Dragomirescu L. Lucrari practice de biostatistica. Editia a III-a revazuta si adaugita, 263 pp.
Editura “Agronomica”, Bucuresti, 2003.

Lp 9 Rezumat

6.7. Clasificări ale testelor statistice

6.7.1. Clasificare ca teste binare

	Se respinge H_0 dacă statistica testului, x, este "mai excentrică" decât α-cuantila sau α-cuantilele corespunzătoare:
• *teste unilaterale dreapta*	$x > x_{\alpha/2}$ (conform obs.2 din 6. 4., Lp 8)
• *teste unilaterale stânga*	$x < x_{1-\alpha/2}$ (conform obs.2 din 6. 4., Lp 8)
• *teste bilaterale*	$x < x_{1-\alpha/2}$ sau $x > x_{\alpha/2}$. În cazul distribuțiilor de eşantionaj simetrice față de origine (normala standard şi t_ν - urile), deoarece $x_{1-\alpha/2} = -x_{\alpha/2}$, condiția se poate scrie sintetic $\mid x \mid > x_{\alpha/2}$.

6.7.2. Clasificare după modul de tratare a variabilelor tip măsurătoare

- **Teste parametrice**. Se aplică doar variabilelor tip măsurătoare (**parametri** în sens larg) utilizându-se întreaga informație. Sunt mai puternice, dar cer ca variabilele să îndeplinească anumite condiții în populație (gaussianitate a distribuției, etc.)
- **Teste neparametrice**. Se aplică variabilelor calitative, rangurilor, precum şi celor tip măsurătoare tratate doar ca ranguri sau ca variabile calitative. Testele neparametrice sunt "libere" de forma distribuției variabilelor în populație, dar sunt mai puțin puternice.

6.7.3. Clasificare după tipul ipotezelor de verificat [14]

1° Teste de conformitate (sau de semnificație în sens restrâns)

Compară un eşantion cu o populație în cel mai restrâns mod.
Verifică dacă un eşantion dat poate fi considerat a fi extras dintr-o anumită populație considerând doar un indicator particular. De exemplu,

compară "o medie empirică cu una teoretică". Sn şi **teste de semnificație în sens restrâns**.

2° Teste de concordanță (sau de ajustare)

Compară un eşantion cu o populație în cel mai complet mod.
Verifică dacă un eşantion dat poate fi considerat a fi extras dintr-o anumită populație (teoretică) considerând distribuțiile acestora.

concordanța ⇒ toate conformitățile posibile
(concordanța este o condiție mai tare decât conformitatea)
cel puțin o nonconformitate ⇒ nonconcordanță.

Testele de concordanță resping mai uşor (des) H_0 decât testele de conformitate, dar cer volume mari de date.
Testele de conformitate resping mai greu (rar) H_0 decât testele de concordanță, dar pot lucra cu volume mai mici de date.

La acelaşi volum de date, NU este posibil ca un test de concordanță să accepte H_0 iar un test de conformitate să respingă H_0.

3° Teste de egalitate (sau de omogenitate sau de comparație)

Compară două sau mai multe populații prin tot atâtea eşantioane considerând doar un indicator particular. De exemplu, compară două medii empirice ori *k* medii empirice.
Observație: Cele 3 categorii de teste se întâlnesc în întreaga statistică (uni, bi şi multivariată). În statistica bivariată a variabilelor calitative se adaugă "testele de independență versus asociere".

Teste de conformitate [3]

Problemă biologică

Un anumit microorganism are un diametru mediu de 50 microni (valoare din literatura de specialitate). Un biolog studiază o cultură şi determină pe 10 microorganisme, diametrul mediu $m_1 = 56$ microni şi dispersia necorectată $s_1^2 = 49$ microni. Poate fi microorganismul din literatură sau este altul?

Ipoteza ştiințifică: "Este altul pentru că 56 diferă cam mult de 50", intuitiv, pe eşantion de 10 unități.
Testul de conformitate (Testul Student pentru compararea unei medii empirice cu o medie teoretică):
Cuplul de ipoteze statistice:

$$\begin{cases} H_0 : \mu_1 = 50 \\ H_0 : \mu_1 \neq 50 \end{cases} \text{sau, mai corect,} \begin{cases} H_0 : \mu_1 = \mu_0 (\mu_0 = 50) \\ H_1 : \mu_1 \neq \mu_0 \end{cases}$$

Adică "Media populației din care a fost extras eşantionul cu media $m_1 = 56$ poate fi *50* sau nu?"

Etapele aplicării testului statistic	**Exemplu**
1. Formularea clară a problemei pentru care se doreşte o decizie.	Provine sau nu eşantionul de medie $m_1 = 56$ dintr-o populație cu media $\mu=50$?
2. Identificarea: **a.** variabilei ca tip şi scală **b.** eşantionului • ca mod de extragere • ca volum **c.** informațiilor despre distribuția variabilei în populație (gaussianitate, simetrizabilitate printr-o anumită transformare etc.).	tip măsurătoare, scală raport • eşantion aleator simplu (cu revenire) • $n = 10 < 30$, volum mic Distribuția diametrului unui microorganism într-o populație omogenă este considerată distribuită gaussian.
3. **a.** Stabilirea distribuției de eşantionaj care impune **b.** statistica testului.	• $t = \dfrac{m-\mu}{s_1/\sqrt{n-1}} \in t_{n-1}$ cu n=10 (Spunem că statistica t urmează o lege Student cu n-1 *grade de libertate*.) • $\lvert t_1 \rvert = \left\lvert \dfrac{m_1-\mu}{s_1/\sqrt{n-1}} \right\rvert$
4. Bazat pe 1-3 **a.** formularea cuplului ipoteza nulă - ipoteza alternativă, care determină **b.** tipul de test (bilateral, unilateral stânga, unilateral dreapta)	• H_0: $\mu = 50$ ($=\mu_0$) H_1: $\mu \neq 50$ • test bilateral
5. Stabilirea nivelului de semnificație α sub care vom respinge ipoteza nulă (etapă denumită impropriu **regula de decizie**).	$\alpha = 0{,}05$

6. **a.** Efectuarea calculelor	$t_1 = \dfrac{m_1 - \mu_0}{s_1/\sqrt{n-1}} = \dfrac{56-50}{7/\sqrt{9}} = \dfrac{6}{7/3} \approx \dfrac{6}{2{,}33} \approx 2{,}58$
b. obținerea deciziei statistice prin: ***I)*** calcul manual: dacă statistica testului este « mai excentrică » decât α-cuantila corespunzătoare (din tabela Student) $\Rightarrow$ se respinge H_0.	$\lvert t_1 \rvert = \lvert 2{,}58 \rvert > 2{,}262^*$ $\Rightarrow$ se respinge H_0
II) calcul automat † dacă riscul de respingere a lui H_0 corespunzător statisticii testului (notat α_c) < α fixat la pct. 5 $\Rightarrow$ se respinge H_0.	$\alpha_c \approx 0{,}0287 \approx 0{,}03 < 0{,}05 \Rightarrow$ se respinge H_0

Observație: Deoarece $0{,}01 < \alpha_c \approx 0{,}03 < 0{,}05$, α_c este în zona de incertitudine şi, deci, se recomandă repetarea testului pe un eşantion cu un volum mai mare.

