

# STATISTICA ANALITICA

C. Băicuș

[www.baicus.ro](http://www.baicus.ro)

p și intervalele de încredere

Dr. Cristian Băicuș

[www.baicus.ro](http://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)

- Placebo: supraviețuire: 10 ani
- Tratament: supraviețuire 10 ani + 1h
- $p = 0,0001$ .

- Tratatamentul îmbunătățește semnificativ supraviețuirea ( $p = 0,0001$ )

**1. Uitați-vă întotdeauna după mărimea efectului!**

- Tratatamentul îmbunătățește *semnificativ* supraviețuirea ( $p=0,0001$ )
  - CI dau mai multe informații decât  $p$ , așadar sunt de preferat
  - $p$ -urile amestecă mărimea efectului cu mărimea eșantionului
  - $p$ -urile nu au ce căuta în medicină
    - Schulz, Grimes. The Lancet Handbook of Clinical Research, 2006
  - 1986: Ken Rothman a interzis  $p$ -urile în *Epidemiology* cat a fost editor șef

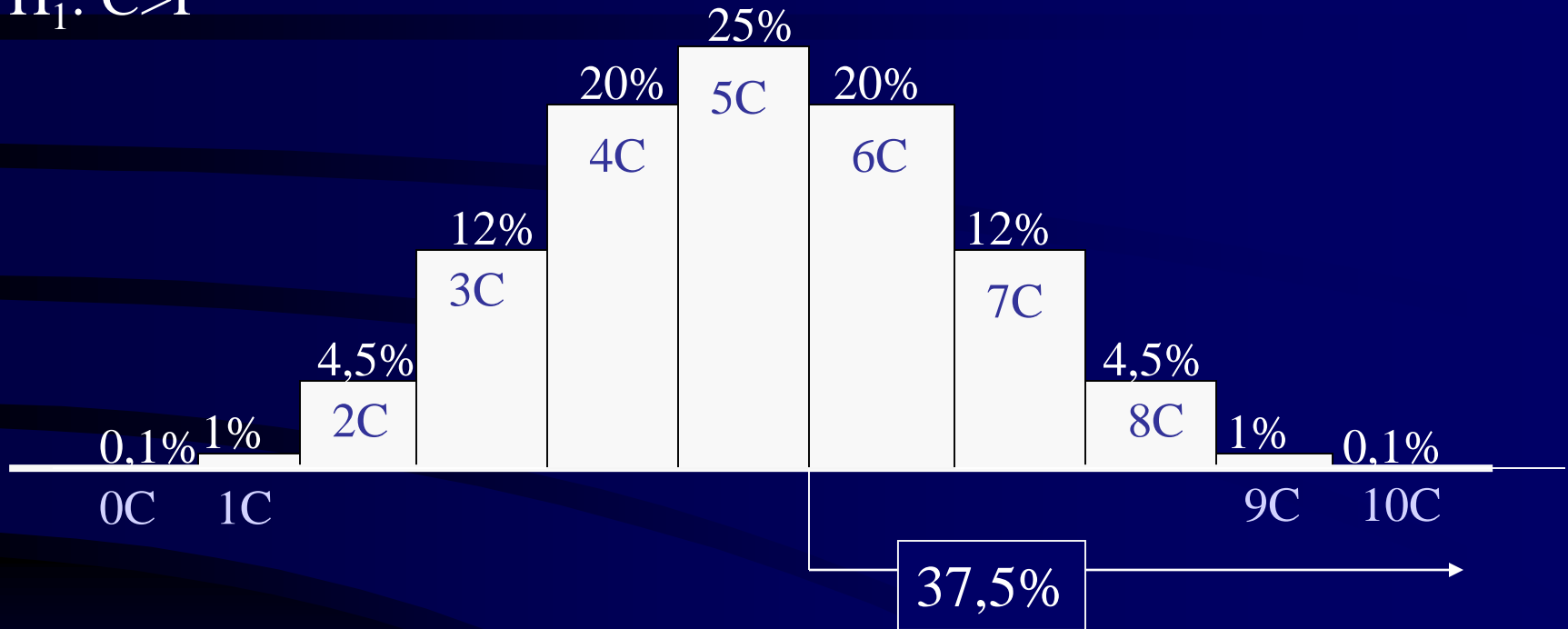


p

10x10.000

$H_0: C=P$

$H_1: C>P$



