Inferenţa statistică pe date calitative
Inferenţa statistică pe date cantitative
Studii de supravieţuire

Inferenţa statistică pe date calitative

Cuprins

- Tabela de contingenţă 2×2
- Riscuri şi raţii
- Testul χ^2 (testarea asocierii în tabela de contingenţă)
- Testul Fisher
- Testul z pentru proporţii
- Testul McNemar

Tabela de contingenţă 2×2

- Scale de tip nominal (dicotomiale: tabela de contingenţă de 2×2) sau ordinal (tabela de contingenţă de r×c)
- Frecvenţa absolută (numărul de evenimente per categorie)
- Tabela de contingenţă de 2×2: 4 categorii
  - AP = adevărat pozitiv
  - FP = fals pozitiv
  - FN = fals negativ
  - AN = adevărat negativ

Tabela de contingenţă 2×2

	Ulcer deschis	Ulcer vindecat	Total
Recurenţă +	AP = 1	FP = 5	= 1+5 = 6
Recurenţă -	FN = 7	AN = 16	= 7+16 = 23
Total	=1+7=8	=5+16=21	= 1+5+7+16 = 29

- Grade de libertate (df) = numărul minim de celule cu numere necesare pentru a calcula restul celulelor.
  - În tabelul de contingenţă de 2×2: dacă avem totalurile de pe rânduri şi coloane putem obţine valorile celorlalte celule.
  - df = (r - 1)(c - 1); r = numărul de rânduri, c = numărul de coloane

Riscuri şi raţii: Mărimi ale asocierii

Denumire	Formula	Definiţie
Rata falşilor pozitivi	=FP/(FP+AN)	Probabilitatea unui test fals + (α)
Rata falşilor negativi	=FN/(FN+AP)	Probabilitatea unui test fals – (β)
Sensibilitate	=AP/(AP+FN)	Probabilitatea unui test real + (1- β)
Specificitate	=AN/(AN+FP)	Probabilitatea unui test real – (1- α)
Acurateţe	=(AP+AN)/n	Probabilitatea generală a unei decizii corecte
Valoarea predictivă pozitivă	=AP/(AP+FP)	Probabilitatea ca un test pozitiv să fie corect
Valoarea predictivă negativă	=AN/(AN+FN)	Probabilitatea ca un test negativ să fie corect
Riscul relativ	=AP(FP+AN)/FN(AP+FP)	
Rata şansei	=(AP·AN)/(FN·FP)	
Riscul atribuabil	=AP/(AP+FP)-FN/(FN+AN)	

Riscuri şi raţii: Mărimi ale asocierii

	Ulcer deschis	Ulcer vindecat	Total
Recurenţă +	AP = 1	FP = 5	= 1+5 = 6
Recurenţă -	FN = 7	AN = 16	= 7+16 = 23
Total	=1+7=8	=5+16=21	= 1+5+7+16 = 29

Denumire	Formula
Rata falşilor pozitivi	= 5/(5+1) = 0,8334
Rata falşilor negativi	= 7/(7+16) = 0,3043
Sensibilitate	= 1/(1+7) = 0,1250
Specificitate	= 16/(16+5) = 0,7619
Acurateţe	= (1+16)/29 = 0,5862
Valoarea predictivă pozitivă	= 1/(1+5) = 0,1667
Valoarea predictivă negativă	= 16/(16+7) = 0,6957
Riscul relativ	= 1(5+16)/7(1+5) = 21/42 = 0,50
Rata şansei	= (1·16)/(7·5) = 0,4571
Riscul atribuabil	= 1/(1+5)-7/(7+16) = 0,1667-0,3043=-0,1376

Testarea asocierii în tabela de contingenţă

- Testul χ^2
  - Nu trebuie utilizat pentru eşantioane de volum mic.
  - Testul este valid doar dacă valoarea expectată (aşteptată) pentru fiecare celulă este cel puţin egală cu 1 şi frecvenţa absolută observată este de minim 5.
  - Dacă aceste condiţii nu sunt îndeplinite se aplică testul exact al lui Fisher (Fisher's Exact Test)

Testul χ^2

- Indică dacă cele două variabile sunt sau nu independente DAR NU cuantifică puterea asocierii dintre ele.

1. Definirea ipotezelor:
2. Definirea parametrului
3. Definirea pragului de semnificaţie
4. Definirea regiunii critice
5. Calcularea valorii observate a parametrului
6. Luarea deciziei

Testul χ^2: Problema

- S-a investigat într-un studiu asocierea dintre obezitatea (ca factor de risc) şi bolile cardio-vasculare la persoanele în etate (> 60 ani). Din totalul de 620 persoane investigate s-au identificat 150 persoane cu obezitate şi boală cardio-vasculară, 230 persoane fără obezitate şi fără boală cardio-vasculară şi 60 persoane fără obezitate dar cu boală cardio-vasculară. Există o asociere între obezitate şi boala cardio-vasculară? (df=1; α=0,05; χ^2_{critic} = 3,84).

Testul χ^2: 1. Definirea ipotezelor

- H_0:
  - Nu există asociere între obezitate şi bolile cardio-vasculare.
  - Obezitatea şi bolile cardio-vasculare sunt independente.
- H_1:
  - Există asociere între obezitate şi bolile cardio-vasculare.
  - Obezitatea şi bolile cardio-vasculare sunt asociate.

Testul χ^2: 2. Definirea parametrului

$$\chi^2 = \sum_{i=1}^{r\cdot c} \frac{(f_i^0 - f_i^t)^2}{f_i^t}$$

urmează o lege cu (r-1)(c-1) grade de libertate

unde:

- χ^2 = parametrul testului χ^2
- f_i^o = frecvenţa observată
- f_i^t = frecvenţa teoretică

Testul χ^2: 3. Definirea pragului de semnificație

- Fie α = 0,05 pragul de semnificație al testului.

Testul χ^2: 4. Definirea regiunii critice

- Regiunea critică este $[\chi_\alpha^2, \infty)$.
- Pentru α = 0,05, χ_α^2 = 3,84.

