

Funcții predefinite Excel (parametrii de dispersie, simetrie și localizare) & Statistica Descriptivă cu Data Analysis

TIMP DE LUCRU: 2 ORE

Obiective:

- Funcții definite de utilizator prin utilizarea referințelor relative: Amplitudinea ($A = X_{\max} - X_{\min}$), Eroarea standard ($ES = s/(\sqrt{n})$); coeficient de variație ($CV = s/\bar{X}$)
- Funcții predefinite Excel: parametrii de dispersie (variația = VAR, deviația standard = STDEV)
- Funcții predefinite Excel: parametrii de simetrie (simetria = SKEW, boltirea = KURT)
- Funcții predefinite Excel: parametrii de localizare (cvartila 1, 2, și 3 = QUARTILE)
- Determinarea parametrilor statisticii descriptive cu Data Analysis.

Problema:

Un medic de familie care are înscriși pe liste un număr de 2542 pacienți a realizat un studiu pentru a identifica riscul cardiac. Datorită resurselor limitate s-a recurs la alegerea aleatoare a unui eșantion de 62 pacienți. A fost inclus în eșantion prin eșantionare sistematică tot al 41-lea pacient care avea vârstă cuprinsă între 20 și 90 ani și a semnat acordul de participare la studiu. Următoarele date au fost culese de la fiecare pacient: sex, profilul dislipidemic (da/nu, din fișa pacientului), consumul afirmativ de alcool (da/nu), fumat (da/nu, afirmativ), vârsta (ani, fișa pacientului). Următoarele măsurători au fost realizate pentru fiecare pacient: greutate (kg), înălțime (cm), TAS (tensiunea arterială sistolică, mmHg), TAD (tensiunea arterială diastolică, mmHg). Următoarele determinări biochimice au fost realizate în urma recoltării unei probe de sânge: glicemie (mg/dl), colesterol (mg/dl), trigliceride (mg/dl) și HDL colesterol (mg/dl). Datele au fost colectate și introduse în fișierul *RiscCardiac.xls*.

Copiați baza de date *RiscCardiac.xls* din folderul laboratorului precedent în folderul laboratorului curent.

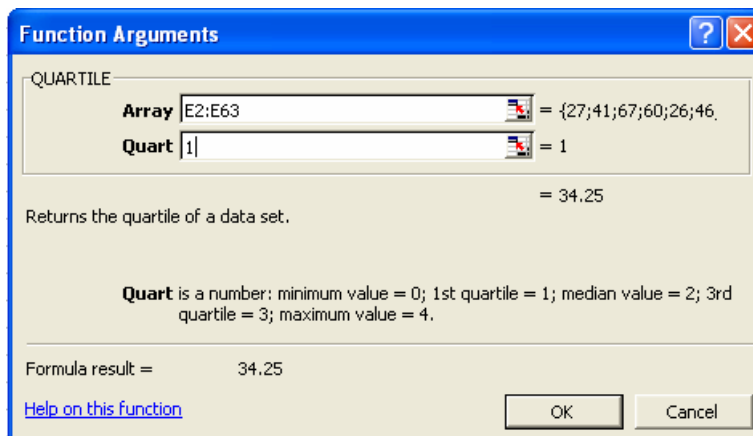
1. Calculați parametrii de dispersie, simetrie și localizare pentru variabilele cantitative continue din baza de date.

- a. În foaia de calcul predefinite, completați structura tabelară de la dreapta tabelului cu date în conformitate cu imaginea de mai jos:

V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM
	dislipidemie	alcool	fumat	varsta	greutate (kg)	inaltime (cm)	IMC	TAS	TAD	TAM	HT	glicemie	diabet	colesterol	trigliceride	HDL colesterol	risc cardiac
COUNTIF=DA	3	12	13								28		4				0
COUNTIF=NU	3	12	13								34		58				62
Media aritmetica				44	79	169	28	128	84	99		88		211	141		45
Mediana				44	79	170	27	130	80	97		88		216	129		45
Modulul				45	76	162	33	130	80	93.3		88		215	136		46
Minimum				17	50	152	19	90	60	70		67		142	48		30
Maximum				71	120	188	41	190	120	143		138		343	420		66
Amplitudinea																	
Valoarea centrală																	
Variația																	
Deviația standard																	
Eroarea standard																	
Coeficient de variație																	
Simetrie																	
Boltire																	
Cvartila 1 (25%) - Q1																	
Cvartila 2 (50%) - Q2																	
Cvartila 3 (50%) - Q3																	
Q2 - Q1																	
Q3 - Q2																	

