

Test d'ipotesi

Test d'ipotesi

Ipotesi statistiche
Ipotesi nulla
Ipotesi nulla
Errori
Scambio di ipotesi
Regione di rifiuto
Potenza
Significatività
Test più potenti
Le tappe
p-value

In molte situazioni una raccolta di dati (=esiti di esperimenti aleatori) viene fatta per prendere delle decisioni sulla base di quei dati. Ad esempio

- sperimentazioni su un nuovo farmaco per decidere se è tossico, se è efficace, se è più efficace di altri in commercio;
- un'associazione di consumatori sostiene che una marca di bottiglie di spumante viene sistematicamente riempita meno del dichiarato e cita perciò in giudizio il produttore. Come farà il giudice ad emettere un verdetto?
- come affermare con sufficiente ragionevolezza che una moneta dà testa con probabilità $= 1/2$ (cioè è equilibrata)?

Test d'ipotesi

Ipotesi statistiche
Ipotesi nulla
Ipotesi nulla
Errori
Scambio di ipotesi
Regione di rifiuto
Potenza
Significatività
Test più potenti
Le tappe
p-value

Una prima idea:

- una prova dell'efficacia potrebbe essere se la maggior parte dei pazienti che testano il farmaco avessero miglioramenti;
- una prova della truffa potrebbe essere se la media del contenuto di un campione di bottiglie risultasse inferiore al contenuto dichiarato;
- una prova che la moneta non sia equilibrata potrebbe essere se il numero di teste fosse diverso dalla metà del numero dei lanci.

Siamo sicuri che sia una buona idea?

Test d'ipotesi

Test d'ipotesi

Ipotesi statistiche
Ipotesi nulla
Ipotesi nulla
Errori
Scambio di ipotesi
Regione di rifiuto
Potenza
Significatività
Test più potenti
Le tappe
p-value

L'idea è da perfezionare, perché:

- **anche con un farmaco inefficace** si possono avere miglioramenti nei pazienti (dovuti ad altre cause non controllabili, dunque aleatorie);
- **anche con un sistema produzione con valore atteso non inferiore al dichiarato** si osservano bottiglie riempite meno del valore atteso (anzi, con il modello normale la metà delle bottiglie ha contenuto inferiore al valore atteso);
- **anche con una moneta equilibrata** si può avere un numero di teste diverso dalla metà del numero dei lanci.

Occorrerà che le “prove” (dell'efficacia, della truffa, del non essere equilibrata...) siano **statisticamente significative**.

Test d'ipotesi

Test d'ipotesi

Ipotesi statistiche
Ipotesi nulla
Ipotesi nulla
Errori
Scambio di ipotesi
Regione di rifiuto
Potenza
Significatività
Test più potenti
Le tappe
p-value

Un test d'ipotesi confronta due *ipotesi statistiche* e prende una decisione fra le due sulla base del campione casuale osservato (come, lo vedremo fra poco).

Si parla di test **statistici** per due motivi:

- 1 la decisione è basata su un campione casuale (anzi, su una *statistica*);
- 2 si fanno dei ragionamenti di probabilità: un'ipotesi viene *rifiutata* se il campione osservato risulta *troppo improbabile* ritenendo vera quella ipotesi.

Ipotesi - test parametrici

Test d'ipotesi

- Ipotesi statistiche
 - Ipotesi nulla
 - Ipotesi nulla
 - Errori
 - Scambio di ipotesi
 - Regione di rifiuto
 - Potenza
 - Significatività
 - Test più potenti
 - Le tappe
 - p-value

Immaginiamo di avere **un preciso modello statistico** per i nostri esperimenti (es: Bernoulli, normale...), **con qualche parametro incognito** (che varia nello spazio dei parametri Θ).

DEFINIZIONE DI IPOTESI STATISTICA

Una **ipotesi statistica** è un'asserzione sul valore vero di un parametro incognito.

