

Le paradigme de l'erreur d'enquête totale

Tiaray RAZAFINDRANOVONA

Séminaire de **Méthodologie Statistique**

22 janvier 2013

Introduction

- Baisse tendancielle des taux de réponse
- Évolutions technologiques
- Pressions sur les coûts

⇒ introduire de nouveaux modes de collecte, collecte multimode ?

Au-delà des questions de coût ...

bien cerner les questions méthodologiques (précision, qualité) !

Plan de la présentation

- **L'erreur d'enquête totale (TSE) : définitions, apports, extensions**
- **La collecte multimode à travers le prisme de la TSE**
- **Une question de recherche importante : l'ampleur des effets de mode ?**

L'erreur totale d'enquête (TSE) : définitions, apports, extensions

- **Définitions**
- **Principales composantes de la TSE**
- **TSE : éléments historiques et apports du concept**
- **Limites et extension : la qualité d'enquête totale**
- **TSE : panorama des questions de recherche**

L'erreur d'enquête totale - *Total Survey Error* (TSE)

- Cadre conceptuel pour optimiser les enquêtes
 - ⇒ en maximisant la précision des données
 - ⇒ sous contraintes (budget, temps)
- TSE : accumulation de toutes les erreurs (plan de sondage, collecte, édition des données).
- Allocation optimale des ressources guidée par la minimisation de la TSE.

Principales composantes de la TSE

- Erreur de couverture
- Erreur d'échantillonnage
- Erreur de non-réponse
- Erreur de mesure, d'observation

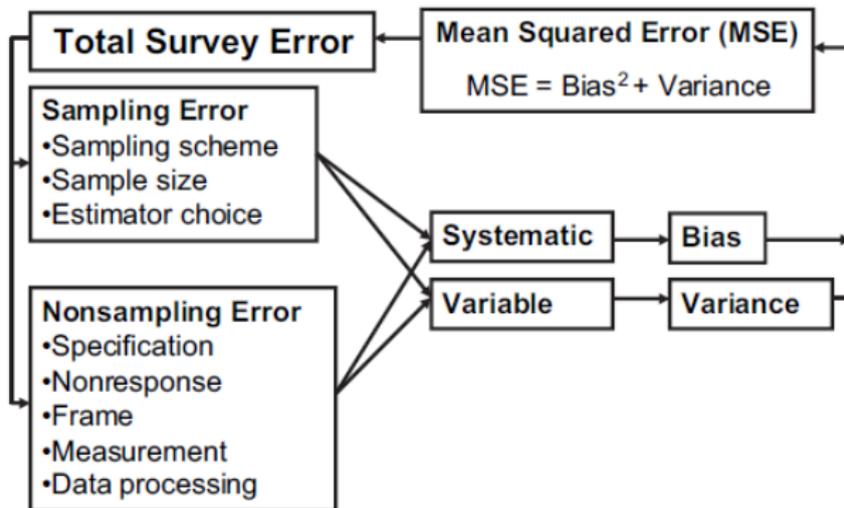
Mais aussi ...

- Erreur de spécification
- Erreur de saisie, codification

Différentes typologies des erreurs

- Erreurs d'échantillonnage vs autres types d'erreurs
- Erreurs d'observation vs autres types d'erreurs
- Erreurs systématiques (biais) vs variables (variance)

TSE, composantes et erreur quadratique moyenne



Biemer (2010)

Principales composantes de la TSE

Erreur de couverture

Erreur liée au fait que la base de sondage ne corresponde pas à la population cible.

Sous-dénombrement : certaines unités de la population d'intérêt ont une probabilité nulle d'être sélectionnées.

Biais de couverture

$$B_{NC} = \frac{N_{NC}}{N} (\bar{Y}_C - \bar{Y}_{NC})$$

avec Y variable d'intérêt, N_{NC} population non-couverte et N population totale.

Erreur d'échantillonnage

Erreur provenant du fait que les estimations reposent sur les individus sélectionnés dans l'échantillon et non sur la population dans son ensemble.

Erreur de non-réponse

Erreur liée à l'absence de réponse de certains individus sélectionnés.

Deux composantes dans cette erreur :

- La proportion de non-répondants
- La différence sur la variable d'intérêt entre répondants et non-répondants

Erreur de non-réponse - version "déterministe"

$$B_{NR} = \frac{N_{NR}}{N} (\bar{Y}_R - \bar{Y}_{NR})$$

avec Y variable d'intérêt, N_{NR} population non-répondante et N population totale.

