



# MatLab: Statistica Inferenziale Parametrica



**Centro E. Piaggio**  
bioengineering and robotics research center



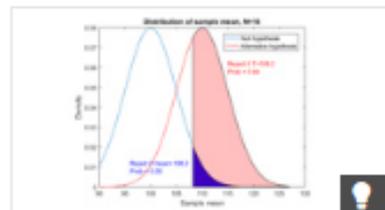
Ing. Vincenzo Catrambone  
vincenzo.catrambone@ing.unipi.it

## Test delle ipotesi

La variazione casuale può rendere difficile determinare se i campioni prelevati in diverse condizioni sono effettivamente differenti. Il test delle ipotesi è uno strumento efficace per analizzare se le differenze da campione a campione sono significative e se richiedono ulteriore valutazione o se sono coerenti con la casualità e la variazione attesa dei dati.

Statistics and Machine Learning Toolbox supporta ampiamente le procedure di test delle ipotesi parametriche e non parametriche, tra cui:

- *T*-test per uno o due campioni
- test non parametrici per un campione, campioni accoppiati e due campioni indipendenti;
- test di distribuzione (chi-quadrato, Jarque-Bera, Lilliefors e Kolmogorov-Smirnov);
- confronto delle distribuzioni (Kolmogorov-Smirnov per due campioni);
- test per autocorrelazione e casualità;
- test delle ipotesi lineari su coefficienti di regressione.



### Selezione della dimensione di un campione (Esempio)

Calcola la dimensione del campione necessaria per un test di ipotesi.

Statistic test,  
Confidence  
Interval,  
p-value:  
differences?

```
>> help ztest
```

```
ztest One-sample Z-test.
```

```
H = ztest(X,M,SIGMA) performs a Z-test of the hypothesis that the data in the vector X come from a distribution with mean M, and returns the result of the test in H. H=0 indicates that the null hypothesis ("mean is M") cannot be rejected at the 5% significance level. H=1 indicates that the null hypothesis can be rejected at the 5% level. The data are assumed to come from a normal distribution with standard deviation SIGMA.
```

```
[H,P] = ztest(...) returns the p-value, i.e., the probability of observing the given result, or one more extreme, by chance if the null hypothesis is true. Small values of P cast doubt on the validity of the null hypothesis.
```

```
[H,P,CI] = ztest(...) returns a 100*(1-ALPHA)% confidence interval for the true mean.
```

```
[H,P,CI,ZVAL] = ztest(...) returns the value of the test statistic.
```

```
>> help icdf
```

```
icdf Inverse cumulative distribution function for a specified distribution.
```

```
X = icdf(NAME,P,A) returns an array of values of the inverse cumulative distribution function for the one-parameter probability distribution specified by NAME with parameter values A, evaluated at the probability values in P.
```

Critic  
values?

```
>> help ttest
```

```
ttest One-sample and paired-sample t-test.
```

```
H = ttest(X) performs a t-test of the hypothesis that the data in the vector X come from a distribution with mean zero, and returns the result of the test in H. H=0 indicates that the null hypothesis ("mean is zero") cannot be rejected at the 5% significance level. H=1 indicates that the null hypothesis can be rejected at the 5% level. The data are assumed to come from a normal distribution with unknown variance.
```

```
H = ttest(X,Y) performs a paired t-test of the hypothesis that two matched samples, in the vectors X and Y, come from distributions with equal means. The difference X-Y is assumed to come from a normal distribution with unknown variance. X and Y must have the same length. X and Y can also be matrices or N-D arrays of the same size.
```

```
ttest2 Two-sample t-test with pooled or unpooled variance estimate.
```

```
H = ttest2(X,Y) performs a t-test of the hypothesis that two independent samples, in the vectors X and Y, come from distributions with equal means, and returns the result of the test in H. H=0 indicates that the null hypothesis ("means are equal") cannot be rejected at the 5% significance level. H=1 indicates that the null hypothesis can be rejected at the 5% level. The data are assumed to come from normal distributions with unknown, but equal, variances. X and Y can have different lengths.
```

## T-test:

- Singolo campione;
- dati appaiati.

