

Alexandre Lourme

# Statistique pour la Licence



**L**avoisier  
hermes

Alexandre Lourme

# Statistique pour la Licence

**L**avoisier  
hermes

[editions.lavoisier.fr](http://editions.lavoisier.fr)

© 2018, Lavoisier, Paris  
ISBN : 978-2-7462-4869-4

# Avant-propos

Ce manuel de statistique est un recueil de méthodes classiques illustrées sous le logiciel R. Les thèmes abordés sont enseignés en premier cycle universitaire et font partie d'un cours standard de statistique appliquée.

Ce livre est divisé en deux parties de trois chapitres chacune. La première partie est consacrée à la statistique univariée : décrire une série, estimer un paramètre, tester une hypothèse. La seconde partie est dédiée à la statistique bivariée : couples de variables qualitatives, mixtes, quantitatives.

Le manuel s'adresse aux étudiants de Licence 1/2/3 de sciences, de sciences expérimentales, de sciences humaines. Les étudiants de PACES, de CPGE, d'IUT et de BTS trouveront également dans cet ouvrage des éléments de leur programme de statistique. Le tableau ci-dessous met en évidence les filières universitaires dans lesquelles sont traités les chapitres du livre.

<i>cursus</i>		<i>chapitre</i>					
		1	2	3	4	5	6
IUT	chimie	✓	✓	✓	✓	✓	✓
	génie biologique	✓	✓	✓	✓	✓	✓
	génie chimique	✓	-	✓	-	-	-
	génie civil	✓	✓	-	-	-	✓
	génie électrique	-	-	✓	-	-	✓
	génie industriel	✓	✓	-	-	-	✓
	génie mécanique	-	✓	✓	-	-	-
	hygiène et sécurité	-	✓	✓	-	-	-
	qualité logistique	✓	✓	✓	✓	✓	✓
	informatique	-	✓	✓	-	-	✓
	statistique & info. décisionnelle	✓	✓	✓	✓	✓	✓
-----							
	économie & gestion	✓	✓	✓	✓	✓	✓
Licences	sciences et technologies	✓	✓	✓	✓	✓	✓
	sciences humaines et sociales	✓	✓	✓	✓	✓	✓
-----							
CPGE	ECE	✓	✓	-	-	-	✓
	ECS	✓	✓	-	-	-	✓
	ECT	✓	✓	-	-	-	-
	ENS D2	-	✓	-	✓	✓	✓
	BCPST	✓	✓	-	-	-	✓
-----							
concours	PACES	-	✓	✓	✓	✓	-

Table 1. Quels chapitres sont abordés dans votre cursus ?

La visée de ce manuel est essentiellement pratique. La plupart des résultats théoriques sont énoncés sans démonstration. Si le lecteur souhaite connaître la justification mathématique des méthodes exposées, il devra se reporter aux ouvrages spécialisés cités en référence.

Généralement, on enseigne d'abord la statistique descriptive puis la statistique inférentielle. Cette disjonction se retrouve dans la plupart des ouvrages universitaires. Nous avons choisi au contraire de regrouper les aspects descriptifs et inférentiels pour offrir une vue d'ensemble des notions abordées.

Chaque chapitre du livre commence par un résumé de cours illustré sous R. Il se termine par une série de Questionnaires à Choix Multiples (QCM) dont la solution est fournie. **Un QCM comporte vingt items. Trois réponses sont proposées à chaque item dont une seule est correcte.** Les QCM permettront aux étudiants d'évaluer leurs connaissances au fil des semestres. Les enseignants y trouveront une source d'inspiration pour leurs propres examens.

Des technologies récentes comme Plickers ou Auto Multiple Choice faciliteront l'implémentation des QCM. Plickers : <https://www.plickers.com/> permet de collecter rapidement, à l'aide d'un simple téléphone mobile, un grand nombre de réponses et d'en faire une synthèse graphique. Auto Multiple Choice : <https://www.auto-multiple-choice.net/> permet de générer plusieurs QCM à partir d'une banque de questions puis de les corriger de façon automatique.

