# Objectif

Statistique descriptive avec TANAGRA.

La statistique descriptive vise à résumer l'information portée par un tableau de données. « Trop d'informations tue l'information » a-t-on coutume de dire. Il est illusoire d'inspecter un tableau contenant des centaines, voire des milliers, d'observations et d'en déduire des tendances.

L'objectif de la statistique descriptive est de nous fournir une image simplifiée de la réalité, en mettant en exergue des caractéristiques qui ne sont pas discernables de prime abord. Elle emmène un nouvel éclairage sur les données. Elle s'appuie pour cela sur des indicateurs et des représentations graphiques qui, pour simples qu'elles soient, sont très souvent pertinentes pour une bonne compréhension de la structure des données.

Nous présenterons les techniques descriptives selon deux axes. Tout d'abord nous ferons la distinction « techniques univariées », qui étudient les variables individuellement, et « techniques bivariées », qui étudient les relations entre 2 variables. Le second axe repose sur la distinction entre les variables catégorielles (qualitatives nominales) et les variables continues (quantitatives).

# Données

### Fichier de données

Les données décrivent les résultats d'une enquête de satisfaction par rapport à un produit pour bébé quelconque auprès de 20 femmes. Elles proviennent du site de <u>Fabrice Mazerolle</u><sup>1</sup> qui, outre les données, propose un cours en ligne d'excellente facture qui peut servir de repère dans ce didacticiel. Par rapport au fichier originel, nous avons supprimé les informations nominatives (c'est le cas généralement dans le traitement des enquêtes).

| M 🕅  | icrosoft Exce           | l - enquete_s  | atisfaction_f          | emmes_1953                     | l. xls              |                  |            |             |
|------|-------------------------|----------------|------------------------|--------------------------------|---------------------|------------------|------------|-------------|
| 1    | <u>Eichier E</u> dition | Affichage Inse | rtion Forma <u>t</u> ( | <u>D</u> utils <u>D</u> onnées | Fenêtre <u>?</u> Ta | nagra Sipina     |            | _ /#  ×     |
|      | e 🖬 🧉                   | 🔉 🌮 🐰 🗉        | b 🖻 🛷 🕞                |                                | Σ fx 🔂 💈            |                  | 🛃 100% 👻 🥳 | 2 • 🚯 🔌 • 💥 |
|      | Al 🝷                    | = Id           |                        |                                |                     |                  |            |             |
|      | A                       | В              | С                      | D                              | E                   | F                | G          | н           |
| 1    | ld                      | NbEnfants      | Revenu                 | Ville                          | Opinion             |                  |            |             |
| 2    | 1                       | 4              | 30000                  | Paris                          | Mauvaise            | -                |            |             |
| 3    | 2                       | 0              | 60000                  | Montpellier                    | Très bonne          | -                |            |             |
| 4    | 3                       | 11             | 9000                   | Rouen                          | Mauvaise            |                  |            |             |
| 5    | 4                       | 11             | 15000                  | Paris                          | Mauvaise            | -                |            |             |
| 6    | 5                       | 2              | 40000                  | Marseille                      | Moyenne             | -                |            |             |
| 7    | 6                       | 2              | 40000                  | Marseille                      | Moyenne             | 2                |            |             |
| 8    | 7                       | 2              | 10000                  | Nice                           | Bonne               | 3                |            |             |
| 9    | 8                       | 3              | 45000                  | Paris                          | Moyenne             |                  |            |             |
| 10   | 9                       | 3              | 80000                  | Nancy                          | Passable            |                  |            |             |
| 11   | 10                      | 4              | 50000                  | Nice                           | Moyenne             |                  |            |             |
| 12   | 11                      | 2              | 60000                  | Nice                           | Passable            | -                |            |             |
| 13   | 12                      | 3              | 55000                  | Marseille                      | Bonne               |                  |            |             |
| 14   | 13                      | 4              | 85000                  | Montpellier                    | Bonne               | -                |            |             |
| 15   | 14                      | 5              | 60000                  | Nice                           | Moyenne             | 2                |            |             |
| 16   | 15                      | 2              | 40000                  | Rouen                          | Bonne               | 3                | -          |             |
| 17   | 16                      | 1              | 10000                  | Nancy                          | Passable            |                  |            |             |
| 18   | 17                      | 3              | 120000                 | Nice                           | Mauvaise            |                  |            |             |
| 19   | 18                      | 3              | 80000                  | Paris                          | Moyenne             | -                |            |             |
| 20   | 19                      | 4              | 18000                  | Paris                          | Très bonne          |                  |            |             |
| 21   | 20                      | 5              | 90000                  | Marseille                      | Passable            | L                |            |             |
| 22   | 1                       |                |                        |                                |                     |                  |            |             |
|      | ► ► \ donn              | ées /          |                        |                                | •                   |                  |            | •           |
| Des  | sin 🔹 🗟 🌀               | Formes automal | tiques 👻 🔪             |                                | 4 🛛 🔌 -             | <u>4 · A</u> · i | = = = •    | <i>•</i> •  |
| Prêt |                         |                |                        |                                | Somme=9             | 997264           | NUM        |             |

<sup>&</sup>lt;sup>1</sup> <u>http://www.mazerolle.fr/stats/stats.htm</u>

## Créer un diagramme dans TANAGRA

Dans un premier temps, il faut initialiser un diagramme de traitements et charger les données dans le logiciel TANAGRA. Le plus simple est d'ouvrir le fichier « enquete\_satisfaction\_femmes\_1953.xls » dans le tableur EXCEL. Nous sélectionnons la plage de données et nous activons le menu TANAGRA/EXECUTE TANAGRA installée à l'aide de la macro complémentaire TANAGRA.XLA livrée avec le logiciel<sup>2</sup>.