Erreur de non-réponse - version "stochastique"

$$B_{NR} = \frac{Cov(\rho, Y)}{\bar{\rho}}$$

avec Y variable d'intérêt, ρ propension à répondre.

Erreur de mesure, d'observation

Erreur car l'information collectée sur l'individu diffère de la "vraie" valeur.

Source ? Enquêté, enquêteur, questionnaire, méthode de collecte ...

Désirabilité sociale : tendance des enquêtés à fournir des réponses

- qui donnent une bonne image d'eux-mêmes
- dont ils croient qu'elles se situent dans la norme sociale

Faible implication du répondant (*satisficing*)

- Liée à l'effort cognitif demandé
- Formes faibles : réponse non optimale, acquiescer, effets d'ordre
- Formes fortes : répondre "Ne sait pas", choix aléatoire des réponses

La "vraie" valeur ?

Hansen et al. (1951)

- la vraie valeur doit être définie de manière unique
- la vraie valeur doit être définie de telle manière que les objectifs de l'enquête soient remplis (cf validité de construction en psychométrie)

Deming (1960) : les vraies valeurs n'existent pas mais la valeur obtenue à partir d'une réalisation préférentielle de technique d'enquête peut être utilisée comme proxy.

TSE : éléments historiques

Quelques références

- Deming (1944) liste les principales sources d'erreurs dans une enquête
- Kish (1965) insiste sur les biais, distingue erreurs d'observation des autres
- Anderson, Kasper, Frankel (1979) "*Total Survey Error*"
- Groves (1989) enrichit le concept avec des éléments provenant de la psychométrie, de l'économétrie
- Biemer et Lyberg (2003) intègrent des éléments de qualité des processus

Les principaux textes sur la théorie des sondages insistent sur l'erreur d'échantillonnage.

⇒ l'enquête par sondage pour assurer sa légitimité, a dû se focaliser sur l'analyse de cette composante.

(par ailleurs, l'erreur d'échantillonnage peut être mesurée sur la plupart des échantillons probabilistes mais autres composantes difficiles à mesurer directement)

TSE : apports du concept

Sarndal et Platek (2003)

"Quand le statisticien d'enquête parle de précision, il se réfère surtout à la seule erreur d'échantillonnage."

Un apport important du concept d'erreur d'enquête totale est de distinguer les composantes de l'erreur d'enquête et de ne pas se limiter à la seule erreur d'échantillonnage.

Cadre adapté pour comparer différents protocoles de collecte, envisager la bascule d'un mode à un autre, etc.

La qualité totale d'enquête (TSQ)

Le cadre de l'erreur d'enquête totale permet d'analyser la précision au sens large, mais quid de la qualité ?

Cadre plus général de la qualité d'enquête totale (*Total Survey Quality*) où la TSE ne serait qu'une composante, en tant qu'indicateur de précision.

Principales dimensions de la TSQ

- précision, exactitude (TSE)
- pertinence
- actualité
- accessibilité, intelligibilité
- comparabilité
- cohérence

Optimiser la qualité d'enquête totale nécessite que les objectifs de qualité pour chaque dimension soient bien définis et que les approches pour atteindre ces objectifs soient optimales et budgétées.

Approche de Biemer et Lyberg (2003) : l'objectif est de maximiser la précision des données sous des contraintes de dimensions "utilisateurs"

TSE : panorama des questions de recherche

- Effets enquêteurs
- Évaluer la qualité des réponses
- Protocoles d'enquêtes adaptatifs
- Comparaison des différents modes de collectes
- Arbitrages erreur de couverture / erreur de non-réponse
- Interactions entre erreurs de non-réponse et erreurs de mesure

La collecte multimode à travers le prisme de la TSE

- **Définitions**
- **Pourquoi le multimode ?**
- **Multimode et TSE**
- **Comparaison entre les modes**
- **La collecte par Internet**
- **Revue de littérature**

Multimode : un peu de vocabulaire ...

- Mode : façon de faire
mode de contact, mode de collecte, mode de suivi
- Enquête multimode :
différents canaux de communication intégrés au processus d'enquête
- Collecte multimode :
différents modes intégrés plus précisément dans la phase de collecte

Principaux modes de collecte : face-à-face, téléphone, papier, internet.

Pourquoi le multimode ?

- Faire face à la baisse tendancielle des taux de réponse
- Diminuer les coûts de collecte
- Optimal point de vue TSE ?

Idée : multimode pour compenser les faiblesses des différents modes

Multimode et TSE

Multimode et TSE

Multimode remet au goût du jour TSE et sa décomposition en erreurs ?

