

SunBurst

1. Introduction

Suite à un sondage réalisé par les étudiants de BUT 1, nous disposons d'une base de données riche en informations sur les habitudes alimentaires des étudiants de l'Université de Villetaneuse. Ces données offrent une opportunité d'explorer les liens entre différents facteurs tels que le sexe ou le temps et les choix alimentaires des usagers du campus de Villetaneuse.

L'objectif principal est de trouver des facteurs pouvant influencer les habitudes alimentaires des individus du campus. Plus spécifiquement, nous cherchons à déterminer si les habitudes alimentaires (type de repas, lieu de consommation, budget) changent en fonction du temps, du sexe et de la fonction sur le campus.

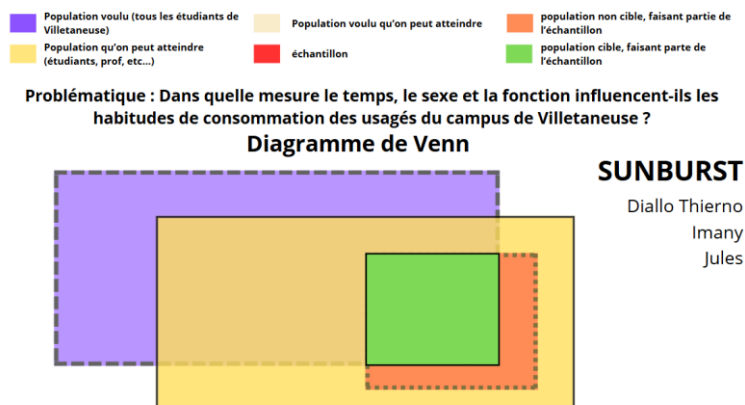
Problématique : Dans quelle mesure le temps, le sexe et la fonction impacte-t-il les habitudes alimentaires des usagers du campus de Villetaneuse ?

1.1 hypothèse/résultat attendu

- **Type de repas consommé:** Nous pensons que, en général, les étudiants consomment plus de fast-food quand il y a un temps de pause court.
- **Le budget moyen:** En général, le prix des aliments dans le campus est de 3.3€, on supposant que l'étudiant a acheté une boisson en plus, le budget moyen doit être au environ de 5€/jour.
- **Proportion de personnes qui consomment en dehors du campus:** Nous pensons que la proportion de personnes qui consomment en dehors du campus est assez faible, soit au tour des 10% (Ce chiffre vient uniquement de nos opinions, selon nos habitudes et observations sur le campus en fonction du temps de pause).

2. Présentation

2.1 Présentation de l'étude



L'échantillon principal, constitué de 1 097 participants interrogés par les étudiants de BUT 1, représente une part significative des usagers du campus de Villetaneuse (**7,83%**), bien qu'il soit majoritairement composé d'étudiants de l'IUT. Les réponses ont été collectées via un questionnaire numérique diffusé par QR Code, avec un effort notable pour équilibrer le genre des répondants. Cependant parmi les personnes sondées, il y a également du personnel du campus.

Cependant, des problèmes de qualités subsistent :

- La surreprésentation des étudiants de l'IUT limite la représentativité des autres départements du campus. En effet les étudiants de IUT ne représentent qu'environ 14,28%, alors que les étudiants de BUT1 (ayant mené l'enquête) ont reconnu que plus de 50% des répondants au questionnaire venaient de l'IUT). Selon notre expérience, en terme de budget, les étudiants de l'IUT et du reste du campus ne sont pas très différents, cependant, en terme de temps de pause, il peut y avoir de grandes différences (IUT : temps moyen de pause : 1h, et autres département : plus de 1h)

- Le questionnaire a principalement été administré au sein de l'IUT et donc cela peut créer un biais car les réponses des étudiants dans d'autres cursus sont moins représentées, par conséquent les réponses globales et moyennes seront fortement axées sur les réponses de l'IUT. Pour nous, cela peut poser problème car nous souhaitons prendre en compte des moyennes sur l'ensemble du campus, de plus on sait que l'IUT a un fonctionnement assez différents des autres institutions du campus, c'est pourquoi nous devons bien garder en tête que les résultats que nous obtiendront seront plus probablement similaires à ceux de l'IUT que ceux du campus dans son entièreté.
- Lors du recueil des données, les étudiants de BUT 1 sont partis de l'hypothèse que la population étudiante de l'IUT suivait un ratio équilibré de 50% : 50% entre filles et garçons. En conséquence, ils ont essayé de respecter cette répartition lors de l'échantillonnage. Cependant, comme nous ne disposons pas d'informations confirmant que ce ratio reflète réellement la composition de la population globale, il est possible que cet équilibre ne soit pas représentatif.

2.2 Présentation des Variables d'étude

2.2.1 Variable d'intérêt

- **Type aliment consommé (question 17) :**
 - type de variable : qualitative nominale
 - Modalités : "Plats type fast-food (burger, sandwich, pasta box, etc.)", "Plats cuisinés (salades, bœuf bourguignon, poisson, etc.)", "Snacking (biscuits, produits sucrés, etc.)", "Autre".
 - code modalité : Fast Food, Cuisiné, Snack, Autre
 - transformation à appliquer : Cette variable sera transformée dans une variable quantitative discrète, où on attribuera la valeur 1 à la modalité encodée "Fast Food" et 0 pour les autres modalités afin de pouvoir déterminer ceux qui mangent majoritairement au Crous
- **Budget (question 32) :**
 - type de variable : quantitative continue
 - Modalités : *Moins de 5€, de 5 à 10€, de 10 à 15€, plus de 15€.*
 - code modalité : *M_DE5, 5A10E, 10A15,*
 - transformation à appliquer : Cette variable sera transformée en variable quantitative discrète (en prenant le centre des classes) afin de pouvoir estimer le budget moyen.
- **Lieux de consommation (question 18) :**
 - Type de variable : Qualitative nominale
 - Modalités : Dans un lieu de restauration (Cantine, Cafétéria du Crous), Dans les couloirs du campus ou dans les espaces extérieurs du campus, Dans une salle de cours ou un bureau, Dans un espace collectif dédié (salle du personnels, BDE, etc.), En dehors du campus ou Chez moi.
 - Code de modalité : RESTAU, AUTRES, SALLES, COLLEC, DEHORS, MAISON, NONREP
 - Transformation à appliquer : Cette variable sera transformée dans une variable quantitative discrète, où on attribuera la valeur 1 à la modalité encodée "RESTAU" et 0 pour les autres modalités afin de pouvoir déterminer ceux qui mangent majoritairement au Crous.

