

TEMA 2

ANALIZA FACTORIALA

Analiza a fost realizata pe un set de date ce contine informatii despre 200 de companii listate la bursa, obtinute din sursa Yahoo Finance, sectiunea Financials (Income Statement & Cash Flow). Datele se refera la TTM (Trailing Twelve Months), oferind o privire de ansamblu asupra performantei financiare a acestor companii pe parcursul ultimelor 12 luni disponibile.

In ceea ce priveste structura setului de date, acesta nu prezinta valori lipsa, iar variabilele au fost notate de la X1 la X10, cu denumiri sugestive in fisierul Excel. In ceea ce priveste outlierii/valorile extreme, acestia au fost eliminati din setul de date, deoarece prezenta acestora afecta negativ analiza, distorsionand structura datelor si reducand relevanta rezultatelor obtinute. Dupa eliminarea valorilor extreme, in setul de date am ramas cu 81 de observatii.

Setul de date este compus din urmatoarele variabile, notate de la X1 la X10 si exprimate in mii USD (\$):

- **Total Revenue/Venit total** – ce reprezinta suma totală a veniturilor generate de o companie din vânzarea bunurilor sau serviciilor sale, fără a ține cont de costurile asociate.

- **Gross Profit/Profitul brut** – este calculat prin scăderea costurilor directe asociate producției bunurilor sau serviciilor din venitul total si reflectă eficiența în generarea profitului din activitatea principală.

- **Operating Income/Venitul operational** – este venitul generat din activitățile de bază ale companiei, excluzând veniturile și cheltuielile non-operaționale; acesta indica profitabilitatea operațiunilor zilnice ale companiei.

- **Net Income/ Venitul net** - reprezintă profitul total obținut de companie după scăderea tuturor cheltuielilor, inclusiv taxe și cheltuieli non-operaționale; este un indicator esențial al sănătății financiare a unei companii.

- **Earnings before interest and taxes (EBIT)** - acest indicator arată profitul companiei înainte de deducerea cheltuielilor cu dobânzile și impozitele, fiind util pentru compararea performanței între companii, indiferent de structura lor de capital.

- **Earnings per share (EPS)** - reprezintă venitul pe acțiune și oferă o măsură a profitabilității unei companii pe acțiune.

- **Operating Cash Flow/Fluxul de numerar operational** - acesta măsoară capacitatea companiei de a genera numerar din activitățile sale operaționale.

- **Investing Cash Flow/Fluxul de numerar din activitati de investitii** - reflectă numerarul cheltuit sau generat din activitățile de investiții ale companiei, inclusiv achiziții de active sau vânzări de active.

- **Financing Cash Flow/Fluxul de numerar din activitati de finantare** - acesta arată fluxurile de numerar rezultate din activitățile de finanțare, cum ar fi emisiunea de acțiuni, împrumuturile și rambursările de datorii.

- **Free Cash Flow/Fluxul de numerar liber** - este un indicator important al capacității unei companii de a genera numerar după ce a acoperit toate cheltuielile necesare.

În ceea ce privește **observatiile**, fiecare linie din setul de date corespunde unei companii listate la bursă, oferind o imagine de ansamblu asupra performanței financiare a acestora prin intermediul variabilelor de mai sus. Aceasta permite compararea companiilor în funcție de diferiți indicatori financiari.

Obiectivul general al analizei factoriale este de a identifica relațiile latente dintre variabilele financiare și de a reduce dimensiunea setului de date, sintetizând informația esențială. Analiza a fost aplicată pe variabile precum Venitul total, Profitul brut, Venitul operațional, Profitul net, EBIT, EPS și fluxurile de numerar (operațional, investițional, de finanțare și liber), pentru a evidenția factorii care influențează performanța financiară a companiilor. Scopul este de a înțelege contribuția acestor variabile la succesul companiilor și de a sprijini deciziile investiționale.

Interpretarea indicatorilor statistici

```
> summary(tema)
```

Companie	X1	X2	X3	X4	X5	X6
Length:81	Min. : 4131	Min. : 3966	Min. : -636864	Min. : -762367	Min. : -727188	Min. : -4.7200
Class :character	1st Qu.: 1275994	1st Qu.: 705928	1st Qu.: -14466	1st Qu.: -83497	1st Qu.: -24339	1st Qu.: -0.4200
Mode :character	Median : 2562440	Median : 1107379	Median : 281000	Median : 116261	Median : 213000	Median : 0.5900
	Mean : 4474807	Mean : 1647343	Mean : 293014	Mean : 89902	Mean : 239688	Mean : 0.9617
	3rd Qu.: 5320059	3rd Qu.: 2123393	3rd Qu.: 506000	3rd Qu.: 280000	3rd Qu.: 439514	3rd Qu.: 2.4400
	Max. : 23813905	Max. : 9646000	Max. : 1550863	Max. : 837880	Max. : 1362945	Max. : 8.8600

X7	X8	X9	X10
Min. : -586000	Min. : -1209300	Min. : -919000	Min. : -614000
1st Qu.: 131885	1st Qu.: -383000	1st Qu.: -335144	1st Qu.: 23889
Median : 384670	Median : -147000	Median : -157094	Median : 181000
Mean : 449780	Mean : -242236	Mean : -194710	Mean : 233963
3rd Qu.: 712000	3rd Qu.: -51200	3rd Qu.: -12411	3rd Qu.: 383000
Max. : 1796100	Max. : 314000	Max. : 414345	Max. : 823000

