Système de pilotage de l'entreprise

Le pilotage de l'entreprise utilise des tableaux de bord, des indicateurs, produits selon des méthodes statistiques. Ces méthodes permettent la synthèse d'une information massive, protégée de l'interprétation par sa masse même, et qui ne peut être utilisable que si elle est résumée de façon intelligente.

Nous allons décrire ici les principales méthodes statistiques utilisées par les entreprises, et les difficultés que rencontre parfois leur utilisation.

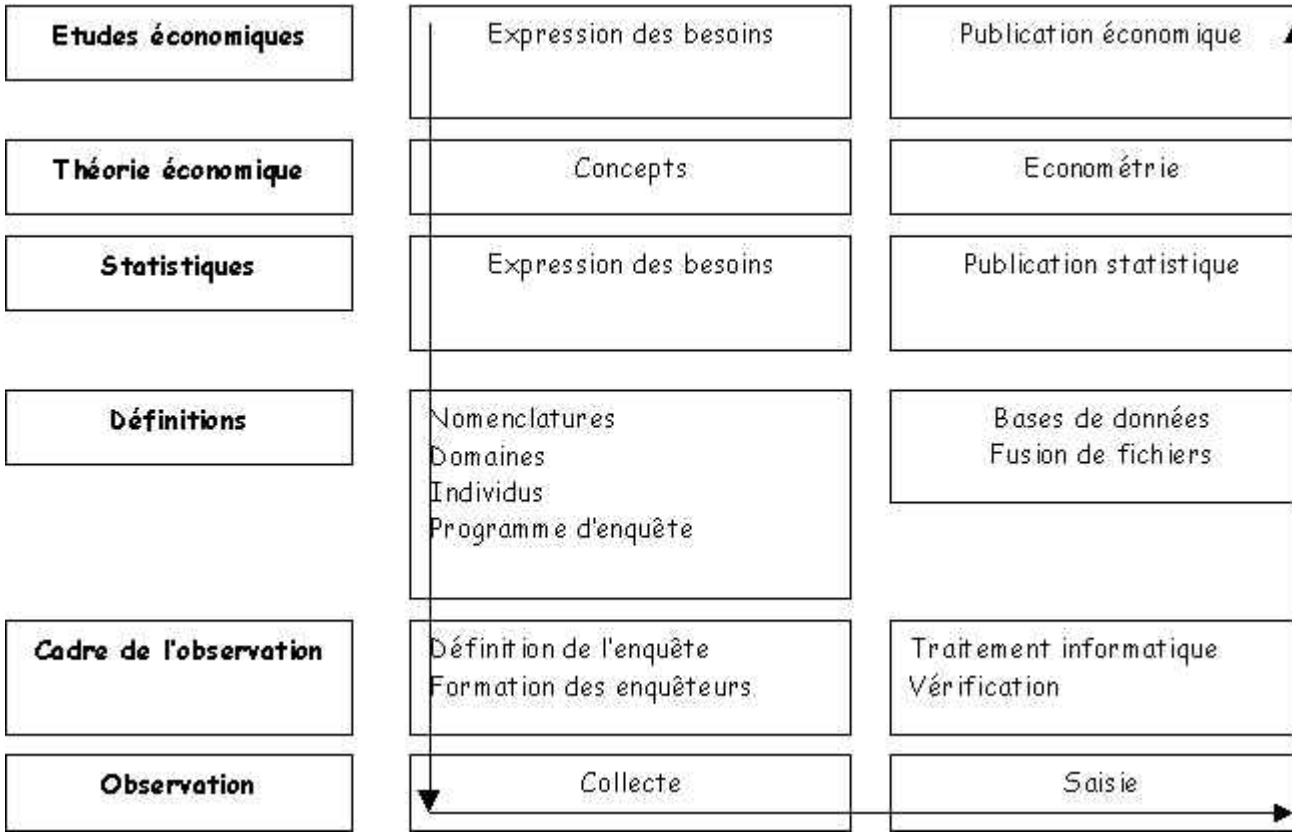
Le terme " statistique " désigne une méthode qui vise à observer puis décrire des populations. Le mot " population " s'entend ici de façon large : il peut s'agir de populations humaines (démographie), mais plus généralement la statistique considère des *ensembles* : sièges*kilomètres de transport de passagers, francs de chiffre d'affaires, tonnes d'acier, appels téléphoniques etc. Il importe, dans chaque cas, de définir la population que l'on considère et les " individus " qui la composent.

En tant qu'outil d'observation, la statistique est analogue à un microscope ou un télescope : elle ne donne un résultat intéressant que si elle est orientée vers un objet intéressant et si elle est utilisée par un opérateur capable d'interpréter l'observation. Il n'est donc pas inutile de savoir pour qui, et pour quoi, l'observation statistique est faite. Le critère de qualité est la *pertinence* - c'est-à-dire l'adéquation aux besoins qu'implique l'action. L'*exactitude* d'une information (c'est-à-dire son aptitude à alimenter un raisonnement exact) importe plus que sa *précision*.

Nota Bene : Le critère de pertinence se distingue du critère d'*objectivité*, que l'on invoque un peu à tort et à travers. Lorsque vous conduisez votre automobile, la présence du concept " signal lumineux " dans votre perception est pertinente, celle du concept " physionomie des passants " ne l'est pas ; il n'en est pas de même si vous flânez sur un trottoir... La présence du concept " signal lumineux " dans la perception du conducteur résulte d'un *choix* et se juge en termes de pertinence. Mais la *valeur* que révèle l'observation (" rouge ", " vert " ou " orange ") ne résulte pas d'un choix : elle est *objective*, dans le cadre du concept choisi. Ainsi l'objectivité n'est pas une " exacte représentation du monde réel ", chimère qui masque un sens commun pétri de préjugés, mais désigne une observation fidèle réalisée *après* un choix conceptuel.

En tant qu'outil de description, la statistique comporte un aspect *éditorial*. Il n'est pas aisé en effet de rendre compte de façon sobre, lisible, intéressante, d'un ensemble d'observations dont le résultat technique constitue une liasse de tableaux de nombres illisibles. La sélection des nombres, ratios, graphiques à publier est un art, ainsi que la rédaction des commentaires

La démarche statistique comporte ainsi des étapes relevant chacune d'une logique différente : définition des concepts puis du programme de l'observation, collecte et traitement de l'information, publication, modélisation et études. On peut les représenter selon un modèle en couches, chaque couche étant reliée à la suivante par une interface qui doit être définie proprement :



Modèle en couches de la statistique et de l'économie

Socle conceptuel

Il faut d'abord définir le *domaine d'observation* (population) et les *individus* qui le composent. Observe-t-on des personnes, des ménages, des entreprises, des établissements, des passagers, des clients ? On aura des déboires si l'on ne sait pas de quoi l'on parle: ainsi, croire que l'on observe des clients alors que l'on observe des passagers peut conduire à des conclusions erronées.

Nota Bene : quelle est, pour un transporteur aérien, la différence entre un passager et un client ? Le passager, c'est la personne qui fait un voyage, que l'on ne connaissait pas avant ce voyage et que l'on ne connaîtra plus après ce voyage. Le client, c'est un passager identifié : on est donc capable de regrouper sous son identifiant tous les voyages qu'il a faits, et d'analyser son comportement. La transition du raisonnement sur le passager au raisonnement sur le client permet de parler de comportement, de mesurer la " valeur " associée à un client (idéalement, cette valeur est égale à la valeur actuelle nette des relations avec ce client, c'est-à-dire à la valeur actualisée du chiffre d'affaires que l'on fera avec lui, diminuée de la valeur actualisée du coût de production des services qu'il consomme; la connaissance de cette valeur suppose que l'on sache prévoir sa consommation, évaluer les coûts des services qu'il consomme, prévoir l'évolution de ces coûts etc. Dans les faits, cette connaissance ne pourra être qu'approchée parce que les données complètes nécessaires pour la construire font défaut).

Pour passer de la connaissance du passager à celle du client, il faut avoir résolu le difficile problème de l'identification du passager.

Identifiants

Il faut pouvoir *identifier* les individus qui composent une population - c'est-à-dire attribuer à chacun un

code qui lui est propre. Il n'est possible de suivre un individu dans le temps que s'il a été identifié. Or les identifiants usuels sont trompeurs (homonymies de l'état-civil, ambiguïté de l'adresse, non univocité du numéro de téléphone) ou instables dans le temps (une entreprise peut déménager, changer de dénomination, de secteur d'activité, de taille ...).

La détermination et la gestion des identifiants est l'un des problèmes délicats du système de pilotage. En ce qui concerne les entreprises et établissements français, on dispose des identifiants Sirene et Siret de l'INSEE, mais ils ne sont d'aucun secours pour identifier les clients situés à l'étranger. L'identification des ménages n'existe pas, où se réduit à celle du logement (il est vrai que le ménage est une unité instable dans le temps) ; l'identification des individus est réalisée en France par le numéro d'état-civil (dit " numéro de Sécurité Sociale ") que souvent l'entreprise n'utilise pas, car ses clients ne comprendraient pas pourquoi on le leur demande. D'autres identifiants sont utilisés pour les équipements, les services etc.

On peut s'efforcer d'identifier les individus à partir *d'informations identifiantes*, dont aucune ne suffit mais dont la conjonction peut permettre d'estimer l'identifiant avec une bonne probabilité de succès. Ainsi, si un client fournit son nom, son prénom, son adresse, son numéro de téléphone, son numéro de carte bancaire et le nom de son entreprise, on peut chercher à retrouver le numéro de sa carte de fidélisation, ou bien lui attribuer un identifiant qui le suivra dans ses relations avec l'entreprise. Si le client fournit non pas toutes les informations ci-dessus, mais seulement certaines d'entre elles, on est en face d'un ensemble d' " informations identifiantes " moins complet et donc moins fiable, mais qui peut toutefois être suffisant en pratique.

