

Curs 11

Cristian Niculescu

1 Testarea semnificației ipotezei 0 I

1.1 Scopurile învățării

1. Să știe definițiile termenilor de testare a semnificației: NHST, ipoteză 0, ipoteză alternativă, ipoteză simplă, ipoteză compusă, nivel de semnificație, putere.
2. Să poată face un test de semnificație pentru date Bernoulli și binomiale.
3. Să poată calcula o p -valoare pentru o ipoteză normală și s-o utilizeze într-un test de semnificație.

1.2 Introducere

Statistica frecvenționistă este adesea aplicată în cadrul testării semnificației ipotezei 0 (NHST). Paradigma [Neyman-Pearson](#) se concentrează pe o ipoteză numită [ipoteza 0](#). Sunt și alte paradigme pentru testarea ipotezelor, dar Neyman-Pearson este cea mai uzuală. Spus simplu, această metodă întreabă dacă datele sunt în afara regiunii unde ne-am aștepta să le vedem sub ipoteza 0. Dacă este așa, respingem ipoteza 0 în favoarea unei a 2-a ipoteze, numită ipoteza alternativă.

Calcululele făcute aici necesită toate funcția de verosimilitate. Sunt 2 diferențe majore între ce vom face aici și ce am făcut la actualizarea Bayesiană:

1. Dovezile datelor vor fi considerate doar prin funcția de verosimilitate, nu vor mai fi ponderate de convingerile noastre a priori.
2. Vom avea nevoie de o noțiune de date extreme, de exemplu 95 de aversuri din 100 de aruncări ale unei monede sau o muscă efemeră care trăiește o lună.

1.2.1 Exemple motivante

Exemplul 1. Presupunem că vrem să decidem dacă o monedă este corectă. Dacă obținem 85 de aversuri din 100 de aruncări, ați considera că moneda este probabil incorectă? Dar la 60 de aversuri? Sau la 52? Majoritatea

oamenilor ar ghici că 85 de aversuri este o dovadă puternică pentru faptul că moneda este incorectă în timp ce 52 de aversuri nu este deloc dovadă. 60 de aversuri este mai puțin clar. Testarea semnificației ipotezei 0 (NHST) este o abordare frecvenționistă pentru a gândi cantitativ despre aceste întrebări.

Exemplul 2. Presupunem că vrem să comparăm un tratament medical nou cu un placebo sau cu standardul curent de îngrijire. Ce fel de dovezi v-ar convinge că noul tratament este mai bun decât placebo sau standardul curent? Din nou, NHST este un cadru cantitativ pentru a răspunde acestei întrebări.

1.3 Testarea semnificației

Listăm ingredientele pentru NHST.

1.3.1 Ingrediente

H_0 : **ipoteza 0**. Aceasta este presupunerea implicită pentru modelul care generează datele.

H_A : **ipoteza alternativă**. Dacă respingem ipoteza 0, acceptăm această alternativă ca cea mai bună explicație pentru date.

X : **statistica testului**. O calculăm din date.

Repartiția 0: repartiția de probabilitate a lui X presupunând H_0 .

Regiunea de respingere: dacă X este în regiunea de respingere, respingem H_0 în favoarea lui H_A .

Regiunea de nerespingere: complementara regiunii de respingere. Dacă X este în această regiune, nu respingem H_0 . Spunem "nu respingem" mai degrabă decât "acceptăm" deoarece cel mai bun lucru pe care-l putem spune este că datele nu susțin respingerea lui H_0 .

Ipoteza 0 H_0 și ipoteza alternativă H_A joacă roluri diferite. Respingem H_0 doar dacă avem destule dovezi împotriva ei.

1.4 Terminologia NHST

În această secțiune vom folosi un exemplu extins pentru a introduce și explora terminologia utilizată în testarea semnificației ipotezei 0 (NHST).

Exemplul 3. Pentru a testa dacă o monedă este corectă o aruncăm de 10 ori. Dacă obținem un număr neașteptat de mare sau de mic de aversuri, vom suspecta că moneda nu este corectă. Precizăm aceasta în limbajul NHST setând ingredientele. Fie θ probabilitatea aversului la aruncarea monedei.

1. Ipoteza 0: $H_0 =$ "moneda este corectă", i.e. $\theta = 0.5$.
2. Ipoteza alternativă: $H_A =$ "moneda nu este corectă", i.e. $\theta \neq 0.5$.

3. Statistica testului: $X = \text{numărul de aversuri în 10 aruncări.}$
4. Repartiția 0: Aceasta este funcția de probabilitate bazată pe ipoteza 0

$$p(x|\theta = 0.5) \sim \text{binomială}(10, 0.5).$$

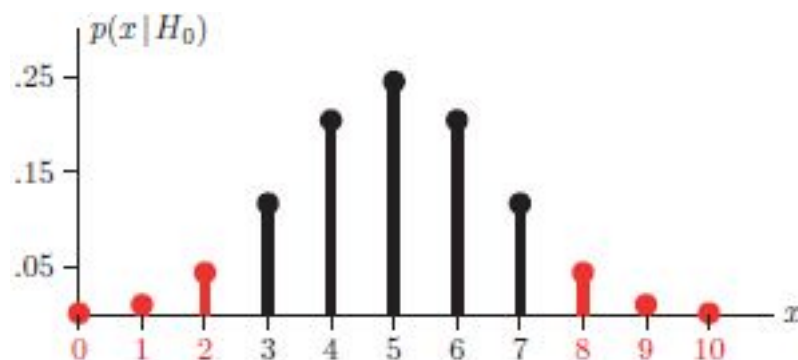
Iată tabelul de probabilitate pentru repartiția 0:

x	0	1	2	3	4	5	6	7	8	9	10
$p(x H_0)$.001	.010	.044	.117	.205	.246	.205	.117	.044	.010	.001

5. Regiunea de respingere: sub ipoteza 0 ne așteptăm să obținem circa 5 aversuri în 10 aruncări. Vom respinge H_0 dacă numărul de aversuri este mult mai mic sau mai mare ca 5. Fie regiunea de respingere $\{0, 1, 2, 8, 9, 10\}$. Adică, dacă numărul de aversuri din 10 aruncări este în această regiune, vom respinge ipoteza că moneda este corectă în favoarea ipotezei că nu este corectă. Putem rezuma toate acestea în graficul și tabelul de probabilitate de mai jos. Regiunea de respingere constă din valorile lui x în roșu. Probabilitățile corespunzătoare sunt pe fond roșu. Arătăm de asemenea repartiția 0 ca o reprezentare cu valorile lui x de respingere în roșu.

x	0	1	2	3	4	5	6	7	8	9	10
$p(x H_0)$.001	.010	.044	.117	.205	.246	.205	.117	.044	.010	.001

Regiunea de respingere și probabilitățile 0 ca un tabel pentru exemplul 3.



Regiunea de respingere și probabilitățile 0 ca o reprezentare pentru exemplul 3.

Observații pentru exemplul 3:

1. Ipoteza 0 este **implicit precaută**: nu vom pretinde că moneda este incorrectă dacă nu avem argumente convingătoare.
2. Regiunea de respingere constă din date care sunt **extreme sub ipoteza 0**. Adică, constă din rezultatele care sunt în coada repartiției 0, departe de centrul de probabilitate mare. Cât de departe, depinde de α , nivelul de semnificație al testului.

3. Dacă obținem 3 aversuri în 10 aruncări, atunci statistica testului este în regiunea de nerespingere. Limbajul științific uzual ar fi să spunem că datele "nu sprijină respingerea ipotezei 0". Chiar dacă avem 5 aversuri, **nu** vom pretinde că datele demonstrează că ipoteza 0 este adevărată.

Întrebare: Dacă avem o monedă corectă, care este probabilitatea că vom decide incorect că este incorectă?

Răspuns: Ipoteza 0 este că moneda este corectă. Se cere probabilitatea ca datele dintr-o monedă corectă să fie în regiunea de respingere. Adică, probabilitatea că vom obține 0, 1, 2, 8, 9 sau 10 aversuri în 10 aruncări. Aceasta este suma probabilităților în roșu. Adică,

$$P(\text{respingerea lui } H_0 | H_0 \text{ este adevărată}) = 0.11.$$

1.4.1 Ipoteze simple și compuse

Definiție: ipoteză simplă: O **ipoteză simplă** este una pentru care putem specifica complet repartiția ei. O ipoteză simplă tipică este că parametrul de interes ia o anumită valoare.

Definiție: Ipoteze compuse: Dacă repartiția nu poate fi specificată complet, spunem că ipoteza este **compusă**. O ipoteză compusă tipică este că parametrul de interes se află într-un domeniu de valori.

În exemplul 3, ipoteza 0 este că $\theta = 0.5$, deci repartiția 0 este binomială(10, 0.5).

Deoarece repartiția 0 este complet specificată, H_0 este simplă. Ipoteza alternativă este că $\theta \neq 0.5$. În realitate sunt multe ipoteze în una: θ ar putea fi 0.51, 0.7, 0.99, etc. Deoarece repartiția alternativă binomială(10, θ) nu este complet specificată, H_A este compusă.

Exemplul 4. Presupunem că avem datele x_1, \dots, x_n . Presupunem că ipotezele noastre sunt

H_0 : datele sunt extrase din $N(0, 1)$

H_A : datele sunt extrase din $N(1, 1)$.

Acestea sunt ambele **ipoteze simple** - fiecare ipoteză specifică complet o repartiție.

Exemplul 5. (Ipoteze compuse.) Acum presupunem că ipotezele noastre sunt

H_0 : datele sunt extrase dintr-o repartiție Poisson cu parametru necunoscut.

H_A : datele nu sunt extrase dintr-o repartiție Poisson.

Acestea sunt ambele ipoteze compuse, deoarece ele nu specifică complet repartiția.

Exemplul 6. Într-un experiment de percepție extrasenzorială (ESP), unui subiect i se cere să identifice culorile (pică, cupă, romb, treflă) a 100 de cărți trase (cu înlocuire) dintr-un pachet de cărți de joc. Fie T , numărul

succeselor. Ipoteza 0 (simplă) că subiectul nu are ESP este dată de

$$H_0 : T \sim \text{binomial}(100, 0.25).$$

Ipoteza alternativă (compusă) că subiectul are ESP este dată de

$$H_A : T \sim \text{binomial}(100, p) \text{ cu } p > 0.25.$$

O altă ipoteză alternativă (compusă) este că se întâmplă ceva în afara șansei pure, i.e. subiectul are ESP sau anti-ESP. Aceasta este dată de

$$H_A : T \sim \text{binomial}(100, p) \text{ cu } p \neq 0.25.$$

Valorile $p < 0.25$ reprezintă ipotezele că subiectul are un fel de anti-ESP.

1.4.2 Tipuri de eroare

Sunt 2 tipuri de erori pe care le putem face. Putem să respingem incorect ipoteza 0 când ea este adevărată sau putem să eșuăm incorect s-o respingem când este falsă. Acestea sunt numite **erori de tipul I** și **de tipul II**. Rezumăm aceasta în următorul tabel.

		True state of nature	
		H_0	H_A
Our decision	Reject H_0	Type I error	correct decision
	'Don't reject' H_0	correct decision	Type II error

Tipul I: respingere falsă a lui H_0 .

Tipul II: nerespingere falsă ("acceptare" falsă) a lui H_0 .

1.4.3 Nivel de semnificație și putere

Nivelul de semnificație și puterea sunt folosite pentru a cuantifica calitatea testului de semnificație. Ideal, un test de semnificație n-ar face erori. Adică, n-ar respinge H_0 când H_0 era adevărată și ar respinge H_0 în favoarea lui H_A când H_A era adevărată. Sunt cu totul 4 probabilități importante corespunzătoare tabelului 2×2 de mai sus:

$$\begin{array}{ll} P(\text{respingem } H_0 | H_0) & P(\text{respingem } H_0 | H_A) \\ P(\text{nu respingem } H_0 | H_0) & P(\text{nu respingem } H_0 | H_A). \end{array}$$

Cele 2 probabilități pe care ne concentrăm sunt:

$$\begin{aligned} \text{Nivelul de semnificație} &= P(\text{respingem } H_0 | H_0) \\ &= \text{probabilitatea să respingem incorect } H_0 \\ &= P(\text{eroarea de tipul I}). \end{aligned}$$

$$\begin{aligned}
\text{Puterea} &= \text{probabilitatea să respingem corect } H_0 \\
&= P(\text{respingem } H_0 | H_A) \\
&= 1 - P(\text{eroarea de tipul II}).
\end{aligned}$$

Ideal, un test al unei ipoteze ar trebui să aibă un nivel de semnificație mic (aproape de 0) și o putere mare (aproape de 1).

