

Workshop Minitab

Settore Chimico-Farmaceutico

PISA - 29/05/2025

Statistical Background

GMSL Scientific Soul

contact@gmsl.it
+39 0331587511

STATISTICAL BACKGROUND

Analisi Descrittiva

Scientific Soul

contact@gmsl.it
+39 0331587511
contact@gmsl.it
+39 0331587511

Analisi dei Dati

Statistica Descrittiva

La statistica descrittiva è la branca della statistica che studia i criteri di rilevazione, classificazione, sintesi e rappresentazione dei dati appresi dallo studio di una popolazione o di una parte di essa (detta campione).

Scientific Soul

contact@gmsl.it
+39 0331587511

Statistica Descrittiva

Rilevazione dei Dati
Indagine su un'intera popolazione (censimento) o su un campione rappresentativo (sondaggio).

Classificazione dei Dati
I dati raccolti possono essere classificati attraverso distribuzioni.

Sintesi dei Dati
I dati raccolti possono essere sintetizzati attraverso famiglie di indici (posizione, variabilità, forma).

Rappresentazione dei Dati
I dati di un'indagine sono rappresentati attraverso grafici (es. istogramma, boxplot, ecc…).

contact@gmsl.it
+39 0331587511

Scientific Soul

STATISTICAL BACKGROUND

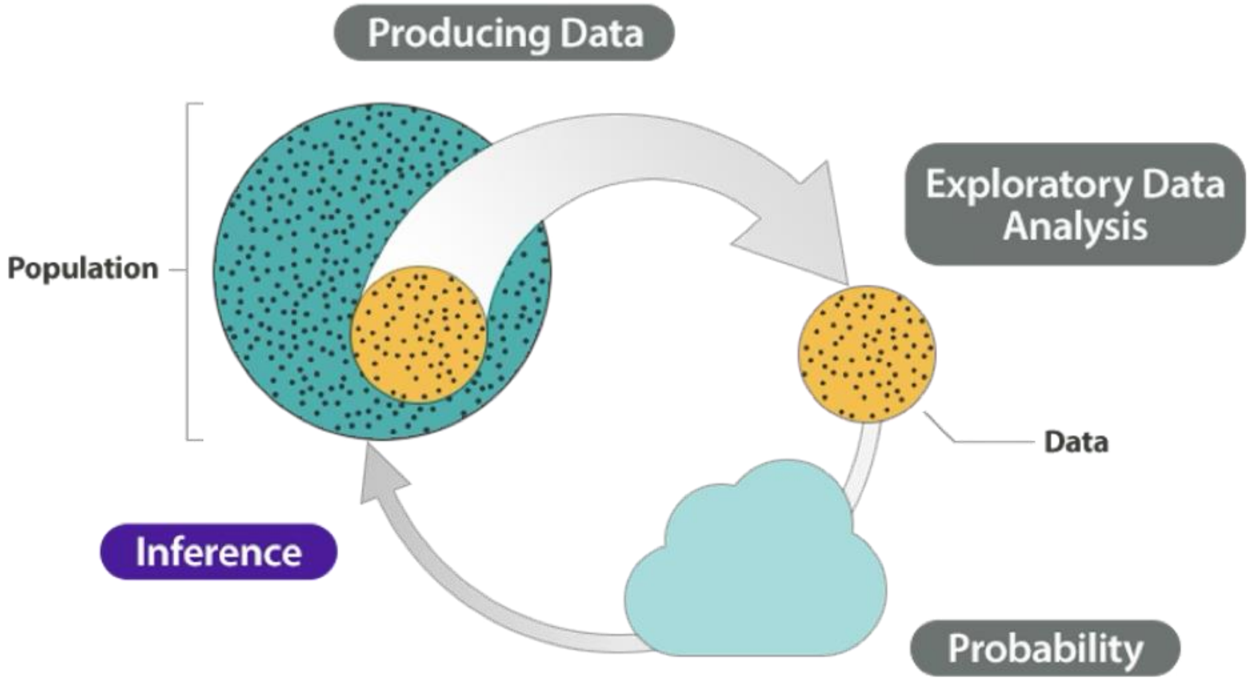
Analisi Inferenziale

Scientific Soul

contact@gmsl.it
+39 0331587511
contact@gmsl.it
+39 0331587511

Statistica Inferenziale

Metodi utilizzati per assumere decisioni o per trattare conclusioni su una popolazione e che per tale scopo si basano sull'informazione contenuta in un campione di tale popolazione.



