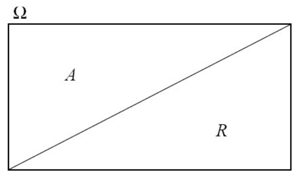
Verifica delle ipotesi

- Un'*ipotesi statistica* è un'affermazione o una congettura riguardante un parametro θ che caratterizza il modello descrittivo della popolazione, *f(x;θ)*, con $\theta \in \Theta$, dove Θ è lo *spazio parametrico*.
- Solitamente, oltre a un'*ipotesi nulla*,
$$H_0 : \theta \in \Theta_0 \subset \Theta$$
si formula un'*ipotesi alternativa* (contrapposta a quella nulla),
$$H_1 : \theta \in \overline{\Theta}_0 \subset \Theta$$
- Un'ipotesi si dice *semplice* se specifica completamente la distribuzione, cioè $\theta = \theta_0$; altrimenti si dice *composta*.

- La ***verifica delle ipotesi*** (***test statistico***) è un procedimento con il quale si decide, alla luce di un campione, se rifiutare o meno H_0 in favore di H_1. Si tratta di dividere lo spazio campionario (Ω) in due sottoinsiemi: *zona di accettazione* (A) e *zona di rifiuto* (R).



Esempio

- Una ditta produce pacchi di pasta che hanno un contenuto medio di 500. Si vuole verificare, sulla base di un campione casuale di 50 pacchi, se il peso medio dei pacchi prodotti è effettivamente pari a 500 g., come dichiarato dalle specifiche del prodotto.
- Sotto l'assunzione che il peso dei pacchi abbia distribuzione $X \sim N(\mu, \sigma^2)$, l'*ipotesi nulla* (che in questo caso è *semplice*) è
$$H_0 : \mu = 500$$
Immaginiamo che ad effettuare la verifica di ipotesi sia un'associazione di consumatori, che risulta danneggiata se il peso è minore di quanto dichiarato. L'*ipotesi alternativa* (*composta*) è quindi
$$H_1 : \mu < 500$$

- Un modo naturale di condurre la *verifica delle ipotesi* è rifiutare H_0 in favore di $H_1 : \mu < 500$ quando la media campionaria è significativamente più piccola della media dichiarata, cioè quando
$$\bar{x} \leq 500 - \varepsilon, \qquad \varepsilon > 0$$
La zona di accettazione e la zona di rifiuto sono pari a:
$$A = \{(x_1, \cdots, x_n) : \bar{x} > 500 - \varepsilon\}, \qquad R = \{(x_1, \cdots, x_n) : \bar{x} \leq 500 - \varepsilon\}$$
- Se l'*ipotesi alternativa* (*composta*) è $H_1 : \mu \neq 500$, è sensato rifiutare se
$$|\bar{x} - 500| \geq \varepsilon, \qquad \varepsilon > 0$$
La zona di accettazione e la zona di rifiuto, in questo caso sono:
$$A = \{(x_1, \cdots, x_n) : |\bar{x} - 500| < \varepsilon\}, \qquad R = \{(x_1, \cdots, x_n) : |\bar{x} - 500| \geq \varepsilon\}$$
- Se l'*ipotesi alternativa* (*composta*) è $H_1 : \mu > 500$, è sensato rifiutare se
$$\bar{x} \geq 500 + \varepsilon, \qquad \varepsilon > 0$$
La zona di accettazione e la zona di rifiuto, in questo caso sono:
$$A = \{(x_1, \cdots, x_n) : \bar{x} < 500 + \varepsilon\}, \qquad R = \{(x_1, \cdots, x_n) : \bar{x} \geq 500 + \varepsilon\}$$

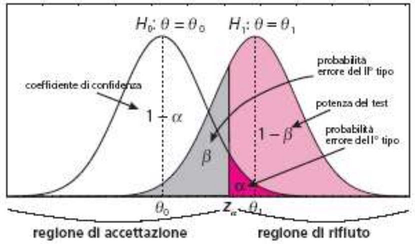
Criteri di valutazione di un test statistico

- Nel condurre un test possiamo incorrere in due tipi di errore:
  - *errore di I tipo*: si rifiuta H_0 quando è vera
  - *errore di II tipo*: si accetta H_0 quando è falsa
- Questi errori possono essere commessi con *probabilità*:
  - $\alpha(\theta) = P_\theta(\mathbf{X} \in R), \quad \theta \in \Theta_0$
  - $\beta(\theta) = P_\theta(\mathbf{X} \in A), \quad \theta \in \overline{\Theta}_0$
- La *potenza del test* è la probabilità di rifiutare H_0 quando è falsa, $\pi(\theta) = 1 - \beta(\theta) = P_\theta(\mathbf{X} \in R), \quad \theta \in \overline{\Theta}_0$

	Decisione	
	Accetto H_0	**Rifiuto** H_0
H_0 è vera	Corretta $1-\alpha$	Errore del I tipo α
H_0 è falsa	Errore del II tipo β	Corretta $1-\beta$

- La potenza *cresce* al crescere della dimensione campionaria e della distanza tra il valore vero del parametro e l'ipotesi nulla.