Analogii

1. Gândiți H_0 ca ipoteza "nu se întâmplă nimic demn de remarcat", i.e. "moneda este corectă", "tratamentul nu este mai bun ca placebo" etc. Și gândiți H_A ca opusa: "ceva interesant se întâmplă". Atunci puterea este probabilitatea de a detecta ceva interesant când este prezent și nivelul de semnificație este probabilitatea de a pretinde greșit că ceva interesant a apărut.

2. În SUA, inculpații sunt presupuși nevinovați până sunt demonstrați vinovați dincolo de un orice dubiu rezonabil. Putem formula aceasta în termeni de NHST ca

H_0 : acuzatul este nevinovat (implicit)

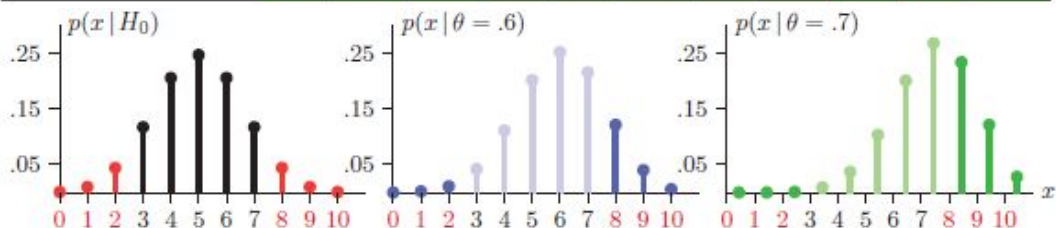
H_A : acuzatul este vinovat.

Nivelul de semnificație este probabilitatea de a găsi vinovată o persoană nevinovată. Puterea este probabilitatea de a găsi corect vinovată o parte vinovată. "Dincolo de orice dubiu rezonabil" înseamnă că ar trebui să cerem ca nivelul de semnificație să fie foarte mic.

Ipoteze compuse

H_A este compusă în exemplul 3, deci puterea este diferită pentru diferite valori ale lui θ . Extindem tabelul anterior de probabilități pentru a include unele valori alternative ale lui θ . Facem același lucru cu reprezentările. Ca întotdeauna în jocul NHST, ne uităm la **verosimilități**: probabilitatea datelor cunoscând ipoteza.

x	0	1	2	3	4	5	6	7	8	9	10
$H_0 : p(x \theta = 0.5)$.001	.010	.044	.117	.205	.246	.205	.117	.044	.010	.001
$H_A : p(x \theta = 0.6)$.000	.002	.011	.042	.111	.201	.251	.215	.121	.040	.006
$H_A : p(x \theta = 0.7)$.000	.0001	.001	.009	.037	.103	.200	.267	.233	.121	.028



Regiunea de respingere și probabilități 0 și alternative pentru exemplul 3. Folosim tabelul de probabilități pentru a calcula nivelul de semnificație și

puterea acestui test.

Nivelul de semnificație = probabilitatea să respingem H_0 când H_0 este adevărată

- = prob. că stat. testului e în regiunea de respingere când H_0 e adev.
- = prob. că stat. testului e în regiunea de respingere pe linia H_0 a tab.
- = suma valorilor căsuțelor roșii din linia $\theta = 0.5$
- = 0.11.

Puterea când $\theta = 0.6$ = probabilitatea să respingem H_0 când $\theta = 0.6$

- = prob. că stat. testului e în regiunea de respingere când $\theta = 0.6$
- = prob. că stat. test. e în regiunea de respingere pe linia $\theta = 0.6$ a tab.
- = suma valorilor căsuțelor albastru închis din linia $\theta = 0.6$
- = 0.18.

Puterea când $\theta = 0.7$ = probabilitatea să respingem H_0 când $\theta = 0.7$

- = prob. că stat. testului e în regiunea de respingere când $\theta = 0.7$
- = prob. că stat. test. e în regiunea de respingere pe linia $\theta = 0.7$ a tab.
- = suma valorilor căsuțelor verde închis din linia $\theta = 0.7$
- = 0.384.

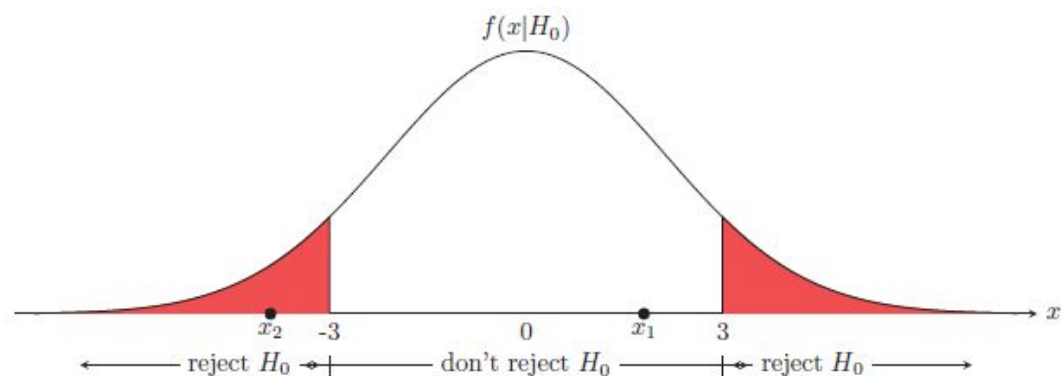
Vedem că puterea este mai mare pentru $\theta = 0.7$ decât pentru $\theta = 0.6$. Aceasta nu este surprinzător deoarece ne așteptăm să fie mai ușor să recunoaștem că o monedă cu $\theta = 0.7$ este incorectă decât este să recunoaștem că o monedă cu $\theta = 0.6$ este incorectă. Tipic, obținem putere mai mare când ipoteza alternativă este mai departe de ipoteza 0. În exemplul 3, ar fi foarte greu să distingem o monedă corectă de una cu $\theta = 0.51$.

1.4.4 Schițe conceptuale

Ilustrăm noțiunile de ipoteză 0, regiune de respingere și putere cu schițe ale pdf-urilor pentru ipotezele 0 și alternative.

Repartiția 0: regiuni de respingere și nerespingere

Prima diagramă de mai jos ilustrează o repartiție 0 cu regiunile de respingere și nerespingere. Sunt arătate de asemenea 2 posibile statistici ale testului: x_1 și x_2 ("reject" = "respingem", "don't reject" = "nu respingem").



Statistica x_1 a testului este în regiunea de nerespingere. Deci, dacă datele noastre au produs statistica testului x_1 , atunci nu vom respinge ipoteza 0 H_0 . Pe de altă parte, statistica testului x_2 este în regiunea de respingere, deci, dacă datele noastre au produs x_2 , atunci vom respinge ipoteza 0 în favoarea ipotezei alternative.

Sunt câteva lucruri de observat în această imagine.

1. Regiunea de respingere constă din valorile departe de centrul repartiției 0.
2. Regiunea de respingere este bilaterală.
3. Ipoteza alternativă nu este menționată. Respingem sau nu respingem H_0 bazați numai pe verosimilitatea $f(x|H_0)$. Ipoteza alternativă H_A ar trebui considerată când alegem o regiune de respingere, dar formal nu joacă niciun rol în respingerea sau nerespingerea lui H_0 .
4. Uneori numim regiunea de nerespingere **regiunea de acceptare**. Aceasta este tehnic incorect deoarece niciodată nu acceptăm cu adevărat ipoteza 0. Sau respingem, sau spunem că datele nu sprijină respingerea lui H_0 . Aceasta este adesea rezumat prin afirmația: **nu putem niciodată demonstra ipoteza 0**.

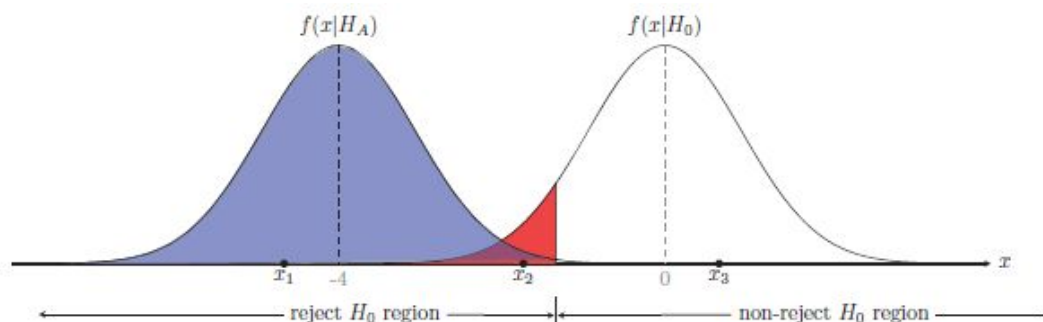
Teste de putere mare sau mică Următoarele 2 figuri arată teste de putere mare, respectiv mică.

Aria umbrită de sub $f(x|H_0)$ reprezintă nivelul de semnificație. Reamintim că nivelul de semnificație este:

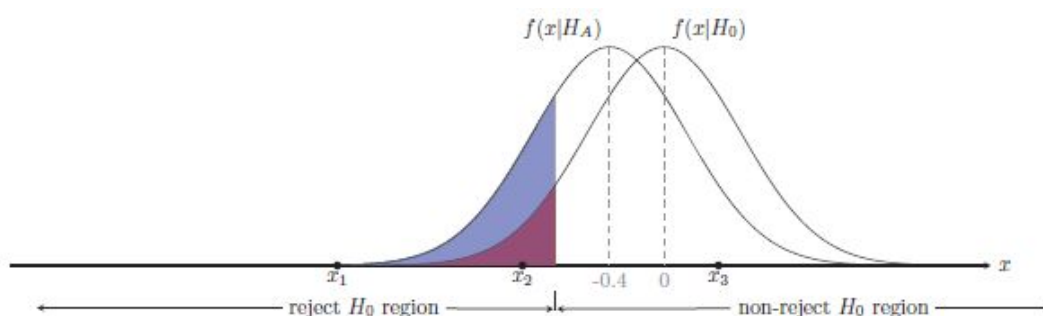
probabilitatea respingerii false a ipotezei 0 când ea este adevărată;

probabilitatea ca statistica testului să cadă în regiunea de respingere chiar dacă H_0 este adevărată.

Analog, regiunea umbrită de sub $f(x|H_A)$ reprezintă puterea, i.e. probabilitatea că statistica testului este în regiunea de respingere a lui H_0 când H_A este adevărată. Ambele teste au același nivel de semnificație, dar dacă $f(x|H_A)$ are o suprapunere considerabilă cu $f(x|H_0)$, puterea este mult mai mică.



Test de putere mare



Test de putere mică

În ambele teste, repartițiile 0 sunt normale standard. Repartiția 0, regiunea de respingere și nivelul de semnificație sunt aceleași. (Nivelul de semnificație este aria roșie/mov de sub $f(x|H_0)$ și de deasupra regiunii de respingere.) În figura de deasupra vedem că mediile celor 2 repartiții sunt la distanță de 4 deviații standard una față de cealaltă. Doarece ariile de sub densități au o foarte mică suprapunere, testul are putere mare. Adică, dacă datele x sunt extrase din H_A , vor fi aproape sigur în regiunea de respingere a lui H_0 . De exemplu x_3 ar fi un rezultat foarte surprinzător pentru repartiția H_A .

În figura de jos vedem că mediile sunt la doar 0.4 deviații standard una de alta. Deoarece ariile de sub densități au multă suprapunere, testul are putere mică. Adică, dacă datele x sunt extrase din H_A , este foarte probabil să se afle în regiunea de nerespingere a lui H_0 . De exemplu, x_3 n-ar fi un rezultat foarte surprinzător pentru repartiția H_A .

Tipic, putem crește puterea unui test crescând numărul de date și, prin aceasta, scăzând dispersia repartițiilor 0 și alternativă. În proiectarea experimentului este important să determinăm dinainte numărul de încercări sau de subiecți de care avem nevoie pentru a atinge o putere dorită.

Exemplul 7. Presupunem că un medicament pentru o boală este comparat cu un placebo. Alegem ipotezele 0 și alternativă astfel:

H_0 = medicamentul nu este mai bun ca placebo;

H_A = medicamentul este mai bun ca placebo.

Puterea testului ipotezei este probabilitatea ca testul să concluzioneze că medicamentul este mai bun, dacă este într-adevăr mai bun. Nivelul de semnificație este probabilitatea ca testul să concluzioneze că medicamentul este mai bun, când de fapt nu este mai bun.

