**TEMA 6. Testul T: compararea a două eşantioane de scoruri corelate / relaţionate, necorelate / nerelaţionate**

**Unităţi de conţinut**

* *Testul t – pentru eşantioanele-perechi.*
* *Testul t – pentru eşantioanele independente.*

**Termeni – cheie:** eşantioane perechi, eşantioane independente, testul t pentru eşantioane perechi şitestul t pentru eşantioane independente.

**La finele studierii temei masteranzii vor fi capabili:**

* să introducă datele și să formateze baza de date;
* să calculeze testul t pentru eşantioanele – perechi şi cele independente;
* să interpreteze rezultatele obţinute.

**Repere teoretice**

**Testul t pentru eşantioanele-perechi**

Testul t este folosit pentru evaluarea semnificaţiei statistice a diferenţei dintre mediile pentru două seturi de scoruri. Cu ajutorul testului t se pune în evidenţă, dacă valoarea medie pentru un set de scoruri diferă în medie de valoarea medie pentru alt set de scoruri.

Testul t are două variante:

1. *prima variantă* este folosită, atunci cînd cele două seturi de scoruri ce trebuie comparate provin dintr-un singur eşantion sau cînd coeficientul de corelaţie între cele două seturi este mare – *testul t, pentru eşantioane-perechi;*

2. *a doua variantă* a testului t este utilizată în momentul în care două seturi diferite de valori provin din grupe diferite de participanţi. – *testul t, pentru eşantioane independente.*

\* Testul t, pentru eşantioane perechi, este optim, dacă distribuţia diferenţelor dintre cele două seturi de valori se prezintă (aproximativ) sub formă de clopot (atunci cînd distribuţia este normală). Dacă distribuţia este foarte diferită de forma de clopot, ar trebui luată în considerare utilizarea unei tehnici statistice relaţionate nonparametrică pentru eşantioane-perechi, cum ar fi testul de perechi *Wilcoxon*.

*Procesarea unui Test t este ilustrată cu datele din tabelul următor unde sunt prezentate numărul de cuvinte pe care aceeaşi copii le-au verbalizat cu mamele lor la 18 luni şi la 24 luni. Numărul de cuvinte verbalizate într-un minut la vîrste diferite.*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Alina** | **Cristina** | **Radmila** | **Corina** | **Sabina** | **Radu** | **Marin** | **Dan** |
| **18 luni** | 3 | 2 | 4 | 5 | 2 | 3 | 2 | 4 |
| **24 luni** | 5 | 4 | 7 | 8 | 3 | 4 | 4 | 6 |

**Introducerea datelor.**

**Pasul 1:** În „Variable View” din „Data Editor” se etichetează primul rînd cu „optsprezece” şi al doilea rînd cu „douăzeci şi patru”.

\* Se înlătură cele două zecimale.

**Pasul 2:** În „Data View” din „Data Editor” se introduc datele în primele două coloane.

**Testul t – pentru eşantioane-perechi**

**Pasul 1:** Se selectează: „Analyze” → „Compare Means”→ „Paired-Samples T Test…”



**Pasul 2:** Se selectează „optsprezece” şi se pune această variabilă lîngă eticheta „Variable 1”, sub „Curent Selections”. Se selectează „douăzeci şi patru” şi se pune această variabilă lîngă eticheta „Variable 2”, sub „Curent Selections”. Se apasă butonul ►, pentru a pune aceste două variabile în lista variabilelor relaţionale. Se apasă „OK”.

****

**Interpretarea output-ului**

Primul tabel arată media, numărul de cazuri şi abaterea standard pentru cele două grupuri. Media pentru „optsprezece” este 3,13 şi abaterea standard este 1,126.

****

Al doilea tabel arată gradul în care cele două seturi de valori sunt corelate. Corelaţia dintre ele 0,94. Aceasta este o corelaţie mare, nivelul de semnificaţie fiind de 0,00.



Primele trei coloane care conţin cifre sunt componentele fundamentale ale calculului unui Test t relaţionat. Media de -2,000 este de fapt diferenţa dintre mediile pe 18 şi 24 luni, deci în realitate este media diferenţei. Valoarea lui t este bazată pe această medie a diferenţei (-2,00), divizată cu eroarea standard a mediei (0,267). Calculul oferă valoarea lui t (-7,483).

****

Rezultatele obţinute ar putea fi prezentate astfel: „Media numărului de cuvinte verbalizate la 18 luni diferă semnificativ de media cuvintelor verbalizate la 24 de luni”.

**Testul t – pentru eşantioanele independente**

Atunci când o investigaţie de tip statistic se efectuează pe un eşantion, orice rezultat obţinut are o valoare relativă, în sensul că datele respective nu numai că nu coincid cu cele referitoare la populaţie, dar nici măcar nu se poate şti cu certitudine care este diferenţa dintre cele două genuri de date, de vreme ce starea populaţiei este, de regulă, necunoscută. Teoria matematică a probabilităţilor oferă însă proceduri pentru evaluarea rezultatelor studiilor selective, permiţînd o estimare, în termeni de probabilitate, a marjei maxime de eroare ce se poate comite prin utilizarea mărimilor din eşantion în locul celor care caracterizează populaţia.

