

Academia de Studii Economice din București  
Facultatea de Cibernetică, Statistică și Informatică Economică  
Analiza afacerilor și controlul performanței întreprinderii



**PROIECT**  
***ȘTIINȚA DATELOR ÎN AFACERİ***  
**Sistemul educațional la nivel mondial**

**Studenți**

Buzescu Raluca - Marinela, 1116  
Burcuță Ștefan – Gabriel, 1116  
Cătălina Andra – Gabriela, 1116  
Ciocâlțeanu Alexandra, 1116

**Prof. Coordonator**

Bizovi Mihai

București, 2023

## ***Curpins***

<i>1. Prezentarea temei alese .....</i>	<i>3</i>
<i>1.1. Importanța problemei .....</i>	<i>3</i>
<i>1.2. Studiile efectuate anterior în domeniul educației .....</i>	<i>4</i>
<i>1.3. Metodologie .....</i>	<i>4</i>
<i>1.4. Datele problemei .....</i>	<i>4</i>
<i>2. Prelucrarea datelor .....</i>	<i>5</i>
<i>2.1. Calcularea și interpretarea statisticilor descriptive .....</i>	<i>6</i>
<i>3. ACP. Extragerea componentelor principale .....</i>	<i>8</i>
<i>3.1. Grafice – pachetul factoextra .....</i>	<i>10</i>
<i>3.2. Criterii pentru determinarea numărului optim de clase – k .....</i>	<i>11</i>
<i>4. Concluzii .....</i>	<i>12</i>

## **1. Prezentarea temei alese**

Am ales această tematică, întrucât este un subiect de interes major la nivel mondial. Pe parcursul evoluției umane, s-a constatat faptul că educația reprezintă baza unei societăți adaptabile și capabile să se dezvolte. Pentru a evidenția principalele curențe ale sistemului educațional, în vederea ameliorării acestora, am identificat o serie de **13 indicatori**.

### **1.1. Importanța problemei**

Educația este un proces gradual de învățare și cultivare a noilor cunoștințe, abilități și valori. Aceasta are rolul de a aduce schimbări pozitive în viața și comportamentul uman, facilitând dezvoltarea unor calități umane și explorarea de noi orizonturi cognitive. Educația înseamnă, de asemenea, să cultivăm valori sociale, învățându-i pe copii și adulți să fie cetățeni activi și eficienți.

Dreptul la educație este un drept fundamental al omului. Fiecare persoană, indiferent de rasă, gen, naționalitate, origine etnică sau socială, religie sau preferințe politice, are dreptul la o educație elementară gratuită. Cu toate acestea, educația rămâne un drept de care nu se pot bucura foarte mulți copii din întreaga lume. Peste 72 de milioane de copii cu vârste între 6 și 12 ani nu merg la școală.

Educația îmbunătățește șansele de reușită ale unui individ în viață și ajută la combaterea sărăciei. De asemenea, educația are un rol esențial în dezvoltarea unei țări, determinând comportamentul oamenilor, modul de gândire al cetățenilor, dar și calitatea sistemelor politice și sociale ale statului. Educația de calitate este fundamentul unei societăți durabile și rezistente, contribuind la realizarea tuturor celorlalte obiective de dezvoltare durabilă.

### **1.2. Studiile efectuate anterior în domeniul educației**

În acest punct al proiectului vom analiza studiile despre educație. Un prim studiu despre care dorim să discutăm este „Global Citizenship Education”, un proiect realizat de European Academy și finanțat de către Uniunea Europeană. În acest studiu se analizează două esanțioane de tineri: primul esanțion este compus din tineri cu studii superioare iar cel de-al doilea, tinerii care au absolvit învățământul obligatoriu. Rezultatul a fost unul evident, și anume, cei care faceau parte din primul esanțion au avut mai multe posibilități de angajare și implicit, o deschidere către o viață mult mai bună în comparație cu cei care faceau parte din al doilea esanțion.

Al doilea studiu despre care dorim să discutăm este realizat chiar de Ministerul Educației Naționale care atinge diferite obiective: rolurile educației, valorile fundamentale, principii fundamentale, beneficiarii serviciului public de educație, obiectivele serviciului public de educație, scenariul propus pentru implementarea viziunii și obiectivele generale ale fiecărei etape din parcursul școlar.

### **1.3. Metodologie**

Metodologia folosită în cadrul proiectului cuprinde următoarele tehnici:

- **Statistica descriptivă** este, împreună cu inferența statistică sau statisticile inferențiale, una dintre cele două mari ramuri ale statisticii. Numele propriu îl indică, încearcă să descrie ceva. Dar nu o descrie în nici un fel, ci într-un mod cantitativ. Luați în considerare greutatea unei cutii de legume, înălțimea unei persoane sau suma de bani câștigată de o companie. Am putea spune multe despre aceste variabile. De exemplu, am putea indica faptul că această cutie de roșii cântărește mult sau cântărește mai puțin decât altele. Continuând cu un alt exemplu, am putea spune că veniturile unei companii variază foarte mult în timp sau că o persoană are o înălțime medie.
- **Corelația. Corelații grafice** se referă la o clasă largă de relații statistice care implică dependența. Exemple familiare de fenomene dependente includ corelații dintre statura fizică a părinților și a copiilor lor, precum și corelații între curba cererii a unui produs și prețul acestuia. Pentru studiul dependenței dintre două variabile, fiecare dintre ele fiind supusă la o împrăștiere aleatoare, se aplică metode de analiză a corelației. Analiza corelației studiază legea medie de comportare a fiecăreia dintre variabile în funcție de valorile celeilalte variabile, precum și măsura dependenței dintre variabilele considerate
- **Analiza componentelor principale.** În recunoașterea formelor, selecția și extragerea caracteristicilor reprezintă o alegere decisivă pentru proiectarea oricărui clasificator. Selecția caracteristicilor poate fi văzută și ca un proces de „compresie de date”, fiind asimilată cu o transformare liniară din spațiul inițial al observațiilor într-un spațiu cu mai puține dimensiuni. O astfel de transformare este necesară deoarece poate conserva o mare parte din informație și permite aplicarea unor algoritmi în timp real, algoritmi eficienți fiind doar în spații cu dimensiuni reduse.
- **Pachetul factoextra** este pachetul cu ajutorul caruia putem într-un mod mult mai ușor să extragem și să vizualizăm output-uri de analiză a datelor
- **Arbore decizionali.** Arborele de decizie reprezintă un instrument complex pentru vizualizarea și evaluarea alternativelor în procesul selecției deciziei. El furnizează o structură eficientă, în care se pot detalia gradat toate alternativele, în funcție de toate opțiunile posibile, obținându-se în final rezultatele asociate cu fiecare curs posibil al alternativelor.

#### 1.4. Datele problemei

- **Ratele de înregistrare în ciclurile: preprimar, primar, secundar și terțiar** sunt indicatori ce reflectă numărul de elevi și studenți înscriși pentru studii raportate la vârsta convențională. Conform obiectivului analizei, aceștia trebuie să aibă valori cât mai mari, pentru a arăta un progres în rata de emancipare din cadrul unei populații, însă pentru a fi relevanți, trebuie să îi corelăm corespunzător cu următorii indicatori.
- **Repetenții din ciclul primar, Rata de abandon școlar printre copii, Rata de abandon școlar printre adolescenți, Rata de analfabetism** sunt indicatori ai regresului educațional. Primii indicatori sunt irelevanți în procesul de sporire a educației, dacă aceste rate sunt crescute.
- **Rata de absolvire a ciclului primar/ a ciclului secundar și Numărul obligatoriu de ani de studiu** sunt indicatori esențiali pentru conturarea analizei. Cu ajutorul lor, identificăm nivelul de pregătire al studenților, implicit eficiența învățământului, dar și

numărul mediu de ani obligatoriu în școli, pentru obținerea unei performanțe cât mai ridicate.

- **Investițiile guvernamentale distribuite pentru sistemul educațional** de pe teritoriul unei țări sunt semne ale unui învățământ tehnologizat, modernizat, ale unor condiții de studiu bune, dar, în prim plan, ale unui interes național crescut pentru educație. După părerea mea, cu cât investițiile sunt mai mari, cu atât interesul pentru învățământ este mai mare, iar societate este mai prosperă.

Adjusted net enrollment rate, primary (% of primary school age children)
Repeaters, primary, total (% of total enrollment)
School enrollment, secondary (% gross)
School enrollment, tertiary (% gross)
School enrollment, preprimary (% gross)
Primary completion rate, total (% of relevant age group)
Progression to secondary school (%)
Government expenditure on education, total (% of GDP)
Literacy rate, adult total (% of people ages 15 and above)
Adolescents out of school (% of lower secondary school age)
Children out of school (% of primary school age)
Compulsory education, duration (years)

## 2. Prelucrarea datelor

**Primul pas** este eliminarea spațiilor libere din setul de date extras. Identificăm țările care nu au valori corespunzătoare și ștergem înregistrările. În **cazul** în care există un număr mic de înregistrări goale pe o coloană, le vom înlocui cu media (calculată cu formula AVERAGE) pe coloana respectivă.

C3

Fig.1 Faza inițială a datelor în Excel

**Al doilea pas** este eliminarea outlieri-lor(=valorile aberante, valori ale datelor care diferă foarte mult de majoritatea unui set de date). Pentru detectarea și eliminare outlieri-lor, folosim un interval de încredere calculat cu ajutorul funcțiilor Excel:

Medie	95.47391045	1.559441979	96.29477322	96.483
Abatere standard	3.539800051	1.237210938	13.36545475	3.4238
Medie-3AbatereST				
Medie+3AbatereST	84.85451029	-2.152190835	56.19840897	86.211
	106.0933106	5.271074794	136.3911375	106.7

Fig.2 Intervalul de încredere pentru indicatorii aleși

Modul de indentificare al outlieri-lor în Excel este exemplificat în Fig. 3 și Fig. 4.

