Master de Psychologie sociale des représentations
PSR73C : Informatique - TD N°2
Analyse de variance

7. Analyse de variance à un facteur de variation. Plan S<A>.

Des groupes indépendants de sujets ont été soumis aux différents niveaux d'un facteur A. On souhaite tester l'effet des différents niveaux du facteur A sur le comportement des sujets, évalué à l'aide d'une variable dépendante X. Le modèle de score est ici :

Score = Moyenne Générale + Effet de A + Résidu aléatoire.

La forme générale du tableau d'analyse de variance correspondant est la suivante :

Sources de variation	Somme des Carrés	ddl	Carrés Moyens	F	p
Facteur A	SC_A	$a-1$	CM_A	$F_{Obs}=\frac{CM_A}{CM_{S(A)}}$	....
Résidu S(A)	$SC_{S(A)}$	$N-a$	$CM_{S(A)}$		
Total	SC_T	$N-1$			

La statistique F suit une loi de Fisher Snedecor à (a-1) et (N-a) degrés de liberté.

7.1. Première méthode

On reprend l'énoncé "Bransfor" :
On demande à des sujets d'écouter un texte dans quatre conditions expérimentales différentes :

Le but visé par Bransford et al. est de montrer l'importance du contexte dans la compréhension et la mémorisation d'un texte. Pour ce faire, ils utilisent quatre groupes expérimentaux:

- Un groupe "sans contexte" entend simplement le texte.
- Le groupe "avec contexte avant" regarde une figure suggérant un contexte approprié pendant qu'il entend le texte.
- Le groupe "avec contexte après" entend le texte puis regarde la figure précédente.
- Le groupe ``avec contexte partiel" regarde une figure suggérant un contexte inapproprié pendant qu'il entend le texte.

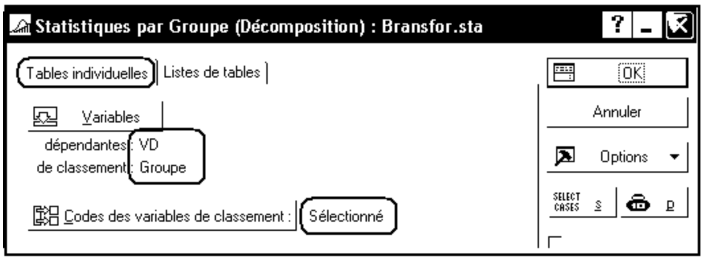
A proprement parler cette étude comprend un groupe expérimental (le groupe 2: contexte pendant) et trois groupes contrôles (les groupes 1, 3 et 4). Les groupes contrôles doivent permettre d'éliminer des explications concurrentes (en particulier, effet facilitateur sur la mémoire de l'imagerie, de l'aspect concret du matériel, etc.). L'expérimentateur s'attend, donc, à observer une performance pour le groupe 2 supérieure aux trois autres groupes.

Il choisit de mesurer le comportement des sujets par la variable dépendante "nombre d'idées correctement rappelées".

GR1	**GR2**	**GR3**	**GR4**
3	5	2	5
3	9	4	4
2	8	5	3
4	4	4	5
3	9	1	4

Définissez un nouveau classeur Statistica et insérez une nouvelle feuille de données dans ce classeur. Saisissez les données selon un plan d'expérience S<A> (c'est-à-dire, définissez une variable "Groupe" et une variable "Variable dépendante" ou "VD"). Enregistrez ensuite le classeur sous le nom Bransfor.stw.

Utilisez ensuite le menu Statistiques - Statistiques élémentaires - Décompositions et ANOVA à un facteur. Sélectionnez l'onglet "Tables individuelles" et indiquez les variables utilisées par l'analyse :

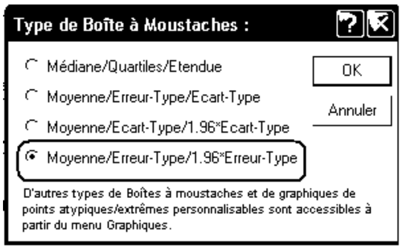


Sélectionnez l'onglet "Base" ou l'onglet "ANOVA et Tests" et cliquez sur le bouton "Analyse de variance". On obtient le résultat suivant :

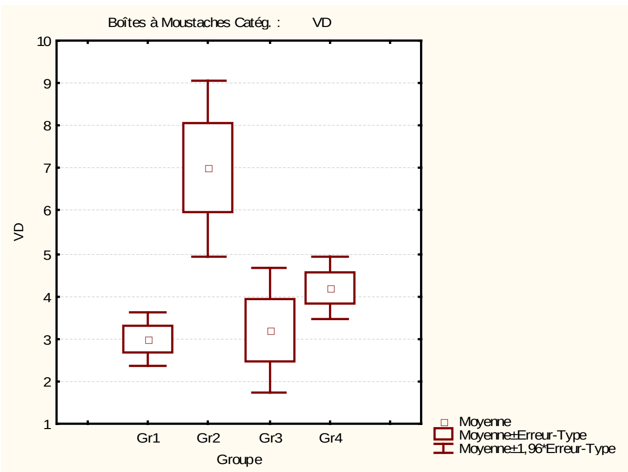
Variable	Analyse de la Variance (Bransfor.sta) Effets significatifs marqués à p < ,05000 SC Effet	dl Effet	MC Effet	SC Erreur	dl Erreur	MC Erreur	F	p
VD	50,95	3	16,98	37,60	16	2,35	7,23	0,002782

Tous les éléments du tableau d'analyse de variance sont présents, mais la disposition n'est pas celle qui est traditionnellement utilisée.

On peut aussi illustrer la situation à l'aide d'un graphique. Par exemple, sous l'onglet "Stats Descriptives", on pourra utiliser le bouton "Boîtes à moustaches catégorisées" et l'item "Moyenne/Erreur-Type/1.96*Erreur-Type" :



On obtient ainsi le graphique suivant :



7.2. Deuxième méthode

On utilise toujours le classeur Bransford.stw.

Utilisez le menu Statistiques - ANOVA, puis "ANOVA à un facteur" et "Spécifications rapides".

Dans le dialogue suivant, indiquez la variable dépendante et le facteur (facteur catégoriel dans la terminologie de Statistica), cliquez ensuite sur le bouton OK, puis le bouton "Tous les effets". Vous devriez obtenir le résultat suivant :

Effet	Tests Univariés de Significativité pour VD (Bransfor.sta) Paramétrisation sigma-restreinte Décomposition efficace de l'hypothèse SC	Degr. de Liberté	MC	F	p
ord. origine	378,4500	1	378,4500	161,0426	0,000000
Groupe	50,9500	3	16,9833	7,2270	0,002782
Erreur	37,6000	16	2,3500		

Lecture du résultat :

On reconnaît les colonnes "somme de carrés", "carrés moyens", "degrés de liberté" et F. Cependant, la présentation du résultat diffère de celle adoptée en cours (et utilisée par la plupart des autres logiciels). Dans la première ligne du tableau, la somme des carrés est égale à :

(moyenne générale)^2 x nombre d'observations

Le test de cette première ligne correspond à l'hypothèse nulle : $\mu = 0$, où μ désigne la moyenne de la VD, avant prise en compte de l'effet du facteur "Groupe".
Les deux lignes suivantes correspondent aux lignes "Inter-groupes" et "Intra-groupes" du tableau d'ANOVA classique. Enfin, Statistica n'affiche pas de ligne de synthèse. Il nous appartient donc de la reconstituer pour obtenir le tableau habituel :

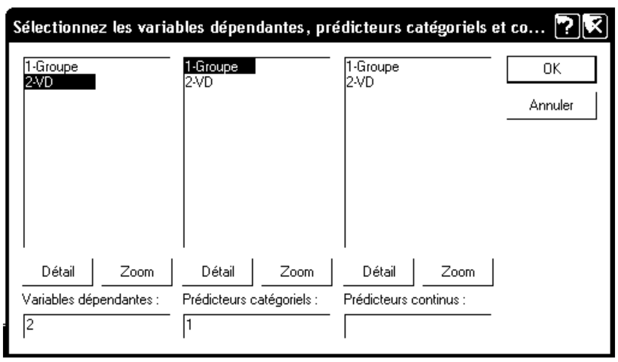
Sources de variation	Somme des Carrés	ddl	Carrés Moyens	F	p
Inter-groupes	50,9500	3	16,9833	7,2270	0,278%
Intra-groupes	37,6000	16	2,3500		
Total	88,5500	16	2,3500		

