



UNIVERSITATEA DE MEDICINĂ, FARMACIE, ȘTIINȚE ȘI TEHNOLOGIE
„GEORGE EMIL PALADE”
DIN TÂRGU MUREȘ

Facultatea de Inginerie și Tehnologia Informației

Master Inteligență Artificială
An 1

Ianău Andrei-Ioan

Introducere

Studiul analizeaza diferiti factori care influenteaza diferiti parametri din domeniul cardiologic.

Eșantionul folosi face parte din baza de date creată de:

1. Hungarian Institute of Cardiology. Budapest: Andras Janosi, M.D.
2. University Hospital, Zurich, Switzerland: William Steinbrunn, M.D.
3. University Hospital, Basel, Switzerland: Matthias Pfisterer, M.D.
4. V.A. Medical Center, Long Beach and Cleveland Clinic Foundation: Robert Detrano, M.D., Ph.D.

Donată de:

David W. Aha (aha '@' ics.uci.edu) (714) 856-8779

1. Descrierea bazei de date individuale (cu menționarea tipului de variabilă și a unității de măsură) și sursa de unde au fost culese datele

Baza de date contine 123 de observatii statistice, respectiv pacienti si urmatoarele variabile.

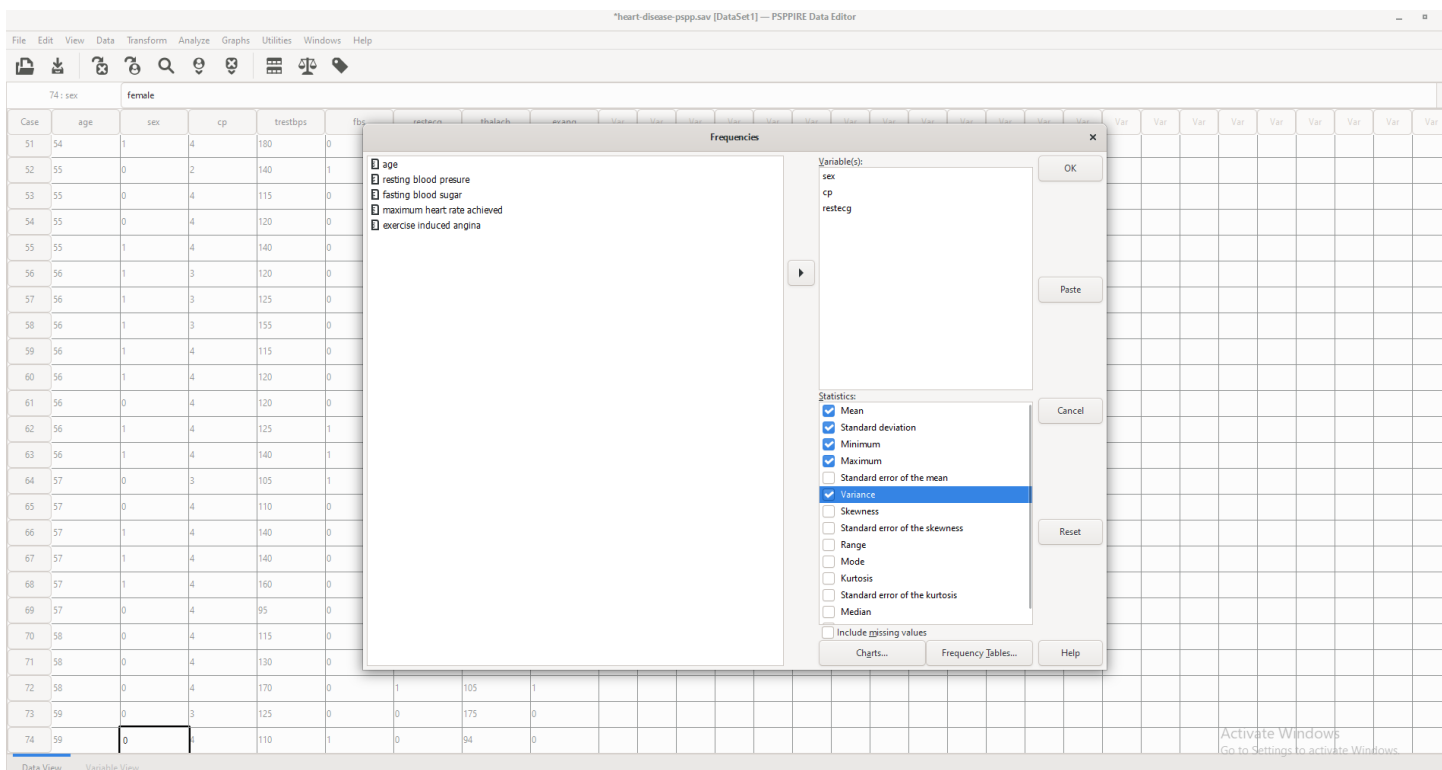
Nr. crt	Denumire variabila	Tipul variabile	Unitate de masura
1	Varsta	Cantitativa numerica continua	ani
2	Sexul pacientului	Calitativa dihotomica	
3	Tipul de durere in piept	Calitativa alternativa	
4	Presiunea sanguina	Cantitativa numerica ordinala	mmHg
5	Glicemie pe nemancate	Cantitativa numerica continua	
6	ECG	Cantitativa numerica continua	
7	Pulsul maxim	Cantitativa numerica continua	
8	Angina indusa prin sport	Calittiva dihotomica	

