

Collection



Statistique  
et probabilités  
appliquées

$$\begin{aligned} \text{variance}(Y) &= \frac{1}{n} \sum_i (y_i - \bar{y})^2 \\ &= \frac{1}{n} \left( \sum_i y_i^2 - n\bar{y}^2 \right) \\ &= \frac{1}{n} \sum_i y_i^2 - \bar{y}^2 \\ &= \text{mean}(Y^2) - \text{mean}^2(Y) \end{aligned}$$

Valentin Rousson

# Statistique appliquée aux sciences de la vie

$$\begin{aligned} \sum_i (y_i - \bar{y})^2 &= \sum_i y_i^2 - 2\bar{y} \sum_i y_i + n\bar{y}^2 \\ &= \sum_i y_i^2 - 2\bar{y} \sum_i y_i + n\bar{y}^2 \\ &= \sum_i y_i^2 - n\bar{y}^2 \end{aligned}$$

$$\text{Var} \left( \sum_i Y_i \right) = \sum_i \text{Var}(Y_i)$$

$$\text{Var}(\hat{\mu}) = \text{Var} \left( \frac{\sum_i Y_i}{n} \right) = \frac{\sum_i \text{Var}(Y_i)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

**L**avoisier  
hermes



# Sommaire

## Préface

vii

## Table des matières

ix

<b>1 Premiers concepts</b>	<b>1</b>
1.1 Population, variable et échantillon . . . . .	2
1.2 Échantillonnage et indépendance . . . . .	3
1.3 Principaux types de variables . . . . .	4
<b>2 Distribution d'une variable</b>	<b>7</b>
2.1 Distribution d'une variable qualitative . . . . .	7
2.2 Distribution d'une variable continue . . . . .	9
2.3 Densité de probabilité . . . . .	11
2.4 Boxplot et quantiles . . . . .	14
2.5 Mesures de tendance centrale . . . . .	18
2.6 Mesures de variabilité . . . . .	20
2.7 Changement d'unités . . . . .	22
2.8 Distribution normale . . . . .	23
2.9 Distribution normale standardisée . . . . .	26
2.10 Variable standardisée et qq-plot . . . . .	29
2.11 Mesures de non-normalité . . . . .	31
2.12 Transformation logarithmique . . . . .	33
2.13 Distribution d'une variable binaire . . . . .	35
<b>3 Estimation</b>	<b>37</b>
3.1 Distribution d'un estimateur . . . . .	38
3.2 Variable aléatoire . . . . .	38
3.3 Distribution de la moyenne d'un échantillon . . . . .	40
3.4 Distribution de la variance d'un échantillon . . . . .	44
3.5 Distribution d'une proportion calculée dans un échantillon . . . . .	46
3.6 Distribution exacte d'une proportion calculée dans un échantillon . . . . .	47

