

Étude sur la langue des outils de commerce électronique des entreprises

Rapport méthodologique

Pour tout renseignement concernant l'Institut de la statistique du Québec (ISQ) et les données statistiques dont il dispose, s'adresser à :

Institut de la statistique du Québec
200, chemin Sainte-Foy, 3^e étage
Québec (Québec) G1R 5T4

Téléphone :
418 691-2401
1 800 463-4090 (Canada et États-Unis)

Site Web : statistique.quebec.ca

Ce document est disponible seulement en version électronique.

Dépôt légal
Bibliothèque et Archives nationales du Québec
2^e trimestre 2025
ISBN 978-2-555-01502-9 (en ligne)

© Gouvernement du Québec, Institut de la statistique du Québec, 2025

Toute reproduction autre qu'à des fins de consultation personnelle est interdite sans l'autorisation du gouvernement du Québec.
statistique.quebec.ca/fr/institut/nous-joindre/droits-auteur-permission-reproduction

Juin 2025

Publication réalisée à l'Institut de la statistique du Québec par :	Mathieu Baillargeon, statisticien
Avec la collaboration de :	Geneviève Renaud, économiste
Sous la coordination de :	Luc Belleau, coordonnateur
Sous la direction de :	Éric Gagnon, directeur
Révision linguistique et édition :	Direction de la diffusion et des communications
Travaux méthodologiques :	Mathieu Baillargeon, statisticien
Étude sous la responsabilité de :	Direction générale des statistiques et de l'analyse économiques Institut de la statistique du Québec
Étude financée par :	Office québécois de la langue française
Pour tout renseignement concernant le contenu de cette publication :	Direction de la méthodologie Institut de la statistique du Québec 200, chemin Sainte-Foy Québec (Québec) G1R 5T4 Téléphone : 418 691-2401 1 800 463-4090 (Canada et États-Unis) Site Web : statistique.quebec.ca

Notice bibliographique suggérée

BAILLARGEON, Mathieu (2025). *Étude sur la langue des outils de commerce électronique des entreprises. Rapport méthodologique*, [En ligne], Québec, Institut de la statistique du Québec, 21 p. [statistique.quebec.ca/fr/fichier/etude-langue-outils-commerce-electronique-entreprises-2022-methodologie.pdf].

Table des matières

Introduction	5
1 Plan d'échantillonnage	6
Population visée	6
Base de sondage	6
Stratification	7
Répartition de l'échantillon	7
2 Collecte des données	10
Prétest	10
Questionnaire	10
Déroulement de l'étude	11
Résultats	11
3 Estimations	13
Pondération	13
Méthode d'estimation	14
Production d'estimations	14
Tests d'association et analyse des résultats	14
4 Qualité des estimations	15
Erreurs dues à l'échantillonnage	15
Erreurs non dues à l'échantillonnage	16
Le taux de réponse global et la non-réponse partielle	16
L'interprétation des questions	17
Les erreurs de saisie	17
Les réponses volontairement faussées	18
5 Confidentialité des résultats	19
6 L'appréciation globale	20

Introduction

L'Institut de la statistique du Québec (ISQ) a été mandaté par l'Office québécois de la langue française (OQLF) pour réaliser la première édition de *l'Étude sur la langue des outils de commerce électronique des entreprises*. Cette étude a pour objectif de dresser un portrait des langues que les entreprises situées au Québec utilisent pour vendre leurs produits et services en ligne et informer leur clientèle.

L'ISQ et l'OQLF ont conjointement travaillé à l'élaboration du questionnaire de l'étude. Il compte exactement 50 questions portant sur :

- le profil de l'entreprise ;
- la vente en ligne ;
- le site Web de l'entreprise ;
- le module de paiement du site Web de l'entreprise
- les documents disponibles sur le site Web de l'entreprise
- l'application de l'entreprise ;
- le site Web ou l'application de l'enseigne, de la chaîne, du groupe ou de la maison mère associé à l'entreprise ;
- les comptes de médias sociaux (Facebook, Instagram, LinkedIn) ;

- l'infolettre numérique ;
- le courriel ou la notification de confirmation de réception de commande ;
- la facture ;
- les raisons du choix des langues utilisées dans ces divers outils de commerce électronique.

L'Étude sur la langue des outils de commerce électronique des entreprises a été conçue de manière à ce que puissent être obtenus des résultats représentatifs de la situation des entreprises au Québec, lesquels seraient détaillés pour quatre catégories de taille de l'entreprise (en termes de nombre d'employés), pour quatre regroupements industriels et pour trois grandes régions.

Le présent rapport expose les grandes lignes de la méthodologie utilisée pour *l'Étude sur la langue des outils de commerce électronique des entreprises*. Il traite notamment de la population visée, de la base de sondage utilisée, du plan de sondage, du déroulement de la collecte des données et de la méthodologie employée pour produire les estimations. Il y est également question de la qualité des estimations et de la confidentialité des résultats. Enfin, on y trouve une appréciation globale de l'étude.

