

# TECHNOSUP

Les FILIÈRES TECHNOLOGIQUES des ENSEIGNEMENTS SUPÉRIEURS

## STATISTIQUES

# Statistique mathématique

Applications commentées

Jean-Pierre BOULAY

ellipses

# TABLE DES MATIERES

## Chapitre I : Echantillonnage

### A - Rappels de cours

1. Lois de probabilités de base rencontrées en statistique	1
1.1 Définitions et caractérisations	1
1.2 Les propriétés de convergence	3
2. Statistiques et distributions d'échantillonnage	5
2.1 Le principe de l'inférence statistique	5
2.2 Cas d'une moyenne	5
2.3 Cas d'une proportion	6
2.4 Cas d'une variance	7
2.5 Récapitulatif concernant espérance, proportion, et variance	8
3. La pratique de l'échantillonnage	10

### B – Applications

1. Distributions d'échantillonnage et propriétés	11
1.1 Moyenne et variance dans le cas d'échantillons gaussiens	11
1.2 Paramètres représentatifs des statistiques décrivant la variance	16
1.3 Distribution d'échantillonnage des rapports de variances	19
1.4 Distribution d'échantillonnage des différences de moyennes	22
1.5 Distribution d'échantillonnage des différences de proportions	26
1.6 La différence entre estimation et estimateur	27
2. Exemples de méthodes d'échantillonnage	30
2.1 Les sondages aléatoires sans replacement (exhaustifs)	30
2.2 Les sondages par stratification	33

### C – Exercices complémentaires

42

## Chapitre II : Estimation

### A – Rappels de cours

1. La problématique de l'estimation statistique	51
2. Propriétés des estimateurs ponctuels	52
2.1 Qualités d'un bon estimateur	52
2.2 Comparaison des estimateurs	52
2.3 Information de FISHER	53
2.4 Inégalité de CRAMER RAO	55
2.5 Statistiques exhaustives	56
2.6 Le cas particulier de la famille exponentielle	57

<b>3. Construction des estimateurs</b>	<b>58</b>
3.1 <i>Théorèmes de RAO-BLACKWELL et LEHMANN-SCHEFFE</i>	58
3.2 <i>Méthode des moments</i>	60
3.3 <i>Méthode du maximum de vraisemblance</i>	61
3.4 <i>Méthode des moindres carrés</i>	64
3.5 <i>Espérance, proportion, variance, et covariance</i>	64
<b>4. Estimation par intervalle de confiance</b>	<b>65</b>
4.1 <i>Construction de l'intervalle de confiance</i>	65
4.2 <i>Le cas d'une moyenne</i>	66
4.3 <i>Le cas d'une proportion</i>	67
4.4 <i>Le cas d'une variance</i>	68
<b>B – Applications</b>	
<b>1. Exemples de modèles et propriétés des estimateurs</b>	<b>68</b>
1.1 <i>Modèle gaussien</i>	68
1.2 <i>Modèle de POISSON</i>	71
1.3 <i>Modèle uniforme</i>	74
1.4 <i>Modélisation d'une hauteur de crue (loi de RAYLEIGH)</i>	78
1.5 <i>Modélisation de la durée de vie de diodes (loi de WEIBULL)</i>	80
1.6 <i>Modèle de PARETO</i>	82
1.7 <i>Modèle exponentiel translaté</i>	85
<b>2. Techniques particulières d'estimation</b>	<b>88</b>
2.1 <i>Modèle mélangé POISSON/Gamma en assurance automobile</i>	88
2.2 <i>Comment estimer un paramètre intime</i>	94
2.3 <i>Comptage des poissons dans un lac (méthode de capture et recapture)</i>	96
2.4 <i>Estimateur du nombre de fraudeurs dans un transport collectif</i>	98
2.5 <i>Evaluation d'une contamination (méthode most powerful number)</i>	100
2.6 <i>Evaluation de <math>\pi</math> à travers deux méthodes de MONTE-CARLO</i>	104
<b>3. Intervalles de confiance</b>	<b>108</b>
3.1 <i>Comparaison des méthodes d'approximation pour une proportion</i>	108
3.2 <i>Sondages de popularité</i>	109
3.3 <i>Contrôle de fabrication par mesures</i>	111
3.4 <i>Intervalles de confiance d'une moyenne pour la loi de POISSON</i>	115
3.5 <i>Une méthode par simulation, le bootstrap</i>	118
<b>C – Exercices complémentaires</b>	<b>122</b>

## Chapitre III : Décision

### A – Rappels de cours

<b>1. Les principes généraux de la décision statistique</b>	<b>139</b>
1.1 <i>L'objet des tests d'hypothèse</i>	139
1.2 <i>Les risques associés</i>	139
1.3 <i>La classification des tests</i>	141
<b>2. Les tests paramétriques</b>	<b>141</b>
2.1 <i>Hypothèses simples et multiples</i>	141
2.2 <i>La construction de la règle de décision</i>	142

