

Lezione I

Francesco Bloise

francesco.bloise@uniroma1.it

Programma del corso

Parte II - ELEMENTI FONDAMENTALI DELL'ANALISI DI REGRESSIONE

4. Regressione lineare con un singolo regressore
5. Regressione con un singolo regressore: verifica di ipotesi e intervalli di confidenza
6. Regressione lineare con regressori multipli
7. Verifica di ipotesi e intervalli di confidenza nella regressione multipla
8. Funzioni di regressione non lineari
9. Valutazione di studi basati sulla regressione multipla

Programma del corso

Parte III - ULTERIORI SVILUPPI DELL'ANALISI DI REGRESSIONE

10. Regressione con dati panel
11. Regressione con variabile dipendente binaria
12. Regressione con variabili strumentali
13. Esperimenti e quasi esperimenti
14. Predizione con molti regressori e grandi dati

Testo di riferimento

Stock, J. H., e Watson, M. W. (2020). *Introduzione all'econometria*, quinta edizione, Pearson Italia

Modalità d'esame

- Esame scritto
- 3 macro domande con sottopunti
- Domande teoriche
- Domande di interpretazione dei risultati di una regressione

A cosa serve l'econometria

- Utilizza un insieme di strumenti statistici applicati all'analisi di possibili relazioni (causali o non) tra variabili
- E' molto spesso utilizzata per predire (prevedere) un determinato fenomeno di interesse
- Lo strumento principale è l'analisi di regressione lineare
- Ulteriori strumenti e stimatori per risolvere alcune delle problematiche che rendono inefficace lo stimatore dei minimi quadrati ordinari (OLS)
- A quali domande risponde l'econometria?

Esempi di possibili domande econometriche

- Ridurre la dimensione delle classi può aumentare l'apprendimento nelle scuole elementari?
- Esiste discriminazione razziale nella concessione dei mutui immobiliari?
- Aumentare la spesa sanitaria migliora l'aspettativa di vita o lo stato di salute degli individui?
- Di quanto crescerà il PIL Italiano il prossimo anno?

Esempi di possibili domande econometriche

- C'è un'associazione fra spot pubblicitari e vendite?
- E' possibile prevedere le vendite future di un'azienda
- La relazione tra due variabili è lineare?
- E' possibile prevedere gli individui (o le famiglie) che diventeranno poveri il prossimo anno?

Struttura dei dati

- Dati cross-sezionali consistono di più entità osservate in un solo periodo.
- Dati longitudinali. Consistono in una singola entità osservata in più periodi. → creazione di una serie storica.
- Dati Panel: più entità studiate nel tempo (ho molti dati osservati nel tempo). È l'evoluzione di un fenomeno.

Possibili fonti di dati in Econometria

- **Dati campionari (da survey)**

Assegno la popolazione questionari creando così un campione casuale e di questa rappresentativo. Ottengo informazioni dai questionari. Abbiamo una quantità di dati ridotti. siamo studiando un fenomeno estremo o poco numeroso.

Esempi:

Indagine sui bilanci della famiglie italiane (Banca d'Italia)

European Union Statistics on Income and Living Conditions (EU-SILC)

- **Dati Amministrativi o fiscali**

Vengono tenute in considerazione tante variabili (solo quelle essenziali). Otteniamo allora molti dati: Rappresentiamo la popolazione.

Esempi:

Archivi amministrativi Istituto Nazionale di Previdenza Sociale (INPS).

Archivi Fiscali dell'Agenzia delle entrate

Caratteristiche dei dati da indagine campionaria

- Interviste specifiche a un campione rappresentativo della popolazione totale
 - Dati campionari registrano molte variabili (istruzione, età, reddito e molte altre caratteristiche individuali o familiari)
 - Limiti nell' osservare «fenomeni estremi» o «poco numerosi»
-

Caratteristiche dei dati amministrativi

- Dati amministrativi riferiti a una specifica popolazione di riferimento – i pensionati, i lavoratori dipendenti, gli iscritti a un Ateneo
 - Dati amministrativi rilevano solo le variabili utili ai fini amministrativi (e.g. istruzione, famiglia e origini familiari non rilevate in INPS)
-

Il modello di regressione lineare

Caso più semplice: Regressione lineare con un solo regressore

$$y_i = \alpha + \beta X_i + \varepsilon_i$$

y_i = variabile dipendente

α = costante

β = pendenza della retta di regressione lineare

X_i = regressore o variabile esplicativa

ε_i = termine di errore

i = unità di analisi

Minimizzazione della somma degli scarti quadratici tra i valori osservati e la retta (valori stimati). Usiamo il quadrato perché le distanze possono essere sia positive che negative.

Partendo dal campione rappresentativo della popolazione si fa inferenza tenendo conto dell'incertezza (la valuteremo con SE, intervalli di confidenza).

Quando con l'OLS ho problemi di endogeneità passerò a tale studio tramite altri modelli.

Varie possibili unità di osservazione

- Individui
- Famiglie
- Imprese
- Nazioni
- Città
- Regioni
- Province
- Banche

Possibili obiettivi dell'analisi econometrica

Obiettivi: Identificazione di

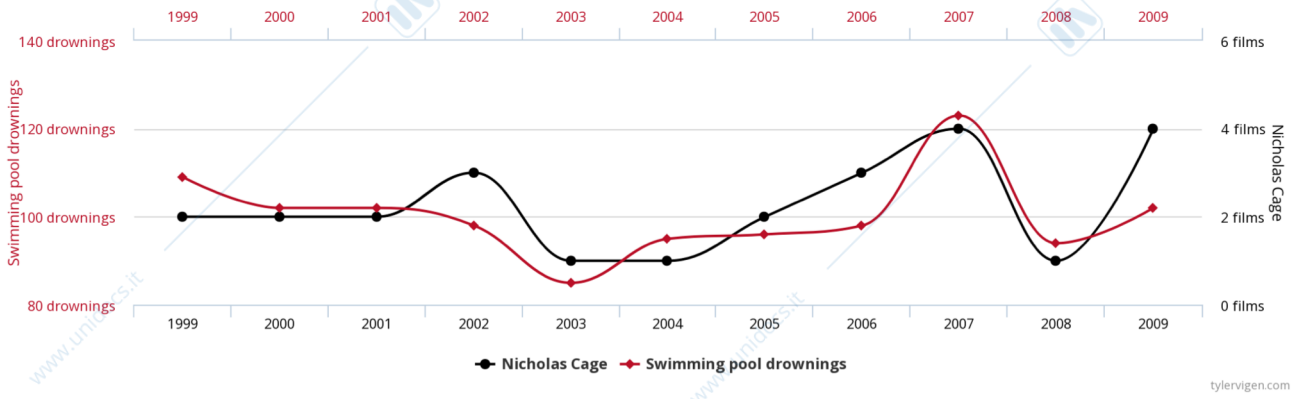
- effetti causali (necessità di distinguere tra associazioni di variabili e rapporto causale) → Una correlazione non implica un effetto causale
- o
- predizione (di un fenomeno)

Quando non c'è relazione causale tra variabili ma sono correlate

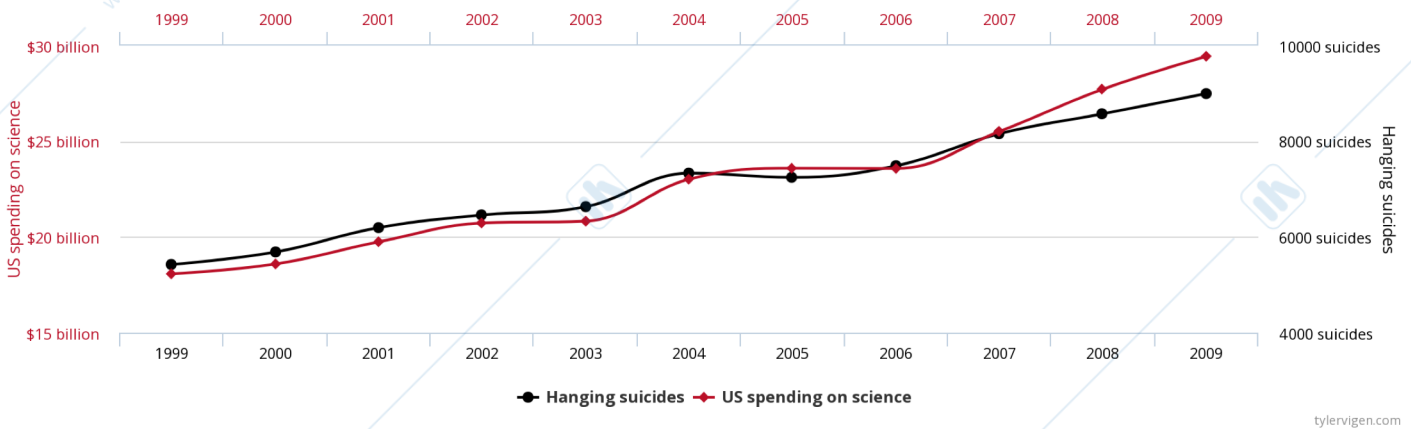
Alcuni esempi di correlazioni spurie

Fonte: [Spurious Correlations \(tylervigen.com\)](http://SpuriousCorrelations.tylervigen.com)

Number of people who drowned by falling into a pool correlates with Films Nicolas Cage appeared in



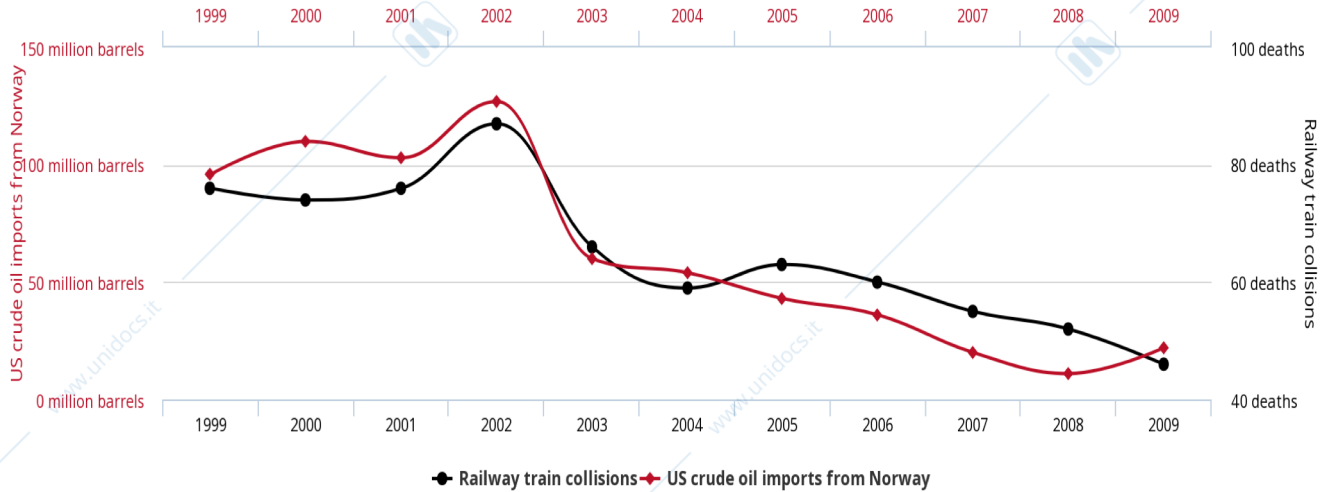
US spending on science, space, and technology correlates with Suicides by hanging, strangulation and suffocation



US crude oil imports from Norway

correlates with

Drivers killed in collision with railway train

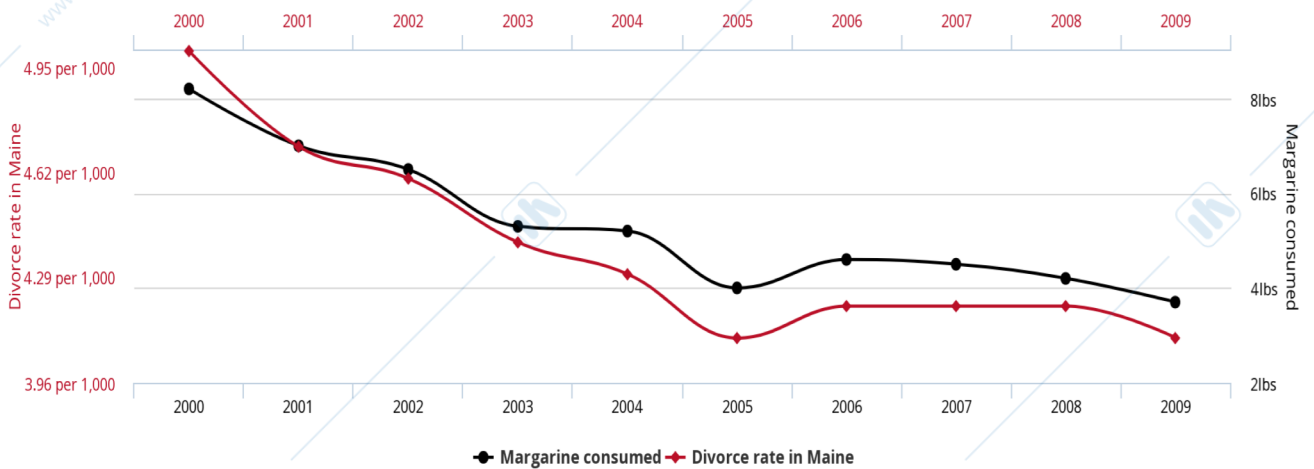


tylervigen.com

Divorce rate in Maine

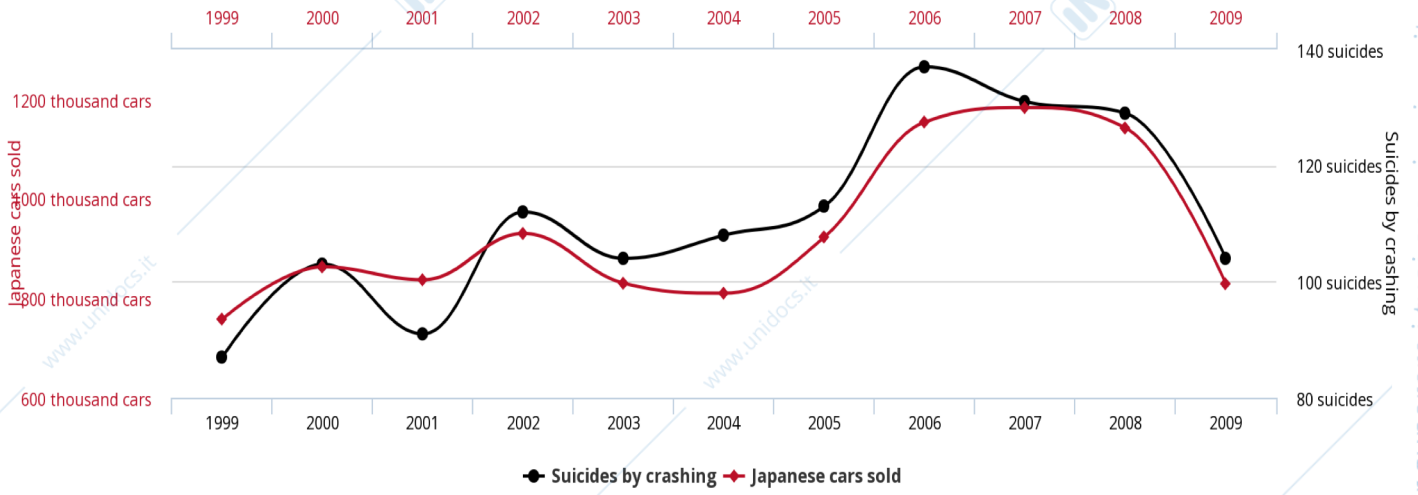
correlates with

Per capita consumption of margarine



tylervigen.com

Japanese passenger cars sold in the US correlates with Suicides by crashing of motor vehicle



tylervigen.com

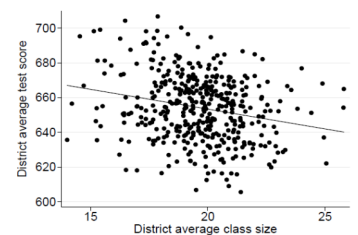
Lezione 2

Francesco Bloise

francesco.bloise@uniroma1.it

RELAZIONE TRA LE VARIABILI

Nella scorsa lezione abbiamo visto il modello di regressione lineare semplice dove abbiamo una variabile dipendente, un termine costante e un regressore (il termine di errore). La differenza con un'equazione lineare semplice è che noi non sappiamo l'associazione che c'è tra le due variabili. Quando scriviamo un'equazione lineare con un termine costante e una variabile x andiamo ad imporre una pendenza di quella curva. Nel nostro caso abbiamo un β da stimare. Per **stimare la relazione tra le due variabili** noi utilizzeremo il modello di regressione semplice con l'utilizzo del metodo dei minimi quadrati. Se guardiamo questa «nuvola di punti noi vediamo che c'è una relazione tra le due variabili: i risultati scolastici in un distretto e la dimensione delle classi medie nel distretto; per trovare la retta che interpreta meglio questi punti utilizzeremo il *metodo dei minimi quadrati* con il quale minimizziamo la distanza tra la somma dei punti e la retta di regressione al quadrato. -> troviamo quindi il β minimizzando le distanze al quadrato e la retta di regressione che interpreta meglio questi punti. (Utilizzeremo il quadrato perché avremo valori sia positivi che negativi). Cerchiamo il minor numero possibile tra i risultati positivi. La relazione tra due variabili è negativa : più è ampia la dimensione delle classi, più abbiamo voti peggiori; Nulla possiamo dire sulla relazione causale (se esiste o meno) delle due variabili; dire che le due variabili si muovono insieme non implica che una va ad incidere sull'altra(i voti peggiori sono dovuti ad una classe ampia). Inoltre non possiamo sapere se l'associazione negativa è statisticamente significativa ossia dobbiamo osservare il livello di incertezza (dobbiamo tener conto del tipo di stimatore che stiamo utilizzando quando calcoliamo il livello di incertezza; ad esempio nel caso dei minimi quadrati).



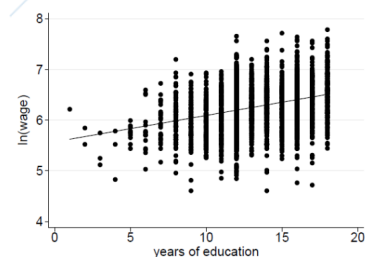
Per tener conto dell'incertezza dobbiamo vedere gli standard error dei beta stimati e degli intervalli di confidenza.