1 Plan d'échantillonnage

Population visée

La population visée par l'étude est constituée des entreprises de cinq employés et plus présentes au Québec qui reçoivent des commandes sur Internet ou effectuent des ventes de produits ou de services sur Internet et qui appartiennent à l'une des neuf industries du Système de classification des industries de l'Amérique du Nord (SCIAN)¹ suivantes :

- la fabrication (SCIAN 31-33);
- le commerce de gros (SCIAN 41);
- le commerce de détail (SCIAN 44-45);
- le transport et l'entreposage (SCIAN 48-49);
- l'industrie de l'information et l'industrie culturelle (SCIAN 51);
- la finance et les assurances (SCIAN 52);
- les services professionnels, scientifiques et techniques (SCIAN 54);
- les arts, spectacles et loisirs (SCIAN 71);
- les services d'hébergement et de restauration (SCIAN 72).

Base de sondage

La base de sondage utilisée pour l'étude a été construite à partir de la Base de données du registre des entreprises (BDRE) de Statistique Canada mise à jour en juin 2021, de laquelle on a extrait les entreprises dont la taille et l'industrie d'appartenance (SCIAN) figuraient parmi celles visées par l'étude. Puisque la BDRE ne contient pas d'information relative à la vente en ligne, il n'était pas possible de disqualifier des entreprises sur cette base lors de la construction de la base de sondage.

Par rapport à d'autres sources d'information sur les entreprises, publiques ou privées, la BDRE présente de nombreux avantages. Les plus importants sont sans doute sa mise à jour régulière et le fait qu'elle soit conçue spécialement pour la réalisation d'études. Ses autres principaux avantages sont le fait que les données sur la structure des sociétés sont désagrégées selon plusieurs niveaux (le niveau de détail le plus fin correspond à l'emplacement) et qu'il y a présence d'une classification industrielle à jour et assez précise (SCIAN). On y trouve aussi une variable catégorielle estimant le nombre d'employés (1 à 4, 5 à 9, etc.) au Québec de chacune des entreprises, variable qui peut s'avérer précieuse pour l'élaboration de plans de sondage.

La base de sondage ainsi construite comprend 55 723 entreprises. Ces entreprises sont situées au Québec et satisfont aux critères d'inclusion dans la population visée relativement au nombre d'employés et à l'industrie

1. Les secteurs suivants sont donc exclus de la population visée : Agriculture, foresterie, pêche et chasse (SCIAN 11); Extraction minière, exploitation en carrière, et extraction de pétrole et de gaz (SCIAN 21); Services publics (SCIAN 22); Construction (SCIAN 23); Services immobiliers et services de location et de location à bail (SCIAN 53); Gestion de sociétés et d'entreprises (SCIAN 55); Services administratifs, services de soutien, services de gestion des déchets et services d'assainissement (SCIAN 56); Services d'enseignement (SCIAN 61); Soins de santé et assistance sociale (SCIAN 62); Autres services, sauf les administrations publiques (SCIAN 81); et Administrations publiques (SCIAN 91).

d'appartenance. Cette base de sondage contient à la fois les entreprises qui effectuent de la vente en ligne et celles qui n'en font pas, puisque, comme mentionné précédemment, la BDRE ne nous permet pas de savoir si une entreprise en fait ou non. Tous les critères d'inclusion dans la population visée ont été vérifiés auprès des répondants lors de la collecte afin que l'étude ne soit pas menée auprès des entreprises non admissibles. Les répondants ont également dû préciser les régions dans lesquelles les entreprises avaient des établissements.

Stratification

Le plan de sondage devait permettre d'obtenir un nombre de répondants suffisant pour la production d'estimations par taille d'entreprise (en termes de nombre d'employés), par industrie et par région. Pour cette raison, une

stratification de la population a été effectuée (on retrouve donc cinq tailles d'entreprise, quatre regroupements industriels et trois grandes régions). Une stratification efficace assure à la fois une meilleure représentativité de l'ensemble de la population et l'obtention d'une certaine précision pour chaque strate. Les variables de stratification employées sont détaillées dans le tableau 1. On notera que certaines modalités de taille désirées pour la production des résultats n'étaient pas disponibles dans la BDRE. C'est pour cette raison que la variable de stratification relative à la taille comporte cinq modalités et non quatre comme souhaité pour la production des résultats². L'adresse principale des entreprises (généralement l'adresse de leur siège social) a été utilisée pour la construction de la variable de stratification régionale. Pour les entreprises dont le siège social est situé à l'extérieur du Québec, on a utilisé l'adresse du plus gros établissement dans la province. Ainsi, chaque entreprise a été associée à une unique région.

Tableau 1

Variables de stratification et domaines d'estimation pour lesquels un niveau de précision fixé est désiré

Variable	Catégorie
Variables de stratification et domaines d'estimation pour lesquels un niveau de précision fixé est désiré	
Taille d'entreprise en termes de nombre d'employés	5 à 9 10 à 19 20 à 29 30 à 49 50 et plus
Région de l'adresse principale de l'entreprise	Île de Montréal Région métropolitaine de recensement de Montréal, Île de Montréal exclue Reste du Québec
Industrie (Code SCIAN à deux chiffres)	Fabrication (31-33) Commerce de gros (41) Commerce de détail (44-45) Services (48, 49, 51, 52, 54, 71, 72)

Source : Institut de la statistique du Québec, *Étude sur la langue des outils de commerce électronique des entreprises*, 2022.