## T-test2:

- dati non appaiati.

```
>> help vartest
```

```
vartest One-sample test of variance.
```

```
H = vartest(X,V) performs a chi-square test of the hypothesis that the data in the vector X come from a normal distribution with variance V, against the alternative that X comes from a normal distribution with a different variance. The result is H=0 if the null hypothesis ("variance is V") cannot be rejected at the 5% significance level, or H=1 if the null hypothesis can be rejected at the 5% level.
```

```
vartest2 Two-sample F test for equal variances.
```

```
H = vartest2(X,Y) performs an F test of the hypothesis that two independent samples, in the vectors X and Y, come from normal distributions with the same variance, against the alternative that they come from normal distributions with different variances. The result is H=0 if the null hypothesis ("variances are equal") cannot be rejected at the 5% significance level, or H=1 if the null hypothesis can be rejected at the 5% level. X and Y can have different lengths.
```

```
>> help anova1
```

```
anova1 One-way analysis of variance (ANOVA).
```

anova1 performs a one-way ANOVA for comparing the means of two or more groups of data. It returns the p-value for the null hypothesis that the means of the groups are equal.

```
P = anova1(X, GROUP, DISPLAYOPT)
```

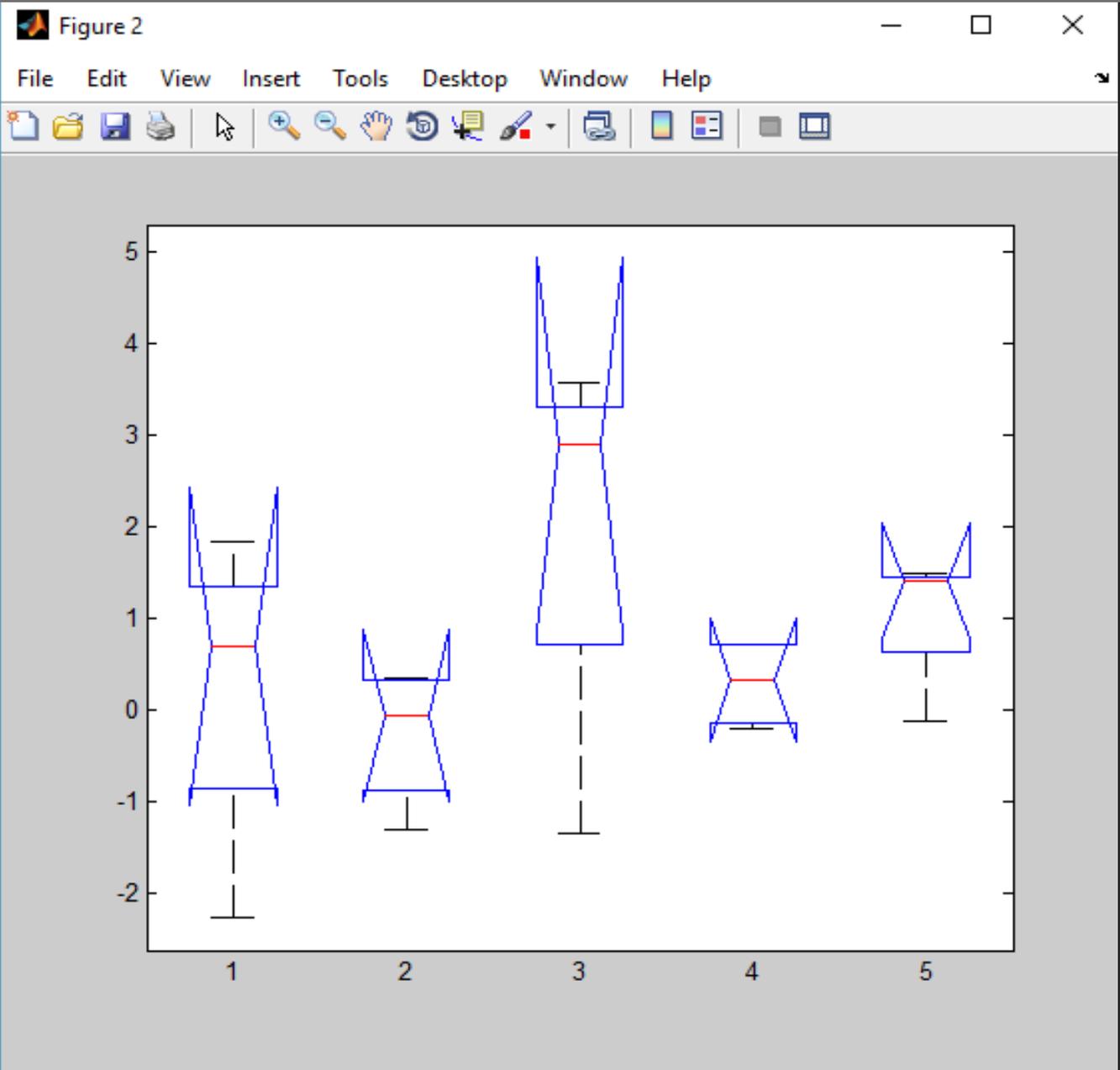
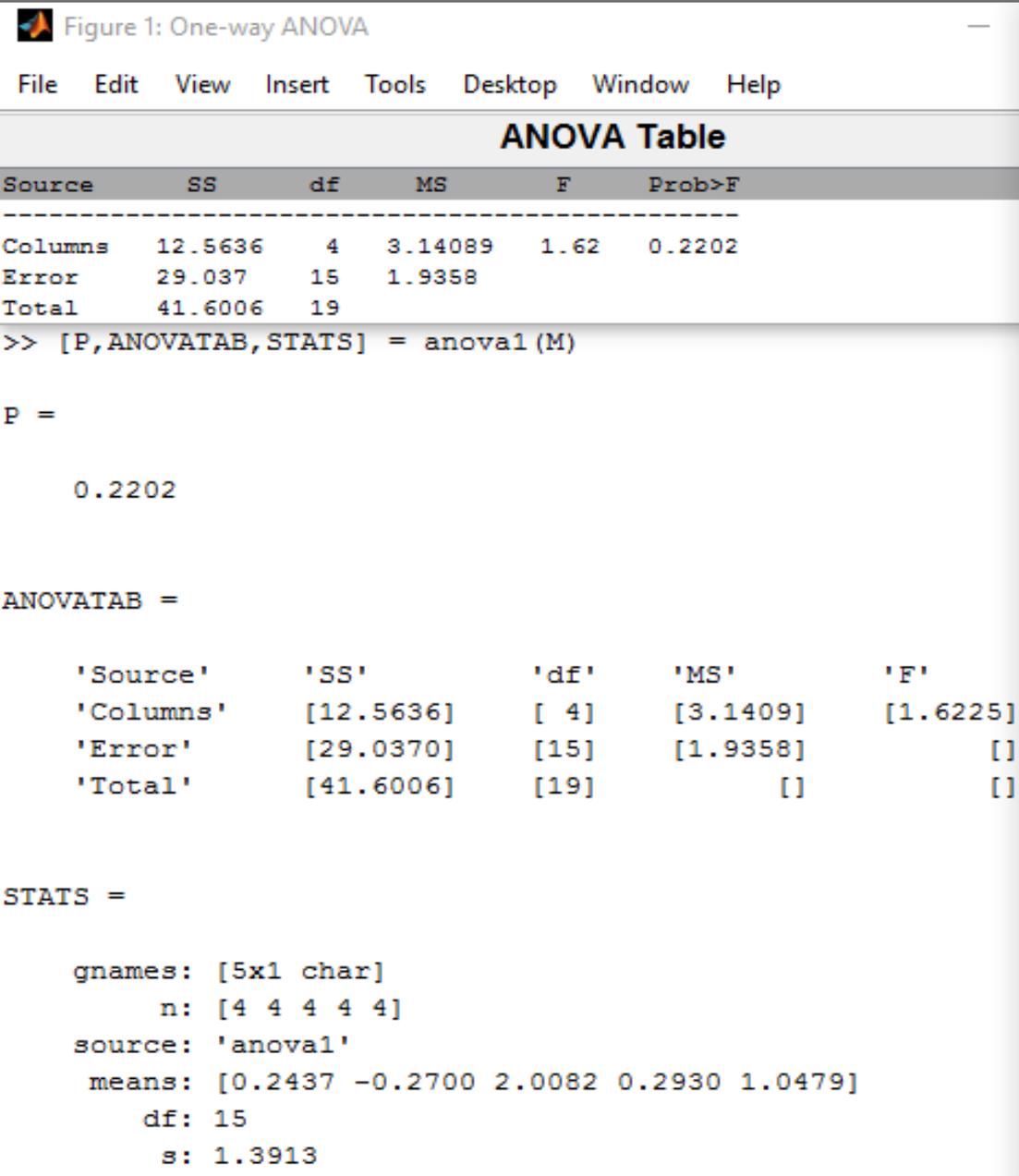
If X is a matrix, anova1 treats each column as a separate group, and determines whether the population means of the columns are equal.

