# MatLab in Biostatistica

Centro E. Piaggio
bioengineering and robotics research center

IN SUPREMÆ DIGNITATIS
1343

Ing. Vincenzo Catrambone
vincenzo.catrambone@ing.unipi.it

# Generatore di numeri casuali

Y = random(name,A,B,C,D,[m,n,...]) returns an m-by-n-by... matrix of random numbers for distributions that require four parameters.

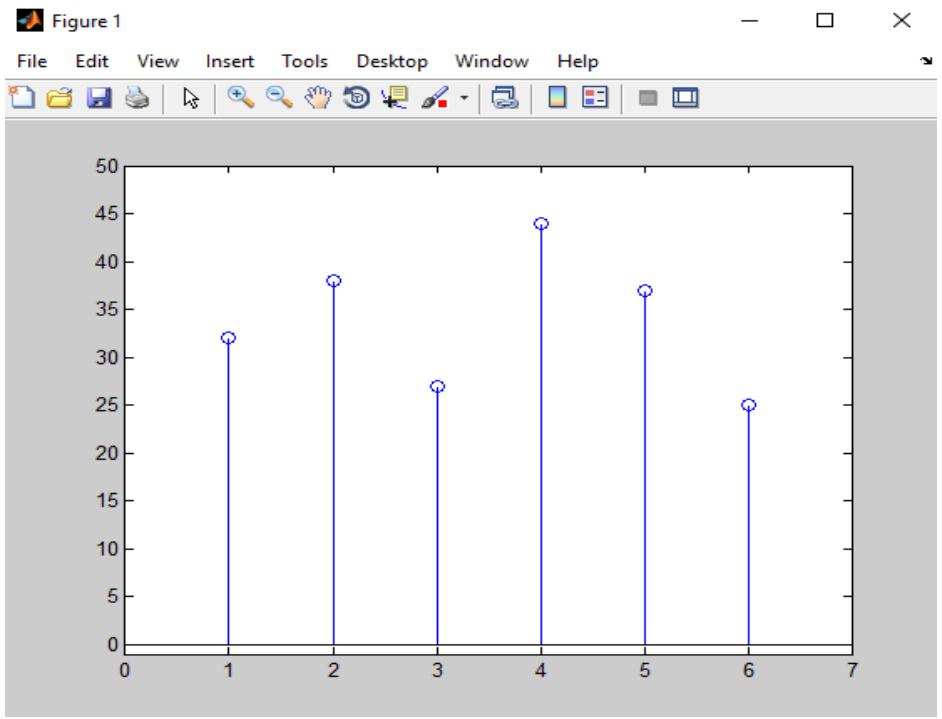
| name | Distribution | A | B |
|---|---|---|---|
| 'bino' or 'Binomial' | Binomial Distribution | n: number of trials | p: probability of success for each trial |
| 'chi2' or 'Chisquare' | Chi-Square Distribution | $\nu$: degrees of freedom | — |
| 'exp' or 'Exponential' | Exponential Distribution | $\mu$: mean | — |
| 'gam' or 'Gamma' | Gamma Distribution | a: shape parameter | b: scale parameter |
| 'logn' or 'Lognormal' | Log-normal Distribution | $\mu$: mean | $\sigma$: standard deviation |
| 'norm' or 'Normal' | Normal Distribution | $\mu$: mean | $\sigma$: standard deviation |
| 'poiss' or 'Poisson' | Poisson Distribution | $\lambda$: mean | — |
| 't' or 'T' | Student's Distribution | $\nu$: degrees of freedom | — |