Justification : Plus le temps de pause est long, plus les étudiants ont la possibilité de choisir entre différents lieux de consommation, comme le CROUS, les repas faits maison ou des restaurants. En revanche, lorsque le temps est limité, les choix se restreignent souvent aux options les plus rapides, comme les fast-foods ou les snacks prêts à consommer, qui sont non seulement moins diversifiés mais également, dans bien des cas, plus coûteux. Cette contrainte de temps accentue ainsi le rôle du budget alimentaire, rendant les repas rapides parfois moins accessibles pour les étudiants aux ressources limitées.

2.2.2 Variables auxiliaires

- **Temps moyen des pauses-déjeuner**
 - type de variable : quantitative continue
 - Modalités : de 0 à 30min, de 30min à 1h, de 1h à 1h30, de 1h30 à 2h, de 2h et plus
 - code modalité : Moins de 30 minutes, Entre 30 minutes et 1h, Entre 1h et 1h30, Entre 1h30 et 2h, Plus de 2h.
 - transformation : Nous allons regrouper les classes en 3 classes. de 0 à 1h et de 1h à 2h et de 2h à plus.

NB (hypothèse sur la répartition des temps de pause) : Après avoir consulté des informations sur le site l'Étudiant nous avons retrouvé le nombre des licences en 1^{er} année (2097) et des masters de première année (1600). Afin de se faire une approximation et de dégager des proportions, nous avons décidé d'être très optimiste en posant l'hypothèse que tous les étudiants de 1^{er} année parviennent à obtenir leur diplôme d'un coup. En faisant cela, nous obtenons un nombre de 6291 étudiant en licence et 3200 en master. Étant donné que les formations sont complémentaires nous les regroupons dans un

même groupe ce qui correspond à 9491 étudiants donc environ 68% du campus de Villetaneuse. D'après nos connaissances du campus ces étudiants ont généralement de 1h à 2h de pause. On sait qu'il y a 2000 étudiants à l'IUT et 1800 ingénieur, après avoir consulté de nombreux emplois du temps nous regroupons c'est deux types d'étudiants dans un seul groupe d'étudiants (28% du campus) mangeant en 1h. Et ensuite le reste du campus comprenant les professeurs, le personnel et les doctorants représentent environ 4% du campus. Nous supposons que ce dernier groupe a en moyenne plus de liberté au niveau du temps de pause et nous le mettons à plus de 2h. Nous supposons également que les professeurs ont le même temps de pause que leurs étudiants.

- raison d'utilisation : Le temps de pause est une caractéristique directement liée à la problématique de l'étude. Cependant, si les pauses de durée courte ou longue sont surreprésentées dans l'échantillon par rapport à la population globale, cela peut influencer la fiabilité des conclusions. Par exemple, si une majorité de répondants proviennent d'un département où les pauses sont courtes (comme l'IUT), cela pourrait biaiser les résultats sur les habitudes alimentaires, en masquant la diversité des pauses des autres étudiants. La variable peut être utilisée pour ajuster les résultats via des techniques comme le redressement proportionnel post stratification.
- **Fonction des individus**
 - type de variable : qualitative nominale
 - modalité : Étudiant.e, Je ne suis que de passage sur le campus, Personnel travaillant à l'université (enseignant, administratif, etc.)
 - Code modalité : E pour étudiant, P pour personnel
 - raison d'utilisation : La fonction des individus est une variable essentielle pour différencier les groupes qui fréquentent le campus. Elle permet de segmenter les réponses en fonction des usages spécifiques (par exemple, besoins et comportements différents des étudiants par rapport au personnel). Cela aide également à identifier les éventuelles disparités ou tendances dans les réponses entre les catégories (étudiants, personnel).
- **Sexe**
 - type de variable : Qualitative nominale ,
 - Modalité : Homme, Femme, Non genré ,
 - Code modalité : H, F, N
 - raison d'utilisation : Le sexe est une caractéristique fondamentale qui permet de s'assurer que l'échantillon reflète la composition réelle de la population . Par exemple, si la proportion homme/femme dans l'échantillon diffère de celle de la population ciblée, cela peut introduire un biais dans les résultats. En cas de déséquilibre entre les proportions observées dans l'échantillon et celles attendues dans la population ciblée, un redressement par post-stratification peut être effectué. Cela consiste à ajuster les poids des observations en fonction des proportions réelles dans la population. **NB** : n'ayant pas trouvé les données spécifiques au campus de Villetaneuse, nous avons cherché des données pour tous les campus de France, nous avons obtenu : 55% de femme, 44% homme et 1% neutre. Cela signifie que, même si la variance est un peu plus importante, l'estimation finale reflète mieux la structure réelle de la population étudiée.

3. Transformation des données

La première étape consiste à examiner la base de données pour identifier et corriger les erreurs potentielles, telles que :

- Les valeurs manquantes ou incohérentes.
- Les doublons ou les saisies incorrectes.
- Certaines variables, bien qu'encodées sous forme qualitative ou textuelle, représentent en réalité des informations quantitatives. Par exemple, la variable décrivant les dépenses alimentaires avec des modalités telles que "Moins de 5€", "de 5 à 10€", "de 10 à 15€", et "plus de 15€" sera transformée en une variable quantitative discrète avec les valeurs correspondantes: 5, 10, 15, 20. Cette transformation permettra d'effectuer des analyses statistiques plus précises.
- Sélection des données pertinentes : Après le nettoyage complet, nous filtrerons les données pour conserver uniquement celles qui sont directement pertinentes pour notre étude. Les observations non pertinentes ou inexploitable seront exclues.

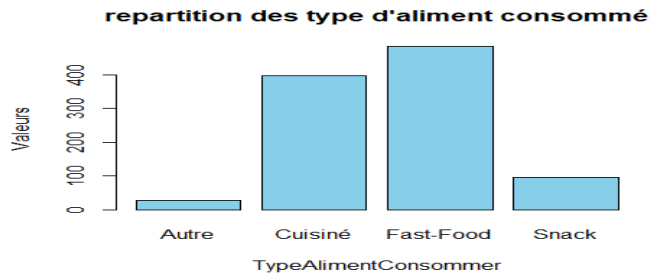
4. Analyse descriptive

Pour chaque variable d'intérêt, nous allons faire un histogramme (pour les variables quantitatives) et un diagramme en barre (pour les variables qualitatives), nous effectuerons également des calculs de moyenne, de variance et de quartile afin de déterminer la distribution de nos variables (discrètes) et des calculs de proportions (variable nominale).