Nomenclatures

Enfin, les concepts selon lesquels on entend réaliser l'observation doivent être définis : ce sont les nomenclatures (produits, zones géographiques, activité économique d'une entreprise, classe de taille d'une entreprise, catégorie socioprofessionnelle d'une personne, classe de revenu d'un ménage, classe d'âge d'une personne, équipements, services etc.). Les nomenclatures se présentent le plus souvent comme une suite de partitions emboîtées : une nomenclature peut avoir plusieurs niveaux (exemple : commune, canton, département, région). Leur construction est une étape essentielle de la démarche. Un concept qui n'est pas compatible avec la nomenclature ne pourra être présent ni dans les statistiques qu'elle permet d'établir, ni dans les raisonnements que l'on pourra bâtir sur ces statistiques. Les nomenclatures délimitent la sphère de pertinence théorique d'une statistique ; la confusion dans les nomenclatures engendre des dommages : si deux entités utilisent des nomenclatures différentes, elles ne pourront échanger l'information qu'à travers des " tables de passage " qui comportent une imprécision.

Administration des données

Une *donnée*, c'est le *couple* formé d'un *concept* et d'une *mesure*.

Administrer *une* donnée, c'est donc :

- définir le *concept* ;
- décrire la méthode de *mesure* ;
- identifier le *propriétaire* de la donnée (celui qui est chargé de la mesure, et qui est donc autorisé à mettre la donnée à jour).

Administrer *les* données d'une entreprise, c'est :

- faire le travail ci-dessus pour chaque donnée et qualifier son rôle (donnée de référence, intermédiaire de calcul, donnée technique, donnée publique etc.),
- vérifier la cohérence des données (pas de synonymes ni d'homonymes etc.),

- éditer une documentation.

Pour comprendre le partage des responsabilités en matière d'administration des données, il faut se référer aux notions de *maître d'ouvrage* (directions opérationnelles, ou encore " métiers "), et de *maître d'œuvre*.

Le maître d'ouvrage est le *client* (interne à l'entreprise) qui met en œuvre les applications informatiques et utilise l'information. Le maître d'œuvre est celui qui coordonne la fourniture (développement) et l'exploitation de ces applications, et en porte la responsabilité devant le maître d'ouvrage.

C'est le maître d'ouvrage qui, dans chaque domaine, fournit le *contenu* de l'administration des données : il est porteur des définitions et des règles de mesure. La *mise en forme* de ce contenu, par contre, est une tâche technique ; elle peut être remplie par l'informatique sous la responsabilité de la maîtrise d'ouvrage et pour son compte. Les *arbitrages*, parfois nécessaires pour désigner le propriétaire d'une donnée, doivent être rendus par une entité rattachée au Président.

Voici des citations extraites de nos entretiens. Elles montrent l'étendue de l'effort à faire pour améliorer l'administration des données :

" Notre première difficulté est de savoir de quoi l'on parle. Dans les réunions, on discute plus de l'écart entre deux données que des actions à mettre en place pour redresser la situation. Plusieurs entités gèrent sous le même nom des notions différentes. Dans le domaine financier, quand on parle d'enveloppe, chacun met ce qu'il veut sous ce terme : enveloppe initiale, résultante, etc.

" On a du mal à trouver une personne capable de prendre une donnée et de la commenter en disant ce qu'elle représente, si elle est fiable, quelles conclusions on peut en tirer, avec quelle certitude.

" Certaines informations conçues soi-disant pour un pilotage n'ont pas la périodicité requise. Est-ce qu'une entité peut piloter correctement ses consommations à partir d'une comptabilité analytique qui sort deux mois après ? On construit parfois des indicateurs qui ne représentent pas ce que l'on veut piloter. Le manager ne peut pas faire son travail.

Dans beaucoup d'entreprises, l'administration des données est défaillante. Or rien n'est possible en statistique - et, d'une façon plus générale, en système d'information - si les données ne sont pas bien administrées.

Certaines données jouent un rôle particulièrement sensible : ce sont les *données de référence*, auxquelles les diverses applications font souvent appel. Cette désignation recouvre les nomenclatures (découpage géographique, organigramme etc.), les taux de change, etc. Le taux de change est modifié de façon pratiquement continue, les découpages en zones géographiques, organigrammes etc. connaissent aussi des changements mais ils sont moins fréquents. Il importe que chaque donnée de référence soit gérée de façon centrale et unique, et que chacune des applications qui l'utilise puisse accéder à une information à jour.

Or il arrive souvent que les données de référence soient entrées séparément dans chaque application, voire plusieurs fois pour une même application. Il est alors nécessaire de procéder à des mises à jour manuelles chaque fois que la donnée de référence doit être modifiée. Inévitablement, ces mises à jour conduisent à des erreurs : elles sont partielles, certaines sont oubliées, etc. Il en résulte des dommages graves lors des exploitations : que l'on pense par exemple à ce que devient la comparabilité des données si la définition des zones géographiques n'est pas mise à jour de façon cohérente.

Il importe donc que les données de référence soient bien traitées comme telles, et non dispersées

dans les applications où elles réclament des mises à jour multiples qui peuvent être sources d'erreurs.

Programme de l'observation statistique

La définition d'un programme d'observation statistique suppose plusieurs choix :

- nomenclatures fournissant son découpage conceptuel,
- périodicité de l'observation (une donnée ne répond pas aux mêmes besoins selon qu'elle est produite de façon quotidienne, hebdomadaire, mensuelle, trimestrielle ou annuelle),
- finesse de l'observation (on peut choisir un niveau de nomenclature plus ou moins agrégé),
- degré de précision (le taux de sondage sera d'autant plus élevé que l'on souhaite davantage de précision),
- délai de disponibilité de l'information.

Un programme statistique se concrétise sous la forme d'une liste d'enquêtes ou d'exploitations répondant chacune à un sous-ensemble des questions posées. Idéalement, on doit choisir les méthodes permettant de satisfaire les besoins au moindre coût, même si le calcul du coût n'est pas explicite.

Ici se pose un problème. Nous avons dit que la statistique était fondée sur des concepts dont la pertinence s'évalue selon leur adéquation à l'action. Tout est simple si le demandeur est un individu. Tout se complique si l'on entend établir une statistique destinée à un ensemble d'individus dont les besoins peuvent différer, et qui peuvent même avoir des intérêts opposés. Le choix conceptuel pertinent pour l'un peut ne pas l'être pour l'autre. Comment une même statistique peut-elle donc correspondre à des besoins divers?

On ne peut pas répondre à cette question de façon logiquement absolue, car il sera toujours possible d'imaginer un cas où des concepts inconciliables sont pertinents pour des acteurs divers. Cependant cette question reçoit souvent une réponse pratique. L'action de chacun suppose en effet la communication avec d'autres. Sauf pathologie annonciatrice d'un éclatement, toute collectivité humaine définit des concepts qui lui permettent de partager observations, raisonnements et décisions. Le partage du langage lui donne sens et cohésion. Le besoin de communiquer contrebalance la diversité des priorités individuelles.

La procédure utilisée pour construire le programme statistique doit donc être attentive à l'obtention d'un consensus, de sorte que les données puissent servir à la communication. La méthode utilisée par l'INSEE mérite ici d'être observée : les diverses composantes de la société civile sont invitées à participer au CNIS, qui rassemble des représentants des organisations professionnelles, des syndicats, des universités, des administrations, des associations etc. L'ordre du jour des réunions est établi par l'INSEE, qui rédige les comptes rendus. Le CNIS joue un rôle consultatif, mais en pratique ses avis sont généralement écoutés dans la limite des ressources budgétaires. Éclairé par des experts qui répondent à toutes ses questions, le CNS donne son avis sur les nomenclatures, les méthodes statistiques, le programme des enquêtes, exploitations et publications etc. Il n'en résulte pas nécessairement un programme excellent (le système statistique français a des lacunes, notamment sur le revenu des fonctionnaires), mais les diverses composantes de la société civile sont invitées à donner un avis dont il est tenu compte. La qualité de la statistique comme langage commun et outil de communication en est améliorée.

On peut imaginer une structure analogue dans l'entreprise. Un " conseil de la statistique et des études" comporterait des représentants mandatés par les diverses directions. Une équipe assurerait l'apport de compétence technique et d'animation. Les méthodes à utiliser (CVS sur les séries chronologiques, évaluation des élasticités prix, segmentation, publication, diffusion), y seraient discutées posément ainsi que les nomenclatures et le programme des enquêtes. Des questions délicates pourraient être ainsi traitées à froid, dans une ambiance sereine.

Collecte et traitement de l'information

L'entreprise dispose de sources diverses : les applications utilisées dans l'activité opérationnelle fournissent une information potentiellement utilisable au fil de l'eau ; des enquêtes permettent d'obtenir des informations complémentaires ; enfin, il est possible d'acheter des sources extérieures sur le marché des bases de données.

A chacun de ces types de sources correspond un mode d'exploitation spécifique - et il est possible, et utile, de les fusionner entre elles. Ainsi l'on rencontre l'ensemble des situations techniques auxquelles un statisticien peut être confronté.

Sources

Sources exhaustives

Les données fournies par les applications opérationnelles constituent une source exhaustive, mécanique en principe (sans intervention humaine, donc sans erreur humaine), utilisable en temps réel. Cependant cette information est limitée aux relations avec les clients de l'entreprise (elle ne permet donc pas d'éclairer sa part de marché). En outre ces sources exhaustives sont volumineuses, et donc difficiles à exploiter sur le plan statistique.

