Università degli studi di Ferrara
Corso di Laurea Magistrale in Matematica – Facoltà di Scienze
Prova di Statistica – 03 Luglio 2012

Problema 1

Una popolazione di quattro unità presenta i seguenti valori del carattere X:

5 9 11 15

a) Costruire la distribuzione dello stimatore "media aritmetica campionaria" relativa ai campioni di ampiezza 2 estratti casualmente e con ripetizione.

Lo spazio campionario è dato da tutte le 16 coppie ordinate (con ripetizione) che si possono ottenere con i numeri 5,9,11,15. Ciascun campione ha la stessa probabilità di essere estratto. Quindi si ottiene:

Campioni	Media Campionaria	Probabilità
(5,5)	5	1/16
(5,9)	7	1/16
(5,11)	8	1/16
(5,15)	10	1/16
(9,5)	7	1/16
(9,9)	9	1/16
(9,11)	10	1/16
(9,15)	12	1/16
(11,5)	8	1/16
(11,9)	10	1/16
(11,11)	11	1/16
(11,15)	13	1/16
(15,5)	10	1/16
(15,9)	12	1/16
(15,11)	13	1/16
(15,15)	15	1/16
		1

Sintetizzando i valori che può assumere lo stimatore "media campionaria" e le probabilità corrispondenti si ottiene:

$$\overline{X} = \begin{cases} 5 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 15 \\ 1/16 & 2/16 & 2/16 & 1/16 & 4/16 & 1/16 & 2/16 & 2/16 & 1/16 \end{cases}$$

b) Verificare che tale stimatore è corretto per la media della popolazione.

Il valore della vera media della popolazione di riferimento è 10

Si calcola il valore atteso della media campionaria e si ottiene:

$$E(\overline{X}) = \sum_{i=1}^{9} \overline{x}_i P(X = x_i) = 10$$

Lo stimatore è non distorto per la media della popolazione.

c) Verificare che relazione sussiste tra la varianza del carattere X e la varianza dello stimatore proposto.

Si deve verificare che sussiste la relazione:

$$Var(\overline{X}) = \sigma^2 / n$$

Dove \sigma^{2} indica la varianza della popolazione di riferimento e con n l'ampiezza campionaria.

La varianza dello stimatore proposto è data da:

$$Var(\overline{X}) = E(\overline{X}^2) - [E(\overline{X})]^2 = 6.5$$

La varianza della popolazione è pari a:

$$Var(X) = \frac{1}{n}\sum x_i^{\ 2} - \mu^2 = 13$$

La dimensione campionaria è pari a due quindi la relazione

$$Var(\overline{X}) = \sigma^2 / n$$

È verificata.

Problema 2

Disponendo di un campione di 13 individui di *Heterocypris incongruens* pescati in un fiume, dei quali sono riportate le lunghezze (in mm),

Individui	1	2	3	4	5	6	7	8	9	10	11	12	13
Lunghezze	1.21	1.39	1.21	1.21	1.21	1.21	1.20	1.18	1.23	1.21	1.23	1.24	1.33

a) Calcolare i cinque numeri di sintesi, costruire il diagramma scatola a baffi e commentare opportunamente circa variabilità forma e posizionamento.

Min=1.18
Max=1,39
Q1=1.21
Q3=1.235
Me=1.21

b) Verificare se alla probabilità P = 0.99 la loro lunghezza media è significativamente differente dalla media di 1,25 mm stimata per la stessa specie nei laghi della regione, in varie ricerche precedenti.(livello di significatività 1%).

Dai 13 dati campionari, devono essere calcolati il valore della media e della deviazione standard:

$$\overline{X} = 1{,}235 \quad S = 0{,}059 \quad n = 13$$

La domanda dell'esempio richiede un test a due code o bilaterale, poiché prima della raccolta dei dati è ugualmente logico che la media del campione abbia un valore sia significativamente minore sia maggiore della media attesa. Indicando con μ la media reale del campione estratto dal fiume e con μ_0 la media della popolazione che vive nei laghi, l'ipotesi nulla è $H_0 : \mu = \mu_0$ e l'ipotesi alternativa $H_1 : \mu \neq \mu_0$

Mediante il test **t**

$$t_{(12)} = \frac{1{,}235 - 1{,}250}{\frac{0{,}059}{\sqrt{13}}} = \frac{-0{,}015}{0{,}01636} = -0{,}917$$

si ottiene un valore di t(12) uguale a -0,917.

