

2025

00

Bases
statistiques
et généralités

Neuchâtel 2025

Statistique sur les télécommunications

Plan de sondage

Rapports de méthodes de la section Méthodes statistiques

Les rapports de méthodes décrivent les méthodes mathématiques et statistiques à la base des résultats et des analyses de la statistique publique. Ils présentent également l'évaluation et le développement de nouvelles méthodes en vue d'une application future. Ces publications visent d'une part à documenter les méthodes utilisées ou envisagées dans un souci de transparence et de rigueur scientifique, et d'autre part à favoriser la collaboration avec le monde scientifique et universitaire.

Les résultats numériques présentés dans les rapports de méthodes illustrent les concepts mathématiques décrits, mais ne sont pas des résultats officiels des enquêtes concernées. De même, les méthodes réellement appliquées peuvent différer légèrement de celles décrites dans ces rapports.

Les rapports de méthodes sont disponibles sous forme électronique sur le site Internet de l'OFS.

Statistique sur les télécommunications

Plan de sondage

Rédaction	Clément Chevalier, OFS
Éditeur	Office fédéral de la statistique (OFS)
	Neuchâtel 2025

Éditeur: Office fédéral de la statistique (OFS)

Renseignements: Clément Chevalier, tél. +41 58 481 00 19
clement.chevalier@bfs.admin.ch

Rédaction: Clément Chevalier, OFS

Contenu: Section méthodes statistiques, OFS

Série: Statistique de la Suisse

Domaine: 00 Bases statistiques et généralités

Langue du texte original: français

Mise en page: Publishing et diffusion PUB, OFS
Ce document a été produit à l'aide de LaTeX.
Il ne répond donc pas aux normes typographiques
des publications de l'OFS.

En ligne: www.statistique.ch

Imprimés: www.statistique.ch
Office fédéral de la statistique, CH-2010 Neuchâtel,
order@bfs.admin.ch, tél. +41 58 463 60 60
Impression réalisée en Suisse

Copyright: OFS, Neuchâtel 2025
La reproduction est autorisée, sauf à des fins commerciales,
si la source est mentionnée.

Numéro OFS: 338-0094

ISBN: 978-3-303-00780-8

Table des matières

Préambule et Résumé	5
1 Introduction	7
2 Données et indicateurs	7
2.1 Données brutes livrées par l'OFCOM	7
2.2 Indicateur de stratification des unités	8
3 Plan de sondage et code R	9
3.1 Stratification et mesures de précision	9
3.2 Présentation du code R	11
4 Résultats préliminaires	12
4.1 Problèmes rencontrés sur un premier jeu de données	12
4.2 Premières expériences sur les données mises à jour	14
5 Résultats définitifs et recommandations	16
5.1 Statistiques sur l'indicateur de stratification	16
5.2 Strates exhaustives et strate négligée	17
5.2.1 Taux de sondage pour la strate 4	17
5.2.2 Taux de sondage des autres strates	19
5.3 Définition retenue pour la strate négligée	20
5.4 Recommandations	22
6 Conclusions et perspectives	24
Bibliographie	25
Rapports de méthodes de la section méthodes statistiques de l'OFS	27

Préambule

L'Office Fédéral de la Communication (OFCOM) interroge chaque année les fournisseurs de services de télécommunications (FST) dans le cadre de la statistique sur les télécommunications. Sur la base des réponses données, l'OFCOM publie un certain nombre de statistiques concernant par exemple les services téléphoniques sur réseaux fixes et mobiles. La statistique sur les télécommunications est une enquête exhaustive sans non-réponse, ce qui signifie que l'ensemble des unités du cadre de sondage participe à l'enquête. La section des méthodes statistiques (METH) de l'Office Fédéral de la Statistique (OFS) a été sollicitée pour étudier la possibilité de remplacer le relevé exhaustif par une enquête par échantillonnage aléatoire. Cette question se pose compte tenu des coûts importants induits par la collecte des données et par le fait qu'un nombre restreint de FST contribue souvent fortement au total de certaines variables d'intérêt. La section METH a effectué des travaux basés sur des données de l'OFCOM allant de 2020 à 2023. Ce rapport décrit les résultats correspondants. La méthodologie de construction du plan se base sur l'historique des réponses passées des FST. Ceci est une approche notablement différente des méthodes basées sur de l'information auxiliaire qui sont utilisées plus souvent à l'OFS. L'auteur remercie les membres de la section METH pour la relecture attentive de ce rapport, Daniel Assoulin, Blandine Bianchi, Daniel Kilchmann, Lionel Qualité et Jean-Pierre Renfer. L'auteur remercie également les collaborateurs de l'OFCOM qui ont relu le rapport et travaillé avec METH à l'élaboration des méthodes, Philippe Hayoz, Patrick Heer, Amalric Oriet et Christine Rohn Hofmann.

Résumé

La statistique sur les télécommunications, publiée chaque année par l'Office Fédéral de la Communication (OFCOM) a fait l'objet d'une étude par la section des méthodes statistiques de l'Office Fédéral de la Statistique (OFS). L'objectif était d'analyser la possibilité de remplacer un relevé exhaustif des unités (fournisseurs de services de télécommunication, FST) par une enquête par échantillonnage aléatoire. Ce rapport décrit les analyses effectuées sur la base de données livrées par l'OFCOM pour les années 2020 à 2023. Force est de constater que pour une forte majorité des variables d'intérêt relevées, un nombre restreint de FST contribue significativement au total de ces variables. Cela a motivé l'étude d'un plan d'échantillonnage dont la particularité est la présence d'une strate négligée, c'est-à-dire échantillonnée avec un taux de sondage de 0%, ce qui peut introduire un biais dans les résultats. Cependant, celui-ci reste limité avec des choix judicieux de la taille de la strate négligée et des FST qui la composent. Les résultats montrent des biais typiquement inférieurs à 5% pour la majorité des variables d'intérêt avec une strate négligée de taille égale à environ à 25% de la taille du cadre de sondage. L'approche proposée n'offre cependant pas de garantie théorique sur le biais futur et l'expertise de l'OFCOM demeure nécessaire pour valider la construction de la strate négligée ainsi que sa mise à jour.

1 Introduction

Dans le cadre de la statistique sur les télécommunications, l'Office Fédéral de la Communication (OFCOM) collecte chaque année, depuis 1998, des données sur les fournisseurs de services de télécommunication (FST). La statistique sur les télécommunications permet à l'OFCOM de publier, entre autres, des informations sur les services téléphoniques sur réseaux fixes et mobiles, sur les infrastructures ou encore sur la diffusion audiovisuelle. Jusqu'en 2024, ces publications se basent sur un relevé exhaustif portant sur environ 500 FST présents en Suisse, à l'aide d'un questionnaire en ligne. Les réponses aux questionnaires permettent de calculer les totaux de plus de 400 variables d'intérêt, dont une centaine sont jugées essentielles par la suite. Les totaux de ces variables sont calculés en sommant les réponses des différents FST et sont publiés sur le site internet de l'OFCOM ¹.

La collecte des données dans la statistique sur les télécommunications a un coût important pour l'OFCOM et génère une charge non-négligeable pour les FST. Or, les données collectées montrent que, pour la grande majorité des variables d'intérêt, un nombre restreint de FST fournissent la quasi-totalité des totaux publiés. Par ailleurs, un certain nombre de FST répondent de manière consistante dans le temps des valeurs très faibles voire nulles et ce pour la totalité ou quasi-totalité des variables d'intérêt. Compte tenu des coûts mentionnés pour la collecte des données de chaque FST et de la charge induite, il est légitime de s'interroger sur la précision qui serait obtenue pour les différentes variables d'intérêt si le relevé exhaustif était remplacé par une enquête non-exhaustive par échantillonnage. Cette question est traitée dans ce rapport sur la base des données fournies par l'OFCOM à l'OFS.

Les données fournies portent sur les enquêtes, exhaustives, des années 2020 à 2023. Une phase exploratoire d'analyse étudie l'usage de plans simples stratifiés classiques, où les FST sont placés dans des strates et où un nombre prédéfini de FST – variable selon la strate – est tiré dans chaque strate. Des simulations basées sur ces données montrent la pertinence d'une stratégie de type "tout ou rien" en termes de taux de sondage, ce qui signifie qu'il est recommandé d'échantillonner certains FST avec un taux de sondage de 100% et d'autres FST avec un taux de sondage de 0%. Une telle approche mène à étudier les potentiels biais induits par l'absence des FST qui ne sont pas du tout échantillonnés.

Ce rapport est organisé comme suit. La section 2 décrit les données livrées par l'OFCOM et présente la construction d'un indicateur qui servira de base à la stratification. La section 3 étudie les plans de sondage simples et stratifiés construits à partir de l'indicateur de stratification et présente des calculs de biais et de variance. Un code implémenté en langage R [R Core Team, 2021] est également décrit dans cette section, avec des résultats préliminaires en section 4. La section 5 présente les résultats et recommandations de l'OFS. Comme expliqué, les résultats démontrent la pertinence d'un plan de sondage stratifié exhaustif sur certaines strates bien choisies et avec un taux de sondage nul sur d'autres strates. Cette approche privilégie l'étude et la limitation du biais par rapport à l'étude de la variance due à l'échantillonnage.

2 Données et indicateurs

2.1 Données brutes livrées par l'OFCOM

L'OFCOM a livré à l'OFS une table de données comportant près d'un million de lignes et cinq colonnes. La table contient les relevés de la statistique sur les télécommunications pour les

1. <https://www.bakom.admin.ch/fr/observatoire-statistique>

années 2020 à 2023, soit 4 ans de données au total. Les colonnes de la table de données sont décrites ci-dessous.

1. Numéro OFCOM, permettant d'identifier un FST de manière unique.
2. Nom de la société, utilisée à titre uniquement informatif. Les analyses et simulations utilisent seulement le numéro OFCOM pour identifier les FST de manière unique.
3. Nom de la variable d'intérêt relevée. Dans les données brutes, plus de 450 variables d'intérêt sont présentes mais les analyses de l'OFCOM ont conduit à se concentrer sur 96 variables d'intérêt jugés "essentiels".
4. Année pour laquelle la variable d'intérêt considérée est relevée : donc ici entre 2020 et 2023 inclus.
5. Réponse relevée pour le FST, l'année et la variable d'intérêt considérée.

Les données ont été livrées dans un fichier au format .Rda, c'est-à-dire importable dans le logiciel R. Les traitements et simulations de l'OFS présentés dans ce rapport ont également utilisé R. Pour faciliter les calculs quatre tables distinctes de données, pour les années 2020 à 2023, ont été créées à partir du fichier de l'OFCOM. Ces tables contiennent un nombre de lignes égal au nombre de FST interrogés dans la statistique sur les télécommunications de l'année a considérée. Ce nombre varie selon l'année. Il est ici compris entre 443 et 539. Les tables créées contiennent le même nombre, $p = 96$, de colonnes, chaque colonne correspondant à une variable d'intérêt importante relevée.

Les analyses de ce rapport vont reposer sur les données de ces 96 variables d'intérêt, mais restent toutes pertinentes dans le cas où ce nombre varie dans le temps. Des recommandations à ce sujet sont données en section 5.4. Pour obtenir une certaine stabilité dans la stratification il est nécessaire que le nombre de variables d'intérêt conserve le même ordre de grandeur d'une année à l'autre.