Esempio: nel caso della moneta il modello per ogni lancio è $B(p)$ con $p \in [0, 1]$ e un'ipotesi statistica è $p = 1/2$.

Ipotesi nulla e alternativa

DEFINIZIONE DI IPOTESI NULLA E ALTERNATIVA

L'ipotesi nulla, si indica con H_0 e viene ritenuta *vera fino a prova contraria*. H_0 è un'affermazione del tipo: $\theta \in \Theta_0$ dove Θ_0 è un sottoinsieme dello spazio dei parametri Θ .

L'ipotesi alternativa, si indica con H_1 ed è l'affermazione: $\theta \in \Theta_0^c$.

In genere quali devono essere le due ipotesi appare evidente, ma la scelta di quale delle due sia H_0 e quale H_1 va ponderata.

Nella costruzione di un test, si sceglie come H_0 l'ipotesi alla quale si è disposti a rinunciare solo in caso di *forte evidenza del contrario*. Equivalentemente, si sceglie come H_1 l'ipotesi che si giudicherà vera solo in presenza di prove *significative*.

Ipotesi nulla e alternativa

DEFINIZIONE DI IPOTESI NULLA E ALTERNATIVA

L'**ipotesi nulla**, si indica con H_0 e viene ritenuta *vera fino a prova contraria*. H_0 è un'affermazione del tipo: $\theta \in \Theta_0$ dove Θ_0 è un sottoinsieme dello spazio dei parametri Θ .

L'**ipotesi alternativa**, si indica con H_1 ed è l'affermazione:
 $\theta \in \Theta_0^c$.

La definizione è generale.

La definizione di Ipotesi non è limitante; se si cerca un test per $g(\theta)$ e si volesse testare l'ipotesi $g(\theta) \in \tilde{\Theta}_0$, sarebbe equivalente testare $\theta \in \Theta_0 = g^{-1}(\tilde{\Theta}_0) := \{\theta : g(\theta) \in \tilde{\Theta}_0\}$.

Test d'ipotesi

Ipotesi statistiche
Ipotesi nulla
Ipotesi nulla
Errori
Scambio di ipotesi
Regione di rifiuto
Potenza
Significatività
Test più potenti
Le tappe
p-value

Il test d'ipotesi porta a una delle due risposte:

- 1 il test **accetta** H_0 (dunque la si ritiene ragionevolmente vera);
- 2 il test **rifiuta** H_0 (dunque la si ritiene ragionevolmente falsa e si ritiene vera H_1).

Nel prendere queste decisioni possono capitare degli errori.

DEFINIZIONE DI ERRORI DI I E II SPECIE

- H_0 è vera ma il test la rifiuta: **errore di prima specie**;
- H_0 è falsa ma il test la accetta: **errore di seconda specie**.
- Se invece H_0 è vera e il test la accetta oppure se H_0 è falsa e il test la rifiuta si ha la **decisione corretta**.

Come scegliere H_0

Test d'ipotesi

Ipotesi statistiche
Ipotesi nulla
Ipotesi nulla
Errori
Scambio di ipotesi
Regione di rifiuto
Potenza
Significatività
Test più potenti
Le tappe
p-value

Riassumendo:

	H_0 è vera	H_0 è falsa
Rifiutiamo H_0	<i>Errore di I specie</i>	Decisione corretta
Accettiamo H_0	Decisione corretta	<i>Errore di II specie</i>

L'errore di I specie è considerato più grave di quello di II specie. In altre parole noi sceglieremo H_0 ed H_1 affinché l'errore di I specie sia quello che vorremmo evitare di commettere.

Il problema è che non conoscendo il valore vero del parametro **non potremo mai sapere con certezza se stiamo commettendo un errore** ma solo **controllarne la probabilità**.

Esempio per ricordare

Test d'ipotesi

Ipotesi statistiche

Ipotesi nulla

Ipotesi nulla

Errori

Scambio di ipotesi

Regione di rifiuto

Potenza

Significatività

Test più potenti

Le tappe

p-value

Consideriamo il processo ad un imputato.