2. Statistica descriptivă (inclusive reprezentări grafice)

2.1.Indicatorii tendinței centrale

Pentru calculul indicatorilor tendinței centrale s-a utilizat soft-ul PSPP 1.6.2-g78a33a\

Comandă PSPP: Analyze -> Descriptive statistics -> Frequencies

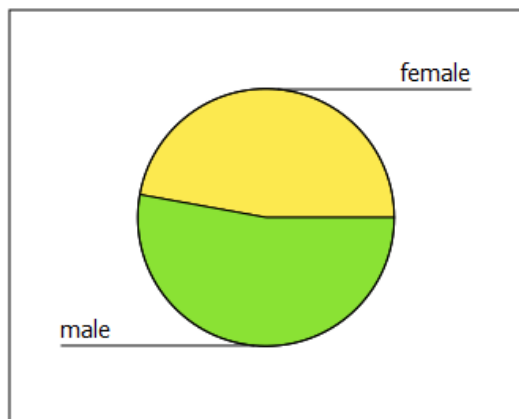


Sexul pacientului:

sex

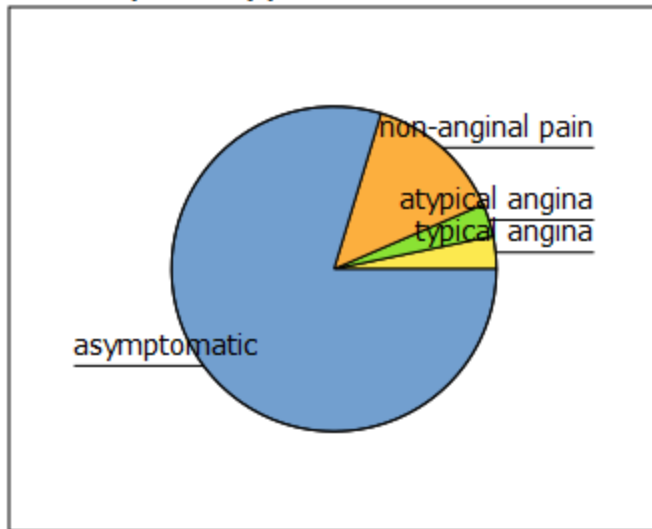
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	female	58	47.2%	47.2%	47.2%
	male	65	52.8%	52.8%	100.0%
Total		123	100.0%		

sex



Se observa ca avem un numar relativ apropiat de femei si de barbati in studiu.

chest pain type

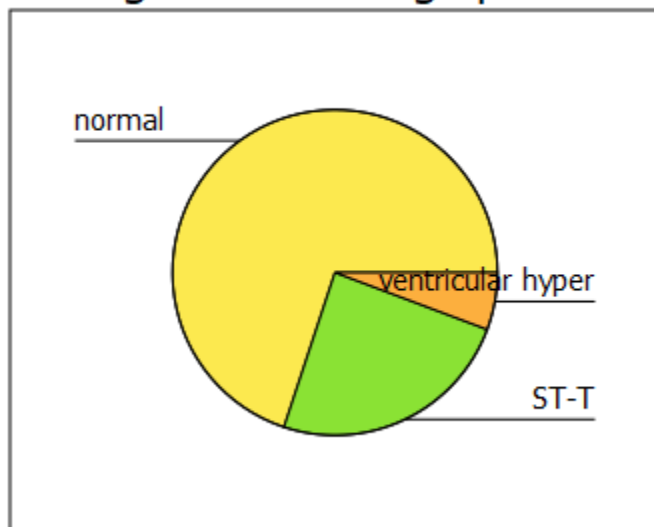


chest pain type

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid typical angina	4	3.3%	3.3%	3.3%
atypical angina	4	3.3%	3.3%	6.5%
non-anginal pain	17	13.8%	13.8%	20.3%
asymptomatic	98	79.7%	79.7%	100.0%
Total	123	100.0%		

Avem majoritar dureri de piept asimptomatice.

resting electrocardiographic results



resting electrocardiographic results

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid normal	86	69.9%	69.9%	69.9%
ST-T	30	24.4%	24.4%	94.3%
ventricular hyper	7	5.7%	5.7%	100.0%
Total	123	100.0%		

Din intregul numar de pacienti, ~70% au EKG-ul normal in timp ce ~25% au ST-T wave abnormality.

fasting blood sugar

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid < 120 mg/dl	104	84.6%	84.6%	84.6%
> 120 mg/dl	19	15.4%	15.4%	100.0%
Total	123	100.0%		

