

Analyse de Contenu et de Données

Premières Analyses

Guillaume Metzler
guillaume.metzler@live.fr

Univ. Lyon, UJM-Saint-Etienne, CNRS, Institut d'Optique Graduate School, Laboratoire
Hubert Curien UMR 5516, F-42023, SAINT-ETIENNE, France

Printemps 2020



Dans l'épisode précédent

Résumé

Vous avez appris

- des définitions fondamentales de statistiques
- à distinguer la variable de la population étudiée
- à traiter les données manquantes
- à traiter les "outliers"
- à distinguer les différents types de variable

Dans l'épisode précédent

Faire la distinction entre les différents types de variables

Nature des Variables

Deux types de variables

1) Quantitatives :

- discrètes : nombre de "pile" obtenu lors d'un lancer de pile ou face
- continues : le salaire net d'un employé

2) Qualitatives :

- nominales : le lieu de résidence
- ordinales : la note attribuée à un produit
- ouvertes : impression sur vos dernières vacances

Nature des Variables - Sphinx

Exemple avec l'enquête "Télévision"

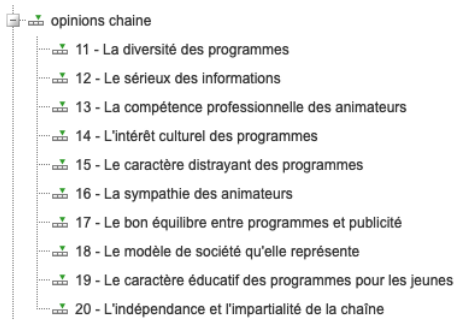
1	-	Fréquence
2	-	Circonstances
3	-	Motivations
4	-	Choix
5	-	Zapping
6	-	Chaînes préférées
7	-	Chaîne la plus regardée
8	-	Emission préférée
9	-	Adjectifs
10	-	Note globale
+	-	opinions chaîne
+	-	opinion tv
31	-	Sexe
32	-	Age
33	-	Situation Profession
34	-	Diplôme
35	-	Nb personnes
37	-	Date de saisie
45	-	Temps de saisie
39	-	Langue

Signification des symboles

- 1 - Fréquence : variable catégorique à réponse unique
- 2 - Circonstances : variable catégorique à réponses multiples
- 4 - Choix : variable ordinale
- 8 - Emission préférée : variable textuelle, i.e. réponse à une question ouverte
- 37 - Date de saisie : variable temporelle de type "date" ou "heure"
- 38 - Temps de saisie : variable numérique
- 39 - Langue : variable encodée

Nature des Variables - Sphinx

Une variable peut en cacher une autre



La présence d'un symbole "+" doit vous inciter à cliquer dessus pour voir quelles sont les variables qui s'y trouvent !

"Opinions chaîne" cache en fait un ensemble de 10 variables ordinales.

Sommaire

- Règles de présentations des données en fonction de la nature de variables
- Construction de tableaux à l'aide de Sphinx Campus
- Construction de graphiques - l'aide Sphinx Campus
- Rédiger une analyse de données

Présentation des données

Plusieurs possibilités

- 1) un tableau
- 2) un tableau ordonné (quand cela a un sens)
- 3) mais aussi des graphiques
 - un diagramme (ordonné ou non)
 - un graphique par secteur
 - une courbe
 - ...

Attention !

Certaines présentations ne sont pas adaptées à certains types de variables !

Variable qualitative

Variable catégorique à réponse unique

La version tableau classique : modalités - effectifs - fréquences

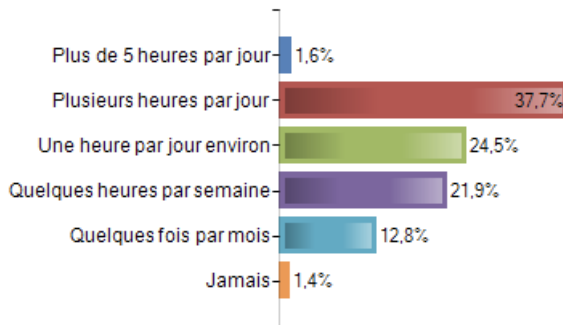
	Effectifs	% Obs.
Plus de 5 heures par jour	8	1,6%
Plusieurs heures par jour	183	37,7%
Une heure par jour environ	119	24,5%
Quelques heures par semaine	106	21,9%
Quelques fois par mois	62	12,8%
Jamais	7	1,4%
Total	485	100%

Nous aurions pu également ordonner les lignes du tableau par effectif *croissant* ou *décroissant* ou encore ...