Aree Principali
- Stima dei Parametri
- Test d'Ipotesi

contact@gmsl.it
+39 0331587511

Test di Ipotesi

Analisi dei dati di un ESPERIMENTO COMPARATIVO

Ipotesi Statistica
Asserzione relativa ai parametri di una o più popolazioni

Test d'Ipotesi
Procedura che porta a una decisione inerente una particolare ipotesi

H_0 Ipotesi Nulla
H_1 Ipotesi Alternativa

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases} \qquad \text{test BILATERALE}$$

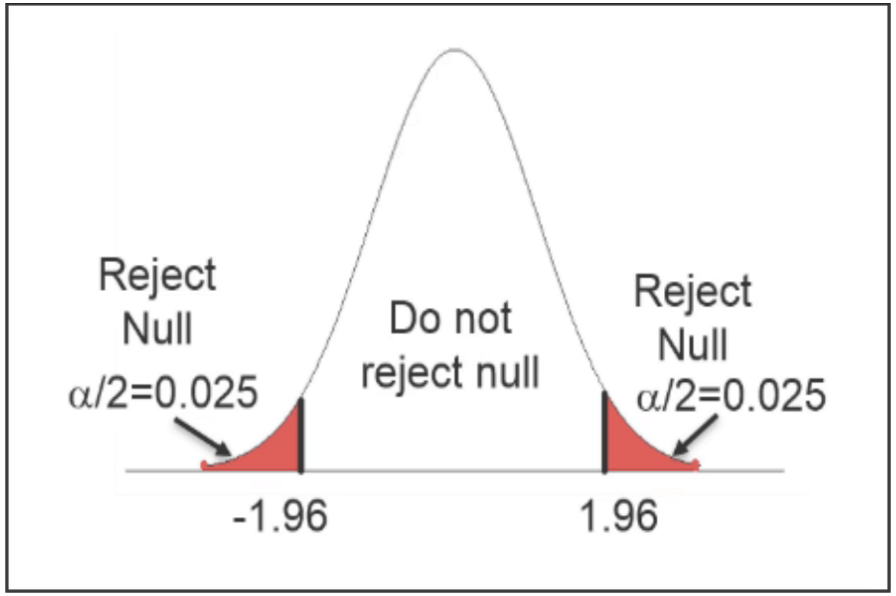
Scientific Soul
contact@gmsl.it
+39 0331587511

Test di Ipotesi

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

test BILATERALE

- Livello di Confidenza -> CI (es. 95%)
- Livello di Significatività -> α (es. 5%)
- p-value -> p



Prendere una decisione:

- $p < \alpha$ -> RIGETTO l'ipotesi nulla (H_0)
- $p > \alpha$ -> FALLISCO NEL RIGETTARE l'ipotesi nulla (H_0)

Scientific Soul

contact@gmsl.it
+39 0331587511

Power Analysis

Power is the ability of a test to detect a difference when one exists. A hypothesis test has the following possible outcomes:

	Null hypothesis	
Decision	**True**	**False**
Fail to reject	Correct decision probability = $1-\alpha$	Type II error probability = β
Reject	Type I error probability = α	Correct decision probability = $1-\beta$ (Power)

The power of the test is the probability that you will correctly reject the null hypothesis, given that the null hypothesis is false. Use a power analysis to determine how much power a test has or to design a new test with adequate power.

