

## QUELQUES RÉFLEXIONS SUR L'ENSEIGNEMENT DES STATISTIQUES DESCRIPTIVES DANS NOS CLASSES

Avant 1990, on ne faisait des statistiques que dans quelques classes de lycée.

Puis, à partir de 1990 les statistiques sont enseignées au collège et en seconde.

Depuis l'an 2000, en Seconde, c'est la grande "SMS" (Sensibilisation à la Modélisation par la Simulation).

Aussi, à l'heure où les statistiques ont une place considérable dans les programmes puisque nous nous sommes décidés (et que nous y prenons même quelques plaisirs !) à les enseigner, et où la fluctuation d'échantillonnage n'a plus aucun secret pour personne (!?), le moment est venu de nous ré-impregnier des statistiques "élémentaires" dites descriptives, de celles qui pourront peut-être (sûrement) servir à nos élèves dans leur vie de citoyen, et de remettre en cause (un peu !) nos pratiques d'enseignant.

Actuellement, de la 3<sup>ème</sup> à la terminale Bac Pro, les statistiques descriptives (séries à une variable) sont au programme, sous cette forme :

	Contenus	Compétences attendues
<b>3<sup>ème</sup> T</b>	<p>Moyenne, médiane d'une série statistique.</p> <p>Les travaux permettront de faire la synthèse des activités antérieures.</p>	<p><b>Savoir lire et exploiter</b> des données statistiques mises sous la forme de tableaux ou de diagrammes.</p> <p>Savoir calculer une moyenne.</p> <p>A partir de données statistiques, calculer les effectifs ou les fréquences, les présenter dans des tableaux et tracer les diagrammes correspondants.</p>
<b>BEPA</b>	<ul style="list-style-type: none"> <li>- variables quantitatives discrètes, variables quantitatives continues.</li> <li>- variables qualitatives.</li> <li>- collecte et organisation des données.</li> <li>- effectifs et effectifs cumulés.</li> <li>- représentations graphiques.</li> <li>- paramètres de position : mode, classe modale, moyenne et médiane.</li> <li>- paramètres de dispersion : variance et écart-type.</li> </ul>	<p>Lire et <b>interpréter</b> des tableaux statistiques.</p> <p>Représenter graphiquement des données statistiques.</p> <p>Déterminer graphiquement la médiane d'une série statistique à une variable continue.</p> <p>Calculer la moyenne et l'écart-type d'une série statistique.</p>
<b>2<sup>nde</sup></b>	<p>Résumé numérique par une ou plusieurs mesures de tendance centrale (moyenne, médiane, classe modale, moyenne élaguée) et une mesure de dispersion (on se restreindra en classe de 2<sup>nde</sup> à l'étendue)...</p>	<p>Utiliser les propriétés de linéarité de la moyenne d'une série statistique. Calculer la moyenne d'une série à partir des moyennes de sous-groupes.</p> <p>Calcul de la moyenne à partir de la distribution des fréquences.</p>

<b>Bac Pro</b>	Généralités sur les séries statistiques. Paramètres de tendance centrale : mode, médiane et moyenne. Paramètres de dispersion : étendue, variance et écart-type.	Maîtriser les différents types de variables statistiques. Savoir visualiser les données par des représentations graphiques. Calculer et <b>interpréter</b> les différents paramètres statistiques.
<b>1<sup>ère</sup> STAE</b>	- fréquence et fréquence cumulée. - mesure de tendance centrale : mode, moyenne, médiane. - mesure de dispersion : étendue et écart-type. - exemples de calculs de quartiles, quintiles, déciles et intervalle interquartile.	Mobiliser les acquis des classes antérieures pour étudier et <b>interpréter</b> des cas concrets. Utiliser la courbe des fréquences cumulées croissantes. <b>Faire preuve d'esprit critique</b> pour les méthodes et les interprétations. Lire un diagramme en boîte issu d'une situation concrète en relation avec les disciplines techniques.

... et dans beaucoup de manuels, les exercices ressemblent à ça !

