

Esercizi in R relativi al libro
Impariamo R
di Stefano Leonardi
D-Scuola

8 agosto 2025

Esercizio 1 Lo scopo di questi due esercizi è quello di imparare a convertire delle formule matematiche in comandi e funzioni di R.

Cercate la formula della media ponderata e calcolate in R la media dei vostri voti pesando il voto degli esami per i crediti dell'esame stesso.

Esercizio 2 Tenendo presente le seguenti definizioni:

- media:

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N}$$

- devianza (o somma degli scarti dalla media quadratici):

$$dev(x) = \sum_{i=1}^N (x_i - \bar{x})^2$$

- varianza:

$$var(x) = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N - 1}$$
$$var(x) = \frac{dev(x)}{N - 1}$$

- varianza con la formula computazionale (o alternativa):

$$var(x) = \sum_{i=1}^N x_i^2 - \frac{\left(\sum_{i=1}^N x_i\right)^2}{n}$$

- deviazione standard:

$$sd(x) = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N - 1}}$$
$$sd(x) = \sqrt{var(x)}$$

- covarianza:

$$cov(x, y) = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{N - 1}$$

- e correlazione di Pearson:

$$cor(x, y) = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left(\sum_{i=1}^N (x_i - \bar{x})^2\right) \left(\sum_{i=1}^N (y_i - \bar{y})^2\right)}}$$

Generate un vettore casuale di 20 numeri interi con una distribuzione uniforme con minimo=0 e massimo=50 utilizzando la funzione `runif()`, Creare le funzioni che svolgano i calcoli dei parametri statistici di cui sopra e confrontate i risultati con quelli delle funzioni apposite di R. In R non esiste una funzione appositata per calcolare la devianza. Per la covarianza e la correlazione è necessario generare anche un secondo vettore di dati. Applicate la vostra funzione ad almeno un vettore e controllate che dia gli stessi risultati della funzione appositata di R (se esiste) .

Esercizio 3 Esercitatevi ad importare i file in R da Excel o da un altro foglio elettronico. Potete provare ad importare il secondo foglio del file `SLA.xls` che abbiamo usato in classe, ma vi consiglio anche di provare con degli altri dati.

Esercizio 4 Esercizio sull'uso dei vettori logici come indici:

1. Creare la funzione `is.pari()` che prenda un vettore numerico come argomento e restituisca dei `TRUE` o `FALSE` per ogni elemento del vettore a seconda che il numero sia rispettivamente pari o dispari ¹.
2. Generare un vettore `x` casuale di 20 numeri distribuiti normalmente con media 5 e deviazione standard 8 e arrotondarlo a due cifre decimali. Con un unico comando sostituire i numeri negativi di `x` con zero.
3. Generare un vettore `a` casuale di 20 numeri interi distribuiti uniformemente da 1 a 50. Utilizzando la vostra funzione `is.pari`,

¹Suggerimento: usate l'operatore per il resto di una divisione

con un solo comando, estrarre i numeri pari di `a` e assegnarli ad un vettore nuovo chiamato `apari`. Fare la stessa cosa con i numeri dispari creando il vettore nuovo `adispari`.

4. Estrarre gli elementi di `x` corrispondenti agli elementi di `a` che sono pari e assegnarli ad un vettore nuovo chiamato `xapari`. Quindi fare la stessa cosa con i dispari e creare il nuovo vettore `xadispari`
5. Concatenate `xapari` e `xadispari` in un vettore chiamato `xa`. Quindi, con l'aiuto delle funzioni `rep` e `as.factor`, generate un nuovo vettore `ParDisp` di fattori lungo tanto quanto `xa` e contenente le stringhe "Pari" e "Dispari" a seconda che il corrispondente elemento di `xa` sia originato da `xapari` o `xadispari`.
6. Creare un dataframe con i due vettori `xa` e `ParDisp`. Quindi effettuare un test T di Student con la funzione `t.test()` su `xa` per confrontare le medie dei due gruppi `Pari` e `Dispari`. Commentare il risultato e presentare un grafico che lo illustri.

