

DES GROUPES D'INDIVIDUS DÉCRITS PAR UNE VARIABLE QUALITATIVE

COMPARER ET DÉCRIRE LES GROUPES

Mots-clés : Variables qualitatives ; Tableau de contingence ; Analyse globale ; analyses locales ; Φ^2 ; χ^2 ; Profils et profil moyen ; Attractions-Répulsions ; Écart relatif ; Taux de liaison ; Densités ; Contributions au Φ^2 ;

Ce document indique comment obtenir les différents résultats avec le logiciel SES-Pegase (**version 6.1**). Cependant, l'objectif est qu'il puisse être utilisé comme guide méthodologique et d'interprétation, quel que soit le logiciel utilisé.

ATTENTION : Mettre les commandes de la version 7.

TYPE DES DONNÉES ANALYSÉES

Nous présenterons l'analyse d'un dossier particulier, le dossier SYMPA. Mais cette analyse s'applique à toutes données de la forme suivante (Tableau 1) :

Tableau 1 : Structure du tableau des données individuelles

| INDIV | GRP | REP |
|-------|-------|-------|
| i1 | 1 | 3 |
| i2 | 1 | 1 |
| i3 | 3 | 4 |
| i4 | 2 | 1 |
| i5 | 2 | 2 |
| i6 | 1 | 2 |
| i7 | 3 | 3 |
| i8 | 2 | 4 |
| (...) | (...) | (...) |

On a recueilli des données sur des individus (i1, i2...). Ces individus peuvent être, des personnes, des pays, des animaux, des voitures...

Parmi les données recueillies, on a :

- l'appartenance des individus à des groupes : la catégorie socioprofessionnelle de la personne, le constructeur de la voiture, le continent où se sont déroulées les observations...
- une variable qualitative observée sur les individus : la réponse à une question, un comportement observé...

Les deux variables sont *qualitatives*, même si elles sont codées numériquement comme dans le Tableau 1. Cela exclut donc des variables de type temps, effectifs, salaires, températures, notes...

Pour les données traitées ici, la variable observée comprend *plus de deux modalités* (plus de deux types de comportement observé, plus de deux réponses possibles à un questionnaire...). Dans le cas contraire, les analyses à mettre en œuvre sont différentes et sont présentées ailleurs, à partir d'un autre exemple.

Tableau Individus × Variables ou Tableau de contingence ?

Il arrive souvent que les données se présentent, non pas sous la forme d'un tableau Individus × Variables (cf. Tableau 1) mais sous la forme d'un tableau de contingence (cf. Tableau 2) qui représente la répartition de 75 personnes interrogées selon leur groupe et leur réponse.

Tableau 2 : Tableau de contingence

| | Rep1 | Rep2 | Rep3 | Rep4 |
|------|------|------|------|------|
| Grp1 | 10 | 5 | 0 | 10 |
| Grp2 | 0 | 3 | 5 | 2 |
| Grp3 | 20 | 10 | 0 | 10 |

Comment saisir les données quand les données se présentent sous cette forme ?

Pour l'exemple ci-dessus, on pourrait reconstituer un tableau de 75 lignes et deux colonnes. Le travail risque d'être quelque peu fastidieux. On peut noter qu'un grand nombre de lignes seraient identiques. On aurait par exemple 10 lignes avec GRP=1 et REP=1.

Une solution consiste à créer un tableau plus synthétique de 12 lignes (cf. Tableau 3) correspondant aux 12 combinaisons des deux variables (3 x 4). Une colonne EFFECTIF ou POIDS indique l'effectif de chacune de ces 12 combinaisons. Ainsi, la première ligne indique que 10 individus du groupe 1 ont donné la réponse 1. La deuxième ligne y indique que 5 individus du groupe 1 ont répondu 2...

Tableau 3 : Tableau pondéré

| | GRP | REP | POIDS |
|-------|-------|-------|-------|
| 1 | 1 | 1 | 10 |
| 2 | 1 | 2 | 5 |
| 3 | 1 | 3 | 0 |
| 4 | 1 | 4 | 10 |
| 5 | 2 | 1 | 0 |
| (...) | (...) | (...) | (...) |
| 11 | 3 | 3 | 0 |
| 12 | 3 | 4 | 10 |

Questions

Les objectifs consisteront :

1. à chercher s'il existe des différences entre les groupes du point de vue de la variable observée,
2. à décrire les particularités de chaque groupe, relativement aux autres,
3. à résumer, par une représentation factorielle des données, les principales différences entre groupes,
4. à chercher à regrouper les groupes en classes homogènes.

UN EXEMPLE : LE DOSSIER SYMPA¹

On a interrogé 1 200 personnes. Ces personnes appartiennent à huit groupes professionnels (GRP) différents : Paysans, Ouvriers, Vendeurs, Commerçants, Employés, Techniciens, Universitaires, Professions Libérales.

Elles avaient à choisir la qualité (QUAL) la plus appropriée pour qualifier un homme sympathique, parmi les 9 *qualités* suivantes : Sérieux, Généreux, Gai, Honnête, Intelligent, Serviable, Courageux, Compréhensif, Discret.

Questions

1. Les qualités citées pour qualifier un homme sympathique (QUAL) diffèrent-elles selon le groupe professionnel (GRP) ?
2. Si oui, quelles sont les particularités de chaque groupe ?

Type et statut des variables

On met ici en relation deux *variables qualitatives* :

- le groupe professionnel (GRP), à 8 modalités,
- la qualité citée (QUAL), à 9 modalités.

La première variable (GRP) a le statut de variable indépendante (VI) au sens large, ou de variable prédictrice. La deuxième variable (QUAL) est la variable dépendante (VD) ou variable à prédire.

¹ Source : D'après Maisonneuve, *Recherches diachroniques sur la représentation sociale*, 1978.

Les données présentées ici sont adaptées. Les données originales sont présentées, analysées et commentées dans Rouanet et B. Le Roux (1993), p.241-247, et Rouanet et B. Le Roux (1995) p. 87-88 et 98-99.

ANALYSES UNIVARIÉES (PAR VARIABLE)

Avant de répondre à la question posée, c'est-à-dire comparer les groupes professionnels du point de vue des qualités citées (analyse bivariée), on commence par analyser séparément chacune des deux variables, groupes et qualités (analyses univariées). Analyser une variable qualitative consiste à examiner la répartition, en effectifs et en pourcentages, des différentes modalités et à s'intéresser en particulier aux modalités les plus fréquentes et les plus faibles.


Lancer SESAnalyse
Menu Fichier - Ouvrir un dossier SES (*.SES)
Sélectionner le dossier SYMPA.SES.