| N 🔀  | licrosoft Exce  | l - enquete_s  | atisfaction_f          | emmes_195             | 3.xls                        | V                       |            | -            |       |
|------|-----------------|----------------|------------------------|-----------------------|------------------------------|-------------------------|------------|--------------|-------|
|      | Eichier Edition | Affichage Inse | rtion Forma <u>t</u> ( | utils <u>D</u> onnées | Fe <u>n</u> être <u>?</u> Ta | nagra Sipina            |            |              | _ 8 × |
| l n  | 🚔 🔲 🙈 [         | Շ 💖 🐰 🖻        | h 🙉 🛷 🕞                | ) + Ci + 🥘            | Σ f* 🕞                       | I ZI In 🛛               | 🛃 100% 👻 🕅 | 2 _ 2%       | 🕭 + 2 |
|      |                 |                |                        |                       | , — , ~ <b>40</b> 12         |                         | •          | 4 4 1 3 4 10 | · ·   |
|      | A               | В              | с                      | D                     | E                            | F                       | G          | Н            | -     |
| 1    | ld              | NbEnfants      | Revenu                 | Ville                 | Opinion                      |                         |            |              | -     |
| 2    | 1               | 4              | 30000                  | Paris                 | Mauvaise                     |                         |            |              |       |
| 3    | 2               | 0              | 60000                  | Montpellier           | Très bonne                   |                         |            |              |       |
| 4    | 3               | 1              | 9000                   | Rouen                 | Mauvaise                     |                         |            |              |       |
| 5    | 4               | 1              | 15000                  | Paris                 | Mauvaise                     |                         |            |              |       |
| 6    | 5               |                | 40000                  | Mana aill a           | 1                            | 1                       |            |              |       |
| 7    | 6               | Execute Ta     | nagra                  |                       |                              |                         |            |              |       |
| 8    | 7               | -              |                        |                       |                              |                         |            |              |       |
| 9    | 8               | Datase         | at range (includin     | g the name of th      | ne attributes fir            | st row):                |            |              |       |
| 10   | 9               | \$A\$          | 1:\$E\$21              |                       |                              |                         | _          |              |       |
| 11   | 10              |                |                        |                       | 1                            |                         | _          |              |       |
| 12   | 11              | -              |                        |                       | ок                           | Cancel                  |            |              |       |
| 13   | 12              | -              |                        |                       |                              |                         |            |              |       |
| 14   | 13              | 5              | 60000                  | Nice                  | Movenne                      |                         |            |              |       |
| 16   | 14              | 2              | 40000                  | Rouen                 | Boppe                        |                         |            |              |       |
| 17   | 15              | 1              | 10000                  | Nancy                 | Passable                     |                         |            |              |       |
| 18   | 17              | 3              | 120000                 | Nice                  | Mauvaise                     |                         |            |              |       |
| 19   | 18              | 3              | 80000                  | Paris                 | Movenne                      |                         |            |              |       |
| 20   | 19              | 4              | 18000                  | Paris                 | Très bonne                   |                         |            |              |       |
| 21   | 20              | 5              | 90000                  | Marseille             | Passable                     |                         |            |              |       |
| 22   |                 |                |                        |                       |                              |                         |            |              |       |
| 4    | ▶ ▶ donne       | ées /          |                        |                       | •                            |                         |            |              |       |
| Des  | sin + 🔓 🌀       | Formes automal | tiques 🗸 🔪             |                       | 4 🗕 🕭 -                      | <u>.</u> • <u>A</u> • : |            | <b>-</b>     |       |
| Prêt |                 |                |                        |                       | Somme=9                      | 997264                  | NUM        |              |       |

TANAGRA est automatiquement lancé, un nouveau diagramme de traitements est mis en place et les données sont disponibles à la racine du diagramme.

