

Simulazione d'esame

Maura Mezzetti

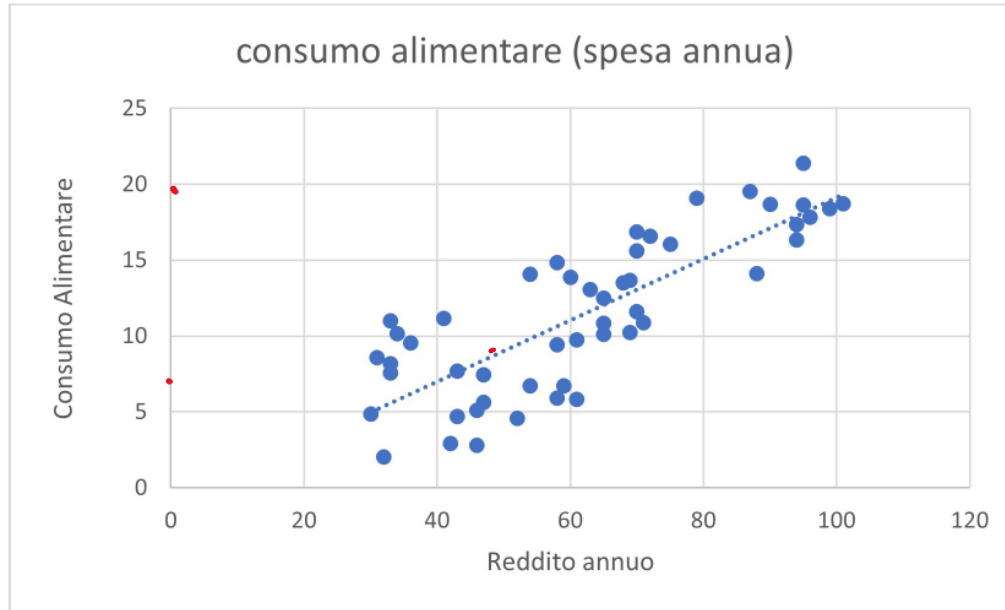
Affermazione	Vero o Falso?
Gli stimatori dei minimi quadrati dei coefficienti della retta di regressione sono stimatori non distorti	V
La somma di due variabili casuali si distribuisce asintoticamente come una Gaussiana	F

Affermazione	Vero o Falso?
Nel caso di popolazione Bernoulliana con parametro p , l'errore quadratico medio della proporzione campionaria è $p(1-p)/n$	V
La somma delle frequenze relative è sempre pari al numero di modalità osservate	F

17. Si indichi se le seguenti affermazioni sono Vere o False (2 punti)

Affermazione	Vero o Falso?
Al crescere di n , la distribuzione Binomiale si può approssimare con una distribuzione Gaussiana	V
La varianza è sempre maggiore della deviazione standard	F

19. (1 punto) Il seguente grafico mostra il diagramma di dispersione del reddito annuale e spesa annuale in beni alimentari di 50 famiglie italiane



Quali dei seguenti parametri vi sembrano più adeguati per spiegare la relazione che lega le due variabili:

- (a) $\beta_0 = 0$ e $\beta_1 = 1$
- ☒ (b) $\beta_0 = -1$ e $\beta_1 = 0.2$
- (c) $\beta_0 = -1$ e $\beta_1 = -0.4$
- (d) $\beta_0 = 5$ e $\beta_1 = 0.4$

Esercizio 2

Il direttore di una catena di negozi di scarpe per corrispondenza vorrebbe sviluppare un modello per prevedere le vendite settimanali (in migliaia di dollari) per i singoli negozi in base al numero di clienti che hanno visitato il sito nell'ultima settimana. Un campione casuale di 5 negozi è stato selezionato tra tutti i negozi della catena.

