

Note: FSDA version 8.7.10.5 or later must be installed

```
% Il codice chiaramente riconducibile a una generazione automatica tramite  
% strumenti di intelligenza artificiale (ad esempio soluzioni prodotte  
% integralmente da modelli linguistici) comporterà una penalizzazione nella  
% valutazione dell'elaborato. Gli studenti sono tenuti a dimostrare una  
% comprensione personale dei metodi utilizzati e a produrre codice che  
% rifletta il proprio ragionamento, la propria struttura e il proprio stile  
% di commento.
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%  
%%% DURATA: 75 minuti
```

Il file Matlab (script), formato m oppure mlx,
va salvato con il vostro nome e cognome (senza spazi e accenti)
e va caricato nella pagina che viene comunicata.
La votazione finale terrà conto della qualità del codice e della
sua chiara presentazione.
Indicare il nickname GitHub
Indicare (se presente) la partecipazione al seminario
Indicare (se presente) la segnalazione di refusi nel libro di testo

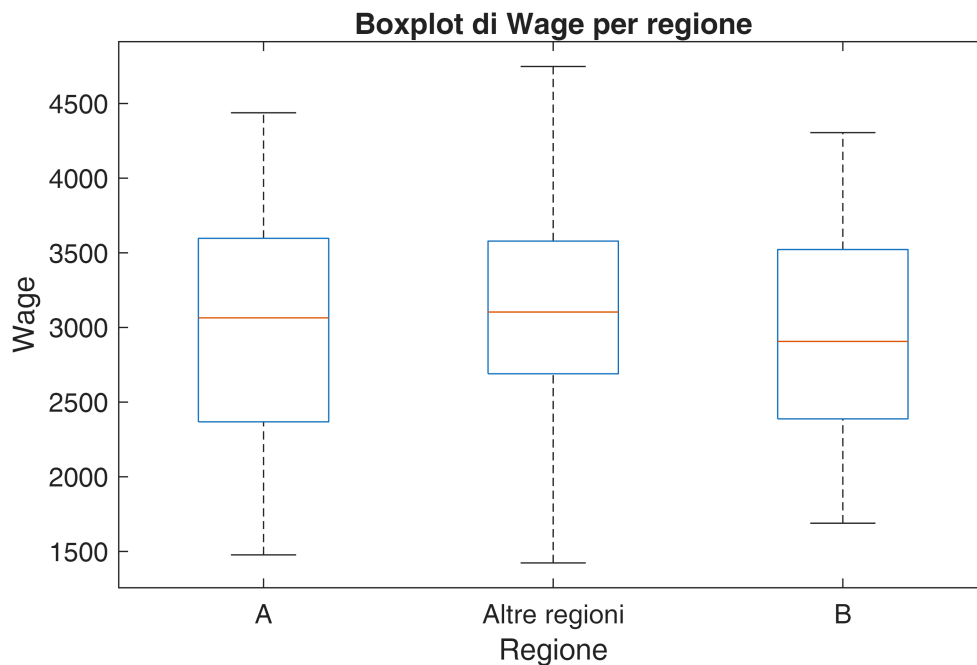
```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

ESERCIZIO I

Caricare in memoria il contenuto del file adm.mat Nella tabella dipendenti le prime 30 righe si riferiscono ai residenti nella regione A e le ultime trenta righe ai residenti nella regione B. Le altre righe si riferiscono ai residenti in altre regioni. Costruire il boxplot per la variabile Wage suddiviso per regione di appartenenza (punti 6)

```
load adm.mat  
  
% Creo la variabile classificatoria  
n=height(dipendenti);  
Regione=strings(n,1);  
Regione(1:n)="Altre regioni";  
Regione(1:30) = "A"; % Assign "A" for the first 30 rows  
Regione(n-29:n) = "B"; % Assign "B" for the last 30 rows  
  
% Costruzione del boxplot per la variabile Wage suddiviso per regione  
boxplot(dipendenti.Wage, Regione);
```

