INTERVALLI DI CONFIDENZA (segue)

Si consideri uno stimatore gaussiano non polarizzato

$$\hat{\vartheta} \sim N(\vartheta^0, \sigma^2_{\hat{\vartheta}})$$

Si è visto (pag. 165) che

$$I_\gamma = [\hat{\vartheta} - z_0 \sigma_{\hat{\vartheta}}, \hat{\vartheta} + z_0 \sigma_{\hat{\vartheta}}].$$

Conoscendo la varianza dello stimatore si ottiene l'intervallo di confidenza.

Per il significato di γ e z_0 si veda la fig. di pag. 165.

A $\gamma = 0{,}95$ corrisponde $z_0 = 1{,}96$, per cui

$$I_{0,95} = [\hat{\vartheta} - 1{,}96\,\sigma_{\hat{\vartheta}}, \hat{\vartheta} + 1{,}96\,\sigma_{\hat{\vartheta}}]$$

Sul grafico della d.d.p. della Gaussiana standard la larghezza di banda che realizza un'area pari a 0,95 è pari a (circa)

$$\pm 2\sigma_{\hat{\vartheta}}$$

Se si considera il caso della media campionaria

$$M_1 \sim N\left(\mu, \frac{\sigma^2}{n}\right),$$ con σ^2 nota ed n grande.

si ottiene:

$$I_\gamma = \left[M_1 - z_0 \frac{\sigma}{\sqrt{n}}, M_1 + z_0 \frac{\sigma}{\sqrt{n}}\right]$$

Si è considerato n "grande" per poter applicare il teorema centrale del limite.

Esercizio 1 Si consideri il seguente caso:

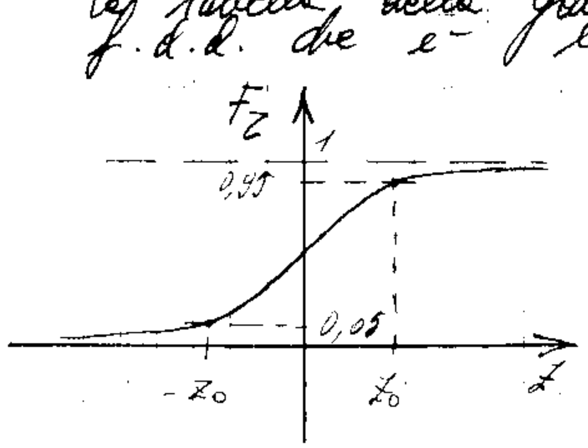
$n = 100$, $\sigma = 1$, $\gamma = 0{,}9$

Si determini l'intervallo di confidenza

Per calcolare I_γ serve la conoscenza di z_0, essendo

$$I_{0,9} = \left[M_1 - z_0 \frac{1}{\sqrt{100}} \;,\; M_1 + z_0 \frac{1}{\sqrt{100}}\right]$$

La tabella della Gaussiana standard opera sulla f.d.d. che è l'integrale della d.d.p.



Se $\gamma = 0,9$, si ha:

$$\frac{1-\gamma}{2} = \frac{1-0,9}{2} = 0,05$$

$$\int_{-\infty}^{-z_0} F_Z(z)\,dz = 0,05$$

$$\int_{-\infty}^{z_0} F_Z(z)\,dz = 0,95$$

Occorre cercare sulle tabelle il valore di z_0, tale che

$$F_Z(z_0) = 0,95$$

La tabella contiene i valori di $F_Z(z)$, considerando solo i valori positivi di z, a partire cioè da $F_Z = 0,5$.
I valori di z (positivi) si leggono sulle intestazioni di riga e colonna.
Il primo valore della tabella è 5000 che significa

$$F_Z(z) = 0,5000$$

In corrispondenza si trovano i valori

0,0+ (sulle intestazioni di colonna)

a cui si deve aggiungere il valore

0,00 (sulle intestazioni delle righe)

ottenendo: $z_0 = 0$

Ricordiamo che si sta considerando una Gaussiana standard ed il punto centrale ($z = 0$) divide la d.d.p. in due parti equivalenti di area 0,5.
Il valore di 0,95 si trova nella quarta colonna ed è situato tra 9495 (cui corrisponde $z = 1,5 + 0,14$) e 9505 ($z = 1,5 + 0,15$).
Interpolando si ottiene: $z_0 = 1,5 + 0,145 = 1,645$
Risulta:

