Università degli Studi di Napoli "Parthenope"

Corso di Laurea in Statistica e Informatica per l'Azienda, la Finanza e le Assicurazioni (SIAFA)

STATISTICA
II MODULO

Sergio LONGOBARDI
longobardi@uniparthenope.it

Test d'ipotesi

Test d'ipotesi

IPOTESI STATISTICA: una congettura riguardante un parametro θ della popolazione (ad esempio la media o la varianza).

Esempio:

Secondo il costruttore di un certo tipo di batterie per autovetture, la durata media è di 3400 ore.

Un cliente, per verificarne la durata, osserva un campione di 30 batterie e formula due ipotesi:

Ipotesi 1 (H_0): le batterie hanno durata media di almeno 3400 ore

Ipotesi 2 (H_1): le batterie hanno durata media inferiore a 3400 ore

3

Test d'ipotesi

Obiettivo:

Attraverso un campione di osservazioni stabilire, con un certo grado di attendibilità, se poter rifiutare o meno l'ipotesi nulla a favore dell'ipotesi alternativa relativa ad un determinato parametro della popolazione

Se l'ipotesi riguarda uno o più parametri della distribuzione di probabilità della popolazione, si parlerà di test parametrico.

L'impostazione data da J.Neyman e E.S.Pearson, nota come test d'ipotesi, prevede la formulazione di un'ipotesi nulla e un'ipotesi alternativa.

4

Test d'ipotesi

In generale seguendo l'approccio di Neyman-Pearson , nei test d'ipotesi si distinguono due ipotesi contrapposte:

- ipotesi nulla, indicata con H_0
- ipotesi alternativa, indicata con H_1

L'ipotesi nulla è preesistente all'osservazione dei dati campionari, ritenuta vera fino a prova contraria.

5

Test d'ipotesi

Ad esempio:

$$\begin{cases} H_0: & \mu \geq 3400 \\ H_1: & \mu < 3400 \end{cases}$$

Generalizzando:

Θ è lo spazio parametrico, ossia l'insieme di tutti i possibili valori che può assumere θ

$\boldsymbol{\Theta_0}$ e $\boldsymbol{\Theta_1}$ sono i sottospazi che formano una partizione dello spazio parametrico.

Indichiamo le due ipotesi con il seguente sistema:

$$\begin{cases} H_0: & \theta \in \Theta_0 \\ H_1: & \theta \in \Theta_1 \end{cases}$$

6

Test d'ipotesi

Un'ipotesi può essere:

- ***semplice:*** quando specifica un singolo valore numerico
- ***composta:*** quando specifica un intervallo di valori

Esempio:

H_0:μ=5 è un'ipotesi semplice

H_0:μ>5 è un ipotesi composta

Un'ipotesi composta può essere:

- ***unidirezionale:*** quando specifica un intervallo di valori
- ***bidirezionale:*** quando specifica due intervalli di valori

Esempio:

H_0:μ>7 è un ipotesi composta ***unidirezionale***,

H_0:μ≠7 è un'ipotesi composta ***bidirezionale***

7

Test d'ipotesi

I sistemi di ipotesi più frequentemente utilizzati sono i seguenti:

$$\begin{cases} H_0: & \theta = \theta_0 \\ H_1: & \theta > \theta_0 \end{cases} \qquad \begin{cases} H_0: & \theta = \theta_0 \\ H_1: & \theta < \theta_0 \end{cases}$$

$$\begin{cases} H_0: & \theta = \theta_0 \\ H_1: & \theta \neq \theta_0 \end{cases}$$

Dove θ_0 è un valore fissato del parametro.

8

Test d'ipotesi

Un test statistico è una regola che permette di discriminare i campioni che portano all'accettazione dell'ipotesi nulla da quelli che portano al suo rifiuto.

Il test si basa sul valore assunto da una statistica test.

L'insieme dei valori della statistica test che portano all'accettazione dell'ipotesi nulla è chiamata regione di accettazione.

L'insieme dei valori della statistica test che portano al rifiuto dell'ipotesi nulla è chiamata regione di rifiuto.

La statistica test è una statistica campionaria la cui distribuzione deve essere completamente nota sotto l'ipotesi nulla.

9

Test d'ipotesi

Nel caso della media *μ, si considera* il corrispondente stimatore $\overline{X}$.

