

Decizie de indexare a faptei de plagiat la poziția 00439 / 06.01.2020 și pentru admitere la publicare în volum tipărit

care se bazează pe:

A. Nota de constatare și confirmare a indiciilor de plagiat prin fișa suspiciunii inclusă în decizie.

Fișa suspiciunii de plagiat / Sheet of plagiarism's suspicion		
Incidența minimă a suspiciunii / Minimum incidence of suspicion		
OS	STOICA, Cristina Maria, ALEXA, Elena Lidia. <i>Cercetări de marketing. Teorie și aplicații</i> . București: C.H.Beck. 2010.	
OA	RATEAU, Patrick. <i>Metodele și statisticile experimentale în științele umane</i> . Iași: Pdirom. 2004.	
P.01	p.166 – p.169	p.140 – p.144
P.02	p.169 – p.170	p.145 – p.146
P.03	p.170 – p.174	p.147 – p.150
P.04	p.173 – p.174	p.150 – p.152
P.05	p.174 – p.175	p.152 – p.153
P.06	p.189 – p.190	p.154 – p.157
P.07	p.192 – p.194	p.158 – p.162
P.08	p.194 – p.197	p.163 – p.167
P.09	p.198 – p.201	p.167 – p.172
P.10	p.201 – p.203	p.172 – p.175
Fișa întocmită pentru includerea suspiciunii în Indexul Operelor Plagiate în România de la Sheet drawn up for including the suspicion in the Index of Plagiarized Works in Romania at www.plagiate.ro		

Notă: Prin „p.72:00” se înțelege paragraful care se termină la finele pag.72. Notăția „p.00:00” semnifică până la ultima pagină a capitoului curent, în întregime de la punctul inițial al preluării.

Note: By „p.72:00” one understands the text ending with the end of the page 72. By „p.00:00” one understands the taking over from the initial point till the last page of the current chapter, entirely.

B. Fișa de argumentare a calificării de plagiat alăturată, fișă care la rândul său este parte a deciziei.

Echipa Indexului Operelor Plagiate în România

Argumentarea calificării faptei de plagiat

Nr. crt.	Descrierea situației care este încadrată drept plagiat	Se confirmă
1.	Preluarea identică a unor pasaje dintr-o operă autentică publicată, fără precizarea întinderii și menționarea provenienței și însușirea acestora într-o lucrare ulterioară celei autentice.	✓
2.	Preluarea identică a unor pasaje dintr-o operă autentică publicată, care sunt rezumate ale unor opere anterioare operei autentice, fără precizarea întinderii și menționarea provenienței și însușirea acestora într-o lucrare ulterioară celei autentice.	
3.	Preluarea identică a unor figuri dintr-o operă autentică publicată, fără menționarea provenienței și însușirea acestora într-o lucrare ulterioară celei autentice.	
4.	Preluarea identică a unor poze dintr-o operă autentică publicată, fără menționarea provenienței și însușirea acestora într-o lucrare ulterioară celei autentice.	
5.	Preluarea identică a unor tabele dintr-o operă autentică publicată, fără menționarea provenienței și însușirea acestora într-o lucrare ulterioară celei autentice.	✓
6.	Republicarea unei opere anterior publicate, prin includerea unui nou autor fără contribuție explicită în lista de autori	
7.	Republicarea unei opere anterior publicate, prin excluderea unui autor din lista inițială de autori.	
8.	Preluarea identică de pasaje dintr-o operă autentică publicată, fără precizarea întinderii și menționarea provenienței, fără nici o intervenție care să justifice exemplificarea sau critica prin aportul creator al autorului care preia și însușirea acestora într-o lucrare ulterioară celei autentice.	✓
9.	Preluarea identică de figuri sau reprezentări grafice dintr-o operă autentică publicată, fără menționarea provenienței, fără nici o intervenție care să justifice exemplificarea sau critica prin aportul creator al autorului care preia și însușirea acestora într-o lucrare ulterioară celei autentice.	
10.	Preluarea identică de tabele dintr-o operă autentică publicată, fără menționarea provenienței, fără nici o intervenție care să justifice exemplificarea sau critica prin aportul creator al autorului care preia și însușirea acestora într-o lucrare ulterioară celei autentice.	✓

Actualizat la 1 decembrie 2019.

Notă: Prin „proveniență” se înțelege informația din care se pot identifica cel puțin numele autorului / autorilor, titlul operei, anul apariției.

Constatarea faptei istorice de plagiat adică fapta de plagiat care se referă la scrieri care au fost deja aduse la cunoștința publicului este posibilă când sunt îndeplinite simultan cerințele ca:

- Condiția de preluare neconformă prin care, în scrierea plagiată, se pot identifica fragmente care nu sunt delimitate în mod explicit și pentru care nu există nici o indicație explicită a provenienței ca referință bibliografică.
- Există o însușire explicită prin care fragmentul preluat apare într-o altă scriere, dată publicității ulterior scrierii autentice, sub numele unei persoane care o revendică în mod implicit ca fiind a sa și/sau că este publicată pentru prima oară.

BCU Cluj-Napoca



LEGAL201112636

Colecția Oeconomica

**Cristina Maria Stoica
Elena Lidia Alexa**

Cercetări de marketing

Teorie și aplicații



Editura C. H. Beck



AVERTISMENT!

Având în vedere amploarea luată de fenomenul fotocopierii lucrărilor de specialitate, mai ales în domeniul Dreptului, atragem atenția că, potrivit art. 14 și 140 din Legea nr. 8/1996 privind dreptul de autor și drepturile conexe, reproducerea operelor sau a produselor purtătoare de drepturi conexe, dacă respectiva reproducere a fost efectuată fără autorizarea sau consimțământul titularului drepturilor recunoscute de legea menționată, constituie infracțiune și se pedepsește cu închisoare sau cu amendă. Prin reproducere, conform legii, se înțelege realizarea, integrală sau parțială, a uneia ori a mai multor copii ale unei opere, direct sau indirect, temporar ori permanent, prin orice mijloace și sub orice formă.

Nu vă faceți părtași la distrugerea cărții!

Editura C.H. Beck este acreditată CNCIS - Consiliul Național al Cercetării Științifice din Învățământul Superior

Cercetări de marketing. Teorie și aplicații
Cristina Maria Stoica, Elena Lidia Alexa

Copyright © 2010 – Editura C.H. Beck

Toate drepturile rezervate Editurii C.H. Beck.

Nicio parte din această lucrare nu poate fi copiată fără acordul scris al Editurii C.H. Beck.

Drepturile de distribuție în străinătate aparțin în exclusivitate editurii.

Descrierea CIP a Bibliotecii Naționale a României
STOICA, CRISTINA MARIA

Cercetări de marketing : teorie și aplicații / Stoica
Cristina Maria, Alexa Elena Lidia. - București : Editura
C.H. Beck, 2010

Bibliogr.

ISBN 978-973-115-777-1

I. Alexa, Elena Lidia

339.138

Editura C.H. Beck

Str. Serg. Nuțu Ion nr. 2, sector 5, București

Tel.: 021.410.08.47; 021.410.08.09;
021.410.08.73; 021.410.08.46

Fax: 021.410.08.48

E-mail: comenzi@beck.ro

Redactor: Mihaela Popescu
Tehnoredactor: Cătălin Mantu
Coperta: Constantin Trif

Cristina Maria Stoica, este profesor universitar doctor, în cadrul Universității "Petre Andrei" din Iași. Este coautoare a lucrărilor: Dossier francophones – Marketing et Finances; Marketing modern. Concepte, tehnici, strategii științifice. De asemenea, este autoare a lucrărilor: Managementul strategic al resurselor umane; Raționalitatea concurențială și strategii competitive (editată în Serbia); Managementul resurselor umane; Comunicarea și negocierea în afaceri internaționale; Managementul modelelor matematice utilizate în macroeconomie; Modelarea comportamentului consumatorului; Modelarea comportamentului producătorului; Cercetări de marketing. Autoarea a publicat 43 de studii și articole în reviste de specialitate și a realizat 29 de studii de cercetare în domeniul marketingului și managementului.