2.3 Tests de conformité à une valeur standard	143
a) Le cas d'une moyenne	143
b) Le cas d'une proportion	145
c) Le cas d'une variance	145
d) Autres tests de conformité	146
e) Le cas des hypothèses composites	147
2.4 Tests de comparaison entre deux échantillons indépendants	148
a) La comparaison de variances (test de FISHER-SNEDECOR)	148
b) La comparaison de moyennes (test t de STUDENT)	149
c) La comparaison de proportions	151
2.5 Tests de comparaison entre deux échantillons appariés	151
2.6 Tests de comparaisons entre K échantillons indépendants, ( $K > 2$ )	152
a) L'analyse de la variance (ANOVA)	152
b) Comparaison de variances (test de BARTLETT)	156
2.7 Tests progressifs	156
<b>3. Les tests non paramétriques</b>	<b>158</b>
<b>3.1 Tests d'adéquation</b>	<b>158</b>
a) Le test du Chi- Deux	158
b) Le test de KOLMOGOROV	159
c) Le test de normalité de SHAPIRO et WILK	160
d) La méthode graphique de la droite de HENRY	161
<b>3.2 Tests de comparaison entre K échantillons indépendants</b>	<b>162</b>
a) Le test d'identité de KOLMOGOROV-SMIRNOV, ( $K=2$ )	162
b) Les tests d'identité de MANN-WHITNEY et WILCOXON, ( $K=2$ )	164
c) Le choix du test approprié	166
d) Le test d'identité de KRUSKAL-WALLIS, ( $K \geq 2$ )	166
<b>3.3 Tests de comparaison entre K échantillons appariés</b>	<b>167</b>
a) Le test d'identité des signes, ( $K=2$ )	167
b) Le test d'identité des rangs « signés » de WILCOXON, ( $K=2$ )	169
c) Le test d'identité de MAC NEMAR, (variables binaires et $K=2$ )	170
d) Le test d'identité de COCHRAN, (variables binaires et $K \geq 2$ )	171
e) le test d'identité de FRIEDMAN, ( $K \geq 2$ )	172
<b>3.4 Tests d'associations, (<math>K=2</math>)</b>	<b>173</b>
a) Le coefficient de corrélation des rangs Rho de SPEARMAN	174
b) Le coefficient de corrélation des rangs Tau de KENDALL	176
c) Le test de contingence de Chi- Deux	178
<b>B – Applications</b>	
1. Tests à un échantillon sous modèle gaussien	179
1.1 Test t de STUDENT et pluviométrie	179
1.2 Test de proportion et étude de marché	180
1.3 Risques client et fournisseur	181
1.4 Test séquentiel de WALD portant sur une moyenne	183
1.5 Ajustements par une loi normale	188
2. Tests à un échantillon sous autres modèles	191
2.1 Test paramétrique pour le modèle de POISSON	191
2.2 Test paramétrique pour le modèle de RAYLEIGH	194
2.3 Tests portant sur un modèle de revenus « PARETO »	199
2.4 Test paramétrique entre deux lois pour une étude de clientèle	202
2.5 Test séquentiel de WALD et contrôle de réception	205

2.6 Ajustement par une loi uniforme	209
2.7 Tests non paramétriques de conformité à une valeur standard	209
3. Tests à deux échantillons sous modèle gaussien	213
3.1 Un exemple utilisant les tests de STUDENT et de FISHER SNEDECOR	213
3.2 Comparaison de moyennes sur échantillons appariés	216
3.3 Comparaison de variances entre deux types de solutions aqueuses	217
3.4 Comparaison de proportions	222
3.5 Tables de contingences (2,2) et échantillons indépendants	224
3.6 Corrélation entre taille et poids (coefficient $r$ de PEARSON)	228
4. Tests à deux échantillons sous autres modèles	231
4.1 Test paramétrique de comparaison sous modèle exponentiel	231
4.2 Comparaison du test de WILCOXON avec le test paramétrique	233
4.3 Au sujet du traitement des ex-aequo dans les tests de rangs	238
4.4 Etude de tendance suivant échantillons indépendants puis appariés	239
4.5 Evaluation de l'efficacité d'un traitement par tests non paramétriques	241
4.6 Etude d'impact suivant le test de MAC NEMAR	246
4.7 Coefficient de contingence	247
4.8 Alternative au « $r$ » de PEARSON, le coefficient « $\tau$ » de KENDALL	249
4.9 Coefficient « $\rho$ » de SPEARMAN	252
5. Tests à plus de deux échantillons	254
5.1 Analyse de variance (test « ANOVA » de FISHER)	254
5.2 Test de KRUSKAL- WALLIS	257
5.3 Test de la médiane généralisée	259
5.4 Test de FRIEDMAN appliqué à un problème d'ergonomie	261
5.5 Comparaisons sur échantillons liés et données binaires (COCHRAN)	263
<b>C – Exercices complémentaires</b>	265
<b>Chapitre IV : Régression</b>	
<b>A – Rappels de cours</b>	
1. Régression linéaire simple	299
1.1 Le modèle	299
1.2 Estimation des paramètres	300
1.3 Erreur moyenne	300
1.4 Interprétation du coefficient de corrélation empirique	301
1.5 Coefficient de détermination et analyse de la variance	301
1.6 Propriétés des estimateurs des coefficients de la droite de régression	303
1.7 Intervalles de confiance et tests pour modèle linéaire gaussien	304
2. Régression linéaire multiple	305
2.1 Le modèle	305
2.2 Estimateurs des moindres carrés	305
2.3 Etude des coefficients et analyse de la variance	307
<b>B – Applications</b>	
1. Modèles à une variable explicative	308

