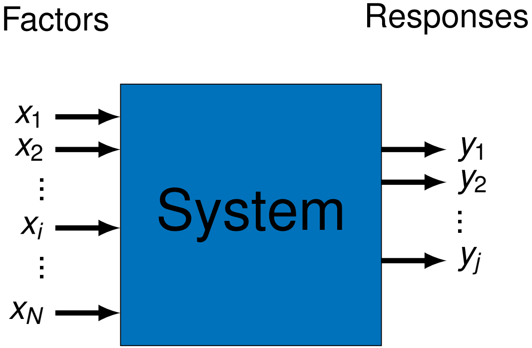
Constant & random coefficient models

Dr J.-M. Fürbringer

November 20, 2024

1. Qualitative factors

1.1 System & Model



$$\vec{Y} = \vec{F}(\vec{X})$$

1.2 Qualitative vs quantitative factors

- A **qualitative factor** is a factor whose levels can not be classified in an order of magnitude. Examples are spieces of plants, categories of products or element of a process.
- A **quantitative factor** is a factor that can be sense-fully classified in an order of magnitude potentially in relation with the response(s) to be analyzed. Examples are the weight of an object in relation with mechanics, the duration of an operation in relation with the effectiveness of a process.

1.3 Case 1: optimization of a workshop

Your company is producing mechanical pieces for the aeronautic industry. An analysis of last semester production has shown that the workshop WA3 has a quality problem.

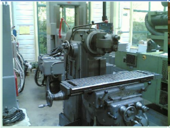
You are in charge of identifying the origin of the problem.

After discussing with the workshop supervisor, you have identified three possible factors that possibly affect the quality of the production:

1. the machines
2. the drills
3. the operators

You want now to evaluate their respective influence on the quality.

1.4 The factors

Operators			
Machines			
Drills			

1.5 The mindmap

Production quality in a workshop

- **Objective**: Identify the dominent factors to manage the production quality of a workshop
- **Type of investigation**: Experiment: factors can be manipulated
- **Factors**:
  1. Operators (3 levels)
  2. Machines(3 levels)
  3. Drills(3 levels)
- **Response(s)**: The quality could be mesured by different observation and measurements: the surface quality is chosen: the site in *mu*m of the bigest defect detected on the surface.
- **Model(s)**: A constant coefficient model (CCM) with or without interactions
- **Strategies**:
  1. Factorial design (all the possible experiments)
  2. Latin square design (based on a magic square)

2. Sweeping

2.1 First strategy

Factorial design: all the possible combinations $\rightarrow$ 27 runs

		Machines		
defect in μm	**Drill**	**Deckel**	**Schaublin**	**Maho**
Charlie	1mm	23.7	27.4	27.3
	5mm	16.7	22.9	22.6
	20mm	9.9	16.5	16.6
Pierre	1mm	22.8	29.7	28.4
	5mm	18.3	23.8	24.0
	20mm	11.4	18.0	18.6
Louis	1mm	24.3	30.7	30.3
	5mm	18.7	24.4	22.3
	0mm	13.3	18.5	18.8

2.2 Model with constant effects

- Synthesis of observations with **predictive capability**
- Discrete variables (qualitative and quantitative)
- The experimenter chooses specifically the levels of the factors: *Charlie*, *Pierre* and *Louis* are the operators of the workshop
- Mathematically represented by:

$$y_{mdoi} = \mu + \gamma_m + \tau_d + \omega_o + \epsilon_{mdoi} \tag{2.1}$$

with $m = 1:3$; $d = 1:3$; $o = 1:3$; $i = 1:N$;

- Here is an example

$$\widehat{y}_{mdo} \approx 21.5 + \left\{ \begin{array}{c} -4.41 \\ 2.57 \\ 1.84 \end{array} \right._m + \left\{ \begin{array}{c} 5.23 \\ 0.19 \\ -5.42 \end{array} \right._d + \left\{ \begin{array}{c} -0.56 \\ 0.51 \\ 0.05 \end{array} \right._o$$

2.3 Model with random effects

- Synthesis of observations with **limited predictive capabilities**
- Discrete variables (qualitative and quantitative)
- The experimenter chooses the levels of some factors randomly: *Charlie*, *Pierre* and *Louis* are members of the group of the operators and have been chosen randomly
- Mathematically represented also by equation 2.1

