



# **Logiciels de statistique et Interactions Humain-Ordinateur dans les processus d'enseignement-apprentissage de la statistique en contexte universitaire**

Jean-Claude Régnier

## **► To cite this version:**

Jean-Claude Régnier. Logiciels de statistique et Interactions Humain-Ordinateur dans les processus d'enseignement-apprentissage de la statistique en contexte universitaire. 11th International Meeting on Information, Knowledge and Action - Cognition and Modeling, Federal University of Jequitinhonha and Mucuri Valleys (UFVJM) and by the Brazilian Society on Cognitive Science (SBCC), with several partnerships and the support of CAPES., Mar 2021, Diamantina, MG - Brasil, Brésil. <hal-02376037>

**HAL Id: hal-02376037**

**<https://hal.science/hal-02376037v1>**

Submitted on 20 Mar 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

## **Logiciels de statistique et Interactions Humain-Ordinateur dans les processus d'enseignement-apprentissage de la statistique en contexte universitaire**

Jean-Claude Régnier

Professeur des universités émérite

UMR 5191 ICAR – Université de Lyon -Lyon2 (France)

Professeur invité

Université d'État de Tomsk (Sibérie – Russie)

[Jean-Claude.Regnier@univ-lyon2.fr](mailto:Jean-Claude.Regnier@univ-lyon2.fr)

## **Softwares de estatística e Interações Humano-Computador nos processos de ensino-aprendizagem da estatística em um contexto universitário**

## **Statistics Software and Human-Computer Interactions in the Teaching-learning Processes of Statistics in a University Context**

### **Résumé**

Cette conférence a pour but de conduire une réflexion sur le rôle des instruments informatiques, tels que les divers logiciels mobilisés pour les traitements et analyses des données, dans les processus d'enseignement et d'apprentissage de la statistique en contextes universitaires. Nous avons adopté une approche basée sur la didactique de la statistique et sur les apports d'études sur les interactions entre les êtres humains et les ordinateurs qui dépassent la logique « presse boutons » et qui jouent un rôle d'amplificateur culturel pour l'apprentissage de la statistique. Les questions de la connaissance mathématique sous-jacente au savoir-faire statistique et du niveau minimum des compétences mathématiques apparaissent aussi centrales surtout dans le cadre du courant statistique sans mathématiques qui considère que les logiciels assument la dimension mathématique des traitements statistiques.

### **Mots clé**

Enseignement-apprentissage de la statistique, raisonnement statistique, traitement statistique assisté par informatique

### **Resumo**

Esta conferência visa conduzir uma reflexão sobre o papel dos instrumentos informáticos, como os diversos softwares mobilizados para os tratamentos e análises de dados, nos processos de ensino-aprendizagem da estatística em contextos universitários. Posicionamo-nos em uma abordagem baseada na didática da estatística e nos aportes dos estudos sobre as Interações Humano-Computador que ultrapassem a lógica de “apertar botões”, e que ofereçam um papel de amplificador cultural para a aprendizagem da estatística. As questões do conhecimento matemático subjacente ao saber-fazer estatístico e do nível mínimo de competências matemáticas aparecem também como centrais sobretudo no quadro da corrente *estatística sem matemática* que considera que os softwares assumem a dimensão matemática dos tratamentos estatísticos.

**Palavras chave :**

Ensino-aprendizagem da estatística, raciocínio estatístico, tratamento estatístico assistido por informática

**Abstract:**

This conference aims to conduct a reflection on the role of computer tools, such as the various software mobilized for the treatment and analysis of data, in the processes of teaching-learning of statistics in university contexts. We position ourselves in a approach based on the didactics of statistics and the contributions of studies on Interactions Human-computer that go beyond the “push buttons” logic, and that offer a role cultural amplifier for learning statistics. Knowledge issues underlying the statistical know-how and the minimum level of competence mathematics also appear as central especially in the statistical current without mathematics that considers that software takes on the mathematical dimension of treatments statistical data.

**Keywords :** Teaching-learning of statistics, statistical reasoning, statistical treatment assisted by informatics

## 1 Introduction

Le propos que nous tenons dans cet article qui fonde la conférence donnée au XI<sup>ème</sup> colloque international Information, Connaissance et Action, s'appuie principalement sur une réflexion conduite sur les produits de nos activités d'enseignement des mathématiques et de la statistique et de recherche en pédagogie et didactique des mathématiques et de la statistique, au cours des 50 années dernières années. Dès 1968, nous avons commencé à nous investir dans le domaine de l'éducation. Force est de constater que le monde, dans les deux sens donnés par Paulo Freire (Freire, 2006) (Régner, 2006, p.165-166) distinguant le *support* et le *monde* comme le *monde des êtres humains*, par exemple, a connu des transformations spectaculaires au moins dans ses dimensions matérielles, techniques et technologiques et ses dimensions sociales, culturelles et économiques. L'accroissement vertigineux des connaissances scientifiques et techniques produit par les découvertes et les inventions réalisées par les êtres humains qui ont poursuivi les questionnements et les investigations sur le *monde*, sans oublier les connaissances déjà acquises, a contribué à la construction d'outils qui participent eux-mêmes au développement des connaissances. Parmi ces outils, nous prenons ici particulièrement en considération les machines électroniques que sont les ordinateurs et les calculatrices portables programmables ou non, dont l'opérationnalité se fonde sur des objets immatériels fournis par les mathématiques et la science informatique. Cette puissance opérationnelle grandissante a apporté et continue à apporter une aide, difficilement mesurable, au développement de la statistique et à sa mise en œuvre.

Ainsi nous tenons comme un fait historique identifié chez les êtres humains et repérable depuis plusieurs millénaires, celui de leur propension à se pourvoir d'outils, d'instruments, d'ustensiles à partir d'objets naturels ou construits à partir de matériaux naturels de leur environnement afin de les aider à résoudre des problèmes soulevés par leur action et leur activité sur cet environnement même, le *monde* comme nous l'avons déjà évoqué plus haut.

Cela peut aller de la pierre utilisée telle quelle, ou des artefacts comme la pierre taillée ou la pierre polie afin d'amplifier la force de frappe de l'être humain, par exemple couper un arbre, jusqu' à ce que d'aucuns ont convenu d'appeler l'« homme augmenté » comme le résultat de modifications visant l'amélioration de ses capacités d'être humain au moyen d'interventions scientifiques et technologiques. Par exemple, un des moyens est le recours aux exosquelettes.

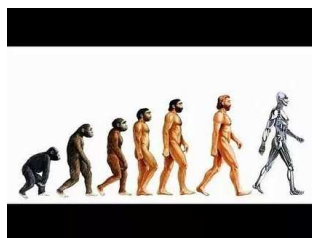


Figure 1 : Une conception de l'évolution humaine... (source<sup>1</sup>)

Ce qui nous intéresse est plus particulièrement les outils développés par les êtres humains pour les aider dans des traitements de calculs numériques et par la suite de calculs symboliques qui vont alors jouer un grand rôle dans l'histoire des mathématiques et des sciences qui sont instrumentées par les mathématiques, comme la statistique qui est le domaine qui nous intéresse ici.

Rappelons un exemple remarquable. En 1642 Blaise Pascal a construit une machine à calculer pour faciliter le travail de son père, commissaire des impôts en Normandie. Pour la première fois, une machine additionne et soustrait en effectuant automatiquement la retenue à l'aide d'une ingénieuse petite pièce : le sautoir. La « Pascaline » fut conçue pour simplifier les opérations d'arithmétique commerciale.

---

<sup>1</sup>[https://static.wixstatic.com/media/f2c212\\_38b4e9b3a7314451acb3823f58c40dd8.jpg/v1/fill/w\\_426,h\\_319,al\\_c,q\\_80,usm\\_0.66\\_1.00\\_0.01/f2c212\\_38b4e9b3a7314451acb3823f58c40dd8.webp](https://static.wixstatic.com/media/f2c212_38b4e9b3a7314451acb3823f58c40dd8.jpg/v1/fill/w_426,h_319,al_c,q_80,usm_0.66_1.00_0.01/f2c212_38b4e9b3a7314451acb3823f58c40dd8.webp)



Figure 2 : Machine arithmétique de Pascal au XVIIème siècle<sup>2</sup>

Trois siècles plus tard, à partir des années 40, nous avons pu assister à l'apparition d'usages systématisés de l'informatique et des ordinateurs dans les traitements des calculs scientifiques.

A partir de 1975, le développement de l'informatique s'accélère. La puissance et la fiabilité des ordinateurs n'ont de cesse de s'accroître. Leur miniaturisation en facilite la diffusion, en les rendant plus portables. Des domaines sont alors institués tel que la bureautique, par exemple, ainsi définie<sup>3</sup> comme l'« *Ensemble des techniques et des moyens tendant à automatiser les activités de bureau et principalement le traitement et la communication de la parole, de l'écrit et de l'image* ». Force est de constater que les outils informatiques impactent la quasi-totalité des pratiques sociales de recherche, conception, fabrication, commercialisation, publication et communication et même de nos jours, jusqu'aux pratiques quotidiennes individuelles. Les constructeurs ayant conçu des architectures pour les connecter, les machines informatiques sont mises en réseaux qui leur permettent d'établir des communications.

À partir des années 1990, l'ampleur du développement de l'intégration de l'informatique, déterminée au moyen de systèmes embarqués, dans les outils les plus divers mis en œuvre dans la vie quotidienne est immense au point d'enivrer certains individus fascinés par les possibilités, au risque d'être alors confrontés, par exemple, à la dialectique : protection des individus et de leur vie privée *versus* contrôle et surveillance systématiques des individus. La

---

<sup>2</sup> Une *pascaline*, signée par Pascal en 1652, visible au musée du Conservatoire national des arts et métiers à Paris. <https://collections.arts-et-metiers.net/?queryId=bdf4539d-9f66-4ac0-855a-e9d50b326de6>

<sup>3</sup> Définition du Journal Officiel de la République Française (arrêté du 22 décembre 1981) du 17 février 1982.

création d'Internet, réseau de réseaux à l'échelle de la planète, conduit à un dispositif technique permettant à ses utilisateurs de se connecter, de communiquer et d'échanger des informations sous forme de texte, d'image, de vidéographie ou de son. Ces connexions et les opérations mises en œuvre par les utilisateurs sont rendues possibles, d'une manière complètement transparente et invisible, par l'intermédiaire de protocoles informatiques opérant au travers de composants électroniques. Nous pourrions citer le protocole World Wide Web et les protocoles TCP(Transmission Control Protocol) /IP(Internet Protocol) qui définissent les règles d'échange des données et celles de la gestion des erreurs qui se produisent lors du transport, indépendamment des machines et logiciels utilisés au sein des réseaux. Ces possibilités offertes aux êtres humains ne sont pas sans soulever de sérieux problèmes éthiques, comme par exemple ceux que posent la diffusion de *fake-news* qui peuvent être fondées sur des documents falsifiés qui auront toutefois une apparence d'authenticité.

Durant ces périodes, l'utilisation des machines informatiques a conduit à prendre en considération par les utilisateurs d'une nouvelle dimension, celle de l'interactivité, en s'intéressant aux interactions entre êtres humains et machines qui se produisent dans l'univers des interfaces.

C'est dans ce décor historique brièvement et succinctement planté que nous plaçons notre réflexion dont l'objet central reste l'apprentissage de la statistique en milieu universitaire basé sur son enseignement dont nous avons assumé la charge durant près de 30 années. Rappelons que les étudiants concernés sont des non spécialistes pour lesquels la statistique est essentiellement une discipline-outil à mettre en œuvre dans les travaux de recherche fondés sur les données construites qui doivent être traitées pour être interprétées. Toutefois en tant qu'éducateur, notre démarche didactique et pédagogique ne peut faire l'économie de nos valeurs téléologiques : c'est aussi la formation citoyenne à la statistique (Régner, 2014b) (Régner, Braga, 2008) que nous visons au travers des situations d'enseignement que nous avons organisées. À cela, nous devons ajouter que les mathématiques tout comme l'informatique constituent deux cadres théoriques de référence

apportant des outils conceptuels et techniques fondamentaux pour la statistique et son développement.

## **2 Retour réflexif sur un premier outil de la fin du XX<sup>ème</sup> siècle : la machine à calculer électronique programmable ou non**

Au cours des années 70, alors que nous étions nous-même étudiant puis, à partir de 1974, professeur de mathématiques de lycée en France, sont apparues les premières calculatrices électroniques. En 1972, l'entreprise américaine Texas Instruments réalisa les premiers modèles. En 1976, l'entreprise américaine Hewlett-Packard commercialisa la première calculatrice électronique programmable. Ces outils ont alors été intégrés à l'enseignement des mathématiques. En 1981, nous avons publié un article qui traitait de la place et du rôle de cet outil pour l'enseignant et pour les élèves.