Esempio :

H_0:*μ=10*

H_1:*μ≠10*

Se H_0:*μ=10 è vera, ci aspettiamo che lo* stimatore $\overline{X}$ assuma un valore "vicino" a *10 (esempio 9,5) :*

- valori di $\overline{X}$ (stime campionarie) "abbastanza vicini" a *10 portano ad accettare* H_0
- valori di $\overline{X}$ "molto distanti" da *10 portano* a rifiutare H_0 e ad accettare H_1:*μ≠10*

10

Test d'ipotesi

In generale:

Per il parametro θ ,*consideriamo* il corrispondente stimatore T

Se H_0 : $\theta = \theta_0$ *è vera, ci aspettiamo che lo* stimatore T assuma un valore "vicino" a θ_0 *:*

- valori di T (stime campionarie) "abbastanza vicini" a θ_0 *portano ad accettare* H_0
- valori di T "molto distanti" da θ_0 *portano* a rifiutare H_0 e ad accettare H_1: $\theta \neq \theta_0$

11

Test d'ipotesi

Le affermazioni "abbastanza vicino" e "molto distante" devono essere definite sulla base della distribuzione dello stimatore T, supponendo vera H_0

Questa distribuzione è suddivisa in due regioni (o aree):
regione A di Accettazione
regione R di Rifiuto (o critica)

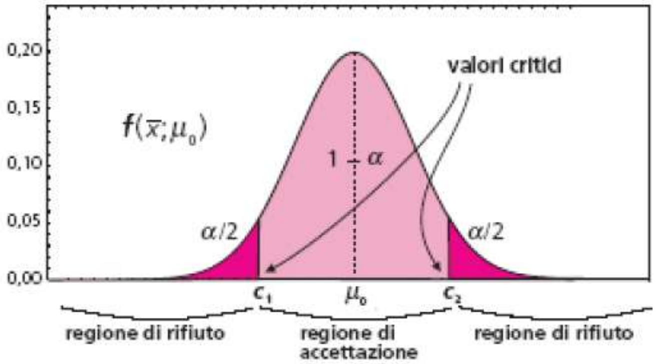
12

Test d'ipotesi

Esempio: Supponiamo che la popolazione sia Normale con media μ incognita e varianza σ nota. Si vuole verificare:

$$\begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu \neq \mu_0 \end{cases}$$

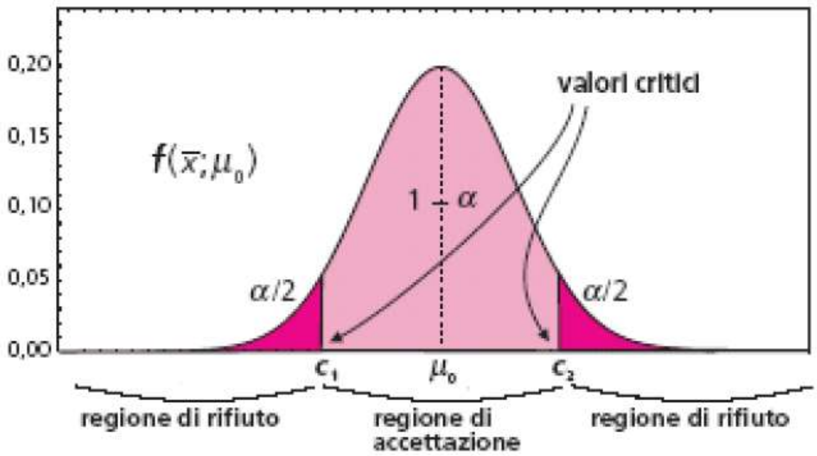
Considerando come statistica test la media campionaria sappiamo che sotto l'ipotesi nulla questa si distribuisce come una Normale con media $\boldsymbol{\mu = \mu_0}$ e varianza: $\boldsymbol{\sigma^2/n}$.



13

Test d'ipotesi

Dalla figura si può vedere che i valori critici definiscono la zona di accettazione e che dipendono dal livello di significatività α : maggiore è il suo valore, più ampia sarà la regione di rifiuto.

