

MODALISA

Création, analyse de questionnaires et d'entretiens pour
MAC OS Classic & Windows 95 et suivants
Version 4.2

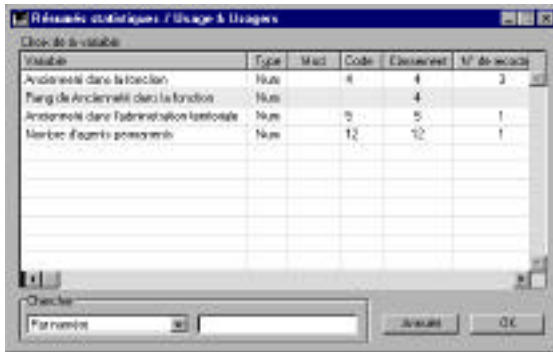


Variables numériques

Vous trouverez dans cette section les informations nécessaires à la réalisation d'analyse des variables numériques. Les recodages sont décrits dans l'opuscule Classements et recodages.

RESUMES STATISTIQUES

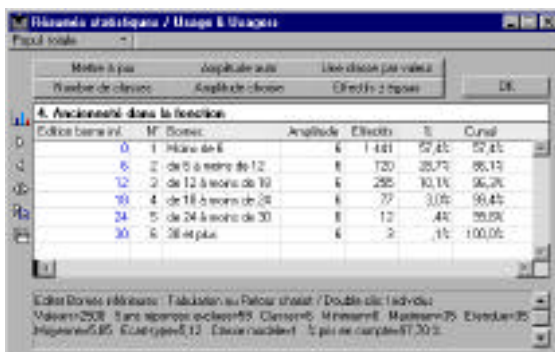
Présentation des questions numériques en classes



Cette option ne permet pas de créer un recodage, mais elle offre des possibilités qui ne sont pas actives sur le format de recodage en classes :

Les boutons variable suivante, variable précédente, atteindre, graphique, copier, dupliquer l'écran sont actives.

Le menu local sous-opulation permet de sélectionner une sous population parmi ceux existantes dans la base.



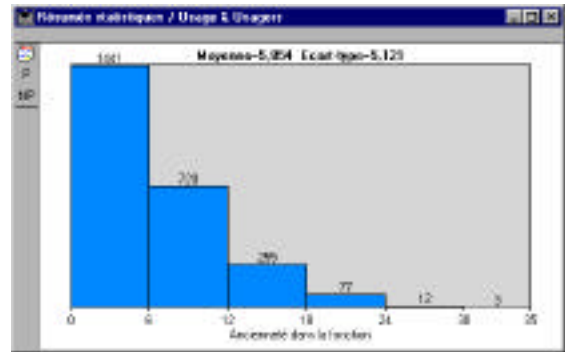
L'utilisateur peut choisir le nombre de classes, l'amplitude, une classe par valeur,...

Dans l'encadré en bas de la fenêtre ci-dessus, Modalisa donne les valeurs de la moyenne, le maximum, le minimum, l'écart-Type, l'étendue, le nombre de classes,...

L'activation de l'icône graphique fait apparaître une fenêtre de dialogue permettant à l'utilisateur de choisir le type de graphique désiré.

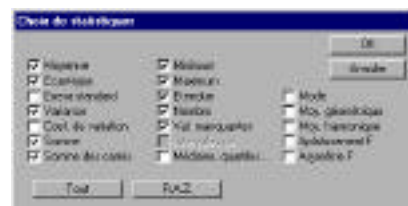


Le choix 4 (Effectifs pour chaque classe) donne le graphique ci-dessous.



