

Statistiques appliquées aux sciences

SCI 1018

Élaboration et optimisation de plans d'échantillonnage

Marc J. Mazerolle

Institut de recherche sur les forêts, Université du Québec en Abitibi-Témiscamingue



© TÉLUQ, 2013

Table des matières

| | | |
|----------|--|-----------|
| 1 | Introduction | 2 |
| 2 | Plans d'échantillonnage | 2 |
| 2.1 | Expérience <i>vs</i> étude d'observation | 2 |
| 2.2 | Témoin | 3 |
| 2.3 | Placébo | 3 |
| 2.4 | Prise de données en aveugle | 4 |
| 2.5 | Répétition, randomisation et répartition spatiale | 5 |
| 3 | Nos meilleurs alliés dans l'élaboration d'un plan d'échantillonnage | 7 |
| 4 | Problèmes associés à certains plan d'échantillonnage | 9 |
| 4.1 | L'absence de réelles répétitions : la pseudoréplication | 9 |
| 4.1.1 | Pseudoréplication simple | 10 |
| 4.1.2 | Pseudoréplication sacrificielle | 11 |
| 4.1.3 | Pseudoréplication temporelle | 11 |
| 4.2 | Contrôler la variabilité temporelle | 12 |
| 4.3 | Sources de confusion | 12 |
| | Conclusion | 14 |
| | Index | 15 |

1 Introduction

La leçon précédente a illustré les concepts d'effets fixes et d'effets aléatoires ainsi que de modèle mixte. L'ANOVA en blocs complets aléatoires a été présentée comme étant le cas le plus simple de modèle mixte et comme l'extension du test t pour les données appariées. Le dispositif en blocs complets aléatoires comporte des avantages importants, notamment sur le plan logistique comparé aux dispositifs complètement aléatoires (p. ex., ANOVA à un critère avec répétitions, ANOVA à deux critères avec répétitions). Dans la présente leçon, nous discuterons de façon plus générale de l'élaboration de dispositifs expérimentaux. Nous fournirons plusieurs conseils afin d'arriver à des conclusions solides.

2 Plans d'échantillonnage

Dans ce texte, nous parcourrons différents éléments qui sont importants dans l'élaboration de **plans d'échantillonnage** aussi appelés **plans d'expérience** ou **dispositifs expérimentaux** (*experimental design, sampling design*) afin de tirer le maximum d'information de votre expérience et d'arriver à des conclusions solides. Avant d'aller plus loin, nous allons définir quelques termes qui reviendront au cours de la leçon. Certains de ces termes ont été présentés dans des leçons antérieures, mais nous tenons à les rappeler ici.

2.1 Expérience *vs* étude d'observation

Une **étude d'observation** consiste à prendre des mesures sur des unités expérimentales sélectionnées aléatoirement, mais où les traitements n'ont pas nécessairement été attribués aléatoirement. Par exemple, dans une étude sur l'effet de différents traitements sylvicoles, on peut choisir des sites ayant subi l'un de trois traitements (coupe totale, coupe partielle, aucune coupe). Bien que l'on ait choisi aléatoirement les sites, les traitements n'ont pas été assignés de façon aléatoire, à moins que l'on ait travaillé en collaboration avec une compagnie d'exploitation forestière. Les résultats ne sont pas aussi clairs que ceux d'une vraie **expé-**

rience aussi appelée **expérience contrôlée**. En effet, il est difficile de contrôler plusieurs facteurs dans une étude d'observation.

À l'opposé, l'expérience consiste à prendre des mesures sur des unités expérimentales sélectionnées aléatoirement, auxquelles on a assigné aléatoirement un certain traitement. Ce genre de dispositif expérimental est celui que nous avons considéré dans le cas de l'ANOVA à un et à deux critères, par exemple. Dans l'expérience, nous contrôlons tous les facteurs et nous ne faisons varier que ceux qui sont d'intérêt, ce qui nous permet d'arriver à des conclusions solides. Bien que l'expérience soit souvent l'approche idéale, elle manque parfois de réalisme, et l'étude d'observation peut procurer des informations supplémentaires à propos de l'effet potentiel de certaines variables sur le système d'intérêt.

