

Seminar 5

Deschideți fișierul "P5.xls" și salvați-l sub numele: "Nume_P5.xls" în directorul dedicat cursului. Fișierul conține mai multe foi (worksheets)!

Testarea ipotezelor

Pentru a calcula statistica-t vom folosi programul EXCEL, care are 3 tipuri de teste-t și un test-z. Acestea se pot găsi în Analysis ToolPak ('Tools', 'Data Analysis'):

t-Test: Paired Two-Sample for Means

t-Test: Two-Sample Assuming Equal Variances

t-Test: Two-Sample Assuming Unequal Variances

z-Test: Two Sample for Means

Programul EXCEL nu are teste pentru un eșantion, deci pentru acest caz, calculele se vor efectua *'manual'*.

I. ("workshet-ul TI 1") O companie din domeniul biotehnologiilor susține că folosirea grâului modificat genetic în combinație cu un anumit ierbicid propriu duce la obținerea unei producții medii de 3,5 tone/hectar.

Pentru a verifica această afirmație a fost făcut un studiu. Astfel, semințele de grâu și ierbicidul au fost testate la 20 de ferme. Recoltele au ajuns la maturitate și producția medie a fost de 2,9 tone /hectar, cu o deviație standard (calculată folosind corecția Bessel) de 0,9 tone/ hectar.

Folosind testul-t calculați *"manual"* dacă există o diferență semnificativă între producția medie declarată de companie și producția medie obținută din teste, pentru un nivel de încredere de 95%.

Pentru a determina valoarea calculată folosiți formula dată la curs:

$$t_{\text{calc}} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{\sqrt{n}}{s} (\bar{x} - \mu)$$

Pentru a determina valoarea critică (t_{crit}) pentru statistica t, folosiți tabelul probabilității statisticii-t sau funcția TINV.

Pentru a determina probabilitatea ca o variabilă aleatoare care urmează distribuția-t să se afle la maxim t unități depărtare de zero (deci între $-t$ și t) se folosește funcția TDIST!

II. Un institut de cercetări medicale a investigat efectul unui nou supliment alimentar asupra nivelului glicemiei în organism. Un număr aleator de analize au fost prelevate de la elevii a două

scoli aflate în același oras, elevii dintr-o școală au utilizat noul supliment alimentar, iar cei din cealaltă școală nu. Rezultatele măsurătorilor de laborator se află în worksheet-ul "TI 2".

Folosind "Analysis ToolPak" din EXCEL, efectuați un **test-t** (variantă egală) asupra datelor. Pentru acest tip de date trebuie folosit testul: "t-Test: Two-sample Assuming Equal Variances".

- În "Variable 1 range" și "Variable 2 range" introduceți domeniul celulelor ce conțin cele două eșantioane de date. Dacă bifați casuța "Labels", puteți include și celulele cu titluri (etichete).

- Selectați o valoare alpha de 0,05 (corespunzătoare nivelului de încredere de 5%).

- Activați "output range" și selectați o celulă goală de pe ecran. Click pe OK pentru a fi afișat tabelul cu rezultate.

- Studiați mărimile care apar în tabel pentru a înțelege ce înseamnă fiecare, în special mărimile "P(T≤t)" și "t Critical" pentru teste mărginite la un capăt sau la două (one-tail și two-tailed).

- Folosind informațiile din tabelul dat de programul EXCEL, decideți dacă acceptați sau respingeți ipoteza nulă (H_0 : nu sunt diferențe între mediile celor două eșantioane) pentru un nivel de încredere de 90%, respectiv 99%.

- Folosind informațiile din tabelul dat de programul EXCEL, decideți dacă acceptați sau respingeți ipoteza nulă (H_0) pentru un nivel de încredere de 95%.

Repetăți **testul-t**, folosind testul care presupune **varianțe inegale**. Comparați tabelul obținut în acest caz cu cel pentru cazul varianțelor egale. Observați diferențele dintre cele două teste. Care test este mai potrivit pentru aceste date?

III. Testul-t pentru eșantioane pereche a fost creat pentru situații în care sunt înregistrate măsurători pereche a aceleiași variabile, de multe ori "înainte" și "după" tratament. În programul EXCEL, acest test se numește "t-Test: Paired Two Sample for Means".

Presupunem că fiecare set de date din worksheet-ul "TI 3" a fișierului EXCEL P5 STAT.xls reprezintă 20 culturi esofagiene, analizate înainte și după tratamentul cu un anumit antimicrobic.

Datele obținute pentru fiecare cultură esofagiană reprezintă eșantioane pereche.

- o Folosind "Data Analysis Tool", efectuați un test-t pereche pentru aceste date. Decideți dacă veți accepta sau respinge ipoteza nulă (nu sunt diferențe între medii), pentru un nivel de încredere de 95%.