2.4 Two types of effects

1. **Constant effects**
   - → Which state of the factor optimizes the answer?
   - → All the population is measured
   - → All factors are fixed: *constant-effect model*
   - → $H_o : \tau_i = 0, \forall i$

2. **Random effects**
   - → Which are the dominant factors?
   - → Measurement on a random sample
   - → All factors are random: *random-effect model*
   - → Otherwise: *mixed-effect model*
   - → $H_o : \sigma_\tau = 0$

2.5 Decomposition by sweeping

- The data corresponds to a vector **Y** of 27 dimensions
- The design corresponds to a set of vectors constituting a base
- The model components are the projections of **Y** on the set of vectors

	Drill	Deckel	Schaublin	Maho
Charlie	1mm	23.7μm	27.4μm	27.3μm
	5mm	16.7μm	22.9μm	22.6μm
	20mm	9.9μm	16.5μm	16.6μm
Pierre	1mm	22.8μm	29.7μm	28.4μm
	5mm	18.3μm	23.8μm	24.0μm
	20mm	11.4μm	18.0μm	18.6μm
Louis	1mm	24.3μm	30.7μm	30.3μm
	5mm	18.7μm	24.4μm	22.3μm
	0mm	13.3μm	18.5μm	18.8μm

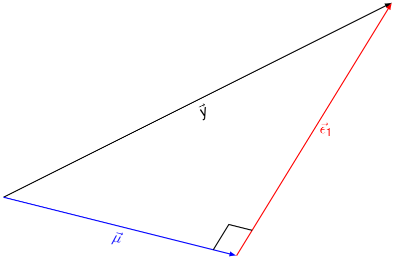
$$\rightarrow \quad y_{mdoi} = \mu + \gamma_m + \tau_d + \omega_o$$

2.6 Step 1: determination of the constant

data = grand mean + first residual

$$\begin{cases} y_{mdoi} = \mu + \epsilon_{mdoi} \\ \mu = \frac{1}{N} \sum_{mdoi} y_{mdoi} \end{cases} \tag{2.2}$$

$$\begin{cases} \vec{y} = \vec{\mu} + \vec{\epsilon}_1 \\ \vec{\mu} \cdot \vec{\epsilon}_1 = 0 \end{cases} \tag{2.3}$$

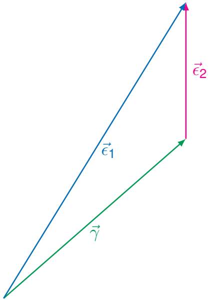


2.7 Step 2: calculation of the effects

- first residual = effect + new residual
- effect = means of first residual by groups of the related factor
- successively for each factor, the order doesn't mater

$$\begin{cases} y_{mdoi} = \mu + \gamma_m + \epsilon_{mdoi,2} \\ \gamma_m = \frac{1}{N_m} \sum_{doi} \epsilon_{m...} \end{cases} \tag{2.4}$$

$$\begin{cases} \vec{y} = \vec{\mu} + \vec{\gamma} + \vec{\epsilon}_2 \\ (\vec{\mu} + \vec{\gamma}) \cdot \vec{\epsilon}_2 = 0 \end{cases} \tag{2.5}$$

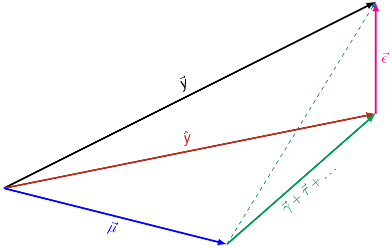


2.8 Step 3: building the model

Measurement Y = mean + effect + residue

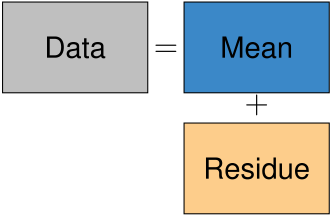
Model $\widehat{Y}$ = mean + effect

$$\begin{cases} \vec{y} = \vec{\mu} + \vec{\gamma} + \vec{\tau} + \ldots + \vec{\epsilon} \\ \hat{y} \cdot \vec{\epsilon} = 0 \end{cases} \tag{2.6}$$