2.2 Témoin

Le **témoin** ou **contrôle** (*control*) d'une expérience est un traitement qui consiste en une absence de traitement. Autrement dit, c'est un traitement qui présente des conditions de référence. Par exemple, dans une expérience qui compare l'effet de deux médicaments sur des patients, on pourrait considérer un troisième groupe de patients auxquels on n'administre pas de médicament. C'est le groupe témoin ou le témoin. L'utilisation de témoin est primordiale afin d'obtenir une moyenne de référence que l'on peut comparer aux moyennes de groupes avec traitements. Le témoin contrôle les changements temporels ou les effets associés à des procédures ou à des manipulations liées à certains traitements.

2.3 Placébo

Le **placébo** (*placebo*) est un traitement témoin qui consiste à administrer une substance neutre à des individus afin de contrôler l'effet de la manipulation. Par exemple, imaginons une expérience visant à déterminer l'effet d'un nouveau médicament. On administre le médicament sous forme de comprimé à un premier groupe tandis que le deuxième groupe reçoit un comprimé de sucre (placébo) et un troisième groupe ne subit aucun traitement (témoin).

Comme autre exemple, dans une expérience qui vise à étudier l'activité d'animaux en présence de prédateurs, on soumet un premier groupe d'animaux à la présence d'un prédateur vivant maintenu dans un enclos. On soumet un deuxième groupe à la présence d'un prédateur empaillé (placébo), tandis qu'un troisième groupe n'est exposé à aucun prédateur (témoin).

Le placebo est important lorsqu'il existe un effet de manipulation (p. ex., administrer un comprimé, la présence d'un prédateur vivant dans l'enclos). La comparaison d'un groupe placebo (avec manipulation) à un groupe témoin (sans manipulation) permet de quantifier l'effet de la manipulation sur les individus utilisés dans l'expérience. La comparaison d'un groupe placebo à un groupe qui a reçu un traitement permet de conclure qu'il y a bel et bien un effet du traitement en tenant compte de l'effet de manipulation.

2.4 Prise de données en aveugle

Un moyen parfois utilisé d'éviter les biais associé à la prise de données est la **prise de données en aveugle** aussi appelée **prise de données à l'insu** (*blinding*). Pour ce faire, les personnes qui récoltent les données ou les sujets de l'expérience ignorent le traitement qui a été appliqué. Ceci permet d'éviter le cas où un technicien trop zélé introduit consciemment ou inconsciemment un biais dans les données en favorisant un des traitements. Le même phénomène peut survenir lorsqu'un patient sachant qu'il reçoit un certain traitement influence sa réponse. Autrement dit, un patient à qui on annonce qu'il reçoit le médicament peut avoir une réponse plus forte au traitement qu'un patient qui ne connaît pas la nature du traitement. La prise de données en aveugle peut régler ce problème. La **prise de données en double aveugle** ou **prise de données à double insu** (*double blinding*) implique que les personnes récoltant les données ainsi que les participants à l'expérience ignorent les traitements qui ont été administrés.

2.5 Répétition, randomisation et répartition spatiale

La **répétition** des traitements ou des combinaisons des traitements contrôle la variabilité due au hasard aussi appelée **stochasticité** lors d'une expérience. En d'autres mots, la présence de répétitions peut protéger contre la présence d'éléments perturbateurs qui agissent sur les unités expérimentales (p. ex., un arbre tombe sur un des bassins contenant des têtards dans une expérience sur la compétition, un prédateur se nourrit des animaux dans un des enclos d'une expérience sur la sélection d'habitat) : si certaines des unités sont contaminées, celles qui restent pourront tout de même être utilisées pour mener à bien l'expérience.

La **randomisation** contrôle le biais potentiel de la sélection, de l'emplacement, et de l'attribution des traitements aux unités expérimentales. La randomisation aide à assurer l'indépendance des observations (ou des erreurs), une condition essentielle à l'utilisation de plusieurs analyses statistiques.