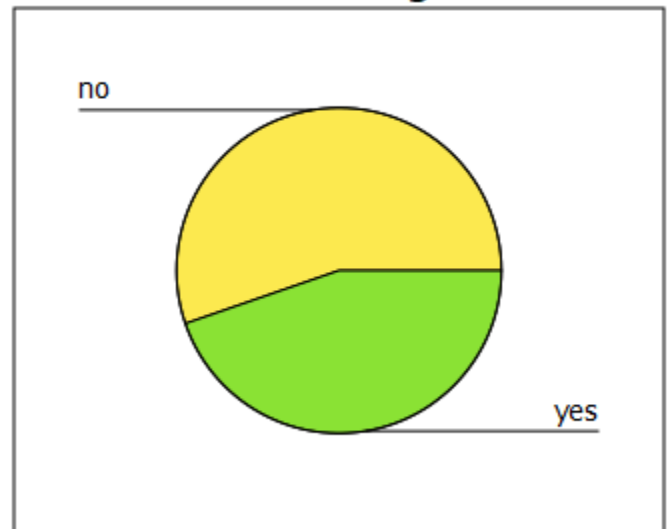
Se observa ca doar 15% din pacienti au glicemia > 120mg/dl.

Pentru angina indusa prin exercitiu valorile sunt relativ apropiate.

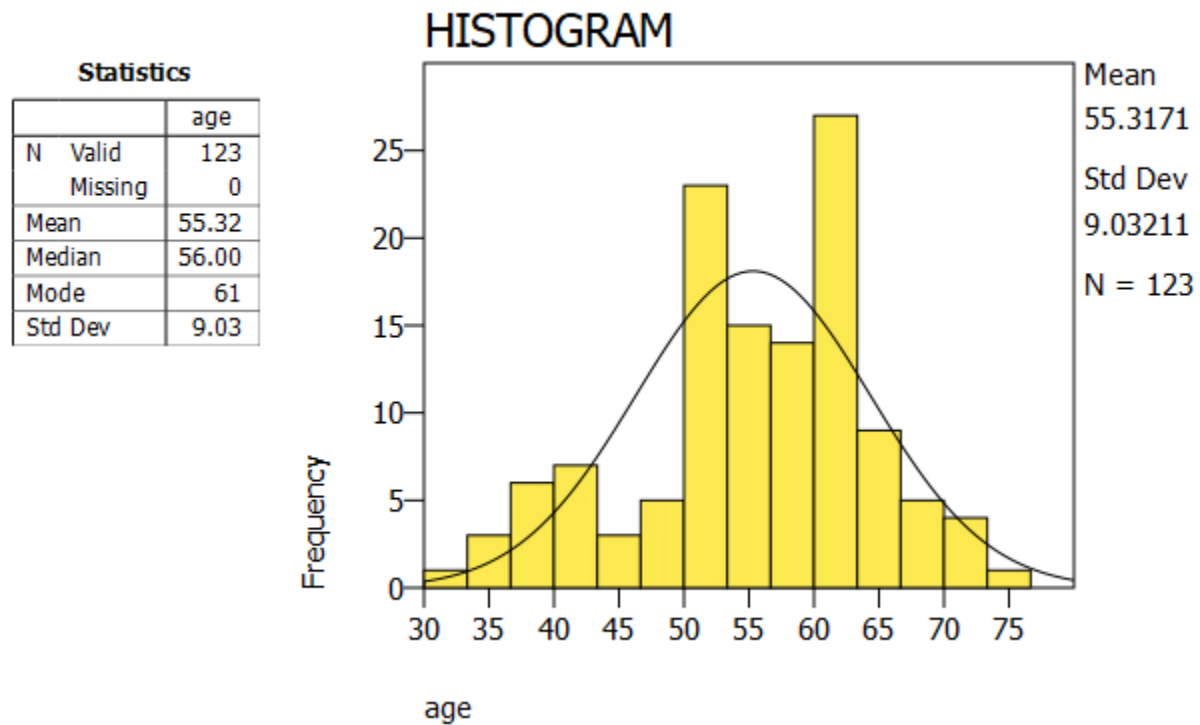
exercise induced angina

	Frequency	Percent	Valid Percent
Valid no	68	55.3%	55.3%
yes	55	44.7%	44.7%
Total	123	100.0%	