Figura 1. Rezultatul comenzii *summary*

```
> describe(tema[1:10])
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
X1	1	81	4474806.54	5230485.32	2562440.00	3345078.97	2334521.23	4131.00	23813905.00	23809774.00	2.07	3.92	581165.04
X2	2	81	1647342.80	1646814.18	1107379.00	1354668.31	975246.87	3966.00	9646000.00	9642034.00	2.25	6.56	182979.35
X3	3	81	293013.65	398911.76	281000.00	272393.32	391702.92	-636864.00	1550863.00	2187727.00	0.52	0.38	44323.53
X4	4	81	89902.35	313272.53	116261.00	100658.46	257585.44	-762367.00	837880.00	1600247.00	-0.32	0.27	34808.06
X5	5	81	239687.85	426923.00	213000.00	228556.34	351878.80	-727188.00	1362945.00	2090133.00	0.33	0.18	47435.89
X6	6	81	0.96	2.41	0.59	0.95	1.76	-4.72	8.86	13.58	0.41	1.27	0.27
X7	7	81	449780.25	447403.06	384670.00	410150.88	382605.69	-586000.00	1796100.00	2382100.00	0.84	0.78	49711.45
X8	8	81	-242235.63	282768.47	-147000.00	-199593.78	172682.87	-1209300.00	314000.00	1523300.00	-1.28	1.43	31418.72
X9	9	81	-194709.60	284965.00	-157094.00	-180020.72	236108.50	-919000.00	414345.00	1333345.00	-0.49	0.16	31662.78
X10	10	81	233962.54	298504.74	181000.00	228332.17	249804.76	-614000.00	823000.00	1437000.00	0.13	-0.13	33167.19

Figura 2. Rezultatul comenzii *describe*

Variabila X1 - Venitul total prezintă o distribuție între 4.131 USD și 23.813.905 USD, indicând variații semnificative între companiile analizate, de la firme mici la lideri ai pieței. Majoritatea companiilor au venituri situate între 1.275.994 USD (1st Qu) și 5.320.059 USD (3rd Qu), dar media de 4.474.807 USD este semnificativ influențată de câteva companii foarte mari. Mediana de 2.562.440 USD, mai mică decât media, indică o distribuție asimetrică spre dreapta, aspect confirmat de coeficientul de asimetrie (skew) de 2,07. Kurtosis-ul de 3,92 sugerează o distribuție leptocurtică, cu o concentrație mare a companiilor în jurul valorilor centrale și câteva valori extreme notabile. Devierea standard de 5.230.485 USD reflectă variații mari între veniturile companiilor analizate, sugerând o piață eterogenă dominată de câțiva actori majori.

Variabila X2 - Profitul brut prezintă o distribuție cu valori între 3.966 USD și 9.646.000 USD, indicând variații semnificative între companiile analizate, de la profituri mici la valori foarte mari. Majoritatea companiilor au profituri între 705.928 USD și 2.123.393 USD. Media de 1.647.343 USD este influențată de câteva companii extrem de profitabile, în timp ce mediana de 1.107.379 USD, mai mică decât media, reflectă o distribuție asimetrică spre dreapta, confirmată de coeficientul de asimetrie de 2,25. Kurtosis-ul de 6,56 sugerează o distribuție leptocurtică, caracterizată prin concentrarea valorilor în jurul medianei și prezența unor valori extreme notabile. Devierea standard de 1.646.814 USD evidențiază o dispersie semnificativă între companii, confirmând că piața este dominată de câțiva actori majori cu profituri semnificative, în timp ce restul companiilor obțin profituri considerabil mai mici.

Variabila X3 - Venitul operațional prezintă o distribuție semnificativ variabilă, cu valori între -636.864 USD și 1.550.863 USD, indicând o gamă largă de performanțe între companiile analizate, de la pierderi semnificative până la profituri operaționale mari. Majoritatea companiilor au venituri operaționale între -14.466 USD și 506.000 USD. Media de 293.014 USD este influențată de câteva companii cu venituri operaționale semnificativ mai mari, în timp ce mediana de 281.000 USD este mai mică decât media, ceea ce sugerează o distribuție asimetrică moderată spre dreapta. Coeficientul de asimetrie (skew) de 0.52 indică o ușoară asimetrie, iar kurtosis-ul de 0.38 sugerează o distribuție plată, cu o concentrație mai scăzută în jurul medianei și mai puține valori extreme. Aceasta sugerează o piață unde majoritatea companiilor au venituri operaționale mai mici, cu câțiva jucători mari care generează venituri semnificativ mai mari.

În ceea ce privește **variabila X4**, reprezentând **venitul net**, valoarea medie în perioada analizată este de 89.902 USD, cu un minim de -762.367 USD și un maxim de 837.880 USD, ceea ce reflectă o variabilitate moderată între companiile incluse în analiză, de la pierderi până la profituri semnificative. Conform medianei, jumătate din observații au venituri nete sub 116.261 USD, în timp ce cealaltă jumătate le depășește. Diferența dintre medie și mediană sugerează o ușoară asimetrie spre stânga, indicată și de coeficientul de asimetrie de -0.32, ceea ce înseamnă că există câteva valori mai mici care influențează distribuția. Kurtosis-ul de 0.27 arată o distribuție relativ plată, cu o concentrație scăzută în jurul medianei și câteva valori extreme, dar nu foarte dominante. Abaterea standard de 313.272,53 USD indică o dispersie semnificativă a valorilor, subliniind diferențele dintre companiile analizate. Această analiză evidențiază faptul că, deși majoritatea companiilor au venituri nete moderate, există variații relevante datorate atât pierderilor, cât și unor performanțe mai ridicate.

Analizând **variabila X5 - EBIT (Earnings Before Interest and Taxes)**, observăm o distribuție moderat variabilă, cu valori cuprinse între -727.188 USD și 1.362.945 USD, ceea ce indică diferențe semnificative în rentabilitatea operațională a companiilor analizate. Media EBIT este de 239.688 USD, iar mediana de 213.000 USD sugerează că jumătate dintre companii au valori sub această medie, iar cealaltă jumătate le depășește. Coeficientul de asimetrie de 0.33 indică o distribuție ușor asimetrică spre dreapta, ceea ce arată că există câteva companii cu valori mari care influențează media. Kurtosis-ul de 0.18 sugerează o distribuție ușor plată, cu o concentrare scăzută a datelor în jurul medianei. Abaterea standard de 426.923 USD reflectă o dispersie semnificativă între valorile EBIT ale companiilor. Analiza sugerează că, în timp ce majoritatea companiilor au rentabilități operaționale mai mici, există câteva cu performanțe mai ridicate, ceea ce evidențiază posibile strategii mai eficiente sau avantaje competitive.

Variabila X6, reprezentând venitul pe acțiune (EPS), prezintă o distribuție variabilă, cu valori cuprinse între -47.200 USD și 88.600 USD, majoritatea valorilor fiind concentrate în jurul medianei de 0.5900 USD. Coeficientul de asimetrie de 0.41 indică o distribuție ușor asimetrică spre dreapta, în care majoritatea valorilor sunt concentrate în partea stângă, dar există câteva valori mai mari care influențează media. Kurtosis-ul de 1.27 sugerează o distribuție moderat leptocurtică, ceea ce înseamnă că există o concentrație relativ mare de valori în jurul medianei, dar și câteva valori extreme ce se depărtează de medie. Din punct de vedere economic, acest lucru sugerează că majoritatea companiilor au un EPS mic, iar câteva reușesc să obțină profituri semnificative pe acțiune, influențând astfel media. Variabilitatea moderată sugerează o diversificare a performanțelor financiare pe piață, cu companii care pot avea rezultate modeste sau semnificativ mai mari.

Variabila X7, fluxul de numerar operațional, variază semnificativ, cu valori cuprinse între -586,000 USD și 1,796,100 USD, reflectând diferențe mari între companiile analizate, de la fluxuri de numerar negative până la valori semnificativ pozitive. Majoritatea companiilor au fluxuri de numerar între 131,885 USD și 712,000 USD, iar media de 449,780 USD este influențată de câteva companii cu performanțe financiare foarte bune. Mediana de 384,670 USD sugerează o ușoară asimetrie spre dreapta, iar coeficientul de asimetrie de 0.84 confirmă această tendință, indicând că există companii cu fluxuri de numerar extrem de mari care trag media în sus. Kurtosis-ul de 0.78 sugerează o distribuție platycurtică, ceea ce înseamnă că valorile sunt distribuite mai uniform, cu mai puține concentrații în jurul mediei. Aceste date sugerează o piață diversificată, cu performanțe financiare influențate de diferite strategii și condiții de piață ale companiilor.

Variabila X8, fluxul de numerar din activitățile de investiții, prezintă o distribuție variabilă, cu valori cuprinse între -1,209,300 USD și 314,000 USD, ceea ce reflectă diferențe semnificative între companiile analizate, de la cheltuieli mari până la fluxuri pozitive asociate unor câștiguri din investiții. 50% dintre observații au fluxuri de numerar între -383,000 USD și -51,200 USD, iar media de -242,236 USD este influențată de câteva companii care au înregistrat fluxuri pozitive, însă majoritatea companiilor au fluxuri negative. Mediana de -147,000 USD, mai mare decât media, confirmă o distribuție asimetrică spre stânga. Coeficientul de asimetrie de -1.28 indică o coadă lungă spre stânga, ceea ce arată că multe companii se confruntă cu cheltuieli

semnificative pentru investiții, dar câteva înregistrează fluxuri pozitive care atenuează media. Kurtosis-ul de 1.43 sugerează o distribuție leptocurtică moderată, cu o concentrație mare de valori în jurul medianei, dar și câteva valori extreme. Analiza arată că majoritatea companiilor se află într-o etapă de investiții semnificative, ceea ce ar putea reflecta strategii de creștere pe termen lung.

Variabila X9, fluxul de numerar din activitățile de finanțare, prezintă o distribuție variabilă, cu valori între -919,000 USD și 414,345 USD. Media de -194,710 USD și mediana de -157,094 USD sugerează că majoritatea fluxurilor de numerar din activitățile de finanțare sunt negative, ceea ce indică o tendință generală de finanțare negativă. Coeficientul de asimetrie de -0.49 indică o ușoară asimetrie spre stânga, cu o concentrare a valorilor în partea dreaptă și o coadă ușor extinsă în stânga. Kurtosis-ul de 0.16 sugerează o distribuție plată, cu o concentrație moderată în jurul medianei și cu câteva valori extreme. Abaterea standard de 284,965 USD subliniază variabilitatea semnificativă între fluxurile de numerar din activitățile de finanțare ale companiilor analizate. Aceste rezultate sugerează că majoritatea companiilor sunt implicate în activități financiare cu fluxuri negative, dar există și câteva companii care beneficiază de fluxuri financiare pozitive din activitățile de finanțare.

Variabila X10, fluxul de numerar liber, variază între -614,000 USD și 823,000 USD, cu o medie de 233,963 USD și o mediana de 181,000 USD. Coeficientul de asimetrie de 0.13 sugerează o distribuție ușor asimetrică spre dreapta, indicând că majoritatea companiilor au fluxuri de numerar mai mici, cu câteva excepții generând valori semnificativ mai mari. Kurtosis-ul de -0.13 arată o distribuție plată, cu o concentrație relativ uniformă a valorilor și puține valori extreme. Abaterea standard de 298,504.74 USD indică o variabilitate semnificativă între fluxurile de numerar liber ale companiilor. Din punct de vedere economic, fluxul de numerar liber sugerează că majoritatea companiilor au un flux de numerar relativ modest, dar există un număr mic de firme cu fluxuri de numerar mai mari, care pot reflecta o performanță financiară puternică și oportunități pentru investiții sau recompensarea acționarilor.

Asadar, concluzia analizei este că există o diversitate semnificativă între companiile analizate, cu o dominanță clară a unor actori mari pe piață. Variabilele financiare examinează diferențele considerabile în veniturile, profiturile și fluxurile de numerar ale companiilor, ceea ce reflectă o piață eterogenă. În ansamblu, analiza sugerează o piață financiară diversificată, unde companiile mari au un impact semnificativ asupra indicatorilor financiari, iar majoritatea companiilor mici și mijlocii au rezultate mai modeste, dar cu un potențial de creștere pe termen lung, în special pentru cele implicate în activități de investiții.

Matricea de corelatie si matricea de covarianta

Pentru a observa mai bine rezultatele si a fi mai usor de interpretat, am standardizat datele utilizand functia scale. Observam ca dupa standardizarea datelor, matricea de covarianta este egala cu matricea de corelatie, toate variabilele avand aceeași deviatie standard (1) si media 0, ceea ce le face comparabile direct între ele. Asadar, valorile observate reflectă doar relațiile dintre variabile, fără influența unității de măsură, astfel încât covarianțele și corelațiile devin mai ușor de interpretat.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	1.0000000	0.9045192	0.5718650	0.2988523	0.4717430	0.12605371	0.6700836	-0.53350678	-0.4915897	0.4187086
X2	0.9045192	1.0000000	0.5752673	0.2847471	0.4123124	0.11844546	0.6509549	-0.50014086	-0.4603218	0.4554152
X3	0.5718650	0.5752673	1.0000000	0.6695824	0.8032117	0.39782999	0.7554865	-0.66274626	-0.5823679	0.5261587
X4	0.2988523	0.2847471	0.6695824	1.0000000	0.9158698	0.74327828	0.5107917	-0.28602568	-0.5455754	0.5513663
X5	0.4717430	0.4123124	0.8032117	0.9158698	1.0000000	0.63149859	0.6425682	-0.45358593	-0.6531158	0.5789148
X6	0.1260537	0.1184455	0.3978300	0.7432783	0.6314986	1.00000000	0.2245358	-0.02663007	-0.3068137	0.3263840
X7	0.6700836	0.6509549	0.7554865	0.5107917	0.6425682	0.22453583	1.0000000	-0.71682955	-0.6125110	0.8096187
X8	-0.5335068	-0.5001409	-0.6627463	-0.2860257	-0.4535859	-0.02663007	-0.7168295	1.00000000	0.2874937	-0.3145959
X9	-0.4915897	-0.4603218	-0.5823679	-0.5455754	-0.6531158	-0.30681372	-0.6125110	0.28749371	1.0000000	-0.5262104
X10	0.4187086	0.4554152	0.5261587	0.5513663	0.5789148	0.32638399	0.8096187	-0.31459593	-0.5262104	1.0000000

Figura 3. Matricea de corelatie

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	1.0000000	0.9045192	0.5718650	0.2988523	0.4717430	0.12605371	0.6700836	-0.53350678	-0.4915897	0.4187086
X2	0.9045192	1.0000000	0.5752673	0.2847471	0.4123124	0.11844546	0.6509549	-0.50014086	-0.4603218	0.4554152
X3	0.5718650	0.5752673	1.0000000	0.6695824	0.8032117	0.39782999	0.7554865	-0.66274626	-0.5823679	0.5261587
X4	0.2988523	0.2847471	0.6695824	1.0000000	0.9158698	0.74327828	0.5107917	-0.28602568	-0.5455754	0.5513663
X5	0.4717430	0.4123124	0.8032117	0.9158698	1.0000000	0.63149859	0.6425682	-0.45358593	-0.6531158	0.5789148
X6	0.1260537	0.1184455	0.3978300	0.7432783	0.6314986	1.00000000	0.2245358	-0.02663007	-0.3068137	0.3263840
X7	0.6700836	0.6509549	0.7554865	0.5107917	0.6425682	0.22453583	1.0000000	-0.71682955	-0.6125110	0.8096187
X8	-0.5335068	-0.5001409	-0.6627463	-0.2860257	-0.4535859	-0.02663007	-0.7168295	1.00000000	0.2874937	-0.3145959
X9	-0.4915897	-0.4603218	-0.5823679	-0.5455754	-0.6531158	-0.30681372	-0.6125110	0.28749371	1.0000000	-0.5262104
X10	0.4187086	0.4554152	0.5261587	0.5513663	0.5789148	0.32638399	0.8096187	-0.31459593	-0.5262104	1.0000000

Figura 4. Matricea de covarianta

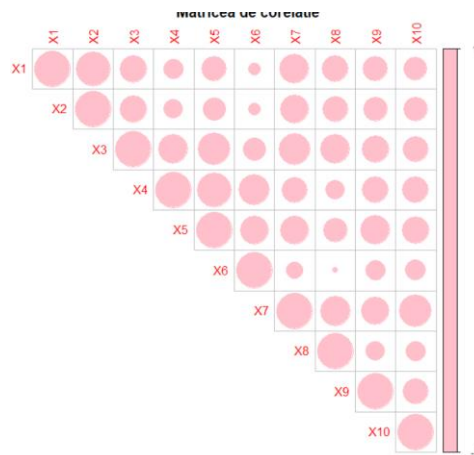


Figura 5. Reprezentarea grafica a matricei de corelatie

Analizând matricea de corelație, putem observa un set complex de relații între variabilele financiare, care oferă o perspectivă asupra interdependențelor dintre indicatori.

Variabila X1 (Venitul total generat) este puternic corelată cu X2 (Profitul brut) și X7 (Fluxul de numerar operațional), indicând o legătură strânsă între creșterea veniturilor și performanța financiară generală. Corelațiile moderat pozitive cu X3 (Venitul operațional) și X5 (EBIT) sugerează că veniturile susțin profitabilitatea operațională și profitul înainte de dobânzi și taxe.

Variabila X2 (Profitul brut) are corelații puternice cu X1 (Venitul total generat), X7 (Fluxul de numerar operațional) și X3 (Venitul operațional), indicând că o creștere a veniturilor și a fluxului de numerar operațional contribuie semnificativ la profitabilitatea brută și operațională a companiei. Corelațiile mai slabe cu X6 (Venitul pe acțiune) și X4 (Venitul net) sugerează că profitul brut nu influențează direct performanța pe acțiune sau profitabilitatea netă, aceste variabile fiind mai sensibile la alți factori economici sau strategii financiare.

Variabila X3 (Venitul operațional) prezintă corelații puternice cu X5 (EBIT) și X7 (Fluxul de numerar operațional), indicând o relație strânsă între performanța operațională și fluxul de numerar generat de activitățile de bază ale companiei. Corelațiile moderat pozitive cu celelalte variabile sugerează că venitul operațional este influențat într-o măsură semnificativă de veniturile totale, profitul brut și fluxul de numerar liber, dar și de variabilele financiare de bază.

Variabila X4 (Venitul net) este puternic corelată cu X5 (EBIT) și X6 (Venitul pe acțiune), indicând o influență semnificativă a profitabilității operaționale și performanței pe acțiune. Corelațiile moderate cu celelalte variabile sugerează o legătură mai slabă, dar încă semnificativă, între venitul net și veniturile totale, profitul brut și fluxurile de numerar. Corelațiile mai slabe cu X1 (Venitul total) și X2 (Profitul brut) sugerează că venitul net este mai puțin influențat direct de acestea din cauza cheltuielilor și taxelor.

Variabila X5 (EBIT) prezintă corelații puternice cu majoritatea variabilelor financiare, ceea ce subliniază faptul că profitul înainte de dobânzi și taxe are un impact semnificativ asupra altor indicatori cheie. Corelația cea mai puternică este cu **X4 (Venitul net)**, ceea ce sugerează că EBIT are un rol determinant în generarea veniturilor nete, indicând o relație strânsă între profitul operațional și profitul final al companiei. Aceasta evidențiază faptul că un EBIT mai mare conduce, în general, la un venit net mai mare, având în vedere impactul relativ mai mic al altor cheltuieli și taxe.

Variabila X6 (EPS - Venitul pe acțiune) are corelații slab și moderat pozitive cu majoritatea variabilelor financiare, ceea ce sugerează că venitul pe acțiune este influențat într-o măsură relativ mică de celelalte indicatori financiari. Cele mai puternice corelații sunt cu X4 (Venitul net) și X5 (EBIT), indicând că venitul pe acțiune este strâns legat de profitabilitatea generală a companiei.

Variabila X7 (Fluxul de numerar operațional) este puternic corelată pozitiv cu majoritatea variabilelor financiare, indicând o influență directă din veniturile totale, profitul brut, venitul operațional și EBIT. Corelația slab pozitivă cu X6 (Venitul pe acțiune) sugerează un

impact indirect asupra veniturilor pe acțiune. Corelațiile negative cu X8 (Fluxul de numerar din activități de investiții) și X9 (Fluxul de numerar din activități de finanțare) sugerează că investițiile și finanțarea prin datorii afectează negativ fluxul de numerar operațional, indicând o utilizare mai puțin eficientă a numerarului.

Variabilele X8 (Investing Cash Flow - Fluxul de numerar din activități de investiții) și X9 (Financing Cash Flow - Fluxul de numerar din activități de finanțare) au corelații negative cu toate celelalte variabile financiare, ceea ce sugerează că, pe măsură ce fluxurile de numerar din investiții și finanțare cresc, performanțele financiare ale companiei (precum veniturile, profiturile și fluxurile de numerar operaționale) tind să scadă. Totuși, X8 și X9 prezintă o corelație pozitivă moderată între ele, ceea ce sugerează că, de obicei, atunci când o companie alocă mai multe fonduri pentru activități de investiții, ea recurge și la surse de finanțare suplimentare (precum împrumuturi sau emisiuni de acțiuni), reflectând un comportament strategic comun de finanțare a expansiunii prin investiții.

Variabila X10 (Fluxul de numerar liber) are cea mai puternică corelație pozitivă cu X7 (Fluxul de numerar operațional), indicând că un flux operațional mai mare generează mai multe fonduri pentru cheltuielile de capital și pentru investiții sau plata datoriilor. De asemenea, X10 prezintă corelații moderat pozitive cu celelalte variabile financiare, sugerând o legătură între fluxul de numerar liber și performanțele financiare ale companiei, influențate de mai mulți factori economici.

Matricea de covarianță, având aceleași valori ca matricea de corelație după standardizare, oferă informații esențiale despre relațiile dintre variabilele economice analizate. Majoritatea covarianțelor sunt pozitive, ceea ce sugerează o relație directă între variabilele financiare. Astfel, pe măsură ce o variabilă crește, celelalte au tendința să crească și ele. Cu toate acestea, covarianțele negative între X8 (fluxul de numerar din activități de investiții) și X9 (fluxul de numerar din activități de finanțare) cu celelalte variabile financiare sugerează că o creștere a acestor fluxuri are un efect negativ asupra altor indicatori, cum ar fi venitul total și profitul brut. Aceasta poate semnala faptul că investițiile și finanțările externe sunt invers proporționale cu profitabilitatea pe termen scurt, resursele fiind alocate pe termen lung, ceea ce afectează negativ fluxurile de numerar operaționale și alte activități economice.

Reprezentari grafice

- Histogramele si densitatile de probabilitate:**

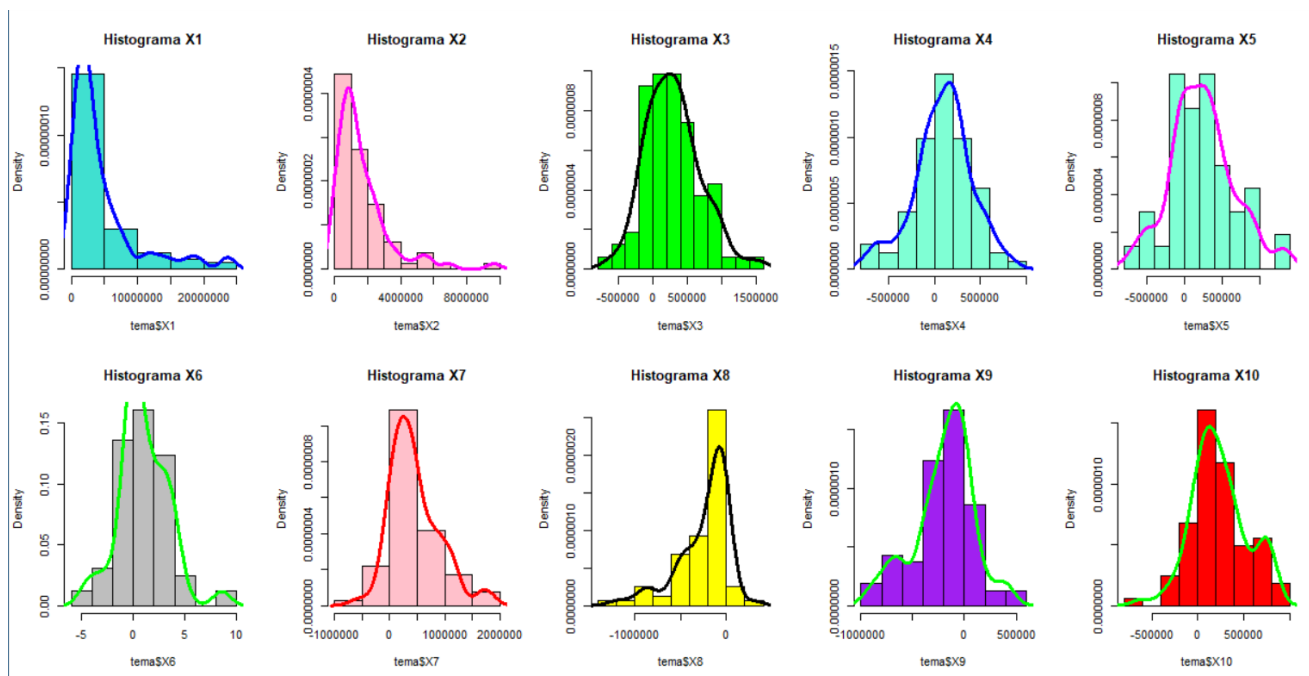


Figura 6. Histogramele si densitatile de probabilitate

Distribuțiile primelor două variabile (Venitul total - X1 și Profitul brut - X2) sunt puternic asimetrice la dreapta, majoritatea valorilor fiind concentrate în partea stângă, la valori mai mici. Totuși, există și câteva valori mai mari, care reprezintă companiile foarte profitabile ce domină piața bursieră, influențând semnificativ formele distribuțiilor.

Variabila X3 (Venitul operațional) prezintă o distribuție ușor asimetrică pozitiv, ceea ce sugerează că majoritatea valorilor se află pe partea stângă a distribuției, iar câteva valori mai mari (outliers) trag coada distribuției spre dreapta. Astfel, majoritatea observațiilor sunt concentrate în jurul valorilor mici, iar valorile mai mari sunt mai rare. De asemenea, distribuția este bimodală, ceea ce indică existența a două grupuri/tendințe dominante.

X4 (Venitul net) prezintă o distribuție ușor asimetrică, cu date distribuite aproape simetric, dar cu o ușoară tendință spre stânga, indicând prezența unor valori extreme mai mici. Majoritatea valorilor se află pe partea dreaptă a distribuției, iar câteva valori mai mici trag coada distribuției spre stânga, ceea ce sugerează o tendință generală de rentabilitate pozitivă, dar cu o mică porțiune de companii care înregistrează pierderi sau performanțe financiare slabe.

X5 (EBIT) are o distribuție ușor asimetrică la dreapta, majoritatea valorilor fiind concentrate pe valori mici, iar câteva valori mai mari trag coada distribuției spre dreapta. De

asemenea, observăm o distribuție bimodală, indicând două grupuri de companii: unele cu performanțe financiare medii și altele cu performanțe extrem de bune.

La fel ca X5, variabilele X6 (Venitul pe acțiune), X7 (Fluxul de numerar operațional) și X10 (Fluxul de numerar liber) prezintă distribuții ușor asimetrice la dreapta, ceea ce înseamnă că majoritatea valorilor sunt concentrate pe partea stângă, cu câteva valori extreme pe partea dreaptă. Aceste distribuții sugerează prezența unor companii cu valori mai mici de venituri sau fluxuri de numerar, dar și câteva companii cu performanțe financiare foarte bune, ce influențează semnificativ aceste variabile.

În schimb, X8 (Fluxul de numerar din activități de investiții) și X9 (Fluxul de numerar din activități de finanțare) au distribuții asimetrice la stânga, X8 fiind mai puternic asimetric decât X9. Aceste distribuții negative sugerează că majoritatea companiilor au fluxuri de numerar din investiții și finanțare mai mici, iar câteva companii care înregistrează valori extreme pozitive ce domină forma distribuției.

- **Dependenta dintre 2 variabile - Dreapta de regresie. Scatter Plot.**

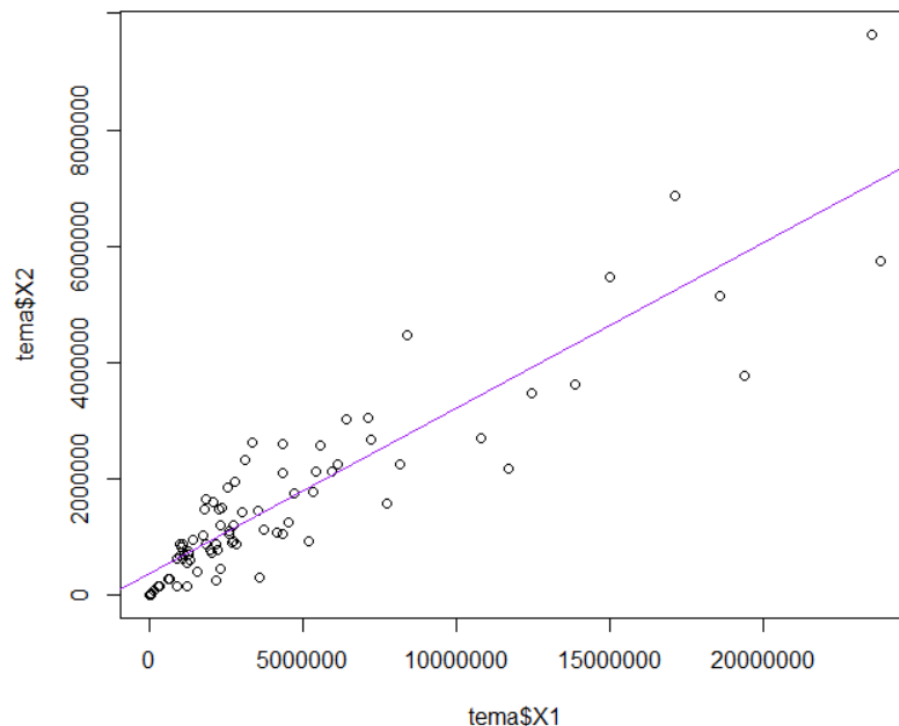


Figura 7. Dreapta de regresie dintre X1 si X2

X1 reprezintă variabila independentă, iar X2 este variabila dependentă. Analizând graficul scatter plot de mai sus, observăm că la început datele urmează o relație liniară, ceea ce sugerează că pentru anumite valori ale lui X1 și X2, rezultatele pot fi aproximativ precise folosind o dreaptă de regresie. Totuși, pe măsură ce valorile cresc, relația devine mai complexă,

iar previziunea devine mai dificilă. X_1 explică doar parțial variabilitatea lui X_2 , existând și alți factori care devin mai relevanți pe măsură ce datele se dezvoltă. De asemenea, putem observa prezența valorilor extreme sau a outlierilor, reprezentate de punctele care se îndepărtează semnificativ de dreapta de regresie, ce reprezintă companiile mari cu valori neobișnuite.

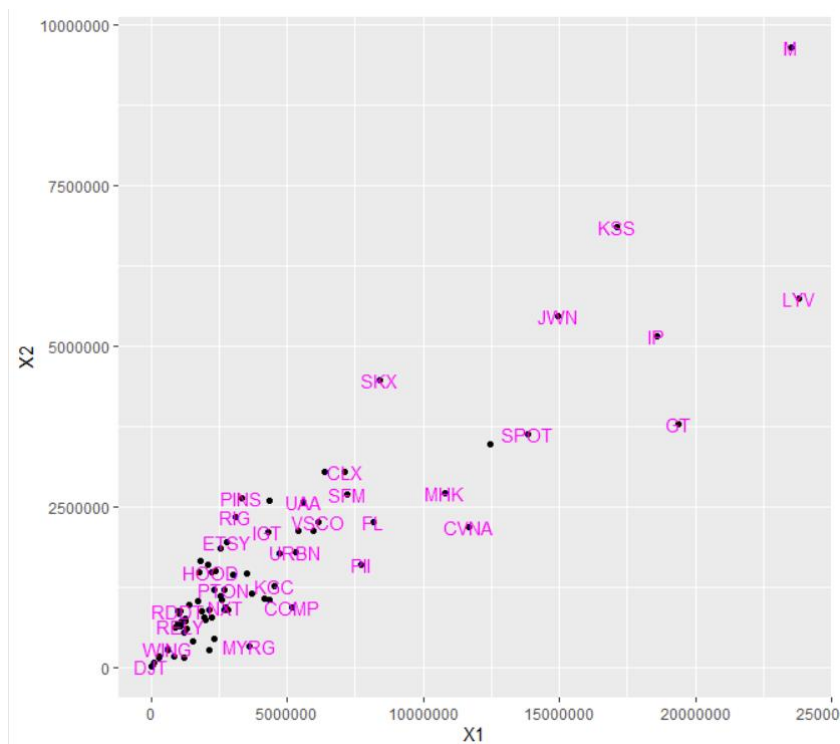


Figura 8. Scatter Plot

Analizând figura 10, putem determina companiile cele mai profitabile, ce domina piața bursieră, având veniturile și profiturile semnificativ mai mari decât ale celorlalte companii. Câteva dintre acestea sunt M, LYV, KSS, JWN, IP, OT, etc.

• Graficele Boxplot

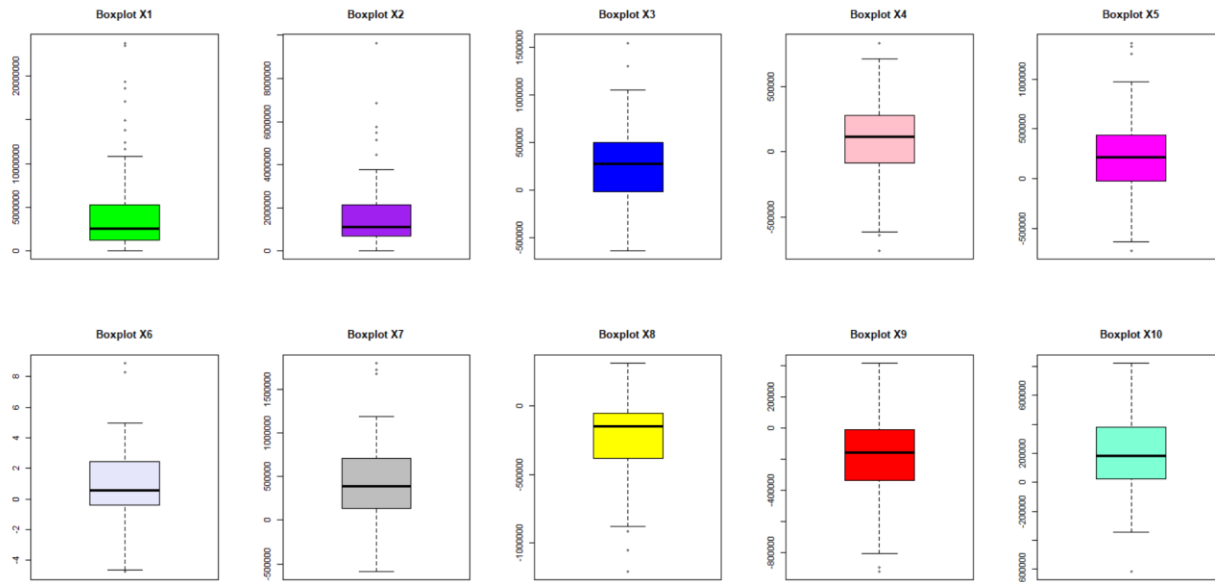


Figura 9. Graficele Boxplot

Pe baza figurii, observam ca X1 si X2 au distributiile puternic asimetrice, majoritatea valorilor fiind concentrate in partea inferioara a graficului (valori mici), si prezentand destul de multe valori extreme in partea superioara/in zona valorilor mari, reprezentate de companii ce performeaza mai bine din punct de vedere financiar. X4 si X9 au distributii asemanatoare, usor asimetrice la stanga, prezentand outlieri in zona valorilor mai mici, asadar majoritatea companiilor prezinta valori ridicate pentru venitul net si fluxul de numerar din activitati de finantare, existand doar cateva companii cu valori scazute. Pe cand X8 are o distributie puternic asimetrica la stanga, prezentand un numar mai mare de valori extreme. Celelalte variabile au distributii usor asimetrice la dreapta, prezentand un numar mic de outlieri in zona valorilor ridicate, reprezentati de companiile ce domina piata bursiera.

ANALIZA FACTORIALA

Ca un prim pas in analiza, am analizat matricea de corelatie si am ales sa elimin variabilele X8 si X9, deoarece au corelatii negative cu toate celelalte variabile, acestea fiind mai putin relevante pentru explicatia variabilitatii totale a datelor. Eliminarea acestora ajuta la imbunatatirea calitatii analizei, concentrandu-se pe variabilele care au o corelatie mai puternica cu celelalte, si, de asemenea, reduce dimensiunea datelor, facilitand interpretarea si identificarea principalilor factori care influenteaza variabilitatea setului de date.

Apoi, am standardizat noul set de date.

Evaluarea factorabilitatii setului de date

Indicele KMO

