

TABLE DES MATIÈRES

CHAPITRE 19	Régressions linéaire multiple, curvilinéaire et robuste	2
19.1	Régression linéaire multiple	2
19.1.1	Modèle de régression linéaire multiple	2
19.1.2	Estimation des coefficients de régression	3
19.1.3	Contribution du modèle à l'explication de Y	6
19.1.4	Inférences concernant les paramètres du modèle et une prédiction	8
19.1.5	Notion de corrélation partielle	10
19.1.6	Test de signification de la contribution marginale d'une variable	12
19.1.7	Contribution et effet des variables explicatives	14
19.1.8	Construction du modèle	19
19.1.8.1	Méthode descendante d'élimination des variables	19
19.1.8.2	Méthode ascendante de sélection des variables	19
19.1.8.3	Méthode progressive de sélection des variables	20
19.1.8.4	Méthode combinatoire de sélection des variables	20
19.1.9	Diagnostic et mesures de correction	24
19.1.9.1	Validité prédictive interne et externe d'un modèle	24
19.1.9.2	Manque d'ajustement du modèle	25
19.1.9.3	Normalité des résidus	27
19.1.9.4	Valeurs exceptionnelles, influentes et leviers	27
19.1.9.5	Constance de la variance des erreurs	31
19.1.9.6	Multicolinéarité	32
19.1.9.7	Indépendance des résidus	35
19.1.9.8	<i>Bootstrap</i>	35
19.2	Régression multiple avec interaction	40
19.3	Régression multiple pondérée	43
19.4	Régression orthogonale	45
19.5	Régression pseudo-orthogonale (ridge)	52
19.6	Régression partielle et des moindres carrés partiels	56
19.7	Régression multiple des moindres écarts absolus	60
19.8	Régression M	62
19.9	Régression multiple non paramétrique de rang	66
19.10	Régression polynomiale	67
19.11	Régression kernel	70

