Verifica di ipotesi e p-value

Un'introduzione minimale

Luca La Rocca[1]

Dipartimento di Scienze Fisiche, Informatiche e Matematiche
Università degli Studi di Modena e Reggio Emilia

Insegnamento di Analisi Statistica dei Dati
Corsi di Laurea Magistrale in Informatica e Matematica
Anno Accademico 2018/2019

[1] `http://personale.unimore.it/rubrica/dettaglio/llarocca`

Ipotesi nulla

Dato un modello statistico, indicizzato dal parametro $\theta \in H$, per il dato $z \in \mathcal{Z}$, si dice ipotesi statistica un suo sottomodello (senz'altro individuato da un sottoinsieme H_0 di H).

Un'ipotesi $H_0 \subset H$ si dice ipotesi nulla quando rappresenti uno "status quo" contro il quale si cerca evidenza nel dato; riterremo $\theta \in H_0$ sino a prova contraria (fornita da z).

Per esempio, se consideriamo il modello esponenziale per i tempi di sopravvivenza non censurati della tabella 1.4, nel testo di riferimento, supponendoli i.i.d. Exp$(1/\mu)$, può interessarci come ipotesi nulla

$$H_0 : \quad \mu = 1000 \quad (\textit{giorni})$$

individuata dal singoletto $\{\mu_0\} = \{1000\}$ di $H = \mathbb{R}_+^* \ldots$

Ipotesi semplici e composite

. . . si dice semplice un'ipotesi statistica individuata da un singolo valore del parametro e composita un'ipotesi statistica che non sia semplice.

Per esempio, se consideriamo $X \sim \text{Binom}(m, \pi)$ e $Y \sim \text{Binom}(n, \psi)$, indipendenti, per i conteggi della tabella 3.1, nel testo di riferimento, definito il rapporto dei pronostici $\rho = \{\pi/(1 - \pi)\}/\{\psi/(1 - \psi)\}$, potrebbe interessarci come ipotesi nulla

$$H_0 : \quad \rho = 1$$

individuata dal sottospazio parametrico $\{(\pi, \psi) \in]0, 1[^2 : \pi = \psi\}$.

Discrepanza dall'ipotesi nulla

Supponiamo che $d_0(z)$ quantifichi la discrepanza tra H_0 e z,
nel senso che $d_0(z) = 0$ se z è perfettamente compatibile con H_0,
mentre $d_0(z) < \underline{d} < 0$ o $d_0(z) > \overline{d} > 0$ indicano scarsa compatibilità.

Avremo tipicamente

$$d_0(z) = \inf_{\theta \in H_0} d(\theta, z),$$

o più semplicemente $d_0(z) = d(\theta_0, z)$ se $H_0 = \{\theta_0\}$,
dove $d(\theta, z)$ quantifica la discrepanza tra θ e z
(con $d(\hat{\theta}(z), z) = 0$ per un opportuno stimatore $\hat{\theta}$ per θ
come in particolare quello di massima verosimiglianza).

Statistica test

Nell'esempio del modello esponenziale, possiamo prendere

$$d(\mu, x_{1:n}) = \frac{\bar{x}_n - \mu}{\mu} = \frac{\bar{x}_n}{\mu} - 1,$$

di modo che dati scarsamente compatibili con μ_0 corrisponderanno a valori particolarmente bassi o alti della statistica test $n\bar{X}_n/\mu_0$ (ottenuta dal pivot $n\bar{X}_n/\mu \sim \text{Gamma}(n, 1)$ ponendo $\mu = \mu_0$).

Valori "estremi" della statistica test ci forniranno evidenza "sufficiente" per rifiutare H_0; altrimenti, visto che $\theta \in H_0$ sino a prova contraria, finiremo per accettare H_0 (almeno provvisoriamente).

Regione di rifiuto

Diremo regione di rifiuto l'insieme dei dati scarsamente compatibili con l'ipotesi nulla:

$$\mathcal{R} = \{z \in \mathcal{Z} \mid d_0(z) < \underline{d} \text{ o } d_0(z) > \overline{d}\}$$

nel caso bilaterale o

$$\mathcal{R} = \{z \in \mathcal{Z} \mid d_0(z) < \underline{d}\}$$

nel caso unilaterale sinistro e

$$\mathcal{R} = \{z \in \mathcal{Z} \mid d_0(z) > \overline{d}\}$$

nel caso unilaterale destro.