```
> KMO(date_AF_std)
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = date_AF_std)
Overall MSA = 0.71
MSA for each item =
  X1  X2  X3  X4  X5  X6  X7  X10
0.63 0.66 0.76 0.73 0.73 0.84 0.73 0.68
```

Figura 10. Rezultatul KMO

Indicele KMO (Kaiser-Meyer-Olkin) are un prag minim de 0.5-0.6 pentru a putea continua cu analiza factorială. Ideal, valoarea ar trebui să fie cât mai aproape de 1. În cazul nostru, valoarea este de 0.71, astfel că putem continua analiza factorială.

Testul de sfericitate Bartlett

[illegible]

Figura 11. Testul Bartlett

Ipotezele testului Bartlett:

- **H0:** Variabilele sunt ortogonale (matricea de corelație este matricea identitate).
- **H1:** Există cel puțin un factor comun care explică corelația dintre variabile.

Se analizează p-value: dacă acesta este mai mic decât 0.05, respingem ipoteza nulă și acceptăm ipoteza alternativă.

Rezultatul ($p\text{-value} < 0.05$) arată că respingem H_0 și acceptăm H_1 , ceea ce indică faptul că există cel puțin un factor comun care explică corelația dintre variabile.

Alegerea numărului de factori:

Pentru alegerea numărului de factori, se vor folosi aceleasi metode ca in cazul analizei componentelor principale (ACP).

In primul rand, am calculat componentele principale, aplicand ACP pe datele standardizate. Apoi am calculat valorile proprii, procentul de informatie explicata, procentul cumulat, si am realizat un data frame cu toate aceste valori pentru a fi mai usor de verificat.

```
> acp = princomp(date_AF_std, cor=T)
> acp
Call:
princomp(x = date_AF_std, cor = T)

Standard deviations:
  Comp.1   Comp.2   Comp.3   Comp.4   Comp.5   Comp.6   Comp.7   Comp.8
2.2001733 1.2577388 0.8327886 0.6660704 0.4583346 0.3495752 0.2606887 0.1997794

8 variables and 81 observations.
```

Figura 12. Componentele principale

	sdev	valp	procent_info	procent_cumulat
Comp.1	2.2001733	4.84076271	60.5095338	60.50953
Comp.2	1.2577388	1.58190676	19.7738346	80.28337
Comp.3	0.8327886	0.69353686	8.6692107	88.95258
Comp.4	0.6660704	0.44364980	5.5456225	94.49820
Comp.5	0.4583346	0.21007062	2.6258827	97.12408
Comp.6	0.3495752	0.12220285	1.5275356	98.65162
Comp.7	0.2606887	0.06795859	0.8494824	99.50110
Comp.8	0.1997794	0.03991182	0.4988977	100.00000

Figura 13. Tabelul cu valori proprii, procentele de informatie explicata si procentele cumulate

Criterii de alegere a numărului de factori

Criteriul Screeplot : Se analizează graficul Scree pentru a identifica punctul de inflexiune/cotitura, unde scăderea devine mai lină si panta se apropie de 0.

Am realizat graficul Scree ce prezinta valorile proprii (varianta explicata de fiecare componenta) pe axa verticala si componentele principale pe axa orizontala.

Acest grafic ajută la determinarea numărului optim de componente. Cautăm punctul de cotitură (locul unde panta graficului începe să se aplatizeze), în cazul nostru acesta fiind în dreptul componentei 5. Punctul de cotitură indică faptul că primele componente au o contribuție semnificativă, iar celelalte adaugă doar variație mai mică. Ducem o dreaptă paralela cu axa verticală din punctul de cotitură (5), iar primul întreg din stanga dreptei va reprezenta numărul de componente pe care le păstrăm, adică 4. Asadar, decidem să păstrăm 4 factori în analiză.

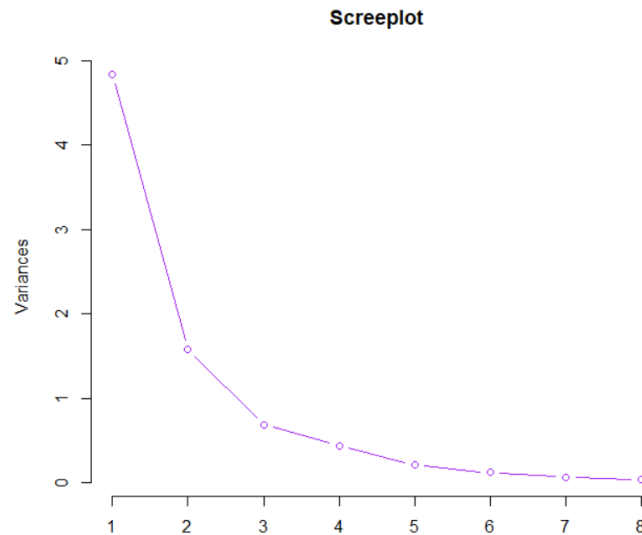


Figura 14. Graficul Scree

Criteriul lui Kaiser: Se rețin doar componentele principale a căror valoare proprie este mai mare sau egală cu 1.

Din figura 13, observăm că valorile proprii sunt mai mari sau egale cu 1 decât pentru primele 2 componente principale, asadar păstrăm 2 factori în analiză.

Criteriul procentului de acoperire/cumulat: Se rețin componentele principale care, împreună, explică între 75-80% din variația totală a datelor.

Conform figurii 13, observăm că procentul de 80% este atins de a doua componentă, asadar vom păstra primele 2 componente în analiză, deoarece împreună ele explică 80.28% din variația totală a datelor.

În concluzie, conform celor trei criterii, numărul optim de factori este 2. Aceasta sugerează că doi factori principali sunt suficienți pentru a explica majoritatea variabilității datelor, ceea ce va permite o reducere semnificativă a dimensiunii setului de date, menținând în același timp informațiile esențiale.

Estimarea modelului factorial

1. Metoda Axelor Principale (PA)

Am realizat analiza factoriala folosind metoda axelor principale (Principal Axis - PA), pe 2 factori.

```
> print(factor1$loadings, cutoff = 0.4)
```

```
Loadings:
      PA1  PA2
X1  0.707 -0.547
X2  0.696 -0.569
X3  0.838
X4  0.817  0.549
X5  0.896
X6  0.512  0.504
X7  0.841
X10 0.696
```

```

      PA1  PA2
SS loadings  4.613 1.359
Proportion Var 0.577 0.170
Cumulative Var 0.577 0.747
```

Figura 15. Incarcaturile factorilor



Figura 16. Matricea incarcaturilor

În primul rând, am afișat încărcăturile/loading-urile factorilor, filtrând valorile mai mici de 0.4 pentru a evidenția relațiile semnificative dintre variabile și factori.

PA1 și PA2 reprezintă componentele principale rezultate în urma analizei factorilor. Fiecare variabilă este asociată cu coeficienți (sau încărcări) care indică cât de mult contribuie acea variabilă la fiecare factor. Astfel, din figura 15, observăm că variabilele X1, X2, X3, X4, X5, X7, și X10 au încărcări mari pe acest factor (majoritatea peste 0.5), ceea ce înseamnă că acestea sunt bine corelate cu PA1. Aceste variabile sunt influențate în mod semnificativ de primul factor. Iar, în ceea ce privește PA2, este explicată doar de câteva variabile, într-o măsură mult mai mică comparativ cu PA1.

SS Loadings reprezintă varianța totală explicată de fiecare factor. PA1 explică 4.613 unități de varianță, iar PA2 explică 1.359 unități de varianță. Iar Proportion Var arată proporția de varianță explicată de fiecare factor. PA1 explică 57.7% din varianța totală, iar PA2 explică 17%. Împreună, cele două factori explică 74.7% din varianța totală a datelor.

Asadar, primul factor (PA1) are o influență semnificativă asupra majorității variabilelor, explicând cea mai mare parte din varianța datelor, în timp ce al doilea factor (PA2) explică o proporție mai mică a varianței și este mai puțin influent asupra variabilelor analizate.

Acest aspect poate fi observat și în figura 16, în matricea încărcăturilor, unde se remarcă corelațiile mai puternice ale variabilelor cu PA1, comparativ cu corelațiile mai slabe, și câteva chiar negative, cu PA2.

Pentru a afișa toate rezultatele analizei factoriale, am rulat *factor 1*.