Elena Lidia Alexa este director de marketing la compania Print Multicolor din Iași, asistent universitar în cadrul Universității "Petre Andrei" din Iași și doctorand în marketing în cadrul Școlii Doctorale de Economie a Universității „Al. I. Cuza” Iași.

COMPARAREA FRECVENȚELOR. TESTUL HI-PĂTRAT

Se numește frecvență (sau proporție) a unei clase dintr-o populație sau dintr-un eșantion, raportul dintre efectivul clasei și efectivul total al populației sau al eșantionului. Fie un eșantion de mărime n și o clasă i a acestui eșantion cu efectivul n_i .

Se numește p_i frecvența clasei i dată de $p_i = n_i / n$. De exemplu, fie un eșantion de $n = 100$ de studenți, dintre care numărăm câți fac parte din clasa i „studenți salariați”. Să presupunem că $n_i = 25$. Frecvența p_i a studenților salariați este deci $p_i = 25/100 = 25\% = 0,25$.

Toți cercetătorii în științe umaniste trebuie să manipuleze observații exprimate sub formă de frecvențe, adesea procentaje. Ca și în cazul mediilor, este deci important să dispunem de teste statistice prin care să putem compara frecvențele.

Putem folosi, în general, două tehnici. Prima se bazează pe o perspectivă parametrică, și este aceeași cu cea folosită la compararea mediilor. Are însă două limite: permite compararea a două frecvențe, și numai a două, și poate fi aplicată doar dacă dispunem de eșantioane de mărime mare, care să asigure normalitatea distribuției în proporția populației de origine.

Acestei prime tehnici i se preferă în general o a doua, care evită ambele inconveniente și nu necesită ajustarea la legea normală (este deci o tehnică bazată pe o perspectivă neparametrică) – de aceea poate fi folosită și pe efective reduse. În plus, permite compararea mai multor frecvențe între ele. Aceasta este cea pe care o vom prezenta în continuare.

TESTUL HI-DOI (χ^2): PRINCIPIUL GENERAL

Testul Hi-doi (sau Hi-pătrat) face posibilă măsurarea distanței între una sau mai multe repartiții de efective observate și respectiva sau respectivele repartiții redistribuite, conform unui model probabilist.

Comparațiile se efectuează asupra efectivelor distribuțiilor de comparat, și nu direct asupra frecvențelor. Întrucât este vorba despre repartiții de efective, trebuie ca datele să fie plasate în categorii. Așadar, testul este destinat în special cazului datelor care provin din scale nominale și ordinale. Dacă datele au fost culese cu o scală de intervale, atunci intervalul de variație al distribuției trebuie transformat în clase. Formula de calcul a lui hi-pătrat este:

$$\chi^2 = \sum \frac{(\text{Efectivul observat} - \text{Efectivul teoretic})^2}{\text{Efectivul teoretic}}$$

Se observă că acest indicator se referă la efective, și nu la proporții. Se observă, de asemenea, că ridicarea la pătrat a numărătorului interzice pentru hi-pătrat valorile negative. Cu alte cuvinte, testul Hi-doi este un test direcțional.

Ca întotdeauna, respingerea sau păstrarea ipotezei nule depinde simultan de situația lui hi-pătrat calculat față de hi-pătrat critic (mai mic, egal sau superior lui) dat de tabelul Hi-doi, de gradul de eroare fixat și numărul de grade de libertate, diferit în funcție de folosirea testului.

- Dacă hi-pătrat calculat este inferior lui hi-pătrat critic, ipoteza nulă se păstrează. Diferența între cele două frecvențe nu este semnificativă.

- Dacă hi-pătrat calculat este mai mare sau egal cu hi-pătrat critic, ipoteza nulă se respinge. Diferența între cele două frecvențe este semnificativă.

Citirea tabelului Hi-doi

Să presupunem că vrem să găsim valoarea critică a lui Hi-doi asociată unui $gdl = 4$, la pragul de încredere de 1%. Se obține:

Prag/gdl	0.20	0.10	0.05	0.02	0.01	0.001	0.0001	0.00001
					↓			
1								
...								
4 →					13.28			
...								

COMPARAREA UNEI FRECVENȚE ȘI A UNEI NORME: HI-PĂTRAT DE AJUSTARE

Aici trebuie studiat dacă o distribuție de efective observate poate fi sau nu considerată echivalentă cu o distribuție teoretică reconstruită. Reconstrucția se poate face fie prin echirepartiția efectivelor modalităților variabilei studiate, fie plecând de la informații cunoscute cu privire la repartiția modalităților acestei variabile. Trebuie înțeles că procedura de calcul nu se schimbă, ci variază numai modul de reconstruire a distribuției efectivelor teoretice, în funcție de informațiile de care dispunem.

Cazul echirepartiției efectivelor teoretice

Echirepartiția efectivelor teoretice înseamnă că fiecare modalitate a variabilei studiate are teoretic aceeași probabilitate de apariție ca și celelalte modalități.

Fie, de exemplu, cazul celor două fețe ale unei monede. Să presupunem că un individ care joacă „cap sau pajură” obține, după 100 de aruncări consecutive, de 41 de ori cap și de 59 de ori pajură.

Se știe că fiecare dintre fețele monedei are, teoretic, o șansă din două să cadă la fiecare aruncare. La 100 de aruncări consecutive, repartiția teoretică a capului și pajurei este: de 50 de ori cap și de 50 de ori pajură. Se pune acum întrebarea dacă diferențele între efectivele observate (41;59) și efectivele teoretice (50;50) sunt suficient de mari ca să putem reține intervenția unei variabile oarecare în timpul aruncărilor (o monedă trucată, o abilitate specială a jucătorului, o greșeală în notarea frecvențelor etc.).

Diferența se calculează cu formula lui hi-pătrat:

$$\chi^2 = \sum \frac{(\text{Efectiv observat} - \text{Efectiv teoretic})^2}{\text{Efectiv teoretic}} = \frac{(41 - 50)^2}{50} + \frac{(59 - 50)^2}{50} = 3,24$$

Nu ne rămâne decât să comparăm acest hi-pătrat cu cel critic, dat de tabel pentru gradul de libertate corespunzător. În cazul comparării unei frecvențe și a unei norme, $gdl = \text{numărul de modalități ale variabilei} - 1$, adică, în cazul nostru: $2 - 1 = 1$. La riscul de eroare de 5% (0,05), hi-pătrat critic = 3,84. Hi-pătrat calculat este deci inferior lui hi-pătrat critic.

Se concluzionează că nu există o diferență semnificativă între distribuția observată a aruncărilor și distribuția lor teoretică. Diferența observată se datorează deci fluctuațiilor întâmplătoare. Să trecem acum la exerciții legate mai direct de științele umaniste.

APLICAȚIE

Directorul unei fabrici de jucării vrea să afle dacă, în general, copiii sunt mai atrași de jucăriile din pluș în formă de iepuraș, broască țestoasă sau elefant. El lansează un studiu în care cele trei jucării sunt prezentate unui număr de 54 de copii de 2 ani.

