

## **Table of Contents • Table des matières**

Table of Contents • Table des matières	2
1998 Annual Meeting • Congrès 1998	3
Acknowledgements • Remerciements	4
Committee Meetings	5
Réunions des comités	6
List of Sessions • Liste des sessions	7
General Information • Information générale	10
Program • Programme	11
Abstracts • Résumés	32
Author Index • Index des auteurs	92

# 1998 Annual Meeting • Congrès 1998

## Organizing Committees • Comités organisateurs

### Program Committee • Comité du programme

**Chair • Président** Louis-Paul RIVEST, Université Laval

**SSC Survey Methods Section • Section de méthodologie d'enquête de la SSC** Pierre LAVALLÉE, Statistics Canada / *Statistique Canada*

**SSC Biostatistics Section • Section de biostatistique de la SSC** Richard J. COOK, University of Waterloo

**Colloque "Méthodologie et applications de la statistique"** Paul BERTHIAUME, Bureau de la statistique du Québec et Stéphane TREMBLAY, Statistics Canada / *Statistique Canada*

**Association des statisticiens et des statisticiennes du Québec (ASSQ)** Mario MONTÉGIANI, Société d'habitation du Québec

### Local Organizing Committee • Comité organisateur local

**Chair • Président** Ernest MONGA, Université de Sherbrooke

**Member • Membre** Bernard COLIN, Université de Sherbrooke

# Acknowledgements • Remerciements

The Statistical Society of Canada gratefully acknowledges the help and support of the following persons and organizations • La Société statistique du Canada tient à manifester sa reconnaissance aux personnes et organisations suivantes pour leur aide et leur soutien:

## Sponsors • Commanditaires

- Université de Sherbrooke
- Centre de recherches mathématiques de l'Université de Montréal
- The Fields Institute for Research in Mathematical Sciences
- Faculté des sciences, Université de Sherbrooke
- Département de mathématiques et d'informatique, Université de Sherbrooke
- Astra Pharma Inc.
- Bristol-Myers Squibb
- Glaxo Wellcome Inc.
- Pharmacia and Upjohn
- Purdue Frederick

## Exhibitors • Exposants

- John Wiley and Sons Canada Ltd.

For his work on translations and on the preparation of the program booklet, we are indebted to • Pour son travail de traduction et pour la préparation du livret du programme, nous sommes reconnaissant à

- Abderrahmane CHAKAK (Université Abdelmalek Essaadi)

For their construction of the WEB version of the program • Pour la réalisation de la version WEB du programme

- Duncan MURDOCH, Queen's University
- Charles PATRICK, Statistics Canada / *Statistique Canada*

For constructing the L<sup>A</sup>T<sub>E</sub>X input files for the program and abstracts • Pour avoir construit les fichiers L<sup>A</sup>T<sub>E</sub>X qui ont servi au programme et aux résumés

- André MONTPETIT

For secretarial assistance, thanks to • Pour son aide à la dactylographie, nous remercions

- Suzanne TALBOT, Université Laval

For their contributions to the local arrangements we thank • Pour leur contribution à l'organisation locale nous remercions

- Service de la sécurité et du stationnement, Université de Sherbrooke
- Faculté d'éducation physique et sportive, Université de Sherbrooke
- Service des résidences, Université de Sherbrooke

# Committee Meetings

Annual General Meeting (SSC)	June 2	17:15	Pavillon Marie-Victorin 422
Annual General Meeting (ASSQ)	June 2	17:15	Pavillon Marie-Victorin 407
Bilingualism Committee	June 2	12:30	Mezzanine, Pavillon Multifonctionnel
Biostatistics Annual General Meeting	June 1	17:15	Pavillon Marie-Victorin 422
Biostatistics Executive	June 1	7:00	Mezzanine, Pavillon Multifonctionnel
Biostatistics Executive	June 2	12:30	Mezzanine, Pavillon Multifonctionnel
Board of Directors	May 31	12:00	Mezzanine, Pavillon Multifonctionnel
Board of Directors	June 4	8:00	Mezzanine, Pavillon Multifonctionnel
Caucus for Women in Statistics	May 31	18:30	Mezzanine, Pavillon Multifonctionnel
Executive Committee	May 30	19:00	Delta Hotel
Executive Committee	June 4	12:00	Mezzanine, Pavillon Multifonctionnel
Finance Committee	May 31	9:00	Pavillon Marie-Victorin 417
Finance Committee	June 2	7:30	Mezzanine, Pavillon Multifonctionnel
Liaison Editorial Board	June 3	12:30	Mezzanine, Pavillon Multifonctionnel
NSERC Luncheon Meeting	June 1	12:30	Pavillon Marie-Victorin 407
Program Committee	June 1	7:30	Mezzanine, Pavillon Multifonctionnel
Public Relations Committee	May 31	9:00	Pavillon Marie-Victorin 418
Public Relations Committee	June 1	12:30	Mezzanine, Pavillon Multifonctionnel
Publications Committee	May 31	10:30	Pavillon Marie-Victorin 418
Publications Committee	June 2	12:30	Mezzanine, Pavillon Multifonctionnel
Regional and Society Cooperation	June 3	12:30	Mezzanine, Pavillon Multifonctionnel
Research Committee	June 2	7:30	Mezzanine, Pavillon Multifonctionnel
Southern Ontario Regional Association	June 1	12:30	Pavillon Marie-Victorin 409
Statistical Education	June 2	12:30	Mezzanine, Pavillon Multifonctionnel
Statistical Consulting	June 3	7:30	Mezzanine, Pavillon Multifonctionnel
Survey Methods Annual General Meeting	June 1	17:15	Pavillon Marie-Victorin 407
Survey Methods Executive	June 1	12:30	Mezzanine, Pavillon Multifonctionnel

## Réunions des comités

Assemblée générale annuelle (SSC)	2 juin	17 :15	Pavillon Marie-Victorin 422
Assemblée générale annuelle (ASSQ)	2 juin	17 :15	Pavillon Marie-Victorin 407
Assemblée générale annuelle de la section de biostatistique	1 juin	17 :15	Pavillon Marie-Victorin 422
Assemblée générale annuelle de la section de méthodologie d'enquête	1 juin	17 :15	Pavillon Marie-Victorin 407
Association régionale du sud de l'Ontario	1 juin	12 :30	Pavillon Marie-Victorin 409
Bilinguisme	2 juin	12 :30	Mezzanine, Pavillon Multifonctionnel
Comité éditorial de Liaison	3 juin	12 :30	Mezzanine, Pavillon Multifonctionnel
Coopération nationale-régionale au sein de la Société	3 juin	12 :30	Mezzanine, Pavillon Multifonctionnel
Conseil d'administration	31 mai	12 :00	Mezzanine, Pavillon Multifonctionnel
Conseil d'administration	4 juin	8 :00	Mezzanine, Pavillon Multifonctionnel
Consultation statistique	3 juin	7 :30	Mezzanine, Pavillon Multifonctionnel
Dîner-rencontre sur le CRSNG	1 juin	12 :30	Pavillon Marie-Victorin 407
Exécutif	30 mai	19 :00	Hotel Delta
Exécutif	4 juin	12 :00	Mezzanine, Pavillon Multifonctionnel
Exécutif de la section de biostatistique	1 juin	7 :00	Mezzanine, Pavillon Multifonctionnel
Exécutif de la section de biostatistique	2 juin	12 :30	Mezzanine, Pavillon Multifonctionnel
Exécutif de section de méthodologie d'enquête	1 juin	12 :30	Mezzanine, Pavillon Multifonctionnel
Finances	31 mai	9 :00	Pavillon Marie-Victorin 417
Finances	2 juin	7 :30	Mezzanine, Pavillon Multifonctionnel
Formation statistique	2 juin	12 :30	Mezzanine, Pavillon Multifonctionnel
Programme	1 juin	7 :30	Mezzanine, Pavillon Multifonctionnel
Promotion de la femme en statistique	31 mai	18 :30	Mezzanine, Pavillon Multifonctionnel
Publications	31 mai	10 :30	Pavillon Marie-Victorin 418
Publications	2 juin	12 :30	Mezzanine, Pavillon Multifonctionnel
Recherche	2 juin	7 :30	Mezzanine, Pavillon Multifonctionnel
Relations publiques	31 mai	9 :00	Pavillon Marie-Victorin 418
Relations publiques	1 juin	12 :30	Mezzanine, Pavillon Multifonctionnel

## List of Sessions • Liste des sessions

Session 1	1997 SSC Gold Medal Address <i>Allocution du récipiendaire 1997 de la médaille de la SSC</i>	June 1 juin	8:45–10:15	D2 1060
Session 2	Special Biostatistics Session of the Fields Institute : Event History Analysis <i>Session spéciale de l'Institut Fields en biostatistique : Analyse de séries d'évènements</i>	June 1 juin	10:45–12:15	PMV 422
Session 3	Effects of Effluents on White Suckers in the Kapuskasing River, Ontario <i>Effets des effluents de la rivière Kapuskasing en Ontario sur le poisson blanc</i>	June 1 juin	10:45–12:15	PMV 409
Session 4	Probability : Stochastic processes <i>Probabilité : Processus stochastiques</i>	June 1 juin	10:45–12:15	PMV 417
Session 5	Survey Methods : Health Surveys <i>Méthodologie d'enquête : Enquêtes de santé</i>	June 1 juin	10:45–12:15	PMV 407
Session 6	Smoothing, Quality Control and Experimental Design <i>Lissage, contrôle de qualité et plans d'expériences</i>	June 1 juin	10:45–12:15	PMV 418
Session 7	Biostatistics : Statistical Genetics, New ideas, New challenges <i>Biostatistique : La statistique génétique, nouvelles idées et nouveaux défis</i>	June 1 juin	13:30–15:00	PMV 422
Session 8	Biostatistics : Risk Analysis and Related Topics <i>Biostatistique : Analyse de risques et sujets connexes</i>	June 1 juin	13:30–15:00	PMV 418
Session 9	Probability : Limit Theorems for Multiparameter and Set-Indexed Processes <i>Probabilité : Théorèmes limites pour des processus multiparamètres indicés par des ensembles</i>	June 1 juin	13:30–15:00	PMV 417
Session 10	Survey Methods : Administrative Data <i>Méthodologie d'enquête : Utilisation de données administratives</i>	June 1 juin	13:30–15:00	PMV 407
Session 11	Extreme Values and Time Series <i>Valeurs extrêmes et séries chronologiques</i>	June 1 juin	13:30–15:00	PMV 409
Session 12	Biostatistics : Cancer Research <i>Biostatistique : Recherche sur le cancer</i>	June 1 juin	15:30–17:00	PMV 422
Session 13	Multidimensional Statistics <i>Statistique multidimensionnelle</i>	June 1 juin	15:30–17:00	PMV 418