Sondages internes

On peut chercher à surmonter la lourdeur d'une source exhaustive en faisant un sondage à l'intérieur de cette source, pour examiner à la loupe ses propriétés sur l'échantillon. Nous appellerons un tel sondage " sondage interne ".

Si l'on a su exploiter correctement la source exhaustive, le sondage interne n'apporte rien sur les totalisations, puisque la mesure qu'il procure comporte une imprécision alors que la mesure exhaustive est exacte. Par contre, il permet d'étudier les *corrélations* entre variables, alors que cette étude serait lourde sur les sources exhaustives. Le sondage interne est donc un bon outil sur le chemin de l'économétrie, fondée sur l'utilisation des corrélations.

Un piège : l'exhaustif partiel

Les ingénieurs, épris de précision, sont parfois mal à l'aise devant la technique des sondages qui fournit des mesures entachées d'incertitude. Ils préfèrent utiliser des données exhaustives, recueillies sur une sous population, et en tirer des leçons. Tel directeur est connu pour généraliser souvent, à toute la population, des évaluations obtenues sur une sous-population précisément connue. Les conclusions abusivement tirées d'une expérience personnelle, ou d'une monographie, relèvent de la même démarche.

Cependant si un exhaustif partiel donne des données certaines, elles sont entachées d'un *biais* lorsqu'on les utilise comme estimateur des données globales : une sous-population n'est représentative que d'elle-même, alors qu'un échantillon est représentatif de la globalité dont il est extrait.

Le carré de l'erreur sur une donnée est somme du carré du biais et de la variance :

$E^2 = B^2 + \sigma^2$

L'exhaustif partiel n'est donc préférable au sondage que si le biais qu'il comporte est inférieur à l'écart-type de l'incertitude associée au sondage. Or souvent il lui est supérieur.

Il faut donc que les responsables apprennent à maîtriser l'inconfort que procure l'utilisation des

intervalles de confiance, et sachent qu'une information incertaine est en général plus exacte qu'une information biaisée.

Sondages externes

Les sources exhaustives et les sondages internes à ces sources ne fournissent pas toute l'information dont a besoin l'entreprise. Ils ne donnent pas d'information sur la satisfaction des clients, ni sur les données contextuelles qui permettent d'expliquer leur comportement (revenu, CSP, taille du ménage ; taille de l'entreprise, chiffre d'affaires, etc.).(NB : voir les éléments de théorie des sondages)

Il faut obtenir ces informations par enquête. Ces enquêtes donnent une information entachée d'incertitude, mais une astuce permettrait d'obtenir des estimations plus précises : celle qui consiste à caler les données fournies par les sondages internes sur des données exhaustives connues de façon certaine.

Supposons en effet que j'associe à un sondage externe (portant sur des données que je ne connais pas) un sondage interne, qui procure sur chacun des individus de l'échantillon la mesure de données dont j'ai par ailleurs une connaissance exhaustive.

Dès lors je peux estimer sur l'échantillon la corrélation entre données internes et données externes. Or je connais avec certitude le total de la donnée interne. Je peux donc, en utilisant la corrélation, estimer la donnée externe. Cette estimation sera certaine si la corrélation est absolue ($\rho^2 = 1$). Si la corrélation n'est pas absolue ($\rho^2 < 1$), j'obtiens une estimation non certaine, mais meilleure que celle fournie par un sondage qui ne serait pas confronté à la source interne, et d'autant meilleure que la corrélation est plus forte.

Panels

Un panel, c'est un échantillon dont on suit l'évolution dans le temps (une " cohorte "). La technique du panel vise à combiner les avantages du sondage (économie sur les coûts d'observation) avec le suivi des comportements permis par la présence permanente des mêmes individus dans le panel. Cette technique est donc séduisante au premier abord. Cependant sa mise en œuvre est des plus difficiles.

La difficulté vient d'une contradiction entre la permanence du panel et l'évolution des populations sondées qui se renouvellent par " naissance ", " décès " et " migration " des individus : un ménage se forme, se dissout, déménage ; une entreprise se crée, fusionne, éclate, change d'activité, etc. La continuité du panel, sa représentativité, sont ainsi rompues par des évolutions de type démographique.

La représentativité de l'échantillon risque d'être altérée après quelques années, car s'il a bien été constitué au départ de façon aléatoire les " décès " ont obéi non à une loi aléatoire, mais à une loi naturelle qui ne frappe pas au hasard. Il faut renouveler l'échantillon pour introduire des représentants des " naissances " : mais ces nouveaux individus n'ont pas de passé, et il faut donc les ignorer dans les calculs d'évolution. Les pondérations à accorder aux individus doivent varier dans le temps, pour donner une image fidèle de la proportion des diverses strates : mais des pondérations variables peuvent avoir des effets surprenants pour l'intuition.

Autres sources

On trouve des bases de données sur le marché. Certaines contiennent des informations utiles pour compléter les données internes, et permettent d'éviter la dépense d'une enquête.

Il faut distinguer parmi ces sources celles qui permettent d'identifier l'individu, et autorisent donc une fusion de fichier puis des calculs de corrélation, et celles qui ne fournissent que des totaux sur des sous-populations, et ne sont utilisables qu'à ce niveau agrégé. Les secondes sont potentiellement

moins riches et moins utiles. Les contraintes imposées par la CNIL peuvent parfois obliger à dégrader une source : elle pourrait donner des informations individuelles, mais on s'interdit de les utiliser et on ne peut se servir que des données agrégées.

La politique d'achat en matière de sources externes requiert une expertise précise. Elle doit être attentive :

- au niveau d'agrégation auquel il est possible de fusionner la source externe et la source interne,
- au prix, comparé à celui d'un sondage procurant une information analogue,
- au prix, comparé à l'utilité du complément d'information par rapport aux sources internes.

Délimitation des domaines de l'exhaustif et du sondage

Nous avons distingué ci-dessus les calculs qui permettent d'évaluer des totaux (ou des moyennes) et ceux qui permettent d'estimer des corrélations.

Si l'on utilise le langage de la mathématique, on distingue parmi les mesures associées aux variables les " moments d'ordre un " (moyennes, totaux) et les " moments d'ordre deux " (variance, corrélation).

Les moments d'ordre deux sont à l'origine des calculs économétriques, des modèles, et de la segmentation qui implique une analyse de la variance.

L'exigence de précision concernant les moments d'ordre un est souvent élevée et leur calcul est simple (il suppose une addition). Le calcul des moments d'ordre deux est plus lourd, et l'exigence de précision moins élevée : en matière de variance ou de corrélation, on se contente souvent d'une estimation.

Ainsi, dans les cas où l'on dispose d'une grande source exhaustive, on utilisera cette source pour calculer moyennes et totaux, et on utilisera un sondage interne à cette source pour calculer variances et corrélations. Le sondage est également utile pour estimer les moments d'ordre un et deux des variables non comprises dans la source exhaustive.

Ceci permet de voir clairement le domaine de validité des sondages : ils sont précieux pour les travaux économétriques ou de modélisation, pour la segmentation, pour étudier les corrélations entre variables internes et variables externes (par exemple entre facture et opinion).

Vérification

Avant d'exploiter une source statistique, il faut la vérifier. La vérification porte sur chaque enregistrement individuel. On distingue vérification logique et vérification sémantique. La vérification logique est simple : il s'agit de voir si le questionnaire est complet, si les codes ont des valeurs admissibles, si les totalisations tombent d'aplomb, si tel ou tel taux est juste (TVA, taux de change etc.). La vérification sémantique est plus subtile, car elle porte sur la vraisemblance des informations ou des ratios, comparés par exemple à leur distribution dans l'ensemble de la population.

L'idéal est de faire ces vérifications au moment de la saisie : on peut alors éditer des messages d'erreur ou d'anomalie, et la personne chargée de la saisie effectue les corrections sur le champ. S'il s'agit d'une source obtenue après saisie, il est indispensable de la soumettre à un programme de vérification avant de l'exploiter.

Exploitation

Exploiter une source statistique (fichier ou enquête), c'est définir la liste des tableaux qui seront produits pour présenter les résultats, puis programmer et faire réaliser par l'ordinateur les opérations

nécessaires pour les obtenir.

Les résultats se présentent sous forme de tableaux, obtenus en faisant des tris dans la population considérée, puis des totalisations. Si l'enquête a été réalisée par sondage, chaque réponse doit avant d'être additionnée aux autres être pondérée par l'inverse du taux de sondage dans la strate à laquelle appartient l'individu. Le calcul de cette pondération peut être délicat : dans un panel, par exemple, la représentativité d'un même individu peut varier dans le temps, et il faut faire évoluer sa pondération.

Toujours dans le cas des sondages, le calcul des intervalles de confiance nécessite le recours à des formules un peu complexes. Si l'on considère une évolution dans le temps, il faut savoir distinguer dans les résultats ce qui revient à l'évolution à population constante (suivi de " cohortes ", panel " cylindré ") et ce qui revient à la démographie (" mort ", " naissance " et " migration " d'individus). Cette distinction est importante dans les statistiques d'entreprise ; la présentation des résultats en est inévitablement alourdie, leur interprétation n'est pas aisée.

L'utilisation de pondérations variables peut avoir des effets surprenants : il peut ainsi arriver que le taux de croissance de la moyenne de deux variables ne soit pas contenu dans l'intervalle des taux de croissance de ces variables (effet de structure). Cet effet est notoire dans les indices de prix, qui sont des indices de Paasche à pondération variable.

Il est souhaitable d'établir le programme d'exploitation au moment même où l'on définit le programme d'enquête, le dessin du questionnaire etc. : on s'assure ainsi que l'on dispose bien de toutes les données dont on aura besoin, et que la collecte n'est pas excessivement riche. Il est vrai que trop souvent les statisticiens ne respectent pas cette règle, et construisent le programme d'exploitation à chaud, alors même qu'ils croulent sous les questionnaires et vérifications ...

