

Méthodes Quantitatives pour les Sciences Humaines et Sociales

Séance 6 : Construire un questionnaire, recoder des données

E. Madinier

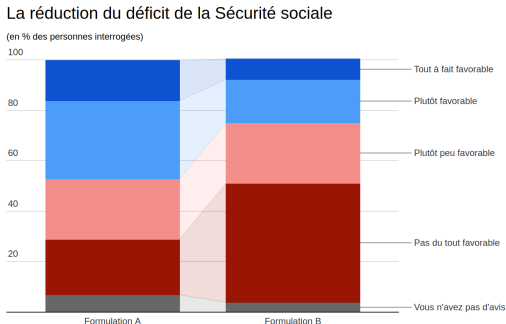
2025

Etes-vous personnellement : favorable/défavorable à

- A la réduction des dépenses de la Sécurité sociale parce que son déficit serait trop important (en l'occurrence, un déficit de 3,4 milliards d'euros) ?
- B une diminution des remboursements de certains médicaments et de certains soins médicaux pour assainir le déficit de la Sécurité sociale ?

Etes-vous personnellement : favorable/défavorable à

- A la réduction des dépenses de la Sécurité sociale parce que son déficit serait trop important (en l'occurrence, un déficit de 3,4 milliards d'euros) ?
- B une diminution des remboursements de certains médicaments et de certains soins médicaux pour assainir le déficit de la Sécurité sociale ?



La formulation A recueille 46% d'opinions défavorables, contre 71% pour la seconde.

La construction de questionnaires : un art complexe

Les enjeux

- Production de données fiables et exploitables
- Minimisation des biais de réponse
- Respect des contraintes légales et éthiques
- Optimisation du taux et de la qualité des réponses

Les étapes de la construction

① Définition des objectifs

- Que cherche-t-on à mesurer ?
- Quelle est la population cible ?

② Conception du questionnaire

- Structure et organisation
- Types de questions

③ Test et validation

- Pré-test sur échantillon réduit
- Ajustements nécessaires

Les types de questions

Questions fermées

- Choix unique
- Choix multiples
- Échelles de Likert
- Questions binaires

Questions ouvertes

- Réponse courte
- Réponse longue
- Numériques
- Dates

Avantages et inconvénients

Le choix dépend de :

- L'objectif de la question
- La capacité de traitement
- Le temps disponible

Les biais courants

- **Biais de désirabilité sociale**
 - Tendance à donner une réponse " socialement acceptable"
- **Biais d'acquiescement**
 - Tendance à être d'accord avec l'énoncé
- **Biais de position**
 - Impact de l'ordre des modalités
- **Biais de cadrage**
 - Influence du contexte de la question

Ordre des questions

- **Questions d'introduction**
 - Simples et engageantes
 - Non sensibles
- **Corps du questionnaire**
 - Thématiques principales
 - Questions sensibles
- **Questions signalétiques**
 - Profil du répondant
 - Fin du questionnaire

La qualité des réponses a tendance à baisser au cours du questionnaire, attention à la fatigue du répondant.

Effets d'ordre et de contexte

Types d'effets

- Effet de halo : tendance à ce qu'une impression générale influence l'évaluation d'autres caractéristiques
- Effet de contamination : influence d'une question sur les réponses aux questions suivantes
- Effet d'ancrage : tendance à s'appuyer sur la première information comme référence pour les jugements suivants
- Effet de consistance : tendance à maintenir une cohérence avec ses réponses précédentes

Organisation thématique

- Regroupement logique des questions
- Transitions entre thèmes
- Équilibre entre sections
- Filtres et sauts conditionnels

Bonnes pratiques de formulation

Règles générales

- Phrases simples et directes
- Éviter les doubles négations
- Questions neutres et non orientées

Pour aller plus loin

Lectures recommandées

- Grémy (1987) - "Les expériences françaises sur la formulation des questions d'enquête"
- Piau (2004) - "Expériences sur la formulation des questions"
- Guide CNRS sur le RGPD