PMV = Pavillon Marie-Victorin au 3<sup>ième</sup> étage (3rd floor); D2 = Pavillon D2 Building

Session14	Statistical Education: Panel Discussion on Probability and Statistics in Canadian Schools <i>Education statistique : Atelier sur l'enseignement de la probabilité et de la statistique dans les écoles canadiennes</i>	June 1 juin	15:30–17:00	PMV 409
Session15	Survey Methods : The Project to Improve Provincial Economic Statistics (PIPES) <i>Méthodologie d'enquête : Le projet d'amélioration des statistiques économiques provinciales (PASEP)</i>	June 1 juin	15:30–17:00	PMV 407
Session16	SSC Presidential Invited Address <i>Allocution de l'invité du président</i>	June 2 juin	8:45–10:15	D2 1060
Session17	Biostatistics : Issues in the Pharmaceutical Industry, Multiple Endpoints in Clinical Trials <i>Biostatistique: Débat dans l'industrie pharmaceutique, les essais cliniques à terminaison variable</i>	June 2 juin	10:45–12:15	PMV 407
Session18	Industrial Statistics <i>Statistique industrielle</i>	June 2 juin	10:45–12:15	PMV 409
Session19	Biostatistics : Linear Models <i>Biostatistique : Modèles linéaires</i>	June 2 juin	10:45–12:15	PMV 417
Session20	Special Address on Survey Methods <i>Conférence spéciale sur la méthodologie d'enquête</i>	June 2 juin	10:45–12:15	PMV 422
Session21	Panel Session <i>Séance par affichage</i>	June 2 juin	12:30–13:30	PMV 419
Session22	Biostatistics : Applications <i>Biostatistique : Applications</i>	June 2 juin	13:30–15:00	PMV 422
Session23	Special Session of the Centre de Recherches Mathématiques : Analysis of shape <i>Session spéciale du Centre de Recherches Mathématiques : Analyse de la forme</i>	June 2 juin	13:30–17:00	PMV 409
Session24	ASSQ Session : Quality Control in Industry <i>Session organisée par l'ASSQ : Contrôle de la qualité en milieu industriel</i>	June 2 juin	13:30–15:00	PMV 418
Session25	Survey Methods : Modeling survey data <i>Méthodologie d'enquête : Modélisation des données d'enquête</i>	June 2 juin	13:30–15:00	PMV 407
Session26	Biostatistics : Mixed topics <i>Biostatistique : Sujets variés</i>	June 2 juin	13:30–15:00	PMV 417
Session27	Biostatistics : Estimation <i>Biostatistique : Estimation</i>	June 2 juin	15:30–17:00	PMV 417
Session28	ASSQ Session : Panel Discussion on Statistical Consulting <i>Session organisée par l'ASSQ : Atelier sur la consultation statistique</i>	June 2 juin	15:30–17:00	PMV 418
Session29	Survey Methods: Variance Estimation, Calibration and Related Topics <i>Méthodologie d'enquête : estimation de la variance, calage et sujets connexes</i>	June 2 juin	15:30–17:00	PMV 407

Session30	Survey Methods and Biostatistics : New Approaches to Data Analysis <i>Méthodologie d'enquête et biostatistique : Nouvelles approches pour l'analyse des données</i>	June 2 juin	15:30–17:00	PMV 422
Session31	Pierre Robillard Award Winner Lecture <i>Allocution du (de la) lauréat(e) du prix Pierre-Robillard</i>	June 3 juin	8:45–9:30	D2 1060
Session32	Canadian Journal of Statistics Award Winner Lecture <i>Allocution du (de la) lauréat(e) du prix de la Revue canadienne de statistique</i>	June 3 juin	9:30–10:15	D2 1060
Session33	Topics in Environmental Statistics <i>Statistique de l'environnement</i>	June 3 juin	10:45–12:15	PMV 418
Session34	Habitat Usage Versus Availability for the Cutthroat Trout in Dutch Creek, Alberta <i>Disponibilité et usage de l'habitat pour la truite "coupe-gorge" dans le Dutch Creek en Alberta</i>	June 3 juin	10:45–12:15	PMV 409
Session35	Theory and applications of classification <i>Théorie et applications en classification</i>	June 3 juin	10:45–12:15	PMV 422
Session36	Probability and Mathematical Statistics <i>Probabilité et statistique mathématique</i>	June 3 juin	10:45–12:15	PMV 417
Session37	Survey Methods : Survey Design <i>Méthodologie d'enquête : Conception d'enquêtes</i>	June 3 juin	10:45–12:15	PMV 407
Session38	Biostatistics : Topics in Environmental Statistics <i>Biostatistique : Statistique de l'environnement</i>	June 3 juin	13:30–15:00	PMV 422
Session39	Mathematical Statistics <i>Statistique mathématique</i>	June 3 juin	13:30–15:00	PMV 417
Session40	Survey Methods : Outliers in Survey Data <i>Méthodologie d'enquête : traitement des valeurs aberrantes</i>	June 3 juin	13:30–15:00	PMV 407
Session41	Committee on Women in Statistics & Caucus for Women in Statistics : Classification <i>Comité des femmes en statistique &amp; Caucus pour les femmes en statistique : Classification</i>	June 3 juin	13:30–15:00	PMV 418

---

PMV = Pavillon Marie-Victorin au 3<sup>ième</sup> étage (3rd floor); D2 = Pavillon D2 Building



## **General Information • Information générale**

### **Conference Halls • Salles des conférences**

Most sessions take place on the fourth floor of the Pavillon Marie-Victorin (PMV). The plenary sessions are in room 1060 of the D2 building (D2) • La plupart des sessions se déroulent au quatrième étage du Pavillon Marie-Victorin (PMV). Les sessions plénières ont lieu à la salle 1060 de l'immeuble D2 (D2).

### **Eating Lunch • Repas du midi**

Lunch is served at the cafeteria of the Pavillon Multifonctionnel (PMF); committee meetings are held on the mezzanine of the Pavillon Multifonctionnel (M PMF). Sandwiches are also available at a food outlet, on the second floor of the Pavillon Marie-Victorin • La cafétéria du Pavillon Multifonctionnel (PMF) sert des repas le midi; des réunions de comité se tiennent à la mezzanine du Pavillon Multifonctionnel (M PMF). On peut également acheter des sandwichs au Pavillon Marie-Victorin, à un comptoir situé au deuxième étage

### **Speakers Identification • Identification des conférenciers**

In the program booklet, speakers are identified by an asterisk, \* • Une astérisque, \*, identifie les conférenciers dans le livret du programme.

# Program • Programme

**Sunday, May 31st • Dimanche 31 mai**

---

- 8:00–17:00 Registration • Inscription** **PVM 419**
- 9:00–17:00** **PVM 422**  
 Biostatistics Workshop • Atelier de biostatistique
- 9:00–17:00** **PVM 407**  
 Sample Surveys Workshop • Atelier de méthodologie d'enquête
- 17:00–23:00 Registration • Inscription** **PVM 419**
- 17:30–18:30 Special Conference • Conférence spéciale** **M PMF**  
 Committee on Women in Statistics & The Caucus for Women in Statistics: Sunday Evening Reception • Comité des femmes en statistique & Caucus pour les femmes en statistique : Réception du dimanche soir  
 Organizer and Chair • Responsable et présidente: Charmaine DEAN  
 Gender and Work: For Richer or Poorer • Le travail des hommes et des femmes : qui s'enrichit ? [MS-1]  
 Nancy GALAM, Statistics Canada
- 19:30–22:00** **PMF**  
 Reception offered by the University of Sherbrooke • Réception offerte par l'Université de Sherbrooke

## Monday, June 1st • Lundi 1 juin

---

**7:30 Registration • Inscription**

**PMV 419**

**8:30–8:45**

**D2 1060**

Welcoming Announcements and Award Presentations • Mots de bienvenue et présentation des lauréats

**8:45–10:15 Session 1**

**D2 1060**

1997 SSC Gold Medal Address • Allocution du récipiendaire 1997 de la médaille d'or de la SSC

Special Lecture • Conférence principale

Chair • Président: Doug WIENS, University of Alberta

Statistical Services — Preparing for the Future • Les services statistiques — Comment se préparer pour l'avenir [MS-2]

Ivan P. FELLEGI\*, Chief Statistician of Canada • Statisticien en chef du Canada

**10:15–10:45 Coffee Break • Pause-café**

**PMV 411**

**10:45–12:15 Session 2**

**PMV 422**

Special Biostatistics Session of the Fields Institute : Event History Analysis • Session spéciale de l'Institut Fields en biostatistique : Analyse de séries d'évènements

Invited Paper Session • Conférences sur invitation

Organizer and Chair • Responsable et président: David E. MATTHEWS, University of Waterloo

10:45 A Frailty Model for the Incidence of Testicular Cancer • Un modèle de "frailty" pour l'incidence du cancer des testicules [MS-3]

Odd. A. AALEN\*, University of Oslo

11:15 Goodness-of-Fit Tests for the Multiplicative Intensity Model • Tests d'adéquation pour des modèles à intensité multiplicative [MS-4]