Fusion de fichiers

La fusion de fichiers est l'une des opérations les plus puissantes de la statistique. Supposons que l'on ait fait deux enquêtes sur une même population, et que l'on ait observé la variable X (n modalités) dans l'une, la variable Y (m modalités) dans l'autre. Fusionner les fichiers revient à constituer un nouveau fichier, indiquant pour chaque individu les valeurs des variables X et Y. Ainsi on peut calculer leur corrélation, et établir le tableau qui les croise (nm modalités) : on peut donc dire de façon très exacte que la fusion de fichiers *multiplie* les possibilités d'exploitation.

La CNIL est souvent hostile aux fusions de fichiers, car cette technique puissante permet de recouper sur un même individu des informations d'origines diverses qui n'avaient pas été collectées dans ce but. Il est vrai que la fusion de fichiers, surtout lorsqu'elle est associée à des techniques d'analyse discriminante (" scoring "), risque de donner trop de puissance à des démarches indiscrètes ou malveillantes. Les contraintes déontologiques doivent être ici particulièrement strictes.

La fusion de fichier est d'ailleurs une technique délicate, car les fichiers sont souvent incomplets, et les identifiants souvent mal contrôlés. En outre la fusion est l'occasion d'une vérification. On procède par étapes : " mise en forme ", " interclassements ", " appariements ", " réintroduction d'unités absentes ", " confrontations ", " mise à niveau des données ". Ces opérations sont coûteuses, lourdes, et demandent un savoir faire spécialisé.

Séries chronologiques

Il est utile, dans un environnement concurrentiel, de percevoir rapidement les réactions du marché, et d'interpréter les inflexions de ses tendances. Il faut pour cela (a) disposer de séries chronologiques de bonne qualité, (b) traiter ces séries de façon à faire apparaître la tendance instantanée.

L'exigence est analogue si l'on entend procéder à des expérimentations : il faut disposer d'une mesure d'assez bonne qualité pour pouvoir évaluer les effets d'une nouvelle tarification, d'une

nouvelle offre que l'on teste sur un échantillon de clients.

Qualité des séries chronologiques

Une mesure exacte si elle est adéquate au concept que l'on entend mesurer. Ainsi, additionner les factures émises pendant un mois n'est pas adéquat pour mesurer la valeur de la production de ce mois, puisque certaines de ces factures peuvent correspondre à la production des mois précédents, et qu'une partie de la production du mois considérée sera facturée plus tard. Evaluer les dépenses d'un mois en considérant les factures reçues le 10 du mois suivant n'est pas adéquat non plus, pour des raisons analogues.

Les montants des factures émises ou reçues ont la faveur des comptables, parce qu'il s'agit de données *précises*. Le problème, c'est qu'elle n'ont pas de valeur économique. L'économiste veut avoir une estimation sans biais de la production du mois, ou des dépenses du mois. Il accepte ainsi une perte en précision pour obtenir un gain en exactitude, au sens que nous avons donné ci-dessus à ce terme.

Autre exemple : les balances comptables mensuelles visent non à donner une estimation des données du mois, mais à faire progresser tout au long de l'année le classement des enregistrements comptables, de façon à faciliter l'arrêté des comptes en fin d'année. Les données mensuelles ainsi inscrites ne sont pas révisées par la suite, et la donnée relative au mois m est la somme de ce que l'on connaît sur le mois m vers le 10 du mois m + 1, plus la somme des corrections et compléments apportés aux données des mois antérieurs : il s'agit en fait plus de la mise à jour partielle d'un cumul depuis le début de l'année que d'une donnée mensuelle.

" Badger " les données

Seules certaines données se prêtent donc à l'établissement de séries chronologiques. Il convient de distinguer les données selon l'utilisation qui peut en être faite. Il existe plusieurs façons de mesurer un effectif, un chiffre d'affaires, une production, etc. Il convient de réserver à certaines de ces mesures (si possible, une seule par concept) le badge " publiable " ; les autres porteront un badge technique : données " comptables ", " de gestion ", " intermédiaires de calcul " etc. En général les données publiables sont celles qui ont un contenu économique, et non celles qui sont les plus faciles à mesurer : la production d'un mois ne peut être connue qu'après des traitements délicats, alors que le montant des factures émis durant un mois (indicateur fallacieux sur le plan économique) est disponible plus vite.

Personne ne peut reprocher à un comptable de calculer le total des factures émises durant un mois, et de l'utiliser à des fins de vérification et de recoupement. Mais ce n'est pas une donnée économique, car elle ne peut pas aider à calculer la production, la productivité, ni à estimer le résultat etc. Il y aura souvent conflit entre la donnée comptable (précise mais trompeuse) et la donnée économique (fidèle mais imprécise). Il faut savoir préférer la seconde, qui seule peut éclairer le raisonnement, et apprendre à supporter l'imprécision de l'évaluation.

Traiter les séries chronologiques

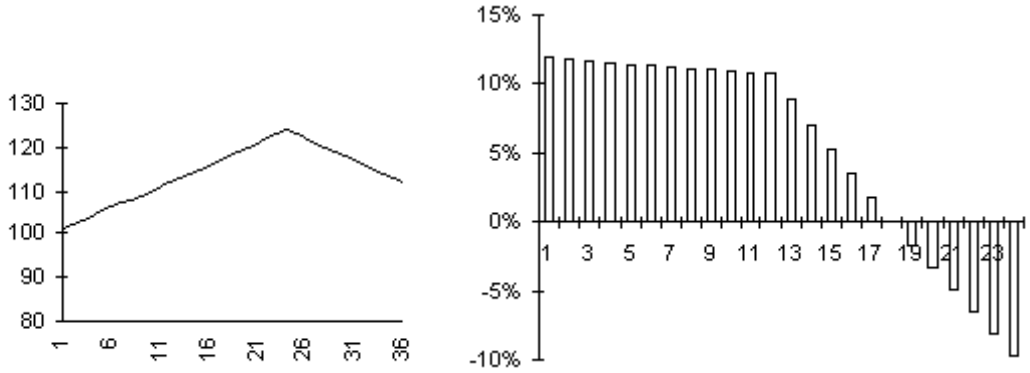
Considérons une série mensuelle publiable. Son interprétation requiert un traitement. En effet, une mesure mensuelle est un produit T = n*I, où n est le nombre de jours d'activité du mois et I l'intensité de la consommation. Si l'on veut dégager cette dernière, il faut appliquer à la donnée brute une correction tenant compte du nombre de jours du mois, voire du nombre de jours ouvrables selon les secteurs économiques considérés.

Enfin les intensités elles-mêmes sont influencées (a) par le mouvement saisonnier propre au marché considéré, (b) par la tendance sous-jacente, qui seule intéresse vraiment le management. Pour détecter la tendance sous-jacente il faut éliminer le mouvement saisonnier. C'est à quoi sert la

correction des variations saisonnières, ou CVS.(NB : voir la théorie de la CVS).

Certains jugent les CVS trop compliquées pour leur goût. Ils préfèrent les données brutes, qu'ils jugent plus " concrètes ", et utilisent pour éliminer l'effet saisonnier la comparaison dite " R/R " : on compare la réalisation d'un mois à celle du mois correspondant de l'année précédente.

Cette méthode est fallacieuse. Considérons une série mensuelle observée sur trois ans, et l'évolution du rapport R/R sur les deux dernières années :



Série chronologique et rapport R/R

La série est croissante pendant les deux premières années, se retourne à la fin de la seconde année et décroît pendant la troisième année. Cependant le rapport R/R reste croissant pendant la première moitié de la troisième année, et ne se retourne qu'au milieu de cette année. Si l'on s'en fie à ce rapport, on se trompe donc pendant six mois sur le signe de l'évolution, et l'on commet une erreur de six mois sur la date du retournement.

D'une façon plus générale, l'évolution du rapport R/R résulte des deux tendances observées lors de l'année en cours et de l'année précédente : elle mêle donc deux informations, et elle est malgré son apparente simplicité plus difficile à interpréter que celle d'une série CVS.

Utiliser les séries brutes ou le rapport R/R peut paraître rassurant à ceux qui n'ont pas de culture statistique ; mais ce confort se paie par des erreurs d'appréciation qui, de la part d'un décideur, peuvent avoir des conséquences graves.

L'empirisme débridé des calculateurs a produit d'autres méthodes plus sophistiquées, mais analogues, qui présentent les mêmes défauts que le rapport R/R ou des défauts plus graves encore : ainsi la méthode qui consiste à associer à un mois le rapport " somme des douze dernier mois, divisée par la somme des mois n - 13 à n - 24 ", équivaut à la suite des opérations suivantes :

- moyenne mobile sur douze mois ;

- rapport R/R ;

- retard de six mois.

L'empirisme, armé de l'ordinateur, peut produire une grande variété d'indicateurs synthétiques, de graphiques, qui ne veulent rien dire et sont même fallacieux. Les graphiques diffusés, étant erronés ou fallacieux, peuvent avoir une influence sur la décision. L'attachement de l'entreprise aux graphiques présentant l'évolution des rapports R/R, graphiques qui n'apportent aucune information

utile et qui peuvent même tromper, est inquiétante.