Testul χ^2: 5. Calcularea valorii observate a parametrului

OBSERVAT	BCV+	BCV-	Total
Obezitate +	AP = **150**	FP = **180**	330
Obezitate -	FN = **60**	AN = **230**	290
Total	210	410	620

TEORETIC	BCV+	BCV-	Total
Obezitate +	**= 330×210/620**	**= 330×410/620**	330
Obezitate -	**= 290×210/620**	**= 290×410/620**	290
Total	210	410	620

Testul χ^2: 5. Calcularea valorii observate a parametrului

OBSERVAT	BCV+	BCV-
Obezitate +	**150**	**180**
Obezitate -	**60**	**230**

TEORETIC	BCV+	BCV-
Obezitate +	**= 112**	**= 218**
Obezitate -	**= 98**	**= 192**

$$\chi^2 = \frac{(150-112)^2}{112} + \frac{(180-218)^2}{218} + \frac{(60-98)^2}{98} + \frac{(230-192)^2}{192}$$

$$\chi^2 = \frac{38^2}{112} + \frac{(-38)^2}{218} + \frac{(-38)^2}{98} + \frac{(38)^2}{192}$$

$$\chi^2 = \frac{1444}{112} + \frac{1444}{218} + \frac{1444}{98} + \frac{1444}{192} = 12,89 + 6,63 + 14,73 + 7,52 = \boxed{41,77}$$

Testul χ^2: 6. Luarea deciziei

- Dacă $\chi^2 \in [3,84, \infty)$ se respinge H_0 cu un risc de eroare de tip I (α).
- Dacă $\chi^2 \notin [3,84, \infty)$ se acceptă H_0 cu un risc de eroare de tip II (β).

- Deoarece $41,77 \in [3,84, \infty)$ se respinge H_0 cu un risc de eroare de 5%.
- **Există asociere între obezitate şi bolile cardio-vasculare.**

Testul χ^2: Corecția Yates

$$\chi^2 = \sum_{i=1}^{r\cdot c} \frac{|f_i^0 - f_i^t|^2 - 0,5}{f_i^t}$$

- 0,5 = corecția Yates (ajustarea mărimilor zecimale)

Testul Fisher

- Corecție a testului χ^2;
- Valoarea p asociată parametrului ne dă probabilitatea ca valoarea observată de independență să fie atribuită doar şansei.
- O valoare p mică indică că există alte cauze decât şansa influențează rezultatul şi astfel cele două variabile investigate nu sunt independente.

Testul z pentru proporții

1. Compararea unei frecvențe observate cu o frecvență teoretică.
2. Testarea egalității a două frecvențe.

Testul z: 1. Compararea unei frecvențe observate cu o frecvență teoretică

- **Scop:** Investigarea semnificației diferenței între o frecvență teoretică *p* (într-o populație) şi o frecvență observată *f* pe un eşantion reprezentativ (variabilă calitativă (binare)).
- **Condiții de aplicare:** Testul este corect aplicat dacă numărul *n* al observațiilor eşantionului este suficient de mare ($n \cdot p$, $n \cdot (1-p) > 10$.
- **Parametrul**:
  - n = volumul eşantionului

$$z = \frac{f - p}{\sqrt{\frac{p(1-p)}{n}}}$$

Testul z: 1. Compararea unei frecvențe observate cu o frecvență teoretică

- Suntem interesați de investigarea prevalenței hepatitei B la personalul care lucrează în laboratoarele clinicilor de boli infecțioase din Transilvania. Se ştie din studii anterioare că prevalența hepatitei B în populația generală din Transilvania este de 9%. S-a luat în studiu un eşantion de 100 persoane şi s-a obținut o prevalență a hepatitei B de 6%. Există diferență semnificativă între frecvența hepatitei B la personalul care lucrează în laboratoarele spitalelor de boli infecțioase din Transilvania față de populația generală?

Testul z: 1. Compararea unei frecvențe observate cu o frecvență teoretică

- f = 0,06, p = 0,09, n = 100
- Ipoteza nulă: Nu există diferență semnificativă între frecvența hepatitei B la eşantionul studiat față de frecvența hepatitei B în populația generală.
- Ipoteza alternativă, test bilateral: Există diferență semnificativă între frecvența hepatitei B la nivelul eşantionului şi prevalența hepatitei B în populația generală.

Testul z: 1. Compararea unei frecvențe observate cu o frecvență teoretică

- f = 0,06; p = 0,09; n = 100
- Pragul de semnificație: $\alpha = 0{,}05$.
- Regiunea critică test bilateral: $(-\infty; -1{,}96\,] \cup [1{,}96; \infty)$

$$z = \frac{f-p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0{,}06-0{,}09}{\sqrt{\frac{0{,}09(1-0{,}09)}{100}}} = \frac{-0{,}03}{\sqrt{\frac{0{,}09 \cdot 0{,}91}{100}}}$$

$$z = \frac{-0{,}03}{\sqrt{\frac{0{,}0819}{100}}} = \frac{-0{,}03}{\sqrt{0{,}000819}} = \frac{-0{,}03}{0{,}029} = -1.05$$

Testul z: 1. Compararea unei frecvențe observate cu o frecvență teoretică

- Concluzia testului:
  - Deoarece parametrul statistic calculat al testului nu aparține regiunii critice, se acceptă ipoteza nulă. Nu există diferență semnificativă între frecvența hepatitei B la eșantionul studiat față de frecvența hepatitei B în populația generală.

Testul z:
2. Testarea egalității a două frecvențe

- **Scop**: Investigarea semnificației diferenței între frecvențele relative și respectiv ale unei valori a unei variabile calitative pe două eșantioane randomizate independente extrase din două populații diferite.
- **Condiții de aplicare:** Testul este aproximativ și se presupune că numărul observațiilor eșantioanelor este suficient de mare (n_1, n_2 > 30) pentru a justifica aproximarea distribuției binomiale prin una normală.

$$z = \frac{(p_1 - p_2)}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad p = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2}$$

Testul z:
2. Testarea egalității a două frecvențe

- S-a studiat statutul HIV pe un eșantion de 170 femei cu vârste cuprinse între 18 și 40 de ani din Moldova, și respectiv un eșantion de 89 femei cu vârste cuprinse între 18 și 40 de ani din Transilvania. Pentru eșantionul din Moldova, Frecvența testelor HIV+ a fost de 10% în eșantionul din Moldova și 2,7% în eșantionul din Transilvania.
- Frecvența infecției cu HIV la femeile cu vârste cuprinse între 18 și 40 de ani din Moldova este diferită față de frecvența infecției la femeile de aceeași vârstă din Transilvania?

Testul z:
2. Testarea egalității a două frecvențe

Datele problemei:
- $p_1 = 0,10$; $p_2 = 0,027$; $n_1 = 170$; $n_2 = 89$.

Ipoteza nulă:
- Nu există o diferență semnificativă între frecvența infecției HIV la femeile din Moldova față de frecvența infecției HIV la femeile din Transilvania.

Ipoteza alternativă, test bilateral:
- Există o diferență semnificativă între frecvența infecției HIV la femeile din Moldova față de frecvența infecției HIV la femeile din Transilvania.

