Suport Curs 7 – Testarea ipotezelor
Lect. dr. Adrian Gorbănescu

7.1 Ipoteza cercetării și ipoteza de nul

Testarea ipotezelor reprezintă o metodă prin care verificăm afirmație sau o presupunere despre parametrul unei populații utilizând date măsurate la nivelul unui eșantion extras din acea populație. Prin această metodă determinăm care este probabilitatea ca afirmația făcută despre populație să fie corectă pornind de la analiza datelor obținute în eșantion. În concluzie, testarea ipotezei reprezintă o modalitate prin care încercăm să confirmăm sau să infirmăm o presupunere despre populație folosind date extrase de la nivelul eșantionului.

Punctul final al testării unei ipoteze este acela de a lua o decizie. În domeniul cercetării psihologice recurgem la testarea ipotezelor deoarece ea ne poate indica un anumit nivel de incertitudine atunci când luăm o decizie. Dacă descoperim, după colectarea datelor, că nivelul de incertitudine este unul scăzut, putem considera că decizia pe care o luăm are un nivel înalt de încredere. Dacă nivelul de incertitudine este mare atunci trebuie luată o altă decizie: fie adunăm mai multe date, fie abandonăm deoarece designul cercetării nu a fost corect. În paginile ce vor urma de acum înainte se va face presupunerea că cititorul a înțeles conceptele referitoare la testarea ipotezelor. Prin urmare, noțiunile și informațiile din acest capitol sunt importante pentru a înțelege testele statistice explicate în capitolele ce vor urma. Fiecare dintre acestea vizează același aspect: testarea unei ipoteze referitoare la una sau mai multe populații.

Să ne imaginăm că ne dorim să cumpărăm înghețată, iar compania producătoare, BI, susține că o cupă de înghețată are 150g. În calitate de cumpărători, ne putem întreba dacă chiar primim cantitatea menționată de vânzător. În această situație, BI susține că media populației este μ = 150g.

Bineînțeles, putem cumpăra o cupă de înghețată, să o cântărim și să o comparăm cu 150. Să presupunem că în urma cântăririi observăm că am primit o cupă care cântărește mai puțin de 150g. A concluziona că „BI își păcălește clienții, cupele de înghețată cântărind sub 150g” nu este corect, deoarece am avea nevoie de mai multe măsurători. Pe baza statisticii inferențiale vom putea trage câteva concluzii referitoare la media populației:

- Cupele de înghețată pe care le primesc clienții cântăresc mai puțin decât susține firma BI.
- Cupele de înghețată au o greutate apropiată de cea asumată de firmă, angajații încercând să pună 150g în fiecare cupă.
- Cupele de înghețată cântăresc mai mult decât susține firma.

Pentru a lua o decizie cu privire la cantitatea de înghețată, vom cântări cupele de înghețată vândute unui eșantion de 40 de clienți. După cântărirea celor 40 de cupe de înghețată obținem m = 153,95g și s = 8,32g. După cum se poate observa, media eșantionului este mai mare decât cantitatea propusă de companie, 150g.

Într-o astfel de cercetare ne bazăm pe estimarea unui rezultat așteptat: greutatea cupelor de înghețată primite de clienți este diferită de media populației (150g). Acest rezultat se numește **ipoteza cercetării (H_1)**. Conform Rumsey (2010) ipoteza cercetării poate fi formulată astfel:

1. Greutate unei cupe de înghețată cumpărate de clienți este diferită de 150g. Într-o astfel de formulare nu se specifică dacă cantitatea unei cupe primite de client este mai mare sau mai mică, cercetătorul admițând posibilitatea ca greutatea unei cupe să fie atât mai mare de 150g, dar și mai mică (decizie bilaterală, lb. engleză **two-tailed**).
2. Greutatea unei cupe de înghețată este mai mică de 150g. De data aceasta cercetătorul consideră că greutatea cupelor este mai mică (decizie unilaterală, lb. engleză **one-tailed**).

3. Greutatea unei cupe de înghețată este mai mare de 150g. În această situație se admite doar posibilitatea ca înghețata cumpărată de clienți să aibă o greutate peste 150g (decizie unilaterală).

