

La régression linéaire multiple dans les essais contrôlés randomisés et les méta-analyses en sciences de l'éducation.

Nathalie ROQUES

Novembre 2023

Table des matières

Résumé	5
Avant propos	7
Objectifs.....	9
Les études primaires.....	9
Les méta-analyses	10
Partie 1. Régression linéaire multiple dans les études primaires	12
1. Etudes expérimentales versus études non-expérimentales	12
2. Le modèle de régression linéaire multiple	13
2.1. Premières notations	14
2.2. Estimations des paramètres β_j	14
2.3. Interprétation des paramètres β_j	16
2.4. Quand les hypothèses ne sont pas satisfaites.....	17
3. Évaluation de la qualité du modèle	17
3.1. Corrélations des variables explicatives.....	18
3.2. Décomposition de la variabilité	18
3.3. Tests d’hypothèses	19
4. Les études expérimentales	20
4.1. Essais contrôlés randomisés au niveau individuel	21
4.1.1. Sans les scores prétests	21
4.1.2. Avec les scores prétests.....	23
4.2. Essais contrôlés randomisés au niveau des clusters	24
4.2.1. Présentation	24
4.2.2. Effets fixes et effets aléatoires	25
4.2.3. Hypothèses	25
4.2.4. Le coefficient de corrélation intra classe.....	26
4.2.5. Les modèles linéaires.....	26
Partie 2. Influence des modérateurs sur les tailles d’effet.....	29
1. Les modèles	29
2. Estimations des tailles d’effets indépendantes.....	30
2.1 Analyse de la variance (ANOVA) ou analyse de sous-groupes	30
2.2 Méta-régression	32
3. Estimations des tailles d’effets non indépendantes.....	34
4. Exemples de méta régressions	35
4.1 Synthèse sur l’enseignement des mathématiques et de la lecture	35

4.2. Synthèse sur l'enseignement des fractions.....	37
Partie 3. Discussion.....	39
1. Modérateurs ou médiateurs ?.....	39
2. Analyse confirmatoire ou analyse exploratoire ?.....	40
3. Pour conclure.....	42
Références.....	43
Annexes	44
Annexe 1. ANOVA, ANCOVA, MANOVA	44
Annexe 2. Exemple des loyers : ANOVA et régression linéaire multiple.....	45
Annexe 3. Évaluation d'une intervention (étude primaire)	49
Annexe 4. Étude de l'intervention 1stClass@Number	52
Annexe 5. Utilisation d'un tableur.....	55
Annexe 6. Estimation des variances pour les études avec randomisation au niveau des clusters.....	57
Annexe 7. Glossaire	58

Résumé

Les outils statistiques mettant en œuvre des régressions linéaires multiples dans le but d'analyser les résultats d'études quantitatives dans le domaine des Sciences de l'éducation sont expliqués en différenciant deux niveaux d'études.

Dans un premier temps, les méthodes utilisées pour décrire les données issues d'études primaires par comparaison de groupes sont détaillées. Les calculs permettant d'estimer les coefficients de régression et les tests statistiques évaluant la qualité des modèles linéaires choisis, concernent plus particulièrement les essais contrôlés randomisés (ECR) avec affectation aléatoire dans les groupes traitement et contrôle d'une part, et les essais contrôlés randomisés au niveau des clusters (groupes de sujets) d'autre part. L'interprétation des coefficients de régression estimés quand certaines variables dépendantes sont catégorielles est illustrée par des exemples numériques.

Le second niveau concerne les synthèses quantitatives d'études primaires, et plus précisément les méta-analyse. Les procédures utilisées pour analyser l'influence de variables modératrices sur les tailles d'effet sont présentées en s'appuyant sur des exemples issus de la littérature scientifiques.

Un dernier chapitre propose quelques commentaires sur la nature des variables explicatives (médiatrices ou modératrices) et sur le type d'analyses (exploratoires ou confirmatoires).

Avant propos

Ce texte a été écrit 2 ans après *Mesurer l'effet d'un traitement. Méta-analyse et enseignement*^a dont l'objectif principal était de rendre les méta-analyses du What Works Clearinghouse (WWC) compréhensibles par le plus grand nombre. Les méta-analyses du WWC ne font pas appel aux techniques de la régression multiple ni à l'analyse de sous-groupes^b décrites dans les pages qui suivent. Ce premier texte s'adressait également à un large public et des personnes ayant un niveau de bachelier scientifique étaient en mesure d'en comprendre les mathématiques. L'objectif et le public ciblés par le texte qui suit sont différents. En effet, il s'agit ici d'initier aux procédures statistiques permettant dans un premier temps d'analyser l'influence de caractéristiques sur les apprentissages des élèves ayant subi une intervention, pour dans un second temps étudier l'influence de modérateurs sur les tailles d'effet dans le cadre d'une méta-analyse. Ces procédures font appel à du calcul matriciel et requièrent de ce fait un niveau en mathématiques plus élevé que celui acquis en fin de secondaire pour être pleinement comprises.

Un glossaire (annexe 7) présente les définitions des mots écrits en gras à leur première apparition dans le texte.

^a ROQUES, 2021. Les notions développées dans ce premier document ne sont pas reprises ici.

^b Cette analyse des sous-groupes a tout de même été en partie traitée dans *Mesurer l'effet d'un traitement. Méta-analyse et enseignement* (p.102) ; elle est présentée ici comme un cas particulier de la méta-régression.

Objectifs

De nombreuses études quantitatives en Sciences de l'éducation ont comme premier objectif d'évaluer l'effet d'une **intervention** (ou encore traitement) sur les apprentissages des élèves. Il s'agit de mesurer l'effet de telle méthode pédagogique ou de telle application numérique sur l'acquisition de compétences ou connaissances préalablement identifiées et d'en informer les enseignants ou les responsables en charge de l'organisation scolaire. Mais d'autres questions se posent parfois, qui concernent les influences que certaines **caractéristiques** (comme l'âge des élèves ou leur niveau socio-économique ou la durée de l'intervention) peuvent avoir sur cet effet. L'objectif de ce texte est de présenter certains modèles statistiques permettant de répondre à ce deuxième type d'interrogation.

On distinguera deux ensembles d'études qui correspondent à deux niveaux d'analyse. Le premier ensemble est constitué d'**études par comparaison** de deux groupes de sujets (les élèves) : un groupe intervention qui a subi une intervention, et un groupe contrôle qui ne l'a pas subi. Quand l'affectation des élèves dans l'un des deux groupes est aléatoire, l'étude est un **essai contrôlé randomisé** (ECR) ; quand ce n'est pas le cas, on définit l'étude comme étant quasi-expérimentale. Une fois que plusieurs **études primaires** (ce sont les études précédemment évoquées) ont été réalisées pour une même intervention, les scientifiques conduisent des synthèses quantitatives qui constituent le second ensemble, les **études secondaires**. Pour comparer et agréger les résultats des études primaires, la première étape consiste à calculer une **taille d'effet** pour chacune d'entre elles. Ces tailles d'effet peuvent ensuite être agrégées dans le but d'estimer l'effet global d'une intervention ; les synthèses qui procèdent de cette manière sont des **méta-analyses**. Mais là aussi d'autres objectifs seront poursuivis pour affiner les conclusions en évaluant l'influence de certaines caractéristiques sur les tailles d'effet.

D'une façon générale et pour ces deux niveaux d'analyse, il ne s'agira donc plus seulement de déterminer si une intervention est efficace (ou non), mais de préciser la portée de son effet en fonction de caractéristiques qui intéressent les chercheurs et les praticiens.

Les études primaires

Les caractéristiques qui influent sur l'effet d'une intervention sont des **modérateurs**. Par définition, ce type de caractéristiques exercent une influence sur l'intensité de la relation entre l'intervention et l'évolution des performances des élèves mais ne sont pas intégrés à la chaîne causale qui relie l'intervention aux performances (voir la discussion dans la partie 3). Elles peuvent être **quantitatives** (ce sont par exemple des mesures de scores) ou **qualitatives**^a (et définies par des catégories). Pour analyser statistiquement leur influence (il s'agira parfois seulement d'en tenir compte), on leur associe des **variables**. Ces variables sont **continues** pour les caractéristiques quantitatives, ou **catégorielles** pour les caractéristiques qualitatives. Elles sont incluses dans un modèle mathématique et s'ajoutent à la variable catégorielle associée à la participation (ou non) de l'élève à l'intervention. Toutes ces variables sont qualifiées d'indépendantes, la mesure des performances de l'élève étant quant à elle la variable dépendante (sous-entendu des autres variables). Deux types d'analyses peuvent être alors menées : des analyses de la variance ou ANOVA (annexe 1) quand les variables indépendantes sont catégorielles, ou des régressions linéaires multiples quand les variables indépendantes sont continues et/ou catégorielles.

^a On parle aussi de **facteurs**

Les méta-analyses

À ce niveau, l'effet de l'intervention sur les élèves a déjà été évalué par plusieurs études primaires et des tailles d'effet ont été calculées pour chacune d'entre elles. La question porte maintenant sur l'influence que certaines caractéristiques peuvent avoir sur ces tailles d'effet. Dans le domaine des sciences de l'éducation, ces caractéristiques concernent souvent la structure des études (comme leur design), l'intervention étudiée (comme sa durée) ou les populations analysées (comme le niveau d'étude des élèves). Par exemple, la question peut être de savoir si une intervention est plus (ou moins) efficace à l'école primaire ou au collège. Ou bien si les tailles d'effet estimées par les essais contrôlés randomisés sont inférieures (ou supérieures) aux tailles d'effet estimées par les études quasi-expérimentales. Ici aussi ces caractéristiques sont des modérateurs car elles ont une influence sur l'intensité de la relation entre l'intervention et les performances des élèves et donc la taille d'effet calculée. Elles seront également traduites en autant de variables (tableau 1), et pour évaluer l'influence de ces dernières sur les tailles d'effet, les statisticiens se sont inspirés des analyses statistiques employées initialement dans le cadre des études primaires. On retrouve ici aussi des analyses de la variance quand les variables indépendantes sont catégorielles : ce sont des analyses de sous-groupes ; et des régressions linéaires multiples quand les variables indépendantes sont continues et/ou catégorielles : ce sont des méta-régression^a. Deux différences majeures peuvent d'ores et déjà être soulignées. Tout d'abord, les caractéristiques sont maintenant observées au niveau des études et non plus au niveau des élèves. En second lieu, la variable dépendante analysée dans une méta-analyse est une statistique (la taille d'effet) et non une mesure (le score d'un élève).

La première partie de ce texte sera consacrée aux régressions linéaires multiples conduites dans le cadre d'une étude primaire. La mise en œuvre de ce type d'analyse statistique dans une méta-analyse sera l'objet de la deuxième partie qui reprendra également quelques éléments concernant les analyses de la variance (ANOVA) ou analyses de sous-groupes. Une vue d'ensemble de ces différents types d'analyse et du vocabulaire utilisé est proposée dans le tableau 2 ci-dessous.

Tableau 1. Caractéristiques et variables associées

Type de caractéristiques	Type de variable associée	Exemples
Quantitative	Continue	Étude primaire. Le niveau en mathématiques d'un élève (caractéristique) est associé au score obtenu après passation d'un test (variable).
		Étude secondaire. L'effet d'une intervention (caractéristique) est associé à la taille d'effet (variable).
Qualitative	Catégorielle	Étude primaire. Le niveau socio-économique d'un élève (caractéristique) est associé à une variable catégorielle qui indique son éligibilité ou non aux repas gratuits ^b .
		Étude secondaire. La durée de l'intervention de chaque étude (caractéristique) est associée à une variable catégorielle.

^a Ce mot a été utilisé pour la première fois dans Greenland, S. (1987). Quantitative methods in the review of epidemiologic literature. *Epidemiologic reviews*, 9(1), 1-30. (d'après Tipton, 2019, p.166).

^b Voir les exemples proposés en annexe 4 à partir des études d'Education Endowment Foundation (EEF)

Tableau 2. Les différents types d'analyses statistiques

	Variables indépendantes catégorielles	Variables indépendantes continues et/ou catégorielles
Étude primaire (étude par comparaison de groupes)	Analyse de la variance ou ANOVA	Régression linéaire multiple
Étude secondaire (méta-analyse)	Analyse de sous – groupes	Méta-régression

Partie 1. Régression linéaire multiple dans les études primaires

La régression linéaire multiple est une méthode d'analyse statistique qui a comme objectif de modéliser une liaison entre une variable continue dépendante à expliquer, dénommée encore endogène, et plusieurs variables indépendantes explicatives, ou exogènes^a qui peuvent être continues ou catégorielles. La régression linéaire multiple est une généralisation du cas où une seule variable explicative est intégrée au modèle, et donc de la régression linéaire « tout court » (souvent présentée en première intention aux étudiants). Avec une seule variable explicative catégorielle, on peut également conduire de manière équivalente une analyse de la variance (ANOVA ou *one-way ANOVA*)^b. Si la complexité des calculs mis en œuvre dans une régression linéaire multiple a pu expliquer leur relatif sous-emploi dans les années 70-80 par rapport à l'ANOVA, l'apparition de logiciels performants auraient dû rendre leur utilisation plus fréquente, car leur domaine d'application est plus large. Elles permettent en effet d'analyser l'influence de variables catégorielles (comme l'ANOVA) mais aussi continues (ce que l'ANOVA ne permet pas). C'est donc un outil statistique puissant qu'il est aujourd'hui aisé de mettre en œuvre grâce aux nombreuses solutions logicielles proposées. Il n'en reste pas moins vrai que si les variables sont catégorielles, une ANOVA reste sans doute plus simple à comprendre et ses résultats plus faciles à interpréter.

1. Etudes expérimentales versus études non-expérimentales

Pour présenter les modèles statistiques qui permettent non seulement d'évaluer l'effet d'une intervention mais aussi d'en préciser les contours en fonction de certaines caractéristiques des élèves, nous avons besoin dans un premier temps d'évoquer les études quantitatives non expérimentales. Les études de ce type analysent l'effet de certaines caractéristiques (comme le niveau socio-économique ou l'âge des élèves) sur le niveau de compétence des élèves, mais le chercheur n'en manipule expérimentalement aucune et se contente de les observer et de les mesurer. Dans toutes ces études, expérimentales comme non expérimentales, la variable dépendante est un score obtenu après passation d'un test (score posttest). Les études expérimentales et non expérimentales se distinguent quand on considère les variables indépendantes qui influent sur la variable dépendante.

Dans les études non expérimentales, toutes les variables indépendantes ont en général le même statut et elles intéressent le chercheur de la même façon. Dans les études expérimentales, la variable traitement est la variable d'intérêt ; elle traduit le fait qu'un élève ait subi ou non le traitement et la question de recherche porte en premier lieu sur son influence. Les autres variables sont des variables **modératrices** ou encore des **covariables** ; il s'agit pour le chercheur d'en tenir compte dans l'analyse des données en produisant des résultats présentés parfois comme « ajustés aux covariables ». La seule relation causale qui puisse être démontrée est celle qui lie le traitement (la cause) à l'évolution du niveau de compétence de l'élève (l'effet) ; on parle ici d'une analyse confirmatoire car il s'agit de confirmer (ou non) une hypothèse. Toutes les autres influences sur la variable dépendante, que ce soit les influences de l'ensemble des variables indépendantes dans les études non expérimentales ou celles des variables indépendantes modératrices (ou des covariables) dans les études expérimentales, ne peuvent se prévaloir de ce statut. L'analyse de ces

^a ou prédicteurs ou régresseurs.

^b Avec plusieurs variables explicatives, on pourra utiliser une ANOVA à plusieurs facteurs, ou encore une *two-way ANOVA*. Voir l'annexe 1 pour une définition de ces acronymes.

influences est qualifiée d'exploratoire, car il s'agit d'explorer des interactions, et donc d'émettre des hypothèses qui devront par la suite être confirmées^a.

En ce qui concerne les études expérimentales qui nous intéressent plus particulièrement, on s'intéressera à deux catégories d'entre elles. La première catégorie rassemble les essais contrôlés randomisés (ECR) avec affectation et analyse au niveau individuel : les élèves sont affectés aléatoirement individuellement au groupe traitement ou au groupe contrôle, les mesures et les analyses se font également au niveau individuel. La seconde concerne encore des ECR, mais cette fois l'affectation aléatoire se fait au niveau de **clusters** d'élèves (par exemple des classes ou des établissements) ; les mesures et analyses quant à elles se font toujours au niveau individuel. Les méthodes analytiques présentées concerneront d'abord les études de la première catégorie, plus simples à comprendre, pour s'appuyer sur cette première description et proposer des méthodes adaptées aux études de la seconde catégorie. Les ECR avec affectation individuelle des élèves sont aujourd'hui délaissés au profit d'études de plus grande ampleur dont la validité interne comme externe est de ce fait augmentée. On peut citer ici les études réalisées au Royaume-Uni par Education Endowment Foundation^b (EEF) où des établissements entiers sont aléatoirement affectés à l'un des deux groupes (Roques, à paraître).

2. Le modèle de régression linéaire multiple

Pour faciliter la compréhension des méthodes présentées ci-dessous, des exemples numériques sont proposés en annexe qui présentent plusieurs modèles linéaires. En annexe 2, on compare l'analyse de la variance (ANOVA) des données avec l'analyse utilisant trois modèles de régression linéaire ; en annexe 3, les données (scores posttests et scores prétests) d'une étude par comparaison de groupes fictive sont analysées à l'aide de 6 modèles linéaires ; en annexe 4, les résultats de l'analyse publiés par Education Endowment Foundation concernant l'intervention *FirstClass@Numbers* (EEF, 2018) sont présentés. Pour chacune de ces annexes, un fichier Excel peut être téléchargé dont les méthodes de calculs employés sont détaillées en annexe 5.

D'une façon générale, deux objectifs sont poursuivis : le premier consiste à évaluer l'influence des caractéristiques sélectionnées, pour par exemple les comparer ou identifier la plus influente ; le second concerne la qualité du modèle qui devra être estimée. Quelques éléments repris ici peuvent également être retrouvés dans *Modélisation statistique appliquée aux sciences sociales* (Bressoux, 2010).