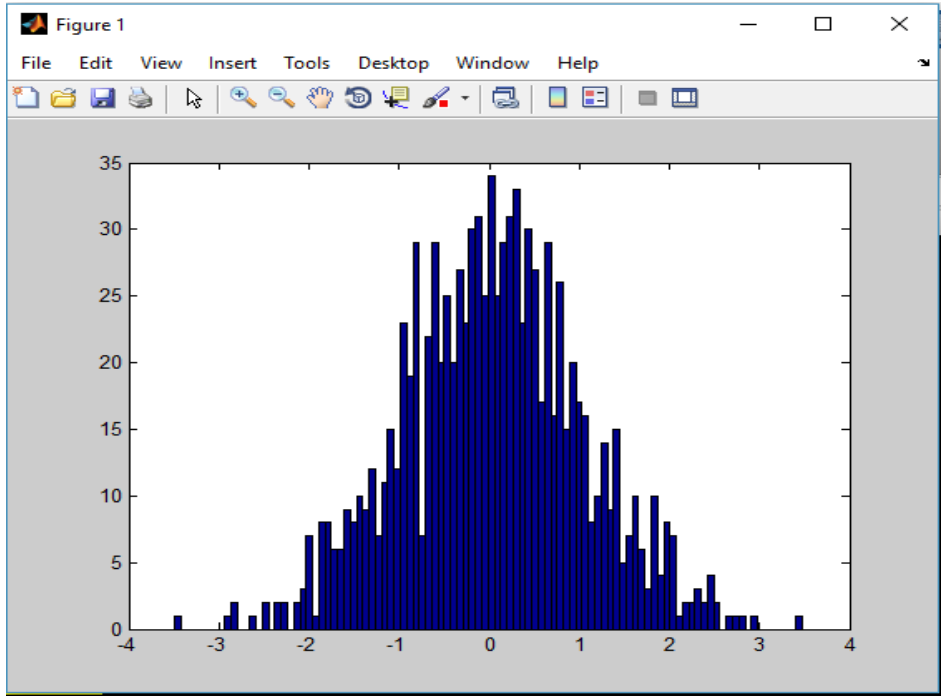
Y = random('norm',0,1,[m,n]) = randn(m,n)

# Rappresentazione grafica dei campioni

Campione proveniente da una distribuzione continua:

Istogramma: hist(x)



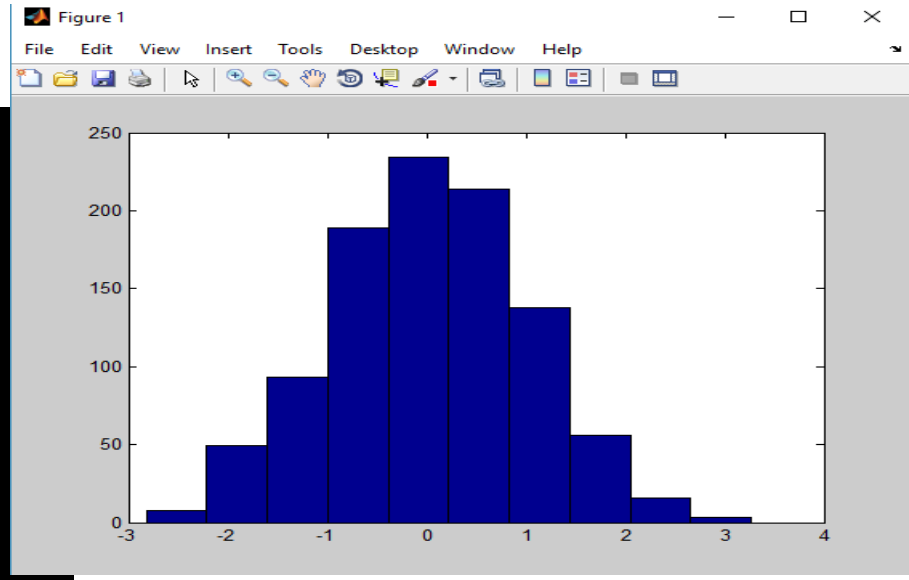

Campione appartenente a una distribuzione discreta:

Diagramma a bastoncini: stem(x)

