Facoltà di Ingegneria - Università di Bologna

Anno Accademico: 2010/11

TECNICA ED ECONOMIA DEI TRASPORTI

Docente: Marino Lupi

IL CALCOLO DELLA DOMANDA DI TRASPORTO

(Capitolo 4)

Modello *descrittivo* per il calcolo della domanda di trasporto aereo fra due città

$$\ln d_{ij} = \beta_1 + \beta_2 \ln(P_i P_j) + \beta_3 \ln F_{ij}$$

Faccio il seguente esperimento:

- Introduco nel membro di destra i valori delle variabili esplicative osservate in un certo anno, le indico con: $\ln\left(\hat{P}_{i,} \hat{P}_j\right), \ln \hat{F}_{ij},$
- Introduco nel membro di sinistra il valore $\ln \hat{d}_{ij}$ della domanda verificatosi in quell'anno

$$\ln \hat{d}_{ij} - \left[\beta_1 + \beta_2 \ln\left(\hat{P}_i \hat{P}_j\right) + \beta_3 \ln \hat{F}_{ij}\right] = e_{ij}$$

Scarto aleatorio dovuto agli attributi trascurati e semplificazioni introdotte nel modello

$$\ln \hat{d}_{ij} = \beta_1 + \beta_2 \ln\left(\hat{P}_i \hat{P}_j\right) + \beta_3 \ln \hat{F}_{ij} + e_{ij}$$

$$y_c = \beta_1 x_{c1} \quad + \quad \beta_2 x_{c2} + \quad \beta_3 x_{c3} + e_c$$

$$x_{c1} = 1$$

In generale ho T equazioni:

$$\begin{cases} y_1 = \beta_1 x_{11} + \dots\dots\dots\dots + \beta_K x_{1K} + e_1 \\ y_2 = \beta_1 x_{21} + \dots\dots\dots\dots + \beta_K x_{2K} + e_2 \\ \dots\dots \quad \dots\dots \quad \dots\dots \quad \dots\dots\dots \quad \dots\dots\dots \\ y_c = \beta_1 x_{c1} + \dots\dots\dots\dots + \beta_K x_{cK} + e_c \\ \dots\dots \quad \dots\dots \quad \dots\dots \quad \dots\dots\dots \\ y_T = \beta_1 x_{T1} + \dots\dots\dots\dots + \beta_K x_{TK} + e_T \end{cases}$$

In forma compatta scrivo: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$

$\mathbf{X}(T \times K)$: matrice delle variabili esplicative od indipendenti, gli elementi di $\mathbf{X}$ sono quantità *deterministiche*.

$\mathbf{y}(T \times 1)$: è il vettore *aleatorio* dei dati campionari.

$\boldsymbol{\beta}(K \times 1)$: è il vettore dei coefficienti (parametri) del modello (sono quantità *deterministiche*).

$\mathbf{e}(T \times 1)$: è il vettore *aleatorio* degli “errori”.

In fase di calibrazione del modello il vettore $\mathbf{y}$ e la matrice $\mathbf{X}$ sono quantità *note*.

Le quantità *non note* sono:

$\boldsymbol{\beta}$: vettore dei coefficienti da stimare (quantità deterministiche).

$\mathbf{e}$: vettore di variabili aleatorie non osservabili.

Ipotesi base del modello di regressione lineare $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$

1° ipotesi

$$E(e_c) = 0 \quad \Leftrightarrow \quad E(y_c) = \beta_1 x_{c1} + \beta_2 x_{c2} + \ldots\ldots + \beta_K x_{cK} \qquad \forall c$$

2° ipotesi

$$\mathrm{var}(e_c) = \mathrm{var}(y_c) = \sigma^2$$

In base a questa ipotesi il modello si dice *omoscedastico*

3° ipotesi

Le variabili y_c, come le variabili e_c, sono indipendenti

$$\mathrm{cov}(y_c, y_{c'}) = 0 \forall c, c' \qquad \mathrm{cov}(e_c, e_{c'}) = 0 \forall c, c'$$

4° ipotesi La matrice $\mathbf{X}$ (TxK) è composta da variabili *deterministiche* ed ha *pieno rango colonna.*

Queste sono le ipotesi fondamentali del *modello di regressione lineare omoscedastico:* l'estimatore del vettore dei coefficienti (ottenuto con il metodo dei minimi quadrati di cui parleremo) è detto *estimatore dei minimi quadrati ordinario* ("*Ordinary Least Squared Estimator", OLSE*)

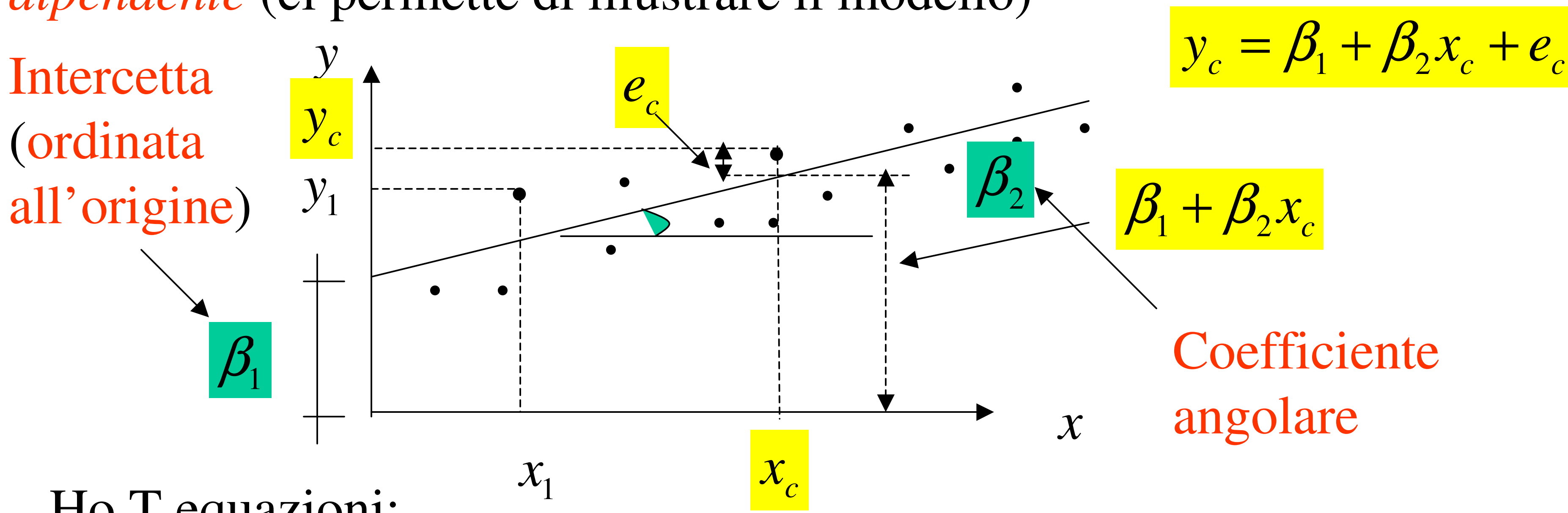
5° ipotesi (questa ipotesi *non è necessaria* in sede di *calibrazione*, lo sarà invece nella fase di corroborazione).

Per la variabile aleatoria e_c, "*errore*", *essendo l'effetto risultante della somma di numerosi fattori non specificati nel modello,* ossia nelle variabili indipendenti, si fa l'ipotesi, basandosi sul *teorema del limite centrale,* che sia distribuito secondo una variabile aleatoria normale.

Teorema del limite centrale: se Z è la somma di n variabili aleatorie, distribuite in modo qualsiasi e fra loro indipendenti , tali che la varianza di ciascuna di esse è trascurabile rispetto alla varianza di Z, al crescere di n la legge di probabilità di Z tende a quella di una variabile aleatoria normale avente media e varianza pari a quella di Z.

Se le variabili aleatorie e_c sono distribuite in modo normale anche le y_c sono variabili aleatorie *normali* . Le variabili e_c e y_c differiscono solo per la media, per il resto sono del tutto uguali.

Caso di un'*unica* variabile *indipendente* e di un'*unica* variabile *dipendente* (ci permette di illustrare il modello)



Ho T equazioni:

$$\begin{cases} y_1 = \beta_1 + \beta_2 x_1 + e_1 \\ y_2 = \beta_1 + \beta_2 x_2 + e_2 \\ \ldots\ldots \quad \ldots\ldots \quad \ldots\ldots \\ y_c = \beta_1 + \beta_2 x_c + e_c \\ \ldots\ldots \quad \ldots\ldots \quad \ldots\ldots \\ y_T = \beta_1 + \beta_2 x_T + e_T \end{cases} \Rightarrow \underbrace{\mathbf{y}}_{T\times 1} = \underbrace{\mathbf{X}}_{T\times 2}\underbrace{\boldsymbol{\beta}}_{2\times 1} + \underbrace{\mathbf{e}}_{T\times 1}$$

La prima colonna della matrice $\mathbf{X}$ è costituita da tutti 1.

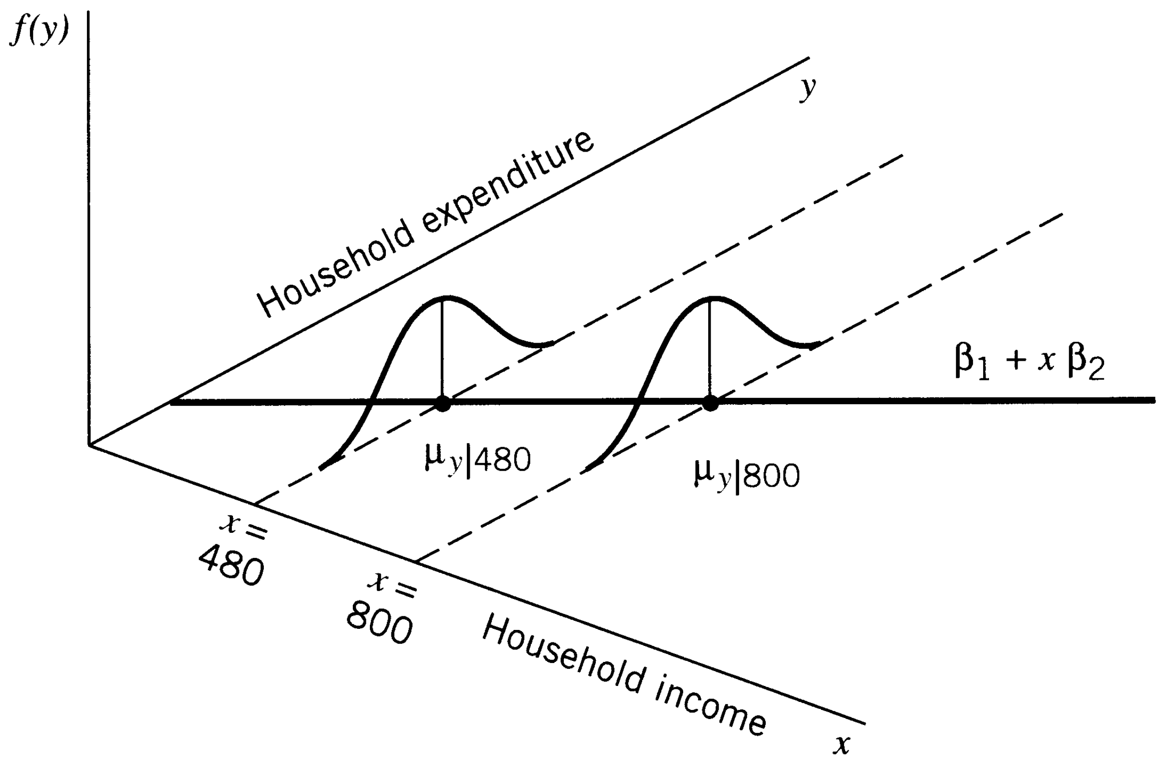
Gli economisti: “Houseold expenditure”

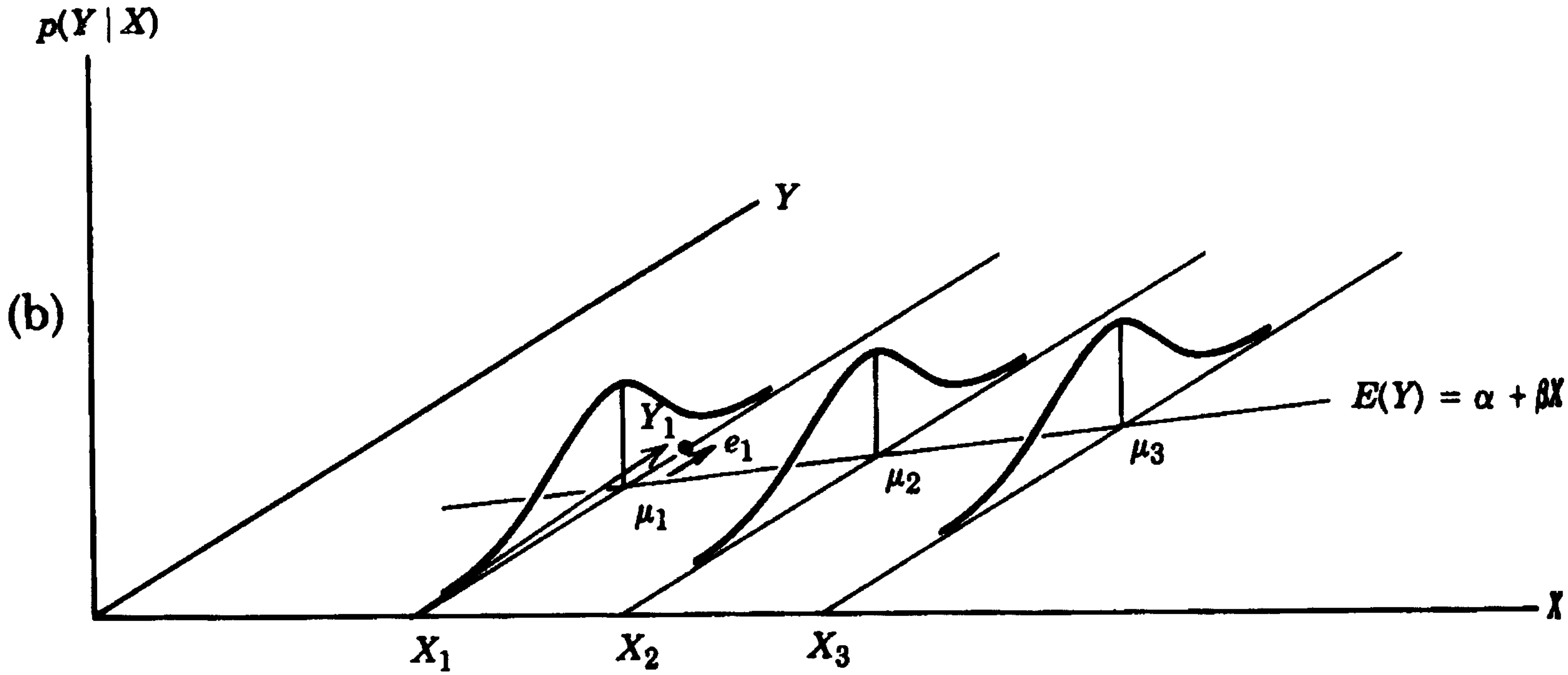
Modello *economico*: $HE = \beta_1 + \beta_2 HI$

