

Esercizio 1.

Consideriamo la seguente successione di coppie di dati che identificano il genere (X) e lo stipendio medio orario espresso in euro (Y), raccolti su una popolazione di 20 unità statistiche.

Le coppie di dati sono: (F, 11.3), (F, 11.5), (F, 9.3), (M, 9.2), (M, 7.5), (F, 12.9), (M, 12.1), (M, 8.4), (F, 9), (M, 7.5), (M, 12.3), (M, 13.4), (M, 10.8), (F, 10.2), (M, 11.8), (M, 9.2), (M, 9.5), (M, 9.7), (F, 9.7), (F, 13).

- a) Costruire la distribuzione delle frequenze assolute congiunte, considerando i seguenti intervalli per il carattere stipendio medio orario $7.5| - 9.5$, $9.5| - 12$, $12| - 13.5$.
- b) Ottenere dalla tabella precedente le frequenze relative congiunte.
- c) Dalla tabella ottenuta al punto a), costruire le distribuzioni di frequenze marginali per il carattere X e Y.
- d) Qual è la percentuale di maschi?
- e) Qual è, tra i maschi, la percentuale di quelli che hanno uno stipendio medio orario compreso nell'intervallo 7.5-9.5?
- f) Qual è la percentuale di femmine tra coloro che hanno uno stipendio medio orario compreso nell'intervallo 7.5 -9.5?
- g) Qual è la percentuale di coloro che hanno uno stipendio medio orario di almeno 9.5 euro?
- h) Qual è la numerosità e qual è la percentuale di coloro che hanno al massimo uno stipendio medio orario pari a 9.5 euro (estremo escluso)?
- i) Calcolare la media aritmetica per il carattere Y utilizzando i valori iniziali della successione e utilizzando la distribuzione in classi. Commentare le differenze ottenute.

A)

	7.5 –9.5	9.5 –12	12 –13.5	
M	5	4	3	12
F	2	4	2	8
				20

B) Frequenze relative congiunte

	7.5 –9.5	9.5 –12	12 –13.5	
M	0.25	0.2	0.15	0.6
F	0.1	0.2	0.1	0.4
				1

C)

X_i	n_i	y_i – y_{i+1}	n_i
M	12	7.5 –9.5	7
F	8	9.5 –12	8
		12 –13.5	5

D) guardando dalla prima tabella, o la ultima fatta notiamo che su un totale di 20 persone, 12 sono maschi:

$$12/20=60\%$$

E) Qual è, tra i maschi, la percentuale di quelli che hanno uno stipendio medio orario compreso nell'intervallo 7.5-9.5?

F) Qual è la percentuale di femmine tra coloro che hanno uno stipendio medio orario compreso nell'intervallo 7.5 -9.5?

G) Qual è la percentuale di coloro che hanno uno stipendio medio orario di almeno 9.5 euro?

$y_i y_{i+1}$	n_i	R_i	f_{Ri}
7.5 9.5	7	7	0.35
9.5 12	8	15	0.75
12 13.5	5	20	1.00

frequenza retro cumulata

	7.5 9.5	9.5 12	12 13.5	
M	5	4	3	12
F	2	4	2	8
	7	8	5	20

E) $5/12=41.7\%$ in questo caso viene domandato quanti, in percentuale, hanno uno stipendio orario all'interno della prima classe rispetto a tutti gli uomini all'interno dell'analisi (12)

F) $2/7=29\%$ DIVERSO! Qui viene domandato quante persone sono donne tra COLORO CHE GUADAGNANO UNO STIPENDIO RIENTRANTE NELLA PRIMA CLASSE (7 persone)

G) $13/20=65\%$

H) Qual è la numerosità e qual è la percentuale di coloro che hanno al massimo uno stipendio medio orario pari a 9.5 euro (estremo escluso)?

7 persone che corrispondono a $7/20=35\%$

Esso può essere fatto anche sottraendo a 1 la percentuale ottenuta nell'esercizio precedente, $1-0.35=0.65=35\%$

l) Calcolare la media aritmetica per il carattere Y utilizzando i valori iniziali della successione e utilizzando la distribuzione in classi. Commentare le differenze ottenute.

