

L'ANALYSE DES DONNEES EPIDEMIOLOGIQUES : QUELQUES CONSEILS PRATIQUES

J.J. BENET*



=====

RESUME : L'analyse des données épidémiologiques nécessite le respect d'une méthode simple, bien qu'elle soit très générale : définition des objectifs, formulation des hypothèses, choix d'une stratégie, détermination des traitements (transformations de variables, statistiques), jeu d'essai, puis analyse proprement dite. Cette chaîne doit être pensée dès l'élaboration du protocole de récolte de données épidémiologiques, à moins qu'elle ne soit appliquée à des données existantes, dont l'étude critique préalable est indispensable. La disposition de logiciels d'analyse de données sur micro-ordinateurs est un atout important, mais inférieur toutefois à l'élaboration d'une bonne procédure de traitement.

SUMMARY : Epidemiological data analysis needs the application of a simple, albeit very general, method. Aims definition, hypothesis proposition, choice of a strategy, treatments propositions (variable transformations, statistics...), tries, and then the real analysis. This whole process must be drawn as soon as the plan of epidemiological data collection is thought, unless the data are already existing. In this last case the critical analysis of data is essential. Existence of data processing programs running on microcomputers is important, but yet not as much as a good thinking of the treatment process.

* * *

Le propos n'est pas de traiter de l'art et la manière d'analyser des données (épidémiologiques) : ce serait l'objet d'un développement bien plus complet, par de véritables spécialistes de la question.

L'objectif est ici bien plus modeste : il s'agit seulement de souligner les principes de base fondamentaux et quelques pièges grossiers que tout utilisateur non spécialiste doit connaître, sous peine d'en faire l'apprentissage à ses dépens.

Ainsi, peut-on l'espérer, la mésaventure suivante sera évitée...

* Ecole nationale vétérinaire d'Alfort - Chaire de maladies contagieuses - 94704 Maisons-Alfort cedex.

PLUS JAMAIS CA !

Une personne s'étant livrée à la collecte de données épidémiologiques, sollicite quelqu'un de sa connaissance ayant la double réputation d'être équipé d'un ordinateur (doté de logiciels de gestion de données y compris statistiques, ce qui n'est pas si commun) et d'être serviable.

"Nous avons accumulé des données dans des dossiers d'enquête. Nous en avons actuellement ... (suit une précision sur la quantité estimée soit en unité de poids, soit en unité de longueur, pour ne pas dire de hauteur).

Nous serions très contents et très reconnaissants si vous acceptiez d'en faire l'analyse."

Cette évocation est tout à fait caricaturale. Elle est même improbable. Pourtant elle est conforme à certaines situations vécues.

Le problème n'est pas celui de la disponibilité éventuelle de celui qui est sollicité pour faire cette analyse (comme s'il attendait impatiemment, pour occuper ses journées encore insuffisamment remplies, qu'on lui confie un tel travail !), mais de savoir pourquoi de tels spécialistes de l'analyse des données à qui on pose ce genre de questions réagissent invariablement par un ensemble de manifestations physiologiques caractéristiques du choc adrénalinique.

Pourquoi ?

"Il fallait venir me voir AVANT" répondent-ils invariablement.

Comme si on pouvait penser à tout dès le départ !...

Et pourtant !

Nous allons voir pourquoi effectivement le problème de l'analyse des données doit être posé le plus tôt possible.

Alors, commençons.

"POSONS LE PROBLEME"

"Je dispose d'une masse de données que je voudrais bien ANALYSER".

◊ "ANALYSER", qu'est-ce que cela signifie ?

Sûrement pas faire n'importe quel test dans n'importe quels sens, de façon à aboutir à une conclusion rassurante : "significatif" (si possible à la "n"ème décimale).

Ces tests sont des outils. Ils ont une indication, un usage bien précis.

Par exemple, on ne saurait se servir d'un ciseau à bois comme d'un tournevis. C'est donc que je sais d'avance à quel traitement je destine les données collectées.

C'est l'évidence. Mais il n'est pas inutile de le rappeler.

Le problème peut donc être reformulé de la façon suivante :

"Je vais appliquer des tests statistiques appropriés à des données collectées afin de les analyser".

◊ QUELS TESTS STATISTIQUES ?

