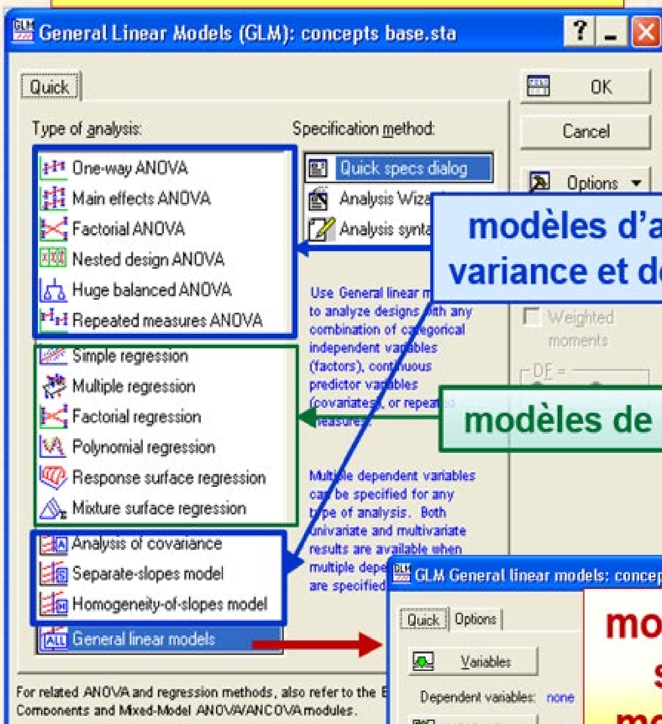
MODÈLES d'ANALYSE de la VARIANCE

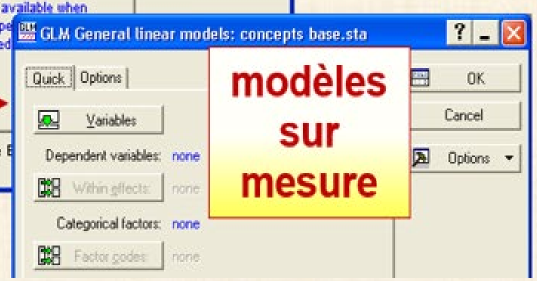
- Ch11 - Anova 1: introduction - Modèles d'Analyse de la Variance
- Ch12 - Anova 2 : 1 facteur
- Ch13 - Anova 3 : plusieurs facteurs
- Ch14 - Anova 4 : mesures répétées (étude longitudinale)
- Ch15 - Anova 5 : facteurs aléatoires

modules Statistica pour modèles d'analyse de variance (GLM)



modèles d'analyse de variance et de covariance

modèles de régression



modèles sur mesure

modules JMP modèles d'analyse de variance

MODÈLES d'ANALYSE de la VARIANCE
Ch11 - introduction

SYSTÈME / PROCESSUS et ANALYSE STATISTIQUE

Facteurs X

caractéristiques contrôlables ou mesurables à l'entrée

PROCESSUS

Réponses Y

caractéristiques mesurables de sortie

$$Y = f(X; \theta) + \varepsilon$$

ε : terme d'erreur $f = ?$

X : variables = facteurs	MODÈLE + DATA + ANALYSE
quantitatifs = numériques = continus	Régression (MR)
catégoriques = qualitatives	Variance (MAV)
catégoriques et continus	Covariance (MACV)

Elément de comparaison	Modèle d'analyse de régression	Modèles d'analyse de variance
But	développement d'un modèle prédictif de la réponse	identification des effets significatifs sur la réponse
Source des données : type d'étude statistique	observationnelles ou historiques	résultat d'un plan d'expérimentation
Nombre d'observations	grand nombre : centaines, milliers ou plus	petit nombre : dizaines
Variables d'entrée	continues	catégoriques var. continues traitées catégoriques
Nombre de valeurs distinctes des variables d'entrée	autant qu'il y a d'observations	nombre restreint généralement moins de 10
Utilisation des variables codage indicatrices (1 et 0) ou codage à effets (1 / 0 / -1)	occasionnelle	employé pour représenter les modalités
Emphase et difficulté	forme et la qualité du modèle	spécification du modèle reflétant la complexité du plan expérimental
Structure des données *Unités expérimentales*	simple observées sans aucun contrôle de l'expérimentateur	complexe plan expérimental : attribution des traitements aux unités expérimentales

éléments distinctifs

élément	qualificatif	commentaire
type d'étude statistique	▪ expérimentale (mode actif) ▪ observationnelle (mode passif)	unités expérimentales : oui attribution des traitements unités d'observations
Contrôle des X	▪ fixés (étudiés, principaux) ▪ mesurés ▪ bloqués (secondaires) ▪ inconnus	*inconnus* : tout ce que l'on ne connait pas = *l'erreur expérimentale* même si les données n'ont pas été générées par un plan expérimental structuré
nombre de X	▪ 1 : simple (one way) ▪ ≥ 2 : multifacteurs (multi way)	
nature des X	▪ Fixés ▪ Aléatoires ▪ hybrides (« mixed ») présence des 2 types	*fixé* : modalités (valeurs) choisies délibérément *aléatoire* : échantillon de modalités extraites de manière aléatoire = aucun contrôle sur les modalités

éléments distinctifs

élément	qualificatif	commentaire
nombre de Y	▪ 1 : ANOVA 1 Y ▪ ≥ 2 : MANOVA 4 Y Multivariate Analysis Of Variance	
design expérimental (plan, devis, protocole) = méthode d'assignation des traitements aux unités (sujets) expérimentales	▪ Aléatoire (CRD) ▪ Blocs (RCBD) ▪ Blocs incomplets (IBD) ▪ Blocs incomplets équilibrés (BIBD) ▪ Carrés Latins et Graeco-Latins ▪ Mesures répétées ▪ Fractionnaires ▪ Emboités (nested) ▪ SplitPlot = parcelles divisées ▪ Crossover ▪ …..	*mesures répétées* variable de réponse mesurée plusieurs fois sur la même unité expérimentale *SplitPlot* unité expérimentale (plot) est subdivisée dans l'assignation des traitements aussi restriction à la randomisation

Caractéristiques des modèles d'analyse de la variance

- Les ensembles de données sont relativement modestes: quelques dizaines d'observations en général;
- Les données proviennent généralement d'expériences planifiées avec un nombre restreint (moins de 10 ?) de variables souvent appelées facteurs (X) considérés comme catégoriques;
- Il y a une différence fondamentale par rapport aux données observationnelles: des unités expérimentales sont identifiées et reçoivent des traitements (combinaison des facteurs) selon un protocole expérimental faisant intervenir la randomisation des traitements dans l'attribution des traitements aux unités expérimentales;
- Il y a différents protocoles expérimentaux: plan complètement randomisé, plan en blocs, …
- Les données n'existaient pas au début du projet : on réalise des expériences pour générer les données de la réponse Y;
- Les facteurs d'expériences sont généralement fixés avec peu de modalités, généralement entre 3 et 5;
- On distingue 2 types de facteurs : les facteurs fixés dont les valeurs sont d'un d'intérêt et qui affectent la moyenne de la variable de réponse; il y a aussi des facteurs aléatoires qui affectent la variance de la réponse;
- Les facteurs sont généralement croisés mais peuvent aussi être emboités;
- Le modèle reliant les facteurs X sur la réponse Y doit être proposé avec soin reflétant le protocole expérimental;
- l'écriture du modèle (équation) est une opération plus complexe que dans le cas des données observationnelles; voir les exemples pages 16-18;
- des unités expérimentales peuvent recevoir des modalités différentes d'un même facteur ou être mesurées plusieurs fois dans le temps - connue sous le nom de mesures répétées ou longitudinales;

ÉTUDES EXPÉRIMENTALES : concepts de base

- **Unité expérimentale :** plus petite unité de matériel (vivante ou non) pouvant recevoir un traitement
- **Facteurs explicatifs = X** (variables catégoriques le plus souvent)

  facteurs expérimentaux = facteurs contrôlés = facteurs manipulés
  modalités (niveaux) choisis par l'expérimentateur

  facteurs observationnels : non contrôlés par expérimentateur
  facteurs d'arrière-plan : associés aux unités expérimentales ou à l'environnement de l'expérience
- **Traitements :** combinaisons des facteurs contrôlés
- **Structure des traitements** : croisés, emboités
  nombre: tous (= complet) / fractionnaires (= sous ensemble)
- **Design expérimental :** règles et procédures de l'assignation aléatoire des traitements aux unités expérimentales (ou vice versa)
- **Mesure du résultat Y** sur les unités expérimentales : output

AUTRES : randomisation
blocage = randomisation sous contrainte
n = # répétitions chaque traitement
t = # traitements   10 var X a 2 modalités t = 1024
N = taille expérience = traitements x répétitions = t * n
N plus petit possible ?   Feature eng.

Principaux Designs Expérimentaux

NOM	DÉFINITION
CRD = Completely Randomized Design	traitements assignés aux unités expérimentales complètement au hasard
RCB = Randomized Complete Block	▪ division des unités expérimentales en blocs homogènes (strates) (b = # blocs) ▪ randomisation des traitements à l'intérieur de chaque bloc (k = # taille bloc)
IBD = Incomplete Block Design	taille du bloc (k) < nombre de traitements (t)
BIBD = Balanced Incomplete Block Design	chaque traitement apparait avec tous les autres traitements le même nombre de fois dans le même bloc facilite la comparaison des paires traitements
MESURES RÉPÉTÉES	même unité expérimentale reçoit plusieurs traitements dans le temps
FACTORIELLE	toutes les combinaisons de traitements
FRACTIONNAIRES	une partie des traitements choix traitements : méthode, règles …
HIÉRARCHIQUE (NESTED)	facteurs emboités souvent : facteurs sont aléatoires

Types d'expériences

complexité

faible

forte

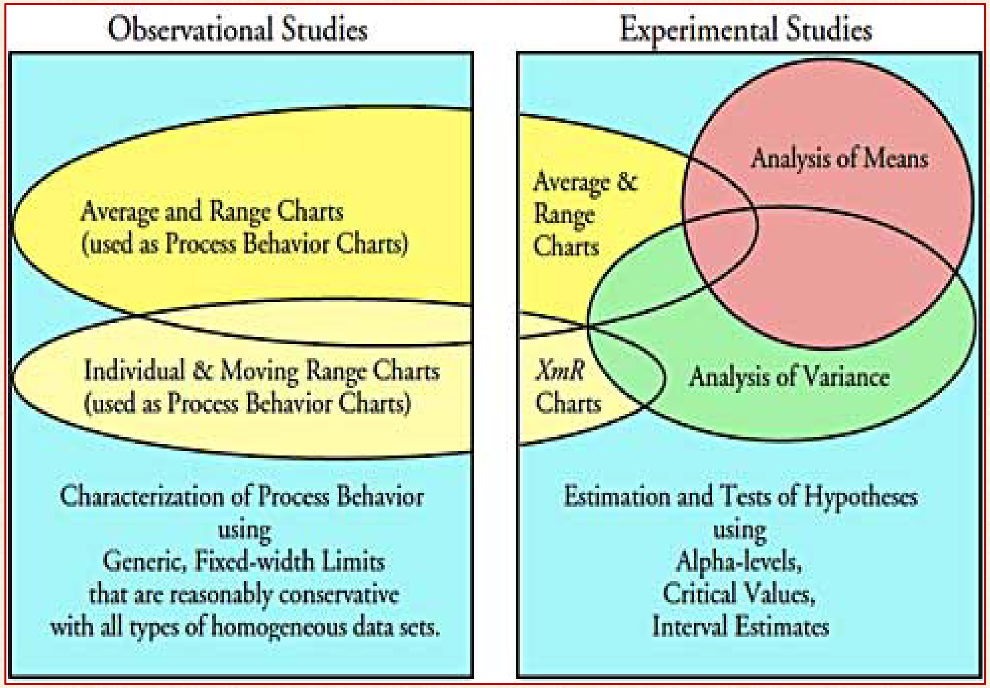
1. Essais et erreurs
2. Production lots sous conditions contrôlées
3. Expériences pilotes
4. Expériences avec 1 facteur
5. Comparaison de 2 méthodes
6. Expériences de 2 à 4 facteurs
7. Expériences de 5 à 20 (et plus) facteurs:

   tamisage (screening)
8. Expériences d'envergure avec plusieurs phases:

   modélisation, optimisation

ÉTUDES STATISTIQUES

SPC | Statistique



Donald J. Wheeler PhD Statistics and SPC 03/03/2014
http://www.qualitydigest.com/

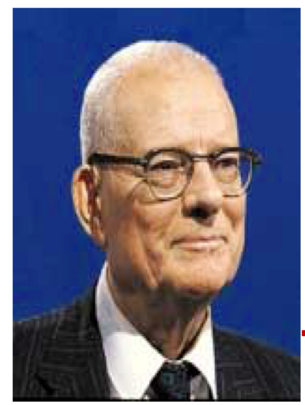
résumé de l'article Wheeler

Autre DISTINCTION

Two Types of Studies

Enumerative Studies:	Analytic Studies:
Census	Selection of suppliers
Inventory	Poll to determine strategy
Exit survey at polls	Experiment to improve performance
Acceptance sampling	Drug testing

W. E. Deming's Two Types of Studies



Chapter 7 from *Some Theory of Sampling*, 1950

The aim of any experiment is to provide a rational basis for action

Enumerative study: an experiment in which action will be taken on the universe.