I diversi errori che si possono commettere:



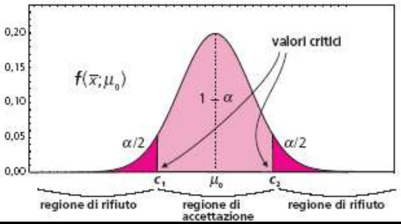
Tra α e β sussiste una relazione inversa: minore è il valore di α, maggiore è il valore di β. Le probabilità di commettere gli errori corrispondono a delle aree.

Esempio

Supponiamo che la popolazione sia Normale con media μ incognita e varianza σ^2 nota. Si vuole verificare:

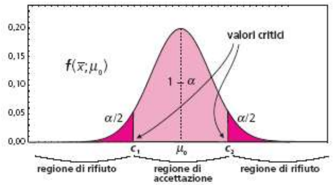
$$\begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu \neq \mu_0 \end{cases}$$

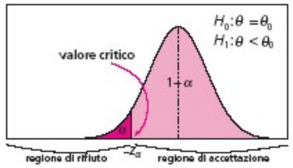
Considerando come statistica test la media campionaria $\overline{X}$ sappiamo che sotto l'ipotesi nulla questa si distribuisce come una Normale con media $\mu = \mu_0$ e varianza σ^2/n.

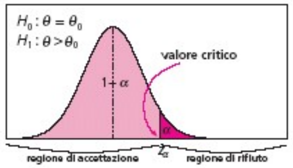


In corrispondenza dell'ipotesi alternativa si possono configurare diverse regioni di rifiuto:

$$\begin{cases} H_0: & \theta = \theta_0 \\ H_1: & \theta \neq \theta_0 \end{cases}$$







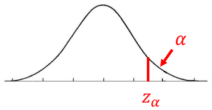
Verifica di ipotesi sulla media
(popolazione normale, varianza nota)

- Si assuma che $X \sim N(\mu, \sigma^2)$, con σ^2 noto.
- Si consideri l'ipotesi nulla H_0: $\mu = \mu_0$ e l'ipotesi alternativa H_1: $\mu > \mu_0$
- Vogliamo determinare la soglia c in modo che la probabilità di rifiutare l'ipotesi nulla quando questa è vera sia pari ad α:

$$P(\bar{X} \geq c | H_0) = \alpha$$

- Standardizzando si ottiene

$$P\left(\frac{\bar{X}-\mu_0}{\sqrt{\sigma^2/n}} \geq \frac{c-\mu_0}{\sqrt{\sigma^2/n}} | H_0\right) = P\left(\frac{\bar{X}-\mu_0}{\sqrt{\sigma^2/n}} \geq z_\alpha | H_0\right) = \alpha$$



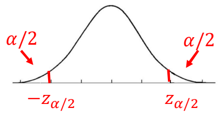
- Si può quindi utilizzare la *statistica test* $Z = \frac{\bar{X}-\mu_0}{\sqrt{\sigma^2/n}}$ che sotto H_0 si distribuisce come una $N(0;1)$
- Se ***H_1: $\mu > \mu_0$, si rifiuta H_0* se $z \geq z_\alpha$** oppure se $\bar{x} \geq \mu_0 + z_\alpha\sqrt{\sigma^2/n}$ dove z_α è il quantile della distribuzione Normale standard tale che *$P(Z \geq z_\alpha) = \alpha$ con $Z \sim N(0,1)$*.

- Se l'ipotesi alternativa fosse invece ***H_1: $\mu < \mu_0$***, dovremmo determinare la soglia c in modo che la probabilità di rifiutare l'ipotesi nulla quando questa è vera sia pari ad α:

$$P(\bar{X} \leq c | H_0) = P\left(\frac{\bar{X}-\mu_0}{\sqrt{\sigma^2/n}} \leq \frac{c-\mu_0}{\sqrt{\sigma^2/n}} | H_0\right) = P\left(\frac{\bar{X}-\mu_0}{\sqrt{\sigma^2/n}} \leq -z_\alpha | H_0\right) = \alpha$$

