lezione n. 6 (a cura di Gaia Montanucci)

METODO MASSIMA VEROSIMIGLIANZA PER STIMARE β_0 E β_1

Distribuzione sui termini di errore ε_i → $\varepsilon_i \sim N(0, \sigma^2)$ ne consegue :

ogni y_i ha ancora distribuzione normale, più precisamente $y_i \sim N(\mu_i, \sigma^2)$ con $\mu_i = \beta_0 + \beta_1 x_i$

Verosimiglianza: L = $\prod_i f(y_i) = \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y_1-\beta_0-\beta_1 x_i}{\sigma}\right)^2} = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2}\Sigma_i\left(\frac{y_1-\beta_0-\beta_1 x_i}{\sigma}\right)^2}$.

Log verosimiglianza: $l = \underbrace{-\frac{n}{2}\log(2\pi\sigma^2)}\ \underbrace{-\ \frac{1}{2}\sum_i}\left(\frac{y_1-\beta_0-\beta_1 x_i}{\sigma}\right)^2$

Parte indipendente da β_0 e β_1, la possiamo trascurare quando facciamo la massimizzazione

Parte dipendente da β_0 e β_1

$$\max_{\beta_1\beta_2} l \leftrightarrow \min_{\beta_1\beta_2} \sum_i (y_1 \quad \beta_0 \quad \beta_1 x_i)^2$$

$$\max_{\beta_1\beta_2} l \leftrightarrow \min_{\beta_1\beta_2} \sum_i S^2$$

CONCLUSIONE: Relativamente a β_0 e β_1, il metodo dei minimi quadrati porta esattamente alla stessa soluzione del metodo della massima verosimiglianza. Inoltre, massimizzare la verosimiglianza equivale a minimizzare S.

Per quanto riguarda la varianza il discorso è diverso:

metodo dei minimi quadrati → $S^2 = \frac{\Sigma_i y_i \widehat{-(y_i)}}{n-2}$

metodo della massima verosimiglianza → $\sigma^2 = \frac{\Sigma i\ (y_i - \widehat{y_i})^2}{n}$

Quindi in conclusione se stimo β_0 e β_1 non c'è differenza tra i due metodi, mentre se devo stimare σ^2 i due approcci portano a conseguenze diverse.

I due metodi sono basati su due idee diverse:

- ➔ Il metodo dei minimi quadrati è basato sulla minimizzazione dell'errore di previsione che è un concetto intuitivo e non richiede l'assunzione della distribuzione normale sulle variabili casuali
- ➔ Il metodo di massima verosimiglianza è più formale, richiede l'assunzione della distribuzione normale per i termini di errore

PROPRIETA' DEGLI STIMATORI:

1) β_0 e β_1 sono stimatori non distorti: $E(\widehat{\beta_0}) = \beta_0$, $E(\widehat{\beta_1}) = \beta_1$

Per stimatore non distorto s'intende che il valore atteso coincide con la quantità da stimare.

Anche S^2 è uno stimatore non distorto della varianza: $E(S^2) = \sigma^2$

Mentre $E(\sigma^{\prime 2}) \neq \sigma^{\prime 2}$ anche se distorsione piccola → σ^2 *(stimatore di massima verosimiglianza della varianza)* è distorto

2) Var ($(\beta_1 0)\hat{}$) = $\sigma^2\left(\frac{1}{n}+\frac{x^2}{dev(x)}\right)$

Var ($\widehat{\beta_1}$) = $\frac{\sigma^2}{dev(x)}$

3) Distribuzione degli stimatori : se è valida la proprietà che i termini di errore hanno distribuzione normale → $\varepsilon_i \sim N(0, \sigma^{\prime 2})$ allora anche β_0 e β_1 avranno distribuzione normale → $\widehat{\beta_0} \sim N\left(\beta_0, \mathbf{Var}\left(\widehat{\beta_0}\right)\right)$ $\widehat{\beta_1} \sim N(\beta_1 1$, Var ($\widehat{\beta_1}$))

Sotto questa condizione posso anche ricavare: $S^2\,\frac{(n-2)}{\sigma^2} \sim \chi^2(n-2)$

Questa proprietà mi serve per fare inferenza : per trovare gli intervalli di confidenza e la verifica delle ipotesi.