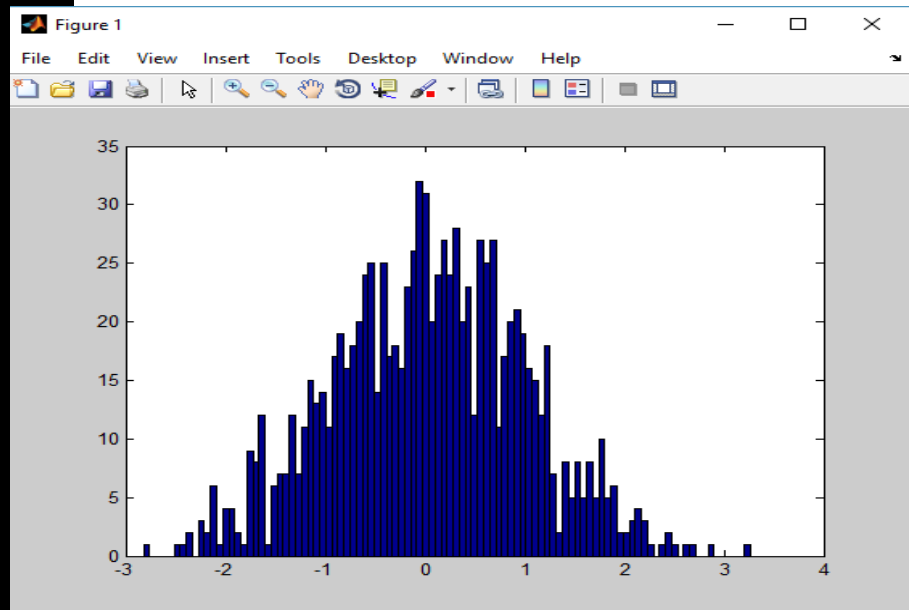
# Rappresentazione grafica (2)

```
>> help hist
 hist  Histogram.
    N = hist(Y) bins the elements of Y into 10 equally spaced containers
    and returns the number of elements in each container.  If Y is a
    matrix, hist works down the columns.

    N = hist(Y,M), where M is a scalar, uses M bins.

    N = hist(Y,X), where X is a vector, returns the distribution of Y
    among bins with centers specified by X. The first bin includes
    data between -inf and the first center and the last bin
    includes data between the last bin and inf. Note: Use HISTC if
    it is more natural to specify bin edges instead.

    [N,X] = hist(...) also returns the position of the bin centers in X.
```

hist(x)

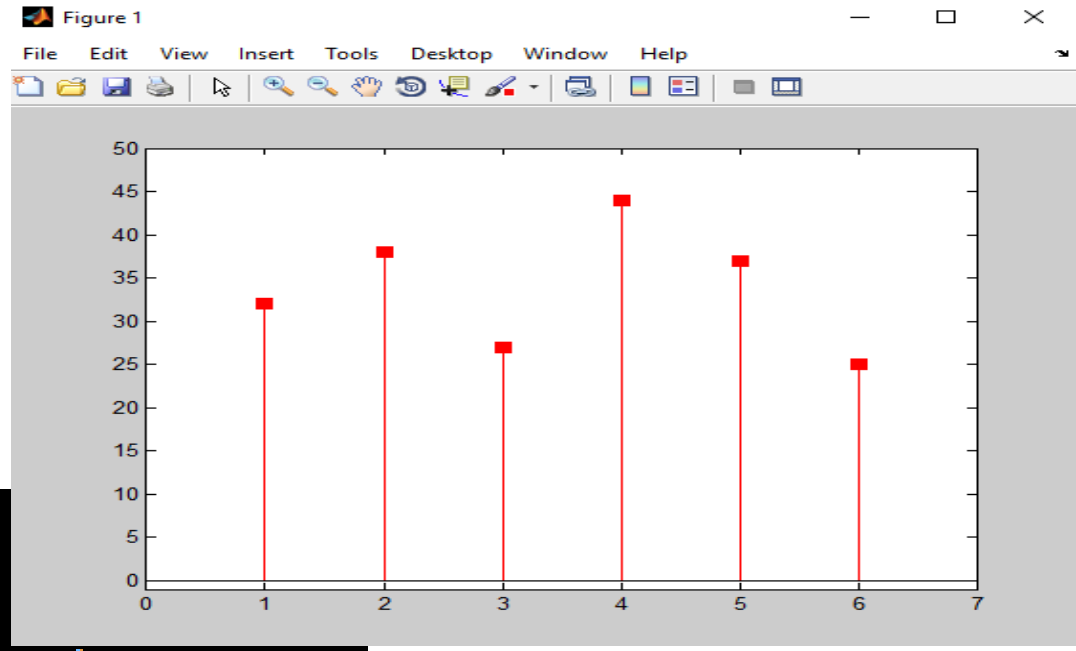



hist(x,100)



# Rappresentazione grafica (3)

stem(y,'filled','sr')