Variable qualitative

Variable catégorique à réponse unique

... utiliser un graphique qui permet de mieux rendre compte de la modalité *dominante*.

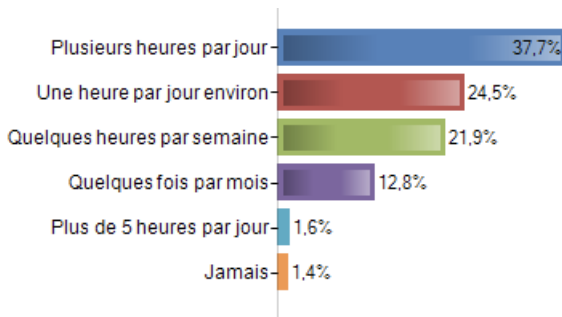


Mais on peut accentuer cela !

Variable qualitative

Variable catégorique à réponse unique

Il suffit d'ordonner les modalités par ordre *croissant* ou *décroissant*.

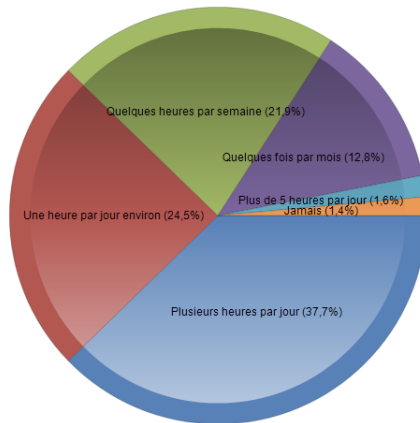


Seul inconvénient des diagrammes : on perd l'information relative à l'effectif total.

Variable qualitative

Variable catégorique à réponse unique

Pour finir, une autre représentation possible.



Variable qualitative

Variable catégorique à réponses multiples

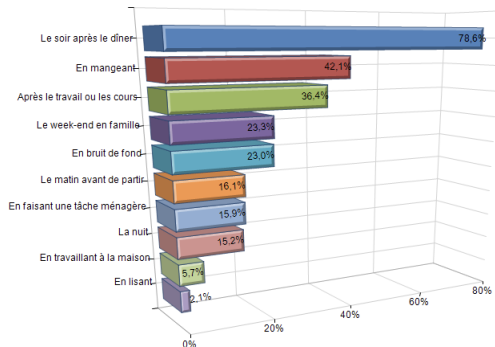
On peut les représenter de la même façon que précédemment, à l'aide d'un tableau ...

	Effectifs	% Obs.
Non-réponse	5	1%
Le matin avant de partir	76	15,9%
Le soir après le dîner	372	77,8%
La nuit	72	15,1%
Après le travail ou les cours	172	36%
En faisant une tâche ménagère	75	15,7%
En lisant	10	2,1%
En travaillant à la maison	27	5,6%
En mangeant	199	41,6%
Le week-end en famille	110	23%
En bruit de fond	109	22,8%
Total	478	

Variable qualitative

Variable catégorique à réponses multiples

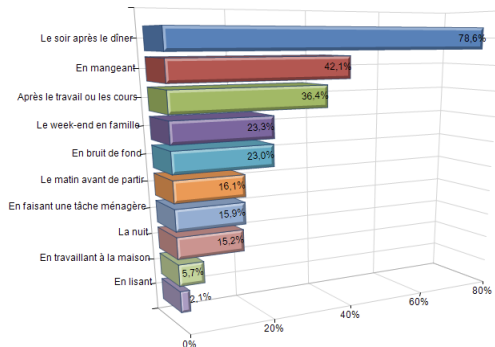
ou encore d'un diagramme à barres ordonnées comme présenté plutôt pour mettre en avant les modalités dominantes.