Les variables auxiliaires étant majoritairement qualitative et quantitative continue, nous effectuerons principalement des graphiques en barres (pour visualiser les répartitions)

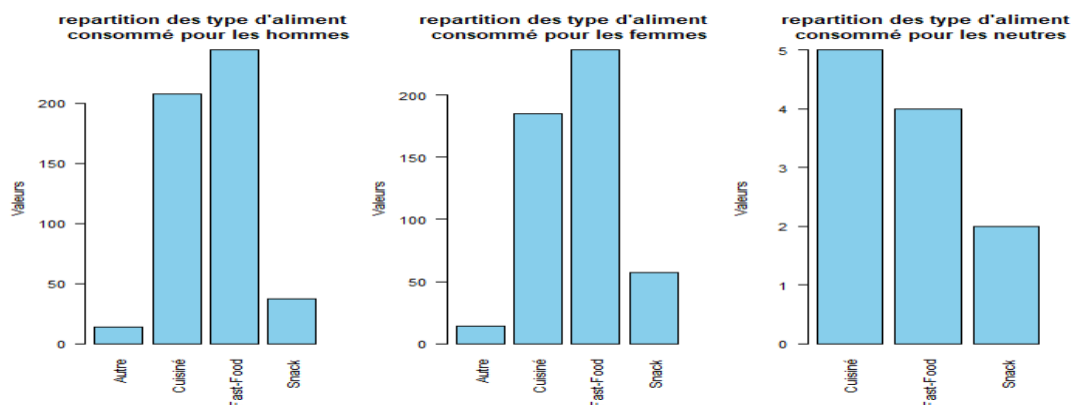
4.1 variable d'intérêt

4.1.1 Distribution des types de repas consommé



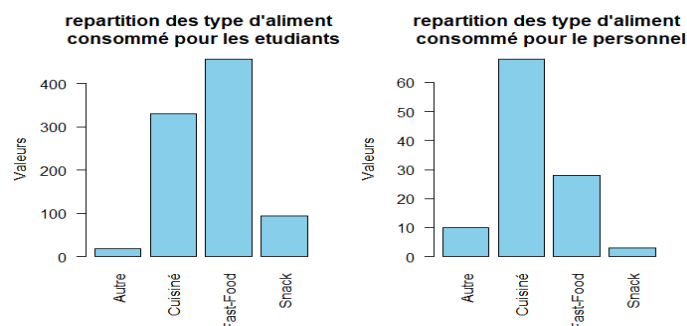
Ci-dessus nous avons un graphique présentant, la répartition des types d'aliments consommés. On observe que les plats les plus consommés sont les fast-food avec 48,11% de répondants, ensuite il y a les plats cuisinés avec 39,59%. Enfin les deux dernières modalités avec le moins de répondants sont les Snacks avec 9,52% et les Autres types de plats 2,78%.

Par sexe



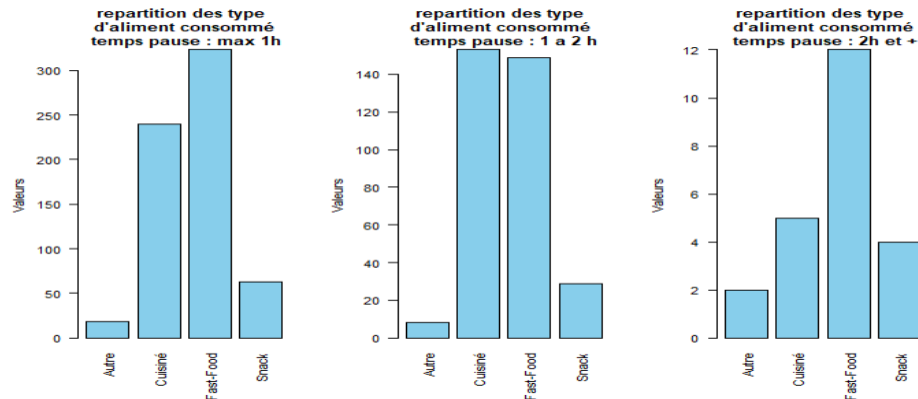
Ici nous avons 3 graphes sur la répartition des types d'aliments consommés pour les sexes. Le premier graphe correspond aux hommes, le deuxième aux femmes et le dernier aux non genrés. Les graphes des hommes et des femmes ont des répartitions assez similaires avec les fast-food et les plats cuisinés en tête, malgré une proportion de snack un peu supérieur chez les femmes. Pour les non genrés la répartition est très différente, on constate que les types d'aliments les plus consommés sont les plats cuisinés suivis des fast food et des snacks, il n'y a aucun autre type d'alimentation. Cependant au vu du nombre de répondant non genré très faible à l'échelle de l'échantillon la répartition des réponses est beaucoup plus marginal.

Par fonction



On a ici la répartition des types d'aliments par rapport à la fonction au sein du campus, nous avons le personnel face aux étudiants. On remarque qu'il y a beaucoup plus d'étudiants que de personnel. Le premier graphe sur les étudiants montre que les types d'aliments les plus consommés sont les fast food suivis par les plats cuisinés et ensuite dans une moindre mesure il y a les snacks et les autres types d'aliments. Le graphe du personnel est très différent car les plats cuisinés sont de loin les plus consommés ensuite il y a les fast food puis les autres types d'aliments et enfin les snacks. On garde en tête que la différence du nombre de répondant entre les deux fonctions peut jouer sur les répartitions des deux graphes.

Par temps de pause



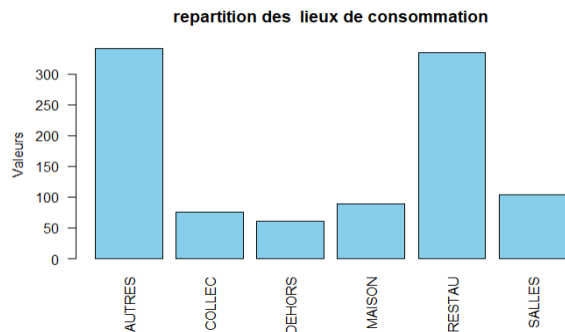
Nous avons les 3 graphes respectifs au temps de pause dans l'ordre suivant : 0 à 1h où l'on remarque que les fast food sont les plats les plus consommés suivis par les plats cuisinés.

