

ANALYSE DES VARIABLES NUMÉRIQUES

MISE EN CLASSE AVEC SPHINX

Mise en classe

- * On va cependant regarder une dernière manipulation : la mise en classe.
- * Il s'agit de créer des classes pour des questions numériques ouvertes ou des variables quantitatives afin d'en faciliter l'étude.
—> **passage à une variable ordinale**

Exemple : il peut être intéressant de créer des catégories d'âge chez des individus ou encore de créer des catégories de personnes selon les revenus ou selon les dépenses effectuées dans une enseigne pour déterminer quelles personnes sont des bons clients.

- * Cela rejoint l'idée du regroupement des modalités pour une variable qualitative nominale/catégorique.
- * Pour cela, rendez-vous sur Sphinx et sélectionnez l'enquête « Automobiles »

Mise en classe

- * Rendez-vous dans l'onglet « Analyse » —> « Nouvelle Analyse » —> « Tris à plat » et sélectionnez la variable 24 : « Entretien2 »
- * Sphinx vous proposera une mise en classe par défaut, avec des choix qui peuvent se révéler inappropriés pour votre étude

24. ENTRETIEN2

	Effectifs	% Obs.
Moins de -87	0	0%
De -87 à 725	55	31,4%
De 726 à 1538	84	48%
De 1539 à 2351	22	12,6%
2352 et plus	14	8%
Total	175	100%

Réponses effectives : 175
Taux de réponse : 100%

Non-réponse(s) : 0
Moyenne : 1132 ; Médiane : 1000 ; Ecart-type : 812.74 ; Min - Max : 0 - 5000 ; Somme : 198100

Mise en classe

- * Vous pouvez tout à fait modifier la mise en classe selon vos besoins en cliquant sur « Utiliser la mise en classe » puis en cliquant sur l'engrenage se trouvant à côté

Tableau **Graphique**

Variable

24 - ENTRETIEN2

Utiliser la mise en classe

Options

Analyses

Effectifs

Pourcentages

Tests statistiques

Options de calcul

Ignorer les non-réponses

Affichage

Titre : Automatique

Afficher le tableau

Afficher le graphique

Afficher un commentaire personnalisé

Mise en classe : "ENTRETIEN2"

Mettre les nombres en classe

de même valeur

de même amplitude

autour de la moyenne

de même effectif

personnalisées

Classes de même effectif

6 Classes

OK Fermer

Mise en classe : les options

- * **De même valeur** : pas l'idéal quand l'idéal prend un nombre important de valeurs différentes
- * **De même amplitude** : permet de définir des classes dont les intervalles ont tous la même longueur - option : nombre de classes.
Intéressant, en pratique, lorsque l'on cherche à comparer des effectifs pour des classes avec des tailles identiques
- * **Autour de la moyenne** : construction des classes selon une gaussienne définie par son écart-type - options : écart-type et nombre de classes.
Une représentation intéressante quand la distribution des données est gaussienne
- * **De même effectif** : avoir des classes avec des effectifs égaux (quand cela est possible) - option : nombre de classes.
Permet de retracer la distribution des données, une façon indirecte de construire les quantiles
- * **Personnalisées** : vous laisse la possibilité de définir vous même les classes à étudier.

Mise en classe : exemple

- * Personnalisées : il faut simplement rentrer les valeurs qui définissent les bornes de chaque classe.
La borne inférieure est incluse dans la classe et la borne supérieure est exclue de la classe pour le calcul des effectifs.

