

Analisi dei dati

Esercizi

4 marzo 2018

1 Parte A

1. Si considerino n osservazioni numeriche positive y_1, \dots, y_n ad esempio ottenute con

```
> set.seed(1)
> n<-30
> y<-1+rpois(n,3)
```

e raggruppate in una tabella di frequenza

$$\begin{array}{cccc} v_1 & v_2 & \cdots & v_k \\ f_1 & f_2 & \cdots & f_k \end{array}$$

ovvero

```
y
 1  2  3  4  5  6  7  9
1  3 11  3  6  3  2  1
```

- (a) Scrivere il codice di una funzione in R che calcola la media armonica e di una funzione in R che calcola la media geometrica.
2. Si considerino n osservazioni ordinate $y_1 \leq \dots \leq y_n$ e l'indice di Gini

$$A = \frac{1}{2} - \sum_{i=1}^n \frac{(F_i - F_{i-1})(Q_i + Q_{i-1})}{2}$$

$$F_i = \frac{i}{n} \quad Q_i = \frac{\sum_{j=1}^i y_j}{\sum_{j=1}^n y_j}$$

- (a) Si scriva il codice una funzione in R che calcola l'indice. A titolo di esempio si consideri il vettore delle osservazioni (NA indica un valore mancante)
- $$1, 2, 3, 10, 15, 15, \text{NA}, 30, 50$$
- (b) Si rappresenti la curva di Lorenz associata utilizzando il *package* `ggplot2`
3. Si consideri il package `Lock5Data` e lo si installi

```
> install.packages("Lock5Data")
> library(Lock5Data)
```

- (a) Carica in R il *dataframe* `SleepStudy`

```
> data("SleepStudy")
e considera l'help
```

`> help(SleepStudy)`

(b) Crea una tabella delle variabili **Stress** e **AlcoholUse**

- i. Qual è la proporzione di studenti in ciascuna categoria di uso di alcol?
- ii. Qual è la proporzione di studenti nel gruppo ad alto stress che presentano un consumo elevato di alcol?
- iii. Visualizza i dati di questa tabella in un grafico a barre che confronta in modo efficace la distribuzione dell'uso di alcol per ciascun gruppo di stress.
- iv. Scrivi il codice R che hai usato per creare questo grafico.
- v. Commenta i risultati che ricavi dall'analisi.

(c) Considera le variabili **Drinks** e **LarkOwl**

- i. Trova la media, la mediana e lo scarto quadratico medio del numero di bevande alcoliche a settimana per l'intero gruppo di studenti.
- ii. Ripeti l'analisi ma separatamente per ogni gruppo di studenti che si classificano come mattinieri (*larks*), nottambuli o nessuno dei due
- iii. Crea un'efficace visualizzazione della variabile Bevande che mostra come la distribuzione del numero di **drinks** a settimana varia tra mattinieri, nottambuli e nessuno dei due.
- iv. Commenta i risultati che ricavi dall'analisi.