Alla probabilità $\alpha = 0.01$ per un test bilaterale con 12 gdl il valore critico riportato è uguale a 3,055. Il valore calcolato in modulo è nettamente inferiore a quello corrispondente riportato nella tavola sinottica; di conseguenza, non si è in grado di rifiutare l'ipotesi nulla.

La dimensione media dei 13 individui della specie *Heterocypris incongruens* pescati nel fiume non è significativamente diversa da quella degli individui della stessa specie che vivono nei laghi della regione.

c) Risolvere il punto b) costruendo un intervallo di confidenza con l'1% di significatività.

Per la stima dell'intervallo di confidenza, dopo il calcolo dei medesimi parametri si deve scegliere il valore del t con 12 gdl

- alla probabilità $\alpha = 0.01$ per un test a due code oppure
- alla probabilità $\alpha = 0.005$ per un test a una coda

trovando in entrambi i casi t0.005; 12 = 3,055.

I valori del limite inferiore l_1 e del limite superiore l_2 dell'intervallo fiduciale

$$\mu = 1{,}235 \pm 3{,}055\frac{0{,}059}{\sqrt{12}} = 1{,}235 \pm 0{,}05203$$

Sono rispettivamente

$$l_1 = 1{,}175 \text{ e } l_2 = 1{,}287.$$

La media della popolazione μ0 uguale a 1,25 è compresa nell'intervallo fiduciale della media campionaria. Pertanto, non esiste una differenza significativa alla probabilità prefissata di α = 0.01 in un test
bilaterale.

d) Commentare i risultati ottenuti.

Problema 3

Sia *(X₁,…,Xₙ)* un campione casuale estratto da una popolazione avente funzione di ripartizione:

$$F(x;\vartheta) = \begin{cases} 1 - e^{-\frac{\sqrt{x}}{\vartheta}} & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases}$$

dove ϑ è un parametro > 0.

a) Determinare lo stimatore di ϑ mediante metodo di massima verosimiglianza e verificarne la correttezza.

La funzione di densità è pari a:

$$f(x;\vartheta) = \begin{cases} (e^{-\frac{\sqrt{x}}{\vartheta}})/2\vartheta\sqrt{x} & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases}$$

Si consideri un campione casuale di ampiezza *n*, conseguentemente la funzione di verosimiglianza sarà limitandosi all'ortante pari a :

$$L(\vartheta;\underline{x}) = (1/2\vartheta)^n (x_1 \cdot \ldots \cdot x_n)^{-1/2} e^{-\frac{\sum_i \sqrt{x_i}}{\vartheta}}$$

Passando al logaritmo:

$$\log L(\vartheta;\underline{x}) = -n\log(2\vartheta) - \frac{\sum_i \sqrt{x_i}}{\vartheta} - (1/2)\log(x_1 \cdot \ldots \cdot x_n)$$

Derivando rispetto al parametro incognito si ottiene:

$$V'_n = \log L(\vartheta;\underline{x}) = -(n/\vartheta) + \frac{\sum_i \sqrt{x_i}}{\vartheta^2}$$

Si verifica che ϑ è estremante se e solo se:

$$\vartheta = \frac{\sum_i \sqrt{x_i}}{n} \text{ quindi lo stimatore è } \hat{\vartheta} = \frac{\sum_i \sqrt{X_i}}{n}$$

Si vuole ora verificare che lo stimatore è corretto.

$$E(\hat{\vartheta}) = E\left(\frac{\sum_i \sqrt{X_i}}{n}\right) = \frac{1}{n}E\left(\sum_i \sqrt{X_i}\right) = \frac{1}{n}E\left(\sum_i \sqrt{X_i}\right) = \frac{1}{n}nE(\sqrt{X}) = \int_0^{+\infty} \sqrt{x} f_X(x;\vartheta)dx$$

$$\int_0^{+\infty} \sqrt{x} f_X(x;\vartheta)dx = \int_0^{+\infty} (1/2\vartheta)e^{-\frac{\sqrt{x}}{\vartheta}} dx$$

Risolvendo per sostituzione ($\frac{\sqrt{x}}{\vartheta}$=t) si ottiene $E(\hat{\vartheta}) = \vartheta$ quindi lo stimatore ML ottenuto è corretto.

Problema 4
Il test chi-quadro per l’indipendenza e la bontà di adattamento: discutere le caratteristiche principali e fornire un esempio.