“Houseold income”

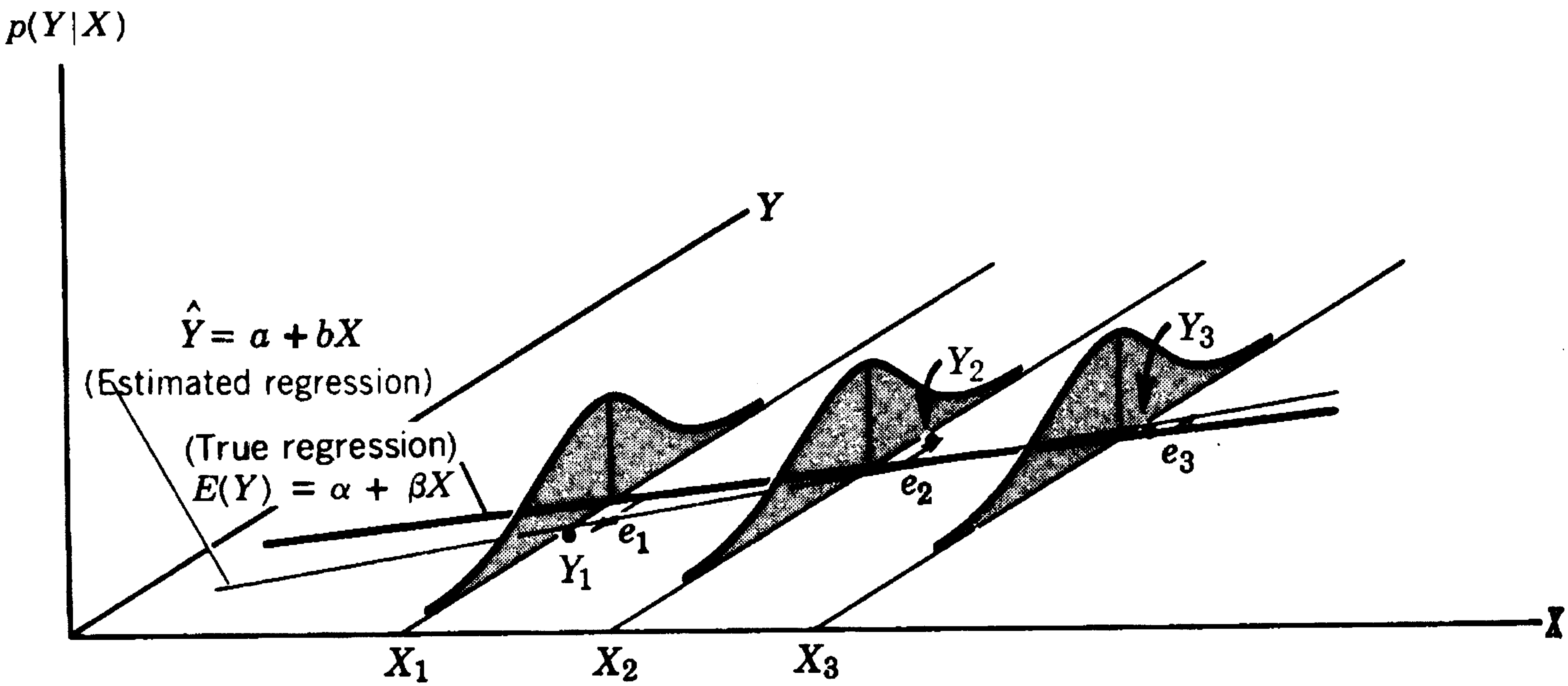
Modello *econometrico*: $HE = \beta_1 + \beta_2 HI + e$

errore del modello





$p(Y \mid X)$
(b)
Y
Y_1
e_1
$E(Y) = \alpha + \beta X$
μ_1
μ_2
μ_3
X
X_1
X_2
X_3



True (population) regression and estimated (sample) regression.

Modello *trasportistico*:

$$\ln d_{ij} = \beta_1 + \beta_2 \ln\left(P_i P_j\right) + \beta_3 \ln F_{ij}$$

Modello *trasportometrico*:

$$\ln d_{ij} = \beta_1 + \beta_2 \ln\left(P_i P_j\right) + \beta_3 \ln F_{ij} + e_{ij}$$

$$y_c = \beta_1 \underbrace{x_{c1}}_{=1} + \beta_2 x_{c2} + \beta_3 x_{c3} + e_c$$

Questa dizione però di fatto non si usa: forse è meglio parlare di *modello trasportistico inquadrato in termini statistici.*

Normalmente nei modelli utilizzati nel campo dei trasporti, come in molti modelli economici, l'esperimento che dà luogo alla variabile *e*, "errore", *non è ripetibile*: ossia non posso osservare più y per degli stessi valori di x . Ma poiché le variabili indipendenti spiegano solo in parte la variabile dipendente, io so che se potessi ripetere l'esperimento, con le stesse x, otterrei una y differente. Perciò modellizzo il fenomeno come se potessi ripetere l'esperimento: ossia considero e ed y come variabili aleatorie.

Indico con **b** la stima del vettore **β** .

Determino quel vettore **b** tale che gli scarti fra quello che dà il modello ed i dati sperimentali sono minimi:

$$Min\sum_{i=1}^{T}\left[y_i-(b_1+b_2x_i)\right]^2$$

Scelgo **b** tale che sia *minima* la *sommatoria del quadrato degli scarti* fra quello che dà il modello ed il dato sperimentale (considero il quadrato per evitare che i residui positivi e negativi si compensino).

Nel caso multivariato:

$$Min \quad \sum_{i=1}^{T} \left[y_i - (b_1 x_{i1} + b_2 x_{i2} + \quad .... + b_K x_{iK}) \right]^2$$

In termini *vettoriali compatti*:

$$Min \underset{(1\times T)}{\left(\mathbf{y} - \mathbf{Xb}\right)'} \underset{(T\times 1)}{\left(\mathbf{y} - \mathbf{Xb}\right)}$$

$$\underset{T\times 1}{\mathbf{y}} \; - \; \underset{T\times K}{\mathbf{X}} \; \underset{K\times 1}{\mathbf{b}}$$

$$S_Q = (\mathbf{y}-\mathbf{X}\mathbf{b})'(\mathbf{y}-\mathbf{X}\mathbf{b}) = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{b} - \mathbf{b}'\mathbf{X}'\mathbf{y} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} =$$

$$= \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\mathbf{b} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}$$

$\underset{1\times K}{\mathbf{b}'}\ \underset{K\times T}{\mathbf{X}'}\ \underset{T\times 1}{\mathbf{y}} \longleftarrow$ E’ uno scalare: perciò coincide con il suo trasposto

$$\mathbf{b}'\mathbf{X}'\mathbf{y} = (\mathbf{b}'\mathbf{X}'\mathbf{y})' = \mathbf{y}'\mathbf{X}\mathbf{b}$$

Calcolo il vettore gradiente (considerato come vettore colonna) di S_Q e lo pongo uguale a 0: in questo modo ottengo il vettore $\mathbf{b}$ che minimizza gli scarti.

$$\nabla S_Q(\mathbf{b}) = -2\mathbf{X}'\mathbf{y} + 2(\mathbf{X}'\mathbf{X})\mathbf{b} = 0 \quad \Rightarrow \quad 2(\mathbf{X}'\mathbf{X})\mathbf{b} = 2\mathbf{X}'\mathbf{y} \quad \Rightarrow$$

$\mathbf{X}$ ha pieno rango (4° ipotesi del modello) $\Rightarrow$ $(\mathbf{X}'\mathbf{X})$ è invertibile

$\Rightarrow$ $\boxed{\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}}$ $\mathbf{b}$ *è l ‘estimatore dei minimi quadrati.*

$$\underset{K\times 1}{\mathbf{b}} = \underbrace{\left(\underset{k\times T}{\mathbf{X}'}\ \underset{T\times K}{\mathbf{X}}\right)}_{K\times K}^{-1} \underset{K\times T}{\mathbf{X}'}\ \underset{T\times 1}{\mathbf{y}}$$

b è un vettore aleatorio perché è funzione di **y** che è un vettore aleatorio

Proprietà dell'estimatore dei minimi quadrati

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

$$E(\mathbf{b}) = E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}\right] = E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}\right]$$

$$E\left[\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}\right] = \boldsymbol{\beta} + E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}\right] = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{e}] = \boldsymbol{\beta}$$

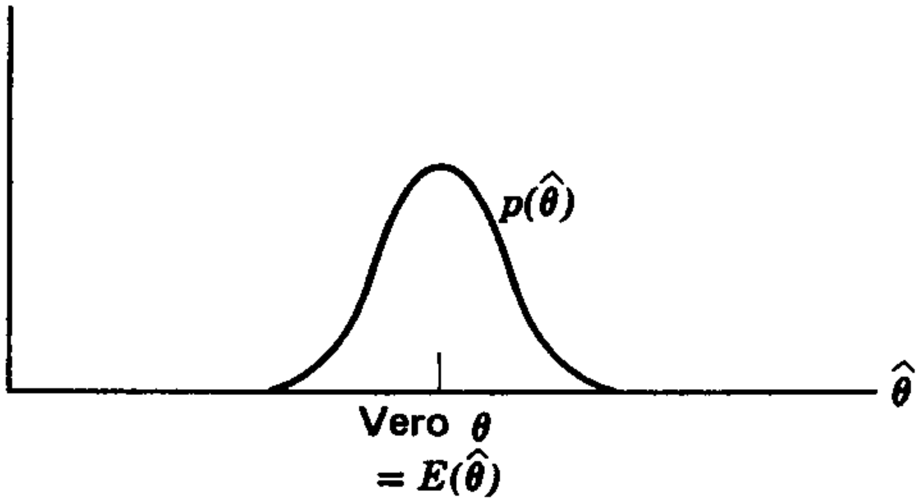
E è un operatore lineare

$= 0$

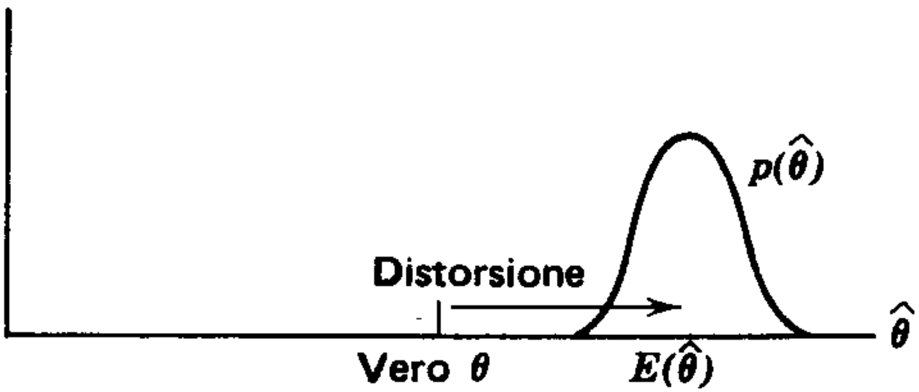
(*1° ipotesi del modello*)

$E[\mathbf{b}] = \boldsymbol{\beta}$

L'estimatore dei minimi quadrati è un *estimatore corretto*



$\hat{\theta}$ è un estimatore *corretto* di θ



$\hat{\theta}$ è un estimatore *distorto* di θ

b è un vettore, invece di parlare di varianza parlo di *matrice di covarianza* (detta anche *matrice di varianza-covarianza*)

Dato un qualsiasi vettore aleatorio $\mathbf{z}\ (k \times 1)$

$$\mathbf{V}[\mathbf{z}] \rightarrow \begin{cases} V[\mathbf{z}]_{ij} = \mathrm{cov}[z_i z_j] & [i \neq j] \\ V[\mathbf{z}]_{ii} = \mathrm{var}[z_i] & [i = j] \end{cases}$$

$$\mathbf{V}[\mathbf{z}] = \begin{bmatrix} \mathrm{var}(z_1) & \mathrm{cov}(z_1 z_2) \ldots\ldots\ldots & \mathrm{cov}(z_1 z_K) \\ \mathrm{cov}(z_2 z_1) & \mathrm{var}(z_2) \ldots\ldots\ldots & \mathrm{cov}(z_2 z_K) \\ \ldots\ldots\ldots & \ldots\ldots\ldots & \ldots\ldots\ldots \\ \ldots\ldots\ldots & \ldots\ldots\ldots & \ldots\ldots\ldots \\ \mathrm{cov}(z_K z_1) & \mathrm{cov}(z_K z_2) \ldots\ldots\ldots & \mathrm{var}(z_K) \end{bmatrix}$$

