	95
7	6	F	37	100	130	122	97
8	7	F	29	75	125	115	105
9	8	M	56	98	138	121	109
10	9	F	27	59	129	110	108
11	10	F	21	55	110	115	105
12	11	F	36	65	113	111	99
13	12	M	38	87	145	133	129
14	13	F	29	82	138	132	130
15	14	M	44	94	132	130	131
16	15	M	42	115	130	126	128
17	16	F	32	97	125	120	105

	Vârsta	Greutate	Glicemie - mom. 0	Glicemie - 3 luni	Glicemie - 6 luni
Vârsta	1				
Greutate	0.387960551	1			
Glicemie - mom. 0	0.069795287	0.066012538	1		
Glicemie - 3 luni	0.029167104	0.051983949	0.801093206	1	
Glicemie - 6 luni	0.11276537	0.190907855	0.643440682	0.748215017	1

Figura 23. Corelații obținute

Se generează de fapt *Matricea Coeficienților de Corelație*, ce conține coeficienții de corelație Pearson între toate perechile posibile de variabile selectate prin comandă.

Rezultatele obținute sunt: corelația între greutate și vârstă este 0.38796, deci directă, rezonabilă; corelațiile între glicemie 0 și vârstă (0.06979), respectiv glicemie 0 și greutate (0.06601), corelațiile între glicemie 3 luni și vârstă (0.02916), respectiv glicemie 3 luni, greutate (0.051984), corelațiile între

glicemie 6 luni, vârstă (0,112765), respectiv glicemie 3 luni și greutate (0.19090) sunt *practic absentă*.

3. Concluzii

În acest material am reușit:

- ✚ să ne familiarizăm cu funcțiile statistice din Insert Function;
- ✚ să utilizăm funcțiile din Data analysis.

Referințe

- Vernic CV, Apostol SA, Frandes M, Mada L, Lungeanu D. APLICATII PRACTICE DE INFORMATICA SI BIOSTATISTICA MEDICALA IN NURSING. Editura Eurostampa, ISBN 978-606-32-0487-6, Colectia Derzelas, 2017:1-216.
- Vernic CV, Timar B, Mada L, Apostol SA. NOȚIUNI DE INFORMATICĂ MEDICALĂ ȘI ABORDĂRI DE BIOSTATISTICĂ ÎN NURSING. Editura Eurostampa, ISBN 978-606-32-0152-3, Colectia Derzelas, 2015:1-282.
- Vernic CV, Timar B, Mada L, Apostol SA *Informatică medicală și metode de biostatistică aplicate în nursing*. Editura Eurostampa, 2014: 1-388.
- Vernic CV, Mada L, Lungeanu D, Muntean C, Apostol SA, Catu CO, Ursoniu S. *Aplicații practice de Informatică Medicală și Biostatistică*. Editura Victor Babes, 2010: 1-229.
- Johnson S. *Microsoft Office. Excel 2003*. Teora, 2004.
- <http://office.microsoft.com/en-us/excel-help/>
- Laing R. *Microsoft Excel Basics*. Editura Flame Tree, 2015.