| Image: Second Statistics     Image: Second Stati   | 💇 TANAGRA 1.4.19 - [Da  | taset (tan22.txt)]       |         |                        |                     |                      |  |  |  |  |  |  |  |
|---|-------------------------|--------------------------|---------|------------------------|---------------------|----------------------|--|--|--|--|--|--|--|
| Analysis       Dataset (tan22.txt)       Parameters         Dataset (tan22.txt)       Parameters       Dataset (tan22.txt)         Results       Dataset (tan22.txt)       Results         Download information       Dataset (tan22.txt)       Results         Download information       Dataset (tan22.txt)       Results         Download information       Dataset (tan22.txt)       Results         Dataset description       Satribute(s)       Satribute(s)       Results         Dataset description       Satribute(s)       Satribute(s)       Results       Results         Dataset description       Satribute(s)       Satribute(s)       Results       Results       Results         Dataset description       Satribute(s)       Satribute(s)       Results       Results       Results         Dataset description       Satribute(s)       Satribute(s)       Results       Results       Results         Dataset description       Sattifices       Nonparametric statistics       Instance selection       Feature construction         Feature selection       Regression       Factorial analysis       PLS       Clustering         Spv learning       Meta-spv learning       Spv learning assessment       Scoring       Association         Components  | Tile Diagram Component  | Window Help              |         |                        |                     | _ 8 ×                |  |  |  |  |  |  |  |
| Dataset (tan22.txt)         Dataset (tan22.txt)         Parameters         Database : C:DOCUME-1Watson/LOCALS-INTemp/tan22.txt         Dataset description         Dataset description         Sattribute(s)         20 example(s)         Components         Components         Components         Components         Construction         Feature selection       Feature construction         Feature selection       Regression         Factorial analysis       PLS       Clustering         Spy learning       Spy learning assessment       Scoring       Association         Components       Clustering         Spy learning       Spy learning assesss  | D 📽 🖪  🍇                |                          |         |                        |                     |                      |  |  |  |  |  |  |  |
| Image: Dataset (tan22.txt)       Parameters         Data vi   |                         | Analysis                 |         |                        |                     |                      |  |  |  |  |  |  |  |
| Parameters         Database : C:DOCUME-1WatsonUOCALS-11TempUtan22.txt         Results         Download information         Datasource processing         Computation time         Omega         Allocated memory         S attribute(s)         20 example(s)   | Dataset (tan22.txt)     | 4                        |         |                        | Dataset (tan22.txt) | ,                    |  |  |  |  |  |  |  |
| Details       Concent Header He |                         |                          |         | Database : CUDG        | Parameters          | n)tan22 tut          |  |  |  |  |  |  |  |
| Results         Download information         Datasource processing         Computation time 0 ms         Allocated memory is KB       Dataset description         Sattribute(s)         20 example(s)         Components         Components         Components         Components         Components         Components         Components         Components         Components         Construction         Feature selection       Regression         Feature selection       Regression       Scoring       Alsociation         Components       Components         Components       Clustering         Spy learning assessment       Scoring       Alsociation         Clustering assessment       Scoring       Association         © Correlation scatterplot       Metaset       Scatterplot dataset       Scatterplot       Scatterplot       Scatterplot       Scatterplot   <   |                         |                          |         | DataDase : C:DC        |                     | ipitalizz.txt        |  |  |  |  |  |  |  |
| Download information         Datasource processing         Computation time       0 ms         Allocated memory       6 KB         Dataset description       5 attribute(s)         20 example(s)       20 example(s)         Components       Clustering         Pata visualization       Regression         Feature selection       Regression         Spulearning       Meta-spulearning         Spulearning       Scatterplot with label         Componenta       Scoring         Association       Scoring   |                         |                          |         |                        | Results             |                      |  |  |  |  |  |  |  |
| Datasource processing<br>Computation time 0 ms<br>Allocated memory 6 KB         Dataset description<br>S attribute(s)<br>2 dexample(s)         Components         Corponents         Corponents         Spy learning         Meta-spy learning         Spy learning         Meta-spy learning         Spy learning         Solaterplot         Solaterplot         Solaterplot         Solaterplot         Solaterplot         Solaterplot         Solaterplot   | Download information    |                          |         |                        |                     |                      |  |  |  |  |  |  |  |
| Dataset consumer       Computation time 0 ms         Allocated memory or KB       Dataset description         S attribute(s)       20 example(s)         20 example(s)       Components         Components       Clustering         Pata visualization       Regression         Feature selection       Regression         Patorial analysis       PLS         Clustering       Sociaterplot with label         Correlation scatterplot       Satterplot with label         Scatterplot       View multiple scatterplot   |                         | Download information     |         |                        |                     |                      |  |  |  |  |  |  |  |
| Computation time 0 ms<br>Allocated memory 6 KB<br>Dataset description<br>5 attribute(s)<br>20 example(s) ♥<br>Teature selection Regression Feature construction<br>Feature selection Regression Factorial analysis PLS Clustering<br>Spv learning Meta-spv learning Spv learning assessment Scoring Association<br>Correlation scatterplot With label<br>Correlation scatterplot With label<br>Scatterplot Line Scotterplot with label<br>Scatterplot Line Scotterplot  |                         |                          |         | Datasource pro         | cessing             |                      |  |  |  |  |  |  |  |
| Allocated memory 6 KB<br>Dataset description<br>5 attribute(s)<br>20 example(s)   |                         |                          |         | Computation time 0 ms  |                     |                      |  |  |  |  |  |  |  |
| Dataset description       Sattribute(s)<br>20 example(s)       Sattribute(s)<br>20 example(s)       Seture selection       Regression       Feature selection       Regression       Factorial analysis       PLS       Clustering       Association       Correlation scatterplot       Costaterplot       Scatterplot   |                         |                          |         | Allocated memory to NB |                     |                      |  |  |  |  |  |  |  |
| 5 attribute(s)<br>20 example(s)       5 attribute(s)<br>20 example(s)     5 attribute(s)<br>20 example(s)       Data visualization     Statistics     Nonparametric statistics     Instance selection     Feature construction       Feature selection     Regression     Factorial analysis     PLS     Clustering       Spv learning     Meta-spv learning     Spv learning assessment     Scoring     Association       Correlation scatterplot     L Scatterplot with label     Scatterplot     Scatterplot   |                         |                          |         | Dataset description    |                     |                      |  |  |  |  |  |  |  |
| Data visualization     Statistics     Nonparametric statistics     Instance selection     Feature construction       Feature selection     Regression     Factorial analysis     PLS     Clustering       Spv learning     Meta-spv learning     Spv learning     Spv learning     Scoring       Correlation scatterplot     Image: Scoring     Association       Export dataset     Image: Scoring     Scoring       Mew dataset     Image: Scoring     Scoring  |                         |                          |         | 5 attributo(c)         |                     |                      |  |  |  |  |  |  |  |
| Components           Data visualization         Statistics         Nonparametric statistics         Instance selection         Feature construction           Feature selection         Regression         Factorial analysis         PLS         Clustering           Spv learning         Meta-spv learning         Spv learning assessment         Scoring         Association           Correlation scatterplot         LL Scatterplot with label         Export dataset         LL Scatterplot   |                         |                          |         | 20 example(s)          |                     | *                    |  |  |  |  |  |  |  |
| Data visualization         Statistics         Nonparametric statistics         Instance selection         Feature construction           Feature selection         Regression         Factorial analysis         PLS         Clustering           Spv learning         Meta-spv learning         Spv learning assessment         Scoring         Association           Correlation scatterplot         Image: Spv learning assessment         Scoring         Association           Spv learning         Image: Spv learning assessment         Scoring         Association           Correlation scatterplot         Image: Spv learning assessment         Scoring         Association           Spv learning         Image: Spv learning assessment         Scoring         Association  | <u>р.</u>               |                          |         | Componente             |                     |                      |  |  |  |  |  |  |  |
| Feature selection     Regression     Factorial analysis     PLS     Clustering       Spv learning     Meta-spv learning     Spv learning assessment     Scoring     Association       Correlation scatterplot     Image: Scatterplot with label     Image: Scatterplot with label     Image: Scatterplot with label       Scatterplot     Image: Scatterplot with label     Image: Scatterplot with label     Image: Scatterplot with label   | Data visualization      | Statistics               | Nonpar  | ametric statistics     | Instance selection  | Feature construction |  |  |  |  |  |  |  |
| Spv learning         Meta-spv learning         Spv learning assessment         Scoring         Association           Correlation scatterplot         Image: Scatterplot with label         Image: Scatterplo  | ,<br>Feature selection  | Regression               | Fac     | torial analysis        | PLS                 | Clustering           |  |  |  |  |  |  |  |
| Correlation scatterplot     Zostterplot with label     Scatterplot     Scatterplot     Scatterplot     Scatterplot  | Spv learning            | Meta-spv learning        | Spv lea | rning assessment       | Scoring             | Association          |  |  |  |  |  |  |  |
| 隆王 Export dataset 國家 View dataset<br>교 Scatterplot 住고 View multiple scatterplot   | Correlation scatterplot | Zeatterplot with label   |         |                        |                     |                      |  |  |  |  |  |  |  |
| Image: Scatterplot         Image: Scatterplot   | Export dataset          | 🚾 View dataset           |         |                        |                     |                      |  |  |  |  |  |  |  |
|   | 🧖 Scatterplot           | 🔛 View multiple scatterp | olot    |                        |                     |                      |  |  |  |  |  |  |  |
|   |                         |                          |         |                        |                     |                      |  |  |  |  |  |  |  |
|   | L                       |                          |         |                        |                     |                      |  |  |  |  |  |  |  |

<sup>&</sup>lt;sup>2</sup> Cette macro est disponible depuis la version 1.4.11 de TANAGRA. Un didacticiel décrit la procédure d'installation (<u>http://eric.univ-lyon2.fr/~ricco/tanagra/fichiers/fr\_Tanagra\_Excel\_AddIn.pdf</u>).

## Statistiques univariées

L'outil privilégié pour décrire une variable qualitative est le diagramme en bâtons. Il s'agit d'énumérer les valeurs (les modalités) prises par la variable, et pour chaque modalité fournir l'effectif correspondant, et éventuellement la proportion (effectif modalité / effectif total) associée.