Publication

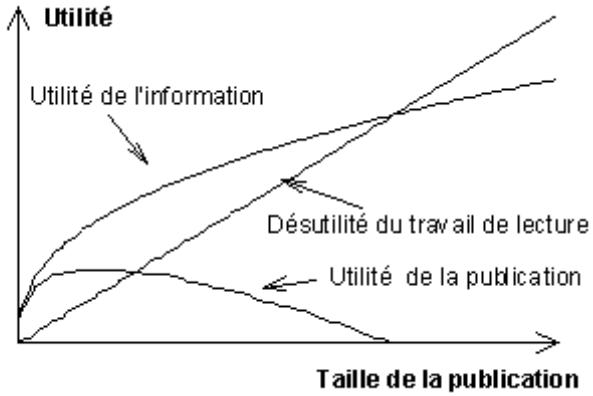
Les publications statistiques se classent en deux familles :

- les compilations qui, comme les cours de bourse, fournissent à des lecteurs habitués un grand nombre de résultats parmi lesquels ces lecteurs savent trier ce qui les intéresse ;

- les tableaux de bord qui fournissent à des lecteurs *a priori* non spécialisés une information synthétique : à cette catégorie appartiennent les EIS qui visent à éclairer des responsables. Les publications statistiques internes à une entreprise relèvent en majorité de cette seconde famille.

Or les exigences qu'impliquent ces deux familles sont différentes. Une compilation fournit beaucoup de nombre non commentés. Beaucoup croient qu'un tableau de bord doit se présenter de la même façon, et qu'il sera d'autant plus utile qu'il apporte plus d'informations. Ce n'est pas exact.

Pour chaque lecteur d'un tableau de bord, on peut en effet considérer l'indicateur le plus intéressant, puis le second etc. L'utilité totale en fonction du nombre d'indicateurs est représentée par une courbe dont la concavité est tournée vers le bas (graphique 2). Par ailleurs, l'effort demandé par la lecture d'une publication est à peu près proportionnel à la taille de celle-ci. Il en résulte que l'utilité de la publication, différence entre l'utilité des indicateurs et l'effort de lecture, passe par un maximum pour un nombre d'indicateurs donné, puis décroît et peut même devenir négative.



L'utilité en fonction de la taille d'une publication

Un tableau de bord doit donc être sélectif, sobre, visuel et commenté.

- *sélectif* : retenir d'abord les indicateurs les plus intéressants.
- *sobre* : ne publier que le nombre d'indicateurs correspondant au maximum d'utilité pour le lecteur.
- *visuel* : la visualisation graphique est indispensable, car les nombres sont difficiles à lire.
- *commenté* : les commentaires facilitent la compréhension des graphiques, en communiquant au lecteur les raisonnements de bon sens dont le statisticien se sert lui-même, ou des indications sur le contexte qui permettent d'interpréter les données.

Il est difficile de respecter ces exigences simples. Il est pénible pour un statisticien de limiter la taille de sa publication, car les indicateurs qu'il a produit lui semblent tous intéressants. Les représentations graphiques sont diverses, et le choix entre courbes, histogrammes, formages, ainsi qu'entre diverses échelles est délicat. Enfin la rédaction du commentaire est embarrassante : la gamme est large de la

paraphrase des nombres à l'explication qui nécessite de les confronter à une théorie ; le rédacteur a peur d'être banal, d'en dire trop, de prendre le risque d'exprimer une opinion qui peut être discutée.

La plupart des publications statistiques et tableaux de bord ne respectent donc pas ces exigences. On rencontre souvent un empilement de nombres, des graphiques absents ou mal choisis, des commentaires absents ou réduits à un charabia technique, faisant souvent la part belle à des subtilités comptables sans signification économique. Rien de tout cela ne peut servir à un décideur.

Modélisation et études

Les trois techniques essentielles utilisées pour interpréter les données sont l'analyse des données, l'économétrie, et la classification (ou segmentation).

L'analyse des données

L'analyse des données formalise les techniques de statistique descriptive (analyse factorielle des correspondances, analyse en composantes principales, analyse discriminante etc.). Elle considère uniquement les relations algébriques entre données et ne suppose aucune hypothèse explicative *a priori*. C'est un instrument d'exploration qui suscite des questions et aide à détecter des phénomènes, ainsi qu'à les décrire et les visualiser, mais ne vise pas à les expliquer.

L'analyse des données est une bonne méthode lorsque l'on doit aborder un gros corpus de données sur lequel on n'a pas d'idées *a priori*, et que l'on veut l'explorer rapidement. Pour traiter les questions que suscite cette exploration et les confronter à des schémas explicatifs, il faut recourir à d'autres méthodes.

Segmentation

La segmentation, ou encore classification, est l'opération par laquelle on définit des classes (segments) dans lesquelles on range les individus appartenant à une population. La construction d'une segmentation suppose des choix : sur une population donnée, plusieurs segmentations sont logiquement possibles. Il importe donc de savoir à quelles fins on construit une segmentation, et de s'assurer que les choix sur lesquels elle repose sont adéquats à ces fins.

Toute action suppose une segmentation, car un cas particulier ne peut être traité que si l'on sait le ranger dans une classe : par l'observation des symptômes, le médecin range un patient dans une classe (la maladie), ce qui lui permet d'établir sa prescription. De même, par l'observation des variables relatives à un client, le commercial classe celui-ci dans un segment, ce qui lui permet d'appliquer le traitement personnalisé convenable. Observons en effet que personnaliser la démarche commerciale, ce n'est pas traiter un client selon les caractéristiques ineffables de son individualité, mais l'affecter à une classe pour laquelle une démarche a été définie *a priori*.

On distingue deux types de segmentation dont les utilisations sont différentes : la segmentation *a priori* comporte peu de classes (quelques dizaines au plus) et repose sur des critères faciles à observer ; la segmentation *a posteriori* résulte d'une utilisation poussée de la statistique, et fournit un grand nombre de classes.

Segmentation a priori

La segmentation *a priori* fournit à l'entreprise un langage et des points de repère. Les classes du découpage, leurs dénomination, deviennent des concepts sur lesquels se construisent raisonnement et communication. Des statistiques agrégées, calculées sur chaque classe par totalisation, permettent de la caractériser.

Segmentation a posteriori

La segmentation *a posteriori* résulte d'une utilisation extensive des bases de données individuelles. En les soumettant à des démarches d'analyse des données, elle délimite quelques centaines de classes. Il ne s'agit donc plus ici de fournir au raisonnement un outil conceptuel, mais de repérer des " niches " sur lesquelles peuvent être engagées des actions spécifiques, et que la segmentation *a priori* n'aurait pas permis de repérer.

Qualité de la segmentation

La segmentation structure souvent l'organisation. En effet des entités spécialisées sont créées dans l'organigramme pour assurer chacune la relation commerciale avec une classe. Dès lors la segmentation prend une existence institutionnelle qui peut masquer ses origines et les choix sur lesquels elle est fondée. Il est important de relativiser ces choix pour éviter que la segmentation ne se fossilise.

Par ailleurs la finalité de la segmentation doit être prise en compte lors de sa définition. Fonder une segmentation sur des critères relatifs à la relation entre l'entreprise et le client serait au rebours des besoins de la modélisation (cf. encadré).

Les besoins de la modélisation

Un modèle qui vise à expliquer la consommation d'un client, et à révéler un potentiel, doit distinguer :

- les variables exogènes qui décrivent le contexte (dans le cas d'une entreprise : activité, taille, localisation, chiffre d'affaires etc.)
- les variables endogènes qui décrivent la consommation et sont modélisées en considérant les exogènes comme des variables explicatives. Il est ensuite intéressant de comparer la consommation effective à la valeur estimée par le modèle pour détecter et interpréter les écarts (sur- ou sous-consommation).

Il importe donc de privilégier les variables exogènes pour définir une segmentation sur laquelle un modèle pourra ensuite s'appuyer. Un segmentation fondée sur la consommation ne peut pas servir à cette fin, puisqu'elle prend comme variables d'entrée celles que la modélisation doit justement expliquer.

Exemple : si on a défini les classes en fonction du volume consommé (gros consommateurs, petits consommateurs etc.), la réussite d'une politique commerciale portant sur les petits consommateurs aura pour effet de vider cette classe au bénéfice de la classe des gros consommateurs, donc éventuellement de diminuer sa consommation totale !

L'économétrie

(NB : voir les éléments de théorie de l'économétrie).

L'interprétation des statistiques, comme celle de toute observation, réclame une confrontation avec un modèle théorique, qui outre le découpage conceptuel propre à la description postule des relations fonctionnelles entre les données observées. Ainsi la statistique fournira la mesure du revenu R et de la consommation C, mais c'est la théorie qui fournira l'hypothèse d'une relation fonctionnelle (comportant éventuellement un aléa) du type C = f(R).

Parfois, la théorie est simple, et il est inutile de l'expliciter : on a alors l'impression que les données parlent d'elles-mêmes. Souvent par contre il est nécessaire pour faire parler les données de se référer à une toile de fond théorique.

Il est alors commode de formuler la théorie sous la forme d'une équation qui permet d'expliquer une donnée à partir des autres et comporte un terme aléatoire, par exemple :

$Y_t = a + b.PIB_t + c.IPI_t + d.EMP_t + e.CONS_t + f.P_t + e_t$

où :

- a, b, c, d, e, f sont des coefficients,
- Y est le taux de croissance annuel du marché considéré (indexé par le temps t),
- les autres variables désignent sous des notations évidentes les taux de croissance du PIB, de la production industrielle, du prix relatif du produit étudié, de l'emploi et de la consommation,
- e_t est assimilé à un bruit blanc.

Les techniques de l'économétrie permettent d'estimer les valeurs les plus plausibles des coefficients. Ici, le coefficient f est l'élasticité de la consommation au prix, le coefficient a est le taux de croissance tendanciel, etc. Ces techniques comportent également des tests qui permettent de vérifier que les données constatées n'invalident pas la spécification de l'équation.

Les économètres testent plusieurs spécifications avant d'en retenir une qui soit à la fois féconde sur le plan théorique, et acceptable du point de vue des tests statistiques.

