t·p·é*

Une collection d'outils pour (s')initier à la publication en éducation et valoriser la transformation pédagogique

* transformer et publier en sciences de l'éducation

Ces outils sont organisés en trois volets :
→ accompagner / structurer pour les accompagnant·es
→ rédiger pour les équipes enseignantes débutant en recherche-action
→ s'enrichir / ressources une série de références et de guides

Ils ont été développés par la Chaire recherche-action sur l'innovation pédagogique de l'Université Paris Saclay et l'institut Villebon - *Georges Charpak,* en collaboration avec l'UQAM, et sont le fruit du travail de Marine Moyon, Frédéric Bouquet, Jeanne Parmentier, et Martin Riopel, et d'Emmanuel Ahr pour l'outil EVA.

L'exploration, la conception de la charte graphique et la mise en forme des outils ont été réalisées par Dalva Rospape et Marie Jouble.

Retrouvez tous les outils sur :
https://cep.villebon-charpak.fr/tpe

→ s'enrichir / ressources

stat 101

Ce document présente une sélection de diapositives issues d'une formation en statistiques appliquées aux sciences de l'éducation, consacrée plus particulièrement aux modèles de régression.

Introduction aux notions de statistiques

Vocabulaire

Variable qualitative versus quantitative

Mener une étude quantitative n'implique pas que toutes les variables soient nécessairement de nature quantitative. En effet, une étude quantitative peut inclure à la fois des variables quantitatives et qualitatives.
Cette distinction est essentielle, car la confusion entre le type d'étude et la nature des variables est une erreur fréquemment observée.

Variable qualitative

- On parle de variable qualitative nominale (ou catégorielle) lorsqu'on ne peut que qualifier les modalités que peut prendre la variable.
- Par exemple, dans le cadre d'un choix d'option que doit effectuer un étudiant, parmi les modalités de réponses suivantes : Astrophysique, Biologie Moléculaire ou Biologie Végétale
- On parle de variable qualitative ordinale lorsque les modalités qu'elle peut prendre présentent une relation d'ordre, ce qui permet de les classer selon un certain critère.

Par exemple, dans le cadre d'un renseignement concernant la fréquence de révision du cours de mathématiques, parmi les 5 modalités de réponses suivantes : Très souvent > Souvent > Parfois > Rarement > Jamais

Les données pourront être résumées et présentées sous forme de fréquence, pourcentage, effectif

Variable quantitative

- On parle de variable quantitative discrète lorsque la variable ne peut prendre qu'un certain nombre de valeurs bien précises. Il s'agit dans ce cas de compter.

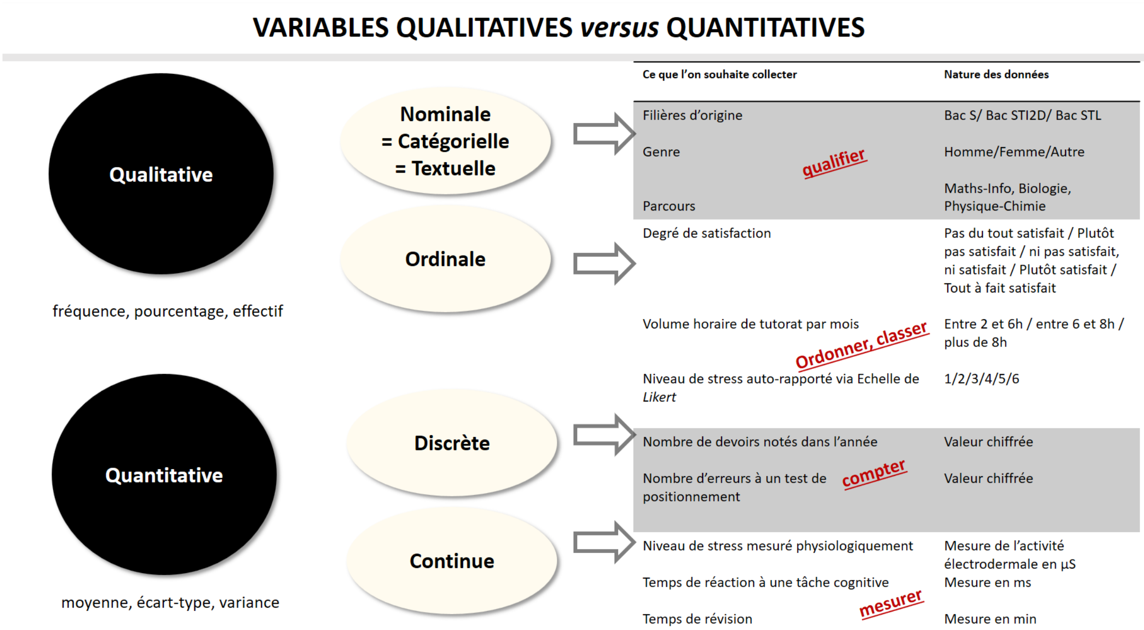
Notons qu'une variable nominale sous-jacente peut être identifiée.

Par exemple, dans le cas d'un dénombrement d'étudiants en situation de handicap, la variable qualitative sous-jacente correspond à la présence d'une situation de handicap, avec deux modalités de réponse : oui versus non.

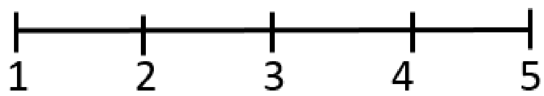
- On parle de variable quantitative continue lorsque la variable peut prendre une infinité de valeurs. Pour avoir accès à ces valeurs, il nous faut les mesurer.

Par exemple, le temps mis par un étudiant pour parcourir un document.
Les données pourront être résumées et présentées sous forme de moyenne, écart-type, variance

Les données pourront être résumées et présentées sous forme de moyenne, écart-type, variance.



Échelle de Likert



L'échelle de type Likert est une technique de mesures unidimensionnelles très populaire Il est demandé au participant de lire attentivement une série d'affirmations et pour chacune d'elle, de se positionner vis-à-vis de son contenu, en sélectionnant une modalité de réponse sur l'échelle.

Les échelles ne sont pas toutes de même nature :

Échelle d'accord : Pas du tout d'accord / en désaccord / ni d'accord, ni pas d'accord /d'accord/fortement d'accord

Échelle d'importance : Sans importance / peu important / modérément important / important / très important

Échelle de probabilité : Presque jamais vrai / habituellement pas vrai / parfois vrai / habituellement vrai / presque toujours vrai

Échelle de fréquence : Jamais / Très Rarement / Rarement / Parfois / Souvent / Très souvent / Toujours

Variable dépendante versus indépendante

VARIABLE DEPENDANTE (VD)

On appelle variable dépendante (VD), la variable mesurée par l'expérimentateur.
ex. score, temps de réponse, nombre d'erreurs commises, performance académique, …

Les VD sont des mesures qui, par hypothèse, sont susceptibles de dépendre du changement de modalité d'une ou plusieurs VI

VARIABLE INDEPENDANTE (VI)

VI = variable contrôlée et manipulée par l'expérimentateur.

- Provoquée (possibilité d'agir dessus) ex. degré de difficulté de la tâche ; familiarité de la tâche ; ludisme de la tâche ; heure de passation de la tâche ; température de la pièce...

- Invoquée (intrinsèque et donc non manipulable, simplement contrôlable) ex. âge ; genre...

Organisation du jeu de données

Structuration des données collectées dans le cadre d'une étude quantitative

La structuration des données est une étape cruciale, à ne surtout pas négliger. Il est important de soigner cette étape, qui conditionne l'analyse statistique subséquente.

→ Il est essentiel de toujours conserver le jeu de données brut intact et de travailler sur une copie afin de préserver l'intégrité des données originales.

→ Il est essentiel que toutes les données relatives à un même participant soient regroupées sur une unique ligne, au sein d'un même onglet, y compris s'il s'agit d'une étude comparative incluant des mesures préalables et postérieures à une intervention. Les données de pré et post-tests doivent être disponibles sur le même onglet.