Deux exercices de statistiques donnés lors du **Brevet 2002**

#### A Grenoble

Une usine teste des ampoules électriques, sur un échantillon, en étudiant leur durée de vie en heures. Voici les résultats :

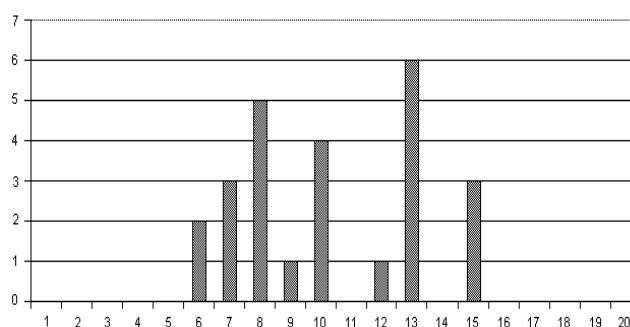
<b>d : durée de vie (en h)</b>	<b><math>1\ 000 \leq d &lt; 1\ 200</math></b>	<b><math>1\ 200 \leq d &lt; 1\ 400</math></b>	<b><math>1\ 400 \leq d &lt; 1\ 600</math></b>	<b><math>1\ 600 \leq d &lt; 1\ 800</math></b>	<b><math>1\ 800 \leq d &lt; 2\ 000</math></b>
<b>Nombre d'ampoules</b>	550	1 460	1 920	1 640	430

- 1) Quel est le pourcentage d'ampoules qui ont une durée de vie de moins de 1 400 h ?
- 2) Calculer la durée de vie moyenne d'une ampoule.

#### A Besançon, Dijon, Lyon, Nancy-Metz, Reims, Strasbourg

Voici le diagramme représentant la répartition des notes obtenues par les élèves d'une classe de troisième lors d'un contrôle de français : les notes sur 20 sont reportées en abscisses, le nombre d'élèves est reporté en ordonnées.

- 1) Quel est l'effectif de cette classe de 3<sup>ème</sup> ?
- 2) Calculer la moyenne des notes obtenues en donnant le résultat sous sa forme décimale exacte.



#### Dans le Déclik Seconde page 190

Dans une usine de conserves, on remplit les boîtes à l'aide d'un système automatisé. Pour en vérifier le réglage, on prélève au hasard 80 boîtes et on pèse leur contenu (masse nette). Le tableau ci-dessous indique la répartition, en classes d'amplitude 10 g.

Masse (en g)	[480 ; 490[	[490 ; 500[	[500 ; 510[	[510 ; 520[	[520 ; 530[	[530 ; 540[	[540 ; 550[
Nombre de boîtes	4	8	45	18	3	1	1

- 1) Déterminer la classe modale et la classe médiane. Calculer la masse médiane. Interprétez ce résultat par une phrase.
- 2) Quel est le pourcentage de boîtes dont la masse nette est supérieure ou égale à 500 grammes ?

Pour notre part, il nous paraît essentiel de favoriser les calculs sur les données brutes et d'éviter tant que faire se peut, les regroupements par classes (choisies arbitrairement le plus souvent), qui sont un résumé de la série, au même titre que la moyenne.

Calculer des paramètres statistiques à partir de ces regroupements ne donnera que des valeurs approchées (parfois même très éloignées). Inutile de faire un dessin, c'est tellement vrai quand on y pense... Au passage, relisez le programme de la classe de quatrième.

D'autre part, comme le recommande le programme de 3<sup>ème</sup>, "Il s'agit (...) d'habituer les élèves à avoir une attitude de lecteurs responsables face aux informations de nature statistique". Et finalement, dans les statistiques que l'on rencontre couramment, nous n'avons pas (ou très peu) de calculs à faire (tout au plus une moyenne, ou une médiane à trouver). Notre travail est surtout d'interpréter des "chiffres", des lettres ou des graphiques. Et c'est ce que nous allons voir !

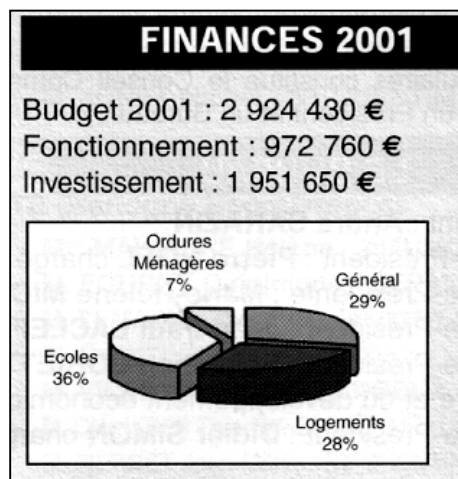
Nous vous proposons deux types d'activités, réalisables avec des élèves de la 3<sup>ème</sup> à la Terminale Bac Pro, sur l'interprétation de statistiques courantes (seuls les exercices où interviennent les déciles, pourraient poser quelques difficultés avant la 1<sup>ère</sup> STAE). La première partie concerne l'interprétation de graphiques et la seconde l'interprétation de données résumées par leurs paramètres statistiques.

