Réflexions avant d'entreprendre un projet expérimental

Bernard Clément, PhD

Introduction

La genèse de ce document trouve sa source dans la correction d'un travail proposé dans le cadre du cours MTH8301 Planification et analyse d'expériences qui est enseigné à Polytechnique Montréal. Dans le cadre du cours, les étudiants doivent développer un projet de recherche de planification d'expérience sur un sujet précis qui leur est proposé. Le présent document propose une série de réflexions et d'éclaircissements indispensables à toute personne ou organisation qui envisage d'entreprendre un projet expérimental. Ils sont de nature universelle en articulant la terminologie et les définitions indispensables avec un objectif de clarifier et de guider toute la démarche expérimentale. Indépendamment du domaine d'application, l'information examine des concepts fondamentaux de la planification et de l'analyse statistique d'expériences.

Périmètre et envergure d'une étude expérimentale

Le projet de planification d'expérience transmis est souvent fut interprété dans un cadre beaucoup plus général que nécessaire. Souvent on analyse le projet en visant des objectifs d'envergure qui dépasse l'objectif poursuivi. C'est louable mais cela va au-delà de la problématique soumise et cela a des conséquences directes sur la durée et les coûts impliqués.

En général, il y a un facteur principal et le seul objectif de l'expérience est d'évaluer son impact sur une variable de réponse. C'est le cas par exemple dans la mise au point d'un nouveau traitement médical, ou le développement d'un nouveau vaccin. Associé au facteur principal, il y a des facteurs associés comme la dose, la durée, … etc. qui viennent préciser avec plus de détails. Il faut répondre à la question : quel est le facteur principal du projet d'expérimentation? Toutefois, il existe une autre catégorie d'expérimentation qui est de tamiser un ensemble de facteurs et de déterminer les plus importants influençant la réponse. Il s'agit d'expérience de tamisage ((« screening experiments »). Ces expériences sont souvent la première phase d'un projet d'expérimentation plus global.

On convient facilement qu'il y de nombreux facteurs qui affecte une variable de réponse. La liste peut-être relativement longue et contient beaucoup de facteurs secondaires, aussi appelés facteurs nuisibles qui devaient être tenus en compte mais ils ne sont pas au cœur du projet. C'est le cas le plus souvent avec un projet d'expérience qui veut essentiellement répondre à une question. Que doit faire avec tous ces facteurs? Il y a essentiellement 3 possibilités : les maintenir constants, les mesurés et enfin appliquer une randomisation des essais. Cette dernière méthode est un des principes fondamentaux de la construction d'un plan d'essais et devrait être généralement appliquer à toute série d'essais, sauf en certaines circonstances.

Comment faire pour que les conclusions de l'expérience soient les plus générales possibles et ne concernent pas sur un horizon trop petit en laissant en suspens trop de questions? Comment établir un équilibre essentiel entre répondre à la question spécifique et généraliser l'étendue de la conclusion? Il n'y a pas une bonne et unique réponse à cette question. L'expérimentateur doit établir un équilibre délicat en faisant des choix judicieux. Ceci fait partie de l'art de l'expérimentation.

En définitive, la définition du périmètre du projet se doit d'être bien circonscrite. Cela relève directement du domaine d'application et des objectifs précis et poursuivis par l'expérience. Ceci a un impact direct sur l'échéancier et le budget. Dans certains cas on peut partir de ces contraintes pour aider à cerner l'ampleur d'un projet expérimental.

Revue de la littérature

Il est souhaitable de prendre connaissance du sujet de toute étude expérimentale en faisant une recherche de la littérature associée au projet en consultant l'internet. Consulter les articles les plus récents est généralement suffisant. Il n'est pas nécessaire d'être exhaustif comme ce serait le cas pour une thèse de doctorat par exemple.

En général, un projet d'expérimentation contient un minimum de références. C'est facile et rapide de nos jours de réaliser une recherche. Cela devient une étape incontournable. Consulter les premières pages d'une recherche google est suffisant. Afin de faire un tri rapide, la lecture du résumé constitue un moyen très efficace pour sélectionner les articles et les présentations sur le thème retenu.

Identification, définition et mesure des variables de réponse Y

Il est fortement conseiller de commencer par la définition des variables de réponse Y du système/processus à l'étude. Ces variables sont directement liées aux objectifs (minimum, maximum, cible) de l'étude expérimentale. La méthode (processus) de mesure de Y peut poser des défis à résoudre préalablement avant même de lancer le projet expérimental et faire même l'objet d'un projet expérimental en soi. Il faut consulter la littérature sur la définition et la mesure de Y et ne pas réinventer la roue. D'ailleurs, nous avons réalisé un document très complet (108 pages) sur le processus de mesurage. Il contient de nombreux exemples et on peut en prendre connaissance :
https://cours.polymtl.ca/mth6301/mth8301/Clement/Clement-Evaluation_Processus_Mesure.pdf

En général, il y a seulement quelques variables de réponse Y, disons 5 ou moins (heuristique). On doit identifier celles qui sont près des objectifs poursuivis. La recherche des facteurs X qui ont une influence sur Y sera mieux ciblée.

Identification des facteurs explicatifs X de(s) la variable de réponse Y

Constituer une liste et choisir les facteurs (variables) explicatifs X constituent un défi. Mais c'est surtout la suite qui présente des décisions critiques pour le projet d'expérimentation. Il faut viser à être exhaustif le plus possible pour ne pas oublier des facteurs potentiellement importants. Il est raisonnable de penser qu'une liste de 20 facteurs ou moins est généralement suffisante. C'est une règle heuristique que je propose et qui est supportée si on consulte de nombreux projets d'expérimentation de la littérature.

Ce sont les étapes et les décisions ultérieures qui présentent des défis. En face d'un facteur on doit prendre des décisions qui ont un impact direct sur la liste des traitements

et sur les designs d'expérimentation qui seront envisagés. Il faut faire la distinction entre deux structures : celle des traitements et celle du protocole (design) de la méthode d'attribution des traitements aux unités expérimentales.

Cette dernière remarque m'a inspiré pour proposer un guide basé sur la nature et le rôle que l'on veut faire jouer aux facteurs explicatifs X. Cette information a un impact direct sur les plans d'expérimentation qui seront envisagés. Je ne crois pas, à ma connaissance, avoir vu un tel guide dans la littérature sur les plans d'expérimentation. Mais avant, précisons des concepts essentiels pour la construction des plans d'expériences planifiées et de leurs analyses.

L'univers des facteurs

Il y a beaucoup à dire sur les facteurs, leur nature, leur rôle, leur traitement statistique etc. Malheureusement, les monographies en planification et analyse statistique d'expériences ne présentent pas suffisamment d'information sur ce sujet important. Nous allons ici tenter de corriger partiellement cet état de fait.

Les facteurs d'une expérience sont représentés par des variables X ayant une influence potentielle sur la (les) variable de réponse Y. La source et l'identification de ces facteurs provient d'une recherche la plus exhaustive possible.
Les grandes catégories sont toujours les mêmes :

MÉTHODES, MACHINES, MATÉRIAUX, MESURES, ENVIRONNEMENT, PERSONNES

Consulter la page 15 pour un exemple de diagramme d'Ishikawa.

La liste que l'on a établie est-elle complète? Il faut plutôt se poser la question : est-elle suffisamment riche pour tenir en compte les facteurs les plus importants. Quelles sont les conséquences si on oublie certains facteurs connus ou inconnus? La réponse à cette dernière question réside dans l'introduction d'un facteur fourretout appelée l'erreur expérimentale. On ne le contrôle pas mais on décèle son influence avec la répétition des essais. Aussi, on peut calculer son influence lors l'analyse statistique des résultats.