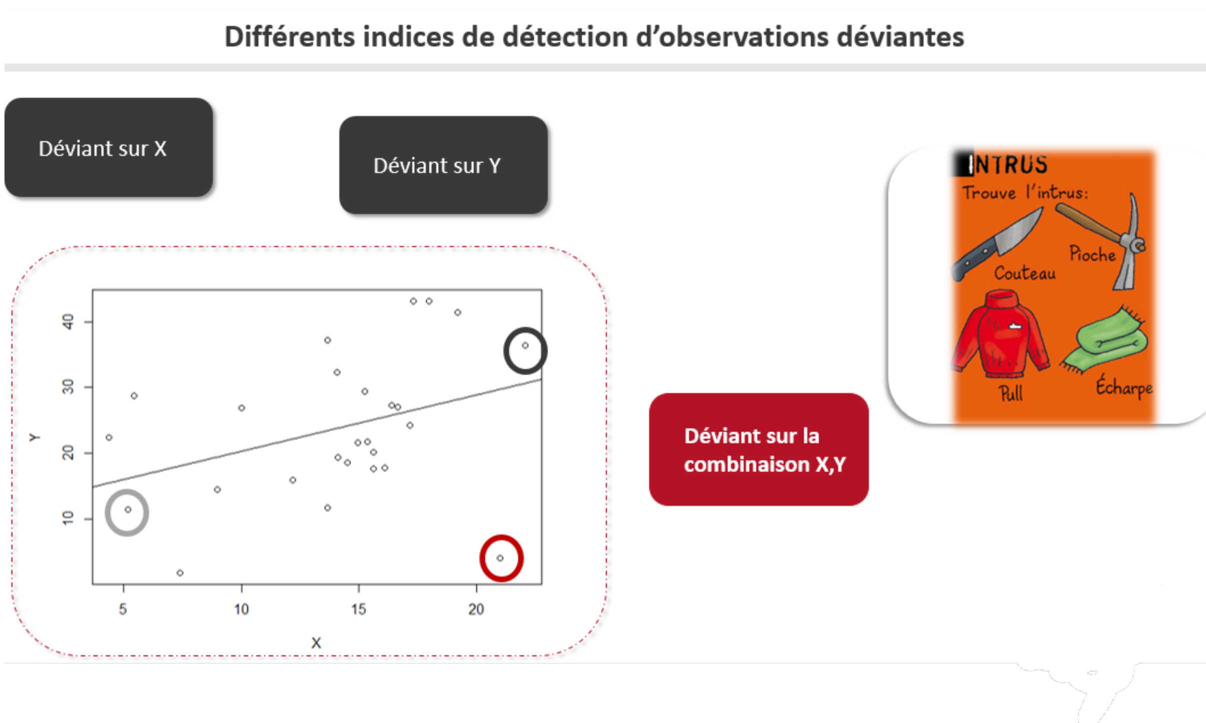
→ La première ligne de chaque colonne doit contenir un intitulé clair, concis, sans caractères spéciaux, accents ni espaces. La nomenclature utilisée pour nommer les colonnes doit suivre une logique uniforme (par exemple : Q1_DIM1_T0 ; Q2_DIM1_T0 ; Q1_DIM1_T1). Il est fortement déconseillé d'utiliser les énoncés complets des questions comme intitulés de colonnes.

Détection de données aberrantes

Il apparait assez aisé de comprendre que la suppression des données, n'est jamais à recommander.

Nettoyer les données des potentiels bruits, artefacts, valeurs aberrantes, constitue une étape particulière, nécessitant une précaution toute particulière.

Avant toute chose, la décision d'exclusion d'un participant d'une étude doit toujours pouvoir être justifiée. Il est déontologiquement et scientifiquement inacceptable de retirer un individu avec pour motif de satisfaire nos attentes en termes de résultat. Différents indices permettent de détecter des observations déviantes, par exemple par des tests de type levier, RSS et cookd.



Dans cet exemple, alors que le participant en gris et le participant en noir semblent de prime abord présenter des valeurs extrêmes, tous deux semblent contribuer à la droite résumant les données. Seul le participant en rouge semble atypique.

Analyses statistiques des données en éducation

Pour personnes débutantes

Après avoir rigoureusement sélectionné nos outils de mesures, collecté nos données, puis organisé minutieusement notre jeu de données, arrive le moment des analyses.

Avant toute chose, il est important de bien comprendre qu'une statistique de test correspond à un standard de publication de résultats scientifiques. Il s'agit d'un indicateur sur lequel nous allons nous appuyer afin de construire une règle décisionnelle. Il nous reviendra donc de décider s'il existe ou non une différence significative entre différentes conditions données. Il ne s'agira aucunement de prouver une réalité. Ainsi, toute décision s'accompagnera systématiquement d'une part d'incertitude.

Commençons par une expérience de pensée.

Soient 2 cohortes d'étudiants (i.e. deux échantillons) de première année universitaire. La première cohorte est constituée de n1 individus ; la seconde est constituée de n2 individus. Nous pouvons nous demander si ces deux cohortes d'étudiants présentent un niveau équivalent en sciences ou au contraire, si l'une de ces deux cohortes dispose d'un meilleur potentiel. Pour répondre à cette question, nous choisissons d'administrer à l'ensemble de ces étudiants, un test de positionnement en sciences, validé dans la littérature.

Deux possibilités de réponse sont alors envisageables, à savoir la présence ou l'absence de différence significative en termes de performances entre ces deux cohortes d'étudiants.

Au passage, notons que le mot « significatif », dans le langage courant véhicule l'idée de grandeur (ex. j'ai gagné une somme d'argent significative). Dans le domaine des statistiques, le mot « significatif » signifiera, « difficilement attribuable au hasard ». **Ainsi, une différence significative sera une différence difficilement attribuable au hasard.**

Notre intuition est la suivante :

→ Si la différence observée en termes de performance est très faible, alors nous pouvons penser que les deux échantillons présentent le même potentiel.

→ Si la différence observée en termes de performance est forte, alors les deux échantillons présentent certainement des potentiels différents.

Tout l'enjeu consiste à savoir à quel endroit positionner la limite.

Le recours aux tests statistiques permet une formalisation de notre intuition.

En réalité, lorsque nous réalisons un test statistique, nous ne commençons pas par fixer de limite. Nous commençons par nous fixer une probabilité de nous tromper. On parle de risque alpha. Dans notre cas, le risque alpha correspond à la probabilité de **prendre la décision qu'il existe une différence significative** de potentiel entre nos deux cohortes d'étudiants, **alors qu'en réalité, il n'en existe pas**.

L'objectif de la statistique de test classique est de **maîtriser la probabilité de se tromper**, autrement dit le risque alpha, que l'on fixe généralement à 5 %. Lorsqu'une statistique de test est supérieure à 1,96 en valeur absolue, **la probabilité qu'un tel résultat soit imputable au hasard devient très faible.** Bien que ce ne soit pas impossible, cela reste peu vraisemblable. L'apparition d'un tel événement – alors qu'en réalité, il n'existe pas de différence significative – est donc associé à une faible probabilité.

De façon conventionnelle, nous fixons ce risque à 5%. Il s'agit là d'un standard de publication.

Ainsi, comparer la statistique de test à 1.96 nous permet de prendre une décision.

Toutefois, dans notre pratique quotidienne, nous avons tendance à appréhender la question dans le sens inverse. En effet, bien qu'il soit tout à fait opérationnel de comparer la statistique de test à la valeur seuil de 1.96, nous avons plutôt tendance à nous baser sur une donnée renvoyée par tout logiciel statistique et que nous nommons **probabilité critique** (ou **p-value**). Cette dernière devient alors le critère principal d'aide à la décision.

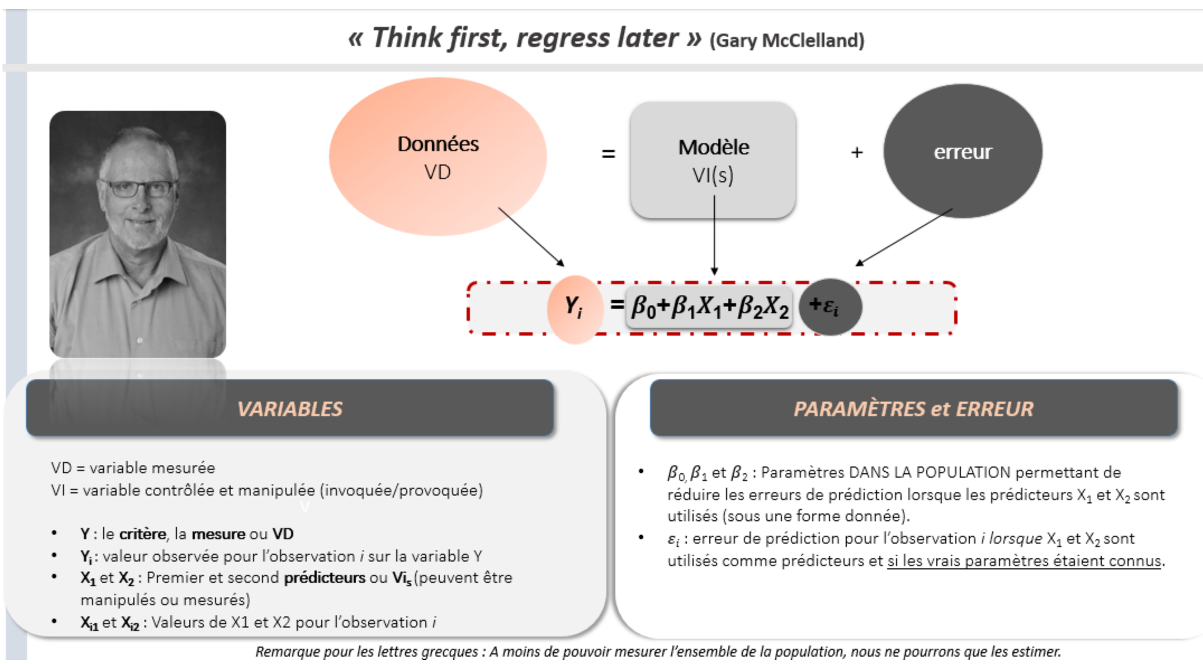
La **probabilité critique**, notée, p, correspond à la probabilité **d'observer ces données (donc d'observer une différence au moins aussi grande), sous le seul effet du hasard**, alors qu'en réalité **il n'existe pas de différence significative.**

La règle de décision est alors la suivante : si la probabilité critique est inférieure à 5%, nous **déciderons** qu'il existe une différence significative entre les conditions comparées (avec un risque alpha à 5% de se tromper). Au contraire, si la probabilité critique est supérieure à 5% nous déciderons qu'il n'existe pas de différence significative entre les conditions comparées.

