

Portrait de l'emploi en statistique et en analyse de données au Québec (1^{ère} partie)

La pénurie de main-d'œuvre a un impact sur un grand nombre de domaines de travail au Québec, incluant celui de la statistique et de l'analyse de données. Plusieurs employeurs témoignent de la difficulté de recruter et de conserver la main-d'œuvre. Il en est de même dans certaines universités, où les étudiants sont peu nombreux dans les programmes de statistique, malgré l'engouement envers les domaines reliés aux sciences des données. De plus, plusieurs choisissent d'aller sur le marché de l'emploi, plutôt que de poursuivre vers des études de deuxième et troisième cycles, en raison de la forte demande de travailleurs. Dans un autre ordre d'idées, on semble percevoir un changement dans les logiciels utilisés par les statisticiens, et on se questionne sur la pertinence de les enseigner à l'université.

C'est dans ce contexte que l'enquête sur les conditions d'emploi en statistique et en analyse de données au Québec est née. Celle-ci a été menée par l'Association des statisticiennes et statisticiens du Québec, avec la collaboration de la firme de recherche SOM.

Cette enquête s'inspire de ce qui est déjà fait depuis quelque temps par l'*American Statistical Association* (ASA) de façon régulière¹. Cette association enquête auprès de ses membres afin de faire le portrait de leurs conditions salariales. Leur bassin de membres étant relativement élevé, ils sont en mesure de présenter des tableaux du salaire ventilés en fonction de plusieurs facteurs. Ils ont l'avantage de pouvoir calculer leur taux de réponse, et de vérifier la représentativité de l'échantillon par rapport à l'ensemble de leurs membres. La dernière enquête de 2020 s'est intéressée, en plus des salaires, à l'impact de la COVID-19 sur le milieu, à la satisfaction envers l'emploi, à l'importance de différents aspects du travail, ainsi qu'au titre d'emploi préféré.

D'autre part, au Québec, une enquête similaire a été menée par TECHNOCompétences. Ils ont publié en 2021 un *Profil de la main-d'œuvre en intelligence artificielle, science des données et mégadonnées au Québec*². Celui-ci comprend entre autres une taxonomie des métiers dans le domaine. On y décrit les tâches effectuées, l'évolution des offres d'emploi, les compétences et les qualifications requises pour différents emplois du milieu. Basée sur un sondage avec 75 répondants, la distribution des salaires est aussi présentée pour chaque emploi. Le « scientifique de (méga) données » est associé à un salaire médian de 73 250\$, contre 91 625\$ pour un poste comme scientifique principal. Quant à l'analyste de (méga) données, le salaire médian rapporté est de 68 000\$.

Méthodologie

L'enquête s'est déroulée du 7 mars au 14 avril 2023. La firme SOM a procédé à l'envoi de l'invitation par courriel à une liste de 218 membres et anciens membres de l'ASSQ. Un échantillonnage « boule de neige » était prévu. Après quelques semaines de collecte, le recrutement était inférieur aux attentes, et semblait suréchantillonner certaines sous-

¹ <https://www.amstat.org/your-career/salary-information>

² Cayrat, C., Sigouin-Lebel, A., Poirier St-Pierre, G. (2021). *Profil de la main-d'œuvre en intelligence artificielle, science des données et mégadonnées au Québec*. TECHNOCompétences, Montréal, 113 p.

populations. Afin de remédier à cela, des envois par courriel ont été faits à près de 130 anciens étudiants et employés de l'Université Laval. L'annonce a été affichée dans divers groupes sur les réseaux sociaux. De plus, l'invitation a été transférée individuellement via LinkedIn à environ 250 personnes satisfaisant les critères d'inclusion. Une attention particulière a été portée afin d'équilibrer l'échantillon pour les sous-populations qui étaient peu représentées à ce moment, principalement les gens à l'extérieur de la région de Québec, ainsi que les entreprises privées.

Les critères d'inclusion de l'étude étaient de travailler ou d'habiter au Québec, ainsi que d'avoir actuellement un emploi dans le domaine de la statistique ou de l'analyse de données. Le premier critère permettait de retenir les Québécois qui travaillent à Statistique Canada, par exemple. Le but était de représenter tous les emplois qui pouvaient être obtenus au Québec, même si l'employeur n'y était pas physiquement.

Les questions ouvertes ont été recodées dans les catégories existantes lorsque possible. Aucun traitement des données manquantes n'a été effectué.

Les résultats descriptifs sont présentés avec fréquences et pourcentages pour les variables qualitatives, alors que les moyennes, écarts-types et percentiles sont présentés pour les variables quantitatives. Des croisements ont aussi été effectués entre certaines variables du questionnaire, à l'aide de tests de Wilcoxon, Kruskal-Wallis, du khi deux et Cochran-Mantel-Haenszel, selon la nature des variables considérées. Pour ces croisements, des regroupements de catégories ont été effectués afin d'en diminuer le nombre et d'éviter les catégories à faibles effectifs. Outre ceux pour le salaire, les tableaux de résultats ne sont pas présentés pour les autres croisements.

Les analyses ont été effectuées avec SAS version 9.4 (SAS Institute Inc., Cary, NC, É.-U.) alors que les graphiques ont été générés dans SPSS (IBM Corp. Released 2022. IBM SPSS Statistics for Windows, Version 29.0. Armonk, NY: IBM Corp).

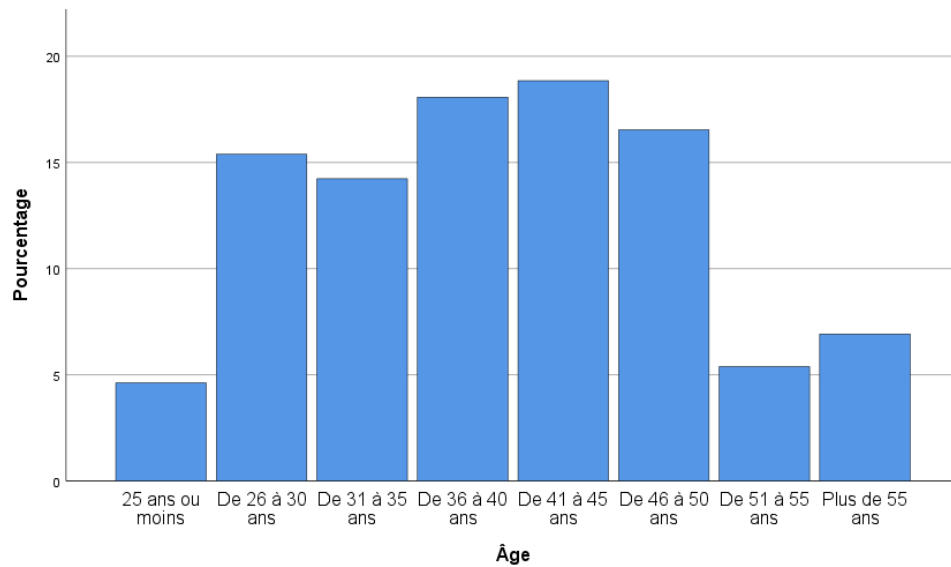
Échantillon

Bien que l'invitation ait été transmise à plus de 600 personnes, 302 participants ont accédé au questionnaire. Parmi ceux-ci, 26 ne respectaient pas les critères d'inclusion et n'ont donc pas pu accéder au reste du sondage. De plus, un répondant avec des réponses non plausibles a été retiré, en plus de 14 participants qui n'avaient pas complété minimalement la première section sur les caractéristiques de leur emploi. Ainsi, 261 répondants ont été inclus pour les analyses.

