

Étude statistique – 20%

Consignes

- Tout le traitement statistique sera réalisé à l'aide d'Excel.
- Le travail doit être remis individuellement sur Léa.
- Date limite : dimanche 4 décembre 23h59.
- Une pénalité de 10% par jour sera appliquée pour les travaux remis en retard.
- Une pénalité allant jusqu'à 5% sera appliquée pour la qualité de la présentation.

Si vous travaillez en équipe de deux, faites très attention au plagiat et mentionnez avec qui vous avez collaboré. Vous aurez besoin, principalement, de toutes les notions de statistiques vues pendant la session. N'hésitez pas à me montrer votre travail au fur et à mesure. Je pourrai vous guider et vous confirmer que tout avance bien. La majorité du travail peut être effectuée dès maintenant, je vous déconseille très fortement d'attendre à la dernière minute pour commencer.

Mise en contexte

« Dans le cadre du Plan d'action Montréal inclusive de la Ville de Montréal, le sondage Baromètre Écho 2020 a permis de rejoindre 2500 Montréalais.e.s, dont plus de 1 250 personnes immigrantes. Les données et le rapport publié viennent combler des lacunes importantes en termes de données récentes décrivant la situation des personnes immigrantes montréalaises de manière intersectionnelle¹ (ADS+²), notamment en matière de discriminations, de sentiment d'appartenance, ainsi que sur l'accès à l'emploi, au logement et aux services de la Ville. »³

« Ce sondage a également été mené dans un objectif de disposer de données récentes et à jour afin de bien arrimer les actions de la Ville de Montréal avec les besoins de la population immigrante montréalaise. [...] »

Cette enquête Web a été effectuée entre le 20 février et le 23 mars 2020 auprès de 2 503 Montréalais, soit 1 252 non-immigrants et 1 251 immigrants.»⁴

But

Dans le fichier de données qui vous est fourni, seulement les Montréalais.e.s immigrant.e.s sont inclus.e.s. Nous cherchons donc à établir un portrait de ces dernier.e.s et comparer leur situation à celle des non-immigrant.e.s. Vous me présenterez vos résultats et vos analyses dans différentes feuilles ajoutées au fichier de données initial. Un travail de qualité professionnelle est attendu.

¹ Modèle d'analyse et approche pour penser les effets conjoints des différents systèmes d'oppression et pour voir comment ces intersections mettent en place des expériences particulières d'oppression et de privilège afin de développer des stratégies inclusives. (https://cmquebec.qc.ca/wp-content/uploads/2020/09/RGFCN_brochureADS_mars020_VF.pdf)

² L'ADS+ consiste à mettre des lunettes égalitaires pour voir la réalité complexe et éviter ou corriger la discrimination. (https://cmquebec.qc.ca/wp-content/uploads/2020/09/RGFCN_brochureADS_mars020_VF.pdf)

³ <https://www.donneesquebec.ca/recherche/dataset/vmtl-sondage-inclusion-personnes-immigrantes/resource/88bb6a18-54fa-4789-82e8-14888a933a4a>

⁴ <https://www.donneesquebec.ca/recherche/dataset/vmtl-sondage-inclusion-personnes-immigrantes>

Étapes de réalisation

1. Choix des variables étudiées

- Créez une feuille nommée *Choix*.
- Définissez l'unité statistique de l'étude.
- Mentionnez une raison pour laquelle un sondage fut réalisé plutôt qu'un recensement.
- Prenez le temps de bien lire (et peut-être même commencer) la 2^e étape afin de choisir adéquatement les 4 variables que vous allez me présenter. Pour chacune d'entre elles, expliquez pourquoi vous l'avez choisie en une à deux phrases. Les questions suivantes peuvent guider vos justifications : À quoi vous attendez-vous ? Avez-vous une impression que vous aimeriez vérifier ? Qu'est-ce qui vous intrigue ? Est-ce une réalité, un enjeu, qui vous touche personnellement ? etc.

Pour chacune des sous-sections ci-dessous, créez une feuille nommée *Feuille X* (avec la lettre appropriée) et faites toutes les démarches et manipulations décrites comme vous l'avez fait dans les laboratoires. Retournez visionner les capsules vidéo appropriées au besoin.

/40

2. Présentation des données

Une fois un tableau créé (et avant de faire un graphique), il vous faudra retirer les non-réponses (cellules vides).