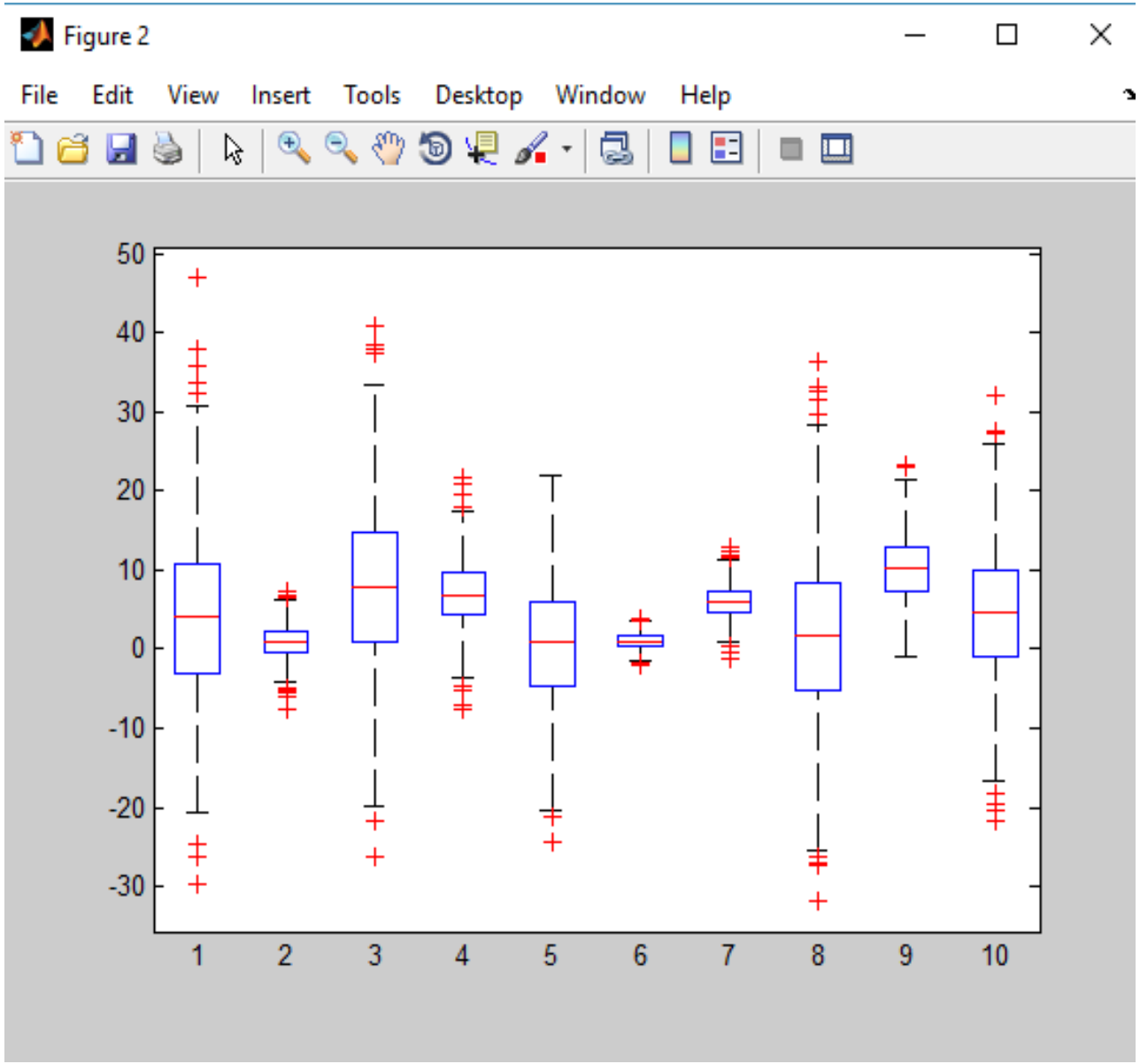



```
>> help stem
 stem   Discrete sequence or "stem" plot.
    stem(Y) plots the data sequence Y as stems from the x axis
    terminated with circles for the data value. If Y is a matrix then
    each column is plotted as a separate series.

    stem(X,Y) plots the data sequence Y at the values specified
    in X.

    stem(...,'filled') produces a stem plot with filled markers.

    stem(...,'LINESPEC') uses the linetype specified for the stems and
    markers.  See PLOT for possibilities.
```