David HARRINGTON\*, Dana-Faber Cancer Institute

Cheryl JONES, Harvard School of Public Health

11:45 Analysis of Event History Data Through Rate Functions Modeling • Analyse des données historiques d'évènements par la modélisation des fonctions de taux [MS-5]

Claude NADEAU\*, Université Laval

12:15 Floor discussion • Discussion plénière

**10:45–12:15 Session 3**

**PMV 409**

Case Studies • Études de cas

Effects of Effluents on White Suckers in the Kapuskasing River, Ontario • Effets des effluents de la rivière Kapuskasing en Ontario sur le poisson blanc

Organizer • Responsable: Jeanette O'HARA-HINES, University of Waterloo

Chair • Présidente: Julie ZHOU, Lakehead University

10:45 Linear and Non-Linear Analysis of Kapuskasing River Data with Longitudinal Elements • Analyse linéaire et non-linéaire des données de la rivière Kapuskasing avec une composante longitudinale

Rafal KUSTRA, University of Toronto

Carmen MAK, University of Toronto

- 11:10 Growth Curve Analysis of Kapuskasing River Fish Data • Analyse des données de la rivière Kapuskasing avec des courbes de croissance  
Ruxandra SPIJAVCA, University of Toronto  
Juan Pablo LEWIGER, University of Toronto
- 11:35 Robust Estimation on Fish Growth Model • Estimation robuste dans un modèle pour la croissance des poissons  
J. LEE, Lakehead University  
C. LIU, Lakehead University  
J. ZHOU, Lakehead University
- 11:55 Visualization of the Effects of Effluents • Visualisation de l'effet des effluents  
Kajitani YOSHIO, University of Western Ontario  
A. Ian MCLEOD, University of Western Ontario

**10:45–12:15 Session 4****PMV 417****Probability : Stochastic Processes • Probabilité : Processus stochastiques**

Invited Paper Session • Conférences sur invitation

Organizer and Chair • Responsable et président: Jean VAILLANCOURT, Université de Sherbrooke

- 10:45 Empirical Processes Based on Pseudo-Observations • Processus empiriques basés sur des pseudo-observations [MS-6]  
Bruno RÉMILLARD\*, Université du Québec à Trois-Rivières
- 11:15 A Probability Model for the Continuity Laws • Un modèle probabiliste pour les lois de continuité [MS-7]  
René FERLAND\*, Université du Québec à Montréal  
Brigitte CHAUVIN, Université du Québec à Montréal  
Gaston GIROUX, Université du Québec à Montréal
- 11:45 Super-Tree Random Measures • Construction de mesures aléatoires par super branchement [MS-8]  
Hassan ALLOUBA\*, Duke University
- 12:15 Floor discussion • Discussion plénière

**10:45–12:15 Session 5****PMV 407****Survey Methods : Health Surveys • Méthodologie d'enquête : Enquêtes de santé**

Invited Paper Session • Conférences sur invitation

Organizers • Responsables: Louise BOURQUE &amp; Robert COURTEMANCHE, Bureau de la Statistique du Québec

Chair • Président: Robert COURTEMANCHE, Bureau de la Statistique du Québec

- 10:45 Global Overview of the Data Quality of the National Population Health Survey • Un aperçu global de la qualité des données de l'Enquête nationale sur la santé de la population [MS-9]  
Yves BÉLAND\*, Statistique Canada  
Janine BOUTROS\*, Statistique Canada
- 11:30 Quality of Data from Direct and by Telephone Interviews in a Survey on Health and Social Security • Qualité des données recueillies par entretiens téléphoniques et en face à face dans une enquête sur la santé et la protection sociale [MS-10]  
Arié MIZRAHI\*, ARGSES, France
- 12:15 Floor discussion • Discussion plénière

**10:45–12:15 Session 6****PMV 418****Smoothing, Quality Control and Designs of Experiment • Lissage, contrôle de qualité et schémas expérimentaux**

Contributed Paper Session • Communications libres

Chair • Présidente: Nancy HECKMAN, University of British Columbia

- 10:45 Adventures in Functional Linear Modeling • Aventures dans la modélisation linéaire fonctionnelle [MS-11]  
J. O. RAMSAY\*, McGill University
- 11:00 Bump Hunting in Regression Revisited • Recherche de bosses en régression [MS-12]  
Jarek HAREZLAK\*, University of British Columbia
- 11:15 Monitoring Process Mean and Variability With One EWMA Chart • Contrôle de la moyenne et de la variabilité d'un processus à l'aide d'un seul diagramme MMEP [MS-13]  
Gemai CHEN\*, University of Regina  
Smiley W. CHENG, University of Manitoba  
Hansheng XIE, University of Manitoba
- 11:30 Defect Rate Estimation Using Imperfect Zero-Defect Sampling with Rectification • Estimation des taux d'articles défectueux par utilisation de l'échantillonnage zéro-un avec rectification [MS-14]  
Pavan WADHWA\*, Mitchell Madison Group
- 11:45 Friedman Type Tests for Cyclic and Other Incomplete Designs • Tests de type Friedman pour les plans cycliques et pour d'autres plans incomplets [MS-15]  
Paul CABILIO\*, Acadia University  
Mayer ALVO, University of Ottawa
- 12:00 Experimental Designs for Orientation Models • Plans d'expérience pour les modèles d'orientation [MS-16]  
Hwa Shin H. SHIN\*, Queen's University
- 12:15 Floor discussion • Discussion plénière

**12:30–13:30 Lunch • Dîner****PMF****12:30–13:30****PMV 407**

NSERC Luncheon Meeting : Question and answer Period (Open to all) • Dîner - Rencontre CRSNG : Questions et réponses (ouvert à tous)

**13:30–15:00 Session 7****PMV 422**

Biostatistics : Statistical Genetics New ideas, New challenges • Biostatistique : La statistique génétique, nouvelles idées et nouveaux défis

Invited Paper Session • Conférences sur invitation

Organizer and Chair • Responsable et président: Neil KLAR, Dana-Farber Cancer Institute and Harvard School of Public Health

- 13:30 Mapping Minor Histocompatibility Genes Involved in Graft-versus-Host Disease: A Novel Application of Discordant Sibpairs Methodology • Images des gènes presque compatibles intervenant dans la maladie greffe-contre-hôte : une nouvelle application de la méthodologie des paires discordantes [MS-17]  
Kathy LUNETTA\*, Dana-Farber Cancer Institute and Harvard School of Public Health
- 14:00 Triangle Constraints in Triangle Affected Sib Pairs Models with Covariates • Contraintes triangulaires dans le modèle de fratries affectées avec covariables [MS-18]  
Celia GREENWOOD\*, University of Toronto
- 14:30 How Genetics Challenges the Ethics of Research • Comment la génétique défie l'éthique de la recherche [MS-19]  
Charles WEIJER\*, Mount Sinai Hospital, Toronto
- 15:00 Floor discussion • Discussion plénière

**13:30–15:00 Session 8****PMV 418****Biostatistics : Risk Analysis and Related Topics • Biostatistique : Analyse de risques et sujets connexes**

Contributed Paper Session • Communications libres

Chair • Président: Cynthia STRUTHERS, University of Waterloo

- 13:30 The Statistics Function in a Canadian Health Protection Agency • Le rôle de la statistique au sein d'un organisme canadien voué à la protection de la santé [MS-20]  
 S. BARTLETT\*, Health Canada  
 D. MACLEOD, Health Canada  
 B. MORI, Health Canada  
 E. ORMSBY, Health Canada  
 L. PAPATHANASOPOULOU, Health Canada  
 M. RAHMAN, Health Canada  
 W.H. ROSS, Health Canada
- 13:45 Determination of Toxicity Health States for Interferon Maintenance Treatment in Multiple Myeloma and 4-Health-State Q-TWiST Evaluation of Interferon Effectiveness • Détermination des niveaux d'états de santé dus à la toxicité pour le traitement par interféron dans l'évaluation de l'efficacité de l'interféron pour des mélanomes multiples selon 4 états de santé et par le TAQSST [MS-21]  
 Tong LI\*, Queen's University  
 Benny ZEE, Queen's University  
 Keith JAMES, Queen's University  
 Michael BRUNDAGE, Queen's University
- 14:00 The Statistical Modelling of Micronutrient Interactions in Nutrient Risk Assessments: The Effect of Calcium on Iron • L'établissement de modèles statistiques des interactions des micronutriments dans l'évaluation du risque associé aux nutriments : l'effet du calcium sur le fer [MS-22]  
 W.H. ROSS\*, Health Canada  
 K. COCKELL, Health Canada  
 S. DUBOIS, Health Canada  
 M. L'ABBE, Health Canada  
 B. JUNKINS, Health Canada  
 D. MILLER, Health Canada
- 14:15 Statistical Issues in the Analysis of Parallel Time Series of Health and Environmental Variables • Débat sur l'analyse statistique des séries chronologiques des variables décrivant des problèmes de santé et des variables environnementales [MS-23]  
 Sabit ÇAKMAK\*, Health Canada  
 R. BURNETT, Health Canada  
 D. KREWSKI, Health Canada
- 14:30 A Quantal Isobologram Model to Identify Joint Action for Chemical Mixtures • Un modèle quantitatif isobologramme pour identifier l'action conjointe de mélanges chimiques [MS-24]  
 Ding G. CHEN\*, Department of Fisheries and Oceans
- 14:45 Stochastic Modelling of HIV/AIDS Patients Migration: Application to Atlantic Canada HIV/AIDS Migration Pilot Study • Modélisation stochastique de la migration d'individus atteints du HIV/SIDA : Application à une étude pilote dans les provinces atlantiques concernant la migration des individus atteints du HIV/SIDA [MS-25]  
 Swarna WEERASINGHE\*, Dalhousie University  
 Lamont SWEET, Dalhousie University  
 Diane MACDONALD, Dalhousie University  
 Andrea SCOTT, Dalhousie University  
 Jerrold YATMAN, Dalhousie University  
 Lynn JOHNSTON, Dalhousie University
- 15:00 Floor discussion • Discussion plénière

