Statistica descrittiva

- Campione di N dati numerici $x = \{x_1, \dots, x_N\}$, $x_i \in \mathbb{R}$, abbiamo definito:
  - frequenza assoluta e relativa di un dato (istogramma)
  - media $\bar{x}$, varianza $var(x)$, scarto quadratico medio $\sigma(x)$
  - funzione di ripartizione $F(t)$
  - percentili (mediana).

- Confronto tra campioni. <u>Correlazione</u>

$x = \{x_1, \dots, x_N\}$, $\quad y = \{y_1, \dots, y_N\}$

$\bar{x}$ media	$\bar{y}$ media
$\sigma(x)$ s.q.m	$\sigma(y)$ s.q.m.
$var(x)$ varianza	$var(y)$ varianza

Def Si definisce COVARIANZA DI x e y il numero

$$cov(x,y) := \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})$$

OSS $cov(x,x) = var(x)$

$$cov(x,y) = \left(\frac{1}{N} \sum_{i=1}^{N} x_i y_i\right) - \bar{x}\,\bar{y} \quad .$$

Def Si chiama COEFFICIENTE DI CORRELAZIONE TRA x e y il numero $\rho(x,y) \in [-1,1]$ dato da

$$\rho(x,y) := \frac{cov(x,y)}{\sigma(x)\,\sigma(y)}$$

OSS $\rho(x,x) = \dfrac{cov(x,x)}{\sigma(x)\,\sigma(x)} = \dfrac{var(x)}{var(x)} = 1$

$\rho(x,-x) = -1$

Si interpreta $\rho(x,y)$ dicendo che x e y sono correlati se $|\rho(x,y)| \geq 0{,}7$ e sono scorrelati se $|\rho(x,y)| \leq 0{,}3$.

ES $\quad x = \{2, 2, 2, 3, 1, 1, 2, 3, 4, 1, 4\}$ $\qquad N = 11$

$y = \{110, 103, 102, 97, 85, 94, 95, 99, 122, 93, 101\}$

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \sim 2{,}27 \quad , \quad \sigma(x) = \sqrt{\operatorname{var}(x)} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2 - \bar{x}^2} \sim 1{,}06$$

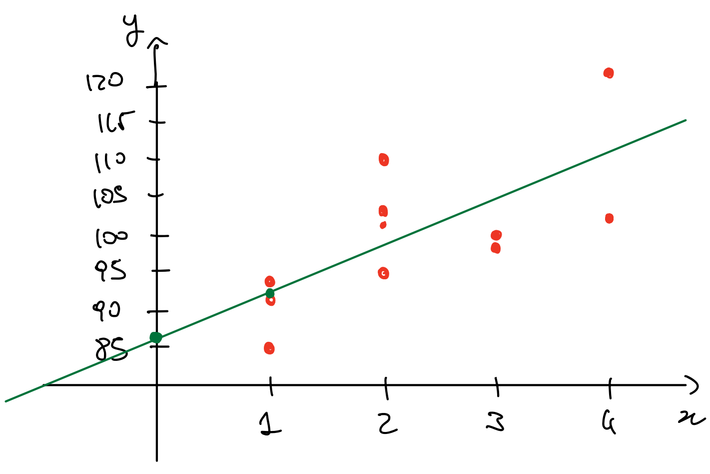
$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i \sim 100{,}09 \quad , \quad \sigma(y) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} y_i^2 - \bar{y}^2} \sim 9{,}27$$

$$\operatorname{cov}(x,y) = \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \bar{x}\,\bar{y} =$$

$$= \frac{1}{11}\Big(2\cdot 110 + 2\cdot 103 + 2\cdot 102 + 3\cdot 97 + 1\cdot 85 + 1\cdot 94 + 2\cdot 95 + 3\cdot 99 + 4\cdot 122 + 1\cdot 93 + 4\cdot 101\Big) - 2{,}27\cdot 100{,}09$$

$$\sim 6{,}62$$

$$\rho(x,y) = \frac{\operatorname{cov}(x,y)}{\sigma(x)\,\sigma(y)} = \frac{6{,}62}{1{,}06\cdot 9{,}27} \sim 0{,}67$$



(x_i, y_i)

$(x_1, y_1) = (2, 110)$

$\vdots \qquad \vdots$

$(x_{11}, y_{11}) = (4, 101)$

$$b^* = \frac{\operatorname{cov}(x,y)}{\operatorname{var}(x)} \sim 5{,}89$$

$$y = 5{,}89x + 86{,}72$$

$$a^* = \bar{y} - b^*\bar{x} \sim 86{,}72$$

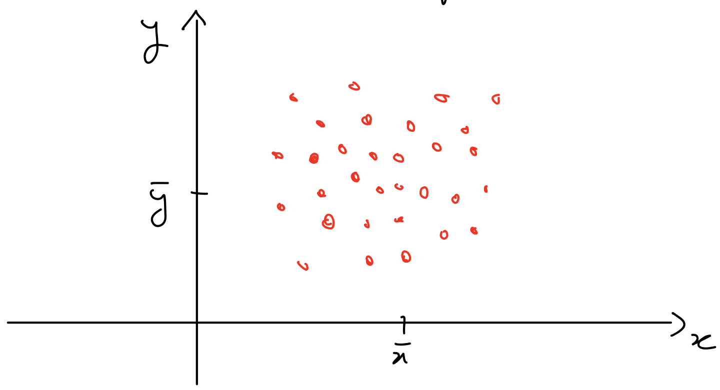
Def Si chiama RETTA DI REGRESSIONE tra x e y, la retta

$$y = b^* x + a^*, \quad \text{dove} \quad b^* = \frac{\operatorname{cov}(x,y)}{\operatorname{var}(x)}, \quad a^* = \bar{y} - b^*\bar{x}$$