```
Standardized loadings (pattern matrix) based upon correlation matrix
  PA1  PA2  h2  u2 com
X1  0.71 -0.55 0.80 0.201 1.9
X2  0.70 -0.57 0.81 0.192 1.9
X3  0.84  0.00 0.70 0.297 1.0
X4  0.82  0.55 0.97 0.031 1.7
X5  0.90  0.34 0.92 0.079 1.3
X6  0.51  0.50 0.52 0.484 2.0
X7  0.84 -0.25 0.77 0.229 1.2
X10 0.70 -0.03 0.49 0.515 1.0

SS loadings          PA1  PA2
Proportion Var       4.61 1.36
Cumulative Var       0.58 0.17
Proportion Explained 0.58 0.75
Cumulative Proportion 0.77 1.00

Mean item complexity = 1.5
Test of the hypothesis that 2 factors are sufficient.

df null model = 28 with the objective function = 8.72 with Chi Square = 666.72
df of the model are 13 and the objective function was 1.72

The root mean square of the residuals (RMSR) is 0.06
The df corrected root mean square of the residuals is 0.09

The harmonic n.obs is 81 with the empirical chi square 17.14 with prob < 0.19
The total n.obs was 81 with Likelihood Chi Square = 129.22 with prob < 0.00000000000000000003

Tucker Lewis Index of factoring reliability = 0.601
RMSEA index = 0.332 and the 90 % confidence intervals are 0.283 0.388
BIC = 72.09
Fit based upon off diagonal values = 0.99
Measures of factor score adequacy

  PA1  PA2
Correlation of (regression) scores with factors 0.99 0.96
Multiple R square of scores with factors        0.98 0.92
Minimum correlation of possible factor scores   0.96 0.85
```

Figura 17. Rezultatele analizei PA

- **Încărcăturile standardizate (Standardized Loadings):** Variabilele X5, X3, X7, X4, X1, X2 și X10 (în această ordine, descrescător după valoarea încărcăturii) prezintă încărcături mari pe prima componentă principală (PA1), ceea ce indică faptul că acestea sunt bine reprezentate de acest factor. Pe rând, PA2 contribuie mai puțin, având încărcături mai mici pentru cele mai multe variabile, cea mai mare valoare fiind de 0.55 pentru X4.
- **Comunalitatea (h²)** reprezintă informația comună tuturor variabilelor. Aceasta reflectă cât din variabilitatea fiecărei variabile este explicată de cei doi factori. O valoare mare a comunalității înseamnă că varianta variabilei este bine explicată de factori. De exemplu, X4 are o comunalitate foarte ridicată (0.97), ceea ce înseamnă că 97% din variabilitatea sa este explicată de factori. Iar X6 și X10 au comunalități mai mici (0.52 și 0.49), deci sunt mai slab reprezentate de cei doi factori.
- **Unicitatea (u²)** reprezintă informația de natură specifică, partea neexplicată de factori. De exemplu, pentru X4 (u² = 0.031), doar 3.1% din variabilitatea sa nu este captată de

factori. Asadar, este una dintre cele mai bine reprezentate variabile. Pe cand, cea mai slab reprezentata este X10, avand unicitatea de 0.515, ceea ce inseamna ca 51.5% din variabilitatea sa nu este captata de cei doi factori.

- **Complexitatea (com) sau Indicele de complexitate Hoffman** masoara numărul de factori latent necesari pentru a explica o variabilă. O valoare mai mică indică o asociere mai clară cu un singur factor. De exemplu, X3 și X10 sunt foarte simple (com = 1.0), în timp ce X6 este mai complexă (com = 2.0).
- **SS Loadings:** PA1 explică 4.61 din variabilitatea totală, ceea ce corespunde la 57.7% din informație. PA2 explică 1.36, adică 17.0% din variabilitatea totală. Împreună, cei doi factori explică 75% din variația totală (Cumulative Var), ceea ce este o acoperire bună pentru analiza factorială.
- **Proportia explicata (Proportion Explained):** PA1 contribuie mai mult la explicație (77% din totalul informației explicate), iar PA2 mai puțin (23%).
- **RMSR (Root Mean Square Residual):** Valoarea mică a RMSR (0.06) indică o bună potrivire între matricea de corelație observată și cea reprodusă de model.
- **Tucker-Lewis Index (TLI):** Valoarea de 0.601 sugerează că modelul poate fi îmbunătățit.
- **RMSEA (Root Mean Square Error of Approximation):** Valoarea ridicată (0.332) indică o potrivire mai puțin ideală.
- **Fit based upon off diagonal values** are o valoare ridicata, de 0.99, ceea ce indica faptul ca modelul se potriveste bine datelor in ceea ce priveste explicarea corelatiilor in afara diagonalei principale din matricea de corelatie.
- **Corelatia scorurilor cu factorii** prezinta valori foarte mari (0.99 pentru PA1 si 0.96 pentru PA2), ceea ce sugerează o estimare precisă și o reprezentare cu acuratete.
- **Coeficientul de determinare R^2 :** PA1 (0.98) și PA2 (0.92) arată că modelul explică bine variabilitatea datelor.

Asadar, modelul identifica 2 factori principali care explica 75% din variabilitatea datelor, PA1 fiind factorul dominant, captand majoritatea corelațiilor dintre variabile. Totuși, unele valori ale indicatorilor de ajustare (RMSEA și TLI) sugerează că modelul poate fi perfecționat.

2. Metoda Verosimilitatii Maxime (ML – Maximu Likelihood)

Am realizat analiza factoriala folosind metoda verosimilitatii maxime (ML), pe 2 factori.

```
> print(factor2$loadings, cutoff = 0.4)
```

Loadings:

	ML1	ML2
X1		0.940
X2		0.938
X3	0.645	0.544
X4	0.938	
X5	0.906	
X6	0.719	
X7	0.444	0.669
X10	0.492	0.426

	ML1	ML2
SS loadings	3.102	2.849
Proportion Var	0.388	0.356
Cumulative Var	0.388	0.744

Figura 18. Incarcaturile factorilor

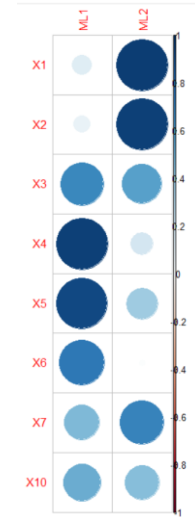


Figura 19. Matricea incarcaturilor

ML1 și ML2 reprezintă componentele principale rezultate în urma analizei factorilor cu metoda ML. Fiecare variabilă este asociată cu coeficienți (sau încărcări) care indică cât de mult contribuie acea variabilă la fiecare factor. Incarcaturile arată corelația dintre fiecare variabilă și factorii principali (ML1 și ML2). Astfel, din figura 18, observăm că variabilele X4 și X5 sunt asociate puternic cu ML1 (0.938 și 0.906), iar X1 și X2 sunt asociate puternic cu ML2 (0.940 și 0.938). Variabila X3 are încărcături semnificative pentru ambii factori, indicând că ea contribuie la mai multe dimensiuni ale analizei. Iar variabila X10 are încărcături mai mici pe ambii factori, sugerând o contribuție mai redusă. Asadar, majoritatea variabilelor sunt asociate puternic cu unul dintre cei doi factori (ML1 sau ML2), indicând o bună separare a dimensiunilor analizate.

SS Loadings reprezintă varianța totală explicată de fiecare factor. ML1 explică 3.102 unități de variație, iar ML2 explică 2.849 unități de variație. Iar Proportion Var arată proporția de varianță explicată de fiecare factor. ML1 explică 38.8% din varianța totală, iar ML2 explică 35.6%. Împreună, cele două factori explică 74.4% din varianța totală a datelor, ceea ce sugerează că modelul este adecvat pentru a reprezenta relațiile dintre variabile.

Observăm că modelul este bine echilibrat, deoarece contribuția celor doi factori este similară (38.8% și 35.6%), ceea ce sugerează că niciun factor nu domină excesiv analiza. Acest echilibru oferă o reprezentare coerentă a structurii datelor, cu ambele dimensiuni având o importanță comparabilă. Acest aspect poate fi observat și în figura 19, în matricea încărcăturilor, unde se remarcă distribuția echilibrată a corelațiilor cu cei doi factori principali.

Pentru a afișa toate rezultatele analizei factoriale, am rulat *factor 2*.

```
Standardized loadings (pattern matrix) based upon correlation matrix
      ML1  ML2  h2    u2 com
X1  0.14  0.94  0.90  0.097 1.0
X2  0.10  0.94  0.89  0.111 1.0
X3  0.65  0.54  0.71  0.288 1.9
X4  0.94  0.18  0.91  0.086 1.1
X5  0.91  0.36  0.95  0.051 1.3
X6  0.72  0.01  0.52  0.483 1.0
X7  0.44  0.67  0.64  0.356 1.7
X10 0.49  0.43  0.42  0.577 2.0

      SS loadings      ML1  ML2
Proportion Var      3.10 2.85
Cumulative Var      0.39 0.36
Proportion Explained 0.52 0.48
Cumulative Proportion 0.52 1.00

Mean item complexity = 1.4
Test of the hypothesis that 2 factors are sufficient.

df null model = 28 with the objective function = 8.72 with Chi Square = 666.72
df of the model are 13 and the objective function was 1.53

The root mean square of the residuals (RMSR) is 0.07
The df corrected root mean square of the residuals is 0.1

The harmonic n.obs is 81 with the empirical chi square 22.29 with prob < 0.051
The total n.obs was 81 with Likelihood Chi Square = 114.89 with prob < 0.000000000000000002

Tucker Lewis Index of factoring reliability = 0.65
RMSEA index = 0.311 and the 90 % confidence intervals are 0.262 0.367
BIC = 57.76
Fit based upon off diagonal values = 0.99
Measures of factor score adequacy