$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ Modello di regressione lineare

Abbiamo fatto l'ipotesi:

$\mathrm{var}(e_c) = \mathrm{var}(y_c) = \sigma^2$ Modello di regressione lineare *omoscedastico*

Ed inoltre che:

$\mathrm{cov}(y_c, y_{c'}) = 0 \ \forall c, c'$ $\mathrm{cov}(e_c, e_{c'}) = 0 \ \forall c, c'$

La matrice di covarianza è la stessa per $\mathbf{y}$ *e* $\mathbf{e}$ ed è data da:

$$\underset{(T\times T)}{\mathbf{V}[\mathbf{y}]} = \underset{(T\times T)}{\mathbf{V}[\mathbf{e}]} = \begin{bmatrix} \sigma^2 & 0 \ldots\ldots\ldots\ldots\ldots\ldots 0 \\ 0 & \sigma^2 \ldots\ldots\ldots\ldots\ldots 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \sigma^2 & \end{bmatrix} = \sigma^2 \mathbf{I}_T$$

$\mathbf{I}_T$ è la matrice unitaria di dimensione $T \times T$

$$\mathbf{I}_T = \begin{bmatrix} 1 & 0 \ldots\ldots\ldots\ldots\ldots 0 \\ 0 & 1 \ldots\ldots\ldots\ldots\ldots 0 \\ 0 \ldots\ldots 1 \ldots\ldots\ldots\ldots 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 1 \end{bmatrix}$$

1 sulla diagonale principale il resto degli elementi sono tutti 0.

$$\mathbf{V}[\mathbf{b}] = \sigma^2 \Big(\underbrace{\underset{K\times T}{\mathbf{X}'} \underset{T\times K}{\mathbf{X}}}_{K\times K} \Big)^{-1}$$

Matrice di covarianza dell'estimatore dei minimi quadrati **b**

(dimostrazione sulle dispense)

Riassumiamo le proprietà dell'estimatore dei minimi quadrati $\mathbf{b}$

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

E' un estimatore *lineare*: ossia ottenuto con una trasformazione lineare del vettore aleatorio $\mathbf{y}$ ($\mathbf{b} = \mathbf{A}\mathbf{y}$ dove $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$)

$$E[\mathbf{b}] = \boldsymbol{\beta}$$

E' un estimatore *corretto*: la media coincide con il vettore dei parametri da stimare $\boldsymbol{\beta}$.

$$\mathbf{V}[\mathbf{b}] = \sigma^2 \, (\mathbf{X}'\mathbf{X})^{-1}$$

La *matrice di covarianza* ha questa espressione

- Teorema di Gauss-Markov

L'estimatore dei minimi quadrati $\boldsymbol{\beta}$ è il *migliore* estimatore *lineare corretto.*

- Classe degli estimatori lineari, ossia del tipo:

$$\hat{\boldsymbol{\beta}} = \mathbf{A}\mathbf{y}$$

(Nel caso dell'estimatore dei minimi quadrati ordinario $\mathbf{b}$: $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$)

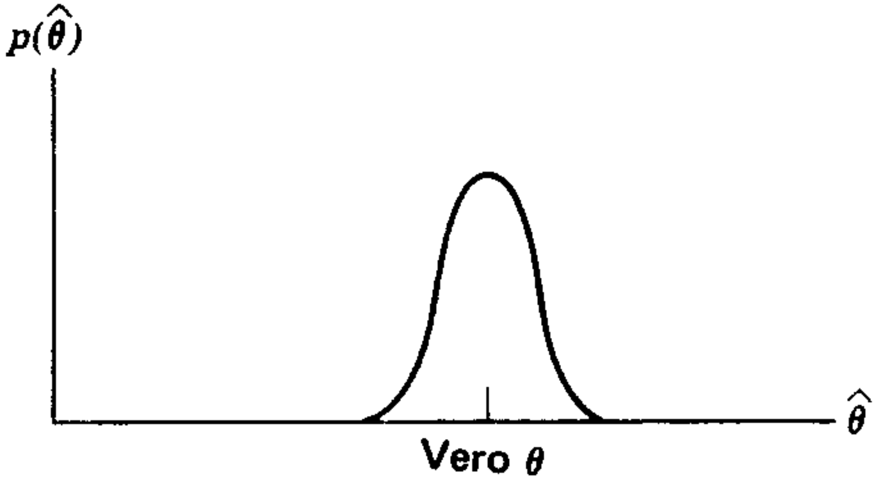
- Classe degli estimatori corretti:

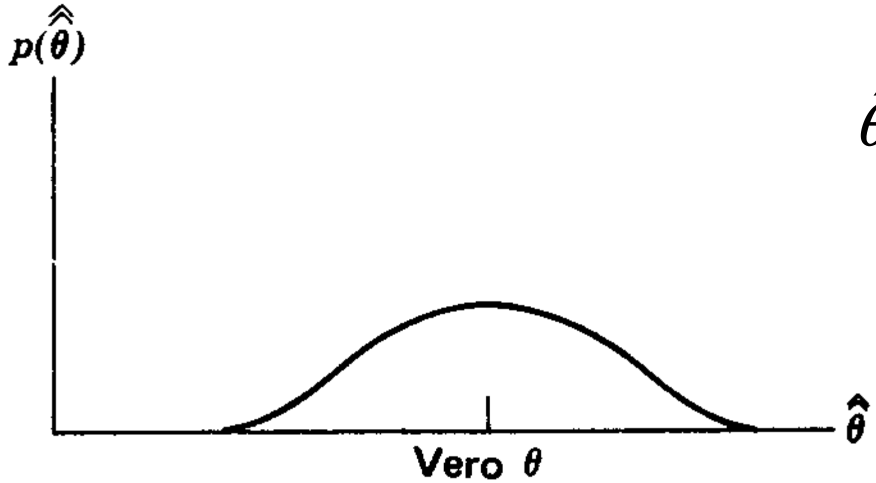
$$E\left[\hat{\boldsymbol{\beta}}\right] = \boldsymbol{\beta}$$

Nell'ambito della classe degli estimatori lineari corretti $\mathbf{b}$ è il *migliore* nel senso che:

$$\boxed{\operatorname{var}(b_i) \leq \operatorname{var}(\hat{\beta}_i) \qquad \forall i}$$

Ogni altro estimatore lineare corretto ha una varianza maggiore, al più uguale, a quella dell'estimatore dei minimi quadrati, ma mai minore.





$\hat{\theta}$ è un estimatore più *efficiente di* $\hat{\hat{\theta}}$ ($\hat{\theta}$ e $\hat{\hat{\theta}}$ sono entrambi estimatori corretti di θ)

Il teorema di Gauss- Markov (risultato teorico rigorosamente dimostrato) insieme alla ragionevolezza dell'approccio, minimizzo il quadrato degli scarti fra modello e dati sperimentali, oltre al collaudato utilizzo, spiegano il *diffuso utilizzo* dell'*estimatore dei minimi quadrati*.

Come posso stimare σ^2 ?

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{e}$$

$\mathrm{var}(e_i) = \sigma^2 \quad \forall i \quad \sigma^2$ è la varianza sia degli errori e_i, sia di y_i

$$\underbrace{\mathrm{var}(e_i)}_{\text{Per definizione di varianza}} = E\left[\left(e_i - \underbrace{E[e_i]}_{=0\ (1° \text{ ipotesi del modello di regressione lineare})}\right)^2\right] = E\left[e_i^{\,2}\right]$$

Per definizione di varianza

=0 (1° ipotesi del modello di regressione lineare)

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$$

$$\downarrow$$

$$\hat{\mathbf{e}} = \mathbf{y} - \mathbf{X}\mathbf{b}$$

Stimo il vettore degli errori $\mathbf{e}$ con il vettore dei residui $\hat{\mathbf{e}}$, ottenuti inserendo al posto di $\boldsymbol{\beta}$ la sua stima $\mathbf{b}$.

Essendo:

$$\sigma^2 = \text{var}(e_i) = E\left[e_i^{\ 2}\right] \quad \forall i$$

Potrei stimare la σ^2 con la $\hat{\sigma}^2$

$$\hat{\sigma}^2 = \frac{\hat{e}_1^2 + \hat{e}_2^2 + \ldots\ldots + \hat{e}_T^2}{T}$$

Media aritmetica dei residui

Si dimostra però che questo estimatore non è corretto, ossia:

$$E\left[\hat{\sigma}^2\right] = E\left[\frac{\hat{e}_1^2 + \hat{e}_2^2 + \ldots\ldots + \hat{e}_T^2}{T}\right] \neq \sigma^2 \longleftarrow$$

Media dell'estimatore diversa dal parametro da stimare

L’estimatore :

$$\hat{\hat{\sigma}}^2 = \frac{\hat{e}_1^2 + \hat{e}_2^2 + ...+ \hat{e}_T^2}{T-K} = \frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T-K}$$

è invece un estimatore corretto:

$$E\left[\hat{\hat{\sigma}}^2\right] = E\left[\frac{\hat{e}_1^2 + \hat{e}_2^2 + ...+ \hat{e}_T^2}{T-K}\right] = \sigma^2$$

Quindi per stimare σ^2 si utilizza l’estimatore $\hat{\hat{\sigma}}^2$

Le fasi di messa a punto di un modello di domanda sono tre:

- *Specificazione*
- *Calibrazione*
- *Corroborazione*

Fase di *corroborazione* nel caso di un modello di regressione lineare $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$

Una volta stimato il vettore dei parametri del modello $\boldsymbol{\beta}$ con il vettore dei minimi quadrati $\mathbf{b}$:

Calcolo il *coefficiente di determinazione* R^2

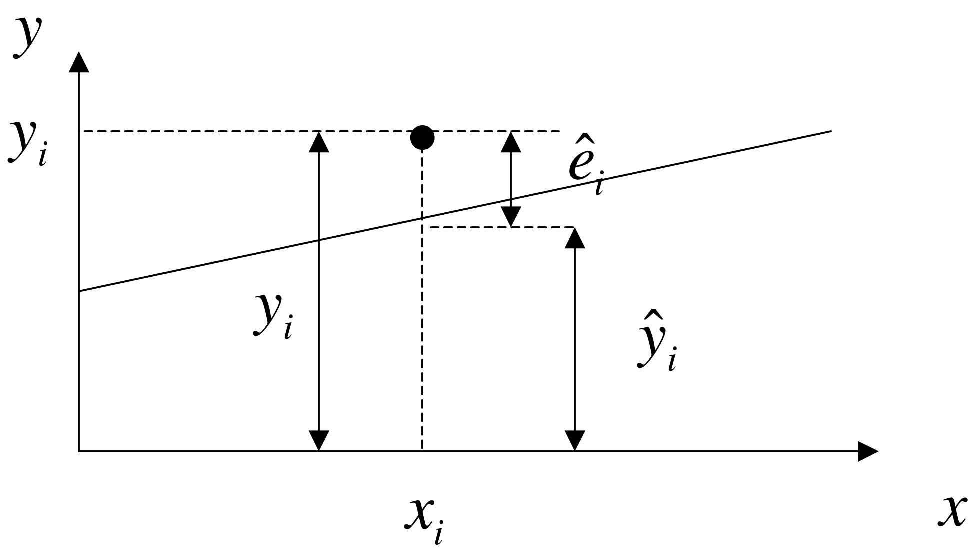
$$\sum_{i=1}^{T}(y_i - \overline{y})^2$$ Misura della variazione di $\mathbf{y}$

($\overline{y} = \dfrac{\sum_i y_i}{T}$ è la media degli y_i)

$\hat{\mathbf{y}} = \mathbf{Xb}$ estimatore di $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$

$$\sum_{i=1}^{T}(\hat{y}_i - \overline{\hat{y}})^2$$ Misura della variazione di $\hat{\mathbf{y}}$

($\overline{\hat{y}} = \dfrac{\sum_i \hat{y}_i}{T}$ è la media degli $\hat{y}_i$)



Se il modello contiene l'intercetta, ossia la prima colonna di X è costituita da tutti 1, risulta: $\overline{\hat{y}} = \overline{y}$. Inoltre si dimostra che:

$$\sum_{i=1}^{T}(y_i - \overline{y})^2 = \sum_{i=1}^{T}(\hat{y}_i - \overline{y})^2 + \hat{\mathbf{e}}'\hat{\mathbf{e}} \quad \Rightarrow \quad \sum_{i=1}^{T}(\hat{y}_i - \overline{y})^2 = \sum_{i=1}^{T}(y_i - \overline{y})^2 - \hat{\mathbf{e}}'\hat{\mathbf{e}}$$

Variazione *totale* = variazione *spiegata dal modello* + variazione *non spiegata dal modello*

$$R^2 = \frac{\text{variazione spiegata dal modello}}{\text{variazione totale}}$$