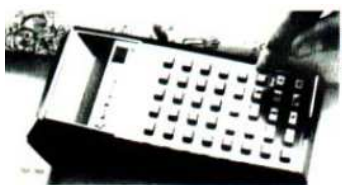


Figure 3 : Machine à calculer programmable des années 80(Régnier, 1981)

À cette époque, certes, les machines à calculer n'offraient pas autant de possibilités et capacités de traitement que les modèles actuels, mais les fonctionnalités dont elles pouvaient déjà être dotées, amplifiaient alors, de manière spectaculaire, les capacités des traitements calculatoires. Nous rapportons ce que nous écrivions il y a 40 ans. « Aujourd'hui la machine à calculer fait partie de l'environnement de l'élève. Il serait aberrant de l'ignorer. Quel que soit les motifs, sans doute avant tout économiques qui ont poussé le législateur à autoriser son introduction dans l'enceinte de la classe, nous ne pouvons ignorer son existence. Mieux, nous devons tout faire pour qu'elle devienne un outil au service de l'individu » (Régnier, 1981). Pour cela, nous envisagions cet objet sous deux dimensions :

- Réservoir de nombres et de tableaux de valeurs de fonctions mémorisés ;



- Réservoir d'algorithmes opératoires permettant la réalisation de calculs numériques (à l'époque)

À cette période, nous percevions l'opérationnalité de la machine à calculer électronique selon deux directions. La première en tant qu'outil qui rend accessibles de nombreux calculs difficilement réalisables à la main, c'est-à-dire un outil qui sert à résoudre des problèmes. C'était d'ailleurs un des arguments publicitaires à l'époque : avec les calculatrices plus de problème... ! Il est attribué à Albert Einstein cette pensée selon laquelle « Les machines, un jour, pourront résoudre tous les problèmes, mais jamais aucune d'entre elles ne pourra en poser un ! ». Nous en convenons. Toutefois, nous envisagions une seconde direction de nature pédagogique et didactique. La machine à calculer en tant qu'outil d'éducation mathématique, outil d'aide à l'apprentissage des mathématiques en ce qu'elle permet de poser des problèmes

Nous postulions alors que l'exposé magistral dogmatique en tant que modalité pédagogique dominante, ne pouvait en aucun cas pallier la nécessité d'organiser des conditions de manipulation, d'expérimentation pour l'élève, c'est-à-dire la mise en œuvre d'un apprentissage fondé sur le tâtonnement expérimental de l'apprenant selon la conception de la Pédagogie de l'école moderne – Pédagogie Freinet (Regnier, 1988, 1994, 2014). Énoncé dans une forme plus actuelle, il s'agissait d'organiser les conditions de production d'interactions de l'apprenant avec la machine à calculer électronique par l'intermédiaire des entrées dans sa forme la plus simple : appuyer sur les touches, et des sorties : affichages à l'écran. L'idée était de ne pas limiter l'usage de la machine à des calculs pour des calculs. Nous considérons qu'« ainsi l'individu doit-il découvrir les potentialités de l'objet mais aussi ses limites ! Par exemple les limites dues à la capacité, mais aussi celles dues à l'individu lui-même lors de la construction de l'algorithme ou du programme, ou bien de la frappe sur les touches. La machine ne rectifie pas d'elle-même ! Limites posées aussi dans le domaine du raisonnement : extrapolation, interpolation, induction abusive. » (Regnier, 1981) L'objectif pédagogique était de conduire l'élève vers une prise de conscience des limites de l'outil qui

l'amènerait à acquérir une conduite de vigilance, de grande rigueur dans la construction des algorithmes de calcul et des programmes pour les machines programmables, de précaution dans la manipulation. Nous soutenions aussi l'idée que, bien que la réalisation de simulations puisse étayer des raisonnements de type inductif en apportant une masse d'informations en temps restreint bien supérieur aux traitements papier-crayon, cette prise de conscience des limites vise à mieux mettre en évidence l'importance des démarches déductives pour le contrôle de la validité des résultats obtenus c'est-à-dire ce qui s'affiche. De là, une conclusion formulée en référence aux domaines des mathématiques et de la statistique que nous maintenons encore aujourd'hui malgré les évolutions vertigineuses des outils électroniques de calcul numérique et même symbolique : « ... la machine à calculer n'autorise nullement de se libérer de la connaissance des opérations élémentaires, de leurs techniques et de leurs propriétés. Au contraire son usage les rendant nécessaires tout en permettant d'aller plus loin dans le même temps, elles auraient tendance à être mieux assimilées. » (Régner, 1981) Nous allons analyser un exemple extrait d'observations systématiques et d'enquêtes par questionnaire au cours des années 1992 à 2018, dans le cadre des cours de statistique en licence et master de sciences de l'éducation et de la formation. La machine à calculer électronique portable de type scientifique, non nécessairement programmable, faisait partie des outils explicitement requis dans le but de réduire les difficultés et les obstacles posés par les calculs nécessairement requis dans les traitements statistiques. Évidemment les calculatrices auxquelles les étudiants avaient désormais accès, comportaient d'importantes évolutions dans leur puissance de calculs en comparaison de celles évoquées des années 80. Ajoutons aussi l'intégration de ces calculatrices dans les téléphones portables, les ordinateurs portables et de bureau. Illustrons rapidement ce propos : dans les années 80, les calculatrices n'affichaient qu'au plus 10 chiffres. La comparaison des deux fractions suivantes devenait un problème :

$$A = \frac{3223806232}{2564895693} \text{ et } B = \frac{1990538546}{1583694358}$$

Le recours à la calculatrice fournissait le résultat suivant qui s'affichait :  $A=B=1,256895647$  mais c'était oublié que l'affichage était limité à 10 chiffres. Un calcul de division euclidienne réalisé à la main donne les valeurs respectives :

$$A = 1,256895647\mathbf{38988} \text{ et } B = 1,256895647\mathbf{63143}$$

En fait, l'information que donne l'affichage, revient à dire que  $Ent[10^{10}A] = Ent[10^{10}B]$ .

Un autre raisonnement sans recours à la calculatrice conduit rapidement à la conclusion de l'inégalité de ces deux fractions. En utilisant le raisonnement fondé sur le produit en croix, nous constatons que la multiplication de 3223806232 par 158369358 est un nombre dont le chiffre des unités est 6 alors que celle de 2564895693 par 1990538546 conduit à un nombre dont le chiffre des unités est 8. C'est-à-dire que A et B sont distincts.

Notons d'ailleurs que cette limitation de l'outil liée à ses capacités d'affichage ne peut de toute façon pas être abolie. Seule la taille des nombres affichés peut être augmentée en accroissant la précision des résultats des calculs. Pour notre propos, cette limitation doit être prise en considération dans les interprétations des résultats issus de traitements calculatoires, en particulier, dans le domaine de la statistique et se trouve liée à la question des approximations que nous avons abordée (Régnier, 1998).

Repartons de la situation problème telle qu'elle fut proposée lors d'une épreuve d'examen en licence de sciences de l'éducation.

Considérez-vous qu'il existe un profil type d'enfants prédisposés à être en difficulté d'apprentissage ?				
Tableaux statistiques de la variable V 07				
Tableau n° T 7	V07 = profil -type	oui	non	Je ne sais pas
« Maîtres généralistes »	Effectifs	108	160	30
« Maîtres E »	Effectifs	52	57	14

Q 401 Donner la proportion des individus déclarant « oui, il existe un profil type... » dans l'**échantillon** des « maîtres généralistes ».

Q 402 Donner une **estimation ponctuelle** de la proportion des individus déclarant « oui, il existe un profil type... » dans la **population** des « maîtres généralistes »

Q 403 Donner une **estimation par intervalle de confiance à 95%** de la proportion des individus « oui, il existe un profil type... » dans la population des « maîtres généralistes »

Tableau 1 : Situation problème de statistique

La formule servant à répondre à la question Q403 est fournie, avec les conditions de validité de son application, aux étudiants comme un moyen

mnémotechnique dont ils disposent durant l'épreuve où tous les documents sont permis.

$$f_n - k\sqrt{\frac{f_n(1-f_n)}{n-1}} \leq \pi \leq f_n + k\sqrt{\frac{f_n(1-f_n)}{n-1}}$$

Si nous décomposons cette forme langagière afin de recourir à une machine à calculer pour réaliser son calcul, nous identifions les quatre opérations arithmétiques : addition, soustraction, multiplication et division ainsi que l'opérateur racine carrée dont le symbole  $\sqrt{\quad}$  n'est autre que la lettre **r** stylisée. L'autre symbole de la relation d'inégalité  $\leq$  n'intervient pas dans le traitement avec la calculatrice. Pris isolément les opérateurs font partie de la formation scolaire de base. Toutefois nous avons pu observer directement l'effet du tout qui rassemble ces cinq opérateurs, au travers des réactions émotionnelles spontanées que la vue ou l'évocation de cette formule suscitent.

En ce qui concerne l'opérateur *Racine carrée*, l'apport de la calculatrice est, de nos jours, fondamental dans la mesure où la connaissance d'un algorithme pour réaliser sur papier-crayon l'extraction de la racine carrée d'un nombre<sup>4</sup> ne fait plus partie de la formation scolaire de base, de même que l'usage des tables numériques<sup>5</sup> dont nous disposions, il y a 40 ans, a complètement disparu. Sans doute est-il possible de les retrouver parmi les documents mis à disposition sur Internet mais cela n'aurait aucun intérêt. Il est possible aussi d'y trouver des applets qui réalisent ce calcul, sous la forme de calculatrices en ligne.

Revenons à la réalisation du calcul de ces expressions numériques qui conduit à l'obtention de la borne inférieure et de la borne supérieure de l'estimation par intervalle de confiance à 0,95 d'une proportion (dont la valeur de l'estimation ponctuelle est 0,3624). Voici un exemple de traitement effectivement produit.

<sup>4</sup> <https://fr.wikihow.com/calculer-une-racine-carr%C3%A9e-%C3%A0-la-main>

<sup>5</sup> [https://fr.wikipedia.org/wiki/Table\\_num%C3%A9rique](https://fr.wikipedia.org/wiki/Table_num%C3%A9rique)

Formule	$f_n \pm h \sqrt{\frac{f_n(1-f_n)}{n-1}}$
Traitement 1	$0,3624 \pm 1,96 \sqrt{\frac{0,3624 \times 0,6376}{298-1}}$
Traitement 2	$= -1,96 \sqrt{\frac{0,3624 \times 0,6376}{297}}$
Traitement 3	$\approx 0,305$
Traitement 4	$0,3624 - 0,305 = 0,0574$
Traitement 5	soit 5,74%
Conclusion	La population "des maîtres généralistes" qui pensent qu'il y a un profil type est comprise dans l'intervalle [5,74%, 77,74%]

Tableau 2 : exemple de traitement mis en œuvre effectivement avec usage d'une calculatrice

Nous pouvons observer que la formule mathématique (Formule) qui devrait conduire aux résultats valides est formellement respectée. Le premier pas du traitement (Traitement 1) est tout à fait correct, consistant à remplacer les symboles  $k$ ,  $n$  et  $f_n$  par les valeurs numériques qu'ils représentent dans cette situation. Le second pas (Traitement 2) fait apparaître un usage de la calculatrice qui ne tient pas compte des conventions d'écriture du langage mathématique. Ici il n'est pas tenu compte des règles de priorité : celle de

$$1,96 \sqrt{\frac{0,3624 \times 0,6376}{298-1}}$$

l'opération multiplication qui doit être effectuée avant l'opération soustraction. Bien évidemment la calculatrice ne signale pas cette erreur de raisonnement et produit une valeur qui correspond à la différence entre 0,3624 et 1,96. Dans cet exemple, le raisonnement mis en œuvre dans la suite est difficilement interprétable à partir des seules valeurs numériques qui sont fournies. Il aurait nécessité la réalisation d'un entretien d'explicitation car nous n'avons pas réussi à reproduire par diverses combinaisons de calculs, la trace laissée par l'étudiant.

Nous avons ici un exemple paradigmatique de la nécessité d'un minimum de connaissances en mathématiques pour avoir un usage pertinent de la machine à calculer comme un amplificateur (Acioly-Régner, 2007, 2008) (Acioly-Régner, Régner, 2007) (Schliemann, Acioly, 1989) des capacités de

l'individu. Un point important est de remarquer que l'ordre du traitement calculatoire de l'expression numérique  $0,3624 - 1,96 \sqrt{\frac{0,3624 \times 0,6376}{298-1}}$  est l'opposé de l'ordre de son énonciation.

$f_n = \frac{108}{298} = 0,3624$	$n = 298$
$1 - 0,3624 = 0,6376$	$298 - 1 = 297$
$0,3624 \times 0,6376$	
$\frac{0,3624 \times 0,6376}{297}$	
$\sqrt{\frac{0,3624 \times 0,6376}{297}}$	
$1,96 \sqrt{\frac{0,3624 \times 0,6376}{297}}$	
$0,3624 - 1,96 \sqrt{\frac{0,3624 \times 0,6376}{297}}$	
0,3077303749388819...	
Approximativement 30,77%	

Tableau 3 : exemple d'organigramme d'algorithme de traitement à mettre en œuvre dans l'usage d'une calculatrice

### 3 Retour sur la statistique, son enseignement et son apprentissage

Sans entrer dans des détails subtils, nous souhaitons maintenant revenir sur notre conception de la statistique pour mieux situer nos propos réflexifs. Celle-ci conditionne aussi nos conceptions pédagogiques et didactiques à l'œuvre dans l'élaboration des situations d'apprentissage sur lesquelles nous fondons l'organisation de nos séquences d'enseignement de la statistique.

