

# Programme-cadre canadien pour les programmes de premier cycle en statistique

---

*Karen Buro, MacEwan University*  
*Sotirios Damouras, University of Toronto Scarborough*  
*Alison Gibbs, University of Toronto*  
*John Petkau, University of British Columbia*  
*John Sheriff, University of Lethbridge*  
*Jim Stallard, University of Calgary*  
*Bethany White, University of Toronto*

## Contexte et motivation

La science statistique connaît actuellement une forte croissance à la fois quant à sa portée et à sa pratique. La disponibilité accrue de données massives et d'une puissance de calcul sans précédent a ouvert de nouvelles voies très prometteuses pour la discipline. Mais ces progrès créent, partant, de nouvelles attentes par rapport à l'éducation et la formation des futurs statisticiens. Si nous voulons que nos diplômés de premier cycle puissent aspirer à des postes de statisticiens, de scientifiques de données ou d'analystes, nous devons évoluer avec la discipline et mettre à jour les programmes d'enseignement sans plus tarder.

En réponse à ces défis, et suite à une initiative similaire de la American Statistical Association (voir <http://www.amstat.org/education/curriculumguidelines.cfm>), un atelier a été organisé en 2015 sur le sujet *Comment faire progresser l'innovation et l'érudition en éducation statistique au Canada*. Celui-ci a débouché sur la création d'un groupe de travail chargé d'élaborer des lignes directrices pour les programmes de premier cycle en statistique au Canada. Les lignes directrices présentées ici ont été élaborées en consultation avec divers intervenants (membres de la Société statistique du Canada, anciens et actuels étudiants, interlocuteurs industriels, etc.) et sont présentées sous la forme d'objectifs d'apprentissage recommandés et de sujets et compétences suggérés. Elles représentent une vision actuelle des compétences professionnelles que devraient posséder nos étudiants de premier cycle à l'obtention de leur diplôme pour être prêts à assumer des rôles professionnels divers et variés.

## Domaines fondamentaux

Pour préparer les étudiants de premier cycle en statistique à des carrières polyvalentes dans des domaines existants et émergents, nous avons identifié quatre domaines fondamentaux à développer dans les programmes d'études :

- *Méthodes statistiques* : La trousse à outils du statisticien s'est considérablement agrandie ces dernières années. La modélisation prédictive et les données d'observation occupent aujourd'hui autant de place dans le travail des statisticiens que les méthodes inférentielles/explicatives et les données expérimentales. Les diplômés en statistique doivent pouvoir sélectionner, appliquer et évaluer les méthodes statistiques adaptées à leurs activités.
- *Analyse informatique de données* : L'analyse pratique de données s'appuie de plus en plus sur le calcul à grande échelle. À mesure que s'améliorent la fonctionnalité et la disponibilité des technologies existantes et que de nouveaux outils technologiques apparaissent, les diplômés en statistique doivent être habiles au calcul statistique, conscients de son évolution et capables de s'y adapter. Cela exige de solides bases en science informatique et une préparation à la réflexion algorithmique, à la programmation en un langage évolué et à de bonnes compétences en gestion et traitement de données.
- *Pratique statistique* : Les diplômés de premier cycle en statistique doivent avoir eu de nombreuses occasions d'appliquer leurs connaissances, de pratiquer et perfectionner leurs compétences et de développer les attributs de statisticiens praticiens efficaces par un apprentissage authentique dans des conditions réalistes. Ils doivent être capables de résoudre des problèmes et de traiter de vraies questions en collectant et analysant de vraies données, en communiquant efficacement leurs résultats

et en employant des pratiques exemplaires et des normes éthiques dans le contexte de problèmes du monde réel.

- **Fondements théoriques** : Les diplômés de premier cycle en statistique doivent maîtriser les outils mathématiques utilisés en statistique et comprendre les bases théoriques de l'inférence et de la prévision statistiques.

