



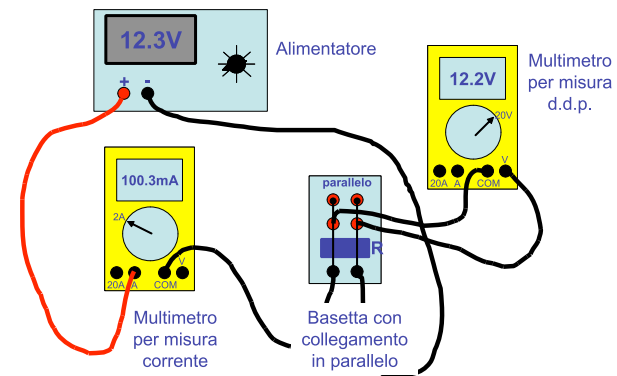
Stima dei parametri Regressioni

Sommario

- Esempi di problemi di stima di modelli lineari
- Stima e Stimatore
- Minimi Quadrati e Massima Verosimiglianza
- Esempio Stimatore MV con modelli lineari

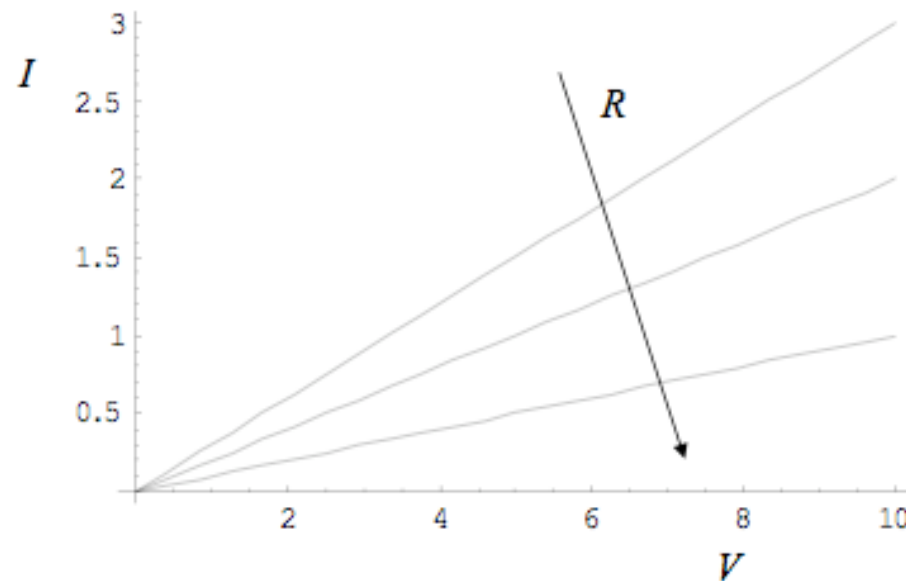
Esempio I

- Vogliamo verificare che la legge di Ohm e determinare la resistività di un circuito semplice.
- Equazione che modella il processo (Legge di Ohm): $I = (1/R) V$
 - V : differenza di potenziale
 - I : corrente
 - R resistività
- Eseguiamo N prove sperimentali fissando la differenza di potenziale e misurando la corrente. Sulla base di tali dati vogliamo verificare la validità della legge e **stimare** R .
 - V : Condizione sperimentale
 - I : Grandezza misurata



Esempio I

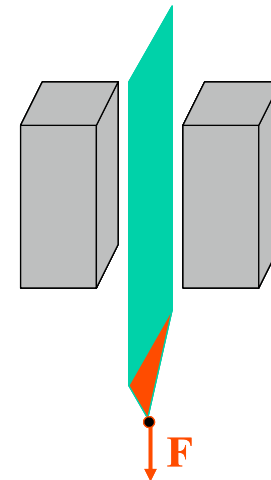
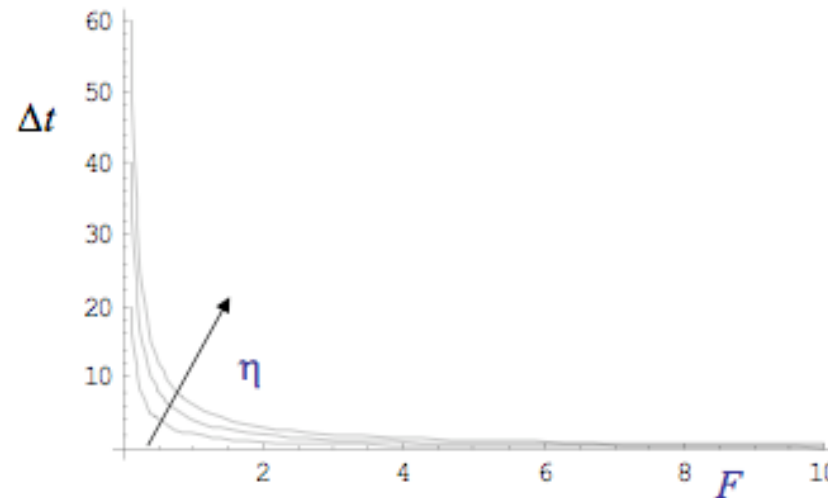
- Diagrammiamo la legge di Ohm per varie resistività



- Data la resistività per ogni V conosco perfettamente la I .
 - Modello deterministico del processo

Esempio 2

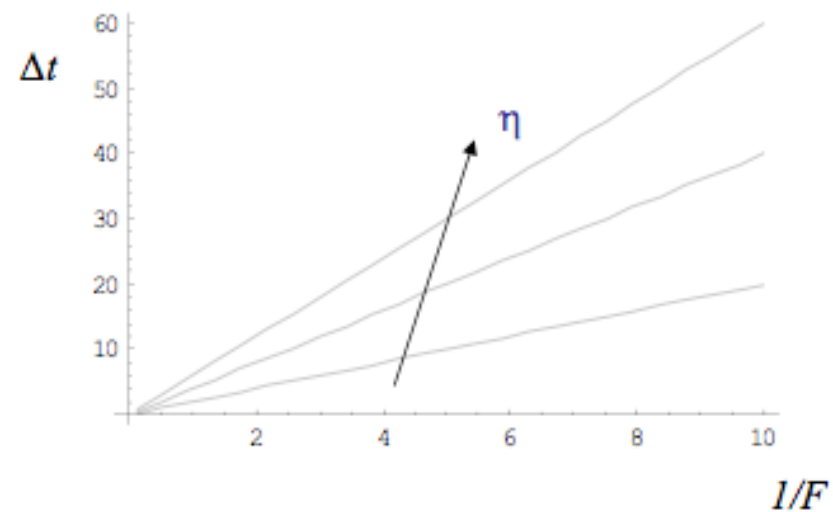
- Vogliamo verificare che un certo liquido si comporti come previsto da Newton e determinare la sua viscosità.
- Usiamo il viscosimetro a banda.
- L'equazione che modella il processo è $\Delta t = \eta \frac{2LS}{\delta} \frac{1}{F}$
 - F condizione sperimentale
 - Δt grandezza misurata