1.5 Proiectarea unui test al ipotezei

Formal, tot ce necesită un test al ipotezei este H_0, H_A , o statistică a testului și o regiune de respingere. În practică proiectarea este adesea făcută folosind pașii următori:

1. Alege ipoteza 0 H_0 .

Adesea alegem H_0 să fie simplă. Sau, adesea alegem H_0 să fie cea mai simplă și cea mai prudentă explicație, i.e. medicamentul nu are efect, fără ESP, moneda este corectă.

2. Decide dacă H_A este unilaterală sau bilaterală.

În exemplul 3 am vrut să știm dacă moneda a fost incorectă. O monedă incorectă poate fi deplasată pentru sau împotriva aversului, deci $H_A : \theta \neq 0.5$ este o ipoteză bilaterală. Dacă ne pasă doar dacă moneda este deplasată sau nu pentru avers am putea folosi o ipoteză unilaterală $H_A : \theta > 0.5$.

3. Alege o statistică a testului.

De exemplu, media de selecție, totalul de selecție sau dispersia de selecție. Unele statistici standard sunt z, t și χ^2 . Repartițiile care merg cu aceste statistici sunt totdeauna condiționate de ipoteza 0, adică vom calcula verosimilități ca $f(z|H_0)$.

4. Alege un nivel de semnificație și determină regiunea de respingere.

Vom folosi de obicei α pentru a nota nivelul de semnificație. Paradigma Neyman-Pearson este de a alege α înainte. Valori tipice sunt 0.1, 0.05, 0.01. Reamintim că nivelul de semnificație este probabilitatea unei erori de tip I, i.e. respingerea incorectă a ipotezei 0 când ea este adevărată. Valoarea pe care o alegem va depinde de consecințele unei erori de tipul I. Odată ce nivelul de semnificație este ales, putem determina regiunea de respingere în coada (cozile) repartiției 0. În exemplul 3, H_A este bilaterală, deci regiunea de respingere este împărțită între cele 2 cozi ale repartiției 0. Această repartiție este dată în următorul tabel:

x	0	1	2	3	4	5	6	7	8	9	10
$p(x H_0)$.001	.010	.044	.117	.205	.246	.205	.117	.044	.010	.001

Dacă punem $\alpha = 0.05$, atunci regiunea de respingere trebuie să conțină cel

mult 0.05 probabilitate. Pentru o regiune bilaterală, obținem

$$\{0, 1, 9, 10\}.$$

Dacă punem $\alpha = 0.01$, regiunea de respingere este

$$\{0, 10\}.$$

Presupunem că schimbăm H_A în "moneda este deplasată în favoarea aver-sului". Acum avem o ipoteză unilaterală $\theta > 0.5$. Regiunea noastră de respingere va fi acum în coada dreaptă, deoarece nu vrem să respingem H_0 în favoarea lui H_A dacă obținem un număr mic de aversuri. Acum, dacă $\alpha = 0.05$, regiunea de respingere este domeniul unilateral

$$\{9, 10\}.$$

Dacă punem $\alpha = 0.01$, regiunea de respingere este

$$\{10\}.$$

5. Determină puterea (puterile)

După cum am văzut în exemplul 3, odată ce regiunea de respingere este stabilită, putem determina puterea testului la diverse valori ale ipotezei alternative.

Exemplul 8. (Consecințele semnificației.) Dacă $\alpha = 0.1$ așteptăm o rată a erorii de tipul I de 10%. Adică, așteptăm să respingem ipoteza 0 în 10% din acele experimente unde ipoteza H_0 este adevărată. Faptul dacă 0.1 este un nivel de semnificație rezonabil depinde de deciziile care vor fi făcute folosindu-l.

De exemplu, dacă faci un experiment pentru a determina dacă ciocolata ta are mai mult de 72% cacao, atunci o rată a erorii de tipul I de 10% este probabil în regulă. Adică, a considera fals că o ciocolată 72% are mai mult de 72%, este probabil acceptabil. Pe de altă parte, dacă laboratorul tău criminalistic identifică amprente pentru un proces de crimă, atunci o rată de 10% a erorii de tipul I, i.e. a pretinde greșit că amprente găsite la locul crimei au aparținut cuiva care în realitate este nevinovat, este categoric inacceptabilă.

Semnificația pentru o ipoteză 0 compusă. Dacă H_0 este compusă, atunci P (eroarea de tipul I) depinde de care membru al lui H_0 este adevărat. În acest caz nivelul de semnificație este definit ca maximum acestor probabilități.

1.6 Valori critice

Valorile critice sunt ca cuantilele cu excepția faptului că se referă la probabilitatea de la dreapta valorii, în loc de cea de la stânga.

Exemplul 9. Folosiți R pentru a afla valoarea critică 0.05 pentru repartiția normală standard.

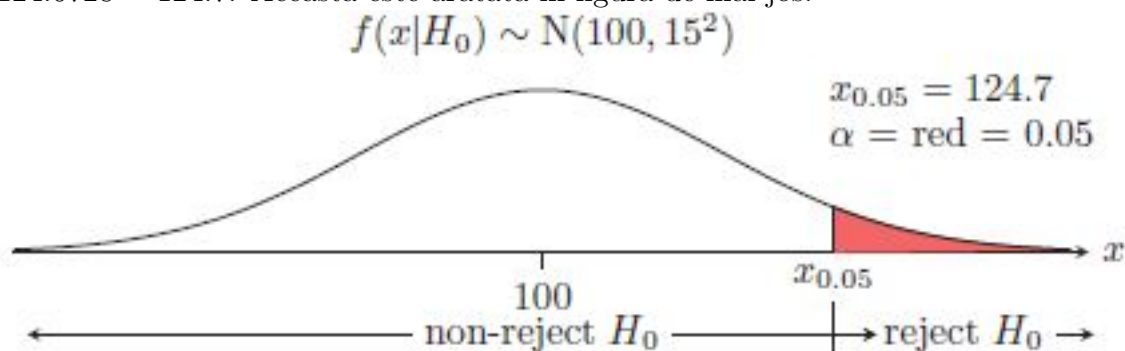
Răspuns. Notăm această valoare critică cu $z_{0.05}$. Valoarea critică $z_{0.05}$ este chiar 0.95 cuantila, i. e. are 5% probabilitate la dreapta ei și de aceea 95% probabilitate la stânga. O calculăm cu funcția R `qnorm(0.95,0,1)` sau `qnorm(0.95)` și obținem 1.644854.

Într-un test de semnificație tipic, regiunea de respingere constă din una sau ambele cozi ale repartiției 0. Valoarea care marchează începutul regiunii de respingere este o [valoare critică](#).

Exemplul 10. Valori critice și regiuni de respingere. Presupunem că statistica x a testului nostru are repartiția 0 $N(100, 15^2)$, i. e.

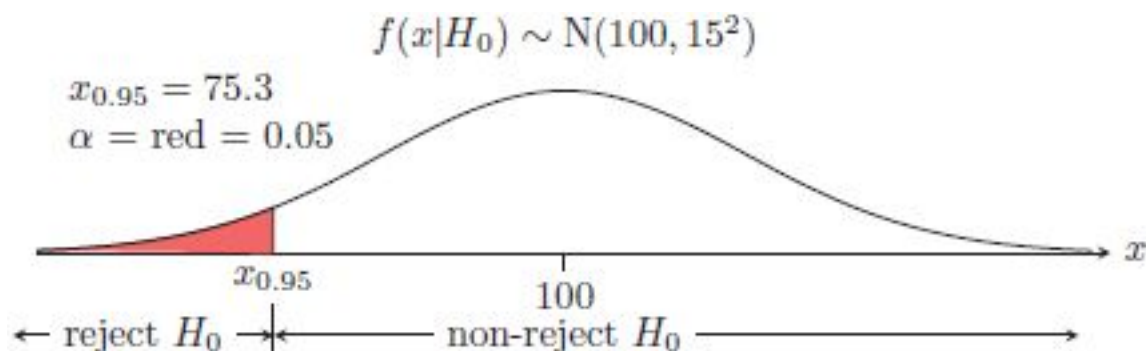
$f(x|H_0) \sim N(100, 15^2)$. Presupunem de asemenea că regiunea noastră de respingere este în dreapta și avem un nivel de semnificație de 0.05. Aflați valoarea critică și schițați repartiția 0 și regiunea de respingere.

Răspuns. Notăția folosită pentru valoarea critică cu coada dreaptă conținând 0.05 probabilitate este $x_{0.05}$. Valoarea critică $x_{0.05}$ este chiar 0.95 cuantila, i. e. are 5% probabilitate la dreapta ei și de aceea 95% probabilitate la stânga. O calculăm cu funcția R `qnorm(0.95,100,15)` și obținem $124.6728 \approx 124.7$. Aceasta este arătată în figura de mai jos.



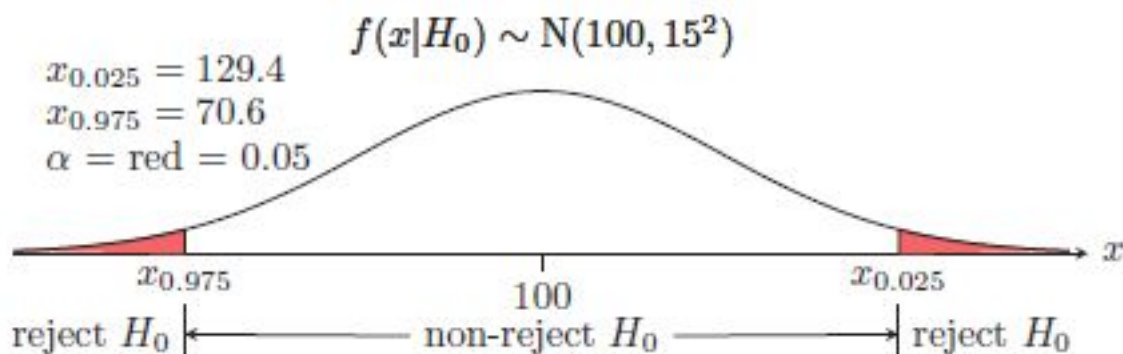
Exemplul 11. Valori critice și regiuni de respingere. Repetați exemplul precedent pentru o regiune de respingere în stânga cu nivelul de semnificație 0.05.

Răspuns. În acest caz valoarea critică are 0.05 probabilitate la stânga și de aceea 0.95 probabilitate la dreapta. Deci o notăm cu $x_{0.95}$. Deoarece este 0.05 cuantila, o calculăm în R cu `qnorm(0.05,100,15)` și obținem $75.3272 \approx 75.3$.



Exemplul 12. Valori critice. Repetați exemplul anterior pentru o regiune de respingere bilaterală. Puneți jumătate din semnificație în fiecare coadă.

Răspuns. Pentru a avea un total de semnificație de 0.05 punem 0.025 în fiecare coadă. Adică, coada stângă începe la $x_{0.975} = q_{0.025}$ și coada dreaptă începe la $x_{0.025} = q_{0.975}$. Calculăm aceste valori cu `qnorm(0.025, 100, 15)` și `qnorm(0.975, 100, 15)`. Valorile aproximative sunt arătate în figura de mai jos.



1.7 Valori p

În practică, oamenii adesea specifică nivelul de semnificație și fac testul de semnificație folosind **valori p** .

Dacă valoarea p este mai mică decât nivelul de semnificație α , atunci respingem H_0 . Altfel nu respingem H_0 .

Definiție. p -valoarea este probabilitatea, presupunând ipoteza 0, de a vedea datele **cel puțin la fel de extreme ca datele experimentale**. Ce înseamnă "cel puțin la fel de extreme" depinde de proiectarea experimentului.

Ilustrăm definiția și utilizarea valorilor p cu un exemplu unilateral. Acest exemplu introduce de asemenea **testul z** . Toate acestea înseamnă că statistica testului nostru este normală standard (sau aproximativ normală standard).

Exemplul 13. testul z pentru ipoteze normale

IQ-ul este repartizat normal în populație conform unei repartiții $N(100, 15^2)$. Presupunem că cei mai mulți dintre studenții FMI au IQ-ul peste medie, deci vom formula următoarele ipoteze:

H_0 = IQ-urile studenților FMI sunt repartizate identic cu cele ale populației generale
= IQ-urile studenților din FMI au o repartiție $N(100, 15^2)$.

H_A = IQ-urile studenților FMI tind să fie mai mari decât cele ale populației generale
= media IQ-urilor studenților din FMI este mai mare ca 100.

Observați că H_A este unilaterală.

Presupunem că testăm 9 studenți și aflăm că au un IQ mediu de $\bar{x} = 112$.

Putem respinge H_0 la nivelul de semnificație $\alpha = 0.05$?

Răspuns. Pentru a calcula p , întâi standardizăm datele: Sub ipoteza 0, $\bar{x} \sim N(100, 15^2/9)$ și de aceea

$$z = \frac{\bar{x} - 100}{15/\sqrt{9}} = \frac{36}{15} = 2.4 \sim N(0, 1).$$

Adică, repartiția 0 pentru z este normală standard. Numim z o [statistică \$z\$](#) , o vom utiliza ca statistica testului nostru.

