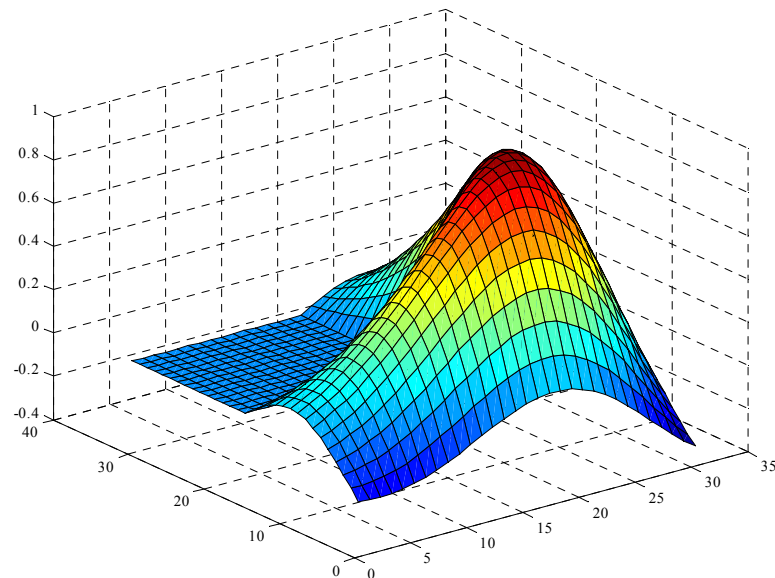
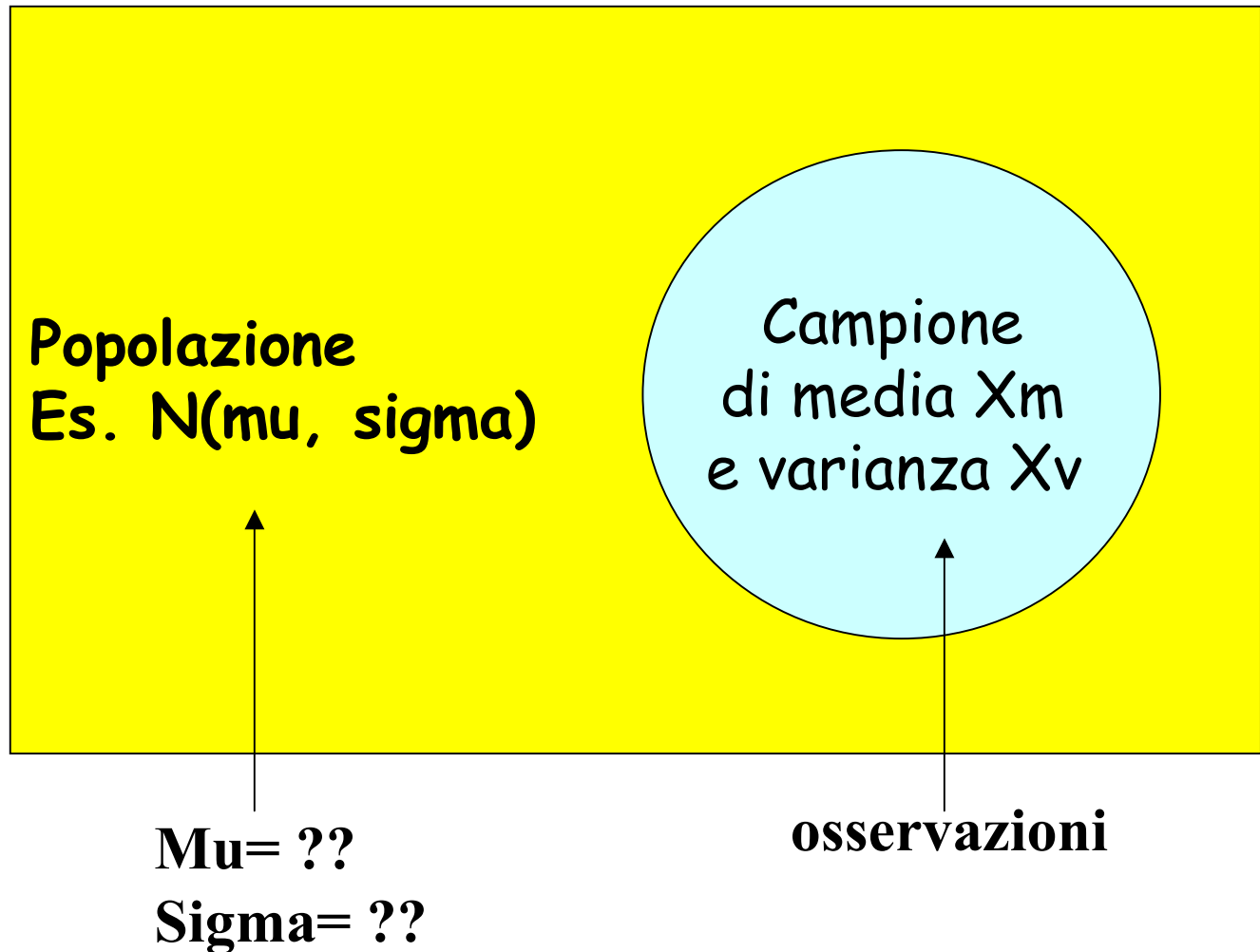


