

**PROIECT ANALIZA MULTIDIMENSIONALA**

Student: Zipisi Vlad-Andrei

Grupa 1096

Prof. coordonator: Prof.univ.dr. Cristina Boboc

**Cuprins**

Contents

[**Introducere** 3](#_Toc91594857)

[**Descrierea bazei de date si a componentelor acesteia** 3](#_Toc91594858)

[**Prelucrarea datelor** 4](#_Toc91594859)

[**Analiza descriptiva a variabilelor folosite in cele doua metode** 4](#_Toc91594860)

[**Analiza cluster** 5](#_Toc91594861)

[Fig.1.Variabilele folosite in analiza cluster 5](#_Toc91594862)

[Fig.2.Cluster Summary 5](#_Toc91594863)

[Fig.3. Statistics for variables Fig.4. Cluster Means 6](#_Toc91594864)

[**Analiza discriminanta** 6](#_Toc91594865)

[Fig.5.Codificarea variantelor de raspuns 7](#_Toc91594866)

[Fig.6.Transformarea tipului variabilei 7](#_Toc91594867)

[Fig.7.Noile variabile 8](#_Toc91594868)

[Fig.8.Variabilele din analiza discriminanta 8](#_Toc91594869)

[Fig.9.Stepwise Selection Summary 9](#_Toc91594870)

[Fig.10.Output clasificare 9](#_Toc91594871)

[Fig.11.Simple Statistics 10](#_Toc91594872)

[**Concluzii** 11](#_Toc91594873)

# **Introducere**

Analiza de fata are ca tema impactul sistemului online educational asupra elevilor din Bangladesh, avand la dispozitie mai multe variabile ce pot motiva sau explica diferite comportamente.

Obiectivele propuse sunt de a obtine componente noi formate din unitatile din cadrul esantionului nostru, aceste componente fiind de dimensiuni reduse. Astfel dorim sa grupam studentii, privind orele alocate pentru studii inainte si dupa noul sistem educational, cel online.

Totodata ne-am propus sa observam comportamentul persoanelor ce si-au imbunatatit notele dupa impunerea noului sistem educational.

Baza de date folosita pentru realizarea acestei analize este preluata de pe site-ul Kaggle.com, mai precis de la adresa <https://www.kaggle.com/jehanbhathena/online-survey-data-of-bangladeshi-students>.

# **Descrierea bazei de date si a componentelor acesteia**

In cadrul acestui set de date sunt descries 8783 de observatii, aceste observatii fiind obtinute in urma unui sondaj. Aceasta baza de date este alcatuita din raspunsuri date de participanti la saptesprezece intrebari prezentate in lista de mai jos:

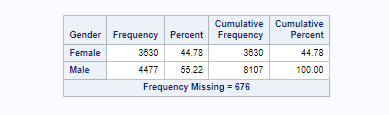
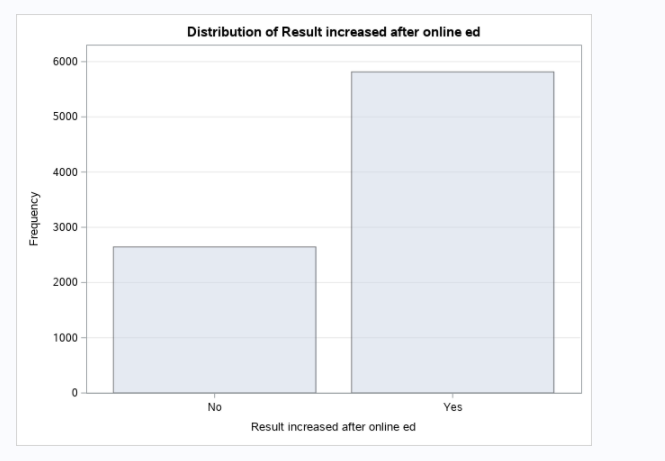
* Ati folosit un smartphone/computer/laptop inainte de cursurile online?
* Notele au crescut dupa ce sistemul online a fost impus?
* Credeti ca v-ati dezvoltat cunostintele dupa trecerea la sistemul online educational?
* Sunteti fericit cu sistemul online educational?
* Aveti acces la internet?
* Vi s-a imbunatatit performanta scolara in noul sistem?
* Ati intampinat vreo problema in timpul cursurilor online?
* Nivelul de studii – Higher Secondary Certificate (**HSC**) sau Honours degree
* Varsta
* Zona institutului educational – urban sau rural
* La ce fel de internet aveti acces – o retea cu performanta ridicata sau internet mobil
* Numarul de ore pe care l-ati alocat pentru studii inainte de noul sistem
* Numarul de ore pe care l-ati alocat pentru studii dupa noul sistem
* Tipul institutului – public sau privat
* Locatia curenta (in timpul studiilor) - urban sau rural
* Genul – masculin sau feminin
* Device-ul preferat pentru cursurile online – computer sau mobile