Dobbiamo tener conto inoltre nell'ambito dell'incertezza dobbiamo tener conto anche del tipo di stimatore che stiamo utilizzando; esempio in quello dei minimi quadrati ordinari gli standard error hanno delle caratteristiche specifiche invece, in quello dei minimi quadrati a due stadi, stiamo utilizzando due diverse regressioni per correggere i problemi legati ai problemi alle relazioni casuali quindi in questo stimatore l'incertezza aumenta, gli standard error aumentano e la varianza aumenta nella stima. Quando il metodo dei minimi quadrati ordinari ha problemi di endogeneità si utilizzano altri stimatori per correggere tali errori.

Qual è l'effetto dell'istruzione sul salario? CORRELAZIONE

Altro esempio che ci mostra come andremo a stimare le rette di regressione lineare è questo esempio con salario-livelli di istruzione. (In modo più preciso abbiamo: livelli di anno di studio che coincidono con le fasi scolastiche e il logaritmo del salario. **Il log perché abbiamo una non linearità** quindi abbiamo una trasformazione della variabile dipendente o del regressore; In termini semplici utilizziamo il log perché possiamo interpretare i coefficienti in termini percentuali perché *la differenza di due logaritmi può essere interpretata come la differenza percentuale tra i salari dei due individui*: es. quanto varia il salario tra i diversi soggetti). (la differenza tra due logaritmi si approssima alla variazione %).

L'associazione tra salario e anni di studio è positivo (d'altronde come ci aspettavamo). Tuttavia abbiamo problemi di endogeneità quindi possiamo interpretare tale grafico come la dimostrazione di un **associazione** delle due variabili.



Le variabili casuali:

Prima di vedere i modelli di regressione lineare univariati (singolo regressore) o multivariati (più regressione) abbiamo bisogno di un ripasso di statistica e di probabilità:

- Una **variabile casuale** (o aleatoria) è una realizzazione numerica di un fenomeno complesso che sono imprevedibili e casuali (es. numeri di giorni in cui pioverà la prossima settimana). Presenta delle realizzazioni numeriche di fenomeni imprevedibili o casuali. Es. I voti all'esame, il numero di giorni in cui nevierà la prossima settimana
- Il numero **realizzazioni** può essere sia ristretto ad un intervallo che andare da 0 a infinito. Es. Il numero di giorni in cui nevierà la prossima settimana è casuale e assume un valore numerico (0,1,2,3,4,5,6 o 7).

Esistono due tipi di variabili casuali:

Una variabile casuale è **discreta** quando assume un numero discreto di valori, come 0,1,2, ... (Assume 2 o più valori). Parliamo in questo caso di variabili categoriali. Utilizzeremo modelli di regressione diversi da quello semplice: es. modelli per variabili dipendenti ordinarie come il logit e progit (?) perché se abbiamo variabili dicotomiche dobbiamo interpretare i risultati dei coefficienti come variazione di probabilità e nel modello di regressione semplice nella maggior parte dei casi non è possibile fare questo (servono modelli più specifici per questo di variabili).

Una variabile casuale **continua** assume un continuum di valori possibili (possono assumere un qualunque valore in un continuo di punti). Utilizzeremo in questo caso il modello di regressione lineare semplice.

Distribuzione di probabilità delle variabili casuali discrete

- Ogni risultato di una variabile casuale discreta si verifica con un determinata probabilità
- Una distribuzione di probabilità di una variabile casuale discreta è l'elenco dei possibili valori della variabile e la probabilità che ogni valore si verificherà
- Ad esempio, ipotizziamo una variabile casuale S che rappresenta il numero di giorni in cui nevierà nella prima settimana di gennaio

Probability distribution of S								
Outcome	0	1	2	3	4	5	6	7
Probability	0.20	0.25	0.20	0.15	0.10	0.05	0.04	0.01

Distribuzione di probabilità cumulata: variabili casuali discrete

- Una distribuzione di probabilità cumulata (CDF) è la probabilità che il valore numerico di una realizzazione della variabile sia minore o uguale a un valore particolare.
- Ad esempio la probabilità che nevichi al massimo s giorni. Abbiamo 8 possibilità che la settimana prossima nevichi e ognuna delle possibilità che la variabile casuale discreta può assumere ha una certa probabilità; abbiamo una possibile realizzazione h a cui abbiamo una probabilità associata definita sul mese di gennaio. Es la probabilità che nevichi 0 giorni la settimana prossima è del 20%. Abbiamo che la distribuzione di prob cum è l'elenco delle possibili realizzazioni della variabile che può assumere e per ogni realizzazione abbiamo una probabilità associata. Mettiamo tali dati nella tabella.
- La probabilità è invece cumulata quando alla variabile s (giorni in cui nevicherà la prima settimana di gennaio), le realizzazioni, le prob associate aggiungiamo un'ulteriore riga : la prob cumulata. Vogliamo vedere con la cumulata, la **probabilità che la realizzazione sia al massimo entro un certo valore** (la realizzazione deve essere minore o uguale a un certo valore: la probabilità che nevichi al massimo un giorno è del 45% (la somma del giorno 0 e 1)

(cumulative) Probability distribution of S								
Outcome	0	1	2	3	4	5	6	7
Probability	0.20	0.25	0.20	0.15	0.10	0.05	0.04	0.01
CDF	0.20	0.45	0.65	0.80	0.90	0.95	0.99	1

Distribuzione di probabilità: variabili casuali continue

- La temperatura che ci sarà domani è un esempio di una variabile casuale continua. La CDF è definita in modo simile a una variabile casuale discreta. (Potenzialmente domani ci potrebbe essere qualsiasi temperatura quindi la realizzazione numerica è una possibile realizzazione di tutte le realizzazioni possibili data la popolazione.
- Una distribuzione di probabilità che elenca tutti i valori e la probabilità di ciascuno valore non è adatto per una variabile casuale continua.

Invece la probabilità in questo caso può essere riassunta in una funzione di densità di probabilità (PDF / densità). Questo per dire che le variabili casuali presentano delle realizzazioni numeriche di fenomeni imprevedibili o casuali. Nella regressione in genere si usano variabili continue con variabili dipendenti lineari uso probit o logit (in questi casi i coefficienti sono da interpretare come variazioni di probabilità).

Nel caso volessimo la probabilità cumulata: abbiamo due problemi (che non si creano invece nella variabile discreta):

- 1) l'intervallo non limitato
- 2) Non è possibile associare una possibile probabilità ad uno specifico valore.

Se ad esempio mi chiedessi qual è la probabilità che a Roma la temperatura sia di 18.9 con il continuo è prossima allo 0. Non è possibile definire una probabilità specifica per una singola realizzazione specifica. Per questo motivo è possibile calcolare la probabilità su dei range (es. prob che la temperatura che sia minore di 20 gradi quindi da 0 a 20 gradi o che la temperatura sia maggiore di una certa temperatura prefissata. Per fare questo dobbiamo rappresentare la distribuzione di probabilità attraverso una funzione di densità di probabilità. (quella che nel caso delle distribuzioni normali ha la forma a campana.)

Il valore atteso di una variabile casuale discreta

$$E(Y) = \sum_{i=1}^n y_i \cdot Pr(Y = y_i) = \mu_Y$$

Number of days it will snow in the last week of January (S)								
Outcome	0	1	2	3	4	5	6	7
Probability	0.20	0.25	0.20	0.15	0.10	0.05	0.04	0.01

$$E(S) = 0 \cdot 0.2 + 1 \cdot 0.25 + 2 \cdot 0.2 + 3 \cdot 0.15 + 4 \cdot 0.1 + 5 \cdot 0.05 + 6 \cdot 0.04 + 7 \cdot 0.01 = 2.06$$

Le y con i sono le singole realizzazioni;

La media: è la sommatoria da 1 a n delle realizzazioni di s moltiplicate per le rispettive probabilità;

Nb. Quando le realizzazioni della variabile hanno tutte identiche probabilità di realizzazione allora non dobbiamo ponderare moltiplicando ogni realizzazione per la sua rispettiva probabilità.

La varianza di una variabile casuale discreta

$$\sigma_Y^2 = Var(Y) = E[(Y - \mu_Y)^2] = \sum_{i=1}^k (y_i - \mu_Y)^2 \cdot Pr(Y = y_i)$$

Number of days it will snow in the last week of January (S)								
Outcome	0	1	2	3	4	5	6	7
Probability	0.20	0.25	0.20	0.15	0.10	0.05	0.04	0.01

$$Var(S) = \frac{(0 - 2.06)^2 \cdot 0.2 + (1 - 2.06)^2 \cdot 0.25 + (2 - 2.06)^2 \cdot 0.2 + (3 - 2.06)^2 \cdot 0.15 + (4 - 2.06)^2 \cdot 0.1 + (5 - 2.06)^2 \cdot 0.05 + (6 - 2.06)^2 \cdot 0.04 + (7 - 2.06)^2 \cdot 0.01}{2.94}$$

La varianza è indice di dispersione : dà un valore sintetico al valore atteso delle differenze al quadrato della media.

Abbiamo che la radice della varianza è sigma, ossia la deviazione standard espressa nella stessa unità di misura della y .

Nota: La deviazione standard è espressa nella stessa unità di misura della variabile originaria Variazioni rispetto alla media (usiamo il quadrato perché le differenze potrebbero essere positive o negative).

L'esempio della Bernoulli

La variabile di Bernoulli è una variabile casuale discreta dicotomica (ossia che può assumere soltanto due valori):

- 1 con probabilità p
- 0 con probabilità $(1-p)$

Es. Testa o croce;

$$E(B) = \mu_B = \sum_{i=1}^k b_i \cdot Pr(B = b_i) = 1 \cdot p + 0 \cdot (1-p) = p$$

$$\begin{aligned} Var(B) = \sigma_B^2 &= \sum_{i=1}^k (b_i - \mu_B)^2 \cdot Pr(B = b_i) \\ &= (0 - p)^2 \cdot (1-p) + (1 - p)^2 \cdot p \\ &= p(1-p) \end{aligned}$$

Valore atteso e varianza di una funzione lineare

$$Y = a + b \cdot X$$

$$E(Y) = \mu_Y = E(a + b \cdot X) = a + b \cdot E(X) = a + b \cdot \mu_X$$

$$\begin{aligned} Var(Y) = \sigma_Y^2 &= E[(Y - \mu_Y)^2] \\ &= E[((a + bX) - (a + b\mu_X))^2] \\ &= E[b^2(X - \mu_X)^2] \\ &= b^2 E[(X - \mu_X)^2] \\ &= b^2 \cdot \sigma_X^2 \end{aligned}$$

Indipendenza

- Due variabili casuali Y e X sono indipendenti se la distribuzione condizionata di y non dipende da x :

$$Pr(Y = y | X = x) = Pr(Y = y)$$

Nb: se due variabili sono indipendenti non c'è legame causale!

- Nel caso del valore atteso:

$$E(Y | X) = E(Y)$$

Tornando all'esempio della relazione tra neve e giorni di freddo:

$$E(Y | X = 1) = E(Y | X = 0) = E(Y)$$

Per quanto riguarda l'indipendenza parliamo poi di covarianza e correlazione..

La covarianza

- La covarianza è misura il grado di associazione tra due variabili: se le due variabili si muovono insieme;
- Nb. Non ci dice l'impatto della x sulla y ; per vedere tale impatto valuteremo anche la relazione tra x (regressore) e l'errore.
- La covarianza misura **l'associazione tra due variabili** ma il valore non è facilmente interpretabile (dipende dall'unità di misura delle due variabili)

$$\begin{aligned} Cov(X, Y) = \sigma_{XY} &= E[(X - \mu_X) \cdot (Y - \mu_Y)] \\ &= \sum_{i=1}^k \sum_{j=1}^l (x_j - \mu_X)(y_i - \mu_Y) \cdot Pr(X = x_j, Y = y_i) \end{aligned}$$

	Very cold ($X = 1$)	Not very cold ($X = 0$)	Total
Snow ($Y = 1$)	0.15	0.07	0.22
No snow ($Y = 0$)	0.15	0.63	0.78
Total	0.30	0.70	1.00

$$\begin{aligned} Cov(X, Y) &= (1 - 0.3)(1 - 0.22) \cdot 0.15 \\ &+ (1 - 0.3)(0 - 0.22) \cdot 0.15 \\ &+ (0 - 0.3)(1 - 0.22) \cdot 0.07 \\ &+ (0 - 0.3)(0 - 0.22) \cdot 0.63 \end{aligned}$$

Lezione 3

Francesco Bloise

francesco.bloise@uniroma1.it

Correlazione

- La correlazione invece varia tra -1 e +1
- Non dipende dall'unità di misura (non ha range a differenza della covarianza)
- Formula :

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

(stiamo standardizzando la covarianza;)

- Due variabili non sono correlate quando la correlazione (e la covarianza) sono pari a 0
($\text{COV}(X,Y)=0 \Rightarrow \text{CORR}(X,Y)=0$)

La legge delle aspettative iterate

$$E(Y) = E[E(Y | X)] = \sum_i E(Y | X = x_i) \cdot Pr(X = x_i)$$

Un possibile esempio:

Il valore atteso del salario del totale lavoratori può essere scomposto nel:

1. Valore atteso del salario degli uomini
2. Valore atteso del salario delle donne

Quindi il valore atteso totale sarà uguale alla somma dei due valori attesi ponderata per la specifica probabilità di essere un lavoratore uomo o un lavoratore donna.

Media, varianza e covarianza della somma di variabili casuali

Means, Variances, and Covariances of Sums of Random Variables

KEY CONCEPT

2.3

Let X , Y , and V be random variables, let μ_X and σ_X^2 be the mean and variance of X , let σ_{XY} be the covariance between X and Y (and so forth for the other variables), and let a , b , and c be constants. Equations (2.29) through (2.35) follow from the definitions of the mean, variance, and covariance:

$$E(a + bX + cY) = a + b\mu_X + c\mu_Y, \quad (2.29)$$

$$\text{var}(a + bY) = b^2\sigma_Y^2, \quad (2.30)$$

$$\text{var}(aX + bY) = a^2\sigma_X^2 + 2ab\sigma_{XY} + b^2\sigma_Y^2, \quad (2.31)$$

$$E(Y^2) = \sigma_Y^2 + \mu_Y^2, \quad (2.32)$$

$$\text{cov}(a + bX + cV, Y) = b\sigma_{XY} + c\sigma_{VY}, \quad (2.33)$$

$$E(XY) = \sigma_{XY} + \mu_X\mu_Y, \quad (2.34)$$

$$|\text{corr}(X, Y)| \leq 1 \text{ and } |\sigma_{XY}| \leq \sqrt{\sigma_X^2\sigma_Y^2} \text{ (correlation inequality)}. \quad (2.35)$$

La varianza di Z (2.31)

$$\begin{aligned} \text{Var}(Z) = \text{Var}(aX + bY) &= E \left\{ [(aX + bY) - (a\mu_X + b\mu_Y)]^2 \right\} \\ &= E \left\{ [a(X - \mu_X) + b(Y - \mu_Y)]^2 \right\} \\ &= E \left\{ \begin{aligned} &a^2(X - \mu_X)^2 + b^2(Y - \mu_Y)^2 \\ &+ 2ab(X - \mu_X)(Y - \mu_Y) \end{aligned} \right\} \\ &= a^2 E \left[(X - \mu_X)^2 \right] + b^2 E \left[(Y - \mu_Y)^2 \right] \\ &\quad + 2ab E \left[(X - \mu_X)(Y - \mu_Y) \right] \\ &= a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{Cov}(X, Y) \end{aligned}$$

Distribuzioni di probabilità rilevanti in econometria

- La normale
- La normale standardizzata
- La distribuzione chi-quadrato (o chi-quadro)
- La t di student
- La distribuzione F

La distribuzione normale

- E' la più frequentemente utilizzata
- La funzione di densità di una distribuzione $N(\mu, \sigma^2)$:

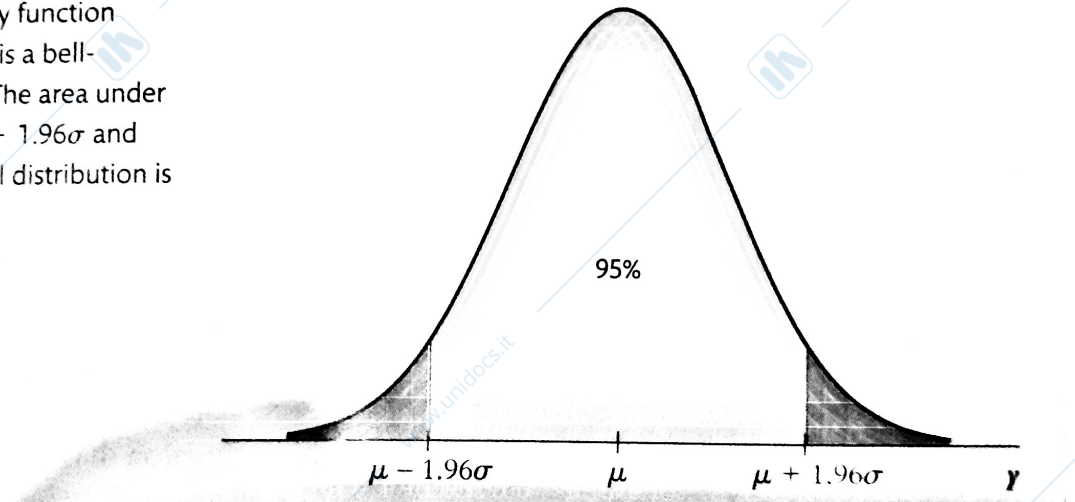
$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right]$$

- Dove μ e σ sono, rispettivamente, il valore atteso e la deviazione standard

La funzione di densità della distribuzione normale

FIGURE 2.5 The Normal Probability Density

The normal probability density function with mean μ and variance σ^2 is a bell-shaped curve, centered at μ . The area under the normal p.d.f. between $\mu - 1.96\sigma$ and $\mu + 1.96\sigma$ is 0.95. The normal distribution is denoted $N(\mu, \sigma^2)$.



La distribuzione normale standardizzata

- E' una distribuzione normale dove μ e σ sono, rispettivamente, pari a 0 e 1. Solitamente definita come distribuzione Z.