STATISTIQUES SUR UNE VARIABLE

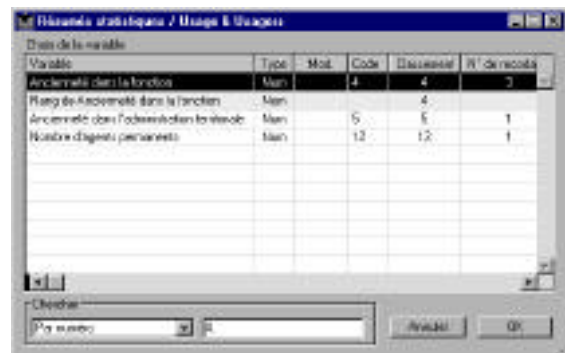
Les formules de calcul des résumés statistiques sont développées en fin de documentation.



Cet écran présente toutes les statistiques disponibles et permet de sélectionner les statistiques à effectuer.

Par défaut, la plupart des traitements sont sélectionnés, mais le bouton "R.A.Z." permet de remettre à zéro les différentes sélections.

Choisir la variable numérique à traiter.

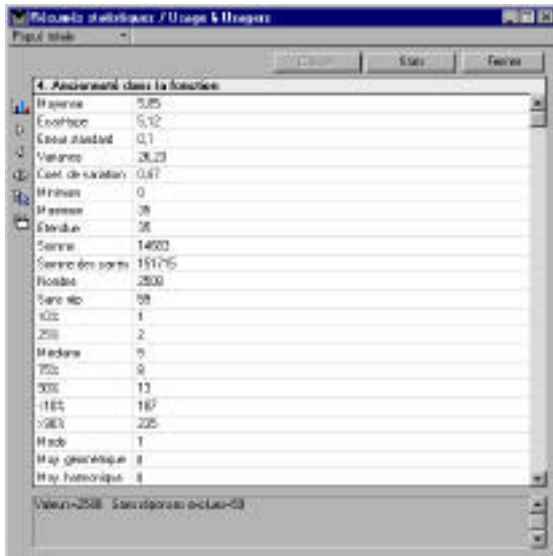


Modalisa™ présente un tableau comportant les résumés statistiques demandés.

Le popup Population totale permet d'obtenir les calculs sur une sous population sélectionnée.

D'autres options sont accessibles par les boutons situés sur le côté gauche de l'écran à savoir : graphique, variable suivante, variable précédente, atteindre, copier et dupliquer la fenêtre.

COMPARAISON DES RESUMES STATISTIQUES DE N VARIABLES

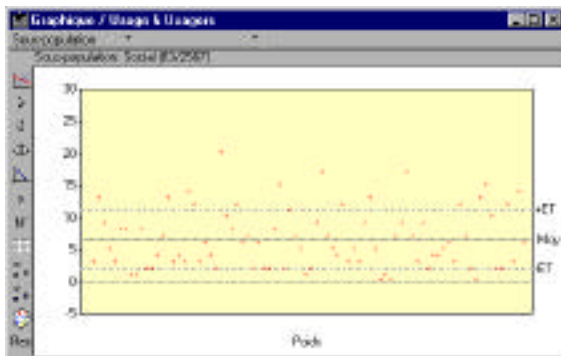


Le bouton « graphique » active l'écran suivant :

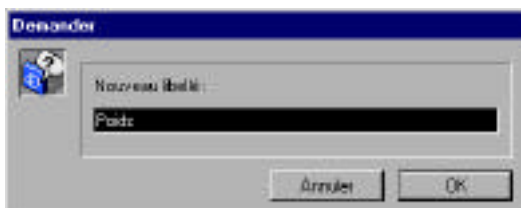


3 types de graphiques peuvent être réalisés : "Moyenne et \pm Ecart-type", "boîte de dispersion" et "Effectifs pour chaque valeur de la variable".