2.2 Indicateur de stratification des unités

Cette section décrit la construction d'un indicateur de stratification relatif aux FST. Soit a une année donnée pour laquelle des données exhaustives sont disponibles, c'est-à-dire que les réponses sont données pour tous les FST du cadre de sondage de l'année a et pour toutes les variables d'intérêt. L'idée initiale pour remplacer un relevé exhaustif par une enquête par échantillonnage est de classer les FST du cadre de l'année $a + 1$ dans des strates sur la base des réponses obtenues lors de l'année a . Des taux de sondage potentiellement différents seront attribués à ces strates. Le fait de stratifier le cadre de sondage sur la base des réponses à l'enquête de l'année précédente est une approche rarement utilisée à l'OFS. Il est en effet plus fréquent de baser la stratification sur des informations auxiliaires, autres que les réponses relevées, disponibles chaque année sur tout le cadre de sondage. Stratifier en fonction des réponses passées est possible sous certaines conditions. Ceci est discuté en section 5.

Soit N le nombre de FST et p le nombre de variables d'intérêt considérées lors d'une année a . Soient $T_1, \dots, T_p$ les totaux de ces variables d'intérêt pour l'année a . Ces totaux ne sont pas estimés, mais simplement déterminés par la somme des valeurs de tous les FST, puisque le relevé de l'année a est supposé exhaustif et sans non-réponse. Soit $y_{i,j}$ la réponse du FST $1 \leq i \leq N$ pour la variable d'intérêt $1 \leq j \leq p$ et soit $\tilde{y}_{i,j}$ la réponse normalisée obtenue en divisant $y_{i,j}$ par le total correspondant T_j . La quantité $\tilde{y}_{i,j}$ représente la contribution relative du FST i au total T_j de la variable d'intérêt j lors de l'année a . Intuitivement, lors d'une enquête par échantillonnage, la précision de l'estimateur du total de la variable d'intérêt j est bonne si les FST qui contribuent

fortement à cette variable d'intérêt ont des taux de sondage élevés. Il est ainsi judicieux de rassembler dans une même strate les FST qui contribuent fortement aux totaux des variables d'intérêt. L'indicateur de stratification proposé ici, noté z_i^* , dépend du FST i et correspond à l'influence maximale suivante :

$$z_i^* := \max_{j=1,\dots,p} |\tilde{y}_{i,j}|. \quad (1)$$

La valeur absolue dans l'équation (1) vient du fait que des réponses strictement négatives peuvent être relevées pour certaines variables d'intérêt. Ceci se produit pour un nombre de variables d'intérêt compris entre 0 et 5 pour les données de 2021 à 2023. Les réponses strictement négatives demeurent faibles en valeur absolue par rapport au total des variables d'intérêt : inférieures à 0.2% de ce total pour toutes les variables d'intérêt en 2021, 2022 et 2023.

3 Plan de sondage et code R

3.1 Stratification et mesures de précision

La variable z_i^* définie dans l'équation (1) peut servir de base pour établir une stratification en vue de réaliser une enquête par échantillonnage. Dans le cas général, le cadre de sondage est divisé en H strates disjointes, chaque FST faisant partie d'une et une seule strate. Pour attribuer une strate de manière unique à un FST, il est proposé de se baser sur l'indicateur de stratification z_i^* et sur des seuils prédéfinis. Par exemple, des simulations sont présentées en section 4.1 à l'aide des seuils à 0.1%, 1% et 5%. Cela signifie que $H=4$ strates sont utilisées et que les FST sont classés dans les strates 1, 2, 3 et 4 respectivement si z_i^* est dans l'intervalle [5%, 100%], [1%, 5%), [0.1%, 1%) et [0, 0.1%) respectivement. Un choix différent de seuil est bien sûr possible. Des analyses à ce sujet sont présentées en section 5.1 et un choix définitif est proposé en section 5.3.

Soit a une année donnée pour laquelle une enquête par échantillonnage, non-exhaustive, est effectuée. Le nombre de FST dans le cadre de l'année a considérée est noté N , et $N_1, \dots, N_H$ sont les tailles des différentes strates, donc

$$N = N_1 + \dots + N_H. \quad (2)$$

L'idée est de tirer dans la strate h un échantillon de taille fixe n_h , choisi par l'utilisateur, en donnant la même probabilité de tirage, $f_h := n_h/N_h$, aux FST de la strate h . Les tirages dans chaque strate sont indépendants. La taille totale de l'échantillon, notée n , est ainsi donnée par $n = n_1 + \dots + n_H$. En considérant l'une des p variables d'intérêt y_j , le total des réponses sur le cadre de sondage est $T_j := y_{1,j} + \dots + y_{N,j}$. En principe, ce total est inconnu et doit être estimé sur la base de l'échantillon de taille n . L'estimateur classique pour estimer le total des réponses T_j à partir d'un échantillon aléatoire, sans connaître toutes les réponses $y_{1,j}, \dots, y_{N,j}$ et en l'absence de non-réponse est l'estimateur de Horvitz-Thompson [Horvitz & Thompson, 1952]. Celui-ci est obtenu en utilisant la somme pondérée des réponses sur l'échantillon tiré, avec des poids égaux à l'inverse de la probabilité de tirage, c'est-à-dire des poids de N_h/n_h pour les FST qui font partie de la strate h . Ainsi, si $S_1, \dots, S_H$ sont les H sous-échantillons tirés dans chacune des strates, le total T_j est estimé par la quantité

$$\hat{T}_j := \sum_{h=1}^H \frac{N_h}{n_h} \left(\sum_{i \in S_h} y_{i,j} \right). \quad (3)$$

Cette quantité est aléatoire car les sous-échantillons $S_1, \dots, S_H$ sont aléatoires. La variance de l'estimateur $\hat{T}_j$ peut être calculée et vaut :

$$V_j := \sum_{h=1}^H \frac{N_h^2}{n_h} (1 - f_h) D_{h,j}, \quad (4)$$

où $D_{1,j}, \dots, D_{H,j}$ sont les variances des valeurs y_j à l'intérieur des strates $1, \dots, H$, c'est-à-dire

$$D_{h,j} = \frac{1}{N_h - 1} \sum_{i \in h} (y_{i,j} - \bar{y}_{h,j})^2$$

et $\bar{y}_{h,j} := 1/N_h \sum_{i \in h} y_{i,j}$ est la moyenne des valeurs y_j dans la strate h .

L'équation (4) montre que chaque strate h contribue à la variance totale. La contribution de la strate h à la variance peut être nulle si la strate est échantillonnée de manière exhaustive, c'est-à-dire si le taux de sondage f_h est égal à 1. Si $f_h < 1$, la contribution à la variance est faible dans le cas où f_h se rapproche de 1, c'est-à-dire si n_h est proche de N_h , ou bien si les valeurs y_j à l'intérieur de la strate h sont homogènes, de sorte que la variance $D_{h,j}$ soit faible.

Le coefficient de variation (CV) du total T_j est défini comme l'écart-type de $\hat{T}_j$, c'est-à-dire la racine-carrée de la variance (4), divisé par le total T_j :

$$CV_j := \frac{\sqrt{V_j}}{T_j}. \quad (5)$$

Si T_j et V_j sont inconnus, le coefficient de variation peut être estimé en remplaçant T_j par $\hat{T}_j$, voir équation (3), et en estimant la variance V_j à l'aide des variances empiriques des réponses à l'intérieur de chaque strate, voir équation (4). Le CV est une grandeur relative qui peut être exprimée en pour cent. Il est utile pour comparer les précisions lorsque plusieurs variables d'intérêt interviennent.

L'équation (4) s'applique si chaque n_h est *strictement* positif, c'est-à-dire si aucun taux de sondage n'est nul. Si certaines strates h ne sont pas du tout échantillonnées, il faut les retirer du calcul de la variance, comme si la strate n'existait pas, et appliquer l'équation (4) uniquement pour les strates h telles que $n_h > 0$. Les strates avec un taux de sondage nul ne génèrent en effet pas de variance, mais induisent un biais. Le biais total B_j de l'estimateur $\hat{T}_j$ de l'équation (3) est égal à la somme des valeurs à l'intérieur des strates qui ne sont pas échantillonnées :

$$B_j := \sum_{i \in h : n_h = 0} y_{i,j}. \quad (6)$$

En présence de plusieurs variables d'intérêt, le biais relatif $B_{\text{rel},j}$ est une grandeur intéressante, car il facilite les comparaisons. Celui-ci est égal au biais divisé par le total des valeurs sur le cadre de sondage :

$$B_{\text{rel},j} := \frac{B_j}{T_j}. \quad (7)$$

Le fait de ne pas échantillonner du tout certaines strates revient à accepter un biais, dans le but de réduire la variance. Cette idée sera exploitée dans ce rapport, voir section 5.2.1. En pratique, si une enquête par échantillonnage est effectuée et que certaines strates h sont négligées, le biais ne peut pas être observé, puisque les réponses $y_{i,j}$ des strates négligées ne sont pas disponibles. Néanmoins, des calculs et simulations peuvent être faits sur la base de données

exhaustives lorsque celles-ci existent pour des années antérieures, et c'est ce qui est fait dans ce rapport, voir section 5.

La mesure classique de précision qui tient compte à la fois du biais et de la variance est l'erreur quadratique moyenne, ou Mean Squared Error (MSE), et en particulier sa racine-carrée : Root Mean Squared Error, RMSE. La MSE est égal à la variance plus le carré du biais, mais en présence d'un nombre important de variables d'intérêt les grandeurs relatives s'avèrent encore une fois plus appropriées pour comparer les précisions des estimations sur ces différentes variables. La RMSE relative, exprimée en pour cent, s'obtient à partir du CV et du biais relatif par la formule suivante :

$$\text{RMSE}_{\text{rel},j} = \frac{\text{RMSE}_j}{T_j} \quad (8)$$

$$= \frac{\sqrt{V_j + B_j^2}}{T_j} \quad (9)$$

$$= \sqrt{\text{CV}_j^2 + B_{\text{rel},j}^2} \quad (10)$$

La RMSE relative est utilisée dans certaines simulations, voir par exemple sections 4.1 et 5.2.1.

3.2 Présentation du code R

Un code R détaillé et documenté a été livré par l'OFS à l'OFCOM. Les annotations intégrées au code visent à en clarifier la logique et à en faciliter la réutilisation par l'OFCOM.

Compte tenu de la stratégie d'échantillonnage décrite en sections 2.2 et 3.1, la première fonction fournie en R s'appelle `stratif_alldata` et attribue une strate aux FST sur la base de l'indicateur de stratification à partir d'un jeu de données complet pour une année a donnée. La fonction prend en entrée un jeu de données sous la forme d'une matrice $N \times p$, où N est le nombre de FST dans le cadre pour l'année a et p le nombre de variables d'intérêt, et des seuils pour classer les FST dans des strates, par exemple 0.1%, 1% et 5% voir section 3.1. La fonction calcule donc l'indicateur de stratification z_i^* de l'équation (1) et assigne les FST à une strate en fonction de celui-ci et des seuils donnés en entrée. Le résultat de la fonction est un vecteur de taille N qui indique le numéro de strate de chaque FST.

Une autre fonction, `biais_CV_alldata` calcule des biais et CV pour chaque variable d'intérêt en fonction de la stratification renseignée et des taux de sondage dans chaque strate. Les calculs se basent sur des données exhaustives, c'est-à-dire des valeurs disponibles sur tout le cadre de sondage. Ceci n'est pas une situation réaliste en pratique si une enquête non-exhaustive par échantillonnage est réalisée. La fonction est utilisée à l'intérieur des fonctions de simulation décrites ci-dessous.