Analytic study: an experiment in which action will be taken on a cause system to improve performance in the future.

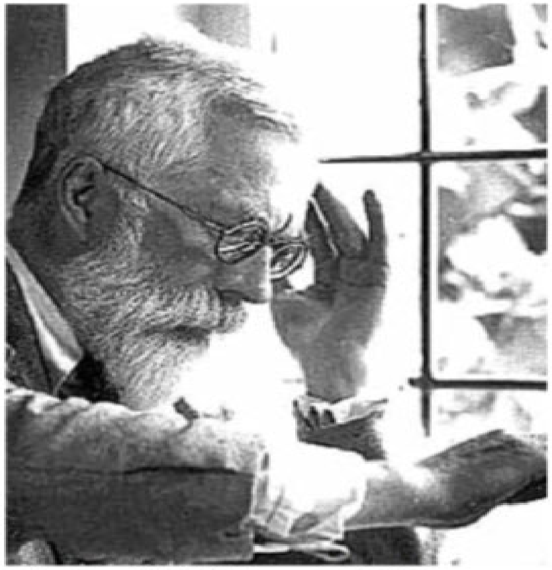
https://cours.polymtl.ca/mth6301/mth8302/mth8302-Cours&Plus/Clement-Etudes_observationnelles_Etudes_experimentales.pdf

C'est quoi le SPC ? https://cours.polymtl.ca/mth6301/Clement/Clement-ControleStatistiqueProcessus.pdf

Planification et analyse statistique d'expériences (DOE) cours à Polytechnique : MTH8301

Méthode extrêmement importante en R&D, innovation,….

Rothamsted Experimental Station



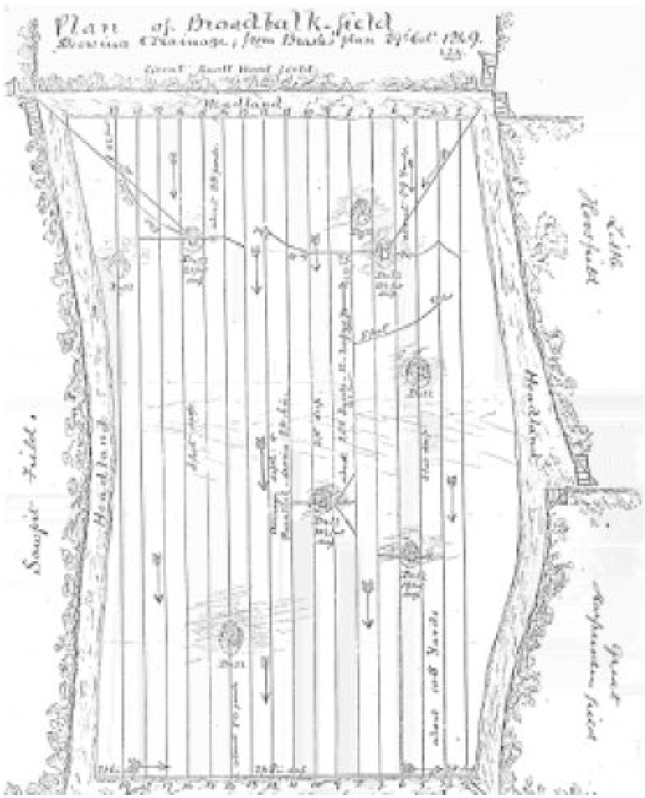
Sir Ronald A. Fisher 1890-1962

développement de nouveaux concepts … essentiels
origine en agriculture …. mais s'appliquent ingénierie, médical …
TOUS les domaines de l'activité humaine

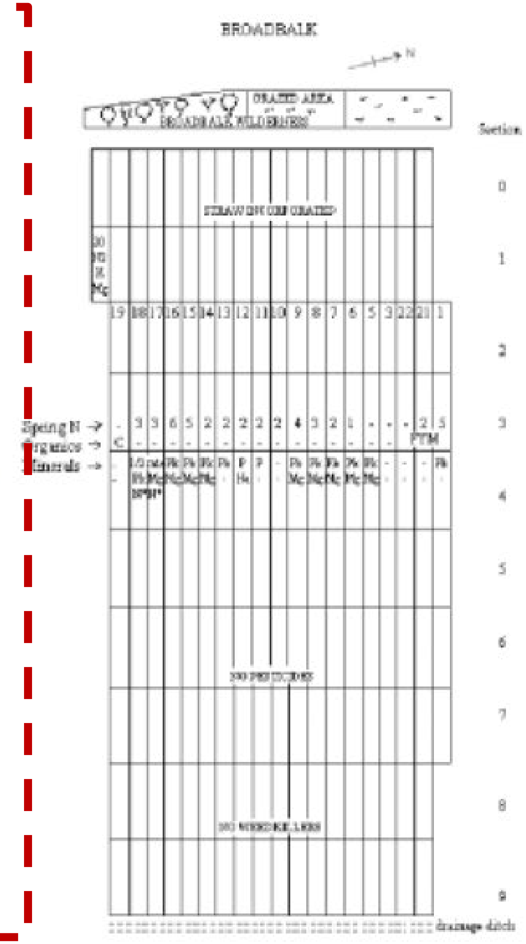
Our success at Amazon is a function of how many experiments we do per year, per month, per week, per day."

Jeff Bezos, who, according to **this *Harvard Business Review* report on how to do smart business experiments**, fired a group of web designers who changed an Amazon page without first doing testing.

LANGUAGE de L’EXPÉRIMENTATION



- Factors
- Background Variables
- Response variables
- Experimental unit
- Randomization
- Replication
- Blocks
- Plots
- Subplots
- Experimental Pattern



remarques / définitions

- **modèle linéaire général : la régression et l'analyse de variance sont des cas particuliers de ce modèle qui est le plus employé en analyse statistique ;**

- **en général, les unités (sujets, individus statistiques) sont choisies au hasard parmi une population d'unités statistique ;**

  lorsque l'on contrôle certains facteurs, on fait l'assignation au hasard des traitements (combinaison des facteurs contrôlés) aux unités :
  c'est la randomisation des traitements
  caractéristique fondamentale et essentielle d'un plan d'expérience;
  sinon les données sont de nature observationnelle;

- **si nécessaire, on fixe (contrôle) des facteurs secondaires sur les unités pour améliorer l'efficacité de l'étude (plans en blocs, …) : c'est le principe du blocage (stratification);**

- **facteur intra (within) : si la même unité statistique est mesurée plusieurs fois (temps, différentes conditions), on a un plan en mesures répétées ;**
  le facteur enfoui dans la réponse est appelé *facteur intra* (« within ») ;
  on traite ce cas en définissant plusieurs variables de réponse et en effectuant avec une analyse de variance multidimensionnelle (MANOVA) ; les réponses sont naturellement dépendantes car elles sont mesurées sur le même sujet ;
  on décompose la variabilité totale en *variabilité intra sujet (within)* et en *variabilité inter sujet (between)*

- **le sujet (unité expérimentale) est un facteur aléatoire**

remarques / définitions

- **facteur inter (between)** : varie sur des groupes distincts de sujets ;
- **analyse de covariance** : on mesure (contrôle indirect) une variable quantitative (X) jouant un rôle de facteur et que l'on est en présence de d'autres facteurs catégoriques (A, B, C…);
- **facteurs emboités (nested)** : si les modalités prisent par un facteur sont spécifiques aux modalités d'un autre facteur, le concept d'interaction entre les 2 facteurs n'est pas défini;
- **reconnaître les structures (traitements, assignation)** présentent dans les données est nécessaire si l'on veut réaliser la **« bonne » analyse statistique** ; c'est plus facile si on a participé à la planification de l'étude et si on connait le domaine (contexte) d'application;

l'utilisation d'un logiciel statistique de donne pas la réponse à ces questions !

L'analyse statistique c'est plus que du « point and click » !

remarques / définitions

- Il faudra caractériser les données par la structure des traitements et la structure d'assignation (aussi appelé *protocole*) des traitements aux unités expérimentales

- Exemples de plans expérimentaux: plan factoriel complet,
  plan fractionnaire, plan complètement aléatoire,
  plan en blocs plan blocs incomplets, plan en carré Latin,
  plan en mesures répétées, plan avec facteurs aléatoires, ….

- On distingue aussi les modèles avec méthode d'analyse statistique employée : ANOVA, MANOVA, analyse de covariance, analyse en mesures répétées, idenfication des facteurs aléatoires si présents,….

- Écriture mathématique des modèles: GROSSE différence entre
  - modèles de régression (MR): design n'intervient pas car on n'assigne pas les traitements aux UE : étude observationnelle
  - modèle d'analyse de variance (MAV) : structure du design expérimental est prise en compte

ÉCRITURE MODÈLE D'ANALYSE DE VARIANCE : exemples

différence majeure par rapport à l'écriture des modèles de régression

Design classification simple : 1 facteur catégorique A

(1) $Y_{ir} = \mu + \alpha_i + \varepsilon_{ir}$ $i = 1, 2, \ldots, I$ (= nombre groupes) $r = 1, 2, \ldots, n_i$

Y_{ir} : valeur de la variable de réponse Y r-ème essai modalité i du facteur A

μ : effet général - comme β_0 dans les modèles de régression

α_i : effet différentiel de la modalité i du facteur $\Sigma_i \alpha_i = 0$

ε_{ir} : erreur aléatoire distribuée $N(0, \sigma^2)$

Design classification double avec 2 facteurs A et B croisés

(2) $Y_{ijr} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijr}$ $i = 1, 2, \ldots, I$ / $j = 1, 2, \ldots, J$ / $r = 1, 2, \ldots, n_{ij}$

α_i : effet principal de A $\Sigma \alpha_i = 0$

β_j : effet principal de B $\Sigma \beta_i = 0$

$(\alpha\beta)_{ij}$: effet d'interaction entre A et B $\Sigma_i (\alpha\beta)_{ij} = 0$ $\Sigma_j (\alpha\beta)_{ij} = 0$

Design classification double avec facteur B emboité dans facteur A

(3) $Y_{ijr} = \mu + \alpha_i + \beta_{j(i)} + \varepsilon_{ijr}$ $i = 1, 2, \ldots, I$ / $j = 1, 2, \ldots, J$ / $r = 1, 2, \ldots, n_{ij}$

α_i : effet facteur A et $\Sigma \alpha_i = 0$

$\beta_{j(i)}$: effet facteur B et $\sum \beta_{j(i)} = 0$ pour tout i

effet d'interaction AxB n'existe pas

remarque: termes d'erreur ε_{ir} ε_{ijr} .... sont emboités dans la structure la plus fine (cellules) des données .... on devrait écrire $\varepsilon_{r(i)}$ $\varepsilon_{r(ij)}$

ÉCRITURE MODÈLE D'ANALYSE DE VARIANCE : exemples

Design d'analyse de covariance : facteur catégorique A + facteur continu X

pente β égale chaque sous-groupes de A = pas d'interaction entre A et X

(4) $Y_{ir} = \mu + \alpha_i + \beta\,(X_{ir} - \overline{X}_{..}) + \varepsilon_{ir}$ $\qquad \overline{X}_{..} = \Sigma\,\Sigma\, X_{ir} / n$

Design d'analyse de covariance : pentes distinctes

pentes β_i distinctes chaque sous-groupes de A = interaction entre A et X

(5) $Y_{ir} = \mu + \alpha_i + \beta_i\,(X_{ik} - \overline{X}_{..}) + \varepsilon_{ir}$

si $\beta_i = \beta$ modèle (5) devient modèle (4)

Design mesures répétées par tous les sujets sur toutes modalités facteur A

(6) $Y_{ir} = \mu + \alpha_i + \beta_r + \varepsilon_{ir}$ $\quad$ i = 1, 2, …, s (# sujets) / r = 1, 2,…, n

μ : effet général

α_i : effet aléatoire du sujet i - indépendantes N $(0, \sigma_\rho^2)$

β_r : effet de la modalité j du facteur fixe A et $\Sigma\, \tau_j = 0$

ε_{ir} : erreur aléatoire N $(0, \sigma^2)$

α_i, ε_{ir} indépendantes

<u>remarque</u> : exemple d'un modèle mixte - nature différentes des facteurs
facteur 1 = sujet = facteur aléatoire
facteur 2 = A = facteur fixe (modalités contrôlées)

ÉCRITURE MODÈLE D'ANALYSE DE VARIANCE : exemples

Design Split-Plot (parcelles divisées) : 2 tailles d'unités exp. - 2 termes d'erreur

(7) $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\gamma\beta)_{ijk} + \varepsilon_{ijk}$

plot (parcelle) + subplot (sous parcelle) + erreur exp.