2. Les quatre modalités de taille désirées pour la production des résultats étaient : 1) 5 à 9 employés, 2) 10 à 24 employés, 3) 25 à 49 employés et 4) 50 employés et plus.

Répartition de l'échantillon

La taille de l'échantillon de l'étude a été fixée à 11 099 entreprises. Pour arriver à calculer cette taille, on a posé des hypothèses relativement aux taux de réponse et aux taux d'admissibilité à l'étude attendus. En effet, plusieurs situations pouvaient entraîner la non-admissibilité d'une entreprise : c'était le cas, par exemple, si elle était fermée, si elle ne faisait pas de vente en ligne, etc. En outre, en raison de la structure de la BDRE, on a dû poser une hypothèse quant à la proportion des entreprises de 20 à 29 employés qui comptaient en fait entre 20 et 24 employés. On a supposé que cette proportion était de 65 %.

Sur la base du prétest et des résultats d'une enquête récente menée auprès des entreprises, soit l'édition 2020 de *l'Enquête sur l'intégration d'Internet aux processus d'affaires*, un taux de réponse de 80 % semblait raisonnable. Étant donné les résultats obtenus lors de cette enquête, il n'y avait pas lieu de croire que le taux d'admissibilité serait influencé par la taille de l'entreprise ou par la région dans laquelle se trouve son adresse principale. En revanche, il semblait exister une association entre le fait de faire de la vente en ligne ou non et l'industrie. Pour cette raison, on a posé une hypothèse différente pour chaque regroupement de secteurs d'activité quant au taux d'admissibilité global, comme nous pouvons le voir au tableau 2.

Des croisements particuliers étaient visés pour la production des estimations relatives à l'étude. Il était prévu dès le départ que les résultats soient ventilés par la taille, par le croisement de la région et de la taille, et par le croisement de l'industrie et de la région. Précisons que les estimations pour les croisements relatifs à une certaine région devaient inclure l'ensemble des entreprises présentes sur ce territoire. Ainsi, si une entreprise est présente dans plus d'une région, elle contribue aux estimations de plusieurs régions. Cela dit, pour effectuer les calculs de taille d'échantillon, on a simplement tenu compte de la variable de stratification régionale, c'est-à-dire de l'adresse principale de l'entreprise sur le territoire québécois.

On a donc réparti l'échantillon en 60 strates correspondant à l'ensemble des croisements des variables de stratification. Un coefficient de variation³ (CV) de 15 % pour les proportions estimées de 15 % ou plus était visé pour chaque modalité de ces croisements. Nous avons déterminé ces seuils en tenant pour acquis que la contribution de certaines entreprises aux estimations de plusieurs régions plutôt qu'une seule augmenterait ultimement le degré de précision de ces estimations.

Tableau 2

Hypothèses quant aux taux d'admissibilité

	Taux d'entreprises faisant du commerce en ligne	Taux d'admissibilité partiel	Taux d'admissibilité global
	%		
Industrie (Code SCIAN à deux chiffres)			
Fabrication (31-33)	24	80	19,2
Commerce de gros (41)	38	80	30,4
Commerce de détail (44-45)	47	80	37,6
Services (48, 49, 51, 52, 54, 71, 72)	40	80	31,8

Source : Institut de la statistique du Québec, *Étude sur la langue des outils de commerce électronique*, 2022.

3. Le coefficient de variation sera défini plus en détail à la section 5.1.

Le tableau 3 présente la taille de la population et de l'échantillon ainsi que la fraction de sondage par taille et par industrie.

Tableau 3

Taille de la base de sondage et de l'échantillon des entreprises

		Base de sondage	Échantillon	Fraction de sondage
		n		%
Taille d'entreprise en termes de nombre d'employés				
5 à 9 employés		21 866	2 871	13,1
10 à 19 employés		15 034	2 196	14,6
20 à 29 employés		6 041	1 296	21,5
30 à 49 employés		5 238	1 286	24,6
50 employés et plus		7 544	3 450	45,7
Région de l'adresse principale et taille de l'entreprise				
Île de Montréal	5 à 49 employés	13 366	1 930	14,4
	50 employés et plus	2 754	1 160	42,1
RMR de Montréal, Île de Montréal exclue	5 à 49 employés	11 163	1 893	17,0
	50 employés et plus	1 718	1 182	68,8
Reste du Québec	5 à 49 employés	23 650	3 826	16,2
	50 employés et plus	3 072	1 108	36,1
Industrie (Code SCIAN à 2 chiffres) et région de l'adresse principale de l'entreprise				
Fabrication (31-33)	RMR de Montréal	3 989	1 930	48,4
	Reste du Québec	4 488	1 608	35,8
Commerce de gros (41)	RMR de Montréal	3 652	1 011	27,7
	Reste du Québec	1 904	987	51,8
Commerce de détail (44-45)	RMR de Montréal	5 935	1 056	17,8
	Reste du Québec	7 454	1 001	13,4
Services (48, 49, 51, 52, 54, 71, 72)	RMR de Montréal	15 425	2 168	14,1
	Reste du Québec	12 876	1 338	10,4
Ensemble du Québec		55 723	11 099	19,9

Source : Institut de la statistique du Québec, *Étude sur la langue des outils de commerce électronique des entreprises*, 2022.

