Test d'ipotesi

- La statistica non si pone solo l'obiettivo di comprendere le caratteristiche di una popolazione a partire dai dati raccolti su un campione rappresentativo, ma si pone anche l'obiettivo di verificare ipotesi fatte a priori su alcuni aspetti di interesse biologico.
- Per far questo utilizziamo i modelli teorici di distribuzione di probabilità che abbiamo appena introdotto.
- In altre parole rappresentiamo i nostri dati con un modello teorico che ben gli si adatti e su questo applichiamo dei ragionamenti che si possano tradurre formalmente in operazioni-procedure a base matematico-statistica.

Test d'ipotesi

- Illustriamo il ragionamento con un esempio teorico.
- Immaginiamo di avere una popolazione nota gaussiana con media μ e sd σ. Su questa popolazione ha agito un certo trattamento sperimentale che dovrebbe aver spostato la media della popolazione sul valore $\nu > \mu$.
- Ora prendiamo un campione di dimensione n da questa popolazione.
- Calcoliamo la media campionaria e troviamo che $\bar{x} > \mu$
- Ci chiediamo: *il valore $\bar{x}$ è effettivamente superiore a μ perché il trattamento ha avuto l'effetto ipotizzato, oppure si tratta di una normale oscillazione legata all'errore di campionamento?*

Test d'ipotesi

- Da un punto di vista pratico:
    1. Si formulano due ipotesi: *l'ipotesi nulla* (corrispondente ad esempio, ad affermare che non vi è effetto) e *l'ipotesi alternativa* (ad esempio $\nu > \mu$ come nel nostro esempio)
    2. Fissato un modello di distribuzione toerico, si calcola la probabilità che l'ipotesi nulla sia vera
    3. Se l'evento osservato ha probabilità inferiore ad una certa soglia α prefissata (generalmente 0.05), si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa.

Test d'ipotesi

- Quando formulo l'ipotesi scelgo una *statistica test* da calcolare sul campione
- E' quella statistica che determina a quale distribuzione di probabilità farò riferimento nel test
- Ad esempio con la media stimata da una popolazione gaussiana o con un campione molto numeroso userò la gaussiana o la t a seconda delle circostanze.

Test d'ipotesi: Esempio

- Poniamo di monitorare con uno strumento di analisi la concentrazione di una determinata sostanza in un pozzo, un metallo ad esempio.
- Le misure finora effettuate (molto numerose) mostrano che la concentrazione media è pari a 250 $\mu g/l$, con una deviazione standard pari a 180 $\mu g/l$
- Improvvisamente, estraiamo un campione d'acqua e facendo tre analisi otteniamo un valore medio pari a 550 $\mu g/l$.
- *Possiamo sospettare che il pozzo si è inquinato o si tratta di una normale oscillazione legata all'errore sperimentale?*

