TEST PARAMETRICI PER UN CAMPIONE

Due tipologie:

- Test sulla media da una popolazione
    - Con varianza nota
    - Con varianza ignota
- Test sulla proporzione da una popolazione

TEST PER LA MEDIA

Utilizzata per verificare l'ipotesi su particolari valori della media di una popolazione.

Si possono fare test unilaterali sinistri/destri o bilaterali. Consideriamo il caso generale di ipotesi composte

➢ sistema di ipotesi

$$\begin{cases} H_0 : \mu=\mu_0 \\ H_1 : \mu<\mu_0 \end{cases} \quad \begin{cases} H_0 : \mu=\mu_0 \\ H_1 : \mu>\mu_0 \end{cases} \quad \begin{cases} H_0 : \mu=\mu_0 \\ H_1 : \mu\neq\mu_0 \end{cases}$$

unilaterale sinistro — unilaterale destro — bilaterale

ASSUMIAMO che la popolazione di interesse segua una distribuzione di tipo **NORMALE**, cioè che il campione provenga da una **V.C. $X \sim N(\mu,\sigma^2)$**

Dobbiamo distinguere due casi (la differenza è solo nella distribuzione della statistica test sotto H0):

1. Varianza nota
2. Varianza incognita

Caso 1: Normale con varianza nota

Supponiamo di estrarre un campione casuale semplice da una variabile casuale normale con media ignota e **varianza nota** pari a σ^2_0

SISTEMA DI IPOTESI

$$\begin{cases} H_0 : \mu=\mu_0 \\ H_1 : \mu\neq\mu_0 \end{cases} \quad (H_1: \mu \geq \mu_0 \text{ ovvero } H_1: \mu \leq \mu_0)$$

N.B. Indichiamo con μ1 un generico possibile valore del parametro sotto l'ipotesi alternativa.

FUNZIONE TEST:

$$t(\underline{x})=\frac{\overline{X}-\mu_0}{\sigma_0/\sqrt{n}} \sim \begin{cases} N(0,1) \text{ sotto } H_0 \\ N\left(\frac{(\mu_1-\mu_0)}{\sigma_0/\sqrt{n}},1\right) \text{ sotto } H_1 \text{ se } \mu=\mu_1\neq\mu_0 \end{cases}$$

VALORE CRITICO

Per calcolare la regione di rifiuto R di H_0 occorre che sia garantita la probabilità di errore di primo tipo. Se H_0 è semplice allora deve essere $\alpha=P(t(\underline{X})\in R|H_0)$ e di conseguenza il valore critico si calcola individuando il percentile corrispondente al livello di significatività (α) relativo alla distribuzione di probabilità che la funzione test assume sotto H_0.

N.B. Oltre ad effettuare un test d'ipotesi con una prefissata significatività α, possiamo sempre calcolare il p-value. Quest'ultima è l'opzione generalmente preferita negli output dei pacchetti statistici.

REGIONI DI RIFIUTO di H_0		
	1) Test BILATERALE	$t(\underline{X}) = \left\lvert \frac{\overline{X}-\mu_0}{\sigma_0/\sqrt{n}} \right\rvert \geq z_{\alpha/2}$
	2) UNILATERALE DX	$t(\underline{X}) = \frac{\overline{X}-\mu_0}{\sigma_0/\sqrt{n}} \geq z_{\alpha}$
	3) UNILATERALE SN	$t(\underline{X}) = \frac{\overline{X}-\mu_0}{\sigma_0/\sqrt{n}} \leq -z_{\alpha}$

Osservazioni:

Se H0è composta, allora R si determina così che la probabilità di errore di I tipo sia α in corrispondenza del valore del parametro punto di frontiera tra le due ipotesi. Nel caso dei test bilaterali μ=μ0.

2) e 3) si applicano non solo al caso in cui H0: μ =μ0, ma anche al caso in cui H0sia, rispettivamente, μ≤μ0 e μ ≥ μ0.

