

Analyse de Contenu et de Données

Valeurs Centrales et Mesures de Dispersion

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



Résumé du cours précédent

Lors de la séance précédente vous avez :

- appris à distinguer les variables quantitatives et qualitatives sur Sphinx
- dresser un tableau et un graphique pour les différents types de variables

Sommaire

- Rappels de quelques notions statistiques fondamentales
- Acquisition et manipulations des grandeurs statistiques sur Sphinx

Rappels en Statistiques :

Moyenne - Médiane - Mode - Variance

Moyenne

Soit x_1, x_2, \dots, x_n un échantillon de n exemples. La **moyenne empirique** \bar{x} de cet échantillon est définie par :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Dans le cas où l'on cherche à étudier les moyennes d'une variable aléatoire X qui admet une densité f , la moyenne (aussi appelée **espérance**), notée $\mathbb{E}[X]$, est définie par :

$$\mathbb{E}[X] = \int x f(x) dx$$

Valeurs centrales

A propos

La moyenne est la caractéristique la plus souvent utilisée pour caractériser la tendance d'un groupe de population. On parle par exemple du prix moyen d'un produit, de la valeur de moyenne des actions BNP Paribas, du temps moyen pour effectuer un trajet.

Exemple

On considère la distribution des âges suivante:

Âge	23	22	15	20	20
-----	----	----	----	----	----

Valeurs centrales

A propos

La moyenne est la caractéristique la plus souvent utilisée pour caractériser la tendance d'un groupe de population. On parle par exemple du prix moyen d'un produit, de la valeur de moyenne des actions BNP Paribas, du temps moyen pour effectuer un trajet.

Exemple

On considère la distribution des âges suivante:

Âge	23	22	15	20	20
-----	----	----	----	----	----

La moyenne est égale à : $\frac{1}{5} (23 + 22 + 15 + 13 + 20) = 20$.

Valeurs centrales

A propos

La moyenne présente cependant un inconvénient majeur, elle est sensible aux valeurs aberrantes ou outliers, i.e. aux valeurs qui sont très différentes des autres valeurs de l'échantillon (beaucoup plus grandes ou plus petites).

Exemple

On cherche à estimer le patrimoine financier moyen des français. On tire alors 100 individus au hasard parmi lesquels se trouve celui de (complètement au hasard) celui de *Bernard Arnault*, PDG de LVMH. Si on considère que le patrimoine financier d'un français lambda est d'environ 200k euros et que celui du PDG de LVMH est de 73.2Md d'euros. Nous aurions alors une moyenne française de égale à environ 730M d'euros.

Valeurs centrales

A propos

Cet exemple est un peu extrême mais ils montrent les limites de cette mesure de la valeur centrale.

Heureusement il existe une autre grandeur statistique permettant de palier à ce genre de problème et biais du aux valeurs extrêmes : **la médiane**.

Remarque : c'est une grandeur qui est plus couramment utilisée lorsque l'on cherche à étudier le salaire des français !

Médiane

Soit x_1, x_2, \dots, x_n un échantillon de n exemples. On note $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ ce même échantillon dont les valeurs sont ordonnées par ordre croissant. La valeur **médiane**, notée m , de cet échantillon est définie par :

$$x_{(1)} \leq \dots \leq x_{(n/2)} \leq m \leq x_{(n/2)+1} \leq \dots \leq x_{(n)}.$$

Autrement dit, c'est un nombre réel tel que 50% des valeurs de l'échantillon sont inférieures à cette valeur et 50% sont supérieurs à cette valeur.

Dans le cas d'une variable X admettant une densité f , la médiane est la valeur m pour laquelle on a :

$$\int_{-\infty}^m f(x)dx = 0.5.$$

Valeurs centrales

A propos

Reprenons notre exemple précédent et regardons quelle est la valeur de cette médiane.

Exemple

Si on considère que le patrimoine financier d'un français lambda est d'environ $200k$ euros et que celui du PDG de LVMH est de $73.2Md$ d'euros. Comme plus de la moitié de l'échantillon a un patrimoine d'environ $200k$ euros, la valeur médiane est alors de $200k$ euros.

Exemple

Pour une distribution gaussienne, la moyenne est égale à la médiane.

On s'intéresse à l'âge médian des étudiants en filière professionnelle. Les données recueillies sur un échantillon de 250 personnes sont les suivantes :

Age	16	17	18	19	20
Effectif	80	20	60	56	34
Effectif cumulé	80	100	160	216	250

Pour étudier la médiane, il est d'usage de calculer les effectifs cumulés croissants (ou fréquences cumulées croissantes).