When to use power analysis

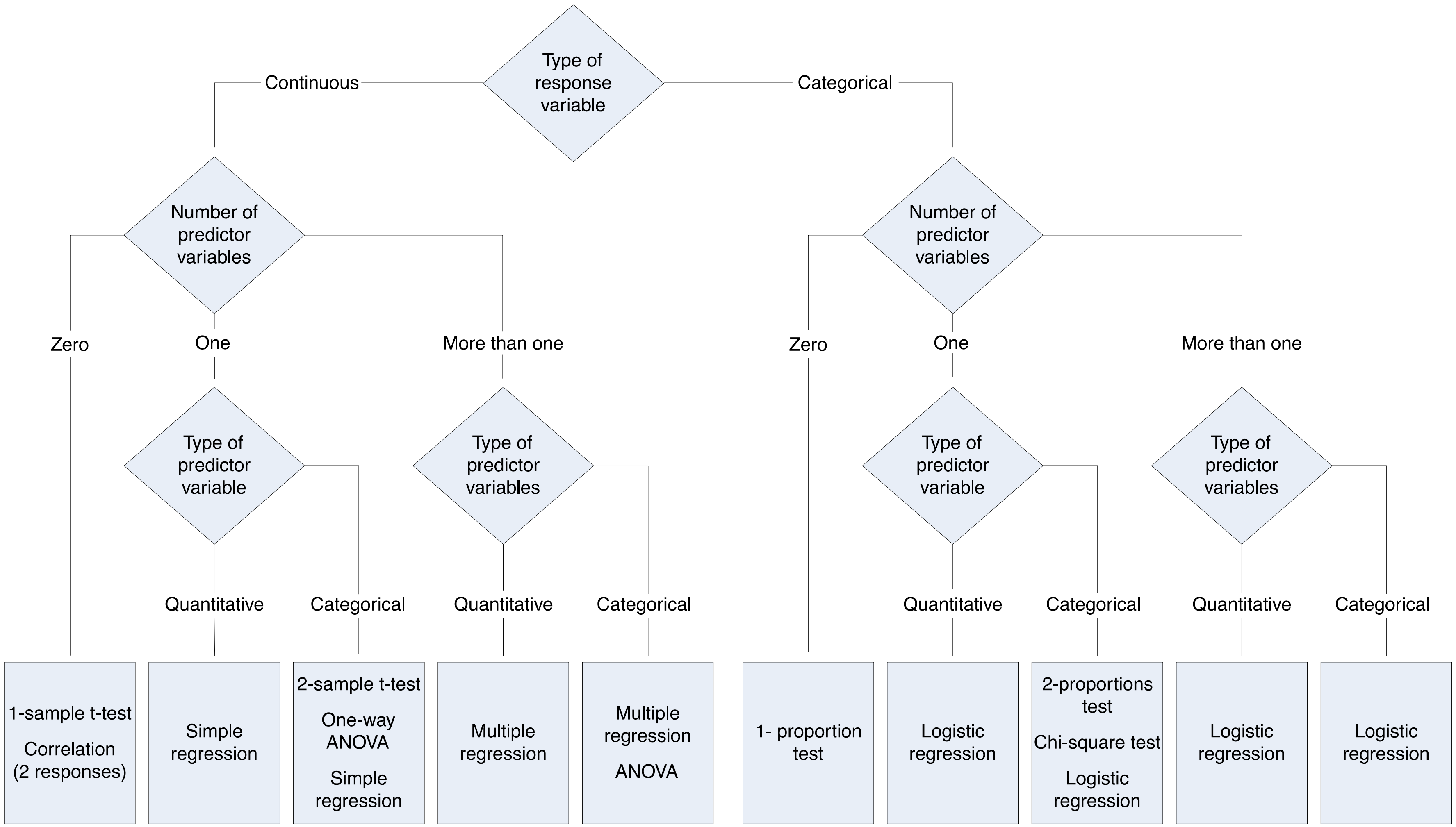
Use a power analysis:

- Before collecting data, to determine the sample size
- After collecting data, to evaluate the power to detect a difference

Scientific Soul

contact@gmsl.it
+39 0331587511

Test e Modelli



Scientific Soul

contact@gmsl.it
+39 0331587511

Scientific Soul

STATISTICAL BACKGROUND

ANOVA

Scientific Soul

contact@gmsl.it
+39 0331587511
contact@gmsl.it
+39 0331587511

ANOVA

L’analisi della varianza (ANOVA, dall’inglese **Analysis of Variance**) comprende una serie di test statistici che rientrano nell’ambito della statistica inferenziale.