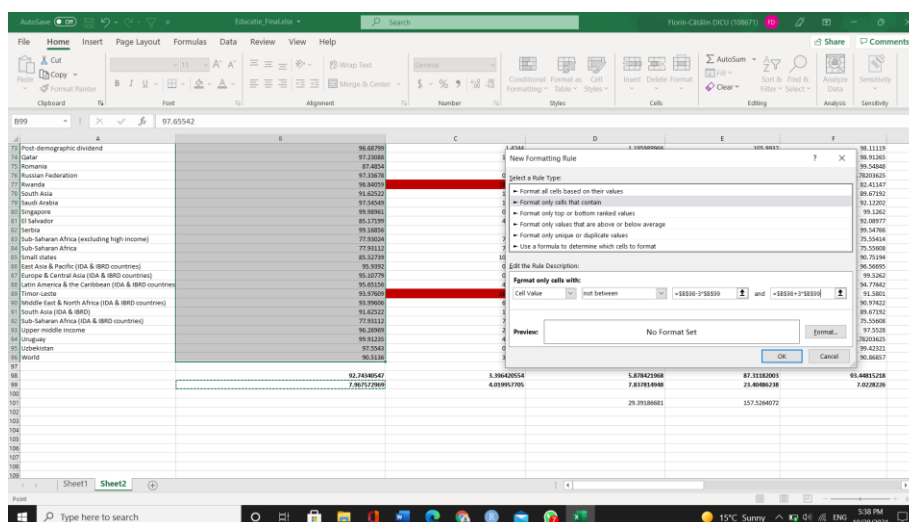


Fig.3 Metoda de aplicare a intervalului de încredere în Excel

Country Name	Adjusted net enrollment rate, primary (% of primary school age children)	Repeaters, primary, total (% of total enrollment)	Children out of school (% of primary school age)	School enrollment, secondary (% gross)	Progression to secondary school (% Primary)
1 Albania	91.6142	0.3819	2.94079935	95.7499934	96.1712
2 Arab World	84.5158	6.17746	15.5500609	76.79079	88.5012
3 Argentina	99.4406	2.34039	0.43239983	100.402337	99.85158
4 Armenia	82.4934	0.7611	7.54226139	86.894041	96.09919
5 Austria	88.8771	2.88471	0.39600007	100.963029	99.76727
6 Azerbaijan	86.8639	0.14844	0.796239991	98.8980441	97.16762
7 Burkina Faso	75.8239	5.60723	24.14706993	55.78713134	76.0944
8 Bulgaria	89.40128	0.07805	10.2520928	99.6777025	99.13268
9 Bahrain	97.9826	0.48922	1.25599047	100.838437	99.3796
10 Belarus	96.31791	0.03289	0.51811988	104.709213	98.01376
11 Belize	93.287	6.8446	0.52770987	97.0801032	96.20051
12 Bolivia	89.84287	2.24885	10.01541996	91.4283923	96.59641
13 Brunel Darussalam	96.6863	0.60008	0.15711993	94.0417631	99.78679
14 Cape Verde	86.9038	12.46839	12.1813794	85.0580806	91.38152
15 Cameroon	94.483	11.7888	4.57844821	80.0581887	91.7820423
16 Colombia	95.6489	1.186	6.17513213	95.0480138	95.7820423
17 Costa Rica	98.3878	3.2392	1.43397008	125.8581621	85.28271
18 Czech Republic	96.8817	0.8612	0.18979911	104.448114	99.36707
19 Germany	90.62417	0.47731	0.51420975	98.8119946	99.5035
20 Djibouti	96.098	8.29897	0.61977	88.1104835	81.9042
21 Denmark	97.9805	0.87805	125.1339716	125.1339716	99.36442
22 Dominican Republic	95.13993	3.61024	4.57922967	79.80252694	92.58356
23 Europe & Central Asia (excluding high income)	96.19937	0.67608	2.86180053	96.84906	99.11496
24 Europe & Central Asia	96.71136	1.27032	1.98434996	104.7957	98.81243
25 Ecuador	90.0649	0.79422	1.87115959	103.701297	99.0281
26 Egypt, Arab Rep.	86.44529	3.10996	6.13300002	85.1484823	94.39786
27 Euro area	98.14906	2.48095	1.03091964	100.1946	96.49019
28 Spain	96.3337	2.99351	0.901349852	126.3506602	99.71225
29 Estonia	93.88065	0.53405	2.81209932	115.034122	99.82892
30 European Union	97.6283	1.27386	1.43839977	133.1485	97.30759
31 Finland	99.12084	0.26785	0.72910008	103.848073	99.89739
32 Greece	97.84221	0.57318	1.51870029	102.004841	98.93364
33 Guatemala	87.12785	6.34544	12.70100018	53.45309318	89.47884
34 Heavily indebted poor countries (HIPC)	90.54475	9.11131	17.42207985	80.5532	78.71728
35 HIPC only	96.2457	1.84777	3.23968008	86.2729	94.2670
36 IDA & IBRD total	89.78409	3.73666	8.877200127	75.95572	90.04418

Fig 4. Outlieri identificați

În urma acestor pași, am **importat** datele în RStudio, folosind comenzile:

```
library(readxl)
education <- read_excel("C:/Users/pc/Desktop/An I, sem I/SDA/education.xlsx")
```

