

# SCRIPT ANOVA

Dataset su cui si lavora: **Top 20 products**

## **MANOVA su vendite in volume (log) e % sconto per PRODOTTO**

Questo test (così come il successivo) ci aiutano a rispondere alla seconda domanda di ricerca e ci serviranno anche dopo quando andiamo a creare i linear mixed models (per includere i prezzi scontati e non scontati come random slopes). Facciamo un manova test per le vendite in volume (log) e lo sconto (in percentuale) per andare a vedere se queste due variabili cambiano in base al prodotto (riferimento ai boxplot, che ci aiutano a visualizzare questa categorizzazione). La conclusione consiste nel rifiutare l'ipotesi nulla ( $p\text{-value} < 2.2e-16$ ), quindi la categorizzazione per prodotto influisce sulle due variabili. Significa che le vendite in volume e la percentuale di sconto differiscono in base al prodotto.

## **MANOVA su vendite in volume (log) e % sconto per BRAND**

Stesso discorso fatto per il modello sopra. Anche qua c'è differenza nelle vendite in volume e nello sconto differiscono in base al brand.

## **ANOVA su vendite in volume (log) per TIPO DI SCONTO**

Questo test ci serve per rispondere alla prima domanda di ricerca. Evidenziamo le varie tipologie di sconto presenti e diciamo che in tutto ci sono 14 combinazioni di tipologie di sconti, infatti spesso succede che le promozioni vengono fatte in concomitanza ad altre. Il risultato di questo test consiste nel rifiutare l'ipotesi nulla ( $p\text{-value} < 2.2e-16$ ), che significa che le vendite in volume differiscono in base alla promozione applicata, si vede bene dal boxplot (che è fatto sulle vendite in volume (log) per tipologia di sconto, magari spiegare la legenda sugli sconti)

Fare riferimento alle weaknesses di questo metodo applicato ai nostri dati: anche se usiamo la trasformazione logaritmica in alcuni gruppi non otteniamo la normalità (in nessuno dei 3 test) e la assunzione sull'uguaglianza delle varianze/covarianze è spesso violata.

## **Risposta ChatGPT alla domanda “Perché usare la scala logaritmica nelle vendite di una time series?”**

È spesso preferibile utilizzare la scala logaritmica quando si lavora con serie temporali di vendite o prezzi per diversi motivi principali:

### **1. Gestione delle Variazioni Percentuali**

Quando si lavora con serie temporali che mostrano variazioni significative nei valori (come vendite o prezzi), le variazioni percentuali sono più rilevanti rispetto alle variazioni assolute. La scala logaritmica tende a comprimere le variazioni assolute più grandi, rendendo più evidenti le variazioni percentuali minori. Questo aiuta a visualizzare meglio le dinamiche relative delle serie temporali.

### **2. Stabilizzazione della Varianza**

Le serie temporali che mostrano varianza crescente nel tempo (eteroschedasticità) possono beneficiare dall'essere trasformate su scala logaritmica. Questa trasformazione può aiutare a stabilizzare la varianza nel tempo, rendendo più uniforme la distribuzione dei dati rispetto alla media. Ciò è particolarmente utile quando si desidera applicare tecniche statistiche che presuppongono omoschedasticità (varianza costante).

### **3. Interpretazione delle Tendenze**

Su una scala lineare, piccoli cambiamenti nei valori possono sembrare insignificanti, mentre su una scala logaritmica, gli stessi cambiamenti possono essere più visibili. Questo può facilitare l'identificazione di tendenze e pattern nascosti nelle serie temporali, consentendo una migliore interpretazione dei dati.

### **4. Comparazione tra Serie Temporali**

La trasformazione logaritmica può rendere più semplice la comparazione tra diverse serie temporali che potrebbero variare notevolmente in termini di ampiezza. Portare i dati su scala logaritmica può

aiutare a standardizzare le serie temporali e a rendere più facile rilevare le differenze relative tra di esse.

## Considerazioni Pratiche

- **Logaritmo Naturale o Decimale:** La scelta tra logaritmo naturale ( $\ln$ ) e logaritmo decimale dipende dal contesto e dalle preferenze dell'analista. Entrambi sono comuni, con il logaritmo naturale che è più frequentemente usato in analisi finanziarie e economiche.
- **Interpretazione dei Risultati:** È importante ricordare che la scala logaritmica modifica la percezione dei dati, quindi l'interpretazione dei risultati dovrebbe riflettere questa trasformazione. Ad esempio, un aumento costante su una scala logaritmica corrisponde a un tasso di crescita percentuale costante su una scala lineare.

In conclusione, l'uso della scala logaritmica è consigliato quando si lavora con serie temporali di vendite o prezzi per migliorare la visualizzazione, stabilizzare la varianza, facilitare l'interpretazione delle tendenze e consentire una comparazione più efficace tra le serie temporali.