Caractéristiques des répondants

La majorité des répondants est constituée d'hommes (57 %). La distribution de l'âge des participants, présentée dans la figure 1, est bien répartie.

Figure 1 : Distribution de l'âge des répondants



L'ensemble des caractéristiques des répondants est présenté dans le tableau 1. Seize participants n'ont pas encore terminé leurs études, mais travaillent déjà dans le domaine. La majorité a un diplôme de maîtrise, et on observe que 9 % n'ont aucun diplôme en statistique. Cinquante-cinq pour cent ont aussi un diplôme dans un autre domaine, principalement en mathématiques, les autres domaines étant également répartis entre l'administration, l'informatique, l'intelligence d'affaires, la finance, la santé, les sciences et génie, ainsi que les sciences sociales. Ces autres diplômes sont surtout des baccalauréats et des maîtrises. La majorité des participants ne sont pas des membres de l'ASSQ. Ils ont en moyenne 13 ans d'expérience (écart-type = 10, $n = 260$) dans le domaine. La distribution de cette variable est présentée à la figure 2.

Figure 2 : Distribution du nombre d'années d'expérience en statistique des répondants

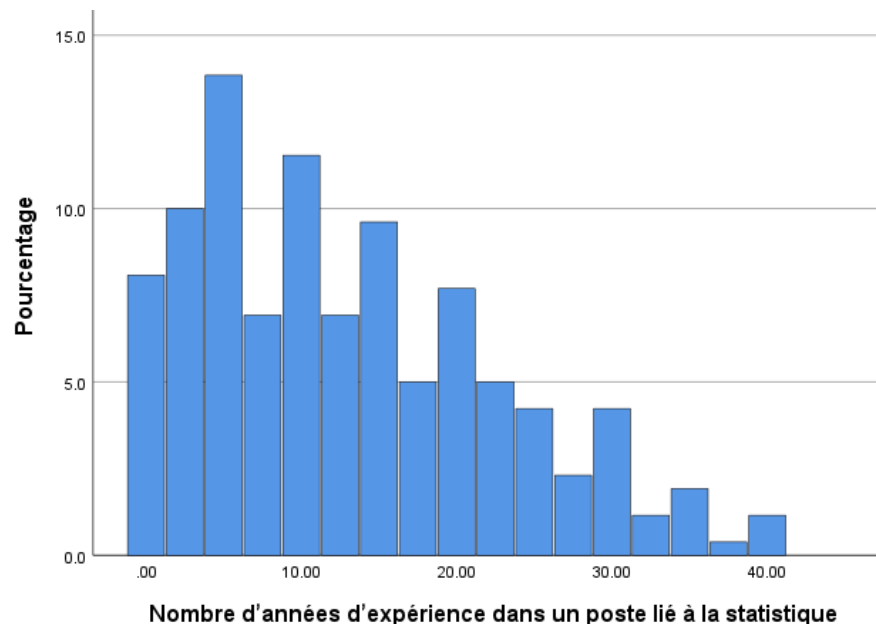


Tableau 1 : Caractéristiques des répondants

Variable (n manquant)	Catégorie	Fréquence (%)
Statut d'emploi	En emploi	245 (93,9)
	En emploi et aux études	16 (6,1)
Hommes (2)		147 (56,8)
Certificat en statistique (2)		6 (2,3)
Baccalauréat en statistique (2)		116 (44,8)
Maîtrise en statistique (2)		152 (58,7)
Doctorat en statistique (2)		21 (8,1)
Combinaison de diplômes en statistique (2)	Aucun	23 (8,9)
	1 ^{er} cycle seulement	71 (27,4)
	1 ^{er} cycle et maîtrise au moins	48 (18,5)
	Maîtrise (avec ou sans doctorat)	104 (40,2)
	Doctorat seulement	13 (5,0)
Plus haut diplôme en statistique (2)	Baccalauréat	71 (27,4)
	Maîtrise	144 (55,6)
	Doctorat	21 (8,1)
Domaine d'un autre diplôme (1)	Aucun	117 (45,0)
	Mathématiques	49 (18,8)
	Autres	94 (36,2)
Plus haut autre diplôme (1), parmi ceux en ayant un	Collégial	4 (2,8)
	Certificat	11 (7,7)
	Baccalauréat	62 (43,7)
	DESS	4 (2,8)
	Maîtrise	46 (32,4)
	Doctorat	15 (10,6)
Membre de l'ASSQ (5)	Oui	69 (27,0)
	Non, ancien membre	71 (27,7)
	Non, jamais membre	116 (45,3)
Accréditation (2)	Aucune	243 (93,8)
	Société statistique du Canada	11 (4,2)
	Autre organisation	5 (1,9)

