

# Laboratorio di Statistica Applicata - 1

Federico Plazzi

10 Novembre 2017

# Introduzione ad R

## Le prime cose da sapere

- ▶ Far partire R
- ▶ Oggetti e funzioni
- ▶ Esempi di funzioni: `quit()`, `history()`
- ▶ Esempi di oggetti: `pi`, `discoveries`
- ▶ Chiedere informazioni ad R su R:
  - ▶ `il ?`
  - ▶ completare con il Tab

## I pacchetti

- ▶ La funzione `install.packages()`
- ▶ La funzione `library()`
- ▶ Pacchetti interessanti: `ggplot2`, `FactoMineR`
- ▶ Pacchetto `ggplot2movies`: il database movies

# Analisi del database `movies`

## Ottenere il database `movies`

- ▶ Far partire R
- ▶ Menu Packages > Install package(s)...
- ▶ Selezionare un mirror (tipicamente il più vicino geograficamente)
- ▶ Selezionare il pacchetto `ggplot2movies`
- ▶ Rispondere di sì ad entrambe le richieste di R
- ▶ Il database è un oggetto R del tipo `"data.frame"` e si chiama `movies`

# Analisi del database movies

## Ottenere il database movies

- ▶ Far partire R
- ▶ Menu Packages > Install package(s)...
- ▶ Selezionare un mirror (tipicamente il più vicino geograficamente)
- ▶ Selezionare il pacchetto `ggplot2movies`
- ▶ Rispondere di sì ad entrambe le richieste di R
- ▶ Il database è un oggetto R del tipo “`data.frame`” e si chiama `movies`

## Il database movies

Il database `movies` è un dataframe (cioè una tabella) che contiene 58.788 film, per ciascuno dei quali sono disponibili alcuni dati forniti dall'Internet Movie DataBase (IMDB; <http://imdb.com>).

A titolo di esempio, per vedere solo le prime venti righe bisogna digitare `movies[1:20,]`.

# Le variabili del database movies

## Cosa c'è nelle colonne?

- ▶ title: il titolo del film
- ▶ year: anno di uscita nelle sale
- ▶ length: la lunghezza in minuti
- ▶ budget: il budget in dollari, se disponibile
- ▶ rating: il voto medio da parte degli utenti IMDB
- ▶ votes: il numero di utenti IMDB che hanno assegnato un voto al film
- ▶ r1-r10: decili della distribuzione dei voti
- ▶ mpaa: classificazione MPAA
- ▶ action, animation, comedy, drama, documentary, romance, short: 1 se il film è classificabile in quel genere, 0 in caso contrario

Per trovare tutte queste informazioni, è sufficiente digitare  
?movies.

# Variabili e vettori

## Dalle variabili ai vettori

- ▶ Creiamo una variabile
- ▶ Creiamo un vettore: la funzione `c()`
- ▶ Sottoinsiemi di vettori: le parentesi quadre

## Tablelle

- ▶ Le tabelle più semplici di R: i data frame
- ▶ La funzione `data.frame()`
- ▶ Richiamare le colonne di un data frame: il `$`
- ▶ La funzione `attach()`
- ▶ Richiamare parte di un data frame (per righe e per colonne)

## Alcune funzioni utili

- ▶ La funzione `sum()`
- ▶ La funzione `mean()`
- ▶ La funzione `median()`
- ▶ La funzione `hist()`
- ▶ Le funzioni `qqnorm()` e `qqline()`
- ▶ La funzione `shapiro.test()`

## Esportare i dati

- ▶ Le funzioni `dev.copy()`
- ▶ La funzione `write.table()`

# Al lavoro!

## Alcuni esercizi. . .

- ▶ Costruire un data frame che escluda i cortometraggi
- ▶ Costruire un data frame che escluda i cortometraggi e che contenga solo i film con almeno 1000 votanti
- ▶ Calcolare il voto medio di almeno due tipi diversi di film
- ▶ Calcolare la varianza del budget nei film d'azione
- ▶ Rappresentare graficamente la distribuzione della durata dei film romantici

# Normalità e test t

## Alcuni esercizi. . .

- ▶ Stabilire se i voti si distribuiscono in modo normale
- ▶ Stabilire se il budget dei film d'azione si distribuisce in modo normale
- ▶ Stabilire se il numero di votanti per le commedie si distribuisce in modo normale
- ▶ Stabilire se i voti dei film d'animazione appartengono ad una popolazione di voti con media pari a 8
- ▶ Stabilire se i film d'azione ricevono voti significativamente più bassi di quelli dei film romantici
- ▶ Stabilire se il budget stanziato per i film d'animazione è significativamente più alto di quello stanziato per i documentari