Généralement notés par la lettre X; les facteurs aussi appelés variables explicatives ou variables d'intrant (input) du système/processus à l'étude. Influencent-ils ou non la variable de réponse Y? On doit obligatoirement répondre à la question car c'est l'enjeu premier de l'expérience. Le terme Les variables explicatives que l'on fera varier à des modalités spécifiques a dans un premier temps, un sens de variable catégorique, même si certains sont de nature quantitative. C'est seulement lorsque l'on veut établir une fonction mathématique entre les X et Y que l'on fait intervenir la nature quantitative continue de X si c'est le cas. On peut même introduire des variables catégoriques dans

la fonction mathématique en adoptant un codage approprié, soit le codage disjonctif complet ou le codage à effets. Toutefois, le premier objectif de toute étude expérimentale est de déterminer si X influence on non Y. On peut ajouter des étapes et des développements mathématiques additionnels si cela convient.

Dans l'étude sur le sommeil profond, il s'agit de la présence ou de l'absence du son spécial qui est identifié comme le facteur primaire. Il est le déclencheur d'un projet d'expérimentation afin élucider si le son proposé influence ou pas le sommeil profond. L'ajout de quelques autres facteurs liés aux unités expérimentales s'imposent naturellement. Dans l'exemple du sommeil profond, le genre (homme, femme) et l'âge (catégories à définir) sont des facteurs qui s'ajoutent naturellement. C'est toujours le cas dans les études expérimentales impliquant des humains. La raison principale est évidente : on doit contrôler la variabilité des unités expérimentales (UE) qui sont relativement hétérogènes. On veut aussi généraliser les conclusions aux humains. C'est typique dans les domaines d'application comme la médecine, les sciences de la vie, en psychologie etc. C'est une différence fondamentale si on compare les domaines des sciences physiques et en ingénierie où les UE sont généralement homogènes et constituées de matériaux non vivants.

Facteurs primaires

Ce sont les facteurs (variables) à l'origine du projet expérimental. Ils sont généralement peu nombreux et souvent il y en a un seul comme dans l'étude de cas qui fut proposée dans l'ANNEXE 1. Par exemple en industrie pharmaceutique le projet expérimental est de démontrer scientifiquement les propriétés bénéfiques d'une nouvelle molécule. Mais il y a des projets pour explorer et modéliser un système/processus, alors on peut en avoir dizaines de facteurs sans distinction entre facteurs primaires et facteurs secondaires. Dans ce cas, ce n'est pas un facteur spécifique qui est le déclencheur du projet expérimental, mais plutôt un objectif de modéliser le système avec les facteurs les plus importants.

Facteur : contrôlable ou facteur non contrôlable?

Cette question est capitale. On expérimente seulement avec des facteurs dont on peut fixer les valeurs des variables quantitatives et les modalités des variables catégoriques. Sinon on n'a pas un projet expérimental. Mais la question est plus complexe telle qu'elle apparait à première vue. Par exemple, des facteurs liés à l'environnement ou à des conditions d'utilisation d'un produit sont évidemment incontrôlables par leur nature même. Il est possible de les contrôler en laboratoire, probablement au prix d'un effort important. On peut ainsi neutraliser leur influence sur le système/processus avec les

facteurs contrôlables. C'est la philosophie de Taguchi pour qui la recherche de la robustesse (performance) des produits est une idée forte.

Dans le cas où les facteurs sont incontrôlables, la solution repose sur le principe de répétition et le principe de randomisation. Le principe de répétition des essais conduit à estimer les effets avec une plus grande précision et, par conséquence, permet de savoir si un facteur a un effet significatif ou non sur la réponse Y au-delà de l'erreur expérimentale. Le principe de randomisation, consiste à conduire les essais dans un ordre aléatoire ou randomisé ce qui du moins théoriquement, distribue uniformément les effets des facteurs incontrôlables sur l'ensemble des essais et évite la possibilité très réelle de biais aussi connue d'erreur systématique. Mais ce principe n'est pas sans faille. Par exemple, si l'instrument de mesure est recalibré en cours des essais, la randomisation ne règle pas la possibilité d'un écart significatif entre certains essais. D'autre part, cette solution ne tient pas en compte, la situation que les changements de modalités de certains facteurs soient plus difficiles pour certains facteurs que pour d'autres facteurs. Ainsi, si on peut conduire les essais dans un ordre plus commode, appelé mode parcelles divisées (« SplitPlot ») qui ne respecte pas complètement le principe de la randomisation.

Facteurs secondaires (facteur nuisibles)

Dans toutes les expériences il y a des facteurs qui ne sont pas au cœur du projet mais qui sont présents dans la conduite des essais. Ce sont les facteurs secondaires, aussi appelés facteurs nuisibles. Ils sont toujours présents dans toutes les expériences. Ils ont une influence potentielle sur Y rendant ainsi plus difficile la détection de l'influence des facteurs primaires. Que faire? La réponse réside en deux mots : blocage et covariable.

Le blocage consiste à identifier des facteurs connus et contrôlables dont peut fixer leurs modalités ou valeurs dans un premier temps. C'est comme un facteur d'expérimentation primaire mais avec une particularité essentielle. On le tient en compte dans le plan avec le principe de stratification. On construit un plan avec des strates à l'intérieur desquelles le facteur est maintenu constant. Au moment de l'analyse, l'influence de ce facteur est dégagée de la réponse Y et ainsi, l'analyse procède comme si le facteur était demeuré constant dans les essais. Ainsi, on peut alors dégager convenablement l'effet des facteurs primaires. L'absence d'interaction entre les facteurs blocs et les facteurs primaires est une hypothèse du modèle lors de l'analyse.

L'autre cas consiste à mesurer les facteurs X secondaires dont on ne peut pas fixer leurs valeurs. Ces facteurs sont des variables continues associées à Y. Comme dans le cas du blocage, on enlève ou corrige la réponse Y en enlevant leur effet (par régressions) sur Y. On obtient ainsi une variable de réponse corrigée comme si les covariables étaient demeurées constantes. On peut alors dégager clairement les effets des facteurs primaires

comme si les facteurs étaient demeurés constants. L'analyse est connue sous le nom de l'analyse de covariance. Une des hypothèses fondamentales pour l'analyse des résultats est que les facteurs secondaires n'ont pas d'effet d'interaction avec les facteurs primaires comme pour le cas des facteurs blocs.

Le blocage et l'analyse de covariance sont des variantes de la même idée. On veut neutraliser l'influence des facteurs secondaires afin de bien dégager l'influence des facteurs primaires, ceux qui sont directement l'objet principal du projet d'expérimentation. Cette stratégie est aussi connue sous le nom de contrôle local dont l'objectif est de minimiser l'influence des facteurs secondaires. Ce principe vient s'ajouter aux deux autres que sont la répétition et la randomisation des essais.

Traitement, unité expérimentale, unité d'observation

Une combinaison des modalités provenant de plusieurs facteurs s'appelle un traitement. Dans toute expérience on applique des traitements à des unités expérimentales (UE). Ils sont aussi souvent appelés sujets expérimentaux dans des expériences impliquant des personnes ou des animaux. L'entité UE est définie comme la plus petite quantité de matière, vivante ou non sur laquelle on applique un traitement. Il est important de bien identifier cette entité dans toute expérience.