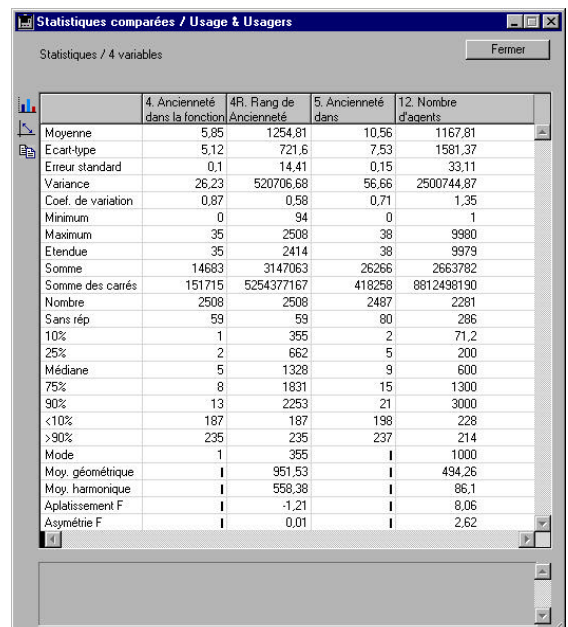
Graphique situant les individus relativement à la moyenne et +/- l'écart-type.



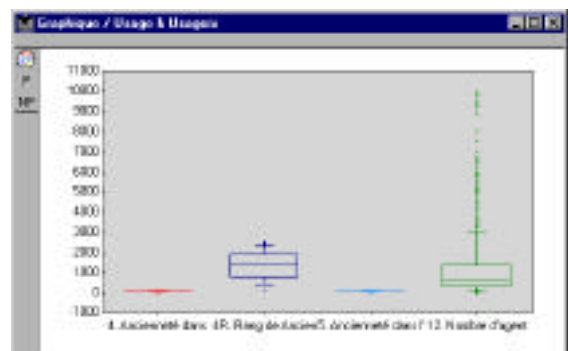
Pour changer de libellé, il suffit de cliquer deux fois sur l'intitulé de la variable.



Sélectionner les questions (variables numériques) à traiter. Modalisa effectue les calculs et les présente sous la forme d'un tableau.



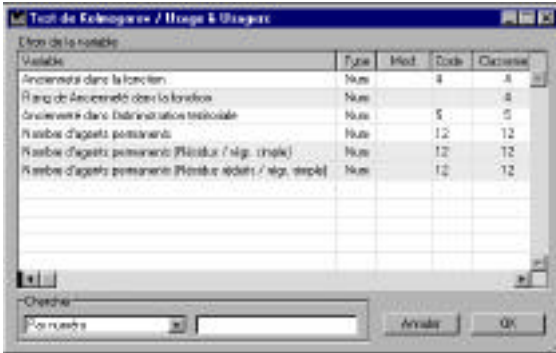
Le bouton graphique présente les boîtes de dispersion correspondante aux variables sélectionnées.



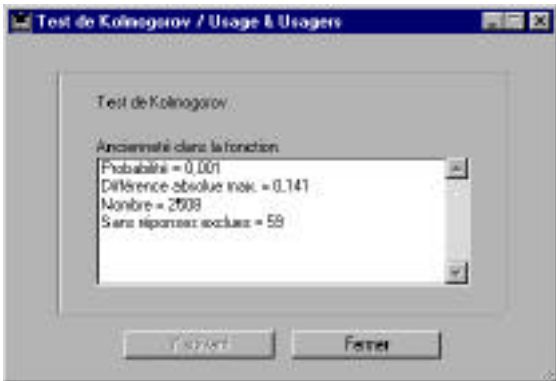
L'utilisateur peut transposer les lignes et les colonnes et peut aussi copier le tableau.

TESTS NON PARAMETRIQUES

Test de Kolmogorov-Smirnov pour 2 échantillons indépendants



Les deux échantillons sont constitués par les individus caractérisés par les deux modalités d'une question à réponse unique.



Edition:

Variables choisies

Différence absolue maximum entre les répartitions

- Effectif de chaque sous-population
- Effectif des non réponses, le cas échéant.
- Sous-population courante, le cas échéant.

