



BBA-1

Analyse de Contenu et de Données Tutorat 1 : Qualité d'un échantillon

Guillaume Metzler

Université de Lyon, Université Lumière Lyon 2
Laboratoire ERIC EA3083, Lyon, France
guillaume.metzler@live.fr

Résumé

Ce premier TD a pour but de vous faire manipuler Sphinx et ainsi de mettre en pratique ce qui a été vu lors du cours introductif.

Ce travail là est effectué ensemble pendant la séance et ne sera pas évalué. Un travail similaire vous sera transmis en fin de séance et sera à rendre avant la séance suivante. Ce dernier sera lui évalué.

Etude de l'enquête Automobile

Cet exercice portera sur l'enquête *Automobile* de Sphinx.

1. Ouvrir l'enquête *Automobile*
2. Combien de personnes ont répondu à cette enquête ?
3. Combien de questions ont été posées au cours de cette enquête ?
4. Combien de variables comportent cette étude ? Que peut-on observer par rapport à la question précédente ?
5. Combien de femmes ont répondu à cette enquête ?
6. En excluant les non-réponses, combien de modalités la variable **Marque** présente-t'elle ?
7. Quel est le nombre de répondant masculin possédant une BMW ?
8. En reprenant la question précédente, que peut-on dire de la puissance de ces voitures ?
9. Toujours par rapport à la question 8, combien de modalités présente la variable **Ancienneté** ?
10. Supprimons maintenant les filtres précédents et trions les données par rapport à la variable **Vitesse** par ordre **ascendant**, on n'affichera que les 5 premières variables pour cela.
 - (a) Quel est le point commun entre les dix premiers répondants ?
 - (b) Qu'observez vous au niveau des réponses fournies par les répondants ? Est-ce justifiée ?
 - (c) Quelle est la nature de la variable **Vitesse** ?
 - (d) Toujours en conservant ce tri, quelle est la première **Marque** de voiture apparaissant dans le tableau ?
 - (e) La question 4 portant sur l' **Appartenance** est-elle une question ouverte ou fermée ?
11. On se concentre maintenant sur la partie *Qualité de l'échantillon*.

- (a) Que peut-on dire du taux de remplissage global de l'enquête? Faire le lien avec la question 10.a).
 - (b) Combien de questions ont été intégralement documentées?
 - (c) Combien de questions numériques peut on dénombrer dans ce questionnaire? En donner un exemple.
 - (d) Globalement, est-il vrai de dire que la majorité des individus a répondu à l'ensemble du questionnaire?
 - (e) Quelle est la nature de la variable la moins bien documentée par l'ensemble des observations?
 - (f) Peut-on dire que les questions d'échelles ont été globalement correctement documentées par l'ensemble des observations.
12. Dernière série de questions. retournons sur notre tableau de données.
- (a) Que peut-on dire concernant le **Type** de véhicule dont l'**Entretien 2** est compris entre 200 et 400 euros?
 - (b) Si l'on considère **entretien1** \geq 400 OU **entretien2** \in [500, 1000], qu'elle est la marque de la voiture qui ressort en premier?
 - (c) En regardant uniquement les dix premiers résultats, peut-on affirmer que les voitures **puissantes** sont celles dont le coût de l'**entretien2** est le plus élevé?
 - (d) Toujours par rapport aux 10 premiers résultats, peut-on affirmer que les **voitures anciennes** sont celles dont l'**entretien2** est le plus élevé?
 - (e) Que peut-on dire du **coût de l'entretien** mais aussi du statut (ou **CSP**) des détenteurs d'une BMW?