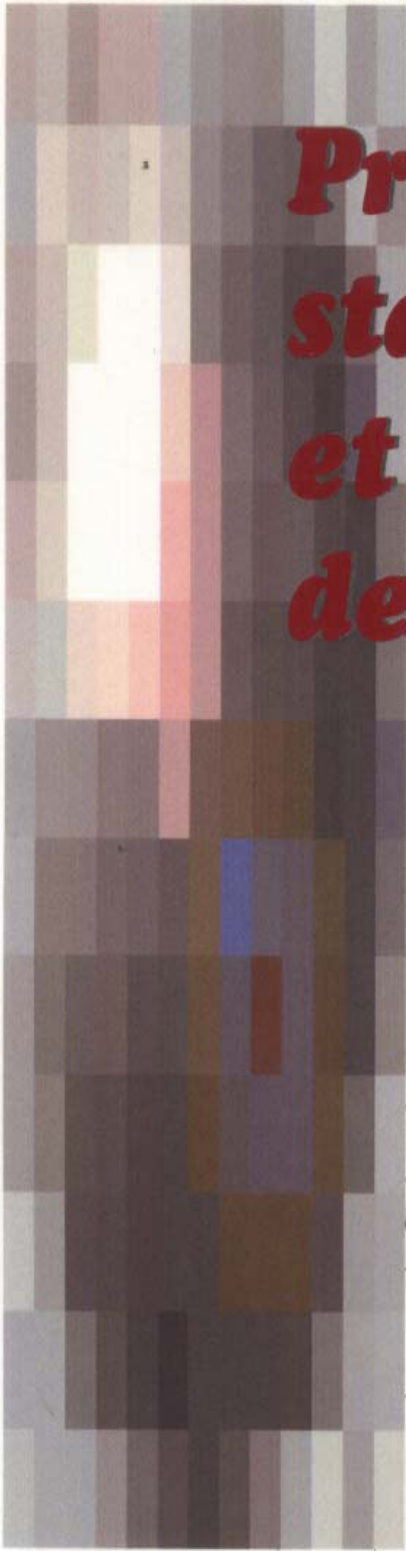
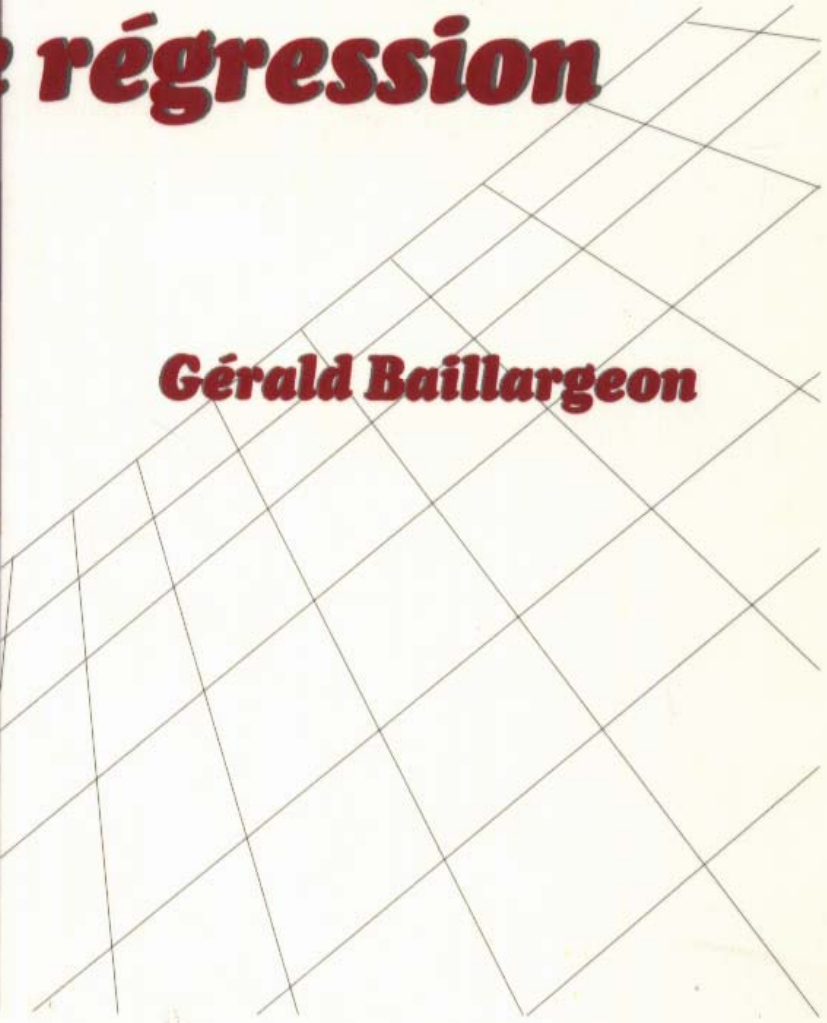