$i = 1, 2, ..., r$ (facteur bloc = répétition) $j = 1, 2, ..., a$ (facteur A) $k = 1, 2, ..., b$ (facteur B)

- α_i : effet facteur bloc - facteur aléatoire distribution $N(0, \sigma_\alpha^2)$
- β_j : effet principal facteur A = facteur plot (plot=unité expérimentale)
- $(\alpha\beta)_{ij}$: erreur plot = interaction répétition x A
- γ_k : effet principal facteur B = facteur subplot (subdivision unité expérimentale)
- $(\alpha\gamma)_{ik}$: interaction répétition x B
- $(\beta\gamma)_{jk}$: interaction A x B
- $(\alpha\gamma\beta)_{ijk}$: erreur subplot = interaction répétition x AB
- ε_{ijk} : erreur expérimentale distribution $N(0, \sigma^2)$

version simplifiée de (7) $(\alpha\gamma)_{ik}$ et $(\alpha\gamma\beta)_{ijk}$ négligeables et incoporées avec ε_{ijk}

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

design Split-Plot
- restriction à la randomisation
- utile avec facteurs difficiles à changer

Syntaxe STATISTICA: procédures GLM, GLZ, PLS, GDA, MANOVA

Simple terms and operators

A Main effect for factor (categorical or continuous) predictor variable A.

A*B A by B interaction effect (for categorical or continuous predictor variables A and B).

A(B) A nested in B; the levels of categorical predictor (factor) A are nested within the levels of categorical predictor (factor) B.

Bar operator

A|B Complete factorial for factors A and B; this expression will be expanded (internally by the syntax interpreter) to A+B+A*B; complete higher order factorial designs can be specified analogously, for example, a complete 3-way design for factors A, B, and C can be specified as: A|B|C.

Factorial degree operator

A|B|C @2 The complete factorial for factors A, B, and C, up to degree 2; this expression will be expanded (internally by the syntax interpreter) to A + B + C + A*B + A*C + B*C; i.e., a factorial design will be constructed only up to the requested degree (to the second degree in this example; the three-way interaction will not be added).

Grouping-of-terms-operator

A | (B+C) The complete factorial for factors A and (main effects) B + C, this expression will be expanded (internally by the syntax interpreter) to A + B + C + A*B + A*C; note that the three-way interaction term A*B*C is not included in this model.

Deletion operator. The deletion operator (-) can be used to remove effects from a factorial design specified via the bar operator (|):

A|B|C-C Complete factorial for factors A, B, and C without the main effect for C; this expression will be expanded (internally by the syntax interpreter) to
A + B + A*B + A*C + B*C + A*B*C
note that main effect C is missing.

A|B|C-|C Complete factorial for factors A, B, and C without all interactions that involve factor C; this expression will be expanded (internally by the syntax interpreter) to
A + B + C + A*B
note that all interactions involving effect C are missing.

A|B|C-(A*C) Complete factorial for factors A, B, and C without the A by C interaction; this expression will be expanded
to A + B + C + A*B + B*C + A*B*C
note that the A by C interaction is missing.

A|B|C-|(A*C) Complete factorial for factors A, B, and C without all higher order interactions that involve the A by C interaction; this expression will be expanded (internally by the syntax interpreter) to
A + B + C + A*B + A*C + B*C
note that the only higher-order interaction involving A by C is the three-way interaction, which will is missing.

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

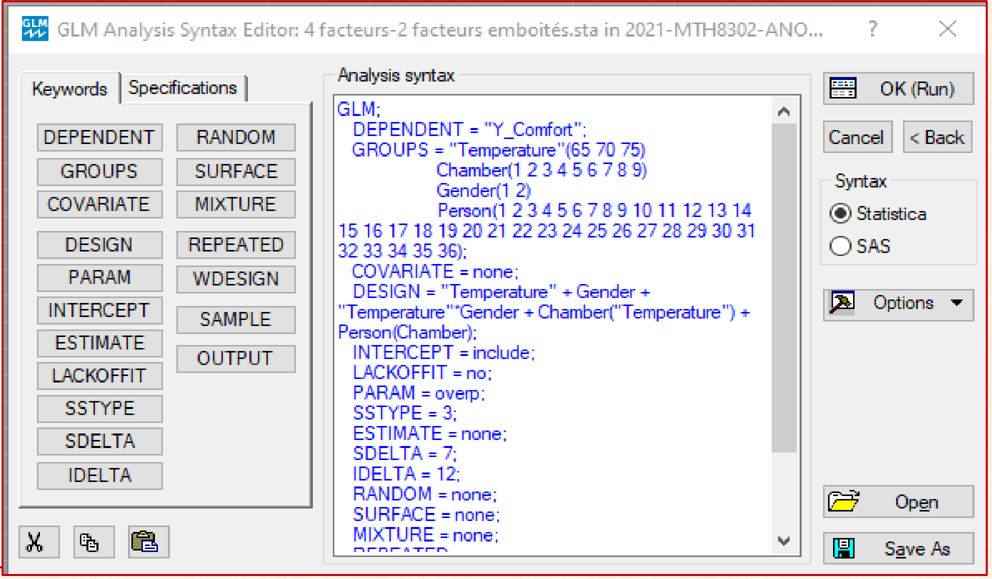
T = temperature
C = chamber
G = gender
P = person

Spécifiction modele

Y_Comfort

= T + G + T*G + C(T) + P(C)

C(T) : C emboité dans T
P(C) : P emboité dans C



CONCEPTS DE BASE : ANALYSE DE LA VARIANCE (ANOVA)

- **But** des **modèles** d'analyse de la variance et de **leur analyse**:
  tester la **présence ou non** de différences significatives entre
  les effets des modalités des facteurs X sur la variable de réponse Y
  En d'autres termes : le facteur X influence t-il la réponse Y?
  La même question s'applique dans une situation multifactorielle

- **L'exemple archi classique**: comparaison de 2 moyennes provenant
  de 2 échantillons indépendants. ….. **test t de Student**

- **L'analyse du modèle** consiste à produire un **tableau d'analyse de la variance** : décomposition de la variabilité totale selon
  les **différentes sources** présentes dans les données

- **Variabilité totale sera décomposée** =
  variabilité due aux écarts entre les moyennes (**INTER groupe**)
  \+ variabilité résiduelle (non expliquée, **INTRA groupe**)

- **Erreur expérimentale :** provient des répétitions + modélisation

- **Généralisation :** études statistiques avec plusieurs facteurs
  intervenant dans des plans expérimentaux complexes

Exemple 1 : 2 groupes avec un seul facteur A avec modalités a b

id	groupe	y-réponse
1	a	1
2	a	2
3	a	3
4	b	5
5	b	6
6	b	7

	groupe a	groupe b
moyenne somme de carrés (SS)	2 2	6 2
moyenne globale somme totale de carrés	4 28	

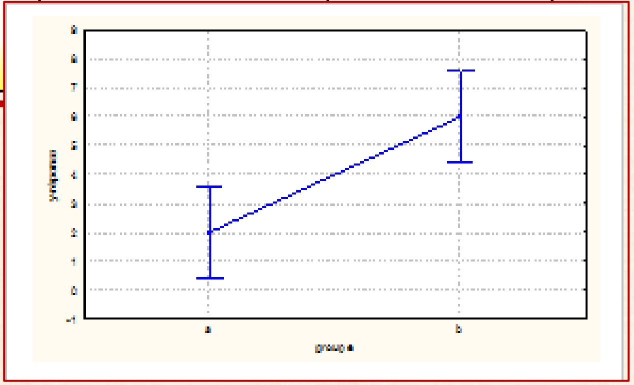
Analyse de la variance (ANOVA) = tableau avec

- identification des sources de variabilité
- calcul des sommes de carrés (SS)
- calcul des degrés de liberté (df)
- calcul des carrés moyens (MS) MS = SS / df
- calcul du ou des ratios F_0 = MS (effet) / MS (erreur)
- évaluation de la probabilité p (p-value): p = Prob ($F \geq F_0$)

On pourrait appelé cela : un théorème de Pythagore sophistiqué …

Exemple : 2 groupes avec un seul facteur A avec modalités a b

ANOVA					
Source variabilité	Degr. de liberté (df)	y-réponse SS	y-réponse MS	y-réponse F	y-réponse p
Intercept	1	96	96	96	0,00061
Inter groupe	1	24	24	24	0,00805
Intra groupe (erreur)	4	4	1		
Totale corrigée	5	28			



groupes sont statistiquement différents car p est « petit »

différence significative peut ne pas être importante en pratique

Tous les calculs sont basés sur les seules valeurs de Y et la structure des sous-groupes

d'où provient la valeur 96 dans le tableau ?

$$SS_{totale} = \sum y_i^2 = \sum(y_i - \bar{y} + \bar{y})^2$$

$$= \sum(y_i - \bar{y})^2 + n\bar{y}^2$$

$$= SS_{totale\ corrigée} + 6x4^2$$

$$= 28 + 96$$

Plusieurs facteurs

- En général, expériences ont plusieurs facteurs, généralement ≤ 10 (?)
- ANOVA : peut tenir en compte plusieurs facteurs ainsi que des designs expérimentaux complexes

Exemple : ajout deuxième facteur, sexe de l'individu

id	groupe	sexe	y-réponse
1	a	hom	1
2	a	hom	2
3	a	hom	3
4	b	hom	5
5	b	hom	6
6	b	hom	7
7	a	fem	3
8	a	fem	4
9	a	fem	5
10	b	fem	7
11	b	fem	8
12	b	fem	9

moyenne	groupe a	groupe b	a + b
hommes	2	6	4
femmes	4	8	6
tous	3	7	5

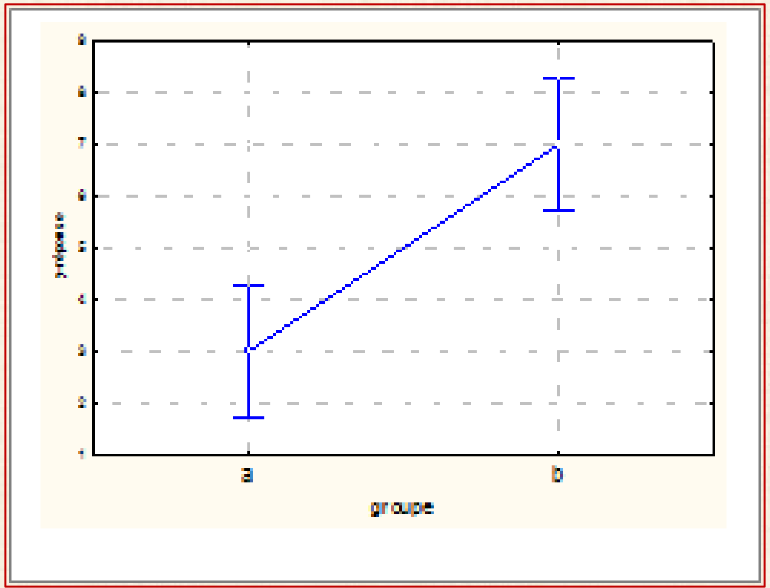
variabilité Y 3 sources

- facteur groupe (INTER)
- facteur sexe (INTER)
- inexpliquée = résiduelle = INTRA

Analyse 1: facteur sexe pas tenu en compte

id	groupe	sexe	y-réponse
1	a	hom	1
2	a	hom	2
3	a	hom	3
4	b	hom	5
5	b	hom	6
6	b	hom	7
7	a	fem	3
8	a	fem	4
9	a	fem	5
10	b	fem	7
11	b	fem	8
12	b	fem	9