# Rappresentazione grafica (4)

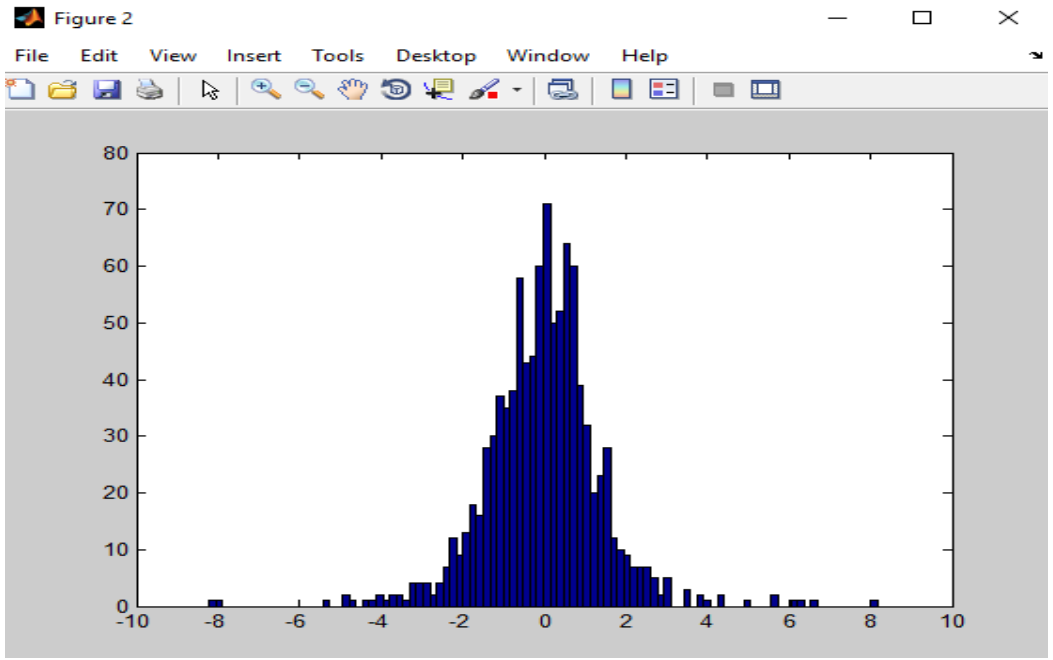
>> help boxplot

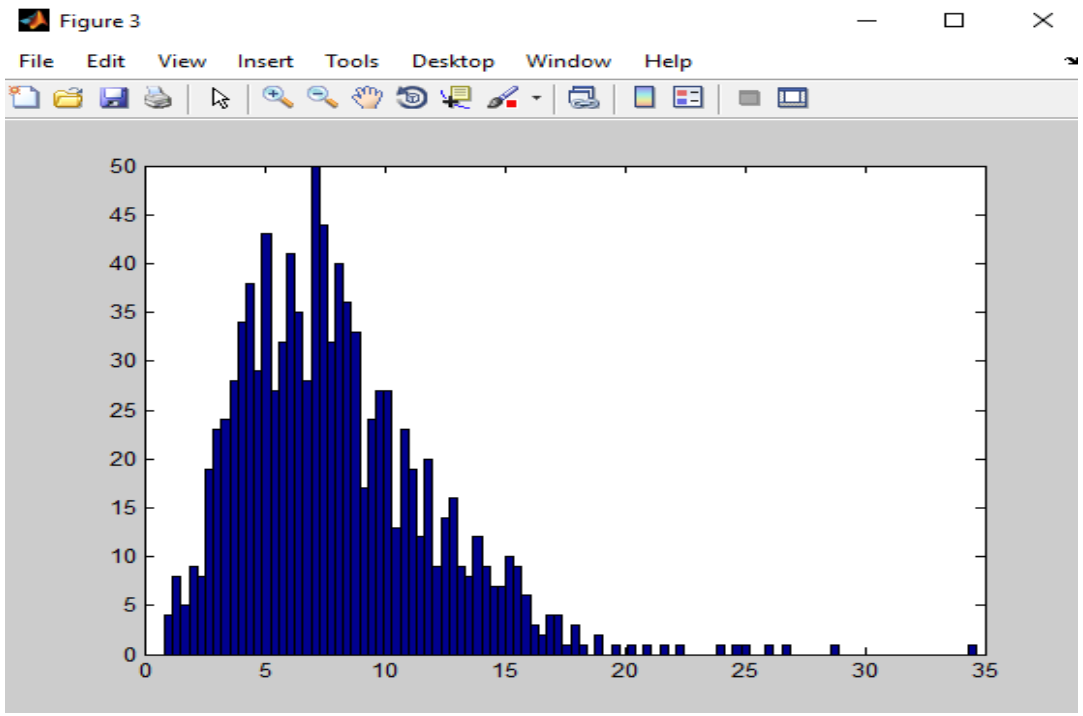
boxplot Displays box plots of multiple data samples. boxplot(X) produces a box plot of the data in X. If X is a matrix there is one box per column, and if X is a vector there is just one box. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme datapoints the algorithm considers to be not outliers, and the outliers are plotted individually.



