MTH8302 Modèles d'analyse de la variance
ch13 – plusieurs facteurs

- Exemples : concepts - définitions
- 2 facteurs : modèles - lien avec régression
- Interaction : définition - analyse
- Test interaction de Tukey si n = 1
- $n_{i\,j}$ inégaux dans cellule *

  * cellules = croisement des modalités facteurs
- Analyse covariance : présence facteur continu
- Plus de 2 facteurs
- Facteur secondaires : bloc - analyse
- Mesures répétées / Mesures longitudinales
- Facteurs emboités
- Facteurs aléatoires - modèles mixtes

Bernard Clément (2010), Modèles d'Analyse de la variance, Génistat Conseils Inc. (125 pages)
https://cours.polymtl.ca/mth6301/mth8302/mth8302-Cours&Plus/Clement-Modeles_Analyse_Variance.pdf

ANOVA avec 2 facteurs ou plus

problématique - analyse

- **FACTEURS** (variables qualitatives expérimentation)
  qualitatifs – quantitatifs – fixés – aléatoires - bloc (secondaire) – croisés - emboités (hiérarchiques)
  inter , intra

- **Protocoles (designs) statistiques**
  - complètement aléatoire (CRB)
  - bloc complet (RCBD)
  - bloc incomplet équilibré (PBIBD)
  - mixtes : effets fixés + effets aléatoires
  - présence de covariables
  - mode parcelles divisées (« SlitPlot »)

- **Tailles échantillonnales n dans les cellules**
  $n_{ijk..}$ égales ….. $n_{ijk..}$ inégales
  $n_{ij..} = 1$ ………… $n_{ij..} = 0$

Exemples données : modèles d'analyse de la variance

Kutner 5 ed. p. 833
influence de 2 facteurs A et B sur les ventes

1 ID	2 unité expérimentale	3 A_hauteur	4 B_largeur	5 rep	6 Y-vente
1	?	bas	régulière	1	47
2	?	bas	régulière	2	43
3	?	bas	large	1	46
4	?	bas	large	2	40
5	?	milieu	régulière	1	62
6	?	milieu	régulière	2	68
7	?	milieu	large	1	67
8	?	milieu	large	2	71
9	?	haut	régulière	1	41
10	?	haut	régulière	2	39
11	?	haut	large	1	42
12	?	haut	large	2	46

facteurs :
fixés ou aléatoires ?

unités expérimentales ?
homogènes ?
hétérogènes

Kutner et all 5 ed. p 1004 - expérience 3 facteurs A B C + répétition

1 ID	2 unité_ex	3 A_genre	4 B_bodyfat	5 C_fumeur	6 rep	7 Y_tol_Effort
1	?	homme	bas	light	1	24,1
2	?	homme	bas	light	2	29,2
3	?	homme	bas	light	3	24,6
4	?	femme	bas	light	1	20,0
5	?	femme	bas	light	2	21,9
6	?	femme	bas	light	3	17,6
7	?	homme	élevé	light	1	14,6
8	?	homme	élevé	light	2	15,3
9	?	homme	élevé	light	3	12,3
10	?	femme	élevé	light	1	16,1
11	?	femme	élevé	light	2	9,3
12	?	femme	élevé	light	3	10,8
13	?	homme	bas	heavy	1	17,6
14	?	homme	bas	heavy	2	18,8
15	?	homme	bas	heavy	3	23,2
16	?	femme	bas	heavy	1	14,8
17	?	femme	bas	heavy	2	10,3
18	?	femme	bas	heavy	3	11,3
19	?	homme	élevé	heavy	1	14,9
20	?	homme	élevé	heavy	2	20,4
21	?	homme	élevé	heavy	3	12,8
22	?	femme	élevé	heavy	1	10,1
23	?	femme	élevé	heavy	2	14,4
24	?	femme	élevé	heavy	3	6,1

3 facteurs A B C à 4 modalités dans un carré latin 4 x4 de 16 essais

1 ID	2 A_Conducteur	3 B_Voiture	4 C_Additif	5 Y consom	6 c5	7 Conducteur2	8 Voiture2	9 Additif2
1	c1	v1	A	21	colonnes	1	1	1
2	c1	v2	B	26	2-3-4	1	2	2
3	c1	v3	D	20	codées	1	3	4
4	c1	v4	C	25	numériques	1	4	3
5	c2	v1	D	23	dans les	2	1	4
6	c2	v2	C	26	colonnes	2	2	3
7	c2	v3	A	20	6-7-8	2	3	1
8	c2	v4	B	27		2	4	2
9	c3	v1	B	15	nécessaire	3	1	2
10	c3	v2	D	13	pour	3	2	4
11	c3	v3	C	16	analyser	3	3	3
12	c3	v4	A	16	avec module	3	4	1
13	c4	v1	C	17	DOE	4	1	3
14	c4	v2	A	15	de STATISTICA	4	2	1
15	c4	v3	B	20		4	3	2
16	c4	v4	D	20		4	4	4

Ex - 7.2 : carré latin 4 x 4

étude 4 additifs essence A – B – C – D facteur principal

réponse : Y = consommation (m/g)

2 facteurs nuisibles : voiture v1– v2 – v3 - v4

conducteur c1– c2 – c3 – c4

plan : carré latin 4 x 4

voiture

conducteur	v1	v2	v3	v4
c1	A / 21	B / 26	D / 20	C / 25
c2	D / 23	C / 26	A / 20	B / 27
c3	B / 15	D / 13	C / 16	A / 16
c4	C / 17	A / 15	B / 20	D / 20

facteur principal / réponse Y

Exemples données : modèles d'analyse de la variance

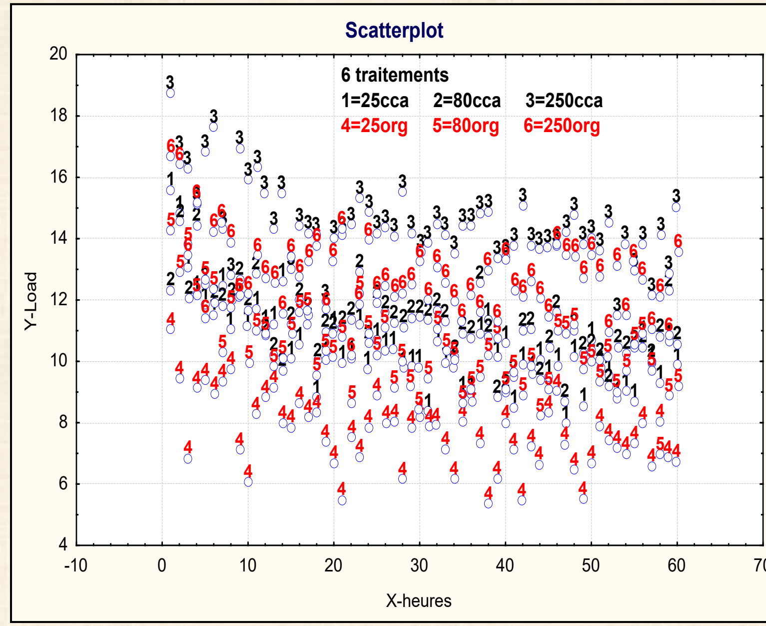
Contexte et données de l'étude CCA « Chromate Copper Arsenate »

Le CCA préserve le bois des dommages causés par l'eau et les insectes.
Le CCA présente des risques pour la santé et l'environnement.
Une nouvelle solution dite « organic » pourrait remplacer le CCA.
Une étude statistique fut conduite pour comparer les deux solutions.
Les données du fichier représentent 360 planches de pin traitées avec
les 2 solutions à 3 niveaux de concentration: 0.25, 0.80 et 2.50 (lbs/pi cu).
Ces niveaux de concentration sont employés pour 3 applications typiques:
bois exposé à l'air, bois de fondation, bois en eau salée.
Les planches traitées furent placées dans des chambres à un vieillissement accéléré durant plusieurs heures.
Chaque heure représente l'équivalent d'une exposition d'une année aux éléments.
Suite à l'application de la solution et du processus de vieillissement, chaque planche fut soumise
à un test de bris à la rupture représenté par la variable « Load ».
Cette variable de réponse est du type « larger the better ».
Les facteurs contrôlés sont : Concentration, Solution et X_Heures est un facteur continu mesuré (covariable).
Objectif de l'expérience : comparer la performance (LOAD) de la solution « organic » avec la solution CCA.
La solution « organic » est-elle aussi bonne que la solution CCA?

1 ID	2 Concentration	3 Conc	4 Solution	5 X_heures	6 trait	7 trait2	8 Y-Load
1	0,25	25	CCA	1,0	25cca	1	15,597
2	0,25	25	CCA	2,0	25cca	1	14,827
3	0,25	25	CCA	3,0	25cca	1	13,037
4	0,25	25	CCA	4,0	25cca	1	12,220
5	0,25	25	CCA	5,0	25cca	1	11,867
6	0,25	25	CCA	6,0	25cca	1	11,520
7	0,25	25	CCA	7,0	25cca	1	12,467
8	0,25	25	CCA	7,9	25cca	1	11,054
9	0,25	25	CCA	9,0	25cca	1	12,162
10	0,25	25	CCA	9,9	25cca	1	11,136

....

353	2,50	250	ORG	53,0	250org	6	13,121
354	2,50	250	ORG	54,0	250org	6	11,492
355	2,50	250	ORG	54,9	250org	6	13,243
356	2,50	250	ORG	55,9	250org	6	12,669
357	2,50	250	ORG	57,0	250org	6	11,031
358	2,50	250	ORG	58,0	250org	6	12,120
359	2,50	250	ORG	59,0	250org	6	10,844
360	2,50	250	ORG	60,1	250org	6	13,569



Influence des facteurs sur Y ?

Exemples données : modèles d'analyse de la variance

Milliken and Johnson, vol 1 p.418 <u>facteurs emboités (nested)</u>

Milliken, G. A., & Johnson, D. E. (1984). *Analysis of messy data:* Vol. I. *Designed experiments.* Van Nostrand Reinhold, Co.
Milliken, G. A., & Johnson, D. E. (1992). *Analysis of messy data*: Vol. 2. *Unreplicated Experiments.* Chapman & Hall.
Milliken, G. A. (2002). *Analysis of messy data:* Vol. 3. *Analysis of covariance*. Chapman & Hall.

A comfort experiment was conducted to study the effects of temperature and gender on a person's comfort.

Researchers were interested only in 3 temperature settings: 65, 70, and 75 deg. Fahrenheit.
Each temperature setting was randomly assigned to 3 of 9 available environmental chambers. (c1, c2, …, c9) .
36 persons (p1, p2, …, p36) 18 males and 18 females were randomly assigned to 9 chambers so that 2 males and 2 females were randomly assigned to each of the 9 chambers.

After the people were subjected to the environmental condition for 3 hours, their comfort was measured.

Y_confort scores : 1 = cold 8 = confortable 15 = hot

modèle

Y_Comfort = Temperature + Gender + Temperature*Gender

+ Chamber(Temperature) + Person(Chamber) + error

influence des facteurs sur Y ?

1 ID	2 Temperature	3 Chamber	4 Gender	5 Person	6 Y_Comfort
1	65	c1	Male	p1	5
2	65	c1	Male	p2	4
3	65	c1	Female	p3	1
4	65	c1	Female	p4	2
5	65	c2	Male	p5	5
6	65	c2	Male	p6	4
7	65	c2	Female	p7	5
8	65	c2	Female	p8	5
9	65	c3	Male	p9	4
10	65	c3	Male	p10	2
11	65	c3	Female	p11	1
12	65	c3	Female	p12	3
13	70	c4	Male	p13	8
14	70	c4	Male	p14	8
15	70	c4	Female	p15	10
16	70	c4	Female	p16	7
17	70	c5	Male	p17	6
18	70	c5	Male	p18	3
19	70	c5	Female	p19	8
20	70	c5	Female	p20	8
21	70	c6	Male	p21	5
22	70	c6	Male	p22	7
23	70	c6	Female	p23	8
24	70	c6	Female	p24	8
25	75	c7	Male	p25	12
26	75	c7	Male	p26	8
27	75	c7	Female	p27	11
28	75	c7	Female	p28	13
29	75	c8	Male	p29	8
30	75	c8	Male	p30	7
31	75	c8	Female	p31	8
32	75	c8	Female	p32	8
33	75	c9	Male	p33	6
34	75	c9	Male	p34	6
35	75	c9	Female	p35	6
36	75	c9	Female	p36	7

Exemples données : modèles d'analyse de la variance

Kutner et all 5 ed. p. 1132
données : évaluation (rang sur 30) de 4 vins (v1 v2 v3 v4) par 6 juges (A B C D E F)
2 analyses possibles : comparer les 4 vins OU comparer les 6 juges
//

1 ID	2 juge	3 ID_vin	4 Y_rang (sur 30)	5 c5	6 juge2	7 Y_v1	8 Y_v2	9 Y_v3	10 Y_v4	11 c11	12 vin	13 Y_A	14 Y_B	15 Y_C	16 Y_D	17 Y_E	18 Y_F
1	A	v1	20	données	A	20	24	28	28	données	v1	20	15	18	26	22	19
2	A	v2	24	désampilées	B	15	18	23	24	désampilées	v2	24	28	29	26	24	21
3	A	v3	28	unité expérimentale	C	18	19	24	23	unité expérimentale	v3	28	23	24	30	28	27
4	A	v4	28	= juge	D	26	26	30	30	= vin	v4	28	24	23	30	26	25
5	B	v1	15		E	22	24	28	26								
6	B	v2	18		F	19	21	27	25								
7	B	v3	23														
8	B	v4	24														
9	C	v1	18														
10	C	v2	19														
11	C	v3	24														
12	C	v4	23														
13	D	v1	26														
14	D	v2	26														
15	D	v3	30														
16	D	v4	30														
17	E	v1	22														
18	E	v2	24														
19	E	v3	28														
20	E	v4	26														
21	F	v1	19														
22	F	v2	21														
23	F	v3	27														
24	F	v4	25														

exemple de mesures répétées

influence des facteurs sur Y ?

