Verifica delle ipotesi

Con la procedura di stima intervallare si cerca definire in modo verosimile il valore di un parametro incognito a partire dalle osservazioni campionarie

Valore campionario → **Procedura di stima intervallare** → Intervallo nel quale cade con un certo valore di probabilità il parametro della popolazione

Verifica delle ipotesi

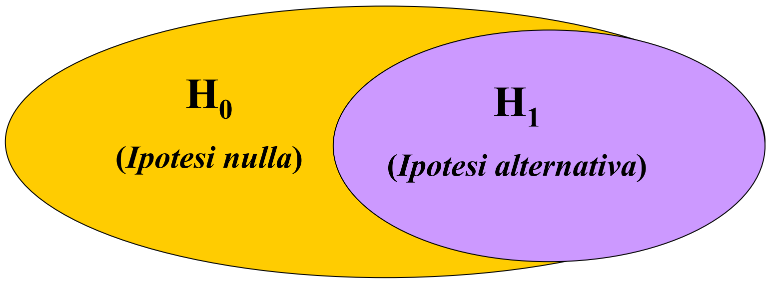
…con la verifica delle ipotesi non ci si prefigge di arrivare a conoscere il valore di uno o più parametri incogniti, ma si vuole accertare se una certa affermazione sulla popolazione debba ritenersi vera o falsa

ipotesi → **TEST STATISTICO** → Decisione probabilistica circa la verità o la falsità dell'ipotesi

Verifica delle ipotesi

Logica del test

I test statistici sono procedure formalizzate con le quali si sottopone a *falsificazione* una certa ipotesi definita *ipotesi nulla*, *l'ipotesi alternativa* o di ricerca contiene una affermazione non compatibile con quella dell'ipotesi nulla.



L' *ipotesi nulla e l'ipotesi alternativa* sono mutualmente esclusive.

Verifica delle ipotesi

Logica del test

Esempio: ci chiediamo se le differenze nella prestazione media di un campione di maschi e di uno di femmine in un particolare compito cognitivo sia da attribuirsi al caso o ad una differenza tra le due popolazioni (maschi e femmine) dalle quali abbiamo estratto i campioni.

Per sottoporre alla "verifica" di un test questa affermazione dobbiamo tradurla in termini idonei.

1. **Individuazione dell'ipotesi da sottoporre al test statistico**
2. **Scelta del test da utilizzare**
3. **Definizione di un livello di significatività**

Verifica delle ipotesi

Logica del test

1 Individuazione dell'ipotesi da sottoporre al test statistico

Per verificare l'ipotesi di ricerca (c'è una differenza tra i maschi e le femmine nella capacità cognitiva rilevata) dobbiamo cercare di falsificare l'ipotesi opposta: le medie delle due popolazioni sono uguali . Tale ipotesi è detta IPOTESI NULLA (H_0)

H_1 (ipotesi alternativa) è data da tutte le affermazioni non compatibili con H_0 (i maschi hanno una media maggiore/le femmine hanno una media maggiore).

In questo caso l'ipotesi di ricerca è detta bidirezionale poiché considera la differenza in termini assoluti.

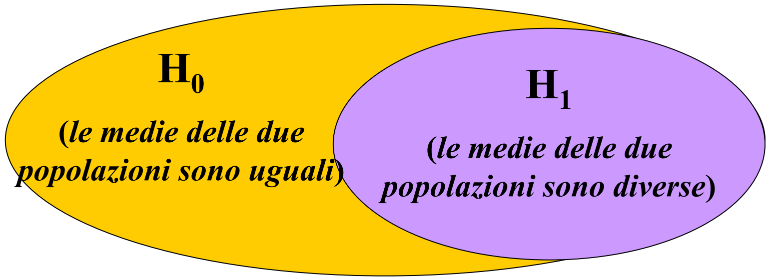
Verifica delle ipotesi

Logica del test

L' ***ipotesi nulla e l'ipotesi alternativa*** **sono mutualmente esclusive.**

Quindi o accettiamo l'una o accettiamo l'altra.