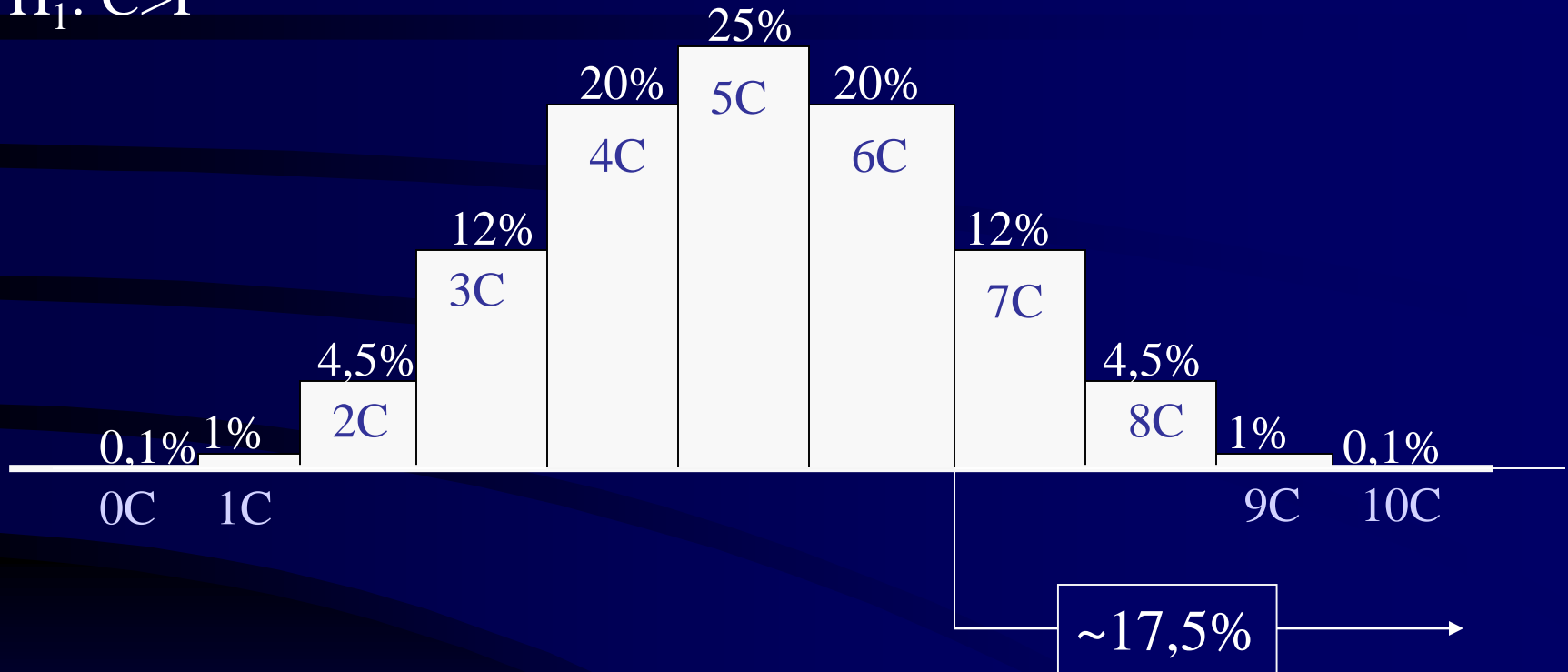
$p=0,375$

unidirectional, *one tailed (one sided)*

p

$H_0: C=P$

$H_1: C>P$



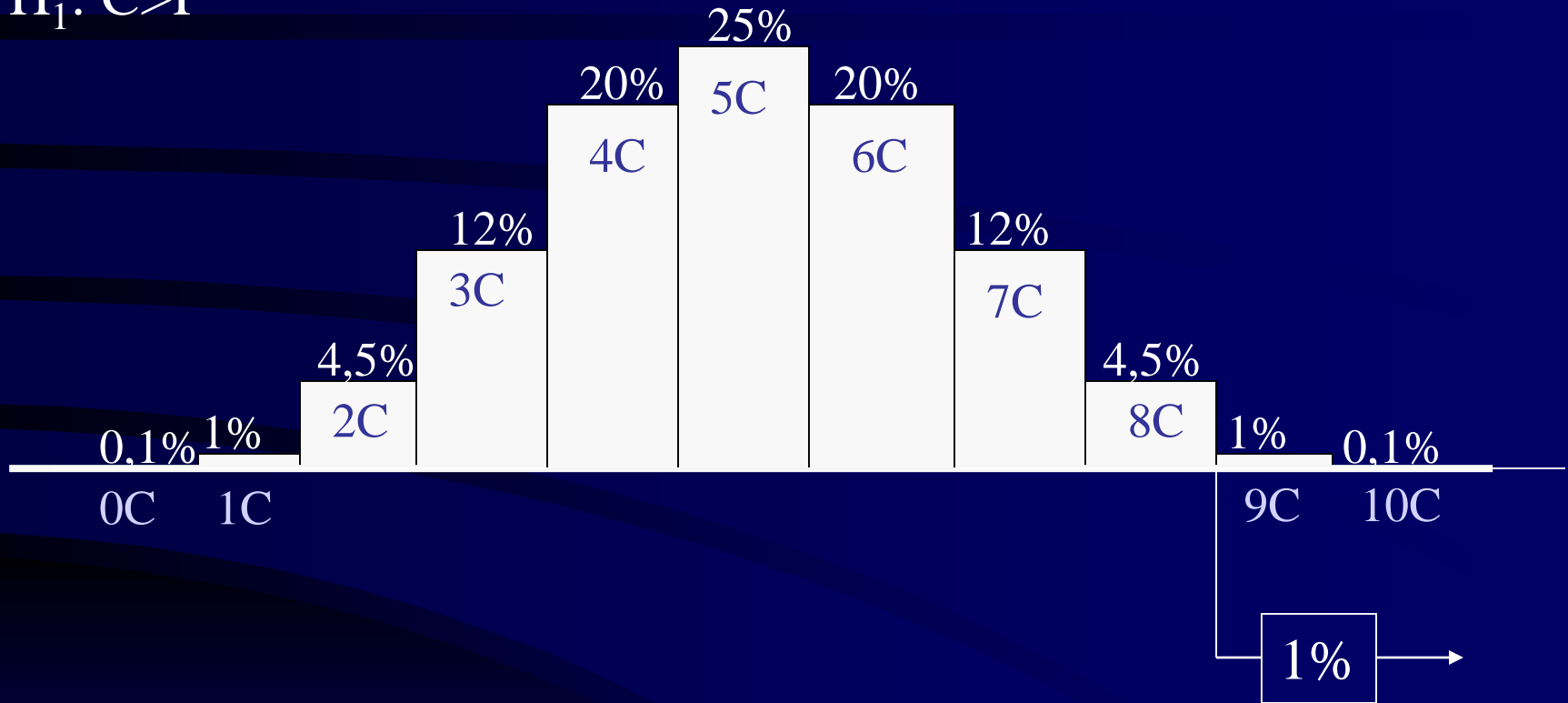
$p=0,175$

unidirectional, *one tailed (one sided)*

p

$H_0: C=P$

$H_1: C>P$



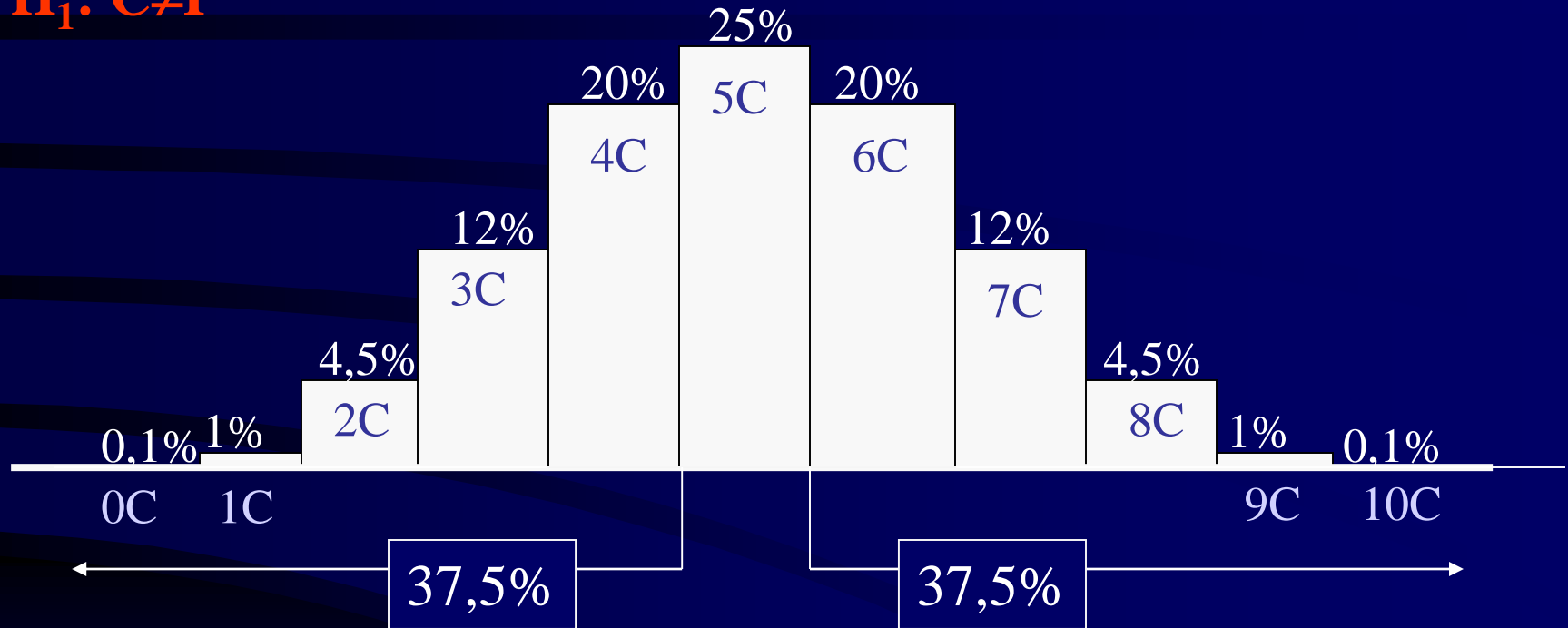
$p=0,01$

unidirectional, *one tailed (one sided)*

p

$H_0: C=P$

$H_1: C \neq P$



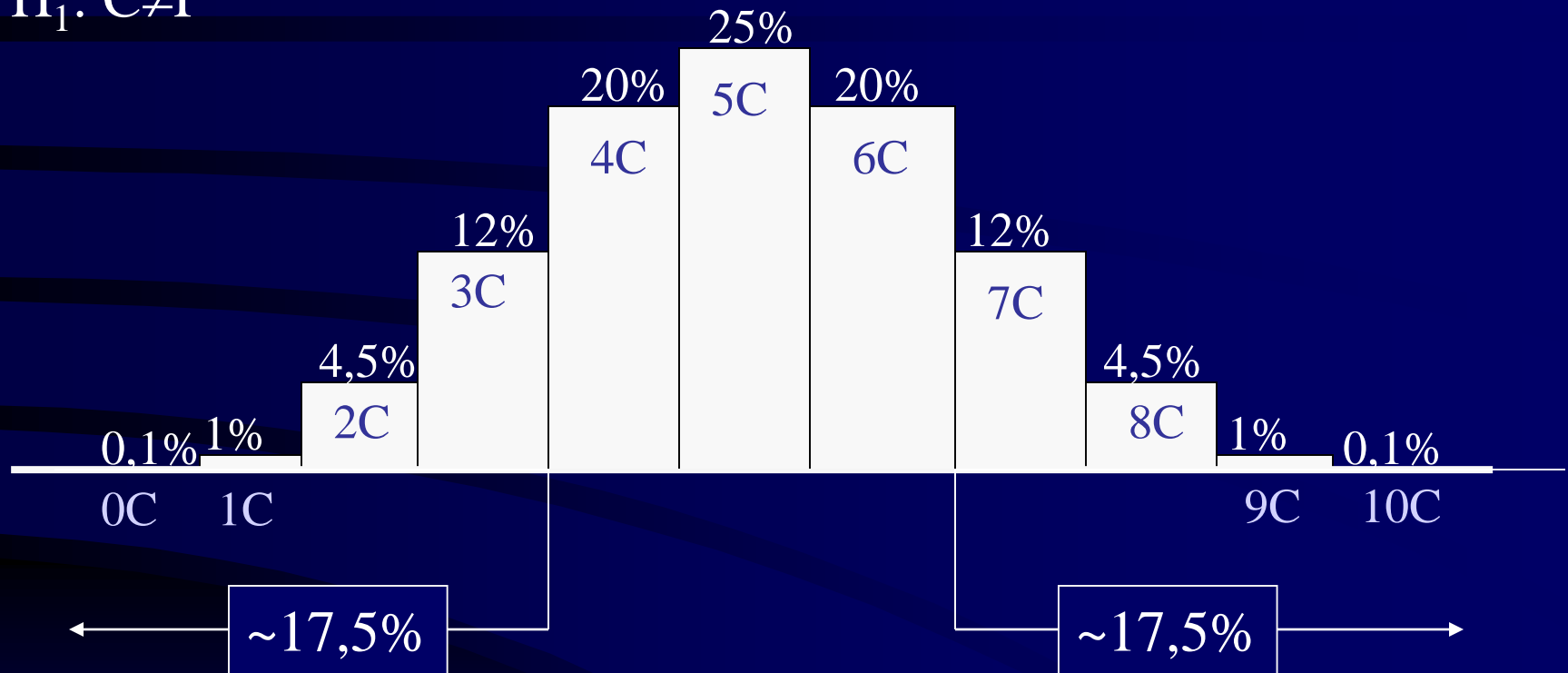
$p=0,75$

*bidirectional, two tailed (two sided)*

