

CASO DI STUDIO #2

Il file `finanziamenti.csv` contiene alcune informazioni riguardo a progetti finanziati nell'anno corrente:

- *id*: identificatore del progetto,
- *TemaPrioritario*: codice che indica il tema prioritario del progetto,
- *FONTE*: area tematica del progetto,
- *CATEGORIA*: categoria del progetto,
- *CodiceCategoria*: codice numerico della *CATEGORIA*,
- *UNITA*: unità di lavoro,
- *FinProvincia*: entità del finanziamento da parte della provincia,
- *FinRegione*: entità del finanziamento da parte della regione,
- *TotSpese*: spese sostenute per il progetto.

Le colonne sono separate dal simbolo ";" e i numeri reali sono stati registrati con il simbolo "." come separatore dei decimali. Per accedere al file, collegatevi al sito `upload.di.unimi.it`, selezionate l'esame di *Statistica e analisi dei dati* per l'appello odierno e scaricate il file `finanziamenti.csv`.

1. Il carattere *CodiceCategoria* è nominale, ordinale o scalare? Giustificate la risposta.
2. Calcolate la tabella delle frequenze assolute del carattere *UNITA*.
3. Tracciate un grafico opportuno per descrivere il carattere *UNITA*.
4. La Figura 1 mostra la funzione di ripartizione empirica per un **sottoinsieme** delle osservazioni relativi al carattere *TotSpese*, in cui gli importi sono indicati in centinaia di migliaia di Euro. Leggendo esclusivamente il grafico:
 - 4.1. indicate quale sottoinsieme di osservazioni è stato utilizzato;
 - 4.2. specificate quale percentuale delle osservazioni visualizzate assume un valore compreso tra uno e due milioni di Euro.
5. Prendiamo in considerazione la quota di finanziamento erogata dalla Provincia.
 - 5.1. Create una variabile (chiamatela `progetti_a`, per indicare i progetti di tipo A) che contenga la parte di dataset relativa ai progetti per i quali la quota provinciale di finanziamento è minore di quella regionale, e un'altra (chiamata `progetti_b`, per indicare i progetti di tipo B) che contenga la parte di dataset relativa ai progetti per i quali la quota provinciale di finanziamento è maggiore di quella regionale.
 - 5.2. Quanti sono progetti di tipo A? Quanti sono progetti di tipo B?

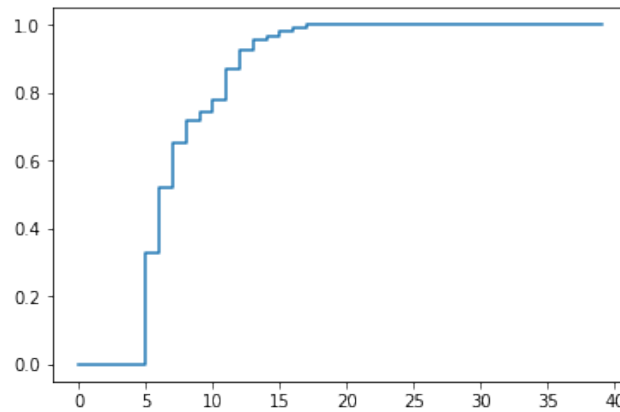


Figura 1: Il grafico della funzione di ripartizione empirica per un sottoinsieme delle osservazioni di *TotSpese*.

6. Concentriamoci sui progetti di tipo A.

- 6.1. Selezionate i progetti di tipo A che hanno ricevuto un finanziamento provinciale compreso tra i 200 e i 1000 euro, estremo sinistro incluso, e salvate questa parte di dataset in una variabile chiamata `selezione_progetti_a`.
- 6.2. Tracciate un istogramma del finanziamento provinciale di tali progetti, imponendo che le classi abbiano ampiezza 100 euro.
- 6.3. Tracciate anche il boxplot per la medesima quantità.
- 6.4. Tra i due grafici appena prodotti, quale ritenete maggiormente informativo? Giustificate la risposta.
- 6.5. Relativamente a tali progetti, qual è stato l'importo medio finanziato dalla provincia? Quale la deviazione standard?
- 6.6. Quanti hanno ricevuto un finanziamento provinciale compreso tra i 500 e i 700 euro?
- 6.7. Esiste una evidente relazione tra finanziamento provinciale e spese sostenute. Descrivete tale relazione nel modo più dettagliato possibile, utilizzando un indice numerico e un metodo grafico.
- 6.8. Nella relazione avrete notato la presenza di almeno un progetto che si discosta notevolmente dall'andamento più generale. Eliminate tali progetti dall'insieme dei dati e rispondete nuovamente alle domande del punto precedente.

7. Ritorniamo al dataset completo. Quanti sono i progetti che non hanno ancora sostenuto spese?