Ils seront exposés par J.C. Fayet, et ne seront pas abordés ici. Sachons seulement qu'ils se résument à peu de choses (même si les modalités sont nombreuses).

- Tests qualitatifs : comparer des distributions (nombre de sujets selon des critères qualitatifs : race, lots...).
- Tests quantitatifs : comparer les valeurs des paramètres de position (moyenne) et de dispersion (écart-type ou variance).

En quoi les tests statistiques conditionnent-ils les données (avant leur collecte) ?

Prenons l'exemple de l'âge des animaux. Il peut être noté de diverses façons :

- Qualitatif

Veau - génisse - génisse pleine - mère - hors d'âge -

- Quantitatif

- . Exact : nombre d'années, de mois - Variable selon le moment de l'enquête, donc introduit une erreur.
- . Précis : date de naissance (qui par calcul donne l'âge).
- . Approximatif : pour les jeunes, selon le format, les cornes, pour les adultes selon le nombre de veaux.

- Semi-quantitatif

Par tranche d'âge : 0 à 3 mois - 3 mois à 1 an - 1 an à 2 ans - 2 à 5 ans - plus de 5 ans.

Cette présentation conduit à une évidence, il est impossible de calculer une moyenne avec la première façon de collecter l'information. Donc il faut choisir judicieusement la modalité de collecter, en fonction du test statistique qui sera appliqué.

Il serait apparemment plus simple d'enregistrer d'emblée les données sous une forme quantitative (puisque l'on peut toujours les transformer en données qualitatives). Les contraintes du terrain ne se satisfont pas toujours de tels "y'à qu'à".

◊ QUELLES HYPOTHESES ?

En effet, selon la nature de la donnée, il est possible (ou non) d'étudier la relation, par exemple entre l'âge et un phénomène étudié, soit pour des variables continues avec un test de régression, soit pour des variables discontinues, avec un χ^2 .

Le choix des outils statistiques répond directement à la formulation des HYPOTHESES, qu'il s'agit de mettre à l'épreuve.

L'hypothèse est avancée d'une relation entre l'âge et le phénomène ; elle est même formulée de façon plus précise par le pari du type de relation qui peut exister (croissante, linéaire...).

Selon le contenu de l'hypothèse, il est évident que le protocole est tout à fait différent, et conditionne la récolte des données.

- S'il s'agit d'un protocole à objectif descriptif (évaluation de la fréquence d'un événement), il est nécessaire de faire au départ, l'hypothèse d'un taux ; c'est le seul moyen de déterminer le nombre de sujets de l'enquête nécessaires à une estimation valable.

- S'il s'agit d'un protocole analytique, de la même façon, pour pouvoir comparer les populations on doit faire l'hypothèse de la différence attendue, qui permet là encore de déterminer le nombre de sujets nécessaires pour pouvoir conclure avec une puissance suffisante.

D'autres éléments interviennent dans la détermination des épreuves statistiques.

◊ QUELS CHOIX STRATEGIQUES ?

L'objectif, les hypothèses, sont formulés. Mais, pour parvenir à ces fins, différents choix sont possibles, qui conditionnent les données.

Exemple : Je veux comparer l'efficacité de deux substances, dont l'effet est mesurable. Par des essais préliminaires, j'ai déterminé un ordre d'idée de la différence attendue. La comparaison à envisager peut être conçue de deux façons : soit entre deux populations soumises chacune à l'une ou l'autre des substances, soit sur une population, dont chaque sujet sert de témoin (en appliquant les deux substances sur le même animal en même temps ou de façon différée). Dans un cas concret dont le détail n'est pas donné ici, il a été possible d'estimer le nombre de sujets nécessaires à 2.500 pour le premier choix stratégique et seulement 60 pour le deuxième (séries appariées).

◊ CONCLUSION

Nous venons de vérifier rapidement que la conception du protocole de récolte des données contient par avance (et par nécessité) l'orientation des modalités de l'analyse des données qui sera ultérieurement entreprise.

On comprend le souci du spécialiste de l'analyse de données d'être consulté avant la phase de récolte des données, et cela dès les prémices du travail, c'est-à-dire dès la formulation des hypothèses.