ESEMPIO 1

Un campione casuale semplice di 100 unità estratto da una variabile **X** distribuita in modo normale, ha fornito i valori sintetizzati nella seguente distribuzione di frequenza:

x_i	1	3	6	7	12
frequenze	20	10	40	10	20

Sapendo che la varianza di **X** è pari a 5, verificare, ad un livello di significatività pari a 0.05, l'ipotesi che la media incognita di **X** sia pari a 5.2, contro l'ipotesi alternativa bilaterale.

- **Sistema di ipotesi**

$$\begin{cases} H_0 : \mu = 5.2 \\ H_1 : \mu \neq 5.2 \end{cases}$$

- **Regione di rifiuto**

$$t(\underline{X}) = \left\lvert \frac{\overline{X}-\mu_0}{\sigma_0/\sqrt{n}} \right\rvert \geq z_{\alpha/2} \Longleftrightarrow \overline{X} \leq \mu_0 - z_{\alpha/2}\sigma_0/\sqrt{n} \text{ oppure } \overline{X} \geq \mu_0 + z_{\alpha/2}\sigma_0/\sqrt{n}$$

Espressa in funzione della statistica test

Espressa in funzione della media campionaria

8

- **Valore critico**

Essendo α=0.05 dalle tavole della normale standard si ricava che $z_{\alpha/2}$=1.96.

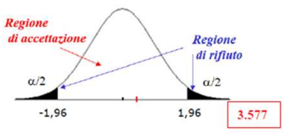
Regione di rifiuto individuata rispetto alla funzione test è pertanto:

$(-\infty,-1.96)\cup(1.96, \infty)$

- **Valore funzione test calcolato**

dai dati del campione: n=100; media campionaria=6;

$$t(\underline{X}) = \frac{\overline{X}-\mu_0}{\sigma_0/\sqrt{n}} = \frac{6-5.2}{\sqrt{5/100}} = 3.577$$



- **Decisione**

Il valore calcolato risulta maggiore del valore critico 3.577 > 1.96 quindi esso appartiene alla regione di rifiuto, per cui la decisione è RIFIUTO H_0.

9

ESEMPIO 2

Dato un campione casuale semplice di n=25 unità estratto da una variabile **X** distribuita in modo normale con varianza nota e pari a 1, con il seguente valore per la media campionaria -7.997, verificare, ad un livello di significatività pari a 0.05, l'ipotesi che la media incognita di **X** sia pari a 4, contro l'ipotesi alternativa bilaterale.

- **Sistema di ipotesi**

$$\begin{cases} H_0 : \mu=4 \\ H_1 : \mu\neq 4 \end{cases}$$

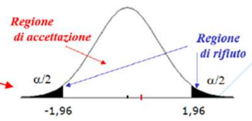
- **Regione di rifiuto e valore critico**

Essendo α=0.05 dalle tavole della normale standard si ricava che $\mathbf{z_{\alpha/2}=1.96}$.

Regione di rifiuto individuata rispetto alla funzione test è pertanto: $(-\infty,-1.96)\cup(1.96,\infty)$

- **Valore funzione test**

$$t(x)=\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}=\frac{-7.997-4}{1/\sqrt{25}}=-59.98$$



- **Decisione**

Il valore calcolato risulta decisamente maggiore del valore critico $-59.98 < -1.96$ quindi la decisione è RIFIUTO H_0.

10

ESEMPIO 3

$\underline{x} = (x_1,x_2,\ldots,x_n)$ da $N(\mu,\sigma^2)$ con σ^2 noto

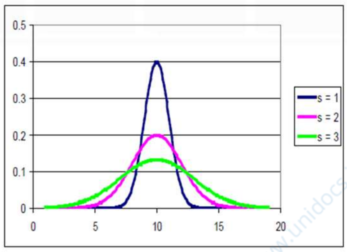
H_0: $\mu\leq 10$ $\quad H_1$: $\mu>10$ $\quad$ con α=0.05

