

Analiza multivariată



Obiectivele cursului

1. Înțelegerea tipurilor de date (simple, multidimensionale).
2. Înțelegerea analizei statistice multivariate.
3. Studiu de caz:

Impactul schimbărilor climatice asupra peisajului - PCA

Metoda de lucru: de la *aspecte teoretice* la *exerciții* utilizând softurile R, Fragstats, ArcMap.

Date folosite în analiza multivariată

A. Tipuri de date

Populația statistică: elemente statistice (exprimate cantitativ sau calitativ) care *au cel puțin o caracteristică comună*. Pentru studii relevante *populația statistică trebuie să fie numeric reprezentativă*.

Variabila – caracteristici **măsurate** sau **observate** (e.g. **colectare teren**) pentru fiecare element din populația statistică. Variabila statistică reprezintă o trăsătură comună tuturor unităților unei populații. *Variabila este un șir de valori sau un vector.*

Variabilele pot fi :

- I.**
1. măsurate (temperatura, ph, umiditate etc)
2. calculate (*metrici* sau / *landscape metrics*)
3. observate (bogăția de specii)

II.

1. categoriale
2. nominale (M, F)
3. ordinale (mic, mare, mediu)

III.

1. omogene
2. heterogene

Se poate verifica această caracteristică făcând media (valoare mică = omogenitate, valoare mare = heterogenitate)

IV.

4.1. cantitative : exprimate pe scara numerică. *Datele cantitative* reprezintă valori numerice , discrete și continue. *Datele discrete*: numărul de indivizi ai unei specii. *Date continue*: date de altitudine.

4.2.calitative : (ce pot fi transformate în valori numerice prin variabile de tipul **dummy/binare** (0/1)

Transformare și dispunerea matricială a datelor calitative

Date binare (*dummy variables*): se pot nota cu 1 cele prezente și cu 0 cele absente Dau de tipul : Da sau NU = **Date calitative**

Management	<i>CPS</i>	<i>ECT</i>	<i>CPL</i>	<i>PMT</i>	<i>UCP</i>
<i>CPS</i>	1	0	0	0	0
<i>ECT</i>	0	1	0	0	0
<i>CPL</i>	0	0	1	0	0
<i>PMT</i>	0	0	0	1	0
<i>UCP</i>	0	0	0	0	1

Conservare (protectie stricta)/CPS ; Ecoturism ECT ; Comercializarea produselor locale CPL/; Practicarea meseriilor tradiționale PMT/; Utilizarea calitatii peisajelor (aspecte estetice: frumusetete, unicitate) pentru atragerea turistilor) UCP

B. Prelucrarea și ordonarea datelor

Distribuția datelor

Datele pot fi diferite (e.g. pe scara numerică o valoare de 0,7 și una de 200 nu se pot analiza d.p.d.v statistic !!!). Se impune **standardizarea** (în R se face automat)

Standardizarea (*media, mediana, abaterea standard, transformarea in valori relative - proportii /procente, logaritmare*)

Media

Abaterea standard - abaterea de la media datelor.

Mediana – 1 1 1 1 -- 2 2 3 4 6

Repartitia gaussiană = reflectă distribuția uniformă a datelor .

Variația sau Varianța?