- Utilizzando la *statistica test* $Z = \frac{\bar{X}-\mu_0}{\sqrt{\sigma^2/n}}$ che sotto H_0 si distribuisce come una $N(0;1)$, ***si rifiuta H_0* se $z \leq -z_\alpha$** oppure se $\bar{x} \leq \mu_0 - z_\alpha\sqrt{\sigma^2/n}$ dove z_α è il quantile della distribuzione Normale standard tale che *$P(Z \leq -z_\alpha) = \alpha$ con $Z \sim N(0,1)$*.
- Se l'ipotesi alternativa fosse invece ***H_1: $\mu \neq \mu_0$***, dovremmo determinare la soglia c in modo che la probabilità di rifiutare l'ipotesi nulla quando questa è vera sia pari ad α:

$$P(|\bar{X}-\mu_0| \geq c | H_0) = P\left(\frac{|\bar{X}-\mu_0|}{\sqrt{\sigma^2/n}} \geq \frac{c}{\sqrt{\sigma^2/n}} | H_0\right)$$
$$= P\left(\frac{\bar{X}-\mu_0}{\sqrt{\sigma^2/n}} \leq -z_{\alpha/2} \text{ oppure } \frac{\bar{X}-\mu_0}{\sqrt{\sigma^2/n}} \geq z_{\alpha/2} | H_0\right) = \alpha$$



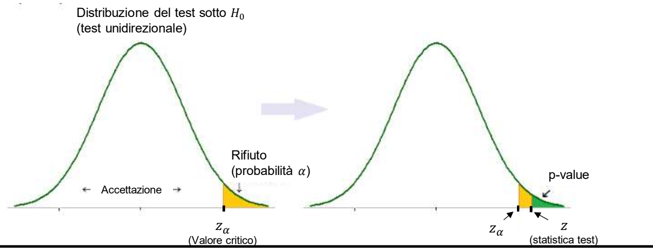
- Utilizzando la *statistica test* $Z = \frac{\bar{X}-\mu_0}{\sqrt{\sigma^2/n}}$ che sotto H_0 si distribuisce come una $N(0;1)$, ***si rifiuta H_0* se $z \leq -z_{\alpha/2}$ o $z \geq z_{\alpha/2}$** oppure se $\bar{x} \leq \mu_0 - z_{\alpha/2}\sqrt{\sigma^2/n}$ o $\bar{x} \geq \mu_0 + z_{\alpha/2}\sqrt{\sigma^2/n}$ dove $z_{\alpha/2}$ è il quantile della distribuzione Normale standard tale che *$P(Z \geq z_{\alpha/2}) = \alpha/2$ con $Z \sim N(0,1)$*.

Un altro modo per evidenziare il risultato del test è quello di calcolare il p-value.

p-value : probabilità di osservare un valore della statistica test uguale o più estremo del valore ottenuto dal campione, sotto l'ipotesi nulla.

E' una quantità che misura l'evidenza fornita dai dati contro l'ipotesi nulla: minore è il valore del p-value, più è forte l'evidenza contro l'ipotesi nulla.

In pratica se p-value < α → rifiuto H_0



Esempio

- Si supponga di voler verificare l'ipotesi che l'*altezza media degli italiani* è 172cm sulla base del campione

170,75 186,14 173,39 185,12 173,39

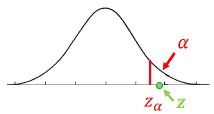
sotto l'assunzione $X \sim N(\mu, \sigma^2)$, con varianza $\sigma^2 = 50$.

- L'ipotesi nulla è H_0: $\mu = 172$ e quindi, dato che $\bar{x} = 177{,}758$, la statistica test Z assume il valore

$$z = \frac{177{,}758 - 172}{\sqrt{50/5}} = \frac{5{,}758}{3{,}162} = 1{,}821$$

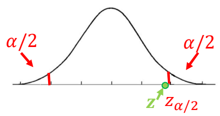
- Se H_1: $\mu > \mu_0$ e $\alpha = 0{,}05$, si rifiuta H_0 in quanto

$$1{,}821 = z \geq z_\alpha = 1{,}645$$



- P-value: $P(Z \geq 1{,}821 | H_0) = 0{,}0344$
- Se H_1: $\mu \neq \mu_0$ e $\alpha = 0{,}05$, non si rifiuta H_0 in quanto

$$1{,}821 = |z| < z_{\alpha/2} = 1{,}96$$



- P-value: $2P(Z \geq 1{,}821 | H_0) = 2 \cdot 0{,}0344$

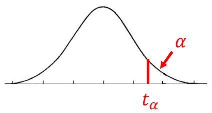
Verifica di ipotesi sulla media
(popolazione normale, varianza non nota)

