

Exploitation et analyse des données appliquées aux techniques d'enquête par sondage.



Introduction.



Etudes et traitements statistiques des données :

le cas illustratif de la démarche par sondage

INTRODUCTION : « *Que cherche t'on à mesurer ?* »

- La nécessité d'une phase exploratoire, pourquoi ?
- La nécessité d'une phase exploratoire, comment ?



Phase 1 : poser les bonnes questions



Phase 2 : identifier les moyens nécessaires à l'étude



Phase 3 : situer les types d'objectifs de l'étude pour exploiter les bons outils



Analyse des données

appliquée aux techniques d'enquête par sondage

INTRODUCTION : « Que cherche t'on à mesurer ? »

- La nécessité d'une phase exploratoire, comment ?

Phase 1 : se poser les bonnes questions (suite)

- Qui sont les acteurs ?
 - Le chargé d'étude et les autres acteurs
 - Le, (les) demandeur(s)
 - Le, (les) décideur(s)
 - Les relais
 - Les opposants



INTRODUCTION : « Que cherche t'on à mesurer ? »

- La nécessité d'une phase exploratoire, comment ?

Phase 2 : identifier les moyens nécessaires à l'étude

- moyens matériels
- moyens financiers
- moyens permettant l'accès aux personnes et aux sources d'informations.
- Moyens en temps
- Moyens d'approche du terrain : la maîtrise d'un réseau d'enquêteurs



INTRODUCTION : « Que cherche t'on à mesurer ? »

- **La nécessité d'une phase exploratoire, comment ?**

Phase 2 (suite): formaliser le cahier des charges

- **Les informations à obtenir**
- **Les sources : quelle base de données ou quelle population interroger**
- **Les protocoles : comment s'y prendre pour collecter les données**
- **Les délais et les résultats attendus**

