Verifica delle Ipotesi 2

Stefania Bartoletti

8 Giugno 2020

Indice

- Confronto tra due medie di di due popolazioni normali
- Confronto tra due proporzioni

Esempio

- Si vuole confrontare il consumo di due automobili di modelli A e B in termini di km/litro. Il numero di km/litro è distribuito come una variabile aleatoria normale con media sconosciuta. Si ottiene tale numero per un campione di n auto per il modello A e m auto per il modello B e si vuole sapere se il consumo è uguale (quindi le due variabili aleatorie hanno la stessa media) o diverso (oppure se una consuma più dell'altra).

Confronto tra due medie di due popolazioni normali

- Studiando gli intervalli di confidenza, abbiamo considerato il caso di due campioni indipendenti estratti da due popolazioni diverse con valori attesi μ_1 e μ_2 e deviazioni standard σ_1 e σ_2.
- Abbiamo ottenuto un intervallo di confidenza bilaterale per la differenza tra le medie delle due popolazioni come

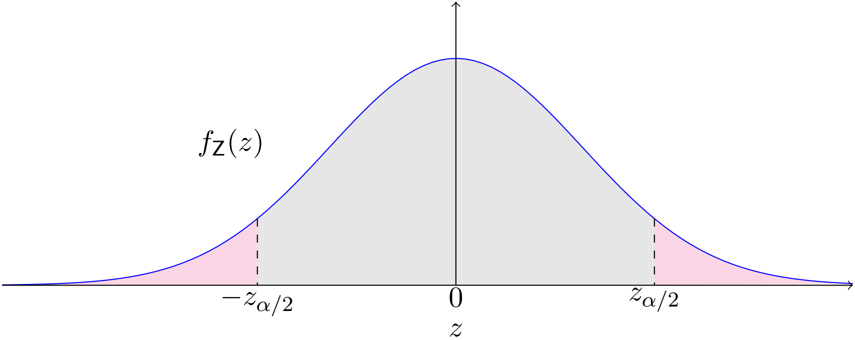
$$\mu_1 - \mu_2 \in \Big(\overline{x} - \overline{y} - z_{\alpha/2}\sqrt{(\sigma_1^2/n + \sigma_2^2/m)},\\ \overline{x} - \overline{y} + z_{\alpha/2}\sqrt{(\sigma_1^2/n + \sigma_2^2/m)}\Big)$$

Confronto tra due medie di due popolazioni normali

- Ricordando il legame tra l'intervallo di confidenza e la verifica di ipotesi, possiamo costruire facilmente un test di ipotesi con significatività α per la differenza tra le medie di due popolazioni normali
- Nel caso in cui σ_1 e σ_2 siano note, consideriamo $H_0 : \mu_1 = \mu_2$ e $H_1 : \mu_1 \neq \mu_2$, o anche $H_0 : \mu_\mathsf{d} = \mu_1 - \mu_2 = 0$ e $H_1 : \mu_\mathsf{d} \neq 0$
- Definite le ipotesi, consideriamo la statistica $\mathsf{Z} = \frac{\overline{\mathsf{X}} - \overline{\mathsf{Y}} - \mu_\mathsf{d}}{\sqrt{(\sigma_1^2/n + \sigma_2^2/m)}} \sim \mathcal{N}(0,1)$
- La statistica Z ha distribuzione normale standard se le popolazioni sono normali o anche nel caso in cui abbiano un'altra distribuzione ma m ed n siano elevati

Confronto tra due medie di due popolazioni normali

- Non ci resta che applicare ciò che abbiamo già visto per il caso di una media, considerando un livello di significatività α

