

Esercitazione pratica

Gaetano Scaduto

2023-10-18

Introduzione

Questo file contiene il testo degli esercizi pratici da svolgere in R (o in altri software statistici, se preferite). Potete cercare online soluzioni ai problemi che incontrerete, e anzi vi incoraggio a farlo. Potete anche interrogare intelligenze artificiali se volete (ma occhio: sbagliano molto più spesso di quanto non sembri). Un'altra utilissima fonte è cercare le soluzioni sulla **traduzione italiana del più semplice e completo manuale di R che ci sia, che trovate a questo link** [clicca qui]. Quasi tutti gli esercizi possono essere svolti con piccole variazioni nel codice visto a lezione. Gli asterischi indicano il livello di difficoltà (*=semplice, **=medio, ***=difficile). Tutti gli esercizi devono essere svolti utilizzando il dataset visto a lezione, che dovrete caricare nell'ambiente Rstudio per svolgere tutti gli esercizi (rivedete il pdf della lezione se non ricordate come si fa e ricordate di caricare la libreria “rio” con il comando `library(rio)`). Ricordate anche rimuovere i missing data col comando

```
data = data[complete.cases(data), ]
```

come visto a lezione prima di iniziare a lavorare, altrimenti vi darà errore!

Non abbiate paura se la prima volta che eseguite un comando vi dà errore. Controllate bene di non aver fatto errori di battitura (anche In mErItO al CaSE). Dopodiché, provate a cercare su internet il testo dell'errore (in inglese!). Provate anche a chiedere a chatgtp. Vale tutto!

Ricordate anche di fare molta attenzione al fatto che R è case sensitive, quindi dovete mettere le maiuscole e le minuscole nel modo giusto!

Vabè dai, vi dò una mano. I primi tre comandi che dovete eseguire sono questi. Ma ricordate che, per funzionare, dovete aver installato la libreria “rio” la scorsa volta con il comando `install.packages()`. Se l'avete fatto l'altra volta e avete settato la cartella giusta la volta scorsa con il comando `setwd()` e non avete spostato file, andrà tutto bene.

```
library(rio)
```

```
## Warning: il pacchetto 'rio' è stato creato con R versione 4.1.3
```

```
data = import("ris2022_lezione.RDS")
data = data[complete.cases(data), ]
```

Esercizio 1*

- Utilizzando il comando `summary()` sul dataset, ispezionate il suo output e riportate, per ogni variabile nel dataset, di che tipo di variabile si tratta (numerica o categoriale).

```
summary(data)
```

- C'è qualche variabile che R “frintende”? Se sì, quale?

R frintende la variabile “istat”, che interpreta come numerica anche se non ha un significato numerico, ma categoriale (si tratta di un codice per identificare i comuni univocamente)

Esercizio 2*

- Per ognuno dei sette partiti presenti nel dataset, calcola la media e la mediana dei voti ricevuti, utilizzando i comandi `mean()` e `median()` come visto a lezione

```
mean(data$Lega)
```

```
## [1] 11.46451
```

```
median(data$Lega)
```

```
## [1] 10.805
```

```
mean(data$Fi)
```

```
## [1] 9.7061
```

```
median(data$Fi)
```

```
## [1] 8.6
```

```
mean(data$AVS)
```

```
## [1] 2.81466
```

```
median(data$AVS)
```

```
## [1] 2.61
```

```
mean(data$M5S)
```

```
## [1] 13.1321
```

```
median(data$M5S)
```

```
## [1] 9.99
```

```
mean(data$PD)
```

```
## [1] 16.79158
```

```
median(data$PD)
```

```
## [1] 15.88
```

```
mean(data$Aziv)
```

```
## [1] 6.350031
```

```
median(data$Aziv)
```

```
## [1] 6.165
```

```
mean(data$Fdi)
```

```
## [1] 29.11733
```

```
median(data$Fdi)
```

```
## [1] 29.83
```

- Per quali partiti la media e la mediana sono più diversi? Riesci ad immaginare il perché?

Lo sono soprattutto per i 5 stelle, che hanno preso tanti voti al sud (che ha meno comuni) e pochi voti al nord (che ha più comuni), quindi la distribuzione è più asimmetrica.