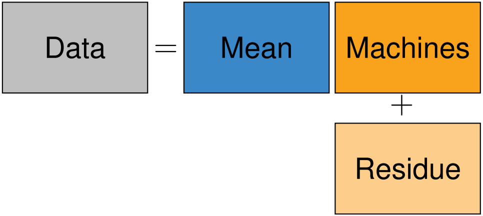


2.9 Effect inference

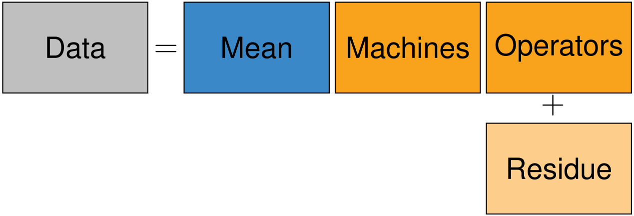
View by blocs of data



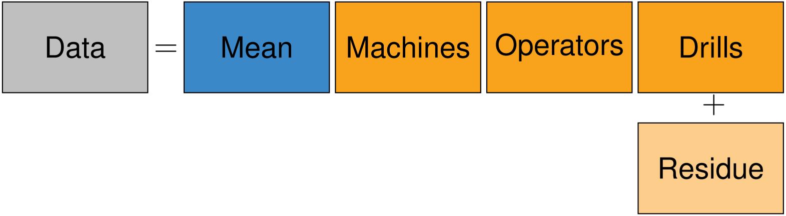
2.9 Effect inference



2.9 Effect inference



2.9 Effect inference



2.10 Effect inference - spreadsheet view

The data correspond to surface defects in μm

Grand mean

21,48	21,48	21,48
21,48	21,48	21,48
21,48	21,48	21,48
21,48	21,48	21,48
21,48	21,48	21,48
21,48	21,48	21,48
21,48	21,48	21,48
21,48	21,48	21,48
21,48	21,48	21,48

Machine

-3,80	2,07	1,73
-3,80	2,07	1,73
-3,80	2,07	1,73
-3,80	2,07	1,73
-3,80	2,07	1,73
-3,80	2,07	1,73
-3,80	2,07	1,73
-3,80	2,07	1,73
-3,80	2,07	1,73

Operator

-1,10	-1,10	-1,10
-1,10	-1,10	-1,10
-1,10	-1,10	-1,10
0,21	0,21	0,21
0,21	0,21	0,21
0,21	0,21	0,21
0,89	0,89	0,89
0,89	0,89	0,89
0,89	0,89	0,89

Drill

5,69	5,69	5,69
0,05	0,05	0,05
-5,74	-5,74	-5,74
5,69	5,69	5,69
0,05	0,05	0,05
-5,74	-5,74	-5,74
5,69	5,69	5,69
0,05	0,05	0,05
-5,74	-5,74	-5,74

Residue 1

2,17	5,91	5,81
-4,79	1,40	1,11
-11,61	-5,00	-4,92
1,31	8,24	6,88
-3,17	2,37	2,56
-10,05	-3,47	-2,84
2,80	9,26	8,79
-2,76	2,89	0,83
-8,17	-2,96	-2,68

Residue 2

5,98	3,84	4,08
-0,98	-0,67	-0,62
-7,80	-7,07	-6,65
5,12	6,17	5,15
0,64	0,30	0,83
-6,24	-5,54	-4,57
6,61	7,19	7,06
1,05	0,82	-0,90
-4,36	-5,03	-4,41

Residue 3

7,08	4,94	5,18
0,12	0,43	0,48
-6,70	-5,97	-5,55
4,91	5,96	4,95
0,43	0,09	0,63
-6,45	-5,75	-4,77
5,72	6,30	6,17
0,16	-0,07	-1,79
-5,25	-5,92	-5,30

Final residue

1,39	-0,75	-0,51
0,06	0,38	0,43
-0,96	-0,23	0,19
-0,78	0,27	-0,74
0,38	0,04	0,57
-0,71	-0,01	0,97
0,03	0,61	0,48
0,10	-0,13	-1,84
0,49	-0,18	0,44

2.11 Inference of the interaction coefficients

$$Y_{mdoi} = \mu + \gamma_m + \tau_d + \omega_o + {\color{red}(\gamma\tau)_{md}} + {\color{red}(\gamma\omega)_{mo}} + {\color{red}(\tau\omega)_{do}} + \epsilon_{mhoi} \tag{1.7}$$

Drill

5,69	5,69	5,69
0,05	0,05	0,05
-5,74	-5,74	-5,74
5,69	5,69	5,69
0,05	0,05	0,05
-5,74	-5,74	-5,74
5,69	5,69	5,69
0,05	0,05	0,05
-5,74	-5,74	-5,74