6.8. Teste de concordanță

(Prezentare mai generală decât în [3] conținând şi (**6. 8. 1.) Testarea normalității**.)

Se poate testa concordanța unei distribuții empirice cu orice distribuție teoretică de parametri definitori egali cu cei ai distribuției empirice, de exemplu (1) distribuția uniformă cu acelaşi volum, (2) distribuția produsă de legea Hardy-Weinberg, cu aceleaşi proporții de alele şi acelaşi volum, (3) distribuția normală cu aceeaşi medie, aceeaşi abatere standard şi acelaşi volum, etc.

Se calculează χ^2 pentru măsurarea concordanței între distribuția empirică formată din *c* frecvențe (de clase sau de valori distincte) şi cea teoretică ca în 3. 7. 7., în care am notat cu litera *c*, în loc de *p*, numărul de frecvențe. Am procedat astfel deoarece litera *p* este consacrată, în statistica inductivă, nivelului de semnificație. În plus, o vom utiliza în continuare şi pentru notarea proporției unei anumite alele într-o populație biologică, pentru a păstra şi această notație tradițională în genetică.

* vezi tabela Sudent, pentru t_9, 0,05-cuantila bilaterală, adică valoarea de pe linia 9 coloana marcată jos cu 0,05.

† (executat aici printr-un program propriu)

1° Numărul de grade de libertate, ν

(1) $\nu = c - 1$, în cazul testării concordanței cu o distribuție uniformă,
(2) $\nu = c - 2$, în cazul testării concordanței cu o distribuție produsă de o lege Hardy-Weinberg,
(3) $\nu = c - 3$, în cazul **testării normalității (gaussianității)**, adică a concordanței cu o distribuție normală (gaussiană). (Vezi explicațiile în « Exercițiile sau problemele rezolvate ».)

În general, ν = numărul de clase, c, minus numărul de legături, adică de valori obținute din datele observate care sunt utilizate în calcularea frecvențelor teoretice.

2° Consultarea tabelei χ^2

Tabela conține în interior α-cuantile, iar pe coloane ariile aflate sub curbă la dreapta acestora (probabilitățile de a se obține, din întâmplare, valori χ^2 mai mari decât valoarea α-cuantilei respective). Pe linia ν sunt α-cuantilele distribuțiilor χ^2 cu ν grade de libertate.

3° Testul χ^2 de concordanță

Cuplul de ipoteze statistice:
H_0: distribuțiile concordă
H_A: distribuțiile nu concordă.

Statistica testului este $\chi^2 = \sum_{j=1}^{c} \frac{o_j^2}{t_j} - n$, în care o_j sunt frecvențele observate (empirice), iar t_j sunt frecvențele distribuției teoretice de același volum n. Se fixează $\alpha = 0{,}05$.

Dacă χ^2 (calculat) > $\chi^2_{\nu;\, 0{,}05}$ (tabelat) $\Rightarrow$ distribuția empirică nu concordă cu distribuția teoretică respectivă, neconcordanța fiind semnificativă (* sau $p < 0{,}05$).

În caz contrar $\Rightarrow$ distribuția empirică concordă cu distribuția teoretică respectivă ($p \geq 0{,}05$).

4° Condiții de validitate ale testului

1. Se aplică numai frecvențelor (observate) absolute (în mod necesar valori întregi pozitive), nu celor relative.
2. Frecvențele teoretice (care pot fi și valori fracționare pozitive) $t_j \geq 5$.

Dragomirescu L. Lucrari practice de biostatistica. Editia a III-a revazuta si adaugita, 263 pp.
Editura "Agronomica", Bucuresti, 2003.

6.9. Teste de comparaţie

6.9.1. Compararea a două eşantioane

1° Testul *t* de comparaţie a două medii empirice de observaţii perechi

Fie seriile de câte n observaţii perechi $x_1, x_2,..., x_n$ şi $x'_1, x'_2,..., x'_n$ extrase prin randomizare dintr-una sau două populaţii distribuite gaussian. Se verifică dacă mediile lor m, respectiv m' diferă sau nu semnificativ, adică, notând cu μ şi μ' mediile populaţiilor de origine, testăm cuplul de ipoteze:

$$H_0: \mu = \mu'$$
$$H_A: \mu \neq \mu'.$$

Se calculează diferenţele $d_i = x_i - x'_i$, media lor, m_d şi abaterea standard a diferenţelor, s_d.

Statistica testului: $|t_d|$ unde $t_d = \dfrac{m_d}{s_d / \sqrt{n-1}}$.

Fixăm α.

Numărul de grade de libertate este $n - 1$, deoarece s-a pierdut un grad de libertate estimând dispersia diferenţelor în populaţie, prin dispersia corectată a şirului de diferenţe provenite din eşantioanele de observaţii perechi.

Decizia statistică: Dacă $|t_d|$ (calculat) > $t_{n-1;\ \alpha/2}$ (tabelat) $\Rightarrow$ respingem ipoteza nulă cu risc $p < \alpha$.

2° Testul *t* de comparaţie a două medii empirice de observaţii independente

Fie $x_1, x_2,..., x_{n_A}$ şi $y_1, y_2,..., y_{n_B}$ două *eşantioane de volume mici* (n_A, n_B < 30) *prelevate independent* din una sau două populaţii statistice în care variabila respectivă este distribuită gaussian. Se presupune şi că abaterile standard în cele două populaţii sunt egale, ceea ce se consideră îndeplinit dacă cea mai mare abatere standard a unui eşantion nu depăşeşte două abateri standard ale celuilalt eşantion [16]. Dorim să verificăm dacă mediile lor m_A şi m_B diferă semnificativ sau nu, adică, notând cu μ_A şi μ_B mediile populaţiilor de origine, testăm cuplul de ipoteze:

H_0: $\mu_A = \mu_B$

H_A: $\mu_A \neq \mu_B$.

Statistica testului: $|t_{AB}|$ unde $t_{AB} = \dfrac{m_A - m_B}{s_e\sqrt{\dfrac{1}{n_A}+\dfrac{1}{n_B}}}$, în care

$s_e^2 = \dfrac{n_A \cdot s_A^2 + n_B \cdot s_B^2}{n_A + n_B - 2}$ este **dispersia comună** (corectată) a celor două eşantioane considerate împreună, iar s_A^2 şi s_B^2 sunt dispersiile eşantioanelor (s_e^2 estimează dispersia fiecărei populații).

Fixăm α.

Decizia statistică: Dacă $|t_{AB}|$ calculat > $t_{\nu;\ \alpha/2}$ (tabelat) $\Rightarrow$ respingem ipoteza nulă cu risc $p < \alpha$ unde numărul de grade de libertate $\nu = n_A + n_B - 2$.