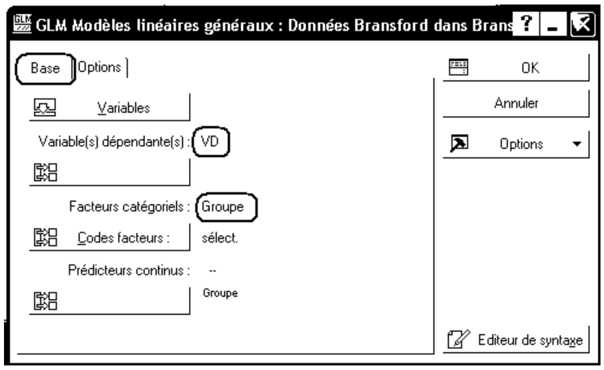
7.3. Troisième méthode

Tous les traitements d'analyse de variance que nous nous proposons d'étudier sont également disponibles dans le module Modèles linéaires / non linéaires avancés - Modèle linéaire général. Bien que l'interface soit un peu plus complexe, il peut être intéressant d'utiliser le même item de menu pour tous les traitements d'ANOVA que nous nous proposons d'étudier.

On utilise toujours le classeur Bransford.stw.

Utilisez le menu Statistiques - Modèles linéaires / non linéaires avancés, puis "Modèle linéaire général" et "Modèles linéaires généraux".
Dans le dialogue suivant, indiquez la variable dépendante et le facteur (facteur catégoriel dans la terminologie de Statistica):





Cliquez ensuite sur le bouton OK, puis le bouton "Tous les effets". On obtient ainsi le résultat sous une forme identique à celle obtenue avec la méthode 2.

8. Traitement d'un plan S*A. Plan à mesures répétées

Dans un plan S*A, ou plan à mesures répétées, un groupe de sujets a été soumis aux différents niveaux d'un facteur A (situation de groupes appareillés). On souhaite tester l'effet des différents niveaux du facteur A sur le comportement des sujets, évalué à l'aide d'une variable dépendante X. Le modèle de score est ici :

Score = Moyenne Générale + Effet de A + Effet "sujet" + Résidu aléatoire.

La forme générale du tableau d'analyse de variance correspondant est la suivante :

Sources de variation	Somme des Carrés	ddl	Carrés Moyens	F	p
Facteur A	SC_A	$a-1$	CM_A	$F_{Obs} = \frac{CM_A}{CM_{AS}}$	....
Facteur S	SC_S	$n-1$	CM_S		
Résidu AS	SC_{AS}	$(a-1)(n-1)$	CM_{AS}		
Total	SC_T	$an-1$			

La statistique F suit une loi de Fisher Snedecor à (a-1) et (a-1)(n-1) degrés de liberté.

8.1. Première méthode

Enoncé du cas:
Dans une expérimentation sur l'inhibition proactive, des sujets apprennent une liste de dix paires de mots, puis doivent se rappeler ces paires deux jours plus tard. Après le rappel, les sujets doivent apprendre une deuxième liste de dix paires dont ils devront se rappeler deux jours plus tard, le rappel de la deuxième liste est suivie de l'apprentissage d'une troisième, etc., jusqu'à la sixième liste. La variable indépendante sera la position ordinale de la liste (e.g., première, seconde, ... , sixième).

La variable dépendante sera le nombre de paires correctement rappelées. Les auteurs de l'expérience prédisent que le rappel se détériorera à mesure que l'on progresse dans la position ordinale (prédiction qui traduit simplement l'effet de l'inhibition proactive).

Ouvrez le classeur Inhibit.stw et rendez active la feuille Inhibit1.

Les données y sont saisies selon la logique "plan d'expérience" : une ligne par observation, les sujets sont considérés comme un facteur, au même titre que "position" :

	Sujet	Position	Nb Mots
1	s1	1	17
2	s2	1	14
3	s3	1	17
...			
9	s1	2	13
10	s2	2	18
11	s3	2	16
...			

Il s'agit là d'une expérience conduite selon un plan S*A, où A est le facteur "position" et S le facteur "sujet".
Pour traiter les données à l'aide de Statistica :
Utilisez le menu Statistiques - ANOVA.
Choisissez l'item ANOVA - Effets principaux .
Spécifiez Nb Mots comme variable dépendante, Sujet et Position comme prédicteurs.
Le résultat fourni par Statistica est le suivant :

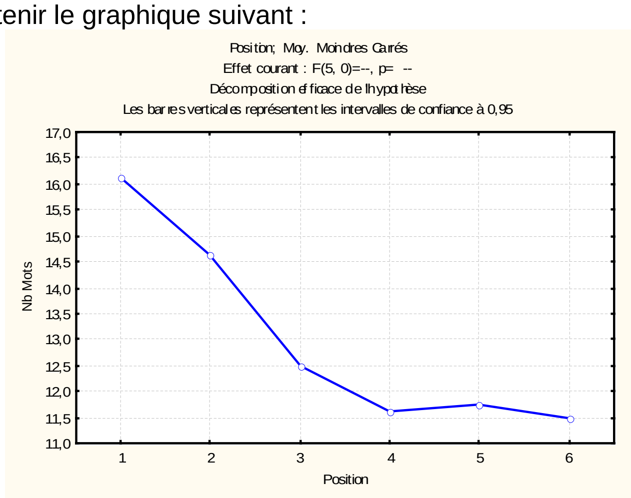
	Tests Univariés de Significativité pour Nb Mots (Données dans Inhibit.stw) Paramétrisation sigma-restreinte Décomposition efficace de l'hypothèse				
Effet	SC	Degr. de Liberté	MC	F	p
ord. origine	8138,021	1	8138,021	2858,431	0,000000
Sujet	52,479	7	7,497	2,633	0,026938
Position	146,854	5	29,371	10,316	0,000004
Erreur	99,646	35	2,847		

Ce résultat permet de recomposer le tableau d'analyse de variance attendu :

```
Analyse de variance pour Nb mots
Source        DL        SC         CM        F        P
Position       5    146,85      29,37    10,32    0,000
Sujet          7     52,48       7,50
Erreur        35     99,65       2,85
Total         47    298,98
```

Il peut également être intéressant de produire un graphique montrant les moyennes observées pour chacun des niveaux du facteur Position.

Dans la fenêtre de dialogue "Résultats", cliquez sur le bouton "Tous effets/Graphs" puis sélectionnez la ligne "Position". Vous devriez obtenir le graphique suivant :



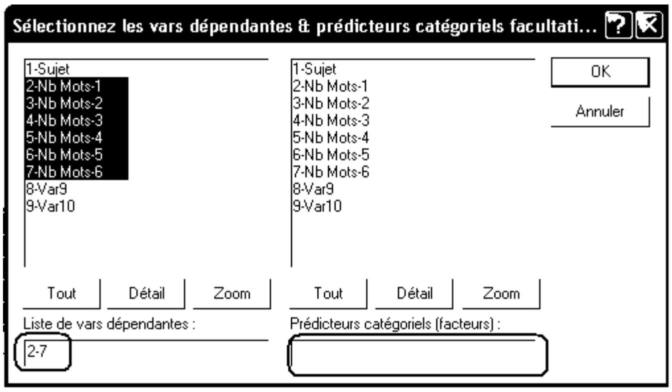
8.2. Deuxième méthode

Une autre façon de saisir les données est celle figurant dans la feuille Inhibit2 :

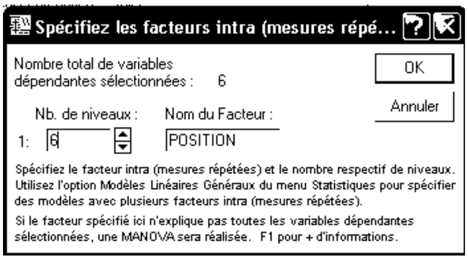
	Sujet	Nb Mots-1	Nb Mots-2	Nb Mots-3	Nb Mots-4	Nb Mots-5	Nb Mots-6
1	s1	17	13	12	12	11	11
2	s2	14	18	13	18	11	12
3	s3	17	16	13	11	15	14
4	s4	18	16	11	10	12	10
5	s5	17	12	13	10	11	13
6	s6	16	13	13	11	11	11
7	s7	14	12	10	10	10	10
8	s8	16	17	15	11	13	11

Chaque ligne correspond alors à un individu statistique (sujet). Autrement dit, les six scores relatifs à un même sujet se trouvent sur une même ligne.
Statistica peut également réaliser l'ANOVA sur des données structurées de cette façon.