Dans TANAGRA, nous devons définir les variables à manipuler. Pour ce faire, nous introduisons le composant DEFINE STATUS en utilisant le raccourci dans la barre d'outils. La boîte de paramétrage apparaît automatiquement, nous sélectionnons en INPUT les variables VILLE et OPINION.

| File     Degram     Component. Window     Help       Image: State | - 8 × |
|---|-------|
| Define attribute statuses   |       |
| Parameters Define status 1 Parameters Attributes : Target Input Illustrative Ville Opinion Ville D Opinion  |       |
| Data visualization Statist  | •n    |
| Feature selection Regres OK Cancel Help   |       |
| Correlation scatterplot [2] Scatterplot with label<br>Export dataset<br>Scatterplot [:_View multiple scatterplot  |       |

Nous introduisons maintenant le composant UNIVARIATE DISCRETE STAT (onglet STATISTICS) dans le diagramme, à la suite du DEFINE STATUS. Nous activons menu VIEW pour accéder aux résultats.



L'indice de GINI permet de situer la concentration des effectifs sur certaines modalités. Une valeur proche de 0 indique que les effectifs sont concentrés sur une des modalités, s'il tend vers (1 - 1/K) où K est le nombre de modalités, nous avons équirépartition c.-à-d. les pourcentages sont les mêmes sur l'ensemble des modalités. Dans le cas de VILLE, la valeur maximale est 1-1/6 = 0.83, la valeur observée est 0.8, cela confirme que les différentes villes sont assez équitablement représentées dans ce fichier.

## Statistiques bivariées – Croisement de variables qualitatives

Nous nous intéressons maintenant à la répartition de l'opinion selon les villes. Nous devons former un tableau croisé, dit de contingence, il doit calculer les effectifs pour chaque association des modalités des variables.

### Fréquences absolues

A la suite du DEFINE STATUS, nous plaçons le composant CONTINGENCY CHI-SQUARE (onglet NONPARAMETRIC STATISTICS). Nous activons le menu PARAMETERS, et nous indiquons que les variables sont énumérées en tant qu'INPUT (Option INPUT LIST). Dans ce cas, un tableau de contingence est élaboré pour chaque couple de variable INPUT.



Nous cliquons sur le menu VIEW et nous obtenons :

| 💯 TANAGRA 1.4.19 - [Contingency Chi-Squ               | TANAGRA 1.4.19 - [Contingency Chi-Square 1] |                |                                     |                    |               |             |             |              |               |          |     |     |
|---|---|----------------|-------------------------------------|--------------------|---------------|-------------|-------------|--------------|---------------|----------|-----|-----|
| Tile Diagram Component Window Help                    |   |                |                                     |                    |               |             |             |              |               |          | -   | ₽ × |
| D 🗳 🔲   👪   |   |                |                                     |                    |               |             |             |              |               |          |     |     |
| Analysis  |   |                |                                     |                    |               |             |             |              |               |          |     | ^   |
| 🖃 🏢 Dataset (tan22.txt)                               |   |                |                                     |                    |               | Results     |             |              |               |          |     |     |
| 🖮 🎇 Define status 1<br>— 📶 Univariate discrete stat 1 | Row<br>(Y)                                  | Column<br>(X)  | ann Statistical indicator Cross-tab |                    |               |             |             |              |               |          |     |     |
| 🔲 🔛 Contingency Chi-Square 1                          |   |                | Stat                                | ¥alue              |               | Mauvaise    | Très        | Moyenne      | Bonne         | Passable | Sum |     |
|   |   |                | Tschuprow's t                       | 0.503409           |               |             | bonne       |              |               |          |     |     |
|   |   |                | Cramer's v                          | 0.532291           | Paris         | 2           | / 1         | 2            | 0             | 0        | 5   |     |
|   |   |                | Phi²                                | 1,133333           | Montpellier   | 0           | 1           | 0            | 1             | 0        | 2   |     |
|   | 1.00  | Opinion        | Chi² (n-value)                      | 22.67              | Rouen         | 1           | 0           | 0            | 1             | 0        | 2   |     |
|   | ville                                       |                | (p)                                 | (0.3055)           | Marseille     | 0           | 0           | 2            | 1             | 1        | 4   |     |
|   |   |                | Lambda                              | 0.200000           | Nice          | 1           | 0           | 2            | 1             | 1        | 5   |     |
|   |   |                | Tau (p-value)                       | 0.1925<br>(0.5682) | Nancy         | 0           | 0           | 0            | 0             | 2        | 2   |     |
|   |   |                | U(R/C) (p-value)                    | 0.3598<br>(0.2192) | Sum           | 4           | 2           | 6            | 4             | 4        | 20  |     |
|   |   |                |                                     |                    |               |             |             |              |               |          |     |     |
|   | LComput                                     | ation time · ( | lms                                 |                    |               |             |             |              |               |          |     | ×   |
|   |   |                |                                     | Component          | 3             |             |             |              |               |          |     |     |
| Data visualization Statistics                         |   | Nonparame      | tric statistics                     | Instance s         | election      | Feature     | constructio | n F          | eature seled  | ction    |     |     |
| Regression Factorial analy                            | rsis  | P              | LS                                  | Cluste             | ring          | Spv         | learning    |              | leta-spv lear | rning    |     |     |
| Spv learning assessment Scoring                       |   | Asso           | ciation                             |                    |               |             |             |              |               |          |     |     |
| 🔄 Categorical r 🛛 🖾 Goodman-K                         | endall's tau                                | ÷.             | Mood Runs                           | Test               | 🖄 Spe         | arman's rho |             | : Wilcox     | on Signed Ra  | anks "   |     |     |
| 🔛 Contingency Chi-Square 🛛 📃 Goodman-K                | ruskal Tau                                  | <u>lat</u> , k | (ruskal-Wallis 1-w                  | vay ANOVA 🛛 🖪      | Partial Theil | U           | 🗾 The       | eil U        |               |          |     |     |
| Friedman's ANOVA by Ranks 👫 Kendall's Co              | oncordance                                  | ew Millin      | Aann-Whitney Co                     | omparison 📑        | Sign Test     |             | 📥 Wal       | ld-Wolfowitz | Runs Test     |          |     |     |
| <   |   |                |                                     |                    |               |             |             |              |               |          |     | >   |
|   |   |                |                                     |                    |               |             |             |              |               |          |     |     |

Le tableau permet la lecture suivante : 2 personnes de « Paris » ont une opinion « mauvaise » du produit, 1 personne de « Nice » a une « bonne » opinion du produit, etc.

