

C19

*filippo*

*May 3, 2018*

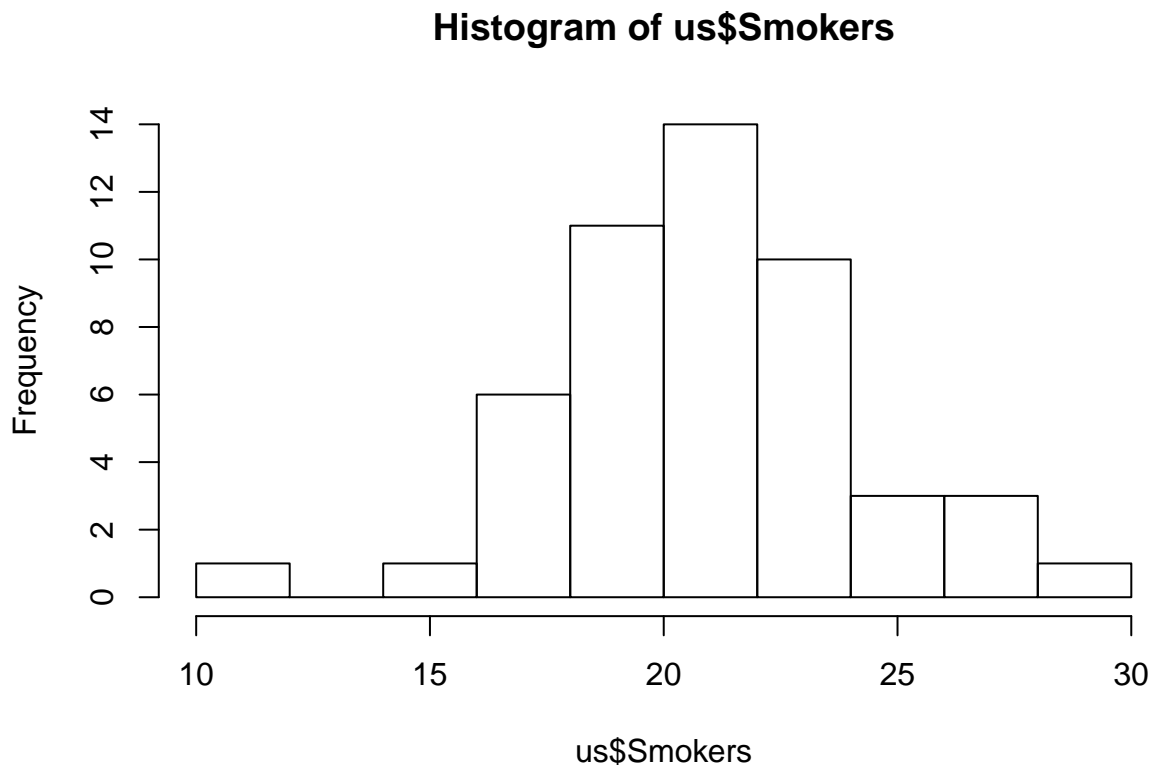
## Es 19 - USStates

Si consideri il file `USStates.csv`

(a)

Si utilizzi la *library* per disegnare un opportuno grafico della variabile `Smokers`, la percentuale di residenti che fuma. Si commenti la forma della distribuzione e il suo valore centrale.

```
hist(us$Smokers)
```



(b)

Si calcoli la proporzione media di fumatori

(c)

Si scriva una funzione in R che estrae 10000 campioni casuali di 5 stati and ritorna come output la proporzione media di fumatori per ogni campione. Si salvi l'output in un vettore `estimates`

**(d)**

Cosa rappresenta l'output di tale funzione ?

**(e)**

Si calcoli la media aritmetica e lo scarto quadratico medio di `estimates` e utilizzate `ggplot2` per rappresentare la densità e la funzione di ripartizione empirica. Si commenti la forma della distribuzione e il suo valore centrale.

**(f)**

Quale dovrebbe essere la relazione teorica tra lo scarto quadratico di `Smokers` e `estimates`. Vi sembra rispettata empiricamente ?

**(g)**

Un campione di cinque stati è Arkansas, Florida, Pennsylvania, California, e Vermont. Si calcoli la proporzione media di questi stati.

**(h)**

Si utilizzi il valore ricavato per costruire un intervallo di confidenza di livello 95% della proporzione media dei fumatori

**(i)**

L'intervallo contiene la 'vera' proporzione media.

**(j)**

Quale proporzione dei 10000 valori di `estimates` in realtà contiene il vero valore ?

**(k)**

Quale proporzione ci si attende ? Si è vicini a questo valore