Ipotesi alternativa

Si dice ipotesi alternativa il complemento dell'ipotesi nulla in H: $H_1 = H \setminus H_0$.

Nell'esempio del modello esponenziale, avremo

$$\mathcal{R} = \{x_{1:n} \in \mathbb{R}_+^n \mid n\bar{x}_n/\mu_0 < \underline{g}_n \text{ o } n\bar{x}_n/\mu_0 > \overline{g}_n\},$$

se $H_1 = \{\mu \in \mathbb{R}_+^* \mid \mu \neq \mu_0\}$, cioè nel caso bilaterale, oppure

$$\mathcal{R} = \{x_{1:n} \in \mathbb{R}_+^n \mid n\bar{x}_n/\mu_0 > \overline{g}_n\},$$

se $H_1 = \{\mu \in \mathbb{R}_+^* \mid \mu > \mu_0\}$, cioè nel caso unilaterale destro; in questo caso $H = [\mu_0, \infty[$ o $H_0 =]0, \mu_0]$.

Analogamente nel caso unilaterale sinistro. Si noti che è l'ipotesi alternativa a determinare il tipo di regione di rifiuto.

Errori di prima e seconda specie

Si dice errore di prima specie l'errore commesso quando $\theta \in H_0$; poiché $\theta \in H_0$ sino a prova contraria, cercheremo di controllare la probabilità di commetterlo

$$\alpha = \mathbb{P}_{\theta_0}\{X_{1:n} \in \mathcal{R}\} = \sup_{\theta \in H_0} \mathbb{P}_{\theta}\{X_{1:n} \in \mathcal{R}\},$$

per esempio richiedendo $\alpha = 0.05$ (livello di significatività).

Si dice errore di seconda specie l'errore commesso quando $\theta \in H_1$; poiché tipicamente $\sup_{\theta \in H_1} \mathbb{P}_{\theta}\{X_{1:n} \in \mathcal{R}\} = \alpha$, dobbiamo accontentarci di valutare la potenza della procedura di verifica di ipotesi

$$\beta = \mathbb{P}_{\theta_1}\{X_{1:n} \in \mathcal{R}\}$$

in corrispondenza di un particolare valore alternativo per θ.

Il p-value

In pratica si preferisce spesso calcolare

$$p = \mathbb{P}\{d_0(Z) > d_0(z^\bullet)\}$$

in caso di alternativa unilaterale destra e

$$p = \mathbb{P}\{d_0(Z) < d_0(z^\bullet)\}$$

in caso di alternativa unilaterale sinistra, dove $z^\bullet$ è il dato osservato, raddoppiando il valore ottenuto in caso di alternativa bilaterale.

In questo modo si ottiene il minimo valore di α per cui si rifiuta H_0 e si finisce per rifiutare H_0 quando il valore p, detto anche livello di significatività osservato, è “piccolo”.

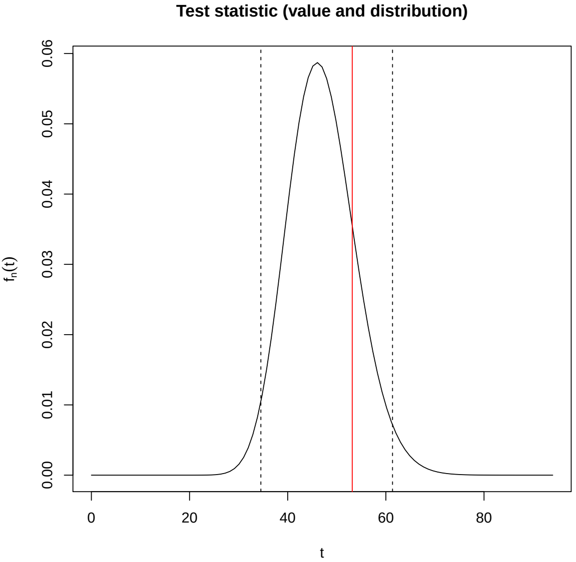
Un p-value

Nell'esempio del modello esponenziale per i tempi di sopravvivenza non censurati della tabella 1.4, nel testo di riferimento, considerando per semplicità l'alternativa unilaterale $\mu > \mu_0$, possiamo calcolare

$$p = \mathbb{P}_{\mu_0}(n\bar{X}_n/\mu_0 > n\bar{x}_n^\bullet/\mu_0) = 0.1819,$$

dove $n\bar{x}_n^\bullet/\mu_0 = 53.146$ è il valore osservato della statistica test $n\bar{X}_n/\mu_0$ ($n = 47$, $\bar{x}_n^\bullet = 1130.766$ e $\mu_0 = 1000$).

