

Analisi dei dati per il marketing (compito 20 maggio 2025)

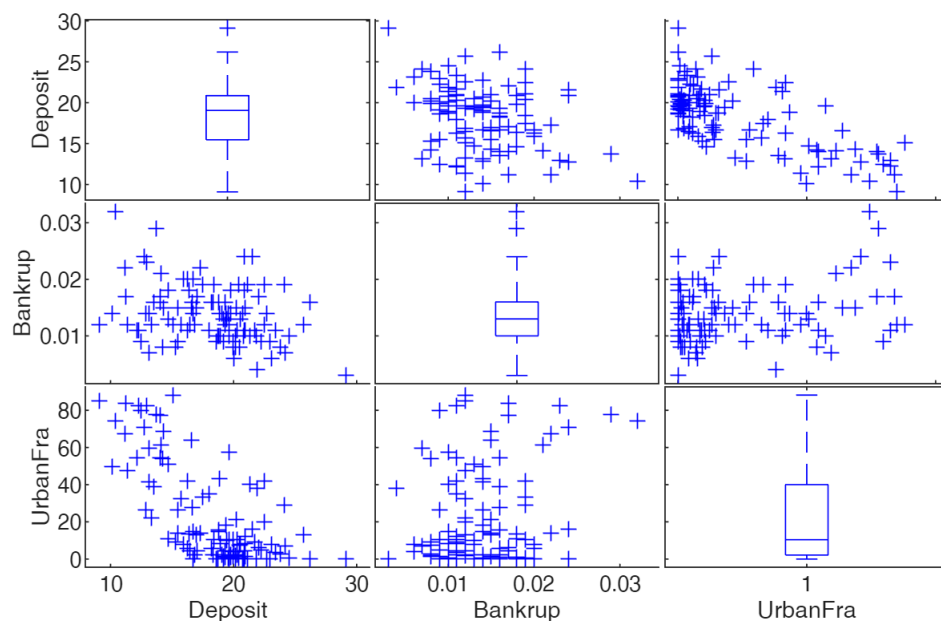
Scrivere nel testo del compito il nome GitHub e l'eventuale partecipazione al seminario

Caricare in memoria la table citiesItaly2024 tramite l'istruzione

```
load citiesItaly2024.mat
```

Costruire la matrice dei diagrammi di dispersione per le prime 3 variabili inserendo i boxplot sulla diagonale principale (punti 6)

```
X=citiesItaly2024;  
spmplot(X(:,1:3),'dispopt','box');
```



Commentare la relazione tra UrbanFra (indicatore di fragilità urbana) e Deposit (indicatore dei depositi), (punti 3)

```
% Relazione decrescente. All'aumentare dei valore dei depositi generalmente  
% l'indicatore di fragilità urbana dimiuisce
```

Calcolare il coefficiente di cograduazione e la sua significatività tra le variabili Paym30D (pagamenti entro 30 giorni) e ElecPar (partecipazione elettorale) (punti 5).

```
[R,Pval]=corr([X.Paym30D X.ElecPar], 'Type', 'Spearman');
```

Mostare il valore del P-value del coeff. di cograduazione nella Command Window (punti 3)

```
disp(['p-value assenza di cograduazione tra pagamenti a 30 giorni e  
partecipazione elettorale =' num2str(Pval(1,2))])
```

p-value assenza di cograduazione tra pagamenti a 30 giorni e partecipazione elettorale =3.2915e-16

Commentare il p-value ottenuto (punti 3)

```
% Rifiuto decisamente l'ipotesi nulla di assenza di cograduazione tra  
% partecipazione elettorale e pagamenti a 30 giorni.  
% In altri termini, le province che presentano elevata partecipazione  
% elettorale sono anche quelle dove i pagamenti a 30 giorni sono molto  
% frequenti.
```

Calcolare la prima componente principale (punti 3).

```
out=pcaFS(X);
```

The first PC already explains more than 0.95^v variability
In what follows we still extract the first 2 PCs
Initial correlation matrix