Testul z:
2. Testarea egalității a două frecvențe

Pragul de semnificație: $\alpha = 0,05$.

Regiunea critică:
- Testul bilateral: (-∞; -1,96] ∪ [1,96; ∞)
- Testul unilateral: [1,645, ∞)

$$z = \frac{(p_1 - p_1)}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0,10 - 0,027}{\sqrt{0,075(1-0,075)\left(\frac{1}{170} + \frac{1}{89}\right)}}$$

$$z = \frac{0,073}{\sqrt{0,075 \cdot 0,925 \cdot (0,006 + 0,011)}} = \frac{0,073}{\sqrt{0,001}} = \frac{0,073}{0,034} = 2,118$$

Testul z:
2. Testarea egalității a două frecvențe

Concluzie:

- Test bilateral:
  - Deoarece parametrul statistic calculat al testului aparține regiunii critice se respinge ipoteza nulă și se acceptă ipoteza alternativă.
  - Există diferență semnificativă între frecvența infecției HIV la femeile din Moldova față de frecvența infecției HIV la femeile din Transilvania.

Testul McNemar

- Evaluarea dependenţei variabilelor calitative perechi (dorim să determinăm dacă o anumită caracteristică este sau nu asociată cu o anumită patologie):
  - Identificăm *n* pacienţi care prezintă patologia de interes (e.g. Cancer bronho-pulmonar) şi *n* pacienţi cu aceleaşi caracteristici ca şi primul grup dar care nu o prezintă.

	Martor = da	Martor = nu
Caz = da	**a**	**b**
Caz = nu	**c**	**d**

Testul McNemar

	Martor = da	Martor = nu
Caz = da	**a**	**b**
Caz = nu	**c**	**d**

$$\chi^2_{1df} = \frac{(|b-c|-1)^2}{b+c}$$

Testul McNemar: fumat vs cancer

	Martor = da	Martor = nu
Cancer = da	a = 2	b = 5
Cancer = nu	c = 0	d = 3

ID	Caz	Martor
1	Da	Nu
2	Nu	Nu
3	Da	Da
4	Da	Nu
5	Da	Nu
6	Nu	Nu
7	Nu	Nu
8	Da	Da
9	Da	Nu
10	Da	Nu

$$\chi^2_{1df} = \frac{(|5-0|-1)^2}{5+0} = \frac{4^2}{5} = \frac{16}{5} = 3{,}2$$

$\chi^2_{critic(\alpha=0,05)} = 3{,}84$

$3{,}2 < 3{,}84 \rightarrow$ acceptăm ipoteza nulă

Fumatul nu este în relaţie cu apariţia cancerului bronho-pulmonar.

De reţinut!

- Aplicarea unui test statistic trebuie făcută în conformitate cu condiţiile acestuia.
- Pe variabile calitative se aplică teste non-parametrice (nu necesită asumpţia distribuţiei normale a datelor).
- Variabile nominale:
  - Un singur eşantion sau eşantioane perechi:
    - Tabelul de contingenţă cu parametrii de tip raţii şi rapoarte
  - Eşantioane perechi: testul Mc Nemar

De reţinut!

- Variabile nominale:
  - Două eşantioane: realizarea tabelului de contingenţă 2×2 şi aplicarea testului Fisher sau χ^2
- În analiza proporţiilor există teste diferite pentru:
  - Compararea unei frecvenţe cu o frecvenţă cunoscută
  - Compararea a două frecvenţe
- Atenţie la calcularea riscurilor şi raţiilor pe tabela de contingenţă!

Inferenţa statistică pe date cantitative

Cuprins

- Variabile cantitative continue:
  - Testul z şi t (o medie sau medii perechi)
  - Testul z şi t (testarea a două medii)
  - ANOVA (≥ 3 medii)
- Ranguri (variabile cantitative discrete sau cantitative care nu îndeplinesc condiţia de normalitate):
  - Testul sumei rangurilor: Wilcoxon
  - Kruskal-Wallis (≥ 3 eşantioane independente)
  - Friedman (≥ 3 eşantioane dependente)

Teste de normalitate: variabile cantitative continue

- Shapiro-Wilk
- Kolmogorov-Smirnov
- Shapiro-Wilk
- Chi-Square Goodness-of-Fit

vezi **cursul 10**

- Dacă datele urmează o distribuţie normală: aplicăm un test parametric
- Dacă datele nu urmează o distribuţie normală: aplicăm un test de comparare al rangurilor

Testul Z de comparare a mediei unui eşantion cu media unei populaţii

- **Scopul testului*:*** compararea mediei unei variabile cantitative continue pe un eşantion reprezentativ extras dintr-o populaţie cu o medie cunoscută. Se presupune că cele două populaţii au aceiaşi variaţie σ^2 care se cunoaşte.

Condiţii de aplicare*:*

1. Este necesar să cunoaştem variaţia populaţiei (dacă nu o cunoaştem, aplicăm testul Student pentru compararea mediei unui eşantion cu media unei populaţii).
2. Testul este corect aplicat dacă populaţia este normal distribuită. Dacă populaţia nu este normal distribuită iar talia eşantionului este mică (< 30) testul dă o valoare orientativă.
3. Talia eşantionului este mare (≥ 30).

Testul Z de comparare a mediei unui eşantion cu media unei populaţii

Ipoteze:
- Ipoteza nulă: nu există diferenţă semnificativă între media eşantionului şi media populaţiei.
- Ipoteza alternativă pentru testul bilateral: există diferenţă semnificativă între media eşantionului şi media populaţiei.

Pragul de semnificaţie: $\alpha = 0{,}05$.

Regiunea critică pentru testul bilateral este
- $(-\infty , -1{,}96\,] \cup [1{,}96 , \infty)$

- Parametrul testului:

$$Z = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

- n = volumul eşantionului
- $\overline{X}$ = media eşantionului
- σ = deviaţia standard a populaţiei.

Testul Z de comparare a mediei unui eşantion cu media unei populaţii

- Studierea agregării familiale a bolilor cardiovasculare (adică prevalenţa bolii printre membrii unei familii este mai mare decât în rândul populaţiei generale) se poate realiza prin studiul legăturii dintre nivelul lipidic sanguin şi aceste boli. Se ştie că nivelul mediu al colesterolului sanguin la copii este de 175 mg/dL. La un eşantion de 10 copii, proveniţi din familii în care tatăl a decedat în urma unei boli cardiovasculare, media colesterolului sanguin este de 200 mg/dL iar deviaţia standard este de 50 mg/dL.
  - Nivelul colesterolului la această populaţie de copii este sau nu mai mare decât cel al populaţiei generale?
  - Este nivelul colesterolului obţinut la acest eşantion semnificativ diferit faţă de cel al populaţiei generale?