MxO

0,16	-0,20	0,04
0,16	-0,20	0,04
0,16	-0,20	0,04
-0,37	0,10	0,27
-0,37	0,10	0,27
-0,37	0,10	0,27
0,21	0,10	-0,31
0,21	0,10	-0,31
0,21	0,10	-0,31

DxO

0,04	0,04	0,04
0,29	0,29	0,29
-0,33	-0,33	-0,33
-0,42	-0,42	-0,42
0,33	0,33	0,33
0,08	0,08	0,08
0,37	0,37	0,37
-0,62	-0,62	-0,62
0,25	0,25	0,25

MxD

0,21	0,04	-0,26
0,18	0,10	-0,28
-0,39	-0,14	0,53
0,21	0,04	-0,26
0,18	0,10	-0,28
-0,39	-0,14	0,53
0,21	0,04	-0,26
0,18	0,10	-0,28
-0,39	-0,14	0,53

Final residue

1,39	-0,75	-0,51
0,06	0,38	0,43
-0,96	-0,23	0,19
-0,78	0,27	-0,74
0,38	0,04	0,57
-0,71	-0,01	0,97
0,03	0,61	0,48
0,10	-0,13	-1,84
0,49	-0,18	0,44

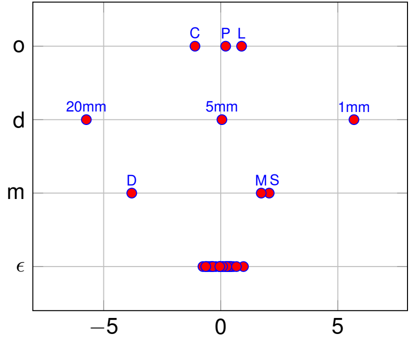
1,22	-0,55	-0,55
-0,10	0,58	0,39
-1,13	-0,03	0,15
-0,41	0,17	-1,01
0,75	-0,06	0,31
-0,34	-0,11	0,70
-0,18	0,51	0,79
-0,10	-0,23	-1,54
0,28	-0,28	0,75

1,18	-0,59	-0,59
-0,39	0,29	0,10
-0,79	0,30	0,49
0,01	0,59	-0,59
0,42	-0,39	-0,02
-0,42	-0,19	0,62
-0,55	0,14	0,42
0,52	0,40	-0,91
0,03	-0,53	0,50

0,97	-0,64	-0,33
-0,57	0,19	0,38
-0,40	0,44	-0,05
-0,21	0,54	-0,34
0,23	-0,49	0,26
-0,03	-0,05	0,08
-0,76	0,09	0,67
0,34	0,30	-0,64
0,43	-0,39	-0,04

2.12 Case 1 linear model

$$\widehat{Y}_{mdo} \approx 21.48 + \left\{ \begin{array}{r} -3.8 \\ 2.07 \\ 1.73 \end{array} \right\}_m + \left\{ \begin{array}{r} 5.69 \\ 0.05 \\ -5.74 \end{array} \right\}_d + \left\{ \begin{array}{r} -1.10 \\ 0.21 \\ 0.89 \end{array} \right\}_o$$



2.13 Case 1 linear model with interactions

$$\widehat{Y}_{mdo} \approx 21.48 + \left\{\begin{array}{r} -3.8 \\ 2.07 \\ 1.73 \end{array}\right\}_m + \left\{\begin{array}{r} 5.69 \\ 0.05 \\ -5.74 \end{array}\right\}_d + \left\{\begin{array}{r} -1.10 \\ 0.21 \\ 0.89 \end{array}\right\}_o$$

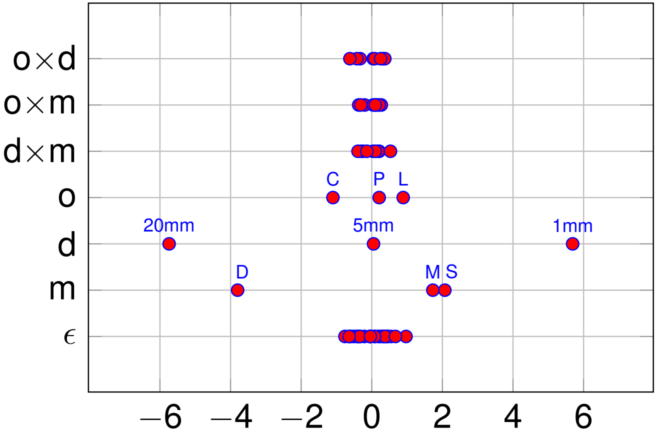
$$+ \left\{\begin{array}{rrr} 0.21 & 0.04 & -0.26 \\ 0.18 & 0.10 & -0.28 \\ -0.39 & -0.14 & 0.53 \end{array}\right\}_{d\times m}$$