# **Prelucrarea datelor**

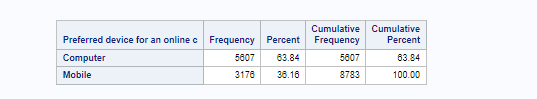
Aici se vor aplica cele doua metode de analiza mutlidimensionala pentru indeplinirea obictivelor mentionate mai sus.

## **Analiza descriptiva a variabilelor folosite in cele doua metode**

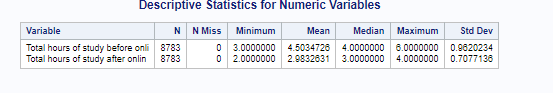
In baza noastra de date, 31.26% dintre studenti au raspuns negativ la intrebarea despre imbunatatirea notelor dupa cursurile online, iar 68.74% au raspuns pozitiv.



Dintre persoanele ce au raspuns la variabila gen, 3630 de studenti sunt femei, iar 4477 sunt barbati, restul de 676 nu au completat o valoare.



In ceea ce priveste device-ul preferat folosit pentru studiile online, 5607 studenti utilizeaza un computer, restul folosind telefonul.



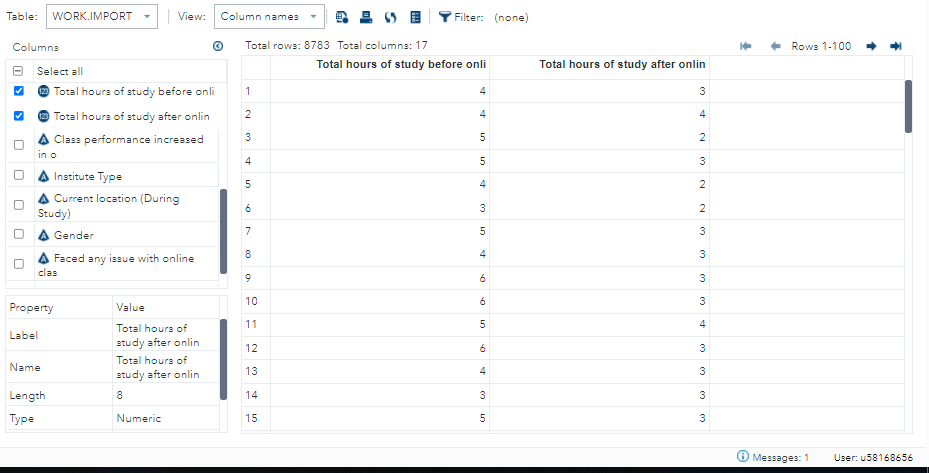
Cele doua variabile de mai sus sunt omogene, pentru ca au coeficientul de variatie sub 30-35%, adica mediile sunt reprezentative pentru acestea.