- b. Următoarele funcții predefinite se utilizează pentru obținerea rezultatelor în tabelul anterior:

- i. VAR = variația
- ii. STDEV = deviația standard
- iii. SKEW = simetria
- iv. KURT = boltirea
- v. QUARTILE = cvartile. În fereastra de dialog a funcției la opțiunea **Array** selectați celulele unde aveți datele pentru variabila de interes iar la opțiunea **Quart** scrieți 1 dacă doriți să obțineți cvartila de 25%, 2 dacă doriți să obțineți cvartila de 50% sau 3 dacă doriți să obțineți cvartila de 75%:



- c. Următoarele formule definite de utilizator prin utilizarea referințelor relative de celule se utilizează pentru obținerea rezultatelor în tabelul anterior:
 - i. Amplitudinea: $=X_{Max} - X_{Min}$. Exemplu: pentru a calcula amplitudinea vârstei selectăm celula Z9 și **=Z8-Z7**
 - ii. Valoarea centrală: $=(X_{Max} + X_{Min})/2$. Exemplu: pentru a calcula valoarea centrală a vârstei selectăm celula Z10 și **=(Z8+Z7)/2**.
 - iii. Eroarea standard: $=s/(\sqrt{n})$ unde s = deviația standard și n = volumul eșantionului. Exemplu: pentru a calcula eroarea standard a vârstei selectăm celula Z13 și **=Z12/SQRT(62)**.
 - iv. Coeficient de variație: $CV=s/\bar{X}$ unde s = deviația standard și \bar{X} = media aritmetică. Exemplu: pentru a calcula coeficientul de variație al vârstei selectăm celula Z14 și **=Z12/Z4**. Cum este variabila investigată din punct de vedere al omogenității? Inerați un comentariu în care să includeți tipul seriei (omogenă, relativ omogenă, eterogenă, relativ eterogenă) pentru fiecare variabilă. Exemplu: valoarea coeficientului de variație pentru variabila vârstă este egală cu 0.29 ceea ce semnifică o serie relativ eterogenă. Pentru a include tipul de serie asociat valorii selectăm celula Z14 – clic dreapta – Insert Comment și în căsuța care apare scriem relativ eterogenă.

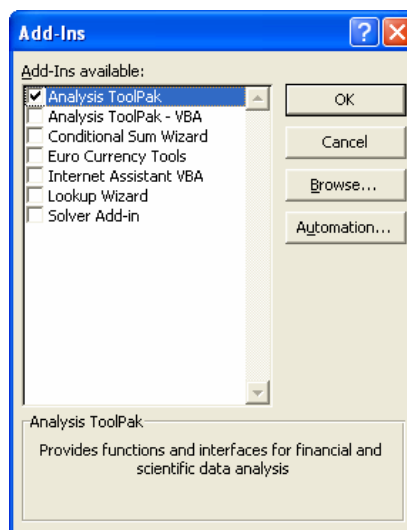
V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM
	dislipidemie	alcool	fumat	varsta	greutate (kg)	inaltime (cm)	IMC	TAS	TAD	TAM	HT	glicemie	diabet	colesterol	trigliceride	HDL colesterol	risc cardiac
COUNTIF=DA		3	12	13							28		4				0
COUNTIF=NU		3	12	13							34		58				62
Media aritmetica				44	79	169	28	128	84	99		88		211	141		45
Mediana				44	79	170	27	130	80	97		88		216	129		45
Modulul				45	76	162	33	130	80	93.3		88		215	136		46
Minimum				17	50	152	19	90	60	70		67		142	48		30
Maximum				71	120	188	41	190	120	143		138		343	420		66
Amplitudinea				54													
Valoarea centrală				44													
Variația				166.085													
Deviația standard				12.8874													
Eroarea standard				1.6367													
Coefficient de variație				0.29495													
Simetrie																	
Boltire																	
Cvartila 1 (25%) - Q1				34.25													
Cvartila 2 (50%) - Q2				43.5													
Cvartila 3 (50%) - Q3				50.75													
Q2 - Q1																	
Q3 - Q2																	

Sorana:
relativ eterogenă

- v. Diferența dintre cvartial 2 și 1: =Q2-Q1. Exemplu: pentru a calcula diferența pentru variabila vârstă selectăm celula Z20 și =Z18-Z17.
- vi. Diferența dintre cvartial 3 și 2: =Q3-Q2. Exemplu: pentru a calcula diferența pentru variabila vârstă selectăm celula Z21 și =Z19-Z18.
- vii. Pe baza diferenței cvartilelor clasificați simetria fiecărei variabile (simetrică, asimetrică). Intrați această informație ca și comentariu asociat denumirii variabilelor (de exemplu, pentru vârstă includeți comentariul asociat celulei Z1). Informații cu privire la simetrie și asimetrie se găsesc în Cursul_08.