## 2.1. Calcularea și interpretarea statisticilor descriptive

Statisticile descriptive au fost calculate cu ajutorul funcției `summary`, cu scopul de a afla informații despre setul de date importat.

```
> summary(education)
Country Name      ER_primary  Repeaters_primary  ER_secondary  Progression_secondary  Primary_completionRate
Length:67      Min.   :85.17      Min.   :0.00208      Min.   : 67.31      Min.   :87.54      Min.   : 87.68
Class :character 1st Qu.:93.71      1st Qu.:0.64116      1st Qu.: 86.95      1st Qu.:93.79      1st Qu.: 93.14
Mode  :character Median :96.36      Median :1.27052      Median : 97.54      Median :98.01      Median : 97.34
              Mean :95.47      Mean :1.55944      Mean : 96.29      Mean :96.48      Mean : 97.44
              3rd Qu.:97.98      3rd Qu.:2.24472      3rd Qu.:105.01      3rd Qu.:99.35      3rd Qu.:100.67
              Max.   :99.99      Max.   :4.63183      Max.   :129.12      Max.   :99.97      Max.   :111.30

Adolescents_OutofSchool  Children_OutofSchool  ER_tertiary  Government_expenditure  Compulsory_education  ER_preprimary
Min.   : 0.05647      Min.   : 0.00014      Min.   : 8.464      Min.   :2.758      Min.   : 6.00      Min.   : 22.21
1st Qu.: 1.48812      1st Qu.: 0.55615      1st Qu.:39.879      1st Qu.:3.962      1st Qu.: 9.00      1st Qu.: 57.89
Median : 4.97799      Median : 1.84649      Median :51.080      Median :4.426      Median :10.00      Median : 73.28
Mean : 5.89733      Mean : 2.95598      Mean :54.188      Mean :4.374      Mean :10.75      Mean : 73.47
3rd Qu.: 8.20832      3rd Qu.: 3.75968      3rd Qu.:69.608      3rd Qu.:4.780      3rd Qu.:12.00      3rd Qu.: 91.95
Max.   :18.70549      Max.   :10.42136      Max.   :88.185      Max.   :5.857      Max.   :15.00      Max.   :112.31

OverAge_primary
Min.   : 0.00681
1st Qu.: 2.56406
Median : 7.08973
Mean : 5.91644
3rd Qu.: 7.30318
Max.   :16.42866
```

Fig 5. Statistici descriptive, Rstudio

- Interpretare **Rata de înscriere în ciclul primar(ER\_primary)** :

Indicatorul arată procentajul(%) copiilor înscriși, ce îndeplinesc criteriul vârstei pentru acest ciclu.

**Valoare minimă** înregistrată a fost de 85.17%, ce aparține statului El Salvador, iar **valoarea maximă** a fost de 99.99% în Singapore. Statele au fost identificate prin următoarele comenzi:

```
> which.min(education$ER_primary)
[1] 58
> which.max(education$ER_primary)
[1] 57
```

### Corelații grafice

- **Funcția `corr`**

În urma aplicării acestei funcții pe setul de date, obținem probabilitățile p-value corespunzătoare coeficientului de semnificație. Dacă p-value este mai mic decât 0.05, respingem ipoteza nulă, deci nu avem o relație de corelație semnificativă între cele două variabile analizate.

```
p
ER_primary  Repeater_primary  ER_secondary  Progression_secondary  Primary_completionRate
0.0771      0.0771      0.0000      0.0134      0.0000
Repeater_primary  0.0000      0.1061      0.0000      0.1811
Progression_secondary  0.0134      0.0000      0.0000      0.0000
Primary_completionRate  0.0000      0.1811      0.0000      0.0072
Adolescents_OutofSchool  0.0000      0.0893      0.0000      0.0000
Children_OutofSchool  0.0000      0.0266      0.0000      0.0000
ER_tertiary  0.0259      0.2738      0.0000      0.0130
Government_expenditure  0.3719      0.9637      0.0083      0.0245
Compulsory_education  0.8971      0.0003      0.7242      0.5210
ER_preprimary  0.7646      0.7839      0.0000      0.0285
OverAge_primary  0.0002      0.0000      0.0100      0.0082

Adolescents_OutofSchool  Children_OutofSchool  ER_tertiary  Government_expenditure  Compulsory_education
0.0000      0.0000      0.0259      0.3719      0.8971
Repeater_primary  0.0893      0.0266      0.2738      0.9637      0.0003
ER_secondary  0.0000      0.0000      0.0000      0.0083      0.7242
Progression_secondary  0.0000      0.0000      0.0000      0.0245      0.5210
Primary_completionRate  0.0000      0.0000      0.0130      0.2622      0.0628
Adolescents_OutofSchool  0.0000      0.0000      0.0000      0.0025      0.8053
Children_OutofSchool  0.0000      0.0002      0.0002      0.0046      0.8398
ER_tertiary  0.0000      0.0025      0.0046      0.0185      0.8023
Government_expenditure  0.0025      0.0046      0.0185      0.0609      0.0609
Compulsory_education  0.8053      0.8398      0.8023      0.0609      0.1676
ER_preprimary  0.0031      0.0172      0.0000      0.0005      0.0750
OverAge_primary  0.0423      0.0461      0.9967      0.9765      0.7887

ER_primary  Repeater_primary  ER_secondary  Progression_secondary  Primary_completionRate
0.0771      0.0771      0.0000      0.0134      0.0000
Repeater_primary  0.0000      0.1061      0.0000      0.1811
Progression_secondary  0.0134      0.0000      0.0000      0.0000
Primary_completionRate  0.0000      0.1811      0.0000      0.0072
Adolescents_OutofSchool  0.0000      0.0893      0.0000      0.0000
Children_OutofSchool  0.0000      0.0266      0.0000      0.0000
ER_tertiary  0.0259      0.2738      0.0000      0.0130
Government_expenditure  0.3719      0.9637      0.0083      0.0245
Compulsory_education  0.8971      0.0003      0.7242      0.5210
ER_preprimary  0.7646      0.7839      0.0000      0.0285
OverAge_primary  0.0002      0.0000      0.0100      0.0082

Adolescents_OutofSchool  Children_OutofSchool  ER_tertiary  Government_expenditure  Compulsory_education
0.0000      0.0000      0.0259      0.3719      0.8971
Repeater_primary  0.0893      0.0266      0.2738      0.9637      0.0003
ER_secondary  0.0000      0.0000      0.0000      0.0083      0.7242
Progression_secondary  0.0000      0.0000      0.0000      0.0245      0.5210
Primary_completionRate  0.0000      0.0000      0.0130      0.2622      0.0628
Adolescents_OutofSchool  0.0000      0.0000      0.0000      0.0025      0.8053
Children_OutofSchool  0.0000      0.0002      0.0002      0.0046      0.8398
ER_tertiary  0.0000      0.0025      0.0046      0.0185      0.8023
Government_expenditure  0.0025      0.0046      0.0185      0.0609      0.0609
Compulsory_education  0.8053      0.8398      0.8023      0.0609      0.1676
ER_preprimary  0.0031      0.0172      0.0000      0.0005      0.0750
OverAge_primary  0.0423      0.0461      0.9967      0.9765      0.7887

ER_primary  Repeater_primary  ER_secondary  Progression_secondary  Primary_completionRate
0.0771      0.0771      0.0000      0.0134      0.0000
Repeater_primary  0.0000      0.1061      0.0000      0.1811
Progression_secondary  0.0134      0.0000      0.0000      0.0000
Primary_completionRate  0.0000      0.1811      0.0000      0.0072
Adolescents_OutofSchool  0.0000      0.0893      0.0000      0.0000
Children_OutofSchool  0.0000      0.0266      0.0000      0.0000
ER_tertiary  0.0259      0.2738      0.0000      0.0130
Government_expenditure  0.3719      0.9637      0.0083      0.0245
Compulsory_education  0.8971      0.0003      0.7242      0.5210
ER_preprimary  0.7646      0.7839      0.0000      0.0285
OverAge_primary  0.0002      0.0000      0.0100      0.0082

Adolescents_OutofSchool  Children_OutofSchool  ER_tertiary  Government_expenditure  Compulsory_education
0.0000      0.0000      0.0259      0.3719      0.8971
Repeater_primary  0.0893      0.0266      0.2738      0.9637      0.0003
ER_secondary  0.0000      0.0000      0.0000      0.0083      0.7242
Progression_secondary  0.0000      0.0000      0.0000      0.0245      0.5210
Primary_completionRate  0.0000      0.0000      0.0130      0.2622      0.0628
Adolescents_OutofSchool  0.0000      0.0000      0.0000      0.0025      0.8053
Children_OutofSchool  0.0000      0.0002      0.0002      0.0046      0.8398
ER_tertiary  0.0000      0.0025      0.0046      0.0185      0.8023
Government_expenditure  0.0025      0.0046      0.0185      0.0609      0.0609
Compulsory_education  0.8053      0.8398      0.8023      0.0609      0.1676
ER_preprimary  0.0031      0.0172      0.0000      0.0005      0.0750
OverAge_primary  0.0423      0.0461      0.9967      0.9765      0.7887
```