- Standardizzazione:
$$Z = \frac{(Y - \mu)}{\sigma}$$

- Si distribuisce quindi come $Z(0,1)$

$$\mu - 1.96\sigma = -1.96$$

$$\mu + 1.96\sigma = +1.96$$

Avremo sulla code 2,5% e nel centro il 95% ;

Il valore centrale è 1,96;

(Costruzione dell'intervallo di confidenza)

La distribuzione chi-quadrato

- La somma di m normali standardizzate e indipendenti tra di loro;
- sarà χ al quadrato con m g.d.l)

Si distribuisce come una chi-quadrato con m gradi di libertà

- La distribuzione chi-quadrato è utilizzata in econometria in alcuni test d'ipotesi

La distribuzione t di student

- Sia Z una distribuzione normale standardizzata e W una distribuzione Chi-Quadrato con m gradi di libertà:
- La distribuzione t di student con m gradi di libertà è la distribuzione della variabile casuale:

$$t = \frac{Z}{\sqrt{W/m}}$$

Ossia è il rapporto tra una $N(0,1)$ e $\sqrt{\frac{\chi^2 m}{m}}$
(N.B. : non è x ma χ)

- E' utilizzata spessissimo in econometria. Ad esempio per valutare la significatività statistica dei parametri in una regressione lineare.
- Quando $m \geq 30$, la t di student è approssimata da una distribuzione normale standardizzata $N(0,1)$. Se m aumenta il denominatore dell t tende a 0.

La distribuzione F

- Sia W una distribuzione Chi-Quadro con m gradi di libertà e V una distribuzione Chi-Quadro con n gradi di libertà:
- La distribuzione F con m e n gradi di libertà, $F_{m,n}$, è la distribuzione della variabile casuale:

$$F_{m,n} = \frac{W/m}{V/n}$$

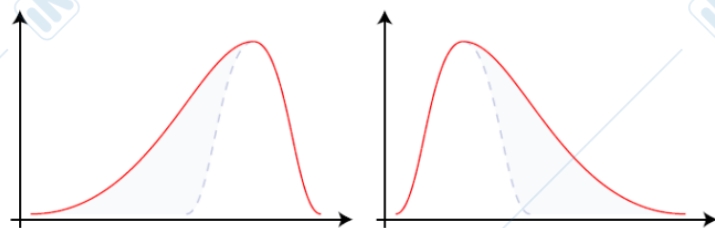
Dove W è $\chi^2 m$ e V è $\chi^2 n$

- La distribuzione F è spesso usata in econometria per i test d'ipotesi congiunti.

La forma di una distribuzione: l'asimmetria

- Formula per l'asimmetria:
$$\frac{E[(Y - \mu_Y)^3]}{\sigma_Y^3}$$

N.B. : Denominatore sempre >0 ;
Dipende quindi dal numeratore:



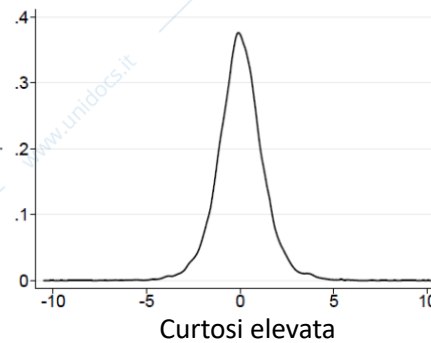
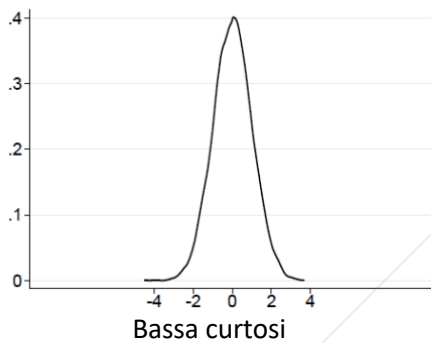
Se $=0$ è una normale

Asimmetria negativa (<0)

Asimmetria positiva (>0)

La forma di una distribuzione: la curtosi

- Formula per la curtosi:
$$\frac{E[(Y - \mu_Y)^4]}{\sigma_Y^4}$$
- Indica lo spessore delle code (dice se c'è quantità più o meno elevata di outliers. Se è bassa non ci sono valori anomali (si concentra attorno alla media) ;
- La curtosi è > 0 o $= 0$



La distribuzione di probabilità su stata

- E' possibile generare delle variabili casuali su stata
- Analizziamo la funzione di densità di probabilità

Lezione 4

Francesco Bloise

francesco.bloise@uniroma1.it

Il campionamento casuale

- Il campionamento casuale consiste nell'estrarre casualmente n elementi da una popolazione e ognuno degli n oggetti ha la stessa probabilità di essere estratto
- La distribuzione di probabilità di ognuno degli n elementi è la stessa essendo stati estratti dalla stessa distribuzione
- Ognuno degli n elementi è indipendente da quella degli altri: conoscere il valore di Y_i non da alcuna informazione sul valore di Y_j per ogni $i \neq j$
- La distribuzione di probabilità di ognuno degli elementi è identica a quella degli altri essendo estratti dalla stessa popolazione di riferimento
- In questo caso si dice che gli n elementi sono variabili casuali i.i.d.

Il campionamento casuale: un esempio

- Sia G il genere di un individuo ($G=1$ se femmina, $G=0$ se uomo)
- Com'è distribuita questa variabile?
- $E(G)=\mu_G=\Pr(G=1)=0.5$
- Ipotizziamo di estrarre casual n elementi dalla popolazione generale
- Ognuno degli G_i elementi è distribuito indipendentemente dagli altri $n-1$ per ogni $i \neq j$
- E se estraessimo gli elementi dagli studenti iscritti alla facoltà di Fisica della Sapienza?

Media e varianza campionaria

- Media campionaria $\bar{Y} = \frac{1}{n} (Y_1 + Y_2 + \dots + Y_n) = \frac{1}{n} \sum_{i=1}^n Y_i$
- Se gli n elementi sono i.i.d, con il valore atteso di Y pari a μ_Y e la varianza di Y pari a σ_Y^2 .

• Allora:
$$E(\bar{Y}) = E\left(\frac{1}{n} \sum_{i=1}^n Y_i\right) = \frac{1}{n} \sum_{i=1}^n E(Y_i) = \frac{1}{n} n E(Y) = \mu_Y$$

• e
$$\begin{aligned} \text{Var}(\bar{Y}) &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n Y_i\right) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(Y_i) + 2 \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \text{Cov}(Y_i, Y_j) \\ &= \frac{1}{n^2} n \text{Var}(Y) + 0 \\ &= \frac{1}{n} \sigma_Y^2 \end{aligned}$$

La distribuzione asintotica

- In econometria si lavora molto spesso con campioni di numerosità elevata, soprattutto nel caso dei microdati (centinaia, ma molto spesso migliaia di elementi estratti casualmente da una stessa popolazione
- La distribuzione asintotica è un'approssimazione distribuzione campionaria di \bar{Y} quando la dimensione campionaria tende a infinito $n \rightarrow \infty$
- Due concetti sono importanti per approssimare la distribuzione della media campionaria in campioni di dimensione elevata
 1. La legge dei grandi numeri
 2. Il teorema del limite centrale

La legge dei grandi numeri

- Se valgono le seguenti due assunzioni
 1. gli n elementi estratti da una popolazione sono i.i.d. con $E(Y_i) = \mu_Y$
 2. La probabilità di avere outlier è ridotta al minimo: $\text{Var}(Y_i) = 0 < \sigma_Y^2 < \infty$

Allora \bar{Y} sarà molta vicina a μ_Y con una probabilità molto alta quando la dimensione campionaria è molto elevata ($n \rightarrow \infty$)

Allora si dice che:

$$\bar{Y} \xrightarrow{p} \mu_Y$$

Cioè la media campionaria converge in probabilità alla media della popolazione.

Il teorema del limite centrale

- Se valgono le seguenti due assunzioni

1. gli n elementi estratti da una popolazione sono i.i.d. con $E(Y_i) = \mu_Y$
2. $\text{Var}(Y_i) = 0 < \sigma_Y^2 < \infty$

Allora, quando $n \rightarrow \infty$:

- I. La distribuzione di \bar{Y} approssima una normale quando $n \rightarrow \infty$
- II. La distribuzione della media campionaria standardizzata approssima una normale standardizzata

La legge dei grandi numeri e il teorema del limite centrale su Stata...

Gli stimatori

- Uno stimatore è una funzione che esprime un valore specifico (una stima) in base ad ogni campione specifico estratto casualmente da una popolazione
- Quindi è una variabile casuale
- La media campionaria è uno stimatore della media della popolazione
- La varianza campionaria è uno stimatore della varianza della popolazione

Tre proprietà rilevanti di uno stimatore

1. La distorsione

$$E(\hat{\mu}_Y) = \mu_Y$$

2. La consistenza

$$\hat{\mu}_Y \xrightarrow{p} \mu_Y$$

3. L'efficienza

$$\text{Var}(\hat{\mu}_Y) < \text{Var}(\bar{\mu}_Y)$$

Esempio:

Supponiamo di essere interessati alla media della popolazione dei laureati in Economia in US (60,000 dollari)

Estraiamo 10 elementi casualmente dalla popolazione dei laureati in Economia

i	W_i
1	47281.92
2	70781.94
3	55174.46
4	49096.05
5	67424.82
6	39252.85
7	78815.33
8	46750.78
9	46587.89
10	25015.71

$$\bar{W} = \frac{1}{10} \sum_{i=1}^{10} W_i = 52618.18$$

$$W_1 = 47281.92$$

$$\tilde{W} = \frac{1}{10} \left(\frac{1}{2} W_1 + \frac{3}{2} W_2 + \dots + \frac{1}{2} W_9 + \frac{3}{2} W_{10} \right) = 49398.82.$$

I tre stimatori sono non distorti:

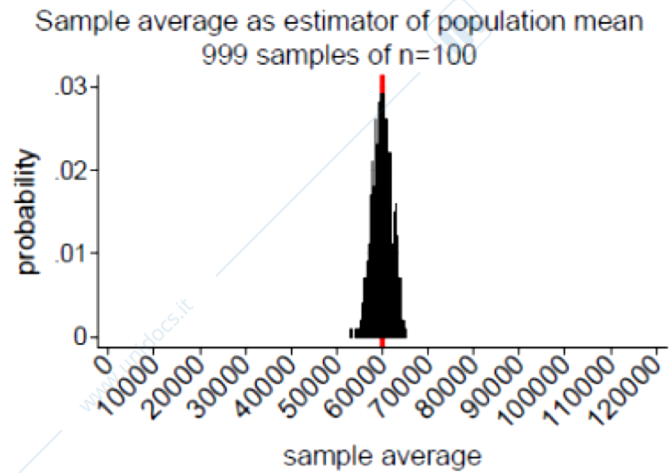
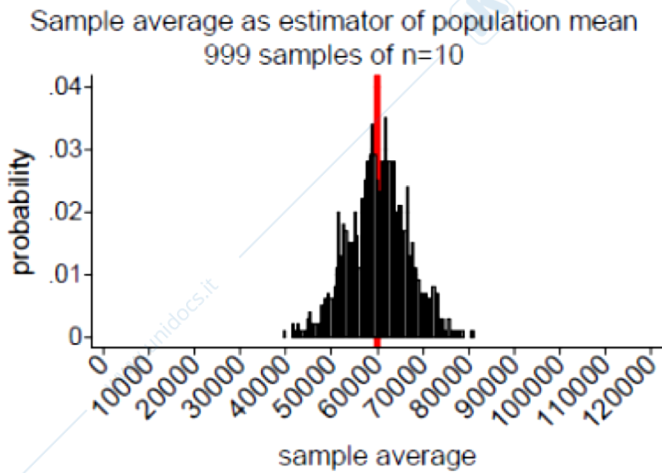
$$E(\bar{Y}) = \mu_Y$$

$$E(Y_1) = E(Y) = \mu_Y$$

$$\begin{aligned} E(\tilde{Y}) &= E\left(\frac{1}{n} \left(\frac{1}{2} Y_1 + \frac{3}{2} Y_2 + \dots + \frac{1}{2} Y_{n-1} + \frac{3}{2} Y_n \right)\right) \\ &= \frac{1}{n} \left(\frac{1}{2} E(Y_1) + \frac{3}{2} E(Y_2) + \dots + \frac{1}{2} E(Y_{n-1}) + \frac{3}{2} E(Y_n) \right) \\ &= \frac{1}{n} \left[\left(\frac{n}{2} \cdot \frac{1}{2} \right) E(Y_i) + \left(\frac{n}{2} \cdot \frac{3}{2} \right) E(Y_i) \right] \\ &= E(Y_i) = \mu_Y \end{aligned}$$

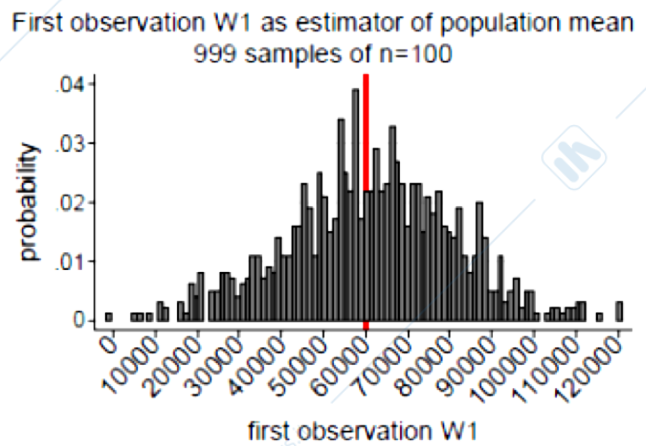
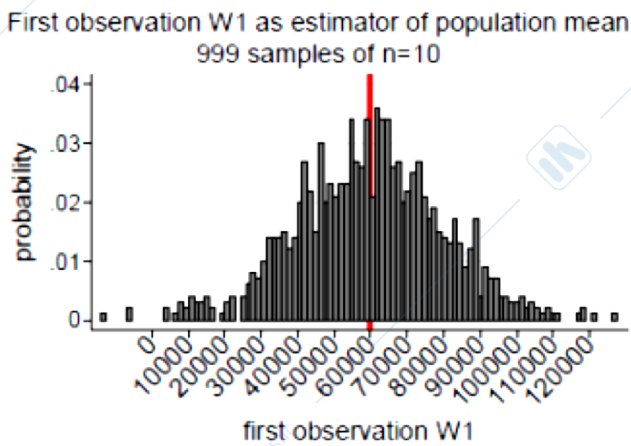
Due stimatori su tre sono consistenti

- Stimatore 1 (\bar{Y}):



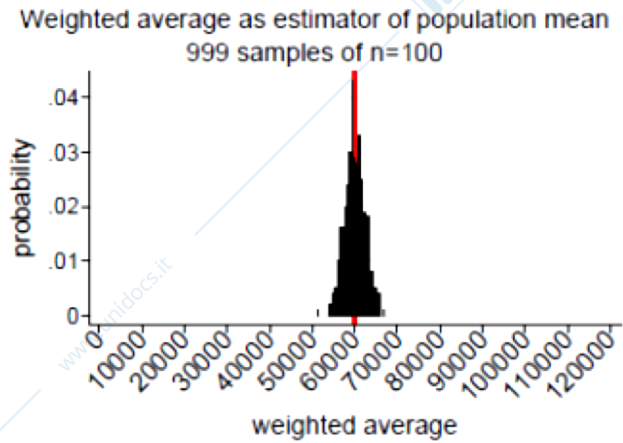
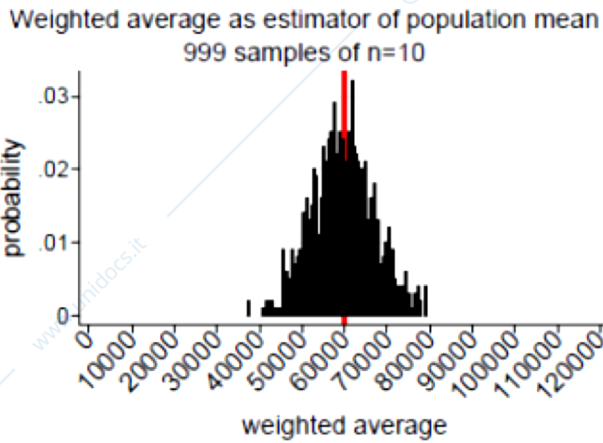
Due stimatori su tre sono consistenti

- Stimatore 2 (Y_1):



Due stimatori su tre sono consistenti

- Stimatore 3 (\tilde{Y}):



Efficienza

- Stimatore 1 (\bar{Y}):

$$\text{Var}(Y_1) = \text{Var}(Y) = \sigma_Y^2$$

- Stimatore 2 (Y_1):

$$\text{Var}(\bar{Y}) = \frac{1}{n} \sigma_Y^2$$

- Stimatore 3 (\tilde{Y}):

$$\text{Var}(\tilde{Y}) = 1.25 \frac{1}{n} \sigma_Y^2$$

Lo stimatore BLUE

- Uno stimatore non distorto che ha la varianza più bassa (più efficiente) rispetto a tutti gli altri stimatori alternativi è detto:

«Best Linear Unbiased Estimator»

Lezione 5

Francesco Bloise

francesco.bloise@uniroma1.it

Test di ipotesi

- Quando si hanno dati campionari, attraverso gli stimatori, ottengo stime sui parametri della popolazione-> posso controllare che il valore stimato sia esattamente pari a un valore specifico tramite l'ipotesi nulla e l'ipotesi alternativa;
- Ogni qual volta che lavoriamo su variabili casuali e lavoriamo su un campione della popolazione noi abbiamo delle stime della popolazione. Dobbiamo verificare che sia vera. Per fare ciò possiamo formulare delle ipotesi specifiche
- Analizziamo il caso del **test d'ipotesi per la media** della popolazione

Esempio: Il salario mensile medio dei lavoratori laureati in Economia è di 1800 euro? ipotesi

Cioè, si testa che il valore di μ della variabile casuale y sia esattamente pari ad uno specifico valore.

Per condurre il test possiamo utilizzare un'ipotesi nulla e un'ipotesi alternativa:

- Ipotesi nulla: $H_0: E(Y) = \mu_{y,0}$ es. ipotizziamo la media della pop. pari a $\mu_y = 1800$
- Ipotesi alternativa: $H_1: E(Y) \neq \mu_{y,0}$ (salario medio mensile diverso da 1800)

Se abbiamo estratto n elementi (ossia un campione della popolazione), sappiamo che possiamo utilizzare la media del campione (la stima della media della popolazione) e potremo dire che la media campionaria è quella della popolazione. Ipotizziamo l'ipotesi nulla e l'alternativa.