Les données se présentent sous la forme d'un tableau pondéré de 72 lignes (8 groupes x 9 qualités). La colonne POIDS indique les nombres d'individus dans chacune des 72 combinaisons.

Les groupes professionnels (GRP)

Menu Nouvelle analyse
- Sélectionner la variable GRP comme variable dépendante (VD).
- Cliquer sur .

Menu Analyser
- Analyser la distribution
- Distribution en effectifs et en pourcentages

Cliquer sur l'icône graphique 
Sélectionner Diagramme à secteurs

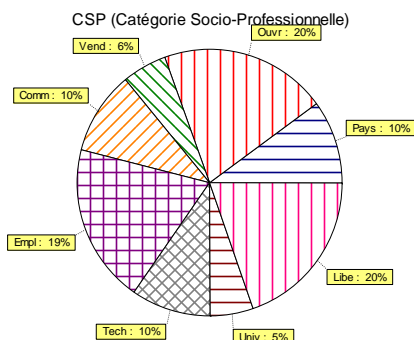


Figure 1 : Répartition en pourcentages, des groupes professionnels (diagramme à secteurs)

Tableau 4 : Répartition, en effectifs et en pourcentages, des groupes

| Groupes | Pays | Ouvr | Vend | Comm | Empl | Tech | Univ | Libe | Total |
|----------|------|------|------|------|------|------|------|------|-------|
| Effectif | 123 | 243 | 66 | 120 | 231 | 117 | 63 | 237 | 1200 |
| % | 10% | 20% | 6% | 10% | 19% | 10% | 5% | 20% | 100 % |

Pour connaître la valeur plus précise (non arrondie) d'un pourcentage : positionner le curseur de la souris sur le pourcentage auquel on s'intéresse.

Le tableau précédent nous indique par exemple que, sur un total de 1200 personnes, 123 sont paysans. Ils représentent environ 10% de l'effectif total (123 / 1200).

On constate que trois groupes professionnels sont plus fortement représentés que les autres dans cet échantillon : les ouvriers (20%), les professions libérales (20%) et les employés (19%).

A l'inverse, deux groupes sont peu représentés² : les vendeurs (6%) et les universitaires (5%).

Les autres groupes (paysans, commerçants, techniciens) représentent chacun 10% de l'échantillon.

² On pourrait s'interroger sur la représentativité de ces différents groupes par rapport à la population générale en comparant ces pourcentages observés aux pourcentages de ces groupes dans la population générale au moment de l'enquête.

Les qualités citées (QUAL)

Menu Nouvelle analyse

- Sélectionner la variable QUAL comme variable dépendante (VD).
- Cliquer sur .

Menu Analyser

- Analyser la distribution
- Distribution en effectifs et en pourcentages

Cliquer sur l'icône graphique



Sélectionner Diagramme à secteurs

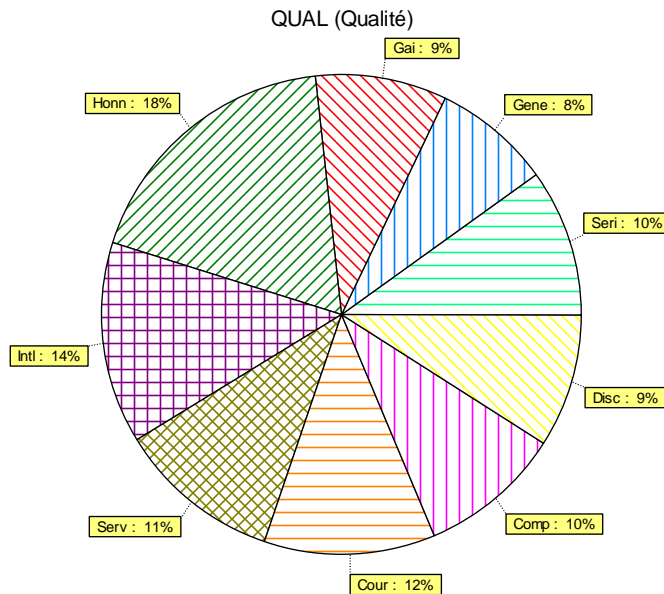


Figure 2 : Répartition en pourcentages des qualités citées (QUAL)

Tableau 5 : Répartition, en effectifs et en pourcentages, des qualités citées (QUAL)

| QUAL | Seri | Gene | Gai | Honn | Intl | Serv | Cour | Comp | Disc | Total |
|-----------------|------|------|-----|------|------|------|------|------|------|-------------|
| Effectif | 120 | 95 | 106 | 220 | 163 | 133 | 138 | 117 | 108 | 1200 |
| % | 10% | 8% | 9% | 18% | 14% | 11% | 12% | 10% | 9% | 100% |

On constate que, tous groupes professionnels confondus :

- les qualités les plus souvent citées sont l'honnêteté (18%), puis l'intelligence (14%).
- les qualités les moins souvent citées sont la générosité (8%), la gaité (9%) et la discrétion (9%).

Cette répartition correspond au *profil moyen* des personnes interrogées (tous groupes professionnels confondus).

Rédiger un compte rendu de l'analyse

Pour cet échantillon de 1 200 personnes adultes,
on constate (cf. Figure 1) que :

- trois groupes professionnels sont plus fortement représentés que les autres dans cet échantillon : les ouvriers (20%), les professions libérales (20%) et les employés (19%).
 - deux groupes sont peu représentés³ : les vendeurs (6%) et les universitaires (5%).
- Les autres groupes (paysans, commerçants, techniciens) représentent chacun 10% de l'échantillon.

< insérer Figure 1 >

On constate également (cf. Figure 2) que,
tous groupes professionnels confondus :

- les qualités les plus souvent citées sont l'honnêteté (18%), puis l'intelligence (14%).
- les qualités les moins souvent citées sont la générosité (8%), la gaité (9%) et la discrétion (9%).

< insérer Figure 2 >

DIFFÉRENCES ENTRE LES GROUPES ?

Cette partie de l'analyse vise à répondre à deux questions :

1. existe-t-il des différences entre les groupes du point de vue des qualités citées pour qualifier un homme sympathique ?
2. peut-on qualifier ces différences de faibles ou d'importantes ?