Valeurs centrales

Médiane

On s'intéresse à l'âge médian des étudiants en filière professionnelle. Les données recueillies sur un échantillon de 250 personnes sont les suivantes :

Age	16	17	18	19	20
Effectif	80	20	60	56	34
Effectif cumulé	80	100	160	216	250

Pour étudier la médiane, il est d'usage de calculer les effectifs cumulés croissants (ou fréquences cumulées croissantes).

La **valeur médiane** est ici égale à 18.

On s'intéresse à l'âge médian des étudiants en filière professionnelle. Les données recueillies sur un échantillon de 250 personnes sont les suivantes :

Age	16	17	18	19	20
Effectif	80	20	60	56	34
Effectif cumulé	80	100	160	216	250

Pour étudier la médiane, il est d'usage de calculer les effectifs cumulés croissants (ou fréquences cumulées croissantes).

La **valeur médiane** est ici égale à 18.

Quid si l'on a une variable qualitative ?

Valeurs centrales

Le mode

Les notions de **moyennes et médianes** n'ont pas de sens lorsqu'on manipule des variables **qualitatives**. En revanche il d'usage de regarder quelle est la modalité (ou la catégorie) la plus citée dans ce cas là.

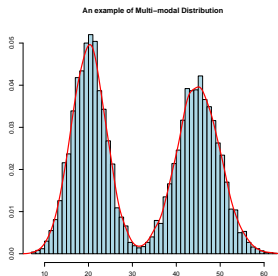
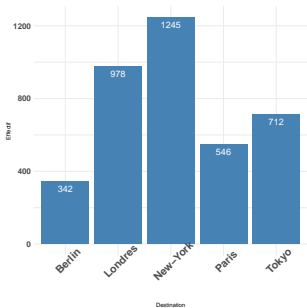
Définition

Le **mode**, en statistiques, est la valeur la plus représentée dans une population lorsque la variable X prend un nombre fini de valeurs. Dans le cas où X est une variable aléatoire à valeurs dans \mathbb{R} , il s'agit de la valeur de x où la densité de la densité de la variable aléatoire X est maximale.

Remarque : Le mode existe aussi pour les variables quantitatives ! Il s'agit de la valeur la plus fréquente (ou probable).

Valeurs centrales

Le mode : exemples



- A gauche, le mode est "*New-York*".
- A droite les modes sont 20 et 45. La distribution est dite **multi-modale**.

Une fois les valeurs centrales calculées, il est également intéressant de regarder comment sont *dispersées* les valeurs, soit de façon générale, soit autour de la valeur centrale. Pour cela on introduit deux notions :

- l'écart-type (ou la variance)
- l'étendue ou la plage des valeurs (valeurs min et max)

Lorsque l'on étudie des variables quantitatives (discrètes ou continues), une bonne mesure de dispersion des valeurs consiste à étudier l'écart-type σ .

Définition

Soit X une variable aléatoire à valeurs réelles admettant un moment d'ordre 2 (i.e. la variance existe), la **la variance** V de la variable aléatoire X est définie par :

$$V[X] = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

L'écart-type σ n'est alors rien d'autre que la racine carrée de la variance, i.e. :

$$\sigma = \sqrt{V}.$$

Application à un échantillon

Soit (x_1, x_2, \dots, x_n) un échantillon i.i.d. issue d'une certaine loi de probabilités. On suppose que pour tout $i, x_i \in \mathbb{R}$ et on note \bar{x} la moyenne de cet échantillon. Alors la variance V est déterminée par :

$$V = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

A nouveau, l'écart-type σ n'est rien d'autres que la racine carrée de la variance :

$$\sigma = \sqrt{V} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

Remarque

La variance peut se définir littéralement comme **la moyenne de la somme des carrés des écarts à la moyenne**.

Certains livres et logiciels statistiques déterminent la variance d'un échantillon par la formule suivante :

$$V = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Ne soyez donc pas surpris de voir les deux. On appelle cela un **estimateur sans biais de la variance**.

Exemple

On considère la distribution des âges suivante:

Âge	23	22	15	20	20
-----	----	----	----	----	----

La moyenne est égale à 20 et la variance est égale :

Exemple

On considère la distribution des âges suivante:

Âge	23	22	15	20	20
-----	----	----	----	----	----

La moyenne est égale à 20 et la variance est égale :

$$\begin{aligned} V &= \frac{1}{5} \left((23 - 20)^2 + (22 - 20)^2 + (15 - 20)^2 + 2 \times (20 - 20)^2 \right), \\ &= \frac{1}{5} (9 + 4 + 25 + 2 \times 0), \\ &= 7.6. \end{aligned}$$

Exemple

On considère la distribution suivante des notes dont on souhaite déterminer la variance.

