

ANALIZA DATELOR

C. Băicuș, 2009

www.baicus.ro

Tipuri de variabile

- Variabila = orice calitate, caracteristică sau constituent al unei persoane sau lucru, care poate fi măsurat
- toată cercetarea medicală = studiul relațiilor dintre variabile
- cercetătorii medicali sunt interesați de studierea fie a asociațiilor, fie a diferențelor dintre variabile.

Tipuri de variabile

- **Variabile nominale**

- variabile sub formă de nume sau alte simboluri reprezentând categorii ce nu pot fi ordonate una în raport cu cealaltă
- numele, grupa sanguină, sexul, rasa, culoarea ochilor, diagnosticul etc.
- **variabile dihotomice** (binare, bimodale) = variabile ce nu pot lua decât două valori
 - mort/viu, fumător/nefumător, prezent/absent, normal/anormal - DA/NU

Tipuri de variabile

- **Variabile ordinale**

- variabile ce sunt clasificate în mai mult de două categorii și la care există o ordine naturală între categorii
- evoluția, stadializări, clasificări, scoruri etc.

Tipuri de variabile

- **Variabile cantitative**

- **continue**

- variabile cu un număr potențial infinit de valori de-a lungul unui continuum
 - înălțimea, greutatea, TA, vârsta etc.

- **discontinue (discrete)**

- variabile ce pot fi descrise numai prin unități întregi ce nu pot fi măsurate în intervale mai mici decât unitatea
 - AV, nr. copii etc.

STATISTICA DESCRIPTIVA

C. Băicuș

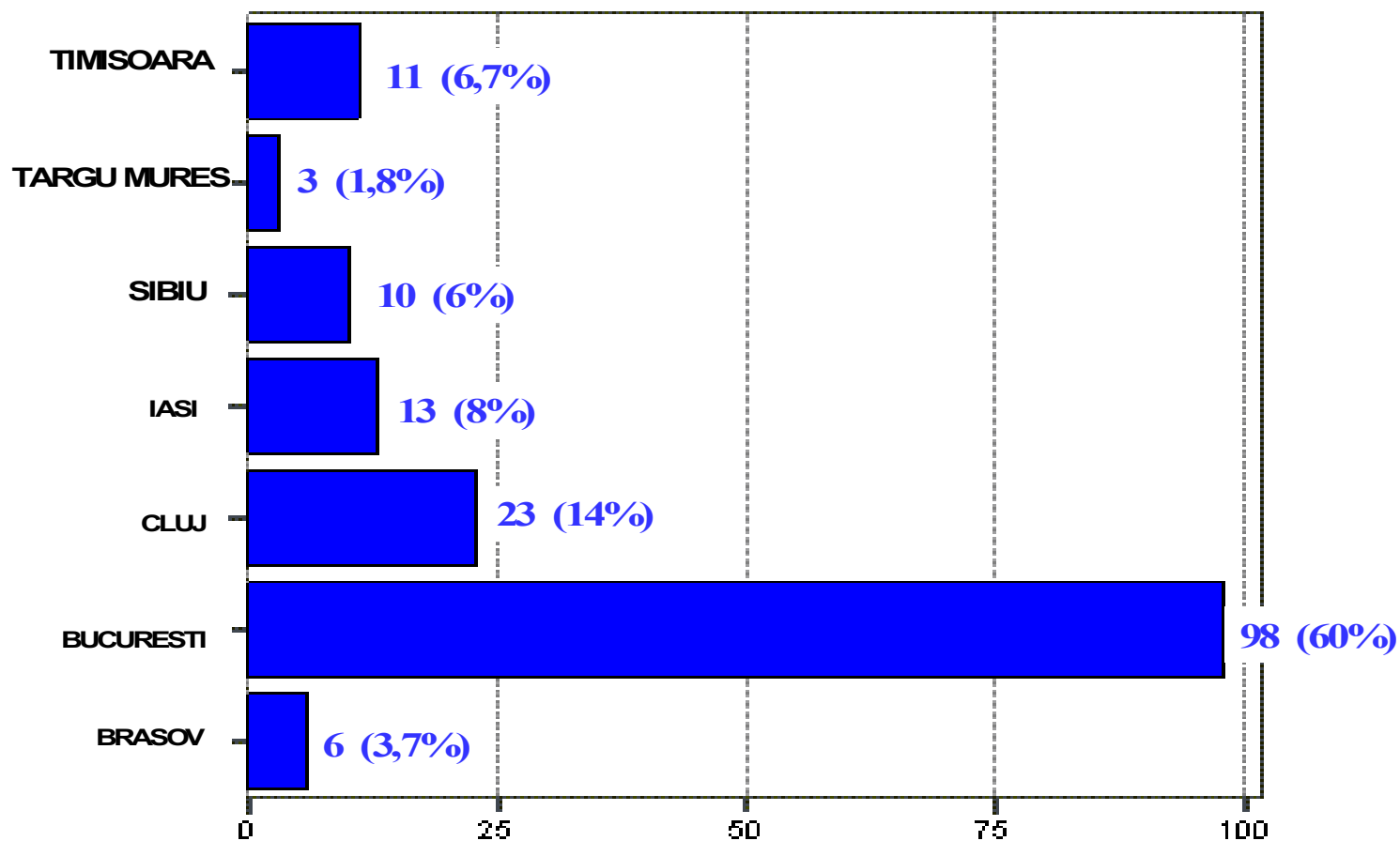
www.baicus.ro

Tipuri de variabile - statistică descriptivă

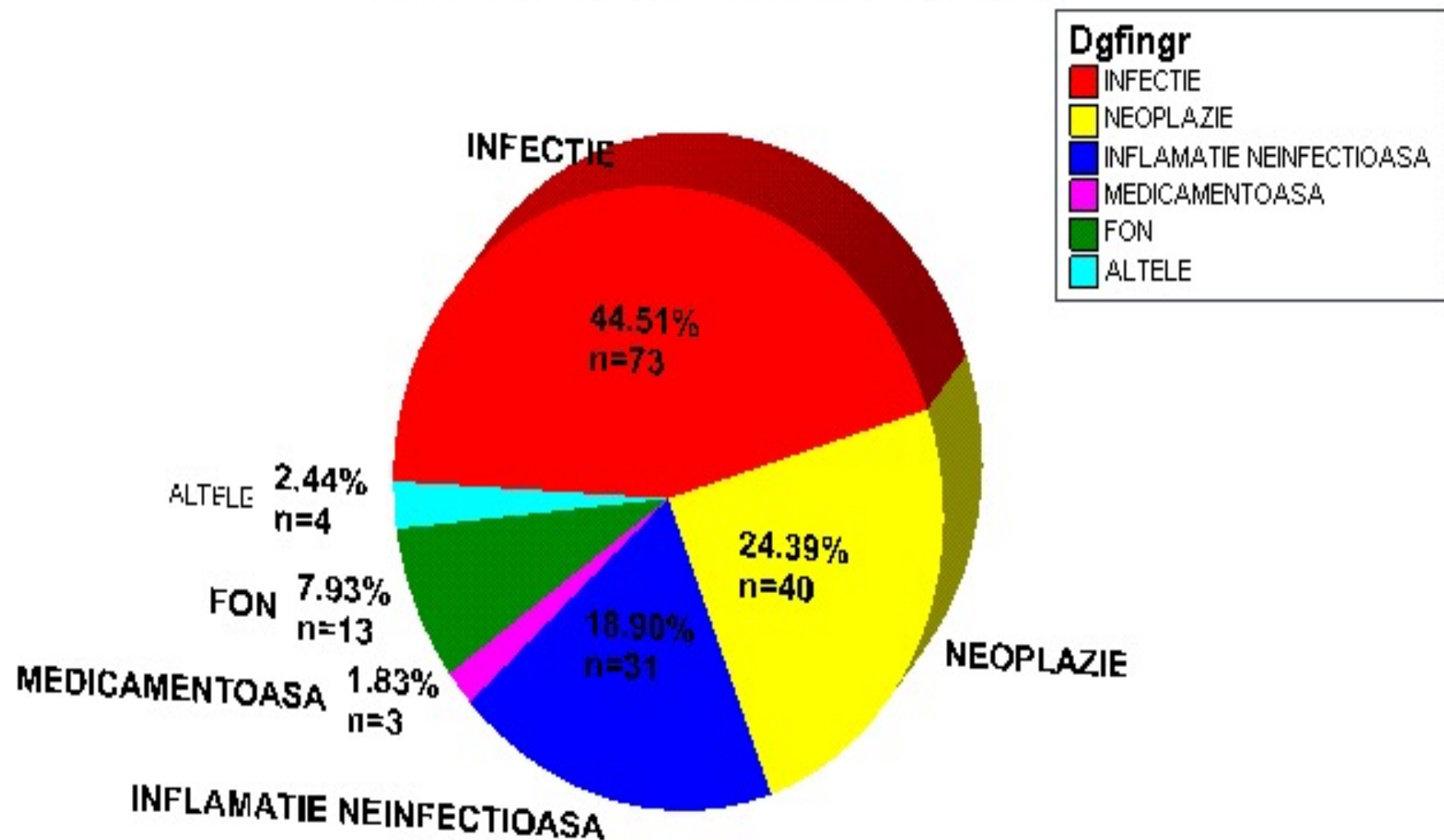
Variabile nominale

- tabel de frecvență (frecvența observațiilor pe categorii)
- grafic: bare, plăcintă
- modul (categoria cea mai frecventă)

Figura II.3. Distribuția pe orașe a cazurilor incluse în studiu



Etiologia FON pe grupe mari diagnostice



Tipuri de variabile - statistică descriptivă

Variabile ordinale

- tabel de frecvență, mediana, modul
- utilitatea mediei este discutabilă (intervalele dintre categorii nu sunt egale ca mărime)
- deseori, media este furnizată - se presupune că intervalele scalei sunt egale sau, cel puțin aproximativ egale

Clasa NYHA	Frecventa	Frecventa relativa	Frecventa cumulata
I	50	23,8%	23,8%
II	70	33,3%	57%
III	60	28,6%	85,7%
IV	30	14,3%	100%
TOTAL	210	100%	

Table 1 Clinical characteristics of the congestive heart failure outpatient cohort (n = 175)

Age (years)	71.1 (11.2)	←
Men	135 (77%)	
NYHA	2.7 (0.53)	
LVEF (%)	36.1 (12.1%)	←
Hyperlipidaemia	99 (55%)	
Smoking	65 (37%)	
Hypertension	101 (58%)	
Diabetes mellitus	62 (35%)	
Ischaemic heart disease	130 (74%)	
Chronic atrial fibrillation	52 (30%)	
TIA/CVA	20 (11%)	
PTCA or CABG	92 (53%)	
ACE inhibitors/ARBs	140 (80%)	
Spironolactone	77 (44%)	
β blockers	128 (73%)	

Data are mean (SD) or number (%).

