6.3 Test di verifica di ipotesi

6.3.1 Il test di verifica di ipotesi: definizione

Un parametro può essere stimato attraverso i metodi per fare stime puntuali visti nei paragrafi precedenti, ma più frequentemente l'interesse è quello di decidere, fra due ipotesi o asserzioni contraddittorie riguardo a un parametro, quale sia quella corretta.

Questi metodi di inferenza statistica vengono detti *test di verifica di ipotesi*. Alcuni esempi.

- Asserzione: 'La media del diametro di alcuni tubi sembra essere 0.75'. È vero oppure no che la media esatta dei tubi è 0.75? Si deve dare una risposta a questa domanda attraverso informazioni basate come sempre su campioni della popolazione che si sta analizzando.
- Asserzione: 'La distribuzione di probabilità della distanza di arresto dei veicoli è di tipo normale'. È vera oppure no?

In questo testo, ci occuperemo prevalentemente di casi simili al primo, cioè di ipotesi sui parametri.

Una *ipotesi statistica*, o semplicemente *ipotesi*, è una asserzione sul valore di un singolo parametro, di più parametri o di una intera distribuzione di probabilità.

Una ipotesi è *semplice* quando è definita da un vincolo di uguaglianza, è *composita* quando è definita da un vincolo di disuguaglianza.

Alcuni esempi.

- L'altezza media degli italiani è $\mu = 1,72$. Questa è una ipotesi semplice sulla media di una popolazione.
- La durata di una certa lampadina è di più di 2000 ore. Questa è un'ipotesi composita ($l \geq 2000$ se l indica la durata in ore).

In ogni test di ipotesi, ci sono due ipotesi contraddittorie fra cui scegliere. Di più, il test di ipotesi è formulato in modo che una delle due ipotesi sia inizialmente 'favorita', perché ritenuta più plausibile. Questa ipotesi inizialmente favorita *può* essere rigettata in favore dell'altra solo se ci sono *evidenti* indizi che la contraddicono.

Il test di ipotesi si può quindi formulare nel modo seguente.

Test di ipotesi. L'*ipotesi nulla* H_0 è l'affermazione inizialmente vera.

L'*ipotesi alternativa* H_a è un'affermazione che contraddice H_0. L'ipotesi nulla *può essere rigettata* in favore di H_a solo se l'evidenza derivante dall'analisi di campioni suggerisce che H_0 è falsa.

Se il campione non *contraddice fortemente l'ipotesi nulla* si continua a sostenere l'ipotesi nulla.

Le due possibili conclusioni di un test di ipotesi sono quindi:

- **rigettare** H_0

oppure

- **non rigettare** H_0.

L'ipotesi nulla viene tipicamente formulata come un'uguaglianza del tipo:

- H_0: $\theta = \hat{\theta}$.

L'ipotesi alternativa è invece composita ed è di solito formulata con una delle tre alternative:

- $\theta > \theta_0$
- $\theta < \theta_0$
- $\theta \neq \theta_0$

Nei primi due casi l'ipotesi si dice *unilaterale*, nell'ultimo caso *bilaterale*.

Nell'esempio precedente riguardante il diametro dei tubi il test di ipotesi può essere formulato così:

- H_0 è l'ipotesi: 'la media $\mu = 1.5$ ';
- H_a è l'ipotesi: 'la media $\mu \neq 1.5$ '

6.3.2 Procedura per un test di ipotesi

La *procedura di test* è una regola, basata su dati campionati, usata per decidere se rigettare o no H_0.

Ci sono due fasi:

1. Un *test statistico* per prendere la decisione;
2. una *regione di rifiuto* costituita da quei valori del risultato del test di ipotesi per cui H_0 viene effettivamente rifiutata in favore di H_a. La regione costituita dai valori del risultato del test di ipotesi per cui H_0 **non** viene rifiutata (regione complementare a quella di rifiuto) viene detta *regione di accettazione*.

Quindi:

L'ipotesi nulla viene rifiutata se e solo se il risultato del test statistico ricade nella regione di rifiuto.

Esempio 6.17 *Un produttore di bevande alla frutta dice che la media di zuccheri presenti nei succhi è di 1,5 g/l. Supponiamo che la distribuzione dello zucchero nella bevanda sia normale con deviazione standard $\sigma = 0.25$ e di avere un campione di 32 elementi. Eseguire un test di ipotesi sull'affermazione del produttore.*

Soluzione.

Allora:

- H_0 è l'ipotesi: 'la media $\mu = 1.5$' ;

- H_a è l'ipotesi: ' la media $\mu > 1.5$'.

Si devono eseguire i seguenti passi:

1. Prendere un campione di n bevande (SRS(n)).
2. Considerare come test di ipotesi: *Calcolo della media campionaria* $\bar{X}$.
3. Scelto un valore di α, calcolare un intervallo di confidenza I_α $100(1-\alpha)\%$ per $\bar{X}$.
4. Se il valore 1.5 **non** sta nell'intervallo di confidenza calcolato, allora *rigetto* H_0.

