Test di ipotesi

Spesso occorre stabilire se il parametro sconosciuto $\psi(\theta)$ sia o no di un certo tipo. Dunque non si deve darne una stima, ma piuttosto formulare su di esso una qualche ipotesi e utilizzare un campione statistico per confermare o rigettare l'ipotesi.

Supponiamo per esempio che una ditta farmaceutica dichiari che ogni compressa di un certo nuovo medicinale contenga mediamente almeno 25 mg di principio attivo. Per verificare questa affermazione, si prende una scatola di 30 pasticche e si misura la quantità di principio attivo in ciascuna pasticca, ottenendo un valore di 24.0 mg con una deviazione standard di 0.5 mg. Questi valori sono compatibili o no con quanto dichiarato?

Se decideremo, in base ai nostri dati, che la media μ di principio attivo in una pasticca sia effettivamente ≥ 25 mg, riterremo il farmaco efficace e lo metteremo in commercio; se invece decideremo di ritenere che $\mu < 25$ mg, giudicheremo il farmaco inefficace e ne vieteremo il commercio.

Tra le due condizioni possibili, una è detta H_I (ipotesi) l'altra è detta H_A (alternativa). Ad esse corrispondono due insiemi D_I, D_A disgiunti, con $D_I \cup D_A = D$, tali che valga H_I se e solo se $\psi(\theta) \in D_I$, e valga H_A se e solo se $\psi(\theta) \in D_A$. Nel caso del farmaco, $H_I: \mu \leq 25$ e $D_I = [0, 25]$, $H_A: \mu \geq 25$ e $D_A = [25, \infty[$.

Formalizziamo il problema. Sia $(\Omega, \mathcal{A}, \{P^\theta : \theta \in D\})$ un modello statistico, sia $X_1, \dots, X_N$ un campione statistico di taglia N e sia $T = t(X_1, \dots, X_N)$ una statistica che sarà uno stimatore del parametro $\psi(\theta)$.

In base ai valori assunti da T sul campione, dovremo decidere tra le due possibilità in esame. Ci saranno dei valori $\psi(\theta)$, appartenenti ad un certo sottoinsieme $K \subset \mathbb{R}$, che ci faranno propendere per H_A, mentre i valori $\psi(\theta)$ appartenenti a K^c ci faranno preferire H_I. Si chiama <u>regione critica</u> (o <u>di rigetto</u>) del test l'evento $\{T \in K\}$, mentre si chiama <u>regione di accettazione</u> l'evento complementare $\{T \in K^c\}$.

La regione critica decide dunque il risultato del test, secondo la regola che, qualunque sia il risultato ω dell'esperimento, si rifiuta l'ipotesi H_I se $T(\omega) \in K$, mentre la si accetta se $T(\omega) \in K^c$. Come dovremo scegliere la regione critica? Osserviamo che, tradizionalmente, si sceglie come ipotesi H_I la "peggiore" o la più sfavorevole delle due, e come ipotesi H_A l'altra. Ora, comunque si scelga la regione critica $\{T \in K\}$, vi è una probabilità positiva che l'ipotesi sia vera anche se l'osservazione ω appartiene $\{T \in K\}$, e che quindi venga respinta a torto l'ipotesi; questo è l'<u>errore di 1ª specie</u>.

Analogamente, vi è una probabilità positiva che l'ipotesi sia falsa anche se l'osservazione ω appartiene a $\{T \in K^c\}$, e che quindi si accetti a torto l'ipotesi errata: questo è l'<u>errore di 2ª specie</u>. Pensando all'esempio del farmaco, si capisce che l'errore di 1ª specie è più grave di quello di 2ª specie.

Osserviamo che, quando $\psi(\theta) \in D_I$, il numero $P^\theta(T \in K)$ è la probabilità di respingere a torto l'ipotesi H_I, supposto che $\psi(\theta)$ sia il vero valore del parametro; quindi, se $\psi(\theta) \in D_I$, $P^\theta(T \in K)$ è esattamente la probabilità di commettere un errore di 1ª specie. Invece, quando $\psi(\theta) \in D_A$, il numero $P^\theta(T \in K^c)$ è la probabilità di accettare l'ipotesi malgrado questa sia falsa, sempre supponendo che $\psi(\theta)$ sia il vero valore del parametro. Quindi, se $\psi(\theta) \in D_A$, $P^\theta(T \in K^c) = 1 - P^\theta(T \in K)$ è la probabilità di commettere un errore di 2ª specie.