**13:30–15:00 Session 9****PMV 417****Probability : Limit Theorems for Multiparameter and Set-Indexed Processes • Probabilité : Théorèmes limites pour des processus multiparamètres indicés par des ensembles**

Invited Paper Session • Conférences sur invitation

Organizer • Responsable: André DABROWSKY, University of Ottawa

Chair • Président: Bruno RÉMILLARD, Université de Québec à Trois-Rivières

13:30 Functional LIL for a Multiparameter Martingale • La loi du logarithme itéré pour une martingale à plusieurs paramètres [MS-26]

André DABROWSKI\*, University of Ottawa

14:00 Poisson Convergence for a Set-Indexed Empirical Process • La convergence de Poisson pour un processus empirique indicé par des ensembles [MS-27]

Gail IVANOFF\*, University of Ottawa

14:30 A Central Limit Theorem for Set-Indexed Martingales • Un Théorème de la Limite Centrale pour les martingales indicées par des ensembles [MS-28]

Dean SLONOWSKY\*, University of Ottawa

15:00 Floor discussion • Discussion plénière

**13:30–15:00 Session 10****PMV 407****Survey Methods : Administrative Data • Méthodologie d'enquête : Utilisation de données administratives**

Invited Paper Session • Conférences sur invitation

Organizer and Chair • Responsable et président: Michel HIDIROGLOU, Statistique Canada

13:30 Use of Administrative Records in the Canadian Survey of Employment, Payrolls and Hours • L'utilisation des rapports administratifs dans l'enquête sur l'emploi canadien, la rémunération et les heures [MS-29]

Éric RANCOURT\*, Statistique Canada

Mike HIDIROGLOU, Statistique Canada

14:00 Administrative Data and Non-Observed Heterogeneity: A Permanent Union for an Efficient Longitudinal Study • Données administratives et hétérogénéité non observée : Un mariage sans divorce pour une analyse longitudinale efficace [MS-30]

Abdellatif DEMNATI\*, Statistique Canada

J.-F. BEAUMONT, Statistique Canada

14:30 Impacts of the Use of a Mixture Two Way Response on Transversal and Longitudinal Estimation of Income • Impacts de l'utilisation d'un mode de collecte mixte sur les estimations transversales et longitudinales du revenu [MS-31]

Martin RENAUD\*, Statistique Canada

Michel LATOUCHE, Statistique Canada

15:00 Floor discussion • Discussion plénière

**13:30–15:00 Session 11****PMV 409****Extreme Values and Time Series • Valeurs extrêmes et séries chronologiques**

Contributed Paper Session • Communications libres

Chair • Président: Philippe CAPÉRAÀ, Université Laval

13:30 Exceedance over High Thresholds: A Guide to Threshold Selection • Modélisation des grandes valeurs : Un guide pour la sélection du seuil [MS-32]

D.J. DUPUIS\*, Dalhousie University

13:45 Estimation Procedure for the Dependence Function of Extreme Value Copulas • Méthode d'estimation de la fonction de dépendance des copules de valeurs extrêmes [MS-33]

Julie JACQUES\*, Université Laval

- 14:00 On a Lag-Window Estimator of the Covariances of Sample Autocovariances for a Strongly Mixing Process  
 • Sur un estimateur à délai pondéré des covariances des autocovariances échantillonnables pour un processus stationnaire fortement mélangeant [MS-34]  
 Daniel MOREAULT\*, Université de Montréal
- 14:15 Animation of the Storm in Osaka, Japan • Animation de l'orage à Osaka au Japon [MS-35]  
 Ryan LAM\*, University of Western Ontario  
 A.I. MCLEOD, University of Western Ontario
- 14:30 Floor discussion • Discussion plénière

**15:00–15:30 Coffee Break • Pause-café****PMV 411****15:30–17:00 Session 12****PMV 422****Biostatistics : Cancer Research • Biostatistique : Recherche sur le cancer**

Invited Paper Session • Conférences sur invitation

Organizer and Chair • Responsable et président: Salomon MINKIN, Ontario Cancer Institute and University of Toronto

- 15:30 Mechanistic Models of Carcinogenesis: What is their Use for Cancer Research and Risk Assessment ? •  
 Modèles mécanistes de cancérogenèse : quels sont leurs utilisations dans la recherche sur le cancer et dans l'évaluation des risques ? [MS-36]  
 Annette KOPP-SCHNEIDER\*, Deutsches Krebsforschungszentrum
- 16:15 Establishing Guidelines for Exposures to Carcinogens • Etablissement de lignes directrices pour l'exposition aux produits cancérogènes [MS-37]  
 David GAYLOR\*, National Center of Toxicological Research
- 17:00 Floor discussion • Discussion plénière

**15:30–17:00 Session 13****PMV 418****Multidimensional Statistics • Statistique multidimensionnelle**

Contributed Paper Session • Communications libres

Chair • Présidente: Debbie DUPUIS, Dalhousie University

- 15:30 Some Comments on Some Selected Matrix Inequalities and Equalities, with Some Statistical Applications-dedicated to the memory of Geoffrey Stuart Watson (1921-1998) • Commentaires concernant un florilège d'égalités et d'inégalités de matrices, avec des applications à la statistique - à la mémoire de Geoffrey Stuart Watson (1921-1998) [MS-38]  
 George P. H. STYAN\*, McGill University
- 15:45 Point Estimation of a Coefficient of Determination • Estimation ponctuelle d'un coefficient de détermination [MS-39]  
 Eric MARCHAND\*, Université du Nouveau-Brunswick
- 16:00 A Robust Test for a Given Variance • Un test robuste pour une variance donnée [MS-40]  
 Martin BILODEAU\*, Université de Montréal
- 16:15 Bivariate Sign Tests Based on the Sup,  $L_1$  and  $L_2$  Norms • Tests de signe basés sur les normes Sup,  $L_1$  et  $L_2$  pour des observations bivariées [MS-41]  
 Denis LAROCQUE\*, Université de Montréal
- 16:30 The Multivariate Linear Model with Matric-T Error Variables • Le modèle linéaire multivarié où les variables erreurs suivent la loi t matricielle [MS-42]  
 B. M. Golam KIBRIA\*, The University of Western Ontario  
 M. Safiul HAQ, The University of Western Ontario
- 16:45 Two Spaces of Projection for Regression • Deux espaces de projection pour la régression [MS-43]  
 Marylène TROUPÉ\*, Université des Antilles-Guyane
- 17:00 Floor discussion • Discussion plénière



**15:30–17:00 Session 14****PMV 409**

Statistical Education : Panel Discussion on Probability and Statistics in Canadian Schools • Éducation statistique : Atelier sur l'enseignement de la probabilité et de la statistique dans les écoles canadiennes

Invited Paper Session • Conférences sur invitation

Organizer and Chair • Responsable et présidente: Maureen TINGLEY, University of New Brunswick

- 15:30 Marginalized by the instructors and quickly forgotten by the students • Marginalisés par les enseignants et vite oubliés par les élèves [MS-44]  
Gordon FICK\*, University of Calgary
- 15:50 Skills for Daily Living - talking back to numbers • Un atout pour la vie quotidienne - savoir répondre aux chiffres [MS-45]  
Shirley MILLS\*, Carleton University
- 16:10 Painless statistics for non-statisticians - how can it be done? • Des statistiques faciles pour les non-statisticiens - Est-ce possible ? [MS-46]  
Hew GOUGH\*, Statistics Canada
- 16:30 Floor discussion • Discussion plénière

**15:30–17:00 Session 15****PMV 407**

Survey Methods: The Project to Improve Provincial Economic Statistics (PIPES) • Méthodologie d'enquête : Le projet d'amélioration des statistiques économiques provinciales (PASEP)

Contributed Paper Session • Communications libres

Organizer • Responsable: Julie BERNIER, Statistique Canada

Chair • Président: Normand LANIEL, Statistique Canada

- 15:30 Project to Improve Provincial Economic Statistics: Objectives and Pilot Surveys • Projet d'amélioration des statistiques économiques provinciales : Objectifs et enquête pilote [MS-47]  
Normand LANIEL\*, Statistique Canada  
D. ROYCE, Statistique Canada
- 15:45 The Business Register • Le Registre des entreprises [MS-48]  
Élaine CASTONGUAY\*, Statistique Canada
- 16:00 Questionnaire Development in PIPES • La conception de questionnaires à travers la vision du PASEP [MS-49]  
Mike SIROIS\*, Statistique Canada  
Sylvain PERRON, Statistique Canada  
Jean DROUIN, Statistique Canada  
Sylvain POIRIER, Statistique Canada
- 16:15 Sampling and Estimation for the Unified Enterprise Survey • Échantillonnage et estimation dans le cadre de l'enquête unifiée sur les entreprises [MS-50]  
Michelle SIMARD\*, Statistique Canada  
Guylaine DUBREUIL, Statistique Canada
- 16:30 Data collection for the Unified Enterprise Survey • La collecte des données pour l'enquête unifiée sur les entreprises [MS-51]  
Sylvain PERRON\*, Statistique Canada  
Mike SIROIS, Statistique Canada  
Jean DROUIN, Statistique Canada  
Sylvain POIRIER, Statistique Canada
- 16:45 Edit and Imputation of Data for the Unified Enterprise Survey • La vérification et l'imputation des données de l'enquête unifiée sur les entreprises [MS-52]  
Marie-Claude DUVAL\*, Statistique Canada  
Julie BERNIER, Statistique Canada

17:00 Calendarization in the Unified Enterprise Survey • La calendrialisation de l'enquête unifiée sur les entreprises  
[MS-53]  
Yanick BEAUCAGE\*, Statistique Canada  
Peter TARASSOFF, Statistique Canada

17:15 Floor discussion • Discussion plénière

**17:15**

**PMV 422**

Biostatistics Annual General Meeting • Assemblée générale annuelle de la section de biostatistique

**17:15**

**PMV 407**

Survey Methods Annual General Meeting • Assemblée générale annuelle de la section de méthodologie d'enquête

## Tuesday, June 2nd • Mardi 2 juin

---

### 8:00 Registration • Inscription

PMV 419

### 8:45–10:15 Session 16

D2 1060

#### SSC Presidential Invited Address • Allocution de l'invité du président

Special Lecture • Conférence principale

Chair • Présidente: Jane GENTLEMAN, Statistics Canada

Numbers and News: How Newsmakers- and the Media - Use and Abuse Economic Data • Comment les gens qui font l'évènement- et les média- utilisent et abusent des données économiques [MS-54]

Bruce LITTLE\*, Globe and Mail

10:15 Floor discussion • Discussion plénière

### 10:15–10:45 Coffee Break • Pause-café

PMV 411

### 10:45–12:15 Session 17

PMV 407

#### Biostatistics: Issues in the Pharmaceutical Industry, Multiple Endpoints in Clinical Trials

• Biostatistique : Débat dans l'industrie pharmaceutique, les essais cliniques à terminaison variable

Invited Paper Session • Conférences sur invitation

Organizer and Chair • Responsable et président: James D. MYLES, Glaxo Wellcome Inc.

