

Test d'Ipotesi /2

TIPICI PROBLEMI DI VERIFICA DI IPOTESI SONO:

- Test per la media
- Test per una proporzione
- Test per la varianza
- Test per due campioni indipendenti
- Test di indipendenza

2

Test per la media (σ noto)

Test per media – **Popolazione Normale** – **Varianza nota**

Statistica test e sua distribuzione sotto l'ipotesi nulla:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

con μ_0 che indica il valore della media ipotizzato in H_0

Ipotesi alternativa	Regione di rifiuto
$H_1 : \mu > \mu_0$	$Z \geq z_\alpha$
$H_1 : \mu < \mu_0$	$Z \leq -z_\alpha$
$H_1 : \mu \neq \mu_0$	$ Z \geq z_{\alpha/2}$

Interpretazione

Nella statistica test precedente il numeratore ci dice di quanto la media (campionaria) osservata differisce dalla media ipotizzata μ_0 (quindi quando l'ipotesi nulla è vera). Al denominatore troviamo l'errore standard della media campionaria

➡ Z ci dice per quanti errori $\bar{X}$ standard differisce da μ_0 : ossia ci indica se lo scostamento osservato rientra nella variabilità media dello stimatore o se invece è troppo grande da poter essere giustificato dalla variabilità campionaria

Test per la media

Test per media – **Popolazione Normale** – **Varianza ignota**

Statistica test e sua distribuzione sotto l'ipotesi nulla:

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t - Student(n-1)$$

con μ_0 che indica il valore della media ipotizzato in H_0
e con S che indica la radice quadrata dello stimatore corretto della varianza σ^2

Ipotesi alternativa	Regione di rifiuto
$H_1 : \mu > \mu_0$	$T \geq t_\alpha$
$H_1 : \mu < \mu_0$	$T \leq -t_\alpha$
$H_1 : \mu \neq \mu_0$	$ T \geq t_{\alpha/2}$

Test per la media

Test per media – **Popolazione non-Normale** – **Varianza ignota**

Statistica test e sua distribuzione sotto l'ipotesi nulla, al tendere di n a infinito:

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

con μ_0 che indica il valore della media ipotizzato in H_0
e con S che indica la radice quadrata dello stimatore corretto della varianza σ^2

Ipotesi alternativa	Regione di rifiuto
$H_1 : \mu > \mu_0$	$Z \geq z_\alpha$
$H_1 : \mu < \mu_0$	$Z \leq -z_\alpha$
$H_1 : \mu \neq \mu_0$	$ Z \geq z_{\alpha/2}$

Test per la media - esempio

Test sulla statura media in un collettivo:

$$H_0 : \mu = 175 \text{ contro } H_1 : \mu > 175$$

Si assume che la statura sia una variabile casuale $X \sim N(\mu, \sigma^2)$ con varianza ignota.

Si estrae un campione di 10 giovani e si trova:

$$\bar{x} = 181,5$$

$$S^2 = 95,5067$$

$$t = \frac{181,5 - 175}{\sqrt{95,5067/10}} = 2,103$$

Ponendo $\alpha = 0,05$ si ottiene dalla t-Student con 9 gradi di libertà:

$$t_{0,05} = 1,8331$$

e quindi si rifiuterà l'ipotesi nulla poiché $t \geq 1,8331$.

Il p-value corrispondente al valore osservato della statistica test è $P(T \geq 2.103 / H_0 \text{ è vera}) = 0,0324$. $\implies$ forte evidenza empirica contro l'ipotesi nulla la quale potrebbe essere accettata soltanto ad un livello di significatività α minore di 0.0324

7 Test per la media – dimensione campionaria

Può essere d'interesse determinare n in maniera tale da garantirsi un test che raggiunga una certa potenza sotto una specificata ipotesi alternativa. Si determina n in funzione di valori obiettivo degli errori α e β (oppure $1-\beta$). Ipotizziamo comunque una n 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 $P(Z \geq z_\alpha) = \alpha$

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

allora:

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

8

Per una ipotesi alternativa unidirezionale

(vedi figura 14.2.2 pag 347 del libro)

I valori critici sono anche uguali rispettivamente nei due casi: $c = \mu_1 - z_\beta \sigma / \sqrt{n}$ e $c = \mu_0 + z_\alpha \sigma / \sqrt{n}$

Consideriamo le distanze: $|\mu_1 - c| = z_\beta \sigma / \sqrt{n}$ e $|\mu_0 - c| = z_\alpha \sigma / \sqrt{n}$

da qui si ricava

$$|\mu_1 - \mu_0| = (z_\beta + z_\alpha) \sigma / \sqrt{n}$$

E quindi n .