La statistique s'est enrichie du développement de l'informatique. Il y a quelques années, les praticiens étaient limités par la performance des ordinateurs et les théoriciens cantonnés à la statistique mathématique. Aujourd'hui, les machines sont puissantes, les logiciels sont nombreux, les forums en ligne facilitent les collaborations et une émulation nouvelle apparaît. Pour être en phase avec une pratique actuelle de la statistique, l'enseignement de la statistique doit être associé à celui d'un logiciel informatique. Nous avons choisi d'illustrer ce manuel sous R car c'est référence parmi les utilisateurs et les professionnels de la statistique.

R est un logiciel libre et gratuit, largement répandu dans le monde de la recherche, de l'enseignement et de l'entreprise. Il est développé par une communauté ouverte de chercheurs, de statisticiens, d'enseignants, d'ingénieurs, etc. Il peut être téléchargé sur la plateforme du CRAN (Comprehensive R Archive Network) à l'adresse <https://cran.r-project.org/>.

R possède un langage de programmation permettant aux utilisateurs expérimentés d'écrire leurs propres scripts. Mais il peut aussi s'employer à un niveau élémentaire avec Rcommander ou dans l'environnement de Rstudio. R comporte un grand nombre de librairies. Chaque librairie contient de nombreux jeux de données et de fonctions. Pour commencer rapidement avec R, nous vous invitons à lire la rubrique : À propos de R.

Des logiciels récents connaissent une forte expansion et pourraient concurrencer R, mais peu d'applications sont aussi adaptées à la statistique que R et les dix-mille librairies de R construites en une vingtaine d'années ne tomberont sans doute pas en désuétude avant longtemps.

Nous remercions chaleureusement Eugen Ursu pour sa lecture attentive de l'ouvrage et ses conseils toujours très judicieux. Nous remercions également par avance, tous ceux qui contribueront à améliorer ce manuel par de leurs remarques et leurs commentaires.

# À propos de R

Voici quelques points de repère sur l'utilisation de R. Pour acquérir une plus grande maîtrise du logiciel, nous recommandons l'ouvrage de Pierre Lafaye de Micheaux, Rémy Drouilhet et Benoît Liqueur : Le logiciel R : Maîtriser le langage. Effectuer des Analyses Statistiques. (2011). Springer.

## *Effectuer des calculs*

Comme une calculatrice R permet de réaliser des opérations élémentaires : `2+3;4-2;2*8;6/5;2.4^2` ou de déterminer des valeurs de fonctions usuelles : `exp(2);log(3);tan(pi/8)`.

Les fonctions de R tiennent leur nom de l'anglais. Ainsi, `mean()` désigne la moyenne, `sd()` l'écart-type (standard deviation en anglais), etc. La commande `builtins()` donne un aperçu des fonctions de base de R. On obtiendra plus de détails sur la fonction `sd` par exemple, avec : `help(sd)`.

R permet d'effectuer des calculs matriciels. La matrice définie par : `a=matrix(c(2,1,3,1),nrow=2,byrow=TRUE)` a pour déterminant `det(a)`, pour trace `sum(diag(a))`, son inverse est `solve(a)`, sa transposée `t(a)` et son carré `a%*%a`.

R permet d'effectuer des calculs dont les arguments ou les résultats ont une valeur logique. Ainsi : `islands>100` indiquera les îles du jeu de données `islands` dont la superficie excède 100 unités de surface.

## *Tracer des graphiques*

R trace la courbe de fonctions d'une variable : `plot(function(x) x*log(x),xlim=c(0,1))`, représente les fonction de deux variables par une surface : `help(persp)` et permet d'obtenir la plupart des diagrammes utiles en statistiques comme le montre la page : <https://www.r-graph-gallery.com/>.

Un graphique tracé sous R pourra être exporté, au format eps par exemple, grâce à : `dev.copy2eps(file='/chemin/figure.eps')`.

## *Utiliser une librairie*

Beaucoup de fonctions de R sont rangées dans des librairies/packages. La fonction `skewness()` de la librairie `moments` par exemple permet de calculer le coefficient d'asymétrie. Pour mesurer l'asymétrie des prix de `EuStockMarkets`, on doit (i) installer le package : `install.packages('moments')` (ii) ouvrir la librairie : `library('moments')` (iii) appliquer la fonction : `skewness(EuStockMarkets)`.