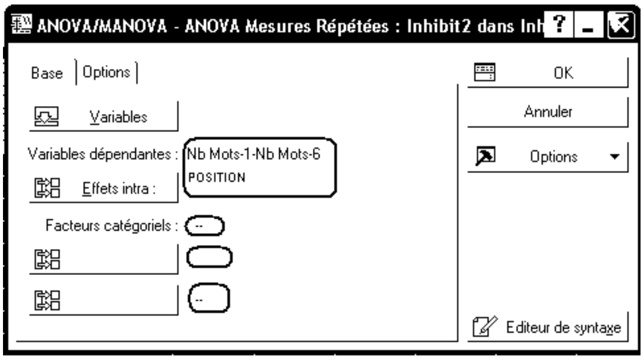
Rendez active la feuille Inhibit2.
Utilisez le menu Statistiques - ANOVA, puis l'item ANOVA - Mesures répétées.
Indiquez NB Mots-1 à Nb Mots-6 comme variables dépendantes, et n'indiquez pas de prédicteur catégoriel :



Validez et cliquez ensuite sur le bouton "Effets intra" ; indiquez que les variables Nb Mots-1 à Nb Mots-6 correspondent aux 6 niveaux du facteur POSITION :



Vous devriez obtenir la description suivante des facteurs impliqués dans l'analyse :



Statistica dispose alors le tableau d'analyse de variance de la façon suivante :

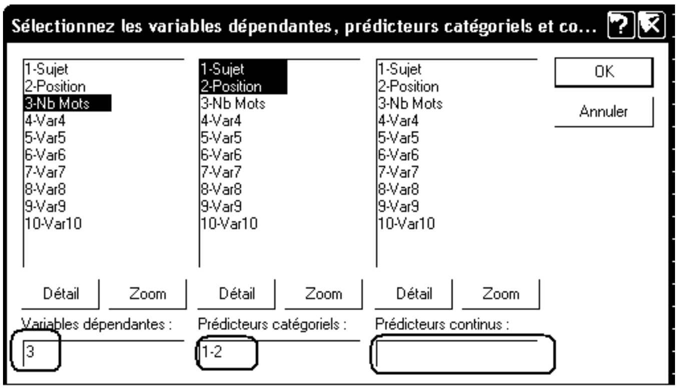
Effet	Analyse de la Variance Mesures Répétées (Données Inhibit dans Inhibit2.stw) Paramétrisation sigma-restreinte Décomposition efficace de l'hypothèse SC	Degr. de Liberté	MC	F	p
ord. origine	8138,021	1	8138,021	1085,500	0,000000
Erreur	52,479	7	7,497		
POSITION	146,854	5	29,371	10,316	0,000004
Erreur	99,646	35	2,847		

8.3. Troisième méthode

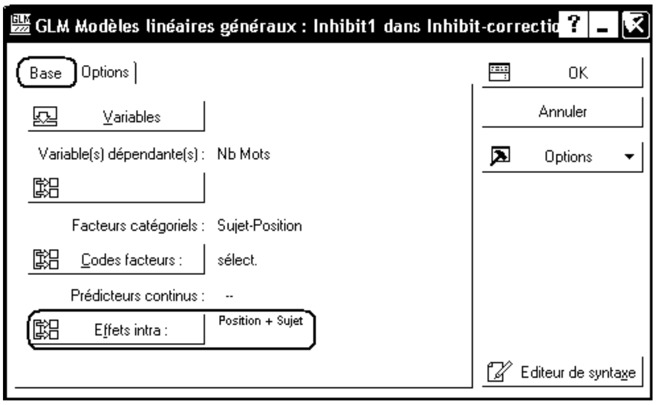
8.3.1 Données saisies "par observation"

Comme précédemment, utilisons le menu <u>Statistiques - Modèles linéaires / non linéaires avancés</u>, puis "<u>Modèle linéaire général</u>" et "<u>Modèles linéaires généraux</u>".

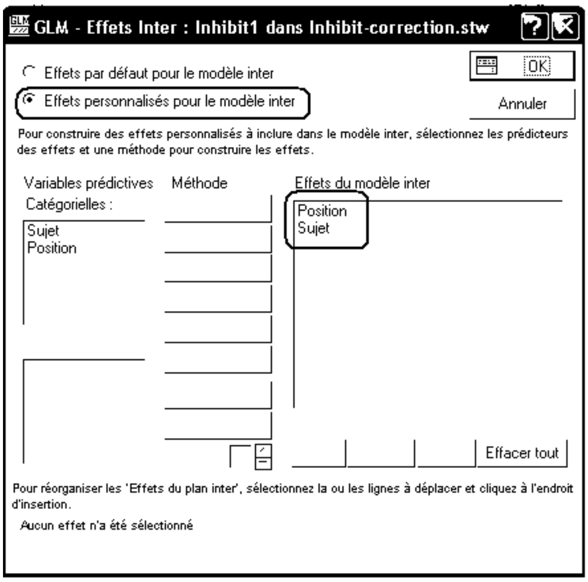
Avec les données structurées "par observation" (feuille Inhibit1 à rendre active), complétez le dialogue "Variables" de la manière suivante:



Cliquez ensuite sur le bouton "Effets Intra":



Cliquez alors sur le bouton radio "Effets personnalisés pour le modèle inter, et utilisez les outils fournis dans la fenêtre de dialogue pour obtenir les spécifications suivantes :

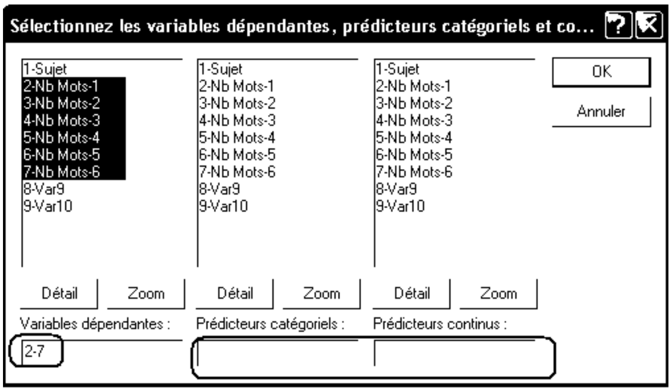


Le résultat s'affiche sous la même forme que dans la méthode 1.

8.3.2 Données saisies "par sujet" :

Activez la feuille Inhibit2. Utilisez le menu <u>Statistiques - Modèles linéaires / non linéaires avancés</u>, puis "<u>Modèle linéaire général</u>" et "<u>Modèles linéaires généraux</u>".

Dans le dialogue suivant, indiquez les variables 2 à 7 comme variables dépendantes et n'indiquez aucun prédicteur catégoriel, ni prédicteur continu.



Cliquez ensuite sur le bouton "Effets Intra" et définissez les 6 variables dépendantes comme niveaux du facteur POSITION :

Le résultat s'affiche sous une forme analogue à celle de la méthode 2.

8.4. Exercice

Dans une étude sur l'effet du bruit sur la discrimination perceptive, on utilise six sujets. On mesure pour chaque sujet le nombre d'erreurs commises dans une tâche de discrimination perceptive. Les sujets sont soumis à trois conditions. Dans la première, les sujets accomplissent la tâche en l'absence de bruit; dans la seconde, le bruit est présenté de façon intermittente (i.e., bruits d'avions) ; dans la dernière, le bruit est présenté de façon continue (bruits de "marteau piqueur") On obtient les résultats suivants

Sujets	Absence de bruit	Bruit Intermittent	Bruit continu
1	117	119	127
2	130	126	131
3	122	118	129
4	123	117	134
5	126	120	137
6	116	120	128

Après avoir identifié la ou les variable(s) indépendante(s), dépendante(s), vous répondrez à la question classique : la variable indépendante influe-t-elle sur les variables dépendantes ?. Réalisez cette étude à l'aide de Statistica et recomposez le tableau d'analyse de variance convenable.