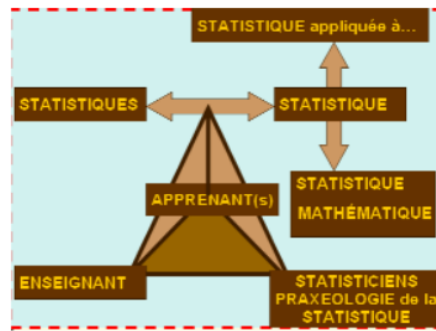


Figure 4 : Schématisation du système didactique usant d'une représentation tétraédrique (Régner, 2005a p.4)

Rappelons ce que nous avons déjà amplement développé dans divers articles et en particulier dans (Régner, 2005a) à propos de nos cadres de références pédagogique et didactique schématisés dans la figure précédente. À côté des trois pôles habituels du schéma systémique triangulaire didactique : enseignant, apprenant(s) et statistique, nous en avons introduit un quatrième, celui des pratiques sociales professionnelles de référence. Ce pôle renvoie aux communautés des statisticiens et à la praxéologie de la statistique. Il représente l'intégration d'un milieu socioculturel et professionnel autre que celui des enseignants, dans le traitement des objets à prendre en considération dans la théorie des situations didactiques de statistique. Le rôle de référence que joue ce milieu peut permettre d'identifier des obstacles macrodidactiques (au sens de Brousseau). La corporéité de ce milieu statistique se manifeste à la fois dans des institutions universitaires et dans des organisations associatives au sein desquelles la science statistique se développe, ses paradigmes se débattent, les pratiques théorisées de la statistique et de son enseignement se confrontent, la formation en statistique y est organisée.

### 3.1 Que peut être la statistique ?

Le pôle *savoir statistique* du système didactique renvoie des questions épistémologiques et méthodologiques récurrentes : qu'est-ce que la statistique ? Parmi les centaines de définitions identifiées depuis plus d'un siècle, celle-ci nous semble parmi les plus pertinentes. La statistique est un cadre théorique qui fournit une sorte de langage commun, méthode générale



reliant divers domaines scientifiques portant sur des ensembles d'individus, de variables et de relations conduisant à des conclusions plutôt vraisemblables et probables que vraies et certaines énonçant des propriétés de groupe valides sur des ensembles parfois mal définis. En tant que domaine scientifique, la statistique se développe sous l'effet d'une tension dialectique entre la statistique mathématique et la statistique appliquée à.... Cette relation dialectique elle-même est en tension dialectique avec un autre pôle constitués par ce qui est désigné par les statistiques au sens des données construites.

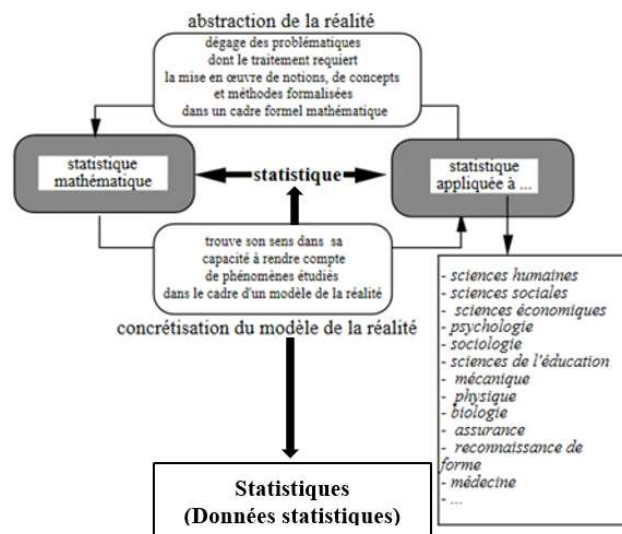


Figure 5 : Schématisation de notre conception de la statistique (Régner, 2002)

Dans les usages courants, nous pouvons d'ailleurs observer l'emploi du terme *statistiques* au pluriel en lieu et place du terme *statistique* singulier, entretenant la confusion entre le cadre théorique et les objets qu'il permet de traiter. Nous avons discuté cette approche dans (Régner, 2002 p.189-191).

### 3.2 Place et rôle de l'informatique dans l'enseignement, l'apprentissage et les usages de la statistique

La place et le rôle de l'informatique peuvent être diversement abordés comme nous tentons de le schématiser dans la figure ci-dessous. L'enseignement de la statistique peut prendre appui sur l'usage de traitements informatiques assistés par ordinateur. Ce recours aux outils informatiques a joué un rôle historique puissant dans le développement même de la statistique. La place qu'ils occupent aujourd'hui, renvoie à une problématique analogue à celle du



rapport entre statistique et mathématiques que nous avons déjà évoqué dans les sections précédentes. Quelles sont donc les connaissances et compétences minimales requises pour que les outils informatiques puissent être utilisés avec efficacité ? Nous avons pu déjà constater, au travers d'observations que nous avons nous-même conduites dans le contexte de l'enseignement de la statistique en sciences de l'éducation, combien le coût de l'apprentissage pour utiliser les outils informatiques peut s'avérer élevé. Cet apprentissage peut même générer une difficulté supplémentaire dans l'apprentissage de la statistique en raison des conditions temporelles imposées des séquences d'enseignement orientées par des critères souvent plus économiques que pédagogiques.

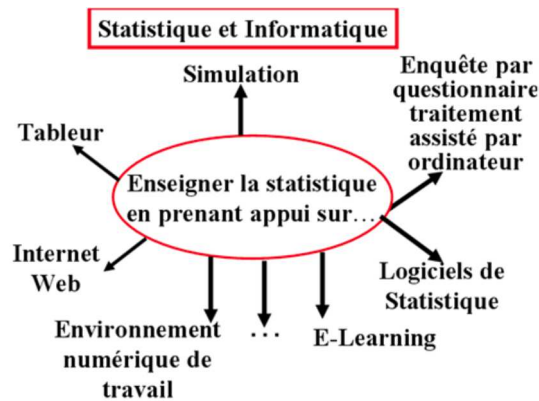


Figure 6 : Informatique dans l'enseignement de la statistique (Régner, 2012)

### 3.2.1 Statistique, Informatique et Simulation

Dans le cadre d'une approche didactique, nous pouvons citer les travaux de Jean-Claude Oriol (2002, 2007) (Oriol et Régner, 2004) qui ont porté sur quelques conditions d'efficacité d'une séquence didactique fondée sur une enquête par questionnaire, ou ceux sur le fonctionnement didactique de la simulation en statistique à propos du concept d'intervalle de confiance (Oriol et Régner, 2003a) ou de celui de corrélation (Oriol et Régner, 2003b).

Dans (Oriol, 2007) Jean-Claude Oriol parvint à cette conclusion :

« Tout au long de ce travail nous avons essayé de montrer comment l'enquête par questionnaires d'une part, et la simulation d'expériences aléatoires d'autre part s'inscrivent dans une perspective constructiviste de la connaissance statistique s'appuyant sur les concepts de schèmes et de situations. L'enquête par questionnaires mobilise la presque totalité des compétences en action dans

le métier de statisticien : collecte des données, nettoyage des données, vérification des cohérences, choix techniques, choix technologiques, traitements, validation auprès des commanditaires. Mais l'appareillage mis en place pour effectuer une enquête est assez conséquent et ne peut donner une proximité suffisante de la variabilité qu'après de multiples enquêtes et des années d'expérience. La simulation permet aux étudiants de dégager des invariants de la variabilité, de comprendre ce que recouvre la convergence, de s'approprier la statistique inférentielle, de concevoir les fluctuations d'échantillonnage.

L'opération est délicate et nécessite d'avoir à batailler sur tous les fronts : celui des signifiés, celui des signifiants et celui des situations de référence. Dans la première enquête (...) on constate que les représentations de la statistique, pour les étudiants du département QLIO qui font peu de statistique, tournent essentiellement autour du calcul numérique et qu'en revanche elles sont multi-référentielles pour les étudiants du département STID qui ont un emploi du temps chargé en statistique, enquêtes et calcul de probabilité. A l'issue de ce travail, nous savons qu'administrer des enquêtes par questionnaires est une tâche fondamentale pour l'appropriation des techniques et du champ de la statistique ; elle ne se résume pas au simple traitement des données et en ce sens permet d'enrichir l'ensemble des situations de référence ; mais nous savons aussi que construire des simulations est une activité fondamentale où le sujet apprenant est obligé de construire et manipuler des signifiants d'objets inconnus, et de ce fait d'accorder une place, à ce que j'ai appelé le non savoir, dans le champ même du savoir.» (p. 258-259)

Jean-Claude Oriol a réalisé dans (Oriol, 2007) un travail minutieux d'analyse sur la simulation et sa mise en œuvre comme moyen d'étayer le processus de conceptualisation chez les étudiants en formation en statistique. S'interrogeant sur la question de la formation citoyenne en statistique, Oriol s'est trouvé confronté à des questions relatives aux concepts à enseigner et aux stratégies pédagogiques fondées sur la transposition didactique et une ingénierie didactique. En abordant le concept de coefficient de corrélation

linéaire, il a construit des situations didactique et pédagogique, non pas en se centrant sur la démonstration mathématique de l'amélioration de la convergence grâce à la transformation de Fisher appliquée à ce coefficient, mais en portant son attention aux explications données aux étudiants dans le programme du DUT STID à l'Institut universitaire de Technologie – IUT – et à leur évolution depuis les années 70. Il a pu alors ainsi constater le passage d'arguments d'autorité couplés à l'outil *abaque*, à une argumentation plus souple fondée en partie sur la simulation. Par ailleurs, il rappelle que la simulation concerne, de nos jours, un nombre sans cesse croissant d'activités industrielles pour deux raisons principales : d'une part, l'accroissement de la puissance de calcul des ordinateurs à moindre coût, d'autre part, celui de l'importance et de l'efficacité des modèles théoriques concernant les divers domaines de la production et de l'exploitation. Pour simplifier, il précise que l'on fait une simulation lorsqu'on expérimente non pas sur l'objet réel mais sur un modèle.

Pour réaliser les simulations, Oriol intègre dans son ingénierie pédagogique « l'outil informatique comme instrument canonique d'une pratique de la statistique. Et en matière d'outil informatique, le tableur [ici Excel] est privilégié dans un premier temps » (Oriol 2007 p.92) L'activité s'organise alors autour des interactions entre l'étudiant ou le groupe d'étudiants et l'ordinateur en tant que calculateur automatique dans l'interface constituée par le tableur.

### **3.2.2 Formation universitaire à la statistique et activités collaboratives médiées par les TICE**

Diane Diaz (Diaz, 2016) s'est intéressée aux facteurs pouvant influencer la réussite des activités collaboratives médiées par les TICE dans une situation de formation universitaire à la statistique. Ses travaux de recherche et leur objet trouvent leur raison d'être dans ce postulat initial « Le développement d'Internet et des TIC produit de profondes mutations sociétales. Les matériels informatiques, devenus courants, engendrent de nouveaux usages dans notre vie quotidienne et professionnelle. L'accroissement du numérique génère ainsi une augmentation exponentielle d'informations collectées, le plus

souvent automatiquement par l'outil numérique qui, tout à la fois, produit, stocke, synthétise et transforme les informations. Il ne fait aucun doute que notre société produit de plus en plus de données qui doivent être synthétisées et analysées à l'aide d'algorithmes fondés sur des modélisations issues de la statistique. Cette dernière permet de donner du sens à des nombres qui, sans leur traitement, resteraient inintelligibles, et aide à la compréhension de phénomènes grâce aux comparaisons et aux combinaisons qu'elle consent. Elle regroupe un ensemble de méthodes et d'outils utilisés pour quantifier, mesurer, décrire et interpréter. Ces divers traitements sont à la fois essentiels pour la compréhension d'un phénomène et pour la prise de décision qu'elle infère. Pour toutes ces raisons, il paraît utile d'enseigner au plus grand nombre les rudiments de la statistique. » (Diaz, 2016, 16) Nous pourrions dire que Diane Diaz a considéré que les outils informatiques à l'œuvre dans leur mise en réseau comme Internet ou encore dans ce qu'il est convenu d'appeler les TIC sont à même de produire des données d'interaction stockables dans des dispositifs informatiques, et dont l'intelligibilité serait améliorée pour le citoyen par une meilleure formation en statistique de celui-ci. Pour ce faire elle a choisi comme terrain d'étude, un cours de formation universitaire à la statistique en Master de Sciences de l'éducation conduit sous la responsabilité d'un professeur et au sein duquel elle a réalisé des observations et des enquêtes par questionnaire et par entretien, répétées quatre années durant. Les données obtenues ont été construites à partir d'activités collaboratives autour de situations problèmes dont la résolution mobilisait un logiciel de statistique implémenté sur un ordinateur dans une salle dédiée de l'université (salle informatique). L'analyse données construites par entretien met en évidence les conditions facilitatrices telles que perçues et exprimées par les étudiants à l'égard de la formation à la statistique. Ces entretiens ont permis d'identifier plusieurs facteurs. « Ceux d'ordre biographique relèvent principalement de l'expérience précédente, la connaissance antérieure et le temps d'exposition ; ceux d'ordre psychologique sont essentiellement la stimulation, l'aide, l'entraide, l'échange et le partage. Enfin, les principaux facteurs cognitifs

évoqués présentent des similitudes avec les notions d'étayage, de conflit sociocognitif et de métacognition. » (Diaz, 2016, p.266)