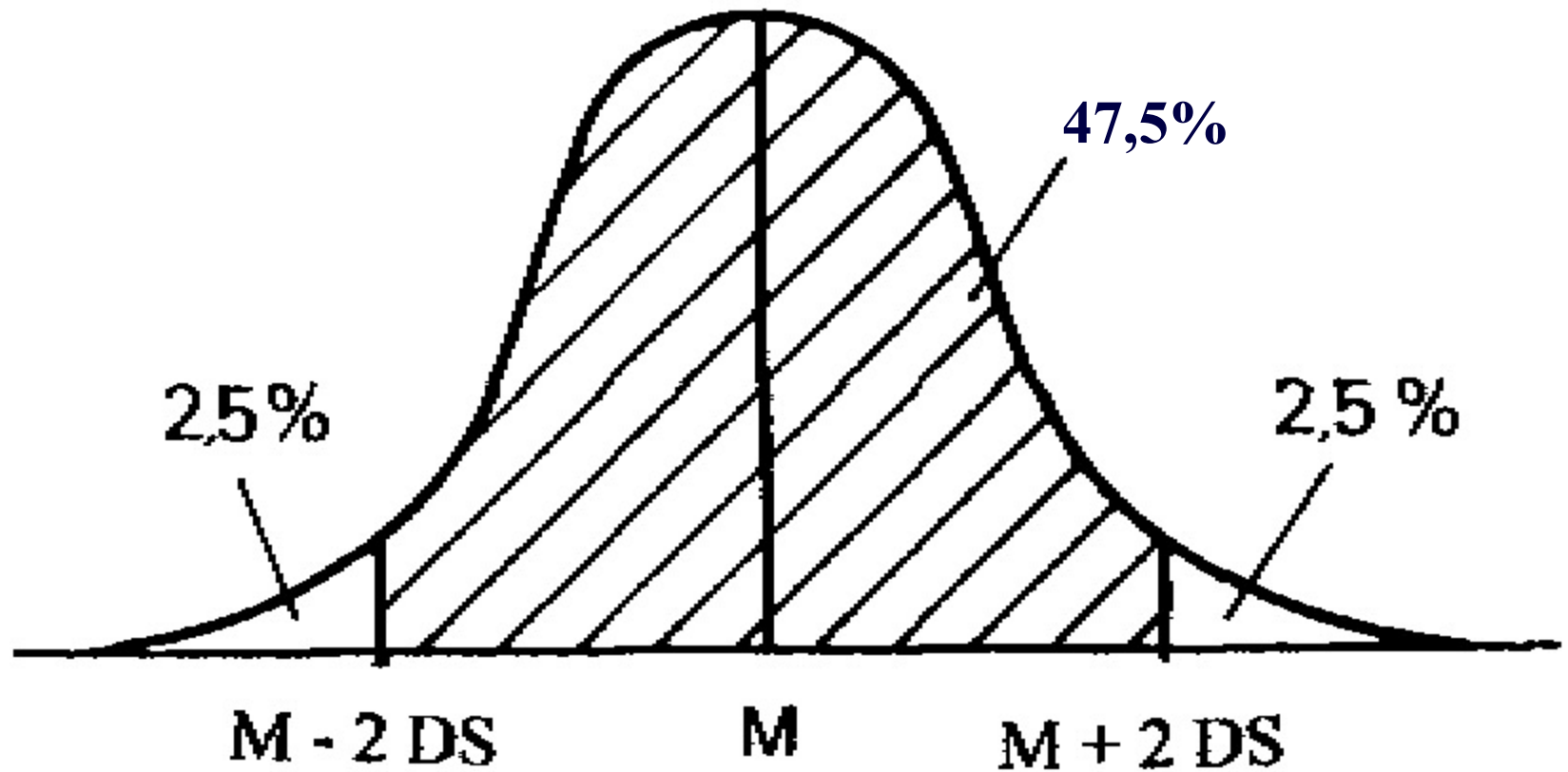
ACE, angiotensin-converting enzyme; ARB, angiotensinogen receptor blockers; CABG, coronary artery bypass surgery; CVA, cerebrovascular accident; LVEF, left ventricular ejection fraction; NYHA, New York Heart Association; PTCA, percutaneous transluminal coronary angioplasty; TIA, transient ischaemic attack.

George J et al.
Heart 2006;92:1420–1424.

Tipuri de variabile - statistică descriptivă

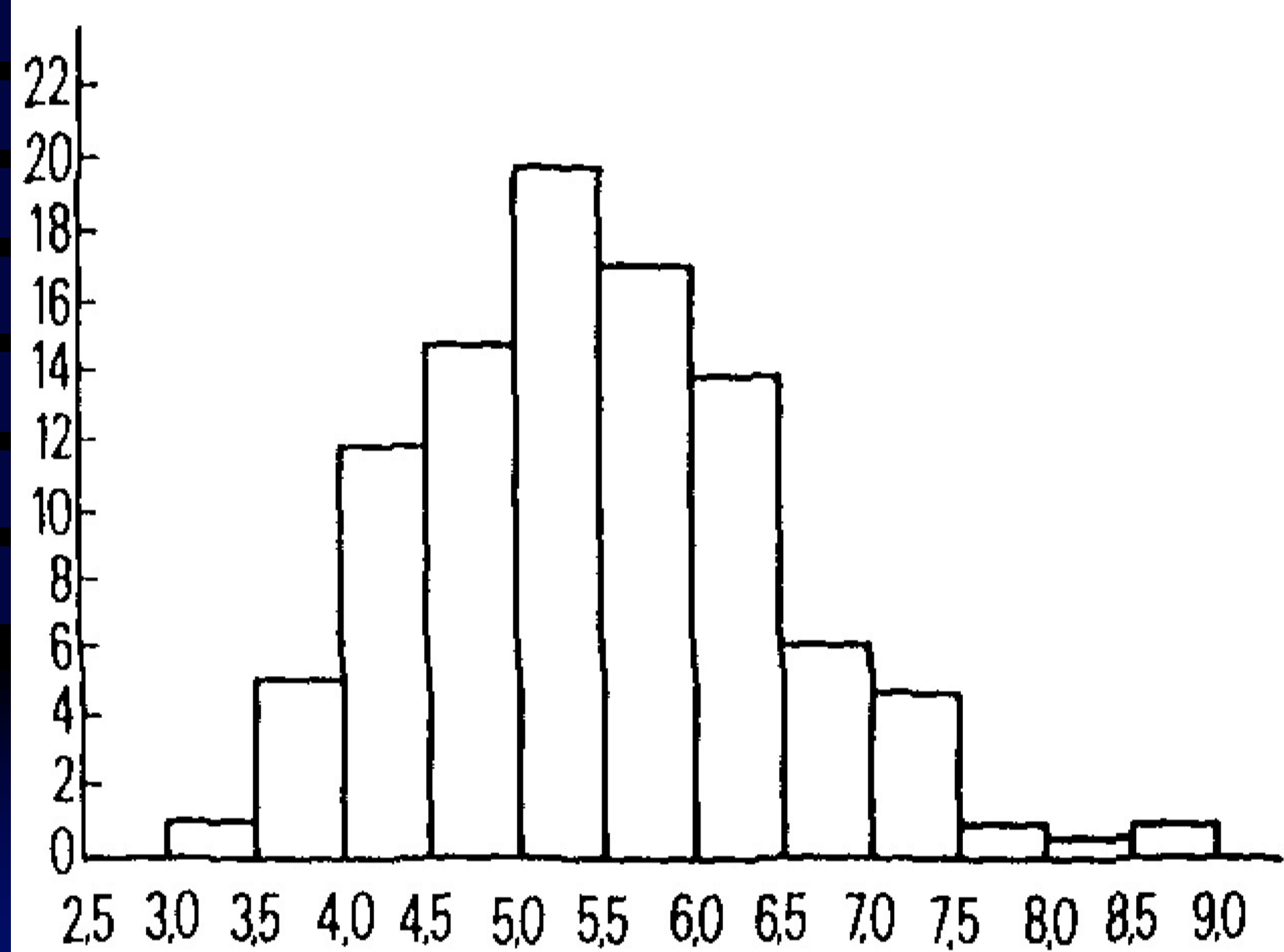
Variabile cantitative

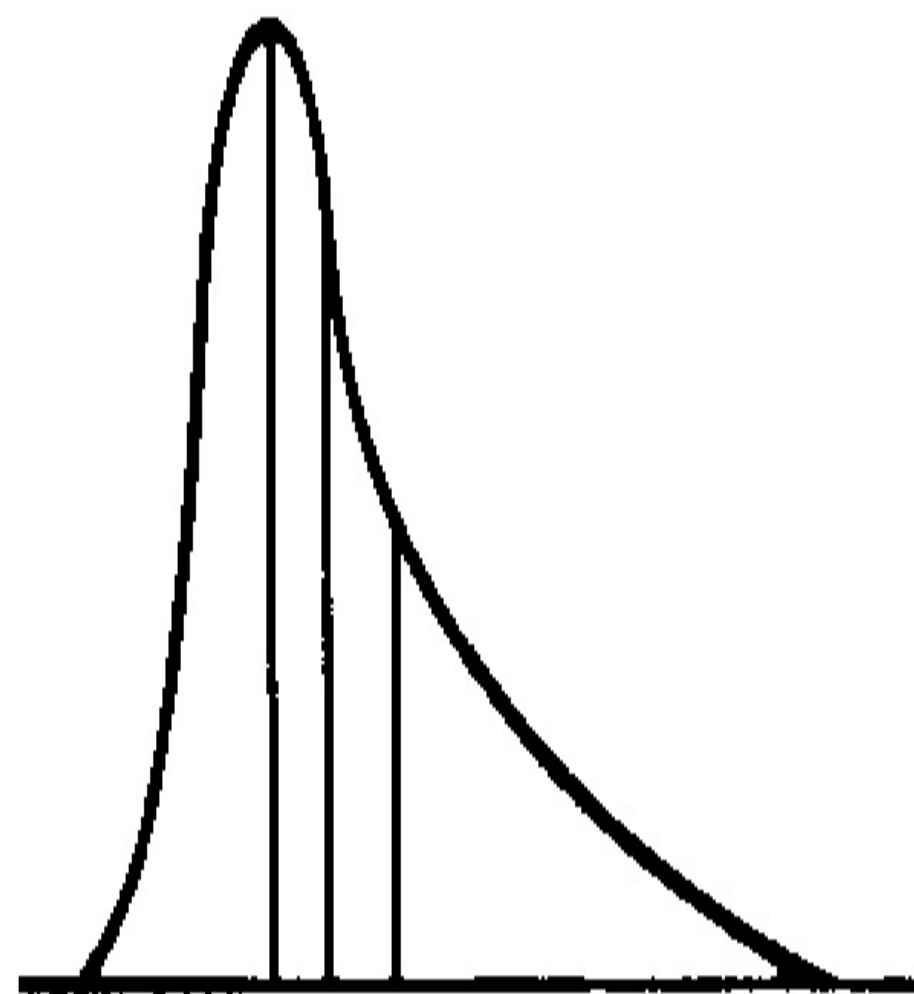
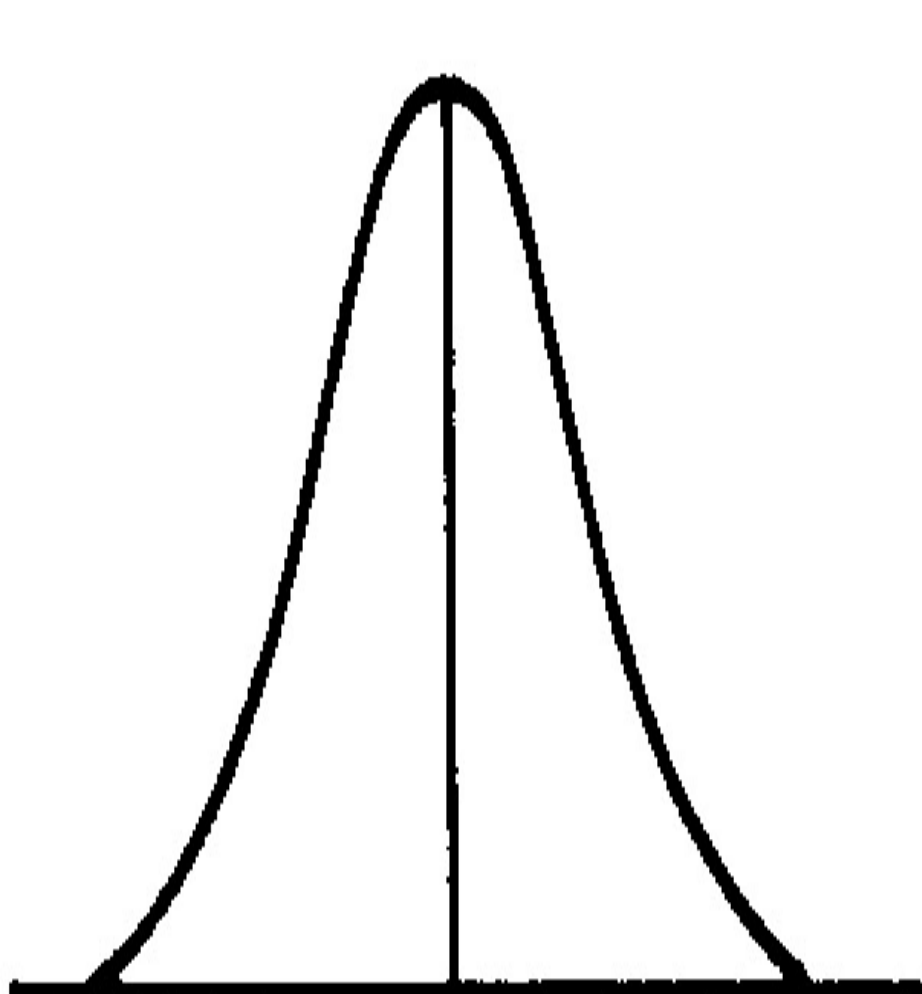
- măsura tendinței centrale (media, mediana, modul)
- măsura dispersiei (deviația standard, varianța)
- distribuția normală

