Statistica I. Ingegneria Gestionale. Scritto del 15/09/2012

Cerchiare, su questo foglio, le risposte corrette e risolvere per esteso gli esercizi sui fogli assegnati.

Esercizio 1. La ditta CARS si occupa del noleggio di autovetture. Negli anni 2010-2011 noleggiava 228.5 giorni-auto al mese, in media, con una deviazione standard pari a 47.2. Dopo una prima parte del 2012 vuole fare confronti e aggiornare le sue statistiche.

i) (3 pt) Per capire se il noleggio medio di giorni-auto ha subito modifiche, la ditta utilizzerà i valori relativi ad n mesi del 2012 ed eseguirà: A = un test bilaterale; B = un test unilaterale destro; C = un test unilaterale sinistro:

$$A, \qquad B, \qquad C$$

Motivare la scelta tra A, B e C, descrivere il test e dire su che base potrebbe scegliere il valore di n.

ii) (4 pt) Scelto $n = 8$, presi i valori dei primi 8 mesi del 2012, riscontra una media aritmetica pari a $\overline{x} = 264.8$. Dire com'è definito il p-value di questo test e, partendo dalla definizione data, calcolare quanto vale (senza usare formule prestabilite o saltare passaggi a mente).

0.0216, 0.0546, 0.0296, 0.0469, 0.0666.

iii) (4 pt) Se il valore medio vero delle vendite fosse aumentato a 250, eseguendo il test precedente al 90%, con che probabilità ce ne accorgeremmo? Scrivere matematicamente la probabilità richiesta e, partendo da lì, ricavare il risultato, usando solamente le formule riportate sotto.

0.3624, 0.4624, 0.2624, 0.1624, 0.5624.

iv) (3 pt) Quanto era precisa la stima della media, 228.5, ottenuta negli anni 2010-2011, al 95%?

8.884, 22.884, 12.884, 28.884, 18.884.

v) (4 pt) Con quella stima (si ignorino i nuovi valori del 2012), quanti giorni-macchina riesce a noleggiare mensilmente, come minimo, al 90%? Spiegare il calcolo tramite un disegno.

68.08, 168.08, 123.08, 118.08, 23.08.

vi) (4 pt) Ogni macchina ha probabilità 0.01 di subire un lieve incidente, durante un generico giorno di noleggio. Che probabilità c'è che ci sia almeno un lieve incidente nell'arco di 20 giorni-auto di noleggio? Impostare il problema usando appropriate variabili aleatorie.

0.0821, 0.2221, 0.1821, 0.3121, 0.2821.

Esercizio 2. Nel seguito, X è gaussiana $N(1,4)$, Y è Bernoulli $B(1,0.2)$. Si suppongano X e Y indipendenti.

i) (2 pt) Calcolare $P(-1 < X < 3)$

0.5826, 0.6826, 0.1832, 0.2832, 0.3832.

ii) (3 pt) Calcolare $P(Y < X)$

0.6532, 0.1532, 0.3532, 0.5234, 0.2234.

iii) (3 pt) Calcolare $E\left[XY^2e^Y\right]$.

0.1522, 0.3522, 0.4522, 0.5437, 0.2437.

Quantili utili per questo compito:

$x =$	-2.9334	-0.5	0.3566	1	1.28	1.645	1.96	2.1753
$\Phi(x) =$	0.0017	0.3085	0.6393	0.8413	0.9	0.95	0.975	0.9852

Formule e teoremi potenzialmente utili. 1) Se X è $B(n,p)$, allora $P(N=k) = \binom{n}{k}p^k(1-p)^{n-k}$, $E[X] = np$, $Var[X] = np(1-p)$; se X è $\mathcal{P}(\lambda)$, allora $P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$, $E[X] = Var[X] = \lambda$; se X è esponenziale di parametro λ, $F(x) = 1-e^{-\lambda x}$, $E[X] = \sigma[X] = \frac{1}{\lambda}$. 2) Se $X_1,...,X_n$ sono $N(\mu,\sigma^2)$, allora $P\left(\left|\overline{X}-\mu\right| \leq \frac{\sigma q_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right) = 1-\alpha$. 3) Se $X_1,...,X_n$ sono $N(\mu,\sigma^2)$, allora $Z = \frac{\overline{X}-\mu_0}{\sigma}\sqrt{n}$ è una $N(d,1)$ con $d = \frac{\mu-\mu_0}{\sigma}\sqrt{n}$.