À l'étape (i) un miroir du CRAN doit être choisi. Cette étape est définitive : une fois le package chargé, il est définitivement installé ; il peut être mis à jour, mais il est inutile de le charger à chaque session. La librairie appelée à l'étape (ii) est ouverte pour la durée de la session : il est inutile d'appeler une librairie à chaque utilisation de l'une de ses fonctions.

`help` permet de connaître l'ensemble des fonctions d'un package. Ainsi `help(package='moments')` `$info` donne la liste des fonctions du package `moments`.

## *Importer des données*

Chaque librairie de R comporte des données propres. Par exemple, le package `likert` dédié aux données ordinales contient un jeu appelé `MathAnxiety`. Une fois le package installé, ces données peuvent être chargées par : `data(MathAnxiety)`. L'ensemble des jeux de données du package `likert` s'obtient avec : `data(package='likert')`.

Beaucoup de données peuvent aussi être téléchargées sur un dépôt n'ayant aucun rapport avec R. Les données Abalone de l'University of California Irvine par exemple sont constituées d'ordres décrits par des variables biologiques. Une explication détaillée de ces données est fournie à l'adresse <https://archive.ics.uci.edu/ml/datasets/abalone>. Les données elles-mêmes peuvent être consultées sous : <https://archive.ics.uci.edu/ml/machine-learning-databases/abalone/abalone.data>. On observe que : la première ligne ne contient pas d'en têtes, les variables sont séparées par des virgules et des points séparent la partie entière de la partie décimale de chaque mesure. Les données seront donc enregistrées sous : `mydata <- read.table(https://archive.ics.uci.edu/ml/machine-learning-databases/abalone/abalone.data,header=FALSE,sep=',',dec='.')`.

### *Manipuler des données*

Les données sont souvent consignées dans R sous forme d'un data frame. Il s'agit d'un tableau dont chaque colonne est une liste d'objets de même nature (nombres, lettres, valeurs logiques, etc.). Par exemple, si `goal=c(2,1,0,3)` indique le nombre de buts marqués par la Lazio au cours des quatre derniers matchs et `time=c(12,46,90,08)` la durée séparant le dernier but de la fin du match, les performances de l'équipe sont enregistrées dans `lazio <- data.frame(goal=goal,time=time)`. La variable `sun=c(TRUE,FALSE,TRUE,FALSE)` indiquant si le match s'est déroulé par beau temps (TRUE) ou non (FALSE) peut être ajoutée au data frame : `lazio[['sun']] <- sun`. Les données ainsi enregistrées peuvent alors être exportées au format csv : `write.csv(lazio,'/chemin/lazio.csv')` pour être exploitées en dehors de R.

### *Définir une fonction*

R permet de définir soi-même une fonction. Supposons que l'on souhaite calculer la moyenne et la variance d'une série privée de la plus petite et de la plus grande de ses valeurs. On pourra définir la fonction :

```
trimmed <- fonction(s){n=length(s) ; ts=sort(s)[-c(1,n)] ; m=mean(ts) ;  
v=var(ts) ; c(m,v)}.
```

La variable d'entrée `s` est une liste de nombres et la variable de sortie une autre liste contenant la moyenne et la variance : `m` et `v` de la série tronquée. Ainsi `trimmed(rivers)` donne la moyenne et la variance tronquées des longueurs de rivières américaines.

### *Écrire un script*

Un script est un ensemble de commandes écrites dans le langage de R et traduisant un algorithme. Voici un script qui donne la moyenne et l'écart-type d'observations et représente leur densité.

```
cat('entrez les observations','\n')  
data <- scan()  
plot(density(data))  
cat('-----','\n')  
cat('moyenne :',mean(data),'\n')  
cat('écart-type :',sd(data),'\n')
```

Ce script peut être copié dans un éditeur de texte puis enregistré sous le nom `short_script` dans un dossier appelé `folder`. Il est alors exécuté par la commande : `source('folder/short_script')`.