p

$H_0: C=P$

$H_1: C \neq P$



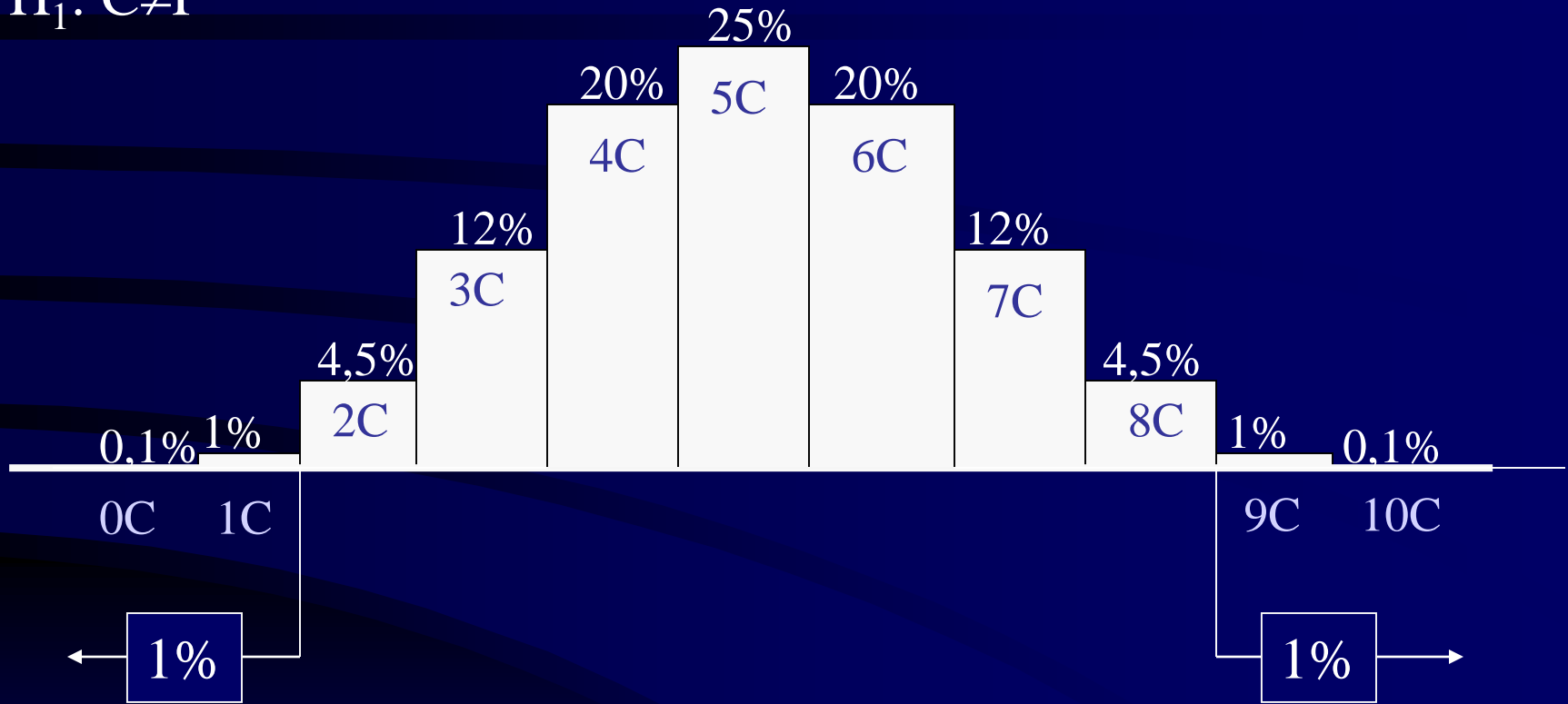
$p=0,35$

*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
<b>500 pts</b>	<b>500 pts</b>
<b>23</b>	<b>23</b>



# Studiu terapeutic

Placebo	Tratament
<b>500 pts</b>	<b>500 pts</b>
<b>23</b>	<b>23</b>
<b>25</b>	<b>21</b>

**p=0,546**

# Studiu terapeutic

Placebo	Tratament
<b>500 pts</b>	<b>500 pts</b>
<b>23</b>	<b>23</b>
<b>25</b>	<b>21</b>
<b>29</b>	<b>17</b>

