

a) Publications

Contributions sans comité de lecture

Binette, O.*, Coache, A.* (2018, septembre) The Significance of the Adjusted R Squared. (Bio)statistics Research and Career Day, Montréal. [Affiche](#).

Coache, A.*, Larose, F.* (2018, juin) "Do schools kill creativity?" Well, they help analyse popularity! Congrès annuel de la SSC, Montréal. [Affiche](#).

Coache, A. (2017, juillet) Estimation non-paramétrique de la fonction quantile. Séminaire étudiant de probabilités et statistique, Montréal. [Exposé oral](#). Organisé par l'Université du Québec À Montréal.

Ferland, R.*, Froda, S.*, Coache, A. (2017, juin) Comparison of surveillance flu data across regions. Congrès annuel de la SSC, Winnipeg. [Affiche](#).

b) Participation à des projets de recherche et à des stages de formation

L'affiche « The Significance of the Adjusted R Squared », réalisé en collaboration avec mon collègue étudiant Olivier Binette, s'inscrit dans un projet de recherche débuté de façon parallèle à mes travaux de maîtrise. Dans le contexte d'un modèle de régression linéaire, une des statistiques pertinentes est le R carré ajusté (R_a^2). Plusieurs versions de cette mesure existent dans la littérature, mais les explications précisant quelle forme privilégier sont d'une portée très limitée. Notre recherche conjointe parvient à soumettre une nouvelle interprétation plus intuitive de ce coefficient et, par le fait même, à considérer des tests exacts de sélection de modèles linéaires généralisés emboîtés. L'idée principale réside dans le fait que la valeur espérée du R_a^2 est invariante sous l'addition de variables considérée comme du bruit. Ces nouveaux résultats permettent de donner une perspective plus pédagogique pour l'étude de ce coefficient et peuvent éclairer les chercheurs dans l'interprétation de leurs analyses de données. Ce projet m'a permis de développer mon autonomie ainsi que ma polyvalence pour mener plusieurs projets parallèles à terme. L'affiche a été présentée par moi et Olivier à l'université McGill dans le cadre d'un congrès sur la recherche en statistique et en biostatistique.

J'ai participé à l'été 2018 au concours d'affiches d'études de cas en analyse de données dans le cadre du congrès annuel de la Société Statistique du Canada (SSC) présenté à Montréal. Ce projet, réalisé en collaboration avec ma collègue étudiante Florence Larose, consistait à analyser un jeu de données réel provenant du site web de TED, une organisation sans but lucratif ni affiliation politique, pour étudier les caractéristiques qui contribuent à la popularité des différentes présentations. En utilisant des techniques d'extraction du contenu de sites Web, communément appelé web scraping, pour obtenir certaines données supplémentaires, nous avons défini une mesure de popularité composite pour quantifier la popularité des vidéos. De plus, en appliquant un algorithme de modélisation de sujet, soit l'allocation de Dirichlet latente, nous sommes parvenus à classer les présentations sous différentes catégories, allant de la génétique à la politique. En procédant à différentes méthodes d'apprentissage machine, telles que les forêts aléatoires, certaines caractéristiques permettent de déceler une présentation propice à être populaire, dont le sujet du vidéo. Notre affiche, nommée « "Do schools kill creativity?" Well, they help analyse popularity! », résume

notre démarche ainsi que nos pistes de réflexion. Ce concours m'a permis de m'introduire à des techniques d'apprentissage machine et de classification. De plus, j'ai pu appliquer mes connaissances acquises lors de mon parcours universitaire à de véritables cas et en apprendre davantage, comme le langage de programmation Python.

L'exposé oral « Estimation non-paramétrique de la fonction quantile » concerne l'estimation de quantiles de distributions diverses. Elle intervient entre autres dans la gestion de portefeuille par l'entremise des mesures de risque. En effet, les mesures de risque bien connues comme la valeur-à-risque et la perte moyenne outre-seuil (expected-shortfall) nécessitent de calculer les quantiles de la distribution statistique des pertes. Or, le calcul analytique de ces quantiles peut être difficile du fait que la loi conjointe des actifs du portefeuille est rendue complexe par la nécessité de reproduire le comportement non gaussien du marché pour les actifs. Le point de départ du projet est d'utiliser la simulation de Monte Carlo pour résoudre partiellement le problème en donnant une estimation des quantiles, même si, par la suite, la sensibilité des mesures aux variations des positions du portefeuille deviennent à leur tour difficile à calculer. Mon intérêt s'est porté sur deux estimateurs non-paramétriques asymptotiquement convergents de la fonction quantile, soit l'estimateur Harrell-Davis et l'estimateur par le noyau, pour lesquels une revue de littérature exhaustive a été effectuée. À travers plusieurs études de simulation, j'ai pu quantifier l'importance du gain de performance en préconisant ces estimateurs plutôt que l'estimateur non-paramétrique traditionnel. La recherche était sous la supervision des professeurs François Watier et René Ferland lors d'un stage se déroulant à l'été 2016.

