

Elementi di Inferenza Statistica

La verifica di ipotesi

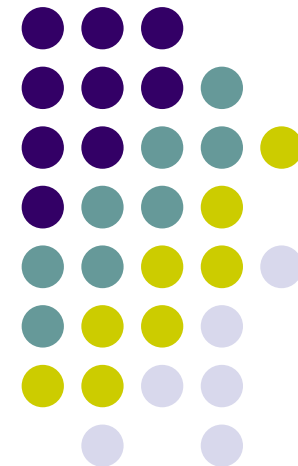
V Scuola Estiva AISV

*La statistica come strumento di analisi nelle
scienze umanistiche e comportamentali*

Soriano nel Cimino (VT), 6 Ottobre 2009

Pier Francesco Perri

*Dipartimento di Economia e Statistica - UNICAL
pierfrancesco.perri@unical.it*



... per iniziare



Caso studio. La *ripetizione di non-parole* è una prova con cui si può valutare la capacità di un bambino di analizzare il segnale acustico, costruire un piano articolatorio e immagazzinarlo nella memoria fonologica a breve termine.

Essa rappresenta una misura della memoria fonologica a breve termine.

La prova consiste nel chiedere al bambino di ripetere, il più fedelmente possibile, delle non-parole ascoltate tramite un registratore (esempio: deccarello < pennarello)

Consideriamo la popolazione di bambini di età compresa tra 6 e 7 anni che frequentano la scuola elementare e supponiamo che il numero medio di non-parole prodotte erroneamente sia pari a 8.



Un gruppo di ricercatori conduce uno studio su un campione di 60 bambini ottenendo un numero medio di non-parole prodotte erroneamente pari a 7.17 con una deviazione standard pari a 3.56. (www.infantiae.org/orcasace.asp?st=1)



- La discrepanza tra ciò che si teorizza e ciò che si osserva è da attribuire alla variabilità campionaria?
- Oppure, la congettura iniziale non è supportata dai fatti e quindi è da ritenersi non vera?

Una risposta può essere data attraverso lo strumento statistico della **verifica di ipotesi**

Ipotesi e test



Ipotesi Statistica: è un'affermazione/congettura su un parametro della popolazione.

Verifica (o Test) di Ipotesi: è un procedimento/regola che utilizza i dati campionari per decidere se un'ipotesi è vera o falsa.

Un test è strutturato su due ipotesi denominate **ipotesi nulla** (H_0) ed **ipotesi alternativa** (H_1).

- ✚ L'ipotesi nulla è quella che riflette lo *status quo* ed è ritenuta vera fino a prova contraria. E' così chiamata perché la sua accettazione non comporta alcuna modifica della realtà.
- ✚ L'ipotesi alternativa è un'affermazione contrastante H_0

Ipotesi e test



Con riferimento al caso studio iniziale, potremmo strutturare un test per le seguenti situazioni

- $H_0: \mu = 8$ vs $H_1: \mu > 8$ (*test unilaterale a destra*)
- $H_0: \mu = 8$ vs $H_1: \mu < 8$ (*test unilaterale a sinistra*)
- $H_0: \mu = 8$ vs $H_1: \mu \neq 8$ (*test bilaterale*)

Un test di ipotesi parte dal presupposto che H_0 sia vera. Pertanto si cerca di verificare se i dati osservati suffragano o meno questa ipotesi. Il test, quindi, è condotto su H_0 e può dar luogo solo a due esiti:

- **Rifiutare H_0** . C'è una forte evidenza empirica che non supporta H_0
- **Accettare H_0** . C'è una forte evidenza empirica che supporta H_0

Errori nella verifica d'ipotesi



- ✚ La decisione se rifiutare o meno H_0 viene presa sulla base di un solo campione: quello selezionato ed osservato.
- ✚ Questa azione ha un limite derivante dal fatto che qualsiasi informazione deducibile dal campione, e connessa ad affermazioni concernenti l'intera popolazione, non può che essere parziale e dipendente dallo specifico campione di cui si dispone.
- ✚ Potrebbe darsi che un altro campione conduca ad una decisione opposta!

