La Statistica e il Metodo Scientifico

Donata Marasini

Piero Quatto

Università degli Studi di Milano – Bicocca

Dipartimento di Economia, Metodi quantitativi e Strategie d’impresa

Metodo Scientifico

Continua

Rappresentazione=Modello

Induzione

Osservazione

Analisi dei dati

Esempi

In campo medico

La ricerca di base e la ricerca pre clinica fanno uso della prima figura

La ricerca clinica, che lavora sui pazienti, fa uso della seconda

In campo economico

Il controllo di qualità fa uso della seconda

Statistica descrittiva e inferenziale

La conoscenza parziale della realtà può essere raggiunta con:

- l'analisi descrittiva e esplorativa dei dati
- l'analisi induttiva realizzata tramite la sperimentazione intesa in senso lato.

Statistica descrittiva nel primo caso

Statistica inferenziale nel secondo.

I Modelli

Un **terzo filone** si identifica nella spiegazione di un fenomeno in funzione di altri, predisponendo modelli in termini probabilistici. Modelli particolarmente impiegati:

- Regressione multipla (Variabile dipendente quantitativa)
- Regressione logistica (Variabile dipendente qualitativa)
- Equazioni strutturali (Variabile dipendente latente)

Argomenti

- Stima puntuale di una grandezza e sue proprietà
- Il supporto dell’intervallo di confidenza alla stima puntuale
- Il modello asintotico
- I test di significatività
- Il P-value
- Il teorema di Bayes
- Punti di criticità

Probabilità: divergenze nell'interpretazione

Frequentista basata sulle frequenze relative nel lungo andare di Richard von Mises (1883-1953) del 1928

Soggettivista coincidente con il grado di fiducia misurato sulla coerenza di una scommessa di Bruno de Finetti (1906-1985) del1935

Assiomatica di Andrei Kolmogorov (1903-1987) del1933.

Basi dell'inferenza statistica

Fenomeno ***X*** che interessa.

Si ipotizza un modello che interpreta ***X*,** ovvero una v.c. (variabile casuale) X

n misure temporali e indipendenti $(x_1,\ldots,x_n)$ ottenute da un esperimento.

Lo strumento di misura può non essere perfetto; le condizioni sperimentali non sono state mantenute costanti.

Continua

La generica rilevazione x_i può considerarsi suddivisa in $x_i = \mu + \varepsilon_i$ dove μ è la “vera” misura e ε_i è l’errore dovuto al caso che agisce e la fa allontanare dal parametro ignoto .

Se l’esperimento viene idealmente ripetuto un numero “infinito” di volte x_i descrive una v.c . X_i identica a X

La presenza dell’errore porta spesso a considerare X_i Normale con media μ e varianza σ^2

Stimatori e stime

Una **stima** è una funzione del campione, cioè una sua sintesi. Nel caso che il parametro sia μ la stima è

$$\overline{x}_n = \sum_i x_i / n$$

Uno **stimatore** è una funzione (misurabile) del campione casuale $\boldsymbol{X} = (X_1,\ldots,X_n)$ con il compito di fornire stime del parametro di interesse. Nel caso di μ

$$T = \overline{X}_n = \sum_i X_i / n$$

è detta **media campionaria** e $\overline{x}_n$ è la stima corrispondente.