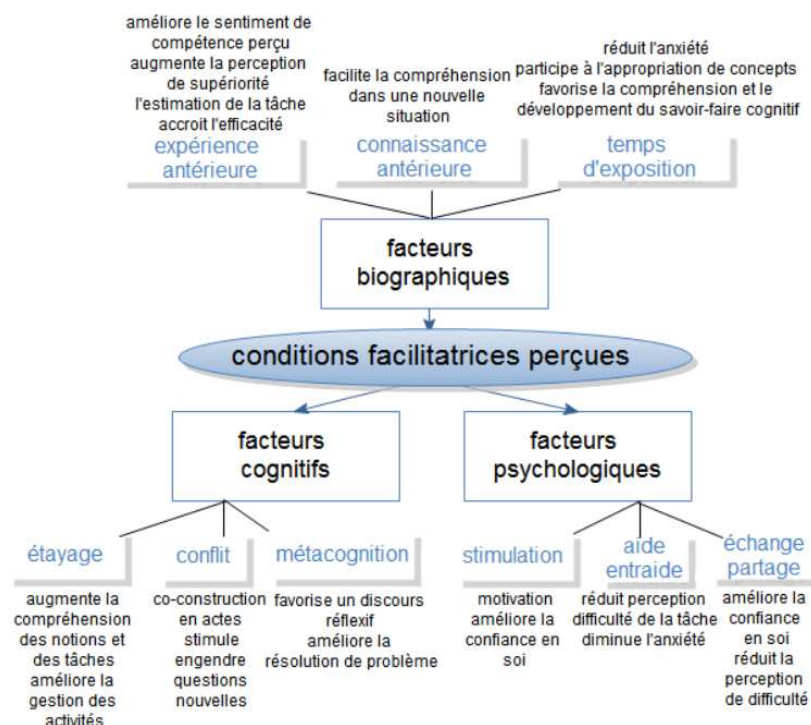


Figure 7 : Les conditions facilitatrices évoquées par les étudiants selon Diaz (2016 p.266)

### 3.2.3 Enseignement-apprentissage de la statistique, TICE et environnement numérique de travail

Elayne de Moura Braga (Braga, 2009) a conduit une recherche sur les effets de l'usage de supports didactiques numériques sur la conceptualisation en statistique. Ces supports constituaient un des moyens techniques de l'ingénierie pédagogique mise en œuvre pour la formation à la statistique dans un dispositif de formation à distance en licence de sciences de l'éducation intégré au campus numérique FORSE<sup>6</sup> dont la création à l'origine de laquelle nous avons contribué en partenariat avec des collègues de l'Université de Rouen et du CNED – Centre National d'Enseignement à Distance de Poitiers, avait débuté en 1997. Nous entendons ici l'ingénierie pédagogique dans un sens assez complet tel que défini dans le J.O. (du 11 septembre 1992 p.

<sup>6</sup> FORSE Formation et Ressources en Sciences de l'éducation. <http://www.sciencedu.org/>

12522). : « l'ensemble des activités de conception, d'étude, de projet, de réalisation, d'aide au fonctionnement et d'évaluation des moyens techniques d'enseignement et de formation. » Pour mieux situer les travaux de Elayne de Moura Braga, nous renvoyons le lecteur à (Régner, 2003, 2005b). Nous tentons cependant d'évoquer le contexte mais de manière toutefois très succincte. Par les deux figures ci-dessous, nous avons schématisé, sous une forme systémique, le dispositif pédagogique et didactique organisé autour de relations triangulaires à partir d'une extension du triangle didactique dont nous avons déjà parlé plus haut.

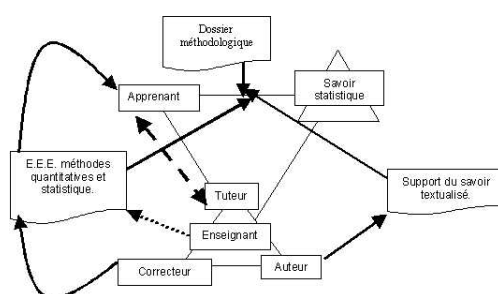


Figure 8 : représentation schématique du système didactique et pédagogique d'enseignement de la statistique dans le dispositif de formation à distance Campus Numérique FORSE (Régner, 2003, 2005a)

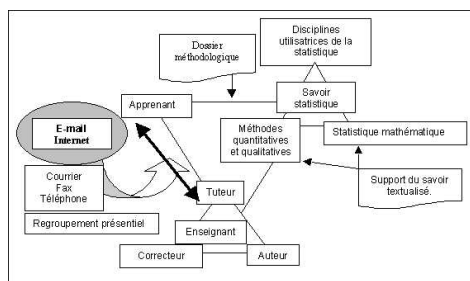


Figure 9 : représentation schématique du système didactique et pédagogique d'enseignement de la statistique dans le dispositif de formation à distance Campus Numérique FORSE intégrant les modalités d'interaction à distance (Régner 2003, 2005a)

En 2000, à l'origine du dispositif FAD<sup>7</sup>, ce qui est dénommé *support du savoir textualisé*, était un support de cours sous la forme traditionnelle d'un ouvrage imprimé dont disposait chaque étudiant. Associé à cet ouvrage, les étudiants disposaient d'un livret de situations-problèmes désigné par EEE (exercice d'entraînement à l'examen) qu'ils devaient résoudre par écrit soumis ensuite à la correction du tuteur du groupe auquel ils étaient rattachés. Des

<sup>7</sup> Formation A Distance

interactions centrées sur les contenus statistiques se produisaient entre étudiants et tuteurs par courriel, téléphone, fax ou même courrier postal. Le développement rapide des TICE intégrées à l'organisation d'un campus numérique a évidemment modifié les modalités des formes d'interaction. Nous avons rapporté dans (Régnier, 2005b) les résultats des analyses de données construites dès le début de l'ouverture du campus numérique FORSE en septembre 2002. Les années suivantes un travail d'élaboration de supports de cours numérique a été conduit par l'équipe des enseignants – auteurs de cours. Le cours de statistique intégré aux cours dénommés *méthodes quantitatives*. Le contenu de l'ouvrage a été mis sous la forme d'un document numérique qui intégrait des liens hypertextes et des tableaux interactifs qui le différenciaient alors du support imprimé ou même du document numérique en pdf sans hyperlien dont ont continué à disposer les étudiants. C'est l'usage de ce support numérique interactif conçu comme un outil d'aide à l'apprentissage de la statistique qui fut un des objets de la recherche conduite par Elayne de Moura Braga. Cette recherche avait pour but d'identifier les « effets médiateurs des supports didactiques numériques dans la conceptualisation en Statistique et leur besoin de règles d'action afin de devenir de bons médiateurs ». Elayne de Moura Braga précise que « la médiation et les règles d'action sont des concepts essentiels dans la conception de l'architecture des logiciels destinés à l'éducation, d'un point de vue cognitif en ergonomie. Ce besoin existe car c'est à travers les règles d'action que le logiciel pourra prendre en compte les représentations affectives et l'attribution de la causalité de l'échec. Le support numérique *Méthodes Quantitatives FORSE*, mis à disposition des étudiants de Licence Formation à Distance (FAD) en Sciences de l'Éducation a été le logiciel que [elle a] utilisé dans cette recherche et auquel [elle conseille] une expansion afin de devenir plus qu'un médiateur suffisamment bon en tant que matériel numérique concernant la statistique, mais un bon médiateur. » (Braga, 2009, p.186). Elayne de Moura Braga nous rappelle alors que « l'insertion des technologies de l'information et de la communication dans l'éducation (TICE) implique un nouveau paradigme dans le processus d'enseignement-

apprentissage. Le rôle de chaque pôle du triangle pédagogique-didactique se modifie pour donner place à un quatrième pôle avec l'insertion de l'environnement didactique numérique. » (*Ibidem*) Partant de l'idée que la conception d'un outil didactique virtuel comme un médiateur doit prendre en compte tout à la fois son rôle de porteur du contenu et celui de facilitateur de l'apprentissage, elle a réalisé une analyse de l'interface de l'outil « Méthodes Quantitatives FORSE », en prenant appui sur le domaine de l'ergonomie cognitive mais sur l'idée de « bon médiateur » dans le sens de Feuerstein. Cette analyse est complétée par celle de données construites par une observation de l'utilisation de l'outil et une enquête par questionnaire portant sur son ergonomie et sur son utilisation. Elle formule alors cette conclusion que « Le support didactique numérique *Méthodes Quantitatives FORSE* est très peu utilisé. Cette donnée était déjà connue tout au début de [la] recherche, [puisque c'est] ce qui [avait] constitué une des motivations pour l'étude de cet outil. » Ainsi il ressort, par exemple, que 79% des individus répondent positivement à la question *Savez-vous qu'il existe un cours de méthodes quantitatives en ligne ?* mais que plus des 3/4 des individus déclarent ne l'avoir jamais l'utilisé. Parmi ceux qui l'ont utilisé, pour la majorité ce ne fut qu'« une fois pour voir ». Elle constate aussi que près 1/3 des individus déclarent ne pas être du tout satisfait du support numérique comme outil d'étude de la Statistique, donnant de mots avec connotation négative de ce support. Parmi les nombreux résultats auxquels elle était parvenue en 2009, nous rappelons cette remarque que « le support didactique numérique *Méthodes Quantitatives FORSE*, pour exercer un rôle de bon médiateur de la Statistique, doit être modifié et adapté aux caractéristiques que notre recherche a pointé à partir de l'ergonomie cognitive et de la définition de la bonne médiation par des systèmes informatiques. » Il s'agissait alors d'une recommandation dont il serait judicieux d'en étudier les effets plus de 10 années après !



### 3.3 Autour de la question de l'interprétation statistique des résultats issus de traitements assistés par ordinateur ou calculatrice électronique

L'utilisation des logiciels de traitement statistique ou des tableurs conduit à obtenir des réponses, les résultats des traitements calculatoires qui vont être convertis dans différents registres de représentations sémiotiques. Pour préciser notre pensée, nous rapportons la synthèse faite par Diane Diaz (2009, p.270-271). Selon ses analyses, pour un même concept de statistique, elle a pu identifier sept formes de représentation qu'elle a ainsi caractérisées :

Forme	Caractérisation
déclarative	Comprend le vocabulaire ainsi que les définitions afférentes aux concepts de statistique.
mathématique	Correspond aux algorithmes, aux formules mathématiques.
graphique	Représente des données au travers de figures illustratives.
opérationnelle	Contient les étapes des opérations à produire.
procédurale	Correspond à la pratique des calculs.
numérique	Se limite à la valeur numérique, exprimée par des nombres
fonctionnelle	Se rapporte à l'analyse et interprétation des résultats.

Tableau 4 : formes d'expression d'un concept de statistique

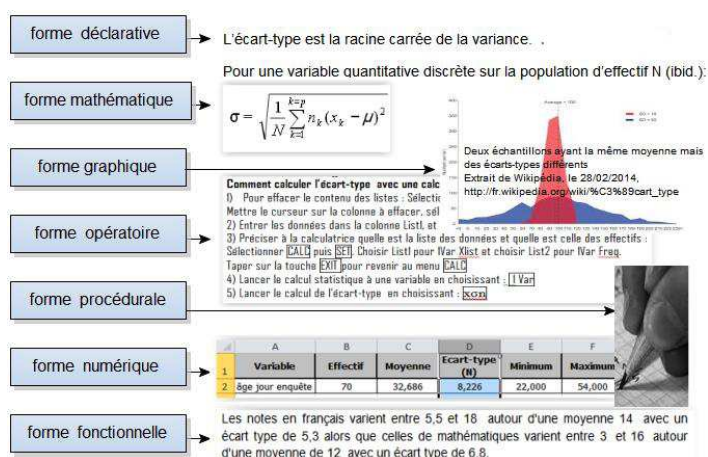


Figure 10 : Les principales formes concernant la notion d'écart-type (d'après Diaz 2009, p.271)

L'utilisateur d'un logiciel dit de statistique, d'un tableur ou même d'une calculatrice électronique qui comportent des fonctions statistiques est alors confronté à l'interprétation des résultats issus des traitements et analyses des données qui apparaissent, par exemple, dans le rapport sous forme de fichiers au format Excel. Dans notre propos, l'utilisateur est un individu en situation

d'apprentissage. Cet apprentissage est supposé se réaliser en partie sur la base de résolutions de problèmes de statistique, non pas sous la forme de problèmes de statistique mathématique mais bien de situations problèmes de statistique appliquée à ... Par exemple, comme dans les travaux de Jean-Claude Oriol (Oriol, 2007) la réalisation complète d'une enquête par questionnaire. Pour revenir à l'interprétation statistique, nous reprenons ce que nous avons développé à plusieurs reprises dans divers articles (Régnier, 2002, 2005a) en nous appuyant sur une idée exprimée par Brigitte Escofier et Jérôme Pagès (Escofier & Pagès 1990 p.217-218), nous la caractérisons selon trois directions :

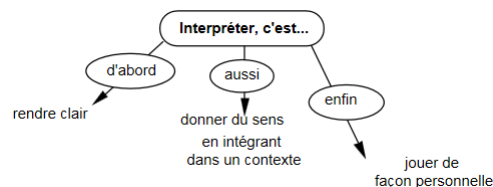


Figure 11 : Interpréter ce peut être ... (Régnier, 2005a)