Rezultatele sunt următoarele: 10 aleg iepurașul, 23 broasca-țestoasă și 21 elefantul. Se poate trage concluzia, cu un risc de eroare de 5%, că există o alegere preferențială din partea copiilor?

În ipoteza independenței stricte a tipului de animal și a alegerii copiilor, ar trebui să avem următoarea repartiție teoretică: $54/3 = 18$ copii care aleg iepurașul, $54/3 = 18$ copii care aleg broasca-țestoasă și $54/3 = 18$ copii care aleg elefantul. Se compară apoi distribuția observată (10;23;21) cu distribuția teoretică (18;18;18)

cu ajutorul testului hi-pătrat. $\chi^2 = \frac{(10-18)^2}{18} + \frac{(23-18)^2}{18} + \frac{(21-18)^2}{18} = 5,45$.

La gradul de libertate $gdl = 3 - 1 = 2$, hi-pătrat critic dat de tabel, la pragul de 5% (0,05), este de 5,99. Hi-pătrat calculat este inferior acestei valori. Nu există deci un efect al tipului de animal asupra alegerii făcute de copii, întrucât distribuția alegerilor este, din punct de vedere statistic, echivalentă cu echirepartiția. Diferențele observate rezultă deci din fluctuațiile de eșantionare.

Plecând de la acest exemplu, se observă că testul hi-pătrat de ajustare permite concluzii globale asupra repartiției efectivelor fiecărei modalități a unei variabile. Absența unei diferențe globale nu înseamnă însă neapărat că toate modalitățile sunt echivalente.

Chiar acesta este cazul aici, fiindcă se observă că alegerile iepurașului sunt inferioare celor ale celorlalte două jucării. La fel, există cazuri în care se poate concluziona asupra unei diferențe semnificative în timp ce toate modalitățile nu sunt diferite. Când numărul de modalități crește, repartiția trebuie adesea analizată mai precis, cu ajutorul analizelor două câte două.

În exemplul de față, se obțin următoarele rezultate:

1. Comparția iepuraș / broască țestoasă:

- Efective observate: 10; 23; Efective teoretice: 16,5; 16,5

$$- \chi^2 = \frac{(10-16,5)^2}{16,5} + \frac{(23-16,5)^2}{16,5} = 5,12$$

- $gdl = 2 - 1 = 1$. Hi-pătrat critic = 3,84 pentru riscul de eroare de 5%

- Diferența este semnificativă, copiii au preferat broasca țestoasă iepurașului.

2. Comparția iepuraș / elefant:

- Efective observate: 10; 21; Efective teoretice: 15,5; 15,5

$$- \chi^2 = \frac{(10 - 15,5)^2}{15,5} + \frac{(21 - 15,5)^2}{15,5} = 3,90$$

- gdl = 2 - 1 = 1. Hi-pătrat critic = 3,84 pentru riscul de eroare de 5%

- Diferența este semnificativă, copiii au preferat elefantul iepurașului.

3. Compararea broască țestoasă / elefant:

- Efective observate: 23; 21; Efective teoretice: 22; 22

$$- \chi^2 = \frac{(23 - 22)^2}{22} + \frac{(21 - 22)^2}{22} = 0,09$$

- gdl = 2 - 1 = 1. Hi-pătrat critic = 3,84 pentru riscul de eroare de 5%

- Diferența nu este semnificativă, copiii au ales atât elefantul, cât și broasca-țestoasă.

Cu alte cuvinte, deși nu există o diferență statistică globală între cele trei jucării, analiza două câte două spune că alegerea copiilor s-a îndreptat semnificativ mai puțin asupra iepurașului decât asupra celorlalte două animale. Directorul fabricii de jucării ar trebui deci să reducă producția de iepurași!

Cazul repartiției efectivelor teoretice independente

Spre deosebire de cazul precedent, fiecare modalitate a variabilei prezintă aici o probabilitate teoretică de apariție independentă de alte modalități. Cu alte cuvinte, unele modalități au o probabilitate de apariție mai mare sau mai mică decât altele.

Ne aflăm în cazul în care dispunem deja de frecvențele teoretice ale modalităților unui fenomen (statisticile demografice, sociale, ale unei populații etc.). Această metodă permite compararea eșantionelor, în raport cu date deja cunoscute, așa cum este adesea cazul în științele umaniste.

Testul statistic este strict același ca și în cazul precedent. Diferența decurge din faptul că distribuția teoretică nu poate fi obținută aici împărțind efectivul la numărul de modalități ale variabilei. Într-adevăr, partea respectivă a fiecărei modalități nu este constantă. Trebuie deci reconstituită distribuția efectivelor teoretice, ținând cont de diferențele proporțiilor. Dacă datele cunoscute sunt exprimate în procente, așa cum este adesea cazul, se aplică o regulă de trei simplă:

$$\text{Efectivul teoretic} = (\text{Efectivul total al eșantionului}) \times (\% \text{ modalității}) / 100$$

Astfel, se convertește distribuția teoretică pentru a face compatibilă scala sa cu cea a distribuției observate.

APLICAȚIE

Un studiu asupra populației afirmă că 30% dintre locuitorii orașului Iași vorbesc engleza, 20% spaniola, 5% germana și 45% nici una dintre aceste limbi. Ce putem spune despre această afirmație dacă știm că, dintr-un eșantion de 400 de ieșeni interogați, 100 au declarat că vorbesc engleza, 100 spaniola și 30 germana?

Se compară frecvența observată cu o normă dată de rezultatele anchetei. Efective observate sunt culese de la 400 de persoane, în timp ce rezultatele anchetei se exprimă în procente. Pentru a le putea compara, cele două distribuții trebuie aduse la scala celor 400 de persoane din eșantion. Se obține:

Limba	Efective observate	Efective teoretice
Engleză	100	$\frac{400 \cdot 30}{100} = 120$
Spaniolă	100	$\frac{400 \cdot 20}{100} = 80$
Germană	30	$\frac{400 \cdot 5}{100} = 20$
Alta	170	$\frac{400 \cdot 45}{100} = 180$

Rămâne de aplicat formula convențională pentru hi-pătrat:

$$\chi^2 = \sum \frac{(\text{Efectiv observat} - \text{Efectiv teoretic})^2}{\text{Efectiv teoretic}}$$

Adică:

$$\chi^2 = \frac{(100-120)^2}{120} + \frac{(100-80)^2}{80} + \frac{(30-20)^2}{20} + \frac{(170-180)^2}{180} = 13,88$$

La pragul de 0.05 și gradul de libertate $gdl = 4 - 1 = 3$, hi-pătrat critic dat de tabel este 7,82. Hi-pătrat calculat este deci superior acestei valori și îi rămâne superior, atât la pragul de 0.02, cât și la pragul de 0.01. Se poate afirma așadar, cu un risc de eroare de 1%, că între distribuția observată și distribuția livrată de anchetă există o diferență semnificativă.

Cu alte cuvinte, se trage concluzia fie că afirmația anchetei este imperfectă, fie că eșantionul considerat nu este reprezentativ pentru populație.

COMPARAREA MAI MULTOR FRECVENȚE: χ^2 DE INDEPENDENȚĂ

P.03

Testul hi-pătrat poate fi folosit pentru evaluarea diferențelor între două sau mai multe variabile. În acest caz, se numește hi-pătrat de independență, întrucât permite evaluarea gradului de independență care există între modalitățile încrucișate ale variabilelor aflate în joc.

Frecvențele calculate pe eșantioane independente

Se studiază aici legătura între mai multe variabile cu statut diferit, ale căror modalități sunt încrucișate. Încrucișarea variabilelor se reprezintă în general printr-un tabel încrucișat, numit și de contingență.

