Anexa 8. Ce înseamnă nivel de semnificație

Gândirea statistică este greu de asimilat, printre altele, şi pentru că se bazează când pe intuiții comune, când pe aspecte specifice, contrare intuiției, fără ca nimeni să marcheze când este un caz sau celălalt. Vom încerca, în continuare, să evidențiem aceste diferențe în cazul conceptului de *semnificație statistică* versus *semnificație în sens comun*.

Semnificație în sens comun

Să pornim de la un exemplu din viața de toate zilele. Să presupunem că am fost la piață şi am cumpărat un kilogram de piersici. Venim acasă şi vrem să verificăm dacă am fost înşelați la cântar. Pentru aceasta cântărim din nou piersicile pe cântarul propriu care are o precizie mai mare. Constatăm greutatea de 950 grame. În altă zi, negustorul ni le-a dat “cu bătaie” şi constatăm greutatea de 1050 grame. Întrebarea este “ne considerăm înşelați în aceste cazuri ori diferențele sunt nesemnificative, adică provin din inerentele erori întâmplătoare de măsurare ?”. Personal aş opta pentru varianta a doua.

Să presupunem, însă, că acasă kilogramul de piersici are doar 800 de grame. În acest caz cred că nu mai suntem dispuşi să acceptăm că diferența este nesemnificativă pentru că este de 200 de grame (a cincia parte din greutate) față de diferențele anterioare de $\pm$ 50 grame (doar a 20-a parte). Spunem că această diferență este semnificativă. În mod asemănător o diferență, să zicem de 250 grame (a patra parte), o putem considera foarte semnificativă şi aşa mai departe. Aceasta reprezintă **semnificație în sens comun**. Ea este legată de specificul fiecărei probleme de viață, în acest caz depinzând de buzunarul fiecăruia şi de cât de “rău de pagubă” este. Astfel, un om sărac şi rău de pagubă poate considera semnificativă şi diferența de 50 de grame, mai ales dacă piersicile sunt scumpe.

Reținem două aspecte comune conceptului de semnificație atât statistică cât şi în sens comun: (1) o diferență este cu atât mai semnificativă cu cât este mai mare şi (2) plasarea nivelului de semnificație (semnificativ, foarte semnificativ, etc.) este o problemă subiectivă.

Semnificație statistică (a posteriori)

Pentru a ne apropia de conceptul de semnificație statistică să dezvoltăm, într-un anumit mod, exemplul anterior. Să presupunem că am cumpărat câte un kilogram de piersici de foarte multe ori de la acelaşi negustor, de fiecare dată cantitatea măsurată acasă fiind cuprinsă între 950 grame şi 1050 grame. Într-o zi însă, kilogramul de piersici cumpărat de la acelaşi negustor are acasă 800 de grame. Deoarece experiența ne-a învățat că negustorul nostru nu ne-a înşelat şi ne-a dat cantități ce diferă de 1 Kg doar în limitele preciziei cântarului său (± 50 g) este greu să credem că diferența de 200 g a apărut la fel. O vom considera deci **semnificativă statistic**, *în sensul că ea ar putea apărea doar cu o mică probabilitate, dintr-o eroare întâmplătoare*. Prin urmare, semnificația statistică este legată de producerea unui rezultat puțin aşteptat să apară din întâmplare şi tocmai de aceea denumit “semnificativ”, adică care aduce ceva nou, ceva cu semnificație. Se observă aici mecanismul intuitiv care stă la baza testelor statistice: atunci când un eveniment are o mică probabilitate de apariție din întâmplare în anumite condiții presupuse, îl considerăm aducător de (o nouă) semnificație, adică decidem că, de fapt, condițiile presupuse nu sunt adevărate.

Revenind la exemplul cu piersicile, în cazul diferenței de 200 g considerăm că negustorul ne-a înşelat, deşi este posibil (dar puțin probabil) ca diferența să fi apărut şi din întâmplare, dintr-o eroare involuntară a negustorului.

Acest exemplu ilustrează ideea de semnificație statistică bazată pe experiență. Putem denumi acest concept **semnificație statistică a posteriori** (experienței se înțelege). În ultimul exemplu am acordat o mică probabilitate apariției diferenței de 200 g bazându-ne pe experiența anterioară: negustorul nostru nu greşea decât în limita a ± 50 g.