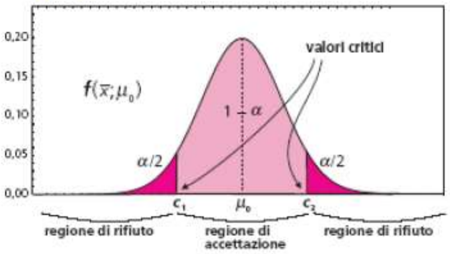


14

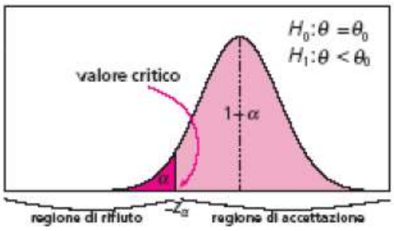
Test d'ipotesi

In corrispondenza dell'ipotesi alternativa si possono configurare diverse regioni di rifiuto:

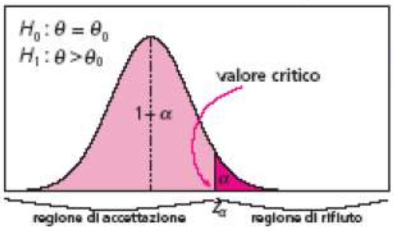
$$\begin{cases} H_0: & \theta = \theta_0 \\ H_1: & \theta \neq \theta_0 \end{cases}$$



$$\begin{cases} H_0: & \theta = \theta_0 \\ H_1: & \theta < \theta_0 \end{cases}$$



$$\begin{cases} H_0: & \theta = \theta_0 \\ H_1: & \theta > \theta_0 \end{cases}$$



15

Test d'ipotesi

Passi per la verifica di ipotesi

1. Definizione del sistema di ipotesi
2. Scelta della statistica test
3. Scelta del livello α di significatività del test e della numerosità campionaria n
4. Definizione della regione di rifiuto
5. Estrazione del campione
6. Calcolo del valore della statistica test sulla base dei dati campionari
7. Si prende la decisione

16

Test d'ipotesi

Si ipotizza di testare la seguente ipotesi relativamente alla media di una popolazione la cui varianza è pari a 25

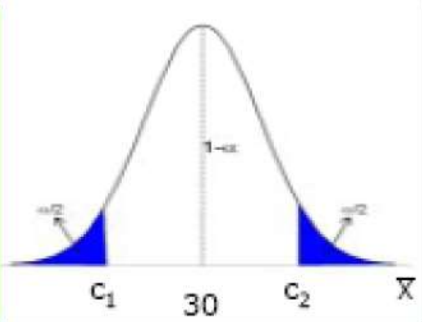
H_0: μ=30

H_1: μ≠30

Si estrae un campione di ampiezza n=150 e la media campionaria risulta pari a 29. Se è vera H_0 (cioè μ =30)

$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n}) \rightarrow \overline{X} \sim N(30, \frac{25}{150})$$

Si dovrebbero calcolare i valori c_1 e c_2, se la media campionaria "cade" all'interno di questo intervallo si può accettare l'ipotesi H_0.

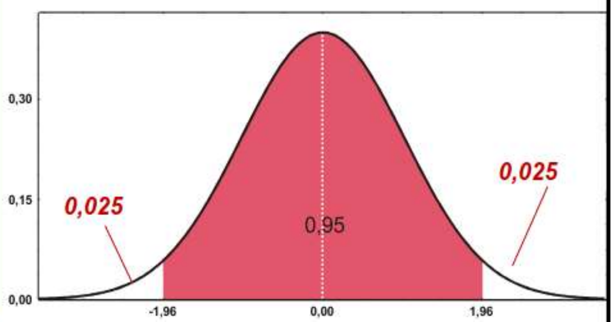


A questo scopo si ricorre alla normale standardizzata Z

17

Test d'ipotesi

Passando alla Z, la regione di rifiuto con un a=5% è costituita dai valori superiori a 1,96 e inferiori a -1,96



Quindi se il valore standardizzato della media campionaria (statistica test) risulta all'interno dell'intervallo si accetterà l'ipotesi nulla

H_0: μ=30

H_1: μ≠30

$$\overline{X} \sim N(30, \frac{25}{150}) \rightarrow Z = \frac{\overline{X} - 30}{\sqrt{25/150}} = -2,44$$

Rifiuto H_0

18

Test d'ipotesi

Test per media – Popolazione Normale – Varianza nota

Statistica test e sua distribuzione sotto l'ipotesi nulla:

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

con μ_0 che indica il valore della media ipotizzato in H_0

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

19

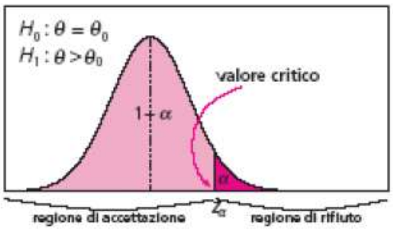
Test d'ipotesi

$$H_0: \ \mu = \mu_0$$
$$H_1: \ \mu > \mu_0$$

Ipotesi di test sulla media con alternativa composta unidirezionale

Si calcola la statistica test

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



Se il valore della statistica test risulta maggiore del valore z_a si rifiuta l'ipotesi nulla

$$Z \geq z_\alpha$$

20

Test d'ipotesi

Test sulla durata dei ferri da stiro espressa in ore :

H_0:μ=1850

H_1:μ>1850

Si assume che la durata dei ferri da stiro sia una variabile casuale $X \sim N(\mu, \sigma^2)$ con varianza nota pari a 6400 .

Si estrae un campione di 100 ferri e si osserva una media pari a 1880. Si effettui il test con un α=5%

Si calcola la statistica test $$Z = \frac{1880-1850}{\sqrt{6400/100}} = 3{,}75$$

Per α=0,05 si ottiene che :$z_{0,05}$=1,645

e quindi si rifiuterà l'ipotesi nulla poiché $Z \geq z_\alpha$ **3,75>1,645**

21

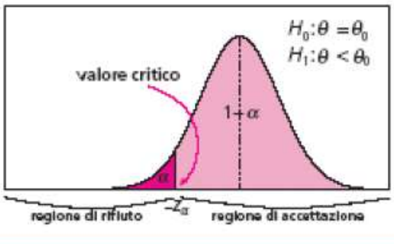
Test d'ipotesi

$$H_0: \ \mu = \mu_0$$
$$H_1: \ \mu < \mu_0$$

Ipotesi di test sulla media con alternativa composta unidirezionale

Si calcola la statistica test

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



Se il valore della statistica test risulta minore del valore $-z_a$ si rifiuta l'ipotesi nulla

$$Z \leq -z_\alpha$$

22

Test d'ipotesi

Test sul peso medio di un collettivo:

H_0:μ=80

H_1:μ<80

Si assume che il peso degli individui sia una variabile casuale

$X \sim N(\mu, \sigma^2)$ con varianza nota pari a 16 .

Si estrae un campione di 64 persone e si osserva una media pari a 79 kg. Si effettui il test con un α=5%

Si calcola la statistica test $$Z = \frac{79-80}{\sqrt{16/64}} = -2$$

Per α=0,05 si ottiene che:$-z_{0,05}$=-1,645

e quindi si rifiuterà l'ipotesi nulla poiché $Z \le -z_\alpha$ **-2<-1,645**

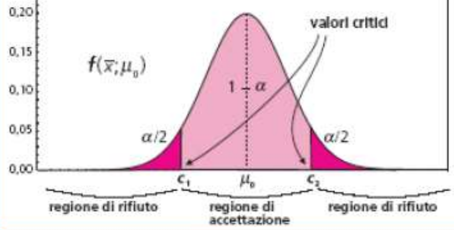
23

Test d'ipotesi

$$H_0: \ \mu = \mu_0$$
$$H_1: \ \mu \neq \mu_0$$

Ipotesi di test sulla media con alternativa composta unidirezionale

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



Se il valore della statistica test risulta al di fuori dell'intervallo compreso tra $-z_{a/2}$ e $z_{a/2}$ si rifiuta l'ipotesi nulla

24

Test d'ipotesi

Test sul voto medio agli esami

H_0:μ=26,4

H_1:μ≠26,4

Si assume che il voto medio agli esami sia una variabile casuale $X \sim N(\mu, \sigma^2)$ con varianza nota pari a 10,24 .