Se il ricercatore riesce a falsificare in modo probabilistico l'ipotesi nulla non resterà che rifiutarla e quindi accettare l'ipotesi alternativa



Verifica delle ipotesi

Logica del test

2 Scelta del test da utilizzare

La scelta del test dipende da 4 elementi:

a) Il tipo di ipotesi da sottoporre a verifica

b) Il numero di campioni e la relazione che esiste tra questi

c) Il livello di scala della variabile

d) La conoscenza della popolazione

Verifica delle ipotesi

Logica del test

a) Il tipo di ipotesi da sottoporre a verifica

a1) ipotesi circa la forma di una distribuzione

Ci si chiede se un insieme di osservazioni campionarie possono considerarsi provenienti da una popolazione aventi una certa distribuzione

a2) ipotesi circa l'uguaglianza di due distribuzioni

Ci si chiede se due distribuzioni sono uguali o diverse in generale senza riferirsi a particolari aspetti

a3) ipotesi relative a medie o mediane (tendenza centrale)

Ci si interroga circa l'indice di tendenza centrale che caratterizza una o più popolazioni

a4) ipotesi sulla dispersione

Ci si interroga circa l'indice di dispersione che caratterizza una o più popolazioni

Verifica delle ipotesi

Logica del test

b) Il numero di campioni e la relazione che esiste tra questi

Se il numero di campioni riguarda almeno due campioni bisogna distinguere le situazioni nelle quali i campioni sono tra loro indipendenti da quelle nelle quali non lo sono

c) Il livello di scala della/e variabile/i coinvolta/e

d) La conoscenza della popolazione

Se conosciamo la forma della distribuzione o se il campione è sufficientemente ampio è possibile ricorrere a test parametrici

Verifica delle ipotesi

Logica del test

3 Livello di errore

Possiamo commettere due tipi di errore

Errore di tipo I , di prima specie o errore α

Rifiutare un'ipotesi nulla (H_0) che in realtà è vera

Errore di tipo II , di seconda specie o errore β

Accettare un'ipotesi nulla (H_0) che in realtà è falsa

	$H_{0=vera}$	$H_{0=falsa}$
Accetto H_0	$p=1-\alpha$	$p=\beta$
Rifiuto H_0	$p=\alpha$	$p=1-\beta$

Verifica delle ipotesi

Logica del test

3 Livello di errore

I due livelli di errore sono inversamente proporzionali e non è possibile controllarli entrambi

Operativamente si fissa a priori un livello di probabilità α in genere pari a 0,05 o a 0,01

La probabilità di commettere un errore α è uguale alla somma delle probabilità dei risultati che ci fanno respingere l'ipotesi nulla, ossia di quei risultati che rientrano nella regione di rifiuto

$1 - \beta$ = potenza del test

Verifica delle ipotesi

Test binomiale

Test Binomiale		
Natura dell'ipotesi	**Test su una probabilità**	
Tipo di test	**Parametrico**	
Tipo di variabile	**Dicotomica**	
Numero di campioni	**1**	
H_0	$P=P_0$	
H_1	$P \# P_0$ **(bidirezionale)**	
Statistica Test	**Numero di successi**	

Verifica delle ipotesi

Test binomiale

Esempio: sospettando che una moneta sia truccata si vuole sottoporre a verifica l'ipotesi (H_0) che la probabilità di ottenere testa (p) sia uguale a quello di ottenere croce =0,5 (p_0)

H_0 = **P=0,5**

H_1 = **P > 0,5 o P < 0,5**

Lanciamo la moneta dieci volte (N=10) e registriamo il numero di successi ottenuti (uscite testa)

Decidiamo un valore α pari a 0,05

Confrontiamo il valore ottenuto con la distribuzione che otterremmo se fosse vera l'ipotesi nulla