Il valore critico è espresso rispetto alla media campionaria

$$\overline{X}\geq \mu_0 + z_\alpha \sigma/\sqrt{n}$$

$$\mathbf{R=\underline{x}\ t.c.\ t(\underline{x})=\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}=\frac{\overline{X}-10}{\sigma/\sqrt{n}}>z_\alpha}$$

	v. critico		
	σ^2=1	σ^2=4	σ^2=9
n			
1	11.64485	13.28971	14.93456
5	10.7356	11.4712	12.2068
10	10.52015	11.0403	11.56044
20	10.3678	10.7356	11.1034
50	10.23262	10.46523	10.69785
100	10.16449	10.32897	10.49346
1000	10.05201	10.10403	10.15604



→A parità di n il valore critico della soglia aumenta al variare di σ^2

→A parità di σ^2 il valore critico diminuisce al crescere di n

Caso 2: Normale con varianza ignota

Supponiamo di estrarre un campione casuale semplice da una variabile casuale normale con media ignota e **varianza ignota** pari a σ^2.

SISTEMA DI IPOTESI

$$\begin{cases} H_0 : \mu=\mu_0 \\ H_1 : \mu\neq\mu_0 \end{cases} \qquad (H_1: \mu \geq \mu_0 \text{ ovvero } H_1: \mu \leq \mu_0)$$

FUNZIONE TEST

$$t(\underline{x})=\frac{\overline{X}-\mu_0}{s/\sqrt{n}} \sim$$

$\boxed{\mathbf{t_{n-1}\ sotto\ H_0}}$

$\mathbf{t'_{n-1}}$ **sotto H_1** se $\mu=\mu_1\neq\mu_0$

t' = t di Student non centrale

VALORE CRITICO

Si calcola individuando il percentile corrispondente al livello di significatività (α) relativo alla distribuzione di probabilità che la funzione test assume sotto H_0 (in questo caso la v.c. t_{n-1}).

Ricordiamo che $\alpha = P(t(\underline{x}) \in R | H_0)$

REGIONI DI RIFIUTO di H_0

1) **BILATERALE** $\left| \frac{\bar{X}-\mu_0}{s/\sqrt{n}} \right| \geq t_{\alpha/2,n-1}$

2) **UNILATERALE DX** $\frac{\bar{X}-\mu_0}{s/\sqrt{n}} \geq t_{\alpha,n-1}$

3) **UNILATERALE SN** $\frac{\bar{X}-\mu_0}{s/\sqrt{n}} \leq -t_{\alpha,n-1}$

La potenza del test dipende da $\delta = \mu - \mu_0$.

ESEMPIO

Dato un campione casuale semplice di n=30 unità estratto da una variabile **X** distribuita in modo normale con varianza ignota, con media campionaria pari a 7.967 e varianza campionaria corretta pari a 0.963, verificare, ad un livello di significatività pari a 0.05, l'ipotesi che la media incognita di **X** sia pari a 8, contro l'ipotesi alternativa bilaterale.

- **Sistema di ipotesi** $\begin{cases} H_0 : \mu=8 \\ H_1 : \mu\neq 8 \end{cases}$

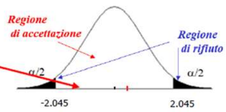
- **Regione di rifiuto e valore critico**

Essendo α=0.05 dalle tavole della t con gdl=29 si ricava che $\mathbf{t_{(29)\alpha/2}=2.045}$.

Regione di rifiuto individuata rispetto alla funzione test è pertanto: $(-\infty,-2.045)\cup(2.045,\infty)$

- **Valore funzione test**

$$t(\underline{X}) = \frac{\overline{X}-\mu_0}{s/\sqrt{n}} = \frac{7.967-8}{\sqrt{0.963/30}} = -0.1656$$



- **Decisione**

Il valore calcolato risulta compreso tra i due valori critici -2.045< -0.1656 < 2.045 quindi la decisione è ACCETTO H_0.