Stimatori corretti

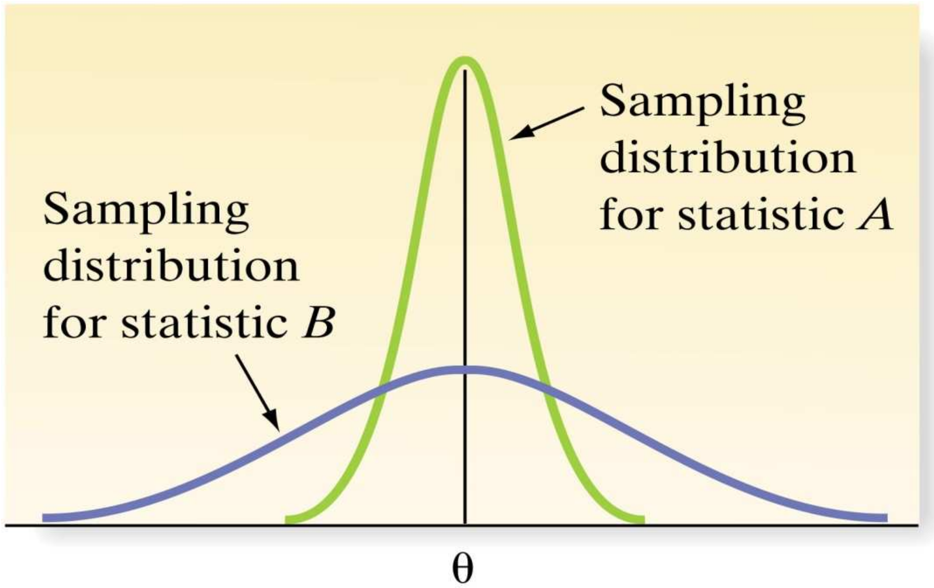
Uno stimatore è **corretto** se in media riesce a cogliere l'ignoto valore del parametro d'interesse θ

$$E(T) = \theta$$

ovvero se ripetendo le misure le sottostime e le sovrastime si bilanciano

Ad esempio

$$E(\sum_i X_i /n) = \mu$$



Errore e livello di confidenza

Non è possibile stabilire se $\left|\bar{x}_n - \mu\right| < \varepsilon$

ma è possibile valutare $P\left(\left|\bar{X}_n - \mu\right| < \varepsilon\right)$

determinando l'**errore** ε al **livello di confidenza prefissato** (1 – *α*).

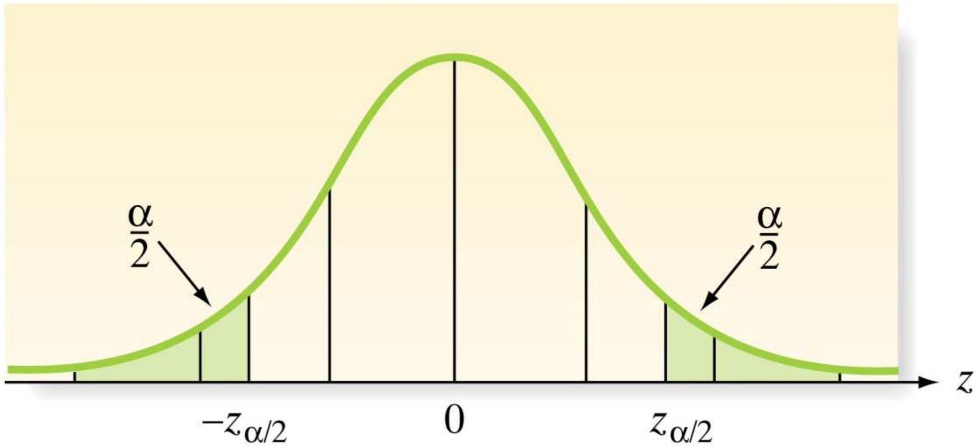
Continua

Si impone l'uguaglianza

$$P\left(\left|\overline{X}_n - \mu\right| < \varepsilon\right) = 1 - \alpha$$

e si identifica l'errore

$$\varepsilon \cong z_{1-\alpha/2} \sqrt{s_n^2 / n}$$



$$z_{1-\alpha/2}$$

è un quantile e

$$s_n^2 = \frac{1}{n-1} \sum_i (x_i - \overline{x})^2$$

Intervallo di confidenza e interpretazione di $(1-\alpha)$

$$\bar{x}_n - \varepsilon < \mu < \bar{x}_n + \varepsilon$$

ossia

$$|z_n| < z_{1-\alpha/2} \Leftrightarrow \bar{x}_n - z_{1-\alpha/2}\sqrt{s_n^2/n} < \mu < \bar{x}_n + z_{1-\alpha/2}\sqrt{s_n^2/n}$$

Se si ripetesse più volte la procedura di stima

ottenendo molti campioni di numerosità *n*, nel 100(1 – *α*)% di questi campioni la media campionaria si discosterebbe per non più di *ε* dal vero (ma ignoto) valore del parametro d'interesse *μ*.