# Esercitazioni di Statistica con Matlab

*Dott. Orietta Nicolis*  
*<http://ingegneria.unibg.it>*



# Inferenza statistica



# Definizione delle variabili

Siano  $x$  e  $y$  due variabili campionarie

$x = [18 \ 25 \ 30 \ 20 \ 28];$

$y = [24 \ 25 \ 23 \ 25 \ 26];$

# Rappresentazione grafica

- `bar(x)`, grid on
- `axis([0 6 15 30])`
- `subplot(2,1,2)`
- `bar(y)`, grid on
- `axis([0 6 15 30])`

# Media

$$m = \frac{\sum_{i=1}^n X_i}{n}$$

- Mediante la funzione **mean**  
`xm=mean(x);`  
`ym=mean(y);`
- Mediante operazioni algebriche  
`xme=sum(x)/length(x);`  
`yme=sum(y)/length(y);`

# Varianza

$$v = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

- Mediante la funzione **var**

```
xv=var(x,1);
```

```
yv=var(y,1);
```

- oppure

```
xve=(sum(x-mean(x)).^2)/length(x);
```

```
yve=(sum(y-mean(y)).^2)/length(y);
```

# Varianza campionaria corretta

$$v = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

```
xv_corr=var(x);
```

```
yv_corr=var(x);
```

```
xve_corr=(sum(x-mean(x)).^2)/(length(x)-1);
```

```
yve_corr=(sum(y-mean(y)).^2)/(length(y)-1);
```

# Forma matriciale

```
z=[x' y'];
```

```
xym=mean(z);
```

```
xyv=var(z,1);
```

% divido per N

```
xyv=var(z);
```

% divido per N-1



# Distribuzione empirica dei dati campionari

`n=hist(Y)`

Calcola l'istogramma suddividendo i dati in 10 classi

Es.

```
y = randn(1000,1);  
hist(y)
```

```
n=hist(Y, x)
```

Calcola l'istogramma mettendo x al centro.