Negozio	Numero di clienti (X)	Vendite (Y)
1	900	10
2	940	13
3	500	8
4	750	9
5	800	10

Soluzione 2

Calcolare i parametri della retta di regressione delle vendite settimanali (Y) sul numero di clienti (X).



	x	y	x ²	xy
	900	10	810000	9000
	940	13	883600	12220
	500	8	250000	4000
	750	9	562500	6750
	800	10	640000	8000
somma	3890	50	3146100	39970
media	778	10	629220	7994

	x	y	x²	xy
	9	10	81	90
	9,4	13	88,36	122,2
	5	8	25	40
	7,5	9	56,25	67,5
	8	10	64	80
somma	38,9	50	314,61	399,7
media	7,78	10	62,922	79,94

Soluzione 2

Calcolare i parametri della retta di regressione delle vendite settimanali (Y) sul numero di clienti (X).

$$\beta_1 = \frac{Cov(x, y)}{Var(x)} = \frac{\frac{\sum_{i=1}^n x_i y_i}{n} - \bar{x}\bar{y}}{\frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2} = \frac{7994 - 778 \times 10}{629220 - 778^2}$$

$$\beta_1 = \frac{214}{23936} = 0.0089$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x} = 10 - 0.0089 \times 778 = 3.0758$$