On remarque que dans le graphe correspondant aux pauses de 1 à 2h la répartition change et les plats cuisinés sont les repas les plus consommés suivis de très près par les fast foods. Pour les pauses de plus de 2h les fast food sont de loin les plats les plus consommés. Ces graphiques rejoignent notre problématique et hypothèse de départ qui disait que le temps de pause influence les habitudes alimentaires. Ici on remarque que entre (0 à 1h) et (1h à 2h) la répartition des aliments consommés change, lorsque le temps de pause a augmenté les plats cuisinés ont été plus consommés.

La où on pourrait se poser des questions c'est au niveau du graphe du temps de pause supérieur à 2h, mais une explication peut être donné quant à la forte représentation des fasts food dans ce graphe. En effet le temps de pauses de + de 2h, favorise les déplacements hors de l'établissement et de prolonger son temps de consommations d'un plat de type fast food, c'est ce que la population majoritairement jeune du campus est enclin à faire.

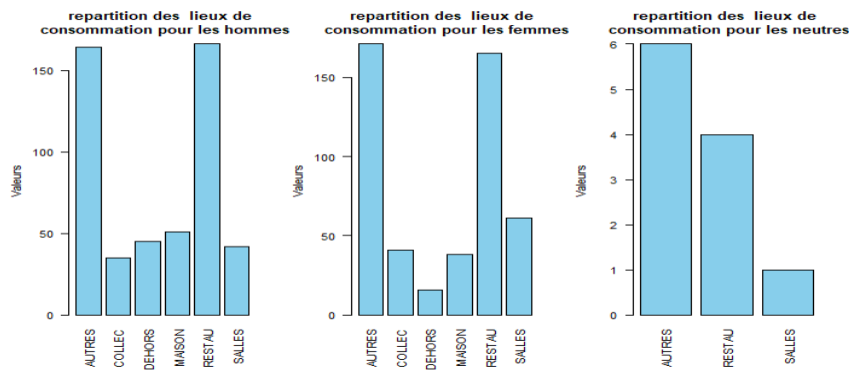
Cela peut expliquer pourquoi on retrouve les fast food et non les plats cuisinés comme type d'aliments plus consommés par ceux qui ont + de 2h de pauses. Globalement on remarque que la répartition des types d'aliments consommés change en fonction du temps de pause.

4.1.2 Distribution des lieux de consommations



Ce graphique nous présentent les distributions des usagés du campus par lieux de consommation de leurs déjeunés. Ainsi, nous constatons que les catégories les plus dominantes sont RESTAU qui représente la cantine Crous et Autre qui représente tous autre lieux n'étant pas présent dans cette liste. On voit que ces deux modalités ont environs chacune plus de 300 personnes. Puis loin derrière, arrive les catégories restant dans cet ordre (Salles, maison, collec, dehors) avec des valeurs comprises entre 100 et 50. Au final, nous voyons une separation nette entre les catégories dominantes et les autres. Ce qui montre une certaine hétérogénéité.

Par sexe



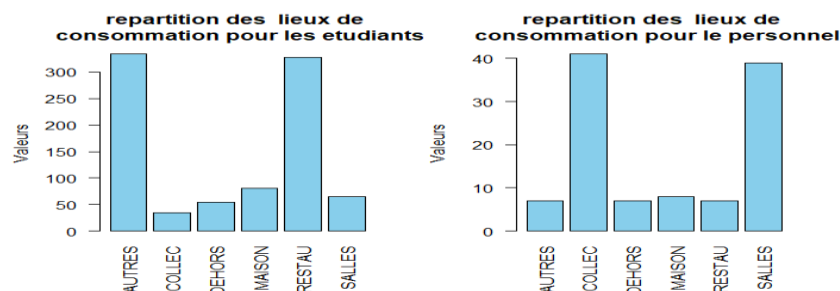
Ces graphiques nous présentent les répartition des lieux de consommation des usagés du campus par rapport au genre.

Pour la répartition du lieu de consommation pour les hommes, nous constatons qu'il est très proche de la tendance globale, avec AUTRES et RESTAU qui domine et les autres catégories qui suivent. cependant on constate une légère augmentation de MAISON et DEHORS mais également une légère baisse de SALLES. Les hommes mangent donc plus souvent chez eux (par rapport à la tendance globale et aux autres)

Pour la répartition du lieu de consommation pour les femmes, nous voyons qu'elle est très proche de la tendance globale (comme celui des hommes) mais avec une baisse de DEHORS et une augmentation de COLLEC et SALLES. Ce qui suggère qu'elles mangent généralement au sein du Campus, mais également entre ami (salles collectif).

Pour la répartition du lieu de consommation pour le genre Neutre, nous constatons qu'il est très différent de la tendance globale. Les deux catégories dominantes sont toujours la (RESTAU et AUTRES). Mais nous voyons que toutes les autres catégories ont disparues excepté SALLES.

Par fonction

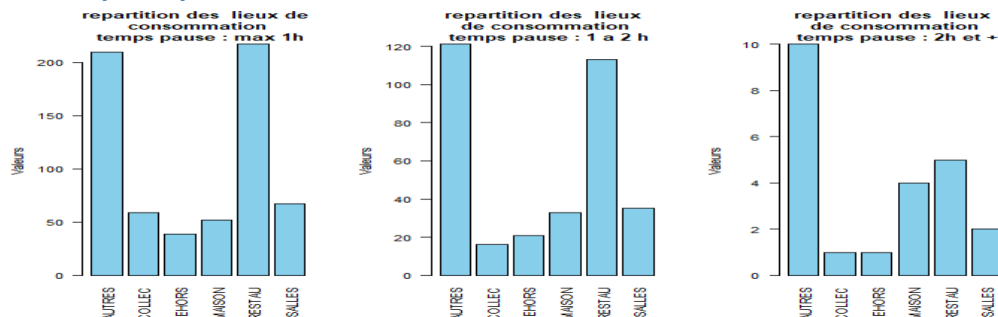


Ces graphiques nous présentent les répartitions des lieux de consommation par rapport à leur fonction (étudiants, personnel). Nous constatons :

Pour les étudiants, ils sont assez proches de la tendance globale. Les catégories dominantes sont toujours AUTRES et RESTAU. Cependant nous voyons une grande augmentation de MAISON et SALLES. Cela peut être du soit au fait qu'ils aient envie de mangé avec des amis ou même peut être afin de pouvoir réviser en vue du cours qui va suivre.

