

Sull, D.N. 1999. Why Good Companies Go Bad? *Harvard Business Review*. July-August 1999: 42-52.

Wang, R.Y. et D.M. Strong. 1996. Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, 12(4): 5-34.

VI: Avoir un travail

Groupe d'indiv = ménage

Unité stat = personne en âge d'activité

VI: dépenses alimentaires

Groupe de log = immeuble / pote de maison

Unité stat = ménage / indiv composant le ménage

VI: Conditions de logement dans les écoles

Groupe étudiants : école

unité stat : étudiant

Enquête : activité organisée et méthodique de collecte de données sur des caractéristiques d'intérêt d'une partie ou de la totalité d'une population, en cas de besoin d'info

Liste de vérification de la planification

Voici certaines considérations à ne pas oublier lors de la planification des étapes de l'enquête :

1. Formulation de l'énoncé des objectifs

- besoins d'information de l'enquête,
- principales utilisations et principaux utilisateurs des données,
- définitions et concepts opérationnels, ménages, groupe, personnes, \Rightarrow Qui / Quoi, où, Quand
- matière de l'enquête, \rightarrow Plan d'éch, méthode de collect
- plan d'analyse (c.-à-d. la structure et le niveau de détail des résultats de l'enquête).

Voir le Chapitre 2 - Formulation de l'énoncé des objectifs

2. Sélection d'une base de sondage

- définition de la population cible, \rightarrow Qui / Quoi / où / Quand Pop Enq \rightarrow \rightarrow Pop Cible \rightarrow l
- définition de la population observée selon les bases de sondage disponibles, Recensement \rightarrow voir toute la population / Coûteux
- base de sondage aréolaire, liste ou base de sondage multiple, Eng-ech \rightarrow cible une partie de la pop / Economique
- utilisation des bases de sondage déjà créées, Combien 2 bases Aréolaire : constructions de recteurs
- coût d'établissement d'une nouvelle base de sondage, ou Plus géographique
- données disponibles pour établir une nouvelle base de sondage, Erreurs d'enquête :
- unités de la base de sondage, unité de référence \rightarrow dues à l'échantillonnage
- qualité de la base de sondage, \rightarrow non dues à l'échantillonnage
- sous-dénombrement, Exclusion des données
- surdénombrement, Ajout des données
- doubles,
- base de sondage à jour?

Taux de sondage

$$f = \frac{N}{n} \times 100$$

$$\Rightarrow \frac{\text{taille pop}}{\text{taille d'éch}} \times 100$$

Voir le Chapitre 3 - Introduction au plan d'enquête.

3. Choix du plan d'échantillonnage

- recensement ou échantillon,
- échantillonnage probabiliste ou non probabiliste pour l'enquête-échantillon,
- si l'échantillonnage probabiliste est choisi :
 - échantillonnage aléatoire simple,
 - échantillonnage aléatoire simple stratifié,
 - échantillonnage par grappes,
 - échantillonnage à plusieurs degrés,
 - échantillonnage à plusieurs phases,
- si l'échantillonnage stratifié est choisi :
 - variables de stratification,
 - méthode de répartition de l'échantillon en strates,
- taille de l'échantillon, \rightarrow base de sondage \rightarrow select' des élmnts de l'éch \rightarrow Collecte des données
- méthode d'estimation,
- degré de précision voulu (variance) des estimations,
- enquête unique ou réitérée?

éch probabiliste \rightarrow Aléatoire simple

\rightarrow Stratifié

\rightarrow Par degré

\rightarrow Systematique

Non probabiliste \rightarrow A priori

\rightarrow Accidentel

\rightarrow Boule de neige

\rightarrow Quotas

Age / Sexe / Form

$$\text{taille de l'éch} = n = \frac{N}{1 + N \times (0.05)^2}$$

Voir le Chapitre 6 - Plans d'échantillonnage, le Chapitre 7 - Estimation et le Chapitre 8 - Calcul de la taille de l'échantillon et répartition.

Conception du questionnaire

Questionnaire : séquence de questions formulées pour obtenir d'un répondant de l'information

Le choix entre ces méthodes dépend de : la disponibilité de l'informant, caractéristiques de la pop, nature des questions, les ressources disponibles

méthodes de collecte :

- assistée par intervieweur, autodénombrement ou observation directe,
- si la méthode assistée par intervieweur est appliquée :
 - interview sur place ou téléphonique,
- si la méthode par autodénombrement est appliquée :
 - méthode de distribution et de collecte des questionnaires,
- utilisation de certaines données administratives pour une partie de la collecte des données?
- recours à des substituts à titre de répondants?
- matière du questionnaire,
- formulation des questions, → revue à l'essai
- genres de questions :
 - réponses ouvertes ou fermées, ouvert : Réponses non prédéfinies
 - ordre des questions.
 - fermé : Réponses prédéfinies

utilisation des données collectées dans des fins administratives

Erreurs de Réponses : liées aux réponses des répondants.
Traitement des données :
Problèmes de Saisie

Voir le Chapitre 4 – Méthodes de collecte des données et le Chapitre 5 – Conception du questionnaire.