**p=0,07**

# Studiu terapeutic

Placebo	Tratament
<b>500 pts</b>	<b>500 pts</b>
23	23
25	21
29	17
30	15

**p=0.022**

# CI

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

# CI

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

# CI

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

# CI

<i>Johannis</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**



**Nesemnificativ statistic:**

**atunci când CI include**

- **1 (pt. RR, OR)**
- **0 (pt. RRR sau RRA)**
- **$\infty$  (pt. NNT)**

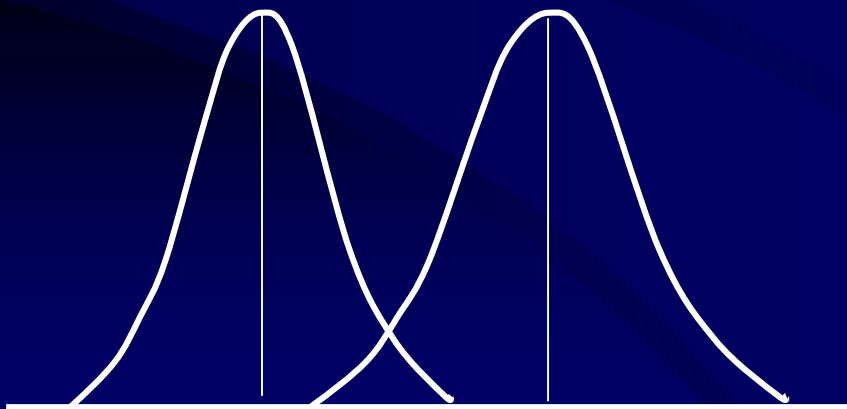
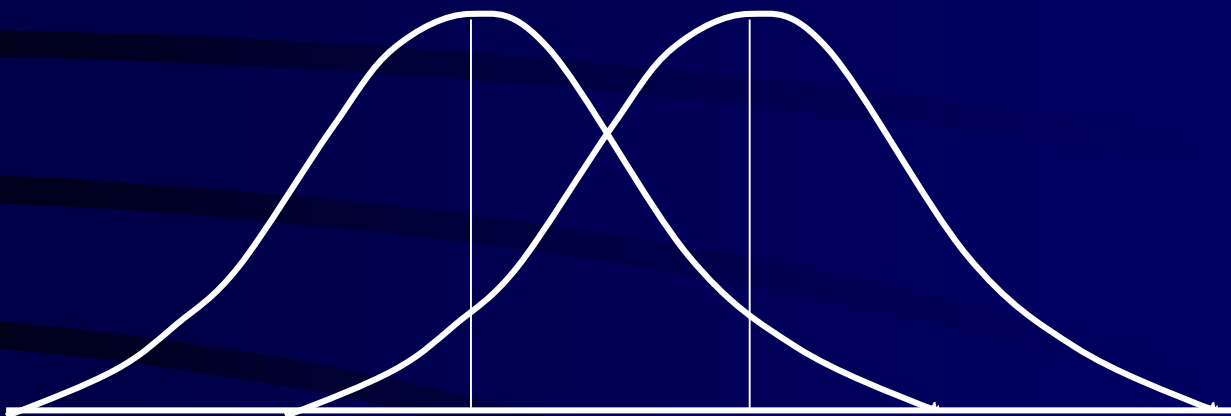
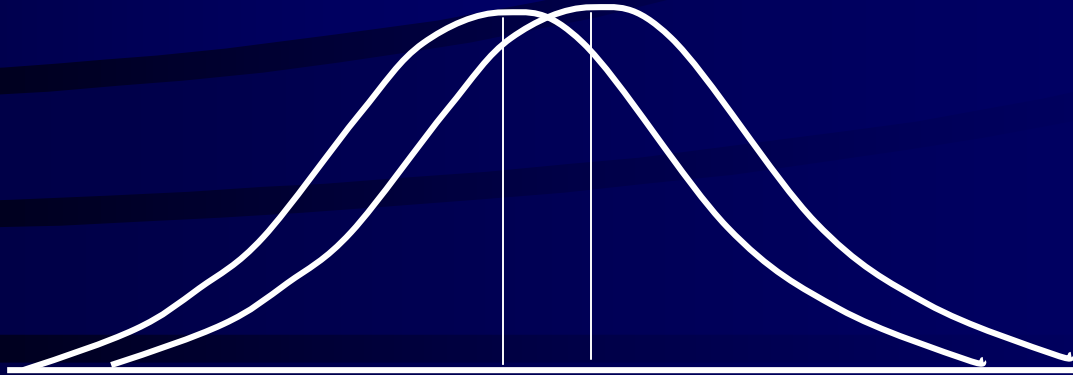
- nesemnificativ statistic =  $H_0$  nu a putut fi înlăturată, nu că aceasta este adevărată
- nu există diferență, sau studiul nu a avut puterea necesară pentru a evidenția această diferență  
⇒ atenție la CI

CI

semnificație statistică



semnificație clinică



- Clasa NYHA: I medic

4, 4, 2, 3, 1, 3, 4, 4, 3, 2

- Clasa NYHA: II medic

3, 2, 2, 4, 2, 4, 2, 3, 3, 2, 1, 4, 2, 2, 3

- Clase NYHA: I medic

1, 2, 2, 3, 3, 3, 4, 4, 4, 4

- Clase NYHA: II medic

1, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4

I: 1:	-14	-14
2:	$(+ 1 - 7) \times 2 =$	-12
3:	$(+ 8 - 3) \times 3 =$	+15
4:	$(+12 \quad ) \times 4 =$	+48
<hr style="border: 0.5px solid black;"/>		
		+37

	DECES		
	+	-	
<b>TRAT</b>	<b>10</b> <i>/15</i>	<b>90</b> <i>/85</i>	<b>100</b>
<b>PLACEBO</b>	<b>20</b> <i>/15</i>	<b>80</b> <i>/85</i>	<b>100</b>
	<b>30</b> (0,15)	<b>170</b> (0,85)	<b>200</b>

	DECES		
	+	-	
<b>TRAT</b>	<b>20</b> <i>/28,57</i>	<b>180</b> <i>/171,42</i>	<b>200</b>
<b>PLACEBO</b>	<b>30</b> <i>/21,42</i>	<b>120</b> <i>/128,57</i>	<b>150</b>
	<b>50</b> (0,14)	<b>300</b> (0,86)	<b>350</b>



# Sumarizarea datelor

Varsta, VSH, autoevaluarea scaderii ponderale si sexul la pacientii cu cancer si fara

Tipul variabilei	Exemplu	Cancer+	Cancer-	p
Cantitativa (distributie Gaussiana)	Varsta	Media (SD) 65 (11)   55 (9)		<0.001
Cantitativa (distributie neGaussiana)	VSH	<b>Mediana (min, max)</b> <b>Mediana (Q25%, Q75%)</b> 76 (12, 157)   45 (12, 122) 76 (23, 119)   45 (19, 87)		0.027
Ordinala	Eval scaderii ponderale scala Lickert	<b>Mediana (min, max)</b> <b>Mediana (Q25%, Q75%)</b> 3 (0, 4)   4 (0, 5)		0.09
Nominala dihotomica	Sexul masculin	Nr (%) 228 (65%)   115 (43%)		0.03

# Sumarizarea datelor

Varsta, VSH, autoevaluarea scaderii ponderale si sexul la pacientii cu cancer si fara

Tipul variabilei	Exemplu	Cancer+	Cancer-	p
Cantitativa (distributie Gaussiana)	Varsta	Media (SD)		<0.001
		65 (11)	55 (9)	