Que dire de ceux qui ramassent des données, sans avoir clairement formulé les hypothèses, absorbés qu'ils sont par les seuls aspects tactiques (le questionnaire), et ayant complètement négligé la stratégie. On peut les comparer à des chiffonniers, glanant quelque probande sur un tas d'ordure : ne nous étonnons pas si à l'arrivée, les résultats ne sont pas enthousiasmants. Selon l'adage anglais : "garbage in, garbage out".*

* *

Toutefois, on n'a pas toujours la possibilité de monter spécialement une opération de récolte de données. D'ailleurs, de plus en plus, les épidémiologistes utiliseront des données existantes : abattoir, laboratoires de diagnostic...

Les beaux principes précédents n'auront donc pas pu être mis en application.

Est-ce grave ?

Comment déjouer les périls éventuels ?

* *

UTILISATION DE DONNEES EXISTANTES

Prenons un exemple banal, celui de l'heureux possesseur d'un stock d'informations (sous forme de fichiers), qu'il soit D.S.V., vétérinaire praticien, directeur de laboratoire, directeur d'un groupement... Ces fichiers existent pour une tâche bien précise pour laquelle ils ont été conçus, le plus souvent de gestion. De fait, de nombreuses recherches épidémiologiques peuvent résulter d'activités de gestion, dont elles sont un sous-produit, moyennant un minimum de précautions.

La tentation est grande de partir des seules données existantes, de les "triturer" pour "voir ce que l'on pourrait bien en sortir".

Les commentaires développés en première partie ont suffisamment montré qu'il n'était pas possible de procéder ainsi, à l'aveuglette.

Il faut donc construire un protocole d'analyse de données, dont chaque étape sera systématiquement confrontée aux contraintes résultant de l'existant, pour autoriser le passage à l'étape suivante, ou déterminer les conditions d'adaptation.

Pour cela, il faut tout d'abord élaborer les grandes lignes du projet, puis faire la description et l'analyse de l'existant, avant de bâtir le protocole d'analyse de données.

* Des ordures dedans, des ordures en sortie.

Ce que nous ferons au travers d'un exemple concret.

A. PRESENTATION DU PROJET D'ANALYSE DE DONNEES (Exemple)

Objectif. Exploiter les données collectées dans les fichiers sanitaires de la D.S.V. pour approfondir l'étude descriptive (de la tuberculose bovine, par exemple à l'échelon du département.

Hypothèse. La connaissance épidémiologique résultant des statistiques globales apparaissant dans les compte-rendus annuels n'est pas assez détaillée, voire non suffisamment conforme à la réalité.

Stratégie. Le cas de chaque exploitation sera étudié. Les données disponibles (résultats de tuberculination et d'abattoir) seront étudiées, de manière à dégager les typologies (différents cas de figures rencontrés dans les manifestations épidémiologiques) les caractéristiques d'évolution dans le temps, et de répartition dans l'espace.

B. DESCRIPTION DE L'EXISTANT

C'est un véritable dossier d'analyse qu'il faut constituer qui permette de reconstituer toutes les étapes, tous les éléments de cette chaîne de traitement de l'information qui aboutit aux données du fichier. Cette connaissance doit être critique, de manière à apprécier la valeur de l'information.

Nous n'entrerons pas dans le détail de l'exemple évoqué. Nous ne ferons que souligner les éventuelles conséquences de certaines caractéristiques de l'existant sur la construction du projet.

◊ Objectif

Les données sont collectées pour répondre à un objectif. Parmi les différentes formulations possibles (car la réalité est complexe), retenons la composante de "contrôle", dans un sens administratif, voire policier (... police sanitaire).

Cet aspect (fragmentaire) peut conduire à des dissimulations, et donc des sous-estimations.

Inversement, dans un objectif de "sanction" (élimination, abattage des animaux infectés), l'indemnisation peut avoir un effet de sur-estimation (abattage totaux).

Les données sont collectées avec plusieurs objectifs, notablement différents, qui peuvent profondément interférer avec le projet.

◊ Hypothèses

Ce sont les hypothèses qui précèdent la saisie des données, qu'elles soient élaborées lors de la conception de la chaîne de traitement, ou qu'elles soient formulées pendant le déroulement de la chaîne.

- Au stade de la conception de la chaîne existante

Par exemple, la réglementation considère que toute exploitation dans laquelle des réactions tuberculiques non négatives sont constatées est réputée infectée de tuberculose.

