Laboratorio di ST1 - Lezione 4

Antonietta di Salvatore

Dipartimento di Matematica
Universitá degli Studi Roma Tre

Outline

- variabili continue
  - t di Student
- intervalli di confidenza
  - media con varianza nota
  - media con varianza non nota
  - varianza

Variabili continue

t di Stdent: $X \backsim t_{(g)}$

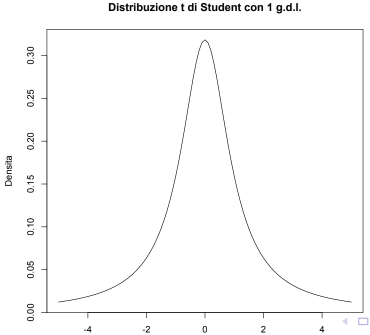
Date la v.a. Z normale standardizzata e la v.a. χ chi quadro con g g.d.l (df) la v.a.

$$T = \frac{Z}{\sqrt{\frac{X}{g}}}$$

é una t di Studentcon g g.d.l.
$E(X) = 0$
$V(X) = g/(g-2)$, per $g > 2$

Disegnamo i valori di una curva t tra -5 e 5 con il comando `curve`

```
curve(dt(x,df=1), -5, 5, ylab = 'Densitá',main = 'Distribuzione
t di Student con 1 g.d.l.')
```

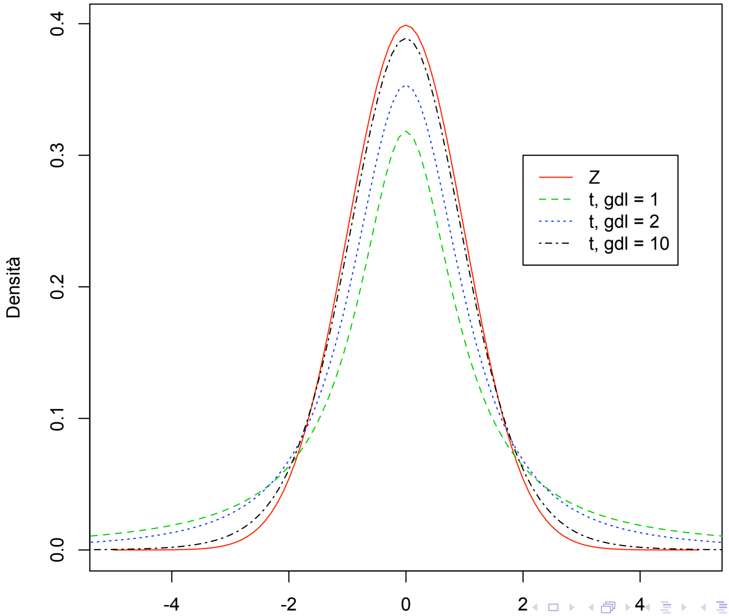


Per $g \to \infty$, la funzione di densitá della v.a. t tende in distribuzione a quella della $Z(0,1)$

```
curve(dnorm(x), -6, 6, ylab = 'f(x)', col = 2)
curve(dt(x, df = 1), -6, 6, lty = 2, col = 3, add= T)
curve(dt(x, df = 2), -6, 6, lty = 3, col = 4, add= T)
curve(dt(x, df = 10), -6, 6, lty = 4, col = 1, add= T)
legend(2,0.3, c('Z', 't, gdl = 1', 't, gdl = 2','t, gdl = 10'),
lty =c(1, 2, 3,4), col=c(2, 3, 4,1))
```

df = degrees of freedom

per $g > 50$ le differenze tra la normale standardizzata e la t di student sono minime



Densità
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
Z
t, gdl = 1
t, gdl = 2
t, gdl = 10

Intervalli di confidenza

dati

```
cm=c(5.25, 2.92, 4.65, 7.47, 6.57, 3.27, 6.68, 8.16, 0.83, 6.10,
5.79, 5.51, 1.89, 3.92, 7.92, 4.68, 4.75, 5.29, 3.28, 4.80,
2.06, 6.08, 2.04, -0.48, 7.18, 8.52, 4.49, 5.24, 3.71, 4.57)
```

Stima per intervallo della media di una popolazione Normale con σ^2 *nota*

Ricordiamo che la funzione qnorm serve per ottenere i quantili della distribuzione Normale Standardizzata, es: `qnorm(0.5), qnorm(0.95)`

La varianza della popolazione σ^2 é nota (nell'esercizio é indicata con $\mathtt{S}^2$).
Costruiamo l'intervallo di confidenza al livello di fiducia (o di significativitá) del $(1-\alpha)\%$

```
XM=mean(cm)♯ STIMA PUNTUALE
S=2 ♯ supponiamo nota la s.d.
n=length(cm)
a=0.05
Z=qnorm(1-a/2) ♯ QUANTILE
CI=c(XM-Z*S/sqrt(n),XM+Z*S/sqrt(n)) ♯ INTERVALLO DI CONFIDENZA
```

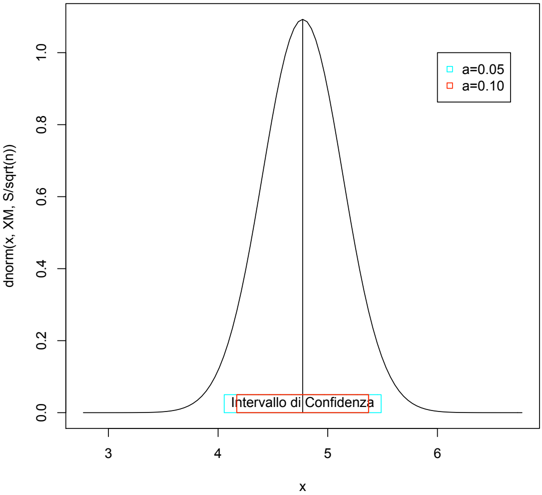
Lintervallo di confidenza ci indica il grado di fiducia che abbiamo sulla stima puntuale.
Tanto piú é ampio l'intervallo, tanto piú la fiducia é alta.