ANOVA					
Source	df	y-réponse SS	y-réponse MS	y-réponse F	y-réponse p
Intercept	1	300	300	150	0,000000
Inter groupe	1	48	48	24	0,000624
Erreur	10	20	2		
Totale	11	68			



groupes sont différents

mais une partie de cette différence est due et confondue au facteur sexe

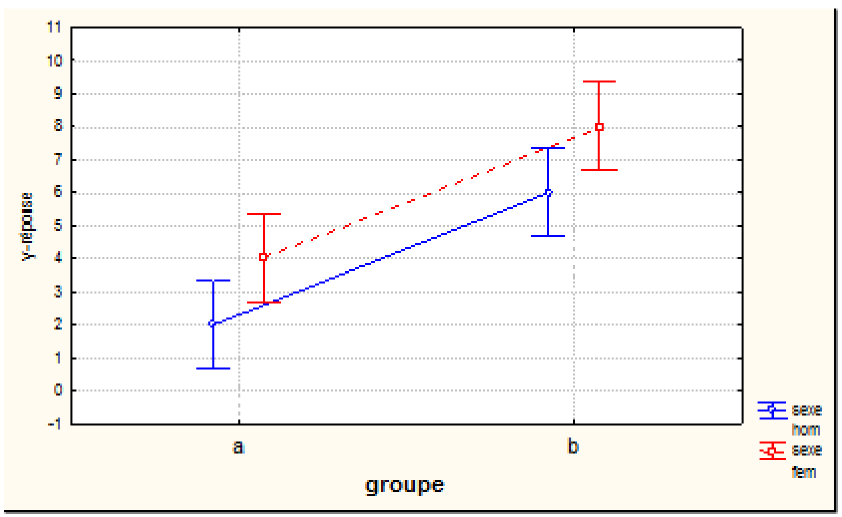
nouvelle analyse à faire avec 3 effets

- Facteur GROUPE
- Facteur SEXE
- interaction GOUPE*SEXE

Exemple 2 : ajout deuxième facteur, sexe de l'individu

Analyse 2 : modèle y = général + groupe + sexe + groupe*sexe

ANOVA					
Source	**df**	**SS**	**MS**	**F**	**p**
Intercept	1	300	300	300	0,0000
groupe	1	48	48	48	0,0001
sexe	1	12	12	12	0,0085
groupe*sexe	1	0	0	0	1,0000
erreur	8	8	1		
total	11	68			



vualisation des données

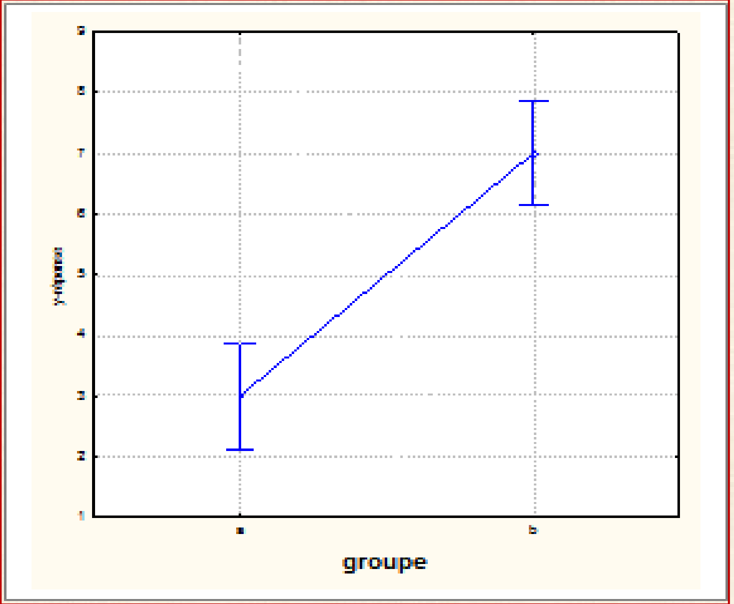
graphique de
l'interaction entre les 2 facteurs

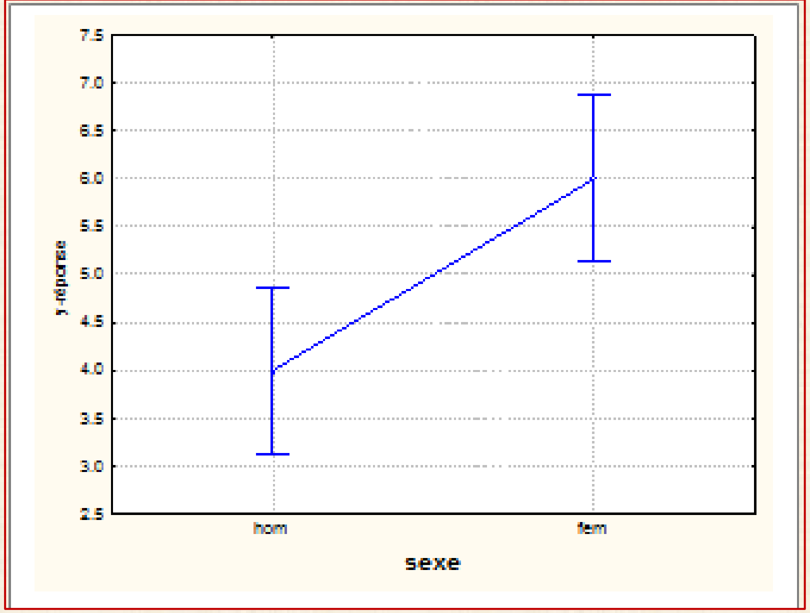
parrallélisme
implique
absence d'interaction
entre les 2 facteurs

Exemple 2 : ajout deuxième facteur, sexe de l'individu

Analyse 3 : modèle simplifié y = général + groupe + sexe

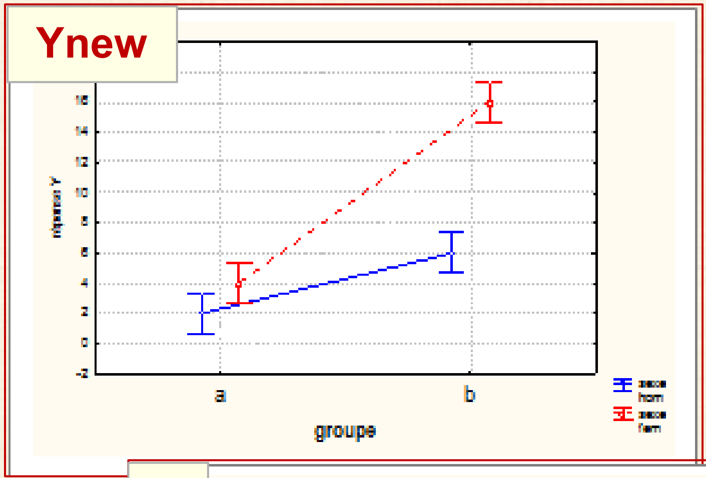
ANOVA					
Source	SS	df	MS	F	p
Intercept	300	1	300	337,5	0,000000
groupe	48	1	48	54,0	0,000043
sexe	12	1	12	13,5	0,005121
erreur	8	9	0.89		
total	68	11			

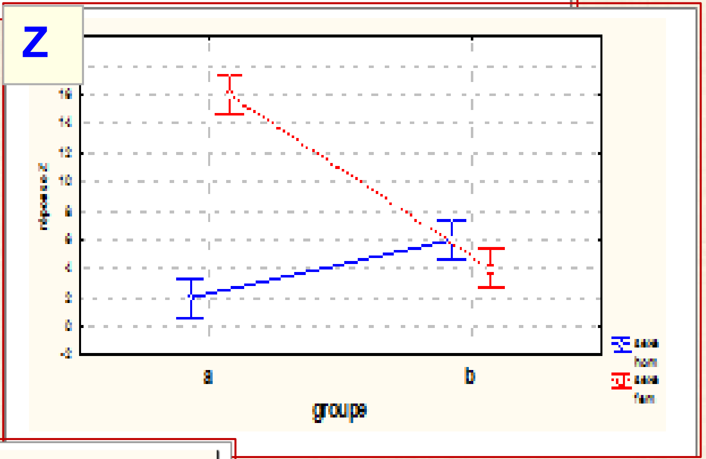




Exemple 3: 2 facteurs présence d'interaction

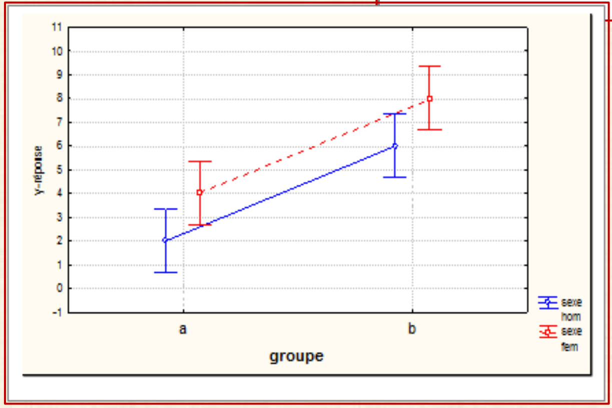
id	groupe	sexe	réponse Ynew	réponse Z
1	a	hom	1	1
2	a	hom	2	2
3	a	hom	3	3
4	b	hom	5	5
5	b	hom	6	6
6	b	hom	7	7
7	a	fem	3	15
8	a	fem	4	17
9	a	fem	5	16
10	b	fem	15	3
11	b	fem	17	4
12	b	fem	16	5





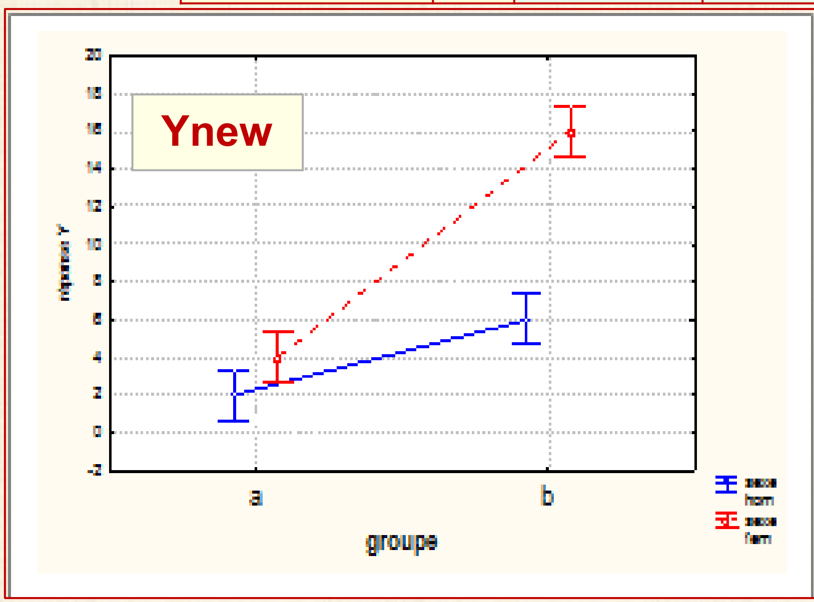
id	groupe	sexe	y-réponse
1	a	hom	1
2	a	hom	2
3	a	hom	3
4	b	hom	5
5	b	hom	6
6	b	hom	7
7	a	fem	3
8	a	fem	4
9	a	fem	5
10	b	fem	7
11	b	fem	8
12	b	fem	9