OSS Se $x = y$, $b^* = 1$, $a^* = 0$.

a^* e b^* realizzano $\displaystyle\min_{a,b\in\mathbb{R}} \sum_{i=1}^{N} (y_i - bx_i - a)^2$

OSS Se x e y sono scorrelati, ci aspettiamo di vedere



Probabilità e Variabili aleatorie

Caso discreto

- Supponiamo di avere una quantità finita o numerabile di valori reali

$$x_1, x_2, x_3, \ldots$$

Una probabilità su $x_1, x_2, x_3, \ldots$ è un insieme di $\{p_i\}$, con la proprietà che $p_i \geq 0 \ \forall i$, $\sum_i p_i = 1$. p_i è la probabilità di x_i.

ES $\{1, 2, 3, 4, 5, 6\}$ valori

$\left\{\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right\}$ è una distribuzione di probabilità.

- Una VARIABILE ALEATORIA DISCRETA X è una funzione $X : \Omega \to \mathbb{R}$, dove Ω è detto universo, con immagine $\{x_1, x_2, x_3, \ldots\}$ insieme discreto finito o numerabile.

ES Ω universo del lancio di un dado

$X : \Omega \to \mathbb{R}$, $\quad \mathrm{Imm}\, X = \{1, 2, 3, 4, 5, 6\}$

$X : \Omega \to \mathbb{R}$, $\quad \mathrm{Imm}\, X = \{-1, +1\}$, $\quad X(\omega) = \begin{cases} +1, & \text{esce numero pari} \\ -1, & \text{esce " dispari} \end{cases}$

La LEGGE di X è la distribuzione di probabilità su $\mathrm{Imm}\, X$ che si ottiene utilizzando una probabilità $\mathbb{P}$ su Ω ponendo

p_i = prob. di $x_i \in \operatorname{Imm} X := \mathbb{P}\{\omega \in \Omega : X(\omega) = x_i\}$
$p(x_i)$ $\quad (= \mathbb{P}(X = x_i))$.

ES Ω = universo del lancio di un dado

1. $X : \Omega \to \mathbb{R}$, $\operatorname{Imm} X = \{1, 2, 3, 4, 5, 6\}$

$p(1) = \mathbb{P}(X = 1) = \frac{1}{6}$

$p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}$.

2. $X : \Omega \to \mathbb{R}$, $\operatorname{Imm} X = \{-1, +1\}$

$$X(\omega) = \begin{cases} +1 & \text{, esce numero pari} \\ -1 & \text{, esce " dispari} \end{cases}$$

$p(-1) = \mathbb{P}(X = -1) = \left(\mathbb{P}\{1,3,5\} = \mathbb{P}\{1\} + \mathbb{P}\{3\} + \mathbb{P}\{5\}\right) = \frac{1}{2}$

$p(+1) = \mathbb{P}(X = +1) = \frac{1}{2}$

ES Ω = universo del lancio ripetuto di una moneta onesta 3 volte.

$= \{CCC, CCT, CTT, CTC, TTT, TTC, TCT, TCC\}$

$\mathbb{P}(CCC) = \frac{1}{8} = \mathbb{P}(CCT) = \mathbb{P}(CTT) = \dots = \mathbb{P}(TCC)$

X è la somma di $+1$ per ogni C e di -1 per ogni T.

$\operatorname{Imm} X \subseteq \{-3, -2, -1, 0, +1, +2, +3\}$

$p(-3) = \mathbb{P}(X = -3) = \mathbb{P}\{\omega \in \Omega : X(\omega) = -3\} = \mathbb{P}\{TTT\} = \frac{1}{8}$

$p(-2) = \mathbb{P}(X = -2) = \mathbb{P}(\emptyset) = 0$

$p(-1) = \mathbb{P}(X = -1) = \mathbb{P}\{CTT, TTC, TCT\} = \frac{3}{8}$

$p(0) = 0$

$p(+1) = \mathbb{P}(X = +1) = \mathbb{P}\{CCT, CTC, TCC\} = \frac{3}{8}$

$p(+2) = 0$

$p(+3) = \mathbb{P}(X = +3) = \mathbb{P}\{CCC\} = \frac{1}{8}$

Def La FUNZIONE DI RIPARTIZIONE DI X DISCRETA è

$F: \mathbb{R} \to [0,1]$ data da $F(t) = \mathbb{P}(X \leq t)$

$\left(= \mathbb{P}\{\omega \in \Omega : X(\omega) \leq t\}\right)$

ES F