$$I_{0,9} = \left[M_1 - 1,645 \frac{1}{\sqrt{100}} \;,\; M_1 + 1,645 \frac{1}{\sqrt{100}}\right] =$$

$$= [M_1 - 0{,}1645 \ , \ M_1 + 0{,}1645]$$

Esercizio 2 Si conoscano

$$n = 100 \qquad \sigma = 1$$

$$I_\gamma = [M_1 - 0{,}1 \ , \ M_1 + 0{,}1]$$

Determinare γ

I_γ si può scrivere

$$I_\gamma = [M_1 - \delta \ , \ M_1 + \delta]$$

con

$$\delta = z_0 \frac{\sigma}{\sqrt{n}} = 0{,}1$$

Si ricava

$$z_0 = \frac{0{,}1\sqrt{n}}{\sigma} = \frac{1}{10}\frac{\sqrt{100}}{1} = 1$$

Conoscendo z_0 si deve calcolare γ

Dalle tavole, a $z_0 = 1 = 1{,}0 + 0{,}00$ corrisponde

$$F_Z(z_0) = 0{,}8413 = \frac{1-\gamma}{2} + \gamma = \frac{1+\gamma}{2}$$

da cui

$$\gamma = 2 \times 0{,}8413 - 1 = \underline{0{,}6826}$$

Esercizio 3 Si conoscano:

$$\gamma = 0{,}95 \qquad \sigma = 1$$

$$I_\gamma = [M_1 - 0{,}1 \ , \ M_1 + 0{,}1]$$

Determinare n

A $\gamma = 0{,}95$ corrisponde $z_0 = 1{,}96$ (da ricordare a memoria)

$$\delta = z_0 \frac{\sigma}{\sqrt{n}} = 0{,}1 \qquad \sqrt{n} = \frac{z_0 \cdot 1}{0{,}1} = 10\, z_0$$

$$n = 100\, z_0^2 = 100 \times (1{,}96)^2 = 384{,}16 = \underline{385}$$

L'arrotondamento va fatto all'intero superiore.

170 Se si desidera che nel 95% dei casi l'errore sia inferiore a 0,1 occorre raccogliere almeno 385 dati.

Per avere un δ 10 volte più piccolo del precedente caso

$$\delta = 0{,}01$$

quanti dati occorre raccogliere?

$$\delta = z_0 \frac{\sigma}{\sqrt{n}} = 0{,}01$$

$$\sqrt{n} = 1{,}96 \frac{1}{0{,}01} = 196$$

$$n = 196^2 = 38416$$

<u>Osservazione</u>: Se si vuole ottenere un δ 10 volte inferiore occorre moltiplicare per 100 il numero di dati raccolti.

n risulta inversamente proporzionale al quadrato di

$$n \equiv \frac{1}{\delta^2}$$

• <u>V.C. t di Student</u>

È una v.c. continua che deve il suo nome allo pseudonimo Student usato da William Sealy Gosset (Canterbury, 1876 - Londra 1937 - statistico inglese) che ideò il omonimo test, mentre la v.c. stessa venne identificata da Ronald Fisher (Wikipedia).

Essa è così definita:

$$t_n = \frac{z_0}{\sqrt{\sum_{i=1}^{n} z_i^2}} \cdot \sqrt{n} = \frac{z_0}{\chi_n} \sqrt{n},$$

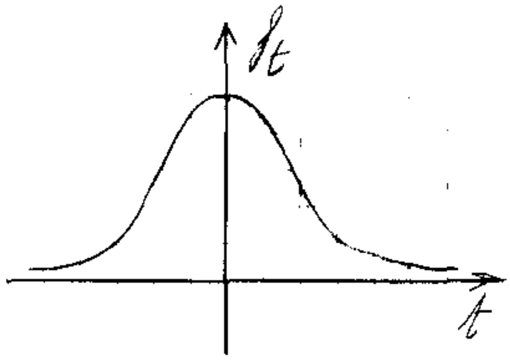
essendo

$$z_i \quad \text{con } i = 0, 1, \dots, n \quad \text{i.i.d.}$$

e gaussiane standard.

La v.c. t_n è detta "<u>t di Student</u>" con n gradi di libertà.

È simile ad una Gaussiana standard ma più panciuta.



Proprietà:

- è simmetrica rispetto all'origine
- $E[t_n]=0$
- $Var[t_n] = \frac{n}{n-2}$
- per $n \to \infty$, t_n tende ad una Gaussiana standard, in distribuzione.