Variable qualitative

Variable catégorique à réponses multiples

ou encore d'un diagramme à barres ordonnées comme présenté plutôt pour mettre en avant les modalités dominantes.

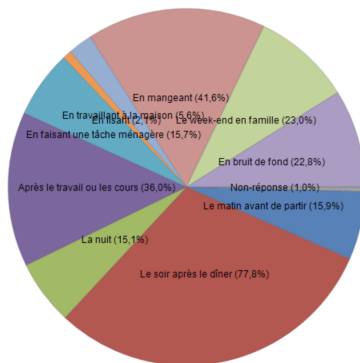


Mais surtout ...

Variable qualitative

Variable catégorique à réponses multiples

Ne présenter JAMAIS vos résultats de la façon suivante !!!



La représentation est contre-intuitive par rapport à la nature de la variable.

Variable qualitative

Variable ordinale

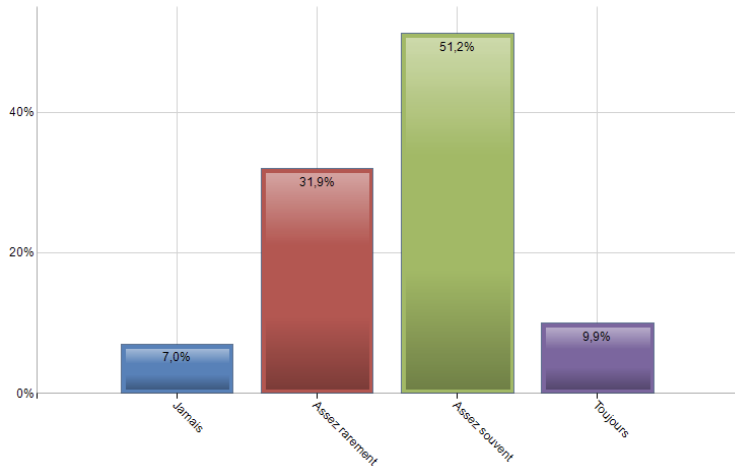
Diagrammes à barres ou de simples tableaux sont tout à fait adaptés à ce type de variables.

	Effectifs	% Rep.
Assez souvent	242	51,2%
Assez rarement	151	31,9%
Toujours	47	9,9%
Jamais	33	7%
Total	473	100%

Par contre : ne pas modifier l'ordre des modalités ! Cela n'aurait pas de sens car elles ont un lien logique entre elles.

Variable qualitative

Variable ordinale



Variable quantitative

Variable numérique

Etude

- un pré-traitement des données est nécessaire : mise en classes, i.e. création d'intervalles de valeurs.
Ex : salaire - montant des dépenses - temps passé à remplir un questionnaire - ...
- représenter la distribution par sa densité (pas faisable sous sphinx)

Dans le cas d'une mise en classes : on traitera la variable comme une variable *ordinaire*.

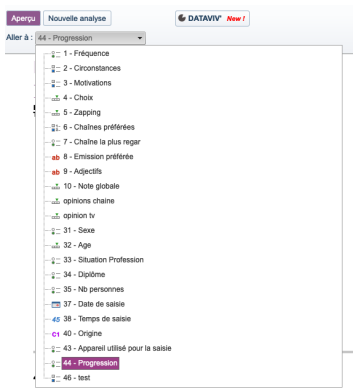
Générer des graphiques à l'aide de Sphinx Campus

Connectez vous sur Sphinx et sélectionnez l'enquête "Télévision".



Nous allons cette fois-ci nous rendre dans l'onglet "Analyses" (tout à droite)

Cliquez ensuite sur "Revenir à l'environnement standard d'analyse".



Vous verrez un tableau et un graphique par variable.

Le menu déroulant vous permet d'accéder directement aux graphiques générés par Sphinx pour le variable sélectionnée (ici "Progression")

Sur cette même page vous pouvez également :

Voir le nombre de total de réponse.

Filtrer les résultats selon les modalités d'une variable comme vu au cours précédent.

Ici on aura filtrer les résultats selon la variable "Sexe" en ne conservant que les résultats pour les répondants du sexe féminin.