## **Analiza cluster**

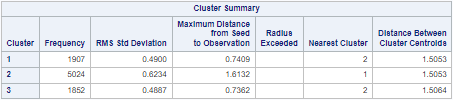
Prima metoda este folosita pentru a grupa esantionul de studenti in trei clusteri diferiti, dupa numarul de ore pe care acestia il petrec pentru studii inainte si dupa implementarea noului sistem educational. Aceasta metoda are ca scop vizualizarea impactului sistemului online educational asupra orelor oferite studiilor, ci anume daca exista studenti ce au acordat mai mult timp sau mai putin timp educatiei.

Construirea celor trei clusteri va fi realizata cu ajutorul a doua variabile (numarul de ore alocat pentru studii inainte de noul sistem, numarul de ore alocat pentru studii dupa noul sistem)

Cum cele doua variabile folosite au aceeasi scala de masura (vezi fig.1.), nu este nevoie de o standardizare.

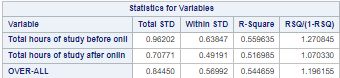
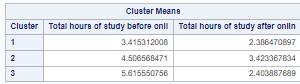


## Fig.1.Variabilele folosite in analiza cluster



## Fig.2.Cluster Summary

In fig.2. vedem cate observatii are fiecare cluster, astfel aflam ca primul si al treilea cluster au dimensiuni relative apropiate, in timp ce al doilea cluster are cea mai mare frecventa, ci anume 5024 de studenti. Clusterul 2 estel cel mai apropiat vecin pentru celelalte doua clustere, iar clusterul 2 are ca vecin pe clusterul 1.

## Fig.3. Statistics for variables Fig.4. Cluster Means

Din fig.3., analizand R-Square, putem spune ca cele doua variabile au o influenta relativ asemanatoare in realizarea celor trei clustere. Din tabelul alaturat putem observa caracteristicile fiecarui cluster. Primul cluster se diferentiaza de celelalte prin cele mai mici valori medii de ore dedicate studiului inainte si dupa sistemul online comparativ cu celelalte clustere, cu o diferenta intre cele doua variabile aproximativ egala cu cea din al doilea cluster. Putem afirma ca in primele doua clustere studentii si-au micsorat numarul de ore in medie cu o ora, dupa aparitia sistemului online.

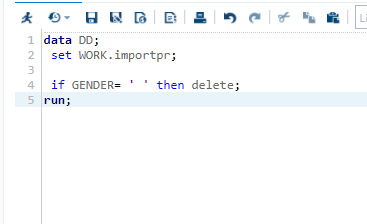
Clusterul 3 se remarca prin cel mai mare numar mediu de ore alocat studiilor inainte de trecerea in online, dar si cea mai mare micsorare a acestuia dupa noul sistem. O concluzie pe care o putem trage din aceasta analiza este ca noul sistem a avut un impact semnificativ asupra timpului alocat de catre studenti pentru educatie , acesta micsorandu-se; 1852 de studenti au inregistrat o micsorare in medie de 3.2 ore, in timp ce restul au inregistrat o micsorare nu atat de ingrijoratoare de aproximativ o ora in medie.

## **Analiza discriminanta**

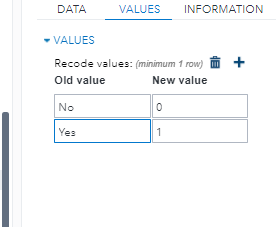
Cea de-a doua metoda, analiza discrimanta, este folosita pentru descrierea studentilor care au observat o imbunatatire asupra notelor obtinute. Dorim sa prezicem criteriile indeplinite de studentii care au note mai mari dupa noul sistem educational, astfel in model folosim drept variabila pentru previzionat raspunsurile de la intrebarea “notele au crescut dupa noul sistem educational?”, cu predictorii:

* numarul de ore pe care l-ati alocat pentru studii inainte de noul sistem
* numarul de ore pe care l-ati alocat pentru studii dupa noul sistem
* device-ul preferat pentru cursurile online
* sunteti fericit cu sistemul online educational?
* Genul

