

Data Analysis et Méthodologie d'enquêtes

Bachelor – Créalise – 2024/2025

Sommaire

- Partie 1: Introduction et rappels théoriques
- Partie 2: Collecte et préparation des données
- Partie 3: Analyse des données d'enquêtes
- Partie 4: Présentation des résultats et conclusions

Partie 1

Introduction générale (15 min)

- Objectifs du cours
- Importance des enquêtes dans la recherche et l'analyse de données
- Différence entre enquête et sondage, échantillonnage, représentativité

Méthodologie d'enquête : rappels théoriques (30 min)

- Les étapes d'une enquête : formulation de la question de recherche, conception de l'enquête, échantillonnage, collecte de données
- Types d'enquêtes (enquêtes par questionnaire, enquêtes qualitatives)
- Techniques de construction de questionnaires

Introduction à l'analyse des données (30 min)

- Types de données (quantitatives, qualitatives)
- Variables (dépendantes, indépendantes, continues, discrètes)
- Statistiques descriptives de base : moyenne, médiane, écart-type

Atelier pratique rapide (15 min)

- Création d'un petit questionnaire simple avec Google Forms ou autre outil
- Discussions sur les choix de questions



Introduction et importance des enquêtes

Objectif

- Présenter les objectifs du cours, qui sont de comprendre les méthodes d'enquête, de maîtriser les techniques de collecte et d'analyse de données, et de développer des compétences pratiques pour concevoir et analyser des enquêtes.
 - Souligner l'importance des enquêtes dans différents domaines (recherche académique, entreprises, secteur public, marketing) et leur rôle dans la prise de décision éclairée.
- 



Introduction et importance des enquêtes

Importance des enquêtes dans la recherche et l'analyse de données :

- Les enquêtes sont une méthode clé pour recueillir des données directement auprès des personnes, notamment lorsqu'il s'agit de comprendre des comportements, des opinions ou des préférences.
 - L'enquête permet de récolter des données primaires, en opposition aux données secondaires (données déjà existantes). Ce type de données est crucial pour répondre à des questions spécifiques et contemporaines.
- 



Introduction et importance des enquêtes

Différence entre enquête, sondage, échantillonnage, représentativité :

- **Enquête** : Processus de collecte de données par des questions posées à un groupe spécifique.
 - **Sondage** : Type particulier d'enquête, souvent à grande échelle, visant à obtenir une vue d'ensemble d'une population en interrogeant un échantillon représentatif.
 - **Échantillonnage** : Processus de sélection d'un sous-ensemble de la population cible pour l'enquête, avec des techniques comme l'échantillonnage aléatoire, stratifié, etc.
 - **Représentativité** : Importance que l'échantillon soit un bon reflet de la population totale afin de garantir que les conclusions puissent être généralisées.
- 

Méthodologie de l'enquête : rappels théoriques

Étapes d'une enquête

Formulation de la question de recherche : Définir clairement l'objectif de l'enquête. Par exemple, « Quel est le degré de satisfaction des clients sur un produit ? »

Conception de l'enquête : Choisissez le type d'enquête (quantitative ou qualitative), la méthode de collecte et les outils.

Échantillonnage : Sélectionnez les participants à l'aide de techniques telles que l'échantillonnage aléatoire ou stratifié.

Collecte de données : Administrez l'enquête au moyen de questionnaires, d'entrevues ou de groupes de discussion.

Questionnaire Construction Techniques

Clarté des questions : Évitez les ambiguïtés, les doubles négations et la complexité.

Types de questions : Ouverte ou fermée (à choix multiples, échelles de Likert).

Ordre des questions : Commencez par des questions simples et passez à des questions plus complexes.

Durée du questionnaire : Établissez un équilibre entre la quantité d'information nécessaire et la fatigue des répondants.

Enquête par questionnaire : Structurée avec des questions fermées (oui/non, échelles) ou ouvertes pour l'analyse quantitative.

Enquête qualitative : Recueil des données qualitatives par le biais d'entrevues approfondies ou de groupes de discussion.

Types d'enquêtes



Introduction à l'analyse des données

Types de données

Quantitatif : Données numériques qui peuvent être mesurées (p. ex., âge, revenu).