```
title('Boxplot di Wage per regione');
xlabel('Regione');
ylabel('Wage');
```



```
% Commentare il grafico che si ottiene (punti 4)
% La distribuzione dei salari è omogenea
% tra le diverse regioni.
% Il primo quartile del salario delle "Altre regioni" è più alto di quello
% delle altre due regioni
```

ESERCIZIO II

Caricare il dataset `smoke.mat`. Costruire la tabella di contingenza tra le due variabili "profession" e "smoke_frequency". La variabile `profession` è qualitativa ed indica il tipo di professione. La variabile `smoke_frequency` è qualitativa e si riferisce all'attitudine verso il fumo. Denominare la tabella di contingenza con le prime 4 lettere del proprio cognome (senza accenti) (punti 2)

```
load smoke
% Costruzione tabella di contingenza
RIAN=pivot(smoke,"Rows","profession","Columns","smoke_frequency", ...
    "RowLabelPlacement","rownames");
```

Calcolare l'indice Chi2, tra queste due variabili, l'indice H e l'indice tauyx (punti 4).

Commentare il valore dell'indice H (punti 4)

Costruire il grafico che mostra il contributo di ogni singola cella della tabella di contingenza al valore dell'indice Chi2 (punti 6)

% v. pagina 385

```
out=corrNominal(RIAN,'plots',1);
```

Chi2 index
16.4416

pvalue Chi2 index
0.1718

Phi index
0.2919

Cramer's V
0.1685

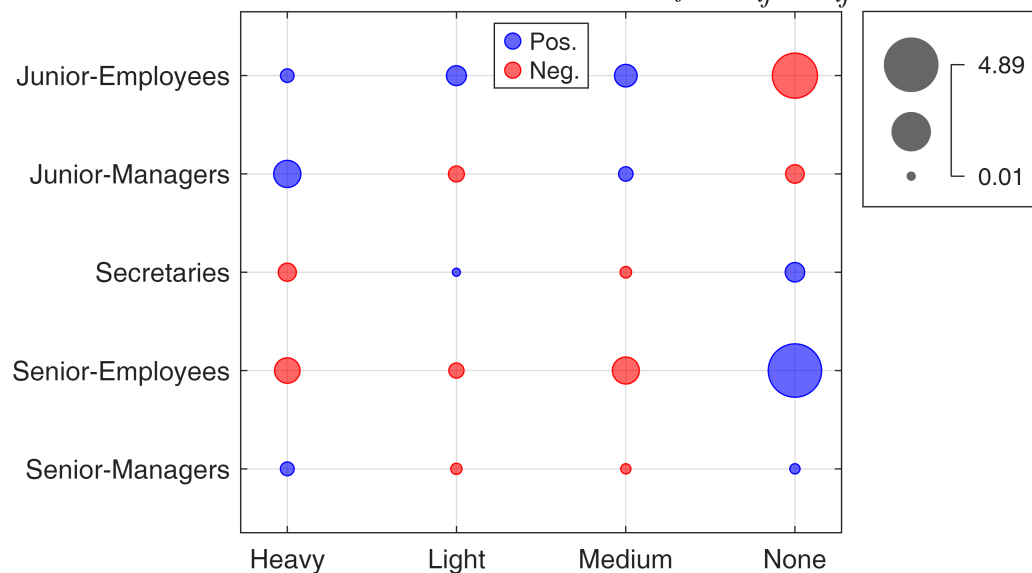
Test of H_0: independence between rows and columns

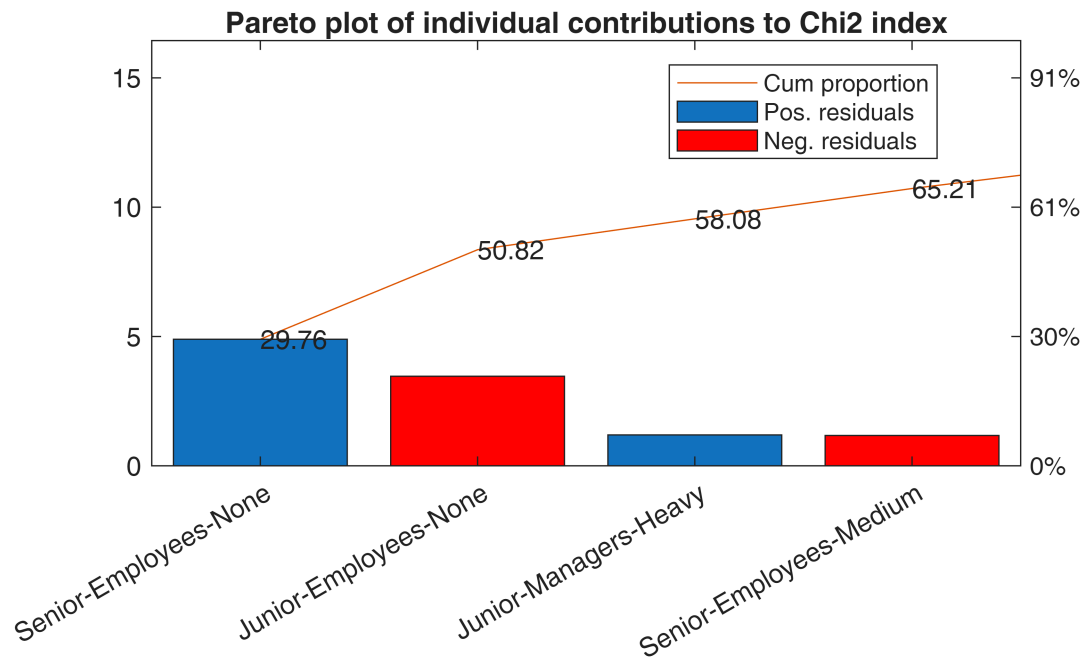
	Coeff	se	zscore	pval
CramerV	0.16851	0.085978	1.96	0.05