### *Pour aller plus loin*

Pour obtenir une description plus précise des fonctionnalités de R, on pourra se référer au site du CRAN dont la page d'entrée est : <https://cran.r-project.org/>. On trouvera également à l'adresse <https://cran.r-project.org/doc/manuals/r-release/R-intro.html> une aide sur les différents types d'objets accessibles sous R et sur leur manipulation.

Certains thèmes abordés par ce manuel sont traités sous R dans d'autres ouvrages. La table suivante indique la correspondance entre les thèmes abordés dans ce manuel et quelques livres de référence.

		<i>référence</i>				
		[1]	[2]	[3]	[4]	[5]
<i>thème</i>	statistique descriptive	✓	-	✓	-	-
	intervalles de confiance	✓	-	✓	-	-
	tests d'hypothèses	✓	-	✓	-	✓
	couples de variables qualitatives	✓	-	✓	-	✓
	couples de variables mixtes	✓	-	✓	-	✓
	couples de variables quantitatives	✓	✓	✓	✓	✓

[1] Pierre-André Cornillon, Arnaud Guyader, François Husson, Nicolas Jégou, Julie Josse, Maela Kloareg, Eric Matzner-Løber, and Laurent Rouvière. *Statistique avec R*. 2008

[2] Pierre-André Cornillon and Eric Matzner-Løber. *Régression avec R*. Springer Science & Business Media, 2010

[3] Pierre Lafaye De Micheaux, Rémy Drouilhet, and Benoît Liquet. *Le logiciel R : Maîtriser le langage-Effectuer des analyses statistiques*. Springer Science & Business Media, 2011

[4] François Husson, Sébastien Lê, and Jérôme Pagès. *Analyse de données avec R*. Presses universitaires de Rennes, 2016

[5] Gaël Millot. *Comprendre et réaliser les tests statistiques à l'aide de R*. Brussels : De Boeck, 2011



# Table des matières

Symboles et notations	13
<b>I Statistique univariée</b>	<b>17</b>
<b>1 Décrire une série</b>	<b>19</b>
1.1 L'essentiel du cours	19
1.1.1 Généralités	19
1.1.2 Séries quantitatives continues	20
1.1.3 Séries quantitatives discrètes	24
1.1.4 Séries qualitatives	26
1.1.5 Avec R	26
1.2 Questionnaires	30
1.2.1 QCM 1 La taille des enfants	30
1.2.2 QCM 2 Durée de travail des médecins	32
1.2.3 QCM 3 Distribution des fréquences	34
1.2.4 QCM 4 Les rendements financiers d'Ekson	38
1.2.5 QCM 5 Temps frais au pays d'Arth	40
1.2.6 QCM 6 Propriétés des statistiques usuelles	41
1.2.7 QCM 7 Les filles sont meilleures que les garçons	42
1.2.8 QCM 8 Genre et groupe sanguin	44
1.2.9 QCM R 1 Les éruptions d'Old Faithful	47
1.2.10 QCM R 2 Les pétales des iris de Fisher	49
1.2.11 QCM R 3 Saveurs d'Italie	50
1.2.12 QCM R 4 La puissance des autos	53
1.2.13 QCM R 5 La taille des familles parisiennes	54
1.2.14 QCM R 6 Quelques séries nominales	56
1.2.15 QCM R 7 Quelques séries ordinales	58
<b>2 Estimer un paramètre</b>	<b>61</b>
2.1 L'essentiel du cours	61
2.1.1 Généralités	61
2.1.2 Méthodes d'estimation	62
2.1.3 Avec R	66
2.2 Questionnaires	69
2.2.1 QCM 1 Qualité d'un estimateur	69
2.2.2 QCM 2 Comparer des estimateurs	71
2.2.3 QCM 3 Estimateurs basés sur la moyenne	72
2.2.4 QCM 4 Méthode des Moments	74
2.2.5 QCM 5 Méthode des Moments	75
2.2.6 QCM 6 Maximum de Vraisemblance	77
2.2.7 QCM 7 Echantillons gaussiens et Maximum de Vraisemblance	78
2.2.8 QCM 8 Intervalles de confiance	79