Altre distribuzioni*

Lognormale

$$\varphi(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log x - \mu}{\sigma}\right)^2} \qquad x > 0$$

Esponenziale negativa

$$\varphi(x) = \lambda e^{-\lambda x} \qquad x > 0, \lambda > 0$$

Teoria asintotica

X_n successione di variabili casuali i.i.d.

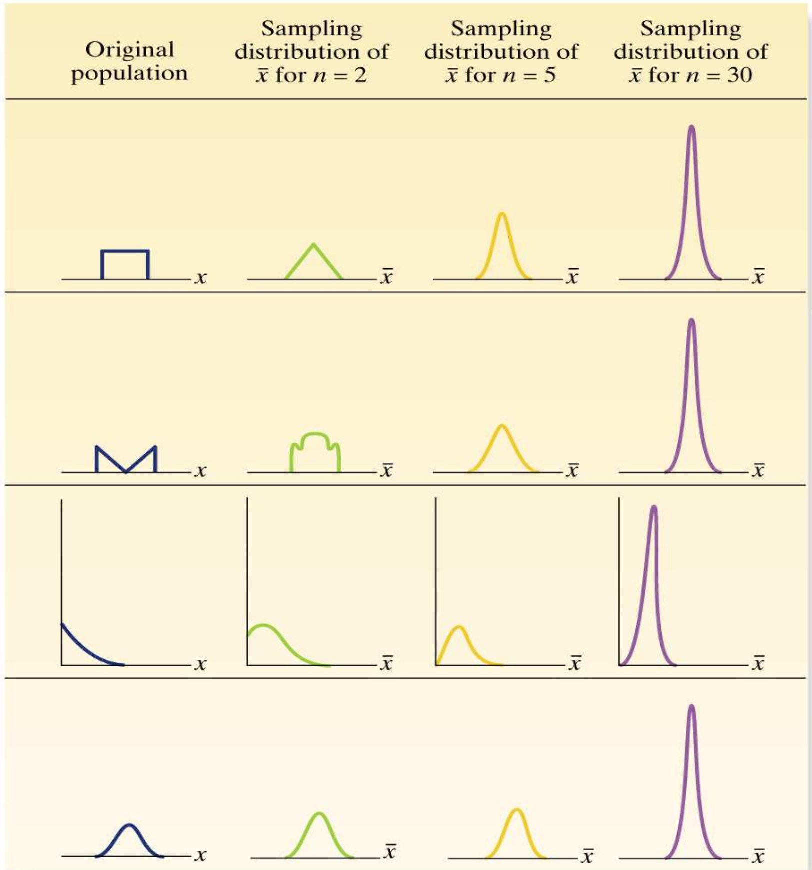
$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Legge dei grandi numeri

$$\overline{X}_n \xrightarrow{P} \mu$$

Teorema centrale del limite

$$\frac{\overline{X}_n - \mu}{\sqrt{\sigma^2/n}} \xrightarrow{D} N(0,1)$$



Test di significatività

Ronald Fisher ha proposto i test di significatività come metodi volti a sondare l’incidenza del caso nei risultati di un esperimento realizzato per la verifica di un’ipotesi di interesse .

Esempio

In **campo fisico**

L’ipotesi sia:

“**nessuna particella è più veloce della luce**”.

Continua

Realizzato l’esperimento e costruita un’opportuna misura di discrepanza tra quanto ipotizzato e quanto osservato, se l’ipotesi dell’azione del caso non può essere confutata allora la ricerca deve andare avanti e formulare nuove Ipotesi, se può esserlo allora è necessario produrre nuove evidenze.

Esempi

Con riguardo all'ipotesi:

"**nessuna particella è più veloce della luce**".

se l'eventuale discrepanza si ritiene dovuta al caso, la ricerca deve andare avanti e formulare nuove ipotesi, altrimenti è necessario produrre nuove evidenze in favore di

"**esistono particelle che hanno una velocità maggiore della luce**"

Continua

In **campo medico**

Un farmaco per l'emicrania sperimentato su due gruppi di pazienti (di controllo e trattati) ha dato effetti positivi.

Al farmaco viene dato molto credito

Controesempio

Un farmaco non specifico sperimentato su analoghi gruppi di pazienti ha dato effetti positivi.

Anche in questo caso viene dato molto credito al farmaco?

Continua

Controllo di qualità

Vi sono esperimenti dove la dualità caso-non caso si trasforma in accettazione o rifiuto dell'ipotesi.