Esempio 2

- Possiamo vedere anche le cose definendo come condizione sperimentale l'inverso della Forza imposta

$$\Delta t = \eta \frac{2LS}{\delta} x$$

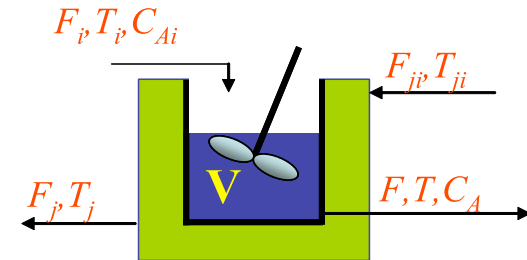


Esempio 3

- CSTR **isotermo**, $A \rightarrow B$ in fase liquida cinetica lineare. Vogliamo caratterizzare la cinetica chimica

$$r = kC_A$$

- Condizione sperimentale c_A (concentrazione di A)
- Grandezza misurata r
- Parametro: Costante cinetica k



- **Ma come misuriamo praticamente r nel CSTR?**
- Dobbiamo modellare il CSTR
 - Il problema è isotermo: NON SERVE UN BILANCIO ENTALPICO
 - La fluidodinamica è ipotizzata (PERFETTA MISCELAZIONE) NON SERVE UN BILANCIO DI QUANTITA' DI MOTO
 - Possiamo solo scrivere bilanci di materia

Esempio 3

- Ipotesi
 - Livello liquido nel CSTR costante: portata in ingresso = portata in uscita. Il bilancio di materia globale è inutile.
- Bilancio di materia sul componente A in **condizioni stazionarie**
 - Q : portata volumetrica
 - V : volume del reattore
 - c_{Ai} : Concentrazione in alimentazione

$$0 = \underbrace{FC_{Ai}}_{\text{Ingresso}} - \underbrace{FC_{Au}}_{\text{Uscita}} - \underbrace{Vr}_{\text{Generazione}}$$

$$r = \frac{F}{V}(C_{Ai} - C_{Au})$$

- Se misuriamo la concentrazione in uscita conosciamo sperimentalmente r .

Esempi

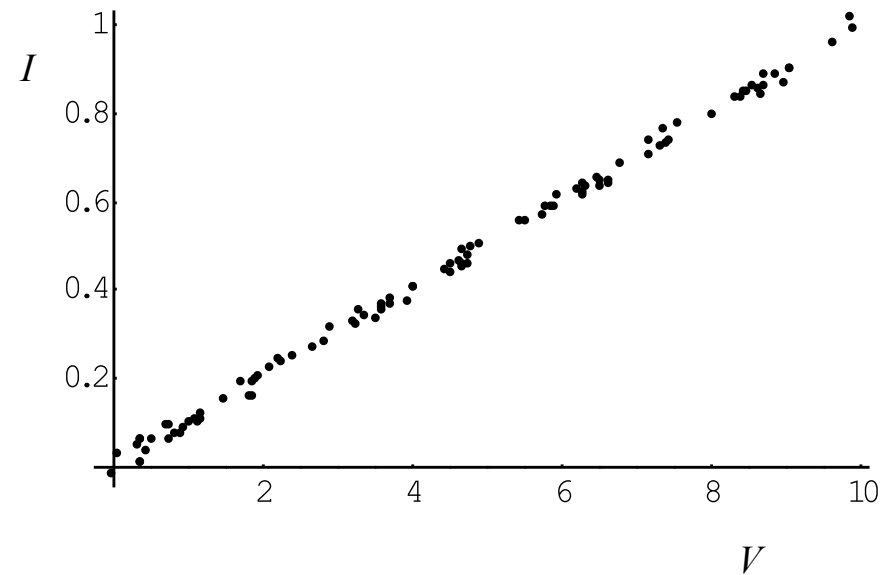
- Tutti e tre i modelli dei processi sono lineari nel parametro incognito.
- Non ci preoccupa una eventuale non linearità nella condizione sperimentale (vedi $1/F$).
- In generale quindi potremmo considerare i tre modelli analoghi, e scrivibili in modo generale come:

$$y = \beta_0 + \beta_1 x$$

- **Nei tre casi considerati l'intercetta β_0 è nulla.**
- Ovviamente noi dobbiamo pervenire ad una stima dei parametri (β_0 e β_1) attraverso la sperimentazione

Esempio I

- Diagrammiamo i dati sperimentali per



- I punti sperimentali ovviamente non giacciono su una retta.
 - Presenza di errori casuali

Usi ed Abusi delle regressioni

USI

- Previsioni
- Interpolazioni
- Fitting di dati
- Verifica dell'effetto di variabili indipendenti su una variabile dipendente
- Determinazione di valori della var. ind. Che massimizzano o minimizzano una risposta
- Controllo di una variabile dipendente attraverso una selezionata variabile indipendente

ABUSI

- Estrapolazione
- Cattiva interpretazione di relazioni spurie
- Sovrastima sulla validità di modelli empirici

Modello dell'esperimento

- Torniamo ad esaminare il modo in cui abbiamo presentato l'esperimento aleatorio. In particolare consideriamo il caso in cui le condizioni sperimentali siano deterministiche.



- Questa descrizione taglia fuori il problema del CSTR. Perché?
- Il singolo esperimento è trattabile con una VA Y (di che tipo?):

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

Deterministico ← $\beta_0 + \beta_1 x$ ← Aleatorio

Modello dell'esperimento

- In questo modello Y è la variabile dipendente (aleatoria), x è la variabile indipendente (non aleatoria)
- ε è l'errore casuale (o residuo)
 - Ammontare della variazione in Y non valutata dalla relazione lineare
- I parametri β_0 e β_1 sono detti coefficienti della regressione e sono grandezze da stimare
- **Assumiamo che la VA ε sia Gaussiana a media zero e con varianza indipendente dalle condizioni sperimentali σ^2 .**
- Quindi, il legame **lineare** tra la VA ε e la VA Y impone che

