

3

La matrice dei dati e le analisi univariate

In questo capitolo vengono richiamate le principali metodologie di base della statistica in modo da riassumere i dati per rendere fruibili le informazioni in essi contenute. Quando si carica un dataset, per verificare la tipologia delle variabili, è importante visualizzare in anteprima la distribuzione di frequenze per le variabili qualitative, e il minimo, massimo e mediana per le variabili quantitative.

Esercizio 3.1

Importare dentro MATLAB il dataset presente nel file `Firm.xlsx` in una table denominata `Xt`. Definire la variabile “Gender” come qualitativa nominale e la variabile “Education” come qualitativa ordinale. Creare un summary dei dati in cui per ogni variabile quantitativa vengano riportati il valore minimo, il valore massimo e la mediana e per ogni variabile categorica (qualitativa nominale o ordinale) la distribuzione di frequenza.

Soluzione

```
1 %% Caricamento dati
2 miofile="Firm.xlsx";
3 Xt=readtable(miofile,"ReadRowNames",true);
4
5 %% Impostazione tipologia delle variabili
6
7 % Definisco la variabili Gender come qualitativa nominale (categorical
8 % senza ordine tra le modalità)
9 Xt.Gender = categorical(Xt.Gender);
10 % Definisco la variabili Education come qualitativa ordinale (categorical
11 % con ordine tra le modalità A<B<C)
12 % Per ulteriori informazioni su questa operazione consultare
13 % https://it.mathworks.com/help/matlab/matlab\_prog/convert-table-variables-containing-strings-to-categorical.html
14 %
15 Xt.Education=categorical(Xt.Education,{'A','B','C'},'Ordinal',true);
16 summary(Xt)
```

L'output è il seguente

```
1 Variables:
```