

## Plan de cours STT 6515, analyse de données multivariées

Professeur : Martin Bilodeau, bureau 4229, martin.bilodeau@umontreal.ca.

Disponibilités : MA 15 :30-16 :30 J 15 :30-16 :30

Horaire des cours : MA 14 :00-15 :30 et J 14 :30-15 :30 4186 Pav. André-Aisenstadt

**Objectif :** Ce cours vise à introduire certaines méthodes d'analyse multivariée. Ces méthodes sont utiles en soi mais elles faciliteront également votre apprentissage que ce soit dans un cours de régression, d'apprentissage automatique où la réduction de la dimension du problème importe. Les analyses proposées reflètent mon choix personnel. Elles constituent tout de même un éventail de méthodes classiques et de méthodes plus récentes. Les tests d'hypothèses sur la moyenne et sur la matrice de covariance, l'analyse discriminante linéaire et quadratique, la régression multivariée sont autant d'exemples de méthodes classiques. L'analyse multivariée robuste regroupe plusieurs méthodes récentes qui font de plus en plus leur apparition dans les progiciels statistiques de même que les modèles graphiques gaussiens. L'utilisation de chaque méthode sera illustrée par l'analyse de données réelles en R.

### Contenu :

1. Distributions multivariées. Propriétés des distributions elliptiques. Fonction caractéristique et densité. Distributions marginales et conditionnelles. Transformations linéaires affines. Distribution multinormale, Student multivariée, puissance exponentielle. Mélanges de distributions multinormales.
2. Estimateurs sans biais de la moyenne et de la matrice de covariance. Estimateurs de vraisemblance maximale pour les distributions elliptiques. Estimateurs robustes équivariants. Identification des observations aberrantes verticales, bons et mauvais points de levier.
3. Coefficients de corrélation simple. Corrélations multiples, partielles et canoniques. Modèles graphique gaussiens.
4. Test de Hotelling sur la moyenne. Intervalles de confiance simultanés. Intervalles de Roy-Bose, Scheffé et Bonferroni. Tests sur la matrice de covariance. Tests du rapport de vraisemblance avec correction pour l'aplatissement. Robustesse des différents tests.
5. Analyse discriminante. Analyse discriminante linéaire et quadratique. Taux d'erreur. Validation croisée. Probabilités a priori et a posteriori des groupes. Règles de Bayes. Analyse discriminante régularisée.
6. Méthodes de rééchantillonnage en analyse multivariée. Tests bootstrap et de permutation non-paramétriques sur la moyenne et la matrice de covariance. Tests d'indépendance.
7. Analyse en composantes principales.
8. Régression multivariée.

### Évaluation :

Deux travaux pratiques 20%

intra 35%

final 45%

Les étudiants inscrits au Bureau de Soutien aux Étudiants en Situation de Handicap (BSESH) désirant bénéficier de mesures d'accommodement aux examens (intra et final) sont priés de consulter le lien suivant pour connaître la procédure à suivre :

<https://safire.umontreal.ca/reussite-et-ressources/mesures-daccommodement-aux-examens-pour-les-etudiants-en-situation-de-handicap/>

### Référence :

- Anderson, T. W. (1984) An introduction to multivariate statistical analysis (2nd edition), Wiley Series in Probability and Mathematical Statistics, ISBN 0-471-88987-3.
- Bilodeau, M., Brenner, D. (1999) Theory of multivariate statistics, Springer texts in statistics, ISBN 0-387-98739-8.
- Flury, B. (1997) A first course in multivariate statistics, Springer texts in statistics, ISBN 0-387-98206-X.
- Johnson, R. A. and Wichern, D. W. (2007) Applied multivariate statistical analysis (6th edition), Prentice Hall, ISBN 0-13-187715-3.
- Kres, H. (1983) Statistical tables for multivariate analysis, a handbook with references to applications, Springer-Verlag, ISBN 0-387-90909-5.
- Rencher, A. C. (1995) Methods of multivariate analysis, Wiley Series in Probability and Mathematical Statistics, ISBN 0-471-57152-0.
- Rencher, A. C. (1998) Multivariate statistical inference and applications, Wiley Series in Probability and Statistics, ISBN 0-471-57151-2.
- Rousseeuw, P. J., Leroy, A. M. (1987) Robust regression and outlier detection, Wiley Series in Probability and Mathematical Statistics, ISBN 0-471-85233-3.
- Srivastava, M. S. and Khatri, C. G. (1979). An introduction to multivariate statistics, North Holland.
- Srivastava, M. S. (2002) Methods of multivariate statistics, Wiley Series in Probability and Statistics, ISBN 0-471-22381-6.

### Quelques rappels :

1. La date limite pour abandonner un cours sans frais, soit le 22 janvier.
2. La date limite pour abandonner un cours avec frais, soit le 13 mars.
3. L'obligation pour l'étudiant de motiver une absence prévisible à une évaluation dès qu'il est en mesure de constater qu'il ne pourra être présent, il appartiendra à l'autorité compétente de déterminer si le motif est acceptable (article 9.9).
4. Le plagiat : attention, c'est sérieux ! L'étudiant est invité à consulter le site [www.integrite.umontreal.ca](http://www.integrite.umontreal.ca).