Coefficiente di *determinazione*

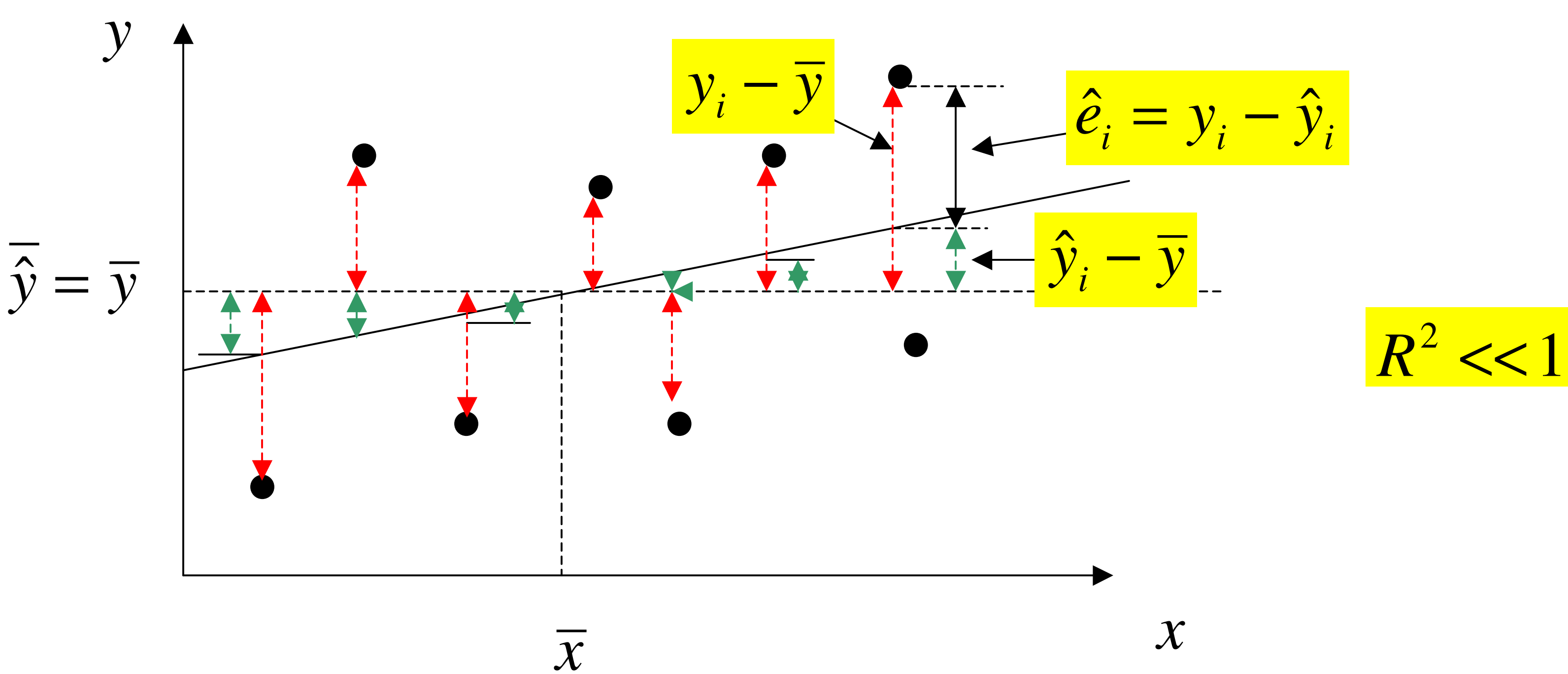
$$R^2 = \frac{\sum_{i=1}^{T}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{T}\left(y_i - \bar{y}\right)^2} = 1 - \frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{\sum_{i=1}^{T}\left(y_i - \bar{y}\right)^2}$$

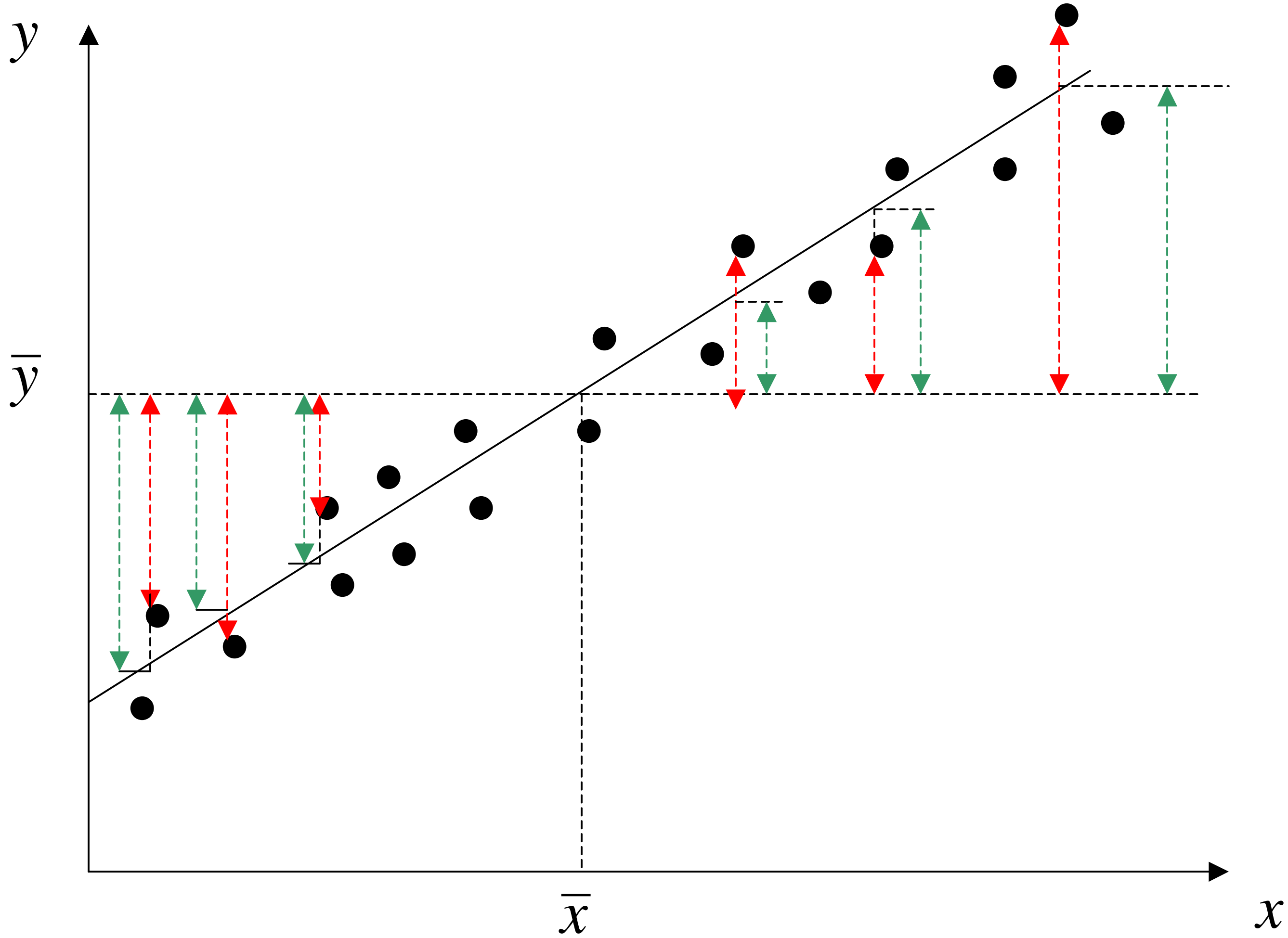
$$1 - \frac{\text{variazione non spiegata dal modello}}{\text{variazione totale}}$$

$$0 \le R^2 \le 1$$

Se $R^2 \cong 0$ Il modello non funziona perché la variazione spiegata dal modello è bassa rispetto alla variazione totale.

Se $R^2 \cong 1$ Il modello (da questo punto di vista) va bene perché la variazione spiegata dal modello è circa uguale alla variazione totale.





$R^2 \rightarrow 1$

Esempio- Domanda di trasporto aereo fra due città

$$\ln d_{ij} = \left[\beta_1 + \beta_2 \ln\left(P_i P_j\right) + \beta_3 \ln T_{ij}\right] + e_{ij}$$

Dopo calibrazione considerando 35 coppie di città (dati campionari, non serie storica)

$$\ln d_{ij} = -4{,}568 + 0{,}925 \ln\left(P_i P_j\right) - 0{,}767 \ln T_{ij}$$

$$R^2 = 0{,}774$$

Nel modello (macromodello) che avevamo visto a proposito del traffico di veicoli pesanti in autostrada:

$$\ln d_t^{vp} = 3{,}62 + 1{,}99 \ln IPI_t - 0{,}76 T_t$$

$$R^2 = 0{,}97$$

Il coefficiente di determinazione R^2 è un *indice* di corroborazione *globale* del modello.

Serve però anche un *indice puntuale*: che ci dica se ciascuna variabile indipendente (attributo) "spiega"(e in quale misura) la variabile dipendente (domanda).

$\mathbf{b} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$ è un vettore normale multivariato (questo è vero se suppongo che $\mathbf{y}$ ed $\mathbf{e}$ siano distribuiti in modo normale multivariato: 5° ipotesi del modello di regressione lineare).

$$\mathbf{b} \sim (\boldsymbol{\beta}, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1})$$

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{b}$$

$$\hat{\mathbf{y}} = b_1\mathbf{x}_1 + b_2\mathbf{x}_2 + \dots\dots\dots\dots + b_K\mathbf{x}_K$$

Seconda colonna della matrice $\mathbf{X}$

Es. $\mathbf{x}_2 = \begin{bmatrix} \ln PIL_1 \\ \ln PIL_2 \\ \dots\dots\dots \\ \ln PIL_T \end{bmatrix}$

$$\mathbf{b} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X'X})^{-1})$$ ← Vettore normale multivariato

$$b_j \sim N(\beta_j, \sigma^2(\mathbf{X'X})_{jj}^{-1})$$

$\begin{cases} \text{Elemento } jj \text{ della matrice} \\ \sigma^2(\mathbf{X'X})^{-1} \end{cases}$

La generica componente j-esima, b_j, di $\mathbf{b}$ è una variabile normale

$$\frac{b_j - \beta_j}{\sqrt{\sigma^2(\mathbf{X'X})_{jj}^{-1}}}$$ ← E' una variabile aleatoria normale standard

Ossia:

$$\frac{b_j - \beta_j}{\sqrt{\sigma^2(\mathbf{X'X})_{jj}^{-1}}} \sim N(0,1)$$

σ^2 non lo conosco, ma lo posso stimare

σ^2 lo stimo con $\hat{\sigma}^2 = \dfrac{\hat{e}_1^2 + \hat{e}_2^2 + ... + \hat{e}_T^2}{T-K} = \dfrac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T-K}$

dove: $\hat{\mathbf{e}} = \mathbf{y} - \mathbf{X}\mathbf{b}$ $\longleftarrow$ vettore dei residui $\hat{\mathbf{e}}$, ottenuti inserendo al posto di $\boldsymbol{\beta}$ la sua stima $\mathbf{b}$.

Si dimostra che se sostituisco σ^2 con $\hat{\sigma}^2$ ottengo una variabile

t di Student con T-K gradi di libertà

Variabile normale standard

$$\downarrow$$

$$\frac{b_j - \beta_j}{\sqrt{\sigma^2 (\mathbf{X}'\mathbf{X})^{-1}_{jj}}}$$

Variabile t di Student con T-K gradi di libertà

$$\downarrow$$

$$\frac{b_j - \beta_j}{\sqrt{\hat{\sigma}^2 \, (\mathbf{X}'\mathbf{X})^{-1}_{jj}}}$$

$$\frac{b_j - \beta_j}{\sqrt{\hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}_{jj}}} = \frac{b_j - \beta_j}{s_j}$$

è una variabile t di student con T-K gradi di libertà (una variabile t di student, con *DF* gradi di libertà, è una variabile aleatoria, a media nulla, che "assomiglia" tanto più ad una normale standard quanto più è alto il numero di gradi di libertà)

(s_j è la stima della deviazione standard di b_j)

Test di ipotesi sui parametri del modello

$\beta_j = 0$ IPOTESI NULLA (la variabile esplicativa j-esima "non pesa" nello spiegare la variabile dipendente)

$\beta_j \neq 0$ IPOTESI ALTERNATIVA (la variabile esplicativa j-esima "pesa" nello spiegare la variabile dipendente)

$$\frac{b_j - \beta_j}{s_j}$$

È la statistica del test (è una t di Student con T-K gradi di libertà)

Se l'ipotesi nulla è vera (ossia $\mathbf{x}_j$ non è statisticamente significativa nello spiegare la y)

$$\frac{b_j}{s_j}$$

è una t di Student con T-K gradi di libertà.