Semnificație statistică (a priori)

Semnificația statistică cu care se operează de obicei este o **semnificație statistică a priori**. Aceasta are toate caracteristicile semnificației a posteriori cu excepția modului de calcul al probabilității apariției unui eveniment. Dacă eşantionul este extras aleator putem calcula această probabilitate indiferent de experiența anterioară, bazându-ne doar pe presupunerile făcute în ipoteza nulă şi, eventual, în condițiile testului.

Să adaptăm acestui caz exemplele anterioare. Presupunem, de pildă, că negustorul nostru de piersici pretinde că fructele sale au în medie un diametru de 6 cm. Putem verifica afirmația lui fie (1) măsurând toate piersicile negustorului fie (2) măsurând doar o parte extrasă aleator.

Ipoteze logice

În cazul (1), putem spune că verificăm ipoteza I_0: $\mu = 6$ cm în opoziție cu ipoteza I_A: $\mu \neq 6$ cm. Evident după măsurarea tuturor piersicilor şi calcularea mediei vom decide cu certitudine care dintre cele două ipoteze este adevărată. Decizia va fi luată în condițiile logicii clasice în care acționează *legea terțului exclus*[*] şi *legea contradicției*[†]. De aceea, propunem pentru cuplul I_0, I_A denumirea de **ipoteze logice** (una fiind, evident, contrara celeilalte).

În cazul (2) să presupunem că extragem aleator din marfa vânzătorului 37 de piersici. Măsurăm diametrele acestora şi obținem media, să spunem de 5,9 cm (notată m_1), ori cea de 6,1 cm (notată m_2). Intuitiv, vom spune referitor la ipoteza I_0 ($\mu = 6$ cm), susținută de negustor că “probabil are dreptate”. Aceasta înseamnă că “este posibil să aibă dreptate” şi că suntem dispuşi să credem că “mai de grabă are dreptate decât că nu are”. Să argumentăm pe rând ultimele două afirmații:

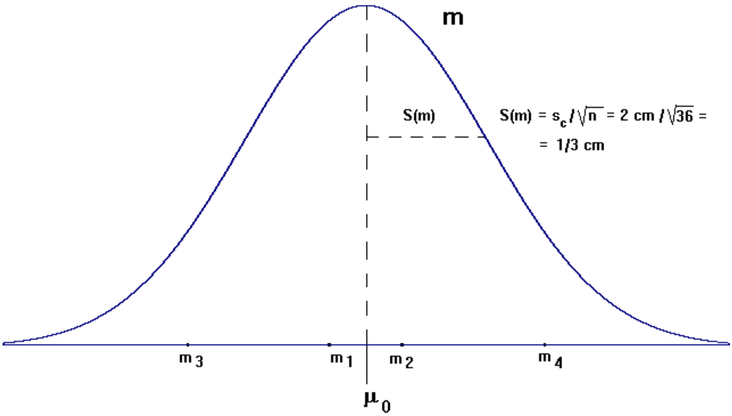
- Spunem că “este posibil să aibă dreptate” pentru că nu am cercetat toată marfa (aşa-numita populație statistică) pentru a obține o certitudine.
- Afirmăm că “mai de grabă are dreptate decât că nu are” pentru că diferențele între 6 cm şi 5,9 cm, respectiv 6,1 cm, adică de $\pm$ 0,1 cm nu ni se par *semnificative*. Altfel spus, le putem atribui întâmplării, adică inerentelor deosebiri între populație şi eşantioane. Spunem că aceste deosebiri se datorează aşa-numitelor **fluctuații de eşantionaj**.

Nu cred că vom reacționa la fel dacă media eşantionului ar fi 5,5 cm (notată m_3) ori 6,5 cm (notată m_4). În aceste ultime exemple au apărut diferențele de $\pm$ 0,5 cm. Acestea fiind relativ mari, este mai greu să le atribuim fluctuațiilor de eşantionaj dintr-o populație cu media $\mu = 6$ cm. Fiind puțin probabile, le considerăm cvasiimposibile în condițiile în care I_0 ar fi adevărată şi vom decide în favoarea celeilalte ipoteze, I_A: $\mu \neq 6$ cm.

[*] “O propoziție este ori adevărată ori falsă, a treia posibilitate este exclusă.”

