

Jean-Michel FORT

**L'ANALYSE DES DONNEES
ET SES APPLICATIONS**

**POUR L'AIDE A LA DECISION
DES SERVICES PUBLICS**

Ce livre a été publié sur www.bookelis.com

ISBN : 979-10-227-4784-4

© Jean-Michel Fort, 2017

Tous droits de reproduction, d'adaptation et de traduction, intégrale ou partielle réservés pour tous pays.
L'auteur est seul propriétaire des droits et responsable du contenu de ce livre.

AVANT- PROPOS

Aux lectrices et aux lecteurs,

Ayant été confronté à de nombreuses problématiques liées surtout à l'analyse financière et économique, aux questions juridiques et administratives, à la gestion des ressources humaines, et soucieux de voir l'optimisation de l'organisation des territoires, j'ai tenu à vous faire partager mon intérêt pour « l'analyse de données » qui obéit à des règles statistiques bien précises.

De plus en plus, la production d'information nous oblige à être très attentif et doit nous permettre de faire des choix judicieux pour prendre les bonnes décisions en limitant les risques d'erreurs dans la gestion des dossiers que nous devons préparer à quelque niveau que ce soit et qui concerne les élus, les agents des Collectivités Territoriales et toute personne du secteur public et du secteur privé susceptible de s'intéresser au traitement de données recueillies avec différentes méthodes complémentaires. J'ai souhaité aussi que cet ouvrage témoigne d'un effort pédagogique dans le but de faciliter l'accès à la compréhension des différents aspects plus ou moins techniques et parfois abstraits de l'analyse de données.

L'étude qui va suivre correspond à un travail d'analyse et de synthèse pour aller à l'essentiel, tout en étant très attentif à présenter le fonctionnement de l'analyse de données le plus concrètement possible dans le cadre des Collectivités Territoriales, avec quelques exemples de simulations qui peuvent être transposés dans la réalité de votre profession et dans des secteurs parfois très différents.

Par ailleurs, je ne peux que vous engager à lire et à étudier parallèlement, pour celles et ceux qui le souhaitent, certains documents techniques qui vous permettront de bien comprendre les fondements mathématiques de l'analyse de données surtout lorsque nous sommes appelés à prendre des décisions où la notion d'incertitude est très présente.

Sans avoir la prétention de traiter le sujet d'une manière exhaustive, je pense que ce dossier est susceptible d'intéresser un ensemble de lectrices et de lecteurs attentifs aux questions relatives à « l'aide à la décision » dans de nombreux domaines parmi lesquels on peut citer : les finances, les questions socio-économiques, l'environnement, l'urbanisme, la gestion des ressources humaines, l'informatique, la communication....mais aussi de nombreux faits de société.

Si, pour beaucoup de lectrices et de lecteurs, certains éléments de cette étude semblent d'une évidence tout à fait élémentaire, je pense qu'il est nécessaire de les préciser, de les situer, car dans un contexte plus général de l'analyse de données, ils peuvent nous aider à prendre conscience de la nécessité d'adopter une démarche très rationnelle pour mener à bien les différentes analyses et recherches souvent en dehors de toute les structures importantes organisées, par exemple comme certains laboratoires, think tanks, etc...

En effet, l'analyse de données est une compétence qui permet de traiter les problématiques d'une manière transversale et s'adapte parfaitement à des situations très variées. Le lecteur en général, peut aussi découvrir comment l'analyse de données réussit très précisément à nous

montrer l'existence d'informations plus ou moins cachées qu'il est souvent difficile, voire impossible d'appréhender uniquement par une approche intuitive.

Par ailleurs, j'ai associé à cet ouvrage, un site internet appelé «listing», actualisé régulièrement, qui contient tous les liens mentionnés, des adresses de sites contenant de nombreuses informations avec la possibilité de télécharger des fichiers, des cours, des programmes, afin de donner à cette étude, « un aspect dynamique et fonctionnel ».

Je souhaite également que chacune et chacun d'entre vous puisse y trouver des informations intéressantes susceptibles de donner des orientations pour obtenir des réponses à des questions posées en utilisant des méthodes que nous n'avons pas toujours étudiées dans le cadre de nos différents cursus de formation.

