

## Demandes de tris croisés (notes de cours)

Sans traiter des étapes préalables de construction du questionnaire et de constitution de l'échantillon de personnes auquel il est soumis, qu'on espère néanmoins effectuées dans les règles de l'art, on présente ci-après quelques notions relatives à l'examen des réponses obtenues, et plus particulièrement à la demande de *tris croisés*.

### Typologies des variables

Les questionnaires relèvent divers types de variables notant les réponses aux questions posées que l'on peut classer selon plusieurs typologies transversales. La terminologie n'est pas totalement fixée, aussi convient-il d'identifier et de comprendre celle utilisée par l'interlocuteur (ou le logiciel employé).

### Typologie relative à l'étude

On peut distinguer

- Les variables du *descriptif* (ex: sexe, âge, CSP, etc., de la personne interrogée).
- Les variables *propres à l'étude* (ex: appréciation gustative du produit, fréquence d'achat, adjectifs le qualifiant le mieux, prix maximum acceptable, etc.)

### Typologies fondées sur la nature des variables

On distingue d'abord les types suivants

- Les *variables quantitatives*, c'est à dire mesurées par des nombres (ex: âge, nombre d'enfants, salaire mensuel net, note attribuée au produit examiné, quantité achetée, etc.)
- Les *variables logiques, qualitatives* ou à *items*, c'est à dire dont les réponses ne sont pas numériques (ex: CSP, sympathie politique, type d'habitat, appréciation qualitative sur le produit, etc.)

Remarques certaines variables, bien qu'utilisant des nombres comme items, mais sans véritable signification numérique (tels les sept codes de CSP du premier niveau ou encore le code postal du département), sont évidemment de type qualitatif.

Les *classements* sont un type de variable quantitative particulier, qui n'ont pas le caractère additif d'un salaire ou simplement mesurable d'une note attribuée à un produit.

Préférant souvent traiter des variables à items, on transforme aisément une variable quantitative en variable de ce type, soit directement si le nombre de valeurs numériques prises est faible (ex: nombre de véhicules, d'enfants, etc.), soit après découpage en classes par intervalles (ex: salaires).

Une troisième classification, concernant plutôt les variables qualitatives, tient compte du nombre de réponses possibles à la question posée, elle envisage

- Les *variables simples*, ne permettant pas plus d'une réponse (ex: CSP, sexe, chaîne de télévision préférée, meilleur adjectif qualifiant le produit, etc.), et parmi celles-ci les variables *dichotomiques*, c'est à dire à une seule réponse possible parmi deux possibilités (ex: oui/non, homme/femme, actif/inactif, etc.)
- Les *variables multiples*, autorisant plusieurs réponses, parmi une liste fermée a priori ou non (ex: chaînes de télévision préférées, magazines achetés régulièrement, adjectifs décrivant le mieux le produit, etc.)

On désigne enfin parfois par *variables groupées*, ou *variable dimensionnée*, une série de variables simples admettant le même ensemble de réponses, et qu'il est souvent commode de traiter parallèlement (ex: appréciations en "très satisfait", "assez satisfait", "peu satisfait" et "pas du tout satisfait" pour le prix, le SAV, la facilité d'emploi, l'esthétique, le sentiment général, etc., d'un certain produit de consommation).

### **Étude des variables**

Les méthodes utilisées dépendent à la fois du nombre de variables examinées conjointement et de leur type.

### **Étude des variables considérées isolément (analyse univariée)**

- Variables quantitatives on rend compte de la distribution d'une telle variable par la construction de sa *courbe de fréquence* et le calcul d'indicateurs synthétiques de *tendance centrale* (*moyenne, médiane, mode*), de *dispersion* (*variance, écart-type*), etc.
- Variables à items celles-ci sont examinées par *tri à plat*, c'est à dire par le calcul des *fréquences* d'apparition des différents items, des *diagrammes* convenables (*bâtonnets, camemberts*, etc.) permettent de présenter les résultats de manière plus attrayante.

Les calculs de fréquences nécessitent que soit précisée la base retenue : le total des personnes interrogées, ou simplement le total de celles qui ont répondu (les *répondants*), c'est à dire diminué des *non-réponses* ou *rebut*.

Dans le cas de questions à réponses multiples, la somme des fréquences diffère généralement de l'unité, ou 100%.

C'est naturellement l'analyse des variables propres à l'étude qui offre de l'intérêt, celle des variables du descriptif pouvant éventuellement servir à contrôler la structure de l'échantillon.