# Generatore di numeri casuali(2)

hist(random('T',4,1000,1),100)



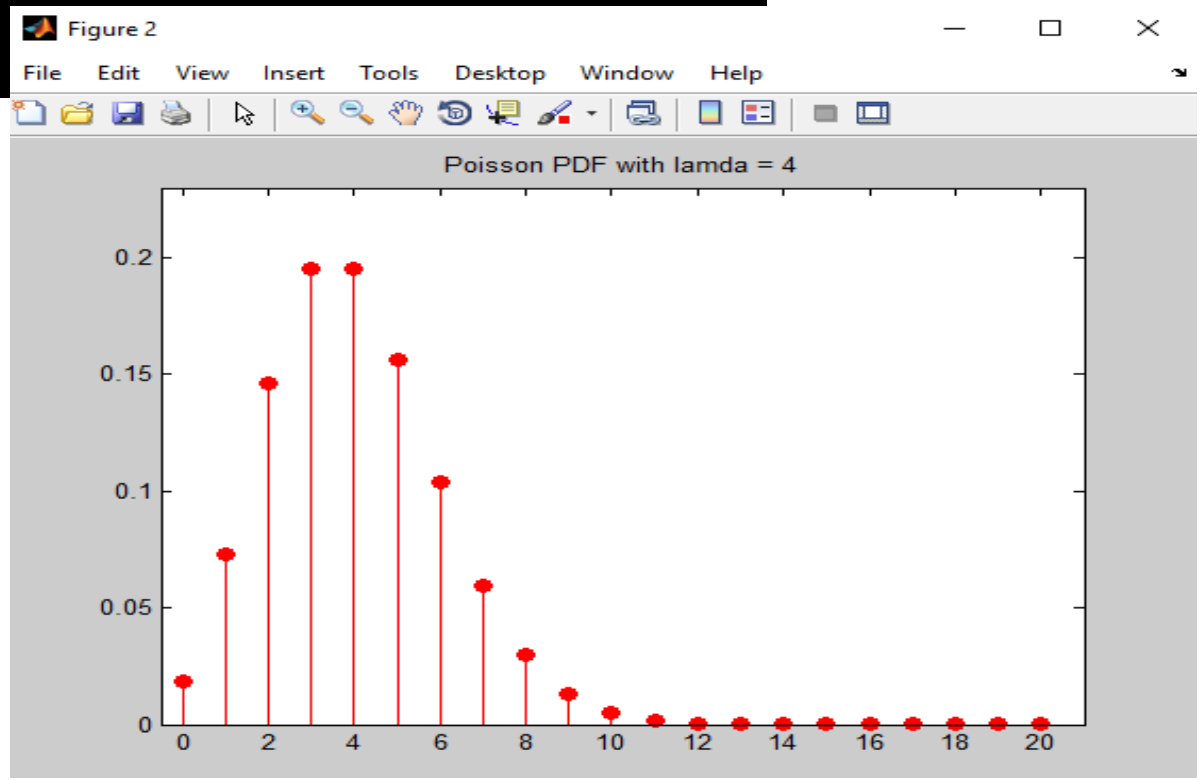

hist(random('gam',4,2,1000,1),100)

# Funzioni Densità di Probabilità

```
>> help pdf
 pdf Density function for a specified distribution.
    Y = pdf(NAME,X,A) returns an array of values of the probability density
    function for the one-parameter probability distribution specified by NAME
    with parameter values A, evaluated at the values in X.

    Y = pdf(NAME,X,A,B) or Y = pdf(NAME,X,A,B,C) returns values of the
    probability density function for a two- or three-parameter probability
    distribution with parameter values A, B (and C)
```

'Beta', 'Binomial', 'Chisquare', 'Exponential', 'F', 'Gamma', 'Geometric', 'Hypergeometric', 'Lognormal', 'Normal', 'Poisson', 'Rayleigh', 'T'.

```
x = 0:20;
y = pdf('Poisson',x,4);
figure;    stem(x,y,'filled','r');
ylim([0 0.23]);   xlim([-0.5 21]);
title('Poisson PDF with lamda = 4')
```