Pentru o ipoteză alternativă la dreapta fraza "datele cel puțin la fel de extreme" este o coadă unilaterală la dreapta lui z . Valoarea p este atunci

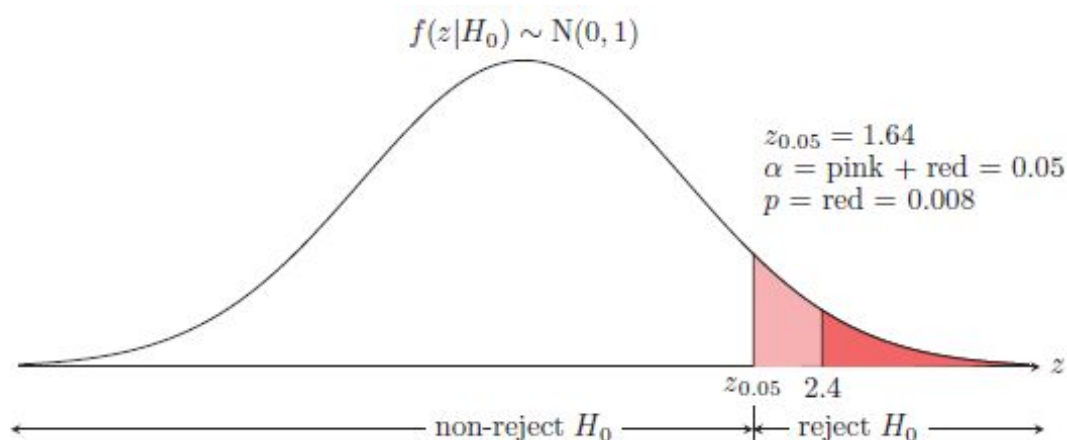
$$p = P(Z \geq 2.4) = 1 - \text{pnorm}(2.4, 0, 1) = 0.008197536.$$

Deoarece $p \leq \alpha$ respingem ipoteza 0. Formulăm concluzia noastră astfel: Respingem ipoteza 0 în favoarea ipotezei că studenții FMI au IQ-uri mai mari în medie. Am făcut aceasta la nivelul de semnificație 0.05 cu o valoare p de 0.008.

Observații: 1. Media $\bar{x} = 112$ este aleatoare: dacă facem experimentul din nou am putea obține o valoare diferită pentru \bar{x} .

2. Am fi putut folosi statistica \bar{x} direct. Standardizarea este preferabilă deoarece, cu practica, vom simți bine sensul diferitelor valori z .

Justificarea respingerii lui H_0 când $p \leq \alpha$ este dată în următoarea figură.



În acest exemplu $\alpha = 0.05$, $z_{0.05} = 1.64$ și regiunea de respingere este domeniul de la dreapta lui $z_{0.05}$. De asemenea, $z = 2.4$ și valoarea p este probabilitatea de la dreapta lui z . Figura ilustrează că $z = 2.4$ este în regiunea de respingere; aceasta este același lucru cu z este la dreapta lui $z_{0.05}$; aceasta este același lucru cu probabilitatea de la dreapta lui z este mai mică decât 0.05, ceea ce înseamnă $p < 0.05$.

1.8 Mai multe exemple

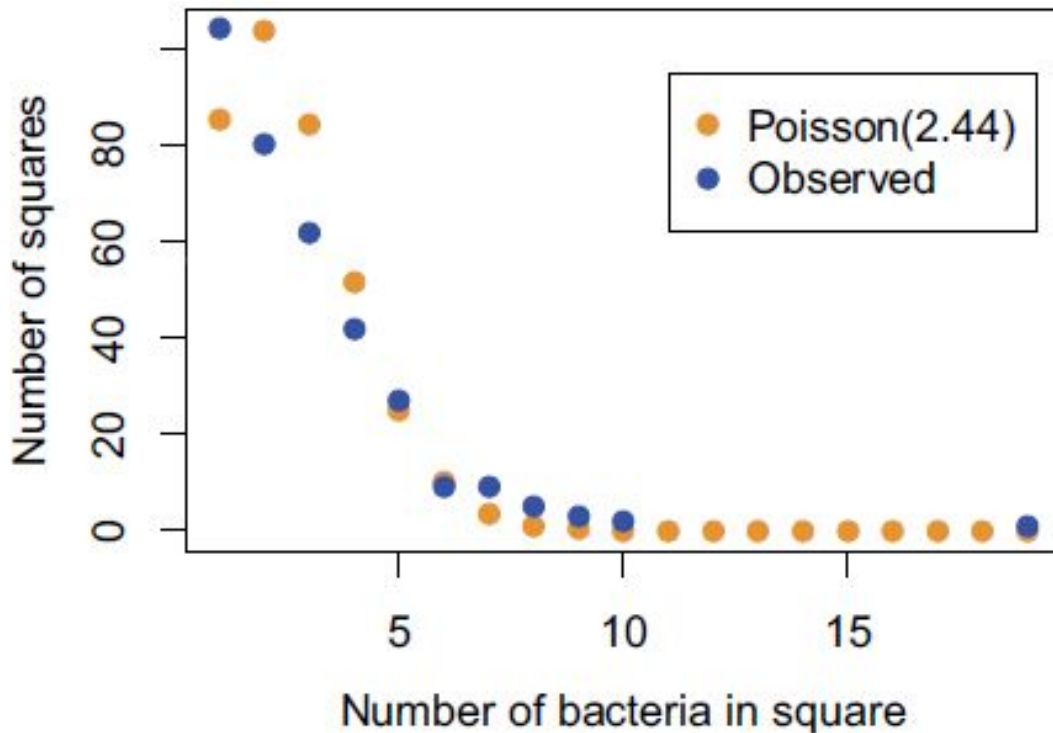
Testarea ipotezelor este larg utilizată în statistica deductivă.

Exemplul 14. Statistica χ^2 și bunătaea potrivirii.

Pentru a testa nivelul contaminării bacteriene, a fost vărsat lapte peste o rețea cu 400 de pătrate. Cantitatea de bacterii din fiecare pătrat a fost numărată. Rezumăm datele în tabelul de mai jos. Ultima linie a tabelului dă numărul de pătrate diferite care aveau o cantitate dată de bacterii.

Amount of bacteria	0	1	2	3	4	5	6	7	8	9	10	19
Number of squares	56	104	80	62	42	27	9	9	5	3	2	1

Cantitatea medie de bacterii pe pătrat este 2.44. Deoarece repartiția Poisson(λ) este folosită pentru a modela numărările evenimentelor relativ rare și parametrul λ este media repartiției, decidem să vedem dacă aceste date ar putea veni dintr-o repartiție Poisson. Pentru a face asta, întâi comparăm grafic frecvențele observate cu cele așteptate din Poisson(2.44).



Facem un test al ipotezei cu

H_0 : datele vin dintr-o repartiție Poisson(2.44).

H_A : datele vin dintr-o repartiție diferită.

Folosim o statistică χ^2 , numită așa deoarece ea are aproximativ o repartiție χ^2 . Pentru a calcula X^2 întâi combinăm ultimele câteva celule în tabel astfel încât numărul așteptat minim este în jur de 5 (o regulă a degetului mare generală în acest joc.)

Numărul așteptat de pătrate cu o anumită cantitate de bacterii vine din considerarea a 400 de încercări dintr-o repartiție Poisson(2.44), de exemplu, cu $l = 2.44$ numărul așteptat de pătrate cu 3 bacterii este $400 \cdot e^{-l} \frac{l^3}{3!} = 84.4$.

Statistica χ^2 este $\sum \frac{(O_i - E_i)^2}{E_i}$, unde O_i este numărul observat și E_i este numărul așteptat.

Number per square	0	1	2	3	4	5	6	> 6
Observed	56	104	80	62	42	27	9	20
Expected	34.9	85.1	103.8	84.4	51.5	25.1	10.2	5.0
Component of X^2	12.8	4.2	5.5	6.0	1.7	0.14	0.15	44.5

Adunând obținem $X^2 = 74.99$.

Doarece media (2.44) și numărul total de încercări (400) sunt fixate, cele 8 celule au doar 6 grade de libertate. Deci, presupunând H_0 , statistica noastră X^2 are (aproximativ) o repartiție χ^2_6 . Folosind această repartiție,

$P(X^2 > 74.99) = 0$. Astfel, respingem decisiv ipoteza 0 în favoarea ipotezei alternative că repartiția nu este Poisson(2.44).

Pentru a analiza mai departe, privim la componentele individuale ale lui X^2 . Sunt contribuții mari în coada distribuției, deci acolo potrivirea nu merge.

Exemplul 15. Testul t al lui Student.

Presupunem că vrem să comparăm un tratament medical pentru creșterea speranței de viață cu un placebo. Dăm la n oameni tratamentul și la m oameni placebo. Fie X_1, \dots, X_n numerele de ani pe care oamenii îi trăiesc după primirea tratamentului. Analog, fie Y_1, \dots, Y_m numerele de ani pe care oamenii îi trăiesc după primirea placebo. Fie \bar{X} și \bar{Y} mediile de selecție. Vrem să știm dacă diferența dintre \bar{X} și \bar{Y} este semnificativă statistic. Formulăm aceasta ca un test al ipotezei. Fie μ_X și μ_Y mediile (necunoscute).

$$H_0 : \mu_X = \mu_Y, \quad H_A : \mu_X \neq \mu_Y.$$

Cu anumite presupuneri și o formulă adecvată pentru eroarea standard unificată s_p statistica testului $t = \frac{\bar{X} - \bar{Y}}{s_p}$ are o repartiție t cu $n + m - 2$ grade de libertate. Deci regiunea noastră de respingere este determinată de un prag t_0 cu $P(t > t_0) = \alpha$.

2 Testarea semnificației ipotezei 0 II

2.1 Scopurile învățării

1. Să poată să listeze pașii comuni tuturor testelor semnificației ipotezei 0.
2. Să poată defini și calcula probabilitatea erorilor de tipul I și II.
3. Să poată să caute și să aplice t -teste pentru unul sau 2 eșantioane.

2.2 Introducere

Fiecare test face unele presupuneri despre date - adesea că sunt provenite dintr-o repartiție normală. Toate testele urmează același tipar. Doar calculul statisticii testului și tipul repartiției 0 se schimbă.

2.3 Configurarea și folosirea unui test de semnificație

Există o mulțime standard de pași care se fac pentru a configura și a utiliza un test al semnificației ipotezei 0.

1. Proiectează un experiment pentru a colecta datele și a alege o statistică x a testului pentru a fi calculată din date. Cerința cheie aici este să știm repartiția 0 $f(x|H_0)$. Pentru a calcula puterea, trebuie să cunoaștem și

repartiția alternativă $f(x|H_A)$.

2. Decide dacă testul este unilateral sau bilateral bazat pe H_A și forma repartiției 0.
3. Alege un nivel de semnificație α pentru respingerea ipotezei 0. Dacă se poate, calculează puterea corespunzătoare a testului.
4. Folosește experimentul pentru a colecta datele x_1, x_2, \dots, x_n .
5. Calculează statistica testului x .
6. Calculează valoarea p corespunzătoare lui x folosind repartiția 0.
7. Dacă $p < \alpha$, respinge ipoteza 0 în favoarea ipotezei alternative.

Observații.

1. În loc de a alege un nivel de semnificație, putem alege o regiune de respingere și să respingem H_0 dacă x este în această regiune. Nivelul de semnificație corespunzător este atunci probabilitatea ca x să fie în regiunea de respingere.
2. Ipoteza 0 este adesea numită "ipoteza precaută". Cu cât punem nivelul de semnificație mai mic, cu atât mai multe "dovezi" vor fi necesare înainte de a respinge ipoteza noastră precaută în favoarea unei alternative. Este o practică standard a publica valoarea p însăși, astfel încât alții să poată trage concluziile lor proprii.

3. **Un punct cheie de confuzie:** Un nivel de semnificație de 0.05 **nu** înseamnă că testul greșește doar în 5% din cazuri. Înseamnă că **dacă ipoteza 0 este adevărată**, atunci probabilitatea ca testul să o respingă din greșeală este 5%. Puterea testului măsoară acuratețea testului când ipoteza alternativă este adevărată. Anume, puterea testului este probabilitatea de a respinge ipoteza 0 **dacă ipoteza alternativă este adevărată**. De aceea probabilitatea de a eșua fals să respingem ipoteza 0 este 1 minus puterea.

Erori. Putem rezuma aceste 2 tipuri de erori și probabilitățile lor după cum urmează:

Eroarea de tipul I = respingerea lui H_0 când H_0 este adevărată.

