Appunti sulla proprietà della mediana come stimatore della tendenza centrale di una popolazione

$$E(\widehat{Me_i}) = M_e = \mu,$$

$$Var(\widehat{Me}) = \frac{1}{4nf(Me)^2},$$

dove $f(M_e)$ è la funzione di densità della popolazione a livello della mediana (uguale a μ nelle distribuzioni simmetriche) e n è la grandezza del campione.

In pratica, la funzione di densità della popolazione a livello della mediana, $f(M_e)$, non è nota; tuttavia se si può assumere come normale, abbiamo

$$\begin{aligned} f(Me = \mu; \mu; \sigma) = &\ \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left\{\frac{1}{2}\frac{(\mu-\mu)^2}{\sigma^2}\right\}} \\ = &\ \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left\{\frac{1}{2}\frac{0}{\sigma^2}\right\}} \\ = &\ \frac{1}{\sqrt{2\pi\sigma^2}} = \frac{1}{\sqrt{2\pi}} \times \frac{1}{\sqrt{\sigma^2}} \\ = &\ 0.3989423 \times \frac{1}{\sqrt{\sigma^2}} \\ \approx &\ 0.4 \times \frac{1}{\sqrt{\sigma^2}} \end{aligned}$$

Il valore di ordinata massima (dove $f(x)$ raggiunge il valore massimo) delle funzioni di densità di probabilità normali è quindi

$$0.4 \times \frac{1}{\sqrt{\sigma^2}}$$

che si semplifica a 0.4 nel caso della normale standard, dove $\sigma^2 = 1$.

- La varianza della distribuzione campionaria della mediana, secondo l'assunzione di normalità nella popolazione, diventa

$$\boxed{Var(\widehat{Me})} = \frac{1}{4n\left(0.4 \times 1/\sqrt{\sigma^2}\right)^2} = \frac{\sigma^2}{4n(0.4)^2} = \frac{1}{4(0.4)^2} \times \frac{\sigma^2}{n}$$

$$= 1.5625\frac{\sigma^2}{n}$$

…Simulazione 1

$$= \boxed{1.25^2\sigma^2_{\bar{x}}}$$

- La formula precedente rivela che la mediana è meno efficiente della media quale operatore di tendenza centrale (ha una varianza campionaria 1.25^2 volte maggiore).
- Esaminiamo i risultati di una simulazione.
- Calcoliamo empiricamente la varianza della distribuzione campionaria di media e mediana di un campione casuale di $n = 200$ osservazioni, estratto (50000 volte) da una popolazione normale con media $\mu = 100$ e deviazione standard $\sigma = 36$.

```
rep <- 50000
n <- 200
 DistrCampMedia <- rep(0, rep)
 DistrCampMediana <- rep(0, rep)
 for (i in 1:rep) {
 samp <- rnorm(n, mean=100, sd=36)
 DistrCampMedia[i]<-mean(samp)
 DistrCampMediana[i]<-median(samp)
 }
```

I risultati confermano quello che ci aspettiamo:

```
var.media <- var(DistrCampMedia)*(n-1)/n
var.media
## 6.499387

36^2/200
## 6.48
```

la varianza della distribuzione campionaria della media è uguale al rapporto tra la varianza della popolazione e la numerosità del campione.

Vediamo ora cosa è successo per la mediana calcolata sui medesimi campioni.

```
var.mediana <- var(DistrCampMediana)*(n-1)/n
var.mediana
## 10.06176

1/(4* dnorm(0,0,1)^2) * 36^2/200
## 10.17876
```

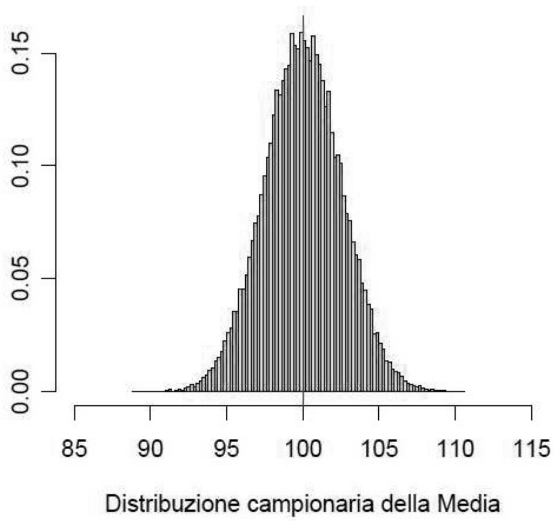
Di quanto è più grande la deviazione standard della distribuzione campionaria della mediana della deviazione standard della distribuzione campionaria della media?

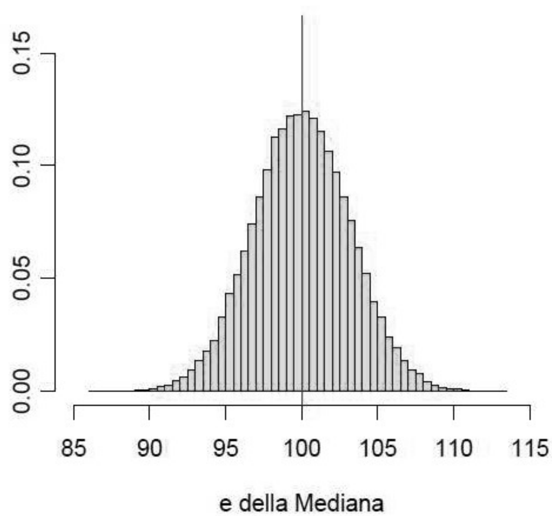
```
 sqrt(var.mediana)/sqrt(var.media)
## [1] 1.24423

 var.mediana/var.media
## [1] 1.548109
 sqrt(1.548109)
## [1] 1.24423
```

Come vi ho mostrato, quando n è grande e la popolazione si assume essere normale, la varianza della distribuzione campionaria della mediana è

$$V(Me_i) = 1.25^2 \frac{\sigma^2}{n} \approx 10.17876$$





- La simulazione precedente rivela che la mediana è meno efficiente della media quale operatore di tendenza centrale (ha una varianza campionaria maggiore).