Caractéristiques de l'emploi

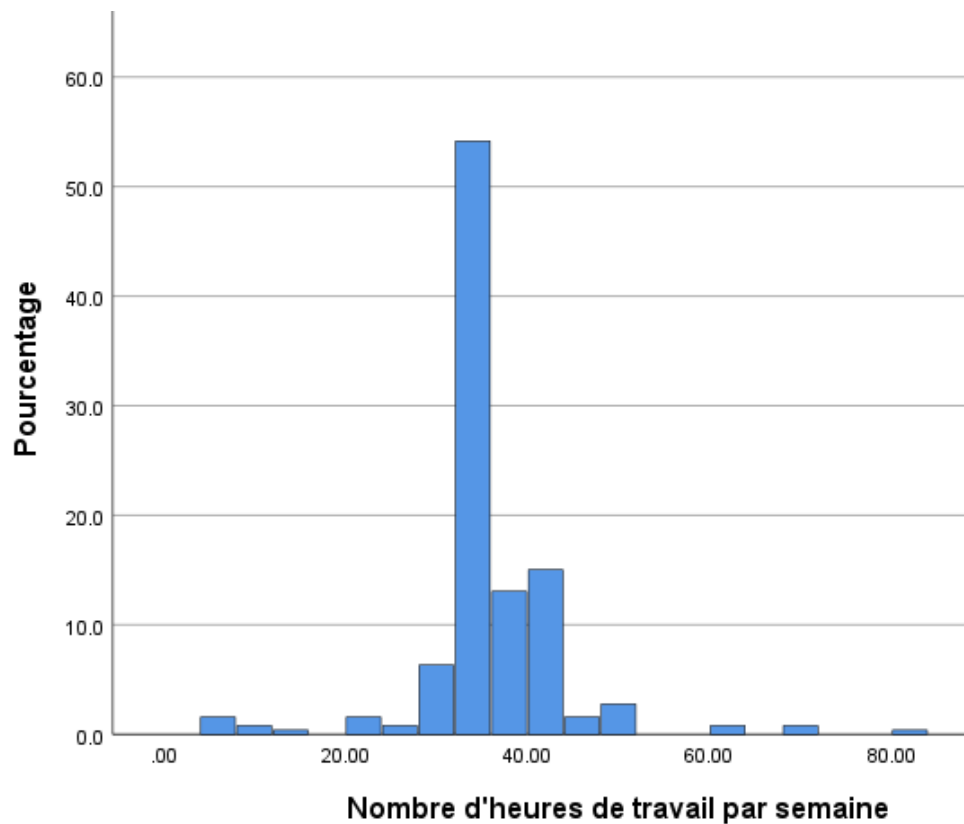
Les participants travaillent surtout au gouvernement (municipal, provincial, ou fédéral) et pour des entreprises privées, tel qu'on le remarque au tableau 2. Ils travaillent principalement dans le domaine de la santé ou des assurances et des banques. Plus de la moitié des répondants travaillent dans la région de Québec. Peu de participants travaillent à temps partiel, et le cinquième occupe un poste de gestion. Les titres d'emploi sont plutôt partagés; les répondants travaillent sous plusieurs titres différents. Quand on leur demande quel titre ils rapportent lorsque quelqu'un leur demande ce qu'ils font dans la vie, la plupart se considèrent comme statisticien. Ils travaillent en moyenne 35,8 heures (écart-type = 7,9; n = 253) par semaine. La distribution du nombre d'heures travaillées par semaine est présentée à la figure 3.

Tableau 2 : Caractéristiques de l'emploi

Variable (n manquant)	Catégorie	Fréquence (%)
Type d'employeur	Travailleur autonome	4 (1,5)
	Établissement d'enseignement	28 (10,7)
	Gouvernement	102 (39,1)
	Entreprise privée	91 (34,9)
	Organisme à but non lucratif	6 (2,3)
	Centre de recherche	30 (11,5)
Principal domaine de travail (5)	Santé	93 (36,3)
	Assurances/Finance/Bancaire	62 (24,2)
	Recherche marketing	5 (2,0)
	Éducation	13 (5,1)
	Nature	12 (4,7)
	Conseils (sujets variés)	23 (9,0)
	Statistique	7 (2,7)
	Enquêtes	11 (4,3)
	Transports	10 (3,9)
	Autres	20 (7,8)
Région de l'employeur	Région de Québec	143 (54,8)
	Région de Montréal	82 (31,4)
	Ailleurs au Québec	13 (5,0)
	Ontario	12 (4,6)
	International	11 (4,2)
Type d'emploi (3)	Temps plein	234 (90,7)
	Temps partiel	24 (9,3)
Poste de gestion (1)		53 (20,4)
Titre de votre poste (3)	Statisticien	51 (19,8)
	Méthodologiste	8 (3,1)
	Scientifique des données	37 (14,3)
	Analyste de données	25 (9,7)
	Professeur/chargé de cours	13 (5,0)

	Biostatisticien	33 (12,8)
	Professionnel	9 (3,5)
	Consultant	12 (4,7)
	Agent de recherche	8 (3,1)
	Conseiller	19 (7,4)
	Analyste	5 (1,9)
	Cadre/chef d'équipe/directeur	22 (8,5)
	Programmeur	11 (4,3)
	Autre titre	5 (1,9)
Appellation personnelle (9)	Statisticien	114 (45,2)
	Biostatisticien	45 (17,9)
	Analyste ou scientifique de données	58 (23,0)
	Gestionnaire	14 (5,6)
	Programmeur	6 (2,4)
	Autre	15 (6,0)

Figure 3 : Distribution du nombre d'heures de travail par semaine



Salaire

Le salaire personnel annuel avant impôt rapporté est en moyenne de 95 991 \$ (écart-type = 34 976 \$; n = 217). Puisque des employés à temps partiel font partie de l'échantillon, et que d'autres travaillent plus de 50 heures par semaine entre autres, un salaire en équivalent 35 heures a été calculé afin de faciliter les comparaisons. La moyenne de ce salaire sur 35 heures est de 93 582 \$ (écart-type = 26 942 \$; n = 213). La distribution des deux variables de salaire est présentée dans les figures 4 et 5. Le salaire est la variable avec le plus de données manquantes. Les gens n'ayant pas rapporté leur salaire ne semblent pas tant se démarquer de ceux qui l'ont fait.

Figure 4 : Distribution du salaire annuel rapporté

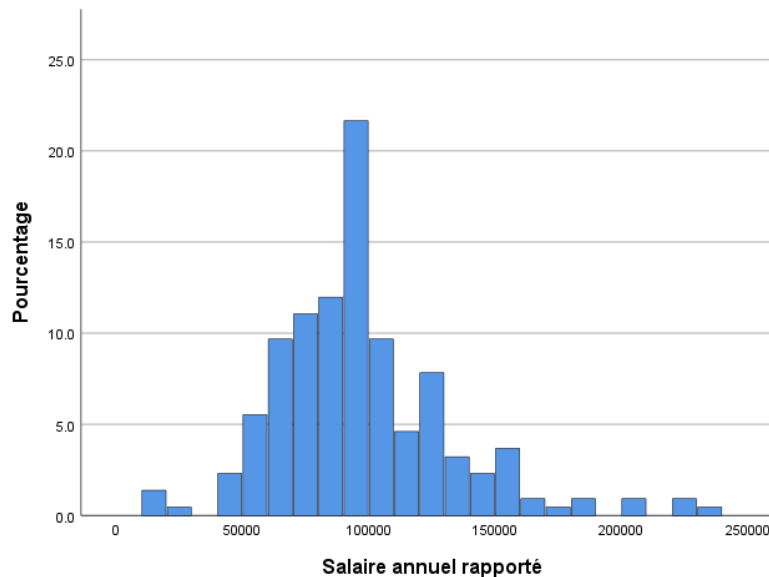
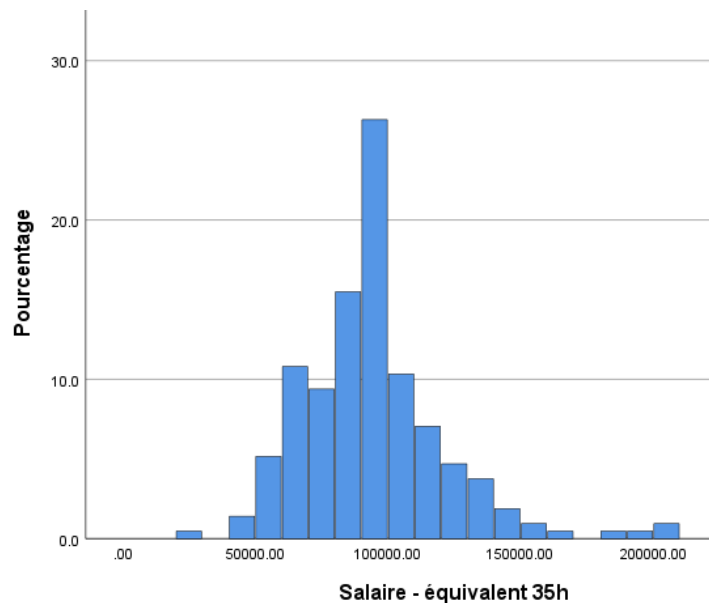


