

## Échantillonnage systématique

Parfois appelé *échantillonnage par intervalles*, l'*échantillonnage systématique* (SYS) signifie qu'il existe un écart, ou un intervalle, entre chaque unité 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 (où **N** est la taille de la population totale).

1. 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 100 unités à partir d'une population de 400, vous auriez besoin d'un intervalle d'échantillonnage de  $400 \div 100 = 4$ .  $K = 4$ , par conséquent. Vous devrez sélectionner une unité sur 4 pour avoir finalement au total 100 unités à l'intérieur de votre échantillon.
2. Sélectionner au hasard un nombre entre 1 et **K**. Ce nombre s'appelle l'origine choisie au hasard et serait le premier nombre inclus dans votre échantillon. À l'aide de l'échantillon fourni ci-dessus, vous sélectionneriez un chiffre entre 1 et 4 à partir d'une table de nombres aléatoires (pris au hasard). Si vous choisissiez 3, la troisième unité incluse dans votre base de sondage serait la première unité comprise dans votre échantillon; si vous choisissiez 2, le début de votre échantillon serait la deuxième unité incluse dans votre base de sondage.
3. Sélectionner chaque **K<sup>e</sup>** (dans ce cas, chaque 4<sup>e</sup>) unité après ce premier nombre. L'échantillon pourrait, par exemple, se composer des unités suivantes de façon à constituer un échantillon de 100 : 3 (l'origine choisie au hasard), 7, 11, 15, 19... 395, 399 (jusqu'à **N**, qui est 400 dans ce cas).

Vous pouvez constater, à l'aide de l'exemple fourni ci-dessus, que dans le cas d'un échantillonnage systématique, seuls quatre échantillons possibles, qui correspondent aux quatre origines choisies au hasard également possibles, peuvent être sélectionnés:

1, 5, 9, 13... 393, 397

2, 6, 10, 14... 394, 398

3, 7, 11, 15... 395, 399

4, 8, 12, 16... 396, 400

Chaque membre de la population ne fait partie que de l'un des quatre échantillons et chaque échantillon a une chance égale d'être sélectionné. Cela nous permet de constater que chaque unité a une chance sur quatre d'être sélectionnée à l'intérieur de l'échantillon. Sa probabilité d'être sélectionnée est la même que si l'on sélectionnait un échantillon aléatoire simple de 100 unités. La principale différence tient au fait que dans le cas d'un échantillonnage aléatoire simple, toute combinaison de 100 unités aurait une chance de constituer l'échantillon, tandis que dans celui d'un échantillonnage systématique, il n'y a que quatre échantillons possibles. Cela nous permet aussi de constater à quel point l'échantillonnage systématique est précis comparativement à l'échantillonnage aléatoire simple. L'ordre de la population incluse dans la base de sondage

déterminera les échantillons possibles pour l'échantillonnage systématique. Si la population est distribuée au hasard dans la base de sondage, un échantillonnage systématique devrait alors produire des résultats similaires à ceux d'un échantillonnage aléatoire simple.

On utilise souvent cette méthode dans l'industrie, où l'on sélectionne une unité pour des essais à partir d'une chaîne de production afin de s'assurer que la machinerie et l'équipement sont d'une qualité uniforme. Un essayeur à l'intérieur d'une usine pourrait, par exemple, soumettre à un contrôle de la qualité chaque 20<sup>e</sup> produit sur une ligne de montage. L'essayeur pourrait choisir une origine au hasard entre les nombres 1 et 20. Cela déterminerait le premier produit à essayer; chaque 20<sup>e</sup> produit serait ensuite soumis à des essais.

Les intervieweurs peuvent utiliser cette technique d'échantillonnage lorsqu'ils interrogent des gens pour une enquête-échantillon. Le responsable d'une étude de marché pourrait sélectionner, par exemple, chaque 10<sup>e</sup> personne qui entrerait dans un magasin, après avoir sélectionné au hasard la première personne. Un enquêteur peut interviewer les occupants de chaque 5<sup>e</sup> maison d'une rue, après avoir sélectionné au hasard l'une des cinq premières maisons.