Choisir une strate ▼ Partager ▼ ? ↗

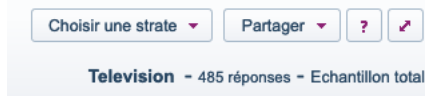
Television - 485 réponses - Echantillon total

Calculs effectués sur la strate : Sexe parmi "Feminin"

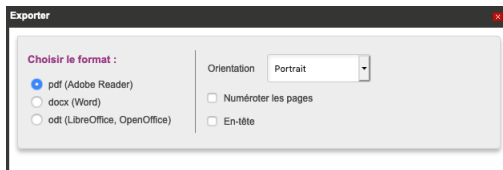
1. Fréquence

	Effectifs	% Obs.
Plus de 5 heures par jour	4	1,7%
Plusieurs heures par jour	105	44,5%
Une heure par jour environ	48	20,3%
Quelques heures par semaine	45	19,1%
Quelques fois par mois	31	13,1%
Jamais	3	1,3%
Total	236	100%

Encore sur cette même page en cliquant sur "Partager" :



Vous pourrez diffuser vos tableaux et graphiques mais également les exporter dans différents formats (pdf ou word).

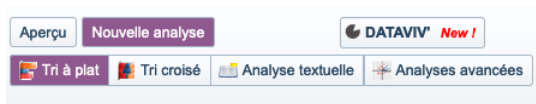


Je vous conseille en "pdf", la qualité des images est meilleure mais c'est moins flexible que le format "word" selon votre éditeur de texte.

Cliquez maintenant sur "Nouvelle analyse" en haut à gauche de la page :



Une nouvelle page va s'ouvrir avec ce nouveau menu au même endroit que le précédent



Par défaut, vous voyez que Sphinx effectue un "Tri à plat", i.e. il étudie une variable seule, indépendamment des autres. Nous verrons les autres options plus tard.

Concentrons nous sur le menu qui se trouve sur la gauche, il comprend deux onglets :

Un menu avec des options vous permettant de personnaliser vos tableaux (avec impacts sur les graphes).

Un menu pour personnaliser les graphiques (style, couleur, forme).

The screenshot shows the 'Tableau' tab selected. It contains several sections for table customization:

- Variable:** A dropdown menu set to '1 - Fréquence' with left and right arrow buttons.
- ☐ Regrouper les modalités
- Options:** A section with a collapsed dropdown arrow.
- Analyses:** A section with checkboxes for 'Effectifs' (checked), 'Pourcentages' (checked), and 'Tests statistiques' (unchecked).
- Options de calcul:** A section with ☐ Ignorer les non-réponses.
- Affichage:** A section with a 'Titre' dropdown set to 'Automatique', and checkboxes for 'Afficher le tableau' (checked), 'Afficher le graphique' (checked), and 'Afficher un commentaire personnalisé' (unchecked).

The screenshot shows the 'Graphique' tab selected. It contains options for graph customization:

- Type de graphique:** Radio buttons for 'Automatique' (selected) and 'Personnalisé'.
- Couleurs:** Radio buttons for 'Palette prédéfinie' (selected), 'Une seule couleur' (unchecked), and 'Deux couleurs' (unchecked). A dropdown menu next to 'Palette prédéfinie' is set to 'Office 2007'.

Sélectionnez la variable 2 "Circonstances".
Vous verrez alors le tableau suivant :

	Effectifs	% Obs.
Non-réponse	5	1%
Le matin avant de partir	76	15,9%
Le soir après le dîner	372	77,8%
La nuit	72	15,1%
Après le travail ou les cours	172	36%
En faisant une tâche ménagère	75	15,7%
En lisant	10	2,1%
En travaillant à la maison	27	5,6%
En mangeant	199	41,6%
Le week-end en famille	110	23%
En bruit de fond	109	22,8%
Total	478	

Retournons sur le menu qui se trouve sur la gauche, cliquez sur "regroupez les modalités" et sur la petite "étoile violette", la page de droite va alors apparaître.