INTERVALLI DI CONFIDENZA β_0

Si distinguono due casi:

1) conosco σ^2 : $\widehat{\beta_0} - z_{\frac{\alpha}{2}}\sqrt{Var\left(\widehat{\beta_0}\right)}$ Estremo di dx, $\widehat{\beta_0} + z_{\frac{\alpha}{2}}\sqrt{Var\left(\widehat{\beta_0}\right)}$ Estremo di sx

2) non conosco σ^2: $\widehat{\beta_0} - t_{\frac{\alpha}{2}}\sqrt{\widehat{Var}\left(\widehat{\beta_0}\right)}$ Estremo di dx, $\widehat{\beta_0} + t_{\frac{\alpha}{2}}\sqrt{\widehat{Var}\left(\widehat{\beta_0}\right)}$ Estremo di sx

Quando utilizzo la t mi devo ricordare di utilizzare n - 2 gradi di libertà

$\widehat{Var}\left(\widehat{\beta_0}\right) = \quad s^2\left(\frac{1}{n}+\frac{\overline{x^2}}{dev(x)}\right)$

VERIFICA DELLE IPOTESI SU β_0

$H_0 : \beta_0 = \beta_{00}$ dove β_0 = intercetta e β_{00} = valore particolare dell'intercetta

Esempio → $H_0 : \beta_0 = 5$

$H_1 : \beta_0 \neq \beta_{00}$

Bisogna distinguere due casi :

1) conosco σ^2 : calcoliamo la statistica test → $z = \frac{\widehat{\beta_0} - \beta_{00}}{\sqrt{\widehat{Var}\left(\widehat{\beta_0}\right)}}$

La statistica test ci dice quanto il valore che stimo è diverso da quello ipotizzato → discrepanza

$|z| \geq z_{\frac{\alpha}{2}}$ → RH_0

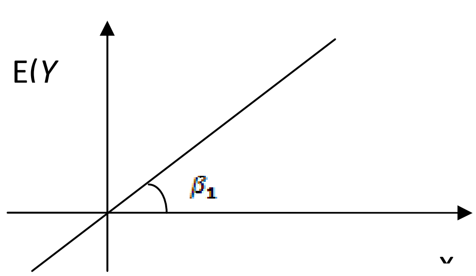
2) non conosco σ^2 : statistica test $t = \frac{\widehat{\beta_0} \quad \beta_{00}}{\sqrt{\widehat{Var}\left(\widehat{\beta_0}\right)}}$

$|t| \geq t_{\frac{\alpha}{2}}$ → RH_0

Caso standard : $H_0 : \beta_0 = 0$

$H_1 : \beta_0 \neq 0$

Questa è l'ipotesi più interessante perché $y_i = \beta_0 + \beta_1 x + \varepsilon_i$ → la *y* è PROPORZIONALE alla *x* se vale la ipotesi nulla e in questo caso abbiamo una retta passante per l'origine:



INTERVALLO DI CONFIDENZA PER β_1

1) conosco σ^2: $\quad \widehat{\beta_1} - z_{\frac{\alpha}{2}}\sqrt{Var(\widehat{\beta_1})} \qquad \widehat{\beta_1} + z_{\frac{\alpha}{2}}\sqrt{Var(\widehat{\beta_1})}$

2) non conosco σ^2: $\quad \widehat{\beta_1} - t_{\frac{\alpha}{2}}\sqrt{\widehat{Var}(\widehat{\beta_1})} \qquad \widehat{\beta_1} + t_{\frac{\alpha}{2}}\sqrt{\widehat{Var}(\widehat{\beta_1})}$

$\widehat{Var}(\widehat{\beta_1}) = \frac{S^2}{Dev(x)}$

VERIFICA DELLE IPOTESI β_1

$H_0 : \beta_1 = \beta_{10}$ dove β_1 = coefficiente angolare β_{10} = preciso valore che deve assumere β_1

Esempio → $H_0 : \beta_1 = -1$

$H_1 : \beta_1 \neq \beta_{10}$

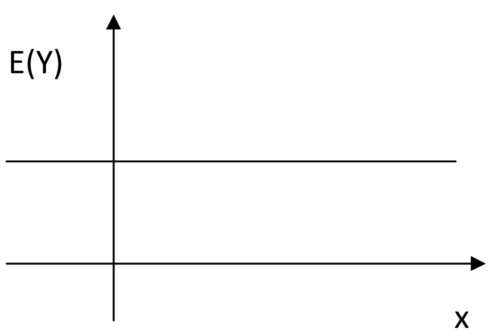
1) conosco σ^2: $\quad z = \frac{\widehat{\beta_1} - \beta_{10}}{\sqrt{Var(\widehat{\beta_1})}} \qquad |z| \geq z_{\frac{\alpha}{2}}$ → RH_0

2) non conosco σ^2: $\quad t = \frac{\widehat{\beta_1} - \beta_{10}}{\sqrt{\widehat{Var}(\widehat{\beta_1})}} \qquad |t| \geq t_{\frac{\alpha}{2}}$ → RH_0

Caso standard : $H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

Se vale l'ipotesi nulla → $y_i = \beta_0 + \beta_1 x + \varepsilon_i$ e ciò equivale a dire che *y* non dipende linearmente dalla *x*