Cette méthodologie apparaît avantageuse, puisque le degré de significativité p nous donne accès à une quantification de l'incompatibilité de l'observation avec l'hypothèse selon laquelle il n'existe pas de différence significative entre les conditions comparées.

Présentation générale du modèle de régression

- Lorsqu'on évoque des analyses telles que l'ANOVA ou l'ANCOVA, il est important de comprendre que celles-ci reposent sur des modèles statistiques sous-jacents
- Face à la diversité des tests statistiques, le choix de l'analyse la plus appropriée dépend directement de la question de recherche.
- Dans le domaine des sciences humaines et sociales, notons que, dans la grande majorité des cas, la question de recherche peut être traduite dans un langage statistique commun, à savoir, la régression

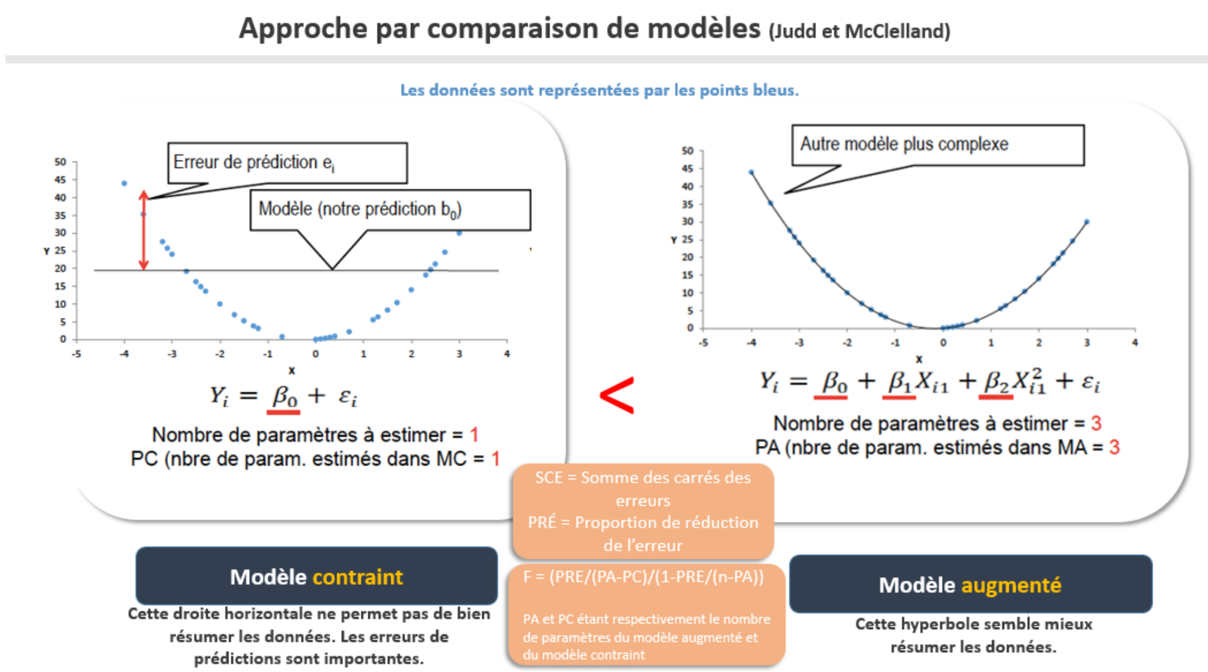


L'objectif est **de résumer et d'expliquer les données (Y) à l'aide d'un modèle parcimonieux**, c'est-à-dire en utilisant le nombre **minimal** de prédicteurs (Xi) nécessaires. Par exemple, pour une cohorte de participants donnée, Y pourrait correspondre à une performance académique, X1 au temps de révision et X2 au niveau de stress auto-rapporté 5 min avant l'épreuve.

Pour cela, il va être nécessaire de coder la relation qu'entretiennent X1 et X2 avec Y, à l'aide d'une série de paramètres, venant pondérer les prédicteurs. Un paramètre (β) doit être appréhendé non pas comme une corrélation, mais plutôt comme une pente liant le prédicteur (X) et la mesure (Y).

Il ne sera pas possible de connaître la valeur réelle des paramètres pour l'ensemble de la population. On ne pourra qu'estimer leur valeur à partir d'un échantillon. C'est pourquoi, nous utiliserons la notation avec des lettres grecques pour l'équation de régression.

L'erreur de prédiction, ε_i correspond à l'erreur liée à la mesure, aux variables contextuelles (e.g. effet de la chronobiologie, état de fatigue, niveau de perturbation sonore). Cette erreur est quantifiable. Les tests les plus couramment utilisés reposent sur une somme des carrés des erreurs.



Dans l'approche par comparaison de modèle, il s'agit de se demander si le fait d'ajouter des paramètres permet d'améliorer le modèle. Un modèle avec peu de paramètre (modèle contraint) sera comparé à un modèle comprenant davantage de paramètres (modèle augmenté). Ainsi, il s'agit de se demander si le modèle augmenté fait mieux que le modèle contraint.

Il s'agit de quantifier l'erreur associée à chaque modèle (un modèle associé à peu d'erreur étant un bon modèle). Il faudra comparer les modèles en confrontant la somme des carrés des erreurs (SCE) du modèle contraint à la SCE du modèle augmenté. On va estimer la proportion de réduction de l'erreur (PRE) lorsque l'on passe d'un modèle contraint à un modèle augmenté. On va regarder le pourcentage de réduction de l'erreur. La question sera alors de savoir si on est satisfait de la réduction de l'erreur.

Dans l'approche par **comparaison de modèles**, l'objectif est d'évaluer si **l'ajout de paramètres conduit ou non à une amélioration** du modèle. Un modèle contraint (avec peu de paramètres) est comparé à un modèle augmenté (intégrant des paramètres supplémentaires) pour déterminer si ce dernier offre une meilleure explication des données.

Cette évaluation repose sur la comparaison de la somme des carrés des erreurs (SCE) associée à chaque modèle ; une SCE plus faible indiquant un meilleur ajustement du modèle aux données.

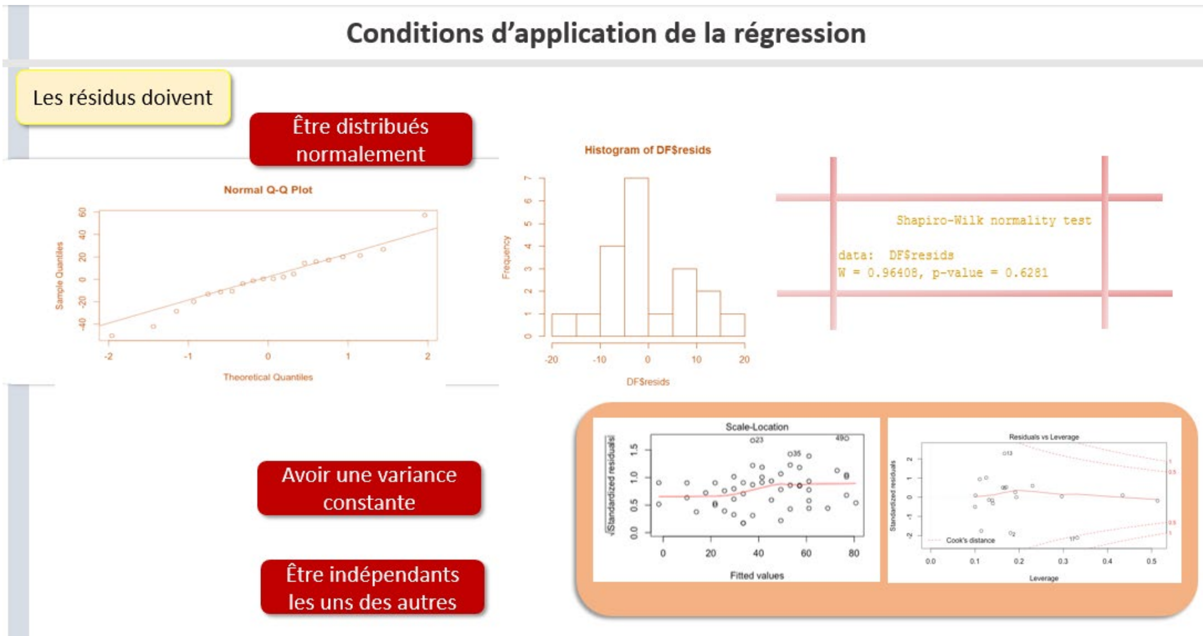
Sera alors estimée la proportion de réduction de l'erreur de prédiction (PRE), correspondant au pourcentage de diminution de la SCE lors du passage du modèle contraint au modèle augmenté. Une PRE élevée suggère que l'ajout de paramètres a significativement amélioré le modèle.

La question centrale est alors de déterminer si cette réduction de l'erreur justifie la complexité accrue du modèle.

Selon la logique habituelle, nous partirons du principe que le modèle contraint est vrai (H0 vrai). Puis nous nous demanderons si le fait d'observer cette proportion de réduction de l'erreur est probable sachant que H0 est vraie. (plus le PRE sera faible et plus nous aurons tendance à décider de conserver le modèle contraint). Nous dirons alors que nous invalidons le modèle augmenté (ce qui ne veut pas dire que nous validons le modèle contraint).