Ces valeurs permettent de lire la table de Smirnov (cf MASSEY F.J., Distribution table for the deviation between two sample cumulatives. The Annals of Mathematical Statistics, 23, pp.435-441)

Test de Kruskal-Wallis

Dans la version précédente de Modalisa, la procédure ne prenait pas en compte l'ajustement nécessaire lorsque le nombre de rangs moyens (ex aequos) est élevé. Dans ce cas, le résultat du test est désormais divisé par : $1 - \frac{(t_i^3 - t_i)}{N^3 - N}$ où t_i est la taille du groupe moyen i . (Cf. Dodge pp 353-355).

N.B. La valeur de « p » est approximée selon la

distribution de la table du Chi2 avec k-1 degrés de liberté (k=nombre d'échantillons). Cf. Dodge, p.354.

LEXIQUE

En raison du caractère très « technique » de cette section, nous avons fait figurer ci-dessous un rappel des éléments de lexique élémentaires.

- Analyse Factorielle des Correspondances ou AFC: Technique d'analyse statistique d'un ou de plusieurs tableaux de contingences permettant une représentation graphique des attractions et des distances entre les modalités des variables choisies.

- Analyse factorielle en Composantes Principales ou ACP: Technique d'analyse statistique portant sur au moins deux variables numériques permettant notamment une représentation graphique des corrélations entre les variables et de la position des individus par rapport aux vecteurs des variables.

- Ascendante hiérarchique indicée (typologie): Mode de calcul d'une classification automatique (typologie) :

. Ascendante : on part des individus pour terminer à l'ensemble des individus. Hiérarchique : chaque association constitue un "noeud" dont la position dans l'arbre (dendrogramme) correspond au moment où l'agrégation a été effectuée.

. Indicée : sur le dendrogramme, la distance relative entre deux noeuds de l'arbre représente la distance entre leurs centres de gravité calculés selon leurs coordonnées sur les axes factoriels.

- Boîte de dispersion: Représentation graphique de résumés numériques affichant les positions relatives de la médiane, de l'écart-type inférieur et supérieur à la médiane, et du premier et du neuvième décile. Et le cas échéant, affichage des valeurs supérieures au neuvième décile et inférieures au premier.

- Booléen (fonction ou argument): Voir Logique.

- Booléenne (variable): Variable fermée à deux modalités (oui/non) créée à partir d'une sous-population. La modalité 'oui' représente les individus appartenant à la sous-population.

- Classe (numérique): Regroupement d'observations (individus) selon leur appartenance à une plage définie par un minimum et un maximum. Selon le choix de l'utilisateur les valeurs minimum et maximum sont incluses ou non dans la classe.

- Classe modale: Numéro de la classe incluant le plus grand nombre d'individus.

S'il y a plus d'une classe de ce type la valeur 0 est affichée.