Errori nella verifica d'ipotesi



In sostanza, dal momento che non si conoscono gli aspetti della popolazione su cui si cerca di inferire, **non si saprà mai con certezza se la decisione di accettare o rifiutare H_0 sia corretta.**

Pertanto, di fronte ad una situazione sperimentale, il ricorso ad un test statistico conduce a quattro possibili azioni:

- ❏ **E1:** si rifiuta H_0 quando è vera (decisione errata)
- ❏ **E2:** si accetta H_0 quando è falsa (decisione errata)
- ❏ **G1:** si accetta H_0 quando è vera (decisione corretta)
- ❏ **G2:** si rifiuta H_0 quando è falsa (decisione corretta)

Errori nella verifica d'ipotesi



Le quattro situazioni sono astratte e definibili solo concettualmente dal momento che solo una delle due ipotesi è quella vera, ma non si conosce quale.

Da ciò discende che non **avremo mai la certezza** di aver preso la decisione corretta dal momento che non è dato conoscere se H_0 è effettivamente vera o falsa (se lo sapessimo il problema non si porrebbe!).

Errore di 1° tipo: Rifiutare H_0 quando è vera (E1)

Errore di 2° tipo: Accettare H_0 quando è falsa (E2)

Questi due tipi di errori sono insiti nella procedura inferenziale e non possono essere eliminati.

Rappresentano il prezzo da pagare nel prendere decisioni in condizioni di incertezza sulla base di un campione.

Errori nella verifica d'ipotesi



Prima della selezione del campione, è possibile associare una probabilità agli errori di 1° e 2° tipo.

La **probabilità dell'errore di 1° tipo** è indicata con α ed è detta **livello di significatività del test**. Essa rappresenta una misura del rischio che si è disposti ad accettare ricorrendo ad un test che potrebbe condurre a rifiutare erroneamente l'ipotesi nulla

La **probabilità dell'errore di 2° tipo** è indicata con β

- ✓ $P(E1) = \alpha$
- ✓ $P(E2) = \beta$
- ✓ $P(G1) = 1 - \alpha$
- ✓ $P(G2) = 1 - \beta$ (**potenza del test**)

Un test è detto **significativo** se conduce al rifiuto di H_0

Errori nella verifica d'ipotesi



- ✚ Nella costruzione di un test si vorrebbe che le probabilità degli errori di 1° e 2° tipo fossero entrambe piccole, addirittura pari a zero.
- ✚ Premesso che ogni test presuppone l'assunzione di un rischio di errore e che entrambe le probabilità non possono essere minimizzate, si procede fissando il valore di α e si definisce una regola per la quale il valore di β è minimo.

In altri termini si preferisce controllare l'errore di 1° tipo in quanto ritenuto più importante.

Modificare la realtà senza alcuna evidente motivazione comporterebbe, infatti, conseguenze dannose.

Tanto più piccolo è il valore di α tanto più evidente dovrà essere la discrepanza tra ciò che viene ipotizzato e ciò che viene osservato per poter rifiutare H_0 .

$\alpha = 0.05$ (5%)

$\alpha = 0.01$ (1%)

$\alpha = 0.001$ (1‰)

La costruzione di un test



- Formulate le ipotesi H_0 e H_1 e fissato il livello di significatività del test (α), occorre decidere, sulla base del campione, se accettare o rifiutare l'ipotesi nulla
- Per poter procedere occorre sintetizzare le informazioni del campione in un'unico valore
- Si **chiama statistica test** una funzione dei dati campionari che consente di prendere una decisione circa H_0
- La statistica test si configura come una v.c. la cui distribuzione è nota se si assume vera H_0

La costruzione di un test



si individua una statistica test



si definisce la **regione di rifiuto del test**, ovvero l'insieme dei valori della statistica test che conducono al rifiuto dell'ipotesi nulla



si osserva un campione e si calcola il valore della statistica test sul campione



se il valore della statistica test ricade nella regione di rifiuto si rigetta l'ipotesi nulla, altrimenti la si accetta

Test per la media di una Normale, varianza nota



Supponiamo che il fenomeno in esame X abbia una distribuzione normale, $X \sim N(\mu, \sigma^2)$.

