

Comprendre et utiliser les statistiques dans les sciences de la vie

- Médecine
- Biologie
- Recherche clinique
- Épidémiologie...

B. FALISSARD

■ ■ 3^e édition

■ ■ MASSON

Table des matières

INTRODUCTION

1. REMARQUES CONCERNANT LA MÉTHODOLOGIE

EN RECHERCHE MÉDICALE	3
Méthode et méthodologie	3
Démonstration et connaissance médicale	4
<i>Le statut de connaissance médicale</i>	4
<i>Démonstration et niveau de preuve</i>	7
Preuve et croyance	7
La question de la reproductibilité en recherche médicale	8
Le caractère aléatoire des résultats expérimentaux	9
« Choisir ce qu'il faut observer »	10
<i>Quelles variables mesurer et chez qui les mesurer ?</i>	10
Mesures catégorielles	10
Mesures dimensionnelles	11
Mesures objectives, mesures subjectives	12
<i>Comment les mesurer ?</i>	13
<i>Mesurer dans quelles circonstances ?</i>	14
Observation, expérimentation	14
Le tirage au sort	15
Rien n'est acquis	16
Conclusion	17

2. DÉFINITIONS

3. GESTION DES DONNÉES

Avant l'étude	21
<i>Les aspects réglementaires</i>	21
<i>Le cahier de recueil des données</i>	21
<i>Mise en route</i>	22
Pendant l'étude	22
Après l'étude	22
<i>Le codage</i>	22
<i>La saisie</i>	23
<i>L'analyse</i>	24
La manipulation des fichiers	24
Le contrôle de consistance	24

MÉTHODES UNIVARIÉES

1. LES REPRÉSENTATIONS GRAPHIQUES

Une variable quantitative	27
Une variable qualitative	30
Deux variables quantitatives	31
Deux variables qualitatives	32
Variable quantitative et variable qualitative	33
Représentation de données temporelles	35

2. L'ESTIMATION

Les mesures de position	37
Les mesures de dispersion	38
Les mesures de liaison	41
<i>Variables aléatoires quantitatives</i>	41
Mesures de dépendance	42

Mesures de dépendance monotone	42
Mesures de concordance.....	44
Variables aléatoires qualitatives	45
Mesures de dépendance	45
Mesures de dépendance monotone	49
Mesures de concordance.....	50
Et les intervalles de confiance ?	55
Intervalle de confiance d'un pourcentage	55
Intervalle de confiance d'une moyenne	56
Intervalle de confiance d'un coefficient de corrélation.....	56
Intervalle de confiance d'un odds-ratio	57
3. LES TESTS D'HYPOTHÈSE	59
Les bases.....	59
Les risques α et β	60
Le choix d'une hypothèse, le calcul de « p »	60
A propos de β	62
Tests d'hypothèse ou de significativité ?	64
Pourquoi 5 % ?	64
Tests uni ou bilatéraux ?	65
Il est permis de critiquer.....	65
L'approche bayésienne.....	66
Applications pratiques	68
Quels tests pour quelles hypothèses ?	68
Comparaison de deux moyennes.....	68
Comparaison d'une moyenne à une moyenne de référence	70
Comparaison de deux pourcentages.....	71
Comparaison d'un pourcentage à un pourcentage de référence.....	73
Test de la nullité d'un coefficient de corrélation	74
Tests paramétriques ou non ?	74
Comparaison de deux moyennes.....	75
Comparaison de deux pourcentages.....	75
Test de nullité d'un coefficient de corrélation	76
Tester la normalité ?	76
Tests appariés, non appariés	78
Le calcul de β et du nombre de sujets à inclure	81
Comparaison de deux pourcentages.....	81
Comparaison de deux moyennes.....	82
Nullité d'un coefficient de corrélation	83
Cas de groupes d'effectifs déséquilibrés.....	84
Quelques écueils à éviter	85
Les analyses intermédiaires.....	85
De l'usage abusif du « p »	86
4. DONNÉES DE SURVIE	87
Définitions	88
Représentations graphiques	88
L'estimation.....	93
Paramètres de position et de dispersion.....	93
Paramètre de liaison	94
Tests d'hypothèses.....	95
Comparaison de deux populations	95
Comparaison de plus de deux populations.....	97
Test d'une association entre la survie et une variable quantitative	97