Qualitatif : Données descriptives non numériques (par exemple, opinions, comportements).

Statistiques descriptives de base

- Moyenne : somme de toutes les valeurs divisée par le nombre total de valeurs.
- Médiane : valeur centrale lorsque les données sont classées par ordre croissant.
- Écart-type : mesure la dispersion des données autour de la moyenne.

Variables

Variables dépendantes et indépendantes :

- Variable indépendante : Facteur manipulé ou observé (p. ex., sexe, âge).
- Variable dépendante : Résultat mesuré ou observé (p. ex., satisfaction du client).

Variables continues et discrètes :

- Variables continues : valeurs infinies (par exemple, taille, poids).
 - Variables discrètes : valeurs finies (p. ex., nombre d'enfants, évaluation sur une échelle).
- 

Statistiques – Types de variables



Atelier pratique rapide



Création d'un questionnaire simple

Objectif : Créer un questionnaire en ligne avec des questions fermées et ouvertes sur un sujet spécifique (par exemple, la satisfaction des étudiants sur un cours).

Étapes:

Définissez le thème et les objectifs.

Rédigez 5 à 7 questions en utilisant différents formats (choix multiples, échelles de Likert, réponses ouvertes).

Structurez les questions pour plus de clarté et d'attention.



Commentaires sur les défis

- Rencontré lors de la création du questionnaire.

Discussion sur les choix de questions

- Importance d'une bonne formulation pour éviter les biais.

- Comprendre comment les réponses peuvent être utilisées dans l'analyse de données.



Conclusion

En conclusion, cette séance a permis d'introduire les enquêtes, les sondages et l'analyse des données.

Nous avons exploré l'importance des enquêtes dans différents domaines et la différence entre les enquêtes, les sondages, l'échantillonnage et la représentativité.

Les rappels théoriques ont porté sur la méthodologie d'enquête, les types d'enquêtes, les techniques de construction des questionnaires, les types de données, les variables et les statistiques descriptives de base.

L'atelier pratique rapide a permis aux participants de créer un questionnaire simple et de discuter des choix de questions et de l'analyse des données.

Cette séance jette les bases d'une exploration plus approfondie des enquêtes et de l'analyse des données.



Sommaire

- Partie 1: Introduction et rappels théoriques
- Partie 2: Collecte et préparation des données
- Partie 3: Analyse des données d'enquêtes
- Partie 4: Présentation des résultats et conclusions

Partie 2

Méthodes de collecte de données (30 min)

- Outils de collecte (sondages en ligne, entretiens, observations)
- Échantillonnage et biais (aléatoire, stratifié, quotas)
- Techniques pour minimiser le biais dans les enquêtes

Préparation des données pour l'analyse (30 min)

- Nettoyage des données : gestion des valeurs manquantes, doublons, outliers
- Codage des données qualitatives
- Introduction à la normalisation et à la transformation des données

Atelier pratique (30 min)

- Importation et nettoyage d'un jeu de données (par exemple avec Excel ou Python)
- Identification et traitement des anomalies dans les données

Méthodes de collecte des données

Outils de collecte



Sondages en ligne

- Utilisez des plateformes telles que Google Forms, SurveyMonkey ou Typeform
- Avantages : rapide, faibles coûts, large échantillonnage
- Limites : accès limité à Internet, faible taux de réponse

Entrevues

- En présentiel, par téléphone ou par vidéoconférence
- Avantages : réponses plus profondes, compréhension des motivations, des attitudes et des émotions
- Limites : chronophage, coûteux, difficulté à normaliser les résultats



Commentaires

- Recueil de données sans interaction directe avec le sujet. Utilisée notamment dans des études de comportements ou d'interactions sociales.
 - Avantages : données plus objectives, non biaisées par la perception du répondant.
 - Limites : pas adapté pour comprendre des processus internes ou des opinions.
- 



Echantillonnage et biais

Types d'échantillonnage :

