Analysing and presenting data: **practical hints**

Giorgio MATTEI

giorgio.mattei@centropiaggio.unipi.it

Course: Fenomeni di trasporto biologico
Date: 06 Oct 2014

Equal or different?
more than ***two samples***

ANalysis Of VAriance (ANOVA)

- **More than 2** groups: **NO pairwise comparisons** (*t-test*)

  ↑ groups → ↑ overall probability that at least one of them is significant
  (e.g. α=0.05 and n=20 → in average 1 group will be significantly different for the case, even if H_0 is true)

$$H_0: \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$$

H_1: not all means are equal

- **all** means **are different**
- ...
- **one** mean **is different** from the others, which are all equals

- **ANOVA**
  - uses **Fisher's distribution** (**F-**distribution)
  - the **sources of variations** on observed values of **two or more groups** can be **decomposed** and **accurately measured**
  - the **source of variation** is called **EXPERIMENTAL FACTOR** (or **TREATMENT**) and can be multi-levelled
  - each **unit or observation** of the experimental factor is called **REPLICATION**

one-way ANOVA: an example
The problem

- Content of iron in air in 3 different zones (A, B, C) of a city (µg/N mc at 0 °C and 1013 mbar)

EXPERIMENTAL FACTOR		
A	B	C
2,71	1,75	2,22
2,06	2,19	2,38
2,84	2,09	2,56
2,97	2,75	2,60
2,55		2,72
2,78		

MATLAB

one-way ANOVA

***p*-value for H_0** (means of the groups are equal)

ANOVA table values

Structure of statistics useful **for performing a multiple comparison of means** with the **MULTCOMPARE** function

[P,ANOVATAB,STATS] = anova1(X,GROUP,DISPLAYOPT)

Matrix with **1 group per column** (requires equal-sized samples)

Character array: one row per column of X, containing the **group names**

Vector of data

Vector: one group name for **each element of X**

'on' (the default) to **display figures containing a standard one-way anova table and a boxplot**, or 'off' to omit these displays

MATLAB
one-way ANOVA: example

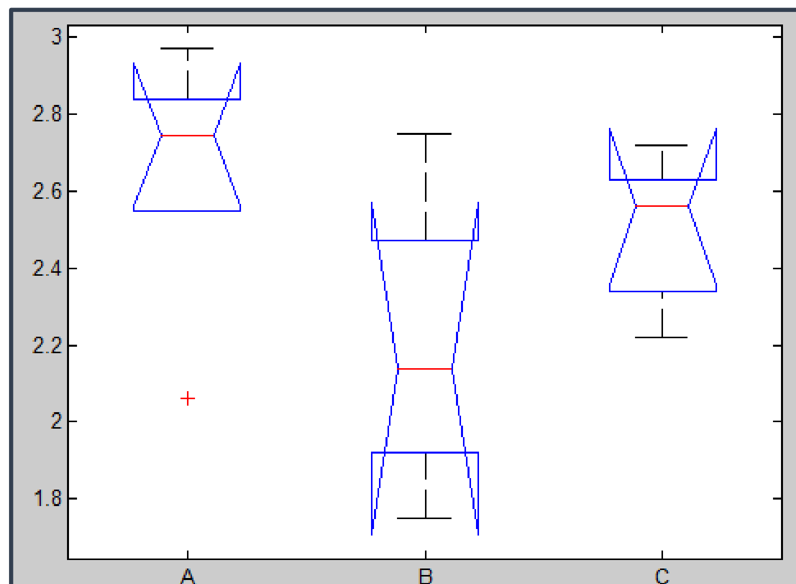
```
>> X=[2.71,2.06,2.84,2.97,2.55,2.78,1.75,2.19,2.09,2.75,2.22,2.38,2.56,2.6,2.72]';
>> GROUP=['A','A','A','A','A','A','B','B','B','B','C','C','C','C','C']';
>> [P,ANOVATAB,STATS] = anova1(X,GROUP)
```

P = 0.1204

```
STATS =  gnames: {3x1 cell}
              n: [6 4 5]
         source: 'anova1'
          means: [2.6517 2.1950 2.4960]
             df: 12
              s: 0.3148
```