Verifica delle ipotesi

Test binomiale

La probabilità che l'evento si presenti esattamente k volte in n prove è data da

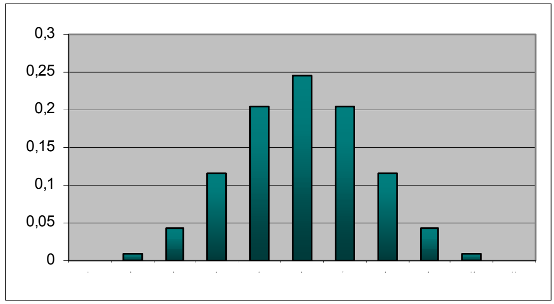
$$P_k = \binom{N}{k} p^k q^{n-k}$$

Verifica delle ipotesi

Test binomiale

Esempio: probabilità di ottenere n volte testa, lanciando una moneta 10 volte

n=10	p=0,5
k	p
0	0,001
1	0,01
2	0,044
3	0,117
4	0,205
5	0,246
6	0,205
7	0,117
8	0,044
9	0,01
10	0,001



Verifica delle ipotesi

Test binomiale

Poiché è stato fissato un livello di errore α pari a 0,05 dobbiamo determinare la regione di rifiuto

VALORI SOGLIA

n=10	p=0,5	
k	p	
0	0,001	
1	0,01	0,011
2	0,044	
3	0,117	
4	0,205	
5	0,246	
6	0,205	
7	0,117	
8	0,044	
9	0,01	
10	0,001	0,011

Probabilità di rifiutare l'ipotesi nulla avendo osservato un valore con essa compatibile

Ovvero probabilità di commettere l'errore alfa

Regione di rifiuto

0,022

Verifica delle ipotesi

Test binomiale

Se il risultato ottenuto cade nell'area di rifiuto rifiuteremo l'ipotesi nulla affermando che la probabilità dei due eventi non è uguale, viceversa se il valore cadrà fuori dall'area di rifiuto accetteremo l'ipotesi nulla affermando che le due probabilità non sono significativamente differenti

Verifica delle ipotesi

Test del chi quadro per la bontà dell'adattamento

Test Chi quadro		
Natura dell'ipotesi	**Test su forma distribuzionale**	
Tipo di test	**Non Parametrico**	
Tipo di variabile	**Categoriale**	
Numero di campioni	**1**	
H_0	**La distribuzione ha una certa forma**	
H_1	**La distr. Non ha la forma specificata**	
Statistica Test	**Test χ^2**	

Verifica delle ipotesi

Test del chi quadro per la bontà dell'adattamento

Lo scopo di questo test è di valutare se esistono differenze significative tra le distribuzioni ottenute empiricamente e quelle che ci saremmo aspettate come conseguenza logica della presenza di alcuni requisiti

Es: uno psicologo del lavoro vuole verificare se la scelta di un prodotto è legata alla marca .

Chiede a 100 soggetti di scegliere la preferita tra 4 maglie identiche differenziate solamente dall'etichetta

Le marche sono connotate come A, B, C, D

Verifica delle ipotesi

Test del chi quadro per la bontà dell'adattamento

Se la marca non influenza i soggetti, la distribuzione delle scelte dovrebbe premiare ugualmente tutte le marche

$$\boxed{H_0 = P(A) = P(B) = P(C) = P(D)}$$

Se la marca influenza i soggetti, la distribuzione delle scelte dovrebbe premiare in modo differente almeno una delle quattro marche

$$\boxed{H_1 = P(A) \neq P(B) \neq P(C) \neq P(D)}$$

Verifica delle ipotesi

Test del chi quadro per la bontà dell'adattamento

Le scelte dei soggetti sono rappresentate nella tabella di frequenza

Marca	F
A	3
B	10
C	24
D	63
totale	100

FREQUENZE OSSERVATE n_k

Sotto l'ipotesi nulla si sarebbe dovuta osservare un sistema di scelte equidistribuito