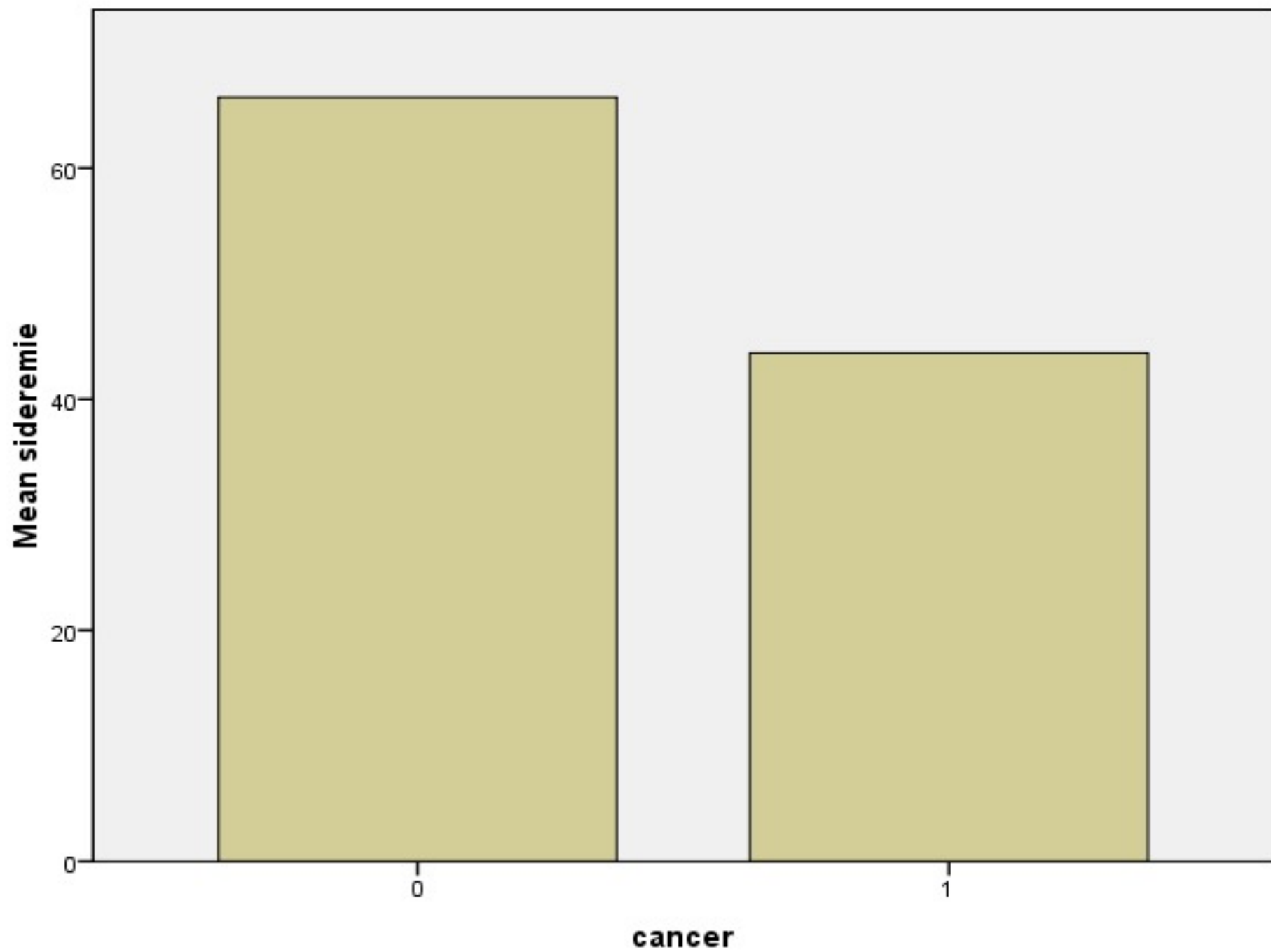


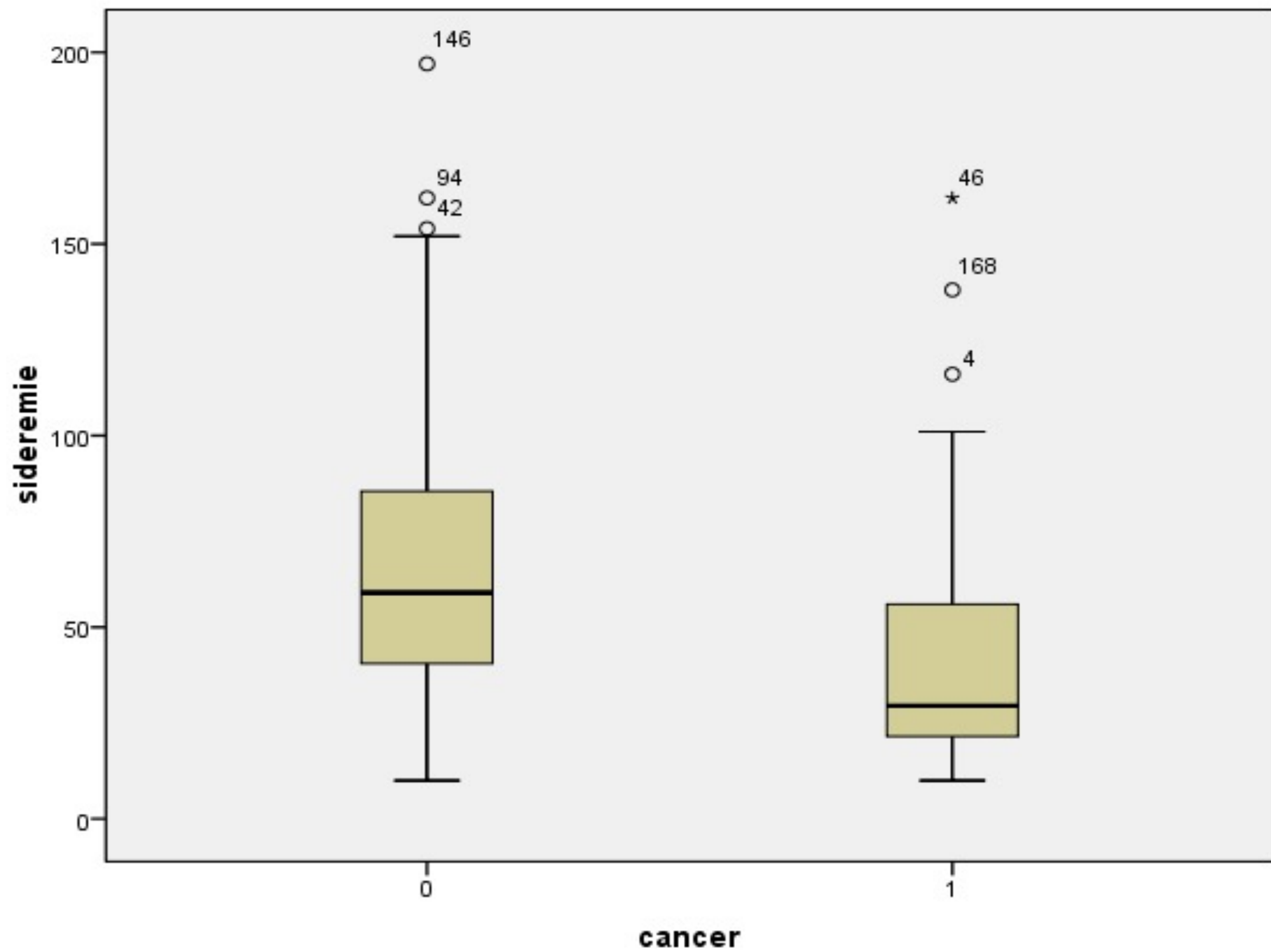
$$SD = \sqrt{\frac{\sum(x - \bar{x})^2}{n-1}}$$

