

## ESERCIZIO 1

Una società di ricerche di mercato ha svolto un'indagine per conoscere il reddito mensile medio dalle famiglie residenti nelle regioni Lazio e Campania. I dati in Euro sono riportati nelle due tabelle seguenti:

REGIONE LAZIO		REGIONE CAMPANIA	
Reddito mensile medio	Frequenze Assolute	Reddito mensile medio	Frequenze Assolute
1'000	27	750	13
1'500	33	1'250	26
2'000	49	1'750	57
2'500	57	2'500	86
3'000	69	3'125	51
3'500	36	3'725	21
4'000	15	4'850	7

Confrontare, analiticamente e graficamente, la forma e la variabilità delle due distribuzioni. Commentare i risultati ottenuti.

### SVOLGIMENTO

Ai fini del confronto tra le due distribuzioni è possibile utilizzare, quali indici di posizione, media, mediana e quartili e calcolare per le due distribuzioni i più comuni indici di variabilità assoluta e relativa (scarto quadratico medio assoluto e relativo, coefficiente di variazione)

La seguente tabella riporta, per la regione Lazio, i calcoli per la determinazione dei momenti primo e secondo della distribuzione, nonché delle frequenze cumulate, utili per il calcolo di mediana e quartili:

REGIONE LAZIO							
Reddito mensile medio	Frequenze		Frequenze		$x_i \times f_i$	$x_i^2$	$x_i^2 \times f_i$
	Assolute	Relative	Assolute	Relative			
1'000	27	0.0944	27	0.0944	94.4056	1'000'000	94'405.59
1'500	33	0.1154	60	0.2098	173.0769	2'250'000	259'615.38
2'000	49	0.1713	109	0.3811	342.6573	4'000'000	685'314.69
2'500	57	0.1993	166	0.5804	498.2517	6'250'000	1'245'629.37
3'000	69	0.2413	235	0.8217	723.7762	9'000'000	2'171'328.67
3'500	36	0.1259	271	0.9476	440.5594	12'250'000	1'541'958.04
4'000	15	0.0524	286	1.0000	209.7902	16'000'000	839'160.84
	<b>286</b>	<b>1.0000</b>			<b>2482.5175</b>	<b>50'750'000</b>	<b>6837412.5874</b>

Analoghi calcoli sono riportati per la regione Campania:

REGIONE CAMPANIA							
Reddito mensile medio	Frequenze		Frequenze		$x_i \times f_i$	$x_i^2$	$x_i^2 \times f_i$
	Assolute	Relative	Assolute	Relative			
750	13	0.0498	13	0.0498	37.3563	562'500	28'017.24
1'250	26	0.0996	39	0.1494	124.5211	1'562'500	155'651.34
1'750	57	0.2184	96	0.3678	382.1839	3'062'500	668'821.84
2'500	86	0.3295	182	0.6973	823.7548	6'250'000	2'059'386.97
3'125	51	0.1954	233	0.8927	610.6322	9'765'625	1'908'225.57
3'725	21	0.0805	254	0.9732	299.7126	13'875'625	1'116'429.60
4'850	7	0.0268	261	1.0000	130.0766	23'522'500	630'871.65
	<b>261</b>	<b>1.0000</b>			<b>2408.2375</b>	<b>58'601'250</b>	<b>6567404.2146</b>

Dai calcoli precedenti è facile arrivare ai seguenti risultati per gli indici che interessano:

REGIONE LAZIO			REGIONE CAMPANIA	
286	n		261	
2'482.5175	$\mu_1$		2'408.2375	
6'837'412.5874	$\mu_2$		6'567'404.2146	
674'519.5364	$\sigma^2$		767'796.1275	
821.2914	$\sigma$		876.2398	
0.3308	CV		0.3639	
0.0012	$\sigma_{rel}$		0.0014	
3'000	Mo		2'500	
2'000	Q1		1'750	
2'500	Q2		2'500	
3'000	Q3		3'125	

Per il calcolo di alcuni indici di forma è necessario unire ai calcoli precedenti i calcoli dei momenti terzi standardizzati, riportati per le due regioni, nelle seguenti tabelle:

REGIONE LAZIO					
Reddito mensile medio	Frequenze		$z_i = (x_i - \mu_i) / \sigma$	$z_i^3$	$z_i^3 \times f_i$
	Assolute	Relative			
1'000	27	0.0944	-1.8051	-	0.56
1'500	33	0.1154	-1.1963	-	0.20
2'000	49	0.1713	-0.5875	-	0.03
2'500	57	0.1993	0.0213	0	0.00
3'000	69	0.2413	0.6301	0	0.06
3'500	36	0.1259	1.2389	2	0.24
4'000	15	0.0524	1.8477	6	0.33
	<b>286</b>	<b>1.0000</b>			<b>-0.1570</b>

REGIONE CAMPANIA					
Reddito mensile medio	Frequenze		$z_i = (x_i - \mu_i) / \sigma$	$z_i^3$	$z_i^3 \times f_i$
	Assolute	Relative			
750	13	0.0498	-1.8924	7	0.34
1'250	26	0.0996	-1.3218	2	0.23
1'750	57	0.2184	-0.7512	0	0.09
2'500	86	0.3295	0.1047	0	0.00
3'125	51	0.1954	0.8180	1	0.11
3'725	21	0.0805	1.5027	3	0.27
4'850	7	0.0268	2.7866	22	0.58
	<b>261</b>	<b>1.0000</b>			<b>0.3005</b>

Dai calcoli precedenti è facile calcolare i principali indici di forma, riportati per le due regioni nella seguente tabella:

REGIONE LAZIO		REGIONE CAMPANIA	
-	0.02	AHS	-
	0	AYB	
-	0.1570	AF	-

(AHS→Indice di Hotelling e Solomon, AYB→Indice di Yule e Bowley e AF→Indice di Fisher)

E' infine possibile calcolare l'indice di curtosi per valutare la normalità delle due distribuzioni. Per il calcolo dell'indice di curtosi è necessario procedere al calcolo dei momenti quarti standardizzati, riportati, per le due regioni, nelle seguenti tabelle:

REGIONE LAZIO					
Reddito mensile medio	Frequenze		$z_i = (x_i - \mu_i) / \sigma$	$z_i^4$	$z_i^4 \times f_i$
	Assolute	Relative			
1'000	27	0.0944	-1.8051	11	1.00
1'500	33	0.1154	-1.1963	2	0.24
2'000	49	0.1713	-0.5875	0	0.02
2'500	57	0.1993	0.0213	0	0.00
3'000	69	0.2413	0.6301	0	0.04
3'500	36	0.1259	1.2389	2	0.30
4'000	15	0.0524	1.8477	12	0.61
	<b>286</b>	<b>1.0000</b>			<b>2.2049</b>

REGIONE CAMPANIA					
Reddito mensile medio	Frequenze		$z_i = (x_i - \mu_i) / \sigma$	$z_i^4$	$z_i^4 \times f_i$
	Assolute	Relative			
750	13	0.0498	-1.8924	13	0.64
1'250	26	0.0996	-1.3218	3	0.30
1'750	57	0.2184	-0.7512	0	0.07
2'500	86	0.3295	0.1047	0	0.00
3'125	51	0.1954	0.8180	0	0.09
3'725	21	0.0805	1.5027	5	0.41
4'850	7	0.0268	2.7866	60	1.62
	<b>261</b>	<b>1.0000</b>			<b>3.1276</b>

Da cui risultano i seguenti indici di curtosi

REGIONE LAZIO			REGIONE CAMPANIA	
2.20	$\mu_4$ (std.)		3.13	
-0.80	AK		0.13	

Un confronto tra le due distribuzioni può essere effettuato anche andando a misurare la categoria degli indici di disuguaglianza (differenze medie con e senza ripetizione).

I calcoli sono riportati nelle seguenti tabelle:

REGIONE LAZIO									
Reddito Mensile Medio	Reddito Mensile Medio	1'000	1'500	2'000	2'500	3'000	3'500	4'000	
	Frekuensi	27	33	49	57	69	36	15	286
1'000	27	0	445'500	1'323'000	2'308'500	3'726'000	2'430'000	1'215'000	
1'500	33	445'500	0	808'500	1'881'000	3'415'500	2'376'000	1'237'500	
2'000	49	1'323'000	808'500	0	1'396'500	3'381'000	2'646'000	1'470'000	
2'500	57	2'308'500	1'881'000	1'396'500	0	1'966'500	2'052'000	1'282'500	
3'000	69	3'726'000	3'415'500	3'381'000	1'966'500	0	1'242'000	1'035'000	
3'500	36	2'430'000	2'376'000	2'646'000	2'052'000	1'242'000	0	270'000	
4'000	15	1'215'000	1'237'500	1'470'000	1'282'500	1'035'000	270'000	0	
	286								

75'816'000	Totale
286	n
81'510	n x (n-1)
930.14	$\Delta$ (senza rip.)
81'796	n <sup>2</sup>
926.89	$\Delta$ (con rip.)