Cependant, pour le personnel, la répartition est totalement différente. Ici les catégories dominantes sont COLLEC et SALLES, ce qui est raisonnable car ils ont une salles de restauration réservé au personnel, mais une grande partie mange aussi directement dans leur bureau. Et les autres modalités ont des valeur proches.

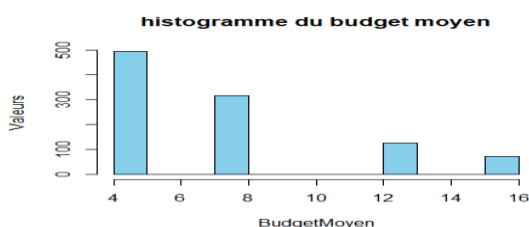
Par temps de pause



Ces graphiques nous présentent les répartitions du lieu de consommation par temps moyen de pause.

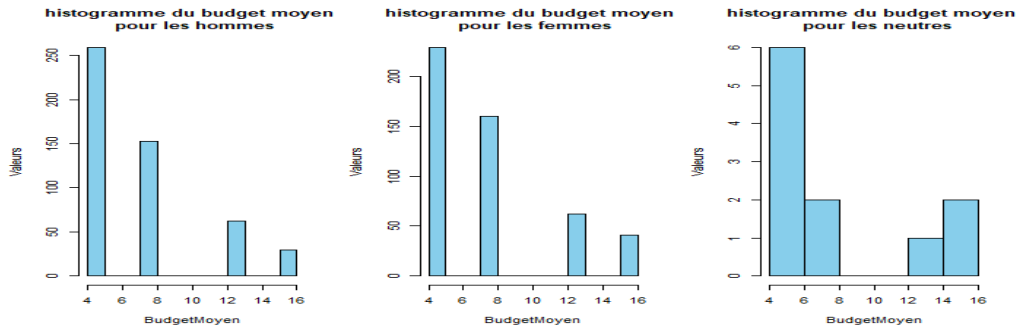
Nous constatons que les répartitions de lieu de consommation pour un temps de pause de maximum 1h et entre 1h et 2h sont très similaires et sont proche de la tendance générale avec AUTRES et RESTAU qui sont les catégories dominantes. Cependant, pour un temps de pause de 2h et plus, nous constatons que seul la catégorie AUTRES contient de nombreuses personnes. Et la catégorie RESTAU a beaucoup baissé. Cela peut être du au fait que quand ils ont plus de temps, ils vont mangé ailleurs ou chez eux.

4.1.3 Distribution du budget



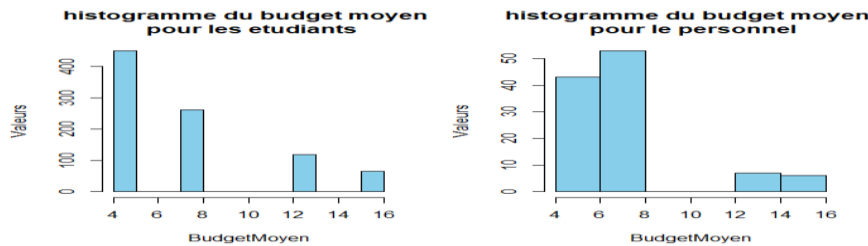
Le graphique de répartition du budget prend en compte les classes centrées que nous avons effectué à partir des réponses aux modalités du questionnaire. On observe que la plupart des répondants disent avoir un budget d'environ 4 à 5€ par repas ou bien 7 à 8€. Une minorité de 12 à 13€ et entre 14 et 16 €.

Par sexe



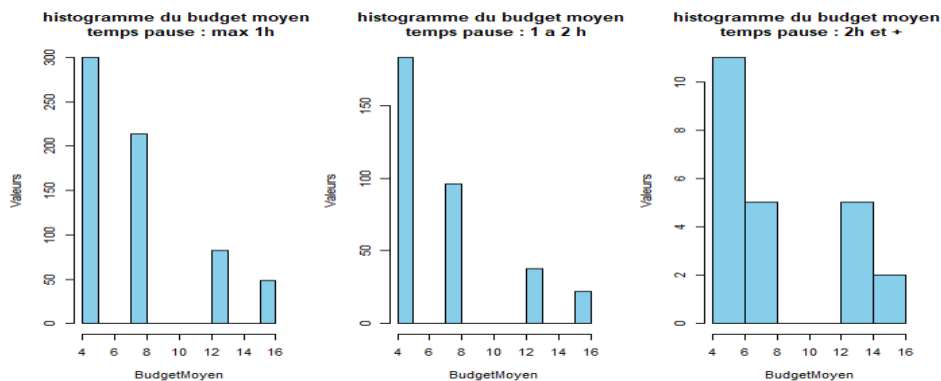
Sur ces graphes du budget moyen par rapport au sexe on observe que le budget moyen des hommes et des femmes est similaire, la majorité dépense aux alentours de 4€, la différence se fait avec un petit peu plus de femmes qui dépensent aux alentours de 16 €. Pour les neutres la répartition est différente la majorité d'entre eux ont un budget de 4€ mais la part de personnes dépensant aux alentours de 16€ est supérieure aux deux autres graphiques et à la répartition de base.

Par fonction



Ici nous avons le graphique sur la répartition du budget entre les étudiants et le personnel. On remarque que le budget des étudiants est similaire au budget de base récolté sur l'échantillon. Contrairement au budget du personnel qui a principalement un budget d'environ 8€ suivi par un budget d'environ 4€. De plus la répartition entre les dépenses aux alentours de 12€ et celles aux alentours de 16€ est très similaire, ce qui diffère du graphique de base. On pourrait expliquer cette répartition des budgets, par le niveau de vie moyen des étudiants et du personnel. Le personnel étant un groupe de salariés ils ont plus d'argent à dépenser que les étudiants qui peuvent parfois être en situation précaire.