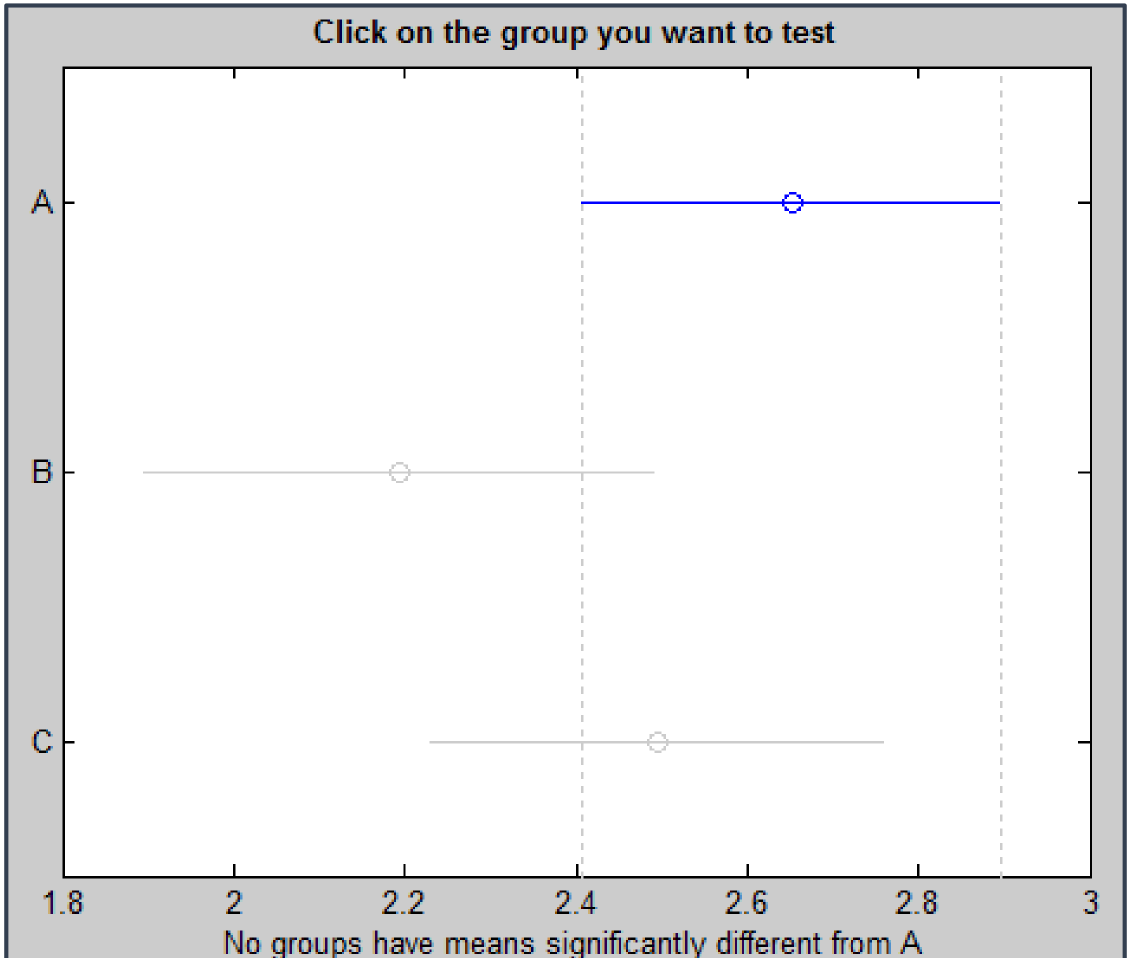
ANOVA Table

Source	SS	df	MS	F	Prob>F
Groups	0.50294	2	0.25147	2.54	0.1204
Error	1.1889	12	0.09908		
Total	1.69184	14			



MATLAB
one-way ANOVA: example

COMPARISON = multcompare(STATS)



two-way ANOVA: an example
The problem

- **Content** of **Pb** in air in **5 different urban zones** revealed every **6 hours** during the day

	TREATMENTS (urban zone)				
BLOCKS (time)	**1**	**2**	**3**	**4**	**5**
6 am	28	25	30	22	26
12 am	34	32	37	31	30
6 pm	22	21	24	20	19
12 pm	36	31	40	33	29