Il est possible de lier par de telles équations les données qui décrivent un domaine de l'économie : on aura alors construit un modèle économétrique. On distingue parmi les données les " variables exogènes ", paramètres fournis en entrée au modèle, et les " variables endogènes ", qui résultent du calcul. Par exemple, dans le cas de l'équation ci-dessus qui constitue un modèle simple, on peut considérer Y comme une endogène, " expliquée " par les variables situées dans le terme de droite de l'équation.

L'économètre est parfois tenté de travailler de façon purement statistique, sans référence aucune à une théorie. Il sont alors tentés de donner foi à des corrélations constatées dans le passé, mais qui sont accidentelles et n'ont donc aucune valeur prédictive. Le surdimensionnement de la flotte navale s'expliquerait par une erreur de ce type.

Lorsqu'il est bien fait, le travail de l'économètre est technique, long et compliqué. Il lui est difficile de communiquer à un non expert les raisonnements qu'il a faits pour sélectionner les spécifications. En outre la seule autorité des résultats qu'il obtient, c'est de ne pas avoir été rejetés par les tests lors de la confrontation avec les données observées.

Certains tirent parti de cette difficulté de communication et de cette autorité limitée pour révoquer en doute les résultats de l'économétrie, et leur préférer d'autres hypothèses qui sans doute ne résisteraient pas à la confrontation avec les données, mais correspondent mieux à leurs préjugés, et qu'ils ne se soucient d'ailleurs pas de soumettre à des tests.

Il n'est pas facile pour un dirigeant d'entreprise d'utiliser convenablement les résultats de l'économétrie, car cela demande savoir faire et expérience. Il est certain en tout cas qu'un dirigeant fait prendre un risque à son entreprise lorsqu'il tourne le dos à ces résultats pour se fier à des affirmations plus séduisantes, mais qui leur sont contraires.

Les prévisions

Si l'on dispose de prévisions sur les exogènes, on peut en faisant tourner le modèle sur ces valeurs prévisionnelles en déduire une prévision des endogènes. La qualité de cette prévision dépendra de celle des exogènes, mais aussi de la validité du modèle en dehors de l'intervalle de temps sur lequel il est étalonné. Il est utile, pour vérifier ce dernier point, d'étalonner une même équation sur des

intervalles de temps différents, et de comparer les estimations des coefficients ainsi obtenues.

Il n'est pas toujours possible de disposer de prévisions sur les exogènes : on dit alors que ce sont des variables explicatives, mais non prédictives. Pour produire un modèle prévisionnel, il faut limiter les spécifications en ne retenant comme exogènes que des variables prédictives.

L'information prévisionnelle est plus précieuse que l'information sur le passé proche: on ne peut rien faire pour corriger le passé, alors que l'on peut réagir si la prévision montre que l'on est devant un obstacle.

Le mot prévision est à comprendre ici en un sens technique précis.

Si vous conduisez une automobile la nuit, les phares portent droit devant vous et indiquent les tournants de la route et les obstacles. Vous manœuvrez de façon à prendre le virage, éviter l'obstacle etc., et sortez donc de la trajectoire que les phares avaient indiquée. Vous n'accusez pas pour autant les phares d'avoir donné des indications erronées.

De même, les prévisions que l'on obtient par l'économétrie, tendancielles en tout ou partie (les exogènes prédictives sont souvent obtenues par extrapolation), montrent des obstacles que le pilotage de l'entreprise s'emploiera à éviter. Les réactions que suscite la prévision font donc - heureusement - que la prévision ne se réalise pas. Mais, si l'on ne comprend pas que la prévision économétrique est analogue aux phares d'une automobile, on croit que le prévisionniste a eu tort.

Il ne faut pas s'étonner si l'expert garde alors pour lui ses prévisions. Il ne les communique, avec prudence, que si elles lui montrent un obstacle qu'il estime vraiment dangereux pour l'entreprise.

Nota bene : le mot " prévision " reçoit parfois un sens différent de celui que nous avons indiqué ici : il désigne des données établies afin de mensualiser la présentation du budget. La finalité politique du budget - que nous ne remettons pas en cause, mais que nous distinguons d'une finalité économique - donne à ces " prévisions " un caractère hautement conventionnel. Ce sont elles qui sont considérées lorsque l'on calcule des rapports R/P (réalisation/prévision) dans la présentation des statistiques, rapports qui n'ont guère plus d'utilité pour évaluer la tendance que les rapports R/R que nous avons évoqués.

Il est utile ici de regrouper diverses remarques concernant les statistiques commerciales :

- les experts disposent d'estimateurs sans biais, de séries corrigées de variations saisonnières et interprétables, d'analyses économétriques, de prévisions tendancielles sur les six mois à venir ;
- l'entreprise leur réclame, et obtient, des données biaisées, présentées sous la forme de rapports R/R ou R/P, accompagnées de lourds commentaires comptables et sans prévisions autres que budgétaires et donc conventionnelles ;
- les données utilisées par l'entreprise sont fallacieuses et peuvent donc conduire à des interprétations et décisions erronées ;
- l'utilisation de données économiques de qualité suppose un changement de l'attitude de l'entreprise vis-à-vis des données et des experts.

Points divers importants

Difficultés propres à la statistique des entreprises

Une population se prête à la statistique :

- si elle est assez nombreuse pour que l'on puisse estimer les moyennes, totaux, dispersions et corrélations des variables qui la décrivent, la classer en segments et procéder aux mêmes

estimations sur les segments ;
- si elle est assez stable dans le temps pour que l'on puisse étudier l'évolution des variables ci-dessus.

Les populations des ménages et petites entreprises répondent à ces exigences. Cela ne veut pas dire qu'elles soient homogènes : elles sont au contraire très diversifiées. Mais ces populations sont assez nombreuses pour que l'on puisse les segmenter, réduire leur diversité à des différences entre classes, et en rendre compte à la fois en description instantanée et en évolution.

La situation n'est pas la même en ce qui concerne les grandes entreprises. La population est alors peu nombreuse ; bien qu'exhaustive, elle a les caractéristiques d'un petit échantillon qui serait tiré dans une population infinie, celle des " entreprises possibles ", population purement idéale et donc hors d'atteinte par l'observation. Si cet " échantillon exhaustif " est de taille trop petite, les moyennes, totaux etc. que l'on peut en tirer ont une valeur purement descriptive, mais ne se prêtent ni à l'analyse économétrique, ni aux calculs prévisionnels.

C'est flagrant en ce qui concerne les grandes entreprises. Saint-Gobain, Rhône-Poulenc, Bouygues, la Lyonnaise des Eaux, EDF sont des entreprises qui ont une forte individualité et qu'il convient d'étudier en tant qu'individus, même si on les compare à leurs concurrents dans certains domaines. France Télécom, Air France, la Société Générale peuvent être situées chacune sur la toile de fond d'un secteur (opérateurs télécoms, transport aérien, banque), mais présentent aussi chacune des particularités uniques et importantes.

Bref : les entreprises, surtout les grandes entreprises, constituent une population qui ne se prête pas ou mal à la description statistique. *Les mesures que l'on peut faire sur ces populations, même parfaitement précises, sont à considérer sur le plan du raisonnement comme des estimations de faible qualité*.

Cela ne veut pas dire que l'on ne puisse rien tirer des données relatives aux grandes entreprises, mais qu'on est plutôt avec elles dans le domaine de la *monographie* que de la statistique.

Portée et limites de la monographie

Une monographie, c'est une étude qui ne concerne qu'un individu, que l'on considère sous divers aspects entre lesquels on cherche à établir des relations. L'approche monographique est utile en statistique, car elle permet par une démarche purement descriptive de préparer le cadre conceptuel de l'observation d'un domaine nouveau. Mais elle est aussi dangereuse, car l'on est souvent tenté de donner à une monographie une portée excessive, en généralisant indûment ses enseignements (c'est un travers fréquent).

Bref : la monographie est à considérer comme une étape préliminaire du travail statistique ; dans certains domaines comme celui des grandes entreprises on doit en rester à cette étape, et interpréter les résultats qu'elle fournit en se gardant de les généraliser. On peut d'ailleurs, en compilant les monographies d'entreprises d'un même secteur, dégager des ratios et moyennes qui, même incertains, donnent des points de repères utiles.

Connaissance d'un marché concurrentiel

L'effet immédiatement visible de la concurrence, c'est la perte de chiffre d'affaires qu'elle induit. Mais il en est un autre, plus insidieux et peut-être plus grave à terme pour une entreprise qui exploite un réseau : l'incertitude sur la demande future est accrue, car d'une part la connaissance de la demande adressée aux concurrents est imparfaite, d'autre part le choix des clients entre offres concurrentes comporte un aléa qui n'existait pas auparavant. Or, si l'on suit le raisonnement qui fonde les règles de dimensionnement, on voit qu'un accroissement de l'incertitude sur la demande provoque toutes choses égales d'ailleurs un accroissement du coût du réseau. *Maintenir la qualité de la connaissance de la demande par delà les obstacles que l'arrivée de la concurrence élève devant cette connaissance*

est un enjeu important.

Pour connaître le marché des concurrents, trois voies se présentent : soit il existe sur ce marché des intermédiaires auxquels on peut acheter l'information (c'est ce que font les transporteurs aériens), soit on passe un accord d'échange d'information avec les concurrents, soit ... on utilise les techniques du renseignement en assumant les risques qu'elles comportent.

Domaines connexes à la statistique : comptabilité et mesure des coûts

La qualité de la comptabilité et de l'évaluation des coûts de production est d'autant meilleure que les méthodes utilisées sont plus proches de celles de la statistique (mot que nous utilisons pour désigner les données propres à alimenter un raisonnement économique). Ainsi, une provision comptable correcte est une estimation sans biais de l'écart entre données connues et données réelles, etc. Cependant le langage de la comptabilité, héritier d'une longue tradition, n'est pas identique à celui de la statistique.