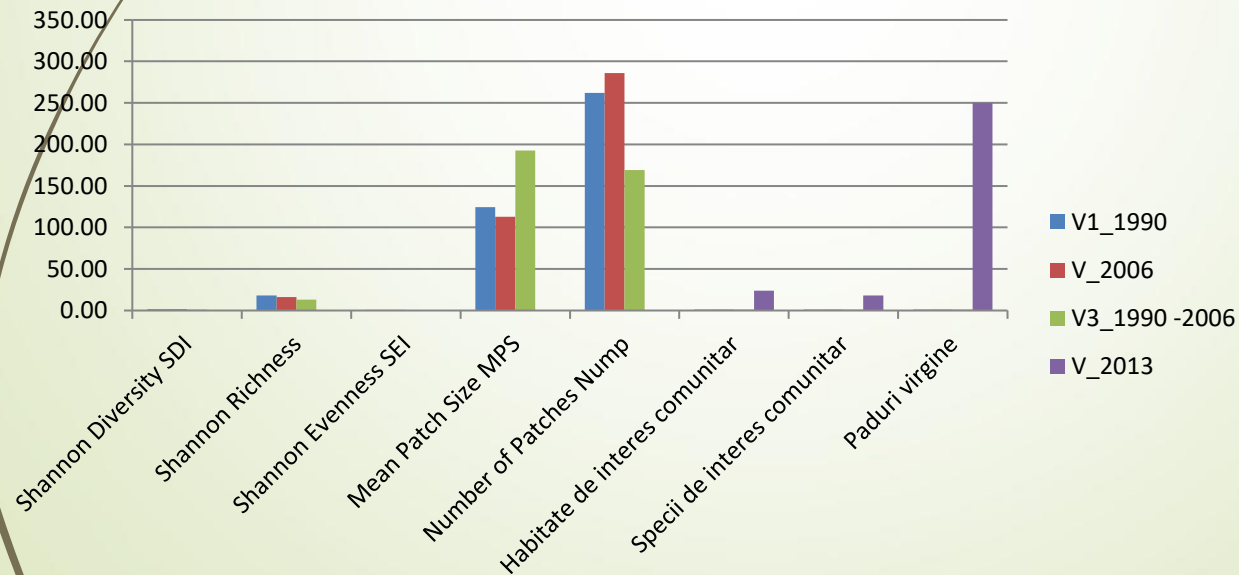
Standardizare sau logaritmare?

Exemplul_1_ Diversitatea peisajelor / LM

Ordonarea și prezentarea datelor (statistică elementară)

Parcul natural Bucegi	V1_1990	V_2006	V3_1990-2006	V_2013
Shannon Diversity SDI	1.64	1.63	1.13 -	0
Shannon Richness	18.00	16.00	13.00	0
Shannon Evenness SEI	0.57	+ 0.59	0.22 -	0
Mean Patch Size MPS	124.34	113.00	192.80	0
Number of Patches Nump	262.00	286.00	169.00 -	0
Habitat de interes comunitar	1	1	1	24
Specii de interes comunitar	1	1	1	18
Paduri virgine	1	1	1	76

1 = prezente



Exemplul_2

Ordonarea și prezentarea datelor (statistică elementară)

Bucegi /Parc natural

Raster/metrici

	LD_90		LD_06	
	V1 S	V2 %	V3 S	V4 %
CLC				
Spatiu urban discontinuu si spatiu rural	10	0.0306	42	0.12
Zone de extractie a minereurilor	122	0.37	140	0.42
Zone in constructie	31	0.0951	0	0
Zone de agrement	25	0.076	0	0
Livezi	31	0.0951	31	0.0951
Pasuni secundare	155	0.475	180	0.55
Zone de culturi complexe	3	0.0092	34	0.104
Terenuri predominant agricole in amestec cu vegetatie natura	28	0.0852	28	0.085
Paduri de foioase	2095	6.42	2097	6.43
Paduri de conifere	11113	34.09	12636	38
Paduri mixte	9060	27.79	7527	23.09
Pajisti naturale	6825	20.93	6807	20.88
Vegetatie subalpina	1741	5.34	1628	4.99
Zone de tranzitie cu arbusti (in general defrisate)	615	1.88	670	2.05
Plaje, dune, renii	15	0.04	15	0.04
Stancarii	311	0.95	446	1.36
Areale cu vegetatie rara	287	0.88	186	0.57
Acumulari de apa	130	0.39	130	0.39

Exemplul_3

Ordonarea și prezentarea datelor în analiza multivariată (e.g.nPCA)

Valori obținute din calcularea landscape metrics (LM)

Măsuratori/ Populație statistică	VARIABLE					NU			
	V1	V2	V3	V4	V5	V6	V7	V8	
1934	136.7939	1.2501	0.6893	34.3511	0.9375	11.9981	59.9791	92.4009	
1954	87.9389	0.9604	0.5642	50.687	2.25	18.5169	67.7289	95.0418	
1968	131.9084	1.119	0.6415	37.5573	1.125	11.2608	75.8134	92.4423	
1985	168.5496	1.2907	0.7092	25.4972	0.9375	10.2238	62.2423	89.4211	
2000	151.4504	1.318	0.7212	21.6794	0.8125	7.1546	62.8356	85.7809	