10:45 Handling Multiple Outcome Measures in Prospective Pharmacoeconomic Studies • Traitement de réponses multiples dans les études prospectives pharmaco-économiques [MS-55]

Jayanti MUKHERJEE\*, Bristol-Mayers Squibb

11:15 Design Issue of Phase II Cancer Clinical Trials with Multiple Endpoints • Débat sur la planification des essais cliniques à terminaison variable pour le cancer en phase II [MS-56]

Benny ZEE\*, NCIC Clinical Trials Group

E. EISENHAUER, NCIC Clinical Trials Group

11:45 Testing Hypotheses with Multiple Endpoints in the Regulatory Context • Tests d'hypothèses à terminaison variable dans le contexte réglementé [MS-57]

Peter A. LACHENBRUCH\*, Food and Drug Administration

12:15 Floor discussion • Discussion plénière

### 10:45–12:15 Session 18

PMV 409

#### Industrial Statistics • Statistique industrielle

Invited Paper Session • Conférences sur invitation

Organizer and Chair • Responsable et président: Gemai CHEN, University of Regina

10:45 A Review of Response Surface Methodology: 1988-1998 • La méthodologie par les surfaces de réponses de 1988 à 1998 [MS-58]

G. Geoffrey VINING\*, University of Florida

11:15 Fractional Factorial Split-Plot Designs for Industrial Experiments • Fraction de plan d'expériences factorielles de type split-plot pour des expériences industrielles [MS-59]

Randy R. SITTER\*, Simon Fraser University

11:45 Current Research Directions in Statistical Process Control • Les directions de recherche actuelles dans le contrôle statistique de la qualité [MS-60]

William H. WOODALL\*, University of Alabama

12:15 Floor discussion • Discussion plénière

### 10:45–12:15 Session 19

PMV 417

#### Biostatistics : Linear Models • Biostatistique : Modèles linéaires

Contributed Paper Session • Communications libres

Chair • Président: Martin BILODEAU, Université de Montréal

- 10:45 Change Point Model for Asthma Mortality • Modèle avec point de rupture pour la mortalité par l'asthme [MS-61]  
 Hong YANG\*, McGill University  
 Robert PLATT, McGill University  
 Sammy SUISSA, McGill University
- 11:00 Bootstrap Methods for the Sensitivity of a Diagnostic Test • Des méthodes bootstrap pour la sensibilité d'un test de diagnostic [MS-62]  
 Robert PLATT\*, McGill University
- 11:15 Simultaneous Modeling of Operative and Long-Term Mortality after Coronary Artery Bypass Surgery, with Application to British Columbia Cardiac Registries Data • Modélisation simultanée de la mortalité opératoire et de la mortalité à long terme suite à la chirurgie coronarienne par contournement d'artères et application aux "British Columbia Cardiac Registries Data" [MS-63]  
 M. GHAHRAMANI\*, Simon Fraser University  
 C. DEAN, University of British Columbia  
 J. SPINELLI, Centre for Health Evaluation and Outcome Sciences
- 11:30 Power Comparisons of Several Tests in the Analysis of Cross-Over Clinical Trials • Comparaison de la puissance de plusieurs tests pour l'analyse de plans croisés [MS-64]  
 José Andrés CORREA\*, McGill University  
 François BELLAVANCE, McGill University
- 11:45 Modeling Effect Size Estimates in Meta-Analysis Using Fixed and Mixed Effects Linear Models • Modélisation de la variable dépendante dans une étude de méta-analyse à l'aide de modèles linéaires comprenant des effets fixes et aléatoires [MS-65]  
 Asmaâ MANSOUR\*, McGill University  
 François BELLAVANCE, McGill University  
 Martin COLE, McGill University
- 12:00 Rasch Models for Polychotomous Scales • Modèles de Rasch pour échelles polychotomiques [MS-66]  
 Jean-Marie TRICOT\*, Université de Bretagne Sud  
 Mounir MESBAH, Université de Bretagne-Sud

12:15 Floor discussion • Discussion plénière

### 10:45–12:15 Session 20

PMV 422

#### Special Address on Survey Methods • Conférence spéciale sur la méthodologie d'enquête

Special Lecture • Conférence principale

Chair • Président: Pierre LAVALLÉE, Statistique Canada

- Dealing with Non-Response Correction by Benchmarking or by Constrained Sampling • La correction de la non-réponse par calage ou par échantillonnage équilibré [MS-67]  
 Jean-Claude DEVILLE\*, I.N.S.E.E.

12:15 Floor discussion • Discussion plénière

### 12:30–13:30 Lunch • Dîner

PMF

**12:30–13:30 Session 21****PMV 419****Poster Session • Séance par affichage****Contributed Paper Session • Communications libres**

A Categorical Constraint Guided Matching Algorithm • Un algorithme de jumelage guidé par des contraintes catégorielles [MS-68]

Tzen-Ping LIU, Statistics Canada

Semi-parametric Estimation of the Distribution of Incubation Time for AIDS Using Cohort Data • Estimation semi-paramétrique de la distribution du temps d'incubation du SIDA à l'aide des données de cohorte [MS-69]

David E. MATTHEWS, University of Waterloo

Jianguo SUN, University of Missouri

Physician Intervention with Respect to their Smoking Patients • Les interventions des médecins face aux patients fumeurs [MS-70]

Niyonsenga T., Université de Sherbrooke

Courteau J., Université de Sherbrooke

Xhignesse M., Université de Sherbrooke

Vanasse N., Université de Sherbrooke

Laplante P., Université de Sherbrooke

Grant A., Université de Sherbrooke

**13:30–15:00 Session 22****PMV 422****Biostatistics: Applications • Biostatistique : Applications****Invited Paper Session • Conférences sur invitation**

Organizers • Responsables: Yves LEPAGE, Université de Montréal & François BELLAVANCE, Centre hospitalier St-Mary

Chair • Président: François BELLAVANCE, Centre hospitalier St-Mary

13:30 Survival Prediction for Terminally Ill Cancer Patients • Prédiction des durées de vie de patients en phase terminale de cancer [MS-71]

Benoît MÂSSE\*, Université Laval

Kamal DESAI, Université Laval

Pierre ALLARD, Université Laval

14:00 Family History Assessment: Development of a Family Risk Scale in the Evaluation of the Risk of Cardiovascular Diseases • Évaluation de l'histoire familiale : développement d'une échelle de risque familial pour les maladies cardiovasculaires [MS-72]

Théophile NIYONSENGA\*, Centre de recherche clinique (CRC-CUSE)

J. COURTEAU, Centre de recherche clinique (CRC-CUSE)

A. CIAMPI, Dept. of Family Medicine of McGill University

S. LUSSIER-CACAN, Institut de Recherche Clinique de Montréal (IRCM)

M. ROY, Institut de Recherche Clinique de Montréal (IRCM)

M. XHIGNESSE, CRC-CUSE, Univ. de Sherbrooke, IRCM

14:30 From Trees to Neural Nets • Des arbres aux réseaux de neurones [MS-73]

Antonio CIAMPI\*, Université McGill

15:00 Floor discussion • Discussion plénière

**13:30–17:00 Session 23****PMV 409****Special Session of the Centre de Recherches Mathématiques: Analysis of Shape • Session spéciale du Centre de Recherches Mathématiques : Analyse de la forme****Invited Paper Session • Conférences sur invitation**

Organizer and Chair • Responsable et président: Christopher SMALL, University of Waterloo

13:30 Detecting Differences in Brain Shapes Using Random Fields • Détection des différences de forme du cerveau en utilisant les champs aléatoires [MS-74]

Keith WORSLEY\*, McGill University

14:15 The Representation and Measurement of Shape and Applications • La représentation et la mesure des formes : théorie et applications [MS-75]  
Huiling LE\*, University of Nottingham

15:00–15:30 Coffee Break • Pause café

15:30 Statistical Shape Analysis of Landmarks and Images • Analyse statistique des formes codées à l'aide de repères et d'images [MS-76]  
Ian DRYDEN\*, University of Leeds

16:15 Discussant • Modérateur  
Christopher SMALL, University of Waterloo

### **13:30–15:00 Session 24**

**PMV 418**

**ASSQ Session : Quality Control in Industry • Session organisée par l'ASSQ : Contrôle de la qualité en milieu industriel**

Invited Paper Session • Conférences sur invitation

Organizer and Chair • Responsable et président: François PAGEAU, Les Technologies Industrielles SNC

13:30 Coordinators "EAQ" and Statistical Methods • Coordonnateur EAQ et méthodes statistiques [MS-77]  
Marc DUCHESNE\*, Via Systems