## 1<sup>ère</sup> Partie : trois exemples de graphiques

Nous avons relevé dans la presse française, plusieurs graphiques statistiques et nous avons sélectionné pour vous, ceux qui nous ont le plus interpellés. Soit pour notre grande incapacité à les interpréter, soit pour les nombreux renseignements qu'ils contiennent.

Pour chacun des quatre graphiques étudiés, nous proposons une série de questions, nos réponses et nos commentaires.

A : un diagramme circulaire (extrait du Guide de la communauté de commune du porcien 2002)  
Aucun commentaire n'accompagnait ce graphique.



### Questions :

- 1) A partir de quelles données calcule-t-on les pourcentages indiqués sur le graphique ?
- 2) Quel volume semble représenter le pourcentage le plus important ?
- 3) Les volumes semblent-ils être proportionnels aux pourcentages ?

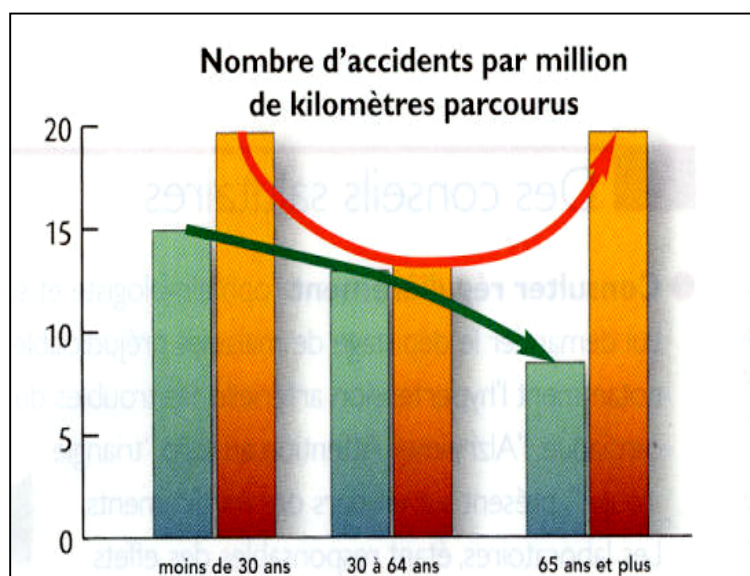
## Réponses :

- 1) On ne sait pas.
- 2) C'est vrai que l'on est un peu de mauvaise foi et que notre vue baisse, mais sans hésiter, nous proposons le quartier le plus foncé.
- 3) La réponse 2) répond à la question 3). La réponse est non.

**Commentaires :** Ce graphique représente-t-il la répartition du budget 2001 ? Celle du fonctionnement ? Qu'entend-on par "général" ? C'est un peu vague ! Non ?

Ce graphique ne dit rien et pourtant, nous avons eu, au premier coup d'œil, l'impression qu'il était simple et clair. Encore une fois, les apparences sont trompeuses... Et tout graphique doit être étudié, épluché avec attention.

B : un graphique (extrait de la revue MAIF infos septembre 2001)



En vert (rectangle de gauche), le nombre de kilomètres parcourus. En orangé (rectangle de droite), le nombre d'accidents.

Bien que roulant moins, les personnes âgées causent autant d'accidents que les jeunes conducteurs (source : SOFRES 1990)

## Questions :

- 1) A quoi correspond la graduation de l'axe des ordonnées ? Aux millions de kilomètres qui sont parcourus ? Au nombre d'accidents (par million de kilomètres parcourus) ?
- 2) Le nombre d'accidents par million de kilomètres parcourus est sensiblement le même chez les moins de 30 ans que chez les plus de 65 ans. Que peut-on dire quant au nombre d'accidents par tranche d'âge ?
- 3) Quel message veulent faire passer les deux flèches ?