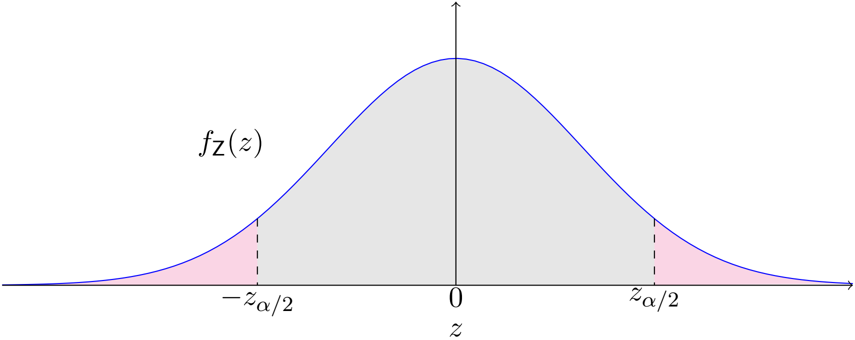


Accetto H_0 se

$$\left| \frac{\overline{x} - \overline{y} - \mu_d}{\sqrt{(\sigma_1^2/n + \sigma_2^2/m)}} \right| \leqslant z_{\alpha/2}$$

Confronto tra due medie di due popolazioni normali

- Non ci resta che applicare ciò che abbiamo già visto per il caso di una media, considerando un livello di significatività α



Accetto H_0 se

$$\mathbb{P}\left\{|\mathsf{Z}| > \frac{\overline{x} - \overline{y} - \mu_{\mathsf{d}}}{\sqrt{(\sigma_1^2/n + \sigma_2^2/m)}}\right\} > \alpha$$

Confronto tra due medie di due popolazioni normali

- Abbiamo anche visto che nel caso in cui le varianze non siano note, ma uguali, l'intervallo di confidenza bilaterale si calcola come

$$\mu_1 - \mu_2 \in \Big(\overline{x} - \overline{y} - t_{n+m-2,\alpha/2}\sqrt{s_{\mathsf{p}}^2(1/n + 1/m)},\\ \overline{x} - \overline{y} + t_{n+m-2,\alpha/2}\sqrt{s_{\mathsf{p}}^2(1/n + 1/m)}\Big)$$

  con $s_{\mathsf{p}}^2 = \frac{(n-1)s_1^2 + (m-1)s_2^2}{n+m-2}$
- Seguendo lo stesso ragionamento, effettueremo il test t, guardando alla statistica $\mathsf{t} = \frac{\overline{\mathsf{X}} - \overline{\mathsf{Y}} - \mu_{\mathsf{d}}}{\sqrt{s_{\mathsf{p}}^2(1/n + 1/m)}} \sim t_{n+m-2}$

Esempio per il caso di popolazioni non indipendenti

- Siamo ancora interessanti al consumo delle automobili
- Questa volta installiamo un dispositivo contro l'inquinamento e ci chiediamo se questo possa influire sui consumi di più automobili
- Un modo per realizzare questo progetto, consiste nel radunare un campione di n auto prive del dispositivo, e provare i consumi di ciascuna prima e dopo l'installazione
- Questa volta ciò che misuriamo come x_i non può essere considerato indipendente da y_i, trattandosi della stessa automobile
- In più possiamo accoppiare le variabili misurate prima dell'installazione con quelle dopo, per ogni automobile (x_i, y_i)

Caso di due popolazioni non indipendenti

- In casi come questi, siamo interessati a confrontare le medie di due popolazioni che però non sono indipendenti. Per esempio, potremmo guardare a due campioni estratti dalla stessa popolazione o varibili che hanno qualche caratteristica in comune
- In questo caso dobbiamo guardare direttamente alla popolazione data dalla differenza tra le variabili con media μ_{d} e varianza σ^2_{d}
- Ancora una volta, distingueremo il caso di varianza nota dal caso di varianza non nota

Caso di due popolazioni non indipendenti

- In alcuni casi siamo interessati a confrontare le medie di due popolazioni che però non sono indipendenti. Per esempio, potremmo guardare a due campioni estratti dalla stessa popolazione o variabili che hanno qualche caratteristica in comune.
- In questo caso supponiamo $n = m$ e dobbiamo guardare direttamente alla popolazione data dalla differenza tra le variabili, i.e.

