Université de Provence

Département des sciences de l'éducation

Année 2007-2008 – Master 1

UE SCER22 :

« Didactique des disciplines scientifiques »

Yves Chevallard
y.chevallard@aix-mrs.iufm.fr

Trois leçons sur la didactique des PER

➔ Leçon 2

Sommaire. – 0. Gestes d'étude et de recherche / 1. Dialectique de l'excription et de l'inscription : « Comment analyser des données ? » / 2. Comment « observer » ?

0. Gestes d'étude et de recherche

a) Rappelons la question Q_0 considérée dans la leçon 1, en la reformulant adéquatement.

> Q_0. Que sont les praxéologies dont les auteurs cités dans la leçon 1 à propos de pourcentages et autres proportions se prévalent dans leurs observations critiques à l'encontre des écrits ou propos dont ils s'y font l'écho ?

➔ Pour répondre à cette question, nous avons envisagé d'étudier une autre question, Q_0' qu'on peut dire « engendrée par Q_0 » ***dans le PER*** amorcé (et non pas dans l'abstrait) :

> Q_0'. Quelles praxéologies trouve-t-on « dans la culture » autour des types de tâches T_i ($1 \leq i \leq 8$) ci-après ?
>
> T_1. Déterminer la valeur g_1 d'une grandeur G qui, partant de la valeur g_0, subit une augmentation de r %.
>
> T_2. Déterminer la valeur g_1 d'une grandeur G qui, partant de la valeur g_0, subit une diminution de r %.
>
> T_3. Déterminer le pourcentage d'augmentation r % d'une grandeur G qui, partant de la valeur g_0, prend la valeur $g_1 > g_0$.
>
> T_4. Déterminer le pourcentage de diminution r % d'une grandeur G qui, partant de la valeur g_0, prend la valeur $g_1 < g_0$.
>
> T_5. Déterminer le pourcentage d'augmentation r % d'une grandeur G qui, partant de la valeur g_0, prend la valeur $g_1 = kg_0$, où k est un nombre entier > 1.
>
> T_6. Déterminer le pourcentage de diminution r % d'une grandeur G qui, partant de la valeur g_0, prend la valeur $g_1 = g_0/k$, où k est un nombre entier > 1.
>
> T_7. Déterminer le nombre entier k tel que $g_1 = kg_0$ lorsqu'une grandeur G, partant de la valeur g_0, prend la valeur g_1 après une augmentation de 100 r %.

T_8. Exprimer [en pourcentage] la proportion, dans un ensemble donné U, des éléments appartenant à une partie V de U.

➔ On semble là « contourner » la question Q_0 : on ne cherche pas à connaître ce que ***sont*** exactement les praxéologies mises en œuvre par les intervenants du site Pénombre, mais ce qu'elles ***pourraient*** être. En cherchant ainsi les praxéologies « portées » par la culture ambiante, on élargit la question pour la rendre moins ***intraitable***.

➔ Pour étudier Q_0' et donc pour amorcer concrètement le fonctionnement de $S(X ; y ; Q_0')$, y a invité X à examiner deux documents, $O^{\%}_{\mathrm{Wf}}$ et $O^{\%}_{\mathrm{We}}$, disponibles sur Internet, et cela afin d'apporter réponse à la question suivante :

Q_0''. Quelles praxéologies trouve-t-on dans les œuvres $O^{\%}_{\mathrm{Wf}}$ et $O^{\%}_{\mathrm{We}}$ autour des types de tâches T_i ($1 \leq i \leq 8$) ?

Une réponse $R_0''^{\heartsuit}$ à Q_0'' doit être construite en tenant compte d'une ***contrainte*** essentielle, déterminée par la perspective dans laquelle on s'est posé la question Q_0'' : répondre à la question Q_0' et, au-delà, à la question Q_0 elle-même.

b) Avant d'aller plus loin dans l'étude de Q_0'', on s'arrête un instant sur cinq « gestes » de base, présents en principe en tout parcours d'étude et de recherche relatif à une question Q. Un premier geste consiste à ***observer*** les réponses $R^{\Diamond}$ à Q déposées dans des ***institutions*** ou portées par des ***personnes***. « Observer » doit être pris en un sens englobant : il désigne le ***repérage***, l'***identification*** et la ***collecte*** d'éléments d'information sur les réponses $R^{\Diamond}$ « observables » qui nous intéressent. Ainsi y a-t-il « observé » les deux documents $O^{\%}_{\mathrm{Wf}}$ et $O^{\%}_{\mathrm{We}}$ et a-t-il ensuite communiqué à X cette double observation (par le biais des annexes à la leçon 1).

➔ Ces observations peuvent être multipliées. Si on recherche des documents appropriés sur le Web, on pourra rencontrer le document dont on a reproduit ci-après le tout début (http://mathscyr.free.fr/formulaire/PourcentagesFORMULAIRE.pdf).

Cours et exercices de mathématiques M.CUAZ

POURCENTAGES - FORMULAIRE

1) Pourcentage d'une quantité par rapport à une autre

Pour prendre les t % d'une quantité a, on effectue le calcul $\frac{t}{100} \times a$

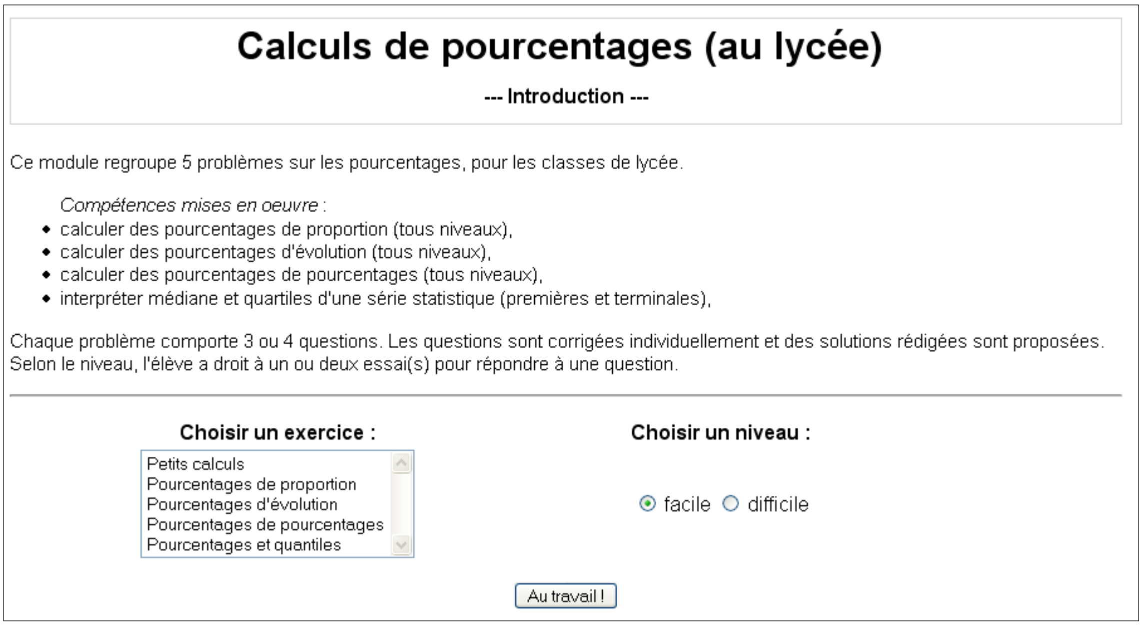
Pour trouver le pourcentage que a représente par rapport à b, on effectue la division $\frac{a}{b}$ et on écrit le résultat sous la forme d'un pourcentage.

2) Augmentation ou diminution d'un pourcentage

x représente une valeur "initiale" (par exemple un prix, une population, une production, etc…)
y représente une valeur "finale".

"y est égal à x augmenté de t%" équivaut à $y = x\left(1+\frac{t}{100}\right)$.
"y est égal à x diminué de t%" équivaut à $y = x\left(1-\frac{t}{100}\right)$.

On trouvera de même le document reproduit ci-après (http://wims.auto.u-psud.fr/wims/wims.cgi?module=H5%2Falgebra%2FpourcentageLycee.fr).


Calculs de pourcentages (au lycée)

--- Introduction ---

Ce module regroupe 5 problèmes sur les pourcentages, pour les classes de lycée.

Compétences mises en oeuvre :

- calculer des pourcentages de proportion (tous niveaux),
- calculer des pourcentages d'évolution (tous niveaux),
- calculer des pourcentages de pourcentages (tous niveaux),
- interpréter médiane et quartiles d'une série statistique (premières et terminales),

Chaque problème comporte 3 ou 4 questions. Les questions sont corrigées individuellement et des solutions rédigées sont proposées. Selon le niveau, l'élève a droit à un ou deux essai(s) pour répondre à une question.

Choisir un exercice :

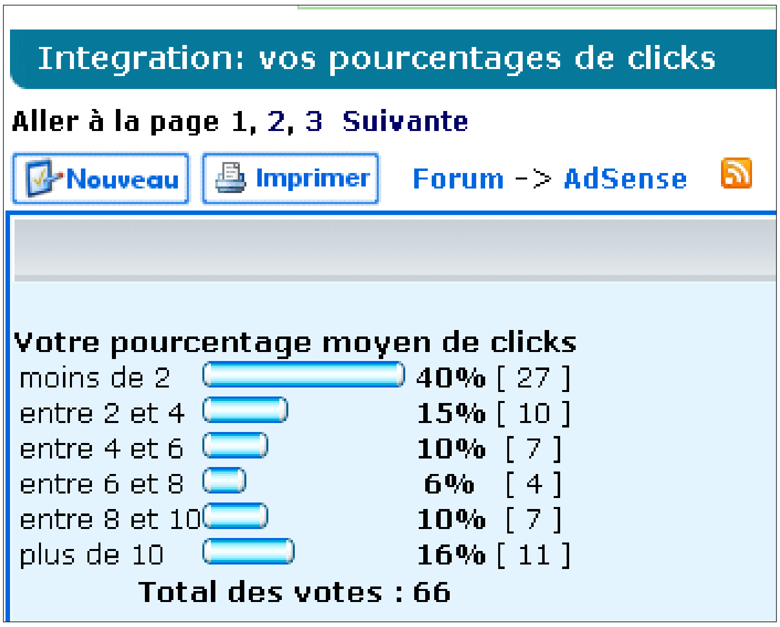
Petits calculs
Pourcentages de proportion
Pourcentages d'évolution
Pourcentages de pourcentages
Pourcentages et quantiles

Choisir un niveau :

facile difficile

Au travail !