9

Esempio

Sempre nel caso del problema dell'altezza siano :

$$H_0 : \mu = 175 \text{ contro } H_1 : \mu > 175$$

In maniera tale che la potenza del test sia almeno $(1-\beta)=0.9$ quando $\mu=178$, con $\alpha=0.05$. Sia una stima iniziale di $\sigma=15.5$.

In tale caso $z_\alpha=1.645$ $z_\beta=1.282$

$$n = \left[\frac{15.5(1.645+1.282)}{178-175} \right]^2 = 228.7$$

per raggiungere la potenza voluta la dimensione campionaria non deve essere inferiore a 229!

Test per una proporzione

Test per una proporzione – **Popolazione Bernoulliana**

Statistica test e sua distribuzione sotto l'ipotesi nulla, al tendere di n a infinito:

$$Z = \frac{\bar{X} - \pi_0}{\sqrt{\pi_0(1 - \pi_0)/n}} \sim N(0, 1)$$

con π_0 che indica il valore della proporzione ipotizzato in H_0

Ipotesi alternativa	Regione di rifiuto
$H_1 : \pi > \pi_0$	$Z \geq z_\alpha$
$H_1 : \pi < \pi_0$	$Z \leq -z_\alpha$
$H_1 : \pi \neq \pi_0$	$ Z \geq z_{\alpha/2}$

Test per una proporzione - esempio

Si vuole verificare che nel 2002 la percentuale degli occupati in Italia nel settore agricolo è la stessa del 1991 pari a 8,4%:

$$H_0 : \pi = 0,084 \text{ contro } H_1 : \pi > 0,084$$

Si estrae un campione di 1000 occupati.

Dei mille estratti, 53 sono occupati nel settore agricolo, pertanto:
Il valore della statistica test è: $\bar{x} = 0,053$

Ponendo $\alpha = 0,01$ si ottiene dalla normale standardizzata:

$$z = \frac{0,053 - 0,084}{\sqrt{(0,084)(0,916)/1000}} = -3,534 \qquad -z_{0,01} = -2,326$$

e quindi si rifiuterà l'ipotesi nulla poiché $z < -2,326$.

Il p-value corrispondente al valore osservato della statistica test è 0,0002. Tale valore mostra una forte evidenza contro l'ipotesi nulla.

Test per la varianza

Test per la varianza – **Popolazione Normale** – **media ignota**

Statistica test e sua distribuzione sotto l'ipotesi nulla:

$$\frac{(n-1)S^2}{\sigma_0^2} \sim \text{Chi-quadra to}(n-1)$$

con σ_0^2 che indica il valore della varianza ipotizzato in H_0 e con S^2 che indica lo stimatore corretto della varianza.

Ipotesi alternativa	Regione di rifiuto
$H_1 : \sigma^2 > \sigma_0^2$	$(n-1)S^2 / \sigma_0^2 \geq \chi_{\alpha}^2$
$H_1 : \sigma^2 < \sigma_0^2$	$(n-1)S^2 / \sigma_0^2 \leq \chi_{1-\alpha}^2$
$H_1 : \sigma^2 \neq \sigma_0^2$	$(n-1)S^2 / \sigma_0^2 \leq \chi_{1-\alpha/2}^2$ $(n-1)S^2 / \sigma_0^2 \geq \chi_{\alpha/2}^2$

Test per la varianza - esempio

Si ipotizzi che la spesa delle famiglie sia una variabile casuale Normale. Si vuole verificare:

$$H_0 : \sigma^2 = 22500 \text{ contro } H_1 : \sigma^2 \neq 22500$$

Si estrae un campione di 61 famiglie.

Dal campione, si osservano: $\bar{x} = 2010$ $S^2 = 20000$

Il valore della statistica test è: $\frac{60 \cdot 20000}{22500} = 53,33$

Ponendo $\alpha = 0,05$ si ottiene dal Chi-quadrato con $n-1=60$ gradi di libertà:

$$\chi_{0,975}^2 = 40,482 \qquad \chi_{0,025}^2 = 83,298$$

quindi il valore osservato non cade nella regione di rifiuto.

Pertanto non si rifiuterà l'ipotesi nulla, ossia i dati osservati non sono difforni all'ipotesi che la varianza è pari a 22500.

14 Il caso di due Popolazioni

E' il caso di due popolazioni diverse X_1 ed X_2 (Normali o Bernoulliane) ed il processo inferenziale si basa sui dati provenienti da due campioni (indipendenti) diversi di dimensione n_1 ed n_2 estratti dalle due popolazioni. Tre sono i casi:

1. Confronto tra le due medie
2. Confronto tra le due proporzioni
3. Confronto tra le due varianze

Questo ultimo caso è molto importante per le aziende impegnate a diminuire la variabilità di processi per assicurare elevata qualità per unità prodotta.