19.12	Régression polynomiale locale	75	20.4.3	Estimation par intervalle de moyennes ajustées et de contrastes pour les plans factoriels déséquilibrés	139
19.13	Régression spline cubique	75	20.5	Plan factoriel avec facteur aléatoire : modèles II et III	140
19.14	Régression non linéaire	80	20.5.1	Modèle I.....	140
19.14.1	Caractérisation du modèle.....	81	20.5.2	Modèle II	140
19.14.2	Estimation des paramètres du modèle.....	83	20.5.2.1	Plan factoriel équilibré à deux facteurs aléatoires	140
19.14.3	Sélection du modèle.....	85	20.5.2.2	Plan factoriel déséquilibré à deux facteurs aléatoires	143
19.14.4	Intervalle de confiance et tests de signification	86	20.5.3	Modèle III ou modèle mixte	149
19.14.4.1	Approche asymptotique	86	20.5.3.1	Plan factoriel équilibré à un facteur aléatoire et à un facteur fixé	150
19.14.4.2	Approche par rééchantillonnage	88	20.5.3.2	Plan factoriel déséquilibré avec facteurs aléatoires et fixés.....	151
RÉFÉRENCES.....		90	20.6	Plan factoriel à plus de deux facteurs	157
CHAPITRE 20 Analyse des plans factoriels et pyramidaux		91	20.7	Plan pyramidal	157
20.1	Plans à deux facteurs croisés et sans répétition	93	20.7.1	Plan pyramidal sans répétition.....	157
20.2	Plans factoriels équilibrés	98	20.7.2	Plan pyramidal équilibré avec répétition ..	158
20.3	Plans factoriels déséquilibrés	104	20.7.3	Plan pyramidal déséquilibré	163
20.3.1	Sources de déséquilibre et problème du déséquilibre.....	105	20.7.3.1	Modèle I : facteurs fixés	163
20.3.2	Plans factoriels déséquilibrés : approche des contrastes.....	107	20.7.3.2	Modèle II : facteurs aléatoires, modèle III : au moins un facteur aléatoire.....	167
20.3.2.1	Principe général de l'approche des contrastes.....	107	20.7.3.3	Codes SAS®	168
20.3.2.2	Test d'interaction	108	20.8	Diagnostic et mesures correctives	170
20.3.2.3	Comparaison de moyennes marginales pondérées par les effectifs	108	RÉFÉRENCES.....		172
20.3.2.4	Comparaison de moyennes marginales équipondérées	111	CHAPITRE 21 Analyse des plans en blocs, stratifiés en unités divisées et avec covariables		173
20.3.2.5	Comparaison de moyennes marginales pondérées par l'effectif efficace	113	21.1	Méthode des blocs	174
20.3.3	Plans factoriels déséquilibrés : approche des moindres carrés	115	21.1.1	Plans en blocs complets sans répétition ...	174
20.3.3.1	Modèle de régression	116	21.1.1.1	Analyse de variance	174
20.3.3.2	Test d'interaction	119	21.1.1.2	Test F non paramétrique et test de Friedman.....	176
20.3.3.3	Test d'interaction en présence de cases vides : analyse de type IV.....	120	21.1.2	Plans en blocs complets avec répétition...	179
20.3.3.4	Comparaison de moyennes marginales équipondérées : analyse de type III.....	120	21.1.3	Plans en blocs incomplets sans répétition.....	180
20.3.3.5	Comparaison de moyennes marginales pondérées par l'effectif efficace : analyse de type II	122	21.2	Stratification	181
20.3.3.6	Comparaison de moyennes marginales pondérées par l'effectif efficace : analyse de type I	123	21.3	Plan factoriel en unités divisées	185
20.3.3.7	Comparaison de moyennes marginales en présence de cases vides : analyse de type IV	124	21.3.1	Plan factoriel équilibré en unités divisées	186
20.3.3.8	Choix du type de somme de carrés	125	21.3.2	Plan factoriel déséquilibré en unités divisées	197
20.3.3.9	Programmation sous SAS®	131	21.4	Analyse de covariance (ANCOVA)	197
20.4	Inférences statistiques et modèle général linéaire	133	21.4.1	Choix des covariables et de y.....	198
20.4.1	Fonctions estimables : cas général.....	133	21.4.2	Moyennes ajustées et paradoxe de Simpson	199
20.4.2	Intervalle de confiance et test de signification d'une combinaison linéaire de paramètres.....	136	21.4.3	Modèle d'analyse de covariance.....	202
			21.4.4	Estimation des paramètres du modèle.....	203
			21.4.5	Contribution des termes du modèle	204
			21.4.6	Construction du modèle et tests d'intérêt	205
			21.4.7	Conditions d'application, diagnostic et remèdes.....	222

21.4.8	L'ANCOVA par régression ou à l'aide de SAS®	222	22.4.1	Intérêt du plan croisé	296
21.4.9	Comparaison de moyennes ajustées et tests de contraste	230	22.4.2	Définitions	297
21.4.9.1	Modèle 7 : Modèle classique d'analyse de covariance, droites parallèles	230	22.4.3	Randomisation et traitements	297
21.4.9.2	Modèle 8 : Droites de même origine	232	22.4.4	Analyse des plans croisés 2x2	298
21.4.9.3	Modèle 5 : Droites différentes	232	22.4.4.1	Analyse paramétrique par contrastes	298
	RÉFÉRENCES	236	22.4.4.2	Analyse non paramétrique par contrastes	301