Si noti che la varianza è maggiore di 1, ma per $n \to \infty$, essa tende ad 1.

- <u>Intervallo di confidenza per la media con varianza sconosciuta</u>

Si considerino delle v.c. $\sim N(\mu, \sigma^2)$, distribuite gaussianamente, indipendenti e con σ^2 sconosciuta.

Proviamo a standardizzare, mediante standardizzazione nel seguente modo, ottenendo la v.c.

$$\frac{M_1 - \mu}{S/\sqrt{n}}$$

Al denominatore dovrebbe comparire $\sigma/\sqrt{n}$, ma non conoscendo la varianza si è usata S, con S^2 la varianza campionaria definita come a pag. 158

$$S^2 := \frac{n}{n-1} d_2$$

Ricordiamo anche che

$$d_2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - M_1)^2 \quad \text{(pag. 155)}$$

e che

$$\frac{n d_2}{\sigma^2} = \chi^2_{n-1} \quad \text{(pag. 156)}$$

Standardizzando otteniamo

$$\frac{M_1 - \mu}{\sqrt{\frac{S^2}{n}}} = \frac{M_1 - \mu}{\sqrt{\frac{d_2}{n-1}}} = \frac{(M_1 - \mu)\sqrt{\frac{n}{\sigma^2}}}{\sqrt{\frac{d_2}{n-1} \frac{n}{\sigma^2}}} =$$

$$= \frac{\dfrac{M_1 - m}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\dfrac{S_2^2\, n}{\sigma^2}}}\sqrt{n-1} = \frac{Z}{\sqrt{\chi^2_{n-1}}}\sqrt{n-1} =$$

$$= t_{n-1}$$

Al numeratore compare

$$\frac{M_1 - m}{\sigma/\sqrt{n}}$$

che è la standardizzazione vera.

Ciò che si è ottenuto è una t di Student ad $n-1$ gradi di libertà.

Concludendo, si può osservare che standardizzando conoscendo σ si ottiene una Gaussiana standard.

Pseudostandardizzando, usando S anziché σ, la variabile standardizzata che si ottiene è una t di Student.

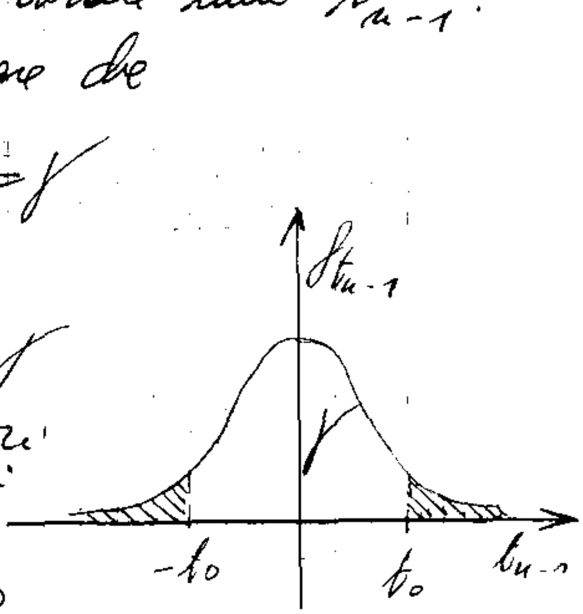
Anziché una $N(0,1)$ si ottiene una t_{n-1}.

Conseguenza: si può affermare che

$$P\left(\left|\frac{M_1 - m}{S/\sqrt{n}}\right| \le t_0\right) = \gamma$$

essendo t_0 tale che

$$P\left(|t_{n-1}| \le t_0\right) = \gamma$$



In altre parole, tra i valori $-t_0$ e t_0 cade un'area pari a γ e la probabilità che la variabile standardizzata nel modo esposto sia, in modulo, minore o eguale a t_0 è proprio γ.

La relazione ottenuta si può scrivere anche così

$$P\left(|M_1 - m| \le t_0 \frac{S}{\sqrt{n}}\right) = \gamma$$

La quantità $t_0 S/\sqrt{n}$ è δ (vedi pag. 164).
Si ottiene
$$I_f = \left[M_1 - t_0 \frac{S}{\sqrt{n}} \ , \ M_1 + t_0 \frac{S}{\sqrt{n}} \right]$$
Con σ^2 noto si è ottenuto
$$\delta = z_0 \frac{\sigma}{\sqrt{n}}$$
Con σ^2 sconosciuto, invece
$$\delta = t_0 \frac{S}{\sqrt{n}}$$
I valori di t_0 si trovano nella tabella della t di Student.