Pentru a găsi un răspuns la problema formulată prin situația de mai sus ne vom folosi de un eșantion de clienți, deoarece este imposibil să cântărim greutatea de înghețată primită de fiecare cumpărător. Am observat mai sus că ipoteza cercetării este scrisă ca o afirmație care prezintă o relație (în cazul nostru, diferență) între eșantion și populație. Astfel, pot exista două tipuri de relație între eșantion și populație:

1. Eșantionul face parte din populație
2. Eșantionul nu face parte din populație, fiind diferit de aceasta (afirmație făcută de ipoteza cercetării).

La prima vedere, greutatea cupelor de înghețată primite de clienți pare să fie mai mare decât cea asumată de firmă. Totuși, cercetarea științifică ne solicită să conștientizăm faptul că generalizarea mediei unui eșantion asupra unei populații implică anumite riscuri. Noi am ales un eșantion de 40 de clienți, acesta fiind doar unul din multitudinea de eșantioane care ar fi putut fi selectate. Cercetătorul trebuie să dovedească că media eșantionului său, care este cu 3,95g mai mare decât cea a populației, nu se încadrează în caracteristica oricărei medii a unui eșantion de a oscila în jurul mediei populației din care a fost extras (vezi teorema limitei centrale).

Astfel, cercetătorul trebuie să calculeze care este probabilitatea ca eșantionul să facă parte din populația din care a fost extras. Acest raționament statistic se numește **ipoteză de nul (H_0)**. Dacă probabilitatea este mică, vom respinge ipoteza de nul și concluzionăm că eșantionul nu face parte din populație și cantitatea de înghețată primită de clienți este diferită de cea asumată de firmă. Dacă probabilitatea este mare vom accepta ipoteza de nul și vom concluziona că eșantionul face parte din populație. După cum se poate observa, respingerea ipotezei de nul presupune acceptarea ipotezei cercetării, în timp ce acceptarea H_0 solicită respingerea H_1.

7.2 Testul z(t) pentru un singur eșantion

Acest test statistic este utilizat pentru a studia semnificația statistică a diferenței dintre media unui eșantion și media populației din care a fost extras.
Am selectat un eșantion de 40 de clienți și în urma cântăririi cupelor de înghețată pe care aceștia le-au primit am observat că s-a obținut m = 153,95 și s = 8,32. Ne amintim că media populației este 150g. Pe baza acestor informații putem calcula scorul z corespunzător eșantionului după formula:

$$z = \frac{m - \mu}{s_m}$$

(formula 6.3)

- m este media eșantionului
- μ este media populației
- s_m este eroarea standard a mediei

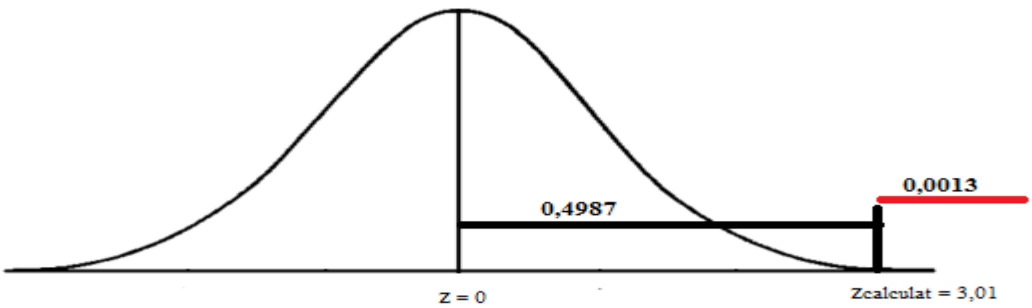
$$s_m = \frac{\sigma}{\sqrt{N}}$$

(formula 6.4)

Pentru a calcula eroarea standard a mediei avem nevoie de abaterea standard a populației, dar deoarece, în cele mai multe situații, nu avem acces la ea o vom aproxima prin utilizarea abaterii standard a eșantionului. Astfel, rezultatele vor fi următoarele:

$$z = \frac{m - \mu}{s_m} \rightarrow z = \frac{m - \mu}{\frac{s}{\sqrt{N}}} \rightarrow z = \frac{153{,}95 - 150}{\frac{8{,}32}{\sqrt{40}}} \rightarrow z = \frac{3{,}95}{1{,}31} \rightarrow z = 3{,}01$$