J'ai aussi contribué à un projet en épidémiologie en collaboration avec les professeurs Sorana Froda et René Ferland. L'affiche « Comparison of surveillance flu data across regions », présentée par Sorana Froda et René Ferland à Winnipeg dans le cadre du congrès annuel de la SSC en juin 2017, prend comme point de départ un jeu de données de syndromes grippaux récoltés à travers les États-Unis depuis 1997 par les Centers for Disease Control and Prevention (CDC). On vise à comparer ces données entre les différentes régions géographiques ainsi qu'entre les années. Plutôt que de comparer les taux d'incidence de la maladie à travers des séries temporelles, on considère la distribution relative des nouveaux cas où la proportion de nouveaux cas dépend de certaines variables, telles que l'âge. Cette approche alternative permet de prendre avantage de la panoplie de variables explicatives disponibles lors des analyses. En plus de me familiariser au domaine de l'épidémiologie et aux techniques déjà existantes dans la littérature, j'ai effectué un nettoyage de la base de données entière fournie par le CDC ainsi que la programmation de multiples scripts en SAS et en R pour l'analyse et la représentation des résultats, ce qui a été une autre belle occasion pour perfectionner mes compétences de programmation.

Finalement, lors de mon baccalauréat, j'ai participé à deux reprises aux concours de bourses de recherche de premier cycle offertes par le Conseil de Recherches en Sciences Naturelles et en Génie (CRSNG) lors des sessions d'étés 2016 et 2017, que j'ai gagnés. Ces bourses m'ont permis non seulement de m'initier à la recherche en milieu académique, mais aussi d'apprendre une multitude de connaissances mathématiques. Les projets étudiés lors des sessions, sous la supervision de François Watier, m'ont donné la possibilité de pouvoir contribuer à l'avancement des recherches dans plusieurs domaines en pleine effervescence, tels que la finance et l'épidémiologie. En effet, tel que mentionné précédemment, un des projets m'a permis de collaborer avec les professeurs René Ferland et Sorana Froda à l'élaboration d'une affiche pour le congrès annuel de la SSC. Aux dires de mon superviseur, ces stages m'auront bien préparé à poursuivre des études supérieures. Étant une personne ambitieuse croyant que j'ai encore beaucoup de connaissances à apprendre et d'aptitudes à parfaire, ces stages de recherche ont été des occasions parfaites pour assouvir ma soif de

connaissance, me transmettre un engouement pour la recherche en statistique ainsi que me faire comprendre que je souhaite vivement oeuvrer dans un domaine de recherche aussi stimulant que ceux abordés lors de ces étés.

c) Autres informations pertinentes

J'ai été auxiliaire d'enseignement pour plusieurs cours au niveau du baccalauréat à l'Université du Québec À Montréal (UQAM), soit modèles de régression (STT2120) à une reprise, laboratoire de statistique (STT2100) à deux reprises et méthodes statistiques (MAT2080) à cinq reprises, pour en faire la liste exhaustive. Ces charges de démonstrations se sont déroulées lors de la fin de mon baccalauréat et tout au long de ma maîtrise. De plus, j'ai été tuteur privé pour certains cours en mathématiques au cégep et un cours de statistique appliquée à l'université, plus particulièrement le cours d'analyse quantitative en psychologie (PSY4031), donné encore une fois à l'UQAM. Cette expérience d'enseignement combiné au tutorat a été bénéfique pour mon parcours académique, puisque non seulement j'ai pu parfaire mes compétences de pédagogie en plus de confirmer mon intérêt pour l'enseignement, j'ai aussi pu perfectionner mes habiletés de communication et de vulgarisation scientifique.

L'excellence de mon dossier académique n'est pas passée inaperçue au sein du département de mathématiques de l'UQAM, puisque j'ai été nommé à maintes reprises sur la liste d'excellence du doyen de la Faculté des sciences. Cette distinction, remise à des étudiants à temps plein présentant des résultats remarquables, m'a été décernée lors de quatre sessions universitaires sur une possibilité de six lors de mon baccalauréat, tel que mentionné dans une section précédente. De plus, j'ai obtenu une mention d'honneur lors de la remise de mon diplôme pour l'excellence de mes résultats lors du baccalauréat en mathématiques, concentration statistique.

Finalement, il n'y a eu aucune circonstance particulière lors de mon parcours universitaire, c'est-à-dire que j'ai suivi le cheminement typique d'un étudiant à temps plein au baccalauréat. J'ai complété mes études de premier cycle sans aucune interruption. Effectivement, les 90 crédits nécessaires à l'obtention de mon baccalauréat ont été obtenus en six sessions universitaires à temps plein. En ce qui a trait à ma maîtrise, je suis en voie de la terminer aussi sans retard, puisque j'ai complété la totalité des cours requis lors de la première année et que mes travaux de recherche sont déjà bien entamés. Les cours auxquels j'ai assisté, autant en tant qu'étudiant inscrit qu'auditeur libre, abordaient plusieurs sujets importants en statistique, comme la simulation, la statistique spatiale, les processus stochastiques ainsi que les modèles de régression, pour n'en nommer que quelques-uns. Ces sujets variés m'ont procuré une vue d'ensemble d'une panoplie de techniques en plus de m'introduire à la multitude d'opportunités de recherche dans de nombreux domaines.