Notes	0	1	2	3	4	5
Fréquences	0.4	0.1	0.2	0.2	0.1	0.1

Exemple

On considère la distribution suivante des notes dont on souhaite déterminer la variance.

Notes	0	1	2	3	4	5
Fréquences	0.4	0.1	0.2	0.2	0.1	0.1

La moyenne est égale à $0.1 + 0.4 + 0.6 + 0.4 + 0.5 = 2$. La variance est donc égale à

$$\begin{aligned} V &= 0.4 \times 2^2 + 0.1 \times 1^2 + 0.2 \times 1^2 + 0.1 \times 2^2 + 0.1 \times 3^2, \\ &= 0.4 \times 4 + 0.1 + 0.2 + 0.1 \times 4 + 0.1 \times 9, \\ &= 1.6 + 0.1 + 0.2 + 0.4 + 0.9, \\ &= 3.2 \end{aligned}$$

On a donc un écart-type σ de 1.8 environ.

Mesures de dispersions

Etendue

Lorsque l'on étudie des variables qualitatives ordinales pour lesquelles la notion de "variance" n'a pas de sens, un critère de dispersion des valeurs souvent employé est l'**étendue**.

Définition

L'**étendue** est une grandeur statistique correspondant à la différence entre la *valeur maximale* et la *valeur minimale* d'un échantillon.

Exemple

L'étendue des valeurs de température sur Terre est de $58 - (-89)$ soit 147 degrés. (Différence entre la température maximale enregistrée en Libye et la température minimale enregistrée en Antarctique).

La façon la plus précise d'avoir une idée de la dispersion des valeurs reste encore de regarder les **quantiles** de notre échantillon.

De façon informelle, un quantile d'ordre $\alpha \in [0, 1]$, notée q_α correspond à la valeur x_i de notre échantillon pour laquelle un ratio α de la population à une valeur inférieure à q_α et un ratio $1 - \alpha$ de la population à une valeur supérieure à q_α .

On retrouve la définition de la médiane dans le cas où $\alpha = 0.5$.

Remarques

Lorsque que l'on a des intervalles de valeurs

Lorsque les valeurs d'une variable quantitative sont regroupées dans des classes, i.e. **des intervalles de valeurs de la forme** $]a_i, b_i]$, la moyenne \bar{x} se calcule de la façon suivante :

$$\bar{x} = \frac{1}{m} \sum_{i=1}^m \alpha_i \times \frac{a_i + b_i}{2},$$

où α_i désigne l'effectif de la classe $]a_i, b_i]$ et m l'effectif total.

Autrement dit pour calculer la moyenne on considère simplement le centre de la classe puis le principe du calcul reste le même.

Intervalles de valeurs

Moyenne

On étudie le salaire mensuel des salariés d'une entreprise (par mois et millier d'euros) que l'on regroupe par classe dans le tableau suivant :

Salaire	$]0, 1]$	$]1, 1.5]$	$]1.5, 2]$	$]2, 2.5]$	$]2.5, 3]$	$]3, 5]$
Effectif	10	4	12	8	6	16

Intervalles de valeurs

Moyenne

On étudie le salaire mensuel des salariés d'une entreprise (par mois et millier d'euros) que l'on regroupe par classe dans le tableau suivant :

Salaire]0, 1]]1, 1.5]]1.5, 2]]2, 2.5]]2.5, 3]]3, 5]
Effectif	10	4	12	8	6	16

La valeur moyenne est alors égale à :

$$\frac{1}{56} \times (10 \times 0.5 + 4 \times 1.25 + 12 \times 1.75 + 8 \times 2.25 + 6 \times 2.75 + 16 \times 4) = 2.3$$

soit un salaire moyen de 2300 euros pour les salariés de cette entreprise.

Cas de la variance

On utilisera le même procédé pour calculer la variance en considérant le centre des classes.

Remarques

Lorsque que l'on a des intervalles de valeurs

En ce qui concerne la médiane, il sera bon de faire la distinction entre **la classe médiane** et **la valeur médiane** (que l'on pourra uniquement extrapoler dans le cas où l'on manipule des classes de valeurs):

- **classe médiane** : elle sépare notre échantillon en deux parts égales, i.e. 50% de l'échantillon à une valeur à la fois inférieure et supérieure aux valeurs de la classe.
- **valeur médiane** : il faut cette fois ci estimée la valeur en utilisant les informations relatives à la classe médiane.