Marca	F
A	25
B	25
C	25
D	25
Totale	100

FREQUENZE TEORICHE $\hat{n}_k$

Verifica delle ipotesi

Test del chi quadro per la bontà dell'adattamento

Per verificare la significatività della differenza tra la distribuzione ottenuta e quella ipotizzata si calcola il Chi quadro

$$X^2 = \sum_{k=1}^{K} \frac{(n_k - \hat{n}_k)^2}{\hat{n}_k}$$

Verifica delle ipotesi

Test del chi quadro per la bontà dell'adattamento

Scelta del livello di errore alfa

Esempio alfa=0,05

ν = k-1 =3

p=(1-alfa)	0.010	0.022	0.050	(…)	0.950	0.975	0.990
g							
1							
2							
3					**7.815**		
4							
5							
6							
(…)							

Verifica delle ipotesi

Test del chi quadro per la bontà dell'adattamento

Marca	F(osservate)	F(attese)	Fo - Fa	(Fo - Fa)2	(Fo - Fa)2/Fa
A	3	25	-22	484	19.36
B	10	25	-15	225	9
C	24	25	-1	1	0.04
D	63	25	38	1444	57.76
totale	100	100	0	0	0
				Chi quadro	86.16

Chi quadro osservato = 86,160

Chi quadro critico = 7,815

Poiché il valore della statistica test osservato è maggiore del valore critico respingiamo l'ipotesi nulla

Verifica delle ipotesi

Test della media su un solo campione

Test Media	
Natura dell'ipotesi	**Test sulla media di una popolazione**
Tipo di test	**Parametrico**
Tipo di variabile	**cardinale**
Numero di campioni	**1**
H_0	$\mu = \mu_0$
H_1	$\mu < \mu_0$ $\quad$ $\mu > \mu_0$
Statistica Test	**z**
Assunti	**Campione casuale distribuzione normale**

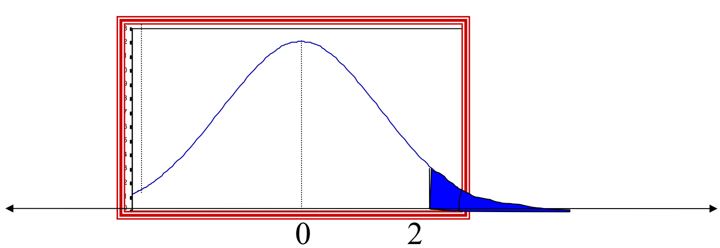
Verifica delle ipotesi

Test della media su un solo campione

Se gli assunti sono veri la distribuzione campionaria della media si distribuisce normalmente con media μ_0 e varianza σ^2/n .

Standardizzando la distribuzione si avrà media=0 e varianza =1

Sappiamo quindi ad esempio che la probabilità di osservare un valore maggiore di 2 è pari a 0.5000 – 0.4772 =0.0228

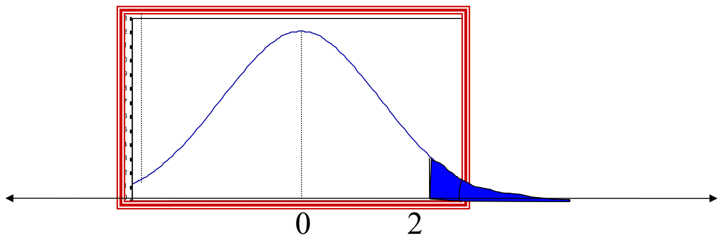


Verifica delle ipotesi

Test della media su un solo campione

Se dunque l'ipotesi nulla H_0 è vera possiamo, standardizzando il valore della media campionaria, calcolare la probabilità che tale valore (o un valore più estremo) appartenga alla distribuzione campionaria delle medie generata da quella popolazione .