9. Traitement d'un plan S<A*B>. Plan factoriel à 2 facteurs

Le plan S<A*B>, ou plan factoriel, correspond au cas où l'on étudie l'effet de deux facteurs croisés, en utilisant des groupes indépendants de sujets dans chacune des conditions définies par le croisement des deux facteurs. Les sources de variation à prendre en compte sont les facteurs A et B et, éventuellement, l'interaction AB. Le modèle de score est ici :

Score = Moyenne Générale + Effet de A + Effet de B + Effet d'interaction + Résidu aléatoire.

La forme générale du tableau d'analyse de variance correspondant est la suivante :

Sources de variation	Somme des Carrés	ddl	Carrés Moyens	F	p
Facteur A	SC_A	$a-1$	CM_A	$F_A = \frac{CM_A}{CM_{S(AB)}}$	....
Facteur B	SC_B	$b-1$	CM_B	$F_B = \frac{CM_B}{CM_{S(AB)}}$	....
Interaction AB	SC_{AB}	$(a-1)(b-1)$	CM_{AB}	$F_{AB} = \frac{CM_{AB}}{CM_{S(AB)}}$	....
Résidu S(AB)	$SC_{S(AB)}$	$ab(n-1)$	$CM_{S(AB)}$		
Total	SC_T	$N-1$			

Trois tests statistiques, correspondant aux trois sources de variation A, B et AB peuvent être effectués. La statistique F_A suit une loi de Fisher Snedecor à (a-1) et ab(n-1) degrés de liberté, la statistique F_B, une loi de Fisher Snedecor à (b-1) et ab(n-1) degrés de liberté et la statistique F_{AB} une loi de Fisher Snedecor à (a-1)(b-1) et ab(n-1) degrés de liberté

9.1. Première méthode : le menu ANOVA - Factorielle

Enoncé du cas:
Cet exemple est basé sur des données fictives présentées par Lindeman (1974).
Vous testez la performance de rats d'origines différentes dans un labyrinthe. La tâche du rat est d'apprendre à se rendre directement à l'endroit où de la nourriture a été placée, sans erreurs. Trois lignées de rats sont utilisées. Pour chacune de ces lignées, vous utilisez 4 animaux élevés dans un environnement libre, et 4 animaux élevés dans un environnement restreint. La variable dépendante est le nombre d'erreurs faites par le rat dans son parcours vers la nourriture.

Extrait des données :

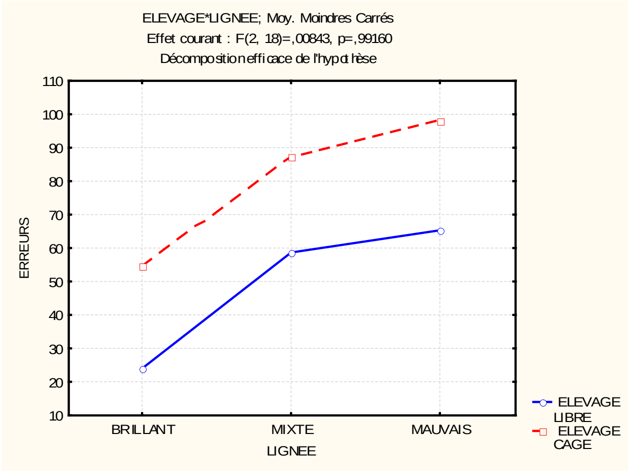
	ELEVAGE	LIGNEE	ERREURS
1	LIBRE	BRILLANT	26,000
2	LIBRE	BRILLANT	14,000
...			
17	CAGE	MIXTE	39,000
...			
24	CAGE	MAUVAIS	124,000

Ouvrez le fichier Rats.stw. Identifiez les facteurs de variation et la variable dépendante.
Il s'agit ici d'un plan S4<A2*B3>.

Utilisez le menu Statistiques - ANOVA.
Choisissez l'item ANOVA - Factorielle.
Spécifiez Erreurs comme variable dépendante, Elevage et Lignée comme prédicteurs.
Le résultat fourni par Statistica est le suivant :

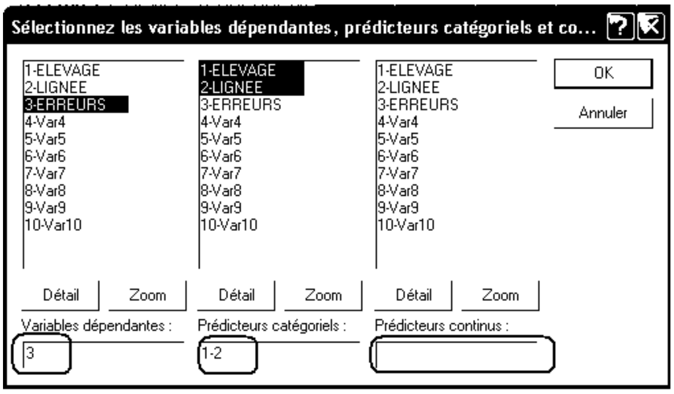
	Tests Univariés de Significativité pour ERREURS (Rats.sta) Paramétrisation sigma-restreinte Décomposition efficace de l'hypothèse				
Effet	SC	Degr. de Liberté	MC	F	p
ord. origine	100233,4	1	100233,4	105,1353	0,000000
ELEVAGE	5551,0	1	5551,0	5,8225	0,026705
LIGNEE	7939,8	2	3969,9	4,1640	0,032635
ELEVAGE*LIGNEE	16,1	2	8,0	0,0084	0,991604
Erreur	17160,8	18	953,4		

L'interaction (ici l'absence d'interaction) entre les deux facteurs étudiés peut être illustrée à l'aide d'un graphique :
Affichez la fenêtre de dialogue "Résultats ANOVA" et cliquez sur le bouton "Tous effets:Graphs".
Sélectionnez ensuite la ligne : "Elevage*Lignée"
Vous devriez obtenir le graphique suivant :

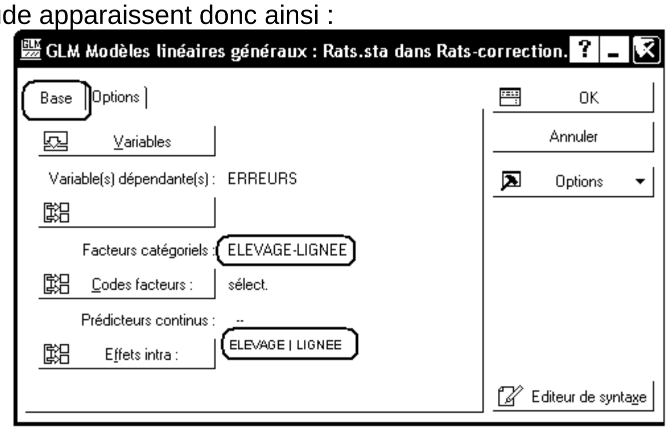


9.2. Deuxième méthode : le module "Modèle linéaire général"

Comme précédemment, utilisons le menu Statistiques - Modèles linéaires / non linéaires avancés, puis "Modèle linéaire général" et "Modèles linéaires généraux".
Compléter le dialogue relatif aux variables comme suit :



Les spécifications de l'étude apparaissent donc ainsi :



Le résultat obtenu est identique au précédent.

