



La mise en place d'un panel de consommateurs :

méthodes mathématiques

Il arrive qu'un industriel soit mal informé des consommations de ses clients : une situation typique est celle de l'eau et du gaz, puisque les facturations sont souvent annuelles ou semestrielles ; les utilisations mensuelles ou trimestrielles sont inconnues. La mise en place d'un panel de consommateurs permet de remédier à cette difficulté. Nous décrivons ici le principe général qui permet la mise en place d'un tel panel, au profit d'un producteur (noté PROD)

1. Constitution du panel

Le PROD doit d'abord définir une "région cible" : ce peut être la France entière, une région, une ville, à son choix. Il y a deux conditions : que les données de consommation soient disponibles sous forme d'une base de données, et ensuite que le PROD puisse équiper certains consommateurs (ceux qui seront désignés) d'instruments permettant de les relever.

La taille de la région cible est indifférente et n'impacte pas la taille du panel (voir plus bas).

Le PROD et SCM se mettent d'accord sur la précision attendue à partir du panel. Si par exemple on dit 1%, cela signifie que la consommation reconstruite à partir du panel ne doit pas différer de plus de 1% de la consommation effectivement relevée (lorsque les relevés sont faits). Dans nos expériences, dans le domaine de l'eau, le cahier des charges exigeait cette précision, mais nous étions parvenus à 0.3%. Nous ne savons pas d'avance quelle précision est exigible pour une production quelconque.

Le nombre de consommateurs constituant le panel est directement lié à la précision souhaitée pour les estimations. Attention : plus la précision demandée est importante, plus le panel doit être important, et augmenter la précision d'un facteur 2 requiert d'augmenter le panel d'un facteur 4.

Il vaut mieux commencer avec une précision relativement modeste (1% est déjà exigeant) et constituer un panel qui fonctionne ; on peut toujours améliorer la précision ensuite, une fois le système convenablement rodé.

Le PROD communique à la SCM la base de données de l'ensemble des consommateurs de la région cible, au format Excel ou compatible, faisant apparaître les relevés de consommation, les dates de relevés, mais aussi (éventuellement) les dates d'entrée et/ou de sortie (début et fin de l'abonnement).

De manière pratique, cela se présente en général sous la forme de couples "date, relevé". Il en faut au moins deux pour évaluer la consommation moyenne par jour, par personne : c'est la donnée qui nous intéresse. A partir de là, la SCM définit les "tranches de consommation" (elles seront au nombre de dix à quinze). Par expérience, nous pouvons affirmer que :

- Les très petits consommateurs (limite à définir) ne participent pas au panel ;
- Les très gros consommateurs (limite à définir) sont tous inclus dans le panel.

La définition des bornes de chaque tranche (limite inférieure et supérieure) se fait par un procédé mathématique lié à la "dispersion" des consommations à l'intérieur de chaque tranche. Nous choisissons les bornes de manière à ce qu'un panel de taille raisonnable soit représentatif de la tranche.

Nous communiquons alors ces tranches au PROD, pour information. Elles peuvent avoir un intérêt commercial (voir plus bas).

Une fois les tranches de consommation définies, à partir des bases de données de consommateurs, la SCM procède à l'extraction aléatoire d'un certain nombre de personnes dans chaque tranche, au sein de la base de données. Ce nombre est de l'ordre de 200 à 300. Il ne peut être fixé a priori, car il dépend de la largeur de chaque tranche ; l'objectif est de respecter la précision demandée (par exemple 1%).

Ces consommateurs, qui devront être relevés mensuellement ou trimestriellement, constituent le "panel". Nous savons par expérience que les relevés ne sont pas toujours faciles (même les appareils modernes peuvent être défectueux), aussi est-il souhaitable de constituer un panel plus gros que nécessaire (par exemple +20%), pour tenir compte d'éventuelles impossibilités dans les relevés.

Nous communiquons au PROD les identités (numéro de référence) des consommateurs choisis pour constituer le panel, et ceci pour chaque tranche.

2. Extrapolation du panel à l'ensemble des consommateurs

Une fois le panel constitué, le PROD procède aux relevés des consommateurs du panel à la fin de chaque mois ou de chaque trimestre et communique les chiffres à la SCM.

A partir de ces relevés de consommation, la SCM procède aux calculs d'extrapolation à l'ensemble du périmètre en tenant compte :

- de la durée de présence moyenne des consommateurs de chaque tranche pendant le mois ou le trimestre concerné ;

- des consommateurs qui ne seraient plus clients (appelés les "sortants") ;
- du nombre de consommateurs de chaque tranche pendant le trimestre concerné. Ce nombre dépend des "sortants", mais également des nouveaux clients ("entrants").