This form of anova1 is appropriate when each group has the same number of elements (balanced ANOVA). GROUP can be a character array or a cell array of strings, with one row per column of X, containing the group names. Enter an empty array ([]) or omit this argument if you do not want to specify group names.

If X is a vector, GROUP must be a categorical variable, vector, string array, or cell array of strings with one group name for each element of X. X values corresponding to the same value of GROUP are placed in the same group.

```
[P, ANOVATAB] = anova1(...) returns the ANOVA table values as the cell array ANOVATAB.
```

# One way ANOVA function: anova1()



## Esercizio 1

Per studiare l'effetto a breve termine di una sostanza stupefacente sulle capacità intellettuali di soggetti (senza esperienze precedenti), alcuni ricercatori hanno verificato su un campione i cambiamenti nei punteggi ad opportuni test dopo aver assunto la sostanza. I risultati sono presentati nella seguente tabella:

<i>Soggetto</i>	1	2	3	4	5	6	7	8	9
<i>Differenza punteggio</i>	5	-17	-7	-3	-7	-9	-6	1	3

Si determini l'intervallo di confidenza per la media della differenza dei punteggi al 99% (Hp di normalità).

A	2.7	2.8	2.9	2.5	2.6	2.7	2.8	---
B	2.2	2.1	2.2	2.3	2.1	2.2	2.3	2.6

## Esercizio 2.

Sia  $t_v$  una variabile aleatoria relativa a realizzazioni di distribuzioni t di Student a  $v$  gradi di libertà, e  $F_{n,m}$  una variabile aleatoria relativa a realizzazioni della distribuzione di Fisher con  $n$  gradi del num. e  $m$  gradi del denominatore. Verificare la relazione  $t_v^2 = F_{1,v}$  utilizzando i campioni di dati A e B (ipotesi di Gaussianità delle v.a. associate ai campioni):

## %% Esercizio 2

```
x1 = [2.7 2.8 2.9 2.5 2.6 2.7 2.8];  
x2 = [2.2 2.1 2.2 2.3 2.1 2.2 2.3 2.6];  
[H_t,p_t,ci_t,stat_t] = ttest2(x1,x2);  
MAT = [[x1; NAN] x2];  
[p_F, ANVTB, stat_F] = anova1(MAT, [], 'off');
```

## %% Esercizio 1

```
x = [5 -17 -7 -3 -9 -7 -6 1 3];  
n = length(x);  
v = n-1;  
alpha = 0.01;  
p = 1-alpha/2;  
t_crit = icdf('t',p,v);  
m = mean(x);  
s = std(x);  
IC99 = [ m - t_crit*s/sqrt(n); m + t_crit*s/sqrt(n) ];
```

## Esercizio 3:

Il gruppo di ricerca di una nota casa farmaceutica vuole valutare l'effetto di un proprio farmaco anestetico rispetto i tempi di reazione ad uno stimolo audio. Per questo, sono stati acquisiti i tempi di reazione di un gruppo di soggetti che ha preso l'anestetico (gruppo A) ed un gruppo di soggetti (diverso dal primo) che non ha preso l'anestetico (gruppo B). Verificare se la differenza di valori tra i due gruppi è dovuta al caso.

A: 142 138 191 143 145 142 131 138 141 156  
167 230 135

B: 245 144 158 139 244 268 289 171 142 256  
165 267 178

## Esercizio 4:

Risolvere l'esercizio precedente, nel caso in cui fossero gli stessi soggetti, momenti diversi, a costituire i gruppi A e B.

## Esercizio 5:

Si supponga che in una certa unità sanitaria locale il tempo medio necessario per ottenere un referto riferito alla cardiologia sia di 2 giorni con una deviazione standard di 1, mentre per ottenere un referto riferito alla ematologia sia di 20 ore con una deviazione standard di 2 giorni.

Considerando un campione di 70 referti cardiologici e 225 referti ematologici effettuati nella stessa USL, calcolare la probabilità che i referti cardiologici siano pronti entro 24 ore dopo rispetto a quelli ematologici.

## Esercizio 6:

Il campione di dati A è stato acquisito da un campione di 7 soggetti sani. Dimostrare che vi è una differenza statisticamente significativa con una significatività del 5% rispetto al campione dei soggetti patologici rappresentato dai gruppi B e C. Si considerino i campioni derivanti da popolazioni con distribuzione gaussiana.

<b>A</b>	<b>B</b>	<b>C</b>
75	41	47
52	30	29
58	38	25
19	20	10
34	28	13
33	26	17
25	15	12

### %% Esercizio 3

```
A = [142 138 191 143 145 142 131 138 141 156 167 230 135];  
B = [245 144 158 139 244 268 289 171 142 256 165 267 178];  
[H,p,ci,stat] = ttest2(A,B);
```

### %% Esercizio 6

```
A = [75 52 58 19 34 33 25];  
B = [41 30 38 20 28 26 15];  
C = [47 29 25 10 13 17 12];  
alpha = 0.05;  
ABC = [A' B' C'];  
[p_6, ANVTB_6, stat_6] = anoval(ABC, [], 'off');  
if p_6 < alpha  
    [H_AB,p_AB,ci_AB,stat_AB] = ttest2(A,B,'alpha',alpha/2);  
    [H_AC,p_AC,ci_AC,stat_AC] = ttest2(A,C,'alpha',alpha/2);  
end
```

### %% Esercizio 4

```
[H,p,ci,stat] = ttest(A,B);
```

### %% Esercizio 5

```
mu1 = 48;          mu2 = 20;  
std1 = 24;         std2 = 48;  
n1 = 70;           n2 = 125;  
  
mu_d = mu1 - mu2;  
std_d = sqrt( std1^2/n1 + std2^2/n2);  
Z_d = (24-mu_d)/std_d;  
p_z = cdf('norm',Z_d);
```