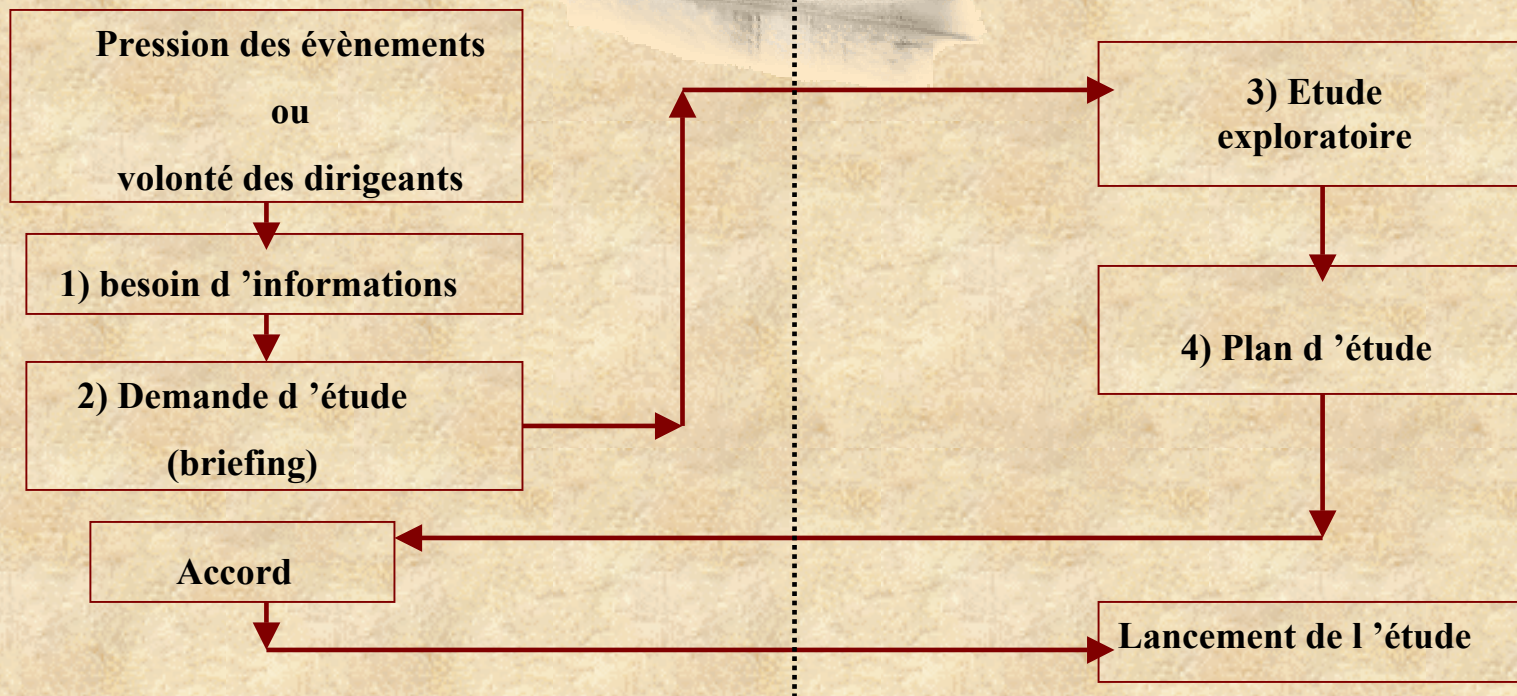


INTRODUCTION : « Que cherche t'on à mesurer ? »

- La nécessité d'une phase exploratoire, comment ? (suite)

Entreprise

Cabinet ou chargé d'études



INTRODUCTION : « Que cherche t'on à mesurer ? »

- La nécessité d'une phase exploratoire, comment ?(suite)

Phase 3 : situer les types d'objectifs de l'étude pour exploiter les bons outils

OBJECTIFS	Contenu	Etudes	Exemples d'études
Décrire	une situation à un moment donné (exemple le climat social d'une entreprise à une date précise)	Documentaire Qualitative Quantitative	SI Observations, entretiens Sondages
	suivre l'évolution d'une tendance	Documentaire Quantitative	SI panels, baromètres