$$+ \left\{\begin{array}{rrr} 0.16 & -0.20 & 0.04 \\ -0.37 & 0.10 & 0.27 \\ 0.21 & 0.10 & -0.31 \end{array}\right\}_{o\times m}$$

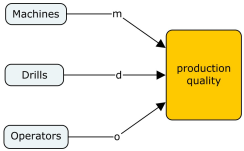
$$+ \left\{\begin{array}{rrr} 0.04 & 0.29 & -0.33 \\ -0.42 & 0.33 & 0.08 \\ 0.37 & -0.62 & 0.25 \end{array}\right\}_{o\times d}$$

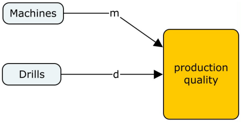
$$s = \sqrt{\frac{\sum {\epsilon_i}^2}{\nu}} = 0.75$$

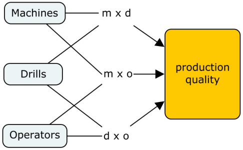
2.13 Case 1 linear model with interactions



2.14 Case 1 causal model







2.15 The direct way to compute the effects

(if r is the number of replicates, n the number of columns, m the number of rows)

The grand mean :

$$\mu = \frac{1}{nmr} \sum_i \sum_j \sum_k x_{ijk} \tag{1.8}$$

The column effects :

$$\alpha_j = \frac{1}{nr} \sum_i \sum_k x_{ijk} - \mu = \mu_j - \mu \tag{1.9}$$

The row effects :

$$\beta_j = \frac{1}{mr} \sum_j \sum_k x_{ijk} - \mu = \mu_i - \mu \tag{1.10}$$

The interactions effects :

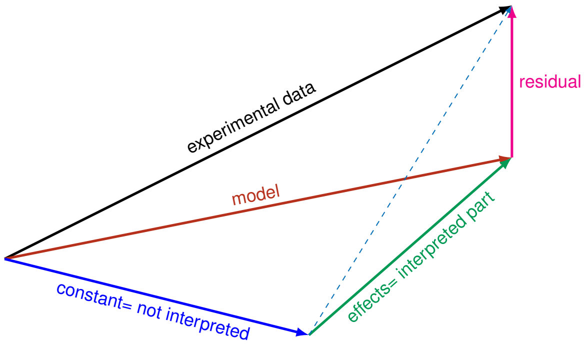
$$\alpha\beta_{ij} = \frac{1}{r} \sum_k x_{ijk} - \mu - \alpha_j - \beta_i \tag{1.11}$$

The residue :

$$\epsilon_{ijk} = x_{ij} - \mu - \alpha_j - \beta_j - \alpha\beta_{ij} \tag{1.12}$$

3. ANOVA

3.1 Distance between model and data



- Compare the norms of these vectors : sum of squares
- The explained part must be significantly greater than the residual (noise)

3.2 Sums of squares

Source	SS
Constant	12454
Machines	196
Drills	588
Operators	18.4
Residue	11.7
Total	13268

But each sub-space has a different dimension

3.3 Degree of freedom and Fisher's ratio

Source	SS	DF	MS	F
Constant	12454	1	12454	-
Machines	196	2	98	167
Drills	588	2	294	502
Operators	18.4	2	9.2	16
Residue	11.7	20	0.6	1
Total	13268	27		

$$MS = \frac{SS}{DF}$$

3.4 ANOVA table

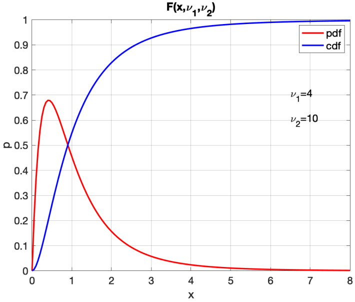
Source	SS	DF	MS	F	P*
Constant	12454	1	12454		
Machines	196	2	98	167	0.000%
Drills	588	2	294	502	0.000%
Operators	18.4	2	9.2	16	0.0%
Residue	11.7	20	0.6		
Total	13268	27			

Matlab

Calculation of p : fcdf(x,df1,df2,'upper')

*p: the probability to get F by chance

3.5 The Fisher distribution



In the ANOVA situation, x is the Fisher ratio. Observe that the probability (the red curve) to be large diminish exponentially.