Se ne deduce che, per $\psi(\theta) \in D_A$, il numero $P^\theta(T \in K)$ è la probabilità di prendere la decisione corretta, ossia di respingere (a ragione) l'ipotesi H_I. Ciò motiva la seguente definizione:

<u>Definizione</u> La funzione $\theta \mapsto P^\theta(T \in K)$, $\psi(\theta) \in D_A$, è detta <u>potenza del test</u>. Si dice che <u>il test è di livello $\alpha > 0$</u> se

$$\sup_{\psi(\theta) \in D_I} P^\theta(T \in K) \leq \alpha .$$

Si richiederà che α sia piccolo (valori tipici: $\alpha = 0.05$, $\alpha = 0.01$).

La relazione che definisce un test di livello α ci garantisce una piccola probabilità di errore di 1ª specie.
Una volta fissato il livello del test, si è determinato K in funzione di α e si è determinata la potenza del test.
Può capitare che per certi valori $\psi(\theta) \in D_A$, la potenza sia bassa: se, ad esempio, per $\psi(\theta_0) \in D_A$ si avesse $P^{\theta_0}(T \in K) = 0.25$, avremmo una probabilità del 25% di respingere H_I quando essa è falsa: troppo poco. Una situazione del genere può capitare per valori di N troppo piccoli (poche osservazioni) o per valori di α troppo piccoli (livello del test troppo basso).
D'altra parte, aumentando N oppure α si accresce la potenza del test, ma aumentano le difficoltà pratiche e i costi (se N cresce) e aumenta il rischio di errori di 1ª specie (se α cresce).
In definitiva, come si è detto, se le nostre osservazioni $T(\omega) \in K$, rifiuteremo l'ipotesi H_I, e sbaglieremo, se $\psi(\theta) \in D_I$, al più con probabilità $\leq \alpha$; se $T(\omega) \in K^c$, accetteremo l'ipotesi H_I.

<u>Esempio</u> Nella settimana successiva al suicidio di un famoso matematico napoletano, in città si sono registrati 12 suicidi, contro una media settimanale di 8. Si può dire che si sia verificato un fenomeno di imitazione?
Sia p la probabilità che un cittadino di Napoli si suicidi, e che ciascun cittadino decida o no di uccidersi in modo

indipendente degli altri. Allora il numero X di suicidi in una settimana sarà una v.a. con legge $\mathcal{B}(n,p)$, ove n è il numero di cittadini. Poiché n è molto grande e p è molto piccolo, possiamo supporre che X abbia legge di Poisson $\mathcal{P}(np) = \mathcal{P}(8)$ (perché $np = \mathbb{E}[X] = 8$ è la media settimanale).

Dire che vi è stato un fenomeno di imitazione vuol dire che X ha legge di Poisson $\mathcal{P}(\lambda)$ con $\lambda > 8$. Stabiliamo che

$$H_I : \lambda = 8 , \qquad H_A : \lambda > 8 ,$$

cioè facciamo l'ipotesi che non vi sia stata imitazione (anche questo caso non è il più sfavorevole).

Un test ragionevole è allora quello di respingere l'ipotesi se il parametro $\overline{X}$ del nostro campione è troppo grande.

Scegliamo come regione critica $\{X \geq k\}$, ove k è tale che

$$\sup_{\lambda = 8} P^{\lambda}(X \geq k) = P^{8}(X \geq k) \leq 0.05$$

(qui $D_I = \{8\}$ e $\alpha = 0.05$). Dunque deve essere

$$P^{8}(X \geq k) = \sum_{h=k}^{\infty} \frac{8^h e^{-8}}{h!} \leq 0.05$$

Stime numeriche della legge di Poisson forniscono

$$P^{8}(X \geq 12) = 0.112, \quad P^{8}(X \geq 13) = 0.064, \quad P^{8}(X \geq 14) = 0.034 .$$

Dunque 14 è il più piccolo dei numeri k tali che $P^{8}(\overline{X} \geq k) \leq 0.05$.

La nostra media empirica è $\overline{X} = 12$, che non appartiene alla regione critica. Non possiamo quindi rifiutare l'ipotesi, e dunque non possiamo dire che vi sia stata imitazione, almeno al livello 0.05. Se i suicidi fossero stati almeno 14, avremmo potuto prendere la decisione opposta.