Test d'ipotesi per la media campionaria

- Supponiamo di aver estratto casualmente n elementi dalla popolazione e abbiamo calcolato la media campionaria \bar{y}
- La media campionaria può essere diversa dal valore previsto all'interno dell'ipotesi nulla per due motivi diversi:
 1. La media della popolazione è diversa dal valore ipotizzato dall'ipotesi nulla (ipotesi nulla non è vera)
 2. A causa del campionamento casuale si ha che $\bar{y} \neq \mu_y = \mu_{y,0}$ (ipotesi nulla è vera)

Per quantificare la seconda motivazione si utilizza il p-value (probabilità compresa tra 0 e 1) -> motivo della discrepanza; questo dice la prob di osservare una differenza tra la media campionaria e il valore dell'ipotesi nulla maggiore rispetto a quella osservata dato che l'ipotesi nulla H_0 sia vera.

Il p-value

- Il p-value è la probabilità di osservare una differenza tra la media campionaria ottenuta e il valore dell'ipotesi nulla (1800) maggiore rispetto a quella osservata ipotizzando che l'ipotesi nulla sia vera. Noi ipotizziamo che l'ipotesi nulla sia vera fissando un valore specifico per l'ipotesi nulla.
- Formula del p-value: è la probabilità che estraendo un altro campione otteniamo la differenza tra l'ipotesi vera e l'ipotesi nulla. È la probabilità che estraendo un altro campione la differenza tra la media campionaria e l'ipotesi nulla è maggiore.

$$p - value = Pr_{H_0} \left[|\bar{Y} - \mu_{Y,0}| > |\bar{Y}^{act} - \mu_{Y,0}| \right]$$

Per calcolare il p-value è necessario conoscere la distribuzione di \bar{Y} . (è necessario conoscere la distribuzione media di y per calcolare che la differenza sia maggiore).

(se dal campione ottengo $\bar{Y}^{act} = \mu_{y,0}$ allora la prob di osservare una diff maggiore è 1)

Più il p value è alto (tende a 1) più il valore consiste nell'ipotesi nulla. Significa che, avendo es. alfa=5%, estraendo un altro campione, nel 95% dei casi, il parametro risiede in $\bar{Y} \pm 1,96 * SE(\bar{Y})$

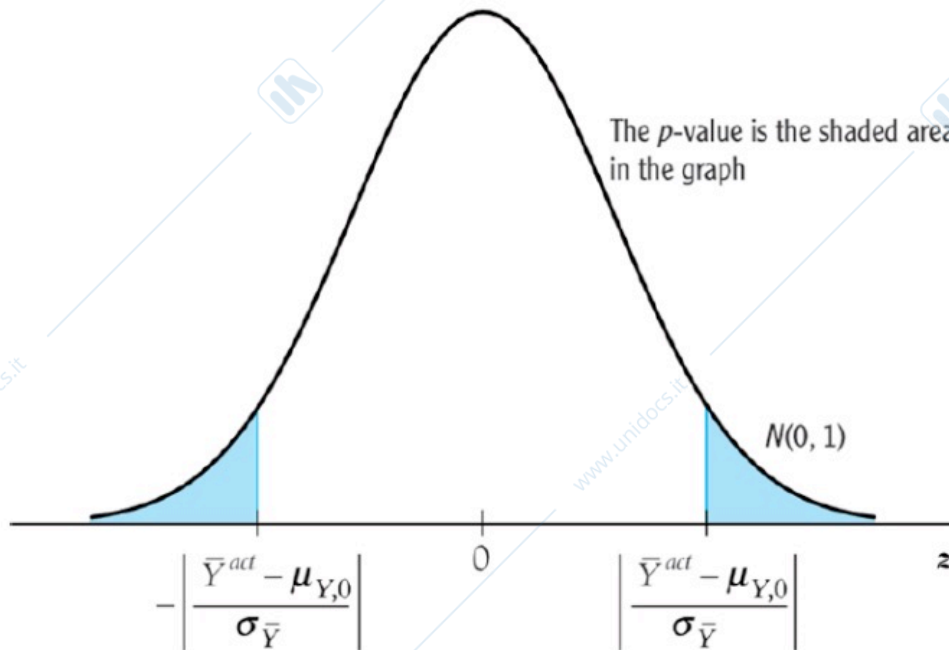
- In generale, se il p-value è alto, allora il valore osservato è consistente con l'ipotesi nulla;
- Il p value è utile per la significatività statistica dei parametri della regressione. Se ipotizzo che $\bar{Y} \sim N(0,1)$ vedo le code della distribuzione per il p value (vedi grafico)

Il caso della normale standardizzata

Più è alta la differenza tra \bar{Y}^{act} e $\mu_{y,0}$ e più le code sono ridotte;

Esempio estremo $\bar{Y}^{act} = \mu_{y,0} \Rightarrow$ il numeratore è 0 (ho tutta l'area sotto la curva allora la prob è 1, ossia il p-value è 1.)

Se estraggo un altro campione ottengo sicuramente una differenza maggiore tra \bar{Y}^{act} e $\mu_{y,0}$.



La varianza e la deviazione standard campionaria

- Se la distribuzione non è nota, si devono stimare i parametri rilevanti. Ricordate l'esempio della media campionaria.

- Nel caso della varianza campionaria:

$$s_Y^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$$

Lo stimatore della varianza campionaria sarebbe distorto dalla varianza reale se utilizzassimo $1/n$, mentre con il -1 riusciamo a correggere tale distorsione.

(nota dal libro : La divisione per $n-1$ nella var campionaria invece che per n è detta correzione per i gradi di libertà: stimare la media consuma una parte dell'informazione contenuta nei dati—consuma cioè un “grado di libertà”—lasciando solo $n - 1$ gradi di libertà.)

- E la deviazione standard è

$$s_Y = \sqrt{s_Y^2}$$

La statistica T

- La media campionaria standardizzata è utilizzata per i test di ipotesi: la statistica test è utilizzata per la verifica di ipotesi.

$$\text{Statistica T: } t = \frac{\bar{Y} - \mu_{Y,0}}{SE(\bar{Y})}$$

Dove SE è standard error di \bar{Y} (dove \bar{Y} è la media campionaria). Questo è uno stimatore della deviazione standard di \bar{Y} . (dal libro: Quando Y_1, \dots, Y_n sono iid $SE(\bar{Y}) = \hat{\sigma}_{\bar{Y}} = s_Y / \sqrt{n}$).

- Se non conosco la distribuzione esatta di \bar{Y} devo usare uno stimatore per la deviazione standard: lo stimatore è lo standard error.
- Per n abbastanza grande la T si distribuisce come una $N(0,1)$ ($n > 30$, ma più è alto è migliore è l'approssimazione. Quando coincideranno sarà una perfetta approssimazione); (dal libro: Quando n è grande, la varianza campionaria si approssima alla varianza della popolazione, perciò per la statistica t la distribuzione $(\bar{Y} - \mu_{Y,0}) / \sigma_{\bar{Y}}^2$ per il TLC si approssima a una normale standard per n grande; quindi sotto H_0 , t si distribuisce come una normale. Il valore della statistica t effettivamente calcolato sarà $t^{act} = \frac{\bar{Y}^{act} - \mu_{Y,0}}{SE(\bar{Y})}$ e si conseguenza per n grande il p value sarà calcolato come $\text{valore-}p = 2\Phi(-|t^{act}|)$).

Errori nei test d'ipotesi

Due tipi di errori possono essere commessi nell'ambito dei test di ipotesi:

1. Errore del primo tipo: rigettare l'ipotesi nulla quando è vera. Se ipotizziamo il salario medio degli studenti di economia è 1800 e sappiamo che è vera ma la rigettiamo ma è vera.
2. Errore del secondo tipo: non rigettare l'ipotesi nulla quando è vera (accetto H_0 quando è falsa). Non rigettare l'ipotesi nulla quando è falsa. Es ipotesi 1800 di salario è falsa e non la rigettiamo.

come lo decidiamo se rigettare l'ipotesi nulla? Attraverso il valore soglia fissato (alfa=0,05; ma possiamo ritenerlo anche pari ad altri valori.). Poi ne calcoliamo il p value e statistica t per decidere se rigettare l'ipotesi. (per i diversi livelli di significatività troviamo i livelli di alfa nella tabella in fondo al libro).

Di solito si fissa un livello di significatività riferito all'errore del primo tipo: rigettare l'ipotesi nulla quando è vera. Spesso significatività al 5% ($\alpha = 0.05$)

Nell'ambito della normale standardizzata, il 5% corrisponde all'area nelle code per valori della statistica t superiori a 1.96 o inferiori a -1.96. (a seconda se ottengo un valore maggiore, minore o tra le due decidiamo se rigettare l'ipotesi.)

Si rigetta l'ipotesi nulla quando il p -value < 0.05 . Cioè si rigetta l'ipotesi nulla quando $|t^{act}| > 1.96$ (rigettiamo perché siamo andati oltre il valore soglia;) (dove t alla act sono i valori della statistica t).

Come decidere se rifiutare o meno l'ipotesi (scritto meglio):

Supponiamo di aver deciso che l'ipotesi nulla venga rifiutata se il valore-p è inferiore a 5%. Siccome l'area nelle code della distribuzione normale al di fuori dell'intervallo $\pm 1,96$ è 5%, si ottiene una semplice regola

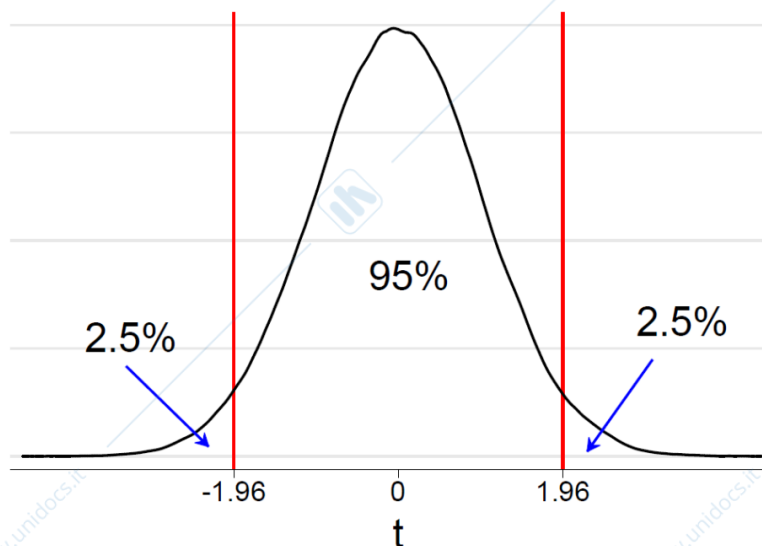
$$\text{rifiutare } H_0 \text{ se } |t^{act}| > 1,96,$$

Se il test rifiuta con un livello di significatività del 5%, si dice che la media della popolazione μ_Y è statisticamente diversa da $\mu_{Y,0}$ al livello di significatività 5%.

La probabilità di rifiutare erroneamente l'ipotesi nulla (rifiutare quando in realtà essa è vera) è 5%.

La distribuzione della T in campioni di grandi dimensioni

quando n è grande si approssima alla normale standard: possiamo ottenere il p value a seconda del valore t . quando la differenza star il valore Queste linee tendono a collapsare. Se andiamo oltre la soglia stiamo escludendo dalle code 0,05 da destra e da sinistra. Se siamo nel centro non stiamo rigettando l'ipotesi nulla.



I diversi step nei test di ipotesi

Definisco l'ipotesi nulla (è un'ipotesi sul valore della media della popolazione). Es. salario 1800

Definisco l'ipotesi alternativa. $H_0 : E(Y) = \mu_{Y,0}$ $H_1 : E(Y) \neq \mu_{Y,0}$

Es. salario maggiore o minore di 1800.

1 STEP: Si calcola la media campionaria \bar{y}

2 STEP: si calcola lo standard error di \bar{y}

3 STEP: si calcola la statistica T

$$t^{act} = \frac{\bar{Y} - \mu_{Y,0}}{SE(\bar{Y})}$$

Quando il valore della statistica t è basso siamo molto vicini alla possibilità di dover rigettare l'ipotesi

- Si rigetta l'ipotesi nulla al 5% di significatività se il p-value < 0.05 ovvero se $|t^{act}| > 1.96$

Gli intervalli di confidenza della media della popolazione

Un intervallo di confidenza è un intervallo che contiene con una certa probabilità il valore reale della media della popolazione. Possiamo avere un intervallo di valori all'interno del quale al 95% troveremo il vero valore della media della popolazione.

Per esempio, un intervallo di confidenza al 95% contiene al 95% il vero valore della statistica d'interesse (intervallo 1-alfa).

$$\left\{ \bar{Y} - 1.96 \cdot SE(\bar{Y}) \quad , \quad \bar{Y} + 1.96 \cdot SE(\bar{Y}) \right\}$$

- $1,96 \cdot s.error$ y campionario. Otteniamo così due soglie. Se il valore ottenuto è maggiore o minore delle due soglie dobbiamo rigettare l'ipotesi.
- Intuizione: un valore di $\mu_{y,0}$ minore o maggiore delle soglie è rigettato considerando un livello di significatività del 5%

Gli intervalli di confidenza associati a diversi livelli di significatività

$$95\% \text{ confidence interval for } \mu_Y = \left\{ \bar{Y} \pm 1.96 \cdot SE(\bar{Y}) \right\}$$

$$90\% \text{ confidence interval for } \mu_Y = \left\{ \bar{Y} \pm 1.64 \cdot SE(\bar{Y}) \right\}$$

$$99\% \text{ confidence interval for } \mu_Y = \left\{ \bar{Y} \pm 2.58 \cdot SE(\bar{Y}) \right\}$$

Il modello di regressione lineare

- Vogliamo andare a vedere la relazione tra due variabili attraverso il modello di regressione

$$Y_i = B_0 + B_1 X_1 + E_i$$

- Y_i = variabile dipendente
- B_0 = termine costante
- B_1 = coefficiente angolare
- E_i = termine di errore
- X_1 = regressore

Come si stimano i parametri b_0 e b_1 ? In base alla loro stima otteniamo diversi punti in analisi: vedi alide modello di reg li (intuizione); noi per la stima cercheremo di minimizzare il quadrato -> metodo dei minimi quadrati; la somma del quadrato delle diverse distanze deve essere minima: cerchiamo di minimizzare in questo modo il termine di errore

Il modello di regressione lineare (2)

- Come diventa il modello quando abbiamo solo la costante?

$$Y_i = B_0 + E_i \quad \text{dove } E_i = Y_i - B_0$$

Minizziamo il quadrato della somma di Y_i

Sommatoria $(E_i)^2 \rightarrow$ dobbiamo minimizzare allora facciamo la derivate prima=0 della sommatoria da 1 a n di $(y_i - b_0)^2$.

$$D' = 2(\text{sommatoria da 1 a n})(y_i - b_0) = 0;$$

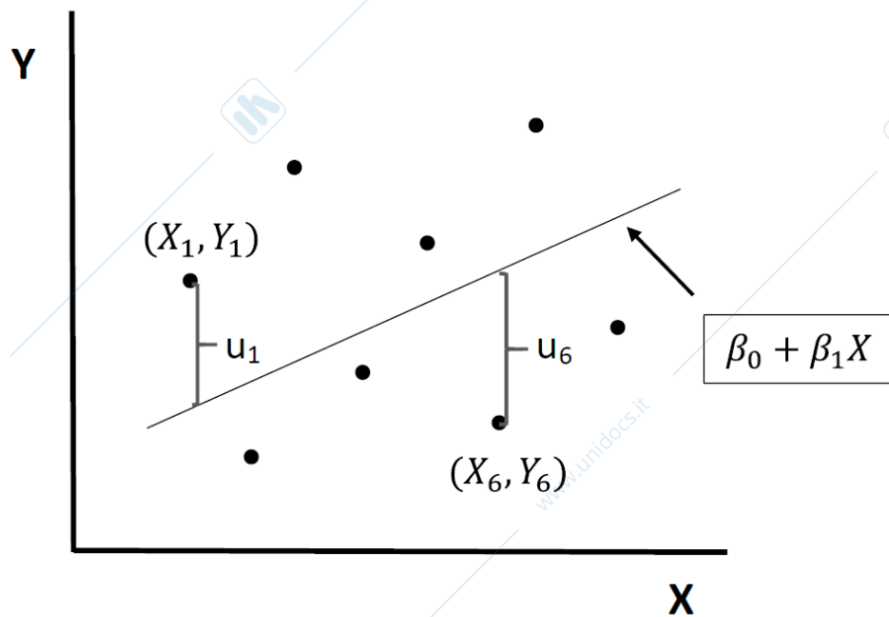
Svolgendo i calcoli \rightarrow (sommatoria di $y_i - b_0$) = 0/2;

Sommatoria di $y_i - n B_0 = 0$ perchè b_0 è una costante quindi n volte ripetuto b_0 ;

Il modello di regressione lineare (2)

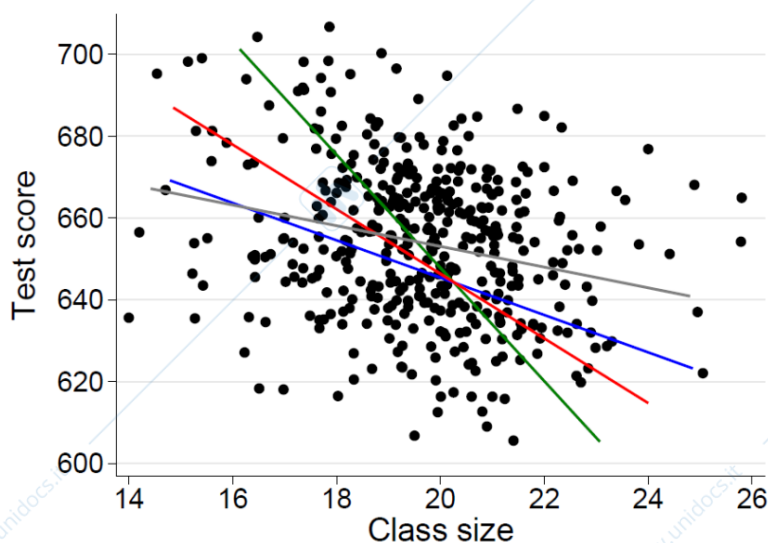
$$\begin{aligned}
 Y_i &= B_0 + \varepsilon_i & \varepsilon_i &= Y_i - B_0 \\
 \sum_i (\varepsilon_i)^2 &= \sum_{i=1}^n (Y_i - B_0)^2 \\
 \sum_{i=1}^n (Y_i - B_0) &= 0 & \sum_{i=1}^n Y_i - n B_0 &= 0 \\
 n B_0 &= \sum_{i=1}^n Y_i \\
 \frac{1}{n} B_0 &= \frac{1}{n} \sum Y_i & B_0 &= \bar{Y}_i
 \end{aligned}$$

Il modello di regressione lineare (intuizione)



Come trovare la retta che si adatta meglio ai dati?