5. Collecte des données

- comment situer les unités sélectionnées et communiquer avec elles,
- sélection des intervieweurs,
- formation des intervieweurs,
- supervision des intervieweurs,
- contrôle des documents (numéro de repérage des questionnaires complétés, etc.),
- procédures de travail sur place,
- contrôle de la qualité du travail sur place :
 - observation des intervieweurs,
 - nouvelles interviews,
- vérifications sur place,
- suivi des non-réponses.

Examen des questionnaires
Formulation du Quest
Révision du Quest
Revue à l'essai
Discrimination
Complétion
Codage

Voir le Chapitre 9 – Opérations de collecte des données.

6. Saisie et codage des données

- saisie des données,
- codage des données :
 - méthode de classification à appliquer,
- méthodes de mesure, de contrôle et de correction des erreurs :
 - assurance de la qualité,
 - contrôle qualitatif.

Codage : affectation d'une valeur numérique aux réponses pour faciliter la saisie et le traitement des données.

→ coûte cher
→ Ode bcp du temps

Voir le Chapitre 10 – Traitement et l'Annexe B – Contrôle qualitatif et assurance de la qualité.

7. Vérification et imputation

- vérifications à faire sur place,
- vérifications à faire après la collecte sur place (vérifications manuelles et automatisées),
- genre de vérifications à faire :
 - vérifications de la validité,

Vérification : pour repérer les entités manquantes, non valables, par contrôle manuel ou par un prog in formatique

Imputation : processus pour déterminer et attribuer des valeurs de remplacement pour résoudre le problème des valeurs manquantes

- vérifications de l'uniformité,
- uniformité des vérifications,
- méthodes d'imputation à appliquer,
- uniformité de l'imputation,
- préparation et mise à l'essai des systèmes d'imputation et de vérification.

Voir le Chapitre 10 – Traitement.

8. Estimation

Ech. 2 phases de π_1 et π_2 : pondérat° $w_d = \frac{1}{\pi_1} \times \frac{1}{\pi_2}$
 probab. de sélection

- calcul des pondérations du plan d'échantillonnage,
- ajustements possibles pour :
 - le total des non-réponses,
 - les données auxiliaires,
- paramètres à estimer (estimations ponctuelles) :
 - totaux,
 - ratios,
 - proportions, etc.,
- estimateurs pour les estimations ponctuelles,
- estimateurs pour la variance d'échantillonnage des estimations ponctuelles.

$$\pi_1 = \pi_2 \quad \pi_1 = \frac{m_1}{N_1}, \quad \pi_2 = \frac{m_2}{N_2}$$

$N = 1000$ réparties en 2 strates prop

$$N_1 = 400 \quad N_2 = 600 \quad \text{ech: } m = 200$$

$$\frac{m}{N} = \frac{1}{4} \Rightarrow m_1 = \frac{N_1}{4} = 100, \quad m_2 = \frac{N_2}{4} = 100$$

$$m = m_1 + m_2 \quad \hat{m} \text{ prop}$$

2 strates mom prop $N = 1100, N_1 = 1000, N_2 = 100$

$$m = 200, m_1 = 200, m_2 = 0, \pi_1 = \frac{1}{11}, \pi_2 = \frac{1}{11}$$

Proba d'inclus: $\pi = \frac{m}{N}$

Fact d'ajust: $\frac{m}{m_r}$

Poids ajusté pour les non-réponses $\left(\frac{1}{\pi}\right) \frac{m}{m_r}$

Poids de base $\left(\frac{1}{\pi}\right) \frac{m}{m_r}$

Voir le Chapitre 7 – Estimation.