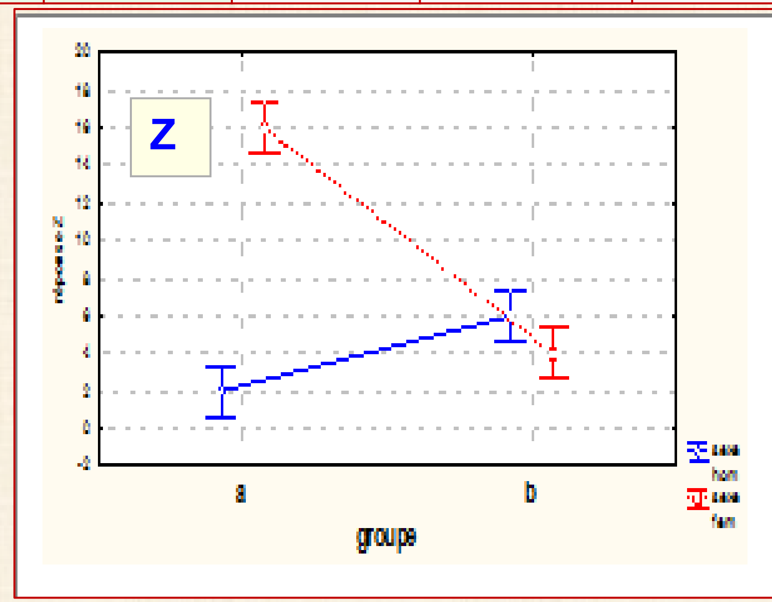
données avant



Exemple 3: 2 facteurs présence d'interaction

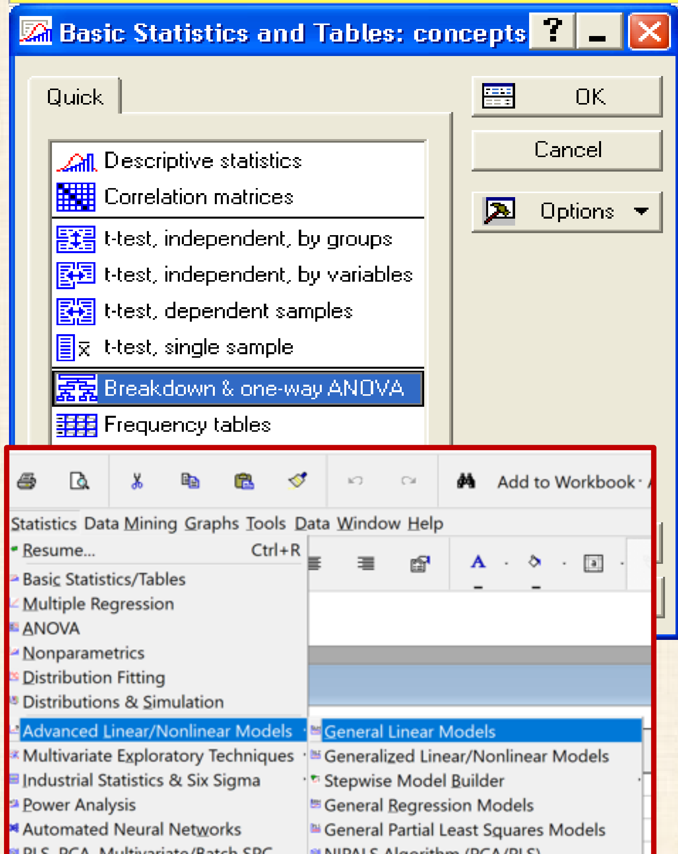
ANOVA							
Source	df	réponse Ynew SS	réponse Ynew MS	réponse Ynew F	réponse Z SS	Réponse Z MS	Réponse Z F
Intercept	1	588	588	588	588	588	588
groupe	1	192	192	192	48	48	48
sexe	1	108	108	108	108	108	108
groupe*sexe	1	48	48	48	192	192	192
erreur	8	8	1		8	1	
Total	11	356			356		





modules Statistica pour analyse des modèles d'analyse de variance

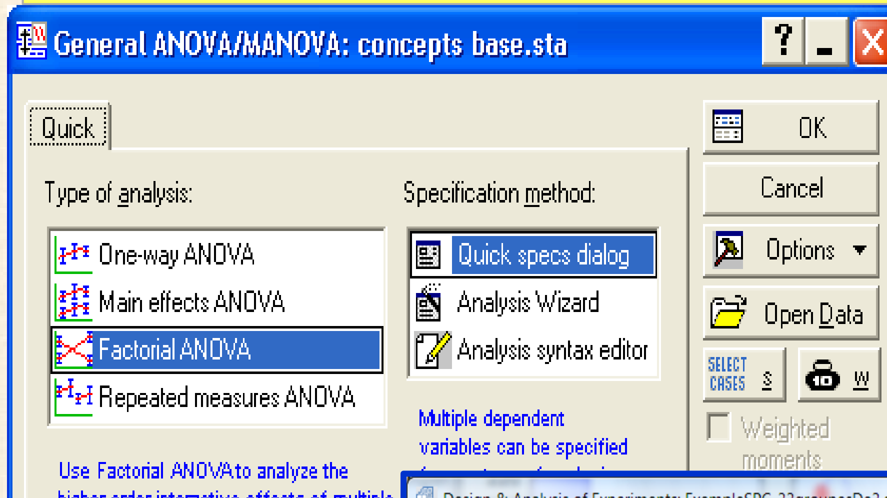
Basic Statistics / Tables …
… Breakdown & one-Way ANOVA
un facteur seulement

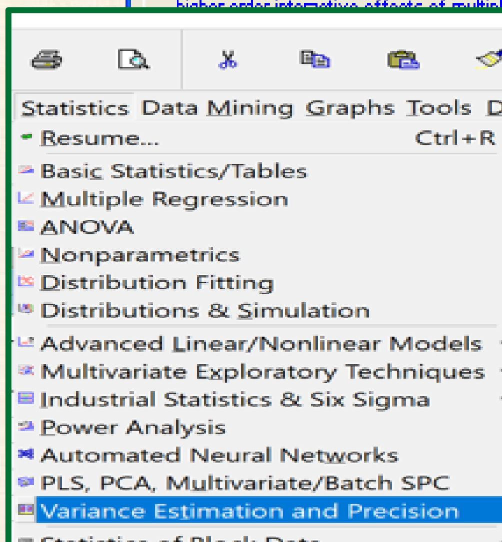


Advanced Linear/NonLinear Models
… General Linear Models (GLM)

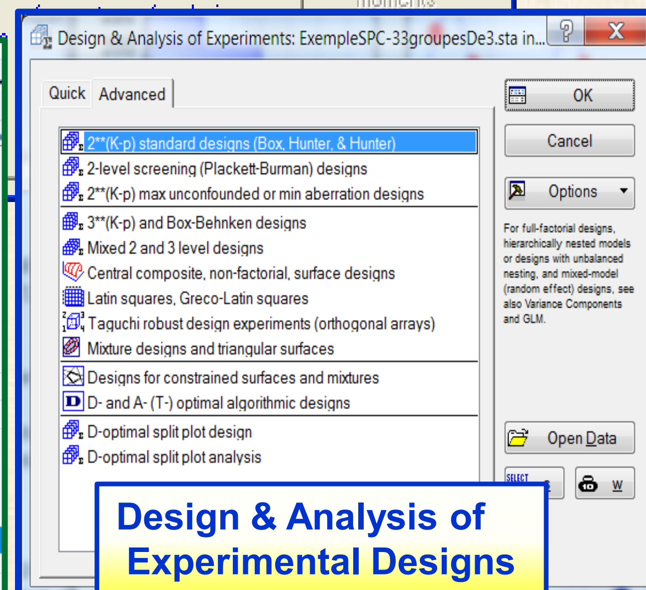
détails page suivante

General ANOVA/MANOVA 4 facteurs ou moins)





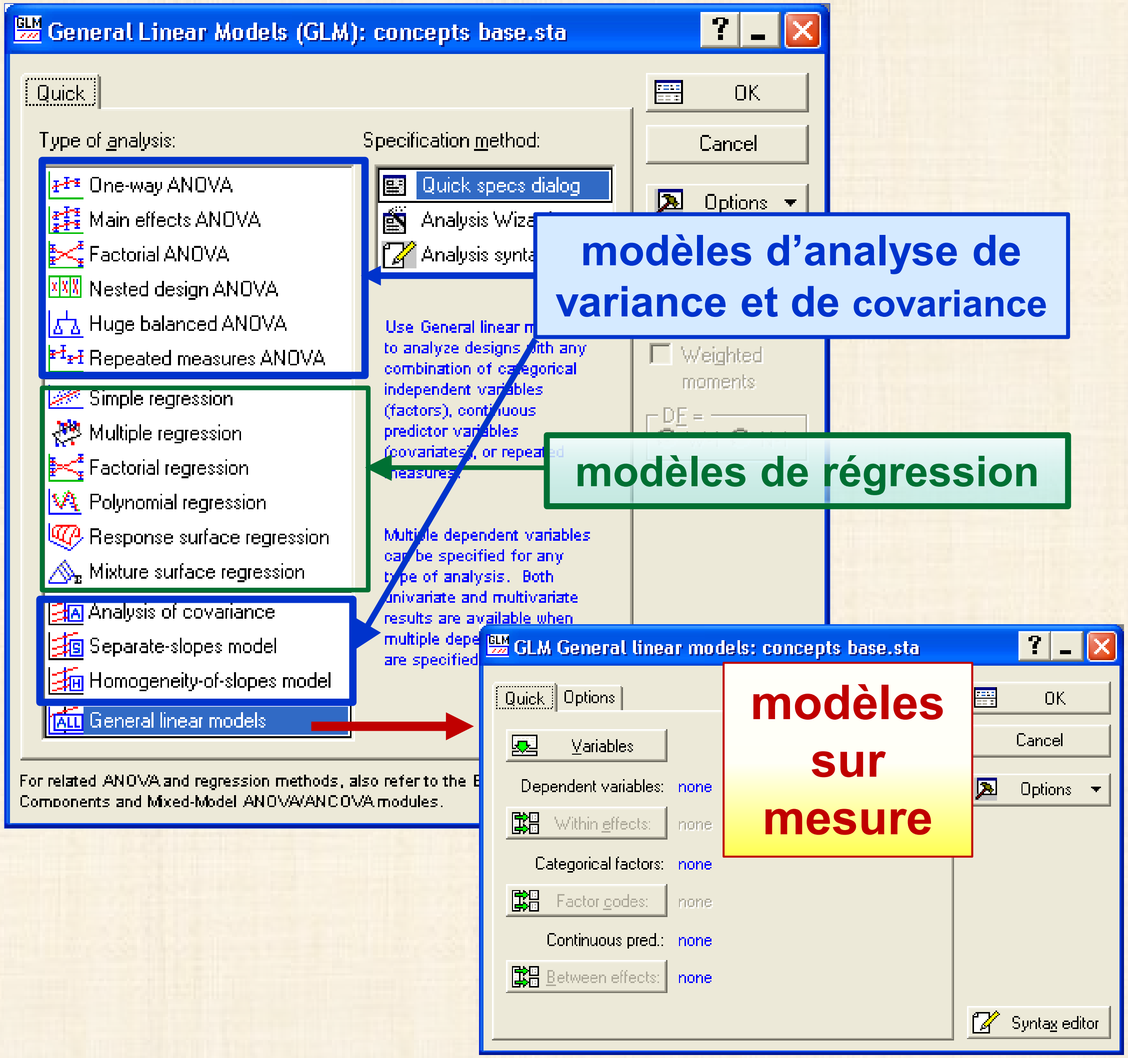
VEPAC modèles mixtes
facteurs aléatoires
+ facteurs fixés



Design & Analysis of Experimental Designs

modules Statistica pour modèles d'analyse de variance (GLM)

modules JMP modèles d'analyse de variance



- One-Way Analysis of Variance
- Two-Way Analysis of Variance
- Two-Way Analysis of Variance with Interaction
- Three-Way Full Factorial
- Analysis of Covariance, Equal Slopes
- Analysis of Covariance, Unequal Slopes
- Two-Factor Nested Random Effects Model
- Three-Factor Fully Nested Random Effects Model
- Simple Split Plot or Repeated Measures Model
- Two-Factor Response Surface Model

modèles d'analyse de variance et de covariance

Between-Subject Designs

- One way ANOVA
- Main effect ANOVA
- Factorial ANOVA
- Nested designs
- Balanced ANOVA
- ~~Simple regression~~
- ~~Multiple regression~~
- ~~Factorial regression~~
- ~~Polynomial regression~~
- ~~Response surface regression~~
- ~~Mixture surface regression~~
- Analysis of covariance (ANCOVA)
- Separate slopes designs
- Homogeneity of slopes
- Mixed-model

autre distinction
Whithin-subject / Between-Subject

Whithin-Subject Designs

chaque Sujet est mesuré plusieurs fois à plusieurs époques (temps) ou selon des conditions différentes; dans les réponses observées, il y a un facteur enfoui dans les réponses (le temps, en général).
Par exemple, des expériences comparatives de type avant-après ou le suivi temporel des unités expérimentales (étude longitudinale)