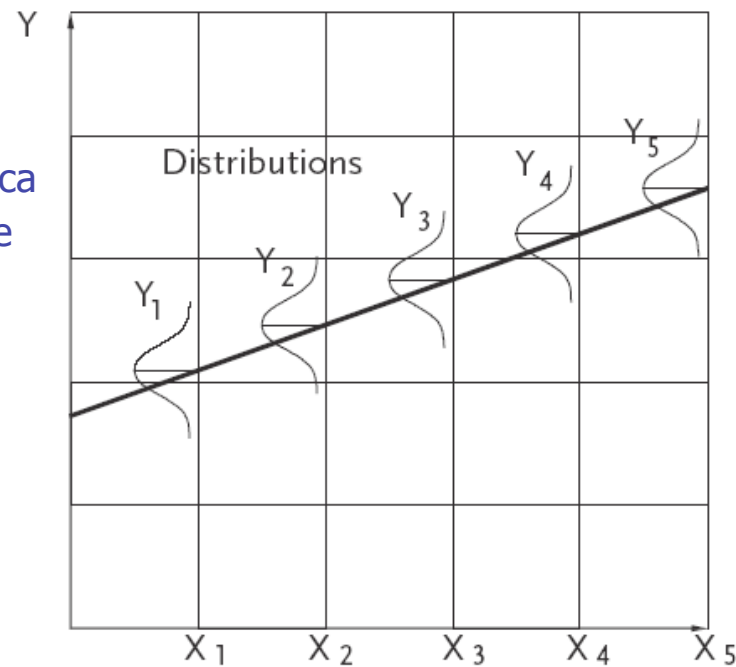
$$Y = N(\beta_0 + \beta_1 x, \sigma^2)$$

Modello dell'esperimento

- Il valore atteso della VA Y è:

$$E(Y) = E(\beta_0 + \beta_1 x + \varepsilon) = \beta_0 + \beta_1 x$$

Interpretazione geometrica
della regressione lineare



Modello dell'esperimento

- Supponiamo di avere N dati (x_i, y_i)
- L'obiettivo è stimare β_0 e β_1 ovvero la VA Y in termini delle N osservazioni
- Se $\hat{\beta}_0$ e $\hat{\beta}_1$ denotano le stime di β_0 e β_1 allora la stima del valore atteso di Y è

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

- Dobbiamo distinguere tra parametri della popolazione e quelli stimati a partire dalle proprietà del campione
- Attraverso i dati raccolti calcoliamo $\hat{\beta}_0$ e $\hat{\beta}_1$ che saranno stime dei parametri della popolazione β_0 e β_1

$$Y_i = \hat{Y}_i + e_i$$

Inferenza parametrica

Residuo

- Il **residuo** osservato è $e_i = Y_i - \hat{Y}_i$
- Come possiamo ottenere le stime $\hat{\beta}_0$ e $\hat{\beta}_1$?
 - Manipolando i dati sperimentali
- Siamo pronti per fare il salto nel mondo della inferenza parametrica.
- Stimare statisticamente significa attribuire un valore “plausibile” ad una grandezza (parametro) non misurabile esattamente.
- Stima puntuale: determinare i parametri della distribuzione sulla base dei dati sperimentali e di una ipotesi sul tipo di distribuzione dell'errore casuale.
 - Si parla di stima puntuale perché per il momento ci limiteremo a determinare singoli valori dei parametri.
- Nell'esempio che stiamo considerando dovremmo essere capaci di dare un valore a $\hat{\beta}_0$ e $\hat{\beta}_1$ ed anche a σ^2 .

Statistiche

- Un campione casuale di dimensione n di una variabile aleatoria X è un vettore di VA con elementi $X_1, \dots, X_n$ indipendenti la cui distribuzione è identica a quella di X .

$$\bar{X} := \sum_{k=1}^n \frac{X_k}{n}$$

- La VA media del campione è una statistica. Essa è caratterizzata da una distribuzione che dipende dalle distribuzioni delle X_i .
 - Se le X_i sono Gaussiane la VA media del campione è Gaussianale. Se le X_i non sono Gaussiane ma n è grande il teorema del limite centrale garantisce la Gaussianità della VA media del campione

$$S^2 := \sum_{k=1}^n \frac{(X_k - \bar{X})^2}{n-1}$$

- La VA varianza del campione è una statistica caratterizzata da una distribuzione che dipende dalle proprietà della VA media del campione e dalle proprietà delle X_i .

Stima e Stimatore

- La statistica (ovvero la funzione che applicheremo ai dati) per pervenire alla stima viene detto **STIMATORE**.
- Tra **STIMATORE** e **STIMA** c'è lo stesso rapporto che abbiamo visto esistere tra VA ed esito dell'esperimento.
 - Posso considerare lo Stimatore senza effettivamente svolgere gli esperimenti.
 - La stima è la singola determinazione dello stimatore $g(y_1, y_2, \dots, y_n)$ essendo $(y_1, y_2, \dots, y_n)$ un campione osservato cioè una determinazione di $(Y_1, Y_2, \dots, Y_n)$
- **Lo Stimatore è una variabile aleatoria** che comprende tutti i valori che la stima può avere.
- La VA media del campione è uno stimatore della media della popolazione (ignota). Allo stesso modo la varianza del campione è uno stimatore della varianza della popolazione.
- Il valore medio degli esiti degli esperimenti, e la loro varianza sono stime puntuali di μ e σ^2 .

Stimatore

- Per comodità useremo la seguente nomenclatura:
 - Parametro θ
 - Valore vero del parametro θ
 - Stimatore del parametro Θ
 - Stima del parametro $\hat{\theta}$
- **Non esiste un unico Stimatore per un singolo parametro.**
 - Pensate a media e mediana come possibili misure della tendenza centrale di una VA.
- Possiamo caratterizzare le proprietà della VA Stimatore in modo da poter scegliere quella che potremmo considerare ottimale in qualche senso.
 - Questa frase può sembrare un po' oscura e vuota di senso ma vedremo rapidamente cosa significa.

Proprietà degli Stimatori

Imparzialità

- Uno stimatore si dice imparziale (unbiased) se il suo valore atteso coincide con il valore vero del parametro: $E(\Theta) = \theta$
 - Sebbene il valore vero non sarà mai noto è possibile valutare il verificarsi della imparzialità.
 - Si definisce distorsione (bias) di uno stimatore: $\text{Bias}[\Theta] = E[\Theta] - \theta$

Efficienza

- È una misura della varianza dello stimatore. Se dispongo di più stimatori devo scegliere quello con varianza minima ovvero quello con la massima efficienza.