Consideriamo l'ipotesi nulla $H_0: \mu = \mu_0$

La statistica test è: $\frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \approx N(0, 1)$

Le regioni di rifiuto, per le diverse configurazioni dell'ipotesi alternativa, sono:

$$R = \begin{cases} z \leq -z_\alpha & \text{se } H_1: \mu < \mu_0 \\ z \geq z_\alpha & \text{se } H_1: \mu > \mu_0 \\ |z| \geq z_{\alpha/2} & \text{se } H_1: \mu \neq \mu_0 \end{cases}$$

Regioni di rifiuto



Così, ad esempio, se si assume $\alpha=0.05$ si rifiuta H_0 se:

- si osserva un valore z più piccolo di $z^*=-1.654$ se $H_1: \mu < \mu_0$
- si osserva un valore z più grande di $z^*=1.654$ se $H_1: \mu > \mu_0$
- si osserva un valore z più piccolo di $z^*=-1.96$ oppure più grande $z^*=1.96$ se $H_1: \mu \neq \mu_0$

I valori soglia (z^*) vengono detti **valori critici**

Il *P*-value



Accade spesso che il test conduca a conclusioni diverse a seconda del livello di significatività assunto.

Così, ad esempio, potremmo rifiutare l'ipotesi nulla con un livello di significatività del 5% ma accettarla ad un livello di significatività del 1% .

D'altra parte, decisori diversi possono avere una differente *avversione al rischio* che si concretizza nella scelta di un livello di significatività che rispecchia il proprio rischio. In altri termini la scelta di α è soggettiva!

Per svincolare la decisione su H_0 dalla scelta di α è possibile ricorrere al ***P*-value** o **livello di significatività osservato**.

Il P-value

Il P-value è una misura dell'incoerenza tra ciò che viene postulato sotto H_0 e ciò che viene osservato.

Assumendo vera H_0 è calcolato come la probabilità che la statistica test assuma un valore più estremo (= nelle code della distribuzione) rispetto al suo valore osservato.

$$P\text{-value} = P(Z > z | H_0)$$

$$P\text{-value} = P(Z < z | H_0)$$

$$P\text{-value} = P(|Z| > z | H_0)$$

Un valore "piccolo" del P-value può essere determinato da due situazioni:

- l'ipotesi nulla è vera e si è verificato un evento raro, poco plausibile
- l'ipotesi nulla non giustifica/supporta quanto osservato e quindi è falsa

Tra le due si opta per la seconda.

Ma quanto piccolo deve essere il P-value?



Il P -value



Ma quanto piccolo deve essere il P -value per rifiutare di H_0 ?

Non esiste una regola assoluta, anche perché la decisione deve rispecchiare il proprio grado di avversione al rischio.

E' consuetudine abbastanza consolidata rifiutare H_0 quando P -value < 0.05 .

Ovviamente, tanto più piccolo è il P -value tanto più piccolo sarà il margine di errore della nostra decisione, ovvero più certi saremo di aver rifiutato correttamente l'ipotesi nulla!!

Esempio

Ritorniamo al caso studio relativo alla *ripetizione di non-parole*. Supponiamo che il numero di non-parole prodotte erroneamente segua una distribuzione Normale con varianza nota pari a 25.

Vogliamo sottoporre a verifica l'ipotesi nulla $H_0: \mu = 8$ assumendo un livello di significatività del 5%.

A tal fine viene selezionato un campione di 60 bambini ottenendo un numero medio di non-parole prodotte erroneamente pari a **7.17**.

Il valore della statistica test è: $z = \frac{7.17 - 8}{5 / \sqrt{60}} = -1.286$

Tale valore deve essere ora confrontato con i valori critici (z^*) determinati dall'ipotesi alternativa



✚ $H_0: \mu = 8$ vs $H_1: \mu > 8$ Poiché $-1.286 < 1.645$ si accetta H_0

✚ $H_0: \mu = 8$ vs $H_1: \mu < 8$ Poiché $-1.286 > -1.645$ si accetta H_0

✚ $H_0: \mu = 8$ vs $H_1: \mu \neq 8$ Poiché $-1.96 < -1.286 < 1.96$ si accetta H_0

Non c'è una sufficiente evidenza empirica per rifiutare l'ipotesi nulla, qualunque sia l'ipotesi alternativa.

A maggior ragione, la decisione circa H_0 non cambia se si assume un livello di significatività inferiore al 5%.