Probabilités, statistique et techniques de régression



Gérald Baillargeon

Les Éditions SMG



Outils d'aide à la décision

PROBABILITÉS, STATISTIQUE ET TECHNIQUES DE RÉGRESSION

**Gérald
Baillargeon**
Professeur associé
Université du Québec
à Trois-Rivières



LES ÉDITIONS SMG

TABLE DES MATIÈRES

CHAPITRE 1

STATISTIQUE DESCRIPTIVE

Sommaire	1
Objectifs pédagogiques	2
La statistique descriptive	3
Quelques notions fondamentales	3
Présentation des données recueillies	7
Dépouillement des observations et distribution de fréquences	7
Considérations pratiques dans l'élaboration d'une distribution de fréquences ...	8
Distribution de fréquences: cas où le caractère étudié est discret	11
Distribution de fréquences avec classes ouvertes	12
Principales représentations graphiques d'une distribution de fréquences	12
Courbe de fréquences cumulées	18
Problèmes	22
Caractéristiques des séries statistiques	28
La moyenne arithmétique, la variance, l'écart-type et le coefficient de variation..	28
Propriétés et remarques sur la moyenne arithmétique, la variance et l'écart-type	33
Calcul de la moyenne et de la variance: observations individuelles et observations groupées	35
Calcul simplifié de la moyenne et de la variance: changements d'origine et d'échelle	37
Autre mesure de tendance centrale: la médiane	41
Le mode	44
Autres mesures de position: les quantiles	45
Caractéristiques de forme: asymétrie et aplatissement	48
Problèmes	51
Problème de synthèse	57

CHAPITRE 2

CALCUL DES PROBABILITÉS

Sommaire	59
Objectifs pédagogiques	60
Introduction	61
Notions d'épreuve, de résultat, d'espace échantillonnal et d'événement	61

La notion de probabilité	64
Axiomes régissant le calcul des probabilités	66
Résumé sur le vocabulaire des événements	67
Probabilités totales: événements ne s'excluant pas	70
Événements liés, probabilités conditionnelles, probabilités composées et événements indépendants	73
Probabilités des causes: Formule de Bayes	79
L'analyse combinatoire, l'échantillonnage et les probabilités	84
Cheminement de réflexion pour résoudre les problèmes sur les probabilités	91
Problèmes	91

CHAPITRE 3

VARIABLES ALÉATOIRES ET LOIS DE PROBABILITÉ

Sommaire	97
Objectifs pédagogiques	98
Généralités sur les notions probabilistes et les notions statistiques	99
Notions de variable aléatoire et de loi de probabilité	100
Représentation graphique de la loi de probabilité (cas discret)	102
Loi de probabilité d'une variable continue	104
Espérance mathématique et variance d'une variable aléatoire	108
Propriétés de l'espérance mathématique et de la variance	113
Inégalité de Tchebycheff	115
Loi à deux dimensions	117
Propriétés de l'espérance mathématique et de la variance dans le cas d'une combinaison de plusieurs variables aléatoires	124
Problèmes	131

