4. ÉCHANTILLONNAGE

4.1 DEUX HISTOIRES DE SONDAGE

4.2 MÉTHODOLOGIE SIMPLIFIEE

4.3 EXEMPLES

Remarque introductive

Le chapitre qui suit présente de façon très pragmatique les principes et méthodes usuels d'échantillonnage.

Une seule formule doit être retenue (sans démonstration), le reste des calculs nécessitant de maitriser les opérations arithmétiques classiques, de savoir calculer des pourcentages, de manipuler des « règles de 3 » et ...

de réfléchir .

4.1 DEUX HISTOIRES DE SONDAGE

1. Histoire chinoise

2. Histoire nord-américaine

1. HISTOIRE CHINOISE

Dans une (grande) boîte :

900 millions de papiers portant les noms, prénoms
et adresses des chinois de plus de 18 ans

et leur opinion sur la peine de mort :

POUR OU CONTRE

1. HISTOIRE CHINOISE

La boîte est brassée (longuement).

On tire au hasard 100 papiers (N= 100)

On compte 62 « POUR » : (P = 62 en %)

• Combien seriez vous disposés à parier qu'**une majorité** des chinois adultes est pour la peine de mort?

1€ ? 100€ ? 1000€ ?

1. HISTOIRE CHINOISE

La boîte est brassée (longuement).

On tire au hasard 100 papiers (N= 100)

On compte 62 « POUR » : (P = 62 %)

- Meilleure estimation du % de « POUR »

  sur les 900 millions de chinois adultes (population d'origine) ?

- Marge d'erreur sur cette estimation ?

1. HISTOIRE CHINOISE

✍ **Il y a 95 % de probabilité** que,
sur les 900 millions de chinois adultes,
le % de « POUR »
soit compris entre : (62 - E) et (62 + E)

ou $$\boxed{E ? 2\sqrt{\frac{PxQ}{N}}}$$ = **« Marge d'erreur »**

avec P = 62, Q = 100 – 62 = 38, N = 100 (taille de l'échantillon!)

NB : Cette formule est applicable si N > 30

✍ E = … ?

INTERVALLE DE CONFIANCE = (62 – E , 62 + E) = (… ? , … ?)

et donc : avez-vous gagné votre pari ?

1. HISTOIRE CHINOISE
ABAQUE

N	P (%)	E (%)	INTERVALLE (%)
30	62	17,0	(45,0 ; 79,0)
100	62	?	(? ; ?)
500	62	?	(? ; ?)
1000	62	3,1	(58,9 ; 65,1)
2000	62	2,2	(59,8 ; 64,2)

1. HISTOIRE CHINOISE

ABAQUE (suite)

Pour N = 1000

P en %	10%	20%	30%	40%	50%	60%	70%	80%	90%
Marge	**1,9%**	**2,5%**	**2,9%**	**3,1%**	**3,2%**	**?**	**?**	**?**	**?**

Conclusion : marge d’erreur maximale pour P = … %

1. HISTOIRE CHINOISE

ABAQUE (suite)

Pour N = 1000

P en %	10%	20%	30%	40%	50%	60%	70%	80%	90%
Marge	**1,9%**	**2,5%**	**2,9%**	**3,1%**	**3,2%**	**3,1%**	**2,9%**	**2,5%**	**1,9%**

⇩

marge d'erreur maximale pour P = 50 %

✍ Pour déterminer la taille d'un échantillon assurant une marge d'erreur donnée : dans la mesure où on ignore P, on applique la formule du cours, en prenant P=50 qui donne une borne supérieure à la marge!

A RETENIR :

Un sondage sur 1000 personnes, en Chine comme à Monaco assure une marge d'erreur d'environ 3 % (en + ou en -)

Nécessité d'être rigoureux dans l'information

En octobre 2000 :

France Soir **: « Al gore : 44% , Bush :41%, …**

A devance B dans les sondages »

France Info **: « Al gore : 44% , Bush : 41%, …**
avec une marge d'erreur de + ou – 3%,
cela s'appelle un coude-à-coude »

Exercice : a) sur quel effectif de répondants, le sondage a-t-il été réalisé?
b) justifiez le commentaire de France Info par un graphique.

Nécessité d’être rigoureux dans l’information

En octobre 2006 :

Le Figaro : « Au second tour Ségolène Royal l’emporte sur Nicolas Sarkozy avec 51% des voix contre 49% »

✍ **Manque de professionnalisme de certains journalistes**

2. HISTOIRE NORD-AMERICAINE

1936 ÉLECTION DU PRÉSIDENT DES ÉTATS-UNIS

✍Vote « à blanc » des lecteurs du « Literary Digest » (retour coupon) + abonnés au téléphone (téléphone)+ possesseurs de voitures (postal)

2 400 000 réponses ✍ **défaite de Roosevelt**

✍Sondage sttistique de Georges Gallup

4 000 réponses ✍ **victoire de Roosevelt**

✍Conclusion ?