14:00 Confidence Limits for Capabilities Indices • Intervalles de confiance pour les indices de capabilités [MS-78]  
Daniel PROTEAU\*, C.M.P. Metal Products

14:30 Prevention of Premature Tool Wear in Metal Forming Using a Control Chart Adapted to the Nonnormality of the Data • Prévention de l'usure prématurée d'outils de formage du métal par une carte de contrôle adaptée à la non-normalité des données [MS-79]  
François PAGEAU\*, Les Technologies Industrielles SNC

15:00 Floor discussion • Discussion plénière

### **13:30–15:00 Session 25**

**PMV 407**

**Survey Methods: Modeling Survey Data • Méthodologie d'enquête : Modélisation des données d'enquête**

Contributed Paper Session • Communications libres

Chair • Président: John KOVAR, Statistics Canada

13:30 A General Framework for Estimating a Regression Model with or without Survey Data • Un cadre de travail pour estimer un modèle de régression avec ou sans données d'enquête [MS-80]  
Phillip S. KOTT\*, U.S. National Agricultural Statistics Service

13:45 Estimation of Populations Totals/Mean in Multiple Characteristic Surveys when Study Variable has a Weaker Relationship with the Size Variable • Estimation de totaux/moyennes dans les sondages à caractéristiques multiples quand la variable étudiée est faiblement liée à la variable taille [MS-81]  
Pranesh KUMAR\*, University of Transkei

14:00 Outlier Detection: A Comparison of Inter-quartile Range versus a Variation of Sigma Gap Using Asymmetric Samples • Détection de données aberrantes sur des échantillons de distributions asymétriques. Comparaison de 2 méthodes : intervalle inter-quartile et règle de l'écart-sigma modifié [MS-82]  
K. NOBREGA\*, Statistics Canada  
J. BERNIER, Statistics Canada

14:15 A Statistical Approach for Disaggregating Mixed-Frequency Economic Time Series • Une approche statistique pour la désagrégation de fréquences mixtes dans les séries chronologiques [MS-83]  
Zhao-Guo CHEN\*, Statistics Canada

14:30 Should Stores be Open on Sunday? The Impact of Sunday Opening on the Retail Trade Sector in New Brunswick • Les conséquences sur le commerce de détail au Nouveau Brunswick de l'ouverture des commerces le dimanche [MS-84]  
Benoît QUENNEVILLE\*, Statistics Canada

Pierre CHOLETTE, Statistics Canada  
 Marietta MORRY, Statistics Canada

14:45 Floor discussion • Discussion plénière

### 13:30–15:00 Session 26

PMV 417

#### Biostatistics : Mixed Topics • Biostatistique : Sujets variés

Contributed Paper Session • Communications libres

Chair • Président: Rollin BRANT, University of Calgary

13:30 Statistics Canada's Environmental Statistics Program: Future Directions • Le programme statistique sur l'environnement de Statistique Canada : Directions futures [MS-85]

Cynthia BAUMGARTEN\*, Statistics Canada

Claude SIMARD, Statistics Canada

13:45 Simultaneous Modelling of Longitudinal and Cross-Sectional Dependence for Binary Outcomes • Modélisation simultanée de la dépendance longitudinale et transversale de réponses binaires [MS-86]

Patrick J. FARRELL\*, Acadia University

Andreas I. SASHEGYI, University of Waterloo

K. Stephen BROWN, University of Waterloo

14:00 Cooking with Statistics: Recipes for Calculating the Probability of Illness from Food Borne Microbial Pathogens • Cuisine statistique : Recettes pour calculer la probabilité de tomber malade à cause de microbes pathogènes dans les aliments [MS-87]

W. H. ROSS\*, Health Canada

S. HAYWARD, Health Canada

14:15 Local Likelihood and Profile Likelihood in Semiparametric Hazard Regression Model • Vraisemblance locale et vraisemblance de profil dans les modèles semi paramétriques de régression de la fonction de risque [MS-88]

Xuewen LU\*, Agriculture and Agri-Food Canada

R.S. SINGH, University of Guelph

14:30 Least Squares By Completing The Square • Moindres carrés en complétant les carrés [MS-89]

Gordon H. FICK\*, University of Calgary

14:45 Actuarial Adjustment for a Mortality Deficient Table • Ajustements actuariels pour une table de mortalité déficiente [MS-90]

Remy, Konga KONGA\*, Compagnie National d'Assurance, Cameroun

15:00 Floor discussion • Discussion plénière

### 15:00–15:30 Coffee Break • Pause-café

PMV 411

### 15:30–17:00 Session 27

PMV 417

#### Biostatistics : Estimation • Biostatistique : Estimation

Contributed Paper Session • Communications libres

Chair • Président: Théophile NIYONSENGA, Université de Sherbrooke

15:30 Weakly Parametric Models for Doubly Censored Data • Modèles faiblement paramétriques pour des données doublement censurées [MS-91]

Min ZHAN\*, University of Waterloo

15:45 Bootstrap Variable Selection and Model Validation for Cox's Regression Model • Sélection de variables pour le modèle de Cox par la méthode du Bootstrap [MS-92]

Feng ZHAO\*, Queen's University

Dongsheng TU, Queen's University

Joe PATER, Queen's University

- 16:00 Choice of a Time Scale in Survival Analysis • Choix d'une échelle de temps en analyse de durées de vie [MS-93]  
T. DUCHESNE\*, University of Waterloo
- 16:15 Estimation of the Sterilization Time for Dental Instruments • Estimation du temps requis pour stériliser les instruments d'examen dentaire [MS-94]  
Marie Claude LAMBERT\*, Université Laval
- 16:30 Estimation of Small Mammals' Biomass Available for Coyotes Living in Forest and Farmland of Eastern Quebec • Estimation de la biomasse de petits mammifères disponible aux coyotes dans les milieux forestier et agro-forestier de l'est du Québec [MS-95]  
Nancy CARON\*, Université Laval
- 16:45 Comparison of Generalized Procrustree Analysis and the Intra-class Correlation Coefficient to Estimate Inter-Rater Reliability • Comparaison de l'analyse Procustéenne généralisée au coefficient de corrélation intra-classe pour estimer la concordance inter-observateurs [MS-96]  
Natalie RODRIGUE\*, McGill University  
François BELLAVANCE, McGill University
- 17:00 Joint Analysis of Survival and Quality of Life in Cancer Disease • Analyse conjointe de la survie et de la qualité de vie dans le traitement du cancer [MS-97]  
Mounir MESBAH\*, Université de Bretagne Sud  
Lucile AWAD, Rhône-Poulenc Rorer
- 17:15 Floor discussion • Discussion plénière

**15:30–17:00 Session 28****PMV 418**

ASSQ Session: Panel Discussion on Statistical Consulting • Session organisée par l'ASSQ : Atelier sur la consultation statistique

Invited Paper Session • Conférences sur invitation

Organizer and Chair • Responsable et président: Sylvain VÉGIARD, Ministère des ressources naturelles du Québec

Panelists • Panélistes [MS-98]

Diane LEROUX, Ministère des transports du Québec

Michel FLUET, SOM recherches et sondages

Hélène CRÉPEAU, Université Laval

**15:30–17:00 Session 29****PMV 407**

Survey Methods: Variance Estimation, Calibration, and Related Topics • Méthodologie d'enquête : Estimation de la variance, calage et sujets connexes

Contributed Paper Session • Communications libres

Chair • Président: Benoit QUENNEVILLE, Statistique Canada

15:30 Enhancements in the Series of Monthly Estimates for the Retail Sales by a Calibration Estimator • Amélioration de la série d'estimations mensuelles des ventes au détail au moyen d'un estimateur par calage [MS-99]

Joël BISSONNETTE\*, Statistique Canada

Isabelle MARCHAND, Statistique Canada

Martin ST-PIERRE, Statistique Canada

Julie TRÉPANIÉ, Statistique Canada

15:45 Parameter Specification for the Model Variance Structure in the Generalized Regression Estimator • Spécification du paramètre de la structure de variance du modèle dans l'estimateur de régression généralisée [MS-100]

Éric PELLETIER\*, Statistique Canada

Éric RANCOURT, Statistique Canada

16:00 Small Area Estimation Using Empirical Bayes Methods • Estimation dans les petites régions par la méthode de Bayes empirique [MS-101]

Ève BELMONTE\*, Université Laval



- 16:15 Robust Estimators by Benchmarking with Constrained Weights • Estimateurs de calage robustes avec poids contraints [MS-102]  
Pierre DUCHESNE\*, Université de Montréal
- 16:30 Matching Data from the Labour Force Survey with the Census • Appariement de données de l'Enquête sur la population active canadienne à celles du recensement de la population [MS-103]  
Isabelle LÉVESQUE\*, Statistique Canada  
Benoit ALLARD, Statistique Canada  
Johane DUFOUR, Statistique Canada
- 16:45 Model to Predict the Probability of a Person Having Employment Activity Limitation Using Data from the Health and Activity Limitation Survey (HALS) • Modèle pour prédire la probabilité d'avoir une limitation d'activité reliée à l'emploi à partir de l'enquête sur la santé et les limitations d'activité [MS-104]  
Daniel HURTUBISE\*, Statistique Canada  
Éric LANGLET, Statistique Canada
- 17:00 Floor discussion • Discussion plénière

**10:45–12:15 Session 30****PMV 422****Survey Methods and Biostatistics: New Approaches to Data Analysis • Méthodologie d'enquête et biostatistique : Nouvelles approches pour l'analyse des données**

Invited Paper Session • Conférences sur invitation

Organizer and Chair • Responsable et président: M. P. SINGH, Statistique Canada

- 15:30 A New Look at Confidence Intervals in Survey Sampling • Un nouveau regard sur les intervalles de confiance en échantillonnage [MS-105]  
V. P. GODAMBE\*, University of Waterloo
- 16:00 Multistage Designs and Estimation in Biostatistics • Plans à plusieurs degrés et estimation en biostatistique [MS-106]  
Jerry LAWLESS\*, University of Waterloo
- 16:30 Design-Free Analysis • Analyse libre du plan d'échantillonnage [MS-107]  
Fritz SCHEUREN\*, Ernst and Young
- 17:00 Floor discussion • Discussion plénière

