

Chapitre3 :

Les méthodes d'échantillonnage

Pour effectuer une étude statistique (enquête, observation ou expérimentation), on se sert généralement d'un échantillon. Celui-ci doit refléter le plus exactement possible l'image de la population. En fait, choisir un échantillon, c'est mettre ensemble un certain nombre d'individus qui composeront une sorte de modèle réduit de la population à laquelle ils appartiennent. Mais comment choisir ces individus? C'est à cette question que nous répondrons dans ce chapitre, en présentant les méthodes d'échantillonnage les plus courantes.

I- Les méthodes aléatoires:

Commençons d'abord par les méthodes dites aléatoires. Le mot "aléatoire" vient du latin aléa, "jeu de dés". Il est employé ici pour signifier que les méthodes abordées ont toutes comme caractéristique d'être basées sur le hasard.

1/ L'échantillonnage aléatoire simple:

L'échantillonnage aléatoire simple consiste à choisir des individus de telle sorte que chaque membre de la population ait une égale chance de figurer dans l'échantillon. Ce choix peut se faire avec ou sans remise: avec remise (on dit aussi non exhaustif), un individu peut être choisi plusieurs fois; sans remise (on dit aussi exhaustif), un individu déjà choisi ne peut l'être de nouveau. L'échantillonnage aléatoire simple est habituellement fait sans remise. En effet, l'échantillon est très souvent de petite taille par rapport à celle de la population, et la possibilité qu'un même individu soit choisi plus d'une fois est alors faible. En outre, l'échantillonnage sans remise est plus facile sur le plan opérationnel.

Pour recourir à l'échantillonnage aléatoire simple, il faut disposer d'une liste à jour de tous les individus d'une population déterminée. On choisit alors l'échantillon à l'aide de nombres aléatoires: un nombre aléatoire est un nombre dont le hasard a déterminé chacun des chiffres qui le composent. Pour y parvenir, on utilise soit une table spécialement conçue à cette fin, soit une calculatrice, soit un programme informatique. L'exemple suivant aidera à comprendre comment procéder.

Exemple :

On veut choisir par échantillonnage aléatoire simple, sans remise, 8 étudiants; parmi un groupe de 60. La sélection de l'échantillon sera faite à l'aide de la table de nombres aléatoires qui se trouve à la page suivante.

Table des nombres aléatoires

92033	80696	28539	82033	85696	33539	46017	09513	27344	20174
48885	18295	35380	38885	23295	40380	24443	11793	12962	04574
04774	56239	41710	50226	61239	46710	02387	13903	16742	14060
60385	11803	81477	50385	16803	86477	30193	27159	16795	02951
56702	16943	35824	46702	21943	40824	28351	11941	15567	04236
32744	07171	46142	22744	12171	51142	16372	15381	07581	01793
63084	35161	89169	53084	40161	94169	31542	29723	17695	08790
96750	08697	97546	86750	13697	12546	48375	32515	28917	02174
99809	57531	13973	89809	62531	18973	49905	04658	29936	14383
48837	32253	97156	38837	37253	10156	24419	32385	12946	08063
46347	90015	62242	36347	95015	67242	23174	20747	12116	22504
65793	54819	89871	55793	59819	94871	32897	29957	18598	13705
40585	83547	35551	30585	88547	40551	20293	11850	10195	20887
48546	02045	19225	38546	07045	24225	24273	06408	12849	00511
62661	07873	00215	52661	12873	05215	31331	01072	17554	01968
35762	94389	86960	25762	99389	91960	17881	28987	08587	23597
35940	85953	54717	25940	90953	59717	17970	18239	08647	21488
96063	43229	00984	86063	48229	05984	48032	00328	28688	10807
29650	55866	60952	19650	60866	65952	14825	20317	06550	13967
85129	68938	62005	75129	73938	67005	42565	20668	25043	17235
50109	88952	74335	40109	93952	79335	25055	24778	13370	22238
37708	49907	91188	27708	54907	96188	18854	30396	09236	12477
72040	49876	02777	62040	54876	07777	36020	00926	20680	12469
06335	19548	02172	36650	24548	07172	03168	00724	12217	04887
11859	66327	22991	01859	71327	27991	05930	07664	00620	16582
54718	01251	55991	44718	06251	60991	27359	18664	14906	00313
38055	63671	12730	28055	68671	17730	19028	04243	09352	15918
79352	16177	95394	69352	21177	10394	39676	31798	23117	04044
94669	73141	26022	84669	78141	31022	47335	08674	28223	18285
20501	93684	87259	10501	98684	92259	10251	29086	03500	23421
41980	06793	31724	31980	11793	36724	20990	10575	10660	01698
37273	92154	85129	27273	97154	90129	18637	28376	09091	23039
58854	67852	50109	48854	72852	55109	29427	16703	16285	16963
79359	05765	37708	69359	10765	42708	39680	12569	23120	01441
57074	10788	72040	47074	15788	77040	28537	24013	15691	02697
47399	14441	55866	37399	19441	60866	23700	18622	12466	03610
81615	88425	68938	71615	93425	73938	40808	22979	23872	22106
11168	61067	88952	01168	66067	93952	05584	29651	00389	15267
87140	24350	72362	77140	29350	77362	43570	24121	25713	06088
80039	77313	17951	70039	82313	22951	40020	05984	23346	19328
12567	84092	27273	02567	89092	32273	06284	09091	00856	21023
51646	92661	84092	41646	97661	89092	25823	28031	13882	23165
72362	04775	92661	62362	09775	97661	36181	30887	20787	01194
17951	55486	37273	07951	60486	42273	08976	12424	02650	13872
47518	78723	72362	37518	83723	77362	23759	24121	12506	19681
70070	07152	17951	60070	12152	22951	35035	05984	20023	01788
73556	43198	07273	63556	48198	12273	36778	02424	21185	10800