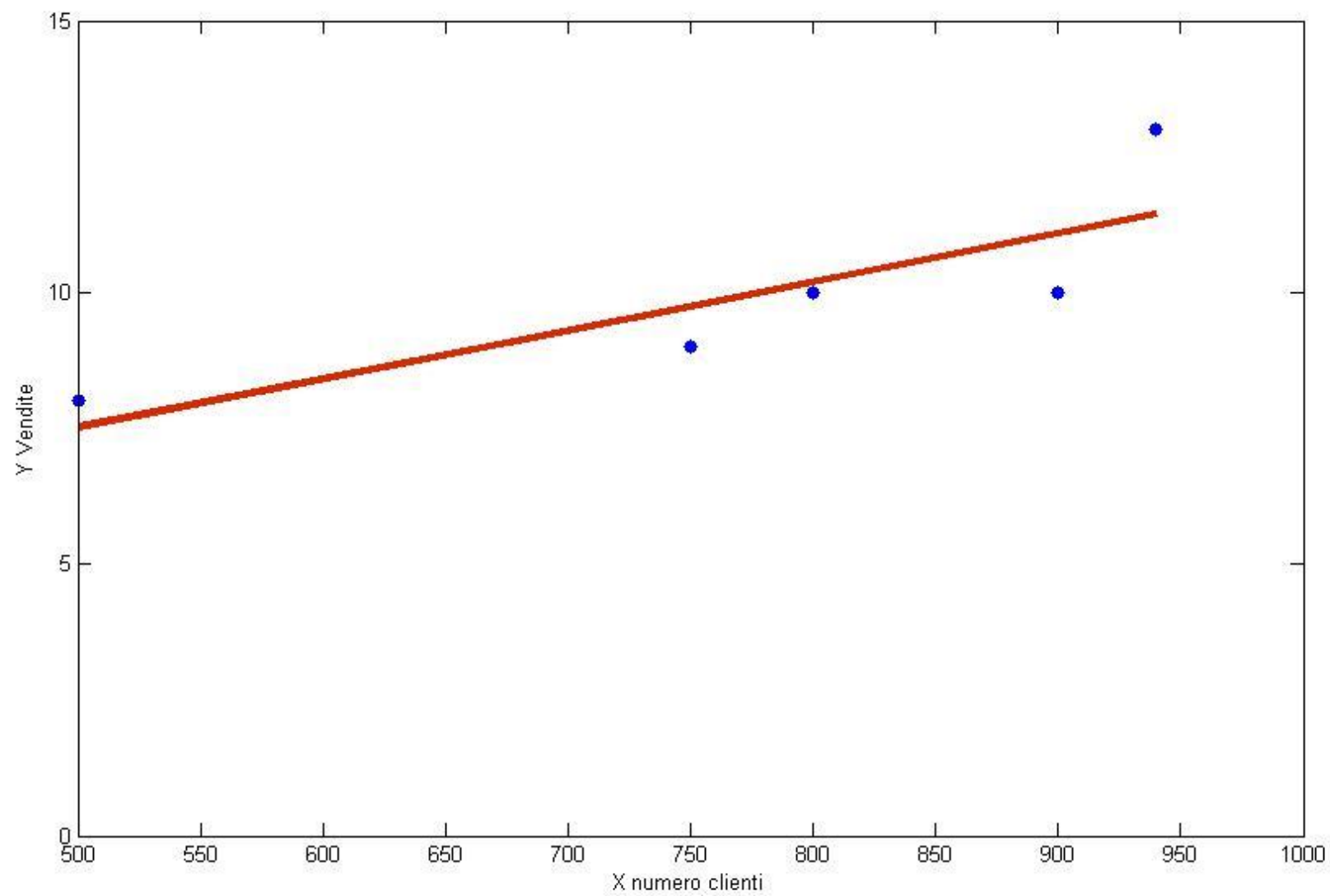
Soluzione 2

Calcolare i parametri della retta di regressione delle vendite settimanali (Y) sul numero di clienti (X).

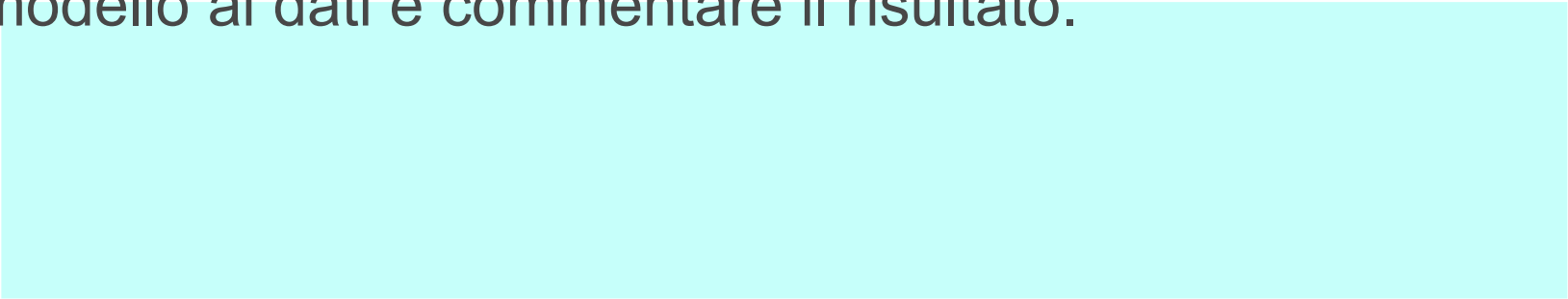
$$\beta_1 = \frac{Cov(x, y)}{Var(x)} = \frac{\frac{\sum_{i=1}^n x_i y_i}{n} - \bar{x}\bar{y}}{\frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2} = \frac{79.94 - 7.78 \times 10}{62.922 - 7.78^2}$$