Rares sont les ouvrages de statistiques qui contiennent des tables de valeurs critiques de PRE. Les tables de F sont beaucoup plus communes.
Nous calculerons alors le F afin :

i. d'examiner la proportion de réduction de l'erreur par paramètre additionnel ajouté dans le modèle (dit autrement, il s'agira de savoir dans quelle mesure le paramètre utilisé permet de davantage réduire l'erreur, en comparaison à tout autre paramètre que nous aurions pu sélectionner) ii.
ii. de comparer la proportion d'erreur qui a été réduite (PRE) avec la proportion d'erreur restante (1-PRE).



Une régression ne peut être appliquée que si les trois conditions d'application suivantes sont respectées :

i. La distribution des résidus (i.e. des erreurs) doit être normale. Pour vérifier visuellement ce critère, il est possible de comparer la distribution de nos résidus avec une droite représentant une distribution normale théorique (diagramme des quantiles-quantiles ou Q-Q plot), et de s'assurer de l'absence d'observations extrêmes. Il est également possible d'afficher l'histogramme des résidus et de s'assurer que la distribution suit bien une gaussienne. Par ailleurs, un test de normalité (Shapiro-Wilk) pourra être conduit pour s'assurer d'une distribution normale des résidus.

Dans le cas où le test de normalité renvoie une p-value $<.05$, il sera possible de se rabattre sur des tests non-paramétriques ou alors de tenter de rendre la distribution normale (par exemple, en appliquant une fonction logarithmique, inverse ou puissance)

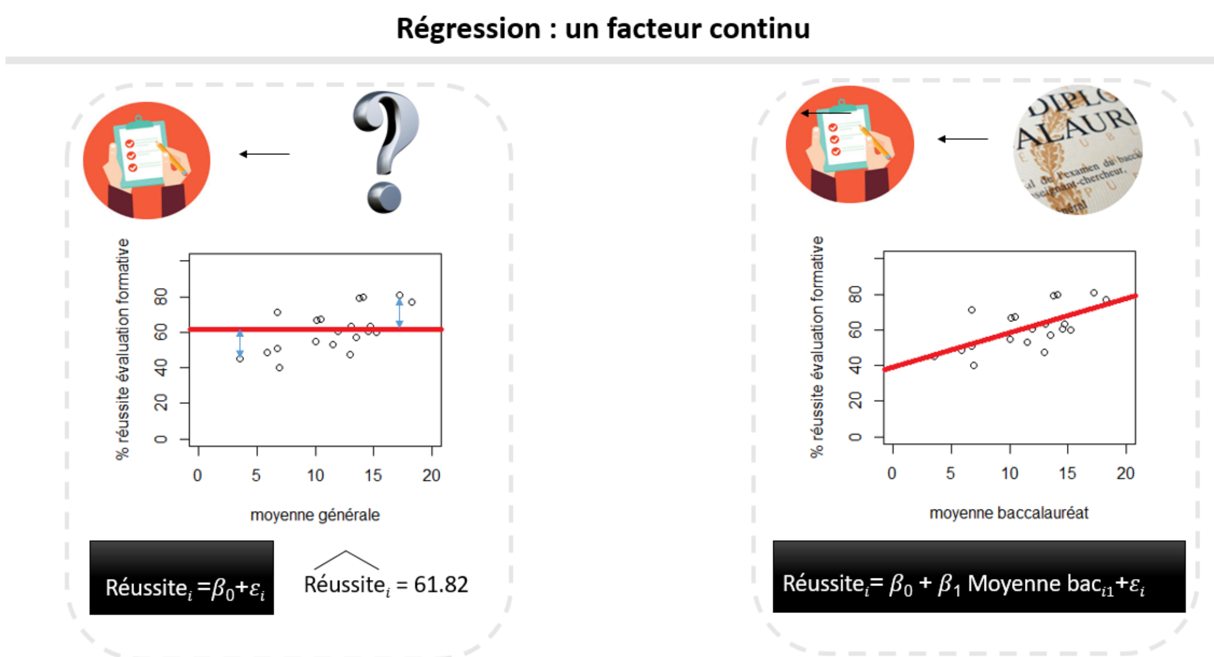
ii. La variance des résidus doit également être constante. On parle d'homoscédasticité. Test de Levene (ou encore le Test de Bartlett)

iii. Les résidus doivent être indépendants les uns des autres.

Modèles de régression, simple et multiple

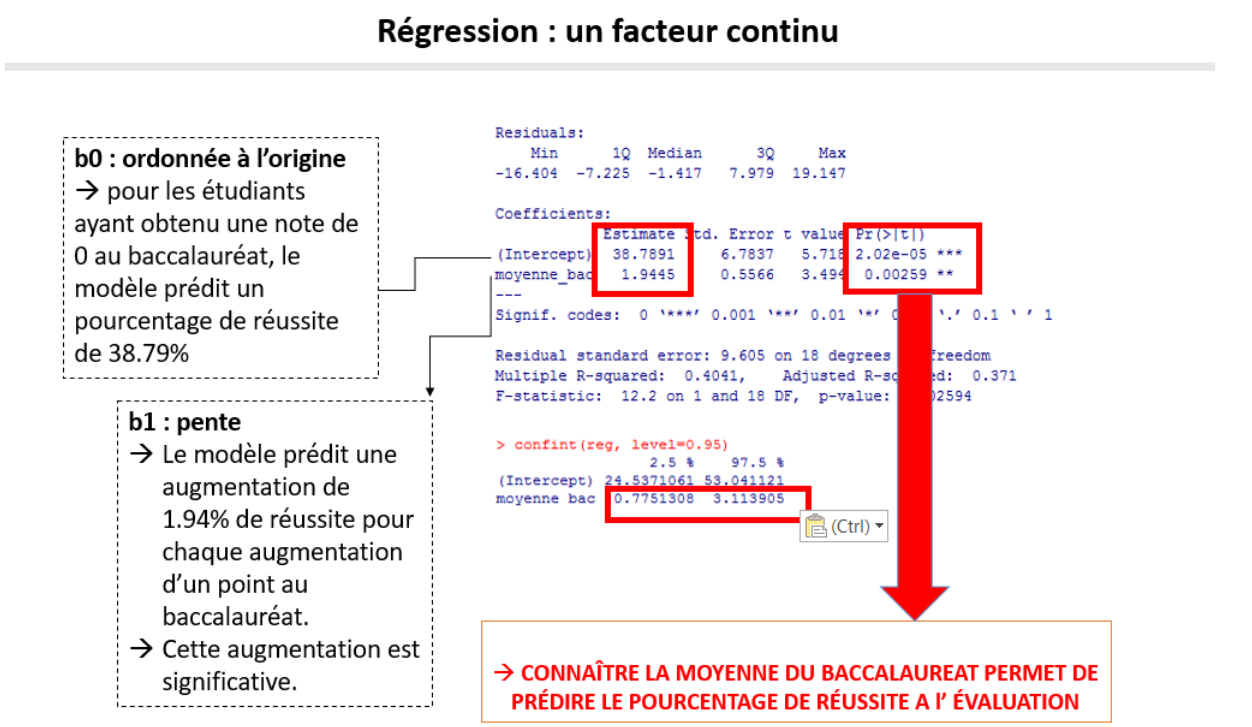
Régression simple : VI continue

Situation 1a :
→ quinze participants sont soumis à une évaluation formative. Nous allons tenter de modéliser leurs performances. Pour cela, nous allons chercher à proposer une prédiction qui produise la plus petite somme des carrés des erreurs.



Concernant le modèle contraint, nous proposons de prédire, pour chaque participant, son pourcentage de réussite à l'évaluation à partir de la moyenne des pourcentages de réussite à l'évaluation obtenue par tous les élèves.
→ Ce modèle apparait non pertinent. La droite présente un coefficient directeur nul, et ne représente pas convenablement notre échantillon.

Concernant le modèle augmenté, nous nous demandons si la moyenne obtenue au baccalauréat par chacun des élèves permettrait de mieux prédire le pourcentage de réussite à l'évaluation formative.
→ La courbe de tendance semble indiquer l'existence d'un lien entre le pourcentage de réussite à l'évaluation et la moyenne individuelle obtenue aux différentes épreuves du baccalauréat. Visuellement, cette variable semble expliquer nos données.