Osservazione: si è ottenuta una t di Student ad $n-1$ gradi di libertà, avendo operato con n dati ed avendo eseguito una pseudo-standardizzazione.

Si noti che nella tabella e nella colonna 0,05 che corrisponde a $\gamma = 0,95$, se ν gradi di libertà tendono all'infinito si ottiene una t_0 che tende a 1,96.

Ovviamente, noto t_0, si calcola I_f.

114 LEZIONE (10) 25/05/09 parte 2ª

MATLAB - (1)

Si consiglia di scaricare appunti dispense ed esercitazioni relative all'uso del programma MATLAB dal sito
http://risdip.unifo.it/lab/didattica/corsi/imead_triennale.html

- help <comando>
fornisce informazioni sul <comando>
Es.: help diary
diary è un file di diario in cui viene registrato tutto ciò che è svolto durante una sessione

- diary on
attiva la funzione diary

Matlab ha la matrice come elemento fondamentale ed usa fondamentalmente la linea di comando.

- inserimento di una matrice
A = [1 2 3
4 5 6]   oppure
A = [1 2 3; 4 5 6]
La matrice A è in memoria. Il comando A mostra la matrice

- il segno di % permette di inserire commenti
% ciao è ininfluente sulle operazioni

- inserimento di una matrice
>> [5 6 7
8 9 10]
Matlab fornisce la risposta con ans
ans =

5 6 7
8 9 10

```
>> [1 2 3 ...
7 8 9]
ans =
     1 2 3 7 8 9
```

... è un segno di continuazione ed è equivalente a non andare a capo.

>> [1 2 3 7 8 9] è un comando equivalente

>> x = 1:10

Inserisce un vettore riga con i numeri da 1 a 10 spaziati di una unità

```
x =
   1 2 3 4 5 6 7 8 9 10
```

>> x = 1:0.5:10

Come il comando precedente ma con passo 0,5.

-) <u>inserimento di una matrice di zeri</u>

I seguenti comandi sono equivalenti

>> Z = [0 0 0; 0 0 0]

>> Z = zeros(3,2)

Viene costruita una matrice di 3 righe e 2 colonne con elementi tutti nulli.

-) <u>mostrare i comandi precedenti</u>

>> ↑ (più volte)

-) <u>inserimento di matrice quadrata</u> di zeri

>> Z = zeros(2)

```
Z =
   0 0
   0 0
```

-) <u>inserimento di una matrice di uni</u> e <u>identità</u>

>> U = ones(3,4)   (matrice 3 × 4) (di uni)

176 >> I = eye(5) matrice identità di ordine 5

-) <u>estrazione di parti di matrice</u>

>> I(:, 2) estrae la seconda colonna della matrice I appena costruita

ans =
0
1
0
0
0

Il simbolo ":" indica che si prendono tutti gli elementi. Nell'esempio si acquisiscono tutte le righe e la colonna 2.

>> I(2, :) estrae la seconda riga e tutte le colonne

ans =
0 1 0 0 0

>> I(3, 3) estrae l'elemento di I nella riga 3 e colonna 3

ans =
1

-) <u>visualizzare il contenuto di un elemento</u>

>> A

A =
1 2 3
4 5 6

>> Y = [7 8 9] definire il vettore Y

-) <u>composizione di matrici</u>

>> B = [A ; Y] Y è incollato sotto la A. Il ";" indica "sotto".

B =
1 2 3
4 5 6
7 8 9

N.B.: A e Y devono avere lo stesso numero di colonne.

-) trasposizione di una matrice

```
>> A'     traspone A
ans =
     1  4
     2  5
     3  6
```

`>> x(:)` prende tutti gli elementi del vettore x (pag. 175) e li dispone in colonna.

-) operazioni tra matrici

`>> [5 6 7; 1 2 3] + [3 3 3; 4 4 4]`

il comando somma le due matrici

N.B. le matrici devono avere dimensioni appropriate

`>> [1 0 2] - 5` toglie 5 a tutti gli elementi del vettore

N.B. il comando non da errore: 5 è interpretato come un vettore.

```
>> A = [1 2; 3 4; 5 6]
>> B = [4 4; 5 5]
```

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} * \begin{bmatrix} 4 & 4 \\ 5 & 5 \end{bmatrix} = \begin{bmatrix} 14 & 14 \\ 32 & 32 \\ 50 & 50 \end{bmatrix}$$

```
>> A * B
ans =
     14  14
     32  32
     50  50
```

-) operazioni su vettori

`>> v = [1 : 6]` costruisce un vettore con numeri da 1 a 6

`>> w = [7 : 12]`

`>> v * w` il comando produce errore per le errate dimensioni. Occorre trasporre w.

```
>> v * w'
ans =
     217
```

`>> v .* w` l'operazione viene eseguita "elemento per elemento" se si usa il ".".

ans =
7 16 27 40 55 72

v e w sono delle stesse dimensioni. L'elemento i-esimo di v è moltiplicato per il corrispondente elemento di w.