IV. În tabelul din worksheetul "TI 4" găsiți date referitoare la îmbolnavirile de bronsită: anotimpul îmbolnavirii, mediul, genul, vârsta, dacă au existat antecedente în ultimul an.

Efectuați un **test- χ^2** , pentru a putea răspunde la următoarea întrebare: *Imbolnavirea de bronsita depinde de mediul de viata?* Dați răspunsul pentru un nivel de încredere de 95%.

Pentru a nu modifica datele originale, realizati o copie a worksheet-ului 'TI 4' in care sa efectuati calculele:

- click-dreapta pe eticheta cu numele 'TI 4' situată în partea de jos a work sheet-ului, alegeți opțiunea 'Move or Copy'; bifați căsuța 'Create a copy';
- renumiți worksheet-ul copiat: faceți dublu-click pe eticheta cu numele și scrieți noul nume ("Tabel contingenta", apoi apăsați tasta 'Enter').
- Scrieți în două celule ale worksheet-ului nou creat ipoteza nulului (H_0) si ipoteza alternativă (H_1).
- Creați un *tabel de contingență*:

trebuie să determinați numărul imbolnavirilor din mediul urban '+' și a celor din mediul rural '-' folosind funcția "`=COUNTIF(range, condition)`".

- într-o parte a foi de lucru, unde doriți să plasați tabelul de contingență scrieți două capuri de tabel, 'Urban' și 'Rural' în două celule alăturate.
- sub celula 'Urban', folosiți funcția COUNTIF pentru a calcula numărul imbolnavirilor din mediul urban: "`=COUNTIF(range,="+")`" [Dacă tastați direct formula, trebuie să introduceți caracterul + între ghilimele; dacă folosiți icoana "formula", atunci introduceți numai caracterul + în câmpul corespunzător]
- Faceți aceiași pași pentru mediul rural, dar introducând caracterul "-" în câmpul corespunzător.
- Cele două numere calculate reprezintă 'frecvențele observate' ale tabelului de contingență, deci etichetați acest rând ' f_o '.
- În rândul următor calculați frecvențele. Etichetați acest rând ' f_e '.
- Calculați statistica χ^2 .
- În rândul următor, calculați valoarea termenului $(f_o - f_e)^2/f_e$ pentru fiecare coloană (creati pentru aceasta o formulă).
- Calculați totalul tuturor valorilor pentru $(f_o - f_e)^2/f_e$, care reprezintă:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

Etichetați suma finală ' χ^2 '.


- Calculați gradul de libertate, alegeți nivelul de semnificație și apoi citiți valoarea critică X^2 din tabelul probabilităților statisticii χ^2 .
- Respingeți sau acceptați ipoteza nulului?

Deși programul EXCEL nu are facilitatea de a efectua testul X^2 , totuși există câteva funcții care ușurează efectuarea testului χ^2 .

Funcția "CHIDIST": calculează probabilitatea asociată cu orice valoare a lui χ^2 .

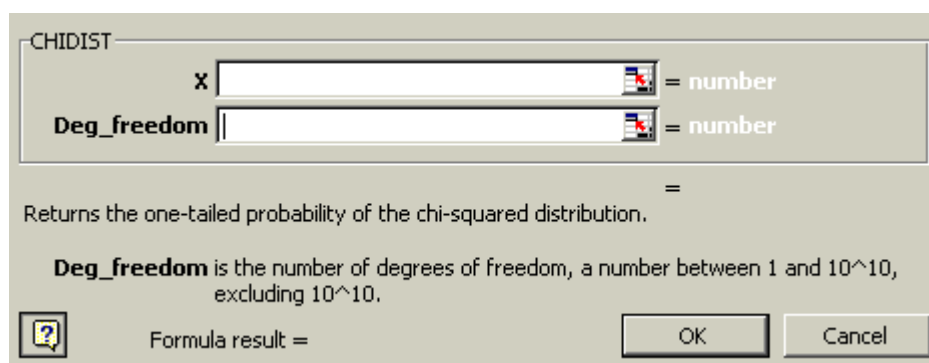
- Folosiți funcția CHIDIST pentru a calcula probabilitatea ca valoarea calculată pentru χ^2 să apară numai datorită șansei (exemplu: prin eșantionarea aleatoare a unei populații care are un număr egal de așezări cu plaje protejate și neprotejate).

- Selectați o celulă goală.

- Click pe icoana pentru formule  și alegeți CHIDIST dintre formulele Statistice.

- Introduceți valoarea pe care ați calculat-o pentru χ^2 , gradul de libertate și apoi click 'OK'.


Notă: în loc de a introduce valoarea lui χ^2 calculat se poate introduce trimitere la celula în care s-a calculat χ^2 - acest lucru este util dacă vreți să faceți modificări la tabelul de contingență: rezultatul funcției CHIDIST pentru noua valoare va apărea automat.