Présentation de l'activité

Objectif

Construire une ou plusieurs questions pour un questionnaire sur un sujet donné

- Travail en groupes de 3-4 étudiants
- 15 minutes de préparation
- 15 minutes de restitution collective

Thème 1 : Économie

Sujet : Les pratiques de consommation des étudiants

- Budget mensuel
- Habitudes d'achat
- Critères de choix
- Impact de l'inflation

Thème 2 : Sociologie

Sujet : Les pratiques culturelles et de loisirs

- Fréquence des activités
- Types de pratiques
- Contexte social
- Budget alloué

Thème 3 : Science Politique

Sujet : L'engagement politique des jeunes

- Formes d'engagement
- Motivations
- Canaux d'information
- Participation électorale

Introduction à la protection des données

Contexte historique

- Évolution des technologies
- Multiplication des collectes de données
- Risques croissants pour la vie privée

Enjeux

- Protection de la vie privée
- Sécurité des données
- Confiance des répondants

Le RGPD : principes fondamentaux

- 1 **Licéité, loyauté et transparence**
- 2 **Limitation des finalités**
- 3 **Minimisation des données**
- 4 **Exactitude**
- 5 **Limitation de la conservation**
- 6 **Intégrité et confidentialité**
- 7 **Responsabilité**

Données personnelles vs données sensibles

Données personnelles

- Nom, prénom
- Adresse
- Email
- Téléphone
- IP

Déroptions possibles pour la recherche.

Données sensibles

- Origine ethnique
- Opinions politiques
- Convictions religieuses
- Données de santé
- Orientation sexuelle

Obligations légales

- **Information des personnes**

- Finalité du traitement
- Destinataires des données
- Durée de conservation

- **Recueil du consentement**

- Libre
- Spécifique
- Éclairé

Droits des personnes

- Droit d'accès
- Droit de rectification
- Droit à l'effacement
- Droit à la limitation
- Droit à la portabilité
- Droit d'opposition

Mentions légales obligatoires

Éléments à inclure

- Identité du responsable
- Finalités du traitement
- Base légale
- Destinataires
- Durée de conservation
- Droits des personnes
- Coordonnées DPO

Sécurité des données

- **Mesures techniques**

- Chiffrement
- Pseudonymisation
- Contrôle d'accès

- **Mesures organisationnelles**

- Procédures internes
- Formation du personnel
- Documentation

Impacts sur la construction du questionnaire

À prendre en compte

- Minimisation des données collectées
- Justification de chaque question
- Protection particulière des données sensibles
- Modalités de conservation et d'anonymisation

Documentation RGPD

Ressources essentielles

- Guide CNRS
- Documentation CNIL
- Modèles Sciences Po

Mise en route

Un exemple de base de données "complexe"

Nature de la décision	Age	Lieu de résidence
Acquitté	25	Paris (75001)
Non-lieu	32	Lyon centre
Condamné à 2 ans	45	Marseille 13001
Condamné avec sursis	19	Paris 75116
Relaxe	28	Lyon

Questions

- Comment analyser ces données en l'état ?
- Quels sont les problèmes évidents ?
- Par où commencer ?

Pourquoi recodons-nous ?

Le recodage n'est pas un "nettoyage"

- La complexité des données n'est pas une "saleté"
- Elle reflète la richesse de la réalité sociale
- Les choix de recodage sont des choix d'analyse

Exemple

Dans notre base :

- "Acquitté" et "Relaxe" : même réalité juridique ?
- Les arrondissements : niveau pertinent d'analyse ?
- Âge : variable continue ou en classes ?