MATLAB
two-way ANOVA

***p*-value for H_0** (means of the groups are equal)

ANOVA table values

Structure of statistics useful **for performing a multiple comparison of means** with the **MULTCOMPARE** function

[P,ANOVATAB,STATS] = anova2(X,REPS,DISPLAYOPT)

Matrix of data (balanced ANOVA → equal number of repetitions)

Columns: 1st factor
Rows: 2nd factor

REPS indicates the **number** of **observations per "cell"**

A **"cell" contains REPS number** of **rows**

'on' (the default) to **display a standard two-way anova table**, or 'off' to skip the display

MATLAB
two-way ANOVA: example

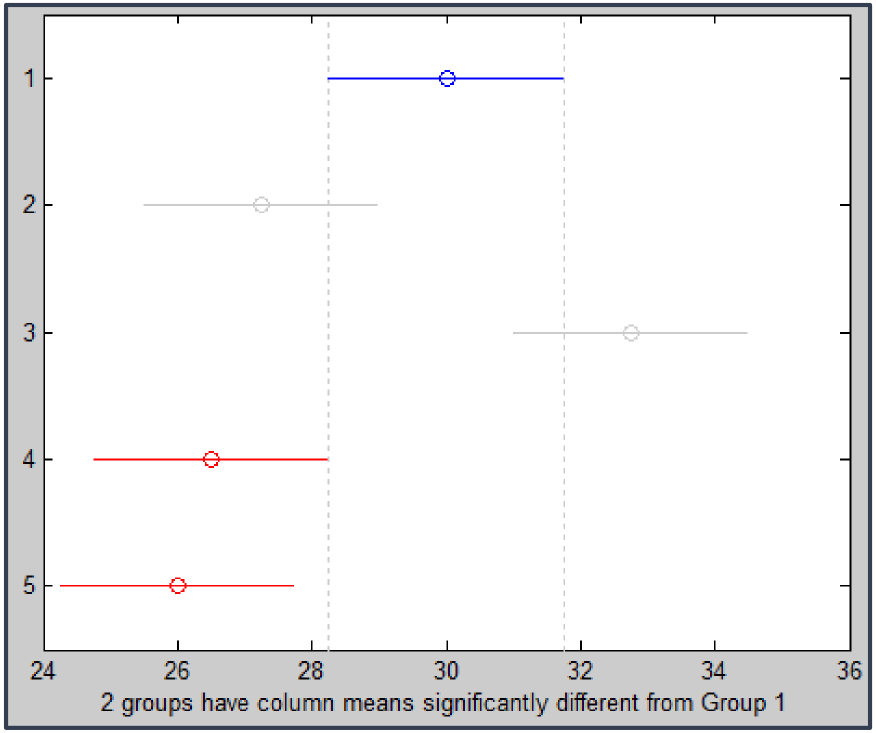
```
>> X=[28 25 30 22 26;
34 32 37 31 30;
22 21 24 20 19;
36 31 40 33 29];
>> [P,ANOVATAB,STATS] = anova2(X)

P =   1.0e-03 *  0.2187    0.0001

STATS =
     source: 'anova2'
    sigmasq: 2.3917
   colmeans: [30 27.2500 32.7500 26.5000 26]
       coln: 4
   rowmeans: [26.2000 32.8000 21.2000 33.8000]
       rown: 5
      inter: 0
       pval: NaN
         df: 12
```