GKlambdayx	0.12977	0.055617	2.3333	0.019633
tauyx	0.032185	0.016286	1.9763	0.048124
Hyx	0.03177	0.015492	2.0507	0.040295

Indexes and 95% confidence limits

	Value	StandardError	ConflimL	ConflimU
CramerV	0.16851	0.085978	0	0.19759
GKlambdayx	0.12977	0.055617	0.020763	0.23878
tauyx	0.032185	0.016286	0.00026561	0.064104
Hyx	0.03177	0.015492	0.0014059	0.062134

Pearson residuals² : $(\pm)(n_{ij} - n_{ij}^*)^2/n_{ij}^*$





% Calcolo dell'indice Chi2, H e tauyx

```
chi2Stat = out.Chi2;
```

```
Hyx = out.Hyx;
```

```
tauyx = out.tauyx;
```

% Commentare il valore di Hyx

% L'intervallo di confidenza di H non contiene

% il valore 0. Il valore è significativo anche se molto basso.

% Per il commento v. p. 384

% La conoscenza della professione lavorativa riduce del 3.18% circa l'eterogeneità

% della previsione dell'attitudine verso il fumo.

% Le due celle che contribuiscono di più al valore dell'indice Chi2 sono

% relative a Senior-Employees-None (il colore è blu di conseguenza questa

% combinazione nel campione osservato è molto più frequente di quella

% attesa nell'ipotesi di indipendenza tra le due variabili) e

% Junior-Managers-None (il colore è rosso di conseguenza questa

% combinazione nel campione osservato è molto meno frequente di quella

% attesa). Queste due celle contribuiscono per un ammontare oltre il 50 per

% cento del valore dell'indice Chi2

Effettuare un'analisi delle corrispondenze tra le due variabili di cui sopra e commentare il grafico. (punti 4)

