Université du Québec À Montréal — Automne 2020
Faculté des Sciences — Département de Mathématiques

STT-3100.MAT-8081 Analyse Multivariée (3 CR.)

Groupe : 20-20

Professeur : Karim Oualkacha

Bureau : PK-5440

Courriel : oualkacha.karim@uqam.ca

Page Web : http://www.karimoualkacha.uqam.ca/

Heures de cours : Le mardi de 11h00 à 12h30, **en line, via zoom**

Le mardi de 13h30 à 15h00, **en line, via zoom**

RENSEIGNEMENTS GÉNÉRAUX :

Ouvrages de références

- Les notes de cours seront données en partie au tableau en classe et sous forme de diapositives qui seront disponibles sur le site web du cours. Autres documents complémentaires seront aussi disponibles au besoin sur le site web du cours. Vous aurez accès au site web du cours via Moodle : https://www.**moodle2.uqam**.ca/
- Livre de référence principale: Les notes de cours sont basées sur le livre "Methods of Multivariate Analysis", (2001), par Alvin C. Rencher, ISBN 0-471-41889-7, QA278.R45 2001.

Notez qu'il n'est pas nécessaire de se procurer ce manuel pour suivre le cours. Cependant, si vous désirez des compléments aux notes, vous pourrez consulter ce manuel disponible à la bibliothèque, ou l'acheter à la librairie scientifique de la COOP UQAM. Ce livre sera aussi réservé pour consultation à la bibliothèque des sciences du pavillon PK dès le début de la session.

Une copie intégrale en format PDF du livre est disponible via internet.

- Autres livres de références (Algèbre Linéaire):
  - Notes de cous d'Algèbre Linéaire 1, MAT1250: http://www.lacim.uqam.ca/~christo/Algèbre_linéaire_1/Algèbre%20linéaire%201.pdf
  - "Matrix Algebra Useful for Statistics", [(1982), 1 ère édit. ou (2006), 2 ème édit.], par Shayle R. Searle.
  - "A Matrix Handbook for Statisticians", [(2007), 1 ère édit.], par George A. F. Seber.

Évaluation

- **Examens** : Deux examens comptant chacun pour 30% de la note finale.
- **Date des examens :** Examen 1 : mardi 27 octobre, de 11h00 à 13h00.

  Examen 2 : mardi 15 décembre, de 11h00 à 13h00.
- **Les devoirs :**
  - Deux devoirs comptant chacun pour 10% de la note finale. Le Devoir 1 portera sur la matière vue en classe durant les 7 premières semaines de cours. Le Devoir 2 portera sur la matière vue en classe entre le 3 novembre et le 8 décembre.
  - Dix questions/exercices hebdomadaires comptant chacune pour 2% de la note finale. Chaque question portera sur la matière vue en classe durant la semaine courante.

Les devoirs et les exercices hebdomadaires peuvent nécessiter l'usage du progiciel *SAS/R*.

Laboratoires : Il y aura quelques ateliers durant lesquelles les étudiants apprendront à utiliser le progiciel *SAS et/ou R*. Ces séances de laboratoire se tiendront de façon virtuelle via zoom, durant quelques séances du cours.

Objectifs Généraux : Le cours vise à

- familiariser l'étudiant avec les données multivariées qui impliquent un nombre *d* de variables mesurées sur n sujets regroupés dans un ou plusieurs groupes.
- familiariser l'étudiant avec l'utilisation du progiciel *SAS* pour appliquer les techniques de l'analyse multivariée sur des jeux de données multivariés réels.

Objectifs Spécifiques : À la fin du cours, l'étudiant devra être capable

- de calculer l'espérance et la variance de combinaisons linéaires d'un vecteur aléatoire qui suit quelques lois multivariées usuelles.
- de vérifier l'hypothèse de normalité multivariée d'un jeu de données construit de vecteurs de *d* mesures prises sur n sujets, à l'aide de *SAS/R*.
- d'ajuster un modèle de copules afin de capter la structure de dépendance non-linéaires entre deux variables.
- d'identifier et manipuler la distribution de Wishart et de Hotelling.
- d'effectuer les tests d'hypothèses multivariés à un échantillon et à deux échantillons manuellement et à l'aide de *SAS*.
- de tester avec *SAS* des hypothèses concernant la forme du vecteur de moyennes d'un vecteur aléatoire lorsque des mesures répétées prises sur un même individu.
- d'effectuer l'analyse de variance multidimensionnelle à un facteur avec *SAS*.

- de simplifier la structure de corrélation d'un jeu de données multivarié à l'aide de l'analyse des composantes principales (ACP).
- d'appliquer l'ACP sur un jeu de données réel en utilisant *SAS* et interpréter les résultats de *SAS*.
- d'effectuer une analyse discriminante sur une population formée de q groupes et utiliser *SAS* pour trouver les fonctions discriminantes.
- de faire l'analyse hiérarchique à l'aide de SAS pour regrouper n sujets dans un certain nombre de groupes ou classes homogènes.