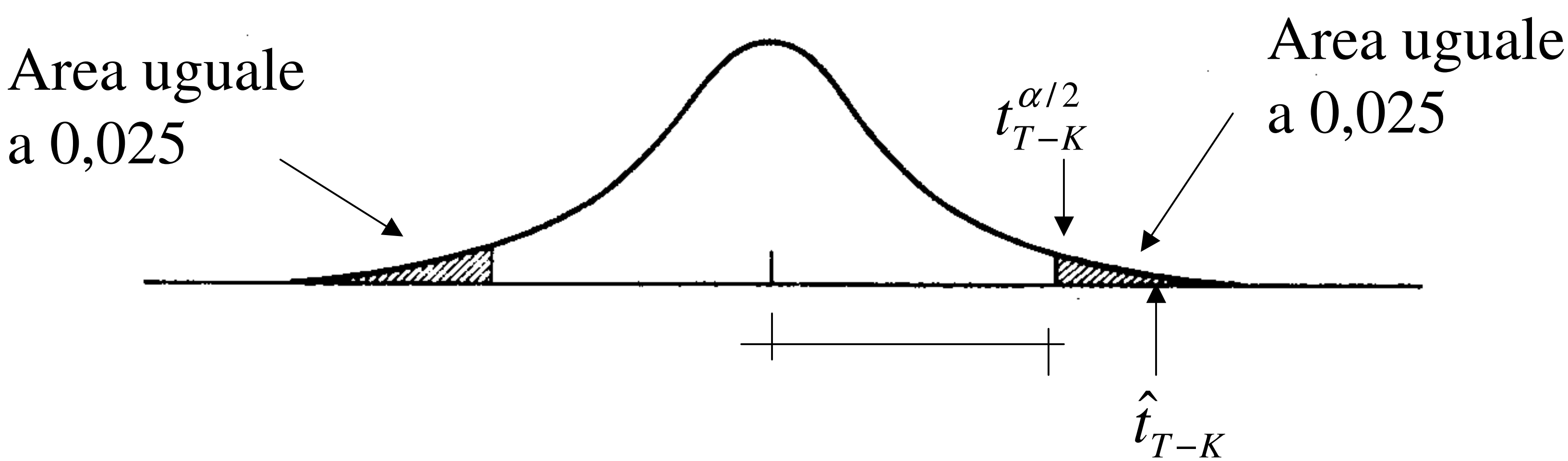
(Se per esempio nel modello ho 20 equazioni e tre parametri da stimare T-K=20-3=17 gradi di libertà)

Fisso una dimensione del test. Per esempio $\alpha = 0{,}05$, oppure $\alpha = 0{,}01$

$$\hat{t}_{T-K} = \frac{b_j}{s_j}$$

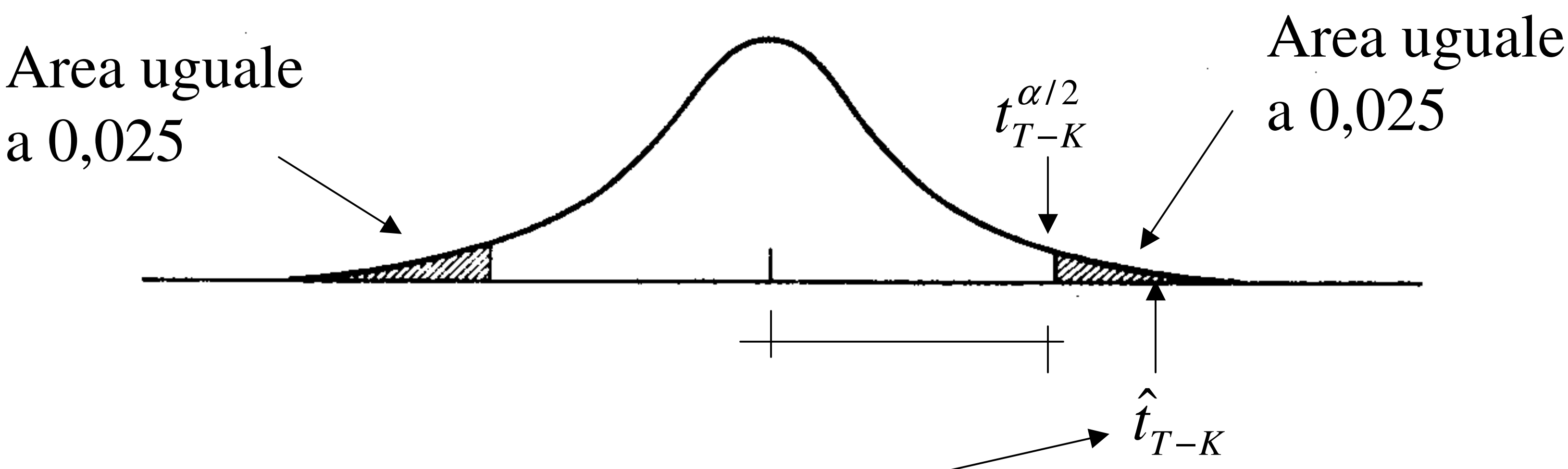
Calcolo il valore della t di Student con T-K gradi di libertà in corrispondenza dei miei dati.

Considero la t di Student con T-K gradi di libertà (è simile ad una normale standard)



(Se $\alpha = 0{,}05$ e T-K=17 $t^{\alpha/2}_{T-K} = t^{0{,}025}_{17} = 2{,}11$)

L'evidenza campionaria mi fa respingere *l'ipotesi nulla*: la variabile x_j è *statisticamente significativa* per spiegare *y*.



Se $\beta_j = 0$ l'evento è *possibile*, *ma improbabile*

Se invece si verifica *questo caso* accetto l'ipotesi nulla di non significatività statistica della variabile x_j

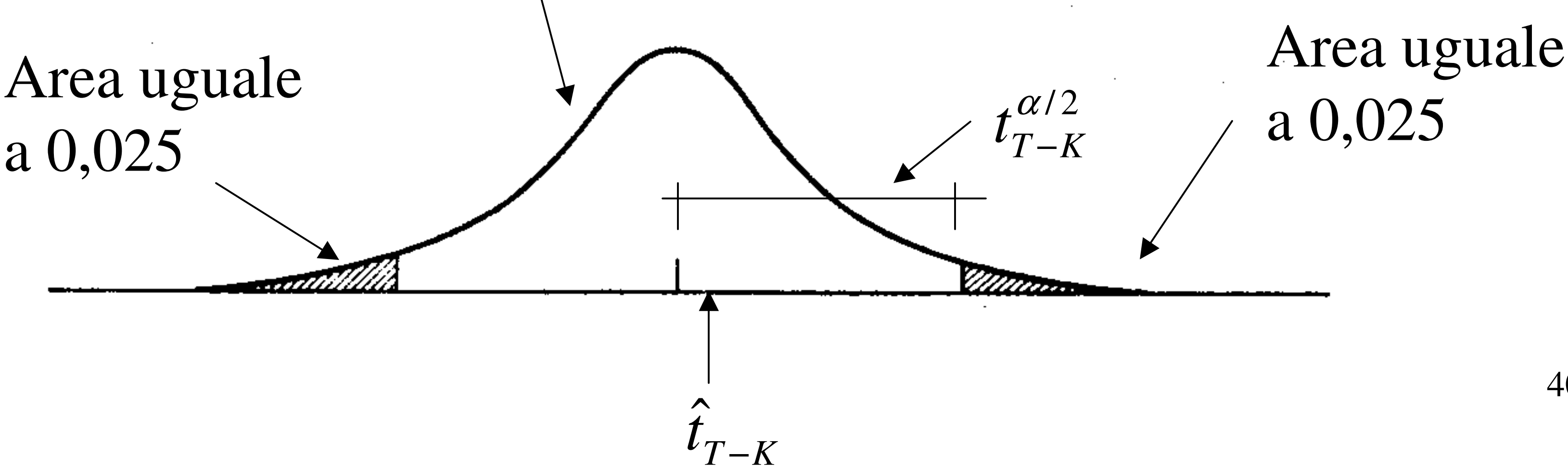


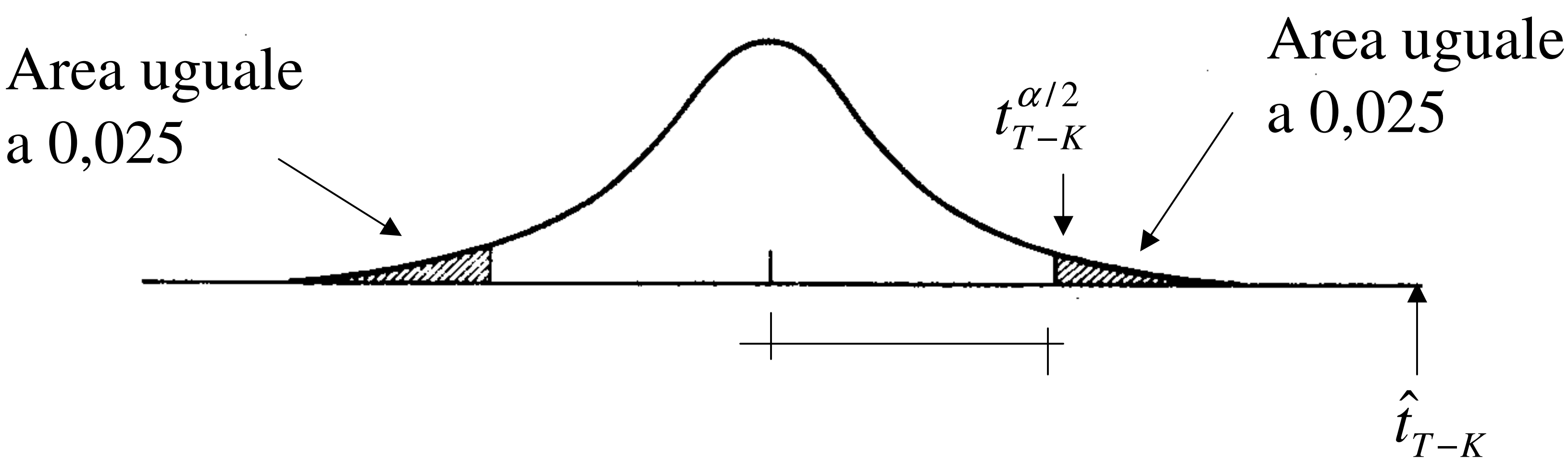
Table 2 **Right-Tail Critical Values for the *t*-distribution**

DF	$\alpha = .10$	$\alpha = .05$	$\alpha = .025$	$\alpha = .01$	$\alpha = .005$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
31	1.309	1.696	2.040	2.453	2.744
32	1.309	1.694	2.037	2.449	2.738
33	1.308	1.692	2.035	2.445	2.733
34	1.307	1.691	2.032	2.441	2.728
35	1.306	1.690	2.030	2.438	2.724
36	1.306	1.688	2.028	2.434	2.719
37	1.305	1.687	2.026	2.431	2.715
38	1.304	1.686	2.024	2.429	2.712
39	1.304	1.685	2.023	2.426	2.708
40	2.303	1.684	2.021	2.423	2.704
50	1.299	1.676	2.009	2.403	2.678
60	1.296	1.671	2.000	2.390	2.660
70	1.294	1.667	1.994	2.381	2.648
80	1.292	1.664	1.990	2.374	2.639
90	1.291	1.662	1.987	2.368	2.632
100	1.290	1.660	1.984	2.364	2.626
110	1.289	1.659	1.982	2.361	2.621
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.326	2.576

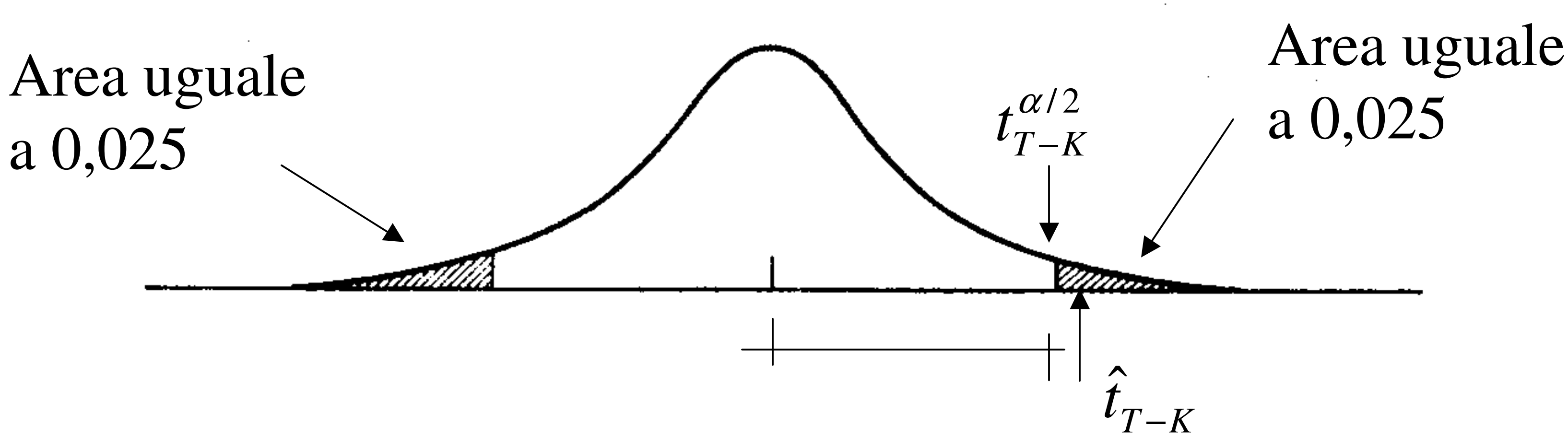
Source: This table was generated using the SAS® function TINV.

