Les initiatives des élèves en statistique descriptive : cas de l'ajustement linéaire en République Centrafricaine

DELON FABRICE RODOUMDJE

CREAD, F-29200 Brest
Université de Bretagne Occidentale
France
malkofabrice@gmail.com

ABSTRACT
Our study concerns students' initiative in mathematics, in the context of the Central African Republic. We focus on double statistical series and linear regression. We refer to the Anthropological Theory of Didactics. We proposed a problem on the theme of linear trend estimation to students in grade 12. In this article, we present and analyse this problem and we also analyse the students' productions, in terms of taking initiatives. The modelling activity offers possibilities, but the students' initiative is reduced, and the difficulties of calculation are an obstacle.

KEYWORDS
Praxeology, semiotic register, students' initiatives, modeling

RÉSUMÉ
Notre étude concerne les prises d'initiatives des élèves en mathématiques, dans le contexte de la République Centrafricaine. Nous nous centrons sur les séries statistiques doubles et l'ajustement linéaire. Nous nous référons à la Théorie Anthropologique du didactique. Nous avons proposé à des élèves de Terminale un problème sur le thème de l'ajustement. Dans cet article, nous présentons et nous analysons ce problème et nous analysons également les productions des élèves, en particulier en termes de prises d'initiatives. L'activité de modélisation offre des possibilités, mais les prises d'initiative des élèves sont réduites et les difficultés de calcul y font obstacle.

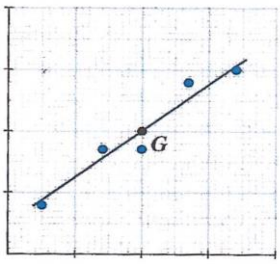
MOTS CLÉS
Praxéologies, registre sémiotique, prises d'initiatives des élèves, modélisation

INTRODUCTION ET CONTEXTE

Notre recherche en didactique des mathématiques se situe dans le contexte de l'enseignement secondaire en République Centrafricaine (RCA dans ce qui suit). La RCA est un pays qui fait face à de nombreuses difficultés politiques et économiques. L'une des conséquences de cette situation est que les classes dans l'enseignement secondaire sont particulièrement surchargées : il n'est pas rare d'avoir environ cent cinquante élèves dans une classe de Terminale. C'est dans ce contexte que nous avons entrepris une recherche sur les prises d'initiatives, possibles et effectives des élèves dans le domaine de la statistique descriptive.

Vue l'importance de l'enseignement de la statistique et la nécessité pour un citoyen de comprendre et d'interpréter les données qu'il croise dans divers domaines (Gattuso, 2011), il est donc important pour les futurs cadres de développer une pensée et un raisonnement statistique (Régnier, 2005). Les statistiques descriptives sont enseignées au secondaire en Centrafrique. En particulier, en classe de Terminale Scientifique, les élèves rencontrent les séries statistiques doubles, et réalisent des ajustements de ces séries afin de résoudre des problèmes d'estimation. Sur le plan mathématique, on peut résumer brièvement cette problématique d'ajustement de la manière suivante : lorsque les points du nuage, associés à cette série statistique double sont sensiblement alignés, on peut alors construire une droite passant au plus près possible de ces points. On dit alors que cette droite réalise un ajustement affine du nuage de points de la série statistique double. Le sens attribué à l'expression « au plus près possible » correspond aux différentes techniques d'ajustement.

FIGURE 1



Ajustement affine

Plusieurs techniques sont possibles pour effectuer un ajustement linéaire, puisqu'il s'agit de faire passer une droite « près » d'un nuage de points, mais nous avons observé dans les manuels (Rodoumdje, 2019) que la technique était presque toujours indiquée dans l'énoncé de l'exercice, réduisant ainsi les prises d'initiatives possibles des élèves. Notons de plus qu'en RCA au lycée, les élèves n'ont pas accès à des ordinateurs équipés de logiciels statistiques ou de tableurs. En réalité, la plupart des élèves n'ont même pas de calculatrice.

Est-il possible, dans ces conditions délicates, que les élèves développent certaines initiatives dans le cours sur l'ajustement linéaire ?

Pour apporter des éléments de réponse à cette question, nous avons conçu un problème ayant un potentiel, du point de vue des prises d'initiatives aux élèves, sur le choix de la technique d'ajustement. Nous avons soumis ce problème aux élèves d'une classe de Terminale S en Centrafrique. Dans cet article, nous analysons ce problème, en particulier en termes d'initiatives possibles pour les élèves. Nous analysons aussi les productions de ces élèves et leurs initiatives effectives.

REVUE DE TRAVAUX

Travaux sur l'enseignement des statistiques

Henry (2010) souligne que l'apprentissage des outils statistiques permet aux futurs citoyens de faire la lecture de tableaux, de graphiques, d'analyses économiques... Cela sert l'individu dans sa vie quotidienne (Jullien & Nim, 1989). Gattuso (2011) montre l'importance et la place de l'enseignement de la statistique au sein de l'enseignement des mathématiques au secondaire. Selon l'auteur, l'enseignement de la statistique favorise le développement de la pensée statistique et la construction de concepts statistiques, en ce sens, il ne faut pas rester sur l'aspect calculatoire mais développer la compréhension conceptuelle.

Lors d'une expérimentation d'enseignement de la statistique (Gattuso & Pannone, 2000), dont le but est d'utiliser les données réelles et officielles pour faire acquérir aux élèves les compétences en raisonnement statistique, force est de constater que les enseignants qui ont pris part à l'expérimentation ont une conception superficielle de la statistique descriptive orientée vers le traitement des données plutôt qu'au niveau de connaissances théoriques. Alors que le thème des statistiques ouvre des possibilités pour un enseignement fondé sur l'investigation : travail sur des données réelles (Goga & Ardilly, 2017), voire même conception d'une enquête et recueil de données par les élèves, choix de la technique de traitement des données… les recherches montrent que ce potentiel n'est pas souvent exploité. De plus, les travaux qui ont proposé et évalué des enseignements de la statistique fondés sur l'investigation ou la résolution de problèmes mobilisent tous des technologies informatiques (Jamie, 2002) qui ne sont pas accessibles en RCA. À notre connaissance, aucun travail de recherche en didactique n'a encore concerné les statistiques en RCA.

