

## Sciences de la décision

### Logiciels statistiques pour l'analyse de données - 6-613-11(Public)

A2016

Groupe 00

#### Enseignant(s)

Sarah Legendre Bilodeau  
Chargé(e) de cours  
[sarah.legendre-bilodeau@hec.ca](mailto:sarah.legendre-bilodeau@hec.ca)  
Bureau : 4.810  
Sections S01 et S02

#### Secrétaire(s)

Jennifer Caron  
Secrétaire  
[jennifer.caron@hec.ca](mailto:jennifer.caron@hec.ca)  
514-340-6473  
Bureau : 4.632  
Disponibilité : Du lundi au vendredi  
de 8h à 12h et de 13h à 16h30

Diane Brousseau  
Secrétaire  
[diane.brousseau@hec.ca](mailto:diane.brousseau@hec.ca)  
514-340-6472  
Bureau : 4.632  
Disponibilité : Du lundi au vendredi  
de 8h à 12h et de 13h à 16h30

#### Présentation du cours

##### Objectifs

L'étudiant apprendra à programmer en SAS et en R afin de nettoyer des jeux de données, de les représenter graphiquement et d'en faire une analyse statistique complexe. En plus de maîtriser le code de base de SAS, l'étudiant apprendra la syntaxe du module ODS qui permet de gérer le contenu des sorties. Il apprendra aussi le langage macro de SAS et s'en servira afin de créer des fonctions permettant des analyses statistiques supplémentaires et des traitements de données optimisés. En R, l'étudiant apprendra les bases du langage qui lui serviront à faire une analyse statistique des données. Il apprendra aussi à écrire des fonctions permettant l'analyse statistique de données et à construire une librairie de fonctions afin de partager les outils d'analyse qu'il aura codés. L'étudiant sera également initié au logiciel Tableau afin d'être en mesure de créer des visualisations de données de base. R et SAS sont basés sur des langages de programmation différents que l'étudiant devra apprendre à maîtriser.

##### Approche pédagogique

Les cours auront lieu dans un laboratoire informatique. Chaque période consistera en plusieurs blocs formés d'une présentation magistrale suivie de travail individuel sur les postes informatiques afin que les étudiants expérimentent sans délai les concepts présentés. Des exercices à faire en classe seront fournis.

## Matériel pédagogique

### Ressources bibliographiques

Braun, J. et al. (2007). *A first course in statistical programming with R*, Cambridge University Press Cambridge.

ISBN:9780511500206

 [Disponible à la bibliothèque](#)

Burlew, Michele M (2006). *SAS macro programming made easy, second edition*, Cary, N.C., SAS Institute. ISBN:9781590478820

 [Disponible à la bibliothèque](#)

Cody, Ronald P (2008). *Cody's data cleaning techniques using SAS, second edition*, Cary, N.C., SAS Institute Inc..

ISBN:9781599946597

 [Disponible à la bibliothèque](#)

Cody, Ronald P (2007). *Learning SAS by example a programmer's guide*, Cary, N.C., SAS Institute. ISBN:9781599941653

 [Disponible à la bibliothèque](#)

Delwiche, L.D. et al. (2008). *The little SAS book: a primer*, SAS Publishing.

ISBN:9781599947259

 [Disponible à la bibliothèque](#)

Maindonald, J.H. et al. (2010). *Data analysis and graphics using R: an example-based approach*, Cambridge University Press; 3 edition.

ISBN:9780521762939

 [Disponible à la bibliothèque](#)

Matloff, Norman S (2011). *The art of R programming : tour of statistical software design*, San Francisco, No Starch Press.

ISBN:9781593273842

 [Disponible à la bibliothèque](#)

Matloff, Norman S (2011). *The art of R programming a tour of statistical software design*, San Francisco, No Starch Press.

ISBN:1593273843

**(Version électronique)**

 [Disponible à la bibliothèque](#)

R Development Core Team (2010). *Writing R Extensions*, R Development Core Team.