Exemples données : modèles d’analyse de la variance

Source STATISTICA

Example data file for repeated measures ANOVA - 3 factors - 6 sujets (s1, s2,...,s6)
1 between factor (INTER) : NOISE values = MEANGFLL WHITE
2 within factor (INTRA) : TIME values = t1, t2, t3
DIAL values = d1, d2
the slowest changing subscript is specyfied first : TIME dans cet exemple
the fastest changing subscript is specify last : DIAL dans cet exemple (intra TIME)

1 ID	2 sujet	3 NOISE	4 dial	5 time	6 Y	7 c7	8 sujet2	9 NOISE	10 Y_t1_d1	11 Y_t1_d2	12 Y_t1_d3	13 Y_t2_d1	14 Y_t2_d2	15 Y_t2_d3	16 Y_t3_d1	17 Y_t3_d2	18 Y_t3_33
1	s1	MEANGFLL	d1	t1	45	données	s1	MEANGFLL	45	53	60	40	52	57	28	37	46
2	s2	MEANGFLL	d1	t1	35	désampilées	s2	MEANGFLL	35	41	50	30	37	47	25	32	41
3	s3	MEANGFLL	d1	t1	60		s3	MEANGFLL	60	65	75	58	54	70	40	47	50
4	s4	WHITE	d1	t1	50		s4	WHITE	50	48	61	25	34	51	16	23	35
5	s5	WHITE	d1	t1	42		s5	WHITE	42	45	55	30	37	43	22	27	37
6	s6	WHITE	d1	t1	56		s6	WHITE	56	60	77	40	39	57	31	29	46
7	s1	MEANGFLL	d2	t1	53												
8	s2	MEANGFLL	d2	t1	41												
9	s3	MEANGFLL	d2	t1	65												
10	s4	WHITE	d2	t1	48												
11	s5	WHITE	d2	t1	45												
12	s6	WHITE	d2	t1	60												
13	s1	MEANGFLL	d3	t1	60												
14	s2	MEANGFLL	d3	t1	50												
15	s3	MEANGFLL	d3	t1	75												
16	s4	WHITE	d3	t1	61												
17	s5	WHITE	d3	t1	55												
18	s6	WHITE	d3	t1	77												
19	s1	MEANGFLL	d1	t2	40												
20	s2	MEANGFLL	d1	t2	30												
21	s3	MEANGFLL	d1	t2	58												
22	s4	WHITE	d1	t2	25												
23	s5	WHITE	d1	t2	30												
24	s6	WHITE	d1	t2	40												
25	s1	MEANGFLL	d2	t2	52												
26	s2	MEANGFLL	d2	t2	37												
27	s3	MEANGFLL	d2	t2	54												
28	s4	WHITE	d2	t2	34												
29	s5	WHITE	d2	t2	37												
30	s6	WHITE	d2	t2	39												
31	s1	MEANGFLL	d3	t2	57												
32	s2	MEANGFLL	d3	t2	47												
33	s3	MEANGFLL	d3	t2	70												
34	s4	WHITE	d3	t2	51												
35	s5	WHITE	d3	t2	43												
36	s6	WHITE	d3	t2	57												
37	s1	MEANGFLL	d1	t3	28												
38	s2	MEANGFLL	d1	t3	25												
39	s3	MEANGFLL	d1	t3	40												
40	s4	WHITE	d1	t3	16												
41	s5	WHITE	d1	t3	22												
42	s6	WHITE	d1	t3	31												
43	s1	MEANGFLL	d2	t3	37												
44	s2	MEANGFLL	d2	t3	32												
45	s3	MEANGFLL	d2	t3	47												
46	s4	WHITE	d2	t3	23												
47	s5	WHITE	d2	t3	27												
48	s6	WHITE	d2	t3	29												
49	s1	MEANGFLL	d3	t3	46												
50	s2	MEANGFLL	d3	t3	41												
51	s3	MEANGFLL	d3	t3	50												
52	s4	WHITE	d3	t3	35												
53	s5	WHITE	d3	t3	37												
54	s6	WHITE	d3	t3	46												

exemple de mesures répétées avec 2 facteurs INTRA (time, dial)
1 facteur INTER (noise)

influence des facteurs sur Y ?

Exemples données : modèles d'analyse de la variance

A study was performed to test 2 new cholesterol drugs A ANB against a control drug.
20 patients with high cholesterol were randomly assigned to each
of 4 treatments (the two experimental drugs, the control, and a placebo).
Each patient's total cholesterol was measured at 6 times during the study:
the first day in April, May, and June in the morning and afternoon.
You are interested in whether either of the new drugs is effective at
lowering cholesterol and in whether time and treatment interact.

1 ID	2 Patient	3 Treatment	4 Time	5 Month	6 AM/PM	7 Days	8 Y	9 c9	10 patient2	11 treatment2	12 Y_Av_AM	13 Y_Av_PM	14 Y_Ma_AM	15 Y_Ma_PM	16 Y_Ju_AM	17 Y_Ju_PM
1	p1	A	April AM	April	AM	0,0	278,0	données	p1	A	278,0	280,0	204,0	208,0	171,3	175,0
2	p1	A	April PM	April	PM	0,5	280,0	désampilées	p2	A	278,0	281,0	195,2	199,0	185,0	189,0
3	p1	A	May AM	May	AM	30,0	204,0		p3	A	276,0	280,0	213,4	219,0	179,0	181,0
4	p1	A	May PM	May	PM	30,5	208,0		p4	A	276,0	281,0	201,3	211,0	183,0	188,9
5	p1	A	June AM	June	AM	61,0	171,3		p5	A	279,0	285,0	188,0	192,0	170,3	174,0
6	p1	A	June PM	June	PM	61,5	175,0		p6	B	266,0	270,0	220,0	224,0	180,0	184,0
7	p2	A	April AM	April	AM	0,0	278,0		p7	B	280,0	284,0	228,0	232,0	200,0	204,0
8	p2	A	April PM	April	PM	0,5	281,0		p8	B	284,0	288,0	233,0	237,0	175,0	179,0
9	p2	A	May AM	May	AM	30,0	195,2		p9	B	273,0	277,0	215,0	219,0	202,0	207,0
10	p2	A	May PM	May	PM	30,5	199,0		p10	B	281,0	285,0	237,0	241,0	199,0	205,0
11	p2	A	June AM	June	AM	61,0	185,0		p11	Control	278,0	282,0	273,0	277,2	281,0	285,0
12	p2	A	June PM	June	PM	61,5	189,0		p12	Control	273,0	277,0	274,0	278,0	280,0	282,0
13	p3	A	April AM	April	AM	0,0	276,0		p13	Control	282,0	285,0	276,0	281,3	279,0	282,0
14	p3	A	April PM	April	PM	0,5	280,0		p14	Control	274,0	277,3	284,5	289,0	274,0	278,0
15	p3	A	May AM	May	AM	30,0	213,4		p15	Control	277,0	280,6	279,1	283,6	284,0	285,0
16	p3	A	May PM	May	PM	30,5	219,0		p16	Placebo	279,0	283,0	278,0	284,0	268,0	272,0
17	p3	A	June AM	June	AM	61,0	179,0		p17	Placebo	277,0	279,0	291,0	291,0	280,0	285,0
18	p3	A	June PM	June	PM	61,5	181,0		p18	Placebo	275,0	279,0	280,0	283,0	281,0	283,0
19	p4	A	April AM	April	AM	0,0	276,0		p19	Placebo	276,0	282,4	277,0	282,0	274,0	279,0
20	p4	A	April PM	April	PM	0,5	281,0		p20	Placebo	282,0	286,0	281,0	285,0	282,0	285,0
21	p4	A	May AM	May	AM	30,0	201,3									
22	p4	A	May PM	May	PM	30,5	211,0									

....

113	p19	Placebo	June AM	June	AM	61,0	274,0
114	p19	Placebo	June PM	June	PM	61,5	279,0
115	p20	Placebo	April AM	April	AM	0,0	282,0
116	p20	Placebo	April PM	April	PM	0,5	286,0
117	p20	Placebo	May AM	May	AM	30,0	281,0
118	p20	Placebo	May PM	May	PM	30,5	285,0
119	p20	Placebo	June AM	June	AM	61,0	282,0
120	p20	Placebo	June PM	June	PM	61,5	285,0

exemple de mesures répétées

influence des facteurs sur Y ?

Exemples données : modèles d'analyse de la variance

4 facteurs : 2 fixés (group, gender) + 2 aléatoires (time, paid)
covariable = X_stress + répétition
2 groups X 2 gender X 3 times X 2 paid X 2 rep = 48 obs.
3 réponses : Y_correct1, Y_correct2, Y_correct3

1 ID	2 A_GROUP	3 B_GENDER	4 C_TIME	5 D_PAID	6 rep	7 X_STRESS	8 Y_CORRECT1	9 Y_CORRECT2	10 Y_CORRECT3
1	EXPERMTL	MALE	BEFORE	NOT_PAID	1	1,41	12	4	6
2	EXPERMTL	MALE	BEFORE	NOT_PAID	2	1,73	3	3	7
3	EXPERMTL	MALE	BEFORE	PAID	1	0,00	7	6	0
4	EXPERMTL	MALE	BEFORE	PAID	2	1,41	11	7	3
5	EXPERMTL	MALE	AFTER_1	NOT_PAID	1	12,83	8	2	7
6	EXPERMTL	MALE	AFTER_1	NOT_PAID	2	2,24	15	1	7
7	EXPERMTL	MALE	AFTER_1	PAID	1	2,24	15	3	3
8	EXPERMTL	MALE	AFTER_1	PAID	2	13,32	11	4	6
9	EXPERMTL	MALE	AFTER_2	NOT_PAID	1	13,74	19	5	4
10	EXPERMTL	MALE	AFTER_2	NOT_PAID	2	13,32	11	6	7
11	EXPERMTL	MALE	AFTER_2	PAID	1	13,00	9	6	5
12	EXPERMTL	MALE	AFTER_2	PAID	2	13,16	10	4	8
13	EXPERMTL	FEMALE	BEFORE	NOT_PAID	1	14,00	8	2	2
14	EXPERMTL	FEMALE	BEFORE	NOT_PAID	2	12,23	6	5	5
15	EXPERMTL	FEMALE	BEFORE	PAID	1	13,22	8	4	4
16	EXPERMTL	FEMALE	AFTER_1	PAID	2	12,23	7	6	0
17	EXPERMTL	FEMALE	AFTER_1	NOT_PAID	1	13,00	5	1	1
18	EXPERMTL	FEMALE	AFTER_1	NOT_PAID	2	13,00	6	3	2
19	EXPERMTL	FEMALE	AFTER_1	PAID	1	16,25	7	5	7
20	EXPERMTL	FEMALE	AFTER_1	PAID	2	14,26	6	6	8
21	EXPERMTL	FEMALE	AFTER_2	NOT_PAID	1	10,00	5	4	7
22	EXPERMTL	FEMALE	AFTER_2	NOT_PAID	2	13,25	7	6	8
23	EXPERMTL	FEMALE	AFTER_2	PAID	1	10,00	8	5	6
24	EXPERMTL	FEMALE	AFTER_2	PAID	2	9,78	6	6	9
25	CONTROL	MALE	BEFORE	NOT_PAID	1	4,25	11	7	5
26	CONTROL	MALE	BEFORE	NOT_PAID	2	4,89	10	8	1
27	CONTROL	MALE	BEFORE	PAID	1	6,26	9	9	5
28	CONTROL	MALE	BEFORE	PAID	2	9,25	8	8	6
29	CONTROL	MALE	AFTER_1	NOT_PAID	1	10,35	4	7	4
30	CONTROL	MALE	AFTER_1	NOT_PAID	2	12,23	5	5	5
31	CONTROL	MALE	AFTER_1	PAID	1	10,25	6	8	1
32	CONTROL	MALE	AFTER_1	PAID	2	12,13	5	6	2
33	CONTROL	MALE	AFTER_2	NOT_PAID	1	4,25	4	5	4
34	CONTROL	MALE	AFTER_2	NOT_PAID	2	4,26	8	8	5
35	CONTROL	MALE	AFTER_2	PAID	1	7,80	12	7	6
36	CONTROL	MALE	AFTER_2	PAID	2	9,40	15	2	7
37	CONTROL	FEMALE	BEFORE	NOT_PAID	1	6,32	14	4	8
38	CONTROL	FEMALE	BEFORE	NOT_PAID	2	3,25	16	5	7
39	CONTROL	FEMALE	BEFORE	PAID	1	4,25	13	4	5
40	CONTROL	FEMALE	BEFORE	PAID	2	5,25	12	2	4
41	CONTROL	FEMALE	AFTER_1	NOT_PAID	1	12,23	16	1	1
42	CONTROL	FEMALE	AFTER_1	NOT_PAID	2	12,25	14	8	2
43	CONTROL	FEMALE	AFTER_1	PAID	1	11,23	10	9	3
44	CONTROL	FEMALE	AFTER_1	PAID	2	10,25	9	8	2
45	CONTROL	FEMALE	AFTER_2	NOT_PAID	1	6,23	8	7	4
46	CONTROL	FEMALE	AFTER_2	NOT_PAID	2	5,25	6	8	5
47	CONTROL	FEMALE	AFTER_2	PAID	1	14,25	7	1	6
48	CONTROL	FEMALE	AFTER_2	PAID	2	13,25	12	2	5

modèle mixte et analyse de covariance

exemple de facteurs aléatoires

influence des facteurs sur Y ?

Modèle ANOVA avec 2 facteurs

CAS : 2 facteurs A, B fixes

A : niveaux a_i $i = 1, 2,..., I$ B : niveaux b_j $j = 1, 2, \ldots, J$

Y_{ijr} r-ème répétition réponse Y chaque combinaison (i, j) = cellule

$r = 1, 2, \ldots, n_{ij} = n > 1$

$N = n\,I\,J$: nombre total d'observations

design = CRD = assignation au hasard des traitements (a_i, b_j) aux unités expérimentales

autres cas : $n_{ij} = 1$ / $n_{ij} > 0$ et inégales / $n_{ij} = 0$ certains (i, j)

Modèle de type MOYENNES DES CELLULES : pas d'effet général

$Y_{ijr} = \mu_{ij} + \varepsilon_{ijr}$ $\varepsilon_{ijr} \sim N(0, \sigma^2)$ modèle moins explicatif

Modèle à type EFFETS .... modèle très explicatif

$Y_{ijr} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijr}$

$\mu = \sum\sum \mu_{ij} / I\,J$ $\mu_{i\cdot} = \sum \mu_{ij} / J$ $\mu_{\cdot j} = \sum \mu_{ij} / I$

$\alpha_i = \mu_{i\cdot} - \mu$ $\sum \alpha_i = 0$ $\beta_j = \mu_{\cdot j} - \mu$ $\sum \beta_j = 0$

$(\alpha\beta)_{ij} = \mu_{ij} - (\mu + \alpha_i + \beta_j) = \mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu$

$\sum (\alpha\beta)_{ij} = 0$ $\sum (\alpha\beta)_{ij} = 0$

Modèle ANOVA avec 2 facteurs

Ajustement du modèle à moyenne de cellules

minimum $Q = \sum\sum (Y_{ijr} - \mu_{ij})^2$

solution: $\widehat{\mu}_{ij} = \overline{Y}_{ij.}$ $\widehat{Y}_{ijr} = \overline{Y}_{ij.}$

résidu : $e_{ijr} = Y_{ijr} - \widehat{Y}_{ij.} = Y_{ijr} - \overline{Y}_{ij.}$

Ajustement du modèle à effets

minimum $Q = \Sigma\Sigma(Y_{ijr} - \mu - \alpha_i - \beta_j - (\alpha\beta)_{ij})^2$