Esercizio 3*

- Utilizzando il comando `table()` come visto a lezione, in quali regioni fratelli d'italia non è il partito più votato nella maggior parte dei comuni.
 - Indizio: per farlo, ti servirà incrociare solo due variabili già presenti nel dataset.

```
table(data$regione, data$most_voted)
```

```
##
##               AVS Aziv  fdi   Fi lega  M5s  PD
##  Abruzzo           1    0  246    5    3   19  31
##  Basilicata         0   13   30    9    5   58  16
##  Calabria           0    1   55   77    2  242  27
##  Campania           2    2  162   18   11  275  80
##  Emilia-Romagna     0    0  216    1    0    0 111
##  Friuli-Venezia Giulia 0    0  197    0    2    0  15
##  Lazio              0    1  352    5    1    4  15
##  Liguria             0    0  200    1    4    0  29
```

##	Lombardia	0	1	1456	0	21	0	25
##	Marche	0	0	202	0	0	1	22
##	Molise	0	0	57	8	2	49	20
##	Piemonte	1	1	1102	6	12	4	53
##	Puglia	1	0	115	4	0	118	18
##	Sardegna	1	0	228	10	3	80	54
##	Sicilia	0	5	108	30	5	223	17
##	Toscana	0	1	170	2	0	0	100
##	Trentino-Alto Adige	0	0	149	0	1	0	15
##	Umbria	0	0	81	0	2	0	9
##	Veneto	0	0	554	0	2	0	7

Come è possibile vedere, in Basilicata, Campania, Calabria, Puglia e Sicilia i 5 Stelle hanno vinto in più comuni rispetto a FDI.

- Forza Italia ha vinto soprattutto nei comuni di una regione in particolare. Sapresti dire quale?
La Calabria

Esercizio 4**

- In classe abbiamo visto come la macroarea (nord, centro, sud) spiega il voto ai Cinque Stelle in maniera molto efficace. Replica la regressione univariata vista in classe in cui regredivamo la percentuale presa dai 5 stelle sulla variabile macroarea per un altro partito a tua scelta e confronta i risultati con quelli dei cinque stelle.
 - Ricorda di caricare la libreria “texreg” col comando `library(texreg)`. Vedi il pdf della lezione.

