Laboratorio di ST1 - Lezione 6

Antonietta di Salvatore

Dipartimento di Matematica
Universitá degli Studi Roma Tre

Outline

- l'intervallo di confidenza: approccio frequentista
- intervalli di confidenza per la differenza di di due medie
  - con varianze note e diverse
  - con varianze note e uguali
  - con varianze non note e uguali
- intervalli di confidenza simultanei

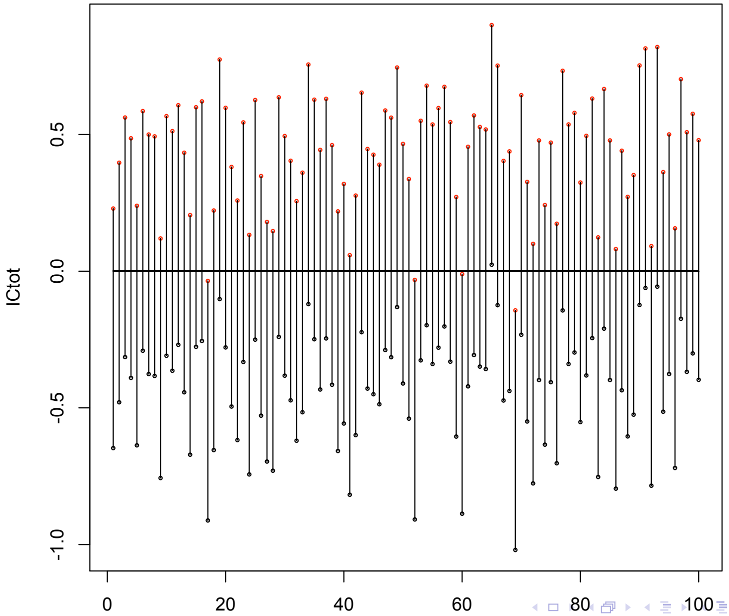
Quando costruiamo un intervallo di confidenza per un parametro θ al livello di fiducia $(1-\alpha)\%$, possiamo affermare che sulla base del campione osservato, riponiamo una fiducia del $(1-\alpha)\%$ che esso sia uno di quelli che contiene θ.
ES: costruiamo 100 intervalli di confidenza per la media di una popolazione Normale con varianza nota a partire da 100 campioni estratti dalla stessa variabile aleatoria ($N(0,1)$)

```
ICtot=data.frame()

for (i in 1:100){
   x=rnorm(20)
   xm=mean(x)
   z=qnorm(0.975)
   IC=xm+c(-1,1)*z*1/sqrt(20)
   ICtot=rbind(ICtot,IC)}
names(ICtot) <- c('c1','c2')

matplot(ICtot,pch=1,cex=0.4,main = 'Simulazione di intervalli di
confidenza al 95%')
for (i in 1:100){
lines(c(i,i),c(ICtot$c1[i],ICtot$c2[i]))}
lines(c(1,100),c(0,0))
```