Figure 5 : Distribution du salaire annuel en équivalent 35 heures



Les salaires en équivalent 35 heures sont présentés en fonction de certaines caractéristiques des employés et de l'emploi dans les tableaux 3 (croisements simples) et 4 (croisements doubles avec l'expérience). Certains regroupements de catégories ont été effectués. Pour celles ne pouvant être regroupées, les résultats ne sont pas présentés lorsque moins de cinq répondants se trouvent dans une cellule. La figure 6 présente de façon graphique les résultats du tableau 4b. On observe que les salaires sont plus élevés dans les entreprises privées qu'au gouvernement ou que chez les autres employeurs, ainsi que pour les employés ayant plus d'expérience ou occupant un poste de cadre.

Tableau 3 : Distribution du salaire en équivalent 35 heures selon certains facteurs

Variable	Catégorie	N	25 ^e percentile	Médiane	75 ^e percentile	90 ^e percentile
Genre	Masculin	119	75 000	90 000	108 500	131 250
	Féminin	92	80 000	91 862	102 500	123 846
Combinaison de diplômes en statistique	Aucun	19	75 000	85 050	108 500	125 000
	1 ^{er} cycle seulement	61	78 400	90 500	106 000	118 222
	1 ^{er} cycle et maîtrise au moins	40	77 400	90 000	99 625	118 333
	Maîtrise (avec ou sans doctorat)	84	73 688	93 667	107 187	131 250
	Doctorat seulement	8	89 333	109 375	136 062	159 444
Plus haut diplôme en statistique	1 ^{er} cycle	61	78 400	90 500	106 000	118 222
	Maîtrise	118	74 800	93 333	102 667	130 375
	Doctorat	14	83 300	93 000	118 125	159 444
Expérience en statistique	0-5 ans	61	61 035	70 000	84 000	97 344
	6-10 ans	46	82 133	97 000	113 750	130 375
	11-15 ans	34	84 000	90 312	105 000	137 000
	16-25 ans	49	90 500	99 625	112 500	126 163
	26+ ans	23	92 000	99 625	126 000	140 000
Type d'employeur	Établissement d'enseignement	20	74 000	91 625	113 125	136 062
	Gouvernement	86	82 000	90 000	99 000	104 000
	Privé	77	81 000	105 000	125 000	140 000
	Autre	30	69 600	80 000	91 724	97 875
Principal domaine de travail	Santé	73	70 000	90 000	108 500	130 375
	Assurances/Finance/Bancaire	54	75 000	90 000	110 000	125 000
	Conseils (sujets variés)	20	79 500	95 000	110 000	130 025
	Autres	62	82 585	90 880	99 625	126 000
Titre du poste	Statisticien	37	84 000	92 000	100 000	130 375
	Scientifique des données	31	80 000	91 500	120 000	128 800
	Analyste de données	24	64 500	74 900	84 025	92 000
	Biostatisticien	27	69 600	87 500	100 000	130 681
	Cadre/chef d'équipe/directeur	19	100 188	108 500	119 737	200 000
	Autre titre	74	78 000	91 799	105 000	120 000
Région de l'employeur	Région de Québec	123	70 000	90 000	99 625	118 222
	Région de Montréal	61	82 133	95 000	108 500	131 250
	Ailleurs au Québec	10	75 000	87 937	94 500	108 333
	Ontario	11	98 000	102 667	126 000	128 000
	International	8	93 654	130 375	161 591	201 250
Type d'emploi	Temps plein	197	78 000	90 625	106 000	128 800
	Temps partiel	16	70 000	91 667	96 505	100 000
Poste de gestion	Oui	41	100 187	113 750	130 375	144 375
	Non	172	70 293	89 025	99 000	112 778

Tableau 4 : Distribution du salaire selon certains facteurs et l'expérience

a) Plus haut diplôme en statistique

Plus haut diplôme	Expérience	N	25 ^e percentile	Médiane	75 ^e percentile	90 ^e percentile
Aucun	0-5 ans	11	72 000	78 000	85 050	105 000
1 ^{er} cycle	0-5 ans	18	50 000	65 833	83 000	102 667
	6-10 ans	11	70 000	98 000	123 846	128 800
	11-15 ans	9	82 250	90 625	93 750	117 600
	16-25 ans	15	90 500	99 500	108 500	116 667
	26+ ans	6	99 167	109 111	126 000	140 000
Maîtrise	0-5 ans	29	60 000	67 275	84 000	97 344
	6-10 ans	25	86 735	99 000	120 000	130 375
	11-15 ans	21	85 000	90 000	109 375	141 867
	16-25 ans	27	93 333	99 625	115 000	131 250
	26+ ans	14	90 000	94 356	99 750	140 000

