

# Analisi dei dati per il marketing (compito 23 gennaio 2025)

Caricare in memoria la table SportHealth l'istruzione

```
load SportHealth.mat
```

Calcolare manualmente (ossia senza passare attraverso l'output di CorAna) la frequenza teorica, nell'ipotesi di indipendenza tra le due variabili della tabella, della coppia di modalità Sometimes e Poor Health.

Mostrare l'output in formato table come indicato dall'anteprima output di seguito

```
table
```

	Poor Health
Sometimes	14.343

(punti 5)

```
N=SportHealth;  
n=sum(N{:, :}, 'all');  
freqTeorica=sum(N{"Sometimes", :})*sum(N{:, "Poor Health"})/n;  
Nteorica=N('Sometimes', 'Poor Health'); % inizializzazione  
Nteorica{:, :}=freqTeorica;  
disp(Nteorica)
```

	Poor Health
Sometimes	14.343

Confrontare tramite assert il risultato ottenuto manualmente con quello estratto dall'output della funzione CorAna (punti 5)

```
out=CorAna(N, 'plots', 0, 'dispresults', 0);  
% v. le prime due righe di pagina 518  
% Estraggo dalla table delle frequenze teoriche la combinazione riga  
% colonna richiesta  
NteoricaDaCorAna=out.Nhatable('Sometimes', 'Poor Health');  
disp(NteoricaDaCorAna)
```

	Poor Health
Sometimes	14.343

```
% Controllo la differenza tra le due implementazioni sia zero
```

```
assert(abs(Nteorica{:,:}-NteoricaDaCorAna{:,:})==0,'Frequenza calcolata in
maniera errata')
```

Commentare la differenza tra la frequenza effettiva e la frequenza teorica della modalità Sometimes e Poor Health (**punti 3**)

```
% La frequenza teorica è superiore al doppio di quella effettiva (pari a 6)
ed indica
% che le persone che hanno dichiarato di essere in cattiva salute e di fare
% attività sportiva salutare sono molto inferiori a quelle che ci si
% attende nell'ipotesi di indipendenza tra i due caratteri. L'indice Chi2 è
% altamente significativo
```

Caricare in memoria il dataset citiesItaly tramite l'istruzione

```
load citiesItaly.mat
```

Calcolare i quantili 0.05, 0.15, 0.25, 0.35, ..., 0.95 per tutte le variabili del dataset

Mostrare il risultato in una table denominata con le prime 3 lettere del cognome come indicato nell'immagine che segue (**punti 8**)

```
disp(RIA)
```

	<u>addedval</u>	<u>depos</u>	<u>pensions</u>	<u>unemploy</u>	<u>export</u>	<u>bankrup</u>	<u>billsover</u>
<b>Q0.05</b>	10413	4121.6	8059.1	2.599	1.592	14.931	13.667
<b>Q0.15</b>	12322	4591	8634.5	3.3565	5.6555	17.736	24.74
<b>Q0.25</b>	13711	5311	9226.6	4.3225	10.685	20.858	29.3
<b>Q0.35</b>	15782	6284.9	9650.1	5.1055	14.655	25.867	33.29
<b>Q0.45</b>	17559	7423.3	9857.1	5.774	19.215	28.025	37.46
<b>Q0.55</b>	19302	8340.4	10148	7.152	24.512	30.843	42.20
<b>Q0.65</b>	20665	8885	10509	10.135	27.953	33.787	48.59
<b>Q0.75</b>	21968	9613.1	10774	15.635	32.53	36.89	55.58
<b>Q0.85</b>	23206	10183	11098	18.864	39.176	40.498	66.75
<b>Q0.95</b>	25737	11639	12217	26.865	54.307	51.965	92.80

```
quant=0.05:0.1:0.95;
QQ=quantile(citiesItaly{:,:},quant);
nomi="Q"+quant;
RIA=array2table(QQ,"RowNames",nomi,"VariableNames",citiesItaly.Properties.VariableNames);
disp(RIA)
```

	<u>addedval</u>	<u>depos</u>	<u>pensions</u>	<u>unemploy</u>	<u>export</u>	<u>bankrup</u>	<u>billsoverd</u>
<b>Q0.05</b>	10413	4121.6	8059.1	2.599	1.592	14.931	13.667

Q0.15	12322	4591	8634.5	3.3565	5.6555	17.736	24.746
Q0.25	13711	5311	9226.6	4.3225	10.685	20.858	29.37
Q0.35	15782	6284.9	9650.1	5.1055	14.655	25.867	33.294
Q0.45	17559	7423.3	9857.1	5.774	19.215	28.025	37.463
Q0.55	19302	8340.4	10148	7.152	24.512	30.843	42.207
Q0.65	20665	8885	10509	10.135	27.953	33.787	48.594
Q0.75	21968	9613.1	10774	15.635	32.53	36.89	55.587
Q0.85	23206	10183	11098	18.864	39.176	40.498	66.757
Q0.95	25737	11639	12217	26.865	54.307	51.965	92.808

Mostrare il terzo quartile della variabile unemploy (tasso di disoccupazione) nella Command Window (Oss. l'output deve essere in formato table) **(punti 2)**

```
disp(RIA("Q0.75", "unemploy"))
```

```

unemploy
-----
Q0.75    15.635

```

Interpretare il terzo quartile per la variabile unemploy **(punti 3)**

% Il 75 per cento delle province italiane presenta un valore del tasso di  
% disoccupazione inferiore a 15.635 ed il rimanente 25 per cento un valore  
% superiore

Calcolare la mediana, la media, il MAD, la standard deviation e la skewness per tutte le variabili del dataset e mostrare l'output in formato table nella Command Window. Nella prima colonna della tabella di output ci devono essere le mediane, nella seconda le medie, nella terza il MAD, nella quarta colonna le std e nella quinta le skewness **(punti 4)**

```
S=grpstatsFS(citiesItaly); % v. file di riepilogo nel sito di Giappichelli
disp(S(:, [2 1 4 3 5])) % riordino le colonne
```

	median	mean	MAD	std	skewness
	-----	-----	-----	-----	-----
addedval	18469	18096	6001.8	4941.6	0.1079
depos	7661.1	7769.1	3150.5	2841.4	1.0734
pensions	9975.2	10044	1170.6	1230.8	0.31692
unemploy	6.42	10.173	4.8778	7.8789	1.0687
export	21.53	23.11	16.457	15.642	0.5797
bankrup	29.62	30.467	11.49	12.11	1.0067
billsoverd	40.6	44.614	19.496	22.783	0.98031