On doit différencier unité expérimentale (UE) et unité d'observation (UO). L'EU peut contenir une ou plusieurs UO sur lesquelles les mesures sont prises. De par leur nature, toutes les unités d'observation reçoivent le même traitement. Dans la grande majorité des circonstances, les deux unités coïncident. Mais dans d'autres, il y a une différence entre EU et UO. Clarifions ces concepts avec deux exemples.

Dans un premier exemple, on applique un traitement sur une pièce, disons un cylindre individuel; par la suite, on mesure la force de rupture du cylindre. Dans ce cas, UE et UO coïncident. Dans un deuxième exemple, supposons que l'on s'intéresse à la mesure du goût d'un gâteau. Les facteurs-ingrédients du gâteau sont mélangés, versés dans un moule et cuits dans un four. Après la cuisson on découpe le gâteau en tranches sur lesquelles on mesure l'indicateur du goût. Dans ce cas, l'UE est le gâteau et les UO sont les tranches. L'UO est la plus petite unité d'observation, ici une tranche de chaque gâteau. Cette distinction est importante car c'est la variation entre les UE qui est la base pour le calcul de l'erreur expérimentale.

L'erreur expérimentale est l'ensemble des tous les facteurs inconnus qui ont un impact potentiel sur la variable de réponse Y. Ce facteur est fondamentalement de nature aléatoire. Il est impératif de faire son estimation si on veut dégager des conclusions sur l'influence ou non des facteurs X sur Y. Dans l'exemple du gâteau on pourrait utiliser la

moyenne des indicateurs de goût, une manifestation de l'erreur expérimentale. En général, dans toute analyse des résultats d'une expérience, l'estimation de l'erreur expérimentale s'obtient par le calcul de la somme de carrés résiduels de la réponse Y après avoir enlevé l'influence des toutes les variables explicatives sur la réponse Y. Il faut donc obligatoirement proposer un modèle statistique approprié reliant les X et Y pour réaliser l'analyse.

Facteur INTER – Facteur INTRA

On doit distinguer entre un facteur INTER et un facteur INTRA dans le design et l'analyse des plans expérimentaux. Un facteur INTER prend des valeurs différentes sur des UE distinctes. Un facteur est déclaré INTRA quand chaque UE se voit attribué toutes les modalités du facteur. Un cas particulier fréquent d'un facteur INTRA est lorsque l'on prend des mesures sur un même UE en fonction du temps; cette situation étant aussi connue sous le nom de mesures longitudinales.

Par exemple, un facteur INTRA pourrait être la position de mesure sur une UE et ainsi, on aurait des mesures à chaque position. En général, les plans en mesures répétées sont particulièrement utiles dans les études statistiques lorsque les UE sont des personnes ou des animaux car on peut contrôler l'effet de la variabilité des unités expérimentales. Mais il faut nécessairement que la mesure ne soit pas destructive de l'UE comme cela peut survenir lors d'expériences en ingénierie ou en science physiques. Dans les expériences avec des sujets vivants, cette situation se présente quand les sujets reçoivent 2 variantes d'un traitement. Il y a la possibilité que le sujet ait un effet résiduel du premier traitement alors qu'il subit un deuxième traitement.

Par ailleurs, les analyses avec des mesures répétées sont plus complexes à cause de la structure de dépendance de la réponse Y car elles sont associées à la même unité expérimentale. Cette particularité doit être obligatoirement tenue en compte lors de l'analyse des résultats. Des informations sont exposées dans L'ANNEXE 2 concernant les différentes structures de dépendance.

Le cas du concept de mesures emboitées et de parcelles divisées (SplitPlot) introduit un élément de dépendance de la réponse Y. Le concept de parcelles divisées est important en expérimentation car il permet de tenir en compte la distinction entre facteurs difficiles à changer (DAC) et les facteurs faciles à changer (FAC). Le plan d'expérimentation recommandé de construire est connu sous le nom de plan en parcelles divisées (Split-plot). Il est très important de reconnaitre cette structure d'expérimentation qui est fréquemment employée dans les expériences scientifiques et industrielles. Dans certains cas, l'utilisateur peut ne pas proposer le bon modèle et alors, les résultats de l'analyse

sont erronés.

La construction d'un tel plan est, en général, relativement facile. Ce qui l'est moins, c'est la méthode d'analyse statistique associée. Souvent, on commet une erreur, lors de l'analyse des résultats, en considérant les données comme provenant d'un plan complètement aléatoire (CRD). Le résultat de l'analyse est inexact dans le cas de mesures répétées s'il repose sur seul terme d'erreur dans le modèle. En particulier, les conclusions concernant les DAC. Le plan en mesures répétées renferme 2 termes d'erreur car il y deux types d'UE : les parcelles (WholePlot) (parcelles) et les parcelles divisées (SplitPlot). Les deux articles suivants contiennent beaucoup d'information pratique et des exemples sur la construction et l'analyse sur les plans en parcelles divisées:

Jones, B. Nachtsheim C. J. (2009) *Split-Plot Designs : What, Why and How,*
Journal of Quality Technology, vol. 41, no4, October, p. 340-388

Kowalski, S. M., Parker, P.A., Vining, G.G. (2007) *Tutorial: Industrial Split-plot Experiments*
Quality Engineering, vol 18, p. 1-15

Mesures répétées : un cas simple

Dans un cours de base sur les méthodes statistiques, on apprend à reconnaitre et à analyser le cas le plus simple de données en mesures répétées. C'est le test t de Student avec des données appariées ou pairées. Cette situation se présente lorsque l'on mesure les mêmes unités expérimentales sous 2 conditions différentes, disons A et B. Par exemple, on a effectué un test sur 15 composants (unités expérimentales) selon 2 conditions représentant les 2 modalités d'un facteur X. Les données peuvent être représentées comme dans le tableau suivant :

Comp. (UE) :	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Y_A :	7.6	10.2	9.5	1.3	3.0	6.3	5.3	6.2	2.2	4.8	11.3	12.1	6.9	7.6	8.4
Y_B :	7.3	9.1	8.4	1.5	2.7	5.8	4.9	5.3	2.0	4.2	11.0	11.0	6.1	6.7	7.5
D = Y_A-Y_B :	0.3	1.1	1.1	-0.2	0.3	0.5	0.4	0.9	0.2	0.6	0.3	1.1	0.8	0.9	0.9

Pour mesurer l'influence du facteur X sur la variable de réponse Y, l'analyse repose sur le test de Student de l'échantillon des différences D entre Y_A et Y_B. En réalité le test élimine la corrélation entre les données de Y_A et celles de Y_B. La méthode des différences pairées est opérationnelle car le facteur X a 2 modalités seulement. Si le facteur X prend 3 modalités ou plus on doit procéder autrement. On impose une structure de corrélation entre les mesures Y pour l'ensemble des modalités.

Exemple : supposons que le facteur X prend 5 modalités A B C D E et est appliqué à toutes les UE donc X est un facteur intra. La matrice de variance-covariance ΣY entre les 5 mesures de réponse Y, notées Y_A, Y_B, Y_C, Y_D et Y_E, prend la forme suivante appelée symétrie composée :

$$\Sigma_Y = \sigma^2 \begin{pmatrix} 1 & \rho & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho & \rho \\ \rho & \rho & 1 & \rho & \rho \\ \rho & \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & \rho & 1 \end{pmatrix}$$

C'est un modèle postulé dans l'analyse de variance en mesures répétées. Le test de Mauchly permet de décider si le modèle est applicable. On trouvera plus d'information sur les modèles statistiques avec une structure de dépendance dans l'ANNEXE 4.