      ML1  ML2
Correlation of (regression) scores with factors 0.98 0.97
Multiple R square of scores with factors 0.96 0.94
Minimum correlation of possible factor scores 0.92 0.89
```

Figura 20. Rezultatele analizei ML

- **Încărcăturile standardizate (Standardized Loadings):** Variabilele X4, X5, X6, X3 (în această ordine descrescătoare după valoarea încărcăturii pe ML1) prezintă încărcături mari pe primul factor (ML1). Acest lucru indică faptul că aceste variabile sunt bine reprezentate de acest factor. În schimb, ML2 contribuie mai mult la variabilele X1, X2 și X7, care au încărcături mari, sugerând că ele sunt mai bine explicate de acest factor. Iar variabila X10 prezintă încărcături moderate pe ambii factori, având o contribuție mai redusă.
- **Comunalitatea (h2)** reflectă cât din variabilitatea fiecărei variabile este explicată de cei doi factori. De exemplu, variabilele X1, X4 și X5 au comunalități foarte mari (0.90, 0.91 și 0.95), ceea ce indică faptul că 90/91/95% din variabilitatea lor este explicată de factori. Variabilele X6 și X10 au comunalități mai mici (0.52 și 0.42), deci sunt mai slab reprezentate de cei doi factori.
- **Unicitatea (u2)** este partea neexplicată de factori. Pentru X5 ($u2 = 0.051$), doar 5.1% din variabilitatea sa nu este captată de factori, ceea ce o face una dintre cele mai bine explicate variabile. X10 ($u2 = 0.577$) are cea mai mare unicitate, ceea ce înseamnă că 57.7% din variabilitatea sa nu este captată de cei doi factori.

- **Indicele de complexitate Hoffman (com)** măsoară numărul de factori necesari pentru a explica o variabilă. Variabilele X1, X2 și X6 sunt foarte simple (com = 1.0), fiind asociate clar cu un singur factor. Variabilele X3 și X10 sunt mai complexe (com = 1.9 și 2.0), având contribuții semnificative la ambii factori.
- **SS Loadings:** ML1 explică 3.10 din variabilitatea totală, ceea ce corespunde la 39% din informație. ML2 explică 2.85, adică 36% din variabilitatea totală. Împreună, cei doi factori explică 74% din variația totală, ceea ce reprezintă o acoperire bună pentru analiza factorială.
- **Proportia explicata (Proportion Explained):** ML1 contribuie mai mult la explicație (52% din totalul informației explicate), iar ML2 contribuie mai puțin (48%).
- **RMSR (Root Mean Square Residual):** Valoarea mică a RMSR (0.07) indică o bună potrivire între matricea de corelație observată și cea reprodusă de model.
- **Tucker-Lewis Index (TLI):** Valoarea de 0.65 sugerează că modelul poate fi îmbunătățit.
- **RMSEA (Root Mean Square Error of Approximation):** Valoarea ridicată (0.311) și intervalele de încredere (0.262–0.367) indică o potrivire mai puțin ideală a modelului.
- **Fit based upon off diagonal values** are o valoare ridicată, de 0.99, ceea ce indică faptul că modelul se potrivește bine datelor în ceea ce privește explicarea corelațiilor în afara diagonalei principale.
- **Corelatia scorurilor cu factorii** prezintă valori foarte mari (0.98 pentru ML1 și 0.97 pentru ML2), indicând o estimare precisă.
- **Coeficientul de determinare R^2 :** ML1 (0.96) și ML2 (0.94) arată că modelul explică bine variabilitatea datelor.

Determinarea celui mai potrivit model

Pe baza outputurilor, observăm că modelul ML (Maximum Likelihood) oferă performanțe mai bune comparativ cu modelul PA (Principal Axis).

Indicatorii de potrivire globală arată că RMSEA este mai mic pentru ML (0.311) decât pentru PA (0.332), deși ambele valori sunt peste pragul acceptabil (<0.08), indicând o potrivire slabă. TLI este de asemenea mai bun pentru ML (0.65 față de 0.601), dar ambele sunt sub pragul de 0.9 pentru o potrivire acceptabilă. În ceea ce privește Chi-pătratul ajustat, ML are o valoare de 22.29 ($p = 0.051$), mai aproape de pragul de semnificație, comparativ cu PA (17.14, $p = 0.19$). Astfel, deși niciun model nu atinge standardele ideale, ML prezintă o potrivire globală ușor mai bună.

La nivelul comunalităților (h^2), ML oferă valori între 0.42 și 0.95, cu o distribuție mai echilibrată, chiar dacă X10 rămâne slab explicată (0.42). În schimb, PA are comunalități între 0.49 și 0.97, însă variabilele X3 și X10 sunt sub pragul de 0.5, ceea ce sugerează o reprezentare slabă. Din punct de vedere al sarcinilor factoriale, ML produce factori mai bine separați și

interpretabili. În schimb, PA prezintă un prim factor dominant, iar al doilea are sarcini scăzute și greu de interpretat.

Deși RMSR este aproape similar pentru ambele modele (0.06 pentru PA și 0.07 pentru ML), ceea ce avantajează ușor PA, proporția variabilității explicate este similară între modele: 75% pentru PA (58% + 17%) și 74% pentru ML (39% + 36%). Totuși, ML oferă o distribuție mai echilibrată între factori.

În concluzie, modelul ML este mai potrivit, datorită factorilor mai bine separați și interpretabili, comunalităților mai echilibrate și unor indicatori globali ușor mai buni comparativ cu PA.

Diagramele modelelor

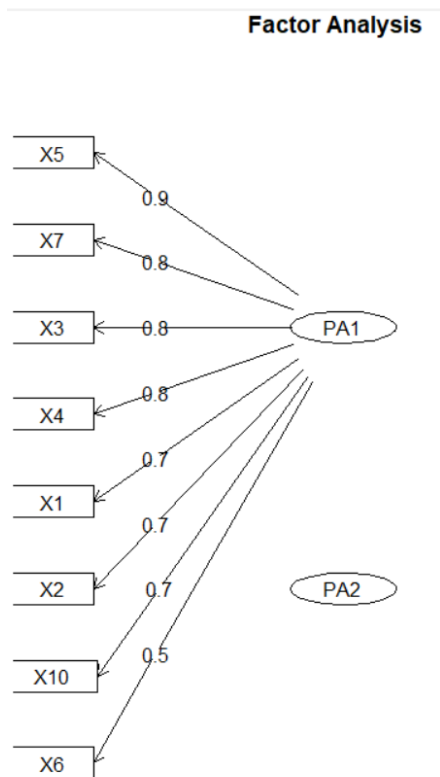


Figura 21. Diagrama modelului PA

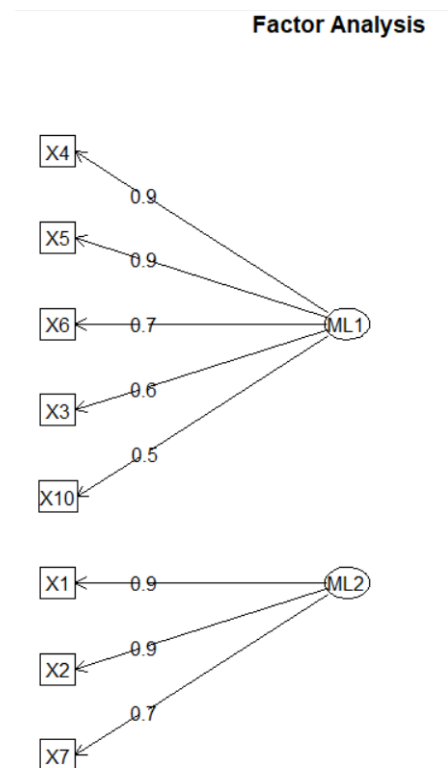


Figura 22. Diagrama modelului ML

Diagrama modelului PA indică faptul că Factorul 1 se corelează puternic cu toate variabilele, având sarcini factoriale mari (peste 0.7), ceea ce sugerează că explică o parte semnificativă din variabilitatea acestora. În schimb, Factorul 2 nu prezintă corelații relevante cu nicio variabilă, având sarcini factoriale mici și inconsistente (sub 0.5), ceea ce indică o capacitate slabă de a explica variabilitatea variabilelor din model.

Pentru modelul ML, ambii factori prezintă corelații puternice cu variabilele: Factorul 1 se corelează semnificativ cu X4, X5, X6, X3 și X10, în timp ce Factorul 2 are o legătură strânsă cu X1, X2 și X7. Factorul 1 are sarcini factoriale ridicate, precum 0.94 pentru X4 și 0.91 pentru X5, indicând că ML1 explică bine variabilitatea acestor variabile. De asemenea, Factorul 2 prezintă sarcini mari, sugerând corelații puternice și o contribuție clară la explicarea variabilelor asociate.

Spre deosebire de primul model, ambii factori din modelul ML explică clar câte un grup distinct de variabile, ceea ce îl face mai bine definit și mai potrivit. Această claritate este evidentă și din structura mai bine organizată a diagramei aferente.

Pe baza structurii sarcinilor factoriale din acest model (ML), putem atribui denumiri sugestive celor doi factori. Factorul 1 se corelează puternic cu variabilele X4, X5, X6, X3 și X10, care sunt legate de profitabilitatea și performanța financiară generală a companiei. Astfel, acesta poate fi denumit „Performanța financiară”. Factorul 2, care se corelează puternic cu X1, X2 și X7, reflectă abilitatea companiei de a genera venituri și de a menține un flux operațional solid, indicând sănătatea operațională și eficiența în generarea profitului brut. Prin urmare, acesta poate fi denumit “Performanța operațională și venituri”.

Deoarece ML este modelul cel mai potrivit, noul set de date va fi salvat cu scorurile factoriale calculate pe baza acestui model. Asadar, am generat scorurile factoriale, le-am adăugat la setul de date original și le-am redenumit sugestiv. Acesta este noul set de date:

	X1	X2	X3	X4	X5	X6	X7	X10	Performanta financiara	Performanță operațională și venituri
1	-0.667892429	-0.472650050	-2.33103597	-2.1192932313	-2.05318491	-1.841847057	-0.92638446	-0.70766227	-2.12044976	-0.364309617
2	1.790693001	1.206971142	-1.04036456	-0.6381100433	-0.74179151	-0.649061628	0.85430741	1.97329346	-1.22096629	1.611272693
3	-0.854734345	-0.997912589	-0.77455891	-0.4727236926	-0.60537112	-0.578408833	-1.02706773	-0.81639421	-0.34485656	-0.913597114
4	-0.844918831	-0.977051219	-0.91891164	-0.6386271650	-0.75570736	-0.557628599	-1.17183875	-1.03482625	-0.51573683	-0.897550317
5	-0.651501598	-0.467732676	-0.70450331	-0.2291913255	-0.53337219	-0.366450447	-0.72233356	-0.48394053	-0.29545101	-0.582677706
6	-0.849558171	-0.989999857	-1.23203352	-0.8100625619	-1.02629012	-0.732182564	-1.21978210	-1.19973485	-0.76954917	-0.894578898
7	-0.690816688	-0.903163709	-0.47131640	-0.2228166842	-0.58719688	-0.391386727	-0.21452747	-0.45748869	-0.23413325	-0.744684722
8	-0.339243194	-0.267694321	-1.75129873	-2.0487029366	-1.59112500	-1.276624693	-1.15305479	-1.07121428	-1.85469965	-0.140019695
9	-0.832534512	-0.954947936	-0.78590478	-0.4505449204	-0.61844139	-0.474507663	-0.88796944	-1.11964567	-0.35243911	-0.885911084
10	-0.427647992	-0.104652247	0.03004761	0.6387334898	0.15298344	-0.270861370	-2.31509422	-2.84070040	0.34703917	-0.572659162
11	-0.267051040	0.416353710	-0.78968255	-1.4010240544	-0.52863831	-0.574252786	-0.77956607	-1.90939191	-1.01520031	0.129208638
12	0.008468326	-0.237332669	1.45743096	1.2867315828	1.51599271	-0.233456949	3.00918761	1.51031924	1.55204491	0.033941501
13	-0.361497343	-0.362119061	-1.18074649	-0.9668972531	-0.81908882	-0.570096739	0.23294376	0.78068259	-0.82983867	-0.205351786
14	-0.483880824	-0.530657808	0.94070516	-0.7356193919	-0.17883284	-0.449571382	1.42829991	-0.38827371	-0.21562247	-0.214867719
15	-0.648673753	-0.618868125	-0.99428670	-0.6184689977	-0.77458196	-0.628281394	-1.02163638	-0.85435675	-0.61044234	-0.628159534
16	-0.690035116	-0.620326698	-1.76855318	-1.4668868359	-1.52760769	-0.890112342	-1.03702743	-0.90181328	-1.44132199	-0.553868507
17	-0.515538306	-0.106385290	0.73089936	1.2201218478	0.85613366	3.041507920	0.14304272	0.24479161	1.32304012	-0.533135418
18	-0.412568894	-0.727318733	-0.38637281	-0.2324249318	-0.34302404	-0.333202072	-0.57600913	-0.70510284	-0.16943732	-0.560261143

Figura 23. Noul set de date

Reprezentari grafice

- **Scatter Plot-ul observatiilor /Score Plot**

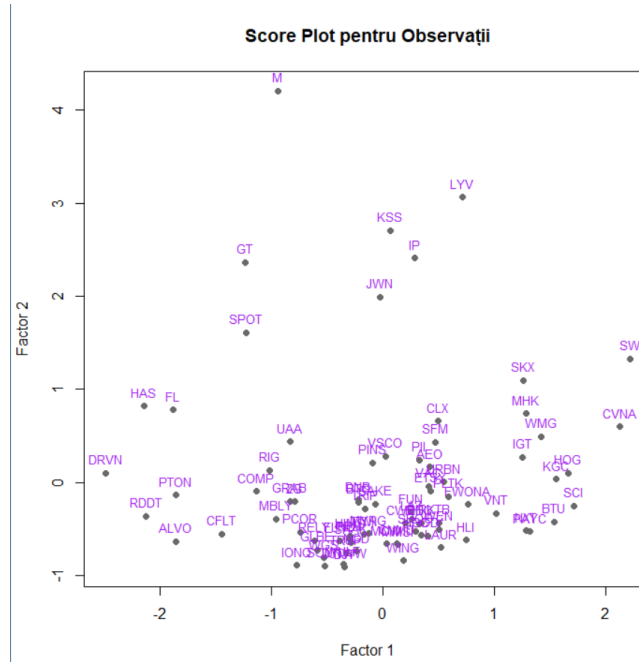


Figura 24. Scatterplot-ul observatiilor

Graficul prezintă prezintă observațiile în funcție de primele două componente principale (sau factori) extrase din analiza factorială. Așadar, axa X reprezintă scorurile pentru primul factor (ML1), iar axa Y reprezintă scorurile pentru al doilea factor (ML2).

Observam ca in partea inferioara a graficului exista un cluster, ceea ce sugereaza faptul ca exista un grup de observatii similare intre ele in ceea ce priveste performantele pe piata bursiera. Acesta indica un segment de companii cu performante asemanatoare. Acest grup se situeaza in apropierea originii (0), ceea ce inseamna ca au valori apropiate de medie pentru cei 2 factori.

Pe de alta parte, observam si cateva observatii raspandite uniform ceea ce inseamna ca se diferentiaza de celelalte companii pe piata, existand o diversitate mare in date. Acestea sunt departe de origine, ceea ce inseamna ca sunt foarte influentate de cei doi factori si difera semnificativ de media generala. Observam si valori extreme, observatii semnificativ diferite fata de restul datelor, cum ar fi companiile M, LYV, KSS, GT, etc.

- **Graficul incarcaturilor/Loading Plot**

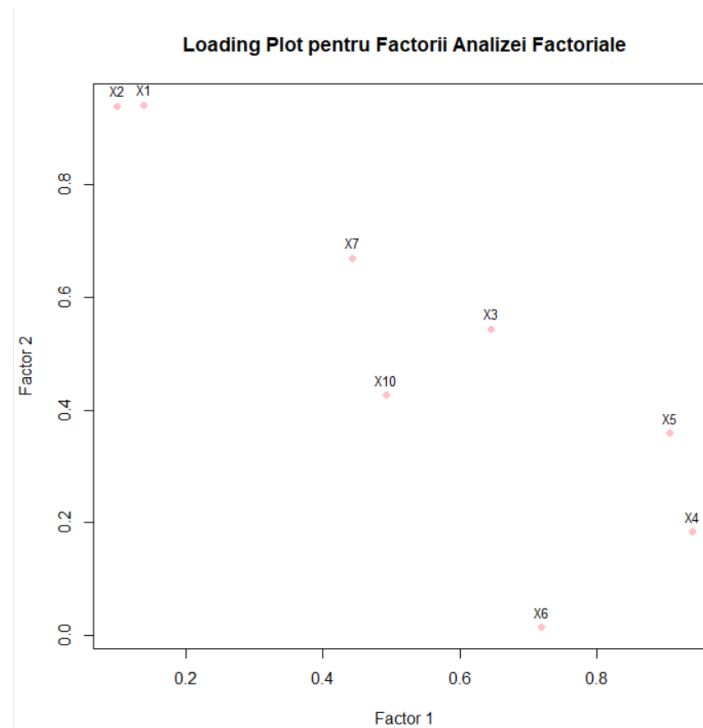


Figura 25. Graficul incarcaturilor

Loading Plot-ul arată relațiile dintre variabilele originale și factorii obținuți în cadrul analizei și ne ajută să înțelegem cum sunt "încărcate" variabilele pe fiecare factor și cât de mult contribuie fiecare variabilă la explicarea factorilor respectivi.

Axa X arată încărcăturile (coeficienții) pentru primul factor (ML1), iar axa Y pentru al doilea factor (ML2). Fiecare punct de pe grafic reprezintă o variabilă originală și poziția sa pe acest grafic indică cât de mult contribuie respectiva variabilă la acești doi factori.

Observăm că X1 și X2 sunt plasate aproape de axa Y (Factorul 2), ceea ce sugerează că aceste variabile (Venitul total și Profitul brut) au o influență puternică asupra lui ML2 și sunt strâns corelate, explicând o mare parte din variabilitatea acestui factor. În același timp, fiind aproape de axa Y, aceste variabile au corelații scăzute cu factorul 1.

Variabila X6 este cea mai apropiată de axa X (Factorul 1), indicând faptul că aceasta (EPS) are cea mai mare influență asupra ML1 și contribuie semnificativ la variabilitatea factorului 1. În schimb, corelația sa cu factorul 2 este slabă.

X7, X3 și X10 se află în centrul graficului, având corelații puternice atât cu ML1, cât și cu ML2, ceea ce sugerează că aceste variabile influențează ambii factori.

În final, X5 și X4 sunt plasate aproape de axa X, indicând o influență mai mare asupra ML1 decât asupra ML2, având corelații mai puternice cu primul factor.

- **Biplot**

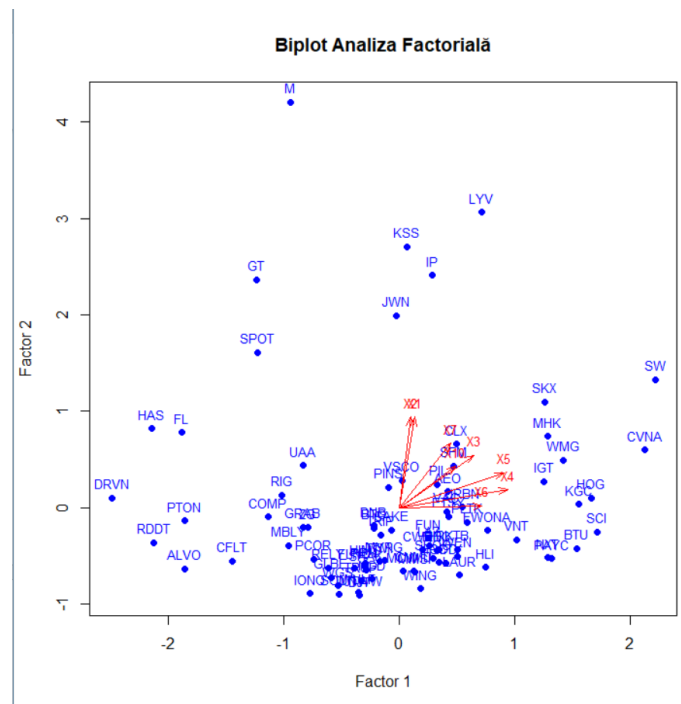


Figura 26. Graficul Biplot

Graficul Biplot combina atât informațiile despre observații, cât și despre variabile, reprezentându-le într-un singur spațiu bidimensional. Punctele reprezintă observațiile, iar săgețile variabilele originale.

Lungimea săgeților indică impactul fiecărei variabile asupra factorilor, iar direcția lor reflectă corelațiile între variabile. Cele mai lungi săgeți sunt pentru variabilele X1, X2, X4 și X5, care au cel mai mare impact asupra factorilor, iar X1 și X2 sunt foarte apropiate, ceea ce sugerează o corelație puternică între ele. Observațiile situate aproape de o săgeată sunt influențate semnificativ de acea variabilă. De exemplu, companiile VSCO și PINS sunt influențate de X1 și X2, iar CLX și SFM de X7, și SFM și AEO de X3. În partea inferioară a graficului, observăm un grup de observații similar, ce include companii asemănătoare din punct de vedere al performanței financiare. Totuși, observațiile mai dispersate (ex: SPOT, JWN, FL, HAS, etc.) sunt mai bine explicate de factori, deoarece reflectă diferențe mai mari între ele.

Concluzii

În urma analizei factoriale, am identificat numărul optim de factori și faptul că modelul cel mai adecvat pentru analiza datelor este modelul verosimilității maxime (ML – maximum likelihood). Aceasta a permis obținerea a doi factori principali, care pot fi denumiți „Performanța financiară” și „Performanța operațională și venituri”, în funcție de variabilele corelate cu fiecare factor. Acești factori explică o parte semnificativă din variabilitatea datelor și sunt esențiali pentru evaluarea generală a performanței companiilor.

De asemenea, am evaluat contribuția fiecărei variabile la factori. Variabilele X1, X2 și X6 au un impact major asupra factorilor, în special X1 (Venit total) și X2 (Profit brut), care sunt puternic corelate și influențează considerabil performanța operațională și veniturile. Variabila X6 (EPS), pe de altă parte, are o influență semnificativă asupra performanței financiare. Analiza a relevat, totodată, existența unui grup (cluster) de companii similare în termeni de performanță, precum și câteva companii dispersate care sunt explicate de factori diferiți. În plus, au fost identificate și câteva companii „outlieri”, care au o performanță mai bună decât majoritatea și care se abat semnificativ de la tendințele generale.

ANALIZA CORESPONDENTELOR

Tabelul de contingență, folosit în analiza corespondentelor, ilustrează numărul de absolvenți din diferite regiuni ale României, în funcție de tipul de liceu absolvit. Fiecare linie reprezintă o regiune a țării, iar fiecare coloană corespunde unui tip de liceu, clasificat după profil: tehnic, resurse, servicii, artistic și educație fizică și sport. Cele 8 regiuni incluse în tabel sunt: Nord-Vest, Centru, Nord-Est, Sud-Est, Sud-Muntenia, București-Ilfov, Sud-Vest Oltenia și Vest. Coloanele reflectă cele 5 tipuri de licee, cu numărul de absolvenți corespunzător fiecărui profil. Fiecare celulă din matrice arată numărul de absolvenți dintr-o regiune specifică, care au terminat un tip de liceu anume. Astfel, tabelul oferă o imagine detaliată a distribuției absolvenților pe regiuni și tipuri de liceu.

Setul de date utilizat provine din sursa INSSE (TEMPO), secțiunea SCL109B - Absolvenți pe niveluri de educație, forme de învățământ, macroregiuni, regiuni de dezvoltare și județe, pentru anul 2022.

	Regiuni	Licee profil tehnic	Licee profil resurse	Licee profil servicii	Licee profil artistic	Licee educatie fizica si sport
1	NORD-VEST	2231	883	2857	614	661
2	CENTRU	2285	981	2557	503	333
3	NORD-EST	2958	1687	4456	619	767
4	SUD-EST	2408	1482	3296	552	669
5	SUD-MUNTENIA	3364	1094	2932	342	599
6	BUCURESTI-ILFOV	2060	554	2861	523	837
7	SUD-VEST OLTENIA	3630	1164	1838	303	534
8	VEST	2434	767	1892	428	271

Figura 27. Tabelul de contingenta

Obiectivul general al analizei este de a explora și înțelege distribuția absolvenților din diferite regiuni ale României, în funcție de tipul de liceu absolvit, utilizând analiza corespondentelor. Prin aceasta, se urmărește identificarea relațiilor și pattern-urilor între regiunile geografice și profilurile liceale, pentru a obține o imagine clară a diversității educaționale la nivel național. Analiza va oferi informații relevante despre preferințele educaționale din fiecare regiune, contribuind astfel la o mai bună înțelegere a tendințelor educaționale și a diversității acestora în contextul regiunilor din România.

Reprezentarea grafica a tabelului de contingenta

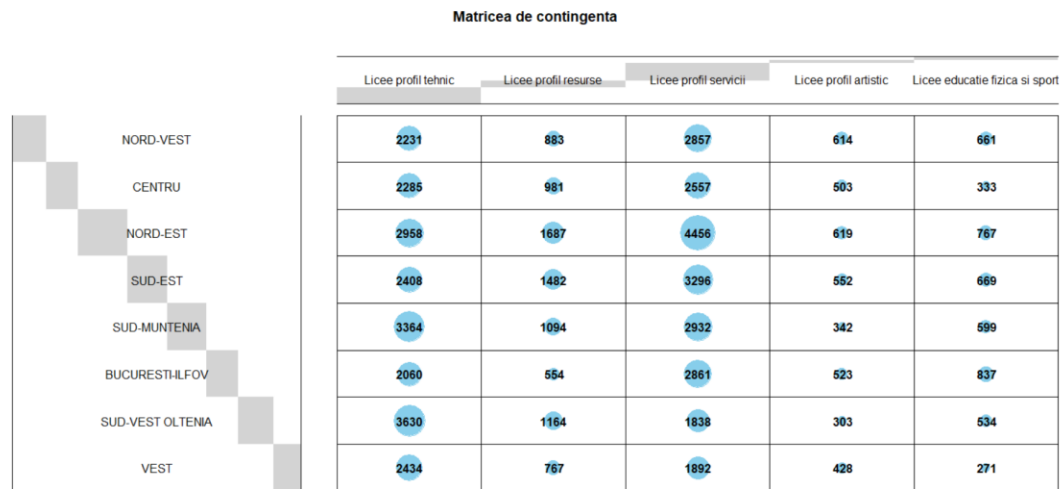


Figura 28. Reprezentarea grafica a tabelului de contingenta

Testul Chi-Patrat

Testul de independenta Chi-patrat este utilizat pentru a verifica dacă există o asociere între variabilele analizate.

Ipotezele testului sunt:

H₀ - nu există asocieri între variabile (liniile și coloanele matricei sunt independente)

H₁ - există asocieri între variabile.

Rezultatul testului a indicat un p-value de 2.2e-16 (0.000000000000000022) , care este semnificativ mai mic decât 0.05, ceea ce duce la respingerea ipotezei nule și acceptarea ipotezei alternative, sugerând astfel că **există o asociere semnificativă între variabile**.

Valoarea statisticii X-squared de 2259.4, care este foarte mare, confirmă această asociere. În plus, gradele de libertate (df) ale testului, calculate ca $(m-1)(n-1) = 7*4 = 28$, reflectă dimensiunea testului și contribuie la validarea rezultatului.

```
> x2=chisq.test(date_ac)
> x2
```

Pearson's Chi-squared test

```
data: date_ac
X-squared = 2259.4, df = 28, p-value < 2.2e-16
```

Figura 29. Testul Chi-patrat

Metoda analizei de corespondenta (AC)

Am realizat analiza corespondentelor (CA – Correspondence Analysis) pe setul de date `date_ac`, utilizand functia CA pentru a explora relațiile între variabilele categorice dintr-un tabel de contingență. Valorile numerice ale rezultatelor analizei vor fi stocate in variabila `rez`.