J'adresse aussi mes plus sincères remerciements d'une part, à toutes les personnes qui ont décidé de lire ce livre, et d'autre part, aux auteurs d'ouvrages, de logiciels, les universitaires, les éditeurs de sites internet qui, par leurs cours et leurs manuels techniques édités en version papier et en version électronique, ont facilité la réalisation de ce travail dont le but principal est de faire connaître au lecteur, en toute modestie, « un aspect parmi d'autres de l'analyse de données ».

Je tiens aussi à remercier mes proches qui m'ont aussi permis de dégager le temps nécessaire à la préparation puis à la réalisation de l'ensemble du projet.

Toute erreur qui se serait glissée dans le présent ouvrage serait bien évidemment involontaire malgré les relectures programmées avant son impression qui sera susceptible de mises à jour ultérieures éventuelles.

Le 29 avril 2017,

L'Auteur,
Jean-Michel Fort,

SOMMAIRE

AVANT- PROPOS	5
SOMMAIRE	7
<hr/>	
PREMIERE PARTIE	15
L'ANALYSE DE DONNEES : LA DEFINITION ET LA MISE EN ŒUVRE	15
CHAPITRE 1 - LA PRESENTATION DE L'ANALYSE DE DONNEES	17
1.1 Une définition à partir d'un exemple	17
1.2 L'analyse de données pour les Services des Collectivités Territoriales	17
1.3 L'analyse de données et la méthodologie	18
1.3.1 Etape 1 : le recueil des données et la mise en œuvre des analyses	18
1.3.1.1 Avec l'utilisation de deux versions de questionnaires modulables	18
1.3.1.1.1 « En version papier »	18
1.3.1.1.2 « En version dématérialisée »	18
1.3.1.2 Avec l'utilisation de sources de données	19
1.3.2 Etape 2 : le traitement des données recueillies	22
1.3.3 Etape 3 : l'analyse des données	22
1.3.3.1 Le 1 ^{er} niveau avec la méthode du tri à plat	22
1.3.3.2 Le 2 ^{ème} niveau, avec la méthode de l'analyse bivariée	22
1.3.3.2.1 Les tris croisés	22
1.3.3.2.2 L'étude de la régression simple ou multiple	22
1.3.3.3 Le 3 ^{ème} niveau avec l'analyse multivariée composée principalement de deux méthodes	22
1.3.3.3.1 La méthode de l'analyse factorielle des correspondances	22
1.3.3.3.2 La méthode de la régression multiple	22
1.4 Les conclusions	23
1.4.1 Comment susciter, maintenir l'appétence des décideurs pour l'analyse de données ?	23
1.4.2 Comment faire un test rapide en deux étapes avec l'analyse de données ?	23
1.4.2.1 Etape 1 : le test du χ^2 dit « d'indépendance »	24
1.4.2.2 Etape 2 : l'analyse graphique	25
CHAPITRE 2 - LE TRAITEMENT DES DOSSIERS DE L'ANALYSE DE DONNEES	27
2.1 La Collectivité traite en interne ses dossiers d'analyses de données	27
2.1.1 Exemple 1	27
2.1.2 Exemple 2	27
2.2 La Collectivité confie ses dossiers d'analyses de données à un prestataire	27
CHAPITRE 3 - UNE SOLUTION POUR REALISER UN PROJET D'ANALYSE DE DONNEES	29
3.1 Les objectifs	29
3.2 Deux cas principaux peuvent se présenter	29
3.2.1 Exemple 1 : la Collectivité dispose d'un panel simple	29
3.2.2 Exemple 2 : la Collectivité n'utilise pas de panel	30