Eroarea de tipul II = nerespingerea lui H_0 când H_A este adevărată.

$$\begin{aligned} P(\text{eroarea de tipul I}) &= \text{probabilitatea de a respinge fals } H_0 \\ &= P(\text{statistica testului este în regiunea de respingere} | H_0) \\ &= \text{nivelul de semnificație al testului} \end{aligned}$$

$$\begin{aligned} P(\text{eroarea de tipul II}) &= \text{probabilitatea de a nu respinge fals } H_0 \\ &= P(\text{statistica testului este în regiunea de acceptare} | H_A) \\ &= 1 - \text{putere.} \end{aligned}$$

Analogii utile. În termeni de testare medicală pentru o boală, o eroare de tipul I este un rezultat fals pozitiv și o eroare de tipul II este un rezultat fals negativ. În termenii unui proces juridic, o eroare de tipul I este condamnarea

unui nevinovat și eroarea de tipul II este achitarea unui vinovat.

2.4 Înțelegerea unui test de semnificație

Întrebări de pus:

1. Cum au colectat datele? Care este schema experimentului?
2. Care sunt ipotezele 0 și alternativă?
3. Ce tip de test de semnificație a fost folosit?

Se potrivesc datele lor cu criteriile necesare pentru a utiliza acest tip de test?

Cât de robust este testul la deviații de la aceste criterii?

4. De exemplu, unele teste compară 2 grupe de date presupunând că grupele sunt provenite din repartiții care au aceeași dispersie. Acest fapt trebuie verificat înainte de aplicarea testului. Adesea verificarea este făcută folosind alt test de semnificație proiectat pentru a compara dispersiile a 2 grupuri de date.

5. Cum este calculată valoarea p ?

Un test de semnificație vine cu o statistică a testului și o repartiție 0. În cele mai multe teste valoarea p este

$$p = P(\text{datele cel puțin la fel de extreme ca ce am obținut} | H_0)$$

Ce înseamnă "datele cel puțin la fel de extreme ca datele pe care le-am văzut"? I.e., este testul unilateral sau bilateral?

6. Care este nivelul de semnificație pentru acest test? Dacă $p < \alpha$, atunci experimentatorul va respinge H_0 în favoarea H_A .

2.5 Teste t

Multe teste de semnificație presupun că datele provin dintr-o repartiție normală, deci înainte de a folosi un astfel de test trebuie să examinăm datele pentru a vedea dacă ipoteza normalității este rezonabilă. Reprezentarea unei histogramme este un start bun. Ca și testul z , testele t cu o selecție și cu 2 selecții pe care le considerăm mai jos pleacă de la această presupunere de normalitate.

2.5.1 Test z

Recapitulăm testul z .

Date: presupunem $x_1, x_2, \dots, x_n \sim N(\mu, \sigma^2)$, unde μ este necunoscută și σ este cunoscută.

Ipoteza 0: $\mu = \mu_0$ pentru o anumită valoare specifică μ_0 .

Statistica testului: $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ = media standardizată.

Repartiția 0: $f(z|H_0)$ este pdf a $Z \sim N(0, 1)$.

Valoare p unilaterală (partea dreaptă): $p = P(Z > z|H_0)$.

Valoare p unilaterală (partea stângă): $p = P(Z < z|H_0)$.

Valoare p bilaterală: $p = P(|Z| > |z||H_0)$.

Exemplul 1. Presupunem că avem datele care au o repartiție normală de medie necunoscută μ și dispersie cunoscută 4. Fie ipoteza 0: $\mu = 0$. Fie ipoteza alternativă $H_A : \mu > 0$. Presupunem că obținem următoarele date:

$$1, 2, 3, 6, -1.$$

La un nivel de semnificație de $\alpha = 0.05$ ar trebui să respingem ipoteza 0?

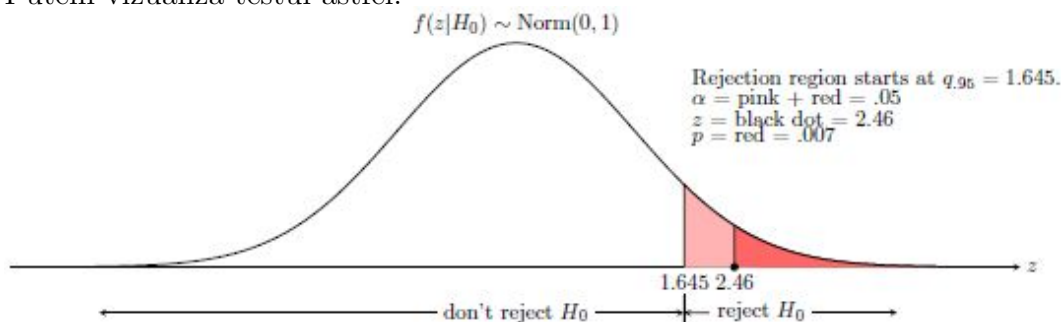
Răspuns. Sunt 5 date cu media $\bar{x} = 2.2$. Deoarece avem date normale cu o dispersie cunoscută ar trebui să folosim un z test. Statistica noastră z este

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{2.2 - 0}{2/\sqrt{5}} = 2.46.$$

Testul nostru este unilateral deoarece ipoteza alternativă este unilaterală. Deci (folosind R) valoarea p a noastră este

$$p = P(Z > z) = P(Z > 2.46) = 1 - \text{pnorm}(2.46) \approx 0.007.$$

Deoarece $p < 0.05$, respingem ipoteza 0 în favoarea ipotezei alternative $\mu > 0$. Putem vizualiza testul astfel:



2.5.2 Repartiția t a lui Student

"Student" este pseudonimul utilizat de William Gosset care a descris primul acest test și această repartiție. Vezi http://en.wikipedia.org/wiki/Student's_t-test.

Repartiția t este simetrică și are formă de clopot ca repartiția normală. Are un parametru df care stă pentru grade de libertate. Pentru df mic repartiția t are mai multă probabilitate în cozile ei ca repartiția normală standard. Când df crește, $t(df)$ devine din ce în ce mai asemănătoare cu repartiția normală

standard.

Iată o aplicație care arată $t(df)$ și o compară cu repartiția normală standard: <http://mathlets.org/mathlets/t-distribution>.

Ca de obicei în R, funcțiile `pt`, `dt`, `qt`, `rt` corespund la cdf, pdf, cuantile, respectiv generarea unui eșantion aleator dintr-o repartiție t . Reamintim că puteți tasta `?dt` în RStudio pentru a vedea fișierul de ajutor care specifică parametrii lui `dt`. De exemplu, `pt(1.65, 3)` calculează probabilitatea ca x să fie mai mic sau egal cu 1.65 dat fiind că x provine dintr-o repartiție t cu 3 grade de libertate, i.e. $P(x \leq 1.65)$ dat fiind că $x \sim t(3)$.

2.5.3 Testul t cu o selecție

Pentru testul z , am presupus că dispersia repartiției datelor este cunoscută. Totuși, adesea nu știm σ și trebuie s-o estimăm din date. În aceste cazuri, folosim un test t cu o selecție în locul unui test z și media **Studentizată** în locul mediei standardizate.

Date: presupunem $x_1, x_2, \dots, x_n \sim N(\mu, \sigma^2)$, unde atât μ cât și σ sunt necunoscute.

Ipoteza 0: $\mu = \mu_0$ pentru o anumită valoare specifică μ_0 .

Statistica testului:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}},$$

unde

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Aici t este numită **media Studentizată** și s^2 este numită **dispersia de selecție**. Ultima este o estimare a adevăratei dispersii σ^2 .

Repartiția 0: $f(t|H_0)$ este pdf pentru $T \sim t(n-1)$, repartiția t cu $n-1$ grade de libertate.*

Valoare p unilaterală (partea dreaptă): $p = P(T > t|H_0)$.

Valoare p unilaterală (partea stângă): $p = P(T < t|H_0)$.

Valoare p bilaterală: $p = P(|T| > |t|)$.

*Există o teoremă (nu o presupunere) că dacă datele sunt normale cu media μ_0 , atunci media Studentizată are o repartiție t . Puteți vedea demonstrația în http://en.wikipedia.org/wiki/Student's_t-distribution#Derivation.

Exemplul 2. Presupunem acum că în exemplul precedent dispersia este necunoscută. Adică, avem date care au o repartiție normală cu medie necunoscută μ și dispersie necunoscută σ^2 . Presupunem că obținem aceleași date:

$$1, 2, 3, 6, -1.$$

Ca mai sus, fie $H_0 : \mu = 0$ și $H_A : \mu > 0$. La nivelul de semnificație de $\alpha = 0.05$ ar trebui să respingem ipoteza 0?

Răspuns. Sunt 5 date cu media $\bar{x} = 2.2$. Deoarece avem date normale cu medie și dispersie necunoscute ar trebui să folosim un t test pentru o selecție. Calculând dispersia de selecție obținem

$$s^2 = \frac{1}{4}((1 - 2.2)^2 + (2 - 2.2)^2 + (3 - 2.2)^2 + (6 - 2.2)^2 + (-1 - 2.2)^2) = 6.7.$$

(Puteam folosi și comenzile în R

`x=c(1,2,3,6,-1)`

`var(x).`)

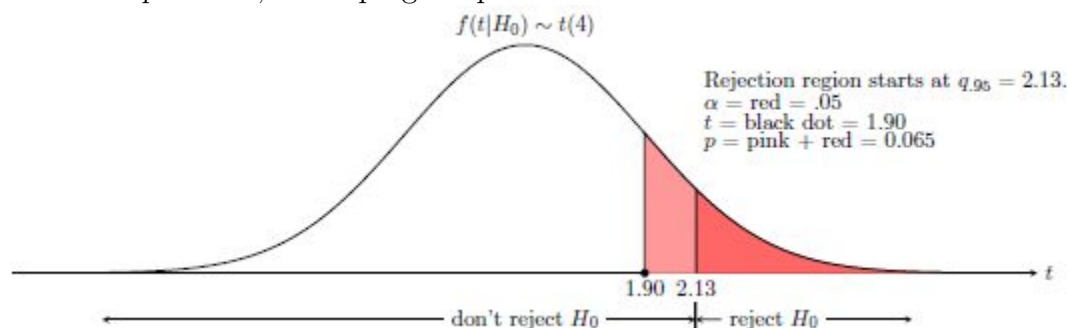
Statistica noastră t este

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{2.2 - 0}{\sqrt{6.7}/\sqrt{5}} = 1.900511.$$

Testul nostru este unilateral deoarece ipoteza alternativă este unilaterală. Deci (folosind R), valoarea p este

$$p = P(T > t) = P(T > 1.900511) = 1 - \text{pt}(1.900511, 4) = 0.06508106.$$

Deoarece $p > 0.05$, nu respingem ipoteza 0. Putem vizualiza testul astfel:



2.5.4 Testul t pentru 2 selecții cu dispersii egale

Considerăm în continuare cazul comparării mediilor a 2 selecții. De exemplu, putem fi interesați în compararea eficiențelor medii ale 2 tratamente medicale.

Date: Presupunem că avem 2 mulțimi de date provenite din repartiții normale

$$x_1, x_2, \dots, x_n \sim N(\mu_1, \sigma^2)$$

$$y_1, y_2, \dots, y_m \sim N(\mu_2, \sigma^2),$$

unde mediile μ_1 și μ_2 și dispersia σ^2 sunt toate necunoscute. Presupunem că cele 2 repartiții au **aceeași dispersie**. Sunt n date în primul grup și m date în al 2-lea.

Ipoteza 0: $\mu_1 = \mu_2$ (valorile lui μ_1 și μ_2 nu sunt specificate).

Statistica testului:

$$t = \frac{\bar{x} - \bar{y}}{s_p},$$

unde s_p^2 este **dispersia combinată**

$$s_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2} \left(\frac{1}{n} + \frac{1}{m} \right).$$

Aici s_x^2 și s_y^2 sunt dispersiile de selecție ale x_i -urilor și respectiv y_j -urilor.

Repartiția 0: $f(t|H_0)$ este pdf pentru $T \sim t(n+m-2)$.

Valoare p unilaterală (partea dreaptă): $p = P(T > t|H_0)$.

Valoare p unilaterală (partea stângă): $p = P(T < t|H_0)$.

Valoare p bilaterală: $p = P(|T| > |t|)$.