Deux fonctions, `simu_nextyear_fromalldata` et `simu_nextyear_fromsample`, sont implémentées pour tester différentes stratégies à l'aide des données livrées. La première fonction de simulation, `simu_nextyear_fromalldata`, suppose que des données exhaustives étaient disponibles pour l'année a et qu'une enquête par échantillonnage est effectuée pour l'année $a + 1$ en se basant sur les réponses de l'année a . La fonction offre certaines options supplémentaires discutées en section 5. Dans cette fonction, les FST de l'année $a + 1$ qui existaient déjà lors de l'année a sont mis dans des strates sur la base de leurs réponses de l'année a à l'aide de la fonction `stratif_alldata`. Par défaut, les "nouveaux" FST de l'année $a + 1$, soit les FST présents dans le cadre de l'année $a + 1$ mais pas dans celui de l'année a , reçoivent un taux de

sondage de 100%. Ce paramètre peut être modifié mais cela n'est pas recommandé du fait de l'augmentation du risque de ne pas échantillonner certains FST pour lesquels aucun historique de réponses passées n'est disponible.

La seconde fonction, `simu_nextyear_fromsample`, effectue également des simulations pour une année $a + 1$ donnée mais en supposant cette fois-ci que les données de l'année a étaient déjà des données non-exhaustives, c'est-à-dire issues d'une enquête par échantillonnage. Dans cette seconde fonction, la question de la mise à jour des strates des différents FST se pose. Pour les FST qui existent lors de l'année a mais qui sont absents de l'échantillon, la mise à jour ne peut pas être effectuée facilement. Le choix proposé est de ne pas échantillonner ces FST pour l'année $a + 1$. La question de la mise à jour des strates est abordée en section 5.2.2.

L'objectif de l'ensemble des simulations est d'évaluer la précision des résultats et en particulier de s'assurer d'un maintien de précisions acceptables dans le temps en termes de biais et CV lorsqu'une enquête par échantillonnage remplace un relevé exhaustif. Des résultats préliminaires sur l'utilisation de ces fonctions sont donnés dans la section suivante.

4 Résultats préliminaires

4.1 Problèmes rencontrés sur un premier jeu de données

Dans tous ce rapport, des biais relatifs et CV sont calculés à l'aide la fonction `biais_CV_alldata` et des équations (4) à (7). Dans l'équation (4) de calcul de variance, lorsque le taux de sondage f_h de la strate h varie, les tailles d'échantillon n_h varient aussi. La variance $D_{h,j}$ des valeurs à l'intérieur de la strate h est basée sur toutes les valeurs du cadre de sondage et ne dépend donc pas du taux de sondage. De même, le total T_j au dénominateur des équations (5) et (7) est basé sur toutes les valeurs. Ainsi, les calculs de CV, biais relatifs ou RMSE relatives ne nécessitent jamais d'effectuer des tirages dans les strates qui recevraient des taux de sondage strictement compris entre 0% et 100%.

Des premières simulations ont été effectuées en 2023 à l'aide des données de 2020, 2021, 2022 – les données de 2023 n'ayant pas encore été livrées à cet instant. Ces simulations se basent sur un plan avec 4 strates définies avec des seuils de 0.1%, 1% et 5%. Les strates 1 et 2, c'est-à-dire les strates contenant les FST vérifiant $z_i^* \geq 1\%$, reçoivent un taux de sondage de 100% et sont donc exhaustives. La strate 3, dans laquelle $z_i^* \in [0.1\%, 1\%)$, reçoit un taux de sondage de 50%. La strate 4 contient les FST vérifiant $z_i^* < 0.1\%$ et reçoit un taux de sondage de 0%. Ainsi, il faut considérer le biais et la variance pour évaluer les estimations, en utilisant par exemple la RMSE relative, voir section 3.1, équation (10).

La première simulation considère un ensemble de 399 variables d'intérêt, ce qui est bien supérieur au nombre $p = 96$ mentionné en section 2.1. Dans cette première simulation, les RMSE relatives sont calculés pour une année a donnée et ce sont les réponses de cette même année a qui sont utilisées pour définir les strates. Les résultats obtenus sont ainsi nécessairement "optimistes". Par exemple, pour l'année 2020, les strates 1, 2, 3, 4 contiennent respectivement 112, 114, 97 et 205 unités. Les unités de la strate 4 contribuent au maximum à 0.1% à l'une des 399 variables d'intérêt testées ici, ce qui garantit un biais relatif faible : en théorie limité à 205 fois le seuil de 0.1% mais en pratique nettement moins.

La figure 1 montre les box plots des RMSE relatives des 399 variables d'intérêt obtenues de cette manière pour les données des années 2020, 2021 et 2022. Les RMSE observées ne dépassent jamais 1.5% pour les années 2020 et 2022. Le seuil de 1.5% n'est dépassé que pour

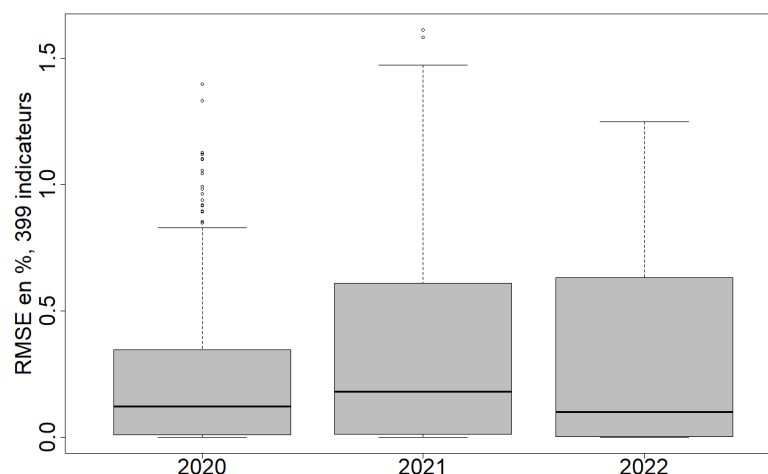


FIGURE 1 RMSE relatives des 399 variables d'intérêt obtenus pour les données des années 2020, 2021, 2022 en stratifiant sur la base de données exhaustives et avec des taux de sondage respectifs de 100%, 100%, 50%, 0% pour les strates 1, 2, 3, 4.

2 variables d'intérêt sur 399 pour l'année 2021. Ces premiers résultats semblent très encourageants, d'autant plus que la stratification utilisée permet d'obtenir des tailles d'échantillons relativement faibles compte tenu des contraintes de l'OFCOM : la taille n de l'échantillon avec la stratification choisie représente respectivement 52%, 50% et 57% de la taille N du cadre de sondage pour les années 2020, 2021 et 2022. Ceci est surtout dû à la strate 4 qui représente une fraction importante du cadre de sondage et pour laquelle un taux de sondage de 0% est utilisé (strate négligée).

Les premières difficultés liées à la stratégie appliquée sont observées lors d'une utilisation simple de la fonction de simulation `simu_nextyear_fromalldata`, qui simule une situation plus réaliste. Dans ce contexte, les données d'une année a sont utilisées pour réaliser une enquête lors de l'année $a + 1$, et pas lors de l'année a . Ainsi, si un FST est classé en strate 4 ($z_i^* < 0.1\%$) sur la base des réponses de l'année a , cela ne garantit aucunement que celui-ci ne puisse pas répondre, pour une ou plusieurs variables d'intérêt, des valeurs bien supérieures à 0.1% du total pour l'année $a + 1$.

La figure 2 montre les RMSE relatives obtenus dans ce contexte, calculés à l'aide des équations (4) à (7). Sur les 399 variables d'intérêt, 15 ont une RMSE relative supérieure à 5% pour l'année 2021 et 17 pour l'année 2022. De plus, la RMSE relative est supérieure à 20% pour 8 variables d'intérêt pour l'année 2021 et 4 pour l'année 2022. Ces résultats, relativement peu satisfaisants, s'expliquent par plusieurs facteurs et deux d'entre eux sont détaillés ci-dessous :

1. À la suite de discussions avec l'OFCOM, il s'avère que les valeurs renseignées de certaines variables d'intérêt étaient partiellement inexactes et/ou que certaines variables d'intérêt ont une importance finalement modérée. Ainsi, les $p = 96$ variables d'intérêt les plus importantes, sur lesquelles les analyses suivantes de ce rapport se basent, ont pu être identifiées par l'OFCOM.
2. Certains FST sont créés lors d'une année a donnée et font donc partie du cadre de sondage de l'enquête de l'année a . Cependant, certains ne démarrent une activité importante que lors de l'année suivante, $a + 1$. Dans ce contexte, les réponses à l'enquête de l'année a ne reflètent pas bien l'importance relative du FST et ne constituent pas forcément une bonne prédiction de l'activité du FST pour l'année $a + 1$.

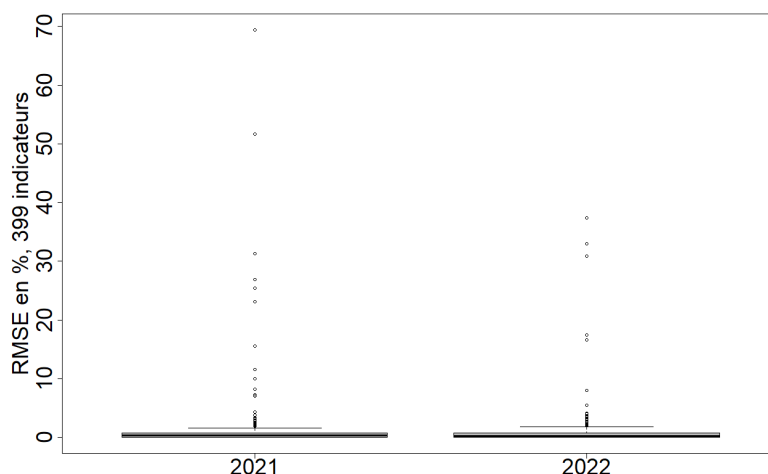


FIGURE 2 RMSE relatives des 399 variables d'intérêt obtenus pour les années 2021, 2022 en stratifiant à l'aide des données exhaustives de 2020 et 2021 respectivement, et avec des taux de sondage respectifs de 100%, 100%, 50%, 0% pour les strates 1, 2, 3, 4.

Les résultats préliminaires présentés dans la section suivante prennent en compte le point 1. ci-dessus mais pas encore le point 2. Des résultats consolidés, basés sur un traitement des deux points ci-dessus, sont donnés en section 5. En particulier, le point 2. ci-dessus est étudié en section 5.3.

4.2 Premières expériences sur les données mises à jour

Les simulations présentées dans la suite de ce rapport ont été effectuées à l'aide des données de 2020, 2021, 2022 et également 2023, nouvellement disponibles. De plus, les $p = 96$ variables d'intérêt jugées importantes, mentionnées en section 4.1, ont été utilisées pour toutes les analyses qui suivent. Des RMSE relatives ont été calculés en stratifiant selon la même méthode que précédemment et en utilisant des taux de sondage de respectivement 100%, 100%, 100%, 0% pour les strates 1, 2, 3 et 4. Par rapport aux simulations précédentes, la strate 4 continue de ne pas être échantillonnée du tout. Ce point sera justifié davantage en section 5.2.1. De plus, le taux de sondage sur la strate 3 passe de 50% dans les simulations de la section 4.1 à 100% et ceci pour des raisons détaillées en section 5.2.2. Dans ces conditions, la variance est nulle et la RMSE relative est égal au biais relatif, voir équation (10).

Les simulations à l'aide de la fonction `simu_nextyear_fromalldata` calculent des RMSE relatives pour l'année $a + 1$ avec une stratification basée sur des données supposées exhaustives de l'année a .