Expliquer	définir des relations entre des variables explicatives et des faits	Qualitative Quantitative	Etudes de motivation Sondages Corrélation, discrimination, tests de dépendance
Prévoir	définir des situations à venir	Quantitative	Marchés tests, régression
Proposer	apporter des solutions alternatives	Quantitatif	Outils aide à décision
Vérifier	face à un problème, contrôler le bien-fondé d'une hypothèse ou d'une action	Qualitatif Qualitatif	Créativité (recherche de nom, par exemple) Tests de produits, par exemple Post-tests



INTRODUCTION : « Que cherche t'on à mesurer ? »

- **La nécessité d'une phase exploratoire, comment ?(suite)**

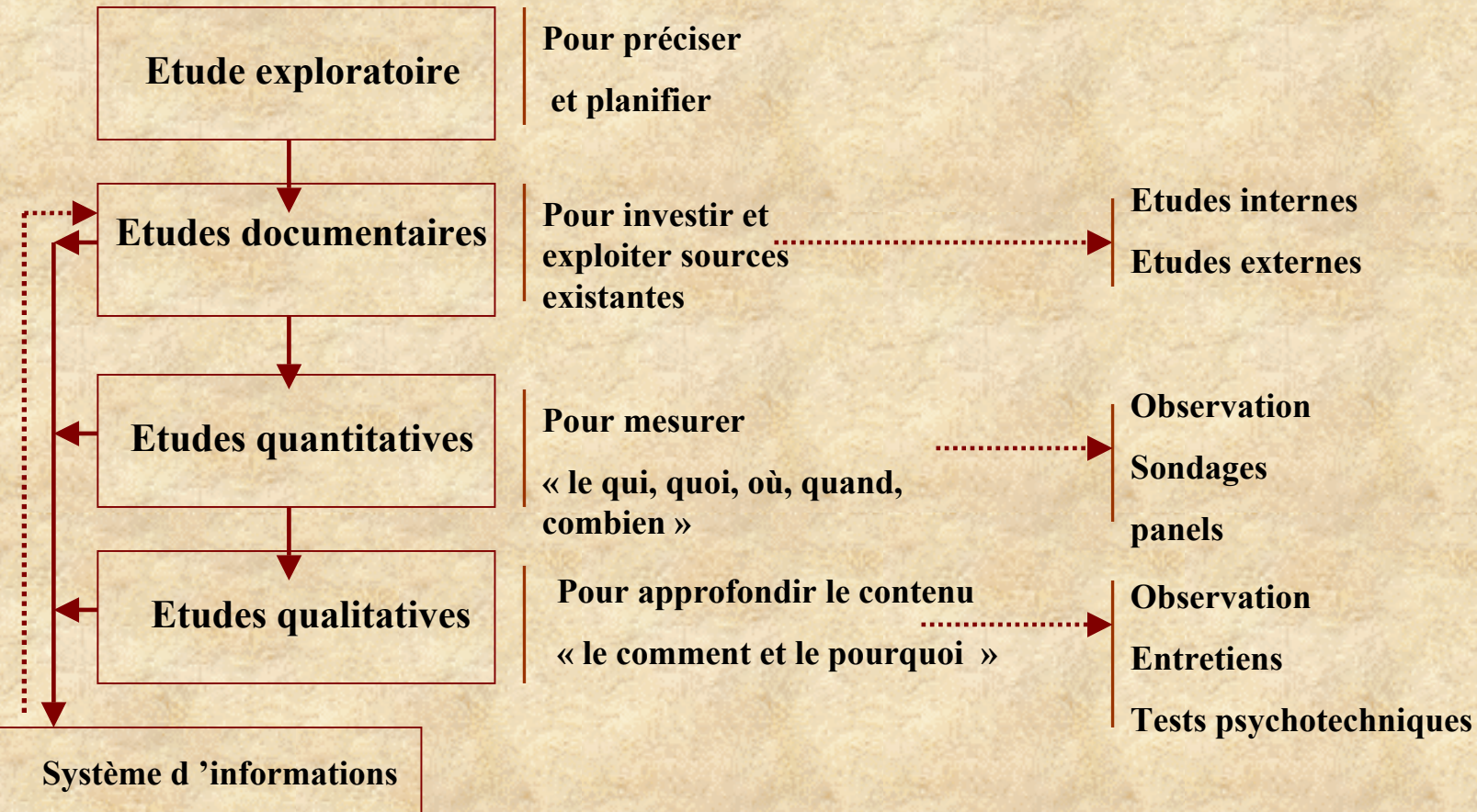
Phase 3 : situer les types d'objectifs de l'étude pour exploiter les bons outils