2.2.9	QCM 9 Intervalles de confiance . . . . .	81
2.2.10	QCM 10 Inférence Bayésienne et lois discrètes . . . . .	83
2.2.11	QCM 11 Inférence Bayésienne et lois continues . . . . .	84
2.2.12	QCM R 1 Maximum de Vraisemblance et lois discrètes . . . . .	86
2.2.13	QCM R 2 Maximum de Vraisemblance et modèle Gaussien . . . . .	88
2.2.14	QCM R 3 Méthode des Moments . . . . .	90
2.2.15	QCM R 4 Intervalles de confiance . . . . .	93
2.2.16	QCM R 5 Estimation Bayésienne . . . . .	96

### **3 Tester une hypothèse 99**

3.1	L'essentiel du cours . . . . .	99
3.1.1	Généralités . . . . .	99
3.1.2	Comparaison d'un paramètre à une valeur de référence . . . . .	102
3.1.3	Comparaison de deux paramètres . . . . .	105
3.1.4	Avec R . . . . .	108
3.2	Questionnaires . . . . .	115
3.2.1	QCM 1 Des ampoules de Lux . . . . .	115
3.2.2	QCM 2 Carnassiers d'eau douce . . . . .	116
3.2.3	QCM 3 Les biscuits de Belem . . . . .	118
3.2.4	QCM 4 La fréquence du gène A . . . . .	119
3.2.5	QCM 5 Comparer une moyenne/une variance à une valeur de référence . . . . .	121
3.2.6	QCM 6 Saveurs du Jura . . . . .	123
3.2.7	QCM 7 Le temps des cerises . . . . .	124
3.2.8	QCM 8 Comparer des proportions . . . . .	125
3.2.9	QCM 9 Test d'hypothèse et taille d'échantillon . . . . .	127
3.2.10	QCM R 1 Comparer une moyenne/une variance à une valeur de référence . . . . .	129
3.2.11	QCM R 2 Comparer une proportion à une valeur de référence . . . . .	131
3.2.12	QCM R 3 Comparer deux proportions . . . . .	133
3.2.13	QCM R 4 Comparer une proportion/une variance à une valeur de référence . . . . .	134
3.2.14	QCM R 5 Test d'hypothèse et simulation . . . . .	136

## **II Statistique bivariée 139**

### **4 Variables qualitatives 141**

4.1	L'essentiel du cours . . . . .	141
4.1.1	Mesurer la dépendance . . . . .	141
4.1.2	Tester l'indépendance (test du $\chi^2$ ) . . . . .	143
4.1.3	Avec R . . . . .	144
4.2	Questionnaires . . . . .	147
4.2.1	QCM 1 Beaux et bons champignons . . . . .	147
4.2.2	QCM 2 Groupe sanguin et facteur rhésus . . . . .	148
4.2.3	QCM 3 Statistiques pour couples de variables qualitatives . . . . .	150
4.2.4	QCM 4 Test d'indépendance du $\chi^2$ . . . . .	152
4.2.5	QCM R 1 Genre, couleur des yeux et des cheveux . . . . .	156
4.2.6	QCM R 2 Les passagers du Titanic . . . . .	158

### **5 Variables mixtes 161**

5.1	L'essentiel du cours . . . . .	161
5.1.1	Mesurer la structuration . . . . .	161
5.1.2	Analyse de la variance (à un facteur) . . . . .	162
5.1.3	Avec R . . . . .	163
5.2	Questionnaires . . . . .	166