9. Analyse des données et présentation des résultats de l'enquête

- mesures de l'erreur d'échantillonnage,
- mesures de l'erreur non due à l'échantillonnage :
 - erreur de couverture,
 - non-réponses (p. ex., taux de non-réponses),
 - erreur de mesure,
 - erreur de traitement (p. ex., taux de rejets à la vérification),
- méthodes de mesure, de contrôle et de correction des erreurs :
 - assurance de la qualité,
 - contrôle qualitatif,
- évaluation de toutes les opérations de l'enquête,
- type d'analyses à faire,
- totalisation des données,
- rapports à produire,
- méthodes de contrôle de la divulgation,
- suppression ou modification de données.

$$\hat{N} = \sum w_i = \sum \frac{1}{\pi_i}$$

$$\hat{Y} = \sum w_i y_i$$

$$\hat{\bar{Y}} = \frac{\hat{Y}}{\hat{N}} = \frac{\sum y_i}{m}$$

$$\hat{p} = \frac{\hat{N}_c}{\hat{N}} \quad G^2 = \frac{(N-1)S^2}{N}$$

$$S^2 = \sum \frac{(y_i - \bar{Y})^2}{N-1}$$

$$\text{Var}(\hat{\bar{Y}}) = \left(1 - \frac{m}{N}\right) \frac{S^2}{m}$$

Voir le Chapitre 11 – Analyse des données de l'enquête, le Chapitre 12 – Diffusion des données et l'Annexe B – Contrôle qualitatif et assurance de la qualité.

10. Diffusion des données

- utilisateurs et utilisations,
- moyens de diffusion :
 - publication sur support papier,
 - discours ou présentation en public,
 - interview à la radio ou à la télévision,

La diffusion des données : la distribution des données de l'enquête aux utilisateurs par l'intermédiaire des moyens de diffusion.

microfiches,

médias électroniques :

- internet;
- fichier de microdonnées,

méthodes de contrôle de la diffusion.

→ protection des données diffusées : obligée par la loi

Voir le Chapitre 12 – Diffusion des données.

11. Documentation

- auditoire cible :
 - direction,
 - personnel technique,
 - planificateurs d'autres enquêtes,
 - etc.,
- rapport d'enquête,
- rapports sur la méthodologie,
- rapports d'évaluation de la qualité des données,
- manuels de formation (p. ex., pour les interviews),
- rapports de rendement des intervieweurs,
- manuels d'instruction (p. ex., pour les répondants),
- échéancier des activités,
- spécifications pour les programmes des systèmes,
- rapport de faisabilité,
- rapports d'état d'avancement,
- rapport d'enquête (qui documente l'application de toutes les étapes de l'enquête),
- rapport d'analyse des données,
- rapport général ou rapports techniques.

Voir le Chapitre 9 – Opérations de collecte des données pour la documentation des opérations sur place et le Chapitre 12 – Diffusion des données pour la documentation en général.

Liste de vérification des coûts

Il faut tenir compte de certains éléments pour évaluer les coûts de l'enquête, notamment :

- la planification,
- la conception et l'élaboration :
 - le plan d'enquête,
 - les procédures de l'enquête (p. ex., la collecte des données),
 - le traitement après l'enquête,
- l'évaluation de l'enquête,
- la documentation,
- la formation du personnel.

Les coûts du traitement de l'enquête comprendraient le temps du personnel, l'achat ou la location du matériel et des logiciels, et d'autres services, p. ex., les bureaux, les meubles, les articles.

Les critères suivants déterminent la conception et l'élaboration d'une application de l'interview assistée par ordinateur (IAO) :

- la longueur et la complexité des questionnaires,
- le nombre et la complexité des vérifications intégrées,
- le logiciel utilisé,
- les exigences d'entrée préalable de l'information tirée d'un cycle précédent,
- les fonctionnalités nécessaires, par exemple les rapports de gestion, de pistage, etc.,
- le système d'échantillonnage, c.-à-d. les listes téléphoniques ou la composition aléatoire, etc.,
- la trousse de formation à intégrer à l'application,
- le nombre de révisions à apporter aux spécifications,
- la période de référence pour l'élaboration,
- les exigences de la mise à l'essai,
- la taille de l'échantillon (c.-à-d. si des mesures spéciales sont nécessaires à cause du nombre important de données).

Les éléments suivants déterminent le coût de la formation du personnel :

- le genre de formation (en classe, études à domicile, etc.),
- les degrés de formation (qui forme qui),
- la durée (heures, jours),
- l'endroit,
- le nombre d'intervieweurs formés,
- la location de matériel (p. ex., téléphone, ordinateurs, visualiseur d'OP).

Les éléments suivants déterminent le coût de la collecte des données :

- la taille de l'échantillon,
- la répartition de l'échantillon,
- la base d'échantillonnage,
- la durée de l'interview,
- la méthode de collecte des données (sur place, au téléphone, par la poste, etc.) :
 - p. ex., le nombre d'intervieweurs s'il s'agit d'une enquête assistée par intervieweur,
- le taux de réponses prévu,
- la stratégie de suivi,

- population cible (interview de substituts ou non),
exigences de pistage,
échancier de la collecte des données,
- vérifications manuelles et codage,
 - exigences de la saisie des données et de la vérification,
 - location de matériel (téléphone, ordinateurs),
 - location de superficies,
 - coût des déplacements des intervieweurs.