>> A2 = A.^2 Il punto può essere usato anche con le matrici.

Ogni elemento di A è elevato al quadrato e collocato nella matrice A2.

A è definita a pag. 127

A2 =
1 4
9 16
25 36

>> v./w (vedi pag. 127) Ogni elemento di v è diviso per il corrispondente elemento di w

ans =
0,1428 0,2500 0,3333 0,4000 0,4545 0,5000
↓ ↓ ↓ ↓ ↓ ↓
1/7 2/8 3/9 4/10 5/11 6/12

>> A = [1 2 0; 0 4 2; 2 2 8]

A =
1 2 0
0 4 2
2 2 8

>> b = ones(3, 1)

b =
1
1
1

>> % calcoliamo la soluzione del sistema di equazioni lineari $Ax = b$

>> % la soluzione è data da $x = A^{-1}b =$
$= inv(A) * b$

affinché esista l'inversa di una matrice, occorre che il suo determinante sia diverso da zero.

>> det (A) calcola il determinante di A

```
>> ans =
   36
>> x = inv(A)*b
   x =
      0.4444
      0.2778
     -0.0556
```

L'operazione richiede il calcolo della matrice inversa ed un prodotto tra matrici. L'operazione può essere molto laboriosa.

È meglio utilizzare il seguente comando che fornisce lo stesso risultato

`>> x = A\b`   Il comando è da interpretare come "b diviso A"

Quando A una matrice così equivale a calcolare il prodotto di b per l'inversa di A.

Abbiamo utilizzato la codoletta inversa sinistra.

Non si può utilizzare la inversa destra date le dimensioni di A e b.

`>> rank(A)`   calcola il rango di A

```
   ans =
      3
```

`>> eig(A)`   calcola gli autovalori di A

```
   ans =
      2
      2
      9
```

N.B: in inglese autovalori si dice eigenvalues

`>> y=[10 -10 20 -30]`

`>> y>0`   operatore logico: è calcolato un vettore binario che ha degli uni tutte le volte che la condizione è soddisfatta

```
   ans =
      1   0   1   0
```

`>> y(y>0)`   del vettore y sono considerati solo i valori positivi.

```
   ans =
      10   20
```

$y>0$ fornisce un valore logico (e.o. zero). Il comando può servire per eliminare i valori negativi.

180 >> %h (h > 170) del vettore h sono conservati solo i valori > di 170.

>> find (f > 0) individua gli elementi di f che sono positivi, fornendo la loro posizione
ans
1 3

>> f (find (f > 0)) come il comando f(f > 0)
ans
10 20

>> a = 6 ad a è assegnato il valore 6

>> c = 11

>> a == 6 == è un operatore logico che risponde vero se a = 6 e falso se a ≠ vero
ans =
1

>> a == 7 la risposta è: falso.
ans =
0

>> a == 6 & c == 11 & è l'AND logico
ans =
1

>> a == 6 | c == 11 | è l'OR logico
ans =
1

>> a ~= 5 ~= significa NON UGUALE
ans =
1

>> a ~= 6
ans =
0

-) operatori sulla memoria

>> who mostra le variabili nella memoria

>> whos come who ma con informazioni sulle dimensioni, sui byte occupati e sulle classe

>> clear A2 B C elimina dalla memoria le variabili indicate

>> clear all   libera tutta la memoria dalle variabili che la occupano

>> x = -20 : .1 : 20;   il punto e virgola finale elimina l'echo a video e non mostra il risultato

>> y1 = x;

>> y2 = x.^2;

>> y3 = x.^3;

>> y4 = x.^4;

-) realizzazione di grafici

>> plot (x, y2)   disegna il grafico di y2 che risulta una parabola

>> plot (x, y2, '*')   come il comando precedente, ma appaiono degli asterischi dei punti

>> plot (x, y2, 'r')   il grafico appare rosso

>> plot (x, y2, 'r*')   appaiono degli asterischi rossi

>> figure (100)   etichetta la figura attiva

>> plot (x, y3, 'g')   disegna in verde il grafico di y3.