On va définir un modèle linéaire dans lequel les variations d'une variable dépendante continue (le score des élèves) dépendent des variations de variables explicatives, qui peuvent être continues ou catégorielles. Ce modèle est linéaire par rapport aux paramètres de la régression (ce sont les coefficients de régression, voir plus loin), et non par rapport aux variables représentant les caractéristiques qui peuvent donc être élevées à une puissance par exemple. Toutes les variables explicatives et à expliquer sont observées à partir d'un échantillon, et on supposera que l'on a plus d'observations que de variables. On dit qu'on fait une régression de la variable dépendante à expliquer sur les variables indépendantes explicatives.

^a En statistiques, démontrer une relation de causalité entre la caractéristique décrite par la variable indépendante et la caractéristique décrite par la variable dépendante requiert l'utilisation d'une expérience qui répartit la première caractéristique de façon aléatoire entre les sujets. Si une intervention peut être aléatoirement répartie entre les élèves, il n'en va pas de même pour leur niveau socioéconomique par exemple.

^b <https://educationendowmentfoundation.org.uk/>

2.1. Premières notations

On note p le nombre de variables explicatives et n la taille de l'échantillon. On a donc en tout $p + 1$ variables et d'après ce qui vient d'être dit on a également $n > p + 1$.

D'une manière générale, on pose Y la variable à expliquer dépendante (le score des élèves), qui sera considérée comme une variable aléatoire (on souhaite inférer sur une population), et $X_1, \dots, X_j, \dots, X_p$ les p variables explicatives indépendantes. Les variables explicatives ne sont pas aléatoires : elles peuvent être contrôlées ou bien leur ensemble de définition est connu et aucune inférence n'est attendue. On dit alors que les caractéristiques auxquelles elles sont associées ont chacune un effet fixe. On verra plus loin certains modèles qui tiennent compte de caractéristiques dont les effets sont aléatoires. Les caractéristiques qualitatives définies par plusieurs catégories devront être préalablement traduites en variables explicatives catégorielles.

Chaque valeur observée de la variable dépendante sera considérée comme la première réalisation d'une des n variables aléatoires « prendre au hasard un élément de l'échantillon, noter la mesure qui intéresse le chercheur, et replacer l'élément dans la population ». On va utiliser dans le modèle linéaire les notations suivantes :

- i est le rang de l'observation, j est le rang de la variable explicative avec $1 \leq i \leq n$ et $1 \leq j \leq p$.
- $Y_1 \dots Y_i \dots Y_n$ sont les n variables dépendantes aléatoires et observées.
- $x_{1,1}, x_{2,1}, \dots, x_{n,1} \dots x_{i,j} \dots x_{1,p}, \dots, x_{n,p}$ sont les $n \times p$ valeurs observées des variables explicatives $X_1, \dots, X_j, \dots, X_p$.

L'un des modèles linéaires qui peut être choisi est un modèle avec constante qui peut s'écrire pour chaque observation i

$$Y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_p x_{i,p} + \varepsilon_i$$

Avec

- $\beta_0 \dots \beta_j \dots \beta_p$ les $p + 1$ paramètres ou coefficients de régression du modèle qui vont être estimés.
- $\varepsilon_1 \dots \varepsilon_i \dots \varepsilon_n$ les n résidus ou erreurs du modèle, variables aléatoires non observées.

Déterminer le modèle c'est calculer les estimations des paramètres β_j (définis au niveau de la population entière, ils sont inconnus et le resteront). Ces estimations seront calculées à partir des données de l'échantillon.

2.2. Estimations des paramètres β_j

On note :

- y_i la valeur observée de la variable Y_i
- b_j l'estimation du paramètre β_j
- e_i l'estimation de la variable ε_i

Avec les valeurs observées et les estimations des paramètres, on a alors

$$\begin{cases} y_1 = b_0 + b_1x_{1,1} + b_2x_{1,2} + \dots + b_px_{1,p} + e_1 \\ \dots \\ y_i = b_0 + b_1x_{i,1} + b_2x_{i,2} + \dots + b_px_{i,p} + e_i \\ \dots \\ y_n = b_0 + b_1x_{n,1} + b_2x_{n,2} + \dots + b_px_{n,p} + e_n \end{cases}$$

Ces relations sont plus succinctement décrites en utilisant les notations matricielles suivantes

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \dots \\ y_n \end{pmatrix} \quad \mathbf{x} = \begin{pmatrix} 1 & x_{1,1} & \dots & x_{1,p} \\ \vdots & & \ddots & \vdots \\ 1 & x_{n,1} & \dots & x_{n,p} \end{pmatrix} \quad \mathbf{x}_j = \begin{pmatrix} x_{1,j} \\ \dots \\ x_{n,j} \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} b_0 \\ \dots \\ b_p \end{pmatrix} \quad \mathbf{e} = \begin{pmatrix} e_1 \\ \dots \\ e_n \end{pmatrix}$$

\mathbf{y} , \mathbf{b} et \mathbf{e} sont des vecteurs^a. La matrice \mathbf{x} est la matrice explicative (*design matrix*) de dimension $n \times (p + 1)$.

Le modèle de régression linéaire multiple s'écrit alors

$$\mathbf{y} = \mathbf{x} * \mathbf{b} + \mathbf{e}$$

Définir le modèle, c'est donc calculer la matrice \mathbf{b} .

On montre que

$$\mathbf{b} = (\mathbf{t}_x * \mathbf{x})^{-1} * \mathbf{t}_x * \mathbf{y}$$

Avec \mathbf{t}_x la matrice transposée de la matrice \mathbf{x} (parfois notée \mathbf{x}') et $(\mathbf{t}_x * \mathbf{x})^{-1}$ la matrice inverse de $\mathbf{t}_x * \mathbf{x}$. Cette dernière doit donc être inversible^b.

Des hypothèses sont formulées qui permettent (1) l'application du modèle de régression choisi et (2) la réalisation des tests d'hypothèses sur les paramètres estimés. En particulier :

- (1) la matrice $\mathbf{t}_x * \mathbf{x}$ doit être inversible, ce qui impose ici que les vecteurs \mathbf{x}_j et le vecteur unité ne soient pas colinéaires, c'est-à-dire que chaque variable apporte une information nouvelle par rapport aux autres. Cela ne signifie pas que les variables explicatives X_j soient statistiquement indépendantes : elles peuvent être par exemple corrélées et leurs covariances peuvent être différentes de zéro.
- (1) les erreurs ε_i sont indépendantes entre elles et indépendantes des variables explicatives, leurs espérances sont nulles et elles ont même variance σ^2 . Cette variance sera estimée par s^2 ; cela revient à dire que les variables Y_i sont indépendantes entre elles et qu'elles ont même variance σ^2 .
- (2) les erreurs ε_i sont normalement distribuées ou encore $\varepsilon_i \sim \mathcal{N}(0, \sigma)$. Cela revient à dire que les variables Y_i sont normalement distribuées autour de leur espérance μ_i avec même variance σ^2 ; $Y_i \sim \mathcal{N}(\mu_i, \sigma)$ et $\mu_i = \beta_0 + \beta_1x_{i,1} + \beta_2x_{i,2} + \dots + \beta_px_{i,p}$

Un modèle sans constante aurait pu être proposé, on aurait alors eu

$$Y_i = \beta_1x_{i,1} + \beta_2x_{i,2} + \dots + \beta_px_{i,p} + \varepsilon_i$$

Les conditions d'applications restent les mêmes, mais l'interprétation des coefficients de régression n'est plus la même. On verra également que certaines méta-régressions utilisent des modèles sans

^a Vecteurs et matrices sont notés en **gras**.

^b Un exemple de calcul matriciel est donné dans le fichier excel associé à l'annexe 2.

constante mais intègre dans la matrice explicative une première colonne de 1, ce qui revient à définir un modèle avec une constante et seulement $p - 1$ variables.

Les estimations des paramètres β_j peuvent être calculées à l'aide d'un tableur (voir annexe 5) et la significativité statistique qu'on peut leur associer sera estimée à l'aide de tests statistiques (voir plus loin).

2.3. Interprétation des paramètres β_j

L'interprétation des paramètres β_j dépend du modèle choisi mais aussi des variables (notamment de l'unité de la variable dépendante et du traitement des variables explicatives qui peuvent par exemple être centrées pour les variables continues). Il est donc difficile de présenter un cadre général permettant d'interpréter ces paramètres dans toutes les situations possibles, et comme on l'a déjà signalé, l'ANOVA reste une option intéressante quand aucune variable explicative n'est continue.

D'une façon générale, on dira que le paramètre β_j représente l'impact ou l'effet de la variable explicative j sur la variable dépendante quand les autres variables explicatives sont maintenues constantes (« toutes choses égales par ailleurs »). On dit aussi que β_j caractérise l'association entre la variable à expliquer et la variable explicative j avec ajustement aux autres variables, notamment quand ces dernières sont des covariables. On admet par la suite qu'une estimation pour chaque paramètre β_j a été calculée.

Les propriétés des fonctions linéaires vont permettre d'interpréter les coefficients en distinguant le cas de caractéristiques décrites par une variable continue (par exemple des scores prétests) et le cas de facteurs définis chacun par plusieurs catégories (par exemple les niveaux d'étude des élèves).

Variable explicative continue

Si la variable explicative X_j est continue, alors le coefficient de régression b_j est l'accroissement calculé selon le modèle de la variable dépendante quand la variable explicative concernée augmente d'une unité et que les autres variables sont maintenues constantes dans le modèle linéaire considéré. On dit aussi que c'est l'effet marginal de la variable concernée. Pour des valeurs des variables explicatives données, on peut écrire de façon simplifiée que la variable dépendante calculée d'après le modèle est égale à $C + b_j * x_j$. On a tout simplement isolé ici le terme $b_j * x_j$ qui nous intéresse plus particulièrement, C représentant le reste de l'équation. Si toutes les variables explicatives restent constantes à l'exception de la variable x_j qui augmente d'une unité, alors la variable dépendante devient $C + b_j * (x_j + 1) = C + b_j * x_j + b_j$. Elle a donc bien augmenté de b_j .

Variables catégorielles ou muettes

Pour chaque facteur qui sont des caractéristiques qualitatives définies par plusieurs catégories, des variables catégorielle ou **muettes** devront être définies et intégrées au modèle linéaire. Supposons qu'un facteur soit défini par k catégories mutuellement exclusives (chaque observation est associée à une catégorie et une seule). On définit dans un premier temps k variables en leur affectant par exemple la valeur 1 si l'observation est associée à la catégorie concernée et la valeur 0 dans le cas contraire (d'autres codages sont possibles comme nous le verrons au chapitre 4.1.1). Ces variables sont dites muettes (*dummy variable*) car les valeurs qui leurs sont attribuées ne représentent pas

des quantités. On dit aussi que ce sont des variables **indicatrices**. Avec le codage défini ci-dessus, pour chaque observation la somme des k variables indicatrices est égale à 1. Et donc, dans les modèles avec constante, pour satisfaire à la condition de non colinéarité des vecteurs de la matrice explicative, l'une de ces variables est exclue du modèle. La catégorie associée à la variable exclue est la catégorie de référence. Reprenons le modèle linéaire simplifié et supposons que la valeur dépendante soit calculée pour un sujet appartenant à la catégorie de référence. Les $k - 1$ variables muettes sont alors toutes égales à zéro. Pour la variable muette x_j particulière, la variable dépendante sera égale à $D + b_j * 0 = D$. Si toutes les variables explicatives restent constantes à l'exception de la variable muette x_j qui prend la valeur 1 (car cette fois l'observation est associée à la catégorie j) alors la variable dépendante devient $D + b_j * 1 = D + b_j$; elle aura augmenté de b_j . Le coefficient de régression b_j est donc la variation de la variable dépendante quand la catégorie concernée est observée, par rapport à la catégorie de référence (voir les annexes 2 et 3 pour des exemples).

On notera que les deux types de variables (continues ou muettes) peuvent être utilisées dans un même modèle, l'interprétation des coefficients n'en sera pas modifiée.

β_0 est l'intercept^a : si on se limite à considérer le modèle linéaire mathématique proposé, c'est la valeur prise par la valeur dépendante quand toutes les variables explicatives sont nulles. Son interprétation quant à elle dépend du contexte et du modèle (notamment de la façon de coder les facteurs). D'une façon générale, on ne peut extrapoler au-delà des limites des domaines d'observation des variables explicatives car ces variables ont des effets fixes. Des modèles avec variables à effets aléatoires permettent d'inférer au-delà des modalités observées, mais ne sont pas étudiées dans ce chapitre.

Dans certains cas, l'interprétation des résultats se fondera uniquement sur la signification statistique que l'on pourra associer à l'estimation des paramètres de régression.

2.4. Quand les hypothèses ne sont pas satisfaites

Le modèle que nous venons de décrire suppose entre autres que les erreurs ε_i sont indépendantes, ce qui revient à dire que les scores sont indépendants entre eux. Cette condition n'est parfois pas vérifiée, notamment dans le cas où les données sont issues de clusters (c'est-à-dire de regroupements d'élèves) qui influent sur la caractéristique dépendante. Par exemple quand les élèves sont regroupés dans plusieurs classes, les élèves d'une même classe ont des scores plus proches que des élèves de classes différentes. On verra plus loin comment analyser certaines études expérimentales avec des élèves regroupés dans des clusters.

Enfin, nous avons implicitement considéré que l'effet d'une variable ne dépendait pas des autres variables. Dans certains cas, cette hypothèse n'est pas vérifiée et il faudra tenir compte d'une interaction entre deux variables (voir l'exemple donné en annexe 4).

3. Évaluation de la qualité du modèle

Un modèle linéaire a été choisi : des variables ont été sélectionnées (et modélisées en ce qui concerne les variables catégorielles), une équation linéaire définie et les coefficients de régression

^a Dans le cas d'une régression avec une seule variable explicative, c'est l'ordonnée à l'origine de la droite de régression. Ce mot anglais a été utilisé dans ce texte car d'usage courant en statistiques.

estimés. Il convient maintenant d'évaluer la qualité du modèle pour déterminer dans quelle mesure il rend compte des observations expérimentales.

3.1. Corrélations des variables explicatives

Certains chercheurs incluent ici dans leur analyse une étude des liens entre les différentes variables explicatives. Le fait que deux variables soient plus ou moins fortement corrélées (liées ou encore dépendantes) peut en effet apporter des informations sur la qualité du modèle linéaire choisi.

Soit \bar{x}_k la moyenne des n valeurs observées de la variable X_k . On calcule $r_{k,j}$, les coefficients de corrélation entre toutes les paires de variables X_j et X_k

$$r_{k,j} = \frac{\sum_{i=1}^n (x_{k,i} - \bar{x}_k) (x_{j,i} - \bar{x}_j)}{\sqrt{\sum_{i=1}^n (x_{k,i} - \bar{x}_k)^2 \sum_{i=1}^n (x_{j,i} - \bar{x}_j)^2}} = \frac{cov(X_k, X_j)}{\sqrt{var X_k \times var X_j}}$$

On a $r_{j,k} = r_{k,j}$ et $r_{j,j} = 1$. Si $r_{j,k} = 0$ alors les variables X_j et X_k ne sont pas corrélées (elles sont indépendantes entre elles). Plus le coefficient est important, plus les variables associées sont fortement corrélées et moins bien les paramètres associés seront estimés.

3.2. Décomposition de la variabilité

Soit \hat{y}_i les n valeurs prédites de la variable dépendante, calculées en utilisant le modèle linéaire et les estimations des coefficients de régression. On a donc :

$$\hat{y}_i = b_0 + b_1 x_{i,1} + b_2 x_{i,2} + \dots + b_p x_{i,p} \text{ donc } y_i = \hat{y}_i + e_i \text{ et } y_i - \hat{y}_i = e_i$$

L'écart entre la valeur observée y_i et sa valeur estimée \hat{y}_i est le résidu ou l'erreur et dépend du modèle choisi. On note \bar{y} la moyenne des n valeurs observées y_i et on définit les trois sommes suivantes :

$$SC_{résidus} = \sum (y_i - \hat{y}_i)^2 = \sum e_i^2 = \text{somme des carrés des résidus (ou des erreurs).}$$

$$SC_{totale} = \sum (y_i - \bar{y})^2 = \text{somme des carrés des écarts des valeurs observées à la moyenne des valeurs observées ou encore somme des carrés totale.}$$

$$SC_{régression} = \sum (\hat{y}_i - \bar{y})^2 = \text{somme des carrés des écarts des valeurs prédites à la moyenne des valeurs observées ou somme des carrés dus au modèle ou somme des carrés de la régression.}$$

On montre que l'on a :

$$SC_{totale} = SC_{régression} + SC_{résidus}$$

Cela signifie que la variabilité totale de Y est décomposée en variabilité expliquée par le modèle et en variabilité résiduelle.

On calcule alors s^2 , une estimation de σ^2 (variance des erreurs ε_i) de la façon suivante

$$s^2 = \frac{SC_{résidus}}{n - p - 1} = \frac{\sum e_i^2}{n - p - 1}$$

Le degré de liberté est égal à $n - p - 1$: la perte de $p + 1$ degrés de liberté s'explique par l'utilisation dans le calcul de $SC_{résidus}$ des $p + 1$ estimations des paramètres $\beta_0 \dots \beta_j \dots \beta_p$.

Pour évaluer la proportion de la variabilité de Y expliquée par les variations des variables explicatives selon le modèle choisi, on calcule le coefficient de détermination R^2 (exprimé en pourcentage) de la façon suivante :

$$R^2 = \frac{SC_{régression}}{SC_{totale}} = \frac{SC_{totale} - SC_{résidus}}{SC_{totale}} = 1 - \frac{SC_{résidus}}{SC_{totale}}$$

Plus R^2 est proche de 100%, c'est-à-dire plus $SC_{résidus}$ (ou de manière équivalente s^2) est petit, meilleur est le modèle.