Mise en classe : "ENTRETIEN2"

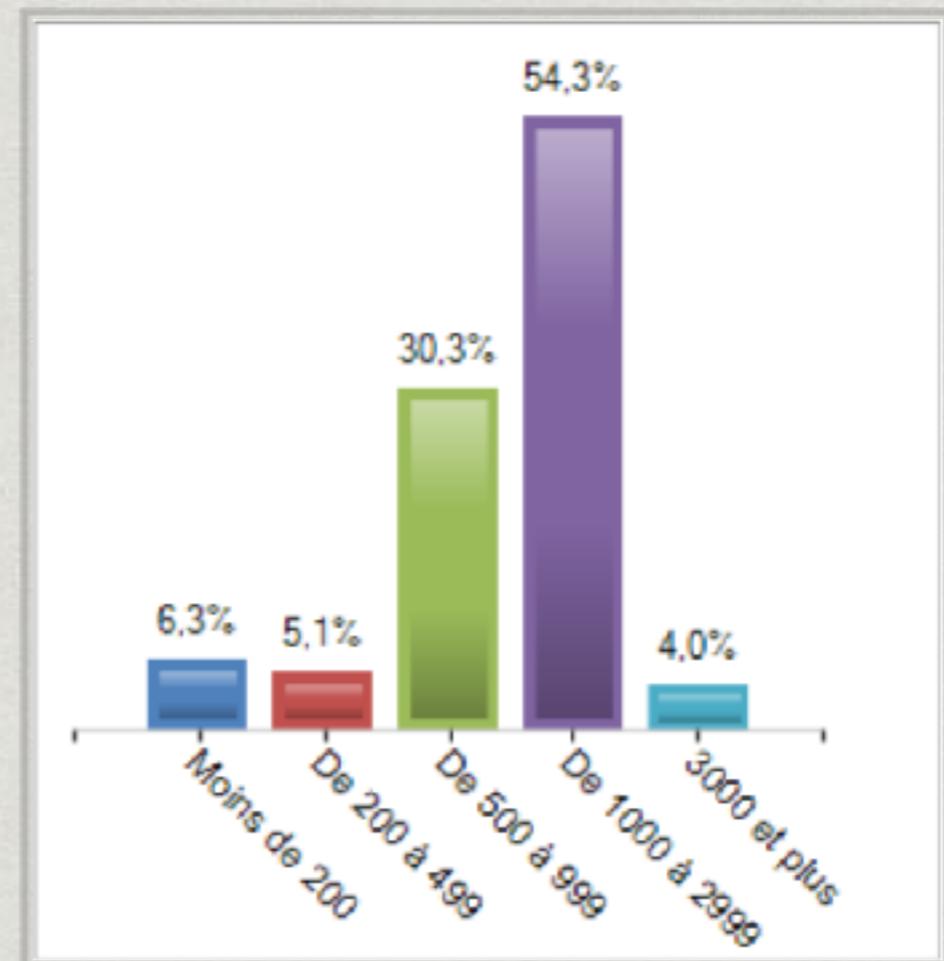
Mettre les nombres en classes personnalisées

Classes personnalisées

200;500;1000;3000

Indiquer les limites de classes en les séparant par des points-virgules, par exemple 1000;5000

OK Fermer



A garder à l'esprit

- * Pour réaliser une « bonne » mise en classe, il faut justifier à la fois le nombre de classes et les bornes des classes —> tout cela dépendra du contexte d'application
- * La mise en classe doit conserver l'information présente dans les données : quantiles - moyenne - mode et la mise en classe doit s'adapter à la distribution
- * Enfin, elle doit faire ressortir les ressemblances/différences entre les individus —> mise en avant de profils différents - catégories sociales. Cela se révèle intéressant pour des études effectuées sur une base clientèle ou des études marketing de façon plus générale
- * Ne pas créer un nombre de classes trop grand (maximum 7) afin de faciliter la lecture et l'analyse des résultats a posteriori