Une pléiade d'indicateurs est affichée. *Grosso modo*, ils indiquent le lien qui existe entre les deux variables du tableau. Pour se donner une idée simple de ce qu'ils véhiculent, prenons l'exemple du T de Tschuprow. Il prend la valeur 0 si les deux variables sont indépendantes c.-à-d. la connaissance de la valeur prise par l'une ne donne aucune indication sur la valeur prise par la seconde variable (en inversement). Si la relation est déterministe, sa valeur est 1. Il existe une littérature abondante sur ce thème<sup>3</sup>.

### Profils lignes et colonnes

La lecture directe des valeurs dans les cases du tableau de contingence apporte peu d'informations en définitive. Il est plus intéressant de s'intéresser aux pourcentages en ligne ou en colonne.

Si l'on reprend la satisfaction selon les villes, il paraît plus judicieux de ramener les données sur un même référentiel en divisant les effectifs par le nombre de personnes associées à chaque ville. Ainsi, les chiffres seront comparables. Prenons un exemple très simple : 1 personne a une bonne opinion du produit à Montpellier, il en est de même à Nice. Apparemment, la situation est identique.

Dans TANAGRA, nous pouvons afficher les profils en modifiant le paramétrage du composant. Nous cliquons sur le menu PARAMETERS. Dans l'onglet ANALYSIS, nous sélectionnons l'option ROW PERCENT. Nous validons en cliquant sur OK.

<sup>&</sup>lt;sup>3</sup> <u>http://www2.chass.ncsu.edu/garson/PA765/assocnominal.htm</u>

| С  | ross-tabulation parameters   |
|----|--|
| Cı | Parameters Analysis Additional information Additional information Additional information Additional information One Row percent Column percent Column percent Column percent Column secontribution Expected values Residuals Contribution thresold |
|    | Highlight contribution greater than 2.00 × average   |
|    |  |
|    | OK Cancel Help   |

En actionnant le menu VIEW, nous obtenons :

| 💯 TANAGRA 1.4.19 - [Contingency Chi-Squa                           | re 1]           |            |                  |                |                      |             |              |                   |             |           |              |       |
|--|-----------------|------------|------------------|----------------|----------------------|-------------|--------------|-------------------|-------------|-----------|--------------|-------|
| Tile Diagram Component Window Help                                 |                 |            |                  |                |                      |             |              |                   |             |           | -            | a x   |
|  |                 |            |                  |                |                      |             |              |                   |             |           |              |       |
| Analysis   | Row             | Column     | Statistical      | indicator      |                      |             |              | Cross-tab         |             |           |              | ^     |
| 🖃 🎹 Dataset (tan22.txt)  | (1)             | ()()       |                  |                |                      |             |              |                   |             |           |              |       |
| 😑 🎦 Define status 1  |                 |            | Stat             | Value          |                      | Mauvaise    | Très         | Moyenne           | Bonne       | Passable  | Sum          |       |
|  |                 |            | Tschuprow's t    | 0.503409       |                      | 2           | Donne        | -                 |             |           |              |       |
| Contingency Chi-Square 1   |                 |            | Cramer's v       | 0.532291       | Paris<br>Montpellier | 40.00%      | 20.00%       | 40.00%            | 0.00%       | 0.00%     | 5<br>100%    |       |
|  |                 |            | Phi <sup>z</sup> | 1,133333       |                      | 0           | 1            | 0                 | 1           | 0         | 2            |       |
|  |                 |            | Chi² (p-value)   | 22.67 (0.3055) |                      | 0.00%       | 50.00%       | 0.00%             | 50.00%      | 0.00%     | 100%         |       |
|  |                 |            | Lambda           | 0.200000       | Pouen                | 1           | 0            | 0                 | 1           | 0         | 2            |       |
|  | Ville           | Oninion    | Tau (p-value)    | 0.1925         | Rouen                | 50.00%      | 0.00%        | 0.00%             | 50.00%      | 0.00%     | 100%         |       |
|  |                 |            |                  | (0.5682)       | Marseille            | 0           | 0            | 2                 | 1           | 1         | 4            |       |
|  |                 |            | U(R/C) (p-value) | 0.3598         | _                    | 0.00%       | 0.00%        | 50.00%            | 25.00%      | 25.00%    | 100%         | =     |
|  |                 |            |                  | (0,2172)       | Nice                 | 20.00%      | 0.00%        | 2<br>40.00%       | 20.00%      | 20.00%    | 200%         |       |
|  |                 |            |                  |                |                      | 0           | 0.00,0       | -0.00,0           | 0           | 2010070   | 2            |       |
|  |                 |            |                  |                | Nancy                | 0.00%       | 0.00%        | 0.00%             | 0.00%       | 100.00%   | 100%         |       |
|  |                 |            |                  |                | 6                    | 4           | 2            | 6                 | 4           | 4         | 20           |       |
|  |                 |            |                  |                | sum                  | 20%         | 10%          | 30%               | 20%         | 20%       | 100%         |       |
|  |                 |            |                  |                |                      |             |              |                   |             |           |              | ~     |
|  |                 |            |                  | Component      | s                    |             |              |                   |             |           |              |       |
| Data visualization Statistics                                      |                 | Nonparamet | tric statistics  | Instance se    | lection              | Feature o   | construction | n Fe              | ature selec | tion      |              |       |
| Regression Factorial analysis                                      |                 | P          | LS               | Cluster        | ing                  | Spv         | learning     | Meta-spv learning |             | ning      |              |       |
| Spv learning assessment Scoring                                    |                 | Assoc      | iation           |                |                      |             |              |                   |             |           |              |       |
| 🖾 Categorical r 🛛 🕹 Goodman-Kruskal Lambda 🛛 🖄 Kendali's tau       |                 |            |                  | <u>.</u>       | Mood Runs T          | est         | 🖾 Spe        | arman's rho       |             | : Wilcoxe | n Signed Rar | nks T |
| 🖽 Contingency Chi-Square 🛛 Goodman-Kruskal Tau 🕍 Kruskal-Wallis 1- |                 |            |                  | ay ANOVA 🛛 🛱   | Partial Theil        | U           | 🗾 The        | ilU               |             |           |              |       |
| Friedman's ANOVA by Ranks 👫 Kendall's Con                          | Nann-Whitney Co | mparison 🏦 | Sign Test        |                | 📥 Wali               | d-Wolfowitz | Runs Test    |                   |             |           |              |       |
| <  |                 |            |                  |                |                      |             |              |                   |             |           |              | >     |
|  |                 |            |                  |                |                      |             |              |                   |             |           |              |       |

Reprenons notre exemple ci-dessus. Nous constatons que 50% des personnes ont une bonne opinion du produit à Montpellier, en revanche, ils ne sont que 20% à Nice. La raison est que l'effectif, on parle d'effectif marginal, n'est pas le même dans les deux villes.

