

## CHAPITRE 1

ESQUISSE D'UNE ÉPISTÉMOLOGIE DE LA MODÉLISATION STATISTIQUE EN SCIENCES SOCIALES .....	11
<b>1.1 Formalisme et modélisation</b> .....	12
1.1.1 <i>Consensus et dissensions dans les sciences sociales</i> .....	12
1.1.2 <i>Faits virtuels et faits actuels</i> .....	13
1.1.3 <i>Modélisation</i> .....	15
1.1.4 <i>Syntaxe et sémantique</i> .....	16
1.1.5 <i>Formalisme et irréalité</i> .....	18
1.1.6 <i>Modèles purs et modèles à déictiques</i> .....	19
<b>1.2 La modélisation statistique</b> .....	20
1.2.1 <i>Deux fonctions des statistiques</i> .....	20
1.2.2 <i>La notion de modèle statistique : l'importance du terme d'erreur stochastique</i> .....	25
1.2.3 <i>Le point de vue <math>\alpha</math>-modélisateur de J.-P. Benzécri</i> .....	26
1.2.4 <i>Modèle statistique et analyse factorielle : des méthodes antinamiques ?</i> .....	36

## CHAPITRE 2

MODÉLISER DES RELATIONS SIMPLES .....	41
<b>2.1 Différentes échelles de mesure</b> .....	41
<b>2.2 Déterminer la fonction qui lie <math>Y</math> à <math>X</math></b> .....	44
2.2.1 <i>Comment déterminer cette fonction ?</i> .....	47

2.2.2	<i>La fonction statistique</i>	48
2.2.3	<i>Modèle théorique et modèle empirique</i>	49
<b>2.3</b>	<b>La droite de régression par les moindres carrés ordinaires (MCO)</b>	51
2.3.1	<i>Propriétés des MCO et signification des coefficients</i>	52
2.3.2	<i>Hypothèses liées au modèle linéaire MCO</i>	54
2.3.3	<i>Inférence statistique</i>	57
2.3.4	<i>Pouvoir explicatif du modèle : le coefficient de détermination <math>R^2</math></i>	82
2.3.5	<i>Diagnostics sur la régression : analyser les résidus</i>	86
<b>2.4</b>	<b>Un cas appliqué d'analyse de régression simple</b>	87
<b>2.5</b>	<b>Annexe au chapitre 2</b>	95

## CHAPITRE 3

<b>MODÉLISER DES RELATIONS MULTIPLES</b>	97	
<b>3.1 Intérêts de la régression multiple</b>	97	
<b>3.2 Expression formalisée du modèle de régression multiple</b>	99	
<b>3.3 Hypothèses liées à la régression multiple</b>	100	
<b>3.4 Signification des hypothèses et conséquences de leur violation</b>	103	
3.4.1	<i>Signification et conséquences d'une violation de H1 (linéarité et additivité)</i>	103
3.4.2	<i>Signification et conséquences d'une violation de H2 (en particulier, variables explicatives supposées mesurées sans erreur)</i>	104
3.4.3	<i>Signification et conséquences d'une violation de H3 (moyenne nulle et erreur indépendante de X)</i>	105
3.4.4	<i>Signification et conséquences d'une violation de H4 (homoscédasticité)</i>	106
3.4.5	<i>Signification et conséquences d'une violation de H5 (indépendance des erreurs)</i>	108
3.4.6	<i>Signification et conséquences d'une violation de H6 (normalité)</i>	109
3.4.7	<i>Signification et conséquences d'une violation de H7 (pas de colinéarité parfaite)</i>	110
<b>3.5 Pouvoir explicatif du modèle et gain de variance expliquée</b>	110	
<b>3.6 Inférence statistique</b>	113	
3.6.1	<i>Significativité globale du modèle</i>	113
3.6.2	<i>Significativité du gain de variance expliquée</i>	114
3.6.3	<i>Significativité des variables</i>	115
<b>3.7 Une variable explicative est quantitative, les autres sont qualitatives : analyse de la covariance</b>	115	
3.7.1	<i>Cas de variables qualitatives dichotomiques</i>	116
3.7.2	<i>Cas de variables qualitatives polytomiques</i>	127
<b>3.8 Plusieurs variables explicatives sont quantitatives</b>	133	