Les résultats détaillés en figure 3 sont plus encourageants. La figure 3 montre que les biais relatifs restent contenus, typiquement inférieur à 5%, sauf pour une variable d'intérêt lors de l'année 2023 où le biais dépasse 8%. Des analyses ont montré que le biais en 2023 est lié à un FST classé en strate 4 sur la base des réponses en 2022 qui "répondrait" à lui seul une valeur correspondant à un peu plus de 8% du total pour une certaine variable d'intérêt. Ce FST existe déjà en 2020. Pour les simulations de 2021, basées sur les réponses de 2020, il est classé en strate 2 et donc échantillonné. Pour les simulations de 2022, basées sur les réponses de 2021, il est également classé en strate 2. Ce n'est que pour les simulations de 2023 que le FST est placé en strate 4, causant un biais important. Cette expérience montre qu'il est probablement

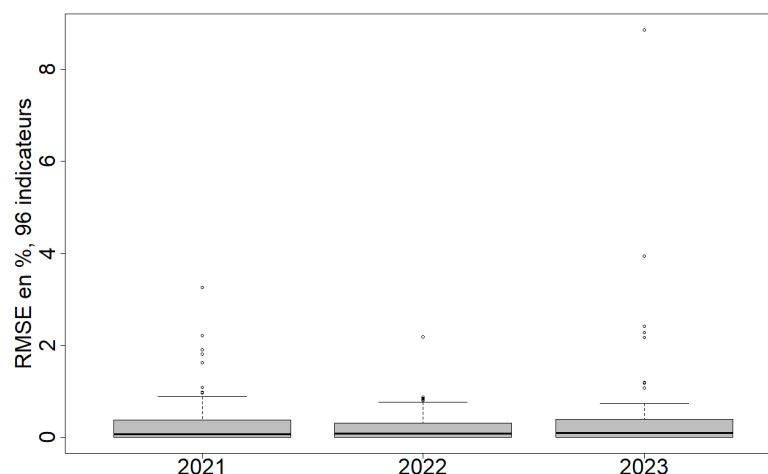


FIGURE 3 RMSE relative de $p = 96$ variables d'intérêt pour les années 2021, 2022, 2023 en stratifiant à l'aide des données exhaustives de 2020, 2021 et 2022 respectivement, et avec des taux de sondage respectifs de 100%, 100%, 100%, 0% pour les strates 1, 2, 3, 4.

nécessaire d'être plus prudent avant de classer un FST en strate 4. Ceci peut être fait de plusieurs manières. Il est par exemple possible de rabaisser le seuil de 0.1% sur l'indicateur de stratification z_i^* décrit en section 2.2. Il est également possible de se baser sur les réponses à l'enquête sur un historique plus long qu'un an. Ce point est étudié en section 5.3.

Les soucis observés dans ces résultats se retrouvent en simulant davantage dans le futur, c'est-à-dire en utilisant la fonction `Rsimu_nextyear_fromsample` qui calcule des résultats pour l'année $a + 1$ sur la base de données *non-exhaustives* de l'année a . Ces données non-exhaustives correspondent à un échantillon qui a lui même été obtenu en stratifiant sur la base de données complètes de l'année $a - 1$. La figure 4 présente les RMSE relatives obtenues pour des simula-

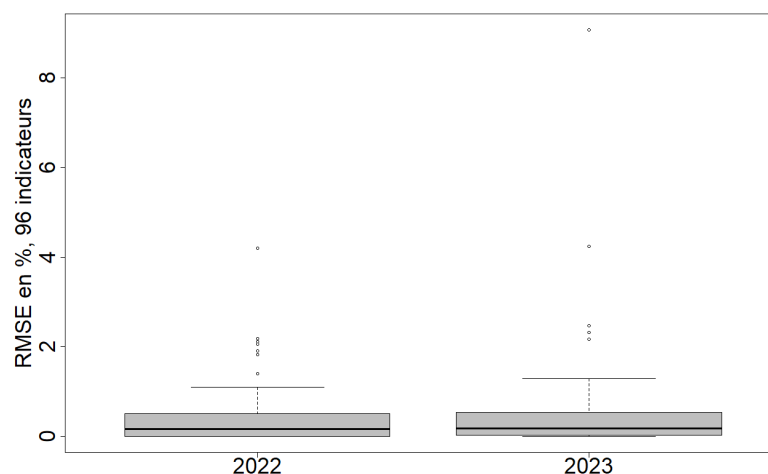


FIGURE 4 RMSE relatives de $p = 96$ variables d'intérêt pour les années 2022 et 2023 en stratifiant sur la base des réponses d'un échantillon pour les années 2021 et 2022.

tions d'enquêtes de 2022 et 2023. L'enquête de 2022 stratifie sur la base des réponses en 2021 d'un échantillon de FST. Les FST classés en strate 4 en 2021, sur la base de réponse en 2020, sont automatiquement placés en strate 4 pour l'année 2022. Les autres FST en strate 4 pour l'année 2022 sont les FST présents dans l'échantillon de 2021 et dont indicateur de stratification

z_i^* est strictement inférieur à 0.1%. L'enquête de 2023 fonctionne selon le même principe.

Il apparaît que, dans le cadre d'enquêtes répétées par échantillonnage, la strate 4 tend à prendre de l'importance avec le temps. En effet, les FST classés en strate 4 sur la base de leur réponse en année a ne sont pas échantillonnés pour l'année $a + 1$ et pas non plus pour les années suivantes, $a + 2, a + 3$, etc. Les FST classés dans les strates exhaustives 1, 2 ou 3 sur la base de leur réponse en année a sont certes échantillonnés pour l'année $a + 1$. Cependant ces FST peuvent tomber dans la strate 4 pour l'année suivante, $a + 2$, si leur réponse lors de l'année $a + 1$ est faible par rapport aux réponses des autres FST des strates 1, 2 et 3. La strate 4 dans sa définition actuelle est ainsi une strate dans laquelle on peut entrer, mais pas sortir. Cette progression de la strate 4 dans le temps explique les résultats de la figure 4 légèrement moins bons que ceux de la figure 3. Dans les simulations de la figure 3, le nombre de variables d'intérêt avec une RMSE relative supérieure à 1% est respectivement de 6, 1 et 9 pour les années 2021, 2022 et 2023. Dans les simulations de la figure 4, ce nombre est respectivement de 11 et 12 pour les années 2022 et 2023. La hausse est due à un plus grand nombre de FST en strate 4. Pour l'année 2022, le pourcentage de FST en strate 4 passe de 40.9% avec la fonction `simu_nextyear_fromalldata` à 46.1% avec la fonction `simu_nextyear_fromsample`. Pour l'année 2023 il passe de 33.8% à 37.0%.

Diverses approches sont explorées pour traiter la problématique liée à la progression de la strate 4. Elles sont détaillées dans la section suivante, tandis que des recommandations opérationnelles figurent en section 5.4.

5 Résultats définitifs et recommandations

5.1 Statistiques sur l'indicateur de stratification

Les résultats présentés s'appuient sur les données livrées par l'OFCOM pour les années 2020 à 2023, pour $p = 96$ variables d'intérêt jugées importantes. La table 1 donne des statistiques sur l'indicateur de stratification z_i^* de l'équation (1). Ces statistiques sont obtenues en utilisant la fonction `stratif_alldata` présentée en section 3.2. Elles permettent de quantifier la proportion de FST dont les réponses sont "faibles" pour chacune des p variables d'intérêt. Les statistiques montrent que chaque année, entre 12% et 16% des FST répondent 0 pour l'ensemble des p variables d'intérêt. Ces FST sont nommés "all zeros". Par ailleurs, la proportion de FST dont les p réponses sont toutes inférieures à 0.01% du total de la variable d'intérêt varie de 25% à 31% environ. Cette proportion augmente si le seuil de 0.01% est relevé. Par exemple la proportion monte à environ la moitié des FST si le seuil est de 0.1%, ce qui est le seuil utilisé pour les résultats préliminaires de la section 4.1.

La table 1 explique le choix initial du seuil de 0.1% pour définir la strate 4 qui reçoit un taux de sondage nul dans les simulations décrites en sections 4.1 et 4.2. Dans les simulations avec la fonction `simu_nextyear_fromalldata` de la section 4.2, la strate 4 représente respectivement 40.8%, 40.9% et 33.8% de la taille du cadre de sondage pour les années 2021, 2022 et 2023 respectivement. Ces chiffres sont légèrement différents de ceux de la table 1 car les simulations de l'année $a + 1$ se basent sur les réponses de l'année a avec un cadre de sondage qui varie, et en particulier des nouveaux FST pour l'année $a + 1$ qui sont échantillonnés de manière exhaustive dans ces simulations. Ici, "l'économie" réalisée sur le nombre de FST dans l'échantillon, supérieure à 30% dans ces simulations, est considérée comme suffisante par l'OFCOM. Il serait même acceptable d'avoir une économie plus faible si les résultats des simulations peuvent être améliorés.

TABLE 1 Nombre N de FST dans le cadre de sondage pour chaque année, proportion de FST dont toutes les réponses sont nulles (all zeros) et proportions de FST dont l'indicateur de stratification z_i^* est inférieur à des seuils donnés.

	2020	2021	2022	2023
N	528	539	443	473
all zeros	13.4%	16.0%	12.2%	13.3%
$z_i^* < 0.01\%$	29.9%	30.8%	25.1%	25.6%
$z_i^* < 0.05\%$	44.1%	44.9%	37.5%	37.8%
$z_i^* < 0.1\%$	49.6%	51.9%	44.9%	45.2%
$z_i^* < 0.5\%$	67.8%	68.5%	64.3%	67.9%
$z_i^* < 1\%$	75.9%	76.6%	73.8%	74.6%

Une amélioration des résultats peut potentiellement être obtenue en choisissant un seuil plus bas que 0.1% pour définir la strate 4 et/ou en définissant des conditions plus strictes pour classer un FST en strate 4. Ce point est discuté en section 5.3. Avant cela, la section suivante est consacrée aux options méthodologiques liées au choix du taux de sondage pour la strate 4 et en particulier de savoir s'il vaut mieux utiliser un taux de sondage faible ou nul.

5.2 Strates exhaustives et strate négligée

5.2.1 Taux de sondage pour la strate 4

Lors des simulations des sections 4.1 et 4.2, le choix du taux de sondage à appliquer à la strate 4 a été étudiée. Les deux options envisageables sont d'utiliser soit un taux de sondage jugé "faible", soit un taux de sondage nul. Un taux de sondage faible, mais strictement positif, donne des estimations sans biais et dans ce contexte seule la variance est à évaluer. Un taux de sondage nul ne génère pas de variance mais génère du biais. Les deux options peuvent être comparées entre elles en considérant la RMSE relative, voir équation (10). Par exemple, la figure 3 en section 4.2 donnait les RMSE obtenus pour les années 2021 à 2023 en utilisant la fonction `simu_nextyear_fromalldata` avec un taux de sondage de 0% pour la strate 4 et 100% pour les autres strates. La figure 5 fournit à nouveau ces RMSE relatives et ajoute celles obtenues en utilisant un taux de sondage pour la strate 4 de 5% (boîtes rouges) et 10% (boîtes oranges).

Sur la figure 5, l'axe des ordonnées est tronqué pour améliorer la visibilité. En 2021, la RMSE relative dépasse 10% pour une variable d'intérêt si le taux de sondage de la strate 4 est de 5%. Le seuil de 10% n'est pas dépassé dans les deux autres cas analysés. En 2023, la RMSE relative dépasse 10% pour deux variables d'intérêt si le taux de sondage de la strate 4 est de 5% ou 10% et il n'y a pas de dépassement si ce taux est de 0%. Des statistiques sur les RMSE relatives obtenues sont données en table 2. Globalement, les RMSE relatives obtenues sont meilleures avec un taux de sondage de 0% sur la strate 4 par rapport à 5% et 10%. Les RMSE relatives obtenues pourraient devenir meilleures si le taux de sondage de la strate 4 dépassait 10%, mais un tel taux poserait la question du gain de ressources obtenu. Ces résultats poussent à choisir l'option "0%" pour le taux de sondage de la strate 4, ce qui revient à négliger cette strate et à accepter un biais qui sera modéré si la strate 4 est convenablement construite.