$\Sigma\, \alpha_i = 0$ $\Sigma\beta_j = 0$ $\Sigma(\alpha\beta)_{ij} = 0$ $\Sigma(\alpha\beta)_{ij} = 0$

paramètre	estimateur
μ	$\widehat{\mu} = \overline{Y}_{...}$ (87)
$\alpha_i = \mu_{i.} - \mu$	$\widehat{\alpha}_i = \overline{Y}_{i..} - \overline{Y}_{...}$ (88)
$\beta_j = \mu_{.j} - \mu$	$\widehat{\beta}_j = \overline{Y}_{.j.} - \overline{Y}_{...}$ (89)
$(\alpha\beta)_{ij} = \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu$	$\widehat{(\alpha\beta)}_{ij} = \overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...}$ (90)

Modèle ANOVA avec 2 facteurs

ANOVA : analyse de la variance 2 facteurs avec n répétitions

Source	SS	df	MS	F	p-value
A	SSA	a - 1	MSA = SSA / (a-1)	MSA / MSE	
B	SSB	b - 1	MSB = SSB / (b – 1)	MSB / MSE	
AB	SSAB	(a – 1)(b – 1)	MSAB = SSAB / (a-1)(b-1)	MSAB / MSE	
erreur	SSE	ab(n-1)	MSE = SSE / ab(n-1)	----	
totale	SStotale	abn - 1	----	-----	

$SSA = nb \sum (\overline{Y}_{i..} - \overline{Y}_{...})^2 \qquad SSB = na \sum (\overline{Y}_{.j.} - \overline{Y}_{...})^2$

$SSAB = n \sum \sum (\overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...})^2$

$SSE = \sum \sum \sum (Y_{ijk} - \overline{Y}_{ij.})^2$

$SStotale = \sum \sum \sum (Y_{ijk} - \overline{Y}_{...})^2$

<u>test</u> de H_0 : $(\alpha\beta)_{ij} = 0$ pour tout (i, j) vs H_a : certains $(\alpha\beta)_{ij}$ ne sont pas zéro

on rejette H_0 si $F = MSAB / MSE \geq F(1 - \alpha ; (a-1)*(b-1), (n-1)*ab)$

<u>test</u> de H_0 : $\alpha_1 = \alpha_1 = ... = \alpha_a = 0$ vs H_a : certains α_i ne sont pas zéro.

on rejette H_0 si $F = MSA / MSE \geq F(1 - \alpha ; a-1, (n-1)*ab)$

<u>test</u> de H_0 : $\beta_1 = \beta_1 = ... = \beta_b = 0$ vs H_a : certains β_j ne sont zéro.

on rejette H_0 si $F = MSA / MSE \geq F(1 - \alpha ; b-1, (n-1)*ab)$

Modèle ANOVA avec 2 facteurs

Exemple : étalage

Facteur — modalités

A_hauteur bas / milieu / haut

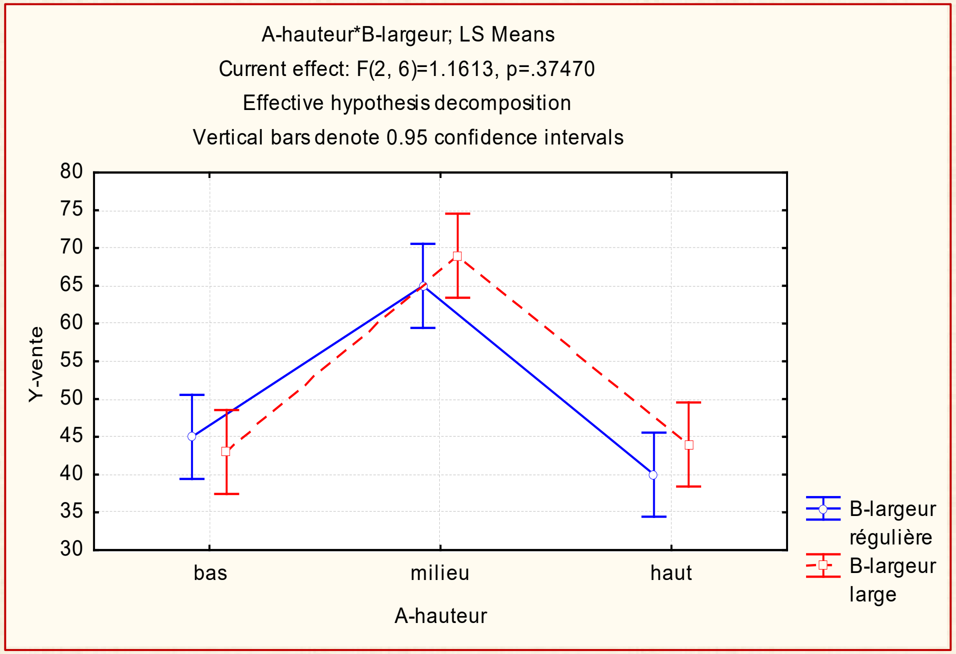
B_largeur régulière / large

Kutner 5 ed. p. 833
influence de 2 facteurs A et B sur les ventes

1 ID	2 unité expérimentale	3 A_hauteur	4 B_largeur	5 rep	6 Y-vente
1	?	bas	régulière	1	47
2	?	bas	régulière	2	43
3	?	bas	large	1	46
4	?	bas	large	2	40
5	?	milieu	régulière	1	62
6	?	milieu	régulière	2	68
7	?	milieu	large	1	67
8	?	milieu	large	2	71
9	?	haut	régulière	1	41
10	?	haut	régulière	2	39
11	?	haut	large	1	42
12	?	haut	large	2	46

ANOVA

	DF	SS	MS	F	p
Intercept	1	31212	31212.0	3020.52	0.00000
A-hauteur	2	1544	772.0	74.71	0.00006
B-largeur	1	12	12.0	1.16	0.32261
A-hauteur*B-largeur	2	24	12.0	1.16	0.37470
Error	6	62	10.3		
Total	11	1642			

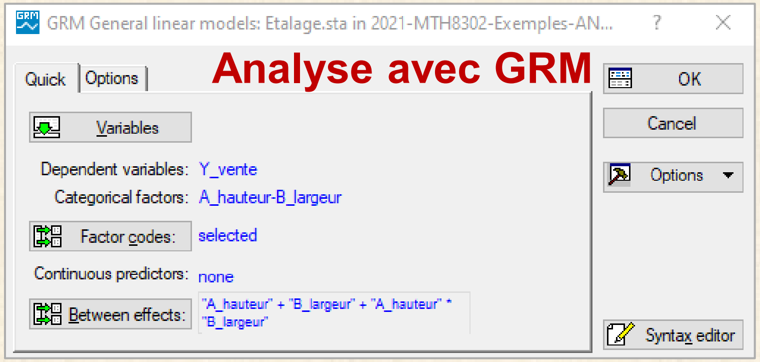


Modèle ANOVA : lien avec la régression

Kutner 5 ed. p. 833
influence de 2 facteurs A et B sur les ventes

1 ID	2 unité expérimentale	3 A_hauteur	4 B_largeur	5 rep	6 Y-vente
1	?	bas	régulière	1	47
2	?	bas	régulière	2	43
3	?	bas	large	1	46
4	?	bas	large	2	40
5	?	milieu	régulière	1	62
6	?	milieu	régulière	2	68
7	?	milieu	large	1	67
8	?	milieu	large	2	71
9	?	haut	régulière	1	41
10	?	haut	régulière	2	39
11	?	haut	large	1	42
12	?	haut	large	2	46

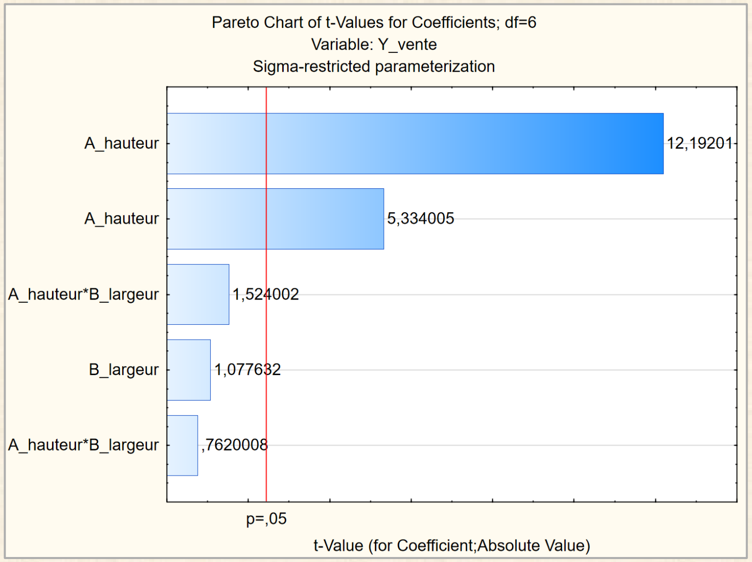
Analyse avec GRM

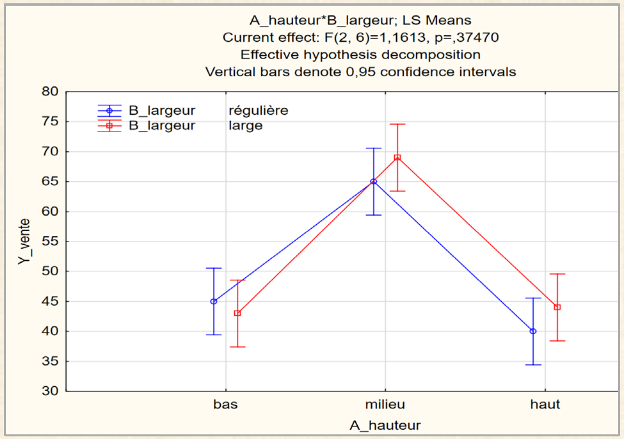


effet	Column	Variable	Level of Variable	versus Level	Variable	Level of Variable	versus Level
Intercept	1						
A_hauteur	2	A_hauteur	bas	haut			
A_hauteur	3	A_hauteur	milieu	haut			
B_largeur	4	B_largeur	régulière	large			
A_hauteur*B_largeur	5	A_hauteur	bas	haut	B_largeur	régulière	large
A_hauteur*B_largeur	6	A_hauteur	milieu	haut	B_largeur	régulière	large

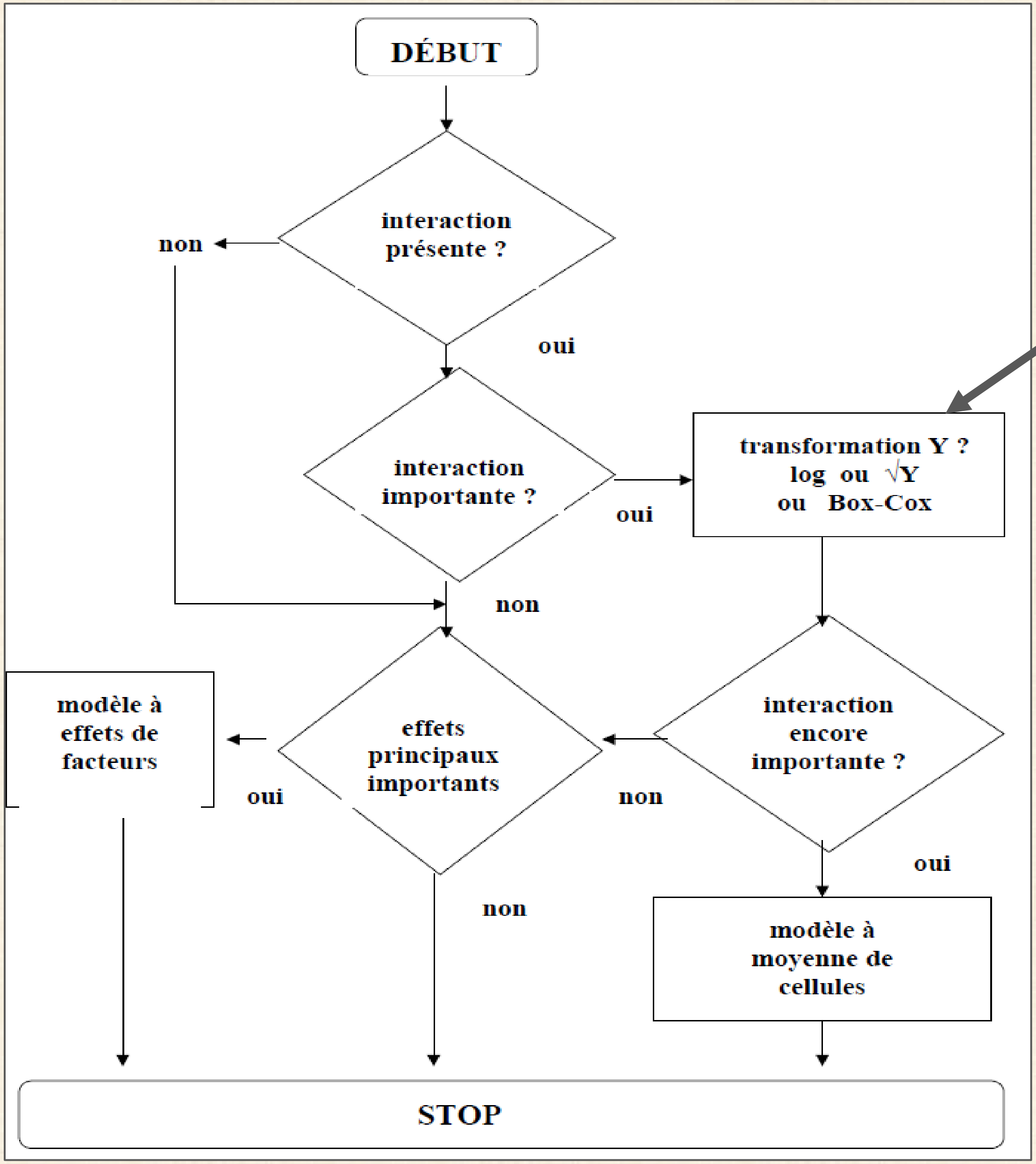
effet	Level of Effect	Column	Y_vente Param.	Y_vente Std.Err	Y_vente t	Y_vente p
Intercept		1	51,00000	0,927961	54,95922	0,000000
A_hauteur	bas	2	-7,00000	1,312335	-5,33401	0,001771
A_hauteur	milieu	3	16,00000	1,312335	12,19201	0,000019
B_largeur	régulière	4	-1,00000	0,927961	-1,07763	0,322605
A_hauteur*B_largeur	1	5	2,00000	1,312335	1,52400	0,178345
A_hauteur*B_largeur	2	6	-1,00000	1,312335	-0,76200	0,474937

effet	df	Y_vente SS	Y_vente MS	Y_vente F	Y_vente p
Intercept	1	31212,00	31212,00	3020,516	0,000000
A_hauteur	2	1544,00	772,00	74,710	0,000058
B_largeur	1	12,00	12,00	1,161	0,322605
A_hauteur*B_largeur	2	24,00	12,00	1,161	0,374697
Error	6	62,00	10,33		
Total	11	1642,00			