 [Version libre de droit en ligne](#)

Refaat, Mamdouh (2007). *Data preparation for data mining using SAS*, Amsterdam, Morgan Kaufmann Publishers.

ISBN:0123735777

 [Disponible à la bibliothèque](#)

Robert, C. and Casella, G. (2011). *Methodes De Monte-Carlo Avec R*, Springer Verlag France.

ISBN:978-2-8178-0180-3

 [Disponible à la bibliothèque](#)

SAS Institute et al. (2009). *SAS certification prep guide base programming for SAS 9*, Cary, N.C. , SAS Institute.

ISBN:9781607640455

 [Disponible à la bibliothèque](#)

SAS Institute (2011). *SAS certification prep guide base programming for SAS 9, third edition*, Cary, N.C. , SAS Institute.

ISBN:9781607649243

 [Disponible à la bibliothèque](#)

Venables, W.N., Smith, D.M. and The R Development Core Team (2009). *An Introduction to R*, , R Development Core Team.

 [Version libre de droit en ligne](#)

## Évaluations

### Sommaire des évaluations

Travail individuel	20 %	
Projet d'équipe	30 %	
Examen final	50 %	Voir <i>HEC en ligne</i> pour date

### Travail individuel (20 %)

---

Individuel / À la maison - Hors classe / Écrit  
Mode de remise : Électronique

#### Description

Au cours de la session, quatre devoirs permettront de mettre en application la matière vue en classe. Le délai de remise sera d'au moins deux semaines. Chaque devoir compte pour 5%.

### Projet d'équipe (30 %)

---

En équipe / À la maison - Hors classe / Écrit  
Mode de remise : Électronique

#### Description

Le travail en équipe consistera en un projet de programmation R, par exemple :

- L'élaboration d'une simulation et l'analyse des résultats obtenus.
- L'analyse descriptive de données réelles.
- L'élaboration d'une bibliothèque de fonctions R.

Les attentes seront proportionnelles au nombre de membres dans l'équipe. Il n'est pas recommandé de former des équipes de plus de 4 ou 5 personnes.

Le projet devra permettre de mettre en application plusieurs éléments vus en classe et être approuvé au préalable.

## Examen final (50 %)

Voir *HEC en ligne* pour date

Individuel / En classe / Écrit  
Mode de remise : Électronique

## Organisation du cours

### Présentation du cours

#### 1 - Présentation du cours

##### Activités/Ressources après la séance

[Site web \(en anglais\) offrant une introduction interactive à R](http://tryR.codeschool.com/) [Document de nature pédagogique]  
(<http://tryR.codeschool.com/>)

[Chaîne YouTube présentant des vidéos d'apprentissage pour R \(en anglais\)](https://m.youtube.com/user/marinstatlectures) [Document de nature pédagogique]  
(<https://m.youtube.com/user/marinstatlectures>)

### Partie I : Apprentissage de R

#### 2 - Nettoyage, préparation de données et production de graphiques haute résolution

##### Description

Initiation à la syntaxe du langage R, à la définition de variables, aux opérations mathématiques, aux énoncés conditionnels et aux boucles. Calculs sur des matrices et utilisation des commandes « apply ».

Traitement des données manquantes, combinaison de jeux de données et production de graphiques haute résolution en format vectoriel pour l'analyse descriptive des données.

##### Activités/Ressources avant la séance

Braun, J. et al. (2007). *A first course in statistical programming with R*, Cambridge University Press  
Cambridge.

ISBN:9780511500206

**Chapitres 1 à 4**

 [Disponible à la bibliothèque](#)

Venables, W.N., Smith, D.M. and The R Development Core Team (2009). *An Introduction to R*, R Development Core Team.

**Chapitres 1 à 9, 12**

 [Version libre de droit en ligne](#)

---

### 3 - Packages et analyses statistiques avec R

---

#### Description

Présentation des modules d'analyse fréquemment utilisés. Installation de modules supplémentaires à partir du CRAN pour des analyses utilisant des méthodes nouvellement développées.