APLICAȚIE

Într-o anchetă asupra timpului liber, un tânăr cercetător vrea să știe dacă există vreo legătură între sex și modul de petrecere a timpului liber. În cursul studiului, întreabă atât femei, cât și bărbați dacă preferă să meargă la cinema sau la teatru. Iată răspunsurile, sub forma unui tabel de contingență:

	Teatru	Cinema	$\sum$ linii
Bărbați	45	55	100
Femei	58	44	102
$\sum$ coloane	103	99	N = 202

Se știe că testul hi-pătrat face posibilă măsurarea distanței între o distribuție observată și distribuția teoretică reconstruită. Or, aici, repartiția efectivelor teoretice trebuie efectuată plecând de la ceea ce s-a observat, și nu pe baza echirepartiției sau a proporțiilor cunoscute în prealabil.

Informația esențială pentru determinarea distribuției efectivelor teoretice se găsește în totalul marginal al liniilor și coloanelor din tabelul de contingență, adică în așa-numitele efective marginale. Se observă, pe de o parte, că există o diferență globală la nivelul modului de petrecere a timpului liber: se pare că teatrul este mai des ales decât cinematograful.

Pe de altă parte însă, se observă că studentul care a făcut ancheta, nu s-a adresat aceluiași număr de bărbați și de femei. Există deci un dezechilibru în efectivele observate, care trebuie păstrat în calculul efectivelor teoretice. Iată metoda: pentru fiecare căsuță din tabel, se efectuează produsul sumei liniilor și al sumei coloanelor la intersecția cărora se află căsuța, iar rezultatul se împarte la efectivul total. Cu alte cuvinte, determinarea efectivului teoretic al fiecărei căsuțe din tabel se efectuează cu formula următoare :

(Efectivul marginal pe linie al căsuței x Efectivul marginal pe coloană al căsuței) / Efectivul total

În exemplul nostru, se obțin următoarele efective teoretice:

	Teatru	Cinema
Bărbați	$\frac{100 \cdot 103}{202} = 50,99$	$\frac{100 \cdot 99}{202} = 49,00$
Femei	$\frac{102 \cdot 103}{202} = 52,00$	$\frac{102 \cdot 99}{202} = 49,99$

Rămâne de aplicat formula convențională pentru hi-pătrat:

$$\chi^2 = \frac{(45 - 50,99)^2}{50,99} + \frac{(55 - 49)^2}{49} + \frac{(58 - 52)^2}{52} + \frac{(44 - 49,99)^2}{49,99} = 2,84$$

Pentru interpretare, hi-pătrat calculat trebuie comparat cu hi-pătrat critic dat de tabel. În acest scop, trebuie calculate gradele de libertate. Datorită încrucișării celor două variabile, se încrucișează și gradele lor de libertate. Altfel spus, $gdl = (\text{numărul de modalități al primei variabile} - 1) \times (\text{numărul de modalități al celei de-a doua variabile} - 1)$. Sau, generalizând potrivit tabelului de contingență:

$$gdl = (\text{numărul de coloane} - 1) \times (\text{numărul de linii} - 1)$$

Aici, $gdl = [(2 - 1) \times (2 - 1)] = 1$. La pragul de eroare de 5% (0.05), hi-pătrat critic = 3,84. Hi-pătrat calculat fiind inferior acestei valori, suntem obligați să păstrăm ipoteza nulă și să concluzionăm că între cele două variabile există o relație de independență. Tânărul cercetător nu poate afirma că răspunsurile bărbaților diferă semnificativ de cele ale femeilor. Diferențele observate pot fi atribuite unor simple fluctuații de eșantionare.

APLICAȚIE

Directorul unei întreprinderi de produse lactate a cerut unei agenții de publicitate să propună un nume nou pentru un iaurt degresat care se va lansa pe piață. În acest scop, agenția îi chestionează direct pe consumatorii potențiali, propunându-le să aleagă un nume dintre patru posibile, după ce au gustat din iaurt. Fiindcă produsul vizează cu predilecție clientela feminină, agenția testează omogenitatea preferințelor asupra numelui, în funcție de clasa de vârstă. Pe un eșantion de 550 de femei, se definesc trei clase de vârstă.

	Alex	Babu	Daria	Maria
25-35 ani	28	13	36	84
35-50 ani	47	21	52	103
peste 50 ani	32	18	47	69

În prima etapă, se determină efectivele marginale pe linie și pe coloană:

	Alex	Babu	Daria	Maria	Σ linii
25-35 ani	28	13	36	84	161
35-50 ani	47	21	52	103	223
peste 50 ani	32	18	47	69	166
Σ coloane	107	52	135	256	550

Se determină apoi efectivele teoretice, aplicând pentru fiecare căsuță formula de mai sus. Rezultatele sunt prezentate în tabelul următor:

	Alex	Babu	Daria	Maria	Σ linii
25-35 ani	31,32	15,22	39,52	74,94	161
35-50 ani	43,39	21,08	54,74	103,79	223
Peste 50 ani	32,29	15,70	40,74	77,27	166
Σ coloane	107	52	135	256	550

Nu rămâne de calculat decât hi-pătrat, măsurând pentru fiecare căsuță din tabel distanța între efectivele observate și efectivele teoretice.

$$\chi^2 = \frac{(28 - 31,32)^2}{31,32} + \frac{(13 - 15,22)^2}{15,22} + \frac{(36 - 39,52)^2}{39,52} + \frac{(84 - 74,94)^2}{74,94} +$$

$$+ \frac{(47 - 43,39)^2}{43,39} + \frac{(21 - 21,08)^2}{21,08} + \frac{(52 - 54,74)^2}{54,74} + \frac{(103 - 103,79)^2}{103,79} +$$

$$+ \frac{(32 - 32,29)^2}{32,29} + \frac{(18 - 15,70)^2}{15,70} + \frac{(47 - 40,74)^2}{40,74} + \frac{(69 - 77,27)^2}{77,27} = 4,71$$

Pentru gradul de libertate $gdl = [(4 - 1) \times (3 - 1)] = 6$, tabelul generează, cu un risc de 5%, un hi-pătrat critic de 12,59. Hi-pătrat calculat fiind inferior acestei valori, suntem obligați să păstrăm ipoteza nulă. Preferințele pentru cele patru nume sunt independente de clasa de vârstă. Altfel spus, în general femeile chestionate au arătat aceleași preferințe.

P.04 Frecvențele calculate pe eșantioane perechi

Dacă două variabile identice sunt supuse aceluiași subiecți, avem un caz de împerechere (două sarcini similare supuse aceluiași indivizi, două momente de evaluare distincte cu același instrument de măsură etc.). Scopul este compararea perechilor de modalități divergente, și numai a acestora, pentru a evita repetițiile. În acest scop, se folosește o formulă care nu verifică decât distanța între perechile divergente.

Atenție însă: acest test, numit hi-pătrat al lui Mac Nemar, nu poate fi aplicat decât în cazul încrucișării a două variabile cu două modalități (tabel de contingență 2×2). În celelalte cazuri, se folosește formula tradițională pentru hi-pătrat, știind că împerecherea nu se ia în considerare. Este așadar de preferat, să se recurgă la descompuneri (mai multe teste succesive) sau la regrouparea modalităților. Fie următorul tabel de contingență:

		Variabila 1	
		Modalitatea 1	Modalitatea 2
Variabila 2	Modalitatea 1	A	B
	Modalitatea 2	C	D

Scopul lui hi-pătrat al lui Mac Nemar constă în evaluarea distanței între perechile divergente. Cu alte cuvinte, pentru calcularea lui hi-pătrat se iau în considerare numai aceste perechi. Referindu-ne la căsuțele tabelului de mai sus, se obține:

$$\chi^2_{\text{MacNemar}} = \frac{(B - C)^2}{B + C}$$

Valoarea obținută se compară cu cea din tabelul hi-pătrat pentru gradul de libertate $gdl = (2 - 1)(2 - 1) = 1$;

- Dacă hi-pătrat calculat este inferior lui hi-pătrat critic, ipoteza nulă se păstrează. Diferența între cele două frecvențe nu este semnificativă.