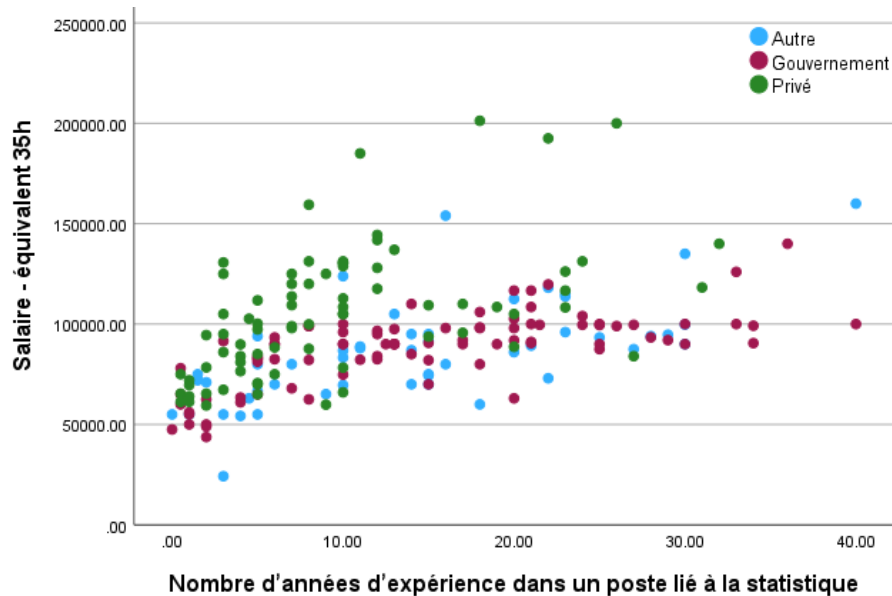
b) Type d'employeur

Employeur	Expérience	N	25 ^e percentile	Médiane	75 ^e percentile	90 ^e percentile
Gouvernement	0-5 ans	19	50 000	62 500	81 250	84 000
	6-10 ans	11	74 800	90 000	99 000	99 000
	11-15 ans	15	82 500	90 000	95 000	97 500
	16-25 ans	25	91 135	98 333	102 667	116 667
	26+ ans	12	92 667	99 396	100 000	126 000
Privé	0-5 ans	29	67 275	78 400	95 000	111 806
	6-10 ans	25	98 000	109 442	125 000	131 250
	11-15 ans	8	113 488	132 500	143 121	185 000
	16-25 ans	11	105 000	110 000	131 250	192 500
Autres	0-5 ans	12	55 000	59 000	71 500	75 000
	6-10 ans	7	69 600	80 000	86 735	123 846
	11-15 ans	11	74 375	88 000	95 000	95 000
	16-25 ans	12	83 000	94 667	114 375	118 125
	26+ ans	7	90 000	94 713	135 000	160 000

c) Domaine

Domaine	Expérience	N	25 ^e percentile	Médiane	75 ^e percentile	90 ^e percentile
Santé	0-5 ans	18	55 000	65 167	78 000	105 000
	6-10 ans	13	74 800	98 933	123 846	130 375
	11-15 ans	14	82 000	86 900	105 000	117 600
	16-25 ans	20	83 000	104 141	115 208	122 144
	26+ ans	8	91 000	97 231	120 000	200 000
Assurances / Finance / Bancaire	0-5 ans	21	65 000	72 000	86 012	95 000
	6-10 ans	15	78 289	99 000	113 750	125 000
	11-15 ans	10	90 000	102 188	137 000	164 688
	16-25 ans	6	90 000	98 000	110 000	119 737
Autres	0-5 ans	21	59 452	71 000	82 585	94 500
	6-10 ans	15	82 133	90 000	120 000	131 250
	11-15 ans	10	88 000	90 000	95 000	95 800
	16-25 ans	22	92 000	99 563	105 000	154 000
	26+ ans	14	93 333	99 396	126 000	140 000

Figure 6 : Dispersion du salaire en équivalent 35 heures selon le nombre d'années d'expérience, selon le type d'employeur



Avantages sociaux

Le tableau 5 présente les avantages sociaux offerts aux travailleurs. Lorsque l'on compare ces avantages sociaux en fonction des caractéristiques des employés et de l'emploi, on observe principalement des différences au niveau du type d'employeur. Les personnes travaillant pour le gouvernement ont l'avantage d'avoir plus d'horaires flexibles et de conciliation travail-famille, ainsi que des régimes de retraite à prestations déterminées. Cependant, elles ont moins de latitude sur leur mode de travail (hybride, en présence, télétravail). Les personnes travaillant au privé font généralement davantage d'heures. Cependant, elles ont généralement des régimes d'assurances plus généreux. Quant aux personnes dans les établissements d'enseignement, elles travaillent davantage en présentiel, et bénéficient d'horaires d'été.

Tableau 5 : Avantages sociaux

Caractéristique (n manquant)	Fréquence (%)
Prime de rareté ou de marché (7)	27 (10,6)
Horaire flexible / conciliation (4)	226 (87,9)
Horaire réduit l'été (6)	31 (12,2)
Mode de travail (1)	
100 % personne	15 (5,8)
Hybride par l'employeur	103 (39,6)
Hybride par l'employé	88 (33,8)
100 % télétravail	54 (20,8)
Semaines de vacances (4)	
2 ou moins	10 (3,9)
3	28 (10,9)
4	159 (61,9)
5 ou plus	60 (23,3)
Régime de retraite (22)	
Non	23 (9,6)
Cotisations déterminées	82 (34,3)
Prestations déterminées	134 (56,1)
Assurance vie (18)	166 (68,3)
Assurance invalidité (18)	185 (76,1)
Assurance soins longue durée (18)	130 (53,5)
Assurance médicament (18)	217 (89,3)
Assurance dentaire (18)	132 (54,3)
Assurance soins professionnels (18)	189 (77,8)
Assurance vue (18)	122 (50,2)

Limites

Contrairement à l'ASA qui mène son étude parmi ses membres seulement, la présente enquête a été menée avec une méthode boule de neige, et on ne connaît pas la population complète. Cela ne permet donc pas de calculer un taux de réponse, ni de vérifier si l'échantillon est représentatif de la population. Cependant, puisqu'il y a davantage de répondants de Québec que de Montréal,

ou encore davantage de personnes travaillant pour le gouvernement que dans le privé, il est possible que le portrait obtenu ne soit pas représentatif de la population. Il faut noter aussi que l'enquête a été diffusée en français seulement. Ainsi, les anglophones n'ont vraisemblablement pas répondu à ce sondage.

En plus de 20 répondants ayant abandonné le questionnaire avant la fin de la première section, et qui ont donc été éliminés de l'analyse, quelques-uns n'ont pas répondu à certaines questions, en particulier à celle du salaire. Aucun traitement de la non-réponse n'a été effectué dans la présente enquête.

Plusieurs facteurs confondants peuvent affecter les différences observées lorsque des croisements ont été effectués. Étant donné la faible taille d'échantillon, il n'est pas possible de faire des croisements en tenant compte de toutes les caractéristiques pertinentes.

Conclusion

En résumé, cette étude basée sur 261 travailleurs dans le domaine de la statistique et de l'analyse de données a permis de brosser un portrait des conditions de travail au Québec, notamment au niveau du salaire. L'enquête comportait aussi une section sur la satisfaction des répondants envers leur emploi, ainsi que sur la nature du travail réalisé et avec quels logiciels. Les résultats de cette partie de l'enquête seront présentés séparément dans la prochaine édition du *Convergence*.