## Réponses :

- 1) Nous, on ne sait pas. Peut-être aux deux, peut-être pas !
- 2) On note  $N$  le nombre d'accidents et  $D$  le nombre de millions de kilomètres parcourus. Sont représentés alors,  $D$  et  $\frac{N}{D}$ .
  - Pour les moins de 30 ans, le nombre d'accidents est donc de  $20 \times 15 = 300$ .
  - Tandis que pour les plus de 65 ans, il est de  $20 \times 8 = 160$ .
  - Et là, bien que ces calculs semblent corrects, beaucoup de questions nous assaillent encore : 300 accidents par an ? C'est tout ! etc.

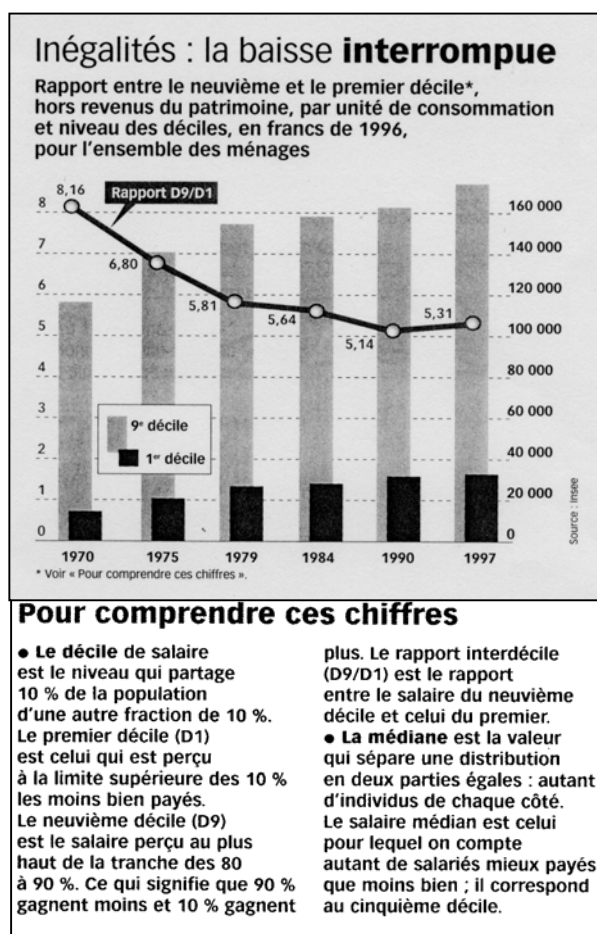
3) Qu'après 65 ans on roule moins (à moins que ce ne soit le nombre de conducteurs qui diminue) ? Que finalement, on est vraiment sage qu'entre 30 et 64 ans ?

Ce doit être ça le message ! Non seulement, le graphique nous laisse perplexe, mais voilà qu'il nous pose des questions métaphysiques...

**Commentaires :** Sans doute avez-vous d'autres interprétations.

Pour notre part, nous n'arrivons pas à nous mettre d'accord sur la finalité de ce graphique. Cela montre à quel point, il faut être vigilant sur la conception d'un graphique et commencer par se poser ces quelques questions qui pourraient être utiles à certains de nos élèves qui usent et abusent de représentations graphiques souvent insensées (au sens propre). Quel message faire passer ? Quel titre donner ? Quel type de représentation choisir ? Finalement, un graphique est-il indispensable ?

C : un diagramme en barres (extrait de Alternatives économiques - Hors-série n°54 - 4<sup>ème</sup> Trimestre 2002)



**Questions :**

- 1) Que veut dire décile ?
- 2) Que veut dire unité de consommation ?
- 3) Comment interprétez-vous le nombre 120 000 donné sur l'échelle de droite du document ?
- 4) En 1990, on peut lire que le premier décile est de 30 000. Interpréter ce nombre.
- 5) Sur ce graphique, on indique un rapport de 5,31. Comment obtient ce résultat ?
- 6) Pour l'année 1990, peut-on lire sur un tel graphique, la valeur de la médiane ? De la moyenne ?
- 7) Comment interprétez-vous le titre donné à ce document ?