ANALYSE DES VARIABLES TEXTUELLES

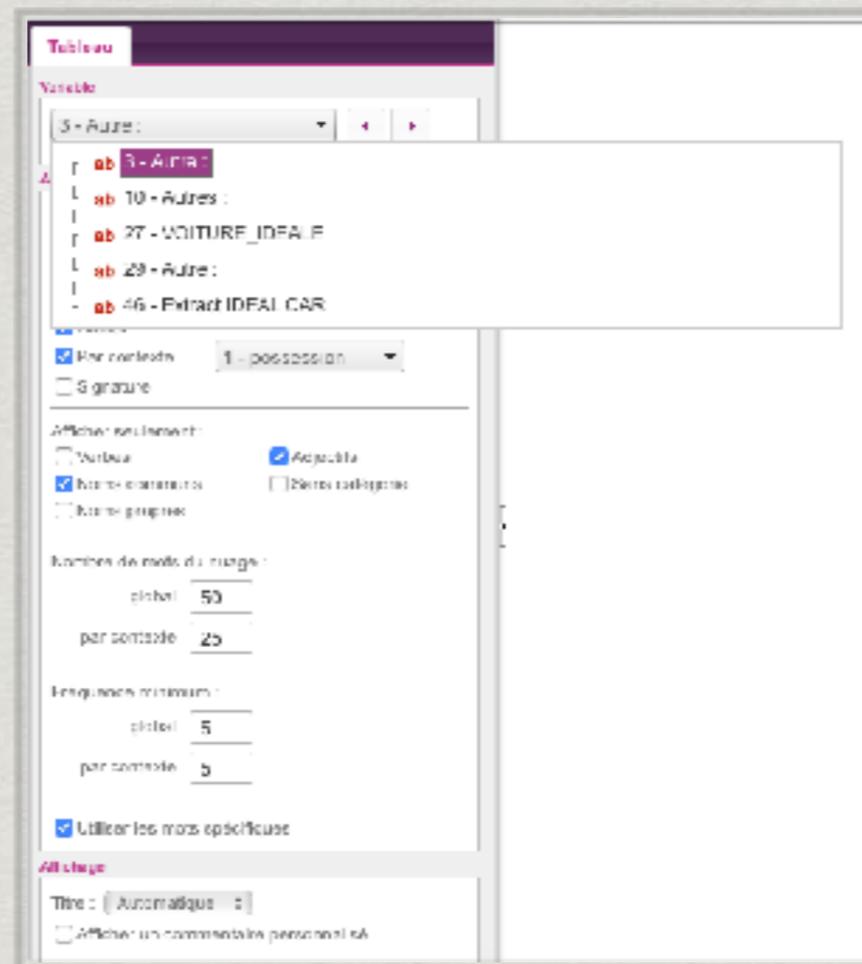
PROCÉDER À UNE ANALYSE AVEC SPHINX

Questions Textes

- * L'analyse des questions ouvertes est très différente de ce que nous avons pu voir plus tôt
 - * Pas de notion de valeur centrale
 - * Pas de notion de dispersion
- * En revanche, Sphinx nous permet malgré tout de mettre en avant les *termes - champs lexicaux* qui reviennent le plus souvent dans les réponses aux questions ouvertes
- * Il propose également des représentations intuitives qui permettent d'illustrer la représentativité de chaque mot au sein des réponses fournies par les répondants.

Questions ouvertes

- * Rendez-vous dans l'onglet « Analyse » —> « Nouvelle Analyse » —> « Analyse Textuelle » et sélectionnez la variable 27 : « Voiture Idéale »
- * Sphinx ne vous proposera l'analyse que des questions ouvertes de type *texte*, donc des questions précédées du sigle **ab** dans Sphinx.



Questions ouvertes

- * Vous trouverez aussi des informations sur la variable comme : le taux de réponse à cette question - le nombre de mots utilisés dans l'ensemble des réponses - la longueur médiane de la réponse
- * Sphinx vous propose également de synthétiser les informations relatives à chaque réponse en vous proposant les thèmes prédominants dans les réponses

27 - VOITURE_IDEALE

Aperçu lexical et sémantique

Documentation : 71 % 

Corpus : 483 mots

Longueur médiane : 3 mots

Principales Idées : Transports routiers - Rapport à la vitesse - Equipement industriel - Coloration - Intoxication - ...