Structure pour analyser des données avec des mesures répétées

On peut adopter 2 structures pour les données pour analyser des données en mesures répétées. La première structure est une forme empilée et connue sous nom d'approche unidimensionnelle. On identifie une seule variable de réponse Y et au moins une autre variable X donnant les modalités du facteur intra. Afin de bien reconnaitre la structure il est souhaitable d'identifier une autre variable qui identifie les unités expérimentales pour montrer quelles sont mesurées au moins 2 fois. Le tableau 1 qui suit est un exemple de cette structure. Les données de Y sont dépendantes et on adopte la structure de dépendance comme dans la matrice plus haut avec seulement 2 modalités A et B pour le facteur INTRA dans cet exemple. Le tableau 2 présente les mêmes données sous une forme désempilée avec 2 variables de réponse. Cette structure est la plus souvent recommandée dans les logiciels statistiques comme Statistica. Elle est plus générale et ne repose pas sur une structure de dépendance particulière pour les variables de réponse. C'est l'approche multidimensionnelle.

Tableau 1 : données empilées

Composant	X (facteur intra)	Y_réponse
C1	A	7,6
C2	A	10,2
C3	A	9,5
C4	A	1,3
C5	A	3,0
C6	A	6,3
C7	A	5,3
C8	A	6,2
C9	A	2,2
C10	A	4,8
C11	A	11,3
C12	A	12,1
C13	A	6,9
C14	A	7,6
C15	A	8,4
C1	B	7,3
C2	B	9,1
C3	B	8,4
C4	B	1,5
C5	B	2,7
C6	B	5,8
C7	B	4,9
C8	B	5,3
C9	B	2,0
C10	B	4,2
C11	B	11,0
C12	B	11,0
C13	B	6,1
C14	B	6,7
C15	B	7,5

Tableau 2 : données désempilées

Composant	Y_A	Y_B
C1	7,6	7,3
C2	10,2	9,1
C3	9,5	8,4
C4	1,3	1,5
C5	3,0	2,7
C6	6,3	5,8
C7	5,3	4,9
C8	6,2	5,3
C9	2,2	2,0
C10	4,8	4,2
C11	11,3	11,0
C12	12,1	11,0
C13	6,9	6,1
C14	7,6	6,7
C15	8,4	7,5

Le tableau 2 est la structure naturelle pour présenter les données provenant d'expériences en mesures répétées car chaque ligne du tableau est associée chaque unité d'expérimentale. Les mesures répétées sur chaque unité (composante) est représentée par autant de variables de réponse correspondant aux facteurs INTRA. Cette structure offre la plus grande flexibilité pour l'analyse statistique.

Le tableau 3 est un exemple typique de données en mesures répétées avec des facteurs INTRA sujet et des facteurs INTER sujet.

Tableau 3 : exemple de données avec 2 facteurs INTRA (TIME et DIAL) et 1 facteur INTER (NOISE2)

	1 ID	2 sujet2	3 NOISE2	4 DIAL	5 TIME	6 Y	7 c7	8 sujet	9 noise	10 Y_t1d1	11 Y_t1d2	12 Y_t1d3	13 Y_t2d1	14 Y_t2d2	15 Y_t2d3	16 Y_t3d1	17 Y_t3d2	18 Y_t3d3
1	1	s1	M	d1	t1	45		s1	M	45	53	60	40	52	57	28	37	46
2	2	s2	M	d1	t1	35		s2	M	35	41	50	30	37	47	25	32	41
3	3	s3	M	d1	t1	60		s3	M	60	65	75	58	54	70	40	47	50
4	4	s4	W	d1	t1	50		s4	W	50	48	61	25	34	51	16	23	35
5	5	s5	W	d1	t1	42		s5	W	42	45	55	30	37	43	22	27	37
6	6	s6	W	d1	t1	56		s6	W	56	60	77	40	39	57	31	29	46
7	7	s1	M	d2	t1	53												
8	8	s2	M	d2	t1	41												

....

.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	1 ID	2 sujet2	3 NOISE2	4 DIAL	5 TIME	6 Y	7 c7	8 sujet	9 noise	10 Y_t1d1	11 Y_t1d2	12 Y_t1d3	13 Y_t2d1	14 Y_t2d2	15 Y_t2d3	16 Y_t3d1	17 Y_t3d2	18 Y_t3d3
49	49	s1	M	d3	t3	46												
50	50	s2	M	d3	t3	41												
51	51	s3	M	d3	t3	50												
52	52	s4	W	d3	t3	35												
53	53	s5	W	d3	t3	37												
54	54	s6	W	d3	t3	46												

Les colonnes 1 à 6 contiennent les données avec une seule variable de réponse Y selon le format empilé. Les colonnes 8 à 18 contiennent les mêmes données mais réorganisées selon le format désempilé. Il y 9 variables de réponse (colonnes 10 à 18) associées aux 2 facteurs INTRA. Ces variables ont une structure de dépendance puisque les 6 sujets de cet exemple sont mesurés à 9 reprises selon différentes conditions des facteurs INTRA. La colonne 9 identifie le facteur INTER et les colonnes 10 à 18 représentent 9 variables de réponse associées aux 2 facteurs INTRA. L'analyse statistique repose sur ces 9 variables de réponse en adoptant une structure de dépendance particulière dont les différents modèles de dépendance seront présentés dans l'ANNEXE 2.

Facteurs FIXÉS – Facteurs ALÉATOIRES – Modèle MIXTE

Une autre distinction concernant les facteurs est relative au contrôle des modalités (valeurs) qu'ils peuvent prendre. Dans une expérience planifiée il y a au moins un facteur dont les traitements sont directement attribués aux unités expérimentales. Le ou les facteurs correspondants sont alors déclarés facteurs FIXÉS. Mais si on a aucun contrôle pour fixer les valeurs prises par les facteurs, on peut considérer que l'on a affaire à des facteurs ALÉATOIRES. Les valeurs observées proviennent du hasard de l'échantillonnage d'une population infinie de valeurs. Si on a des données collectées passivement sans intervention sur les unités statistiques, on est en présence de données observationnelles. On traite les variables comme des facteurs fixés qui affectent la moyenne de Y de la variable de réponse. C'est le cas de l'analyse statistique supervisée avec des modèles prédictifs comme la régression, les réseaux de neurones, les arbres de classification et support vector machine (SVM). Dans ce cas, les données sont considérées comme un échantillon aléatoire provenant d'un processus associé générant les données. On peut considérer les données comme des instants d'observations sur les unités statistiques sur lesquelles on recueille les informations.

Toutefois dans le cas où les données proviennent d'un protocole (design) expérimental, la distinction entre facteur FIXÉ et facteur ALÉATOIRE a un impact distinct sur la variable de réponse Y. Les facteurs FIXÉS influencent la moyenne de Y tandis que les facteurs ALÉATOIRES influencent la variance de Y. Le modèle statistique pour analyser les données contient des termes aléatoires qui s'ajoutent au terme d'erreur expérimental. Ce dernier est toujours aléatoire par sa nature même. Avec des facteurs des deux types, on est en présence d'un modèle mixte. Des éléments de la théorie de ces modèles seront présentés dans l'ANNEXE2.

Parcelles divisées et unités statistiques

La nature des unités dépend en partie du domaine d'application de l'étude. Doit-on faire intervenir explicitement ces unités statistiques dans les modèles d'analyse ?

Dans le domaine des sciences de la vie (médical, pharmaceutique, biologie, etc.) les unités sont généralement hétérogènes (âge, genre, taille, etc.). Cette hétérogénéité est tenue en compte en associant des variables explicatives X qui seront explicites dans les modèles statistiques.