➔ Chaque fois, il convient de déterminer la nature (l'origine, etc.) du document – ce qui ne semble pas trop difficile pour les documents précédents. Peut-être en va-t-il autrement pour le document ci-après que tel moteur de recherche propose en réponse à une requête des plus innocentes sur les pourcentages (http://www.webrankinfo.com/forums/viewtopic_82534.htm).



Sans doute, encore, y aura-t-il un effort de à faire pour comprendre ce qu'est cet autre document dont on a reproduit ci-après un extrait peut-être significatif (http://blog.alsacreations.com/2005/01/09/109-histoires-de-pourcentages-calcul-arrondi).

> Une petite expérience (fr) réalisée par FlorentG (fr) tend à montrer qu'il existe bien une différence entre les navigateurs dans le calcul des arrondis des pourcentages :
>
> - Internet Explorer fait un vrai arrondi mathématique (si 33% correspondrait à 19.7px, alors IE considère la valeur comme étant 20px)
> - Mozilla Firefox se contente simplement de supprimer la partie décimale (si 33% correspondrait à 19.7px, alors FF considère la valeur comme étant 19px)
>
> Dans le cas exposé sur le forum, les différence de calcul arrondi font que le total calculé par Internet Explorer est de 512px, ce qui dépasse du conteneur dont la taille est de 511px. Le dernier élément du menu est défini à 16%, soit 81.76px (arrondi à 82 sous IE et 81 sous FF).
>
> Cette différence dans le calcul arrondi des pourcentages (*qui s'applique d'ailleurs également aux unités en "em"*) pourrait expliquer certains problèmes courants de débordements dans les designs fluides.
>
> Je n'ai malheureusement trouvé que très peu de sources sur cette différence de calcul à part les deux liens et pistes suivantes :

➔ L'observation suppose alors le recueil, le classement, etc., ce qui pose des problèmes dans lesquels on n'entrera pas ici.

c) Au-delà de l'observation, c'est-à-dire de la constitution d'un ***corpus de données*** (prenant une forme quelconque : textes, nombres, images, sons, vidéos, etc.), une série de gestes restent à accomplir.

➔ Il convient bien sûr d'***analyser*** les données recueillies pour dresser un tableau des réponses $R^{\lozenge}$ sur lesquelles elles nous informent.

➔ Il convient alors d'***évaluer*** les $R^{\lozenge}$ analysés, c'est-à-dire d'en apprécier la ***valeur*** pour ce ***projet*** que constitue le PER dans lequel le recueil de ces données prend place.

➔ À partir de là, il convient de ***développer*** une réponse propre, $R^{\heartsuit}$, ce qui est un « geste » certes complexe mais qui doit constamment constituer ***l'horizon*** du travail dans lequel on se trouve concrètement engagé. (Dans le cas de la question $Q_0^{”}$, les types de tâches T_1 à T_8 balisent cet horizon : c'est à ces balises que nous nous référerons ci-dessous.)

➔ Il faudra ensuite ***défendre & illustrer*** la réponse $R^{\heartsuit}$ élaborée – situation classique que vit l'étudiant soutenant le mémoire dans lequel est exposée sa réponse $R^{\heartsuit}$ à la question qu'il a choisi d'étudier.

d) Tous ces « gestes » – ***observer***, ***analyser***, ***évaluer*** des réponses $R^{\lozenge}$, ***développer*** puis ***défendre & illustrer*** sa réponse $R^{\heartsuit}$ – appelleraient de multiples commentaires : on en verra quelques-uns au fil du travail restant à accomplir. Mais on revient ici à l'étude particulière esquissée : quelle réponse $R_0^{”\heartsuit}$ apporterons-nous à la question $Q_0^{”}$, « Quelles praxéologies trouve-t-on « dans les œuvres $O^{\%}_{\mathrm{Wf}}$ et $O^{\%}_{\mathrm{We}}$ » autour des types de tâches T_i ($1 \leq i \leq 8$) ? » On s'arrête d'abord sur le document $O^{\%}_{\mathrm{Wf}}$.

➔ Le premier souci est bien sûr d'identifier une réponse éventuelle à la question (technologique) « Qu'est-ce qu'un pourcentage ? » C'est par cela que commence, de fait, $O^{\%}_{\mathrm{Wf}}$. Mais, si la formulation donne l'***idée*** au cœur de la notion, elle ne donne d'abord que cela.

> Un **pourcentage** est une façon d'exprimer une proportion ou une fraction d'un ensemble.
>
> On compare une valeur particulière (ou une population partielle, population étant entendu au sens statistique) à une valeur de référence (ou une population totale) et on cherche à déterminer ce que vaudrait cette valeur particulière (ou cette population partielle) si la valeur de référence (ou la population totale) était ramenée à 100 sachant que les proportions sont respectées.
>
> Ainsi, si, dans une population de 400 personnes, 56 d'entre elles ont une particularité P, on rencontrerait, si les proportions étaient respectées, dans une population de 100 personnes, 14 personnes ayant la particularité P. On écrit alors que, parmi les 400 personnes (population de référence) 14 % (lu 14 pour cent) d'entre elles vérifient la particularité P.

Le nombre 14 semble résulter d'un « raisonnement proportionnel » : si une population de 400 individus contient 56 individus d'une certaine sorte, une population de 100 individus en aura, « si les proportions sont respectées », 4 fois moins, soit 56 ÷ 4 = 14. Mais il y a ici un cercle : la notion de pourcentage repose sur la notion de proportion, qui elle-même n'est pas définie !

➔ On peut commencer à ébaucher, de façon conjecturale, une réponse $R'^{\heartsuit}_0$: dans une population U de N individus, la proportion des individus appartenant à une sous-population $V \subset U$ ***serait*** égale ***par définition*** à (la valeur de) la fraction

$$\frac{n}{N}.$$

Cette définition (que $O^{\%}_{\mathrm{Wf}}$ omet d'expliciter) est en accord avec ce passage de $O^{\%}_{\mathrm{Wf}}$.

> Le calcul de ce pourcentage revient à trouver le numérateur d'une fraction dont le dénominateur serait 100 et qui serait égale à $\frac{56}{400}$.

Dans le cas pris en exemple, la proportion des 56 individus ayant la propriété P (sous-population V) dans la population U de 400 individus est donc donnée par la fraction

$$\frac{56}{400}.$$

Selon le passage cité de $O^{\%}_{\mathrm{Wf}}$, on doit ainsi déterminer le nombre x tel que

$$\frac{x}{100} = \frac{56}{400}.$$

➔ Pour calculer de x, $O^{\%}_{\mathrm{Wf}}$ donne implicitement une première technique (non explicitement justifiée) : on ***aurait*** ce que voici :

$$x = \frac{56}{400} \times 100 \ (= \frac{56}{4} = 14).$$

Mais $O^{\%}_{\mathrm{Wf}}$ enveloppe cela dans un développement qu'on peut trouver passablement obscur que l'on reproduit ici.

Le calcul de ce pourcentage revient à trouver le numérateur d'une fraction dont le dénominateur serait 100 et qui serait égale à $\frac{56}{400}$. C'est ainsi que l'on confond souvent la fraction de dénominateur 100 avec le pourcentage et donc le pourcentage avec le nombre décimal 0,14. Cette confusion, très pratique en mathématique, induit parfois des incompréhensions dans le domaine technique puisque l'on rencontre souvent l'indication de calcul suivante : pourcentage de personne ayant la particularité P : $\frac{56}{400} \times 100 = 14$.

On voit qu'il suffirait de remplacer « pourcentage » par « nombre » (ou par « effectif ») pour que l'énoncé incriminé soit innocenté :

… ***nombre*** de personne***s*** ayant la particularité P :

$$\frac{56}{400} \times 100 = 14.$$

La mention de confusions possibles, et la référence aux mathématiques peut engendrer une confusion : la fraction $\frac{14}{400}$, écrite 14 %, ne désignerait-elle pas le pourcentage cherché ? Celui-ci est-il autre chose que le nombre décimal 0,14 ? Pourquoi ces identifications conduiraient-elles à confondre 14 et 14 %, c'est-à-dire – selon l'interprétation « mathématique » – un nombre avec un nombre ***cent fois plus petit ?*** On va, pour le moment, laisser les choses en suspens sur tout cela.

➔ Le développement consacré à l'histoire du symbole % ne semble pas contribuer clairement à la construction de techniques τ_1, …, τ_8. On notera seulement les règles concernant l'emploi du signe % en usage dans les typographies française et anglaise.

Le signe « % » en typographie française doit être précédé et suivi d'une espace forte [2].

Dans d'autres langues, et notamment en anglais, le signe est collé au chiffre.

En d'autres termes, on écrit, en français, 14 % ; en anglais, on écrira 14%. Mais $O^{\%}_{\mathrm{Wf}}$ contient-il d'autres indications quant au type de tâches T_8 (« Exprimer [en pourcentage] la proportion, dans un ensemble donné U, des éléments appartenant à une partie V de U ») ? Sous la rubrique *Calculer un pourcentage*, on lit ceci.