Io ho deciso di farlo con la Lega

```
library(texreg)
```

```
## Warning: il pacchetto 'texreg' è stato creato con R versione 4.1.3
```

```
## Version: 1.38.6
## Date: 2022-04-06
## Author: Philip Leifeld (University of Essex)
##
## Consider submitting praise using the praise or praise_interactive functions.
## Please cite the JSS article in your publications -- see citation("texreg").
```

```
regr = lm(data=data, Lega ~ macroarea)
screenreg(regr)
```

```
##
## =====
##                               Model 1
## -----
## (Intercept)                8.44 ***
##                               (0.15)
## macroareaNorth-East        5.46 ***
##                               (0.19)
## macroareaNorth-West        6.90 ***
```

```
##                                (0.17)
## macroareaSouth                -1.47 ***
##                                (0.17)
## -----
## R^2                            0.41
## Adj. R^2                       0.41
## Num. obs.                      7700
## =====
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

- L'R quadro è più alto o più basso? Cosa significa?

L'R-quadro è più basso rispetto a quello dei 5 Stelle, ma comunque rilevante

- Come interpreti gli effetti descritti dal modello?

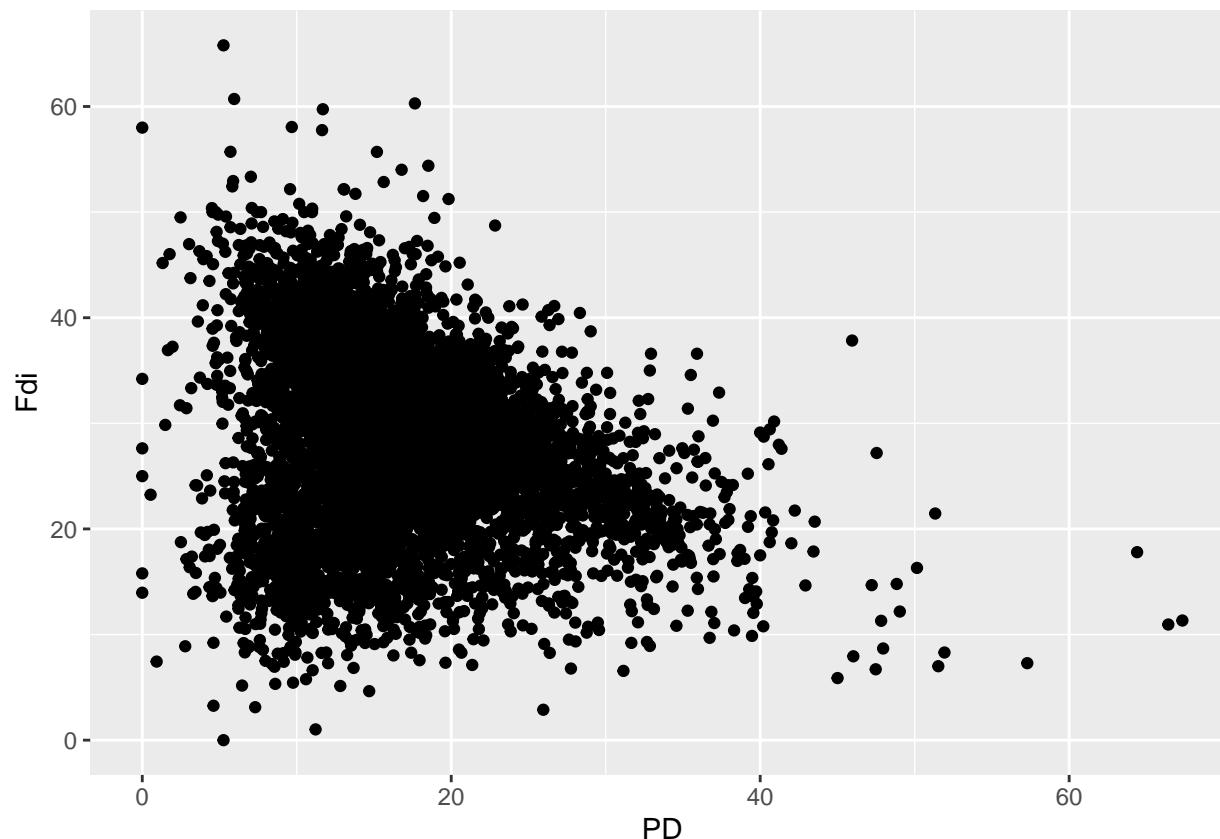
Che la Lega prende più voti al nord rispetto che al sud e al centro (e grazie)

Esercizio 5**

- Utilizzando la libreria ggplot2 ed il comando visto a lezione (`geom_point()`), fai uno scatterplot (grafico coi puntini) dei voti del PD contro i voti di Fratelli d'Italia. Riesci a vederci una relazione?
 - Ricordati di caricare ggplot2 col comando `library(ggplot2)`

```
library(ggplot2)

ggplot(data, aes(x=PD, y=Fdi))+
  geom_point()
```



Sembrerebbe proprio che ci sia una relazione di proporzionalità inversa.

Esercizio 6***

- Utilizzando il comando `max()` trova qual è la massima percentuale presa dal Movimento 5 stelle in un comune.

```
max(data$M5S)
```

```
## [1] 85.12
```

- Una volta ottenuto questo numero, usalo per trovare il nome del comune in cui i 5 Stelle hanno preso il massimo
- Indizio: puoi usare tanti diversi modi per farlo. Uno lo abbiamo visto a lezione (ma non sta nel pdf). Altri puoi trovarli cercando su internet o ragionandoci un po' su.

Qui hai due opzioni: opzione facile fai `View(data)`, poi clicchi su filter e scrivi 85.12 nella colonna corrispondente a M5S, e poi leggi il nome del comune. Opzione difficile (ma più veloce):

```
data[data$M5S==85.12, ]$comune
```

```
## [1] "VOLTURARA APPULA"
```

- Che cos'ha questo comune di speciale?

È il paese di Giuseppe Conte.