H_0 : l'imputato è innocente

H_1 : L'imputato è colpevole otterremmo:

	Se è innocente	Se è colpevole
Condannato	<i>Errore di I specie</i>	Decisione corretta
Assolto	Decisione corretta	<i>Errore di II specie</i>

Ritenendo che sia più grave condannare un innocente rispetto a lasciare un colpevole in libertà, dobbiamo scegliere come ipotesi nulla H_0 : *l'imputato è innocente*.

Scambio di ipotesi

Test d'ipotesi

- Ipotesi statistiche
- Ipotesi nulla
- Errori
- Scambio di ipotesi
- Regione di rifiuto
- Potenza
- Significatività
- Test più potenti
- Le tappe
- p-value

Abbiamo già osservato che la scelta di quale ipotesi sia H_0 e quale H_1 non è arbitraria:

H_0 viene ritenuta valida *fino a prova contraria*

H_1 viene ritenuta valida *solo in caso di forte evidenza*.

Quindi se in un test accettate H_0 , non è detto che scambiando i ruoli delle due ipotesi e ripetendo il test (con gli stessi dati), questo nuovo test concluda rifiutando la nuova H_0 (e dunque accettando la vecchia).

Esempio per ricordare

Test d'ipotesi

Ipotesi statistiche

Ipotesi nulla

Ipotesi nulla

Errori

Scambio di ipotesi

Regione di rifiuto

Potenza

Significatività

Test più potenti

Le tappe

p-value

Un esame universitario è in fondo un test statistico: lo studente deve rispondere a varie domande e anche se ha appreso bene può rispondere in maniera errata (emozione, interpretazione errata del problema...), mentre anche se non ha studiato può azzeccare la risposta per puro caso.

Il professore confronta due ipotesi: “lo studente ha appreso bene” e “lo studente non ha appreso bene”.

È ugualmente facile passare l'esame se H_0 : “lo studente ha appreso bene” oppure H_0 : “lo studente non ha appreso bene”?

Esempio per ricordare

Test d'ipotesi

Ipotesi statistiche
Ipotesi nulla
Ipotesi nulla
Errori
Scambio di ipotesi
Regione di rifiuto
Potenza
Significatività
Test più potenti
Le tappe
p-value

Se H_0 : “lo studente ha appreso bene” salvo grandi strafalcioni il professore farà passare l'esame allo studente (fino a prova contraria si rimane dell'opinione H_0).

Se invece H_0 : “lo studente non ha appreso bene” per passare l'esame lo studente dovrà fornire *prove evidenti* di aver appreso il programma (solo con forte evidenza statistica si passa a ritenere vera H_1).

Dunque uno studente mediamente preparato potrebbe passare l'esame nel primo caso e non passarlo nel secondo!

Quando rifiutare H_0

Test d'ipotesi

Ipotesi statistiche

Ipotesi nulla

Ipotesi nulla

Errori

Scambio di ipotesi

Regione di rifiuto

Potenza

Significatività

Test più potenti

Le tappe

p-value

Il test d'ipotesi opera in questo modo:

- 1 si sceglie **una statistica** $T = t(X_1, \dots, X_n)$ (in genere è uno stimatore del parametro θ);
- 2 si sceglie **una regola di decisione**: “se $t(x_1, \dots, x_n) \in I$ allora rifiuto H_0 ”
(dove $t(x_1, \dots, x_n)$ è il valore della statistica calcolato dal campione e I è un sottoinsieme di \mathbb{R}).

L'idea è che I è scelto in modo che se H_0 è vera, allora è *molto improbabile* che $T \in I$.

Regione di rifiuto

DEFINIZIONE DI REGIONE DI RIFIUTO

L'insieme \mathcal{R} delle realizzazioni campionarie che portano a rifiutare H_0 , cioè

$$\mathcal{R} = \{(x_1, \dots, x_n) : t(x_1, \dots, x_n) \in I\}$$

è detta **regione di rifiuto**, o **critica** del test.