```
> rez$eig
      eigenvalue percentage of variance cumulative percentage of variance
dim 1 0.024593327          66.644312          66.64431
dim 2 0.007868536          21.322579          87.96689
dim 3 0.003114519           8.439890          96.40678
dim 4 0.001325983           3.593219         100.00000
```

Figura 30. Output valori proprii, procentul de varianta si procentul cumulat

- **Valorile proprii** pentru cele patru dimensiuni/variabile sunt 0.02, 0.007, 0.003 și 0.001. Dimensiunea 1 are cea mai mare valoare proprie, ceea ce înseamnă că aceasta capturează cea mai mare parte din variația totală a datelor. În schimb, dimensiunea 4 are cea mai mică valoare proprie, indicând o importanță scăzută.
- În ceea ce privește **procentul de varianta** explicat, dimensiunea 1 explică 66.64% din variația totală, dimensiunea 2 21.32%, dimensiunea 3 8.44%, iar dimensiunea 4 doar 3.59%. Astfel, prima componentă este cea mai semnificativă.
- **Procentul cumulat** al variației arată că, după prima componentă (66.64%), adăugarea celei de-a doua dimensiuni aduce un total de 87.97%, iar cu toate cele patru dimensiuni se acoperă 100% din variație. Totuși, ultimele componente contribuie foarte puțin la explicarea variației.
- Numărul de dimensiuni este determinat de formula $\min((m-1), (n-1)) = \min(7, 4) = 4$ dimensiuni (unde m - nr de linii, n – nr de coloane).
- **Inertia totala**

```
> s=sum(eig[,1])
> s
[1] 0.03690236
```

Figura 31. Valoarea inertiei totale

Suma valorilor proprii reprezintă ineria totală. Inertia totală de 0.03690236 indică faptul că întreaga variabilitate a datelor este de aproximativ 0.0369, adică valoarea totală a tuturor valorilor proprii.

Graficul Scree-plot

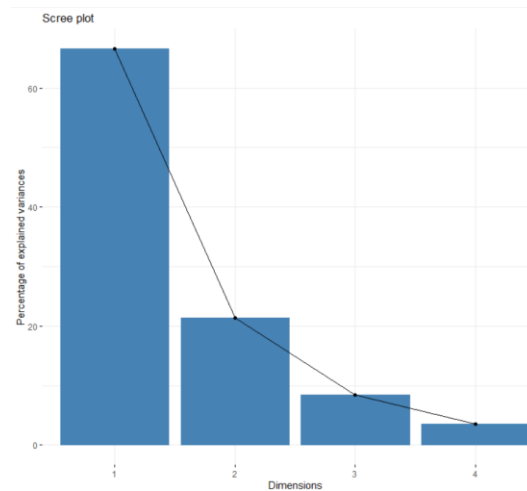


Figura 32. Scree-plot

Conform graficului Scree, se poate observa că prima dimensiune explică mai mult de 60% din variația totală, acoperind majoritatea variabilității în date. După aceasta, urmează o scădere abruptă, ceea ce sugerează că dimensiunile următoare contribuie cu o cantitate mică de variabilitate suplimentară.

Indicatorii pentru dimensiunea linie

```
> rez$row
$coord
      Dim 1      Dim 2      Dim 3      Dim 4
NORD-VEST  0.11308506  0.056804605 -0.03851964  0.038124787
CENTRU     0.00626054  -0.069548091 -0.08511516 -0.011941324
NORD-EST   0.12182630  -0.086756997  0.03304808 -0.026352953
SUD-EST    0.09485720  -0.093759821  0.04433036  0.044287446
SUD-MUNTENIA -0.11008079  0.026702750  0.03555470 -0.064853514
BUCURESTI-ILFOV 0.17388563  0.195343619  0.01447493 -0.003195431
SUD-VEST_OLTENIA -0.31760826  0.029369392  0.05080037  0.037999958
VEST       -0.14422812  -0.004734154 -0.11186734 -0.002483143

$contrib
      Dim 1      Dim 2      Dim 3      Dim 4
NORD-VEST  6.15397309  4.85328292  5.6381401  12.97297114
CENTRU     0.01733323  6.68574234  25.2985740  1.16960782
NORD-EST   10.33665929  16.38439802  6.0064350  8.97090158
SUD-EST    5.02375574  15.34068591  8.6639660  20.31087233
SUD-MUNTENIA 6.70450966  1.23304609  5.5228639  43.16089335
BUCURESTI-ILFOV 13.72501199  54.13861873  0.7510077  0.08596539
SUD-VEST_OLTENIA 50.03718571  1.33728067  10.1080960  13.28479796
VEST       8.00157128  0.02694532  38.0109172  0.04399042

$cos2
      Dim 1      Dim 2      Dim 3      Dim 4
NORD-VEST  0.674760332  0.1702574787  0.078289499  0.0766926909
CENTRU     0.003196065  0.3944231900  0.590752960  0.0116277853
NORD-EST   0.614431781  0.3116022590  0.045215147  0.0287508134
SUD-EST    0.414356125  0.4048243845  0.090497313  0.0903221773
SUD-MUNTENIA 0.662140043  0.0389617791  0.069074997  0.2298231809
BUCURESTI-ILFOV 0.440664266  0.5561333143  0.003053607  0.0001488125
SUD-VEST_OLTENIA 0.953790356  0.0081556627  0.024400746  0.0136532354
VEST       0.623841072  0.0006721385  0.375301873  0.0001849172

$inertia
[1] 0.002242969 0.001333771 0.004137365 0.002981756 0.002490201 0.007659884 0.012902006 0.003154413
```

Figura 33. Indicatorii pentru dimensiunea linie

- **Cos²:**

```
> rez$row$cos2
```

	Dim 1	Dim 2	Dim 3	Dim 4
NORD-VEST	0.674760332	0.1702574787	0.078289499	0.0766926909
CENTRU	0.003196065	0.3944231900	0.590752960	0.0116277853
NORD-EST	0.614431781	0.3116022590	0.045215147	0.0287508134
SUD-EST	0.414356125	0.4048243845	0.090497313	0.0903221773
SUD-MUNTENIA	0.662140043	0.0389617791	0.069074997	0.2298231809
BUCURESTI-ILFOV	0.440664266	0.5561333143	0.003053607	0.0001488125
SUD-VEST OLTENIA	0.953790356	0.0081556627	0.024400746	0.0136532354
VEST	0.623841072	0.0006721385	0.375301873	0.0001849172

Figura 34. Cos2 pentru dimensiunea linie

Cos² reflectă calitatea explicării fiecărei variabile (regiuni) pe dimensiunile respective, indicând cât de bine se poate prezenta fiecare regiune pe acele dimensiuni. Pe dimensiunea 1, majoritatea variabilelor (cu excepția regiunii CENTRU) au valori foarte mari, ceea ce înseamnă că această dimensiune explică foarte bine variația acestora. Pe dimensiunea 2, valorile sunt moderate, cu valori mai mari înregistrate pentru regiunile București-Ilfov și Sud-Est. Dimensiunea 3 explică foarte bine variabila CENTRU, spre deosebire de dimensiunea 1, care avea o valoare extrem de mică pentru aceasta. În ceea ce privește celelalte variabile, acestea sunt explicate într-o proporție mică. Dimensiunea 4 explică cel mai puțin variația variabilelor, având valori scăzute, cu excepția regiunii Sud-Muntenia, care înregistrează o valoare mai ridicată.

- **Contributia;**

```
> rez$row$contrib
```

	Dim 1	Dim 2	Dim 3	Dim 4
NORD-VEST	6.15397309	4.85328292	5.6381401	12.97297114
CENTRU	0.01733323	6.68574234	25.2985740	1.16960782
NORD-EST	10.33665929	16.38439802	6.0064350	8.97090158
SUD-EST	5.02375574	15.34068591	8.6639660	20.31087233
SUD-MUNTENIA	6.70450966	1.23304609	5.5228639	43.16089335
BUCURESTI-ILFOV	13.72501199	54.13861873	0.7510077	0.08596539
SUD-VEST OLTENIA	50.03718571	1.33728067	10.1080960	13.28479796
VEST	8.00157128	0.02694532	38.0109172	0.04399042

Figura 35. Contributia pentru dimensiunea linie

Contribuția reflectă cât de mult fiecare unitate (regiune) contribuie la explicarea variabilității pe dimensiunea respectivă. Pe dimensiunea 1, regiunea Sud-Vest Oltenia are o contribuție semnificativă (50,04%), având un rol dominant în descrierea acestei dimensiuni. București-Ilfov are o contribuție mare pe dimensiunea 2 (54,14%), fiind o regiune semnificativă.

Pe dimensiunea 3, regiunile Vest și Centru prezintă cele mai puternice contribuții în definirea acestei dimensiuni. De asemenea, observăm că regiunile Nord-Est și Sud-Est au contribuții semnificative pe mai multe dimensiuni, indicând un comportament mai echilibrat în descrierea acestora pe diverse dimensiuni.

- **Coordonatele:**

```
> rez$row$coord
```

	Dim 1	Dim 2	Dim 3	Dim 4
NORD-VEST	0.11308506	0.056804605	-0.03851964	0.038124787
CENTRU	0.00626054	-0.069548091	-0.08511516	-0.011941324
NORD-EST	0.12182630	-0.086756997	0.03304808	-0.026352953
SUD-EST	0.09485720	-0.093759821	0.04433036	0.044287446
SUD-MUNTENIA	-0.11008079	0.026702750	0.03555470	-0.064853514
BUCURESTI-ILFOV	0.17388563	0.195343619	0.01447493	-0.003195431
SUD-VEST OLTENIA	-0.31760826	0.029369392	0.05080037	0.037999958
VEST	-0.14422812	-0.004734154	-0.11186734	-0.002483143

Figura 36. Coordonatele pentru dimensiunea linie

Coordonatele reflectă poziția fiecărei regiuni în spațiul dimensional și sunt esențiale pentru a înțelege cum se alocă regiunile în raport cu axele componente principale. Pe dimensiunea 1, regiunile București-Ilfov, Nord-Est și Nord-Vest au cele mai mari valori, indicând că sunt cele mai influente pe această componentă principală. Pe dimensiunea 2, cea mai semnificativă influență o are regiunea București-Ilfov. Coordonatele pe dimensiunea 3 sunt mici pentru majoritatea regiunilor, cu cele mai influente regiuni fiind Sud-Vest Oltenia și Sud-Est, având valori de 0,05, respectiv 0,04. Totuși, aceste valori sunt destul de scăzute, ceea ce sugerează că această dimensiune nu are o contribuție semnificativă. Pe dimensiunea 4, coordonatele sunt cele mai mici, ceea ce indică o contribuție redusă a acestei dimensiuni în descrierea regiunilor.

- **Inertia:**

```
> rez$row$inertia
```

[1]	0.002242969	0.001333771	0.004137365	0.002981756	0.002490201	0.007659884	0.012902006	0.003154413
-----	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------

Figura 37. Inertia pentru dimensiunea linie

Inertia reflectă contribuția fiecărei regiuni la variabilitatea totală a setului de date. Regiunile cu o inertie mare sunt Sud-Vest Oltenia (0.012) și București-Ilfov (0.007), ceea ce sugerează că acestea au o influență semnificativă asupra variabilității totale, fiind probabil regiuni cu valori considerabil diferite față de celelalte pe mai multe dimensiuni. Regiunile cu

inertie medie, precum Vest (0.003) și Nord-Est (0.004), contribuie moderat la variabilitate, arătând diversitate în valori, dar fără o dominanță clară pe vreo dimensiune. Regiunile cu inertie mică, cum sunt Centru (0.001) și Nord-Vest (0.002), au valori apropiate de centru și nu se diferențiază semnificativ de celelalte regiuni, indicând o variabilitate mai mică.

Matricea factor pentru linii

```
> mf_linii <- rez$row$coord
> print(mf_linii)
```

	Dim 1	Dim 2	Dim 3	Dim 4
NORD-VEST	0.11308506	0.056804605	-0.03851964	0.038124787
CENTRU	0.00626054	-0.069548091	-0.08511516	-0.011941324
NORD-EST	0.12182630	-0.086756997	0.03304808	-0.026352953
SUD-EST	0.09485720	-0.093759821	0.04433036	0.044287446
SUD-MUNTENIA	-0.11008079	0.026702750	0.03555470	-0.064853514
BUCURESTI-ILFOV	0.17388563	0.195343619	0.01447493	-0.003195431
SUD-VEST OLTENIA	-0.31760826	0.029369392	0.05080037	0.037999958
VEST	-0.14422812	-0.004734154	-0.11186734	-0.002483143

Figura 38. Matricea factor pt linii

Coordonatele reflectă poziția fiecărei regiuni în spațiul dimensional și sunt esențiale pentru a înțelege cum se alocă regiunile în raport cu axele componente principale. Pe dimensiunea 1, regiunile București-Ilfov, Nord-Est și Nord-Vest au cele mai mari valori, indicând că sunt cele mai influente pe această componentă principală. Pe dimensiunea 2, cea mai semnificativă influență o are regiunea București-Ilfov. Coordonatele pe dimensiunea 3 sunt mici pentru majoritatea regiunilor, cu cele mai influente regiuni fiind Sud-Vest Oltenia și Sud-Est, având valori de 0,05, respectiv 0,04. Totuși, aceste valori sunt destul de scăzute, ceea ce sugerează că această dimensiune nu are o contribuție semnificativă. Pe dimensiunea 4, coordonatele sunt cele mai mici, ceea ce indică o contribuție redusă a acestei dimensiuni în descrierea regiunilor.

Indicatorii pentru dimensiunea coloana

```
> rez$col
$coord
      Dim 1      Dim 2      Dim 3      Dim 4
Licee profil tehnic    -0.19714750  0.03837092 -0.012055154 -0.008213291
Licee profil resurse   -0.03873937 -0.17723100  0.050319057  0.040605135
Licee profil servicii   0.14408784 -0.01439454 -0.005479095 -0.032940760
Licee profil artistic   0.15709195  0.03659223 -0.158211078  0.085810868
Licee educatie fizica si sport 0.14286259  0.19070873  0.120547621  0.051365894

$contrib
      Dim 1      Dim 2      Dim 3      Dim 4
Licee profil tehnic    55.1612226  6.5309902  1.6286329  1.775682
Licee profil resurse   0.8583332  56.1504941  11.4351571  17.490096
Licee profil servicii  31.2836110  0.9758462  0.3571956  30.325561
Licee profil artistic   6.3655207  1.0795113  50.9830884  35.228169
Licee educatie fizica si sport 6.3313124  35.2631582  35.5959260  15.180492

$cos2
      Dim 1      Dim 2      Dim 3      Dim 4
Licee profil tehnic    0.95844590  0.036306919  0.003583692  0.001663487
Licee profil resurse   0.04045952  0.846827632  0.068262256  0.044450593
Licee profil servicii   0.94012228  0.009382649  0.001359400  0.049135671
Licee profil artistic   0.42248600  0.022923568  0.428527029  0.126063405
Licee educatie fizica si sport 0.27599454  0.491818193  0.196508227  0.035679040

$inertia
[1] 0.014154142 0.005217380 0.008183702 0.003705432 0.005641707
```

Figura 39. Indicatorii pentru dimensiunea coloana

- **Cos²:**

```
> rez$col$contrib
      Dim 1      Dim 2      Dim 3      Dim 4
Licee profil tehnic    0.95844590  0.036306919  0.003583692  0.001663487
Licee profil resurse   0.04045952  0.846827632  0.068262256  0.044450593
Licee profil servicii   0.94012228  0.009382649  0.001359400  0.049135671
Licee profil artistic   0.42248600  0.022923568  0.428527029  0.126063405
Licee educatie fizica si sport 0.27599454  0.491818193  0.196508227  0.035679040
```

Figura 40. Cos2 pentru dimensiunea coloana

Cos2 pentru coloanele referitoare la tipurile de licee indică modul în care fiecare profil de liceu este reprezentat pe diferitele dimensiuni. „Licee profil tehnic” este foarte bine reprezentat pe Dimensiunea 1 (0.958), dar mai slab pe celelalte dimensiuni. „Licee profil resurse” este foarte bine reprezentat pe Dimensiunea 2 (0.847). „Licee profil servicii” are o reprezentare puternică pe Dimensiunea 1 (0.94). „Licee profil artistic” este distribuit între Dimensiunea 1 (0.422) și Dimensiunea 3 (0.429), indicând o contribuție semnificativă pe ambele dimensiuni. În final, „Licee educație fizică și sport” sunt reprezentate cel mai bine pe Dimensiunea 2 (0.492), dar și într-o măsură mai mică pe Dimensiunea 3 (0.197).

- **Contributia**

```
> rez$col$contrib
```

	Dim 1	Dim 2	Dim 3	Dim 4
Licee profil tehnic	55.1612226	6.5309902	1.6286329	1.775682
Licee profil resurse	0.8583332	56.1504941	11.4351571	17.490096
Licee profil servicii	31.2836110	0.9758462	0.3571956	30.325561
Licee profil artistic	6.3655207	1.0795113	50.9830884	35.228169
Licee educatie fizica si sport	6.3313124	35.2631582	35.5959260	15.180492

Figura 41. Contributia pentru dimensiunea coloana

Contribuțiile pe fiecare dimensiune arată rolul diferitelor tipuri de licee în explicarea variației datelor. Pe Dimensiunea 1, „Licee profil tehnic” domină, având o contribuție de 55.16%, ceea ce înseamnă că această categorie explică cea mai mare parte a variației pe această dimensiune. Pe Dimensiunea 2, „Licee profil resurse” este principalul contribuitor, cu 56.15%, urmat de „Licee profil educație fizică și sport”, care are o contribuție de 35.26%. Pe Dimensiunea 3, „Licee profil artistic” este dominantă, explicând 50.98% din variație. În ceea ce privește Dimensiunea 4, „Licee profil artistic” (35.23%) și „Licee profil servicii” (30.33%) au contribuții semnificative la variabilitatea pe această dimensiune.

- **Coordonatele**