Ce qui implique une série de contraintes pour l'éleveur, source de sous-estimation, si le vétérinaire choisit de mettre en place un protocole particulier d'observation dans cette exploitation de manière à préciser son diagnostic, sans forcément transmettre immédiatement l'information à la D.S.V.

- Au stade de la réalisation

L'interprétation de la réalité constitue une hypothèse, aussi longtemps que l'observation n'est pas suffisamment confirmée.

Exemple : L'observation "réactions tuberculiniques positives" peut être interprétée "cheptel infecté de tuberculose". Un doute peut toutefois exister en fonction des conditions d'observation. Selon les cas, la constatation de lésions tuberculeuses sur les animaux abattus confirme cette hypothèse, ou inversement, l'absence (sur tous les animaux abattus) la remet en question.

Exemple : Partant de l'a priori que tout animal éliminé pour motif de tuberculose (marqué du T) est effectivement tuberculeux, certains sont amenés à déclarer la constatation de lésions tuberculeuses à l'abattoir, par souci d'efficacité (prudence à l'égard du consommateur, impact psychologique vis-à-vis de l'éleveur, pour qui l'élimination était ainsi justifiée), mais éventuellement aux dépens de la réalité (= même en l'absence de lésions visibles).

On voit donc que les hypothèses peuvent fortement conditionner la validité des données.

◊ Les conditions de réalisation

Le déroulement des étapes de circulation de l'information peut aussi interférer avec le projet.

Ainsi, le devenir de tous les animaux d'une exploitation n'est pas toujours connu. La perte de l'identification individuelle, la succession des intermédiaires commerçants peuvent empêcher de répercuter l'information concernant l'animal jusqu'à l'exploitation.

◊ Les données disponibles

Les aspects précédents concernent la validité des données; il faut aussi, bien sûr, faire une description des données.

Sur un plan qualitatif, quelles sont les différentes variables existant :

- . identifiant : nom, adresse...
- . effectif
- . observations pathologiques
- . dates
- . etc.

Sous quelle forme sont notées ces informations :

Exemple : Nombre de réagissants sur effectif total, ou sur nombre de non réagissants.

Sur un plan quantitatif, quel est le volume d'information à traiter au total par exploitation.

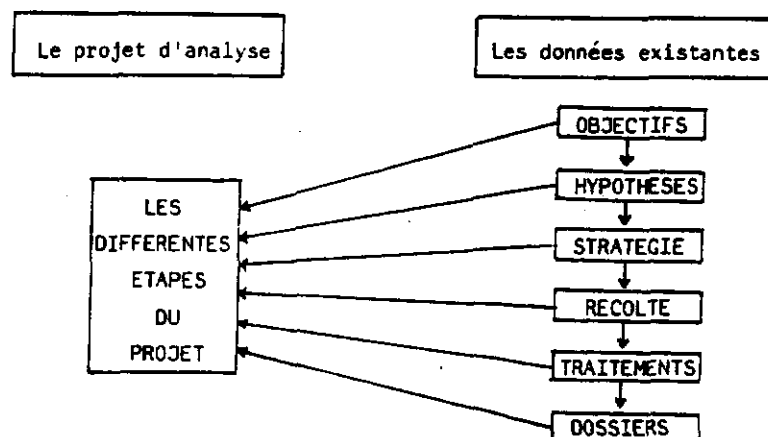
Pour chaque variable, il faut en approfondir la connaissance, sans devoir pour autant y consacrer une partie trop importante du temps de l'étude. Il s'agit d'appréhender les principaux paramètres, même de façon approximative : les écarts (plus petit, plus grand), le score le plus fréquent (approximation du mode), le score moyen (un ordre d'idée de la moyenne), ou bien les effectifs de fréquence par classe (en valeur absolue cumulée, éventuellement en regroupant des classes). Toute présentation pouvant aider cette perception est utile (tableaux, histogrammes de fréquence, voire simplement emploi de stylos de couleurs différentes sur les tableaux de données).

* * *

Ces quelques exemples illustrent la nécessité de connaître précisément les données existantes, leur validité, et la répercussion de leur valeur sur l'ensemble du projet (figure 1). Au point, qu'il peut être salutaire de se poser la question de l'opportunité d'entreprendre une telle analyse, dont les résultats pourraient donner une image éloignée de la réalité...