Simulazione di intervalli di confidenza al 95%



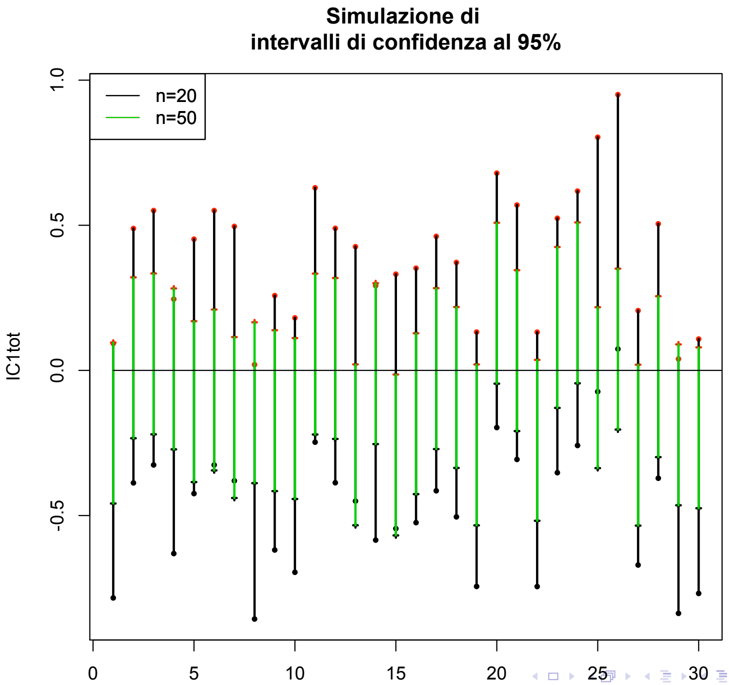
Vediamo ora l’impatto della numerositá campionaria

```
ICtot=data.frame()
IC1tot=data.frame()

rip=30
for (i in 1:rip){
   x=rnorm(50)
   x1m=mean(x[1:20])
   xm=mean(x)
   z=qnorm(0.975)
   IC1=x1m+c(-1,1)*z*1/sqrt(20)
   IC=xm+c(-1,1)*z*1/sqrt(50)
   ICtot=rbind(ICtot,IC)
   IC1tot=rbind(IC1tot,IC1)}
names(ICtot) <- c('c1','c2')
names(IC1tot) <- c('c1','c2')

matplot(IC1tot, pch=1, cex=0.4, lwd=2, main = 'Simulazione di
intervalli di confidenza al 95%')
matpoints(ICtot, pch=3, cex=0.4, lwd=2)

for (i in 1:rip){
   lines(c(i,i),c(IC1tot$c1[i],IC1tot$c2[i]), lwd=2)
   lines(c(i,i),c(ICtot$c1[i],ICtot$c2[i]), col=3, lwd=2)}
lines(c(1,100),c(0,0))
```


Simulazione di
intervalli di confidenza al 95%
n=20
n=50
IC1tot
1.0
0.5
0.0
-0.5
0
5
10
15
20
25
30

Intervalli di confidenza per la differenza tra due medie

Siano X e Y due variabili casuali indipendenti tali che $X \backsim N(\mu_X, \sigma_X^2)$ e $Y \backsim N(\mu_Y, \sigma_Y^2)$. Allora valgono i seguenti risultati

$$X + Y \backsim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2), \qquad X - Y \backsim N(\mu_X - \mu_Y, \sigma_X^2 + \sigma_Y^2)$$

La somma e la differenza di due variabili casuali gaussiane sono ancora variabili casuali gaussiane aventi come valore atteso rispettivamente la somma e la differenza dei valori attesi e come varianza la somma delle varianze in entrambi i casi. Siano

$X_1, \ldots, X_n$ e $Y_1, \ldots, Y_m$ due campioni casuali indipendenti di numerositá n e m estratti rispettivamente da X e Y. Siamo $\bar{X}_n$ e $\bar{Y}_m$ le rispettive stime delle medie campionarie, allora si ha che

$$\bar{X}_n + \bar{Y}_m \backsim N(\mu_X + \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}), \qquad \bar{X}_n + \bar{Y}_m \backsim N(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m})$$

X=c(24, 24, 21, 23, 27, 24, 16, 25, 23, 29, 32, 25, 23, 26, 15, 27, 15, 24, 21, 27, 26, 20, 22, 28, 20, 31, 33, 19, 27, 30, 29, 25,18, 28, 23, 32, 32, 20, 32, 28, 24, 33, 24, 19, 24, 23, 29, 22, 24, 29, 23, 23, 22, 25, 27, 25, 22, 14, 25, 29, 28, 23, 24, 23, 34, 27, 23, 18, 20, 29)

Y=c(27, 18, 16, 25, 20, 20, 19, 16, 19, 20, 20, 18, 22, 23, 19, 15, 18, 25, 22, 24, 14, 23, 21, 17, 18, 18, 23, 19, 25, 20, 23, 17,12, 22, 17, 20, 23, 25, 22, 20, 20, 16, 22, 18, 17, 21, 22, 21, 19, 21)