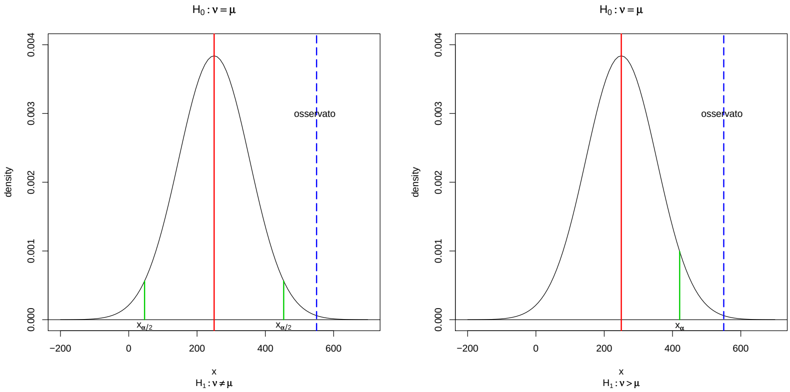
Test d'ipotesi: Esempio

- In altre parole: quanto è probabile il risultato che abbiamo ottenuto per il pozzo che stiamo analizzando?
- Per rispondere ci avvaliamo di quanto già sappiamo: le misure fatte in passato sul pozzo si distribuiscono come una normale di media 250 e deviazione standard 180. Sappiamo anche che la media campionaria estratta da una normale si distribuisce come una normale con media uguale a quella della popolazione di provenienza e sd$= \frac{\sigma}{\sqrt{n}}$
- Quindi per rispondere dobbiamo calcolare la probabilità di osservare il valore 550 in una popolazione $N(250, \frac{180}{\sqrt{3}})$
- $Prob(550) = 0.001946209$ molto improbabile e quindi possiamo dire che il pozzo è inquinato.

Test d'ipotesi

- Nel nostro esempio possiamo formalizzare il problema di test d'ipotesi in due modi:
  1. $H_0 : \nu = 250$; $H_1 : \nu \neq \mu$ ovvero l'ipotesi nulla assume che non sia successo niente e che la media osservata assuma quel valore solo per effetto del caso, l'ipotesi alternativa dice solo che questo non è vero
  2. $H_0 : \nu = 250$; $H_1 : \nu > \mu$ l'ipotesi nulla è come prima mentre l'ipotesi alternativa dichiara esplicitamente che il valore osservato maggiore della media usuale.
- La forma delle ipotesi alternative implica un modo operativo diverso di distribuire l'errore sulla distribuzione di riferimento.

Test d'ipotesi: Esempio



Test d'ipotesi: errori

- Il test d'ipotesi ci ha condotto alla decisione di rifiutare l'ipotesi nulla. Ma siamo proprio sicuri che la nostra decisione corrisponda alla veritá?
- No, perché anche se abbiamo deciso di rifiutarla, l'ipotesi nulla potrebbe essere vera; infatti questo evento non é impossibile, ma solo improbabile (ha una probabilità di 0.0019 di verificarsi).
- Questo tipo di errore (rifiutare erroneamente l'ipotesi nulla) si dice **errore di prima specie** e la sua probabilità nel nostro esempio é pari a 0.0019.
- Esiste anche il cosiddetto **errore di seconda specie**, che chiariremo con un esempio.

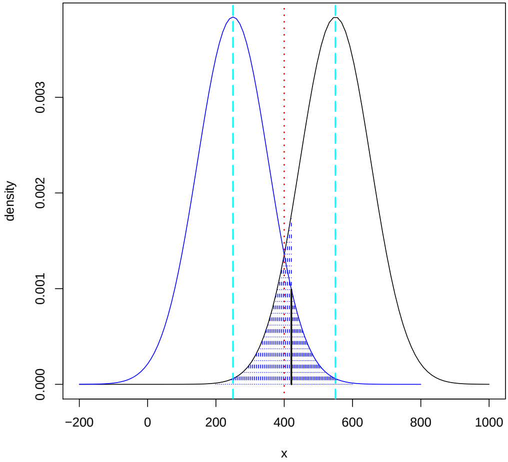
Test d'ipotesi: errori

- L'errore di seconda specie è l'errore che si commette quando accettiamo l'ipotesi nulla e questa non è vera.
- Immaginiamo un problema di ipotesi un po' diverso dal precedente:

$$H_0 : \nu = 250; \;\; H_1 : \nu = 550$$

- Stiamo dicendo che la media vera è quella osservata.
- Le distribuzioni di riferimento relative alle due ipotesi sono due normali con la stessa deviazione standard ma con medie diverse.
- Queste distribuzione si sovrappongono in parte.

Test d'ipotesi:errori



Test d'ipotesi: errori

- Normalmente fissiamo l'errore di prima specie, $\alpha = 0.05$ nel nostro esempio, a cui corrisponde un valore specifico sull'asse x che nel nostro esempio è $x_c = 420.9382$ ed è detto **valore critico**
- L'errore di seconda specie è la probabilità sotto la distribuzione di H_1 a sinistra del valore critico e nel nostro caso è $\beta = 0.107$
- Non è possibile rendere minimi entrambi gli errori simultaneamente. Di solito si fissa l'errore di prima specie.