3.3. Tests d'hypothèses

Test de significativité globale de la régression

On va tester l'hypothèse H_0 : « $\beta_0 = \dots = \beta_j = \dots = \beta_p = 0$ » contre l'hypothèse H_1 : « au moins l'un des paramètres β_j est différent de 0 ». L'hypothèse H_0 signifie qu'il n'y a pas de liaison entre aucune des variables explicatives et la variable à expliquer et l'hypothèse H_1 signifie qu'il existe une liaison entre au moins l'une des variables explicative et la variable à expliquer. On utilise la statistique suivante :

$$F(p, n - p - 1) = \frac{\frac{SC_{régression}}{p}}{\frac{SC_{résidus}}{n - p - 1}}$$

Cette statistique est distribuée sous H_0 selon une loi de Fisher à p et $n - p - 1$ degrés de liberté. On conclura en comparant la valeur calculée à une valeur de référence, ou par un calcul de la valeur-p ou d'un écart type et d'un intervalle de confiance, en ayant auparavant fixé un niveau de confiance (généralement 95 %).

Test de significativité individuelle des coefficients β_j

On va tester l'hypothèse H_0 : « $\beta_j = 0$ » contre H_1 : « $\beta_j \neq 0$ ». L'hypothèse H_0 signifie qu'il n'y a pas de liaison entre la variable explicative j et la variable à expliquer et l'hypothèse H_1 signifie qu'il y a une liaison entre la variable explicative j et la variable à expliquer. On note s_{β_j} l'écart-type de β_j et on calcule la statistique suivante^a :

$$t_j = \frac{b_j - 0}{s_{\beta_j}} = \frac{b_j}{s_{\beta_j}}$$

Cette statistique est distribuée sous H_0 selon une loi de Student à $n - p - 1$ degrés de liberté, et on conclura en comparant la valeur calculée à deux valeurs de référence, ou par un calcul de la valeur-p ou par le calcul d'une erreur standard et d'un intervalle de confiance, en ayant auparavant fixé un niveau de confiance (généralement 95 %).

On a donc besoin de calculer les écarts-types s_{β_j} de chaque paramètre β_j , ou leur variances notées $var\beta_j$ ($var\beta_j = s_{\beta_j}^2$). Pour se faire, on calcule la matrice symétrique des variances – covariances de dimension $(p + 1) \times (p + 1)$ des coefficients de la façon suivante :

^a On ne différencie pas ici le paramètre β_j (un nombre réel inconnu) de son estimateur (une variable aléatoire) par souci de simplicité et l'écart-type considéré ici est bien entendu celui de l'estimateur ; de la même façon, on ne distingue pas la statistique observée t_j de la variable aléatoire T_j dont on évoque la distribution.

$$s^2 * (\mathbf{t}_x * \mathbf{x})^{-1} = \begin{pmatrix} var\beta_0 & cov(\beta_0, \beta_1) & \cdots & cov(\beta_0, \beta_p) \\ cov(\beta_1, \beta_0) & var\beta_1 & & \\ \vdots & & \ddots & \vdots \\ \cdots & & \cdots & var\beta_p \end{pmatrix}$$

4. Les études expérimentales

On va s'intéresser ici à deux catégories d'études : les essais contrôlés randomisés avec affectation et analyse au niveau des élèves d'une part et les essais contrôlés randomisés avec affectation au niveau des clusters (soit des classes, soit des établissements) et analyse au niveau des élèves d'autre part. L'objectif est d'évaluer l'effet d'une intervention sur des élèves en tenant compte de certaines caractéristiques. Pour quantifier cet effet, on estime une taille d'effet δ qui est la différence des moyennes des scores posttests des deux populations (traitement et contrôle), standardisée. Si on note μ_T la moyenne des scores posttests de la population traitée et μ_C la moyenne des scores posttests de la population contrôle et σ l'écart-type des scores de la population^a, on a alors

$$\delta = \frac{\mu_T - \mu_C}{\sigma}$$

On a donc $\mu_T = \delta \times \sigma + \mu_C$, ce qui revient à considérer δ comme le nombre d'écarts-types qu'il faut ajouter à la moyenne des scores des élèves de la population contrôle pour obtenir la moyenne des scores de la population traitée.

Pour les deux types de design, la répartition aléatoire assure que seule la caractéristique « condition » (avoir ou non subi l'intervention) différencie les deux groupes d'élèves. Mais l'influence du niveau scolaire des élèves sur leurs apprentissages est telle que, y compris dans les essais contrôlés randomisés, il est recommandé de nos jours de prendre en compte cette caractéristique dans le modèle (WWC, 2022). C'est pour cette raison que les compétences sont mesurées par un test avant l'expérience puis par un test après l'expérience. Chaque élève obtient donc un score prétest et un score posttest.

En résumé, on cherche à analyser l'influence de la participation au traitement ou non sur la variable dépendante score posttest, en tenant compte de la variable indépendante scores prétests. Cette dernière est une variable modératrice qui a le statut de covariable, car c'est bien l'influence du traitement qui intéresse le chercheur. Pour chacun des deux types de design étudiés ci-après, un modèle sans covariable est présenté en première intention, et les scores prétests seront intégrés au modèle dans un second temps.

On notera enfin qu'avec un même jeu de données, plusieurs modèles peuvent être utilisés qui répondent à des questions différentes ou dont les conditions d'applications peuvent différer. Le choix du statisticien se fondera notamment sur l'analyse du design de l'étude, de la nature des données et du type de question de recherche posée.

^a La population traitée est théorique ; les variances des scores de deux populations (traitée et contrôle) sont supposées être identiques.

4.1. Essais contrôlés randomisés au niveau individuel

Un exemple numérique est donné en annexe 3 à partir de scores d'élèves fictifs (le fichier Excel peut également être téléchargé sur le site de l'auteure). On verra que d'autres variables modératrices peuvent être intégrées au modèle dans le chapitre concernant les essais contrôlés randomisés au niveau des clusters.

Les élèves sont répartis de façon aléatoire entre un groupe traitement et un groupe contrôle. Participer au traitement ou ne pas participer au traitement constituent les deux catégories du facteur condition. Pour satisfaire à la condition de non colinéarité des vecteurs de la matrice explicative, on utilise une seule variable indicatrice, la variable T ou encore variable traitement. La catégorie « groupe contrôle » servira alors de référence.

On pose les hypothèses suivantes :

- les scores posttests du groupe traitement sont normalement distribués autour de μ_T la moyenne des scores de la population traitement
- les scores posttests du groupe contrôle sont normalement distribués autour de μ_C la moyenne des scores de la population contrôle
- ces deux distributions ont la même variance σ^2 .

4.1.1. Sans les scores prétests

On définit le modèle linéaire suivant :

$$Y_i = \beta_0 + \beta_1 T_i + \varepsilon_i$$

Avec

- i le rang de l'élève
- T_i la valeur de la variable T pour l'observation i .
- Y_i la variable indépendante aléatoire et observée « score posttest de l'élève i » ; dans le groupe traitement, $Y_i \sim \mathcal{N}(\mu_T, \sigma)$ et dans le groupe contrôle $Y_i \sim \mathcal{N}(\mu_C, \sigma)$
- β_1 est le coefficient de régression de la variable T et représente l'impact ou l'effet du traitement sur le score post-test.
- β_0 est l'intercept et son interprétation dépend du codage de la variable T
- $\varepsilon_1 \dots \varepsilon_i \dots \varepsilon_n$ les n résidus ou erreurs ; $\varepsilon_i \sim \mathcal{N}(0, \sigma)$

Dans les rapports d'EEF la variables indicatrice T prend la valeur 1 si l'élève appartient au groupe traitement, et elle prend la valeur 0 dans le cas contraire ; Hedges quant à lui a choisi de fixer cette valeur à 0,5 si l'élève appartient au groupe traitement, et à $-0,5$ dans le cas contraire (Hedges, 2010 a).

En utilisant les propriétés de l'espérance d'une variable aléatoire, et en notant $E(Y_i)$ l'espérance de la variable Y_i , on a alors :

$T_i = 0$ ou bien $T_i = 1$	$T_i = -0,5$ ou bien $T_i = 0,5$
Dans le groupe traitement $Y_i = \beta_0 + \beta_1 + \varepsilon_i$ $E(Y_i) = \beta_0 + \beta_1 = \mu_T$	Dans le groupe traitement $Y_i = \beta_0 + 0,5\beta_1 + \varepsilon_i$ $E(Y_i) = \beta_0 + 0,5\beta_1 = \mu_T$
Dans le groupe contrôle $Y_i = \beta_0 + \varepsilon_i$ $E(Y_i) = \beta_0 = \mu_C$	Dans le groupe contrôle $Y_i = \beta_0 - 0,5\beta_1 + \varepsilon_i$ $E(Y_i) = \beta_0 - 0,5\beta_1 = \mu_C$
Donc $\beta_0 = \mu_C$ $\beta_1 = \mu_T - \mu_C$	Donc $\beta_0 = \frac{(\mu_T + \mu_C)}{2}$ $\beta_1 = \mu_T - \mu_C$

Si les groupes traitements et contrôles sont de même taille, alors $\beta_0 = \frac{(\mu_T + \mu_C)}{2}$ est la moyenne générale des scores posttests.

En utilisant les données expérimentales et les estimations, on a

$T_i = 0$ ou bien $T_i = 1$	$T_i = -0,5$ ou bien $T_i = 0,5$
Dans le groupe traitement $y_i = b_0 + b_1 + e_i$ $\bar{y}_T = b_0 + b_1 + \bar{e}_i = b_0 + b_1$	Dans le groupe traitement $y_i = b_0 + 0,5b_1 + e_i$ $\bar{y}_T = b_0 + 0,5b_1 + \bar{e}_i = b_0 + 0,5b_1$
Dans le groupe contrôle $y_i = b_0 + e_i$ $\bar{y}_C = b_0 + \bar{e}_i = b_0$	Dans le groupe contrôle $y_i = b_0 - 0,5b_1 + e_i$ $\bar{y}_C = b_0 - 0,5b_1 + \bar{e}_i = b_0 - 0,5b_1$
Donc $b_0 = \bar{y}_C$ $b_1 = \bar{y}_T - \bar{y}_C$	Donc $b_0 = (\bar{y}_C + \bar{y}_T)/2$ $b_1 = \bar{y}_T - \bar{y}_C$

Avec \bar{y}_T la moyenne des scores posttests du groupe traitement, \bar{y}_C la moyenne des scores posttests du groupe contrôles et \bar{e}_i les moyenne des erreurs, qui sont à chaque fois négligeables.

On calcule alors b_1 , l'estimation de β_1 et s , l'estimation de σ .

Si on s'intéresse à la taille d'effet δ , on a alors dans les deux cas

$$\delta = \frac{\mu_T - \mu_C}{\sigma} = \frac{\beta_1}{\sigma}$$

Pour les deux modèles, le d de Cohen qui est une estimation de cette taille d'effet se calcule de la même façon

$$d = \frac{b_1}{s}$$

Les résultats d'un modèle sans constante avec deux variables indicatrices sont également proposés en annexe 2. Les coefficients de régression sont alors les moyennes conditionnelles (moyennes des groupes traitement d'une part et contrôle d'autre part).

Avec un modèle encore plus simple, sans variable traitement (on a alors $Y_i = \beta_0 + \varepsilon_i$), on estime la moyenne des scores posttests. Cette remarque sera utilisée dans la partie s'intéressant aux méta-analyses.

4.1.2. Avec les scores prétests

Les hypothèses restent les mêmes. On ajoute au modèle précédent une variable explicative continue X qui est la variable score prétest. X_i est alors défini comme le score pré-test de l'élève i . Le modèle s'écrit

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + \varepsilon_i$$

On reprend la liste des indications précédemment données avec les modifications suivantes :

- β_1 est la différence des moyennes des scores des deux populations ajustée aux scores prétests
- β_2 est le coefficient de régression de la variable X_i , que l'on interprète comme l'accroissement du score posttest quand le score prétest augmente d'une unité.
- $\varepsilon_i \sim \mathcal{N}(0, \sigma_A)$ avec σ_A^2 la variance intra-cluster ajustée aux scores prétests.

On se limite dorénavant au codage de la variable T utilisé par EEF. En utilisant les données expérimentales et les estimations, on a

Dans le groupe traitement	Dans le groupe contrôle
$y_i = b_0 + b_1 + b_2 X_i + e_i$ $\bar{y}_T = b_0 + b_1 + b_2 \bar{X}_T + \bar{e}_i = b_0 + b_1 + b_2 \bar{X}_T$ en notant \bar{X}_T la moyenne des scores prétests du groupe traitement et \bar{y}_T la moyenne des scores posttests du groupe traitement et \bar{e}_i la moyenne des erreurs, qui est négligeable.	$y_i = b_0 + b_2 X_i + e_i$ $\bar{y}_C = b_0 + b_2 \bar{X}_C + \bar{e}_i = b_0 + b_2 \bar{X}_C$ en notant \bar{X}_C la moyenne des scores prétests du groupe contrôle, \bar{y}_C la moyenne des scores posttests du groupe contrôle et \bar{e}_i la moyenne des erreurs, qui est négligeable.

Donc $b_1 = \bar{y}_T - \bar{y}_C - b_2(\bar{X}_T - \bar{X}_C)$ qui est la différence des moyennes des scores posttests ajustées aux scores prétests^a.

Les paramètres comme les estimations diffèrent du modèle vu au chapitre 4.1.1, mais les notations restent les mêmes, sauf en ce qui concerne l'écart-type des scores posttests de la population. Là aussi le paramètre et son estimation diffèrent de ceux du modèle précédent car ils sont maintenant ajustés aux scores prétests, mais contrairement aux autres paramètres, les notations ont été modifiées. En effet, l'objectif premier du chercheur reste encore et toujours le calcul de la taille d'effet. Et si tenir compte des scores prétests dans le calcul des moyennes (et de leur différence) est souhaitable, l'écart-type des scores de la population doit quant à lui rester non ajusté aux covariables : on souhaite traduire les différences des moyennes en nombres d'écart-types des scores de la population, cet écart-type étant considéré ici comme une unité. Et pour comparer

^a On retrouve la différence des moyennes ajustées calculées dans une ANCOVA où b_2 est le coefficient directeur de la droite de régression intra.

entre elles (ou agréger) des différences de moyennes issues d'études différentes, on se doit de les calculer dans la même unité.

Si les scores prétests sont centrés sur les moyennes de leurs groupes, on retrouve les résultats du modèle sans scores prétests (voir annexe 3).

La taille d'effet au niveau de la population est donc donnée par l'équation

$$\delta = \frac{(\mu_T - \mu_C)_{ajustée}}{\sigma} = \frac{\beta_1}{\sigma}$$

Cela implique donc que s_A , l'estimation de la variance ajustée calculée dans le modèle avec covariable ne permet pas d'estimer δ (Feingold, 2013 ; Hedges, 2010 a ; WWC, 2022). On devra utiliser s , l'estimations obtenue dans le premier modèle, sans covariable (on parle de modèle vide, *empty model*). Une estimation de la taille d'effet ajustée aux scores prétests est donc donnée par le d de Cohen

$$d = \frac{b_1}{s}$$

4.2. Essais contrôlés randomisés au niveau des clusters

Dans les ECR de grande ampleur, par exemple pour les études d'efficacité en conditions réelles conduites par Education Endowment Foundation (EEF) au Royaume-Uni (Roques, à paraître), l'objectif est d'étudier l'effet d'une intervention sur un très grand nombre d'élèves. Dans 70% des ECR conduits par EEF, les tailles d'échantillon sont supérieures à 1000. Les études sont alors conduites sur plusieurs établissements que l'on nommera des clusters. Les élèves regroupés dans des clusters ne sont pas indépendants les uns des autres : les scores obtenus à partir d'un même établissement sont plus semblables entre eux que les scores obtenus à partir d'établissements différents. Le terme groupe sera réservé par la suite au groupe traitement et au groupe contrôle. En annexe 4, les équations d'un modèle de régression linéaire multiple utilisé pour évaluer une intervention en milieu scolaire au Royaume-Uni sont données et certains résultats sont recalculés.

4.2.1. Présentation

Deux types de design tenant compte des clusters sont couramment utilisés en Sciences de l'éducation. Le premier rassemble des essais contrôlés randomisés au niveau des clusters (*cluster randomised controled trial* ou C-RCT) où des clusters intacts d'élèves (par exemple des établissements entiers) sont affectés aléatoirement au groupe traitement ou au groupe contrôle ; on parle également de modèle hiérarchique. Le second réunit des essais contrôlés randomisés multisites (*multi-site randomised controled trials*, MS-RCT, ou encore *randomized-block design*) où les élèves sont affectés aléatoirement au groupe traitement ou contrôle de façon individuelle à l'intérieur de leur cluster (et par conséquent, dans chaque établissement, il y a à la fois des élèves traitement et des élèves contrôles). Pour chacun de ces deux types de design, le niveau d'analyse reste toujours le niveau élève. Nous nous limiterons ici à décrire le modèle statistique utilisé pour analyser les études du premier ensemble, car ce sont ces études qui sont le plus couramment conduites par EEF de nos jours (ce design limitant notamment le risque de contamination entre

élèves de groupes différents). Deux niveaux seront considérés : le niveau élève (niveau 1), et le niveau cluster (niveau 2). Des analyses avec 3 niveaux de regroupements sont également possibles mais ont été laissées de côté ici (Hedges, 2010 a ; EEF, 2018).