W=c(35, 33, 19, 48, 31, 24, 27, 13, 27, 16, 18, 19, 34, 24, 34, 41, 23, 25, 20, 27, 30, 44, 16, 25, 24, 31, 34, 41, 36, 25, 36, 30, 12, 26, 28, 35, 24, 38, 37, 46, 25, 25, 30, 31, 13, 28, 51, 36, 19, 27, 22, 22, 21, 31, 29, 35, 22, 51, 36, 44)

caso 1 - varianze note

```
XM=mean(X)
YM=mean(Y)
S2x=16 ♯ supponiamo nota
S2y=9 ♯ supponiamo nota
n1=length(X)
n2=length(Y)
```

IC al livello di significativitá 0.95%

```
a=0.05
Z=qnorm(1-a/2)
IC1=XM-YM + c(-1,1)*Z*sqrt(S2x/n1+S2y/n2)
```

IC al livello di significativitá 0.99%

```
a=0.01
Z=qnorm(1-a/2)
IC2=XM-YM + c(-1,1)*Z*sqrt(S2x/n1+S2y/n2)
```

Si osserva che $IC1 \subset IC2$

caso 1 - varianze note e uguali

```
WM=mean(W)
S2w=9 ♯ supponiamo nota
n3=length(W)
```

IC al livello di significativitá 0.95%

```
a=0.05
Z=qnorm(1-a/2)
IC1=YM-WM+c(-1,1)*Z*sqrt(S2y/n2+S2w/n3)
```

IC al livello di significativitá 0.99%

```
a=0.01
Z=qnorm(1-a/2)
IC2=YM-WM+c(-1,1)*Z*sqrt(S2y/n2+S2w/n3)
```

Si osserva che $IC1 \subset IC2$

caso 2 - varianze non note ma uguali

Supponiamo di sapere che i campioni X e W provengono da due variabili Normali con stessa varianza incognita. Costruiamo l'intervallo di confidenza per la differenza delle medie.
Una stima della varianza campionaria comune é data dalla varianza campionaria *pooled*.

```
n3 = length(W)
WM = mean(W)
Vp = (var(Y)*n2+var(W)*n3)/(n2+n3-2)
a = 0.05
g=n2+n3-2
t = qt(1-a/2,g)
ICc = YM-WM+ c(-1,1)*t*sqrt(Vp*(1/n2+1/n3))
```

otteniamo lo stesso risultato usando il comando

```
t.test(Y,W,var.equal=T)
```

Osservazioni:

1) la perdita d'informazione sulle varianze comporta IC piú ampi a paritá di fiducia

```
IC1[2]-IC1[1]
```

```
ICc[2]-ICc[1]
```

2) dato che $n2 + n3 - 2 > 100$, si ha che $t_{1-\frac{\alpha}{2}, n_2+n_3-2} \approx z_{1-\frac{\alpha}{2}}$

```
t = qt(1-a/2,g)
z=qnorm(1-a/2)
```

Quindi *per grandi campioni* possiamo utilizzare anche il seguente intervallo di confidenza

```
IC = YM-WM+ c(-1,1)*z*sqrt(Vp*(1/n2+1/n3))
```

intervalli di confidenza simultanei

Dato il campione $X_1, \ldots, X_n$ estratto da una normale $N(\mu, \sigma)$, vogliamo trovare un IC simultaneo per media e varianza.
Consideriamo le quantitá pivotali

$$Q_1 = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \quad \text{and} \quad Q_2 = \frac{(n-1)S^2}{\sigma^2}$$