Test d'ipotesi

- Analizziamo un primo esempio pratico di test d'ipotesi: nella sperimentazione biologica si ha spesso interesse a considerare due popolazioni per scoprire se queste sono diverse per il carattere o i caratteri considerati.
- Se poi rappresentiamo ciascuna popolazione con la sua media, saremo interessati a rispondere al quesito *se l'eventuale differenza rilevata tra le due medie è da ritenersi una differenza reale, effettiva e con un preciso significato biologico*
- in termini statistici, dovremo stabilire se la differenza tra le medie è *significativa* oppure *da attribuire a fattori casuali e quindi non significativa.*

Test d'ipotesi: esempio II

- Si vogliono confrontare due varietà di grano con diverse caratteristiche delle cariossidi (VICTO e LUCREZIA), per valutare quale delle due ha peso ettolitrico più alto.
- Vengono coltivate 350000 piante uniformemente suddivise tra le due varietà.
- Si estraggono 5 campioni casuali per ciascuna varietà e se ne analizza il contenuto di Cariossidi (peso ettolitrico)
- Useremo la distribuzione t di student per costruire gli Intervalli di Confidenza dato che i campioni sono poco numerosi

VICTO	LUCREZIA
65	70
68	71
69	74
71	78
78	84

Test d'ipotesi: esempio II

Risultati I

	VICTO	LUCREZIA
media	70.2	75.4
sd	4.87	5.73
t 4 gdl p=0.975	2.77	2.77
estremo inf IC	64.15	68.29
estremo sup IC	76.25	82.51

Notare che i due IC si sovrappongono parzialmente

Test d'ipotesi: esempio II

- Il peso ettolitrico medio di LUCREZIA è maggiore di quello di VICTO, ma è anche vero che dei cinque dati relativi a LUCREZIA, solo uno è superiore a tutti quelli relativi a VICTO.
- Gli intervalli di confidenza delle due medie si sovrappongono parzialmente.
- Come risolvere $\rightarrow$faccio un test rigoroso usando la distribuzione t come modello di riferimento.
- la domanda è: *Possiamo affermare che la varietà Lucrezia ha un peso ettolitrico più alto di Victo?*

Test d'ipotesi: costruiamo il test

- La decisione dovrà essere basata su due aspetti:
  1. **l'ampiezza della differenza tra le medie**: più la differenza tra le due medie è alta e più è probabile che essa sia significativa;
  2. **deviazione standard delle medie**. Più è elevata la variabilità dei dati e quindi l'errore di stima e più è bassa la probabilità che le differenze osservate tra le medie siano significative.
- Il test lo costruisco usando proprio queste due considerazioni

Test t

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_{\bar{x}_1 - \bar{x}_2}}$$

Test d'ipotesi: costruiamo il test

- La statistica test, t è definita dal rapporto tra la differenza tra le medie delle due popolazioni e la deviazione standard della differenza. Quest'ultima si ottiene come media ponderata delle varianze delle due popolazioni:

$$S_{\bar{x}_1 - \bar{x}_2} = \sqrt{S_m^2 \frac{n_1 + n_2}{n_1 n_2}}$$

dove

$$S_m^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}$$

S_1^2 ed S_2^2 sono le varianze ed n_1, n_2 le numerosità dei due campioni. Così si calcola la varianza nel caso di assunzione di eguale varianza per le due popolazione.

Test d'ipotesi: costruiamo il test

- Ora occorre definire la struttura delle ipotesi. In generale diremo (ipotesi alternativa bidirezionale):

$$
\begin{aligned}
H_0 &: \quad \mu_1 = \mu_2 \\
H_1 &: \quad \mu_1 \neq \mu_2
\end{aligned}
$$

- potremmo anche costruire un'ipotesi alternativa del tipo (ipotesi alternativa monodirezionale)

$$H_1 : \mu_1 > \mu_2 \;\; \textit{oppure} \;\; \mu_1 < \mu_2$$

  questo nel caso avessimo la convinzione che una media è maggiore (o minore) dell'altra.