Esempio:

```
x = -3:.1:3;
```

```
y = randn(10000,1);
```

```
n=hist(y,x)
```

Oppure

```
[n xout]=hist(y,x)
```

n = colonna delle frequenze

xout =colonna dei dati

```
bar(xout, n)
```

# minimo, max e percentili

- `min(z)`            % valore minimo
- `max(z)`            % valore massimo
- `prctile(z, 50)`    % la mediana
- `prctile(z, 25)`    % il 1° quartile
- `prctile(z, 75)`    % il 3° quartile
- `sort(z)`            % ordina i dati in senso  
                         crescente

# Altre misure di dispersione

- `range(z)`
- `iqr(z)`            % Range interquartile D3-D1
- `std(z,1)`            % Deviazione standard
- `mad(z)`            % Deviazione media assoluta

# Esercizio

- Al fine di analizzare l'andamento dei voti nell'esame di statistica è stato estratto il seguente campione:

$Y=[22\ 23\ 20\ 19\ 26\ 28\ 28\ 30\ 22\ 23\ 04\ 29\ 23\ 24\ 24$   
 $25\ 26\ 27\ 26\ 24\ 22\ 22\ 23\ 25\ 24\ 26\ 21\ 24\ 24\ 27$   
 $25\ 25\ 25\ 28];$

Determinare la distribuzione di frequenza, la media, la moda, la mediana e la deviazione standard.

(tempo = 5min.)

# Boxplot

`boxplot(x, NOTCH)`

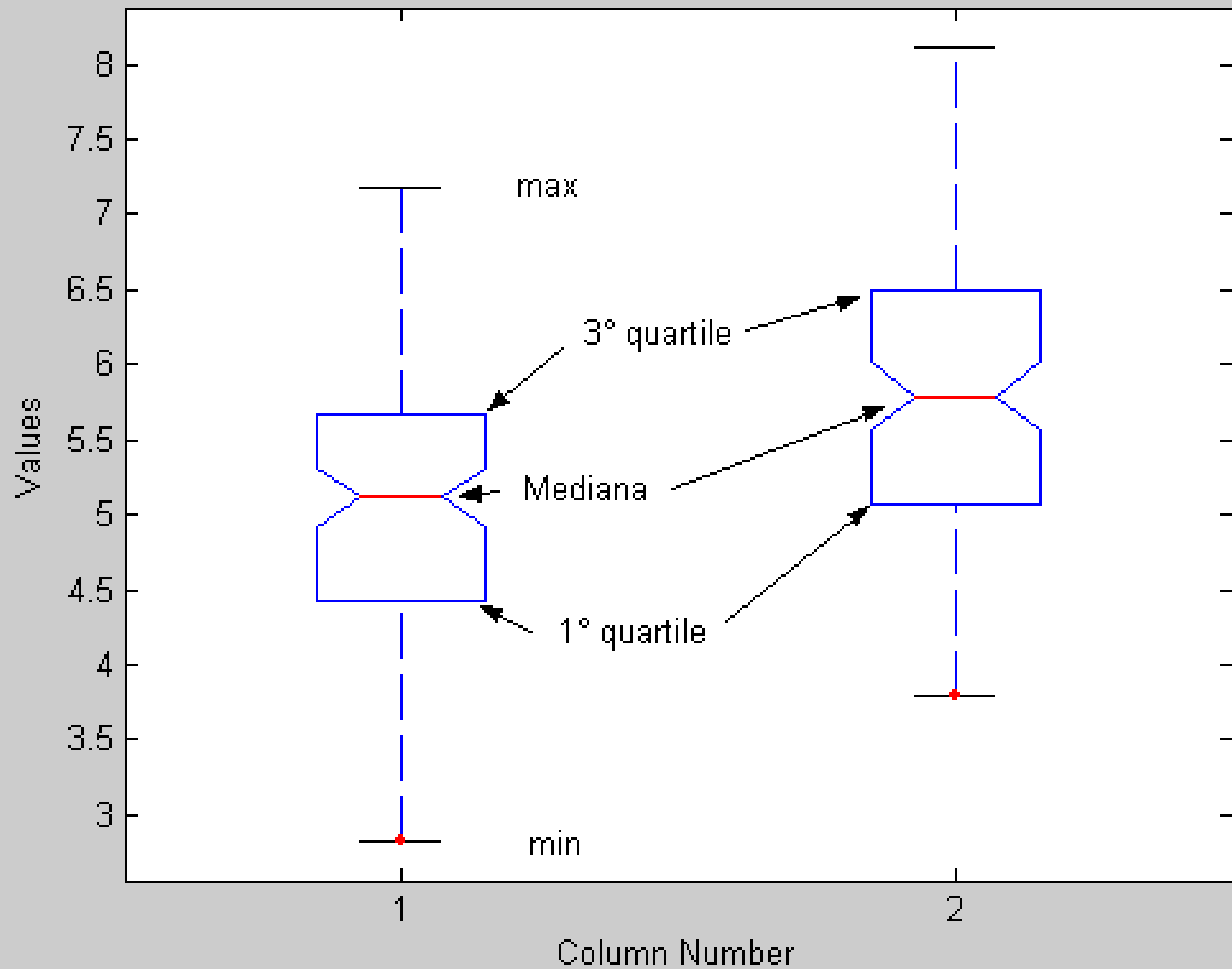
Esempio:

```
x1 = normrnd(5,1,100,1);
```

```
x2 = normrnd(6,1,100,1);
```

```
x = [x1 x2];
```

```
boxplot(x,1)
```