Testul Z de comparare a mediei unui eşantion cu media unei populaţii

- **Ipoteza nulă**: nu există diferenţă semnificativă între media colesterolului pentru eşantion faţă de media populaţiei.
- **Ipoteza alternativă** pentru testul bilateral: există diferenţă semnificativă între media colesterolului la eşantion şi respectiv la populaţia generală.
- **Prag de semnificaţie**: $\alpha = 0{,}05$
- **Regiunea critică pentru testul bilateral**:
  - $(-\infty; -1{,}96\,] \cup [1{,}96; \infty)$

$$Z = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{200 - 175}{\frac{50}{\sqrt{10}}} = \frac{25}{\frac{50}{3{,}162}} = \frac{25}{15{,}811} = 1{,}58$$

Testul Z de comparare a mediei unui eşantion cu media unei populaţii

- **Regiunea critică pentru testul bilateral**:
  - (-∞; -1,96] ∪ [1,96; ∞)

$$Z = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{200 - 175}{\frac{50}{\sqrt{10}}} = \frac{25}{\frac{50}{3,162}} = \frac{25}{15,811} = 1,58$$

- **Concluzie pentru testul bilateral**:
  - Deoarece parametrul statistic calculat al testului nu aparţine regiunii critice respingem ipoteza nulă.
  - Există o diferenţă semnificativă între media colesterolului la eşantionul ales şi populaţia generală.

Testul t de comparare a unei medii cu o medie cunoscută (variaţii necunoscute)

- **Scopul** testului este investigarea semnificaţiei diferenţei dintre media unui eşantion şi o medie standard cunoscută.
- **Ipoteza nulă**: nu există diferenţă semnificativă între media eşantionului şi media standard.
- **Ipoteza alternativă** pentru testul bilateral: există diferenţă semnificativă între media eşantionului şi media standard.
- **Condiţii de aplicare**
  - Testul se poate aplica atunci când variaţia σ^2 nu este cunoscută iar estimarea s_2 a acesteia se realizează pentru un eşantion mic (n < 30) care respectă o distribuţie normală. Dacă această condiţie de normalitate nu este satisfăcută atunci testul îşi pierde validitatea.
  - Dacă se cunoaşte variaţia populaţiei σ^2, şi $n \geq 30$ se aplică testul Z care este un test mult mai puternic.

Testul t de comparare a unei medii cu o medie cunoscută (variaţii necunoscute)

- **Numărul de grade de libertate (df)**: df = n-1
- **Pragul de semnificaţie**: α = 0,05.
- **Regiunea critică** pentru testul bilateral este:

$$(-\infty; -t_{n-1,\frac{\alpha}{2}}] \cup [t_{n-1,\frac{\alpha}{2}}; +\infty)$$

$$(-\infty; -t_{n-1;0,025}] \cup [t_{n-1;0,025}; +\infty)$$

- **Parametrul testului:**

$$t = \frac{\overline{X} - \mu_0}{\frac{s}{\sqrt{n}}}$$

- n = volumul eşantionului
- μ_0 = media standard
- $\overline{X}$ = media eşantionului
- s = deviaţia standard a eşantionului.

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{X})^2}{n-1}}$$

Testul t de comparare a unei medii cu o medie cunoscută (variaţii necunoscute)

- **Problema**: Nivelul mediu al colesterolului sangvin la femeile cu vârstă între 21 şi 40 de ani din România are o distribuţie normală şi o valoare medie de 190 mg/dL cu o deviaţie standard de 40mg/dL. S-au efectuat teste de sânge pe un eşantion de 10 femei din mediul rural cu vârste cuprinse între 21 şi 40 de ani şi s-a obţinut o medie a colesterolului de 181,52 mg/dL cu o deviaţie standard de 40 mg/dL.
  - Este nivelul colesterolului femeilor cu vârstă între 21 şi 40 de ani din rural semnificativ diferit de nivelul colesterolului populaţiei României?
  - Presupunem că nivelul colesterolului la femeile cu vârste cuprinse între 21 şi 40 de ani, din mediul rural este normal distribuit.

Testul t de comparare a unei medii cu o medie cunoscută: Soluţia

- **Datele problemei**:
  - μ_0 = 190; n = 10,
  - $\overline{X}$ = 181,52; s = 40
- **Ipoteza nulă**: media colesterolului la femeile din mediul rural nu diferă faţă de media colesterolului populaţiei femeilor din României.
- **Ipoteza alternativă pentru testul bilateral**: media colesterolului la femeile din mediul rural diferă faţă de media colesterolului populaţiei feminine a României.
- **Pragul de semnificaţie**:
  - α = 0,05.
- **Numărul de grade de libertate**: df = n-1 = 10-1 = 9
- **Regiunea critică**:

$$(-\infty; -t_{9;0,025}] \cup [t_{9;0,025}; +\infty)$$

$$(-\infty; -2,262] \cup [2,262; +\infty)$$

Testul t de comparare a unei medii cu o medie cunoscută: Soluţia

$$t = \frac{\overline{X} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{181,52 - 190}{\frac{40}{\sqrt{10}}} = \frac{-8,48}{\frac{40}{3,16}} = \frac{-8,48}{12,66} = -0,67$$

Concluzia:

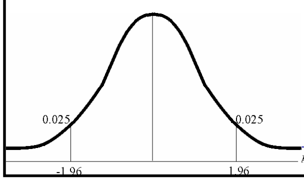
- Deoarece valoarea parametrului statistic calculat al testului nu aparţine regiunii critice ipoteza nulă se acceptă. Aceasta înseamnă că nivelul mediu al colesterolului la femeile din mediul rural nu diferă semnificativ faţă de media colesterolului în populaţia de sex feminin a României.

Testul Z de comparare a mediilor a două populaţii (variaţii cunoscute şi inegale)

- **Scopul testului:** compararea mediile pentru o variabilă cantitativă continuă în două populaţii, cunoscând variaţia în fiecare dintre aceste populaţii.
- **Condiţii de utilizare*:***
  - Populaţiile trebuie să aibă variaţii cunoscute. Dacă variaţiile nu sunt cunoscute, se aplică un test de tip Student pentru compararea mediilor a două populaţii.
  - Testul este corect numai dacă populaţiile sunt normal distribuite. Dacă populaţiile nu sunt normal distribuite, testul dă doar o valoare orientativă.
- **Ipoteza nulă**: diferenţa mediilor celor două populaţii este egală cu zero.
- **Ipoteza alternativă** pentru testul bilateral: diferenţa mediilor celor două populaţii este diferită de zero.