15

Se non c'è scritto "varianza campionaria CORRETTA", devo utilizzare S^{2*} e devo correggerla io

Osservazioni:

1. Se n è «grande» (al solito> 30/50) è assumibile la normalità della funzione test sotto H0 e quindi il valore critico si troverà sulle tavole della Normale
2. Se la distribuzione di probabilità non è nota e la numerosità campionaria non consente l'applicazione del tlc (n<<50) è possibile ricorrere ai test non parametrici (non li facciamo)

Il test con R:

In R la funzione per eseguire il test sulla media di un campione è

```
t.test(x,mu=,alternative="")
```

dove:
- x = dati del campione
- mu = valore della media sotto H_0
- alternative = tipo di test (unilaterale sx o dx o bilaterale)

➢ La funzione esegue sempre quello per media con varianza INCOGNITA (dato che nelle applicazioni la situazione di varianza nota è improbabile)

➢ L'output è costituito da:
  - valore della statistica test
  - numero di gradi di libertà
  - p-value
  - estremi dell'IC al 95%
  - valore della media campionaria

Esempio 1

Dati simulati da una Normale con media=8 e varianza=1 - numerosità n=40 (piccolo campione)

```
#### 1 campione
#dati simulati N(mu, sigma)
data<-rnorm(40,8,1)
```

- test unilaterale destro

```
>t.test(data,mu=5,alternative="greater")

        One Sample t-test

data:  data
t = 16.238, df = 39, p-value < 2.2e-16
alternative hypothesis: true mean is greater than 5
95 percent confidence interval:
 7.432385      Inf
sample estimates:
mean of x
  7.71399
```

Rifiuto H_0 per ogni valore di α

Deldossi-Paroli - Statistica Applicata 18

- test unilaterale sinistro

```
> t.test(data,mu=5,alternative="less")

        One Sample t-test

data:  data
t = 16.238, df = 39, p-value = 1
alternative hypothesis: true mean is less than 5
95 percent confidence interval:
     -Inf 7.995595
sample estimates:
mean of x
  7.71399
```

Accetto H_0 per ogni valore di α

- test bilaterale

```
> t.test(data,mu=5,alternative="two.sided")

        One Sample t-test

data:  data
t = 16.238, df = 39, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 5
95 percent confidence interval:
 7.375924 8.052057
sample estimates:
mean of x
  7.71399
```

Rifiuto H_0 per ogni valore di α

Deldossi-Paroli - Statistica Applicata 19

Esempio 2

Dati simulati da una Normale con media=8 e varianza=1 - numerosità n=100 (grande campione)

```
#### 1 campione
#dati simulati N(mu, sigma)
data<-rnorm(100,8,1)

>t.test(data,mu=5,alternative="greater")

        One Sample t-test

data:  data
t = 34.693, df = 99, p-value < 2.2e-16
alternative hypothesis: true mean is greater than 5
95 percent confidence interval:
 8.018421      Inf
sample estimates:
mean of x
 8.170144
```

Rifiuto H_0 per ogni valore di α

Deldossi-Paroli - Statistica Applicata 20

Note:

- df= gradi li bertà
- rnorm (40,8,1)= ci permette di simulare dati casuali da una normale (numero dati, media, varianza)
- Più il p-value è piccolo più accettiamo H1

TEST PER LA PROPORZIONE

Si utilizzano per verificare l'ipotesi su particolari valori della proporzione π di una popolazione bernoulliana. Si possono fare test unilaterali sinistri/destri o bilaterali

➤ **sistema di ipotesi**

$\begin{cases} H_0 : \pi = \pi_0 \\ H_1 : \pi < \pi_0 \end{cases}$	$\begin{cases} H_0 : \pi = \pi_0 \\ H_1 : \pi > \pi_0 \end{cases}$	$\begin{cases} H_0 : \pi = \pi_0 \\ H_1 : \pi \neq \pi_0 \end{cases}$
unilaterale sinistro	unilaterale destro	bilaterale

ASSUMIAMO che la popolazione di interesse segua una distribuzione di tipo **BERNOULLIANO**, cioè che il campione provenga da una **V.C. X~Bin(1,π).**

Ricordiamo che:

- La proporzione π coincide con il calcolo della media nel caso di caratteri di tipo dicotomico (codificabili quindi come 0 o1).
- La varianza risulta essere funzione della proporzione π

  Var(X) =π(1-π)

  e non un valore indipendente da π.