Fig 6. Funcția *rcorr* pe setul de date, RStudio

Drept exemplu putem lua corelația dintre numărul de repetenți din ciclul primar și rata de înscriere în ciclul secundar. În acest caz, p-value este 0.1061, valoare mai mare decât 0.05, fiind o corelație semnificativă. Există și cazuri contrare, unde corelația este nesemnificativă, precum investițiile guvernului în educație și rata de înscriere în ciclul terțiar(0.018).

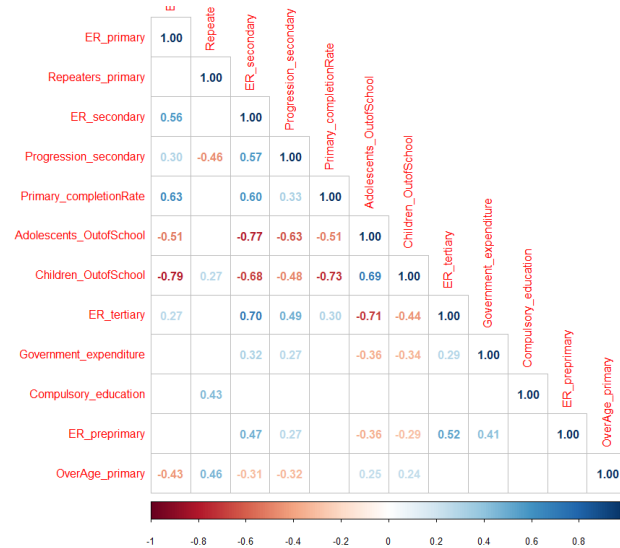


Fig 7. Corelațiile, RStudio

În diagrama din fig. 7, sunt afișați coeficienții semnificativi, iar cei contrari acestei condiții sunt marcați în alb. Cu cât culoarea este mai puternică, de exemplu relația dintre abandonul printre adolescenți și rata de înscriere în ciclul secundar (-0.77), cu atât corelația este mai stransă, altfel culoarea coeficientului este pală(0.27) între rata de înscriere în ciclul preprimar și rata de absolvire în ciclul secundar.