>> xlabel ('Asse X')   etichetta gli assi

ylabel ('Asse Y')

>> title ('x^3 - cubica')   mette un titolo

N.B.: quando si utilizzano stringhe è necessario racchiuderle tra apici.

>> subplot (2, 2, 2)   permette di realizzare più grafici nella stessa figura.

Il comando indica di costruire una griglia di grafici di 2 righe e 2 colonne e poi occupare la seconda cella con il seguente comando

>> plot (x, y1)

I seguenti comandi disegnano nella figura attiva quattro grafici

182 >> figure (100)

>> subplot (2,2,1), plot (x, y1)

>> subplot (2,2,2), plot (x, y2)

>> subplot (2,2,3), plot (x, y3)

>> subplot (2,2,4), plot (x, y4)

N.B.: per collocare più istruzioni sulla stessa riga si usa la virgola.

I grafici possono essere ingranditi o rimpiccioliti.

Quando si disegna un grafico nella figura attiva viene sostituito il grafico precedente.

Per sovrapporre più grafici nella stessa figura il comando seguente permette di mantenere anche il grafico precedente

>> figure (10)

>> plot (x, y2)

>> hold on   non cancella l'ultimo grafico

>> plot (x, y1)

>> hold off   annulla il precedente comando

>> grid   crea una griglia nel grafico

-) Operazioni su numeri casuali

>> xg = randn (10, 1);   produce un vettore di 10 righe ed 1 colonna di Gaussiane standard indipendenti tra loro.

I numeri generati sono pseudocasuali. Per renderli più "casuali" occorre modificare il valore dell'opzione seed, cioè del seme che genera i numeri.

Si può memorizzare il seme per riprodurre la serie di dati.

>> xg = randn (1000, 1);

>> mean (xg)   calcola la media

```
ans =
    0,0540
```

La media non è nulla, essendo una media campionaria.

```
>> median (xg)          calcola la mediana
ans =
    0,0596
>> var (xg)             calcola varianza campionaria
ans =
    1,0801
>> std (xg)             calcola la deviazione standard
ans =
    1,0393
>> xu = rand (1000, 2);   genera una serie di 1000 numeri casuali distribuiti su due colonne.
>> mean (xu)            crea la media delle due colonne (media campionaria)
ans =
    0,5021    0,4993
>> xg = randn (10, 2);
>> mean (xg)            calcola la media delle due colonne
>> mean (xg, 2)         calcola la media ~~della seconda colonna~~ lungo la seconda di misurazione (la media è calcolata riga per riga)
                        In pratica vengono creati 10 valori.
>> cov (xg)             stima la matrice di covarianza
ans =                   In teoria, le variabili dovrebbero essere indipendenti e dare una matrice così fatta
    0,8081    0,2237        [1  0]
    0,2237    0,9244        [0  1]

>> xg = randn (1000, 2);
>> cov (xg)
```

ans =
1,0263 -0,0085
-0,0085 0,9941

I risultati si avvicinano di più ai valori teorici poiché le serie sono più numerose.

>> var (xg)

ans =
1,0263 0,9941

>> corrcoef (xg)

calcola il coefficiente di correlazione

ans =
1,0000 -0,0084
-0,0084 1,0000

>> xg = rand (1000, 1)

>> close all

chiude tutte le figure

>> figure (1)

>> hist (xg, 50)

crea l'istogramma sui dati xg e 50 bins

>> a = -10 : 10

>> h = a + randn (size (a))

size (a) fornisce la dimensione di a, in questo caso 1, 21 (1 riga e 21 colonne)

il quando costruisce un vettore di dimensioni casuali con le stesse dimensioni del vettore a.

>> scatter (a, h)

crea lo scatter plot di a e h

h non è uguale ad a poiché è presente un addendo casuale, se non ci fosse i dati di a e h starebbero su una retta.

- gestione dei file di dati

>> save prova.mat xg xn

del file prova.mat vengono salvati i vettori xg e xn

>> clear

viene cancellata la memoria di lavoro

>> load prova

carica il file prova con i dati memorizzati.

-) uso dei cicli con condizioni

```
>> s=0
>> for i=1:10, s=s+i, end
>> if s<60, b=100, end
b =
    100
>> if s>60, b=100, else b=-100, end
b =
    -100
```