Testul Z de comparare a mediilor a două eşantioane (variaţii inegale)

- **Pragul de semnificaţie** considerat este $\alpha = 0,05$.
- **Regiunea critică** pentru testul bilateral: $(-\infty; -1,96\,] \cup [1,96; \infty)$



- **Parametrul testului**:

$$z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- $\overline{X}_1$ = media primului eşantionului;
- n_1 = volumul primului eşantion;
- s_1^2 = variaţia primului eşantion;
- $\overline{X}_2$ = media celui de-al doilea eşantion;
- n_2 = volumul celui de-al doilea eşantion;
- s_2^2 = variaţia celui de-al doilea eşantion.

Testul Z de comparare a mediilor a două eşantioane: Exemplu

- Se ştie că nivelul seric al magneziului urmează legea normală cu o variaţie de cu o variaţie de 1 mg/100 ml la persoanele din România şi respectiv cu o variaţie de 2,3 mg/100 ml la persoanele din Moldova. Nivelul mediu al magneziului seric, obţinut pe un eşantion de 12 persoane cu vârste cuprinse între 25 şi 35 de ani din România este de 2 mg/100 ml. S-au efectuat teste serologice la un eşantion de 8 persoane cu vârste cuprinse între 25 şi 35 de ani, din Moldova şi media magneziului seric a fost de 2,5 mg/100 ml. Există diferenţă între nivelul seric al magneziului la persoanele din Moldova faţă de persoanele din România.

Testul Z de comparare a mediilor a două eşantioane: Exemplu

- **Datele problemei**:
  - $n_1 = 12$; $n_2 = 8$
  - $m_1 = 2$; $m_2 = 2,5$
  - $s_1^2 = 1$; $s_2^2 = 2,3$
- **Ipoteza nulă**: Diferenţa mediilor magneziului seric la cele două eşantioane nu este semnificativ diferită de zero.
- **Ipoteza alternativă** pentru testul bilateral: Diferenţa mediilor magneziului seric la cele două eşantioane este semnificativ diferită de zero.
- **Pragul de semnificaţie**: $\alpha = 0,05$.
- **Regiunea critică** pentru testul bilateral:
  - $(-\infty; -1,96\,] \cup [1,96; \infty)$

Testul Z de comparare a mediilor a două eşantioane: Exemplu

$$z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{2 - 2,5}{\sqrt{\frac{1}{12} + \frac{2,3}{8}}} = \frac{-0,5}{\sqrt{0,083 + 0,288}}$$

$$z = \frac{-0,5}{\sqrt{0,371}} = \frac{-0,5}{0,609} = -0,821$$

Concluzie:

- Pentru testul bilateral: Deoarece parametrul statistic calculat al testului nu aparţine regiunii critice se acceptă ipoteza nulă, adică diferenţa mediilor magneziului seric pentru cele două eşantioane nu diferă semnificativ de zero.

Testul t de comparare a două medii (variaţii necunoscute şi egale)

- **Ipoteza nulă**: Diferenţa mediilor celor două populaţii este egală cu zero.
- **Ipoteza alternativă** pentru testul bilateral: Diferenţa mediilor celor două populaţii este diferită de zero.
- **Condiţii de aplicare**
  - Variabila de analizat în cele două populaţii este normal distribuită şi variaţiile celor două populaţii sunt egale.
  - Dacă aceste condiţii nu sunt satisfăcute atunci testul îşi pierde validitatea.
  - Dacă se cunoaşte variaţia populaţiei σ^2, se aplică testul Z care este un test mult mai puternic.

Testul t de comparare a două medii (variaţii necunoscute şi egale)

- **Numărul de grade de libertate (df):**
  - $df = n_1 + n_2 - 2$
- **Pragul de semnificaţie**: α = 0,05.
- **Regiunea critică** pentru testul bilateral

$$(-\infty;-t_{n_1+n_2-2;\frac{\alpha}{2}}]\cup[t_{n_1+n_2-2;\frac{\alpha}{2}};+\infty)$$

- **Parametrul statistic al testului**

$$t=\frac{\bar{X}_1-\bar{X}_2}{\sqrt{s\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$s=\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$$

Testul t de comparare a două medii: Exemplu

- Dorim să studiem dacă există o diferenţă semnificativă între cantitatea de acid uric sangvin la femeile din mediul urban faţă de cele din mediul rural. Pe un eşantion de 16 femei cu vârste cuprinse între 30 şi 50 de ani din mediul urban, media acidului uric este de 5 mg/100 ml, cu o variaţia egală cu 2 mg/100 ml. S-a determinat media acidului uric la un eşantion de 16 persoane de sex feminin cu vârste cuprinse între 30 şi 50 de ani din mediul rural, având o valoare de 4 mg/100 ml cu o variaţia de 2 mg/100 ml.

Testul t de comparare a două medii: Exemplu

- **Datele problemei**:
  - $n_1 = 16$; $n_2 = 16$
  - $m_1 = 5$; $m_2 = 4$
  - $s^2 = 2$.
- **Ipoteza nulă**: Nu există diferenţă semnificativă între mediile acidului uric la cele două eşantioane.
- **Ipoteza alternativă** pentru testul bilateral: Există o diferenţă semnificativă între mediile acidului uric la cele două eşantioane.
- **Numărul de grade de libertate**:

$df = n_1+n_2-2 = 16+16-2=30$

- **Pragul de semnificaţie**: α = 0,05.
- **Regiunea critică** pentru testul bilateral:

$$(-\infty;-t_{n_1+n_2-2;0,025}]\cup[t_{n_1+n_2-2;0,025};+\infty)$$

$$(-\infty;-2{,}04]\cup[2{,}04;+\infty)$$

Testul t de comparare a două medii: Exemplu

$$s=\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}=\sqrt{\frac{(16-1)2+(16-1)2}{16+16-2}}=\sqrt{\frac{60}{30}}=1{,}41$$

$$t=\frac{\bar{X}_1-\bar{X}_2}{\sqrt{s\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}=\frac{5-4}{\sqrt{1{,}41\left(\frac{1}{16}+\frac{1}{16}\right)}}=\frac{1}{\sqrt{1{,}41\cdot 0{,}25}}=\frac{1}{\sqrt{0{,}3525}}=\frac{1}{0{,}5937}=1{,}68$$

Concluzie:

- Deoarece parametrul testului nu aparţine regiunii critice, se acceptă ipoteza nulă. În concluzie nu există o diferenţă între mediile acidului uric la femeile cu vârste cuprinse între 30 şi 50 de ani din mediul urban şi respectiv mediul rural.