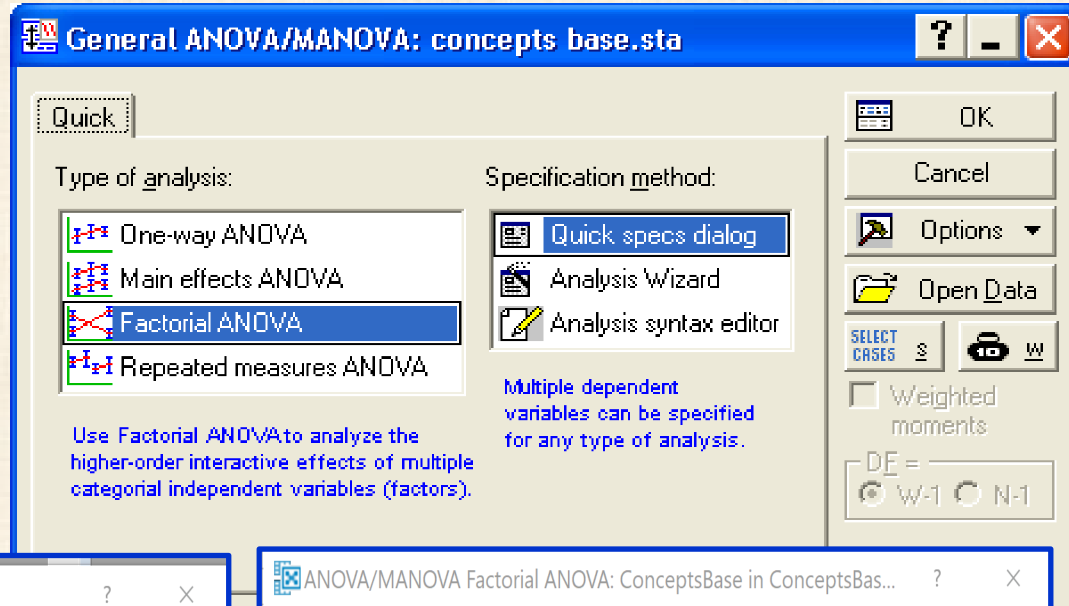
- One-way / Multi-way
- Repeated Measures …. connue aussi sous le nom mesures longitudinales (dans le temps)

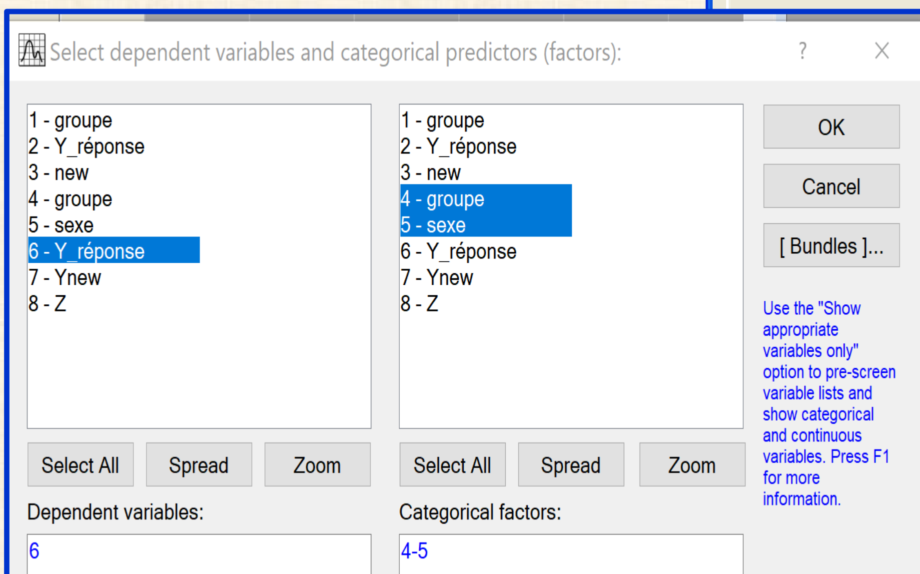
Between-Subject Designs

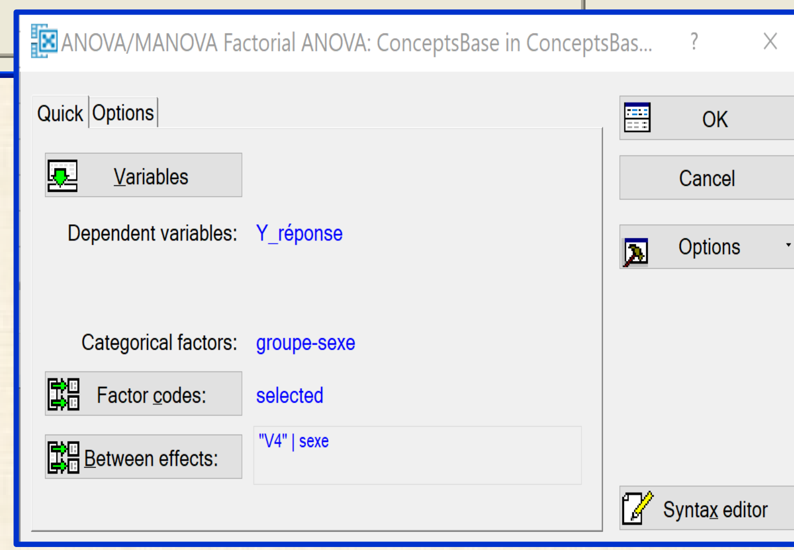
chaque sujet (unité expérimentale) est mesuré une seule fois; les traitements (combinaisons de facteurs) varient d'un sujet à l'autre.

ANALYSE : exemple 3 d'introduction

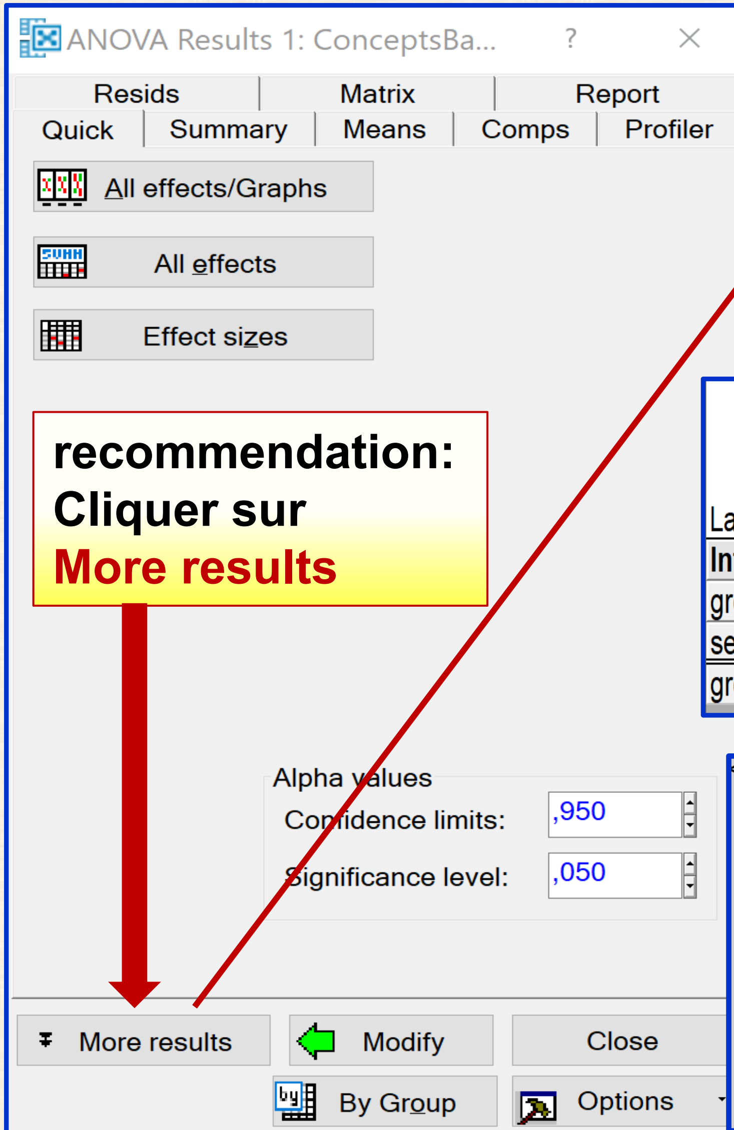
4 groupe	5 sexe	6 Y_réponse	7 Ynew	8 Z
a	hom	1	1	1
a	hom	2	2	2
a	hom	3	3	3
b	hom	5	5	5
b	hom	6	6	6
b	hom	7	7	7
a	fem	3	3	15
a	fem	4	4	17
a	fem	5	5	16
b	fem	7	15	3
b	fem	8	17	4
b	fem	9	16	5



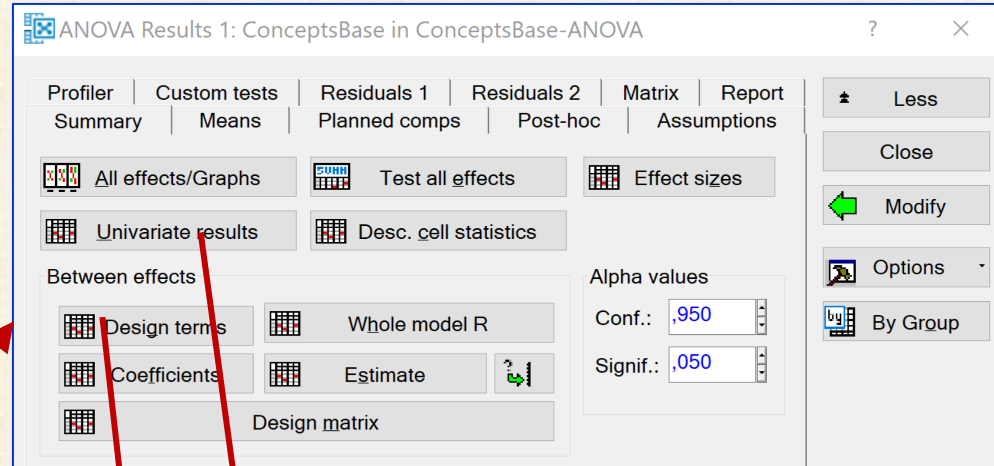




ANALYSE de l'exemple 3 d'introduction



recommendation:
Cliquer sur
More results



Column Labels (ConceptsBase in ConceptsBase-ANOVA)
Labels for the columns of the design matrix X

Label	Column	Variable	Level of Variable	versus Level	Variable	Level of Variable	versus Level
Intercept	1						
groupe	2	groupe	a	b			
sexe	3	sexe	hom	fem			
groupe*sexe	4	groupe	a	b	sexe	hom	fem

Univariate Results for Each DV (ConceptsBase in ConceptsBase-ANOVA)
Sigma-restricted parameterization
Effective hypothesis decomposition

Effect	Degr. of Freedom	Y_réponse SS	Y_réponse MS	Y_réponse F	Y_réponse p
Intercept	1	300,0000	300,0000	300,0000	0,000000
groupe	1	48,0000	48,0000	48,0000	0,000121
sexe	1	12,0000	12,0000	12,0000	0,008516
groupe*sexe	1	0,0000	0,0000	0,0000	1,000000
Error	8	8,0000	1,0000		
Total	11	68,0000			

MODÈLES LINÉAIRES STATISTIQUES et variables catégoriques

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \qquad \varepsilon \sim N(0, \sigma^2)$

Y : réponse $X_1, X_2, \ldots, X_k$: variables explicatives catégoriques

différentes méthodes pour le codage des variables catégoriques

Codage : méthode 1 - disjonctif complet

1 variable catégorique M avec k modalités m1 m2 … mk peut être représentée par k variables indicatrices Xi, chacune prenant la valeur 1 si modalité = mi et 0 autrement

attention : multicolinéarité car ∑Xi = 1 - utilisation dans certains modèles ANCOVA

Codage : méthode 2 - disjonctif incomplet : méthode 1 avec une variable en moins

1 variable catégorique M avec k modalités m1 m2 … mk peut être représentée par k – 1 variables indicatrices, chacune prenant les valeurs 0 et 1

Codage : méthode 3 – codage à effet - le plus employé et recommandé

1 variable catégorique M avec k modalités m1 m2 … mk peut être représentée par k -1 variables X_i prenant les valeurs 1 / 0 / - 1

UNE modalité de référence est choisie disons mk :