Allora, calcolata la media campionaria $\bar{X}$, calcoliamo l'intervallo di confidenza:

$$I_\alpha = \left(\bar{X} - \zeta_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + \zeta_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

Se $1.5 \notin I_\alpha$ allora H_0 viene rigettata.

Esempio 6.18 *In questo script* R *, si fa una simulazione analoga all'esempio precedente, tranne per il fatto che il campione viene simulato come SRS(n) da una distribuzione normale con media $\mu = 1.5$ e deviazione standard $\sigma = 0.25$.*

Soluzione.

In questo caso sappiamo quindi che l'ipotesi nulla è vera. Il risultato del test di verifica è infatti: '**non** rigettare H_0'.

```
> sigma <- 0.25
> n <- 64
> mu <- 1.5
> alpha <- 0.05
> mu0 <- 1.5
> x <- rnorm(n, mu, sigma)
> m <- mean(x)
> zalfa <- qnorm(1 - alpha / 2, 0, 1)
> c1 <- m - zalfa * sigma / sqrt(n)
> c2 <- m + zalfa * sigma / sqrt(n)
> c1

## [1] 1.411451

> c2

## [1] 1.533949
```

```
> if (mu0 < c1 | mu0 > c2)
+ {
+   message("ipotesi nulla rigettata")
+ } else
+ {
+   message("ipotesi nulla non rigettata")
+ }

## ipotesi nulla non rigettata
```

Nello script successivo (di cui non si riporta il codice, essendo praticamente uguale al precedente), il valore della variabile `mu = 1.55`, sempre con $\alpha = 0.05$.

Il risultato dell'esecuzione è:

```
> c1

## [1] 1.368076

> c2

## [1] 1.490574

> if (mu0 < c1 | mu0 > c2)
+ {
+   message("ipotesi nulla rigettata")
+ } else
+ {
+   message("ipotesi nulla non rigettata")
+ }

## ipotesi nulla rigettata
```

Se adesso ripetiamo la simulazione con $\alpha = 0.01$ abbiamo questo risultato:

```
> c1

## [1] 1.421221

> c2

## [1] 1.58221

> if (mu0 < c1 | mu0 > c2)
+ {
+   message("ipotesi nulla rigettata")
```

```
+ } else
+ {
+   message("ipotesi nulla non rigettata")
+ }

## ipotesi nulla non rigettata
```

Quindi, cambiando il valore di α si ottengono risultati diversi.

Esempio 6.19 *Eseguire il seguente est di verifica di ipotesi:*

- H_0: *'la media della profondità del lago Huron è di 580 piedi'.*
- H_a: *'la media della profondità del lago Huron è inferiore a 580 piedi'*

Soluzione.

Si considerano i dati nel file `LakeHuron` contenuto nel pacchetto `datasets` di R , che rappresentano le altezze del lago Huron dal 1875 al 1972. Possiamo supporre che la distribuzione dei valori sia normale, ma non se ne conosce la deviazione standard. L'intervallo di confidenza viene quindi calcolato, utilizzando le formule viste nella sezione precedente, utlizzando la deviazione standard campionaria S invece della deviazione standard esatta σ.

```
> n <- 98
> x <- LakeHuron
> mu_0 <- 580
> alfa <- 0.001
> barx <- mean(x)
> zalfa <- qnorm(1 - alfa / 2, mean = 0, sd = 1)
> S <- sd(x)
> c1 <- barx - zalfa * S / sqrt(n)
> c2 <- barx + zalfa * S / sqrt(n)
> if (mu_0 < c1 | mu_0 > barx)
+ {
+   message("rigettata ipotesi nulla")
+ }

## rigettata ipotesi nulla
```

6.3.3 Errori nel test di verifica di ipotesi

La regione di rifiuto viene di solito scelta anche sulla base degli errori che possono essere fatti nel test di ipotesi. Gli errori nascono dal fatto che non è possibile esaminare l'intera popolazione, ma solo dei campioni che variano.

Ci sono due tipi di errore possibili.

- Si dice *errore di primo tipo* (E1) il rifiuto di H_0 quando è vera.
- Si dice *errore di secondo tipo* (E2) il *non rifiuto* di H_0 quando è falsa.

Nell'esempio precedente, un errore del secondo tipo consiste nel non rifiutare l'ipotesi H_0 che la media sia $\mu = 1.5$ quando invece non lo è. Per esempio, la media campionaria osservata è di $\mu = 1.55$ e quindi viene accettata l'ipotesi H_0 anche se non è vera.

Gli errori del primo tipo sono considerati 'peggiori' degli errori del secondo tipo. Sarebbe auspicabile rendere sia E1 che E2 più piccoli possibile in modo da ridurre la probabilità di commettere entrambi gli errori. La seguente proposizione dice che non è possibile scegliere contemporaneamente E1 ed E2 piccoli. Vale infatti il seguente risultato.

Supponiamo che in un esperimento siano fissati sia la dimensione n del campione che il test statistico con cui verificare l'ipotesi. Allora decrescere la regione di rifiuto per ottenere valori più piccoli per E1 comporta avere valori più grandi di E2.