Travaux sur la modélisation

Dans cette partie nous nous intéressons aux travaux de recherche en didactique sur la modélisation. En effet, les problèmes de statistiques descriptives sont en particulier des problèmes de modélisation : on part de données « réelles » ou le plus souvent présentées comme telles, et on élabore un modèle mathématique qui va permettre de tirer diverses informations de ces données.

Nous nous référons à la notion de modèle telle que définie par Henry (2001, p. 151) : « un modèle est une interprétation abstraite, simplifiée et idéalisée d'un objet du monde réel, ou d'un système de relations, ou d'un processus évolutif issus d'une description de la réalité. » Le processus de modélisation, en mathématiques, comporte plusieurs étapes. Il y à la source de ce processus une situation extra-mathématique ; mais bien souvent, comme le souligne Coulange (1997), les élèves n'ont accès qu'à un modèle pseudo-concret, avec une situation très simplifiée par rapport à la vie courante. Par exemple en statistiques il s'agira de traiter un nombre réduit de valeurs (alors que les statisticiens professionnels manipulent des fichiers comportant des milliers de données), et il n'y a aucun travail à faire sur ces valeurs de départ (alors que les statisticiens doivent supprimer les valeurs aberrantes, décider comment traiter les données manquantes etc.).

À partir de ce modèle pseudo-concret, il faut déterminer un modèle mathématique : par exemple dans notre cas une droite d'ajustement. On réalise ensuite des traitements au sein de ce modèle mathématiques. Finalement on retourne au modèle pseudo-concret en interprétant les résultats obtenus. Ces différentes étapes : détermination du modèle mathématique, traitement dans le modèle, interprétation et retour au modèle pseudo-concret peuvent être plus ou moins à la charge de l'élève.

Coulange (1997) s'intéresse à la modélisation dans le cas des systèmes linéaires. Elle constate en analysant des manuels que la seule responsabilité des élèves dans la résolution de problèmes concrets, est de procéder à l'écriture d'un bon système linéaire à partir de l'énoncé du problème, en rapport avec le modèle mathématique. Nous avons de même dans nos analyses de manuel que le choix de l'échelle de représentation graphique du nuage de points est à la charge des élèves.

CADRE THÉORIQUE ET QUESTIONS DE RECHERCHE

Théorie Anthropologique du didactique

Nous nous plaçons dans le cadre de la théorie anthropologique du didactique (TAD) de Chevallard (1998), et utilisons plus précisément la notion de praxéologie.

La praxéologie, ou organisation praxéologique vient du grec praxis, qui signifie « pratique » (savoir-faire) et logos qui signifie « raison » (savoir). Une praxéologie comporte quatre éléments, elle est notée **[T/t/ θ/Θ].** Pour un type de tâches **T** donné, une praxéologie relative à **T** précise une manière d'accomplir, de réaliser les tâches: à une telle manière de faire, τ, on donne ici le nom de technique (du grec tekhnê, savoir-faire). (Chevallard, 1998). On entend par technologie, et on note généralement θ, un discours rationnel (logos) sur la technique la tekhnê τ, discours ayant pour objet premier de justifier « rationnellement » la technique τ, en nous assurant qu'elle permet bien d'accomplir les tâches du type **T**, c'est-à-dire de réaliser ce qui est prétendu (Chevallard, 1998). Donc la technologie a pour rôle de justifier, d'expliquer la technique utilisée. Et toute technologie a besoin à son tour d'une justification, que l'on appelle la théorie associée à cette technologie et qu'on note Θ. Nous pouvons utiliser la TAD pour décrire les activités de l'élève dans la situation d'apprentissage dans une institution. Il y a une négociation de la répartition des tâches et des responsabilités entre le professeur et les élèves. Ainsi la notion de *topos de l'élève* est définie comme l'ensemble des gestes que l'élève aura à accomplir en *autonomie didactique* (Chevallard, 2001) : le topos de l'élève peut se réduire à l'application d'une technique, ou peut être beaucoup plus important.

Les registres de représentations sémiotiques

Duval (1993) appelle registre de représentation tout système sémiotique permettant d'accomplir les trois activités cognitives fondamentales à la sémiosis : la formation d'une représentation identifiable, le traitement et la conversion. La formation d'une représentation identifiable, constitue une trace ou un assemblage de traces perceptibles qui soient identifiables comme une représentation de quelque chose dans un système déterminé. Le traitement, c'est une transformation interne à un registre. Et convertir les représentations produites dans un système en représentations d'un autre système, de telle façon que ces dernières permettent d'expliciter d'autres significations relatives à ce qui est représenté, c'est ce que Duval appelle la conversion. Dans le domaine des séries statistiques à deux variables, les représentations jouent un rôle très important. La série est souvent donnée au départ sous forme d'un tableau. Mais le processus d'ajustement se fait en passant par le registre graphique. Par conséquent, les élèves doivent pouvoir passer du registre tableau au registre graphique ou algébrique, en vue de faire des estimations.

Les initiatives des élèves

Dans cette partie, nous précisons ce que nous appelons ici *initiatives des élèves*. Nous avons vu que le « topos » de l'élève désigne dans la TAD ce qui est à la charge de l'élève dans une praxéologie : le choix d'une technique parmi plusieurs, ou simplement l'application d'une technique qui est indiquée ; la justification de cette technique etc.. Dans ce topos il peut y avoir des activités techniques qui ne requièrent pas d'initiatives de la part des élèves. Par ailleurs, il peut y avoir des prises d'initiatives des élèves qui ne relèvent pas du modèle praxéologique. C'est pourquoi nous proposons une définition spécifique.

En lien avec la modélisation, nous considérerons qu'il y a des initiatives des élèves possibles à différentes étapes. Lorsqu'il faut choisir un modèle mathématique ; lorsqu'il faut prendre des informations dans l'énoncé pour leur appliquer le modèle mathématique choisi, parfois en transformant les données de l'énoncé ; lorsqu'il faut interpréter les résultats fournis par le modèle mathématique.