- **La réflexion « théorique et les modèles » : *constituer un groupe de travail.***
- **La recherche documentaire : *exploiter les connaissances et informations disponibles.***
- **La valorisation du système d'information existant : *c'est toujours moins coûteux.***
- **Produire des informations nouvelles : *uniquement si celles-ci n'existent pas sous une autre forme.***



INTRODUCTION : « Comment garantir la précision et la fiabilité des résultats ? »

Planifier rigoureusement sa démarche d'étude et choisir ses outils :



INTRODUCTION : « Comment garantir la précision et la fiabilité des résultats ? »



Respecter les étapes clés d'une enquête par sondage :



INTRODUCTION : « Comment déterminer la précision et la fiabilité des résultats ? »

Exploiter une analyse inférentielle et descriptive des données :

I.

- L'analyse inférentielle

pour

- Permettre de généraliser à la population globale les résultats obtenus sur l'échantillon.

- Mesurer le degré de précision de cette généralisation.



- L'analyse descriptive

par

- L'analyse univariée des données.

- L'analyse bivariée des données.

- L'analyse multivariée des données.

II.



.I. Garantir la précision et l'objectivité des résultats d'une enquête :



L'échantillonnage



.I. LA THEORIE DES SONDAGES ET L 'ECHANTILLONNAGE

1.1 La notion de théorie des sondages



1.2 Quelle méthode d'échantillonnage utiliser ?

1.3 Comment calculer la taille optimale d 'un échantillon ?

1.4 Erreurs d 'échantillonnage et redressements d 'échantillon

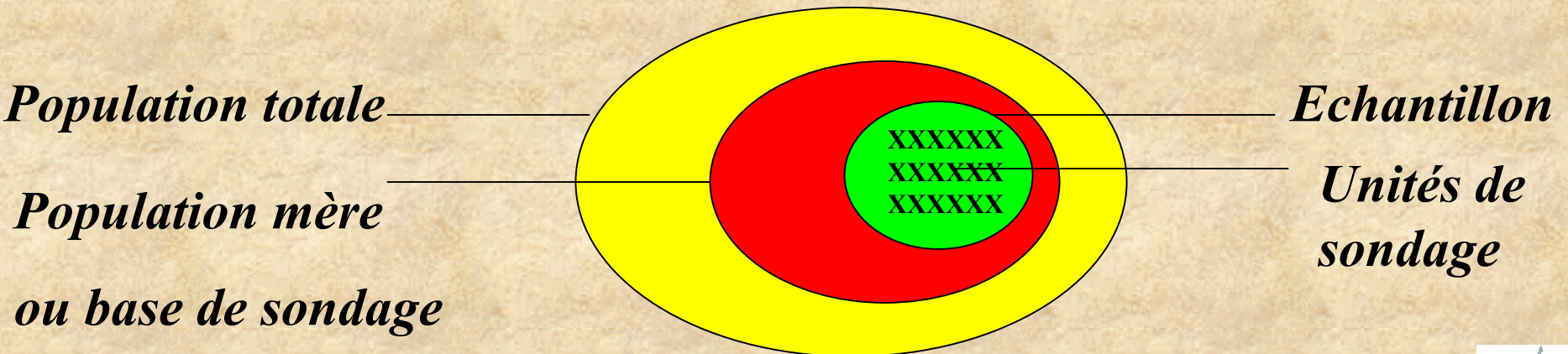


1.1 La notion de théorie des sondages

➤ Pourquoi travailler à partir d'un échantillon ?

- Contrainte de taille
- Contrainte de délais
- Logique statistique

➤ Vocabulaire propre à l'échantillonnage



1.2 Quelle méthode d'échantillonnage utiliser ?

- Par nécessité et, ou par volonté se reposer sur le hasard :

1.21 La constitution d'un échantillon aléatoire

- *par nécessité* : on ignore tout de la population mère
- *par volonté* : lois statistiques d'équiprobabilité et précision

- Par nécessité et, ou simplicité établir un échantillon à partir de critères raisonnés :

1.22 La constitution d'un échantillon non probabiliste

- *par nécessité* : contraintes coûts - délais
- *par volonté* : connaissance population mère permettant constitution échantillon représentatif



1.21 La constitution d'un échantillon aléatoire

➤ Quand utiliser le sondage aléatoire simple ?

© quand la population possède des caractéristiques homogènes

© quand on dispose d'une base de sondage (fichier)

➤ Comment utiliser le sondage aléatoire simple ?

© tirage aléatoire

© tables de nombres au hasard



1.21 La constitution d'un échantillon aléatoire

EXTRAITS D'UNE TABLE DE NOMBRES AU HASARD

(Kendall et Babington Smith, table tirée de Christian Labrousse, Statistique, Tome2, Dunod, Paris, 1962)

02 22 85 19 48 74 55 24 89 69 15 53 00 20 88 48 95 08
85 76 34 51 40 44 62 93 65 99 72 64 09 34 01 13 09 74
00 88 96 79 38 24 77 00 70 91 47 43 43 82 71 67 49 90
64 29 81 85 50 47 36 50 91 19 09 15 98 75 60 58 33 15
94 03 80 04 21 49 54 91 77 85 00 45 68 23 12 94 23 44
42 28 52 73 06 41 37 47 47 31 52 99 89 82 22 81 86 55
09 27 52 72 49 11 30 93 33 29 54 17 54 48 47 42 04 79
54 68 64 07 85 32 05 96 54 79 57 43 96 97 30 72 12 19
25 04 92 29 71 11 64 10 42 23 23 67 01 19 20 58 35 93
28 58 32 91 95 28 42 36 98 59 66 32 15 51 46 63 57 10
64 35 04 62 24 87 44 85 45 68 41 66 19 17 13 09 63 37
61 05 55 88 25 01 15 77 12 90 69 34 36 93 52 39 36 23
98 93 18 93 86 98 99 04 75 28 30 05 12 09 57 35 90 15
61 89 35 47 16 32 20 16 78 52 82 37 26 33 67 42 11 93
94 40 82 18 06 61 54 67 03 66 76 82 90 31 71 90 39 27
54 38 58 65 27 70 93 57 59 00 63 56 18 79 85 52 21 03
63 70 89 23 76 46 97 70 00 62 15 35 97 42 47 54 60 60
61 58 65 62 81 29 69 71 95 53 53 69 20 95 66 60 50 70
51 68 98 15 05 64 43 32 74 07 44 63 52 38 67 59 56 69
59 25 41 48 64 79 62 26 87 86 94 30 43 54 26 98 61 38
85 00 02 24 67 85 88 10 34 01 54 53 23 77 33 11 19 68
01 46 87 56 19 19 19 43 70 25 24 29 48 22 44 81 35 40
42 41 25 10 87 27 77 28 05 90 73 03 95 46 88 82 25 02
03 57 14 03 17 80 47 85 94 49 89 55 10 37 19 50 20 37
18 95 93 40 45 43 04 56 17 03 34 54 83 91 69 02 90 72



1.21 La constitution d'un échantillon aléatoire

➤ Quand utiliser le sondage systématique ?