#### Activités/Ressources avant la séance

Maindonald, J.H. et al. (2010). *Data analysis and graphics using R: an example-based approach*, Cambridge University Press; 3 édition.

ISBN:9780521762939

 [Disponible à la bibliothèque](#)

Venables, W.N., Smith, D.M. and The R Development Core Team (2009). *An Introduction to R*, R Development Core Team.

**Chapitre 11**

 [Version libre de droit en ligne](#)

---

### 4 - Simulations, bootstrap et bibliothèques de fonctions

---

#### Description

Planification et programmation de scénarios simulés pour soutenir la prise de décisions. Structure, syntaxe et règles à suivre pour la création d'une bibliothèque de fonctions R.

#### Activités/Ressources avant la séance

Braun, J. et al. (2007). *A first course in statistical programming with R*, Cambridge University Press Cambridge.

ISBN:9780511500206

**Chapitre 5**

 [Disponible à la bibliothèque](#)

Robert, C. and Casella, G. (2011). *Methodes De Monte-Carlo Avec R*, Springer Verlag France.

ISBN:978-2-8178-0180-3

 [Disponible à la bibliothèque](#)

R Development Core Team (2010). *Writing R Extensions*, , R Development Core Team.

 [Version libre de droit en ligne](#)

Venables, W.N., Smith, D.M. and The R Development Core Team (2009). *An Introduction to R*, , R Development Core Team.

### Chapitre 13

 [Version libre de droit en ligne](#)

## Partie II : Apprentissage de SAS

### 5 - Base et nettoyage

#### Description

Syntaxe de base du langage SAS, création de jeux de données, importation de données, traitement des différents types de variables. Stratégies de validation et de nettoyage des données.

#### Activités/Ressources avant la séance

Cody, Ronald P (2007). *Learning SAS by example a programmer's guide*, Cary, N.C. , SAS Institute. ISBN:9781599941653

Chapitres 1 à 7

 [Disponible à la bibliothèque](#)

Delwiche, L.D. et al. (2008). *The little SAS book: a primer*, , SAS Publishing. ISBN:9781599947259

Chapitres 1 à 3

 [Disponible à la bibliothèque](#)

#### Ressources générales

Cody, Ronald P (2008). *Cody's data cleaning techniques using SAS, second edition* , Cary, N.C. , SAS Institute Inc.. ISBN:9781599946597

 [Disponible à la bibliothèque](#)

### 6 - Manipulations (1)

#### Description

Fusionner des fichiers de données (concaténer, juxtaposer). Ordonner des observations. Contrôler les attributs des variables et les fichiers de données produits. Gestion des dates, des variables numériques et des chaînes de caractères. Traitement des valeurs manquantes.

#### Activités/Ressources avant la séance

Cody, Ronald P (2007). *Learning SAS by example a programmer's guide*, Cary, N.C. , SAS Institute.

ISBN:9781599941653

Chapitres 9 à 12

 [Disponible à la bibliothèque](#)

Delwiche, L.D. et al. (2008). *The little SAS book: a primer*, , SAS Publishing.

ISBN:9781599947259

Chapitre 6

 [Disponible à la bibliothèque](#)

---

## 7 - Manipulations (2)

---

### Description

Séparation de jeux de données. Agréger des observations. Techniques pour transposer des données et faire référence à d'autres observations.

### Activités/Ressources avant la séance

Cody, Ronald P (2007). *Learning SAS by example a programmer's guide*, Cary, N.C. , SAS Institute.

ISBN:9781599941653

Chapitres 21 à 24

 [Disponible à la bibliothèque](#)

---

## 8 - Reporting

---

### Description

Utilisation des procédures permettant de présenter des résultats telles que MEANS, FREQ, TABULATE et REPORT et de leurs principales options. Présentation de la syntaxe du code relié aux commandes du module ODS et de possibilités d'interactions avec d'autres applications.