**Réponses :**

- 1) La définition est donné dans le commentaire.
- 2) Dans le graphique de la page suivante cette définition est aussi donnée. L'INSEE divise les revenus en parts appelées "unités de consommation". Le premier adulte vaut une part pleine, le second 0,5 et les enfants de moins de 14 ans valent 0,3. Exemple si le montant des revenus annuels d'un ménage composé de deux adultes et d'un jeune enfant est de 17 000 €, le montant annuel des revenus par unité de consommation est de 9 444 €
- 3) 120 000 est le montant annuel des revenus par unité de consommation exprimé en francs.

- 4) En 1990, 10 % des ménages avaient un revenu annuel par unité de consommation inférieur à 30 000 F.
- 5) 5,31 s'obtient en divisant le 9<sup>ème</sup> décile qui est de l'ordre de 170 000 par le premier décile de l'ordre de 30 000.
- 6) Il n'est pas possible à partir du 1<sup>er</sup> décile et du 9<sup>ème</sup> décile de connaître la médiane et la moyenne.
- 7) Dans le titre il n'est pas question des écarts entre le 9<sup>ème</sup> décile et le premier. On s'aperçoit que cet écart est pratiquement toujours en augmentation. La baisse indiquée dans le titre se rapporte aux rapports entre le 9<sup>ème</sup> et le premier décile qui remonte à partir de l'année 1990 pour passer de 5,14 à 5,31.

### Commentaires :

Il s'agit ici d'essayer de montrer toute la richesse des questions que l'on peut poser à partir de certains graphiques.

## 2<sup>ème</sup> Partie : des "chiffres" à interpréter

Dans la presse ou via Internet, nous avons recueilli des "chiffres" à l'état brut, qui pourraient être la source d'exercices d'interprétation des paramètres statistiques.

A : les revenus des français (Alternatives économiques - Hors-série n°54 - 4<sup>ème</sup> Trimestre 2001)

### Revenus des français en 2001

#### Questions :

Revenu le plus élevé des	Ménages		Unités de consommation	
	En francs 2001	En euros 2001	En francs 2001	En euros 2001
10 % les plus bas	4 007	611	2 859	436
20 %	6 093	929	4 237	646
30 %	7 776	1 184	5 346	815
40 %	9 499	1 448	6 370	971
50 %	11 380	1 735	7 393	1 127
60 %	13 525	2 062	8 547	1 303
70 %	16 123	2 458	10 325	1 574
80 %	19 652	2 996	11 820	1 802
90 %	25 903	3 949	15 185	2 315
95 %	32 530	4 959	19 174	2 923

Source : Insee, calculs Alternatives économiques

1 735 euros (11 380 francs) par mois, voilà ce que gagne le ménage français médian, celui qui se trouve au milieu de la distribution des revenus. Au bas de l'échelle, un tiers des ménages dispose de moins de 1 200 euros (7 871 francs). Souvent des personnes seules (des jeunes qui commencent leur vie active ou des veufs à la retraite) ou des familles dont au moins l'un des conjoints est au chômage ou ne travaille pas. Pour tenir compte de la composition des ménages, l'Insee divise leurs revenus en parts appelées « unités de consommation ». Le premier adulte vaut une pleine part, le second 0,5 et les enfants de moins de 14 ans valent 0,3. Selon ce système, l'adulte situé au niveau médian touche 1 130 euros (7 412 francs). Cette méthode revient à considérer qu'un enfant réduit le niveau de vie de façon strictement proportionnelle au revenu, quel que soit celui-ci, ce qui est faux.

- 1) Quel est le salaire médian ? Comment peut-on interpréter ce nombre ?
- 2) Quelle interprétation peut-on donner au chiffre "4 959" du tableau ?
- 3) Le texte dit que "un tiers des ménages dispose de moins de 1 200 €". Cette information paraît-elle cohérente ?
- 4) Que penser de l'affirmation : "le troisième quartile est de 15 784 francs" ? (On n'arrive pas à se faire à l'euro, à notre âge, c'est trop tard !).
- 5) Justifiez que "le revenu moyen est supérieur au salaire médian d'au moins 300 francs

(soit environ 45 euros. Là, ça va !)".

- 6) Et pour le plaisir... Quel est le deuxième quintile de cette série ?
- 7) Comment pourrait-on représenter cette série ? Quel message passera ?