Tale valore di probabilità rappresenta la probabilità di commettere un errore di tipo I (errore α) rifiutando l'ipotesi nulla.



Verifica delle ipotesi

Test della media su un solo campione

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Una volta calcolato z si deve confrontare il valore ottenuto con il valore z_c (valore di z individuato in relazione al livello di errore α - *ipotesi monodirezionale* o al valore $\alpha/2$ – ipotesi bidirezionale)

Verifica delle ipotesi

Test della media su un solo campione

Esempio

La produttività degli operai di una fabbrica è in media di 1800 pezzi a testa al giorno con una varianza pari a 10000 . Uno psicologo del lavoro sottopone un campione di 50 soggetti ad un corso di formazione che mira ad aumentare la capacità degli operai al lavoro di squadra. Dopo una settimana di corso registra l'attività produttiva di questi 50 soggetti, osservando che la produzione media è pari a 1850 unità . Si può ritenere l'intervento formativo efficace ad un livello di significatività dello 0.01 ?

Verifica delle ipotesi

Test della media su un solo campione

Esempio

$$z = \frac{1850 - 1800}{100/\sqrt{50}} = 3.55$$

z	0	1	2	3
0.0				
0.2				
…				
2.3				.4901

Ipotesi monodirezionale

Z_c=2.33

Verifica delle ipotesi

Test della media su un solo campione

Esempio

Poiché il valore z osservato 3.55 è maggiore del valore critico 2.33 possiamo rifiutare l'ipotesi nulla e sostenere che l'intervento formativo ha avuto degli effetti positivi sul livello di produttività degli operai

Verifica delle ipotesi

Test sulla varianza in un solo campione

Natura dell'ipotesi	**Test sulla varian di una popolazione**
Tipo di test	**Parametrico**
Tipo di variabile	**cardinale**
Numero di campioni	**1**
H_0	$\sigma^2 = \sigma^2_0$
H_1	$\sigma^2 < \sigma^2_0$ $\sigma^2 > \sigma^2_0$
Statistica Test	**z**
Assunti	**Campione casuale distribuzione normale**

Verifica delle ipotesi

Test sulla varianza in un solo campione

Se gli assunti sono rispettati e l'ipotesi nulla è vera il rapporto

$$G = \frac{(n-1)\,S^2}{\sigma^2}$$

È una variabile χ^2 con n-1 gradi di libertà

Ipotesi bidirezionale

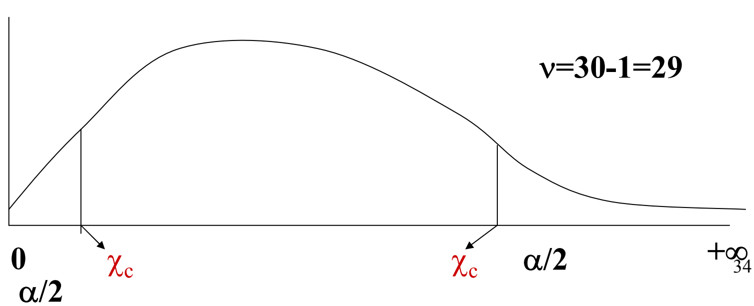
Verifica delle ipotesi

Test sulla varianza in un solo campione

Esempio

Ritornando all'esempio precedente, immaginiamo che la varianza della produzione di 30 operai sottoposti al corso di formazione sia=9500 , mentre quella della popolazione sia =10000

Poniamo alfa=0.05



Verifica delle ipotesi

Test sulla varianza in un solo campione

Esempio

χ_c=16 ⟷ χ_c=45.7

$$G = \frac{(29)9500}{10000} = \frac{275000}{10000} = 27{,}5$$

Poiché G <u>non</u> cade nell'area di rifiuto non possiamo rifiutare l'ipotesi nulla e dobbiamo sostenere che il corso <u>non</u> ha modificato la variabilità delle prestazioni