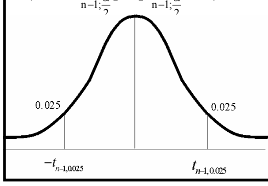
Testul t de comparare a mediilor a două eşantioane perechi

- **Scopul testului:** compararea pentru o variabilă cantitativă continuă media ei aritmetică pentru două eşantioane perechi (observaţii ale aceleiaşi variabile cantitative realizate pe elementele unui eşantion înainte şi după acţiunea unui factor).
- **Condiţii de aplicare:** fiecărei observaţii din primul eşantion îi corespunde o observaţie pereche din al doilea eşantion iar diferenţele dintre valorile perechi sunt normal distribuite.
- **Ipoteza nulă**: Media diferenţei valorilor perechi din eşantioanele perechi nu este semnificativ diferită de zero.
- **Ipoteza alternativă** pentru testul bilateral: Media diferenţei valorilor perechi din eşantioanele perechi este semnificativ diferită de zero.

Testul t de comparare a mediilor a două eşantioane perechi

- **Numărul de grade de libertate (df):** df = n – 1.
- **Pragul de semnificaţie este**: α = 0,05.
- **Regiunea critică:**

$$(-\infty;-t_{n-1;\frac{\alpha}{2}}]\cup[t_{n-1;\frac{\alpha}{2}};+\infty)$$



- **Parametrul statistic al testului**

$$t=\frac{\bar{d}}{\frac{s}{\sqrt{n}}}$$

$$\bar{d}=\frac{(d_1+d_2+\ldots+d_n)}{n}$$

- s = deviaţia standard a diferenţelor
- n = volumul eşantionului

Testul t de comparare a mediilor a două eşantioane perechi: Problema

ID	Tensiunea arterială sistolică la femei înainte de utilizarea contraceptivelor orale (x_{i1})	Tensiunea arterială sistolică la femei după utilizarea contraceptivelor orale (x_{i2})	$d_i(x_{i1} - x_{i2})$
1	115	128	13
2	112	115	3
3	107	106	-1
4	119	128	9
5	115	122	7
6	138	145	7
7	126	132	6
8	105	109	4
9	104	102	-2
10	115	117	2

Testul t de comparare a mediilor a două eşantioane perechi: Soluţie

- **Ipoteza nulă**: nu există diferenţă semnificativă între tensiunea arterială sistolică înainte şi respectiv după utilizarea contraceptivelor orale.
- **Ipoteza alternativă** pentru testul bilateral: există diferenţă semnificativă între tensiunea arterială sistolică înainte şi respectiv după utilizarea contraceptivelor orale.
- **Numărul de grade de libertate**: df = n – 1 = 10-1 = 9
- **Pragul de semnificaţie**: α = 0,05.
- **Regiunea critică** pentru testul bilateral: $(-\infty;-2{,}262]\cup[2{,}262;+\infty)$

Testul t de comparare a mediilor a două eşantioane perechi: Soluţie

$$\bar{d} = \frac{13+3-1+9+7+7+6+4-2+2}{10} = \frac{48}{10} = 4{,}8$$

$$s = \sqrt{\frac{(13-4{,}8)^2+(3-4{,}8)^2+(-1-4{,}8)^2+(9-4{,}8)^2+2\cdot(7-4{,}8)^2+(6-4{,}8)^2+(4-4{,}8)^2+(-2-4{,}8)^2+(2-4{,}8)^2}{10-1}}$$

$$s = \sqrt{\frac{8{,}2^2+(-1{,}8)^2+(-5{,}8)^2+(4{,}2)^2+2\cdot 2{,}2^2+1{,}2^2+(-0.8)^2+(-6{,}8)^2+(-2{,}8)^2}{10-1}}$$

$$s = \sqrt{\frac{67{,}24+3{,}24+33{,}64+(4{,}2)^2+2\cdot 4{,}84+1{,}44+0{,}64+46{,}24+7{,}84}{10-1}} = \sqrt{\frac{187{,}60}{9}} = \sqrt{20{,}84} = 4{,}57$$

$$t = \frac{\bar{d}}{\frac{s}{\sqrt{n}}} = \frac{4{,}8}{\frac{4{,}57}{\sqrt{9}}} = \frac{4{,}8}{\frac{4{,}57}{3}} = \frac{4{,}8}{1{,}52} = 3{,}15$$

Testul t de comparare a mediilor a două eşantioane perechi: Soluţie

Concluzie (testul bilateral):

- Deoarece parametrul testului aparţine regiunii critice ipoteza nulă se respinge. Se poate trage concluzia că utilizarea contraceptivelor orale se asociază cu creşterea tensiunii arteriale sistolice.

Testul ANOVA: compararea mediilor a mai multe eşantioane

- H_0 = toate mediile sunt egale.
- H_1 = nu toate mediile sunt egale.

Condiţii de aplicare:

1. Datele sunt independente unele faţă de celelalte.
2. Datele fiecărui grup sunt normal distribuite.
3. Deviaţia standard este aceeaşi pentru toate grupurile.

Testul ANOVA: compararea mediilor a mai multe eşantioane

Id	Medicament					
	A	B	C	D	E	F
1	5	6	7	10	5	9
2	6	8	8	11	8	8
3	7	7	9	13	6	7
4	8	9	11	12	4	5
5	9	10	10	9	7	6
Suma	35	40	45	55	30	35
Media	7	8	9	11	6	7

m=(7+8+9+11+6+7)/6

m=8

$(7-8)^2+(8-8)^2+(9-8)^2+(11-8)^2+(6-8)^2+(7-8)^2 = (-1)^2+0^2+1^2+3^2+(-2)^2+(-1)^2 = 1+0+1+9+4 = 16$

Testul ANOVA: compararea mediilor a mai multe eşantioane

- m=(7+8+9+11+6+7)/6
- m=8
- $(7-8)^2+ (8-8)^2+ (9-8)^2+ (11-8)^2+ (6-8)^2+ (7-8)^2 =$
  $= (-1)^2+ 0^2+ 1^2+ 3^2+ (-2)^2+ (-1)^2 = 1 + 0 + 1 + 9 + 4 = 16$
- Suma pătratelor (între) = ∑(media grupului – media generală)^2×N(numărul de grupuri)
- Suma pătratelor (în) = ∑(valoarea individuală – media grupului)^2
- F = (suma pătratelor(între))/(suma pătratelor(în))

Testul ANOVA: compararea mediilor a mai multe eşantioane

- Suma pătratelor (între) = 16×5 = 80
- Suma pătratelor (în) = $(5-7)^2+(6-7)^2+(7-7)^2+(8-7)^2+(9-7)^2+....+(9-7)^2+(8-7)^2+(7-7)^2+(5-7)^2+(6-7)^2 = 60$
- Cu cât diferenţa dintre suma pătratelor între grupuri este mai mare comparativ cu suma pătratelor în interiorul fiecărui grup cu atât diferenţa între grupurile investigate e mai mare.

