REGRESSIONE → cercare di capire che relazione c'è tra alcuni dati che controllo in entrata e l'uscita che misuro con un certo errore → $X_1 \dots X_N$ in entrata e $Y = f(X_1, \dots, X_N)$ in uscita

↓

problema: determinare f come forma oppure i suoi coefficienti → REGRESSIONE LINEARE → f funzione lineare

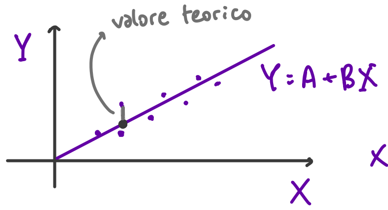
↓

REGRESSIONE LINEARE SEMPLICE → $N = 1$ (una sola v.c. d'ingresso)

$Y = \alpha + \beta X$ in teoria

$Y = \alpha + \beta X$ + errore in pratica → α e β incogniti

fare esperimento più volte, misurare i dati in uscita e cercare di trovare una retta (caso lineare) o una curva (non-lineare)



X_u sono valori non v.c. ← non media campionaria!

per scegliere la retta che interpola meglio i dati usiamo due stimatori → per considerare le distanze tra valore teorico e valore effettivamente misurato

↓

somma dei quadrati delle distanze tra valore misurato e punto ∈ retta

↓

$$SS^2 = \sum_{u=1}^{M} (Y_u - (A + BX_u))^2$$

A → α, B → β: scelti in modo da minimizzare SS

↓

derivare rispetto ad A e portarla a 0, poi stessa cosa x B

NOTAZIONE

$$\bar{X} = \sum_{u=1}^{M} \frac{X_u}{M} \qquad \bar{Y} = \sum_{u=1}^{M} \frac{Y_u}{M}$$

$$\begin{cases} \dfrac{\partial SS^2}{\partial A} = 0 \\ \dfrac{\partial SS^2}{\partial B} = 0 \end{cases} = \begin{cases} \sum_{u=1}^{M} 2(Y_u - A - BX_u)(-1) = 0 \\ \sum_{u=1}^{M} 2(Y_u - A - BX_u)(-X_u) = 0 \end{cases} = \begin{cases} \sum_{u=1}^{M} Y_u - \underbrace{\sum_{u=1}^{M} A}_{= AM} - B \sum_{u=1}^{M} X_u = 0 \\ \sum_{u=1}^{M} Y_u X_u - A \sum_{u=1}^{M} X_u - B \sum_{u=1}^{M} X_u^2 = 0 \end{cases} =$$

- divido per M ⇒ $\bar{Y} - A - B\bar{X} = 0$, $A = \bar{Y} - B\bar{X}$

$$\sum_{u=1}^{M} Y_u X_u - A \overbrace{\sum_{u=1}^{M} X_u}^{= \bar{X}M} - B \sum_{u=1}^{M} X_u^2 = 0 \longrightarrow \sum_{u=1}^{M} Y_u X_u - M\bar{X}(\bar{Y} - B\bar{X}) - B \sum_{u=1}^{M} X_u^2 = 0 \to \sum_{u=1}^{M} Y_u X_u - M\bar{Y}\bar{X} = B\left(\sum_{u=1}^{M} X_u^2 - M\bar{X}^2\right)$$

$$\begin{cases} B = \dfrac{\sum_{u=1}^{M} Y_u X_u - M\bar{Y}\bar{X}}{\sum_{u=1}^{M} X_u^2 - M\bar{X}^2} \\ A = \bar{Y} - B\bar{X} \end{cases} \to$$

① X è una quantità che posso controllare praticamente senza errore, le $X_1 \dots X_n$ NON sono v.c. ma valori scelti da chi fa l'esperimento, mentre Y sarà una v.c. (gaussiana) con errore

↓

② $Y_u \sim N(\alpha + \beta X_u, \sigma^2)$ → $Y_u = \alpha + \beta X_u$ + errore con errore $\sim N(0, \sigma^2)$

③ $Y_1, \dots Y_M$ sono V.C. INDIPENDENTI

↓

questo ci dice che B è una v.c. perché è una combinazione lineare delle v.c. Y_u dato che le Y_u sono gaussiane indip. x la riproducibilità $B \sim N$