Du point de vue des registres, nous considérerons les changements de registre également comme des prises d'initiatives des élèves. Ces prises d'initiatives sont limitées si la nécessité de changement de registre est indiquée dans l'énoncé, ou habituelle pour un certain type d'exercices. Elle est plus importante si l'élève décide par lui-même de changer de registre.

Nous proposons de considérer les initiatives des élèves comme des processus issus des actions et décisions de l'élève, en rapport avec le savoir. La prise d'initiative des élèves consiste à construire des connaissances, et à formuler des stratégies lors de la résolution des problèmes.

Questions de recherche
C’est en s’appuyant sur les différents cadres théoriques précédemment présentés, que nous en arrivons aux questions suivantes :

- Dans un problème de statistiques descriptives conçu pour permettre des prises d’initiatives des élèves, quelles praxéologies mettent-ils en œuvre ?
- Quelles initiatives prennent-ils, en particulier en termes de modélisation ?

MÉTHODOLOGIE

Dans cette partie, nous présentons tout d’abord nos choix pour l’élaboration d’un problème en statistique descriptive. Ensuite nous présentons nos outils pour le recueil de données en classe lors du travail des élèves sur ce problème, puis pour l’analyse de productions d’élèves issues de cette séance.

Conception et analyse a priori d’une situation
Nous avons centré notre étude sur le thème de l’ajustement statistique, qui permet de fournir des estimations lors de l’étude de séries statistiques doubles. En effet, ce thème est propice à un travail de modélisation. Lorsqu’une estimation est demandée, au sujet d’une série statistique double, les élèves peuvent prendre l’initiative de déterminer une droite fournissant un ajustement du nuage de points correspondant à la série, et d’utiliser cette droite pour fournir une estimation. Mais dans la plupart des exercices de manuels scolaires sur ce thème (Rodoumdje, 2018, 2019) la technique d’ajustement est indiquée : « en utilisant la technique des moindres carrés » ou « par la technique de Mayer ». Nous avons choisi dans notre problème (voir la présentation et l’analyse du problème dans la partie suivante) de ne donner aucune indication de technique.

Analyse a priori de l’énoncé
L’énoncé proposé s’inspire d’un exercice du manuel mathématiques (Roche & Barny, 2000). Nous avons modifié les questions, pour laisser le choix de la technique d’ajustement à la charge des élèves.

FIGURE 2

Exercice

Le tableau suivant indique en euros les montants des remboursements annuels y_i effectués de 2013 à 2018 par un ménage ayant contracté plusieurs emprunts.

Année	2013	2014	2015	2016	2017	2018
Rang de l’année x_i	1	2	3	4	5	6
Montant des remboursements y_i	3568	4561	5502	6693	8241	9120

Le ménage pense qu’en 2019, il dépensera moins de 10000 euros. Que pensez-vous de cette prévision ? Vous expliquerez en détail la méthode que vous avez choisie. Puis justifier pourquoi un ajustement affine est correct (ou approprié).

Exercice proposé aux élèves

Notre objectif dans cette partie du travail, est de faire l’analyse praxéologique de cet exercice. Et préciser le rôle de la calculatrice, et les registres qui peuvent être utilisés par les élèves.

Analyse Praxéologique

Pour répondre à la question qui est : « Que pensez-vous de cette prévision ? », il faut tout d'abord interpréter la question mathématiquement comme « il faut calculer une estimation de la dépense en 2019 ». Le premier type de tâche (implicite ici) est donc « interpréter la question mathématiquement », ce qui requiert une première initiative des élèves.

Le second type de tâches pour cette question est de: « faire une estimation en l'an 2019 ». Ici il n'y a aucune indication de technique, sur la manière de réaliser cette estimation. Ainsi nous considérons que les élèves prennent une initiative en choisissant cette technique. Le choix attendu peut être décrit par les sous-types de tâches suivants :

- Réaliser un ajustement du nuage de points ;
- Réaliser un ajustement affine.

Pour cet ajustement affinee, il est possible que les élèves optent pour une résolution graphique en plaçant une droite qui passe « près » du nuage. Dans ce cas, de nouveaux sous-types de tâches interviennent :

- Choisir un repère adapté pour représenter les points de coordonnées (x_i, y_i)
- Représenter ces points ;
- Tracer une droite d'ajustement ;
- Déterminer le rang de l'année 2019 ;
- Lire l'ordonnée du point correspondant de la droite.

Nous notons en particulier que, pour le sous-type de tâches « tracer une droite d'ajustement », plusieurs techniques sont possibles. Les élèves peuvent réaliser un ajustement visuellement, en plaçant la règle « au plus près » des points du nuage. Ils peuvent aussi joindre deux points particuliers du nuage, comme les deux points d'abscisses 1 et 6. Ceci est également une possibilité de prise d'initiative. Si les élèves ne choisissent pas une résolution graphique, ils vont s'orienter vers une résolution par le calcul. Dans ce cas le sous-type de tâches suivant est :

- Trouver l'équation d'une droite qui ajuste le nuage de points.

Cependant il faut souligner que les élèves ont récemment travaillé ce chapitre, ce qui réduit la prise d'initiative nécessaire. Du point de vue praxéologique, dans la technique des moindres carrés on calcule d'abord le coefficient directeur a de la droite. Les éléments technologiques qui permettent de faire ces calculs sont : la définition de la moyenne arithmétique permet de calculer $\bar{x}$ et $\bar{y}$. Puis la définition de la covariance pour le calcul σ_{xy} et ensuite la définition de la variance permet de faire le calcul de σ_x^2. Le coefficient directeur est obtenu par le quotient de la covariance de x et y avec la variance de x .