Obiettivo

Valutare gli effetti di uno o più fattori di controllo su una variabile di interesse

ANOVA vs t-test

La procedura ANOVA (analisi della varianza) è una generalizzazione del t-test per campioni indipendenti: è possibile utilizzare l’ANOVA per analizzare le medie su più di due gruppi contemporaneamente.

Ipotesi sui gruppi di Dati

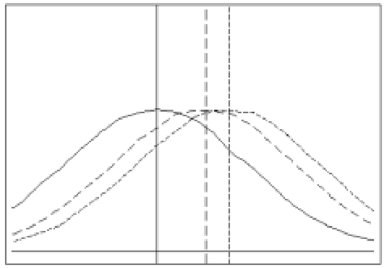
- Normalità dei dati
- Omoschedasticità

contact@gmsl.it
+39 0331587511

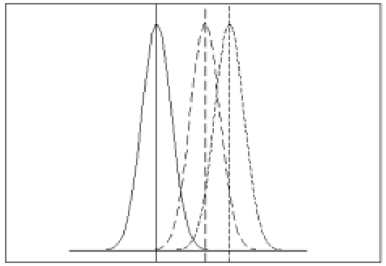
Come funziona ANOVA?

La logica alla base di ANOVA è che la variazione all'interno dei gruppi sia dovuta solamente all'errore casuale:

- se l'ammontare della variazione tra i gruppi è simile a quella all'interno dei gruppi, è probabile che la media dei gruppi differisca solamente a causa dell'errore casuale.



- se la variazione tra gruppi è più ampia rispetto a quella all'interno del gruppo, è probabile che le differenze tra le medie siano causate dalle differenze sui livelli dei fattori.



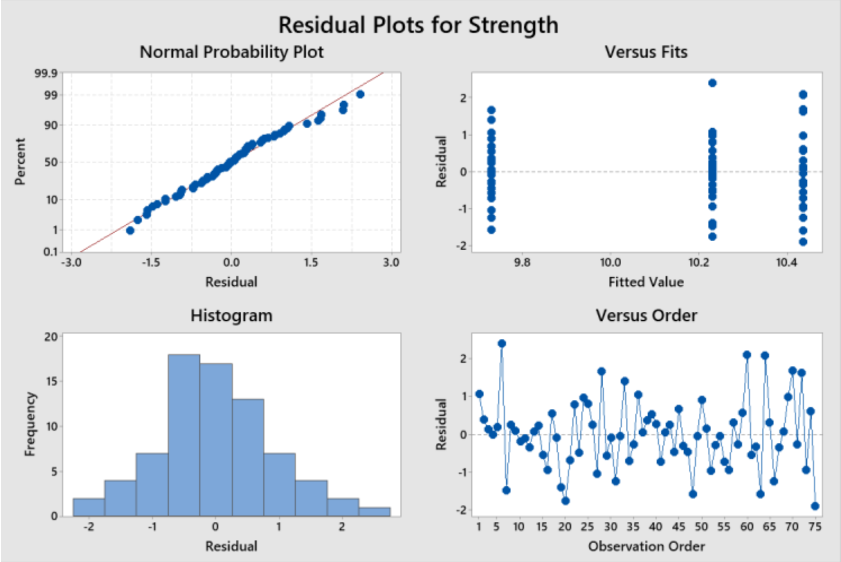
contact@gmsl.it
+39 0331587511

Quando si usa ANOVA?

