Laboratorio di ST1 - Lezione 7

Antonietta di Salvatore

Dipartimento di Matematica
Universitá degli Studi Roma Tre

Outline

- La costruzione del test statistico
  - errore di prima specie α e seconda specie β
- Esercizi sui test

Relazione tra α e β

A paritá di n (numerositá campionaria), la relazione che sussiste tra α e β é generalmente esprimibile mediante una funzione decrescente.

Consideriamo di aver estratto un camione di n=50 osservazioni da una $X \backsim N(\mu, 36)$ e vogliamo verificare il seguente sistema di ipotesi

$H_0 : \mu = 0$
$H_1 : \mu = 3$

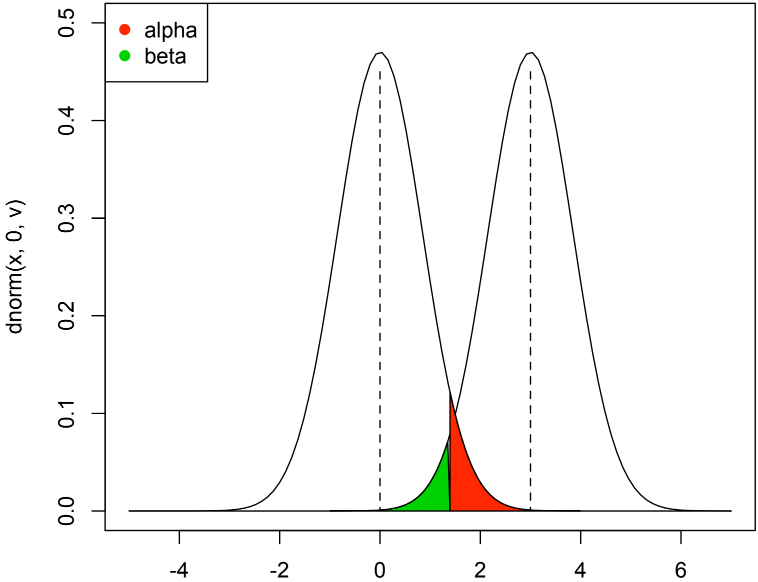
```
v=sqrt(36/50)
curve(dnorm(x,0,v),-5,7, axes = TRUE, ylim =c(0,.5))
lines(c(0,0),c(0.45,0),lty=2)
curve(dnorm(x,3,v),-5,7, ylim =c(0,.5), add=T)
lines(c(3,3),c(0.45,0),lty=2)
```

disegnamo α

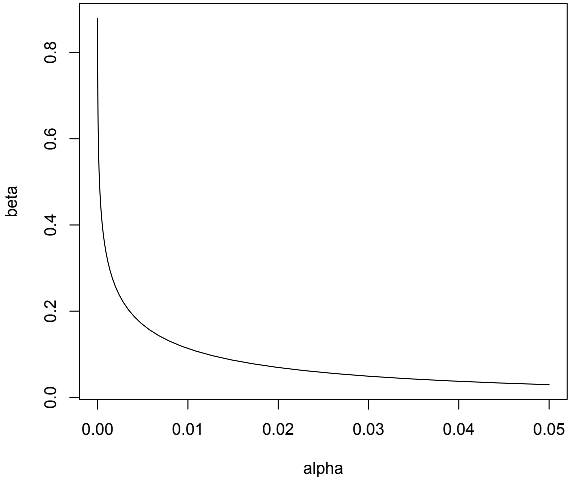
```
z=qnorm(0.95)
zv=z*v
x1=seq(zv,4,0.05)
y1=dnorm(x1,0,v)
polygon(c(x1,zv),c(y1,0), col=2)
```

e ora disegnamo β

```
x2=seq(-1,zv,0.05)
y2=dnorm(x2,3,v)
polygon(c(x2,zv),c(y2,0), col=3)
legend('topleft', c('alpha','beta'), col=c(2,3), pch=19)
```


alpha
beta
dnorm(x, 0, v)
0.0
0.1
0.2
0.3
0.4
0.5
-4
-2
0
2
4
6

```
alpha=1-pnorm(x1,0,v)
beta=pnorm(x1,3,v)
plot(alpha,beta,type='l')
```



Potenza del test γ

Visualizziamo la potenza del test per l'esercizio precedente

```
curve(dnorm(x,0,v),-5,7, axes = TRUE, ylim =c(0,.5))
lines(c(0,0),c(0.45,0),lty=2)
curve(dnorm(x,3,v),-5,7, ylim =c(0,.5), add=T)
lines(c(3,3),c(0.45,0),lty=2)
z=qnorm(0.95)
```

Sappiamo che la potenza del test test γ é la pobabilitá di rifiutare H_0 quando questa é falsa, quindi si ha che

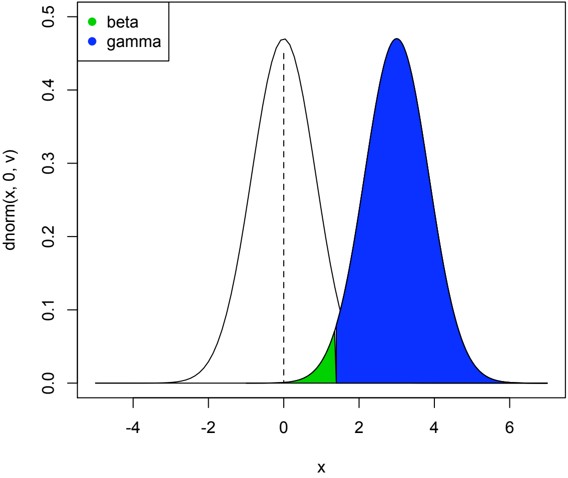
$$\gamma = 1 - \beta$$

disegnamo β

```
x2=seq(-1,zv,0.05)
y2=dnorm(x2,3,v)
polygon(c(x2,zv),c(y2,0), col=3)
```

disegno gamma

```
gamma=pnorm(zv,3,v)
xg=seq(zv,7,0.05)
yg=dnorm(xg,3,v)
polygon(c(xg,zv),c(yg,0), col=4)
legend('topleft', c('beta','gamma'), col=c(3,4), pch=19)
```


beta
gamma
dnorm(x, 0, v)
0.0
0.1
0.2
0.3
0.4
0.5
-4
-2
0
2
4
6
x

Intervalli di confidenza e verifica delle ipotesi

Esercizio 1.