<b>1.1 Autour de la droite de régression</b>	308
<b>1.2 Parabole des moindres carrés et distance de freinage</b>	312
<b>1.3 Equations non linéaires se ramenant au modèle linéaire (gaz parfait)</b>	314
<b>1.4 Modèle de régression (taille, poids)</b>	315
<b>2. Modèles à plusieurs variables explicatives</b>	316
<b>2.1 Illustration autour d'un modèle à deux variables explicatives</b>	316
<b>2.2 Matrices et régression linéaire multiple</b>	320
<b>C – Exercices complémentaires</b>	326
<b>Annexes</b>	
Table des valeurs de la loi normale centrée réduite	334
Table des valeurs de la loi de STUDENT	335
Table des valeurs de la loi du chi- deux de PEARSON	336
Tables de la loi de FISHER-SNEDECOR	337
Test de SHAPIRO et WILK	339
Test binomial	339
Test de WILCOXON, MANN, et WHITNEY	340
Test des rangs signés de WILCOXON	340
Test de KOLMOGOROV (pour un échantillon)	341
Test de FRIEDMAN	341
Test de KOLMOGOROV-SMIRNOV (pour deux échantillons)	342

<b>Bibliographie</b>	343
----------------------	-----

## Index

<b>Index alphabétique</b>	345
---------------------------	-----

Alphabétique	Discrète	Occurrence d'un caractère au cours de n épreuves de BERNOULLI indépendantes (à savoir le tirage sans remise d'un échantillon de taille n dans une population de taille N).	Values : {0, 1, 2, ..., n}	P(X=x) = $\frac{C_n^x p^x q^{n-x}}{n!}$
POISSON	Discrète	Occurrence des événements relativement rares	Values : N	P(x) = $e^{-\lambda} \frac{\lambda^x}{x!}$

\* Pour rappel, l'épreuve de BERNOULLI est une épreuve dans laquelle, seuls sont possibles, les résultats C (avec la probabilité  $p$ ) et  $\bar{C}$  (avec la probabilité complémentaire  $q = 1 - p$ ).

La collection TECHNOSUP dirigée par Claude Chèze est une sélection d'ouvrages dans toutes les disciplines, pour les filières technologiques des enseignements supérieurs.

- Niveau A    Approche (éléments, résumés ou travaux dirigés)  
Niveau B    Bases (cours avec exercices et problèmes résolus)  
Niveau C    Compléments (approfondissement, spécialisation)

- IUT - BTS - 1<sup>er</sup> cycle  
IUP - Licence  
Écoles d'ingénieurs, Master

### *L'ouvrage : niveau B (IUP - Licence)*

S'appuyant sur le calcul des probabilités, dont les techniques usuelles ont été développées dans un précédent ouvrage de l'auteur, ce manuel présente les principales méthodes utilisées en statistique mathématique, à travers des rappels de cours et plus d'une centaine de problèmes corrigés qui les illustrent au moyen de cas concrets.

L'ouvrage s'adresse à un large public qui est celui des écoles d'ingénieurs et des I.U.T, mais aussi des écoles de commerce et des universités dans des spécialités aussi diverses que l'ingénierie, la médecine, la biologie, l'agriculture, la gestion, l'économie... Il traite une grande variété de modèles et donne un riche aperçu des techniques statistiques autour des sujets classiques, échantillonnage, estimation, décision, et régression. Il trouve ainsi un juste compromis entre la rigueur mathématique et la pratique effective.

La présentation est particulièrement fournie quant à l'estimation et la décision statistique où les tests non paramétriques trouvent une place importante.

### *L'auteur :*

*Jean-Pierre Boulay, Ingénieur de la Ville de Paris, enseigne le calcul des probabilités et la statistique à l'École Spéciale des Travaux Publics à Paris, où il assure également un cours de recherche opérationnelle.*

Illustration de couverture : Dessin de Léonard de Vinci.



9 782729 856021

[www.editions-ellipses.fr](http://www.editions-ellipses.fr)