Statistiques inférentielles

Introduction

L'objet de la statistique inférentielle est de généraliser les observations d'une variable statistique (ou variable aléatoire) d'un échantillon à une population dont est extrait cet échantillon.

L'exemple le plus connu est celui des sondages. Pour prévoir les résultats des élections, on effectue un sondage auprès d'un échantillon « représentatif » de la population et on essaie d'en déduire les résultats des élections.

Exemples :

1- Un fabricant souhaite vérifier la qualité des ampoules électriques produites par une nouvelle chaîne de production. Il faut donc évaluer la durée moyenne de fonctionnement des ampoules. Comment évaluer cette durée moyenne? On ne peut pas tester toutes les ampoules!

2- Le responsable d'un parti politique souhaite estimer la proportion des militants favorables à la candidature de Mr X pour la prochaine élection présidentielle. Comment calculer la popularité d'un candidat au sein d'une population? Interroger tous les militants est trop coûteux et nécessite beaucoup de temps.

Définitions :

La population : l'ensemble de tous les éléments considérés dans une étude.

Paramètres : mesures ou caractéristiques utilisées pour décrire une population.

L'échantillon est un sous ensemble fini de la population. La taille de l'échantillon est le nombre d'éléments sélectionnés pour constituer l'échantillon.

Observations de statistiques ou réalisations : mesures ou caractéristiques utilisées pour décrire un échantillon

Exemples :

1- Le fabricant d'ampoules. Il prélève un échantillon constitué de 100 ampoules. Pour chaque ampoule, il mesure la durée de fonctionnement. La moyenne de l'échantillon vaut 36 000 heures. Une **estimation** pour la population est 36 000 heures.

2- Le responsable du parti. Il constitue un échantillon de taille 400. Parmi les personnes sélectionnées, 250 sont favorables au candidat proposé. Une **estimation** de la proportion de la population favorable à Mr X est 250/400 = 0.625.

Question : Quelle est la qualité de ces deux estimations?

Exemple où l'on connait le résultat exact attendu (retour sur le calcul de probabilité !)

Considérons le cas d'un dé à 6 faces, que l'on suppose parfaitement équilibré : la population est $P = \{1, 2, 3, 4, 5, 6\}$. L'équilibre parfait de ce dé signifie qu'il n'y a aucune raison physique d'observer une face plus qu'une autre dans une série de lancers.

En jetant le dé n fois, on obtient bien sûr n faces : à chaque jet, la probabilité d'obtenir {1} est égale à 1/6, et la face obtenue au i^e jet n'a aucune incidence sur les autres faces obtenues : il y a équiprobabilité, et les lancers sont indépendants.

Dans ces conditions, l'expérience montre que, pour n suffisamment grand, la proportion de faces {1} va tourner autour de 1/6. De même la proportion de faces {2}, de faces {3} etc.

Exemple : *nous avons effectué n = 600 lancers d'un dé parfaitement équilibré. Les numéros ont tous été observés dans une proportion voisine de 1/6.*

	numéros					
	n° 1	*n° 2*	*n° 3*	*n° 4*	*n° 5*	*n° 6*
	102	*103*	*99*	*92*	*102*	*102*
Proportion	*0.17*	*0.171*	*0.165*	*0.153*	*0.17*	*0.17*

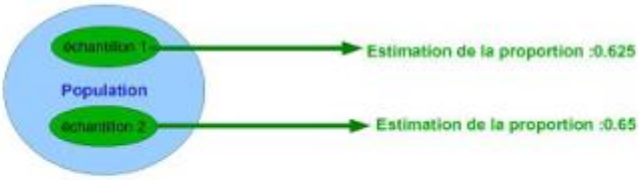
Alors que : $\frac{1}{6} = 0.166$

L'expérience du dé peut être schématisée à l'aide d'une urne contenant six boules numérotées de 1 à 6. Pour que les tirages soient indépendants, il suffit de remettre chaque boule tirée dans l'urne : les tirages sont donc effectués « avec remise ». On peut généraliser l'expérience en tirant dans une urne contenant un nombre quelconque de boules numérotées à partir de 1.

Axiome de la loi des grands nombres : On considère une population contenant N unités statistiques. On y effectue n tirages avec remise et on compte le nombre n_A de réalisations d'un événement A donné d'effectif N_A. La proportion observée n_A / n converge vers la probabilité N_A / N de l'événement A lorsque le nombre de tirages augmente indéfiniment.

On voit, à travers cet exemple, qu'un « échantillon » de 600 lancers donne une estimation du résultat pour obtenir l'une des 6 valeurs proche de 1/6 mais pas exactement égale à 1/6. En effet, un échantillon donne une estimation du résultat recherché et non une valeur exacte d'où la nécessité de :

1- **« Bien choisir » l'échantillon**
Erreur d'échantillonnage : Elle résulte de l'utilisation d'un sous ensemble de la population (l'échantillon) et non de la population toute entière. Exemple du responsable du parti : deux échantillons différents vont fournir des estimations différentes.



2- **Evaluer l'erreur de l'estimation ou encore l'écart entre l'estimation et la valeur réelle.**

Echantillonnage :

On veut, à partir d'un échantillon de la population, déduire des informations sur cette population. Le problème qui se pose alors est le suivant: comment choisir une partie de la population qui reproduit le plus fidèlement possible ses caractéristiques. C'est le problème de l'échantillonnage.

Différentes méthodes sont utilisées. On en cite :

1. **Echantillonnages sur la base des méthodes empiriques**

La Méthode des quotas (respect de la composition de la population pour certains critères) est la plus utilisée.

2. **Echantillonnages aléatoires**

Quand la probabilité de sélection de chaque élément de la population est déterminée avant même que l'échantillon soit choisi. Dans ce cas , on peut juger objectivement la valeur des estimations.

2.1 Echantillonnage aléatoire simple

On tire au hasard et avec remise les unités dans la population concernée

2.2 Echantillonnage stratifié

Subdiviser d'abord la population en sous-ensembles (strates) relativement homogènes. Extraire de chaque strate un échantillon aléatoire simple. Regrouper tous ces échantillons.

2.3 Echantillonnage par grappes

Choisir un échantillon aléatoire d'unités qui sont elles-mêmes des sous-ensembles de la population (grappes)

(Exemple : diviser la ville en quartiers; un certain nombre de quartiers sont choisis pour faire partie de l'échantillon; on fait l'enquête auprès de toutes les familles résidant dans ces quartiers).