Consistenza:

- È una proprietà dello stimatore al variare del numero di prove sperimentali. Uno stimatore si dice consistente se: $\lim_{N \rightarrow \infty} \hat{\theta} = \theta$
 - Convergenza in probabilità

Stimatore

- Uno **Stimatore consistente** non è necessariamente imparziale.
- La varianza del campione è uno stimatore consistente della varianza della popolazione ma come vedremo esso non è imparziale.
- Consideriamo una situazione molto semplice che abbiamo già visto spesso. Misuriamo la viscosità di un polimero nel viscosimetro a nastro applicando sempre lo stesso peso e sempre alla stessa temperatura.
- Lo stimatore della viscosità può essere la VA media del campione

$$E[\bar{X}] = E\left[\sum_{k=1}^n \frac{X_k}{n}\right] = \frac{1}{n} \sum_{k=1}^n E[X_k] \stackrel{\text{i.d.}}{=} \frac{1}{n} \sum_{k=1}^n \mu = \frac{1}{n} n\mu = \mu.$$

Lo stimatore è imparziale

Errore medio quadratico

- L'errore medio quadratico (mean square error) di uno stimatore Θ di un parametro θ incognito di una popolazione è dato da

$$MSE[\Theta] = E[(\Theta - \theta)^2]$$

- Si vede facilmente che

$$MSE[\Theta] = VAR[\Theta] + (Bias[\Theta])^2$$

- Quindi se lo stimatore è imparziale $MSE[\Theta] = VAR[\Theta]$
- **EFFICIENZA:** Se Θ_1 e Θ_2 sono due stimatori per lo stesso parametro allora Θ_1 è relativamente più efficiente (o semplicemente migliore) se

$$MSE[\Theta_1] < MSE[\Theta_2]$$

Regressione

- Abbiamo visto negli esempi relazioni funzionali tra variabili
 - Possiamo anche accorgerci dai dati della sussistenza di un rapporto funzionale.
- Una regressione tenta di ricostruire il legame funzionale tra due grandezze
- Per il momento ci limitiamo a legami che coinvolgano linearmente i parametri.
 - La non linearità della variabile indipendente non conta, e può essere risolta cambiando variabile.
- Quindi ci occuperemo per ora di **regressioni lineari**

Stimatore da MASSIMA VEROSIMIGLIANZA

- La **VEROSIMIGLIANZA** di avere come esito di n esperimenti $y_1, y_2, \dots, y_n$ è la densità di probabilità corrispondente a quei valori fissato il valore dei parametri della distribuzione

$$L(\theta; y_1, y_2, \dots, y_n) = f_Y(y_1, y_2, \dots, y_n; \theta)$$

- Tenendo fissi i valori dei dati **la stima di Massima Verosimiglianza (MV)** è quella che si ottiene scegliendo il valore dei parametri che massimizza la Verosimiglianza:

$$\hat{\beta}_{MV} = \max_{\beta} L(\beta; y_1, \dots, y_n)$$

- NB: Il valore dei parametri derivati dalla massimizzazione della verosimiglianza è quello che rende i dati del campione i più probabili, non è il valore più probabile dei parametri sulla base dei dati del campione.
- Aspetto tecnico: Massimizzare L equivale a massimizzare il logaritmo di L , cosa in genere più facile da fare.

Stimatore MV

- Dal punto di vista pratico lo stimatore basato sul **criterio della massima verosimiglianza** si è mostrato il migliore di quelli conosciuti.
- **Tale stimatore è consistente, è il più efficiente degli stimatori conosciuti, ma non è imparziale.**
- Il criterio della massima verosimiglianza è **arbitrario**.
- In alcuni casi lo stimatore basato sulla MV coincide con quello basato sui minimi quadrati (MQ).

Parametri di una Gaussiana

- Supponiamo che $y_1, y_2, \dots, y_n$ siano le determinazioni di un campione casuale di una $N(\mu, \sigma^2)$ con media e varianza incognite.

- Quale è la stima basata sulla MV di tali parametri?

-

$$L[\theta; y_1, \dots, y_n] = f_{\mathbf{Y}}(y_1, \dots, y_n, \mu, \sigma)$$

$$f_{\mathbf{Y}}(y_1, \dots, y_n, \mu, \sigma) = \prod_{i=1}^n f_{Y_i}(y_i, \mu, \sigma)$$

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

$$\log(f_{Y_i}) = -\log(\sqrt{2\pi}) - \log(\sigma) - \frac{1}{2} \frac{(y_i - \mu)^2}{\sigma^2}$$

$$\log(L) = -n \log(\sqrt{2\pi}) - n \log(\sigma) - \frac{1}{2\sigma^2} \left[(y_1 - \mu)^2 + (y_2 - \mu)^2 \dots + (y_n - \mu)^2 \right]$$

Parametri di una Gaussiana

- Le derivate rispetto ai parametri sono:

$$\frac{\partial \log(L)}{\partial \mu} = \frac{1}{\sigma^2} [(y_1 - \mu) + (y_2 - \mu) \dots + (y_n - \mu)] = \frac{n}{\sigma^2} (\bar{y} - \mu)$$

$$\frac{\partial \log(L)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} [(y_1 - \mu)^2 + (y_2 - \mu)^2 \dots + (y_n - \mu)^2] = -\frac{n}{\sigma^3} \left(\sigma^2 - \frac{1}{n} \sum_{i=1}^n (y_i - \mu)^2 \right)$$

- Eguagliando a zero le derivate e risolvendo il sistema si ottiene:

$$\hat{\mu} = \bar{y}$$
$$\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}$$

- La media del campione è la stima basata sulla massima verosimiglianza della media della popolazione.
- La deviazione standard del campione basata su n gradi di libertà è la stima di MV della deviazione standard della popolazione

Stimatore MV

ESEMPIO

- Stima con la MV della viscosità con il viscosimetro a banda.
- Assumeremo che la f_Y abbia una certa forma che poi andrà verificata.
- Modello dell'esperimento:

$$g(x, \underline{\beta}) = \Delta t = \eta \frac{a}{F} + b$$

$$Y_i = \Delta t_i + \epsilon_i$$

$$\underline{x} = \begin{pmatrix} \dots \\ \frac{a}{F} \\ \dots \end{pmatrix} \quad \underline{\beta} = \begin{pmatrix} \eta \\ b \end{pmatrix}$$

$$\underline{Y} = g(\underline{x}, \underline{\beta}) + \underline{\epsilon}$$

Ovviamente ci aspettiamo che b sia nullo

Stimatore MV

- Con la sperimentazione faremo N misure. La VA Y sarà caratterizzata da una congiunta f_Y .
- Supponiamo che $\varepsilon_i := N(0, \sigma^2)$.
- Stiamo ipotizzando che ciascun esperimento sia caratterizzato dalla stessa precisione e che non commettiamo errori sistematici (media zero).
- Quindi la VA Y_i : $Y_i := N(g(x_i, \underline{\beta}), \sigma^2)$

Stimatore MV

- La precedente distribuzione è una delle marginali.
- Se la sperimentazione è stata condotta eseguendo esperimenti indipendenti la conoscenza delle marginali comporterà la conoscenza della congiunta.
- Ricapitoliamo le ipotesi fatte:

Modello del processo vero

$$Y_i = g(x_i, \underline{\beta}) + \varepsilon_i \quad \text{Modello dell'esperimento}$$

$$\varepsilon_i := N(0, \sigma^2) \quad \text{Tipo di esperimento}$$

$$\varepsilon_i \text{ indipendente da } \varepsilon_j \quad \text{Tipo di esperimento}$$