$$\beta_1 = \frac{2.14}{2.3936} = 0.89$$

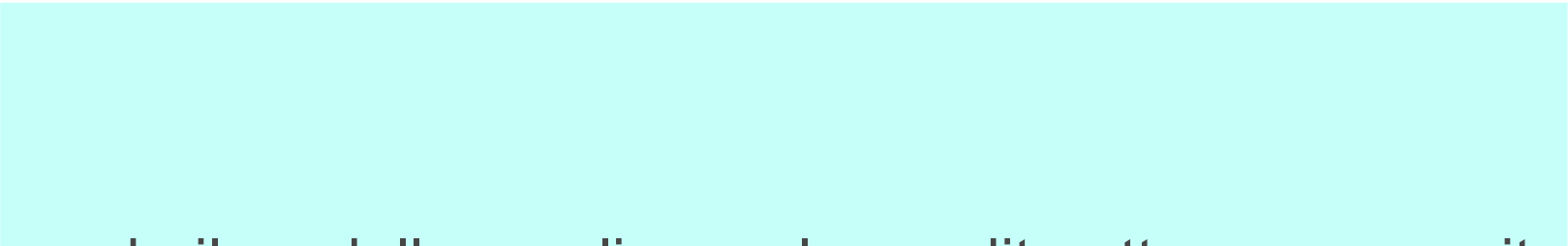
$$\beta_0 = \bar{y} - \beta_1 \bar{x} = 10 - 0.89 \times 7.78 = 3.0758$$



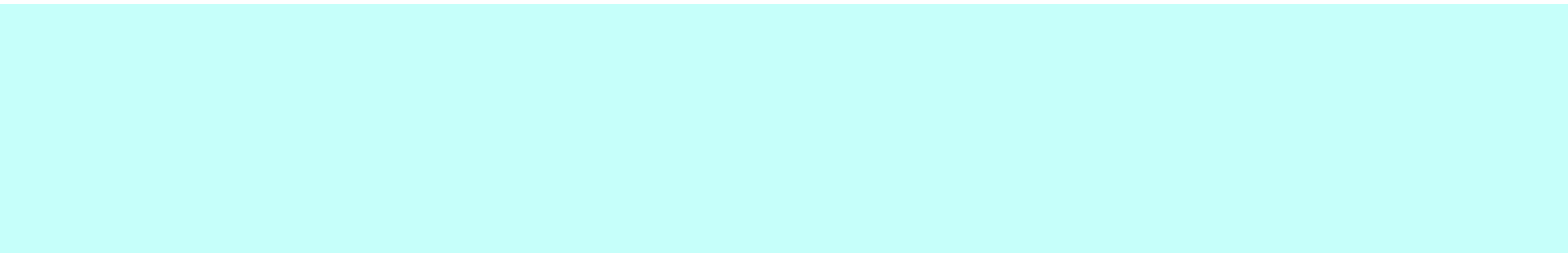
calcolare un opportuno indice di bontà di adattamento del modello ai dati e commentare il risultato.



sulla base dei risultati ottenuti al punto precedente dire se che conclusioni può trarre il direttore



secondo il modello, quali sono le vendite attese per un sito con 600 visitatori nell'ultima settimana?



	x	y	x ²	xy	y ²
	900	10	810000	9000	100
	940	13	883600	12220	169
	500	8	250000	4000	64
	750	9	562500	6750	81
	800	10	640000	8000	100
somma	3890	50	3146100	39970	514
media	778	10	629220	7994	102.8

Soluzione 2

calcolare un opportuno indice di bontà di adattamento del modello ai dati e commentare il risultato

$$R^2 = \frac{Cov(x, y)^2}{Var(x)Var(y)} = \frac{214^2}{23936 \times (102.8 - 10^2)} = \frac{214^2}{23936 \times 2.8} = 0.6833$$

Soluzione 2

secondo il modello, quali sono le vendite attese per un sito con 600 visitatori nell'ultima settimana?