Se il campione osservato sta in \mathcal{R} (e quindi il corrispondente valore della statistica T sta in I) rifiuteremo H_0 .

Osservazione

La regione di rifiuto è un sottoinsieme dei valori che il campione casuale può assumere, dunque in generale è un sottoinsieme di \mathbb{R}^n . Inoltre

$$((X_1, \dots, X_n) \in \mathcal{R}) = (T \in I) = (\text{rifiuto } H_0).$$

La bontà di un test

Test d'ipotesi

- Ipotesi statistiche
- Ipotesi nulla
- Errori
- Scambio di ipotesi
- Regione di rifiuto
- Potenza**
- Significatività
- Test più potenti
- Le tappe
- p-value

Volendo costruire un test che confronti H_0 e H_1 abbiamo una vasta scelta per T e per I : quale scelta sarà la migliore?

Il test dovrebbe idealmente essere capace di individuare quando H_0 è vera e quando è falsa, dunque dovrebbe **limitare al massimo le probabilità di commettere l'errore di prima specie e l'errore di seconda specie.**

Tuttavia nella maggiorparte dei test limitare la probabilità dell'errore di una specie, aumenta la probabilità di commettere l'errore dell'altra specie.

Potenza

Definiamo una funzione che ci aiuterà a capire quanto “buono” sia il nostro test.

DEFINIZIONE DI FUNZIONE POTENZA DI UN TEST

La **funzione potenza** $\text{Pot} : \Theta \rightarrow [0, 1]$ è definita nel seguente modo:

$$\text{Pot}(\theta) := \mathbb{P}_{\theta}(T \in I).$$

In altre parole ad esempio $\text{Pot}(5)$ è la **probabilità di rifiutare H_0 quando il valore vero del parametro θ è 5.**

Se $\theta \in \Theta_0$ (dove $H_0 : \theta \in \Theta_0$)

$\text{Pot}(\theta)$ è la probabilità dell'errore di I specie.

Se $\theta \in \Theta_0^c$ (dove $H_1 : \theta \in \Theta_0^c$)

$1 - \text{Pot}(\theta)$ è la probabilità dell'errore di II specie.

Il test ideale

Test d'ipotesi

Ipotesi statistiche
Ipotesi nulla
Ipotesi nulla
Errori
Scambio di ipotesi
Regione di rifiuto
Potenza
Significatività
Test più potenti
Le tappe
p-value

Il test “ideale” dovrebbe avere

$$\text{Pot}(\theta) = 0, \text{ se } \theta \in \Theta_0$$

cioè probabilità di rifiutare H_0 quando è vera (errore di I specie) pari a 0, e

$$\text{Pot}(\theta) = 1, \text{ se } \theta \in \Theta_0^c$$

cioè probabilità di rifiutare H_0 quando è falsa (**non** commettere errore di II specie) pari a 1.

In realtà questa situazione ideale è (quasi) sempre irraggiungibile: si **fissa** perciò **la probabilità dell'errore di I specie**, e si cerca un test che **limiti il più possibile la probabilità dell'errore di II specie**.

Significatività

DEFINIZIONE DI LIVELLO DI SIGNIFICATIVITÀ DI UN TEST

Il **livello di significatività** α del test con regione critica \mathcal{R} è definito come

$$\alpha = \sup_{\theta \in \Theta_0} \text{Pot}(\theta) = \sup_{\theta \in \Theta_0} \mathbb{P}(T \in I) = \sup_{\theta \in \Theta_0} \mathbb{P}(\text{rifiutare } H_0).$$

In altre parole

α è la **massima** probabilità di commettere un errore di I specie.