- Dacă hi-pătrat calculat este mai mare sau egal cu hi-pătrat critic, ipoteza nulă se respinge. Diferența între cele două frecvențe este semnificativă.

APLICAȚIE

Într-un studiu de docimologie, se analizează rezultatele la admiterile de la Politehnică și de la Facultatea de Litere. Nu se iau în considerare decât rezultatele celor 300 de candidați care s-au prezentat la ambele admiteri: 60 au fost admiși numai la Facultatea de Litere, 44 numai la Politehnică și 16 la amândouă. Se poate trage concluzia că ambele examene sunt la fel de dificile?

În primul rând, trebuie reconstituit tabelul de contingență:

		Politehnică	
		Admis	Respins
Facultatea de Litere	Admis	16	60
	Respins	44	180

Se înțelege că este inutil să comparăm candidații admiși la ambele examene cu cei respinși la ambele. Singura comparație care interesează este cea între candidații admiși la Facultatea de Litere și cei respinși la Politehnică, sau cea între candidații admiși la Politehnică și cei respinși la Facultatea de Litere. Aplicarea testului hi-pătrat a lui Mac Nemar ne generează rezultatul:

$$\chi^2 \text{ MacNemar} = \frac{(60 - 44)^2}{60 + 44} = \frac{256}{104} = 2,46$$

Pentru gradul de libertate $gdl = (2 - 1)(2 - 1) = 1$, tabelul dă, cu un risc de 5%, un hi-pătrat critic de 3,84. Hi-pătrat calculat fiind inferior acestei valori, suntem obligați să păstrăm ipoteza nulă. Cu alte cuvinte, ambele examene sunt la fel de dificile.

LIMITELE FOLOSIRII ȘI INTERPRETĂRII TESTULUI HI-PĂTRAT

P.05

Necesitatea unui număr minim de observații

Valoarea lui hi-pătrat depinde mult de numărul de observații efectuate. Dacă numărul este mic, riscă să falsifice valoarea reală a distanței între efectivele observate și efectivele teoretice. Acest test nu este aplicabil atunci când unul dintre efectivele teoretice este mai mic decât 1. În plus, trebuie ca efectivele tuturor căsuțelor din tabelul de contingență să fie mai mari sau egale cu 5. În caz contrar, crește riscul de a mări artificial valoarea lui hi-pătrat calculat și putem ajunge la concluzii greșite.

Corecțiile aduse testului

Dacă unul dintre efectivele observate este mai mic decât 10, în formula hi-pătrat trebuie introdusă o corecție: corecția lui Yates.

- În cazul eșantioanelor independente, formula corectată este:

$$\chi^2 \text{ corectat} = \sum \frac{[|\text{Efectivul observat} - \text{Efectivul teoretic}| - 0,5]^2}{\text{Efectivul teoretic}}$$

- În cazul eşantioanelor perechi, testul hi-pătrat Mac Nemar corectat este:

$$\chi^2_{\text{Mac Nemar corectat}} = \sum \frac{|B - C| - 1}{B + C}$$

Limitele interpretării testului hi-pătrat

Testul nu este eficient atunci când vrem să analizăm variabile cu număr mare de modalități: cu cât numărul de încrucișări ale modalității crește, cu atât interpretarea este mai dificilă. În cazul unui tabel de contingență cu un număr foarte mare de căsuțe, avem la dispoziție două soluții: fie reunim modalitățile sau facem comparații descompuse pe perechi, care au dezavantajul că reduc informația sau multiplică ipotezele, fie recurgem la tehnici de analiză mai complexe, care, pentru a fi prezentate, ar necesita o întreagă lucrare (de exemplu, analiza factorială a corespondențelor).

EXEMPLU DE UTILIZARE A TESTULUI HI-PĂTRAT Folosind pachetul SPSS

În general, testul hi-pătrat este folosit pentru evaluarea existenței unei diferențe semnificative între două sau mai multe eşantioane care sunt formate din date de frecvență (date nominale). Cu alte cuvinte, este testul statistic uzual care analizează tabelele de asociere sau contingență pe baza a două variabile categoricale nominale.

De asemenea, poate fi folosit și pentru testarea faptului că un singur eşantion este semnificativ diferit de o populație cunoscută. Această aplicație este cea mai puțin comună, deoarece caracteristicile unei populații sunt rareori cunoscute în cercetare.

Este extrem de important de reținut că testul hi-pătrat analizează frecvențe. Acestea nu trebuie **niciodată** să fie transformate în **procente** pentru a fi introduse în SPSS, deoarece vor da rezultate eronate atunci când se vor calcula valoarea și semnificația lui hi-pătrat. Trebuie să existe o distincție foarte clară între acest lucru și folosirea procentelor atunci când încercăm să interpretăm ce se întâmplă în cadrul unui tabel de contingență.

De asemenea, trebuie să avem grijă că o analiză hi-pătrat trebuie să includă datele fiecărui individ o singură dată. Așadar, frecvențele totale trebuie să fie egale cu numărul de persoane folosite în analiză.

Analizarea și interpretarea tabelelor de contingență 2x2 sunt procese directe. Interpretarea unor tabele de contingență mai mari nu este la fel de ușoară, deoarece poate necesita împărțirea tabelului într-o serie de tabele mai mici. Partiționarea testului hi-pătrat, așa cum este cunoscut procedeul, necesită de obicei adaptarea la nivelele de semnificație pentru a lua în considerare și numărul de subanalize efectuate.

Atunci când datele nu îndeplinesc condițiile pentru hi-pătrat (în special atunci când frecvențele anticipate sunt prea mici) se folosește testul Fisher. Este discutat și testul Mac Nemar pentru semnificația modificărilor, care se află în strânsă legătură cu hi-pătrat.

Unele variante ale testului hi-pătrat sunt folosite ca măsuri ale parametrului **goodness-of-fit** în unele dintre cele mai avansate tehnici statistice cum ar fi regresia

TESTUL DE CORELAȚIE LINEARĂ SAU TESTUL BRAVAIS-PEARSON

Folosirea acestui test se face cu următoarele condiții:

- datele trebuie să provină din scale de intervale sau de raport;
- legătura între distribuții trebuie să fie tendențial lineară.

Dacă ambele condiții sunt simultan reunite, se poate aplica testul coeficientului de corelație Bravais-Pearson (notat cu r), a cărui formulă este următoarea:

$$r = \frac{\sum [(Valorile \text{ es.1} - Media \text{ es.1}) \cdot (Valorile \text{ es.2} - Media \text{ es.2})]}{\sqrt{\sum (Valorile \text{ es.1} - Media \text{ es.1})^2 \cdot \sum (Valorile \text{ es.2} - Media \text{ es.2})^2}}$$

Valoarea calculată pentru r se interpretează în comparație cu valoarea r critică dată de tabelul legii r Bravais-Pearson pentru gradul de libertate egal cu (numărul de cupluri - 2):

- dacă r calculat < r critic, ipoteza nulității corelației se păstrează, între cele două variabile nu există corelație;
- dacă r calculat > r critic, ipoteza nulă se respinge. Cele două variabile sunt corelate. În acest caz, trebuie să facem raportarea la semnul lui r calculat pentru a determina sensul corelației, pozitiv sau negativ.