9.3. Exercice

On demande aux sujets de mémoriser des listes comportant 12, 24 ou 48 mots (facteur A, avec trois modalités). Ces mots peuvent se regrouper par paires en catégories (par exemple pomme et orange se regroupent en "fruits"). On demande aux sujets d'apprendre les mots, et on leur montre le nom des catégories à ce moment en leur précisant qu'ils n'ont pas à apprendre le nom de ces catégories. Au moment de l'épreuve de rappel qui a lieu immédiatement après l'apprentissage on crée deux conditions. Dans un cas, on présente aux sujets la liste des catégories. Dans l'autre cas, on ne leur présente pas cette liste

(facteur B présentation de la liste des catégories au moment de l'apprentissage versus absence de présentation). Dans cette reprise d'une expérience de Tulving et Pearlstone (1966), la variable dépendante sera le nombre de mots rappelés. En examinant les deux variables indépendantes, la première (nombre de mots de la liste) est, clairement, triviale. Il semble superfétatoire de construire une expérimentation pour montrer que plus une liste de mots est longue, plus on peut en retenir. Cette remarque indique que les auteurs de cette expérience s'intéressaient d'emblée à un effet d'interaction.
On interroge dix sujets par condition expérimentale. Voici les résultats

	Facteur A Nombre de mots par liste		
Facteur B	a1:12	a2:24	a3:48
	10 6	13 15	17 16
	8 11	18 13	20 23
b1	12 10	19 9	22 19
	8 9	13 8	13 20
	7 9	8 14	21 19
	12 10	12 13	31 29
	12 12	20 12	30 32
b2	7 10	19 13	26 24
	9 7	14 15	29 24
	9 12	16 6	28 27

Etudiez à l'aide d'une analyse de variance quels sont les facteurs dont l'effet est significatif. Recomposez le tableau d'analyse de variance et illustrez la situation proposée à l'aide d'un graphe d'interaction et commentez-le.

10. Traitement d'un plan S<A>*B. Plan à mesures partiellement répétées

Le plan S<A>*B correspond au cas où des groupes indépendants de sujets (facteur "groupe" A, emboîté dans les sujets) ont été observés dans deux ou plusieurs conditions, définies par les niveaux du facteur B, chaque sujet passant par tous les niveaux du facteur B. Un tel plan est qualifié de "plan à mesures partiellement répétées".
Les sources de variation à prendre en compte sont les facteurs A et B et, éventuellement, l'interaction AB. Le modèle de score est ici :

Score = Moyenne Générale + Effet de A + Effet "sujet" + Effet de B + Interaction AB + Résidu aléatoire.

La forme générale du tableau d'analyse de variance correspondant est la suivante :

Sources de variation	Somme des Carrés	ddl	Carrés Moyens	F	p
Entre les sujets					
Facteur A	SC_A	$a-1$	CM_A	$F_A = \dfrac{CM_A}{CM_{S(A)}}$	....
Facteur S(A)	$SC_{S(A)}$	$a(n-1)$	$CM_{S(A)}$		
Dans les sujets					
Facteur B	SC_B	$b-1$	CM_B	$F_B = \dfrac{CM_B}{CM_{BS(A)}}$	....
Interaction AB	SC_{AB}	$(a-1)(b-1)$	CM_{AB}	$F_{AB} = \dfrac{CM_{AB}}{CM_{BS(A)}}$	....
Résidu BS(A)	$SC_{BS(A)}$	$a(b-1)(n-1)$	$CM_{BS(A)}$		
Total	SC_T	$N-1$			

Pour un plan S<A>*B, avec A et B facteurs fixes, le rapport F relatif au facteur A se calcule en utilisant comme dénominateur le carré moyen relatif à S(A), tandis que les rapports F relatifs à B et AB utilisent le carré moyen du résidu.
Trois tests statistiques, correspondant aux trois sources de variation A, B et AB peuvent être effectués. La statistique F_A suit une loi de Fisher Snedecor à (a-1) et a(n-1) degrés de liberté, la statistique F_B, une loi de Fisher Snedecor à (b-1) et a(b-1)(n-1) degrés de liberté et la statistique F_{AB} une loi de Fisher Snedecor à (a-1)(b-1) et a(b-1)(n-1) degrés de liberté.

10.1. Première méthode : données saisies "par sujet"

Enoncé du cas:
En 1986, King a étudié l'activité motrice chez le rat après injection d'un médicament appelé midazolam. La première injection du médicament entraîne généralement une diminution nette de l'activité motrice. Mais une certaine tolérance se développe rapidement. King souhaitait savoir si cette tolérance acquise pouvait s'expliquer sur la base d'une tolérance conditionnée.
Il a utilisé trois groupes et n'a recueilli les données (présentées dans le tableau ci-dessous) que le dernier jour, jour du test. Durant le pré-test, deux groupes d'animaux ont reçu à plusieurs reprises des injections de midazolam réparties sur plusieurs jours, tandis que le groupe témoin recevait des injections d'une solution saline physiologique.

Le jour du test, un groupe (le groupe "même") a reçu une injection de midazolam dans le même environnement qu'auparavant. Le groupe "différent" a également reçu une injection de midazolam, mais dans un environnement différent. Enfin, le groupe témoin a reçu, pour la première fois, une injection de midazolam. Ce groupe témoin devrait donc manifester la réaction initiale classique au médicament (comportement ambulatoire réduit), tandis que le groupe "même" devrait présenter l'effet normal de tolérance. Par contre, si King a raison, le groupe "différent" devrait réagir de la même façon que le groupe témoin; en effet, ces animaux allaient cette fois recevoir l'injection dans un environnement différent, et les éléments nécessaires pour susciter une tolérance conditionnée ne seraient pas présents. La variable dépendante du tableau ci-dessous est une mesure du comportement ambulatoire, en unités arbitraires.

Comme le médicament se métabolise sur une période d'environ 1 heure, King a enregistré ses données par blocs (ou intervalles) de 5 minutes. Le tableau donne les valeurs observées pour les 6 premiers blocs de données.

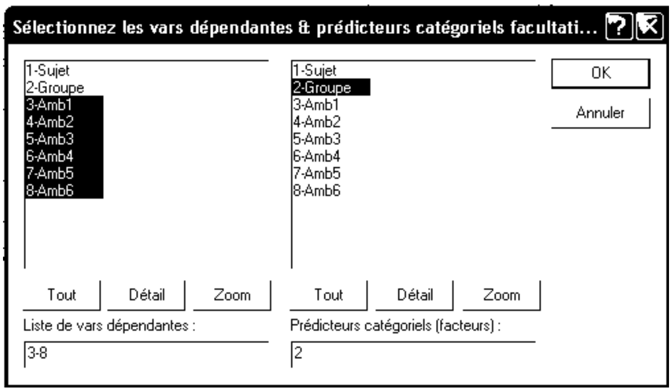
Ouvrez le fichier King.stw et rendez active la feuille King1.

Le plan d'expériences utilisé par King est du type S8<A3>*B6. Avec les notations utilisées dans la feuille de données, il s'agit du plan Sujet_8<Groupe_3>*Intervalle_6.

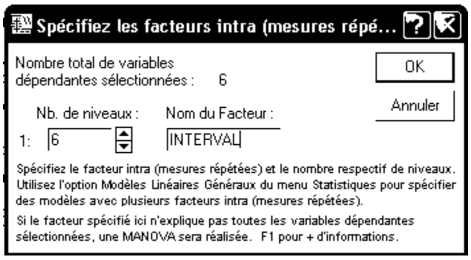
Dans la feuille King1, les données ont été saisies "par sujet" :

	Sujet	Groupe	Amb1	Amb2	Amb3	Amb4	Amb5	Amb6
1	S1	Témoin	150	44	71	59	132	74
2	S2	Témoin	335	270	156	160	118	230
...								
9	S9	Même	346	175	177	192	239	140
10	S10	Même	426	329	236	76	102	232
...								

Utilisez le menu Statistiques - ANOVA.
Choisissez l'item ANOVA - Mesures répétées.
Indiquez Amb1 à Amb6 comme variables dépendantes et Groupe comme prédicteur catégoriel.