```
> rez$col$coord
```

	Dim 1	Dim 2	Dim 3	Dim 4
Licee profil tehnic	-0.19714750	0.03837092	-0.012055154	-0.008213291
Licee profil resurse	-0.03873937	-0.17723100	0.050319057	0.040605135
Licee profil servicii	0.14408784	-0.01439454	-0.005479095	-0.032940760
Licee profil artistic	0.15709195	0.03659223	-0.158211078	0.085810868
Licee educatie fizica si sport	0.14286259	0.19070873	0.120547621	0.051365894

Figura 42. Coordonatele pentru dimensiunea coloana

Coordonatele reflectă poziția fiecărui tip de liceu în spațiul generat de dimensiunile principale, indicând cât de puternic și în ce direcție contribuie fiecare tip de liceu la o anumită dimensiune. Pe Dimensiunea 1, „Licee profil tehnic” (-0.197) și „Licee profil resurse” (-0.03) au contribuții negative, ceea ce înseamnă că nu influențează semnificativ această dimensiune, în timp ce celelalte trei tipuri de licee contribuie semnificativ la definirea acestei dimensiuni. Pe Dimensiunea 2, „Licee educație fizică și sport” este cea mai semnificativă, cu celelalte tipuri de licee având contribuții scăzute. Aceeași tendință se observă și pe Dimensiunea 3, unde „Licee educație fizică și sport” joacă un rol dominant. În ceea ce privește Dimensiunea 4, toate contribuțiile sunt scăzute, cu „Licee profil artistic” având cea mai mare valoare (0.08).

- Inertia

```
> rez$col$inertia  
[1] 0.014154142 0.005217380 0.008183702 0.003705432 0.005641707
```

Figura 43. Inertia pentru dimensiunea coloana

Cele mai mari valori ale inerteiei sunt asociate cu „Licee profil tehnic” (0.0142), „Licee profil servicii” (0.0082) și „Licee profil educație fizică și sport” (0.0056). Aceste categorii au un rol semnificativ în explicarea variației totale din setul de date, indicând că acestea contribuie considerabil la diversitatea și variabilitatea observațiilor.

Matricea factor pentru coloane

```
> mf_coloane <- rez$col$coord  
> print(mf_coloane)
```

	Dim 1	Dim 2	Dim 3	Dim 4
Licee profil tehnic	-0.19714750	0.03837092	-0.012055154	-0.008213291
Licee profil resurse	-0.03873937	-0.17723100	0.050319057	0.040605135
Licee profil servicii	0.14408784	-0.01439454	-0.005479095	-0.032940760
Licee profil artistic	0.15709195	0.03659223	-0.158211078	0.085810868
Licee educatie fizica si sport	0.14286259	0.19070873	0.120547621	0.051365894

Figura 44. Matricea factor pt coloane

Coordonatele reflectă poziția fiecărui tip de liceu în spațiul generat de dimensiunile principale, indicând cât de puternic și în ce direcție contribuie fiecare tip de liceu la o anumită dimensiune. Pe Dimensiunea 1, „Licee profil tehnic” (-0.197) și „Licee profil resurse” (-0.03) au contribuții negative, ceea ce înseamnă că nu influențează semnificativ această dimensiune, în timp ce celelalte trei tipuri de licee contribuie semnificativ la definirea acestei dimensiuni. Pe Dimensiunea 2, „Licee educație fizică și sport” este cea mai semnificativă, cu celelalte tipuri de licee având contribuții scăzute. Aceeași tendință se observă și pe Dimensiunea 3, unde „Licee educație fizică și sport” joacă un rol dominant. În ceea ce privește Dimensiunea 4, toate contribuțiile sunt scăzute, cu „Licee profil artistic” având cea mai mare valoare (0.08).

Reprezentari grafice

- **Graficul Biplot**

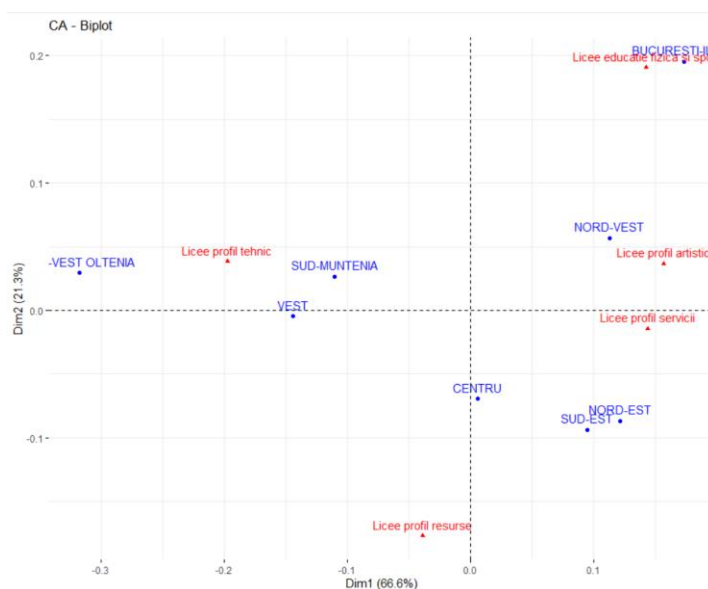


Figura 45. Biplot

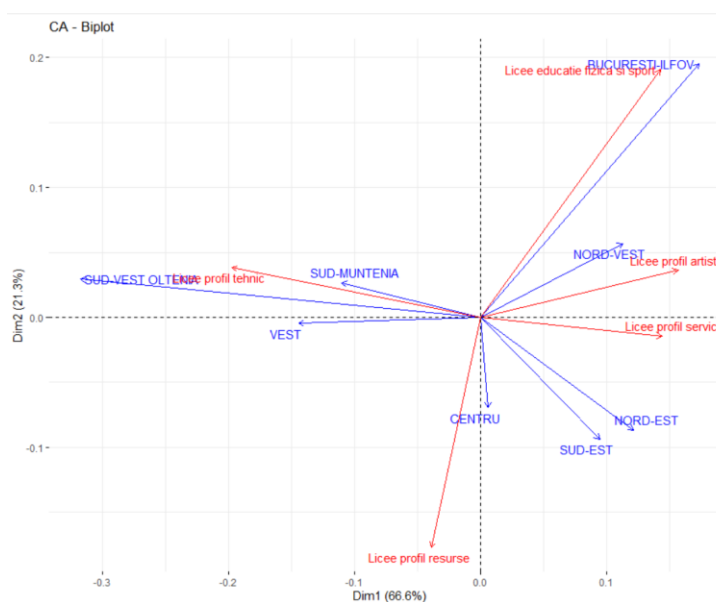


Figura 46. Biplot cu sageti

Graficul Biplot in analiza corespondentelor reprezintă o vizualizare a relațiilor dintre variabilele categorice (de pe linii și coloane) într-un spațiu dimensional redus (pe primele două dimensiuni principale).

Liniile din grafic reprezintă unitățile de observație, adică regiunile din România, iar plasarea acestora în funcție de valorile lor pe dimensiunile principale reflectă asemănările și diferențele dintre regiunile respective. Astfel, regiunile NORD-EST și SUD-EST, respectiv SUD-MUNTENIA și VEST, fiind apropiate între ele pe grafic, sugerează că aceste regiuni sunt mai asemănătoare în ceea ce privește distribuția tipurilor de licee sau profilurile educaționale. Acest lucru poate însemna că au o structură similară a educației sau că tipurile de licee din aceste regiuni sunt relativ comparabile. În schimb, BUCUREȘTI-ILFOV și SUD-VEST OLTENIA sunt plasate mai departe de celelalte regiuni, ceea ce indică diferențe semnificative în ceea ce privește numărul de absolvenți sau profilurile educaționale, poate și datorită diferențelor economice sau de dezvoltare ale acestor regiuni.

Coloanele reprezintă variabilele categoriale, adică tipurile de licee, care sunt plasate în funcție de contribuția lor la variabilitatea datelor pe cele două dimensiuni principale. Direcția în care sunt plasate aceste coloane indică modul în care fiecare tip de liceu influențează structura generală a datelor. O coloană apropiată de o linie sugerează o asociere puternică între tipul de liceu respectiv și regiunea respectivă, în sensul că acest tip de liceu este predominant în acea regiune. De exemplu, BUCUREȘTI-ILFOV este strâns legat de Liceele de educație fizică și sport, ceea ce sugerează că acest tip de liceu este mai frecvent în această regiune, în concordanță cu specificul urban și orientarea către educația fizică. SUD-MUNTENIA are o legătură puternică cu Liceele cu profil tehnic, ceea ce poate reflecta o tradiție mai puternică a educației tehnice în această regiune. CENTRU este legată de Liceele cu profil resurse, ceea ce poate indica o concentrare mai mare de licee de acest tip în centrul țării.

În schimb, NORD-EST și SUD-EST sunt plasate la o distanță mare de Liceele cu profil tehnic, ceea ce sugerează o lipsă de asociere între aceste regiuni și acest tip de liceu. Aceste regiuni pot avea o distribuție mai echilibrată a tipurilor de licee sau o tendință de a se orienta spre alte domenii educaționale.

Lungimea unui vector este un alt indicator important în interpretarea graficului, deoarece reflectă cât de mult o variabilă contribuie la definirea dimensiunii respective. Un vector mai lung sugerează o variabilă importantă pe acea dimensiune, în timp ce un vector mai scurt semnifică o influență mai redusă. În cazul dimensiunii 1, BUCUREȘTI-ILFOV are o lungime mare a vectorului, ceea ce sugerează că această regiune joacă un rol semnificativ în definirea primei dimensiuni, adică are un impact mare asupra diversității profilurilor educaționale. Pe de altă parte, CENTRU, având un vector mai scurt, contribuie mai puțin la această dimensiune, ceea ce poate însemna că structura educațională a acestei regiuni este mai uniformă sau mai puțin variată în raport cu dimensiunea 1.

În ceea ce privește dimensiunea 2, SUD-VEST OLTENIA se remarcă printr-o contribuție mare, ceea ce sugerează că această regiune influențează semnificativ structura educațională pe această dimensiune. În schimb, VEST are o contribuție mică, ceea ce poate indica faptul că profilurile educaționale din această regiune sunt mai echilibrate sau mai puțin specifice față de celelalte regiuni. Aceste observații sugerează că structura educațională a regiunilor analizate este variabilă și că există regiuni cu un profil educațional bine definit, în timp ce altele sunt mai echilibrate în ceea ce privește tipurile de licee.

- **Reprezentarea grafica a contributiilor**

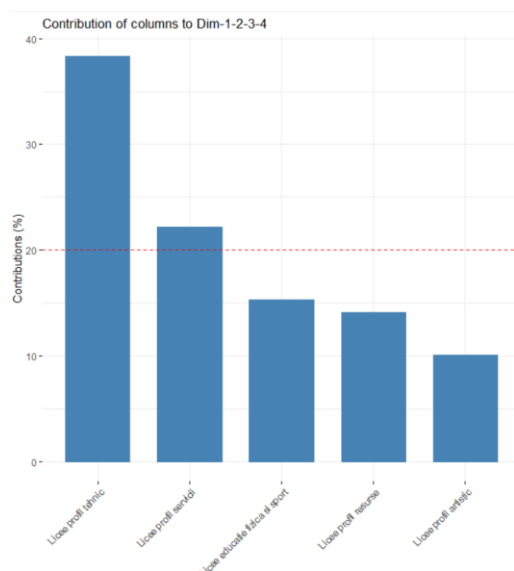


Figura 47. Contributia coloanelor

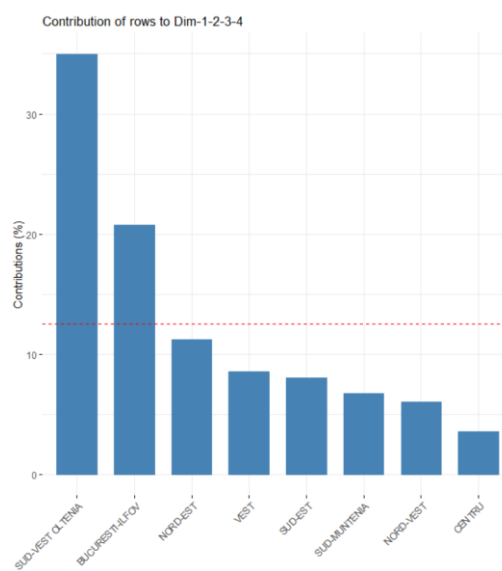


Figura 48. Contributia randurilor

Graficul din figura 44 ilustrează contribuția procentuală a fiecărui tip de liceu la variabilitatea totală a datelor. Această analiză este esențială pentru a înțelege care dintre variabilele categorice joacă un rol dominant în definirea dimensiunilor principale ale analizei. Din grafice, reiese că „Liceele cu profil tehnic” au cea mai mare influență asupra variabilității generale, ceea ce indică faptul că acest tip de licee prezintă o diversitate mai mare sau valori mai diferite față de alte categorii. Următoarele ca importanță sunt „Liceele cu profil servicii”, care contribuie și ele semnificativ la variația datelor. În schimb, celelalte tipuri de licee au contribuții considerabil mai mici, cea mai mică fiind asociată „Liceelor cu profil artistic”. Aceasta sugerează că profilurile artistice sunt mai omogene sau mai puțin relevante în structura generală a datelor.

În ceea ce privește graficul din figura 45, acesta reflectă contribuția fiecărei regiuni la variabilitatea totală. Observăm că regiunea SUD-VEST OLTEA are cea mai mare contribuție, ceea ce sugerează o diversitate mare sau valori extreme în datele corespunzătoare acestei regiuni. Această regiune joacă un rol central în definirea dimensiunilor principale, influențând semnificativ analiza. Pe de altă parte, regiunea CENTRU are cea mai mică contribuție, ceea ce indică o omogenitate mai mare sau valori mai apropiate de media generală.

- **Reprezentarea grafica a calitatii reprezentarii**

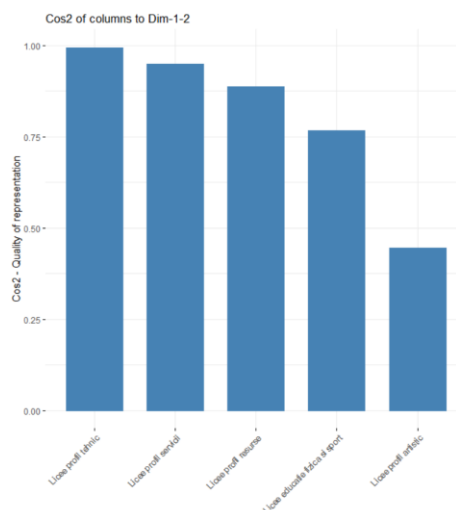


Figura 49. Calitatea reprezentării coloanelor

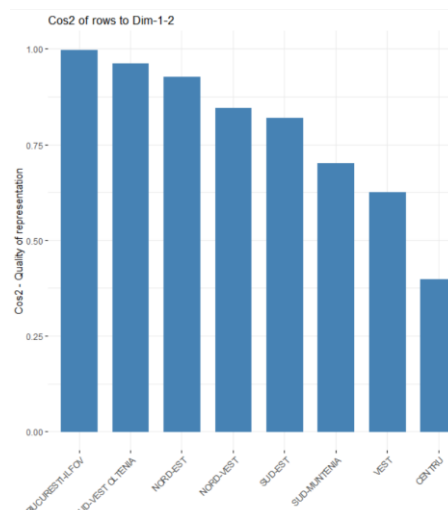


Figura 50. Calitatea reprezentării liniilor

Reprezentarea grafică a calității reprezentării, prezentată în figura 46, arată că majoritatea tipurilor de licee sunt bine reprezentate, în special „Liceele cu profil tehnic” și „Liceele cu profil servicii”, care au un impact semnificativ asupra variabilității totale. Aceste tipuri de licee sunt plasate de-a lungul dimensiunilor principale, indicând o contribuție mare la variabilitatea generală. Pe de altă parte, „Liceele cu profil artistic” sunt mai slab reprezentate, având o contribuție mai mică în definirea dimensiunilor, ceea ce sugerează că aceste licee nu au o influență semnificativă asupra structurii generale a datelor.

În ceea ce privește regiunile, observăm că regiunile BUCUREȘTI-ILFOV, SUD-VEST OLTEA și NORD-EST sunt cele mai bine reprezentate, având o contribuție semnificativă în definirea dimensiunilor principale. Aceste regiuni au o diversitate mai mare de tipuri de licee, ceea ce le face mai reprezentative în analiza globală. În schimb, regiunea CENTRU are o reprezentare mai slabă, având o contribuție mică la variabilitatea totală. Aceasta sugerează că, în comparație cu alte regiuni, CENTRU prezintă o diversitate mai mică în ceea ce privește tipurile de licee sau caracteristicile acestora.

Concluzii

Analiza arată că există o legătură clară între tipurile de licee și regiunile în care acestea sunt mai proeminente. De exemplu, „Liceele cu profil tehnic” sunt asociate cu regiunea SUD-MUNTENIA, iar „Liceele cu profil educație fizică și sport” cu BUCUREȘTI-ILFOV. Acest lucru indică o distribuție neuniformă a absolvenților și reflectă cerințele sau orientările educaționale specifice fiecărei regiuni. „Liceele cu profil tehnic” și „Liceele cu profil servicii” contribuie cel mai mult la variabilitatea totală, indicând o diversitate semnificativă a absolvenților în aceste categorii. În schimb, „Liceele cu profil artistic” și alte tipuri de licee au contribuții mai mici, sugerând o omogenitate mai mare sau o reprezentare mai limitată în dataset.

Regiunile SUD-VEST OLTENIA și BUCUREȘTI-ILFOV se evidențiază ca outliers, având cele mai mari contribuții la variația totală. Acestea prezintă caracteristici distincte comparativ cu alte regiuni. În contrast, regiunile CENTRU și NORD-VEST au contribuții reduse, fiind mai aproape de media generală, ceea ce indică o mai mică diferențiere a acestora față de alte regiuni.

Dimensiunea 1 este dominată de „Liceele cu profil tehnic” și „Liceele cu profil servicii,” reflectând importanța acestor tipuri de licee în explicarea variației. Dimensiunea 2 este influențată în principal de „Liceele cu profil educație fizică și sport,” sugerând o direcție diferită a variabilității în date, specifică anumitor regiuni.

Aceste corelații ar putea reflecta fie priorități educaționale regionale, fie nevoile pieței muncii din respectivele zone. De exemplu, asocierile dintre SUD-MUNTENIA și liceele tehnice ar putea indica o orientare industrială mai pronunțată, în timp ce BUCUREȘTI-ILFOV este corelat cu liceele sportive, posibil din cauza infrastructurii și oportunităților din capitală.

Această analiză oferă o perspectivă valoroasă asupra legăturii dintre distribuția absolvenților pe tipuri de licee și specificitatea regională, evidențiind atât diferențele, cât și potențialele direcții de intervenție educațională.

ANEXA