Revenus fiscaux en 1998 (chiffres de l'INSEE de Bourgogne 2002)

	<b>Bourgogne</b>	France entière
<b><i>Foyers fiscaux imposés</i></b>		
Nombre de foyers fiscaux imposés	<b>454 195</b>	16 758 555
Pourcentage de foyers fiscaux imposés	<b>52,3</b>	52,7
Revenu moyen net annuel imposable (en euros)	<b>20 088</b>	21 844
Montant de l'impôt net moyen (en euros)	<b>2 235</b>	2 699
<b><i>Foyers fiscaux non imposés</i></b>		
Nombre de foyers fiscaux non imposés	<b>414 666</b>	15 043 655
Pourcentage de foyers fiscaux non imposés	<b>47,7</b>	47,3
Revenu moyen net annuel imposable (en euros)	<b>5 978</b>	5 789

Source : Direction Générale des impôts

### Questions :

- 1) Quel est le montant de l'impôt sur le revenu perçu par l'État en Bourgogne en 1998 ?
- 2) Quel est le revenu moyen net annuel imposable en Bourgogne ?
- 3) On classe les foyers fiscaux par ordre croissant du montant de l'impôt net annuel moyen. Le foyer fiscal médian est-il imposé ?
- 4) Dans une autre région, le montant de l'impôt net médian est nul. Comment peut-on interpréter cela ?

### B : *pluviométrie*

Voici des résumés des résultats de deux stations météorologiques de lycée agricole, l'un situé dans la région Poitou-Charentes et l'autre en Bretagne.

Le caractère étudié est la hauteur de pluie (en mm) tombée chaque jour. L'étude porte sur deux années consécutives. Le logiciel pour dépouiller les données, fournit les résultats suivants :

Lycée 1		Lycée 2	
<b>Tendances centrales</b>		<b>Tendances centrales</b>	
Moyenne	2,24	Moyenne	3,34
Médiane	0,00	Médiane	0,40
1er quartile	0,00	1er quartile	0,00
3ème quartile	2,20	3ème quartile	3,80
1er décile	0,00	1er décile	0,00
9ème décile	6,80	9ème décile	10,60
1er centile	0,00	1er centile	0,00
99ème centile	22,00	99ème centile	25,00
2.5ème centile	0,00	2.5ème centile	0,00
97.5ème centile	16,80	97.5ème centile	22,20
5ème centile	0,00	5ème centile	0,00
95ème centile	11,80	95ème centile	17,40
<b>Dispersion</b>		<b>Dispersion</b>	
Minimum	0,00	Minimum	0,00
Maximum	37,20	Maximum	43,40
Amplitude	37,20	Amplitude	43,40
IQR	2,20	IQR	3,80
Écart type	4,75	Écart type	6,05

**Questions :**

- 1) A quoi dans cette série peut servir la moyenne ?
- 2) Comment peut-on interpréter la différence sur les écart-types ?
- 3) L'intervalle interquartile IQR est de 3,80 dans la seconde série, comment interpréter ce nombre ?
- 4) Le minimum et le maximum sont-ils des événements rares ?
- 5) Justifier que dans cette série le couple médiane/intervalle interquartile est plus pertinent que le couple moyenne/écart-type.

**Réponses :**

- 1) Uniquement à calculer la hauteur d'eau annuelle.
- 2) On ne sait pas dans ce type de série.
- 3) On peut dire que 3 jours sur 4 il tombe moins de 3,8 mm d'eau par jour.
- 4) Plus d'un jour sur deux il ne pleut pas au lycée 1 et on peut même dire qu'il fait bon vivre. Le minimum est donc souvent obtenu, alors que le maximum est très peu souvent atteint, le centile 99 le confirme avec une valeur de 25.
- 5) On peut se demander si la moyenne et l'écart type sont plus intéressants pour résumer ces deux séries que la médiane et l'intervalle interquartile noté IQR Ici la moyenne ne dit pas grand chose si ce n'est que par an, il tombe beaucoup plus d'eau dans le second lycée. Mais ici la moyenne n'est pas d'une autre utilité et le calcul qui a été fait pour l'obtenir semble bien inutile puisque l'on revient au nombre qui nous intéressait, à savoir la hauteur d'eau annuelle. En fait dans ce type de phénomène il est plus important de savoir comment tombe cette hauteur d'eau annuelle. La médiane, ainsi que d'autres paramètres tels que quartile ou quintile sont plus pertinents.