Figure 1 : Interaction des données existantes sur l'analyse des données.

Dans le cas de l'utilisation de données existantes (collectées dans un fichier par exemple), dans un objectif qui n'avait pas été envisagé initialement, il importe d'analyser chacune des étapes de la procédure de collecte de ces données existantes (tant en ce qui concerne la conception que la réalisation) car chacune d'entre elles est susceptible d'avoir un retentissement sur l'une ou l'autre des étapes du projet d'analyse de ces données.



* * *

CONCEPTION ET MISE EN OEUVRE D'UN PROTOCOLE D'ANALYSE (figure 2)

◊ Objectif

Certes, l'objectif de l'analyse a déjà été formulé. Mais, l'étude du dossier peut conduire à en modifier le contenu... et cela même alors que le travail est profondément engagé. On doit toujours conserver à l'esprit que les objectifs initiaux peuvent être remis en question, et, si nécessaire, les amender consciemment, plutôt que de subir involontairement une "dérive" résultant de leur modification inconsciente.

Figure 2 : Conception et mise en oeuvre d'un protocole d'analyse de données.

Un travail d'analyse de données se décompose en deux étapes, conception, réalisation, qui doivent impérativement être distinguées.

La conception passe par une étape initiale de récolte d'informations (1) (de terrain, de bibliographie), qui permettra de définir les caractéristiques des étapes suivantes.

La définition des objectifs (2) constitue à la fois le point de départ du projet, et le point d'arrivée (8).

La troisième étape (3) consiste en une (des possibles...) formulation sous une forme scientifique des questions (d'intérêt scientifique) qui doivent permettre d'atteindre l'objectif, ainsi que des hypothèses nécessaires (soit que l'on veuille les vérifier, soit qu'elles soient indispensables à la poursuite du travail).

Ex. : Objectif = connaissance des taux d'infection... une des questions = quels taux de prévalence des cheptels, des animaux infectés, des animaux infectés par cheptels infectés.

Hypothèse = on pense que les taux espérés sont, respectivement, de l'ordre de 10 %, 1 %, 20 %.

La stratégie (4) est la façon dont on compte s'y prendre pour répondre à ces questions : c'est le choix entre différentes options possibles.

Ex. : Soit traitement exhaustif, soit échantillonnage, selon certaines règles (sondage sur 1/10, ou bien on ne retient que les élevages de naisseurs...).

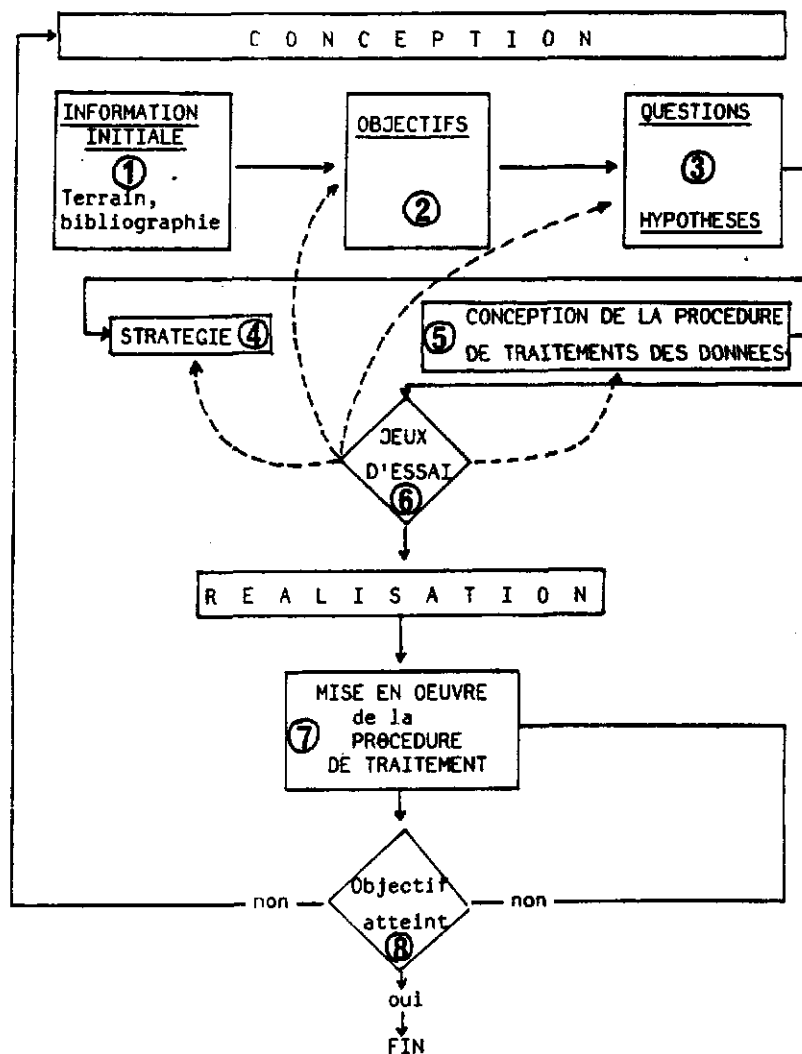
La conception de la procédure de traitement (5) est plus directement liée aux données, c'est pourquoi elle est souvent privilégiée aux dépens des étapes précédentes. Elle est expliquée plus en détail dans la figure 3.