# Statistica descrittiva: indici di posizione

```
>> help mean
 mean   Average or mean value.
    For vectors, mean(X) is the mean value of the elements in X. For
    matrices, mean(X) is a row vector containing the mean value of
    each column.  For N-D arrays, mean(X) is the mean value of the
    elements along the first non-singleton dimension of X.

    mean(X,DIM) takes the mean along the dimension DIM of X.
```

median(x)
mode(x)

```
>> help quantile
 quantile Quantiles of a sample.
    Y = quantile(X,P) returns quantiles of the values in X.  P is a scalar
    or a vector of cumulative probability values.  When X is a vector, Y is
    the same size as P, and Y(i) contains the P(i)-th quantile.  When X is
```

# Statistica descrittiva: indici di dispersione e forma

var(x) std(x) range(x) = max(x) – min(x) iqr(x)

kurtosis(x) skewness(x) moment(x,order)

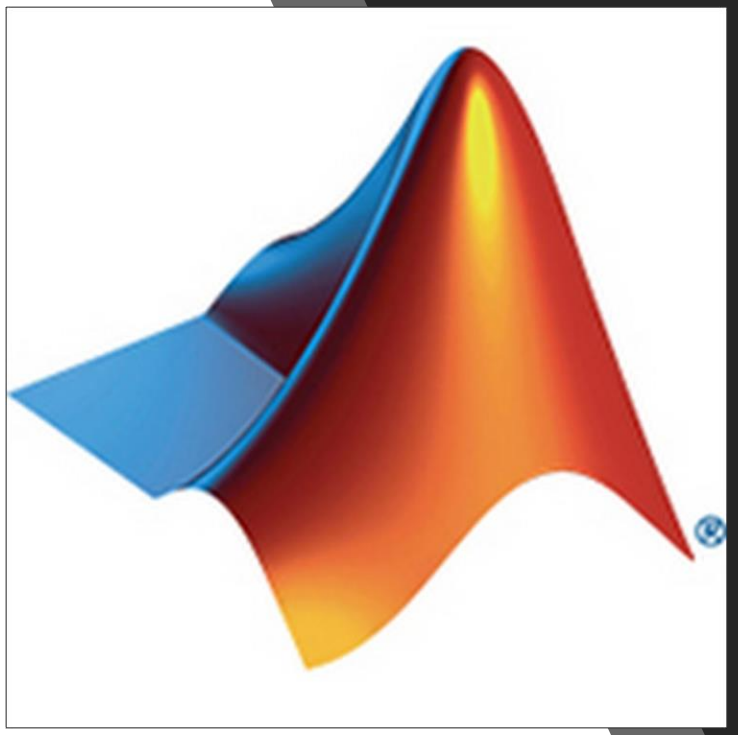
```
>> help moment
  moment Central moments of all orders.
    SIGMA = moment(X,ORDER) returns the ORDER-th central sample moment of
    the values in X.  For vector input, SIGMA is MEAN((X-MEAN(X)).^ORDER).
    For a matrix input, moment(X,ORDER) returns a row vector containing the
    central moment of each column of X.  For N-D arrays, moment operates
    along the first non-singleton dimension.

    moment(X,ORDER,DIM) takes the moment along dimension DIM of X.

    The first central moment is exactly zero. The second central moment is
    the variance, using a divisor of N instead of N-1, where N is the
    sample size.
```

# Statistica descrittiva: indici di frequenza

```
freq_ass(k) = sum(X == k);
freq_rel(k) = sum(X == k)/length(X);
freq_cum_ass(k) = cumsum(freq_ass);
freq_cum_rel(k) = cumsum(freq_rel);
```

```
>> help cumsum
 cumsum Cumulative sum of elements.
    For vectors, cumsum(X) is a vector containing the cumulative sum of
    the elements of X. For matrices, cumsum(X) is a matrix the same size
    as X containing the cumulative sums over each column.  For N-D
    arrays, cumsum(X) operates along the first non-singleton dimension.

    cumsum(X,DIM) works along the dimension DIM.
```