= standardizare

Măsuratori/ Populație statistică	VARIABLE				DA
	V1	V2	V3	V4	
1934	136.7939	1.2501	0.6893	34.3511	
1954	87.9389	0.9604	0.5642	50.687	
1968	131.9084	1.119	0.6415	37.5573	
1985	168.5496	1.2907	0.7092	25.4972	
2000	151.4504	1.318	0.7212	21.6794	

Minim 4 variabile

C. Analiza multivariată*

Datele sunt în general eterogene și interdependente (*condiționate de alte fenomene sau procese cu care se află în dependență reciprocă*).

În domeniul fenomenelor și proceselor (e.g. ecologice), iau naștere o serie de **legături, de interdependențe**, determinate de acțiunea unor cauze și condiții diferite, care influențează mai mult sau mai puțin fenomenele existente.

Analiza multivariată ne ajută să înțelegem:

- (i) complexitatea fenomenelor
- (ii) relațiile de cauzalitate
- (iii) direcția și gradul de intensitate a legăturilor, ce se manifestă pe scara timpului, sau în dinamica spațială (Buttler, 2013).

* Buttler, A., (2013), Ecologie numérique. site EPFL <http://www.epfl.ch/index.en.html>

În analiza multivariată în funcție de rolul lor în analiză, datele se grupează în:

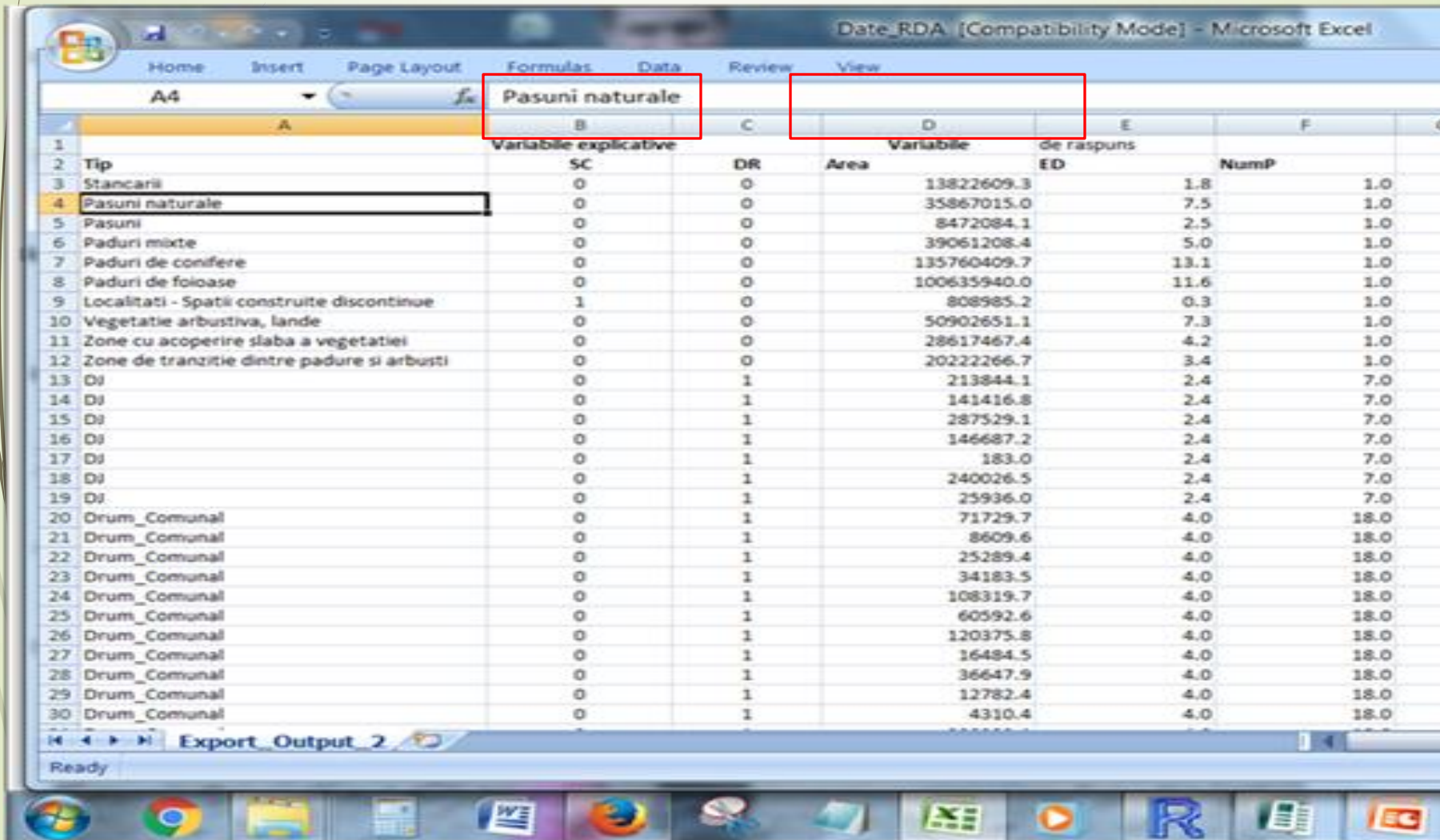
variabile explicative / *explanatory variables* / *variables explicatives*
(e.g. geologia, solul, panta, altitudinea, clima)