- **Échantillonnage aléatoire** : Tous les membres de la population ont une chance égale d'être sélectionnés. Avantage : garantit une représentativité statistique.
 - **Échantillonnage stratifié** : La population est divisée en sous-groupes (strates) homogènes, puis un échantillon aléatoire est pris dans chaque strate. Avantage : plus précis si les strates sont bien définies.
 - **Échantillonnage par quotas** : Échantillon construit pour refléter certaines caractéristiques de la population (exemple : âge, genre). Avantage : contrôle des quotas pour assurer la représentativité, mais moins rigoureux qu'un échantillonnage aléatoire.
- 



Echantillonnage et biais

Biais dans les enquêtes :

- **Biais de sélection** : Se produit lorsque l'échantillon ne reflète pas correctement la population.
 - **Biais de réponse** : Survient lorsque les répondants donnent des réponses inexactes ou biaisées (par désir de plaire, par crainte de jugement).
 - **Biais de non-réponse** : Lorsque certains groupes de la population ne répondent pas à l'enquête, créant ainsi un manque de représentativité.
- 



Echantillonnage et biais

Techniques pour minimiser le biais :

- S'assurer que l'échantillonnage est aléatoire lorsque possible.
 - Utiliser des questions neutres pour éviter de diriger les réponses.
 - Offrir une confidentialité totale ou partielle pour réduire les biais liés à la pression sociale
- 



Préparation des données pour l'analyse

Nettoyage des données

- **Gestion des valeurs manquantes :**
 - Identifier les valeurs manquantes à l'aide d'outils comme Excel ou Python (pandas).
 - Stratégies :
 - **Imputation** : Remplacer les valeurs manquantes par la moyenne, la médiane, ou une autre valeur calculée.
 - **Suppression** : Supprimer les lignes ou colonnes où les valeurs manquent si l'impact est minime.
 - Importance de vérifier si les valeurs manquantes sont aléatoires ou suivent un pattern qui pourrait révéler des biais.
 - **Doublons :**
 - Identification des réponses en double. Dans des outils comme Excel, utiliser la fonction "Supprimer les doublons". En Python, la méthode `drop_duplicates()` est utilisée pour enlever les lignes dupliquées.
 - Importance : Les doublons peuvent fausser l'analyse, surtout dans les enquêtes où chaque participant ne doit répondre qu'une seule fois.
 - **Outliers (valeurs aberrantes) :**
 - Détection des valeurs aberrantes avec des visualisations (boîtes à moustaches, histogrammes) ou des méthodes statistiques (éliminer les points à plus de 3 écarts-types de la moyenne).
- 



Préparation des données pour l'analyse

Codage des données qualitatives

- Pour les questions ouvertes, le codage consiste à transformer les réponses textuelles en catégories ou thèmes.
 - **Exemple :** Pour la question "Quels sont vos loisirs ?", les réponses "football", "basketball", "natation" peuvent être codées sous un thème "Sports".
 - Méthodes de codage : manuel (analyse des réponses et regroupement) ou automatique (text mining avec des outils comme Python - NLTK ou spaCy)
- 



Préparation des données pour l'analyse

Introduction à la normalisation et la transformation des données

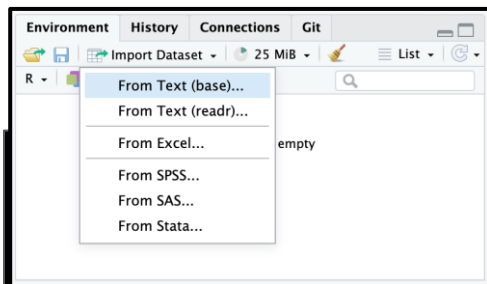
Normalisation : Transformation des données pour qu'elles aient une échelle comparable (souvent entre 0 et 1). Utile pour les algorithmes qui sont sensibles aux échelles (par exemple, les algorithmes d'apprentissage automatique).

- Techniques : min-max scaling, Z-score normalization (soustraction de la moyenne et division par l'écart-type).

Transformation des données : Parfois, les données doivent être transformées pour mieux correspondre aux hypothèses des méthodes d'analyse. Exemple : logarithmisation des données avec une forte asymétrie.