	Deposit	Bankrup	UrbanFra	Paym30D	ElecPar	QualLif	Protest	SalaryA
Deposit	1.00	-0.33	-0.70	0.72	0.58	0.76	-0.14	0.74
Bankrup	-0.33	1.00	0.26	-0.39	-0.13	-0.44	0.33	-0.18
UrbanFra	-0.70	0.26	1.00	-0.77	-0.74	-0.74	0.23	-0.74
Paym30D	0.72	-0.39	-0.77	1.00	0.69	0.81	-0.35	0.74
ElecPar	0.58	-0.13	-0.74	0.69	1.00	0.58	-0.16	0.64
QualLif	0.76	-0.44	-0.74	0.81	0.58	1.00	-0.37	0.70
Protest	-0.14	0.33	0.23	-0.35	-0.16	-0.37	1.00	-0.14
SalaryA	0.74	-0.18	-0.74	0.74	0.64	0.70	-0.14	1.00
SpendingA	0.75	-0.22	-0.81	0.79	0.78	0.77	-0.31	0.85
Employment	0.79	-0.32	-0.90	0.80	0.73	0.83	-0.31	0.80
AddedVa	0.79	-0.24	-0.70	0.63	0.55	0.72	-0.11	0.86
LowISEE	-0.76	0.40	0.83	-0.82	-0.62	-0.86	0.39	-0.66

Explained variance by PCs

	Eigenvalues	Explained_Variance	Explained_Variance_cum
PC 1	8.02	66.82	66.82
PC 2	1.29	10.77	77.60
PC 3	0.77	6.45	84.05
PC 4	0.52	4.37	88.42
PC 5	0.36	3.00	91.42
PC 6	0.27	2.24	93.66
PC 7	0.24	1.98	95.64
PC 8	0.17	1.44	97.08
PC 9	0.13	1.04	98.12
PC10	0.10	0.82	98.94
PC11	0.09	0.73	99.66
PC12	0.04	0.34	100.00

Loadings = correlations between variables and PCs

PC1 PC2

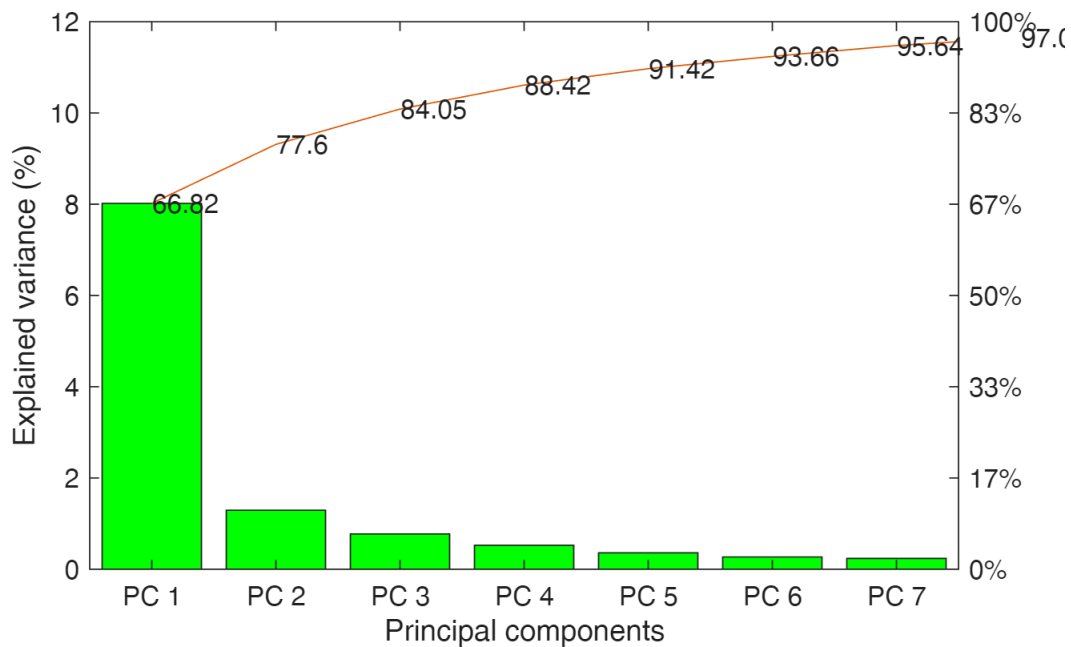
	-----	-----
Deposit	-0.86	-0.10
Bankrup	0.38	-0.69
UrbanFra	0.89	0.10
Paym30D	-0.89	0.11
ElecPar	-0.77	-0.22
QualLif	-0.89	0.18
Protest	0.34	-0.73
SalaryA	-0.86	-0.27
SpendingA	-0.93	-0.12
Employm	-0.96	-0.04
AddedVa	-0.84	-0.24
LowISEE	0.90	-0.17

Communalities

	PC1	PC2	PC1-PC2
	-----	-----	-----
Deposit	0.73	0.01	0.74
Bankrup	0.15	0.47	0.62
UrbanFra	0.80	0.01	0.81
Paym30D	0.79	0.01	0.80
ElecPar	0.59	0.05	0.64
QualLif	0.80	0.03	0.83
Protest	0.11	0.53	0.65
SalaryA	0.74	0.07	0.82
SpendingA	0.86	0.01	0.88
Employm	0.91	0.00	0.92
AddedVa	0.71	0.06	0.76
LowISEE	0.82	0.03	0.85