- La première étape consiste à numéroter chaque étudiant à l'aide de deux chiffres afin que chacun des étudiants ait la même chance d'être choisi: 01, 02, 03..., 59, 60.

- La deuxième étape consiste à choisir un processus de déplacement dans la table et à déterminer au hasard un premier nombre de deux chiffres (car il y a deux chiffres dans 60) qui sera le point de départ. L'échantillon sera alors constitué des huit premiers nombres de deux chiffres n'excédant pas 60, tout nombre apparaissant une deuxième fois étant éliminé.

Supposons que l'on ait choisi comme point de départ de la table les deux premiers chiffres à l'intersection de la 5ème ligne et de la 3ème colonne. Supposons aussi que l'on ait décidé d'aller de haut en bas jusqu'à la fin de la colonne en ne prenant toujours que les deux premiers chiffres de chaque nombre, puis de reprendre le même processus en haut de la colonne suivante.

L'échantillon se composerait alors des étudiants portant les numéros suivants: 35,46,13,19,54,60,02 et 22.

En résumé, L'échantillonnage aléatoire simple consiste à choisir n unités statistiques parmi N unités d'une population dont on possède la liste.

1. On numérote tous les individus de la liste avec des nombres comportant un même nombre de chiffres.
2. En utilisant une table de nombres aléatoires, une calculatrice ou un programme informatique, on obtient des nombres aléatoires comportant le nombre de chiffres désiré.
3. En rejetant les nombres qui ne font pas partie de la liste ou qui se répètent, on s'arrête après avoir sélectionné n individus.

2/ L'échantillonnage systématique:

L'échantillonnage systématique est une méthode demandant moins de manipulations que l'échantillonnage aléatoire simple. Cependant, elle suppose aussi l'existence d'une liste de la population où chaque individu est numéroté de 1 jusqu'à N . Notons n , le nombre d'individus que doit comporter l'échantillon. L'entier voisin de N/n sera noté r et appelé la raison de sondage (ou le pas de sondage). Choisissons ensuite au hasard un entier d entre 1 et N : cet entier sera le point de départ. Pour former l'échantillon, il s'agira de choisir un premier individu comme point de départ; ce sera l'individu dont le numéro correspond à d . Pour sélectionner (les autres, il suffit d'ajouter à d la raison de sondage: les individus choisis seront alors les individus dont les numéros correspondent à: d ; $d + r$; $d + 2r$; $d + 3r$; etc.

Il faudra reprendre au début lorsque la liste sera épuisée.

Exemple : On veut choisir par échantillonnage systématique 6 étudiants parmi un groupe de 60. La raison de sondage (r) sera 10, car $N/n = 10$. (Si l'on devait trouver un échantillon de taille (n) 7, la raison serait plutôt 9 puisque $60/7 = 8,57$, arrondi à 9.) Si l'origine choisie au hasard est disons, 3, les étudiants inclus dans l'échantillon correspondront aux numéros suivants:

3, c'est-à-dire le point de départ ; 13, c'est-à-dire $3 + 10$;
23, c'est-à-dire $3 + (2 \times 10)$ 33, c'est-à-dire $3 + (3 \times 10)$ 43, c'est-à-dire $3 + (4 \times 10)$ 53, c'est-à-dire $3 + (5 \times 10)$

Notre échantillon est maintenant complet; il sera composé des étudiants portant les numéros suivants: 3, 13,23, 33, 43, 53.