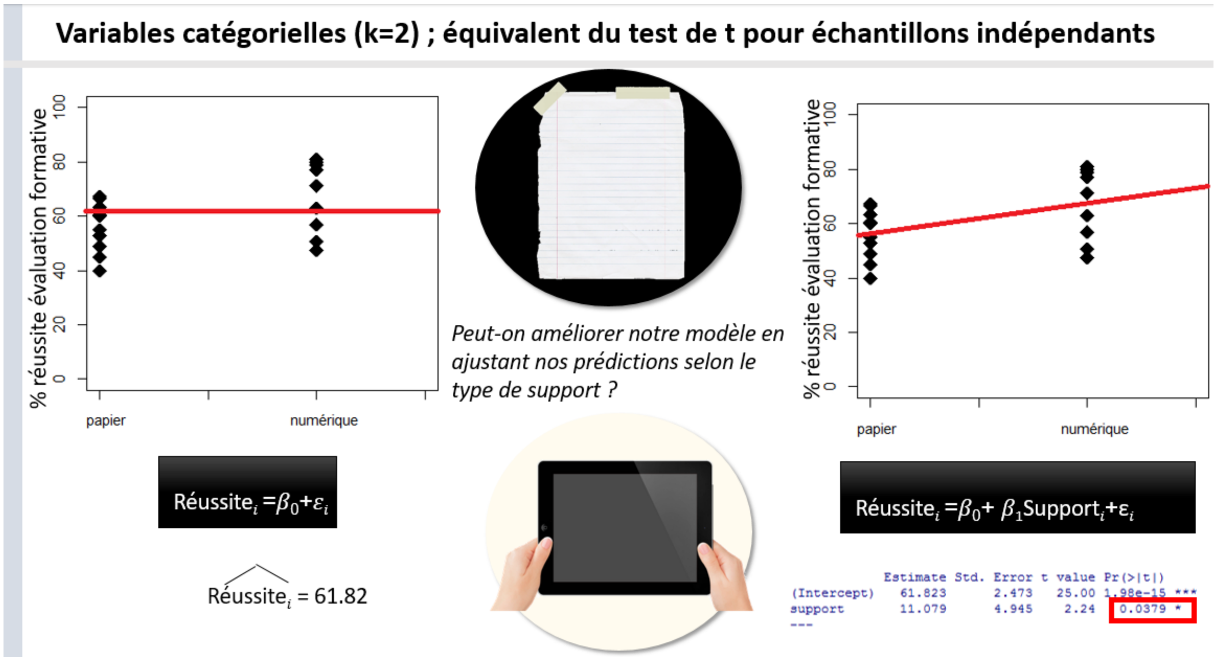
Ajouter la moyenne obtenue au baccalauréat permet de réduire l'erreur associée au modèle. Observer ce pourcentage de réduction d'erreur, alors que H0 est vrai, apparaît très peu probable ; nous choisissons de rejeter le modèle contraint.

Régression simple : VI catégorielle

Situation 1b :

→ nous repartons de la situation 1a. Nous nous intéressons toujours à la même mesure, à savoir le % de réussite à l'évaluation formative. La moyenne de l'échantillon reste inchangée.

Parmi les participants, la moitié a réalisé l'évaluation au format papier et l'autre moitié a réalisé l'évaluation au format numérique. Nous pouvons nous demander si la condition de passation (le support) pourrait permettre de prédire le pourcentage de réussite à l'évaluation formative.



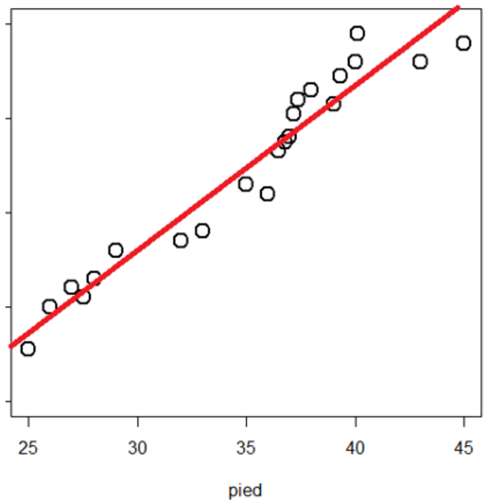
Lorsque l'on passe de la condition support papier à support numérique, le % de réussite à l'évaluation augmente de 11.08% et ce, de façon significative. La moyenne à l'évaluation est de 56% pour la condition papier et de 67% pour la condition numérique. La droite associée au modèle augmenté, passe par la moyenne de chacune des deux conditions.

Régression multiple : deux VI continues

Situation 2 :

→ nous cherchons à prédire le niveau d'anglais de nos élèves. Nous nous demandons si ce niveau pourrait être prédit par la taille de leurs pieds.

Régression multiple : variables continues



$$\text{Niveau anglais}_i = \beta_0 + \beta_1 \text{Pied}_i + \varepsilon_i$$

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -73.0915     7.1159  -10.27 2.01e-09 ***
pied          3.5009     0.2014   17.39 1.54e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.246 on 20 degrees of freedom
Multiple R-squared:  0.9379,    Adjusted R-squared:  0.9348
F-statistic: 302.3 on 1 and 20 DF,  p-value: 1.537e-13
```

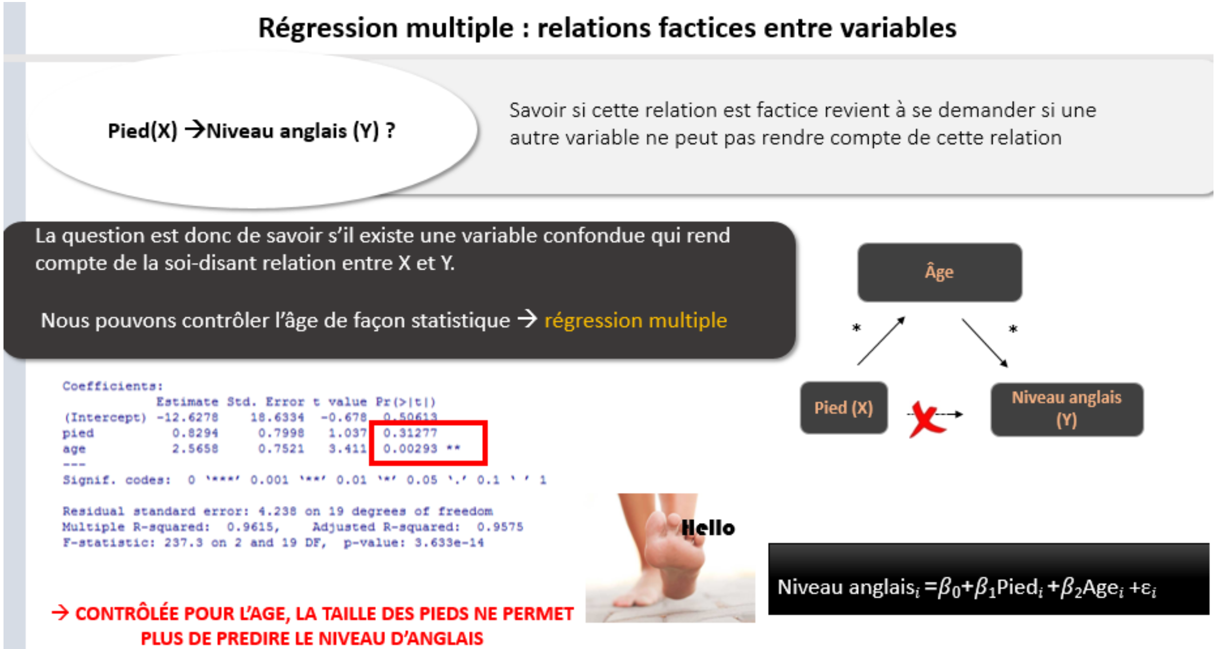
La taille des pieds permet de prédire le niveau d'anglais

Peut-on affirmer que :
Pied(X) → Niveau anglais (Y) ?

Nous nous apercevons alors qu'il existe un lien entre ces deux variables. Nous retrouvons une relation linéaire et significative.

> Intercept → performance d'anglais que nous allons prédire pour les étudiants qui ont une taille de pied de 0 (donc personne)
>
> Pied → à chaque fois que nous augmentons d'une unité de taille de pied, la performance d'anglais augmente de 3.5 points.

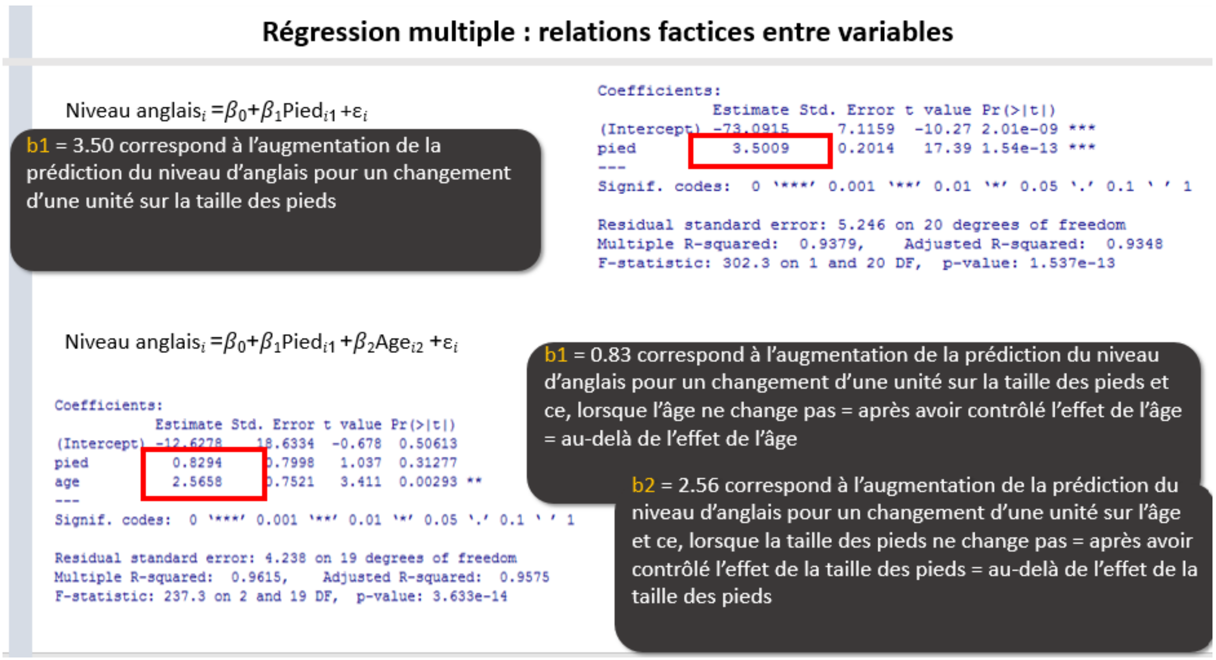
Nous nous demandons toutefois si cette relation ne pourrait pas s'expliquer par une nouvelle variable X2. Autrement dit, nous nous demandons s'il existe une variable confondue qui pourrait rendre compte de la soi-disant relation entre la variable X1 (taille des pieds) et la variable Y (niveau d'anglais). Nous choisissons d'introduire dans le modèle, la variable (X2) âge.



Nous nous apercevons alors qu'une fois l'effet de l'âge contrôlé, la taille des pieds ne permet plus de prédire le niveau d'anglais. La relation observée précédemment était donc factice.

L'ajout de la variable âge dans le modèle supprime la part de variance commune qui existait entre le niveau d'anglais et la taille des pieds. Ainsi, une fois l'effet de l'âge contrôlé, la taille des pieds ne permet plus de prédire le niveau d'anglais. La relation observée précédemment était donc factice.

Ainsi savoir si une relation est factice revient à se demander si une autre variable ne pourrait pas rendre compte de cette relation.



Intercept → pour des individus qui ont 0 en âge et 0 en taille de pieds, nous pouvons prédire une performance d'anglais de -12

Pied → Pour un changement d'une unité sur la taille de pied, la performance en anglais augmente de 0.83 point, pour un niveau d'âge contrôlé, en maintenant l'âge constant, au-delà de l'effet de l'âge.

Age → Pour toute augmentation d'une année sur l'âge, la performance d'anglais augmente de 2.56 pt, au-delà de l'effet de la taille des pieds.

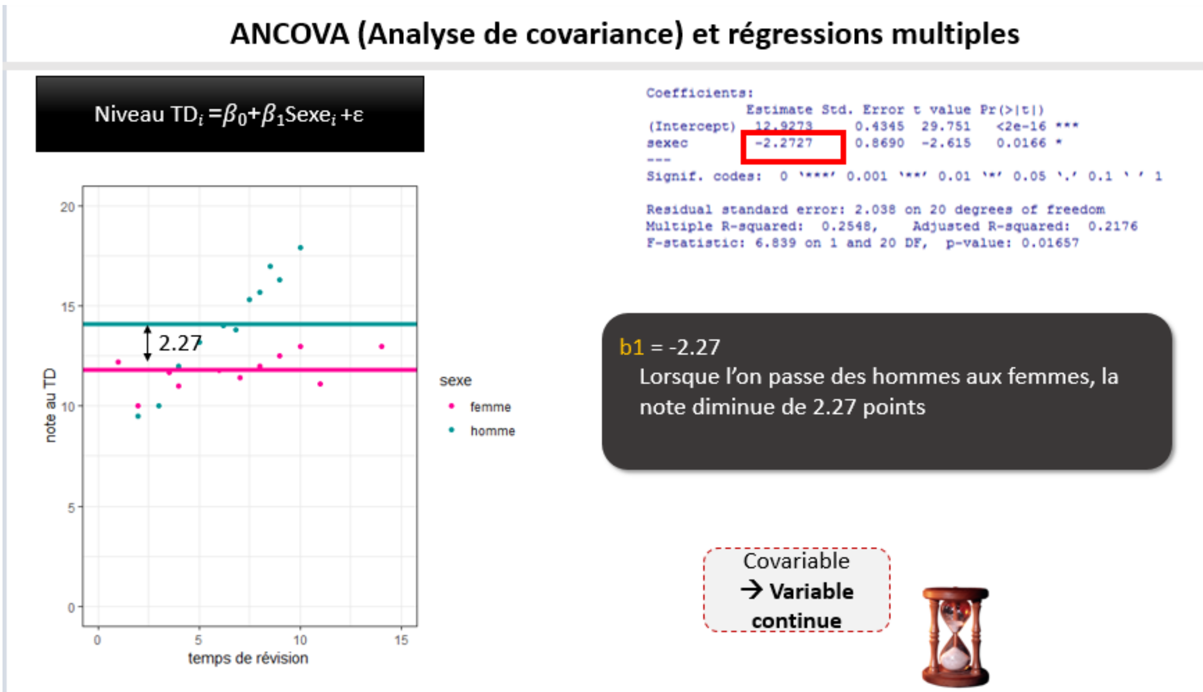
Régression multiple : une VI continue et une VI catégorielle

Situation 3 :

→ des étudiants inscrits en 3ère année universitaire obtiennent en fin de semestre leur note moyenne au TD de pharmacologie.

Nous nous demandons s'il existe un effet du sexe sur cette note moyenne de TD ?

Nous décidons alors d'appliquer un modèle de régression simple, avec la variable catégorielle sexe comme prédicteur.



Intercept → ici l'intercept va correspondre à la moyenne générale hommes et femmes (du fait du codage centré). Il s'agit donc de la valeur moyenne de tous les individus.

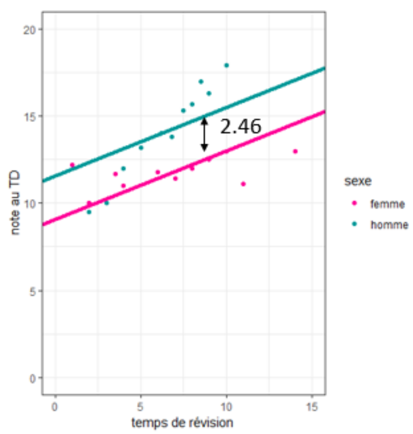
Nous observons que l'effet est significatif, avec les hommes performant mieux que les femmes.

Nous nous demandons si une deuxième variable ne permettrait pas d'expliquer mieux le modèle.

Cette fois, nous allons intégrer au modèle, une variable continue, à savoir le temps d'apprentissage personnel.

Ajustement de l'ANCOVA

Niveau $TD_i = \beta_0 + \beta_1 Sexe_i + \beta_2$ temps apprentissage$_i$ + ε_i