-
- Classement Manipulation des données brutes en vue d'en favoriser le traitement.
 - Coefficient d'aplatisse-ment de Fisher: Mesure permettant d'évaluer la forme [normale (=0), aplatie (<0) ou aiguë (>0)] de la distribution des valeurs d'une variable par rapport à une distribution théorique. Cf. JAMBU (1989, pp. 157-177)
 - Coefficient d'asymétrie de Fisher: Mesure permettant d'évaluer la forme [symétrie (=0), étalement à gauche (<0) ou à droite(>0)] de la distribution des valeurs d'une variable par rapport à une distribution théorique. Cf. JAMBU (1989, pp. 157-177)
 - Coefficient de corrélation (PEARSON): Indicateur permettant de juger la qualité de la liaison entre deux variables numériques. Une valeur positive (maximum :+1) indique une variation simultanée dans le même sens et une valeur négative (minimum -1), une variation simultanée en sens inverse.
 - Coefficient de corrélation des rangs (SPEARMAN): Coefficient de corrélation calculé sur les numéros d'ordre des valeurs de deux variables.
 - Coefficient de détermina-tion (r²): Carré du coefficient de corrélation. Pour une régression linéaire
ce coefficient permet d'évaluer la part de la variance de la variable
Y expliquée par la régression.
 - Coefficient de redresse-ment: Rapport du pourcentage souhaité au pourcentage observé. Ce coefficient est appliqué pour le calcul des effectifs sur les variables logiques chaque fois qu'un individu est concerné par la ou les modalités faisant l'objet du redressement.
 - Coefficient de variation: Indicateur de dispersion utilisé dans l'analyse d'une variable numérique, égal au rapport de l'écart-type à la moyenne.
 - Constante: Pour les variables numériques désigne une valeur stable prise en compte lors des calculs sur les données de chaque questionnaire.
 - Contingences (Tableau de): Tableau à deux entrées représentant pour chaque case les effectifs corres-pondant aux modalités croisées de deux variables
 - Contribution absolue: Indice utilisé en Analyse factorielle . Cet indice résume le poids d'une modalité (ou variable) par rapport à l'ensemble des modalités (ou variables) expliquées par un facteur.
 - Contribution relative: Indice utilisé en Analyse factorielle. Cet indice résume le poids d'une modalité (ou variable) par rapport à l'ensemble des modalités (ou variables) expliquées relativement à l'ensemble des facteurs.
 - Cooccurrence: Fréquence d'apparition de deux termes (ou expressions) dans le même questionnaire.
 - Covariance: Moyenne du produit des écarts de deux variables à leur moyenne respective.
 - Critère: Elément discriminant
 - Décile: Valeurs (9) d'une variable numérique divisant la population en dix groupes égaux.
 - Définitions pour l'expression: Affectation de variables ou de résumés statistiques à des lettres qui seront utilisées dans une formule de calcul.
(Voir Formule de calcul)
 - Degrés de liberté: Nombre d'informations non redondantes nécessaires et suffisantes pour compléter sans ambiguïté une série de valeurs dont on connaît tel ou tel paramètre (par exemple : les marges d'un tableau de contingence).
 - Dendrogramme: Présentation graphique d'une typologie ascendante hiérarchique indicée.
 - Discriminante (variable): Le programme permet de classer les questions/variables selon leur caractère discriminant d'après le critère de l'écart à l'indépendance par un inventaire systématique des écarts positifs à l'indépendance, du Khi² partiel et du test du Khi² entre toutes les questions choisies.
Principe : Plus la somme des écarts inventoriés et/ou la somme des khi² partiels et/ou le nombre de tests du khi² significatifs pour une variable sont élevés, plus la variable a de chances d'être discriminante relativement au groupe des varia-bles choisies pour l'analyse.
Voir aussi Ecart à l'indépendance et Profil de modalités.
 - Droite de régression linéaire: Représentation graphique d'une équation de régression linéaire. Synonyme en pratique de la 'Droite des moindres carrés'.
 - Droite des moindres carrés: Equation (ou sa représentation graphique) résumant la liaison entre deux variables numériques de telle sorte que le cumul des écarts positifs des valeurs des variables à la droite soit rigoureusement compensé par le cumul des écarts négatifs.
La somme des carrés des écarts à la droite est minimale.
 - Ecart Type: Racine carrée de la variance
 - Écarts à l'indépendance: Pour un tableau de contingences un écart à l'indépendance représente pour chaque case la différence entre les effectifs observés et les valeurs théoriques.
 - Écarts non pondérés: Cf. Écarts à l'indépendance
 - Écarts pondérés: Cf. Khi² par case
 - Effectifs: Nombre d'individus caractérisés par une modalité d'une variable logique ou comptabilisés dans une classe d'une variable numérique.
 - Enquête: Terme générique regroupant l'ensemble des éléments d'une étude : Questionnaire et données.
-

- Étendue: Différence entre la plus grande et la plus petite valeur d'une variable numérique.

- Facteur: Dans une analyse factorielle un facteur représente une des étapes de l'extraction des valeurs et vecteurs propres. Représente aussi un des axes de visualisation des résultats.

