

PRIMA PROVA INTERMEDIA DI STATISTICA
CLEA (COD. 5047/4038)
5 Novembre 2003

*Cognome**Nome**Numero di matricola*

COMPITO B1

Ai fini della valutazione si terrà conto solo ed esclusivamente di quanto riportato negli appositi spazi. Al termine della prova, è OBBLIGATORIO consegnare il presente foglio ed il foglio di brutta (DI CUI NON SI TERRÀ CONTO AI FINI DELLA VALUTAZIONE).

APPROSSIMARE TUTTI I CALCOLI ALLA QUARTA CIFRA DECIMALE

Una Società distributrice di acque minerali ha deciso di comparare le caratteristiche attuali di alcune marche di acqua per mettere in relazione caratteristiche organolettiche delle acque e giudizio espresso dai consumatori, partendo da alcuni dati disponibili in rete. Le informazioni raccolte su 9 acque minerali sono quindi le seguenti:

- RF residuo fisso a 180°C (espresso in g/l)
 PR prezzo per una confezione da 6 bottiglie in €
 MAGN quantità di magnesio disciolta in un litro d'acqua (bassa, media, alta)
 OLIG acqua oligominerale (sì, no)
 C quantità di calcio disciolta in un litro d'acqua
 G_ETI voto medio dei giudizi dei 750 consumatori intervistati sulle caratteristiche presentate nell'etichetta della bottiglia d'acqua (1= voto minimo; 5=voto massimo).

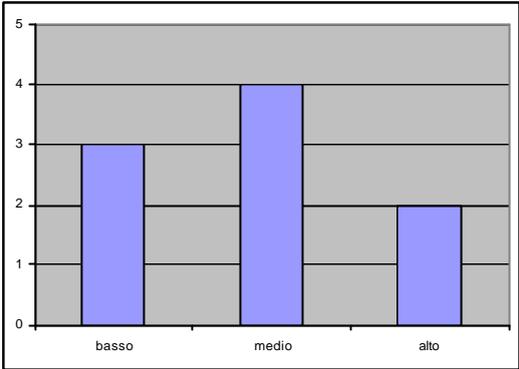
Di seguito sono riportati i dati, con alcuni calcoli utili.

	RF	PR	MAGN	C	OLIG	G_ETI	RF ²	PR ²	RF*PR
San Viandante	1,07	1,52	alto	208	no	3,8	1,1449	2,3104	1,6264
Ulivento	0,89	2,28	medio	190	no	4,1	0,7921	5,1984	2,0292
S. Benevento	0,25	1,59	medio	46	no	4,3	0,0625	2,5281	0,3975
Tronchetta	0,17	3,24	basso	59	sì	3,9	0,0289	10,4976	0,5508
Vanna	0,13	2,7	basso	30	sì	4,1	0,0169	7,2900	0,3510
Ferrareffe	1,27	1,69	medio	362	no	3,6	1,6129	2,8561	2,1463
Fuggi	0,12	3,75	basso	16	sì	3,6	0,0144	14,0625	0,4500
Pera	0,16	2,94	medio	36	sì	4,4	0,0256	8,6436	0,4704
Ontario	0,63	1,54	alto	133	no	4	0,3969	2,3716	0,9702
Totale	4,69	21,25		1080		35,8	4,0951	55,7583	8,9918

1. (2 punti) Quali indicatori si possono calcolare per i seguenti caratteri? Segnare con ✓ dove si possono calcolare.

	OLIG	C	RF	G_ETI
Coefficiente di variazione	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Media	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Mediana	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Moda	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

2. (3 punti) Determinare la distribuzione di frequenza per il carattere MAGN (quantità di magnesio) e rappresentarla graficamente e calcolarne infine la mediana.

<p>(1 punto) Distribuzione di frequenza</p> <table border="1" style="width: 100%; border-collapse: collapse; margin-top: 10px;"> <thead> <tr style="background-color: #003366; color: white;"> <th style="text-align: left;">MAGN</th> <th style="text-align: center;">Frequenze assolute</th> <th style="text-align: center;">Frequenze relative</th> </tr> </thead> <tbody> <tr> <td style="background-color: #cccccc;">basso</td> <td style="text-align: center;">3</td> <td style="text-align: center;">33,33%</td> </tr> <tr> <td style="background-color: #cccccc;">medio</td> <td style="text-align: center;">4</td> <td style="text-align: center;">44,44%</td> </tr> <tr> <td style="background-color: #cccccc;">alto</td> <td style="text-align: center;">2</td> <td style="text-align: center;">22,22%</td> </tr> </tbody> </table>	MAGN	Frequenze assolute	Frequenze relative	basso	3	33,33%	medio	4	44,44%	alto	2	22,22%	<p>(1 punto) Grafico</p> 
MAGN	Frequenze assolute	Frequenze relative											
basso	3	33,33%											
medio	4	44,44%											
alto	2	22,22%											
<p>(1 punto) Mediana</p> <p style="text-align: center; margin-left: 300px;">medio</p>													

3. (3 punti)

(a) Calcolare i coefficienti della retta di regressione di PR (prezzo) su RF (residuo fisso).