- Si assuma che $X \sim N(\mu,\sigma^2)$, con σ^2 non noto.
- Si consideri l'ipotesi nulla H_0: $\mu = \mu_0$ e l'ipotesi alternativa H_1: $\mu > \mu_0$
- Vogliamo determinare la soglia c in modo che la probabilità di rifiutare l'ipotesi nulla quando questa è vera sia pari ad α:

$$P(\bar{X} \geq c|H_0) = \alpha$$

- Utilizzando S^2 come stimatore di σ^2 e standardizzando si ottiene

$$P\left(\frac{\bar{X}-\mu_0}{\sqrt{S^2/n}} \geq \frac{c-\mu_0}{\sqrt{S^2/n}}|H_0\right) = P\left(\frac{\bar{X}-\mu_0}{\sqrt{S^2/n}} \geq t_\alpha|H_0\right) = \alpha$$



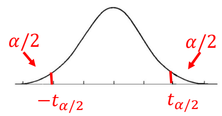
- Si può quindi utilizzare la *statistica test* $T = \frac{\bar{X}-\mu_0}{\sqrt{S^2/n}}$ che sotto H_0 si distribuisce come una t di Student con $n-1$ g.d.l.
- Se ***H_1: $\mu > \mu_0$*, *si rifiuta H_0* se t $\geq t_\alpha$** oppure se $\bar{x} \geq \mu_0 + t_\alpha\sqrt{s^2/n}$ dove t_α è il quantile della distribuzione t di Student con $n-1$ g.d.l. tale che $\mathrm{P}(T_{n-1} \geq t_\alpha) = \alpha$ con $T_{n-1} \sim t_{n-1}$.

- Se l'ipotesi alternativa fosse invece ***H_1: $\mu < \mu_0$*** dovremmo determinare la soglia c in modo che la probabilità di rifiutare l'ipotesi nulla quando questa è vera sia pari ad α:

$$P(\bar{X} \leq c|H_0) = P\left(\frac{\bar{X}-\mu_0}{\sqrt{S^2/n}} \leq \frac{c-\mu_0}{\sqrt{S^2/n}}|H_0\right) = P\left(\frac{\bar{X}-\mu_0}{\sqrt{S^2/n}} \leq -t_\alpha|H_0\right) = \alpha$$

- Utilizzando la *statistica test* $T = \frac{\bar{X}-\mu_0}{\sqrt{S^2/n}}$ che sotto H_0 si distribuisce come una t di Student con $n-1$ g.d.l., ***si rifiuta H_0* se t $\leq -t_\alpha$** oppure se $\bar{x} \leq \mu_0 - t_\alpha\sqrt{s^2/n}$.
- Se l'ipotesi alternativa fosse invece ***H_1: $\mu \neq \mu_0$*** dovremmo determinare la soglia c in modo che la probabilità di rifiutare l'ipotesi nulla quando questa è vera sia pari ad α:

$$P(|\bar{X}-\mu_0| \geq c|H_0) = P\left(\frac{|\bar{X}-\mu_0|}{\sqrt{S^2/n}} \geq \frac{c}{\sqrt{S^2/n}}|H_0\right) =$$
$$P\left(\frac{\bar{X}-\mu_0}{\sqrt{S^2/n}} \leq -t_{\frac{\alpha}{2}} \text{ oppure } \frac{\bar{X}-\mu_0}{\sqrt{S^2/n}} \geq t_{\frac{\alpha}{2}}|H_0\right) = \alpha$$



- Utilizzando la *statistica test* $T = \frac{\bar{X}-\mu_0}{\sqrt{S^2/n}}$ che sotto H_0 si distribuisce come una t di Student con $n-1$ g.d.l., ***si rifiuta H_0* se t $\leq -t_{\alpha/2}$** o **t $\geq t_{\alpha/2}$** oppure se $\bar{x} \leq \mu_0 - t_\alpha\sqrt{s^2/n}$ o $\bar{x} \geq \mu_0 + t_\alpha\sqrt{s^2/n}$.