Esercizio 5 Produrre un vettore `weight` (grammi) di 15 dati casuali, estratti da una distribuzione normale con $\mu = 40$ e $\sigma = 10$ utilizzando la funzione `rnorm()`, creare una funzione per standardizzarlo, cioè che tolga la media e divida il risultato per la deviazione standard. Confrontate il risultato con quello della funzione `scale()`.

Esercizio 6 Spiegare la differenza fra le seguenti espressioni in R:

```
rep(1:3,4) ,  
rep(1:3,c(4,4,4)) ,  
rep(1:3,rep(4,3)),  
rep(1:3,each=4) e  
rep(1:3,1:3)  
e perchè rep(1:3,rep(3,4)) non funziona?
```

Esercizio 7 Rispondere alle seguenti domande, riportare i comandi e l'output di R e commentate:

- Se `x<-c(2,2,3)` e `y<-1:6`. Cosa succede se provo `x * y`?
- Perché il risultato di `x * y`, dove `x<-c(2,2,3)` e `y<-1:7`, è così ed ho un `warning`? Perché nel caso precedente non ho il `warning`?
- Cosa succede se moltiplico/sommo/divido/sottraggo due vettori di lunghezza diversa? Questa è la "**Regola del riciclaggio dei vettori**" che si applica quando i vettori sono di lunghezza diversa. Provate a spiegare con parole vostre questa regola.

Esercizio 8 Scrivere due funzioni che calcolino la traccia di una matrice quadrata (somma degli elementi sulla diagonale). Nella prima usare la funzione apposita di R per estrarre la diagonale. Nella seconda usare un ciclo `for` per accedere agli elementi della matrice. Applicarle ad una matrice quadrata 12×12 con numeri con due decimali estratti casualmente da una distribuzione normale con $\mu = 10$ e $\sigma = 5$ e verificarne il corretto funzionamento.

Esercizio 9 Con un ciclo `for` calcolare le medie e la deviazione standard della lunghezza dei sepali per ciascuna delle tre specie di iris nel dataset `iris`, ottenibile con il comando `data(iris)`. Un modo per risolverlo è quello di effettuare un `subset()` del database originale ad ogni ciclo del `for` e selezionare i dati relativi ad una sola specie creando un dataframe temporaneo. Quindi calcolare la media e la deviazione standard e immetterle in strutture-dati adeguate a contenere le statistiche richieste. Confrontare i risultati con quelli del comando `tapply()`.

Esercizio 10 Caricare la libreria MASS, che tutti avete a disposizione, con il comando `library(MASS)`, quindi caricate il dataset pronto per l'uso `immer` con il comando `data(immer)`. Il data set contiene la produzione di 30 campi di orzo in 6 località definite dal vettore `Loc`. In ciascuna località sono coltivate 5 varietà di orzo definite dal vettore `Var`. La produzione dei campi è misurata in quintali in due anni successivi (anno 1931 e anno 1932) rappresentata rispettivamente nei due vettori `Y1` e `Y2`.

Creare un nuovo dataframe utilizzando i dati del precedente, aggiungendo un nuovo fattore che definisca (utilizzate la funzione `rep()`) l'anno di produzione. Aggiungete un unico vettore per le produzioni concatenando i due vettori delle produzioni del dataset precedente. Riportate anche i dati relativi alle località e alle varietà nel nuovo dataframe.

Effettuate un test statistico adatto per verificare se le produzioni differiscono in modo significativo fra i due anni. Effettuate anche un test non parametrico adatto.

Esercizio 11 Ripetere il test di permutazione che ha mostrato John Rauser nel suo video disponibile su youtube o sul sito `elly.scvsa.unipr.it`: *"Statistics Without the Agonizing Pain"*. I dati sono visibili attorno al minuto 3:37 del video. Effettuate almeno 50 mila ricombinazioni e abbondate con commenti in modo da far capire che avete afferrato bene il concetto e la parte tecnico-informatica del test.