Intervalles de valeurs

Médiane

On étudie le salaire mensuel des salariés d'une entreprise (par mois et millier d'euros) que l'on regroupe par classe dans le tableau suivant :

Salaire	$]0, 1]$	$]1, 1.5]$	$]1.5, 2]$	$]2, 2.5]$	$]2.5, 3]$	$]3, 5]$
Effectif	10	4	12	8	6	16
Effectif cumulé	10	14	26	34	40	56

Intervalles de valeurs

Médiane

On étudie le salaire mensuel des salariés d'une entreprise (par mois et millier d'euros) que l'on regroupe par classe dans le tableau suivant :

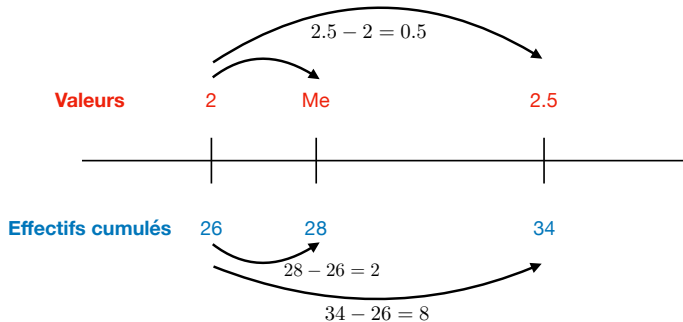
Salaire	$]0, 1]$	$]1, 1.5]$	$]1.5, 2]$	$]2, 2.5]$	$]2.5, 3]$	$]3, 5]$
Effectif	10	4	12	8	6	16
Effectif cumulé	10	14	26	34	40	56

L'effectif total étant égal à 56, il faut donc trouver l'intervalle dans lequel se trouve au moins 28 individus. Dans cet exemple, la classe médiane est égale à $]2, 2.5]$.

La valeur médiane peut-être estimée par une méthode linéaire.

Intervalles de valeurs

Médiane



On calcule les différences « extrêmes »

Puis la différence entre la valeur à gauche et la valeur de l'effectif associé à la médiane

On calcule le produit en croix suivant :

$$2 + \frac{8}{0.5} \left| \frac{2}{?} \right. = + \frac{0.5 \times 2}{8} = 2.125$$

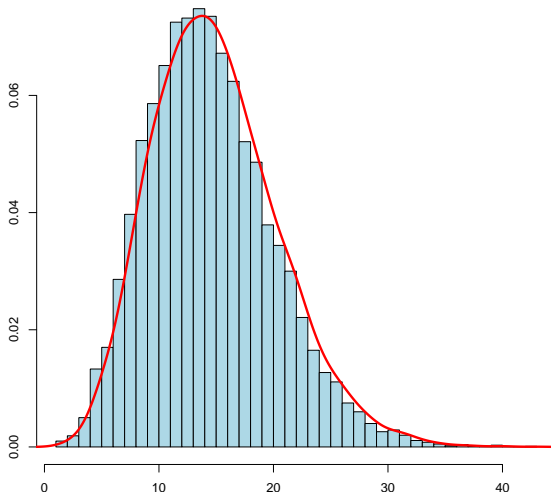
Définition

La **Skewness** est une mesure de l'**asymétrie** d'une densité de probabilité, pour une variable aléatoire prenant des valeurs réels. Elle permet de mesurer où se situe la moyenne de cette variable par rapport à sa médiane (mesure d'asymétrie de Pearson).

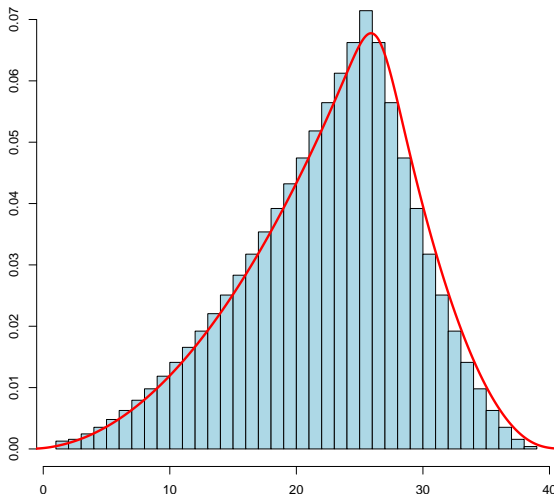
Pour une distribution uni-modale :

- $\text{Skewness} < 0$: la queue de la distribution est étalée vers **la gauche**, i.e. **la moyenne est plus petite que la médiane**.
- $\text{Skewness} > 0$: la queue de la distribution est étalée vers **la droite**, i.e. **la moyenne est plus petite que la médiane**.