	SP	df	Media pătratelor	F = $MP_{între}/MP_{în}$
Între	80	5	= 80/5 = 16	= 16/2,5 = 6,4
În	60	24	= 60/24 = 2,5	
Total	140	29	-	

Testul sumei rangurilor: Wilcoxon

- Aplicat pentru:
  - Un set de observaţii provenite dintr-o valoare ipotetică comună
  - Perechi de observaţii pe aceiaşi indivizi (înainte şi după)
- Utilizat şi pentru a verifica dacă distribuţia diferenţelor are mediana egală sau nu cu zero

Testul sumei rangurilor: Wilcoxon

- Medicaţia intraoculară determină modificarea semnificativă a bătăilor cardiace?
- Suma rangurilor pentru diferenţele negative = 2+3+1 = 6
- Suma rangurilor pentru diferenţele pozitive = 5+4+6+8+7 = 30
- Probabilitatea asociată intersecţiei dintre suma rangurilor negative egală cu 6 cu volumul eşantionului egal cu 8 = 0.109

Înainte	După	Diferenţa	Rang
64	66	-2	2
66	58	8	5
67	68	-1	3
71	65	6	4
72	75	-3	1
76	67	9	6
78	59	19	8
89	74	15	7

Kruskal-Wallis (≥ 3 eşantioane independente)

- Test de ranguri aplicate pe mai mult de 3 eşantioane
- H = parametrul testului
- n = suma volumelor eşantioanelor studiate (n_1, n_2, n_3 ..., n_k)
- T_k = suma rangurilor

$$H = \frac{12}{n(n+1)}\left(\frac{T_1^2}{n_1}+\frac{T_2^2}{n_2}+...\frac{T_k^2}{n_k}\right)-3(n+1)$$

Kruskal-Wallis (≥ 3 eşantioane independente)

- Valoarea antigenului prostatic este diferită la pacienţii cu hipertrofie prostatică benignă, biopsie pozitivă pentru cancer prostatic, biopsie negativă la pacienţi indemni.

	PSA	5.3	7.9	8.7	4.3	6.6	6.4			SumRank
HTPB	Rank	10	17	18	7	13.5	11			76.5
	PSA	7.1	6.6	6.5	14.8	17.3	3.4	13.4	7.6	
BioPoz	Rank	15	13.5	12	21	22	6	20	16	125.5
	PSA	11.4	0.5	1.6	2.3	3.1	1.4	4.4	5.1	
BioPoz	Rank	19	1	3	4	5	2	8	9	51.0

$$H = \frac{12}{22(22+1)}\left(\frac{76,5^2}{6}+\frac{125,5^2}{8}+\frac{51}{8}\right)-3(22+1)$$

$$H = \frac{12}{506}\left(\frac{5852,25}{6}+\frac{15750,25}{8}+\frac{2601,00}{8}\right)-3\cdot 23$$

$$H = 0,02(975,38+1968,78+325,13)-69$$

$$H = 0,03\cdot 3269,28-69 = 77,53-69 = 8,53$$

Kruskal-Wallis (≥ 3 eşantioane independente)

- df = k-1 (k = numărul de eşantioane)
- df = 3-1 = 2
- $H_{critic(\alpha=0,05)} = 5,99$
- H = 8,53 > 5,99 → nivelul PSA este diferit la pacienți cu hipertrofie prostatică benignă, biopsie pozitivă şi respectiv biopsie negativă

Friedman (≥ 3 eşantioane dependente)

- Design randomizat de tip bloc: trei sau mai multe tratamente sunt aplicate aceluiaşi eşantion (extensie a tipului de studiu pe eşantioane perechi)

1. Definirea numărului de tratamente k
2. Obținerea rangurilor pentru fiecare tratament
3. Sumarea rangurilor fiecărui tratament
4. Calcularea parametrului FRIEDMAN (urmează o distribuție χ^2)
5. Dacă $Fr > Fr_{critic}$ → respingem H_0

$$Fr = \frac{12}{n \cdot k(k+1)}\left(T_1^2 + T_2^2 + ... + T_k^2\right) - 3 \cdot n(k+1)$$

Friedman (≥ 3 eşantioane dependente)

- Antigenul prostatic rămâne neschimbat post-terapeutic în cancerul de prostată? PSA a fost măsurat trei ani consecutiv după tratamentul cancerului de prostată la un eşantion de 9 pacienți.

ID	PSA1 (ng/ml)	PSA2 (ng/ml)	PSA3 (ng/ml)	Rang-PSA1	Rang-PSA2	Rang-PSA3
1	3,9	3,95	4,95	1	2	3
2	3,7	3,8	0,1	2	3	1
3	1,8	1,86	3,03	1	2	3
4	0,8	0,3	0,3	3	1,5	1,5
5	3,8	7,68	13,5	1	2	3
6	1,8	2,1	2,54	1	2	3
7	1,8	1,67	0,8	3	2	1
8	3,6	4,51	6,8	1	2	3
9	0,62	0,65	0,42	2	3	1
			T	15	19,5	19,5
			T^2	225	380,25	380,25

Friedman (≥ 3 eşantioane dependente)

- $n = 9$; $k = 3$; $T_1^2 = 225$; $T_2^2 = 380,25$; $T_3^2 = 380,25$

$$Fr = \frac{12}{9 \cdot 3(3+1)}\left(225 + 342,25 + 380,25\right) - 3 \cdot 9(3+1)$$

$$Fr = \frac{12}{108} \cdot 985,50 - 108 = 1,5$$

- $Fr_{critic} = 5,99$
- Fr < Frcritic → nivelul PSA nu creşte în primii 3 ani după intervenția asupra cancerului de prostată

De reținut!

