

Éléments pour l'analyse des résultats et la construction des tableaux et graphiques

- **analyse des tris à plats**
 - Il n'existe pas vraiment de méthode pour l'analyse des tris à plats (c'est-à-dire les résultats bruts). La méthode suivante que je vous propose est le minimum à faire (certains sociologues n'analysent même pas les tris à plats jugeant qu'il n'y a rien à y apprendre ; or, il me semble que nous pouvons déjà dégager quelques éléments qui seront approfondis par les tris croisés ou par des entretiens).
 - Le plus simple consiste à prendre les résultats tableaux par tableaux (pour les questions fermées ; les questions ouvertes doivent être recodées préalablement) et d'entourer le chiffre le plus grand. Faites une phrase à partir de ce chiffre. Ensuite, vous pouvez noter le second chiffre le plus important ainsi que le chiffre le moins grand. Ce dernier peut nous intéresser, dans une analyse sociologique, car il peut marquer des pratiques sociales à la marge ou "anormales". Une fois que vous avez identifié les chiffres qui vous semblent pertinents, résumez le tableau en quelques phrases (2-3 lignes selon l'importance de la question ou du tableau).
 - Ensuite, pour approfondir votre première lecture, vous pouvez observer les autres données (les plus fortes et les moins fortes). Pensez à faire des commentaires sur ces chiffres également.
 - Ayez deux choses à l'esprit quand vous analysez vos tris à plats :
 - D'une part, tous les résultats bruts ne sont pas à analyser car ils ne peuvent rien nous apprendre (cela dépend de la formulation de la question, du choix des items et de la façon dont a été construit le masque de saisie). Avec de l'expérience, vous identifierez plus vite ces questions et vous ne perdrez pas votre temps avec ;
 - D'autre part, tout en analysant vos tris à plats, pensez déjà aux tris croisés que vous allez faire par la suite. Ainsi, pour une question, vous identifiez deux pratiques différentes et vous aimeriez savoir à quel type de population elles correspondent, vous pouvez anticiper en pensant avec quelles questions vous allez les croiser pour approfondir votre analyse.

- **Sélection et analyse des tris croisés.**
 - Dans un tableau de tris croisés, on essaye de déterminer en quoi une variable, un déterminant social, va jouer sur une autre variable. On dira qu'il existe un lien statistique entre deux variables si la répartition des données d'une variable varie de façon significative selon les modalités d'une autre variable : on parlera de dépendance entre les deux variables (et inversement).

Nous allons donc croiser une variable indépendante (âge, sexe, le milieu social, etc.) –qui, selon nos hypothèses va jouer, va être un facteur influant la situation– et une variable dépendante –sensée subir l’action de la première.

- Au moment où vous sélectionnez vos tris croisés (sur le logiciel de traitement statistique ; Sphinx à Rennes 2), pensez toujours à mettre la variable indépendante en ligne et la variable dépendante en colonne. Et les pourcentages doivent être mis en lignes donc en fonction de la variable indépendante. (Pour d’autres consignes sur la construction des tableaux, cf. ci-dessous). Ainsi, le total de chacune de vos lignes fera 100% dans vos tableaux de tris croisés.
- Pour l’analyse des tableaux de tris croisés, je vous propose de suivre certaines étapes proposées par François De Singly¹ que j’ai adapté en fonction de ma propre méthode et expérience professionnelle. Tout d’abord, faites une première lecture en colonne. On souligne les chiffres les plus importants mais aussi les moins importants. Le nombre en soi importe peu. C’est sa comparaison que nous allons faire avec les autres (que cela soit ceux des autres variables mais aussi ceux de la ligne "Total") qui nous importe pour le moment.
- Dans la lecture du tableau, les catégories de la variable indépendante dont les pourcentages sont similaires (à plus ou moins cinq points) seront regroupées. C’est une règle d’équivalence que nous adopterons. Ensuite, nous pourrions les distinguer dans notre analyse (même si, compte tenu de ce faible écart, ces différences ne sont pas vraiment pertinentes). En règle générale, dans ma pratique professionnelle, j’écarte les données dont l’écart est inférieur à 5 points par rapport à la ligne Total.
Cette lecture proposée par De Singly semble intéressante car la lecture des pourcentages en lignes gomme certains particularismes. Ainsi, dans une lecture par ligne, la variable la plus commune va forcément ressortir. Par contre, une lecture par colonne des pourcentages va nous permettre de mettre en valeur les pratiques rares, atypiques, les minorités qui peuvent nous être utiles dans notre analyse ou, par la suite, dans la phase des entretiens.
- Une fois les deux étapes précédentes réalisées, il est temps de commencer à rédiger deux ou trois phrases résumant les grandes lignes du tableau. Résumer les grandes idées en faisant des comparaisons. L’idée est de faire ressortir les premiers éléments du tableau avant de rentrer dans l’analyse fine.
- Après, nous passons à la deuxième lecture qui se fait ligne par ligne mais toujours en comparant les lignes du tableau. Cela permet d’observer s’il y a des pratiques minoritaires –qui peuvent être illégitimes ou de distinction– lorsque les pourcentages sont dissemblables ou, au contraire, si les avis sont assez semblables d’une catégorie à une autre (pour la variable indépendante)

¹ François De Singly, *L’enquête et ses méthodes : le questionnaire*, Paris : Nathan Université (Coll. « 128 »), 1992, 128p