Stratégie pour l'étude de 2 facteurs (Kutner & all p. 848)



peut faire disparaitre l'interaction

interprétation simplifiée :

effets principaux seulement

INTERACTION significative ? importante (« grande ») ? transformation ?

exemple : Y : durée (minute) apprentissage tâche
2 facteurs : genre (homme, femme) âge (jeune, moyen, agé)
3 versions de la réponse :
Y1 aucune interaction Y2 interaction importante (forte) Y3 interaction faible

données

exemples : Y1 aucune interaction Y2 interaction importante Y3 interaction faible

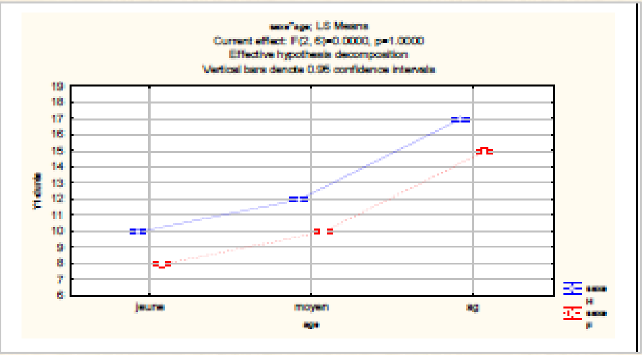
	sexe	age	Y1 durée	Y2 durée	Y3 durée		sexe	age	Y1 moy	Y 2 moy	Y3 moy
1	H	jeune	9.95	8.95	9.75		H	jeune	10	9	9.8
2	H	moyen	11.95	11.95	11.95		H	moyen	12	12	12.0
3	H	agé	16.95	17.95	17.15		H	agé	17	18	17.2
4	F	jeune	7.95	8.95	8.15		F	jeune	8	9	7.8
5	F	moyen	9.95	9.95	9.95		F	moyen	10	10	10.0
6	F	agé	14.95	13.95	14.75		F	agé	15	14	15.2
7	H	jeune	10.05	9.05	9.85						
8	H	moyen	12.05	12.05	12.05		H		13	13	13
9	H	agé	17.05	18.05	17.25		F		11	11	11
10	F	jeune	8.05	9.05	8.25			jeune	9	9	9
11	F	moyen	10.05	10.05	10.05			moyen	11	11	11
12	F	agé	15.05	14.05	14.85			agé	16	16	16
							gen		12	12	12

$$\text{inter } Y = \bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}$$

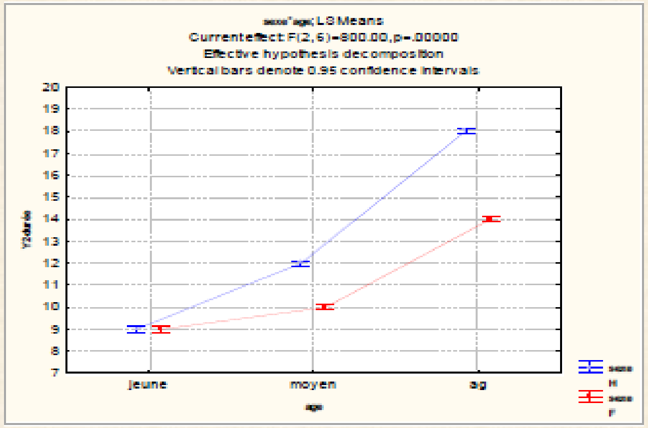
	sexe	age	Y1 moy	Y2 moy	Y3 moy		inter Y1	inter Y2	inter Y3
1	H	jeune	10	9	9.8		0	-1	-0.2
2	H	moyen	12	12	12.0		0	0	0
3	H	agé	17	18	17.2		0	1	0.2
4	F	jeune	8	9	8.2		0	1	-0.2
5	F	moyen	10	10	10.0		0	0	0
6	F	agé	15	14	14.8		0	-1	0.2

INTERACTION significative ? importante (« grande ») ? transformation ?

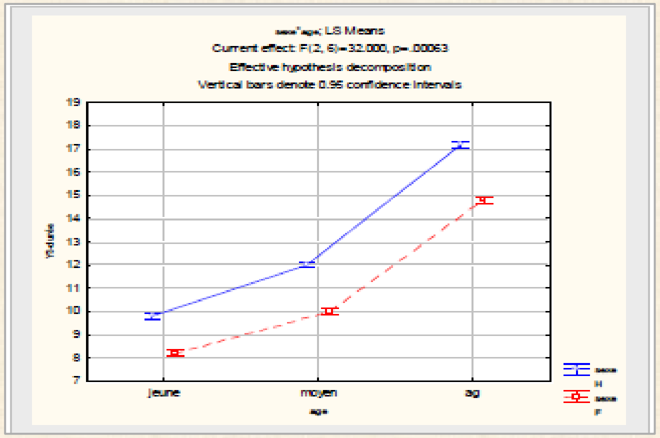
ANOVA : Y1					
Source	DF	SS	MS	F	p-value
Intercept	1	1728.00	1728	345600	0.0000
sexe	1	12.00	12	2400	0.0000
age	2	104.00	52	10400	0.0000
sexe*age	2	0.00	0	0	1.0000
Error	6	0.03	0.005		
Total	11	116.030			



ANOVA : Y2					
Source	DF	SS	MS	F	p-value
Intercept	1	1728	1728	345600	0.00000
sexe	1	12	12	2400	0.00000
age	2	104	52	10400	0.00000
sexe*age	2	8	4	800	0.00000
Error	6	0.03	0.005		
Total	11	124.03			



ANOVA : Y 3					
	DF	SS	MS	F	p-value
Intercept	1	1728	1728	345600	0.00000
sexe	1	12	12	2400	0.00000
age	2	104	52	10400	0.00000
sexe*age	2	0.32	0.16	32	0.00063
Error	6	0.03	0.005		
Total	11	116.35			



Élimination d'interaction par transformation: 2 cas spéciaux

modèle multiplicatif $\mu_{ij} = \mu\, \alpha_i\, \beta_j$

$$\log(\mu_{ij}) = \log(\mu) + \log(\alpha_i) + \log(\beta_j)$$

$$\mu'_{ij} = \mu' + \alpha'_i + \beta'_j$$

Avec un modèle multiplicatif, une analyse de Y détecterait la présence d'interaction.
Une analyse de $Y' = \log(Y)$ donne un modèle additif sans interaction.

modèle racine $\mu_{ij} = (\sqrt{\alpha_i} + \sqrt{\beta_j})^2$

$= \alpha_i + \beta_j + 2\sqrt{\alpha_i}\sqrt{\beta_j}$: modèle avec interaction

$\mu'_{ij} = (\mu_{ij})^{0.5} = \sqrt{\alpha_i} + \sqrt{\beta_j} = \alpha'_i + \beta'_j$

modèle sans interaction

Une analyse de $Y' = \sqrt{Y}$ donne un modèle additif sans interaction.

Transformation de Box-Cox : $Y' = Y^{\lambda}$ peut « éliminer » une interaction

INTERACTION significative ? importante (« grande ») ? transformation ?

Analyse des effets avec un modèle additif : facteurs sans interaction sur Y

$$E(Y_{ijk}) = \mu_{ij} = \mu + \alpha_{i.} + \beta_{.j}$$

modèle ajusté : $\widehat{Y}_{ij} = \widehat{\mu} + \widehat{\alpha}_{i.} + \widehat{\beta}_{.j}$

$$= \widehat{\mu} + (\widehat{\mu}_{i.} - \widehat{\mu}_{..}) + (\widehat{\mu}_{.j} - \widehat{\mu}_{..})$$

$$= \overline{Y}_{...} + (\overline{Y}_{i..} - \overline{Y}_{...}) + (\overline{Y}_{.j.} - \overline{Y}_{...})$$

contraste facteur A: $L = \sum c_i \mu_i \qquad \sum c_i = 0$

$$\widehat{L} = \sum c_i \widehat{\mu}_{i.} = \sum c_i \overline{Y}_{i..}$$

$$s^2(\widehat{L}) = (MSE / bn) \sum c_i^2$$

$(\widehat{L} - L) / s(\widehat{L}) \sim$ Student avec $(n-1)*ab$ degrés de liberté

Intervalle confiance L : $\widehat{L} \pm t(1-\alpha\,;(n-1)*ab)$

comparaison pairée : $D = \mu_{i..} - \mu_{i'.} \qquad i \neq i'$

procedure de Tukey : $H_0: D = 0$ vs $H_a: D \neq 0$

rejet de H_0 si $|\widehat{D}| = |\overline{Y}_{i..} - \overline{Y}_{i'..}| > (MSE/bn)^{0.5}\, q(1-\alpha\,;a\,;(n-1)*ab)$

$q(1-\alpha; a; (n-1)*ab)$: "studentized range"

<u>autres procedures</u> : Bonferronni, Scheffé

<u>Aussi</u> : formules analogues pour le facteur B

INTERACTION significative ? importante (« grande ») ? transformation ?

données

exemples : Y1 aucune interaction Y2 interaction importante Y3 interaction faible

	sexe	age	Y1 durée	Y2 durée	Y3 durée		sexe	age	Y1 moy	Y 2 moy	Y3 moy
1	H	jeune	9.95	8.95	9.75		H	jeune	10	9	9.8
2	H	moyen	11.95	11.95	11.95		H	moyen	12	12	12.0
3	H	agé	16.95	17.95	17.15		H	agé	17	18	17.2
4	F	jeune	7.95	8.95	8.15		F	jeune	8	9	7.8
5	F	moyen	9.95	9.95	9.95		F	moyen	10	10	10.0
6	F	agé	14.95	13.95	14.75		F	agé	15	14	15.2
7	H	jeune	10.05	9.05	9.85						
8	H	moyen	12.05	12.05	12.05		H		13	13	13
9	H	agé	17.05	18.05	17.25		F		11	11	11
10	F	jeune	8.05	9.05	8.25			jeune	9	9	9
11	F	moyen	10.05	10.05	10.05			moyen	11	11	11
12	F	agé	15.05	14.05	14.85			agé	16	16	16
							gen		12	12	12

Tukey HSD test; variable Y1-durée
Approximate Probabilities for Post Hoc Tests
Error: Between MS = .00500, df = 6

	sexe	age	{1} 10	{2} 12	{3} 17	{4} 8	{5} 10	{6} 15
1	H	jeune		0.0002	0.0002	0.0002	1.0000	0.0002
2	H	moyen	0.0002		0.0002	0.0002	0.0002	0.0002
3	H	agé	0.0002	0.0002		0.0002	0.0002	0.0002
4	F	jeune	0.0002	0.0002	0.0002		0.0002	0.0002
5	F	moyen	1.0000	0.0002	0.0002	0.0002		0.0002
6	F	agé	0.0002	0.0002	0.0002	0.0002	0.0002	

INTERACTION significative ? importante (« grande ») ? transformation ?

Analyse des effets en présence d'interactions importantes

Si on ne peut éliminer une interaction importante entre 2 facteurs par une transformation (logarithmique, racine) alors on peut analyser les effets des facteurs avec le **modèle à moyenne de cellules.**

données

exemples :	Y1 aucune interaction				Y2 interaction importante				Y3 interaction faible		
	sexe	age	Y1 durée	Y2 durée	Y3 durée		sexe	age	Y1 moy	Y 2 moy	Y3 moy
1	H	jeune	9.95	8.95	9.75		H	jeune	10	9	9.8
2	H	moyen	11.95	11.95	11.95		H	moyen	12	12	12.0
3	H	agé	16.95	17.95	17.15		H	agé	17	18	17.2
4	F	jeune	7.95	8.95	8.15		F	jeune	8	9	7.8
5	F	moyen	9.95	9.95	9.95		F	moyen	10	10	10.0
6	F	agé	14.95	13.95	14.75		F	agé	15	14	15.2
7	H	jeune	10.05	9.05	9.85						
8	H	moyen	12.05	12.05	12.05		H		13	13	13
9	H	agé	17.05	18.05	17.25		F		11	11	11
10	F	jeune	8.05	9.05	8.25			jeune	9	9	9
11	F	moyen	10.05	10.05	10.05			moyen	11	11	11
12	F	agé	15.05	14.05	14.85			agé	16	16	16
							gen		12	12	12

Scheffe test; variable Y2-durée Probabilities for Post Hoc Tests
Error: Between MS = .00500, df = 6

	sexe	age	{1} 9	{2 12	{3} 18	{4} 9	{5} 10	{6} 14
1	H	jeune		0.0000	0.0000	1.0000	0.0002	0.0000
2	H	moyen	0.0000		0.0000	0.0000	0.0000	0.0000
3	H	agé	0.0000	0.0000		0.0000	0.0000	0.0000
4	F	jeune	1.0000	0.0000	0.0000		0.0002	0.0000
5	F	moyen	0.0002	0.0000	0.0000	0.0002		0.0000
6	F	agé	0.0000	0.0000	0.0000	0.0000	0.0000	

« Pooling » lorsque l'interaction n'est pas significative

Avec 2 facteurs on postule toujours pour commencer un modèle avec un effet d'interaction (voir eq. (78)). Si à l'analyse on constate que l'interaction n'est pas significative, on peut réviser le modèle et postuler un modèle sans interaction c.-à-d. sans le terme $(\alpha\beta)_{ij}$ dans l'équation (78). Cette opération peut se faire à condition que le ratio F soit petit, disons < 2.

conséquences

SSE (modèle révisé) = SSE (modèle complet) + SSAB

df (modèle révisé) = (a – 1) (b-1) + (n - 1)ab = nab – a – b – 1

MSE (modèle révisé) = SSE (modèle révisé) / df (modèle révisé)

En général, le MSE (modèle révisé) devrait être « voisin » du MSE (modèle complet).

Les résultats des tests de signification des effets principaux des facteurs ne devraient pas changer suite à cette opération.