```
tema <- Date_AD
```

```
View(tema)
```

```
# Eliminarea outlierilor:
```

```
# Iterăm prin coloanele setului de date (excluzând prima coloană)
```

```
for (col in colnames(tema)[-1]) { # Excludem prima coloană folosind [-1]
```

```
  # Identificăm outlierii pentru fiecare coloană
```

```
  outliers <- boxplot(tema[[col]], plot = F)$out
```

```
  # Excludem liniile care conțin outlieri
```

```
  tema <- tema[-which(tema[[col]] %in% outliers), ]
```

```
}
```

Iordan Maria-Alexandra
Grupa 1080-A

```
View(tema)
```

```
# Calcularea indicatorilor statistici
```

```
summary(tema)
```

```
install.packages("psych")
```

```
library(psych)
```

```
describe(tema[-1])
```

```
# Matricea de corelatie si matricea de covarianta
```

```
cor(tema[-1])
```

```
cov(tema[-1])
```

```
#Pentru a observa mai bine rezultatele, vom standardiza datele
```

```
tema_std = scale(tema[-1], scale = T)
```

```
View(tema_std)
```

```
# Recalculam corelatia si covarianta
```

```
matrice_corelatie <- cor(tema_std)
```

```
matrice_covarianta <- cov(tema_std)
```

```
View(matrice_corelatie)
```

```
View(matrice_covarianta)
```

```
# Reprezentarea grafica a matricei de corelatie
```

Iordan Maria-Alexandra
Grupa 1080-A

```
install.packages("corrplot")  
  
library(corrplot)  
  
windows()  
  
corrplot(matrice_corelatie, method = "circle", type = "upper", col = "pink", title =  
"Matricea de corelatie" )
```

Reprezentari grafice

```
options(scipen=999)
```

```
windows()
```

```
par(mfrow=c(2,5)) #afisarea graficelor in aceeasi fereastră
```

Distributia variabilei X1 - histograma + densitate de probabilitate

```
hist(tema$X1, col = "turquoise", main = "Histograma X1", freq = F)
```

```
lines(density(tema$X1), col="blue", lwd=3)
```

Distributia variabilei X2 - histograma + densitate de probabilitate

```
hist(tema$X2, col = "pink", main = "Histograma X2", freq = F)
```

```
lines(density(tema$X2), col="magenta", lwd=3)
```

Distributia variabilei X3 - histograma + densitate de probabilitate

```
hist(tema$X3, col = "green", main = "Histograma X3", freq = F)
```

```
lines(density(tema$X3), col="black", lwd=3)
```

Distributia variabilei X4 - histograma + densitate de probabilitate

```
hist(tema$X4, col = "aquamarine", main = "Histograma X4", freq = F)
```

```
lines(density(tema$X4), col="blue", lwd=3)
```

```
# Distributia variabilei X5 - histograma + densitate de probabilitate
hist(tema$X5, col = "aquamarine", main = "Histograma X5", freq = F)
lines(density(tema$X5), col="magenta", lwd=3)

# Distributia variabilei X6 - histograma + densitate de probabilitate
hist(tema$X6, col = "grey", main = "Histograma X6", freq = F)
lines(density(tema$X6), col="green", lwd=3)

# Distributia variabilei X7 - histograma + densitate de probabilitate
hist(tema$X7, col = "pink", main = "Histograma X7", freq = F)
lines(density(tema$X7), col="red", lwd=3)

# Distributia variabilei X8 - histograma + densitate de probabilitate
hist(tema$X8, col = "yellow", main = "Histograma X8", freq = F)
lines(density(tema$X8), col="black", lwd=3)

# Distributia variabilei X9 - histograma + densitate de probabilitate
hist(tema$X9, col = "purple", main = "Histograma X9", freq = F)
lines(density(tema$X9), col="green", lwd=3)

# Distributia variabilei X10 - histograma + densitate de probabilitate
hist(tema$X10, col = "red", main = "Histograma X10", freq = F)
lines(density(tema$X10), col="green", lwd=3)

windows()

# Dependenta dintre 2 variabile -> regresia
plot(tema$X1, tema$X2, col="black")
```

Iordan Maria-Alexandra
Grupa 1080-A

```
abline(lm(tema$X2~tema$X1), col="purple")
```

```
# Scatter plot
```

```
install.packages("ggplot2")
```

```
library(ggplot2)
```

```
ggplot(tema, aes(x=X1, y=X2))+
```

```
  geom_point()+
```

```
  geom_text(label=tema$Companie, color="magenta",
```

```
            nudge_x = 0.25, nudge_y = 0.25,
```

```
            check_overlap = T)
```

```
par(mfrow=c(2,5))
```

```
# Boxplot
```

```
boxplot(tema$X1, col="green", main="Boxplot X1")
```

```
boxplot(tema$X2, col="purple", main="Boxplot X2")
```

```
boxplot(tema$X3, col="blue", main="Boxplot X3")
```

```
boxplot(tema$X4, col="pink", main="Boxplot X4")
```

```
boxplot(tema$X5, col="magenta", main="Boxplot X5")
```

```
boxplot(tema$X6, col="lavender", main="Boxplot X6")
```

```
boxplot(tema$X7, col="grey", main="Boxplot X7")
```

```
boxplot(tema$X8, col="yellow", main="Boxplot X8")
```

```
boxplot(tema$X9, col="red", main="Boxplot X9")
```

```
boxplot(tema$X10, col="aquamarine", main="Boxplot X10")
```

```
# ANALIZA FACTORIALA
```

```
tema_AF = tema
```

Iordan Maria-Alexandra
Grupa 1080-A

```
View(tema_AF)
```

```
cor(tema_AF[-1])
```

```
# Elimin X8 si X9
```

```
date_AF = cbind(tema_AF[,2:8], tema_AF[,11])
```

```
# Standardizarea datelor
```

```
date_AF_std = scale(date_AF, scale=TRUE)
```

```
View(date_AF_std)
```

```
# Testul KMO
```

```
install.packages("psych")
```

```
library(psych)
```

```
KMO(date_AF_std)
```

```
#pragul minim pentru a merge mai departe cu analiza factoriala este 0.5-0.6
```

#ideal este ca valoarea sa fie cat mai aproape de 1, in cazul nostru este 0.71, asadar
continuuam analiza factoriala

```
# Testul de sfericitate Bartlett
```

```
# H0-variabilele sunt ortogonale(matricea de corelatie este matricea identitate)
```

```
# H1- exista cel putin 1 factor comun care sa-mi explice corelatia dintre variabile
```

Ne uitam la p-value-> daca e < de 0.05 => respingem ipoteza nula si acceptam ipoteza
alternativa

Iordan Maria-Alexandra
Grupa 1080-A

```
R=cor(date_AF_std)
cortest.bartlett(R,n=81,diag=TRUE)
```

```
#respingem H0, acceptam H1: exista cel putin un factor comun
```

```
#Criterii de alegere al factorilor
```

```
#Criteriul Kaiser-retin in analiza comp principale care au suma >=1
```

```
#Criteriul procentului de acoperire-retin un nr de componente principala care impreuna  
contin un procent informational 75-80%
```

```
#Criteriul lui Screeplot
```

```
#Alegerea numarului de factori
```

```
acp = princomp(date_AF_std, cor=T)
acp
```

```
sdev = acp$sdev
valp = sdev*sdev
procent_info = (valp/sum(valp))*100
procent_cumulat=cumsum(procent_info)
```

```
X = data.frame(sdev, valp, procent_info, procent_cumulat)
X
View(X)
```

```
# Criteriul lui Screeplot
```

```
windows()
screeplot = prcomp(date_AF_std)
```


Iordan Maria-Alexandra
Grupa 1080-A

```
plot(screepplot, type = "l", main = "Screeplot", col = "purple")
```

#Pe axa orizontala avem componentele principale, iar pe axa verticala avem varianta/valorile proprii

#Acest screeplot il folosim ca un prim criteriu pt alegerea unui nr potrivit de comp principale

#Prima data ducem o dreapta paralela cu axa verticala din punctul de cotitura (5) -repr punctul in care panta incepe sa se apropie de 0 - si primul intreg din stanga drepteii va repr nr de componente pe care le pastram

#4 factori

Criteriul lui Kaiser - vom retine in analiza componentele principale care au valorile proprii mai mari sau egale cu 1

#Observam ca valp sunt ≥ 1 pt primele 2 componente principale

#2 factori

Criteriului procentului de acoperire - procent cumulat

#ar trebui sa avem un procent de acoperire in jur de 70-75-80%

#ne oprim la comp 2 deoarece are val de 80.28%

#2 comp principale

Pastram in final in analiza 2 componente principale/2 factori

Estimarea modelului factorial

1. Metoda axelor principale

```
install.packages("GPArotation")
```

```
library(GPArotation)
```

```
library(psych)
```

```
factor1 = fa(date_AF_std, nfactors = 2, rotate = "none", fm = "pa")
```

Iordan Maria-Alexandra
Grupa 1080-A

```
print(factor1$loadings, cutoff = 0.4)
```

```
install.packages("corrplot")
```

```
library(corrplot)
```

```
corrplot(factor1$loadings)
```

```
factor1
```

```
# 2. Metoda verosimilitatii maxime
```

```
factor2 = fa(date_AF_std, nfactors = 2, rotate = "none", fm = "ml")
```

```
factor2 = fa(date_AF_std, nfactors = 2, rotate = "varimax", fm = "ml")
```

```
print(factor2$loadings, cutoff = 0.4)
```

```
corrplot(factor2$loadings)
```

```
factor2
```

```
# Diagramele modelelor
```

```
windows()
```

```
fa.diagram(factor1)
```

```
fa.diagram(factor2)
```

```
# Generarea scorurilor factoriale
```

```
scoruri_ml <- factor2$scores
```

```
# Aduagam scorurile la setul de date original
```

Iordan Maria-Alexandra
Grupa 1080-A

```
date_noi <- cbind(date_AF_std, scoruri_ml)

# Redenumim coloanele

colnames(date_noi)[(ncol(date_noi)-1):ncol(date_noi)] <- c("Performanta financiara",
"Performanță operațională și venituri" )

# Salvarea noului set de date ca fisier csv

write.csv(date_noi, "date_ML.csv", row.names=FALSE)


View(date_noi)


# Asociem numele companiilor cu scorurile pentru a le afisa etichetele pe grafic

companii <- tema_AF$Companie

rownames(factor2$scores) <- companii


# Reprezentari grafice


windows()


# Scatterplot-ul observatiilor/Score plot


plot(factor2$scores[, 1], factor2$scores[, 2],
      col = "#696969", pch = 16,
      main = "Score Plot pentru Observații",
      xlab = "Factor 1", ylab = "Factor 2")

text(factor2$scores[,1], factor2$scores[,2], labels=rownames(factor2$scores),
col="purple", cex=0.8, pos=3)


# Loading plot/Graficul incarcaturilor fiecarei variabile pe fiecare factor


plot(factor2$loadings[,1], factor2$loadings[,2],
```

Iordan Maria-Alexandra
Grupa 1080-A

```
xlab = "Factor 1", ylab = "Factor 2",  
main = "Loading Plot pentru Factorii Analizei Factoriale",  
pch = 16, col = "pink")  
  
text(factor2$loadings[,1], factor2$loadings[,2], labels = rownames(factor2$loadings),  
col="black", cex=0.8, pos=3)  
  
# Biplot  
  
plot(factor2$scores[,1], factor2$scores[,2],  
xlab = "Factor 1", ylab = "Factor 2",  
main = "Biplot Analiza Factorială",  
pch = 16, col = "blue")  
  
text(factor2$scores[,1], factor2$scores[,2], labels=rownames(factor2$scores), col="blue",  
cex=0.8, pos=3)  
  
arrows(0, 0, factor2$loadings[,1], factor2$loadings[,2],  
col = "red", length = 0.1, angle = 20)  
  
text(factor2$loadings[,1], factor2$loadings[,2], labels = rownames(factor2$loadings),  
col = "red", cex = 0.8, pos = 3)  
  
# ANALIZA CORESPONDENTELOR  
  
install.packages("FactoMineR")  
library(FactoMineR)  
install.packages("factoextra")  
library(factoextra)  
library(ggplot2)  
install.packages("corrplot")  
library(corrplot)
```

Iordan Maria-Alexandra
Grupa 1080-A

```
date = date_AC  
date_ac = date[,-1]  
date_ac=as.table(as.matrix(date_ac))  
rownames(date_ac)=date$Regiuni  
View(date_ac)
```

Reprezentarea grafica

```
install.packages("gplots")  
library(gplots)  
windows()  
balloonplot(t(date_ac),main="Matricea de contingenta",xlab="",ylab="",  
            label=T,show.margins=F)
```

Testul de independenta χ^2

Ipoteze:

H0: nu exista asocieri intre variabile (liniile si coloanele matricii sunt independente)

H1: exista asocieri intre variabile

```
x2=chisq.test(date_ac)
```

x2

p-value = 0.00000000000000022 < 0.05 => respingem H0, acceptam H1

=> Testul Chi-patrat sugereaza ca exista o asociere semnificativă între variabile,
deoarece p-value este extrem de mic, putem spune ca este foarte puțin probabil ca observațiile sa
fi aparut intamplator si ca exista o legatura semnificativa intre variabilele testate

X-squared = 2259.4

Valoarea mare sugereaza ca exista o asociere semnificativa intre variabilele implicate

Iordan Maria-Alexandra
Grupa 1080-A

df reprezinta gradele de libertate ale testului, si se calculeaza ca $(m-1)*(n-1) = 7*4 = 28$

#Rezultate AC

rez=CA(date_ac,graph=F)

rez

#Valorile proprii

eig=get_eigenvalue(rez)

eig

rez\$eig

#Inertia totala

s=sum(eig[,1])

s

Inertia totala de 0.03690236 sugereaza ca întreaga variabilitate a datelor este de aproximativ 0.0369 (valoarea totală a tuturor valorilor proprii)

Scree-plot

windows()

fviz_screplot(rez)

Conform graficului, putem observa ca prima dimensiune explica mai mult de 60% din variatia totala, adica explica majoritatea variabilitatii in date, fiind urmata apoi de o cadere brusca, ce indica faptul ca urmatoarele componente nu adauga foarte multa variabilitate suplimentara

summary(rez,nb.dec=2)

Indicatorii pt dimensiunea linie

rez\$row

#Coordonatele

rez\$row\$coord

#Contributiile

rez\$row\$contrib

#Calitatea reprezentarii

rez\$row\$cos2

#Inertia

rez\$row\$inertia

#Matricea factor pt randuri ????

cor(date_ac, rez\$row\$coord)

#Indicatorii pentru dimensiunea coloana

rez\$col

#Coordonatele

rez\$col\$coord

Iordan Maria-Alexandra
Grupa 1080-A

```
#Contributiile
```

```
rez$col$contrib
```

```
#Calitatea reprezentarii
```

```
rez$col$cos2
```

```
#Inertia
```

```
rez$col$inertia
```

```
#Matricea factor pentru coloane
```

```
mf_coloane <- rez$col$coord
```

```
print(mf_coloane)
```

```
#Matricea factor pentru linii
```

```
mf_linii <- rez$row$coord
```

```
print(mf_linii)
```

```
# Reprezentari grafice
```

```
windows()
```

```
#Biplot
```

```
fviz_ca_biplot(rez)
```

```
fviz_ca_biplot(rez,map="row principal",arrow=c(T,T),repel=T)
```

```
windows()
```



```
# Contributia coloanelor la cele patru dimensiuni
```

```
fviz_contrib(rez,choice="col",axes=1:4)
```

```
# Contributia randurilor la cele patru dimensiuni
```

```
corrplot(rez$row$contrib,is.corr=F)
```

```
fviz_contrib(rez,choice="row",axes=1:4)
```

```
# Calitatea repr pentru coloane
```

```
fviz_cos2(rez, choice = "col", axes = 1:2)
```

```
# Calitatea repr pentru linii
```

```
fviz_cos2(rez, choice = "row", axes = 1:2)
```

```
#Proiectia coloanelor pe primele 2 dim
```

```
fviz_ca_col(rez, axes=1:2)
```

```
#Proiectia randurilor pe primele 2 dim
```

```
fviz_ca_row(rez, axes=1:2)
```