- Troviamo la nuvola di punti avendo la problematica risultati della classe e dimensione della classe.



Stima su stata

- 1 Generiamo 1000 osservazioni -> set obs 1000
- 2 generiamo una variabile casuale normale (media, deviazione standard) -> gen y=normal(18000, 5000)
- 3 facciamo statistiche descrittive per vedere i parametri di questa variabile -> sum y
- Ci comparirà la tabella con media, d.standard, min e max;
- Regrediamo la variabile : stiamiamo la regressione con il metodo dei minimi quadrati solo sulla costante : ottiamo la media di y: il valore di beta0 è il valore «coef» 18116,12 -> comando : reg y

Rigetto l'ipotesi nulla perché $p > t$ è 0.00 quindi minore di 0,05

```
. sum y
```

Variable	Obs	Mean	Std. Dev.	Min	Max
y	1,000	18116.12	4809.176	3457.429	32239.44

```
. reg y
```

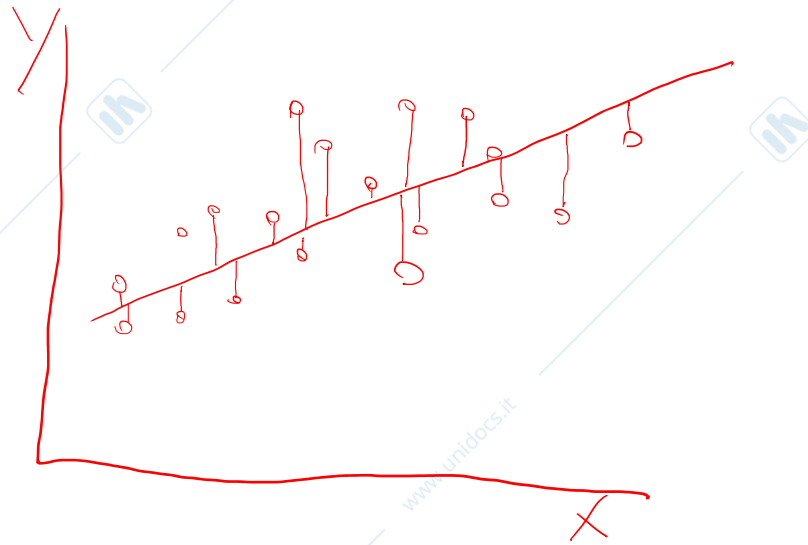
Source	SS	df	MS	Number of obs	=	1,000
Model	0	0	.	F(0, 999)	=	0.00
Residual	2.3105e+10	999	23128170.2	Prob > F	=	.
Total	2.3105e+10	999	23128170.2	R-squared	=	0.0000
				Adj R-squared	=	0.0000
				Root MSE	=	4809.2

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
_cons	18116.12	152.0795	119.12	0.000	17817.69 18414.56

- Non c'è relazione tra le due variabili quando rigetto l'ipotesi;

METODO DEI MINIMI QUADRATI ORDINARI

La retta di regressione rappresenta la relazione tra le due variabili. I punti sono dati simulati e le linee dai punti alla retta sono le distanze. Vogliamo studiare la relazione tra le 2 variabili: vediamo la minima distanza. Metodo dei minimi quadrati (quadrati perché le distanze sono sia negative che positive).



MODELLO CON SOLA COSTANTE

$$Y_i = \beta_0 + \varepsilon_i \quad \rightarrow \quad \varepsilon_i = Y_i - \beta_0$$

$$\varepsilon_i^2 = (Y_i - \beta_0)^2$$

MINIMIZZIAMO LA SOMMA DEI TERMINI DI ERRORE AL QUADRATO

$$\sum_{i=1}^n (Y_i - \beta_0)^2$$

→ DERIVIAMO AL SPETTO A β_0 E PONIAMO UGUALE A 0

$$-2 \sum_{i=1}^n (Y_i - \hat{\beta}_0) = 0$$

$$\sum_{i=1}^n Y_i - \sum_{i=1}^n \hat{\beta}_0 = 0$$

$$\sum_{i=1}^n Y_i - n \hat{\beta}_0 = 0$$

$$\frac{1}{n} \sum_{i=1}^n Y_i - \frac{1}{n} n \hat{\beta}_0 = 0$$

$$\bar{Y} - \hat{\beta}_0 = 0$$

$$\hat{\beta}_0 = \bar{Y}$$

MODELLO CON 1 SOLO REGRESSORE

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

$$\epsilon_i = Y_i - \beta_0 - \beta_1 X_i$$

$$\epsilon_i^2 = (Y_i - \beta_0 - \beta_1 X_i)^2$$

MINIMIZZIAMO LA SOMMA DI ϵ_i^2 DERIVATA RISPETTO A $\beta_0 \rightarrow = 0$

$$\sum_{i=1}^n (Y_i - \hat{\beta}_0 - \beta_1 X_i)^2 \rightarrow -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \beta_1 X_i) = 0$$

$$\sum_{i=1}^n Y_i - n \hat{\beta}_0 - \beta_1 \sum_{i=1}^n X_i = 0 \rightarrow \frac{1}{n} \sum_{i=1}^n Y_i - \frac{1}{n} \sum_{i=1}^n X_i = \hat{\beta}_0$$

$$\bar{Y} - \hat{\beta}_0 - \beta_1 \bar{X} = 0 \rightarrow -\hat{\beta}_0 = \beta_1 \bar{X} - \bar{Y}$$

$$\hat{\beta}_0 = \bar{Y} - \beta_1 \bar{X}$$

DERIVATA RISPETTO A $\beta_1 \rightarrow = 0$

$$\sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2 \rightarrow -2 X_i \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i) = 0$$

DIVIDENDO PER $-2X_i$ ENTRAMBI I MEMBRI:

$$\sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i) = 0$$

SOSTITUIAMO β_0 CON $\hat{\beta}_0$ NELLA PRECEDENTE EQUAZIONE

$$\sum_{i=1}^n (Y_i - \bar{Y} + \beta_1 \bar{X} - \beta_1 X_i) = 0$$

$$\sum_{i=1}^n \left[(Y_i - \bar{Y}) + \hat{\beta}_1 (\bar{X} - X_i) \right] = 0$$

$$\sum_{i=1}^n (Y_i - \bar{Y}) + \hat{\beta}_1 \sum_{i=1}^n (\bar{X} - X_i) = 0 \rightarrow \hat{\beta}_1 \sum_{i=1}^n (\bar{X} - X_i) = - \sum_{i=1}^n (Y_i - \bar{Y})$$

$$\hat{\beta}_1 \sum_{i=1}^n (X_i - \bar{X}) = \sum_{i=1}^n (Y_i - \bar{Y}) \rightarrow \hat{\beta}_1 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})}$$

DA LLA SLIDE PRECEDENTE

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})} = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})} = \frac{1}{n-1} \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})}$$

$\frac{S_{xy}}{S_{xx}}$ → VARIAZA CAMPIONARIA
 $\frac{S_{xy}}{S_{xx}}$ → COVARIANZA CAMPIONARIA

Posso fare il test d'ipotesi sui parametri:

1) $H_0: \beta_{2,0} = 0$
 $H_2: \beta_{2,1} \neq 0$

la statistica t sarà:

$$t_{\beta_2} = \frac{\hat{\beta}_2 - \beta_{2,0}}{SE(\hat{\beta}_2)} = \frac{\hat{\beta}_2}{SE(\hat{\beta}_2)}$$

2) $H_0: \beta_{0,0} = 0$
 $H_2: \beta_{0,1} \neq 0$

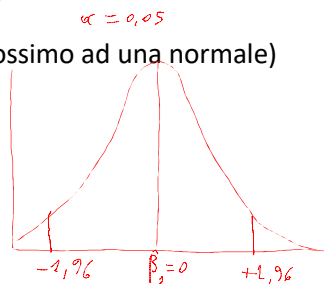
$$t_{\beta_0} = \frac{\hat{\beta}_0 - \beta_{0,0}}{SE(\hat{\beta}_0)} = \frac{\hat{\beta}_0}{SE(\hat{\beta}_0)} = \frac{\hat{\beta}_0}{SE(\hat{\beta}_0)}$$

Oltre un certo valore soglia l'ipotesi nulla sarà rigettata

$\alpha = 0,05$ $t \pm 1,96$

(B è significativamente diverso da 0 al livello di significatività scelto (es. 5% → $\alpha = 0,05$))

Al 5% il valore critico è $\pm 1,96$ in quanto approssimativo ad una normale)



R^2 dice la bontà di adattamento del modello; è detto coefficiente di determinazione; è una stima dell'accuratezza della predizione. La sua formula è:

$$R^2 = \frac{\frac{1}{n-2} \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\frac{1}{n-2} \sum_{i=1}^n (Y_i - \bar{Y})^2} = \frac{ESS}{TSS} \quad \text{oppure} \quad 1 - \frac{RSS}{TSS}$$

Al numeratore abbiamo la varianza predetta e al denominatore la varianza campionaria.

$0 \leq R^2 \leq 1$ SE=0 la X non spiega nulla della Y (la varianza spiegata è 0).

Non è detto che un valore di R quadro alto sia ottimale per previsioni future (su altri dati, fuori dal campione).

$\mathcal{D}_0 \sqrt{B}$

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

$$0 \leq R^2 \leq 1$$

COEFFICIENTE
DI
DETERMINAZIONE

È il modello stimato della y grazie alla stima fatta sui parametri β_0 e β_1

MISURA DELL'ACCURATEZZA
DELLA PREDIZIONE

Lezione 7, 8, 9

23/03

Francesco Bloise

francesco.bloise@uniroma1.it

Ricordiamo che...

Nei test d'ipotesi dobbiamo formulare un ipotesi nulla e un ipotesi alternativa.

Per i coefficienti della regressione lineare si ha

L'ipotesi nulla la possiamo formulare per il parametro b_1 o per la costante b_0 ; l'ipotesi nulla che facciamo è di b_1 sotto l'ipotesi nulla $=0$ oppure è b_0 sotto l'ipotesi nulla $=0$

Stiamo testando di quanto eventualmente il nostro coefficiente stimato si allontana dallo zero dell'ipotesi nulla. Sulla base di questo possiamo formulare delle ipotesi sulla significatività dei parametri stimati secondo un livello di significatività α che stabiliamo ex ante

Essendo l'ipotesi nulla per i coeff b_1 e $b_0=0$ l'ipotesi alternativa sarà che i due parametri sotto l'ipotesi alternativa sono diversi da

1) Ipotesi nulla: $H_0: \beta_1 = \beta_{1,0} = 0$

Ipotesi alternativa: $H_1: \beta_1 = \beta_{1,1} \neq 0$

2) Ipotesi nulla: $H_0: \beta_0 = \beta_{0,0} = 0$

Ipotesi alternativa: $H_1: \beta_0 = \beta_{0,1} \neq 0$

Nel caso della regressione lineare:

$$t_{\hat{\beta}_1} = \frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)}$$

Come si stima nel caso della regressione lineare la statistica t per il parametro b0 e b1 ed i p-value? La statistica t si costruisce (abbiamo visto nel caso della media campionaria) come differenza tra il valore stimato meno il valore dell'ipotesi nulla, tutto diviso lo SE. (stimatore della deviazione standard del parametro stimato). Nel caso della media campionaria noi avevamo una stima della media campionaria meno l'h.nulla che poteva variare di caso in caso (nel nostro caso era 1200) tutto diviso lo standard error. Nel caso di b1, la statistica t per b1 = beta 1 stimato meno il b1 sotto l'ipotesi nulla formulato/ s.e stimato. Essendo l'ipotesi nulla b1=0 nell'ipotesi formulata all'inizio avremo che la statistica t sarà b1 stimato/s.e b1 stimato;

$$t_{\hat{\beta}_0} = \frac{\hat{\beta}_0 - \beta_{0,0}}{SE(\hat{\beta}_0)} = \frac{\hat{\beta}_0 - 0}{SE(\hat{\beta}_0)} = \frac{\hat{\beta}_0}{SE(\hat{\beta}_0)}$$

In modo analogo lavoriamo sul b0: la statistica t sarà b0 stimato meno ipotesi nulla di b0 (b0 teorico=0) / s.e di b0

$$P\text{-value} = Pr_{H_0} [|\hat{\beta}_1 - \beta_{1,0}| > |\hat{\beta}_1^{act} - \beta_{1,0}|]$$

Ricordiamo che il p-value si calcola ipotizzando che l'ipotesi nulla sia vera. Calcoliamo allora che la probabilità che il beta1 stimato meno il b1 teorico siano maggiori rispetto al b1 del campione su cui lavoriamo meno il beta teorico. Essendo l'ipotesi nulla b1 pari a 0, se il nostro beta stimato nel nostro campione è =0 o prossima allo zero, allora la prob che un altro b1 stimato sia maggiore del nostro beta 1 teorico è pari a 1. (perché sicuramente sarà maggiore dello 0 visto data la distribuzione di b1.

Se alla prima formula del p-value e dividiamo per lo s.e. di b1 stimato ottengo la formula della statistica t nella formula del p-value. Quindi riformulo la formula del b value come la probabilità sotto l'ipotesi nulla che la statistica t di b1 stimato in un altro campione sia maggiore rispetto alla statistica t su uno specifico campione su cui stiamo lavorando. (stesso discorso lo possiamo fare anche su b0 in modo analogo sostituendo dove c'è b1, b0).

$$P\text{-value} = Pr_{H_0} \left[\left| \frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)} \right| > \left| \frac{\hat{\beta}_1^{act} - \beta_{1,0}}{SE(\hat{\beta}_1)} \right| \right] = Pr_{H_0} \left[|t_{\hat{\beta}_1}| > |t_{\hat{\beta}_1}^{act}| \right]$$

A partire da una statistica t che ha una formulazione nota, possiamo calcolare il p-value, data la distribuzione di b1 stimato (perché per calcolare la probabilità dobbiamo per forza avere una distribuzione di b1 stimato in mente. Altrimenti quando otteniamo che il beta 1 stimato =0 non sappiamo effettivamente quale sia effettivamente la probabilità per ogni valore di b1 che sia maggiore rispetto a quello ottenuto nel nostro specifico campione). Possiamo quindi lavorare sulla statistica t ipotizzando che il b1 si approssimi a una normale standardizzata (che è la distribuzione della t quando abbiamo un campione di grande numerosità). Possiamo utilizzare quindi i valori della distribuzione della normale standardizzata come valori soglia dato il livello di significatività alfa prefissato prima di iniziare l'analisi.

Nel caso della regressione lineare:

• Avendo s.e di b_1 e il valore della statistica t possiamo ottenere gli intervalli di confidenza per b_1 stimato. L'intervallo di confidenza per b_1 stimato ha dei valori entro i quali il b_1 teorico deve risiedere (2 valori soglia). Calcoliamo l'intervallo sommando e sottraendo il valore della statistica t coincidente con il livello di significatività che ci siamo prefissati.

• Intervallo di confidenza di $\hat{\beta}_1 = \hat{\beta}_1 \pm [t_{1-\alpha} * SE(\hat{\beta}_1)]$

• Avendo s.e e statistica t possiamo ottenere gli intervalli di confidenza:

• Esempi:

$$90\%: \hat{\beta}_1 \pm [1.64 * SE(\hat{\beta}_1)]$$

$B1 \pm$ (valore critico della statistica t di $\alpha(10\%) * s.e.$ stimato)

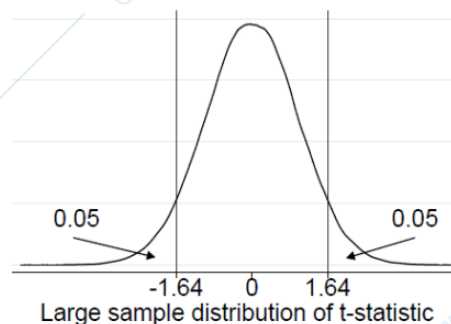
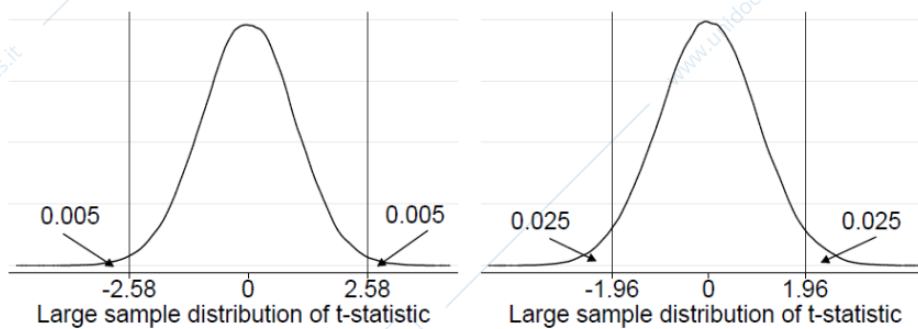
Statistica t ($\alpha = 1 - \alpha$)

$B1$ al 90% stà in quel valore stimato.

$$95\%: \hat{\beta}_1 \pm [1.96 * SE(\hat{\beta}_1)]$$

$$99\%: \hat{\beta}_1 \pm [2.58 * SE(\hat{\beta}_1)]$$

I valori critici della statistica T dipendono da α



- Nel caso a sinistra abbiamo 1% : la probabilità calcolata come l'area che si trova tra -2,58 e 2,58 (l'area totale meno le code) prende il 99% dei possibili valori. Il beta è diverso da 0.

I valori critici della statistica T: valori soglia della statistica t

quando i gradi di libertà tendono a infinito avremo un valore soglia di 1,96, con il 10% è 1,64%. Siamo sempre nel caso 2 sided.