Deux autres arguments peuvent être formulés en faveur du taux de sondage à 0% par rapport à l'option 5% ou 10%. Premièrement, un taux de sondage non-nul pose la question du tirage de l'échantillon et en particulier de la coordination des tirages. Cet argument est repris et détaillé en section 5.2.2. Deuxièmement, un taux de sondage faible fournit des échantillons de très petite

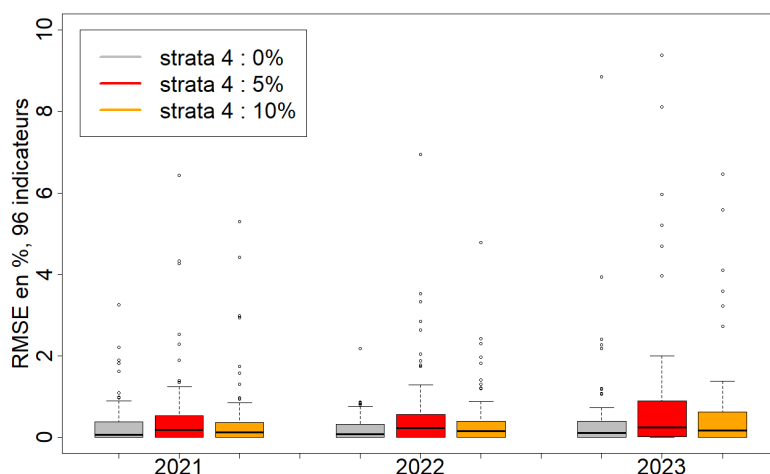


FIGURE 5 RMSE relatives de $p = 96$ variables d'intérêt obtenus pour les années 2021, 2022, 2023 en stratifiant à l'aide des données exhaustives de 2020, 2021 et 2022 respectivement, et avec des taux de sondage exhaustifs pour les strates 1, 2, 3 et un taux de sondage de 0% (boîtes grises), 5% (boîtes rouges) et 10% (boîtes oranges) pour la strate 4.

TABLE 2 Statistiques sur les simulations de la figure 5. L'un des paramètres est le taux de sondage de la strate 4. Les moyennes, médianes, maximum, portent sur les 96 RMSE relatives obtenues, pour une année donnée et un taux de sondage fixé en strate 4.

	strate 4 : 0%			strate 4 : 5%			strate 4 : 10%		
	2021	2022	2023	2021	2022	2023	2021	2022	2023
N	539	443	473	539	443	473	539	443	473
n	319	262	313	330	271	321	341	280	329
n / N (%)	59.2	59.1	66.1	61.2	61.2	67.9	63.3	63.2	69.6
n strate 4	0	0	0	11	9	8	22	18	16
moyenne $RMSE_{rel}$ (%)	0.29	0.21	0.40	0.71	0.51	1.35	0.49	0.35	0.93
médiane $RMSE_{rel}$ (%)	0.07	0.08	0.10	0.18	0.22	0.24	0.13	0.15	0.17
max $RMSE_{rel}$ (%)	3.26	2.18	8.85	14.20	6.94	38.55	9.77	4.78	26.53
nb. $RMSE_{rel} > 1\%$	6	1	9	14	12	24	8	9	14
nb. $RMSE_{rel} > 3\%$	1	0	2	5	3	8	3	1	7
nb. $RMSE_{rel} > 5\%$	0	0	1	3	1	6	2	0	4

taille et ne permet pas d'éliminer les problèmes de biais dans le cas où les réponses d'un ou plusieurs FST de la strate 4 sont plus importantes que prévu. En effet, si le taux de sondage en strate 4 vaut par exemple 5%, alors le poids de sondage des FST tirés dans cette strate devrait être égal à 20, voir section 3.1. Or, il est peu réaliste en pratique de multiplier par 20 une réponse qui serait nettement plus élevée que prévue – par exemple qui serait supérieure à 1% du total. Dans ce cas, le poids des FST en strate 4 serait ainsi de 20 si les réponses sont proches 0 et une robustification des poids aurait lieu dans le cas contraire. La robustification des poids, voir par exemple Beaumont et al. [2013], permettrait de limiter la variance de l'estimation, mais génère du biais. Ceci est un peu contradictoire avec le principal bénéfice obtenu en utilisant un taux de sondage non-nul qui est justement l'absence de biais.

Au final, compte tenu des résultats montrés ici et des arguments avancés, il semble préférable d'accepter le biais lié à un taux de sondage nul dans la strate 4. Néanmoins, le problème de la

progression de la strate 4 soulevé en section 4.2 demeure. Celui-ci est traité en section 5.4.

5.2.2 Taux de sondage des autres strates

Les simulations illustrées dans les sections 4.1 et 4.2 utilisent un taux de sondage de respectivement 50% et 100% pour la strate 3 (indicateur de stratification z_i^* compris entre 0.1% et 1%). Ce taux de sondage est par la suite maintenu à 100%. L'objectif de cette section est de discuter et justifier le choix d'un recensement de la strate 3. Une des première raisons est liée à la taille de l'échantillon : le fait de ne pas échantillonner du tout la strate 4 permet déjà d'atteindre des économies jugées satisfaisantes concernant la taille de l'échantillon. Ceci n'est cependant pas la seule raison.

D'une part, utiliser un taux de sondage dans l'intervalle $(0, 1)$, donc strictement supérieur à 0% et strictement inférieur à 100%, soulève la question du tirage des unités et en particulier de la coordination des tirages d'une année à l'autre. L'OFS dispose d'un système de coordination de ses échantillons, décrit dans Qualité [2009] et Graf & Qualité [2014], qui permet d'introduire une dépendance temporelle entre les échantillons de plusieurs enquêtes de l'OFS d'une année à l'autre. Par exemple, il est possible d'introduire une coordination "négative" d'une année à l'autre afin de renouveler l'échantillon le plus possible. Il est possible au contraire d'utiliser une coordination "positive" afin de garder le plus possible les mêmes unités dans l'échantillon et d'augmenter ainsi l'aspect panel de l'échantillon. Il est également possible de combiner les deux approches en divisant l'échantillon en sous-échantillons et en appliquant une coordination positive à certains sous-échantillons et négative à d'autres. Un tel algorithme de coordination des échantillons serait souhaitable pour l'OFCOM mais son implémentation nécessite un effort de développement conséquent.

Une autre question importante liée à l'utilisation de taux de sondage dans l'intervalle $(0, 1)$ est celle de la mise à jour des strates d'une année à l'autre. Dans un contexte où seuls des taux de sondage de 0% et 100% sont utilisés, cette mise à jour est relativement aisée. En effet :

- les FST qui étaient précédemment en strate 4, donc pas échantillonnés l'année précédente, restent en strate 4 et ne seront jamais échantillonnés ;
- les FST échantillonnés l'année précédente, donc dans une strate exhaustive, sont classés en fonction de leur réponse et éventuellement d'autres paramètres telles que les réponses plus loin dans le passé. Selon tous ces critères, ils sont classés en strate 4 ou non.

Si des taux de sondage différents de 0% et 100% sont utilisés, les calculs de variance de l'équation (5) devraient tenir compte du fait que la stratification dépend d'un tirage aléatoire du passé, autrement dit qui se produit avec une probabilité différente de 0% ou 100%. Ce souci pourrait être réglé en ne mettant jamais à jour la strates des FST qui font partie de strates dont le taux de sondage est strictement compris entre 0% et 100%. Une telle contrainte semble cependant peu applicable en pratique puisqu'elle interdirait de déplacer les FST qui deviennent importants dans des strates exhaustives, et qu'elle interdirait également de déplacer les FST qui répondent des valeurs faibles ou nulles dans la strate négligée. Compte tenu des points soulevés ci-dessus et du fait que les tailles d'échantillon sont suffisamment faibles, il est recommandé de n'utiliser que des taux de sondage de 0% et 100%. Dévier de cette recommandation est possible, mais nécessite un travail supplémentaire pour traiter ces points. Il est aussi important de mentionner qu'avec la méthode choisie, les estimations sont très simples : il s'agit de la somme des valeurs observées et la variance est absente, ce qui est aussi vrai pour les évolutions entre deux années.

5.3 Définition retenue pour la strate négligée

Cette section traite de la construction retenue pour la strate “négligée”, qui est ici la strate 4, qui reçoit un taux de sondage de 0%. La section 5.1 évoquait deux leviers différents dans la construction de cette strate. D’une part, le seuil pour l’indicateur de stratification z_i^* peut être choisi. Ce seuil était fixé à 0.1% dans les sections précédentes, ce qui signifie que les FST pour lesquelles $z_i^* < 0.1\%$ étaient classés en strate 4, la strate négligée. D’autre part, il est possible de ne pas stratifier sur la base des données d’une seule année, mais sur 2 ans par exemple. Dans le détail et en appliquant ces deux leviers, les FST classés en strate 4 seraient les unités qui satisfont toutes les conditions suivantes lors de l’année a :

- être présent dans le cadre de sondage de l’année $a - 1$ et dans le cadre de sondage de l’année $a - 2$, de sorte que les réponses collectées dans le passé soient disponibles pour une durée d’au moins 2 ans,
- sur la base des réponses à ces deux enquêtes, avoir un indicateur z_i^* inférieur à un seuil, à définir, à la fois pour l’année $a - 1$ et $a - 2$.

Comme des données sont disponibles de 2020 à 2023 inclus, les simulations suivantes ont été effectuées :

1. Pour l’année 2022, des simulations sont effectuées en se basant sur des données exhaustives de 2020 et 2021. Ces données exhaustives permettent de construire la strate négligée pour 2022 selon les nouvelles conditions décrites ci-dessus.
2. Pour l’année 2023, il est supposé que seules des données partielles pour 2022 sont disponibles. Cela signifie que les FST qui n’ont pas été échantillonnés en 2022, car en strate négligée, ne le sont pas non plus en 2023. La strate négligée pour 2023 est ainsi formée de l’union entre les FST non-échantillonnés en 2022 et les FST qui satisfont les conditions énumérées ci-dessus pour l’année 2023

La figure 6 présente le résultat des simulations pour 3 seuils distincts sur l’indicateur z_i^* . Ces seuils sont de 0.1% (boîtes grises), 0.05% (boîtes rouges) et 0.01% (boîtes oranges).

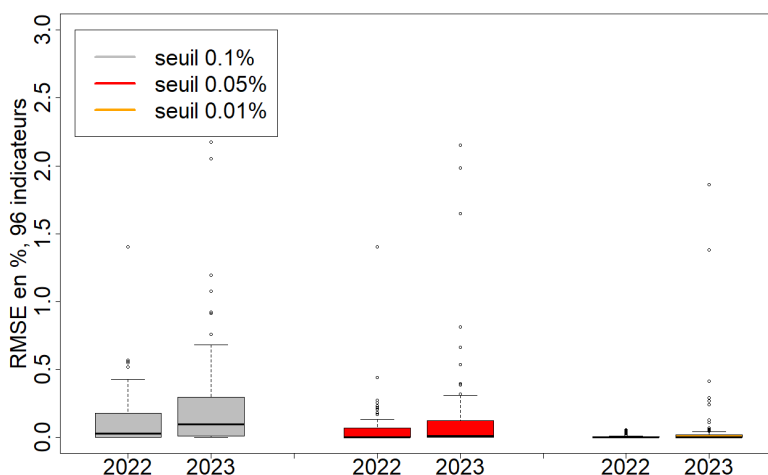


FIGURE 6 RMSE relatives de $p = 96$ variables d’intérêt obtenues pour les années 2022 et 2023 en construisant la strate négligée à l’aide de deux ans d’historique de réponse et des seuils de 0.1% (boîtes grises), 0.05% (boîtes rouges) et 0.01% (boîtes oranges).