exercise induced angina

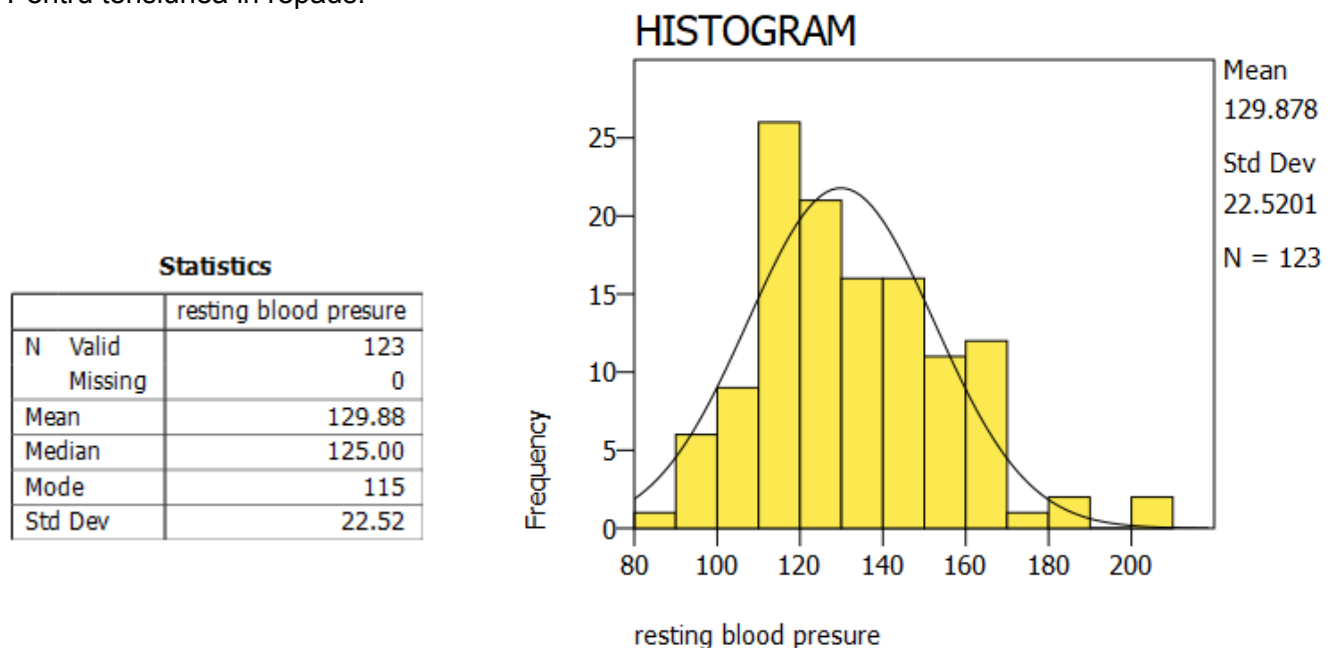


Pentru varsta, avem urmatoarele informatii:



Se observa ca avem o curba gauss de normalitate destul de normala dar o discrepanta la nivelurile de 45-50 de ani si 60-65 de ani. Cei mai multi participanti au 61 de ani. Asta indica o prevalenta de a fi luate in calcul mai mult informatii pentru varstele de 50-63 de ani.

Pentru tensiunea in repaus:



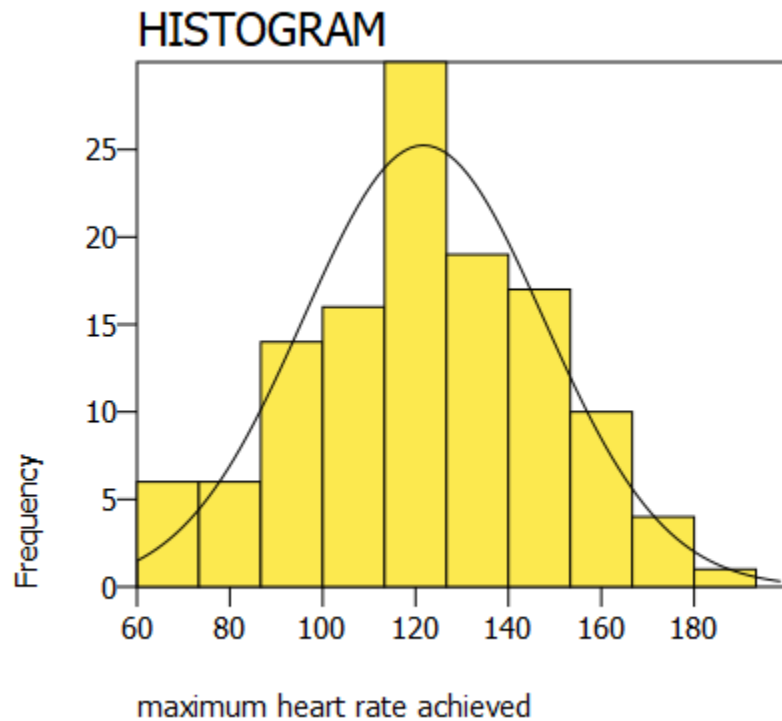
Curba Gauss asigura ca avem o normalitate aproximativa insa si niste outliers (la valorile de 200) si faptul ca sunt un numar ridicat de pacienti ce au 115 ca si tensiune. Acestea reies si din Mode care e 115.

Pentru ritmul maxim al batailor inimii:

Statistics		maximum heart rate achieved
N	Valid	123
	Missing	0
Mean		121.70
Median		121.00
Mode		120
Std Dev		25.92
Variance		671.77

Avem un mode de 120 si media si mediana fiind pe la 121 ca si valoare.

O curba normala gauss desul de descriptiva a datelor.



2.2 Indicatorii ai dispersiei

Descriptive Statistics

	N	Mean	Std Dev	Kurtosis	S.E. Kurt	Skewness	S.E. Skew	Range	Minimum	Maximum
age	123	55.32	9.03	-.11	.43	-.50	.22	42.00	32	74
resting blood pressure	123	129.88	22.52	.38	.43	.60	.22	120.00	80	200
maximum heart rate achieved	123	121.70	25.92	-.22	.43	-.07	.22	122.00	60	182
Valid N (listwise)	123									
Missing N (listwise)	0									

Pentru **varsta** avem faptul ca aceasta variaza de la medie cu ± 9 ani.