## Variables quantitatives

### Statistiques univariées

Dans TANAGRA, les outils sont plus riches concernant les variables quantitatives. Nous disposons de deux composants qui répondent à des spécifications différentes.

Plaçons le composant DEFINE STATUS à la racine du diagramme, toujours en utilisant le raccourci de la barre d'outils, ce qui est le plus simple. Nous plaçons en INPUT maintenant les variables NBENFANTS et REVENU (calculer des statistiques sur les identifiants n'est pas très pertinent).

| TANAGRA 1.4.19 - [Contingency Chi-Squar   | e 1]   |   |  |                   |
|---|--|---|--|-------------------|
| File Diagram Component Window Help  |  |   |  | _ 8 ×             |
| Analysis  Dataset (tan22.txt)   | Row         Column         Statistical indicator           (Y)         (X)         Statistical indicator | Cross   | tab  | ^                 |
| 🖃 🏭 Define status 1   | Define attribute statuses  | loyer   | nne Bonne Passable S                           | um                |
| Lui Univariate discrete stat 1<br>Contingency Chi-Square 1  | Parameters   | 40  | 2 0 0<br>00% 0.00% 0.00%                       | 5<br>100%<br>2    |
|   | C Id<br>C NbEnfants  | Target     Input     Illustrative     0       NbEnfants       | 00% 50.00% 0.00%<br>0 1 0<br>.00% 50.00% 0.00% | 100%<br>2<br>100% |
|   |  | 50  | 2 1 1<br>.00% 25.00% 25.00%                    | 4<br>100%         |
|   |  | 40  | 2 1 1<br>.00% 20.00% 20.00%                    | 5<br>100%         |
|   |  | 0   | 0 0 2<br>.00% 0.00% 100.00%                    | 2<br>100%         |
|   |  |   | 6 4 4<br>30% 20% 20%                           | 20<br>100%        |
|   |  |   |  | <u> </u>          |
| Data visualization Statistics   |  |   | Feature selection                              |                   |
| Regression Factorial analysis   |  | OK Cancel Help  | Meta-spv learning                              |                   |
| Spy learning assessment Scoring   | HSSOCIACION  |   |  |                   |
| Categorical r  Contingency Chi-Square  Goodman-Kru  Good | skal Lambda 🛛 🧭 Kendall's tau<br>skal Tau 🚛 Kruskal-Wallis 1-way ANOVA                                   | Mood Runs Test  Spearman's  Partial Theil U  Theil U  Theil U | rho 💷 Wilcoxon Sig                             | gned Ranks T      |
| E Friedman's ANUVA by Ranks HI Kendalis Con   | cordance w 🔐 wann-Whitney Comparison   | - <u>•</u> Sign Fest  | witz Kuns Test                                 | >                 |
|   |  |   |  |                   |

Puis nous insérons le composant UNIVARIATE CONTNUOUS STAT (onglet STATISTICS). Nous cliquons sur le menu VIEW.

| 💯 TANAGRA 1.4.19 - [Uni   | ivariate continuous : | stat 1]  |                       |                     |            |            |                |                      |       |  |
|---------------------------|-----------------------|--|-----------------------|---------------------|------------|------------|----------------|----------------------|-------|--|
| Tile Diagram Component    | Window Help           |  |                       |                     |            |            |                |                      | - 8 × |  |
| 🗅 📽 🔚   👪                 |                       |  |                       |                     |            |            |                |                      |       |  |
| Analysis                  |                       | Univariate continuous stat 1                           |                       |                     |            |            |                |                      |       |  |
| 🖃 🏢 Dataset (tan22.txt)   |                       |  |                       |                     |            |            | Parameters     |                      |       |  |
| 🖃 🚰 Define status 1       |                       | Attributes   | :2                    |                     |            |            |                |                      |       |  |
| Univariate dis            | screte stat 1         | Examples :   | 20                    |                     |            |            |                |                      |       |  |
| Define status 2           | Chi-Square 1          |  |                       |                     |            |            | Doculto        |                      |       |  |
| Lipivariate co            | optipuous stat 1      |  |                       |                     | -          |            | Results        | 2                    |       |  |
|                           |                       | Attribute  | Min                   | Max                 | Average    | Sta-dev    | Std-deviavg    |                      |       |  |
|                           |                       | NDENTANTS  | 0                     | 5                   | 2.7000     | 1.3803     | 0.5112         |                      |       |  |
|                           |                       | Revenu   | 9000                  | 120000              | 49050.0000 | 30525.7010 | 0.6124         | J                    |       |  |
|                           |                       | Computatio<br>Created at                               | n time<br>19/06.      | : 0 ms,<br>(2007 1: | 3:41:40    |            |                |                      |       |  |
|                           |                       |  |                       |                     | Component  | :s         |                |                      |       |  |
| Data visualization        | Statistics            | Nor  | nparan                | netric s            | tatistics  | Instanc    | e selection    | Feature construction |       |  |
| Feature selection         | Regression            |  | Facto                 | rial ana            | lysis      |            | PLS            | Clustering           |       |  |
| Spv learning              | Meta-spv learning     | Spv  | learn                 | ing ass             | essment    | So         | coring         | Association          |       |  |
| 🛱 Bartlett's test         | 🔚 Group explorati     | on   | ۸                     | ormalit             | y Test     | Liite      | T-Test         |                      |       |  |
| 🚰 Brown - Forsythe's test | 🕁 Levene's test       |  | , <mark>M</mark> il C | Ine-way             | / ANOVA    | Luiu.      | T-Test Unequ   | ual Variance         |       |  |
| Fisher's test             | 🖄 Linear correlati    | lation 🔄 One-way MANOVA 📉 🖾 Univariate continuous stat |                       |                     |            |            |                |                      |       |  |
| Group characterization    | 🗥 More Univariate     | cont stat  | <u>14</u> P           | aired T             | -lest      | LUL        | Univariate dis | crete stat           |       |  |
|                           |                       |  |                       |                     |            |            |                |                      |       |  |
|                           |                       |  |                       |                     |            |            |                |                      |       |  |

Ce composant propose quelques indicateurs simples telles que la moyenne, le min, l'écart type. Il est adapté lorsque nous voulons calculer ces indicateurs sur un très grand nombre de variables et donne une idée très simplifiée de la structure des données. Par exemple, si l'écart type est égal à zéro ou, c'est la même chose, si le min est égal au max, nous pouvons affirmer que la variable n'est composée que d'une seule valeur, c'est une constante.