```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         10.3291     0.7713  13.391 3.98e-11 ***
sexec               -2.4692     0.6781  -3.641  0.00174 **
temps_apprentissage  0.3928     0.1048   3.747  0.00136 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.586 on 19 degrees of freedom
Multiple R-squared:  0.5715,    Adjusted R-squared:  0.5264
F-statistic: 12.67 on 2 and 19 DF,  p-value: 0.0003187
```

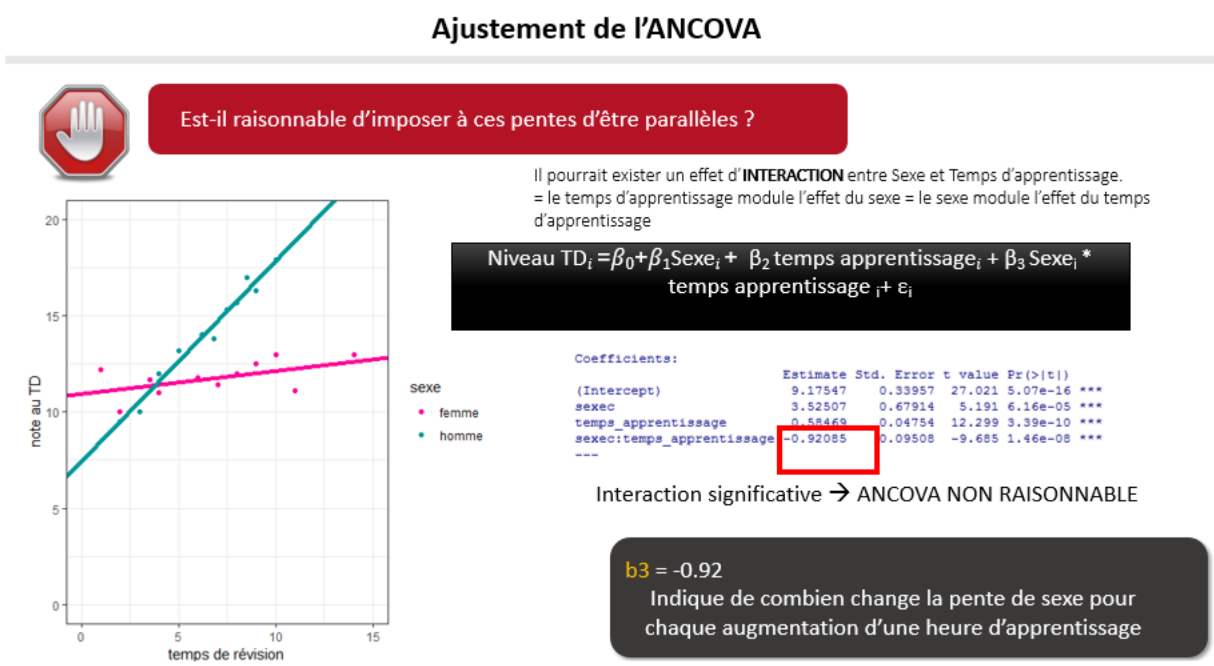
b1 = -2.47
Lorsque l'on passe des hommes aux femmes, la note de TD diminue de 2.47 points et ce, pour un temps d'apprentissage constant

b2 = 0.39
Lorsque l'on augmente le temps de révision d'une heure, la note de TD augmente de 0.39, indépendamment du sexe

Un effet significatif est retrouvé pour chacune des deux variables. Ces deux variables apparaissent donc importantes à considérer. Ici, il n'est pas question de variable confondante ; l'effet du sexe demeurant significatif.

> Intercept → Lorsque l'on passe des hommes aux femmes la note de TD diminue de 2.47 points et ce pour un temps d'apprentissage constant.
>
> Tps d'apprentissage → lorsque nous augmentons d'une heure le temps d'apprentissage, la note de TD augmente de 0.39, indépendamment du sexe de la personne apprenante

Toutefois attention, jusqu'à présent, nous avons un postulat implicite assez fort, qui consiste à dire que les pentes ne vont pas varier, que l'effet du temps d'apprentissage sur la note de TD ne va pas varier en fonction du sexe. Mais est-ce vraiment justifié et raisonnable d'imposer à ces pentes d'être parallèles ? Un modèle dans lequel la pente du temps de révision pourrait varier dépendamment du sexe ne permettrait-il pas de mieux décrire nos données ? En d'autres termes, n'existe-t-il pas un effet d'interaction entre le sexe et le temps de révision ? Cela revient à tester le produit de deux facteurs.



Dans ce cas, l'ANCOVA correspond au modèle contraint et l'interaction au modèle augmenté.
Notons que dans le cas de l'ANCOVA, il est toujours nécessaire de vérifier que l'interaction ne soit pas significative dans le modèle. Si l'interaction est significative, alors l'ANCOVA n'est pas raisonnable.

Ici l'interaction est significative, c'est-à-dire que le temps de révision modère (ou module) l'effet du sexe sur la note, mais également que l'effet du sexe modère (ou module) le temps de révision. Cet effet d'interaction (ou modération, ou modulation) est bidirectionnel.

> Intercept → il s'agit d'une prédiction pour les individus qui ont passé 0h à réviser (X=0, Z= 0, X*Z=0)
>
> sexec → correspond à l'effet du sexe, mais UNIQUEMENT lorsque le temps d'apprentissage est de 0h.
>
> Temps d'apprentissage → correspond à la pente de temps d'apprentissage
>
> sexec:temps → indique de combien change la pente (l'effet) de sexe pour chaque augmentation d'1h d'apprentissage.

Effets simples