Valoarea cea mai mica este de 32 de ani si

cea mai mare este de 74 de ani.

Amplitudinea e de 42 de ani.

Aplatizarea e de $-.11$ ceea ce indica un varf ce este sub media valorii normale Gauss (observatie ce se poate vedea si in histograma de mai sus).

Asimetria este de -0.5 ceea ce indica o "coada" a distributiei spre valori mai mici decat medie. (observatie ce se poate vedea si in histograma de mai sus).

Pentru tensiunea in repaus putem spune ca variaza fata de medie cu 22.52 de unitati.

Valoarea cea mai mica e de 80 si

cea mai mare fiind de 200.

Amplitudinea este de 120.

Aplatizarea are valoarea de 0.38 ceea ce indica un varf al distributiei peste varful distributiei normale normale.

Asimetria este de $.60$ ceea ce indica o "coada" a distributiei catre valori mai mari ceea ce se poate observa.

Ritmul cardiac maxim atins variaza in medie cu ~ 26 de unitati si

are o valoarea maxima de 182 si

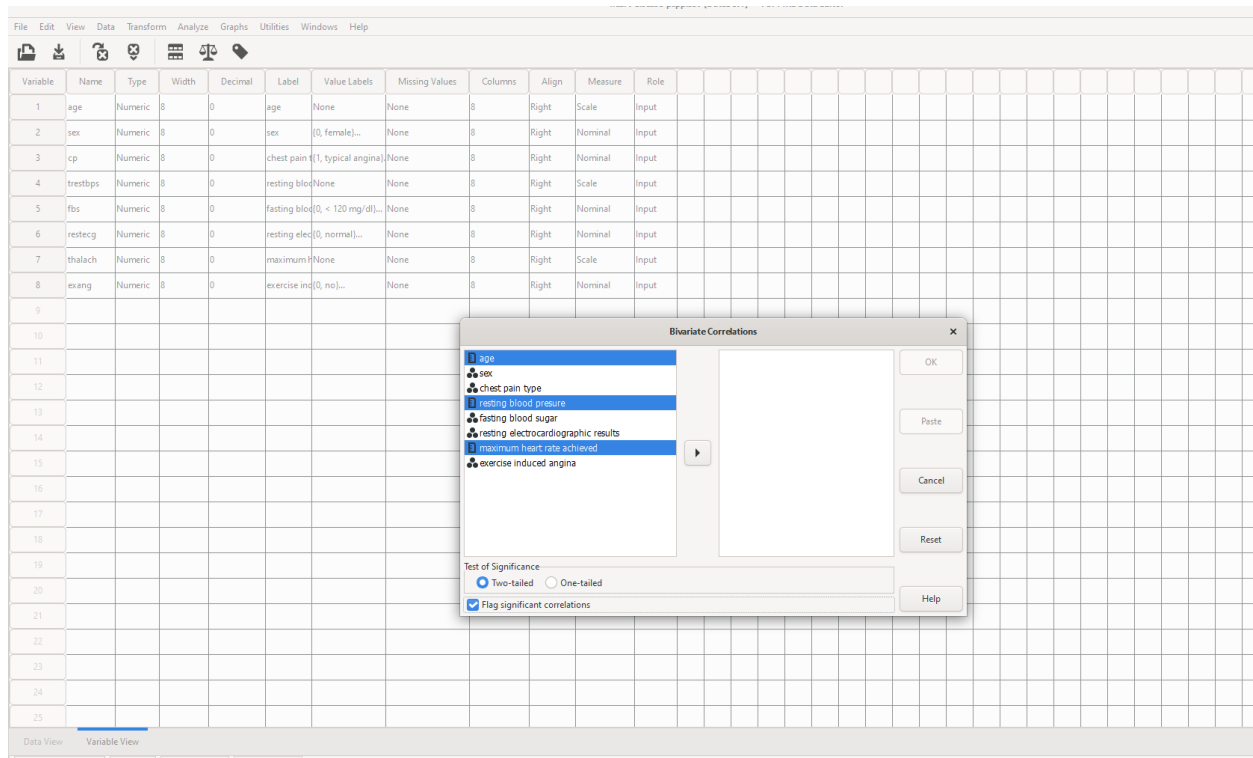
minima de 60, ceea ce face ca

amplitudinea sa aiba valoarea de 122.

Aplatizarea este de -0.22 si asimetria fiind de -0.07 .

3. Corelatii

Comanda se executa astfel: Analyze -> Bivariate correlate. Se bifeaza two-tailed si flag significant correlations.