- Atenție la condițiile de aplicare ale fiecărui test!
- Dacă variabilele sunt cantitative continue se verifică inițial normalitatea distribuției.
- Teste de normalitate: Shapiro-Wilk; Kolmogorov-Smirnov; Shapiro-Wilk; Chi-Square Goodness-of-Fit.
- Compararea mediei unui eşantion cu media unei populații (σ): testul Z
- Compararea mediei unui eşantion cu media o medie cunoscută (s): testul t
- Compararea mediilor a 3 sau mai multe eşantioane: ANOVA

De reținut!

- Compararea mediilor a două populații (σ): testul Z
- Compararea mediilor a două eşantioane (s): testul t
- Compararea mediilor a două eşantioane perechi (s): testul t
  - ATENȚIE! Parametrul testului pentru compararea mediilor a două eşantioane nu este acelaşi cu cel pentru compararea a două eşantioane perechi!

De reţinut!

Ranguri

- Distribuţia datelor nu are importanţă!
- Un eşantion sau eşantioane perechi: testul sumei rangurilor (Wilcoxon)
- Trei sau mai multe eşantioane: Kruskal-Wallis
- Trei sau mai multe eşantioane perechi: Friedman

Studii de supravieţuire

Cuprins

Definiţie

- Colecţie de procedee statistice a căror variabilă de interes este o variabilă de supravieţuire

Variabilele de supravieţuire

- Variabilă cantitativă continuă de tip TIMP
  - Timpului scurs între includerea unui subiect într-un studiu şi apariţia unui eveniment predefinit
- Eveniment predefinit:
- Decesul
- Apariţia (unei boli, a unei complicaţii, a unui simptom, a unui semn, a unei metastaze)
- Dispairiţia (unui simptom, a unui semn, etc)
- Remisiunea
- Vindecarea

Utilitatea studiului de supravieţuire

- Să determine gradul în care o nouă medicaţie, o nouă procedură, ar putea avea un efect mai favorabil decât una cunoscută
  - efectele imediate
  - rezultatele de lungă durată
- Durata de timp scursă de la luarea în observaţie până la producerea evenimentului prestabilit

Ce?

- **DESCRIPTIV**: şansa de supravieţuire într-o afecţiune (probabilitatea)
- **COMPARATIV**: comparăm şansa de supravieţuire în situaţii diferite (procedee terapeutice)
- **PREDICTIV**: stabilim legătura între factorii care ar putea fi asociaţi cu timpul de supravieţuire în vederea calculării unor indicatori statistici predictivi

Diagrama de supravieţuire

- x = producerea evenimentului prestabilit;
- o = pierderea subiectului din studiu

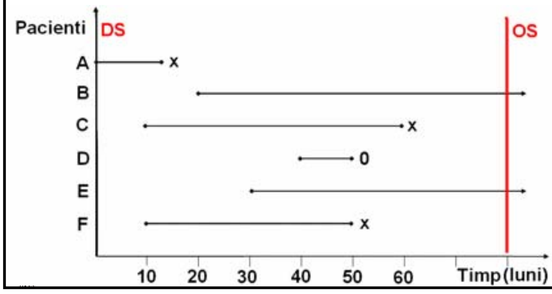


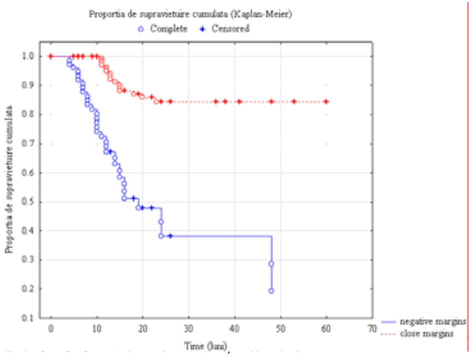
Diagrama de supravieţuire

- **Date de origine**: intrare a subiectului în studiu
- **Timp de participare**: durata de supraveghere a unui subiect care participă în estimarea unei curbe de supravieţuire
- **Timp de recul**: timpul de supraveghere a subiectului luat în studiu în cazul în care NU s-a produs evenimentul prestabilit până la data finală a studiului
- **Data finală**: data încheierii studiului
- **Data ultimei înregistrări**: data la care pentru ultima dată s-au colectat informaţii despre subiectul fără ca evenimentul prestabilit să fi avut loc

Cenzurare

- Pacienţii care nu au ajuns la evenimentul prestabilit
- Observaţiile pe aceşti pacienţi se numesc observaţii **cenzurate la dreapta** (nu se ştie peste cât timp se va produce evenimentul prestabilit):
  - Excluşi în viaţă: la sfârşitul studiului nu s-a produs evenimentul prestabilit
  - Pierduţi din vedere: absenţa urmăririi

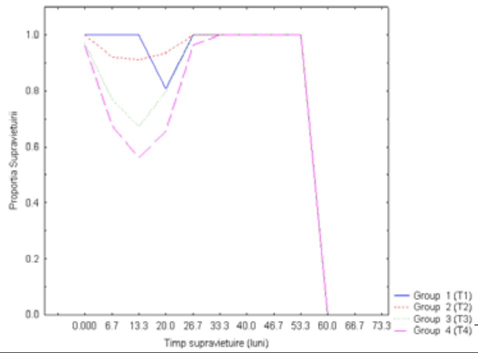
Metode Kaplan Meier



Metode Kaplan Meier

- 2 etape de calcul:
  - calculul probabilităţii de supravieţuire într-un interval
  - calculul probabilităţii de supravieţuire la sfârşitul intervalului
- Furnizează probabilitatea de supravieţuire EXACTĂ, punând la dispoziţie timpul exact de supravieţuire
- Foarte ilustrativ când se doreşte reprezentarea evoluţiei mai multor grupuri pe un acelaşi grafic

Metoda actuarială



Metoda actuarială

- Intervalele **sunt alese arbitrar de cercetător** (număr şi durată).
- Furnizează <u>**probabilitate de supravieţuire APROXIMATIVĂ**</u>
- Două etape de calcul
  - identice cu metoda Kaplan Meier
  - formulele diferă

Compararea curbelor de supravieţuire

- Vizuală
- Teste:
  - Testul Gehan (Wilcoxon generalizat)
  - Testul Logrank
  - Testul Mantel Haenzel

De reţinut!

- Variabila de supravieţuire
- Observaţii cenzurate
- Domenii de aplicare:
  - **Descriptiv**: Calculează şansa de supravieţuire într-o afecţiune (probabilitatea)
    - Metode: **Kaplan Maier** (metodă exactă) & **Actuarială** (metodă aproximativă)
    - Reprezentare grafică: curbe de supravieţuire
  - **Comparativ**: Compară şansa de supravieţuire în situaţii diferite (terapii) prin curbe de supravieţuire şi teste specifice