Les résultats tendent à s’améliorer lorsque le seuil choisi diminue. Ce point est intuitif puisqu’un seuil plus bas conduit classer moins de FST dans la strate 4. L’amélioration des résultats ob-

tenue en abaissant le seuil doit être rapportée au coût en nombre de FST dans l'échantillon. Des statistiques sur les simulations de la figure 6 sont données en table 3. Celles-ci indiquent notamment, dans chaque cas, la taille n de l'échantillon.

TABLE 3 Statistiques sur les simulations de la figure 6. Les moyennes, médianes, maximum, portent sur les 96 RMSE relatives obtenues, pour les années 2022 ou 2023 et pour 3 différents seuils sur l'indicateur de stratification.

	seuil 0.1%		seuil 0.05%		seuil 0.01%	
	2022	2023	2022	2023	2022	2023
N	443	473	443	473	443	473
n	311	331	337	359	389	413
n/N (%)	70.2	70.0	76.1	75.9	87.8	87.3
taille strate 4	132	142	106	114	54	60
taille strate 4 / N (%)	29.8	30.0	23.9	24.1	12.2	12.7
moyenne $RMSE_{rel}$ (%)	0.12	0.29	0.06	0.18	0.004	0.1
médiane $RMSE_{rel}$ (%)	0.03	0.10	0	0.01	0	0
max $RMSE_{rel}$ (%)	1.40	3.94	1.40	3.94	0.06	3.94
nb. $RMSE_{rel} > 0.1\%$	36	48	15	27	0	9
nb. $RMSE_{rel} > 1\%$	1	6	1	4	0	3
nb. $RMSE_{rel} > 3\%$	0	1	0	1	0	1

La table 3 montre que les RMSE relatives ont tendance à être moins bonnes en 2023 par rapport à 2022. Ceci est dû à une situation asymétrique. Les simulations pour 2022 se basent sur des données exhaustives de 2020 et 2021, alors que les simulations de 2023 se basent sur les décisions prises en 2022 et donc sur des données partielles, en plus des données exhaustives de 2021. La table 3 montre aussi que le choix d'un seuil à 0.1% pour l'indicateur z_i^* permet des économies substantielles, puisque environ 30% des FST ne sont pas interrogés. Les RMSE relatives sont en moyenne inférieures à 1% et sont supérieures à 1% pour 6 variables d'intérêt lors de l'enquête 2023. L'abaissement du seuil à 0.05% pour l'indicateur z_i^* réduit la taille de la strate négligée à environ 24% de la taille du cadre de sondage et améliore la RMSE relative moyenne. Le nombre de variables d'intérêt avec une RMSE relative supérieure à 1% descend à 4. Enfin, un seuil encore plus bas à 0.01% réduit considérablement la taille de la strate négligée, à environ 12% des FST. Les RMSE relatives sont faibles, mais dépassent tout de même 1% pour 3 variables d'intérêt en 2023.

À noter que la RMSE relative la plus élevée, 3.94%, ne parvient pas à être réduite en diminuant le seuil. La RMSE relative est ici égale au biais, et après investigation ce biais se produit sur une variable d'intérêt à cause d'un seul FST. Ce FST est présent dans les cadres de sondages 2020 à 2023. Pour 2020, 2021 et 2022, l'indicateur de stratification z_i^* de ce FST vaut respectivement 0.0019%, 0.0016% et 0.0016% : donc des valeurs inférieures aux 3 seuils testés ici. Ce FST répond en 2023, sur une variable d'intérêt, une valeur égale à 3.94% du total, après avoir répondu 0 les 3 années précédentes sur cette même variable d'intérêt. Comme ce FST est classé dans la strate négligée en 2022 et 2023, un biais apparaît en 2023 pour la variable d'intérêt concernée. Cet exemple montre que les deux leviers utilisés pour améliorer les RMSE relatives, c'est-à-dire le seuil utilisé et l'historique de réponse de deux ans, ne peuvent éviter complètement des cas avec un biais élevé. Pour limiter davantage ce risque, des recommandations sont émises en section 5.4 avec en particulier une recommandation d'effectuer périodiquement une enquête exhaustive.

Il apparaît au final que les résultats demeurent satisfaisants si une économie d'environ 25% sur la taille de l'échantillon est visée, c'est à dire si le ratio n/N est égal à 75% environ. Le biais relatif dépasse rarement les 1% (pour moins de 5% des variables d'intérêt) et ne dépasse jamais 5% en simulation. Par ailleurs, les résultats sont meilleurs que ceux donnés en table 2, ce qui montre l'intérêt de construire la strate négligée sur la base d'un historique de réponses de deux années au lieu d'une. Pour mieux analyser l'effet précis d'un historique de deux ans de réponse, la table 4 présente les résultats de simulations identiques à celles de la table 3 mais en utilisant seulement un an d'historique de réponses au lieu de deux pour classer dans la strate négligée. Avec un seuil de 0.05%, le fait d'utiliser une seule année d'historique de réponse donne une strate négligée qui contient plus de 31% des FST, contre environ 24% si deux années d'historique sont prises en compte. Les RMSE relatives obtenues passent presque du simple au double en moyenne.

TABLE 4 Statistiques sur des simulations où la strate 4 est construite sur la base d'une seule année d'historique de réponse et non deux. Les moyennes, médianes, maximum, portent sur les 96 RMSE relatives obtenues, pour les années 2022 ou 2023 et pour 3 différents seuils sur l'indicateur de stratification.

	seuil 0.1%		seuil 0.05%		seuil 0.01%	
	2022	2023	2022	2023	2022	2023
N	443	473	443	473	443	473
n	262	298	294	324	355	379
n/N (%)	59.1	63.0	66.4	68.5	80.1	80.1
taille strate 4	181	175	149	149	88	94
taille strate 4 / N (%)	40.9	37.0	33.6	31.5	19.9	19.9
moyenne $RMSE_{rel}$ (%)	0.21	0.49	0.11	0.29	0.03	0.18
médiane $RMSE_{rel}$ (%)	0.08	0.18	0	0.07	0	0.01
max $RMSE_{rel}$ (%)	2.18	9.07	2.18	4.25	0.81	3.94
nb. $RMSE_{rel} > 0.1\%$	46	55	27	44	6	20
nb. $RMSE_{rel} > 1\%$	1	12	1	7	0	5
nb. $RMSE_{rel} > 3\%$	0	2	0	1	0	1

5.4 Recommandations

Les simulations effectuées dans ce rapport et les méthodes recommandées au final ne donnent pas de garanties formelles sur le biais. Celui-ci pourrait parfaitement dépasser les 5% pour certaines variables d'intérêt, même si cela ne se produit pas lors des simulations illustrées dans la section 5.3, sur le jeu de données livré par l'OFCOM. Par ailleurs, il est important de réaliser qu'en pratique le biais ne peut pas être directement observé si l'enquête n'est pas exhaustive. L'ensemble des calculs de biais effectués ici se basent sur des données exhaustives qui ne seront pas disponibles en pratique. Les simulations effectuées dans ce rapport servent ainsi seulement à donner un ordre de grandeur sur le biais qui pourrait être subi. Les recommandations qui suivent sont aussi utiles pour limiter les risques de biais importants.

Concernant les données livrées par l'OFCOM, il serait intéressant de "poursuivre" les simulations avec les données de 2024, en construisant la strate négligée sur la base des simulations de 2023 et en continuant d'exclure les FST déjà exclus en 2023. En effet, comme expliqué dans les sections précédentes, il n'est pas possible pour un FST de "sortir" de la strate négligée : si le FST est dans cette strate pour l'année a , il le sera automatiquement en année $a + 1$. Cette règle

d'impossibilité de sortie de la strate négligée est bien sûr arbitraire et l'OFCOM doit pouvoir en dévier s'il estime cela opportun. Deux recommandations peuvent être effectuées en ce sens.

D'une part la strate négligée peut être construite sur la base des données, calculs et règles présentés dans ce rapport avec notamment la règle de deux années consécutives où l'indicateur de stratification z_i^* est inférieur à un seuil qui pourrait être fixé à 0.05%. Cependant, il est important que l'OFCOM analyse les résultats de ces calculs et décide au cas par cas du maintien ou non de chaque FST dans la strate négligée ainsi construite. L'OFCOM pourrait en effet disposer d'informations auxiliaires qui justifient la décision de sortir un FST de la strate négligée. D'une manière générale, les calculs présentés devraient servir uniquement à la construction d'une première version de la strate négligée. Cette première version devrait être analysée par l'OFCOM et non acceptée automatiquement. Par ailleurs, la construction du seuil doit bénéficier de toute l'expertise de l'OFCOM et de sa connaissance des FST.

D'autre part, en plus du travail d'analyse individuel sur les FST, il serait possible de diminuer les risques de biais en choisissant de réaliser une enquête exhaustive avec une certaine périodicité, par exemple tous les 5 ans. Avec les données livrées, il n'est pas possible d'émettre facilement des recommandations sur la périodicité exacte des enquêtes exhaustives. Des recommandations plus précises sur cette durée pourraient être faites avec la disponibilité de données pour 2024 et 2025. De bons résultats en simulation avec ces données mises à jour ainsi qu'une bonne application de la recommandation précédente, notamment les contrôles sur la strate 4, devraient permettre d'utiliser une périodicité de 5 ans avec des risques modérés sur le biais et des biais typiquement inférieurs à 5% sur la totalité, ou au moins la très grande majorité des variables d'intérêt importantes.

Lorsqu'une enquête exhaustive est effectuée périodiquement, disons en année $a, a + 5, a + 10$, etc., la construction de la strate négligée en année $a + 1$ pourrait obéir à des règles légèrement différentes. Au lieu d'exiger deux années consécutives de réponses faibles, les FST dans la strate négligée en année $a - 1$ et qui répondent des valeurs faibles en année a seraient classés à nouveau dans la strate négligée en année $a + 1$, sur la base de leur seule réponse en année a et non sur la base d'un historique de deux années de réponse. De tels FST seraient donc remis dans la strate négligée en année $a + 1$ après y avoir été dans le passé et après avoir montré en année a que leurs réponses demeurent faibles.

Les analyses réalisées dans ce rapport sont effectuées en considérant $p = 96$ variables d'intérêt jugées comme importantes. Il est naturel de penser que les variables d'intérêt de l'enquête évolueront dans le temps, c'est-à-dire que certaines variables d'intérêt sont susceptibles de disparaître ou d'apparaître. La disparition de variables d'intérêt ne pose pas de problèmes particuliers. Celles-ci peuvent simplement être retirées lors du calcul de l'indicateur de stratification z_i^* . L'apparition de nouvelles variables d'intérêt soulève de nouvelles problématiques. Il est en effet en théorie possible que des FST soient dans la strate négligée mais contribuent de manière importante aux nouvelles variables d'intérêt introduites. Une solution face à ce problème serait d'une part de sortir un FST de la strate négligée si une contribution importante à une des nouvelles variables d'intérêt peut être anticipée à l'aide d'informations auxiliaires disponibles. D'autre part, l'introduction de nouvelles variables d'intérêt pourrait dans la mesure du possible se faire en même temps que les enquêtes exhaustives introduites périodiquement. Ceci permettrait de collecter des données exhaustives pour les nouvelles variables d'intérêt, au moins pour une année, et de les intégrer dans les bases de décision.