**17:15****PMV 422****SSC Annual General Meeting • Assemblée générale annuelle de la SSC****17:15****PMV 407****ASSQ Annual General Meeting • Assemblée générale annuelle de l'ASSQ****19:00 Banquet****Hôtel Delta Hotel, Salle Sherbrooke III Room**

## Wednesday, June 3th • Mercredi 3 juin

---

**8:00 Registration • Inscription**

**PMV 419**

**8:45–9:30 Session 31**

**D2 1060**

**Pierre Robillard Award Winner Lecture • Allocution du (de la) lauréat(e) du prix Pierre-Robillard**

Special Lecture • Conférence principale

Chair • Président: Mary LESPERANCE, University of Victoria

Speaker's name and abstract will be available June 1 • Le nom du conférencier ou de la conférencière et le résumé seront disponibles le premier juin

**9:30–10:15 Session 32**

**D2 1060**

**Canadian Journal of Statistics Award Winner Lecture • Allocution du (de la) lauréat(e) du prix de la Revue Canadienne de Statistique**

Special Lecture • Conférence principale

Chair • Présidente: Gerarda DARLINGTON, Cancer Care Ontario

Speaker's name and abstract will be available June 1 • Le nom du conférencier ou de la conférencière et le résumé seront disponibles le premier juin

**10:15–10:45 Coffee Break • Pause-café**

**PMV 411**

**10:45–12:15 Session 33**

**PMV 418**

**Topics in Environmental Statistics • Statistique de l'environnement**

Invited Paper Session • Conférences sur invitation

Chair • Président: Claude LABERGE, STATEX Québec

Organizer • Responsable: Louis-Paul RIVEST, Université Laval

10:45 Use of Ratio Estimators to Estimate the Size of Common Eider Populations in Winter • L'estimation par le quotient de la taille des populations hivernantes d'Eider à duvet [MS-108]

Nathalie PLANTE\*, Service canadien de la Faune

Daniel BORDAGE, Service canadien de la Faune

André BOURGET, Service canadien de la Faune

Sylvain PARADIS, Service canadien de la Faune

11:15 ANOVA Involving Qualitative and Quantitative Factors: Environment Impact on Mussel Growth • Croisement de facteurs qualitatif et quantitatif en ANOVA : Impact du milieu sur la croissance des moules [MS-109]

Gaétan DAIGLE\*, Université Laval

E. BOURGET, Université Laval

L.-P. RIVEST, Université Laval

11:45 Network of Monitoring Stations for Detecting Trends in the Quality of Water in Lakes in the Presence of Spatial Correlation • Rationalisation d'un réseau d'échantillonnage pour la détection de tendance dans la qualité de l'eau des lacs en présence d'autocorrélation spatiale [MS-110]

Claude LABERGE\*, STATEX Québec

12:15 Floor discussion • Discussion plénière

**10:45–12:15 Session 34****PMV 409**

Case Studies • Études de cas  
 Habitat Usage Versus Availability for the Cutthroat Trout in Dutch Creek, Alberta •  
 Utilisation, disponibilité et usage de l'habitat pour la truite "coupe-gorge" dans le Dutch  
 Creek en Alberta

Organizer and Chair • Responsable et présidente: Jeanette O'HARA-HINES, University of Waterloo

10:45 Data Presentation • Présentation des données

10:55 Comparison of Habitat Availability to Habitat Usage by Cutthroat Trout • Comparaison de l'utilisation et de  
 la disponibilité de l'habitat pour la truite "coupe-gorge"

Dieter AYERS, University of British Columbia

Jeevanantham RAJES, University of British Columbia

Steven WANG, University of British Columbia

11:15 Tracking trout: a Cutthroat Analysis • Suivre les truites : une analyse coupe-gorge

Joanna BIRNACKA, McMaster University

Keshav RATHI, McMaster University

Roman VIVEROS, McMaster University

Tara SRINIVASAN, McMaster University

Peter MACDONALD, McMaster University

11:35 An Analysis of the Relationship Between Depth Usage and Depth Availability for Cutthroat Trout • Une  
 analyse de la relation entre l'utilisation et la disponibilité de la profondeur pour la truite "coupe-gorge"

Nicky ROSLIN, University of Guelph

Bill KOKIOPOULOS, University of Guelph

11:55 Discussion • Discussion

**10:45–12:15 Session 35****PMV 422**

Theory and Applications of Classification • Théorie et applications en classification

Invited Paper Session • Conférences sur invitation

Organizer and Chair • Responsable et présidente: Pascale ROUSSEAU, Université de Québec à Montréal

10:45 Phylogenetics and Genomics • Phylogénétique et Génomique [MS-111]

David SANKOV\*, Université de Montréal

11:15 Classification Tests with Simply, Doubly, and Triply Repeated Measures • Tests de classifications avec  
 mesures répétées simples, doubles et triples [MS-112]

Pierre DUTILLEUL\*, McGill University

11:45 How to Validate our Classifications? • Comment valider nos classifications ? [MS-113]

François-Joseph LAPOINTE\*, Université de Montréal

12:15 Floor discussion • Discussion plénière

**10:45–12:15 Session 36****PMV 417**

Probability and Mathematical Statistics • Probabilité et statistique mathématique

Contributed Paper Session • Communications libres

Chair • Président: Paul CABILIO, University of Acadia

10:45 Some Quadratic Estimating Function For the Analysis of Proportions • Quelques fonctions quadratiques  
 d'estimation pour l'analyse des proportions [MS-114]

S.R. PAUL\*, University of Windsor

11:00 On Approximated Moments of Order Statistics from Progressively Censored Samples • Les moments appro-  
 chés des statistiques d'ordre d'échantillons censurés progressivement [MS-115]

Uditha BALASOORIYA\*, Memorial University of Newfoundland  
Sutaip S. SAW, Memorial University of Newfoundland

- 11:15 The Gibbs Sampler Converges Slowly ! • L'échantillonneur de Gibbs converge lentement ! [MS-116]  
Claude BÉLISLE\*, Université Laval
- 11:30 Bootstrapping the Convex Hull • La méthode du bootstrap appliquée à l'enveloppe convexe [MS-117]  
Mahmoud ZAREPOUR\*, University of Ottawa
- 11:45 The Vervaat Process: Theory and Applications • Le processus de Vervaat : théorie et applications [MS-118]  
Ričardas ZITIKIS\*, University of Ottawa
- 12:00 Non Parametric Estimation of Functions Associated to a Copula • Estimation non-paramétrique de fonctions associées à une copule [MS-119]  
Abderrahmane CHAKAK\*, Université Abdelmalek Essaadi  
M. EZZERG, Université Abdelmalek Essaadi  
L. IMLAHI, Université Abdelmalek Essaadi
- 12:15 Floor discussion • Discussion plénière

### 10:45–12:15 Session 37

**PMV 407**

**Survey Methods : Survey Design • Méthodologie d'enquête : Conception d'enquêtes**  
Contributed Paper Session • Communications libres

Chair • Président: Robert COURTEMANCHE, Bureau de la Statistique du Québec

- 10:45 An Example of the Challenges of the Special Business Surveys • Un exemple des défis des enquêtes spéciales d'entreprises [MS-120]  
Tony LABILLOIS\*, Statistique Canada
- 11:00 Electronic Methods of Data Collection for a Survey Studying Internet Users in Quebec • Méthode électronique de collecte des données pour une enquête auprès des internautes québécois [MS-121]  
Eric LACROIX\*, Bureau de la Statistique du Québec
- 11:15 Workplace and Employee Survey Methodology Overview • Enquête sur le milieu de travail et les employés : aperçu de la méthodologie [MS-122]  
B. LAPIERRE\*, Statistique Canada  
Z. PATAK, Statistique Canada  
J. MOREL, Statistique Canada
- 11:30 Treatment of Enterprise Restructuring in Business Surveys • Traitement des restructurations d'entreprises dans les panels économiques [MS-123]  
Pierre LAVALLÉE\*, Statistique Canada  
Peter STRUIJS, Eurostat
- 11:45 Redesign of the Survey of Household Spending : Increase in Data Quality and Redistribution of Response Burden • Remaniement de l'enquête sur les dépenses des ménages : amélioration de la qualité de l'information retenue et redistribution du fardeau de réponse [MS-124]  
Sophie ARSENAULT\*, Statistique Canada  
Johanne TREMBLAY, Statistique Canada
- 12:00 Data Quality of Earning Questions in the Canadian Labour Force Survey • Qualité des données sur le revenu pour l'Enquête canadienne sur la population active [MS-125]  
François BRISEBOIS\*, Statistique Canada  
C. WALKER, Statistique Canada
- 12:15 Experience in the Bias Analysis in the Canadian International Travel Survey • Analyse du biais de l'enquête canadienne sur les voyages internationaux [MS-126]  
Stéphane TREMBLAY\*, Statistique Canada  
François LAFLAMME, Statistique Canada
- 12:30 Floor discussion • Discussion plénière

**12:30–13:30 Lunch • Dîner**

**PMF**

**13:30–15:00 Session 38****PMV 422****Biostatistics : Topics in Environmental Statistics • Biostatistique : Statistique de l'environnement**

Invited Paper Session • Conférences sur invitation

Organizer and Chair • Responsable et président: Roman VIVERO-AGUILERA, McMaster University