La **répartition spatiale** ou **temporelle** (selon le cas) contrôle la variation spatiale ou temporelle des propriétés des unités expérimentales. Autrement dit, les conditions de l'emplacement où se trouve une unité expérimentale peuvent être très différentes des conditions à un autre emplacement. En disposant les parcelles à différents endroits sur le site d'étude, on obtient une meilleure estimation de la variabilité naturelle des unités expérimentales. Ceci permet de tirer des conclusions plus générales. De plus, établir les unités expérimentales à une distance minimale les unes des autres assure aussi l'indépendance des observations. Le prochain exemple illustre certains problèmes liés à la répétition, la randomisation et la répartition spatiale qui peuvent survenir lors de l'élaboration d'un plan d'échantillonnage.

Exemple 9.1 On s'intéresse à la décomposition des feuilles à différentes profondeurs dans un lac. On émet l'hypothèse scientifique que la décomposition des feuilles est plus lente en profondeur que près de la surface. Pour tester notre hypothèse, on place un sac de feuilles de bouleau blanc (*Betula papyrifera*) à 0.5 m de

profondeur à un certain endroit d'un lac et un autre sac à 5 m de profondeur au même endroit. Si chaque sac fournit une seule mesure de décomposition, le nombre de répétitions sera insuffisant et on ne pourra pas tester d'hypothèse. On devra donc modifier le plan d'échantillonnage afin d'incorporer plus de répétitions.

On reprend l'expérience, cette fois-ci en augmentant la répétition et en disposant 10 sacs de feuilles à chacune des deux profondeurs au même endroit dans le lac. Malgré la présence de répétitions à chaque profondeur, un problème est toujours évident : le dispositif modifié permettra uniquement de déterminer s'il y a des différences entre les deux profondeurs à cet endroit précis. On ne pourra pas tirer de conclusion sur le lac en entier. En effet, les conditions (p. ex., concentration en oxygène, turbidité, température, ensoleillement) à un point donné du lac peuvent être différentes de celles à d'autres endroits du même lac. Afin de pouvoir décrire la décomposition des feuilles dans le lac, il est nécessaire de répartir les sacs à différents endroits du lac (c.-à-d., répartition spatiale) de façon aléatoire afin d'obtenir une meilleure estimation de la variabilité de la décomposition à une profondeur donnée.

Dans le but d'améliorer la répartition spatiale, on reprend (encore) l'expérience en disposant 10 sacs à un endroit à 0.5 m de profondeur et les 10 autres sacs à 5 m de profondeur à un deuxième emplacement. Avec un tel dispositif, il sera impossible de distinguer l'effet de la profondeur de celui de l'effet de l'emplacement du sac. En d'autres mots, l'emplacement et la profondeur seront deux variables confondues l'une avec l'autre. Il se peut que les conditions ne soient pas homogènes entre les deux emplacements, ce qui rendra difficile la comparaison. Si on trouve une différence, il n'y a aucun moyen de prouver que cela est dû uniquement à la différence de profondeur, l'emplacement pourrait aussi expliquer cette différence. La solution ultime (et le meilleur plan d'échantillonnage) consistera à disposer

chaque sac à un endroit du lac différent (déterminé aléatoirement) selon les deux profondeurs, afin d'obtenir 20 sacs à 20 endroits différents du lac (10 sacs à 0.5 m et 10 sacs à 5 m de profondeur), ce qui correspondrait à un dispositif complètement aléatoire. Contrairement aux plans d'échantillonnage présentés plus haut, cette dernière approche permettra d'obtenir une mesure de la décomposition à l'une ou l'autre des profondeurs (0.5 m ou 5 m) sans ambiguïtés.