6 Conclusions et perspectives

Dans le cadre d'une potentielle révision méthodologique de la statistique sur les télécommunications, l'OFCOM a mandaté l'OFS afin d'évaluer l'introduction de méthodes d'échantillonnage aléatoire dans leur relevé. Les résultats obtenus confirment l'intérêt d'effectuer une enquête par échantillonnage en lieu et place d'un relevé exhaustif. Des recommandations sont données dans ce rapport sur la construction du plan de sondage. En particulier, les tests effectués montrent l'importance de séparer les fournisseurs de services de télécommunications (FST) en strates et d'échantillonner celles-ci soit de manière exhaustive, soit pas du tout. Cette approche a suscité l'analyse de plusieurs scénarios et des effets qui en découlent. Des résultats probants sont obtenus en se concentrant d'une part sur la contribution des FST au total des différentes variables d'intérêt et d'autre part sur la consistance dans le temps de cette contribution. Les simulations effectuées dans ce travail reviennent in fine à considérer seulement deux strates, l'une exhaustive et l'autre négligée. Les simulations pourraient éventuellement être complétées de données pour l'année 2024, ce qui permettrait d'investiguer davantage la périodicité d'une enquête exhaustive qui a été recommandée.

Le code R livré à l'OFCOM est facile à utiliser sur de nouveaux jeux de données, ce qui devrait permettre à l'OFCOM d'avancer de manière autonome sur les simulations et de calculer des ordres de grandeurs sur les biais. L'indicateur de stratification utilisé pour construire les strates est sensible dans le sens où il suffit d'une seule réponse d'influence élevée pour une seule variable d'intérêt pour que l'indicateur soit élevé. Les utilisateurs qui souhaitent donner une importance différente aux variables d'intérêt peuvent utiliser cet indicateur en appliquant éventuellement un facteur d'importance dans l'équation (1), disons α_j , qui dépendrait de la variable d'intérêt j . Une autre possibilité est de considérer une statistique moins sensible que le maximum sur toutes les variables d'intérêt. Ces approches n'ont pas été testées et pourraient faire l'objet de futurs développements.

Bibliographie

- BEAUMONT, J-F. & HAZIZA, D. & RUIZ-GAZEN, A. (2013). A unified approach to robust estimation in finite population sampling. *Biometrika* **100(3)**, 555–569.
- GRAF, E. & QUALITÉ, L. (2014). Sondage dans des registres de population et de ménages en Suisse : coordination d'échantillons, pondération et imputation. *Journal de la Société Française de Statistique* **155(4)**, 95–133.
- HORVITZ, D. G. & THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association* **47(260)**, 663–685.
- QUALITÉ, L. (2009). Unequal probability sampling and repeated surveys. *University of Neuchâtel*, Ph.D. Thesis.
- R CORE TEAM (2021). R : A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Rapports de méthodes de la section méthodes statistiques de l'OFS
Methodenberichte der Sektion Statistische Methoden des BFS
Methodology reports published by the FSO's Statistical Methods Section

- Chevalier, C. (2025). Statistique sur les télécommunications. Plan de sondage. Numéro OFS : 338-0094
- Potterat, J. (2025). Erhebung leichte Nutzfahrzeuge (LWE) 2023. Stichprobenplan, Gewichtung und Schätzverfahren.
BFS-Nummer : 338-0093
- Chevalier, C. (2025). Statistique de la valeur ajoutée 2021. Révision de la pondération. Numéro OFS : 338-0092
- Bianchi, B., Leuenberger, M., Massiani, A., Panchard, C. (2025). Enquête PIAAC 2022-2023. Plan de sondage, pondération, estimation de variance et analyse du biais dû à la non-réponse. Numéro OFS : 338-0091
- Meskaldji, D.-E., Panchard, C., Qualité, L., Saliba, J. (2024). Enquêtes thématiques 2018-2023. Plans de sondage, tirages, pondération et estimation de variance. Numéro OFS : 338-0090
- Chevalier, C. (2024). Statistique de la valeur ajoutée 2021. Révision du plan de sondage. Numéro OFS : 338-0089
- Nedyalkova, D., Assoulin, D. (2024). Swiss structural business statistics (STATENT) 2015-2021. Estimation and precision of the full-time equivalents. FSO number : 338-0088
- Qualité, L. (2023). Relevé structurel de la population suisse 2018. Révision de la pondération et sélections répétées. Numéro OFS : 338-0087
- Potterat, J., Assoulin, D. (2023). Forschung und Entwicklung 2021. Stichprobenplan, Gewichtung und Schätzverfahren. BFS-Nummer : 338-0086
- Chevalier, C. (2023). Méthodes de correction des effets calendrier en usage à l'OFS en 2023. Construction, choix et application des régresseurs mis à disposition. Numéro OFS : 338-0085
- Mariétan, R. (2022). Comportement en matière de voyages 2020. Révision de la pondération. Numéro OFS : 338-0084
- Massiani, A. (2021). Enquête suisse sur la population active 2015-2020 : Révision de la pondération et de l'estimation de la variance. Numéro OFS : 338-0083
- Chevalier, C., Qualité, L. (2021). Enquête suisse sur la structure des salaires 2018. Révision de la pondération. Numéro OFS : 338-0082
- Qualité, L., Potterat, J. (2020). Enquête suisse sur la structure des salaires 2018. Révision du plan de sondage. Numéro OFS : 338-0081
- Potterat, J., Qualité, L., Assoulin, D. (2019). Strukturhebung der eidgenössischen Volkszählung. Stichprobenplan und -ziehung, Gewichtung, Schätzverfahren und Pooling 2010-2018. BFS-Nummer : 338-0080
- Potterat, J., Assoulin, D., Nicoletti, J.-M. (2019). Beschäftigungsstatistik BESTA. Revision 2015 : Stichprobenrahmen, Stichprobenplan und Hochrechnung. BFS-Nummer : 338-0079
- Massiani, A. (2017). Estimation de la couverture du nouveau recensement en Suisse en 2012. Numéro OFS : 338-0078
- Nedyalkova, D., Assoulin, D. (2017). Construction of full-time equivalents for the Swiss structural business statistics. FSO number : 338-0077
- Ferster, M. (2016). Energieverbrauchsstatistik EVS2014 - Stichprobe, Hochrechnung und Vergleichbarkeit mit der EVS2013. BFS-Nummer : 338-0076
- Panchard, C. (2014). Enquête sur la participation aux activités culturelles 2008. Plan de sondage et estimations. BFS-Nummer : 338-0073

Assoulin, D. (2014). Wertschöpfungsstatistik. Revision 2009 : Statistische Datenaufbereitung und Hochrechnung. BFS-Nummer : 338-0072

Wilhelm, M. (2014). Echantillonnage boule de neige. La méthode de sondage déterminé par les répondants. Numéro OFS : 338-0071

Assoulin, D. (2013). Wertschöpfungsstatistik. Revision 2009 : Stichprobenrahmen und Stichprobenplan. BFS-Nummer : 338-0070

Ferster, M. (2013). EVS I - Energieverbrauchsstatistik 2002 bis 2007 : Stichprobenplan und Hochrechnung. BFS-Nummer : 338-0069

Assoulin, D. (2013). Zusatzerhebung für die landwirtschaftliche Betriebszählung 2010 : Stichprobenplan und Hochrechnung. BFS-Nummer : 338-0068

Potterat, J. (2012). Use of conversion keys for NOGA2002-2008. FSO number : 338-0067-05

Andrade, B., Salamin P.-A. (2012). Enquête sur la situation sociale et économique des étudiant-e-s des hautes écoles suisses 2009. Cadre de sondage, plan d'échantillonnage et méthodes d'estimation. Numéro OFS : 338-0066

Potterat, J., Panchard, C., Kilchmann, D. (2012). Umweltschutzausgaben der Unternehmen 2009 (UWSA2009). Stichprobenplan, Einsetzungen, Gewichtung und Schätzverfahren. BFS-Nummer : 338-0065

Potterat, J. (2012). Benutzung der Umsteigeschlüssel NOGA 2002-2008. BFS-Nummer : 338-0064

Potterat, J. (2011). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe im Jahr 2009 (KNBB09). Stichprobenplan, Gewichtung und Schätzverfahren. BFS-Nummer : 338-0063

Kilchmann, D., Potterat, J., Genoud, S. (2011). Gütertransporterhebung 2008. Stichprobenplan, Datenaufbereitung, Gewichtung und Schätzverfahren. BFS-Nummer : 338-0062

Graf, E. (2010). Enquête suisse sur la santé 2007. Plan d'échantillonnage, pondérations et analyses pondérées des données. Numéro OFS : 338-0061

Eichenberger P., Hulliger B., Potterat J. (2010). Describing the Anticipated Accuracy of the Swiss Population Survey. FSO number : 338-0060

Graf, E. (2010). Étude empirique de l'attrition du Panel Suisse de Ménages : vers une caractérisation du profil du non-répondant. Numéro OFS : 338-0059

Graf, E. (2009). Weightings of the Swiss Household Panel : SHP_I wave 9, SHP_II wave 4, SHP_I et SHP_II combined. FSO number : 338-0058

Graf, E. (2009). Pondérations du Panel Suisse de Ménages : PSM_I vague 9, PSM_II vague 4, PSM_I et PSM_II combinés. Numéro OFS : 338-0057-05

Qualité, L., Tillé, Y. (2009). Estimation de la précision d'évolutions dans l'enquête sur la valeur ajoutée. Numéro OFS : 338-0056

Renaud, A., Panchard, C. et Potterat, J. (2008). Statistique de l'emploi. Révision 2007 : méthodes d'estimation. Numéro OFS : 338-0055

Graf, E. (2008). Pondérations du PSM. PSM_I vague 8, PSM_II vague 3, PSM_I et PSM_II combinés. Numéro OFS : 338-0054

Andrade, B., Graf, M. (2008). Enquête suisse sur la structure des salaires 2006. Aspects méthodologiques du modèle des salaires "Salarium". Numéro OFS : 338-0053

Renaud, A. (2008). Statistique de l'emploi. Révision 2007 : cadre de sondage et échantillonnage. Numéro OFS : 338-0052

Graf, E. (2008). Pondérations du SILC pilote. SILC_I vague 2, SILC_II vague 1, SILC_I et SILC_II combinés. Numéro OFS : 338-0051

Kilchmann, D. (2008). Statistik der sozialmedizinischen Institutionen 1999-2004 und Krankenhausstatistik 1999-2002. Einsetzungen für fehlende Daten. BFS-Nummer : 338-0050