2. HISTOIRE NORD-AMERICAINE

1936 ÉLECTION DU PRÉSIDENT DES ÉTATS-UNIS

Exercice : précisez le biais de la première méthode de sondage :

……

……

……

……

4.2 MÉTHODOLOGIE SIMPLIFIEE

1. Choix d’une méthode (aléatoire / par quotas)

2. Détermination de la taille de l’échantillon

3. Stratification

1. CHOIX D'UNE MÉTHODE

A) IL EXISTE UNE « BASE DE SONDAGES » FIABLE
(c.à.d. une liste ou un fichier contenant l'identification de la totalité des individus ou unités de la population – univers total- dont on souhaite extraire un échantillon)

✍ **ÉCHANTILLON « PROBABILISTE » OU ALÉATOIRE**
par tirage au sort dans la base

Exemples de bases de sondages :

……

……

……

……

1. CHOIX D'UNE MÉTHODE

B) IL N 'EXISTE PAS DE «BASE DE SONDAGES» FIABLE

✍ MÉTHODE DES QUOTAS

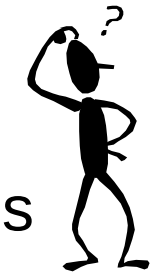
Quotas = petit nombre de variables

- aux statistiques connues sur l'univers total
- corrélées avec les informations recherchées

✍ On s'impose de respecter, sur l'échantillon,

les répartitions des quotas (connues sur l'univers total)

QUOTAS: exemple concret



• Un pays compte 6 millions d ’habitants dont 5 millions d’adultes (âgés de 18 ans et plus), sont répartis en 4 régions.
Un recensement récent fournit les statistiques suivantes :

Nord :	1,2 million	d’adultes,
Est :	0,8 million	«
Ouest :	1 million	«
Sud :	2 millions	«

Enquête par quotas sur 1000 personnes d’adultes avec
2 quotas (non croisés) : 1) Homme , Femme
et 2) 18/34 ans , 35/54 ans , 55 ans et plus

Question : **quelles consignes donner aux enquêteurs, devant chacun interroger 20 personnes en face à face, concernant le sexe et l’âge de ces personnes?**

QUOTAS suite

✍ Nombre d ’enquêteurs par région ?

Région	Nombre d’enquêtés	Nombre d’enquêteurs
Nord		
Est		
Ouest		
Sud		

QUOTAS suite

Nombre d ’enquêteurs par région :

Région	Nombre d’enquêtés*	Nombre d’enquêteurs
Nord	**240**	**12**
Est	**160**	**8**
Ouest	**200**	**10**
Sud	**400**	**20**

** On respecte la proportion des nombres d’adultes par région.*

QUOTAS suite
nombres d'adultes en milliers
(issus du dernier recensement)

	F	H	18/34 ans	35/54 ans	55 ans et +
Nord	620	580	320	380	500
Est	**420**	**380**	**300**	**270**	**230**
Ouest	500	500	290	310	400
Sud	1050	950	490	640	870
Total	**2590**	**2410**	**1400**	**1600**	**2000**

QUOTAS suite

Il existe un très grand nombre de solutions donnant des consignes à chaque enquêteur permettant de respecter les quotas totaux d'âge et de sexe dans chaque région.

Suit une répartition possible sur la région Est donnée pour faire comprendre le mode de calcul.
Cette répartition s'efforce de donner des consignes équilibrées, c.à.d. consistant, par exemple, à'interroger autant de femmes que d'hommes pour un maximum d'enquêteurs, afin d'éviter des «biais de grappe ».

QUOTAS suite

Consignes données pour les 160 enquêtes de la région EST :

Au total :

✍ **Même proportion H/F que dans les statistiques du recensement :**

F : 160 x (420/800) = 84 enquêtes

H : 160 x (380/800) = 76 enquêtes

✍

15/34 ans : ?

34/54 ans : ?

55 ans et + : ?

QUOTAS suite

✍ **Consignes données aux 8 enquêteurs E1, E2, …, E8 pour les 160 enquêtes de la région EST ?**

Enquêteur	F	H	18/34 ans	35/54 ans	55 ans et +
E1	?	?	?	?	?
E2	?	?	?	?	?
---	---	---	---	---	---
E6	?	?	?	?	?
E7	?	?	?	?	?
E8	?	?	?	?	?
Total	**84**	**76**	**60**	**54**	**46**

QUOTAS suite
Exemples de consignes « équilibrées » données aux 8 enquêteurs de la région EST :

Enquêteur	F	H	18/34 ans	35/54 ans	55 ans et +
E1	10	10	7	7	6
E2	10	10	7	7	6
---	---	---	---	---	---
E6	10	10	7	7	6
E7	10	10	9	6	5
E8	14	6	9	6	5
Total	**84**	**76**	**60**	**54**	**46**

2. DÉTERMINATION DE LA TAILLE D'UN ÉCHANTILLON (N)

1) Précision statistique souhaitée
   sur variable(s) fondamentale(s)
   ✍ N

2) Coût d'un entretien x N = budget acceptable ?