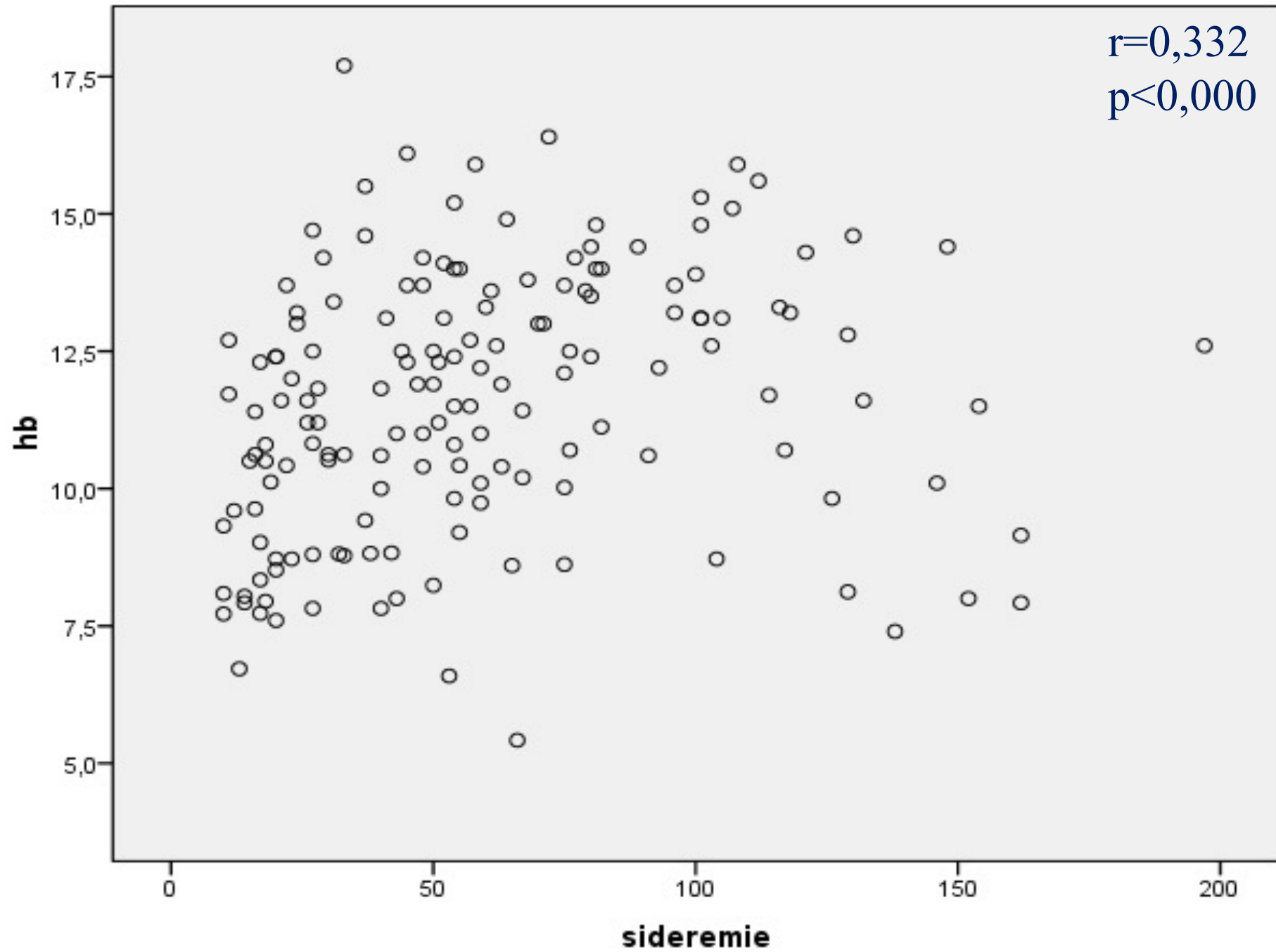
Acid uric (mg/100ml)	Frecvențe absolute	Frecvențe relative (%)	Frecvențe relative cumulate (%)
3,0-3,4	2	0,8	0,8
3,5-3,9	15	5,6	6,4
4,0-4,4	33	12,4	18,7
4,5-4,9	40	15,0	33,7
5,0-5,4	54	20,2	53,9
5,5-5,9	47	17,6	71,5
6,0-6,4	38	14,2	85,8
6,5-6,9	16	6,0	91,8
7,0-7,4	15	5,6	97,4
7,5-7,9	3	1,1	98,5
8,0-8,4	1	0,4	98,9
8,5-8,9	3	1,1	100,0

