Domande (capziose e artificiali) per verificare la comprensione del significato di p-valore.

N.B. Spesso le domande contengono informazioni irrilevanti.

Quesito 1. Ripetiamo 2 volte lo stesso T-test a due code con campioni di dimensione $n = 25$. Assumendo vera H_0, qual è la probabilità che in almeno uno di questi test il p-valore risulti ≤ 0.05 ?

Si scelga tra le seguenti opzioni la più opportuna.

1. La probabilità è $= \ldots$ (specificare)
2. La probabilità è $< \ldots$ (specificare)
3. La probabilità è $> \ldots$ (specificare)
4. Non ci sono sufficienti informazioni per stimare questa probabilità.

Risposta 1. La probabilità è $= 1 - (0.95)^2 = 0.0975$.

Quesito 2. Abbiamo fatto un T-test a due code con un campione di dimensione $n = 25$ e abbiamo ottenuto come p-valore 0.04. Assumendo vera H_0, qual è la probabilità che, ripetendo il test una seconda volta con un campione di dimensione doppia, il p-valore risulti ≤ 0.08 ?

Si scelga tra le seguenti opzioni la più opportuna.

1. La probabilità è $= \ldots$ (specificare)
2. La probabilità è $< \ldots$ (specificare)
3. La probabilità è $> \ldots$ (specificare)
4. Non ci sono sufficienti informazioni per stimare questa probabilità.

Risposta 1. La probabilità è $= 1 - 0.08 = 0.92$.

Quesito 3. Abbiamo fatto un T-test coda superiore con un campione di dimensione $n = 25$ e abbiamo ottenuto come p-valore 0.02. Assumendo vera H_A, qual è la probabilità che ripetendo il test una seconda volta con un campione della stessa dimensione il p-valore risulti di nuovo ≤ 0.02 ?

Nel caso non sia possibile determinare il valore esatto ma solo un limite superiore/inferiore. Si scelga tra le seguenti opzioni la più opportuna.

1. La probabilità è $= \ldots$ (specificare)
2. La probabilità è $< \ldots$ (specificare)
3. La probabilità è $> \ldots$ (specificare)
4. Non ci sono sufficienti informazioni per stimare questa probabilità.

Risposta 4. Non ci sono sufficienti informazioni per stimare questa probabilità.

Quesito 4. Preleviamo un campione di rango $n = 25$ da una popolazione con distribuzione $N(\mu, \sigma^2)$. Sappiamo che la deviazione standard è $\sigma = 5$. La media μ invece potrebbe avere uno qualsiasi dei tre valori 1, 2, o 4.

Vogliamo testare $H_0 : \mu = 2$ contro $H_A : \mu \in \{1, 4\}$.

1. Che test facciamo?
2. Se la media del campione di cui sopra è $\bar{x} = 3$, quant'è il p-valore?
3. Dato questo risultato, la probabilità di $\mu \in \{1, 4\}$ è ... (si scelga la risposta corretta ta le seguenti)

   (a) = p-valore; (b) = 1− p-valore; (c) 2/3; (d) Non ci sono sufficienti informazioni per dirlo.

Esprimere il risutato numerico tramite (solo) le funzioni elencate in calce.

Risposta

Facciamo uno z-test a due code. Risposta 1

$$\Pr\left(|\bar{X}| \geq 3\right) = \Pr\left(\left|\frac{\bar{X}-2}{\sigma/\sqrt{n}}\right| \geq \frac{|3-2|}{\sigma/\sqrt{n}}\right) = \Pr\left(|Z| \geq 1\right)$$

= `2 * norm.cdf(-1)` = `0.317` Risposta 2

(d) Non ci sono sufficienti informazioni per dirlo. Risposta 3

Quesito 5. Preleviamo un campione di rango $n = 9$ da una popolazione con distribuzione $N(\mu, \sigma^2)$. Sappiamo che la deviazione standard è $\sigma = 2$. La media μ invece potrebbe avere uno qualsiasi dei tre valori 2, 6, o 7.

Vogliamo testare $H_0 : \mu = 2$ contro $H_A : \mu \in \{6, 7\}$.

1. Che test facciamo?
2. Se la media del campione di cui sopra è $\bar{x} = 3$, quant'è il p-valore?

Esprimere il risutato numerico tramite (solo) le funzioni elencate in calce.

Risposta

Facciamo uno z-test coda superiore. Risposta 1

$$\Pr\left(\bar{X} \geq 3\right) = \Pr\left(\frac{\bar{X}-2}{\sigma/\sqrt{n}} \geq \frac{3-2}{\sigma/\sqrt{n}}\right) = \Pr\left(Z \geq 3/2\right)$$

= `1 - norm.cdf(3/2)` = `0.134` Risposta 2

Si assuma noto il valore delle seguenti funzioni della libreria `scipy.stats`

`norm.cdf(z)` $= \Pr\left(Z < z\right)$ per $Z \sim N(0, 1)$

`norm.ppf(`α`)` $= z_\alpha$ dove z_α è tale che $\Pr\left(Z < z_\alpha\right) = \alpha$ per $Z \sim N(0, 1)$