Dans une assemblée de 50 personnes, il y a 31 femmes ; celles-ci représentent 62 % de l'assemblée car :

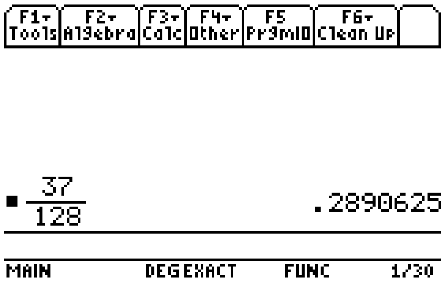
$$\frac{31}{50} = 0{,}62 = \frac{62}{100}$$

La ***technique*** τ_8 qui semble ici mise en œuvre serait la suivante : on écrit la fraction donnant la proportion (soit ici 31/50) et on effectue la division du numérateur par le dénominateur, obtenant ainsi un nombre « décimal » (on obtient ici 0,62) que l'on multiplie alors pour obtenir le numérateur de la fraction de dénominateur 100 cherchée (en l'espèce, donc, $\frac{62}{100}$). Que l'on dise alors (comme le rédacteur de $O^{\%}_{\mathrm{Wf}}$ semble le faire) que, à la fraction $\frac{62}{100}$, « correspond » un pourcentage de 62 % ou que l'on dise (comme il en irait « en mathématique***s*** », selon le même rédacteur) que l'on a

$$\frac{62}{100} = 62\ \%,$$

la conclusion est la même : le pourcentage des femmes dans l'assemblée est de 62 %.

➔ C'est là une technique qu'on peut inclure (sous bénéfice d'inventaire) dans $R_0'^{\heartsuit}$: pour calculer sous forme d'un pourcentage une proportion de, disons, 37/128, on divise le numérateur par le dénominateur (en utilisant de préférence une calculatrice), on multiplie par 100 le quotient obtenu, et on fait suivre ce résultat, convenablement arrondi, d'une espace forte suivie du signe %. Ainsi, on obtient d'abord ce que montre l'écran de la calculatrice ci-après.



On en tire aussitôt ceci : $\frac{37}{128} =_{\text{calc}} 0{,}2890625 = \frac{28{,}90625}{100} \approx \frac{28{,}91}{100} = 28{,}91\ \%$.

➔ La suite de $O^{\%}_{\mathrm{Wf}}$ fournit explicitement une seconde technique – qu'on notera τ_8' – déjà rencontrée plus haut ; en voici la présentation.

> On peut aussi voir le problème comme la recherche d'une quatrième proportionnelle : il s'agit de trouver *t* tel que :
>
> $\frac{31}{50} = \frac{t}{100}$ soit $t = \frac{31 \times 100}{50}$.

La désignation de cette technique mobilise une notion technologique supposée connue (apparemment), celle de « quatrième proportionnelle » (cette notion est classiquement étudiée en 5^{e}). On voit aussi que cette technique suppose d'introduire une inconnue (ici, *t*) et de « poser » une équation, avant de la résoudre par un geste stéréotypé ; on aura ainsi :

$$\frac{t}{100} = \frac{37}{128} \Leftrightarrow t = \frac{37 \times 100}{128} =_{\text{calc}} 28{,}90625 \approx 28{,}91.$$

La proportion considérée est donc de 28,9 % (environ). Notons que cette technique est plus exigeante, au double plan technologique et technique, que la technique τ_8. Bien entendu, on ne l'écarte nullement de $R_0'^{\heartsuit}$, qui doit proposer un ***inventaire*** et non pas une ***sélection***.

➔ La suite du document $O^{\%}_{\mathrm{Wf}}$ généralise la situation où un pourcentage exprime une proportion (une proportion est inférieure ou égale à 1).

> Quand on compare une valeur particulière à une valeur de référence, il est possible d'obtenir des pourcentages dépassant 100 %. Si le coût d'un produit passe de 30 euros à 48 euros et si on considère que le premier prix est une valeur de référence, le second prix représente 160 % du premier prix car :

$$\frac{48}{30} = 1{,}6 = \frac{160}{100}.$$

Cet aspect du pourcentage est particulièrement utilisé en économie dans la notion d'indice.

Sauf erreur, ce type de situations n'est pas rencontré dans les textes examinés dans la leçon 1. Mais on y rencontre le type de situations où, un prix passant de 30 € à 78 €, l'***augmentation*** du prix est de 48 €, et donc où le pourcentage d'augmentation est :

$$\frac{48}{30} = 1{,}6 = \frac{160}{100}.$$

Selon le choix que l'on a fait, on écrira que l'on a $\frac{160}{100}$ = 160 %, ou, comme le ferait le rédacteur de $O^{\%}_{\mathrm{Wf}}$, on dira que le pourcentage « associé » au décimal 1,6 est de 160 %.

➔ Le rédacteur de $O^{\%}_{\mathrm{Wf}}$ évoque alors un type de tâches – « Appliquer un pourcentage » – que l'on n'a pas rencontré jusqu'ici :

T'_8. Connaissant le pourcentage des éléments d'un ensemble V dans un ensemble $U \supset V$ et l'effectif de U, déterminer l'effectif de V.

Le texte est le suivant.

Appliquer un pourcentage, c'est retrouver la valeur étudiée (ou la population partielle) connaissant le pourcentage et la valeur (ou la population) de référence. Cette valeur étudiée se détermine en multipliant la valeur de référence par le décimal associé au pourcentage.

Si une assemblée de 120 personnes compte 15 % de femme *(sic)*, alors il y a 18 femmes dans cette assemblée car :

$$120 \times 0{,}15 = 18.$$

On peut aussi voir le problème comme la recherche d'une quatrième proportionnelle : il faut trouver n tel que :

$$\frac{n}{120} = \frac{15}{100} \text{ ce qui conduit à } n = 120 \times \frac{15}{100}.$$

À nouveau, on a affaire à un type de tâches qui n'apparaît pas dans les textes examinés lors de la leçon 1 : on n'ira donc pas plus loin ici sur ce sujet. Il en va de même pour le passage qui prolonge le développement précédent et que voici.

Le prix hors taxes d'un objet est 120 €. Le taux de TVA est de 5 %. Celle-ci s'élève donc à 6 € car :

$$120 \times 5\ \% = 120 \times \frac{5}{100} = 120 \times 0{,}05 = 6.$$

On rangera dans la même rubrique le développement qui suit.

Retrouver la valeur de référence **[modifier]**

Cette valeur de référence se trouve en divisant la valeur étudiée ou la population partielle par le décimal associé au pourcentage.

Dans une assemblée il y a 36 femmes, elles représentent 30 % de l'assemblée donc l'assemblée est formée de 120 individus car :

$\frac{36}{0,3} = 120.$

On peut aussi voir le problème comme la recherche d'une quatrième proportionnelle : il faut trouver N tel que :

$\frac{36}{N} = \frac{30}{100}$ soit $36 \times 100 = 30 \times N$ soit $N = \frac{36 \times 100}{30}$

Le prix TTC d'un objet est de 198 €. Ce prix représente 119,6 % du prix HT. Le prix HT (hors taxe) est donc de 165,55 € car

$\frac{198}{1,196} \approx 165,55.$

e) La suite de $O^{\%}_{\mathrm{Wf}}$ s'attache au thème du ***Pourcentage d'augmentation et de réduction***, plus proche de notre propre questionnement.

En économie et dans les taux d'intérêts, l'étude porte sur des variations en pourcentage, des augmentations ou des réductions. On peut tout à fait décomposer le calcul en deux temps : calcul de l'augmentation ou de la réduction, puis calcul de la valeur finale en effectuant une addition ou une soustraction. Mais il est préférable de voir ces augmentations ou ses réductions comme issues de l'application d'un coefficient multiplicateur. Seul cet aspect des choses permet de retrouver efficacement une valeur de référence ou d'appliquer des augmentations successives.

Calculer la valeur finale [modifier]

Une **augmentation** de t % se traduit par une multiplication par $1 + \frac{t}{100}$, et une **diminution** de t % par une multiplication par $1 - \frac{t}{100}$.

➔ L'idée soulignée est ici de passer à la notion de ***multiplicateur***. S'agissant du type de tâches T_1 (« Déterminer la valeur g_1 d'une grandeur G qui, partant de la valeur g_0, subit une augmentation de r % »), la technique τ_1 préconisée consiste à utiliser la formule

$$g_1 = \left(1 + \frac{r}{100}\right) g_0.$$

Nous pouvons, muni de cette technique, revenir vers plusieurs des problèmes soulevés par les textes trouvés sur le site de l'association Pénombre. « En quatre ans, écrivait un journaliste, en France, le montant des transactions a quadruplé : 400 % d'inflation, ce n'est pas mal ! » Si la variable G passe de la valeur g_0 à la valeur g_1 en augmentant de r % = 400 %, d'après τ_1 on a :

$$g_1 = \left(1 + \frac{r}{100}\right) g_0 = \left(1 + \frac{400}{100}\right) g_0 = (1 + 4)\, g_0 = 5\, g_0.$$

Si « l'inflation » avait été de 400 %, il y aurait eu quintuplement, et non quadruplement.

➔ Le même passage de $O^{\%}_{\mathrm{Wf}}$ fournit une technique τ_2 relative au type de tâches T_2 (« Déterminer la valeur g_1 d'une grandeur G qui, partant de la valeur g_0, subit une diminution de r % ») : elle revient à appliquer la formule

$$g_1 = \left(1 - \frac{r}{100}\right) g_0.$$

Un autre rédacteur de Pénombre écrivait qu'il serait plus indiqué « quand la baisse est importante, de parler d'une division par 2, plutôt que d'une baisse de 50 %, d'une division par 3, plutôt que d'une baisse de 66,6666666… % ». Si l'on prend $r\%$ = 50 % dans la formule précédente, il vient

$$g_1 = \left(1 - \frac{r}{100}\right) g_0 = \left(1 - \frac{50}{100}\right) g_0 = \ldots = \frac{g_0}{2}.$$

De même, si l'on prend $r\ \% = \frac{200}{3}\ \%$ (plutôt que 66,6666666… %), on obtient

$$g_1 = \left(1 - \frac{r}{100}\right) g_0 = \left(1 - \frac{200}{300}\right) g_0 = \ldots = \frac{g_0}{3}.$$

Il y a bien, dans le premier cas division par 2, dans le second division par 3.

