***Le troisième chapitre : Choisir l’échantillon***

***Concepts généraux***

***Population mère*** : l’ensemble des éléments sur lequel porte l’étude.

La population mère est définie par les caractéristiques des individus qui les rendent aptes à participer à l’enquête.

***L’échantillon*** : sous ensemble d’éléments d’une population donnée.

***Le sondage*** : sonder (ou échantillonner) c’est choisir une partie (l’échantillon) pour représenter le tout (la population mère).

***Technique d’échantillonnage***: l’ensemble des opérations permettant de sélectionner un sous-ensemble d’une population en vue de constituer un échantillon.

***L’échantillon représentatif***

Pour s’assurer de la bonne qualité dans le choix de l’échantillon, trois éléments sont à prendre en compte.

1)- Pour être représentatif, un échantillon doit être obtenu par une méthode rigoureuse qui permet de choisir des enquêtés sans biais. Cela revient à assurer à toute unité de la population la possibilité de figurer dans l’échantillon.

2)- Même obtenu par une méthode rigoureuse de choix. Un échantillon sera plus ou moins précis selon sa taille. Pour améliorer la précision du sondage, on augmente donc la taille de l’échantillon.

3)- L’échantillon doit représenter exactement les mêmes caractéristiques que sa population mère.

***Echantillonnage probabiliste***

L’échantillonnage probabiliste entraine la sélection d’un échantillon à partir d’une population, sélection qui repose sur le principe de la randomisation (la sélection au hasard ou aléatoire) ou la chance.

La méthode qu’on choisira dépendra d’un certain nombre de facteurs, comme la base de sondage dont on disposera, la façon dont la population sera distribuée, ce que sonder les membres de la population coutera et la façon également dont les utilisateurs analyseront les données.

Lorsque vous choisirez un plan d’échantillonnage probabiliste, votre but devrait consister à réduire le plus possible l’erreur d’échantillonnage des estimations pour les variables d’enquête les plus importantes, tout en réduisant le plus possible également le délai et le cout de réalisation de l’enquête.

Un échantillon aléatoire avec base de sondage se compare à une loterie dans laquelle tous les numéros ont une chance égale d’être pigés.

***Échantillon aléatoire simple***

Dans un échantillon aléatoire simple, chaque membre d’une population a une chance égale d’être inclus à l’intérieur de l’échantillon.

Donc la première opération consiste à établir la liste exhaustive des individus de la population.

Une fois la base de sondage dressé, les éléments en sont numérotés de 1 à N, puis le tirage de n numéros est réalisé (si l’échantillon doit comporter n éléments) les sujets correspondant aux n numéros « sortis » constitueront l’échantillon, il faut ensuite aller interroger nommément les personnes désignées par le sort.

Comment sélectionner les numéros ? La méthode de la loterie peut être utilisée : par exemple une main innocente tire successivement dans un chapeau n fiches nominatives parmi l’ensemble des fiches convenablement mélangées. Il est également possible d’utiliser une table de nombres au hasard ou des nombres générés par ordinateur.

***Echantillon aléatoire systématique***

Parfois appelé échantillonnage par intervalles, l’échantillonnage systématique signifie qu’il existe un écart, ou un intervalle, entre chaque unités sélectionnée, qui est incluse dans l’échantillon. Vous devez suivre les étapes énumérées ci-dessous pour sélectionner un échantillon systématique :

1)- numéroté de 1 à N les unités incluses dans votre base de sondage (ou N est la taille de la population totale).

2)- déterminer l’intervalle d’échantillonnage (K) en divisant le nombre d’unités incluses dans la population par la taille de l’échantillon que vous désirez obtenir. Par exemple, pour sélectionner un échantillon de 200 unités à partir d’une population de 800, vous auriez besoin d’un intervalle d’échantillonnage de 800/ 200 = 4, K = 4.

3)- sélectionner au hasard un nombre entre 1et K. Ce nombre s’appelle l’origine choisie au hasard et serait le premier nombre inclus dans votre échantillon. A l’aide de l’échantillon fourni ci-dessus, vous sélectionneriez un chiffre entre 1 et 4 pris au hasard. Si vous choisissez 2, la deuxième unité incluse dans votre base de sondage serait la première unité comprise dans votre échantillon.

4)- sélectionner chaque Ke (dans ce cas, chaque 4e) unité après ce premier nombre. L’échantillon pourrait, par exemple, se composer des unités suivantes de façon à constituer un échantillon de 200 : 2 (l’origine choisie au hasard), 6, 10, 14, …. , 794, 798.

***Echantillonnage stratifié***

Lorsqu’on utilise l’échantillonnage stratifié, on divise la population en groupe homogène (appelés strates), qui sont mutuellement exclusifs, puis on sélectionne à partir de chaque strate des échantillons indépendants.