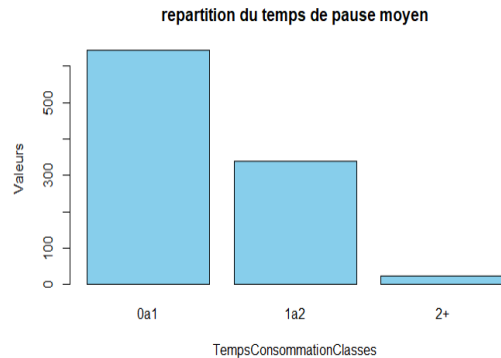
Par temps de pause



Ici on a la représentation du budget moyen par rapport au temps de pause. On observe que les graphes du temps de pause de 0 à 1h et de 1h à 2h sont très similaires avec pour légère différence un budget d'environ 8€ plus représenté sur le graphique de 0 à 1h. Pour le graphique de 2h ou plus on constate que le budget de 4€ reste majoritaire mais le budget d'environ 8€ et celui aux alentours de 12€ sont également représentés. On pourrait croire que plus le temps de pause est court plus les personnes auront tendance à dépenser de l'argent pour gagner du temps. Ces graphiques ne permettent ni de confirmer cette hypothèse ni de la réfuter.

4.2 variable auxiliaire

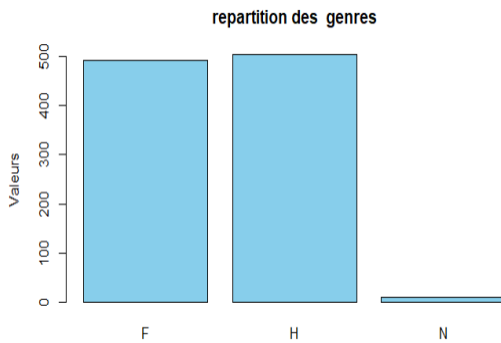
4.2.1 Distribution du temps de pause



Ce graphique montre que la plupart des individus prennent des pauses courtes, inférieures à une heure (catégorie "0a1"). Un nombre moins important de personnes prend des pauses d'une à deux heures (catégorie "1a2"), et très peu de personnes prennent des pauses supérieures à deux heures (catégorie "2+"). Cela suggère une tendance générale à privilégier des pauses courtes dans le groupe étudié.

Cela peut-être du au fait que la majorité des répondants sont des étudiants de l'IUT (selon notre hypothèse, ils ont en moyenne 1h de pause au max), on voit bien que notre échantillon ne respecte pas les vrais proportions de la répartition du temps de pause, à savoir, 28% au maximum 1h, 64% entre 1 et 2 heure et 8% plus de 2h. Un redressement avec cette variable auxiliaire s'avère donc intéressant.

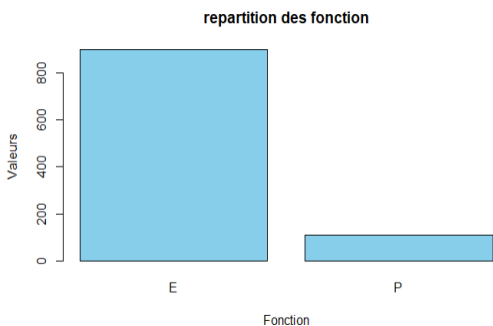
4.2.2 Sexe des individus



Les genres féminins (F) et masculins (H) sont très largement majoritaires et semblent être représentés en nombre à peu près égal. Les barres associées à ces deux catégories atteignent toutes deux une hauteur similaire, autour de 500. La catégorie "N" (neutre) est très peu représentée. La barre associée à cette catégorie est nettement plus courte que celles des deux autres, indiquant un nombre beaucoup plus faible d'individus se déclarant de ce genre.

Nous constatons également que notre échantillon ne respecte pas la vraie répartition des différents genres sur le campus de Villeteuse (selon notre hypothèse et nos recherches). Dans notre échantillon la répartition est 49% femme, 50% homme et 1% neutre. Alors que la vraie répartition est de 55% femme, 44% homme et 1% neutre. Un redressement avec cette variable auxiliaire s'avère donc intéressant.

4.2.3 Fonction des individus



La barre représentant la catégorie "Étudiant" est nettement plus haute que celle représentant la catégorie "Personnel". Cela signifie qu'il y a un nombre significativement plus important d'étudiants que de personnel dans le groupe étudié. La barre représentant la catégorie "Personnel" est beaucoup plus courte, indiquant une proportion nettement plus faible de personnel par rapport aux étudiants.

La répartition Personnel et Étudiant sur le campus est de 95% étudiant et 5% personnel (selon nos hypothèses et données trouvées sur le site de l'université). Dans notre échantillon les répartitions sont de 89% de étudiant et 11% de personnel. On constate donc que notre échantillon ne respecte pas la distribution réelle du campus. Un redressement avec cette variable auxiliaire s'avère donc intéressant.

5. Analyse inférentiel

Dans le cadre de cette étude, notre objectif est d'estimer la proportion d'étudiants du campus de Villeteuse qui consomment régulièrement des fast-foods, leurs lieux de consommation et leurs budget moyen. Ces estimations nous permettront non seulement de mieux comprendre les comportements alimentaires sur le campus, mais aussi de mettre en lumière les variables influençant ces habitudes (temps de pause, sexe, fonction).

Tout au long de ce travail, nous nous concentrerons sur l'estimation des proportion/moyenne de nos variables d'intérêt. Et nous chercherons à affiner ces estimations en appliquant une méthode de post-stratification proportionnel avec les différentes variables auxiliaires (temps de pause, sexe, fonction). L'analyse inclura une évaluation de la variance de l'estimateur (plan simple), suivie de l'impact du redressement post-stratifié sur la précision des résultats. Et nous prendrons soin de discuter des hypothèses nécessaires à la validité de ces estimations, tout en mettant en lumière les erreurs possibles et la qualité de l'analyse statistique.

estimateur : $\bar{X} = \frac{1}{N} \sum_{i=0}^N X_i$ (Notre estimateur suit une loi normale de moyenne μ et de variance S^2/n)

variance (plan simple) : $var(\bar{X}) = \left(1 - \frac{1}{N}\right) \frac{S^2}{n}$

moyenne post stratifié : $\bar{X}_{post} = \sum_{h=1}^H \frac{N_h}{N} \bar{X}_h$

variance post stratifié :
$$\text{Var}(\hat{\mu}_{post}) := \frac{1}{n} \left(1 - \frac{n}{N}\right) \frac{1}{N} \sum_{h=1}^H N_h S_h^2 + \frac{1}{n^2} \left(1 - \frac{n}{N}\right) \sum_{h=1}^H \frac{N - N_h}{N} S_h^2$$