CAS où le nombre de répétition n = 1

modèle : $Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$

ANOVA : 2 facteurs avec n = 1

Source	SS	df	MS	F	p-value
A	SSA	a-1	MSA = SSA / (a -1)	MSA / MSAB	
B	SSB	b-1	MSB = SSB / (b-1)	MSB / MSAB	
erreur	SSAB	(a-1)(b-1)	MSAB = SSAB / (a-1)(b-1)	------	
totale	SStotale	ab - 1	------	------	

$SSE = \sum\sum\sum (Y_{ijk} - \overline{Y}_{ij.})^2$ n'existe pas

SSAB estime l'erreur expérimentale

test d'hypothèse de la nullité du facteur A est basé sur
F = SSA /SSAB qui suit une loi F (a – 1, (a-1)(b-1))

résultat analogue s'applique pour le facteur B

CAS où le nombre de répétition n = 1

modèle : $Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$

test de Tukey: détecter présence d'une interaction si n = 1

remarque: ne pas confondre avec la procédure HSD de Tukey

Il est possible de tester la présence d'une interaction avec n = 1

On suppose que l'interaction est de la forme d'un produit comme dans un polynôme :

$$(\alpha\beta)_{ij} = D\, \alpha_i \beta_j$$

où D est une constante. L'estimateur de D est

$$\hat{D} = \left[\sum\sum \hat{\alpha}_i \hat{\beta}_j Y_{ij}\right] / \left[\sum \hat{\alpha}_i^2 \sum \hat{\beta}_j^2\right]$$

$$= \left[\sum\sum (\overline{Y}_{i.} - \overline{Y}_{..})(\overline{Y}_{.j} - \overline{Y}_{..}) Y_{ij}\right] / C$$

où $$C = \sum (\overline{Y}_{i.} - \overline{Y}_{..})^2 \sum (\overline{Y}_{.j} - \overline{Y}_{..})^2$$

La somme de carrés associée à cette forme d'interaction est:

$$SSAB^* = \left[\sum\sum (\overline{Y}_{i.} - \overline{Y}_{..})(\overline{Y}_{.j} - \overline{Y}_{..}) Y_{ij}\right]^2 / C$$

La somme de carrés résiduelle SSresid est :

$$SSresid = SStotale - SSA - SSB - SSAB^*$$

On peut monter que $F^* = SSAB^* / (SSresid / (ab - a - b))$

suit une loi de fisher-Snedecor $F(1, ab - a - b)$

test de $H_0 : D = 0$ vs $Ha : D \neq 0$

on rejette H_0 si $F^* > F(1 - \alpha ; 1, ab - a - b)$

Test de Tukey : détection interaction avec n = 1

données

exemples :	Y1 aucune interaction			Y2 interaction importante				Y3 interaction faible			
	sexe	age	Y1 durée	Y2 durée	Y3 durée		sexe	age	Y1 moy	Y 2 moy	Y3 moy
1	H	jeune	9.95	8.95	9.75		H	jeune	10	9	9.8
2	H	moyen	11.95	11.95	11.95		H	moyen	12	12	12.0
3	H	agé	16.95	17.95	17.15		H	agé	17	18	17.2
4	F	jeune	7.95	8.95	8.15		F	jeune	8	9	7.8
5	F	moyen	9.95	9.95	9.95		F	moyen	10	10	10.0
6	F	agé	14.95	13.95	14.75		F	agé	15	14	15.2
7	H	jeune	10.05	9.05	9.85						
8	H	moyen	12.05	12.05	12.05		H		13	13	13
9	H	agé	17.05	18.05	17.25		F		11	11	11
10	F	jeune	8.05	9.05	8.25			jeune	9	9	9
11	F	moyen	10.05	10.05	10.05			moyen	11	11	11
12	F	agé	15.05	14.05	14.85			agé	16	16	16
							gen		12	12	12

Exemple : d'apprentissage Y = Y2
avec les observations 1 à 6
avec les observations 7 à 12

17 c17	18 sexe	19 age	20 Y2_durée	21 c22	22 sexe	23 age	24 Y2_durée
1à6	H	jeune	8,95	7à12	H	jeune	9,05
	H	moyen	11,95		H	moyen	12,05
	H	agé	17,95		H	agé	18,05
	F	jeune	8,95		F	jeune	9,05
	F	moyen	9,95		F	moyen	10,05
	F	agé	13,95		F	agé	14,05

$\overline{Y}_{..} = 11.95$ $\qquad \overline{Y}_{1.}\text{(homme)} = 12.95$ $\qquad \overline{Y}_{2.}\text{(femme)} = 10.95$

$\overline{Y}_{.1}\text{(jeune)} = 8.95$ $\qquad \overline{Y}_{.2}\text{(moyen)} = 10.95$ $\qquad \overline{Y}_{.3}\text{(agé)} = 15.95$

$\sum\sum(\overline{Y}_{i.} - \overline{Y}_{..})^2 = 2$ $\qquad \sum\sum(\overline{Y}_{.j} - \overline{Y}_{..})^2 = 26$

$\sum\sum(\overline{Y}_{i.} - \overline{Y}_{..})(\overline{Y}_{.j} - \overline{Y}_{..})\,Y_{ij} = -22$

$\hat{D} = -22$ $\qquad SSAB^* = (-22)^2 / 2*26 = 9{,}31$

SSresid = 61.95 - 6 - 52 - 9,31 = 3,31 $\qquad$ F* = 9,31 / 3.31 = 2,81

H_0 rejetée.

Cas où le nombre de répétitions n_{ij} sont inégales mais > 0

Études observationnelles et, occasionnellement dans les études expérimentales, les tailles échantillonnales n_{ij} sont généralement inégales

Conséquence : l'orthogonalité n'est plus présente décomposition unique des sommes de carrés de l'ANOVA n'existe plus : l'adition des sommes de carrés des effets (principaux et interactions) n'est plus égale à la somme totale des carrés; unicité décomposition n'est plus présente

Situation analogue : en régression où l'on a rarement l'orthogonalité dans la structure des données. Les tailles inégales ont pour conséquence que l'unicité du modèle n'est plus assurée.

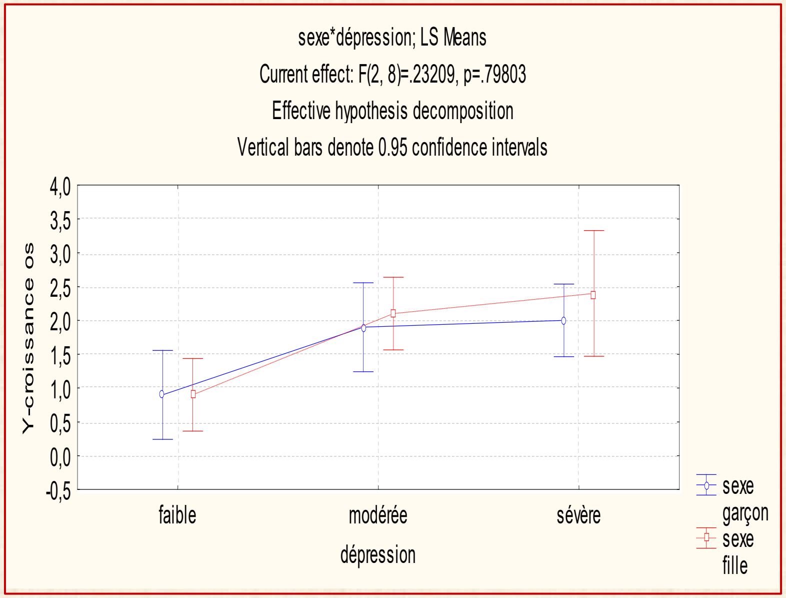
Solution au problème: analyse avec modèle linéaire général de régression avec des variables indicatrices (codage à effet) pour les variables catégoriques

Cas où le nombre de répétitions n_{ij} sont inégales mais > 0

Exemple : données sur la croissance des os (Y)
Kutner et all 5 ed. p. 954
facteurs : sexe enfant (garçon, fille)
dépression (sévère, modérée, faible)

données déséquilibrées
n_{ij} = 3 ou 2 ou 1
selon le cas

1 sexe	2 dépression	3 rep	4 Y-croissance os	5 new	6 X1_sexe	7 X2_depres	8 X3_depres	9 X1_sexe*X2_depres	10 X1_sexe*X3_depres
garçon	sévère	1	1,4		1	1	0	1	0
garçon	sévère	2	2,4		1	1	0	1	0
garçon	sévère	3	2,2		1	1	0	1	0
garçon	modérée	1	2,1		1	0	1	0	1
garçon	modérée	2	1,7		1	0	1	0	1
garçon	faible	1	0,7		1	-1	-1	-1	-1
garçon	faible	2	1,1		1	-1	-1	-1	-1
fille	sévère	1	2,4		-1	1	0	-1	0
fille	modérée	1	2,5		-1	0	1	0	-1
fille	modérée	2	1,8		-1	0	1	0	-1
fille	modérée	3	2,0		-1	0	1	0	-1
fille	faible	1	0,5		-1	-1	-1	1	1
fille	faible	2	0,9		-1	-1	-1	1	1
fille	faible	3	1,3		-1	-1	-1	1	1



Cas où le nombre de répétitions n_{ij} sont inégales mais > 0

Exemple : données sur la croissance des os (Y)   Kutner et all 5 ed. p. 954
facteurs : sexe enfant (garçon, fille)   dépression (sévère, modérée, faible)

$X1 = 1$ si sexe = garçon $X1 = -1$ si sexe = fille
$X2 = 1$ si dépression = sévère $X2 = 0$ si modérée $X2 = -1$ si faible
$X3 = 1$ si dépression = modérée $X3 = 0$ si sévère $X2 = -1$ si faible

codage à effets

Le modèle à effets principaux et d'interaction est :

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

$i = 1, 2$ $j = 1, 2, 3$ $k = 1, 2, \ldots, n_{ij}$
$n_{11} = 3$ $n_{12} = 2$ $n_{13} = 2$
$n_{21} = 1$ $n_{22} = 3$ $n_{23} = 3$

Le modèle équivalent (modèle complet) écrit avec les variables indicatrices est:

$$\text{MC}: Y_{ijk} = \mu + \alpha_1 X_{ijk1} + \beta_1 X_{ijk2} + \beta_2 X_{ijk3} + (\alpha\beta)_{11} X_{ijk1} X_{ijk2} + (\alpha\beta)_{12} X_{ijk1} X_{ijk3} + \varepsilon_{ijk}$$

La procédure pour l'examen des effets de A (sexe), de B (dépression) et de l'interaction AB est basée sur l'ajustement de 4 modèles de régression : MC, MR1, MR2, MR3:

1. test de la nullité des effets d'interaction $(\alpha\beta)_{11}$ et de $(\alpha\beta)_{12}$

$$H_1: (\alpha\beta)_{11} = (\alpha\beta)_{12} = 0 \quad \text{vs} \quad H_a: \text{non } H_0$$

$$\text{MC} + \text{H1}: \quad \text{MR1}: Y_{ijk} = \mu + \alpha_1 X_{ijk1} + \beta_1 X_{ijk2} + \beta_2 X_{ijk3} + \varepsilon_{ijk} \qquad (130)$$

2. test de la nullité de l'effet principal du facteur A

$$H_2: \alpha_1 = 0 \quad \text{vs} \quad H_a: \text{non } H_0$$

$$\text{MC} + \text{H2}: \quad \text{MR2}: Y_{ijk} = \mu + \beta_1 X_{ijk2} + \beta_2 X_{ijk3} + (\alpha\beta)_{11} X_{ijk1} X_{ijk2} + (\alpha\beta)_{12} X_{ijk1} X_{ijk3} + \varepsilon_{ijk} \qquad (131)$$

3. test de la nullité de l'effet principal du facteur B

$$H_3: \beta_1 = \beta_2 = 0 \quad \text{vs} \quad H_a: \text{non } H_0$$

$$\text{MC} + \text{H3}: \quad \text{MR3}: Y_{ijk} = \mu + \alpha_1 X_{ijk1} + (\alpha\beta)_{11} X_{ijk1} X_{ijk2} + (\alpha\beta)_{12} X_{ijk1} X_{ijk3} + \varepsilon_{ijk} \qquad (132)$$

Cas où le nombre de répétitions n_{ij} sont inégales mais > 0

<u>Exemple</u> : données sur la croissance des os (Y) Kutner et all 5 ed. p. 954
facteurs : sexe enfant (garçon, fille) dépression (sévère, modérée, faible)

MR1, MR2, MR3 constituent des modèles réduits

contiennent moins de paramètres que le modèle complet MC

tests d'hypothèses H_1, H_2, H_3 réalisés par un test F basé

sur la réduction de la somme de carrés de l'erreur(résiduelle)

SSE calculée sous le modèle complet SSE(C) et la somme

de carrés de l'erreur SSE calculée sous le modèle réduit SSE(R).

$$F^* = [\,(SSE(R) - SSE(C)) / (df(R) - df(C))\,] / [\,SSE(C) / df(C)\,]$$

$$= [\text{delta } SSE / \text{delta } (df)\,] / MSE(C)$$

Le ratio F* suit une loi F (df(R) - df(C) , df(C))

Cas où le nombre de répétitions n_{ij} sont inégales mais > 0

Exemple : données sur la croissance des os (Y)   Kutner et all 5 ed. p. 954
facteurs :   sexe enfant (garçon, fille)   dépression (sévère, modérée, faible)

résultats de ces analyses de modélisation

modèle	constante	α_1	β_1	β_2	$(\alpha\beta)_{12}$	$(\alpha\beta)_{12}$
MC	1,70	0,10	0,10	0,50	0,30	0,0
MR1	1,68	0,086	0,467	0,327	0	0
MR2	1,69	0	0,444	0,328	- 0,067	- 0,017
MR3	1,63	0,019	0	0	0,067	- 0,193

SSTO = 5.77 avec 13 degrés de liberté

modèle	SSmodèle	SSE	df(mod)	df(C)	R^2	R^2 ajusté
MC	4,47	1,30	5	8	0,77	0,63
MR1	4,40	1,38	3	10	0,76	0,69
MR2	4,35	1,42	4	9	0,75	0,64
MR3	0,28	5,49	3	10	0,55	0,35

Exemple : données sur la croissance des os (Y)   Kutner et all 5 ed. p. 954
facteurs : sexe enfant (garçon, fille)   dépression (sévère, modérée, faible)

test	delta SSE / delta (df)	MSE	Ratio F	p value
H_1	0,037 / 2	1,3 / 8	0,23	0,80
H_2	0,120 / 1	1,3 / 8	0,74	0,41
H_3	2,09 / 2	1,3 / 8	12,89	0,003

H_3 est rejetée: facteur dépression est significatif

H_2 n'est pas rejetée: facteur sexe n'est pas significatif

H_1 n'est pas rejetée: il n'y a pas d'effet d'interaction

ANOVA

	df	SS	MS	F	p-value
Intercept	1	34.680	34.680	213.42	0.0000
sexe	1	0.120	0.120	0.74	0.4152
dépression	2	4.190	2.095	12.89	0.0031
sexe*dépression	2	0.075	0.038	0.23	0.7980
Error	8	1.300	0.162		
Total	13	5.774			

FACTEURS BLOCS

- **Blocage de essais** : méthode pour traiter facteurs «nuisibles» aussi appelés facteurs «secondaires»
- **Facteurs nuisibles typiques**
  lots de matière première, unités expérimentales hétérogènes, opérateurs, équipement, temps, …
- **Facteurs « nuisibles » = facteurs secondaires**
  - effet probable sur la réponse Y
  - ne présentent aucun intérêt en soi
  - non liés aux objectifs (facteurs primaires) de l'expérience
- **Stratégie** : contrôler / minimiser l'impact des facteurs nuisibles
- **Beaucoup d'expériences industrielles implique le blocage**
- **Si absence de contrôle** des facteurs nuisibles (blocage) :
  peut conduire à ne pas détecter les effets importants
  des facteurs primaires

Traitement des facteurs nuisibles

FACTEURS NUISIBLES		
connu	contrôle	méthode
oui	oui	1. bloquer les essais
oui	mesurable	2. analyse de covariance : enlever effet sur la réponse
non	non	3. randomiser les essais

« block what you can and randomize what you cannot »

Sir Ronald A. Fisher

<u>blocs</u> peuvent être définis par plusieurs facteurs nuisibles

<u>exemples</u> plans en carrés latins (2 facteurs nuisible)
plans Gréco-Latins (3 facteurs nuisibles)
plans blocs incomplets équilibrés (PBIBD)

Exemple: test de dureté Rockwell

pointe	matériau type m1	m2	m3	m4
type A	7.69	10.77	15.75	12.82
B	10.61	9.81	2.03	12.22
C	12.22	13.02	13.55	15.36
D	13.18	14.65	14.96	17.15

BLOCS : matière première

- Objectif : comparer 4 types de pointe

  différences entre A B C D?