De plus, on répond à chacune de ces deux questions en deux étapes :

a/ qu'en est-il dans l'échantillon (l'ensemble des personnes interrogées) ?

b/ peut-on généraliser la conclusion à la population (l'ensemble des personnes d'où provient cet échantillon) ?

```
SES-Pegase
Menu Nouvelle analyse
- Sélectionner la variable GRP comme VI et QUAL comme VD.
- Cliquer sur .
Menu Analyser
- Analyser la liaison GRP*QUAL
- Écart global à l'indépendance
- Coefficients de contingence
- Cliquer sur le bouton 
```

Existence de différences entre les groupes ?

a/ dans l'échantillon

Tableau 6 : Carré moyen de contingence (Φ^2)

| | |
|------------------|------|
| Phi ² | 0.13 |
| Phi | 0.36 |

Pour cet échantillon de 1 200 personnes adultes, de groupes professionnels différents,
du point de vue des qualités citées pour qualifier un homme sympathique,
il existe des différences entre les groupes professionnels ($\Phi^2 = 0.13, > 0$).

b/ dans la population

Tableau 7 : Test du χ^2

| | |
|------------------|--------|
| n | 1200 |
| ddl | 56 |
| Khi ² | 155.78 |
| p | <.0001 |

³ On pourrait s'interroger sur la représentativité de ces différents groupes par rapport à la population générale en comparant ces pourcentages observés aux pourcentages de ces groupes dans la population générale au moment de l'enquête.

Le test du Khi^2 est déclaré « significatif », car p est inférieur $< .05$. Ce résultat permet de conclure que, dans la population d'où provient cet échantillon de 1200 personnes, il existe des différences entre les groupes professionnels ($Khi^2 = 155.8$, $ddl = 56$, $p < .0001$).

Si le test n'avait pas été significatif (si $p > .05$) on n'aurait conclu... qu'on ne peut pas conclure à l'existence de différences dans la population.

Ampleur des différences entre les groupes ?

```
Menu Analyser
- Analyser la liaison GRP*QUAL
- Écart global à l'indépendance
- Coefficients de contingence
```

a/ dans l'échantillon

On cherche ici à mesurer l'ampleur de la liaison globale entre les deux variables. Peut-on qualifier la liaison entre ces deux variables de faible/modérée/forte ?

Tableau 8 : V^2 et V de Cramer

| | |
|-------|------|
| V^2 | 0.02 |
| V | 0.14 |

On situe la valeur du V de Cramer par rapport aux deux valeurs repères (.20 et .40)⁴

Il apparaît alors que, pour cet échantillon de 1 200 personnes adultes, de groupes professionnels différents, du point de vue des qualités citées pour qualifier un homme sympathique, les différences entre les groupes apparaissent faibles (V de Cramer = $0.14 < 0.20$).

b/ dans la population

A ce stade, il faudrait calculer un intervalle de confiance sur le V de Cramer pour pouvoir conclure si, comme dans l'échantillon, les différences entre les groupes sont faibles dans la population.

Les principes de calcul d'un tel intervalle sont présentés par exemple dans Smithson, M. (2003). Cette procédure n'est pas disponible dans la version actuelle (6.1) de SES-Pegase.

Rédiger un compte rendu de l'analyse

Pour l'échantillon de 1 200 personnes adultes, de groupes professionnels différents, il existe des différences entre les groupes professionnels ($Phi^2 = 0.13$, > 0) du point de vue des qualités citées pour qualifier un homme sympathique, Toutefois ces différences apparaissent faibles⁵ (V de Cramer = $0.14 < 0.20$).

Il semble que,

dans la population d'où provient cet échantillon de 1200 personnes,

il existe bien des différences entre les groupes professionnels ($Khi^2 = 155.8$, $ddl = 56$, $p < .0001$).

Toutefois, les procédures disponibles ne permettent pas de se prononcer sur l'importance de ces différences dans la population.

⁴ Le V de Cramer peut être considéré comme une corrélation. On prend donc les mêmes valeurs repères pour qualifier les différences de faible (si $< .20$) ou forte (si $> .40$)

⁵ Le V de Cramer peut être considéré comme une corrélation, au sens large. Il est donc possible de prendre les mêmes valeurs repères pour qualifier les différences de faible (si $< .20$) ou forte (si $> .40$)

DÉCRIRE LES GROUPES

Les indices précédents nous disent qu'il existe une différence entre les groupes et que, dans l'échantillon, cette différence est faible. Mais ces indices ne nous disent pas quelle est la nature de ces différences. En particulier, même si les différences entre groupes sont globalement faibles, il n'est pas impossible que, pour certains groupes ou certaines qualités citées, les différences soient importantes. Pour cela il faut analyser, case par case, le contenu de ce tableau.

On parle d'*analyses globales* pour qualifier les analyses précédentes, et d'*analyses locales* pour les analyses, case par case, qui suivent.

Effectifs observés

SES-Pegase
Menu Analyser
- Voir la distribution croisée GRP*QUAL
- Effectifs conjoints (observés)

Tableau 9 : Répartition des 1 200 réponses selon le groupe professionnel et la qualité citée.

| | Seri | Gene | Gai | Honn | Intl | Serv | Cour | Comp | Disc | Total |
|--------------|------------|-----------|------------|------------|------------|------------|------------|------------|------------|-------------|
| Pays | 20 | 9 | 9 | 27 | 10 | 16 | 20 | 4 | 8 | 123 |
| Ouvr | 42 | 10 | 22 | 51 | 18 | 28 | 38 | 12 | 22 | 243 |
| Vend | 11 | 2 | 5 | 14 | 8 | 7 | 5 | 8 | 6 | 66 |
| Comm | 8 | 9 | 12 | 23 | 14 | 16 | 14 | 12 | 12 | 120 |
| Empl | 19 | 10 | 16 | 52 | 32 | 25 | 22 | 25 | 30 | 231 |
| Tech | 10 | 5 | 12 | 23 | 20 | 13 | 11 | 13 | 10 | 117 |
| Univ | 2 | 8 | 7 | 6 | 15 | 6 | 6 | 9 | 4 | 63 |
| Libe | 8 | 42 | 23 | 24 | 46 | 22 | 22 | 34 | 16 | 237 |
| Total | 120 | 95 | 106 | 220 | 163 | 133 | 138 | 117 | 108 | 1200 |


Ce tableau nous indique, par exemple, que les paysans citent Sérieux 20 fois (sur un total de 123 réponses) les ouvriers citent sérieux 42 fois (sur un total de 243 réponses).

Synonymes : tableau de contingence / tableau de correspondances / tableau de dépendance / tri croisé.

Il est important de construire le tableau précédent qui visualise l'ensemble des réponses et les distributions marginales (cf. dernière ligne et dernière colonne). Toutefois, pour l'analyse, on construira d'autres tableaux dérivés. En effet, les effectifs observés dans ce tableau ne sont guère comparables. Par exemple, sérieux a été cité deux fois plus souvent par les ouvriers (42 fois) que par les paysans (20 fois). Mais (cf. dernière colonne) les ouvriers étant deux fois plus nombreux (243) que les paysans (123) il n'est pas possible d'en conclure que sérieux est plus souvent cité par les ouvriers que par les paysans. En effet, si on rapporte les effectifs observés (20 et 42) aux totaux de la ligne correspondante (123 et 243) on constate que cette qualité a été citée approximativement dans les mêmes proportions par les paysans ($20/123 = 16\%$) et par les ouvriers ($42/243 = 17\%$).