2

Collecte des données

Prétest

Puisqu'il s'agissait de la première édition de l'étude, un prétest était nécessaire car nous devions à la fois vérifier nos hypothèses et déterminer si des éléments devaient être modifiés dans le questionnaire.

En ce qui concerne nos hypothèses, avant le prétest, on supposait que 50 % des entreprises de 20 à 29 employés avaient en fait entre 20 et 24 employés. On supposait également que des taux d'admissibilité de 21 % pour les entreprises appartenant au secteur de la fabrication, de 34 % pour celles appartenant au secteur du commerce de gros, de 28 % pour celles appartenant au secteur des services et de 38 % pour celles appartenant au secteur du commerce de détail étaient appropriés. Comme on peut le constater en comparant ces informations à celles présentées au tableau 2, ces hypothèses ont été revues au moment de construire l'échantillon pour l'étude.

Lors du prétest, certaines informations ont été recueillies par l'entremise de questions ouvertes. Lors de la révision du questionnaire pour l'étude, ces questions ouvertes ont été transformées en des questions à choix de réponse construites selon les réponses fournies au prétest.

Au moment du prétest, les entreprises appartenant aux secteurs d'activité de la gestion de sociétés et d'entreprises (SCIAN 55) et des services immobiliers et des services de location et de location à bail (SCIAN 53) étaient incluses dans la population visée. En raison de taux d'admissibilité jugés trop faibles (car trop peu de ces entreprises faisaient de la vente en ligne), il a été décidé de les retirer de l'étude.

Questionnaire

Le questionnaire de l'étude a été conçu conjointement par l'ISQ et l'OQLF. Il est composé de 50 questions réparties en 8 sections dont les titres sont :

1. Profil de l'entreprise ;
2. Vente en ligne ;
3. Site Web de l'entreprise ;
4. Application de l'entreprise ;
5. Site Web ou application de l'enseigne, de la chaîne, du groupe ou de la maison mère associé à l'entreprise ;
6. Médias sociaux ;
7. Autres outils de commerce électronique ;
8. Raisons de l'utilisation des langues.

La première section permet, entre autres, de déterminer les régions dans lesquelles chaque entreprise est présente. Elle permet également de recueillir d'autres informations de nature générale à leur sujet. La deuxième section porte quant à elle sur la vente en ligne. La plus importante question de cette section est la question 15, c'est-à-dire « Votre entreprise reçoit-elle des commandes sur Internet ou effectue-t-elle des ventes en ligne à l'aide des outils suivants ? », puisqu'elle permet de diriger le répondant vers les sections appropriées dans la suite du questionnaire. Les sections 3 à 7 portent sur les langues utilisées pour effectuer la vente en ligne et contiennent sensiblement toutes les mêmes questions, mais portent sur des outils de commerce électronique différents d'une section à l'autre. Les questions de la section 8 portent sur les raisons qui font qu'une entreprise utilise davantage l'anglais que le français, ou vice versa, dans l'un ou l'autre de ses outils de commerce électronique.

Déroulement de l'étude

Processus de filtrage

Afin de vérifier l'admissibilité des entreprises, de déterminer quelle personne était en mesure de répondre au questionnaire, à savoir un responsable des ventes ou de la mise en marché ou un dirigeant de l'entreprise, et d'obtenir les coordonnées de cette personne, un processus de filtrage a été effectué avant l'expédition des documents de l'étude (lettre d'introduction et questionnaire). Ce processus s'est déroulé sur la même période que la collecte des données étant donné la grande quantité d'entreprises visées. Concrètement, on a demandé aux entreprises de nous fournir, au téléphone ou sur le Web, les informations dont nous avons besoin avant d'envoyer les documents, lesquels pouvaient être expédiés par courriel ou par la poste (au choix du répondant). Environ un mois avant la fin de la collecte, ces documents ont été expédiés aux 2 648 entreprises qui n'avaient pas pu être jointes après plusieurs tentatives. La taille de l'entreprise et l'industrie d'appartenance étaient alors vérifiées avant l'entrevue téléphonique ou au début du questionnaire Web.

Collecte des données

Toutes les entreprises échantillonnées pour l'étude avaient le choix de répondre au questionnaire par téléphone ou sur le Web. Un questionnaire et une lettre qui contenait un identifiant et un lien vers le questionnaire Web ont été envoyés par courriel ou par la poste à tous les participants, selon leur préférence (qu'ils ont pu indiquer s'ils étaient allés jusqu'au bout du processus de filtrage). La collecte des réponses pouvait cependant se faire par téléphone si le participant le souhaitait. Des suivis par courriel et par téléphone ont été effectués

afin de maximiser le nombre de répondants. Durant le dernier mois de la collecte, les entreprises appartenant aux strates dans lesquelles les taux de réponse observés étaient les plus faibles ont été priorisées.

L'envoi du questionnaire par courriel ou par la poste a permis aux participants d'effectuer toutes les recherches préalables nécessaires pour bien le remplir. Les critères de validation (propres à la cohérence et à la logique) ont été programmés dans le logiciel utilisé par les intervieweurs et celui utilisé sur le Web. Ainsi, les réponses du participant pouvaient être validées au moment où il répondait au questionnaire. Cela évitait d'avoir à le recontacter pour corriger les incohérences ou pour recueillir les données manquantes auprès de lui.