Lorsque comparés à ceux présentés par l'ASA dans son plus récent rapport, les salaires au Québec sont nettement inférieurs. Si on regarde plutôt les salaires médians rapportés par TECHNOCompétence en 2021 pour deux titres d'emploi, ils sont légèrement supérieurs dans la présente étude.

Lors de la plus récente assemblée générale annuelle de l'ASSQ, certains membres ont soulevé leur intérêt envers la mise à jour de cette enquête dans quelques années, afin de suivre l'évolution de la situation. On pourrait peut-être saisir l'occasion pour d'essayer de pallier certaines limites observées lors de la présente collecte. C'est un dossier à suivre!

Anne-Sophie Julien

Pensée de masse zéro...

« Je ne suis rien, je le sais, mais je compose mon rien avec un petit morceau de tout. »

Victor Hugo, Le Rhin

Portrait de l'emploi en statistique et en analyse de données au Québec

PARTIE 2

L'enquête de l'ASSQ sur l'emploi en statistique et en analyse de données au Québec s'est déroulée du 7 mars au 14 avril 2023. Une première série de résultats provenant des données de 261 participants a été publiée dans l'édition précédente de *Convergence*, incluant la méthodologie de l'enquête. Dans le présent article, les résultats sur la satisfaction, les tâches effectuées dans le cadre du travail et les logiciels utilisés sont présentés.

Satisfaction

Le tableau 1 présente la distribution de la satisfaction envers trois caractéristiques de l'emploi, mesurée à l'aide d'une échelle de Likert de 1=très insatisfait à 10=très satisfait. La satisfaction est très bonne, mais un peu moins au niveau de la rémunération. La complexité des tâches (1=très simples, à 10=très complexes) et la charge de travail (1=pas du tout, à 10=trop élevée) sont considérées comme étant élevées, mais l'emploi demeure très stimulant (1=pas du tout, à 10=très stimulant).

Tableau 1 : Satisfaction

Variable	N	Moyenne	Écart-type	Minimum	25 ^e percentile	Médiane	75 ^e percentile	Maximum
Satisfaction – Tâches liées à votre emploi	252	8,07	1,56	1,00	7,00	8,00	9,00	10,00
Satisfaction – Rémunération globale	251	7,56	1,91	1,00	7,00	8,00	9,00	10,00
Satisfaction - conditions de travail	252	8,45	1,46	1,00	8,00	9,00	10,00	10,00
Complexité des tâches	250	7,26	1,56	1,00	6,00	8,00	8,00	10,00
Travail stimulant	251	7,82	1,74	1,00	7,00	8,00	9,00	10,00
Charge de travail	250	7,31	1,85	1,00	6,00	8,00	9,00	10,00

En croisant ces scores entre les différents groupes de l'échantillon, on remarque que la complexité des tâches, la charge de travail, mais aussi la stimulation au travail, augmentent avec l'expérience. Les tâches sont plus complexes pour les personnes avec un diplôme plus élevé. La satisfaction envers la rémunération et les autres conditions de travail est meilleure au privé qu'au gouvernement. Dans le milieu académique, les tâches sont plus complexes et la charge de travail est plus élevée.

Tâches

La médiane du nombre de tâches effectuées par les répondants dans le cadre de leur travail est de 8 (minimum=0, Q1=5, Q3=10, maximum=17). Sans surprise, le traitement de données, les analyses descriptives, la visualisation de données et les analyses statistiques sont les tâches les plus fréquentes, tel que le rapporte le tableau 2.

Tableau 2 : Tâches effectuées dans le cadre du travail (n = 251)

Tâches	Fréquence (%)
Création de banques de données	81 (32,3)
Gestion de banques de données	87 (34,7)
Nettoyage de banques de données	134 (53,4)
Traitement de données	198 (78,9)
Requêtes et extractions de données	152 (60,6)
Tableaux de bord	87 (34,7)
Analyses descriptives	199 (79,3)
Visualisation de données	177 (70,5)
Automatisation de processus	93 (37,1)
Simulation	73 (29,1)
Planification d'enquêtes, d'études, de sondages, d'expériences	105 (41,8)
Analyses statistiques	211 (84,1)
Rédaction de rapports ou d'articles scientifiques	139 (55,4)
Préparation/animation d'ateliers de formation	74 (29,5)
Consultation avec clients	119 (47,4)
Gestion d'employés	59 (23,5)
Gestion financière	32 (12,7)

Analyses statistiques

La médiane du nombre d'analyses statistiques distinctes effectuées est de 5 (minimum=0, Q1=3, Q3=7, maximum=11). Les modèles linéaires et non linéaires, ainsi que les tableaux de fréquences, font partie des analyses les plus fréquentes, telles que présentées au tableau 3. Les équations structurelles ainsi que les séries chronologiques sont les analyses les moins utilisées par les répondants.

Tableau 3 : Analyses statistiques effectuées (n = 240)

Analyses	Fréquence (%)
Calcul de taille d'échantillon	125 (52,1)
Analyse de données d'enquête et de sondage	117 (48,8)
Développement et validation de questionnaires et d'instruments	73 (30,4)
Équations structurelles, analyses factorielles	54 (22,5)
Modèles linéaires et non linéaires	187 (77,9)
Analyses multivariées	147 (61,3)
Tableaux de fréquences	184 (76,7)
Classification	111 (46,3)
Réduction de dimensions	81 (33,8)
Séries chronologiques	55 (22,9)
Apprentissages machine	81 (33,8)

Logiciels statistiques

La médiane du nombre de logiciels utilisés par les répondants est de 3 (minimum=0, Q1=2, Q3=5, maximum=11). L'utilisation des logiciels est présentée dans le tableau 4. Dans l'ordre, on retrouve Excel, SAS, R, SQL et Python comme les logiciels les plus fréquemment utilisés. Parmi les autres logiciels rapportés, on note entre autres Julia, Minitab, Sudaan, NQquery et Tableau qui reviennent quelques fois.

Tableau 4: Utilisation des logiciels statistiques

Logiciels (n = 257)	Fréquence (%)
Excel	191 (74,3)
SAS	189 (73,5)
R	151 (58,8)
SQL	89 (34,6)
Python	84 (32,7)
PowerBi	54 (21,0)
SPSS	43 (16,7)
Autres	29 (11,3)
G Power	21 (8,2)
STATA	20 (7,8)
MPLUS	15 (5,8)
PASS	14 (5,4)
Matlab	5 (1,9)
GraphPad Prism	3 (1,2)
SYSTAT	2 (0,8)
JMP	2 (0,8)

La fréquence d'utilisation des quatre logiciels les plus utilisés est présentée dans le tableau 5. Excel est utilisé de façon très fréquente. Parmi les logiciels de traitement statistique, les utilisateurs de SAS ont une utilisation plus régulière que les utilisateurs de R ou de Python.