1 Soluzioni

Esercizio 1. i) A, in quanto non ha pregiudizi circa la possibile variazione della media, in negativo o positivo. L'ipotesi nulla è che la media sia 228.5. Quando avrà i dati $x_1, ..., x_n$, scelta la significatività α, dovrà calcolare $\overline{x}$, $z = \frac{\overline{x}-228.5}{47.2}\sqrt{n}$, $q_{1-\frac{\alpha}{2}}$ e vedere se $|z| < q_{1-\frac{\alpha}{2}}$ (in tal caso rifiuterà l'ipotesi nulla, cioè affermerà che la media è cambiata). Oppure calcolerà il valore p (punto (ii)). Specificando una certa potenza relativamente ad un valore della media μ che gli sembri critico può trovare n che permette di avere tale potenza.

ii) Il p-value è la probabilità che l'indicatore $Z = \frac{\overline{X}-228.5}{47.2}\sqrt{8}$, sotto l'ipotesi nulla (secondo tale ipotesi Z è $N(0,1)$), assuma valori più estremi (nel senso bilaterale) del valore sperimentale z,

$$z = \frac{264.8 - 228.5}{47.2}\sqrt{8} = 2.1753$$

quindi

$$\begin{aligned} p &:= P(|Z| > 2.1753) = 1 - P(|Z| < 2.1753) \\ &= 1 - P(-2.1753 < Z < 2.1753) \\ &= 1 - \Phi(2.1753) + \Phi(-2.1753) \\ &= 2 - 2 \cdot \Phi(2.1753) = 2 - 2 \cdot 0.9852 = 0.029\,6. \end{aligned}$$

iii) Dobbiamo calcolare $P_{\mu=250}(|Z| > q_{0.95})$ dove $Z = \frac{\overline{X}-228.5}{47.2}\sqrt{8}$. Risulta

$$Z \sim N(d, 1), \qquad d = \frac{250 - 228.5}{47.2}\sqrt{8} = 1.\,2884$$

in quanto la media vera ora è 250, $\overline{X}$ proviene da una $N(250, 47.2)$. Vale

$$\begin{aligned} P(|Z| > q_{0.95}) &= 1 - P(|Z| < q_{0.95}) = 1 - P(-q_{0.95} < Z < q_{0.95}) \\ &= 1 - \Phi(q_{0.95} - d) + \Phi(-q_{0.95} - d) \\ &= 1 - \Phi(1.645 - 1.\,2884) + \Phi(-1.645 - 1.\,2884) \\ &= 1 - \Phi(0.356\,6) + \Phi(-2.\,933\,4) \\ &= 1 - 0.6393 + 0.0017 = 0.3624. \end{aligned}$$

iv) Per la teoria dell'intervallo di confidenza, la precisione δ è data da

$$\delta = \frac{\sigma q_{1-\frac{\alpha}{2}}}{\sqrt{n}}$$

che approssimiamo a ($\alpha = 0.05$, $\frac{\alpha}{2} = 0.025$, $1 - \frac{\alpha}{2} = 0.975$)

$$\delta = \frac{47.2 \cdot 1.96}{\sqrt{24}} = 18.884.$$

v) Si deve tracciare una gaussiana X di media 228.5 e deviazione 47.2, e cercare il valore λ per cui $P(X > \lambda) = 0.9$. Il punto λ è quindi alla sinistra di

228.5. Tracciando anche la gaussiana canonica per confronto, in cui l'analogo di λ sarebbe il quantile $q_{0.1} = -q_{0,9} = -1.28$, si riconosce che

$$\lambda = 228.5 - 47.2 \cdot 1.28 = 168.08.$$

vi) Indichiamo con $X_1, ..., X_{20}$ delle v.a. di Bernoulli tali che $X_i = 1$ se l'i-esimo giorno-auto c'è un lieve incidente, cosa che accade con probabilità 0.01. La somma $S = X_1 + ... + X_{20}$ è una $B(20, 0.01)$ e rappresenta il numero di giorni-auto con lievi incidenti. Dobbiamo calcolare

$$P(S \geq 1) = 1 - P(S = 0) = 1 - \binom{20}{0} 0.01^0 0.99^{20} = 0.1821.$$

Esercizio 2. i)

$$\begin{aligned} P(-1 < X < 3) &= \Phi\left(\frac{3-1}{2}\right) - \Phi\left(\frac{-1-1}{2}\right) = \Phi(1) - \Phi(-1) \\ &= 2 \cdot \Phi(1) - 1 = 2 \cdot 0.8413 - 1 = 0.6826. \end{aligned}$$

ii)

$$\begin{aligned} P(Y < X) &= P(Y < X | Y = 0) P(Y = 0) + P(Y < X | Y = 1) P(Y = 1) \\ &= 0.8 \cdot P(X > 0) + 0.2 \cdot P(X > 1) \\ &= 0.8 \cdot \left(1 - \Phi\left(-\frac{1}{2}\right)\right) + 0.2 \cdot \left(1 - \Phi\left(\frac{1-1}{2}\right)\right) \\ &= 0.8 \cdot (1 - 0.3085) + 0.2 \cdot (1 - 0.5) = 0.6532. \end{aligned}$$

iii)

$$\begin{aligned} E\left[XY^2e^Y\right] &= E[X] E\left[Y^2e^Y\right] = 1 \cdot \left(0 \cdot e^0 \cdot P(Y = 0) + 1 \cdot e^1 \cdot P(Y = 1)\right) \\ &= e \cdot 0.2 = 0.5437. \end{aligned}$$