APLICAȚIE. Un psiholog școlar vrea să afle dacă există o legătură între capacitatea de raționament a liceenilor și performanțele lor școlare. În acest scop, administrează un test de raționament (notat de la 0 la 10) unui eșantion de 15 elevi și studiază eventuala legătură între scorurile obținute și mediile generale anuale ale acestor elevi.

Elevi	Nota la test	Media generală
1	6	18
2	2	4, 5
3	5	13
4	4	9, 5
5	4	12, 25
6	6	16
7	5	14
8	3	7, 5
9	4	12, 5
10	5	10
11	3	9, 5
12	4	12
13	6	14, 5
14	4	11
15	5	10, 25

TESTE NEPARAMETRICE. COMPARAREA A DOUĂ EȘANTIOANE INDEPENDENTE. TESTUL MEDIANEI. TESTUL "U" MANN-WHITNEY

Majoritatea testelor statistice precedente, cer ca distribuțiile să fie normale. Când însă nu este cazul, se preferă structuri mai puțin rigide ale scalelor de măsură și se aplică un test **neparametric**.

Principala problemă a acestor teste constă în pierderea informației: se lucrează pe ranguri sau semne, și nu pe datele brute. Deși sunt adesea desconsiderate în științele umaniste, majoritatea testelor sunt robuste și fiabile.

Vom analiza succesiv mai multe teste neparametrice, adaptate tipului de organizare a datelor culese, și în special, caracterului independent sau pereche al măsurătorilor. Pentru fiecare caz vom prezenta două teste, dintre care al doilea este întotdeauna cel mai eficient, dar nu întotdeauna realizabil.

COMPARAREA A DOUĂ EȘANTIOANE INDEPENDENTE TESTUL MEDIANEI

Când poate fi folosit ?

- Când datele s-au prelevat cu ajutorul unei scale de intervale și distribuția lor nu este normală. În acest caz, studiarea mediilor distribuțiilor nu are sens. Preferăm să studiem mediana lor.

- De exemplu, când analizăm relația între sexul elevilor și performanțele lor școlare, se poate întâmpla ca un număr de „supradotați” cu note de 20/20 să influențeze exagerat media generală și abaterea standard. Poate fi deci mai pertinent să luăm în considerare performanțele numai ca fiind inferioare sau superioare medianei generale a clasei.

- Când datele s-au prelevat cu ajutorul unei scale ordinale.

Demers

- Se reunesc valorile celor două eșantioane e_1 și e_2 pentru a constitui un nou eșantion E de mărime $N = n_1 + n_2$.

- Se calculează mediana noului eșantion E .

- Se repartizează valorile lui e_1 și e_2 în funcție de poziția lor față de mediană; pe de o parte, valorile inferioare medianei, iar de cealaltă parte, valorile superioare medianei.

- Se constituie tabelul de contingență, încrucișând în linie eșantioanele și în coloană cele două modalități ale variabilei recodate.

Se obține astfel un tabel de tipul următor:

	< mediana	> mediana	$\sum$ linii
e_1	n_1	n_2	$n_1 + n_2$
e_2	n_3	n_4	$n_3 + n_4$
$\sum$ coloane	$n_1 + n_3$	$n_2 + n_4$	Total

Ipoteza nulă pe care o vom testa este următoarea: „Frecvența valorilor superioare mediane în el este egală cu frecvența valorilor superioare mediane în e2”. Fie $H_0 = \frac{n2}{n1 + n2} - \frac{n4}{n3 + n4} = 0$.

Se constată că am ajuns astfel la compararea a două frecvențe calculate pe eșantioane independente, și se aplică deci testul Hi-doi:

$$\chi^2 = \sum \frac{(\text{Efectivul observat} - \text{Efectivul teoretic})^2}{\text{Efectivul teoretic}}$$

Se știe că informația esențială pentru determinarea distribuției efectivelor teoretice se află în totalurile marginale ale liniilor și coloanelor din tabelul de contingență, adică în așa-numitele efective marginale. Metoda de calcul a efectivelor teoretice constă, pentru fiecare căsuță din tabel, în efectuarea produsului sumei liniilor și a sumei coloanelor la intersecția cărora se află căsuța, iar produsul se împarte la efectivul total. Cu alte cuvinte, determinarea efectivului teoretic al fiecărei căsuțe din tabel se obține cu formula:

(Efectiv marginal pe linie al căsuței x Efectiv marginal pe coloană al căsuței) / Efectiv total

Hotărârea de respingere sau de păstrare a ipotezei nule depinde simultan de faptul că hi-pătrat calculat este mai mic, egal sau mai mare decât hi-pătrat critic dat de tabel, de pragul de eroare fixat și de numărul de grade de libertate, care aici este $gdl = (2 - 1)(2 - 1) = 1$.

Astfel, pragul de eroare definit:

- dacă hi-pătrat calculat este inferior lui hi-pătrat critic, ipoteza nulă se păstrează. Diferența între cele două frecvențe nu este semnificativă;
- dacă hi-pătrat calculat este mai mare sau egal cu hi-pătrat critic, ipoteza nulă se respinge. Diferența între cele două frecvențe este semnificativă.

Dacă unul dintre efectivele observate este mai mic decât 10, trebuie aplicată și corecția lui Yates:

$$\chi^2_{\text{corectat}} = \sum \frac{[|\text{Efectiv observat} - \text{Efectiv teoretic}| - 0,5]^2}{\text{Efectiv teoretic}}$$

Observație. În cazul în care avem valori egale cu mediana, se construiesc două tabele de contingență:

	$\leq$ mediana	$>$ mediana
e1		
e2		

	$<$ mediana	$\geq$ mediana
e1		
e2		

Se calculează H_i -doi pe tabelul în care riscul de eroare este cel mai mic, știind că H_i -doi este cu atât mai mic cu cât diferența între $\frac{n_2}{n_1 + n_2}$ și $\frac{n_4}{n_3 + n_4}$ este mai mică.

APLICAȚIE

Un profesor emite ipoteza că reușita la un examen diferă în funcție de tipul examenului (oral sau scris) impus studenților lui, în număr de 102. Mediana notelor se situează la 10,5 și împarte efectivul în două părți de 48 și 54 de studenți.

Încrucișăm cele două variabile într-un tabel de contingență:

	Valori < 10,5	Valori > 10,5	$\sum$ linii
Scris	18	33	51
Oral	30	21	51
$\sum$ coloane	48	54	102

Pentru determinarea efectivelor teoretice se aplică pentru fiecare căsuță formula:

(Efectiv marginal pe linie al căsuței x Efectiv marginal pe coloană al căsuței) / Efectivul total. Rezultatele se află în următorul tabel:

	Valori < 10,5	Valori > 10,5
Scris	24	27
Oral	24	27

Calculăm acum hi-pătrat:

$$\chi^2 = \frac{(18 - 24)^2}{24} + \frac{(33 - 27)^2}{27} + \frac{(30 - 24)^2}{24} + \frac{(21 - 27)^2}{27} = 5,66$$

Pentru gradul de libertate $gdl = 1$, tabelul dă, cu un risc de 5%, un hi-pătrat critic de 3,84. Hi-pătrat calculat este superior acestei valori, deci respingem ipoteza nulă. Tipul de examen a avut într-adevăr un efect. Tabelul de contingență arată că studenții cu note mai mari decât mediana generală sunt semnificativ mai numeroși în condiția „examen scris” decât în celelalte.

TESTUL „U” MANN-WHITNEY

P.08

Când poate fi folosit ?

