PROVA SCRITTA DI STATISTICA (COD 4038-5047-371-377)
4 Giugno 2004

MODALITÀ A

SOLUZIONI

APPROSSIMARE TUTTI I CALCOLI ALLA QUARTA CIFRA DECIMALE

ESERCIZIO 1 (7 punti)

Un prospetto finanziario elenca 10 fondi di investimento esistenti nel mercato. Per ciascun fondo viene indicato:

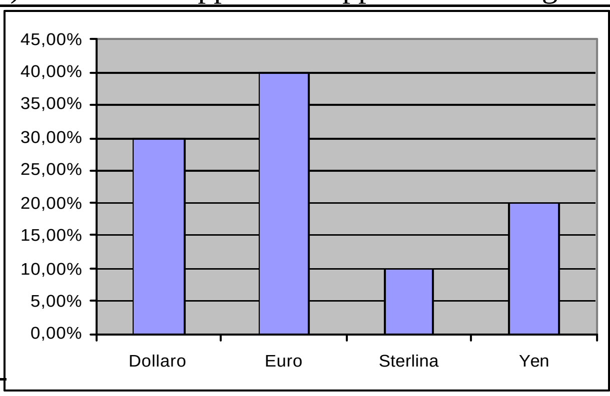
VALUTA	Moneta di denominazione
VALORE	Quotazione odierna
REND	Rendimento ottenuto nell'ultimo anno
TITOLI	Numero di titoli di cui è composto
RISCHIO	Grado di rischio

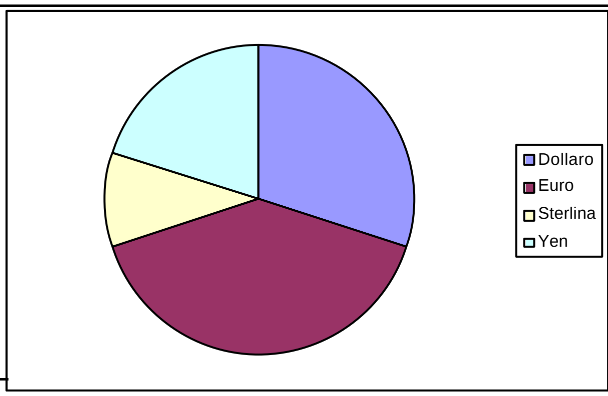
fondo	VALUTA	VALORE	REND	TITOLI	RISCHIO	VALORE^2	REND^2	TITOLI^2
1	Euro	85	0,02	90	Basso	7225	0,0004	8100
2	Euro	90	0,06	200	Medio	8100	0,0036	40000
3	Dollaro	110	0,13	150	Medio	12100	0,0169	22500
4	Euro	105	0,11	160	Basso	11025	0,0121	25600
5	Sterlina	120	0,03	60	Medio	14400	0,0009	3600
6	Dollaro	95	0,11	100	Alto	9025	0,0121	10000
7	Yen	130	0,23	50	Medio	16900	0,0529	2500
8	Euro	102	0,08	220	Basso	10404	0,0064	48400
9	Dollaro	110	0,07	180	Alto	12100	0,0049	32400
10	Yen	160	0,31	120	Alto	25600	0,0961	14400
somma		**1107**	**1,15**	**1330**		**126879**	**0,2063**	**207500**

1) Calcolate il campo di variazione e il coefficiente di variazione del carattere TITOLI.

Campo di variazione (TITOLI) = max– min = 220-50 = **170**
Coefficiente di variazione (TITOLI) = dev.std. /|media| = 55,3263/133 = **0,4160**

2) Fornite un'opportuna rappresentazione grafica del carattere VALUTA.





3) Il rendimento di un fondo dipende dal suo rischio? Rispondete calcolando la funzione di regressione di RENDIMENTO su RISCHIO.

M (RENDIMENTO | RISCHIO = basso) = 0,07
M (RENDIMENTO | RISCHIO = medio) = 0,1125
M (RENDIMENTO | RISCHIO = alto) = 0,1633
Mediamente, il rendimento aumenta all'aumentare del grado di rischio. La funzione di regressione indica quindi una dipendenza.

