Statistica I modulo - CdS SIAFA *(2023/24)*
Esame 17 gennaio 2024 - Soluzioni

1. I seguenti valori rappresentano il numero di nuovi ricercatori reclutati nell'anno 2022 in 64 enti di ricerca:

Numero ric. (x_i)	Enti (n_i)
0	8
1	24
2	16
3	8
4	8
TOT	64

(a) Calcolare la media.

La media aritmetica è data da:

$$\bar{x} = \frac{0 \cdot 8 + 1 \cdot 24 + \ldots + 4 \cdot 8}{64} = \frac{112}{64} = 1,75$$

(b) Individuare la mediana e identificare la moda.

Poiché $n = 64$ *è pari, la mediana è la semisomma dei valori in posizione* $n/2 = 32$ *e* $n/2 + 1 = 33$*. Dopo aver calcolato le frequenze cumulate, si osserva che il valore in posizione 32 è 1, il valore in posizione 33 è 2, dunque la mediana è pari a 1,5. La moda è pari a 1.*

(c) Calcolare la deviazione standard.

La deviazione standard è data da:

$$\sigma = \sqrt{\frac{(0 - 1,75)^2 \cdot 8 + \ldots + (4 - 1,75)^2 \cdot 8}{64}} = 1,20$$

2. Si consideri la seguente tabella a doppia entrata riferita a 210 ricercatori con i caratteri X (*Fruizione di una borsa di studio* con modalità Si, No) e Y (*Velocità nella carriera* con modalità *Bassa, Elevata*):

X/ Y	**Bassa**	**Elevata**
Si	23	35
No	64	88

(a) Considerando i ricercatori che non hanno usufruito di una borsa di studio, calcolare la probabilità che la Velocità nella carriera sia elevata.

Si riscrive la tabella con i totali di riga e colonna.

X/ Y	**Bassa**	**Elevata**	TOT
Si	23	35	58
No	64	88	152
TOT	87	123	210

$$P(Elevata|No) = \frac{88}{152} = 0,579$$

(b) Ricavare la distribuzione marginale della variabile X.

La distribuzione marginale è la seguente:

X	Frequenze
Si	58
No	152
TOT	210

(c) Calcolare l'indice di associazione Chi-quadrato.

L'indice χ^2 *è*

$$\chi^2 = \sum_{i=1}^{H}\sum_{j=1}^{K}\frac{(n_{ij}-n'_{ij})^2}{n'_{ij}}$$

dove n'_{ij} *indica la generica frequenza congiunta teorica, calcolata moltiplicando il totale di riga per il totale di colonna e dividendo per il numero complessivo di osservazioni. Le frequenze teoriche sono riportate nella tabella seguente.*

X / Y	Bassa	Elevata	TOT
Si	$(58{\cdot}87)/210 = 24,03$	$(58{\cdot}123)/210 = 33,97$	58
No	62,97	89.03	152
TOT	87	123	210

L'indice χ^2 *è dunque pari a*

$$\chi^2 \quad = \quad \frac{(23-24,03)^2}{24,03} + \ldots + \frac{(88-89,03)^2}{89,03} = 0,10$$

(d) Calcolare l'indice V di Cramer e giudicare il risultato ottenuto.

L'indice V *di Cramer è dato da*

$$V = \sqrt{\frac{\chi^2/n}{min(H-1,K-1)}} = \sqrt{\frac{0,10/210}{1}} = 0,022$$

Si rileva un grado quasi nullo di dipendenza tra i due caratteri.

3. Per 6 ricercatori si registrano per l'ultimo quinquennio l'entità dei finanziamenti (in migliaia di euro) ricevuti (X) e il numero di pubblicazioni (Y):

Finanziamenti (X)	Pubblicazioni (Y)
10	24
60	53
30	66
20	45
44	73
46	81

(a) Rappresentare il diagramma di dispersione.

Il diagramma di dispersione è riportato in Figura 1.

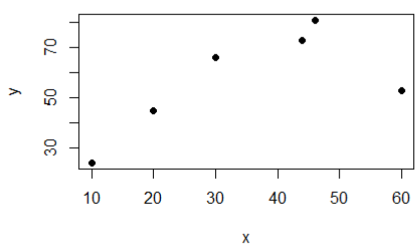

Figura 1: Diagramma di dispersione.

(b) Identificare i coefficienti della regressione (con Y variabile dipendente) e rappresentare la retta di regressione.

Per individuare la retta di regressione si svolgono i calcoli sotto riportati dopo aver individuato $\bar{x} = 35$ *e* $\bar{y} = 57$.

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	Prodotto	$(x_i - \bar{x})^2$
10	24	-25	-33	825	625
60	53	25	-4	-100	625
30	66	-5	9	-45	25
20	45	-15	-12	180	225
44	73	9	16	144	81
46	81	11	24	264	121
Tot				1268	1702

Dunque

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{6}[(x_i - \bar{x})(y_i - \bar{y})]}{\sum_{i=1}^{6}(x_i - \bar{x})^2} = \frac{1268}{1702} = 0,745$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 57 - 9,745 \cdot 35 = 30,925$$

L'equazione della retta di regressione è dunque

$$\hat{Y} = 30,925 + 0,745X$$

La retta di regressione è riportata in Figura 2.

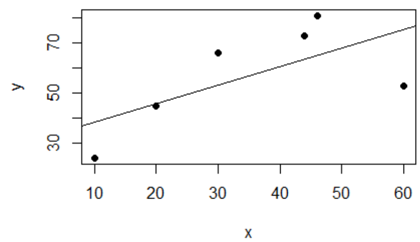

Figura 2: Diagramma di dispersione e retta di regressione.

(c) Che tipo di relazione esiste tra X e Y? Se i finanziamenti aumentassero di 10 (migliaia), qual è la variazione attesa del numero di pubblicazioni?

Poiché $\hat{\beta}_1 > 0$*, tra* X *e* Y *esiste concordanza (ovvero una relazione diretta). Se i finanziamenti aumentassero di 10 (migliaia), la variazione attesa del numero di pubblicazioni è 7,45.*