Dans le domaine des sciences physiques, chimiques et en ingénierie, les unités sont généralement formées de matériaux et sont généralement homogènes. On conçoit alors qu'il n'y a pas de variables explicatives X associées aux unités dans la formulation des modèles statistiques?

Il demeure toutefois une situation qui est indépendante de la nature homogène ou hétérogène des unités statistiques. Est-ce que l'on tient en compte le concept de la facilité / difficulté de changer les modalités des facteurs? En d'autres termes, l'expérience est-elle conduite dans un mode de randomisation restreinte avec des facteurs FAC et DAC? C'est souvent le cas dans les expériences industrielles avec des procédés complexes. Les essais sont conduits afin de faciliter leur conduite ou réduire la durée des essais. Si on fixe un facteur et que l'on fait varier les modalités d'un autre facteur, on a alors affaire à un plan en parcelles divisées (« SplitPlot »). En conséquence, il y 2 grandeurs d'unités statistiques : les grandes parcelles (« WholePlot ») et les parcelles divisées (« SplitPlot »). De plus, on doit ajouter dans le modèle statistique d'analyse un terme d'erreur aléatoire associé aux grandes parcelles ainsi qu'une relation d'emboitement (« nested ») entre les 2 facteurs impliqués.

Les plans en parcelles divisés sont fréquents et on doit tenir compte de cette particularité importante dans la spécification du modèle d'analyse statistique, sinon l'analyse sera erronée. Afin de clarifier le plan en parcelles divisées et le modèle correspondant nous allons illustrer avec un exemple.

Dans une expérience, on veut mesurer la résistance à la corrosion (Y) de barres d'acier en fonction de 2 facteurs. Le premier facteur est la température (T) que l'on fera varier à 3 niveaux 360, 370 380 (^{o} C). Le deuxième facteur est le type de recouvrement (C) selon 4 modalités : C1, C2, C3, C4. On a donc 12 traitements possibles que l'on doit assigner aux barres. On pourrait faire l'assignation en mode complètement aléatoire (CRD). Il faut alors 12 barres d'acier ou un multiple de 12 si on veut répéter l'expérience. Cela implique que la température, qui est un facteur DAC, va changer plusieurs fois en cours des tests et allonge évidemment durée des essais.

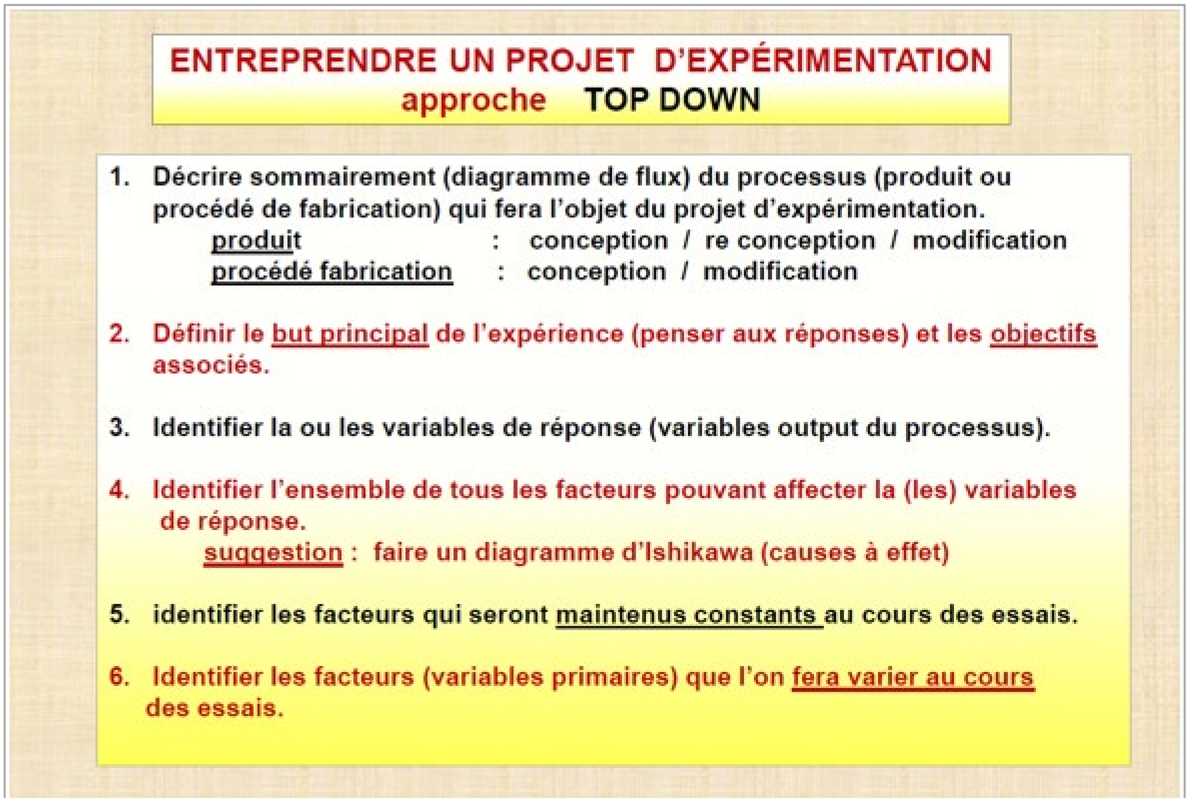
On décide donc de procéder d'une autre manière. On fixe la température T = 360 et ensuite on applique les 4 modalités de C en quatre positions sur la barre. On fait de même avec les 2 autres niveaux de température. On voit les avantages de ce plan : la température est changée 2 fois seulement, mais changera 9 fois (voir la colonne 12) dans le plan complètement randomisé (CRD). Le tableau 4 illustre la comparaison entre les 3 modes d'attribution des traitements aux unités expérimentales : aucune randomisation, complètement randomisé, parcelles divisées. Le mode parcelles divisées est particulièrement utile pour tenir en compte un facteur difficile à changer (DAC), ici la température.

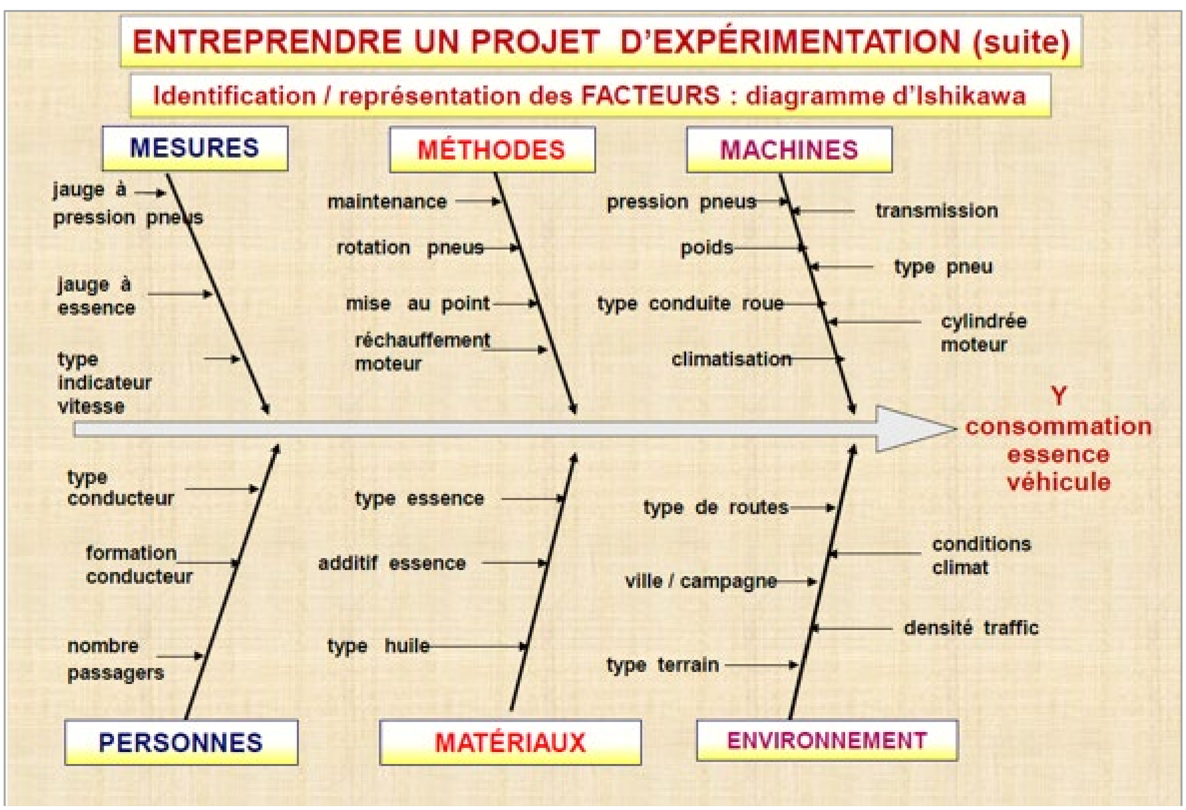
Tableau 4 : comparaison de 3 protocoles expérimentaux