4) Se la funzione di regressione fosse costante, RENDIMENTO e RISCHIO sarebbero indipendenti statisticamente? Motivate la risposta.

No. Se la funzione di regressione fosse costante, allora le due variabili presenterebbero una indipendenza regressiva, che NON implica l'indipendenza statistica.

5) Fornite la definizione di primo quartile e calcolatelo per il carattere VALORE.

Il primo quartile è il valore rispetto al quale (almeno approssimativamente) il 25% delle osservazioni sono più piccole e il 75% sono più grandi.
$Fr\{X \leq q_1\} \geq 0{,}25$ e $Fr\{X \geq q_1\} \geq 0{,}75$

Ordinando in modo crescente i valori del carattere VALORE e considerando la posizione (N+1)/4≈3 (o alternativamente, considerando il valore che per primo raggiunge o supera la frequenza cumulata del 25%), si ottiene:

primo quartile(VALORE) = **95**

ESERCIZIO 2 (3 punti)

Una variabile continua ha una media di 100 e una varianza pari a 25. Dimostrate che la frequenza con cui la variabile appartiene all'intervallo (80 , 120) è superiore al 90%.

Utilizzando la disuguaglianza di Cebyshev:

$$Fr(m - k < X < m + k) \geq 1 - \frac{s^2}{k^2}$$

dove μ e s^2 sono rispettivamente la media e la varianza di X, e ponendo k = 20, si ottiene:

$$Fr(80 < X < 120) \geq 1 - \frac{25}{20^2} = 0{,}9735$$

da cui risulta che la frequenza con cui X appartiene all'intervallo dato è superiore al 90%.

ESERCIZIO 3 (5 punti)

Un'azienda che vuole produrre componenti elettronici acquista due stabilimenti: A e B. Indichiamo con *a* la quota di produzione del primo stabilimento e con *b* la quota del secondo (naturalmente, *a*+*b* =1). Le percentuali di componenti difettosi prodotti sono del 4% per A e del 2% per B.

1) Enunciate il teorema di Bayes.

Se C_i, con i =1,…,n, è una collezione di eventi che costituiscono una partizione di Ω, cioè tali che:
$C_i \cap C_j = \varnothing$, per ogni i ≠ j
$C_1 \cup C_2 \cup \ldots \cup C_n = \Omega$
e se D è un evento tale che P(D)>0, allora:

$$P(C_i \mid D) = \frac{P(D \mid C_i)P(C_i)}{\sum_{i=1}^{n} P(D \mid C_i)P(C_i)}$$

2) Determinate i valori di a e b in modo tale che la probabilità che un componente provenga da A, nell'ipotesi che sia difettoso, sia dello 0,75.

Definiamo gli eventi D="componente difettoso", "A=componente proviene da A", "B=componente proviene da B". Utilizzando il teorema di Bayes, la probabilità che un componente provenga da A, nell'ipotesi che sia difettoso, è:

$$P(A \mid D) = \frac{P(D \mid A)P(A)}{P(D \mid A)P(A) + P(D \mid B)P(B)} = \frac{0.04P(A)}{0.04P(A) + 0.02[1 - P(A)]}$$

Impostando l'equazione uguale a 0.75, si ottiene:

P(A) = **0.6**

P(B) = 1-0.6 = **0.4**

ESERCIZIO 4 (4 punti)

Si consideri il seguente stimatore:

$$T_n = \frac{2X_1 + (n-2)X_n}{n}$$

1) Definite in generale cosa si intende per stimatore consistente in media quadratica.

Uno stimatore T_n è **consistente** (in media quadratica) per un parametro $q \in \Theta$ se:

$$\lim_{n\to\infty} E[(T_n - q)^2] = 0, \forall q \in \Theta$$

Equivalentemente, condizione necessaria e sufficiente affinché T_n sia uno stimatore **consistente** (in media quadratica) per q è che:

$$\lim_{n\to\infty} E(T_n) = q, \qquad \forall q \in \Theta$$
$$\lim_{n\to\infty} VAR(T_n) = 0, \qquad \forall q \in \Theta$$

2) Sapendo che E(X) = VAR(X) = q, stabilite se lo stimatore T_n sia o meno consistente in media quadratica per q.