La collecte des données s'est effectuée du 3 février au 24 mai 2022. La participation des entreprises sélectionnées était obligatoire en vertu de la Loi sur l'Institut de la statistique du Québec, laquelle confère au statisticien en chef le droit de déterminer qu'une demande de renseignements ait un caractère obligatoire s'il est d'avis que cela est nécessaire pour assurer la fiabilité des statistiques produites.

Résultats

Au total, 2 929 entreprises ont répondu au questionnaire de l'étude et, parmi celles-ci, 92 % l'ont fait sur le Web. Le taux de réponse global pondéré⁴ est de 74,6 %. Parmi les 11 099 entreprises échantillonnées, 6 421 ont été considérées comme inadmissibles à l'étude puisqu'elles n'appartenaient pas à la population visée. Le taux d'admissibilité global pondéré est de 33,7 %. Les taux de réponse et d'admissibilité pondérés obtenus par taille et par industrie sont présentés au tableau 4.

4. Comme les entreprises n'avaient pas toutes la même probabilité d'être choisies dans l'échantillon, il est préférable d'examiner les taux de réponse sur la base de données pondérées par l'inverse de cette probabilité. La pondération est abordée à la section 5.1.

Parmi les entreprises ayant pris part à l'étude, six ont rempli un seul questionnaire pour plusieurs entités, car le répondant affirmait qu'il était impossible de fournir des réponses distinctes pour chacune d'elles. Ainsi, les réponses fournies au sujet de ces 6 entreprises s'appliquaient aussi à 17 autres entreprises qui y étaient reliées. Les réponses ont simplement été reportées dans chacun

des questionnaires concernés (à l'exception des réponses à la question 7⁵, pour laquelle on a effectué des imputations). Certains répondants étaient aussi en mesure de fournir des réponses propres à chaque entreprise pour certaines questions, même si les réponses aux autres questions étaient identiques.

Tableau 4

Nombre d'entreprises répondantes, taux d'admissibilité pondéré et taux de réponse pondéré

		Entreprises répondantes	Taux d'admissibilité	Taux de réponse
		n	%	
Taille d'entreprise en termes de nombre d'employés				
5 à 9 employés		498	24,1	71,4
10 à 19 employés		552	34,7	74,4
20 à 29 employés		403	42,9	75,8
30 à 49 employés		424	47,4	78,9
50 employés et plus		1052	42,5	75,9
Région de l'adresse principale et taille de l'entreprise				
Île de Montréal	5 à 49 employés	482	34,2	71,9
	50 employés et plus	258	38,2	67,1
RMR de Montréal, Île de Montréal exclue	5 à 49 employés	467	33,4	70,0
	50 employés et plus	353	41,3	74,8
Reste du Québec	5 à 49 employés	928	30,6	78,0
	50 employés et plus	441	46,9	83,0
Industrie (Code SCIAN à deux chiffres) et région de l'adresse principale de l'entreprise				
Fabrication (31-33)	RMR de Montréal	339	22,5	74,6
	Reste du Québec	328	24,3	79,2
Commerce de gros (41)	RMR de Montréal	279	33,7	74,5
	Reste du Québec	261	29,3	81,8
Commerce de détail (44-45)	RMR de Montréal	402	41,4	77,9
	Reste du Québec	375	38,4	80,3
Services (48, 49, 51, 52, 54, 71, 72)	RMR de Montréal	540	35,7	66,0
	Reste du Québec	405	32,7	77,3
Ensemble du Québec		2 929	33,7	74,6

Source : Institut de la statistique du Québec, *Étude sur la langue des outils de commerce électronique des entreprises*, 2022.

5. La question 7 est : « Quel est le chiffre d'affaires de votre entreprise pour le dernier exercice financier complet ? ».

3 Estimations

Pondération

La pondération a pour objectif fondamental de permettre l'inférence des résultats observés à la population étudiée, et ce, conformément au plan de sondage. Le poids accordé à une entreprise reflète en quelque sorte le nombre d'entreprises qu'elle représente dans la population. Il est fonction de trois éléments : la probabilité de sélection, le traitement des entreprises inadmissibles et le traitement de la non-réponse. Initialement, on associe à chaque entreprise un poids brut correspondant à l'inverse de sa probabilité de sélection.

Il est à noter que pour la présente étude, on compte un total de 2 929 entreprises répondantes, de 6 421 entreprises participantes inadmissibles, de 482 entreprises non répondantes admissibles et de 1267 entreprises non participantes dont le statut d'admissibilité est inconnu. Or, pour effectuer une pondération adéquate, il faut être en mesure d'estimer le nombre total d'entreprises inadmissibles dans la population. On peut uniquement confirmer le statut d'admissibilité des entreprises qui sont allées jusqu'au bout du processus de filtrage. Cela dit, il existe certainement des entreprises qui n'ont pas pu être jointes lors de ce processus et qui sont en fait inadmissibles.