Stimatore MV

- Ne consegue che la congiunta sarà:

$$f_Y = \prod_{i=1}^N f_{Y_i} = \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2} \frac{(y_i - g(x_i, \underline{\beta}))^2}{\sigma^2} \right]$$

- La funzione verosimiglianza è:

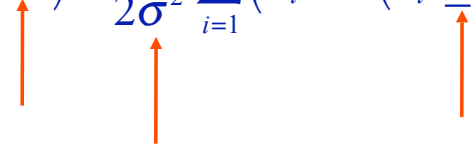
$$L(\underline{\beta}, \sigma^2) = f_Y(\underline{y})$$

- Una volta effettuati gli esperimenti, tale funzione dipende solo dai parametri e dalla varianza sperimentale. In genere la massimizzazione si effettua sul logaritmo della funzione L

$$\max_{\underline{\beta}, \sigma^2} L(\underline{\beta}, \sigma^2) = \max_{\underline{\beta}, \sigma^2} \log [L(\underline{\beta}, \sigma^2)]$$

Stimatore MV

- Con un po' di algebra si ottiene:

$$\log(f_Y) = -\frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - g(x_i, \underline{\beta}))^2$$


- **NB: Questa funzione dipende dalle ipotesi che abbiamo fatto.**
- Cominciamo lo studio dalla stima della varianza sperimentale.
 - Supponiamo di aver già determinato il massimo della L al variare del vettore dei parametri e di conoscere quindi la stima di θ .
- Massimizziamo L rispetto alla varianza:

$$\frac{d}{d\sigma^2} \log(L) = -\frac{N}{2} \frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^N (y_i - g(x_i, \hat{\underline{\beta}}))^2$$

$$\frac{d}{d\sigma^2} \log(L) = 0$$

Stimatore MV

- Per la varianza si ottiene quindi:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^N \left(y_i - g(x_i, \hat{\underline{\beta}}) \right)^2}{N} \quad \text{STIMA}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^N \left(Y_i - g(x_i, \underline{B}) \right)^2}{N} \quad \text{STIMATORE}$$

- C'è una dipendenza implicita da Y_i anche nello stimatore dei parametri.
- $\hat{\sigma}^2$ è una **variabile aleatoria**. Possiamo quindi considerarne sia il valore atteso che la varianza, che per la massima verosimiglianza sarà la più piccola possibile. Il tipo di VA ricorda una χ^2 .

Stimatore MV

STIMA DEI PARAMETRI

- Il modello del processo è **lineare** nei parametri.
- Ricordiamo le ipotesi fatte

Modello del processo vero

$$Y_i = g(x_i, \underline{\beta}) + \varepsilon_i \quad \text{Modello dell'esperimento}$$

$$\varepsilon_i \coloneqq N(0, \sigma^2) \quad \text{Tipo di esperimento}$$

$$\varepsilon_i \text{ indipendente da } \varepsilon_j \quad \text{Tipo di esperimento}$$

$$\log(f_Y) = -\frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - g(x_i, \underline{\beta}))^2$$

**Il massimo di questo termine
corrisponde al minimo della
somma degli scarti quadratici**

Stimatore MV

- Il massimo della funzione verosimiglianza al variare dei parametri corrisponde al minimo della funzione:

$$\Phi(\underline{\beta}) = \sum_{i=1}^N [y_i - g(x_i, \underline{\beta})]^2 = \sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_i)^2$$

MINIMI QUADRATI!

- La somma non ha i pesi perché abbiamo ipotizzato che tutte le varianze siano uguali.

Metodo dei Minimi quadrati

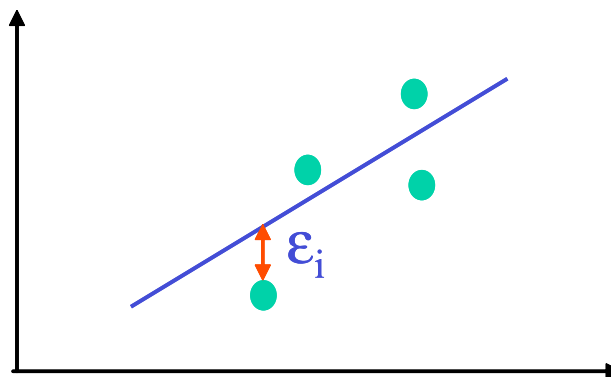
- Il **metodo dei minimi quadrati** (Least squares)
- Questo metodo minimizza la somma degli scarti quadratici tra i valori previsti ed i valori misurati della variabile dipendente.
- Il metodo si basa sul principio che i migliori stimatori per β_0 and β_1 siano quelli che minimizzano la somma dei quadrati degli errori, SS_E :

$$SS_E = \sum_{i=1}^N e_i^2 = \sum_{i=1}^N (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

Stimatore Minimi Quadrati

- Gauss, Theoria Motus, 1809: *“Il valore più probabile dei parametri da stimare sarà quello per cui la somma dei quadrati delle differenze tra i valori osservati ed i valori calcolati moltiplicato per il grado di precisione della misura sia minimo”*
- Il metodo di Gauss dei minimi quadrati è utilizzabile quando
 - Le condizioni sperimentali sono fissate con precisione assoluta e non affette da errori.
 - Il modello è lineare nei parametri
 - Le misure sperimentali sono statisticamente indipendenti
- Come vedremo effettuare una regressione coinvolge anche sempre lo svolgimento di test statistici.
- In realtà il metodo dei MQ non è sempre la migliore scelta.

Modello dell'esperimento



- L'errore è tutto nella variabile dipendente

Ricapitoliamo

- Supponiamo che $\varepsilon_i := N(0, \sigma^2)$.
 - Stiamo ipotizzando che ciascun esperimento sia caratterizzato dalla stessa precisione e che non commettiamo errori sistematici (media zero).

$$Y_i := N(\beta_0 + \beta_1 x, \sigma^2)$$

Stima con i Minimi Quadrati

- Il minimo della funzione:

$$SS_E = \sum_{i=1}^N e_i^2 = \sum_{i=1}^N (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

si determina analiticamente ovvero:

$$\begin{cases} \frac{\partial SS_E}{\partial \hat{\beta}_0} = 0 \\ \frac{\partial SS_E}{\partial \hat{\beta}_1} = 0 \end{cases} \Leftrightarrow \begin{cases} N\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^N x_i = \sum_{i=1}^N y_i \\ \hat{\beta}_0 \sum_{i=1}^N x_i + \hat{\beta}_1 \sum_{i=1}^N x_i^2 = \sum_{i=1}^N y_i x_i \end{cases}$$

- La soluzione del sistema fornisce le stime di $\hat{\beta}_0$ e $\hat{\beta}_1$:

$$\hat{\beta}_0 = \frac{\sum_{i=1}^N y_i}{N} - \bar{x} \hat{\beta}_1, \quad \hat{\beta}_1 = \frac{\sum_{i=1}^N (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2}$$

Stima con i Minimi Quadrati

- Le precedenti relazioni determinano le stime puntuali dei parametri una volta effettuati gli esperimenti. D'altra parte prima della sperimentazione le stesse espressioni rappresentano gli stimatori:

$$B_0 = \frac{\sum_{i=1}^N Y_i}{N} - \bar{x} B_1,$$

$$B_1 = \frac{\sum_{i=1}^N (Y_i - \bar{Y})(x_i - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2}$$

- Entrambi gli stimatori sono funzioni lineari delle Y , quindi se le Y sono VA Normali anche gli stimatori saranno tali.**
- E' anche evidente che i due stimatori **non** sono tra loro indipendenti.