REGIONE CAMPANIA								
Reddito Mensile Medio	Reddito Mensile Medio	750	1'250	1'750	2'500	3'125	3'725	4'850
	Frekuensi	13	26	57	86	51	21	7
750	13	0	87'750	321'750	796'250	1'296'750	2'018'250	1'287'000
1'250	26	87'750	0	214'500	955'500	1'852'500	3'139'500	2'106'000
1'750	57	1'154'250	470'250	0	698'250	2'436'750	4'916'250	3'591'000
2'500	86	3'483'000	2'838'000	2'107'000	0	2'967'000	3'096'000	1'935'000
3'125	51	2'926'125	2'734'875	2'811'375	1'816'875	439'875	688'500	669'375
3'725	21	1'545'075	1'541'925	1'775'025	1'466'325	1'050'525	170'100	86'625
4'850	7	727'650	773'850	977'550	937'650	893'550	340'200	89'250
	261							

72'036'300	Totale
261	n
67'860	n x (n-1)
1'061.54	$\Delta$ (senza rip.)
68'121	n <sup>2</sup>
1'057.48	$\Delta$ (con rip.)

Il confronto grafico tra le due distribuzioni può essere effettuato ricorrendo alla costruzione dei boxplot o degli istogrammi normalizzati: i primi possono essere calcolati sui dati disponibili mentre i secondi richiedono una previa categorizzazione degli stessi.

Le statistiche necessarie per la costruzione dei boxplot sono quasi tutte state calcolate in precedenza: esse sono riassunte nella seguente tabella (unitamente a quelle mancanti relative al tracciamento dei baffi e agli individuali valori mancanti):

REGIONE LAZIO			REGIONE CAMPANIA
1'000	Min		750
4'000	Max		4'850
2'000	Q1		1'750
2'500	Q2		2'500
3'000	Q3		3'125
500	a	-	313
3'500	b		3'813
1000	$\alpha$		750
3500	$\beta$		3725

Per la costruzione dell'istogramma normalizzato è necessario passare dalla distribuzione di frequenze alla distribuzione in classi: si utilizzano, per la costruzione di tale distribuzione, cinque classi equiampie. La seguente tabella riporta il calcolo dell'ampiezza di classe da utilizzare per ciascuna regione.

REGIONE LAZIO			REGIONE CAMPANIA
1'000	Min		750
4'000	Max		4'850
286	n		261
5	Classi		5
600	Ampiezza		820

Le due tabelle seguenti, infine, riportano la distribuzione in classi (5 classi equiampie) per i due casi. Viene riportato il calcolo della densità di frequenza: tale colonna va usata per il calcolo dell'istogramma normalizzato di frequenza, grazie al quale è possibile effettuare il confronto grafico tra le due distribuzioni

REGIONE LAZIO						
	Estr. Inf.	Estr. Sup.	Classi	Frequenza	Ampiezza	Densità Frequenza
1	1'000	1'600	1000 - 1600	60	600	0.10
2	1'600	2'200	1600 - 2200	49	600	0.08
3	2'200	2'800	2200 - 2800	57	600	0.10
4	2'800	3'400	2800 - 3400	69	600	0.12
5	3'400	4'000	3400 - 4000	51	600	0.09
				<b>286</b>		

REGIONE CAMPANIA						
	Estr. Inf.	Estr. Sup.	Classi	Frequenza	Ampiezza	Densità Frequenza
1	750	1'570	750 - 1570	39	820	0.05
2	1'570	2'390	1570 - 2390	57	820	0.07
3	2'390	3'210	2390 - 3210	137	820	0.17
4	3'210	4'030	3210 - 4030	21	820	0.03
5	4'030	4'850	4030 - 4850	7	820	0.01
				<b>261</b>		