CHAPITRE 4

LES LOIS BINOMIALE, DE POISSON ET HYPERGÉOMÉTRIQUE

Sommaire	139
Objectifs pédagogiques	140
Lois de probabilité importantes	141
La loi binomiale et l'expérience de Bernoulli	141
Epreuves de Bernoulli et distribution binomiale	143
Synthèse de diverses relations utilisées dans le calcul des probabilités binomiales	151
Représentation graphique de la distribution binomiale	154

Problèmes	155
La loi de Poisson	158
Représentation graphique de la distribution de Poisson	160
Problèmes	162
Approximation de la loi binomiale par la loi de Poisson	164
Problèmes	167
Notions sur le processus de Poisson	168
Problèmes	171
La loi hypergéométrique	172
Problèmes	175

CHAPITRE 5

LA LOI NORMALE ET APPLICATIONS

Sommaire	179
Objectifs pédagogiques	180
La distribution normale	181
Propriétés de la loi normale	182
Loi normale centrée réduite	183
Résultats importants concernant l'aire sous la courbe normale	192
Combinaison de variables aléatoires normales	197
Approximation de la loi binomiale par la loi normale	200
Problèmes	203

CHAPITRE 6

ECHANTILLONNAGE ET ESTIMATION DE PARAMÈTRES

Sommaire	211
Objectifs pédagogiques	212
Introduction	213
Les sondages: Echantillonnage d'une population	213
L'échantillon aléatoire	214
Principe de la construction d'un échantillon	215
Construction de l'échantillon à l'aide d'une table de nombres aléatoires	215
Construction de l'échantillon par tirage systématique	218
Fluctuations d'échantillonnage d'une moyenne	220

Théorème central limite: Forme de la distribution de $\bar{X}$	224
Transformation de la variable $\bar{X}$ en une variable aléatoire centrée réduite	225
Propriétés de la distribution d'échantillonnage de $\bar{X}$	226
L'estimation de paramètres: Objectif fondamental de l'échantillonnage d'une population	228
L'estimation ponctuelle	229
Propriétés des estimateurs ponctuels	230
Estimation par intervalle de confiance	232
Estimation d'une moyenne par intervalle de confiance: Construction de l'intervalle	234
Marge d'erreur associée à l'estimation et taille d'échantillon requise pour ne pas excéder la marge d'erreur	238
La distribution de Student	240
Estimation de μ par intervalle: Population normale, variance inconnue et petit échantillon	243
Problèmes	246
Fluctuations d'échantillonnage d'une proportion	252
Estimation d'une proportion p par intervalle de confiance	254
Marge d'erreur associée à l'estimation de p et taille d'échantillon requise	256
Problèmes	259
Annexe mathématique sur les estimateurs, la méthode des moindres carrés et la méthode du maximum de vraisemblance	263

CHAPITRE 7

TESTS SUR UNE MOYENNE, UNE PROPORTION ET UNE VARIANCE

Sommaire	269
Objectifs pédagogiques	270
Principe d'un test d'hypothèse	271
Concepts importants dans l'élaboration d'un test d'hypothèse	271
Formulation des hypothèses H_0 et H_1 et type de test	273
Test sur une moyenne: détermination de la règle de décision	275
Tableau-synthèse des tests sur une moyenne	280
Comment exécuter un test d'hypothèse: démarche à suivre	283
Risques de première espèce et de deuxième espèce	288
Schématisation des deux risques d'erreur sur la distribution d'échantillon- nage de $\bar{X}$	289
Test sur une proportion	296
Tableau-synthèse - Test relatif à une proportion	297
Inférence sur une variance et la loi de khi-deux	300

Distribution d'échantillonnage de la variance d'échantillon	302
Test d'hypothèse et intervalle de confiance: variance d'une population normale	303
Problèmes	308

CHAPITRE 8

TESTS SUR DEUX MOYENNES, DEUX PROPORTIONS ET DEUX VARIANCES

Sommaire	317
Objectifs pédagogiques	318
Introduction	319
Distribution des fluctuations d'échantillonnage de la différence de deux moyennes	320
Test sur l'égalité de deux moyennes $H_0: \mu_1 = \mu_2$	326
Tableau-synthèse des tests de comparaison sur deux moyennes	327
Test de comparaison de deux échantillons appariés	330
Comparaison de deux proportions: $H_0: p_1 = p_2$	333
Comparaison de variances et la distribution de Fisher-Snedecor	337
Distribution d'échantillonnage du quotient de deux variances	339
Test sur l'égalité de deux variances	340
Problèmes	344