```
Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                   7.41293    0.54604  13.576 6.76e-11 ***
sexedH                        3.52507    0.67914   5.191 6.16e-05 ***
temps_apprentissage           1.04511    0.08002  13.060 1.28e-10 ***
sexedH:temps_apprentissage -0.92085    0.09508  -9.685 1.46e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6537 on 18 degrees of freedom
Multiple R-squared:  0.931,	Adjusted R-squared:  0.9195
F-statistic: 80.97 on 3 and 18 DF,  p-value: 1.208e-10
```

Homme

```
Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                  10.93801    0.40381  27.087 4.85e-16 ***
sexedF                        3.52507    0.67914   5.191 6.16e-05 ***
temps_apprentissage           0.12426    0.05135   2.420   0.0263 *
sexedF:temps_apprentissage -0.92085    0.09508  -9.685 1.46e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6537 on 18 degrees of freedom
Multiple R-squared:  0.931,	Adjusted R-squared:  0.9195
F-statistic: 80.97 on 3 and 18 DF,  p-value: 1.208e-10
```

Femme

b2 = effet simple (pente) du temps d'apprentissage pour les hommes / pour les femmes uniquement

Il est également possible de s'intéresser aux effets simples et de regarder l'effet du temps d'apprentissage chez les hommes uniquement ; ou chez les femmes uniquement.

```
Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                   7.41293    0.54604  13.576 6.76e-11 ***
sexedH                        3.52507    0.67914   5.191 6.16e-05 ***
temps_apprentissage           1.04511    0.08002  13.060 1.28e-10 ***
sexedH:temps_apprentissage -0.92085    0.09508  -9.685 1.46e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6537 on 18 degrees of freedom
Multiple R-squared:  0.931,	Adjusted R-squared:  0.9195
F-statistic: 80.97 on 3 and 18 DF,  p-value: 1.208e-10
```

Intercept → Il s'agit d'une prédiction pour les individus du sexe masculin et qui ont passé 0h à réviser

sexedH → correspond à l'effet du sexe, mais UNIQUEMENT lorsque le temps d'apprentissage est de 0h.

temps d'apprentissage → correspond à la pente de temps d'apprentissage pour les hommes. Donc qd les hommes augmentent d'une heure leur temps d'apprentissage, la note augmente de 1.05

sexedH:temps → indique de combien change la pente (l'effet) de sexe pour chaque augmentation d'1h d'apprentissage.

```
Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                  10.93801    0.40381  27.087 4.85e-16 ***
sexedF                        3.52507    0.67914   5.191 6.16e-05 ***
temps_apprentissage           0.12426    0.05135   2.420   0.0263 *
sexedF:temps_apprentissage -0.92085    0.09508  -9.685 1.46e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6537 on 18 degrees of freedom
Multiple R-squared:  0.931,	Adjusted R-squared:  0.9195
F-statistic: 80.97 on 3 and 18 DF,  p-value: 1.208e-10
```

Intercept → Il s'agit d'une prédiction pour les individus du sexe féminin et qui ont passé 0h à réviser

sexedF → correspond à l'effet du sexe, mais UNIQUEMENT lorsque le temps d'apprentissage est de 0h.

temps d'apprentissage → correspond à la pente de temps d'apprentissage pour les femmes. Donc qd les femmes augmentent d'une heure leur temps d'apprentissage, la note augmente de 0.12

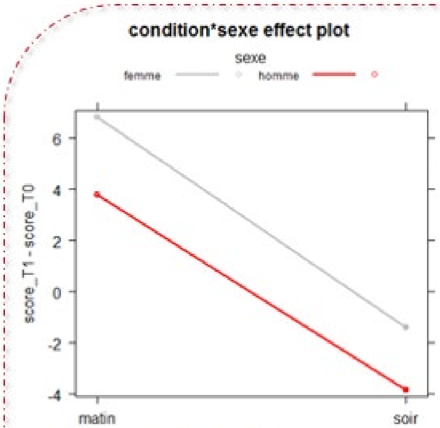
sexedF:temps → indique de combien change la pente (l'effet) de sexe pour chaque augmentation d'1h d'apprentissage.

Remarque : la moyenne de 0.12 et 1.05, c'est bien 0.58.

Régression multiple : deux VI catégorielles

De manière similaire, il est possible de proposer un modèle de régression multiple comprenant deux prédicteurs catégoriels.

Par exemple, vous pourriez vous demander si le progrès en termes de performance à un test psychométrique (score T1 – score T0) pourrait être prédit en fonction du moment de passation du test lors du T1 (matin versus soir) et en fonction du sexe du participant.



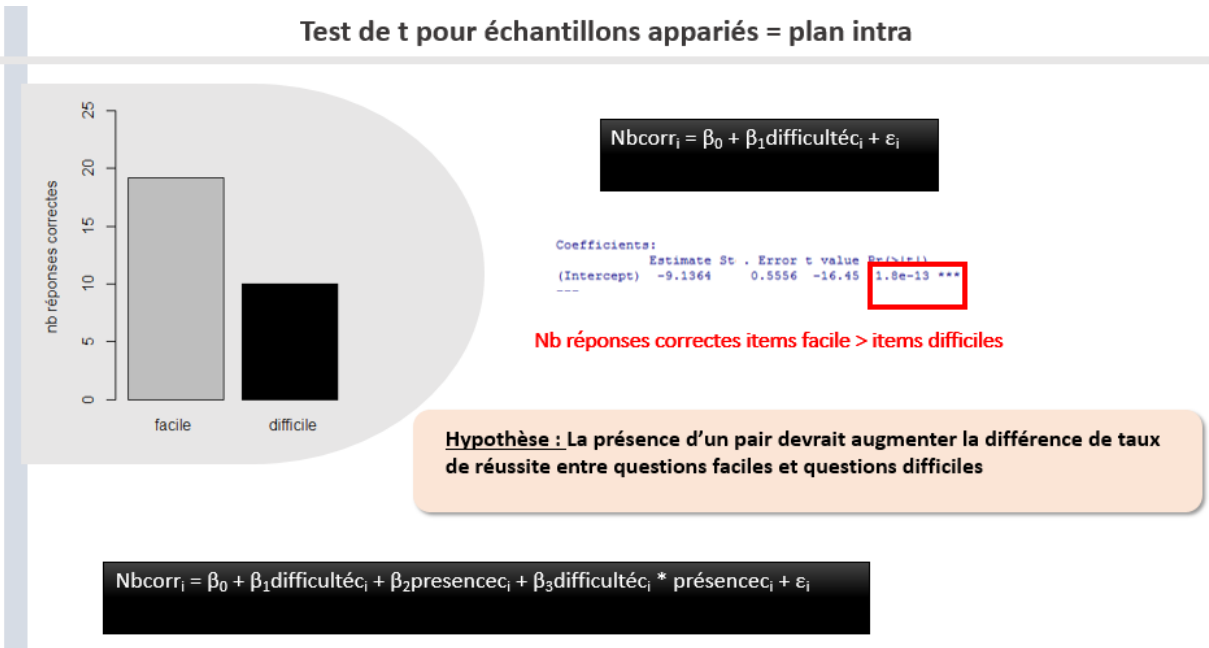

```
Anova Table (Type III tests)

Response: score_T1 - score_T0
                  Sum Sq Df F value    Pr(>F)
(Intercept)        39.76  1  2.6941 0.1180736
conditionc        343.30  1 23.2598 0.0001364 ***
sexec              40.75  1  2.7611 0.1138997
conditionc:sexec    0.49  1  0.0333 0.8573257
Residuals         265.67 18
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ici, seul l'effet de la condition (matin versus soir) est retrouvé significatif.