### Activités/Ressources avant la séance

Cody, Ronald P (2007). *Learning SAS by example a programmer's guide*, Cary, N.C. , SAS Institute.

ISBN:9781599941653

Chapitres 14 à 19

 [Disponible à la bibliothèque](#)

Delwiche, L.D. et al. (2008). *The little SAS book: a primer*, , SAS Publishing.

ISBN:9781599947259

Chapitres 4, 5 et 10

 [Disponible à la bibliothèque](#)

---

## 9 - Graphiques avec SAS/GRAPH - Initiation au logiciel Tableau

---

## Description

Utilisation de procédures graphiques de SAS/GRAPH telles que GPLOT et GCHART. Présentation de techniques de mise en page automatisées. Initiation au logiciel Tableau pour la création de visualisations de données dynamiques.

## Activités/Ressources avant la séance

Cody, Ronald P (2007). *Learning SAS by example a programmer's guide*, Cary, N.C. , SAS Institute.  
ISBN:9781599941653

Chapitre 20

 [Disponible à la bibliothèque](#)

Delwiche, L.D. et al. (2008). *The little SAS book: a primer*, , SAS Publishing.

ISBN:9781599947259

Chapitre 8

 [Disponible à la bibliothèque](#)

## 10 - Boucles et arrays

---

### Description

Utilisation des boucles et arrays dans SAS.

### Activités/Ressources avant la séance

Cody, Ronald P (2007). *Learning SAS by example a programmer's guide*, Cary, N.C. , SAS Institute.  
ISBN:9781599941653

Chapitres 8 et 13

 [Disponible à la bibliothèque](#)

## 11 - Variables macro

---

### Description

Introduction à la programmation à l'aide du langage macro de SAS. Utilisation des variables macro.

### Activités/Ressources avant la séance

Cody, Ronald P (2007). *Learning SAS by example a programmer's guide*, Cary, N.C. , SAS Institute.  
ISBN:9781599941653

Chapitre 25

 [Disponible à la bibliothèque](#)

Delwiche, L.D. et al. (2008). *The little SAS book: a primer*, , SAS Publishing.

ISBN:9781599947259

Chapitre 7

 [Disponible à la bibliothèque](#)

Burlew, Michele M (2006). *SAS macro programming made easy, second edition*, Cary, N.C. , SAS Institute.  
ISBN:9781590478820

Chapitres 1 à 3

 [Disponible à la bibliothèque](#)

## 12 - Fonctions macro

---

### Description

Programmation de fonctions macro et efficacité des programmes. Exemples de programmation macro en SAS.

### Activités/Ressources avant la séance

Burlew, Michele M (2006). *SAS macro programming made easy, second edition*, Cary, N.C. , SAS Institute.  
ISBN:9781590478820

Chapitres 4 à 8

 [Disponible à la bibliothèque](#)

## 13 - PROC SQL

---

### Description

Présentation de la syntaxe de base pour l'utilisation du langage SQL à l'intérieur d'une procédure SAS.

### Activités/Ressources avant la séance

Cody, Ronald P (2007). *Learning SAS by example a programmer's guide*, Cary, N.C. , SAS Institute.  
ISBN:9781599941653

Chapitre 26

 [Disponible à la bibliothèque](#)

## Règlements de HEC Montréal

### Plagiat

Les étudiants sont priés de prendre connaissance des actes et des gestes qui sont considérés comme étant du plagiat ou une autre infraction de nature pédagogique, de la procédure et des sanctions, qui peuvent aller jusqu'à la suspension et même l'expulsion de HEC Montréal. Toute infraction sera analysée en fonction des faits et des circonstances, et une sanction sera appliquée en conséquence. [En savoir plus sur le plagiat...](#)

### Calculatrices

Les étudiants sont priés de prendre connaissance de la politique d'utilisation de calculatrices lors d'examens lorsque celles-ci sont autorisées. [En savoir plus sur la politique d'usage de calculatrices...](#)