Test di Student

Questo test riguarda la media di una popolazione (senza conoscerne la varianza). Supponiamo di osservare un campione di N v.a. indipendenti $X_1, \ldots, X_N$ e di voler stabilire se la media μ del campione è uguale, oppure no, ad un fissato valore μ_0. Dobbiamo dunque realizzare un test con

$$H_I : \mu = \mu_0, \qquad H_A : \mu \neq \mu_0.$$

A questo scopo, consideriamo la media empirica $\bar{X}$, che è uno stimatore corretto di μ, e la varianza empirica S^2, che è uno stimatore corretto della varianza σ. Se il campione è gaussiano, o se N è sufficientemente grande per usare l'approssimazione normale, allora $T = \frac{\bar{X}-\mu_0}{S}\sqrt{N}$ ha legge $t(N-1)$. Poiché è ragionevole supporre che, per $\mu \neq \mu_0$, la statistica $|T|$ tenda ad assumere valori grandi, cerchiamo una regione critica della forma $\{|T| > \delta\}$. Fissato il livello α del test, cercheremo δ tale che

$$\sup_{\mu=\mu_0} P^{\mu}(|T|>\delta) = P^{\mu_0}(|T|>\delta) \leq \alpha.$$

Si ha

$$P^{\mu_0}(|T|>\delta) \leq P^{\mu_0}(T\geq\delta) + P^{\mu_0}(T\leq-\delta) =$$
$$= 1 - P^{\mu_0}(T\leq\delta) + 1 - P^{\mu_0}(T\leq\delta) =$$
$$= 2 - 2P^{\mu_0}(T\leq\delta) = \alpha$$

se e solo se

$$P^{\mu_0}(|T|\leq\delta) = 1-\alpha/2$$

ovvero

$$\delta = t_{1-\alpha/2}(N-1).$$

Si ha dunque $P^{\mu_0}(|T|>\delta)\leq\alpha$ se e solo se $|T|\geq t_{1-\alpha/2}(N-1)$.
Basterà allora calcolare la statistica T sul campione e confrontarla col quantile $t_{1-\alpha/2}(N-1)$: se quest'ultimo è minore, dovremo rigettare l'ipotesi $\mu=\mu_0$.

Se si vuole invece verificare

$$H_I: \mu\leq\mu_0 \quad \text{contro} \quad H_A: \mu>\mu_0,$$

cerchiamo una regione critica della forma $\{T\geq\delta\}$, ove ancora $T=\frac{\bar{X}-\mu_0}{S}$. Però il vero valore della media è μ; ponendo $T'=\frac{\bar{X}-\mu}{S}\sqrt{N}$; se l'ipotesi è vera, ossia $\mu\leq\mu_0$, possiamo scrivere $T = T' + \frac{\mu-\mu_0}{S}\sqrt{N} \leq T'$, e dunque

$$P^{\mu}(T\geq\delta) \leq P^{\mu}(T'\geq\delta) = 1 - P^{\mu}(T'\leq\delta) = \alpha$$

se e solo se $P^{\mu}(T'\leq\delta) = 1-\alpha$, ossia $\delta = t_{1-\alpha}(N-1)$; dunque

$$\sup_{\mu\leq\mu_0} P^{\mu}(T\geq\delta)\leq\alpha \iff T\geq t_{1-\alpha}(N-1).$$

Se dunque il campione T supera $t_{1-\alpha}(N-1)$, dovremo rigettare l'ipotesi $\mu \leq \mu_0$.

Esempio Nel 1957 l'altezza media degli uomini di un certo paese era di $\mu_0 = 170$ cm.. Dieci anni dopo, un campione di 81 uomini mostrò un'altezza media $\bar{X} = 171$ cm, con varianza $S^2 = 16$ cm^2. Si può dire che l'altezza media in 10 anni sia cresciuta, a un livello di fiducia $\alpha = 0.05$?

Si ha $H_I : \mu = \mu_0$, $H_A = \mu > \mu_0$. Calcoliamo

$$|T| = \left|\frac{\bar{X} - \mu_0}{S}\right| \sqrt{N} = \frac{171 - 170}{4}\sqrt{81} = \frac{9}{4} = 2.25,$$

mentre $t_{0.975}(80) = 1.88$. Dunque, rigettiamo l'ipotesi: almeno a livello 0.05, l'altezza media è cresciuta.

Test di Fischer-Snedecor

Questo test riguarda la varianza σ^2 di un campione $X_1, \dots, X_N$ di v.a. indipendenti con media μ sconosciuta.