La degenza ospedaliera in giorni per il trattamento di una certa malattia é stata controllata su un campione di 10 pazienti ottenendo i seguenti dati:
5.4, 11.1, 8.2, 7.5, 4.7, 9.4, 10.3, 2.7, 3.3, 7.7.
Fissato un livello di significativitá del 5%:

1. testare sul campione l'ipotesi che la media sia 0 contro l'alternativa che non lo sia e trovare l'intervallo di confidenza per la media μ.
2. testare sul campione l'ipotesi che la media sia 7 contro l'alternativa che non lo sia
3. calcolare la potenza del test se l'ipotesi alternativa é media = 8
4. calcolare la potenza del test se l'ipotesi alternativa é media = 18
5. cosa accade se il campione é di 50 elementi con sd=2.1?

Detto deg il campione

1. t.test(deg, alternative = 'two.sided')
2. t.test(deg, mu = 7, alternative = 'two.sided')
3. power.t.test(n=10,sd=2.9, 0.05, delta=1)
4. power.t.test(n=10,sd=2.9, 0.05, delta=10)
5. power.t.test(50,sd=2.1, 0.05, delta=1)

Esercizio 2.

In uno studio per accertare l'efficacia dell'esercizio fisico per dimagrire sono stati riportati i seguenti risultati, in termini di riduzione di peso in Kg, relativi ad un gruppo di 9 persone a cui era stato assegnato un certo programma di esercizi per un mese:
5.9, 3.2, -0.5, 2.3 , 1.4 , 0.9 , -0.5 , 0 , 3.6
assumendo che la riduzione di peso segua una distribuzione normale

1. si verifichi l'efficacia del programma al livello dell'1$\%$
2. si calcoli la potenza del test con alternativa media=1
3. cosa accade alla potenza del test se la sd é 5?

Imposto il problema come media=0 vs media > 0.
Detto fis il campione

```
t.test(fis, mu = 0, alternative = 'two.sided')
sd(fis)
power.t.test(n=9,sd=2.152, 0.05, delta=1)
power.t.test(n=9,sd=5, 0.05, delta=1)
```

Esercizio 3.

Il ph dell'acqua che fuoriesce da un impianto di depurazione é specificato al livello di 7.0. Un campione di acqua selezionato da un impianto é
5.2, 5.7, 5.8, 6.5, 6.1, 5.9, 6.9, 6.0, 6.7, 6.7
C' é ragione di dubitare circa la specificazione del livello di ph per l'impianto al livello del 5$\%$?

Esercizio 4.

Ad un esame di statistica un campione di 20 studenti che hanno frequentato le esercitazioni riporta i seguenti voti
30, 24, 23, 29, 30, 27, 30, 25, 33, 27, 27 ,30, 27, 24 ,30, 27, 30, 24, 25, 25.
Un campione di chi non ha seguito le esercitazioni ha riportato i seguenti
24, 20, 19, 21, 20, 22, 20, 19, 20, 19, 19, 19, 23, 22, 23, 18 ,19, 17, 22, 20. a) Si verifichi l'ipotesi che la partecipazione alle lezioni non influisce sul voto.
b) Si verifichi l'ipotesi che le due popolazioni abbiano stessa varianza.

I comandi da usare sono

```
t.test()
```

```
var.test
```

Esercizio 5.

Il percorso usuale di una persona da casa al lavoro richiede in media 40 minuti. Un amico suggerisce un altro percorso, sostenendo che é piú breve. La persona prova il nuovo percorso 10 volte coi seguenti tempi (in minuti): 44, 38.5, 37.5, 39, 38.2, 36, 42, 36.5, 36, 34. C' é evidenza sufficiente per confermare che il nuovo percorso é piú veloce, ai livelli 1, 5 e 10 $\%$?

Esercizio 6.

In un campione di 100 misure della temperatura di ebollizione di un certo liquido si é trovata la media campionaria di $\bar{x} = 100$ C, con una varianza campionaria di $S_n^2 = 0.0098C^2$. Supponendo che le osservazioni provengono da una popolazione normale:

1. qual é il livello di significativitá minimo che porta a rifiutare l'ipotesi che la varianza della misura sia inferiore alla soglia di 0.015?
2. Sulla base della risposta del punto precedente, se il livello del test é fissato pari a 0.01 cosa siete portati a decidere?
3. Si rappresenti graficamente la funzione potenza del test di ampiezza $\alpha = 0.01$.

1. Impostiamo il sistema di ipotesi
   $H_0 = \sigma^2 \leq 0.015$
   $H_1 = \sigma^2 > 0.015$
   La regione critica di questo problema di verifica é

   $$D = (\frac{S^2(n-1)}{\sigma^2}) > \chi_{99,(1-\alpha)}$$

   ```
   n=100
   S2=0.0098
   S0=0.015
   d= (n-1)*S2/S0
   a=1-pchisq(d,99)
   ```

2. ```
   alfa=0.01
   c=qchisq( 1-alfa, df=n-1 )
   ```

3. ```
   curve(1-pchisq(c*S0/x,99),0,0.06)
   ```