Positive Skewed with mean = 14.95 and median = 14



Negative Skewed with mean = 22.92 and median = 24



Passons à la pratique

Enquête Tourisme

Variables quantitatives

Ouvrez l'enquête "Tourisme" -> Analyses -> Nouvelle Analyse -> Tri à plat et sélectionnez la variable 7 : "Dépenses totales".

Vous devriez voir le tableau ci-dessous :

	Effectifs	% Obs.
Non-réponse	246	44,6%
Moins de -279	0	0%
De -279 à 2180	107	19,4%
De 2181 à 4640	145	26,3%
De 4641 à 7100	32	5,8%
7101 et plus	22	4%
Total	552	100%

Regardons maintenant les valeurs qui se trouvent sous le tableau.

Focus

Réponses effectives : 306
Taux de réponse : 55,4%

Non-réponse(s) : 246
Moyenne : 3410.62 ; Médiane : 3000 ; Ecart-type : 2460.17 ; Min - Max : 500 - 15000 ; Somme : 1043650

Pour des variables **quantitatives**, le logiciel vous retourne par défaut :

- **La moyenne**, correspond ici à la moyennes des dépenses par répondant
- **La médiane**, i.e. la valeur telle que 50% aient dépensé moins que ce montant
- **L'écart-type**, i.e. une mesure de dispersion des dépenses
- **Min-Max**, i.e. plus grande et la plus petite dépense totale
- **Somme**, i.e. la somme des dépenses totales

Sphinx vous propose une mise en classe par défaut pour des valeurs continues :

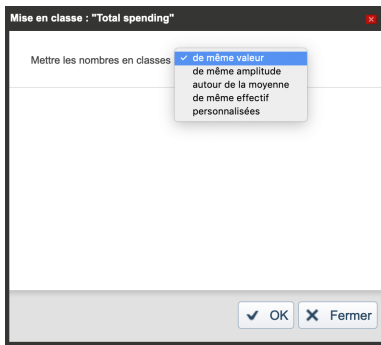
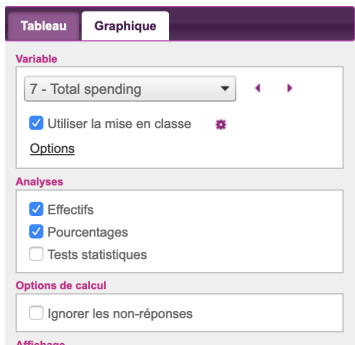
	Effectifs	% Obs.
Non-réponse	246	44,6%
Moins de -279	0	0%
De -279 à 2180	107	19,4%
De 2181 à 4640	145	26,3%
De 4641 à 7100	32	5,8%
7101 et plus	22	4%
Total	552	100%

Il est cependant possible de la modifier selon vos besoins :

Enquête Tourisme

Mise en classe

Il est cependant possible de modifier la mise en classe selon vos besoins en cliquant sur "Utiliser la mise en classe" puis en cliquant sur la petite étoile violette :



Plusieurs options de mise en classe :

- **de même valeur** : pas l'idéal quand la variable prend un nombre important de valeurs
- **de même amplitude** : permet de définir des classes avec des amplitudes identiques - option : nombre de classes
- **autour de la moyenne** : définit les classes autour de la moyenne en fonction de l'écart-type - options : nombre de classes et écart-type (pour l'étendue de classe)
- **de même effectif** : lorsque l'on souhaite avoir des classes avec des effectifs identiques - option : nombre de classes
- **personnalisées** : lorsque l'on souhaite définir soit même les différentes classes

Enquête Tourisme

Variables qualitatives - catégorielles ou nominales

Ouvrez l'enquête "Tourisme" -> Analyses -> Nouvelle Analyse -> Tri à plat et sélectionnez la variable 3 : "Mode d'hébergement".

Vous devriez voir le tableau ci-dessous :

	Effectifs	% Obs.
Non-réponse	16	2,9%
Hôtel	104	18,8%
Camping	113	20,5%
Location / gîte	140	25,4%
Famille / amis	179	32,4%
Total	552	100%

Regardons maintenant les valeurs qui se trouvent sous le tableau.