Test per due campioni indipendenti - medie

Test per medie – **Popolazioni Normali** – **varianze note**

Statistica test e distribuzione sotto l'ipotesi nulla: $H_0 : \mu_1 = \mu_2$

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0,1)$$

con σ_1^2 e σ_2^2 che indicano il valore delle varianze nelle due popolazioni e con n_1 e n_2 le dimensioni dei due campioni.

Ipotesi alternativa	Regione di rifiuto
$H_1 : \mu_1 > \mu_2$	$Z \geq z_\alpha$
$H_1 : \mu_1 < \mu_2$	$Z \leq -z_\alpha$
$H_1 : \mu_1 \neq \mu_2$	$ Z \geq z_{\alpha/2}$

E' immediato comprendere la precedente statistica test perché si riconduce a quello per una media; infatti:

$$H_0: \mu_1 = \mu_2 \iff H_0: \mu_\delta = \mu_1 - \mu_2 = 0$$

Nel caso in cui le due varianze siano note, per stimare μ_δ si utilizza il seguente stimatore caratterizzato dalla sua distribuzione sotto l'ipotesi nulla:

$$\bar{X}_1 - \bar{X}_2 \square N\left(0, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

Facile da verificare!

Test per due campioni indipendenti - medie

Test per medie – Popolazioni Normali – varianze ignote e uguali

Statistica test e distribuzione sotto l'ipotesi nulla: $H_0 : \mu_1 = \mu_2$

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p^2(1/n_1 + 1/n_2)}} \sim t\text{-Student}(n_1 + n_2 - 2)$$

con S_p^2 che indica lo stimatore congiunto della varianza, ossia:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Ipotesi alternativa	Regione di rifiuto
$H_1 : \mu_1 > \mu_2$	$T \geq t_\alpha$
$H_1 : \mu_1 < \mu_2$	$T \leq -t_\alpha$
$H_1 : \mu_1 \neq \mu_2$	$ T \geq t_{\alpha/2}$

Test per due campioni indipendenti - medie

Test per medie – Popolazioni Normali – var. ignote ma no uguali

**Statistica test e distribuzione sotto l'ipotesi nulla: $H_0 : \mu_1 = \mu_2$,
al tendere di n_1 e n_2 a infinito:**

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \sim N(0,1)$$

con S_1^2 e S_2^2 che indicano gli stimatori corretti delle varianze.

Ipotesi alternativa	Regione di rifiuto
$H_1 : \mu_1 > \mu_2$	$Z \geq z_\alpha$
$H_1 : \mu_1 < \mu_2$	$Z \leq -z_\alpha$
$H_1 : \mu_1 \neq \mu_2$	$ Z \geq z_{\alpha/2}$

19 Test per due campioni indipendenti - esempio

Test sull'efficacia di un trattamento per il controllo della pressione sanguigna. Due popolazioni: la prima formata dagli individui trattati e la seconda formata da individui non trattati:

$$H_0 : \mu_1 = \mu_2 \text{ contro } H_1 : \mu_1 < \mu_2$$

Si suppone che la pressione in entrambe le popolazioni sia una variabile casuale Normale con varianza nota pari a 200:

Due campioni: $n_1=15$ individui trattati e $n_2=12$ individui non trattati:

$$\bar{x}_1 = 135,87 \quad \bar{x}_2 = 170,58$$

Il valore della statistica test è:
$$z = \frac{135,87 - 170,58}{\sqrt{(200/15) + (200/12)}} = -6,34$$

Ponendo $\alpha = 0,01$ si ottiene il valore della normale standardizzata:

$$-z_{0,01} = -2,326$$

quindi si deve rifiutare l'ipotesi nulla.

Test per due campioni indipendenti - varianze

Test per varianze – Popolazioni Normali

Statistica test e distribuzione sotto l'ipotesi nulla: $H_0 : \sigma_1^2 = \sigma_2^2$

$$\frac{S_1^2}{S_2^2} \sim f - \text{Fisher}(n_1 - 1, n_2 - 1)$$

con S_1^2 e S_2^2 che indicano gli stimatori corretti delle varianze.