Grafico

```
n=length(cm)
curve(dnorm(x,XM,S/sqrt(n)),XM-2,XM+2)
v=seq(CI[1],CI[2],by=0.001)
x=c(CI[1],v,CI[2])
y=c(0,rep(0.05,length(v)),0)
polygon(x,y,border=5)
text(XM,0.03,'Intervallo di Confidenza')
lines(c(XM,XM),c(0,dnorm(XM,XM,S/sqrt(n))))
```

cabiamo il livello di fiducia

```
b=0.10
Zb=qnorm(1-b/2)
CIb=c(XM-Zb*S/sqrt(n),XM+Zb*S/sqrt(n))
v=seq(CIb[1],CIb[2],by=0.001)
x=c(CIb[1],v,CIb[2])
y=c(0,rep(0.05,length(v)),0)
polygon(x,y,border=2)
legend(6,1, c('a=0.05','a=0.10'),col=c(5,2),pch=22)
```


a=0.05
a=0.10
Intervallo di Confidenza
dnorm(x, XM, S/sqrt(n))
x
0.0
0.2
0.4
0.6
0.8
1.0
3
4
5
6

L'ampiezza dell'intervallo di confidenza al livello di fiducia $(1-\alpha)\%$ e con numerositá campionaria n é:

$$a_n = 2 * Z_{(\frac{\alpha}{2})} * S/sqrt(n)$$

```
ampiezza= CI[2]-CI[1] ♯ ampiezza intervallo
aa=2*Z*S/sqrt(n) ♯ ampiezza intervallo
```

Osserviamo

- fissato Z (e quindi la fiducia associata all'intervallo), all'aumentare della numerositá campionaria l'ampiezza dell'intervallo di confidenza diminuisce

  ```
  aa2=2*Z*S/sqrt(50)
  ```

- per a_n fissato, la fiducia aumenta al crescere di n

  ```
  z=ampiezza*sqrt(50)/(2*S)
  q=pnorm(z)
  ```

Osserviamo che z é il quantile della distribuzione normale standardizzata che isola alla sua sinistra un'aria di probabilitá pari a q.

Quale deve essere la numerositá n del campione affinché l'intervallo al 95$\%$ abbia ampiezza pari a 1?

$$2 * 1.96 * S/sqrt(n) = 1 \rightarrow n = (2 * 1.96 * S)^2$$

Stima per intervallo della media di una popolazione Normale con σ^2 *non nota*

La varianza della popolazione σ^2 non é nota, quindi deve essere stimata. Costruiamo l'intervallo di confidenza con coefficiente di confidenza $1 - \alpha$

```
XM=mean(cm)
n=length(cm)
S2c=var(cm)*n/(n-1)♯ VARIANZA CAMPIONARIA CORRETTA
a=0.05
t=qt(1-a/2,n-1)
CIc=c(XM-t*sqrt(S2c/n),XM+t*sqrt(S2c/n))
```

osserviamo che a paritá di α la mancato conoscenza della varianza della popolazione comporta un ampiamento dell'intervallo di confidenza (CI $\subset$ CIc).

Stima per intervallo della varianza σ^2 di una popolazione Normale

Costruiamo l'intervallo di confidenza con coefficiente di confidenza $1-\alpha$

```
n=length(cm)
S2c=var(cm)*n/(n-1) ♯ VARIANZA CAMPIONARIA CORRETTA
a=0.05
chi1=qchisq(a/2,n-1)
chi2=qchisq(1-a/2,n-1)
CI=c((n-1)*S2/chi2,(n-1)*S2/chi1)
```

Si osservi che rispetto al valore S2 che costituisce la stima puntuale di σ^2, questo intervallo non é simmetrico, ma é stato costruito in modo da avere code uguali.

In R usando la funzione `t.test` ed usando l'argomento opzionale `conf.level` si ottengono intervalli di confidenza:

```
x <- c(1,2,3,4,5,6)
mean(x) +c(-1,1)*qt(0.975,5)*sd(x)/sqrt(length(x))
t.test(x,conf.level=0.95)
```

Per fare lintervallo di confidenza per la differenza fra le medie di due campioni:

```
y <- c(2,3,4,5,6,7,8)
```

se assumo che le varianze delle 2 popolazioni sono ingognite ma UGUALI allora uso lo stimatore della varianza POOLED:

```
t.test(x,y,var.equal=T)
```

se lipotesi di uguaglianza delle 2 varianze non é valida allora R usa il metodo introdotto da WELCH:

```
t.test(x,y)
```