Dans cet exemple, aucun témoin n'a été présenté. L'inclusion d'un témoin améliorerait le dispositif. Un témoin dans le contexte de cette expérience pourrait s'avérer être un groupe de sacs à une profondeur de 0 m (directement à l'interface de l'eau et de l'air) ou encore des feuilles provenant d'une espèce alternative de feuillu ou de conifère. Le cas échéant, les sacs témoins devraient être répartis aléatoirement dans le lac. Ceci permettrait de tester efficacement l'effet de la profondeur sur la décomposition des feuilles en contrôlant la variabilité des endroits dans le lac (répartition spatiale), en intégrant la randomisation et la répétition. Un dispositif alternatif consisterait à utiliser des blocs complets aléatoires en sélectionnant aléatoirement 10 endroits sur le lac et en disposant un trio de sacs à chaque endroit (1 sac à 0.5 m, 1 sac à 5 m, 1 sac témoin).

3 Nos meilleurs alliés dans l'élaboration d'un plan d'échantillonnage

Lors de l'élaboration d'un plan d'échantillonnage, nos meilleurs alliés sont :

- une hypothèse scientifique claire et bien définie ;
- l'échantillonnage aléatoire des observations parmi la population statistique d'intérêt ;
- l'assignation aléatoire des traitements aux unités expérimentales ;

- l'utilisation de témoins ;
- l'utilisation de blocs (complets aléatoires) lorsque le milieu est hétérogène ;
- la stratification lorsque des conditions différentes sont présentes dans la population (p. ex., différents habitats, différentes régions géographiques) ;
- l'utilisation de variables additionnelles (p. ex., pH, type de sol) si les conditions hétérogènes sont susceptibles d'influencer la variable réponse et qu'on veut quantifier cet effet ;
- la répétition, la répétition et la répétition (une bonne taille d'échantillon peut pallier plusieurs problèmes) ;
- l'équilibre du nombre d'unités expérimentales par traitement (dispositif équilibré).

Avant de réaliser une expérience et de récolter les données, il est très important de répondre à certaines questions, notamment :

- Quelle est l'hypothèse scientifique à tester ?
- Quel est le meilleur dispositif qui permettra de tester simplement et efficacement l'hypothèse ?
- Est-ce une expérience (avec manipulations) ou une étude d'observation (sans manipulations) ?
- Quelle est la variable réponse et quelles sont les variables explicatives ?
- Le dispositif que j'ai choisi est-il un dispositif complètement aléatoire ou y a-t-il des unités expérimentales avec une structure particulière (p. ex., observations groupées entre elles) ?
- Y a-t-il de l'hétérogénéité environnementale ou temporelle ? Le cas échéant, comment cette hétérogénéité a-t-elle été prise en compte ou mesurée ?
- Les traitements ont-ils été attribués aléatoirement aux unités expérimentales ?
- Mon dispositif est-il balancé (équilibré) et ai-je suffisamment de répétitions ?
- Quelle analyse statistique permettra de tester le plus simplement et efficacement l'hypothèse scientifique ?

- Pour l'analyse statistique choisie, quelles sont les suppositions à respecter afin que les conclusions soient valides ?
- Mes observations sont-elles réellement indépendantes ? Si elles ne le sont pas, comment l'analyse tient-elle compte de la non-indépendance des données ?
- Y a-t-il des variables confondantes (*counfounding variable*, *counfounding effect*) qui rendront potentiellement difficile l'interprétation des résultats ? On entend par variable confondante, une variable qui est intimement liée à une autre et qui rend difficile voire impossible de distinguer l'effet des deux variables sur la variable réponse. Par exemple, considérons une expérience où l'on a trois sites auxquels on assigne un traitement différent à chacun. Si on trouve une différence entre les traitements au stade de l'analyse, cette différence peut aussi bien être due à l'effet du traitement ou à l'effet du site car chaque site a reçu un traitement différent. On dira donc que l'effet du traitement et du site sont confondus. Sans plus d'information (répétitions des sites) ou l'utilisation de blocs (p. ex., l'application de tous les traitements à chaque site), on ne peut rien conclure et on aura dépensé inutilement du temps et de l'argent.