Degrees of Freedom	Significance Level				
	20% (2-Sided) 10% (1-Sided)	10% (2-Sided) 5% (1-Sided)	5% (2-Sided) 2.5% (1-Sided)	2% (2-Sided) 1% (1-Sided)	1% (2-Sided) 0.5% (1-Sided)
1	3.08	6.31	12.71	31.82	63.66
2	1.89	2.92	4.30	6.96	9.92
3	1.64	2.35	3.18	4.84	5.84
4	1.53	2.13	2.78	3.75	4.60
5	1.48	2.02	2.57	3.36	4.03
6	1.44	1.94	2.45	3.14	3.71
7	1.41	1.89	2.36	3.00	3.50
8	1.40	1.86	2.31	2.90	3.36
9	1.38	1.83	2.26	2.82	3.25
10	1.37	1.81	2.23	2.76	3.17
11	1.36	1.80	2.20	2.72	3.11
12	1.36	1.78	2.18	2.68	3.05
13	1.35	1.77	2.16	2.65	3.01
14	1.35	1.76	2.14	2.62	2.98
15	1.34	1.75	2.13	2.60	2.95
16	1.34	1.75	2.12	2.58	2.92
17	1.33	1.74	2.11	2.57	2.90
18	1.33	1.73	2.10	2.55	2.88
19	1.33	1.73	2.09	2.54	2.86
20	1.33	1.72	2.09	2.53	2.85
21	1.32	1.72	2.08	2.52	2.83
22	1.32	1.72	2.07	2.51	2.82
23	1.32	1.71	2.07	2.50	2.81
24	1.32	1.71	2.06	2.49	2.80
25	1.32	1.71	2.06	2.49	2.79
26	1.32	1.71	2.06	2.48	2.78
27	1.31	1.70	2.05	2.47	2.77
28	1.31	1.70	2.05	2.47	2.76
29	1.31	1.70	2.05	2.46	2.76
30	1.31	1.70	2.04	2.46	2.75
60	1.30	1.67	2.00	2.39	2.66
90	1.29	1.66	1.99	2.37	2.63
120	1.29	1.66	1.98	2.36	2.62
∞	1.28	1.64	1.96	2.33	2.58

Values are shown for the critical values for two-sided (\neq) and one-sided ($>$) alternative hypotheses. The critical value for the one-sided ($<$) test is the negative of the one-sided ($>$) critical value shown in the table. For example, 2.13 is the critical value for a two-sided test with a significance level of 5% using the Student t distribution with 15 degrees of freedom.

La non distorsione del beta (1)

Interpretazione del coefficiente beta 1.

Dalla lezione precedente ricordiamo che il b1 stimato lo abbiamo ottenuto da (covarianza campionaria/varianza campionaria).

Il valore della cov/ var è un'associazione tra il valore delle variabili x e y.

Non possiamo però interpretare x e y come casualmente relazionali (si muovono insieme le variabili ma una non è causa dell'altra). Non possiamo quindi dire tramite b1 se la variabile x ha un impatto sulla y.

Possiamo solo dire che le due variabili si muovono insieme. È importante sapere se una variabile ha causato l'altra perché così potremmo interpretare la x come una politica per modificare la variabile di interesse y. Ipotizziamo di avere il tasso di occupazione come variabile dipendente, se vogliamo fare una politica per aumentare o diminuire il tasso di occupazione potremmo pensare a una politica x t.c. deve poter influenzare la nostra variabile dipendente. Il problema è che dovremo dare un'interpretazione casuale al nostro b1. Nelle lezioni precedenti abbiamo visto che tra le **proprietà di uno stimatore** c'è la **non distorsione**. Questa si ottiene come il valore atteso dello stimatore che deve essere uguale al valore teorico del nostro stimatore. Nel nostro caso dovremo avere che il valore dello stimatore deve essere = b1. **$E(B1 \wedge) = B1$** ;

Se vale questa ipotesi possiamo dire che il nostro b1 stimato è non distorto. **Se non distorto allora significa che lo possiamo interpretare come fattore causale** (almeno sul campione su cui stiamo lavorando) come un fattore causale della variabile x sulla variabile y. Come valutiamo se il b1 è distorto o no? Vediamo il valore atteso del b1 stimato e vediamo se è = b1 (vediamo se l'uguaglianza $e(b1) = b1$ è vera).

VEDI FORMULA SLIDE SUCCESSIVA (B CAPPELLO 1)

Ricordiamo per la seconda parentesi al numeratore la forma della regressione lineare $y = b_0 + b_1 x + \text{termine di errore}$; dobbiamo sostituire a xi e x medio il valore della regressione.

$$B_1 = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Nota diapositiva precedente

- Dal modello di regressione lineare con due parametri, tramite le derivate rispetto ai parametri, abbiamo ottenuto la formula di b1 (lez 6 del 17/3); abbiamo detto che b1 ci dice la relazione tra le due variabili (ossia se si muovono insieme, ma non se la relazione è causale o meno). Sicuramente si può dire che se B1 predetto (cappello) è diverso da 0 allora c'è correlazione tra x e y. Se è NON DISTORTO allora si potrebbe ipotizzare l'effetto causale. Affinché sia non distorto deve valere la relazione **$E(B1 \wedge) = B1$** .

Dato Y_i :

posso riscrivere $B1^\wedge$ come segue:

$$E(\hat{\beta}_1) = \beta_1$$

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x}) (\beta_0 + \beta_1 x_i + \epsilon_i - \beta_0 - \beta_1 \bar{x} - \bar{\epsilon})}{\sum (x_i - \bar{x})^2} =$$

$$\frac{\sum (x_i - \bar{x}) [\beta_1 (x_i - \bar{x}) + (\epsilon_i - \bar{\epsilon})]}{\sum (x_i - \bar{x})^2} = \frac{\beta_1 \sum (x_i - \bar{x})^2 + \sum (x_i - \bar{x}) (\epsilon_i - \bar{\epsilon})}{\sum (x_i - \bar{x})^2}$$

$$\beta_1 + \frac{\sum (x_i - \bar{x}) (\epsilon_i - \bar{\epsilon})}{\sum (x_i - \bar{x})^2} = \beta_1 + \frac{\sum (x_i - \bar{x}) \epsilon_i - \sum (x_i - \bar{x}) \bar{\epsilon}}{\sum (x_i - \bar{x})^2}$$

Nell'ultima la sommatoria = 0 perché è la somma degli scarti della media moltiplicata per una costante. In media ho $E(B1^\wedge)$.

.. Continuando vedi slide successiva

La non distorsione del beta (2)

$$E(\hat{\beta}_1) = \frac{E[\sum (x_i - \bar{x}) \epsilon_i]}{E[\sum (x_i - \bar{x})^2]} = \beta_1 + E\left[\frac{E[\sum (x_i - \bar{x}) \epsilon_i | x_i]}{\sum (x_i - \bar{x})^2} \right]$$

Posso dire che il valore medio di $b1$ teorico è $b1$. Per la legge delle aspettative ho la prima frazione.

Ho potuto fare l'ultimo passaggio per via di una proprietà;

$$\beta_1 + \frac{E\left\{ \sum (x_i - \bar{x}) E(\epsilon_i | x_i) \right\}}{\sum (x_i - \bar{x})^2} = \beta_1$$

A Se vale $E(\epsilon_i | x_i) = 0$

Devo ora rimanere con il solo $b1$: il valore atteso di $b1$ stimato è = $b1$

Non posso dire che $b1$ è non distorto; condizione necessaria affinché sia non distorto è $E(\hat{\beta}_1) = \beta_1$. Solo se assumo che il valore atteso di y_i dato x_i posso arrivare alla conclusione che il mio valore atteso è di $b1$ stimato è di $b1$.

Conclusione: nel caso dei minimi quadrati ordinari, affinché il beta 1 sia non distorto (quindi beta predetto=beta teorico reale) è necessario che $E(\epsilon_i | X_i) = 0$ (ossia è necessario che tutti i fattori che spiegano la y siano indipendenti da x ; le due variabili non devono essere correlate [condizione che ritroveremo anche nella consistenza -> cov=0])

La consistenza del beta

Altre proprietà del beta : la consistenza. la consistenza è un altro modo di valutare b_1 e l'effetto causale. Prima abbiamo lavorato sul campione ma se vogliamo lavorare in termini asintotici devo aggiungere altre 2 condizioni. Allora possiamo valutare la consistenza di $\hat{\beta}_1$ Se:

1. gli n elementi x_i e y_i estratti dalla popolazione sono i.i.d. (tutti estratti dalla stessa popolazione, l'estrazione è fatta con ripetizione (le palline estratte vengono rimesse nell'urna e rifaccio sempre l'estrazione.)
2. La probabilità di avere outlier deve essere ridotta al minimo (Ricordiamo che dobbiamo vedere la Curtosi, che è il momento quarto \rightarrow come si vedono gli outlier). Come avevamo assunto dalla non distorsione, anche qui assumiamo che x e y siano indipendenti, ossia che la covarianza (momento 4°) sia zero (prima avevamo assunto che le due variabili non fossero correlate). Se la $cov=0$ allora $B_1^{\wedge} \rightarrow B_1$.

Riportando il concetto in termini asintotici quindi, possiamo dire che se $n \rightarrow$ infinito (n ossia popolazione grande) allora b_1 stimato converga in probabilità a quello teorico all'aumentare del nostro campione verso infinito.

In termini probabilistici $\text{plim } \hat{\beta}_1$.

Se $n \rightarrow$ infinito, B_1^{\wedge} converge a $\text{plim } B_1^{\wedge}$. Ossia plim dice $\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}} \xrightarrow{P} \frac{COV(x,y)}{VAR(x)}$

- Lavoriamo sulla consistenza perché possiamo dire che il nostro campione è sufficientemente numeroso per approssimare i dati della popolazione
- beta ricordiamo era di formula in termini probabilistici converge a

$$\hat{\beta}_1 = \frac{S_{xy}}{s_{xx}} \xrightarrow{P} \frac{COV(x,y)}{VAR(x)} =$$

La varianza e cov campionaria per n che tende ad infinito, se valgono le proprietà viste (outlire ridotte al minimo e abbiamo iid) allora la formula del beta tende in termini asintotici alla cov/var. \rightarrow parliamo della relazione causale in termini di consistenza. Possiamo riportare quello detto fin ora per la non distorsione sulla consistenza attraverso la seguente formula: 3° rigo della slide successiva ($cov(x_i, b_0 + b_1 + e_i) / var(x_i)$).

Per le proprietà della covarianza possiamo riscrivere.. Inoltre ricordiamo che

- La cov per un termine, con la costante è 0;
- La cov di un termine per se stesso=1

La consistenza del beta (o non distorsione asintotica)

$$\hat{\beta}_1 \xrightarrow{P} \beta_1$$

$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}} \xrightarrow{P} \frac{\text{COV}(X, Y)}{\text{VAR}(X)} =$$

$$\frac{\text{COV}(X_i, \beta_0 + \beta_1 X_i + \epsilon_i)}{\text{VAR}(X_i)} = \frac{\text{COV}(X_i, \beta_0)}{\text{VAR}(X_i)} + \beta_1 \frac{\text{COV}(X_i, X_i)}{\text{VAR}(X_i)} + \frac{\text{COV}(X_i, \epsilon_i)}{\text{VAR}(X_i)}$$

$$\hat{\beta}_1 \xrightarrow{P} \beta_1$$

$\beta_1 + \frac{\text{COV}(X_i, \epsilon_i)}{\text{VAR}(X_i)} \xrightarrow{0}$
 \downarrow
 $\text{COV}(X_i, \epsilon_i) = 0$

Ma il meno prof?

Da Michele Bellipanni a Tutti:
prof può ripetere il ragionamento appena fatto sulla quantità uguale a 0

scusi ma la quantità di sinistra è uguale ... per quale motivo non è zero anche l'altra parte?

Da Daniele Rossignoli a Tutti:
perché la somma degli scarti dalla media non è moltiplicata per una costante (infatti moltiplica "epsilon i")

Da Michele Bellipanni a Tutti:
non avevo indicizzato ad (i) epsilon, grazie Daniele

A: Tutti

Esempio in cui $\hat{\beta}_1$ non è consistente: esempio di regressione con il salario dell'individuo. $\text{SALARIO}_i = b_0 + b_1(\text{istruzione})_i + \epsilon_i$.

$\hat{\beta}_1$ non è la causale di b_0 ma il termine di errore ha un impatto sull'istruzione e anche sul salario : potrebbe ad esempio essere l'abilità dell'individuo, il reddito di famiglia, ecc.

$\hat{\beta}_1$ non può essere interpretato come effetto causale in quanto nell'errore e teniamo conto di altri fattori(abilità, ecc).

Ciò causa la covarianza tra il termine di errore e b_1 è diverso da 0; $\text{COV}(X_i, \epsilon_i)$ diverso da 0. In particolare notiamo che è >0 allora il termine che vado a stimare è maggiore di quello teorico ($\hat{\beta}_1 > \beta_1$). L'importanza dell'istruzione è sopravvalutata.

$$\text{Salario}_i = \beta_0 + \beta_1 \text{ISTRUZIONE}_i + \epsilon_i$$

• NOTA: IL BETA PUO' ESSERE NON CONSISTENTE ANCHE IN CASO DI VARIABILE EMESSA.-> SLIDE 35: INCONSISTENZA DA VAR EMESSA.

Il modello di regressione lineare con variabile binaria

Il caso in cui la variabile esplicativa (x) è una variabile binaria (ossia può assumere solo due valori: 1 o 0).

- La variabile esplicativa è una variabile binaria (x è binario) che può assumere soltanto due valori, 0 oppure 1 a seconda della specifica categoria della variabile.

- Es. Il sesso di un individuo:

1: Femmina

0: Maschio

Vediamo come diventa il modello in questo caso

Nb. Il caso se maschio è il caso con solo la costante: il beta 0 stimato in questo caso è la media di y . (lo avevamo dimostrato).

Se l'individuo è femmina devo fare anche qui la minimizzazione : $\sum (y_i - \beta_0 - \beta_1)^2$, e poi ne faccio la derivata rispetto a β_1 . i successivi calcoli sono banali.

La stima del modello con variabile esplicativa binaria

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Se FEMMINA : $y_i = \beta_0 + \beta_1 + \varepsilon_i$

Se MASCHIO : $y_i = \beta_0 + \varepsilon_i \longrightarrow \hat{\beta}_0 = \bar{y}$

$$\sum_{i=1}^n (y_i - \beta_0 - \beta_1)^2$$

$$\frac{d}{d\beta_1} - 2 (y_i - \beta_0 - \beta_1) = 0$$

Vediamo il modello descritto con Y_1 .

Tramite i passaggi in questa slide e nella successiva riusciamo ad arrivare al β_1 stimato (che in questo caso è espresso come differenza delle medie).

$$\sum_{i=2}^m y_i - \sum_{i=2}^m \bar{y} - \sum_{i=2}^m \beta_1 = 0$$

$$\bar{y}_{i,1} - \bar{y}_{i,0} - \beta_1 = 0$$

$$-\hat{\beta}_1 = \bar{y}_{i,0} - \bar{y}_{i,1}$$

$$\hat{\beta}_1 = \bar{y}_{i,1} - \bar{y}_{i,0}$$

- Il caso della stima con variabile esplicativa binaria è un caso in cui possiamo interpretare il beta 1 stimato come **differenza tra le medie dei due gruppi** perché quando abbiamo a stimare il modello per gli individui che hanno la variabile esplicativa pari a 0 (i maschi) otterrò la media del sottogruppo 0. ottengo quindi y medio per il sottogruppo 0.

Gli standard error della regressione e del parametro $\hat{\beta}_1$

Nell'ambito della regressione lineare abbiamo 2 standard error da analizzare: SE e SER (togliendo 2 gdl per non avere distorsione).

- **Lo standard error della regressione:**

$$SER = \sqrt{\frac{1}{n-2} \sum_{i=1}^N \varepsilon_i^2} = \sqrt{\frac{1}{n-2} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

Nota come dalle slide precedenti capisci che l'errore può essere riscritto come la differenza delle medie.

Rappresenta una stima della relazione standard del termine di errore. Come la calcoliamo la stima della relazione standard del termine di errore? Il termine di errore è epsilon i e siccome noi dobbiamo calcolare una deviazione standard lo possiamo scrivere come la radice quadrata della varianza del termine errore (che non è altro che il termine di errore meno la sua media tutto elevato al quadrato). E quindi il valore atteso del termine di errore non abbiamo nel caso specifico però non abbiamo la possibilità di costruire la varianza di correlazione ma usiamo la varianza campionaria. La varianza campionaria per il termine errore si ottiene con la sommatoria del termine di errore meno la media del termine di errore al quadrato.

- **Lo standard error del coefficiente stimato $\hat{\beta}_1$:**

$$SE(\hat{\beta}_1) = \sqrt{\hat{\sigma}_{\hat{\beta}_1}^2} = \sqrt{\frac{\frac{1}{n-2} \sum_{i=1}^N (x_i - \bar{x}_i)^2 \varepsilon_i^2}{\left[\frac{1}{n} \sum_{i=1}^N (x_i - \bar{x}_i)^2 \right]^2}}$$

Lo s.e. lo troviamo nella statistica t e ci serve anche nel calcolo degli intervalli di confidenza. Lo s.r del beta stimato lo troviamo nell'output della regressione (quando usiamo es. stata). Lo s.e del coefficiente beta 1 stimato : sappiamo che è una stima della deviazione standard. Anche per il caso del beta 1 stimato coliamo la sua varianza. La formulazione che vediamo si ottiene dal b1 stimato (così otteniamo la seconda frazione) . Se facciamo la varianza del numeratore e del denominatore si ottiene questa formula.