Mai departe, în tabelul din Anexa 1 vom citi probabilitatea asociată lui z = 3,01 și vom constata că probabilitatea tuturor valorilor posibile între media populației și media eșantionului este 0,4987. Astfel, putem interpreta că probabilitatea ca eșantionul să facă parte din populație este de 0,0013. Cu alte cuvinte, probabilitatea de a găsi un alt eșantion cu o greutate medie mai mare a cupelor de înghețată este de 0,0013. Deci, probabilitatea ca media greutății cupelor de înghețată de la cei 40 de clienți să fie mai mare de 150g din întâmplare este de 0,0013. În concluzie, media eșantionului nu face parte din distribuția de eșantionare, ci este rezultatul unui factor care o face să se îndepărteze semnificativ de media populației.



7.3 Pragul de semnificație statistică

După aplicarea testului statistic, cercetătorul trebuie să decidă dacă greutatea medie a cupelor de înghețată primite de clienți este rezultatul întâmplării sau aceștia primesc mai multă înghețată decât ar dori firma. Deci, trebuie să decidem dacă eșantionul de cupe cântărite face parte din populație (dacă acceptăm ipoteza de nul) sau dacă vom considera că este diferit de aceasta (vom respinge ipoteza de nul).

O soluție general acceptată este aceea de a utiliza **pragul alfa**. În științele sociale acest prag (numit și **prag critic** sau **nivel de semnificație statistică**) corespunde unei probabilități de 0,05. Dacă obținem o probabilitate mai mică sau egală cu 0,05 ca eșantionul să facă parte din populație ne aflăm în situația de a respinge ipoteze de nul și de a considera acea diferență ca fiind semnificativă statistic. Dacă probabilitatea este mai mare decât 0,05 vom accepta ipoteza de nul. Comunitatea științifică nu permite stabilirea unui prag critic mai mare de 0,05. În schimb, pot fi situații în care, în funcție de nivelul de risc pe care dorim să ni-l asumăm, putem stabili un prag critic mai mic, de exemplu **0,01**.

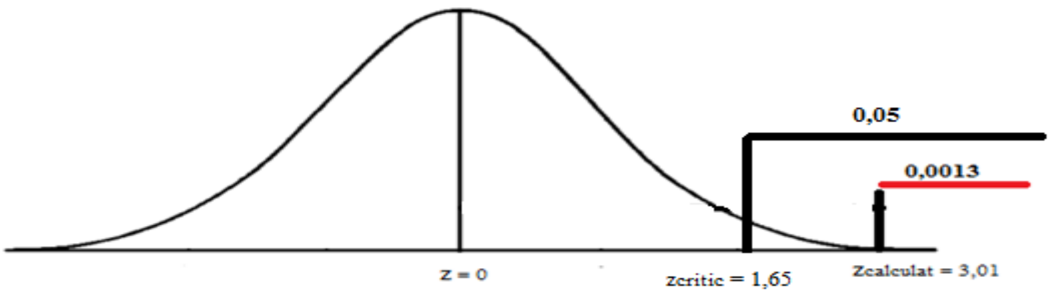
Pe curba normală, fiecărei probabilități îi corespunde o anumită valoare z. În cazul nostru, probabilității 0,05 (probabilitatea critică) îi corespunde unui scor $\mathbf{z_{critic} = 1{,}65}$. Având în vedere că probabilitatea de 0,05 ne indică care este șansa că noi să găsim un alt eșantion cu un scor asemănător, din anexă vom citi valoarea scorului z care are această probabilitate în afara lui. Acesta este algoritmul care ne conduce la scorul critic z = 1,65.

Atunci când **valoarea calculată este mai mare sau egală cu valoarea critică** înseamnă că probabilitatea ca eșantionul să facă parte din populație este mai mică sau egală cu 0,05. În consecință vom respinge H_0 și vom accepta H_1. Cu alte cuvinte diferența dintre media eșantionului și media populației este semnificativă statistic.

Atunci când valoarea calculată este mai mică decât valoarea critică vom concluziona că probabilitatea ca eșantionul să facă parte din populație este mai mare de 0,05 și vom accepta ipoteza nul, respingând-o pe cea a cercetării. Astfel, diferența dintre media eșantionului și media populației este nesemnificativă statistic.