# Esercizio:

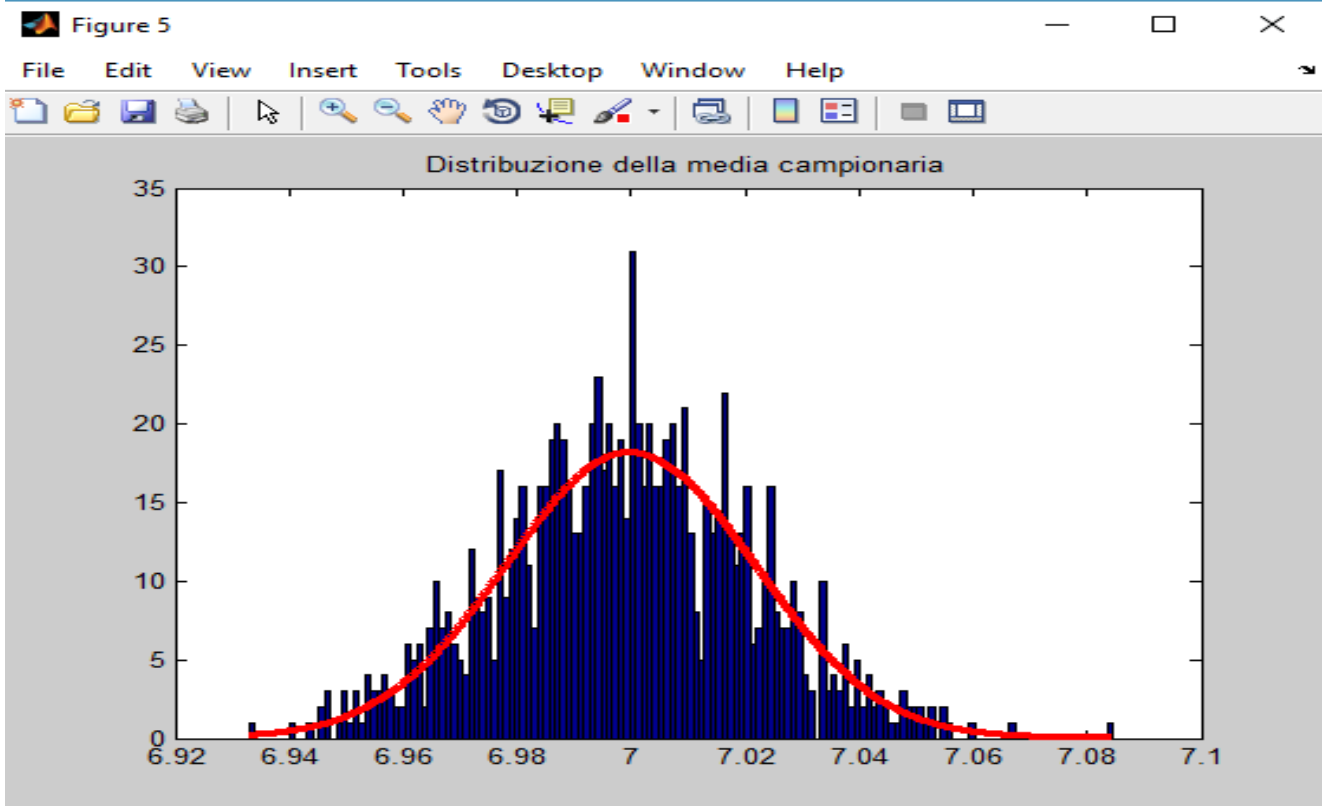


- Costruire 1000 campioni di numerosità N=10000, appartenenti ad una distribuzione Gaussiana con media $\mu = 7$ e deviazione standard $\sigma = 2.3$ .

- Valutare visivamente con un istogramma come si distribuiscono le variabili aleatorie media campionaria e varianza calcolate sui 100 campioni creati in precedenza e caratterizzarli statisticamente.

# Soluzione:

```
samples = 1000;
N = 10000;
X = random('norm',7,2.3,[N,samples]);
mu = = mean(X);
sigma_2 = var(X);
```




```
figure; hist(mu, 100); title('Distribuzione della media campionaria');
figure; hist(sigma_2, 100); title('Distribuzione della varianza campionaria');

x_mu = linspace(min(mu), max(mu), 1000);
distr_mu = pdf('norm', x_mu, 7, std(mu));
figure; hist(mu, 150); title('Distribuzione della media campionaria');
hold on; plot(x_mu, distr_mu, 'r', 'LineWidth', 3);

figure; hist(sigma_2,100); title('Distribuzione della varianza campionaria');
```