➔ Le même rédacteur posait cette question : une inflation de 8000 %, cela fait une multiplication par combien ? La technique τ_1 permet de répondre :

$$g_1 = \left(1 + \frac{r}{100}\right) g_0 = \left(1 + \frac{8000}{100}\right) g_0 = (1 + 80)\, g_0 = 81\, g_0.$$

On voit que, de la technique τ_1, on peut déduire une technique τ_7 pour accomplir les tâches du type T_7 (« T_7. Déterminer le nombre entier k tel que $g_1 = kg_0$ lorsque qu'une grandeur G, partant de la valeur g_0, prend la valeur g_1 après une augmentation de 100 r % ») : on a en effet plus généralement

$$g_1 = \left(1 + \frac{100r}{100}\right) g_0 = (1 + r)\, g_0.$$

➔ Le texte que nous suivons précise alors ceci.

> Des variations successives à taux fixe conduisent à des progressions géométriques. Ainsi, augmenter 35 fois de 2 % revient à multiplier par $1{,}02^{35}$, c'est-à-dire 1,99989, soit quasiment par 2. Et diminuer 35 fois de 2 % revient à multiplier par 0,9835, c'est-à-dire à diviser par 2,028, soit un peu plus de 2.

Ce passage est plus obscur pour qui n'est pas familier du domaine. Il semble en outre ne renvoyer à aucune des interrogations soulevées par les textes examinés. On le laissera donc de côté, avec l'idée d'y revenir éventuellement à partir des apports de $O^{\%}_{\mathrm{Wf}}$.

f) Ce qui manque, à ce stade, ce sont des techniques permettant d'accomplir les tâches des types suivants.

T_3. Déterminer le pourcentage d'augmentation r % d'une grandeur G qui, partant de la valeur g_0, prend la valeur $g_1 > g_0$.

T_4. Déterminer le pourcentage de diminution r % d'une grandeur G qui, partant de la valeur g_0, prend la valeur $g_1 < g_0$.

T_5. Déterminer le pourcentage d'augmentation r % d'une grandeur G qui, partant de la valeur g_0, prend la valeur $g_1 = kg_0$, où k est un nombre entier > 1.

T_6. Déterminer le pourcentage de diminution r % d'une grandeur G qui, partant de la valeur g_0, prend la valeur $g_1 = g_0/k$, où k est un nombre entier > 1.

➔ La technique τ_1 permet de produire, moyennant un petit travail algébrique, une technique τ_3 relative au type de tâches T_3 (« Déterminer le pourcentage d'augmentation r % d'une grandeur G qui, partant de la valeur g_0, prend la valeur $g_1 > g_0$ »). L'égalité

$$g_1 = \left(1 + \frac{r}{100}\right) g_0$$

s'écrit en effet $1 + \frac{r}{100} = \frac{g_1}{g_0}$ et donc $\frac{r}{100} = \frac{g_1}{g_0} - 1 = \frac{g_1 - g_0}{g_0}$; d'où : $r = 100\,\frac{g_1 - g_0}{g_0}$.

➔ On déduit de là aussitôt une technique τ_5 relative à T_5 (« Déterminer le pourcentage d'augmentation r % d'une grandeur G qui, partant de la valeur g_0, prend la valeur $g_1 = kg_0$, où k est un nombre entier > 1 »). Si $g_1 = kg_0$, il vient en effet

$$r = 100\,\frac{g_1 - g_0}{g_0} = 100\,\frac{kg_0 - g_0}{g_0} = 100(k-1).$$

Si $k = 4$ (la quantité quadruple), le taux d'augmentation est de r % = 300 %, etc.

➔ La technique τ_2 permet de produire de même une technique τ_4 relative à T_4 (« Déterminer le pourcentage de diminution r % d'une grandeur G qui, partant de la valeur g_0, prend la valeur $g_1 < g_0$ »). L'égalité

$$g_1 = \left(1 - \frac{r}{100}\right) g_0$$

s'écrit en effet $1 - \frac{r}{100} = \frac{g_1}{g_0}$ et donc $\frac{r}{100} = 1 - \frac{g_1}{g_0} = \frac{g_0 - g_1}{g_0}$; d'où $r = 100\,\frac{g_0 - g_1}{g_0}$. On peut ici revenir sur cette affirmation d'un journaliste du Figaro : « … il faut à peine 10 mètres pour s'arrêter en freinage d'urgence sur ce substrat ultra-râpeux, contre plus de 14 mètres sur une chaussée classique, ce qui représente une diminution de plus de 40 % de la distance d'arrêt. » On a ici $g_0 = 14$, $g_1 = 10$, et donc

$$r = 100\,\frac{g_0 - g_1}{g_0} = 100\,\frac{14 - 10}{14} = \frac{400}{14} \approx 28{,}6.$$

La diminution est inférieure à 30 % ; elle n'est donc pas « de plus de 40 % ».

➔ On déduit de τ_4 une technique τ_6 relative à T_6 (« Déterminer le pourcentage de diminution r % d'une grandeur G qui, partant de la valeur g_0, prend la valeur $g_1 = g_0/k$, où k est un nombre entier > 1 »). Si $g_1 = g_0/k$, il vient en effet

$$r = 100\,\frac{g_0 - g_1}{g_0} = 100\,\frac{g_0 - g_0/k}{g_0} = 100\,\frac{k-1}{k}.$$

Si $k = 2$ (la quantité est divisée par deux), le taux de diminution est r % = 50 % ; si $k = 3$ (la quantité est divisée par trois), le taux de diminution est $r \,\% = \frac{200}{3} \,\% = 66{,}66666\ldots \,\%$.

➔ Une des erreurs épinglées par les rédacteurs de Pénombre est le fait de parler d'une baisse « entre 300 et 400 % de notre niveau de vie », ou d'une « érosion du pouvoir d'achat de 1000 % ». Si grande soit la valeur de g_0 et si petite la valeur de $g_1 < g_0$, on aura un pourcentage de baisse de

$$r\,\% = \frac{g_0 - g_1}{g_0}.$$

Ce pourcentage de baisse est d'autant plus grand que g_1 est petit. Si le « pouvoir d'achat » a baissé jusqu'à être nul, donc si $g_1 = 0$, on aura

$$r\,\% = \frac{g_0 - g_1}{g_0} = \frac{g_0}{g_0} = 1 = 100\,\%.$$

Ainsi la baisse du pouvoir d'achat ne peut-elle être supérieure à 100 % : il est absurde de parler d'une baisse de 400 % par exemple.

g) Le travail réalisé sur $O^{\%}_{\mathrm{Wf}}$ doit être contrôlé de différentes façons.

➔ Une possibilité offerte est d'examiner $O^{\%}_{\mathrm{We}}$. Nous y relevons d'abord la déclaration suivante, qui contraste avec le point de vue exprimée dans $O^{\%}_{\mathrm{Wf}}$.

> In mathematics, a percentage is a way of expressing a number as a fraction of 100 (per cent meaning "per hundred"). It is often denoted using the percent sign, "%". For example, 45% (read as "forty-five percent") is equal to 45/100, or 0.45.

➔ Arrêtons-nous encore sur le passage suivant.

> In general, if an increase of x percent is followed by a decrease of x percent, the final amount is $(1 + 0.01x)(1 - 0.01x) = 1 - (0.01x)^2$ times the initial amount — thus the net change is an overall decrease by x percent of x percent (the square of the original percent change when expressed as a decimal number).

➔ On peut traduire ce qui précède dans le langage employé jusqu'ici. Soit G une grandeur qui prend d'abord la valeur g_0 ; elle subit alors une ***augmentation*** de r % et prend ainsi la valeur g_1 donnée par

$$g_1 = \left(1 + \frac{r}{100}\right) g_0.$$

Elle subit ensuite une ***baisse*** de r % et prend la valeur g_2 donnée par

$$g_2 = \left(1 - \frac{r}{100}\right) g_1.$$

On a ainsi : $g_2 = \left(1 - \frac{r}{100}\right) g_1 = \left(1 - \frac{r}{100}\right)\left(1 + \frac{r}{100}\right) g_0$. Un peu d'algèbre élémentaire permet de voir que l'on a

$$g_2 = \left(1 - \frac{r^2}{10\,000}\right) g_0.$$

On ne retrouve donc pas g_0 : voir son salaire augmenter de 10 % puis baisser aussitôt de 10 % équivaut à une ***baisse*** de 1 %.

➔ On peut maintenant revenir à un passage de $O^{\%}_{\mathrm{Wf}}$ délaissé plus haut.

> Des variations successives à taux fixe conduisent à des progressions géométriques. Ainsi, augmenter 35 fois de 2 % revient à multiplier par $1{,}02^{35}$, c'est-à-dire 1,99989, soit quasiment par 2. Et diminuer 35 fois de 2 % revient à multiplier par 0,9835, c'est-à-dire à diviser par 2,028, soit un peu plus de 2.

Traduction : on part de g_0 ; une augmentation de r % = 2 % = 0,02 conduit à

$$g_1 = (1 + 0{,}02)\, g_0 = 1{,}02\, g_0.$$

Une deuxième augmentation de 2 % conduit alors à

$$g_2 = 1{,}02\, g_1 = 1{,}02\,(1{,}02\, g_0) = 1{,}02^2\, g_0.$$

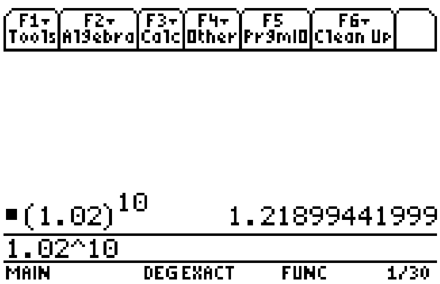
Une troisième augmentation de 2 % fait passer à

$$g_3 = 1{,}02\, g_2 = 1{,}02\,(1{,}02^2\, g_0) = 1{,}02^3\, g_0.$$

On imagine que, plus généralement, s'il se produit n augmentations successives, la grandeur G prendra la valeur

$$g_n = 1{,}02^n\, g_0.$$

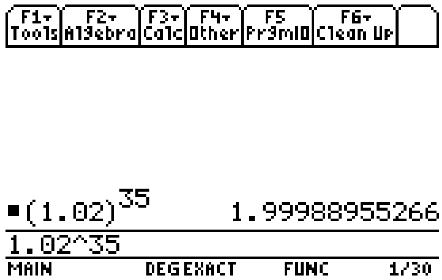
Pour n = 10, par exemple, on obtient ce que montre l'écran de calculatrice ci-après.