<b>4</b>	<b>Intervalle de confiance</b>	<b>49</b>
4.1	Méthode de Wald . . . . .	49
4.2	Intervalle de confiance de Wald pour une moyenne . . . . .	52
4.3	Intervalle de confiance de Student pour une moyenne . . . . .	53
4.4	Niveau nominal et niveau réel d'un intervalle de confiance . . . . .	56
4.5	Intervalle de confiance et intervalle de prédiction . . . . .	59
4.6	Transformation logarithmique . . . . .	61
4.7	Intervalle de confiance pour une variance . . . . .	62
4.8	Intervalle de confiance de Wald pour une proportion . . . . .	63
4.9	Intervalle de confiance de Wilson pour une proportion . . . . .	64
<b>5</b>	<b>Comparaison de deux distributions</b>	<b>67</b>
5.1	Différence de moyenne . . . . .	68
5.2	Intervalle de confiance de Wald pour une différence de moyenne . . . . .	70
5.3	Intervalle de confiance de Student pour une différence de moyenne . . . . .	71
5.4	Intervalle de confiance de Welch pour une différence de moyenne . . . . .	73
5.5	Validité des intervalles de confiance pour différence de moyenne . . . . .	73
5.6	Transformation logarithmique . . . . .	74
5.7	Différence de moyenne standardisée . . . . .	77
5.8	Quotient de variance . . . . .	79
5.9	Différence de proportion . . . . .	81
<b>6</b>	<b>Principe d'un test statistique</b>	<b>83</b>
6.1	L'hypothèse nulle et l'hypothèse alternative . . . . .	83
6.2	Erreurs de première et de seconde espèce . . . . .	84
6.3	Concept de valeur $p$ . . . . .	85
6.4	Tests multiples . . . . .	88
6.5	Statistique de test . . . . .	89
<b>7</b>	<b>Tests du khi-deux pour tables de contingence</b>	<b>91</b>
7.1	Comparaison de distributions de variables qualitatives . . . . .	91
7.2	Comparaison d'une distribution qualitative avec distribution de référence . . . . .	96
7.3	Comparaison de deux proportions . . . . .	97
7.4	Comparaison d'une proportion avec valeur de référence . . . . .	100
<b>8</b>	<b>Test statistique sur la valeur d'un paramètre</b>	<b>103</b>
8.1	Test unilatéral et test bilatéral . . . . .	103
8.2	Test statistique <i>versus</i> intervalle de confiance . . . . .	108
8.3	Test d'équivalence . . . . .	111
<b>9</b>	<b>Tests de Wald et de Student</b>	<b>115</b>
9.1	Méthode de Wald . . . . .	115
9.2	Test de Wald pour une proportion . . . . .	116
9.3	Test de Wald pour une différence de proportion . . . . .	117
9.4	Test de Student pour une moyenne . . . . .	119

9.5	Test de Welch pour une différence de moyenne . . . . .	120
9.6	Test de Student pour une différence de moyenne . . . . .	122
9.7	Test de Student pour données paires . . . . .	123
<b>10</b>	<b>Calcul de taille d'échantillon</b> . . . . .	<b>127</b>
10.1	Valeur <i>p</i> versus taille de l'échantillon . . . . .	128
10.2	Puissance d'un test statistique . . . . .	129
10.3	Exemples de calculs de taille d'échantillon . . . . .	133
<b>11</b>	<b>Tests exacts avec statistique de test discrète</b> . . . . .	<b>139</b>
11.1	Test binomial pour une proportion . . . . .	139
11.2	Comparaison des tests binomial et du khi-deux . . . . .	142
11.3	Test exact de Fisher . . . . .	147
11.4	Test de McNemar . . . . .	150
11.5	Test du signe . . . . .	152
11.6	Test de Mann-Whitney . . . . .	153
11.7	Comparaison des tests de Welch, Student et Mann-Whitney . . . . .	158
11.8	Test de Wilcoxon . . . . .	161
11.9	Récapitulatif des tests statistiques . . . . .	162
<b>12</b>	<b>Analyse de corrélation</b> . . . . .	<b>163</b>
12.1	Diagramme de dispersion . . . . .	163
12.2	Covariance . . . . .	166
12.3	Corrélation de Pearson . . . . .	169
12.4	Corrélation versus causalité . . . . .	173
12.5	Corrélation et choix de la population . . . . .	174
12.6	Distribution normale bivariée . . . . .	178
12.7	Corrélation de Spearman . . . . .	180
12.8	Inférence sur la corrélation . . . . .	184
<b>13</b>	<b>Régression linéaire simple</b> . . . . .	<b>187</b>
13.1	Droite de régression . . . . .	187
13.2	Droite de régression sur la population . . . . .	192
13.3	Variance prédite et variance résiduelle . . . . .	193
13.4	Hypothèse de linéarité . . . . .	196
13.5	Interprétation des paramètres de la droite de régression . . . . .	200
13.6	Modèle de régression linéaire simple . . . . .	203
13.7	Inférence sur la droite de régression . . . . .	206
13.8	Intervalle de prédiction . . . . .	212
13.9	Régression vers la moyenne . . . . .	215
<b>14</b>	<b>Régression linéaire multiple</b> . . . . .	<b>219</b>
14.1	Hyperplan de régression . . . . .	220
14.2	Hypothèse de linéarité . . . . .	222
14.3	Interprétation des paramètres . . . . .	224
14.4	Ajustement pour les variables confondantes . . . . .	228