The screenshot shows the 'Analyses' menu on the left side of the Tableau interface. It is divided into several sections:

- Variable:** A dropdown menu showing '2 - Circonstances'. Below it are two checked options: 'Regrouper les modalités' (with a gear icon) and 'Afficher les sous-ensembles'. There is also an 'Options' link.
- Analyses:** A section with three options: 'Effectifs' (checked), 'Pourcentages' (checked), and 'Tests statistiques' (unchecked).
- Options de calcul:** A section with one option: 'Ignorer les non-réponses' (unchecked).
- Affichage:** A section with a 'Titre' dropdown set to 'Automatique' and three options: 'Afficher le tableau' (checked), 'Afficher le graphique' (checked), and 'Afficher un commentaire personnalisé' (unchecked).

The screenshot shows the 'Regroupements : "Circonstances"' dialog box. It has two tabs: 'Regrouper' (selected) and 'Ordonner'. The main area is titled 'Définition des regroupements par modalité' and contains a table with two columns: 'Modalités' and 'Regroupement'.

Modalités	Regroupement
Le matin avant de partir	Journée
Le soir après le dîner	Soir
La nuit	Soir
Après le travail ou les cours	Soir
En faisant une tâche ménagère	Accessoire
En lisant	Accessoire
En travaillant à la maison	Accessoire
En mandeant	Accessoire

At the bottom right of the dialog box are two buttons: 'OK' (with a checkmark icon) and 'Fermer' (with an 'X' icon).

Après avoir validé vos modifications, un bandeau jaune va apparaître sur votre page. Cliquez-dessus pour valider vos modifications :

	Effectifs	% Obs.
Non-réponse	5	1%
Journée	76	15,9%
Soir	616	128,9%
<i>Le soir après le dîner</i>	372	77,8%
<i>La nuit</i>	72	15,1%
<i>Après le travail ou les cours</i>	172	36%
Accessoire	420	87,9%
<i>En faisant une tâche ménagère</i>	75	15,7%
<i>En lisant</i>	10	2,1%
<i>En travaillant à la maison</i>	27	5,6%
<i>En mangeant</i>	199	41,6%
<i>En bruit de fond</i>	109	22,8%
Week-end	110	23%
Total	478	

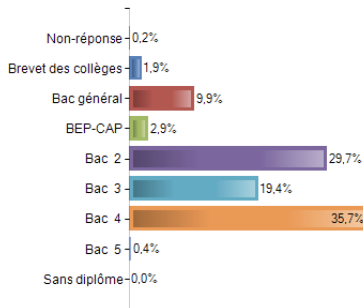
On va faire abstraction des résultats dans un premier temps ... pourquoi le regroupement n'est pas approprié à votre avis ?

Attention

- Ne pas représenter les résultats d'une variable qualitative à réponses multiples par un graphique par secteur
 - Ne pas chercher à regrouper les modalités (si vous avez une fréquence supérieure à 100% ... c'est que vous avez fait de la ...)
-
- Sphinx retourne une nouvelle table avec les nouvelles modalités issues du regroupement
 - Les anciennes modalités (sous-ensembles) apparaissent grisées
 - Il est possible de supprimer les sous ensembles de la visualisation en décochant sur "Afficher les sous-ensembles" dans le menu de gauche.

Sélectionnez la variable 34 "Diplôme". Vous verrez alors le tableau et graphique suivants :

	Effectifs	% Obs.
Non-réponse	1	0,2%
Brevet des collèges	9	1,9%
Bac général	48	9,9%
BEP-CAP	14	2,9%
Bac 2	144	29,7%
Bac 3	94	19,4%
Bac 4	173	35,7%
Bac 5	2	0,4%
Sans diplôme	0	0%
Total	485	100%



Retournons sur le menu qui se trouve sur la gauche, cliquez sur "Options" puis trier les lignes. Vous pourrez trier les modalités par ordre "croissant" , "décroissant" ou par "ordre alphabétique".

