

Esercizio 1– regressione lineare

Si considerino i seguenti dati campionari raccolti per stimare il tempo che impiega un computer a processare dati: x è il numero di dati e y il tempo, in secondi, impiegato per processarli.

x	211	332	322	435	275
y	112	155	131	208	138

1. Si calcoli il coefficiente di correlazione tra le variabili x e y ;
2. Usare il metodo dei minimi quadrati per determinare la retta di regressione. Fare un plot dei dati e della retta di regressione;
3. Si calcoli il tempo previsto per processare 200, 300, 400, 500 dati.

Esercizio 2– regressione lineare

I seguenti dati di deformazione lineare (ε_Y) e longitudinale (ε_X) sono stati ottenuti sottoponendo a sforzo delle sbarre fabbricate con una lega sperimentale:

ε_X	0.3	0.4	0.2	0.5	0.6
ε_Y	0.11	0.14	0.06	0.16	0.22

1. Si calcoli il coefficiente di correlazione tra le variabili ε_X e ε_Y .
2. È ragionevole supporre che ci sia una relazione lineare tra le due quantità? Gustificare la risposta;
3. Usare il metodo dei minimi quadrati per determinare la retta di regressione. Fare un plot dei dati e della retta di regressione;
4. Usare la retta di regressione trovata per predire la deformazione laterale corrispondente a una deformazione longitudinale pari a $\varepsilon_X = 0.35$ e a $\varepsilon_X = 0.8$. Quale delle due previsioni ritenete più affidabile? Perché?

Esercizio 3– intervalli di confidenza

Dato il campione, proveniente da una distribuzione normale $N(\mu, \sigma^2)$:

$$\{0.39, 0.68, 0.82, 1.35, 1.38, 1.62, 1.70, 1.71, 1.85, 2.14, 2.89, 3.69\}.$$

1. calcolare \bar{X}_n , $\bar{\sigma}_n$, *skewness* e curtosi. Sono coerenti con i valori di una distribuzione normale?
2. Stimare i valori dei parametri μ e σ in base al campione;
3. Calcolare degli intervalli di confidenza al livello del 95% e 99% per la media e la varianza.

Esercizio 4– intervalli di confidenza

Si sono fatti dei test per confrontare l'affidabilità di tre differenti marche di floppy disk. Per ciascuna delle tre marche, A, B, C, si è considerato un campione di 100 esemplari, che sono stati sottoposti a varie prove. Per ogni dischetto si è misurato il tempo, in ore, dopo il quale si è verificato il primo errore. I dati sono i seguenti:

Marca A	Marca B	Marca C
$\bar{X}_n = 49\text{h}$	$\bar{X}_n = 55\text{h}$	$\bar{X}_n = 53\text{h}$
$S = 8.2\text{h}$	$S = 10.1\text{h}$	$S = 7.5\text{h}$

Costruire intervalli di confidenza al livello del 95% per il tempo medio di funzionamento senza errori dei dischetti delle tre marche, in modo da eseguire un confronto di prestazioni. Cosa si può concludere? Per la validità del procedimento seguito è necessario supporre che la v.a. “Tempo di funzionamento senza errori” sia distribuita normalmente?