CHAPITRE 22 Séries temporelles, plans en

groupes parallèles et croisés	237
22.1 Analyse d'une série chronologique	239
22.1.1 Graphique	239
22.1.2 Lissage et filtrage	239
22.1.2.1 Lissage par la moyenne mobile	240
22.1.2.2 Lissage par régression	240
22.1.2.3 Lissage par régression spline cubique	241
22.1.2.4 Lissage exponentiel	242
22.1.3 Extraction de tendances et recherche de stationnarité	244
22.1.3.1 Extraction de tendance par régression	244
22.1.3.2 Effacement de tendances par différenciation	245
22.1.3.3 Extraction de tendance par la moyenne mobile	245
22.1.3.4 Stabilisation de la variance	245
22.1.4 Corrélogrammes et variogrammes	247
22.1.4.1 Corrélogramme	247
22.1.4.2 Corrélogramme partiel	248
22.1.4.3 Variogramme	248
22.1.4.4 Spectres croisés	249
22.1.5 Estimation de la période d'une variation périodique	251
22.1.5.1 Périodogramme de Whittaker et Robinson	251
22.1.5.2 Périodogramme de Schuster	251
22.1.6 Modélisation d'une série chronologique à des fins de prévisions	254
22.1.6.1 Modèle de régression	254
22.1.6.2 Modèles de lissage exponentiel	255
22.1.6.3 Modèles AR, MA, ARMA et ARIMA	260
22.2 Analyse de n séries chronologiques	271
22.3 Plans en groupes parallèles	272
22.3.1 Analyse de variance pour plan factoriel en unités divisées	272
22.3.2 Analyse de variance multidimensionnelle (MANOVA)	278
22.3.3 Analyse de covariance (ANCOVA)	282
22.3.4 Analyse d'antédépendance	283
22.3.5 Modèle mixte	286
22.3.6 Modèles à effets fixes utilisant le maximum de vraisemblance	291
22.3.7 Modèles linéaires à coefficients aléatoires	292
22.4 Plans croisés	295

22.4.1	Intérêt du plan croisé	296
22.4.2	Définitions	297
22.4.3	Randomisation et traitements	297
22.4.4	Analyse des plans croisés 2x2	298
22.4.4.1	Analyse paramétrique par contrastes	298
22.4.4.2	Analyse non paramétrique par contrastes	301
22.4.4.3	Analyse de variance en mesures répétées	303
22.4.4.4	Analyse de variance à quatre facteurs	304
22.4.4.5	Analyse de variance avec le modèle mixte	305
22.4.4.6	Tests paramétriques ou non paramétriques pour échantillons appariés	306
22.4.4.7	Équivalence de termes	306
22.4.4.8	Interprétation des effets	307
21.4.4.9	Attitude en présence d'effet de rémanence et de tendances temporelles	307
22.4.5	Analyse des plans croisés $p \times p$	308
22.4.6	Analyse des plans croisés incomplets : $p \times h$ avec $h > p$	310

RÉFÉRENCES	311
------------	-----

CHAPITRE 23 Modélisation de données catégorielles, de dénombrement et non normales

23.1 Modèle linéaire généralisé	314
23.1.1 Fonction de lien	314
23.1.2 Estimation des paramètres du modèle linéaire généralisé	316
23.1.3 Choix et justesse du modèle	317
23.2 Modèle logistique	322
23.2.1 Modélisation des données binaires	322
23.2.2 Modèle logistique linéaire	324
23.2.3 Estimation des paramètres du modèle	324
23.2.4 Mise en œuvre du modèle logistique et interprétation des paramètres	325
23.2.4.1 Variable X binaire	325
23.2.4.2 Variable X qualitative polytomique	327
23.2.4.3 Variable X ordinale	328
23.2.4.4 Variable X quantitative	328
23.2.4.5 Interaction entre deux variables explicatives	329
23.2.4.6 Interaction mixte	332
23.2.5 Inférences et tests de signification des coefficients β et de p	333
23.2.5.1 Rapport de cotes et tests associés	333
23.2.5.2 Inférences basées sur la variance de B (test de Wald)	335
23.2.5.3 Test du rapport de vraisemblance	335
23.2.5.4 Intervalle de confiance d'une probabilité prédite par le modèle	336
23.2.5.5 Estimation du rapport de cotes à partir du modèle de régression logistique	337
23.2.5.6 Tests de contraste	337
23.2.6 Sélection du modèle	341