↓

$$B = \frac{\sum_{u=1}^{M} [(X_u - \bar{X})Y_u]}{\sum_{u=1}^{M} X_u^2 - M\bar{X}^2}$$

$$E[B] = E\left[\frac{\sum_{u=1}^{M} [(X_u - \bar{X})Y_u]}{\sum_{u=1}^{M} X_u^2 - M\bar{X}^2}\right] = \frac{\sum_{u=1}^{M} [(X_u - \bar{X})E[Y_u]]}{\sum_{u=1}^{M} X_u^2 - M\bar{X}^2} = \frac{\sum_{u=1}^{M} [(X_u - \bar{X})(\alpha + \beta X_u)]}{\sum_{u=1}^{M} X_u^2 - M\bar{X}^2} = \frac{\alpha \sum_{u=1}^{M} (X_u - \bar{X})}{\sum_{u=1}^{M} X_u^2 - M\bar{X}^2} + \frac{\beta \sum_{u=1}^{M} (X_u^2 - \bar{X}X_u)}{\sum_{u=1}^{M} X_u^2 - M\bar{X}^2} =$$

$$= \frac{\alpha\left[\overbrace{\left(\sum_{u=1}^{M} X_u\right)}^{= M\bar{X}} - M\bar{X}\right]}{\sum_{u=1}^{M} X_u^2 - M\bar{X}^2} + \frac{\beta\left(\left(\sum_{u=1}^{M} X_u^2\right) - \bar{X}\overbrace{\sum_{u=1}^{M} X_u}^{= M\bar{X}}\right)}{\sum_{u=1}^{M} X_u^2 - M\bar{X}^2} = \cancel{\frac{\alpha \cdot 0}{\dots}} + \beta \frac{\left(\sum_{u=1}^{M} X_u^2 - M\bar{X}^2\right)}{\sum_{u=1}^{M} X_u^2 - M\bar{X}^2} = \beta \Rightarrow E[B] = \beta \to \text{STIMATORE CORRETTO!}$$

$$E[A] = E[\bar{Y} - B\bar{X}] = E[\bar{Y}] - E[B]\bar{X} = E\left[\sum_{u=1}^{M} \frac{Y_u}{M}\right] - \beta\bar{X} = \frac{1}{M}\sum_{u=1}^{M} E[Y_u] - \beta\bar{X} = \frac{1}{M}\sum_{u=1}^{M} (\alpha + \beta X_u) - \beta\bar{X} =$$

$$= \frac{1}{M}\left(\sum_{u=1}^{M} \alpha + \beta \sum_{u=1}^{M} X_u\right) - \beta\bar{X} = \frac{\cancel{M}\alpha}{\cancel{M}} + \beta \frac{\overbrace{\sum_{u=1}^{M} X_u}^{= \bar{X}}}{M} - \beta\bar{X} = \alpha \Rightarrow E[A] = \alpha$$

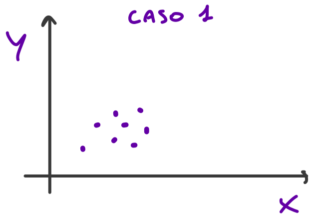
(SEMPRE REGRESSIONE LINEARE)

$$Var(B) = Var\left(\frac{\sum_{u=1}^{M}(X_u - \bar{X})Y_u}{\sum_{u=1}^{M} X_u^2 - M\bar{X}^2} \right) \underset{\substack{\downarrow \\ \text{indip.} \\ Y_u}}{=} \frac{\sum_{u=1}^{M}(X_u-\bar{X})^2 Var(Y_u)}{\left(\sum_{u=1}^{M} X_u^2 - M\bar{X}^2\right)^2} = \sigma^2 \frac{\sum_{u=1}^{M}(X_u-\bar{X})^2}{\left(\sum_{u=1}^{M} X_u^2 - M\bar{X}^2\right)^2} = \sigma^2 \frac{\sum_{u=1}^{M}(X_u-\bar{X})^2}{\left(\sum_{u=1}^{M}(X_u-\bar{X})^2\right)^2} =$$

$$= \frac{\sigma^2}{\sum_{u=1}^{M}(X_u-\bar{X})^2}$$

→ noi vogliamo una varianza piccola per B affinché sia uno stimatore, quindi il denominatore deve essere grande → $\sum_{u=1}^{M}(X_u-\bar{X})^2 \gg 1$

CASO 1



X_u tutti vicini

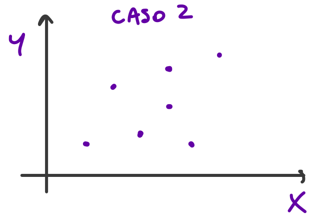
↓

$\sum_{u=1}^{M}(X_u-\bar{X})^2$ piccolo

↓

più complesso stimare il coeff. angolare della retta che interpola i dati

CASO 2



X_u tutti distanti

↓

$\sum_{u=1}^{M}(X_u-\bar{X})^2$ grande

↓

più semplice

NOTA: $\bar{Y} = \sum_{u=1}^{M} \frac{Y_u}{M}$ non è una media campionaria perché le Y_u non vengono dalla stessa popolazione

↓

$\mu_u = \alpha + \beta X_u$ e varia al variare di u

METODO PER LA SIMULAZIONE DI V.C. A PARTIRE DA $U \sim U(0,1)$

X v.c. continua con $F_X(a)$ → $Y = F_X^{-1}(U)$

$$F_Y(a) = P(Y \le a) = P(F_X^{-1}(U) \le a) = P(U \le F_X(a)) = \int_0^{F_X(a)} 1\, du = F_X(a) \Rightarrow X \doteq Y \rightarrow X = F_X^{-1}(U)$$