Tableau **Graphique**

Variable

34 - Diplôme

☐ Regrouper les modalités

Options

☒ Afficher les totaux

☒ Trier les lignes **Ordre décroissant**

Options de calcul

☐ Ignorer les non-réponses

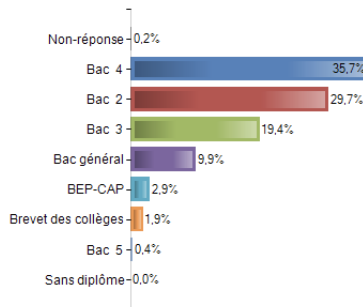
Affichage

Titre : Automatique

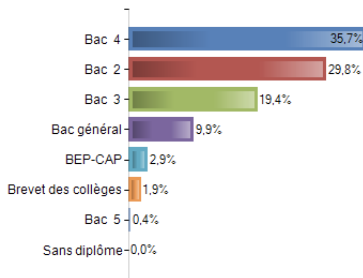
☒ Afficher le tableau

☒ Afficher le graphique

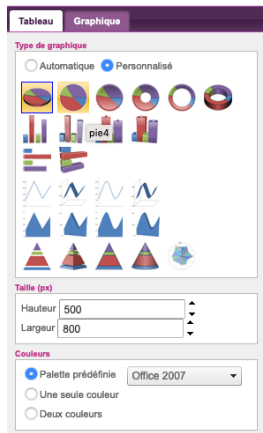
☐ Afficher un commentaire personnalisé



- Cette manipulation a un impact aussi bien sur le graphique que sur le tableau généré.
- Par défaut la modalité "non-réponse" se trouve toujours en première ligne.
- Il est cependant possible de les supprimer en cliquant sur "Ignorer les non-réponses" dans le menu de gauche, ce qui va modifier vos pourcentages.



Regardons maintenant un peu ce que nous cache l'onglet "Graphique" puis cliquez sur "Personnalisé"



Rappelez-vous, les regroupements ou filtres effectués lorsque vous étiez dans l'onglet tableau sont pris en compte dans cette partie.

Vous pourrez alors :

- Modifier vos le style de vos graphiques : par secteur - bâton (horizontaux, verticaux en 2D ou 3D) - courbe - pyramidale (ou hiérarchique) - radar (ou étoile).
- Modifier les couleurs et taille de vos graphiques

Analyser les données

- Rédiger une analyse des données **claire** et **concise** est primordiale lorsque vous devez faire un compte-rendu.
- Il est important d'être synthétique et d'extraire les informations importantes pour répondre à la problématique posée.
- **Il ne s'agit en aucun cas de présenter toutes les données en long, en large et en travers !**
- Outre la synthèse des données, vos commentaires et interprétations sont aussi importants pour tenter d'expliquer ou émettre des hypothèses quant à certaines observations effectuées.

La rédaction de votre analyse de données comprendra essentiellement trois parties :

- *une introduction* : de façon générale, elle va présenter le contexte de l'étude. Elle permettra au lecteur de bien appréhender le sujet afin de bien comprendre vos résultats.
- *un développement* : c'est ici que vous mettrez vos tableaux/graphiques ainsi que l'analyse a proprement parlé de ces différents éléments.
- *une conclusion* : elle doit servir à **conclure** en répondant à la problématique posée. Une ouverture est toujours appréciable.

Les différents éléments

Votre introduction doit présenter :

- le contexte de l'étude
- la problématique de l'étude, i.e. qu'est-ce que l'on cherche à étudier
- une présentation de vos données : nombre d'individus et éventuellement le profil des répondants quand cela est pertinent pour l'étude
- les variables étudiées pour répondre à la problématique
- les éventuels outils statistiques utilisés

Les différents éléments

Votre développement doit comprendre :

- un paragraphe dans lequel on aura extrait les informations pertinentes issues d'un tableau ou d'un graphique que l'on aura cité !
- le(s) tableau(x) et/ou graphique(s) utilisées.
 - 1) **Eviter les représentations redondantes**
 - 2) **Mettre un titre + légende à chaque graphique/tableau.**
Gardez à l'esprit qu'un graphe ou table doit se suffire à lui/elle même, donc être indépendante du texte
- ce même paragraphe comprendre aussi vos conclusions suites aux informations extraites mais aussi d'éventuels commentaires.

On rédige en général un paragraphe par variable ou couples de variables étudié(es).

Les différents éléments

Dans votre conclusion :

- Reprenez les éléments de vos différentes conclusions partielles.
- Faites une synthèse globale des informations extraites des différentes variables.
- Répondez à la problématique posée