- type pointe : facteur primaire de l'expérience
- comparaison des 4 types de pointes faite avec plusieurs type de matériau = unités expérimentales

  unités expérimentales hétérogènes

  but : étendre les conclusions
- matériau : 2ième facteur

  facteur secondaire

  formation de blocs d'unités

  BLOC facteur aléatoire (échantillonnage)
- analyse :

  éliminer influence facteur bloc

  pour comparer les pointes

Exemple : carré latin 4 x 4

étude de 4 additifs essence A - B - C - D facteur principal

réponse : Y = consommation (m/g)

2 facteurs nuisibles : voiture v1– v2 – v3 - v4

conducteur c1– c2 – c3 – c4

plan : carré latin 4 x 4

	voiture			
conducteur	v1	v2	v3	v4
c1	A / 21	B / 26	D / 20	C / 25
c2	D / 23	C / 26	A / 20	B / 27
c3	B / 15	D / 13	C / 16	A / 16
c4	C / 17	A / 15	B / 20	D / 20

facteur principal / réponse y

<u>propriétés du plan carré latin 4 x 4</u>

- seulement 16 des 4x4x4 = 4^3 = 64 combinaisons

  Additif x Voiture x Cond

- plan fractionnaire : 4^{3-1}
- chaque ligne (conducteur) - 4 additifs présents
- chaque colonne (voiture) - 4 additifs présents

Exemple : réaction chimique

2 facteurs primaires : A_concentration (- +) B_catalyseur (- +)

1 facteur bloc : C_laboratoire 1 - 2 - 3

Y: rendement réaction

A	B	Y	C
-1	-1	28	1
1	-1	36	1
-1	1	18	1
1	1	31	1
-1	-1	25	2
1	-1	32	2
-1	1	19	2
1	1	30	2
-1	-1	27	3
1	-1	32	3
-1	1	23	3
1	1	29	3

apparence : plan 2^2 n = 3

quoi de différent?

- randomisation des essais chaque laboratoire

- pas d'effet d'interaction AC AB

- modèle : A + B + AB + C

facteur C

effet principal seulement

pas d'interaction AC BC

Analyse plan en bloc complet avec 1 facteur principal

définition : plan en bloc avec 1 facteur primaire (principal)

- facteur primaire A avec a modalités : $i = 1, 2, \ldots, a$
- facteur bloc (secondaire) B avec b modalités : $j = 1, 2, \ldots, b$
- randomisation des essais à l'intérieur de chaque bloc
- en général : pas de répétitions dans chaque cellule (i, j)
- modèle sans interaction entre A et B

Résumé : comme un plan avec 2 facteurs mais pas d'interaction

modèle

réponse = effet général + effet traitement + effet bloc + erreur

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \qquad i = 1,2,\ldots, a \quad j = 1,2,\ldots, b$$

tableau des données

A \ B	1	2	3	…	b	moyennes
1	y_{11}	y_{12}	y_{13}	…	y_{1b}	$\bar{y}_{1.}$
2	y_{21}	y_{22}	y_{23}	…	y_{2b}	$\bar{y}_{2.}$
i	y_{i1}	y_{12}	y_{i3}	…	y_{ib}	$\bar{y}_{i.}$
a	y_{a1}	y_{a2}	y_{a3}	…	y_{ab}	$\bar{y}_{a.}$
Moyennes	$\bar{y}_{.1}$	$\bar{y}_{.2}$	$\bar{y}_{.3}$	…	$\bar{y}_{.b}$	$\bar{y}_{..}$

$$\bar{y}_{i.} = \sum y_{ij} / b$$

$$\bar{y}_{.j} = \sum y_{ij} / a$$

$$\bar{y}_{..} = \sum \sum y_{ij} / ab$$

CRÉATION de PLANS EN BLOCS avec des répétitions (n > 1)

départ : plan complet de facteurs primaires

b = nombre modalités du (des) facteur(s) nuisibles (secondaires)
= nombre blocs = nombre de répétitions (aussi noté n)

N = nombre essais plan facteurs primaires

N_B = N x b = nombre d'essais plan en blocs … peut être grand

PLAN en blocs

croisement du plan de facteurs primaires par lui-même b fois

(plan facteurs primaires) X (plan facteurs primaires) X ··· X (plan facteurs primaires)

APPLICATION : plans 2^k avec n (=b) répétitions

- répétitions employées pour définir le facteur bloc
- bloc = facteurs secondaires

autres cas : plans en bloc sans répétitions (n = 1)

- plans 2^k / plans 2^{k-p}
- plans central composite (surface réponse)

ANOVA avec facteur bloc

$$\sum\sum(y_{ij} - \bar{y}_{\cdot\cdot})^2 = \sum\sum \left[(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}) + (\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot}) + (y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot}) \right]^2$$

$$= b \sum(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 + a \sum(\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2$$

$$\sum\sum (y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot})^2$$

$SS_{tot} = \sum\sum(y_{ij} - \bar{y}_{..})^2$ **variabilité totale**

$SS_{trait} = b\sum (\bar{y}_{i.} - \bar{y}_{..})^2$ **variabilité traitement**

$SS_{bloc} = a\sum (\bar{y}_{.j} - \bar{y}_{..})^2$ **variabilité bloc**

$SS_{err} = \sum\sum (y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot})^2$ **erreur (résiduelle)**

$$SS_{tot} = SS_{trait} + SS_{bloc} + SS_{err}$$

ANOVA avec facteur bloc

Source	SS	df	MS	F
traitement	SS_{trait}	$a-1$	$SS_{trait}/(a-1)$	$F_0 = MS_{trait} / MS_{err}$
bloc	SS_{bloc}	$b-1$	$SS_{bloc}/(b-1)$	$F_1 = MS_{bloc} / MS_{err}$
resid	SS_{err}	$(a-1)(b-1)$	$SS_{err} / (a-1)(b-1)$	--------
total	SS_{tot}	$ab-1$	----------	--------

hypothèse principale à tester $H_0 : \tau_1 = \tau_2 = \ldots = \tau_a = 0$

rejet si $F_0 > F_{df1,\, df2,\, 1-\alpha}$ $df1 = a-1$ $df2 = (a-1)(b-1)$

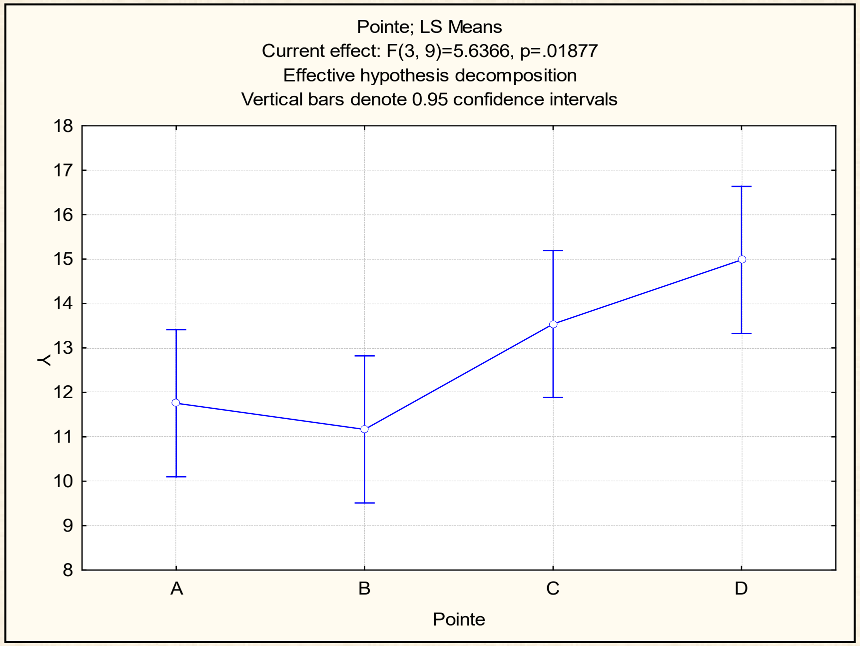
hypothèse H_1 : $\beta_1 = \beta_2 = \ldots = \beta_b = 0$ **pas d'effet bloc**

rejet si $F_1 > F_{df3,\, df2,\, 1-\alpha}$ $df3 = b-1$ $df2 = (a-1)(b-1)$

en général, cette hypothèse sera rejetée

on a bloqué cette source de variabilité potentielle afin d'obtenir une analyse du facteur principal en éliminant une source de variabilité nuisible

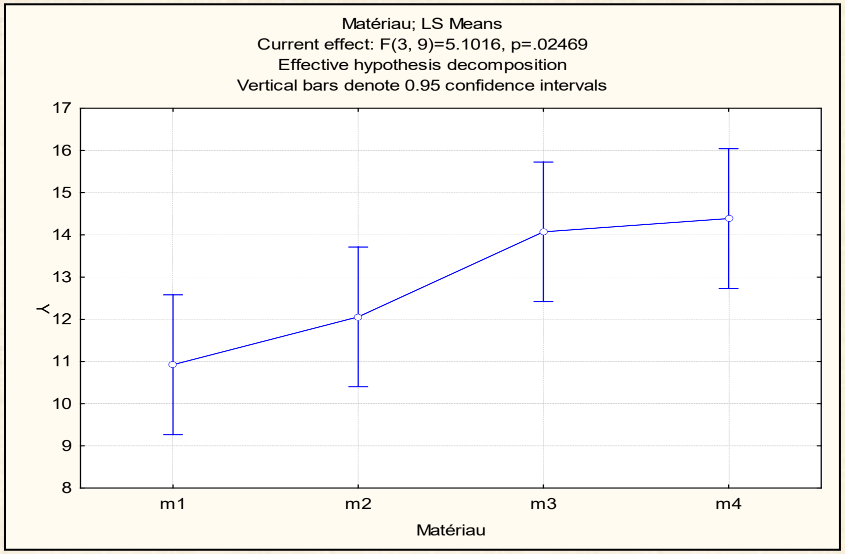
Exemple : test de dureté Rockwell



ANOVA

	df	Y SS	Y MS	Y F	Y p
Intercept	1	2646.59	2646.59	1236.07	0.00000
Pointe	3	36.21	12.07	5.64	0.01877
Matériau	3	32.77	10.92	5.10	0.02469
Error	9	19.27	2.14		
Total	15	88.25			

type de pointe : différences
lesquelles ?
comparaisons Tukey



__différence entre__
pointe A et pointe D

pointe B et pointe D

matériaux : différents … on savait

Plans en carrés latins : exemple plans en blocs

applications

- comparer k modalités d'un facteur principal en présence de
  2 facteurs nuisibles avec chacun k modalités
- modalités des facteurs nuisibles : agrégation pour en avoir k

modèle d'analyse

- effets principaux seulement
- aucun effet d'interaction

2 effets nuisibles : rangée , colonne

carrés latins standards

3 x 3

A	B	C
B	C	A
C	A	B

4 x 4

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

5 x 5

A	B	C	D	E
B	C	D	E	A
C	D	E	A	B
D	E	A	B	C
E	A	B	C	D

6 x 6

7 X 7

A	B	C	D	E	F	G
B						
C						
D						
E						
F						
G						

autres carrés latins : permutation des rangées et colonnes

Exemple : carré latin 4 x 4

étude de 4 additifs essence A – B – C – D facteur principal

réponse : Y = consommation (m/g)

2 facteurs nuisibles : voiture v1– v2 – v3 - v4

conducteur c1– c2 – c3 – c4

plan : carré latin 4 x 4

facteur principal / y

voiture

conducteur	v1	v2	v3	v4
c1	A / 21	B / 26	D / 20	C / 25
c2	D / 23	C / 26	A / 20	B / 27
c3	B / 15	D / 13	C / 16	A / 16
c4	C / 17	A / 15	B / 20	D / 20

moyenne additif	A : 18	B : 22	C : 21	D : 19
moyenne conducteur	c1 : 23	c2 : 24	c3 : 15	c4 : 19
moyenne voiture	v1 : 19	v2 : 20	v 3 : 19	v4 : 22