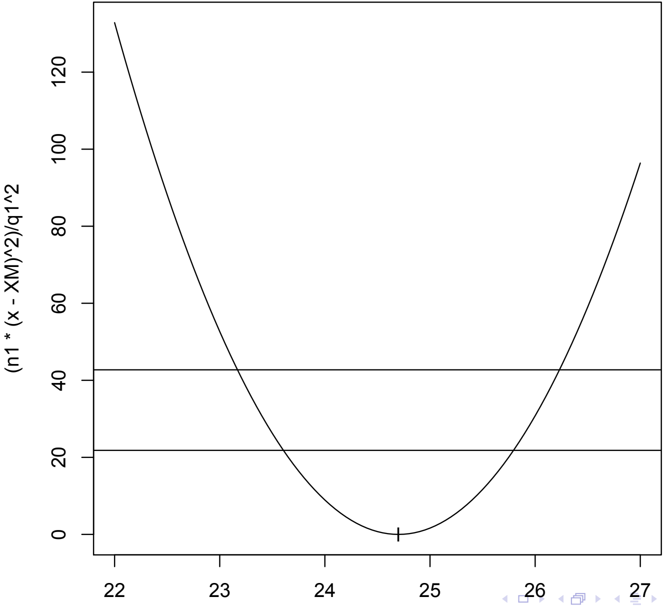
definiamo numeri q_1, q_2' and q_2'' tali che

$$P[-q_1 < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < q_1] = \gamma_1 \quad \text{and} \quad P[q_2' < \frac{(n-1)S^2}{\sigma^2} < q_2''] = \gamma_2$$

Poiché Q_1 e Q_2 sono indipendenti, possiamo costruire il seguente intervallo di confidenza simultaeo

$$P\left[-q_1 < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < q_1; q_2' < \frac{(n-1)S^2}{\sigma^2} < q_2''\right] = \gamma_1\gamma_2$$

```
alpha1=0.05
gamma1=1-alpha1
q1=qnorm(1-alpha1/2)
curve((n1*(x-XM)^ 2)/q1^ 2,22,27)
gamma2=0.95
S2=var(X)*n1/(n1-1)
q21=qchisq(0.025,n1-1)
q22=qchisq(0.975,n1-1)
lines(c(20,29),c((n-1)*S2/q21,(n-1)*S2/q21))
lines(c(20,29),c((n-1)*S2/q22,(n-1)*S2/q22))
gamma=gamma1*gamma2
points(XM,0, pch='l')
```



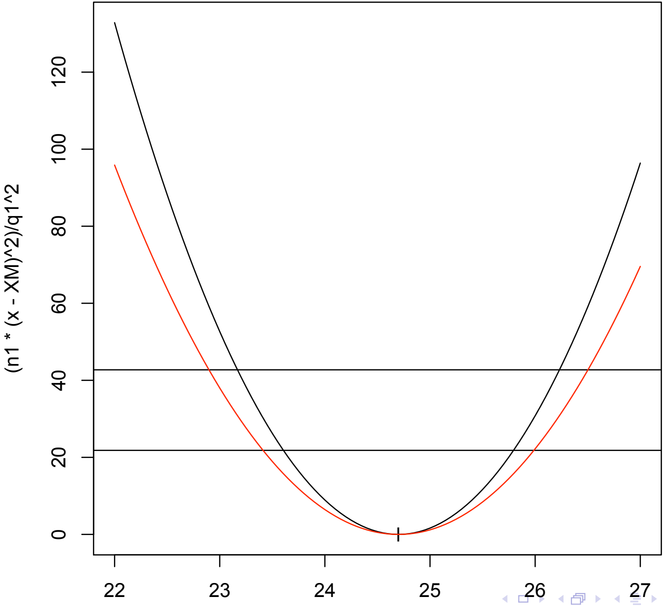
(n1 * (x - XM)^2)/q1^2
0
20
40
60
80
100
120
22
23
24
25
26
27

aumentiamo gamma

```
gamma=0.93
gamma1=gamma/gamma2
```

Si osserva che $\gamma < \gamma 2$

```
alpha1=1-gamma1
q1=qnorm(1-alpha1/2)
curve((n1*(x-XM)^ 2)/q1^ 2,22,27, col=2)
```



(n1 * (x - XM)^2)/q1^2
0
20
40
60
80
100
120
22
23
24
25
26
27

Esercizio: ripetere l'esercizio precedente mantenedo fisso γ_1 e cambiando i valori di γ_2