

## Testul $t$ împerecheat

Testele statistice „împerecheate” (*paired*) se aplică atunci când populațiile de comparat nu sunt independente (adică nu am luat, la întâmplare, 10 studenți de la ASE și medicină), ci i-am împerecheat – de obicei împerecherea se face pe baza altor factori de risc pentru efectul pe care îl studiem, pentru a fi mai siguri că diferența dintre TA ale studenților de la cele două facultăți este datorată facultății, și nu altor variabile, care în acest caz ar fi factori de confuzie (de exemplu, vârsta – care poate fi ceva mai mare la medicină, sexul, greutatea, consumul de sare etc.), și atunci, pentru fiecare student selecționat de la ASE cu anumite vârstă, sex, greutate, consum de sare, selecționăm câte un student de la medicină care să aibă caracteristici apropiate.

Baza de date arată la fel, atât în InStat, cât și în SPSS, numai că noi știm că datele sunt împerecheate, și atunci vom alege un test  $t$  (Student) împerecheat, iar rezultatele vor fi cele din Figura 1.

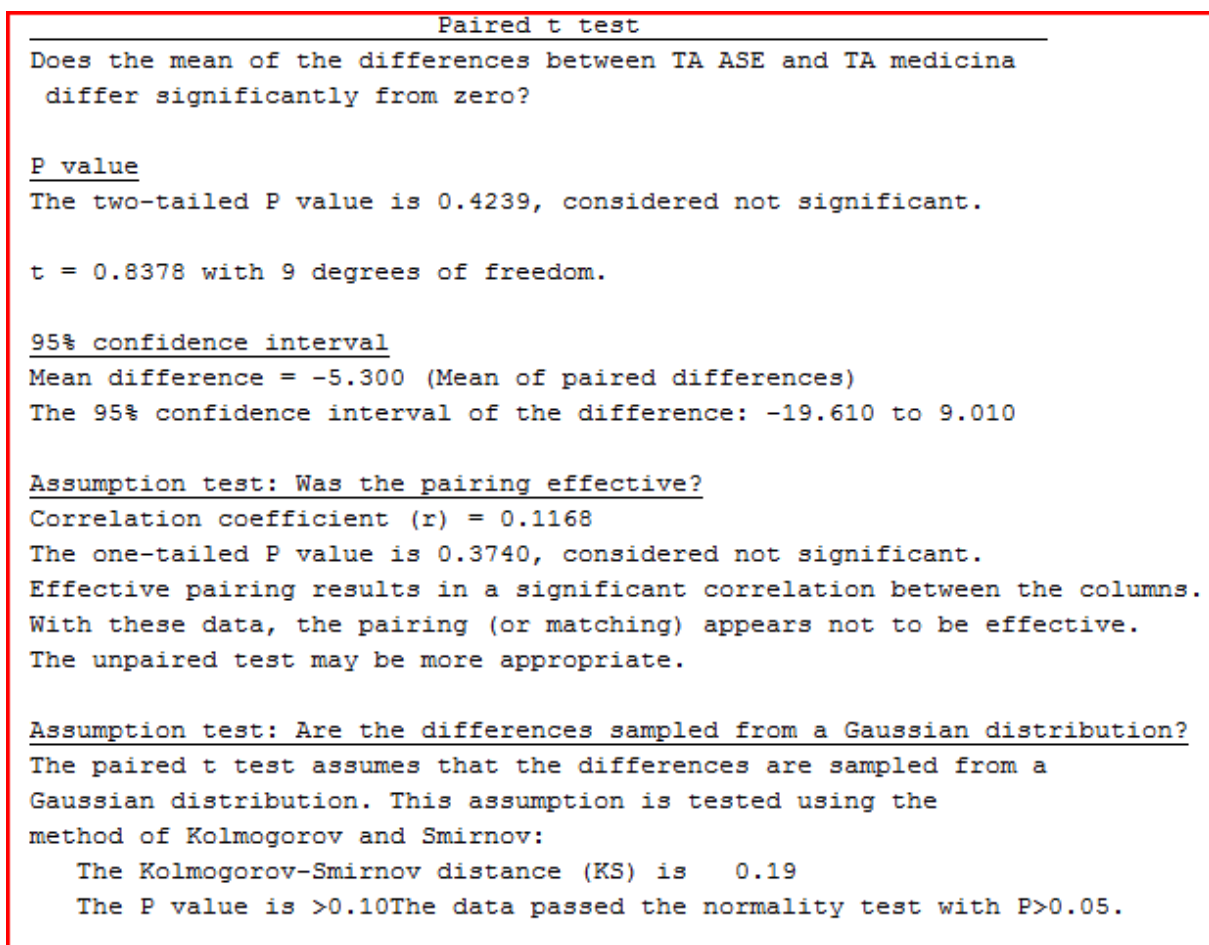


Figura 1. Rezultatele testului  $t$  împerecheat, InStat.

Pe aceleași date, după testul  $t$  împerecheat,  $p=0,42$ , deci nesemnificativ, dar puțin mai mic decât în cazul aplicării  $t$  neîmperecheat. Diferența este mică, însă în cazul eşantioanelor mai mari, diferența crește și ea, așadar testul împerecheat este mai puternic, iar atunci când datele sunt împerecheate, avem tot interesul să îl practicăm pe acesta.

Testul neîmperecheat compară TA (media și deviația standard) ale celor două eşantioane (ASE și medicină); drept dovadă, pentru a compara cele două eşantioane nici nu avem nevoie de datele fiecărui student, ci este de ajuns să știm mărimea eşantionului, media și deviația standard pentru cele două loturi pentru a obține  $p$  și intervalul de încredere al diferenței mediilor.