Correlations

		age	resting blood pressure	maximum heart rate achieved
age	Pearson Correlation	1.000	.361 _a	-.275 _a
	Sig. (2-tailed)		.000	.002
	N	123	123	123
resting blood pressure	Pearson Correlation	.361 _a	1.000	-.213 _a
	Sig. (2-tailed)	.000		.018
	N	123	123	123
maximum heart rate achieved	Pearson Correlation	-.275 _a	-.213 _a	1.000
	Sig. (2-tailed)	.002	.018	
	N	123	123	123

a. Significant at .05 level

Interpretare pt **age-resting blood pressure**: exista corelatie directa (semn algebric pozitiv) de intensitate slaba ($0.20 < 0.361 < 0.39$) semnificativa din punct de vedere statistic pentru toti pacientii din studiu ($p\text{-value} = 0.000$).

Interpretare pt **age-maximum heart rate achieved**: exista corelatie indirecta (semn algebric negativ) de intensitate slaba ($0.20 < 0.275 < 0.39$) semnificativa din punct de vedere statistic pentru 99.8% pacientii din studiu ($p\text{-value} = 0.002$).

Interpretare pt **resting blood pressure-maximum heart rate achieved**: exista corelatie indirecta (semn algebric negativ) de intensitate slaba ($0.20 < 0.213 < 0.39$) semnificativa din punct de vedere statistic pentru 97.2% pacientii din studiu ($p\text{-value} = 0.018$).

4. Odds Ratio și Risk Ratio

Comanda se executa astfel: Odd Ratio: Analyze -> Descriptive statistics -> Crosstab. In Crosstab, la Statistics se bifeaza "chi square" si "risk"

Nu a iesit nimic.

5. Chi Square bivariat (pentru tabele de contingență 2 x 2 sau k x 2)

Comanda se executa astfel: Analyze -> Descriptive statistics -> Crosstab. In Crosstab, la Statistics se bifeaza "chi square"

Pentru un $p\text{-value} > 0.05$ nu vom face interpretari caci $p\text{-value} > 0.05$ inseamna ca nu exista o semnificatie statistica.

De exemplu: pentru "resting ecg x sex":

Chi-Square Tests			
	Value	df	Asymptotic Sig. (2-tailed)
Pearson Chi-Square	1,07	2	,585
Likelihood Ratio	1,11	2	,574
Linear-by-Linear Association	,71	1	,399
N of Valid Cases	123		

Avem $p\text{-value}=0.585$ care e mult mai mare decat 0.05 deci nu avem semnificatie statistica.

Dar, pe de alta parte, pentru "resting blood pressure x sex":

Chi-Square Tests			
	Value	df	Asymptotic Sig. (2-tailed)
Pearson Chi-Square	37,84	19	,006
Likelihood Ratio	45,55	19	,001
Linear-by-Linear Association	10,70	1	,001
N of Valid Cases	123		

TODO: PUNE TABELU DE CROSTAB care insoteste.

Avem un $p\text{-value} < 0.05$ ceea ce inseamna ca testul este semnificativ statistic.

Pentru aceasta situatie, ipoteza nula este:

H_0 = Nu exista diferente semnificative statistic la valorile de presiunea a sangelui in repaus in functie de sex.

Pentru a accepta sau respinge H_0 se compara valoarea calculata a statisticii chisquare (in cazul nostru 37.84) cu valoarea teoretica din tabele de distributie teoretica a statisticii chi square pentru 1 grad de libertate (df = degree of freedom) si $p\text{-value}=0.000$.

Critical values of chi-square (right tail)

Degrees of freedom (df)	Significance level (α)							
	.99	.975	.95	.9	.1	.05	.025	.01
1	-----	0.001	0.004	0.016	2.706	3.841	5.024	6.635

Cum chi square calculat > chi square teoretic -> se respinge H_0 deci exista diferente semnificativ statistic la valorile de presiunea a sangelui in repaus in functie de sex.

6. Regresie liniara simpla

Comanda se executa astfel:
Analyze -> Regression -> Linear

Se selectează variabila dependentă/efect (una singura) și una sau mai multe variabile indepedente/cauze.

Model Summary (resting blood pressure)

R	R Square	Adjusted R Square	Std. Error of the Estimate
,38	,15	,10	21,35

ANOVA (resting blood pressure)

	Sum of Squares	df	Mean Square	F	Sig.
Regression	8977,84	6	1496,31	3,28	,005
Residual	52895,33	116	455,99		
Total	61873,17	122			

Coefficients (resting blood pressure)

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	138,56	15,20	,00	9,12	,000
sex	11,00	4,03	,24	2,73	,007
chest pain type	-1,20	3,01	-,04	-,40	,692
fasting blood sugar	1,82	5,40	,03	,34	,736
resting electrocardiographic results	6,59	3,41	,17	1,93	,056
maximum heart rate achieved	-,12	,08	-,14	-1,51	,134
exercise induced angina	4,29	4,13	,10	1,04	,300

Coefficientul R square indica faptul ca acest model este valid doar pentru 15% din subiecti. Insa modelul este semnificativ statistic conform tabelului ANOVA care indica un p-value de $0.005 < 0.05$.