P-value: $P(Z < -1.286) = 0.099$



Test per la media di una Normale, varianza incognita



Supponiamo che il fenomeno in esame X abbia una distribuzione normale, $X \sim N(\mu, \sigma^2)$, con μ e σ^2 incogniti

Consideriamo l'ipotesi nulla $H_0: \mu = \mu_0$

La statistica test è: $T = \frac{\bar{X} - \mu_0}{S / \sqrt{n}} \approx t\text{-Student con } n-1 \text{ g.d.l.}$

Le regioni di rifiuto, per le diverse configurazioni dell'ipotesi alternativa, sono:

$$R = \begin{cases} t \leq -t_{\alpha}^{(n-1)} & \text{se } H_1: \mu < \mu_0 \\ t \geq t_{\alpha}^{(n-1)} & \text{se } H_1: \mu > \mu_0 \\ |t| \geq t_{\alpha/2}^{(n-1)} & \text{se } H_1: \mu \neq \mu_0 \end{cases}$$

Caso studio



Ritorniamo al caso studio relativo alla *ripetizione di non-parole*. Supponiamo che il numero di non-parole prodotte erroneamente segua una distribuzione Normale con varianza incognita.

Vogliamo sottoporre a verifica l'ipotesi nulla $H_0: \mu = 8$ assumendo un livello di significatività del 5%.

A tal fine viene selezionato un campione di 60 bambini ottenendo un numero medio di non-parole prodotte erroneamente pari a **7.17** e una deviazione standard pari a **3.56**.

Il valore della statistica test è:
$$t = \frac{7.17 - 8}{3.56 / \sqrt{60}} = -1.784$$

Tale valore deve essere ora confrontato con i valori critici (t^*) determinati dall'ipotesi alternativa

Caso studio



Nella tabella che segue riportiamo i valori critici del t-test che conducono al rifiuto di H_0 assumendo diversi livelli di significatività.

α	$H_1: \mu \neq 8$	$H_1: \mu > 8$	$H_1: \mu < 8$
0.05	2.001	1.671	-1.671
0.01	2.662	2.391	-2.391
0.001	3.463	3.234	-3.234

Rifiutiamo l'ipotesi nulla solo con un livello di significatività al 5% in favore dell'ipotesi alternativa $H_1: \mu < 8$

Test per una proporzione



Supponiamo ora di voler verificare un'ipotesi sulla proporzione di individui che presentano un determinato attributo nella popolazione. L'ipotesi nulla è $H_0: p = p_0$

La statistica test da utilizzare é $Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \approx N(0,1)$

$$R = \begin{cases} z \leq -z_\alpha & \text{se } H_1: p < p_0 \\ z \geq z_\alpha & \text{se } H_1: p > p_0 \\ |z| \geq z_{\alpha/2} & \text{se } H_1: p \neq p_0 \end{cases}$$



Esempio. Un gruppo di ricercatori vuole verificare se la proporzione di bambini in età prescolare affetti da difetti dell'apprendimento è aumentata o diminuita rispetto ai dati di 5 anni prima (7.9%).

A tal fine viene condotto uno studio su 200 bambini rilevando che 17 di essi, ovvero l'8.5%, prestano il disturbo.

Si decide di condurre un test bilaterale con un livello di significatività dell'1%.

$$H_1: p \neq 0.079$$

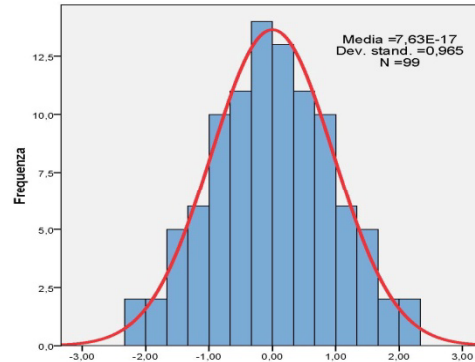
Il valore critico del test è $|z^*|=2.576$

Valore della statistica test

$$z = \frac{0.085 - 0.079}{\sqrt{\frac{0.085(1 - 0.085)}{200}}} = 0.304$$

Accetto H_0 : non c'è stato un cambiamento significativo rispetto a 5 anni fa.