- Facteurs à extraire: Pour une analyse factorielle l'utilisateur peut choisir le nombre de facteurs à extraire et à représenter dans la limite des contraintes du type de données analysées et du mode d'analyse choisi. Voir aussi : AFC, ACP et le chapitre 'Formules'.

- Fiable (test du Khi2): Pour que le test du Khi2 sur un tableau de contingences soit fiable au niveau de signification choisi, il faut que toutes les valeurs du tableau soient au moins éga-les à cinq.

- Formule (de calcul):

. Au sens restreint: fonction opérant soit entre deux variables, soit entre une variable et une constante.

. Au sens large : Combinaison d'opérations arithmétiques et de fonctions mathématiques, statistiques et logiques. Les arguments peuvent être des constantes (numériques ou dates), des variables numériques ou de type texte (pour les dates) et des sous-populations.

- Grapheur: Programme (externe ou Module interne à Modalisa™) permettant de réaliser des graphiques en deux ou trois dimensions.

- Henry (droite de): Présentation graphique permettant d'évaluer la "normalité" d'une distribution.

. Principe : la courbe de fréquence cumulée est une droite si la distribution statistique suit une loi Normale.

- Intervalle de confiance (sur une moyenne): Marge d'erreur lorsque on veut attribuer à une population une moyenne calculée sur un échantillon représentatif.

Calcul pour un risque de 5 % :

$$\pm 1,96 s / \text{Racine carrée} (n)$$

Pour $s =$ Ecart-type, $n =$ nombre d'individus

- Intervalle de confiance (sur une proportion): Marge d'erreur lorsque on veut attribuer à une population une proportion (pour-centage) observée sur un échantillon représentatif.

Calcul pour un risque de 5 % :

$$\pm 1,96 \text{ Racine carrée} (f (1-f) / n)$$

Pour $f =$ fréquence (de 0 à 1), $n =$ nbre d'individus.

- Khi2: Pour un tableau de contingences (tri croisé) le Khi2 global est calculé en faisant la somme des Khi2 par cases, c'est-à-dire des écarts à l'indépendance pondérés par rapport à la marge ligne et colonne du tableau.

- Khi2 par case: Le Khi2 par case représente la proportion du Khi2 au croisement d'une modalité

colonne avec une modalité ligne par rapport au Khi2 global d'un tableau de contingences (tri croisé). Il s'agit d'une pondération de l'écart à l'indépendance relativement à la marge ligne et à la marge colonne du tableau.

Mapping: Map = carte en anglais. Représentation graphique des attractions et des distances entre les éléments constitutifs d'une Analyse factorielle : modalités, variables, individus...

- Marge: Pour un tableau de contingences (tri croisé) les marges correspondent aux totaux lignes et colonnes.

- Matrice de corrélation: Tableau symétrique présentant les coefficients de corrélation entre les variables numériques analysées.

- Médiane: Valeur d'une variable numérique divisant la population en deux effectifs égaux.

- Moyenne (arithmétique): Quotient de la somme des valeurs d'une variable par leur nombre. Dans ce lexique 'moyenne' est équivalent de 'moyenne arithmétique'.

- Moyenne et \pm écart-type (graphique): Graphique représentant la position des individus relativement à la moyenne plus (+) un écart-type et à la moyenne moins (-) un écart-type.

- Moyenne géométrique: Moyenne égale à la racine $n^{\text{ème}}$ du produit des valeurs de la variable. Principalement utilisée dans l'analyse de la croissance d'un phénomène.

- Moyenne harmonique: Inverse de la Moyenne arithmétique des inverses ($1/X$)

Des valeurs d'une variable.

- Multi-process: Voir Multitâche.

- Multi-thread: Voir Multitâche.

- Multitâche: Au sens strict (appelé parfois multi-process ou multi-thread), le multitâche intra-application permet l'exécution simultanée de tâches diverses au sein d'une même application.