Profils des groupes et profil-moyen

Les analyses qui suivent (comparaison et description des groupes, analyse factorielle, classification automatique) s'appuieront essentiellement sur les *profils* (profils des groupes et profil-moyen).

SES-Pegase
Menu Analyser
- Comparer les GRP
- Profils (en %) des GRP.
- Cliquer sur l'icône 

Des groupes d'individus décrits par une variable qualitative
Le dossier SYMPA
Comparer et décrire les groupes

Tableau 10 : Profils des groupes professionnels selon les qualités citées.

| | Seri | Gene | Gai | Honn | Intl | Serv | Cour | Comp | Disc | Total |
|----------------|------------|-----------|-----------|------------|------------|------------|------------|------------|-----------|-------------|
| Pays | 16% | 7% | 7% | 22% | 8% | 13% | 16% | 3% | 7% | 100% |
| Ouvr | 17% | 4% | 9% | 21% | 7% | 12% | 16% | 5% | 9% | 100% |
| Vend | 17% | 3% | 8% | 21% | 12% | 11% | 8% | 12% | 9% | 100% |
| Comm | 7% | 7% | 10% | 19% | 12% | 13% | 12% | 10% | 10% | 100% |
| Empl | 8% | 4% | 7% | 23% | 14% | 11% | 10% | 11% | 13% | 100% |
| Tech | 9% | 4% | 10% | 20% | 17% | 11% | 9% | 11% | 9% | 100% |
| Univ | 3% | 13% | 11% | 10% | 24% | 10% | 10% | 14% | 6% | 100% |
| Libe | 3% | 18% | 10% | 10% | 19% | 9% | 9% | 14% | 7% | 100% |
| Moyenne | 10% | 8% | 9% | 18% | 14% | 11% | 12% | 10% | 9% | 100% |

Ce tableau se lit ligne par ligne. Chaque ligne indique la répartition, en pourcentages, des qualités citées. Par exemple, les paysans citent sérieux dans 16% des cas (20/123, cf. Tableau 9), généreux dans 7% des cas (9/123)... La dernière ligne indique la répartition en pourcentages des qualités citées, tous groupes confondus (ces pourcentages sont ceux, déjà commentés, du Tableau 5). Le total des pourcentages de chaque ligne fait 100%.

L'ensemble des pourcentages d'une ligne constitue le *profil* du groupe professionnel. La dernière ligne constitue le *profil moyen* des personnes interrogées.

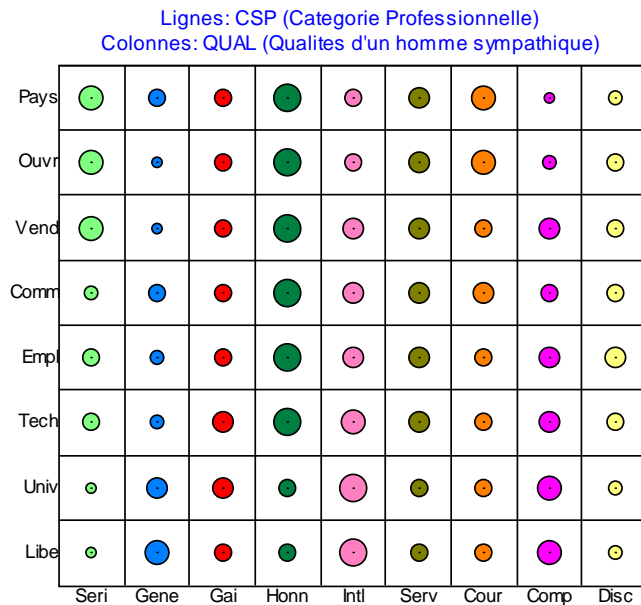


Figure 3 : Représentation graphique des profils des groupes

Sur ce graphique⁶ la taille des points est proportionnelle à la fréquence des qualités citées pour un groupe donné (cf. Tableau 10).

Le profil d'un groupe professionnel correspond à une ligne du tableau. Or une couleur est associée à une colonne du tableau (une qualité) et non pas à une ligne (un groupe). Pourquoi ? Parce que la comparaison des profils se fait colonne par colonne. Ainsi, pour comparer le profil des Paysans et le profil des Ouvriers, on va comparer les deux fréquences de Sérieux (cf. colonne 1) manifestement proches, puis les deux fréquences de Généreux (cf. colonne 2) légèrement plus élevée chez les paysans (...) les deux fréquences de Discret (dernière colonne) plus forte chez les ouvriers.

L'analyse va consister à comparer les profils, soit entre groupes, soit entre chaque groupe et le profil moyen.

⁶ Si des couleurs différentes ont été associées aux différentes qualités (cf. SES-Data) celles-ci apparaissent dans les graphiques comme celui-ci et en facilitent la lecture.

La comparaison des profils des groupes entre eux (cf. par exemple la première colonne du Tableau 10 et de la Figure 3) montre que :

- sérieux est cité plus souvent par les paysans, les ouvriers et les vendeurs que par les autres groupes
- généreux est plus souvent cité par les universitaires et les professions libérales (cf. deuxième colonne du Tableau 10 et de la Figure 3).

Une autre approche pour l'analyse de ces profils consiste à comparer le profil de chaque groupe au profil moyen (cf. ligne « Moyenne » du Tableau 10).

Ainsi, pour les professions libérales (cf. dernière ligne du Tableau 10), on constate que :

- certaines qualités sont peu citées, relativement à l'ensemble des personnes. C'est le cas, par exemple, de Sérieux (3% comparé à 10%) Honnête (10% comparé à 18%).
- certaines qualités sont souvent citées, relativement à l'ensemble des personnes. C'est le cas, par exemple, de Généreux (18% comparé à 8%), Intelligent (19% comparé à 14%).