Observație: Acest test necesită volume mai mari decât cel pentru observații perechi, deoarece la variabilitatea cercetată s-a adăugat « zgomotul » variabilității aduse de folosirea a două mulțimi de unități statistice, în locul uneia singure.

Lp 9 Teste, exerciții şi probleme

TG9. Durata 100'' pe calculator.

Alegeti propozitia corecta:
Testele statistice sunt de urmatoarele categorii in functie de tipul ipotezelor de verificat:
1. teste bilaterale, unilaterale dreapta si unilaterale stanga;
2. teste parametrice si neparametrice:
3. teste de ajustare (concordanta), de independenta, de conformitate, de egalitate (omogenitate);

Compararea unei medii cu o medie teoretica se poate face prin:
1. testul chi2
2. testul F
3. interval de confidenta sau testul z sau Student

Dispunem de concentratiile de azotati in N lacuri din Delta, in anul 1980 si respectiv in anul 1995, din aceleasi N lacuri.
Pentru a stabili daca, in acest interval de timp, s-a modificat concentratia calculam mediile acestora in cei doi ani si folosim:

1. analiza variantei
2. testul Student pentru observatii perechi
3. testul Student de comparare a doua medii empirice din esantioane independente

Ni se reclama cresterea poluarii cu azotati intr-o zona peste o anumita limita precizata. Pentru testarea afirmatiei se culeg n (> 30) probe si se aplica:
1. testul Student unilateral
2. testul Student bilateral
3. analiza variantei

TC9. Durata 5'.

1. Testele statistice ca teste binare sunt fie teste _______________, fie teste _______________.

2. Testele statistice se construiesc pe baza unor populații statistice _______________, iar testele biomedicale pe baza unor populații statistice ____________.
3. Testele statistice aplicate măsurătorilor (parametrilor în sens larg) care folosesc doar informația referitoare la ranguri sau numai cea corespunzătoare unor variabile calitative, se numesc teste _______________.
4. Utilizarea metodelor parametrice necesită îndeplinirea condiției de ________________ a distribuției caracterului în populație.
5. Testele neparametrice necesită volume mai ________ de date decât testele parametrice pentru acceptarea ________.
6. Cu acelaşi volum de date, testele parametrice pot considera ca semnificative diferențe _____________.
7. Se numesc teste de semnificație în sens larg testele ______________, iar în sens restrâns testele de ________________.
8. Compararea unei medii empirice cu o valoare teoretică este un test de ___________________.
9. Testele de _________________ compară o populație cu un eşantion pe baza maximului de informație, adică prin compararea _______________ acestora.
10. Testarea normalității se realizează prin intermediul statisticii __________ cu ________ grade de libertate.
11. Testul *t* pentru observații perechi este un test de _________________.
12. Testul *t* pentru observații perechi este mai _________________ decât testul *t* de comparație a 2 medii independente.

Exerciții sau probleme rezolvate

1° Teste de conformitate

1.

Să se verifice dacă notarea studenților din anul II Biochimie, în anul universitar 1998-99 la susținerea pentru prima oară a colocviului la disciplina "Teoria probabilităților și statistică matematică" a fost <u>normală ori indulgentă sau exigentă</u> (vezi problema 1 din Lp 4) știind că media generală a celor 48 de studenți prezenți a fost m_1 = 5,75, iar abaterea standard a fost s_1 = 1,59.

Rezolvare:
Trebuie să verificăm conformitatea mediei empirice m_1 = 5,75 cu media teoretică μ_0 = 5,5, valoare care rezultă din problema 1 din Lp 4. Deoarece ne interesează normalitatea versus anormalitatea (indulgență sau exigență

semnificative) testul va fi bilateral, cuplul de ipoteze statistice de verificat fiind H_0: $\mu_1 = \mu_0$; H_A: $\mu_1 \neq \mu_0$, în care μ_1 este media populației statistice din care provine eşantionul cu media m_1. Statistica testului este:

$$|t_1| = \left|\frac{m_1 - \mu_0}{s_1 / \sqrt{n-1}}\right| =$$

$$\left|\frac{5{,}75 - 5{,}5}{1{,}59 / \sqrt{48-1}}\right| \cong \left|\frac{0{,}25}{1{,}59 / 6{,}85}\right| = \left|\frac{0{,}25 \cdot 6{,}85}{1{,}59}\right| = \left|\frac{1{,}7125}{1{,}59}\right| \approx |1{,}08| = 1{,}08..$$

Deoarece n = 48 (> 30) este un volum mare şi σ este estimată suntem în cazul 2.1. (conform 5. 6. 2. din Lp 7) şi deci statistica testului se distribuie normal standard. Ca atare respingem ipoteza nulă dacă t calculat este mai mare în modul decât 0,05-cuantila bilaterală superioară a distribuției normale standard (1,96 – vezi Anexa 2). În cazul de aici deoarece 1,08 < 1,96 acceptăm ipoteza nulă. În concluzie, testul a fost calibrat normal, altfel spus, media 5,75 nu este semnificativ diferită de 5,5.

2° Teste de concordanță

2.

Dintr-o biocenoză s-a extras aleator un eşantion format din 4 specii care au următoarea distribuție de abundențe:

$$\begin{pmatrix} s_1 & s_2 & s_3 & s_4 \\ 30 & 27 & 27 & 16 \end{pmatrix}$$

Să se verifice cu risc α = 0,2 dacă putem considera că speciile din biocenoză au o distribuție echitabilă.

Rezolvare:

O distribuție echitabilă sau regulată (termeni din ecologie) se numeşte în statistică distribuție uniformă. Deci trebuie testată concordanța acestei distribuții empirice cu distribuția uniformă de acelaşi volum n. Ipoteza nulă se poate enunța în acest caz sub forma “*distribuția concordă cu o distribuție uniformă*” sau, mai precis “*distribuția concordă cu distribuția uniformă cu acelaşi volum*”. Analog se reformulează şi H_A. Numărul de grade de libertate $\nu = c - 1$ deoarece din c, numărul de frecvențe (de valori distincte), se scade o doar o unitate pentru că există o singură condiție de legătură: volumul distribuției teoretice trebuie să fie egal cu cel al distribuției empirice (al eşantionului).

I. Calculăm volumul total $n = \sum_{j=1}^{c} o_j$ = 30 + 27 + 27 + 16 = 100.

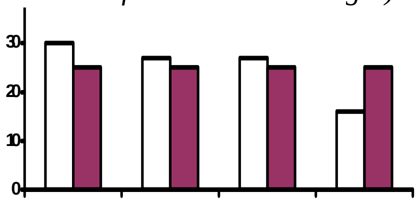
II. Calculăm frecvențele teoretice în ipoteza de uniformitate: $t_j = \dfrac{100}{4} = 25$.