4 Problèmes associés à certains plan d'échantillonnage

4.1 L'absence de réelles répétitions : la pseudoréplication

Au milieu des années 1980, un terme est apparu pour désigner des problèmes liés à la répétition de traitements d'une expérience. Hurlbert (1984¹) a introduit le terme **pseudoréplication** dans un article qui a été énormément cité en écologie. En gros, la pseudoréplication désigne une expérience mal élaborée qui ne possède pas de réelles répétitions. Trois différents types de pseudoréplication ont été présentés dans l'article original, soit la pseudoréplication simple, la pseudoréplication sacrificielle et la pseudoréplication temporelle. Nous décrivons chaque type dans les prochaines lignes.

1. Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* 54 :187-211.

4.1.1 Pseudoréplication simple

La **pseudoréplication simple** (*simple pseudoreplication*) est le cas où on a plusieurs observations, mais seulement une vraie unité expérimentale par traitement (absence de réelles répétitions). On peut avoir plusieurs mesures dans deux parcelles, mais les observations ne sont pas indépendantes. Par exemple, si on s'intéresse à l'effet d'un herbicide sur la densité de coléoptères, on pourrait diviser un champ en deux. On pourrait établir des quadrats dans chaque moitié du champ et mesurer la densité de coléoptères avant d'appliquer un herbicide à l'une des moitiés du champ. Une seconde visite aux quadrats permettrait de déterminer le changement avant et après le traitement. Même avec 200 quadrats dans chaque moitié du champ, on a seulement deux réelles unités expérimentales : les deux moitiés du champ. Effectivement, les quadrats dans la même moitié du champ ne sont pas indépendants. Si on détecte une différence entre les deux moitiés du champ, il se peut que ce soit dû à une différence des caractéristiques des deux moitiés de champ (ensoleillement, drainage, humidité) ou de l'application de l'herbicide : les deux effets sont confondus.

Dans cet exemple, il est nécessaire d'augmenter le nombre d'unités expérimentales plutôt que les observations sur une même unité. Pour ce faire, il faut disposer aléatoirement les quadrats dans le champ et assigner aléatoirement le traitement à chacun pour obtenir un dispositif complètement aléatoire (1 quadrat = 1 unité expérimentale). Si ceci est logistiquement impossible, on pourrait diviser le champ en huit sections assez grand pour permettre le passage d'un tracteur pour épandre l'herbicide, en appliquant l'herbicide sur quatre sections du champ et en établissant des quadrats dans chaque section. Le cas échéant, on aurait huit unités expérimentales puisque les quadrats provenant de la même section du champ ne seraient pas considérés comme indépendants. Toutefois, avec cette structure, il faudrait utiliser les moyennes de chaque section de champ dans les analyses ou utiliser une analyse qui peut tenir compte de structures plus complexes².

2. Les analyses de dispositifs plus complexes ne feront pas l'objet de ce cours.

4.1.2 Pseudoréplication sacrificielle

La **pseudoréplication sacrificielle** (*sacrificial pseudoreplication*) survient lorsqu'on a plusieurs observations sur les mêmes unités expérimentales et que l'on combine ces observations dans la même analyse sans tenir compte de leur origine. Ici, nous avons une répétition des traitements (plusieurs unités par traitement), mais on considère à tort que chaque observation est une vraie unité expérimentale. Cette erreur est souvent justifiée par des énoncés tels que « Les sites ne différaient pas entre eux et nous avons combiné les données de chacun dans la même analyse ... » La réalité, en milieu naturel, est qu'il est pratiquement impossible d'observer deux sites identiques et il est incorrect de simplement regrouper toutes les observations ensembles. Généralement, les observations qui proviennent d'une même unité expérimentale (p. ex., un site, un quadrat ou un arbre) sont plus similaires entre elles que deux observations provenant de deux unités expérimentales différentes. Comme solution, on peut appliquer les mêmes remèdes utilisés pour faire face à la pseudoréplication simple, soit augmenter le nombre de réelles unités expérimentales, travailler sur les moyennes des unités expérimentales (qui elles sont indépendantes les unes des autres), ou encore utiliser une analyse statistique qui permet d'incorporer les observations répétées dans chaque unité expérimentale.