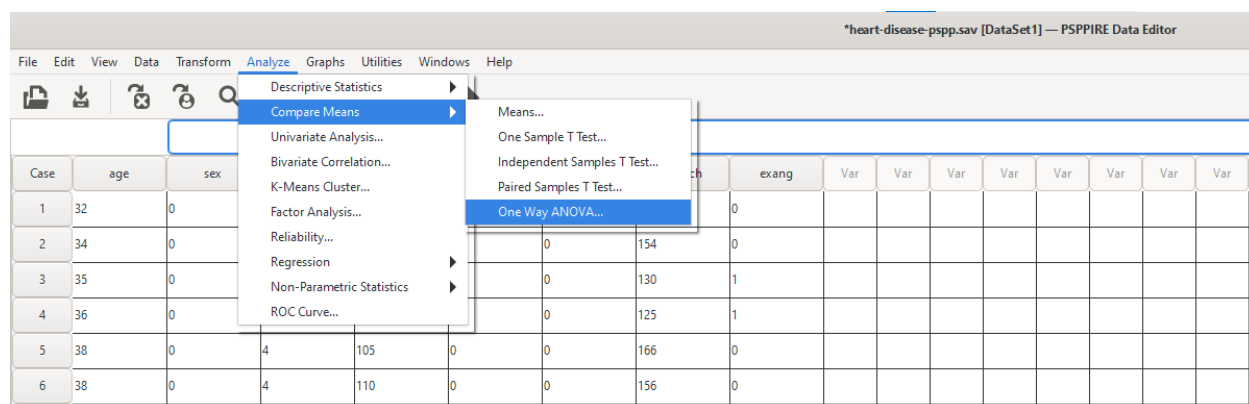
Din tabelul de coeficienti se deduce faptul ca doar variabila sex are o contributie semnificativa in modelul de regresie fiindca are p-value $0.007 < 0.05$. O alta variabila care s-ar apropia de aceasta valoare este "resting ecg".

Tot pe baza informatiilor din ultimul tabel din coloana "Unstandardized coefficients" putem scrie ecuatia modelului de regresie luand fiecare variabila si inmultind-o cu fiecare coeficient:

Resting blood pressure = $11 * \text{sex} - 1.2 * \text{chest pain type} + 1.82 \text{ fasting blood sugar} + 6.59 * \text{resting ecg} - 0.12 \text{ maximum heart rate achieved} + 4.29 \text{ exercise induced angina}$.

7. ANOVA

Comanda se executa astfel: Analyze -> Compare means -> One way ANOVA



```
ONEWAY /VARIABLES= trestbps thalach BY sex
/STATISTICS=DESCRIPTIVES .
```

Descriptives

	sex	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
resting blood presure	female	58	122,84	24,08	3,16	116,51	129,18	80,00	200,00
	male	65	136,15	19,12	2,37	131,42	140,89	105,00	200,00
	Total	123	129,88	22,52	2,03	125,86	133,90	80,00	200,00
maximum heart rate achieved	female	58	128,69	24,13	3,17	122,34	135,03	63,00	182,00
	male	65	115,46	26,04	3,23	109,01	121,91	60,00	176,00
	Total	123	121,70	25,92	2,34	117,07	126,33	60,00	182,00

ANOVA

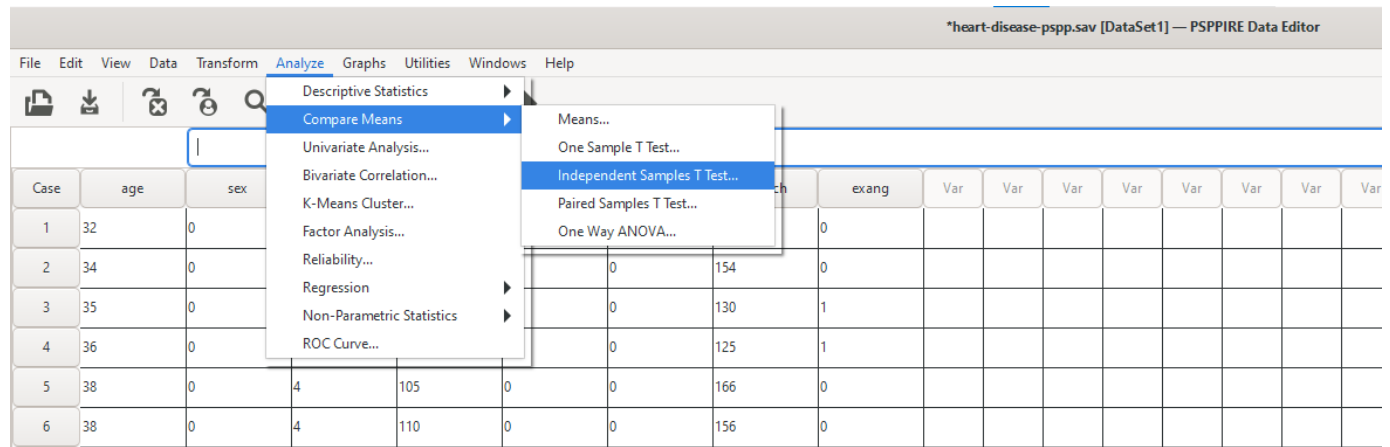
		Sum of Squares	df	Mean Square	F	Sig.
resting blood presure	Between Groups	5429,11	1	5429,11	11,64	,001
	Within Groups	56444,06	121	466,48		
	Total	61873,17	122			
maximum heart rate achieved	Between Groups	5363,30	1	5363,30	8,47	,004
	Within Groups	76592,57	121	633,00		
	Total	81955,87	122			

Primul tabel furnizat de output-ul aferent One way ANOVA, respectiv Descriptives oferă informații detaliate referitoare la indicatorii tendinței centrale și ai împrăștierii din fiecare grup aferent factorului de grupare (sexul masculin sau feminin).