Selon la raison de sondage et le point de départ choisi, il peut arriver que le nombre obtenu se situe à l'extérieur de la liste. En pareil cas, il faut revenir au début de la liste. Si L'on reprenait l'exemple précédent et si L'on devait continuer à prélever des étudiants, on obtiendrait le numéro 63, c'est-à-dire $3 + (6 \times 10)$. Comme ce numéro r se situe à l'extérieur de la liste, on reviendrait au début de la liste et on choisirait le numéro 3 (puisque $63 - 60 = 3$).

En résumé, L'échantillonnage systématique consiste à choisir n unités statistiques parmi N unités d'une population dont on possède la liste.

1. On numérote tous les individus de la liste.
2. On calcule la raison de sondage (r), c'est-à-dire l'entier le plus proche de N/n .
3. On choisit au hasard un entier d entre 1 et N.
4. A partir de d, on ajoute autant de fois r que cela est nécessaire pour sélectionner n unités statistiques.

3/ L'échantillonnage stratifié:

Contrairement aux deux méthodes précédentes, dans L'échantillonnage stratifié, on tient compte des renseignements qu'on pourrait déjà posséder sur la population, renseignements obtenus en particulier lors d'un recensement. La méthode repose en effet sur une division de la population en groupes relativement homogènes, appelés strates, puis sur la sélection d'un échantillon dans chaque strate. C'est une méthode qui permettra d'obtenir un échantillon représentatif, c'est-à-dire un échantillon qui possédera les mêmes caractéristiques que la population dont il a été extrait. Illustrons le processus à L'aide d'un exemple.

Exemple :

On veut choisir par échantillonnage stratifié 10 élèves dans un groupe de 60, en tenant compte du fait que 50 % d'entre eux sont en première année, 30 % en 2e année et 20 % en 3e année. Chaque année sera une strate dans laquelle on ira chercher des élèves en tenant compte des pourcentages qu'occupe chaque strate dans la ' population. Ainsi, on choisira au hasard:

- 5 élèves en 1ère année, puisque $10 \times 50\% = 5$;
- 3 élèves en 2ème année, puisque $10 \times 30\% = 3$;
- 2 élèves en 3ème année, puisque $10 \times 20\% = 2$.

Il ne reste plus qu'à sélectionner un échantillon dans chaque strate, ce qui pourrait être fait par échantillonnage aléatoire simple ou systématique.

Dans une population peu homogène, le découpage en strates sera d'autant plus avantageux qu'existera une certaine homogénéité à L'intérieur de chaque strate.

4/ L'échantillonnage par grappes:

Dans chacune des méthodes précédentes, les unités de l'échantillon étaient choisies individuellement. L'échantillonnage par grappes consiste plutôt à choisir plusieurs individus en même temps, c'est-à-dire par groupes. Par exemple, prenons comme population les habitants d'une ville à partir desquels on désire constituer un échantillon de 600 individus. Selon les méthodes précédentes, il faudrait choisir 600 individus disséminés dans toute la ville. Suivant l'échantillonnage par grappes, on pourra choisir les 600 individus dans une vingtaine d'immeubles choisis au hasard, où tous les occupants auront été retenus.

Exemple :

Au moyen de L'échantillonnage par grappes, il s'agit de choisir 12 étudiants; dans un groupe de 60. On demande aux étudiants de se regrouper par 6. On choisit ensuite au hasard deux regroupements, par exemple les grappes numéros 4 et 7. En retenant tous les individus de ces deux grappes, on constitue un ; échantillon de 12 étudiants.

Dans l'exemple précédent, la situation était assez simple. Dans les faits, les choses sont: beaucoup plus compliquées. Par exemple, dans une ville, les quartiers et les immeubles ne sont pas composés d'un nombre égal d'individus. Comment peut-on procéder alors pour effectuer un échantillonnage par grappes? Puisque le nombre d'habitants et le nombre d'immeubles et de logements de chacun des quartiers sont généralement connus, il est possible de quadriller la ville en un grand nombre de secteurs (qu'on appelle îlots) ayant des populations à peu près semblables. Pour constituer un échantillon, il suffira alors de choisir certains de ces secteurs et, par conséquent, toutes les personnes qui y habitent.

En résumé, L'échantillonnage par grappes consiste à choisir n unités statistiques parmi N unités d'une population.