Objectifs de la séance

- **Comprendre les enjeux**
 - Scientifiques
 - Techniques
 - Pratiques
- **Maîtriser les outils**
 - Fonctions de base d'Excel
 - Techniques de recodage
 - Bonnes pratiques
- **Développer un regard critique**
 - Sur nos choix de catégorisation
 - Sur les données existantes

Typologie des variables et recodages

Variables qualitatives

- Nominales
 - Harmonisation des libellés
 - Regroupement de modalités
- Ordinales
 - Respect de l'ordre
 - Échelles de mesure

Variables quantitatives

- Discrètes
 - Regroupement en classes
 - Traitement des valeurs rares
- Continues
 - Découpage en intervalles
 - Choix des seuils

Avant de commencer : les règles d'or

① Ne jamais modifier les données brutes

- Toujours créer une nouvelle colonne
- Garder une trace des modifications

② Documenter ses choix

- Justification des regroupements
- Description des procédures

③ Vérifier la cohérence

- Pas de données manquantes
- Pas d'erreurs de recopie

La démarche d'analyse préalable

Étape 1 : Le tri à plat

- Lister toutes les modalités
- Compter leurs occurrences
- Repérer les cas problématiques

Exemple : Nature de la décision

Modalité	Effectif
Acquitté	45
Relaxe	3
Non-lieu	12
Condamné à 2 ans	8
Condamné avec sursis	32

Les choix de recodage

Questions à se poser

- Quelles modalités regrouper ?
- Comment nommer les nouvelles catégories ?
- Quel niveau de détail garder ?

Points d'attention

- Balance entre précision et lisibilité
- Cohérence avec les objectifs de recherche
- Effectifs suffisants pour l'analyse

Méthode 1 : Rechercher/Remplacer

Procédure

- 1 Dupliquer la colonne (ex : "Decision_V2")
- 2 Utiliser CTRL+F ou cmd+F
- 3 Remplacer une modalité à la fois
- 4 Vérifier après chaque opération

Exemple

Dans notre base :

- "Acquitté" et "Relaxe" → "Non condamné"
- "Condamné*" → "Condamné" (avec *)

Focus : L'importance des vérifications

Les pièges courants

- Variantes orthographiques
 - Majuscules/minuscules
 - Accents
 - Espaces superflus
- Cas particuliers
 - Abréviations
 - Formulations alternatives

Méthode 2 : Tri et recodage par bloc

Procédure

- 1 Dupliquer la colonne
- 2 Trier par ordre alphabétique
- 3 Identifier les plages à recoder
- 4 Appliquer le nouveau code à toute la plage

Attention !

- Sélectionner toutes les colonnes avant de trier
- Vérifier la cohérence des lignes après tri
- Faire attention aux ruptures dans les plages

L'analyse préalable des distributions

Les indicateurs clés

- Minimum et maximum
- Moyenne et médiane
- Quartiles
- Distribution (histogramme)

Le choix des seuils

Seuils statistiques

- Quartiles
- Déciles
- Moyennes $\pm$ écart-type

Seuils naturels

- Âge légal (18 ans)
- Décennies (20, 30, 40...)
- Standards du domaine

Question clé

Quel type de seuil est le plus pertinent pour votre analyse ?

Méthode 1 : Tri et recodage manuel

Procédure

- 1 Dupliquer la colonne
- 2 Trier par ordre croissant
- 3 Identifier les points de rupture
- 4 Appliquer les codes par plage

Méthode 2 : La fonction SI()

Structure de base

$= SI(\text{test_logique}; \text{si_vrai}; \text{si_faux})$

Par exemple pour des tranches d'âge

$= SI(A2 < 25; "18 - 24"; "25 + ")$

$= SI(A2 < 25; "18 - 24";$
 $SI(A2 < 35; "25 - 34";$
 $SI(A2 < 45; "35 - 44"; "45 + ")))$

Exercice pratique

Objectif

Recoder une variable de votre choix dans la base de données

Discussion collective

- **Présentez vos choix**

- Quelles catégories ?
- Pourquoi ces regroupements ?
- Quelle méthode utilisée ?

- **Débat**

- Avantages/inconvénients des choix
- Solutions alternatives
- Implications pour l'analyse

Points clés à retenir

① Le recodage est un choix analytique

- Pas une simple opération technique
- Des implications importantes

② La méthode compte

- Documentation systématique
- Vérifications régulières