Si estrae un campione di 9 studenti e si osserva una media pari a 27,2. Si effettui il test con un α=5%

Si calcola la statistica test $Z = \frac{27{,}2 - 26{,}4}{\sqrt{10{,}24/9}} = 0{,}75$

Per α=0,05 si ottiene che: $z_{a/2}=z_{0,025}=1{,}96$

e quindi si accetterà l'ipotesi nulla poiché il valore della statistica test è compreso nell'intervallo (-1,96 ; 1,96)

25

Test d'ipotesi

Test per media – Popolazione Normale – Varianza ignota

Statistica test e sua distribuzione sotto l'ipotesi nulla:

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

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**
$H_1: \ \mu > \mu_0$	$T \geq t_\alpha$
$H_1: \ \mu < \mu_0$	$T \leq -t_\alpha$
$H_1: \ \mu \neq \mu_0$	$\lvert T \rvert \geq t_{\alpha/2}$

26

Test d'ipotesi

Test sulla statura media in un collettivo:

H_0:μ=175

H_1:μ>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

$$\bar{x} = 181{,}5 \qquad S^2 = 95{,}5067 \qquad t = \frac{181{,}5 - 175}{\sqrt{95{,}5067/10}} = 2{,}103$$

Ponendo α=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>1,8331**

27

Test d'ipotesi

Test sulla statura media in un collettivo:

H_0:μ=180

H_1:μ≠180

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

$$\bar{x} = 181{,}5 \qquad S^2 = 95{,}5067 \qquad t = \frac{181{,}5 - 180}{\sqrt{95{,}5067/10}} = 0{,}485$$

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

$t_{0,025}$=2,262

e quindi si accetterà l'ipotesi nulla poiché **-2,262≤ t≤2,262**

28

Test d'ipotesi

Test per una proporzione – Popolazione Bernoulliana

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

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

con π_0 che indica il valore della proporzione ipotizzato in H_0

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

Test d'ipotesi

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

$H_0{:}\pi=0{,}084$

$H_1{:}\pi<0{,}084$

Si estrae un campione di 1000 occupati e si osserva che 53 sono occupati nel settore agricolo, pertanto:

$$\bar{x} = 0{,}053$$

Il valore della statistica test è:

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

Ponendo α=0,01 si ottiene dalla normale standardizzata:

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

e quindi si rifiuterà l'ipotesi nulla poiché $z< -z_{0,01}$ $-3{,}534< -2{,}326$

Test d'ipotesi

Test per la varianza – Popolazione Normale – media ignota

Statistica test e sua distribuzione sotto l'ipotesi nulla:

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

con σ^2_0 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**
$H_1:\ \sigma^2 > \sigma_0^2$	$(n-1)S^2/\sigma_0^2 \geq \chi^2_{\alpha}$
$H_1:\ \sigma^2 < \sigma_0^2$	$(n-1)S^2/\sigma_0^2 \leq \chi^2_{1-\alpha}$
$H_1:\ \sigma^2 \neq \sigma_0^2$	$(n-1)S^2/\sigma_0^2 \leq \chi^2_{1-\alpha/2}$ $(n-1)S^2/\sigma_0^2 \geq \chi^2_{\alpha/2}$

31

Test d'ipotesi

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

$H_0: \sigma^2 = 22500$

$H_1: \sigma^2 \neq 22500$

Si estrae un campione di 61 famiglie.

Dal campione, si osservano:

$$\bar{x} = 2010 \qquad S^2 = 20000$$

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

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

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

quindi il valore osservato non cade nella regione di rifiuto e si accetterà l'ipotesi nulla, ossia che la varianza è pari a 22500.

32

Test d'ipotesi

TABELLA 14.4.1 Statistiche test: schema riassuntivo.

Popolazione	Parametri noti	Parametri ignoti	Dimensione del campione	Statistica test	Distribuzione campionaria
			Verifica d'ipotesi per la **media**		
Normale	σ^2	μ	qualsiasi	$\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$	$N(0, 1)$
Normale		μ, σ^2	piccolo	$\frac{\bar{X}-\mu_0}{S/\sqrt{n}}$	*t-Student* con $n-1$ *g.l.*
Qualsiasi		μ, σ^2	grande	$\frac{\bar{X}-\mu_0}{S/\sqrt{n}}$	$N(0, 1)$
			Verifica d'ipotesi per una **proporzione**		
Bernoulli		π	grande	$\frac{\bar{X}-\pi_0}{\sqrt{\pi_0(1-\pi_0)/n}}$	$N(0, 1)$
			Verifica d'ipotesi per la **varianza**		
Normale		μ, σ^2	qualsiasi	$\frac{(n-1)S^2}{\sigma_0^2}$	χ^2 con $n-1$ *g.l.*

33