Atelier pratique

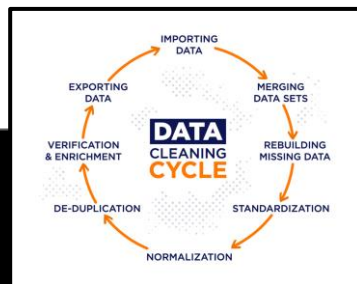


Importation et nettoyage d'un ensemble de données

Étape 1 : Importation du jeu de données à l'aide d'Excel ou de Python (pandas)

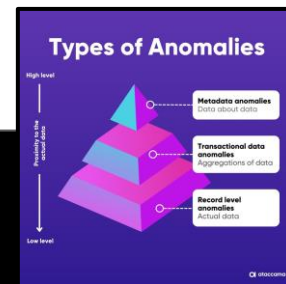
Étape 2 : Identification des anomalies (valeurs manquantes, doublons, valeurs aberrantes)

Étape 3 : Nettoyage du jeu de données (suppression des doublons, gestion des valeurs manquantes, gestion des valeurs aberrantes)



Nettoyage de l'ensemble de données

- Suppression des doublons
- Gestion des valeurs manquantes
- Gestion des valeurs aberrantes



Identification et traitement des anomalies de données

- Vérification des valeurs manquantes
- Vérification des doublons
- Visualiser la distribution des données pour identifier les valeurs aberrantes



Conclusion

En conclusion, les méthodes de collecte de données telles que les sondages en ligne, les entretiens et les commentaires fournissent des informations précieuses pour l'analyse.

La préparation des données pour l'analyse implique des techniques de nettoyage des données telles que la gestion des valeurs manquantes, la suppression des doublons et la gestion des valeurs aberrantes.

L'atelier pratique permet aux participants de s'entraîner à importer et à nettoyer des ensembles de données, ainsi qu'à identifier et à traiter les anomalies de données.

En suivant ces étapes, les chercheurs et les analystes peuvent s'assurer de la qualité et de l'exactitude de leurs données pour une analyse plus approfondie.



Sommaire

- Partie 1: Introduction et rappels théoriques
- Partie 2: Collecte et préparation des données
- Partie 3: Analyse des données d'enquêtes
- Partie 4: Présentation des résultats et conclusions

Partie 3

Analyse exploratoire des données (EDA) (30 min)

- Visualisation des données : graphiques, histogrammes, boîtes à moustaches
- Analyse des distributions et des corrélations

Tests statistiques de base (30 min)

- Tests de significativité : Chi2, T-test
- Comparaison de moyennes, analyse de variance (ANOVA)

Atelier pratique (30 min)

- Réalisation d'une analyse exploratoire avec un logiciel (Excel, R ou Python)
- Visualisation et interprétation des résultats



Introduction

Dans cette session sur l'analyse des données d'enquête, nous explorerons le processus d'analyse exploratoire des données (EDA), les tests statistiques de base et mènerons un atelier pratique à l'aide d'Excel.

La session vise à fournir aux participants les outils et les techniques nécessaires pour analyser et interpréter efficacement les données d'enquête.



Analyse exploratoire des données (EDA) (30 min)

Visualisation des données

- **Graphiques et histogrammes dans Excel :**
 - **Histogramme :** Excel permet de créer facilement un histogramme à partir d'une série de données (par exemple, l'âge des répondants).
 - **Exemple pratique :** Les participants peuvent utiliser la fonctionnalité "Insertion > Graphique > Histogramme" pour visualiser la répartition des âges.
 - **Graphique à barres :** Utile pour représenter des variables catégorielles (par exemple, la répartition de la satisfaction des clients). Il est possible de le créer à partir des réponses à un questionnaire.
 - **Diagrammes circulaires :** Utilisés pour illustrer la répartition des catégories (par exemple, la proportion de clients satisfaits vs insatisfaits).
- **Boîtes à moustaches (boxplots) :**
 - En Excel, les boîtes à moustaches (disponibles via "Graphique à boîtes et moustaches") permettent de visualiser la distribution d'une variable (comme le revenu). Cela montre la médiane, l'étendue des données et les outliers éventuels.
 - **Exemple :** Un boxplot des revenus peut être créé pour observer si certains clients ont des revenus particulièrement élevés ou bas.