14.5	Corrélation partielle . . . . .	231
14.6	Modèle de régression linéaire multiple . . . . .	232
14.7	Analyse des résidus . . . . .	234
14.8	Inférence sur l'hyperplan de régression . . . . .	240
14.9	Estimation du pourcentage de la variance prédite . . . . .	242
14.10	Tests sur la nullité de plusieurs paramètres . . . . .	243
14.11	Multicolinéarité . . . . .	246
14.12	Intervalle de prédiction . . . . .	249
14.13	Choix du modèle . . . . .	252
14.14	Valeurs aberrantes et points leviers . . . . .	255
<b>15</b>	<b>Régression avec prédicteurs binaires</b>	<b>259</b>
15.1	Comparaison de deux groupes . . . . .	259
15.2	Comparaison de deux groupes dans une étude observationnelle	262
15.3	Comparaison de deux groupes dans un essai clinique . . . . .	264
15.4	Planification d'une expérience . . . . .	267
15.5	Analyse de variance . . . . .	270
15.6	Analyse de covariance . . . . .	273
<b>16</b>	<b>Régression logistique</b>	<b>279</b>
16.1	Odds et odds-ratio . . . . .	279
16.2	Étude cas-témoins . . . . .	282
16.3	Inférence sur l'odds-ratio . . . . .	284
16.4	Régression logistique simple . . . . .	286
16.5	Régression logistique multiple . . . . .	291
16.6	Ajustement pour les variables confondantes . . . . .	296
16.7	Comparaison de deux groupes dans un essai clinique . . . . .	300
16.8	Sensibilité, spécificité et courbe ROC . . . . .	302
16.9	Vérification du modèle . . . . .	307
16.10	Méthode du maximum de vraisemblance . . . . .	309
<b>A</b>	<b>Tableaux</b>	<b>313</b>
	<b>Bibliographie</b>	<b>318</b>



Collection  
**Statistique  
et probabilités  
appliquées**

**Dirigée par  
Yadolah Dodge**

COMITÉ ÉDITORIAL:

**Aurore Delaigle**

Université de Melbourne, Australie

**Christian Genest**

Université Laval, Québec

**Marc Hallin**

Université libre de Bruxelles, Belgique

**Ludovic Lebart**

Télécom-ParisTech, Paris

**Christian Mazza**

Université de Fribourg, Suisse

**Stephan Morgenthaler**

EPFL, Lausanne

**Louis-Paul Rivest**

Université Laval, Québec

**Gilbert Saporta**

CNAM, Paris

**Valentin Rousson**

# **Statistique appliquée aux sciences de la vie**

*Cette collection met à la disposition du public intéressé par la statistique (étudiants, enseignants, chercheurs) des ouvrages qui concilient effort pédagogique et travail permanent de mise à jour.*

*Cette démarche implique de prendre en compte de façon sélective et critique les renouvellements des concepts, des champs d'application et des outils de traitement. Seules une compréhension profonde et une appropriation des connaissances permettront de s'adapter aux évolutions qui n'ont pas fini de bouleverser cette discipline.*

Cet ouvrage propose une introduction à la statistique sans qu'aucune connaissance préalable ne soit nécessaire. À partir du concept central de « variabilité », l'auteur aborde les notions de distribution, de statistique descriptive, d'estimation, d'intervalle de confiance, de test statistique, de corrélation et de modélisation statistique (régression linéaire et logistique), tout en recherchant un certain équilibre entre une description littérale des concepts et un minimum de formalisme mathématique.

Des problématiques plus techniques comme le calcul de la taille d'un échantillon, la question de la validité d'un intervalle de confiance, le principe d'un test d'équivalence ou le choix d'un modèle de régression sont également présentées.

Ce texte a été écrit à l'intention des étudiants des sciences de la vie (par exemple biologie ou médecine) mais s'adresse aussi aux étudiants et chercheurs d'autres domaines désirant s'initier à la statistique et se préparer dans les meilleures conditions à aborder des ouvrages statistiques plus avancés.