CHAPITRE 9

TEST D'AJUSTEMENT ET TABLEAU DE CONTINGENCE

Sommaire	353
Objectifs pédagogiques	354
Introduction	355
Application du test de Pearson: Test de conformité entre deux distributions	355
Test de conformité: Hypothèses statistiques et règle de décision	357
Vérification de la qualité d'ajustement pour les distributions binomiale, de Poisson et normale	360
Utilisation du khi-deux pour tester l'indépendance de deux caractères dans un tableau de contingence	369
Test d'indépendance de deux caractères: Hypothèses statistiques et règle de décision	375
Test d'homogénéité de plusieurs populations	379
Problèmes	386

CHAPITRE 10

CORRÉLATION LINÉAIRE SIMPLE

Sommaire	399
Objectifs pédagogiques	400
Introduction	401
Nuage de points: diagramme de dispersion	402
Calcul du coefficient de corrélation linéaire	404
Autres formules pour calculer le coefficient de corrélation	405
Entre quelles valeurs peut varier le coefficient de corrélation?	406
La corrélation est-elle significative?	408
Test de l'hypothèse $H_0: \rho = 0$	409
Test de l'hypothèse $H_0: \rho = \rho_0$ et estimation par intervalle de confiance	413
Problèmes	418
Annexe sur la loi normale à deux dimensions	424

CHAPITRE 11

RÉGRESSION LINÉAIRE SIMPLE

Sommaire	429
Objectifs pédagogiques	430
Introduction	431
Variable dépendante et variable explicative	433
Le modèle linéaire simple	434
Composantes du modèle linéaire simple	435
Conditions d'application du modèle linéaire simple	436
Schématisation des conditions d'application du modèle linéaire simple	437
Estimation de la droite de régression $E(Y) = \beta_0 + \beta_1 X$	439
Résidus et mesure de dispersion	446
Problèmes	449

CHAPITRE 12

INFÉRENCE EN RÉGRESSION LINÉAIRE SIMPLE

Sommaire	455
Objectifs pédagogiques	456
Introduction	457

Propriétés des estimateurs	458
Distribution d'échantillonnage de b_1	458
Estimation de β_1 par intervalle de confiance	460
Test statistique sur β_1 , la pente de la droite	461
Inférence concernant le paramètre β_0	467
Covariance de b_0 et b_1	468
Inférence concernant la moyenne de la distribution conditionnelle de Y à $X = X_h: E(Y_h)$	469
Prévision d'une valeur de la variable dépendante pour une nouvelle observation de X et intervalle de prévision	474
Quelques considérations pratiques dans l'application des méthodes d'analyse de régression	480
Problèmes	481
Annexe sur les propriétés des estimateurs b_0 et b_1	488

CHAPITRE 13

LA RÉGRESSION LINÉAIRE SIMPLE ET L'ANALYSE DE LA VARIANCE

Sommaire	491
Objectifs pédagogiques	492
Introduction	493
Décomposition de la variation dans les observations: calcul des sommes de carrés	493
Le coefficient de détermination simple: r^2	497
Autre façon de tester si la régression est significative: l'analyse de variance	499
Test sur $\beta_1 = 0$ versus $\beta_1 \neq 0$: Test de Fisher	502
Synthèse des résultats d'une analyse de régression linéaire simple	505
Problèmes	506

CHAPITRE 14

RÉGRESSION LINÉAIRE MULTIPLE

Sommaire	515
Objectifs pédagogiques	516
Introduction	517
Modèle linéaire à plusieurs variables explicatives	517
Hypothèses fondamentales	518
Démarche à suivre dans l'élaboration d'un modèle de régression multiple	521