Cliquez ensuite sur le bouton "Effets intra" et indiquez que les variables Amb1 à Amb6 correspondent aux 6 niveaux du facteur Intervalle :



N.B. Le facteur sujet est spécifié, de façon implicite, par la disposition des données : les mesures d'une même ligne sont relatives à un même sujet.
On obtient le résultat suivant :

	Analyse de la Variance Mesures Répétées (Feuille de données160 dans King.stw) Paramétrisation sigma-restreinte Décomposition efficace de l'hypothèse				
Effet	SC	Degr. de Liberté	MC	F	p
ord. origine	4113798	1	4113798	224,5511	0,000000
Groupe	285815	2	142908	7,8006	0,002928
Erreur	384722	21	18320		
INTERVAL	399737	5	79947	29,8524	0,000000
INTERVAL*Groupe	80820	10	8082	3,0178	0,002164
Erreur	281199	105	2678		

qui correspond au tableau d'analyse de variance ci-dessous:

```
Analyse de la variance pour Ambulato, en utilisant la SC ajustée pour les tests

Source             DL      SC           CM        F      P
Entre les sujets
Groupe              2     285815       142908    7.80   0.003
Sujet(Groupe)      21     384722        18320
Dans les sujets
Interval            5     399737        79947   29.85   0.000
Groupe*Interval    10      80820         8082    3.02   0.002
Erreur            105     281199         2678
Total             143    1432293
```

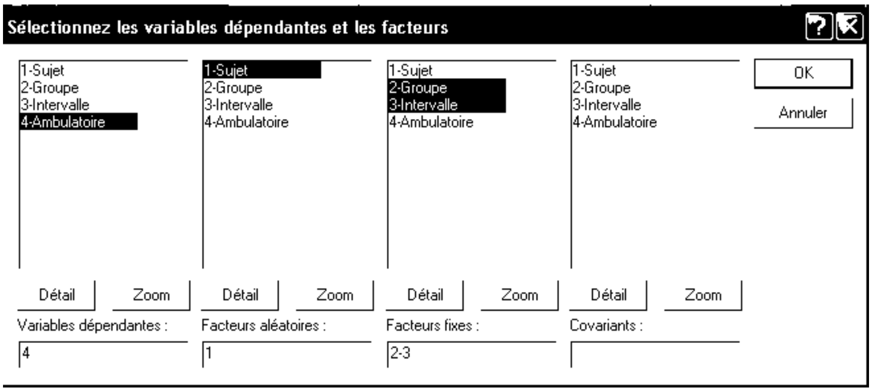
10.2. Deuxième méthode : données "par observation"

La feuille King2 contient les mêmes données, mais saisies selon la logique "plan d'expérience" : chaque facteur est représenté par une variable, et chaque ligne correspond à une observation.

Voici un extrait des données :

	1 Sujet	2 Groupe	3 Intervalle	4 Ambulatoire
1	S1	Témoin	1	150
2	S2	Témoin	1	335
3	S3	Témoin	1	149
4	S4	Témoin	1	159

Utilisez le menu : Statistiques - Modèles Linéaires/Non linéaires avancés - Décomposition de la Variance. Indiquez "Ambulatoire" comme variable dépendante, Sujet comme facteur aléatoire, Groupe et Intervalle comme facteurs fixes:



Cliquez sur le bouton OK, puis sur le bouton "Synthèse : Décomposition de la variance.
Vous obtenez le résultat suivant :

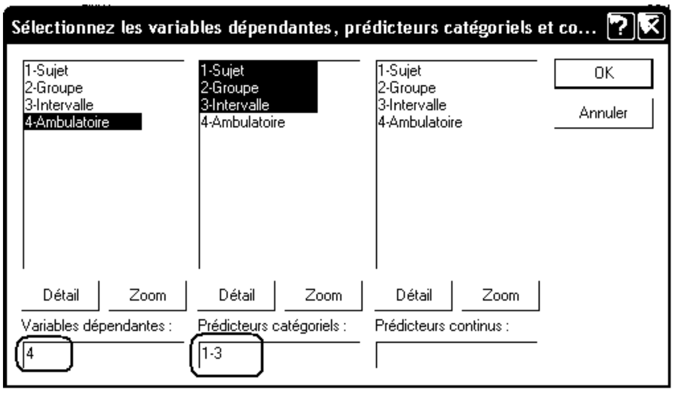
MC, Type I	Résultats ANOVA des Erreurs Synthétisées : Ambulatoire (King2.sta) dl de l'Erreur calculés par la méthode Satterthwaite * Ces tests considèrent les effets fixes impliqués nuls Effet (F/Al.)	dl Effet	MC Effet	dl Erreur	MC Erreur	F	p
{1}Groupe	*Fixe	2	142907,5	21	18320,10	7,80	0,002928
{2}Intervalle	*Fixe	5	79947,3	105	2678,09	29,85	0,000000
{3}Sujet	Aléat.	21	18320,1	105	2678,09	6,84	0,000000
1*2	Fixe	10	8082,0	105	2678,09	3,02	0,002164
1*3	Aléat.	0	0,0				
2*3	Aléat.	105	2678,1				
1*2*3	Aléat.	0	0,0				

En remettant les lignes dans le bon ordre, on retrouve ainsi le tableau d'analyse de variance.

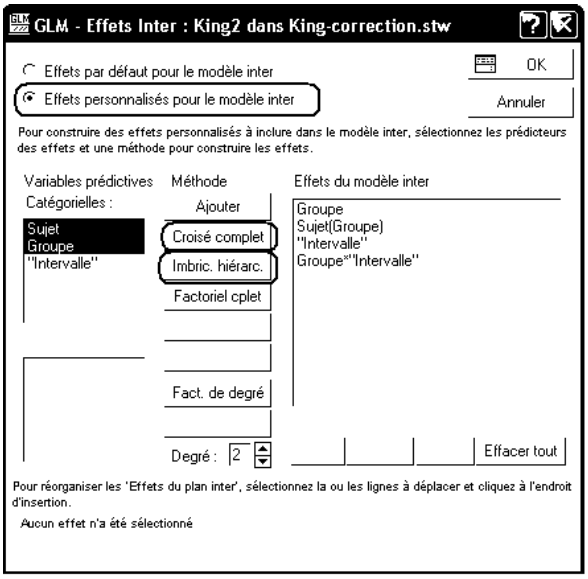
10.3. Troisième méthode : le module "Modèle linéaire général"

10.3.1 Données saisies "par observation"

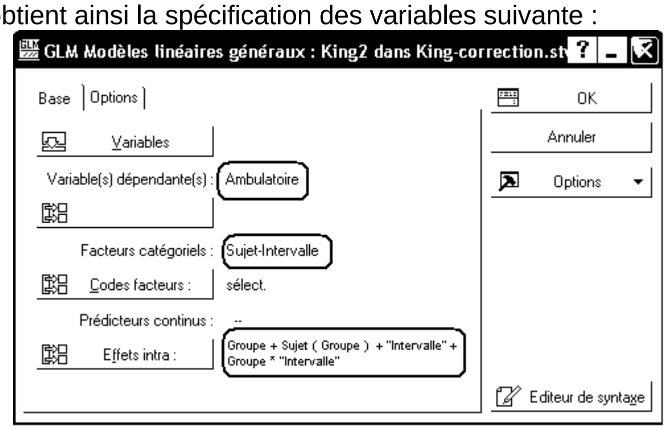
On utilise la feuille de données King1. et le menu Statistiques - Modèles linéaires / non linéaires avancés, puis "Modèle linéaire général" et "Modèles linéaires généraux".
Complétez le dialogue relatif aux variables comme suit :



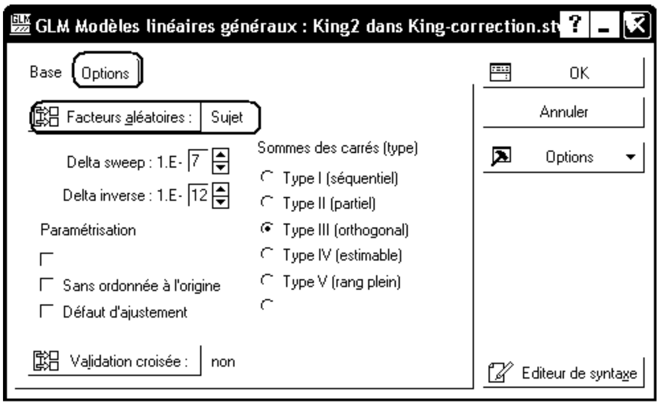
Cliquez ensuite sur le bouton "Effets Intra", puis sur le bouton radio "Effets personnalisés pour le modèle inter". Utilisez le bouton "Imbric. hiérarc." pour ajouter les effets Groupe et Sujet(Groupe). Ajoutez ensuite l'effet "Intervalle" et enfin, utilisez le bouton "Croisé complet pour ajouter l'effet d'interaction Groupe * Intervalle.