5.2.1	QCM 1 ANOVA sur la truite . . . . .	166
5.2.2	QCM 2 Engrais et rendements . . . . .	167
5.2.3	QCM 3 Test d'ANOVA . . . . .	168
5.2.4	QCM R La légèreté des hommes et des femmes . . . . .	171
<b>6</b>	<b>Variables quantitatives</b>	<b>173</b>
6.1	L'essentiel du cours . . . . .	173
6.1.1	Décrire un nuage de points . . . . .	173
6.1.2	Analyse en composantes principales (rôle symétrique des variables) . . . . .	175
6.1.3	Ajustement linéaire (rôle asymétrique des variables) . . . . .	177
6.1.4	Avec R . . . . .	178
6.2	Questionnaires . . . . .	182
6.2.1	QCM 1 Lumière et chaleur des villes suédoises . . . . .	182
6.2.2	QCM 2 Statistiques pour couple de variables qualitatives . . . . .	183
6.2.3	QCM 3 Le chant des oiseaux (ACP) . . . . .	184
6.2.4	QCM 4 Analyse en composantes principales . . . . .	186
6.2.5	QCM 5 Sur les couples salariés (ACP) . . . . .	188
6.2.6	QCM 6 Aux débuts de la vie . . . . .	188
6.2.7	QCM 7 Chiffre d'affaire et nombre de salariés . . . . .	190
6.2.8	QCM R 1 Les éruptions d'Old Faithful . . . . .	191
6.2.9	QCM R 2 Age et circonférence des arbres . . . . .	193
6.2.10	QCM R 3 John's place et Jack's bar . . . . .	195



# Symboles et notations

## Statistique descriptive

---

$x_i$	observation $i$
$n$	nombre d'observations
$\bar{x}$	moyenne des observations
$s^2$	variance observée
$s'^2$	variance corrigée
$x_{(i)}$	statistique d'ordre de rang $i$
$c_k$	borne supérieure de la classe $k$
$f_k$	fréquence associée à la classe $k$
$[z]$	partie entière de $z$
$q_\alpha$	quantile d'ordre $\alpha$
$m_e$	médiane des observations
$cv$	coefficient de variation
$EMA$	écart moyen absolu
$IQ$	écart inter-quartiles
$e$	étendue
$\beta$	coefficient d'aplatissement
$\gamma$	coefficient d'asymétrie
$f$	fonction de probabilité/densité
$F$	fonction de répartition
$\mathbb{1}$	fonction indicatrice
$\max$	maximum
$\min$	minimum

## Estimation et intervalles de confiance

---

$x_i$	observation $i$
$n$	nombre d'observations
$\mathcal{X}$	espace des observations
$f(\bullet, \theta)$	fonction de probabilité/densité des observations
$\theta$	paramètre à estimer
$\Theta$	espace du paramètre
$E(\hat{\theta})$	espérance de l'estimateur $\hat{\theta}$
$B(\hat{\theta})$	biais de l'estimateur $\hat{\theta}$
$V(\hat{\theta})$	variance de l'estimateur $\hat{\theta}$
$MSE(\hat{\theta})$	erreur quadratique moyenne de l'estimateur $\hat{\theta}$
$I(\theta)$	information de Fisher
$L(\theta; x_1, \dots, x_n)$	vraisemblance de $\theta$
$\ell(\theta; x_1, \dots, x_n)$	log-vraisemblance de $\theta$
$\hat{\theta}_{MV}$	estimateur du Maximum de Vraisemblance

$\hat{\theta}_{MM}$	estimateur issu de la méthode des moments
$IC_{1-\alpha}(\theta)$	intervalle de confiance de niveau $1 - \alpha$ de $\theta$
$\mu$	moyenne théorique
$\sigma^2$	variance théorique
$\bar{x}$	moyenne empirique
$s^2$	variance observée
$s'^2$	variance corrigée
$p$	fréquence théorique
$f$	fréquence empirique
$p(x_1, \dots, x_n; \theta)$	vraisemblance (cadre Bayésien)
$p(\theta)$	loi a priori (cadre Bayésien)

## Tests d'hypothèses

---

$\mathcal{H}_0$	hypothèse nulle
$\mathcal{H}_1$	hypothèse alternative
$\alpha$	seuil de significativité
$\beta$	risque de seconde espèce
SdT	statistique de test
ZdR	zone de rejet
$p$ -val	$p$ valeur
$\theta_0$	valeur du paramètre $\theta$ sous $\mathcal{H}_0$
$z_{\text{obs}}$	valeur observée de la statistique $z$