Testul t, în esenţă, este o procedură de testare a semnificaţiei diferenţei dintre două medii. Ca urmare, el este potrivit, atunci cînd variabila dependentă este măsurată pe scală cantitativă (interval-raport). Distribuţia teoretică de referinţă (distribuţia de nul) este cea normală, pentru eşantioane mai mari de 30 de subiecţi, şi distribuţia t (Student), pentru eşantioane mai mici de 30 de subiecţi. Chiar dacă formulele de calcul sunt diferite, forma de prezentare a rezultatelor şi raţionamentul decizional sunt similare pentru ambele situaţii.

Testul t, pentru eşantioane independente, este utilizat pentru a calcula dacă mediile pentru două seturi de variabile sunt diferite semnificativ una faţă de cealaltă.

\* Testul t pentru eşantioane independente este cel mai des folosit.

Testul t, pentru eşantioane independente, este utilizat, atunci cînd cele două seturi de variabile provin din două eşantioane diferite de oameni.

*Procesarea unui Test t, pentru eşantioane independente, este ilustrată cu datele din tabelul următor, care arată valorile la un test de emotivitate pentru 10 copii care provin din familii biparentale şi 10 copii care provin din familii monoparentale. Scopul analizei este de a aprecia, dacă valorile emotivităţii sunt diferite la copii care provin din familii cu doi părinţi faţă de copii care provin din familii monoparentale****.***

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Fam. biparentale** | **12** | **18** | **14** | **10** | **19** | **8** | **15** | **11** | **10** | **13** |
| **Fam. monoparentale** | **6** | **9** | **4** | **13** | **14** | **9** | **8** | **12** | **11** | **9** |

**Introducerea datelor**

**Pasul 1:** În „Variable View” din „Data Editor”, se etichetează pe rînd „Familii”. Aceasta va defini cele două tipuri de familii. Se etichetează al doilea rînd „Emotivitate” (aici vor fi introduse rezultatele la testul de emotivitate).

****

Se înlătură cele două zecimale.

**Pasul 2:** În „Data View” din „Data Editor” se introduc valorile pentru cele două variabile în primele două coloane.

\* Se observă că sunt două coloane de date: A doua coloană conţine cele 20 de valori ale testului de emotivitate de la ambele grupe de copii. Datele nu sunt păstrate separat pentru cele două grupe; valorile 1, din prima coloană, indică copii proveniţi din familiile biparentale şi valorile 2 – indică copii proveniţi din familiile monoparentale. Astfel, este utilizată o singură variabilă dependentă (în cazul nostru „emotivitatea”) şi altă coloană – pentru variabila independentă (familia). Cu alte cuvinte, fiecare rînd în parte reprezintă un anumit copil şi variabilele sale dependente şi independente sunt introduse separat în „Data Editor”.

****

**Efectuarea Testului t, pentru eşantioane independente**

**Pasul 1:** Se selectează: „Analyze” → „Compare Means” → „Independent Samples T test…”



**Pasul 2:** Se selectează „Emotivitate” şi apoi se apasă butonul ►, pentru a introduce această variabilă în lista variabilelor de test. Se selectează „Familie” şi apoi se apasă butonul ◄, pentru a introduce această variabilă în căsuţa „Grouping Variable”.

Se selectează „Define Groups…”, pentru a defini cele două grupuri.

****

**Pasul 3:** Se introduce:

- valoarea 1 (codul pentru familiile biparentale) alături de eticheta „Group 1”;

- valoarea 2 (codul pentru familiile monoparentale) alături de eticheta „Group 2”.

Se selectează „Continue”. Se apasă „OK” din ecranul precedent, care reapare.



**Interpretarea output-ului.**

Primul tabel arată, pentru fiecare grup în parte, numărul de cazuri, media şi abaterea standard. Media pentru familiile biparentale este 13,00. După cum se observă, există o diferenţă între cele două tipuri de familii, dar, întrebarea este dacă mediile diferă semnificativ.



Valoarea lui t – este media diferenţei (3,500) divizată cu eroarea standard a diferenţei (1,493), diviziune ce produce valoarea 2,345.



Dacă valoarea semnificaţiei pentru Testul Levene este mai mare de 0,05, lucru care se întamplă aici (0,642), se foloseşte informaţia de pe acest prim rînd. Dacă valoarea semnificaţiei pentru Testul Levene este mai mică de 0,05, se foloseşte informaţia de pe al doilea rînd (al doilea rînd oferă cifrele pentru cazul în care varianţele sunt diferite semnificativ).

Pentru varianţe egale, t este 2,345, care la 18 grade de libertate este semnificativ la 0,031, pentru nivelul two-tailed.

Rezultatele obţinute se pot raporta astfel: „Media pentru valorile testelor de emotivitate ale copiilor ce provin din familii cu doi părinţi este semnificativ mai mare decît cea a copiilor proveniţi din familiile cu un singur părinte”.

Dacă se preferă folosirea intervalelor de încredere, se poate scrie: „Diferenţa dintre valorile testelor de emotivitate ale copiilor ce provin din familii cu doi părinţi (M = 13,00, SD = 3,55) şi cei proveniţi din familii cu un singur părinte (M = 9,50, SD = 3,10) este de 3,50. Intervalul de încredere de 95% pentru această diferenţă este de la 0,36 la 6,63. Deoarece intervalul nu conţine punctul 0,00, diferenţa este statistic semnificativă la nivelul de semnificaţie two-tailed, de 5%.