CHAPITRE 4 - LE DEROULEMENT D'UNE ENQUETE AVEC LE PARTENARIAT PUBLIC / PRIVE	31
4.1 Un schéma technique en huit étapes.....	31
4.1.1 Présentation d'une possibilité d'un fonctionnement partenarial	31
4.2 Le Fonctionnement des huit étapes.....	32
4.2.1 Cas où la Collectivité a retenu un Prestataire P .	32
4.2.1.1 Etape 1 : la phase de préparation avec 2 options de fonctionnement.....	32
4.2.1.1.1 Option 1 : espace assistance personnalisé .	32
4.2.1.1.2 Option 2 : autre possibilité	32
4.2.1.2 Etape 2 : la mise au point du questionnaire et le lancement de l'enquête	32
4.2.1.3 Etape 3 : les réponses à l'enquête	33
4.2.1.4 Etape 4 : l'enregistrement des réponses et la préparation de leur traitement.....	33
4.2.1.5 Etape 5 : le 1 ^{er} traitement des données recueillies	33
4.2.1.6 Etape 6 : le 2 ^{ème} traitement des données recueillies	33
4.2.1.7 Etape 7 : la rédaction du rapport d'analyse des données	33
4.2.1.8 Etape 8 : la mise à disposition du rapport d'analyse des données	33
4.2.2 Cas où la Collectivité décide de procéder par elle-même.....	34
4.2.3 Quelques informations complémentaires	34
<hr/>	
DEUXIEME PARTIE	35
L'ANALYSE DE DONNEES : CAS PRATIQUES, SIMULATIONS.....	35
 CHAPITRE 5 - SIMULATION 1 POUR LE SECTEUR DES FINANCES PUBLIQUES	 37
5.1 L'objectif : l'analyse par fonction, des dépenses des administrations publiques	37
5.2 La méthode utilisée : l'analyse en composante principale	37
5.3 La mise en œuvre de la collecte des informations.....	37
5.4 L'analyse des données et les résultats graphiques.....	38
5.5 Explications et commentaires synthétiques	45
5.6 Eléments de conclusion	47
5.7 Réflexion générale	47
 CHAPITRE 6 - SIMULATION 2 POUR LE SECTEUR DU CONTROLE DE GESTION	 49
6.1 L'objectif : la mesure d'un niveau de satisfaction.....	49
6.2 La méthode utilisée : une application particulière de la régression multiple	49
6.3 La mise en œuvre de la collecte des informations.....	49
6.4 L'analyse des données et les résultats graphiques (la matrice importance / performance) ..	50
6.5 Explications et commentaires	51
6.6 Eléments de conclusion	52
6.7 Réflexion générale	52
 CHAPITRE 7 - SIMULATION 3 POUR LE SECTEUR SOCIO-ECONOMIQUE	 53
7.1 L'objectif : vérifier l'existence ou non d'un lien entre la catégorie.....	53
socioprofessionnelle et le secteur d'activité économique d'une population.....	53

7.2 La méthode utilisée : l'analyse factorielle des correspondances	53
7.3 La mise en œuvre de la collecte des informations	53
7.4 L'analyse des données et le calcul du Khi^2	54
7.5 Explications et commentaires des informations fournies par le tableau et le graphe	55
7.6 Eléments de conclusion	55
7.7 Réflexion générale	56
CHAPITRE 8 - SIMULATION 4 POUR LE SECTEUR DE L'ENVIRONNEMENT	57
8.1 L'objectif : réaliser l'étude du comportement des catégories socio - professionnelles	57
8.2 La méthode utilisée : les correspondances multiples	57
8.3 La mise en œuvre de la collecte des informations	57
8.4 L'analyse des données et le résultat graphique	57
8.5 Explications et commentaires	58
8.6 Eléments de conclusion	59
8.7 Réflexion générale	60
CHAPITRE 9 - SIMULATION 5 POUR LE SECTEUR DE LA FORMATION	61
9.1 L'objectif : la recherche de « l'interdépendance entre deux questions »	61
9.2 La méthode utilisée : le test du khi^2 d'indépendance - cas général	61
9.3 La mise en œuvre de la collecte des informations	62
9.4 L'analyse des données et les résultats	62
9.5 Explications et commentaires des informations fournies par les tableaux	66
9.6 Eléments de conclusion	67
9.7 Réflexion générale	67
CHAPITRE 10 - SIMULATION 6 POUR LE SECTEUR DES ELECTIONS	69
10.1 L'objectif : on cherche à savoir si les votes sont indépendants du sexe des électeurs.	69
10.2 La méthode utilisée : le Test exact de Fisher	69
10.3 La mise en œuvre de la collecte des informations	69
10.4 L'analyse des données et le résultat	69
10.5 Explications et commentaires	71
10.6 Eléments de conclusion	71
10.7 Réflexion générale	71
CHAPITRE 11 - SIMULATION 7 POUR LE SECTEUR CONTENTIEUX	73
11.1 L'objectif : Il s'agit de faire un essai de prévision	73
11.2 Méthode utilisée : la régression linéaire multiple	73
11.3 La mise en œuvre de la collecte des informations	74
11.3.1 Cas 1 : avec des variables significatives	74
11.3.2 Cas 2 : avec des variables non significatives (voir annexe 4 page 132)	75