3) Ajustement de N pour obtenir le meilleur compromis
   budget acceptable / précision acceptable

EXEMPLE 1

• **Sondage téléphonique** grand public, « sans filtre », « court » (10 minutes)

✍ Coût de l'entretien = 30 €

1) Précision souhaitée : 1%

✍ N = 4000 (en appliquant la formule du cours avec P=50%)

2) Coût total = 4000 x 30 € = 120 000 €

✍ **inacceptable**

3) Meilleur compromis précision / coût = ? ...

EXEMPLE 1

Meilleur compromis ?

N = 1000 ✍ E ~ 3,2 % et coût = 30 000 €

Ou

N = 2000 ✍ E ~ 2,3 % et coût = 60 000 €

EXEMPLE 2

• **Sondage** « face à face », sur 1 000 personnes, grand public, pour mesurer un % de notoriété de marque

1) Marge d'erreur sur le total de l'échantillon = … ?

2) Marge d'erreur sur les hommes(500) = … ?
et sur les femmes (500) = … ?

3) Marge d'erreur sur les hommes
dans chacune de 5 régions (d'effectif 100 chacune) = … ?

4) Variation de coût pour obtenir une marge d'erreur divisée par 2,
si coût unitaire du questionnaire = 40 €

✍ Coût multiplié par … ?

EXEMPLE 2

• **Marges d’erreur :**

1) sur le total de l’échantillon **~ 3,2 %**

2) sur les hommes(500) **~ 4,5%** et les femmes (500) **~ 4,5 %**

3) sur les hommes dans chacune de 5 régions (d’effectif 100) **~ 10 %**

---

4) Coût multiplié par 4 , soit, sur la totalité de l’échantillon :

1000 x 40 € ✍ 4000 x 40 €

3. STRATIFICATION
EXEMPLE

100 000 clients entreprises
dont 5% de grandes entreprises (GE)
et 95% de PME

Enquête de satisfaction auprès de 400 clients

Sondage aléatoire ?

EXEMPLE STRATIFICATION suite

Solution 1 : Sondage aléatoire simple
Nombre d'entreprises interrogées (sur 400)

GE = …… ?

PME = …… ?

Marge d'erreur sur les GE = … ?

EXEMPLE STRATIFICATION suite

Solution 1 : Sondage aléatoire

Clientes	**Nombre d'entreprises interrogées (sur 400)**
GE = **5 000**	✍ **20**
PME = **95 000**	✍ **380**

Marge d'erreur sur les GE **~ 23%** , inacceptable

Attention : limites d'application de la formule ! N doit être > 30

Solution 2 : sondage aléatoire stratifié

=> On fixe a priori un effectif pour chaque strate (PME et GE),

avec un minimum de 100, si possible

EXEMPLE suite

Strate	Effectif univers	Effectif échantillon	Marge sur un %
GE	5 000	100	= … ?
PME	95 000	300	= … ?
Total	100 000	400	= … ?

EXEMPLE 3 suite : estimation d'un % de satisfaction global

Strate	% sur échantillon	% sur univers	Calculs
GE	60%	= … ?	… ?
PME	50%	= … ?	… ?
Total			= … ?

EXEMPLE 3 suite : estimation d'un % de satisfaction global

Strate	% sur échantillon	% sur univers	Calculs
GE	60%	= 60%	**60% x 5**
PME	50%	= 50%	**+ 50% x 95**
Total			**/ 100** **= 50,5%**

STRATIFICATION

- Découpage a priori de la population étudiée en classes disjointes (strates) homogènes relativement au phénomène étudié.

- Adaptation de la méthodologie d'échantillonnage à chaque strate.

- « Redressement » des résultats pour estimation de variables sur la population totale.

NOTION DE REDRESSEMENT

Sur un échantillon de 1 000 personnes, on interroge,

pour des raisons « techniques », 600 femmes et 400 hommes

alors que la proportion est de 50 % / 50 % dans l'univers étudié

PONDÉRATION

Si, sur ce même échantillon, on constate :
65 % de « pour » chez les femmes
55 % de « pour » chez les hommes

Quelle est l'estimation du % de « pour » sur l'ensemble
de la population de référence ?

NOTION DE REDRESSEMENT

Solution 1 :
Moyenne pondérée =
(65% x 500/1000) + (55% x 500/1000) = 60%

Alors que la moyenne non pondérée donne
(65% x 600/1000) + (55% x 400/1000) = 61%

Solution 2 :

Moyenne non pondérée avec

poids des femmes : 500/600 = 0,833

poids des hommes :500/400 = 1,250

= (65% x 0,833 x 600/1000) + (55% x 1,25 x 400/1000) = 60%

(Utilisée dans la pratique quand redressement sur plusieurs critères)

Exercice : Identifiez des sites d'instituts décrivant des méthodologies d'études

site	Objectif de l'étude	Méthode utilisée (taille et mode d'échantillonnage, type de recueil)
……	……	……
……	……	……
……	……	……