variabile de răspuns / *response variables* / *variables réponses*
pentru analizele din landscape ecology / pot fi **patternurile**
peisajul.

În funcție de de tipul lor (*cantitative sau calitative*) se va alege
metoda potrivită de analiză multivariată (e.g. *PCA, RDA, CCA*)

Ordonarea și prezentarea datelor

Variabile explicative și de răspuns



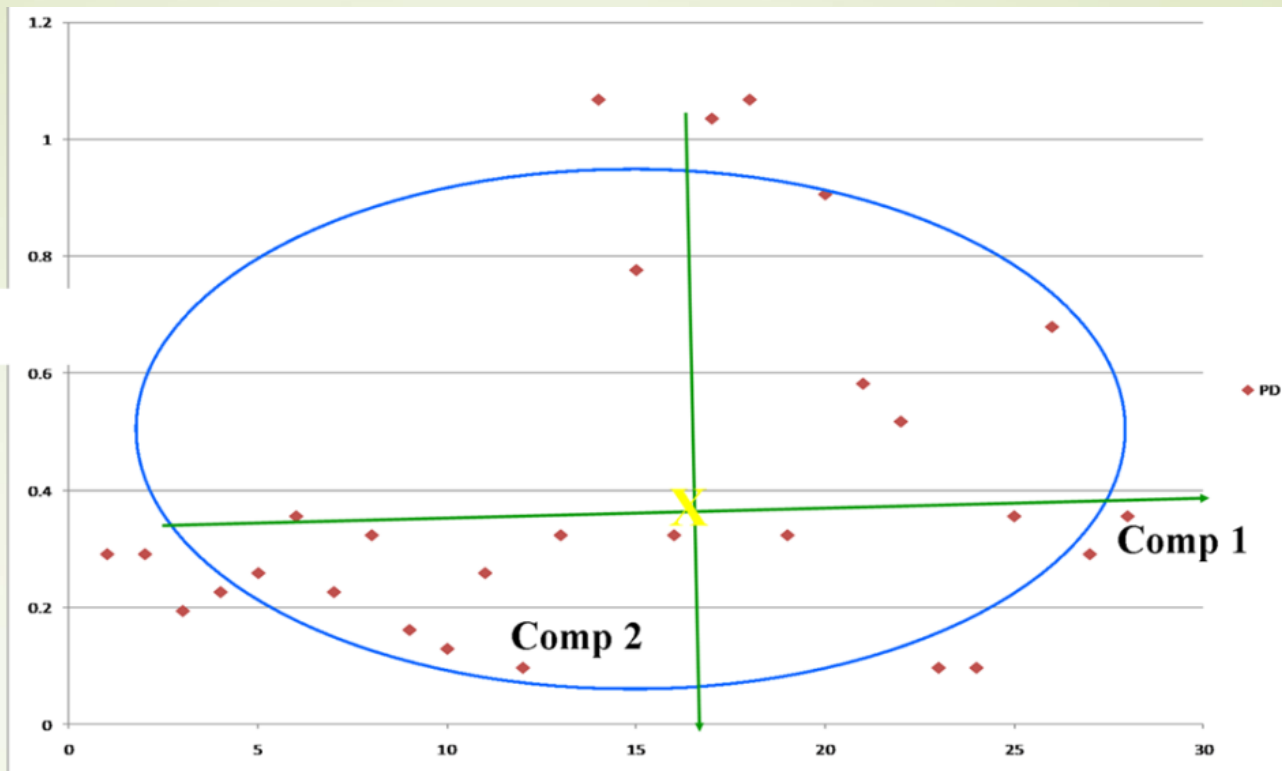
The screenshot shows a Microsoft Excel spreadsheet titled "Date_RDA [Compatibility Mode] - Microsoft Excel". The active cell is A4, containing the text "Pasuni naturale". The spreadsheet is organized into columns: A (Tip), B (Variable explicative), C (DR), D (Variable de raspuns), E (ED), and F (NumP). The data rows are numbered 1 through 30. The row for "Pasuni naturale" is highlighted in yellow. The table below represents the data shown in the spreadsheet.