MODÈLES MULTIVARIÉS

I RÉGRESSION LINÉAIRE ET ANALYSE DE VARIANCE	101
Introduction	101
Régression, prédiction et ajustement.....	101
Régression linéaire et analyse de variance	104
Une différence de traditions.....	104
L'interaction de deux facteurs	106
L'essentiel.....	106
La régression linéaire.....	106
L'analyse de variance (ANOVA)	111
Une ANOVA à un facteur.....	111
La méthode des blocs.....	115
Un plan factoriel d'ordre 2	117
Les carrés latins	120
Le modèle mixte	122
L'analyse de covariance.....	126
A propos de la régression linéaire.....	130
Corrélation et régression linéaire	130
Ajuster une corrélation : le coefficient de corrélation partielle	131
Le test t : un cas particulier de régression	132
Le codage de variables qualitatives	134
Le prix à payer.....	136
La vérification des hypothèses sous-jacentes	136
Vérification de la normalité des résidus e_i	137
Homoscédasticité	138
Indépendance des résidus	139
Le modèle est-il robuste ?	140
Les variables explicatives ne sont-elles pas redondantes ?	143
II MODÈLE LINÉAIRE GÉNÉRALISÉ :	
RÉGRESSIONS LOGISTIQUE ET DE POISSON	145
La régression logistique	145
Introduction	145
L'essentiel.....	150
Premiers contacts, premiers problèmes.....	150
Quand « ajuster » est toujours le maître mot.....	152
A propos de la régression logistique	156
Odds-ratio et régression logistique	156
Un odds-ratio ajusté.....	157
Le test du chi-2 : un cas particulier de régression logistique	158
Le prix à payer.....	159
Le modèle est-il robuste ?	159
Les variables explicatives ne sont-elles pas redondantes ?	166
Quand le modèle ne converge pas	166
La régression logistique multinomiale.....	167
La régression logistique conditionnelle, le test de Mantel-Haentzel.....	172
La régression de Poisson	175
LE MODÈLE DE COX	179
Le test du log-rank comme cas particulier du modèle de Cox	180
Le modèle de Cox : un test du log-rank avec possibilité d'ajustement.	181
Modèle de Cox avec stratification	182
Modèle de Cox avec covariables dépendantes du temps	183

Modèle de Cox : le prix à payer.....	184
<i>Hypothèse des risques proportionnels.....</i>	185
<i>Hypothèse de loglinéarité.....</i>	186
<i>Recherche de sujets marginaux.....</i>	187
4. MESURES RÉPÉTÉES.....	189
La variable à expliquer est quantitative.....	189
<i>Les données sont corrélées.....</i>	189
<i>Un facteur « sujet » à effet aléatoire ?.....</i>	193
<i>Une MANOVA ?.....</i>	195
<i>Une ANOVA sur mesures répétées ?.....</i>	197
<i>Des résidus à covariance structurée ?.....</i>	198
La variable à expliquer est binaire.....	200
La variable à expliquer est censurée.....	201
5. MODÈLES STRUCTURAUX.....	203
Introduction.....	203
Le modèle de régression linéaire : un cas particulier de modèle structural.....	207
Exemple d'application.....	209
6. MODÈLES MULTIVARIÉS : DISCUSSION.....	217
Comment construire un modèle ?.....	217
<i>Le choix de la variable à expliquer.....</i>	218
<i>Le choix des variables explicatives.....</i>	218
Pourquoi ne pas utiliser un maximum de variables explicatives ?.....	219
La modélisation « pas à pas ».....	219
En pratique quelle démarche adopter ?.....	220
<i>Codage des variables explicatives.....</i>	221
Codage des variables explicatives qualitatives.....	221
Codage des variables explicatives quantitatives, ajustement sur <i>spline</i>	223
<i>Les interactions.....</i>	226
Comment interpréter l'effet d'une variable explicative en présence d'un terme d'interaction qui la concerne ?.....	226
Quelles interactions introduire dans un modèle ?.....	228
<i>Stratification, ajustement, score de propension.....</i>	229
Stratifier ou ajuster ?.....	229
L'analyse des données stratifiées.....	230
Score de propension.....	233
<i>Gestion des données manquantes.....</i>	233
<i>Nombre de sujets à inclure.....</i>	238
Interpréter les résultats d'une modélisation.....	239
<i>L'inférence statistique.....</i>	239
Tests ou sommes de carrés de type 1, 2 et 3.....	239
Test de Wald, du score et test du rapport de vraisemblance.....	240
Méthodes de rééchantillonnage, le <i>bootstrap</i>	242
Tests et comparaisons multiples.....	246
<i>Evaluer la qualité d'un modèle.....</i>	249
Le diagnostic de régression.....	249
L'adéquation d'un modèle.....	250