23.2.6.1	Test de signification du modèle : test du rapport de vraisemblance	341	23.6.4.1	Test de Wald	391
23.2.6.2	Tests d'adéquation du modèle.....	342	23.6.4.2	Test d'association marginale de Brown ...	391
23.2.6.3	Choix des termes du modèle	343	23.6.4.3	Test d'association partielle de Brown.....	391
23.2.7	Validité du modèle.....	349	23.6.5	Construction de modèle	395
23.2.7.1	Types de résidus.....	349	23.6.5.1	Sélection a priori d'un modèle à des fins d'analyse confirmative	395
23.2.7.2	Valeur levier.....	351	23.6.5.2	Méthode combinatoire	396
23.2.7.3	Mesure de la contribution de la i^{e} observation à la déviance ou au khi carré...352		23.6.5.3	Sélection descendante ou rétrograde des termes d'un modèle	396
23.2.7.4	Mesure de l'influence de la i^{e} observation sur le j^{e} paramètre.....	352	23.6.5.4	Sélection a posteriori des termes d'un modèle incluant des variables à expliquer et des facteurs dont on cherche l'effet	397
23.2.7.5	Mesure de l'influence de la i^{e} observation sur l'ensemble des paramètres.....	353	23.6.5.5	Sélection ascendante des termes d'un modèle	397
23.2.7.6	Mesure de la contribution de X à l'explication de Y	353	23.6.5.6	Choix de la méthode de sélection	398
23.2.7.7	Diagnostic	353	23.6.6	Diagnostic, validité du modèle et analyse des résidus	401
23.2.8	Emploi de la régression logistique dans les études cas-témoins.....	356	23.6.7	Problème des cases vides et de l'éparpillement.....	403
23.2.9	Calcul d'effectifs d'échantillons et de puissance.....	358	23.6.8	Relation avec le modèle logistique	404
23.3	Modèle logistique polytomique	359	23.6.9	Relation avec le rapport de cotes et le test de Cochran-Mantel-Haenszel	405
23.4	Modèle logistique ordinal (régression des cotes).....	363	23.6.10	Puissance et taille d'échantillon.....	407
23.4.1	Méthodes d'analyse de données catégorielles ordonnées.....	364	RÉFÉRENCES.....		408
23.4.2	Régression des cotes et test de Mann-Whitney	365	CHAPITRE 24 Analyse de survie		409
23.4.3	Modèle de régression des cotes et interprétation des paramètres.....	367	24.1	Estimation de la courbe de survie.....	412
23.4.4	Mise en œuvre du modèle de régression des cotes.....	368	24.1.1	Fonction empirique de survie.....	412
23.4.5	Justesse du modèle.....	371	24.1.2	Tableau de survie.....	412
23.4.6	Puissance.....	372	24.1.2.1	Estimation actuarielle de la survie	412
23.5	Régression de Poisson	373	24.1.3	Estimation de Kaplan-Meier	414
23.6	Modèle log-linéaire.....	375	24.1.4	Estimation de la médiane et des fractiles de temps de survie	416
23.6.1	Paramètres du modèle.....	375	24.2	Fonctions utilisées en analyse de survie.....	417
23.6.1.1	Tableau de contingence 2×2	377	24.2.1	Fonction de survie : $S_{(t)}$	417
23.6.1.2	Tableau de contingence $k_A \times k_B$	379	24.2.2	Fonction de répartition : $F_{(t)}$	417
23.6.1.3	Modèle log-linéaire hiérarchique et non hiérarchique	379	24.2.3	Fonction de densité de probabilité : $f_{(t)}$	418
23.6.1.4	Tableau de contingence $k_A \times k_B \times k_C$	380	24.2.4	Fonction de risque instantané : $h_{(t)}$	418
23.6.1.5	Tableau de contingence à plus de trois dimensions.....	384	24.2.5	Fonction de risque instantané cumulé : $H_{(t)}$	420
23.6.1.6	Système de notation.....	384	24.2.6	Notions de risques proportionnels et de risques relatifs instantanés	425
23.6.2	Estimation des paramètres du modèle.....	386	24.2.7	Fonction d'excès de risque.....	426
23.6.2.1	Estimateurs du maximum de vraisemblance et des moindres carrés.....	387	24.3	Comparaison de courbes de survie.....	426
23.6.2.2	Analyse de données catégorielles par la régression pondérée	388	24.3.1	Tests non paramétriques de comparaison de deux groupes	426
23.6.3	Mesures et tests d'adéquation du modèle.....	388	24.3.1.1	Test du log-rank	427
23.6.3.1	Khi carré de Pearson	388	24.3.1.2	Test de Gehan-Wilcoxon	428
23.6.3.2	Déviance et statistique du rapport de vraisemblance	389	24.3.1.3	Test de Tarone-Ware	428
23.6.3.3	Critère d'information d'Akaike	390	24.3.1.4	Test de Prentice.....	428
23.6.4	Tests de signification des termes d'un modèle	391	24.3.1.5	Test de Peto-Peto et Peto-Peto modifié...428	
			24.3.1.6	Test de Fleming-Harrington.....	429
			24.3.1.7	Test de Savage	429
			24.3.1.8	Test de Renyi.....	429
			24.3.2	Comparaison de plus de deux groupes.....	431