Tip	Variable explicative	DR	Variable de raspuns	ED	NumP
Stancarii	0	0	13822609.3	1.8	1.0
Pasuni naturale	0	0	35867015.0	7.5	1.0
Pasuni	0	0	8472084.1	2.5	1.0
Paduri mixte	0	0	39061208.4	5.0	1.0
Paduri de conifere	0	0	135760409.7	13.1	1.0
Paduri de foioase	0	0	100635940.0	11.6	1.0
Localitati - Spatii construite discontinue	1	0	808985.2	0.3	1.0
Vegetatie arbustiva, lande	0	0	50902651.1	7.3	1.0
Zone cu acoperire slaba a vegetatiei	0	0	28617467.4	4.2	1.0
Zone de tranzitie dintre padure si arbusti	0	0	20222266.7	3.4	1.0
Dj	0	1	213844.1	2.4	7.0
Dj	0	1	141416.8	2.4	7.0
Dj	0	1	287529.1	2.4	7.0
Dj	0	1	146687.2	2.4	7.0
Dj	0	1	183.0	2.4	7.0
Dj	0	1	240026.5	2.4	7.0
Dj	0	1	25936.0	2.4	7.0
Drum_Comunal	0	1	71729.7	4.0	18.0
Drum_Comunal	0	1	8609.6	4.0	18.0
Drum_Comunal	0	1	25289.4	4.0	18.0
Drum_Comunal	0	1	34183.5	4.0	18.0
Drum_Comunal	0	1	108319.7	4.0	18.0
Drum_Comunal	0	1	60592.6	4.0	18.0
Drum_Comunal	0	1	120375.8	4.0	18.0
Drum_Comunal	0	1	16484.5	4.0	18.0
Drum_Comunal	0	1	36647.9	4.0	18.0
Drum_Comunal	0	1	12782.4	4.0	18.0
Drum_Comunal	0	1	4310.4	4.0	18.0

D. Analiza Componentelor Principale / Principal Component Analysis (PCA) /Analyse en Composantes Principales (ACP)

1. Modelul PCA ne arată cum variază datele.
1. PCA evidențiază relațiile semnificative de interdependență dintre variabile ce **NU** pot fi surprinse prin simpla examinare a datelor.
2. Scopul acestei analize este de a simplifica/reduce complexitatea datelor, prin identificarea unor variabile/ caracteristici ce definesc fenomenul/procesul analizat.
3. Prin această selecție se pot pierde o parte din valoarea datelor brute, dar rămân informații despre *variabilitatea datelor*.
4. PCA se face doar cu date cantitative **și nu calitative** chiar dacă ele sunt transformate în cantitative (de tipul variabile binare /0/1)
5. Dacă valorile variabilelor sunt prea diferite se recomandă standardizarea/logaritmare.