CONTENU DU COURS

1. Algèbre linéaire (environ 1 semaine)

Matrices et vecteurs, Operations sur les matrices, Propriétés du produit matriciel, Déterminant, trace et inverse d'une matrice carrée. Vecteurs linéairement indépendants et rang d'une matrice. Valeurs propres et vecteurs propres d'une matrice. Écriture de la moyenne et de la variance échantillonnale d'un échantillon multivarié.

2. La loi normale multivariée (environ 1 semaine)

Densité de la loi normale multivariée. Quelques propriétés : standardisation, additivité, transformation linéaire, indépendance, distribution marginale et conditionnelle, forme quadratique. Exemples et illustrations graphiques.

3. Tests d'hypothèses multivariés (environ 2 semaines)

Test d'hypothèse multivarié à un échantillon avec matrice de variance-covariance connue. Test d'hypothèse multivarié à un échantillon avec matrice de variance-covariance inconnue (distribution échantillonnales : la loi de Hotelling et de Wishart). Test d'hypothèse multivarié à deux échantillons. Test d'égalité des matrices des variance-covariance et test de normalité multivariée. Exemples et illustration avec *SAS*.

4. Analyse de profil à un échantillon (environ 1 semaine)

Test d'hypothèse multivarié concernant la forme du vecteur de moyenne dans le cas des mesures répétées. Test de l'effet linéaire. Exemples et illustration avec *SAS*.

5. Analyse de variance multidimensionnelle à un facteur (environ 2 semaines)

Analyse de variance multidimensionnelle à un facteur : test d'hypothèse multivarié de k vecteurs moyens de k échantillons, décomposition des sommes de carrées totale. Exemples et illustration avec *SAS*.

Examen 1 : 11h00, 27 octobre 2020

6. **Analyse des composantes principales (environ 2 semaines)**

Mise en situation. Introduction à l'analyse des composantes principales. Quelques techniques algébriques pour calculer les composantes principales. Écriture matricielle, structure de corrélation, variation expliquée. Aspects géométrique et graphique de l'ACP. L'ACP avec *SAS*. Interprétation de l'ACP et choix du nombre de composantes. Quelques compléments.

7. **Analyse discriminante (environ 2 semaines)**

Mise en situation et objectifs de l'analyse discriminante. Échantillon de formation, matrice de données, identification des groupes, fonction discriminante, erreurs de classement. L'analyse discriminante pour deux groupes. L'analyse discriminante d'un point de vue stochastique, le cas de populations normales. Exemples et illustrations avec *SAS*.

8. **Introduction à l'analyse multivariée en présence des données de grandes dimensions (environ 1 semaine)**

Mise en situation et objectifs de l'analyse des composantes principales et de l'analyse discriminante dans le cas de données de grandes dimensions. Introduction aux techniques d'optimisation pour résoudre les problèmes sous-jacents. Exemples de l'utilisation dans un cadre pratique (analyse de données génétiques).

9. **Classification (environ 2 semaine)**

Mise en situation et objectifs de la classification. Lien avec l'analyse discriminante. Principe de la classification, mesure de distance, indice de similarité, exemple de distance, illustration avec des exemples. Stratégies de classification: algorithmes ascendants et descendants. Exemples de classification à l'aide de *SAS*.

Examen 2 : 11h00, 15 décembre 2020

PLAGIAT

Règlement no 18 sur les infractions de nature académique

Tout acte de plagiat, fraude, copiage, tricherie ou falsification de document commis par une étudiante, un étudiant, de même que toute participation à ces actes ou tentative de les commettre, à l'occasion d'un examen ou d'un travail faisant l'objet d'une évaluation ou dans toute autre circonstance, constituent une infraction au sens de ce règlement.

- La liste non limitative des infractions est définie comme suit :
- la substitution de personnes ;
- l'utilisation totale ou partielle du texte d'autrui en le faisant passer pour sien ou sans indication de référence ;
- la transmission d'un travail pour fins d'évaluation alors qu'il constitue essentiellement un travail qui a déjà été transmis pour fins d'évaluation académique à l'Université ou dans une autre institution d'enseignement, sauf avec l'accord préalable de l'enseignante, l'enseignant ;
- l'obtention par vol, manœuvre ou corruption de questions ou de réponses d'examen ou de tout autre document ou matériel non autorisés, ou encore d'une évaluation non méritée ;
- la possession ou l'utilisation, avant ou pendant un examen, de tout document non autorisé ;
- l'utilisation pendant un examen de la copie d'examen d'une autre personne ;
- l'obtention de toute aide non autorisée, qu'elle soit collective ou individuelle ;
- la falsification d'un document, notamment d'un document transmis par l'Université ou d'un document de l'Université transmis ou non à une tierce personne, quelles que soient les circonstances;
- la falsification de données de recherche dans un travail, notamment une thèse, un mémoire, un mémoire-création, un rapport de stage ou un rapport de recherche.

Les sanctions reliées à ces infractions sont précisées à l'article 3 du Règlement no 18.

Pour plus d'information sur les infractions académiques et comment les prévenir:
www.integrite.uqam.ca