Softul R ne va indica direct probabilitatea asociată valorii calculate a testului, adică probabilitatea ca ipoteza de nul să fie adevărată. Astfel, vom compara probabilitatea afișată de programul statistic cu pragul de semnificație alfa. Atunci când probabilitatea valorii calculate este **mai mare** decât pragul de semnificație **α** vom accepta H_0. Dacă probabilitatea este **mai mică sau egală** cu α vom respinge H_0.

Testarea ipotezelor este similară cu decizia juraților în cadrul unui proces. Într-un corp de jurați, H_0 este similară cu sentința nevinovat, iar H_1 reprezintă sentința vinovat. În cazul unui proces acuzatul se bucură de prezumția de nevinovăție până când pe baza unor dovezi evidente se stabilește că el este vinovat. Dacă juriul stabilește că probele existente la dosar sunt evidente, va respinge H_0 (nevinovat) și va accepta H_1 (vinovat).



Imaginea 7.1 – Decizia statistică unilaterală

Pe baza rezultatelor obținute ($z_{calculat} = 3{,}01$ mai mare decât $z_{critic} = 1{,}65$) decizia statistică este de a respinge ipoteza de nul și de acceptare a ipotezei cercetării (greutatea cupelor de înghețată primite de clienți este semnificativ statistic mai mare decât cea specificată de companie).

Să ne imaginăm că greutatea cupelor de înghețată ar fi fost 146,05 grame. Aplicând formula de calcul a testului z pentru un singur eșantion am fi obținut o valoarea calculată z = -3,01. În acest caz, dacă ipoteza conform căreia greutatea cupelor de înghețată este mai mare decât greutatea specificată de companie trebuie respinsă și acceptată ipoteza de nul, deși există o diferență semnificativă între cele două medii, dar în sensul că greutatea de înghețată primită de clienți este semnificativ mai mică. Formulând ipoteza cercetării în partea pozitivă, am pierdut din vedere faptul că diferența dintre medii poate fi și negativă.

În consecință vom formula ipoteza bilateral și vom verifica ipoteza la ambele capete ale distribuției. Și în cazul deciziei bilaterale se va menține același nivel al pragului α, dar el se va împărți în mod egal la ambele capete ale distribuției. Astfel, dacă vom alege un $\alpha = 0{,}05$, vom avea o probabilitate critică de 0,025 în fiecare extremitate a curbei. Pentru a afla valoarea critică, din tabelul de la **Anexa 1**, vom citi scorul z care are această probabilitate de a obține un scor mai mare (z = 1,96). În cazul unei decizii bilaterale ne așteptăm ca diferența dintre medii să fie atât pozitivă, cât și negativă. În consecință valoarea critică **z** va fi **±1,96**, deoarece vom stabili un prag critic pentru partea stângă a curbei și o valoare critică pentru partea dreaptă a acesteia.

Imaginea 7.2 – Decizia statistică bilaterală

În cazul unei decizii bilaterale probabilitatea de a accepta ipoteza de nul este mai mare deoarece pragul critic se mută mai mult înspre extremitățile curbei și acesta devine mai greu de depășit. Prin urmare, putem înțelege că testarea bilaterală a unei ipoteze presupune mai multă rigoare, un nivel mai scăzut de asumare a riscului. Testarea unilaterală sau bilaterală a unei ipoteze depinde de decizia cercetătorului, dar literatura de specialitatea recomandă utilizarea deciziilor bilaterale. După cum vom putea observa, programul statistic SPSS este setat implicit pe testarea bilaterală.

7.4 Testul z(t) pentru un singur eșantion cu R

Un director de școală este interesat să analizeze dacă există o diferență semnificativă la evaluarea națională între rezultatele obținute de elevii școlii pe care o administrează și media populației de elevi. La nivel național media la Matematică este 24.86, iar media la Lb. Română este 25.14. Datele colectate au fost sintetizate în baza date date bdtestt.xlsx

Ipoteza cercetării (H_{1a}): Există diferențe semnificative la nivelul rezultatelor obținute la evaluarea națională la Matematică între media elevilor și media populației.
Ipoteza cercetării (H_{1b}): Există diferențe semnificative la nivelul rezultatelor obținute la evaluarea națională la Lb. Română între media elevilor și media populației.
Ipoteza de nul (H_{0a}): Nu există diferențe semnificative la nivelul rezultatelor obținute la evaluarea națională la Matematică între media elevilor și media populației.
Ipoteza de nul (H_{0b}): Nu există diferențe semnificative la nivelul rezultatelor obținute la evaluarea națională la Lb. Română între media elevilor și media populației.