4.2.2. Effets fixes et effets aléatoires

Afin d'explicitier les procédures et d'interpréter les résultats, la notion d'effet fixe ou d'effet aléatoire d'une variable (ou de la caractéristique associée) doit être précisée. On a déjà vu qu'une variable qui possède un ensemble de modalités connues qui seules intéressent le chercheur, a ce qu'il est convenu d'appeler un effet fixe sur la variable dépendante. C'est le cas de l'effet de la variable traitement ou de celui de la variable scores prétests : les valeurs que ces variables peuvent prendre sont connues et on ne souhaite pas inférer au-delà. Quand les modalités d'une variable ont été choisies de façon aléatoire dans une population, on dit que cette variable a un effet aléatoire. Les établissements participant à l'expérience sont un échantillon représentatif d'une population plus vaste, et l'objectif est de conduire une analyse inférentielle (Givord & Guillerm, 2016 ; Bressoux, 2007). L'effet « établissement » est donc un effet aléatoire. On en tiendra compte en ajoutant dans le modèle des termes inobservés dont on suppose qu'ils ont une distribution normale d'espérance nulle et de même variance que l'on cherchera à estimer. On dit des modèles qui intègrent à la fois des effets fixes et des effets aléatoires que ce sont des modèles mixtes. C'est ce type de modèle qui a été notamment retenu par EEF (voir également Hedges, 2007 et 2010 a a) car l'objectif des études en Sciences de l'éducation est de pouvoir inférer au-delà des établissements participants à l'étude, ce que ne permet pas le modèle à effets fixes.

Les scores sont toujours mesurés au niveau individuel mais ils ne sont plus indépendants comme nous en avons fait l'hypothèse précédemment, et il en va de même pour les erreurs ε_i telles qu'elles étaient définies dans le modèle précédent.

4.2.3. Hypothèses

Des hypothèses sont posées et concernent les paramètres de la population et plus précisément les distributions autour de ces paramètres. On considère que

- dans chaque cluster, les scores posttests sont normalement distribués autour de la moyenne de leur cluster avec une variance σ_W^2 (variance intra, *within-cluster variance*) commune à tous les clusters.
- dans chaque groupe (traitement ou contrôle), les moyennes des scores posttests de chacun des clusters sont normalement distribuées autour de la moyenne générale des scores posttests du groupe (μ_T pour le groupe traitement ; μ_C pour le groupe contrôle) avec une même variance σ_B^2 (variance inter, *between-cluster variance*).

On définit la variance totale σ_T^2 par

$$\sigma_T^2 = \sigma_W^2 + \sigma_B^2$$

On suppose également qu'il y a le même nombre d'élèves dans chaque cluster et que les groupes traitement et contrôle ont le même nombre de clusters (Hedges, 2010 a).

4.2.4. Le coefficient de corrélation intra classe

Ce coefficient est la proportion de la variance totale qui est due au regroupement des élèves dans un cluster. On note ce coefficient ρ (ou ρ_{ICC} ou ICC). On a

$$\rho = \frac{\sigma_B^2}{\sigma_W^2 + \sigma_B^2}$$

Quand toute la variance vient du niveau élève, $\rho = 0$ et la variance totale est égale à la variance intra (la variance inter est nulle). Les observations dans un établissement ne sont pas plus semblables entre elles que les observations entre des établissements différents. Cela correspond à la version la plus simple de l'essai contrôlé randomisé avec affectation aléatoire et analyse au niveau des élèves.

Quand toute la variance vient du niveau cluster, $\rho = 1$ et la variance totale est égale à la variance inter (la variance intra est nulle). Chaque élève représente à lui seul son cluster (tous les élèves d'un cluster ont le même score, et ce pour tous les clusters).

4.2.5. Les modèles linéaires

Ici aussi on présente d'abord un modèle sans covariable avant d'analyser le modèle avec la covariable score prétest.

Sans les cores prétests

Soit i le rang de l'observation et j le rang du cluster.

On souhaite déterminer Y_{ij} (score posttest de l'élève i dans le cluster j , variable aléatoire observée) comme une fonction de la variable T (variable traitement). On pose

$$\text{Au niveau élève (niveau 1)} \quad Y_{ij} = \beta_{oj} + \varepsilon_{ij}$$

$$\text{Au niveau cluster (niveau 2)} \quad \beta_{oj} = \gamma_{00} + \gamma_{01}T_j + \eta_{0j}$$

Les deux équations précédentes donnent finalement

$$Y_{ij} = \gamma_{00} + \gamma_{01}T_j + \eta_{0j} + \varepsilon_{ij}$$

Avec

- Dans chaque cluster j , Y_{ij} normalement distribué autour de β_{oj} moyenne des scores du cluster, avec une variance commune σ_W^2 , donc $Y_{ij} \sim \mathcal{N}(\beta_{oj}, \sigma_W)$
- T_j variable associée au traitement, prend la valeur 1 quand le cluster appartient au groupe traitement et la valeur 0 dans le cas contraire.
- ε_{ij} les erreurs au niveau élève indépendamment et normalement distribuées avec une moyenne nulle et une variance commune σ_W^2 , la variance intra-cluster ; $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_W)$
- η_{0i} les erreurs au niveau cluster indépendamment et normalement distribuées avec une moyenne nulle et une variance commune σ_B^2 , la variance inter-cluster ; $\eta_{0j} \sim \mathcal{N}(0, \sigma_B)$

- Dans le groupe traitement, $\beta_{oj} \sim \mathcal{N}(\mu_T, \sigma_B)$ et dans le groupe contrôle, $\beta_{oj} \sim \mathcal{N}(\mu_C, \sigma_B)$
- $\gamma_{01} = \mu_T - \mu_C$
- $\gamma_{00} = \mu_C^a$

On calcule alors c_1 , l'estimation de γ_{01} et s_W et s_B , les estimations de σ_W et σ_B . On cherche toujours à estimer δ la taille d'effet au niveau de la population, avec

$$\delta = \frac{\mu_T - \mu_C}{\sigma_T} = \frac{\mu_T - \mu_C}{\sqrt{\sigma_W^2 + \sigma_B^2}}$$

Là aussi le d de Cohen est une estimation de cette taille d'effet

$$d = \frac{c_1}{\sqrt{s_W^2 + s_B^2}}$$

Les équations permettant de calculer s_W^2 et s_B^2 à partir des scores observés sont présentées en annexe 6.

Avec les scores prétests

On se limitera ici au cas d'une seule covariable modélisée au niveau de l'élève, le score prétest. Des modèles plus complexes intègrent plusieurs covariables au niveau élève mais aussi au niveau des clusters (Hedges, 2010 a). A titre d'exemple, un modèle linéaire publié par EEF et permettant de tenir compte du milieu socio-économique ou du genre des élèves est présenté en annexe 4. On ajoute au modèle précédent une variable explicative continue X qui est la variable score prétest. X_{ij} est alors défini comme le score pré-test de l'élève i dans le cluster j . Le modèle s'écrit

Au niveau **élève** (niveau 1) $Y_{ij} = \beta_{oj} + \beta_1 X_{ij} + \varepsilon_{ij}$

Au niveau **cluster** (niveau 2) $\beta_{oj} = \gamma_{00} + \gamma_{01} T_j + \eta_{oj}$

Les deux équations précédentes donnent finalement

$$Y_{ij} = \gamma_{00} + \gamma_{01} T_i + \beta_1 X_{ij} + \eta_{oj} + \varepsilon_{ij}$$

On reprend la liste des indications précédemment données avec les modifications suivantes :

- β_{oj} est la moyenne du cluster j ajustée aux scores prétests,
- γ_{01} est la différence des moyennes des scores des populations traitement et contrôle ajustée aux scores prétests
- β_1 est le coefficient de régression de la variable X_{ij}
- $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_{AW}^2)$ avec σ_{AW}^2 la variance intra-cluster ajustée aux scores prétests.
- $\eta_{oj} \sim \mathcal{N}(0, \sigma_{AB}^2)$ avec σ_{AB}^2 la variance inter-cluster ajustée aux scores prétests.

On notera que si les scores prétests sont centrés sur les moyennes de leurs clusters, on retrouve les résultats du modèle sans scores prétests (voir annexe 3 pour une simulation sans clusters). Hedges

^a Dans le groupe traitement, $T = 1$ et $E(\beta_{oj}) = \gamma_{00} + \gamma_{01} = \mu_T$ et dans le groupe contrôle, $T = 0$ et $E(\beta_{oj}) = \gamma_{00} = \mu_C$

(2010 a) propose de centrer les scores prétests, mais ne poursuit pas jusqu'au calcul d'une taille d'effet (et ne mentionne pas que le coefficient de régression alors obtenu est « ajusté aux scores prétests »). Dans les études EEF, il n'est pas mentionné que les scores prétests sont centrés. Sullivan (1999) propose plusieurs interprétations du coefficient β_{0j} selon que les covariables soient centrées ou non ; dans les modèles proposés par cet auteur, le coefficient β_1 est dépendant du cluster (et est noté β_{1j}).

On calcule alors c_1 , l'estimation de γ_{01} et s_{AW} et s_{AB} , les estimations de σ_{AW} et σ_{AB} .

Ici aussi, les notations sont les mêmes dans les deux modèles (avec ou sans scores prétests), bien que les paramètres soient différents, à l'exception des variances ajustées (la lettre A a été ajoutée pour le signaler). En effet, on cherche toujours à estimer δ la taille d'effet au niveau de la population et, pour les mêmes raisons que celle évoquées dans l'analyse des essais contrôlés randomisés au niveau individuel, on utilise au dénominateur un écart-type non ajusté aux covariables.

La taille d'effet au niveau de la population est donnée par l'équation

$$\delta = \frac{(\mu_T - \mu_C)_{ajustée}}{\sigma_T} = \frac{(\mu_T - \mu_C)_{ajustée}}{\sqrt{\sigma_W^2 + \sigma_B^2}} = \frac{\gamma_{01}}{\sqrt{\sigma_W^2 + \sigma_B^2}}$$

Une estimation de cette taille d'effet est donnée par le d de Cohen

$$d = \frac{c_1}{\sqrt{s_W^2 + s_B^2}}$$

Cela implique donc que les estimations des variances ajustées σ_{AW}^2 et σ_{AB}^2 ne permettent pas d'estimer ce paramètre.

Tests d'hypothèse

Dans les modèles mixtes qui tiennent compte d'effets fixes et d'effets aléatoires, les tests porteront sur les coefficients de régression pour les effets fixes et sur les variances pour les effets aléatoires. La qualité du modèle choisi n'est jamais évoquée dans les publications, ce qui explique que les statistiques R^2 et F dont nous avons parlé aux chapitres 3.2 et 3.3 ne sont jamais publiés (voir les exemples des annexes 3 et 4).

Partie 2. Influence des modérateurs sur les tailles d'effet

On revient maintenant à notre second niveau d'analyse dans le but d'évaluer l'influence sur les tailles d'effet de certains modérateurs sélectionnés par le chercheur dans une méta-analyse.

On va distinguer deux situations différentes. La première où chaque étude estime une seule taille d'effet, et où toutes les estimations des tailles d'effet sont indépendantes. Et la seconde où ces estimations des tailles d'effets ne sont pas indépendantes. Par exemple, certaines études ont calculé plusieurs tailles d'effet (scores des élèves en mathématiques et en lecture ; scores obtenus tout de suite après le traitement puis 3 mois plus tard ; ...). Ou bien des études différentes ont été menées par la même équipe de recherche. Nous allons d'abord analyser la première situation et proposer un modèle sur lequel on s'appuiera pour, dans un second temps, évoquer des situations réelles plus complexes qui correspondent mieux à la réalité du terrain en Sciences de l'éducation. Il s'agira ici de présenter les bases de certaines procédures plutôt que d'en expliciter les détails, car la complexité liée à la fois à la diversité des contextes (tant au niveau des études synthétisées que des questions de recherche) et aux traitements mathématiques experts ne permet pas dans le cadre de ce texte d'aller au-delà. À ce niveau, le recours à une application informatique devient indispensable. Dans les exemples issus de la littérature proposés au chapitre 3, les équations linéaires ne sont pas explicitées et les interprétations des coefficients de régression peuvent de ce fait manquer de clarté (voir la discussion à la fin).

La rédaction de ce chapitre s'est appuyée sur deux livres : *Introduction to meta-analysis* (Borenstein & al., 2009) et *The Handbook of Research Synthesis and Meta-analysis* (Cooper & al., 2019). C'est l'occasion ici de souligner l'importance des travaux de Hedges qui ont contribué de façon majeure au développement des procédures statistiques utilisées de nos jours par les scientifiques ; les articles de ce chercheur restent indubitablement des références en la matière^a.

C'est une version simplifiée des notations utilisées dans le *Handbook* qui sera employée ici.

1. Les modèles

Contrairement aux études primaires, la condition (avoir subi le traitement ou non) n'est plus une caractéristique à intégrer au modèle, son impact ayant déjà été évalué par le calcul d'une taille d'effet qui est ici la variable dépendante. Les variables explicatives que l'on considère sont donc toutes des variables modératrices ou encore des covariables, et les analyses sont donc toutes de type exploratoire.

De la même façon que dans une étude primaire, les modérateurs peuvent avoir des effets fixes ou des effets aléatoires. Considérons par exemple la variable catégorielle niveau d'étude qui permet de distinguer les études qui n'ont enrôlés que des élèves scolarisés au collège de celles n'ayant étudié que des élèves scolarisés au lycée. L'effet de cette variable est fixe, car si on souhaite reproduire cette analyse, les mêmes variables catégorielles devront être utilisées et aucune inférence n'est attendue. Mais si on regroupe les études en fonction du pays où sont scolarisés les élèves, les effets du modérateur « pays » sont aléatoires car les pays représentés dans la méta-analyse constituent un échantillon d'une population plus vaste, et l'analyse est inférentielle. Dans le domaine des sciences de l'éducation, c'est le premier cas qui sera le plus souvent rencontré. On notera également que la plupart du temps les variables étudiées sont catégorielles (à l'image de nos deux exemples précédents) même s'il arrive que des variables continues soient utilisées.

^a On notera par exemple qu'il a donné son nom à une taille d'effet aujourd'hui utilisée par tous les méta-analystes en sciences de l'éducation, le g de Hedges.

On suppose ici que les études ont été sélectionnées et que des tailles d'effet ont été calculées. Il va s'agir maintenant de les agréger (par exemple en constituant des sous-groupes d'études), et pour se faire on peut choisir entre deux modèles : le modèle de l'effet fixe (*fixed model*) et le modèle des effets aléatoires (*random effects*). Le premier considère que pour un même jeu de variables explicatives (par exemple pour un même sous-groupe d'études), les tailles d'effets calculées estiment une seule et même taille d'effet réelle. Dans le second modèle, pour un même jeu de variables explicatives, les tailles d'effets réelles sont normalement distribuées autour d'une taille d'effet réelle globale.

On aboutit donc finalement à 3 catégories de modèles ou d'analyses (tableau 3).

Tableau 3. Types d'analyses (inspiré de Borenstein & al., 2009, p.183) :

Type d'analyse	Modèle pour les tailles d'effet	Effet des modérateurs	# chapitre dans Cooper
Analyse des effets fixes (<i>fixed-effects analysis</i>)	Effet fixe	Effets fixes	12.3.1 (ANOVA) 12.4.1 (méta-régression)
Analyse des effets mixtes (<i>mixed-effects analysis</i>)	Effets aléatoires	Effets fixes	12.3.2 (ANOVA) 12.4.2 (méta-régression)
Analyse des effets aléatoires (<i>random-effects analysis</i>)	Effets aléatoires	Effets aléatoires	Non décrit

D'après ce qui a été dit plus haut, c'est l'analyse des effets mixtes qui sera la plus souvent mise en œuvre en sciences de l'éducation.

2. Estimations des tailles d'effets indépendantes

On va supposer que les données de chaque étude ont été analysées, qu'une seule taille d'effet par étude a été estimée et qu'elles sont toutes indépendantes. Des modérateurs ont été identifiés qui sont traduits soit

- uniquement par des variables catégorielles : dans ce cas on pourra mener des analyses de la variance (ANOVA) pour chacune des variables, ou encore des analyses de sous-groupes,
- des variables continues et/ou plusieurs variables catégorielles que l'on souhaite analyser conjointement : dans ce cas on doit conduire une méta-régression (nom donné à une régression linéaire multiple menée dans le cadre d'une méta-analyse). Ce second cas généralise le premier.

2.1 Analyse de la variance (ANOVA) ou analyse de sous-groupes

Quand une seule variable indépendante catégorielle est analysée à la fois, on mène le plus souvent une analyse de la variance ou ANOVA. On procède alors de la même manière que pour une étude primaire, mais en tenant compte des caractéristiques particulières des estimations des tailles d'effet (voir également Borenstein et al., 2009, p.149). On va par exemple regrouper les études en fonction de l'année de leur publication, ou en fonction de certains types d'intervention ou encore du niveau d'étude des élèves. Quand les influences de plusieurs variables catégorielles sont conjointement analysées, on se tournera plus volontiers vers la méta-régression.

On a déjà vu que pour estimer des tailles d'effets globales, deux modèles peuvent être utilisés : le modèle de l'effet fixe ou le modèle des effets aléatoires. Dans le cas d'une analyse de sous-groupes d'études, le modèle choisi s'applique aux études d'un même sous-groupe. Ainsi on utilisera le

modèle des effets fixes quand on suppose que, dans chacun des sous-groupe, les tailles d'effets réelles des études sont égales ; et on choisira le modèle des effets aléatoires quand on considère que les tailles d'effets réelles sont normalement distribuées autour d'une taille d'effet globale, toujours à l'intérieur d'un même sous-groupe.

La plupart du temps, c'est le modèle des effets aléatoires qui convient aux synthèses quantitatives d'études dans le domaine qui nous intéresse ; ce modèle s'appuie sur le modèle plus simple de l'effet fixe que nous évoquerons donc en premier lieu.

Notations utilisées

On note n le nombre total de tailles d'effet estimées (c'est donc aussi le nombre d'études), et i le rang de l'étude considérée. On a donc $1 \leq i \leq n$. On définit alors :

- θ_i la taille d'effet réelle de l'étude i
- T_i l'estimation de la taille d'effet réelle θ_i
- ν_i la variance^a de T_i
- ε_i l'erreur intra-étude (c'est l'erreur d'échantillonnage)

Pour chacun des deux modèles (effet fixe et effets aléatoires), on va dans un premier temps considérer l'ensemble des études d'une synthèse, puis dans un second temps supposer que ces études ont été réparties dans plusieurs sous-groupes.