Dans Modalisa, multitâche et multi-fenêtres sont activés/inactivés simultanément. Quand le multitâche est activé chaque tâche s'exécute dans une nouvelle fenêtre.

- Normale (loi): Pour un échantillon de taille suffisamment élevée, la distribution des valeurs autour de la moyenne tend à suivre une courbe en forme de cloche (courbe de Gauss). Cette distribution sert de référence pour évaluer notamment les inter-valles de confiance sur les moyennes et les proportions.

- Numérique (Question ou variable ou donnée): Données susceptibles de faire directement l'objet d'une mesure. 'Variable numérique' est équivalent de 'Variable quantitative'.

- Occurrence de mots: Fréquence d'apparition d'un terme (ou expression).

- Pearson: Cf. coefficient de corrélation

- P.E.M. local (Pourcentage de l'écart maximum): Pour un case d'un tableau de contingences : rapport entre l'écart maximum et l'écart observé.

Le calcul de la liaison maximum consiste à mettre dans la case du tableau croisé "l'effectif le plus fort possible qui soit compatible avec les marges du tableau et à calculer l'écart à l'indépendance maximum qu'il y aurait dans ce cas. La force de la liaison sera simplement calculée en examinant le rapport entre l'écart à l'indépendance observé et l'écart maximum à l'indépendance calculé précédemment. Ce rapport étant mis en pourcentages, on parlera donc de Pourcentage de l'Ecart Maximum ou P.E.M.." Ph. Cibois, 1993, p.45.

- P.E.M. global (Pourcentage de l'écart maximum): Généralisation du P.E.M. local pour l'ensemble d'un tableau de contingences. Pour le mode de calcul du P.E.M. global, voir Ph. Cibois, 1993, p.59-62.

- PICT: Format des graphiques (AFC, ACP, régressions...) copiés sur demande vers le presse papier, récupérables sur des programmes lisant ce format (MacDraw). Les éléments constitutifs de ces graphiques conservent leurs indépendance (droi-tes, points, libellés...)

- Plan (d'analyse ou de recodage): Fonction permettant d'automatiser une série de traitements. L'exécution d'un plan peut être différée.

- Pourcentage de l'écart maximum: Voir P.E.M..

- Pourcentage du Khi2 par case: Pour le Khi2 global calculé sur un tableau de contingences (tri croisé) le % du Khi2 par cases représente la proportion du Khi2 calculé pour la case correspondante.

- Pourcentage sur les interrogés: Pour une question à réponse multiples pourcentage réalisé sur l'ensemble des individus ayant répondu à l'enquête.

- Pourcentage sur les répondants: Pour une question à réponse multiples pourcentage réalisé sur le total des individus ayant fournis au moins une réponse.

- Pourcentage sur les réponses: Pour une question à réponse multiples pourcentage réalisé sur le total des réponses effectivement données. Le total peut donc être inférieur ou le plus souvent supérieur à 100 %.

- Pourcentage en colonne: Pourcentage des effectifs calculés sur la somme de chaque colonne

- Pourcentage en ligne: Pourcentage des effectifs calculés sur la somme de chaque ligne

- Profil de modalité: Inventaire de l'ensemble des écarts à l'indépendance pondérés ou non pondérés pour chaque modalité d'une variable logique.

- Quantiles: Valeurs (n-1) d'une variable numérique divisant la population en n groupes égaux.

- Quartiles: Valeurs (3) d'une variable numérique divisant la population en 4 groupes égaux.

- Quatrième Dimension: Environnement de programmation de base pour le développement de Modalisa™

- r^2 : Cf. Coefficient de détermination

- Rang: Pour une variable numérique le rang d'une valeur correspond à sa position dans la série préalablement triée. Le rang de plusieurs valeurs égales est la moyenne des rangs disponibles entre ceux des 2 valeurs inférieure et supérieure.

- Rang moyen: Pour les questions à réponses ordonnées : moyenne des notes obtenues pour chaque modalité de la question.