Si deve perciò scegliere un compromesso fra E1 ed E2 e l'approccio più comune è quello di avere un valore più grande per E1. Questo rende E2 più piccolo.

Definiamo la probabilità α di commettere un errore del primo tipo e la denominiamo *livello di significatività del test*. Livelli tradizionali di significatività sono 0.1, 0.05, 0.01.

La corrispondente procedura viene detta *un test a livello* α.

Le regioni di rifiuto vengono calcolate a partire livello di significatività del test.

Utilizzando le stime di intervalli viste nel paragrafo precedente, a partire dalle probabilità α richiesta, possiamo costruire intervalli di confidenza rispetto alle stime ottenute con il test statistico che rappresentano la regione di accettazione, e di conseguenza si determina la regione di rifiuto.

Riprendendo una delle simulazioni precedenti, abbiamo visto che per $\alpha = 0.05$ l'ipotesi nulla NON era rigettata, mentre per $\alpha = 0.01$ l'ipotesi nulla era rigettata. Quindi al variare del livello di significatività α si hanno diversi risultati del test. Questo ovviamente non è auspicabile. Non sappiamo, per un fissato test, quale sia il valore di α che 'separa' i risultati positivi da quelli negativi del test. Questo livello ha una rilevanza particolare e viene definito nel modo seguente.

Il *valore p* (*p-value*) di una verifica di test di ipotesi è il *minimo* livello di significatività del test per cui l'ipotesi nulla viene rigettata. Ovvero: fissato il test, l'ipotesi nulla viene rigettata per ogni valore $\alpha > p$ e NON viene rigettata per ogni valore $\alpha \leq p$.

Per esempio, per calcolare il valore p relativo all'esempio precedente della media dell'altezza del lago Huron, dobbiamo calcolare il valore di α per cui il quantile $z_{\alpha/2}$ produce il valore del nostro test statistico , $\bar{X} = 579.0041$.

6.3.4 Test sulla media di una popolazione

Soffermandoci in particolare sulla *media* come parametro su cui fare il test di ipotesi, a partire dall'ipotesi nulla H_0: $\mu = \mu_0$, negli esempi precedenti abbiamo considerato come statistica per la stima la media campionaria $\bar{X}$.

Caso di distribuzione normale con deviazione standard nota.

Sappiamo che se la popolazione esaminata ha una distribuzione normale con media μ e deviazione standard σ, allora $\bar{X} \approx$ `norm(`μ`, `$\sigma/\sqrt{n}$.

Consideriamo adesso un'altra statistica:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

Sappiamo che Z ha una distribuzione normale standard.

Per esempio, se $\mu_0 = 100$, $\sigma/\sqrt{n} = 10/\sqrt{(}25) = 2.0$ e la media campionaria calcolata $\bar{X} = 102$, allora:

$$Z = (102 - 100)/2 = 1.$$

La statistica Z è una misura della distanza fra $\bar{X}$, estimatore del parametro μ del test e μ_0, valore atteso quando H_0 è vera. Se questa distanza è troppo grande nella direzione dell'ipotesi alternativa H_a, allora l'ipotesi nulla viene rigettata.

Questo test, che si può fare solo quando è nota la deviazione standard σ della distribuzione, viene detto *z-test*.

Caso di distribuzione normale con devizione standard non nota di grande dimensione.

In questo caso, si può applicare il teorema del Limite Centrale per cui la variabile aleatoria

$$Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

in cui S è la deviazione standard campionaria ha *approssimativamente* una distribuzione normale standard. Quindi possiamo usare lo z-test anche in questo caso, sostituendo S a σ.

Il valore di n per cui di solito si considera valido il teroema del Limite Centrale e quindi questa approssimazione è $n > 40$.

Caso di distribuzione normale con devizione standard non nota di piccola dimensione

Se n è piccolo (diciamo $n < 40$ anche se non c'è un valore rigoroso ovviamente), allora considero la variabile aleatoria

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

che ha in questo caso distribuzione t di student con 1 grado di libertà, cioè `t(df=1)`. Quindi il calcolo dell'intervallo di confidenza, come abbiamo visto nella sezione precedente, si differenzia rispetto al caso della distribuzione normale perché si devono utilizzare i quantili $t_{\alpha/2}$ della distribuzione t di student anzichè della distribuzione normale standard.

Questo test viene chiamato *t-test*.

In R, la funzione `t.test` esegue il t-test.

Esempio 6.20 *Eseguiamo il test di ipotesi sulla media a partire da un campione SRS(32) estratto da una distribuzione normale con media* $\mu = 2$*, utlizzando la funzione* `t.test` *di* R*.*

Soluzione.

```
> n <- 32
> x <- rnorm(n, 2, 3)
> t.test(x, mu = 0, conf.level = 0.9,
+        alternative = "greater")

##
##  One Sample t-test
##
## data:  x
## t = 4.2027, df = 31, p-value = 0.0001037
## alternative hypothesis: true mean is greater than 0
## 90 percent confidence interval:
##  1.192132      Inf
## sample estimates:
## mean of x
##   1.73169
```