Modèle des effets fixes

Si on considère l'ensemble de toutes les tailles d'effet de la méta-analyse (sans partition en groupes), les tailles d'effets réelles θ_i sont toutes égales, donc

$$\theta_1 = \dots = \theta_i = \dots = \theta_n = \theta \quad \text{et} \quad T_i = \theta_i + \varepsilon_i = \theta + \varepsilon_i.$$

On suppose que T_i est normalement distribuée autour de θ . On a donc

$$T_i \sim \mathcal{N}(\theta, \nu_i) \quad \text{et} \quad \varepsilon_i \sim \mathcal{N}(0, \nu_i)$$

Si on partitionne l'ensemble des tailles d'effet en plusieurs groupes (par exemple en regroupant les études en fonction des dates de leur publication) alors, dans un même groupe, les tailles d'effets réelles θ_i sont égales et les remarques précédemment énoncées s'appliquent à chacun des groupes. En notant k le numéro du groupe on a pour l'étude i appartenant à ce groupe :

$$T_i = \theta_k + \varepsilon_i ; \quad T_i \sim \mathcal{N}(\theta_k, \nu_i) \quad \text{et} \quad \varepsilon_i \sim \mathcal{N}(0, \nu_i)$$

Une fois les estimations des tailles d'effet calculées, une analyse de l'hétérogénéité pourra être menée (Roques, 2021).

Modèle des effets aléatoires

Si on considère dans un premier temps l'ensemble de toutes les tailles d'effet de la méta-analyse (sans partition en groupes), les tailles d'effets réelles θ_i sont normalement distribuées autour de la taille d'effet globale réelle θ . On note ξ_i la variation de θ_i autour de θ . On a donc

$$\theta_i = \theta + \xi_i$$

^a Ici aussi, estimation (un nombre réel) et estimateur (une variable aléatoire) ne sont pas distingués, et la notion de variance est évidemment associée à l'estimateur.

La variance de θ_i est égale à la variance de ξ_i et sera notée τ^2 (c'est la variance inter-étude), on a donc $\theta_i \sim \mathcal{N}(\theta, \tau^2)$. On a toujours $T_i = \theta_i + \varepsilon_i$ et donc $T_i = \theta + \xi_i + \varepsilon_i$ avec la variance de T_i égale à $\nu_i + \tau^2$.

Maintenant, si on partitionne l'ensemble des tailles d'effet en plusieurs groupes (par exemple en regroupant les études en fonction des dates de leur publication) alors, dans un même groupe, les tailles d'effets réelles θ_i sont normalement distribuées autour d'une taille d'effet globale avec une même variance inter-étude. Plusieurs modèles peuvent être appliqués pour calculer une estimation des variances inter-étude^a. Et là aussi, une fois les estimations des tailles d'effet calculées, une analyse de l'hétérogénéité pourra être menée (Roques, 2021).

2.2 Méta-régression

Quand certains modérateurs sont des variables continues ou quand plusieurs variables catégorielles sont intégrées dans un même modèle, l'analyse de leur influence sur le lien causal entre l'exposition au traitement et l'évolution des apprentissages va être menée en conduisant une méta-régression, analogue aux régressions linéaires multiples vues au chapitre précédent. Il faudra cependant prendre en compte un élément nouveau concernant la variable dépendante. En effet, si le score d'un élève n'a pas de variance car c'est une mesure, il n'en va pas de même pour la taille d'effet calculée qui est l'estimation d'un paramètre. Pour calculer les coefficients de régression on intègre dans le modèle la matrice diagonale des poids de chaque estimation des tailles d'effet (voir plus loin). L'estimation d'une taille d'effet globale et l'analyse de sous-groupes peuvent alors être considérés comme des méta-régressions particulières.

À nouveau, quand on considère les tailles d'effet réelles des études, on utilise soit le modèle de l'effet fixe, soit le modèle des effets aléatoires. Les numéros des équations données ci-dessous correspondent à la numérotation indiquée par Hedges dans Cooper (2019). Et à nouveau, le premier modèle reste un modèle théorique qui permet de jeter les bases du second modèle plus complexe mais plus adapté à la réalité des expériences.

Modèle de l'effet fixe

Dans ce modèle, on considère que les tailles d'effets réelles des études sont égales pour un même jeu de variables indépendantes donné.

L'objectif est ici de calculer des estimations des coefficients de régression et d'évaluer la signification statistique qu'on peut leur associer. Aux notations précédemment définies, on rajoutera les suivantes :

- On a p covariables X_j (ou variables prédictives) avec $1 \leq j \leq p$; on a plus d'observations que de variables, donc $n > p + 1$; on suppose que les vecteurs de la matrice explicative sont non colinéaires.
- x_{ij} est la valeur de la $j^{\text{ème}}$ covariable pour la taille d'effet i .

On a toujours $T_i = \theta_i + \varepsilon_i$; $\varepsilon_i \sim \mathcal{N}(0, \nu_i)$ et $T_i \sim \mathcal{N}(\theta_i, \nu_i)$

Les erreurs ε_i comme les tailles d'effet T_i sont indépendantes entre elles.

Le modèle linéaire s'écrit alors de la façon suivante :

$$\theta_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}$$

^a Pour plus de détails, voir Hedges dans Cooper (2019, p.259).

$$T_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i \quad (12.56)$$

Dans la première équation, un terme constant β_0 est ajouté à cette formule dans l'article de Hedges dans Cooper (2019, p.266, équation 12.55). Ce terme a été omis ici pour plus de cohérence car il n'a pas été repris dans la suite de l'article comme le montre la seconde équation (12.56).

On définit les vecteurs et la matrice suivants :

$$\mathbf{T} = \begin{pmatrix} T_1 \\ \dots \\ T_n \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} x_{1,1} & \dots & x_{1,p} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \dots & x_{n,p} \end{pmatrix} \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \dots \\ \beta_p \end{pmatrix} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_n \end{pmatrix}$$

La matrice \mathbf{X} est la matrice explicative de dimension $n \times p$. Hedges indique qu'il est habituel de considérer la première colonne de cette matrice comme constituée de 1, ce qui revient à interpréter le premier paramètre de régression comme une constante (ou encore intercept).

On a alors :

$$\mathbf{T} = \mathbf{X} * \boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

On note \mathbf{b} l'estimation de $\boldsymbol{\beta}$ et \mathbf{W} la matrice diagonale des poids w_i de chaque étude (avec $w_i = \frac{1}{v_i}$).

On montre que

$$\mathbf{b} = (\mathbf{t}_X * \mathbf{W} * \mathbf{X})^{-1} * \mathbf{t}_X * \mathbf{W} * \mathbf{T} \quad (12.57)$$

Modèle des effets aléatoires

Dans ce modèle, on considère que les tailles d'effet réelles des études sont normalement distribuées autour d'une même taille d'effet globale pour un même jeu de variables indépendantes donné.

On introduit dans le modèle précédent le terme ξ_i qui représente la variation des tailles d'effet réelles θ_i autour de la taille d'effet réelle globale. La variance de ξ_i est égale à τ^2 ; c'est la variance inter-étude des tailles d'effets réelles. Les erreurs ξ_i sont supposées indépendantes les unes des autres. Les estimations des tailles d'effet réelles T_i sont toujours normalement distribuée autour des tailles d'effets réelles θ_i avec une variance connue v_i . On a donc :

$$T_i = \theta_i + \varepsilon_i ; \quad \varepsilon_i \sim \mathcal{N}(0, v_i) \text{ et } T_i \sim \mathcal{N}(\theta_i, v_i).$$

Le modèle linéaire s'écrit alors de la façon suivante :

$$\theta_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \xi_i \quad (\text{non numérotée, p.269})$$

$$T_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \xi_i + \varepsilon_i \quad (12.64)$$

En utilisant les notations du chapitre précédent et en rajoutant un vecteur $\boldsymbol{\xi} = \begin{pmatrix} \xi_1 \\ \dots \\ \xi_n \end{pmatrix}$, on a :

$$\mathbf{T} = \mathbf{X} * \boldsymbol{\beta} + \boldsymbol{\varepsilon} + \boldsymbol{\xi}$$

Si on note \mathbf{W}^* la matrice diagonale des poids w_i^* de chaque taille d'effet i , (avec $w_i^* = \frac{1}{v_i + \tau^2}$), on a^a :

^a L'astéris est traditionnellement réservé au modèle des effets aléatoires.

$$\mathbf{b}^* = (\mathbf{t}_X * \mathbf{W}^* * \mathbf{X})^{-1} * \mathbf{t}_X * \mathbf{W}^* * \mathbf{T} \quad (12.80)$$

En choisissant des variables muettes adaptées aux catégories analysées, on retrouve les équations proposées pour une ANOVA du chapitre 2.1.

Comme nous l'avons déjà signalé, le paramètre β_0 (ou intercept) est le plus souvent absent des formules proposées et la première colonne constituée des $x_{1,1}, x_{2,1}, \dots, x_{n,1}$ est une colonne de 1 (ce qui revient à considérer β_1 comme une constante). Les deux formules (avec ou sans coefficient constant) coexistent, y compris à l'intérieur d'un même article. C'est le cas par exemple dans l'article de Hedges (2010 b) : la formule générale est donnée sans terme constant (p. 41^a) et celle utilisée dans l'exemple p.50 est $T_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$ et comporte donc un terme constant.

L'interprétation des coefficients reste la même que dans le modèle utilisé pour une étude primaire : β_j représente l'impact ou l'effet de la variable indépendante X_j sur la variable dépendante T_i . Dans le cas de modèles avec constante et de variables muettes, le coefficient est la variation de la taille d'effet quand la catégorie concernée est observée, par rapport à la catégorie de référence.

Des tests statistiques sont également menés sur ces coefficients afin d'en évaluer la précision : des tests Z pour chacun des coefficient (l'hypothèse nulle étant alors associée à un coefficient particulier) ou une analyse de la statistique Q (l'hypothèse nulle étant alors associée à l'ensemble des coefficients)^b.

3. Estimations des tailles d'effets non indépendantes

Deux types de situations peuvent mener à des estimations de tailles d'effet non indépendantes.

Pour la première, il s'agit par exemple d'études qui ont testé un même groupe d'élèves à des moments différents, ou qui ont comparé plusieurs groupes traitement et un même groupe contrôle, ou qui ont mené des tests de nature différente (français, mathématiques, ...). Dans ce cas, ce sont les erreurs ε_i ou encore les estimations des tailles d'effets réelles T_i qui ne sont pas indépendantes. La seconde situation concerne par exemple des études différentes menées par un même chercheur, et ici ce sont les tailles d'effet réelles θ_i qui ne sont pas indépendantes. Cette deuxième source de dépendance a reçu moins d'attention de la part des chercheurs et n'est pas traité par Hedges dans Cooper (2019) qui décrit les analyses mises en œuvre pour le premier type de situation dans le chapitre 13. Ce qui suit ne concerne donc pas cette dernière situation.

Dire que les estimations des tailles d'effet ne sont pas indépendantes, c'est dire qu'elles sont corrélées, et donc que les covariances ou encore les coefficients de corrélation entre ces tailles d'effet (ou entre les erreurs) ne sont pas nuls. Par exemple, si on considère dans une même étude des scores en mathématiques et des scores en lecture, ces scores ne sont pas indépendants (ce sont les mêmes élèves qui ont été testés) et les estimations des tailles d'effets dérivées de ces scores ne sont pas indépendantes également. On considérera alors que ces tailles d'effet comme les scores sont corrélées, et le coefficient de corrélation devra être intégré aux calculs (voir les exemples ci-dessous).

Plusieurs méthodes d'analyses ou modèles sont présentées par Hedges dans Cooper (2019). Le modèle complet multivarié (*full multivariate method*) suppose que les coefficients de corrélation (qui sont souvent notés ρ) sont connus. C'est le cas général qui est proposé dans le chapitre 13.3, c'est-à-dire que les tailles d'effets suivent le modèle des effets aléatoires ; l'estimation de la taille

^a $T_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i$

^b Borenstein (2009, p.191).

d'effet globale sera alors menée en utilisant comme matrice explicative le vecteur unité (un intercept est calculé sans variable prédictive^a) et l'analyse suivant le modèle de l'effet fixe est alors considérée comme un cas particulier. Un modèle simplifié également proposé (13.3.4) n'impose pas de connaître précisément la matrice des covariances^b et consiste à utiliser une matrice approchée ou encore dite « de travail ». Dans ce cas, une analyse de la sensibilité doit être conduite (plusieurs valeurs du coefficient ρ sont testées). Dans le chapitre 13.4, le calcul d'une estimation robuste de la variance (*Robust Variance Estimation, RVE*) est proposé, l'avantage de cette méthode étant que les coefficients de corrélation n'ont pas à être connus. C'est cette procédure qui est privilégiée de nos jours, même si un nombre important (au moins 10 à 20 études par covariable) est requis.

4. Exemples de méta régressions

L'interprétation des coefficients de régression dépend du modèle choisi et du codage des modérateurs (et donc finalement de la matrice explicative). Les résultats de 2 synthèses quantitatives donnés en exemple ci-dessous s'appuient à chaque fois sur l'exploitation de plusieurs modèles, le choix de ces derniers dépendant des questions posées comme de la disponibilité des données. Aucune équation linéaire n'est proposée dans ces deux publications.

4.1 Synthèse sur l'enseignement des mathématiques et de la lecture

Une synthèse quantitative a été menée par Dietrichson & al., (2021) sur les interventions en mathématiques et en lecture en direction des élèves en difficulté à l'école primaire. 607 études ont été rassemblées et plusieurs analyses réalisées afin de répondre à plusieurs questions de recherche. Trois types de caractéristiques (ou modérateurs) ont été considérés : des caractéristiques pédagogiques au nombre de 8, des domaines d'apprentissage (6 en lecture, 7 en mathématiques et 3 dans des domaines plus généraux) et enfin trois caractéristiques concernant le design des études (les études quasi-expérimentales versus les essais contrôlés randomisés, les études concluant à partir de tests généraux versus tests par sous-domaines ou bien à partir de tests en mathématiques versus tests en lecture, le niveau d'étude des élèves). Les deux premiers ensembles rassemblent des modérateurs substantiels des interventions (voir la discussion plus loin) qui ne sont pas mutuellement exclusifs : plusieurs caractéristiques pédagogiques peuvent coexister au sein d'une même intervention par exemple. Les 351 coefficients de corrélations entre les différentes paires de modérateurs ont été calculés (Dietrichson & al., 2021, tableau 2, p.32). Parmi eux, 12 sont significativement positifs : 6 entre des variables du domaine mathématiques, 6 entre des variables du domaine lecture ; un coefficient est significativement négatif, entre les variables « apprentissage soutenu par les pairs » (*peer-assisted instruction*) et « apprentissage en petits groupes » (*small group instruction*), modérateurs qui font partie du premier ensemble. Les autres coefficients ne sont pas statistiquement différents de 0.

Deux types d'analyses ont ensuite été menées : des analyses de sous-groupes unifactorielles (un seul facteur est analysé à chaque fois) puis des méta-régressions multifactorielles.

Les analyses de sous-groupes concernent chacun des modérateurs des deux premiers ensembles. Pour chaque groupe, une méta-régression dont l'équation linéaire se réduit au seul intercept a été réalisée ; l'intercept calculé est alors l'estimation de la taille d'effet globale (notée ES ci-dessous) du sous-groupe considéré. Les trois tailles d'effet (avec leurs intervalles de confiance à 95 %) les plus

^a Voir un exemple de calcul de moyenne dans l'annexe 3 pour une étude primaire.

^b C'est la matrice symétrique constituée des covariances des paires des variables considérées (ici les tailles d'effet) ; les éléments de la diagonale sont les variances de ces variables.

importantes concernent le domaine des fractions (ES=0,50 [0,15 ; 0,86]), les caractéristiques pédagogiques « apprentissage soutenu par les pairs » (ES=0,44 [0,28 ; 0,61]) et « apprentissage en petits groupes » (ES=0,38 ; [0,31 ; 0,44]). Une seule taille d'effet n'est pas significativement différente de zéro (elle concerne le domaine des compétences socio-émotionnelles).

Quatre méta-régressions multifactorielles avec estimation robuste de la variance ont également été menées. Les variables continues^a ont été centrées et les calculs menés avec un coefficient de corrélation entre les tests d'une même étude fixé à 0,8 (une analyse de la sensibilité a été conduite). Le modèle de la première méta-régression ne comporte pas de constante et n'intègre pas les variables caractérisant les études (modèle 1, voir tableau ci-dessous). Pour le second modèle (modèle 2), une constante a été intégrée ainsi que les variables caractérisant les études ; les 3 indicateurs « autres méthodes pédagogiques », « plusieurs domaines de lecture » et « plusieurs domaines de mathématiques » ont été exclues et constituent ainsi des catégories de référence (ce sont en fait des catégories « autres » et cette exclusion évite la colinéarité des variables). Pour la troisième méta-analyses (modèles 3) les tailles d'effet incluses sont basées sur les tests en lecture et pour la quatrième (modèle 4) sur les tests en mathématiques. Ces deux modèles avec constante n'incluent pas les modérateurs « domaines » et « caractéristiques des études », et l'indicateur « autre méthode » a également été exclu (pour les mêmes raisons que pour le modèle 2). Les coefficients de régression (Dietrichson & al., 2021, tableau 5, p.36) statistiquement significatifs sont présentés dans le tableau 4 ci-dessous (les modérateurs sélectionnés sont soit statistiquement significatifs, soit exclus dans au moins un des modèles).

Dans ces méta-analyses, des modérateurs de nature différente sont intégrés dans un même modèle : les caractéristiques pédagogiques sont traitées de la même façon que les caractéristiques des études dans les modèles 2, 3 et 4 (voir la discussion dans la dernière partie).