Il p-value così ottenuto <u>non</u> è abbastanza piccolo da rifiutare H_0, perché dovremmo accettare un errore di prima specie del 18%; lo sarebbe ancora meno con l'alternativa bilaterale ($p = 0.3638$).



Test statistic (value and distribution)
0.00
0.01
0.02
0.03
0.04
0.05
0.06
$f_n(t)$
0
20
40
60
80
t

Il test esatto di Fisher

Nel caso dell'ipotesi nulla

$$H_0 : \quad \rho = 1$$

sul rapporto dei pronostici $\rho = \{\pi/(1-\pi)\}/\{\psi/(1-\psi)\}$
relativo a $X \sim \text{Binom}(m, \pi)$ e $Y \sim \text{Binom}(n, \psi)$, indipendenti,
si può usare come statistica test $X \mid X + Y = x^\bullet + y^\bullet$
la cui distribuzione, se $\rho = 1$, è ipergeometrica
($x^\bullet + y^\bullet$ estrazioni, $m + n$ biglie in tutto, m biglie bianche)
qualunque sia il valore di $\pi = \psi$.

Significatività asintotica

Se $T_n = g_n(X_{1:n})$ è uno stimatore asintoticamente normale (di ordine $1/\sqrt{n}$) per $\xi = h(\theta)$, con errore standard asintotico $\sqrt{V_n/n}$, possiamo usare

$$Z_n = \frac{T_n - \xi_0}{\sqrt{V_n/n}} \approx \mathrm{Norm}(0, 1)$$

come statistica test per calcolare un p-value approssimato contro l'ipotesi nulla $h(\theta) = \xi_0$.

Troveremo $p = 2\{1 - \Phi(|z_n^\bullet|)\}$, con l'alternativa bilaterale $h(\theta) \neq \xi_0$, dove $z_n^\bullet$ è il valore osservato di Z_n e Φ la funzione di ripartizione normale standard.

Dualità con gli intervalli di confidenza

Se $d_0(z) = d(\theta_0, z)$ e $d(\theta, Z)$ è un pivot,
la cui inversione produce l'intervallo di confidenza $[\underline{T}, \overline{T}]$ per θ,
al livello γ, allora

$$Z \in \mathcal{R} \qquad \Leftrightarrow \qquad \theta_0 \notin [\underline{T}, \overline{T}]$$

e $\mathbb{P}_{\theta_0}\{Z \in \mathcal{R}\} = 1 - \gamma$.

Si noti che $\mathbb{P}_\theta\{\underline{T} \leq \theta \leq \overline{T}\} = \gamma$ è un'affermazione probabilistica relativa a un evento auspicabilmente osservato, valida per qualunque valore del parametro θ, mentre $p = \mathbb{P}\{d_0(Z) > d_0(z^\bullet)\}$ è un'affermazione probabilistica relativa a un evento non osservato, valida per un valore del parametro che stiamo mettendo in dubbio...

Una prassi discutibile

…ciò nonostante si è affermato l'uso di richiedere $p < 0.05$ come precondizione per sostenere la validità di una nuova scoperta.

Al riguardo, occorre tenere presente che ci sono buone ragioni per abbassare la soglia di un ordine di grandezza

https://www.nature.com/articles/s41562-017-0189-z

e più in generale per abbandonare del tutto la prassi in questione

http://andrewgelman.com/2017/09/26/abandon-statistical-significance/

ridimensionando il ruolo del p-value nell'analisi dei dati.

Due caveat e uno spunto di lettura

- Il p-value non è la probabilità che l'ipotesi nulla sia vera.
- Il p-value non può fornire evidenza in favore dell'ipotesi nulla.
- The American Statistical Association's statement on p-values

  http://amstat.tandfonline.com/doi/abs/10.1080/00031305.2016.1154108#.WfKCl4bONE4

  is a recent synthesis on null hypothesis significance testing.