- Renaud, A. (2008). Technologies de l'information et de la communication. Estimations sur la base de la statistique de la valeur ajoutée. Numéro OFS : 338-0049
- Assoulin, D. (2007). Wertschöpfungsstatistik. Einsetzungsversuche für fehlende Antworten grosser Unternehmen. BFS-Nummer : 338-0048
- Kilchmann, D. (2007). Beherbergungsstatistik Campingplätze. Stichprobenrahmen und Schätzverfahren 2005/06. BFS-Nummer : 338-0047
- Gabler, S., Häder, S. (2007). Haushalts- und Personenerhebungen. Machbarkeit von Random Digit Dialing in der Schweiz. BFS-Nummer : 338-0046
- Ferrez, J., Graf, M. (2007). Enquête suisse sur la structure des salaires. Programmes R pour l'intervalle de confiance de la médiane. Numéro OFS : 338-0045
- Renaud, A. (2007). Harmonisation de la scolarité obligatoire en Suisse (HarmoS). Design général de l'enquête et échantillon des écoles. Numéro OFS : 338-0044
- Potterat, J. (2007). Betriebszählung 2005. Statistische Methoden zur Schätzung der provisorischen Ergebnisse. BFS-Nummer : 338-0043
- Hulliger, B. (2006). Umweltschutzausgaben der Unternehmen 2003, Stichprobenplan, Datenaufbereitung und Schätzverfahren. BFS-Nummer : 338-0042
- Renfer, J.-P. (2006). Enquête sur les chiffres d'affaires du commerce de détail. Plan d'échantillonnage et méthodes d'estimation. Numéro OFS : 338-0041
- Salamin, P.-A. (2006). Statistique de l'aide sociale dans le domaine de l'asile. Plan de sondage et extrapolations pour l'enquête pilote 2005. Numéro OFS : 338-0040
- Renaud, A. (2006). Statistique suisse des bénéficiaires de l'aide sociale. Pondération des communes 2004. Numéro OFS : 338-0039
- Graf, M. (2006). Swiss Earnings Structure Survey 2002-2004. Compositional data in a stratified two-stage sample : Analysis and precision assessment of wage components. FSO number : 338-0038
- Potterat, J. (2006). Pensionskassenstatistik 2004. Statistische Methoden zur Schätzung der provisorischen Ergebnisse. BFS-Nummer : 338-0037
- Potterat, J. (2006). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe im Jahr 2004. Stichprobenplan, Gewichtung und Schätzverfahren. BFS-Nummer : 338-0036
- Kilchmann, D. (2006). Vierteljährliche Wohnbaustatistik. Stichprobenplan, statistische Datenaufarbeitung und Schätzverfahren 2005. BFS-Nummer : 338-0035
- Kilchmann, D. (2006). Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2004. Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. BFS-Nummer : 338-0034
- Kilchmann, D., Eichenberger, P., Potterat, J. (2005). Volkszählung 2000. Statistische Einsetzungsverfahren Band 2. BFS-Nummer : 338-0033
- Kilchmann, D., Eichenberger, P., Potterat, J. (2005). Volkszählung 2000. Statistische Einsetzungsverfahren Band 1. BFS-Nummer : 338-0032
- Graf, M., Matei, A. (2005). Enquête suisse sur la structure des salaires 2002. La précision du salaire brut standardisé médian. Numéro OFS : 338-0031
- Graf, E., Renfer, J.-P. (2005). Enquête suisse sur la santé 2002. Plan d'échantillonnage, pondération et estimation de la précision. Numéro OFS : 338-0030
- Potterat, J. (2005). Mietpreis-Strukturerhebung 2003. Gewichtung und Schätzverfahren. BFS-Nummer : 338-0029
- Potterat, J. (2005). Landwirtschaftliche Betriebszählung 2003. Schätzverfahren für die Zusatzerhebung. BFS-Nummer : 338-0028
- Renaud, A. (2004). Coverage estimation for the Swiss population census 2000. Estimation methodology and results. FSO number : 338-0027

Kilchmann, D. (2004). Revision des Schweizerischen Lohnindex. Schätzmethode der Lohnindizes und deren Varianzschätzer. BFS-Nummer : 338-0026

Graf, M. (2004). Enquête suisse sur la structure des salaires 2002. Plan d'échantillonnage et extrapolation pour le secteur privé. Numéro OFS : 338-0025

Renaud, A. (2004). Analyse de données d'enquêtes. Quelques méthodes et illustration avec des données de l'OFS. Numéro OFS : 338-0024

Renaud, A., Potterat, J. (2004). Estimation de la couverture du recensement de la population de l'an 2000. Echantillon pour l'estimation de la sous-couverture (P-sample) et qualité du cadre de sondage des bâtiments. Numéro OFS : 338-0023

Graf, M. (2004). Fusion de données. Etude de faisabilité. Numéro OFS : 338-0022

Potterat, J. (2003). Mietpreis-Strukturerhebung 2003. Entwicklung des Stichprobenplans und Ziehung der Stichprobe. BFS-Nummer : 338-0021

Potterat, J. (2003). Landwirtschaftliche Betriebszählung 2003. Stichprobenplan der Zusatzerhebung. BFS-Nummer : 338-0020.

Renaud, A. (2003). Estimation de la couverture du recensement de la population de l'an 2000. Echantillon pour l'estimation de la sur-couverture (E-sample). Numéro OFS : 338-0019

Hulliger, B. (2003). Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. Erhebung über F+E in der schweizerischen Privatwirtschaft 2000. BFS-Nummer : 338-0018

Renfer, J.-P. (2003). Enquête 2000 sur la recherche et le développement dans l'économie privée en Suisse. Plan d'échantillonnage. Numéro OFS : 338-0017

Potterat, J. (2003). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe. Schätzverfahren. BFS-Nummer : 338-0016

Graf, M., Matei, A. (2003). Stratégie de choix des modèles de désaisonnalisation. Application aux séries de l'emploi total. Numéro OFS : 338-0015

Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Methoden für die Datenbereinigung. BFS-Nummer : 338-0014

Renaud, A. (2002). Programme international pour le suivi des acquis des élèves (PISA). Plans d'échantillonnage pour PISA 2000 en Suisse. Numéro OFS : 338-0013

Renfer, J.-P. (2002). Enquête 2001 sur les coûts et l'utilité de la formation des apprentis du point de vue des établissements. Plan d'échantillonnage. Numéro OFS : 338-0012

Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Stichprobenplan und Schätzverfahren für die provisorischen Ergebnisse. BFS-Nummer : 338-0011

Graf, M. (2002). Enquête suisse sur la structure des salaires 2000. Plan d'échantillonnage, pondération et méthode d'estimation pour le secteur privé. Numéro OFS : 338-0010

Renaud, A., Eichenberger P. (2002). Estimation de la couverture du recensement de la population de l'an 2000. Procédure d'enquête et plan d'échantillonnage de l'enquête de couverture. Numéro OFS : 338-0009

Kilchmann, D., Hulliger, B. (2002). Stichprobenplan für die Obstbaumzählung 2001. BFS-Nummer : 338-0008

Graf, M. (2002). Passage du concept établissement au concept entreprise. Numéro OFS : 338-0007

Salamin, P.A. (2001). La technique de la double enquête pour la statistique du transport routier de marchandise. Numéro OFS : 338-0006

Peters, R., Renfer, J.-P. et Hulliger, B. (2001). Statistique de la valeur ajoutée 1997-1998. Procédure d'extrapolation des données. Numéro OFS : 338-0005

Potterat, J., Hulliger, B. (2001). Schätzung der Sägereiproduktion mit der Sägerei-Erhebung PAUL. BFS-Nummer : 338-0004

Programme des publications de l'OFS

En tant que service statistique central de la Confédération, l'Office fédéral de la statistique (OFS) a pour tâche de rendre les informations statistiques accessibles à un large public. Il utilise à cet effet plusieurs moyens et canaux et diffuse ses informations statistiques par domaines thématiques.

Les domaines statistiques

- 00 Bases statistiques et généralités
- 01 Population
- 02 Espace et environnement
- 03 Travail et rémunération
- 04 Économie nationale
- 05 Prix
- 06 Industrie et services
- 07 Agriculture et sylviculture
- 08 Énergie
- 09 Construction et logement
- 10 Tourisme
- 11 Mobilité et transports
- 12 Monnaie, banques, assurances
- 13 Sécurité sociale
- 14 Santé
- 15 Éducation et science
- 16 Culture, médias, société de l'information, sport
- 17 Politique
- 18 Administration et finances publiques
- 19 Criminalité et droit pénal
- 20 Situation économique et sociale de la population
- 21 Développement durable, disparités régionales et internationales

Sélection de publications

Annuaire statistique de la Suisse



L'Annuaire statistique, que publie l'Office fédéral de la statistique (OFS), constitue depuis 1891 l'ouvrage de référence de la statistique suisse. Il résume les principaux résultats statistiques concernant la population, la société, l'État, l'économie et l'environnement de la Suisse.

Statistique: 175 ans au service de l'État fédéral moderne

Cette brochure a été éditée à l'occasion de la 10 000^e publication de l'Office fédéral de la statistique. Reprenant dix publications et événements appartenant au passé, elle retrace l'histoire de l'OFS, de ses publications et de la statistique en Suisse. La brochure compte 52 pages et est disponible en quatre langues (français, allemand, italien et anglais).



www.statistique.ch – sources d'informations des plus utiles

Le portail Statistique suisse est un outil moderne et attrayant qui vous donne accès aux informations statistiques actuelles. Nous attirons ci-après votre attention sur les offres les plus prisées.

Catalogue général



Presque tous les documents publiés par l'OFS depuis 1860 sont disponibles gratuitement sous forme électronique sur le portail Web Statistique suisse (www.statistique.ch). Les publications sur papier peuvent être commandées par téléphone au +41 58 463 60 60 ou par e-mail à l'adresse suivante order@bfs.admin.ch.

www.statistique.ch → Statistiques → Catalogue

Vous souhaitez être parmi les premiers informés?



Abonnez-vous à un Newsmail et vous recevrez par e-mail des informations sur les résultats les plus récents et les activités actuelles concernant le thème de votre choix.

www.news-stat.admin.ch

Aperçu des données publiées



Pour donner une meilleure vue d'ensemble de tous les jeux de données publiés en continu par l'OFS, nous les avons regroupés sur cette page. Elle contient également un aperçu des données de la statistique fédérale déjà accessibles sur la plateforme opendata.swiss ainsi que des données de la National Summary Data Page (NSDP).

www.data.bfs.admin.ch

Tableau de bord des votations Suisse



Le tableau de bord de l'OFS utilise les données publiques ouvertes (OGD) pour présenter rapidement les résultats des votations sous une forme visuelle et conviviale. Toutes les données sont librement accessibles et disponibles en cinq langues.

votations.admin.ch/fr/overview

Informations complémentaires

Service de renseignement

+41 58 463 60 11, info@bfs.admin.ch



www.statistique.ch



Facebook



Youtube



X



LinkedIn

La statistique sur les télécommunications, publiée chaque année par l'Office Fédéral de la Communication (OFCOM) a fait l'objet d'une étude par la section des méthodes statistiques de l'Office Fédéral de la Statistique (OFS). L'objectif était d'analyser la possibilité de remplacer un relevé exhaustif des unités (fournisseurs de services de télécommunication, FST) par une enquête par échantillonnage aléatoire. Ce rapport décrit les analyses effectuées sur la base de données livrées par l'OFCOM pour les années 2020 à 2023. Force est de constater que pour une large majorité des variables d'intérêt relevées, un nombre restreint de FST contribue significativement au total de ces variables. Cela a motivé l'étude d'un plan d'échantillonnage dont la particularité est la présence d'une strate négligée, c'est-à-dire échantillonnée avec un taux de sondage de 0%, ce qui peut introduire un biais dans les résultats. Cependant, celui-ci reste limité avec des choix judicieux de la taille de la strate négligée et des FST qui la composent. Les résultats montrent des biais typiquement inférieurs à 5% pour la majorité des variables d'intérêt avec une strate négligée de taille égale à environ à 25% de la taille du cadre de sondage. L'approche proposée n'offre cependant pas de garantie théorique sur le biais futur. L'expertise de l'OFCOM demeure nécessaire pour valider la construction de la strate négligée ainsi que sa mise à jour.

En ligne

www.statistique.ch

Imprimés

www.statistique.ch
Office fédéral de la statistique
CH-2010 Neuchâtel
order@bfs.admin.ch
tél. +41 58 463 60 60

Numéro OFS

338-0094

ISBN

978-3-303-00780-8

Les informations publiées ici contribuent à mesurer
la réalisation des objectifs de développement durable (ODD).



Système d'indicateurs MONET 2030

www.statistique.ch → Statistiques → Développement durable
→ Système d'indicateurs MONET 2030

**La statistique
compte pour vous.**

www.la-statistique-compte.ch