1 info	2 trait_ id	3 T	4 C	5 nb ch T	6 Y1	7 info	8 ordre CRD	9 trait_ id	10 T	11 C	12 nb ch T	13 Y2	14 info	15 WP	16 T	17 C	18 trait_ id	19 nb ch T	20 Y3
plan non	t1	360	C1		.	plan CRD	2	t7	370	C3		.	plan splitPlot	wp1	360	C3	t3		.
randomisé	t2	360	C2		.		8	t1	360	C1	1	.		wp1	360	C2	t2		.
	t3	360	C3		.	complètement	11	t6	370	C2	2	.	randomisation	wp1	360	C1	t1		.
	t4	360	C4		.	aléatoire	4	t4	360	C4	3	.	de C à chaque	wp1	360	C4	t4		.
	t5	370	C1	1	.		7	t11	380	C3	4	.	changement de	wp2	370	C3	t7	1	.
	t6	370	C2		.		3	t8	370	C4	5	.	température	wp2	370	C4	t8		.
	t7	370	C3		.		1	t5	370	C1	5	.		wp2	370	C1	t5		.
	t8	370	C4		.		6	t2	360	C2	6	.		wp2	370	C2	t6		.
	t9	380	C1	2	.		10	t10	380	C2	7	.		wp3	380	C3	t11	2	.
	t10	380	C2		.		9	t9	380	C1	7	.		wp3	380	C1	t9		.
	t11	380	C3		.		5	t3	360	C3	8	.		wp3	380	C4	t12		.
	t12	380	C4		.		12	t12	380	C4	9	.		wp3	380	C2	t10		.

La conséquence de cette randomisation partielle qui fut appliqué dans le plan en parcelles divisées fait en sorte que le plan a 2 types d'unités expérimentales : les grandes parcelles, ici les barres, et les parcelles divisées représentées par les positions sur lesquelles l'on applique les recouvrements. La matrice des essais du plan en parcelles divisées est définie par les colonnes 15 à 20 du tableau 4. On remarque une colonne WP (colonne15) qui identifie les grandes parcelles. Chacune est composée de 4 sous parcelles c-à-d les parcelles divisées. Le facteur WP est déclaré est aléatoire et il vient s'ajouter aux 2 facteurs fixés T et C. On est automatiquement en présence d'un modèle mixte pour faire l'analyse.

ANNEXE 1 : Guide et Étapes d'un projet expérimental





ENTREPRENDRE UN PROJET D'EXPÉRIMENTATION (suite)

7. Préciser s'il y a des facteurs (variables) secondaires nuisibles connus qui seront contrôlés. remarque : différence entre 6 et 7 ?
   Facteurs primaires : ceux qui sont à l'origine du projet
   Facteurs secondaires nuisibles : ceux qui varient en cours d'expérimentation et que l'on ne peut pas maintenir constants.
   Si on peut les contrôler (fixer) alors on peut construire un plan en blocs qui permet de neutraliser leurs effets réels ou non sur la réponse.

8. Identifier les variables (facteurs) non contrôlées mais que l'on peut mesurer. Ils sont tenu en compte à titre de covariables lors de l'analyse des données de l'expérience.

9. Préciser la valeur minimale et la valeur maximale (intervalle de variation) de chaque facteur primaire quantitatif que l'on fera varier au cours des essais.
   remarque : explorer le plus grand espace possible mais éviter les régions problématiques

10. Préciser la liste des modalités de chaque facteur primaire qualitatif.
    remarque : 9 et 10 constitue l'espace d'expérimentation qui sera explorer avec les essais

ENTREPRENDRE UN PROJET D'EXPÉRIMENTATION (suite)

11. Anticiper la relation (augmentation / diminution) de la réponse avec chaque facteur
12. Préciser comment seront mesurés les variables de réponse.

    remarque : une étude du processus de mesurage est-elle nécessaire?

13. Selon l'état des connaissances sur le processus, proposer un ou plusieurs plans :

    plan de tamisage pour séparer les facteurs importants et ceux qui ne le sont pas

    séparation claire et nette des effets principaux et des effets d'interaction

    plan pour l'optimisation des réponses

14. Déterminer le nombre de répétition (n) de chaque essai.
15. Considérer l'ajout d'essais au centre de l'espace expérimental.
16. Existe-t-il des relations mathématiques connues entre la réponse et les facteurs?
17. Y a t –il des **RESTRICTIONS À LA RANDOMISATION** complète (ordre au hasard) des essais? remarque : certains facteurs sont-ils difficiles à changer?
18. Préciser tous les détails (protocole expérimental) pour l'exécution des essais.
19. Prévoir un budget et un échéancier pour le projet.

    Inclure des tests pour valider la ou les solutions résultant du projet.

ENTREPRENDRE UN PROJET D'EXPÉRIMENTATION (suite)

X → **Processus système** → **Y**

Identifier le RÔLE (entrée / sortie) des variables

RÔLE Y : variables dépendantes (sortie) qui seront mesurées
X : variables contrôlées que l'on fera varier selon un plan d'expérimentation