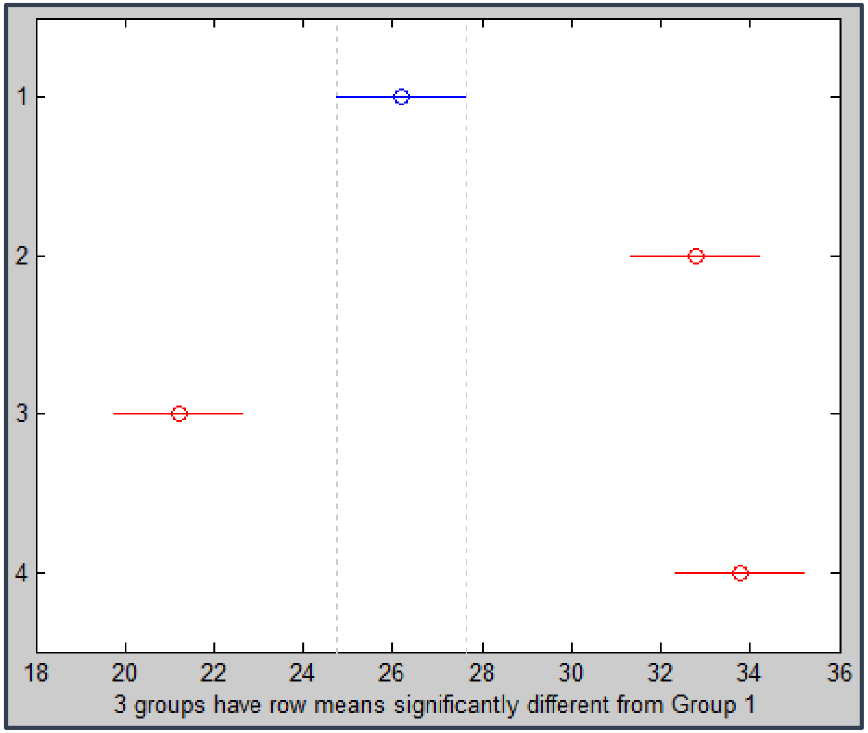
ANOVA Table

Source	SS	df	MS	F	Prob>F
Columns	128.5	4	32.125	13.43	0.0002
Rows	525.8	3	175.267	73.28	0
Error	28.7	12	2.392		
Total	683	19			

MATLAB
one-way ANOVA: example

COMPARISON = multcompare(STATS, 'estimate', 'column' (default) or 'row')

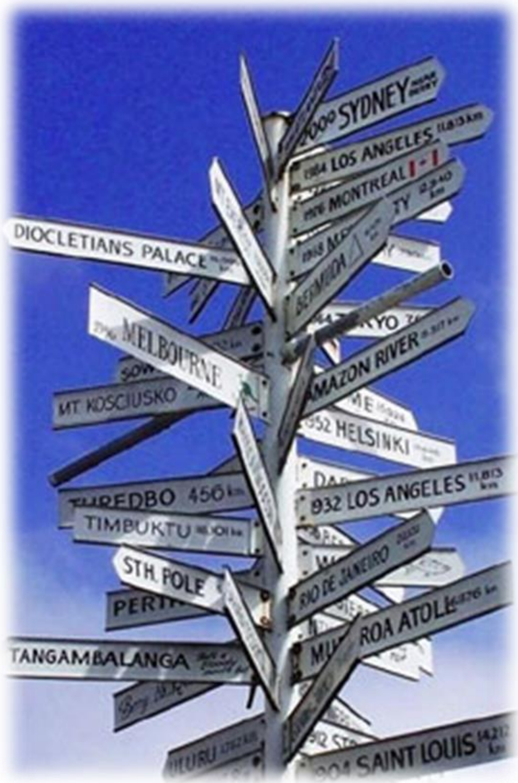




Columns (i.e. **urban zones**)

Rows (i.e. **times**)

MATLAB
anovan: N-way *analysis of variance*



MATLAB
Practical exercises

Sex (M=0, F=1)	Weight (kg)	Height (m)	Age (y)	Cardiac rate (bpm)	Respiratory rate (bpm)	Wrist circumference (cm)
0	65	1.81	21	61	18	17
1	50	1.64	21	59	11	14
1	63	1.7	21	74	16	15
1	50	1.5	21	72	16	15
1	60	1.75	21	67	17	15
0	75	1.69	22	73	15	20
0	60	1.72	21	53	15	18
1	53	1.55	21	57	16	15
1	53	1.62	21	64	16	16
0	78	1.85	21	64	14	17
0	84	1.83	21	70	17	18
0	60	1.7	21	70	17	16

- Plot **data distributions**
- Calculate **mean** and **standard deviation**
- Plot each **parameter versus the wrist circumference** and perform a **linear fit**
- Compute the **covariance matrix** using cov(X), evaluate the **correlation coefficients** using corrcoef(X), **plot variable correlations** using corrplot (X) and determine the parameter most correlated with the wrist circumference
- For each parameter **test the statistic significance between male and female**