Un second outil fournit des informations détaillées. Nous insérons le composant MORE UNIVARIATE CONTNUOUS STAT (onglet STATISTICS) à la suite du DEFINE STATUS 2. Nous visualisons les résultats.

| 💯 TANAGRA 1.4.19 - [M                 | ore Univariate cont                         | stat 1]      |                              |                      |                             |             |             | (          |       |
|---------------------------------------|---|--------------|------------------------------|----------------------|-----------------------------|-------------|-------------|------------|-------|
| Tile Diagram Componen                 | t Window Help                               |              |                              |                      |                             |             |             |            | _ 8 X |
| D 📽 🖪  🎎                              |   |              |                              |                      |                             |             |             |            |       |
| Analysis                              |   | Attribute    | S                            | tats                 |                             |             | ^           |            |       |
| 🖃 🥅 Dataset (tan22.txt)               |   |              | Sta                          | listics              | Values                      | Count       | Percent     | Histogram  |       |
| 🖃 🛟 Define status 1                   |   |              | Average                      | 2.7000               | x_<_0.5000                  | 1           | 5.00%       |            | -     |
| - III Univariate d                    | Ontingency Chi-Square 1                     |              | Median                       | 3.0000               | 0.5000_=<_x_<_1.0000        | 0           | 0.00%       |            |       |
| Define status 2                       | y chi-square i                              |              | Std dev. [Coef of variation] | 1.3803 [0.5112]      | 1.0000 =< x < 1.5000        | 3           | 15.00%      |            |       |
| - K Univariate o                      | continuous stat 1                           |              | MAD [MAD/STDDEV]             | 1,1300 [0,8187]      | 1 5000 -< x < 2 0000        | -           | 0.00%       |            |       |
| - 🕂 More Univa                        | riate cont stat 1                           |              | Min*Max [Full range]         | 0.00 * 5.00 [5.00]   | 0.0000_01_2.0000            | -           | 0.00%       | _          |       |
| 1                                     |   | NbEnfants    | 1st*3rd quartile [Range]     | 2.00 * 4.00 [2.00]   | 2.0000_=<_x_<_2.5000        | 5           | 25.00%      |            |       |
| /'                                    |   |              | Skewness (std-dev)           | -0.0640 (0.5121      | 2.5000_=<_x_<_3.0000        | 0           | 0.00%       | (          |       |
|                                       |   |              | Kurtosis (std-dev)           | -0.6171 (0.9924      | 3.0000_=<_x_<_3.5000        | 5           | 25.00%      |            |       |
|                                       |   |              |                              |                      | 3.5000_=<_X_<_4.0000        | 0           | 0.00%       |            |       |
|                                       |   |              |                              |                      | 4.0000_=<_x_<_4.5000        | 4           | 20.00%      |            |       |
|                                       |   |              |                              |                      | x>=_4.5000                  | 2           | 10.00%      | <u> </u>   | ~     |
|                                       |   | <            |                              |                      | Ш                           |             |             |            |       |
|                                       |   |              |                              | Components           |                             |             |             |            |       |
| Data visualization                    | Statistics                                  | Nor          | parametric statistics        | Instance selection   | Feature construction        | Featur      | e selection | Regression |       |
| Factorial analytis                    | PLS   |              | Clustering                   | Spv learning         | Meta-spv learning           | Spv learnii | ng assessme | nt Scoring |       |
| Association                           |   |              |                              |                      |                             |             |             |            |       |
| 🕁 Bartlett's test 🔠 Group exploration |   |              | 🕂 Normality Test             | 🗽 T-Test             |                             |             |             |            |       |
| 🚰 Brown - Forsythe's test             | 🚰 Brown - Forsythe's text 🛛 🚟 Levene's test |              | 📶 One-way ANOVA              | 🌆 T-Test Unequal V   | ariance                     |             |             |            |       |
| Fisher's test                         | Fisher's test                               |              | 🔄 One-way MANOVA             | KUnivariate contin   | 述Univariate continuous stat |             |             |            |       |
| Group characterization                | K More Univaria                             | te cont stat | Paired T-Test                | 🎹 Univariate discret | e stat                      |             |             |            |       |
|                                       |   |              |                              |                      |                             |             |             |            |       |
|                                       |   |              |                              |                      |                             |             |             |            |       |

En plus des indicateurs précédents, de nouveaux indicateurs sur la forme de la distribution sont proposés : les coefficients d'asymétrie et d'aplatissement, etc. Un histogramme de fréquence permet de préciser la répartition des données.

La largeur et le nombre des intervalles sont automatiquement définis par TANAGRA, il n'est pas possible de les modifier pour afficher des histogrammes avec des amplitudes inégales entre autres.

#### Statistiques bivariées - Croisement de variables quantitatives

#### Graphique nuage de points

A la différence des variables qualitatives, il existe des outils graphiques pour la visualisation des couples de variables quantitatives dans TANAGRA. Nous voulons par exemple le lien qui pourrait exister entre le nombre d'enfants et le revenu des personnes.

Nous plaçons en dessous du DEFINE STATUS 2 le composant SCATTERPLOT (onglet VISUALIZATION). En cliquant sur VIEW, le nuage de point apparaît, il est possible de modifier à la volée les variables en abscisse et en ordonnée. Pour une bonne visualisation, la taille des points dans le graphique est paramétrable.



Lorsque le REVENU augmente, les femmes ont tendance à avoir un plus grand nombre d'enfants. A moins que ce soit l'inverse. C'est bien là d'ailleurs les limites de la statistique. Nous pouvons observer des régularités. Après, les transformer en causalité repose en grande partie sur l'expert du domaine, le sociologue, l'économiste, etc.

#### Indicateurs numériques

Pour analyser la relation entre deux variables continues, l'outil graphique est à privilégier car le lien peut receler des situations très complexes. Il existe néanmoins des indicateurs numériques qui peuvent être intéressants lorsqu'un grand nombre de variables est en jeu.

Parmi les outils disponibles figure le coefficient de corrélation (LINEAR CORRELATION, onglet STATISTICS). Nous l'insérons à la suite du DEFINE STATUS 2, nous indiquons dans le paramétrage que les variables sont toutes en INPUT.

| TANAGRA 1.4.19 - [Scatterplot 1]   |  |              |
|------------------------------------|--|--------------|
| Tile Diagram Component Window Help |  | . <i>B</i> > |
| 0 📽 🖬 🙀                            | Correlation options                                      |              |
| Analysis                           |  | ~            |
| 🖃 🥅 Dataset (tan22.txt)            | Parameters   |              |
| 🖃 🎎 Define status 1                | Sort results   |              |
|                                    | C Sort by  |              |
| Contingency Chi-Square 1           | Y attribute name   |              |
| li liniupristo continuous stat 1   | OVer the second  | *            |
| - I More Univariate cont stat 1    | O X attribute hame                                       |              |
| Scatterplot 1                      | Or-value   |              |
| Linear correlation 1               | ◯  r -value  |              |
|                                    | Input list<br>Target and Input<br>O Cross Input          | 120 000      |
| Data visualization Statistics      |  |              |
| Feature selection Regression       |  |              |
| Spv learning Meta-spv learning     | OK Cancel Help   |              |
| Bartlett's test                    |  | 1            |
| Brown - Forsythe's test            | 📶 One-way ANOVA 🗽 T-Test Unequal Variance                |              |
| Group obstacterization             | · (한국) Une-way ///ANUVA / 전국) Univariate continuous stat |              |
|                                    | usua emeranea i-rest mutumanate discrete stat            |              |
|                                    |  |              |

Les résultats montrent alors que le lien n'est pas à négliger, le coefficient de corrélation est égal à 0.4298. Plus ce coefficient s'éloigne de 0 en valeur absolue, plus forte est la liaison linéaire. La valeur maximale est 1 (en valeur absolue).