## Comment utiliser ces lignes directrices

Les présentes lignes directrices contiennent des suggestions de haut niveau pour les programmes et cours de premier cycle en statistique (p. ex., programmes de majeure en statistique). Elles sont exprimées en termes d'objectifs d'apprentissage et de compétences et sujets associés, regroupés selon les quatre domaines fondamentaux décrits ci-dessus. Elles ne se veulent pas normatives, mais ont été conçues dans l'espoir de constituer un cadre suffisamment souple pour être utile lors de la réforme des programmes et des discussions sur la conception de cours, quel que soit l'établissement ou le contexte. Nous considérons ces quatre domaines fondamentaux comme étant essentiels, mais ils n'auront pas forcément la même valeur de pondération et ils ne sont pas présentés ici par ordre d'importance. Par ailleurs, ces objectifs d'apprentissage ne seront pas forcément atteints dans le cadre des cours; certaines compétences et connaissances pourront être développées et renforcées par l'association, tout au long du programme, d'enseignements en classe, de projets pédagogiques, d'occasions de consultation/recherche et/ou de stages. Le groupe de travail espère que ces lignes directrices seront utiles aux institutions canadiennes alors qu'elles planifient leurs propres programmes en tenant compte de leurs forces universitaires, de leur effectif professoral, de leurs ressources et de leurs plans stratégiques.

## Objectifs d'apprentissage de premier cycle

**1. Méthodes statistiques** : À l'obtention de leur diplôme, les étudiants ayant complété un programme de premier cycle en statistique doivent être capables de bien appliquer, comparer et évaluer les méthodes de collecte et d'analyse de données. Ils doivent notamment savoir :

- 1.1 concevoir des méthodes pour la collecte de données permettant de répondre à des questions,
- 1.2. explorer des données pour comprendre leur comportement et leurs caractéristiques,
- 1.3. sélectionner les techniques et modèles appropriés pour traiter une question statistique donnée,
- 1.4. mettre en place diverses méthodes statistiques et tirer des conclusions de leurs résultats,
- 1.5. évaluer des hypothèses et reconnaître les limites de leurs choix concernant la collecte et l'analyse de données.

**2. Analyse informatique de données** : À l'obtention de leur diplôme, les étudiants ayant complété un programme de premier cycle en statistique doivent être capables d'organiser un grand projet d'analyse statistique dans l'environnement de programmation de leur choix. Ils doivent notamment savoir :

- 2.1. stocker et récupérer des données de divers types et formats et les manipuler efficacement,
- 2.2. explorer, visualiser et résumer des données efficacement et effectuer des tâches de modélisation, d'inférence et/ou de prévision statistique,
- 2.3. écrire du code efficace et bien documenté qui permette des recherches reproductibles, et notamment écrire des algorithmes et effectuer des simulations,
- 2.4. changer d'échelle de calcul pour traiter des données massives.

**3. Pratique statistique** : À l'obtention de leur diplôme, les étudiants ayant complété un programme de premier cycle en statistique doivent être capables de traduire des problèmes réels en questions statistiques et de communiquer leurs conclusions de manière accessible. Ils doivent notamment savoir :

- 3.1. distiller des questions statistiques à partir de l'information contextuelle du problème en question,
- 3.2. recommander des plans d'étude statistiques adaptés au traitement de problèmes réels,
- 3.3. communiquer clairement leurs concepts, méthodes et résultats à des publics divers, verbalement ou par écrit,
- 3.4. fonctionner efficacement dans un environnement de collaboration et/ou interdisciplinaire,
- 3.5. démontrer leur adhésion à la pratique éthique lors de la collecte et du traitement de données, de l'analyse statistique et de la présentation de leurs résultats [1].