Esempio

- Nel caso della verifica dell'ipotesi H_0: $\mu = 172$ considerata in precedenza, si supponga di non conoscere la varianza.
- La *media campionaria* e la *varianza campionaria* sono pari a

$$\bar{x} = 177{,}758 \quad e \quad s^2 = 52{,}932$$

da cui la statistica test T assume il valore

$$t = \frac{177{,}758 - 172}{\sqrt{52{,}932/5}} = \frac{5{,}758}{3{,}254} = 1{,}770$$

- Se H_1: $\mu > \mu_0$ e α *=0,01,* non si rifiuta H_0 in quanto

$$1{,}770 = t < t_\alpha = 3{,}747$$

- Se H_1: $\mu \neq \mu_0$ e α *=0,05,* non si rifiuta H_0 in quanto

$$1{,}770 = |t| < t_{\alpha/2} = 2{,}776$$

Verifica di ipotesi sulla media
(popolazione qualsiasi, grandi campioni)

- Per una *popolazione qualsiasi, nel caso di grandi campioni* ($n \geq 30$), se l'ipotesi nulla è H_0: $\mu = \mu_0$ si può utilizzare la statistica test

$$Z = \frac{\bar{X}-\mu_0}{\sqrt{S^2/n}} \sim N(0,1) \quad (sotto\ H_0)$$

- Se H_1: $\mu > \mu_0$, *si rifiuta H_0* se

$$z \geq z_\alpha \quad oppure\ se \quad \bar{x} \geq \mu_0 + z_\alpha S/\sqrt{n}$$

- Se H_1: $\mu < \mu_0$, *si rifiuta H_0* se

$$z \leq -z_\alpha \quad oppure\ se \quad \bar{x} \leq \mu_0 - z_\alpha S/\sqrt{n}$$

- Se H_1: $\mu \neq \mu_0$, *si rifiuta H_0* se

$$|z| \geq z_{\alpha/2} \quad oppure\ se \quad \bar{x} \leq \mu_0 - z_{\alpha/2} S/\sqrt{n} \quad o \quad \bar{x} \geq \mu_0 + z_{\alpha/2} S/\sqrt{n}$$

Esempio

- Si consideri l'ipotesi H_0: $\mu = 1,2$ contro H_1: $\mu < 1,2$
- Si supponga che il campione di n = 50 prodotti che è stato estratto ha media e varianza pari rispettivamente a

$$\bar{x} = 1{,}059 \quad e \quad s^2 = 0{,}158$$

e quindi la *statistica test* assume il valore

$$z = \frac{1{,}059 - 1{,}22}{\sqrt{0{,}158/50}} = -\frac{0{,}141}{0{,}0562} = -2{,}509$$

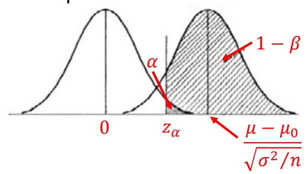
- Al livello di significatività α = 0,01, si rifiuta H_0 in quanto

$$-2{,}509 = z \leq -z_\alpha = -2{,}326$$

Potenza di un test

- Si assuma che $X \sim N(\mu,\sigma^2)$, con σ^2 noto.
- Si consideri l'ipotesi nulla H_0: $\mu = \mu_0$ e l'ipotesi alternativa H_1: $\mu > \mu_0$
- Qual è la distribuzione della statistica test sotto l'ipotesi alternativa?

$$\bar{X} \sim N\left(\mu; \frac{\sigma^2}{n}\right) \rightarrow \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}} \sim N\left(\frac{\mu - \mu_0}{\sqrt{\sigma^2/n}}; 1\right)$$



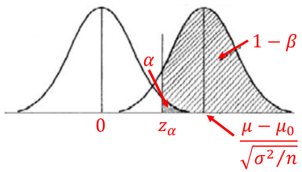
- Qual è la potenza del test?

$$1 - \beta = P\left(\frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}} \geq z_\alpha | H_1\right) = P\left(\frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}} - \frac{\mu - \mu_0}{\sqrt{\sigma^2/n}} \geq z_\alpha - \frac{\mu - \mu_0}{\sqrt{\sigma^2/n}} | H_1\right) =$$

$$= P\left(Z \geq z_\alpha - \frac{\mu - \mu_0}{\sqrt{\sigma^2/n}}\right)$$

- Quindi possiamo determinare la capacità del test di riconoscere un'ipotesi nulla falsa per un particolare valore di μ_1 di μ

- La potenza di un test aumenta al diminuire di σ^2, all'aumentare di n o all'aumentare della distanza tra μ e μ_0 (la distribuzione del test sotto H_1 risulta traslata verso destra)
- La potenza di un test aumenta all'aumentare del livello del test α (la soglia di rifiuto si sposta verso sinistra



Esempio

Un determinato investimento ha un rendimento medio giornaliero dello 0,1% e una volatilità dello 0,05%. Si vuole verificare se questo rendimento sia diminuito nel tempo, ad un livello di significatività dell'1%. Se si utilizzasse un campione di 50 rendimenti giornalieri, quale sarebbe la probabilità di accorgersi di una diminuzione del rendimento medio giornaliero, se questo fosse diventato dello 0,08%? E se si usasse un campione di 100 rendimenti?