Observația 1. Unii autori folosesc o notație diferită. Ei definesc dispersia combinată ca

$$s_{p\text{-alți-autori}}^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}$$

și ce am numit dispersia combinată ei arată că este dispersia estimată a lui $\bar{x} - \bar{y}$. Adică,

$$s_p^2 = s_{p\text{-alți-autori}}^2 \cdot \left(\frac{1}{n} + \frac{1}{m} \right) \approx s_{\bar{x}-\bar{y}}^2.$$

Observația 2. Există o versiune a t -testului cu 2 selecții care permite celor 2 grupe să aibă dispersii diferite. În acest caz statistica testului este mai complicată, dar R o tratează la fel de ușor.

Exemplul 3. Următoarele date vin dintr-un studiu real în care 1408 femei au fost internate la o maternitate pentru (i) motive medicale sau (ii) prin internare de urgență. Durata sarcinii este măsurată în săptămâni complete de la începutul ultimului ciclu. Putem rezuma datele astfel:

Medicale: 775 de observații cu $\bar{x}_M = 39.08$ și $s_M^2 = 7.77$.

Urgențe: 633 de observații cu $\bar{x}_E = 39.6$ și $s_E^2 = 4.95$.

Configurați și folosiți un test t cu 2 selecții pentru a investiga dacă durata medie diferă pentru cele 2 grupuri.

Ce presupuneri ați făcut?

Răspuns. Dispersia combinată pentru aceste date este

$$s_p^2 = \frac{774 \cdot 7.77 + 632 \cdot 4.95}{1406} \left(\frac{1}{775} + \frac{1}{633} \right) = 0.01866256.$$

Statistica t pentru repartiția 0 este

$$\frac{\bar{x}_M - \bar{x}_E}{s_P} = -3.806429.$$

Avem 1406 grade de libertate. Folosind R pentru a calcula valoarea p bilaterală, obținem

$$p = P(|T| > |t|) = 2 * \text{pt}(-3.806429, 1406) = 0.000147048.$$

p este foarte mică, mult mai mică decât $\alpha = 0.05$ sau $\alpha = 0.01$. De aceea respingem ipoteza 0 în favoarea alternativei că există o diferență între ratele medii.

În loc să calculăm valoarea p bilaterală exact folosind o repartiție t am fi putut observa că, cu 1406 de grade de libertate, repartiția t este în esență normală standard și 3.806429 este aproape 4 deviații standard. Deci

$$p = P(|T| \geq 3.806429) \approx P(|Z| \geq 3.806429) < 0.001.$$

Am presupus că datele au fost normale și că cele 2 grupe au avut dispersii egale. Dată fiind diferența mare dintre dispersiile de selecție, această presupunere nu poate fi garantată.

De fapt, sunt alte teste de semnificație care testează dacă datele sunt aproximativ normale și dacă cele 2 grupe au aceeași dispersie. În practică se pot aplica acestea întâi pentru a determina dacă un test t este potrivit în primul rând.

http://en.wikipedia.org/wiki/Normality_test

http://en.wikipedia.org/wiki/F-test_of_equality_of_variances

3 Testarea semnificației ipotezei 0 III

3.1 Scopurile învățării

1. Fiind date ipotezele și datele, să poată identifica un test de semnificație potrivit dintr-o listă de teste uzuale.
2. Fiind date ipotezele, datele și un test de semnificație sugerat, să știe cum să caute detaliile și să aplice testul.

3.2 Introducere

Toate testele de semnificație au același model de bază în proiectarea și implementarea lor.

Proiectarea unui test al semnificației ipotezei 0 (NHST):

Specificăm ipoteza 0 și ipotezele alternative.

Alegem o statistică a testului ale cărei repartiție 0 și repartiție alternativă (repartiții alternative) sunt cunoscute.

Specificăm o regiune de respingere. Cel mai adesea acest lucru este făcut implicit specificând un nivel de semnificație α și o metodă pentru calculul p -valorilor bazată pe cozile repartiției 0.

Calculăm puterea folosind repartiția alternativă (repartițiile alternative).

Folosirea unui NHST:

Colectăm datele și calculăm statistica testului.

Verificăm dacă statistica testului este în regiunea de respingere. Cel mai adesea acest lucru este făcut implicit verificând dacă $p < \alpha$. Dacă este așa, ”respingem ipoteza 0 în favoarea ipotezei alternative”. Altfel concluzionăm ”datele nu sprijină respingerea ipotezei 0”.

Observați formularea prudentă: când eșuăm în a respinge H_0 , nu concluzionăm că H_0 este adevărată. Eșecul respingerii poate avea alte cauze. De exemplu, poate nu avem destule date pentru a distinge clar H_0 și H_A , în timp ce mai multe date ar indica respingerea lui H_0 .

3.3 Parametrii populației și statistici de selecție

Exemplul 1. Dacă alegem aleator 10 bărbați dintr-o populație și le măsurăm înălțimile, spunem că **am selectat înălțimile** din populație. În acest caz **media de selecție**, să spunem \bar{x} , este media înălțimilor selectate. Este o statistică și îi știm explicit valoarea. Pe de altă parte, adevărata înălțime medie a populației, să spunem μ , este necunoscută și putem doar să estimăm valoarea ei. Numim μ **parametru al populației**.

Principalul scop al testării semnificației este folosirea statisticilor de selecție pentru a trage concluzii despre parametrii de selecție. De exemplu, putem testa dacă înălțimea medie a bărbaților dintr-o populație dată este mai mare ca 1.78 m.

3.4 O galerie a testelor de semnificație uzuale legate de repartiția normală

Toate aceste teste presupun **date normale**.

Repartițiile 0 pentru aceste teste sunt **toate legate de repartiția normală** prin formule explicite. Argumentele sunt accesibile oricui știe analiză matematică și este interesat de înțelegerea lor. Dat fiind numele oricărei repartiții, putem căuta online detaliile construcției și proprietăților ei. De asemenea putem folosi R pentru a explora repartiția numeric și grafic.

Când analizăm datele cu oricare din aceste teste, un lucru de importanță cheie este să verificăm că presupunerile sunt adevărate sau cel puțin aproximativ adevărate. De exemplu, nu ar trebui să folosim un test care presupune că datele sunt normale până nu am verificat că datele sunt aproximativ normale.

3.4.1 Testul z

Utilizare: Testăm dacă media populației este egală cu o medie ipotetică.

Date: x_1, x_2, \dots, x_n .

Presupuneri: Datele sunt selecții normale independente:

$$x_i \sim N(\mu, \sigma^2) \text{ unde } \mu \text{ este necunoscută, dar } \sigma \text{ este cunoscută.}$$

H_0 : Pentru un μ_0 specificat, $\mu = \mu_0$.

H_A :

Bilaterală: $\mu \neq \mu_0$;

unilaterală mai mare: $\mu > \mu_0$;

unilaterală mai mică: $\mu < \mu_0$.

Statistica testului: $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$.

Repartiția 0: $f(z|H_0)$ este pdf a lui $Z \sim N(0, 1)$.

valoarea p :

Bilaterală: $p = P(|Z| > |z|) = 2 * (1 - \text{pnorm}(\text{abs}(z), 0, 1));$

unilaterală mai mare: $p = P(Z > z) = 1 - \text{pnorm}(z, 0, 1);$

unilaterală mai mică: $p = P(Z < z) = \text{pnorm}(z, 0, 1).$

Codul R: Nu pare a fi o singură funcție R pentru a folosi un test z . Desigur, se poate folosi R pentru a calcula z și valoarea p .

Exemplul 2. Vezi exemplul 13 din prima parte a acestui curs.

3.4.2 Testul t de o selecție pentru medie

Utilizare: Testăm dacă media populației este egală cu o medie ipotetică.

Date: x_1, x_2, \dots, x_n .

Presupuneri: Datele sunt selecții normale independente:

$$x_i \sim N(\mu, \sigma^2) \text{ unde atât } \mu \text{ cât și } \sigma \text{ sunt necunoscute.}$$

H_0 : Pentru un μ_0 specificat, $\mu = \mu_0$.

H_A :

Bilaterală: $\mu \neq \mu_0$;
 unilaterială mai mare: $\mu > \mu_0$;
 unilaterială mai mică: $\mu < \mu_0$.

Statistica testului: $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$,

unde s^2 este dispersia de selecție: $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$.

Repartiția 0: $f(t|H_0)$ este pdf a lui $T \sim t(n-1)$ (repartiția t a lui Student cu $n-1$ grade de libertate).

Valoarea p :

Bilaterală: $p = P(|T| > |t|) = 2 * (1 - \text{pt}(\text{abs}(t), n-1));$
 unilaterială mai mare: $p = P(T > t) = 1 - \text{pt}(t, n-1);$
 unilaterială mai mică: $p = P(T < t) = \text{pt}(t, n-1).$

Exemplu de cod R: Pentru datele $x = 1, 3, 5, 7, 2$ putem face un test t de o selecție cu $H_0: \mu = 2.5$ folosind comenzile în R:

```
x=c(1,3,5,7,2)
```

```
t.test(x, mu=2.5)
```

Acestea vor da câteva informații, incluzând media datelor, valoarea t și valoarea p bilaterală. Vezi ajutorul pentru această funcție pentru alte setări ale argumentelor.

Exemplul 3. Vezi exemplul 2 din partea a 2-a a acestui curs.

3.4.3 Testul t cu 2 selecții pentru compararea mediilor

Cazul dispersiilor egale

Începem descriind testul [presupunând dispersii egale](#).

Utilizare: Testăm dacă mediile populației din 2 populații diferă printr-o cantitate ipotetică.

Date: x_1, x_2, \dots, x_n și y_1, y_2, \dots, y_m .

Presupuneri: Ambele grupe de date sunt selecții normale independente:

$$x_i \sim N(\mu_x, \sigma^2),$$

$$y_j \sim N(\mu_y, \sigma^2),$$

unde atât μ_x cât și μ_y sunt necunoscute și posibil diferite. Dispersia σ^2 este necunoscută, dar aceeași pentru ambele grupe.

H_0 : Pentru un μ_0 specificat: $\mu_x - \mu_y = \mu_0$.

H_A :

Bilaterală: $\mu_x - \mu_y \neq \mu_0$;
 unilaterială mai mare: $\mu_x - \mu_y > \mu_0$;
 unilaterială mai mică: $\mu_x - \mu_y < \mu_0$.

Statistica testului: $t = \frac{\bar{x} - \bar{y} - \mu_0}{s_p}$, unde s_x^2 și s_y^2 sunt dispersiile de selecție ale datelor x , respectiv y și s_p^2 este (uneori numită) dispersia de selecție combinată:

$$s_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2} \left(\frac{1}{n} + \frac{1}{m} \right) \text{ și } df = n+m-2.$$

Repartiția 0: $f(t|H_0)$ este pdf a lui $T \sim t(df)$, repartiția t cu $df = n+m-2$ grade de libertate.

Valoarea p :

$$\begin{aligned} \text{Bilaterală:} \quad p &= P(|T| > |t|) = 2 * (1 - \text{pt}(\text{abs}(t), df)); \\ \text{unilaterală mai mare:} \quad p &= P(T > t) = 1 - \text{pt}(t, df); \\ \text{unilaterală mai mică:} \quad p &= P(T < t) = \text{pt}(t, df). \end{aligned}$$

Codul R: Funcția din R `t.test` va rula un test t cu 2 selecții cu dispersii egale setând argumentul `var.equal=TRUE`.

Exemplul 4. Vezi exemplul 3 din a 2-a parte a acestui curs.

Observații. 1. Cel mai adesea testul este făcut cu $\mu_0 = 0$. Adică, ipoteza 0 este că **mediile sunt egale**, i.e. $\mu_x - \mu_y = 0$.

2. Dacă datele x și y au aceeași lungime, n , atunci formula pentru s_p^2 devine mai simplă:

$$s_p^2 = \frac{s_x^2 + s_y^2}{n}.$$

Cazul dispersiilor inegale

Există o formă a testului t pentru cazul **când dispersiile nu sunt presupuse egale**. Uneori este numit **testul t al lui Welch**.

Acesta arată exact la fel cu excepția unor mici schimbări în presupuneri și în formula dispersiei combinate:

Utilizare: Testăm dacă mediile populației din 2 populații diferă printr-o cantitate ipotetică.

Date: x_1, x_2, \dots, x_n și y_1, y_2, \dots, y_m .

Presupuneri: Ambele grupe de date sunt selecții normale independente:

$$\begin{aligned} x_i &\sim N(\mu_x, \sigma_x^2), \\ y_j &\sim N(\mu_y, \sigma_y^2), \end{aligned}$$

unde atât μ_x cât și μ_y sunt necunoscute și posibil diferite. Dispersiile σ_x^2 și σ_y^2 sunt necunoscute și **nu sunt presupuse a fi egale**.

H_0, H_A : Exact aceleași ca în cazul dispersiilor egale.