Dans le cadre de l'enseignement universitaire, la question de l'interprétation statistique émerge rapidement chez les étudiants non spécialistes de la statistique. Prenons comme exemple, celui d'une variable quantitative : poids d'un individu sur une période donnée. Celle-ci peut être considérée comme une variable quantitative discrète ou continue, mais le recours à un logiciel de statistique repose sur des algorithmes de calcul informatisés et de ce fait les traitements calculatoires seront toujours de type discret. Les valeurs numériques (poids en kg) qui vont alors figurer dans la base de données vont être soumises aux traitements algorithmiques informatisés d'une manière totalement invisible pour l'utilisateur sans aucune gêne pour lui, car le dispensant de la réalisation d'opérations numériques fastidieuses quand elles sont faites à la main avec papier-crayon. Si nous nous limitons aux valeurs habituelles résumant la variation sur la base de l'échantillon construit telle la moyenne (poids moyen), la variance et l'écart-type, il s'agit d'algorithmes de calcul des opérations représentées symboliquement (forme mathématique selon Diaz)

$m = \frac{1}{N} \sum_{i=1}^N o_i = \frac{1}{N} \sum_{k=1}^p n_k x_k = \sum_{k=1}^p f_k x_k$ <p>Moyenne</p>	$V(x) = \frac{1}{N} \sum_{k=1}^p n_k (x_k - m)^2$ <p>Variance</p>	$\sigma = \sqrt{\frac{1}{N} \sum_{k=1}^p n_k (x_k - m)^2}$ <p>Écart-type</p>
---	---	--

Tableau 5 : Formules mathématiques de la moyenne, variance et écart-type d'une variable quantitative (forme mathématique selon Diaz)

Imaginons une enquête réalisée par un individu ayant procédé à une mesure quotidienne de son poids au lever durant une période d'un an. Supposons qu'il ait recueilli un échantillon de 277 mesures. Enfin ayant eu recours à outil informatique, par exemple, avec le logiciel Tableur Excel, il peut recourir aux fonctions programmées avec leur dénomination dans la langue française :

=MOYENNE(B2:B279) Moyenne	=VAR.P(B2:B279) Variance	=STDEVPA(B2:B279) Écart-type	=RACINE(VAR.P(B2:B279) Écart-type
------------------------------	-----------------------------	------------------------------	-----------------------------------

Tableau 6 : Fonctions programmées dans le logiciel Tableur Excel pour calculer moyenne, variance et écart-type d'une variable quantitative (forme opératoire selon Diaz)

Notons que la seconde procédure de calcul de l'écart-type requiert un niveau de connaissance et de compétence de l'utilisateur plus élevé en mathématique que pour le recours à la première procédure, car elle repose sur la composition de deux fonctions. Supposons maintenant qu'il obtienne les valeurs suivantes affichées dans les cellules d'un tableau du Tableur :

Moyenne m=	Variance v(x)=	Écart-type σ=
91,3711191335742	0,61988791721513	0,7873296115447

Tableau 7 : valeurs de la moyenne, variance et écart-type d'une variable quantitative (forme numérique selon Diaz)

Comment sont alors lues et interprétées ces valeurs numériques ? que représentent-elles ? que signifient ces nombres ?

Nous avons pu observer lors des nombreuses séances de cours que nous avons assumées durant les trente dernières années au cours desquelles les outils informatiques ont connu un développement spectaculaire, que dès que le niveau de formation en statistique permet de maîtriser quelques-uns des outils conceptuels et techniques de base de la statistique, alors les étudiants soulèvent une série de questions qui ont en filigrane celle de l'interprétation des résultats des traitements des données construites et analysées résumés sous divers énoncés.

Le discours tenu pourrait être ainsi résumé : *A partir du tableau de données construites par pesage quotidien, j'ai calculé la moyenne, la variance et l'écart-type ! Mais les nombres que j'obtiens, j'en fais quoi ? ça veut dire quoi ? ...* Nous avons même pu observer que, dans le rapport d'étude que cet utilisateur aura réalisé, ces nombres figureront tels quels au nom de la rigueur scientifique et du profond respect fondé sur la croyance de ce qu'une machine électronique informatique ne peut se tromper ni ne saurait être taxée de subjectivité dans ce qu'elle produit, c'est-à-dire que *le poids moyen est de 91,3711191335742 kg avec une fluctuation moyenne de 0,7873296115447 kg !*

Poursuivons alors notre exposé avec l'exemple ci-dessus qui peut d'ailleurs correspondre à une pratique quotidienne de contrôle du poids chez un individu. En ce qui concerne la moyenne de la série des 277 mesures en kg, l'algorithme de son calcul est largement maîtrisé par la population des étudiants auxquels nous nous adressons. La moyenne est d'un usage culturel très répandu, déjà dans l'espace scolaire très tôt dans la vie des individus scolarisés. Ils savent que cela revient à additionner les 277 nombres puis à diviser la somme obtenue par le nombre 277. Notons alors combien le rôle d'un outil informatique est celui d'un amplificateur culturel (Acioly-Régnier, 2007, 2008) (Acioly-Régnier, Régnier, 2007) (Schliemann, Acioly, 1989) dans la mesure où il permet à l'utilisateur de réaliser des tâches portant sur des objets culturels (les concepts de statistique) plus facilement qu'avec un papier-crayon. Certes l'utilisation d'un ordinateur ou d'une calculatrice requiert de construire la base de données en entrant les 277 nombres, mais ce travail n'est à réaliser qu'une seule fois. Les saisies se font sur un clavier mais sans doute peuvent-elles être réalisées vocalement. Le traitement manuel avec papier-crayon est, bien sûr dans ce cas, réalisable mais il est facile d'en évaluer le coût bien plus élevé. La différence se situe au niveau de l'explicitation de la procédure d'enchaînement des opérations d'addition puis de la division. Rappelons que cette dernière opération reste, malgré l'allongement de la formation scolaire, une opération arithmétique mal maîtrisée. Nous avons conduit une recherche à la fin des années 80 sur cette

question (Régnier, Denis 1990) de même que Jeanne-Gabrielle Guet qui est partie du constat qu'« effectuer une division arithmétique est un objectif qui se heurte à des obstacles liés à l'apprentissage et à la technique opératoire de la division elle-même. » (Guet, 1994) La tâche de calcul de la variance et de l'écart-type requiert un niveau de connaissance et de compétence en calcul supérieur à ce qui est requis pour le calcul de la moyenne. Par ailleurs, dans la forme prédicative (Vergnaud, 2001) de la connaissance, la variance et l'écart-type conduit à une énonciation qui va dans le sens opposé à ce que requiert la forme opératoire (Vergnaud, 2001) de cette connaissance. Pour résoudre le problème du calcul de l'écart-type qui se déduit de la variance, nous pouvons réaliser l'algorithme suivant comme nous l'avons déjà abordé dans le tableau 3.

Calculer la moyenne	$m = \frac{1}{N} \sum_{i=1}^N o_i = \frac{1}{N} \sum_{k=1}^p n_k x_k = \frac{\sum_{k=1}^p f_k x_k}{N}$
Calculer les écarts des N valeurs $o_i$ à la moyenne	$o_i - m$
Calculer les carrés des écarts des N valeurs $o_i$ à la moyenne	$(o_i - m)^2$
Calculer la somme des carrés des écarts des N valeurs $o_i$ à la moyenne	$\sum_{i=1}^N (o_i - m)^2$
Calculer la moyenne des carrés des écarts des N valeurs $o_i$ à la moyenne. Le résultat obtenu est la variance	$\frac{1}{N} \sum_{i=1}^N (o_i - m)^2$
Calculer la racine carrée de la variance. Le résultat obtenu est l'écart-type	$\sqrt{\frac{1}{N} \sum_{i=1}^N (o_i - m)^2}$

Tableau 8 : algorithme du calcul de la variance et écart-type d'une variable quantitative discrète (forme déclarative et forme mathématique selon Diaz)

Les valeurs numériques rapportées dans le tableau 7 sont les produits de la mise en œuvre de ces algorithmes. Mais alors que signifient ces nombres ? Le calculateur électronique qui opère d'une manière invisible pour l'utilisateur ne peut en aucun cas apporter une réponse. L'utilisateur pourrait s'engager dans une simulation des effets des variations des valeurs numériques d'entrée sur les valeurs de sortie, c'est-à-dire en recourant à l'interactivité possible entre l'être humain qu'il est et la machine. Mais la probabilité qu'il puisse ainsi retrouver les algorithmes mathématiques qui produisent ces résultats est vraiment quasi-nulle. C'est pour nous, une situation qui, pour être banale,

n'en conduit pas moins à une question fondamentale qui nous préoccupe depuis fort longtemps : quelles sont les connaissances et compétences minimales en mathématiques au sein desquelles nous plaçons les théories des probabilités, en statistique et en informatique requises pour faciliter la compréhension et l'interprétation des résultats issus de traitements et d'analyses de données assistés par ordinateur dans le cadre théorique de la statistique appliquée à des objets et des faits d'un autre domaine de connaissance ou d'action ?

À ce jour, nous ne sommes toujours pas parvenu à fournir une réponse précise et stable, et encore moins exhaustive, à cette question. Elle est pourtant fondamentale dans la conception des situations d'enseignement-apprentissage de la statistique pour identifier et expliciter les prérequis cognitifs minimaux pour faciliter l'engagement des apprenants dans le processus d'apprentissage. Selon nous, cette question doit être abordée d'une manière multidisciplinaire et multiréférentielle prenant en considération les trois dimensions cognition, affectivité et culture comme le propose et le développe Nadja Acioly-Régnier (Acioly-Régnier, 2010) En fait, nous la (re)formulons ici pour attirer l'attention des concepteurs de situations d'enseignement de la statistique intégrant les machines informatiques pour soutenir le processus d'apprentissage.

Essayons de réaliser une analyse *a priori* de cette situation d'interprétation statistique d'un résultat à la lumière d'une grille que nous allons présenter ci-après mais aussi des données d'observation, des données d'interaction et des données d'enquête par questionnaires construites durant nos séquences d'enseignement universitaires.

Nous nous appuyons sur les notions de modèle et modélisation en statistique telles que nous les avons abordées dans (Régnier, 2005a). En fait, d'un point de vue tant épistémologique que méthodologique, il nous semble que nous n'étudions le monde que par l'intermédiaire de modèles de cette réalité. Le concept même de modèle statistique et le processus de modélisation peuvent être abordés à différents niveaux de conceptualisation et selon diverses perspectives. Nous revenons à notre conception de la statistique afin d'y

ajouter une précision sur l'orientation décisionnelle qui détermine notre rapport à la statistique et à ses usages. Cela revient à compléter la définition donnée plus haut en considérant la statistique comme un domaine réunissant un ensemble de méthodes permettant de prendre des décisions « suffisamment bonnes » en situation incertaine. Dans ce cadre de l'incertitude et d'un point de vue praxéologique, cette orientation s'appuie sur l'idée générale que les méthodes statistiques pour étudier un phénomène consistent à associer un modèle aléatoire à ce phénomène, préciser ce modèle par l'observation, utiliser ce modèle pour prendre une décision.

Explicitons alors les cadres théoriques auxquels sont empruntés les éléments constitutifs du modèle construit par le chercheur pour étudier un fait ou un phénomène, objet central d'une problématique, elle-même résultat d'une construction humaine. Dans notre contexte universitaire, le chercheur est alors un étudiant qui conduit des travaux d'étude et de recherche qui constituent ensuite la matière du mémoire de Master ou de la thèse de Doctorat en sciences humaines et sociales.

Dans (Régnier, 2005a), nous avons trois cadres théoriques imbriqués :

- $\mathcal{ET}(T)$  qui est celui dans lequel est posée la problématique, par exemple, la didactique des mathématiques, la biométrie, etc. ;
- $\mathcal{ET}(S)$ , celui de la statistique ;
- $\mathcal{ET}(M)$ , celui des mathématiques auquel est inclus les théories des probabilités

Nous y ajoutons ici celui de l'informatique  $\mathcal{ET}(I)$  en tant que domaine d'activité scientifique et technique

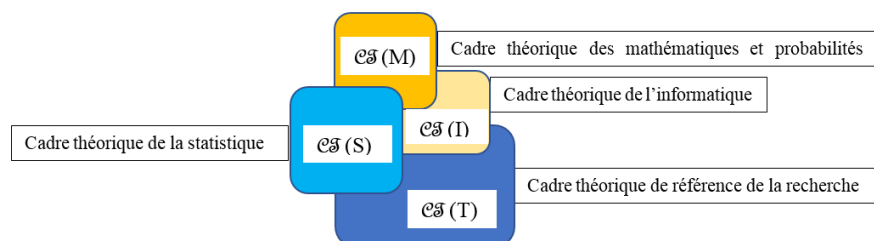


Figure 12 : Schématisation des cadres théoriques de référence imbriqués à prendre en compte pour interpréter

La conduite d'une recherche qui se fonde sur une approche statistique instrumentée par l'informatique passe globalement par l'exécution des tâches suivantes.