Livello di significatività

Ipotesi $H_0 : \mu = \mu_0$

Misura del divario $$d_n = \frac{\left|\bar{x}_n - \mu_0\right|}{\sqrt{s_n^2 / n}}$$

Regione di rifiuto $d_n \geq z$

Livello di significatività $\alpha = P\left(\text{rifiutare } H_0 \mid H_0 \text{ vera}\right)$

Continua

Variabile casuale divario D_n

Livello di significatività $\alpha = P\left(D_n \geq z \mid H_0 \text{ vera}\right) \cong P\left(\left|Z\right| \geq z\right)$

Quantile $z \cong z_{1-\alpha/2}$

Livello di significatività e “sigma”

I livelli di significatività α sono convenzionalmente espressi in termini di multipli della **deviazione standard** σ di una variabile casuale **Normale** X con media μ.

$$\alpha = P(|X - \mu| \geq z)$$

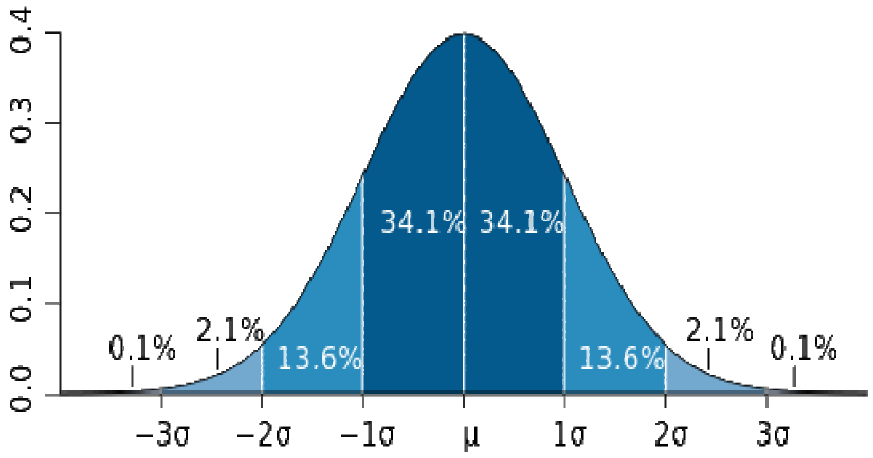
$$\alpha = 0.317310508 \quad z = \sigma$$

$$\alpha = 0.045500264 \quad z = 2\sigma$$

$$\alpha = 0.002699796 \quad z = 3\sigma$$

$$\alpha = 0.000063343 \quad z = 4\sigma$$

$$\alpha = 0.000000573 \quad z = 5\sigma$$



“P-value”

Divario osservato

$$d_n = \frac{\left|\bar{x}_n - \mu_0\right|}{\sqrt{s_n^2 / n}}$$

Livello di significatività osservato o **“P-value”**

$$\hat{\alpha} = P\left(D_n \geq d_n \mid H_0 \text{ vera}\right) \cong P\left(\left|Z\right| \geq d_n\right)$$

“P-value” e livello di significatività

livello di significatività $\alpha = P\left(D_n \geq z \mid H_0 \text{ vera}\right)$

“P-value”

$$\widehat{\alpha} = P\left(D_n \geq d_n \mid H_0 \text{ vera}\right)$$

$$\widehat{\alpha} \leq \alpha \iff d_n \geq z$$

Sintesi

I **test di significatività** sono impiegati per stabilire la compatibilità di un'ipotesi, denominata **ipotesi nulla**, con il campione osservato, in modo che la probabilità di rifiutare tale ipotesi quando è vera non superi un prefissato livello α, detto **livello di significatività**.

Test di Neyman e Pearson

I **test di Neyman e Pearson** sono impiegati per decidere tra un'**ipotesi nulla** H_0 e un'**ipotesi alternativa** H_1 in modo che

(1) la probabilità di rifiutare H_0 quando è vera (probabilità dell'**errore di prima specie**) non superi un prefissato livello α

(2) la probabilità di accettare H_0 quando è falsa (probabilità dell'**errore di seconda specie** e indicata con β) sia minima tra i test con livello α (eventualmente soddisfacenti a ulteriori condizioni).