Analyse exploratoire des données (EDA) (30 min)

Analyse quantitative et corrélations

- **Distributions** : Les participants pourront utiliser les histogrammes et les boxplots pour observer les distributions de leurs données.
 - **Exemple pratique** : Dans Excel, ils peuvent examiner si la distribution des âges est normale ou asymétrique (ce qui peut avoir un impact sur la représentativité de l'échantillon).
- **Corrélation** : Excel propose aussi des outils pour calculer la corrélation entre deux variables quantitatives (par exemple, entre le revenu et la satisfaction).
 - **Exemple pratique** : Utilisation de la fonction "CORREL" dans Excel pour analyser si les clients avec des revenus plus élevés sont plus satisfaits.

Tests statistiques de base (30 min)

Tests de significativité avec Excel

- **Test du Chi2 (Chi-square test) :**
 - Ce test peut être réalisé dans Excel à l'aide des fonctions intégrées ou via des outils en ligne.
 - **Exemple :** Tester s'il y a une relation significative entre la satisfaction des clients et leur tranche de revenu (par exemple, clients satisfaits sont-ils plus souvent dans une tranche de revenu spécifique ?).
 - **Application en Excel :** Les élèves peuvent créer un tableau croisé dynamique (ou tableau de contingence) pour visualiser les résultats et utiliser la fonction "CHITEST".
- **T-test (Test de Student) :**
 - Comparer deux moyennes pour déterminer si la différence entre les groupes est significative.
 - **Exemple pratique :** Comparer la moyenne des revenus entre deux groupes (clients satisfaits vs insatisfaits) en utilisant la fonction "TEST.T" dans Excel.

Tests statistiques de base (30 min)

Analyse de variance

- **ANOVA (Analyse de Variance) :**
 - Ce test permet de comparer les moyennes de plusieurs groupes (par exemple, les niveaux de satisfaction en fonction de l'âge).
 - **Exemple pratique :** Utiliser la fonction "ANOVA" dans Excel via l'onglet "Analyse des données" pour comparer les moyennes de satisfaction entre plusieurs tranches d'âge.

Atelier pratique (30 min)

Réalisation d'une analyse exploratoire avec Excel :

- - Importez le jeu de données fourni dans Excel pour analyse.
 - - Exemple : Téléchargez le fichier CSV et ouvrez-le dans Excel.
 - - Créez des visualisations, telles que des histogrammes et des boîtes à moustaches, pour explorer les données.
 - - Exemple : créez un histogramme d'âge et une boîte à moustaches pour visualiser la répartition des revenus.
- Utilisez des tableaux croisés dynamiques pour examiner les relations entre les variables.

Exemple : Créez un tableau croisé dynamique pour analyser la répartition des niveaux de satisfaction entre les groupes d'âge.



Conclusion

En conclusion, cette session a fourni un aperçu des techniques d'analyse exploratoire des données (EDA), des tests statistiques de base et un atelier pratique utilisant Excel.

En utilisant la visualisation des données, en analysant les distributions et les corrélations, et en effectuant des tests statistiques, les participants ont acquis des compétences précieuses pour analyser et interpréter efficacement les données d'enquête.



Sommaire

- Partie 1: Introduction et rappels théoriques
- Partie 2: Collecte et préparation des données
- Partie 3: Analyse des données d'enquêtes
- Partie 4: Présentation des résultats et conclusions

Partie 4

Interprétation et communication des résultats (30 min)

- Présentation des résultats d'enquête : graphiques, tableaux
- Comment bien interpréter et communiquer les résultats d'une enquête

Éthique et limites des enquêtes (15 min)

- Confidentialité des données
- Biais et erreurs d'interprétation

Atelier final (45 min)

- Création d'un rapport synthétique sur les résultats d'une enquête fictive ou réelle (basée sur les données nettoyées et analysées)
- Présentation et discussion en groupe



Introduction

La présentation efficace des résultats est cruciale pour transmettre les résultats d'une analyse.

Cette section vous guidera sur la façon de structurer votre présentation, de résumer le contexte de l'enquête, de mettre en évidence les résultats importants et d'utiliser des visualisations pour améliorer la compréhension.