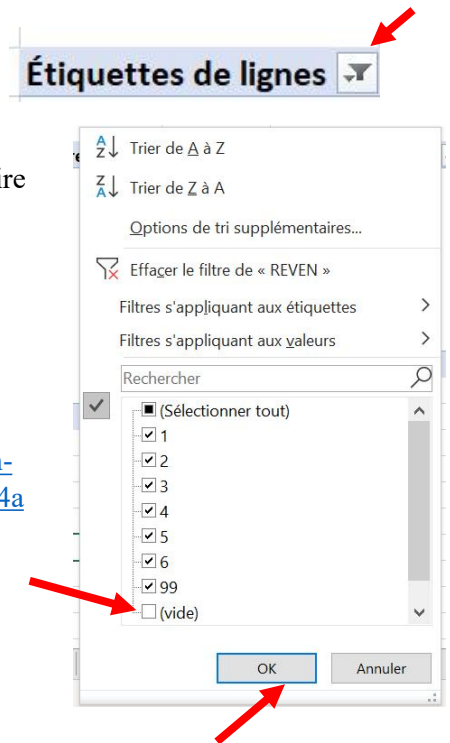
1. Cliquez sur le petit entonnoir à côté de «Étiquettes de lignes»
2. Décochez (*vide*) et cliquez sur *OK*
Attention ! La liste des valeurs peut être longue, il faut bien la faire défiler.

Note 1 : La taille d'échantillon changera donc d'une variable à l'autre.

Note 2 : Portez une attention particulière aux règles de présentation.

La source à citer est la suivante :

Sondage Écho 2020 sur l'inclusion des personnes immigrantes,
<https://www.donneesquebec.ca/recherche/dataset/vmtl-sondage-inclusion-personnes-immigrantes/resource/88bb6a18-54fa-4789-82e8-14888a933a4a>



A) Analyse d'une variable qualitative nominale (Feuille A)

- Faire un tableau de fréquences.
- Créer un graphique approprié.
- Calculer et interpréter dans le contexte une statistique descriptive parmi les suivantes :
Mode, médiane, moyenne, étendue, écart-type, coefficient de variation, cote z

B) Analyse d'une variable qualitative ordinale (Feuille B)

- Faire un tableau de fréquences.
- Créer un graphique approprié.
- Calculer et interpréter dans le contexte une statistique descriptive parmi les suivantes :
Mode, médiane, moyenne, étendue, écart-type, coefficient de variation, cote z

C) Analyse d'une variable quantitative discrète avec peu de valeurs distinctes (Feuille C)

- Faire un tableau de fréquences.
- Créer un graphique approprié.
- Calculer et interpréter dans le contexte les trois mesures de tendance centrale.
- Quelle mesure de tendance centrale est-il préférable de conserver ? Justifier et commenter.
- Calculer et interpréter dans le contexte le coefficient de variation.
- La distribution est-elle homogène ? Justifier et commenter.

D) Analyse d'une variable quantitative continue ou quantitative discrète avec beaucoup de valeurs distinctes (Feuille D)

- Faire un tableau de fréquences.
- Créer un histogramme et une ogive.
- Calculer et interpréter dans le contexte les trois mesures de tendance centrale.
- La distribution est-elle symétrique ou asymétrique ? Justifier et commenter.
- Calculer et interpréter dans le contexte l'écart-type.

/20 3. Inférence statistique

E) Inférence sur une moyenne à l'aide d'un intervalle de confiance (Feuille E)

- Choisir une variable adéquate.
- Choisir un niveau de confiance parmi les suivants : 90%, 95% ou 99%.
- Calculez l'intervalle de confiance pour la moyenne de la variable choisie au niveau de confiance choisi, incluez vos démarches au long.

F) Inférence sur une proportion à l'aide d'un intervalle de confiance (Feuille F)

- Choisir une variable adéquate.
- Choisir un niveau de confiance **différent du précédent** parmi les suivants : 90%, 95% ou 99%.
- Calculez l'intervalle de confiance pour une proportion liée à la variable choisie au niveau de confiance choisi, incluez vos démarches au long.

/20 4. Corrélation

G) Corrélation (Feuille G)

- Choisir deux variables et expliquant brièvement le lien possible entre elles.
- Copier-coller, dans les colonnes A et B de la feuille, les données recueillies pour ces deux variables.
- Faire le nuage de points.
- Au besoin, retirer les couples des données aberrantes et les noter ailleurs sur la feuille.
- Justifier brièvement le retrait de ces couples de données. *Venir me voir au besoin.*
- Refaire le nuage de points et en faire deux copies. *Attention aux règles de présentation.*
- Tester les deux types de corrélation qui semblent les plus adaptés.
- Dans chacun des cas, afficher la courbe de régression, son équation et le coefficient de détermination.
- Laquelle des corrélations s'applique le mieux aux données ? Justifier.
- Dans le modèle retenu, interpréter les coefficients de l'équation de régression ainsi que le coefficient de détermination.