Test di Neyman e Pearson sulla media

Ipotesi nulla $H_0 : \mu = \mu_0$

Ipotesi alternativa $H_1 : \mu = \mu_1 \qquad (\mu_1 > \mu_0)$

Statistica test $T_n = \left(\overline{X}_n - \mu_0\right) \Big/ \sqrt{S_n^2 / n}$

Regione di rifiuto $T_n \geq z$

Livello del test $\alpha = P(\text{rifiutare } H_0 \mid H_0 \text{ vera})$

$$\alpha = P(T_n \geq z \mid H_0 \text{ vera})$$
$$\beta = P(T_n < z \mid H_1 \text{ vera})$$

Interpretazioni del P-value*

Se il P-value è **piccolo** o si è verificato un evento improbabile, oppure la differenza tra realtà e osservazione non può essere attribuita al fatto che sia venuto un campione piuttosto che un altro (cioè al caso).

Se il P- value è **alto** non si può dire che l'ipotesi è convalidata ma che il test non consente di ritenere che l'ipotesi non sia valida

Ancora sul P-value

Il P-value **non** è la probabilità che l'ipotesi nulla sia vera

Cinque sigma che corrisponde a un valore P di 3×10 elevato alla -7, pari circa a una su 3,5 milioni, non è la probabilità che il bosone di Higgs non esista. (Una coda)

(1−P-value) **non** è la probabilità che sia vera l'ipotesi alternativa

La probabilità che esista l'Higgs è del 99,9 per

cento"

Il P-value **non** è la probabilità di rifiutare l'ipotesi vera

P-value e probabilità delle ipotesi

H_0 = {teoria standard}

E = {osservazioni empiriche}

$$P(H_0 \mid E) = P(E \mid H_0)\frac{P(H_0)}{P(E)}$$

Teorema di Bayes

Ipotesi (incompatibili) H_n

Osservazioni $E \subseteq \bigcup_n H_n$

Probabilità a priori $P(H_m)$

Verosimiglianze $P(E \mid H_m)$

Probabilità a posteriori

$$P(H_m \mid E) = \frac{P(E \mid H_m) P(H_m)}{\sum_n P(E \mid H_n) P(H_n)}$$

ELISA (Italia, 1995)*

I = {Infezione da HIV} $P(\mathrm{I}) = 0.000025$ (prevalenza)

T = {test ELISA positivo} (25 per milione)

$P(\mathrm{T} \mid \mathrm{I}) = 0.993$ (sensibilità)

$P(\overline{\mathrm{T}} \mid \overline{\mathrm{I}}) = 0.9999$ (specificità)

$$\begin{aligned} P(\mathrm{T}) &= P(\mathrm{T} \mid \mathrm{I})\, P(\mathrm{I}) + P(\mathrm{T} \mid \overline{\mathrm{I}})\, P(\overline{\mathrm{I}}) \\ &= P(\mathrm{T} \mid \mathrm{I})\, P(\mathrm{I}) + [1 - P(\overline{\mathrm{T}} \mid \overline{\mathrm{I}})][1 - P(\mathrm{I})] \\ &= 0.000125 \end{aligned}$$

$P(\mathrm{I} \mid \mathrm{T}) = P(\mathrm{T} \mid \mathrm{I})\, P(\mathrm{I}) \,/\, P(\mathrm{T}) = 0.000025 \,/\, 0.000125 = 0.2$

$P(\overline{\mathrm{I}} \mid \mathrm{T}) = 1 - P(\mathrm{I} \mid \mathrm{T}) = 0.8$ (80% di falsi positivi)

ELISA (Italia, 1995)*

$I = \{\text{Infezione da HIV}\}$ $\qquad$ $P(\,I\,) = 0.000025$ (prevalenza)

$T = \{\text{test ELISA positivo}\}$

$P(\,T \mid I\,) = 0.993$ $\qquad$ (sensibilità)

$P(\,\overline{T} \mid \overline{I}\,) = 0.9999$ $\qquad$ (specificità)