L'ipotesi nulla può essere respinta *con* *"più o meno forza"*

"più forza"

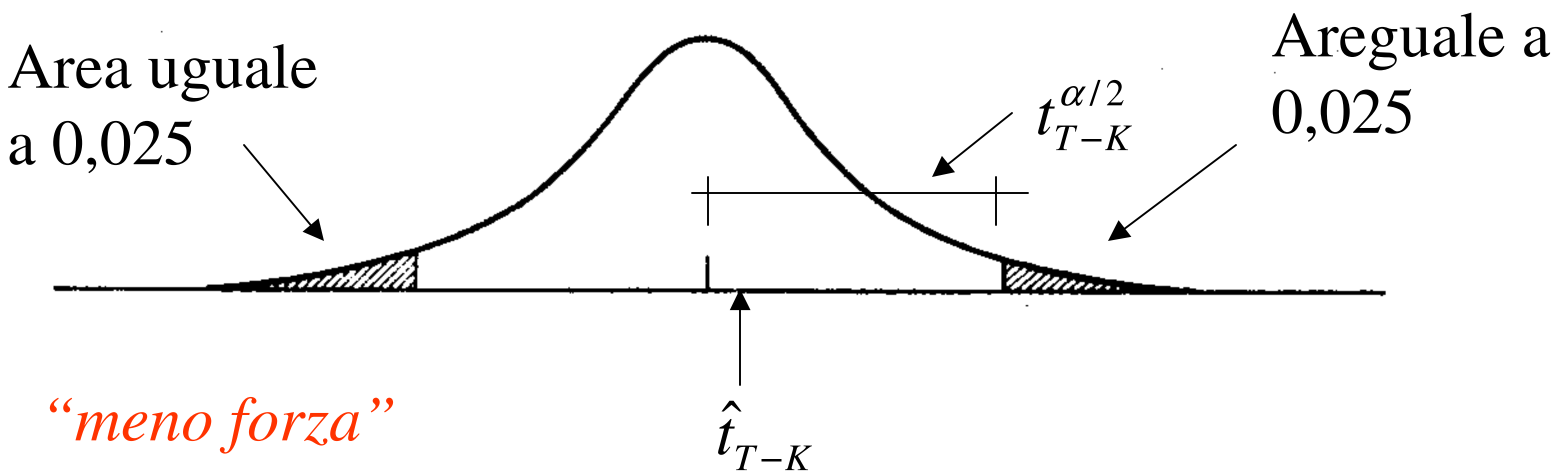


"meno forza"

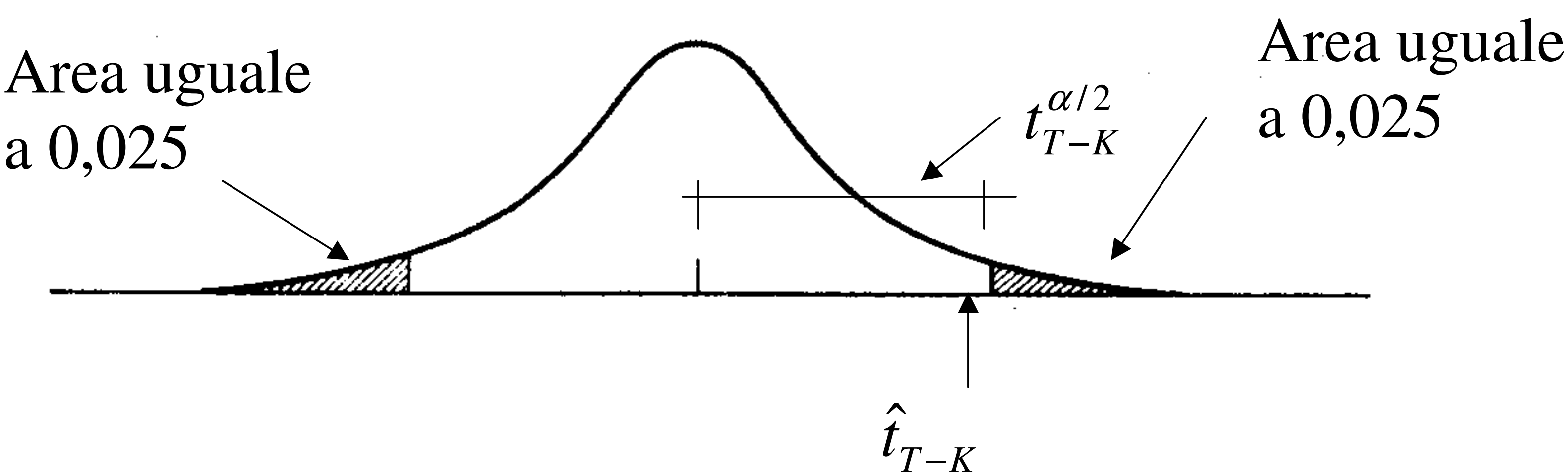


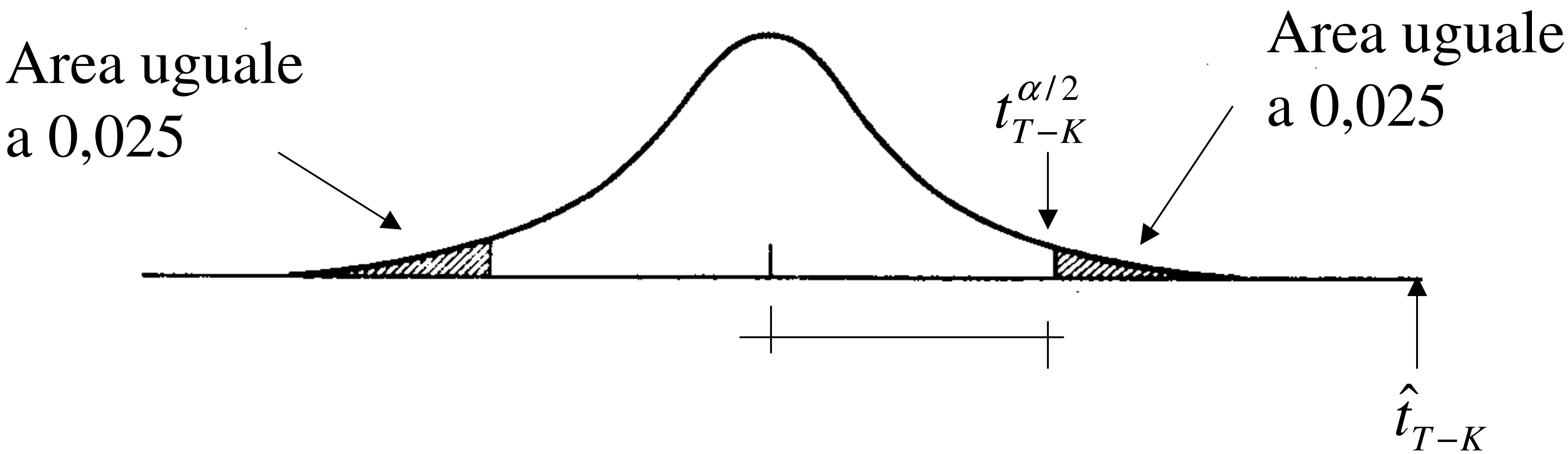
Così l'ipotesi nulla può essere accetta con *"più o meno forza"*

"più forza"



"meno forza"





L'"analista" spera che si verifichi questo caso: l'ipotesi nulla viene respinta con forza; la variabile è statisticamente significativa nello spiegare la variabile dipendente (la domanda). In questo caso vuol dire che è stata scelta bene la variabile esplicativa (indipendente) j-esima.

Il test t di Student sui parametri del modello *mi permette di selezionare* le variabili esplicative ad una ad una. Invece con il coefficiente R^2 le promuovo, o le "boccio", in blocco (il coefficiente di determinazione è una misura sulla "bontà" *globale* del modello).

Traffico autostradale italiano 1983/2000 (veicoli pesanti)

$$\ln d_t^{vp} = 3{,}62 + 1{,}99 \ln IPI_t - 0{,}76 T_t \qquad R^2 = 0{,}97$$

$$\hat{t}_{15}^{\ln IPI_t} = 13{,}42 \qquad \hat{t}_{15}^{T_t} = -3{,}88 \qquad \hat{t}_{15}^{0,025} = 2{,}131$$

Italia – Anni: 1970-1992 - Traffico aereo interno

$$\ln d_t = -17{,}1 + 1{,}89 \ln SE_t - 0{,}67 \ln T_t \qquad R^2 = 0{,}96$$

d_t = passeggeri imbarcati in un anno/popolazione

SE_t = prodotto interno lordo pro capite nell'anno (espresso in lire 1992)

T_t = rapporto fra gli incassi totali e i pass-km realizzati (espresso in cents/pass-km con valore 1992)

$$\hat{t}_{19}^{\ln SE_t} = 10{,}55 \qquad \hat{t}_{19}^{\ln T_t} = -2{,}33 \qquad \hat{t}_{19}^{0,025} = 2{,}093$$

Metodi utilizzati per la calibrazione di un modello di domanda

Modelli *descrittivi* → Estimatore dei *minimi quadrati*

Modelli *comportamentali* → Estimatore di *massima verosimiglianza*

$$V_i^n(\mathbf{X}_i^n) = \sum_{j=1}^{K} \beta_j x_{ij}^n$$

Vettore degli attributi dell'*i*-esima alternativa per l'*n*-esimo utente (K componenti)

Calibrare il modello $\longrightarrow$ *Determinare* il vettore dei *parametri incogniti* $\boldsymbol{\beta}$ (stimare i parametri incogniti)

Metodo della massima verosimiglianza

Intervisto N utenti

Per ogni utente n ho un vettore di attributi $\mathbf{X}_i^n \quad \forall\, i \in I^n$ (che conosco facendo le interviste)

Ogni utente n sceglie un'unica alternativa i

Introduco la variabile y_i^n

= 0 se l'utente non scegli i

= 1 se l'utente scegli i

Ammettiamo che: $I^n = \begin{cases} \text{aereo} \\ \text{treno} \\ \text{auto} \end{cases}$

Ammettiamo che l'utente *n* scegli l'alternativa aereo, sarà:

$$y^n_{aereo} = 1 \qquad y^n_{treno} = 0 \qquad y^n_{auto} = 0$$

La *probabilità di scelta dell'alternativa* aereo (*alternativa effettivamente scelta dall'utente*) secondo il modello la posso scrivere così:

$$\left(p^n_{aereo}\right)^1 \left(p^n_{treno}\right)^0 \left(p^n_{auto}\right)^0 = \left(p^n_{aereo}\right)^{y^n_{aereo}} \left(p^n_{treno}\right)^{y^n_{treno}} \left(p^n_{auto}\right)^{y^n_{auto}}$$

$$= \prod_{i \in I^n} \left(p^n_i\right)^{y^n_i} \longleftarrow \text{Forma compatta}$$

$$\prod_{n=1}^{N}\prod_{i\in I_n}\left(p_i^n\right)^{y_i^n} \longleftarrow$$

Probabilità congiunta di osservare le scelte campionarie

Nel caso di un modello logit $\longrightarrow$

$$p_i^n = \frac{e^{\boldsymbol{\beta}'\mathbf{X}_i^n}}{\sum_{j\in I^n} e^{\boldsymbol{\beta}'\mathbf{X}_j^n}}$$

p_i^n dipende dal vettore $\boldsymbol{\beta}$ da stimare

$$L(\beta_1, \beta_2, \ldots .\beta_K) = \prod_{n=1}^{N}\prod_{i\in I_n}\left(p_i^n\right)^{y_i^n}$$

Funzione di *verosimiglianza* (*likelihood).* Esprime la probabilità congiunta di osservare le scelte campionarie

E' una funzione che ha come variabili indipendenti le componenti del vettore $\boldsymbol{\beta}$.

Il *metodo della massima verosimiglianza* consiste nell'assumere come stima di $\boldsymbol{\beta}$ quel vettore dei parametri $\hat{\boldsymbol{\beta}}$ che *massimizza la funzione di verosimiglianza.*

Per semplificare i calcoli, invece di massimizzare la funzione di verosimiglianza ("likelihood") *massimizzo* il suo *logaritmo* (*funzione* denominata *"log-likelihood"*).

$$\mathscr{L}(\beta_1, \beta_2, \ldots. \beta_K) = \ln L(\beta_1, \beta_2, \ldots. \beta_K)$$

Essendo la funzione logaritmo una funzione strettamente crescente i due punti di massimo (della funzione di "likelihood" e "log-likelihood") coincidono.

$$\mathscr{L}(\boldsymbol{\beta}) = \sum_{n=1}^{N} \sum_{i \in I_n} y_i^n \ln p_i^n$$

Ho una sommatoria, invece di avere una produttoria come nella funzione di "likelihood".

L’estimatore di massima verosimiglianza gode di *buone proprietà asintotiche*: ossia valgono rigorosamente quando l’ampiezza campionaria $N \to \infty$, in pratica (per un ingegnere) basta che N “sia grande”.

- L’estimatore di massima verosimiglianza è un estimatore *consistente* : all’aumentare dell’ampiezza campionaria, N, si distribuisce con varianza sempre più bassa intorno al parametro da stimare.

- L’estimatore di massima verosimiglianza è un estimatore *efficiente (*minima varianza).