$$\begin{cases} H_0:\ \mu = 0{,}001 \\ H_1: \mu < 0{,}001 \end{cases}$$

$\sigma = 0{,}0005$
$n = 50$
$\mu_1 = 0{,}0008$
$\alpha = 0{,}01$

$$1 - \beta = P\left(\frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}} \leq -z_\alpha | H_1\right) = P\left(\frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}} - \frac{\mu_1 - \mu_0}{\sqrt{\sigma^2/n}} \leq -z_\alpha - \frac{\mu_1 - \mu_0}{\sqrt{\sigma^2/n}} | H_1\right) =$$

$$= P\left(Z \leq -z_\alpha - \frac{\mu_1 - \mu_0}{\sqrt{\sigma^2/n}}\right) = P\left(Z \leq -2{,}33 - \frac{0{,}0008 - 0{,}001}{\sqrt{0{,}0005^2/50}}\right) = P(Z \leq 0{,}50) = 0{,}6915$$

$n = 100$

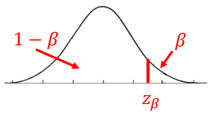
$$1 - \beta = P\left(Z \leq -2{,}33 - \frac{0{,}0008 - 0{,}001}{\sqrt{0{,}0005^2/100}}\right) = P(Z \leq 1{,}67) = 0{,}9525$$

Esempio

Con riferimento all'esempio precedente, che numerosità campionaria avremmo dovuto utilizzare per ottenere una potenza del test del 99%?

$$\begin{cases} H_0: \mu = 0{,}001 \\ H_1: \mu < 0{,}001 \end{cases}$$

$\sigma = 0{,}0005$
$\mu_1 = 0{,}0008$
$\alpha = 0{,}01$
$1-\beta = 0{,}99$
$n = ?$



$$1-\beta = 0{,}99 = P\left(\frac{\bar{X}-\mu_0}{\sqrt{\sigma^2/n}} \le -z_\alpha | H_1\right) = P\left(\frac{\bar{X}-\mu_0}{\sqrt{\sigma^2/n}} - \frac{\mu_1-\mu_0}{\sqrt{\sigma^2/n}} \le -z_\alpha - \frac{\mu_1-\mu_0}{\sqrt{\sigma^2/n}} | H_1\right) =$$

$$= P\left(Z \le -z_\alpha - \frac{\mu_1-\mu_0}{\sqrt{\sigma^2/n}}\right) = P(Z \le z_\beta) \quad \text{da cui}$$

$$-z_\alpha - \frac{\mu_1-\mu_0}{\sqrt{\sigma^2/n}} = z_\beta \quad \rightarrow \quad \frac{\mu_1-\mu_0}{\sqrt{\sigma^2/n}} = -z_\alpha - z_\beta \quad \rightarrow \quad \frac{n(\mu_1-\mu_0)^2}{\sigma^2} = (-z_\alpha - z_\beta)^2$$

$$\rightarrow \quad n = \frac{\sigma^2(z_\alpha+z_\beta)^2}{(\mu_1-\mu_0)^2} = \left[\frac{\sigma(z_\alpha+z_\beta)}{\mu_1-\mu_0}\right]^2 = \left[\frac{0{,}0005\cdot(2{,}33+2{,}33)}{0{,}0008-0{,}001}\right]^2 \approx 136$$

Determinazione della numerosità campionaria

Ipotizziamo una dimensione ampia del campione tale da garantire l'applicazione dell'approssimazione alla Normale.

La procedura segue i seguenti passi:

1. **specificare il livello di significatività α**
2. **specificare il valore di μ_1 e il corrispondente valore di β**
3. **selezionare una stima iniziale di σ**
4. **calcolare la numerosità campionaria**

Sia

z_α il valore per cui $P(Z \ge z_\alpha) = \alpha$

z_β il valore per cui $P(Z < z_\beta) = 1-\beta$

allora:

$$H_1: \ \mu \ne \mu_0 \qquad n = \left[\frac{\sigma(z_{\alpha/2}+z_\beta)}{\mu_1-\mu_0}\right]^2$$

$$\begin{matrix} H_1: \ \mu > \mu_0 \\ H_1: \ \mu < \mu_0 \end{matrix} \qquad n = \left[\frac{\sigma(z_\alpha+z_\beta)}{\mu_1-\mu_0}\right]^2$$

Verifica di ipotesi sulla probabilità di successo

(popolazione Bernoulliana, grandi campioni)

- Si assume che $X \sim Bin(1, \pi)$ (*popolazione Bernoulliana*).
- Si considera l'ipotesi nulla H_0: $\pi = \pi_0$ e quindi la statistica test

$$Z = \frac{\bar{X}-\pi_0}{\sqrt{\pi_0(1-\pi_0)/n}}$$

  che per grandi campioni ($n \ge 30$), ha approssimativamente distribuzione Normale standard $N(0,1)$, sotto H_0.