Si ha $H_I : \sigma^2 \leq \sigma_0^2$, $H_A : \sigma^2 > \sigma_0^2$,

ove σ_0 è un valore fissato. Consideriamo la varianza empirica S^2: cerchiamo una regione critica della forma

$$\{S^2 \geq \sigma_0^2 (1+\delta)\}.$$

Consideriamo $W = \frac{S^2}{\sigma^2}(N-1)$, che ha legge $\chi^2(N-1)$: allora

$$\{S^2 \geq \sigma_0^2(1+\delta)\} = \left\{W \geq \frac{\sigma_0^2}{\sigma^2}(N-1)(1+\delta)\right\},$$

e quindi, se l'ipotesi $\sigma^2 \leq \sigma_0^2$ è vera,

$$P^\sigma(S^2 \geq \sigma_0^2(1+\delta)) = P^\sigma\left(W \geq \frac{\sigma_0^2}{\sigma^2}(N-1)(1+\delta)\right) \leq$$

$$\leq P^\sigma(W \geq (N-1)(1+\delta)) = \alpha$$

se e solo se

$$1 - P^\sigma(W \leq (N-1)(1+\delta)) = \alpha,$$

ossia

$$P^\sigma(W \leq (N-1)(1+\delta)) = 1-\alpha,$$

vale a dire

$$(N-1)(1+\delta) = \chi^2_{1-\alpha}(N-1).$$

Perciò, se l'ipotesi è vera, e se $\delta = \frac{\chi^2_{1-\alpha}(N-1)}{N-1} - 1$, si ha

$$\sup_{\sigma^2 \leq \sigma_0^2} P^\sigma(S^2 \geq \sigma_0^2(1+\delta)) \leq \alpha,$$

e quindi rigetteremo l'ipotesi se per la varianza empirica si ha $S^2 \geq \sigma_0^2(1+\delta)$ con tale δ.

Se si vuole invece verificare

$$H_I: \sigma^2 = \sigma_0^2, \text{ contro } H_A: \sigma^2 \neq \sigma_0^2,$$

si prende una regione critica della forma

$$\{S^2 \geq \sigma_0^2(1+\delta)\} \cup \{S^2 \leq \sigma_0^2(1-\beta)\};$$

Sotto l'ipotesi $\sigma^2 = \sigma_0^2$, la v.a. $W = \frac{S^2}{\sigma^2}(N-1) = \frac{S^2}{\sigma_0^2}(N-1)$ ha legge $\chi^2(N-1)$ e si ha

$$P^{\sigma_0}\left(\{S^2 \geq \sigma_0^2(1+\delta)\} \cup \{S^2 \leq \sigma_0^2(1-\beta)\}\right) = P^{\sigma_0}\left(\{W \geq (N-1)(1+\delta)\} \cup \{W \leq (N-1)(1-\beta)\}\right) =$$

$$= 1 - P^{\sigma_0}\left(W \leq (N-1)(1+\delta)\right) + P^{\sigma_0}\left(W \leq (N-1)(1-\beta)\right) = 1 - \left(1-\frac{\alpha}{2}\right) + \frac{\alpha}{2} = \alpha,$$

purché

$$\delta = \frac{\chi^2_{1-\alpha/2}(N-1)}{N-1} - 1, \quad \beta = 1 - \frac{\chi^2_{\alpha/2}(N-1)}{N-1}.$$

Se la varianza empirica del campione verifica $S^2 \geq \sigma_0^2(1+\delta)$ oppure $S^2 \leq \sigma_0^2(1-\beta)$ con tali δ e β, dovremo rigettare l'ipotesi.

<u>Esempio</u> Una macchina che riempie i barattoli di caffè funziona correttamente se il peso dei barattoli ha una varianza $\sigma^2 \leq 15 g^2$. Su un campione di 20 barattoli si rileva una varianza empirica $S^2 = 25\ g^2$. Si può dire, a livello 0.01, che vi sarà un cattivo funzionamento della macchina?

Poniamo $H_I : \sigma^2 \leq \sigma_0^2 = 15$, $H_1 = \sigma^2 > 15$.

Dobbiamo confrontare $S^2 = 25$ con $\sigma_0^2(1+\delta) = 15\,\frac{\chi^2_{0.99}(19)}{19}$.

Risulta $\chi^2_{0.99}(19) = 36.191$, dunque

$$\sigma_0^2(1+\delta) = 28.57.$$

Poiché $S^2 = 25 < 28.57$, l'ipotesi non può essere rigettata e non si può dire che vi sarà malfunzionamento.