Standardizare?



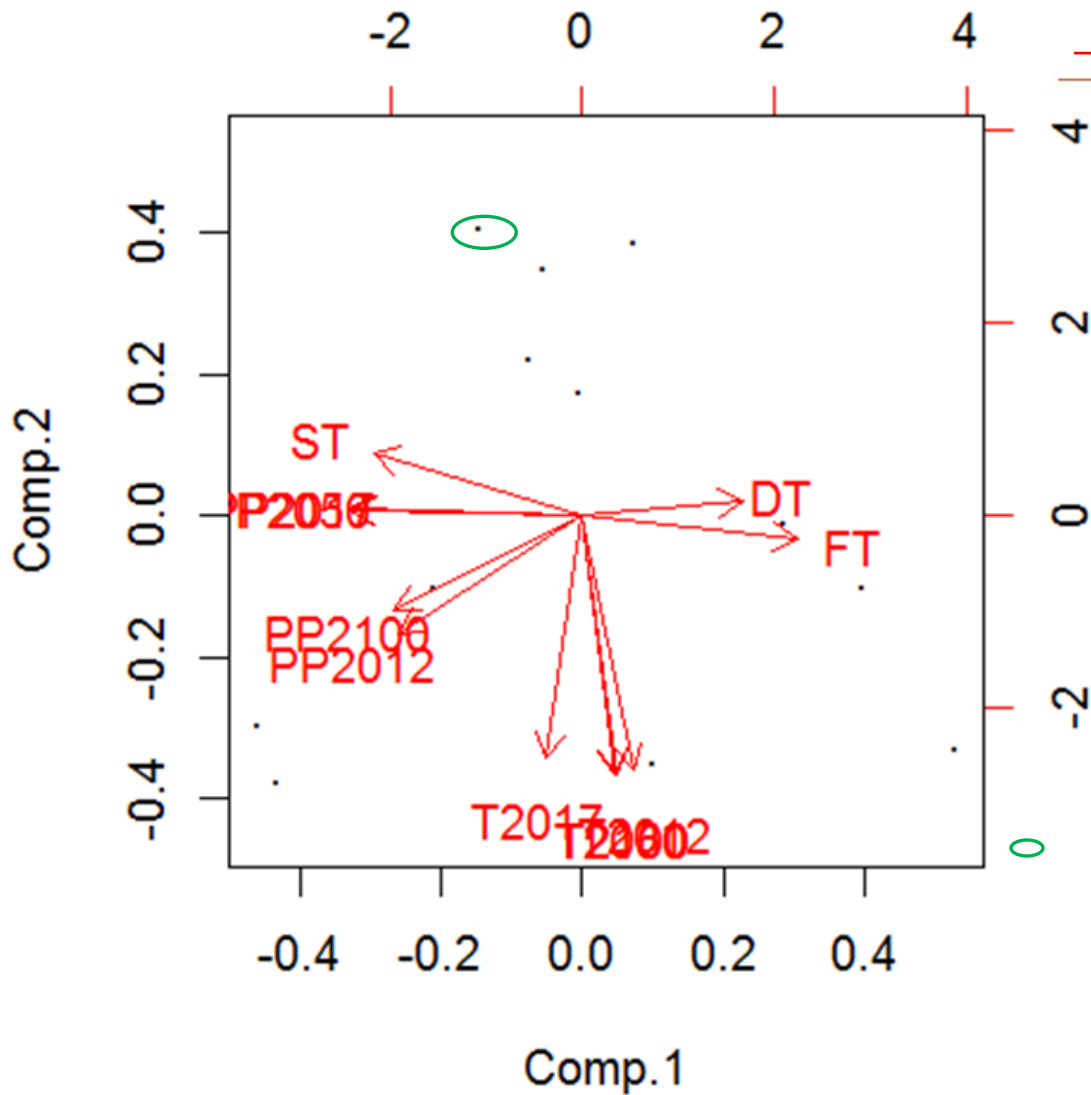
Species Name	DT	ST	FT	T2012	T2017	T2050	T2100	PP201	50	PP2100		
<i>Abies alba</i>		0.23	5	3	-6.8	1.93891	1.60623	1.76427	99.44	60.08079	73.84448	54.29743
<i>Acer campestre</i>		0.25	3	2	-8.6	2.49842	2.72505	3.47093	58.86	51.01936	58.03812	55.14544
<i>Acer pseudoplatanus</i>		0.17	4	2	-0.9	6.07482	6.93351	6.62562	17.15	65.97978	56.53873	64.66414
<i>Carpinus betulus</i>		0.25	4	2	5	11.3651	12.2217	11.7881	152.1	64.3497	60.7608	73.1064
<i>Fagus sylvatica</i>		0.25	5	1	9.6	16.3321	16.9082	16.5533	187.14	72.20303	70.3142	72.5152
<i>Fraxinus excelsior</i>		0.08	3	2	14.6	10.7605	21.49	21.3366	148.42	75.3171	79.1664	83.6964
<i>Picea abies</i>		0.15	3	3	18	10.8823	24.0682	23.5304	59.03	63.14576	55.46675	58.37207
<i>Pinus sylvestris</i>		0.37	1	4	15	11.3064	22.2128	21.3841	71.11	41.19125	47.62871	57.01272
<i>Quercus petraea</i>		0.247	2	3	12	9	16.7821	16.5013	49.84	47.577	49.3188	36.8436
<i>Quercus pubescens</i>		0.41	2	3	8	9.1905	9.99429	11.0258	56.14	52.65815	53.51065	62.26195
<i>Quercus robur</i>		0.165	2	3	3	4.68168	5.90871	5.34384	41.69	67.6623	63.2112	69.1989
<i>Tilia platyphyllos</i>		0.25	3	2	-3.9	2.35812	2.68634	4	68.34	73.69692	58.9279	52.5