La comparaison systématique des profils des groupes au profil moyen, est conduite en deux temps :

1. on analyse le sens des différences (cf. les attractions-répulsions)
2. on analyse l'ampleur des différences (cf. les écarts relatifs et les contributions au Φ^2)

Sens des écarts à la moyenne ?

a/ dans l'échantillon (description)

SES-Pégase

Menu Analyser

- Analyser la liaison GRP*QUAL

- Écarts locaux à l'indépendance - Attractions-répulsions

- Si le système de l'ordinateur est un système 32 bits, cliquer sur [Inférer](#)

Tableau 11 : Attractions-Répulsions entre groupes et qualités citées.

| | Seri | Gene | Gai | Honn | Intl | Serv | Cour | Comp | Disc |
|-------------|------|------|-----|------|------|------|------|------|------|
| Pays | + | = | - | + | - | + | + | - | - |
| Ouvr | + | - | = | + | - | + | + | - | = |
| Vend | + | - | = | + | = | = | - | + | = |
| Comm | - | = | + | = | - | + | = | = | + |
| Empl | - | - | - | + | = | = | - | + | + |
| Tech | - | - | + | + | + | = | - | + | = |
| Univ | - | + | + | - | + | = | - | + | - |
| Libe | - | + | + | - | + | - | - | + | - |

Un signe positif, par exemple pour Paysan-Sérieux, signifie que les paysans citent sérieux plus souvent que l'ensemble des personnes interrogées (16% > 10% : surreprésentation ou attractions).

Un signe égal, pour Paysans-Généreux par exemple, signifie que les paysans citent cette qualité avec approximativement la même fréquence que l'ensemble des personnes interrogées (7% ≈ 8%).

Un signe négatif, pour Paysans-Gai par exemple, signifie que les paysans citent cette qualité moins souvent que l'ensemble des personnes interrogées (7% < 9% : sous-représentation ou répulsion).

b/ dans la population (inférence)

Pour obtenir ce résultat SES-Pégase fait appel au logiciel BAYACT
(cf. Bernard, J.M., 1998, pp. 159-226).

Tableau 12 : Inférence sur les attractions-répulsions entre groupes et qualités citées.

| | Seri | Gene | Gai | Honn | Intl | Serv | Cour | Comp | Disc |
|-------------|---------|---------|-----|---------|---------|------|---------|---------|---------|
| Pays | +(0.99) | ? | ? | ? | -(0.96) | ? | ? | -(1.00) | ? |
| Ouvr | +(1.00) | -(0.99) | ? | ? | -(1.00) | ? | +(0.98) | -(1.00) | ? |
| Vend | ? | ? | ? | ? | ? | ? | ? | ? | ? |
| Comm | ? | ? | ? | ? | ? | ? | ? | ? | ? |
| Empl | ? | -(0.99) | ? | +(0.96) | ? | ? | ? | ? | +(0.99) |
| Tech | ? | ? | ? | ? | ? | ? | ? | ? | ? |
| Univ | -(0.96) | ? | ? | -(0.96) | +(0.98) | ? | ? | ? | ? |
| Libe | -(1.00) | +(1.00) | ? | -(1.00) | +(1.00) | ? | ? | +(0.99) | ? |

La présence d'un "?" signifie qu'il n'est pas possible de se prononcer sur l'existence, dans la population, d'une attraction ou d'une répulsion pour ce couple de modalités avec une garantie suffisante (> 0,95).
La présence d'un "+" ou d'un "-" signifie qu'il est possible de se prononcer sur l'existence d'une attraction ou d'une répulsion pour ce couple de modalités avec une garantie suffisante (> 0,95).

Ampleur des écarts à la moyenne ?

Pour évaluer l'ampleur des écarts il existe deux catégories d'indices :

- les écarts relatifs (ou taux de liaison) et les densités. Écarts relatifs et densités ne sont en fait que deux présentations différentes d'un même indice.

- les contributions au Φ^2


Ces deux indices ne donnent pas nécessairement les mêmes résultats⁷. C'est la raison pour laquelle on les calculera tous les deux afin de les confronter.

Écarts relatifs ou Taux de liaison

SES-Pegase

Menu Analyser

- Analyser la liaison GRP*QUAL
- Écarts locaux à l'indépendance
- Taux de liaison (écarts relatifs)

- Cliquer sur l'icône 

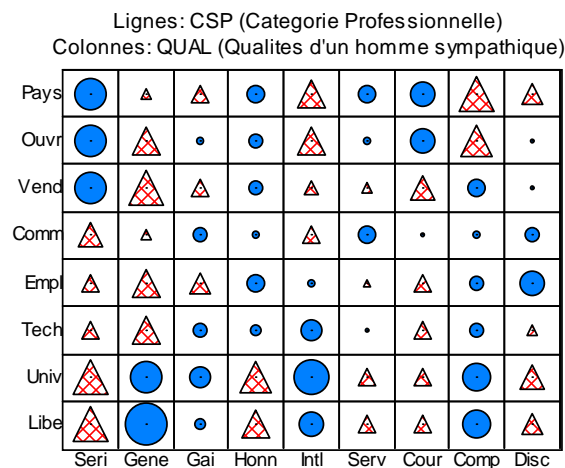


Figure 4 : Représentation graphique des écarts relatifs (taux de liaison) entre profils des groupes et profil moyen

Ce graphique permet de visualiser où se situent les écarts relatifs importants, qu'ils soient positifs (cf. disques bleus) ou négatifs (cf. triangles rouges)⁸. La surface des figures est proportionnelle à l'ampleur des écarts.

On voit, par exemple, que les professions libérales citent relativement souvent (par rapport aux autres groupes) généreux pour qualifier un homme de sympathique. De même on voit que les universitaires et les professions libérales citent peu souvent sérieux.

Après cette première lecture graphique des écarts relatifs, on revient au tableau pour obtenir les valeurs précises de ces écarts

⁷ Comme la moyenne et la médiane, pourtant tous deux indices de tendance centrale, ne prennent la même valeur que dans certains cas.

⁸ Les couleurs associées à positif (bleu) et à négatif (rouge) peuvent être modifiées par l'utilisateur.

Des groupes d'individus décrits par une variable qualitative
Le dossier SYMPA
Comparer et décrire les groupes

Tableau 13 : Écarts relatifs (taux de liaison) entre profils et profil moyen.

| | Seri | Gene | Gai | Honn | Intl | Serv | Cour | Comp | Disc |
|-------------|------|-------|------|------|------|------|------|------|------|
| Pays | +63% | -8% | -17% | +20% | -40% | +17% | +41% | -67% | -28% |
| Ouvr | +73% | -48% | +2% | +14% | -45% | +4% | +36% | -49% | +1% |
| Vend | +67% | -62% | -14% | +16% | -11% | -4% | -34% | +24% | +1% |
| Comm | -33% | -5% | +13% | +5% | -14% | +20% | +1% | +3% | +11% |
| Empl | -18% | -45% | -22% | +23% | +2% | -2% | -17% | +11% | +44% |
| Tech | -15% | -46% | +16% | +7% | +26% | +0% | -18% | +14% | -5% |
| Univ | -68% | +60% | +26% | -48% | +75% | -14% | -17% | +47% | -29% |
| Libe | -66% | +124% | +10% | -45% | +43% | -16% | -19% | +47% | -25% |

Prenons deux exemples, un écart positif et un écart négatif.