Questions ouvertes

- * Une représentation graphique est également proposée pour synthétiser l'ensemble des réponses fournies.
- * Elle accorde une taille proportionnellement à la fréquence d'apparition de ce mot au sein des réponses fournies (synonymes compris).



Questions ouvertes

- * Il est possible de synthétiser les réponses fournies sous différentes formes :
- * **Nuage de mots clefs**
(représentation par défaut)
- * **Verbatim** : représentation des réponses sous forme de tableau dans l'ordre alpha-numérique
- * **Codification** : permet de classer soit même les réponses fournies —> création d'une nouvelle variable avec ses propres catégories

Variable
27 - VOITURE_IDEALE

Analyses

Nuages de mots-clés
 Verbatim
 Codification

Global
 Par contexte 1 - possession
 Signature

Afficher seulement :

Verbes Adjectifs
 Noms communs Sans catégorie
 Noms propres

Nombre de mots du nuage :

global 50
par contexte 25

Fréquence minimum :

global 5
par contexte 5

Utiliser les mots spécifiques

Questions ouvertes

- * Concentrons nous sur la représentation à l'aide d'un nuage de mots clefs - **Global**
 - * Présentation par défaut, analyse globale.
 - * **Choix du type de mot** pris en compte : ici **adjectifs** et **noms communs**
 - * **Nombre maximum de mots** dans le nuage : ici 50 mots
 - * **Fréquence minimale** pour que les mots soient représentées dans le nuage : le mot doit apparaître 5 fois dans le corpus pour être représenté

Variable

27 - VOITURE_IDEALE

Analyses

Nuages de mots-clés

Verbatim

Codification

Global

Par contexte

Signature

Afficher seulement :

Verbes

Adjectifs

Noms communs

Sans catégorie

Noms propres

Nombre de mots du nuage :

global 50

Fréquence minimum :

global 5

Affichage

Titre : Automatique

Afficher un commentaire personnalisé

Questions ouvertes

- * Passons maintenant à la représentation à l'aide d'un nuage de mots clefs - **Par contexte**
- * Affichage d'un nuage de mots selon le contexte, *i.e.* selon les modalités d'une autre variable : ici le sexe. Un nuage de mots clefs par modalité de la variable de contexte
- * Les autres options de représentation restent inchangées.

Nuages de mots-clés
 Verbatim
 Codification
 Global
 Par contexte 31 - SEXE ▼
 Signature

Afficher seulement :

Verbes Adjectifs
 Noms communs Sans catégorie
 Noms propres

Nombre de mots du nuage :