# Test t (Student)

226 (55%)

115 (45%)

Variabile numerice cu distributie neGaussiana

Variabile ordinale:

# Test Mann-Whitney $U$

Cantitativa (distributie neGaussiana)	VSH	<b>Mediana (min, max)</b> <b>Mediana (Q25%, Q75%)</b> 76 (12, 157)   45 (12, 122) 76 (23, 119)   45 (19, 87)	0.027
Ordinala	Eval scaderii ponderale scala Lickert	<b>Mediana (min, max)</b> <b>Mediana (Q25%, Q75%)</b> 3 (0, 4)   4 (0, 5)	0.09
Nominala dihotomica	Sexul masculin	<b>Nr (%)</b> 228 (65%)   115 (43%)	0.03

# Sumarizarea datelor

Varsta, VSH, autoevaluarea scaderii ponderale si sexul la pacientii cu cancer si fara

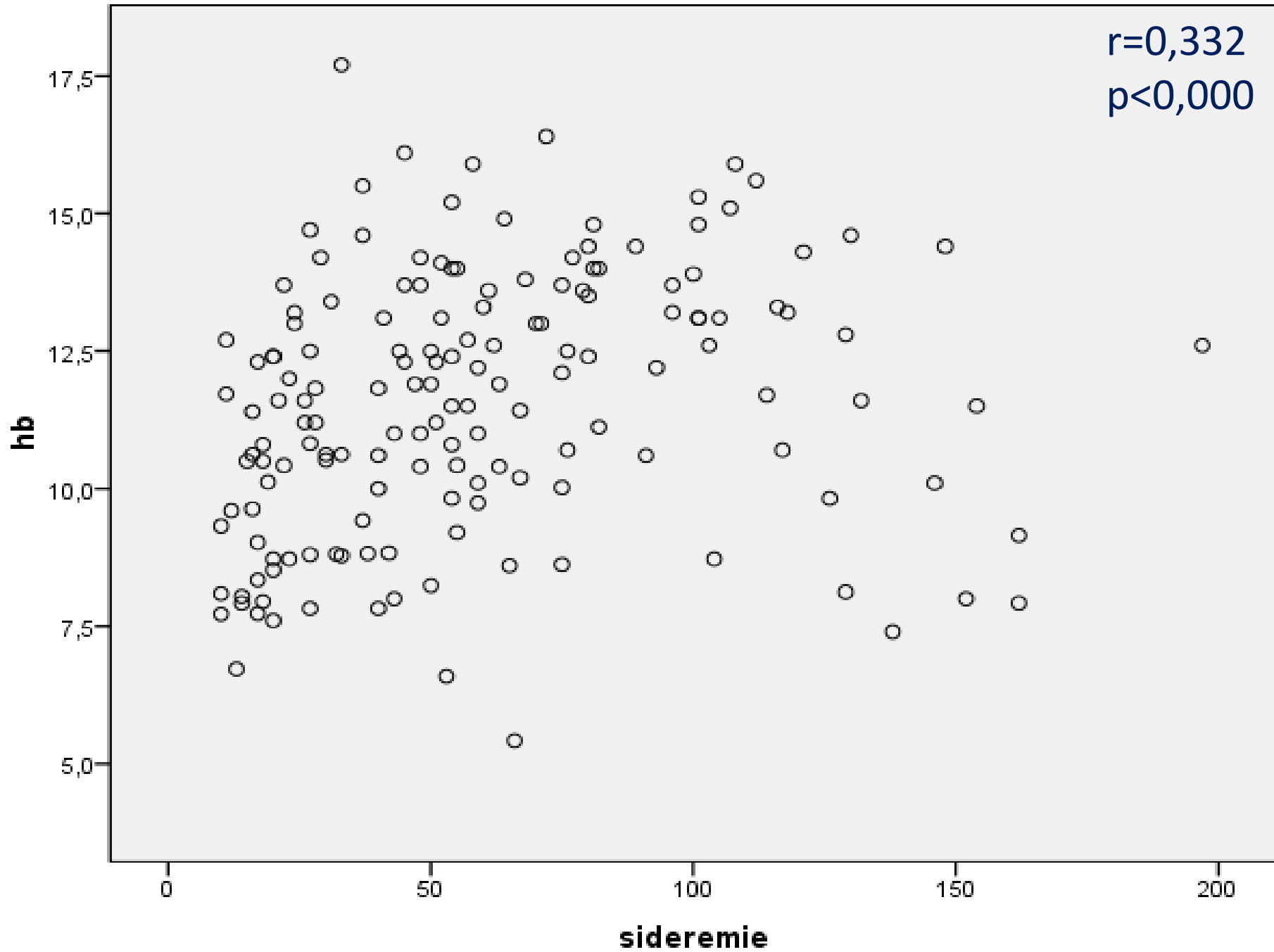
Tipul variabilei	Exemplu	Cancer+	Cancer-	p
Cantitativa (distributie Gaussiana)	Varsta	Media (SD)		<0.001