Un taux de liaison de +63 % (pour la case Pays-Séri) signifie que :

- les paysans citent sérieux plus souvent (+) que l'ensemble des personnes interrogées (sur-représentation ou attraction).

- la fréquence avec laquelle ils la citent est 63% plus grande que la fréquence chez l'ensemble des personnes interrogées.

Un taux de liaison de -68 % (pour la case Univ-Séri) signifie que :

- les universitaires citent sérieux moins souvent (-) que l'ensemble des personnes interrogées (sous-représentation ou répulsion).

- la fréquence avec laquelle ils citent sérieux est 68% plus faible que la fréquence chez l'ensemble des personnes interrogées.

Pour l'interprétation des écarts relatifs, on ne prendra en compte que les écarts les plus importants, supérieurs en valeur absolue à 50%.

Il apparaît ainsi que :

- *Sérieux* est, relativement aux autres groupes, souvent cité par les paysans, les ouvriers et les vendeurs, et peu cité par les universitaires et les professions libérales,

- *Généreux* est, relativement aux autres groupes, peu cité par les vendeurs et souvent cité par les universitaires et les professions libérales,

- *Intelligent* est, relativement aux autres groupes, souvent cité par les universitaires,

- *Compréhensif* est, relativement aux autres groupes, peu cité par les paysans.

Les densités

Menu Analyser

- Analyser la liaison GRP*QUAL
- Écarts locaux à l'indépendance
- Densité (O/T)

Tableau 14 : Densités entre groupes et qualités citées.

| | Seri | Gene | Gai | Honn | Intl | Serv | Cour | Comp | Disc |
|-------------|------|------|------|------|------|------|------|------|------|
| Pays | 1.63 | 0.92 | 0.83 | 1.20 | 0.60 | 1.17 | 1.41 | 0.33 | 0.72 |
| Ouvr | 1.73 | 0.52 | 1.02 | 1.14 | 0.55 | 1.04 | 1.36 | 0.51 | 1.01 |
| Vend | 1.67 | 0.38 | 0.86 | 1.16 | 0.89 | 0.96 | 0.66 | 1.24 | 1.01 |
| Comm | 0.67 | 0.95 | 1.13 | 1.05 | 0.86 | 1.20 | 1.01 | 1.03 | 1.11 |
| Empl | 0.82 | 0.55 | 0.78 | 1.23 | 1.02 | 0.98 | 0.83 | 1.11 | 1.44 |
| Tech | 0.85 | 0.54 | 1.16 | 1.07 | 1.26 | 1.00 | 0.82 | 1.14 | 0.95 |
| Univ | 0.32 | 1.60 | 1.26 | 0.52 | 1.75 | 0.86 | 0.83 | 1.47 | 0.71 |
| Libe | 0.34 | 2.24 | 1.10 | 0.55 | 1.43 | 0.84 | 0.81 | 1.47 | 0.75 |

Les densités sont une autre manière, parfaitement équivalente aux écarts relatifs (taux de liaison), d'exprimer les écarts entre profils et profil moyen.

Lorsque l'on parle d'une augmentation d'un prix de +100%, il est plus simple de dire que le prix a doublé (multiplié par 2). Une augmentation de 200% correspond à un triplement du prix. Si un prix a diminué de la moitié (-50%) il est équivalent de dire qu'il a été divisé par deux (multiplié par 0.50).


De la même manière, alors que les écarts relatifs (taux de liaison) expriment les écarts entre profils et profil moyen sous forme d'un pourcentage (positif ou négatif) les densités expriment ces écarts sous forme d'un coefficient multiplicateur (supérieur à 1 ou inférieur à 1).

A titre indicatif, lorsqu'un écart relatif est positif et supérieur à 100%, on préférera exprimer l'écart sous forme de densité. Ainsi, chez les professions libérales, la fréquence de citation de *généreux* est de 124% plus grande (taux de liaison) que la fréquence de cette qualité chez l'ensemble des personnes interrogées. On dira plutôt que cette fréquence de citation de *généreux* chez les professions libérales est 2.24 fois plus grande (densité) que chez l'ensemble des personnes interrogées. De même, lorsqu'un écart relatif est négatif et inférieur à -50%, on préférera exprimer l'écart sous forme d'une densité. Ainsi, toujours chez les professions libérales, la fréquence de citation de *sérieux* est 66% inférieure (taux de liaison) à la fréquence de cette qualité chez l'ensemble des personnes interrogées. On dira plutôt que cette fréquence de citation de la qualité *sérieux* chez les professions libérales est 3 fois plus petite (densité = 0.34) que chez l'ensemble des personnes interrogées.

Les contributions au Φ^2

Menu Analyser

- Analyser la liaison GRP*QUAL
- Écart local à l'indépendance
- Contributions relatives (signées) au Φ^2

- Cliquer sur l'icône .

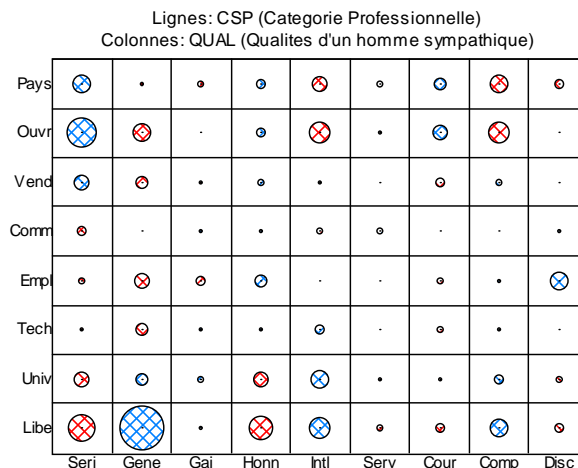


Figure 5 : Représentation graphique des contributions au Φ^2

Ce graphique permet de voir rapidement où se situent les contributions au Φ^2 importantes.

Une contribution au Φ^2 est toujours positive. Cependant :

- un disque bleu indique que la contribution est due à un pourcentage plus élevé que le pourcentage moyen.
- un disque rouge indique que la contribution est due à un pourcentage plus faible que le pourcentage moyen.

On constate ainsi que les Professions libérales citent relativement fréquemment (relativement aux autres GRP) Généreux. De même, mais dans une moindre mesure, les Ouvriers citent relativement fréquemment Sérieux.

A l'inverse, les Professions libérales citent Sérieux moins fréquemment que les autres groupes, etc.