Advances in Landscape Ecology, Curs_4/ Ileana Stupariu, 2020-2021

Date ANM

Date specii după:

După Gillet, 2008, Peringer et al. 2013



- Axa 1, 2 /componenta 1 și componenta 2
- Măsuratori /populație statistică/ variabile
- Vector/descriptor pentru variabile/ (1, 1, 1, 1, 1, 1)

> summary(arc.pca1)

Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Standard deviation	2.0520523	2.0065174	1.0615384	0.78187335	0.71270222
Proportion of Variance	0.3828108	0.3660102	0.1024422	0.05557508	0.04617677
Cumulative Proportion	0.3828108	0.7488210	0.8512631	0.90683822	0.95301499

R Console

Loadings :

	Comp.1	Comp.2	Comp.3	Comp.4
DT	0.292		0.691	-0.296
ST	-0.382	0.115	0.268	0.413
FT	0.395		-0.231	-0.395
T2012		-0.477	-0.141	
T2017		-0.454	0.297	
T2050		-0.485	-0.103	0.142
T2100		-0.484		0.144
PP2012	-0.335	-0.220	0.430	-0.139
PP2017	-0.424		-0.267	
PP2050	-0.428		-0.100	-0.310
PP2100	-0.344	-0.176		-0.646

Bibliografie

- Buttler, A., (2013), *Ecologie numérique*. site EPFL <http://www.epfl.ch/index.en.html>
- Năstase I.I., Pătru-Stupariu I., Kienast F., (2019) *Landscape Preferences and Distance Decay Analysis for mapping the Recreational Potential of an Urban Area*. Sustainability, 11, 3620; doi:10.3390/su11133620
- Pătru-Stupariu I., Stupariu M.S., Stoicescu I., Peringer A., Buttler A., Fürst Ch .(2017) *Integrating geo-biodiversity features in the analysis of landscape patterns*, Ecological Indicators 80: 363–375

Dacă utilizezi informații, scripturi (R), idei, date, metode, interpretarea metricilor, imagini, din acest curs, citează-l ca resursă online
<https://unibuc.ro/user/ileana.stupariu/>