III. Verificăm condiția $t_j \geq 5$ pentru orice j. Într-adevăr, $25 \geq 5$.

IV. Pregătim calculul lui χ^2 pentru testarea concordanței, în tabelul următor:

Nr. frecvență	o_j	o_j^2	t_j	o_j^2 / t_j
1	30	900	25	36,00
2	27	729	25	29,16
3	27	729	25	19,16
4	16	256	25	10,24
Totaluri:	n = 100		n = 100	104,56

Diagrame în batoane pentru cele două distribuții (empirică - batoane albe şi teoretică uniformă - batoane negre):



V. Calculăm $\chi^2 = \sum_{j=1}^{4} \dfrac{o_j^2}{t_j} - n$ = 104,56 –100 = 4,56

VI. Numărul de grade de libertate ν = 4 – 1 = 3. Consultând tabela χ^2 din Anexa 4 pe linia 3 şi coloana lui α = 0,2 obținem valoarea 4,642, notată $\chi^2_{3;\ 0,2}$. Deoarece valoarea calculată (4,56) este mai mică decât cea tabelată pentru nivelul de semnificație α = 0,2 (adică 4,642), acceptăm ipoteza nulă, deci acceptăm concordanța cu distribuția uniformă respectivă.

VII. În concluzie, putem considera că biocenoza are o distribuție echitabilă sau că nu avem suficiente date pentru a considera, eventual, contrariul ($p \geq 0,2$).

3.

Dintr-o biocenoză s-a extras aleator un eşantion format din 5 specii care au următoarea distribuție de abundențe:

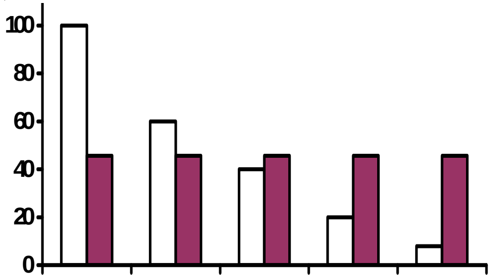
$$\begin{pmatrix} s_1 & s_2 & s_3 & s_4 & s_5 \\ 100 & 60 & 40 & 20 & 8 \end{pmatrix}$$

Să se verifice dacă putem considera că speciile din biocenoză au o distribuţie echitabilă.

Rezolvare:

(I) n = 228. (II) t_j = 45,6. (III) 45,6 ≥ 5.

(IV-V) $\chi^2 = \sum_{j=1}^{5} \frac{o_j^2}{t_j} - n$ = 343,51 – 228 = 115,51. (VI) ν = 5 – 1 = 4. Deoarece valoarea 115,51 este foarte mare consultăm tabela χ^2 din Anexa 4 la cel mai înalt nivel de semnificaţie şi anume α = 0,001. $\chi^2_{4;\,0,001}$ = 18,467.



Deoarece valoarea calculată (115,51) este mai mare decât cea tabelată ($\chi^2_{4;\,0,001}$ = 18,467), respingem ipoteza nulă cu risc sub 1‰. Deci putem afirma cu risc sub 1‰ că “distribuţia empirică nu concordă cu distribuţia uniformă cu acelaşi volum” sau putem afirma că “între distribuţia empirică şi cea uniformă cu acelaşi volum există o discordanţă înalt semnificativă (*** p < 0,001)”. (VII) În concluzie, putem considera cu risc sub 1‰ că biocenoza nu are o distribuţie echitabilă (*** p < 0,001).

4.

Într-un eşantion extras aleator dintr-o populaţie biologică, frecvenţele absolute ale genotipurilor AA, Aa şi aa sunt $\begin{pmatrix} AA & Aa & aa \\ 100 & 120 & 30 \end{pmatrix}$. Să se verifice dacă împerecherea în cadrul populaţiei se face la întâmplare.

Rezolvare de principiu:

În **genetica populațiilor** [7] arată că dacă împerecherea se face la întâmplare, frecvențele <u>relative</u> genotipice trebuie să aibă, conform legii Hardy-Weinberg, distribuția $\begin{pmatrix} AA & Aa & aa \\ p^2 & 2pq & q^2 \end{pmatrix}$, în care p, respectiv q sunt frecvențele relative în populație ale alelelor A, respectiv a.

Din punct de vedere statistic, verificarea cerută pentru genotipuri presupune testarea concordanței distribuției de frecvențe <u>absolute</u> observate (o_j) cu distribuția de frecvențe <u>absolute</u> teoretice în ipoteza aplicării legii Hardy-Weinberg (t_j), distribuție cu acelaşi volum n. Cuplul de ipoteze de verificat este: *H_0: distribuțiile concordă; H_A: distribuțiile nu concordă*, ceea ce, în acest caz, înseamnă pentru H_0 că “*distribuția concordă cu distribuția produsă de legea Hardy-Weinberg, distribuție cu aceleaşi proporții de alele şi acelaşi volum*”.

Statistica testului este $\chi^2 = \sum_{j=1}^{3} \frac{o_j^{\,2}}{t_j} - n$. Dacă χ^2 calculat este mai mare decât cel tabelat la $\nu = 3 - 2 = 1$ grad de libertate, se respinge ipoteza nulă. (În caz contrar, se acceptă.) Din 3, numărul de frecvențe (de valori distincte), se scad două unități deoarece există două condiții de legătură: prima este dată de faptul că volumul distribuției teoretice trebuie să fie egal cu cel al distribuției empirice (al eşantionului) n, iar a doua provine din faptul că frecvențele teoretice au fost calculate pe baza frecvenței p estimată prin datele eşantionului. (Deoarece q derivă din p fiind egal cu $1 - p$, calculul lui q nu aduce o nouă legătură).

Revenind în cadrul **geneticii populațiilor**, vom accepta împerecherea întâmplătoare atunci când se acceptă ipoteza nulă şi o vom respinge în caz contrar.

Rezolvare efectivă (Etape de calcul):

I. Estimăm p şi q în populație prin valorile lor din eşantion:

Indivizi	*Alele A*	*Alele a*	*Total alele:*
100 *AA* 120 *Aa* 30 *aa*	$2 \cdot 100 =$ 200 120	120 $2 \cdot 30 = 60$	
Totaluri:	320	180	500

Calculăm frecvențele relative ale alelelor:

$p = 320 / 500 = 0{,}64$

$q = 180 / 500 = 0{,}36$

(Verificare: $p + q = 0{,}64 + 0{,}36 = 1$.)

