# Probabilità e Statistica per l’informatica Progetto R

## Il contesto

Si vuole valutare se ampliare o meno l’area assistenza e riparazioni della sezione ciclismo di Decathlon Milano, valutando se l’ipotesi del nostro superiore per cui, il servizio di assistenza influisce positivamente sulle vendite dei componenti, sia vera. Sono stati osservati i valori di vendita e di ricavo per 52 settimane relativamente alla vendite di diversi prodotti tra cui la vendita delle componenti e la vendita dei servizi di riparazione e manutenzione. Sono state analizzate le colonne relative alla vendita delle componenti e alla vendita dei servizi di riparazione e manutenzione utilizzando R.

## Ipotesi

L’ipotesi è che il servizio di assistenza influisce positivamente sulle vendite delle componenti.

## Analisi dei dati

### Indici di posizione e dispersione

|  |  |  |
| --- | --- | --- |
|  | Vendite componenti | Servizi di riparazione e manutenzione |
| Media campionaria | 42.48 | 38.48 |
| Primo quartile | 31 | 25.8 |
| Mediana | 38.5 | 38.5 |
| Terzo quartile | 55 | 53 |
| Varianza | 214.73 | 178.14 |
| Deviazione standard | 14.65 | 13.35 |

Notiamo che sia la media che gli indici di dispersione non sono molto diversi fra loro.

Possiamo analizzare i dati anche attraverso i relativi grafici di dispersione e boxplot.

A picture containing text, screenshot, diagram, line

Description automatically generatedA picture containing text, screenshot, diagram, line

Description automatically generated

Graficamente possiamo notare una possibile correlazione lineare fra i due campioni e dai boxplot una distribuzione similare.

A picture containing text, screenshot, diagram, rectangle

Description automatically generated

Come ulteriore verifica è stato calcolato il coefficiente di correlazione lineare ottenendo un valore di 0.90, essendo superiore a 0.7 è un’indicazione della correlazione dei due campioni.

## Ipotesi e Verifica

Formuliamo quindi l’ipotesi secondo cui le vendite dei servizi di riparazione e le componenti vendute tramite servizio di assistenza siano positivamente correlate.

Utilizzeremo il test di indipendenza chi-quadro per verificare se le due variabili sono dipendenti.

Consideriamo la vendita componenti come X e il servizio di riparazione e manutenzione come Y.

Ipotesi nulla H0: X e Y siano indipendenti

Ipotesi alternativa H1: X e Y sono dipendenti

I dati sono stati divisi in 8 classi con i seguenti range

10 – 20 – 30 – 40 – 50 – 60 -70 -80

Sono stati dispositi in una matrice di contingenza ed è stato effettuato un test chi quadro, i risultati ottenuti sono i seguenti

A black background with white text

Description automatically generated with low confidence

Il p/value indica la soglia oltre la quale rifiutiamo l’ipotesi nulla, quindi avendo un p-value inferiore a 0.05 possiamo rifiutare l’ipotesi H0 per cui X e Y siano indipendenti e affermare che c’è evidenza statistica che i due campioni siano dipendenti con una confidenza del 95 %. Il p-value è estremamente basso quindi potremmo rifiutare l’ipotesi nulla anche con una confidenza maggiore.

## Conclusioni

Dau dati raccolti e dalle analisi effettuate è possibile sostenere che ci sia una forte evidenza statistica che il servizio di assistenza influisca positivamente sulla vendita delle componenti, non è possibile affermare che uno sia causa dell’altro