Ipotesi alternativa	Regione di rifiuto
$H_1 : \sigma_1^2 > \sigma_2^2$	$S_1^2 / S_2^2 \geq f_\alpha$
$H_1 : \sigma_1^2 < \sigma_2^2$	$S_1^2 / S_2^2 \leq f_{1-\alpha}$
$H_1 : \sigma_1^2 \neq \sigma_2^2$	$S_1^2 / S_2^2 \leq f_{1-\alpha/2}$ $S_1^2 / S_2^2 \geq f_{\alpha/2}$

Anche la precedente statistica è di facile comprensione. Il test di uguaglianza tra varianze di due P indipendenti (ipotesi di **omoschedasticità**) può essere espresso come:

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = \sigma_R^2 = 1 \quad \text{contro} \quad \begin{cases} \sigma_R^2 > 1 \\ \sigma_R^2 < 1 \\ \sigma_R^2 \neq 1 \end{cases}$$

ed è immediato che la statistica test corrispondente da utilizzare sarà:

$$\frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} = \frac{S_1^2}{S_2^2} \sim F(n_1 - 1; n_2 - 1) \quad (\text{sotto } H_0 \text{ ed ipotesi di normalità})$$

Test per due campioni indipendenti - proporzioni

Test per proporzioni – **Popolazioni Bernoulliane**

Statistica test e distribuzione sotto l'ipotesi nulla: $H_0 : \pi_1 = \pi_2$,
al tendere di n_1 e n_2 a infinito:

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\bar{X}_p(1 - \bar{X}_p)(1/n_1 + 1/n_2)}} \sim N(0,1)$$

con $\bar{X}_p$ lo stimatore congiunto della proporzione:

$$\bar{X}_p = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2}$$

Ipotesi alternativa	Regione di rifiuto
$H_1 : \pi_1 > \pi_2$	$Z \geq z_\alpha$
$H_1 : \pi_1 < \pi_2$	$Z \leq -z_\alpha$
$H_1 : \pi_1 \neq \pi_2$	$ Z \geq z_{\alpha/2}$

23 Test per due campioni indipendenti - esempio

Si vuole comparare il tasso di occupazione π_1 dei laureati in Legge con il tasso di occupazione π_2 dei laureati in Economia:

$$H_0 : \pi_1 = \pi_2 \text{ contro } H_1 : \pi_1 < \pi_2$$

Sono estratti due campioni di persone laureate nell'anno precedente:

$$n_1 = 80 \quad \text{laureati in Legge} \quad n_2 = 120 \quad \text{laureati in Economia}$$

risultano rispettivamente 59 occupati nel primo campione e 93 nel secondo. Pertanto: $\bar{x}_1 = 0,738$ $\bar{x}_2 = 0,775$ $\bar{x}_p = 0,76$

$$z = \frac{0,738 - 0,775}{\sqrt{0,76(1 - 0,76)(1/80 + 1/120)}} = -0,6$$

Ponendo $\alpha = 0,05$ si ottiene il valore della normale standardizzata:

$$-z_{0,05} = -1,645$$

e l'ipotesi nulla non può essere rifiutata.

Test di indipendenza (non parametrico)

Il test di indipendenza permette di verificare se tra due variabili sussiste o meno associazione. Il test può essere applicato sia su variabili quantitative (suddivise in classi) sia su variabili qualitative.

Date due variabili X e Y rispettivamente con H e K modalità, nel caso di **indipendenza** la probabilità p_{ij} di osservare congiuntamente la modalità i -esima della X e la modalità j -esima della Y è data da

$$p_{ij} = p_i p_j \quad \text{(si ricordi che in ambito "descrittivo" era: } n_{ij} = \frac{n_{i.} \times n_{.j}}{n} \text{)}$$

e dividendo per n : $f_{ij} = f_i f_j$)

Quindi, l'ipotesi nulla d'indipendenza è:
contro l'ipotesi alternativa:

$$H_0 : p_{ij} = p_i p_j$$

$$H_1 : p_{ij} \neq p_i p_j$$

La statistica test utilizzata è:

$$\chi^2 = \sum_{i=1}^H \sum_{j=1}^K \frac{(n_{ij} - n'_{ij})^2}{n'_{ij}}$$

che, sotto l'ipotesi nulla, si distribuisce come un Chi-quadrato con $(H-1)(K-1)$ gradi di libertà.

Ipotesi alternativa	Regione di rifiuto
$H_1 : p_{ij} \neq p_i p_j$	$\chi^2 \geq \chi^2_{\alpha}$

Test di indipendenza - esempio

In un'indagine di marketing (139 interviste) si è chiesto di indicare la preferenza tra 3 alimenti liquidi e 3 alimenti solidi da consumare al mattino a colazione:

	Biscotti	Merendina	Fette biscottate	Totale
Caffelatte	45	8	5	58
Tè	7	5	31	43
Succo di frutta	5	27	6	38
Totale	57	40	42	139

La statistica test vale: $\chi^2 = 101,57$ (calcolando l'indice chi-quadrato)

Ponendo $\alpha = 0,01$ il valore critico per la distribuzione Chi-quadrato con $(3-1)(3-1)=4$ gradi di libertà è

$$\chi_{0,01}^2 = 13,227$$

Si rifiuta quindi l'ipotesi nulla di indipendenza tra i due caratteri.