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \times 600$$

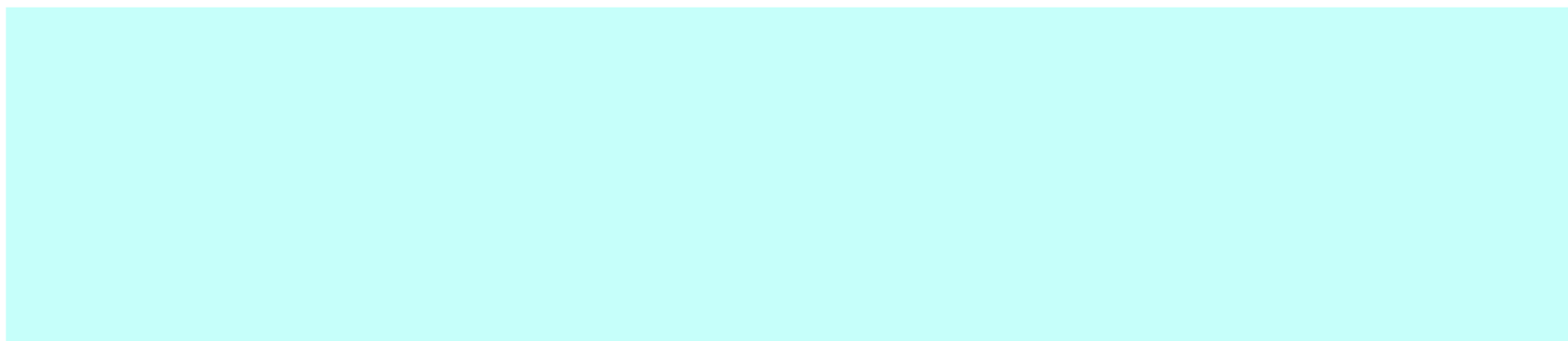
$$\hat{y} = 3.0758 + 0.0089 \times 600 = 8.4158$$

$$\hat{y} = 3.0758 + 0.89 \times 6 = 8.4158$$

Valutare la somma dei residui



Valutare la somma dei residui al quadrato



Valutare la somma dei residui

La somma dei residui è sempre nulla!

Valutare la somma dei residui al quadrato

Valutare la somma dei residui

La somma dei residui è sempre nulla!

Valutare la somma dei residui al quadrato

$$1 - \frac{\sum (Y_i - \hat{Y}_i)^2}{\sum (Y_i - \bar{Y})^2} = R^2$$

$$\sum (Y_i - \hat{Y}_i)^2 = (1 - R^2) \sum (Y_i - \bar{Y})^2$$

somma dei residui al quadrato = $(1 - 0.6833) * 5 * 2.8 = 4.4388$

Esercizio 3

Data una popolazione con media μ e varianza σ^2 e un campione casuale semplice di quattro variabili, considerare il seguente stimatore per μ

$$T_1 = \frac{1}{3}X_1 - \frac{2}{3}X_2 - X_3 + \frac{5}{3}X_4$$

- Indicare se T è uno stimatore corretto per μ
- Calcolare l'errore quadratico medio di T

$$T_1 = \frac{1}{3}X_1 - \frac{2}{3}X_2 - X_3 + \frac{5}{3}X_4$$

$$\begin{aligned} E(T_1) &= \frac{1}{3}E(X_1) - \frac{2}{3}E(X_2) - E(X_3) + \frac{5}{3}E(X_4) = \\ &= \frac{1}{3}\mu - \frac{2}{3}\mu - \mu + \frac{5}{3}\mu = \frac{1}{3}\mu \end{aligned}$$

$$B(T_1) = E(T_1) - \mu = \frac{1}{3}\mu - \mu = -\frac{2}{3}\mu$$

$$Var(T_1) = \left(\frac{1}{3}\right)^2 Var(X_1) + \left(-\frac{2}{3}\right)^2 Var(X_2) + (-1)^2 Var(X_3) + \left(\frac{5}{3}\right)^2 Var(X_4)$$

$$Var(T_1) = \sigma^2 \left(\frac{4}{9} + \frac{1}{9} + 1 + \frac{25}{9} \right) = \frac{39}{9} \sigma^2 = \frac{13}{3} \sigma^2$$

$$MSE(T_1) = \frac{13}{3} \sigma^2 + \frac{4}{9} \mu^2$$

Esercizio 5

Si vuole fare inferenza sul reddito medio annuale μ di una certa popolazione di famiglie. Selezionato un campione di 121 famiglie risulta che il reddito medio del campione è pari a 25 migliaia di euro mentre la varianza campionaria corretta s^2 è pari a 9.

a) Determinare una stima puntuale di μ .

