

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.** 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 2.** Ripetiamo 3 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  $= 1 - (0.95)^3 = 0.142625$ .

**Quesito 3.** 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 4.** 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** 3. La probabilita e  $> 0.05$ .

**Quesito 5.** [da rivedere]

Il disturbo  $D$  è causato dalla presenza di almeno uno dei fattori  $F_1, \dots, F_4$ . Ovvero possiamo assumere che  $D = F_1 \cup \dots \cup F_4$ .

Un gruppo di ricercatori, non sospettando nemmeno l'esistenza dei fattori  $F_i$ , ipotizza invece ci sia un associazione tra  $D$  e un altro fattore  $A$ . La prevalenza  $\Pr(D)$  nella popolazione generale è nota e i ricercatori vogliono fare il seguente test di ipotesi

$$H_0 : \Pr(D) = \Pr(D|A) \qquad H_A : \Pr(D) < \Pr(D|A)$$

Quindi viene isolato un campione casuale dimensione 50 di individui con fattore  $A$  e misurano la proporzione  $p_0$  di individui affetti dal disturbo  $D$ . Viene fatto un test binomiale.

Si assuma che  $A$  sia indipendente da  $D$ . Per semplificare assumiamo che  $A, F_1, \dots, F_4$  siano tra loro mutualmente indipendenti e che  $\Pr(F_i) = p$ . Si calcoli la probabilità che venga rifiutata l'ipotesi nulla con una significatività  $\alpha = 2\%$ .

**Quesito 6.** Preleviamo un campione di rango  $n = 9$  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 9. Vogliamo testare  $H_0 : \mu = 6$  contro  $H_A : \mu \in \{2, 9\}$ .

1. Che test facciamo?
2. Se la media del campione di cui sopra è  $\bar{x} = 5$ , quant'è il p-valore?
3. Data questa media campionaria, la probabilità che  $\mu \in \{2, 9\}$  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(-1)`

Risposta 2

(d) Non ci sono sufficienti informazioni per rispondere.

Risposta 3

**Quesito 7.** 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  $\mu$  invece potrebbe avere uno qualsiasi dei due valori 3 o 8. Vogliamo testare  $H_0 : \mu = 3$  contro  $H_A : \mu = 8$ .

1. Che test facciamo?
2. Se la media del campione di cui sopra è  $\bar{x} = 4$ , quant'è il p-valore?
3. Data questa media campionaria, la probabilità che  $\mu = 8$  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 coda superiore.

Risposta 1

`1 - norm.cdf(1)`

Risposta 2

(d) Non ci sono sufficienti informazioni per rispondere.

Risposta 3

**Quesito 8.** 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  $\Pr(P < 0.05)$  ?
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  $\alpha = 5\%$  ?
3. If we run the tests 4 times (independently), how small do we have to make the cutoff ( $\alpha$  above) to lower to 5% the probability of incorrectly rejecting at least once the null hypotheses?

### Risposta

0.05

Risposta 1

0.1855

Risposta 2

1.2741%

Risposta 3

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

`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  $Z \sim N(0, 1)$