Le rôle de la calculatrice est très important dans la réalisation de ce type de tâches. L'utilisation de la calculatrice est obligatoire. Et dans les classes de terminale S en Centrafrique, les calculatrices scientifiques avec les fonctions statistiques sont autorisées (cependant peu d'élèves en disposent). Pour le calcul de la variance et la covariance, les élèves qui n'ont pas de calculatrice et qui utilisent leur téléphone qui n'a pas de touche « moyenne », pour faire des calculs, peuvent utiliser la parenthèse pour effectuer le premier calcul, moins le produit du deuxième terme. Ou bien ils peuvent recalculer à chaque fois jusqu'à ce qu'ils aboutissent au résultat, c'est-à-dire ils vont recalculer chaque terme au carré pour la variance ou le produit pour la covariance qui est dans la parenthèse, puis recalculer le deuxième avant de faire la différence des résultats.

Les registres utilisés dans la technique des moindres carrés sont le registre algébrique des variables et le registre numérique des variables quantitatives discrètes pour la détermination de l'équation de la droite de régression. Les élèves peuvent aussi présenter leurs résultats intermédiaires dans un tableau statistique.

Les élèves peuvent aussi utiliser la technique de Mayer, qui consiste à séparer le nuage en deux : ici les trois premiers points, puis les trois suivants. Il s'agit ensuite de déterminer les points moyens $G_1(\bar{x}_1, \bar{y}_1)$ et $G_2(\bar{x}_2, \bar{y}_2)$ des deux sous-nuages, et l'équation de la droite déterminée par ces deux points.

En ce qui concerne la deuxième question de l'énoncé, qui consiste à justifier la pertinence de l'ajustement affine, les élèves peuvent calculer le coefficient de corrélation entre les variables x et y. $r = 0{,}99 \approx 1$, r est proche de 1, on peut envisager un ajustement affine. On peut aussi envisager une technique reposant sur la représentation graphique du nuage de points. Si les points semblent alignés, alors une droite d'ajustement affine est appropriée.

Expérimentation en classe et méthodologie pour l'analyse de productions d'élèves

Séance observée en classe

L'activité que nous avons élaborée a été mise en œuvre le 25 avril 2019 dans la classe de terminale S, au lycée d'application de l'école normale supérieure de Bangui (LAENS). L'effectif des élèves était de 146, mais 84 élèves ont accepté de prendre part à l'activité. Les élèves interagissent en groupes pour résoudre le problème en temps limité de 1 heure. Pendant la séance le professeur n'intervient que pour les questions d'ordre organisationnel. Nous avons 16 groupes. Les élèves sont répartis en nombres de 5 ou 6 par groupes. Il y a au moins une calculatrice par groupe. Nous avons ainsi filmé la séance, et récupéré les productions écrites des élèves.

Liste des critères d'analyse de productions d'élèves

Nous sommes partis des critères, présentant une liste d'items relatifs à la démarche entreprise, comme outil méthodologique pour l'analyse de chaque copie. Les items se présentent comme suit :

- Valeur trouvée pour l'estimation en 2019 : la valeur est acceptable / non acceptable ;
- Erreur dans la procédure de résolution ;
- Technique choisie :
- À propos de la mise en œuvre de la technique : application d'une technique apprise en cours / présence d'explication sur la technique / plus autres rubriques pertinentes ?
- Présence d'un graphique : oui / non ;
- Si présence d'un graphique, à propos du graphique : choix de l'échelle pertinent oui / non ;
- Légende sur les axes oui / non ;
- Représentation du nuage de points / de la droite d'ajustement / plus autres rubriques pertinentes ;
- Présence d'une justification : cette justification est- elle correcte, suffisante ?
- Autres remarques sur la production.

ANALYSE DES PRODUCTIONS D'ÉLÈVES

Nous avons choisi une situation de classe où les élèves travaillent en petits groupes, et interagissent entre eux pour faciliter la construction du savoir.

Dans un premier temps, nous décrivons l'analyse de productions des élèves d'une manière globale, puis par la suite nous ferons l'analyse détaillée des réponses formulées par les élèves dans les différents groupes. Nous chercherons aussi à voir leurs prises d'initiative dans la résolution de l'exercice.

Analyse globale

Onze groupes sur seize ont choisi de réaliser un ajustement affine. Ils ont pris des initiatives, pour répondre à la question : en effet ils ont d'eux-mêmes déduit que pour répondre, il fallait estimer la dépense effectuée en 2019. Pour chercher l'estimation, ils ont décidé d'utiliser un ajustement. Et ces onze groupes ont choisi la technique des moindres carrés, pour trouver l'équation de la droite. Les cinq autres groupes n'ont pas réussi à proposer une modélisation adéquate. Ils ont représenté le nuage de points et calculé le coefficient de corrélation linéaire. Ils ont donné des réponses incomplètes et n'ont pas proposé d'estimation. Aucun groupe n'a choisi la technique de Mayer, ni la technique graphique en traçant une droite qui approche le nuage. Ceci s'explique par le fait que ces techniques ne sont pas beaucoup travaillées en classe. En ce qui concerne la détermination des coefficients a et b, de l'équation de la droite de la forme : $y = ax + b$, par la technique des moindres carrés, parmi les onze groupes qui ont choisi la technique des moindres carrés, deux groupes ont utilisé la technique qui consiste à appliquer directement la formule pour déterminer les coefficients a et b. Et les neuf autres groupes ont disposé leurs résultats dans un tableau statistique. Six groupes sur onze, ont pu déterminer l'équation de la droite, qui n'est pas correcte.

Concernant les six groupes qui ont déterminé l'équation de la droite, seuls cinq groupes ont réussi à faire une estimation. Pour justifier la pertinence de l'ajustement affine, cinq groupes ont utilisé comme technique, le coefficient de corrélation linéaire entre les variables x et y. Trois groupes ont utilisé la représentation graphique du nuage de points comme technique, pour justifier l'ajustement affinee. Sept groupes n'ont pas abordé la question. Et un groupe a fait une justification théorique par rapport à la notion de coefficient de corrélation linéaire sans effectuer le calcul.