## Lois de probabilité

---

$\phi$	fonction de répartition de la loi normale centrée réduite
$\mathcal{N}(\mu; \sigma^2)$	loi normale de moyenne $\mu$ et de variance $\sigma^2$
$\mathcal{B}(n; p)$	loi binomiale de paramètres $n$ et $p$
$\mathcal{F}_\nu$	loi de Student à $\nu$ degrés de liberté
$\chi_\nu^2$	loi de chi deux à $\nu$ degrés de liberté
$\chi_\nu^2(\lambda)$	loi de chi deux décentrée à $\nu$ degrés de liberté et de paramètre de non centralité $\lambda$
$\mathcal{F}_{r,s}^2$	loi de Fisher à $r$ et $s$ degrés de liberté
$\mathcal{F}_{r,s}^2(\lambda)$	loi de Fisher décentrée à $r$ et $s$ degrés de liberté et de paramètre de non centralité $\lambda$
$z_\alpha$	quantile d'ordre $\alpha$ de la loi normale centrée réduite
$b_{n;p;\alpha}$	quantile d'ordre $\alpha$ de la loi binomiale de paramètres $n$ et $p$
$t_{\nu;\alpha}$	quantile d'ordre $\alpha$ de la loi de Student à $\nu$ degrés de liberté
$\chi_{\nu;\alpha}^2$	quantile d'ordre $\alpha$ de la loi de $\chi^2$ à $\nu$ degrés de liberté
$w_{r,s;\alpha}$	quantile d'ordre $\alpha$ de la loi de Fisher à $r$ et $s$ degrés de liberté

## Croisement de variables qualitatives

---

$n_{i,j}$	nombre d'observations $(x_i, y_j)$ du couple $(x, y)$
$n_{i,\bullet}$	nombre d'observations de la valeur $x_i$ de la variable $x$
$n_{\bullet,j}$	nombre d'observations de la valeur $y_j$ de la variable $y$
$f_{i,j}$	fréquence de la valeur $(x_i, y_j)$ du couple $(x, y)$
$f_{i,\bullet}$	fréquence de la valeur $x_i$ de la variable $x$
$f_{\bullet,j}$	fréquence de la valeur $y_j$ de la variable $y$
$g_{j i}$	fréquence de la valeur $y_j$ de $y$ conditionnellement à la valeur $x_i$ de $x$
$h_{i j}$	fréquence de la valeur $x_i$ de $x$ conditionnellement à la valeur $y_j$ de $y$
$n_{i,j}^*$	effectif théorique pour la valeur $(x_i, y_j)$ du couple $(x, y)$
$\mathcal{X}^2$	statistique du test de $\chi^2$ d'indépendance

$n_i$	effectif du groupe $i$
$\bar{x}_i$	moyenne empirique dans le groupe $i$
$s_i^2$	variance observée dans le groupe $i$
$\bar{x}$	moyenne marginale
$s^2$	variance totale
$s_F^2$	variance factorielle
$s_R^2$	variance résiduelle
$SC_R$	somme des carrés résiduels
$SC_F$	somme des carrés factoriels
$SC_T$	somme des carrés totaux
$\eta^2$	rapport de corrélation

$x_i$	une observation de $x$
$y_i$	une observation de $y$
$z_i$	le couple $(x_i, y_i)'$
$\bar{x}$	moyenne des observations $x_i$
$\bar{y}$	moyenne des observations $y_i$
$\bar{z}$	point moyen
$s_x^2$	variance des observations $x_i$
$s_y^2$	variance des observations $y_i$
$s_x$	écart-type des observations $x_i$
$s_y$	écart-type des observations $y_i$
$c_{x,y}$	covariance des couples $(x_i, y_i)$
$r_{x,y}$	coefficient de corrélation linéaire des couples $(x_i, y_i)$
$\mathcal{I}_\omega$	inertie du nuage par rapport à $\omega$
$\ u\ $	norme euclidienne de $u$
$x_i^\circ$	observation issue du centrage de $x_i$
$y_i^\circ$	observation issue du centrage de $y_i$
$x^\circ$	la série des données centrées de $x$
$y^\circ$	la série des données centrées de $y$
$c_{x^\circ, y^\circ}$	covariance des observations centrées
$r_{x^\circ, y^\circ}$	coefficient de corrélation des observations centrées
$x_i^*$	observation issue du centrage et de la réduction de $x_i$
$y_i^*$	observation issue du centrage et de la réduction de $y_i$
$x^*$	la série des données centrées et réduites de $x$
$y^*$	la série des données centrées et réduites de $y$
$c_{x^*, y^*}$	covariance des observations centrées réduites
$r_{x^*, y^*}$	coefficient de corrélation des observations centrées réduites
$Z$	matrice des observations
$Z^0$	matrice des observations centrées
$Z^*$	matrice des observations centrées réduites
$S$	matrices des covariances observées
$T$	matrice des écarts-types observés
$R$	matrice des corrélations observées