La SCM communique alors au PROD le résultat de l'extrapolation ; celui-ci s'énonce très simplement : pour chaque tranche, pour la période considérée, la consommation est de tant. L'extrapolation se fait donc tranche par tranche, et la consommation totale est la somme des tranches.

3. Réactualisations et vérifications

Le panel a vocation à être réactualisé chaque année, à cause des entrants et des sortants, mais aussi à cause des consommateurs qui changent de tranche. Notre expérience est que environ 10 à 20% du panel est à reprendre chaque année (sous la forme d'une nouvelle extraction aléatoire), mais les tranches n'ont pas vocation à être modifiées : elles sont en principe déterminées une fois pour toutes, sauf changement considérable dans le périmètre du PROD.

Chaque année, un travail de validation du panel doit être fait : on compare (une fois l'année écoulée) les anticipations faites aux observations réalisées. Si la divergence devait, pour une tranche, être supérieure à 1%, cela signifierait que pour cette tranche-là le panel n'est pas représentatif. Nous ne voyons pas de raison a priori pour que cela se produise, mais il est nécessaire de vérifier.

4. Bénéfices attendus du panel

Outre le fait que le panel est représentatif des consommations de l'ensemble de la population (il est fait pour cela), on peut l'utiliser pour chaque tranche individuellement. On peut par exemple voir si la consommation augmente ou baisse, plus ou moins vite, pour tel type de consommateur. Ceci peut être couplé à d'autres informations (météorologie, environnement, économie, etc.) pour décider d'actions marketing ciblées.

Bien entendu, les consommateurs ont droit à une facture individualisée une fois par semestre, ou plus souvent, mais du point de vue global du chiffre d'affaires, le panel remplace tous les relevés ! Autrement dit, pour évaluer son chiffre d'affaires, le PROD n'aura plus à relever l'ensemble de ses abonnés.

5. Remarques

On remarque que la taille du panel est indépendante de la taille de la population. Elle ne dépend que du profil de population (répartition petits/gros consommateurs) et de la précision voulue. Le profil de population n'est pas connu a priori, mais il se trouve dans la base de données des relevés annuels.

Les probabilités sont une science exacte : si les tranches sont convenablement définies, si l'extraction est convenablement faite (réellement aléatoire), la précision annoncée sera respectée.

On peut essayer de passer d'une région plus petite à une région plus grosse, en se servant d'une formule correctrice fondée sur un historique, mais une telle méthode relève du "bricolage" et ne peut avoir de valeur scientifique.

Un point important est le suivant : pour constituer le panel, on ne cherche pas à mettre en évidence des consommateurs spécifiques qui seraient représentatifs de la population (les journalistes parlent souvent, par exemple, d'un village "qui vote comme la France entière"). Ceci constitue une erreur méthodologique. Il faut au contraire choisir des gens de manière réellement aléatoire. C'est leur diversité qui donne le résultat final, par un effet de moyenne : c'est la loi des grands nombres. On peut évidemment reconstituer ce panel à n'importe quel moment, par une extraction instantanée dans la base de données, la seule difficulté concrète étant que les membres du panel doivent être spécifiquement équipés pour être relevés : il ne suffit pas de les désigner, il faut les interroger.

6. Livres publiés par la SCM

[MPPR] Bernard Beuzamy : Méthodes Probabilistes pour l'étude des phénomènes réels. SCM SA, ISBN 2-9521458-0-6, ISSN 1767-1175, mars 2004 ; seconde édition, juin 2016.

[RDM] Bernard Beuzamy et Olga Zeydina : Méthodes probabilistes pour la reconstruction de données manquantes. SCM SA, ISBN : 2-9521458-2-2, ISSN : 1767 – 1175, avril 2007.

[PIT] Olga Zeydina et Bernard Beuzamy : Probabilistic Information Transfer. SCM SA. ISBN : 978-2-9521458-6-2, ISSN : 1767-1175, mai 2013.

7. Fiches de Compétences

[SCM_QI] La Qualité de l'Information : fiche de compétence de la SCM
http://www.scmsa.eu/fiches/SCM_Qualite_Information.pdf

[SCM_MR] Méthodes robustes : fiche de compétences de la SCM
http://scmsa.eu/fiches/SCM_methodes_robustes.pdf

8. Nos réalisations

- Veolia Environnement (2008 à 2011)
- SAUR (2012 à 2014).