Les jeux d'essai (6) constituent l'étape décisive de la conception, puisqu'elle permet de prendre connaissance d'erreurs, ou d'insuffisances relatives à la conception... Certaines ne pourront apparaître qu'à l'issue de la réalisation (8), mais il vaut mieux s'en apercevoir le plus tôt possible. D'autre part, cette épreuve permet d'appréhender les délais, les difficultés éventuelles.

Cette étape est capitale, car elle seule permet la remise en question des étapes précédentes, sans trop de dommage. Elle est indispensable, car elle permet progressivement de les redéfinir, ce qui est schématiquement représenté par les flèches, illustrant le "feed-back". C'est donc une procédure par approximations successives, qui est réaliste, et non celle d'un déroulement idéalement linéaire.

La réalisation succède à la conception... c'est-à-dire que s'il est nécessaire à ce stade de modifier la conception, c'est que celle-ci n'a pas respecté l'étape des jeux d'essai. Elle est détaillée dans la figure 3, en ce qui concerne la procédure de traitement (7).

On vérifie en final si les objectifs sont atteints (8) : d'où la nécessité qu'ils aient été clairement définis, afin d'éviter une inadéquation, ou une éventuelle dérive résultant d'une reformulation (inconsciente) des objectifs au fur et à mesure du dépouillement des résultats. Dans la négative, il peut être nécessaire d'affiner l'analyse des résultats, ou pire, de réviser la conception de la procédure.



◇ Hypothèse

Le dossier des hypothèses s'est sûrement enrichi de l'analyse de l'existant. Au point que celle-ci est une méthode d'aide à leur élaboration.

Exemple : On en vient par exemple à soupçonner le rôle d'un facteur "abattoir".

◊ Stratégie

L'étude préalable de l'existant peut conduire à envisager de modifier la stratégie initialement prévue, en la complétant par exemple par quelques études approfondies ponctuelles, ou la mise en place d'un protocole spécial pour résoudre certaines questions particulières.

◊ Comment faire ? La procédure de traitement (figure 3)