…quando si dispone di dati in risposta continua per due o più livelli fissi di un singolo fattore…

N.B. Prima di accettare i risultati di una ANOVA, bisogna verificare che gli errori nei dati siano:

- indipendenti (casuali)
- normalmente distribuiti
- a varianza costante (su tutti i livelli dei fattori).



contact@gmsl.it
+39 0331587511

Test d'Ipotesi nell'ANOVA

L'**ipotesi nulla** prevede che i dati di tutti i gruppi nella popolazione abbiano la stessa media, e che le differenze osservate nel campione tra le medie dei gruppi siano dovute solo al caso.

L'**ipotesi alternativa** è invece che ci sia una differenza significativa tra i gruppi. In altre parole, che almeno un gruppo abbia un valore medio significativamente diverso dagli altri.

$$H_0 : \mu_1 = \mu_2 = \dots \mu_p$$

H_1: almeno due delle medie sono tra loro differenti

Scientific Soul

contact@gmsl.it
+39 0331587511

STATISTICAL BACKGROUND

Correlazione e Regressione

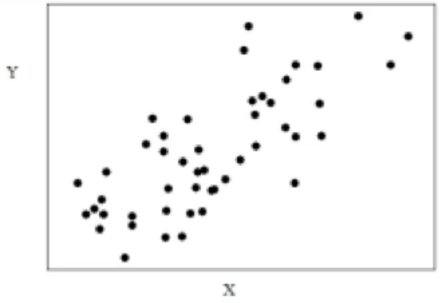
contact@gmsl.it
+39 0331587511
contact@gmsl.it
+39 0331587511

Correlazione

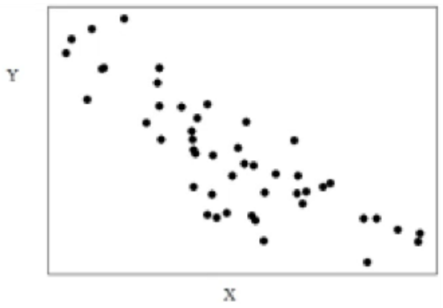
La correlazione è una relazione tra due variabili tale che a ciascun valore della prima corrisponda un valore della seconda, seguendo una certa regolarità. La correlazione non dipende da un rapporto di causa-effetto quanto dalla tendenza di una variabile a cambiare in funzione di un'altra.

La correlazione è una misura del grado di associazione lineare tra due variabili
(il grado con il quale una variabile cambia con l'altra)

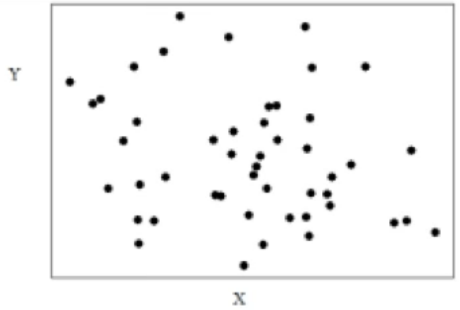
Associazione Lineare Positiva



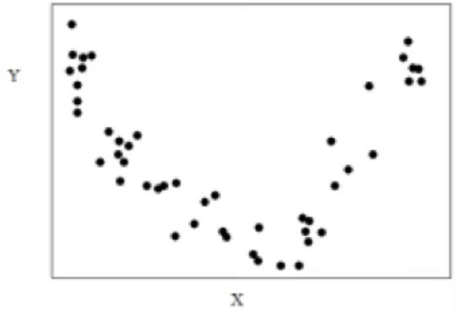
Associazione Lineare Negativa



Nessuna Associazione



Associazione Non Lineare



Scientific Soul

contact@gmsl.it
+39 0331587511

Come misuro la Correlazione?