Cel de-al doilea tabel ANOVA, ne indică faptul că cele două variabile resting blood pressure sau maximum heart rate achivied au medii semnificativ statistic diferite pentru fiecare dintre grupurile considerate. (avand p-value 0.001 si 0.004)

8. Testul t Student (pentru eșantioane independente)

Comanda se executa astfel: Analyze -> Compare means -> Independent Samples T Test



```
T-TEST /VARIABLES= sex fbs trestbps restecg  
/GROUPS=exang (0,1) /MISSING=ANALYSIS  
/CRITERIA=CI (0.95) .
```

Group Statistics

	Group	N	Mean	Std. Deviation	S.E. Mean
sex	no	68	,50	,50	,06
	yes	55	,56	,50	,07
fasting blood sugar	no	68	,18	,38	,05
	yes	55	,13	,34	,05
resting blood presure	no	68	127,65	22,17	2,69
	yes	55	132,64	22,85	3,08
resting electrocardiographic results	no	68	,41	,63	,08
	yes	55	,29	,53	,07

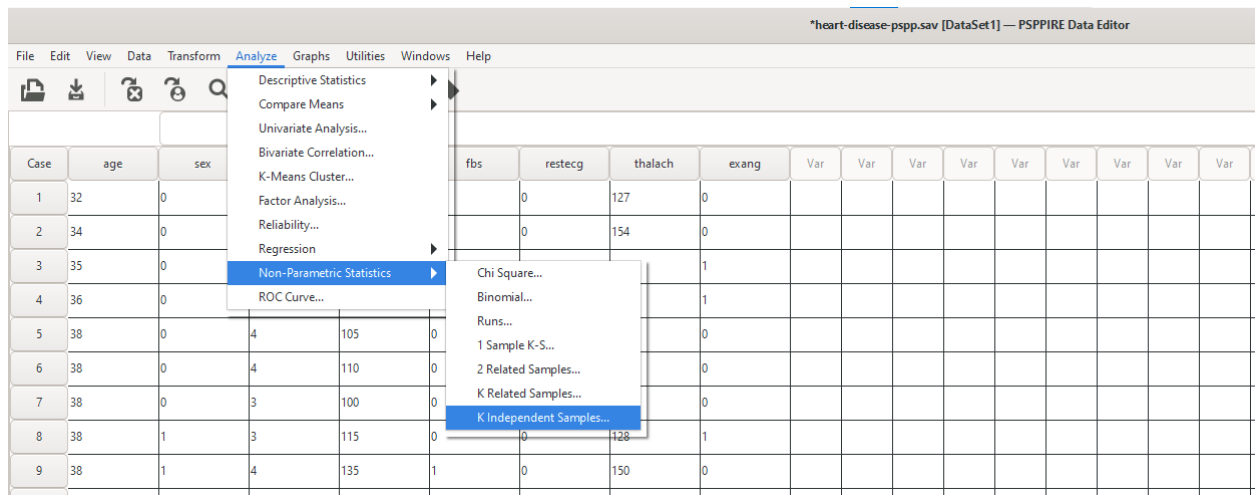
Independent Samples Test

		Levene's Test for Equality of Variances		T-Test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
sex	Equal variances assumed	1,10	,296	-,70	121,00	,486	-,06	,09	Lower	-,24
	Equal variances not assumed			-,70	115,98	,486	-,06	,09	Upper	,12
fasting blood sugar	Equal variances assumed	2,29	,133	,75	121,00	,457	,05	,07	Lower	-,08
	Equal variances not assumed			,76	120,20	,451	,05	,07	Upper	,18
resting blood presure	Equal variances assumed	,06	,800	-1,22	121,00	,223	-4,99	4,08	Lower	-13,06
	Equal variances not assumed			-1,22	114,18	,225	-4,99	4,09	Upper	3,11
resting electrocardiographic results	Equal variances assumed	4,04	,047	1,13	121,00	,259	,12	,11	Lower	-,09
	Equal variances not assumed			1,15	120,71	,251	,12	,10	Upper	,33

Din tabelul Independent Samples Test se deduce că doar pentru media variabilei “resting electrocardiographic results” s-a înregistrat diferențe semnificative statistic ($p\text{-value} < 0.05$) și influențează semnificativ (comparativ cu celelalte varriabile din tabel) “exercise induced angina”.

9. Test neparametric

Comanda se executa astfel: Analyze -> Non-parametric statistics -> K Independent Samples



Eroare.