On effectue donc ensuite une étude de l'inadmissibilité, laquelle permet d'établir des classes expliquant la variation de l'admissibilité selon des caractéristiques des entreprises. Mentionnons entre autres que les entreprises ne sont pas admissibles dans les mêmes proportions selon la strate à laquelle elles appartiennent. Par exemple, les entreprises de 5 à 9 employés ont été plus souvent déclarées inadmissibles à l'étude que celles de 50 employés et plus. Il est important de tenir compte de la structure de l'inadmissibilité lorsque l'on procède à l'ajustement de la pondération puisqu'une partie des entreprises qui ne vont pas jusqu'au bout du processus de filtrage sont probablement inadmissibles. Cette étude de l'admissibilité permet alors de répartir le poids brut des entreprises ne s'étant pas rendues jusqu'à la fin de ce processus entre les entreprises non répondantes et les entreprises inadmissibles. Lors de cette répartition, on considère que le profil des entreprises participantes est semblable à celui des entreprises non participantes.

Par la suite, on effectue un traitement pour la non-réponse sur la base d'une analyse semblable à celle que l'on a faite pour l'admissibilité. La structure de la non-réponse est également différente selon les caractéristiques des entreprises ; il est donc pertinent de la modéliser selon les modalités des variables de stratification qui sont connues pour toutes les entreprises. Ultimement, cette modélisation permet de redistribuer adéquatement le poids brut des entreprises non répondantes aux entreprises répondantes.

Méthode d'estimation

Production d'estimations

Les estimations effectuées dans le cadre de la présente étude ont été faites de manière à ce que les particularités du plan de sondage soient respectées. Elles sont présentées sous forme de tableaux de fréquences. Les estimations sont accompagnées de leur précision estimée. Elles ont été produites à partir des données pondérées et peuvent donc être inférées à la population visée. L'intervalle de confiance et le coefficient de variation accompagnant les estimations ont été calculés selon le plan de sondage stratifié à l'aide du logiciel SUDAAN.

Toutes les estimations sont produites selon un domaine qui diffère en fonction de la question puisque les entreprises n'avaient pas à répondre à toutes les questions. Le domaine sur lequel porte une estimation est indiqué clairement dans chacun des tableaux produits.

Tests d'association et analyse des résultats

Un test du khi-carré avec ajustement de Satterthwaite a été effectué pour chacun des tableaux croisés. Le seuil de signification statistique a été fixé à 5 % pour tous les tests.

Les analyses effectuées sont multivariées. En d'autres termes, les questions ont parfois été mises en relation avec plus d'une variable de croisement à la fois. Chaque question a été analysée en fonction des variables de croisement suivantes : la taille de l'entreprise, l'industrie à laquelle elle est rattachée et les régions dans lesquelles l'entreprise est présente.

Dans le cas où le test du khi-carré était significatif, c'est-à-dire lorsqu'une association entre une question et les variables de croisement était détectée, des tests de comparaisons multiples ont été effectués.

4 Qualité des estimations

Tout résultat tiré d'une étude ou d'un recensement est potentiellement entaché d'erreurs. Ces erreurs se répartissent en deux catégories. D'une part, il y a les erreurs dues à l'échantillonnage, c'est-à-dire celles découlant du fait que seule une partie de la population visée est sondée par échantillon aléatoire. Ces erreurs sont contrôlables a priori par la définition du plan de sondage et la détermination de la taille de l'échantillon. De plus, elles sont quantifiables a posteriori et permettent de prendre en considération la variabilité des estimations produites lors de leur analyse.

D'autre part, il y a les erreurs qui ne sont pas dues à l'échantillonnage, notamment les différents types de biais pouvant influencer sur les résultats : biais causé par la non-réponse, biais dû aux activités ou au mode de collecte, biais de réponse, surcouverture ou sous-couverture de la population par la base de sondage, etc. Les erreurs non dues à l'échantillonnage, elles, ne sont pas formellement quantifiables. Il est donc important de réduire le plus possible les risques de biais et de tenter, le plus possible, de détecter leur présence et d'évaluer, le cas échéant, leur ampleur et leur incidence.

On peut statuer sur la qualité des estimations en examinant certaines statistiques, notamment la précision et la fiabilité des estimations. Les taux de réponse, au total et par question, permettent en outre de porter un jugement éclairé sur la fiabilité des résultats obtenus. Par ailleurs, la marge d'erreur (ME) et le coefficient de variation (CV) associés à une estimation permettent de juger de la précision de celle-ci. En fonction de ces observations, l'utilisateur peut juger de l'importance des erreurs de mesure, des risques de biais et de la portée possible de l'inférence statistique.

Erreurs dues à l'échantillonnage

Les erreurs dues à l'échantillonnage sont mesurées à l'aide de l'intervalle de confiance et du CV. L'intervalle de confiance illustre l'étendue approximative des valeurs que peut prendre la variable étudiée dans la population. Cela signifie que si l'échantillonnage était reproduit un très grand nombre de fois, chaque répétition engendrerait un nouvel intervalle de confiance. Si le niveau de confiance était de 95 %, 19 intervalles sur 20 contiendraient la vraie valeur du paramètre.