Analyse détaillée des productions d'élèves

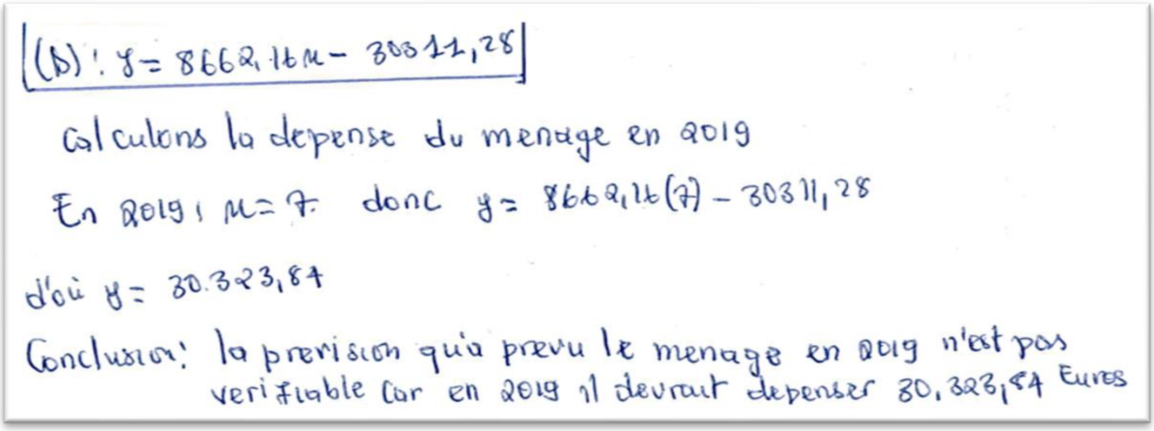
Concernant la valeur de l'estimation en 2019.

Concernant la valeur de l'estimation en 2019, aucun groupe n'a trouvé une valeur acceptable. Nous entendons par valeur acceptable, une valeur qui approche la valeur trouvée par la droite des moindres carrés à 1 % près (soit 100 euros en plus ou en moins, ce qui conduit à la même réponse sur le fait de dépasser ou non 10000). Par contre, six groupes d'élèves sur les onze ont trouvé des valeurs non acceptables. Et les cinq autres groupes ont donné des réponses incomplètes. Ils sont restés sur la détermination des coefficients de l'équation de la droite, donc ils ne proposent pas d'estimation.

Sur ces six groupes qui ont trouvé l'équation de la droite, les groupes nommés groupes 7, 9, 11 et 16 ont fait des erreurs dans la détermination de l'équation de la droite. Les valeurs des indicateurs statistiques (la variance et la covariance) mobilisés pour trouver l'équation de la droite sont erronés, donc ces groupes n'ont pas trouvé des valeurs acceptables de l'estimation, et ces élèves ne contrôlent pas l'ordre de grandeur.

Cependant les groupes 7, 9 et 11 ont des initiatives dans la réalisation des tâches qui leur sont demandées : c'est-à-dire, ils ont réussi à proposer un modèle pour faire l'estimation en l'an 2019. Ils ont fait des erreurs sur la formule de la covariance, mais leur modèle de l'estimation est juste. Pour le groupe 5, les élèves ont trouvé l'équation de la droite, mais ils n'ont pas trouvé une valeur juste de l'estimation. Ils n'ont pas utilisé la bonne valeur de la variable x. Nous constatons que dans ce groupe, les élèves ont appliqué les techniques apprises en classe pour déterminer l'équation de la droite. Mais il leur manque des prises d'initiatives dans la détermination de la valeur du rang de l'année en 2019 qui est $x = 7$. La procédure utilisée par le groupe 8 est bonne, mais la valeur de l'estimation est non acceptable. Cela est dû à des erreurs de calculs. Ils ont des prises d'initiatives, pour trouver la valeur de la variable $x = 7$, et de réaliser la tâche qui leur est demandée.

FIGURE 3


(D) : y = 8662,16x − 30311,28

Calculons la depense du menage en 2019

En 2019, x = 7 donc y = 8662,16(7) − 30311,28

d'où y = 30.323,84

Conclusion: la prevision qu'à prevu le menage en 2019 n'est pas verifiable car en 2019 il devrait depenser 30.323,84 Euros

Extrait de copie du groupe 7

Il semble que l'erreur de calcul s'explique par le fait que, dans certains groupes il y a qu'un seul élève qui a la calculatrice. Et s'il se trompe dans la manipulation des calculs élémentaires, alors cela va jouer sur les résultats du travail du groupe. Mais dans le groupe 8, on voit une prise d'initiatives dans la réalisation de la tâche demandée. Ils ont réussi à proposer une modélisation adéquate.

Pour la technique choisie

Onze groupes ont choisi de travailler avec la technique des moindres carrés. Cette technique a été beaucoup plus travaillée en classe pendant les exercices.

Le groupe 1 a utilisé la technique apprise en classe pour déterminer les valeurs des indicateurs statistiques comme la variance, les moyennes et la covariance. Pour ce faire, ils utilisent la calculatrice pour remplir le tableau statistique. Dans ce groupe, ils n'ont pas pu finir l'exercice, faute de temps. Ils ont une mauvaise gestion du temps, ils passent plus de temps à interagir entre eux, sans pouvoir mettre par écrit leurs idées.

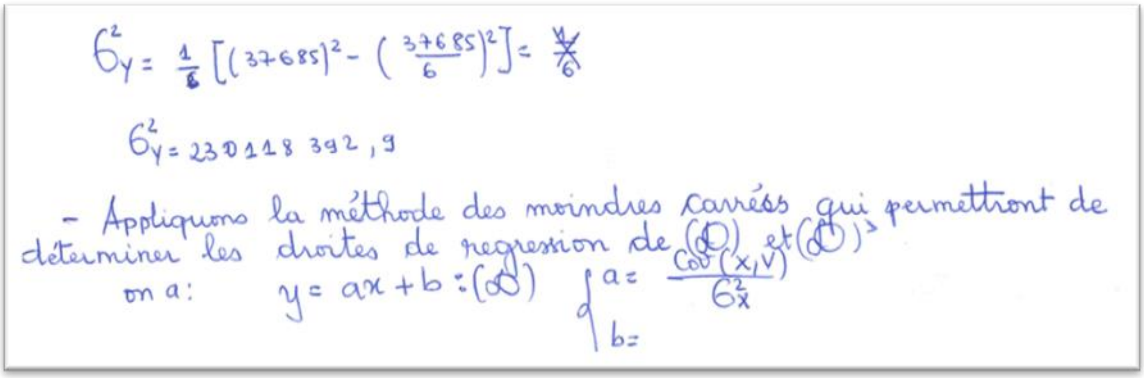
Le groupe 3 utilise aussi la technique du tableau statistique pour trouver les valeurs des résumés statistiques. Et on voit l'utilisation du registre algébrique pour trouver l'équation de la droite. On constate que la technique était travaillée en classe, mais les élèves de ce groupe n'ont pas fini l'exercice. Leur prise d'initiative est moyenne dans la réalisation de l'exercice.