Enquête Tourisme

Variables qualitatives - catégorielles ou nominales

Focus

Réponses effectives : 536
Taux de réponse : 97,1%

Non-réponse(s) : 16
Modalité la plus citée : Famille / amis

Pour des variables **qualitatives** avec plusieurs modalités, le logiciel vous retourne :

- uniquement le **mode** qui est la seule "grandeur statistique" ayant un sens pour ce type de variables.

Enquête Tourisme

Variables qualitatives - ordinales

Ouvrez l'enquête "Automobiles" -> Analyses -> Nouvelle Analyse -> Tri à plat et sélectionnez la variable 19 : "Importance de la formule tout inclus".

Vous devriez voir le tableau ci-dessous :

	Effectifs	% Obs.
Pas d'accord du tout	66	12%
Plutôt pas d'accord	125	22,6%
Cela dépend	171	31%
Plutôt d'accord	121	21,9%
Tout à fait d'accord	69	12,5%
Total	552	100%

Regardons maintenant les valeurs qui se trouvent sous le tableau.

Focus

Réponses effectives : 552
Taux de réponse : 100%

Non-réponse(s) : 0
Modalités les plus citées : Cela dépend; Plutôt pas d'accord; Plutôt d'accord

Pour des variables **qualitatives ordinales**, le logiciel vous retourne :

- uniquement le **mode** qui est la seule "grandeur statistique" ayant un sens pour ce type de variables. Les grandeurs étudiées sont identiques à celles des variables qualitatives dites catégorielles.

On peut cependant les étudier d'un autre point de vue !

Enquête Tourisme

Variables ordinales → Variables quantitatives

Transformation des variables ordinales en variables quantitatives.
Idée : un jugement de valeur peut-être perçu comme une note.
Cliquer sur "Traiter les échelles en nombres"

The screenshot shows a software interface with two tabs at the top: "Tableau" and "Graphique". The "Graphique" tab is selected. Below the tabs, there is a section titled "Variable" in pink. It contains a dropdown menu showing "23 - Future intention" with left and right arrow buttons. Below the dropdown are two checkboxes: "Traiter les échelles en nombre" (checked) and "Regrouper les modalités" (unchecked). Below these is a link labeled "Options".

Below the "Variable" section is a section titled "Analyses" in pink. It contains three checkboxes: "Effectifs" (checked), "Pourcentages" (checked), and "Tests statistiques" (unchecked).

Below the "Analyses" section is a section titled "Options de calcul" in pink. It contains one checkbox: "Ignorer les non-réponses" (unchecked).

Enquête Tourisme

Variables ordinales → Variables quantitatives

Le tableau reste inchangé ...

	Effectifs	% Obs.
Pas d'accord du tout	66	12%
Plutôt pas d'accord	125	22,6%
Cela dépend	171	31%
Plutôt d'accord	121	21,9%
Tout à fait d'accord	69	12,5%
Total	552	100%

En revanche, si l'on regarde ce qui se trouve plus bas

Focus

Réponses effectives : 552
Taux de réponse : 100%

Non-réponse(s) : 0
Modalité la plus citée : Cela dépend ; Moyenne : 3 ; Ecart-type : 1.19 ; Min - Max : 1 - 5

On retrouve des grandeurs statistiques différentes !

- le mode : cela dépend
- la moyenne : 3
- l'écart-type : 1.19
- l'étendue : 1 - 5

Avec 1 : *Pas du tout d'accord* et 5 : *Tout à fait d'accord*.

La nature de la valeur centrale et du critère de dispersion dépend de la nature de la variable et des données :

- Variables quantitatives : on utilisera la *moyenne* (ou *médiane* en présence de valeurs extrêmes) ainsi que *l'écart-type*
- Variables qualitatives dites catégorielles : seul *le mode* est calculable
- Variables qualitatives ordinales :
 - 1) on s'intéresse à la modalité la plus citée (*le mode*)
 - 2) si les échelles sont traités en nombres : *même caractéristiques que pour les variables quantitatives*

La nature de la valeur centrale et du critère de dispersion dépend de la nature de la variable et des données :

- Variables quantitatives : on utilisera la *moyenne* (ou *médiane* en présence de valeurs extrêmes) ainsi que *l'écart-type*
- Variables qualitatives dites catégorielles : seul *le mode* est calculable
- Variables qualitatives ordinales :
 - 1) on s'intéresse à la modalité la plus citée (*le mode*)
 - 2) si les échelles sont traités en nombres : *même caractéristiques que pour les variables quantitatives*

Quid des questions ouvertes ?