Stima μ (1 punti)

b) Definire le proprietà dello stimatore utilizzato. (1 punti)

- corretto
- consistente
- efficiente
- asintoticamente gaussiano

Soluzione 5

$$\bar{x} = 25$$

c) Utilizzando i dati del campione costruire l'intervallo di confidenza al 95% per μ



d) Come cambia l'ampiezza dell'intervallo di confidenza se, a parità di altre condizioni, diminuiamo il livello di confidenza?



Intervallo di confidenza per la media della popolazione μ

- X distribuzione arbitraria $E(X)=\mu$
 $\text{Var}(X)=\sigma^2$ valore noto – n grande

$$\left(\bar{x}_n - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} ; \bar{x}_n + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

- σ^2 valore NON noto - n grande

$$\left(\bar{x}_n - z_{1-\alpha/2} \frac{s_c}{\sqrt{n}} ; \bar{x}_n + z_{1-\alpha/2} \frac{s_c}{\sqrt{n}} \right) \quad s_c^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

$$t_{1-\alpha/2}^{n-1} = t_{0.975}^{120} = 1.98 \approx z_{0.975} = 1.96$$

$$\begin{aligned} 95\%CI &= \left[\bar{x} - 1.96 \frac{s}{\sqrt{n}}; \bar{x} + 1.96 \frac{s}{\sqrt{n}} \right] = \\ &= \left[25 - 1.96 \times \frac{3}{11}; 25 + 1.96 \times \frac{3}{11} \right] = [24.465, 25.535] \end{aligned}$$

e) Verificare, ad un livello di significatività $\alpha = 0.05$, l'ipotesi che il reddito medio non sia cambiato rispetto all'anno precedente dove era pari a 30 migliaia di euro [$H_0: \mu = 30$ vs $H_1: \mu \neq 30$]. (2 punti)

c) Intervallo di confidenza

d)

e) Verifica di ipotesi

E' necessario ottenere regione di rifiuto?

$$R = \{\bar{x} \leq k_1 \cup \bar{x} \geq k_2\}$$

$$R = \left\{ \frac{\bar{x} - \mu_0}{s / \sqrt{n}} \leq k_1 \cup \frac{\bar{x} - \mu_0}{s / \sqrt{n}} \geq k_2 \right\}$$

$$R = \left\{ \left| \frac{\bar{x} - \mu_0}{s / \sqrt{n}} \right| \geq z_{1-\alpha/2} \right\} = \left\{ \left| \frac{\bar{x} - 30}{s / 11} \right| \geq 1.96 \right\}$$

$$\left| \frac{25 - 30}{3 / 11} \right| = 18.33 \geq 1.96 \quad \text{RIFIUTO IPOTESI NULLA}$$

$$p\text{-value} = 2 \times P(\bar{X} \leq 25)$$

$$P\left(\frac{\bar{X} - \mu_0}{s / \sqrt{n}} \leq \frac{\bar{x} - \mu_0}{s / \sqrt{n}}\right)$$

$$2 \times P(Z \leq -18.33) \approx 0$$

Esercizio 6

Il risultato di un test di apprendimento segue una distribuzione gaussiana con media 98 e deviazione standard 9. Gli studenti il cui risultato è inferiore a 80 sono considerati bisognosi di aiuto.