- Se H_1: $\pi > \pi_0$, *si rifiuta* H_0 se

$$z \ge z_\alpha \quad \textit{oppure se} \quad \bar{x} \ge \pi_0 + z_\alpha\sqrt{\pi_0(1-\pi_0)/n}$$

- Se H_1: $\pi < \pi_0$, *si rifiuta* H_0 se

$$z \le -z_\alpha \quad \textit{oppure se} \quad \bar{x} \ge \pi_0 - z_\alpha\sqrt{\pi_0(1-\pi_0)/n}$$

- Se H_1: $\mu \ne \mu_0$, *si rifiuta* H_0 se

$$|z| \ge z_{\alpha/2} \quad \textit{oppurese} \quad \bar{x} \le \pi_0 - z_{\alpha/2}\sqrt{\pi_0(1-\pi_0)/n} \quad o \quad \bar{x} \ge \pi_0 + z_{\alpha/2}\sqrt{\pi_0(1-\pi_0)/n}$$

Esempio

- Si supponga di voler verificare l'ipotesi H_0: $\pi = 0{,}52$ contro H_1: $\pi > 0{,}52$ al livello $\alpha = 0{,}10$ sulla base di un campione di dimensione $n = 50$ in cui ci sono 32 successi.

- La media campionaria è

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{32}{50} = 0{,}64$$

  e quindi la *statistica test* assume il valore

$$z = \frac{0{,}64-0{,}52}{\sqrt{0{,}52(1-0{,}52)/50}} = \frac{0{,}12}{0{,}0707} = 1{,}698$$

- Si rifiuta H_0 in quanto

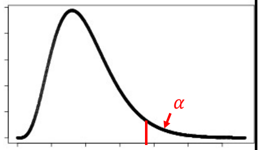
$$1{,}698 = z \ge z_\alpha = 1{,}282$$

Verifica di ipotesi sulla varianza

(popolazione normale, media non nota)

- *X* ~ *N*(μ,σ^2) con μ *non noto*
- Se l'*ipotesi nulla* è H_0: $\sigma^2 = \sigma^2_0$ e l'ipotesi alternativa è H_1: $\sigma^2 > \sigma^2_0$, dobbiamo determinare quella soglia g tale che:
$$\mathrm{P}(S^2 \geq g|H_0) = \alpha$$
- Sapendo che $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$ possiamo scrivere

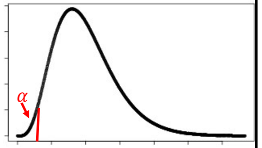
$$\mathrm{P}(S^2 \geq g|H_0) = \mathrm{P}\left(\frac{(n-1)S^2}{\sigma_0^2} \geq \frac{(n-1)g}{\sigma_0^2}|H_0\right) =$$
$$= \mathrm{P}\left(\frac{(n-1)S^2}{\sigma_0^2} \geq \chi^2_\alpha|H_0\right) = \alpha$$



- Si può quindi utilizzare la *statistica test* $C = \frac{(n-1)S^2}{\sigma_0^2}$ che sotto H_0 si distribuisce come un χ^2 con $n-1$ g.d.l.
- Se H_1: $\sigma^2 > \sigma_0^2$, ***si rifiuta** H_0* **se** $\mathbf{c} \geq \chi^2_\alpha$ dove χ^2_α è il quantile della distribuzione χ^2 con $n-1$ g.d.l. tale che $\mathrm{P}(\chi^2 \geq \chi^2_\alpha) = \alpha$.

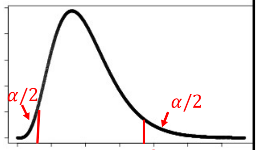
- *Se l'ipotesi alternativa fosse invece* H_1: $\sigma^2 < \sigma^2_0$, dovremmo determinare quella soglia g tale che:

$$\mathrm{P}(S^2 \leq g|H_0) = \mathrm{P}\left(\frac{(n-1)S^2}{\sigma_0^2} \leq \frac{(n-1)g}{\sigma_0^2}|H_0\right) =$$
$$= \mathrm{P}\left(\frac{(n-1)S^2}{\sigma_0^2} \leq \chi^2_{1-\alpha}|H_0\right) = \alpha$$