# Esercizio

Caricate il file `altezze.mat` che contiene i valori delle altezze di un campione di 50 maschi e 50 femmine di studenti universitari.

- Determinare la media, la varianza (semplice e corretta), mediana, quartili e percentili.
- La distribuzione dei dati campionari e il boxplot.

(tempo = 5min.)



# Esercizio

Da una distribuzione normale con media 20 e scarto quadratico medio 2, generare 3 campioni di numerosità 10, 100, 1000.

- Calcolare medie e varianze
- Confrontare le distribuzioni empiriche

(tempo = 5min.)

# Stima di massima verosimiglianza

Verosimiglianza di  $\theta$

$$L(\theta) = \prod_{i=1}^n f_{\theta}(x_i)$$

In matlab:

```
[phat, pci]=mle('dist', data, alpha)
```

```
[phat, pci]=mle('dist', data, alpha, p1*)
```

...fit

betafit, binofit, gamfit, normfit, weibfit

\*Si mette p1 (n. di prove) solo per la distribuzione binomiale

# *Stima di una proporzione*

Generiamo un campione di 10 elementi da una distribuzione binomiale di parametri  $n=100$  e  $p=0.25$

$Rb = \text{binornd}(100, 0.25, 10, 1)$

(Rb da il n° di successi in 100 prove)

La stima di massima verosimiglianza di  $p$  è

$[\text{phat } pci] = \text{mle}(\text{'binomial'}, Rb, 0.05, 100)$

*phat = stima di massima verosimiglianza del parametro vero della popolazione (es. 0.25)*

*pci = intervallo di confidenza di  $p$  ad un livello 0.05.*

# binofit

Al posto di mle è possibile utilizzare la funzione

```
[phat pci]=binofit(x, n, alpha)
```

Infatti,

```
[phat, pci]=binofit(Rb, 100, 0.05)
```

da gli stessi risultati

# *Stima della media e varianza*

```
Rn=normrnd(5, 1.5, 10, 1)
```

Genera un campione casuale di 10 elementi da una distribuzione normale con media  $\mu=5$  e  $\sigma=1.5$ .

```
[phat, pci]=mle('normal', Rn, 0.05)
```

oppure

```
[mu, sigma, muc1, sigmac1]=normfit(Rn, 0.05)
```

**Attenzione!** La dev.std. stimata MLE non è la dev.std corretta che si determina con NORMFIT.

**Attenzione!** Non confondere NORMFIT con NORMSTAT.

# *Esempio*

- *Dopo aver estratto un campione bernulliano di 20 unità da una popolazione esponenziale di parametro  $\lambda=0.25$  (media=4), stimare il parametro  $\lambda$  con il metodo della massima verosimiglianza.*