II. Calculăm volumul total de indivizi $n = \sum_{j=1}^{3} o_j = 100 + 120 + 30 = 250.$

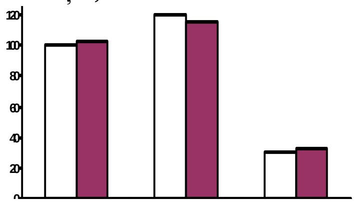
III. Calculăm pentru genotipuri:

a. frecvențele <u>relative</u> teoretice în ipoteza aplicării legii Hardy-Weinberg:	$p^2 = 0{,}64^2 = 0{,}4096$	$2 \cdot p \cdot q = 2 \cdot 0{,}64 \cdot 0{,}36 = 0{,}4608$	$q^2 = 0{,}36^2 = 0{,}1296$	*(Verificare: $p^2+2pq+q^2 = 1$)*
b. frecvențele <u>absolute</u> teoretice în ipoteza aplicării legii Hardy-Weinberg, prin multiplicarea cu *n* (= 250) a frecvențelor relative:	$t_1 = p^2 \cdot n = 0{,}4096 \cdot 250 = 102{,}4$	$t_2 = 2 \cdot p \cdot q \cdot n = 0{,}4608 \cdot 250 = 115{,}2$	$t_3 = q^2 \cdot n = 0{,}1296 \cdot 250 = 32{,}4$	*(Verificare: $t_1 + t_2 + t_3 = n$)*

IV. Verificăm condiția $t_j \geq 5$ pentru orice *j*. Într-adevăr, $102{,}4 \geq 5$ $115{,}2 \geq 5$; $32{,}4 \geq 5$.

V. Pregătim calculul lui χ^2 pentru testarea concordanței, în tabelul următor:

Genotip	o_j	o_j^2	t_j	o_j^2 / t_j
AA	100	10000	102,4	97,6563
Aa	120	14400	115,2	125,0000
aa	30	900	32,4	27,7778
Totaluri:	250		250	250,4341



VI. Calculăm $\chi^2 = \sum_{j=1}^{3} \frac{o_j^2}{t_j} - n \approx 250{,}434 - 250 = 0{,}434.$

VII. Numărul de grade de libertate $\nu = 1$. Consultând tabela χ^2 din Anexa 4 pe linia 1 şi coloana lui $\alpha = 0{,}05$ obținem valoarea 3,841 notată $\chi^2_{1;0,05}$. Deoarece valoarea calculată (0,434) este mai mică decât cea tabelată pentru nivelul de semnificație standard $\alpha = 0{,}05$ (adică 3,841), acceptăm ipoteza nulă, deci acceptăm concordanța cu distribuția dată de legea Hardy-Weinberg.

Observația 1: În acest caz $0{,}434 < \chi^2_{1;0,2} = 1{,}642$, deci respingerea s-ar face cu un risc şi mai mare, $p \geq 0{,}2$.

(*Observația 2:* A nu se confunda această probabilitate *p*, cu proporția *p* din distribuția dată de legea Hardy-Weinberg, ambele notații tradiționale.)

VIII. În concluzie, putem considera că împerecherea se produce la întâmplare în populația respectivă ($p \geq 0{,}05$, de fapt, ținând cont de observația 1, $p \geq 0{,}2$) sau că nu avem suficiente date pentru a considera, eventual, contrariul.

Observație: Referitor la legea Hardy-Weinberg pentru două alele pe același locus putem formula următoarea regulă simplă:
Dacă $\chi^2 < 3{,}841$ considerăm că împerecherea se produce la întâmplare în populația respectivă sau că nu avem date suficiente pentru a considera, eventual, contrariul. Dacă $\chi^2 \geq 3{,}841$ afirmăm cu un risc sub 5% că împerecherea nu se produce la întâmplare în populația respectivă.

5.

Într-o populație biologică frecvențele absolute ale genotipurilor *MM, MN* și *NN* sunt $\begin{pmatrix} MM & MN & NN \\ 200 & 300 & 200 \end{pmatrix}$. Să se verifice dacă împerecherea în cadrul populației se face la întâmplare.

Rezolvare:

Considerăm populația biologică drept un eșantion extras aleator de către natură dintr-o populație statistică pentru care vrem să verificăm existența legii Hardy-Weinberg și aplicăm metodologia din problema anterioară.
(I) $p = 0{,}5$; $q = 0{,}5$ (II) $n = 700$. (IIIa) $p^2 = 0{,}25$; $2\cdot p\cdot q = 0{,}5$; $q^2 = 0{,}25$. (IIIb) $t_1 = 175$; $t_2 = 350$; $t_3 = 175$. (IV) $175 \geq 5$; $350 \geq 5$; $175 \geq 5$. (V-VI) $\chi^2 = 14{,}286$. (VII) Deoarece valoarea calculată (14,286) este mai mare decât cea tabelată în Anexa 4 linia 1 ($\chi^2_{1;\ 0{,}001} = 10{,}827$) respingem ipoteza nulă cu risc sub 1‰. Deci putem afirma cu risc sub 1‰ că “distribuția empirică nu concordă cu distribuția cu aceeași proporție și același volum dată de legea Hardy-Weinberg” sau, mai precis, că între acestea “există o discordanță înalt semnificativă (*** $p < 0{,}001$).”. (VIII) Afirmăm cu risc sub 1‰ că împerecherea nu se produce la întâmplare în populația respectivă (*** $p < 0{,}001$).

6.

Să se testeze “normalitatea” (vezi problema 1 din Lp 4) distribuției notelor finale obținute de anul II Biochimie, în anul universitar 1998-99, la susținerea pentru prima oară a colocviului la disciplina “Teoria probabilităților și statistică matematică”:

$$\begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 1 & 1 & 12 & 5 & 13 & 10 & 5 & 0 & 1 \end{pmatrix}.$$

Rezolvare:

Cuplul de ipoteze de verificat este: H_0*: distribuțiile concordă;* H_A*: distribuțiile nu concordă.* Pentru H_0 putem spune și că “*distribuția concordă cu o distribuție normală*” ceea ce înseamnă că “*distribuția concordă cu distribuția* *normală cu aceeaşi medie, aceeaşi abatere standard şi acelaşi volum”.* Statistica testului este $\chi^2 = \sum_{j=1}^{c} \frac{o_j^{\ 2}}{t_j} - n$, în care o_j sunt frecvențele observate (aici cele de mai sus), t_j sunt frecvențele teoretice ale distribuției normale de aceeaşi medie şi abatere standard, iar n reprezintă volumul fiecărei distribuții. Dacă χ^2 calculat este mai mare decât cel tabelat pentru $\nu = c - 3$ (c fiind numărul de frecvențe de clase) se respinge ipoteza nulă. În caz contrar, se acceptă. Observăm că se pierd 3 grade de libertate prin utilizarea (1) mediei, (2) abaterii standard corectate ale eşantionului pentru estimarea mediei şi abaterii standard ale distribuției normale, precum şi (3) volumului eşantionului în calculul frecvențelor teoretice ale distribuției normale.