Utilizzando i valori della successione:

$$\mu = \frac{1}{N} \sum_{h=1}^N y_h =$$

$$\mu = 1/20 \cdot (11.3 + 11.5 + 9.3 + 12.9 + 12.1 + 12.1 + 8.4 + 9 + 7.5 + 12.3 + 13.4 + 10.8 + 10.2 + 11.8 + 9.2 + 9.5 + 9.7 + 9.7 + 13) = 10.415$$

Utilizzando la distribuzione in classi

$$\mu = \frac{1}{N} \sum_{i=1}^K c_i \cdot n_i = \frac{1}{20} \cdot (59.5 + 86 + 63.75) = 10.463$$

con c_i = valore centrale della classe
 n_i = frequenze assolute

y_i - y_{i+1}	n_i	c_i	$c_i \cdot n_i$
7.5 - 9.5	7	8.5	59.5
9.5 - 12	8	10.75	86
12 - 13.5	5	12.75	63.75

Le due medie sono diverse, perché come ipotesi di lavoro della seconda media attribuiamo al punto centrale della classe gran parte del carattere della classe, così facendo perdiamo informazioni.

es 1 10/12

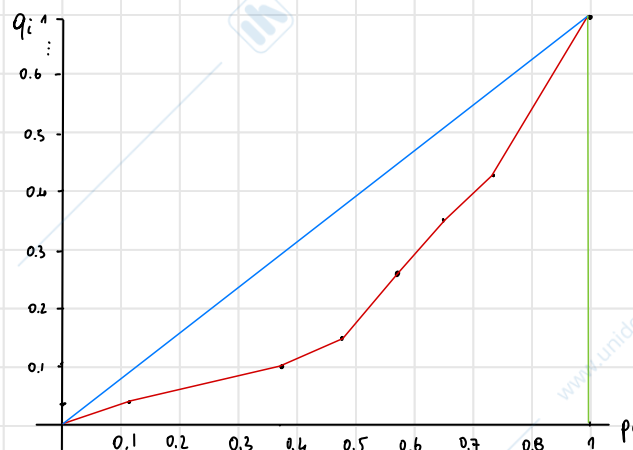
Calcolare il rapporto di concentrazione e disegnare diagramma Lorenz

Classi di investimento	1- 15	15- 30	30- 45	45- 60	60- 75	75- 90	90- 150
Imprese edili	14	23	10	12	6	7	28
N_i	14	37	47	59	65	72	100
$N_i/N \rightarrow P_i$	0.14	0.37	0.47	0.59	0.65	0.72	1
C_i	8	22.5	37.5	52.5	67.5	82.5	120
$C_i \cdot h_i$	112	517.5	375	630	405	577.5	3360
$C_j N_j$	112	629.5	1006.5	1636.5	2039.5	2617	5977
$q_i = \frac{N_i C_i}{N \mu}$	0.0187 $\rightarrow 112/5977$	0.1053	0.1681	0.2735	0.3412	0.4378	1
$p_{i+1} - p_i$	0.14 $\hookrightarrow 0.14 - 0$	0.23 $\hookrightarrow 0.37 - 0.14$	0.10	0.12	0.06	0.07	0.28
$q_{i+1} + q_i$	0.0187	0.1241	0.2736	0.4415	0.6147	0.7791	1.4378

$$R = 1 - \sum_{i=0}^{k-1} (q_{i+1} + q_i) (p_{i+1} - p_i)$$

$$R = 1 - [(0.14 \cdot 0.0187) + (0.23 \cdot 0.1241) \dots + (0.28 \cdot 1.4378)] = 0.395$$

diagramma Lorenz



Esercizio 2

In un questionario online, 168 individui hanno dichiarato di sostenere una ONLUS. Nella tabella è riportato il numero di individui suddivisi per zona di residenza (X) e per tipo di associazione ONLUS sostenuta (Y). Dall'indagine emerge che:

Zona di residenza	Tipo di ONLUS					Tot
	Amnesty International	Telefono Azzurro	Mani Tese	Lila	WWF	
Nord est	20	5	10	5	2	42
Nord ovest	7	20	10	20	5	62
Centro	2	10	5	5	4	26
Sud	5	12	8	6	7	38
Tot	34	47	33	36	18	168

Si valuti, mediante il calcolo di un indice opportuno, l'eventuale grado di dipendenza esistente tra i due caratteri. Commentare il risultato ottenuto.

$$\chi^2 = \sum_{i=1}^h \sum_{j=1}^k \left(\frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}} \right)$$

con $\hat{n}_{ij} = \frac{n_{.j} \cdot n_{i.}}{N}$

\hat{n}_{ij} $\frac{12 \cdot 34}{168}$

	AI	TA	MT	LILA	WWF
NE	8.5	11.75	8.25	9	4.5
NO	12.55	17.35	12.18	13.29	6.64
C	5.26	7.27	5.11	5.57	2.79
S	7.69	10.63	7.46	8.14	4.07

$$\chi^2 = \left(\frac{(20 - 8.5)^2}{8.5} + \dots + \frac{(7 - 4.07)^2}{4.07} \right) = 37.483$$

Interpretazione del risultato: sappiamo che il valore può assumere il valore zero e il valore massimo dato da $N \cdot \min \{k-1, h-1\}$ e corrisponde a 504.

Se io metto a rapporto il valore ottenuto con la numerosità (N) ottengo 0.22

Dividendo invece il valore per il massimo che esso può assumere ottengo 0,07 il quale varierà tra 0 e 1

Quindi tra i due caratteri, seppur minima una condizione c'è