/14 5. Analyse des résultats

H) Comparaisons (Feuille H)

- Effectuer au moins 5 comparaisons entre les immigrants et les non-immigrants. Utiliser le traitement statistique effectué ainsi que les informations fournies à la fin de ce document.
- Qu'est-ce qui vous surprend le plus dans les résultats obtenus ? Expliquer.
- Retourner lire la mise en contexte et lancer des pistes de réflexion. Nommer des éléments qui semblent problématiques à la lumière des résultats obtenus.

Les variables

<p>LANGU : Quelle est la langue que vous avez apprise en premier lieu à la maison dans votre enfance et que vous comprenez toujours?</p> <p>1 Français 2 Anglais 96 Autre 99 Je préfère ne pas répondre</p>	<p>SCOL : À quel niveau se situe la dernière année de scolarité que vous avez terminée?</p> <p>1 Primaire (7 ans ou moins) 2 Secondaire (formation générale ou professionnelle – 8 à 12 ans) 3 Collégiale (formation pré-universitaire, formation technique, certificats, attestations ou diplômes de perfectionnement) 4 Universitaire - certificats et diplômes 5 Universitaire - 1er cycle (Baccalauréat) 6 Universitaire - 2e cycle (Maîtrise) 7 Universitaire - 3e cycle (Doctorat) 9 Je préfère ne pas répondre</p>
<p>Q2r1 : Je comprends le français...</p> <p>1 Parfaitement 2 Très bien 3 Plutôt bien 4 Pas très bien 5 Pas du tout</p>	<p>Q2r5 : Je comprends l'anglais...</p> <p>1 Parfaitement 2 Très bien 3 Plutôt bien 4 Pas très bien 5 Pas du tout</p>
<p>FOY1 : En vous incluant, combien de personnes habitent dans votre logement, en incluant les adultes et les enfants?</p>	<p>Q4 : Êtes-vous actuellement :</p> <p>1 Citoyen.ne canadien.ne 2 Résident.e permanent.e 2 Résident.e non permanent.e</p>
<p>OCCUP : Quelle est votre occupation principale actuelle ? (On parle de votre occupation courante, même si vous êtes actuellement en congé)</p> <p>1 Employé de bureau 2 Personnel spécialisé dans la vente 3 Personnel spécialisé dans les services 4 Travailleur manuel 5 Ouvrier spécialisé ou semi-spécialisé 6 Travailleur des sciences et technologies 7 Professionnel 8 Gestionnaire/Administrateur/Propriétaire 9 Au foyer 10 Étudiant 11 Retraité 12 Sans emploi 96 Autre 99 Je préfère ne pas répondre</p>	<p>Q14 : Avez-vous actuellement plus d'un emploi?</p> <p>1 Oui 2 Non</p>

<p>Q15r1 : Combien d'heures par semaine travaillez-vous en moyenne habituellement en tant que travailleur salarié, entrepreneur ou travailleur autonome?</p>	<p>REVEN : Parmi les catégories suivantes, laquelle reflète le mieux le revenu total avant impôt de tous les membres de votre foyer pour l'année 2018?</p> <p>1 19 999\$ et moins 2 de 20 000\$ à 39 999\$ 3 de 40 000\$ à 59 999\$ 4 de 60 000\$ à 79 999\$ 5 de 80 000\$ à 99 999\$ 6 100 000\$ et plus 99 Je préfère ne pas répondre</p>
<p>Q25 : Votre emploi actuel correspond-il à votre domaine d'études ou professionnel?</p> <p>1 Oui 2 Non 8 Je ne sais pas 9 Je préfère ne pas répondre</p>	<p>Q28 : Dans quel type de logement vivez-vous?</p> <p>1 Habitation à loyer modique (HLM) 2 Coopération d'habitation 3 Organisme à but non lucratif d'habitation 4 Logement locatif privé 5 Maison 6 Condo 96 Autre 98 Je ne sais pas 99 Je préfère ne pas répondre</p>
<p>Q31r1 : Combien est-ce que vous ou votre famille payez au total par mois pour votre hypothèque ou votre loyer?</p>	<p>Q34r1 : Combien de chambres à coucher (pièces fermées) compte votre logement?</p>
<p>Q35r1 : Comment évaluez-vous, par rapport à l'emplacement de votre logement ou maison, la proximité à distance de marche d'une station de métro?</p> <p>1 Moins de 5 minutes à pieds 2 Entre 5 et 10 minutes à pieds 3 Entre 10 et 20 minutes à pieds 4 Plus de 20 minutes à pieds 8 Je ne sais pas</p>	<p>Q39r1 : Depuis combien de temps habitez-vous dans votre logement actuel?</p>
	<p>Q40r1 : Combien de fois avez-vous déménagé au cours des 5 dernières années?</p>
<p>Q44 : Dans quelle mesure diriez-vous que les Montréalais sont accueillants envers les personnes immigrantes?</p> <p>1 Très accueillants 2 Plutôt accueillants 3 Pas vraiment accueillants 4 Pas accueillants du tout 8 Je ne sais pas 9 Je préfère ne pas répondre</p>	<p>Q48 : À quel point vous sentez-vous en sécurité dans votre quartier? Vous sentez-vous...</p> <p>1 Tout à fait en sécurité 2 Plutôt en sécurité 3 Pas vraiment en sécurité 4 Pas du tout en sécurité 8 Je ne sais pas 9 Je préfère ne pas répondre</p>