I. Calculăm media şi dispersia distribuției empirice, prin formulele de calcul rapid şi exact pentru o distribuție de frecvențe absolute (vezi observația din EGA 1, Lp 3), precum şi abaterea standard:

o_j	x_j	$x_j^{\ 2}$	$o_j \cdot x_j$	$o_j \cdot x_j^{\ 2}$
1	2	4	2	4
1	3	9	3	9
12	4	16	48	192
5	5	25	25	125
13	6	36	78	468
10	7	49	70	490
5	8	64	40	320
1	10	100	10	100
n = 48	Sume:		T_1 = 276	T_2 = 1708

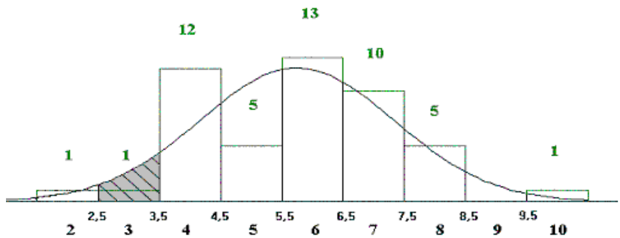
$$m = \frac{T_1}{n} = \frac{276}{48} = 5{,}75$$

$$s^2 = \frac{T_2}{n} - M^2 = \frac{1708}{48} - 5{,}75^2 = $$
$$\approx 35{,}5833 - 33{,}0625 = 2{,}5208$$

$$s = \sqrt{2{,}5208} \approx 1{,}6$$

II. Determinăm frecvențele distribuției normale de medie 5,75 şi abatere standard 1,6 corespunzătoare intervalelor notelor. Pentru aceasta:

a. stabilim limitele claselor (intervalelor), y_j, care reprezintă notele (de exemplu, nota 2 înseamnă scor < 2,5; nota 3 înseamnă scor în intervalul [2,5; 3,5), …, nota 10 înseamnă scor ≥ 9,5.



b. calculăm scorurile z_j pentru fiecare limită de intervale, conform formulei $z_j = \dfrac{y_j - m}{s}$ (vezi Lp 3) şi pentru fiecare scor z_j consultăm Anexa 2, determinând astfel ariile relative aflate la dreapta limitelor de intervale:

c.

y_j	z_j	*arii relative (α_i) la dreapta pct. z_i*	*arii relative între limite consecutive $d_i=(\alpha_{i-1}-\alpha_i)$*	*arii (frecvențe) absolute între limite consecutive ($t_j = n \cdot d_i = 48 \cdot d_i$)*
-∞	-∞	1		
			0,0212	1,02
2,5	-2,03	0,9788		
			0,0581	**2,79**
3,5	-1,41	0,9207		
			0,1384	6,64
4,5	-0,78	0,7823		
			0,2187	10,50
5,5	-0,16	0,5636		
			0,2444	11,73
6,5	0,47	0,3192		
			0,1813	8,70
7,5	1,09	0,1379		
			0,0952	4,57
8,5	1,72	0,0427		
			0,0331	1,59
9,5	2,34	0,0096		
			0,0096	0,46
		Totaluri de control:	1	48

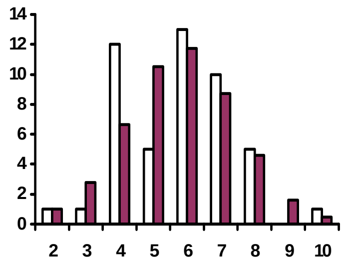
În penultima coloană am calculat ariile relative cuprinse între două limite consecutive (de exemplu aria gri, haşurată de mai sus care este **2,79**), arii ce se determină scăzând aria mai mică din cea mai mare, iar în ultima coloană am scris ariile absolute sub distribuția normală de medie 5,75, abatere standard 1,6 şi volum 48, adică frecvențele distribuției normale care ne interesează. Acestea se determină prin amplificarea cu volumul n (= 48) a datelor din penultima coloană.

III. În vederea calculării lui χ^2:

a. centralizăm rezultatele în tabelul următor, după modelul din [3] pagina 157:

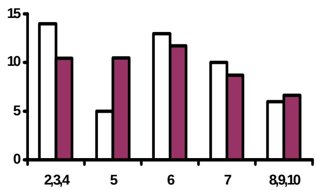
Note	*Intervale*	*Frecvențe observate* o_j	*Frecvențe teoretice (în ipoteza de normalitate)* t_j
2	$(-\infty, 2{,}5)$	1	$\rightarrow$1,02
3	[2,5; 3,5)	1 }14	$\rightarrow$ 2,79 }10,45
4	[3,5; 4,5)	12	6,64
5	[4,5; 5,5)	5	10,50
6	[5,5; 6,5)	13	11,73
7	[6,5; 7,5)	10	8,70
8	[7,5; 8,5)	5	$\rightarrow$ 4,60
9	[8,5; 9,5)	0 }6	$\rightarrow$1,59 }6,65
10	$[9{,}5; +\infty)$	1	$\rightarrow$ 0,46

Diagrame în batoane pentru cele două distribuții (empirică - batoane albe și normală - batoane negre):



b. dar deoarece există frecvențe teoretice mai mici decât 5 (cele marcate cu semnul $\rightarrow$), pentru ca testul χ^2 să fie valid, comasăm intervalele cu frecvențele marcate prin acoladă, obținând următorul tabel la care am adăugat coloanele necesare calculului lui χ^2 (vezi, de asemenea, modelul din [3] pagina 157):

Notele	Intervalele	o_j	o_j^2	t_j	o_j^2 / t_j
2, 3, 4	$(-\infty; 4{,}5)$	14	196	10,45	18,76
5	[4,5; 5,5)	5	25	10,50	2,38
6	[5,5; 6,5)	13	169	11,73	14,41
7	[6,5; 7,5)	10	100	8,70	11,49
8, 9, 10	$[7{,}5; +\infty)$	6	36	6,65	5,41
	Totaluri:	48		48,03	52,45



IV. Calculăm $\chi^2 = \sum_{j=1}^{5} \frac{o_j^{\,2}}{t_j} - n$ = 52,45 – 48 = 4,45.

V. Numărul de grade de libertate ν = 5 – 3 = 2. Consultând tabela χ^2 din Anexa 4 pe linia 2 şi coloana lui α = 0,05 obținem valoarea $\chi^2_{2;\ 0,05}$ = 5,991. Deoarece 4,45 < $\chi^2_{2;\ 0,05}$ = 5,991, acceptăm ipoteza nulă, deci acceptăm concordanța cu distribuția normală respectivă ($p \geq 0{,}05$). Deoarece 4,45 < $\chi^2_{2;\ 0,1}$ = 4,605, concordanța este chiar mai bună ($p \geq 0{,}1$).

VI. În concluzie, notele obținute se distribuie normal, ceea ce trebuie să se întâmple atunci când un lot este pregătit omogen, iar modul de notare este bine calibrat. (Acest lucru s-a produs deoarece nota fiecărui student este o medie ponderată a câte 3 note primite la fiecare lucrare practică şi a trei note din colocviu. Astfel, notele finale nu au un caracter conjunctural, ci descriu comportamentul pe întregul semestru. Prin combinarea a foarte multe teste aplicate continuu s-a redus la minimum falsul pozitiv şi cel negativ al notării finale, care astfel poate fi normală în condițiile de mai sus: omogenitatea pregătirii şi buna calibrare.)

3° Teste de comparație

7.