Quel critère de stratification choisir ? Il est nécessaire de choisir un caractère dont la valeur est connue pour chaque sujet avant l’enquête. Par exemple une population d’étudiants interrogés sur ses projets d’avenir pourra être stratifiée selon la filière d’étude principale, donnée par les listes d’inscription à l’université… Il faut aussi que le critère de stratification soit en relation avec l’objet de l’étude.

***Exercice d’application***

Une entreprise contient 82 cadres, 360 Agents de maitrise et 720 Agents d’exécution. On veut choisir un échantillon de 120 unités.

On opte pour un échantillonnage stratifié, on calcule d’abord la somme de la population mère de cette entreprise

82+360+720= 1162.

Le pourcentage de chaque strate par rapport à la taille de la population mère ;

Pour les cadres

1162 → 100 %

82 → X %

 = $\frac{82x100\%}{1162}$ = 7.05 %

Pour les agents de maitrise

1162 → 100 %

360 → X %

= $\frac{360x100\%}{1162}$ = 30.98 %

Pour les agents d’exécution

1162 → 100 %

720 → X %

= $\frac{720x100\%}{1162}$ = 61.96 %

On choisit un échantillon de 120 unités divisées comme suit :

120 → 100 %

Les cadres représentent 7,05 %

Donc, le nombre de cadres dans notre échantillon = $\frac{7.05\% x 120}{100\%}$ = 8.46 ≈ 9.

Les agents de maitrise représentent 30.98 %

Donc, le nombre d’agents de maitrise dans notre échantillon

= $\frac{30.98\% x 120}{100\%}$ = 37.17 ≈ 37.

Les agents d’exécution représentent 61.96%.

Donc, le nombre d’agents d’exécution dans notre échantillon

= $\frac{61.96\% x 120}{100\%}$ = 74.35 ≈ 74.

Pour finir, on doit choisir 9 unités parmi 82 cadres de façon aléatoire, 37 unités parmi 360 agents de maitrise et 74 unités parmi 720 agents d’exécution.

***L’échantillonnage non probabiliste***

On emploie l’échantillonnage non aléatoire (ou non probabiliste) quand on ne dispose pas d’une base de sondage. Dans un tel cas, on ne connait pas la probabilité qu’un élément de la population fasse partie de l’échantillon. En fait, sans base de sondage, on ne connait pas non plus la taille exacte de la population à étudier, même si on est capable de définir cette population avec précision. Cela signifie qu’on doit choisir l’échantillon de façon arbitraire, en prenant des décisions qui entrainent un risque d’erreur plus difficile à évaluer que dans le cas où on procède de façon aléatoire. Ce type d’échantillonnage est considéré comme non scientifique, ou moins valable sur le plan scientifique, même si on doit souvent l’utiliser pour des raisons pratiques.

***L’échantillonnage accidentel***

L’échantillonnage accidentel (ou à l’aveuglette) est le plus rudimentaire. Il consiste à prélever un échantillon de façon totalement arbitraire, sans préparation ni planification. C’est ainsi qu’on procède lors de sondages dans la rue… Ce type d’échantillonnage donne des résultats acceptables si la population est très homogène, mais le risque d’erreur est élevé.

L’échantillonnage accidentel est utile dans la cadre d’une recherche exploratoire précédant une recherche plus approfondie, ou pour connaitre l’ensemble des opinions sur une question donnée.

***L’échantillon typique (ou par choix raisonné ou intentionnel)***

Cette technique se fonde sur un choix raisonné fait par le chercheur. Ce dernier veut orienter sa recherche sur un type de phénomène ou d’individus qui se distinguent des autres selon certaines caractéristiques.

L’utilisation de cette technique se justifie donc par la pertinence des choix raisonnés qui la sous-tendent. Ainsi, l’échantillon retenu sera aussi pertinent que l’est le choix raisonné du chercheur. Cette technique est très utilisée par des adeptes des méthodes qualitatives qui recherchent moins la représentativité que l’exemplarité de leur échantillon.

***L’échantillonnage en boule de neige***

Dans certaines recherches, trouver des participants n’est pas facile. Mener une enquête auprès de revendeurs de drogue, de prostitués ou d’ex-toxicomanes n’a rien à voir avec mener une enquête auprès de collègues étudiants. L’échantillonnage en boule de neige mise sur le succès du premier échange : le chercheur établit un lien de confiance avec un informateur privilégié qui le mettra en contact avec d’autres personnes de son groupe d’appartenance. De personne en personne enquêtée, le chercheur peut parvenir à ratisser assez large. Si le chercheur est capable, sans jamais mentir aux répondants présentés, de s’intégrer au groupe étudié, cette technique peut donner accès à des personnes qui en diront plus qu’on aurait pu le penser sur le fonctionnement social de leur organisation.