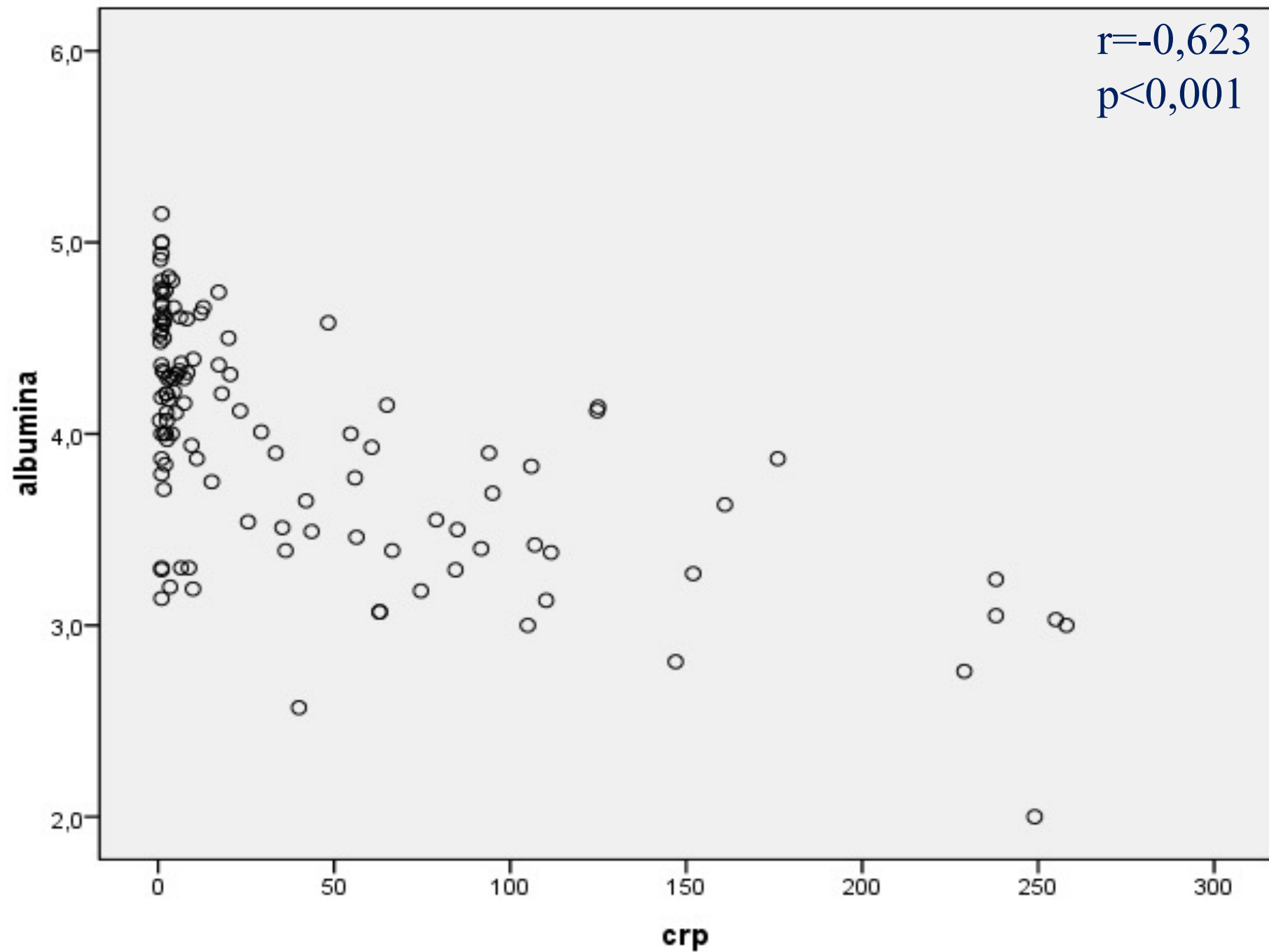


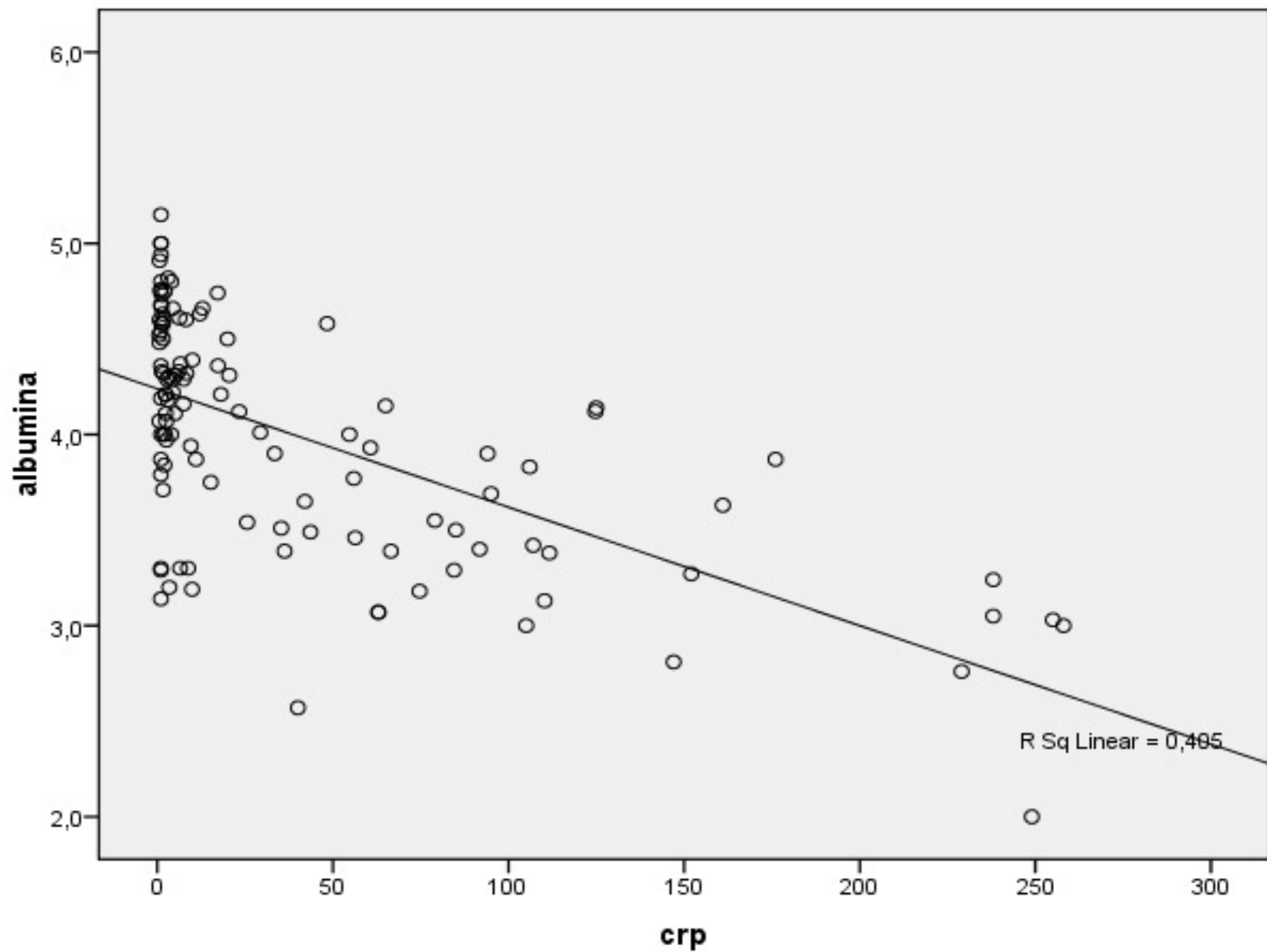


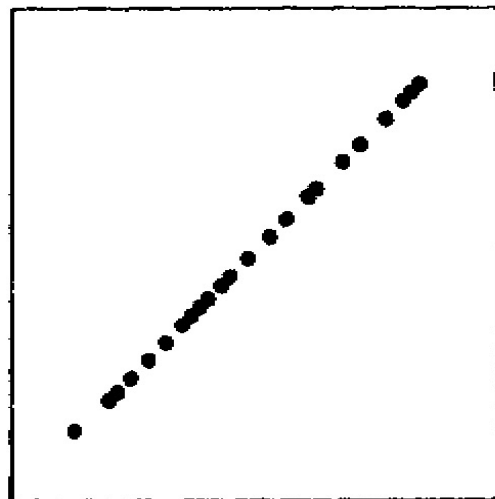






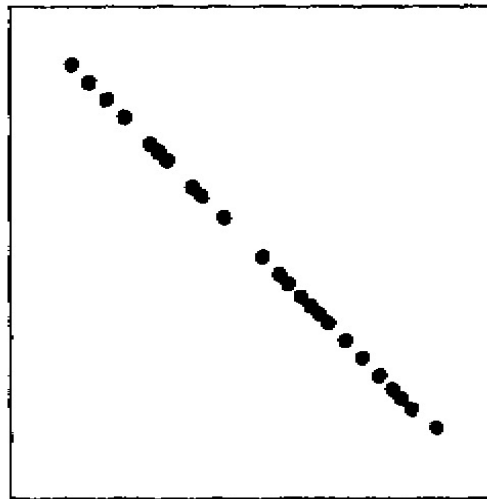






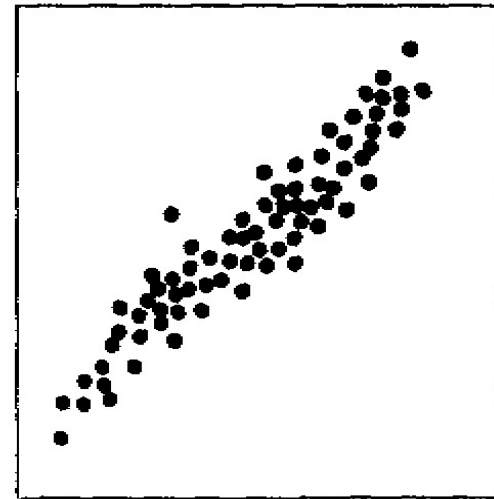
$r = 1.00$

A



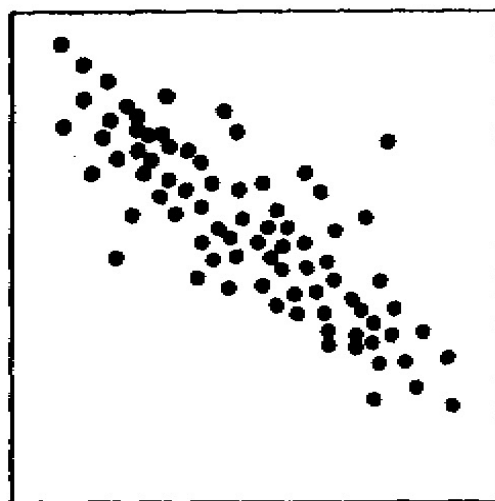
$r = -1.00$

B



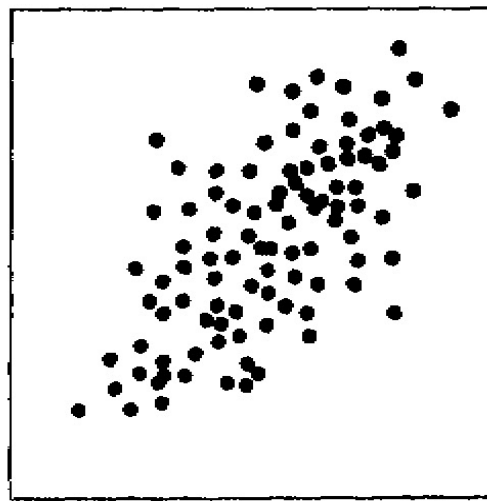
$r = 0.94$

C



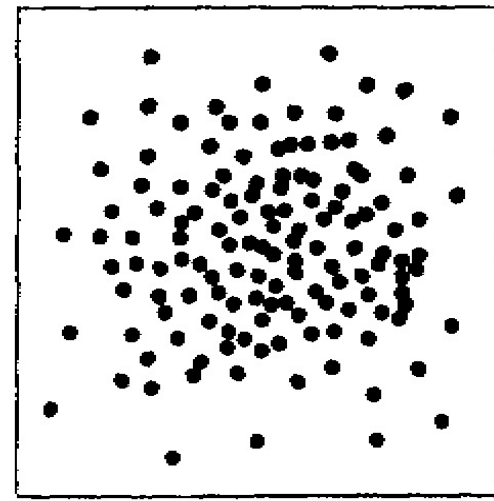
$r = -0.85$

D



$r = 0.44$

E



$r = 0.00$

F

Figura 8: Grafice ilustrând diferite grade de corelație lineară

- SD = deviatia standard
 - arata gradul de dispersie al valorilor
- SEM = eroarea standard a mediei
 - CI95% al mediei
 - $SEM = SD / \sqrt{n}$

STATISTICA ANALITICA

C. Băicuș

www.baicus.ro

p și intervalele de încredere

Dr. Cristian Băicuș

www.baicus.ro

populație

eșantionare

p, CI



eșantion

Incertitudinea statistica

- Incertitudinea (prezenta chiar intr-un eșantion reprezentativ) care apare atunci când folosim datele eșantionului pentru a trage concluzii asupra populației

Măsurarea incertitudinii statistice

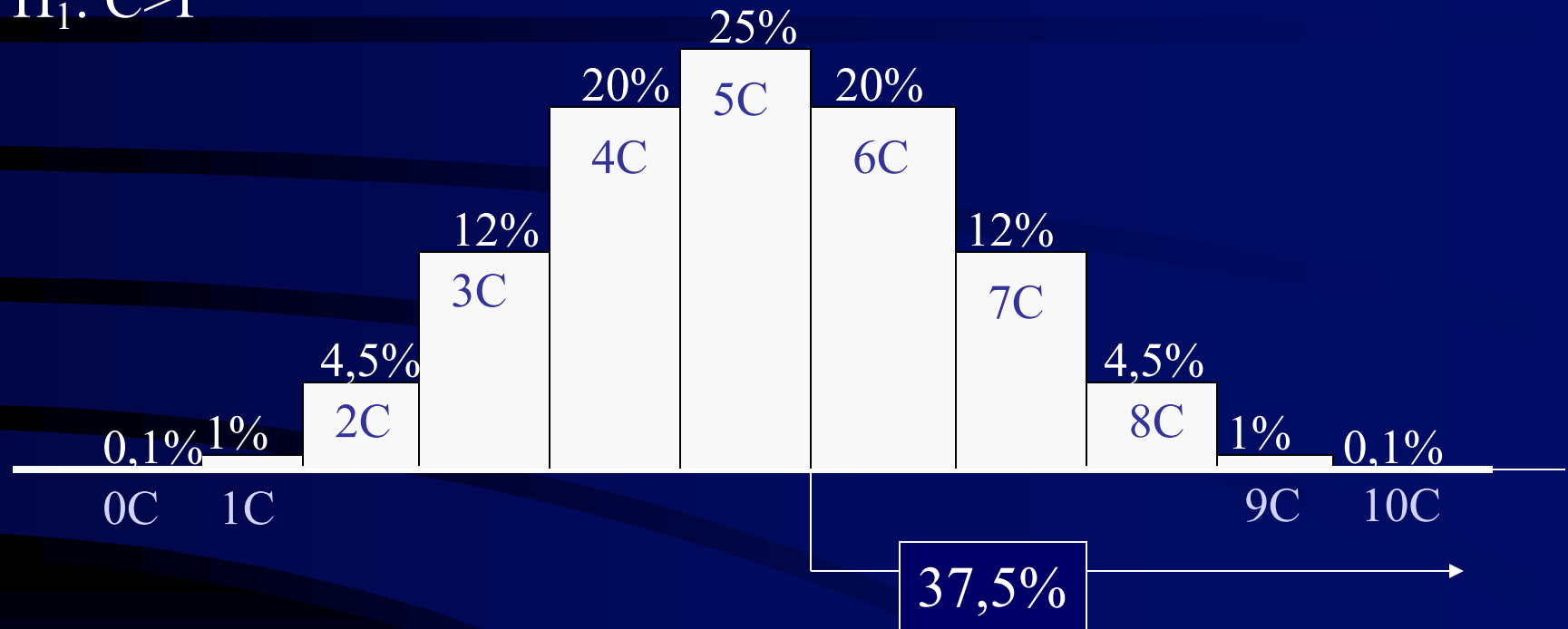
- Incertitudinea statistica poate fi cuantificata prin:
 - testarea ipotezei (cu calculul valorii p)
 - calculul intervalelor de incredere (CI)

