

TD 1 AVEC RÉPONSES

Exercice 1 Dans une étude statistique qu'est-ce que sont la population et un échantillon ?

RÉPONSE.

La population est la collection complète de tous les individus à étudier.

Un échantillon est une sous-collection de membres sélectionnés dans une population.

Exercice 2 Qu'est-ce que sont un paramètre et une statistique ?

RÉPONSE.

Un paramètre est une mesure numérique décrivant certaines caractéristiques de la population.

Une statistique est une mesure numérique décrivant certaines caractéristiques d'un échantillon.

Exercice 3 Déterminez si les variables énumérées ci-dessous sont quantitatives ou catégorielles :

a. Le nombre de chaussures que vous possédez.

RÉPONSE : quantitative

b. La pointure des chaussures.

RÉPONSE : catégorielle

c. La quantité d'eau dans une bouteille.

RÉPONSE : quantitative

d. L'âge d'une personne.

RÉPONSE : quantitative

e. Le type de médicament anti-douleur.

RÉPONSE : catégorielle

f. La température d'une tasse de thé.

RÉPONSE : quantitative

g. Les parfums de glace.

RÉPONSE : catégorielle

h. Le pays d'origine.

RÉPONSE : catégorielle

i. Le salaire d'un enseignant.

RÉPONSE : quantitative

j. Le numéro de vol d'un vol d'avion.

RÉPONSE : catégorielle

k. Le nombre de passagers d'un vol.

RÉPONSE : quantitative

Exercice 4 Déterminez si les variables numériques ci-dessous sont discrètes ou continues.

- | | |
|--|---------------------------|
| a. La pression sanguine. | RÉPONSE : continue |
| b. La longueur d'un écran de téléphone. | RÉPONSE : continue |
| c. Le nombre de mots dans un livre. | RÉPONSE : discrète |
| d. La densité de l'air. | RÉPONSE : continue |
| e. Les battements cardiaques par minute. | RÉPONSE : discrète |
| f. Le nombre de globules rouges d'une personne. | RÉPONSE : discrète |
| g. Degrés centigrades de la température extérieure. | RÉPONSE : continue |
| h. La vitesse d'une voiture mesurée en kilomètres/heure. | RÉPONSE : continue |

RAPPEL

Échantillon aléatoire : les membres de la population sont sélectionnés par un mécanisme aléatoire.

Échantillonnage systématique : on choisit un point de départ et on sélectionne ensuite chaque k -ième (par exemple chaque 50-ème) élément de la population.

Échantillonnage de convenance : nous sélectionnons un échantillon avec une méthode facile à mettre en place.

Échantillonnage stratifié : on subdivise la population en au moins deux sous-groupes différents (strates) de sorte que les sujets d'un même sous-groupe partagent les mêmes caractéristiques (comme le sexe ou la tranche d'âge), puis on tire un échantillon de chaque sous-groupe.

Échantillonnage en grappes : nous divisons d'abord la population en sous-groupes (grappes), puis nous sélectionnons au hasard certaines de ces grappes, et enfin nous choisissons tous les membres de ces grappes sélectionnées.

Exercice 5 Identifiez le type d'échantillonnage (aléatoire, systématique, de convenance, stratifié, en grappe) utilisé lorsqu'un échantillon des 637.133 actionnaires est obtenu comme décrit.

- | | |
|--|-------------------------------|
| a. Une liste complète de tous les actionnaires est compilée et chaque 500-ème nom est sélectionné. | RÉPONSE : systématique |
| b. Lors de l'assemblée annuelle des actionnaires, un sondage est effectué auprès de tous les participants. | RÉPONSE : convenance |
| c. Cinquante courtiers boursiers sont choisis au hasard et un sondage est effectué auprès de tous leurs clients qui possèdent des actions d'IBM. | RÉPONSE : en grappes |
| d. Un fichier informatique de tous les actionnaires d'IBM est compilé de façon à ce qu'ils soient tous numérotés consécutivement, puis des numéros aléatoires générés par l'ordinateur sont utilisés pour sélectionner l'échantillon d'actionnaires. | RÉPONSE : aléatoire |

- e. Tous les codes postaux des actionnaires sont recueillis, et 5 actionnaires sont sélectionnés au hasard dans chaque code postal. **RÉPONSE** : stratifié

Exercice 6 Identifiez le type d'échantillonnage (aléatoire, systématique, de convenance, stratifié, en grappe) utilisé dans chacun des cas suivants.

- a. Nous recueillons des données d'échantillon en choisissant au hasard 20 pages différentes dans une version imprimée d'un dictionnaire, puis en calculant la longueur moyenne de tous les lemmes présents sur ces pages.

RÉPONSE : en grappes

- b. Nous recueillons des données d'échantillon en choisissant de manière aléatoire 5 phrases différentes dans chaque chapitre du roman *Harry Potter à l'école des sorciers*, puis nous calculons le nombre de mots dans toutes les phrases choisies.

RÉPONSE : stratifié

- c. Dans un sondage, 1007 adultes ont été appelés après que leur numéro de téléphone ait été généré au hasard par un ordinateur, et 85% des personnes interrogées ont pu identifier correctement ce qu'est Instagram.

RÉPONSE : aléatoire

- d. Lorsqu'il recueille des données à différents endroits dans un lac, un chercheur utilise la "méthode du transect linéaire" en tendant une corde à travers le lac et en prélevant des échantillons d'eau à tous les 5 mètres.

RÉPONSE : systématique

- e. La chaîne BFM TV obtient des opinions en interrogeant les voisins d'une personne qui fait l'objet d'un reportage.

RÉPONSE : convenance

Exercice 7 Dans le sondage du magazine *Literary Digest* sur les élections présidentielles américaines de 1936 il y avait eu deux types de biais. Énoncez ces biais et décrivez-les brièvement.

RÉPONSE. Le premier biais était un biais dû au fait que les personnes riches étaient largement plus représentées que les non-riches.
Le deuxième biais était un biais d'auto-sélection causé par le fait que seuls ceux qui souhaitaient répondre au sondage y ont participé.