Stimatore Minimi Quadrati

- Se non eseguiamo la sperimentazione possiamo considerare lo stimatore ovvero la VA e verificarne le proprietà.
- **Imparzialità:**

$$\begin{aligned} E(B_1) &= E\left(\frac{\sum_{i=1}^N (\beta_0 + \beta_1 x_i + \varepsilon_i)(x_i - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2}\right) = E\left(\frac{\sum_{i=1}^N (\beta_0 + \beta_1 x_i)(x_i - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2}\right) + E\left(\frac{\sum_{i=1}^N \varepsilon_i (x_i - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2}\right) = \\ &= E\left(\frac{\beta_0 \sum_{i=1}^N (x_i - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2}\right) + E\left(\frac{\beta_1 \sum_{i=1}^N x_i (x_i - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2}\right) = \frac{\beta_1 \sum_{i=1}^N x_i (x_i - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2} = \beta_1 \frac{\sum_{i=1}^N x_i (x_i - \bar{x}) - \bar{x} \sum_{i=1}^N (x_i - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2} = \beta_1 \end{aligned}$$

IMPARZIALE

- Lo stimatore di β_1 è quindi imparziale.
- Analogamente si può dimostrare l'imparzialità dello stimatore per β_0 .

Stimatore Minimi Quadrati

- Il modello dell'esperimento è:
 - x deterministiche
 - Y Gaussiane con media $\beta_0 + \beta_1 x$ e varianza uguale a quella dell'errore sperimentale σ^2
 - Varianza dell'errore sperimentale indipendente da x .
 - Esperimenti indipendenti
- Consideriamo complessivamente gli esperimenti
 - La VA Y sarà caratterizzata da una congiunta f_Y
 - La congiunta sarà il prodotto delle f di ciascun esperimento (perché?)

$$\begin{aligned}g(x, \beta) &= \beta_0 + \beta_1 x \\ Y_i &= g(x_i, \beta) + \varepsilon_i \\ \mathbf{x} &= \begin{bmatrix} \dots \\ x_i \\ \dots \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \\ Y &= g(\mathbf{x}, \beta) + \varepsilon\end{aligned}$$

$$f_Y = \prod_{i=1}^N f_{Y_i} = \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2} \frac{(y_i - g(x_i, \beta))^2}{\sigma^2} \right]$$

Stimatore Minimi Quadrati e Massima Verosimiglianza

- Una volta effettuati gli esperimenti, tale funzione dipende solo dai parametri e dalla varianza sperimentale.
- Tenendo fissi i valori dei dati la stima di Massima Verosimiglianza (MV) è quella che si ottiene scegliendo il valore dei parametri ($\beta_0, \beta_1, \sigma^2$) che massimizza la Verosimiglianza.
 - Nelle condizioni che stiamo adottando per la modellazione dell'esperimento tale criterio coincide con la stima basata sui minimi quadrati.
 - Non sempre le due stime coincidono
 - Il valore dei parametri derivati dalla massimizzazione della verosimiglianza è quello che rende i dati del campione i più probabili, non è il valore più probabile dei parametri sulla base dei dati del campione.
 - Dal punto di vista pratico lo stimatore basato sul criterio della massima verosimiglianza si è mostrato il migliore tra quelli conosciuti.
 - Tale stimatore è consistente, è il più efficiente degli stimatori conosciuti, ma non è sempre imparziale.

Stimatore Minimi Quadrati e Massima Verosimiglianza

- Per la stima della varianza sperimentale.
 - Conosciamo già la stima di β .
- Massimizziamo la verosimiglianza rispetto alla varianza:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^N (y_i - g(x_i, \hat{\beta}))^2}{N} \quad \text{STIMA}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^N (Y_i - g(x_i, \mathbf{B}))^2}{N} \quad \text{STIMATORE}$$

- Si può dimostrare che $E(\hat{\sigma}^2) = \frac{N-2}{N} \sigma^2$

– **Lo stimatore per la varianza non è imparziale!**

– Uno stimatore imparziale è:
$$s^2 = \frac{\sum_{i=1}^N (Y_i - g(x_i, \mathbf{B}))^2}{N-2} = \frac{SS_E}{N-2}$$

Riassumendo

- Ipotesi 1: La relazione tra x ed y è lineare
- Ipotesi 2: La variabile indipendente non cambia se l'esperimento è replicato
- La dimensione del campione è grande rispetto al numero dei parametri
- Proprietà dell'errore
 - Media zero
 - Le Varianze sono uguali (omoschedasticità)
 - Se non è così ci sono dati più “pesanti” di altri: MQP
 - Indipendenza ij
 - Problemi con le serie temporali
 - Il modello è adeguato
 - L'errore è gaussiano

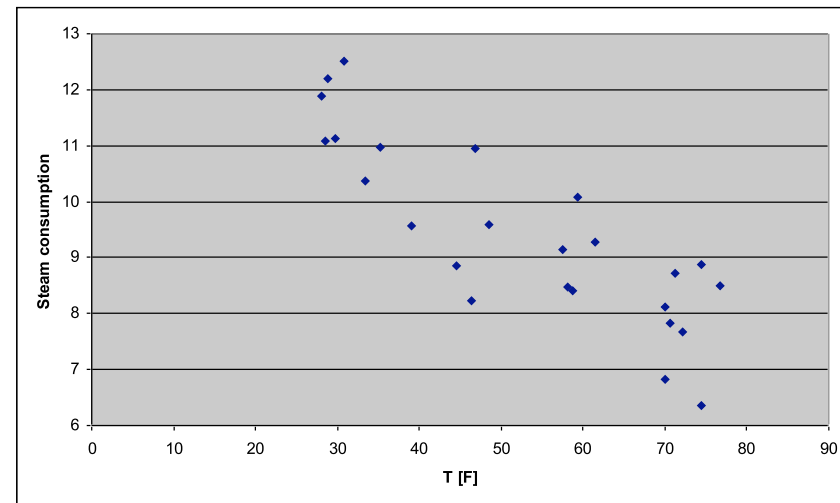
Regressione lineare con Excel

- Selezionare Strumenti (Tools) dal menu
- Selezionare Analisi Dati (Data analysis)
- Selezionare Regressione
- Selezionare un intervallo per le y (Compreso l'eventuale titolo)
- Selezionare un intervallo per le x (Compreso l'eventuale titolo)
- Click sul radio button dopo Etichette
- Selezionare uno spazio vuoto per l'Output
- Procedere