Analyse des questions ouvertes

Différences

L'analyse aux 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

Mais il est toujours possible de les étudier et Sphinx va nous permettre de faire cela.

Prenons tout de suite un exemple.

Ouvrez l'enquête "Télévision" -> Analyses -> Nouvelle Analyse -> Analyse Textuelle et sélectionnez la variable 8 : "Emission préférée".

Vous trouverez cet en-tête :

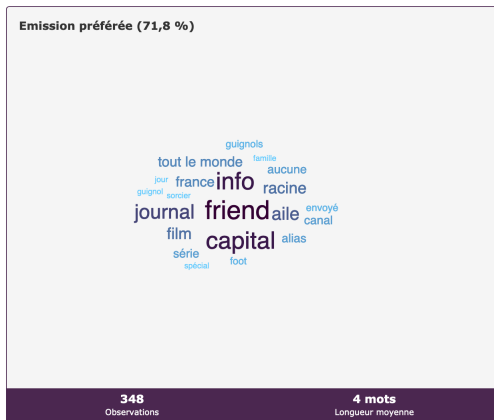
8 - Emission préférée



- taux de réponse à la question
- nombre de mots identifiés dans les réponses
- longueur médiane des réponses

Analyse des textes

Enquête Télévision



Met en avant les mots du corpus (synonymes compris) les plus mentionnés parmi l'ensemble des réponses.

Pour cette analyse, seules les variables précédées du sigle **ab** sont disponibles à l'étude, i.e. les questions textes ouvertes. Il est possible de représenter les réponses de différentes façon :

- Nuages de mots-clefs (représentation par défaut)
- Verbatim (tableau contenant les différentes réponses par ordre alpha-numérique)
- Codification : créer sa propre codification -> création d'une nouvelle variable.

Analyse des textes

Nuages de mots clefs - Global

Variable

8 - Emission préférée ▼ ◀ ▶

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

Fréquence minimum :

global

Affichage

Titre : Automatique ▼

☐ Afficher un commentaire personnalisé

- Global -> présentation par défaut : on ne prend aucune autre variable en compte
- Choix du type de mot pris en compte pour l'étude : ici les **adjectifs** et **noms communs**
- Nombre de mots dans le nuage : ici 50 mots maximum
- Fréquence minimale pour la prise en compte : seuls les mots apparaissant au minimum 5 fois dans le corpus sont pris en compte

Analyse des textes

Nuages de mots clefs - Par contexte

Variable

8 - Emission préférée

Analyses

☒ 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 25

Fréquence minimum :

par contexte 5

☒ Utiliser les mots spécifiques

Affichage

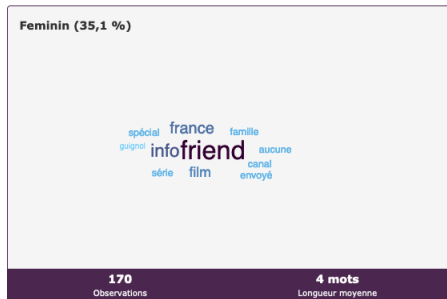
Titre : Automatique
☐ Afficher un commentaire personnalisé

- Affiche un nuage de points selon le contexte, i.e. en fonction des modalités d'une autre variable, ici le "Sexe" des répondants
- Les autres options restent inchangées

Analyse des textes

Nuages de mots clefs - Par contexte

Voilà ce que l'on vous retourne :



Ce qui montre que les hommes et femmes aiment des émissions différentes : "Capital" et des émissions culturelles économiques vs. séries pour les femmes dans la population échantillonnée.

Analyse des textes

Verbatim - Généralités

The screenshot shows a configuration window titled 'Tableau' with a purple header. It is divided into three sections: 'Variable', 'Analyses', and 'Affichage'. In the 'Variable' section, a dropdown menu shows '8 - Emission préférée'. The 'Analyses' section contains several options: 'Nuages de mots-clés' (radio button), 'Verbatim' (radio button with a blue dot), 'Codification' (radio button), 'Signature' (checkbox with a blue checkmark), 'Evaluation' (checkbox), and 'Simplifier' (checkbox). There are also two dropdown menus: one next to 'Verbatim' showing 'Tableau' and another next to 'Signature'. The 'Affichage' section has a dropdown for 'Titre' set to 'Automatique' and a checkbox for 'Afficher un commentaire personnalisé' which is unchecked.

- Signature : affiche la modalité d'une autre variable associée aux différentes réponses à la question ouverte.
- Evaluation : affiche le score de la variable ordinale sélectionnée et associée aux réponses de la question ouverte
- Simplifier : permet de réduire le nombre de lignes du tableau

Analyse des textes

Verbatim - Tableau

8 - Emission préférée

Rechercher les textes contenant

▼ Emission préférée

?