24.3.3	Comparaison de deux groupes avec ajustement à un facteur de confusion : tests stratifiés.....	432	25.2.1.2	Quelques rappels d'algèbre linéaire pour comprendre le détail des opérations.....	475
24.3.4	Test de tendance.....	432	25.2.1.3	Principe de calcul des valeurs et vecteurs propres.....	476
24.4	Modèle de Cox.....	435	25.2.1.4	Représentation des n éléments dans \mathbb{R}^p	478
24.4.1	Présentation du modèle.....	435	25.2.1.5	Signification des valeurs propres.....	480
24.4.2	Estimation des paramètres du modèle	437	25.2.1.6	Représentation des p variables dans \mathbb{R}^n : rôle des variables.....	481
24.4.2.1	Construction de la fonction de vraisemblance.....	437	25.2.2	Quelques aides à la réalisation et à l'interprétation d'une ACP.....	487
24.4.2.2	Maximisation de la fonction de vraisemblance.....	438	25.3	Analyse factorielle des correspondances.....	497
24.4.2.3	Estimation de $h_0(t_i)$	438	25.3.1	La méthode du barycentre.....	498
24.4.3	Termes du modèle.....	438	25.3.2	Profils des lignes et des colonnes.....	500
24.4.3.1	Variable quantitative.....	438	25.3.3	Principe de calcul de l'A.F.C.....	502
24.4.3.2	Variable qualitative ou catégorielle.....	439	25.3.3.1	ACP sur les profils.....	502
24.4.3.3	Interaction entre deux variables catégorielles.....	439	25.3.3.2	Notion de distance et de métrique.....	503
24.4.3.4	Interaction mixte.....	440	25.3.3.3	Centrage du nuage.....	504
24.4.4	Contribution d'un ou de plusieurs termes du modèle.....	440	25.3.3.4	Matrice des variances et covariances.....	505
24.4.4.1	Test du rapport de vraisemblance.....	441	25.3.3.5	Calcul des coordonnées sur les nouveaux axes.....	506
24.4.4.2	Test du score.....	441	25.3.3.6	Notion d'inertie.....	506
24.4.4.3	Test de Wald.....	442	25.3.3.7	Analyse dans \mathbb{R}^p : problème dual.....	508
24.4.5	Construction du modèle.....	442	25.3.4	Propriétés et aides à l'interprétation de l'AFC.....	509
24.4.6	Estimation du risque relatif.....	443	25.4	Analyse factorielle des correspondances multiples (AFCM).....	514
24.5	Modélisation paramétrique.....	448	25.4.1	Le tableau de données.....	515
24.5.1	Loi exponentielle.....	448	25.4.2	Distance.....	517
24.5.2	Loi de Weibull.....	449	25.4.3	Calcul des axes factoriels.....	517
24.5.3	Formes de la fonction de risque selon les lois.....	450	25.4.4	Conduite et interprétation d'une AFCM..	518
24.5.4	Estimation des paramètres.....	451	25.4.4.1	Préparation des données.....	518
24.5.5	Justesse du modèle et outils diagnostiques.....	456	25.4.4.2	Choix de l'analyse.....	520
24.6	Modélisation des risques non proportionnels.....	459	25.4.4.3	Première analyse : AFCM sur les variables extrinsèques.....	520
24.6.1	Evolution des risques par groupe.....	459	25.4.4.4	Deuxième analyse : AFCM sur les variables intrinsèques.....	522
24.6.2	Modèle de Cox morcelé.....	460	25.4.4.5	Troisième analyse : AFCM sur les variables extrinsèques et intrinsèques.....	523
24.7	Variables explicatives dépendantes du temps (VEDT).....	460	25.4.4.6	Quatrième analyse : AFC croisant les variables extrinsèques avec les intrinsèques.....	523
24.8	Modèle des temps accélérés d'échec.....	461	25.4.4.7	Cinquième analyse : ACC ou ACPVI.....	523
24.9	Modèle log-linéaire de survie.....	464	25.4.5	Propriétés des méthodes de groupement et de classification.....
24.10	Calcul de puissance.....	465	25.5	Méthodes de groupement et de classification.....	528
24.10.1	Taille de l'échantillon.....	465	25.5.1	Coefficients de similarité.....	529
24.10.2	Puissance.....	466	25.5.1.1	Problème du double zéro.....	529
24.11	Procédures SAS®.....	467	25.5.1.2	Coefficients binaires de similarité versus coefficients quantitatifs.....	529
RÉFÉRENCES.....	467		25.5.1.3	Coefficients binaires de similarité incluant les doubles zéros.....	530
CHAPITRE 25 Méthodes factorielles, analyses discriminantes, canoniques et méthodes de groupement.....	469		25.5.1.4	Coefficients binaires de similarité excluant les doubles zéros.....	531
25.1	Rudiments d'analyse générale.....	470			
25.2	Analyse en composantes principales.....	472			
25.2.1	Principe de calcul et propriétés des composantes principales.....	472			
25.2.1.1	Principe général de l'ACP.....	473			