### 3. ACP. Extragerea componentelor principale

- Utilizăm funcția *princomp*

```
> stdev=pca$sdev
> valP=stdev^2
> procent_info=valP*100/12
> procent_cumulat=cumsum(procent_info)
> X=round(data.frame(stdev, valP, procent_info, procent_cumulat),4)
> X
```

	stdev	valP	procent_info	procent_cumulat
Comp.1	2.2282	4.9647	41.3722	41.3722
Comp.2	1.3913	1.9356	16.1299	57.5021
Comp.3	1.1852	1.4046	11.7049	69.2070
Comp.4	0.9142	0.8358	6.9651	76.1720
Comp.5	0.8397	0.7050	5.8754	82.0474
Comp.6	0.7915	0.6264	5.2203	87.2677
Comp.7	0.7337	0.5383	4.4857	91.7534
Comp.8	0.5441	0.2960	2.4669	94.2203
Comp.9	0.5103	0.2604	2.1698	96.3901
Comp.10	0.4501	0.2026	1.6886	98.0787
Comp.11	0.3822	0.1461	1.2171	99.2959
Comp.12	0.2907	0.0845	0.7041	100.0000

Fig.8 Statistici descriptive



Din fig. 8, observăm că pe *prima coloană* este *abaterea standard* a componentelor principale, pe cea *de-a doua coloană*, sunt valorile proprii ale matricei de covarianță, adică varianțele componentelor ( ex.  $\text{Var}(W1) = 4.9647$ ,  $\text{Var}(W2) = 1.9356$ ). Aceste valori sunt în ordine descrescătoare, fiind numere reale și pozitive, iar suma acestor valori este egală cu 12, numărul de indicatori, conform *proprietății 1* (***Cele “n” componente conservă integral informația prin cele n variabile originale, în sens de varianță totală.***).

*Coloana 3*, *procent\_info*, arată ponderea de informație pe care fiecare componentă o preia din variabilele ordinale. Putem observa că prima componentă preia 41.3722% din informație, pe când componenta a doua preia 16.1299%. Astfel, putem aplica *proprietatea 2* (***Componentele sunt necorelate două câte două, adică nu există suprapunere informațională.***). Prima componentă preia **maximul** din informația din toate variabilele inițiale, iar a doua preia **maximul** din informația **nepreluată** de prima. Procentajul de 41.3722% este calculat sub forma  **$(\lambda_1/12)*100$** .

```
> (4.9647/12)*100
[1] 41.3725
```

Pe *coloana 4*, *procent\_cumulat*, ponderea de informație preluată din X de fiecare componentă *decrește* și reprezintă ponderea de informație preluată din variabilele originale de primele **k** componente. Primele 5 componente preiau 82.0474% din informație.

- Utilizăm funcția *prcom*

Aportul de informație adus de fiecare componentă principală îl putem vedea în următorul grafic **Scree Plot**.

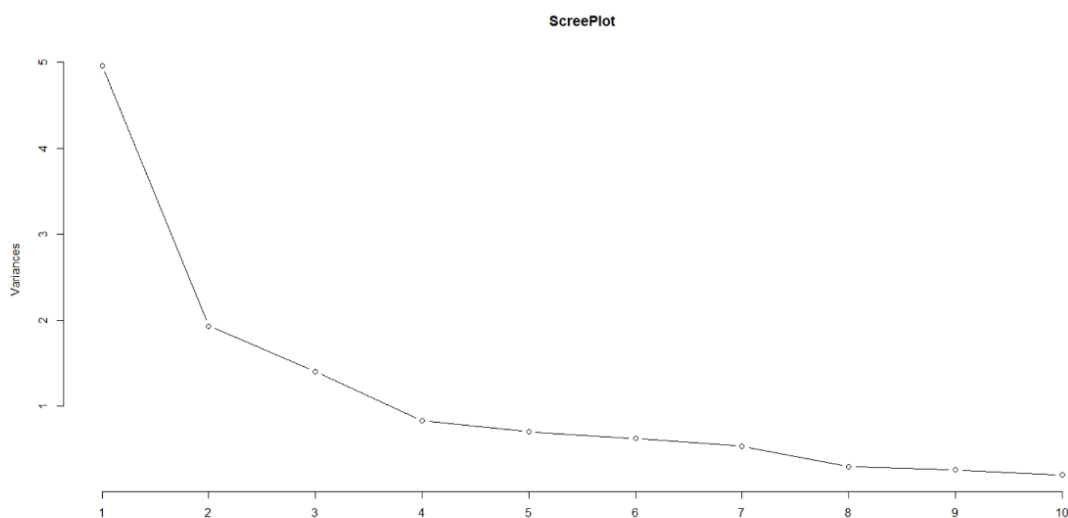


Fig. 9 Scree Plot

Graficul Screeplot ne ajută în determinarea unui număr corespunzător de componente principale. Pe axa Oy sunt reprezentate *valorile proprii*, iar pe Ox numărul *componentelor principale*.

### 3.1. Grafice – pachetul factoextra

- Punctele de pe grafic reprezintă scorurile principale și sunt corelate în funcție de calitatea reprezentării.
- Cu cât punctele au o culoare spre albastru( ex. 48, 36, 66), acele țări sunt mai slab reprezentate de primele două componente principale
- Observațiile 63, 67 au o culoare roșie-portocaliu, deci sunt mai bine reprezentate de cele două dimensiuni.
- Acest grafic ne arată cât de bine sunt reprezentate țările în planul principal