11.4 L'analyse des données et les résultats graphiques	76
11.4.1 Le cas 1 avec des variables significatives	76
11.4.1.1 Etape 1 : l'analyse des coefficients dans le tableau T2	77
11.4.1.1.1 Informations – étape 1	78
11.4.1.2 Etape 2 : le tableau T5	79
11.4.1.2.1 Informations – étape 2	79
11.4.1.3 Etape 3 : le test de Student	80
11.4.1.3.1 Informations – étape 3	82
11.4.1.4 Etape 4 : l'analyse des résidus	83
11.4.1.4.1 Informations – étape 4	88
11.4.1.5 Etape 5 : des examens complémentaires	89
11.4.1.5.1 Informations – étape 5	91
11.4.2 Le cas 2 : avec des variables non significatives (voir annexe 4 page 132)	94
11.5 Une méthode pratique d'analyse	94
11.5.1 Phase 1 : éléments de préparation de l'analyse des données	94
11.5.2 Phase 2 : détermination d'un niveau de fiabilité - tableaux de résultats	95
11.6 Eléments de conclusion concernant le modèle de la régression linéaire multiple	97
11.7 Réflexion générale	97
 CHAPITRE 12 - SIMULATION 8 POUR LE SECTEUR DES SCIENCES HUMAINES	 99
12.1 L'objectif : une étude des scores de stress en fonction des postes de travail	99
12.2 La méthode utilisée : la variance dite à « un facteur »	99
12.3 La mise en œuvre de la collecte des informations	99
12.4 L'analyse des données et le résultat graphique	99
12.5 Eléments de conclusion	101
12.6 Réflexion générale	102
 CHAPITRE 13 - SIMULATION 9 POUR LE SECTEUR FAITS DE SOCIETE	 103
13.1 L'objectif : essai de détermination d'un niveau de délinquance avec deux délits	103
13.2 La méthode utilisée : l'analyse de la variance dite à « deux facteurs »	103
13.3 La mise en œuvre de la collecte des informations	103
13.4 L'analyse des données et les résultats	103
13.5 Explications et commentaires des informations fournies par les tableaux	106
13.6 Réflexion générale	106
 TROISIEME PARTIE	 107
L'ANALYSE DE DONNEES ET L'INFORMATIQUE	107
 CHAPITRE 14 - L'INFORMATIQUE AU SERVICE DE L'ANALYSE DE DONNEES	 109
14.1 Les logiciels à utiliser	109
14.2 Dans quels cas utiliser Excel en interne pour un traitement optimum des données	109