Statistica testului: $t = \frac{\bar{x} - \bar{y} - \mu_0}{s_p}$, unde s_x^2 și s_y^2 sunt dispersiile de selecție

ale datelor x , respectiv y și s_p^2 este (uneori numită) dispersia de selecție combinată:

$$s_p^2 = \frac{s_x^2}{n} + \frac{s_y^2}{m} \text{ și } df = \frac{(s_x^2/n + s_y^2/m)^2}{(s_x^2/n)^2/(n-1) + (s_y^2/m)^2/(m-1)}.$$

Repartiția 0: $f(t|H_0)$ este pdf a lui $T \sim t(df)$, repartiția t cu df grade de libertate.

Valoarea p : Exact aceeași ca în cazul dispersiilor egale.

Codul R: Funcția din R `t.test` va rula un test t în acest caz setând argumentul `var.equal=FALSE`.

Testul t cu 2 selecții în perechi

Când datele vin firesc în perechi (x_i, y_i) , putem folosi [testul \$t\$ cu 2 selecții în perechi](#) (după de verificăm că presupunerile sunt valide!)

Exemplul 5. Pentru a măsura eficiența unei medicații de scăderea colesterolului, putem testa fiecare subiect înainte și după tratamentul cu medicamentul. Deci pentru fiecare subiect avem o pereche de măsurători: x_i = nivelul de colesterol înainte de tratament și y_i = nivelul de colesterol după tratament.

Exemplul 6. Pentru a măsura eficiența unui tratament pentru cancer putem pune în pereche fiecare subiect care a primit tratamentul cu unul care nu l-a primit. În acest caz am vrea să punem în perechi subiecții care sunt similari în termeni de stadiu al bolii, vârstă, sex, etc.

Utilizare: Testăm dacă diferența medie dintre valorile perechilor dintr-o populație este egală cu o valoare ipotetică.

Date: x_1, x_2, \dots, x_n și y_1, y_2, \dots, y_n trebuie să aibă aceeași lungime.

Presupuneri: [Diferențele](#) $w_i = x_i - y_i$ sunt independente dintr-o repartiție normală $N(\mu, \sigma^2)$, unde μ și σ sunt necunoscute.

OBSERVAȚIE: Acesta este chiar un test t de o selecție folosind w_i .

H_0 : Pentru un μ_0 specificat, $\mu = \mu_0$.

H_A :

Bilaterală: $\mu \neq \mu_0$;
unilaterală mai mare: $\mu > \mu_0$;
unilaterală mai mică: $\mu < \mu_0$.

Statistica testului: $t = \frac{\bar{w} - \mu_0}{s/\sqrt{n}}$,

unde s^2 este dispersia de selecție: $s^2 = \frac{1}{n-1} \sum_{i=1}^n (w_i - \bar{w})^2$.

Repartiția 0: $f(t|H_0)$ este pdf a lui $T \sim t(n-1)$ (repartiția t a lui Student cu $n-1$ grade de libertate).

Valoarea p :

Bilaterală: $p = P(|T| > |t|) = 2 * (1 - \text{pt}(\text{abs}(t), n - 1));$
 unilaterală mai mare: $p = P(T > t) = 1 - \text{pt}(t, n - 1);$
 unilaterală mai mică: $p = P(T < t) = \text{pt}(t, n - 1).$

Cod R: Funcția `t.test` din R va face un test cu 2 selecții în perechi dacă punem argumentul `paired=TRUE`. Putem de asemenea să facem un test t de o selecție pentru $x - y$.

Exemplul 7. Acest exemplu este din John Rice, *Mathematical Statistics and Data Analysis*, 2nd edition, p. 412.

Pentru a studia efectul fumatului asupra agregării trombocitelor, P.H. Levine (An acute effect of cigarette smoking on platelet function, *Circulation*, 48, 619-623, 1973) a luat sânge de la 11 subiecți înainte și după ce au fumat o țigară și a măsurat gradul în care trombocitele s-au agregat. Iată datele:

Before	25	25	27	44	30	67	53	53	52	60	28
After	27	29	37	56	46	82	57	80	61	59	43
Difference	2	4	10	12	16	15	4	27	9	-1	15

Ipoteza 0 este că fumatul nu a avut efect asupra agregării trombocitelor, i.e. diferența ar trebui să aibă media $\mu_0 = 0$. Iată codul R pentru un test t cu 2 selecții pentru a testa această ipoteză:

```
before.cig=c(25,25,27,44,30,67,53,53,52,60,28)
after.cig=c(27,29,37,56,46,82,57,80,61,59,43)
t.test(after.cig,before.cig,mu=0,paired=TRUE)
```

Iată rezultatul:

Paired t-test

```
data: after.cig and before.cig
t = 4.2716, df = 10, p-value = 0.001633
alternative hypothesis: true difference in means is not equal to
0
95 percent confidence interval:
4.91431 15.63114
sample estimates:
mean of the differences
10.27273
```

Obținem aceleași rezultate cu testul t de o selecție:

```
t.test(after.cig-before.cig,mu=0)
```

3.4.4 ANOVA unidirecțională (Testul F pentru medii egale)

Utilizare: Testăm dacă mediile populațiilor din n grupe coincid.

Date (n grupe, m date din fiecare grup)

$$\begin{array}{cccc} x_{11}, & x_{12}, & \dots, & x_{1m} \\ x_{21}, & x_{22}, & \dots, & x_{2m} \\ & & \dots & \\ x_{n1}, & x_{n2}, & \dots, & x_{nm} \end{array}$$

Presupuneri: Datele pentru fiecare grup sunt independente normale din repartiții cu medii (posibil) diferite, dar **aceeași dispersie**:

$$\begin{array}{l} x_{1j} \sim N(\mu_1, \sigma^2) \\ x_{2j} \sim N(\mu_2, \sigma^2) \\ \dots \\ x_{nj} \sim N(\mu_n, \sigma^2) \end{array}$$

Mediile grupurilor μ_i sunt necunoscute și posibil diferite. Dispersia σ este cunoscută, dar aceeași pentru toate grupurile.

H_0 : Toate mediile sunt identice $\mu_1 = \mu_2 = \dots = \mu_n$.

H_A : Nu toate mediile sunt egale.

Statistica testului: $w = \frac{MS_B}{MS_W}$, unde

$$\begin{aligned} \bar{x}_i &= \text{media grupului } i \\ &= \frac{x_{i1} + x_{i2} + \dots + x_{im}}{m}. \end{aligned}$$

$$\bar{x} = \text{media tuturor datelor.}$$

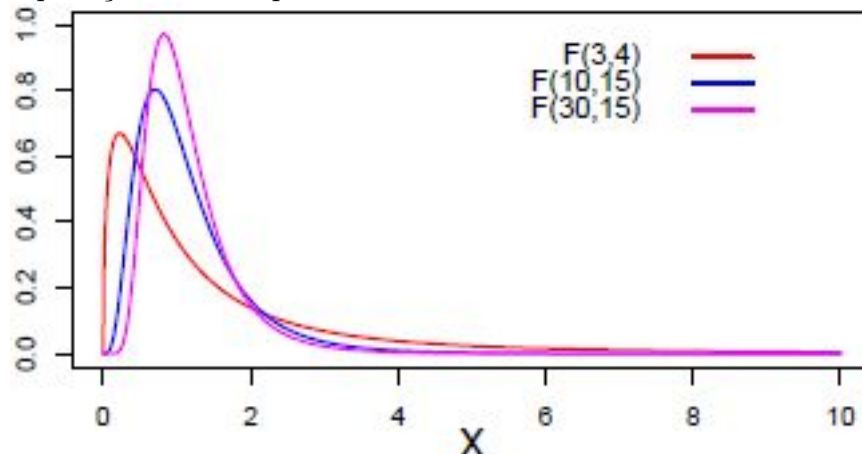
$$\begin{aligned} s_i^2 &= \text{dispersia de selecție a grupului } i \\ &= \frac{1}{m-1} \sum_{j=1}^m (x_{ij} - \bar{x}_i)^2. \end{aligned}$$

$$\begin{aligned} MS_B &= \text{dispersia între grupuri} \\ &= m \cdot \text{dispersia de selecție a mediilor grupurilor} \\ &= \frac{m}{n-1} \sum_{i=1}^n (\bar{x}_i - \bar{x})^2. \end{aligned}$$

$$\begin{aligned} MS_W &= \text{media dispersiilor grupurilor} \\ &= \text{media de selecție a } s_1^2, \dots, s_n^2 \\ &= \frac{s_1^2 + s_2^2 + \dots + s_n^2}{n}. \end{aligned}$$

Ideea: Dacă μ_i sunt toate egale, acest raport ar trebui să fie aproape de 1. Dacă ele nu sunt egale, atunci MS_B ar trebui să fie mai mare, în timp ce

MS_W ar trebui să rămână cam aceeași, deci w ar trebui să fie mai mare.
 Repartiția 0: $f(w|H_0)$ este pdf a lui $W \sim F(n-1, n(m-1))$.
 Aceasta este repartiția F cu $n-1$ și $n(m-1)$ grade de libertate. Câteva repartiții F sunt reprezentate astfel:



Valoarea p : $p = P(W > w) = 1 - \text{pf}(w, n-1, n*(m-1))$.

Observații. 1. ANOVA testează dacă toate mediile sunt egale. Nu testează dacă unele din medii sunt egale.

2. Există un test unde dispersiile nu sunt presupuse egale.

3. Există un test unde grupurile nu au toate același număr de date.

4. R are o funcție `aov()` pentru teste ANOVA. Vezi:

<https://personality-project.org/r/r.guide/r.anova.html#oneway>

<http://en.wikipedia.org/wiki/F-test>.

Exemplul 8. Tabelul arată nivelul de durere perceput de pacienți (pe o scară de la 1 la 6) după 3 proceduri medicale diferite.

T_1	T_2	T_3
2	3	2
4	4	1
1	6	3
5	1	3
3	4	5

(1) Faceți un test F care să compare mediile acestor tratamente.

(2) Bazându-vă pe test, ce puteți concluziona despre tratamente?

Răspuns. Folosind codul de mai jos, statistica F este 0.325 și valoarea p este 0.729. La orice nivel rezonabil de semnificație nu respingem ipoteza 0 că nivelul mediu de durere este același pentru toate 3 tratamentele.

Nu este rezonabil să concluzionăm că ipoteza 0 este adevărată. Cu doar 5

date pe procedură ne poate lipsi puterea de a distinge mediile diferite.

Codul R pentru test

```
# DATE ----
```

```
T1=c(2,4,1,5,3)
```

```
T2=c(3,4,6,1,4)
```

```
T3=c(2,1,3,3,5)
```

```
procedure=c(rep('T1',length(T1)),rep('T2',length(T2)),rep('T3',length(T3)))
```

```
pain=c(T1,T2,T3)
```

```
data.pain=data.frame(procedure,pain)
```

```
aov.data=aov(pain~procedure,data=data.pain) # face analiza dispersiei
```

```
print(summary(aov.data))# arată tabelul rezumatului
```

Rezumatul arată o valoare p (arătată ca $\Pr(>F)$) de 0.729. De aceea nu respingem ipoteza 0 că toate 3 mediile sunt egale.

3.4.5 Testul χ^2 pentru bunătatea potrivirii

Acesta este un test despre cât de bine o repartiție de probabilitate ipotetică se potrivește cu datele. Statistica testului este numită *statistică χ^2* și repartiția 0 asociată statisticii χ^2 este *repartiția χ^2* . Ea este notată cu $\chi^2(df)$, unde parametrul df este numărul de *grade de libertate*.

Presupunem că avem o funcție masă de probabilitate necunoscută dată de următorul tabel.

Outcomes	ω_1	ω_2	...	ω_n
Probabilities	p_1	p_2	...	p_n

În testul χ^2 pentru bunătatea potrivirii ipoteza este o mulțime de valori pentru probabilități. Tipic, ipoteza este că probabilitățile au o repartiție cunoscută cu anumiți parametri, de exemplu binomială, Poisson, multinomială. Apoi testul încearcă să determine dacă această mulțime de probabilități ar fi putut genera rezonabil datele colectate.

Utilizare: Testăm dacă date discrete se potrivesc cu o anumită funcție masă de probabilitate finită.

Date: Câte un număr observat O_i pentru fiecare rezultat posibil ω_i .

Presupuneri: Niciuna.

H_0 : Datele provin dintr-o anumită repartiție discretă.

H_A : Datele provin dintr-o repartiție diferită.