Nous allons regrouper ici des remarques déjà énoncées de façon éparse ci-dessus. Pour éviter tout malentendu, disons clairement que nous considérons les approches comptables et statistiques comme *également légitimes*, chacune dans son ordre. Les difficultés viennent de ce que l'on confond souvent ces deux approches, et que l'on utilise indûment des données comptables à des fins statistiques.

Une entreprise doit savoir être polyglotte en matière de données : elle doit savoir parler le langage de la comptabilité avec ses actionnaires, ses créanciers, l'administration fiscale etc., et le langage de la statistique (qui sera souvent purement interne) pour interpréter sa situation économique et préciser sa démarche marketing. Bien souvent les managers refusent cette complexité, qu'ils jugent superflue. Et comme les données comptables sont nécessairement établies à des fins réglementaires, elles s'imposent au delà de leur cercle de validité, comme si elles fournissaient une représentation économique correcte de l'entreprise.

Statistique et comptabilité

La comptabilité est essentiellement une méthode de classement, permettant de dégager des soldes et des totaux utiles pour la gestion de l'entreprise ainsi que pour la fiscalité.

Statistique et comptabilité produisent toutes deux de l'information, et partagent certaines méthodes (découpage conceptuel, classement etc.). Cependant elles diffèrent sur des points essentiels. Il serait donc erroné de vouloir supprimer à toute force les écarts entre données statistiques et données comptables, car cela reviendrait à " caler la statistique sur la comptabilité ", donc à altérer la qualité de la statistique.

La comptabilité procède par classement des recettes et dépenses attestées par des documents, des " effets de commerce ". Elle est maladroite lorsqu'il s'agit de procéder à des estimations qui ne sont pas fondées sur de tels documents, et applique en outre un " principe de prudence " qui biaise les évaluations.

Balances mensuelles

L'exercice comptable a le plus souvent une durée annuelle. C'est à la fin de l'exercice que l'on " arrête " les comptes, opération coûteuse qu'il n'est pas souhaitable de renouveler souvent. Les balances mensuelles n'ont pas pour but premier de fournir une évaluation mensuelle de l'activité de l'entreprise, mais de faire progresser le classement comptable tout au long de l'exercice.

Les données comptables relatives au mois m recouvrent ainsi les comptes relatifs à ce mois, selon l'image qu'en donnent les pièces disponibles au moment de sa clôture, additionnés à la somme des

corrections apportées aux comptes des mois précédents en classant les pièces relatives à ces mois mais parvenues après la clôture du compte du mois m - 1.

Il en résulte que les données mensuelles fournies par la comptabilité ne constituent pas de véritables séries chronologiques, et sont impropres à l'analyse des tendances (sauf toutefois si les provisions font l'objet d'un calcul rigoureux, ce qui est rarement le cas).

Exactitude et précision

Le comptable équilibre les comptes au centime près, car l'expérience lui a montré qu'un petit écart pouvait être l'indice d'une importante erreur de classement. Les procédures d'estimation, qui rendent la mesure imprécise (" intervalle de confiance "), ne lui conviennent donc pas.

A la précision, le statisticien préfère l'exactitude. Si le comptable évalue les dépenses du mois m en se fondant sur les factures reçues à la clôture du compte mensuel, le statisticien complète cette évaluation en estimant le montant des factures qui restent à recevoir. Alors que le comptable classe des montants figurant dans les documents disponibles, le statisticien complète cette information en estimant les montants qui figureront dans des documents que l'on n'a pas encore.

Dans le meilleur des cas, le comptable estime les informations manquantes en évaluant des " provisions ", calculées à partir de données observables, souvent physiques, corrélées avec les données comptables. Si le calcul des provisions est bien fait, ce qui est rare, il équivaut à une estimation statistique.

Principe de prudence

Une estimation statistique doit avant tout être " sans biais ", c'est-à-dire telle que si elle peut s'écarter sur un cas particulier de la valeur vraie connue *a posteriori*, sur un grand nombre de cas la somme des estimations sera proche de la somme des valeurs vraies (" loi des grands nombres ").

Pour un comptable, l'écart entre évaluation et valeur vraie n'a pas les mêmes conséquences selon qu'il induit un excès d'optimisme ou de pessimisme : le principe de prudence veut que l'on estime la valeur des stocks au coût de production, non au prix de vente ; que l'on estime la valeur d'un actif à son coût d'acquisition (diminué de l'amortissement), non au cours du jour. Les plus-values latentes ne sont pas comptabilisées dans l'actif. Le comptable estime qu'il faut surtout éviter l'excès d'optimisme, jugé plus dangereux que l'excès de pessimisme.

Il en résulte que lors d'une cession ou d'une fusion de l'entreprise, le calcul de l'actif net (mesure comptable de la valeur de l'entreprise) doit être complété par une expertise afin de réévaluer les actifs. L'écart entre les valeurs ainsi établies et la valorisation comptable peut être important, notamment en ce qui concerne les actifs immatériels (valeur du réseau commercial, du savoir faire, de la marque etc.).

Risques de la comptabilité analytique

La comptabilité analytique vise à fournir aux responsables d'unités décentralisées le moyen de calculer le résultat de leur activité, tout en permettant de calculer le coût de production des divers produits de l'entreprise. Elle implique que des prix de cession interne soient associés aux biens que l'entreprise produit pour sa propre consommation.

Il importe que les " signaux prix " qui sont ainsi envoyés aux responsables opérationnels soient exacts, en ce sens qu'ils induisent des comportements favorables à l'entreprise considérée dans son ensemble. Cette définition de l'exactitude peut guider, mieux qu'un prétendu " réalisme " des coûts, le choix des conventions qui permettent le calcul des prix de cession interne. Ces prix doivent être déterminés par arbitrage, et non par négociation : sinon le risque est fort que l'énergie des

responsables soit accaparée par des négociations qui n'apportent rien à l'entreprise, au détriment du temps qu'ils doivent consacrer aux clients et à la conquête des marchés.

L'organisation d'une entreprise est souvent représentée par un organigramme hiérarchique. On est alors tenté d'imposer à la comptabilité analytique la même présentation : les comptes d'une entité de niveau n doivent s'obtenir par addition des comptes des entités de niveau n - 1, etc. Il en résulte une simplicité reposante pour l'esprit, mais qui nécessite des conventions pour " ventiler " entre entités de niveau n - 1 des dépenses engendrées par le fonctionnement des entités de niveau n et au dessus. Les comptes de chaque entité opérationnelle sont alors lestés de " frais généraux ", " frais de siège " et autres, sur lesquels les décisions du responsable opérationnel n'ont aucun effet.

Aucune logique indiscutable ne préside à ces ventilations, qui comportent toujours une part d'arbitraire. Il est tentant pour un responsable opérationnel, s'il se sent en position de force, de chercher à obtenir une convention qui lui soit favorable. Les discussions, contestations et négociations sont sans fin, et constituent une déperdition d'énergie ruineuse. La meilleure solution, c'est en fait de demander à chaque entité de dégager une marge - différence entre les dépenses et recettes qui lui sont directement imputables -, la marge de niveau n étant la somme des marges du niveau n - 1 diminuée des dépenses induites par le niveau n lui-même. Cependant cette solution est jugée parfois trop compliquée ...

Mesure des coûts

La mesure des coûts de production est l'un des objectifs les plus délicats de l'économie de l'entreprise, qu'il s'agisse des coûts destinés à fonder les prix de vente à des clients ou des " prix d'ordre " utilisés pour les cessions internes à l'entreprise.

Les ingénieurs s'imaginent souvent que le coût de production d'un bien est une donnée aussi " réelle " que son poids. Or il n'en est rien. La définition du coût dépend du point de vue sous lequel on considère ce bien. Le " réalisme " de la mesure du coût est une illusion dangereuse.

Prenons l'exemple du réseau télécom. Il est dimensionné pour acheminer le trafic de l'heure de pointe avec un taux de perte des appels socialement acceptable. Son coût est donc fonction de la définition de l'heure de pointe, de l'estimation du trafic anticipé pendant l'heure de pointe, du taux de perte accepté, et du coût des unités d'œuvre utilisées pour le construire. Le coût d'une communication en dehors de l'heure de pointe est nul, puisque le trafic induit par cette communication n'entre pas dans la fonction de coût du réseau. Par contre le coût d'une communication pendant l'heure de pointe est élevé.

Il s'agit donc d'un coût conventionnel, qui dépend de la définition de l'heure de pointe et du taux de perte. Ajoutons que ce coût se décompose entre raccordement de l'abonné, utilisation des ressources de calcul et de mémoire des commutateurs, et équipements de transmission.

La répartition du coût entre abonnés suppose des péréquations que l'on peut pousser plus ou moins loin, selon que l'on tient compte ou non de la distance de l'abonné à son commutateur de rattachement (qui détermine le coût de son raccordement), de la densité de la zone à laquelle il appartient (qui détermine le coût de la ressource locale de commutation), de la nature de ses appels, etc.

Ainsi, sans aller jusqu'à dire comme Claude Riveline que " le coût d'un bien n'existe pas ", il faut reconnaître que c'est une notion construite, et qui peut l'être de diverses façons selon le but que l'on se donne.

Il faut s'habituer à travailler avec des coûts de production divers : prix de cession interne visant à induire des comportements favorables à l'entreprise, coûts de production destinés aux autorités de tutelle, coûts utilisés pour la détermination des prix, etc. Les degrés de liberté que comporte les

péréquations et ventilations de frais généraux doivent être ici utilisés au mieux.