Tableau 5 : Fréquence d'utilisation des logiciels statistiques les plus courants

Logiciel (n manquant)	Tous les jours	Toutes les semaines	Tous les mois	Quelques fois par années	Encore moins souvent
Excel (2)	104 (55,0)	68 (36,0)	14 (7,4)	2 (1,1)	1 (0,5)
SAS (4)	93 (50,3)	39 (21,1)	23 (12,4)	20 (10,8)	10 (5,4)
R (1)	39 (26,0)	44 (29,3)	23 (15,3)	39 (26,0)	5 (3,3)
Python (1)	20 (24,1)	16 (19,3)	21 (25,3)	23 (27,7)	3 (3,6)

L'utilisation des logiciels SAS, R et Python a été croisée avec certaines caractéristiques des employés et de l'emploi. SAS est surtout utilisé au gouvernement, dans le milieu de la santé et de la consultation, par des gens avec plus d'expérience. R est utilisé par presque tous les employés d'un établissement d'enseignement, et par près de la moitié des employés au privé et au gouvernement. R est davantage utilisé chez les gens avec moins d'expérience. Quant à Python, il est utilisé surtout dans le privé, dans le domaine des assurances, par ceux ayant moins d'expérience dans le milieu.

Limites

Les mêmes limites que celles rapportées dans la première partie du rapport publié à l'automne dernier s'appliquent aussi ici.

De plus, le portrait des logiciels utilisés sur le marché du travail n'indique pas nécessairement ce qui devrait être enseigné. On voit que les plus récents diplômés (qui ont moins d'expérience) utilisent davantage R et Python, et que SAS est bien implanté au gouvernement. Avec la hausse annoncée des coûts des licences, il est possible que cette forte présence de SAS diminue dans le futur.

Conclusion

En résumé, cette étude basée sur 261 travailleurs dans le domaine de la statistique et de l'analyse de données a permis de connaître la satisfaction des répondants, la nature du travail réalisé et avec quels logiciels ces derniers travaillent.

Anne-Sophie Julien



Comparaison des emplois en statistique selon leur titre

Lors d'un jeudi de l'ASSQ en 2022, Antoine Girard a présenté son « Enquête sur les profils data » réalisée en France. Les résultats sont disponibles en ligne¹. Cette enquête a été accomplie via son réseau LinkedIn et a inclus 150 répondants, qui étaient majoritairement des « data analysts » (incluant les chargés d'étude) ou des « data scientists » (incluant les statisticiens). Les logiciels utilisés, les types de tâches et d'analyses statistiques réalisées et préférées ont été comparés entre les deux groupes. Leur profil démographique était similaire. Python était le logiciel le plus utilisé, surtout chez les scientifiques de données. Les outils de visualisation de données ou de « business intelligence » étaient plus utilisés par les analystes des données. Ces derniers faisaient davantage de tableaux de bord, de visualisation et de requêtes, alors que les scientifiques faisaient davantage d'analyses statistiques. Quant aux tâches préférées par les employés, les scientifiques aimaient surtout prédire et modéliser, alors que les analystes aimaient mieux visualiser les données.

Cette étude a été un des motifs menant à l'enquête sur les conditions d'emploi en statistique et en analyse de données au Québec réalisée à l'hiver 2023 par l'ASSQ. Les premiers résultats descriptifs de cette enquête québécoise, ainsi que sa méthodologie, ont été présentés dans l'édition de septembre 2023 de *Convergence*. La dernière partie des résultats est présentée dans un autre article de la présente édition de *Convergence*. Ensuite, nous nous sommes intéressés à effectuer une comparaison similaire à celle réalisée en France, afin de mieux comprendre la distinction et l'utilisation des principaux titres d'emplois dans le domaine.

Pour faire cette comparaison, le titre d'emploi tel que défini par l'employeur est utilisé (par opposition au titre d'emploi utilisé par le participant pour décrire son emploi). Les quatre principaux titres présents dans l'échantillon et retenus pour les analyses actuelles, sont les statisticiens ($n = 51$), les biostatisticiens ($n = 33$), les analystes de données ($n = 25$) et les scientifiques des données ($n = 37$). Les autres titres sont exclus, comme les consultants, les méthodologistes, les professeurs, les cadres, etc.

Les caractéristiques des employés et de leur emploi sont présentées dans le tableau 1 selon le titre d'emploi. La parité des genres est atteinte chez les statisticiens, mais chez des scientifiques et les analystes des données, les trois quarts sont des hommes. Les analystes, et scientifiques dans une moindre mesure, sont plus jeunes que les autres groupes. La majorité des biostatisticiens a une maîtrise en statistique, contrairement aux autres groupes, comme les analystes qui ont seulement un baccalauréat. La majorité des statisticiens travaille au gouvernement, les scientifiques et les analystes des données sont plutôt au privé et surtout en

1

<https://public.tableau.com/app/profile/antoine.girard/viz/Rsultatsenquatedata2021/RsultatsdenqueteData>

assurances/finances/bancaire, alors que les biostatisticiens se retrouvent surtout dans les centres de recherche et autres organismes en santé.

Tableau 1 : Caractéristiques des employés et de leur emploi selon le titre d'emploi

Caractéristiques		Statisticiens	Biostatisticiens	Analystes de données	Scientifiques des données
Genre masculin		25 (49)	20 (61)	19 (76)	27 (75)
Âge	30 ans et -	7 (14)	NP	15 (60)	13 (35)
	31-40 ans	14 (27)	10 (31)	7 (28)	17 (46)
	41 ans et +	30 (59)	18 (56)	NP	7 (19)
Degré du plus haut diplôme en statistique	Aucun	NP	NP	NP	5 (14)
	Baccalauréat/Cert.	15 (29)	NP	10 (43)	10 (27)
	Maitrise	31 (61)	28 (85)	11 (48)	20 (54)
	Doctorat	NP	NP	NP	NP
Employeur	Privé	12 (24)	8 (24)	12 (48)	27 (73)
	Gouvernement	32 (63)	8 (24)	9 (36)	NP
	Autres	7 (14)	17 (51)	NP	NP
Domaine de l'employeur	Assurances / finance / bancaire	10 (20)	NP	14 (56)	23 (62)
	Santé	15 (31)	32 (97)	5 (20)	NP
	Autres	24 (49)	NP	6 (24)	12 (32)

Les valeurs sont présentées sous la forme *N* (% par titre d'emploi).

NP : non présenté lorsque *N* < 5.

Qu'en est-il du travail effectué par chacun de ces groupes? Les logiciels employés et les tâches effectuées dans le cadre de l'emploi ont aussi été comparés. On remarque au tableau 2 que les statisticiens et les biostatisticiens effectuent davantage de planification d'enquêtes et d'expériences, et de rédaction de rapports ou d'articles scientifiques que les deux autres groupes. Les analystes de données font davantage de tableaux de bord et de visualisation de données, mais moins d'analyses statistiques. Les scientifiques des données utilisent moins SAS et davantage Python que les trois autres groupes. Finalement, les analystes et les scientifiques font davantage de requêtes et extractions de données ou d'automatisation de processus.