- Attention dans tous les commentaires à ne pas faire de généralités. Il est nécessaire de préciser toujours comment l'échantillon a été construit et donc en quoi il est représentatif de la population totale. Le risque est d'avoir une sur-représentation (ou une sous-représentation) dans notre échantillon et donc de fausser les résultats. Il faut donc prendre toujours des précautions dans l'énoncé des résultats et dans l'analyse. Il faut tenir compte aussi du contexte dans lequel l'enquête a été réalisée. Ce contexte peut jouer dans les résultats. Toutes ces précisions sont donc à indiquer dans le rapport. De même, il faut préciser si les questions sont à choix unique ou multiple. Dans ce dernier cas, nous pouvons avoir des effectifs supérieurs à notre échantillon. Il faut également préciser les modifications qui ont été apportées au tableau (suppression de lignes ou de colonnes telles que les non-réponses par exemple).
- Le test du khi deux permet de mesurer la dépendance entre deux variables. Vous pouvez trouver sa formule mathématique dans le livre de Gilles Grenon et Suzanne Viau (p.196²) : $\chi^2 = \sum (f_o - f_e)^2 / f_e$. Ce test sert à mesurer l'écart entre des fréquences observées (f_o) et des fréquences espérées (f_e). Le résultat du khi deux est toujours positif ou nul. S'il est égal à 0, cela signifie que les variables sont indépendantes. Ainsi, plus il est important, plus les variables sont dépendantes. Pour déterminer si le résultat du khi deux est significatif et s'il y a un lien entre les deux variables dans la population, il faut déterminer le nombre de degrés de liberté, noté v (prononcé nu), et la valeur critique. Pour calculer le nombre de degré de liberté, il faut utiliser la formule suivante : $v = (\text{nombre de modalités de la variable A} - 1) * (\text{nombre de modalités de la variable B} - 1)$. En fonction du résultat, on va déterminer la valeur critique. Cette valeur critique est une valeur théorique créée à partir du modèle de Karl Pearson (en se reportant au tableau suivant, on regarde quelle est la valeur critique en face du résultat de v) :

V	Valeur critique	V	Valeur critique
1	3.84	16	26.3
2	5.99	17	27.59
3	7.81	18	28.87
4	9.49	19	30.14
5	11.07	20	31.41
6	12.59	21	32.67
7	14.07	22	33.92
8	15.51	23	35.17
9	16.92	24	36.42
10	18.31	25	37.65
11	19.68	26	38.89
12	21.03	27	40.11
13	22.36	28	41.34
14	23.68	29	42.56
15	25	30	43.77

² Gilles Grenon, Suzanne Viau, Méthodes quantitatives en sciences humaines, Volume 1 : de l'échantillon vers la population, 2^e édition, Paris : Gaëtan Morin, 1999

Si le khi deux est supérieur à la valeur critique, on peut écarter l'hypothèse d'indépendance entre les deux variables dans la population (pour plus d'éléments, cf. Grenon et Viau pp190-207). Par contre, si le résultat est inférieur, nous ne pouvons pas écarter l'hypothèse de l'indépendance des deux variables dans la population (le lien que nous avons pu trouver dans l'échantillon peut être alors le fruit du hasard).

- **Pour la construction et la présentation des tableaux :**

- Numérotez les tableaux (en chiffres arabes) ;
- Donner un titre assez clair (cela permet de souligner la donnée importante ; il en va de même pour les graphiques) ;
- Intitulez les colonnes avec des noms assez clairs : par exemple, celle de gauche par le nom de la variable, celle de droite par "pourcentage" ;
- Indiquer une ligne "Total". Au cas où la somme des pourcentages ne serait pas égale à 100%, indiquer la raison en note de bas de page (pour les tableaux des tris croisés, il faut également une colonne "Total").
- Avoir, dans la plupart des cas pour les tableaux de tris à plats, une colonne présentant le nombre de cas (cela permet de donner une idée du nombre de réponses sur lesquelles ont été construits les pourcentages ; en outre, nous travaillons avec des humains, donc, les couper sous forme de pourcentages...) ;
- Arrondir tous les pourcentages de la même manière avec, au maximum, deux chiffres après la virgule ;
- Ne pas mettre nécessairement de symbole "%" dans la colonne des pourcentages étant donné que l'on sait, en fonction du nom de la colonne, que ce sont des pourcentages (cela est valable pour les tableaux de tris à plats ; pour les tableaux de tris croisés, il vaut mieux laisser le symbole %) ;
- Sous le tableau ou en note de bas de page, notamment pour les tableaux des tris croisés, mettre une phrase de commentaire pour dire comment on lit ce tableau (par exemple, sur 100 personnes ayant telle particularité, X ont telle opinion). Précisez aussi si nous avons affaire à une sous-population, s'il y a eu des regroupements de modalités, si vous avez supprimé les non-réponses, etc.
- Pour les tableaux des tris croisés, le nombre de cases ne doit jamais dépasser le dixième de la population enquêtée (soit 50 cases pour un échantillon de 500 personnes). Dans les tris croisés, il faudra tenir compte de cela car il sera peut-être nécessaire de regrouper des catégories. Il faut donc réfléchir aussi sur l'utilité de certains croisements (notamment avec les questions ouvertes et les réponses aux items "Autres").

- **Quelques règles sur la présentation des diagrammes en bâtons :**

- Pour les présentations sous forme de graphiques, au-delà de 7 valeurs, pour une question de lisibilité, il vaut mieux privilégier le diagramme en bâtons (verticaux ou horizontaux) au diagramme circulaire (camembert ou anneau).
- Nommez les diagrammes par "figure" (ou "graphique") ;
- Donnez un titre clair et présentant l'idée principale du diagramme ;

- Utiliser les mêmes échelles pour des diagrammes présentant des données semblables (par exemple pour les questions sur les modes de transport utilisés selon différents moments).