Pentru varianta împerecheată, în schimb, avem nevoie de datele tuturor studenților, pentru că testul pornește de la compararea diferențelor de TA din cadrul fiecărei perechi. De aceea, spre deosebire de testul neîmperecheat, în care nu avea importanță ordinea în care scriam valorile TA în baza de date, dacă vrem să aplicăm testul împerecheat trebuie să trecem în baza de date, pe verticală, valorile două câte două, așa cum au fost studenții împerecheați, câte unul de la fiecare facultate.

Împerecherea perfectă este atunci când împerechem fiecare subiect cu el însuși, și aceasta se realizează în cazul studiilor înainte/după. Să ne imaginăm, de exemplu, că nu comparăm între ei studenții de la două facultăți diferite, ci comparăm TA ale studenților de la medicină, înainte și după examenul de metodologia cercetării (sau înainte și după un tratament). Atunci fiecare subiect este relativ constant în privința tuturor factorilor de risc, cu excepția expunerii la stresul examenului. Baza de date este identică (numai variabilele se vor numi altfel, de exemplu TA înainte= $TA_0$ , și TA după= $TA_1$ ), și rezultatele sunt identice -  $p=0,42$ , intervalul de încredere 95% (CI95%) = (-19,6 , 9 mmHg).

Uitându-ne mai jos la rezultate, vedem că programul statistic testează și dacă împerecherea a fost eficientă. Pentru aceasta, ar trebui ca valorile perechilor să se coreleze, și acesta se poate stabili prin calculul coeficientului de corelație Pearson ( $r$ ); în cazul nostru,  $r$  este mic (deci valorile nu se corelează), lăsând la o parte faptul că nici nu este semnificativ statistic. Aceasta înseamnă că împerecherea pe care am realizat-o este total nesatisfăcătoare, dacă vorbim de primul exemplu, cu studenți de la două facultăți împerecheați pe baza celorlalți factori de risc pentru TA crescută (vârstă, sex, indice de masă corporală, consum de sare). Dacă ne referim la al doilea exemplu, în care am comparat fiecare student cu el însuși (înainte și după examen), aici nu aveam cum să nu obținem o împerechere bună și deci un coeficient de corelație mare, așadar datele nu sunt împerecheate, ceea ce este adevărat, pentru că studiul nu este real și datele le-am inventat eu pentru a vă da un exemplu!

NB. Mai multe despre corelație și  $r$  vom vorbi la analiza bivariată în cazul a două variabile numerice.