Chez les non immigrants

Statistiques descriptives

	Mode	Médiane	Moyenne	Étendue	Écart-type
LANGU	1,0	1,0	17,5	98,0	35,8
SCOL	3,0	4,0	4,1	8,0	1,5
Q2r1	1,0	1,0	1,4	4,0	0,8
Q2r5	1,0	2,0	2,0	4,0	1,1
FOY1	2,0	2,0	2,7	7,0	1,1
Q4	1,0	2,0	1,8	2,0	0,8
OCCUP	11,0	7,0	11,7	98,0	20,5
Q14	2,0	2,0	1,8	1,0	0,4
Q15r1	40,0	37,0	36,5	99,0	10,1
REVEN	2,0	3,0	13,6	98,0	29,5
Q25	1,0	1,0	1,7	8,0	1,5
Q26	1,0	1,0	1,7	7,0	1,4
Q31r1	1000,0	850,0	1007,4	15000,0	740,0
Q34r1	2,0	2,0	2,3	18,0	1,2
Q35r1	4,0	3,0	2,9	7,0	1,4
Q39r1	2,0	6,0	10,7	67,0	11,4
Q40r1	1,0	2,0	2,1	9,0	1,3
Q44	2,0	2,0	2,2	8,0	1,6
Q48	1,0	2,0	1,7	8,0	1,0

Distribution (très rapide) de variables (ayant peu de valeurs/modalités distinctes)

LANGU ▼ Nombre 1 832 2 205 96 207 99 8 Total 1252	SCOL ▼ Nombre 1 7 2 215 3 331 4 118 5 324 6 213 7 36 9 8 Total 1252	Q2r1 ▼ Nombre 1 914 2 194 3 86 4 46 5 12 Total 1252	Q2r5 ▼ Nombre 1 546 2 309 3 256 4 121 5 20 Total 1252	FOY1 ▼ Nombre 1 37 2 452 3 175 4 152 5 47 6 10 7 3 8 2 Total 878
Q4 ▼ Nombre 1 207 2 116 3 105 Total 428	OCCUP ▼ Nombre 1 186 2 41 3 72 4 39 5 36 6 66 7 188 8 87 9 42 10 153 11 232 12 43 96 53 99 14 Total 1252	Q14 ▼ Nombre 1 192 2 590 Total 782	REVEN ▼ Nombre 1 158 2 251 3 228 4 147 5 119 6 216 99 133 Total 1252	Q25 ▼ Nombre 1 462 2 288 8 13 9 19 Total 782
Q26 ▼ Nombre 1 163 2 110 4 4 8 11 Total 288		Q34r1 ▼ Nombre 1 291 2 452 3 303 4 90 5 30 6 8 7 3 9 1 19 1 Total 1179	Q35r1 ▼ Nombre 1 227 2 293 3 264 4 430 8 38 Total 1252	Q40r1 ▼ Nombre 1 230 2 168 3 107 4 34 5 17 6 4 8 3 9 1 10 2 Total 566
Q44 ▼ Nombre 1 350 2 691 3 118 4 22 8 55 9 16 Total 1252		Q48 ▼ Nombre 1 574 2 564 3 80 4 16 8 10 9 8 Total 1252		