$$d_i = x_i - y_i, \text{ con } \mathsf{d}_i \sim \mathcal{N}(\mu_{\mathsf{d}}, \sigma^2_{\mathsf{d}})$$

- Ricordiamo che nel caso di variabili dipendenti la varianza non è la somma delle varianze delle popolazioni singole
- Ancora una volta, distingueremo il caso di varianza nota dal caso di varianza non nota

Caso di due popolazioni non indipendenti; varianza nota

- Nel caso in cui σ_{d}^2 sia nota, basterà considerare $H_0 : \mu_{\mathsf{d}} = 0$ e $H_1 : \mu_{\mathsf{d}} \neq 0$
- Definite le ipotesi, consideriamo la statistica

$$\mathsf{Z} = \frac{\bar{d} - \mu_{\mathsf{d}}}{\sigma_{\mathsf{d}}/\sqrt{n}}$$

- e ci baseremo quindi su un test z

Caso di due popolazioni non indipendenti; varianza non nota

- Nel caso in cui σ^2_{d} non sia nota, le ipotesi rimarranno $H_0 : \mu_{\mathsf{d}} = 0$ e $H_1 : \mu_{\mathsf{d}} \neq 0$
- Definite le ipotesi, consideriamo la statistica

$$\mathsf{t} = \frac{\overline{d} - \mu_{\mathsf{d}}}{s_{\mathsf{d}}/\sqrt{n}} \sim t_{n-1}$$

dove $s^2_{\mathsf{d}} = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \overline{d})^2$

- e ci baseremo quindi su un test t
- Si calcolino, per esercizio, i rispettivi intervalli di confidenza

Esempio di confronto tra due proporzioni

- Immaginiamo di volere confrontare due diversi metodi di fabbricazione per un prodotto.
- Indichiamo con p_1 e p_2 le probabilità (incognite) che un pezzo prodotto con i metodi 1 e 2 sia difettoso, rispettivamente
- raccogliamo due campioni di numerosità n e m e indichiamo con x e y il numero di pezzi difettosi trovati. In questo modo x e y sono variabili aleatorie binomiali indipendenti, con parametri (n, p_1) e (m, p_2), rispettivamente.

Confronto tra due proporzioni

- In questi casi, siamo interessati al confronto tra due popolazioni guardando la proporzione di casi con una certa caratteristica
- Consideriamo di avere due campioni con elementi x_i e y_i, di ampiezza n e m dalle due popolazioni, dove la singola variabile è descritta come una Bernoulli con parametri p_1 e p_2, rispettivamente
- Definiamo un test delle ipotesi con $H_0 : p_1 = p_2$ e $H_1 : p_1 \neq p_2$ (ovviamente si procede analogamente per la derivazione per il caso unilaterale)

Confronto tra due proporzioni

- Anche in questo caso seguiremo l'approccio precedentemente utilizzato per la definizione di un intervallo di confidenza (che avevamo visto per una singola popolazione)
- Definiamo un test delle ipotesi con $H_0 : p_1 = p_2$ e $H_1 : p_1 \neq p_2$ (ovviamente si procede analogamente per la derivazione per il caso unilaterale)
- Approssimiamo la distribuzione con una normale, andando a definire la statistica $\mathsf{Z} = \frac{\hat{\mathsf{p}}_1 - \hat{\mathsf{p}}_2}{\sqrt{\overline{\mathsf{p}}(1-\overline{\mathsf{p}})(1/n+1/m)}}$ con $\hat{\mathsf{p}}_1 = \overline{x}$, $\hat{\mathsf{p}}_2 = \overline{y}$, e $\overline{\mathsf{p}} = \frac{n\overline{x}+m\overline{y}}{n+m}$.
- Sotto H_0, $\mathsf{Z} \sim \mathcal{N}(0,1)$ e procediamo quindi con un test z

Schema per la verifica di ipotesi