- a) Qual è la percentuale di studenti che hanno bisogno di aiuto?
- b) In una classe di 20 studenti, qual è la probabilità che il valore medio sia maggiore di 100?
- c) In un campione casuale di dimensione 4, qual è la probabilità che gli studenti che hanno bisogno di aiuto siano esattamente due?

a)

b)

c)

$$X \approx N(98, \sigma = 9)$$

$$a) P(X \leq 80) \quad ?$$

$$b) P(\bar{X} \geq 100) \quad ?$$

$$c) W_i = 1 \quad X \leq 80$$

$$Y = \sum_{i=1}^x W_i \quad P(Y = 2) \quad ?$$

z	Seconda cifra decimale di z									
	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981

$$X \approx N(98, \sigma = 9)$$

$$a) P(X \leq 80) = P\left(\frac{X - 98}{9} \leq \frac{80 - 98}{9}\right)$$

$$P(Z \leq -2) = 1 - 0.9772 = 0.0228$$

$$b) P(\bar{X} \geq 100) = P\left(\frac{\bar{X} - 98}{9 / \sqrt{20}} \geq \frac{100 - 98}{9 / \sqrt{20}}\right)$$

$$P(Z \geq 0.99) = 1 - 0.8389 = 0.1611$$

$$c) Y = \sum_{i=1}^x W_i \quad P(Y = 2)$$

$$X \approx N(98, \sigma = 9)$$

$$a) P(X \leq 80) = 0.0228$$

$$W_i = \begin{cases} 0 & X > 80 \\ 1 & X \leq 80 \end{cases} \quad W_i = \text{Bernoulli}(0.0228)$$

$$Y = \sum_{i=1}^x X_i \quad W_i = \text{Binomiale}(4, 0.0228)$$

$$P(Y = 2) = \binom{4}{2} 0.0228^2 (1 - 0.0228)^2$$

$$Y = \sum_{i=1}^4 X_i \quad W_i = \text{Binomiale}(4, 0.0228)$$

$$P(Y = 2) = \frac{4!}{2!2!} 0.0228^2 \times 0.9772^2 = \frac{4 \times 3 \times 2}{2 \times 2} \times 0.00052 \times 0.9549$$

$$P(Y = 2) = 0.00298$$

Esercizio 1

Un editore è interessato al numero di libri tascabili di fantascienza acquistati in un mese. L'editore conduce un sondaggio. Nel sondaggio, a ciascun adulto viene chiesto il numero di tascabili di fantascienza che lui / lei ha acquistato il mese precedente. I risultati sono riportati nella seguente tabella

Numero di libri	n_j
0	16
1	12
2	10
3	8
4	4

- a) Determinare media, moda, mediana e varianza del “Numero di libri acquistati in un mese”

- b) La frequenza relativa con cui un adulto compra meno di 2 libri al mese

Esercizio 1

Numero di libri	n_j	$x_j * n_j$	f_j	F_j	x_j^2	$x_j^2 * n_j$
0	16	0	0,32	0,32	0	0
1	12	12	0,24	0,56	1	12
2	10	20	0,2	0,76	4	40
3	8	24	0,16	0,92	9	72
4	4	16	0,08	1	16	64
	50	72				188

Soluzione 1

Determinare media, moda, mediana e varianza del “Numero di libri acquistati in un mese” (4 punti)

Media

$$\bar{x} = \frac{\sum_{j=1}^k x_j^* n_j}{\sum_{j=1}^k n_j} = \frac{72}{50} = 1.44$$

Mediana

$$\mathbf{Me=1}$$

Moda

$$\mathbf{Moda=0}$$

Varianza

$$\sigma_X^2 = \frac{\sum_{j=1}^k x_j^{*2} n_j}{\sum_{j=1}^k n_j} - \bar{x}^2 = 3.76 - 1.44^2 = 1.6864$$

Soluzione 1

La frequenza relativa con cui un adulto compra meno di 2 libri al mese

Frequenza

$$\begin{aligned} \textit{Freq}(X < 2) &= \textit{Freq}(X = 0) + \textit{Freq}(X = 1) = \\ &= 16 + 12 = 28 \\ &= 0.32 + 0.24 = 0.56 \end{aligned}$$

Esercizio 4

Enunciare e dimostrare il teorema di Bayes

Esercizio 7

Un dirigente di una compagnia di assicurazioni ha sviluppato un test attitudinale per agenti assicurativi. Sa che dall'attuale gruppo di agenti il 65% ha ottenuto buoni risultati di vendita ed il restante 35% ha ottenuto risultati scarsi. Dà il suo test all'intero gruppo di agenti e trova che il 73% di coloro che hanno ottenuto buoni risultati passa il test e che il 78% di coloro che hanno ottenuto scarsi risultati sbaglia il test.