25.5.1.5	Coefficients quantitatifs de similarité incluant les doubles zéros	531	25.7.2	Analyse factorielle discriminante pour variables qualitatives	548
25.5.1.6	Coefficients quantitatifs excluant les doubles zéros	532	25.7.3	Classement en groupe ou algorithme de détermination d'un attribut	548
25.5.1.7	Coefficients probabilistes.....	532	25.7.3.1	Cas de deux groupes	549
25.5.2	Distances.....	534	25.7.3.2	Cas de trois groupes et plus	550
25.5.3	Coefficients de corrélation et d'association	535	25.8	Analyse canonique (AC)	552
25.5.4	Méthodes de groupement et de classification	535	25.8.1	Principe général de calcul de l'analyse canonique.....	553
25.5.4.1	Classification hiérarchique.....	536	25.8.2	Propriétés et aide à l'interprétation de l'AC	554
25.5.4.2	Classification non hiérarchique.....	539	25.9	Analyse canonique des correspondances (ACC) ou de redondance (ACR)	556
25.5.5	Propriétés des méthodes de groupement ou de classification	541	RÉFÉRENCES.....		559
25.6	Analyse factorielle des proximités et des similarités	544	INDEX		561
25.7	Analyse discriminante	544	Annexe IV Tableaux de données		
25.7.1	Analyse factorielle discriminante.....	545	* À consulter sur le Web à l'adresse suivante :		
25.7.1.1	Approche géométrique.....	545	www.cheneliere.ca/Scherrer		
25.7.1.2	Principe de calculs des axes factoriels discriminants.....	546			
25.7.1.3	Sélection des axes et des variables.....	547			
25.7.1.4	Propriétés de l'AFD et aide à l'interprétation	548			