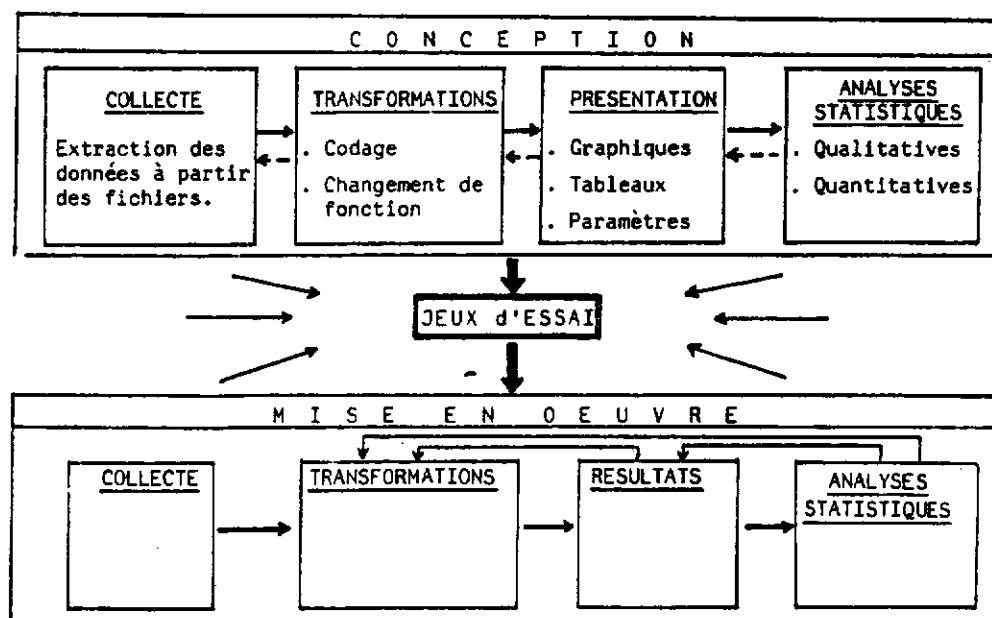
Figure 3 : La procédure de traitement des données.

La chaîne de traitement de données peut être représentée en 4 maillons (collecte, transformations des données, présentation, analyses statistiques), identiques dans les deux étapes distinctes de conception et de réalisation, séparées par l'épreuve des jeux d'essai.

Au stade de la conception, tous les maillons sont indépendants puisque c'est en fonction de l'une des étapes suivantes que seront effectuées les opérations d'une étape donnée. Cette contrainte est d'autant plus évidente que l'on utilise des outils comportant des exigences particulières dans la forme des données par exemple, et spécialement pour les systèmes informatisés.

C'est pourquoi la réalisation de jeux d'essais (soit avec des données fictives, ou mieux avec quelques échantillons de données réelles) permet de mieux articuler cet ensemble d'étapes, représentées ici de façon linéaire, mais qui en réalité est plutôt un réseau aux interactions complexes.

Dans ces conditions, la mise en oeuvre de la chaîne permet de produire des résultats avec un maximum d'efficacité. Cela n'exclut pas pour autant à ce stade une réorganisation des données en fonction des nécessités de traitements dont l'opportunité apparaît au fur et à mesure du dépouillement des résultats... Ce qui suppose que dès la conception cette organisation des données ait été prévue.



Supposons résolus les problèmes de collecte des données à partir des fichiers (bien que ses modalités conditionnent cette étape), car son étude nous entrainerait trop loin : la méthode est celle de la conception d'un questionnaire, d'un bordereau de saisie, et de la définition d'un éventuel plan d'échantillonnage.

Nous disposons (enfin) des données brutes, il est douteux que l'on puisse les utiliser telles quelles pour vérifier les hypothèses.

Il faut tout d'abord transformer les données.

- Transformation

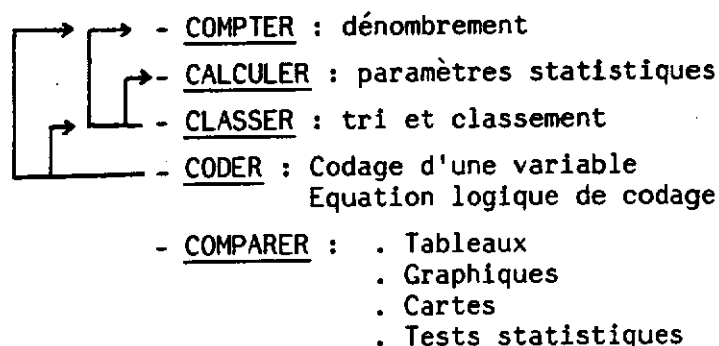
Il peut s'agir d'un codage (une commune = un numéro, par exemple), ou d'une opération élaborée.

Exemple : On dispose de résultats de tuberculination et d'abat-toir à différentes dates. Les données brutes n'ont de valeur qu'en fonction de l'interprétation qu'on va leur donner (qui découle elle-même des hypothèses formulées initialement). A moins de disposer d'un ordinateur doté de précieuses fonctions de codage du type "si telle variable a telle valeur et telle autre variable à telle valeur (et ainsi de suite), alors telle dernière variable a telle valeur", on doit remplacer l'informa-tique insuffisante à la force de ses neurones... ce qui peut être une importante source d'erreurs pour des dossiers importants et complexes.