Code	Tâches à réaliser
[TS01]	Problématiser dans un cadre théorique T pertinent pour l'étude du phénomène qui requiert un modèle Mod(T)
[TS02]	Construire un modèle Mod(S) dans le cadre de la statistique qui intègre un modèle Mod(M) du cadre des mathématiques.
[TS03]	Formuler des énoncés hypothétiques dont la mise à l'épreuve par une approche statistique est pertinente
[TS04]	Construire les données valides, fiables et pertinentes à partir d'un protocole explicite et congruent aux modèles Mod(S) et Mod(T) et pouvant recourir à des modèles Mod(I)
[TS05]	Traiter les données dans le cadre du modèle Mod(S) et du modèle sous-jacent Mod(M) mais aussi éventuellement des modèles Mod(I) compte tenu que les logiciels de statistique comportent la possibilité à l'utilisateur de programmer lui-même dans un langage informatique des scripts d'algorithme (par exemple en langage Python ou R)
[TS06]	Interpréter dans le cadre du modèle Mod(S) et du modèle sous-jacent Mod(M) en tenant compte des modèles Mod(I)
[TS07]	Interpréter dans le cadre du modèle Mod(T)
[TS08]	Décider dans le cadre de modèle Mod(S) et du modèle sous-jacent Mod(M)
[TS09]	Décider dans le cadre du modèle Mod(T)

Tableau 9 : Tâches à réaliser dans une approche statistique pour l'étude d'un phénomène.

Revenons à notre exemple d'étude relative à la variabilité du poids d'un individu. Quelle interprétation pouvons-nous donner au résultat : la moyenne de la série des mesures pondérales en kilogramme avec une précision d'un hectogramme est 91,37111913357420 kg. Au niveau du cadre théorique des mathématiques  $\mathcal{E}(M)$ , il ne s'agit que d'un nombre obtenu par le produit des deux opérations addition et division. Ce nombre peut aussi être interprété comme la fraction  $\frac{25401,1}{277}$  et donc ce que fournit le calculateur informatique est une valeur décimale approchée à  $10^{-13}$  près. Rien d'autre ne peut être dit dans ce cadre  $\mathcal{E}(M)$ . C'est un nombre en tant que concept mathématique dont nous pouvons toutefois comprendre qu'il correspond à la mesure d'une quantité. Dans le cadre théorique de l'informatique  $\mathcal{E}(I)$ , l'interprétation de ce nombre en tant que produit du traitement des données par des opérations



programmées dans des langages informatiques mis en œuvre par l'intermédiaire de la fonction =MOYENNE(B2:B279) qui demeure la seule partie visible, relève d'une approche que nous ne développerons pas ici mais qu'il ne faut pas oublier comme nous l'avons déjà évoqué à propos des limites des outils informatiques. Par exemple le nombre exact représenté sous la forme fractionnaire  $\frac{25401,1}{277}$  peut aussi l'être sous la forme d'une écriture décimale illimitée. Le logiciel informatique ne peut en aucune façon opérer sur la valeur exacte du nombre. Nous avons étudié cette question des risques d'erreurs dues aux approximations dans le cas de l'usage de calculateurs électroniques (Régner, 1998)

Au niveau du cadre théorique de la statistique  $\mathcal{C}\mathcal{I}(\mathcal{S})$ , il s'agit alors de la moyenne qui est un concept de statistique. Ici plus particulièrement, il s'agit de la moyenne arithmétique, c'est-à-dire qu'il en existe d'autres comme la moyenne géométrique ou la moyenne harmonique ! Comme nous l'avons déjà signalé, la connaissance d'une procédure de calcul de la moyenne semble se construire très tôt dans le cadre des pratiques scolaires d'évaluation : les enseignants attribuent des notes considérées comme des nombres sur une échelle allant de 0 à 20 ou de 0 à 10 dans le contexte culturel scolaire français, puis calculent des moyennes de notes et les élèves calculent la moyenne de leurs notes. Comment font-ils ? par la sommation des notes suivie de la division par le nombre de notes additionnées ! Mais cela ne semble pas suffisant pour parvenir à un niveau de compréhension suffisant qui facilite l'interprétation. Il nous semble que cela passe par un niveau de conceptualisation qui intègre une question historique et épistémologique : la recherche d'un nombre situé à une position optimale dans une série de nombres, c'est-à-dire qu'en considérant les écarts entre ce nombre et chacun de ceux de la série, leur somme soit minimale. Pour résoudre ce problème, il faut alors retourner au cadre  $\mathcal{C}\mathcal{I}(\mathcal{M})$  et recourir aux représentations symboliques du langage mathématique. Nous explicitons la procédure en prenant un exemple limité mais qui vaut généralité. Considérons la série des cinq nombres suivants : 2 ; 5 ; 5 ; 7 ; 8. Nous recherchons alors un nombre que nous nommerons  $m$  tel que la somme suivante soit minimale :  $S_{\text{écarts}} = (m -$

$2)+(m-5)+(m-5)+(m-7)+(m-8)$  que nous pourrions aussi écrire :  $S_{\text{écarts}} = (m-2)+2(m-5)+(m-7)+(m-8)$  ou encore  $S_{\text{écarts}} = 5m-27$ . La valeur minimale de  $S_{\text{écarts}}$  est 0. Cette valeur est obtenue en résolvant une équation du 1<sup>er</sup> degré à une inconnue  $5m-27=0$  c'est-à-dire  $m=\frac{27}{5}$  ou encore  $m=5,4$ . En généralisant la démarche à une série quelconque de N nombres  $O_i$  nous retrouvons une formule présentée dans le tableau :  $\frac{1}{N} \sum_{i=1}^{i=N} O_i$

Il s'agit en fait d'un point d'équilibre que le physicien appelle centre de gravité ou le mathématicien, barycentre. Peut-on aller au-delà de cette explicitation ? Pour notre part, nous ne savons pas de quelle façon.

Au niveau du cadre théorique  $\mathcal{CS}(T)$  dans lequel est réalisée la recherche, ici ce pourrait être celui de l'éducation à la santé ou autre domaine dans lequel le poids d'un individu est une des variables d'intérêt pertinentes, il s'agit toujours d'un nombre résultant d'un traitement mathématique informatisé qui correspond à un résumé d'information statistique en tant que centre de gravité de la série des données numériques. Toutefois son statut se précise puisqu'il s'agit d'une quantité qui traduit une mesure de poids d'un corps humain en unité kg et sous-unité hg. Cette précision est celle donnée par une balance pèse-personne habituelle, elle-même outil électronique informatisé, affichant par exemple



Figure 13 : Affichage de la mesure

La procédure de calcul de la moyenne conduit à l'obtention d'un nombre avec les mêmes unités de mesure. C'est-à-dire que la moyenne est donnée en unité kg. Force est alors de constater que le traitement par le logiciel conduit à un résultat donné sous la forme d'un nombre décimal avec 13 chiffres après la virgule, c'est-à-dire une précision de  $10^{-13}$ . Dit autrement, le poids moyen est donné ainsi dix mille milliardièmes de kg soit un dixième de picogramme. Cela ne peut prendre aucun sens. Ainsi l'utilisateur doit procéder à la tâche [TS07] Interpréter dans le cadre du modèle Mod(T) en remplaçant dans le

contexte de l'étude pour donner du sens (Figure 11). Une conclusion pertinente qui a un sens dans ce contexte d'étude, est d'indiquer que le poids moyen vaut entre 91,3kg et 91,4kg. La suite de l'interprétation dépend alors de la problématique posée issue de [TS01]. Mais si nous intéressions aux pratiques sociales observables, nous pourrions entendre dans les propos quotidiens, des discussions sur les questions du corps et, en particulier, de son poids, qui puisent des arguments dans cette information apportée par le poids moyen. Nous n'évoquerons pas les outils connectés capteurs quotidiens de multiples données mesurables. Que font les utilisateurs de ces données massives qui les concernent directement ? Si les outils connectés possèdent des logiciels intégrés de traitement de données, quelle marge de contrôle l'utilisateur détient-il ? Quelle fiabilité, validité et pertinence ont les résultats issus de traitements dont les algorithmes sont totalement invisibles ? Sur quels critères, la décision formulée est prise ? Ces questions surgissent dans l'air et l'esprit du temps. Ces phénomènes désignés par données massives, *big data*, réorientent même la définition du domaine de la statistique appelée science des données, *data science* et le statisticien devient un *data scientist*. Nous revenons donc à la question des connaissances et compétences minimales qu'un étudiant mais aussi un citoyen doit acquérir et développer dans les domaines des mathématiques, de la statistique et de l'informatique. Il conviendrait de poursuivre l'analyse commencée sur la moyenne en nous focalisant sur la variance et l'écart-type. Des difficultés supplémentaires apparaissent, non pas au niveau calculatoire puisque nous restons avec les quatre opérations arithmétiques pour la variance et la racine carrée pour l'écart-type, mais au niveau de la signification de l'interprétation en tant que mesure de dispersion. Par exemple, le fait que la variance s'exprime avec le carré des unités de la variable, pose une difficulté d'interprétation pour des étudiants.

Dans les traitements statistiques assistés par ordinateur, il convient de conserver une attitude de vigilance et un esprit critique à l'égard des résultats produits en se référant au cadre théorique  $\mathcal{CS}(I)$  de l'informatique. Le cas de ce qui est appelé l'Intelligence Artificielle est exemplaire à ce propos quand

on laisse à la machine le soin de prendre des décisions à la place de l'être humain en oubliant que les seuils qui déterminent les décisions sont définis par des êtres humains.

### **3.3.1 La production de représentations graphiques : entre infographie et figures géométriques**

Les traitements statistiques des données par des logiciels permettent de produire très facilement des représentations graphiques. Cette forme de représentation sémiotique a un sens particulier en mathématiques et en statistique qui renvoie au cadre théorique de référence qu'est la géométrie. Une des questions que nous nous posons est celle du statut des *graphiques* comme il est d'usage de désigner ces objets.

Ces *graphiques* peuvent avoir le statut de représentation graphique de la géométrie respectant des propriétés métriques, algébriques et topologiques pour traduire des informations de nature quantitative dans un registre graphique dans le sens développé par Duval (Duval, 1988, 1993, 1995). Sans aller plus loin dans le propos, nous rappelons qu'un registre de représentation doit permettre, selon Duval, la réalisation de trois activités cognitives fondamentales liées à la sémosis :

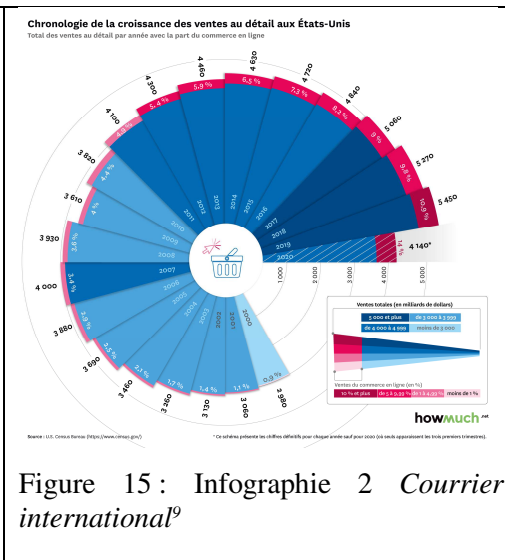
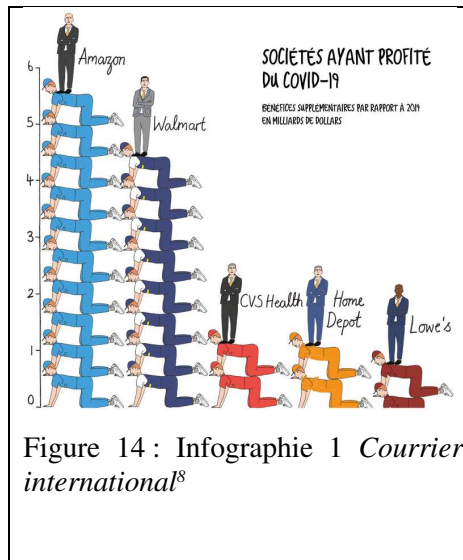
- La production d'une représentation identifiable comme une représentation dans un registre donné.
- Le traitement d'une représentation, c'est-à-dire sa transformation dans le registre même où elle a été formée.
- La conversion d'une représentation, c'est-à-dire la transformation de cette représentation en une représentation d'un autre registre en conservant la totalité ou une partie seulement du contenu de la représentation.

Dans le domaine de la statistique, nous rencontrons des représentations sémiotiques dans les registres graphiques, tabulaires, algébriques, numériques, discursifs. Elles interviennent dans les raisonnements statistiques, en particulier, soit en cours de traitements même, soit dans la formulation des résultats de ces traitements et analyses de données, ainsi que

dans la formulation des interprétations et conclusions et dans leur communication.

Ainsi nous pouvons observer que sont à l'œuvre des opérations de conversion de résultats de traitements exprimés dans le registre numérique et celui de la langue naturelle vers un registre graphique. Par ailleurs, cette production de graphiques a été grandement facilitée avec l'apparition des outils informatiques et des logiciels *grappeurs* qui sont maintenant intégrés aux logiciels de statistique ou aux *tableurs* comme le logiciel Excel.

Mais ces *graphiques* peuvent aussi avoir le statut de représentation infographique comme nous pouvons le voir quotidiennement dans les médias. Donnons ici deux exemples dans cette catégorie :



Revenons à l'usage des représentations graphiques dans le domaine de la statistique que peut en faire un étudiant en contexte universitaire dans le cadre de ses travaux d'étude et de recherche. Les trois représentations graphiques les plus élémentaires qui sont employées habituellement, sont les diagrammes en bâtons, les diagrammes circulaires et les histogrammes.