Validez ce dialogue. On obtient ainsi la spécification des variables suivante :



Important : Activez ensuite l'onglet Options et spécifiez Sujet comme facteur aléatoire :

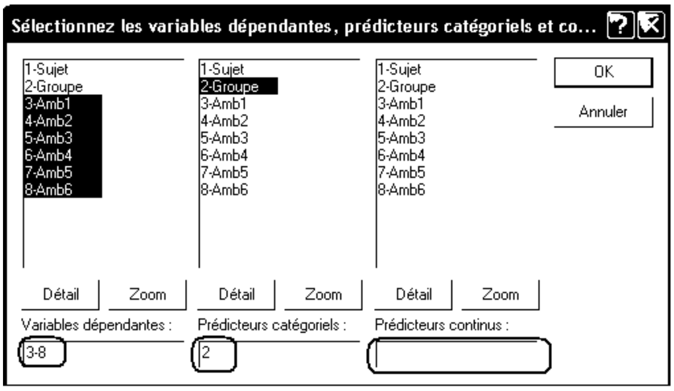


Statistica émet un message d'avertissement puis produit le résultat suivant :

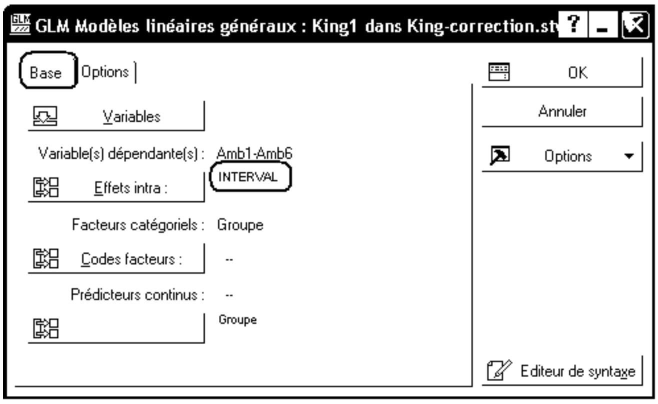
Effet	Effet (F/A)	SC	Degré de Liberté	MC	Syn.Dén. dl Err.	Syn.Dén. MC Erreu	F	p
	Tests Univariés de Significativité de Ambulatoire (King2 dans King-correction.stw) Modèle sur-paramétré Décomposition de Type III							
Ord.Orig.	Fixe	4113798	1	4113798	21	18320	224,55	0,000000
Groupe	Fixe	285815	2	142908	21	18320	7,80	0,002928
Sujet(Groupe)	Aléat.	384722	21	18320	105	2678	6,84	0,000000
Intervalle	Fixe	399737	5	79947	105	2678	29,85	0,000000
Groupe*Intervalle	Fixe	80820	10	8082	105	2678	3,02	0,002164
Erreur		281199	105	2678				

10.3.2 Données saisies "par sujet"

On utilise évidemment la feuille de données King1. et le menu Statistiques - Modèles linéaires / non linéaires avancés, puis "Modèle linéaire général" et "Modèles linéaires généraux".
Complétez le dialogue relatif aux variables comme suit :



Cliquez ensuite sur le bouton "Effets Intra" pour indiquer que les variables dépendantes Amb1 à Amb6 correspondent aux 6 niveaux du facteur "Intervalle :



10.4. Exercice

Dans une reprise d'une expérience de Conrad (1971), on veut mettre en évidence l'hypothèse de recherche suivante "les enfants jeunes n'utilisent pas un codage phonologique en mémoire à court terme". Pour ce faire, on sélectionne cinq enfants de 5 ans et 5 enfants de 12 ans (Variable A, avec deux modalités). On montre à chaque enfant un certain nombre de paires d'images représentant des objets dont on s'est assuré auparavant qu'ils sont nommés d'une seule manière par les enfants. On montre les images aux enfants. Puis on retourne les images (les enfants ne voient plus que le dos des images). Ensuite, on donne aux enfants une paire d'images identiques à celles retournées. Enfin, on leur demande de placer ces nouvelles images comme les images retournées sur la table. Pour la moitié des paires d'images les noms des objets se ressemblent (e.g., noix et doigt). Pour l'autre moitié, les noms des objets ne se ressemblent pas (e.g., maison et cheval). Conrad prédit que les enfants les plus vieux réussiront dans l'ensemble mieux que les enfants les plus jeunes, mais également que les enfants les plus vieux utiliseront un codage phonologique comme mnémonique (i.e., "la parole intérieure"). De ce fait, les enfants les plus vieux devront commettre plus d'erreurs lorsque les noms se ressemblent acoustiquement que lorsque les noms diffèrent. On présente à chaque enfant cinquante paires d'images correspondant à la modalité b1 (dissemblance acoustique), et cinquante paires d'images correspondant à la modalité b2 (ressemblance acoustique); la variable dépendante choisie est le nombre de paires d'images correctement reconstituées. L'ordre de présentation est "aléatorisé" pour chaque passation (Pourquoi cette précaution ?).

Essayer de traduire l'hypothèse de recherche en prédiction sur les sources de variation de l'analyse de variance.
Vous avez dû conclure que, d'une part, on s'attend à un effet principal de l'âge (qui est trivial), et, d'autre part, à un effet d'interaction c'est le point d'importance, ou si vous préférez, le point crucial de la théorie. On retrouve, ici, le rôle essentiel de l'interaction "comme test de théorie".

	a1b1	a1b2
s1	15	14
s2	23	20
s3	12	11
s4	16	17
s5	14	13
	a2b1	a2b2
s6	40	33
s7	38	23
s8	31	21
s9	36	26
s10	30	22

Etudiez ces données à l'aide d'une analyse de variance, recomposez le tableau d'analyse de variance sous sa forme classique et commentez les résultats obtenus. Illustrez l'étude à l'aide d'un graphe d'interaction.

11. Contrastes orthogonaux et tests post hoc

L'analyse de variance montre que, globalement, les différentes moyennes sont significativement différentes. Mais elle n'indique pas quelles sont les différences significatives qui existent dans cet ensemble de moyennes. les résultats relatifs aux moyennes prises deux à deux. Différentes méthodes ont été proposées pour répondre à cette interrogation.

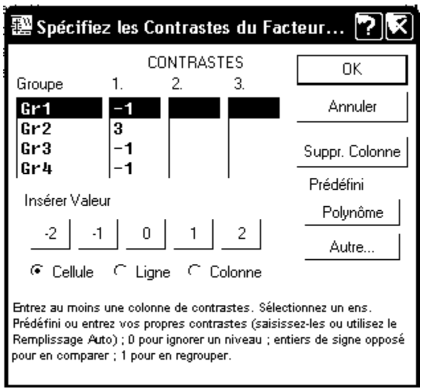
11.1. Contrastes orthogonaux

On reprend les données Bransford (fichier Bransford.stw). L'étude peut être poursuivie à l'aide de la méthode des contrastes orthogonaux.

La première étape consiste opposer le groupe 2 aux trois autres groupes en testant l'hypothèse nulle :

$$3\mu_2 = \mu_1 + \mu_3 + \mu_4$$

Refaites le calcul de l'ANOVA sur les données Bransford. Puis, reprenez l'analyse et affichez l'onglet "Comps". Cliquez sur le bouton "Contrastes de moyennes MC" et entrez les coefficients suivants :



Cliquez ensuite sur le bouton OK, puis le bouton Calculer.
On obtient les deux tableaux de résultats suivants (N.B. le premier tableau n'est affiché que dans un rapport. Demandez donc à ce que les résultats du traitement soient copiés dans un rapport) :

	Estimations de Contrastes (Bransfor.sta) Variable dépendante : VD					
Contraste	Estimati	Err-Ty.	t	p	-95,00% Lmt Cnf.	+95,00% Lmt Cnf.
CNTRST1	10,60000	2,374868	4,463405	0,000392	5,565504	15,63450

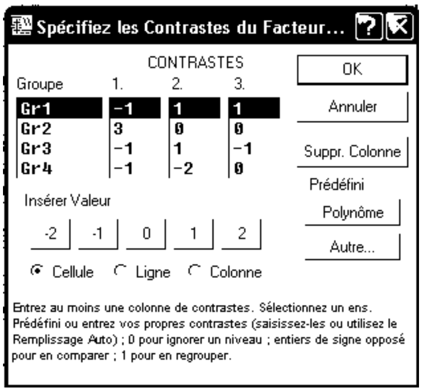
	Test Univarié de Significativité pour les Comp. Plannifiées (Bransfor.sta) Variable dépendante : VD				
Source	Somme Carrés	Degr. de Liberté	Moyen. Carré	F	p
Effet	46,81667	1	46,81667	19,92199	0,000392
Erreur	37,60000	16	2,35000		

Détails des calculs : voir la fiche de TD de statistiques.