On aura ainsi : $g_{10} \approx 1{,}22\, g_0 = \left(1 + \frac{22}{100}\right) g_0$. Ainsi 10 augmentations successives de 2 % produisent-elles une augmentation de près de 22 %. Dans le cas de 35 augmentations successives de 2 %, on a

$$g_{35} = 1{,}02^{35}\, g_0.$$

On a ceci.



Il vient donc : $g_{35} \approx 2\, g_0 = \left(1 + \dfrac{100}{100}\right) g_0$. L'augmentation est voisine de 100 % : la valeur est multipliée par 2.

➔ Pour des diminutions successives de 2 %, on obtient d'abord

$$g_1 = (1 - 0{,}02)\, g_0 = 0{,}98\, g_0$$

puis

$$g_2 = 0{,}98\, g_1 = 0{,}98(0{,}98\, g_0) = 0{,}98^2\, g_0$$

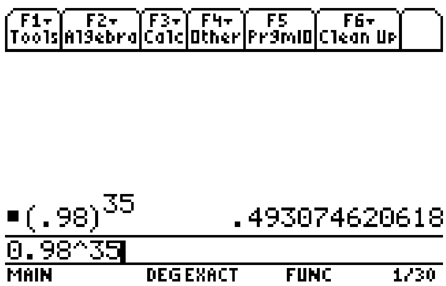
et, plus généralement,

$$g_n = 0{,}98^n\, g_0.$$

Pour 35 diminutions successives, on obtient

$$g_{35} = 0{,}98^{35}\, g_0.$$

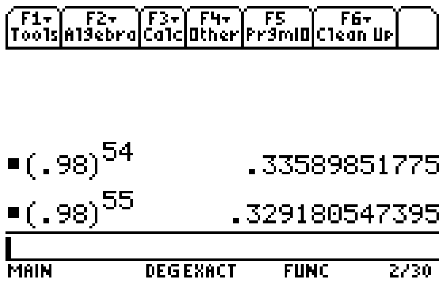
On a ceci.



Il vient : $g_{35} \approx 0{,}49\, g_0 = (1 - 0{,}51)\, g_0 = \left(1 - \dfrac{51}{100}\right) g_0$. La diminution atteint presque 51 % : la valeur de G est divisée par un peu plus de 2. Pour diviser par 3 au moins la valeur g_0 il faudrait prendre n de façon que

$$g_n = 0{,}98^n\, g_0 \leq \frac{1}{3}\, g_0$$

c’est-à-dire tel que $0{,}98^n \le \frac{1}{3}$. Si l’on ignore les logarithmes, on peut, à l’aide d’une calculatrice, déterminer le premier entier n vérifiant cette inégalité ; la calculatrice montre qu’il s’agit de $n = 55$.



h) Le travail d’enquête accompli jusqu’ici permet de nous faire une idée de ce que ***pourraient être*** les praxéologies mises en jeu par les rédacteurs de Pénombre. Bien entendu, ces conclusions restent conjecturales. Elles peuvent encore être mises à l’épreuve sur un texte laissé de côté lors de la leçon 1 et que l’on reproduit ici.

> Il fut un temps en France où la TVA sur les voitures était de 33 %, ce qui faisait dire à de nombreux acheteurs qu’un tiers de leur facture allait à l’État. En fait, ils ne versaient qu’un quart à l’État. En effet, une voiture dont le prix hors taxes était de 100, était vendue 133 et ce prix de vente comportait donc 25 % de taxes (33/133). À l’heure actuelle ce taux, le taux normal, est de 19,6 %, soit environ un cinquième du prix hors taxes, ce qui conduit à un sixième du prix de vente (19,6/119,6).

➔ Dans le cas de la TVA, nous serions portés à dire que, si une voiture a pour prix hors taxes g_0, le prix TVA incluse est de $1{,}33\, g_0$. Du prix payé par le consommateur, $1{,}33\, g_0$, le pourcentage qui « va à l’État » est donné par

$$\frac{0{,}33\, g_0}{1{,}33\, g_0}.$$

On voit que ce taux ***ne dépend pas*** de g_0 : d’où la technique « élémentaire » consistant à « se ramener à 100 » : la voiture est payée 133 euros, tandis que 33 euros « vont à l’État », etc. Le rapport précédent est égal à

$$\frac{33}{133} \approx 0{,}248 = 24{,}8\ \%.$$

La part qui « va à l’État » est donc même un peu inférieure à 25 %. Pour une TVA de 19,6 %, la même part vaut

$$\frac{19{,}6}{119{,}6} \approx 0{,}164 = 16{,}4\ \%.$$

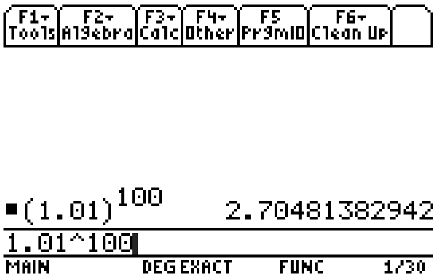
Elle est un peu inférieure au sixième du prix payé par le consommateur puisque

$$\frac{1}{6} \approx 16{,}7\ \%.$$

➔ On met ainsi en évidence une technique qu'on n'avait pas observée explicitement encore : se ramener à une « base » 100, puis effectuer des « règles de trois » et autres petits bricolages calculatoires : si je gagne 100 et qu'on m'augmente de 10 %, je gagne désormais 110 ; mais si on me diminue de 4 %, je perdrai d'abord 4 (sur 100) puis (sur 10), 0,4, soit en tout 4,4, en sorte que je serai ramené à 110 – 4,4 = 105,6 : l'augmentation « finale » sera de 5,6 %. Par la technique « algébrique », ce même résultat s'obtiendrait ainsi (voir plus haut) :

$$g_2 = \left(1 - \frac{4}{100}\right) g_0 = 0{,}96\, g_1 = 0{,}96 \left(1 + \frac{10}{100}\right) g_0 = 0{,}96 \times 1{,}1\, g_0 = 1{,}056\, g_0 = \left(1 + \frac{5{,}6}{100}\right) g_0.$$

On voit ainsi que la variation finale du salaire consiste en une augmentation de 5,6 %. Il est facile de s'assurer – en considérant par exemple le problème de calculer l'augmentation correspondant à 100 augmentations successives de 1 % (la quantité est multipliée par 2,7 environ : voir ci-après) – que la technique « simple » consistant à « se ramener à 100 » a une moindre ***portée*** que la technique algébrique.



➔ On n'est jamais sûr d'avoir exploré suffisamment les réponses $R^\lozenge$ présentes dans la culture ambiante – ce qu'on appellera l'***offre praxéologique***. Par exemple, ici, on peut avoir appris (et n'avoir pas oublié) la technique « intermédiaire » qui repose sur la définition suivante (laquelle ignore la notion de ***multiplicateur*** mise en avant par $O^{\%}_{\mathrm{Wf}}$) : le pourcentage d'augmentation d'une grandeur G qui passe de la valeur g_0 à la valeur g_1 est donné par

$$\frac{g_1 - g_0}{g_0}.$$

Si donc on a (par exemple) $g_1 = 3\, g_0$ il vient : $\frac{g_1 - g_0}{g_0} = \frac{3g_0 - g_0}{g_0} = \frac{2g_0}{g_0} = 2 = 200\ \%$. Si $g_1 = 0{,}2\, g_0$, de même, on a : $\frac{g_1 - g_0}{g_0} = \frac{0{,}2g_0 - g_0}{g_0} = \frac{-\,0{,}8\, g_0}{g_0} = -\,0{,}8 = -\,80\ \%$. La ***diminution*** est de 80 %. Si $g_1 = 0\, g_0$, on a : $\frac{g_1 - g_0}{g_0} = \frac{0\, g_0 - g_0}{g_0} = \frac{-\, g_0}{g_0} = -\,1 = -\,100\ \%$. Cette fois, la diminution est de 100 %, et c'est le « pire » que l'on puisse faire !

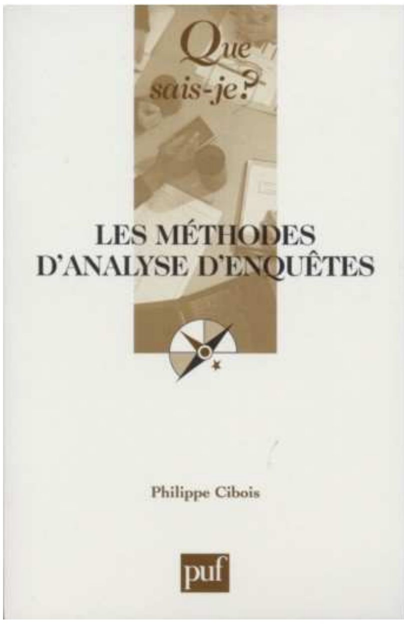
1. Dialectique de l'excription et de l'inscription : « Comment analyser des données ? »

a) Les deux dialectiques illustrées dans la leçon 1 – la dialectique ***des médias et des milieux*** et la dialectique ***des boîtes noires et des boîtes claires*** – constituent deux points faibles traditionnels de l'équipement didactique ***personnel*** et ***institutionnel*** construit à l'école. On abordera ici concrètement une troisième dialectique qu'on a pratiqué dans ce qui précède, dont le statut scolaire est sans doute un peu meilleur, mais qui s'évanouit rapidement hors des exigences intellectuelles et didactiques de l'institution scolaire.

b) On suppose pour cela un étudiant en sciences de l'éducation, *x*, qui souhaite apprendre à « analyser des données », type de tâches d'ampleur gigantesque mais difficilement évitable. Cet étudiant se pose donc la question $Q^\#$ suivante :

$Q^\#$. Comment analyser des données ?