Pour le groupe 5, ils ont utilisé le registre tableau et algébrique pour trouver l'équation de la droite de régression. Et l'explication fournie par ce groupe sur la valeur de l'estimation est juste, mais la valeur de l'estimation trouvée n'est pas acceptable. Il me semble que les élèves de ce groupe connaissent bien la technique apprise en cours sur la technique des moindres carrés. Mais au lieu du rang de l'année, ils ont mis la valeur 10000 de la dépense maximum prévue en 2019. Et c'est une difficulté liée à la modélisation, ici à l'interprétation de l'énoncé.

Le groupe 6 a mis en œuvre, la technique des moindres carrés, en utilisant directement les formules et les propriétés sur la technique des moindres carrés apprise en cours. Après avoir calculé les résumés statistiques, ils ont expliqué la technique, permettant de trouver l'équation de la droite de régression. En plus, on voit l'utilisation du registre tableau pour calculer l'effectif total, et ils ont une prise d'initiative dans la réalisation de la tâche.

Les groupes 7, 8, 9 et 11, utilisent la même technique du tableau statistique pour le calcul des concepts statistiques et le registre algébrique pour avoir l'équation. Dans leur procédure, les groupes 7, 8 et 9 n'ont pas expliqué la technique utilisée. Seul le groupe 11 explique la technique utilisée pour trouver la prévision.

FIGURE 4

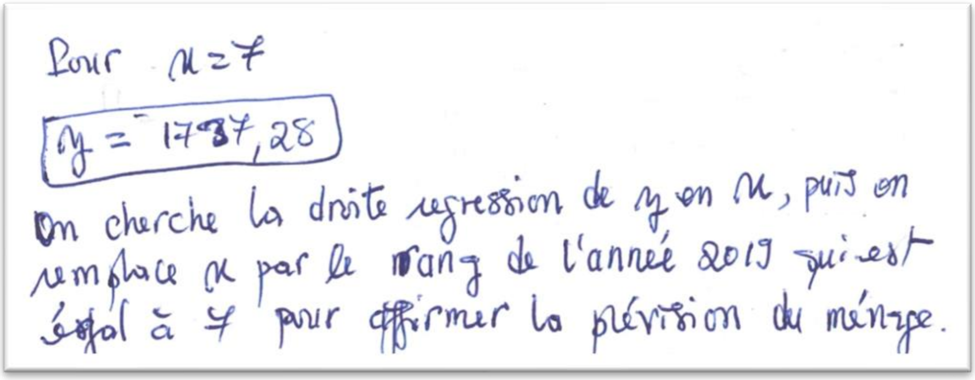

$\sigma_Y^2 = \frac{1}{6}\left[(37685)^2 - \left(\frac{37685}{6}\right)^2\right] = \cancel{\frac{y}{6}}$

$\sigma_Y^2 = 230118392,9$

- Appliquons la méthode des moindres carrés qui permettront de déterminer les droites de regression de (D) et (D'),
on a: $y = ax + b : (D)$ $\begin{cases} a = \frac{Cov(x,y)}{\sigma_x^2} \\ b = \end{cases}$

Extrait de copie du groupe 6

FIGURE 5


Pour $x = 7$

$\boxed{y = 1737,28}$

On cherche la droite regression de y en x, puis on remplace x par le rang de l'année 2019 qui est égal à 7 pour affirmer la prévision du ménage.

Extrait de copie du groupe 11

Et on voit une prise d'initiative dans le calcul de la valeur de prévision en année 2019, ils ont d'abord trouvé le rang de l'année, puis calculé la valeur de la prévision.

Pour les groupes 10, 14 et 16, la technique utilisée consiste à calculer les concepts statistiques sans utiliser le tableau statistique. Ils se sont arrêtés sur le calcul de la moyenne et la covariance. Seul le groupe 16 est arrivé au calcul de l'équation de droite. Ces groupes manquent d'initiatives dans leur façon de procéder, pour trouver la prévision.

Concernant le graphique.

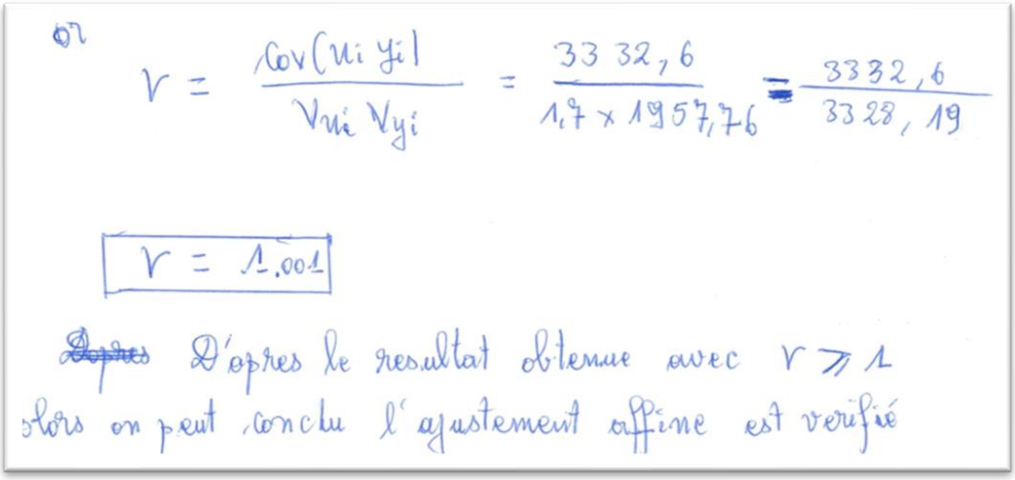
Treize groupes sur les seize ont fait un graphique. Seul quatre groupes ont utilisé le graphique pour justifier l'ajustement affine.