- L’estimatore di massima verosimiglianza è distribuito in modo *normale.*

$$\hat{\boldsymbol{\beta}} = \arg\max \mathscr{L}(\boldsymbol{\beta}) = \arg\max \sum_{n=1}^{N} \sum_{i \in I^n} y_i^n \ln p_i^n$$

$$p_i^n = \frac{e^{\boldsymbol{\beta}'\mathbf{X}_i^n}}{\sum_{j \in I^n} e^{\boldsymbol{\beta}'\mathbf{X}_j^n}} \longleftarrow \quad \text{modello logit}$$

$$\mathscr{L}(\boldsymbol{\beta}) = \sum_{n=1}^{N} \sum_{i \in I^n} y_i^n \ln p_i^n \;= \sum_{n=1}^{N} \sum_{i \in I^n} y_i^n \left(\boldsymbol{\beta}'\mathbf{X}_i^n - \ln \sum_{j \in I^n} e^{\boldsymbol{\beta}'\mathbf{X}_j^n} \right)$$

$$\mathscr{L}(\boldsymbol{\beta}) : R^K \to R$$

$$\frac{\partial \mathscr{L}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{n=1}^{N} \sum_{i \in I^n} y_i^n \left(\mathbf{X}_i^n - \frac{\sum_{j \in I^n} e^{\boldsymbol{\beta}'\mathbf{X}_j^n} \mathbf{X}_j^n}{\sum_{j \in I^n} e^{\boldsymbol{\beta}'\mathbf{X}_j^n}} \right)$$

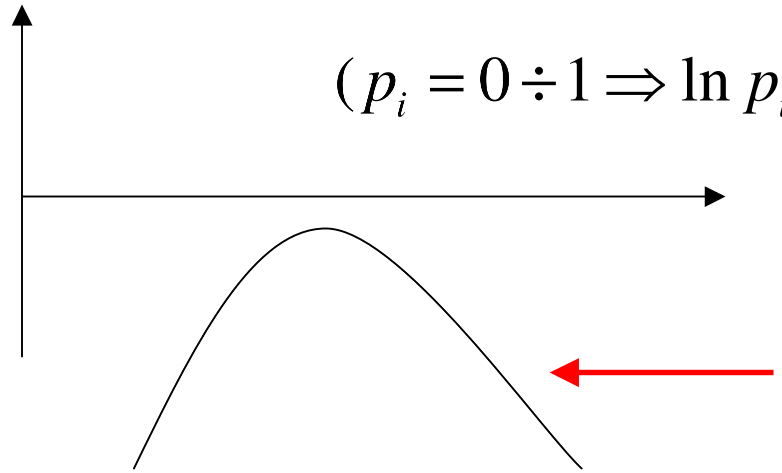
Vettore *gradiente* (k x 1)

$$\mathbf{H}(\boldsymbol{\beta}) = \begin{bmatrix} \dfrac{\partial^2 \mathcal{L}}{\partial \beta_1^2} \cdots\cdots\cdots\cdots\cdots\cdots \dfrac{\partial^2 \mathcal{L}}{\partial \beta_1 \partial \beta_K} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \dfrac{\partial^2 \mathcal{L}}{\partial \beta_K \partial \beta_1} \cdots\cdots\cdots\cdots\cdots\cdots \dfrac{\partial^2 \mathcal{L}}{\partial \beta_K^2} \end{bmatrix}$$

La matrice Hessiana è strettamente negativa: $\mathbf{z}'\mathbf{H}(\boldsymbol{\beta})\mathbf{z} < \mathbf{0} \quad \forall \mathbf{z} \in R^K e \neq \mathbf{0}$

$\Rightarrow$ La funzione di log-likelihood è strettamente concava: il massimo se esiste è unico

$(p_i = 0 \div 1 \Rightarrow \ln p_i \leq 0; \ln p_i = 0 \Leftrightarrow p_i = 1)$



Esempio di funzione strettamente concava

Nel caso dell'estimatore dei minimi quadrati riesco ad "esplicitare" l'estimatore: $\mathbf{b} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$. Nel caso dell'estimatore di massima verosimiglianza non riesco ad arrivare ad una *formula esplicita* dell'estimatore

Devo risolvere il problema il seguente problema di massimo (il problema della stima di massima verosimiglianza si riduce a risolvere un problema di massimo di una funzione a più variabili):

$$\hat{\boldsymbol{\beta}} = \arg\max \mathscr{L}(\boldsymbol{\beta}) = \arg\max \sum_{n=1}^{N} \sum_{i \in I^n} y_i^n \ln p_i^n$$

La funzione è *strettamente concava* inoltre non ho *vincoli* sui parametri β_j .

Passi dell'algoritmo di massimizzazione del logaritmo della funzione di verosimiglianza (log-likelihood).

Passo 0: *Determinazione del vettore di partenza .*

$$\boldsymbol{\beta}^1 = \begin{bmatrix} 0 \\ 0 \\ . \\ . \\ . \\ 0 \end{bmatrix} \qquad c(\text{contatore}) = 1$$

Uso questo vettore, a meno che non ne abbia uno migliore (per esempio: modelli analoghi, tentativi di stima precedenti).

Passo 1: *Calcolo.*

$$\mathcal{L}\left(\beta_1^c, \beta_2^c, \ldots.\beta_K^c\right) \qquad \nabla\mathcal{L}(\boldsymbol{\beta}^c)$$

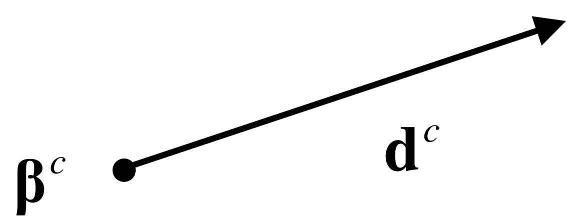
Passo 2: *Determino una direzione di "salita".*

Posso considerare il gradiente

$$\mathbf{d}^c = \nabla\mathcal{L}(\boldsymbol{\beta}^c) \qquad \text{oppure}$$

$$\mathbf{d}^c = -\mathbf{H}^{-1}\left(\boldsymbol{\beta}^c\right) \nabla\mathcal{L}(\boldsymbol{\beta}^c) \quad \text{direzione di “Newton-Raphson”.}$$

Con Newton-Raphson il procedimento converge in meno interazioni, però devo calcolare $\mathbf{H}$ e $\mathbf{H}^{-1}$.



Passo 3: *Massimizzazione monodimensionale.*

$$s^c = \arg\max_{0 \le s \le s_{\max}} \mathscr{L}\left(\boldsymbol{\beta}^c + s\mathbf{d}^c\right)$$

Unica variabile è lo scalare s

$$\mathscr{L}\left(\boldsymbol{\beta}^c + s\mathbf{d}^c\right): R \rightarrow R$$

Massimizzo la funzione ristretta alla direzione $\mathbf{d}^c$ (massimizzo la funzione lungo la direzione $\mathbf{d}^c$)

Passo 4: *Punto iniziale della iterazione successiva.*

$$\boldsymbol{\beta}^{c+1} = \boldsymbol{\beta}^c + s^c\mathbf{d}^c$$

Passo 5: *Test di arresto.*

Metodi "classici" → Sulla funzione (la differenza fra due iterazioni è piccola).

Metodi "classici" → Sul gradiente (la norma del gradiente è circa uguale a 0).

Ma a noi però interessano i coefficienti β_j

$$\frac{1}{K}\sum_{j=1}^{K}\frac{\left|\beta_j^{c+1}-\beta_j^{c}\right|}{\left|\beta_j^{c}\right|}<\varepsilon_1$$

La media degli scarti percentuali, fra le componenti di $\boldsymbol{\beta}$, è inferiore ad un certo valore.

$$\max_{\forall j} \frac{\left|\beta_j^{c+1} - \beta_j^c\right|}{\left|\beta_j^c\right|} < \varepsilon_2$$

Il valore massimo degli scarti percentuali, fra le componenti di $\boldsymbol{\beta}$, è inferiore ad un certo valore

Per esempio: $\varepsilon_1 = 0{,}01 \quad e \quad \varepsilon_2 = 0{,}05$

Se il test di arresto non è soddisfatto si pone $c = c + 1$ e si ritorna al passo 1.

Corroborazione di un modello di utilità aleatoria

Calcolo la statistica ρ^2 (è analoga a R^2 del modello di regressione lineare)

$$\rho^2 = 1 - \frac{\mathcal{L}(\hat{\boldsymbol{\beta}})}{\mathcal{L}(\mathbf{0})} \qquad\qquad 0 \le \rho^2 \le 1$$

$$\rho^2 = 0 \quad \Rightarrow \quad \mathcal{L}(\hat{\boldsymbol{\beta}}) = \mathcal{L}(\mathbf{0})$$ I parametri sono tutti nulli

$$p_i^n = \frac{e^{\boldsymbol{\beta}'\mathbf{X}_i^n}}{\sum_{j \in I_n} e^{\boldsymbol{\beta}'\mathbf{X}_j^n}} = \frac{1}{1+1+\text{.........}+1} = \frac{1}{\text{n. alternative}}$$

Il modello non fornisce nessuna ulteriore informazione rispetto ad una ipotesi di equiprobabilità delle alternative.

$$I^n = \begin{cases} \text{aereo} \\ \text{treno} \\ \text{auto} \end{cases} \qquad p^n_{aereo} = \frac{1}{3} \qquad p^n_{treno} = \frac{1}{3} \qquad p^n_{auto} = \frac{1}{3}$$

Se invece:

$$\rho^2 = 1 \quad \Rightarrow \quad \mathscr{L}\left(\hat{\boldsymbol{\beta}}\right) = 0$$

$$\mathscr{L}(\boldsymbol{\beta}) = \sum_{n=1}^{N} \sum_{i \in I_n} y_i^n \ln p_i^n = \sum_{n=1}^{N} 1 \quad \ln p_{\bar{i}}^n = 0$$

$\bar{i}$: alternativa effettivamente scelta dall'utente

$y_i^n = 0$ per le alternative non scelte dall'utente

$$\sum_{n=1}^{N} 1 \quad \ln p_{\bar{i}}^n = 0 \quad \Rightarrow \quad p_{\bar{i}}^n = 1 \ \forall n$$

Il modello prevede (per ogni utente) una probabilità di scelta pari ad 1 per l'alternativa $\bar{i}$ effettivamente scelta dall'utente: *modello "perfetto".*

Quanto più $\rho^2 \rightarrow 1$ tanto più il modello funziona.

- Altro test di corroborazione: *percentuale "right".*

La percentuale "right" è la *percentuale* di utenti del campione che ha scelto *l'alternativa di massima probabilità per il modello*: nel caso del logit quella di massima utilità sistematica.

Questa percentuale dovrebbe essere calcolata su un campione di utenti non utilizzato per calibrare il modello ("hold out sample").

- Terzo test di corroborazione: confronto la *percentuale di scelta* di una l'*alternativa* rilevata nel *campione* con quella prevista dal *modello* (anche questo test andrebbe eseguito su un campione "hold-out").

Percentuale di scelta riscontrata nel campione $\longrightarrow$ $\dfrac{\text{n. utenti che ha scelto } i}{\text{n. utenti che ha } i \text{ a disposizione } (i \in I^n)}$

Percentuale di scelta prevista dal modello $\longrightarrow$ $\displaystyle\sum_{n \in \mathcal{N}_i} \frac{p_i^n}{N_{\mathcal{N}_i}}$

$\mathcal{N}_i$ l'insieme degli utenti che hanno l'alternativa i a disposizione

$N_{\mathcal{N}_i}$ numerosità dell'insieme $\mathcal{N}_i$

Per esempio: ho intervistato 1000 utenti, 700 di questi hanno la patente e la disponibilità del mezzo privato.

- Test t di Student sui *parametri del modello*

Verifica *puntuale* sui parametri di un modello

Voglio verificare la significatività dell' attributo $\mathbf{X}_j$ nello spiegare le scelte effettuate dagli utenti. Faccio un test di ipotesi sul parametro β_j (*analogo* a quello utilizzato nella fase di corroborazione di un *modello di regressione lineare*)

$\beta_j = 0$ IPOTESI NULLA (la variabile esplicativa j-esima "non pesa" nello spiegare le scelte degli utenti)

$\beta_j \neq 0$ IPOTESI ALTERNATIVA (la variabile esplicativa j-esima "pesa" nello spiegare le scelte degli utenti)

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2_{\hat{\beta}_j})$$

La matrice di covarianza $\mathbf{\Omega}$ dell'estimatore di massima verosimiglianza si stima con:

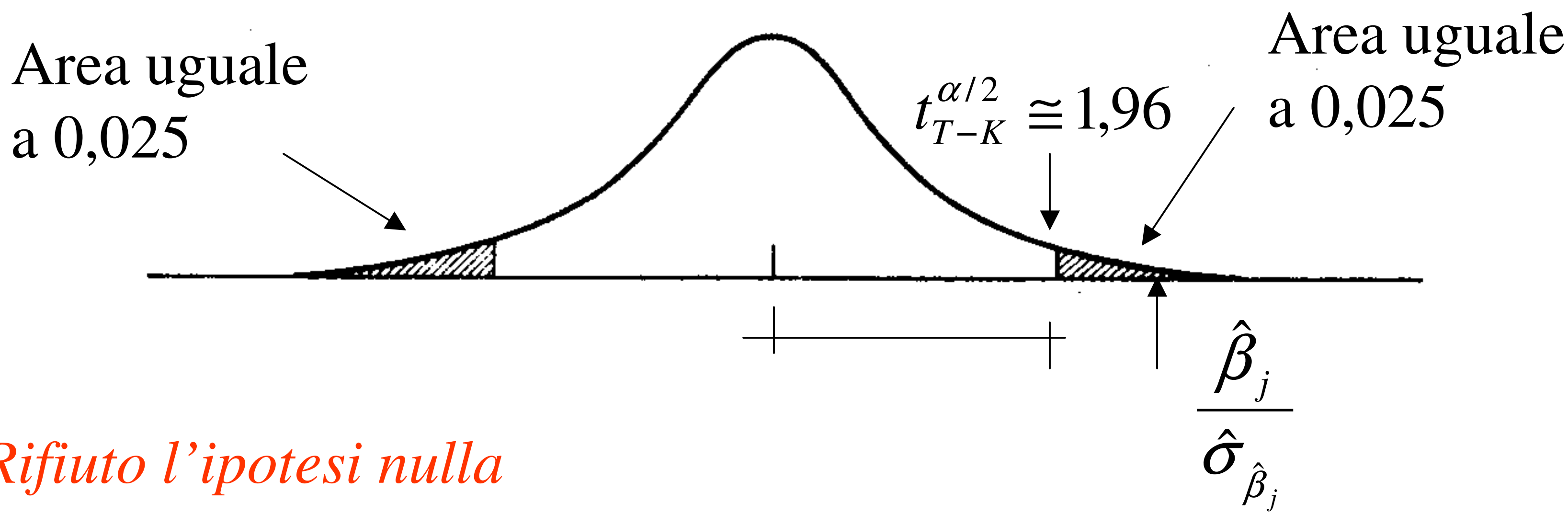
$$\hat{\mathbf{\Omega}} = -\left[\mathbf{H}(\hat{\boldsymbol{\beta}})\right]^{-1}$$

La stima è valida in termini asintotici (in pratica per N, ampiezza campionaria, "grande")