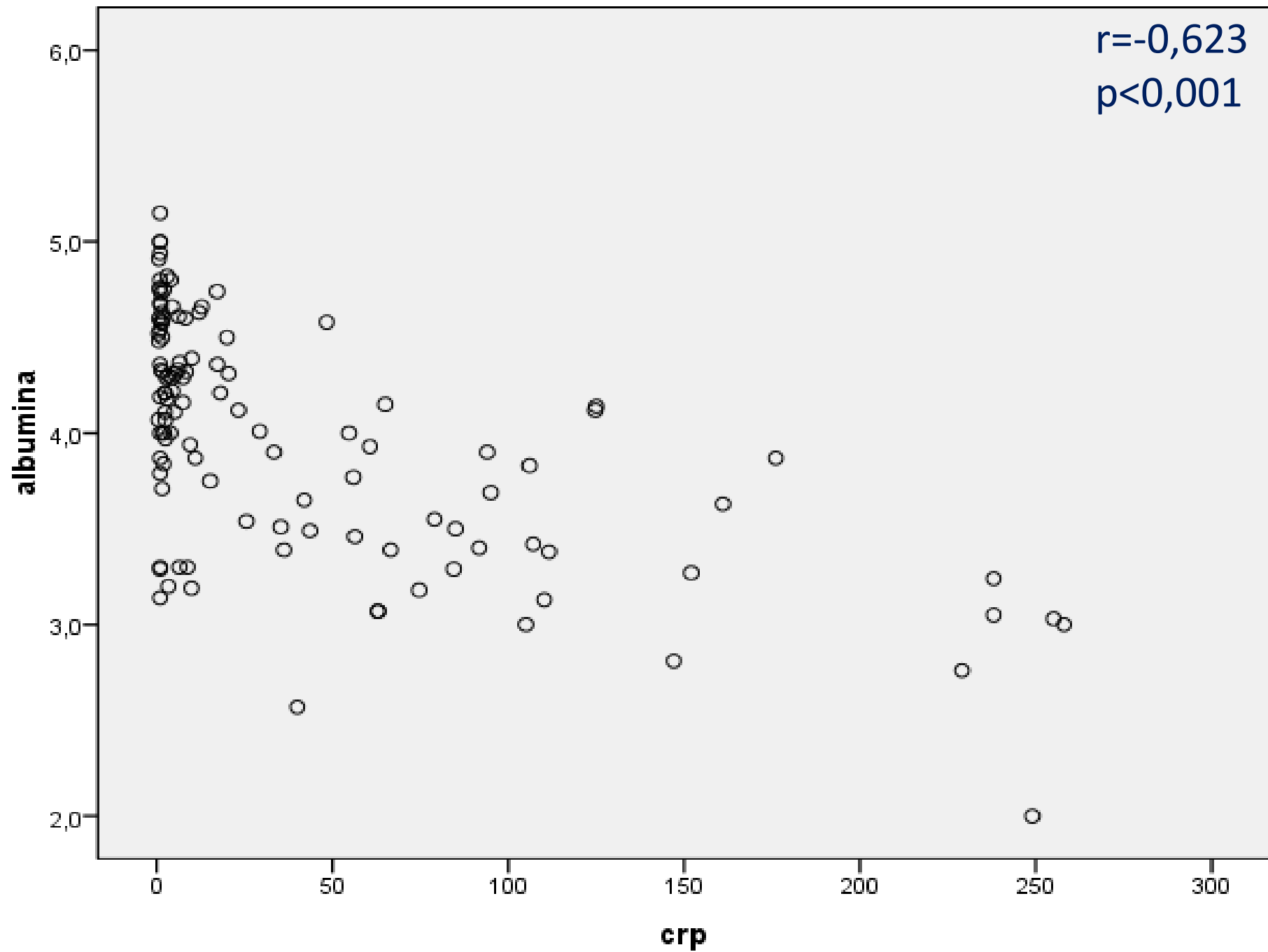
**Variabile nominale:**

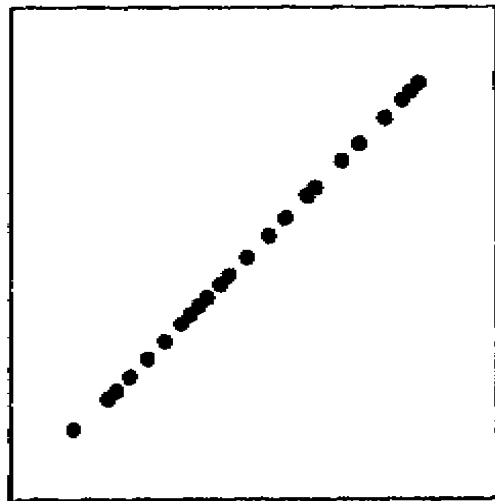
**Testul X2**

**Testul exact al lui Fisher**

Nominala dihotomica	Sexul masculin	Nr (%)		0.03
		228 (65%)	115 (43%)	