Tableau 4. Principaux résultats des méta-régressions (coefficients de régression et intervalle de confiance à 95%)

Modérateurs		Modèle 1	Modèle 2	Modèle 3	Modèle 4
Caractéristiques pédagogiques	Coaching	non significatif	non significatif	non significatif	-0,23 [-0,40 ; -0,06]
	Apprentissage soutenu par les pairs	0,41 [0,24 ; 0,59]	0,40 [0,13 ; 0,67]	0,42 [0,07 ; 0,76]	0,38 [0,12 ; 0,65]
	Apprentissage en petits groupes	0,36 [0,26 ; 0,47]	0,34 [0,12 ; 0,56]	0,32 [0,07 ; 0,58]	0,24 [0,11 ; 0,37]
	Autres méthodes	non significatif	Exclue	Exclue	Exclue
Domaines	Plusieurs domaines lecture	non significatif	Exclue	x	x
	Plusieurs domaines mathématiques	non significatif	Exclue	x	x
	Fractions	0,33 [0,01 ; 0,65]	0,37 [0,06 ; 0,67]	x	x
Caractéristiques des études	Niveau d'étude (grades)	x	-0,03 [-0,06 ; -0,00]	-0,04 [-0,07 ; -0,02]	non significatif
	Constante	x	non significatif	non significatif	0,19 [0,01 ; 0,37]

X : non considéré par le modèle.

^a Ce sont les variables caractérisant le dosage des interventions.

4.2. Synthèse sur l'enseignement des fractions

Schumacher (2020) a conduit une synthèse quantitative sur les interventions portant sur l'enseignement des fractions aux élèves en difficulté et scolarisés du CE2 à la seconde. Les influences de 6 caractéristiques pédagogiques (comme l'enseignement explicite ou les révisions cumulatives) sur les apprentissages des élèves ont été analysées en conduisant une méta-régression avec un coefficient de corrélation égal à 0,8 (une analyse de la sensibilité a été menée) en utilisant le modèle des effets aléatoires avec une estimation robuste de la variance. Là encore, les caractéristiques pédagogiques ne sont pas exclusives : une intervention peut par exemple s'appuyer sur un enseignement explicite et des révisions cumulatives. Pour être retenue dans la méta-régression, une caractéristique pédagogique devait être analysée par au moins 10 études différentes. Toutes les caractéristiques sont incluses simultanément dans un seul modèle et contrôlées pour des caractéristiques importantes liées aux études ou aux interventions : taille des groupes d'élèves, niveau d'étude, durée de l'intervention et nature du groupe contrôle. Les coefficients de régression^a (Schumacher, 2020, tableau 5, p.44) sont interprétés par l'auteure comme des tailles d'effet. Seule la taille d'effet associée à la variable « enseignement et utilisation du langage mathématique » est statistiquement significative avec $ES = 0,57$ et un intervalle de confiance à 95 % égal à $[0,01 ; 1,14]$ ^b.

Les influences de 10 variables catégorielles liées aux caractéristiques des études et des interventions ont été explorées. Le nombre d'études étant relativement faible par rapport au nombre de variables catégorielles, des modèles unifactoriels ont été utilisés en lieu et place d'un modèle multifactoriel. Là encore cela revient à proposer des tailles d'effet (Schumacher, 2020, tableau 6, p.45). Pour chaque variable catégorielle, une variable indicatrice a été exclue et ainsi choisie comme catégorie de référence. Les résultats ne sont pas statistiquement significatifs pour quatre variables catégorielles après correction des valeurs-p en suivant la méthode de Bonferroni^c. Pour les 6 autres variables, les résultats significatifs sont présentés dans le tableau 5 ci-dessous. L'auteure interprète ces résultats en les comparant à la catégorie de référence. Par exemple, elle remarque que l'effet des interventions en direction des élèves scolarisés en primaire est significativement supérieur à l'effet des interventions au collège ou que l'effet des interventions conduites par le personnel de l'établissement est significativement inférieur à l'effet des interventions menées par des membres de l'équipe de chercheurs.

^a Le terme « coefficient de régression » n'est pas employé par l'auteure de l'article.

^b Ici aussi les tailles d'effet sont les g de Hedges.

^c Il s'agit ici de tenir compte de l'inflation de l'erreur de type I liée à la multiplicité des tests d'hypothèse.

Tableau 5. Coefficients de régression significatifs avec intervalles de confiance à 95%

Domaines mathématiques	
Connaissance des nombres rationnels	Exclue
Magnitude des nombres rationnels	0,47 [0,22 ; 0,72]
Addition et soustraction de fraction	0,52 [0,14 ; 0,90]
Niveau d'étude	
Collège	Exclue
Primaire	0,43 [0,22 ; 0,63]
Taille des groupes	
Plus de 6 élèves	Exclue
De 2 à 6 élèves	0,47 [0,28 ; 0,66]
Intervenant	
Membre de l'équipe de recherche	Exclue
Personnel de l'établissement	-0,43 [-0,67 ; -0,20]
Durée	
Moins de 9 heures	Exclue
Plus de 19 heures	0,45 [0,24 ; 0,66]
Formation continue	
Non	Exclue
Oui	0,48 [0,29 ; 0,67]

Partie 3. Discussion

La discussion concernera essentiellement les méta-analyses et va porter sur trois points. Les deux premiers concernent des éléments évoqués précédemment : la nature des variables explicatives pour le premier et la nature des questions de recherche pour le deuxième. La troisième est d'ordre plus général et porte sur la difficulté qu'il y a à comprendre mais aussi à interpréter les résultats d'une méta-régression.

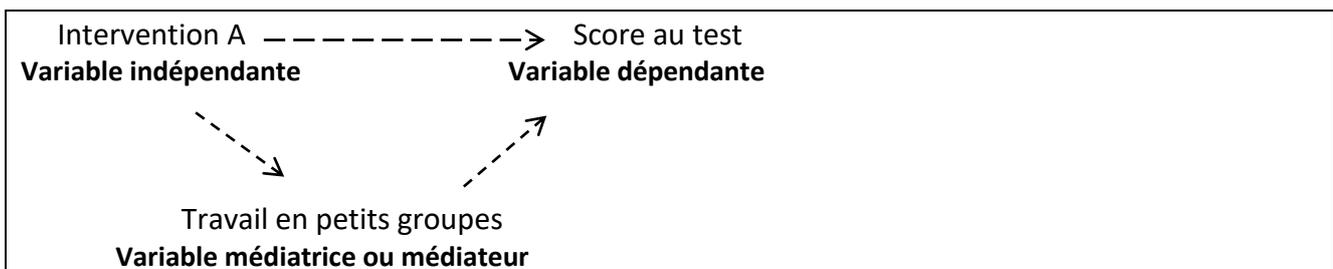
1. Modérateurs ou médiateurs ?

Pour mieux comprendre la question qui se pose quand des variables indépendantes de nature différentes sont intégrées dans une même méta-régression, nous allons devoir ajouter à notre panoplie de variables une nouvelle catégorie : les variables **médiatrices** (ou **médiateurs**). Ces variables médiatrices sont la plupart du temps définies en opposition aux variables modératrices (ou modérateurs) dont nous avons analysé les effets tout au long des deux parties précédentes.

Les médiateurs

Les variables médiatrices sont associées à des caractéristiques qui expliquent la relation entre une variable indépendante et une variable dépendante. Elles permettent de répondre à des questions du type « comment ? » ou « pourquoi ? ». C'est par exemple une caractéristique intégrée au mécanisme qui permet de comprendre comment une variable indépendante influence une variable dépendante. Elles interviennent dans la relation causale, comme résultat causal de la variable indépendante et comme antécédent causal de la variable dépendante. La relation entre les variables indépendante et dépendante n'existe pas en l'absence du médiateur.

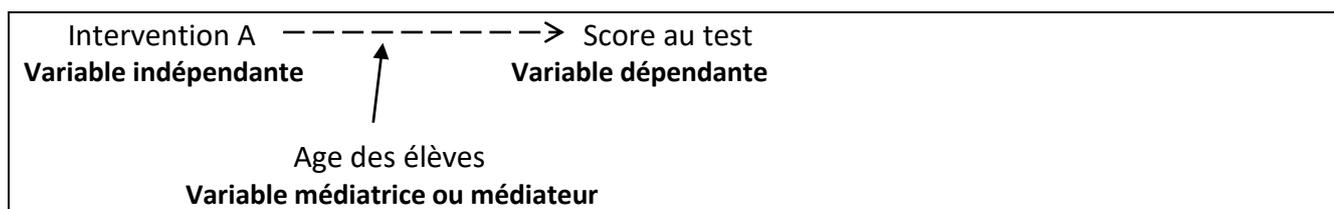
Exemple d'une variable médiatrice



Les modérateurs

Les variables modératrices sont associées à des caractéristiques qui influent sur l'intensité de la relation entre une variable indépendante et une variable dépendante, pouvant parfois même (en théorie) en inverser le sens. Leur identification permet de répondre à des questions du type « quand ? » ou « pour qui ? ». Pour être considérée comme modératrice d'un traitement, la caractéristique concernée doit être indépendante du traitement, ce qui signifie entre autres qu'elle ne soit pas corrélée à ce traitement (Wu & Zumbo, 2008). La relation entre les variables indépendante et dépendante existe même en l'absence de modérateur.

Exemple d'une variable modératrice



Deux catégories de modérateurs peuvent être définies (Tipton & al., 2019) : les modérateurs substantiels qui sont liés à l'intervention ou à la population étudiée (comme une caractéristique pédagogique, le moment de passation du test ou l'âge des élèves) et les modérateurs méthodologiques qui sont liés à l'étude ou au traitement des résultats (comme le design de l'étude, le type de test, la date de publication).

Médiateur ou modérateur ?

Dans la méta-analyse de Dietrichson & al. (2021), la caractéristique pédagogique « travailler en petits groupes » a été considérée comme une variable modératrice. On peut légitimement se demander si cette caractéristique ne devrait pas plutôt être considérée comme une variable médiatrice. Un examen plus précis du rôle joué par cette caractéristique pédagogique au sein même de l'intervention étudiée devrait permettre de trancher. Deux éventualités peuvent être envisagées :

- l'enseignement en petits groupes est une caractéristique majeure de l'intervention, par exemple son absence de l'enseignement dénaturerait profondément ce dernier voire même le rendrait caduque, alors la caractéristique « enseignement en petits groupes » est une variable médiatrice ;
- l'enseignement en petits groupes est une caractéristique mineure de l'intervention, par exemple son absence de l'enseignement ne ferait que modifier l'intensité du lien entre l'intervention et les apprentissages des élèves, alors cette caractéristique est une variable modératrice.

C'est donc, dans certains cas, au niveau de la définition même de l'intervention et des concepts théoriques qui la structurent que la nature d'une variable peut être déterminée. La question du traitement statistique de ces deux types de variables reste encore posée, car si des modèles statistiques différents selon la nature de la variable (modératrice ou médiatrice) ont pu être identifiés pour des études primaires (WU, 2008), les méta-régressions conduites dans le domaine des sciences de l'éducation semblent traiter de la même façon les caractéristiques pédagogiques et les modérateurs tels que le design de l'étude ou l'année de la publication de l'étude.

2. Analyse confirmatoire ou analyse exploratoire ?

Les auteurs de méta-analyses comme les auteurs d'études primaires doivent préciser si leur analyse est de type confirmatoire ou de type exploratoire. Dans le premier cas, une question de recherche est posée, et l'expérience est conçue pour y répondre : une hypothèse est formulée et les critères qui permettent de conclure sont définis *a priori*, c'est-à-dire avant d'avoir connaissance des résultats de l'expérience. Dans le second cas, la question de recherche n'est pas uniquement basée sur des éléments définis en amont de l'expérience, mais aussi au moins partiellement sur une

observation des résultats (qui peut par exemple amener les chercheurs à conduire une analyse de sous-groupe). Or seules les analyses confirmatoires sont à même de prouver un lien de cause à effet. Les analyses exploratoires sont destinées à poser des hypothèses qui devront être vérifiées dans un second temps (tableau 6.) C'est pour cette raison que les résultats de ces dernières doivent être clairement séparés des résultats des analyses du premier type, car leurs conclusions n'ont pas la même portée^a.

Tableau 6. Analyses confirmatoires ou exploratoires : exemples de questions de recherches

	Analyse confirmatoire	Analyse exploratoire
Étude primaire	« L'intervention est-elle efficace ? » (EEF, 2018)	« L'effet de l'intervention est-il différent pour les élèves FSM ? » (EEF, 2018)
Étude secondaire	« Quelle efficacité globale peut-on attribuer aux interventions étudiées dans cette méta-analyse ? » (Dietrichson & al., 2021)	« Quelles sont les caractéristiques associées à l'efficacité des interventions ? » (Dietrichson & al., 2021)

Trois prismes de lecture permettent de différencier les deux types d'analyses. Comme on l'a vu, la question de recherche constitue le premier d'entre eux. Le design des études en est un autre. Ainsi, pour une étude primaire, c'est la répartition aléatoire des élèves dans le groupe traitement et le groupe contrôle qui permet, après une analyse correcte des résultats, de mettre à jour une relation de cause à effet entre un facteur (ici les différentes conditions de traitement) et une évolution des scores. Cette relation causale entre le traitement et les scores est protégée par le hasard de la répartition, et cette protection perdure quand un effet global est estimé dans une méta-analyses (Borenstein et al., 2009, p.209). C'est également le cas quand on analyse l'effet d'une caractéristique de l'enseignement (comme une méthode pédagogique) si on la considère comme intimement liée à l'intervention (et donc comme une variable modératrice substantielle ou comme une variable médiatrice). On retiendra que pour mener une analyse de type confirmatoire, les essais contrôlés randomisés constituent le type même d'études à privilégier.

En troisième lieu, c'est le type d'analyses statistiques menées sur les résultats (aussi bien pour une étude primaire que pour une méta-analyse) qui permet aussi de distinguer une analyse de type confirmatoire d'une analyse de type exploratoire. Comme nous l'avons déjà mentionné, les analyses de sous-groupes ou les régressions linéaires multiples (méta-régressions pour les méta-analyses) ne permettent pas de démontrer l'existence d'un lien de cause à effet entre une variable indépendante (la date de publication des études par exemple) et une variable dépendante (la taille d'effet). En effet, ce sont les premiers résultats obtenus par le chercheur qui permettent de définir des sous-groupes, aussi bien pour les études primaires (ce sont des groupes d'élèves) que pour les méta-analyses (ce sont des groupes d'études). Et prendre une décision en fonction des premiers résultats d'une étude puis mener une analyse à partir de cette décision permet d'explorer des liens de causalité mais non de les confirmer.

Ceci étant dit, les analyses statistiques corrélationnelles peuvent dans certains cas, en s'appuyant par exemple sur l'expertise et/ou les connaissances accumulées dans un domaine, conduire le chercheur à aller au-delà de ce que le seul modèle statistique permet. C'est le contexte qui peut faire ici la différence.

^a Il est recommandé de nos jours de ne pas publier les résultats de tests statistiques (comme les valeurs-p) pour les estimations issues des questions de recherche exploratoires.

3. Pour conclure

Dans « *A history of meta-regression : technical, conceptual and practical developments between 1978 and 2018* » publié en 2019, Tipton trace les grandes lignes de l'évolution des théories et des pratiques liées à la conduite de méta-régression. Certaines questions sont encore d'actualité qui concernent, sinon les limites de ce type d'analyse, du moins les précautions à prendre quant à l'interprétation de leurs résultats. Une attention particulière devrait être portée à la corrélation éventuelle entre les modérateurs (que ce soit entre les modérateurs substantiels les uns avec les autres, ou entre des modérateurs substantiels et certaines caractéristiques méthodologiques des études) et plus particulièrement aux biais de confusion qui peuvent en résulter. Ces questions restent encore ouvertes et la sélection des modérateurs devrait faire l'objet de futures recherches. Un autre sujet d'intérêt concerne la multiplicité des tests car l'inflation de l'erreur de type I (le risque α) qui en est la conséquence doit être prise en compte par les chercheurs et des corrections apportées. Le nombre de questions de recherche de type confirmatoire devrait être fixé en se basant sur les résultats d'analyse de puissance menée en amont ; dans le même ordre d'idée, un nombre de 20 études minimum par modérateur est préconisé.

On terminera enfin par quelques réflexions sur un élément de fond essentiel puisqu'il s'agit de la finalité de ces recherches complexes et coûteuses. Les méta-régressions conduites dans le domaine des sciences de l'éducation ont des objectifs pragmatiques ambitieux : permettre aux praticiens (enseignants, formateurs, responsables administratifs) d'améliorer l'environnement des élèves, par exemple en ciblant certaines pratiques pédagogiques ou une organisation de la scolarité particulière qui favoriseraient leurs apprentissages. Mais comprendre les résultats de ces analyses statistiques qui emploient une terminologie peu claire, des notations disparates, et bien entendu des concepts et des calculs mathématiques complexes, n'est pas à la portée de tous. Les auteurs ne prennent pas toujours le temps d'explicitier leurs méthodes : on a pu constater dans les exemples proposés que les modèles linéaires ne sont pas donnés, ni le codage des variables indépendantes explicité, ce qui rend les résultats des méta-régressions difficiles (voire impossibles) à comprendre, obligeant le lecteur à s'en remettre aux conclusions de l'auteur et à abandonner son esprit critique. Quitter le monde des équations linéaires à une seule variable ou des regroupements de données dans des catégories mutuellement exclusives pour celui plus ténébreux des modèles comportant plusieurs variables, ne peut mener qu'à une complexification des représentations et par conséquent des interprétations et des conclusions. Et vouloir tenir compte de très nombreux paramètres (variables explicatives de premier plan comme modérateurs plus obscurs) ne peut se faire sans convoquer des modèles compliqués et désincarnés^a au prix d'une perte de sens voire même d'un affaiblissement des conclusions. Ces éléments de réflexions concernent évidemment les chercheurs dont nous avons largement évoqué les travaux ici, mais aussi les personnes en charge de l'organisation de la recherche en sciences de l'éducation et de l'enseignement publique au niveau des États.