  Dunque se non è noto π non potrà essere nota neppure la varianza. Non avremo due casi come per i test sulla media di una normale.

FUNZIONE TEST APPROSSIMATA (n>50):

$$t(\underline{x}) = \frac{P-\pi_0}{\sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}}} \sim N(0,1) \text{ sotto } H_0$$

P = proporzione campionaria

REGIONE DI RIFIUTO

R si individua in modo analogo a quanto fatto nel caso normale. I valori critici si ottengono individuando il percentile corrispondente al livello di significatività (α) relativo alla distribuzione di probabilità che la funzione test assume sotto H_0 (in questo caso la approx N(0,1)).

Ricordiamo che $\alpha = P(t(\underline{x}) \in R | H_0)$

REGIONE DI RIFIUTO (per n>50):

1) BILATERALE $\left| \frac{P-\pi_0}{\sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}}} \right| \geq z_{\alpha/2}$

2) UNILATERALE DX $\frac{P-\pi_0}{\sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}}} \geq z_\alpha$

3) UNILATERALE SN $\frac{P-\pi_0}{\sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}}} \leq -z_\alpha$

ESEMPIO

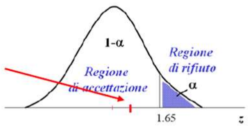
1) Si vuole stabilire se la percentuale di voto a favore del SI in un referendum sia inferiore al 30%.

$$\begin{cases} H_0: \pi \leq 0.3 \\ H_1: \pi > 0.3 \end{cases}$$ test unilaterale dx

Sia n=100 e p=0.32.

Si rifiuta H_0 con significatività 5%?

$$t(x) = \frac{0.32-0.3}{\sqrt{0.3 \cdot 0.7}/\sqrt{100}} = .4364$$



Poiché $0.4364 < z_{0.05} = 1.65$, si conclude che **non** si rifiuta H_0.

2) Qual è il valore minimo di P che consente di rifiutare al 5% l'ipotesi nulla?

Affinché t(x) cada nella regione di rifiuto deve essere > 1.64, cioè:

$$t(x) = \frac{P-0.3}{\sqrt{0.3 \cdot 0.7}/\sqrt{100}} > 1.64$$ risolvo rispetto a P

$$P > 0.3 + 1.64 \cdot \sqrt{0.3 \cdot 0.7}/10 = 0.3752$$

3) Se osserviamo p=0.32 per quale ampiezza campionaria n questo risultato consente di rifiutare l'ipotesi nulla al 5%?

$$t(x) = \frac{0.32-0.3}{\sqrt{0.3 \cdot 0.7}/\sqrt{n}} > 1.64$$ risolvo rispetto a n

$$\sqrt{n} > 1.64 \cdot \sqrt{0.3 \cdot 0.7}/0.02$$

$$n > 1.64^2 \cdot 0.3 \cdot 0.7/(0.02)^2 \approx 1412$$

Il test con R

In R la funzione per eseguire il test sulla media di un campione è

```
>prop.test(x,n,p=,alternative="")
```

dove:

- x = numero di successi nel campione
- n = numerosità del campione
- p = valore della proporzione sotto H_0
- alternative = tipo di test (unilaterale sx o dx o bilaterale)

➢ L'output è costituito da:
  - valore della statistica test
  - p-value (calcolato secondo una approssimazione particolare)
  - estremi dell'IC al 95%
  - valore della proporzione campionaria

ESEMPIO PRECEDENTE

```
> prop.test(32,100,p=0.3,alternative="greater")

        1-sample proportions test with continuity correction
data:  32 out of 100, null probability 0.3
X-squared = 0.10714, df = 1, p-value = 0.3717
alternative hypothesis: true p is greater than 0.3
95 percent confidence interval:  0.2443231 1.0000000
sample estimates:    p 0.32
```