- Régression linéaire (simple): Modèle statistique permettant notamment de calculer l'équation de la droite des moindres carrés entre deux variables numériques.

- Sauvegarde: Fonction permettant d'exporter tout ou partie d'une enquête dans un fichier au format de Modalisa™.

- Score (variable):

. Variable numérique faisant la somme des coefficients attribués à une série quelconque de modalités.

. Variable numérique faisant la somme des coefficients attribués à une série de sous-populations.

- Seuil minimal: Valeur à partir de laquelle un écart à l'indépendance (pondéré ou non) est pris en compte. Et ce, soit au moment du calcul soit au moment de l'affichage.

- Somme des carrés: Somme des valeurs d'une variable multipliées par elles-mêmes.

- Spearman: cf. Coefficient de corrélation des rangs.

- Tableau de nombres (à importer): Option permettant, lors d'une AFC ou d'une ACP, d'importer directement des nombres répartis en lignes et en colonnes. Chaque valeur/colonne est séparée par une tabulation (code ASCII 9) chaque ligne se termine par un retour chariot (code ASCII 13).

- Tableau: Feuille de calcul de type Excel.

- Tâche: Ensemble de traitements exécutés lors du choix d'un article de menu dans la barre de menus principale. Un tâche se termine par la fermeture de la fenêtre affichant soit son état d'avancement, soit le résultat du traitement demandé.

Voir aussi Multitâche.

- Test non paramétrique: Test d'hypothèse reposant sur aucune connaissance a priori de la population-mère. L'hypothèse classique de la normalité de la population est ignorée. Ces tests sont à utiliser lorsque les variables sont ordinales ou lorsque l'hypothèse de normalité de la distribution n'est pas vérifiée. Test non paramétriques dans Modalisa : Mann-Whitney, Corrélation Spearman, Kruskal-Wallis, Friedman.

Voir aussi Coefficient de corrélation des rangs

- Tri: Signification propre au traitement statistique d'une enquête : dénombrement des effectifs pour les modalités d'une variable logique.

- Tri à plat: Ce dénombrement effectué pour les modalités d'une seule variable.

- Tri croisé: Ce dénombrement effectué pour les modalités croisées de 2 variables.

- Typologie: Regroupement d'individus dans des classes les plus hétérogènes entre elles par rapport aux variables prises en compte dans le calcul.

- Valeurs théoriques (ou Effectifs théoriques): Pour un tableau de contingences : effectifs calculés de telle sorte que chaque case représente à la fois la proportion du total colonne et du total ligne par rapport à la somme du tableau.

- Variable: Terme habituellement utilisé en statistiques pour désigner les valeurs observées ou théoriques pour un caractère (dans le cas d'une enquête ; réponses données à une question), dans la mesure où elles peuvent faire directement l'objet d'un comptage ou d'un calcul arithmétique ou statistique.

- Variable centrée réduite: Variable calculée dont la différence de ses valeurs à sa moyenne est divisée par son écart-type. Appelée aussi 'Variable normée'.

- Variable de contrôle: Lors d'un tirage aléatoire de questionnaires il est possible de choisir une question fermée et de déterminer pour chaque modalité de la question le pourcentage souhaité dans la sélection finale.

Cette option permet notamment d'opérer un redressement sur l'échantillon par duplication/suppression de questionnaires.

Voir aussi Aléatoire (tirage).

Voir aussi Redressement.

- Variance: Indicateur de dispersion d'une variable numérique égal à la moyenne des carrés des écarts à la moyenne.

Cf. chapitre "Description des formules utilisées" pour le calcul exact permettant d'estimer la variance de la population mère à partir de données recueillies sur un échantillon.

Kynos 3, rue des Montiboefus 75020 Paris

Tél : 33 (01) 40 30 23 23

Fax : 33 (01) 40 30 24 52

<http://www.modalisa.com>

email : jlvi@modalisa.com