### **Étude de couples de variables (analyse bivariée)**

On peut résumer les différents cas en un tableau croisé...

Type de variable	Quantitative	Qualitative
Quantitative	Corrélation Régression simple	Moyennes partielles Analyse de la variance
Qualitative	Moyennes partielles Analyse de la variance	Tris croisés Chi-deux

Le résultat d'un tri croisé, portant sur deux variables à items, est donc un *tableau croisé* ou *tableau de contingence*, donnant en un tableau rectangulaire les effectifs ou les fréquences des diverses combinaisons de caractères observées.

Sauf à vouloir vérifier la structure de l'échantillon, on évitera encore de croiser deux variables du descriptif entre elles, on pourra en revanche procéder aux deux autres types de croisement.

### Étude de plus de deux variables simultanément (analyse multivariée)

Selon les cas, on recourra aux méthodes plus complexes que sont la *régression multiple*, *l'analyse en composantes principales*, *l'analyse des correspondances multiples*, *l'analyse discriminante*, etc.

### Tris croisés

On donne sur un exemple fictif simple : une variable descriptive, le sexe, croisée avec une variable propre à l'étude, une préférence touristique, les principales définitions.

Sexe/Destination	Espagne	%	Italie	%	USA	%	Total	%
Femmes	50	20,0	150	60,0	50	20,0	250	55,6
Hommes	25	12,5	75	37,5	100	50,0	200	44,4
Total	75	16,7	225	50,0	150	33,3	450	100,0

Le choix de la variable en ligne et de celle en colonne est évidemment libre, mais peut être guidé par des considérations de lisibilité ou d'esthétique. On recommande généralement une présentation au *format paysage*, comme ci-dessus.

La partie gauche de chaque case indique ici l'effectif, les totalisations de la dernière ligne et de la dernière colonne indiquant les *distributions marginales*, c'est à dire les tris à plat associés à chacune des deux variables considérée isolément.

Nous intéressant à la distribution des préférences touristiques selon le sexe, on a calculé et fait figurer les *pourcentages en ligne*, c'est à dire les *fréquences conditionnelles* des trois destinations selon le sexe. Il est clair qu'ici les *pourcentages en colonne* n'auraient pas d'intérêt, dépendant de la structure par sexe de l'échantillon interrogé.

La comparaison des distributions conditionnelles entre elles, ainsi éventuellement qu'à la distribution marginale, permettent d'apprécier l'influence de la variable condition sur l'autre. Ici, par exemple, on note que les femmes sont plus portées sur l'Italie que les hommes. Des

graphiques convenables permettent comme souvent de transmettre plus aisément l'information contenue dans le tableau.



### **Filtrage**

Il arrive qu'on ne souhaite utiliser qu'une partie de l'échantillon, définie par une condition logique plus ou moins complexe, ou *filtre* (ex: les clients ne connaissant pas encore le produit, les ménagères de moins de 50 ans titulaires d'un diplôme de l'enseignement supérieur non confessionnel, etc.)

### **Redressement, pondérations**

Les difficultés rencontrées lors de la constitution de l'échantillon peuvent éloigner celui-ci de la structure désirée a priori (ex: on souhaitait autant d'hommes que de femmes, mais deux fois moins d'hommes ont répondu à l'enquête). Il est possible de remédier à cela en *redressant* l'échantillon, c'est à dire en affectant à chaque observation une *pondération* convenable. Si les pondérations sont évidentes dans les cas simples, tel celui de l'exemple précédent, il est nécessaire de mettre en œuvre des méthodes de calcul itératives pour redresser un échantillon suivant plusieurs critères simultanément (ex : autant d'hommes que de femmes et deux fois plus de provinciaux que de parisiens).

### **Formulation des demandes**

Lorsque les tris ne sont pas effectués par ceux qui les utiliseront, il est nécessaire d'indiquer avec précision ce qui est demandé, de manière à obtenir exactement ce qui est souhaité. Il est commode de recourir à une formulation standard qu'on peut présenter en tableau.

On en donne le modèle avec deux exemples de demandes.

Tête de ligne	Tête de colonne	Filtre	Calculs demandés	Remarques
Sexe	Destination	-	Effectif % horizontaux	-
Q4b motivation codes 1-8	Q35 âge	Q4a = 2 (acheteurs nouveaux)	Effectifs % verticaux	Regrouper codes 5-6 et 7-8

(18.03.2011)