Détermination de l'équation de régression: estimation des paramètres du modèle	521
Analyse de variance en régression multiple	523
Formules requises pour une analyse de régression multiple à deux variables explicatives	525
Tests d'hypothèses en régression multiple	533
Comment déterminer si la régression est significative dans son ensemble?	533
La régression multiple avec un programme informatique	536
Problèmes	536
Distribution d'échantillonnage des coefficients de régression	541
Détermination des variances des coefficients de régression	542
Test sur la contribution marginale de chaque variable explicative	542
Autre façon de tester la contribution marginale: le "F partiel"	545
Estimation de $E(Y_h)$ par intervalle de confiance	549
Intervalle de prévision pour Y_h	551
Problèmes	552
Utilisation de l'information qualitative dans un modèle de régression multiple: variables auxiliaires	559
Comment tester la nullité d'un sous-ensemble de paramètres β_j du modèle?	566
La corrélation partielle	568
Corrélation partielle et test de contribution marginale	573
Colinéarité	575
Autres modèles statistiques linéaires	577
Problèmes	577
BIBLIOGRAPHIE	585
TABLES STATISTIQUES	586
RÉPONSES AUX PROBLÈMES	616
INDEX	627

Probabilités, statistique et techniques de régression

Gérald Baillargeon

Le but premier de cet ouvrage consiste à développer des aptitudes à l'analyse statistique de problèmes concrets, à permettre la réflexion sur l'utilité pratique des concepts présentés et à susciter chez le lecteur un esprit de synthèse sur différents aspects du raisonnement statistique. L'exposé des différents sujets se veut systématique et intégré à diverses applications que l'on peut retrouver dans différents secteurs de l'entreprise tout en assurant une continuité et une synthèse des notions présentées.

Structure pédagogique de l'ouvrage

- a) La présentation structurée de diverses méthodes statistiques avec une approche volontairement pratique, intuitive et très visuelle facilite l'apprentissage et la compréhension.
- b) La diversité et les contextes des applications en font un ouvrage utilitaire à caractère professionnel.
- c) Tous les concepts importants sont définis avec clarté et précision et sont cadrés pour en dégager l'importance, susciter la réflexion et faciliter la révision.
- d) Les nombreuses remarques font ressortir, soit certains points importants, soit certains points complémentaires, soit une réflexion sur certaines notions qui seront présentées ultérieurement.
- e) Tous les exemples qui sont présentés sont titrés pour donner une idée immédiate du concept que nous voulons traiter.
- f) Le sommaire de chaque chapitre de l'ouvrage ainsi que les objectifs pédagogiques qui y sont précisés donnent une vue d'ensemble et très détaillée des notions à acquérir.
- g) Chaque chapitre se termine par de nombreux exercices d'applications diverses.
- h) Pour assurer une certaine efficacité et une évaluation du travail effectué, le lecteur trouvera, à la fin de l'ouvrage, les réponses aux exercices.
- i) On trouvera également, soit sous formes d'exemples, soit en annexe, certaines justifications mathématiques de résultats importants.

Sommaire de l'ouvrage

- 1. **Statistique descriptive**
- 2. **Calcul des probabilités**
- 3. **Variables aléatoires et lois de probabilité**
- 4. **Les lois binomiale, de Poisson et hypergéométrique**
- 5. **La loi normale et applications**
- 6. **Échantillonnage et estimation de paramètres**
- 7. **Tests sur une moyenne, une proportion et une variance**
- 8. **Tests sur deux moyennes, deux proportions et deux variances**
- 9. **Test d'ajustement et tableau de contingence**
- 10. **Corrélation linéaire simple**
- 11. **Régression linéaire simple**
- 12. **Inférence en régression linéaire simple**
- 13. **La régression linéaire simple et l'analyse de la variance**
- 14. **Régression linéaire multiple**
- Tables statistiques - Réponses aux exercices - Index**

ISBN 2-89094-038-1

Les Éditions SMG

5365 boul Jean XXIII, suite 203,

Trois-Rivières ouest, G8Z 4A6

Tel.: (819) 376-5650

Télécopie: (819) 373-2904

Distributeur exclusif pour tous les pays (sauf le Canada) :

Technique et Documentation Lavoisier

14 rue de Provigny

F-94236 Cachan cedex France