(2 punti)

Definita la retta di regressione $PR = a + b \cdot RF$, si ha

$$a = 3.0182$$

$$b = -1.2609$$

(b) Prevedere il prezzo di una nuova acqua minerale, sapendo che ha un residuo fisso pari a 1.

(1 punto)

1.7573

4. (2 punti) Calcolare la funzione di regressione per RF (residuo fisso) su OLIG (acqua oligominerale, sì/no). Cosa si può concludere sulla dipendenza in media (dipendenza regressiva) di RF rispetto a OLIG?

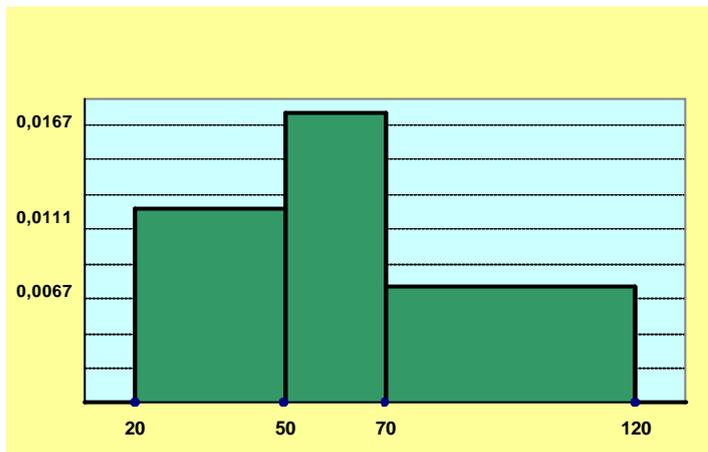
Avendo definito con $m(\text{olig})$ la funzione di regressione di RF su OLIG, si ha

$$m(RF|OLIG) = \begin{cases} 0.822 & \text{olig} = \text{"no"} \\ 0.145 & \text{olig} = \text{"sì"} \end{cases}$$

RF dipende regressivamente da OLIG.

5. (3 punti) Dato un carattere X, quantitativo continuo, si è deciso di rappresentarlo in 3 classi equifrequenti ($p_1=p_2=p_3$), i cui estremi sono indicati nel grafico ([20, 50); [50, 70); [70, 120)). Disegnare l'istogramma nel grafico sottostante, indicando chiaramente i valori sull'asse delle ordinate. Ricavare la mediana del carattere X.

(2 punti) Grafico



(1 punto) Mediana

60

6. (2 punti) Nelle 150 stanze di un pensionato studenti ci sono complessivamente 600 libri. Scrivere la distribuzione del carattere X = 'numero di libri per stanza' nei due casi estremi di equidistribuzione (equipartizione) e di concentrazione massima.

(1 punto) Caso equidistribuzione

x_i	n_i	p_i
4	150	1

(1 punto) Caso max concentrazione

x_i	n_i	p_i
0	149	0.993
600	1	0.007

7. (3 punti) Definire l'indice \tilde{J}^2 . Fornire un esempio di tabella a doppia entrata in cui $\tilde{J}^2 = 1$ e $\tilde{J}^2 \neq \pm 1$.

(1 punto) Definizione dell'indice

$$\tilde{J}^2 = \frac{\sum_{i=1}^k \sum_{j=1}^h \frac{[(p_{XY}(x_i, y_j) - p_X(x_i)p_Y(y_j))]^2}{p_X(x_i)p_Y(y_j)}}{\min\{(k-1), (h-1)\}}$$

con k =numero righe, h numero colonne della tab. a doppia entrata, $p_{XY}(\cdot)$ funzione di distribuzione delle frequenze congiunte, $p_X(\cdot)$ e $p_Y(\cdot)$ funzioni di distribuzione delle frequenze marginali risp. di X e Y .

(2 punti) Esempio

$x \setminus y$	2	5
0	0,3	
1		0,2
4	0,5	

8. (4 punti) In una stazione italiana la probabilità che un treno sia in ritardo è pari a 0,1. La probabilità per un soggetto AAA di perdere il treno è pari a 0,3 se il treno è in orario e a 0,1 se il treno è in ritardo. Calcolare la probabilità per il soggetto in questione di perdere il treno. Calcolare poi, sapendo che il soggetto ha perso il treno, la probabilità che tale treno fosse in ritardo.

(2 punti) Probabilità di perdere il treno

$$0.1 * 0.1 + 0.3 * 0.9 = 0.28$$

(2 punti) Probabilità che, sapendo che il soggetto ha perso il treno, il treno fosse in ritardo

$$\frac{0.1 * 0.1}{0.28} = 0.0357$$

9. (2 punti) In una libreria ci sono 10 libri per ciascuna delle 5 tipologie: giallo, avventura, saggio, attualità, comico. Si estraggono con reimmissione 20 libri. Scrivere la funzione di probabilità della variabile aleatoria che descrive il numero di volte che si estrae un libro di avventura sul totale di libri estratti.

Avendo definito X il numero di libri di avventura estratti su N tentativi, la sua distribuzione di probabilità seguirà una legge binomiale, di parametri $N=20$ e $p=0.2$, per cui

$$\Pr(X = x) = p(x) = \begin{cases} \binom{20}{x} 0.2^x 0.8^{20-x} & x = 0, 1, 2, \dots, 20 \\ 0 & \text{altrove} \end{cases}$$