On ne retient pour l'analyse que les contributions les plus importantes. Comment fixer la limite d'une contribution importante ? En prenant le critère du pourcentage moyen : si les contributions étaient également réparties sur les 72 (8 x 9) cases, on aurait 1.4% pour chaque case. On retiendra donc les contributions au moins égales à 2%. On les a fait apparaître en rouge et en gras dans le tableau.

Des groupes d'individus décrits par une variable qualitative
Le dossier SYMPA
Comparer et décrire les groupes

Tableau 15 : Contributions relatives au Φ^2 .

| | Seri | Gene | Gai | Honn | Intl | Serv | Cour | Comp | Disc | Total |
|--------------|---------------|----------------|-----------|---------------|---------------|-----------|---------------|---------------|---------------|-------------|
| Pays | 3% (+) | 0% (-) | 0% (-) | 1% (+) | 2% (-) | 0% (+) | 2% (+) | 3% (-) | 1% (-) | 11% |
| Ouvr | 8% (+) | 3% (-) | 0% (+) | 1% (+) | 4% (-) | 0% (+) | 2% (+) | 4% (-) | 0% (+) | 22% |
| Vend | 2% (+) | 1% (-) | 0% (-) | 0% (+) | 0% (-) | 0% (-) | 1% (-) | 0% (+) | 0% (+) | 4% |
| Comm | 1% (-) | 0% (-) | 0% (+) | 0% (+) | 0% (-) | 0% (+) | 0% (+) | 0% (+) | 0% (+) | 2% |
| Empl | 0% (-) | 2% (-) | 1% (-) | 1% (+) | 0% (+) | 0% (-) | 1% (-) | 0% (+) | 3% (+) | 8% |
| Tech | 0% (-) | 1% (-) | 0% (+) | 0% (+) | 1% (+) | 0% (+) | 0% (-) | 0% (+) | 0% (-) | 3% |
| Univ | 2% (-) | 1% (+) | 0% (+) | 2% (-) | 3% (+) | 0% (-) | 0% (-) | 1% (+) | 0% (-) | 10% |
| Libe | 7% (-) | 18% (+) | 0% (+) | 6% (-) | 4% (+) | 0% (-) | 1% (-) | 3% (+) | 1% (-) | 40% |
| Total | 23% | 27% | 2% | 10% | 14% | 1% | 6% | 12% | 4% | 100% |

Le total des contributions en pourcentages fait 100%. Chaque pourcentage indique la contribution du couple de modalités au Φ^2 total. Une contribution au Φ^2 est toujours positive. Cependant :

- un signe (+) indique que la contribution est due à un pourcentage plus élevé que le pourcentage moyen,
- un signe (-) indique que la contribution est due à un pourcentage plus faible que le pourcentage moyen.

Rédiger un compte rendu de l'analyse

Il semble que,

pour l'ensemble de la population d'où provient l'échantillon de 1 200 personnes adultes,

il existe des différences entre les groupes professionnels

du point de vue des qualités citées pour qualifier un homme sympathique (cf. Tableau 12).

< insérer Tableau 12 >

Relativement aux autres groupes professionnels (on note g la garantie bayésienne) :

- les paysans citent relativement plus souvent *sérieux* ($g = 99\%$) et citent relativement moins souvent *intelligent* ($g = 96\%$) et *compréhensif* ($g = 100\%$)
- les ouvriers citent relativement plus souvent *sérieux* ($g = 100\%$) et *courageux* ($g = 98\%$) et citent relativement moins souvent *généreux* ($g = 99\%$), *intelligent* ($g = 99\%$) et *compréhensif* ($g = 100\%$)
- les employés citent relativement plus souvent *honnête* ($g = 96\%$) et citent relativement moins souvent *généreux* ($g = 99\%$) et *discret* ($g = 99\%$).
- les universitaires citent relativement plus souvent *intelligent* ($g = 98\%$) et citent relativement moins souvent *sérieux* ($g = 96\%$) et *honnête* ($g = 96\%$).
- les professions libérales citent relativement plus souvent *intelligent* ($g = 100\%$) et *compréhensif* ($g = 99\%$) et citent relativement moins souvent *sérieux* ($g = 100\%$) et *honnête* ($g = 100\%$).

< insérer Figure 4 >

Si l'on s'interroge maintenant sur l'importance de ces écarts,

on constate que,

sur l'échantillon de 1 200 personnes adultes,

relativement autres groupes professionnels :

- les paysans citent beaucoup plus souvent *sérieux* ($txl = +63\%$) et très peu souvent *compréhensif* ($txl = -67\%$),
- les ouvriers citent beaucoup plus souvent *sérieux* ($txl = +73\%$),
- les vendeurs citent beaucoup plus souvent *sérieux* ($txl = +67\%$) et très peu souvent *généreux* ($txl = -62\%$),
- les universitaires citent beaucoup plus souvent *généreux* ($txl = +60\%$) et *intelligent* ($txl = +75\%$) et citent très peu souvent *sérieux* ($txl = -68\%$),
- les professions libérales citent beaucoup plus souvent *généreux* ($txl = +124\%$) et citent très peu souvent *sérieux* ($txl = -66\%$).

ANNEXE

Contributions absolues (signées) au Φ^2

Menu **Analyse**
 - Analyser la liaison entre GRP et QUAL
 - Écartts locaux à l'indépendance
 - Contributions absolues (signées) au Φ^2 .

Tableau 16 : Contributions absolues au Φ^2

| | Seri | Gene | Gai | Honn | Intl | Serv | Cour | Comp | Disc | Total |
|--------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| Pays | 0.0040 (+) | 0.0000 (-) | 0.0003 (-) | 0.0007 (+) | 0.0022 (-) | 0.0003 (+) | 0.0020 (+) | 0.0044 (-) | 0.0007 (-) | 0.0148 |
| Ouvr | 0.0107 (+) | 0.0037 (-) | 0.0000 (+) | 0.0008 (+) | 0.0057 (-) | 0.0000 (+) | 0.0030 (+) | 0.0048 (-) | 0.0000 (+) | 0.0288 |
| Vend | 0.0024 (+) | 0.0017 (-) | 0.0001 (-) | 0.0002 (+) | 0.0001 (-) | 0.0000 (-) | 0.0007 (-) | 0.0003 (+) | 0.0000 (+) | 0.0056 |
| Comm | 0.0011 (-) | 0.0000 (-) | 0.0002 (+) | 0.0000 (+) | 0.0003 (-) | 0.0005 (+) | 0.0000 (+) | 0.0000 (+) | 0.0001 (+) | 0.0022 |
| Empl | 0.0006 (-) | 0.0031 (-) | 0.0008 (-) | 0.0018 (+) | 0.0000 (+) | 0.0000 (-) | 0.0007 (-) | 0.0002 (+) | 0.0034 (+) | 0.0107 |
| Tech | 0.0002 (-) | 0.0016 (-) | 0.0002 (+) | 0.0001 (+) | 0.0009 (+) | 0.0000 (+) | 0.0004 (-) | 0.0002 (+) | 0.0000 (-) | 0.0036 |
| Univ | 0.0024 (-) | 0.0015 (+) | 0.0003 (+) | 0.0022 (-) | 0.0040 (+) | 0.0001 (-) | 0.0002 (-) | 0.0011 (+) | 0.0004 (-) | 0.0123 |
| Libe | 0.0087 (-) | 0.0240 (+) | 0.0002 (+) | 0.0073 (-) | 0.0049 (+) | 0.0006 (-) | 0.0008 (-) | 0.0043 (+) | 0.0011 (-) | 0.0518 |
| Total | 0.0302 | 0.0357 | 0.0020 | 0.0132 | 0.0182 | 0.0016 | 0.0078 | 0.0154 | 0.0058 | 0.1298 |

