Université du Québec À Montréal — Automne 2024
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, salle PK-6605
		Le mardi de 13h30 à 15h00, salle PK-6605

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://moodle.uqam.ca/
- Livre de référence principale: Les notes de cours sont basées sur le livre "Methods of Multivariate Analysis", (2012), par Alvin C. Rencher & William C. Christensen.

  Une copie intégrale en format PDF du livre est disponible via internet :
  https://doi.org/10.1002/9781118391686.fmatter.

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 numérique, ou l'acheter à la librairie scientifique de la COOP UQAM.

- Autres livres de références (Algèbre Linéaire):
  - "Matrix Algebra Useful for Statistics", [(2017), 2ème édit.], par Shayle R. Searle & Andre I. Khuri.
  - "A Matrix Handbook for Statisticians", [(2007), 1 ère édit.], par George A. F. Seber.

Évaluation

- **Examen**s : Deux examens comptant chacun pour 35% de la note finale.
- **Date des examens :** Examen 1 : mardi 22 octobre, de **11h00** à **13h00**.

  Examen 2 : mardi 10 décembre, de **11h00** à **13h00**.
- **Les devoirs :** Deux devoirs comptant chacun pour 15% 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 28 octobre et le 6 décembre. Les deux devoirs peuvent nécessiter l'usage du progiciel *SAS/R*.

Laboratoires : Il y aura quelques séances (habituellement une) de laboratoire informatique durant lesquelles les étudiants apprendront à utiliser le progiciel *SAS*. Ces séances de laboratoire se tiendront aux salles informatiques du LAMISS, Pavillon PK.

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 des progiciels *SAS* (*et R*) 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'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**, 22 octobre 2024

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. 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*.

9. 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).

Examen 2 : **11h00**, 10 Décembre 2024



« 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 »
(R18, art. 2.1, définition d'une infraction).

Liste non limitative des infractions mentionnées dans le R18 :

- la substitution de personnes ou l'usurpation d'identité (art. 2.2 a) ;
- le plagiat : l'utilisation totale ou partielle du texte d'autrui ou de la production d'autrui en le faisant passer pour sien ou sans indication de référence (art. 2.2 b) ;
- le recyclage/la réutilisation de travaux : le dépôt d'un travail aux fins d'évaluation alors que ce travail constitue en tout ou en partie un travail qui a déjà été soumis par la personne étudiante, aux fins d'évaluation académique à l'UQAM ou dans une autre institution d'enseignement, sauf avec l'accord préalable de la personne enseignante à qui ce travail est soumis (art. 2.2 c) ;
- la possession ou l'obtention par vol, manœuvre ou corruption de questions ou de réponses d'examen (art. 2.2 d) ;
- la possession ou l'utilisation de tout document ou matériel non autorisé préalablement, pendant un examen ou lors de la réalisation de travaux, incluant le recours aux outils informatiques ou moyens technologiques (art. 2.2 e) ;
- l'utilisation pendant un examen de la copie d'examen ou de tout autre matériel provenant d'une autre personne (art. 2.2 f) ;
- l'obtention de toute aide non autorisée, qu'elle soit collective ou individuelle (art. 2.2 g) ;
- l'obtention d'une évaluation non méritée notamment par corruption, chantage, intimidation ou toute forme de harcèlement ou la tentative d'obtenir une telle évaluation (art. 2.2 h) ;
- la falsification d'un document ou la création d'un faux document, notamment d'un document transmis à l'Université ou d'un document de l'Université transmis ou non à une tierce personne, quelles que soient les circonstances (art. 2.2 i) ;
- 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 (art. 2.2 j).

Les sanctions reliées à ces infractions sont précisées aux articles 3 et 5 du *Règlement nº 18 sur les infractions de nature académique*.

Pour éviter de vous exposer à des sanctions :

1. Consultez le site **r18.uqam.ca** pour plus d'information sur l'intégrité académique et le R18;
2. Développez les bonnes pratiques en matière de recherche documentaire et de rédaction des travaux via l'outil **Infosphère** et les **formations offertes par le Service des bibliothèques**.