Centralisation/décentralisation

L'organisation de l'observation statistique doit obéir à des impératifs contradictoires : les exigences formelles de cohérence du programme d'observation, des concepts et méthodes militent en faveur d'une centralisation ; les exigences de pertinence militent en faveur d'une décentralisation, l'adéquation des concepts à l'action étant plus aisée si l'on est proche de cette dernière.

L'appareil statistique public donne ici un exemple intéressant : les statistiques relatives aux entreprises sont collectées par divers ministères (ministère de l'industrie pour les entreprises industrielles, ministère de l'agriculture pour les exploitations agricoles et les industries agroalimentaires, ministère de l'équipement pour les entreprises du BTP, etc.) L'INSEE remplit par rapport à ces ministères une fonction de coordination en ce qui concerne les méthodes, et anime le conseil national de la statistique qui joue un rôle consultatif dans la détermination du programme d'observations. L'INSEE a en outre la responsabilité des travaux de synthèse : calcul de l'indice de la production industrielle et de l'indice des prix à la production, fusion de fichiers entre les données fournies par les enquêtes et les sources fiscales (déclarations BIC des entreprises), alimentation des comptes nationaux. Les questionnaires sont visés à la fois par l'INSEE et par le ministère qui réalise l'enquête. L' " enquête annuelle d'entreprise " est réalisée sous une forme analogue par divers ministères. L'ensemble des sources relatives aux entreprises est organisé au sein d'un " système de statistiques d'entreprises ".

La collecte doit être proche du terrain pour être pertinente et acceptable par ses utilisateurs. Les méthodes doivent être cohérentes, et une coordination est nécessaire.

Cependant articuler décentralisation de la collecte et coordination des méthodes ne va pas de soi. L'un des premiers soucis des entités soumises à une coordination, c'est d'échapper à l'autorité du coordinateur. Elles la contestent en faisant état de leur expérience du terrain, qu'elles seraient seules à connaître à fond, des difficultés pratiques, qu'elles seraient seules à supporter, et qualifient volontiers les exigences de la coordination de " théoriques ", terme auquel elles donnent un acception péjorative. La légitimité des coordinateurs sera toujours contestée. Pour équilibrer ces tendances centrifuges, il importe donc de confier la coordination à des personnes aux compétences indiscutables et qui ont acquis sur le terrain une expérience reconnue.

L'équilibre entre des exigences centrifuges et centripètes se gère dans la durée : imaginer qu'on puisse l'obtenir d'un coup, par le moyen de dispositions judicieuses, serait une illusion. Par contre il importe de rendre cette gestion *possible*.

État des lieux

Les besoins d'information

Les besoins d'information de l'entreprise concernent essentiellement le marché, qu'il s'agisse de la demande ou de l'offre concurrente. La connaissance du marché suppose qu'on le découpe en segments, de façon à évaluer les forces et faiblesses de l'entreprise sur chaque segment. Puis il faut pouvoir positionner l'entreprise par rapport à ses concurrents. Cela suppose que l'on connaisse le marché de chacun de ces concurrents.

L'objet de ces observations est de définir des priorités pour les forces de vente, d'évaluer l'effet des actions marketing et commerciales, enfin d'alimenter le plan à moyen terme.

Les informations doivent être disponibles par produit, secteur et zone géographique. Elles doivent permettre aux entités de se comparer à la moyenne et de négocier leurs objectifs.

Elles concernent le chiffre d'affaires par segment, les consommations en volume, enfin les parts de marché par rapport à la concurrence.

Il faut disposer d'outils facilitant à partir des bases de données les extractions, tris et regroupements, la production de tableaux statistiques et les présentations graphiques. Il faut pouvoir connaître les marges d'erreur autour des évaluations dans le cas des sondages et panels. Le partage de l'information suppose une gestion des droits d'accès et de la confidentialité, impliquant la mise en place de fonctions d'administrateur des bases de données.

Besoins d'études

Esquissons ici ce que pourrait être le programme d'études d'une unité de recherche opérationnelle, destiné à alimenter en analyses les responsables opérationnels du programme, de la tarification et de l'action marketing et commerciale :

Connaître les coûts

Une modélisation, fondée sur les règles de dimensionnement et sur des anticipations de demande réaliste, doit permettre d'évaluer les divers types de coût (coût de production, coût de distribution, coût du système d'information ; coût moyen, marginal, incrémental), et d'évaluer l'impact des politiques de partenariat sur les coûts (économies d'échelle, amélioration du rapport de force avec les fournisseurs).

Passer des coûts aux prix

Un outil de simulation tarifaire s'appuyant sur les caractéristiques du client (CSP, âge, habitat, taille du ménage ; montant de la facturation à l'entreprise, répartition du trafic), pourrait permettre de calculer le coût qui doit être affecté à un client et de l'utiliser pour construire des options tarifaires et des formules personnalisées (par segment de clientèle), ainsi que des cadres contractuels et des formules d'intéressement (fidélisation). Le capital de réflexions et d'expertise accumulé par l'entreprise doit être ici situé dans une perspective d'ensemble.

Il faudrait lier les études marketing à la connaissance des coûts lors du lancement des nouveaux produits, et promouvoir les services qui utilisent le réseau sans pour autant obliger à accroître son dimensionnement.

Passer des prix au trafic

Il faut disposer d'un outil de mesure permettant d'évaluer selon une segmentation pertinente les effets des baisses de prix, des options tarifaires, des mesures ponctuelles de promotion, des contrats, de la fidélisation, et de la concurrence.

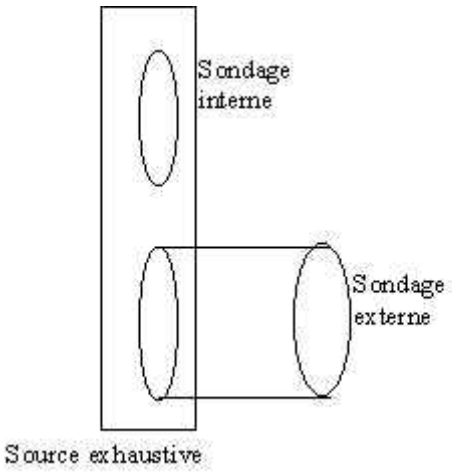
Vers l'approche du client

Il faut classer les clients selon leur contribution au profit actuel et futur de l'entreprise (" scoring ", " life time value "), et construire le système d'information nécessaire au marketing relationnel envers les bons clients. Des segmentations a posteriori, redéfinies fréquemment, doivent permettre des actions ponctuelles adaptées à des niches de clientèle.

Une amélioration technique

Caler les sondages et panels sur des données exhaustives permet de fournir des estimations plus robustes. La confrontation avec d'autres sources externes permet d'enrichir les exploitations, notamment en utilisant la technique de fusion de fichiers. Par exemple la " Sirétisation " des fichiers d'entreprises permet de rapatrier les identifiants de l'INSEE et les informations qui les accompagnent (code APE de l'établissement, classe de taille, code géographique), puis de fusionner les fichiers avec

les autres sources disponibles utilisant le même identifiant.



Articulation des sondages et sources exhaustives

Il faut d'organiser le stockage de l'information de façon à faciliter la constitution de séries chronologiques, outils indispensables pour la modélisation et l'évaluation car l'évolution des données apporte plus d'information que leur niveau instantané. Ce procédé fait d'ailleurs partie de la démarche du datawarehouse.

La représentativité des sondages dans le domaine des entreprises doit être contrôlée grâce à une connaissance des mouvements de l'ensemble des établissements, obtenue sur les sources exhaustives. Les sondages doivent être diversifiés et coordonnés : coordination des échantillons (gestion des inclusions et exclusions), choix des domaines, périodicités des enquêtes et des mises à jour des questions.

L'interprétation doit être enrichie par la modélisation et l'économétrie : les données ne parlent pas d'elles-mêmes, et ne sont interprétables que si elles sont confrontées à des modèles dont l'adéquation est vérifiée par l'économétrie.

Une exploitation enrichie

La production d'information doit être organisée sous plusieurs formes :

- un programme annuel d'exploitation systématique, planifié à l'avance après consultation du comité de statistique ;
- la mise à disposition de bases de données, dont le contrôle d'accès est administré, avec des interfaces permettant de traiter rapidement les demandes simples ;
- des études à la demande réalisées dans un délai limité sur la base d'un devis accepté par le client

Pour définir le programme, les bases de données et le cadre contractuel des études, il est nécessaire de s'organiser pour connaître les besoins.

Le conseil statistique représentatif des utilisateurs examinera les statistiques sur les demandes d'information et d'études et validera le programme annuel d'exploitation ainsi que le design des outils de consultation.

Les utilisateurs disposeront d'une messagerie (ou d'un forum) pour exprimer leurs demandes, poser

des questions et échanger leurs avis. Une hot line téléphonique complétera ce support électronique. La documentation nécessaire (supports de formation aux techniques statistiques, nomenclatures, description des bases de données, documents préparatoires aux réunions du conseil statistique) sera installée sur une base documentaire.

Le maître d'ouvrage des études statistiques ne pourra pas à lui seul réaliser l'intégralité des prestations techniques nécessaires : il devra avoir recours à des sous-traitants pour compléter son offre propre.

Les relations avec les sous-traitants devront être définies de façon à éliminer certains problèmes : ainsi il importe que le maître d'ouvrage soit propriétaire des sources (questionnaires, fichiers informatiques) construites à l'occasion des enquêtes qu'il sous-traite, sous peine d'être dépendant des sous-traitants pour les travaux ultérieurs. Il faudra récupérer les sources des travaux réalisés dans le passé par l'entreprise, au besoin en négociant avec les sous-traitants à l'occasion de la passation de nouveaux contrats.