Si la liaison n'est pas linéaire mais reste monotone, nous pouvons utiliser les coefficients de SPEARMAN et KENDALL (onglet NONPARAMETRIC STATISTICS).

## Croisement de variables qualitatives et quantitatives

Croiser des variables quantitatives et des variables qualitatives peut emmener d'autres types d'informations. Dans notre exemple, essayons de comprendre la satisfaction à travers le nombre d'enfant de la personne. Est-ce qu'il y a un lien entre ces deux variables ?

Nous utiliserons l'outil GROUP CHARACTERIZATION pour cela. Insérons tout d'abord un troisième DEFINE STATUS dans le diagramme. Nous plaçons en TARGET la variable OPINION, et en INPUT NBENFANTS.



Nous plaçons alors le composant GROUP CHARACTERIZATION (onglet STATISTICS).

| 💯 TANAGRA 1.4.19 - [Grou   | p characterization 1]   |  |                            |                            |                              |   |  |                           |  |  |  |
|----------------------------|-------------------------|--|----------------------------|----------------------------|------------------------------|---|--|---------------------------|--|--|--|
| Tile Diagram Component V   | Window Help             |  |                            |                            |                              |   |  | _ 8 ×                     |  |  |  |
| 🗅 📽 🖪   👪                  |                         |  |                            |                            |                              |   |  |                           |  |  |  |
| Analysis                   |                         | Opinion-   | faunaice                   | Opinion                    | -Très bopp                   | Opinion-                                  | Opini                                    |                           |  |  |  |
| 🖃 🏢 Dataset (tan22.txt)    | Ev                      | amples   | 120 0 %1 4                 | Examples                   | 110.0 %1                     | 2 Examples                                | 130.0 %1 4                               | Evamples                  |  |  |  |
| 🖃 🚰 Define status 1        |                         | Test   | [20.0 /0] 4                | Examples<br>Te             | -+                           | Z   | [ 30.0 %] 0                              | Examples                  |  |  |  |
| - 🛄 Univariate disci       | rete stat 1 Ati         | t - Desc value   | Group Overral              | Att - Desc va              | st Group Overr<br>lue        | al Att-Desc value                         | Group Overral                            | Att - Desc val            |  |  |  |
| E Contingency of           | Co<br>(St               | ntinuous attril<br>:dDev)                              | outes : Mean               | Continuous at<br>(StdDev)  | tributes : Mean              | Con <mark>i</mark> nuous attr<br>(StdDev) | ibutes : Mean                            | Continuous at<br>(StdDev) |  |  |  |
| - More Univariate          | e cont stat 1           | Enfants -0.7   | 2.25 2.70<br>(1.50) (1.38) | NbEnfants -(               | 0.7 2.00 2.7<br>(2.83) (1.38 | 0<br>() OEnfants 1.0                      | 3.17 2.70<br>(1.17) (1.38)               | NbEnfants C               |  |  |  |
| Scatterplot 1              | ion 1 Ac                | , rete attribut<br>curac <mark>y</mark>                | tes : [Recall]             | Discrete attri<br>Accuracy | ibutes : [Recar              | Discrete attribu<br>Accuracy              | iscrete attributes : [Recall]<br>couracy |                           |  |  |  |
| 🖻 🚰 Define status 3        |                         |  |                            |                            |                              |   |  |                           |  |  |  |
| Group characte             | erization 1             | Computation time : 0 ms.                               |                            |                            |                              |   |  |                           |  |  |  |
|                            | Crea                    | Created at 19/06/2007 13:56:10                         |                            |                            |                              |   |  |                           |  |  |  |
| <u></u>                    |                         |  |                            |                            |                              | J   |  | 2                         |  |  |  |
|                            |                         |  | Compo                      | nents                      |                              |   |  |                           |  |  |  |
| Data visualization         | Statistics              | Nonparam   | etric statistics           | Instanc                    | e selection                  | Feature const                             | ruction                                  |                           |  |  |  |
| Feature selection          | Regression              | Factor   | ial analysis               |                            | PLS                          | Clusterin                                 | Ig.                                      |                           |  |  |  |
| Spv learning               | Meta-spv learning       | Spv learnii  | ng assessment              | Sc                         | oring                        | Associati                                 | on                                       |                           |  |  |  |
| 🛱 Bartlett's test          | 🔚 Group exploration     | 🕂 No   | ormality Test              | ենել                       | T-Test                       |   |  |                           |  |  |  |
| 🔆 Brown - Forsy the's test | 🕁 Levene's test         | 📶 Or   | ne-way ANOVA               | Luiu,                      | T-Test Unequal               | Variance                                  |  |                           |  |  |  |
| 🖾 Fisher's test            | 🖄 Linear correlation    | Or 🛃   | ne-way MANOV               | A 😾                        |                              |   |  |                           |  |  |  |
| Group characterization     | ik™More Univariate cont | cont stat ≟ Paired T-Test III Univariate discrete stat |                            |                            |                              |   |  |                           |  |  |  |
|                            |                         |  |                            |                            |                              |   |  |                           |  |  |  |

Nous constatons essentiellement que les femmes ayant une opinion très tranchée, bonne ou mauvaise, ont en moyenne moins d'enfants (resp. 2.25 et 2.0) que la population globale (2.7).

Bien entendu, tout ceci n'est qu'un exercice de style, les effectifs sont trop faibles pour conclure à quoique ce soit avec ces données.

# Conclusion

Dans la présentation des résultats d'un traitement de données, des indicateurs basiques, des tableaux et des graphiques simples sont au moins aussi souvent pertinents, tout du moins persuasifs, que les méthodes statistiques compliquées, assez obscures pour les non-spécialistes. C'est pour cette raison que les statistiques descriptives tiendront toujours une place prépondérante dans les rapports.