Statistica I - A.A. 2016-2017

Prova scritta - 17 febbraio 2017

Problema 1. (pt 21=4+4+4+5+4) Questo esercizio tratta della grandezza aleatoria L = livello di gradimento di un certo prodotto e descrive ciò che fa un esaminatore che svolge indagini di mercato su L; nella realtà una delle difficoltà sta nel fatto che le risposte numeriche degli intervistati, su L, possono non essere confrontabili; qui per semplicità supponiamo che lo siano, ad esempio grazie al fatto che sia stata fissata una scala, tipo 10=... Quindi trattiamo L come una qualsiasi grandezza misurabile, aleatoria. L'esaminatore, per semplicità, supporrà che L sia gaussiana. Siano θ e σ la sua media teorica e deviazione standard teorica, ovviamente sconosciute ed eventualmente variabili a seconda del periodo storico o del prodotto esaminato. [Nella risoluzione, in particolare nella scrittura delle ipotesi dei test, si usino i simboli ora introdotti, ovvero L, θ, σ ecc., non simboli generici non interpretabili da parte del docente.]

1. L'esaminatore, nel 2015, svolge una prima indagine di mercato: fa provare il prodotto a 100 persone e chiede a ciascuna di esse di indicare il proprio valore di L, ovvero il proprio livello di gradimento. Le 100 persone forniscono le loro valutazioni numeriche $l_1, ..., l_{100}$. L'esaminatore calcola $\bar{l} = 32.6$, $s^2 = (12.1)^2$. L'errore relativo $\left|\frac{32.6-\theta}{32.6}\right|$ al massimo quanto può essere grande, escludendo il 5% dei casi più sfavorevoli? Il risultato, come si può osservare, non è esaltante, considerato che la numerosità del campione è elevata; abbassando il livello di confidenza a 0.5 migliora (verificare); ma che significato intuitivo ha, secondo voi, in termini di frequenza relativa, la frase <<a livello 50% l'errore relativo è...>>?

2. Quanto detto al punto 1 fu solo una prima indagine; seguirono poi, nel 2015 e 2016, indagini più numerose ed accurate, con migliaia di intervistati, con le quali si arrivò a stabilire con ragionevole certezza che la media θ del gradimento era 35.2, per quel prodotto, con deviazione $\sigma = 10.5$. Un collega chiede all'intervistatore: <<ogni 1000 persone, quante ti aspetti che abbiano un livello di gradimento maggiore di 50?>>. Cosa rispondereste, con un calcolo veloce? Alla fine dei calcoli, rispondete proprio letteralmente, nella forma <<ogni 1000 persone ci aspettiamo che ... abbiano un livello di gradimento maggiore di 50>>.

3. L'esaminatore immagina di interrogare 1000 nuove persone ed ottenere i valori $L_1, ..., L_{1000}$. Ogni volta che un intervistato dichiarerà più di 50, l'intervistatore registrerà un "successo". L'intervistatore, prima di fare l'indagine, vuole stimare la probabilità di avere più di 90 successi. Come fa?

4. A fine 2016 l'azienda ha scoperto una variante del prodotto che potrebbe piacere di più. Per farsi un'idea preliminare, nel gennaio 2017 la fa provare

a 100 persone e chiede loro il livello di gradimento, ottenendo una media empirica pari a 43.5. Può affermare che c'è un miglioramento, al 95%? Risolvere il problema sia in modo unilaterale che bilaterale, confrontando i risultati. Specificare bene le ipotesi e le regioni di rifiuto, nei due casi.

5. Si scoprirà più avanti nel tempo, dopo molte indagini, che il vero livello medio di gradimento del nuovo prodotto è 39.1. Scrivere e poi calcolare la probabilità che avava il test unilaterale del punto precedente di accorgersi di un tale cambiamento. Risolvere l'eseercizio in modo letterale, non traducendo il problema ed usando formule preconfezionate.

Problema 2. (pt 9) Si chiama triangolare di parametro $A \in (0,1)$, su $[0,1]$, una densità di probabilità definita come segue: è nulla fuori da $[0,1]$ e vale

$$f_A(x) = \begin{cases} C_1 x & \text{per } x \in [0,A] \\ C_2(1-x) & \text{per } x \in [A,1] \end{cases}$$

con $C_1, C_2 > 0$ scelti in modo che: i) f_A sia continua in $x = A$, ii) f_A sia una densità di probabilità.

1. Tracciare il grafico di f_A e calcolare C_1, C_2 in funzione di A.

2. Calcolare valore atteso e varianza di una v.a. X con densità f_A, osservando a livello grafico la posizione di $E[X]$ rispetto ad A.

3. Dette X, Y, Z tre variabili aleatorie indipendenti con densità f_A, calcolare $E[XY^2 + 3Z]$, spiegando a cosa è dovuto ogni passaggio.

1 Soluzioni

Esercizio 1.