Le Tableau 16 est l'intermédiaire de calcul vers les contributions relatives au Φ^2 (cf. Tableau 15).

Chaque valeur indique ainsi la contribution du couple de modalités au Φ^2 total. Le total des contributions (cf. en bas à droite du tableau) est donc égal au Φ^2 (0,1298).

Pour une case donnée, en divisant la contribution absolue par le Φ^2 , on obtient la contribution relative (cf. Tableau 15). Par exemple, pour la case en haut à gauche : $0.0040 / 0.1298 = 0.03 = 3\%$.

RÉFÉRENCES

- Bernard, J.-M. (1998). Bayesian Inference for Categorized Data. In H. Rouanet & J.-M. Bernard & M.-C. Bert & B. Lecoutre & M.-P. Lecoutre & B. L. Roux (Eds.), *New Ways in Statistical Methodology - From Significance Tests to Bayesian Inference* (pp. 159-226). Berne: Peter Lang.
- Corroyer, D., & Wolff, M. (2003). *L'Analyse Statistique des Données en Psychologie; Concepts et Méthodes de base*. Paris: Armand Colin (Cursus).
- Escoffier, B., & Pagès, J. (1998). *Analyses factorielles simples et multiples. Objectifs, méthodes et interprétation* (3ème ed.). Paris: Dunod.
- Lebart, L., Piron, M., & Morineau, A. (2006). *Statistique exploratoire multidimensionnelle - Visualisation et inférence en fouilles de données* (4 ed.). Paris: Dunod.
- Rouanet, Le Roux (1993). *Analyse des données multidimensionnelles*, Paris: Dunod.
- Rouanet, Le Roux (1995). *Exercices et Solutions. Statistique en Sciences Humaines*, Paris: Dunod.
- Smithson, M. (2003). *Confidence intervals*. London: Sage.

LISTE DES TABLEAUX

| | |
|----------------------------------------------------------------------------------------------------|----|
| Tableau 1 : Structure du tableau des données individuelles | 1 |
| Tableau 2 : Tableau de contingence | 1 |
| Tableau 3 : Tableau pondéré | 2 |
| Tableau 4 : Répartition, en effectifs et en pourcentages, des groupes | 3 |
| Tableau 5 : Répartition, en effectifs et en pourcentages, des qualités citées (QUAL)..... | 4 |
| Tableau 6 : Carré moyen de contingence (Φ^2)..... | 5 |
| Tableau 7 : Test du χ^2 | 5 |
| Tableau 8 : V^2 et V de Cramer | 6 |
| Tableau 9 : Répartition des 1 200 réponses selon le groupe professionnel et la qualité citée. | 7 |
| Tableau 10 : Profils des groupes professionnels selon les qualités citées. | 8 |
| Tableau 11 : Attractions-Répulsions entre groupes et qualités citées..... | 9 |
| Tableau 12 : Inférence sur les attractions-répulsions entre groupes et qualités citées. | 9 |
| Tableau 13 : Écarts relatifs (taux de liaison) entre profils et profil moyen. | 11 |
| Tableau 14 : Densités entre groupes et qualités citées. | 11 |
| Tableau 15 : Contributions relatives au Φ^2 | 13 |
| Tableau 16 : Contributions absolues au Φ^2 | 14 |

LISTE DES FIGURES

| | |
|--------------------------------------------------------------------------------------------------------------------------|----|
| Figure 1 : Répartition en pourcentages, des groupes professionnels (diagramme à secteurs)..... | 3 |
| Figure 2 : Répartition en pourcentages des qualités citées (QUAL)..... | 4 |
| Figure 3 : Représentation graphique des profils des groupes | 8 |
| Figure 4 : Représentation graphique des écarts relatifs (taux de liaison) entre profils des groupes et profil moyen..... | 10 |
| Figure 5 : Représentation graphique des contributions au Φ^2 | 12 |

SOMMAIRE

| | |
|------------------------------------------------------------------------|-----------|
| Type des données analysées | 1 |
| <i>Tableau Individus × Variables ou Tableau de contingence ?</i> | <i>1</i> |
| <i>Questions</i> | <i>2</i> |
| Un exemple : le dossier SYMPA | 2 |
| <i>Questions</i> | <i>2</i> |
| <i>Type et statut des variables</i> | <i>2</i> |
| Analyses univariées (par variable)..... | 3 |
| <i>Les groupes professionnels (GRP)</i> | <i>3</i> |
| <i>Les qualités citées (QUAL)</i> | <i>4</i> |
| <i>Rédiger un compte rendu de l'analyse</i> | <i>5</i> |
| Différences entre les groupes ? | 5 |
| <i>Existence de différences entre les groupes ?</i> | <i>5</i> |
| <i>Ampleur des différences entre les groupes ?</i> | <i>6</i> |
| <i>Rédiger un compte rendu de l'analyse</i> | <i>6</i> |
| Décrire les groupes..... | 7 |
| <i>Effectifs observés</i> | <i>7</i> |
| <i>Profils des groupes et profil-moyen</i> | <i>7</i> |
| <i>Sens des écarts à la moyenne ?</i> | <i>9</i> |
| <i>Ampleur des écarts à la moyenne ?</i> | <i>10</i> |
| <i>Rédiger un compte rendu de l'analyse</i> | <i>13</i> |
| Annexe | 14 |
| <i>Contributions absolues (signées) au Φ^2</i> | <i>14</i> |
| Références | 15 |
| Liste des tableaux..... | 15 |
| Liste des figures..... | 15 |