

TP3 : Analyses statistiques et modélisation

CIV8760 - Gestion de données en transport
Frédéric Chabot & Nicolas Saunier

27 octobre 2023

Ce troisième travail pratique vise à vous faire appliquer différentes méthodes d'analyses statistiques de données de transports ainsi qu'à vous faire travailler avec un modèle de régression linéaire.

1 Jeu de données

Les données utilisées sont les [temps de parcours](#) collectés par la Ville de Montréal à l'aide de capteurs Bluetooth disponible sur le portail des données ouvertes. Les données sont collectées pour des paires origine-destination pré-définies, décrites dans le fichier des [segments routiers](#) de collecte des temps de parcours. Les données de temps de parcours sont disponibles pour plusieurs années. Les attributs des deux fichiers de données sont décrits sur leurs pages respectives.

2 Mandats

Ce travail pratique est divisé en deux grandes étapes afin de mettre en pratique différentes compétences.

À noter que pour chaque question, une méthodologie doit être donnée de façon claire et concise. L'usage de requêtes SQL et de la programmation est fortement recommandé et le cas échéant, il serait approprié de les présenter dans votre rapport. Autrement, vous êtes libres d'utiliser l'outil que vous voulez pour manipuler les données et faire vos graphiques.

2.1 Analyse d'un tronçon

Pour les deux sections qui suivent, il faut sélectionner **UN** segment routier à étudier qui va dans les deux directions de déplacement (de A vers B **ET** de B vers A, par exemple "Sherbrooke_N06: Curatteau a SaintDonat" et "Sherbrooke_S07: SaintDonat a Curatteau") de longueur supérieure à **1000 m**. Indiquer clairement le segment sélectionné. Il est important de sélectionner un segment qui contient des données pour chacune des années (2016 à 2019).

2.1.1 Analyse des temps de parcours et vitesses

Décrire les temps de parcours et les vitesses dans chaque direction par des statistiques descriptives et des figures représentant l'ensemble de la distribution par année. Chacune des années ne contient pas nécessairement la même quantité d'information pour chaque segment. Il est donc important dans vos analyses/interprétations de tenir compte de cet aspect. De plus, en vous basant sur les analyses précédentes, proposez une méthode pour filtrer les

temps de parcours aberrants. La méthode de filtrage doit être appliquée à toutes les données utilisées par la suite.

Enfin, une fois les données filtrées, faire un test d'adéquation à la loi normale (ou autre loi théorique qui vous semble plus adaptée) des vitesses de parcours dans chaque direction sur votre tronçon. Utilisez la méthode vue en classe et décrivez les différentes étapes (vous pouvez comparer avec d'autres méthodes automatisées sur les différents outils à votre disposition tels que R ou Python). Justifiez votre choix de loi théorique et expliquez clairement l'hypothèse initiale de votre test et le résultat obtenu. Donnez tous résultats intermédiaires et commentez sur la fiabilité d'un tel test.

2.1.2 Analyse des observations et temps de parcours en fonction du temps

En choisissant une direction de déplacement, décrire la répartition temporelle des nombres d'observation et des temps de parcours, selon les mois, jours de la semaine (incluant les jours de fin de semaine) et heures de la journée. Justifier la statistique utilisée pour représenter les temps de parcours. Faire au moins six graphiques présentant les informations suivantes par année (2016 à 2019) et commenter :

- Nombre d'observations et temps de parcours en fonction des mois;
- Nombre d'observations et temps de parcours par jour de la semaine;
- Nombre d'observations et temps de parcours par heure de la journée selon le type de jour (semaine/fin de semaine).

Pour ce dernier point, il vous est recommandé de faire un graphique pour les jours de semaine et un second pour les jours de fin de semaine. Ceci facilitera la visualisation et l'analyse.

Enfin, vous devez étudier la corrélation des temps de parcours entre deux années différentes de votre choix (inclure une figure). Pour ce faire, il s'agit de grouper les valeurs de temps de parcours par heure de la journée, et ce, pour deux années distinctes. Vous aurez ainsi une valeur de temps par année pour chaque heure de la journée. Ensuite, dans un graphique Excel (nuage de points), mettez simplement les valeurs d'une année en "x" et celles de l'autre année en "y". Vous pourrez donc ajouter une ligne de régression linéaire et ajouter le coefficient. Notez qu'il y a des moyens similaires de faire un graphique en y ajoutant une ligne de régression ainsi que de calculer le coefficient de détermination (ou R^2) à l'aide de Python.

2.2 Étude des facteurs associés aux conditions de déplacement

Pour cette dernière partie, vous devez considérer que l'année 2016 et cinq segments routiers dans des quartiers différents de Montréal. Présentez brièvement ces quartiers. Vous devez

ensuite créer une variable décrivant les conditions de déplacement. Par exemple, le ratio de la vitesse moyenne sur le tronçon et sa limite de vitesse (vitesses limites disponibles sur [OpenStreetMap](#)). Vous pourriez aussi penser à une autre façon de dériver les conditions de déplacement sur les tronçons. Cette dernière sera votre variable dépendante, soit la variable à décrire par le modèle.

De plus, avant de procéder au développement du modèle, vous devez enrichir les données avec des conditions météorologiques ([Données historiques - Gouvernement du Canada](#)) et des caractéristiques pertinentes des segments routiers comme le nombre de voies, son orientation (points cardinaux), le nombre de carrefours, la distance au centre-ville, etc. Expliquez bien les démarches afin d'ajouter ces nouvelles caractéristiques aux tronçons sélectionnés pour l'analyse. Vous devrez ainsi avoir au minimum une variable associée à la température, une associée au déplacement et une dernière associée au segment routier dans votre modèle. Cela veut dire que vous devrez ajouter plusieurs variables de chaque type et les tester dans le modèle.

Ainsi, vous serez en mesure d'étudier les facteurs temporels, météorologiques et les caractéristiques des segments associés aux conditions de déplacement sur l'ensemble des segments choisis à l'aide d'un modèle de régression linéaire. Il est important que vous décriviez/illustriez votre processus de développement et de vérification de votre modèle. Enfin, commentez votre modèle final (notamment la significativité et le signe des coefficients) et si les conditions d'estimations de ce dernier sont vérifiées (notamment, concernant les résidus).

3 Modalités

Ce travail se fait avec les mêmes équipes qu'au précédent travail. Veuillez contacter le chargé de laboratoire en cas de problème. Un rapport au format PDF ou Word, ne dépassant pas **20 pages**, doit faire état des mandats de ce travail pratique. La date de rendu est le **9 novembre 2023 à 23h59**. Le fichier doit être déposé en format électronique sur moodle. En cas de besoin, vous pouvez par exemple déposer un fichier Excel contenant vos différents modèles linéaires et les figures d'analyse s'y rattachant.

Le nom du fichier doit porter la nomenclature suivante : EQ{numéro d'équipe}-TP{numéro du TP}-{semestre d'étude (A, H ou E)}{année d'étude}. Par exemple, "EQ01-TP1_A23".

Une attention particulière sera portée à la rédaction (les fautes de français seront sanctionnées tout comme une mauvaise organisation générale du travail), comptant pour un point (5%) sur la note finale.

Veillez consulter le [Guide de rédaction pour ingénieur civil](#) disponible sur Moodle à la section Ressources.