Test d'ipotesi: esempio II

- Riprendiamo l'esempio: $\bar{x}_1 = 70.2$, $\bar{x}_2 = 75.4$, $S_1^2 = 23.7$, $S_2^2 = 32.8$ e $n_1 = n_2 = 5$ i gradi di libertà sono $n_1 + n_2 - 2 = 8$
- assumendo varianze uguali per le due popolazioni abbiamo che $S_{\bar{x}_1 - \bar{x}_2} = \sqrt{S_m^2 \frac{n_1+n_2}{n_1 n_2}} = 3.76$
- il valore della statistica test è

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_{\bar{x}_1 - \bar{x}_2}} = -1.5469$$

- la distribuzione di riferimento è la t con 8 gdl con $t_{8,0.0025} = -2.31$ circa quindi essendo la nostra statistica in valore assoluto più piccola del valore critico accettiamo H_0 con un errore di prima specie $\alpha = 0.05$.

Test d'ipotesi: la F di Fisher Snedecor

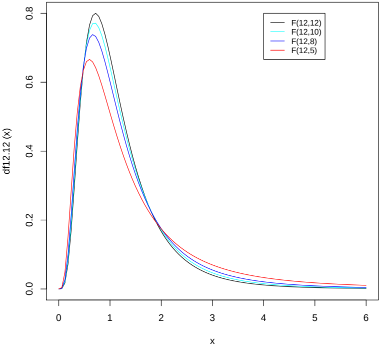
- Nel costruire la statistica test per il confronto tra le medie abbiamo calcolato la deviazione standard della differenza tra le medie assumendo che la variabilità delle due popolazioni a confronto fosse la stessa
- Questa assunzione va verificata. Indichiamo con σ_1^2 e σ_2^2 le varianze delle due popolazioni a confronto.
- Volendo confrontare due varianze costruiamo un problema di ipotesi come segue:

$$\begin{aligned} H_0 &: \sigma_1^2 = \sigma_2^2 \\ H_1 &: \sigma_1^2 > \sigma_2^2 \end{aligned}$$

- Come statistica test usiamo il rapporto tra le due varianze mettendo a numeratore la varianza più grande.

Test d'ipotesi: la F di Fisher Snedecor

- Quindi supponendo che $\sigma_1^2 > \sigma_2^2$ costruiamo $F = \frac{\sigma_1^2}{\sigma_2^2}$
- Questa statistica si distribuisce come una *F di Fisher-Snedecor*



Test d'ipotesi: la F di Fisher Snedecor

- La distribuzione F è caratterizzata dai gradi di libertà del numeratore e del denominatore
- Formalmente si indica $F \sim F_{(gdl_1),(gdl_2)}$
- Nell'esempio del grano stimiamo le due varianze con la varianza campionaria: $s^2_{Victo} = 23.7$ $s^2_{Lucrezia} = 32.8$, ciascuno con $gdl = n - 1 = 4$ quindi

$$F = \frac{s^2_{Lucrezia}}{s^2_{Victo}} \sim F_{4;4}$$

- Il valore della statistica è $F = \frac{32.8}{23.7} = 1.384$ mentre a livello $\alpha = 0.05$ (errore di prima specie) il valore critico del test è $F_{4;4;0.05} = 6.39$
- La nostra statistica è più piccola del valore critico e quindi accettiamo H_0, le varianze sono uguali ed è quindi corretto usare l'espressione della statistica t che abbiamo scelto.