<sup>8</sup> <https://www.courrierinternational.com/grand-format/infographie-la-pandemie-profite-aux-grosses-entreprises>

<sup>9</sup> <https://www.courrierinternational.com/grand-format/consommation-vingt-ans-de-vente-au-detail-en-un-coup-doeil>

Si nous nous référons au logiciel Excel, nous disposons d'une boîte à outils donnant accès à un large éventail de type de graphiques :

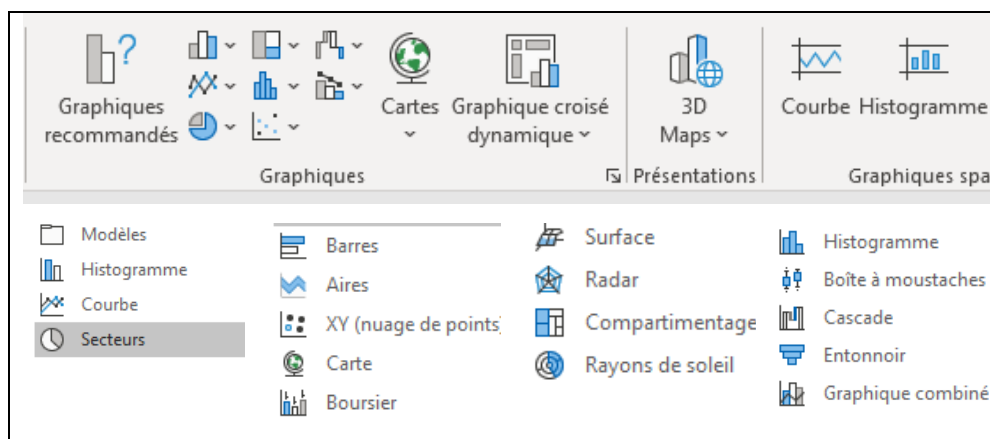


Figure 16 : Boîtes à outils pour construire des graphiques sous Excel

Prenons appui sur une enquête que nous avons récemment réalisée. Il s'agit d'une enquête par questionnaire en ligne Lime Survey réalisée en début mars 2021 auprès d'un groupe de 47 étudiants de master en science de l'éducation et de la formation dans le cadre d'un cours de statistique réalisé à distance organisé autour de la réalisation d'un dossier avec utilisation du logiciel de statistique SPAD 9.2 pour traiter les données. En ce qui concerne les occurrences de rencontrer des situations requérant des connaissances et des compétences en statistique.

Modalités	V03_1_Jamais	V03_2_Rarement	V03_3_Quelquefois	V03_4_Souvent	V03_5_Très souvent	Ensemble
Effectifs	1	11	13	5	1	31
Fréquence	3,2%	35,5%	41,9%	16,1%	3,2%	100,0%

Tableau 10 : Tableau de distribution des effectifs et des fréquences de la variable qualitative ordinaire V03 - Vous êtes-vous déjà trouvé(e) confronté(e) à des situations qui demandent des connaissances en statistique ?

Le résultat est alors fourni sous la forme de représentation sémiotique du registre Tableau que nous pouvons convertir dans le registre graphique à l'aide du logiciel Excel. C'est-à-dire dire la même chose autrement.

Choisissons les deux formes les plus usuelles que le diagramme en barres et le diagramme circulaire. Le premier traduisant la distribution des effectifs (information absolue) et le second, celle des fréquences (information

relative). Voilà ce que nous obtenons et qui pourrait être facilement construit à la main avec un papier-crayon.

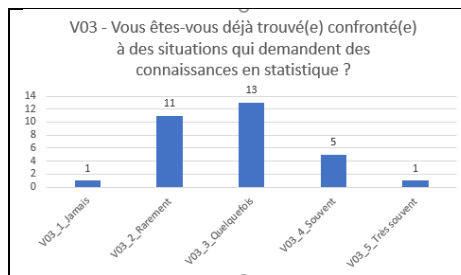


Figure 17 : diagramme en bâtons

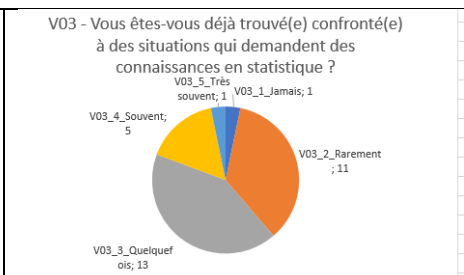


Figure 18 : diagramme circulaire

Ces représentations graphiques respectent les proportions entre les valeurs numériques mais ce n'est nullement le cas de choix des représentations influencées par l'esthétique infographique.

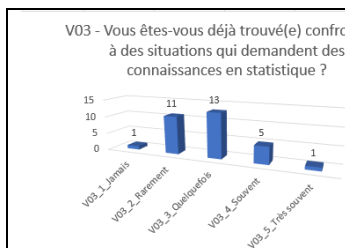


Figure 19 : diagramme en bâtons sous forme infographique

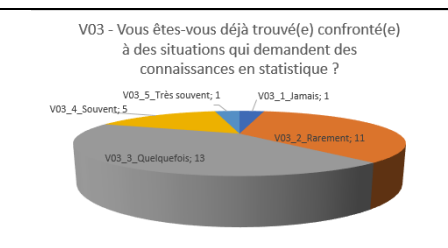


Figure 20 : diagramme circulaire sous forme infographique

Or dans le raisonnement statistique sous-jacent à la formulation des interprétations et des conclusions, c'est bien évidemment la préservation des propriétés lors de la conversion d'un registre vers l'autre qui est prioritaire. La dimension esthétique doit alors respecter ces contraintes de préservation. Une fois de plus se pose la question des connaissances et compétences minimales en mathématiques, statistique et informatique pour faire un choix judicieux et éclairé parmi l'ensemble des outils proposés par le logiciel Excel ou d'autres.

Abordons sans entrer dans les détails, la troisième représentation graphique importante qu'est l'histogramme qui concerne les variables quantitatives continues. Dans les usages courants des traitements statistiques, ces variables sont modélisées par des intervalles de l'ensemble des nombres réels. Nous renvoyons à un article que nous avons publié à la fin des années 90 (Régner, 1998a) sur cette représentation graphique qu'est l'histogramme. Il

conviendrait de poursuivre l'étude sur l'histogramme en tant qu'objet d'enseignement et d'apprentissage et d'outil graphique de représentation et de traitements de données. Dans le diagramme en bâtons, la hauteur des bâtons (segments) est proportionnelle à l'effectif ou la fréquence de la modalité ou de la valeur qu'elle représente. Dans l'histogramme, la fréquence ou l'effectif est représenté par une surface. Dans les usages habituels, les variables quantitatives continues sont modélisées par des intervalles numériques, par exemple, dans le cas de l'âge, on parle de tranches d'âge. Dans ce modèle et pour des raisons théoriques – répartition uniforme des valeurs sur l'intervalle – l'effectif ou la fréquence des valeurs observables sur un intervalle est représenté par la surface d'un rectangle dont la mesure de la base est proportionnelle à l'amplitude de l'intervalle et la mesure de la hauteur est proportionnelle à la densité d'effectif ou de fréquence et non pas à l'effectif ou à la fréquence comme nous le trouvons souvent écrit (Régner, 1998a).

Prenons l'exemple de l'âge des étudiants de l'échantillon auprès duquel nous avons conduit notre enquête. Cette valeur obtenue à partir de la date de naissance est calculée au jour au cours duquel l'étudiant a complété le questionnaire à l'aide des fonctions du tableur Excel. Nous avons ensuite procédé à une transformation sous la forme de 11 intervalles – tranches d'âge – dont l'amplitude est le 1/11 de l'amplitude totale qui va de 21,25 ans à 43,68 ans, c'est-à-dire environ 2 ans. L'âge est exprimé en unité année et 1/100 d'année, soit une précision d'environ 3 jours. Nous rappelons que le bilan de carrière professionnelle, la durée des états de service est exprimée en année, mois et jour !

Tranches d'âge (Intervalles)	Effectif	%	Densité
[21,25 ; 23,29[	10	32,26	0,158
[23,29 ; 25,32[	10	32,26	0,158
[25,32 ; 27,36[	4	12,90	0,063
[27,36 ; 29,4[	2	6,45	0,032
[29,4 ; 31,44[	2	6,45	0,032
[31,44 ; 33,48[	0	0,00	0,000
[33,48 ; 35,52[	1	3,23	0,016
[35,52 ; 37,56[	0	0,00	0,000
[37,56 ; 39,6[	0	0,00	0,000
[39,6 ; 41,64[	1	3,23	0,016
[41,64 ; 43,68[	1	3,23	0,016

Tableau 11 : Tableau de distribution des effectifs et des fréquences de la variable quantitative continue V02 – âge



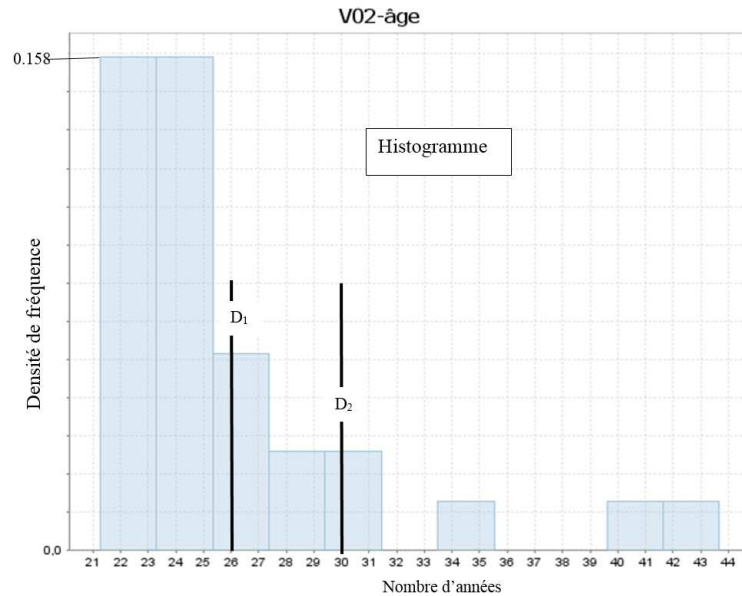


Figure 21 : représentation graphique de la distribution des fréquences d'une variable quantitative continue

Une telle représentation graphique rend possible le calcul de l'estimation de la fréquence des individus dans la tranche d'âge 26 ans à 30 ans, c'est-à-dire la fréquence des observations sur l'intervalle  $[26 ; 30[$  à partir de la mesure de l'aire de la portion géométrique de l'histogramme située entre les deux droites  $D_1$  et  $D_2$ .

Dans l'usage du tableur Excel, force est de constater que le nom *histogramme* ne désigne nullement un histogramme au sens que nous avons défini ci-dessus mais bien un diagramme en bâtons. Delà un utilisateur peu informé ou trop confiant en l'outil informatique peut recourir à cette procédure de construction et obtenir alors les graphiques suivants :

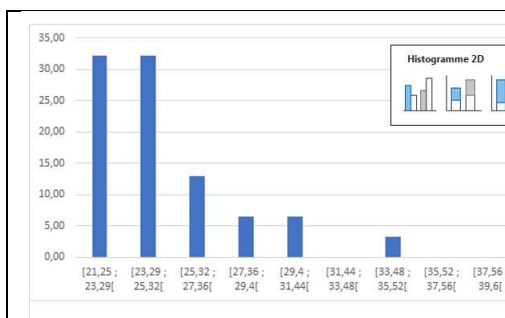


Figure 22 : diagramme en bâtons dénommé *histogramme* dans Excel

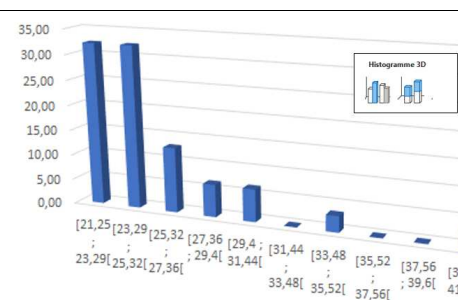


Figure 23 : diagramme en bâtons 3D dénommé *histogramme* dans Excel – représentation infographique

Nous sommes curieux de connaître la procédure géométrique et algébrique qui permettrait de résoudre le problème d'estimation de la fréquence sur l'intervalle  $[26 ; 30[$  à partir du graphique de la figure 22 et l'énigme s'accroît avec celui de la figure 23.

## 4 Conclusion

Répondant à l'invitation du comité de programme du colloque 11th International Meeting on Information, Knowledge and Action, en novembre 2019, nous avons proposé de porter un regard réflexif sur la thématique de la place et du rôle que peut jouer l'usage de logiciels de statistique ainsi que les interactions induites entre l'être humain utilisateur et la machine qu'est l'ordinateur ou la calculatrice électronique, dans les processus d'enseignement-apprentissage de la statistique en contexte universitaire. Nous nous sommes de nous-même mis devant un défi que nous avons tenté de relever. L'approche que nous avons adoptée n'est pas fondée sur une revue de la littérature ni un état actuel de l'art. Elle est fondée sur un retour réflexif sur nos pratiques réfléchies d'enseignant de statistique, sur nos propres travaux de recherche en didactique de la statistique et en statistique (dans le domaine de l'Analyse statistique implicative) en tant que chercheur et ceux conduits par des étudiants sous notre direction en master et doctorat en sciences de l'éducation et de la formation durant les 30 dernières années. Elle est aussi guidée par notre formation en mathématique et didactique des mathématiques.