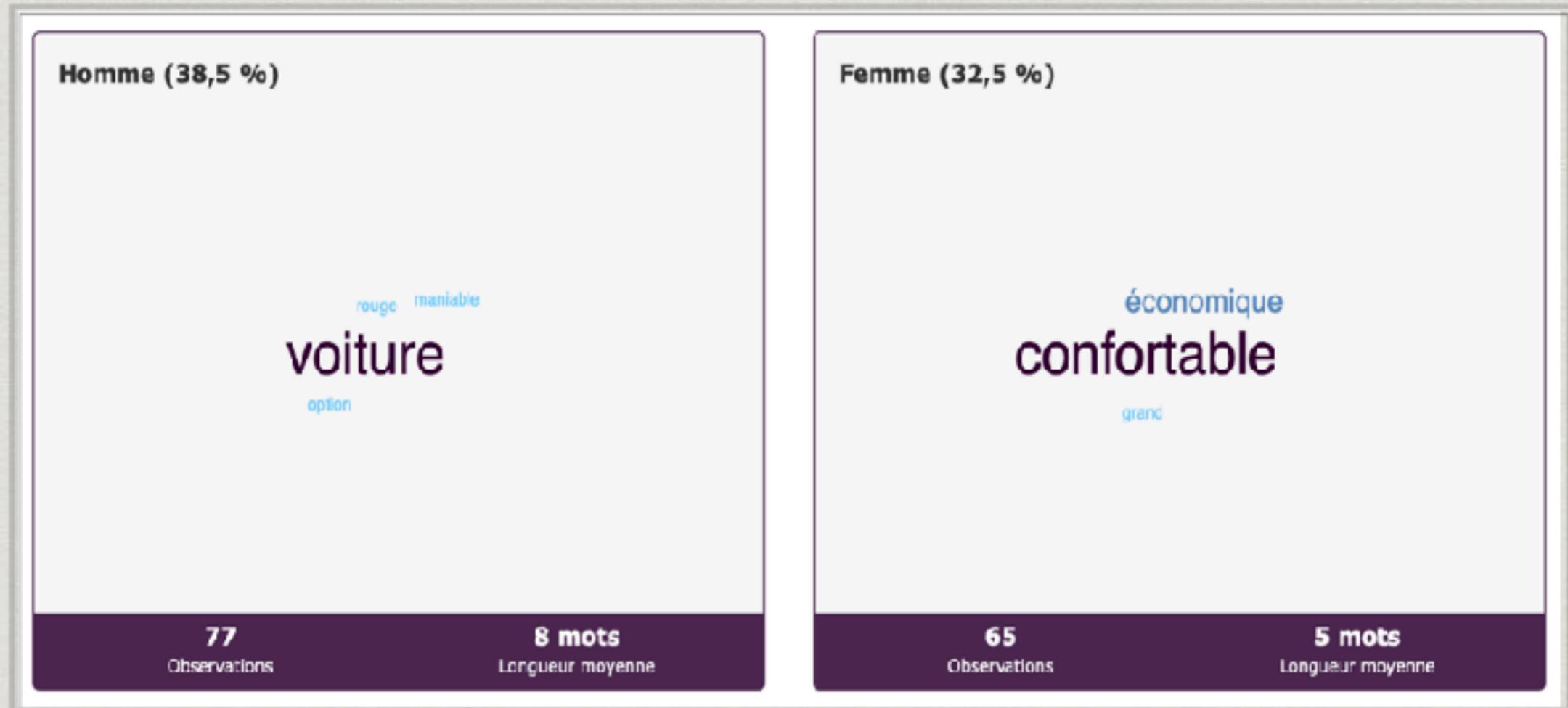
par contexte

Fréquence minimum :

par contexte

Utiliser les mots spécifiques

Questions ouvertes



- * Ici, on remarque que les femmes accordent plus d'importance au « confort » et au côté « économique » alors que facteur déterminant pour l'achat d'une voiture chez les hommes sont les options. On pourrait donc opposer des esprits plus rationnels et concentrés sur les critères essentiels aux personnes qui apprécient les « gadgets » ou qui accordent de l'importance au paraître.

Questions ouvertes

27 - VOITURE_IDEALE	
Rechercher les textes contenant <input type="text"/>	
▼ VOITURE_IDEALE	↕ SEXE
8 grandes places	Femme
aérodynamique	Homme
aérodynamique, non polluante	Femme
Amphibie mi-auto mi-bateau, avion aussi	Homme
anti-PV	Homme
avec des chromes éclatants , super-rapide	Homme
avec une jolie fille dedans	Homme
avec un tableau de bord en bois laqué et un chauffeur en tenue	Homme

- * **Verbatim** : affiche l'ensemble des réponses sous forme de tableau structuré, les réponses à la questions sont affichés dans l'ordre alpha-numérique
- * Il est cependant possible d'utiliser l'option **signature** pour afficher les réponses selon les modalités d'une autre variable : ici le sexe du répondant
- * Cette représentation n'a que peu d'intérêt car elle ne synthétise pas les résultats obtenus mais ne fait que les lister

Questions ouvertes

 3 / 5
Seat
 1 / 5
Lada
 5 / 5
Audi

- * **Verbatim** : affiche l'ensemble des réponses sous **forme de liste** dans le cas présent (variable 3 - Autre). Il est cependant possible d'utiliser l'option **évaluation** pour afficher les réponses selon les modalités d'une autre variable : **une variable numérique** : ici l'importance de la sécurité
- * On remarque ici que les personnes qui ont acheté une Audi l'ont fait car cette marque est pour eux un gage de sécurité contrairement aux propriétaires d'une Seat.