Qualsiasi valore di *x* che prendo ottengo sempre lo stesso valore atteso

$t = \dfrac{\widehat{\beta_1}}{\sqrt{\widehat{Var}(\widehat{\beta_1})}}$ dove $\sqrt{\widehat{Var}(\widehat{\beta_1})}$ è una misura di errore associata alla stima, più grande è tale valore meno precisa è la stima

OUTPUT DI R → METODO DI REGRESSIONE SEMPLICE

est	s.e.	t - stat	p - value
$\widehat{\beta_0}$	$\sqrt{\widehat{Var}(\widehat{\beta_0})}$	$\dfrac{\widehat{\beta_0}}{\sqrt{\widehat{Var}(\widehat{\beta_0})}}$	p – value (β_0)
$\widehat{\beta_1}$	$\sqrt{\widehat{Var}(\widehat{\beta_1})}$	$\dfrac{\widehat{\beta_1}}{\sqrt{\widehat{Var}(\widehat{\beta_1})}}$	p – value (β_1)

La prima colonna ci permette di capire come la *x* influisce sulla *y.*

$\widehat{\beta_0}$ = valore atteso della *y* quando *x* = 0

$\widehat{\beta_1}$ = incremento del valore atteso quando

Ci dice se la soluzione è vera o è dovuta al caso.

I valori per noi più significativi sono quindi *est* e *p – value*. Facciamo un esempio:

est	p - value
10	0.03
3	0.15

Con 0.03 → $RH_0 : \beta_0 = 0$ quindi l'intercetta è significativamente diversa da 0 e ne consegue che 10 non è dovuto al caso ma è un valore plausibile.

Con 0.15 → $AH_0 : \beta_1 = 0$ quindi il coefficiente angolare non è significativamente diverso da 0 e ne consegue che non posso dire che 3 sia dovuto al caso

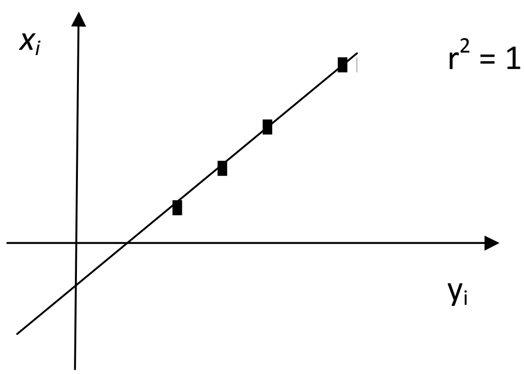
r^2 = INDICE DI DETERMINAZIONE e serve a valutare la bontà dell'adattamento del modello e si calcola : $r^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} = 1 - \frac{DR}{Dev(y)}$ dove DR = devianza residua

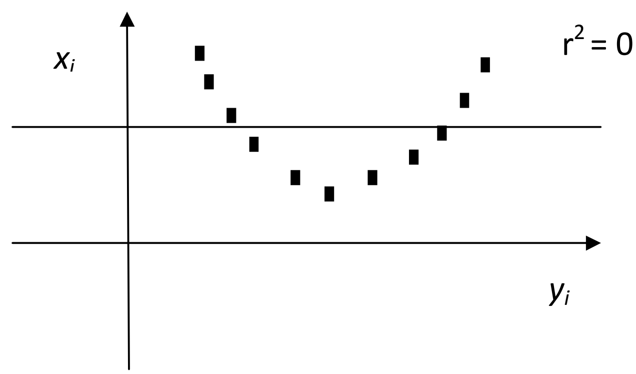
r^2 può assumere i seguenti valori: $0 \leq r^2 \leq 1$

se $r^2 = 0$ vuol dire che la devianza residua è uguale alla devianza di y → ho un ADATTAMENTO PESSIMO: vuol dire che il modello è inutile a livello di previsioni (situazione peggiore)

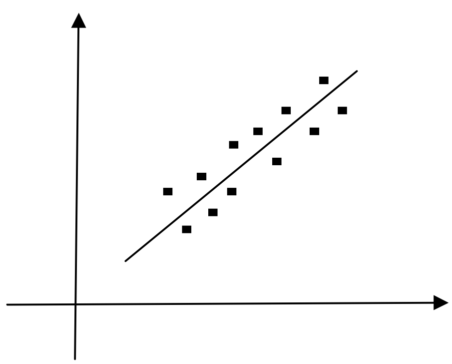
se $r^2 = 1$ vuol dire che la devianza residua è uguale a 0 e quindi ogni valore osservato è uguale a quello previsto → DR =0 $\leftrightarrow y_i = \hat{y}_i$ e abbiamo un ADATTAMENTO PERFETTO: riesco a prevedere ogni situazione

Esempio: $r^2 \geq 0.75$ il modello ha una buona capacità previsiva





I casi che troveremo saranno, usualmente, quelli intermedi:



$r^2 = 0.8$