MÉTHODES EXPLORATOIRES MULTIDIMENSIONNELLES

1. ANALYSE DISCRIMINANTE	253
A la croisée des chemins : analyse discriminante, régression linéaire et régression logistique	253
Un exemple avec plus de deux catégories à discriminer	262
2. MODÈLE LOG-LINÉAIRE	265
Introduction	265
<i>Indépendance marginale, indépendance partielle</i>	265
<i>L'ordre d'une interaction</i>	267
<i>Le modèle log-linéaire</i>	267
Un exemple	268
3. MÉTHODES DE CLASSIFICATION, ANALYSE EN CLUSTERS	275
Classifications hiérarchiques ascendantes	275
La méthode des nuées dynamiques	287
4. MÉTHODES DE SEGMENTATION, CART	297
En quelques mots	297
En pratique	299
5. ANALYSE EN COMPOSANTES PRINCIPALES	303
Une réduction de dimensions	303
<i>Des composantes principales</i>	304
<i>Une représentation graphique optimisée</i>	310
<i>Points sujets ou points variables supplémentaires</i>	312
<i>De l'intérêt de normaliser les données</i>	313
Les composantes principales en pratique	315
<i>Expliquer la variabilité</i>	315
<i>Régression sur composantes principales</i>	317
Les représentations graphiques en pratique	321
ACP et ajustement	326
Une ACP avec variable à expliquer et variables explicatives	328
6. ANALYSE DES CORRESPONDANCES	331
7. ÉCHELONNEMENT MULTIDIMENSIONNEL	337
Un exemple de déploiement non linéaire	338
Une représentation de similitudes	340
Une représentation de préférences	343
8. ANALYSE DES CORRÉLATIONS CANONIQUES	347
9. ANALYSE FACTORIELLE	353
L'analyse factorielle : un modèle linéaire avec variables latentes	353
Analyse factorielle, analyse en composantes principales et classification de variables	360
BIBLIOGRAPHIE	367
INDEX	369

Comprendre et utiliser les statistiques dans les sciences de la vie

B. FALISSARD

L'ouvrage

- Les sciences de la vie doivent recourir à des méthodes objectives pour valider des résultats expérimentaux et les statistiques restent la principale méthode pour y parvenir. Laissant de côté l'aspect purement mathématique du sujet, l'auteur se concentre sur l'utilisation des statistiques (en s'appuyant sur des outils informatiques simples) avec deux objectifs : former le lecteur aux méthodes statistiques et les lui faire comprendre. Le recours à des exemples réels ainsi qu' à de nombreux schémas contribue à rendre cet ouvrage pratique et accessible à tous.
- Cet aspect didactique est renforcé par une partie méthodologique fruit des réflexions pédagogiques de l'auteur et particulièrement adaptée aux étudiants avec des démonstrations et exemples réalisés à l'aide de logiciels courants, dont un est téléchargeable sur Internet et accessible à tous.
- Le lecteur trouvera également des cours multimédias qui résument les points principaux de l'ouvrage et les fichiers sources ayant servi à la réalisation des exemples sur www.masson.fr (cliquer dans la rubrique « Télécharger » de la fiche du livre).

Le public

- Chercheurs et étudiants dans les disciplines des sciences de la vie (biologie, médecine, etc.), médecins et autres acteurs des sciences de la vie.

L'auteur

Bruno Falissard est professeur des universités et praticien hospitalier (P.U.-P.H) en biostatistique à la faculté de médecine Paris-sud - Hôpital Paul-Brousse. Il est aussi responsable d'une unité INSERM.

Retrouvez
tous les ouvrages Masson sur
www.masson.fr

ISBN 978-2-294-01850-3



9 782294 018503