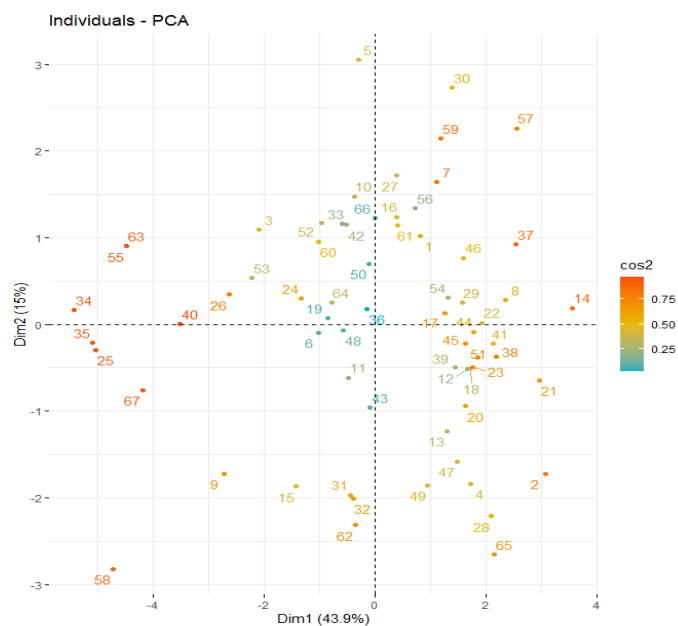


Fig. 10 Pachetul Factoextra-graficul 1

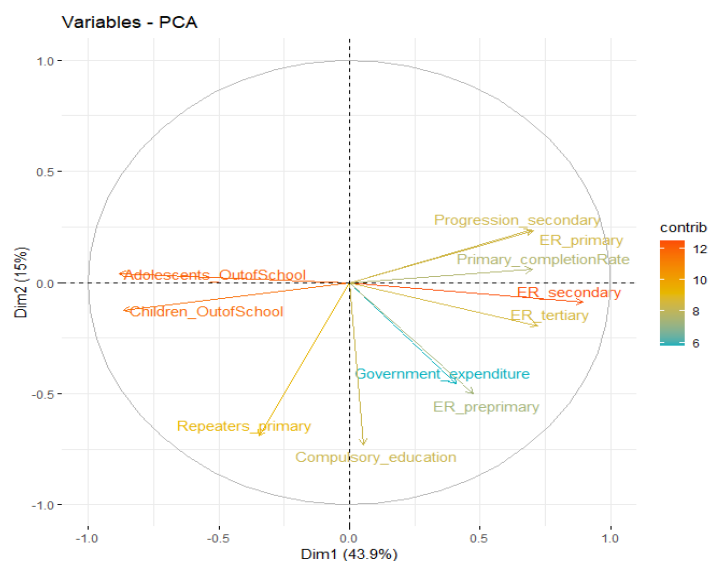


Fig. 11 Pachetul Factoextra-graficul 2

- Cel de-al doilea grafic arată contrinunția variabilelor la formarea dimensiunilor
- *Rata de abandon școlar printre copii/adolescenți* contribuie foarte mult la dimensiunea a II-a
- *Rata de absolvire în ciclul secundar* contribuie destul de puțin la dimensiunea I
- În fig.11, observăm ca distanța dintre o țară și ea însăși este nulă. Dacă punctul are o nuanță de gri mai deschis, atunci sunt forme similare din punctul de vedere al variabilelor explicative, precum Letonia – Belarus (0.8910730), Spania – Danemarca (0.7551769). Dacă punctele au tonuri mai închide de gri, disimilaritatea între țări este mai mare, de exemplu Bolivia – Kazakstan (6.141236), Argentina – Asia Sudică (7.3356527).