On suppose que, pour tenter d'apporter de premiers éléments de réponse à $Q^\#$, *x* se procure un exemplaire d'un petit ouvrage assez récemment paru (en janvier 2007) dans la collection « Que sais-je ? » (n° 3782, PUF), signé de Philippe Cibois et intitulé *Les méthodes d'analyse d'enquêtes*.



c) Dans ce qui suit, on amorce l'examen du chapitre I de cet ouvrage, intitulé *Repérer les questions pertinentes*, et cela avec, en tête, la question $Q^\#$ ci-dessus Au fur et à mesure de la lecture, on s'arrête sur les points jugés obscurs – les boîtes noires, ou d'un niveau de gris soutenu – comme sur autant de ***questions auxiliaires*** que l'on envisage alors d'étudier sans délai ou, au contraire, que l'on accepte comme telle – en prenant ***dans chacun des cas*** un risque quant à la progression du PER dans lequel on est engagé. Voici donc le début du chapitre I : on en présentera ci-après un bref commentaire.

Chapitre I

REPÉRER LES QUESTIONS PERTINENTES

On suppose une enquête déjà existante, soit issue d'une recherche, soit en vue de l'analyse secondaire d'une enquête disponible et rendue accessible aux chercheurs. On suppose aussi que les données de cette enquête sont utilisables par le biais soit d'un logiciel international comme SAS ou SPSS, soit d'un logiciel libre comme Trideux, développé par l'auteur et dont les exemples de ce livre sont issus. Les méthodes de dépouillement présentées ici sont indépendantes des logiciels : les aspects pratiques en dépendent évidemment et il faudra s'y reporter pour plus de détails.

➔ L'auteur évoque d'abord la disponibilité des résultats d'une enquête et fait ensuite référence à des logiciels d'analyse de données (SAS, SPSS, Trideux). Toutefois la remarque selon laquelle « les méthodes de dépouillement présentées ici sont indépendantes des logiciels » permet au lecteur de considérer que, ***pour le moment***, il peut laisser à ces logiciels le statut de ***boîtes noires*** – même s'il est prévenu qu'il « faudra s'y reporter pour plus de détails ».

➔ À propos de la notion d'***analyse secondaire*** utilisée par l'auteur, on trouve dans l'article "Secondary data" de Wikipedia l'explicitation suivante :

> Secondary data analysis is commonly known as second-hand analysis. It is simply the analysis of preexisting data in a different way or to answer a different question than originally intended. Secondary data analysis utilizes the data that was collected by someone else in order to further a study that you are interested in completing.

Cette indication simple peut suffire – admettons-le, du moins – pour une première étude ***excriptrice*** du texte, c'est-à-dire une étude qui s'attache à expliciter le texte, à ***« excrire »*** ce qui y a été ***inscrit*** par l'auteur.

d) Voici maintenant le paragraphe suivant.

> On suppose donc que l'on a un fichier d'individus dont le nombre est variable, qui peut aller de quelques dizaines à plusieurs centaines de milliers : il peut sembler paradoxal d'envisager un dépouillement d'enquêtes avec moins de cent individus mais regarder attentivement le contenu d'un fichier est un objectif valable même si la possibilité d'étendre les résultats obtenus à une population de référence est faible. Quand on fait une enquête, quel que soit le nombre d'individus, on veut légitimement avoir une description de la population enquêtée : si l'effectif en est faible, on ne pourra que constater l'état de la population ; si l'effectif est important on pourra généraliser les résultats à la population dont l'enquête est issue, sous réserve que l'échantillon a été prélevé de manière raisonnée, par exemple par la méthode des quotas ou en sélectionnant des populations spécifiques. Il faut bien distinguer la description des données d'une part, des résultats qui peuvent être généralisés à l'ensemble de la population étudiée d'autre part. Pour pouvoir généraliser, on utilisera des tests statistiques, essentiellement celui du khi-deux que l'on supposera connu : on se souviendra que le khi-deux étant sensible aux effectifs, dès qu'une population d'enquêtés devient importante, il devient rare que le khi-deux d'un tableau croisé ne soit pas significatif.

➔ Ce paragraphe est structuré autour des trois termes d'***individu***, d'***échantillon*** (fichier) et de ***population***. La première notation de l'auteur concerne l'***effectif*** de l'échantillon : le public auquel il pense s'adresser est sans doute réputée friand de « gros » effectifs (de l'ordre de plusieurs centaines) et il le met donc en garde contre le fait de s'obnubiler sur ce point : il convient, souligne-t-il, de « regarder attentivement » même un « petit » fichier de données.

➔ Une deuxième notation concerne bien entendu le fait de passer des conclusions établies sur l'échantillon à des conclusions sur la population « enquêtée ». C'est rappeler d'abord que l'échantillon correspondant au fichier de données n'est là ***en principe*** que pour essayer de mieux connaître la ***population*** dont il est issu. Un point paraît toutefois obscur dans le texte de l'auteur : « si l'effectif en est faible, y lit-on, on ne pourra que constater l'état de la population… » Si l'effectif est faible, on pourra constater l'état de cette sous-population qu'est ***l'échantillon***, sans pouvoir remonter jusqu'à la population. Sur ce point, l'enquête sur le texte examiné devrait continuer…

➔ Pour pouvoir passer de l'échantillon à la population, précise alors l'auteur, il convient que l'échantillon ait été extrait « de manière raisonnée ». La mention de la ***méthode des quotas*** ressemble ici à celle d'une boîte noire. En voici tout de même une petite explication empruntée au site Internet d'Ipsos (http://www.ipsos.fr/CanalIpsos/cnl_static_content.asp?rubId=35#02).

Quels sont les avantages et les inconvénients de la méthode des quotas ?

Par rapport à la méthode aléatoire, celle des quotas a l'avantage d'être plus rapide. Avec l'aléatoire, les sondés ne sont pas « interchangeables ». Cela signifie que la personne tirée au sort doit être recontactée autant de fois que nécessaire. Grâce aux quotas, il est possible de remplacer un sondé par un autre qui a les mêmes caractéristiques socio-démographiques. Cela permet de réaliser un sondage dans des délais plus courts.

L'inconvénient majeur de la méthode des quotas est de ne pas permettre de calculer scientifiquement la marge d'erreur du sondage. Les lois statistiques qui permettent de la déterminer ne valent théoriquement que pour les sondages aléatoires. En pratique, on considère cependant que la marge d'erreur des sondages par quotas est égale ou inférieure à celle des sondages aléatoires.

Ce passage a le mérite d'attirer indirectement l'attention sur ce fait que l'auteur que nous suivons semble curieusement oublier la ***« méthode aléatoire »***, c'est-à-dire le tirage au hasard, dans la population, des individus composant l'échantillon. Par ailleurs, le travail d'excription à effectuer à propos de la mention de la technique de sélection de ***« populations spécifiques »*** exigerait une enquête que l'on ne fera pas ici – ce qui, à nouveau, laisse une boîte noire dans le texte examiné.

➔ L'auteur insiste à nouveau sur le travail de ***description*** des données, possible même quand l'effectif est faible, ainsi que sur les quelques résultats de ce travail que l'on peut envisager de généraliser à la population. L'outil pour « généraliser », ajoute-t-il, est la théorie des ***tests statistiques***. L'auteur fait un sort particulier au test du χ^2 qu'il suppose explicitement être pour le lecteur une « boîte claire », ce qui n'est rien moins que sûr ! Si l'on ne veut pas faire qu'une boîte claire ***supposée*** demeure en vérité une boîte noire, il convient de lancer ici un ***PER auxiliaire*** pour répondre à la question : « Qu'est-ce que le test du χ^2 ? ».

e) On suppose, ***dans le cadre de cette leçon***, qu'une enquête a été menée (par y ou par des $x \in X$), dont on présente ici quelques-uns des résultats – que, par cela, on ***inscrit*** donc en un texte plus ou moins organisé.

➔ Le point de départ tient dans un schéma de raisonnement, d'abord déterministe, ensuite aléatoire, qui va éclairer d'une façon générale ce que l'on fait dans un ***test statistique***. Soit deux propositions p et q. On a le schéma de raisonnement suivant (que l'on peut appeler « raisonnement par contraposition ») :

Si p était vraie, alors q serait fausse. [Cette implication est supposée établie et connue.]
Or q est vraie. [C'est un constat empirique.]
Donc p est fausse. [Conclusion qui résulte des deux « prémisses ».]

Modifions un peu le vocabulaire et les notations. On suppose que l'on considère une certaine hypothèse H_0 (dite en statistique ***hypothèse nulle***) et un certain événement E, liés par l'énoncé suivant :

Si H_0 était vraie, alors l'événement E ne pourrait pas se produire. [Démontré]
Or on a observé la survenue de E. [Constaté]
Donc H_0 est fausse. [Déduit de ce qui précède.]

On écarte maintenant la conclusion touchant H_0 pour ne retenir que le raisonnement suivant.

Si H_0 était vraie, alors l'événement E ne pourrait pas se produire. [Démontré]
Or on a observé la survenue de E. [Constaté.]
Donc on a observé la survenue d'un événement impossible. [Déduit.]

Passons maintenant au cas ***non déterministe***. On suppose ici que l'on dispose d'un ***modèle probabiliste*** qui fournit la probabilité de survenue de l'événement E sous l'hypothèse H_0, ce qu'on note classiquement

$$\mathrm{P}(E \mid H_0)$$

et qu'on lit dans un langage abrégé : « probabilité de E sachant H_0 » (= sachant que H_0 est vraie). On a alors le schéma de raisonnement suivant, où δ est un nombre positif inférieur ou égal à 1 :

Si H_0 était vraie, alors l'événement E aurait une probabilité $\mathrm{P}(E \mid H_0)$ inférieure à δ. [Démontré dans le modèle probabiliste.]
Or on a observé la survenue de E. [Constaté empiriquement.]
Donc on a observé la survenue d'un événement de probabilité inférieure à δ. [Déduit de ce qui précède.]

Supposons alors que l'on se soit fixé la règle d'action suivante : si la probabilité

$$\mathrm{P}(E \mid H_0)$$

est inférieure ou égale à, disons, 1/10 000 = 0,0001, et si E se produit, je considérerai que ce résultat n'est pas « croyable », en ce sens que ***je ne peux pas croire*** que je viens de voir se produire sous mes yeux un événement si improbable ; je déciderai alors ***de tenir H_0 pour fausse*** : on dira que je ***rejette*** H_0 (tout en étant conscient que je peux faire erreur : il est des événements plus rares encore qui se produisent !).