- Quindi se H_1: $\sigma^2 < \sigma_0^2$, ***si rifiuta** H_0* **se** $\mathbf{c} \leq \chi^2_{1-\alpha}$ dove $\chi^2_{1-\alpha}$ è il quantile del χ^2 con $n-1$ g.d.l. tale che $\mathrm{P}(\chi^2 \leq \chi^2_{1-\alpha}) = \alpha$.
- *Se l'ipotesi alternativa fosse infine* H_1: $\sigma^2 \neq \sigma^2_0$, dovremmo determinare quelle soglie g_1 e g_2 tali che:

$$\mathrm{P}(S^2 \leq g_1 \text{ o } S^2 \geq g_2|H_0) =$$
$$\mathrm{P}\left(\frac{(n-1)S^2}{\sigma_0^2} \leq \frac{(n-1)g_1}{\sigma_0^2} \text{ o } \frac{(n-1)S^2}{\sigma_0^2} \geq \frac{(n-1)g_2}{\sigma_0^2}|H_0\right)$$
$$= \mathrm{P}\left(\frac{(n-1)S^2}{\sigma_0^2} \leq \chi^2_{1-\alpha/2} \text{ o } \frac{(n-1)S^2}{\sigma_0^2} \geq \chi^2_{\alpha/2}|H_0\right) = \alpha$$



- Quindi se H_1: $\sigma^2 \neq \sigma_0^2$, ***si rifiuta** H_0* **se** $\mathbf{c} \leq \chi^2_{1-\alpha/2}$ o $\mathbf{c} \geq \chi^2_{\alpha/2}$.

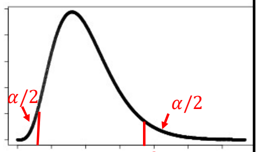
Esempio

- Si supponga di voler verificare l'ipotesi H_0: $\sigma^2 = 75$ contro H_1: $\sigma^2 \neq 75$ al livello α = 0,10 sulla base di un campione (già considerato in precedenza)

  170,75 186,14 173,39 185,12 173,39

  sotto l'assunzione *X* ~ *N*(μ,σ^2), con media μ non nota.

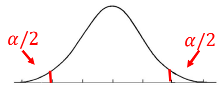
- La varianza campionaria è
$$s^2 = 52{,}932$$
  e quindi la *statistica test* assume il valore
$$c = \frac{4 \cdot 52{,}932}{75} = 2{,}832$$



- Non si rifiuta H_0 in quanto
$$0{,}71 = \chi^2_{1-\alpha/2} \leq c \leq \chi^2_{\alpha/2} = 9{,}49$$

Verifica delle ipotesi e intervalli di confidenza

- Esiste una relazione ben precisa tra verifica delle ipotesi bidirezionale e intervalli di confidenza
- Consideriamo per semplicità la verifica di un'ipotesi sulla media di una popolazione normale con σ^2 noto.
- Utilizzando la *statistica test* $Z = \frac{\bar{X}-\mu_0}{\sqrt{\sigma^2/n}}$ che sotto H_0 si distribuisce come una $N(0;1)$, ***si rifiuta** H_0* **se** $\mathbf{z} \leq -\mathbf{z}_{\alpha/2}$ o $\mathbf{z} \geq \mathbf{z}_{\alpha/2}$.
- La zona di accettazione sarà allora rappresentata da tutti quei campioni per i quali $-z_{\alpha/2} \leq z \leq z_{\alpha/2}$ ossia $-z_{\alpha/2} \leq \frac{\bar{X}-\mu_0}{\sqrt{\sigma^2/n}} \leq z_{\alpha/2}$, che con un po' di passaggi può essere scritta come $\bar{X} - z_{\alpha/2}\sqrt{\sigma^2/n} \leq \mu_0 \leq \bar{X} - z_{\alpha/2}\sqrt{\sigma^2/n}$



- Quindi se l'intervallo di confidenza ottenuto per un determinato campione contiene il valore μ_0, quello stesso campione porterà a non rifiutare il valore μ_0 in una verifica delle ipotesi. Analogamente, un campione per il quale non si rifiuta il valore μ_0, è un campione per il quale l'intervallo di confidenza contiene al suo interno il valore μ_0.

Dove e come studiare

- Libro di testo: S. Borra, A. Di Ciaccio (2014), Cap. 13 (tranne 13.7.1 e 13.8) e Cap. 14 (tranne 14.5 e 14.6)
- Svolgere ‘Esercitazione 8’.
- Svolgere esercizi su foglio di lavoro Excel ‘Esercizi di Inferenza’.
- Altri esercizi di riepilogo sono a disposizione sul sito di Statistica degli anni accademici precedenti (Compiti passati appelli)