Plans en carrés latins : modèle et analyse

modèle carré latin p x p

réponse = effet général + effet traitement + effet rangée + effet colonne + erreur

$$y_{ij} = \mu + \tau_i + \alpha_j + \beta_k + \varepsilon_{ijk}$$

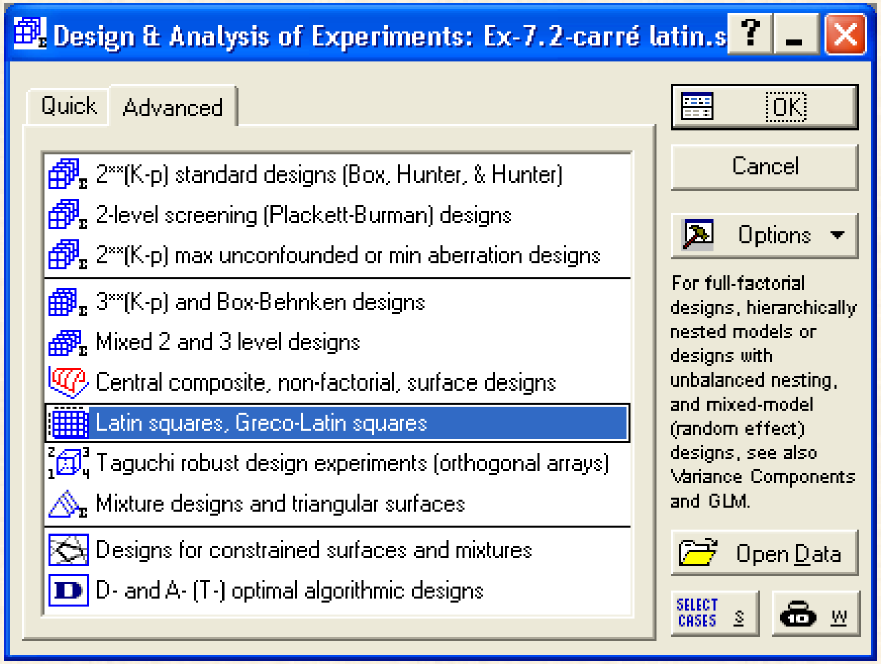
$i = 1,2,\ldots, p$ $j = 1,2,\ldots p$ $k = 1, 2,\ldots p$

p^2 triplets (i, j, k) utilisés parmi p^3 cas

ANOVA

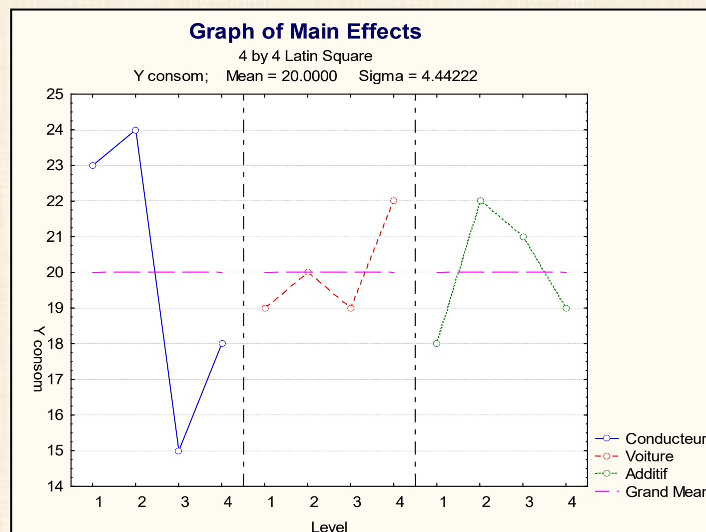
Source	SS	df	MS = SS/df	F
traitement	SS_{trait}	$p-1$	MS_{trait}	MS_{trait} / MS_E
rangée	$SS_{rangée}$	$p-1$	$MS_{rangée}$	------
colonne	$SS_{colonne}$	$p-1$	$MS_{colonne}$	------
erreur	SS_E	$(p-2)(p-1)$	MS_E	------
Total	SS_{tot}	$p^2 - 1$	-----------	

Plans en carrés latins : design et analyse avec *STATISTICA*



	Cond	Voiture	Additif	Y
1	1	1	1	21
2	1	2	2	26
3	1	3	4	20
4	1	4	3	25
5	2	1	4	23
6	2	2	3	26
7	2	3	1	20
8	2	4	2	27
9	3	1	2	15
10	3	2	4	13
11	3	3	3	16
12	3	4	1	16
13	4	1	3	17
14	4	2	1	15
15	4	3	2	20
16	4	4	4	20

Source	SS	df	MS	F	p
Conducteur	216	3	72.000	27.00	0.0007
Voiture	240	3	8.000	3.00	0.1170
Additif	40	3	13.333	5.00	0.0452
Residual	16	6	2.667		



Modèle général plan en blocs

A, B ,… facteurs primaires

$Y = \gamma_j + \beta_0 + \beta_1 A + \beta_2 B + \beta_{12} AB + \dots$ $j = 1, 2, \dots, b$ b = nombre blocs

β_0 : effet général γ_j : effet bloc j $\gamma_1 + \gamma_2 + \gamma_3 + \cdots + \gamma_b = 0$

pas d'interaction entre facteurs primaires et le facteur bloc

Y_{ij} : observation i bloc j $i = 1, 2, \dots, k$ $j = 1, 2, \dots, b$

$\overline{Y}_{.j}$ = moyenne bloc j

$\hat{\beta}_0 = \overline{Y}_{..}$ = grande moyenne = moyenne sur les blocs

$\overline{Y}_{.j} - \overline{\overline{Y}}$: écart bloc

$SS_{bloc} = k \sum (\overline{Y}_{.j} - \overline{\overline{Y}}_{..})^2$ variabilité bloc

$SS_{tot} = \sum \sum (Y_{ij} - \overline{Y}_{..})^2$ variabilité totale

$SS_{tot} - SS_{bloc}$ = variabilité facteurs primaires

Exemple : réaction chimique

2 facteurs primaires : A_concentration (- +) B_catalyseur (- +)

1 facteur bloc : C_laboratoire 1 -2 – 3

Y: rendement réaction

A	B	Y	BLOC
-1	-1	28	1
1	-1	36	1
-1	1	18	1
1	1	31	1
-1	-1	25	2
1	-1	32	2
-1	1	19	2
1	1	30	2
-1	-1	27	3
1	-1	32	3
-1	1	23	3
1	1	29	3

bloc	1	2	3
moyenne	28.25	26.50	27.75
moyenne globale		27.50	
écart	0.75	-1.00	0.25

effet 1 bloc = -1.00 - (0.75+ 0.25) = - 2.00

effet 2 bloc = 0.75 - 0.25 = 0.50

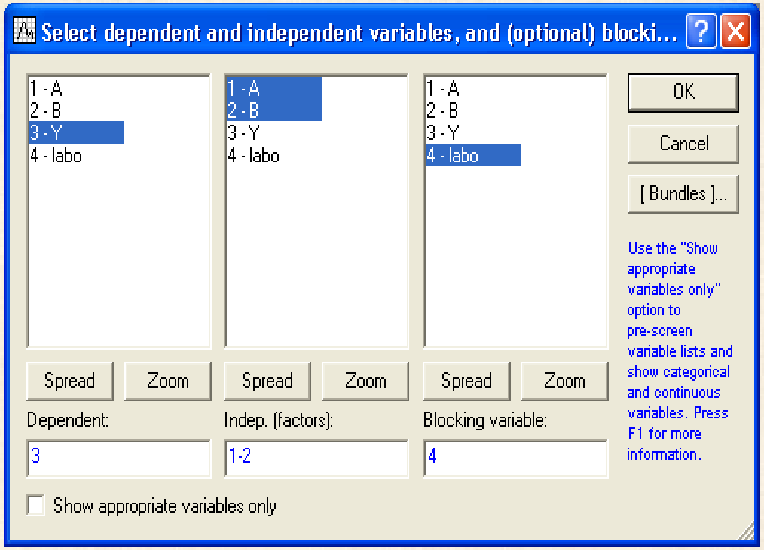
$SS_{bloc} = 4\,[\,0.75^2 + 1.00^2 + 0.25^2\,] = 6.50$

$SS_{tot} = (28.5-27.5)^2 + (36 - 27.5)^2 + \ldots + (29-27.5)^2$

$= 323$

$SS_{tot} - SS_{bloc} = 323 - 6.5 = 316.5$

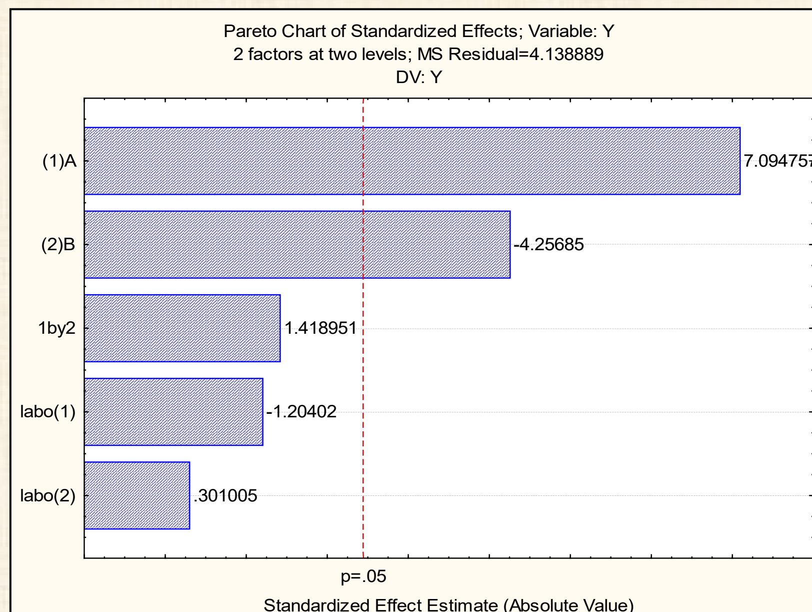
Exemple : réaction chimique



ANOVA; Var.:Y; R-sqr=0.92; Adj:0.86

	SS	df	MS	F	p
Blocks	6.50	2	3.25	0.785	0.4978
(1)A	208.33	1	208.33	50.336	0.0004
(2)B	75.00	1	75.00	18.121	0.0053
1 by 2	8.33	1	8.33	2.013	0.2057
Error	24.83	6	4.14		
Total SS	323.00	11			

	Effect	p	Coeff.
Mean/Interc.	27.500	0.0000	27.500
(1)A	8.333	0.0004	4.167
(2)B	- 5.000	0.0053	-2.500
labo(1)	- 2.000	0.2739	-1.000
1 by 2	1.667	0.2057	0.833
labo(2)	0.500	0.7736	0.250



Plans en blocs incomplets

a : nombre de traitements   k : nombre d'unités chaque bloc
b : nombre de blocs   r : nombre répétitions traitement
N = nombre total d'observations = b k = a r

Cas : # d'unités bloc < nombre traitements   k < a
cause   manque unités expérimentales

Exemple   bloc (matière première)   réponse Y

trait	1	2	3	4
1	-	73	71	74
2	67	-	72	75
3	68	73	-	75
4	72	75	75	-

a = 4 traitements
b = 4 blocs
k = 3 unités / bloc

cas important : Balanced Incomplete Bloc Design (BIBD)

- chaque traitement apparaît le même nombre de fois
- chaque paire de traitements apparaît le même nombre de fois
- liste de BIBD : a < 11 et k < 7   Box, Hunter, Hunter, pp. 269-274
- analyse : Montgomery 6 ed. p. 146

cas particuliers BIBD : plans combinatoires

b = nombre blocs = a ! / [k! (k – a) !]
nombre de blocs b grand si a > 5 et k ≈ a / 2

exemple: a = 6, k = 3 donne b = 20   N =20*3 = 6*10 = 60
plan BIBD non combinatoire avec b = 10   N = 10*3 = 6*5 = 30

PLANS EN BLOCS : résumé

- facteurs nuisibles sont généralement présents dans toutes les expériences industrielles et autres domaines
- si on peut identifier et contrôler les facteurs nuisibles, on peut construire un plan en bloc
- envisager un nombre restreint de modalités pour les facteurs nuisibles
- plan en bloc permet de dégager l'influence des facteurs primaires indépendamment des facteurs blocs
- plan en bloc, la randomisation est faite à l'intérieur de chaque bloc
- facteurs blocs n'ont pas d'effet d'interaction avec facteurs primaires
- existence catalogues (listes) de plans en blocs

Analyse de covariance

contexte : variables continues + variables catégoriques

- modèles de régression (var continues)
  + modèles d'analyse de la variance (var catégoriques)
- idée de base
  augmenter modèles d'analyse de variance contenant effets facteurs catégoriques (qualitatifs) avec une ou plusieurs variables continues potentiellement reliées à la réponse
- but: réduire la variance du terme d'erreur dans le modèle
  augmente la sensibilité à détecter des effets significatifs

X : variable concomitante (ou covariable)

- variable quantitative mesurée reliée potentiellement à Y non contrôlée (fixée) mais mesurée
- covariable : facteur nuisible (environnement)
- corrélée variable de réponse Y
- observées avant / durant / après l'expérience
- pas reliée aux autres facteurs catégoriques
  XA, XB,.. de l'expérience

Analyse de covariance

Exemple : étude CCA

ÉTUDE CCA : « Chromate Copper Arsenate »

CCA préserve le bois des dommages causés par l'eau et les insectes.
présente des risques pour la santé et l'environnement.

Nouvelle solution : «ORGANIC» remplacer le CCA?

Étude expérimentale : comparer solution CCA et ORGANIC.

données : 360 planches de pin traitées avec les 2 solutions (CCA et ORGANIC) à 3 niveaux de concentration: 0.25, 0.80, 2.25 (lbs /pi cu).

Niveaux de concentration employés pour 3 applications typiques:
bois exposé à l'air, bois de fondation, bois en eau salée.

Planches placées dans chambre à un vieillissement accéléré durant plusieurs heures

Chaque heure représente l'équivalent d'une exposition d'une année aux éléments.

Chaque planche : soumise test de bris à la rupture : réponse Y «Load».

3 facteurs : SOLUTION CONCENTRATION (facteurs contrôlés)
HEURE facteur continu mesuré (covariable).

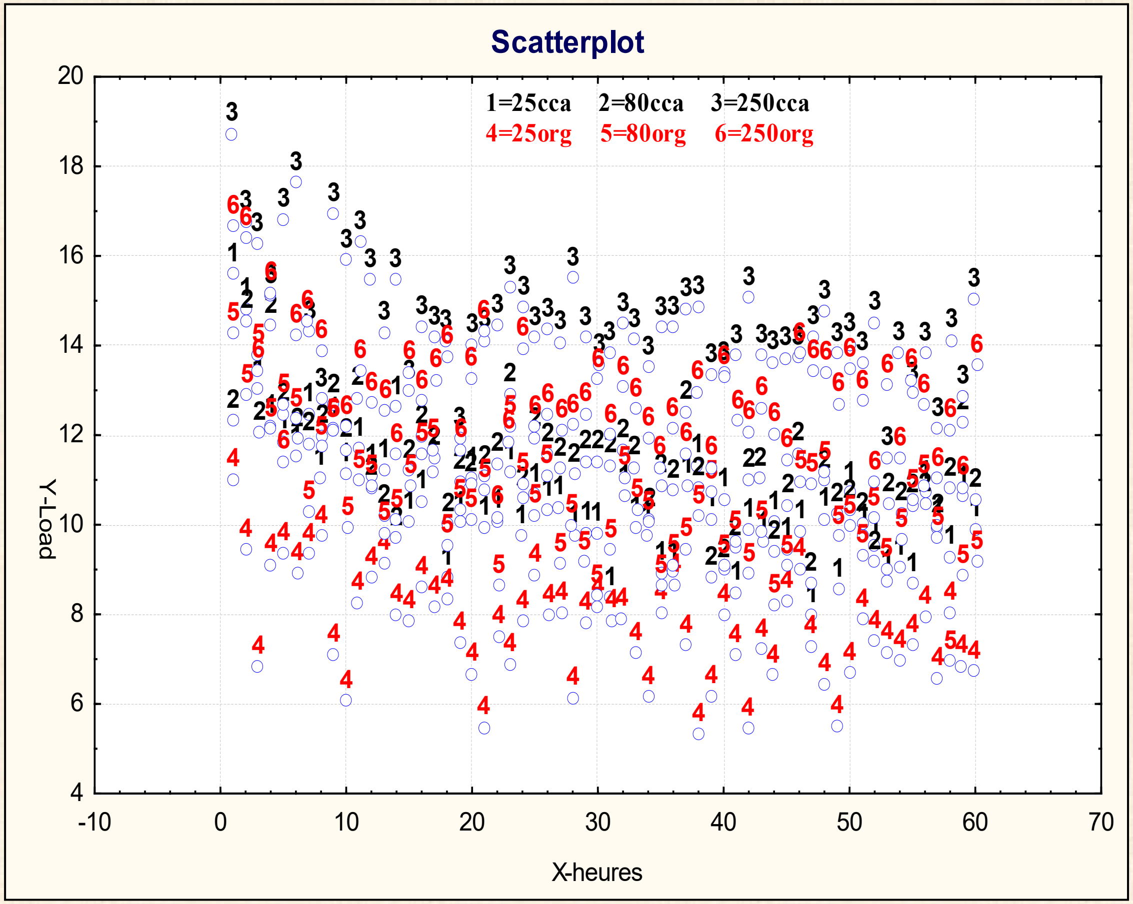
Objectif : comparer la performance de «ORGANIC » avec CCA.