1. Scegliendo un agente a caso e sottoponendogli il test, qual è la probabilità che passi il test?
2. Scegliendo un agente a caso e sottoponendogli il test, qual è la probabilità che chi passa il test abbia ottenuto buone vendite?
3. Scegliendo un agente a caso e sottoponendogli il test, qual è la probabilità che chi non passa il test non abbia ottenuto buone vendite?

$A = \{\text{l'agente ottiene buoni risultati di vendita}\}$

$A^c = \{\text{l'agente NON ottiene buoni risultati di vendita}\}$

$T = \{\text{l'agente PASSA il test}\}$

$T^c = \{\text{l'agente NON PASSA il test}\}$

$P(A) = 0.65$

$P(A^c) = 0.35$

$P(T|A) = 0.73$

$P(T^c | A^c) = 0.78$

1. Scegliendo un agente a caso e sottoponendogli il test, qual è la probabilità che passi il test?

$$P(T)$$

2. Scegliendo un agente a caso e sottoponendogli il test, qual è la probabilità che chi passa il test abbia ottenuto buone vendite?

$$P(A|T)$$

3. Scegliendo un agente a caso e sottoponendogli il test, qual è la probabilità che chi non passa il test non abbia ottenuto buone vendite?

$$P(A^c|T^c)$$

Tabular Approach

events	Prior Probabilities	Conditional probabilities
A_i	$P(A_i)$	$P(T A_i)$
A	0.65	0.75
A^c	0.35	0.22
<i>sum</i>	1	

Tabular Approach

Step 2 Nella Colonna 4 otteniamo le probabilità congiunte. Moltiplicando le probabilità a priori per le probabilità condizionate

$$P(A_i \cap T) = P(A_i) P(T|A_i).$$

Tabular Approach

events	Prior probabilities	Conditional probabilities	Joint probabilities
A_i	$P(A_i)$	$P(T A_i)$	$P(A_i \cap T)$
A	0.65	0.75	0.4875
A^c	0.35	0.22	0.077
<i>sum</i>	1		

Tabular Approach

Step 3 Sommando le probabilità congiunte otteniamo la probabilità di B

$$P(T) = P(T \cap A_1) + P(T \cap A_2) + \dots + P(T \cap A_n)$$

Tabular Approach

events	Prior probabilities	Conditional probabilities	Joint Probabilities
A_i	$P(A_i)$	$P(T A_i)$	$P(A_i \cap T)$
A	0.65	0.75	0.4875
A^c	0.35	0.22	0.077
<i>sum</i>	1		$P(T)=0.5645$

Tabular Approach

Step 4 Infine nell'ultima Colonna, la Colonna 5, otteniamo le probabilità a posteriori

$$P(A_i | T) = \frac{P(A_i \cap T)}{P(T)}$$

Tabular Approach

events	Prior probabilities	Conditional probabilities	Joint probabilities	Posterior probabilities
A_i	$P(A_i)$	$P(T A_i)$	$P(A_i \cap T)$	$P(A_i T)$
A	0.65	0.75	0.4875	0.8636
A^c	0.35	0.22	0.077	0.1364
<i>sum</i>	1		$P(T)=0.5645$	

9. Si indichi se le seguenti affermazioni sono Vere o False (2 punti)

Affermazione	Vero / Falsa?
Due eventi incompatibili con probabilità positiva non possono essere indipendenti.	
La media degli scarti dalla media ha valore nullo	