Le CV est une mesure qui permet de quantifier la précision de l'estimation et s'obtient en divisant l'erreur-type de l'estimation par la valeur de l'estimation. Plus le CV est élevé, moins l'estimation est précise et vice versa. Il faut être très prudent lors de l'utilisation d'estimations associées à un CV élevé (supérieur ou égal à 25 %), puisque cela indique une très grande variabilité. Ces estimations devraient être fournies à titre indicatif seulement. Le CV sert également à comparer la précision de différentes estimations. Afin que le lecteur puisse juger rapidement de la valeur d'une estimation, dans les tableaux de résultats, les CV sont remplacés par des cotes de précision. Ces cotes sont définies au tableau 5.

Tableau 5

Relation entre la valeur d'un coefficient de variation et la précision de l'estimation associée

Coefficient de variation	Cote	Précision de l'estimation
< 5 %	A	Excellente
$5 \% \leq CV < 10 \%$	B	Très bonne
$10 \% \leq CV < 15 \%$	C	Bonne
$15 \% \leq CV < 25 \%$	D	Passable
$\geq 25 \%$	E	Faible

Source : Institut de la statistique du Québec.

Dans le cadre de cette étude, les objectifs de précision ayant trait aux CV ont été atteints pour l'ensemble du Québec.

Pour l'ensemble des entreprises, la grande majorité des estimations de proportions produites sont associées à un CV inférieur à 15 %, ce qui signifie qu'elles sont assez précises et qu'elles peuvent être utilisées sans crainte. Toutefois, lorsque l'estimation d'une proportion est très petite, sa précision tend à être moins satisfaisante ($CV \geq 15 \%$). De plus, les estimations de proportions portant sur des domaines particuliers sont parfois moins précises ($CV \geq 15 \%$) en raison du nombre moindre de répondants admissibles. La prudence est alors de mise dans l'interprétation des résultats. Ces estimations devraient toujours être accompagnées de leur CV exprimé en cote ou de leur intervalle de confiance.

Erreurs non dues à l'échantillonnage

Comme mentionné précédemment, les erreurs non dues à l'échantillonnage peuvent être de différentes natures : un mauvais taux de réponse (global ou partiel), des questions difficiles à interpréter, des erreurs de saisie, des erreurs d'interprétation de l'intervieweur ou des fausses réponses (données volontairement ou non).

Le taux de réponse global et la non-réponse partielle

Le taux de réponse global à une étude est un indicateur de sa qualité, notamment au regard des biais attribuables à la non-réponse qui pourraient être introduits dans les estimations. En effet, plus la non-réponse est élevée, plus il y a un risque que des biais soient introduits dans les estimations qu'on infère à la population, car les non-répondants peuvent présenter des caractéristiques différentes de celles des répondants. L'ISQ veille à limiter les risques de biais dans toutes ses études. C'est pourquoi il cherche à atteindre le meilleur taux de réponse possible et qu'il s'abstient de procéder à l'inférence statistique lorsque ce taux n'atteint pas un niveau satisfaisant. Dans un tel cas, il analyse les biais afin de mesurer les risques d'erreur, et apporte des corrections avant de produire les estimations.

Dans le cas présent, un taux de réponse global satisfaisant, soit de 74,6 %, a été obtenu. Les taux de réponse obtenus par croisement des variables de stratification sont aussi satisfaisants : seulement deux taux sont inférieurs à 70 %, soit celui obtenu pour les entreprises de 50 employés ou plus dont l'adresse principale se trouve sur l'île de Montréal (67,1 %) et celui obtenu pour les entreprises du secteur des services dont l'adresse principale se trouve dans la RMR de Montréal (66,0 %). Ces taux nous permettent d'être relativement rassurés quant aux risques de biais potentiels, surtout si l'on prend en considération les traitements effectués pour la non-réponse, qui tiennent compte des différentes caractéristiques des entreprises répondantes et des entreprises non répondantes pour l'ajustement des poids.

La non-réponse partielle est faible pour la majorité des questions. Toutefois, elle dépasse 5 % pour la question 49 (à choix de réponse), qui était : « Pour quelle(s) raison(s) votre entreprise utilise-t-elle uniquement ou davantage le français dans certains de ses outils de commerce électronique ? ». Aucun traitement n'a été effectué pour réduire le risque de biais associé à cette non-réponse, l'hypothèse sous-jacente étant que les non-répondants auraient répondu de la même façon que les répondants. La non-réponse partielle s'élève à environ 8 % à cette question.

► La base de sondage

Bien qu'efficace, la base de sondage comporte quelques limites. Tout d'abord, si on peut supposer qu'elle contient la presque totalité des entreprises qui étaient en activité au moment de sa mise à jour (juin 2021), il est aussi presque assuré que certaines de ces entreprises ont cessé leurs activités. De plus, on sait qu'une grande partie des entreprises qu'elle contient ne font pas de vente en ligne, alors que pour être admissibles à l'étude, elles devaient en faire. Ce phénomène, connu sous le nom de surcouverture, est donc très présent dans le cadre de l'étude.

La surcouverture fait qu'il est difficile de connaître la taille exacte de la population étudiée. Bien entendu, on peut faire des hypothèses : dans certaines études, il est possible de d'établir le nombre total d'unités échantillonnées qui ne sont pas admissibles et d'extrapoler ce nombre à l'ensemble de la population étudiée.