1. $$\left|\frac{32.6-\theta}{32.6}\right| \leq \frac{12.1\cdot q_{0.975}}{32.6\sqrt{100}} = \frac{12.1\cdot 1.96}{32.6\cdot 10} = 0.07.$$
   Abbassando il livello di confidenza a 0.5 si trova
   $$\left|\frac{32.6-\theta}{32.6}\right| \leq \frac{12.1\cdot q_{0.75}}{32.6\sqrt{100}} = \frac{12.1\cdot 0.67}{32.6\cdot 10} = 0.025.$$
   Però l'interpretazione è: se ripetiamo un certo numero di volte il campionamento, la metà delle volte la stima dell'errore trovata è falsa.

2. Traduciamo ogni 1000 persone, quante con che probabilità c'è di e poi moltiplichiamo per 1000. Vale allora
   $$P^{(35.2,10.5)}(L>50) = P^{(35.2,10.5)}\left(\frac{L-35.2}{10.5} > \frac{50-35.2}{10.5}\right) = P(Z>1.4095)$$
   $$= 1-\Phi(1.41) = 1-0.9207 = 0.0793.$$
   Pertanto, risponderemmo circa 79 su 1000.

3. Introduciamo, per $i=1,...,1000$, $X_i = 1$ se $L_i > 50$. Le v.a. X_i sono Bernoulli indipendenti, con $p = P(L>50) = 0.0793$. La domanda chiede
   $$P(X_1+...+X_{1000} > 90) = P\left(\frac{X_1+...+X_{1000}-1000\cdot p}{\sqrt{1000}\sqrt{p(1-p)}} > \frac{90-1000\cdot p}{\sqrt{1000}\sqrt{p(1-p)}}\right)$$
   $$\overset{TLC}{\sim} P(Z>1.2522) = 1-\Phi(1.25) = 1-0.8943 = 0.1057.$$

4. Soluzione unilaterale: $\mathcal{H}_0)\ \theta \leq 35.2$, $\mathcal{H}_1)\ \theta > 35.2$, regione di rifiuto
   $$\left\{\overline{X} > 35.2 + \frac{10.5\cdot q_{0.95}}{\sqrt{100}}\right\} = \left\{\overline{X} > 36.922\right\}$$
   $$\frac{10.5\cdot q_{0.95}}{\sqrt{100}} = \frac{10.5\cdot 1.64}{10} = 1.722$$
   quindi il test rifiuta l'ipotesi nulla, c'è stato un cambiamento. Soluzione bilaterale: $\mathcal{H}_0)\ \theta = 35.2$, $\mathcal{H}_1)\ \theta \neq 35.2$, regione di rifiuto
   $$\left\{\left|\overline{X}-35.2\right| > \frac{10.5\cdot q_{0.975}}{\sqrt{100}}\right\} = \left\{\left|\overline{X}-35.2\right| > 2.058\right\}$$
   $$\frac{10.5\cdot q_{0.975}}{\sqrt{100}} = \frac{10.5\cdot 1.96}{10} = 2.058$$
   ed anche in questo caso il test rifiuta l'ipotesi nulla. Il valore p unilaterale, se lo si calcola, è più basso, migliore.

5.

$$P^{(39.1,10.5)}\left(\overline{X} > 36.922\right) = P^{(39.1,10.5)}\left(\frac{\overline{X}-39.1}{10.5}\sqrt{100} > \frac{36.922-39.1}{10.5}\sqrt{100}\right)$$
$$= P\left(Z > -2.0743\right) = \Phi\left(2.07\right) = 0.9807.$$

Esercizio 2.

1. I due segmenti di retta devono avere la stessa altezza in $x = A$; inoltre si conosce la formula per l'area di un triangolo. Usando questi due fatti si ha immediatamente

$$C_1 A = C_2\left(1-A\right)$$
$$\frac{C_1 A}{2} = 1$$

da cui $C_2\left(1-A\right) = 2$,

$$C_1 = \frac{2}{A}$$
$$C_2 = \frac{2}{1-A}$$

2.

$$E\left[X\right] = \int_0^A C_1 x^2 dx + \int_A^1 C_2\left(1-x\right) x dx$$
$$= C_1\frac{A^3}{3} + C_2\left(\frac{1}{2}-\frac{A^2}{2}\right) - C_2\left(\frac{1}{3}-\frac{A^3}{3}\right)$$

$$E\left[X^2\right] = \int_0^A C_1 x^3 dx + \int_A^1 C_2\left(1-x\right) x^2 dx = ...$$
$$Var\left[X\right] = E\left[X^2\right] - E\left[X\right]^2 = ...$$

3. Per linearità
$$E\left[XY^2 + 3Z\right] = E\left[XY^2\right] + 3E\left[Z\right]$$

poi per indipendenza tra X ed Y^2

$$= E\left[X\right]E\left[Y^2\right] + 3E\left[Z\right]$$

dove poi vanno sostituiti i risultati del punto precedente.