3.6 Anova table

Anova factorial design

Source	SS	DF	MS	F	P
Constant	12453,7	1	12453,67		
Machine	195,9	2	97,97	167,2	0,000%
Drill	587,9	2	293,97	501,60	0,000%
Operator	18,4	2	9,21	15,7	0,0%
Residue	11,7	20	0,59	1	
Total	13267,7	27			

3.7 Matlab : ANOVAN routine

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} \begin{pmatrix} 'Charly' \\ 'Pierre' \\ 'Louis' \\ \vdots \end{pmatrix}$$

```
[p,table,stats]=anovan(data,{m,d,o},...
                'display','on',...
                'varname',{ 'Machine';'Drill';'Operator' },...
                'model','linear');
```

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Machine	206.577	2	103.289	126.39	0
Tool	544.079	2	272.039	332.88	0
Operator	0.381	2	0.191	0.23	0.794
Error	16.345	20	0.817		
Total	767.382	26			

Constrained (Type III) sums of squares.

3.8 Summary of ANOVA

- Decompose Y in orthogonal components
- Compute the sum of the squares
- Determine the degrees of freedom
- Compute the mean squares
- Compare with the residuals
- Disqualify the insignificants effects
- Compute again the error probability
- Analyse pairs of effects to determine significant contrasts

3.9 The concept of contrast

- Often the standard hypothesis $H_o : \mu_1 = \mu_i = 0$ is not answering the question of the investigator
- What is important is the comparison between treatments such as $H_o : \mu_3 = \mu_4$
- It is equivalent to $H_o : \mu_3 - \mu_4 = 0$
- A contrast is defined as (a is the nb of treatments)

$$\Gamma = \sum_{i=1}^{a} c_i \; \mu_i \tag{3.1}$$

- The t-statistics is then

$$t_o = \frac{\sum_{i=1}^{a} c_i \; \bar{y}_{i.}}{\sqrt{\frac{MS_E}{n} \sum_{i=1}^{a} c_i^2}} \tag{3.2}$$

3.10 Contrast confidence interval and LSD

- The confidence interval (CI) of a contrasts can be evaluated by

$$\Delta = t_{\alpha/2,\nu} \sqrt{\frac{MS_E}{n} \sum_{i=1}^{a} c_i^2} \tag{3.3}$$

ν being the DF of the model and then

$$\sum_{i=1}^{a} c_i \, \bar{y}_{i.} - \Delta \leqslant \Gamma \leqslant \sum_{i=1}^{a} c_i \, \bar{y}_{i.} + \Delta \tag{3.4}$$

n being the number of samples for each treatment, N being the total number of observations

- The least significant difference (LSD) is defined as

$$LSD = t_{\alpha/2,\nu} \sqrt{\frac{2MS_E}{n}} \tag{3.5}$$

3.11 LSD for factorial and latin square design

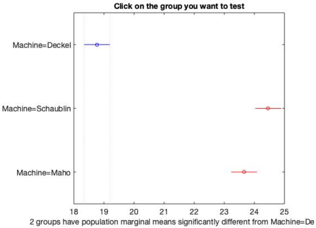
Factorial design

- Nb of obs $N = 27$
- Nb of obs by level $n = 9$
- Significance $\alpha = 5\%$
- $t_{0.975,20} = 2.1$
- $MS_E = 0.6$

$$LSD \approx 2.1 \times \sqrt{\frac{2 \times 0.6}{9}} \approx 0.42$$

3.12 Matlab multcompare routine

```
[c1,m1]=multcompare(stats_factorial,...
    "Alpha",0.05,...
    "CType","scheffe",...
    "Dimension",1,...
    "Estimate","column");
```




```
Pair comparison
   1.0000    2.0000   -6.5465   -5.6908   -4.8352    0.0000
   1.0000    3.0000   -5.7470   -4.8914   -4.0357    0.0000
   2.0000    3.0000   -0.0562    0.7995    1.6551    0.0698
Effects
  18.7679    0.2289
  24.4588    0.2289
  23.6593    0.2289
```