p

10x10.000

$H_0: C=P$

$H_1: C>P$



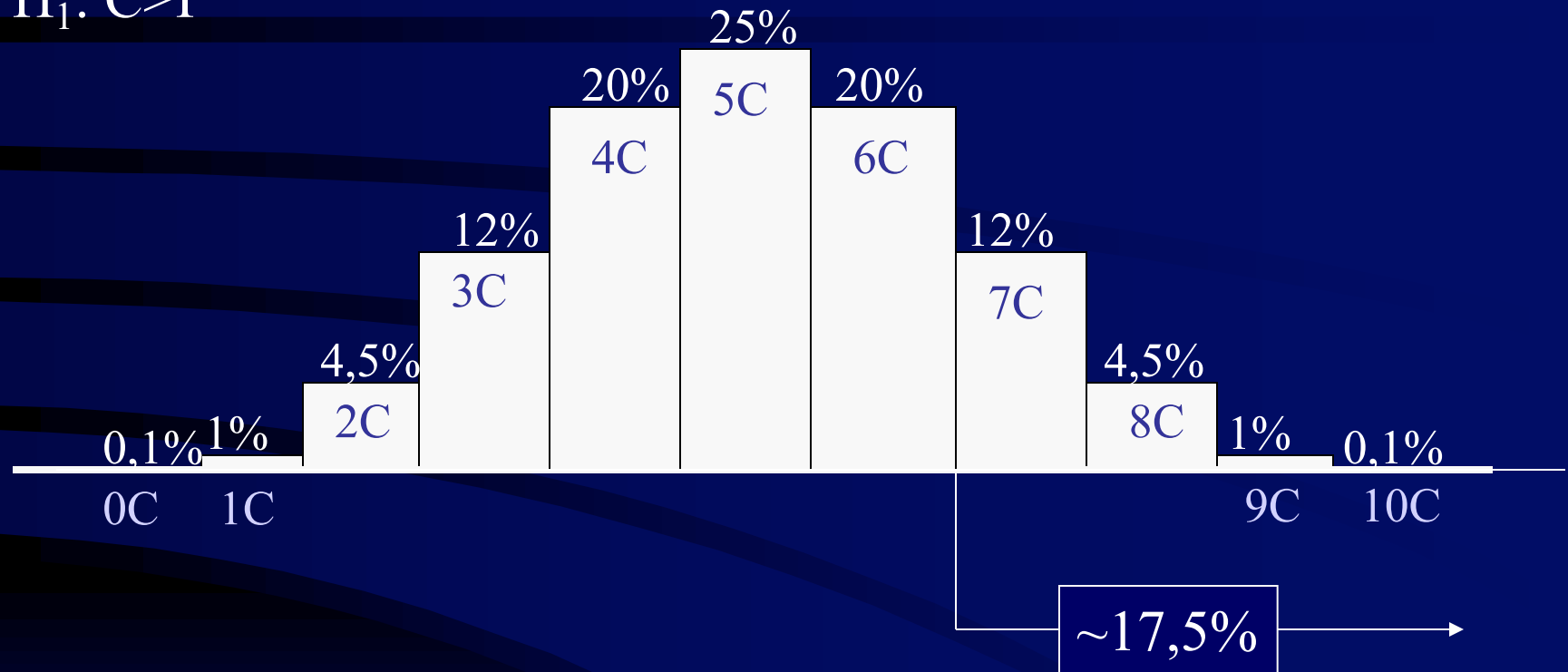
$p=0,375$

unidirectional, *one tailed (one sided)*

p

$H_0: C=P$

$H_1: C>P$



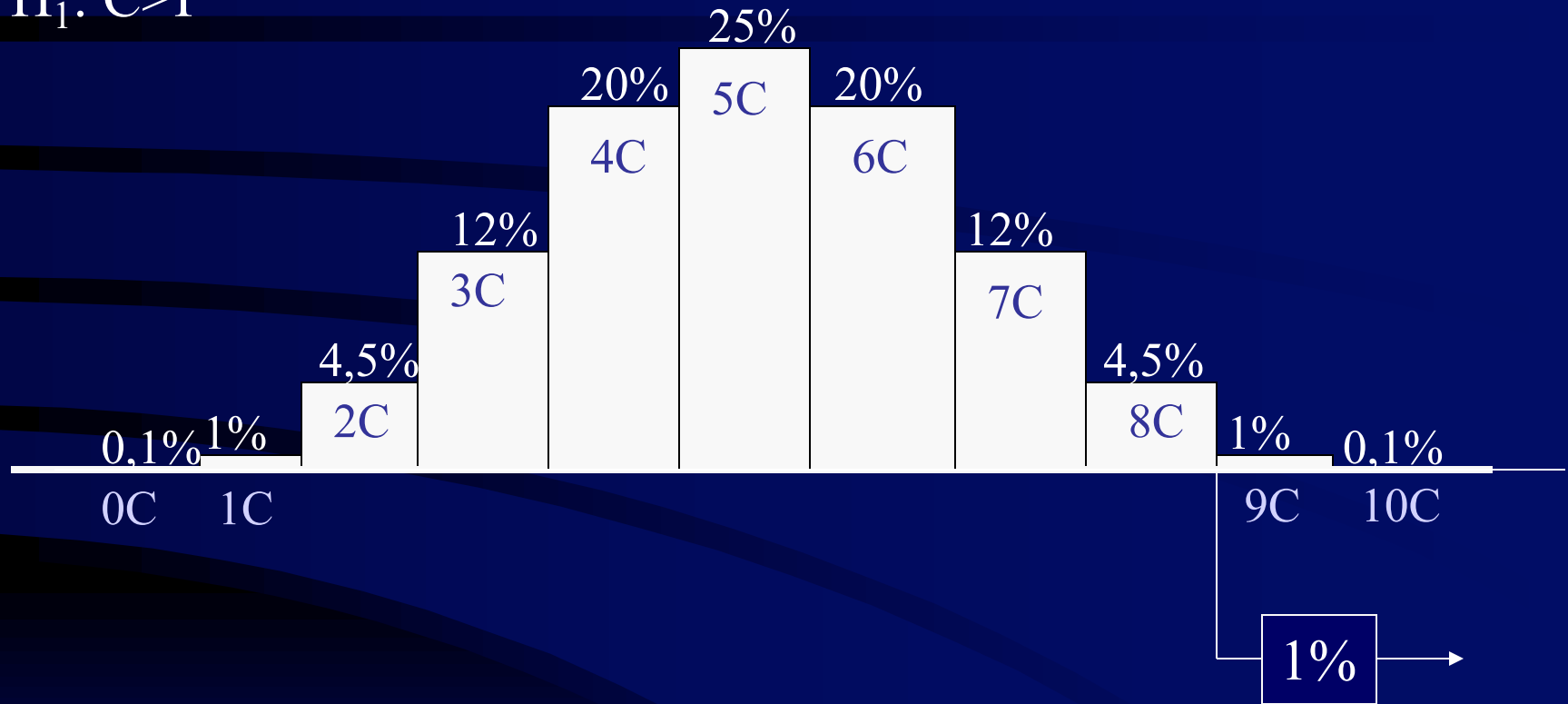
$p=0,175$

unidirectional, *one tailed (one sided)*

p

$H_0: C=P$

$H_1: C>P$



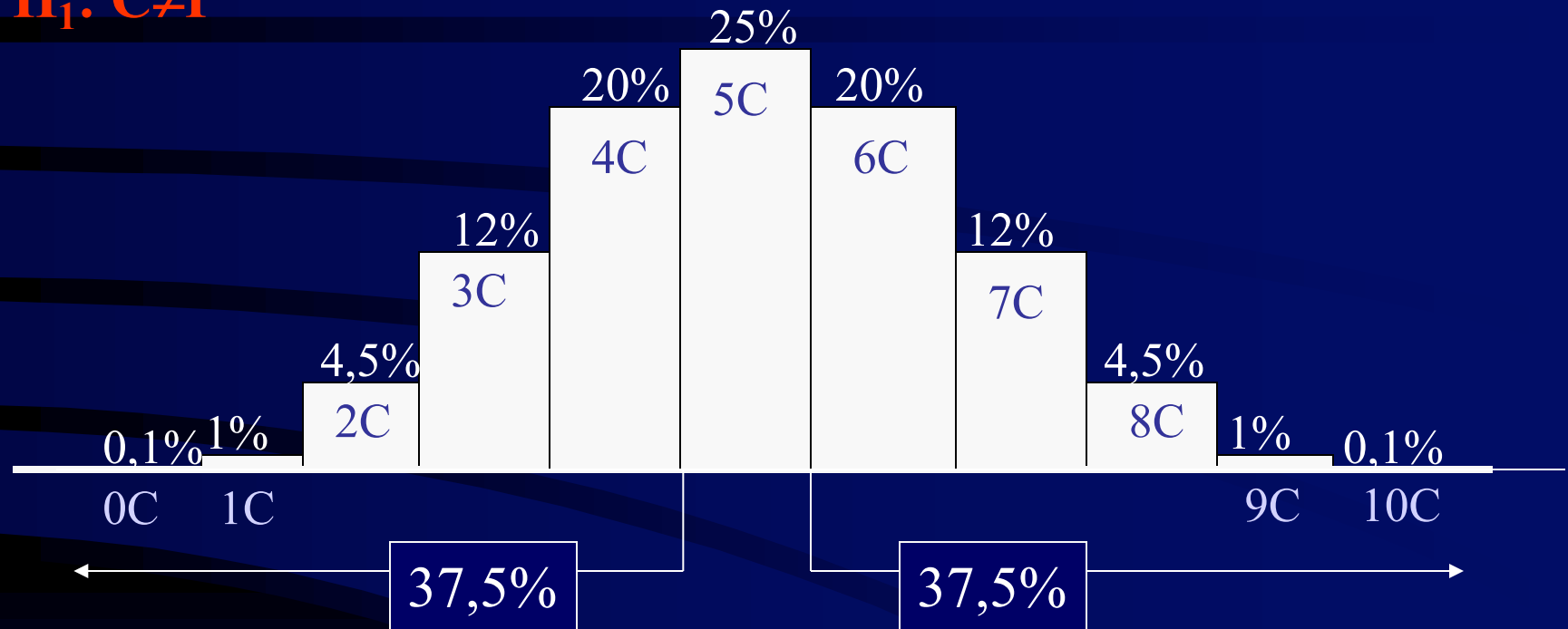
$p=0,01$

unidirectional, *one tailed (one sided)*

p

$H_0: C=P$

$H_1: C \neq P$



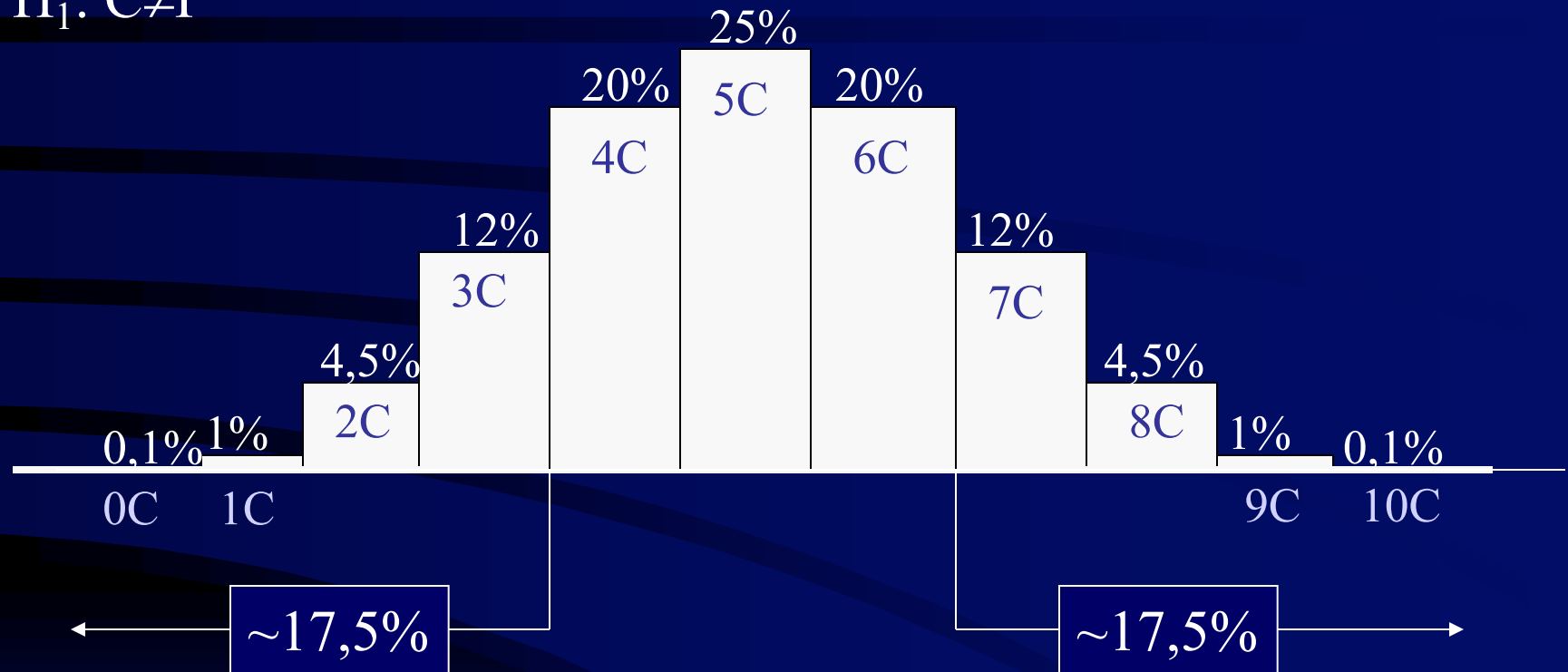
$p=0,75$

bidirectional, *two tailed (two sided)*

p

$H_0: C=P$

$H_1: C \neq P$



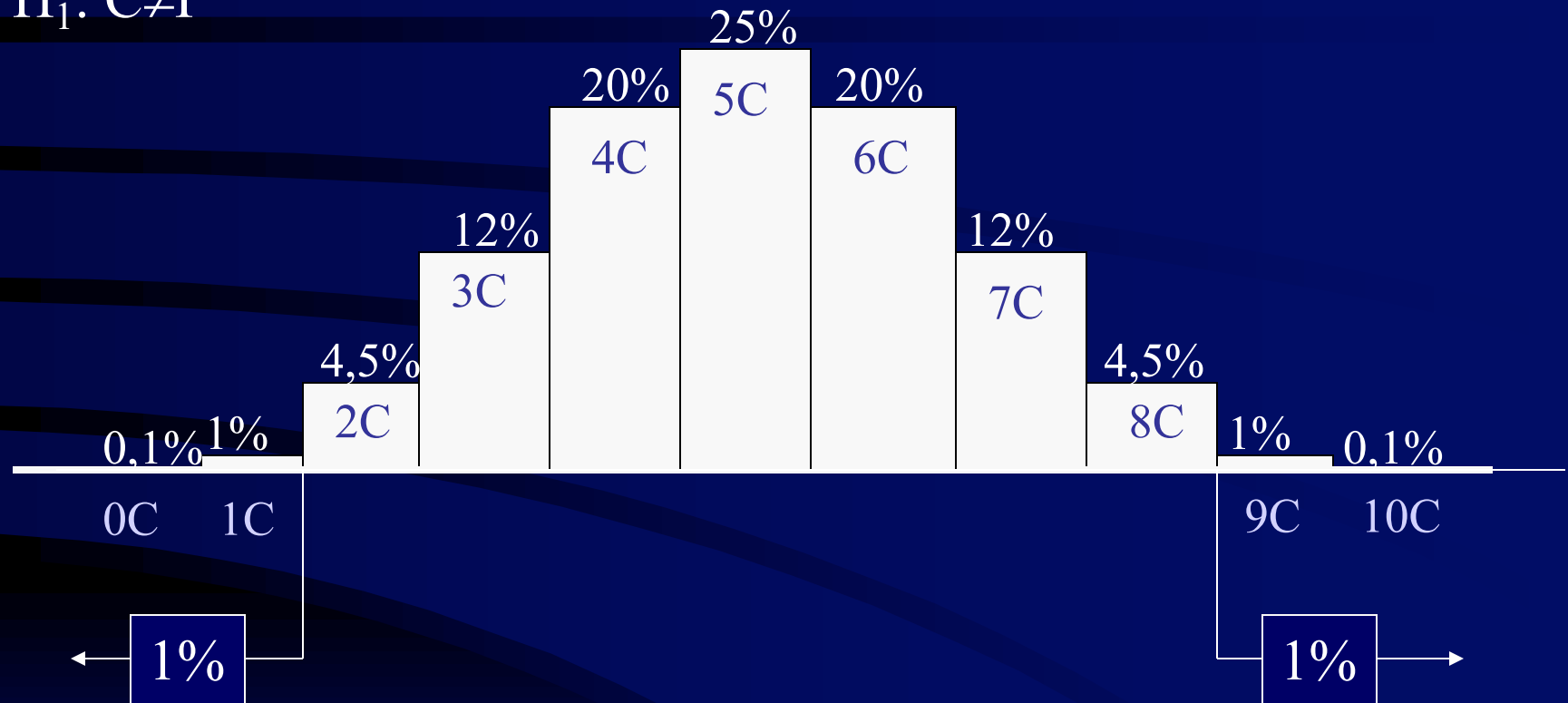
$p=0,35$

bidirectional, two tailed (two sided)

p

$H_0: C=P$

$H_1: C \neq P$



$p=0,02$

bidirectional, *two tailed (two sided)*

Studiu terapeutic

Placebo	Tratament
500 pts	500 pts
23	?

Studiu terapeutic

Placebo	Tratament
500 pts	500 pts
23	23

Studiu terapeutic

Placebo	Tratament
500 pts	500 pts
23	23
25	21

$p=0.64$

Studiu terapeutic

Placebo	Tratament
500 pts	500 pts
23	23
25	21
29	17

$p=0.09$

Studiu terapeutic

Placebo	Tratament
500 pts	500 pts
23	23
25	21
29	17
30	15

$p=0.03$

CI

<i>Basescu</i>	<i>Procent</i>	<i>CI</i>
6/10	60%	[27 , 86]

CI

<i>Basescu</i>	<i>Procent</i>	<i>CI</i>
6/10	60%	[27%, 86%]
24/40	60%	[43%, 74%]

CI

<i>Basescu</i>	<i>Procent</i>	<i>CI</i>
6/10	60%	[27% 86%]
24/40	60%	[43% 74%]
120/200	60%	[53%, 67%]

CI

<i>Basescu</i>	<i>Procent</i>	<i>CI</i>
6/10	60%	[27% 86%]
24/40	60%	[43% 74%]
120/200	60%	[53% 67%]
240/400	60%	[55% 65%]
1200/2000	60%	[58% 62%]

p și CI