- Définition de la procédure d'analyse

L'analyse utilise un nombre limité d'opérations (figure 4).

Figure 4 : Les opérations du traitement de données.



. COMPTER

Les dénombrements de catégories fournissent la matière aux tests qualitatifs.

. CALCULER

Le calcul des paramètres statistiques usuels donne moyenne, écart type.

- CLASSER

Les opérations de tri et de classement précèdent et enrichissent les fonctions précédentes.

- CODER

C'est une opération capitale dans les procédures automatiques.

- COMPARER

Pour cela, la présentation des données est indispensable (tableaux, graphiques, cartes). La comparaison est validée par des épreuves statistiques.

La définition de la procédure d'analyse consiste à déterminer le moyen de répondre à une question précise que l'on se pose.

En pratique, cela revient à préparer les tableaux d'analyse (tableaux de contingence), les graphiques qu'il sera intéressant de réaliser, les paramètres statistiques qu'il faudra calculer, les cartes...

Comme pour de nombreuses étapes, celle-ci est élaborée en interaction avec l'étude elle-même, et donc soumise à amélioration continuelle.

- Jeu d'essai

Il serait dangereux de ne pas faire un essai préalable, pour vérifier la faisabilité du projet, sa pertinence par rapport aux questions posées.

- Passage en vraie grandeur

Comme toutes les tâches importantes, et complexes, celle-ci nécessite de se prémunir contre le risque d'erreurs : la rigueur est un des remèdes (méthode de travail standardisée, enregistrement systématique de tous les événements survenus dans le travail dans un journal de bord), les contrôles en sont un autre (vérification systématique par une même personne un jour différent, selon un ordre différent, ou par une autre personne).

◊ Ce qu'il ne faut pas faire

Parmi la multitude des erreurs à ne pas commettre, un certain nombre ont déjà été soulignées dans cet exposé.

Ajoutons quelques points, dont deux anticipent sur l'exposé de J.C. Fayet.

- Ne pas respecter les conditions de validité des tests

Certains tests statistiques ont des exigences, comme la taille minimale des effectifs, ou la distribution des valeurs selon une loi gaussienne. Pour contourner cet obstacle, on regroupe des effectifs, ou l'on transforme les variables.

- Manquer de puissance statistique

L'interprétation d'un résultat "non significatif" doit être mûrement réfléchie, afin de ne pas conclure (à tort) que s'il n'y a pas de différence, c'est que "c'est pareil", alors que l'on manque simplement de puissance pour révéler la différence.

- Remettre en question la conception même du travail

Finir l'analyse, et aboutir à la conclusion qu'il aurait mieux valu collecter telle donnée autrement est non seulement déprimant, mais sans doute significatif d'une faille dans les étapes d'analyse de l'existant, d'élaboration du protocole, ou de réalisation des jeux d'essai.

- Etre persuadé que le recours à l'ordinateur est un garant de bonne fin des travaux

Seul le respect des différentes étapes est le véritable garant.

A défaut de posséder un bon progiciel, l'ordinateur peut être un frein : dans ce cas, il peut être préférable d'envisager un traitement manuel, pourvu que la procédure soit judicieuse ; pour un certain volume de données (à apprécier en temps de travail), cette solution peut être tout compte fait satisfaisante, en comparaison du temps nécessaire à l'élaboration d'un logiciel spécifique (sa mise au point...) puis à la saisie, avant de permettre l'analyse.

- Prévoir la fin de l'étude pour fin décembre

Dites : fin mars, ou fin juin. Ce sera plus réaliste.

* *

CONCLUSION

La prétention de cet exposé était modeste : donner quelques conseils de méthode dans l'analyse des données, fruits de l'expérience d'un non spécialiste, s'adressant à d'autres non spécialistes.

Bien sûr, de super-ordinateurs, avec des superlogiciels, et une équipe spécialisée (informatique, statistiques...) font beaucoup mieux, plus vite, et sans doute par des voies différentes. Mais pour les problèmes bien plus modestes, mais nombreux, et quand même importants, il faut savoir faire avec ce que l'on a : c'est là que la méthode montre son efficacité, comme c'est le cas pour celle-ci.

*
* *