La question majeure en filigrane organisatrice reste celle-ci : quelles sont les connaissances et compétences minimales en mathématiques au sein desquelles nous plaçons les théories des probabilités, en statistique et en informatique requises pour faciliter l'usage de méthodes statistiques et par suite, faciliter la compréhension et l'interprétation des résultats issus de traitements et d'analyses de données assistés par ordinateur dans le cadre théorique de la statistique appliquée à des objets et des faits d'un autre domaine de connaissance ou d'action ?

Nous ne parvenons pas à fournir une liste exhaustive mais au fil de nos propos, l'analyse des exemples simples que nous avons choisis met en évidence des connaissances et de compétences à l'œuvre.

Considérant, au cours de 50 dernières années, l'immense apport de l'informatique et des machines électroniques dans le développement même de la statistique en tant que discipline -objet mais aussi en tant que discipline-outil, discipline de service pour les travaux de recherche, par exemple, en sciences de l'éducation et de la formation, il convient de poursuivre la réflexion sur le thème de leur place et rôle dans le processus d'apprentissage de la statistique en contexte universitaire mais pas seulement. Nous avons évoqué le recours à la simulation qui tient une place et un rôle important comme a pu le présenter Oriol (2007) au sein d'un dispositif pédagogique et didactique.

En nous plaçant au moins implicitement dans une approche qui s'articule autour des trois dimensions cognition, affectivité et culture (Acioly-Régnier, 2010), il nous semble qu'un des objectifs de la formation en statistique est le développement de la posture [PSC5] dans (Régnier, 2005a) qui se manifeste « par une acceptation conditionnelle éclairée fondée sur l'idée que la statistique est une science sérieuse en laquelle le citoyen doit être formé avec un esprit critique. » C'est cette posture critique éclairée qu'il convient d'encourager chez les étudiants quand ils font usage de logiciels de statistique pour traiter leurs données construites. L'ordinateur est au service de l'être humain, et pas le contraire. D'une certaine façon, il s'agit de déjouer les effets de la pensée magique qui attribuerait la vérité à tout ce que produit un ordinateur : c'est vrai puisque c'est l'ordinateur qui le « dit » !. Pour l'heure, le fonctionnement de l'ordinateur est plutôt déterministe par l'application d'algorithmes qui sont les produits directs ou indirects de l'être humain. Et c'est bien le cas de ceux qu'utilisent les étudiants auxquels nous nous sommes intéressés.

L'attention que nous portons au cadre théorique des mathématiques  $\mathcal{C}\mathfrak{I}(\mathcal{M})$ , tient au fait de la prise en compte de la dimension affective dans le processus d'apprentissage de la statistique qui est sous-jacente à nos propos. Le rapport

des étudiants à la statistique est largement déterminé, dans la majorité des cas que nous avons étudiés par des enquêtes par questionnaire, par le rapport aux mathématiques (Régnier, 2005a)

En conclusion, il nous semble que l'efficacité des usages des logiciels de statistique dans le processus d'apprentissage de la statistique et donc dans la formation en statistique est d'autant plus forte que la distance critique est élevée dans une attitude intégrée à la posture [PSC5] que nous avons décrite plus haut.

## 5 Références

- Acioly-Régnier, N. M. (2007) Quand les facteurs culturels font obstacle à de la relation éducative. Approche dans la perspective culture et cognition A. Cognard (Dir) *Pour qu'éduquer ne soit pas un monologue*, 76420 Bihorel : Centon Editions, pp.85-108
- Acioly-Régnier, N.M (2008). Des instruments techniques aux instruments psychologiques : béquilles intellectuelles ou aides à la conceptualisation en mathématiques ? *Carrefours de l'éducation*. N° 26 – juillet-décembre 2008. p. 115-128 (halshs-00361438).
- Acioly-Régnier, N. M. (2010) *Culture et cognition : Domaine de recherche, Champ conceptuel, Cadre d'intelligibilité et Objet d'étude fournissant des instruments pour conduire des analyses conceptuelles et méthodologiques en psychologie et en sciences de l'éducation* Note de synthèse HDR – Université Lyon 2 <https://hal.archives-ouvertes.fr/tel-01982260>
- Acioly-Régnier, N. M., Régnier, J-C. (2007). Analyse cohésitive et interprétations des données dans le champ de l'éducation. *4e Rencontres sur l'Analyse Statistique Implicative*, Oct 2007, Castellon, Espagne. <https://halshs.archives-ouvertes.fr/halshs-00405180>
- Braga, E. M. (2009) *Enseignement-apprentissage de la statistique, TICE et environnement numérique de travail. Étude des effets de l'usage de supports didactiques numériques sur la conceptualisation en statistique*. Doctorat, Université de Lyon-Lyon 2 (Dir. : J.-C. Régnier). [http://theses.univ-lyon2.fr/documents/lyon2/2009/de\\_moura\\_braga\\_e/download](http://theses.univ-lyon2.fr/documents/lyon2/2009/de_moura_braga_e/download)
- Diaz, D. (2016) *Les facteurs influençant la réussite des activités collaboratives médiées par les TICE dans une situation de formation universitaire à la statistique*. Thèse de Doctorat, Université de Lyon-Lyon 2 (Dir. : J.-C. Régnier). [http://theses.univ-lyon2.fr/documents/lyon2/2016/diaz\\_d/download](http://theses.univ-lyon2.fr/documents/lyon2/2016/diaz_d/download)

[lyon2.fr/documents/lyon2/2016/diaz\\_d/pdfAmont/diaz\\_d\\_these\\_udl\\_t1.pdf](http://lyon2.fr/documents/lyon2/2016/diaz_d/pdfAmont/diaz_d_these_udl_t1.pdf)

- Duval R. (1988). Graphiques et équations. *Annales de Didactique et de Sciences cognitives*. ULP, IREM Strasbourg. 1, 235-253.
- Duval R. (1993). Registres de représentations sémiotiques et fonctionnement cognitif de la pensée. *Annales de Didactique et de Sciences Cognitives*, ULP, IREM Strasbourg. 5, 37-65.
- Duval R. (1995). *Sémiosis et pensée humaine. Registres sémiotiques et apprentissages intellectuels*. Berne, Peter Lang.
- Escofier, B., Pagès, J., (1990) *Analyses factorielles simples et multiples : objectifs, méthodes et interprétation*. Paris : Dunod
- Freire, P. (2006) *La Pédagogie de l'Autonomie. Savoirs nécessaires à la pratique éducative*. Ouvrage traduit de *Pedagogia da autonomia : saberes necessários à prática educativa*. São Paulo : Paz e Terra 16<sup>ème</sup> édition 2000 et commenté par Jean-Claude Régnier. Toulouse : Editions Eres, <https://hal.archives-ouvertes.fr/hal-02412418>
- Guiet, J.G. (1994) *La division : une longue souffrance*. Thèse de Doctorat, Université Paris 8 (Dir. : G. Vergnaud)
- Oriol, J.-C. (2002) Réaliser une enquête par questionnaires : un outil didactique pour la statistique inférentielle à l'Université. In *3ème Rencontre Europe-Amérique-Latine sur la formation technologique et professionnelle*, La Havane, Cuba.
- Oriol, J.-C. (2007) *Formation à la statistique par la pratique d'enquêtes par questionnaires et la simulation : étude didactique d'une expérience d'enseignement dans un département d'IUT*, Thèse de Doctorat, Université de Lyon-Lyon 2 (Dir. : J.-C. Régnier). <https://tel.archives-ouvertes.fr/tel-00191166/document>
- Oriol, J.-C., Régnier, J.-C. (2003a) Fonctionnement didactique de la simulation en statistique. Exemple de l'enseignement du concept d'intervalle de confiance, *Actes des 35<sup>ème</sup> Journées de Statistique de la SFdS*, 743-750.
- Oriol, J.-C., Régnier, J.-C. (2003b) Fonctionnement didactique de la simulation en statistique dans l'enseignement du concept de corrélation, *Espace Mathématique Francophone 2003*, Tozeur, Tunisie.
- Oriol, J.-C., Régnier, J.-C. (2004) Loi de Benford et Pratique d'enquêtes par questionnaires : un exemple de situations didactiques en statistique, *Actes des 35<sup>ème</sup> Journées de Statistique de la SFdS*.
- Régnier, J.-C. (1981). Et les machines à calculer!!!. *Animation & Education* (ISSN 0395-0840), 1981, pp.17. <https://halshs.archives-ouvertes.fr/halshs-00406933>

- Régnier, J-C. (1988). Étude didactique d'une méthode d'apprentissage fondé sur le tâtonnement expérimental de l'apprenant. *Annales de Didactiques et de Sciences Cognitives*, 1988, 1, pp.255-279. <https://halshs.archives-ouvertes.fr/halshs-00363454>
- Régnier, J-C. (1994). Tâtonnement expérimental et Apprentissage en mathématiques. P. Clanché, E. Debarbieux. *La pédagogie Freinet, mises à jour et perspectives*, Presses Universitaires de Bordeaux, pp.135-153. <https://halshs.archives-ouvertes.fr/halshs-00506711>
- Régnier, J-C. (1998). Danger ! Approximations.... *Enseigner la Statistique du CM à la Seconde Pourquoi? Comment ?*, IREM de Lyon Université Lyon1, pp.99-105. <https://halshs.archives-ouvertes.fr/halshs-00405999>
- Régnier, J-C. (1998a). Histogramme. *Enseigner la Statistique du CM à la Seconde Pourquoi? Comment ?*, Irem de Lyon Université Lyon1, pp.21-42. <https://halshs.archives-ouvertes.fr/halshs-00405993>
- Régnier, J-C. (2002) A propos de la formation en statistique. Approches praxéologiques et épistémologiques de questions du champ de la didactique de la statistique. *Revue Questions éducatives. L'école et ses marges*. CRE Université de Saint-Étienne (France), n°22-23 (pp. 157-201) <https://hal.inria.fr/halshs-00363427>
- Régnier, J-C. (2003) Statistical Education and E-learning. *Statistics Education: Statistics and Internet*, Aug 2003, Berlin, Germany. pp.1-10. <https://halshs.archives-ouvertes.fr/halshs-00405176>
- Régnier, J-C. (2005a) Formation de l'esprit statistique et raisonnement statistique. Que peut-on attendre de la didactique de la statistique ? In C. Castela & C. Houdement (Dir.) *Actes du séminaire national de didactique des mathématiques*. Paris : ARDM-Irem de Paris VII - p. 13-37 <https://halshs.archives-ouvertes.fr/halshs-00391741>
- Régnier, J-C. (2005b) Étude des difficultés d'apprentissage de la statistique dans le cadre d'un enseignement à distance. *MESURER*, Paris : L'Harmattan, pp.15-47 <https://halshs.archives-ouvertes.fr/halshs-00361957>
- Régnier, J-C. (2006a). Prolongement réflexif sur la pédagogie de l'autonomie de Paulo Freire. : Traduire un discours pédagogique sans trahir une pensée. In Paulo Freire (2006) *La Pédagogie de l'Autonomie*, Toulouse : Éditions ERES, pp.155-184. <https://halshs.archives-ouvertes.fr/halshs-00362042>
- Régnier, J-C. (2012) Enseignement et apprentissage de la statistique : entre un art pédagogique et une didactique scientifique. *Statistique et Enseignement*, Société Française de Statistique, 2012, 3 (1), pp.19-36. <https://hal.inria.fr/hal-00716002>

- Régnier, J-C. (2014a) Mathématiques : réflexions historiques, épistémologiques, pédagogiques et didactiques pour une contribution à la refondation de l'enseignement scientifique. Martine Boudet, Florence Saint-Luc. *Le système éducatif à l'heure de la société de la connaissance*, Presses Universitaires du Mirail, pp.321-355. Questions d'éducation, 978-2-8107-0324-1. <https://hal.inria.fr/hal-01094731>
- Régnier, J-C. (2014b) Instrumentalisation technocratique des statistiques et alternatives citoyennes. Martine Boudet, Florence Saint-Luc. *Le système éducatif à l'heure de la société de la connaissance*, Presses Universitaires du Mirail, pp.83-106, Questions d'éducation. <https://hal.inria.fr/hal-01094725>
- Régnier, J-C., Braga, E. M. (2008) Instrumentos estatísticos para uma leitura do mundo: Formação do espírito estatístico e cidadania. *Conjectura - Filosofia e Educação (ISSN 0103-1457)*, 2008, 13 (2), pp.9-41. <https://halshs.archives-ouvertes.fr/halshs-00361453>
- Régnier, J-C., Denis, A. (1990) Au-delà des vides d'un questionnaire à trous. *Revue du Centre de Recherche en Éducation*, 1990, (2), pp.1-79. <https://halshs.archives-ouvertes.fr/halshs-00363444>
- Schliemann, A. D., Acioly, N. M. (1989) Mathematical Knowledge Developed At Work: The Contribution of Practice versus The Contribution Of Schooling. *Cognition and Instruction*, New Jersey, v. 6, n.3, p. 185-221, 1989. Lawrence Erlbaum Associates Publishers. ISSN : 0737-0008. (p.185-221).
- Vergnaud, G. (2001) Forme opératoire et forme prédicative de la connaissance. Jean Portugais (Ed) Actes du Colloque GDM-2001. *La notion de compétence en enseignement des mathématiques, analyse didactique des effets de son introduction sur les pratiques et sur la formation*. Québec Montréal [http://recherches.philippeclauzard.com/vergnaud\\_forme\\_connaissance.pdf](http://recherches.philippeclauzard.com/vergnaud_forme_connaissance.pdf)