Questions ouvertes

- * **Codification** : permet de ranger les différentes réponses dans des catégories
 - * **prédéfinies** par Sphinx ou encore **par contexte** pour ajouter un filtre supplémentaire selon une variable choisie
 - * **personnalisées** : la plus intéressante car elle vous permet de créer vos propres catégories dans lesquelles vous aller pouvoir ranger vos réponses

Variable

27 - VOITURE_IDEALE   

Analyses

Nuages de mots-clés

Verbatim

Codification

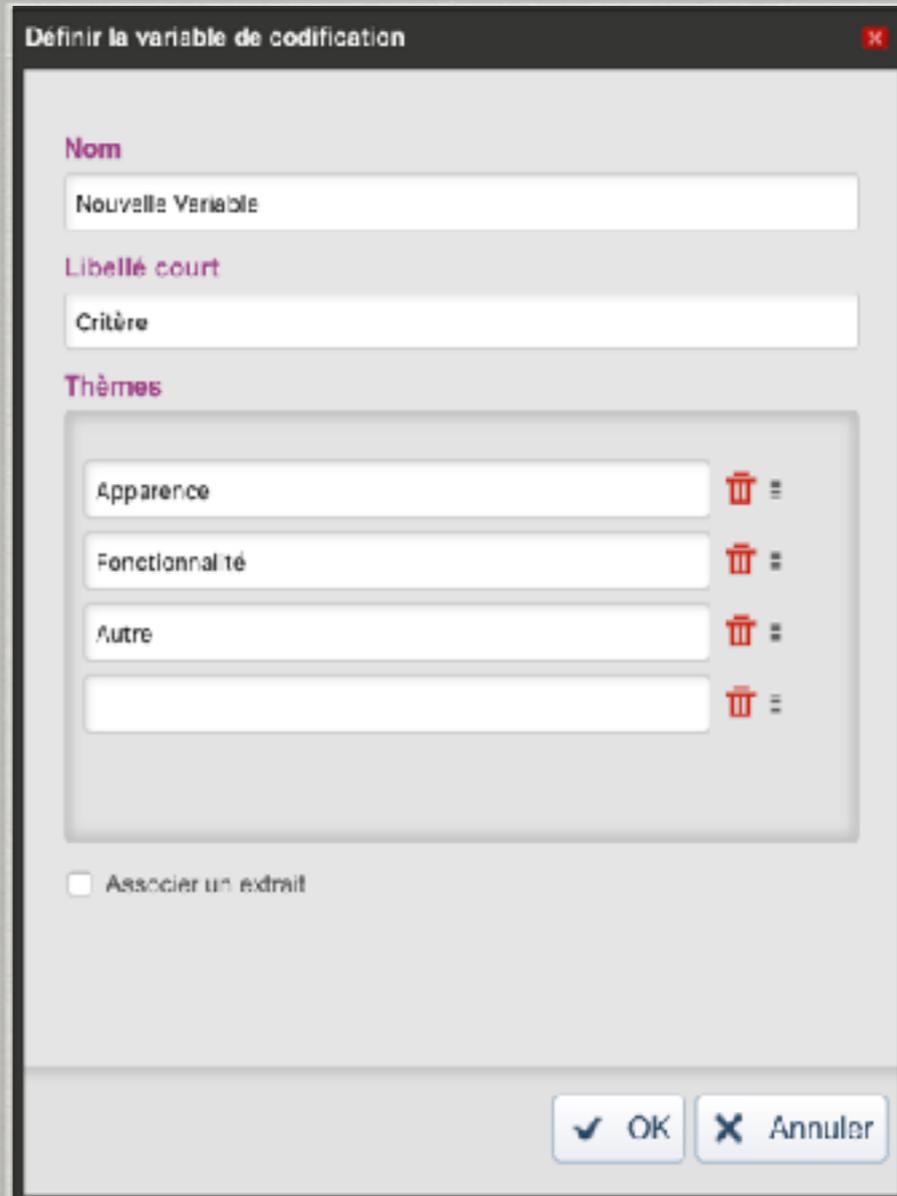
Par contexte

IDEAL CAR   

 Construire une grille thématique

Questions ouvertes

- * **Nom** : nom associé à la nouvelle variable
- * **Libellé court** : nom tel qu'il apparaîtra dans la liste des variables
- * **Thèmes** : vous définissez vos différents thèmes/catégories, *i.e.* les modalités de la variable en cours de création
- * Quand vous avez terminé, vous n'avez plus qu'à cliquer sur Ok



The screenshot shows a dialog box titled "Définir la variable de codification" with a close button (X) in the top right corner. The dialog is divided into several sections:

- Nom**: A text input field containing "Nouvelle Variable".
- Libellé court**: A text input field containing "Critère".
- Thèmes**: A list of themes with input fields and delete icons (trash can) to the right of each field:
 - Apparence
 - Fonctionnalité
 - Autre
 - An empty field
- At the bottom left, there is a checkbox labeled "Associer un extrait" which is currently unchecked.
- At the bottom right, there are two buttons: "OK" with a checkmark icon and "Annuler" with an X icon.

Questions ouvertes

Autre :

Audi

Critère

Apparence Fonctionnalité Autre

Enregistrer

- * Il nous reste plus qu'à ranger les différentes réponses dans les différents thèmes définis par vos soins.
- * Bien que cette étape soit fastidieuse à effectuer, elle vous permet de synthétiser les résultats sous la forme qui vous semble la plus appropriée pour votre étude