1. On subdivise la population en grappes, autant que possible nombreuses et de taille équivalente. Par exemple, si L'on doit répartir 307 individus dans 10 grappes, on pourra faire 7 grappes de 31 personnes et 3 grappes de 30 personnes.
2. On calcule combien il faut de grappes pour constituer L'échantillon.
3. On choisit au hasard les grappes qui serviront à construire L'échantillon.
4. On sélectionne tous les individus des grappes choisies.

II-Les méthodes non aléatoires :

Les méthodes de la section précédente sont dites aléatoires. Le processus de sélection est alors basé sur le hasard, et les individus faisant partie de la population ont tous une égale possibilité de faire partie de L'échantillon. On oppose aux méthodes aléatoires les méthodes dites non aléatoires: ce sont des méthodes où le concept de chances égales est absent. En utilisant des méthodes non aléatoires, il faut toujours craindre un manque de fiabilité. Un des seuls moyens de mesurer l'exactitude de leurs résultats est souvent de les contrôler par d'autres études. Doit-on alors penser que les méthodes non aléatoires sont rarement employées? Pour différentes raisons, on ne peut souvent pas faire autrement que d'y recourir. On s'en sert pour des études

exploratoires; pour réduire les coûts; parce que le recours à une méthode aléatoire n'est ni possible ni même envisageable; ou encore lorsque l'homogénéité de la population est telle qu'il serait inutile d'utiliser des méthodes aléatoires.

1/ L'échantillonnage à l'aveuglette:

La personne qui déguste des échantillons de vin pour déterminer lequel est le meilleur ou le journaliste qui, au moyen d'entrevues dans (a rue, sonde L'opinion du grand public sur un sujet d'actualité, mettent en pratique la méthode de L'échantillonnage à l'aveuglette. Le grand avantage de cette méthode est qu'elle demande peu de planification. D'application restreinte, elle peut malgré tout donner de bons résultats si la population observée est homogène. Par exemple, si L'on désire évaluer la concentration d'un produit chimique dans un lac ou le taux de glycémie d'une personne, il est raisonnable de supposer que les résultats devraient être assez semblables d'un échantillon à L'autre.

2/ L'échantillonnage de volontaires:

Dans le cas d'expériences psychologiques ou médicales, d'enquêtes sur les habitudes de consommation, il ne serait pas pratique de choisir au hasard des individus dans toute la population. Comme l'enquête sera longue, exigeante, quelquefois même désagréable, on préfère réunir des volontaires, d'où le nom d'échantillonnage de volontaires. Néanmoins, il faut toujours craindre un écart entre les caractéristiques des volontaires et celles de la population, surtout en matière d'opinion. Les maisons de sondage font souvent appel à un échantillon fixe, appelé panel, formé de volontaires. Il faut savoir que cette façon de procéder peut être une source importante d'erreurs même si on essaie de l'améliorer en se servant d'un échantillon d'assez grande taille. Par exemple, on a souvent noté la tendance à être systématiquement favorables ou neutres sur certaines d'actualité, alors que les opinions de la population sur les mêmes questions étaient beaucoup plus diversifiées.

3/ L'échantillonnage par quotas:

L'échantillonnage par quotas est largement utilisé dans les enquêtes d'opinion et les études de marché, notamment lorsqu'on ne dispose pas de liste exhaustive des individus de la population. La méthode porte également le nom d'échantillonnage dirigé ou par choix raisonné. On parle aussi d'échantillonnage représentatif. On demande en effet aux enquêteurs de faire un nombre déterminé d'entrevues dans divers groupes établis en fonction du secteur géographique, de l'âge, du sexe ou d'autres caractéristiques. Les échantillons de départ peuvent être obtenus par sélection au hasard, par exemple par téléphone, tout comme pour un échantillon aléatoire. La différence essentielle réside toutefois dans la sélection des individus à l'étape ultérieure: dans la méthode des quotas, chaque enquêteur doit suivre des instructions quant au nombre de personnes à interroger suivant le sexe, l'âge, etc.; il doit écarter certains individus, en choisir d'autres, et ce jusqu'à l'obtention de ses quotas. Ceux-ci ont été calculés à partir de données disponibles, de manière que les sexes, les groupes d'âge et les catégories socio-économiques soient correctement représentés dans l'échantillon. L'échantillon devra avoir à peu près la même composition que la population totale étudiée. On se base sur certains critères basique comme le sexe, l'âge, la catégorie socioprofessionnelle, la région, la ville, le niveau d'études... L'hypothèse est que l'échantillon étant représentatif du point de vue des critères retenus, il y a de grandes chances pour qu'il le soit également concernant les caractéristiques de l'enquête.

La méthode des quotas est généralement moins coûteuse et plus facile à réaliser que celle d'un échantillonnage aléatoire.