$P(\,\overline{T}\,) = 1 - P(\,T\,) = 1 - 0.000125 = 0.999875$

$$
\begin{aligned}
P(\,I \mid \overline{T}\,) &= P(\,\overline{T} \mid I\,)\, P(\,I\,) \,/\, P(\,\overline{T}\,) \\
&= [1 - P(\,T \mid I\,)]\, P(\,I\,) \,/\, P(\,\overline{T}\,) \\
&= 0.000000175 \,/\, 0.999875 = \text{175 per miliardo}
\end{aligned}
$$

$P(\,\overline{I} \mid \overline{T}\,) = 1 - P(\,I \mid \overline{T}\,) = 99.99998\%$

A cavallo fra i secoli XIX e XX

Bruno de Finetti 1906-1985;

Ronald Fisher 1890-1962;

Corrado Gini 1884-1965;

Jerzy Neyman 1894-1981;

Egon Pearson 1895- 1980;

Jimmy Savage 1917- 1971;

Richard von Mises 1883-1953;

Abram Wald 1902-1950

A cavallo fra i secoli XX e XXI

George P. Box 1919-2013;

Bradley Efron 1938;

Erich Leo Lehmann 1917- 2009.

«*Something new is happening in the twenty-first century….*» (B. Efron)

Big Prize, Big Data: Shaw Prize 1 Milione di dollari

David Donoho Professore di Statistica alla Stanford University

Punti critici

E’ giusto eliminare la fase ipotetico-deduttiva del **Metodo scientifico** nel caso della Statistica?

Ha senso parlare di **falsificazione** di un’ipotesi quando viene confutata? (Test di significatività)

Ha senso parlare di confronto fra **paradigmi** quando nella teoria dei test si mettono a confronto due ipotesi e si opta per una delle due? (Test di Neyman-Pearson)

Bibliografia

Bachmann C, Luccio R., Salvadori E. (2005), La verifica della significatività dell'ipotesi nulla, Firenze University Press.

Bellini G. (2005) Divulgazione e metodo scientifico, Dipartimento di Fisica Università di Milano, INFN - Sezione di Milano.

Costantini D., Geymonat L. (1982), Filosofia della probabilità, Feltrinelli, Milano.

D'Agostini G. (2012), Scoperte scientifiche annunciate a colpi di sigma, Dipartimento di Fisica, Roma La Sapienza.

D'Agostini G. (2011), Probably a discovery: bad mathematics means rough scientific communications, ArXiv:1112.3620.v2.

de Cristofaro R. (1998), "I principi della statistica", Giappichelli, Torino.

de Cristofaro R. (2003), The analytical solution to the problem of statistical induction, Statistica Vol. LXIII, n.2.

Continua

de Finetti B. (1970), Teoria della probabilità, Boringhieri, Torino.

Efron B. (2008), Microarrays, empirical Bayes and the two groups model, Statistical Science, Vol. 23, n.1.

Fisher R.A. (1956), Statistical methods and scientific inference, Oliver and Boyd, London.

Giorgi G.M. (2011), Corrado Gini: the man and the scientist, Metron Vol.LXIX, n. 1.

Kuhn T.S. (1962) The structure of scientific revolutions, University of Chicago Press.

Landenna G., Marasini D. (1986), "Uno sguardo alle principali concezioni probabilistiche", Giuffrè, Milano.

Continua

Landenna G., Marasini D, Ferrari P. (1997), “Teoria della stima”, Il Mulino, Bologna.

Landenna G., Marasini D., Ferrari P. (1998), “La verifica di ipotesi statistiche”, Il Mulino, Bologna.

Lehmann E.L. (1933), The Fisher, Neyman-Pearson theories of testing hypoteses, Journal of American Statisticaòl Association, Vol. 88, n.424.

Muliere P., Capuzzo E., Statistica, Enciclopedia Italiana-Treccani

Neyman J. (1957) Inductive behavior as a basic concept in phylosophy of science, Revue de l’Institut International de Statistique.

Continua

Neyman J., Pearson S.E. (1932), The testing of statistical hypotheses in relation to probabilities a priori, Proceedings of the Cambridge Philosophy Society XXIX.

Popper K. (1959) The logic of scientific discovery, Routeridge, Taylor &Francis Group.

Sandrini M.G.(2009), Filosofia dei metodi induttivi, Firenze University Press.

Scardovi I., Statistica applicata alle scienze sociali, Enciclopedia delle Scienze Sociali.