IPOTESI MODELLO MULTIVARIATO PER EFFETTI CAUSALI

- 1) $E(\varepsilon_i | X_j) = 0 \quad \forall j$ LARGO SAMPLE
 $\text{COV}(\varepsilon_i, X_j) = 0$
- 2) $y_i, X_j \quad \forall j \quad \text{i.i.d.}$ $\forall j$
- 3) OUTLIERS (VALORI ESTREMI) DEVONO
ESSERE IMPROBABILI O ASSENTI (MOMENTI QUARTI)
- 4) PERFETTA MULTICOLLINEARITÀ TRA I REGRESSORI $\rightarrow (X'X)^{-1}$ DEVE ESSERE INVERTITA

Per interpretare il parametro stimato nel modello di regressione come un effetto causale abbiamo bisogno di inserire alcune ipotesi :

- il termine di errore deve essere indipendente dalla variabile per ogni j .

$$E(\epsilon_i | X_j) = 0$$

$$\text{COV}(\epsilon_i, X_j) = 0$$

In termini di consistenza diremo (per grandi campioni)

- Per ogni regressore a richiesta l'indipendenza rispetto al termine di errore

$$y_i, X_j$$

$$\forall j \quad i.i.d$$

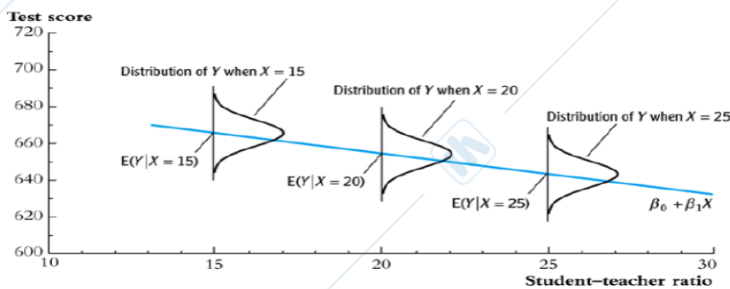
$$\text{COV}(\epsilon_i, X_j) = 0$$

- Gli outliers (valori estremi) Devono essere improbabili o assenti (studiamo il momento quarto)
- È Richiesta l'assenza di perfetta multicollinearità tra i regressori altrimenti non possiamo invertire la matrice $(X'X)^{-1}$. La multicollinearità comporta la non invertibilità, la matrice non sarebbe singolare o $\text{num } k > m$. In questi casi il beta non può essere stimato). Tuttavia è ammessa multicollinearità non perfetta.->

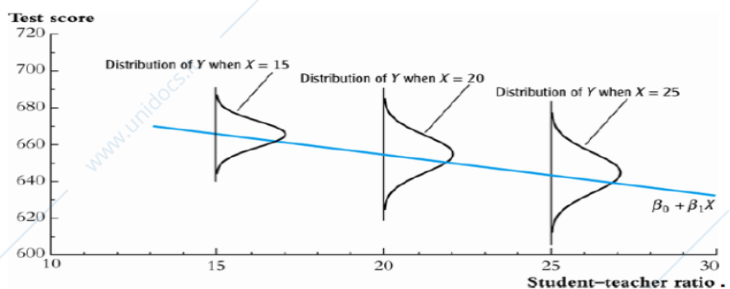
La multicollinearità Se non è perfetta allora esiste dipendenza lineare tra i regressori. Le x sono correlate tra di loro ma non perfettamente . è possibile stimare il beta ma all'aumentare della collinearità gli SE tenderebbero ad aumentare perché diminuirebbe la variabilità . comporta una diminuzione della statistica t e un aumento degli intervalli di confidenza (abbiamo problemi di efficienza della stima ; problema di significatività dei coefficienti trovati).

Vediamo i concetti di Omoschedasticità e Eteroschedasticità

Example of **homoskedasticity** $Var(u_i|X_i) = \sigma_u^2$:



Example of **heteroskedasticity** $Var(u_i|X_i) \neq \sigma_u^2$



La nostra sola assunzione circa la distribuzione condizionata di u_i data X_i è che essa abbia media nulla (la prima assunzione dei minimi quadrati). Se, in aggiunta, la varianza di questa distribuzione condizionata non dipende da X_i , allora si dice che gli errori sono omoschedastici.

nel secondo caso la varianza non è fissa, dipende da x . La variabilità è alta.

In caso di omoschedasticità la formula dello $SE(\beta_1)$ nelle slide precedenti si semplifica.

Nella figura sopra, tutte queste distribuzioni condizionate hanno la stessa dispersione; più precisamente, la varianza di queste distribuzioni è la stessa per i diversi valori di x . Cioè, la varianza condizionata di u_i dato che $X_i = x$ non dipende da x , e quindi gli errori illustrati nella figura sopra sono omoschedastici.

L'omoschedasticità non dipende dal regressore ma dalla varianza!

Implicazioni: 1) Siano gli errori di regressione eteroschedastici o omoschedastici, lo stimatore OLS è non distorto, consistente e asintoticamente normale. 2) se alle assunzioni dei minimi quadrati si aggiunge l'omoschedasticità abbiamo che β_0 e β_1 sono BLUE \rightarrow TEOMARKOW.

A seconda di etero/omoschedasticità alcune formule cambiano:

Gli errori standard robusti all'eteroschedasticità:

Lo stimatore della varianza di $\hat{\beta}_0$ è

$$\hat{\sigma}_{\hat{\beta}_0}^2 = \frac{1}{n} \times \frac{\sum_{i=1}^n \hat{H}_i^2 \hat{u}_i^2}{\left(\frac{1}{n} \sum_{i=1}^n \hat{H}_i^2\right)^2}$$

dove $\hat{H}_i = 1 - [\bar{X} / \frac{1}{n} \sum_{i=1}^n X_i^2] X_i$. L'errore standard di $\hat{\beta}_0$ è $SE(\hat{\beta}_0) = \sqrt{\hat{\sigma}_{\hat{\beta}_0}^2}$. Il ragionamento dietro lo stimatore $\hat{\sigma}_{\hat{\beta}_0}^2$ è lo stesso di $\hat{\sigma}_{\hat{\beta}_1}^2$ e segue dalla sostituzione delle aspettative con le medie campionarie.

Le varianze per il caso di omoschedasticità pura. Sotto l'assunzione di omoschedasticità, la varianza condizionata di u_i data X_i è una costante, ovvero, $var(u_i|X_i) = \sigma_u^2$. Se gli errori sono omoschedastici, le formule del concetto chiave 4.4 si semplificano come

$$\hat{\sigma}_{\hat{\beta}_0}^2 = \frac{\sigma_u^2}{n\sigma_X^2} \quad \text{e} \quad (4.60)$$

$$\hat{\sigma}_{\hat{\beta}_1}^2 = \frac{E(X_i^2)}{n\sigma_X^2} \sigma_u^2 \quad (4.61)$$

Per derivare la (4.60), si scriva il numeratore della (4.14) come $var[(X_i - \mu_X)u_i] = E\{[(X_i - \mu_X)u_i]^2\} = E\{[(X_i - \mu_X)u_i]^2\} = E[(X_i - \mu_X)^2 u_i^2] = E[(X_i - \mu_X)^2 var(u_i|X_i)]$, dove la seconda uguaglianza deriva dal fatto che $E[(X_i - \mu_X)u_i] = 0$ (per la prima ipotesi dei minimi quadrati) e dove l'uguaglianza finale deriva dalla legge delle aspettative iterate (sezione 2.3). Se u_i è omoschedastico, allora $var(u_i|X_i) = \sigma_u^2$ e quindi $E[(X_i - \mu_X)^2 var(u_i|X_i)] = \sigma_u^2 E[(X_i - \mu_X)^2] = \sigma_u^2 \sigma_X^2$. Il risultato nella (4.60) deriva sostituendo tale espressione nel numeratore della (4.14) e semplificando. Un simile calcolo fornisce la (4.61).

Gli errori standard per il caso di omoschedasticità pura. Gli errori standard per il caso di omoschedasticità pura si ottengono sostituendo le medie e le varianze della popolazione nella (4.60) e nella (4.61) con le medie e le varianze campionarie e stimando la varianza di u_i con il quadrato del SER . Gli stimatori di tali varianze per il caso di omoschedasticità pura sono

$$\hat{\sigma}_{\hat{\beta}_0}^2 = \frac{s_u^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (\text{per il caso di omoschedasticità pura}) \quad \text{e} \quad (4.62)$$

$$\hat{\sigma}_{\hat{\beta}_0}^2 = \frac{\left(\frac{1}{n} \sum_{i=1}^n X_i^2\right) s_u^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (\text{per il caso di omoschedasticità pura}), \quad (4.63)$$

dove s_u^2 è dato dalla (4.40). Gli errori standard per il caso di omoschedasticità pura sono le radici quadrate di $\hat{\sigma}_{\hat{\beta}_0}^2$ e $\hat{\sigma}_{\hat{\beta}_1}^2$.

Inconsistenza da variabile omessa

Siamo interessati all'effetto causale di x su y . Nel caso di variabile omessa abbiamo che la condizione di consistenza viene meno. Vediamo perché:

per avere la condizione $\hat{\beta}_1 \xrightarrow{p} \beta_1$ serve che $\text{COV}(X_i, \epsilon_i) = 0$.

Assumiamo di avere il modello reale (1) $Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \epsilon_i$ allora la cov sarà

$$\text{COV}(X_{1,i}, X_{2,i}) \neq 0$$

abbiamo che il nostro modello dipende da 2 variabili: X_1 e X_2 ; X_1 avrà effetto su y e sarà correlata anche con X_2 . (X_1 è correlata con X_2 , ma X_2 non è inserita nel modello)

Se invece che stimare questo modello, stimassimo questo secondo modello: (2) $Y_i = \beta_0 + \beta_1 X_{1,i} + u_i$

dove u_i è $u_i = \beta_2 X_{2,i} + \epsilon_i$ avremo che stiamo omettendo un variabile. Il problema è che il termine di errore si nasconde dentro u_i .

$$p \lim \hat{\beta}_1 = \frac{\text{COV}(Y_i, X_{1,i})}{\text{VAR}(X_{1,i})} = \frac{\text{COV}(\beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \epsilon_i, X_{1,i})}{\text{VAR}(X_{1,i})}$$

$$= \frac{\beta_1 \text{VAR}(X_{1,i}) + \beta_2 \text{COV}(X_{2,i}, X_{1,i}) + \text{COV}(X_{2,i}, \epsilon_i)}{\text{VAR}(X_{1,i})}$$

$\text{Cov}(y_i, x_i) \rightarrow$ sostituisco la y_i che ho nel modello 2. L'ultima cov si annulla perché abbiamo posto la $\text{cov}(X_1, X_2) = 0$ nel modello. Di conseguenza si annulla la frazione. Ho consistenza del beta stimato ponendo tale condizione. Se la condizione vale allora anche stimando il modello ridotto ho che la stima è consistente. Se ho che è diverso da zero no (non ho inserito la variabile correlata ad x quindi avrò un β_1 distorto o inconsistente). Es. salario in funzione dell'istruzione: se non inserisco l'abilità potrei avere una stima inconsistente

Nota diapositiva precedente

$\text{Cov}(y_i, x_i) \rightarrow$ sostituisco la y_i che ho nel modello 2.

L'ultima cov si annulla perché abbiamo posto la $\text{cov}(X_1, X_2) = 0$ nel modello. Di conseguenza si annulla la frazione. Ho consistenza del beta stimato ponendo tale condizione. Se la condizione vale allora anche stimando il modello ridotto ho che la stima è consistente. Se ho che è diverso da zero no (non ho inserito la variabile correlata ad x quindi avrò un β_1 distorto o inconsistente).

Es. salario in funzione dell'istruzione: se non inserisco l'abilità potrei avere una stima inconsistente.

Modello di regressione multipla

Considero un vettore di regressori e avrò beta da stimare. Avrò che il modello in forma matriciale (eliminando il pedice «i») è $y = XB + \epsilon$ dove y ha n osservazioni quindi

Y avrà dimensioni $n \times 1$;

X : $n \times k$ (n osservazioni dal campione e k variabili)

B : $k \times 1$ (ho k parametri da stimare : B_1, B_2, \dots, B_k)

ϵ : $n \times 1$ (termine di errore)

Prodotto
Matriciale:

$$\begin{array}{c}
 \text{VETTORE } y \\
 \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}
 \end{array}
 \quad
 \text{MATRICE } X
 \quad
 \begin{bmatrix} 1 & x_{1,1} & \dots & x_{1,k} \\ 1 & x_{2,1} & \dots & x_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \dots & x_{n,k} \end{bmatrix}
 \quad
 \begin{array}{c}
 \text{VETTORE } \epsilon \\
 \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}
 \end{array}$$

Come si trova il B ? Minimizziamo i minimi quadrati come nella regressione semplice però la struttura è diversa (ho la derivata di un prodotto).

$$\epsilon = y - X\beta \rightarrow \epsilon'\epsilon = (y - X\beta)'(y - X\beta)$$

Derivando rispetto a B e ponendo $=0$ si ottiene

$$\begin{array}{l}
 X'(y - X\hat{\beta}) = 0 \\
 X'y - X'X\hat{\beta} = 0 \rightarrow -X'X\hat{\beta} = -X'y \\
 X'X\hat{\beta} = X'y \\
 \hat{\beta} = (X'X)^{-1}(X'y)
 \end{array}$$

Nb: ' significa TRASPOSTO!

Per esempio se volessimo stimare il salario dell'individuo i esimo, sceglieremo il genere del lavoratore e poi mettiamo il parametro per dire se l'individuo è laureato o no.

- Se assumiamo che la variabile dicotomica assume 1 o 0 in base a se è femmina e maschio invece l'altra variabile assume 0 o 1 se è laureato o no.
- Questo è un modello multivariato con due regressori (in generale si possono avere k regressori. In quel caso è meglio usare la combinazione matriciale del modello. Se volessimo utilizzare le k variabili il modello diventerebbe del tipo-> slide successiva

$$SALARIO_i = \beta_0 + \beta_1 \text{GENERE}_i + \beta_2 \text{LAUREATO}_i + \epsilon_i$$

β_1

$$\frac{\Delta \text{SALARIO}}{\Delta \text{GENERE}}$$

FORMA MATRICIALE

$$Y = X\beta + \epsilon$$

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon_i$$

$$\text{SALARIO}_i = \beta_0 + \beta_1 \text{GONERSE}_i + \beta_2 \text{LAUREATO}_i + \varepsilon_i$$

↓
 $\hat{\beta}_1$

$$\frac{\Delta \text{SALARIO}}{\Delta \text{GONERSE}}$$

FORMA MATRICIALE

$$Y = X\beta + \varepsilon$$

$m \times 1$ $m \times k+1$ $k+1$ $m \times 1$

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i$$

$$(Y - X\beta)'(Y - X\beta)$$

DERIVATA RISPETTO A β
E PONDENDO LA = 0

$$X'(Y - X\beta) = 0 \quad X'Y - X'X\beta = 0 \rightarrow \hat{\beta} = (X'X)^{-1} X'Y$$

Dall'ultima espressione abbiamo trovato lo stimatore di una regressione a k variabili.

Qual è il problema del beta stimato in presenza di k regressori? Dobbiamo invertire la matrice

I problemi legati a ciò sono

1) Il beta stimato in presenza di k regressori non può essere stimato quando abbiamo perfetta multilinearità tra i regressori.

2) Quando k è maggiore di n ($k > n$), cioè quando i regressori sono di più numerosi del nostro campione non possiamo ottenere il beta

Per invertire la matrice dobbiamo avere il determinante diverso da zero (matrice non singolare). Quando è singolare? Quando possiamo esprimere un vettore in combinazione lineare di un altro regressore. I vettori sono linearmente dipendenti quindi il det è 0. altro caso è quello in cui abbiamo più di k regressori perché otterremo una matrice non invertibile (es. campione è di 50, i regressori sono più di 50). Ci sono altri modelli che ci permettono di ottenere una matrice X trasposta X se $k > n$.

N OSSERVAZIONI

K VARIABILI / REGRESSORI

Come si ottiene y_i prefetto? Stimando i parametri.
Prevedendo i parametri (derivate)

$$SALARIO_i = \beta_0 + \beta_1 GONERE_i + (\beta_2 LAUREATO_i) + \epsilon_i$$

$$\beta_1 = \frac{\Delta Y}{\Delta X} \rightarrow \hat{\beta}_1 = \frac{\Delta SALARIO}{\Delta GONERE}$$

FORMA MATRICIALE

$$m \times 1 \quad m \times k+1 \quad k+1 \quad m \times 1$$

$$Y = X\beta + \epsilon$$

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \epsilon_i$$

$$(Y - X\beta)' (Y - X\beta)$$

DERIVATA RISPETTO A β
E PONDENDOLA = 0

$$X'(Y - X\beta) = 0 \quad X'Y - X'X\beta = 0 \rightarrow \hat{\beta} = (X'X)^{-1} X'Y$$

\downarrow
 $\begin{matrix} k+1, n & n, k+1 \\ k+1, & k+1 \end{matrix}$

$$R^2 = \frac{ESS}{TSS} \rightarrow 0 \leq R^2 \leq 1 \rightarrow 1 - \frac{RSS}{TSS}$$

$$TSS = \sum_{i=1}^n (y_i - \bar{y})^2$$

$$ESS = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}$$

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$TSS = ESS + RSS$

\downarrow
 PARTE SPIEGATA DELLA VARIABILITA' DI Y INTORNO ALLA MEDIA

\rightarrow PARTE NON SPIEGATA

PROBLEMA R^2 :

AUMENTA STRUTTURALMENTE ALL'AUMENTARE DEI
REGRESSORI ANCHE SE I REGRESSORI NON
DANNO UN'EFFETTIVA INFORMAZIONE SULLA
VARIABILE Y

$$\text{ADJUSTED } R^2 = 1 - \frac{n-1}{n-k-1} \cdot \frac{RSS}{TSS}$$

$$\text{ORIGINALE } R^2 = 1 - \frac{RSS}{TSS}$$

R^2 è il coefficiente di determinazione;

È una misura per vedere quanto il nostro modello riesce a spiegare della nostra variabile y . Spiega l'accuratezza della predizione. Dice la porzione di varianza totale che il modello della nostra variabile y riesce a spiegare. Dice quanto la varianza della variabile predetta è spiegata.

R^2 È utile per prevedere l'evoluzione di un fenomeno o variabile ma in un'area diversa da quella studiata fino a quel momento.

È 0 quando non riusciamo a spiegare nulla del modello. La variabilità del modello non viene spiegata.

È 1 quando riusciamo a spiegare tutto del modello. La varianza è spiegata bene. $\rightarrow Y \text{ predetto} = Y_i$

Se è alto Significa che non si è capaci di prevedere la variazione della variabile y su altri dati.

Problema dell' R^2 : aumenta strutturalmente all'aumentare dei regressori, anche quando questi non danno informative aggiuntive sulla y .

Si passa allora all' R^2 Adjusted.

$TSS = ESS + RSS$ (ESS è la parte che riusciamo a spiegare invece RSS è la parte non spiegata).