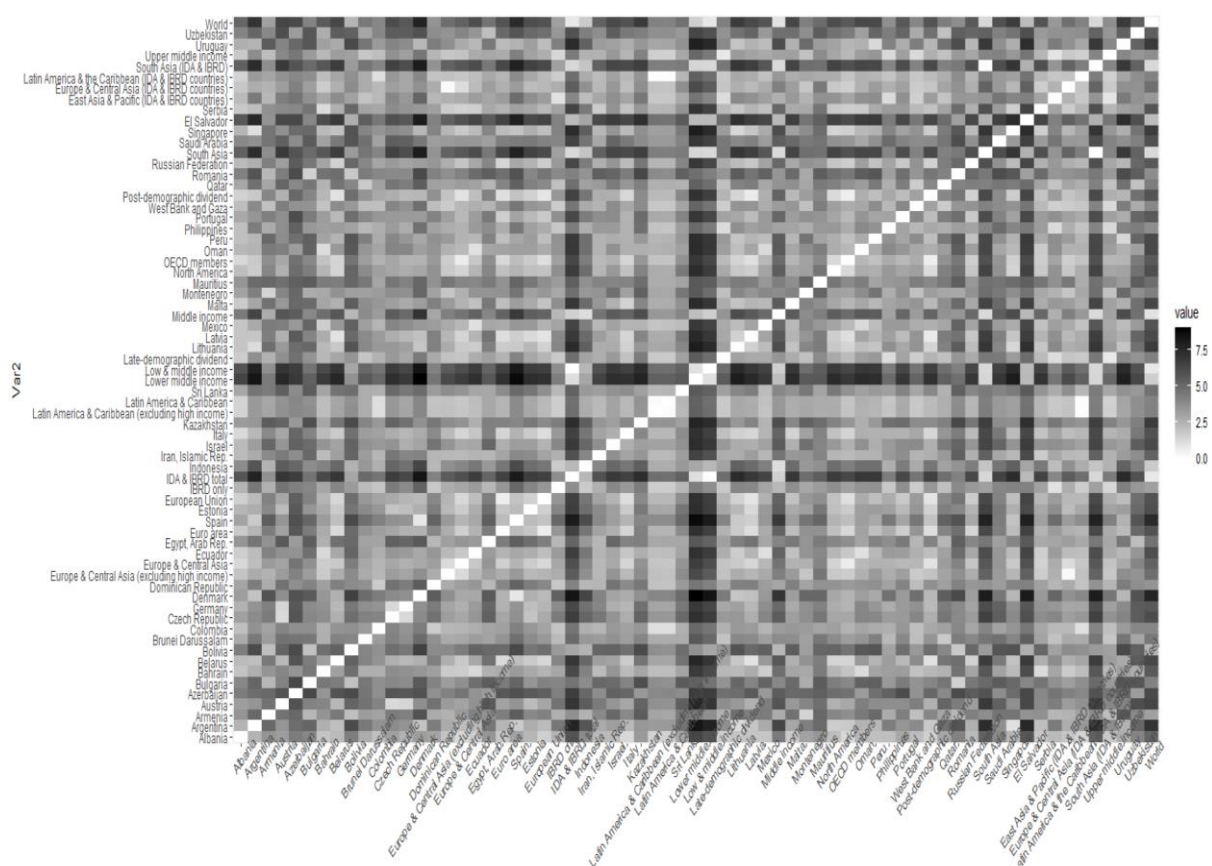


Fig. 12 Matricea de contingență

### 3.2. Criterii pentru determinarea numărului optim de clase – k

Primul criteriu are la bază **dendrograma**, se realizează o tăietură în dendrogramă, acolo unde distanța dintre doi pași consecutivi de clusterizare este cea mai mare - numărul de intersecții ale tăieturii realizate cu dendrograma este *numărul de clase*.

Distanța dintre doi pași consecutivi este, în metoda grafică, reprezentată printr-o bară scurtă, verticală.

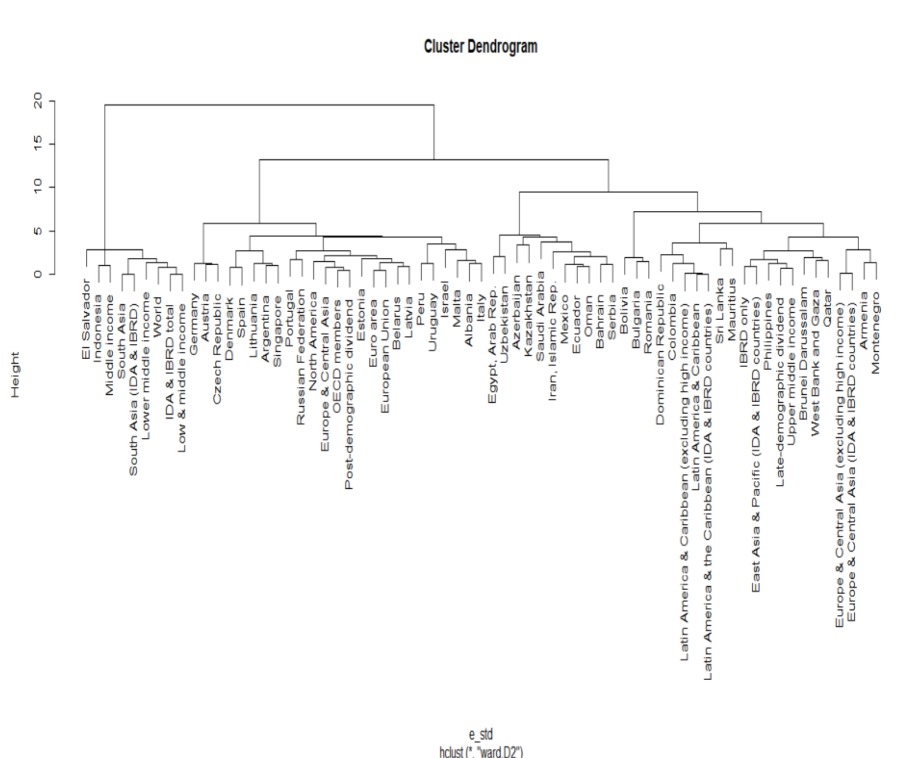


Fig.13. Dendrograma – Realizarea tăieturii

În fig.13, am evidențiat tăietura dendrogramei în punctul în care am identificat distanța maximă între 2 clase consecutive. În urma tăieturii, deduce faptul că există un număr optim de  $k = 6$ . (numărul de intersecții ale tăieturii cu dendrograma).

#### 4. Concluzii

Prin intermediul metodei Analizei Componentelor Principale, dimensiunea setului de date este redus. În cazul setului de date ales, cele 13 variabile au fost reduse la 2 componente principale:

- *Componenta 1* se corelează puternic cu Rata de înscriere în ciclul primar, Rata de înscriere în secundar, Rata de absolvire în ciclul secundar, Rata de absolvire a ciclului primar, Rata de abandon școlar printre adolescenți, Rata de abandon școlar printre copii, Rata de înscriere în terțiar. Numele acestei componente este **Capital uman școlar**
- *Componenta 2* se corelează cu numărul de repetenți din ciclul primar, Rata de înscriere în ciclul preprimar, Numărul de ani obligatoriu de studiu și elevii peste media de vârstă înscriși în ciclul primar. Numele acestei componente este **Rata de alfabetismul funcțional**.

Această componentă are un rol vital în societatea unei țări. Alfabetismul funcțional este un fenomen ce determină mentalitatea unei societăți, iar identificarea și rezolvarea lui ar putea aduce schimbări majore pentru acea națiune. În ciclul primar se pun bazele cele mai importante ale educației sociale, dar și cognitive, precum scrisul, cititul, înțelegerea unui text și abilitatea de a purta/înțelege o conversație. Astfel, odata ce **crește rata de înscriere în ciclul preprimar**, **numărul de ani de studiu obligatoriu crește**, iar **numărul elevilor peste**

*media de vârstă înscriși în ciclul primar **scade***, demersul de studiu elevului tinde să fie mai lung și mai eficient.