Quant aux types d'analyses statistiques, les statisticiens et les biostatisticiens effectuent davantage de calculs de taille d'échantillon, des analyses de données d'enquête, du développement et validation de questionnaires, de modèles d'équations structurelles, et des analyses multivariées que les deux autres groupes. Les analystes font moins de modèles linéaires et non linéaires que les trois autres groupes. Les scientifiques de données font moins de tableaux de fréquences, mais plus de classification, de réduction de dimensions, de séries chronologiques et d'apprentissage machine que les trois autres groupes.

Si on regarde les caractéristiques quantitatives de l'emploi au tableau 3, on remarque que les analystes et les scientifiques des données ont moins d'expérience; ce sont de nouveaux titres d'emploi qui semblent être utilisés pour recruter de jeunes travailleurs.

Tableau 2 : Logiciels utilisés, tâches et analyses effectuées selon le titre d'emploi

Caractéristiques	Statisticiens	Biostatisticiens	Analystes de données	Scientifiques des données
Logiciels				
R	29 (59)	24 (75)	12 (48)	25 (69)
SAS	46 (94)	25 (78)	14 (56)	13 (36)
Python	12 (24)	6 (19)	10 (40)	29 (81)
Tâches				
Création de banques de données	15 (32)	10 (31)	8 (33)	10 (29)
Gestion de banques de données	17 (36)	19 (59)	11 (46)	10 (29)
Nettoyage de banques de données	27 (57)	21 (66)	14 (58)	17 (49)
Traitement de données	38 (81)	25 (78)	21 (88)	30 (86)
Requêtes et extractions de données	27 (57)	19 (59)	18 (75)	27 (77)
Tableaux de bord	11 (23)	8 (25)	20 (83)	13 (37)
Analyses descriptives	40 (85)	31 (97)	21 (88)	26 (74)
Visualisation de données	34 (72)	21 (66)	24 (100)	30 (86)
Automatisation de processus	16 (34)	5 (16)	12 (50)	24 (69)
Simulation	17 (36)	10 (31)	4 (17)	9 (26)
Planification d'enquêtes, d'études, de sondages, d'expériences	29 (62)	19 (59)	4 (17)	4 (11)
Analyses statistiques	43 (91)	32 (100)	17 (71)	29 (83)
Rédaction de rapports ou d'articles scientifiques	37 (79)	22 (69)	7 (29)	11 (31)
Préparation/animation d'ateliers de formation	12 (26)	14 (44)	4 (17)	6 (17)
Consultation avec clients	26 (55)	18 (56)	7 (29)	14 (40)
Gestion d'employés	7 (15)	5 (16)	2 (8)	3 (9)
Analyses				
Calcul de taille d'échantillon	33 (72)	28 (88)	4 (19)	5 (14)
Analyse de données d'enquête et de sondage	29 (63)	15 (47)	7 (33)	7 (20)
Développement et validation de questionnaires et d'instruments	21 (46)	11 (34)	1 (5)	0 (0)
Équations structurelles, analyses factorielles	13 (28)	12 (38)	2 (10)	3 (9)
Modèles linéaires et non linéaires	33 (72)	31 (97)	9 (43)	29 (83)
Analyses multivariées	33 (72)	27 (84)	7 (33)	18 (51)
Tableaux de fréquences	39 (85)	28 (88)	18 (86)	13 (37)
Classification	20 (43)	16 (50)	5 (24)	26 (74)
Réduction de dimensions	15 (33)	11 (34)	2 (10)	21 (60)
Séries chronologiques	6 (13)	5 (16)	3 (14)	15 (43)
Apprentissage machine	12 (26)	10 (31)	8 (38)	28 (80)

Les valeurs sont présentées sous la forme N (% par titre d'emploi).

Le nombre de logiciels utilisés et de tâches effectuées est similaire entre les quatre titres d'emploi. Cependant, le nombre de méthodes statistiques utilisées par les analystes est inférieur à celui des autres titres d'emploi. Aucune différence notable n'est observée quant à la perception de la complexité des tâches, de la stimulation apportée par l'emploi et de la charge de travail.

Tableau 3 : Caractéristiques quantitatives de l'emploi selon le titre d'emploi

Caractéristiques	Titre d'emploi	Minimum	25 ^e percentile	Médiane	75 ^e percentile	Maximum
Nombre d'années d'expérience	Statisticien	0	8	13	25	38
	Biostatisticien	1	4,5	14	21	30
	Analyste	0,5	3	5	7	17
	Scientifique	0,5	3	5	9	20
Nombre de logiciels	Statisticien	0	2	4	5	11
	Biostatisticien	1	2	3,5	5	8
	Analyste	1	3	4	5	7
	Scientifique	1	3	3	4	9
Nombre de tâches	Statisticien	0	6	9	10	15
	Biostatisticien	0	7	9	10	17
	Analyste	0	6	8	10	13
	Scientifique	0	6	7	9	13
Nombre de méthodes statistiques	Statisticien	0	4	5,5	8	10
	Biostatisticien	2	4	6	7,5	11
	Analyste	1	2	2	4	7
	Scientifique	0	3	5	6	9
Complexité des tâches*	Statisticien	1	6	7	8	10
	Biostatisticien	4	7	8	8	10
	Analyste	4	5	7	8	9
	Scientifique	2	6	7	8	10
Travail stimulant*	Statisticien	1	7	8	9	10
	Biostatisticien	3	8	9	9	10
	Analyste	5	6,5	8	9	10
	Scientifique	1	7,5	8	9	10
Charge de travail*	Statisticien	2	6	7	8	10
	Biostatisticien	4	6	7	8	10
	Analyste	5	6,5	8	8	10
	Scientifique	1	5	7	8	10

*Ces trois caractéristiques sont évaluées à l'aide d'une échelle de Likert de 1 à 10. La valeur 10 correspond à des tâches plus complexes, un travail plus stimulant, et une charge de travail plus importante.

En comparant les résultats actuels à ceux de l'étude réalisée en France par Antoine Girard, on constate que les titres d'analystes et de scientifiques de données sont plus fréquents en France qu'au Québec. L'utilisation des logiciels est différente, alors que Python a pris de l'avance en France. Quant à la comparaison des tâches et analyses effectuées, les résultats sont comparables dans les deux enquêtes, bien que les titres comparés ne soient pas exactement les mêmes. Il pourrait être intéressant de faire la même enquête et reprendre ces comparaisons dans quelques années afin de voir l'évolution des titres d'emploi dans ce domaine en constante évolution.

Anne-Sophie Julien