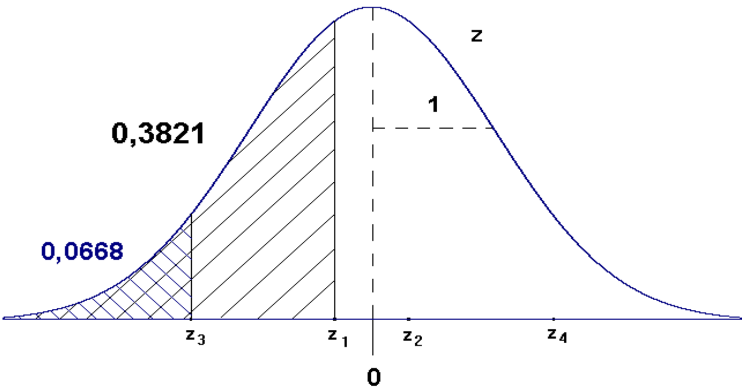
[†] “O propoziție nu poate fi în acela]i timp şi adevărată şi falsă.”

Prin urmare, în cazul utilizării unui eşantion, alegerea între cele două ipoteze se poate face numai dacă (I) ştim să calculăm probabilitatea rezultatului produs de eşantionul respectiv în condiţiile în care ipoteza I_0 ar fi adevărată şi dacă (II) stabilim un prag sub care considerăm mică o probabilitate.



Statistica matematică a rezolvat problema (I) pentru medie, dispersie, frecvenţă, etc. în diverse situaţii. De exemplu, dacă eşantionul este extras aleator cu revenire, iar volumul eşantionului este mai mare de 30 de unităţi, atunci media de eşantionaj, *m*, se va distribui gaussian în jurul mediei populaţiei specificată în ipoteza I_0.

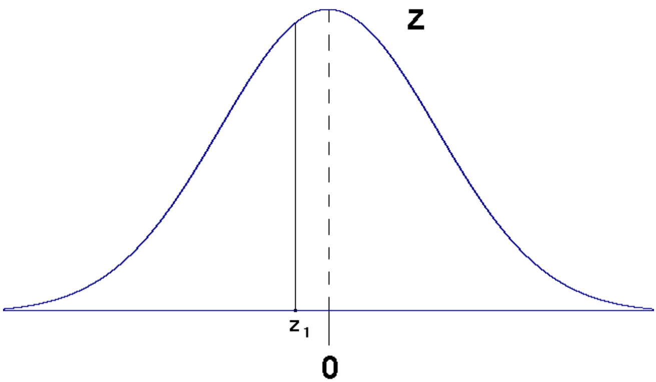
În consecinţă, probabilitatea se poate calcula cu ajutorul tabelei distribuţiei normale standard. Pentru aceasta se calculează scorul $z_i = \dfrac{m_i - \mu_0}{s_c / \sqrt{n}}$, în care m_i este media eşantionului extras (aici 5,9 cm, 6,1 cm, 5,5 cm sau 6,5 cm), μ_0 este media



populaţiei specificată în ipoteza I_0 (aici 6 cm), s_c este abaterea standard a eşantionului (să spunem 2 cm) şi *n* este volumul eşantionului (aici 37). În acest exemplu se obţin scorurile z_i: -0,3; 0,3; -1,5, respectiv 1,5. Scorul z urmează o lege normală standard şi reprezintă un model matematic nemărginit şi continuu pentru o problemă care este mărginită şi discontinuă.

În exemplul nostru media de eşantionaj este mărginită inferior căci nu poate fi negativă, şi superior, pentru că nu poate depăşi o anumită valoare specifică fructului (să spunem 15 cm). Este, de asemenea, discontinuă pentru că dacă luăm în considerare diametrele tuturor piersicilor negustorului, acestea sunt în număr finit şi deci şi toate mediile de câte 37 de unităţi vor fi, de asemenea, în număr finit şi deci discret, discontinuu.

Revenind la modelul de lege normală standard pentru scorul z (în condiţiile în care I_0 este adevărată) observăm că nemărginirea înseamnă că z şi prin urmare şi m, media de eşantionaj, pot lua orice valoare. Prin urmare, cozile distribuţiei vor trebui tăiate pentru ca să devină mărginită. Cum vom face acest lucru, va rezulta din celălalt defect al modelului – continuitatea. Într-adevăr, distribuţia fiind continuă, probabilitatea de a se obţine exact o anumită medie m_i, probabilitate notată $p\ (m = m_i)$ şi egală cu $p\ (z = z_i)$, va avea aceeaşi valoare pentru orice m_i şi anume zero. Justificarea este imediată: $p\ (z = z_i)$ este proporţia ariei segmentului vertical plasat în z_i între axa de coordonate orizontală şi curba normală standard, din toată aria de sub curbă (vezi figura alăturată). Aceasta arie relativă este zero, aria absolută a unui segment fiind întotdeauna nulă.