Supponiamo ad esempio che $\alpha = 0.05$ e H_0 **sia vera**:

la probabilità di rifiutare H_0 (e dunque sbagliare!) è 0.05

\Rightarrow se ripetessi il test molte volte so che (Legge dei Grandi Numeri) al massimo nel 5% delle volte (circa) rifiuterei H_0 .

Test d'ipotesi

Ipotesi statistiche
Ipotesi nulla
Ipotesi nulla
Errori
Scambio di ipotesi
Regione di rifiuto
Potenza
Significatività
Test più potenti
Le tappe
p-value

Test più potenti

Fra due test con uguale livello di significatività (uguale massima probabilità dell'errore di I specie) si preferisce un test che abbia minore probabilità dell'errore di II specie.

DEFINIZIONE DI TEST PIÙ POTENTE

Dati due test di livello di significatività α , il primo con potenza Pot_1 e il secondo con potenza Pot_2 , se per ogni $x \in \Theta_0^c$ si ha

$$Pot_1(x) \geq Pot_2(x)$$

diciamo che il primo test è **più potente** del secondo.

Ricordiamo che $H_1 : \theta \in \Theta_0^c$ e se $\theta = x \in \Theta_0^c$ la probabilità di commettere un errore di II specie è $1 - Pot(x)$.

Fra due test con lo stesso livello di significatività

quello più potente ha minor probabilità di commettere un errore di II specie.

Le tappe di un test

Test d'ipotesi

Ipotesi statistiche
Ipotesi nulla
Ipotesi nulla
Errori
Scambio di ipotesi
Regione di rifiuto
Potenza
Significatività
Test più potenti
Le tappe
p-value

- 1 Sulla base del tipo di esperimento si sceglie un **modello statistico** per gli esiti, con uno o più parametri θ incogniti.
- 2 Si scelgono H_0 e H_1 (affermazioni sul valore di θ) in modo che sia più grave commettere l'errore di I specie che quello di II (equivalentemente, H_1 sarà l'ipotesi che riterremo vera **solo** in presenza di **forte** evidenza).
- 3 Si sceglie una **statistica** $T = t(X_1, \dots, X_n)$ (in genere stimatore di θ).
- 4 Si sceglie il **livello di significatività** $\alpha \in [0, 1]$, cioè la **massima** probabilità che tolleriamo di commettere un errore di I specie.

Le tappe di un test

Test d'ipotesi

Ipotesi statistiche
Ipotesi nulla
Ipotesi nulla
Errori
Scambio di ipotesi
Regione di rifiuto
Potenza
Significatività
Test più potenti
Le tappe
p-value

- 5 In dipendenza da T e α si sceglie la **regione di rifiuto**
 $\mathcal{R} = ((x_1, \dots, x_n) : t(x_1, \dots, x_n) \in I)$.
- 6 Si esegue il **campionamento** (= si fanno gli esperimenti e si osservano dei dati, realizzazioni del campione casuale).
- 7 Si **calcola** la statistica T sulla base del campione osservato.
- 8 Si **decide**: se il valore calcolato della statistica sta in I allora **rifiutiamo** H_0 (e accettiamo H_1) \implies **conclusione forte**.
Altrimenti **accettiamo** H_0 (e rifiutiamo H_1), dicendo che *non c'è evidenza statistica che H_0 sia falsa*. \implies **conclusione debole**.

Test d'ipotesi

Ipotesi statistiche
Ipotesi nulla
Ipotesi nulla
Errori
Scambio di ipotesi
Regione di rifiuto
Potenza
Significatività
Test più potenti
Le tappe
p-value

Per i vari modelli statistici si trovano sui libri i test d'ipotesi per la verifica delle ipotesi statistiche più comuni (e di solito si tratta dei test più potenti che si possano inventare), quindi noi ci troveremo di fronte a tabelle del tipo:

H_0	Se vale	decisione
$\theta \geq \theta_0$	$t(\mathbf{x}_1, \dots, \mathbf{x}_n) < g(\alpha)$	Rifiutiamo H_0
	$t(\mathbf{x}_1, \dots, \mathbf{x}_n) \geq g(\alpha)$	Accettiamo H_0

dove θ_0 è un numero fissato, T è la statistica scelta, α è il livello di significatività e $g(\cdot)$ è una funzione opportuna. In altre parole, la regione di rifiuto è

$$\mathcal{R} = \{(\mathbf{x}_1, \dots, \mathbf{x}_n) : t(\mathbf{x}_1, \dots, \mathbf{x}_n) < g(\alpha)\}.$$

Test d'ipotesi

Ipotesi statistiche
Ipotesi nulla
Ipotesi nulla
Errori
Scambio di ipotesi
Regione di rifiuto
Potenza
Significatività
Test più potenti
Le tappe
 p -value

DEFINIZIONE DI p -value

Dato un test d'ipotesi con regione critica \mathcal{R} , e date le osservazioni del campione x_1, \dots, x_n , il p -value è

$$\inf\{\alpha \in [0, 1] : (x_1, \dots, x_n) \in \mathcal{R}\}.$$

In altre parole

è il più basso livello di significatività per cui i dati osservati portano a rifiutare H_0 .

Notiamo che si tratta di un numero, che dipende dai dati osservati e dà un'indicazione di quanto forte sia l'evidenza che H_0 sia falsa.

Test d'ipotesi

Ipotesi statistiche
Ipotesi nulla
Ipotesi nulla
Errori
Scambio di ipotesi
Regione di rifiuto
Potenza
Significatività
Test più potenti
Le tappe
p-value

p-value e significatività

La **significatività** α rappresenta quanto siamo esigenti nel cercare prove che H_0 sia falsa. Infatti più α è piccolo, meno siamo disposti a commettere un errore di I specie.

Errore di I specie = rifiutare H_0 quando essa è vera



meno probabilità di rifiutare H_0 = rifiuto solo se le prove che
sia falsa sono **molto evidenti**.

Test d'ipotesi

- Ipotesi statistiche
- Ipotesi nulla
- Ipotesi nulla
- Errori
- Scambio di ipotesi
- Regione di rifiuto
- Potenza
- Significatività
- Test più potenti
- Le tappe
- p-value

p-value e significatività

Pensiamo al livello di significatività α come ad un'assicella del salto in alto e alle evidenze che H_0 è falsa come al saltatore: se il saltatore supera l'assicella allora rifiuto H_0 .



Più α è piccolo più l'assicella è alta. Se abbassiamo α , per rifiutare H_0 le prove devono “saltare” più in alto (= essere più evidenti).

p -value e significatività

Il p -value è il più basso livello di significatività per cui i dati osservati portano a rifiutare H_0 .

In termini dell'assicella, ci indica l'altezza più alta che i nostri dati avrebbero superato.



Se il livello è lo stesso, il p -value di sinistra è più basso di quello di destra, anche se in entrambi i casi rifiutiamo H_0 .

Test d'ipotesi

Ipotesi statistiche

Ipotesi nulla

Ipotesi nulla

Errori

Scambio di ipotesi

Regione di rifiuto

Potenza

Significatività

Test più potenti

Le tappe

p -value

Valori numerici per il p -value

Test d'ipotesi

Ipotesi statistiche
Ipotesi nulla
Ipotesi nulla
Errori
Scambio di ipotesi
Regione di rifiuto
Potenza
Significatività
Test più potenti
Le tappe
 p -value

Ricordiamo che più è basso il livello di significatività e più risulta difficile rifiutare H_0 (cioè ci vogliono prove più evidenti).

Perciò un **p -value basso** (0.01 o 0.001...) indicherà che è **molto probabile che H_0 sia falsa** (*anche se potrebbe essere vera!* pensate alla moneta equilibrata che dia 700 teste in 1000 lanci...).

Un **p -value alto** (0.10 o 0.30...) indicherà che è **poco (o per nulla) evidente che H_0 sia falsa**.