◎ quand on dispose d'une base de sondage (fichier)

◎ pour enquêtes postales, téléphoniques

➤ Comment utiliser le sondage systématique ?

◎ tirage aléatoire

◎ à partir d'une règle systématique
et d'une fréquence d'interrogation



1.21 La constitution d'un échantillon aléatoire

➤ Quand utiliser le sondage stratifié ?

- ◎ quand on dispose d'une base de sondage (fichier)
- ◎ quand la population est caractérisée par différentes strates aux caractéristiques homogènes
- ◎ quand la taille de chaque strate est relativement importante par rapport à la taille de la population totale (chaque strate devenant une nouvelle population à étudier indépendamment)
- ◎ pour réduire la dispersion dans l'échantillon

➤ Comment utiliser le sondage stratifié ?

- ◎ pour segmenter la population mère
- ◎ tirage aléatoire dans chaque strate



1.21 La constitution d'un échantillon aléatoire

⇒ Quand utiliser le sondage à plusieurs degrés ?

- Quand le territoire d'enquête est géographiquement large
- Quand on ne dispose pas de base de sondage

(simple et économique mais moins précis que l'échantillon aléatoire simple)

⇒ Comment utiliser le sondage à plusieurs degrés ?

- par tirages au sort successifs : *la population de référence étant composée de sous-ensembles appelés grappes ou unitaires primaires de sondage*



1.21 La constitution d'un échantillon aléatoire

➤ Quand utiliser le sondage en grappes ?

- pas de base de sondage nécessaire
- pour mesure des effets de groupes (enquêtes psychosociologiques)
- quand les unités de chaque grappe sont représentatives

➤ Comment utiliser le sondage en grappes ?

- Choix aléatoires de sous-groupes ou grappes dans la population totale
- interrogation des unités de chaque grappe



1.22 La constitution d 'un échantillon non probabiliste

⇒ Pourquoi utiliser le sondage par quotas ?

- l'échantillon est une maquette à échelle réduite de la population étudiée
- méthode économique

⇒ Quand utiliser le sondage par quotas ?

- bonne connaissance de la population mère permettant la constitution d 'un échantillon représentatif sur certains critères clefs (variables de contrôle)

⇒ Comment utiliser le sondage par quotas ?

- détermination des variables de contrôle
- constitution des quotas de manière descendante
- administration de l 'enquête à l 'aide d'une grille de suivi



1.22 La constitution d'un échantillon non probabiliste

⇒ Quand utiliser le sondage par itinéraire ?

- mesure d'effets de groupe
- enquêtes à optique géo-marketing
- souci de vérification de la qualité de la collecte

⇒ Comment utiliser le sondage par itinéraire ?

- par itinéraire type



1.3 Comment calculer la taille optimale d'un échantillon ?

1.31 Notion de taille optimale



- définition
- taux de sondage
- exhaustivité et représentativité

1.32 Précision et intervalle de confiance :

$$n = t^2 pq/e^2$$

- e = marge d'erreur qui donne la précision recherchée ou intervalle de confiance (en général, 3 ou 4% pour les études marketing)
- p = degré d'homogénéité de la population et q = degré de non homogénéité de la population ($p = 1-q$); p est souvent connu à partir d'une étude antérieure
- le seuil de probabilité retenu qui détermine la valeur de t (*en général $t=1,96$ cad 95% de fiabilité de l'échantillon*)



1.3 Comment calculer la taille optimale d'un échantillon ?

Les valeurs de t selon les seuils de probabilité :

- **t = 1.65** quand le seuil de probabilité retenu est de 90%
- **t = 1.96** quand le seuil de probabilité retenu est de 95%
- **t = 2.58** quand le seuil de probabilité retenu est de 99%

Donc plus le seuil de probabilité cad de certitude est élevé, plus la taille de l'échantillon est importante.

Rque : la taille d'un échantillon aléatoire peut être déterminée également à partir de lecture d'une table des intervalles de confiance ou d'abaques



1.3 Comment calculer la taille optimale d'un échantillon ?

Application : une entreprise souhaite évaluer l'évolution du degré de satisfaction des clients en terme de service après achat; une étude précédente démontrait que 80% des clients se déclareraient satisfaits du service après achat.