Verifica dell'ipotesi di Normalità



$H_0: F(x) = \text{Normale}$ vs $H_1: F(x) \neq \text{Normale}$

Oltre al test di K-S, può essere utilizzato il test di Shapiro-Wilk (*ad hoc* per la normalità)

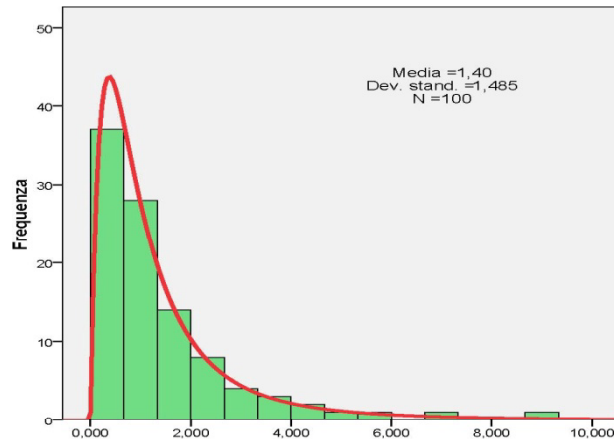
Test di normalità

Sig. = *P*-value

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistica	df	Sig.	Statistica	df	Sig.
normale	,010	99	,200	,998	99	1,000

In entrambi i test *P*-value > 0.05 e quindi **si accetta** l'ipotesi di normalità

Verifica dell'ipotesi di Normalità



$H_0: F(x) = \text{Normale}$ vs $H_1: F(x) \neq \text{Normale}$

Test di normalità

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistica	df	Sig.	Statistica	df	Sig.
normale	,174	100	,000	,742	100	,000

In entrambi i test $P\text{-value} \approx 0$ e quindi si **rifiuta** l'ipotesi di normalità

*Dati_Simulati_2.sav [InsiemeDati1] - SPSS Data Editor

File Modifica Visualizza Dati Trasforma **Analizza** Grafici Strumenti Finestra Aiuto



7:

	normale	fisher
1	-2,33	0,
2	-2,05	0,
3	-1,88	0,
4	-1,75	0,
5	-1,64	0,
6	-1,55	0,
7	-1,48	0,
8	-1,41	0,
9	-1,34	0,
10	-1,28	0,
11	-1,23	0,
12	-1,18	0,
13	-1,13	0,
14	-1,08	0,
15	-1,04	0,

- Report ▶
- Statistiche descrittive ▶
- C**onfronta medie ▶
- Modello lineare generalizzato ▶
- Modelli lineari generalizzati ▶
- Modelli misti ▶
- Correlazione ▶
- Regressione ▶
- Loglineare ▶
- Classifica ▶
- Riduzione dati ▶
- Scala ▶
- Test non parametrici ▶
- Serie storiche ▶
- Sopravvivenza ▶
- Risposte multiple ▶
- Controllo qualità ▶
- Curva ROC...

- M**edie... ▶
- t** Test T: campione unico... ▶
- t** **A-B** Test T: campioni indipendenti... ▶
- t** **A₁-A₂** Test T: campioni appaiati... ▶
- F** **A** ANOVA univariata... ▶

Dati



15:

	normale	fisher	classi	VOT	var	var	var	var	var
1	-2,33	0,000	1,00	18,27					
2	-2,05	0,066	1,00	13,34					
3	-1,88	0,096	1,00	20,50					
4	-1,75	0,121	1,00	21,15					
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17	-0,95	0,343	1,00	24,27					
18	-0,92	0,359	1,00	20,24					
19	-0,88	0,374	1,00	19,62					
20	-0,84	0,389	1,00	16,67					
21	-0,81	0,404	1,00	21,18					
22	-0,77	0,420	1,00	14,66					
23	-0,74	0,435	1,00	22,86					
24	-0,71	0,450	2,00	26,49					
25	-0,67	0,466	2,00	17,23					
26	-0,64	0,482	2,00	23,43					

Test T per un campione

Variabili oggetto del test:
VOT

Valore oggetto del test: 15

Opzioni...

OK Incolla Reimposta Annulla Aiuto

$H_0: \mu = 15$

Test T per un campi...

Intervallo di confidenza: 95 %

Valori mancanti

Esclusione casi analisi per analisi

Esclusione listwise

Continua Annulla Aiuto



Statistiche per un campione

	N	Media	Deviazione std.	Errore std. Media
VOT	50	20,1573	3,90390	,55210

Test per un campione

	Valore oggetto del test = 15					
	t	df	Sig. (2-codice)	Differenza fra medie	Intervallo di confidenza per la differenza al 95%	
					Inferiore	Superiore
VOT	9,341	49	,000	5,15730	4,0478	6,2668