Pentru a lansa acest test în R avem la dispoziție funcția **t.test()**.

t.test(bdtestt$Matematică, mu = 24.86, alternative = “two.sided”), unde:

- t.test este funcția dedicată testului t;
- bdtestt reprezintă baza de date;
- Matematică este variabila analizată;
- mu solicită introducere mediei populației;
- alternative precizează dacă ipoteza este unilaterală sau bilaterală (în acest caz este bilaterală).

Rezultatele obținute ne indică faptul că media obținută de elevi la matematică, 27.98, este semnificativ mai mare decât cea a populației, deoarece *p* este mul mai mic decât 0.05.

- $t = 17.144$ este valoarea calculată a testului și nu se interpretează în absența valorilor critice;
- df = 119 sunt gradele de libertate (sunt egale cu numărul de participanți – 1);
- *p* este probabilitatea asociată testului; atunci când *p* este mai mare decât 0.05 se acceptă ipoteza de nul și se concluzionează că nu există o diferență semnificativă între media eșantionului și media populației; dacă $p \leq 0.05$ se respinge ipoteza de nul și se afirmă că există o diferență semnificativă între media eșantionului și cea a populației.
- mean of x = 27.98 indică media eșantionului.
- 95 percent confidence interval reprezintă intervalul de încredere pentru media eșantionului; în acest caz intervalul este 27.62 – 28.34; cu alte cuvinte, dacă directorul ar alege alte 100 de eșantioane de elevi din școala sa, cel puțin 95% dintre ele ar avea media cuprinsă între 27.62 și 28.34. În cazul în care media populației nu se regăsește între limitele acestui interval avem un indiciu că ipoteza de nul este respinsă. Dacă media are o valoare care face parte din acest interval indiciul este că ipoteza de nul trebuie acceptată.

În cazul nostru *p* este mai mic decât pragul 0.05, ceea ce ne face să respingem ipoteza de nul și să acceptăm ipoteza cercetării; există o diferență semnificativă între media eșantionului de elevi și media populației.

```
t.test(bdtestt$Matematica, mu=24.86, alternative="two.sided")

        One Sample t-test

data:  bdtestt$Matematica
t = 17.144, df = 119, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 24.86
95 percent confidence interval:
 27.62259 28.34408
sample estimates:
mean of x 
 27.98333 
```

Pentru a testa ipoteza referitoare la rezultatele obținute de elevi la Limba Română vom proceda la fel ca în cazul de mai sus, cu excepția faptului că vom schimba variabila analizată și media populației.

```
t.test(bdtestt$Română, mu=25.14, alternative="two.sided")
```

```
	One Sample t-test

data:  bdtestt$Română
t = 27.009, df = 119, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 25.14
95 percent confidence interval:
 28.59345 29.13988
sample estimates:
mean of x
 28.86667
```

Și în acest caz *p* este mai mic decât pragul α (0.05), rezultat care se traduce în faptul că ipoteza de nul este respinsă. Mai specific, media elevilor la Limba Română (28.87) este semnificativ mai mare decât media populației.

Raportarea rezultatelor obținute în urma aplicării testului statistic se face într-o formă clară astfel încât să se înțeleagă contextul cercetării în care testul a fost aplicat. Pentru exemplul de mai sus, raportarea rezultatelor se va face precum în exemplul de mai jos.

S-a aplicat testul t pentru un singur eșantion pentru a testa semnificația statistică a diferenței dintre mediile obținute la Evaluarea Națională de elevii de clasa a IV-a ai unei școli și media populației. În acest sens, se cunoaște faptul că media populației (μ) la Matematică este egală cu 24.46, iar la Limba Română este egală cu 25.14. Media eșantionului la Matematică este 27.98, iar la Limba Română este 28.87. Testul t pentru un singur eșantion are o probabilitate asociată $p < .05$ atât pentru Matematică, cât și pentru Limba Română. Aceste rezultate ne fac să respingem ipoteza de nul și să afirmăm că există diferențe semnificative între rezultatele obținute de elevii școlii și populația de elevi atât la Matematică, cât și la Limba Română.