1) Selon les marges d'erreur et les seuils de probabilité suivants, calculez la taille de l'échantillon à interroger – Que constatez-vous ?

Marge d'erreur	1%	2%	4%	8%
Seuil de probabilité				
90%				
95%				
99%				



1.3 Comment calculer la taille optimale d'un échantillon ?

1) correction

Marge d'erreur	1%	2%	4%	8%
Seuil de probabilité				
90%	4536	1082	272	68
95%	6144	1536	384	96
99%	11664	2916	729	182

- plus la marge d'erreur se réduit, plus la taille de la population à interroger est importante
- plus le seuil de probabilité est élevé, plus l'échantillon doit avoir une taille importante pour obtenir un niveau d'erreur donné
- pour diminuer l'erreur par d on doit augmenter la taille de l'échantillon de d^2 ; *exemple : pour diminuer l'erreur de 8% à 2% cad par 4, on doit augmenter la taille de l'échantillon par 16 cad 4^2*
- donc une marge d'erreur qui dépend uniquement de la taille de l'échantillon et en aucun cas de la taille de la population mère



1.3 Comment calculer la taille optimale d'un échantillon ?

2) On suppose qu'il s'agit d'une première étude et donc p est inconnu (par prudence, on considère alors que la population n'est pas homogène cad $p=q=0,5$). Dans l'hypothèse énoncée, calculez la taille de l'échantillon à interroger

Marge d'erreur	1%	2%	4%	8%
Seuil de probabilité 95%				

3) En supposant un coût moyen de 15 euros par enquête, au seuil de probabilité de 95%, calculez l'évolution du coût de l'étude en fonction de la marge d'erreur

Marge d'erreur	1%	2%	4%	8%
Coûts en euros				



1.3 Comment calculer la taille optimale d'un échantillon ?

Correction

2)

Marge d'erreur	1%	2%	4%	8%
Seuil de probabilité 95%	9 604	2 401	600	150

3)

Marge d'erreur	1%	2%	4%	8%
Coûts en euros	144 060	36 015	9 000	2 250



1.4 Erreurs d'échantillonnage et redressement d'échantillon

1.41 Les erreurs d'échantillonnage :

- **erreur liée à la précision du % obtenu selon l'importance de l'effectif de l'échantillon**
- **une marge d'erreur qui varie en fonction du % trouvé**
- **une erreur directement liée à l'échantillonnage**
- **l'erreur dite de liste**
- **l'erreur de non réponse ou d'inadéquation**
- **les erreurs liées aux outils de mesure**



1.4 Erreurs d'échantillonnage et redressement d'échantillon

1.42 Le redressement de l'échantillon :

- pour rectifier le non respect des quotas imposés

Coefficient de redressement = échantillon préconisé/échantillon retenu

<i>ancienneté client</i>	<i>.-1an</i>	<i>1 à 2 ans</i>	<i>+.2 ans</i>	TOTAL
Echantillon préconisé	22	34	17	73
Echantillon obtenu	18	34	21	73
Coefficient de redressement	1,22	1	0,81	

Si à partir de cet échantillon, on observe que, par exemple, 25% des clients de moins d'un an effectuent + 5 achats par mois, alors il faudra redresser ce résultat : $25\% \times 1.22$ soit 30.5%



1.4 Erreurs d'échantillonnage et redressement d'échantillon

1.42 Le redressement de l'échantillon (suite) :

Sexe	% obtenus	Effectifs obtenus
Féminin	61	1220
Masculin	37	740
Non réponse	2	40
		2000

% à obtenir
55% (1100)
45% (900)

⇒ par extraction :

Élimination de toutes les unités au delà du % à obtenir (cad que les 740 hommes doivent représenter 45% de l'échantillon soit éliminer 316 femmes et les non réponses)

Sexe	Effectif corrigé
Féminin	904
Masculin	740
Non réponse	
	1644

Sexe	Effectif corrigé
Féminin	1100
Masculin	900
Non réponse	
	2000

⇒ par substitution

On remplace toute unité en surnombre d'une modalité sur-représentée par une unité choisie au hasard dans une modalité sous représentée (qui est donc dupliquée) cad « un + d'un côté compensé par un moins de l'autre »