- **p** : probabilitatea ca diferența dintre tratamente observată în studiul nostru să fie datorată întâmplării, iar în realitate (la nivelul populației) să nu existe nici o diferență între tratamente.
- **interval de încredere**: intervalul în care se află parametrul respectiv în realitate (la nivelul populației) (cu o probabilitate de 95%).

RR, OR, RRR, RRA, NNT

- Selecție întâmplătoare dintr-o populație
(sau cel puțin sunt reprezentativi
pentru acea populație);
- fiecare subiect este selectat
independent de ceilalți.

p și CI

Studiu terapeutic, $RR=0,8$

- dacă la nivelul populației nu există nici o diferență între mortalitățile la cei tratați și cei netratați, probabilitatea ca în studiul (eșantionul) nostru să apară un $RR \leq 0,8$ este p
- la nivelul populației (în realitate), $RR \in [0.5, 0.9]$.

p și CI

Studiu etiologic de cohortă, $RR=7$

- dacă la nivelul populației nu există nici o diferență între incidențele cancerului pulmonar la fumători și nefumători, probabilitatea ca în studiul (eșantionul) nostru să apară un $RR \geq 7$ este p
- la nivelul populației (în realitate), $RR \in [3, 15]$.

- intervalul de încredere dă aceleași informații ca și p (de fapt chiar mai multe)

Nesemnificativ statistic:

atunci când CI include

- **1 (pt. RR, OR)**
- **0 (pt. RRR sau RRA)**
- **∞ (pt. NNT)**

- nesemnificativ statistic = H_0 nu a putut fi înlăturată, nu că aceasta este adevărată
- nu există diferență, sau studiul nu a avut puterea necesară pentru a evidenția această diferență
⇒ atenție la CI

CI

semnificație statistică



semnificație clinică

Variabile numerice cu distributie neGaussiana

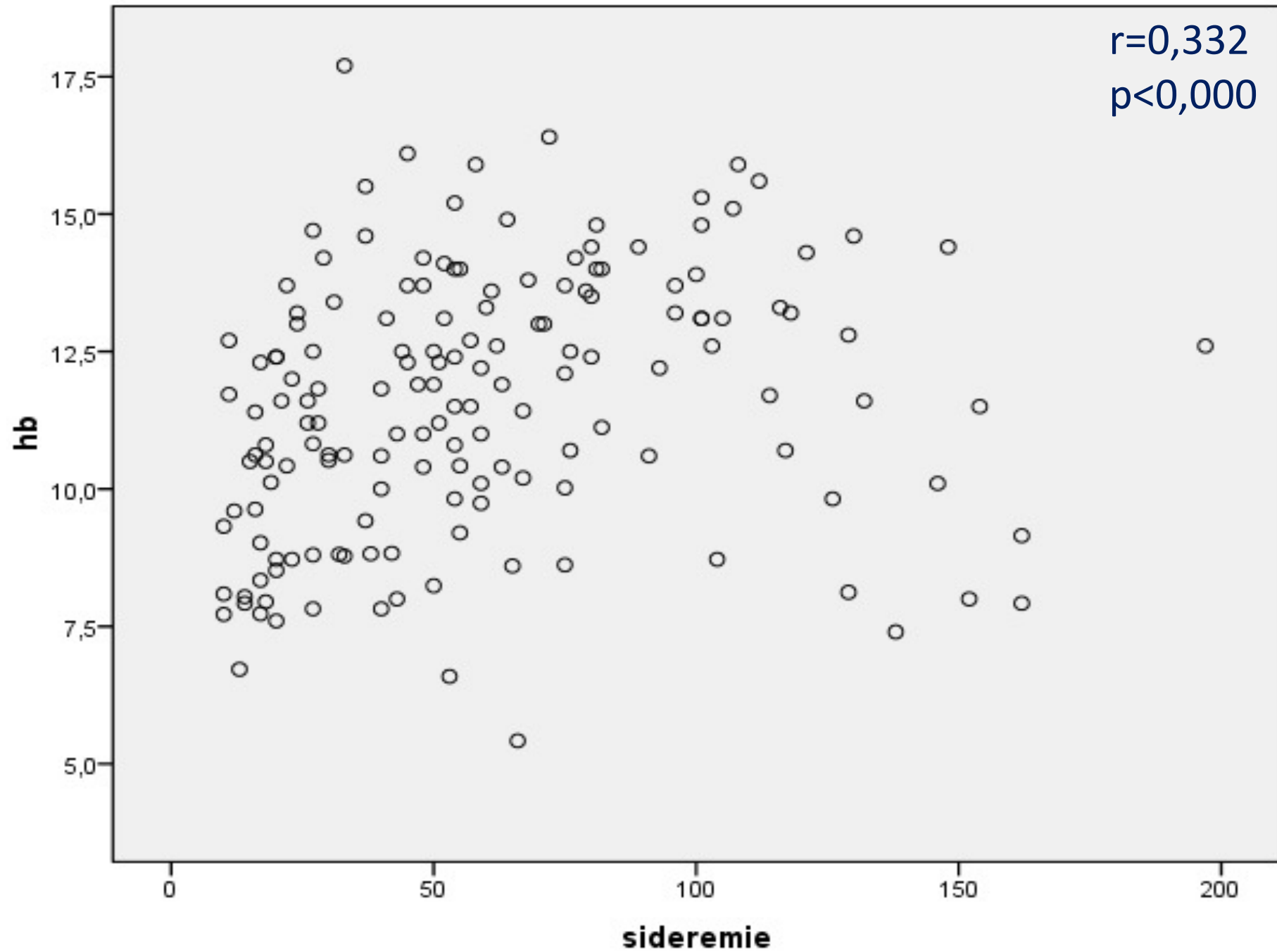
Variabile ordinale:

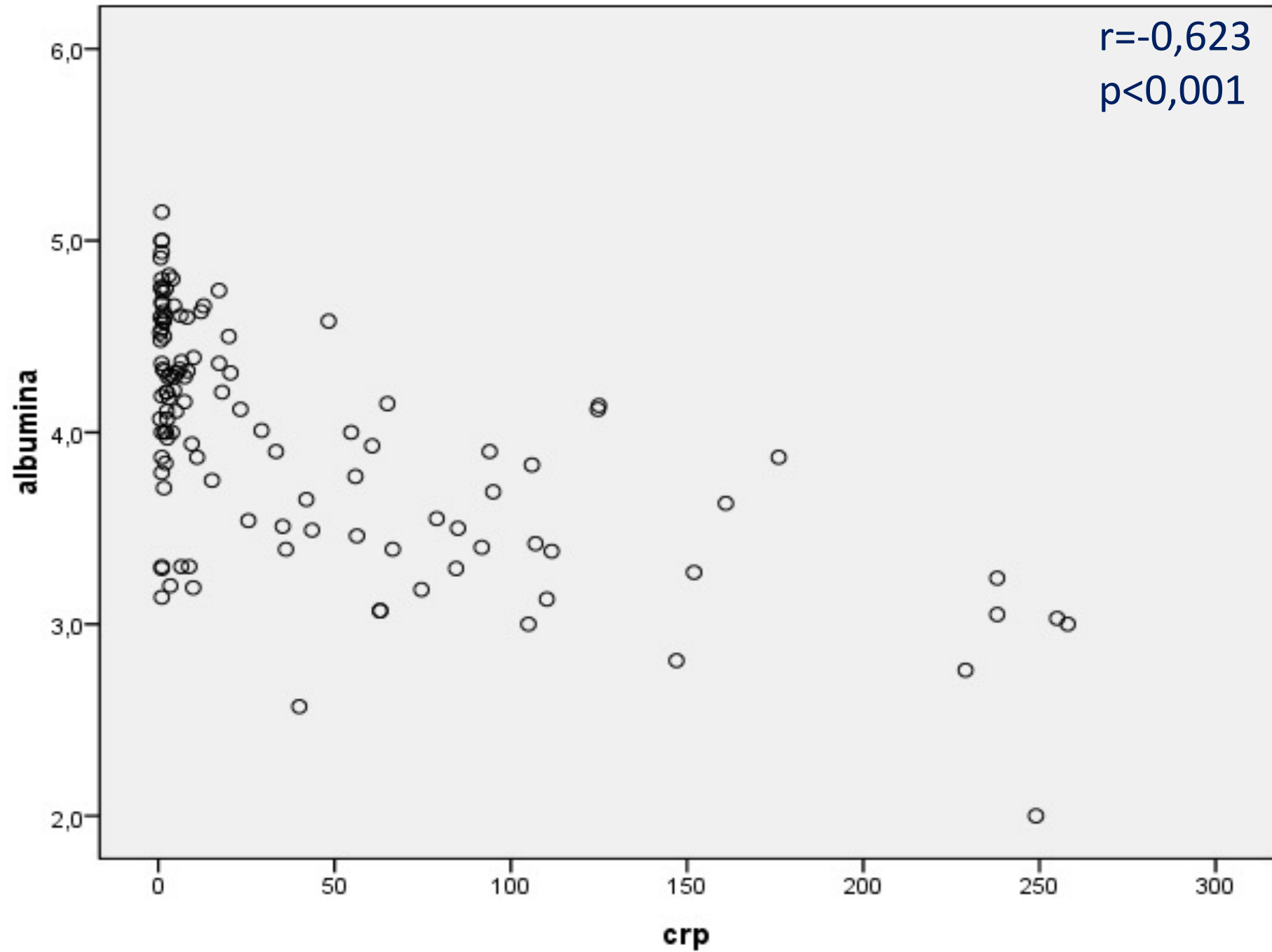
Test Mann-Whitney U

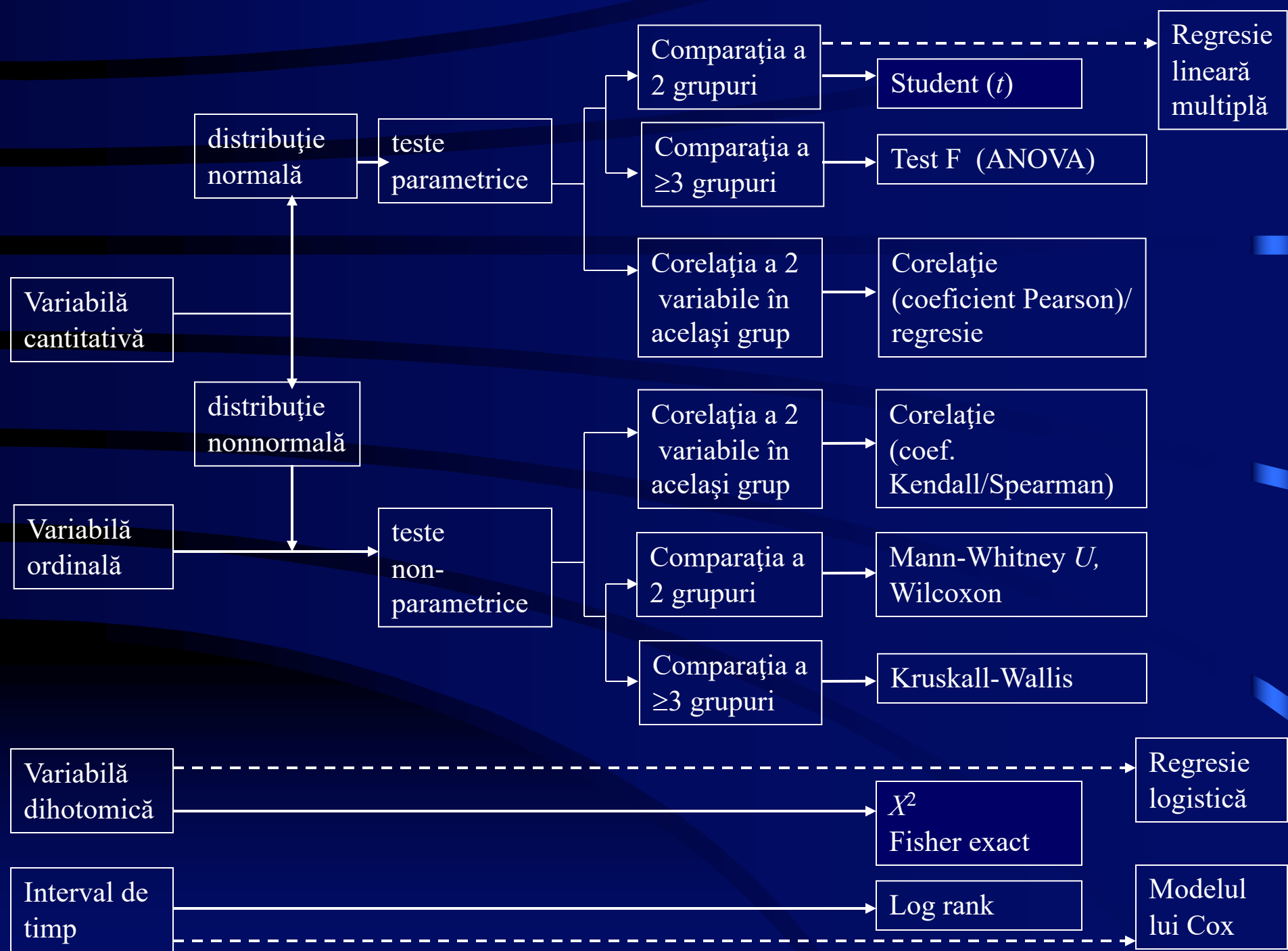
Variabile nominale:

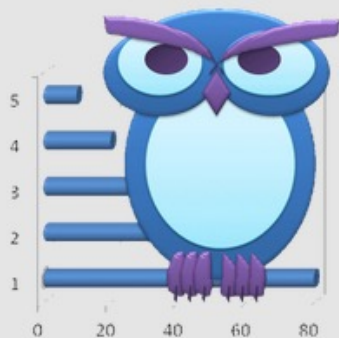
Testul χ^2

Testul exact al lui Fisher









App4Stats SPSS Statistics Free

Javier Párraga - Enrique Bernal - 8 februarie 2014

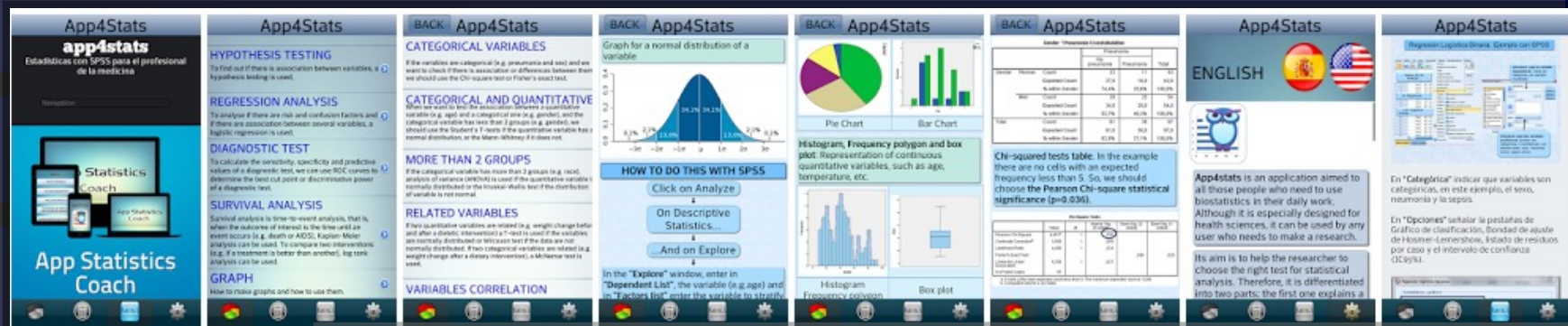
Domeniul medical

Aplicație instalată

Această aplicație este compatibilă cu toate dispozitivele dvs.

★★★★★ (89)

+60 Recomandă acest material pe Google



Medical Statistics: A Guide to Data Analysis and Critical Appraisal

Jennifer Peat, Belinda Barton

ISBN: 978-0-7279-1812-3

338 pages

July 2005, BMJ Books

[View Most Recent Edition of This Title](#)

Alegerea între un test parametric/nonparametric - cazurile simple

Alegem un test neparametric în 2 situații:

- efectul este o variabilă ordinală și populația este clar non-Gaussiană (de ex. notele studenților, scorul Apgar, scala vizuală analogică pt. Durere etc.)
- efectul este o variabilă cantitativă și suntem siguri că nu are o distribuție gaussiană în populație (transformare: log, reciproca, radical)
- efectul este o variabilă cantitativă cu distribuție gaussiană, dar dispersia este mult diferită