- * Une fois que cela est fait, la variable nouvellement créée apparaîtra dans la liste des variables et vous pourrez en visualiser le tableau/graphique associé à cette nouvelle variable (catégorique).

Questions ouvertes

Tableau

Variable

3 - Autre : [dropdown] [left arrow] [right arrow]

Analyses

- Nuages de mots-clés
- Verbatim
- Codification
- Par contexte

Nouvelle Variable [pencil icon] [bar chart icon] [close icon]

+ Construire une grille thématique [Voir les résultats tooltip]

Tri à plat

Critère

Calculs effectués sur la strate : Autre : n'est pas non-réponse

	Effectifs	% Obs.
Non-réponse	16	80%
Apparence	2	10%
Fonctionnalité	2	10%
Autre	0	0%
Total	20	

Réponses effectives : 4
Taux de réponse : 20%

Non-réponse(s) : 16
Modalité la plus citée : Apparence

[Fermer]

Passage à une variable nominale voire ordinaire si les questions ouvertes expriment un avis

Pour résumer

- * Votre mise en classe doit vous permettre en avant des ressemblances/différences au sein des individus interrogés pendant l'enquête.
Penser à limiter le nombre de classes afin de rendre la présentation de vos résultats lisible.
- * Pour l'analyse des textes, penser à utiliser l'analyse par contexte pour mettre en avant d'éventuelles différences entre différentes catégories de répondants.
- * De façon générale, **les manipulations effectuées sur Sphinx doivent vous servir à extraire de l'information.** Elles vous permettront d'extraire des informations sur le profil des répondants.

A suivre

- * Vous allez maintenant mettre en pratique les différentes règles et manipulations afin de procéder à l'analyse de questions ouvertes (numériques et textes)
- * Pour cela, vous utiliserez les enquêtes *Young* et *Tourisme*
- * Le travail s'effectuera **seul** et fera l'objet **d'une présentation orale** devant l'ensemble des étudiants pour certains d'entre vous