<b>3.9 Coefficients de régression standardisés</b> .....	135
<b>3.10 Mauvaise spécification de modèle, par un choix erroné dans les variables explicatives</b> .....	137
3.10.1 Une ou plusieurs variables non pertinentes sont incluses dans le modèle .....	138
3.10.2 Une ou plusieurs variables pertinentes sont exclues du modèle .....	139
<b>3.11 Diagnostics et remédiation</b> .....	141
3.11.1 Analyse des résidus .....	141
3.11.2 Les problèmes de multicolinéarité .....	142
<b>3.12 Un cas appliqué d'analyse de régression multiple</b> .....	146
<b>3.13 Conclusion du chapitre: la démarche de spécification des modèles</b> .....	153
3.13.1 Une autre démarche de spécification possible .....	154
3.13.2 Éviter de procéder selon les recettes de manuels .....	154
<b>3.14 Annexe au chapitre 3</b> .....	156

## CHAPITRE 4

### MODÉLISER DES INTERACTIONS ET DES NON-LINÉARITÉS .....

<b>4.1 Modéliser des interactions</b> .....	157
4.1.1 Cas d'une interaction où la variable modératrice est qualitative .....	159
4.1.2 Cas d'une interaction où la variable modératrice est quantitative .....	174
<b>4.2 Modéliser des non-linéarités</b> .....	191
4.2.1 Modéliser des non-linéarités simples .....	192
4.2.2 Modéliser des interactions et des non-linéarités complexes .....	210
<b>4.3 Annexe au chapitre 4</b> .....	218
4.3.1 Effet de la transformation des variables explicatives par l'ajout d'un terme constant dans un modèle sans terme d'interaction .....	218
4.3.2 Programme SAS permettant de spécifier un modèle avec interaction, avec édition de la matrice de variance-covariance des coefficients de régression .....	218

## CHAPITRE 5

### MODÉLISER DES VARIABLES-RÉPONSES QUALITATIVES : LA RÉGRESSION LOGISTIQUE .....

<b>5.1 Le modèle linéaire de probabilité</b> .....	222
<b>5.2 Le modèle linéaire généralisé</b> .....	227
<b>5.3 Le modèle logistique pour variables-réponses dichotomiques ou modèle logit binaire</b> .....	228
5.3.1 La notion d' <b>odds</b> .....	229

5.3.2	Présentation du modèle logistique (ou logit)	231
5.3.3	Inférence statistique	240
5.3.4	Pouvoir explicatif du modèle	245
5.3.5	Coefficients standardisés	248
5.3.6	Analyse des résidus	250
5.3.7	Les problèmes de multicollinéarité	251
5.3.8	Modéliser des interactions et des non-linéarités avec le modèle logit	251
<b>5.4</b>	<b>Le modèle logistique pour variables-réponses polytomiques</b>	<b>255</b>
5.4.1	L'interprétation en termes d'odds	259
5.4.2	L'interprétation en termes de probabilités estimées	260
5.4.3	Modèle logit multinomial ou ensemble de modèle logit binaires ?	262
<b>5.5</b>	<b>Le modèle logistique pour variables-réponses ordinales</b>	<b>263</b>
5.5.1	Mettre en pratique le modèle logit multinomial pour variables ordinales	268
<b>5.6</b>	<b>Annexe au chapitre 5</b>	<b>269</b>
5.6.1	Programmes SAS pour l'estimation des modèles logit	269

## CHAPITRE 6

### MODÉLISER DES DONNÉES HIÉRARCHISÉES : LES MODÈLES MULTINIVEAUX 271

<b>6.1</b>	<b>L'analyse de contexte</b>	<b>273</b>
6.1.1	Quelques distinctions conceptuelles et leur opérationnalisation	273
6.1.2	L'erreur écologique ou biais d'agrégation	274
<b>6.2</b>	<b>Les modèles utilisant les moindres carrés ordinaires: des <i>a priori</i> inadaptés à l'analyse des effets de contexte</b>	<b>276</b>
6.2.1	La non-indépendance des résidus	276
6.2.2	L'hétéroscédasticité	277
6.2.3	Agrégation versus désagrégation	278
6.2.4	L'hétérogénéité des relations	285
6.2.5	Effets fixes et effets aléatoires	286
<b>6.3</b>	<b>Les modèles multiniveaux : une solution adéquate à l'analyse des effets de l'environnement sur les individus</b>	<b>290</b>
6.3.1	Le modèle « vide » équivalant à une ANOVA avec effets aléatoires	291
6.3.2	Le modèle à constantes aléatoires	295
6.3.3	Estimer les effets de groupe aléatoires ( <b>posterior means</b> )	301
6.3.4	Le modèle multiniveau complet : constantes et pentes aléatoires	302
6.3.5	Etude des composants de la variance et modélisation de l'hétéroscédasticité	310