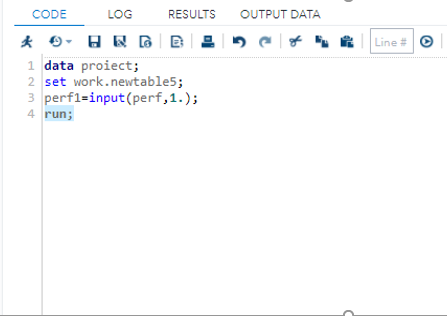
Cum in baza noastra de date exista inregistrari unde nu s-a raspuns la variabila gen, am ales sa stergem aceste randuri, folosind scriptul de mai jos, astfel obtinand 8107 de inregistrari.



Stergerea inregistrarilor acolo unde lipsesc valorile despre gen

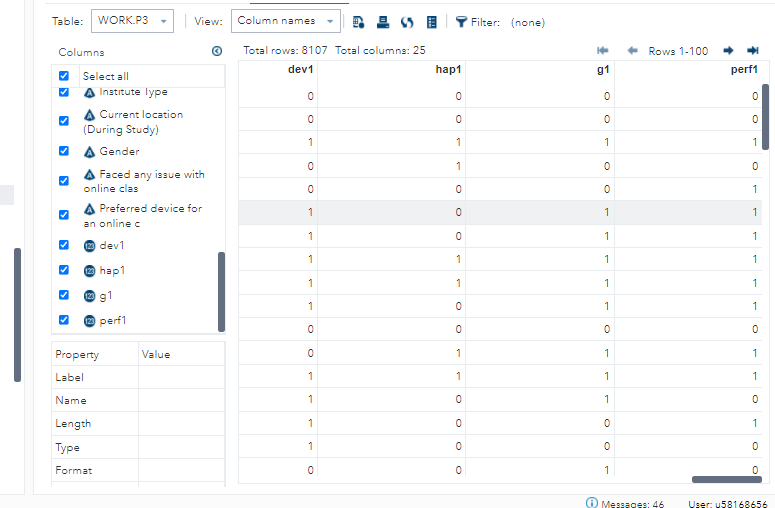


## Fig.5.Codificarea variantelor de raspuns



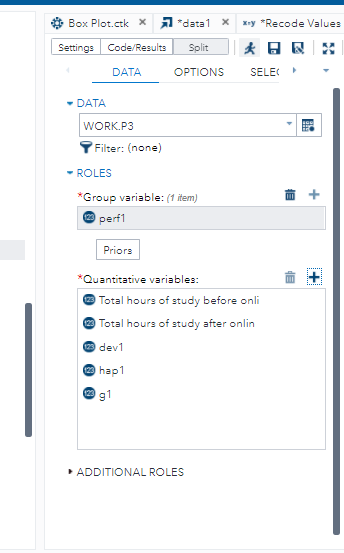
## Fig.6.Transformarea tipului variabilei

Am pornit aceasta analiza prin recodificarea variabilelor, astfel pentru variabilele cu variantele de raspuns da si nu am atribuit 0 si 1, exact ca in figura nr.5, pentru variabila device-ul preferat am codificat varianta de raspuns mobil cu 0 si computer cu 1, iar pentru gen am folosit 0 si 1 reprezentand masculin si feminin. Variabilele cu numarul de ore au ramas neschimbate. Al doilea pas a fost sa transformam variabilele caracter in variabile numerice, pentru acest lucru folosind scriptul din figura nr.6, astfel am obtinut cele patru noi variabile din figura nr.7: dev1, hap1, g1 si perf1 (variabila care se doreste a fi previzionata).

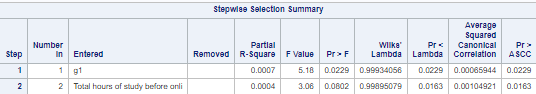


## Fig.7.Noile variabile

Prin urmarare vom aplica metoda analizei discriminante pe noul set de date exact ca in fig.8.