4.1.3 Pseudoréplication temporelle

La **pseudoréplication temporelle** (*temporal pseudoreplication*) se produit lorsqu'on prend des mesures successives sur les mêmes unités expérimentales étalées dans le temps, et que l'on considère ces mesures à différentes périodes comme des observations indépendantes. Les observations qui proviennent d'une même unité expérimentale à différentes périodes ne sont pas indépendantes. Par exemple, des mesures de recouvrement d'arbres feuillus dans 20 sites à trois différentes périodes ne constituent pas des observations indépendantes. Une solution à ce problème consiste à analyser chaque période séparément ou à utiliser une analyse qui tient compte des mesures répétées (p. ex., ANOVA pour mesures répétées, modèles

mixtes).

4.2 Contrôler la variabilité temporelle

Un type particulier de dispositif expérimental concerne l'application de traitements dans le temps, et la prise de mesure typiquement **avant et après** le traitement (*before and after controlled impact design (BACI)*). Par exemple, dans une expérience en milieu forestier, on sélectionne 12 peuplements indépendants et l'on mesure le nombre de pousses de régénération d'arbres feuillus. On applique ensuite l'un de deux traitements sylvicoles (coupe partielle et coupe totale) aux 12 peuplements et on rééchantillonne les peuplements l'année suivante pour dénombrer les pousses de régénération de feuillus. Bien que chaque peuplement puisse agir comme son propre témoin (conditions avant *vs* après), l'application d'un traitement dans le temps présente un problème : il est impossible de distinguer l'effet du traitement d'un effet temporel si toutes les unités reçoivent un traitement.

Toute expérience ou étude d'observation s'échelonnant sur le temps nécessite un témoin temporel. En d'autres mots, on a besoin d'unités expérimentales qui n'ont reçu aucun traitement dans le temps afin de procurer une mesure de référence temporelle (c.-à-d., ce qui se passe en l'absence de traitement entre les périodes de mesures). Dans l'exemple ci-haut sur la régénération de feuillus dans 12 peuplements, il serait important d'inclure un troisième traitement (témoin), afin d'obtenir quatre répétitions de chaque traitement. Cette modification du dispositif expérimental permettrait de distinguer les différences avant et après l'application de chaque traitement, en tenant compte des différences dues uniquement au temps écoulé entre deux mesures (le témoin avant et après).

4.3 Sources de confusion

Lorsqu'on effectue une expérience ou une étude d'observation, certains éléments peuvent influencer la qualité des conclusions, ce que Hurlbert (1984) désigna comme des **sources de confusion**. Parmi celles-ci, on note les changements temporels qui surviennent pendant l'ex-

périence modifiant les propriétés des unités expérimentales. L'utilisation de témoins permet de pallier ce problème. Les **effets liés aux procédures**, comme la manipulation d'animaux ou la perturbation du sol lors de la prise d'une mesure, sont d'autres sources de confusion. Ici encore, l'utilisation de témoins atténue ce problème. Le **biais associé à l'expérimentateur** est une autre source de confusion qui peut se manifester, par exemple, lorsqu'on est plus minutieux dans la prise de mesure de certains traitements. Cette dernière source de confusion peut être minimisée en assignant aléatoirement les traitements aux unités expérimentales, en randomisant la séquence de prise de mesure, ou en prenant les mesures en aveugle (p. ex., le traitement est inconnu de la personne prenant la mesure).

La **variabilité de l'expérimentateur** (erreur aléatoire) est une source de confusion associée naturellement à la prise de mesure en tant que telle. On peut réduire cette source de confusion en augmentant le nombre de répétitions. La **variabilité des unités expérimentales** est une autre source de confusion découlant du fait que chaque unité expérimentale a des caractéristiques qui lui sont propres et que cette variabilité reflète la variabilité à l'intérieur de la population. La répétition des traitements, une bonne répartition spatiale des unités expérimentales, ainsi que la récolte de variables secondaires (p. ex., pH, humidité) pour décrire plus directement les caractéristiques de chaque unité expérimentale peuvent prévenir les problèmes associés à cette source de confusion.