În vederea evaluării rezistenței la un agent poluant a două specii de peşti, două loturi extrase aleator din cele două specii au fost supuse unui tratament cu aceeaşi concentrație de agent poluant. S-au notat timpii de supraviețuire ai fiecărui individ (exprimați în ore) şi au rezultat seriile următoare:

Specia A: 12 10 14 11 12 15;
Specia B: 17 15 14 20 18 17.

Cele două specii rezistă la fel la concentrația dată a agentului poluant sau nu?

Rezolvare cu tabele:
Tratare parametrică, deoarece distribuția caracterului în populație este gaussiană (legea toleranței). Notând cu μ_A, respectiv μ_B mediile celor două populații, cuplul de ipoteze statistice de verificat este: H_0: $\mu_A = \mu_B$; H_A: $\mu_A \neq \mu_B$ (adică trebuie să aplicăm un test bilateral). Volumele sunt mici (< 30) şi dacă $max\ \{s_A, s_B) \leq 2 \cdot min\ \{s_A, s_B)$ vom aplica *testul t de comparație a două medii empirice de observații independente*.

Calcul:
Pentru calculul simultan al mediilor m_A şi m_B şi al dispersiilor s_A^2 şi s_B^2 alcătuim tabela următoare:

Nr. crt.	A_i	$(A_i)^2$	B_i	$(B_i)^2$
1	12	144	17	289
2	10	100	15	225
3	14	196	14	196
4	11	121	20	400
5	12	144	18	324
6	15	225	17	289
Totaluri	$T_{1A} = 74$	$T_{2A} = 930$	$T_{1B} = 101$	$T_{2B} = 1723$

Notând cu n_A, respectiv cu n_B volumele celor două eşantioane avem:

$$n_A = 6; \qquad m_A = \frac{T_{1A}}{n_A} = \frac{74}{6} = 12{,}33;$$

$$s_A^2 = \frac{T_{2A}}{n_A} - m_A^2 = \frac{930}{6} - 12{,}33^2 = 155 - 152{,}0289 = 2{,}9711;$$

$$s_A = \sqrt{s_A^2} = \sqrt{2{,}9711} \approx 1{,}72;$$

$$n_B = 6; \qquad m_B = \frac{T_{1B}}{n_B} = \frac{101}{6} = 16{,}83;$$

$$s_B^2 = \frac{T_{2B}}{n_B} - m_B^2 = \frac{1723}{6} - 16{,}83^2 = 287{,}1667 - 283{,}2489 = 3{,}9178;$$

$$s_B = \sqrt{s_B^2} = \sqrt{3{,}9178} \approx 1{,}98;$$

Observăm că $1{,}98 \le 2\cdot 1{,}72 = 3{,}14$. Deci putem aplica testul *t*.
Calculăm *dispersia comună*

$$s_e^2 = \frac{n_A \cdot s_B^2 + n_A \cdot s_B^2}{n_A + n_B - 2} = \frac{6 \cdot 2{,}7911 + 6 \cdot 3{,}9178}{6+6-2} = \frac{16{,}7466 + 23{,}5068}{10} = 4{,}2534.$$

Deci:

$$s_e = \sqrt{s_e^2} = \sqrt{4{,}2534} \approx 2{,}06$$

şi $$t_{AB} = \frac{m_A - m_B}{s_e\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} = \frac{12{,}33 - 16{,}83}{2{,}06\sqrt{\frac{1}{6} + \frac{1}{6}}} \approx \frac{-4{,}5}{2{,}06\sqrt{0{,}3333}}$$

$\approx \frac{-4{,}5}{2{,}06 \cdot 0{,}58} \approx \frac{-4{,}5}{1{,}1948} =$ - **3,**766. Deci $|t_{AB}| = 3{,}766$. Numărul de grade de libertate $\nu = n_A + n_B - 2 = 6 + 6 - 2 =$ **10**.

Consultăm tabela Student (vezi Anexa 3) pe linia corespunzătoare numărului de grade de libertate (10 aici) şi determinăm cele două „valori critice" între care se plasează ca mărime valoarea calculată $|t_{AB}|$ (3,766 aici). Vom găsi valorile 3,169 şi 4,144. Citim nivelurile de semnificaţie α corespunzătoare testului bilateral (vezi linia de jos a tabelei). Acestea sunt 0,01 respectiv 0,002. în concluzie putem respinge ipoteza nulă (egalitatea mediilor aici) cu un risc < 0,01 (dar nu mai mic şi decât 0,002). Altfel spus, putem afirma o diferenţă semnificativă cu un risc sub 1%. Dacă luăm în consideraţie pragurile standard (0,05; 0,01; 0,001) putem afirma existenţa unei diferenţe foarte semnificative (p < 0,01) sau (**).

Decizie statistică finală: Deoarece $|t_{AB}| = 3{,}766 > 3{,}169$, se respinge ipoteza nulă cu risc $p < 0{,}01$.

Decizie de specialitate: „Putem afirma cu un risc p sub 1% că rezistenţa celor două specii la agentul poluant testat diferă", sau că „rezistenţa celor două specii la agentul poluant testat diferă foarte semnificativ (**)".

Notă: Dacă valoarea calculată ar fi fost mai mare decât cel mai mare prag tabelat (4,587 în acest caz), atunci am fi afirmat că „există diferenţă înalt semnificativă (p < 0,001) sau (***)". Dacă valoarea calculată ar fi fost mai mică decât cel mai mic prag tabelat (0,879 aici) spuneam că „diferenţa nu este semnificativă ori nu avem date suficiente pentru a demonstra, eventual, contrariul".

Rezolvare prin programul EpiInfo:

Se utilizează comanda MEANS din modulul ANALYSIS. Programul execută *ANOVA* (test pentru compararea simultană a mai multor medii) care este echivalent, în cazul a două eşantioane, cu *testul t de comparaţie a două medii empirice de observaţii independente*. Rezultatul final este α_c indicat de program sub denumirea „p-value". În această problemă aceasta are valoarea 0,003278. Astfel, vom putea formula mai precis deciziile finale:

Decizie statistică finală: Se respinge ipoteza nulă cu riscul ***p* < 0,003278** < 0,01.

Decizie de specialitate: „Putem afirma cu un risc ***p* < 0,003278**, sub 1%, că rezistenţa celor două specii la agentul poluant testat diferă", sau că „rezistenţa celor două specii la agentul poluant testat diferă foarte semnificativ (**, *p* < **0,003278**)".

8 (Enunţ din [1] modificat).

Un lot de 10 hipertensivi au primit un tratament destinat diminuării tensiunii arteriale. S-au observat următoarele valori ale tensiunii arteriale sistolice, măsurate în *cm* Hg:

Număr subiect	1	2	3	4	5	6	7	8	9	10
Tensiune sistolică înainte de tratament	15	18	17	20	21	18	17	15	19	16
Tensiune sistolică după tratament	12	16	17	18	17	15	18	14	16	18

Fixând α = 0,025 (conform observației 2 din 6. 4., Lp 8), este eficace acest tratament ?