5.1 Variable 1 : type de repas consommé

Modalité : $\bar{X}_i = 1$ si ils consomment fréquemment des fast-foods, 0 sinon. Donc $\mu = p$ et $S^2 = p(1-p)$

Estimation

Calcul de la moyenne et variance de notre échantillon

J'ai d'abord transformé la variable en variable de bernoulli (0 ou 1), puis effectué le calcul de la moyenne et de la variance. on voit que 48,11% des répondants au questionnaire reconnaissent consommer fréquemment des fast-food à la pause de midi, ce qui représente près de la moitié des usagers du campus interrogés. Cette proportion met en évidence une habitude alimentaire répandue parmi les étudiants, suggérant un intérêt particulier pour ce type de restauration rapide. Cela peut-être du au court temps de pause, seul option disponible, ou même par choix. on a également une variance de 0.2499. Cette valeur reflète la dispersion des réponses autour de la proportion estimée. Cette variance élevée indique une hétérogénéité importante dans les comportements alimentaires des usagers du campus.

variance de l'estimateur : plan simple

On sait que notre population total est de 14000 ($N = 14000$)

Pour une estimation normale (plan simple), nous obtenons une moyenne de 48,11%. La variance de l'estimateur de 0,0002305519 dans un plan d'échantillonnage simple indique une très faible dispersion de l'estimation autour de la moyenne réelle de la population. Cela signifie que l'estimation de la proportion des étudiants qui consomment fréquemment des fast-foods est très précise, avec peu de variation attendue entre différents échantillons tirés de la même population.

Redressement

Par sexe

Pour un redressement par sexe, nous obtenons une moyenne de 48,09%. Avec une variance de l'estimateur de 0,0002313166 après le redressement par stratification selon le sexe, on remarque une légère augmentation de la variance par rapport à celle obtenue dans le plan d'échantillonnage simple (0,0002305519). Cette petite augmentation pourrait s'expliquer par une variabilité introduite par les pondérations utilisées dans le redressement car l'échantillon n'était pas très représentatif de la population. Cependant, il ne faut pas oublier que, n'ayant pas trouvé les données spécifiques au campus de Villetaneuse, nous avons chercher des données pour tous les campus de france, nous avons obtenu : 55% de femmes, 44% d'hommes et 1% neutre. Cela signifie que, même si la variance est un peu plus importante, l'estimation finale reflète mieux la structure réelle de la population étudiée.

Par temps moyen de pause

Pour un redressement par temps moyen de pause, nous obtenons une moyenne de 46,02%. Avec une variance de l'estimateur de 0,0002297833 après stratification par le temps moyen de pause, on observe une légère diminution de la variance par rapport à celle obtenue dans le plan d'échantillonnage simple (0,0002305519). La diminution de la variance montre que cette stratégie de redressement améliore la précision de l'estimation sans introduire de biais important (si notre hypothèse sur la répartition est vrai). Cela peut être dû au fait que les comportements alimentaires, comme la consommation de fast-foods, sont en partie influencés par la durée des pauses disponibles. La stratification par le temps moyen de pause permet une estimation plus précise et peut être considérée comme une meilleure stratégie pour réduire la variance de l'estimateur.

Par fonction

Pour une stratification par fonction, nous obtenons une moyenne de 49,58%. Avec une variance de l'estimateur de 0,0002283693 après stratification par la fonction (étudiant, professeur), on constate une réduction de la variance par rapport au plan d'échantillonnage simple (0,0002305519) et aux autres stratifications précédentes. La stratification par fonction (étudiant ou professeur) semble être la stratégie la plus efficace pour réduire la variance de l'estimateur et améliorer la précision des résultats. Cette méthode permet de mieux tenir compte des différences structurelles entre ces deux groupes sur le campus.

Conclusion et calcul de l'intervalle de confiance

Les résultats montrent que la variance de l'estimateur dans le plan d'échantillonnage simple est de 0,0002305519, tandis que les stratifications par SEXE(0,0002313166), temps moyen de pause (0,0002297833) et fonction (0,0002283693) ont produit des variances légèrement différentes. Bien que la stratification par fonction ait permis une réduction plus significative de la variance, ces gains en précision restent très faibles et ne justifient pas nécessairement l'ajout de complexité.

De plus, compte tenu de ces incertitudes sur les hypothèses que nous avons émises (les vrais valeurs de la répartition des sexes, des fonctions, ou même les vrais ratios des temps de pause), et du fait que la variance du plan simple est déjà suffisamment faible pour répondre à nos besoins, le plan simple s'avère être l'approche la plus adaptée. Sans oublier que nous avons également obtenu des moyennes assez similaires (48%). Nous décidons donc de privilégier le plan simple.

L'**intervalle de confiance** de [0.48066 ; 0.48157] indique que la vraie valeur de la proportion se situe entre 48,066% et 48,157% avec une probabilité de 95%, confirmant ainsi que la moyenne estimée est autour de 48%

5.2 Variable 2 : Lieux de consommation repas

Modalité : $\bar{X}_i = 1$ si DEHORS, 0 sinon. Donc $\mu=p$ et $S^2=p(1-p)$

Estimation

Variance de l'estimateur : plan simple

Il faut garder en tête que la moyenne de l'estimateur par plan simple est (0.06063618) et sa variance est 5.260348e-05. Cette valeur étant très faible cela témoigne de sa précision et qu'il s'agit déjà d'une méthode assez pertinente d'autant plus que nous sommes certains de la qualité des données grâce aux informations réelles que nous avons pu obtenir.

Redressement

Lorsque l'on compare la variance simple de l'estimateur (5.260348e-05) avec celle des estimateurs stratifiés

Par sexe

Par sexe : (4.918686e-05) avec pour moyenne post stratifié (0.05725). On remarque que les valeurs de la variance et de la moyenne sont légèrement inférieures à celles du plan simple. Cette stratification est donc légèrement plus précise que l'estimation du plan simple.

Par fonction des individus

Par fonction : variance (5.249674e-05) avec pour moyenne post stratifié (0.060402). On observe que les valeurs de la variance et de la moyenne sont très légèrement inférieures à celles du plan simple. Elles restent très sensiblement similaires.

Par le temps moyen de pause

Par temps moyen de pause : variance (5.298139e-05) avec pour moyenne post stratifié (0.06082). On constate que les valeurs de la variance et de la moyenne sont légèrement supérieures à celles du plan simple. Étant donnée la complexité de la mise en place de cette stratification, le léger manque de précision par rapport au plan simple, cette stratification n'est donc pas pertinente.