Pour l'étude qui nous occupe, il n'a pas été possible de déterminer le statut de toutes les entreprises. Dans certains cas, il n'était pas facile d'attribuer un statut d'admissibilité à l'étude. Voici quelques exemples de non-réponse associée à un statut ambigu : pas de réponse au numéro composé, mauvais numéro de téléphone, ligne occupée. Il fallait donc faire l'hypothèse suivante : parmi ces cas se trouvent un certain nombre de cas admissibles et un certain nombre de cas inadmissibles. Ces nombres ont pu être estimés à l'intérieur de chaque classe du modèle d'ajustement à l'aide de la répartition des entreprises dont le statut était connu. En déterminant un nombre d'entreprises non participantes admissibles, il a été possible de donner un poids aux entreprises répondantes pour qu'elles représentent tous les cas admissibles de l'échantillon. De cette façon, une estimation de la population admissible à l'étude a été produite. Étant donné

que le statut d'admissibilité n'a pas été vérifié pour 11,4 % des entreprises de l'échantillon et que la surcouverture de la base de sondage est grande (65,3 % des entreprises participantes ont été déclarées inadmissibles), l'ajustement ne permet pas de réduire complètement le risque de biais. En effet, à cause de ces deux éléments, l'importance de l'hypothèse selon laquelle les entreprises n'étant pas allées jusqu'au bout du processus de filtrage ont le même profil que les entreprises se trouvant dans le cas contraire devient plus grande.

L'interprétation des questions

De manière générale, les questions posées semblent avoir été bien comprises, et les réponses fournies étaient pertinentes. Il n'y a donc pas de raisons de croire qu'une mauvaise interprétation des questions ait pu causer des erreurs majeures.

Les erreurs de saisie

Tout au long de la collecte, des contrôles de qualité ont été effectués afin d'assurer le respect des normes de confidentialité et de qualité de l'ISQ. La collecte des données a été réalisée à l'aide du logiciel Interviewer de Voxco (de type ITAO [interview téléphonique assistée par ordinateur] et de type IWAO [interview Web assistée par ordinateur]). Ce logiciel permet d'effectuer les sauts de questions de façon automatique, de sorte que selon la réponse donnée à une question, le participant n'a à répondre qu'aux questions subséquentes le concernant. Lorsque l'entrevue se fait au téléphone, l'intervieweur doit valider la réponse saisie pour chaque question avant de passer à la question suivante. Sur le Web et au téléphone, plusieurs types de validation permettent de confirmer certaines données auprès du répondant. Dans le cadre de cette étude, il s'agit de vérifications de la cohérence entre les questions. Une telle validation apparaît notamment lorsqu'on demande dans quelle(s) région(s) l'entreprise possède au moins un établissement. Si un répondant indique que l'entreprise possède au moins un établissement dans la région de Lanaudière, puis mentionne à la question suivante qu'elle n'a, dans la même région, aucun établissement (ni à l'intérieur, ni à l'extérieur de la RMR de Montréal), il reçoit un message d'erreur.

Par ailleurs, un programme informatique a été conçu afin de valider, à la fin du processus de collecte, toutes les données recueillies.

Grâce à toutes ces précautions, il n'y a pas lieu de craindre outre mesure que des erreurs de saisie aient biaisé les résultats.

Les réponses volontairement faussées

Il n'existe pas de moyen absolu de confirmer la véracité de toutes les réponses reçues. Cependant, rien ne laisse supposer que des réponses erronées aient été délibérément fournies dans cette étude. Néanmoins, quelques questionnaires qui présentaient plusieurs incohérences jugées importantes ont été rejetés lors de la validation des données. Dans ces cas-là, les entreprises ont donc été considérées comme des entreprises non répondantes.

5

Confidentialité des résultats

L'ISQ est tenu de protéger la confidentialité des renseignements fournis par les répondants. Les risques de divulgation de renseignements confidentiels à partir de résultats présentés sous forme de tableaux sont donc examinés. Il est possible de déceler deux types de risque de divulgation : 1) une cellule d'un tableau contient peu de répondants, 2) un petit nombre de répondants contribue à un grand pourcentage d'une estimation présentée dans le tableau. Cependant, puisqu'aucun total n'a été produit lors de la présente étude, aucune vérification pour le second type de divulgation n'a été nécessaire.

L'examen des tableaux produits pour la présente étude a permis de déterminer que le premier type de divulgation se produisait. Afin d'éliminer ces risques, des cellules de tableaux ont été masquées.

6 L'appréciation globale

Pour cette étude, comme c'est le cas pour toute étude, il y a lieu de faire preuve de prudence quant à l'interprétation des résultats qui sont accompagnés d'un coefficient de variation supérieur ou égale à 15 %. Il est important de mettre les bémols appropriés dans l'analyse et la présentation de ces données lorsque nécessaire. Un risque de biais est présent en raison d'un nombre relativement important d'entreprises n'étant pas allées jusqu'au bout du processus de filtrage. Malgré cela, on peut conclure que le potentiel d'analyse des données est très bon et qu'il n'existe pas de problème apparent à inférer les résultats à la population étudiée.

« Une organisation
statistique performante
au service d'une société
québécoise en évolution »

statistique.quebec.ca