?

100 minutes pour convaincre

20h10 pétantes

24

24H

24 heures chrono

24 heures chrono

24 HEURES CHRONO

Actuellement: Newport Beach et C dans l'air

alias

Alias

ALIAS

Alias

alias

Alias

Alias

Alias

alias

Analyse des textes

Verbatim - Signature "Sexe"

8 - Emission préférée

Rechercher les textes contenant

▼ Emission préférée	◆ Sexe
?	Feminin
?	Feminin
100 minutes pour convaincre	Masculin
20h10 pétantes	Masculin
24	Masculin
24H	Feminin
24 heures chrono	Feminin
24 heures chrono	Masculin
24 HEURES CHRONO	Feminin
Actuellement: Newport Beach et C dans l'air	Masculin
alias	Masculin
Alias	Feminin
ALIAS	Masculin
Alias	Feminin
alias	Feminin
Allas	Feminin
Alias	Masculin
Alias	Feminin
alias	Feminin

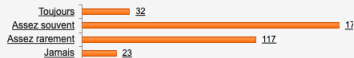
Analyse des textes

Verbatim - Evaluation "Note Globale"

8 - Emission préférée

2,62 / 4

Calculé sur la base de 345 avis



Rechercher les textes contenant



★★★★★ 3 / 4

Urgences

★★★★★ 3 / 4

dismissed sur mtv

★★★★★ 3 / 4

Allas (série)

★★★★★ 3 / 4

capital

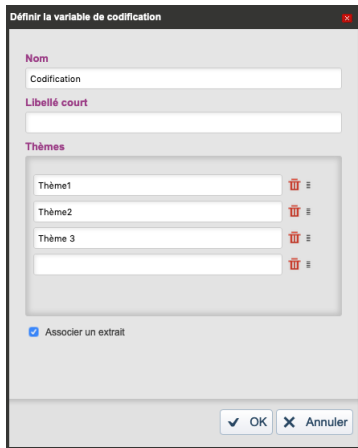
Permet d'encoder soit même les réponses aux questions ouvertes

The screenshot shows a software interface with a purple header bar labeled 'Tableau'. Below it, a 'Variable' section contains a dropdown menu currently showing '8 - Emission préférée'. Underneath, the 'Analyses' section has four radio button options: 'Nuages de mots-clés', 'Verbatim', 'Codification' (which is selected with a blue dot), and 'Par contexte'. At the bottom of the 'Analyses' section, there is a 'Codification' label followed by three icons: a pencil, a bar chart, and a close button. Below this is a button with a plus icon and the text 'Construire une grille thématique'.

- choix des catégories
- classement des réponses selon des critères **subjectifs**
- création/suppression d'une nouvelle variable avec des réponses encodées
- analyse de cette nouvelle variable

Analyse des textes

Codification - Procédure



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

- Nom**: A text input field containing the word "Codification".
- Libellé court**: An empty text input field.
- Thèmes**: A list of three themes, each with a text input field and a delete icon (a trash can) to its right:
 - Thème1
 - Thème2
 - Thème 3
- ☒ Associer un extrait: A checkbox that is currently checked.
- Buttons**: At the bottom right, there are two buttons: "OK" with a checkmark icon and "Annuler" with an 'X' icon.

- **Nom** : Nom de la nouvelle variable - ici Codification
- **Libellé court** : Un nom plus court qui apparaîtra lors de l'analyse
- **Thèmes** : noms des différentes modalités données lors de la phase d'encodage - ici Thème 1, Thème 2, Thème 3

Puis cliquer sur "Ok"

Viens ensuite l'étape fastidieuse où il faudra encoder chaque réponse, i.e. les placer dans une catégorie.

Votre émission ou série préférée...

capital, zone interdite

test

☒ Thème1

☐ Thème2

☐ Thème 3

Extrait test

✓ Enregistrer

Une fois la tâche effectuée, il est possible d'analyser les résultat en retournant dans la partie "Analyse" -> Tri à plat et en sélectionnant la variable créée (elle porte le nom écrit dans **libellé court**)

Analyse des textes

Codification - Résultat

	Effectifs	% Rep.
Thème1	5	22,7%
Thème2	12	54,5%
Thème 3	5	22,7%
Total	22	

Réponses effectives : 22
Taux de réponse : 4,5%

Non-réponse(s) : 463
Modalité la plus citée : Thème2