➔ On peut s'excentrer de ce schéma « subjectif » ***en l'objectivant***. Fixons-nous un ***seuil*** entre ***rejet*** de H_0 et ***acceptation*** de H_0 : on choisit pour cela un nombre ε tel que $0 < \varepsilon < 1$, par exemple $\varepsilon = 1/1000 = 0{,}001$. Si on a observé la survenue d'un événement E tel que

$$\mathrm{P}(E \mid H_0) < \varepsilon$$

on dira alors que ***H_0 est rejeté au seuil ε***. Si ε est grand, par exemple $\varepsilon = 0{,}5 = 50$ %, on va ***rejeter H_0 à tort*** en bien des cas : c'est là ce qu'on nomme le risque de ***première espèce***. Si ε est très petit, par exemple $\varepsilon = 10^{-6} = 0{,}000001$, on va souvent ***accepter H_0 à tort*** : c'est le risque de ***deuxième espèce***.

➔ Voici maintenant un exemple d'utilisation du test du χ^2. On considère une ***population*** de personnes. On suppose qu'on en a extrait au hasard un échantillon de 6800 personnes sur lesquelles on a prélevé chaque fois deux informations, la couleur de yeux, X, et la couleur des cheveux, Y. On pense que ces deux caractères ***ne sont pas indépendants*** – que par exemple, quand on a les cheveux blonds, on a plus souvent les yeux bleus, etc. Considérons le tableau suivant.

	Blond	Brun	Noir	Roux	Totaux
Bleu					2811
Gris-vert					3132
Brun					857
Totaux	2829	2632	1223	116	6800

On va alors formuler l'hypothèse nulle H_0 selon laquelle X et Y ***seraient indépendants*** – avec, bien sûr, l'idée que cette hypothèse H_0 est ***fausse***, et l'espoir de pouvoir ***la rejeter***, même en étant exigeant, par exemple en prenant un seuil ε égal à 1/10 000 = 0,0001. Comme opère-t-on ? Si X et Y étaient indépendants, la distribution des différentes couleurs de cheveux parmi les personnes aux yeux bleus serait en principe ***la même*** que dans ***l'échantillon tout entier*** : selon ce principe, il y aurait par exemple une proportion de

$$\frac{2829}{6800} \approx 41{,}6\ \%$$

des enquêtés qui auraient les yeux bleus, soit exactement

$$\frac{2829}{6800} \times 2811$$

personnes ayant les yeux bleus. On aurait pu aussi bien calculer la proportion de personnes ayant les cheveux blonds parmi les personnes ayant les yeux bleus : elle devrait être la même que dans l'échantillon tout entier, soit $\frac{2811}{6800}$, en sorte que le nombre de personnes ayant à la fois les yeux bleus et les cheveux blonds serait égal à

$$\frac{2811}{6800} \times 2829.$$

On voit que l'on retrouve ainsi le même effectif « théorique », que l'on peut écrire ainsi :

$$c_{11} = \frac{2829 \times 2811}{6800} \approx 1170.$$

On remplirait de même les autres cellules du tableau :

	Blond	Brun	Noir	Roux	Totaux
Bleu	c_{11}	c_{12}	c_{13}	c_{14}	2811
Gris-vert	c_{21}	c_{22}	c_{23}	c_{24}	3132
Brun	c_{31}	c_{32}	c_{33}	c_{33}	857
Totaux	2829	2632	1223	116	6800

(On verra plus loin qu'on n'a plus aujourd'hui à calculer soi-même les effectifs théoriques c_{ij}.) Voici maintenant le tableau « empirique » formés par les effectifs ***observés***.

	Blond	Brun	Noir	Roux	Totaux
Bleu	1768	807	189	47	2811
Gris-vert	946	1387	746	53	3132
Brun	115	438	1223	16	857
Totaux	2829	2632	1223	116	6800

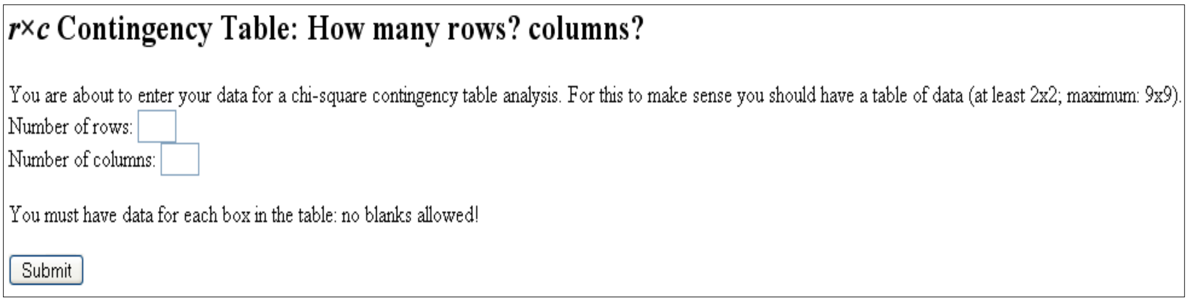
On voit que, dans le cas des personnes à yeux bleus et cheveux blonds, ce tableau observé ***s'éloigne*** du tableau théorique : l'effectif observé (1768) est, en l'espèce, assez nettement ***supérieur*** à l'effectif théorique (1170). Cet écart entre effectif théorique et effectif empiriquement observé, ainsi que les autres écarts (en plus ou en moins) que nous pourrions calculer, sont-ils alors explicables par la seule ***fluctuation d'échantillonnage***, qui peut certes expliquer de petits écarts entre effectifs observés et effectifs théoriques ? C'est là qu'intervient le teste du χ^2.

➔ Le χ^2 est un ***indicateur numérique*** qui évalue en un certain sens la ***dissemblance*** entre les deux tableaux : si on note n_{ij} les effectifs ***observés***, on a de façon générale

$$\chi^2 = \sum_{i,j} \frac{(n_{ij} - c_{ij})^2}{c_{ij}}.$$

Ici, le calcul donne : $\chi^2 \approx 1075$. L'événement E évoqué plus haut est le fait que $\chi^2 \geq 1075$. Or il se trouve que la probabilité d'observer un tel événement – une distance ***aussi grande que*** 1075 – sous l'hypothèse H_0 d'indépendance ***est très faible***. En fait, on va voir que cette probabilité est ***quasi nulle***.

➔ Comment le savoir ? On peut aujourd'hui utiliser un calculateur en ligne comme celui utilisé ci-après (http://www.physics.csbsju.edu/stats/contingency_NROW_NCOLUMN_form.html). Voici le premier écran que l'on rencontre.



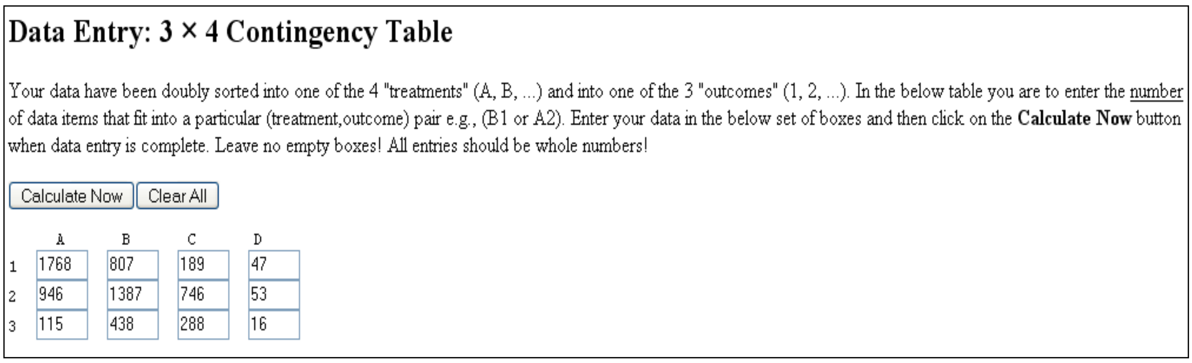
r×*c* Contingency Table: How many rows? columns?

You are about to enter your data for a chi-square contingency table analysis. For this to make sense you should have a table of data (at least 2x2; maximum: 9x9).
Number of rows:
Number of columns:

You must have data for each box in the table: no blanks allowed!

Submit

Le nombre de lignes (*rows*) est ici 3, le nombre de colonnes (*columns*) est 4. En cliquant sur *Submit*, on voit apparaître un écran comportant une table 3×4 où l'on saisit alors les effectifs empiriques observés.



Data Entry: 3 × 4 Contingency Table

Your data have been doubly sorted into one of the 4 "treatments" (A, B, ...) and into one of the 3 "outcomes" (1, 2, ...). In the below table you are to enter the number of data items that fit into a particular (treatment,outcome) pair e.g., (B1 or A2). Enter your data in the below set of boxes and then click on the **Calculate Now** button when data entry is complete. Leave no empty boxes! All entries should be whole numbers!

Calculate Now | Clear All

	A	B	C	D
1	1768	807	189	47
2	946	1387	746	53
3	115	438	288	16

En cliquant sur *Calculate Now*, on obtient alors ceci.

r × *c* Contingency Table: Results

The results of a contingency table X^2 statistical test performed at 06:34 on 21-MAR-2008

data: contingency table

	A	B	C	D	
1	1768	807	189	47	2811
2	946	1387	746	53	3132
3	115	438	288	16	857
	2829	2632	1223	116	6800

expected: contingency table

	A	B	C	D
1	1.169E+03	1.088E+03	506.	48.0
2	1.303E+03	1.212E+03	563.	53.4
3	357.	332.	154.	14.6

chi-square = 0.107E+04
degrees of freedom = 6
probability = 0.000

Le logiciel a calculé : 1. les effectifs ***marginaux***, 2. les effectifs ***théoriques***, 3. la valeur de χ^2, 4. le nombre de ***degrés de liberté*** (*degrees of freedom* : ce nombre vaut ici (3 – 1)(4 – 1) = 6), 5. La probabilité de l'événement $\chi^2 \geq 0{,}107 \cdot 10^4$, qu'il trouve égale à… 0,000. Dans cette situation, on ne peut guère que ***rejeter*** l'hypothèse d'indépendance entre les caractères *X* et *Y*. On est ici, en vérité, dans un cas qui illustre cette remarque de l'auteur des *Méthodes d'analyse d'enquêtes* : « On se souviendra que le khi-deux étant sensible aux effectifs, dès qu'une population d'enquêtés devient importante, il devient rare que le khi-deux d'un tableau croisé ne soit pas significatif. »

➔ De fait, à l'aide d'un autre calculateur en ligne, on obtient par exemple ceci (http://www.fourmilab.ch/rpkp/experiments/analysis/chiCalc.html).