Units with the 5 largest values of (combined) score and orthogonal distance

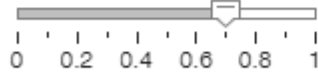
15 21 89 58 60



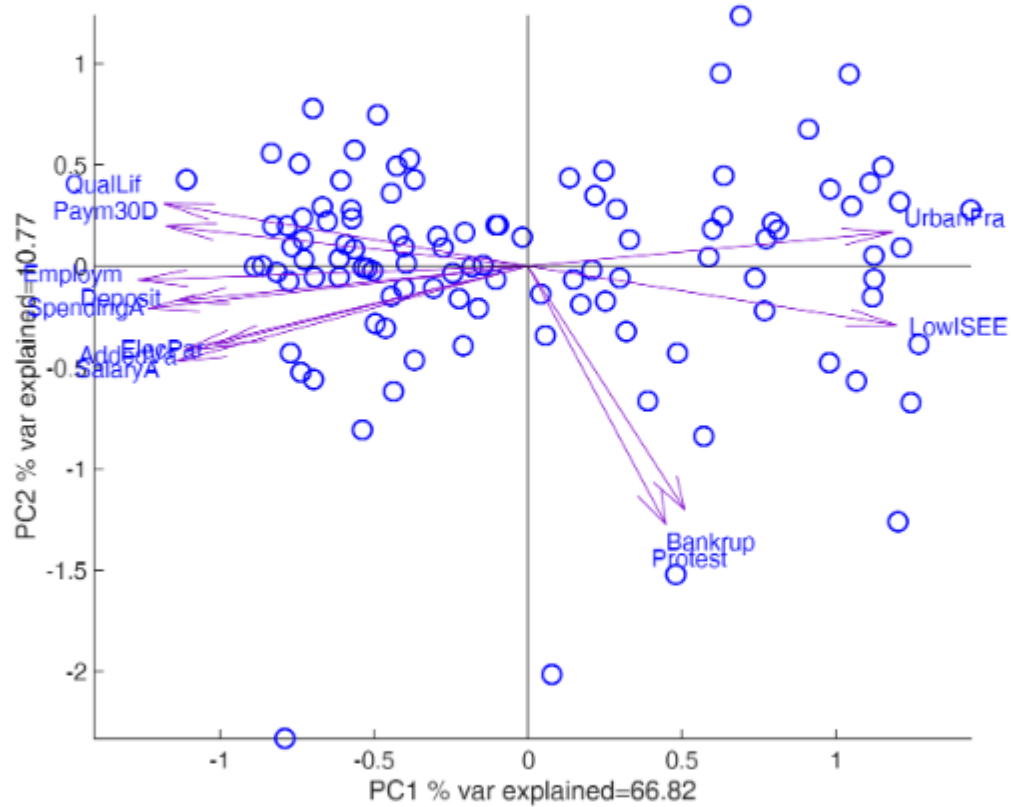
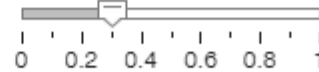
Row points : $(\sqrt{n-1})^\omega U \Gamma^\alpha$

Arrow coordinates : $(\sqrt{n-1})^{1-\omega} V \Gamma^{1-\alpha}$

ω



α



- | | | | |
|--|---|--|---|
| ☒ showRowPoints | ☐ Hide SVD sliders | ☐ Robust analysis | ☐ Brush units |
| ☐ color prop to orthogonal distance | ☐ showRowNames | ☒ showArrows | ☐ Show change sign panel |

Mostrare nella Command Window la varianza spiegata dalla prima PC (punti 3)

```
disp(['La prima PC spiega il ' num2str(out.explainedT.Explained_Variance(1))  
' della varianza totale'])
```

La prima PC spiega il 66.825 della varianza totale

Commentare il significato della prima componente principale (punti 3).

```
% La prima PC è collegata in maniera diretta agli indicatori di disagio ed  
% in maniera inversa agli indicatori di ricchezza.
```

% La prima PC è quindi un indicatore di povertà

In base alla prima PC, mostrare nella Command Window la provincia dove si vive meglio (punti 4)

% La provincia dove si vive meglio è quella con il valore più basso della
% prima PC

```
[~,posmin]=min(out.scoreT{:,1});  
bestProv=X.Properties.RowNames(posmin);
```

```
disp(['La provincia dove si vive meglio è: ' bestProv{:}])
```

La provincia dove si vive meglio è: Bolzano

Per ulteriori informazioni su questo dataset si veda il video youtube



<https://youtu.be/yx569ZBGd58>