$R^2 = \frac{ESS}{TSS} \rightarrow 0 \leq R^2 \leq 1 \rightarrow 1 - \frac{RSS}{TSS}$

$TSS = \sum_{i=1}^n (y_i - \bar{y})^2$

$ESS = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$

$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2$

$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$

$TSS = ESS + RSS$

$SER = \frac{1}{n-2} \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2}$

PARTI NON SPIEGATE DELLA VARIABILITÀ DI Y INTORNO ALLA MEDIA

PARTE SPIEGATA DELLA VARIABILITÀ DI Y INTORNO ALLA MEDIA

```

.gen y=100+(0.5*x)+errore
.
.
.
.
end of do-file

.do "C:\Users\flbl\AppData\Local\Temp\STD3b4c_000000.tmp"

.reg y x
    
```

Source	SS	df	MS	Number of obs	=	100
Model	576669117	1	576669117	F(1, 98)	=	2237.46
Residual	25257884.1	98	257733.511	Prob > F	=	0.0000
Total	601927001	99	6080070.72	R-squared	=	0.9580
				Adj R-squared	=	0.9576
				Root MSE	=	507.67

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x	.4953291	.0104717	47.30	0.000	.4745484 .5161098
_cons	124.2494	116.7418	1.06	0.290	-107.421 355.9197

```
. reg y x
```

Source	SS	df	MS	Number of obs	=	100
Model	576669117	1	576669117	F(1, 98)	=	2237.46
Residual	25257884.1	98	257733.511	Prob > F	=	0.0000
Total	601927001	99	6080070.72	R-squared	=	0.9580
				Adj R-squared	=	0.9576
				Root MSE	=	507.67

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x	.4953291	.0104717	47.30	0.000	.4745484 .5161098
_cons	124.2494	116.7418	1.06	0.290	-107.421 355.9197


```
. reg y x, level(90)
```

Source	SS	df	MS	Number of obs	=	100
Model	576669117	1	576669117	F(1, 98)	=	2237.46
Residual	25257884.1	98	257733.511	Prob > F	=	0.0000
Total	601927001	99	6080070.72	R-squared	=	0.9580
				Adj R-squared	=	0.9576
				Root MSE	=	507.67

y	Coef.	Std. Err.	t	P> t	[90% Conf. Interval]
x	.4953291	.0104717	47.30	0.000	.4779404 .5127178
_cons	124.2494	116.7418	1.06	0.290	-69.60643 318.1052

```
. reg y x
```

Source	SS	df	MS	Number of obs	=	100
Model	576669117	1	576669117	F(1, 98)	=	2237.46
Residual	25257884.1	98	257733.511	Prob > F	=	0.0000
Total	601927001	99	6080070.72	R-squared	=	0.9580
				Adj R-squared	=	0.9576
				Root MSE	=	507.67

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x	.4953291	.0104717	47.30	0.000	.4745484 .5161098
_cons	124.2494	116.7418	1.06	0.290	-107.421 355.9197


```
. reg y x, level(90)
```

Source	SS	df	MS	Number of obs	=	100
Model	576669117	1	576669117	F(1, 98)	=	2237.46
Residual	25257884.1	98	257733.511	Prob > F	=	0.0000
Total	601927001	99	6080070.72	R-squared	=	0.9580
				Adj R-squared	=	0.9576
				Root MSE	=	507.67

y	Coef.	Std. Err.	t	P> t	[90% Conf. Interval]
x	.4953291	.0104717	47.30	0.000	.4779404 .5127178
_cons	124.2494	116.7418	1.06	0.290	-69.60643 318.1052

```

. ****valore soglia al 1% della statistica t=2.58*****
. reg y x, level(99)

Source |           SS       df       MS       Number of obs   =       100
-----|-----
Model | 576669117         1   576669117   F(1, 98)         =       2237.46
Residual | 25257884.1       98   257733.511   Prob > F          =       0.0000
-----|-----
Total | 601927001       99   6080070.72   R-squared         =       0.9580
                                           Adj R-squared    =       0.9576
                                           Root MSE        =       507.67

-----+-----
y |           Coef.   Std. Err.   t    P>|t|   [99% Conf. Interval]
-----+-----
x |   .4953291   .0104717   47.30  0.000   .4678207   .5228375
_cons | 124.2494   116.7418    1.06  0.290  -182.4234   430.9221

. gen ci_up_99=.4953291+(2.58*.0104717)
. gen ci_low_99=.4953291-(2.58*.0104717)
. sum ci_up_99 ci_low_99

Variable |           Obs       Mean       Std. Dev.       Min       Max
-----+-----
ci_up_99 |           100   .5223461           0   .5223461   .5223461
ci_low_99 |           100   .4683121           0   .4683121   .4683121

. ****ESS TSS RSS R2 adj_R2 SER*****

```

Lezione 10

Francesco Bloise

francesco.bloise@uniroma1.it

La non distorsione nel caso multivariato

Il modello migliore nel modello multivariato è quello nella forma matriciale:
(stiamo dimostrando che B stimato = B reale)

MODELLO IN FORMA MATRICIALE:

$$Y = X\beta + \epsilon$$

\downarrow \downarrow \downarrow
 $n \times 1$ $n, k+1$ $k+1, 1$

E sappiamo che in questo caso B stimato sarà:

$$\hat{\beta} = (X'X)^{-1} X'Y$$

SOSTITUENDO ALLA Y LA (1)

$$\hat{\beta} = (X'X)^{-1} X'(X\beta + \epsilon) \rightarrow \beta + (X'X)^{-1} X'\epsilon$$

1) Dimostrazione 1... Nel caso della non distorsione abbiamo che

$$E(\hat{\beta}) = \beta + E\left[(X'X)^{-1} \cdot X'\epsilon\right];$$

allora B sarà:

$$\beta + \underbrace{E\left[(X'X)^{-1}\right]}_0 \cdot \underbrace{E(X'\epsilon)}_0$$

se $E(X'\epsilon) = 0$ (se consideriamo l'ipotesi di indipendenza qui indicata avremo che il valore atteso di quello è = 0)

Modo alternativo di dimostrare la consistenza è farlo attraverso le medie iterate :

2) Dimostrazione 2 :

$$\beta + E \left[(x'x)^{-1} x' \varepsilon \right]$$

CON ASPETTATIVE ITERATE

$$\beta + E \left\{ \left[(x'x)^{-1} x' \mid X \right] (\varepsilon \mid X) \right\}$$

$$\beta + E \left[(x'x)^{-1} x' (\varepsilon \mid X) \right]$$

0 → PER IPOTESI (1)
ossia $E(\varepsilon_i \mid X_j) = 0$

E SI OTTIENE
che il beta non è distorto

$$E(\hat{\beta}) = \beta$$

TEST DI IPOTESI M. MULTIVARIATO

vediamo se i coefficienti sono significativi.
Possiamo vedere se i singoli coefficienti sono significativi

LA STATISTICA T → $t_j = \frac{\hat{\beta}_j - 0}{SE(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}$

Oppure vedere la significatività congiunta dei parametri visto che siamo nel caso multivariato;

SIGNIFICATIVITÀ CON GIUSTA

L'ipotesi nulla è che tutti i parametri siano = 0

$$H_0 = \beta_1 = \beta_2 = \beta_3 = \dots = \beta_k = 0$$

IPOTESI NULLA

Ipotesi alternativa è che almeno un parametro sia diverso da 0

$$H_1 = \text{almeno un } \beta_j \neq 0$$

IPOTESI ALTERNATIVA

Questo test d'ipotesi si distribuisce come una F. Utilizziamo allora la statistica F.

STATISTICA F

La statistica F

Può assumere diverse formulazioni:

$$1) F = \frac{1}{2} \left(\frac{t_1^2 + t_2^2 - 2\hat{\rho}_{t_1 t_2} t_1 t_2}{1 - \hat{\rho}_{t_1 t_2}^2} \right)$$

MODELLO CON 2 SOLI REGRESSORI

CASO SPECIFICO

$$R^2 = 1 - \frac{RSS}{TSS}$$

si passa dal 2° al 3° modello con questi passaggi

$$\left(\frac{SSR_R}{TSS} - \frac{SSR_U}{TSS} \right) / q$$

$$2) F_{homoskedasticity} = \frac{(SSR_{restricted} - SSR_{unrestricted}) / q}{SSR_{unrestricted} / (n - k_{unrestricted} - 1)}$$

numero regressori
numero obs. osservazioni
num. reg. coeff. ≠ 0

CASI GENERALI SOTTO IPOTESI DI

OMOSCHEDASTICITÀ (assumiamo stimatore BLUE)
VAR (ε.) È FISSA
β NON VARIA AL VARIARE X

$$\left(\frac{SSR_U}{TSS} \right) / (n - k - 2)$$

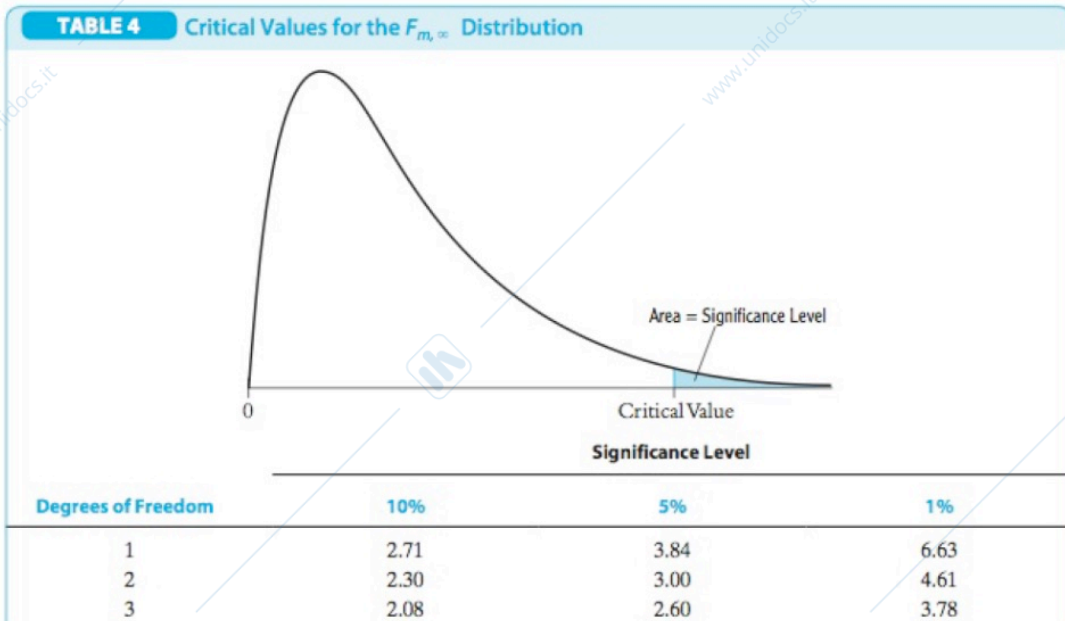
$$\left(R^2 + \frac{1}{q} + R_U^2 - \frac{1}{q} \right) / q$$

$$3) F_{homoskedasticity} = \frac{(R_{unrestricted}^2 - R_{restricted}^2) / q}{(1 - R_{unrestricted}^2) / (n - k_{unrestricted} - 1)}$$

POSSO USARE QUESTA

FORMULAZIONE PER CALCOLARE LA F

$$\left(\frac{R_U^2 - R_R^2}{1 - R_U^2} \right) / q$$



Lezione 11

$$Y_i = \beta_0 + \beta_1 X_{1i} + \epsilon_i$$

$$\frac{\partial Y}{\partial X_1} = \beta_1$$

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

$$\frac{\partial Y}{\partial X_1} = \beta_1 \rightarrow \text{SI OTTIENE TENENDO FISSA } X_2$$

$$\frac{\partial Y}{\partial X_2} = \beta_2 \rightarrow \text{SI OTTIENE TENENDO FISSA } X_1$$

NON LINEARITÀ

1) PRESENZA DI POLINOMI DI GRADO R

SE PONIAMO $R=2$

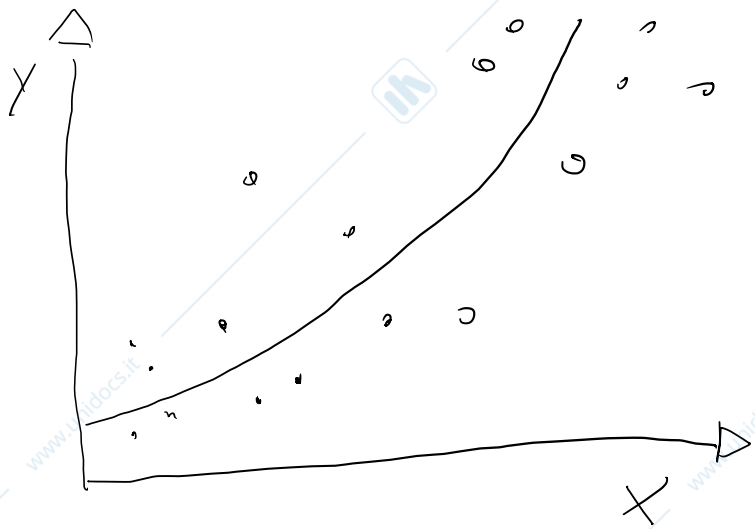
$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}^2 + \epsilon_i$$

$$\frac{\partial Y_i}{\partial X_{2i}} = \beta_1 + \underbrace{2 \beta_2 X_{2i}}_{\substack{\downarrow \\ \beta_2} X_2}$$

\downarrow β_1 \uparrow β_2

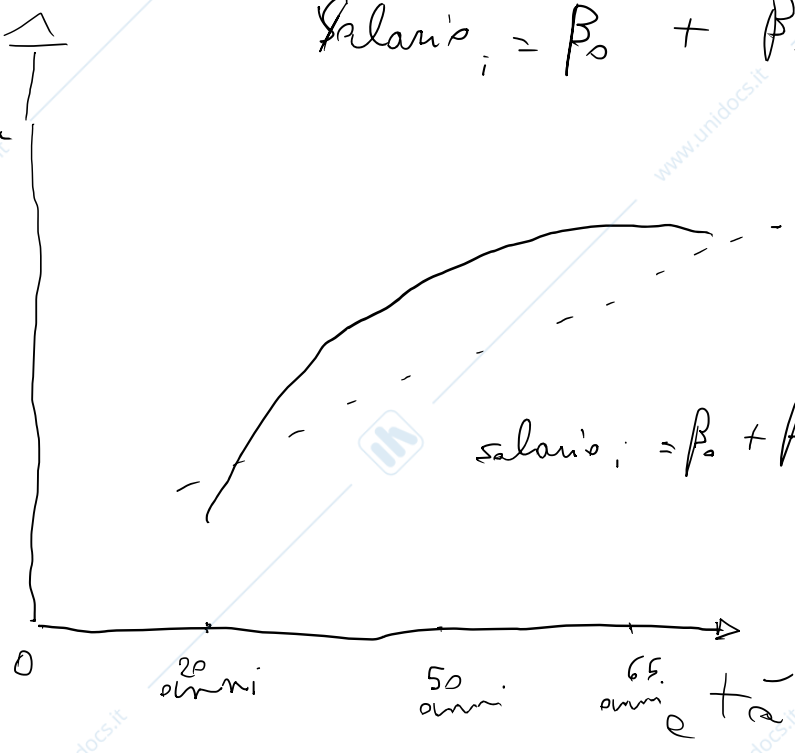
$E \hat{\beta}_1 > 0$ e $\hat{\beta}_2 > 0$

con diversi valori di X_2



Salario Individuale

$Salario_i = \beta_0 + \beta_1 et\bar{a}_i + \epsilon_i$
 $E \beta_2 > 0$



$salario_i = \beta_0 + \beta_1 et\bar{e}_i + \beta_2 et\bar{a}_i^2 + \epsilon_i$

$\beta_1 > 0$ $\beta_2 < 0$

2) i modelli con logaritmo

- log - lineare \rightarrow la variabile dipendente in log
- lineare, log \rightarrow il regressore che è espresso in log
- log - log \rightarrow entrambe le variabili sono espresse in log

$$\Delta \log(Y_i) \approx \frac{\Delta Y}{Y}$$

$$\Delta \log(X_i) \approx \frac{\Delta X}{X}$$

$$\log(Y_i) = \beta_0 + \beta_1 \log(X_1) + \beta_2 \log(X_2) + \epsilon_i$$

$$\beta_2 = \frac{\Delta \log(Y_i)}{\Delta \log(X_{2i})} \approx \frac{\Delta Y / Y}{\Delta X_2 / X_2} \quad \text{log - log}$$

MODELLO LOG - LINEARE

$$\log(Y_i) = \beta_0 + \beta_1 X_{2i} + \varepsilon_i$$

$$\beta_1 = \frac{\Delta \log(Y_i)}{\Delta X_1} \approx \frac{\Delta Y / Y}{\Delta X_1} = 0,5 \quad \text{esempio}$$

MODELLO LINEARE - LOG

$$Y_i = \beta_0 + \beta_1 \log(X_{2i}) + \varepsilon_i$$

$$\beta_1 = \frac{\Delta Y_i}{\Delta \log(X_{2i})} \approx \frac{\Delta Y_i}{\Delta X_{2i} / X_{2i}} = 4 \quad \text{esempio}$$

3) MODELLO CON INTERAZIONI

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 (X_1 \cdot X_2) + \varepsilon_i$$

MODELLO POLINOMIALE DI SECONDO GRADO

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 (X_1^2 \cdot X_2) + \varepsilon_i$$

$$\Delta \frac{\partial Y}{\partial X_2} = \beta_1 + \beta_2 X_2$$

$$\hat{\beta}_1 + \hat{\beta}_2 X_2$$

esempio
modello

$$x_2 = \text{anni di studio}$$

con
interazione

$$x_2 = \text{genere o del lavoratore}$$

$$x_2 \begin{cases} \rightarrow 0 & \rightarrow \text{MASCHIO} \\ \rightarrow 1 & \rightarrow \text{FEMMINA} \end{cases}$$

CASO con $x_2 = 0$

$$y_i = \beta_0 + \beta_2 x_2 + \epsilon_i$$

CASO con $x_2 = 1$

$$y_i = \beta_0 + \beta_2 x_2 + \beta_2 x_2 + \epsilon_i$$

$$y_i = \beta_0 + x_2 (\beta_2 + \beta_2) + \epsilon_i$$

$$y = f(x) + \epsilon_i$$

funzione
può

generica di x
includere sia relazioni
lineari che non
linearità