**4. Fondements théoriques :** À l'obtention de leur diplôme, les étudiants ayant complété un programme de premier cycle en statistique doivent posséder la sophistication théorique requise pour bien évaluer et interpréter les modèles et méthodes statistiques. Ils doivent notamment savoir :

- 4.1. appliquer les résultats fondamentaux du calcul et de l'algèbre linéaire,
- 4.2. utiliser la théorie des probabilités pour décrire et analyser des phénomènes aléatoires,
- 4.3. identifier les enjeux importants qui se présentent lors de la modélisation et de l'analyse de données (p. ex., données aléatoires, sources de variation, association et causalité, incertitude des modèles) et comprendre les fondements et les implications de différentes approches de l'analyse de données.

[1] Conformément au Code de déontologie statistique de la Société statistique du Canada (SSC) et aux exigences éthiques décrites dans l'Énoncé de politique des trois Conseils : Éthique de la recherche avec des êtres humains (EPTC) et dans les normes du Conseil canadien de protection des animaux (CCPA).

## Sujets et compétences suggérés

**1. Méthodes statistiques :** Méthodes d'analyse de données descriptives, inférentielles et prédictives.

- **Collecte de données :** Données expérimentales et d'observation, méthodes de collecte de données expérimentales, d'enquête et d'observation.
- **Statistique descriptive :** Résumés graphiques et numériques de données permettant l'exploration de différents types de variables et de leurs relations.
- **Modélisation :** Modélisation des relations entre variables de réponse et explicatives (régression linéaire simple, multiple et généralisée), modèles pour données multivariées et de séries chronologiques, méthodes de simulation, méthodes non paramétriques et d'apprentissage machine pour la classification et la prévision (arbres de décision, régression non paramétrique, réseaux neuronaux, machines à vecteurs de support).
- **Statistique inférentielle :** Estimation et vérification d'hypothèses pour paramètres de modèles statistiques.

**2. Analyse informatique de données :** Outils informatiques nécessaires à l'analyse de données réelles volumineuses et/ou complexes.

- **Traitement de données :** Outils permettant d'accéder à des données volumineuses structurées (SQL) et non structurées (XML, texte, pages Web, etc.) et de les manipuler.
- **Logiciels statistiques :** Maîtrise d'au moins une langue ou un environnement d'analyse de données (R, SAS, Python, etc.).
- **Programmation :** Concepts de base de la programmation (p. ex., structures de contrôle, récursion, algorithmes de structures de données élémentaires, analyse des temps de parcours) et méthodes de calcul intensif.
- **Scalabilité :** Outils permettant le stockage et le traitement parallèles/distribués de données (p. ex., Hadoop, Spark).

**3. Pratique statistique :** Possibilités régulières et fréquentes pour les étudiants de mettre en pratique leurs connaissances et compétences statistiques, en particulier via l'apprentissage expérientiel.

- **Expérience pratique :** Analyse de données réelles volumineuses et désordonnées dans le cadre de projets de couronnement, de stages, d'occasions de consultation ou de projets de recherche interdisciplinaire.
- **Communication :** Compétences en rédaction technique et en présentation orale, possibilités de collaboration et de travail collectif avec des collaborateurs en statistique et autres disciplines.
- **Maîtrise de la statistique :** Évaluation critique de rapports statistiques : évaluation des hypothèses, identification des forces/faiblesses et évaluation des conclusions et de l'impact.
- **Pratiques exemplaires :** Élaboration et promotion de normes professionnelles et éthiques pour la pratique statistique.

**4. Fondements théoriques** : Connaissances en mathématiques nécessaires à l'élaboration et à l'analyse de modèles statistiques et de leurs propriétés.

- **Calcul et algèbre linéaire** : Différentiation et intégration sur plusieurs variables, opérations matricielles, systèmes linéaires, espaces euclidiens et décompositions de matrices.
- **Probabilité** : Distribution de probabilités discrètes et continues sur plusieurs variables, (in)dépendance et conditionnement, processus stochastiques.
- **Théorie de l'inférence statistique** : Méthode scientifique, inférence causale et descriptive, modélisation prédictive, concepts de vraisemblance, perspectives fréquentistes et bayésiennes, approches paramétriques et non paramétriques