Statistica I - A.A. 2016-2017

Prova scritta - 11 settembre 2017

Problema 1. (pt 21) Una piccola azienda di pregio costruisce pezzi su misura a richiesta, impiegando generalmente numerose ore lavorative. Sia N il numero di ore lavorative necessarie per completare un pezzo; consideriamo N aleatorio perché varia a caso da un pezzo all'altro, e per semplicità lo consideriamo gaussiano.

1. Un dirigente ha bisogno di dichiarare quante ore servono per completare un pezzo. Raccoglie i dati $n_1, ..., n_{20}$ di 20 pezzi trovando il valor medio empirico di $\overline{n} = 38.5$ ore lavorative, con deviazione standard empirica $s = 12.5$ ore. La dichiarazione più neutra che può fare è dire che in media servono 38.5 ore. Se però vuole dichiarare un valore n^* avente la seguente proprietà: il 90% delle volte bastano meno di n^* ore, come calcola n^*?

2. Il valore $\overline{n} = 38.5$ non è ovviamente la media vera μ di N. Quanto dista $\overline{n}$ da μ? Ovviamente non c'è risposta al problema formulato così. Spiegate voi al dirigente quale sia un problema ben formulato, circa la distanza tra $\overline{n}$ e μ e risolvetelo numericamente.

3. Un nuovo operaio lavora molto velocemente, così ad occhio. Si osserva che gli bastano in media $\overline{n}_{new} = 32$ ore con una deviazione $s = 12.5$ ore; questi valori sono emersi osservando 15 sue lavorazioni. Se prendiamo il valore 38.5 come se fosse la media vera delle lavorazioni degli altri operai, possiamo affermare, al 90%, che il nuovo operaio è più veloce? Dichiarare con chiarezza le ipotesi e la regione di rifiuto. Applicare un test unilaterale.

4. Ripetere il test in modo bilaterale usando le gaussiane invece che le t di Student e calcolare anche il valore p, partendo da una definizione di valore p ed illustrando il calcolo con un disegno.

5. Un cliente chiede 40 pezzi. Supponiamo che vengano lavorati secondo i valori medi del punto 1, in sequenza da uno stesso operaio. Che probabilità c'è che servano più di $40 \cdot 40$ ore?

Problema 2. (pt 9)

1. Siano X, Y due gaussiane indipendenti, entrambe con media uguale ad 1. Se $Var[X+Y] = 5$ e $E[X^2] = 2$, quanto valgono $Var[X]$ e $Var[Y]$?

2. Cosa accade se invece di supporre $Var[X+Y] = 5$ si suppone $Var[Y-X] = 5$?

3. Calcolare (nel caso del punto 1) $P(X+Y > 4)$.

1 Soluzioni

Esercizio 1.

1. Cerchiamo n^* tale che $P(N < n^*) = 0.9$; quindi

$$\begin{aligned} 0.9 &= P\left(\frac{N-38.5}{12.5} < \frac{n^*-38.5}{12.5}\right) = \Phi\left(\frac{n^*-38.5}{12.5}\right) \\ \frac{n^*-38.5}{12.5} &= q_{0.9} = 1.28 \\ n^* &= 38.5 + 1.28 \cdot 12.5 = 54.5. \end{aligned}$$

2. Una formulazione corretta consiste nel dichiarare una confidenza e cercare l'errore (o l'intervallo) relativo a quella confidenza. Ad esempio, se fissiamo 95%, ovvero $\alpha = 0.05$, $1 - \frac{\alpha}{2} = 0.975$, $q_{0.975} = 1.96$, l'errore massimo è

$$\frac{\sigma q_{0.975}}{\sqrt{n}} = \frac{12.5 \cdot 1.96}{\sqrt{20}} = 5.478.$$

   Un risultato più corretto si ottiene con la t di Student: $t_{19,0.025} = 2.093$

$$\frac{s t_{19,0.025}}{\sqrt{n}} = \frac{12.5 \cdot 2.093}{\sqrt{20}} = 5.850.$$

3. L'ipotesi nulla è H_0) media ≥ 38.5, ipotesi alternativa H_1) media < 38.5, regione di rifiuto: $\left\{(n_1, ..., n_{15}) : \frac{\overline{n}-38.5}{s}\sqrt{n} < -t_{14,0.1}\right\}$. Vale

$$\frac{\overline{n}-38.5}{s}\sqrt{n} = \frac{32-38.5}{12.5}\sqrt{15} = -2.014$$

   mentre $-t_{14,0.1} = -1.345$, quindi si rifiuta l'ipotesi nulla.

4. Ipotesi nulla H_0) media $= 38.5$, ipotesi alternativa H_1) media $\neq 38.5$, regione di rifiuto: $\left\{(n_1, ..., n_{15}) : \left|\frac{\overline{n}-38.5}{s}\sqrt{n}\right| > q_{0.95}\right\}$,

$$\left|\frac{\overline{n}-38.5}{s}\sqrt{n}\right| = \left|\frac{32-38.5}{12.5}\sqrt{15}\right| = 2.014$$

   mentre $q_{0.95} = 1.64$, quindi si rifiuta l'ipotesi nulla. Il valore p è

$$\begin{aligned} P\left(\left|\frac{\overline{N}-38.5}{12.5}\sqrt{15}\right| > \left|\frac{32-38.5}{12.5}\sqrt{15}\right|\right) &= P(|Z| > 2.014) = 2P(Z > 2.014) \\ &= 2(1 - \Phi(2.014)) = 2(1 - 0.978) \\ &= 0.044. \end{aligned}$$

5. Chiamiamo $N_1, ..., N_{40}$ i tempi di lavorazione dei 40 pezzi, che supponiamo indipendenti e $N(38.5, 12.5^2)$. Interessa calcolare

$$P(N_1 + ... + N_{40} > 40 \cdot 40) = P\left(\frac{N_1 + ... + N_{40} - 40 \cdot 38.5}{\sqrt{40} \cdot 12.5} > \frac{40 \cdot 40 - 40 \cdot 38.5}{\sqrt{40} \cdot 12.5}\right)$$
$$= P(Z > 0.76) = 1 - \Phi(0.76) = 1 - 0.776 = 0.224$$

dove abbiamo usato il fatto che le combinazioni lineari di gaussiane indipendenti sono gaussiane.

Esercizio 2.

1. Per l'indipendenza, $Var[X + Y] = Var[X] + Var[Y]$, quindi $Var[X] + Var[Y] = 5$. Inoltre, $Var[X] = E[X^2] - E[X]^2 = 2 - 1 = 1$, quindi infine $Var[Y] = 5 - 1 = 4$.

2. Non cambia nulla. Infatti, Y e $-X$ sono indipendenti, quindi

$$Var[Y - X] = Var[Y] + Var[-X] = Var[Y] + Var[X] = Var[X + Y].$$

3. La v.a. $X + Y$ è gaussiana (perché somma di gaussiane indipendenti), ha media 2 e varianza 5. Quindi

$$P(X + Y > 4) = P\left(\frac{X + Y - 2}{\sqrt{5}} > \frac{4 - 2}{\sqrt{5}}\right) = P(Z > 0.894)$$
$$= 1 - \Phi(0.894) = 1 - 0.8133 = 0.1867.$$