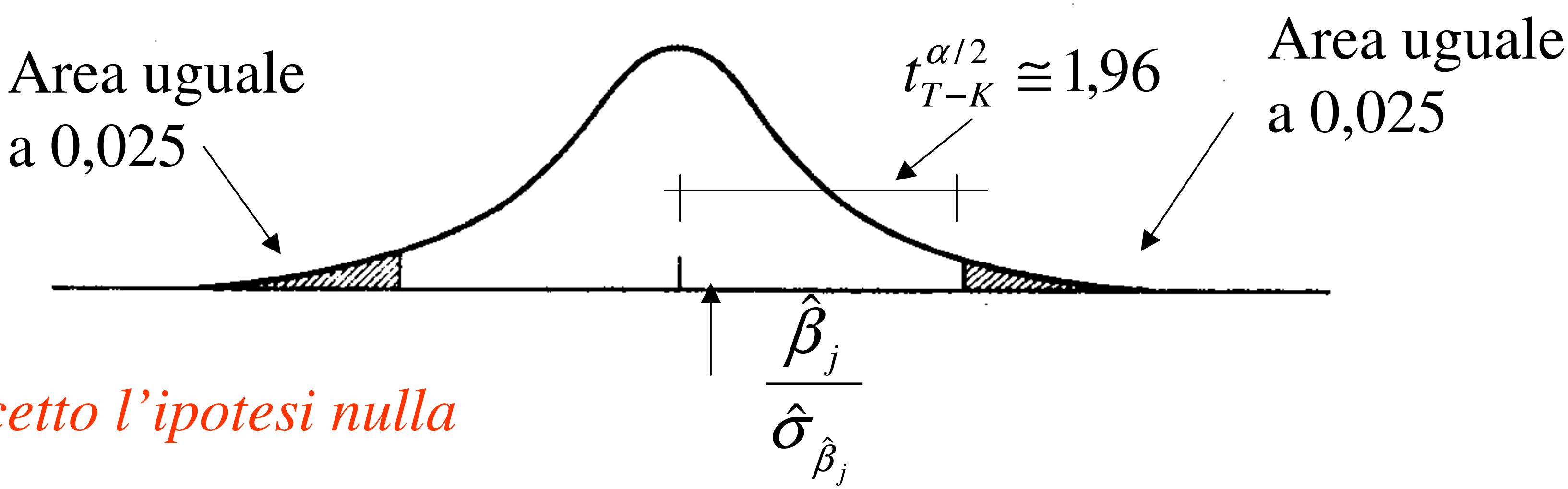
$$\frac{\hat{\beta}_j - 0}{\hat{\sigma}_{\hat{\beta}_j}}$$

E' una t di Student con N-K gradi di libertà (in pratica, dato l'alto numero di gradi di libertà, è una variabile aleatoria normale standard).

Dimensione del test 0,05



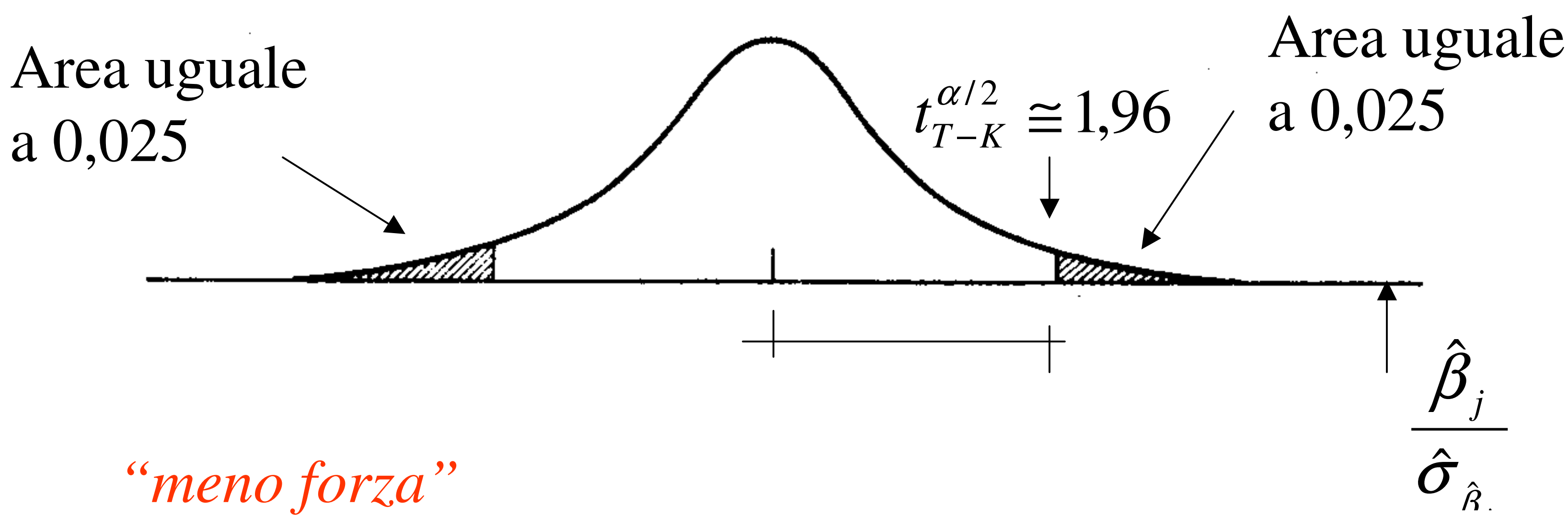
Rifiuto l'ipotesi nulla



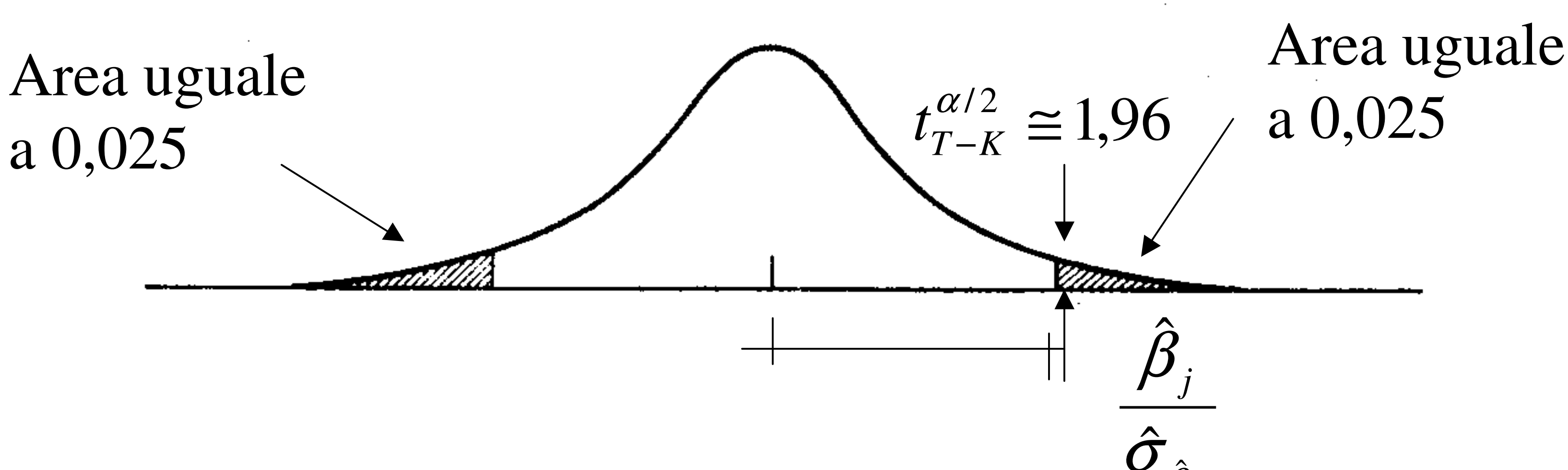
Accetto l'ipotesi nulla

Al solito l'ipotesi nulla può essere respinta *con più o meno forza:*

"più forza"

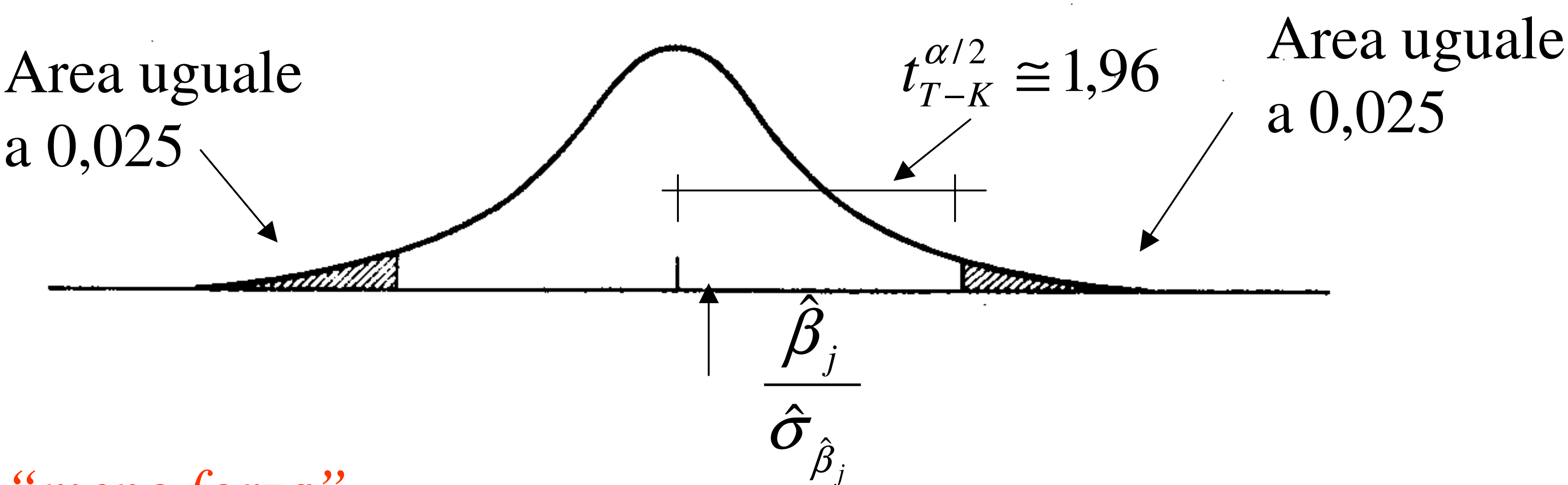


"meno forza"



Così l’ipotesi nulla può essere accetta con *più o meno forza:*

“più forza”



“meno forza”

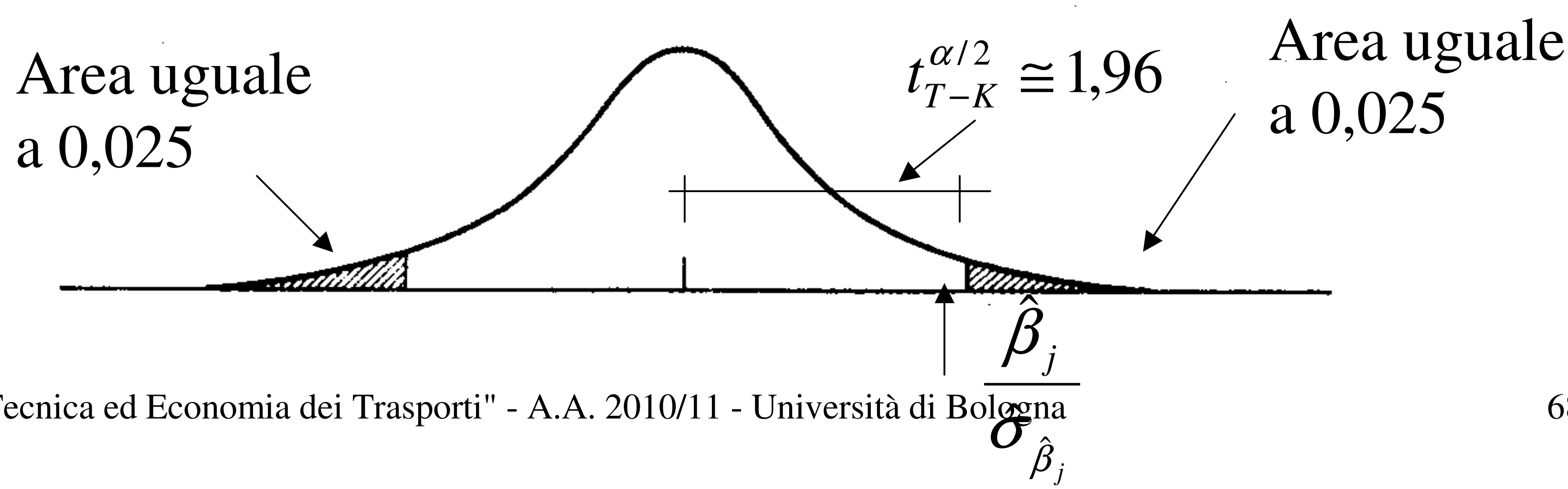


Table 4.6
Washington, D.C., binary logit model

Variable number	Variable name	Coefficient estimate	Asymptotic standard error	t statistic
1	Auto constant	1.45	0.393	3.70
2	In-vehicle time (min)	−0.00897	0.00630	−1.42
3	Out-of-vehicle time (min)	−0.0308	0.0106	−2.90
4	Auto out-of-pocket cost (¢)	−0.0115	0.00262	−4.39
5	Transit fare	−0.00708	0.00378	−1.87
6	Auto ownership (specific to auto mode)	0.770	0.213	3.16
7	Downtown workplace Dummy (specific to auto mode)	−0.561	0.306	−1.84

Summary statistics

Number of observations = 1476

Number of cases = 1476

$\mathscr{L}(0) = -1023$

$\mathscr{L}(\hat{\beta}) = -347.4$

$\rho^2 = 0.660$