<b>6.4 Le pouvoir explicatif du modèle</b> .....	313
<b>6.5 Inférence statistique</b> .....	316
6.5.1 <i>Significativité de chaque paramètre séparément</i> .....	316
6.5.2 <i>Maximum de vraisemblance « complet » et maximum de vraisemblance restreint</i> .....	317
6.5.3 <i>La significativité testée à l'aide de la déviance</i> .....	318
<b>6.6 Analyser la validité des modèles multiniveaux</b> .....	321
6.6.1 <i>Les hypothèses des modèles multiniveaux</i> .....	321
6.6.2 <i>Examiner les propriétés du modèle</i> .....	322
6.6.3 <i>Conditions d'utilisation du modèle multiniveau</i> .....	325
<b>6.7 Quelques problèmes dans l'estimation des modèles multiniveaux</b> .....	326
6.7.1 <i>Pourquoi certaines estimations sont parfois strictement égales à zéro</i> .....	326
6.7.2 <i>Problèmes de non-convergence</i> .....	326
6.7.3 <i>La question de la centration des variables</i> .....	327
<b>6.8 Un cas appliqué d'analyse multiniveau</b> .....	330
<b>6.9 Dans quels cas les modèles multiniveaux sont-ils vraiment nécessaires ?</b> .....	334
<b>6.10 Annexe au chapitre 6</b> .....	337

## CHAPITRE 7

<b>MODÉLISER DES DONNÉES COMPLEXES : LES EXTENSIONS DES MODÈLES MULTINIVEAUX</b> .....	339
<b>7.1 Modéliser une variance complexe au niveau 1</b> .....	339
<b>7.2 Le modèle à trois niveaux (ou plus)</b> .....	345
<b>7.3 Les modèles aléatoires croisés</b> .....	352
7.3.1 <i>Modéliser des structures non strictement hiérarchisées</i> .....	352
7.3.2 <i>Présentation du modèle aléatoire croisé</i> .....	353
<b>7.4 Les modèles de croissance</b> .....	359
7.4.1 <i>Pertinence d'une structure hiérarchisée comme cadre d'analyse longitudinale</i> .....	359
7.4.2 <i>Présentation du modèle multiniveau de croissance</i> .....	364
7.4.3 <i>Modéliser des fonctions non linéaires de croissance</i> .....	376
7.4.4 <i>L'intérêt des régresseurs qui varient dans le temps</i> .....	380
7.4.5 <i>Tester des structures alternatives de la matrice de variance-covariance des erreurs</i> .....	395
<b>7.5 Modéliser des variables-réponses qualitatives : les modèles logit multiniveaux</b> .....	405
7.5.1 <i>Principes des modèles logit multiniveaux</i> .....	405
7.5.2 <i>Le modèle logit multiniveau avec constantes aléatoires</i> .....	407
7.5.3 <i>Le modèle logit multiniveau complet (avec constantes et pentes aléatoires)</i> .....	417

7.5.4 Conclusion sur les modèles logit multiniveaux.....	419
<b>7.6 Annexe au chapitre 7</b> .....	<b>420</b>
7.6.1 Programmes SAS pour estimer des modèles multiniveaux.....	420
7.6.2 Sites de logiciels pour l'analyse multiniveau (utilisés dans les chapitres 6 et 7).....	422
<b>BIBLIOGRAPHIE</b> .....	<b>425</b>
<b>INDEX DES AUTEURS</b> .....	<b>433</b>
<b>INDEX DES TERMES</b> .....	<b>437</b>
<b>Liste des tableaux</b> .....	<b>445</b>
<b>Liste des figures</b> .....	<b>451</b>
<b>Table des matières</b> .....	<b>459</b>