Pour le groupe 3, les élèves ont fait deux erreurs : un mauvais choix d'échelle et la confusion des axes. Pourtant l'énoncé précisait bien quelles valeurs étaient considérées comme x et lesquelles comme$y$. Les groupes 2, 4, 5, 8, 9, 10, 13, 15 et 16, ont fait une représentation graphique, sans pouvoir préciser l'échelle. Tous ces groupes ont fait une représentation du nuage de points. Et seuls les groupes 5 et 15 ont représenté la droite d'ajustement. Le groupe 6 a fait une représentation graphique sans présenter l'échelle. L'ensemble des points de coordonnées M_i (x_i, y_i), qui forme le nuage de points est représenté sur le repère. Pour les groupes 1 et 7, ils ont fait un choix pertinent de l'échelle, c'est une prise d'initiative pour les élèves de ces deux groupes. Le nuage de points est aussi représenté.

Dans la plupart des représentations, nous constatons que les élèves ont centré leur représentation graphique sur le nuage de points, sans éventuellement penser à la représentation d'une droite d'ajustement affine.

Concernant la justification de la pertinence de l'ajustement affine
Dix groupes sur seize n'ont pas répondu à la question. Le groupe 12, a donné une justification qui n'est pas juste à cause de la valeur du coefficient de corrélation linéaire qui est faux. En effet, ce groupe trouve un coefficient de valeur 33,95. Il déduit que l'ajustement affine n'est pas valable « car le coefficient n'est pas proche de 1 ». Mais il ne rend pas compte que cette valeur est forcément fausse. Les groupes 4 et 5 trouvent un coefficient de corrélation strictement supérieur à 1. C'est une erreur grave, normalement un étudiant qui trouve un coefficient de corrélation strictement supérieur à 1 ($r > 1$) se rend compte que son calcul est faux, et refait le calcul. Suite à des erreurs d'arrondi ce groupe (groupe 5) trouve un coefficient de corrélation supérieur à 1, mais il ne se corrige pas, au contraire il conclut que comme le coefficient est supérieur à 1 l'ajustement est bien justifié !

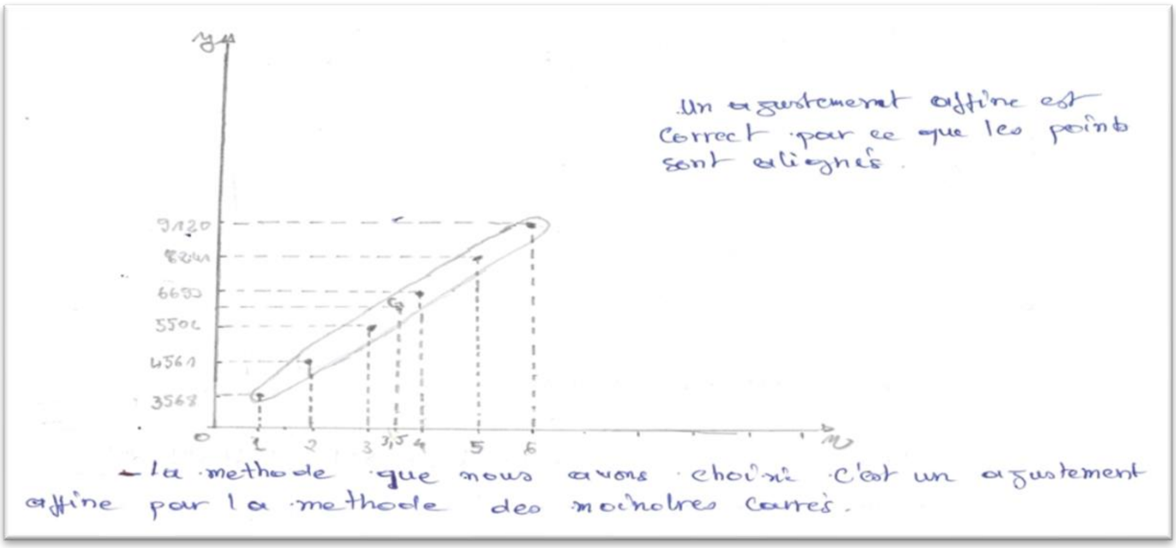
FIGURE 6



Extrait de copie du groupe 5

Par contre, les autres groupes (7, 10 et 13), ont utilisé la représentation graphique du nuage de points, pour la justification de la pertinence de l'ajustement affine.

FIGURE 7



Extrait de copie du groupe 10

Les élèves de ces cinq groupes (5, 7, 9, 10 et 13) ont une prise d'initiative dans la tâche qui leur a été demandée, de justifier la pertinence d'un ajustement affine.

CONCLUSION

Nous commençons par rappeler nos questions de recherche :

- Dans un problème de statistiques descriptives conçu pour permettre des prises d'initiatives des élèves, quelles praxéologies mettent-ils en œuvre ?
- Quelles initiatives prennent-ils, en particulier en termes de modélisation ?

Concernant le problème, nous retenons que certains groupes d'élèves parviennent à proposer un modèle adéquat pour faire l'estimation. Les élèves ont tous interprété la question comme relevant d'une estimation. Une majorité a recherché un ajustement affine avec la technique des moindres carrés. Ceci ne relève pas vraiment d'une prise d'initiative : c'est un effet du contrat didactique (Brousseau, 1998), car les élèves ont travaillé ce type d'exercice avec la technique des moindres carrés. Certains élèves parviennent à trouver une équation correcte de la droite d'ajustement par la technique des moindres carrés. Pour donner une estimation correcte, il faut qu'ils manifestent une prise d'initiative, en notant que 2019 est l'année qui correspond au rang 7. Certains élèves se trompent, en prenant la valeur 2019 ou bien la valeur de 10000. L'ordre de grandeur du résultat obtenu n'est pas réaliste, mais ils ne font pas cette observation pour corriger leur calcul. Par ailleurs, nous avons noté que les erreurs de calcul constituent un obstacle important.