Correlazione Pearson

Il coefficiente di correlazione (r oppure ρ) viene calcolato usando la formula:

$$\rho = \frac{\sum_{j=1}^{N}(X_j - \bar{X}) \cdot \sum_{k=1}^{N}(Y_k - \bar{Y})}{\sqrt{\sum_{j=1}^{N}(X_j - \bar{X})^2 \cdot \sum_{k=1}^{N}(Y_k - \bar{Y})^2}}$$

Il valore di r è sempre compreso tra –1 and 1:

- 1 indica una perfetta correlazione positiva.
- 0 indica nessuna correlazione.
- –1 indica una perfetta correlazione negativa.

P-value

Il p-value verifica le seguenti ipotesi:

H_0: Il coefficiente di correlazione (ρ o rho) della relazione tra popolazioni è uguale a zero.

H_1: ρ non è uguale a zero.

Osservazioni:

- Una forte correlazione non implica necessariamente una relazione cause-and-effect. Una forte correlazione tra due variabili può essere dovuta all'influenza di una terza variabile non considerata.
- Un coefficiente di correlazione vicino allo zero non necessariamente indica nessuna associazione; può significare che l'associazione non è lineare.
- La correlazione assume che i valori di entrambe le variabili siano liberi di variare.

contact@gmsl.it
+39 0331587511

Regressione

La regressione formalizza e risolve il problema di una relazione funzionale tra variabili misurate sulla base di dati campionari estratti da un'ipotetica popolazione infinita.

La Regressione semplice esamina la relazione tra variabile di risposta continua (Y) e la variabile predittore (X). L'equazione generale per la regressione semplice è:

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

dove Y è la risposta, X è il predittore, β_0 è l'intercetta (il valore di Y quando X è uguale a zero), β_1 è il coefficiente angolare, e ε è l'errore casuale.

Nota Bene

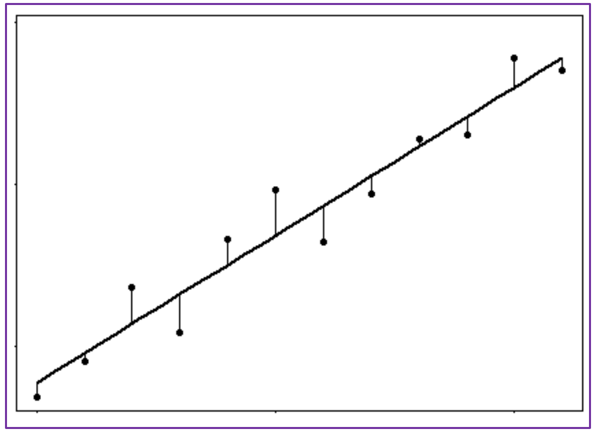
Per confermare la validità dell'analisi bisogna verificare che i residui (valori di Y ottenuti nel campionamento che differiscono da quelli predetti dal modello di regressione) siano:

- indipendenti (casuali)
- normalmente distribuiti
- a varianza costante (su tutti i livelli dei fattori).

contact@gmsl.it
+39 0331587511

Interpretazione del modello di Regressione

- Retta di regressione
- Indici S, R^2, R^2 predicted
- Risultati di ANOVA per valutare se il modello di regressione semplice è vantaggioso.



- Modello di regressione: $Y = \beta_0 + \beta_1 X + \varepsilon$
- Modello ridotto: $Y = \beta_0 + \varepsilon$

Il modello ridotto dice che i cambiamenti in Y sono dovuti solamente ad errori casuali (ε). Questo è equivalente ad un modello di regressione semplice con un coefficiente angolare (β_1) pari a zero. Quindi, le ipotesi per l'ANOVA sono:

H_0: β_1 è uguale a zero.

H_1: β_1 non è uguale a zero.

Il P-value deve essere interpretato come:

- Se il p-value è minore o uguale a α, rifiutare H_0. Il modello di regressione spiega significativamente più variabilità nella risposta rispetto al modello ridotto. β_1 non è uguale a zero.
- Se il p-value è più grande, non si potrà rifiutare H_0. β_1 non è significativamente differente da zero.

POSSIAMO OTTENERE UNA PREVISIONE?

CI vs PI

Scientific Soul

contact@gmsl.it
+39 0331587511