14.2.1 L'application Colmip v1.0 programmée avec Excel, pour quel objectif ?	109
14.2.2 Rappel schématique de son mode de fonctionnement	110
14.2.3 Les paramétrages de Colmip v1.0 en fonction de ses utilisations	110
14.2.3.1 Une utilisation en interne.....	110
14.2.3.2 Une utilisation à caractère externe	111
14.2.4 Les caractéristiques principales de Colmip v1.0.....	112
14.2.4.1 Les points clés de l'application Colmip v1.0.....	112
14.2.4.2 La spécificité de Colmip v1.0 avec les outils de communication.....	112
14.2.4.3 La pratique des enquêtes en interne.....	112
14.2.5 Colmip v1.0 et ses utilisations spécifiques	113
CHAPITRE 15 - DES OUTILS DE CALCUL AU SERVICE DE L'ANALYSE DE DONNEES.....	115
15.1 Un lien pour télécharger des fichiers, des applications	115
15.2 Quelques orientations de réflexions	115
<hr/>	
QUATRIEME PARTIE.....	117
LES ANNEXES	117
➤ ANNEXE 1A.....	118
POUR LA SIMULATION 3 - SECTEUR SOCIO-ECONOMIQUE ET POUR LA SIMULATION 5	
SECTEUR FORMATION - <u>LE TEST DU KHI²</u>.....	118
A - OBJECTIF : CE TEST PERMET LA COMPARAISON DE DEUX TABLEAUX D'EFFECTIFS (REELS ET THEORIQUES).	118
B - PRINCIPE – MODE DE FONCTIONNEMENT	118
1. Exemple concret	118
2. La table du Khi ²	119
3. Les calculateurs en ligne : test du khi ²	120
➤ ANNEXE 1B.....	122
POUR LA SIMULATION 5 - SECTEUR DE LA FORMATION - <u>PROGRAMME DE CALCUL DU KHI²</u>.....	122
➤ ANNEXE 2.....	124
POUR LA SIMULATION 6 - SECTEUR DES ELECTIONS - <u>LE TEST EXACT DE FISHER</u>.....	124
ETAPE 1 : APRES AVOIR OUVERT L'APPLICATION EXCEL	124
ETAPE 2 : PREPARATION DU TEST.....	125
ETAPE 3 : LANCEMENT DE L'ANALYSE.....	125
<u>LE TEST EXACT DE FISHER</u> : AUTRE EXEMPLE DE CALCUL ET INTERPRETATION.....	128
A - OBJECTIF : CE TEST EST UTILISE POUR ANALYSER UN TABLEAU DE CONTINGENCE 2 x 2	128
B - PRINCIPE - MODE DE FONCTIONNEMENT	128
1. Un exemple de résultats obtenus avec l'application Xlstat	128
2. La prise de décision	129
3. Les calculateurs en ligne pour le test exact de Fisher	129

➤ ANNEXE 3	131
POUR LA SIMULATION 7 - SECTEUR CONTENTIEUX - RESSOURCES HUMAINES	
<u>LA TABLE DE FISHER</u>	131
 ➤ ANNEXE 4	132
POUR LA SIMULATION 7- SECTEUR CONTENTIEUX - RESSOURCES HUMAINES	
<u>LA REGRESSION LINEAIRE ET LE CAS 2, AVEC DES VARIABLES NON SIGNIFICATIVES</u>	132
ETAPE 1 : L'ANALYSE DES COEFFICIENTS	133
> Informations - étape 1	133
ETAPE 2 : LE F DE FISHER	133
> Informations - étape 2	134
ETAPE 3 : LE TEST DE STUDENT	135
> Informations - étape 3	135
ETAPE 4 : SYNTHÈSE DES GRAPHES ET DOCUMENTS DE CALCUL	135
> Informations – étape 4	138
ETAPE 5 : DES EXAMENS COMPLÉMENTAIRES	138
> Informations – étape 5	140
ETAPE 6 : PRÉPARATION DU TABLEAU SYNTHÉTIQUE	140
 ➤ ANNEXE 5	143
POUR LA SIMULATION 3 - SECTEUR SOCIO-ECONOMIQUE - <u>LE DEGRÉ DE PRÉCISION DE LA REPRÉSENTATION GRAPHIQUE</u>	
	143
 ➤ ANNEXE 6	145
POUR LA SIMULATION 4 - SECTEUR DE L'ENVIRONNEMENT - <u>TABLEAU DES RÉPONSES AU QUESTIONNAIRE</u>	
	145
 ➤ ANNEXE 7	146
POUR LA SIMULATION 7 - SECTEUR CONTENTIEUX - RESSOURCES HUMAINES	
<u>LA REGRESSION LINEAIRE</u>	146
 ➤ ANNEXE 8 A	147
POUR LA SIMULATION 8 - SECTEUR DES SCIENCES HUMAINES - <u>LA VARIANCE A UN FACTEUR</u>	
	147
ETAPE 1	147
ETAPE 2	148
ETAPE 3	148
ETAPE 4	148
ETAPE 5	150
 ➤ ANNEXE 8 B	151
POUR LA SIMULATION 9 - SECTEUR FAITS DE SOCIÉTÉ - <u>LA VARIANCE A DEUX FACTEURS</u>	
	151