Statistica testului: Datele constau din numerele observate O_i pentru fiecare ω_i . Din tabelul de probabilitate al ipotezei 0 obținem o mulțime de numere

așteptate E_i . Sunt 2 statistici pe care le putem folosi:

$$\text{Statistica raportului de verosimilitate } G = 2 \sum O_i \ln \left(\frac{O_i}{E_i} \right)$$

$$\text{Statistica } \chi^2 \text{ a lui Pearson } X^2 = \sum \frac{(O_i - E_i)^2}{E_i}.$$

Există o teoremă că sub ipoteza 0 $X^2 \approx G$ și ambele sunt aproximativ χ^2 . Înainte de calculatoare, X^2 era folosită deoarece era mai ușor de calculat. Acum, este mai bine să folosim G deși încă veți vedea X^2 folosită destul de des.

Grade de libertate df : Pentru testele χ^2 numărul gradelor de libertate poate fi un pic complicat. În acest caz $df = n - 1$. Este calculat ca numărul de celule care pot fi completate liber în concordanță cu statisticile folosite pentru a calcula numerele așteptate din celule presupunând H_0 .

Repartiția 0: Presupunând H_0 , ambele statistici au (aproximativ) o repartiție χ^2 cu df grade de libertate. Adică atât $f(G|H_0)$ cât și $f(X^2|H_0)$ au aceeași pdf ca $Y \sim \chi^2(df)$.

Valoarea p :

$$\begin{aligned} p &= P(Y > G) &= 1 - \text{pchisq}(G, df) \\ p &= P(Y > X^2) &= 1 - \text{pchisq}(X^2, df) \end{aligned}$$

Codul R: Funcția R `chisq.test` poate fi folosită pentru a face calcule pentru un test χ^2 folosind X^2 . Pentru G , trebuie să facem calculele "cu mâna" (ajutați de R) sau să găsim un pachet care are o funcție. (Probabil va fi numită `likelihood.test` sau `G.test`.)

Observație. Când statistica raportului de verosimilitate G este folosită, testul este de asemenea numit **test G** sau **testul raportului de verosimilitate**.

Exemplul 9. Primul exemplu χ^2 . Presupunem că avem un experiment care produce date numerice. Pentru acest experiment rezultatele posibile sunt 0, 1, 2, 3, 4, 5 sau mai mult. Facem 51 de încercări și numărăm frecvența fiecărui rezultat, obținând următoarele date:

Outcomes	0	1	2	3	4	≥ 5
Observed counts	3	10	15	13	7	3

Presupunem că ipoteza noastră H_0 este că datele provin din 51 de încercări ale repartiției binomiale(8,0.5) și ipoteza noastră alternativă H_A este că datele provin dintr-o altă repartiție.

1. Faceți un tabel al numerelor observate și așteptate.
2. Calculați statistica raportului de verosimilitate G și statistica χ^2 a lui Pearson X^2 .

3. Calculați numărul de grade de libertate ale repartiției 0.
4. Calculați valorile p corespunzătoare lui G și X^2 .

Răspuns. 1. Presupunând H_0 , datele provin dintr-o repartiție binomială(8,0.5). Avem 51 de observații, deci numărul așteptat pentru fiecare rezultat este 51·probabilitatea lui. Folosind R se obține:

Rezultate	0	1	2	3	4	≥ 5
Numere observate	3	10	15	13	7	3
Probabilități H_0	0.00390625	0.03125	0.109375	0.21875	0.2734375	0.3632812
Numere așteptate	0.1992188	1.59375	5.578125	11.15625	13.9453125	18.52734

2. Folosind formulele de mai sus calculăm $X^2 = 116.4056$ și $G = 66.08122$.

3. Singura statistică folosită în calculul numerelor așteptate a fost numărul total de observații 51. Deci, numărul de grade de libertate este 5, i.e. putem pune 5 dintre numerele din celule liber și ultimul este determinat de cerința ca totalul să fie 51.

4. Valorile p sunt $p_G = 1 - \text{pchisq}(G, 5)$ și $p_{X^2} = 1 - \text{pschisq}(X^2, 5)$. Ambele valori p sunt efectiv 0. Pentru aproape orice nivel de semnificație respingem H_0 în favoarea H_A .

Exemplul 10. ([Grade de libertate](#).) Presupunem că avem aceleași date ca în exemplul precedent, dar ipoteza 0 a noastră este că datele vin din încercări independente ale repartiției binomiale(8, θ), unde θ poate fi oricât între 0 și 1. H_A este că datele vin dintr-o altă repartiție. În acest caz trebuie să estimăm θ din date, de exemplu folosind MLE. În total am calculat 2 valori din date: numărul total de date și estimarea lui θ . Deci, numărul de grade de libertate este $6 - 2 = 4$.

Exemplul 11. Experimentele genetice ale lui Mendel. (Adaptat din Rice, *Mathematical Statistics and Data Analysis*, 2nd ed., exemplul C, p. 314).

În unul din experimentele lui pe mazăre, Mendel a încrucișat 556 masculi netezi, galbeni cu femele verzi zbârcite. Presupunând că genele netede și zbârcite apar cu frecvență egală ne-am aștepta ca 1/4 din populația de mazăre să aibă 2 gene netede (SS), 1/4 să aibă 2 gene zbârcite (ss) și 1/2 rămasă să fie heterozigoți Ss . De asemenea așteptăm aceste fracții pentru gene galbene (Y) și verzi (y). Dacă genele de culoare și netezime sunt moștenite independent și cele netede și galbene sunt ambele dominante ne-am aștepta la următorul tabel de frecvențe pentru fenotipuri.

	Yellow	Green	
Smooth	9/16	3/16	3/4
Wrinkled	3/16	1/16	1/4
	3/4	1/4	1

Tabelul de probabilitate pentru ipoteza 0

Deci, din 556 de încrucișări, numărul așteptat de păstăi netede galbene este $556 \cdot 9/16 = 312.75$. Analog pentru celelalte posibilități. Iată un tabel care dă numerele observate și așteptate din experimentele lui Mendel.

	Observed count	Expected count
Smooth yellow	315	312.75
Smooth green	108	104.25
Wrinkled yellow	102	104.25
Wrinkled green	31	34.75

Ipoteza 0 este că numerele observate sunt selecții aleatoare repartizate conform tabelului de frecvențe de mai sus. Folosim numerele să calculăm statisticile noastre.

Statistica raportului de verosimilitate este

$$\begin{aligned}
 G &= 2 \sum O_i \ln \left(\frac{O_i}{E_i} \right) \\
 &= 2 \left(315 \ln \left(\frac{315}{312.75} \right) + 108 \ln \left(\frac{108}{104.25} \right) + 102 \ln \left(\frac{102}{104.25} \right) + 31 \ln \left(\frac{31}{34.75} \right) \right) \\
 &= 0.6184391.
 \end{aligned}$$

Statistica χ^2 a lui Pearson este

$$\begin{aligned}
 X^2 &= \sum \frac{(O_i - E_i)^2}{E_i} \\
 &= \frac{(315 - 312.75)^2}{312.75} + \frac{(108 - 104.25)^2}{104.25} + \frac{(102 - 104.25)^2}{104.25} + \frac{(31 - 34.75)^2}{34.75} \\
 &= 0.6043165.
 \end{aligned}$$

Cele 2 statistici sunt foarte apropiate. Acesta este cazul de obicei. În general statistica raportului de verosimilitate este mai robustă și ar trebui preferată. Numărul de grade de libertate este 3, deoarece sunt 4 cantități observate și o relație între ele: suma lor este 556. Deci, sub ipoteza 0, G are o repartiție $\chi^2(3)$. Folosind **R** pentru a calcula valoarea p obținem

$$p = 1 - \text{pchisq}(0.6184391, 3) = 0.8921985.$$

Nu vom respinge ipoteza 0 la orice nivel rezonabil de semnificație.

Valoarea p folosind statistica lui Pearson este 0.8954435 - aproape identică.

3.4.6 Testul χ^2 pentru omogenitate

Acesta este un test pentru a vedea dacă unele seturi independente de date provin din aceeași repartiție. (Înțelesul omogenității în acest caz este că toate repartițiile sunt aceleași.)

Utilizare: Testăm dacă m mulțimi independente de date discrete provin din aceeași repartiție.

Rezultate: $\omega_1, \omega_2, \dots, \omega_n$ sunt rezultate posibile. Acestea sunt aceleași pentru fiecare set de date.

Date: Presupunem m seturi independente de date dând numere pentru fiecare rezultat posibil. Adică, pentru setul de date i avem un număr observat O_{ij} pentru fiecare rezultat posibil ω_j .

Presupuneri: Niciuna.

H_0 : Fiecare set de date provine din aceeași repartiție. (Nu specificăm care este această repartiție.)

H_A : Seturile de date nu provin din aceeași repartiție.

Statistica testului: Vezi exemplul de mai jos. Sunt mn celule conținând numerele pentru fiecare rezultat pentru fiecare set de date. Folosind Repartiția 0 putem estima numerele așteptate pentru fiecare din seturile de date. Statisticile X^2 și G sunt calculate exact ca mai sus.

Numărul de grade de libertate $df = (m - 1)(n - 1)$. (Vezi exemplul de mai jos.)

Repartiția 0: $\chi^2(df)$. Valorile p sunt calculate ca în testul χ^2 pentru bunătatea potrivirii.

Codul R: Funcția R `chisq.test` poate fi folosită pentru a face calcule pentru un test χ^2 folosind X^2 . Pentru G , trebuie să facem calculele "cu mâna" (ajutați de R) sau să găsim un pachet care are o funcție. (Probabil va fi numită `likelihood.test` sau `G.test`.)

Exemplul 12. Cineva pretinde că a găsit o piesă de teatru pierdută de mult a lui William Shakespeare. Vă cere să testați dacă piesa a fost sau nu scrisă de Shakespeare.

Mergeți la <http://www.opensourceshakespeare.org> și alegeți aleator 12 pagini din *Regele Lear* și numărați folosirea cuvintelor uzuale. Faceți același lucru pentru "piesa pierdută de mult". Obțineți următorul tabel.

Word	a	an	this	that
<i>King Lear</i>	150	30	30	90
Long lost work	90	20	10	80

Folosind aceste date, faceți un test de semnificație pentru pretenția că piesa pierdută de mult este de William Shakespeare. Folosiți un nivel de semnificație

de 0.1.

Răspuns. H_0 : Pentru cele 4 cuvinte numărate piesa pierdută de mult are aceleași frecvențe relative ca numerele luate din *Regele Lear*.

Totalul cuvintelor numărate în ambele piese este 500, deci estimarea de verosimilitate maximă a frecvențelor relative presupunând H_0 este numărul total pentru fiecare cuvânt împărțit la 500.

Word	a	an	this	that	Total count
<i>King Lear</i>	150	30	30	90	300
Long lost work	90	20	10	80	200
totals	240	50	40	170	500
rel. frequencies under H_0	240/500	50/500	40/500	170/500	500/500

Acum, numerele așteptate pentru fiecare piesă sub H_0 sunt numerele totale pentru acea piesă înmulțite cu frecvențele relative din tabelul de mai sus. Următorul tabel dă numerele: (observate, așteptate) pentru fiecare piesă. Pe ultima linie, în loc de "249" se va citi "240".

Word	a	an	this	that	Totals
<i>King Lear</i>	(150, 144)	(30, 30)	(30, 24)	(90, 102)	(300, 300)
Long lost work	(90, 96)	(20, 20)	(10, 16)	(80, 68)	(200, 200)
Totals	(249, 240)	(50, 50)	(40, 40)	(170, 170)	(500, 500)

Statistica χ^2 este

$$\begin{aligned}
 X^2 &= \sum \frac{(O_i - E_i)^2}{E_i} \\
 &= \frac{6^2}{144} + \frac{0^2}{30} + \frac{6^2}{24} + \frac{12^2}{102} + \frac{6^2}{96} + \frac{0^2}{20} + \frac{6^2}{16} + \frac{12^2}{68} \\
 &\approx 7.904412.
 \end{aligned}$$

Sunt 8 celule și toate numerele marginale sunt fixate deoarece a fost nevoie de ele pentru a determina numerele așteptate. Am putea să punem liber valorile din 3 celule din tabel, de exemplu cele 3 celule albastre, apoi restul celulelor sunt determinate pentru a face corecte totalele marginale. Astfel, $df = 3$. (Sau ne puteam reaminti că $df = (m - 1)(n - 1) = 3 \cdot 1 = 3$, unde m este numărul de coloane și n este numărul de linii.)

Folosind R găsim $1 - pchisq(7.904412, 3) = 0.04802909$. Deoarece aceasta este mai mică decât nivelul nostru de semnificație de 0.1, respingem ipoteza 0 că frecvențele relative ale cuvintelor sunt aceleași în ambele piese.

Dacă facem în plus presupunerea că toate piesele lui Shakespeare au frecvențe similare ale cuvintelor (ceea ce este ceva ce putem verifica) concluzionăm că piesa pierdută probabil nu este a lui Shakespeare.