Regressione lineare con Excel

- Esercizio

	T [F]	Steam consumption	
1	35.3	10.98	
2	29.7	11.13	
3	30.8	12.51	
4	58.8	8.4	
5	61.4	9.27	
6	71.3	8.73	
7	74.4	6.36	
8	76.7	8.5	
9	70.7	7.82	
10	57.5	9.14	
11	46.4	8.24	
12	28.9	12.19	
13	28.1	11.88	
14	39.1	9.57	
15	46.8	10.94	
16	48.5	9.58	
17	59.3	10.09	
18	70	8.11	
19	70	6.83	
20	74.5	8.88	
21	72.1	7.68	
22	58.1	8.47	
23	44.6	8.86	
24	33.4	10.36	
25	28.6	11.08	



Regressione lineare con Excel

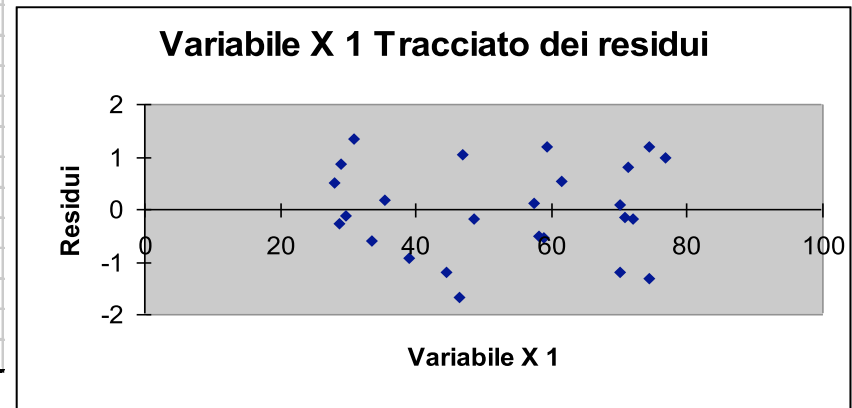
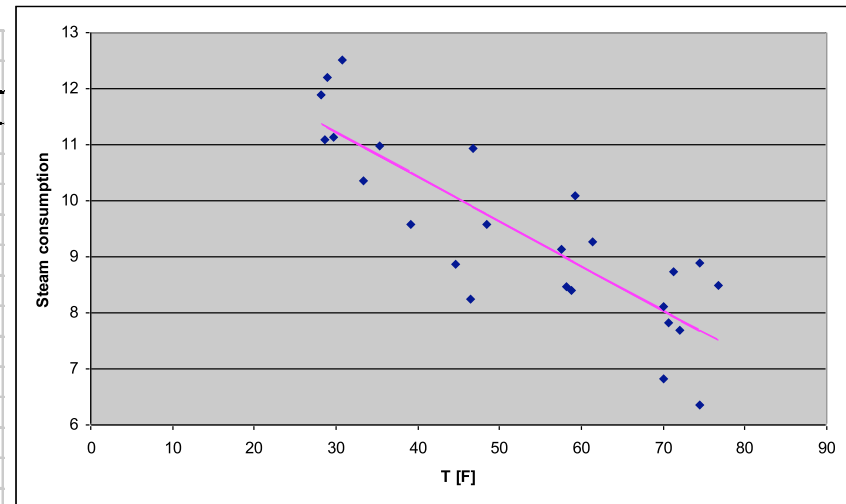
- Stima
 - Le celle in giallo contengono le stime dei due parametri: Intercetta e pendenza del modello lineare

OUTPUT RIEPILOGO								
<i>Statistica della regressione</i>								
R multiplo	0.84524406							
R al quadrato	0.714437521							
R al quadrato corretto	0.702021761							
Errore standard	0.890124516							
Osservazioni	25							
ANALISI VARIANZA								
	<i>gdl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>Significatività F</i>			
Regressione	1	45.59240195	45.59240195	57.54279428	1.05495E-07			
Residuo	23	18.22339805	0.792321654					
Totale	24	63.8158						
	<i>Coefficienti</i>	<i>Errore standard</i>	<i>Stat t</i>	<i>Valore di significatività</i>	<i>Inferiore 95%</i>	<i>Superiore 95%</i>	<i>Inferiore 95.0%</i>	<i>Superiore 95.0%</i>
Intercetta	13.62298927	0.581463494	23.42879545	1.49679E-17	12.42014039	14.82583814	12.42014039	14.82583814
Variabile X 1	-0.079828693	0.010523581	-7.585696691	1.05495E-07	-0.101598379	-0.058059008	-0.101598379	-0.058059008

Regressione lineare con Excel

- Previsioni del modello e residui

OUTPUT RESIDUI			
Osservazione	Y prevista	Residui	Residui standard
1	10.80503639	0.174963606	0.200788419
2	11.25207708	-0.122077077	-0.140095783
3	11.16426551	1.345734486	1.54436632
4	8.929062101	-0.529062101	-0.607152228
5	8.721507499	0.548492501	0.629450575
6	7.931203435	0.798796565	0.916699783
7	7.683734486	-1.323734486	-1.519119097
8	7.500128491	0.999871509	1.147453602
9	7.979100651	-0.159100651	-0.182584076
10	9.032839403	0.107160597	0.122977615
11	9.918937899	-1.678937899	-1.92675091
12	11.31594003	0.874059969	1.003072145
13	11.37980299	0.500197014	0.574026623
14	10.50168736	-0.93168736	-1.0692054
15	9.887006421	1.052993579	1.208416546
16	9.751297643	-0.171297643	-0.196581356
17	8.889147755	1.200852245	1.378099308
18	8.034980736	0.075019264	0.086092186
19	8.034980736	-1.204980736	-1.382837169
20	7.675751616	1.204248384	1.38199672
21	7.86734048	-0.18734048	-0.214992134
22	8.984942187	-0.514942187	-0.590948199
23	10.06262955	-1.202629546	-1.380138941
24	10.95671091	-0.596710912	-0.684786074
25	11.33988864	-0.259888639	-0.298248478



STIMATORI - Riparametrizzazione

- Procediamo per il momento sul modello **riparametrizzato**:

$$Y_i = p_0 + p_1 (x_i - \bar{x}) + \varepsilon_i$$

La riparametrizzazione riguarda solo la definizione dei parametri.