➤ ANNEXE 9	152
POUR LA SIMULATION 7 - SECTEUR CONTENTIEUX - RESSOURCES HUMAINES - LA REGRESSION LINEAIRE - <u>LA LOI NORMALE POUR LE MODE DE CALCUL DES RESIDUS DITS NORMALISES</u>	152
➤ ANNEXE 10	155
POUR LA SIMULATION 9 - SECTEUR FAITS DE SOCIETE - <u>CONFIRMATION DE RESULTATS AVEC SAS 9.0</u>	155
➤ ANNEXE 11	156
PRINCIPAUX SITES INTERNET SUSCEPTIBLES DE FOURNIR DES DONNEES UTILES POUR PREPARER DES ANALYSES DOCUMENTEES (BASES DE DONNEES, ETC...).....	156
➤ ANNEXE 12	158
FICHE TECHNIQUE DU FONCTIONNEMENT DE L'APPLICATION COLMIP V1.0	158
PHASE ECRAN 1 :	158
PHASE ECRAN 2 :	159
PHASE ECRAN 3 :	160
➤ ANNEXE 13	161
FICHE DU SITE INTERNET LISTING (INDEX DE CLASSEMENT ET CODE D'ACCES)	161
<hr/>	
CINQUIEME PARTIE - LES APPLICATIONS INFORMATIQUES UTILISEES	163
<hr/>	
SIXIEME PARTIE - LES SITES INTERNET VISITES	163
<hr/>	
SEPTIEME PARTIE - LA BIBLIOGRAPHIE	163
<hr/>	
HUITIEME PARTIE - LE LEXIQUE	164
<hr/>	

PREMIERE PARTIE

L'analyse de données : la définition et la mise en œuvre

Cette étude tient compte de l'ensemble de la réglementation appliquée en France ; cependant, il n'est pas impossible d'en transposer certains aspects dans des contextes différents de pays étrangers, notamment en ce qui concerne les schémas de fonctionnement technique.

Chapitre 1 - La présentation de l'analyse de données

1.1 Une définition à partir d'un exemple

Supposons que nous ayons collecté les données a, b, c, d, e, f, g, h, et que l'on cherche à répondre à la question suivante : quelles relations peuvent bien exister entre ces différentes données qui pourraient nous fournir des informations intéressantes susceptibles de nous aider à prendre une décision dans tel ou tel projet ?

Cela revient donc à essayer de trouver des informations plus ou moins cachées et souvent difficiles à déceler en raison de leur multiplicité et de leur diversité.

Pour résumer, c'est avec « un ensemble de méthodes essentiellement géométriques » que des réponses le plus fréquemment graphiques peuvent nous être apportées grâce à l'utilisation des mathématiques, de la statistique et des programmes informatiques.

A partir de plusieurs exemples, nous allons donc essayer de voir comment les Collectivités Territoriales et d'une manière plus générale, les décideurs, les observateurs, les chercheurs de solutions, peuvent bénéficier facilement et concrètement de la mise en œuvre de l'analyse de données.

1.2 L'analyse de données pour les Services des Collectivités Territoriales

Sans avoir la prétention d'être exhaustif, on peut étudier plusieurs thématiques qui concernent particulièrement l'activité des services de l'ensemble des Collectivités Territoriales, à savoir :

En matière **d'aménagement du territoire, d'environnement** : par exemple avec le stationnement en zone urbaine, la circulation, la qualité de l'air, les nuisances sonores ;

En matière de **ressources humaines dans les services publics** : avec l'analyse de plusieurs baromètres comme la formation, le climat social, la gestion du stress ;

Dans le secteur **des finances, de l'économie, des faits de société** : avec l'appréciation des critères de mise en place d'un contrôle de gestion, l'analyse comparative des indicateurs concernant les différentes catégories de Collectivités Territoriales, l'étude des comportements des populations ;

Dans le cadre de **la e-administration**, par l'évaluation de « la qualité de la communication via les sites internet / intranet » à partir d'enquêtes réalisées auprès de plusieurs publics (les administrés, les élus, les services, les clients, les fournisseurs, les touristes, etc...).

Ces thématiques concernent toujours l'ensemble des acteurs de toutes les sociétés. Elles sont intemporelles et pour les événements du passé, du présent mais aussi du futur proche et lointain, l'analyse de données peut s'appliquer avec une certaine pertinence en tenant compte évidemment des évolutions techniques, comportementales, organisationnelles, qui peuvent se produire au fil du temps.