Rezolvare:

Cele două eşantioane sunt eşantioane de observații perechi extrase din populații care pot fi considerate distribuite gaussian: populația dinaintea tratamentului pentru că marea majoritate a parametrilor fiziologici sunt astfel distribuiți, iar cea de după tratament, pentru că este vorba de reacția unei populații biologice la un anumit factor de mediu - tratamentul (legea toleranței). Notând cu μ, respectiv μ' mediile celor două populații (înainte, respectiv după tratament) cuplul de ipoteze statistice de verificat este: H_0: $\mu = \mu'$; H_A: $\mu > \mu'$ (adică trebuie să aplicăm un test unilateral dreapta). Populațiile fiind gaussiene putem aplica testul Student pentru observații perechi. Statistica testului este:

$$t_d = \frac{m_d}{s_d / \sqrt{n-1}},$$

m_d fiind media diferențelor valorilor perechi, s_d abaterea standard corectată, a aceloraşi diferențe, iar n numărul perechilor. Pentru calculul celor două statistici vom utiliza tabelul:

Nr.pereche	x_i	x'_i	$d_i = x_i - x'_i$	$d_i - m_d$	$(d_i - m_d)^2$
1	15	12	3	1,5	2,25
2	18	16	2	0,5	0,25
3	17	17	0	-1,5	2,25
4	20	18	2	0,5	,25
5	21	17	4	2,5	6,25
6	18	15	3	1,5	2,25
7	17	18	-1	-2,5	6,25
8	15	14	1	-,5	,25
9	19	16	3	1,5	2,25
10	16	18	-2	-3,5	12,25
Totaluri:	176	161	15	*Verificare:* 0	34,5

n = 10, deci

$$m_d = \frac{\sum d_i}{n} = \frac{15}{10} = 1{,}5\,;\; m = \frac{\sum x_i}{n} = \frac{176}{10} = 17{,}6\,;\; m' = \frac{\sum x'_i}{n} = \frac{161}{10} = 16{,}1\,;$$

$$s_d^2 = \frac{\sum (d_i - m_d)^2}{n} = \frac{34{,}5}{10} = 3{,}45 \Rightarrow s_d = \sqrt{3{,}45} = 1{,}86 \Rightarrow$$

$$t_d = \frac{m_d}{s_d / \sqrt{n-1}} = \frac{1{,}5}{1{,}86/3} = 2{,}42\,.$$

Din tabela Student (vezi Anexa 3 din acest volum) se citeşte 0,025 cuantila unilaterală superioară pentru distribuția t cu ν grade de libertate, unde ν (= n – 1) = 9. Adică se citeşte tabela pe linia 9 şi coloana corespunzătoare lui α = 0,025 citit de sus în jos. Se obține $t_{9;0,025}$ =2,262.

Decizie statistică finală: Deoarece t = 2,42 > 2,262, se respinge ipoteza nulă cu risc p < 0,025.

Decizie de specialitate: Putem afirma cu un risc p sub 2,5% că tratamentul este eficace.

Exerciții sau probleme propuse

1° Teste de conformitate

9.

Deoarece media generală a notelor studenților din anul II Biochimie (în anul universitar 1998-99 la susținerea pentru prima oară a colocviului la disciplina "Teoria probabilităților şi statistică matematică") a fost 5,75, să se verifice dacă notarea a fost <u>prea indulgentă</u>, media generală pentru un test bine calibrat fiind 5,5 (conform problemei 1 din Lp 4). Abaterea standard a notelor a fost s_1 = 1,59.

2° Teste de concordanță

10.

Dintr-o biocenoză s-a extras aleator un eşantion care are următoarea distribuție de abundențe:

$$\begin{pmatrix} s_1 & s_2 & s_3 & s_4 \\ 30 & 20 & 10 & 4 \end{pmatrix}$$

Să se verifice dacă putem considera că biocenoza are o distribuție echitabilă.

11.

Dintr-o biocenoză s-a extras aleator un eşantion care are următoarea distribuţie de abundenţe:

$$\begin{pmatrix} s_1 & s_2 & s_3 & s_4 & s_5 & s_6 \\ 20 & 22 & 17 & 18 & 19 & 24 \end{pmatrix}$$

Să se verifice dacă putem considera că biocenoza are o distribuţie echitabilă.

12.

Un zar a fost aruncat de 120 de ori obţinându-se distribuţia din problema anterioară. Să se decidă dacă zarul este măsluit.

13.

Un ban a fost aruncat de 30 de ori obţinându-se distribuţia $\begin{pmatrix} Stema & Banul \\ 14 & 16 \end{pmatrix}$.

Să se decidă dacă banul este măsluit.

14.

Un ban a fost aruncat de 3000 de ori obţinându-se acelaşi raport între numărul de apariţii ale stemei şi cel de apariţii ale banului ca la problema anterioară, adică 14 / 16. Să se decidă dacă banul este măsluit.

15.

Într-un eşantion extras aleator dintr-o populaţie biologică, frecvenţele absolute ale genotipurilor *AA*, *Aa* şi *aa* sunt $\begin{pmatrix} AA & Aa & aa \\ 500 & 300 & 200 \end{pmatrix}$. Să se verifice dacă împerecherea în cadrul populaţiei se face la întâmplare.

16.

Într-un eşantion extras aleator dintr-o populaţie biologică, frecvenţele absolute ale genotipurilor *AA*, *Aa* şi *aa* sunt $\begin{pmatrix} AA & Aa & aa \\ 350 & 500 & 150 \end{pmatrix}$. Să se verifice dacă împerecherea în cadrul populaţiei se face la întâmplare.

17.

Să se testeze "normalitatea" distribuţiei notelor următoare:

$$\begin{pmatrix} 7 & 8 & 9 & 10 \\ 15 & 20 & 35 & 30 \end{pmatrix}.$$

3° Teste de egalitate

18.

În vederea evaluării rezistenţei la un agent poluant a două specii de peşti, două loturi extrase prin randomizare din cele două specii au fost supuse unui tratament cu aceeaşi concentraţie de agent poluant. S-au notat timpii de supravieţuire ai fiecărui individ (exprimaţi în ore) şi au rezultat seriile următoare:
Specia A: 2 4 7 10 12 14 11 15 15 21 4 9 12 12 18 28;
Specia B: 13 4 7 12 11 17 25 16 17 21 11 31 23 23 35.
Cele două specii rezistă la fel la concentraţia dată a agentului poluant sau nu ?

19.

Pentru testarea eficacităţii unei culegeri de probleme s-au selecţionat prin randomizare 10 studenţi. S-au notat mediile obţinute de aceştia înaintea accesului la culegere, precum şi mediile notelor obţinute după ce studenţii au început să utilizeze culegerea. Mediile au fost următoarele:

Număr subiect	1	2	3	4	5	6	7	8	9	10
Media înaintea utilizării culegerii de probleme	5	6	7	7	8	9	8	6	5	8
Media după utilizarea culegerii de probleme	7	7	5	7	9	8	10	9	8	10

Este eficace culegerea de probleme ?