Solution «ORGANIC» est-elle aussi bonne que solution CCA?

Analyse de covariance

Exemple : étude CCA

	Co	Sol	X-heu	Y
1	0.25	CCA	1.0	15.597
2	0.25	ORG	1.0	11.025
3	0.80	CCA	1.0	12.329
4	0.80	ORG	1.0	14.275
5	2.50	CCA	0.9	18.739
6	2.50	ORG	1.0	16.668
7	0.25	CCA	2.0	14.826
8	0.25	ORG	2.0	9.447
9	0.80	CCA	2.1	14.553
10	0.80	ORG	2.1	12.894
.	.	.	.	.
358	0.80	ORG	60.1	9.183
359	2.50	CCA	59.9	15.035
360	2.50	ORG	60.1	15.569



intrerprétation

- effet temps (heures) ?
- effet solution ?
- effet concentration ?

Analyse de covariance

Modèle d'analyse de covariance avec un facteur catégorique

$$Y_{ij} = \mu + \tau_i + \gamma\,(X_{ij} - \overline{X}_{..}) + \varepsilon_{ij}$$

$i = 1, 2, \ldots, g \quad j = 1, 2, \ldots, n_i$

μ : moyenne générale

τ_i : effet différentiel du facteur catégorique A

contrainte $\sum \tau_i = 0$

γ : coefficient de régression (pente) entre X et Y

X_{ij} : constantes connues de la covariable X

ε_{ij} : terme d'erreur $\sim N(0, \sigma^2)$

Remarque

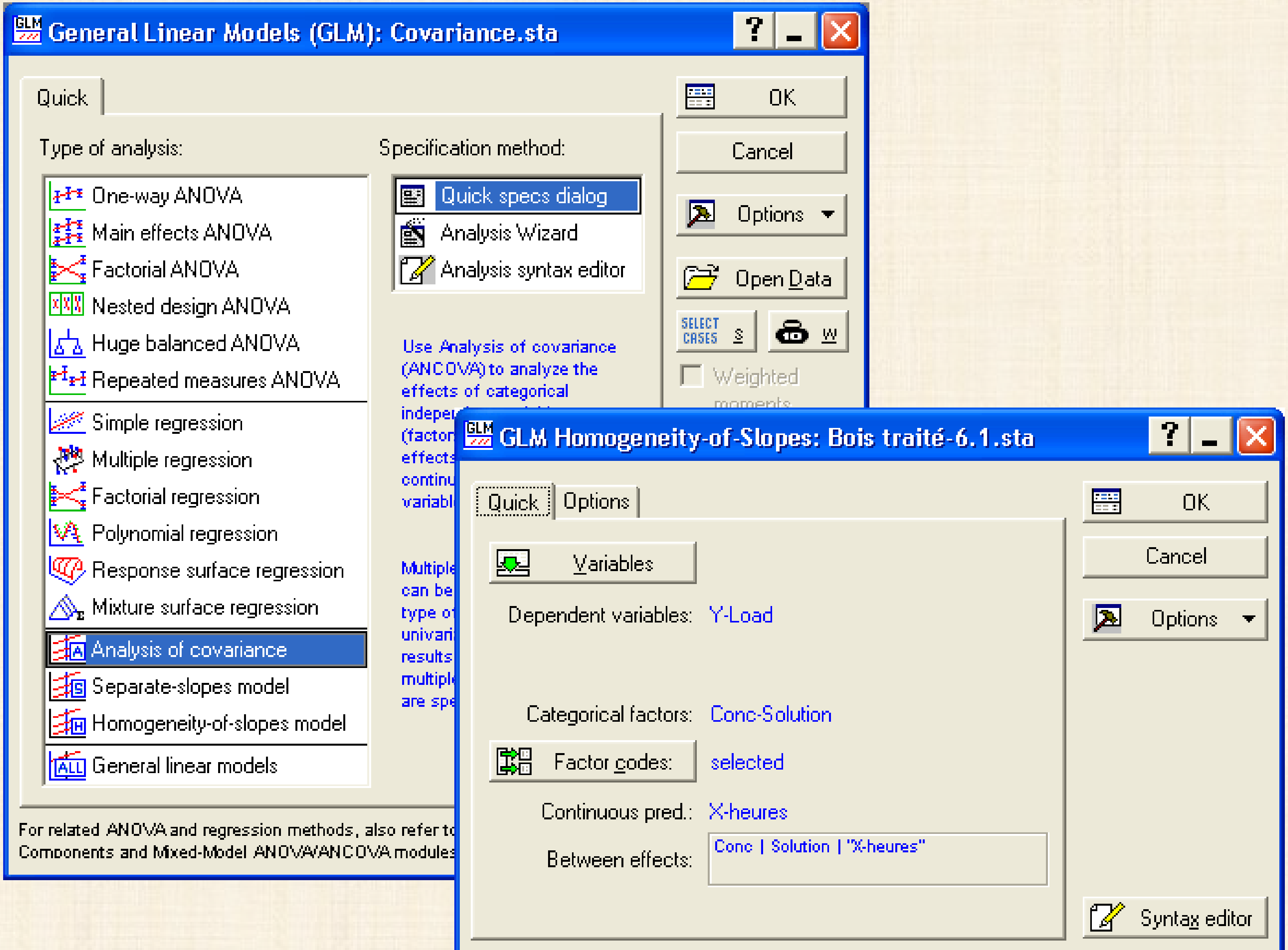
- γ constante - pas reliée au facteur d'intérêt
- pas d'interaction entre la covariable X et le facteur d'intérêt A

Inférence d'intérêt

$H_0 : \tau_1 = \tau_2 = \ldots = \tau_{g-1} = 0$

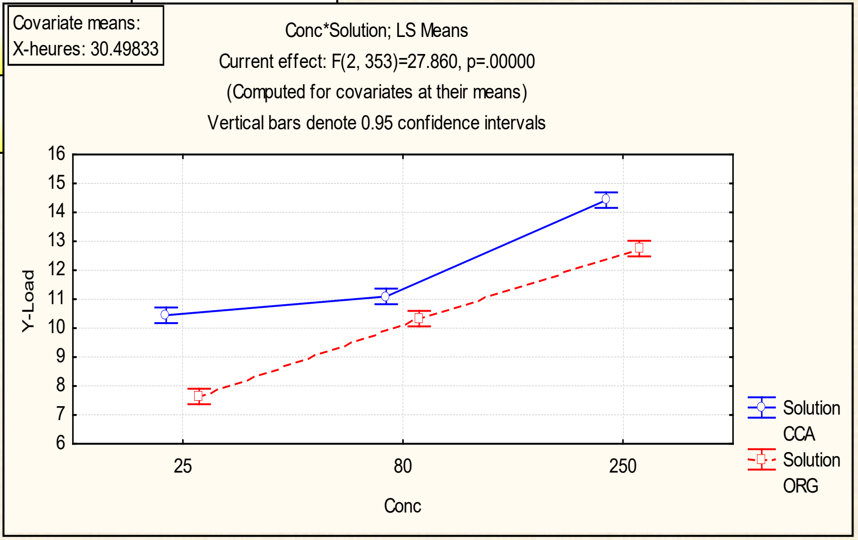
$H_1 : \tau_i$ ne sont pas tous nuls

Analyse de covariance : *STATISTICA*



Analyse de covariance

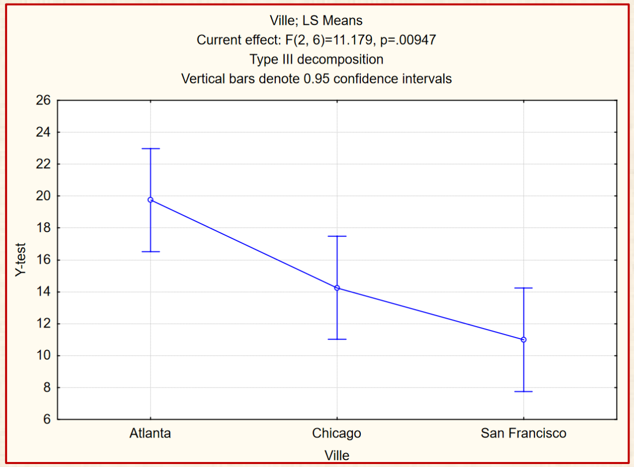
Modèle avec pente égale : analyse de covariance					
	DF	SS	MS	F	p-value
Intercept	1	13411.64	13411.64	11934.61	0.000000
X-heures	1	180.93	180.93	161.01	0.000000
Conc	2	1268.41	634.21	564.36	0.000000
Solution	1	275.61	275.61	245.26	0.000000
Conc*Solution	2	62.62	31.31	27.86	0.000000
Error	353	396.69	1.12		
Total	359	2184.32			

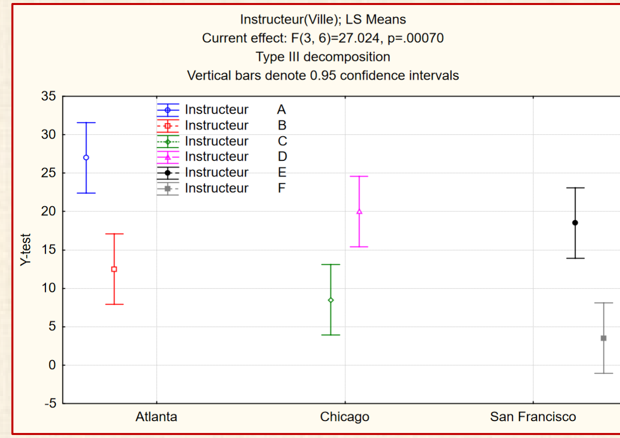


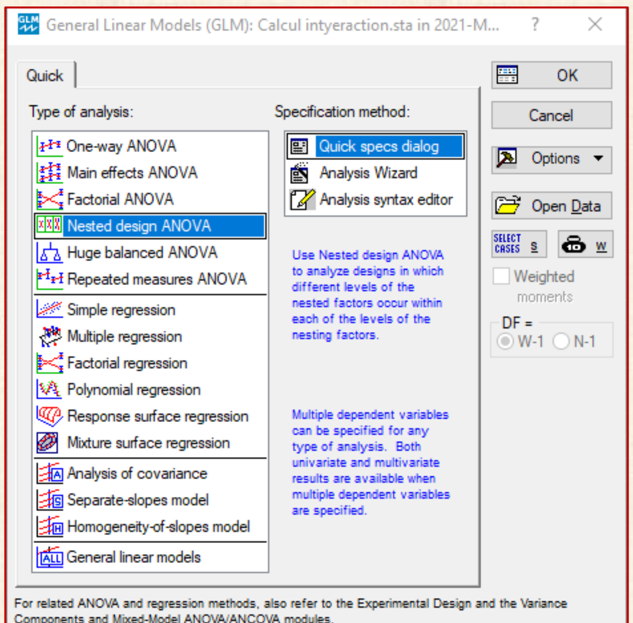
Facteurs emboités

Facteurs emboités
- effets principaux seulement
- concept d'interaction n'existe pas

1 A_Ville	2 B_Instructeur	3 groupe	4 Y-test
Atlanta	A	1	25
Atlanta	A	2	29
Atlanta	B	1	14
Atlanta	B	2	11
Chicago	C	1	11
Chicago	C	2	6
Chicago	D	1	22
Chicago	D	2	18
San Francisco	E	1	17
San Francisco	E	2	20
San Francisco	F	1	5
San Francisco	F	2	2







effet	Degr. of Freedom	Y-test SS	Y-test MS	Y-test F	Y-test p
Intercept	1	2700,00	2700,00	385,71	0,00000
Ville	2	156,50	78,25	11,18	0,00947
Instructeur(Ville)	3	567,50	189,17	27,02	0,00070
Error	6	42,00	7,00		
Total	11	766,00			

Tukey HSD test; variable Y-test (emboité 2 facteurs-6.1.sta)
Approximate Probabilities for Post Hoc Tests
Error: Between MS = 7.0000, df = 6.0000

Ville	Instructeur	{1} 27.000	{2} 12.500	{3} 8.5000	{4} 20.000	{5} 18.500	{6} 3.5000
Atlanta	A		0,0116	0,0034	0,2180	0,1156	0,0010
Atlanta	B	0,0116		0,6713	0,1765	0,3295	0,0937
Chicago	C	0,0034	0,6713		0,0342	0,0620	0,4837
Chicago	D	0,2180	0,1765	0,0342		0,9899	0,0061
San Franciscc	E	0,1156	0,3295	0,0620	0,9899		0,0098
San Franciscc	F	0,0010	0,0937	0,4837	0,0061	0,0098	

Facteurs aléatoires - modèles mixtes

Montgomery - Design and Analysis of Experiments p. 508

LOOM : matériau (rouleau) dans industrie textile
facteur aléatoire car échantillonnage au hasard dans lot

	LOOM	Y
1	L1	98
2	L1	97
3	L1	99
4	L1	96
5	L2	91
6	L2	90
7	L2	93
8	L2	92
9	L2	96
10	L3	95
11	L3	97
12 etc	L3	95

étude répétabilité et reproductibilité :
(R&R) pour évaluer un processus de mesure

	pièce	operateur	rep	Y
1	p1	op1	1	21
2	p2	op1	1	24
3	p3	op1	1	20
4	p4	op1	1	27
5	p5	op1	1	19
6	p6	op1	1	23
7	p7	op1	1	22
8	etc			

Milliken and Johnson, vol 1 p.285
modèle mixte 2 facteurs
Machine (M1 M2 M3) - facteur fixe
Person (P1 P2 P3 P4 P5 P6) - facteur aléatoire
Repeat (1 2 3) - pas un facteur

	Machine	Person	repeat	Score
1	M1	P1	1	52,0
2	M1	P1	2	52,8
3	M1	P1	3	53,1
4	M1	P2	1	51,8
5	M1	P2	2	52,8
6	M1	P2	3	53,1
7	M1	P3	1	60,0
	etc			