Xi = 1 si M = mi Xi = 0 si M = mj j ≠ i Xi = - 1 si M = mk

Cas particulier méthode 3 : codage à effet si k = 2

UNE SEULE variable quantitative X est nécessaire avec 2 valeurs -1 et 1

X = -1 si m1 et X = 1 si m2 valeur 0 pas nécessaire

Méthodes de codage facteurs catégoriques

Exemple 1: Y : ventes influence des 3 facteurs sur Y

X_1 : dépenses publicité - facteur continu

F : incorporation (oui, non) - facteur catégorique

M : management (expérimenté, novice) - facteur catégorique

Situation de modèle d'analyse de covariance

codage de F $X_2 = 1$ si incorporée et $X_2 = -1$ si non incorporé

Codage de M $X_3 = 1$ si expérimenté et $X_3 = -1$ si novice

modèle 1 : 3 effets principaux $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$

modèle 2 : général = 3 effets principaux + 3 interactions

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_2 + \beta_5 X_1 X_3 + \beta_6 X_2 X_3$$

modèle 3 : covariance $\beta_4 = \beta_5 = 0$ pas d'inter entre X1*F et X1*M

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_6 X_2 X_3 + \varepsilon$$

modèle 1 : $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 = \gamma_k + \beta_1 X_1$

k	F	X2	M	X3
1	oui	1	expérimenté	1
2	non	-1	expérimenté	1
3	non	1	novice	-1
4	oui	-1	novice	-1

$Y = \gamma_k + \beta_1 X_1$

$\gamma_1 = \beta_0 + \beta_2 + \beta_3$

$\gamma_2 = \beta_0 - \beta_2 + \beta_3$

$\gamma_3 = \beta_0 + \beta_2 - \beta_3$

$\gamma_4 = \beta_0 - \beta_2 - \beta_3$

Méthodes de codage facteurs catégoriques

Exemple 2: Y : usure outil X_1 : vitesse d'opération
M : manufacturier (M_1, M_2, M_3, M_4)

codage 1 de M : 3 variables X2 X3 X4 à valeurs (0/1)

M	X_1	X_2	X_3	X_4	$X_2 + X_3 + X_4$
M_1	x_{i1}	1	0	0	1
M_2	x_{i1}	0	1	0	1
M_3	x_{i1}	0	0	1	1
M_4	x_{i1}	0	0	0	0

modèle $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$

Si $M = M_4$: $Y = \beta_0 + \beta_1 X_1$

β_0 n'est plus l'effet général mais l'effet de M4

aspect moins intérressant de ce codage

Si M = M1 : $Y = (\beta_0 + \beta_2) + \beta_1 X_1 = \gamma_1 + \beta_1 X_1$

Si $M = M_2$: $Y = (\beta_0 + \beta_3) + \beta_1 X_1 = \gamma_2 + \beta_1 X_1$

Si $M = M_3$: $Y = (\beta_0 + \beta_4) + \beta_1 X_1 = \gamma_3 + \beta_1 X_1$

4 droites avec pentes égales : β_1

test de M : test d'égalité des 3 γ H_{01} : $\gamma_1 = \gamma_2 = \gamma_3$

+ test de nullité H_{02} : $\beta_0 = 0$

Méthodes de codage facteurs catégoriques

Exemple 2: Y : usure outil
X_1 : vitesse d'opération
M : manufacturier (M_1, M_2, M_3, M_4)

codage 2 de M : disjonctif complet (0 / 1) X2 X3 X4 X5

M	X_2	X_3	X_4	X_5	$X_2 + X_3 + X_4 + X_5$
M_1	1	0	0	0	1
M_2	0	1	0	0	1
M_3	0	0	1	0	1
M_4	0	0	0	1	1

modèle $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + e$

variables X2 X3 X4 X5 sont colinéaires

$$X_2 + X_3 + X_4 + X_5 = 1$$

modèle est sur paramétré :

imposible d'estimer les coefficients β

Méthodes de codage facteurs catégoriques

Exemple 2: Y : usure outil
X_1 : vitesse d'opération
M : manufacturier (M_1, M_2, M_3, M_4)

codage 3 : codage à effets: 1 / 0 / -1 création 3 variables X2 X3 X4

M	X_2	X_3	X_4
M_1	1	0	0
M_2	0	1	0
M_3	0	0	1
M_4	-1	-1	-1

M4 = modalité référence

modèle $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + e$

variables X2 X3 X4 ne sont pas colinéaires

Interprétation de β_2 β_3 β_4 ?

β_2 différence d'effet sur Y entre M1 et M4
β_3 différence d'effet sur Y entre M2 et M4
β_4 différence d'effet sur Y entre M3 et M4
test de l'effet de M : test de nullité de $\beta_2 = \beta_3 = \beta_4 = 0$
possibilité d'ajouter effets interactions avec d'autres facteurs

Exemples de modèles d'analyse de variance

Example 3 1 facteur

3 fertilisants employés sur des plantes individuelles.

3 modalités (niveaux) du facteur A *Fertilisant.*

codage de A: 2 variables de codage à effet (1 / 0 / -1)

la matrice X

$$X = \begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \end{array} \begin{array}{c} \begin{array}{ccc} X_0 & X_1 & X_2 \end{array} \\ \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{bmatrix} \end{array}$$

Exemple 4 2 facteurs

étude sur garçons et filles dans 2 groupes d'âge

2 (genre) X 2 (groupe d'âge).

design factoriel complet : toutes les combinaisons des modalités des variables catégoriques. X1 X2

cas : de 2 variables catégoriques à 2 modalités,

matrice *X* matrix pour ce design est avec

X3 = X1 * X2 effet d'interaction

$$X = \begin{array}{c} \\ A_1B_1 \\ A_1B_2 \\ A_2B_1 \\ A_2B_2 \end{array} \begin{array}{c} \begin{array}{cccc} X_0 & X_1 & X_2 & X_3 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \end{array}$$

Exemples de modèles d'analyse de variance

Example 5 : 3 facteurs

3 facteurs catégoriques A, B, C chacun avec 2 modalités

A = (A1, A2) representé par variable X1

B = (B1, B2) représenté par variable X2

C = (C1, C2) représenté par variable X3

design factoriel 2 x 2 x 2

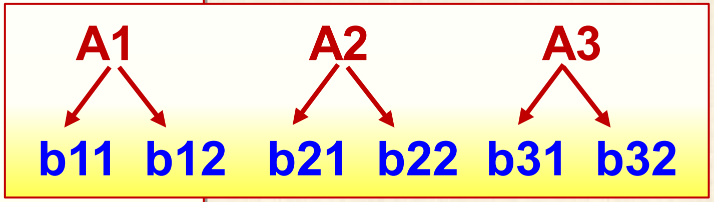
modèle : effets principaux + effets interactions d'ordre 2

	gen	X1	X2	X3	X1*X2	X1*X3	X2*X3
A1B1C1	1	1	1	1	1	1	1
A1B1C2	1	1	1	-1	1	-1	-1
A1B2C1	1	1	-1	1	-1	1	-1
A1B2C2	1	1	-1	-1	-1	-1	1
A2B1C1	1	-1	1	1	-1	-1	1
A2B1C2	1	-1	1	-1	-1	1	-1
A2B2C1	1	-1	-1	1	1	-1	-1
A2B2C2	1	-1	-1	-1	1	1	1

Exemples de modèles d'analyse de variance

Exemple 6   2 facteurs A et B - B emboité dans A

facteur A : 3 modalités (A1, A2, A3)
facteur B : 2 modalités distinctes
pour chaque modalité de A
si A = A1 alors B = b11 et b12
si A = A2 alors B = b21 et b22
si A = A3 alors B = b31 et b32



facteur A codé avec X1 X2 X3   disjonctif complet (1 / 0)
facteur B codé avec X4 X5 X6 X7 X8 X9
design : 3 effets principaux pour A + 6 effets principaux B

$$\mathbf{X} = \begin{array}{c} \\ A1b11 \\ A1b12 \\ A2b21 \\ A2b22 \\ A3b31 \\ A3b32 \end{array} \begin{array}{c} \begin{array}{cccccccccc} X_0 & X_1 & X_2 & X_3 & X_4 & X_5 & X_6 & X_7 & X_8 & X_9 \end{array} \\ \left[\begin{array}{cccccccccc} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \end{array}$$

Exemples de modèles d'analyse de variance

Exemple 7: design d'analyse de covariance
variable catégorique A avec 3 modalité + variable continue X

Modèle 1 : pas interaction entre A et X
pentes égales = droites parallèles
A facteur catégorique
avec 3 modalités
codé avec X2 et X3
codage à effet (1 / 0 / -1)

X1 = X facteur continu

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$

si A = A1
$Y = \beta_0 + \beta_1 X_1 + \beta_2 \qquad = \gamma 1 + \beta_1 X_1$

si A = A2
$Y = \beta_0 + \beta_1 X_1 + \beta_3 \qquad = \gamma 2 + \beta_1 X_1$

si A = A3
$Y = \beta_0 + \beta_1 X_1 - \beta_2 - \beta_3 \quad = \gamma 3 + \beta_1 X_1$

$$X = \begin{array}{cccc} X_0 & X_1 & X_2 & X_3 \\ \end{array}$$
$$X = \begin{bmatrix} 1 & 7 & 1 & 0 \\ 1 & 4 & 1 & 0 \\ 1 & 9 & 0 & 1 \\ 1 & 3 & 0 & 1 \\ 1 & 6 & -1 & -1 \\ 1 & 8 & -1 & -1 \end{bmatrix}$$

Exemples de modèles d'analyse de variance

Exemple 7 : facteur catégorique A avec 3 modalités + facteur continue X1
modèle d' analyse de covariance avec interaction entre A et X1
Conséquence : modèle avec pentes distinctes

hypothèses

facteur A : codage disjonctif complet
avec 3 variables X2 X3 X4 à valeurs (1/0)

facteur continu représenté par X1

interaction entre A et X1

modèle

$$Y = \beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_3 X3 + \beta_4 X4 + \beta_5 X1X2 + \beta_6 X1X3 + \beta_7 X1X4$$

si A = A1

$$Y = \beta_0 + \beta_1 X1 + \beta_2 + \beta_5 X1$$
$$= (\beta_0 + \beta_2) + (\beta_1 + \beta_5) X1 = \gamma 1 + \gamma 2\, X1$$

si A = A2

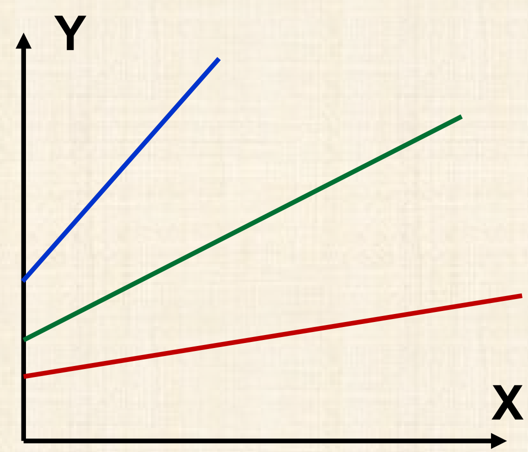
$$Y = (\beta_0 + \beta_3) + (\beta_1 + \beta_6) X1 = \gamma 3 + \gamma 4\, X1$$

si A = A3

$$Y = (\beta_0 + \beta_4) + (\beta_1 + \beta_7) X1 = \gamma 5 + \gamma 6\, X1$$

3 droites avec pentes distinctes
et 3 valeurs distinctes a X = 0

X0	X1	X2	X3	X4	X1X2	X1X3	X1X4
1	7	1	0	0	7	0	0
1	4	1	0	0	4	0	0
1	9	0	1	0	0	9	0
1	3	0	1	0	0	3	0
1	6	0	0	1	0	0	6
1	8	0	0	1	0	0	8



Autres applications variables indicatrices et codages

- **modèle de régression par morceaux**
- **Tenir en compte des facteurs d'intervention dans séries chronologiques**
- **Simplification structure des banques de données en remplacement des valeurs par quelques valeurs types**

  - variables continues : points milieux d'intervalles

  - variables catégoriques : regroupement de catégories
- **perte d'information mais compréhension facilitée des données**