Calculate X^2 from probability *Q* and *d*

To determine the chi-square value indicating a probability *Q* of non-chance occurrence for an experiment with *d* degrees of freedom, enter *Q* and *d* in the boxes below and press **Calculate**.

Given probability *Q*= 0.01 and *d*= 6

Calculate

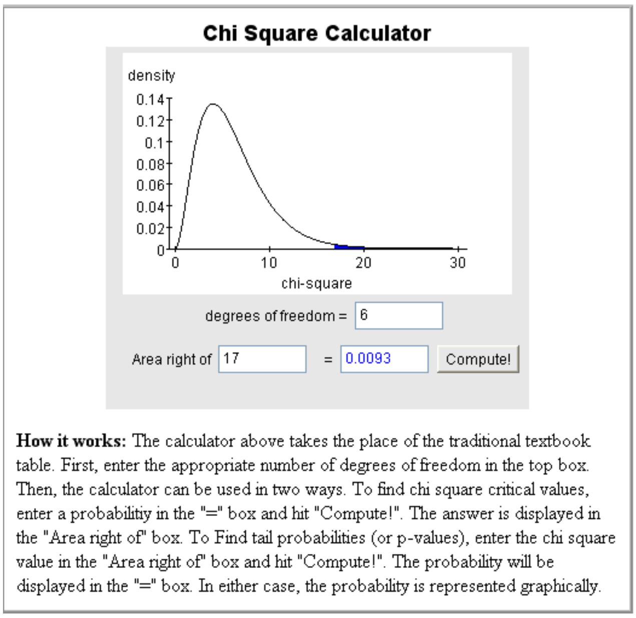
The X^2 value is: 16.8118

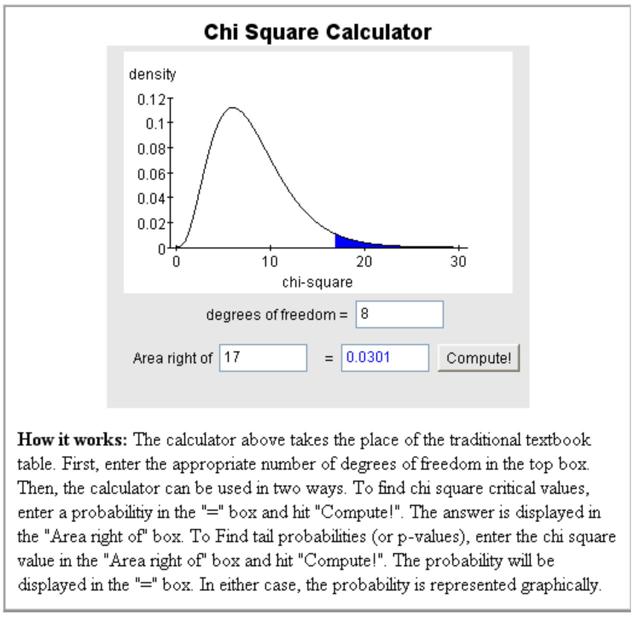
Autrement dit, pour que la différence entre les deux tableaux – théorique et observé – soit significative au seuil de 1 %, il suffit que χ^2 atteigne ou dépasse la valeur 16,8118. Voici un tableau de valeurs que l'on obtient avec le même calculateur (pour 6 degrés de liberté).

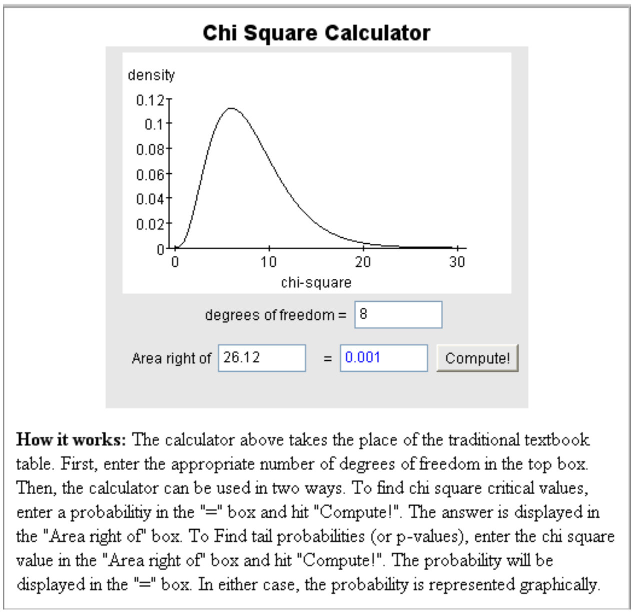
Prob. ≤ …	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
$\chi^2 \geq$ …	16,8118	22.4577	27.8563	33.1070	38.2583	43.3292	48.1824	51.6164

Observer un χ^2 de plus de 52 est, sous l'hypothèse d'indépendance de X et Y, un événement dont la probabilité est déjà inférieure à ***un milliardième*** ; observer un χ^2 de plus de 1070 a donc une probabilité infinitésimale !

➔ On pourra mieux comprendre le type de phénomènes numériques auquel on a affaire en s'en tenant une interprétation graphique. En sollicitant le logiciel que l'on trouvera en ligne à l'adresse http://www.stat.sc.edu/~west/applets/chisqdemo.html, on obtient par exemple ce qui suit.







f) Pour la séance prochaine, chacun établira un relevé des difficultés rencontrées dans la suite du document examiné.

2. Comment « observer » ?

a) On a souligné le rôle cardinal joué par l'observation de « réponses » $R^{\Diamond}$, c'est-à-dire par l'observation de (fragments de) praxéologies soit institutionnelles, soit personnelles. Mais comment observer des praxéologies ? Le travail accompli jusqu'ici permet d'apporter à cette question de premiers éléments de réponse.

➔ Nous nous sommes interrogés d'abord sur les praxéologies mises en œuvre par des personnes fréquentant une certaine institution – le site de l'association Pénombre – à propos de pourcentages : en ce cas, nous nous sommes efforcés d'observer des praxéologies mobilisées par des personnes qui ***« font quelque chose »***, dans une situation ***que nous n'avons pas suscitée***.

➔ L'examen des « œuvres » $O^{\%}_{\mathrm{Wf}}$ et $O^{\%}_{\mathrm{We}}$ a mis à profit un type de situations – également non suscitées par nous – où une ***intention didactique*** qui les porte et qu'elles portent conduit des personnes (ici, anonymes) à émettre un discours (écrit) présentant et explicitant (dans une mesure variable) des praxéologies. À cette observation, on peut joindre celle, très partielle encore, du « discours didactique » tenu par l'auteur – non anonyme, lui – de l'ouvrage *Les méthodes d'analyse d'enquêtes*.

b) Pour recueillir du matériel praxéologique, il est un autre type de techniques qui paraît fort prisé dans les sciences humaines et sociales et qui consiste pour l'enquêteur à procéder, avec leur accord, à un ***interrogatoire*** des personnes dont on étudie les praxéologies, par exemple par ***entretien*** ou par ***questionnaire***, et cela à propos des questions que l'on se pose à leur propos. Il convient de ***bien distinguer*** ce genre de situations – que l'on suscite – d'un autre type de situations, que l'on suscite également mais dans lequel le chercheur, au lieu d'***interroger*** des sujets, leur demande de ***faire quelque chose***.

➔ Ce « quelque chose à faire » peut consister à écrire un bref commentaire sur tel passage d’un article de journal qui dirait que « ce qui se passe depuis un mois dans la rubrique des transferts donne le vertige. En quatre ans en France, le montant des transactions a quadruplé : 400 % d’inflation, ce n’est pas mal ! »

➔ Notons que, « faire faire quelque chose », c’est exactement ce que fait le médecin qui, après avoir procédé à l’***interrogatoire*** de son patient, procède à un ***examen clinique*** et, dans ce cadre, lui demande de tousser, de respirer fort, etc.

➔ Soulignons une situation « limite » entre « interrogatoire » et « observation clinique » : le « quelque chose à faire » peut consister à répondre à des questions de l’enquêteur, mais à des questions ***qui ne sont pas celles que l’on se pose au sujet de l’enquêté***, quoiqu’elles aient pour objectif de rendre sensible (audible, visible, etc.) des éléments de réponse à ces questions, cela sans forcément que le sujet en ait pleine conscience.

➔ Notons que faire ***faire***, y compris faire ***dire***, certaines choses est la stratégie traditionnelle de l’examinateur face au candidat, qui ne demande pas au candidat « Savez-vous calculer le pourcentage d’augmentation d’un avoir placé au taux de 2,6 % pendant 4 ans ? » mais lui demande de « calculer le pourcentage d’augmentation d’un avoir placé au taux de 2,6 % pendant 4 ans ».

c) Nous ferons sans plus attendre l’expérience du type de situations où l’enquêteur (*y*) demande aux enquêtés (*X*) de « faire quelque chose », qui consistera en l’espèce à attribuer un score (de 0 à 6) à un mot (que l’on va préciser) sur chacune des échelles d’un ensemble d’échelles donné, et cela même (et surtout) quand l’échelle proposée peut sembler « sans rapport » avec le mot.

➔ Voici l’une des dix échelles proposées où un score a été porté.

Passif O———O———O———O———O———O———O Actif

Ici, le sujet a coché le score 5 : le mot qui lui a été proposé sonne pour lui comme quelque chose de fortement actif (le maximum est 6).

➔ Voici successivement les mots à « noter » sur les 10 échelles retenues : Université / Statistique / Éducation / Internet / Mémoire.

➔ Les données collectées seront examinées lors de la prochaine séance.