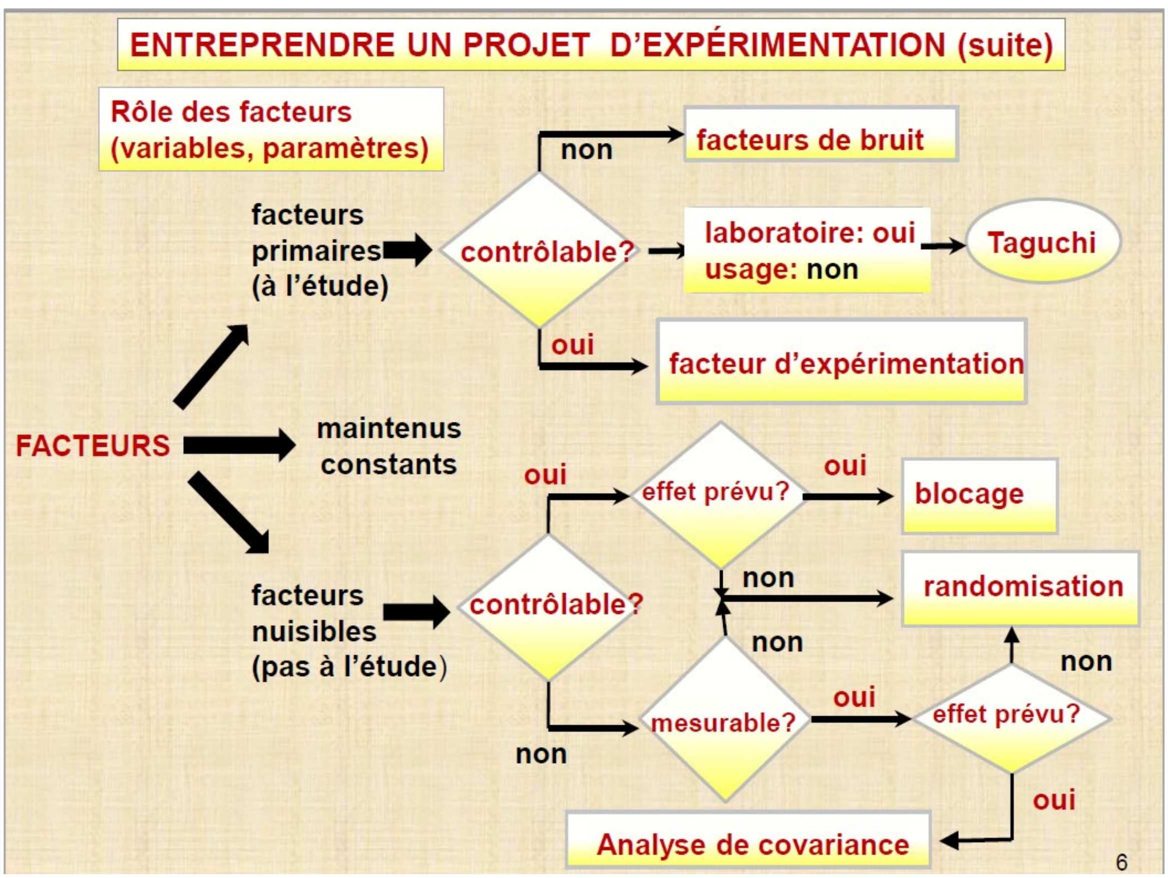
Préciser le TYPE (continues / catégoriques) des variables impliquées

Synonymes ingénieurs et scientifiques emploie le terme PARAMÈTRE plutôt que VARIABLE

Y : variables dépendantes = variables à expliquer
= variables de réponse = variable de sortie

X : variables indépendantes = variables explicatives
= variables d'input = facteurs d'expérimentation

Remarque: en statistique, terme PARAMÈTRE est réservé pour les constantes inconnues dans les modèles



ANNEXE 2 : modèles mixtes et structure de dépendance

Le modèle mixte est défini par l'équation suivante en notation matricielle :

$$Y = X\beta + Zu + \varepsilon$$

Où Y vecteur connu des valeurs de réponses

X matrice connue du design d'effet fixés
β vecteur inconnu d'effets fixés à estimer
Z matrice connue du design d'effets aléatoires
u vecteur inconnu d'effets aléatoires
ε vecteur non observé d'erreurs aléatoires

On suppose que

$u \sim N(0, G)$
$\varepsilon \sim N(0, R)$
$\mathrm{Cov}[u, \varepsilon] = 0$
Notation : $\sim N$: distribuée selon distribution normale multidimensionnelle

Où

G matrice variance-covariance de u
R matrice variance-covariance de ε

La matrice de variance-covariance de Y, notée V, est calculée par

$$\begin{aligned} V &= \mathrm{Var}[Y] \\ &= \mathrm{Var}[X\beta + Zu + \varepsilon] \\ &= 0 + \mathrm{Var}[Zu + \varepsilon] \\ &= ZGZ' + R \end{aligned}$$

Le but de l'analyse d'un modèle mixte est de tester les paramètres β. Mais avant il faut estimer les paramètres inconnus (β, G, and R). L'estimation de G et R repose sur des modèles de structure matricielle. Elles seront proposées plus loin dans ce texte.

Les hypothèses suivantes sont faites pour l'exécution des test F

1. **La variable de réponse est continue.**
2. **Les unités statistiques (pas les observations sur Y !) sont indépendantes.**
3. **Le terme d'erreur suit une distribution normale avec une moyenne de zéro.**

L'argument distinctif et le plus important du modèle mixte par comparaison du modèle linéaire usuel (effets fixés seulement) est la très grande flexibilité qu'Il offre pour modéliser différentes situations : hétérogénéité des variances, corrélation entre les observations, mesures répétées etc. L'estimation des matrices G et R peut se faire par la méthode de vraisemblance maximale (ML) ou par la méthode vraisemblance maximale restreinte (REML).

Formulation du modèle mixte au niveau des unités statistiques

$$Y_i = X_i \beta + Z_i u_i + \varepsilon_i \qquad i = 1, \ldots, N \text{ (nombre de sujets)}$$

y_i : $n_i \times 1$ vecteur des données du sujet i

X_i : matrice $n_i \times p$ des effets fixés du sujet i (p est le nombre de colonnes de X)

β : vecteur p×1 des paramètres de régression

Z_i : matrice $n_i \times q$ des effets aléatoires du sujet i

U_i : vecteur q×1 des effets aléatoires du sujet I ayant une moyenne de zéro et une matrice de variance-covariance Gsub

ε_i : vecteur n_i×1 des erreurs du sujet i avec une moyenne de zéro et une matrice de covariance Ri avec n_i meures répétées sur le sujet i.

N : nombre de sujets

e_i = $y_i - X_i\beta$: vecteur des résidus du sujet i

V_i = Var (y_i) = Z_i Gsub Z_i^t + R_i t : transposition

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}, \quad X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{pmatrix}, \quad Z = \begin{pmatrix} Z_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & Z_N \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix}$$

Il existe deux cas extrêmes de modèles mixtes : les modèles à effets fixés seulement et le modèle contenant que des effets aléatoires. Le cas général du modèle mixte contient les deux catégories d'effets. Un cas important du modèle mixte est le plan en mesurées. L'effet d'un facteur aléatoire n'est pas testé mais on s'intéresse à l'estimation de sa variance.

Un autre cas important de la nécessité des modèles mixtes est lorsque l'on mesure une réponse sur un sujet à plusieurs reprises en fonction du temps à des intervalles fixes ou aléatoires. On modélise avec deux types de sous -modèles : les modèles à coefficients aléatoires et les modèles de covariance ayant des formes spécifiques seront proposés plus loin.

Dans les modèles à coefficients aléatoires on s'intéresse à la relation entre la réponse et le temps quand le temps n'est pas fixé à intervalles réguliers. Le temps est alors un facteur aléatoire. On inclut alors le temps comme une covariable et en modélisant chaque sujet avec un intercepte et un terme d'interaction sujet*temps

comme la pente associée à chaque sujet. Il est raisonnable de spécifier ces termes comme des effets aléatoires et on utilise la distribution bi normale. On tient compte alors de la corrélation entre ces effets aléatoires.

$$\begin{pmatrix} subject_k \\ (subject * time)_k \end{pmatrix} \sim N(0, \mathbf{G}),$$

$$\mathbf{G} = \begin{pmatrix} \sigma^2_{subject} & \sigma_{subject,subject*time} \\ \sigma_{subject,subject*time} & \sigma^2_{subject*time} \end{pmatrix}.$$

Ces modèles à coefficients aléatoires sont différents des modèles classiques usuels quand on ajuste des droites car dans ce cas les estimateurs de la pente et de l'intercepte ne sont pas indépendants.