^a Les calculs requièrent l'utilisation d'un logiciel statistique qui devrait rester un outil réservé aux statisticiens.

Références

- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.
- Bressoux, P. (2007). L'apport des modèles multiniveaux à la recherche en éducation. *Éducation et didactique*, (1-2), 73-88.
- Bressoux, P. (2010). Modélisation statistique appliquée aux sciences sociales [Statistical modeling applied to social sciences]. *De Boeck Supérieur, Louvain-la-Neuve*.
- Cooper, H., Hedges, L., & Valentine, J. (2019). *The Handbook of Research Synthesis and Meta-Analysis*. Russell Sage Foundation.
- Dietrichson, J., Filges, T., Seerup, J. K., Klokke, R. H., Viinholt, B. C. A., Bøg, M., & Eiberg, M. Targeted school - based interventions for improving reading and mathematics for students with or at risk of academic difficulties in Grades K - 6: A systematic review. *Campbell Systematic Reviews*. 2021;e1152. <https://doi.org/10.1002/cl2.1152>
- Education Endowment Foundation (2018), *1stClass@Number. Evaluation report and executive summary*, July 2018, NUNES T., BARROS, R. , EVANGELOU M., STRAND S., MATHERS S., SANDERS-ELLIS D. <https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/1stclassnumber>
- Feingold, A. *A Regression Framework for Effect Size Assessments in Longitudinal Modeling of Group Differences*. *Rev Gen Psychol*. 2013 March ; 17(1): 111–121. doi:10.1037/a0030048.
- Givord, P. & Guillermin, M. (2016). Les modèles multiniveaux. INSEE. <https://www.insee.fr/fr/statistiques/2022152>
- Hedges, L. V. (2007). Effect sizes in cluster-randomized designs, *Journal of Educational and Behavioral Statistics*, 32(4) : 341-447.
- Hedges, L. V., & Rhoads, C. (2010 a). Statistical power analysis in education research. *EE. UU.: IES National Center for Special Education Research*. Retrieved July, 6, 2016.
- Hedges, L., Tipton, E., & Johnson, M. (2010 b). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Method*, 1(1), 39-65. <https://doi.org/10.1002/jrsm.5>
- Roques N. (2021). Mesurer l'effet d'un traitement. Les méta-analyses en sciences de l'éducation. www.mathadoc.fr
- Roques, N. (à paraître). Les essais contrôlés randomisés au Royaume-Uni : évaluer les interventions efficaces pour favoriser l'apprentissage des mathématiques. *Éducation & Didactique*.
- Schumacher, R., Korwaka, S., Haymond, K., Newman-Gonchar, R. & Gersten, R. (2020). *A systematic review and meta-analysis of rational number interventions for students experiencing difficulty with mathematics*, Los Alamitos, CA : Instructional Research Group
- Tipton, E., Pustejovsky, J., Ahmadi, H. (2019). A history of meta - regression: Technical, conceptual, and practical developments between 1974 and 2018. *Research Synthesis Methods*, 10(2), 161-179. <https://doi.org/10.1002/jrsm.1338>
- Wu, A. D., & Zumbo, B. D. (2008). Understanding and using mediators and moderators. *Social Indicators Research*, 87, 367-392.

Annexes

Annexe 1. ANOVA, ANCOVA, MANOVA

ANOVA : analyse de la variance (*ANalysis Of Variance*). On analyse la différence des moyennes d'une variable continue dépendante répartie dans plusieurs groupes. La variable qui distingue les groupes est un facteur (ou variable explicative catégorielle ou variable indépendante catégorielle). Dans le cas de groupes d'élèves, ce sera un facteur inter-sujet.

One-way ANOVA (ou ANOVA) : étude d'un seul facteur. On parle aussi d'ANOVA unidirectionnelle ou d'ANOVA à un facteur.

Two-way ANOVA : étude de 2 (ou plusieurs) facteurs. On parle d'ANOVA à 2 (ou plusieurs) facteurs.

ANCOVA : analyse de la covariance. On analyse une variable dépendante en fonction d'un facteur et d'une variable continue indépendante aussi appelée covariable (*covariate*). Par exemple, les scores pré-test des élèves sont des covariables.

MANOVA : analyse de 2 (ou plus) variables dépendantes continues. De la même façon que pour l'ANOVA, on peut avoir une one-way MANOVA et une two-way MANOVA.

Annexe 2. Exemple des loyers : ANOVA et régression linéaire multiple

Cet exemple numérique est issu de *Pratique de la Régression Linéaire Multiple. Diagnostic et sélection de variables* de Rakotomala^a. Il s'agit d'analyser les 15 loyers de logements en fonction de leur situation géographique : centre-ville (centre), banlieue ou campagne.

Les formules générales utilisées pour une ANOVA sont présentées puis appliquées à l'exemple numérique puis trois modèles linéaires sont proposés.

Les notations sont celles de Rakotomala.

Un fichier excel est disponible sur le site mathadoc.fr (avec notamment les calculs des coefficients à partir des matrices explicatives).

Analyse descriptive

On a les données suivantes :

Loyers (: 100)	Situation géographique			Moyennes	n
	banlieue	campagne	centre		
6,9	OUI	NON	NON	6,72	5
6,3	OUI	NON	NON		
6,7	OUI	NON	NON		
6,5	OUI	NON	NON		
7,2	OUI	NON	NON		
5,6	NON	OUI	NON	5,27	3
4,9	NON	OUI	NON		
5,3	NON	OUI	NON		
7	NON	NON	OUI	7,69	7
7,5	NON	NON	OUI		
8	NON	NON	OUI		
7,2	NON	NON	OUI		
8,4	NON	NON	OUI		
7,4	NON	NON	OUI		
8,3	NON	NON	OUI		
TOUT				6,88	15

Les moyennes des loyers (divisés par 100) des logements situés en banlieue, à la campagne, au centre, sont respectivement 6,72 ; 5,27 et 7,69. La moyenne de tous les loyers est de 6,88.

Formules générales d'une ANOVA

On analyse les données en répartissant les n observations de la variable à expliquer (ici le montant des loyers) dans k groupes. On note i le rang de l'observation et j le rang du groupe, on a donc $1 \leq i \leq n$ et $1 \leq j \leq k$.

On définit les sommes suivantes^b :

$$SC_{intra} = \sum (y_{ij} - \bar{y}_j)^2 = \text{Somme des carrés des écarts à la moyenne de chaque groupe}$$

$$SC_{total} = \sum (y_{ij} - \bar{y})^2 = \text{Somme des carrés des écarts à la moyenne générale}$$

^a Rakotomalala R., *Pratique de la Régression Linéaire Multiple. Diagnostic et sélection de variables*. Version 2.1, p.84 https://eric.univ-lyon2.fr/ricco/cours/cours/La_regression_dans_la_pratique.pdf

^b Pour ne pas alourdir inutilement le texte, la notation utilisée ici ne suit pas le schéma traditionnel des notations « somme ».

$SC_{inter} = \sum n_j (\bar{y}_j - \bar{y})^2 =$ Somme des carrés des écarts des moyennes de chaque groupe à la moyenne générale pondérée par le nombre de données de chaque groupe.

On montre que :

$$SC_{total} = SC_{intra} + SC_{inter}$$

Dans le but de mener un test d'hypothèse et de tester l'égalité globale des moyennes conditionnelles, on calcule alors la statistique $F(k - 1, n - k)$ de la façon suivante :

$$F(k - 1, n - k) = \frac{\frac{SC_{inter}}{k - 1}}{\frac{SC_{intra}}{n - k}}$$

ANOVA des 15 loyers

Avec les données de l'exemple ($n = 15$ et $k = 3$), on a :

Sommes	SC_{total}	15,02	SC_{inter}	12,48	SC_{intra}	2,54
Degrés de liberté	$n - 1$	13	$k - 1$	2	$n - k$	12

$$F = \frac{12,48}{2} \times \frac{12}{2,54} = 29,44$$

Cette valeur est alors comparée à une valeur critique ou une valeur-p est calculée (et on montre ici que l'hypothèse nulle sur l'égalité des moyennes des loyers est rejetée).

Régression linéaire multiple

La variable explicative « situation géographique » est catégorielle ; ses trois modalités sont : banlieue, campagne et centre. Ces modalités définissent trois catégories de logement. Pour appliquer un modèle linéaire, on va définir trois variables indicatrices :

$Z_{banlieue}$, $Z_{campagne}$, et Z_{centre} .

Situation	$Z_{banlieue}$	$Z_{campagne}$	Z_{centre}
Banlieue	1	0	0
Campagne	0	1	0
Centre	0	0	1

Par exemple, si l'appartement est en banlieue, alors la variable $Z_{banlieue}$ est égale à 1, et les deux autres variables sont égales à 0. Soit i le rang de l'appartement. On remarque que pour tout appartement i , on a : $Z_{i,banlieue} + Z_{i,campagne} + Z_{i,centre} = 1$

Trois modèles linéaires sont proposés. Les résultats ont été calculés par Excel en utilisant la fonction DROITEREG (voir annexe 5).

Modèle 1 : modèle avec constante et trois variables indicatrices

$$Loyer = a_0 + a_1 \times Z_{banlieue} + a_2 \times Z_{campagne} + a_3 \times Z_{centre} + \varepsilon$$

Les vecteurs de la matrice explicative sont colinéaires ce qui interdit d'utiliser ce modèle plus avant.

Modèle 2 : modèle sans constante et trois variables indicatrices

On propose le modèle suivant :

$$Loyer = a_1 \times Z_{banlieue} + a_2 \times Z_{campagne} + a_3 \times Z_{centre} + \varepsilon$$

On obtient avec Excel les résultats suivants :

a_3	a_2	a_1
s_3	s_2	s_1
R^2	s	#
F	ddl	#
SC_{reg}	$SC_{résidus}$	#

centre	campagne	banlieue
7,68	5,26	6,72
0,17	0,26	0,20588177
0,996	0,460	#N/A
1136,34	12	#N/A
722,49	2,543	#N/A

Les trois valeurs de la première ligne sont les moyennes conditionnelles de l'ANOVA.

Dans les modèles sans constante, les algorithmes d'Excel ne permettent pas de retrouver R^2 , F et SC_{reg} tels que nous les avons définis dans les chapitres 3.2 et 3.3 de la partie 1. Pour comprendre les calculs conduits par Excel dans les modèles sans constante, voir l'annexe 5.

Modèle 3

On utilise deux variables indicatrices en excluant la troisième variable (qui peut se déduire des deux autres). Ici, $Z_{centre} = 1 - Z_{banlieue} - Z_{campagne}$ et sera la variable exclue.

Situation	$Z_{banlieue}$	$Z_{campagne}$
Banlieue	1	0
Campagne	0	1
Centre	0	0

Le modèle s'écrit alors :

$$Loyer = a_0 + a_1 \times Z_{banlieue} + a_2 \times Z_{campagne} + \varepsilon$$

Les résultats suivants sont calculés :

a_2	a_1	a_0
s_2	s_1	s_0
R^2	s	#
F	ddl	#
SC_{reg}	$SC_{résidus}$	#

campagne	banlieue	constante
-2,41904762	-0,96571429	7,68571429
0,31768246	0,26956251	0,17400185
0,83072164	0,46036562	#N/A
29,4445776	12	#N/A
12,4807619	2,5432381	#N/A

On a finalement :

$$Loyer = 7,69 - 0,97 \times Z_{banlieue} - 2,42 \times Z_{campagne} + \varepsilon$$

Si $Z_{banlieue} = 1$, alors $Z_{campagne} = 0$ et on a :

$$Loyer = 7,69 - 0,97 + \varepsilon = 6,72 + \varepsilon$$

L'estimation de a_0 est la moyenne conditionnelle de la variable exclue (loyer centre). L'estimation de a_1 est la différence entre la moyenne conditionnelle banlieue et la moyenne conditionnelle de la

variable exclue, et c'est la même chose pour a_2 qui est la différence entre la moyenne conditionnelle campagne et la moyenne conditionnelle centre.

On peut montrer que $SC_{reg} = SC_{inter}$ et $SC_{résidus} = SC_{intra}$ avec SC_{inter} et SC_{intra} définis précédemment (voir les résultats de l'ANOVA ci-dessus). On remarque que cette fois le calcul de F donne le même résultat que l'ANOVA.

Remarque sur les tests

Avec le modèle 2, les tests d'hypothèses conduisent à étudier la nullité des paramètres c'est-à-dire des moyennes. Avec le modèle 3, les tests d'hypothèses conduisent à étudier la nullité des paramètres qui sont cette fois des différences de moyennes. Et c'était bien l'objectif premier de l'analyse de cet exemple (et l'objectif d'une analyse de la variance).

D'autres codages sont possibles. Le type de codage retenu conditionne l'interprétation des coefficients.

Annexe 3. Évaluation d'une intervention (étude primaire)

On exploite des données fictives pour calculer et interpréter les coefficients de régression pour 6 modèles linéaires. Les scores posttests et prétests sont ceux d'un exemple publié dans le document *Mesurer l'effet d'un traitement. Méta-analyse et enseignement* (ROQUES, 2021, exemple 8).

Les résultats de six modèles de régressions sont présentés. Ils ont été calculés à l'aide de la fonction DROITEREG d'Excel (voir annexe 5). Dans les modèles sans constante, les valeurs calculées par Excel pour R^2 , F et $SC_{régression}$ ne peuvent pas être interprétés comme nous l'avons vu aux chapitres 3.2 et 3.3 de la partie 1 (voir l'annexe 5 pour des explications à ce sujet). Le fichier Excel peut être téléchargé sur le site internet www.mathadoc.fr¹.

Modèles	1	2	3	4	5	6
Variabes explicatives	• Variable = 1	• Variable traitement	• Variable traitement • Variable contrôle	• Variable traitement • Scores prétests non centré	• Variable traitement • Variable contrôle • Scores prétests non centré	• Variable traitement • Scores prétests centrés
Constante	Non	Oui	Non	Oui	Non	Oui

Modèle 1. Variable = 1, pas de constante. Calcul de la moyenne

On calcule la moyenne et l'écart-type des scores posttest en utilisant un modèle linéaire. Avec un modèle sans constante et une variable explicative qui a la même valeur 1 pour toutes les données, on a un modèle du type $y = b_1 + e$.

Résultats du tableur	
6	0
0,18068435	#N/A
0,906304729	1,93762379
1102,71028	114
4140	428

b1	6,000	Moyenne des scores posttest
s	1,938	Ecart-type des scores posttests

Modèle calculé : $y = 6,00 + e$

Modèle 2. Variable traitement et une constante. Calcul de la taille d'effet non ajustée

Modèle : $y = b_0 + b_1*t + e$ avec $t=1$ pour le groupe traitement et $t=0$ pour le groupe contrôle

Résultats du tableur	
0,766	5,627
0,356	0,248
0,039	1,907
4,630	113,000
16,846	411,154

b1	0,766	Différence des moyennes
b0	5,627	Moyenne du groupe contrôle
s1	0,356	
s0	0,248	
R ²	0,039	
s	1,907	Erreur type échantillon = écart-type groupé
F	4,630	

Modèle calculé : $y = 5,63 + 0,766*t + e$

¹ <http://www.mathadoc.fr/2021/04/19/statistiques/>

$R^2 = 0,039$ ou 3,9%. Le modèle de régression n'explique que pauvrement les variations des scores posttests.

On vérifie par le calcul que b_0 est la moyenne du groupe contrôle, que b_1 est la différence des moyennes et que s est l'écart-type groupé des deux groupes.

Taille d'effet = $b_1/s = 0,766 / 1,97 = 0,401$. C'est le même résultat que le d de Cohen calculé à partir des moyennes posttest sans ajustement aux scores prétests.

On a également $F = 4,630$ qui est égal au F calculé par une ANOVA.

Modèle 3. Deux variables (traitement et contrôle) et pas de constante. Calcul de la taille d'effet non ajustée

Modèle : $y = b_1*t + b_2*c + e$ avec $t=1$ pour le groupe traitement et $t=0$ pour le groupe contrôle, $c=0$ pour le groupe traitement et $c=1$ pour le groupe contrôle.

5,62712	6,39286
0,24833	0,25490
0,90999	1,90749
571,22624	113,00000
4156,84625	411,15375

b1	6,393
b2	5,627
s1	0,248
s0	0,255
s	1,907
ddl	113,000
SC résidus	411,154

Moyenne groupe traitement

Moyenne du groupe contrôle

écart-type échantillon = écart-type groupé

Modèle calculé : $y = 6,393*t + 5,627*c + e$

On calcule également la taille d'effet non ajustée aux scores prétests.

Modèle 4. Variable traitement, variable scores prétest non centrés et constante. Calcul de la taille d'effet ajustée

Modèle : $y = b_0 + b_1*t + b_2*x + e$ avec $t=1$ pour le groupe traitement, $t=0$ pour le groupe contrôle et x le score prétest.

Résultats du tableur		
0,668	0,761	2,194
0,076	0,275	0,437
0,430	1,476	#N/A
42,168	112,000	#N/A
183,847	244,153	#N/A

b2	0,668
b1	0,761
b0	2,194
s2	0,076
s1	0,275
s0	0,437
R ²	0,430
s	1,476

Différence ajustée des moyennes

Pas d'interprétation

ne pas utiliser cette valeur pour le calcul de la taille d'effet

Modèle calculé : $y = 2,19 + 0,761*t + 0,668*x + e$

$R^2=0,43$ ou 43 %. Introduire la variable prétest améliore la qualité du modèle.

Taille d'effet ajustée = $b_1/s = 0,761 / 1,907 = 0,399$ avec $s=1,907$ (calculé dans le modèle sans la covariable score prétest).