```
CA=CorAna(RIAN);
```

French symmetrical model: rows and cols in principal coordinates.

Plot of $X = D_r^{-1/2}U\Gamma$ and $Y = D_c^{-1/2}V\Gamma$



Summary

	Singular_value	Inertia	Accounted_for	Cumulative
dim_1	0.27342	0.074759	0.87756	0.87756
dim_2	0.10009	0.010017	0.11759	0.99515
dim_3	0.020337	0.00041357	0.0048547	1

ROW POINTS

Results for dimension: 1

	Scores	CntrbPnt2In	CntrbDim2In
Junior-Employees	-0.23295	0.33097	0.94193
Junior-Managers	-0.25896	0.083659	0.5264
Secretaries	0.20109	0.070064	0.86535
Senior-Employees	0.38059	0.51201	0.99903
Senior-Managers	0.065768	0.0032977	0.092232

Results for dimension: 2

	Scores	CntrbPnt2In	CntrbDim2In
Junior-Employees	0.057744	0.15177	0.057876
Junior-Managers	-0.2433	0.55115	0.46468
Secretaries	0.078911	0.080522	0.13326
Senior-Employees	-0.01066	0.0029976	0.00078372
Senior-Managers	-0.19374	0.21356	0.80034

COLUMN POINTS

Results for dimension: 1

	Scores	CntrbPnt2In	CntrbDim2In
Heavy	-0.29378	0.14954	0.6844
Light	-0.099456	0.03085	0.32673
Medium	-0.19632	0.16562	0.98185
None	0.39331	0.654	0.99402

Results for dimension: 2

	<u>Scores</u>	<u>CntrbPnt2In</u>	<u>CntrbDim2In</u>
Heavy	-0.19777	0.50575	0.31015
Light	0.14106	0.46317	0.65729
Medium	0.0073591	0.0017368	0.0013796
None	-0.030492	0.029336	0.0059745

Overview ROW POINTS

	<u>Mass</u>	<u>Score_1</u>	<u>Score_2</u>	<u>Inertia</u>	<u>CntrbPnt2In_1</u>	<u>CntrbPnt2In_2</u>
Junior-Employees	0.45596	-0.23295	0.057744	0.026269	0.33097	0.15177
Junior-Managers	0.093264	-0.25896	-0.2433	0.011881	0.083659	0.55115
Secretaries	0.12953	0.20109	0.078911	0.006053	0.070064	0.080522
Senior-Employees	0.26425	0.38059	-0.01066	0.038314	0.51201	0.0029976
Senior-Managers	0.056995	0.065768	-0.19374	0.0026729	0.0032977	0.21356

Overview COLUMN POINTS

	<u>Mass</u>	<u>Score_1</u>	<u>Score_2</u>	<u>Inertia</u>	<u>CntrbPnt2In_1</u>	<u>CntrbPnt2In_2</u>	<u>CntrbDim2In</u>
Heavy	0.12953	-0.29378	-0.19777	0.016335	0.14954	0.50575	0.68
Light	0.23316	-0.099456	0.14106	0.0070588	0.03085	0.46317	0.326
Medium	0.32124	-0.19632	0.0073591	0.01261	0.16562	0.0017368	0.981
None	0.31606	0.39331	-0.030492	0.049186	0.654	0.029336	0.994

Legend

Row scores in principal coordinates

Column scores in principal coordinates

CntrbPnt2In = relative contribution of points to explain total Inertia of the latent dimension
The sum of the numbers in a column is equal to 1

CntrbDim2In = relative contribution of latent dimension to explain total Inertia of a point
CntrbDim2In_1+CntrbDim2In_2+...+CntrbDim2In_K=1

% Come sottolineato dal precedente grafico Senior-Employees è vicino a
% None e Junior-Managers è vicino a Heavy e queste due coppie si trovano
% agli estremi opposti dell'asse x (che si può interpretare come la
% riluttanza al fumo). Tanto più ci si sposta da sx verso dx l'attitudine
% al fumo diminuisce.