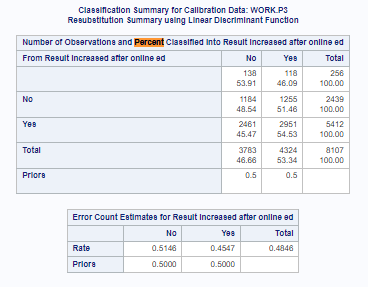


## Fig.8.Variabilele din analiza discriminanta



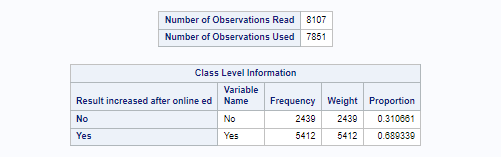
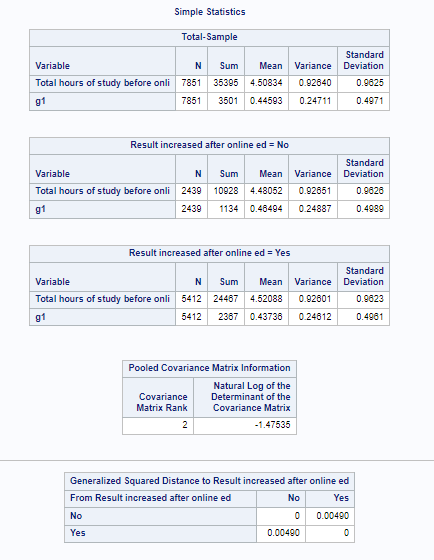
## Fig.9.Stepwise Selection Summary

Din cele 5 variabile incluse in analiza, semnificative pentru impartirea in cele doua grupe sunt genul si numarul de ore de studiu alocat inainte de sistemul online, ce este mai putin semnificativ (p-value este 0.08 > 0.05).



## Fig.10.Output clasificare

Din fig.10. putem spune ca 51.54% dintre studentii din model sunt bine clasificati per total, din cei ce au observat imbunatatiri asupra rezultatelor 55% sunt bine identificati, iar dintre cei ce nu au evoluat in privinta notelor 49% sunt bine identificati. Astfel variabilele folosite in model contribuie intr-o masura putin mai ridicata la identificarea celor ce au inregistrat o imbunatatire a notelor dupa sistemul educational online.



## Fig.11.Simple Statistics

Numarul de ore alocate pentru studii sunt in medie aproximativ egale pentru cei ce au observat rezultate mai mari dupa trecerea in online si cei care nu. In randul studentilor cu imbunatatiri, exista putin mai multi de gen masculin, ci anume 56.3%.

Daca nu am fi aplicat analiza discriminanta, procentul studentilor ce obtin imbunatatiri in privinta notelor este de 68.93%, in cazul in care o folosim vom identifica corect 55% dintre acestia, reusind sa identificam corect si 49% din cealalta categorie. Totusi modelul nu este unul performant.

# **Concluzii**

In analiza cluster s-a constatat ca timpul pe care studentii din esantion l-au alocat pentru studii dupa cursurile online s-a micsorat. O parte din studenti au inregistrat o miscorare de o ora in medie, iar 1852 de student au alocat in medie cu peste 3 ore mai putin studiilor.

Din analiza discriminanta am observat ca variabilele din baza noastra de date nu sunt cele mai sugestive in ceea ce priveste previzionarea studentilor ce au obtinut note mai mari in timpul cursurilor online, dar putem afirma ca cei ce au un raspuns pozitiv la intrebarea despre note au avut si un numar total mediu de ore alocat pentru studii inainte de cursurile online putin mai mare decat cei ce au raspuns negativ. Tot aici am aflat ca 55% din studentii ce au raspuns pozitiv au fost identificati corect. Un motiv pentru cresterea notelor ar putea fi si incidenta crescuta a comportamentului de trisare la examene si teste, acest lucru fiind mult mai greu detectat in mediul online.