---

$u_j^*$	vecteur unitaire dirigeant l'axe factoriel $j$
$l_j^*$	inertie du nuage projeté sur l'axe factoriel $j$
$x_i$	abscisse du point $i$ sur le premier axe factoriel
$y_i$	abscisse du point $i$ sur le second axe factoriel
$\bar{x}$	moyenne des observations $x_i$
$\bar{y}$	moyenne des observations $y_i$
$s_x^2$	variance de la première composante principale
$s_y^2$	variance de la seconde composante principale
$t_i^x$	contribution du point $i$ à la première composante principale
$t_i^y$	contribution du point $i$ à la seconde composante principale
$c_{x,x}$	covariance de $x$ et $x$
$r_{x,x}$	coefficient de corrélation linéaire de $x$ et $x$

## Régression linéaire

---

$\beta_0$	ordonnée à l'origine de la droite de régression
$\beta_1$	pende de la droite de régression
$\epsilon_i$	erreur $i$
$SCF$	somme des carrés factoriels
$SCR$	somme des carrés résiduels
$SCT$	somme des carrés totaux
$R^2$	coefficient de détermination

## Divers

---

$\mathbf{1}$	vecteur dont les composantes valent 1
$\text{tr}(M)$	trace de la matrice $M$
$M'$	transposée de la matrice $M$
$\Gamma$	fonction gamma d'Euler
$\mathbb{1}$	fonction indicatrice
$E$	espérance
$V$	variance
$\mathbb{P}(A)$	probabilité de l'événement $A$

---

Ce manuel s'adresse à tous les étudiants de premier cycle universitaire dont le cursus comporte un enseignement de statistique: Licence, DUT, PACES, CPGE. Il s'inspire de la démarche statistique dans un contexte pratique: décrire une série, estimer un paramètre, tester une hypothèse, etc. Il traite aussi bien de séries univariées que bivariées.

Les notions abordées font l'objet de rappels de cours concis, structurés et sans démonstrations. Ces notions sont illustrées sous R, un logiciel de statistique très largement utilisé par une vaste communauté de praticiens. Enfin, de nombreuses études de cas sont proposées sous la forme de Questionnaires à Choix Multiples de difficulté variable. Les données étudiées sont diverses : biologiques, médicales, sociologiques, économiques, financières. Ainsi, chacun trouvera dans ce manuel des applications de la statistique qui lui sont familières.

Les étudiants trouveront dans ce livre:

- des références théoriques sous forme de rappels de cours succincts mais complets;
- un point de vue original sur les notions abordées mêlant aspects descriptifs et inférentiels;
- des QCMs corrigés permettant d'évaluer leurs connaissances en statistique;

Les enseignants trouveront dans cet ouvrage:

- une source d'inspiration pour leurs examens, leurs travaux dirigés ou leurs travaux pratiques sous R,
- des QCMs sur lesquels leurs étudiants peuvent travailler directement et de façon autonome, à la maison ou en cours,
- des questions inédites mettant en évidence l'intérêt d'un logiciel de statistique.

Alexandre Lourme est professeur agrégé de mathématiques et docteur en statistique appliquée. Après avoir exercé en IUT, dans un département de biologie, il enseigne aujourd'hui les mathématiques et la statistique au sein de l'UFR de Science Économique & AES de Bordeaux. Il est intervenu et intervient encore dans différentes universités, écoles d'ingénieurs ou organismes de formation: Conservatoire National des Arts et Métiers, Universités Technologiques, TELECOM Lille, réseau Polytech, etc.