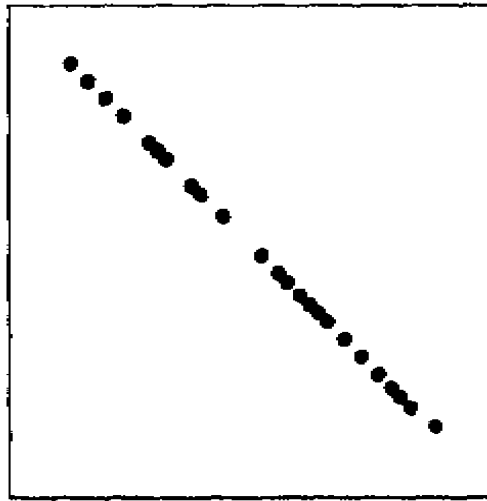






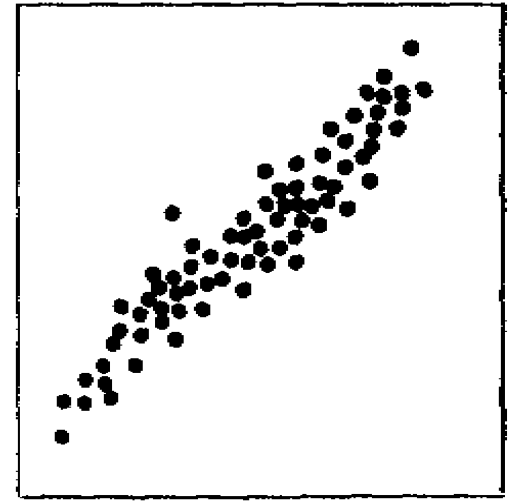
$r = 1.00$

A



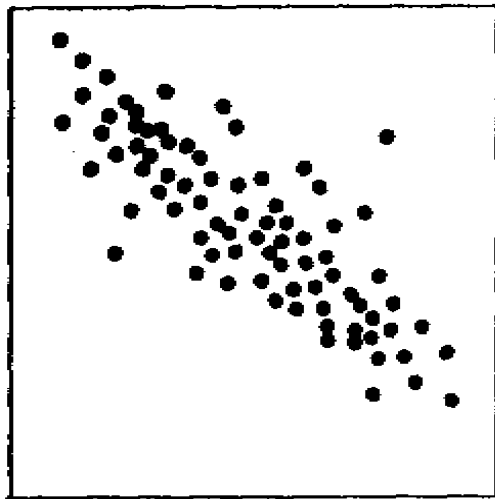
$r = -1.00$

B



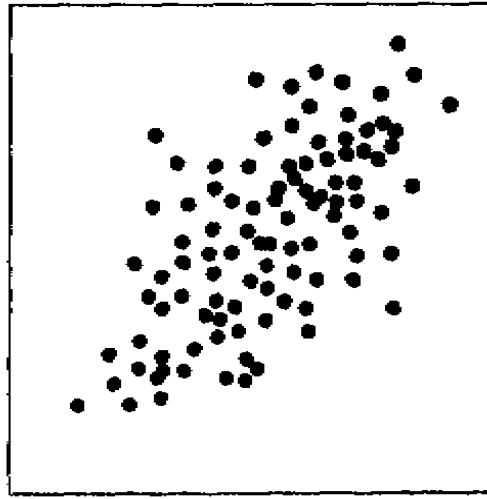
$r = 0.94$

C



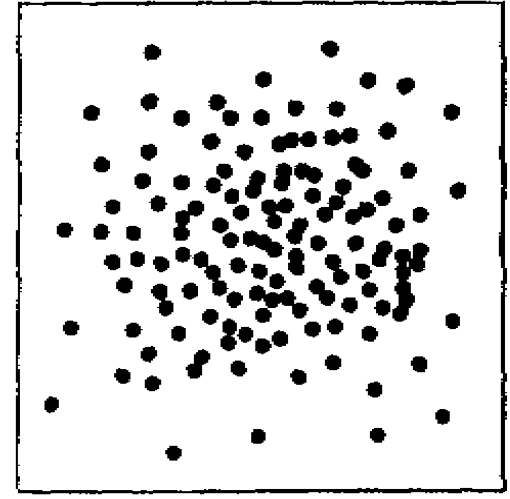
$r = -0.85$

D



$r = 0.44$

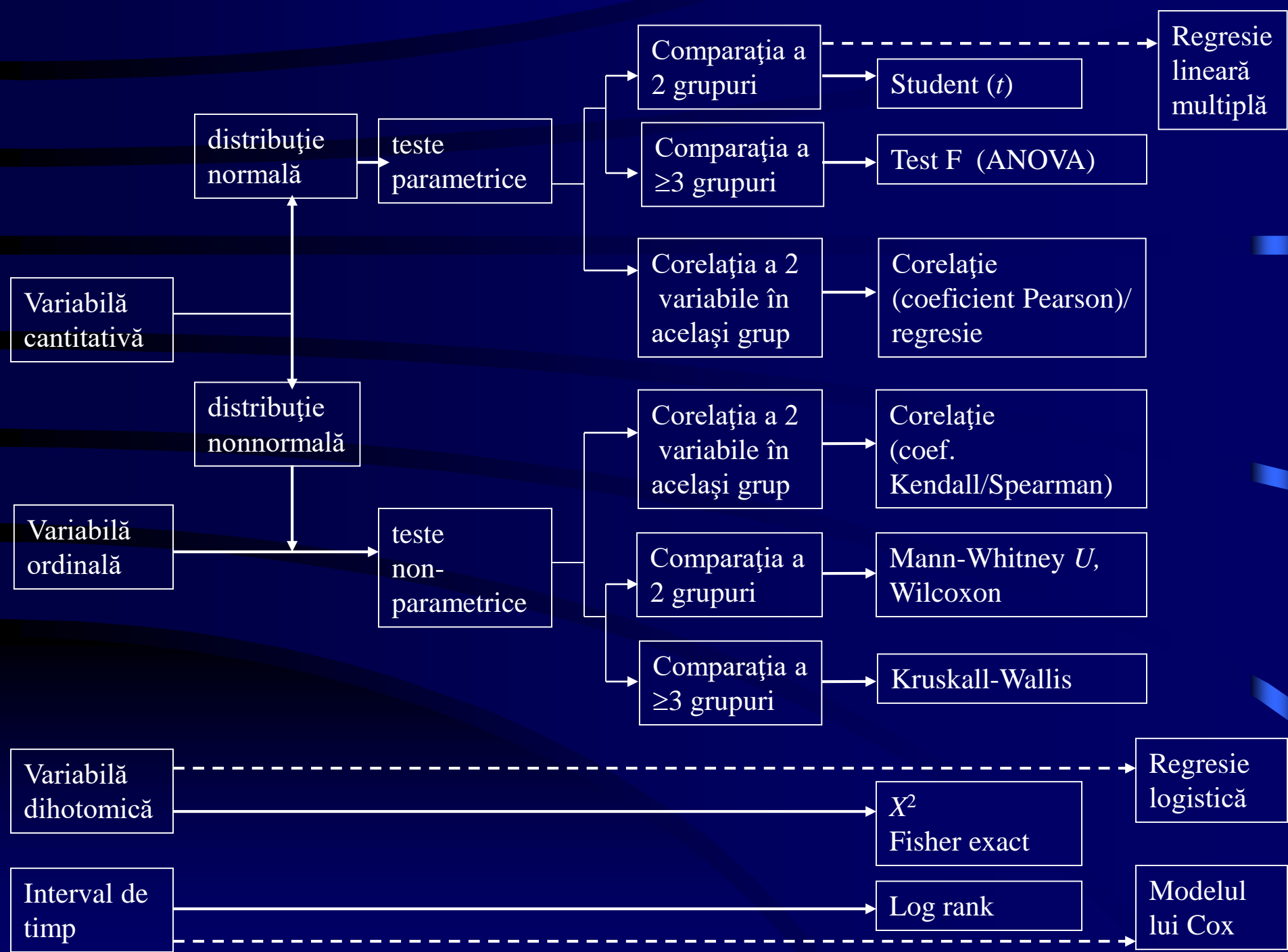
E



$r = 0.00$

F

*Figura 8: Grafice ilustrând diferite grade de corelație lineară*





# 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