Quelques exemples STATISTICA - modèles ANOVA

- Analyses ANOVA ==========
- UN FACTEUR=============
  - Concepts base.sta
  - Formation.sta
  - Hauteur.sta
  - Rouille.sta
  - Flux.sta
  - Flux2.sta
  - Pannes.sta
  - Emballage.sta
- DEUX FACTEURS et PLUS =====
  - Etalage.sta
  - Apprentissage-2F.sta
  - Apprentissage-3F.sta
  - Stress.sta
  - Carré latin.sta
  - BoisTraite.sta
  - Croissance os.sta
  - Ventes.sta
- FACTEURS EMBOITÉS =======
  - 2 facteurs emboite.sta
  - 4 facteurs-2 facteurs emboités.sta
- MESURES RÉPÉTÉES =
  - BodyWeights.sta
  - Wine.sta
  - BodyPart.sta
  - BloodFlow.sta
  - Wheat.sta
  - Dials.sta
  - Cadran.sta
  - Publicité.sta
  - Diètes.sta
  - Opérateurs.sta
  - Cholesterol.sta
- FACTEURS ALÉATOIRES ======
  - 1 facteur aléatoire-Loom.sta
  - 2 facteurs aléatoires-1 fixé +1 aléa.sta
  - 2 facteurs aléatoires-étude R&R.sta
  - 3 facteurs emboités.sta
  - 4 facteurs-2 facteurs emboités.sta
  - Papier (splitPlot).sta
  - 1 facteur aléatoire-2 réponses (Hays484).sta
  - 2 facteurs aléatoires (Miliken238).sta
  - 4 facteurs-modèle mixte-covariable-3 Y-me
  - 2 facteurs-modèle mixte- ProductivityScore
  - 4 facteurs-2 facteurs emboités-Comfort.sta

exemples - concepts de base

1 ID	2 groupe	3 réponse Y	4 new	5 groupe	6 sexe	7 réponse Y2	8 new	9 groupe	10 sexe	11 réponse Y3	12 réponse Z
1	a	1		a	hom	1		a	hom	1	1
2	a	2		a	hom	2		a	hom	2	2
3	a	3		a	hom	3		a	hom	3	3
4	b	5		b	hom	5		b	hom	5	5
5	b	6		b	hom	6		b	hom	6	6
6	b	7		b	hom	7		b	hom	7	7
7				a	fem	3		a	fem	3	15
8				a	fem	4		a	fem	4	17
9				a	fem	5		a	fem		
10				b	fem	7		b	fem		
11				b	fem	8		b	fem		
12				b	fem	9		b	fem		

1 ID	2 Sport	3 i	4 j	5 X1	6 X2	7 X3	8 Y-hauteur saut (cm)
1	Soccer	1	1	1	0	0	38
2	Soccer	1	2	1	0	0	43
3	Soccer	1	3	1	0	0	33
4	Soccer	1	4	1	0	0	40
5	Soccer	1	5	1	0	0	35
6	Tennis	2	1	0	1	0	45
7	Tennis	2	2	0	1	0	53
8	Tennis	2	3	0	1	0	38
9	Tennis	2	4	0	1	0	55
10	Tennis	2	5	0	1	0	50
11	Football	3	1	0	0	1	55
12	Football	3	2	0	0	1	68
13	Football	3	3	0	0	1	43
14	Football	3	4	0	0	1	45
15	Football	3	5	0	0	1	53
16	Basketball	4	1	-1	-1	-1	60
17	Basketball	4	2	-1	-1	-1	65
18	Basketball	4	3	-1	-1	-1	55
19	Basketball	4	4	-1	-1	-1	53
20	Basketball	4	5	-1	-1	-1	55

1 ID	2 rep	3 Genre	4 Age	5 IQ	6 Y1	7 Y2	8 Y3
1	1	homme	jeune	élevé	9	10,5	10,0
2	1	homme	moyen	élevé	12	12,5	12,0
3	1	homme	agé	élevé	18	16,0	15,5
4	1	homme	jeune	normal	19	17,5	18,0
5	1	homme	moyen	normal	20	19,5	20,0
6	1	homme	agé	normal	21	23,0	23,5
7	1	femme	jeune	élevé	9	9,5	10,0
8	1	femme	moyen	élevé	10	10,5	11,0
9	1	femme	agé	élevé	14	13,0	13,5
10	1	femme	jeune	normal	19	18,5	18,0
11	1	femme	moyen	normal	20	19,5	19,0
12	1	femme	agé	normal	21	22,0	21,5
13	2	homme	jeune	élevé	8	9,5	9,0
14	2	homme	moyen	élevé	11	11,5	11,0
15	2	homme	agé	élevé	17	15,0	14,5
16	2	homme	jeune	normal	18	16,5	17,0
17	2	homme	moyen	normal	19	18,5	19,0
18	2	homme	agé	normal	20	22,0	22,5
19	2	femme	jeune	élevé	8	8,5	9,0
20	2	femme	moyen	élevé	9	9,5	11,0
21	2	femme	agé	élevé	13	12,0	12,5
22	2	femme	jeune	normal	18	17,5	17,0
23	2	femme	moyen	normal	19	18,5	18,0
24	2	femme	agé	normal	20	21,0	20,5
25	3	homme	jeune	élevé	10	11,5	11,0
26	3	homme	moyen	élevé	13	13,5	13,0
27	3	homme	agé	élevé	19	17,0	16,5
28	3	homme	jeune	normal	20	18,5	19,0
29	3	homme	moyen	normal	21	20,5	21,0
30	3	homme	agé	normal	22	24,0	24,5
31	3	femme	jeune	élevé	10	10,5	11,0
32	3	femme	moyen	élevé	11	11,5	12,0
33	3	femme	agé	élevé	15	12,0	14,5
34	3	femme	jeune	normal	20	19,5	19,0
35	3	femme	moyen	normal	21	20,5	20,0
36	3	femme	agé	normal	22	23,0	22,5

1 ID	2 ville	3 intervalle	4 Y_durée entre pannes	5 Y_Rang	6 logY
10	B	5	1,61	1	0,207
1	A	1	4,41	2	0,644
8	B	3	7,35	3	0,866
6	B	1	8,24	4	0,916
9	B	4	12,29	5	1,090
3	A	3	14,45	6	1,160
12	C	2	33,83	7	1,529
15	C	5	44,33	8	1,647
4	A	4	47,13	9	1,673
13	C	3	78,88	10	1,897
7	B	2	81,16	11	1,909
5	A	5	85,21	12	1,930
2	A	2	100,65	13	2,003
11	C	1	106,19	14	2,026
14	C	4	342,81	15	2,535

Quelques exemples STATISTICA - modèles ANOVA

Contexte et données de l'étude CCA - 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 : Retention, Solution et
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 Y-Load	7 v7	8 trait	9 trait2	10 c10
1	0,25	25	CCA	1,0	15,597		25cca	1	
2	0,25	25	CCA	2,0	14,827		25cca	1	
3	0,25	25	CCA	3,0	13,037		25cca	1	
4	0,25	25	CCA	4,0	12,220		25cca	1	

Milliken and Johnson, vol 1 p.418 facteurs emboités (nested)
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 degrees Fahrenheit.
Each temperature setting was randomly assigned to 3 of 9 available environmental chambers.
18 males and 18 females were randomly assigned to chambers so that 2 males and 2 females
were assigned to each of the 9 chambers.
After the people were subjected to the environmental condition for three hours, their comfort was measured.
Y_confort scores: 1 = cold 8 = confortable 15 = hot
model Y_Comfort = Temperature + Gender + Temperature*Gender + Chamber(Temperature) + Person(Chamber) + error
//

1 ID	2 Temperature	3 Chamber	4 Gender	5 Person	6 Y_Comfort	7 c7
1	65	1	Male	1	5	
2	65	1	Male	2	4	
3	65	1	Female	3	1	
4	65	1	Female	4	2	
5	65	2	Male	5	5	

NCSS 11 Statistical Software (2016). NCSS, LLC. Kaysville, Utah, USA, ncss.com/software/ncss.
mesure du niveau de cholesterol Y durant une période de 3 mois (april, may, june) 120 obs.
selon 4 types de traitements : A B control placebo - 20 patients p1 p2 ... p20
//

1 ID	2 Patient	3 Treatment	4 Time	5 Month	6 AM/PM	7 Days	8 Y
1	p1	A	April AM	April	AM	0,0	278,0
2	p1	A	April PM	April	PM	0,5	280,0
3	p1	A	May AM	May	AM	30,0	204,0
4	p1	A	May PM	May	PM	30,5	208,0
5	p1	A	June AM	June	AM	61,0	171,3
6	p1	A	June PM	June	PM	61,5	175,0
7	p2	A	April AM	April	AM	0,0	278,0

4 facteurs: 2 fixes (group, gender) 2 aléatoires (time, paid)
covariable = stress
répétition = 2
2 groups X 2 gender X 3 times X2 paid X 2 rep = 48 obs
3 réponses: correct1, correcte2, correct3
modèle mixte et Analyse de covariance

1 ID	2 GROUP	3 GENDER	4 TIME	5 PAID	6 rep	7 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

Quelques exemples JMP - modèles ANOVA

- 2 Factors Crossed.jmp
- 2 Factors Nested.jmp
- 2x3x4 Factorial.jmp
- 3 Factors Crossed & Nested.jmp
- 3 Factors Crossed.jmp
- 3 Factors Nested & Crossed.jmp
- 3 Factors Nested.jmp
- Animals Subset.jmp
- Animals.jmp
- Auto Raw Data.jmp
- Bacteria.jmp
- Big Class Families.jmp
- Big Class.jmp
- Bladder Cancer.jmp
- Drug Measurements.jmp
- Drug Toxicity.jmp
- Drug.jmp
- DrugLBI.jmp
- Exercise.jmp

. . . .

- Reactor 8 Runs.jmp
- Reactor 20 Custom.jmp
- Reactor 32 Runs.jmp
- Reactor Augment Data.jmp
- Reactor Augment Design.jmp
- Reactor Factors.jmp
- Reactor Half Fraction.jmp
- Reactor Response.jmp
- Reactor.jmp
- Singularity.jmp
- Tool Wear.jmp
- Tool-Life.jmp

	Pattern	Feed Rate	Catalyst	Stir Rate	Temperature	Concentration	Percent Reacted
1	++-+-	15	2	100	180	3	93
2	+++++	15	2	120	180	6	82
3	-++--	10	2	120	140	3	54
4	-+--+	10	2	100	140	6	70
5	-++++	10	2	120	180	6	81
6	-+---	10	2	100	140	3	63
7	++--+	15	2	100	140	6	65
8	--+-+	10	1	120	140	6	59

	species	subject	miles	season
1	FOX	1	0	fall
2	FOX	1	0	winter
3	FOX	1	5	spring
4	FOX	1	3	summer
5	FOX	2	3	fall
6	FOX	2	1	winter
7	FOX	2	5	spring
8	FOX	2	4	summer
9	FOX	3	4	fall
10	FOX	3	3	winter
11	FOX	3	6	spring
12	FOX	3	2	summer
13	COYOTE	1	4	fall
14	COYOTE	1	2	winter
15	COYOTE	1	7	spring
16	COYOTE	1	8	summer
17	COYOTE	2	5	fall
18	COYOTE	2	4	winter
19	COYOTE	2	6	spring
20	COYOTE	2	6	summer
21	COYOTE	3	7	fall
22	COYOTE	3	5	winter
23	COYOTE	3	8	spring
24	COYOTE	3	9	summer

	Run	F	Ct	A	T	Cn	Pattern	Y
1	1	-1	-1	-1	-1	-1	-----	61
2	2	1	-1	-1	-1	-1	+----	53
3	3	-1	1	-1	-1	-1	-+---	63
4	4	1	1	-1	-1	-1	++---	61
5	5	-1	-1	1	-1	-1	--+--	53
6	6	1	-1	1	-1	-1	+-+--	56
7	7	-1	1	1	-1	-1	-++--	54
8	8	1	1	1	-1	-1	+++--	61
9	9	-1	-1	-1	1	-1	---+-	69
10	10	1	-1	-1	1	-1	+--+-	61
11	11	-1	1	-1	1	-1	-+-+-	94
12	12	1	1	-1	1	-1	++-+-	93
13	13	-1	-1	1	1	-1	--++-	66
14	14	1	-1	1	1	-1	+-++-	60
15	15	-1	1	1	1	-1	-+++-	95
16	16	1	1	1	1	-1	++++-	98
17	17	-1	-1	-1	-1	1	----+	56
18	18	1	-1	-1	-1	1	+---+	63
19	19	-1	1	-1	-1	1	-+--+	70
20	20	1	1	-1	-1	1	++--+	65
21	21	-1	-1	1	-1	1	--+-+	59
22	22	1	-1	1	-1	1	+-+-+	55
23	23	-1	1	1	-1	1	-++-+	67
24	24	1	1	1	-1	1	+++-+	65
25	25	-1	-1	-1	1	1	---++	44
26	26	1	-1	-1	1	1	+--++	45
27	27	-1	1	-1	1	1	-+-++	78
28	28	1	1	-1	1	1	++-++	77
29	29	-1	-1	1	1	1	--+++	49
30	30	1	-1	1	1	1	+-+++	42
31	31	-1	1	1	1	1	-++++	81
32	32	1	1	1	1	1	+++++	82