

Domande (qualcuna capziosa e artificiale) per verificare la comprensione del significato di p-valore (ed implicitamente anche del FWER).

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  $\leq 0.05$  ?

Si scelga tra le seguenti opzioni la più opportuna.

1. La probabilità è  $= \dots$  (specificare)
2. La probabilità è  $< \dots$  (specificare)
3. La probabilità è  $> \dots$  (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 0.05. 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  $\geq 0.1$  ?

Si scelga tra le seguenti opzioni la più opportuna.

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

**Risposta** 1. La probabilità è  $= 1 - 0.1 = 0.9$ .

**Quesito 3.** Abbiamo fatto un T-test a due code con un campione di dimensione  $n = 25$  e abbiamo ottenuto come p-valore 0.05. Assumendo vera  $H_A$  con effect size 0.1, qual è la probabilità che ripetendo il test una seconda volta con un campione della stessa dimensione il p-valore risulti di nuovo  $\leq 0.05$  ?

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

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

**Quesito 4.** Abbiamo fatto un T-test coda inferiore con un campione di dimensione  $n = 25$  e abbiamo ottenuto come p-valore 0.05. Assumendo vera  $H_A$  con effect size 0.1, qual è la probabilità che ripetendo il test una seconda volta con un campione della stessa dimensione il p-valore risulti di nuovo  $\leq 0.05$  ?

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

**Risposta** 3. La probabilità è  $> 0.05$ .

**Quesito 5.** Preleviamo un campione di rango  $n = 16$  da una popolazione con distribuzione  $N(\mu, \sigma^2)$ . Sappiamo che la deviazione standard è  $\sigma = 2$ . La media  $\mu$  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  $\bar{x} = 1$ , quant'è il p-valore?
3. Data questa media campionaria, la probabilità che  $\mu \in \{1, 4\}$  è (si scelga tra le seguenti)  
(a) = p-valore; (b) =  $1 - \text{p-valore}$ ; (c)  $2/3$ ; (d) Non ci sono sufficienti informazioni per rispondere.

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

**Risposta**

Facciamo uno z-test a due code.

Risposta 1

`2 * norm.cdf(-2)`

Risposta 2

(d) Non ci sono sufficienti informazioni per rispondere.

Risposta 3

**Quesito 6.** Preleviamo un campione di rango  $n = 4$  da una popolazione con distribuzione  $N(\mu, \sigma^2)$ . Sappiamo che la deviazione standard è  $\sigma = 3$ . La media  $\mu$  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  $\bar{x} = 4$ , quant'è il p-valore?

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

### Risposta

Facciamo uno z-test coda superiore.

Risposta 1

`1 - norm.cdf(4/3)`

Risposta 2

**Quesito 7.** Assume the null hypothesis is true and denote by  $P$  the random variable that gives the p-value you would get if you run a test.

1. What is the probability that  $\Pr(P < 4)$  ?
2. If we run the tests 4 times (independently), what is the probability of incorrectly rejecting at least once the null hypotheses with a significance  $\alpha = 4\%$  ?
3. If we run the tests 1/4 times (independently), how small do we have to make the cutoff ( $\alpha$  above) to lower to 1/4% the probability of incorrectly rejecting at least once the null hypotheses?

**Risposta**

1/4

Risposta 1

100

Risposta 2

4%

Risposta 3

---

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

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

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