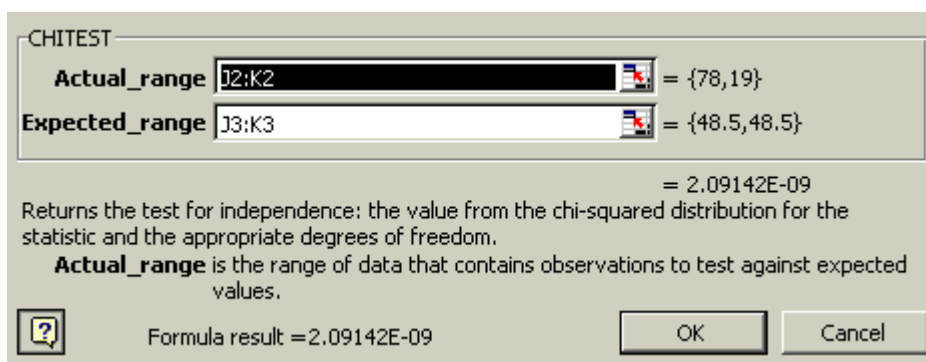


Funcția CHITEST: ia în considerare șirul de date din tabelul de contingență, calculează valoarea χ^2 și afișează probabilitatea asociată cu valoarea χ^2 .

Pentru aceasta trebuie ca datele din tabelul de contingență să fie astfel așezate încât toate valorile observate să fie într-un domeniu bine definit, și toate valorile așteptate să fie în alt domeniu, iar perechile corespondente de valori (observate și așteptate) să se afle în aceeași poziție în cele două domenii.

Folosiți funcția CHITEST pentru a testa dacă imbolnavirea de bronsita este asociată cu mediul rural sau cu cel urban:

- Verificați dacă tabelul de contingență îndeplinește cerințele de mai sus.
- Click pe icoana pentru formule  și alegeți CHITEST dintre formulele Statistice.
- Selectați domeniile pentru "Actual Range" (valorile observate) și "Expected Range" (valorile așteptate), apoi click OK:



- Rezultatul trebuie să fie egal cu cel obținut la aplicarea funcției CHIDIST.
- Nu uitați să salvați fisierul în care lucrați!

Testarea relației dintre îmbolnavirea de bronsită și vârsta

Testați dacă bronsita depinde de vârsta. Efectuați aceiași pași ca în cazul anterior, pentru a crea un tabel de contingență având ca și categorii vârsta (1 - tanar, 2 - adult sub 50 ani, 3 - adult peste 50 ani), apoi folosiți funcția CHITEST.

Testarea relației dintre îmbolnavirea de bronsită și existența antecedentelor în ultimul an

Testați dacă bronsita este relaționată cu existența antecedentelor.

Testarea relației dintre îmbolnavirea de bronsită și gen

Testați dacă între îmbolnavirea de bronsita și gen există o relație semnificativă din punct de vedere statistic.

Testul- χ^2 pentru două eșantioane

Investigați dacă există o diferență semnificativă între bronsita de iarnă și cea de vară. Acest test este un test pentru două eșantioane, deoarece comparăm caracteristicile îmbolnavirilor din iarnă cu cele din vară. De aceea este nevoie să se facă o numărătoare separată pentru iarnă și vară.

- Pentru a crea *tabelul de contingență*, datele trebuie sortate după sezon:
 - Copiați datele într-un nou worksheet și denumiți-l "summer-winter".
 - Folosiți mouse-ul pentru a selecta toate datele, inclusiv capul de tabel. Alegeți opțiunea "Data" din meniu, apoi "Sort". Verificați să fie bifată căsuța "Header Row", apoi alegeți în fereastra de sortare "Sort by" varianta 'Anotimp':
 - Creați un tabel al frecvențelor observate pentru fiecare varsta. Pentru a calcula frecvențele observate folosiți funcția COUNTIF. Tabelul obținut trebuie să arate ca cel de jos:

	tanar <18	adult<50	adult >50	Total
Iarna				
Vara				
Total				

- Completați *tabelul de contingență*, calculând valorile așteptate, Tabelul obținut trebuie să arate ca următorul:

	tanar <18	adult<50	adult >50	Total
Iarna (fo)				
Vara (fo)				
Total				
Iarna (fe)				
Vara (fe)				

- Folosiți funcția CHITEST pentru a obține valoarea probabilității corespunzătoare lui χ^2 calculat.
- Nu uitați să verificați dacă nu a fost încălcată una din regulile ce trebuie îndeplinite pentru a putea aplica testul χ^2 . Calculați gradul de libertate pentru testul pe care l-ați efectuat.