⇒ cadre particulièrement adapté pour analyser, comparer différentes possibilités.

Composantes d'erreur et solutions multimodes :

- Réduire erreur de couverture
⇒ combiner mode à faible coût marginal mais ne couvrant pas toute la population d'intérêt avec face-à-face
- Réduire erreur de non-réponse
⇒ mode alternatif offert aux non-répondants
- Réduire erreur de mesure
⇒ mode différent (auto-administré) pour un sous-ensemble de questions sensibles

Comparaison entre les modes

Couverture

Internet : taux d'équipement en France d'environ 75 %

Taux de réponses

Face-à-face > Téléphone > Papier > Internet

Coûts

Face-à-face > Téléphone > Papier > Internet

Erreur de mesure

Dichotomie entre modes auto-administrés et modes avec enquêteurs

La collecte par Internet

Avantages de la collecte par Internet

- Rapidité de collecte
- Coûts
- Interactivité
- Choix du remplissage à discrétion du répondant
- Contenus variés (image, son, vidéo) a priori possibles

Mais ...

- Problèmes de non-réponse, d'auto-sélection
- Problèmes de couverture
- Problèmes techniques à gérer (compatibilité des navigateurs web)

Revue de littérature

- **Taux de réponse et multimode impliquant Internet**

Millar, Dillman (2011) : offrir un choix simultané web/papier n'augmente pas les taux de réponse (comparé au papier seul), en revanche, le multimode séquentiel (web puis papier) donne de meilleurs résultats.

Holmberg, Lorenc, Werner (2010) : pour un multimode séquentiel web puis papier, retarder l'envoi ou la mention du questionnaire papier, a des conséquences limitées sur le taux de réponse global, mais joue énormément sur la proportion de réponses web.

- **Questionnaires et multimode**

de Leeuw et al (2008), Dillman et al (2009), 4 approches possibles : construction unimode, spécifique au mode, spécifique au mode préférentiel, approche généralisée.

- **Qualité des réponses sur Internet**

Heerwegh et Loosveldt (2008) : plus de non-réponses partielles sur Internet qu'en face-à-face, moindre différenciation sur les échelles de réponse.

Heerwegh (2009) : biais de désirabilité sociale en face-à-face moins fort qu'attendu mais plus important que sur Internet.

Une question de recherche importante : l'ampleur des effets de mode ?

- **Introduction**
- **Comparaison naïve**
- **Contrôle par les observables**
- **Expérimentation contrôlée**
- **Autres exemples de protocoles**
- **Quelles recommandations ?**

Effets de mode : différences dans les réponses obtenues selon le mode de collecte

État de la recherche sur le sujet :

Couper (2010)

" Il n'y a pas encore de consensus sur l'étendue des différences d'erreurs de mesure entre les différents modes et sur ce (si besoin est) qu'il est nécessaire de faire à ce propos "

Comparaison naïve

Considérons une variable d'intérêt Y et 2 types de questionnaires collectés :

- des questionnaires recueillis via le mode de collecte α
- des questionnaires recueillis via le mode de collecte β

⇒ comparer alors \bar{Y}_α et \bar{Y}_β ne permet pas de conclure quant aux effets de mode (liés à l'erreur de mesure), car ce ne sont pas forcément les mêmes types de personnes qui répondent (problèmes de sélection voire de couverture).

⇒ le souci avec ce genre d'analyse est qu'on ne distingue pas l'erreur de mesure des erreurs de couverture ou de sélection.

Contrôle par les observables

On essaie de rendre les deux groupes de répondants "comparables" en contrôlant par des caractéristiques observables Z . On peut utiliser, par exemple :

- les modèles de régression
- les pondérations par score de propension (en modélisant $P(M = \alpha|Z)$)
- les méthodes de matching

Hypothèse forte sous-jacente : la sélection des individus s'explique entièrement par les caractéristiques observables.

Problème : les individus qui répondent (par exemple sur Internet) gardent un profil spécifique que des variables de contrôle ne parviennent pas toujours à corriger.

Expérimentation contrôlée

Pour évaluer les effets de mode, une solution est l'expérimentation contrôlée : affecter aléatoirement un mode de collecte, une fois que l'individu fait le choix de répondre à l'enquête.

Problème : difficile de contraindre quelqu'un à répondre !

Un exemple d'expérimentation en laboratoire, Chang et Kroskik (2010) :

- Deux modes de réponse : auto-administré sur ordinateur ou entretien (interphone)
- Affectation aléatoire des modes de réponse
- Résultats : effets d'ordre des réponses (*recency*) plus marqués en entretien, réponses moins socialement désirables sur ordinateur

Validité externe ? Les participants sont des étudiants en psychologie.