Probabilitatea obţinerii valorii z_i fiind nulă, dar interesându-ne tocmai tăierea cozilor, vom considera **probabilitatea ca z să fie mai excentric decât z_i** [de exemplu $p\ (z > z_2)$, $p\ (z > z_4)$, respectiv $p\ (z < z_1)$ şi $p\ (z > z_3)$]. Luand ca referinta aria de sub intreaga curba, această probabilitate va fi proporţia reprezentata de aria de sub curbă aflată la dreapta punctului z_i, în cazul punctelor z_2 şi z_4, respectiv aflata la stânga punctului z_i, în cazul punctelor z_1 şi z_3. mk

În concluzie, statistica matematică ne furnizează în cadrul punctului (I) de mai sus, în locul “probabilității rezultatului produs de eşantion”, “probabilitatea unui rezultat mai excentric decât cel produs de eşantion”*. Această probabilitate se notează, în literatura de specialitate, p. În [3] am notat-o α_c. Consultând tabela din Anexa 2 obținem pentru z_2, respectiv z_4 probabilitatea 0,3821, respectiv 0,0668. Datorită simetriei față de zero a distribuției normale standard, obținem aceleaşi rezultate pentru z_1, respectiv z_3.

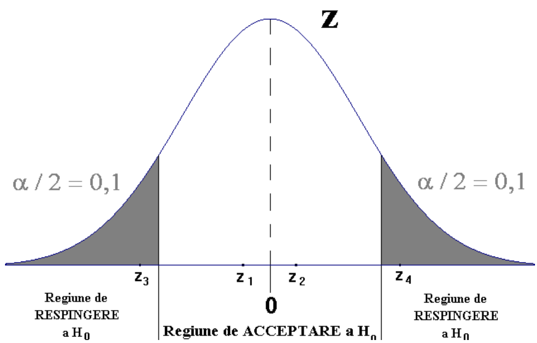
Nivel de semnificație

Având probabilitatea notată p putem lua o decizie doar dacă rezolvăm punctul (II) de mai sus, adică dacă fixăm un nivel α sub care vom considera că probabilitatea respectivă este mică. Nivelul respectiv se numeşte **nivel de semnificație**. Pentru a fi mai vizibile figurile, să fixăm α la valoarea 0,2. Aceasta o vom repartiza în mod egal pe ambele cozi ale distribuției, dacă dorim să luăm în considerare atât probabilitatea unei medii a diametrelor piersicilor negustorului semnificativ mai mică de 6 cm, cât şi una mai mare. Prin urmare, vom considera semnificative acele valori z_i care produc probabilități mai mici ca $\alpha / 2 = 0,1$, în exemplul nostru z_3 şi z_4.

Ipoteze statistice şi teste bilaterale

Să examinăm din punct de vedere statistic construcția anterioară.

- Am formulat ipoteza logică I_0 (μ = 6 cm), care, în cazul utilizării unui eşantion în locul cercetării întregii populații, înseamnă că presupunem că media, notată μ_i, a populației din care am extras eşantionul de medie m_i este egală cu o



* probabilitate egală, evident, şi cu “probabilitatea unui rezultat egal cu cel al eşantionului sau mai excentric decât acesta”. Vom prefera însă, pentru concizie, prima exprimare.

valoare, notată μ_0, egală aici cu 6 cm. Am formulat astfel ipoteza statistică corespunzătoare H_0: $\mu_i = \mu_0$. Aceasta se poate rescrie sub forma H_0: $\mu_i - \mu_0 = 0$, ceea ce înseamnă că diferența între cele două medii este nulă. De aici provine denumirea de **ipoteză** (statistică) **nulă** sau **ipoteză** (statistică) **de nul**.

- Am considerat modul în care se distribuie media de eşantionaj, respectiv, scorul *z*, în contextul în care H_0 ar fi adevărată, iar extragerea a fost aleatoare.
- Am stabilit că ne interesează situația în care media diferă semnificativ atât prin lipsă cât şi prin adaos, adică am formulat ca *ipoteză (statistică) alternativă*, contrara lui H_0 şi anume H_A: $\mu_i \neq \mu_0$.
- Astfel, nivelul de semnificație α s-a repartizat în mod egal (pentru că nu există nici-un motiv să fie altfel) pe ambele cozi ale distribuției de eşantionaj, *z*. Prin urmare, sub *z* apar două regiuni: **regiunea de acceptare** a H_0 şi **regiunea de respingere** (a H_0), plasată bilateral, adică pe ambele cozi ale distribuției.