Pour la question sur la validité de l'ajustement, les élèves qui ont fait une représentation graphique ont donné un argument correct. Pour ceux qui sont passés par le coefficient de corrélation linéaire, certains ont un résultat correct, mais on retrouve le problème des erreurs de calcul et de l'absence de contrôle de l'ordre de grandeur. Certains élèves trouvent des coefficients de corrélation linéaires plus grands que 1 et ne notent pas leur erreur.

Nous avions observé dans nos travaux précédents (Rodoumdje, 2019) que l'enseignement de l'ajustement affine en RCA ne laisse que très peu de place aux initiatives des élèves. Il n'est donc pas surprenant que ces élèves aient rencontré des difficultés, et aient manifesté une prise d'initiative limitée, pour ce problème qui leur laissait des choix pour la modélisation et la technique à employer.

Pour la suite de notre travail, nous souhaitons mettre en œuvre un enseignement de l'ajustement affine laissant plus de place aux initiatives des élèves. Nous évaluerons ensuite les conséquences de cet enseignement sur le travail des élèves, et leur capacité à prendre des initiatives. Dans le contexte complexe de la RCA, il nous semble important que la recherche en didactique puisse contribuer à l'évolution de l'enseignement.

RÉFÉRENCES

Brousseau, G. (1998). *Théorie des situations didactiques*. Grenoble: La Pensée Sauvage.

Chevallard, Y. (1998). Analyse des pratiques enseignantes et didactique des mathématiques : L'approche anthropologique. In *Actes de l'Université d'été de la Rochelle* (pp. 91-120). Clermont-Ferrand, France: IREM de Clermont-Ferrand.

Chevallard, Y. (2001). Organiser l'étude 1. Structures et fonctions. In J.-L. Dorier, M. Artaud, M. Artigue, R. Berthelot & R. Floris (Dirs.), *Actes de la Xième École d'été de didactique des mathématiques, Corps* (pp. 3-32). Grenoble: Pensée Sauvage.

Coulange, L. (1997). Les problèmes concrets à mettre en équations dans l'enseignement. *Petit x, 47*, 33-58.

Duval, R. (1993). Registres de représentation sémiotique et fonctionnement cognitif de la pensée. In *Annales de Didactique et de Sciences Cognitives 5* (pp. 37-65). Strasbourg: IREM de Strasbourg.

Gattuso, L. (2011). L'enseignement de la statistique : Où, quand, comment, pourquoi pas ? *Statistique et Enseignement, 2*(1), 5-30.

Gattuso, L., & Pannone, N. (2002). teacher's training in a statistics teaching experiment. In *Proceedings of the Sixth International Conference on Teaching Statistics* (pp. 685-692). Durban, South Africa: International Statistical Institute.

Goga, C., & Ardilly, P. (2017). Présentation du challenge : « Graine de sondeurs ». *Statistiques et Enseignement, 8*(2) 103-111.

Henry, M. (2010). Évolution de l'enseignement secondaire français en statistique et probabilités. *Statistique et Enseignement, 1*(1), 35-45,

Jamie, D. M. (2002). Using computer simulation methods to teach statistics: A review of the literature. *Journal of Statistics Education, 10*(1), 4. Retrieved from https://doi.org/10.1080/10691898.2002.11910548.

Jullien, M., & Nim, G. (1989). L'E.D.A au secours de l'O.G.D ou quelques remarques concernant l'enseignement de la statistique dans les collèges. *Petit x, 19*, 29-41.

Régnier, J.-C. (2005) Formation de l'esprit statistique et raisonnement statistique. Que peut-on attendre de la didactique de la statistique ? *Recherche en Didactique des Mathématiques, 1*, 13-38.

Roche, F., & Barny, F. (2000). *Mathématiques comptabilité et gestion Informatique de gestion*. Paris: Hachette Éducation.

Rodoumdje, F. (2018). Les initiatives des élèves dans l'apprentissage des statistiques en Centrafrique. In *Actes du Deuxième Colloque de l'Association de Didacticiens des Mathématiques Africains - ADIMA2* (pp. 267-277). Dangbo, Bénin : Institut de Mathématiques et de Sciences Physiques de Dangbo.

Rodoumdje, F. (2019). *Initiatives des élèves dans l'apprentissage des statistiques descriptives : Étude de manuels*. Communication par poster, École d'été de didactique des mathématiques, Autrans, France.

ANNEXE 1 : SOLUTION DE L'EXERCICE PROPOSÉ AUX ÉLÈVES

Solution

Technique des moindres carrés.

Équation de la droite de régression de $y$en x est : $y = ax + b$ avec $a = \frac{\sigma_{xy}}{\sigma_x^2}$ et $b = \bar{y} - a\bar{x}$

- Calcul de $\bar{x}$ et $\bar{y}$, $\bar{x} = \frac{1}{n}\sum_{i=1}^{6} x_i = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3,5$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{❑} y_i = \frac{1}{6}(3568 + 4561 + 5502 + 6693 + 8241 + 9120) = 6295,83$$

- Calcul de la variance de :x σ_x^2

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{6} x_i^2 - \bar{x} = \frac{1}{6}(91) - (3,5)^2 = 2,916$$

- Calcul de la covariance : σ_{xy}

$$\sigma_{xy} = \frac{1}{n}\sum_{i=1}^{6} x_i y_i - \bar{x}\bar{y} = \frac{1}{6}(151893) - (3,5 * 6295,83) = 3295,08$$

$a = \frac{\sigma_{xy}}{\sigma_x^2} = 1129,5$ et $b = \bar{y} - a\bar{x} = 2342,6$

L'équation de la droite de régression de y en x est donc :$y = 1129,5x + 2342,6$

L'an 2019 correspond à $x = 7$ d'où $y = 10249,1$

En 2019 le ménage va dépenser plus de 10 000 euros, pour le remboursement.

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = 0,99$$

Le coefficient de corrélation est proche de 1, on peut envisager un ajustement affinee.