- Când datele s-au prelevat cu ajutorul unei scale de intervale, însă distribuția lor nu este normală. În acest caz, studierea mediilor distribuțiilor nu are sens. Preferăm să studiem mediana lor.
- De exemplu, când analizăm relația între sexul elevilor și performanțele lor școlare, se poate întâmpla ca doi „supradotați” cu note de 20/20 să influențeze

exagerat media generală și abaterea standard. În schimb, dacă notele se transformă în ranguri, cei doi vor fi clasati primul și al doilea, și nu se ia în considerare distanța față de nota celui de-al treilea.

- Când datele se prezintă sub forma unui clasament.

Demers

Se reunesc valorile celor două eșantioane e_1 și e_2 , pentru a constitui un nou eșantion E de mărime $N = n_1 + n_2$. Se atribuie un rang fiecărei valori. Transformarea în rang a valorilor *ex aequo* (egale între ele) se efectuează astfel. Dacă patru indivizi sunt clasati toți în poziția a treia, nu li se atribuie rangul 3 tuturor. Se consideră că, dacă nu ar fi avut exact aceeași notă, ar fi ocupat rangurile 3, 4, 5 și 6.

Li se atribuie deci rangul mediu corespunzător, adică $(3 + 4 + 5 + 6)/4 = 4,5$. În eșantioanele inițiale se înlocuiește fiecare valoare prin rangul ei. Pentru fiecare eșantion, se calculează suma rangurilor: T_1 pentru eșantionul e_1 și T_2 pentru eșantionul e_2 .

Ipoteza nulă testată este aceea potrivit căreia suma rangurilor eșantionului 1 este egală cu suma rangurilor eșantionului 2, adică $H_0 = T_1 - T_2 = 0$. Putem avea două situații:

a. n_1 și/sau n_2 sunt mai mici sau egale cu 20. În acest caz, Mann și Whitney au construit un tabel al valorilor T . T este suma rangurilor eșantionului cu efectivul cel mai mic.

Tabelul dă în linie n_1 (adică mărimea posibilă a lui e_1 , acesta fiind întotdeauna eșantionul cu efectivul cel mai mic), iar în coloană n_2 (adică mărimea posibilă a lui e_2 , adică eșantionul cu efectivul cel mai mare). În corpul tabelului se obțin limitele între care trebuie să fie cuprins T pentru ca H_0 să fie acceptată.

b. n_1 și/sau n_2 sunt mai mari decât 20. Atunci trebuie calculată:

$$U_1 = (n_1 n_2) + \left[\frac{n_1 (n_1 + 1)}{2} \right] - T_1$$

și

$$U_2 = (n_1 n_2) + \left[\frac{n_2 (n_2 + 1)}{2} \right] - T_2$$

Se păstrează valoarea cea mai mică dintre cei doi indici. Acest indice se distribuie aproximativ după o lege normală de medie și abatere standard:

$$m_U = \frac{n_1 n_2}{2} \quad \text{și} \quad \sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

Indicele U se poate transforma așadar în variabilă normală centrată și redusă cu formula: $z = \frac{x - m}{\sigma}$, care devine:

$$z = \frac{U_1 \text{ sau } U_2 - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

Pentru interpretarea acestei valori, vom face raportarea la tabelul legii normale centrate și reduse (pentru prezentarea legii normale, a legii normale centrate și reduse și a procedurii de citire a tabelului):

- dacă frecvența calculată este mai mare decât 0.05, se estimează că riscul de eroare este prea mare pentru a respinge ipoteza H_0 . Se consideră deci că $T_1 - T_2 = 0$, adică cele două eșantioane nu diferă semnificativ;
- dacă frecvența calculată este mai mică sau egală cu 0.05, ipoteza H_0 se respinge și se consideră că cele două eșantioane diferă semnificativ.

APLICAȚIE

Un profesor emite ipoteza că există diferențe în reușita la un examen, în funcție de tipul examenului (parțiale, notate G1, sau examen final, notat G2) impus celor 100 de studenți ai lui. Efectivele obținute pentru fiecare tip de examen sunt:

Note	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Efectiv G1	0	2	3	7	2	5	1	4	3	5	10	2	1	5
Efectiv G2	3	1	1	7	7	11	3	5	3	2	4	0	1	2

Începem cu reunirea efectivelor celor două grupuri:

Note	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Efectiv G1+ Efectiv G2	3	3	4	14	9	16	4	9	6	7	14	2	2	7

Se atribuie un rang fiecărei valori în funcție de efectivul ei:

Nota	Rang
5	2
6	5
7	8,5
8	17,5
9	29
10	41,5
11	51,5
12	58
13	65,5
14	72
15	82,5
16	90,5
17	92,5
18	97

Ne întoarcem la grupurile inițiale și înlocuim fiecare valoare cu rangul ei (vezi tabelul care urmează). Suma rangurilor pentru fiecare grup:

$$T1 = 2847 \text{ și } T2 = 2203$$

n_1 și n_2 fiind mai mari decât 20, se calculează U_1 și U_2 :

$$U_1 = (50 \cdot 50) + \left[\frac{50(50+1)}{2} \right] - 2847 = 928$$

$$U_2 = (50 \cdot 50) + \left[\frac{50(50+1)}{2} \right] - 2203 = 1572$$

Pentru a-l calcula pe z , se păstrează cea mai mică valoare dintre ele, adică U_2 :

$$z = \frac{928 - \frac{(50 \cdot 50)}{2}}{\sqrt{\frac{(50 \cdot 50)(50 + 50 + 1)}{12}}} = -2,22$$

Efectiv x Rang (G1)	Efectiv x Rang (G2)
0 x 2	3 x 2
2 x 5	1 x 5
3 x 8,5	1 x 8,5
7 x 17,5	7 x 17,5
2 x 29	7 x 29
5 x 41,5	11 x 41,5
1 x 51,5	3 x 51,5
4 x 58	5 x 58
3 x 65,5	3 x 65,5
5 x 72	2 x 72
10 x 82,5	4 x 82,5
2 x 90,5	0 x 90,5
1 x 92,5	1 x 92,5
5 x 97	2 x 97

Tabelul legii normale centrate și reduce indică, pentru $z = 2,22$, frecvența de 0.02642, care este mai mică decât 0.05. Putem deci respinge ipoteza H_0 și considera că tipul de examen are efect asupra rezultatelor. Comparând suma rangurilor celor două grupuri, se constată că parțialele sunt mai benefice pentru studenți (suma rangurilor este mai mare).

Valori	Efectiv	Efectiv cumulat crescător	Efectiv cumulat descrescător
8	5	5	40
9	6	11	35
10	4	15	29
11	2	17	25
12	6	23	23
13	8	31	17
14	3	34	9
15	5	39	6
16	1	40	1

Pentru efectivele cumulate crescătoare, $N/2 = 20$ se situează între pozițiile 17 și 23, adică între valorile 11 și 12. Deci: $11 + (1 \times 3)/7 = 11,42$. Pentru efectivele cumulate descrescătoare, 20 se situează între pozițiile 17 și 23, adică între valorile 13 și 12. Deci: $13 - (1 \times 3)/7 = 12,57$. Așadar mediana distribuției este: $(11,42 + 12,57)/2 = 11,99$. Se repartizează valorile celor 4 eșantioane de o parte și de cealaltă a acestei mediane:

	< 11,99	> 11,99	Σ linii
Artă modernă	5	5	10
Artă contemporană	5	5	10
Artă medievală	3	7	10
Artă antică	4	6	10
Σ coloane	17	23	40

Efectivele teoretice, determinate cu ajutorul formulei convenționale, sunt:

	< 11,99	> 11,99
Artă modernă	4,25	5,75
Artă contemporană	4,25	5,75
Artă medievală	4,25	5,75
Artă antică	4,25	5,75

Rămâne de calculat hi-pătrat. Efectivele fiind mai mici decât 10, trebuie aplicată corecția lui Yates:

$$\chi^2 = \frac{(|5 - 4,25| - 0,5)^2}{4,25} + \frac{(|5 - 5,75| - 0,5)^2}{5,75} + \frac{(|5 - 4,25| - 0,5)^2}{4,25} +$$

$$+ \frac{(|5 - 5,75| - 0,5)^2}{5,75} + \frac{(|3 - 4,25| - 0,5)^2}{4,25} + \frac{(|7 - 5,75| - 0,5)^2}{5,75} +$$

$$+ \frac{(|4 - 4,25| - 0,5)^2}{4,25} + \frac{(|6 - 5,75| - 0,5)^2}{5,75} = 0,30$$

La gradul de libertate $(2 - 1)(4 - 1) = 3$, tabelul Hi-doi ne dă, pentru pragul de 0.05, un hi-pătrat critic de 7,82, cu mult superior valorii lui hi-pătrat calculat. Se trage concluzia că eșantioanele sunt echivalente și că, prin urmare, nu există o legătură între specializarea aleasă și rezultatele studenților la cursurile obligatorii comune.

TESTUL KRUSKAL-WALLIS

P.10

Când poate fi folosit ?

- Când datele s-au prelevat cu ajutorul unei scale de intervale, însă distribuția lor nu este normală. În acest caz, studierea mediilor distribuțiilor nu are sens. De exemplu, dacă analizăm relația între categoria socio-profesională a taților elevilor și performanțele școlare ale acestora, se poate ca doi „supradotați” să aibă numai nota 20/20 și să influențeze exagerat media și abaterea standard. În schimb, dacă notele se transformă în ranguri, cei doi vor fi clasati primul și al doilea și nu se ia în considerare distanța față de nota celui de-al treilea.

- Când datele se prezintă sub forma unui clasament.

Demers

- Se reunesc valorile celor două eșantioane e_1 și e_2 și se constituie un nou eșantion E de mărime $N = n_1 + n_2 + \dots + n_k$.

- Se atribuie un rang fiecărei valori. Transformarea în rang a valorilor *ex aequo* (egale între ele) se efectuează astfel. Dacă patru indivizi sunt clasati toți în poziția a treia, nu li se atribuie rangul 3 tuturor. Se consideră că, dacă nu ar fi avut exact aceeași notă, ar fi ocupat rangurile 3, 4, 5 și 6. Li se atribuie deci rangul mediu corespunzător, adică $(3 + 4 + 5 + 6) / 4 = 4,5$.

- În eșantioanele inițiale, se înlocuiește fiecare valoare prin rangul ei.

- Pentru fiecare eșantion, se calculează suma rangurilor T_j .

- Ipoteza nulă testată este aceea potrivit căreia suma rangurilor eșantionului 1 este egală cu suma rangurilor eșantionului 2, adică $H_0 = T_1 = T_2 = \dots = T_k$. Pentru a testa această ipoteză, Kruskal și Wallis au definit variabila H astfel încât:

$$H = \left(\frac{12}{N(N+1)} \cdot \sum \frac{T_j^2}{n_j} \right) - 3(N+1)$$

unde, 12 și 3 sunt constante; N = efectivul total; n_j = efectivul unui eșantion oarecare; T_j = suma rangurilor unui eșantion oarecare. H se interpretează diferit în funcție de două cazuri:

a) $k = 3$ și toate $n_j < 5$

Se folosește atunci tabelul H Kruskal și Wallis în a cărei primă coloană se găsește mărimea fiecărui eșantion, în a treia coloană găsim pragul de probabilitate, iar în a doua coloană, valorile critice ale lui H peste care se respinge ipoteza nulă. Astfel:

- dacă $H \text{ calculat} > H \text{ critic}$, se respinge ipoteza nulă, ceea ce înseamnă că, la pragul de încredere considerat, clasamentele diferă semnificativ;
- dacă $H \text{ calculat} < H \text{ critic}$, se păstrează ipoteza nulă, ceea ce înseamnă că nu putem considera clasamentele statistic diferite.

b) $k > 3$ și/sau unul dintre $n_j > 5$

În acest caz, indicele H urmează aproximativ legea Hi-doi. Va fi deci interpretat cu ajutorul tabelului Hi-doi, în funcție de pragul de probabilitate ales și de gradul de libertate $gdl = k - 1$:

- dacă $H \text{ calculat} > H\text{-pătrat critic}$, se respinge ipoteza nulă, ceea ce înseamnă că, clasamentele diferă semnificativ;
- dacă $H \text{ calculat} < H\text{-pătrat critic}$, se păstrează ipoteza nulă, ceea ce înseamnă că nu putem considera clasamentele statistic diferite.

APLICAȚIE. Se reia aplicația precedentă în care un profesor de istoria artei se întreabă dacă există o legătură între specializarea pe care o alege 40 de studenți de anul III (în artă modernă, contemporană, medievală sau antică) și rezultatele lor la cursurile obligatorii comune. Notele obținute la cursurile obligatorii comune sunt:

- studenții la artă modernă: 15, 8, 14, 10, 10, 12, 11, 13, 14, 9;
- studenții la artă contemporană: 9, 10, 11, 13, 13, 13, 12, 15, 8, 8;
- studenții la artă medievală: 8, 9, 9, 12, 12, 12, 13, 14, 15, 15;
- studenții la artă antică: 9, 10, 16, 13, 13, 13, 12, 15, 8, 9.

Se reunesc valorile celor 4 eșantioane și se distribuie fiecăreia un rang în funcție de efectiv:

Valori	8	9	10	11	12	13	14	15	16
Efectiv	5	6	4	2	6	8	3	5	1
Rang	3	8,5	13,5	16,5	20,5	27	33	37	40

Ne raportăm apoi la eșantioanele inițiale și înlocuim fiecare valoare cu rangul ei. Calculăm apoi suma rangurilor fiecărui eșantion.

Artă modernă	Artă contemporană	Artă medievală	Artă antică
3	3	3	3
8,5	3	8,5	8,5
13,5	8,5	8,5	8,5
13,5	13,5	20,5	13,5
16,5	16,5	20,5	20,5
20,5	20,5	20,5	27,5
27,5	27,5	27,5	27,5

33	27,5	33	27,5
33	27,5	37	37
37	37	37	40
T1 = 206	T2 = 184,5	T3 = 216	T4 = 213,5

Pentru a aplica formula lui Kruskal și Wallis, trebuie calculată suma $\sum \frac{T_j^2}{n_j}$.

$$\sum \frac{T_j^2}{n_j} = \frac{206^2}{10} + \frac{184,5^2}{10} + \frac{216^2}{10} + \frac{213,5^2}{10} = 16871,45$$

Se aplică apoi formula lui H:

$$H = \left(\frac{12}{40 \cdot 41} \cdot 16871,45 \right) - (3 \cdot 41) = 0,44$$

Ne aflăm în cazul în care $k > 3$ și în care toate eșantioanele au un efectiv peste 5. H se interpretează deci cu ajutorul tabelului Hi-doi, la gradul de libertate $gdl = (4 - 1) = 3$. Pentru pragul de 0.05, Hi-doi critic = 7,82. H calculat fiind inferior acestei valori, ipoteza nulă se păstrează și tragem concluzia că nu există o legătură statistică între specializarea aleasă și rezultatele studenților la cursurile obligatorii comune.