Test d'ipotesi: confronto tra medie

- Se il test da esito negativo, ovvero non possiamo affermare che le due varianze sono diverse solo per effetto del caso, allora ne dobbiamo tenere conto nella costruzione della statistica test.
- In questo caso è la stima di $S_{\bar{x}_1 - \bar{x}_2}$ che cambia:

$$S_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

- La differenza più sostanziale però la vediamo nel calcolo dei gradi di libertà da considerare. Si dimostra che questi sono:

$$gdl = \frac{(s_1^2/n_1 + s_2^2/n_2)}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

Nei software questo calcolo dei gradi di libertà è detto *Modifica di Welch*

Test d'ipotesi: confronto tra medie

Altro esempio

- Un prodotto in grado di diminuire la crescita delle piante (brachizzante) viene spruzzato su quattro parcelle di orzo e messo a confronto con quattro parcelle trattate solo con acqua.
- La media dell'orzo trattato è $\bar{x}_1 = 83.75$ mentre quella dell'orzo non trattato è $\bar{x}_2 = 94.75$, le varianze sono $s_1^2 = 31.58$ e $s_2^2 = 24.25$
- Effettuiamo un test sulle varianze e la funzione var. test di R restituisce `F = 1.3024, num df = 3, denom df = 3, p-value = 0.4166` la grandezza che ci interessa è il p-value che è la probabilità della nostra statistica rispetto al modello teorico scelto. In questo caso è maggiore della soglia $\alpha = 0.05$ scelta e quindi accettiamo H_0 le due varianze sono uguali.

Test d'ipotesi: confronto tra medie

Altro esempio

- Ora possiamo condurre il nostro test sulle medie, in R la funzione t.test con alternativa `two.sided`, ovvero $H_1 : \mu_1 \neq \mu_2$, ci dice: `t = -2.9443, df = 6, p-value = 0.02580`
- Il p-value è minore di 0.05 quindi rifiutiamo H_0
- Noi però pensiamo che il trattamento abbia ridotto l'altezza dell'orzo quindi ripetiamo il test con alternativa `less` ovvero $H_1 : \mu_1 < \mu_2$ e otteniamo: `t = -2.9443, df = 6, p-value = 0.01290`
- Il p-value è ancora più piccolo quindi confermiamo il nostro rifiuto di H_0.

Test d'ipotesi: Significatività statistica e significatività biologica

- Quando si rifiuta un'ipotesi nulla (nessun effetto o nessuna differenza), significa che vi è una buona evidenza che un effetto sia presente. Ma quest'effetto può essere molto piccolo.
- **Attenzione:** *Se i campioni sono molto numerosi anche le deviazioni più piccole dall'ipotesi nulla risulteranno significative*.
- *Occorre decidere se la significatività statistica osservata ha un significato biologico*.
- Per esempio, se misurassimo 100 lumache in ciascuna di 2 popolazioni, quasi certamente troveremmo che le dimensioni delle 2 popolazioni sono diverse. Ma se la differenza nella dimensione media fosse dell'1%, avremmo difficoltà a spiegare il significato biologico di questa piccola differenza.

Test d'ipotesi: Tavola riassuntiva

Conclusione/Verità	**H_0 vera**	**H_0 falsa**
H_0 vera	GIUSTO $P = 1 - \alpha$ livello di protezione	ERRORE di II specie $P = \beta$
H_0 falsa	ERRORE di I specie $P = \alpha$ livello di significatività	GIUSTO $P = 1 - \beta$ Potenza del test

P-values: Cosa è

- In statistica inferenziale il *valore p* (o *p-value*, in inglese) di un test di verifica d'ipotesi indica la probabilità di ottenere un risultato pari o più estremo di quello osservato, supposta vera l'ipotesi nulla . Talvolta viene anche chiamato *livello di significatività osservato*.
- Va usato con consapevolezza!

P-values: misunderstandings

- Il p-value NON è la probabilità che l'ipotesi nulla sia vera e NON è la probabilità che sia vera l'ipotesi alternativa..
- il p-value NON è la probabilità che un certo risultato sia semplicemente un "caso fortuito"
- il p-value NON è la probabilità di rifiutare l'ipotesi nulla quando questa è vera.
- il p-value NON è la probabilità che una replica dell'esperimento non darà lo stesso risultato osservato
- Il livello di significatività di un test NON è determinato dal p-value, va deciso a priori prima di effettuare il test.