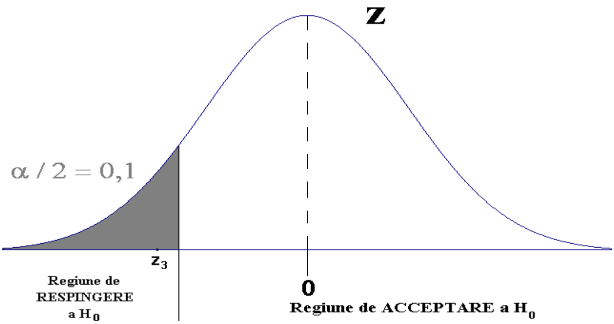
Observăm că în această situație, a testelor bilaterale, ipotezele statistice coincid cu cele logice, ipoteza alternativă fiind contrara celei nule, ceea ce nu se va mai întâmpla în cazul testelor unilaterale.

Ipoteze statistice, ipoteze logice şi teste unilaterale

În problema anterioară, din punctul nostru de vedere, ne afectează neplăcut doar cazul în care media diametrelor piersicilor ar fi mai mică, adică situația în care ipoteza alternativă ar fi H_A: $\mu_i < \mu_0$.

Ipoteza (statistică) nulă va rămâne aceeaşi ca şi în cazul unui test bilateral pentru că putem calcula probabilitatea unui rezultat mai excentric decât cel dat de eşantion doar pentru o valoare fixată, μ_0, a mediei. Deci cuplul de **ipoteze statistice** va fi H_0: $\mu_i = \mu_0$; H_A: $\mu_i < \mu_0$.

Acum zona de respingere se va plasa numai unilateral pe coada din stânga a distribuției lui *z*, în ipoteza că H_0 este adevărată (adică H_0: $\mu_i = \mu_0$). De aceea testul se va numi **test unilateral stânga**.



Observație: Pentru a nu apărea contradicția semnalată de noi în observația 2 din Lp 9, în [4] se propune utilizarea, în cazul testelor unilaterale, a nivelului $\alpha / 2$, α fiind nivelul stabilit în cazul testelor bilaterale.
Revenind la exemplul nostru, vom respinge H_0 doar în cazul valorii m_3, deoarece $p(m < m_3) = p(z < z_3) = 0{,}0668 < 0{,}1$.

Observăm că în cazul unui test unilateral cele două ipoteze statistice nu mai sunt una contrara celeilalte. Situația este contrară intuiției pentru că natural este să alegem între variantele aflate în opoziție una față de alta, I_0: $\mu_i \geq \mu_0$; I_A: $\mu_i < \mu_0$ şi denumite, de noi, **ipoteze logice**. Alegerea între aceste două variante nu se poate face, însă, decât în cazul în care cercetăm întreaga populație statistică, adică nu mai facem statistică inductivă. În cazul cercetării printr-un eşantion, adică al statisticii inductive, "dilema" este mai ciudată, alternativele nefiind în opoziție logică (în cazul testelor unilaterale).

În următoarea tabelă am centralizat cuplurile de ipoteze logice corespunzătoare cuplurilor de ipoteze statistice.

Observații ↓	*Ipoteze statistice:*	*Ipoteze logice ataşate:*
H_0 şi I_0, respectiv H_A şi I_A sunt identice	H_0: $\mu_i = \mu_0$; H_A: $\mu_i \neq \mu_0$	I_0: $\mu_i = \mu_0$; I_A: $\mu_i \neq \mu_0$
H_A şi I_A sunt identice	H_0: $\mu_i = \mu_0$; H_A: $\mu_i < \mu_0$	I_0: $\mu_i \geq \mu_0$; I_A: $\mu_i < \mu_0$
H_A şi I_A sunt identice	H_0: $\mu_i = \mu_0$; H_A: $\mu_i > \mu_0$	I_0: $\mu_i \leq \mu_0$; I_A: $\mu_i > \mu_0$
Observații →	H_0 este aceeaşi.	I_A este opusul lui I_0.

Încheiem cu observația că, aparent contrar intuiției, <u>un nivel de semnificație α este considerat cu atât mai înalt cu cât are o valoare mai mică</u>. Justificarea este bazată pe faptul că o diferență este cu atât mai semnificativă cu cât este mai mare (in valoare absoluta). Dar cu cât este mai mare diferența, cu atât este mai mică probabilitatea de apariție a sa. De exemplu, în cea de-a doua figură din această anexă probabilitatea obținerii unei valori mai excentrice decât z_3 (0,0668) este mai mică decât cea corespunzătoare lui z_1 (0,3821) pentru că $|z_3| > |z_1|$.