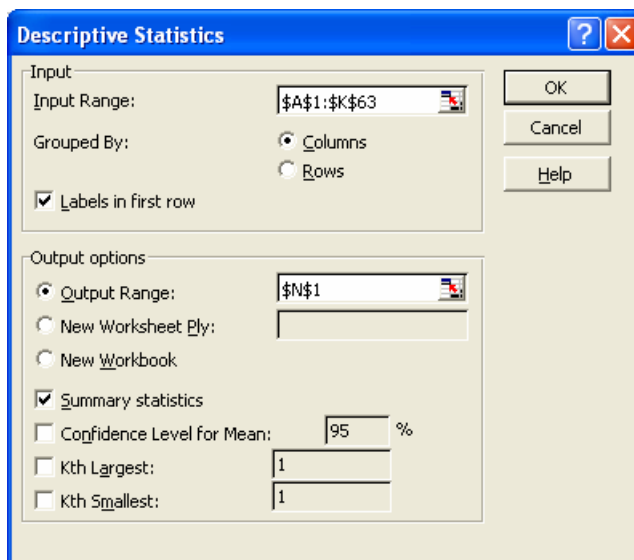
2. Determinați parametrii statistici descriptivi prin utilizarea opțiunii Descriptive Statistics din [Tool – Data Analysis]

- a. Intrați o nouă foaie de calcul în fișierul **RiscCardiac.xls**. Denumiți foaia nou inserata **statistica descriptiva**.
- b. Copiați în foaia de calcul **statistica descriptiva** toate variabilele cantitative din baza de date.
- c. Calculați parametrii statisticii descriptive pentru aceste variabile cantitative:
 - i. Activați opțiunea Data Analysis din meniul Tools: [Tools – Add-Ins – Analysis ToolPak]



ii. [Tools – Data Analysis – Descriptive Statistics]:

1. În căsuța corespunzătoare la **Input Range** selectați toate celulele care conțin valori ale variabilelor cantitative.
2. Specificați că aveți denumiri de coloană în primul rând (bifați **Labels in first row**).
3. În secțiunea **Output options**: alegeți să afișați rezultatele începând cu celula N1 și explicați programului să calculeze toți parametrii statisticii (bifați **Summary statistics**).
4. Pentru detalii vedeți imaginea de mai jos:



5. Formatați rezultatele obținute astfel încât la final să arate ca și în imaginea de mai jos:

N	O	P	Q	R	S	T	U	V	W	X	Y
	<i>varsta</i>	<i>greutate (kg)</i>	<i>inaltime (cm)</i>	<i>IMC</i>	<i>TAS</i>	<i>TAD</i>	<i>TAM</i>	<i>glicemie</i>	<i>colesterol</i>	<i>trigliceride</i>	<i>HDL colesterol</i>
Mean (media aritmetică)	43.69	79.18	168.65	27.75	128.23	83.79	98.60	88.48	210.81	141.37	45.34
Standard Error (eroarea standard)	1.64	1.98	1.16	0.59	2.64	1.63	1.87	1.99	4.93	9.69	0.96
Median (mediana)	43.5	78.5	170	27.20	130	80	96.67	88	215.5	128.5	45
Mode (valoarea modală)	45	76	162	32.89	130	80	93.33	88	215	136	46
Standard Deviation (deviația standard)	12.89	15.62	9.13	4.61	20.79	12.86	14.73	15.70	38.82	76.33	7.54
Sample Variance (variația)	166.08	243.95	83.41	21.27	432.05	165.32	216.87	246.38	1506.85	5826.17	56.85
Kurtosis (boltirea)	-0.61	-0.38	-1.04	0.27	0.33	-0.15	0.15	2.25	0.97	3.88	0.58
Skewness (asimetria)	0.28	0.31	-0.12	0.55	0.71	0.36	0.46	1.43	0.49	1.76	0.62
Range (amplitudinea)	54	70	36	22.36	100	60	73.33	71	201	372	36
Minimum (minim)	17	50	152	19.05	90	60	70	67	142	48	30
Maximum (maxim)	71	120	188	41.41	190	120	143.33	138	343	420	66
Sum (suma)	2709	4909	10456	1720.74	7950	5195	6113.33	5486	13070	8765	2811
Count (volumul eșantionului)	62	62	62	62	62	62	62	62	62	62	62