Le F de Fisher associé à ce contraste vaut 19.92. Les degrés de liberté sont 1 et 16. Le résultat est donc significatif d'un comportement du groupe 2 différent de celui des autres groupes.

La méthode peut être poursuivie en opposant le groupe 4 aux groupes 1 et 3 (coefficients appliqués aux quatre moyennes : 1, 0, 1, -2) puis en opposant les groupes 1 et 3 (coefficients appliqués : 1, 0, -1, 0).

Statistica permet de réaliser les trois calculs en une seule étape. Reprenez la fenêtre de dialogue"Contrastes de moyennes MC" et complétez-la comme suit :



On obtient alors les résultats suivants :

Contraste	Estimations de Contrastes (Bransfor.sta) Variable dépendante : VD Estimati	Err-Ty.	t	p	-95,00% Lmt Cnf.	+95,00% Lmt Cnf.
CNTRST1	10,60000	2,374868	4,46341	0,000392	5,56550	15,63450
CNTRST2	-2,20000	1,679286	-1,31008	0,208666	-5,75993	1,35993
CNTRST3	-0,20000	0,969536	-0,20628	0,839171	-2,25532	1,85532

Source	Test Univarié de Significativité pour les Comp. Plannifiées (Bransfor.sta) Variable dépendante : VD Somme Carrés	Degr. de Liberté	Moyen. Carré	F	p
Effet	50,95000	3	16,98333	7,226950	0,002782
Erreur	37,60000	16	2,35000		

Statistica calcule un t de Student pour chacun des contrastes, et un F de Fisher pour l'ensemble des trois contrastes. On voit que seul le premier contraste conduit à un résultat significatif.

En fait, Statistica permet de ne rentrer que 1, 0 ou -1 comme coefficients. Il se charge de calculer lui-même les pondérations nécessaires. En revanche, il permet de rentrer des jeux de coefficients qui ne correspondent pas à des contrastes orthogonaux.

Pourquoi s'agit-il de contrastes *orthogonaux* ?

Réponse : Les "vecteurs" associés aux coefficients des trois contrastes, à savoir V1=(-1, 3, -1, -1), V2=(1, 0, 1, -2), V3=(1, 0, -1, 0) sont deux à deux orthogonaux ce qui garantit l'indépendance des résultats des trois tests.

11.2. Test de Tukey pour une ANOVA à 1 facteur

11.2.1 Principe

On dispose de données recueillies selon un plan S<G>. On a réalisé une ANOVA qui conclut à une différence entre les groupes. On souhaite répondre à la question suivante : "quelles sont les paires de groupes pour lesquelles les différences sont significatives ?"

Notations :
* r : nombre de groupes
* n : effectif de chaque groupe (N.B. la méthode fonctionne aussi pour des groupes non équilibrés).

Tableau d'ANOVA :

	SC	ddl	CM	F
Entre groupes	SC_G	r-1	CM_G	F
Résidu	$SC_{S(G)}$	r(n-1)	$CM_{S(G)}$	
Total	SC_T	rn-1		

Pour chaque groupe, l'erreur standard estimée est :

$$E = \sqrt{\frac{CM_{S(G)}}{n}}$$

On ordonne les moyennes par ordre décroissant. Soient $\overline{x}_{Max}$ et $\overline{x}_{Min}$ la plus grande et la plus petite valeur de cet ensemble de moyennes. Lorsque *r > 2*, la quantité :

$$Q = \frac{\overline{x}_{Max} - \overline{x}_{Min}}{E}$$

(ou plutôt la quantité $Q/\sqrt{2}$) ne suit pas une loi de Student, car les deux moyennes prises en compte ne sont pas choisies au hasard. On montre que cette quantité suit la loi des écarts studentisés qui prend comme paramètres le nombre de groupes r et le nombre de ddl (celui figurant dans la ligne "résidu" du tableau d'ANOVA).

Dans le test de Newman-Keuls, on calcule la quantité Q pour deux moyennes quelconques, et on considère la loi des écarts studentisés avec comme paramètre r, le nombre "d'échelons" : deux moyennes adjacentes sont distantes de deux échelons, deux moyennes séparées par une 3è sont distantes de 3 échelons, etc. En fait, dès que le nombre de paires augmente, le risque de commettre une erreur de type I dépasse ainsi largement le seuil α fixé par l'expérimentateur.

Le test de Tukey, ou test de la différence franchement significative (HSD : honestly significative difference) consiste à calculer l'expression Q pour toutes les différences de moyennes, en conservant comme loi de distribution, celle de paramètres r et ddl. Autrement dit, on calcule les quantités :

$$Q = \frac{\overline{x}_i - \overline{x}_j}{E} \quad où \quad \overline{x}_i > \overline{x}_j$$

et on conserve pour Q la distribution des étendues studentisées de paramètres r et ddl.
Si $Q_{Obs} > Q_{Crit}$, on conclut à une différence significative entre les deux moyennes constituant la paire.

7.1.2 Le test de Tukey avec Statistica

On reprend le classeur Bransford.stw :

On utilise le menu "Modèle linéaire général", puis les traitements : "Synthèse - Tous les effets" et "Autres résultats - Post Hoc - HSD de Tukey". On obtient alors les résultats suivants :

Test HSD de Tukey ; variable VD (Données Bransford dans Bransford-correction.stw)
Probabilités Approximatives des Tests Post Hoc
Erreur : MC Inter = 2,3500, dl = 16,000

	Groupe	{1}	{2}	{3}	{4}
		3,0000	7,0000	3,2000	4,2000
1	GR1		0,0041	0,9968	0,6132
2	GR2	0,0041		0,0061	0,0476
3	GR3	0,9968	0,0061		0,7341
4	GR4	0,6132	0,0476	0,7341	

Autrement dit, Statistica nous indique les niveaux de significativité des différences entre les moyennes des quatre groupes, en utilisant la distribution des écarts studentisés. Ici, les seules différences significatives sont celles du groupe 2 avec les autres groupes.

7.3. Estimation de l'intensité de l'effet

Une autre grandeur intéressante est le coefficient (souvent noté η^2, Statistica le note R^2) d'estimation de l'intensité de l'effet de la variable indépendante. Dans le cas d'une analyse de variance à un facteur, il est défini par : R^2 = SCinter / SCtotal.

Affichez l'onglet "Synthèse" et cliquez sur le bouton "R modèle complet". On obtient :

Dépendnt Variable	Test de la SC du modèle entier vs. SC Résiduels (Bransfor.sta)							
	Multiple R	Multiple R²	Ajusté R²	SC Modèle	dl Modèle	MC Modèle	SC Résidu	dl Résidu
VD	0,758539	0,575381	0,495765	50,95000	3	16,98333	37,60000	16

Signification : 58% de la variance de la variable dépendante est expliquée par l'effet de la variable indépendante (les différentes conditions expérimentales).

R^2 est aussi le carré d'un coefficient de corrélation. Il peut en effet être obtenu comme coefficient de la corrélation entre l'ensemble des données observées d'une part, et la série de données obtenue en remplaçant chaque observation par la moyenne de son groupe d'autre part.

12. Travail à rendre par courrier électronique

Réalisez les études demandées dans les exercices des paragraphes 2, 3 et 4. Faites parvenir votre travail (rapport Statistica ou fichier Word dans chacun des cas) par mail à votre enseignant.