Conclusion et calcul de l'intervalle de confiance

On observe que la variance la plus faible est celle avec la stratification par sexe suivi de celle par fonction.

On pourrait donc opter pour la variance stratifiée par sexe étant donnée qu'elle est la moins éloignée de la proportion de l'estimateur. Cependant il faut garder en tête que les valeurs que nous avons utilisées sont basées sur des hypothèses que nous avons émises à partir de nos recherches, elles ne garantissent pas une qualité fiable à 100%. De plus les valeurs des différentes moyennes et variances sont assez proches les unes des autres.

En prenant en compte le contexte que nous avons déjà expliqué, il reviendrait donc d'utiliser le plan simple car ce plan nous permet une meilleure garantie de la qualité des données, et en ajoutant la proximité par rapport aux valeurs des autres plans on ne perd pas beaucoup en précision.

En conclusion dans notre contexte, dans lequel nous avons eu accès à très peu de données réelles et dans lequel nous sommes principalement basés sur des hypothèses, le plan simple nous semble le plus pertinent pour calculer nos estimations sur la proportion d'individus mangeant en dehors du campus afin d'éviter les risques de trop s'éloigner de la réalité.

L'**intervalle de confiance** de [0.06053308 ; 0.0673929] indique que la vraie valeur de la proportion se situe entre 6,053% et 6,739% avec une probabilité de 95%, confirmant ainsi que la moyenne estimée est autour de 6%

5.3 Variable 3: Budget moyen

Estimation

Calcul de la moyenne et de la variance

Nous constatons après calcul que le budget moyen (montant consommé par un usagé du campus) par jour est de 7€, ce qui semble assez élevé vu que les repas consommés au crous sont de 3.30€.

On a également une variance de 14.07. Cette valeur reflète la dispersion des réponses autour de la proportion estimée. Cette variance élevée indique une hétérogénéité importante dans les dépenses journalière.

Variance de l'estimateur: plan simple

La faible variance de l'estimateur dans un plan simple est de 0,01297882. Cela montre que la moyenne de 7€ est une estimation fiable du budget quotidien moyen des usagers du campus. Si les données sont représentatives de la population, afin d'en être sûr, nous effectuerons des redressement par rapport à différentes variables dont nous connaissons les répartitions.

Redressement

Par sexe

Pour une stratification par sexe, nous obtenons une moyenne d'environ 7€. La nouvelle variance est légèrement plus élevée (0,0131 > 0,01298), ce qui indique que le redressement par sexe a introduit un peu plus de variabilité dans l'estimation. Cela peut être dû à des déséquilibres dans l'échantillon initial (sous-représentation des femmes).

Par fonction des individus

Pour une stratification par fonction, nous obtenons une moyenne de d'environ 7€. La variance de l'estimateur après stratification par fonction (étudiant, professeur) est de 0,0132, ce qui est légèrement supérieure à celle obtenue avec le plan d'échantillonnage simple (0,01297882) et à celle observée après stratification par sexe (0,0130949733484864).

Par le temps moyen de pause

Pour une stratification par le temps de pause moyen, nous obtenons une moyenne de 7€. La variance de l'estimateur après redressement est de 0,01286, ce qui représente une légère réduction par rapport au plan d'échantillonnage simple (0,012978) et aux autres méthodes de stratification précédentes, comme celle par sexe (0,01309497) ou par fonction (0,0132335). Cette réduction de la variance montre que la stratification par temps de pause moyen est efficace pour capter la variabilité des comportements de consommation sur le campus. En prenant en compte cette variable, on peut mieux expliquer les disparités dans les dépenses moyennes des usagers.

Conclusion et calcul de l'intervalle de confiance

Les résultats montrent que la variance de l'estimateur dans le plan d'échantillonnage simple est de 0,01297882, tandis que les stratifications par sexe (0,01309497), temps moyen de pause (0,01286096) et fonction (0,01323351) ont produit des variances légèrement différentes. Bien que la stratification par temps de pause moyen ait permis une légère réduction de la variance, les gains en précisions sont minimes et ne justifient pas nécessairement l'ajout de complexité à l'échantillonnage. De plus, en tenant compte des incertitudes liées aux hypothèses formulées sur les proportions réelles de sexe, de fonction ou même des temps de pause, et étant donné que la variance du plan d'échantillonnage simple est déjà suffisamment faible pour répondre à nos besoins, nous concluons que le plan simple reste la méthode la plus adaptée. De plus, les moyennes obtenues sont assez similaires, avec seulement des différences mineures (10^{-2}) près. Nous décidons donc de privilégier le plan d'échantillonnage simple, qui est à la fois plus simple à mettre en œuvre et suffisamment précis pour nos objectifs. L'**intervalle de confiance** de [6.99 ; 7.04] indique que la vraie valeur du budget moyen se situe entre au tour de 7€ avec une probabilité de 95%.

6. Conclusion

Pour faire un bilan de cette modélisation, nous avons compris différentes choses. Tout d'abord la phase préliminaire de l'enquête est une des plus importantes pour la suite du sondage, il faut au préalable se renseigner au maximum afin d'avoir des données réelles sur la population étudiée et ciblée. Prendre le temps de choisir un axe d'analyse réalisable. Ensuite savoir adapté le plan de sondage à utilisé en fonction de la situation et des résultats.

Dans notre travail, on a constaté que le redressement n'est pas toujours la meilleure option à privilégier notamment car nous nous sommes basé sur des hypothèses pour les répartitions réelles de chaque strate. La qualité des données n'a donc pas pu être garantie à ce niveau là. Avec nos résultats nous pouvons répondre à notre problématique en affirmant que les facteurs du statut et du temps de pauses ont bel et bien une influence sur les habitudes alimentaires. Par exemple, on remarque que la fonction sur le campus et le temps de pause peut influencer sur le type d'aliment consommé et le lieux de consommation. Cependant pour le sexe nous ne pouvons rien dégager de nos résultats.

Nous gardons toujours en tête que nos calculs et nos analyses peuvent être biaisé car ils reposent intégralement sur l'échantillon recueilli par les BUT1 car nous avons opter pour un plan simple.

Si l'enquête était à refaire, il serait pertinent d'améliorer la collecte des données pour mieux représenter la population cible et envisager des plans d'échantillonnage plus complexes, comme la stratification ou la pondération, afin de mieux contrôler les biais.