	Moyennes		Différences (Trait-Contr)	Différence posttest ajustée
	Traitement	Contrôle		
posttest	6,393	5,627	0,766	0,761
prétest	5,143	5,136	0,007	

Modèle 5. Variable traitement, variable contrôle, scores prétest non-centrés et pas de constante.
Calcul de la taille d'effet ajustée

Modèle : $y = b_1*t + b_2*c + b_3*x + e$ avec $t=1$ pour le groupe traitement, $t=0$ pour le groupe contrôle, $c=0$ pour le groupe traitement et $c=1$ pour le groupe contrôle et x le score prétest.

0,66843	2,19434	2,95522
0,07637	0,43677	0,43953
0,94655	1,47646	#N/A
661,15650	112,00000	#N/A
4323,84660	244,15340	#N/A

b3	0,668
b2	2,194
b1	2,955
s2	0,076
s1	0,437
s0	0,440
s	1,476
ddl	112,000

Moyenne ajustée aux scores prétest du groupe contrôle
Moyenne ajustée aux scores prétest du groupe traitement

ne pas utiliser cette valeur pour le calcul de la taille d'effet

Modèle calculé : $y = 2,955*t + 2,194*c + 0,668*x + e$

Modèle 6. Variable traitement, scores prétest centrés et une constante. Calcul taille d'effet non ajustée.

Modèle : $y = b_0 + b_1*t + b_2*x + e$ avec $t=1$ pour le groupe traitement, $t=0$ pour le groupe contrôle et x le score prétest centré sur les moyennes de groupes (traitement ou contrôle).
On retrouve les mêmes résultats qu'avec un modèle sans scores prétest.

0,668	0,766	5,627
0,076	0,275	0,192
0,430	1,476	#N/A
42,168	112,000	#N/A
183,847	244,153	#N/A

b2	0,668
b1	0,766
b0	5,627
s2	0,076
s1	0,275
s0	0,192
R ²	0,430
s	1,476

Différence **NON** ajustée des moyennes
Moyenne du groupe contrôle

ne pas utiliser cette valeur pour le calcul de la taille d'effet

Modèle calculé : $y = 5,627 + 0,766*t + 0,668*x + e$

On calcule également une taille d'effet non ajustée aux scores prétests.

Annexe 4. Étude de l'intervention 1stClass@Number

L'intervention 1stClass@Number¹ (EEF, 2018) a fait l'objet d'une étude d'efficacité en conditions réelles conduite par EEF (Roques, à paraître). Consulter également *Education Endowment Foundation. L'enseignement fondé sur des preuves au Royaume-Uni* (Roques, 2022, www.mathadoc.fr) pour une présentation générale des études conduites par EEF.

Modèle linéaire

Les élèves ont passé deux tests, le Quantitative Reasoning Test (QRT) et le Key Stage 1 Mathématiques (KS1) qui font l'objet de deux analyses statistiques distinctes. Pour chaque test, le score posttest sera la variable dépendante. Les variables indépendantes sont le score prétest (effet fixe) et l'appartenance au groupe traitement ou non (effet fixe). Ce sont les écoles qui sont réparties aléatoirement dans le groupe contrôle ou le groupe traitement.

Ici i identifie l'élève et j son école. L'équation proposée est celle utilisée p.10 du plan d'analyse statistique de ce rapport (voir aussi l'annexe 6 du rapport).

$$Outcome\ measure_{ij} = \beta_0 + \beta_1 * Pre - test_{ij} + \beta_2 Group\ allocation_{ij} + u_{oj} School + e_{ij} \quad (1)$$

avec

- $Outcome\ measure_{ij}$: score posttest de l'élève (obtenu après avoir passé l'un des deux tests)
- $Group\ allocation_{ij} = 1$ pour les écoles du groupe traitement et 0 pour les écoles du groupe contrôle.
- β_1 : pente pour l'effet prétest (effet fixe, supposé ne pas varier entre les écoles)
- β_2 : pente pour l'appartenance au groupe intervention par rapport au groupe contrôle (au niveau de l'école, effet fixe). C'est la différence entre la moyenne conditionnelle du groupe traitement et la moyenne conditionnelle du groupe contrôle.
- $u_{oj} \sim N(0, \Omega_u)$
- $e_{ij} \sim N(0, \Omega_e)$

Remarque : Ω_u et Ω_e sont des variances (écarts-types élevés au carré). Les scores posttest sont normalement distribués autour de leur moyenne avec comme variance Ω telle que $\Omega = \Omega_u + \Omega_e$.

Les auteurs calculent une taille d'effet en utilisant la formule (non publiée)

$$ES = \frac{\beta_2}{\sqrt{\Omega_u + \Omega_e}}$$

Un intervalle de confiance de β_2 [$Inf\ \beta_2$; $Sup\ \beta_2$] est publié, qui permet de recalculer l'intervalle de confiance de ES en utilisant l'équation (non publiée)

$$\left[\frac{Inf\ \beta_2}{\sqrt{\Omega_u + \Omega_e}} ; \frac{Sup\ \beta_2}{\sqrt{\Omega_u + \Omega_e}} \right]$$

Résultats et recalculs pour les scores du Quantitative Reasoning Test (QRT)

Dans le résumé du rapport de l'intervention la taille d'effet présentée est $g = 0,18$ avec comme intervalle de confiance $[-0,08 ; 0,43]$ pour un niveau de confiance de 0,95. Cette taille d'effet

¹ Education Endowment Foundation (2018)

correspond selon EEF à 2 mois de progrès pour les élèves du groupe traitement. Un niveau de confiance (*security rating*) élevé¹ est associé à cette estimation.

Les résultats de la régression linéaire multiples sont donnés dans le tableau ci-dessous (tableau 1 p.131 du rapport) :

Paramètres	Estimation	Erreur standard	Intervalle de confiance à 95 %
β_0	6,36	0,52	[5,34 ; 7,38]
β_1	0,55	0,08	[0,39 ; 0,71]
β_2	0,68	0,49	[-0,29 ; 1,65]

Les valeurs suivantes sont également calculées par EEF (tableau 1 p.131) :

$$\Omega_u = 3,921 \text{ et } \Omega_e = 14,695.$$

Si on utilise les équations données ci-dessus, on recalcule alors *ES* et son intervalle de confiance, et on trouve :

$$ES = \frac{0,68}{\sqrt{3,92 + 14,695}} = 0,16$$

avec comme intervalle de confiance :

$$\left[\frac{-0,29}{\sqrt{3,921 + 14,695}} ; \frac{1,65}{\sqrt{3,921 + 14,695}} \right] = [-0,07 ; 0,38]$$

En utilisant uniquement la variance élève, on trouve

$$ES = \frac{0,68}{\sqrt{14,695}} = 0,18$$

avec comme intervalle de confiance

$$\left[\frac{-0,29}{\sqrt{14,695}} ; \frac{1,65}{\sqrt{14,695}} \right] = [-0,08 ; 0,43]$$

Ce sont ces résultats qui sont également publiés par EEF dans son tableau 3 p.131 : sur la première ligne sont présentés les résultats calculés en utilisant uniquement la variance élève et sur la seconde ligne les résultats calculés en utilisant la variance totale. C'est le second résultat qui est publié dans le résumé.

Deux analyses de sous-groupe sont également conduites. L'une concerne les élèves éligibles (ou non) aux repas scolaires gratuits (*FSM, Free School Meal*), la seconde le genre des élèves (fille ou garçon). Dans les deux cas, on suppose que l'effet du traitement est dépendant de l'appartenance à l'un de ces groupes. C'est pour cette raison qu'un terme d'interaction (le produit des deux

¹ Ce niveau est de 4 sur 5.

variables) est ajouté au modèle (voir ci-dessous pour une courte explication et Beckel, n.d.¹). Pour la première, l'équation de régression suivante est proposée

$$\begin{aligned} \text{Outcome}_{ij} &= \beta_0 + \beta_1 \text{PreTest}_{ij} + u_{0j} \text{School} + \beta_2 \text{Group allocation}_{ij} + \beta_3 \text{FSM}_{ij} \\ &+ \beta_4 \text{Group allocation} * \text{FSM}_{ij} + e_{ij} \end{aligned}$$

Aux indications déjà données ci-dessus, on rajoutera :

- β_3 : pente pour élève éligibles au FSM
- β_4 : pente d'interaction entre l'appartenance au groupe traitement versus contrôle et l'éligibilité au FSM.

Explication du terme d'interaction

Imaginons que l'on considère dans un premier temps que l'effet du traitement soit indépendant de l'appartenance au groupe FSM, on aurait eu le modèle suivant

$$\text{Outcome}_{ij} = \beta_0 + \beta_1 \text{PreTest}_{ij} + u_{0j} \text{School} + \beta_5 \text{Group allocation}_{ij} + \beta_3 \text{FSM}_{ij} + e_{ij}$$

Si l'effet du traitement β_5 est finalement considéré dépendant de la variable FSM, alors on a $\beta_5 = \beta_2 + \beta_4 \text{FSM}_{ij}$, et le modèle devient alors

$$\text{Outcome}_{ij} = \beta_0 + \beta_1 \text{PreTest}_{ij} + u_{0j} \text{School} + (\beta_2 + \beta_4 \text{FSM}_{ij}) * \text{Group allocation}_{ij} + \beta_3 \text{FSM}_{ij} + e_{ij}$$

$$\begin{aligned} \text{Outcome}_{ij} &= \beta_0 + \beta_1 \text{PreTest}_{ij} + u_{0j} \text{School} + \beta_2 \text{Group allocation}_{ij} + \beta_4 \text{FSM}_{ij} * \text{Group allocation}_{ij} \\ &+ \beta_3 \text{FSM}_{ij} + e_{ij} \end{aligned}$$

On retrouve bien le modèle utilisé par EEF.

¹ Bickel, J.-F. (n.d.). Les effets interaction
https://commonweb.unifr.ch/artsdean/pub/gestens/f/as/files/4665/9547_131825.pdf

Annexe 5. Utilisation d'un tableur

Régression linéaire multiple

On saisit dans un tableau les n valeurs observées de Y (première colonne) puis les $n \times p$ valeurs observées des p variables X_j (dans les p colonnes suivantes). Ensuite, on sélectionne une plage de cellules de 5 lignes et $p + 1$ colonnes, puis

1. saisir =DROITEREG($Y;X;VRAI;VRAI$)
2. CTRL + ↑ + « entrée » (pour obtenir des accolades)

avec X le tableau des $n \times p$ valeurs observées des variables X_j . Le premier VRAI indique la présence d'une constante (il faut choisir FAUX pour un modèle sans constante). Le second VRAI permet de déployer 6 résultats supplémentaires sous la ligne des estimations et la ligne des écarts-types (quand on choisit FAUX, seules les deux premières lignes apparaissent). Voir les exemples en annexe 2 et 3.

On obtient alors comme résultats

Ligne 1 : les $p + 1$ valeurs b_j (par ordre décroissant d'indice)

Ligne 2 : les $p + 1$ valeurs s_{β_j} (par ordre décroissant d'indice)

Ligne 3 : R^2 et s^2

Ligne 4 : F et ddl

Ligne 5 : $SC_{régression}$ et $SC_{résidus}$.

Dans les modèles avec constante

1. $SC_{résidus}$ est calculé comme la somme des écarts entre les valeurs prédites et les valeurs réelles élevés au carré
2. SC_{total} est calculé comme la somme des écarts entre les valeurs réelles et la moyenne de ces valeurs, élevés au carré.
3. $SC_{régression} = SC_{total} - SC_{résidus}$
4. $R^2 = SC_{régression} / SC_{total}$
5. $F = \frac{p}{\frac{SC_{résidus}}{ddl}}$ avec p le nombre de variables dépendantes.
6. $ddl = n - p - 1$

Dans les modèles sans constante

1. $SC_{résidus}$ est calculé comme la somme des écarts entre les valeurs prédites et les valeurs réelles élevés au carré
2. **SC_{total} est calculé comme la somme des écarts entre les valeurs réelles (on ne soustraie pas la moyenne) élevés au carré**
3. $SC_{régression} = SC_{total} - SC_{résidus}$
4. $R^2 = SC_{régression} / SC_{total}$
5. $F = \frac{p}{\frac{SC_{résidus}}{ddl}}$ avec p le nombre de variables dépendantes.
6. **$ddl = n - p$**

Cela explique les différences obtenues à partir du même jeu de données dans les résultats obtenus avec des modèles avec constante ou des modèles sans constante (annexe 2 et annexe 3).

Exemple : avec $p = 3$, on a trois variables X_1, X_2, X_3 . Les résultats sont donnés sous la forme d'un tableau de ce type

b_3	b_2	b_1	b_0
s_3	s_2	s_1	s_0
R^2	s	#	#
F	ddl	#	#
$SC_{régression}$	$SC_{résidus}$	#	#

Dans les modèles sans constante, les résultats de R^2 , F et $SC_{régression}$ ne peuvent pas être interprétés comme nous l'avons vu aux chapitres 3.2 et 3.3 de la partie 1.

On peut également utiliser l'utilitaire d'analyse (à activer dans les options d'Excel) en laissant décochée la case « intersection avec l'origine » pour un modèle avec constante (et en cochant cette case dans le cas inverse).

Coefficients de corrélations

Pour calculer les coefficients de corrélations entre deux variables, saisir dans Excel

=COEFFICIENT.CORRELATION(X_j ; X_k)

avec X_j la colonne des n valeurs de la variable j et X_k la colonne des n valeurs de la variable k .

Annexe 6. Estimation des variances pour les études avec randomisation au niveau des clusters

À partir des scores observés, on définit la variance intra (*within*) S_W^2 , la variance inter (*between*) S_B^2 et la variance totale S_T^2 de la manière suivante :

$$S_W^2 = \frac{\text{Somme des carrés des différences entre les scores et la moyenne de leur cluster}}{N - M}$$

$$S_B^2 = \frac{\text{Somme des carrés des différences entre les moyennes des clusters et la moyenne de leur groupe}}{M - 2}$$

$$S_T^2 = \frac{\text{Somme des carrés des différences entre les scores et la moyenne de leur groupe}}{N - 2}$$

Hedges (2007) propose les estimations suivantes

$$\text{Estimation de } \sigma_W^2 = S_W^2$$

$$\text{Estimation de } \sigma_B^2 = S_B^2 - \frac{S_W^2}{n}$$

$$\text{Estimation de } \sigma_T^2 = \text{estimation de } \sigma_W^2 + \text{estimation de } \sigma_B^2 = S_W^2 + S_B^2 - \frac{S_W^2}{n} = S_B^2 + \frac{n-1}{n} S_W^2 \quad (4)$$

Et par conséquent, une estimation de δ est le d de Cohen défini par

$$d = \frac{b}{\sqrt{S_B^2 + \frac{n-1}{n} S_W^2}}$$

Annexe 7. Glossaire

Caractéristique (nom) : élément permettant de préciser la description du sujet (pour les études primaires) ou des études primaires (pour les études secondaires)

Catégorielle (adjectif) : se dit d'une variable discrète pouvant prendre des valeurs réelles qui ne mesurent pas une grandeur

Cluster : regroupement de sujets en ensembles homogènes (classes ou établissements)

Continue (adjectif) : se dit d'une variable pouvant prendre toutes les valeurs d'un intervalle réel donné

Covariable : variable dont on souhaite contrôler les effets

Étude par comparaison de groupes : étude expérimentale qui évalue quantitativement l'effet d'une intervention (ou traitement) en comparant deux groupes d'élèves : le groupe traitement et le groupe contrôle (qui ne subit pas l'intervention)

Étude primaire : étude expérimentale ou quasi-expérimentale qui évalue l'effet d'une intervention (ou traitement) sur des élèves scolarisés. Une étude primaire peut être incluse dans une méta-analyse (qui est une étude secondaire)

Étude secondaire : analyse d'un ensemble d'études primaires collectées pour répondre à une question de recherche

Essai contrôlé randomisé : étude expérimentale par comparaison de groupes dans laquelle les groupes sont constitués de façon aléatoire, soit au niveau des élèves, soit au niveau de clusters

Facteur : caractéristique qualitative se traduisant par la répartition des sujets ou des études dans plusieurs catégories

Indicatrice (nom ou adjectif) : voir catégorielle (adjectif)

Intervention : programme scolaire ou ensemble de pratiques qualifiées souvent de méthode pédagogique ou produit commercial (comme une application informatique), dont l'objectif est d'améliorer les compétences des élèves.

Médiateur (nom) : caractéristique incluse dans la chaîne causale qui relie une intervention et l'évolution des compétences mesurées.

Médiatrice (adjectif) : se dit d'un médiateur (et donc d'une caractéristique) ou de la variable associée au médiateur.

Méta-analyse : Synthèse systématique d'études quantitatives indépendantes répondant à une même question de recherche qui se traduit notamment par le calcul d'indicateurs mesurant l'effet d'une intervention avec comme objectif principal d'aboutir à une conclusion plus fiable ou plus solide que celle pouvant être tirée d'études distinctes. Une méta-analyse est une étude secondaire rétrospective ; les études incluses dans la méta-analyse sont des études primaires prospectives.

Modérateur (nom) : caractéristique influençant l'intensité de la relation entre l'intervention et l'évolution des compétences mesurées.

Modératrice (adjectifs) : se dit d'un modérateur (et donc d'une caractéristique) ou de la variable associée au modérateur.

Muette (adjectif) : voir catégorielle

Qualitative (adjectif) : se dit d'une caractéristique traduite par des valeurs numériques pour lesquelles des opérations mathématiques (comme le calcul d'une somme, d'une moyenne, ...) n'ont pas de sens

Quantitative (adjectif) : se dit d'une caractéristique traduite par des valeurs numériques pour lesquelles des opérations mathématiques (comme le calcul d'une somme, d'une moyenne, ...) ont un sens

Taille d'effet : différence standardisée des moyennes ou encore écart entre les moyennes des scores des groupes traitement et contrôle divisé par l'estimation de l'écart-type des scores de la population

Variable : symbole que l'on peut remplacer par plusieurs valeurs numériques