Autres exemples de protocoles

Vannieuwenhuyze, Loosveldt et Molenberghs (2010) : comparent une collecte multimode (2 modes) avec une collecte faisant intervenir un seul mode pour bien distinguer les effets sur la mesure, des effets de sélection.

Buelens, van der Laan, Schouten et al. (2012) : leur expérimentation fait intervenir 2 vagues, une première où différents modes sont assignés aléatoirement (web, papier, téléphone et face-à-face) et une deuxième vague où tout le monde répond avec le même mode (face-à-face).

Quelles recommandations ?

Ampleur des effets de mode

Pas encore vraiment de consensus sur la manière de mesurer les effets de mode et la nature de leur ampleur.

Résultats difficilement généralisables car dépendent du questionnaire et des thèmes abordés.

⇒ nécessité d'expérimenter dès lors que l'on souhaite modifier en profondeur un protocole d'enquête et introduire un nouveau mode de collecte.

Corrections, ajustements

Peut-on agréger les réponses quand elles proviennent de modes différents ?

Là aussi, pas de consensus sur ce qu'il est nécessaire de faire ou pas concernant les effets de mode.

Conclusion

- TSE (et TSQ) : cadres généraux adaptés pour évaluer les nouveaux protocoles d'enquête, l'introduction de nouveaux modes de collecte.
- La pertinence de la collecte multimode devrait s'analyser dans ce cadre.
- En pratique, la mesure des composantes de la TSE est complexe, généralisation pas toujours possible.

⇒ **Programme d'expérimentations Internet et multimode à l'Insee**

Bibliographie (TSE)

- ANDERSON, KASPER, FRANKEL (1979) *Total Survey Error : Applications to Improve Health Surveys*
- BIEMER (2010) *Total Survey Error : Design, Implementation and Evaluation Public Opinion Quarterly*
- BIEMER, LYBERG (2003) *Introduction to Survey Quality*
- DEMING (1944) *On Errors in Surveys American Sociological Review*
- DEMING (1960) *Sample Design in Business Research*
- GROVES (1989) *Survey Errors and Survey Costs*
- GROVES, LYBERG (2010) *Total Survey Error : Past, Present and Future Public Opinion Quarterly*
- HANSEN, MORRIS, HURWITZ, MARKS, MAULDIN (1951) *Response Errors in Surveys Journal of the American Statistical Association*
- KISH (1965) *Survey Sampling*
- LYBERG (2012) *La qualité des enquêtes Techniques d'enquête*
- PLATEK, SARNDAL (2001) *Can a Statistician Deliver ? Journal of Official Statistics*

Bibliographie (Internet et multimode)

- CHANG, KROSNICK (2010) Comparing Oral Interviewing with Self-Administered Computerized Questionnaires *Public Opinion Quarterly*
- COUPER (2011) The Future of Modes of Data Collection *Public Opinion Quarterly*
- DILLMAN, SMYTH, CHRISTIAN (2009) *Internet, Mail and Mixed-Mode Surveys*
- de LEEUW, HOX, DILLMAN (2008) *International Handbook of Survey Methodology*
- BIEMER (2010) Total Survey Error : Design, Implementation and Evaluation *Public Opinion Quarterly*
- BUELENS, VAN DER LAAN, SCHOUTEN, VAN DER BRAKEL, BURGER, KLAUSCH (2012) Disentangling Mode-Specific Selection and Measurement Bias in Social Surveys *Statistics Netherlands Discussion Papers*
- FRIPPIAT, MARQUIS (2010) Les enquêtes par Internet en sciences sociales *Population*
- HEERWEGH (2009) Mode Differences between Face-to-face and Web Surveys *International Journal of Public Opinion Research*
- HEERWEGH, LOOSVELDT (2008) Face-to-face versus Web Surveying in a High-Internet-Coverage Population *Public Opinion Quarterly*

Bibliographie (Internet et multimode)

HOLMBERG, LORENC, WERNER (2010) Contact Strategies to Improve Participation via the Web in a Mixed-Mode Mail and Web Survey *Journal of Official Statistics*

MILLAR, DILLMAN (2011) Improving Response To Web and Mixed-Mode Surveys *Public Opinion Quarterly*

VANNIEUWENHUYZE, LOOSVELDT, MOLENBERGHS (2010) A Method for Evaluating Mode Effects in Mixed-Mode Surveys *Public Opinion Quarterly*