Certaines sources de confusion sont difficiles ou impossibles à contrôler, notamment l'effet d'**événements aléatoires** sur l'expérience. Par exemple, lors d'une expérience sur le terrain, le passage d'un ouragan sur les sites d'études peut modifier la variable réponse que l'on mesurera plus tard. Il en va de même avec les pannes électriques qui mettent hors service les systèmes de filtration ou d'éclairage dans des chambres de croissance ou des aquariums. Si l'événement influence toutes les unités expérimentales, le problème est moins grave. Toutefois, si seulement quelques unités expérimentales sont touchées, les observations provenant des unités affectées seront difficiles à interpréter et à comparer au reste des observations : à la limite, certaines données devront être écartées de l'analyse. Le meilleur moyen de réduire

cette source de confusion est d'augmenter le nombre de répétitions et de maintenir une bonne répartition spatiale des unités expérimentales.

Malgré toutes ces précautions, certaines sources de confusion demeurent problématiques, comme le cas d'un **agent extérieur qui intervient délibérément** sur les unités expérimentales d'un traitement particulier et épargne celles des autres traitements. Pour illustrer, considérons le cas d'un site d'étude sur lequel on a disposé des enclos d'animaux où l'on estime la survie de petits mammifères selon différentes densités (nombre d'individus/m²). Ce site est fréquenté par un prédateur qui vise spécifiquement les enclos d'animaux avec les plus grandes densités. Si cette prédation passe inaperçue, on arrivera à des conclusions erronées. Un mauvais plan d'échantillonnage, comme la présence d'effets confondants ou un gradient qui n'a pas été mesuré, correspond aussi à une source de confusion qui ne permet pas de tester de façon efficace une hypothèse scientifique. Dans de tels cas, le facteur qui agit délibérément sur les unités expérimentales, c'est l'expérimentateur. Il faut donc rester vigilant dans l'élaboration des plans d'échantillonnage.

Conclusion

Dans ce texte, nous avons vu plusieurs éléments importants à considérer lors de l'élaboration de plans d'échantillonnage, certains problèmes incluant la pseudoréplication et des sources de confusion atténuant ou invalidant les conclusions, ainsi que des solutions. La planification du plan d'échantillonnage est la partie la plus importante de votre projet et requiert une bonne réflexion. Cette étape détermine si l'analyse sera simple ou plus complexe. Un bon dispositif permet de tester l'hypothèse scientifique simplement et facilite énormément l'analyse. On peut pallier certains problèmes de dispositif expérimental avec des analyses sophistiquées, mais, dans certains cas, l'analyse des données s'avère impossible. Avec une analyse inappropriée réalisée à partir d'un plan d'échantillonnage solide, on peut toujours réanalyser les données, mais avec un mauvais dispositif, il faut refaire l'expérience.

Index

- contrôle, [3](#)
- dispositifs expérimentaux, [2](#)
- étude d'observation, [2](#)
- expérience, [2](#)
- placébo, [3](#)
- plans d'échantillonnage, [2](#)
 - problèmes, [9](#)
- plans d'expérience, [2](#)
 - conseils, [7](#)
- prise de données à double insu, [4](#)
- prise de données à l'insu, [4](#)
- prise de données en aveugle, [4](#)
- prise de données en double aveugle, [4](#)
- pseudoréplication, [9](#)
- pseudoréplication sacrificielle, [11](#)
- pseudoréplication simple, [10](#)
- pseudoréplication temporelle, [11](#)
- randomisation, [5](#)
- répartition spatiale, [5](#)
- répartition temporelle, [5](#)
- répétition, [5](#)
- sources de confusion, [12](#)
 - agent extérieur intervenant délibérément,
[14](#)
- biais associé à l'expérimentateur, [13](#)
- changements temporels, [12](#)
- effets liés aux procédures, [13](#)
- événements aléatoires, [13](#)
- variabilité de l'expérimentateur, [13](#)
- variabilité des unités expérimentales, [13](#)
- stochasticité, [5](#)
- témoin, [3](#)
- variabilité temporelle, [12](#)