**Exemple n° 4 :** Imaginez que vous devez mener une enquête pour votre collègue ou votre université sur les logements pour les étudiants. Dix mille (10 000) étudiants sont inscrits dans votre établissement d'enseignement et vous voulez en prélever un échantillon systématique de 500. Pour ce faire, vous devez premièrement déterminer ce que serait votre intervalle d'échantillonnage (**K**) :

Population totale ÷ taille de l'échantillon = intervalle d'échantillonnage

$$N \div n = K$$

$$= 10\,000 \div 500$$

$$= 20$$

Il faudrait attribuer un numéro séquentiel à chaque étudiant pour entreprendre cet échantillonnage systématique. On choisirait le point de départ en sélectionnant un numéro au hasard entre 1 et 20. Si ce numéro était 9, on sélectionnerait alors le 9<sup>e</sup> étudiant inscrit sur la liste et chaque 20<sup>e</sup> étudiant par la suite. L'échantillon d'étudiants serait constitué de ceux qui correspondraient aux numéros d'étudiant 9, 29, 49, 69... 9 929, 9 949, 9 969 et 9 989.

Dans les exemples utilisés jusqu'ici, l'intervalle d'échantillonnage **K** était toujours un nombre entier, mais ce n'est pas toujours le cas. Par exemple, si vous prélevez un échantillon de 30 unités d'une population qui en compte 740, votre intervalle d'échantillonnage (ou **K**) sera 24,7. Dans de tels cas, il existe quelques possibilités de faire en sorte que le nombre soit plus facile à utiliser. Vous pouvez arrondir le nombre, c'est-à-dire au nombre entier supérieur le plus rapproché (le plus près) ou au nombre entier inférieur le plus rapproché. Le fait de l'arrondir au nombre entier inférieur le plus rapproché vous assurera de sélectionner au moins le nombre d'unités que vous vouliez à l'origine (et vous pourrez ensuite supprimer certaines unités pour obtenir la taille exacte de l'échantillon de votre choix). Il existe des techniques pour adapter l'échantillonnage systématique dans les cas où **N** (la population totale) n'est pas un multiple de **n** (la taille de l'échantillon), mais donne encore un échantillon qui est exactement le même que les unités **n**. Nous ne traiterons pas ici de ces techniques.

Les avantages de l'échantillonnage systématique tiennent au fait que la sélection de l'échantillon ne peut être plus facile (vous n'obtenez qu'un seul nombre aléatoire – l'origine choisie au hasard –

et le reste de l'échantillon suit automatiquement) et que l'échantillon est distribué dans des proportions égales à l'intérieur de la population répertoriée. Le plus gros inconvénient de la méthode d'échantillonnage systématique tient au fait que les échantillons possibles risquent de ne pas être représentatifs de la population s'il existe un certain cycle sur le plan du mode d'ordonnement de la population inscrite sur une liste et si ce cycle coïncide d'une quelconque façon avec l'intervalle d'échantillonnage. C'est ce que l'on peut constater dans l'exemple qui suit :

**Exemple n° 5 :** Supposez que vous dirigez une épicerie de grande surface et que vous possédez une liste des employés de chacune de ses sections. L'épicerie est divisée entre les 10 sections suivantes : le comptoir de charcuterie, la boulangerie, les caisses, les stocks, le comptoir des viandes, les fruits et légumes, la pharmacie, le magasin de photographie, le magasin de fleurs et le nettoyage à sec. Chaque section compte 10 employés, y compris un gérant (ce qui fait 100 employés au total). Votre liste est ordonnée par section, le gérant y étant énuméré le premier et les autres employés y étant ensuite inscrits dans l'ordre décroissant d'ancienneté.

Si vous voulez sonder vos employés au sujet de leurs réflexions sur leur milieu de travail, vous pourriez choisir un petit échantillon pour répondre à vos questions. Si vous utilisiez un échantillonnage systématique et si votre intervalle d'échantillonnage était 10, vous pourriez alors ne sélectionner finalement que les gérants ou que les employés de chaque section ayant le moins d'ancienneté. Ce type d'échantillon ne vous donnerait pas un portrait complet ni approprié des réflexions de vos employés.