Exemples de structures des matrice G et R

Matrice G

$$\mathbf{G} = \begin{pmatrix} G_{sub} & 0 & 0 & \cdots & 0 \\ 0 & G_{sub} & 0 & \cdots & 0 \\ 0 & 0 & G_{sub} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & G_{sub} \end{pmatrix}$$

Structures G_{sub}

diagonale

$$\mathbf{G}_{sub} = \begin{pmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \sigma_3^2 & \\ & & & \sigma_4^2 \end{pmatrix}$$

pas de structure

$$\mathbf{G}_{sub} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{pmatrix}$$

Matrices R

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_1 & 0 & 0 & \cdots & 0 \\ 0 & \mathbf{R}_2 & 0 & \cdots & 0 \\ 0 & 0 & \mathbf{R}_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \mathbf{R}_N \end{pmatrix}$$

$$\mathbf{R}_{Sub} = \begin{pmatrix} \sigma_1^2 & & & & \\ & \sigma_2^2 & & & \\ & & \sigma_3^2 & & \\ & & & \sigma_4^2 & \\ & & & & \sigma_5^2 \end{pmatrix},$$

Diagonal

Homogeneous

$$\begin{pmatrix} \sigma^2 & & & \\ & \sigma^2 & & \\ & & \sigma^2 & \\ & & & \sigma^2 \end{pmatrix}$$

Heterogeneous

$$\begin{pmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \sigma_3^2 & \\ & & & \sigma_4^2 \end{pmatrix}$$

Correlation

$$\begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}$$

Compound Symmetry

Homogeneous

$$\begin{pmatrix} \sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \sigma^2 \end{pmatrix}$$

Heterogeneous

$$\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho\sigma_1\sigma_3 & \rho\sigma_1\sigma_4 \\ \rho\sigma_2\sigma_1 & \sigma_2^2 & \rho\sigma_2\sigma_3 & \rho\sigma_2\sigma_4 \\ \rho\sigma_3\sigma_1 & \rho\sigma_3\sigma_2 & \sigma_3^2 & \rho\sigma_3\sigma_4 \\ \rho\sigma_4\sigma_1 & \rho\sigma_4\sigma_2 & \rho\sigma_4\sigma_3 & \sigma_4^2 \end{pmatrix}$$

Correlation

$$\begin{pmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{pmatrix}$$

AR(1)

Homogeneous

$$\begin{pmatrix} \sigma^2 & \rho\sigma^2 & \rho^2\sigma^2 & \rho^3\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \rho^2\sigma^2 \\ \rho^2\sigma^2 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ \rho^3\sigma^2 & \rho^2\sigma^2 & \rho\sigma^2 & \sigma^2 \end{pmatrix}$$

Heterogeneous

$$\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho^2\sigma_1\sigma_3 & \rho^3\sigma_1\sigma_4 \\ \rho\sigma_2\sigma_1 & \sigma_2^2 & \rho\sigma_2\sigma_3 & \rho^2\sigma_2\sigma_4 \\ \rho^2\sigma_3\sigma_1 & \rho\sigma_3\sigma_2 & \sigma_3^2 & \rho\sigma_3\sigma_4 \\ \rho^3\sigma_4\sigma_1 & \rho^2\sigma_4\sigma_2 & \rho\sigma_4\sigma_3 & \sigma_4^2 \end{pmatrix}$$

Correlation

$$\begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}$$

Toeplitz

Homogeneous

$$\begin{pmatrix} \sigma^2 & \rho_1\sigma^2 & \rho_2\sigma^2 & \rho_3\sigma^2 \\ \rho_1\sigma^2 & \sigma^2 & \rho_1\sigma^2 & \rho_2\sigma^2 \\ \rho_2\sigma^2 & \rho_1\sigma^2 & \sigma^2 & \rho_1\sigma^2 \\ \rho_3\sigma^2 & \rho_2\sigma^2 & \rho_1\sigma^2 & \sigma^2 \end{pmatrix}$$

Heterogeneous

$$\begin{pmatrix} \sigma_1^2 & \rho_1\sigma_1\sigma_2 & \rho_2\sigma_1\sigma_3 & \rho_3\sigma_1\sigma_4 \\ \rho_1\sigma_2\sigma_1 & \sigma_2^2 & \rho_1\sigma_2\sigma_3 & \rho_2\sigma_2\sigma_4 \\ \rho_2\sigma_3\sigma_1 & \rho_1\sigma_3\sigma_2 & \sigma_3^2 & \rho_1\sigma_3\sigma_4 \\ \rho_3\sigma_4\sigma_1 & \rho_2\sigma_4\sigma_2 & \rho_1\sigma_4\sigma_3 & \sigma_4^2 \end{pmatrix}$$

Correlation

$$\begin{pmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_3 & \rho_2 & \rho_1 & 1 \end{pmatrix}$$

Toeplitz(2)

Homogeneous

$$\begin{pmatrix} \sigma^2 & \rho_1\sigma^2 & & \\ \rho_1\sigma^2 & \sigma^2 & \rho_1\sigma^2 & \\ & \rho_1\sigma^2 & \sigma^2 & \rho_1\sigma^2 \\ & & \rho_1\sigma^2 & \sigma^2 \end{pmatrix}$$

Heterogeneous

$$\begin{pmatrix} \sigma_1^2 & \rho_1\sigma_1\sigma_2 & & \\ \rho_1\sigma_2\sigma_1 & \sigma_2^2 & \rho_1\sigma_2\sigma_3 & \\ & \rho_1\sigma_3\sigma_2 & \sigma_3^2 & \rho_1\sigma_3\sigma_4 \\ & & \rho_1\sigma_4\sigma_3 & \sigma_4^2 \end{pmatrix}$$

Correlation

$$\begin{pmatrix} 1 & \rho_1 & & \\ \rho_1 & 1 & \rho_1 & \\ & \rho_1 & 1 & \rho_1 \\ & & \rho_1 & 1 \end{pmatrix}$$

Unstructured

Homogeneous

$$\begin{pmatrix} \sigma^2 & \rho_{12}\sigma^2 & \rho_{13}\sigma^2 & \rho_{14}\sigma^2 \\ \rho_{21}\sigma^2 & \sigma^2 & \rho_{23}\sigma^2 & \rho_{24}\sigma^2 \\ \rho_{31}\sigma^2 & \rho_{32}\sigma^2 & \sigma^2 & \rho_{34}\sigma^2 \\ \rho_{41}\sigma^2 & \rho_{42}\sigma^2 & \rho_{43}\sigma^2 & \sigma^2 \end{pmatrix}$$

Heterogeneous

$$\begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 & \rho_{14}\sigma_1\sigma_4 \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 & \rho_{24}\sigma_2\sigma_4 \\ \rho_{31}\sigma_3\sigma_1 & \rho_{32}\sigma_3\sigma_2 & \sigma_3^2 & \rho_{34}\sigma_3\sigma_4 \\ \rho_{41}\sigma_4\sigma_1 & \rho_{42}\sigma_4\sigma_2 & \rho_{43}\sigma_4\sigma_3 & \sigma_4^2 \end{pmatrix}$$

Correlation

$$\begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{14} \\ \rho_{21} & 1 & \rho_{23} & \rho_{24} \\ \rho_{31} & \rho_{32} & 1 & \rho_{34} \\ \rho_{41} & \rho_{42} & \rho_{43} & 1 \end{pmatrix}$$