Régression multiple : modèle mixte (une VI catégorielle intra-individuelle ; une VI catégorielle inter-individuelle)

Situation

Nous décidons d'administrer à deux cohortes d'étudiants universitaires, un questionnaire composé de 50 items dont la difficulté a été manipulée. 25 items sont jugés de niveau facile et 25 items de niveau difficile. La passation est individuelle. Toutefois, les conditions de passations sont différentes entre les deux cohortes. Alors que les étudiants de la première cohorte répondent au questionnaire dans un box expérimental dans lequel ils sont laissés seuls. Les étudiants de la seconde cohorte répondent au questionnaire dans un box expérimental dans lequel ils sont laissés avec un pair qui, par-dessus l'épaule du participant, prend connaissance des réponses apportées.



Le graphique représente le nombre de réponses correctes aux items jugés faciles versus difficiles. Descriptivement, l'histogramme manifeste moins de réponses correctes pour les items difficiles, comparativement aux items de niveau facile.

Dans ce cas-ci, nous pouvons commencer par proposer un modèle de régression avec u

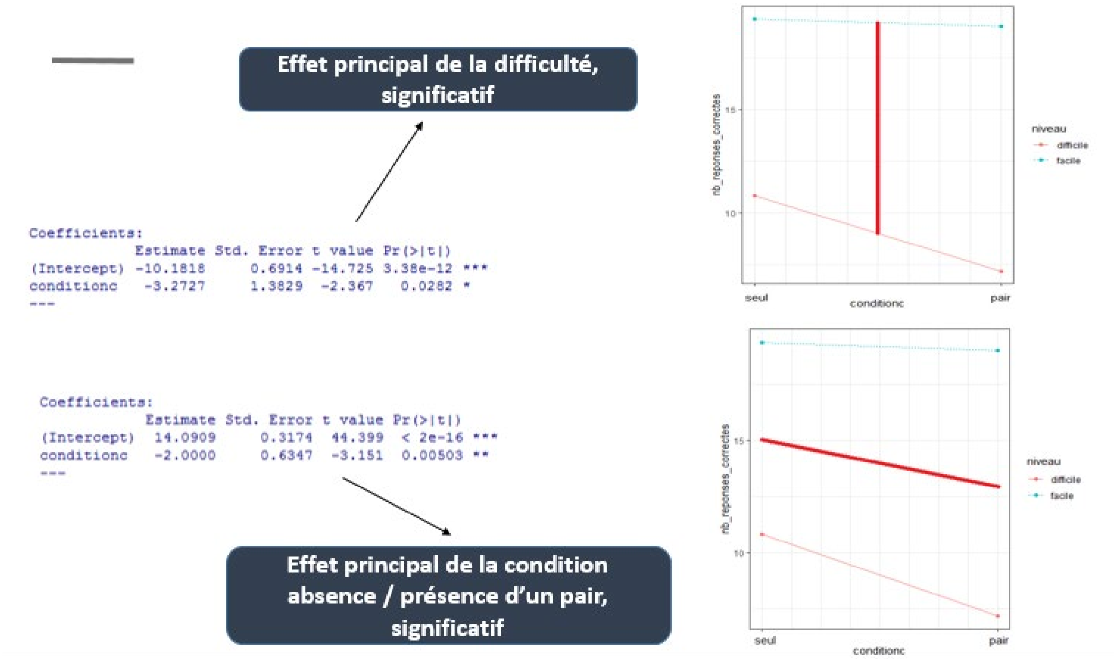
Le modèle contraint, correspondrait à prédire pour chaque participant, son nombre de réponses correctes à partir de la moyenne du nombre de réponses correctes obtenu par l'ensemble des apprenants.

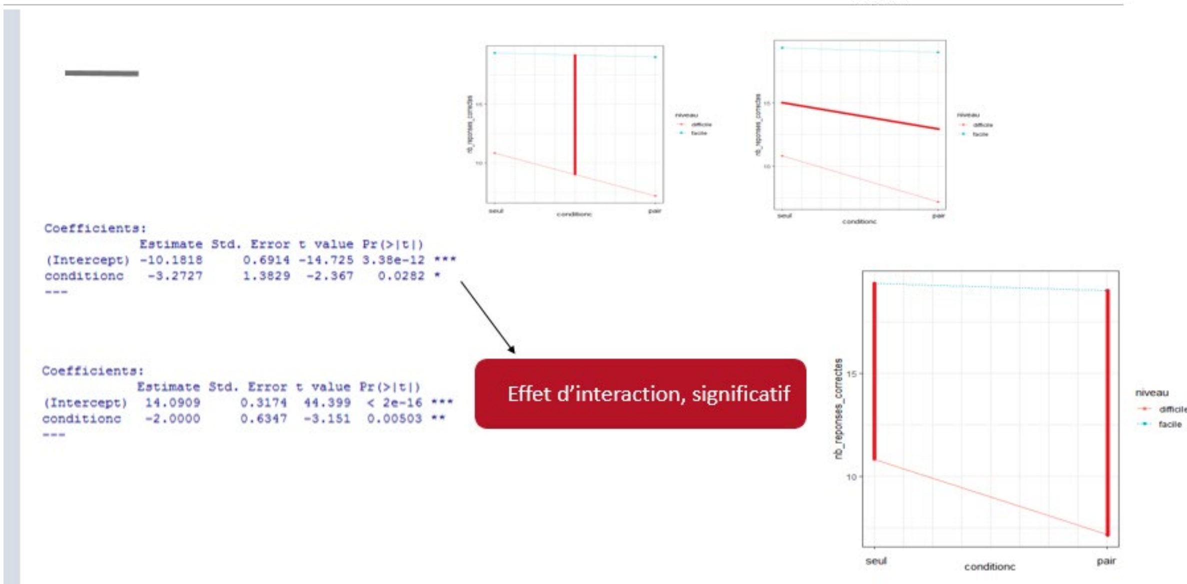
Pour le modèle augmenté, nous pouvons nous demander si le niveau de difficulté de l'item pourrait permettre de prédire le nombre de réponse correctes au questionnaire. Il s'agit donc d'un modèle de régression avec un prédicteur (le niveau de difficulté).

Ce modèle correspond à un t-test pour échantillons appariés. Contrairement au point xx, pour lequel les participants étaient assignés à l'une des deux conditions, ici, chaque participant réalise les 2 conditions, à savoir les items faciles et les items difficiles.
Une différence significative est retrouvée entre la condition facile et la condition difficile, avec une diminution de la performance de 9.14 lorsque l'on passe des items faciles à difficiles.

À ce moment, nous pouvons nous demander si la présence d'un pair dans le box expérimental pourrait amplifier la différence du taux de réussite entre la condition items facile et la condition items difficiles.

Ainsi, nous pouvons ajouter dans le modèle de régression l'effet présence/absence d'un pair ainsi que le produit des deux variables condition :
(présence d'un pair versus seul) * niveau de difficulté (facile/difficile)

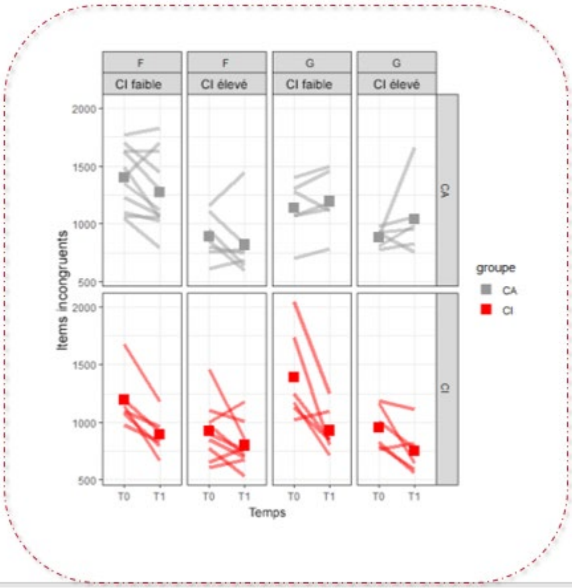




L'effet principal de la difficulté est significatif → il existe une différence significative du nombre de réponses correctes entre la condition facile et la condition difficile et ce, indépendamment de la présence ou non d'un pair. Donc le nombre de réponses correcte en condition difficile est plus faible qu'en condition facile.

Effet principal de la condition (présence/absence d'un pair) → les items sont moins bien réussis en condition « présence d'un pair », qu'en condition « seul » et ce, indépendamment du niveau de difficulté.

L'effet d'interaction est significatif, indiquant que la différence entre questions faciles et questions difficiles est plus faible en condition seul qu'en condition « présence d'un pair ». Donc la différence entre les 2 niveaux de difficultés varie significativement en fonction de la présence ou non d'un pair. Il s'agit d'un effet d'interaction.



Conseil : Il est bon de se limiter à 2 voire 3 VI.

Lorsque le modèle de régression comprend plus de trois prédicteurs – par exemple, un modèle qui tenterait d'expliquer une performance à un certain type d'items, en fonction du temps (pré versus post-test), du groupe expérimental (CI versus CA), du sexe (G versus F), du niveau initial en termes de CI (CI faible * CI élevé) – l'interprétation devient vraiment très compliquée.