Il valore atteso e la varianza dello stimatore sono rispettivamente:

$$E(T_n) = E\left(\frac{2X_1 + (n-2)X_n}{n}\right) = \frac{2E(X_1) + (n-2)E(X_n)}{n} = \frac{2q + (n-2)q}{n} = \frac{n}{n}q = q$$

$$VAR(T_n) = VAR\left(\frac{2X_1 + (n-2)X_n}{n}\right) = \frac{4VAR(X_1) + (n-2)^2 VAR(X_n)}{n^2} = \frac{4q + (n-2)^2 q}{n^2} = \frac{n^2 - 4n + 8}{n^2}q$$

La condizione necessaria e sufficiente per la consistenza NON è interamente verificata, infatti:

$$\lim_{n\to\infty} VAR(T_n) = q \neq 0$$

Quindi lo stimatore **NON è consistente** in media quadratica.

ESERCIZIO 5 (6punti)

Sia X una variabile casuale distribuita secondo una normale N(m, s^2). Sulla base di un campione di ampiezza n = 25, si osservano i seguenti risultati:

$$\sum_{i=1}^{25} x_i = 200; \qquad \sum_{i=1}^{25} x_i^2 = 6250$$

1) Indicate uno stimatore corretto per la media m e calcolatene la stima.

Uno stimatore corretto per la media m è la media campionaria $\overline{X}$, la cui stima è $\bar{x} = \frac{200}{25} = 8$.

2) Indicate uno stimatore corretto per la varianza s^2 e calcolatene la stima.

Uno stimatore corretto per la varianza s^2 è la varianza campionaria corretta S_C^2, la cui stima è

$$s_C^2 = \frac{n}{n-1}\left(\frac{\sum x^2}{n} - \bar{x}^2\right) = 193.75.$$

3) Determinate l'intervallo di confidenza per m con a = 0,1.

$$ic_{1-a}(m) = \left(\bar{x} \pm t_{1-\frac{a}{2}}^{(n-1)}\sqrt{\frac{s_c^2}{n}}\right) = \left(8 \pm t_{0.95}^{(24)}\sqrt{\frac{193.75}{25}}\right) = \left(8 \pm 1.711\sqrt{\frac{193.75}{25}}\right) = (3.2368, 12.7632)$$

4) Costruite la regione di rifiuto per un test di dimensione a = 0,05 e decidete tra le seguenti ipotesi:

H_0: $m \geq 11$ $\qquad$ H_1: $m < 11$

$$R = \left((x_1, \ldots, x_n) : \bar{x} < m_0 - t_{1-a}^{(n-1)}\sqrt{\frac{s_c^2}{n}}\right) = \left((x_1, \ldots, x_n) : \bar{x} < 11 - t_{0.95}^{(24)}\sqrt{\frac{193.75}{25}}\right) =$$

$$= \left((x_1, \ldots, x_n) : \bar{x} < 6.2368\right)$$

Siccome la media campionaria vale 8, ACCETTO l'ipotesi nulla.

ESERCIZIO 6 (2 punti)

Prima di lanciare una nuova rivista sul mercato, una casa editrice decide di fare un test per verificare quante copie (in media) della rivista potrebbero essere vendute per ciascuna edicola.
La casa editrice stabilisce che occorrono in media 20 copie ad edicola perché il prodotto possa generare profitto.
Stabilite le ipotesi H_0 e H_1 del test in questione e giustificate la vostra scelta.

Definendo con m la media delle copie vendute in edicola, le ipotesi del test sono:

H_0 : $m \leq 20$ $\qquad$ H_1: $m > 20$

Infatti, l'ipotesi nulla corrisponde a quella che porta alle conseguenze più gravi se rifiutata quando vera. Nel contesto dell'esercizio, possiamo ritenere più grave per la casa editrice intraprendere il lancio della rivista e quindi sostenere dei costi per un investimento fallimentare, piuttosto che non intraprendere un investimento che avrebbe portato benefici.