- Ovviamente i nuovi parametri sono funzioni lineari dei vecchi parametri β_0 e β_1 :

$$\begin{cases} \beta_1 = p_1 \\ \beta_0 = p_0 - p_1 \bar{x} \end{cases}$$

- Stimare i nuovi parametri è più facile perché le due equazioni lineari cui si perviene sono disaccoppiate, inoltre **i due stimatori sono indipendenti ed imparziali**:

$$P_0 = \frac{\sum_{i=1}^N Y_i}{N}, \quad P_1 = \frac{\sum_{i=1}^N Y_i (x_i - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2}$$

STIMATORI – Varianza stimatori

- Per determinare la varianza degli stimatori appena trovati ricordiamo la seguente proprietà delle variabili aleatorie:

$$Z = aY \quad \sigma_Z^2 = a^2 \sigma_Y^2$$

e

$$\sigma_{Y_1+Y_2}^2 = \sigma_{Y_1}^2 + \sigma_{Y_2}^2$$

- E' quindi facile dedurre: $\sigma_{P_0}^2 = \frac{1}{N} \sigma^2$
- Quindi se cresce il numero di prove sperimentali diminuisce la varianza di tale stimatore (la dipendenza da N è iperbolica quindi ...).
- Inoltre la precisione sull'intercetta è migliore di quella delle singole misure.

STIMATORI – Varianza stimatori

- Per p_1 :
$$\sigma_{P_1}^2 = \frac{1}{\left[\sum_{i=1}^N (x_i - \bar{x})^2 \right]^2} \sum_{i=1}^N \sigma^2 (x_i - \bar{x})^2 = \frac{\sigma^2}{\sum_{i=1}^N (x_i - \bar{x})^2}$$

- Per aumentare la precisione di P_1 il denominatore deve crescere quindi le misure devono essere disperse attorno alla media di x .

STIMATORI – Varianza stimatori

- Torniamo alla formulazione originaria

$$\begin{cases} \beta_1 = p_1 \\ \beta_0 = p_0 - p_1 \bar{x} \end{cases}$$

- quindi:

$$\sigma_{B_0}^2 = \sigma_{P_0}^2 + \sigma_{P_1}^2 \bar{x}^2, \quad \sigma_{B_1}^2 = \frac{\sigma^2}{\sum_{i=1}^N (x_i - \bar{x})^2}$$

- **Gli stimatori di a e b non sono indipendenti.** Per la covarianza:

$$\sigma_{B_0 B_1} = E[(B_1 - \beta_1)(B_0 - \beta_0)]$$

STIMATORI – Varianza stimatori

- Come sappiamo la covarianza può essere negativa.
- Le varianze non possono, ovviamente, essere negative. La matrice di covarianza deve essere positiva definita.
- **Va notato che tutte le varianze contengono il valore vero della varianza sperimentale. Per arrivare a stime bisogna sostituire la stima della varianza.**

Stimatore MV – Modello lineare

- IPOTESI

Modello VERO

$$g(x, \underline{p}) = ax + b$$

Modello del processo lineare

$$Y_i = g(x_i, \underline{p}) + \varepsilon_i$$

Modello dell'esperimento

$$\varepsilon_i := N(0, \sigma^2)$$

Tipo di esperimento

ε_i indipendente da ε_j

Tipo di esperimento

- Abbiamo già visto che in queste condizioni per lo Stimatore dei parametri MV e MQ coincidono.

- Stimatori

$$B_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \sigma_{B_1}^2 = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$B_0 = \bar{Y} - B_1 \bar{x} \quad \sigma_{B_0}^2 = \frac{\sigma^2}{n} + \sigma_{B_1}^2 \bar{x}^2$$

Stimatore MV – Modello lineare

- Una volta stimati i parametri possiamo anche usare il modello con i parametri per avere previsioni ed il relativo Stimatore

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$
$$Y = B_0 + B_1 x$$

- Stima della varianza. La MV porta ad uno stimatore parziale (o distorto):

$$\sigma^2 = \frac{\sum_{i=1}^N \left(Y_i - (\beta_0 + \beta_1 x_i) \right)^2}{N}$$

- Una versione imparziale (non distorta) è:

$$s^2 = \frac{\sum_{i=1}^N \left(Y_i - (\beta_0 + \beta_1 x_i) \right)^2}{N - 2}$$

VA Stimatore MV Varianza – Modello lineare

- Che tipo di variabile aleatoria è lo stimatore della varianza?
- Si può dimostrare che lo stimatore imparziale della varianza è proporzionale ad una χ^2_{n-2}
- La dimostrazione non viene qui riportata. Essa si basa sul teorema della partizione di una χ^2

VA Stimatore MV Varianza – Modello lineare

- Proviamo a costruire una chi-quadro che somigli allo stimatore della varianza:

$$\chi_N^2 = \frac{\sum_{i=1}^N (Y_i - \beta_0 - \beta_1 x_i)^2}{\sigma^2}$$

Valori veri

Costante

- Questa chi-quadro pur assomigliando alla s^2 è diversa da questa perché vi compaiono **i valori veri del modello** del processo e non quelli stimati.
- Il teorema della partizione permette di ripartire questa chi-quadro in modo che un pezzo sia della ripartizione coincida con la nostra s^2 .

VA Stimatore MV Varianza – Modello lineare

- Il teorema della partizione garantisce che:

$$Q_1 = \chi_{N-2}^2 = \frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{\sigma^2}$$

$$Q_2 = \chi_2^2 = \frac{\sum_{i=1}^N (\hat{Y}_i - B_0 + B_1 x_i)^2}{\sigma^2}$$

- Ed inoltre garantisce che Q_1 sia indipendente da Q_2 (conclusione non banale)
- Infine il teorema garantisce che lo stimatore della varianza sia indipendente da quelli dei parametri.

VA Stimatore MV Varianza – Modello lineare

- Quindi, usando Q_1 possiamo studiare lo stimatore della varianza:

$$s^2 = \frac{\sum_{i=1}^N (Y_i - B_0 - B_1 x_i)^2}{N-2} = \frac{\sigma^2}{N-2} \frac{\sum_{i=1}^N (Y_i - B_0 - B_1 x_{ii})^2}{\sigma^2} = \frac{\sigma^2}{N-2} \chi_{N-2}^2$$

– ovvero s^2 è una VA proporzionale ad una chi-quadro a $N-2$ gradi di libertà.

- Possiamo ora facilmente verificare che s^2 sia uno stimatore unbiased:

$$E(s^2) = \frac{\sigma^2}{N-2} E(\chi_{N-2}^2) = \frac{N-2}{N-2} \sigma^2$$

Concetti importanti

- Modelli lineari nei parametri
- Modello dell'esperimento
- Modello per l'errore sperimentale
- Inferenza parametrica
- Stimatore e stima
- Proprietà degli stimatori
- Minimi quadrati
- Massima verosimiglianza
- Varianza degli stimatori
- Tipo VA varianza sperimentale