De plus, nous vous fournirons des conseils pour une bonne communication des résultats.



Structurer la présentation des résultats

Introduction : Résumer le contexte de l'enquête

- Donnez un bref aperçu de l'objectif et de l'échantillon de l'enquête.
- Définir le contexte de l'analyse et établir la pertinence des résultats.

Visualisations : Accompagnez les résultats de graphiques clairs

- Utilisez des graphiques appropriés tels que des histogrammes, des boîtes à moustaches et des graphiques à barres pour représenter visuellement les données.
- Choisissez des titres clairs et concis pour les graphiques afin de faciliter la compréhension.
- Exemple : Utilisez un graphique pour montrer la répartition de la satisfaction par groupe d'âge.

Résultats principaux : Mettre en évidence les statistiques descriptives importantes et les résultats des tests statistiques

- Mettez l'accent sur les statistiques descriptives clés telles que les moyennes, les médianes et les écarts-types.
- Discutez des résultats des tests statistiques et de leur signification.
- Exemple : « Les clients ayant un revenu supérieur à 50 000 euros ont un niveau de satisfaction nettement plus élevé. »

Conseils pour une bonne communication des résultats

Clarté : évitez le jargon technique et expliquez les termes complexes

- Utilisez un langage simple et évitez un jargon technique excessif.
- Fournir des explications pour des termes complexes tels que la valeur p et la corrélation.

Concision : se concentrer sur les points clés de l'analyse

- Hiérarchisez les principales constatations et évitez les détails inutiles.
- Gardez la présentation concise et attrayante pour maintenir l'attention du public.

Simplicité : utilisez des graphiques simples et évitez la surcharge d'informations

- Optez pour des cartes et des diagrammes simples et faciles à comprendre.
- Évitez de submerger le public avec des informations excessives.
- Concentrez-vous sur la présentation des aspects les plus importants de l'analyse.

Interprétation et implications des résultats

✱ Interprétation des résultats des tests

Dans cette section, nous discuterons de la façon d'interpréter les résultats en fonction des hypothèses initiales et d'explorer leurs implications pour l'enquête. Les résultats significatifs doivent être interprétés en termes pratiques, tels que : « Il existe une corrélation significative entre le revenu et la satisfaction des clients, ce qui suggère que les clients les plus riches ont tendance à être plus satisfaits. » Les résultats non significatifs doivent également être abordés, par exemple : « Il n'y a pas de différence significative de satisfaction en fonction de l'âge du client ».

✱ Biais et limites

Discutez des biais potentiels liés à l'échantillonnage ou à la collecte de données, par exemple : « L'échantillon peut ne pas être représentatif de tous les groupes d'âge, ce qui limite la généralisabilité des résultats. » De plus, soulignez les limites des outils d'analyse utilisés, par exemple : « L'analyse n'a pas pris en compte les facteurs externes tels que la localisation du client, qui pourraient influencer la satisfaction. »

Interprétation et implications des résultats



Interprétation des résultats des tests

- **Tests de significativité** : Si un test est significatif ($p < 0,05$), expliquer ce que cela signifie en pratique.
 - **Exemple** : "Il existe une corrélation significative entre le revenu et la satisfaction des clients, ce qui suggère que les clients plus riches tendent à être plus satisfaits."
- **Résultats non significatifs** : Si un test n'est pas significatif, il est aussi important d'expliquer cela et de discuter des raisons possibles.
 - **Exemple** : "Il n'y a pas de différence significative dans la satisfaction en fonction de l'âge des clients."



Biais et limites

- **Biais potentiels dans l'enquête** : Parler des éventuels biais liés à l'échantillonnage ou à la collecte des données.
 - **Exemple** : "L'échantillon n'est pas représentatif de toutes les tranches d'âge, ce qui peut limiter la généralisation des résultats."
- **Limites des méthodes utilisées** : Discuter des limites des outils ou des techniques d'analyse.
 - **Exemple** : "L'analyse n'a pas pris en compte des facteurs externes comme la localisation des clients, qui pourrait influencer la satisfaction."