13:30 The Accuracy of Ranked Set Sampling • Sur la précision d'échantillonnage d'ensembles rangés [MS-127]  
Abdel H. EL-SHAARAWI\*, National Water Research Institute14:00 Public Perception of Environmental Risks and Pre-hoc Statistical Analysis • La perception du public des risques environnementaux et les analyses statistiques 'Pre-hoc' [MS-128]  
Lara J. WOLFSON\*, University of Waterloo14:30 Statistical Issues in the Development of Regional and National Scale Environmental Monitoring Programs • Les aspects statistiques dans le développement de programmes de contrôles environnementaux à l'échelle régionale et nationale [MS-129]  
Anthony R. OLSEN\*, US EPA National Health and Environmental Effects Research Laboratory

15:00 Floor discussion • Discussion plénière

**13:30–15:00 Session 39****PMV 417****Mathematical Statistics • Statistique mathématique**

Contributed Paper Session • Communications libres

Chair • Président: Claude BÉLISLE, Université Laval

13:30 Two Barrier Wiener Process Model for Length of Stay in the Hospital • Un processus de Wiener à deux barrières pour le temps de séjour des patients à l'hôpital [MS-130]  
Julie HORROCKS\*, University of Waterloo13:45 Reducing Delays in Preemptive Priority Queues • Réduction du retard pour les files répétées de priorité de préemption [MS-131]  
Steve DREKIC\*, University of Western Ontario  
David A. STANFORD, University of Western Ontario14:00 Shrinking the Polychotomous Logistic Regression via the Lasso • Le rétrécissement de la régression logistique polychotomique avec le 'Lasso' [MS-132]  
Carmen MAK\*, University of Toronto14:15 Some Practical Designs for Biased Bivariate Regression • Quelques schémas faisables pour la régression bivariée biaisée [MS-133]  
Giseon HEO\*, University of Alberta14:30 Robust Confidence Intervals and the Bootstrap • Intervalles de confiance robustes et le bootstrap [MS-134]  
Matias SALIBIAN-BARRERA\*, University of British Columbia14:45 Comparative Study of Some Deconvolution Methods for Densities • Etude comparative de méthodes de déconvolution de densité [MS-135]  
Stéphane GERMAIN\*, Université Laval

15:00 Floor discussion • Discussion plénière

**13:30–15:00 Session 40****PMV 407****Survey Methods : Outliers in Survey Data • Méthodologie d'enquête : Traitement des valeurs aberrantes**

Invited Paper Session • Conférences sur invitation

Organizer and Chair • Responsable et président: Hynshik LEE, Westat

13:30 Influential Observations : Identification and Treatment by M-estimators • Observations influentes : identification et traitement par les M-estimateurs [MS-136]  
Jean-Philippe GWET\*, Westat

14:00 On Winsorization in Business Surveys • Sur la winsorisation dans les enquêtes entreprises [MS-137]

Phil KOKIC\*, University of Southampton  
Ray CHAMBERS, University of Southampton  
Paul SMITH, Office for National Statistics

14:30 Outlier Robust Generalized Regression Estimator • Estimateur généralisé par la régression robuste à la présence de valeurs extrêmes [MS-138]

Hyunshik LEE\*, Westat  
Zdenek PATAK, Statistics Canada  
Sarah FRANKLIN, Statistics Canada

15:00 Floor discussion • Discussion plénière

### **13:30–15:00 Session 41**

**PMV 418**

**Committee on Women in Statistics & The Caucus for Women in Statistics: Classification • Comité des femmes en statistique & Caucus pour les femmes en statistique : Classification**

Invited Paper Session • Conférences sur invitation

Organizer and Chair • Responsable et présidente: Judy Ann CHAPMAN, Women's College Hospital in Toronto

13:30 Clustering and characterization of HI Astronomical Objects Using Density Estimation • Une approche basée sur l'estimation de la densité pour la caractérisation et la classification d'objets astronomiques [MS-139]

Nadia GHAZZALI\*, Université Laval

14:15 Logistic Regression Trees, Reduced Logistic Regression and Generalized Individualized Regression • Arbre de régression logistique, modèle logistique réduit et régression individuelle généralisée [MS-140]

Pascale ROUSSEAU\*, Université de Québec à Montréal

15:00 Floor discussion • Discussion plénière

## Abstracts • Résumés

### Session

Committee on Women in Statistics & The  
Caucus for Women in Statistics: Sunday  
Evening Reception

*Comité des femmes en statistique & Caucus  
pour les femmes en statistique : Réception  
du dimanche soir*

[MS-1]

**Gender and Work: For Richer or Poorer**

*Le travail des hommes et des femmes : qui s'enrichit ?*

Nancy GALAM, Statistics Canada

Men and women have long performed very different kinds of work. Despite the massive influx of women into the labour market over the past three decades, both paid and unpaid work in Canada continues to be segregated according to sex. Using the most recent available statistics, including data from the 1996 Census, this presentation examines gender differences in the Canadian population with respect to labour force activity, income and earnings, and unpaid work. These differences are discussed in the context of broader social trends, such as female participation in the labour market, family structure and living arrangements, and attitudes towards gender, work and family.

*Les hommes et les femmes ont longtemps eu des tâches très différentes. Malgré l'entrée massive des femmes sur le marché du travail au cours des trois dernières décennies, le travail rémunéré et non rémunéré au Canada continue à connaître une ségrégation sur la base du sexe. Utilisant les statistiques les plus récentes, incluant les données du recensement de 1996, cette présentation examine les différences homme-femme dans la population canadienne selon l'activité, le revenu, et le travail non payé. Ces différences sont discutées dans un cadre social plus large, tel que la participation des femmes au marché du travail, la structure et les arrangements familiaux, et les attitudes vis à vis du sexe, du travail et de la famille.*

Nancy GALAM, Housing, Family and Social Statistics Division 7th floor, Jean Talon Building Statistics Canada Ottawa, Ontario K1A 0T6

GHALNAN@statcan.ca

### Session 1

### 1997 SSC Gold Medal Address

*Allocution du récipiendaire 1997 de la médaille d'or de la SSC*

[MS-2]

**Statistical Services — Preparing for the Future**

*Les services statistiques — Comment se préparer pour l'avenir*

Ivan P. FELLEGI\*, Chief Statistician of Canada • Statisticien en chef du Canada

On this ceremonial occasion, I want to look forward: to the challenges facing statistical agencies and how to prepare for them. I will first of all review the context within which statistical offices must evolve: first the main forces at work that are modifying our economy and society; second, specific policy issues that need to be addressed; and third, changes in the nature of government and in expectations from it.

I will then try to outline an internal strategy for statistical offices derived from the foregoing analysis. On the one hand, this will require the development of new types of data system that are needed. I will illustrate these with reference to recent initiatives by Statistics Canada. While crucial, these new data system will probably only account for a relatively small proportion of our expenditure. Hence a second important component of our strategy designed to cope with external social and economic changes must be to ensure a high level of adaptability of our core activities. Such adaptability requires specific initiatives.

Finally, I will describe the elements of an <<external strategy>> : how to manage our interactions with the world around us. Three elements will be touched upon: achieving and maintaining a high level of relevance; the issue of political objectivity and its perception; and international collaboration. In each case I will try to outline specific measures that I consider essential.

*En cette occasion solennelle, je désire regarder en avant et examiner les défis auxquels feront face les organismes statistiques ainsi que la façon dont ils doivent se préparer à les affronter. En premier lieu, je présenterai le contexte dans lequel évoluent les organismes statistiques : d'abord, les principales forces qui modifient notre économie et notre société ; ensuite, les questions de politique générale auxquelles il faut s'attaquer ; enfin, les changements que subit la nature du gouvernement et les attentes qu'on peut en avoir.*

*En second lieu, je tenterai, à partir de l'analyse qui aura précédé, d'esquisser une stratégie interne à l'intention des bureaux statistiques. D'une part, cela nécessitera l'élaboration de nouveaux types de systèmes de données. J'illustrerai ce point en me basant sur de récentes initiatives de Statistique Canada. Bien qu'essentiels, ces nouveaux systèmes de données ne représenteront probablement qu'une*

*proportion relativement faible de nos dépenses. D'où le deuxième élément fondamental de notre stratégie, conçu en vue de faire face aux changements sociaux et économiques qui se produisent à l'extérieur, qui consiste à s'assurer que nos activités de base sont facilement adaptables. Une telle adaptabilité exige des initiatives bien particulières.*

*En dernier lieu, je décrirai les éléments d'une <<stratégie externe>> : comment gérer nos interactions avec le monde qui nous entoure. À ce propos, je traiterai de trois points : la façon d'atteindre et de maintenir un niveau élevé de pertinence ; la question de l'objectivité politique et la perception qu'on en a ; et la collaboration à l'échelle internationale. Dans chaque cas, je tenterai de donner un aperçu des mesures particulières que j'estime essentielles.*

Ivan P. FELLEGI, Statisticien en chef du Canada Statistique Canada  
Immeuble R.H. Coats Ottawa, Ontario K1A 0T6

## Session 2

### Special Biostatistics Session of the Fields Institute : Event History Analysis

#### *Session spéciale de l'Institut Fields en biostatistique : Analyse de séries d'évènements*

#### [MS-3]

#### **A Frailty Model for the Incidence of Testicular Cancer**

#### *Un modèle de "frailty" pour l'incidence du cancer des testicules*

Odd. A. AALEN\*, University of Oslo

There are two distinct statistical features concerning the incidence of cancer of the testis. Firstly, the incidence has been increasing for many years. Secondly, the incidence is highest among younger men and then declines with age.

A statistical model to account for these features is suggested. The model is based on the frailty thinking which is now common in survival analysis. Here, individuals are assumed to have different risks according to whether a certain damage has occurred during foetal life.

Data from the Norwegian Cancer registry are used for analysis.

*Il y a deux aspects statistiques différents concernant l'incidence du cancer des testicules. Premièrement, l'incidence a été en croissance depuis plusieurs années. Deuxièmement, l'incidence est grande parmi les jeunes adultes et ensuite diminue avec l'âge.*

*Un modèle statistique qui tient compte de ces caractéristiques est suggéré. Le modèle est basé sur le concept de fragilité, lequel est d'usage courant dans les analyses des durées de survie. Ici, les individus sont supposés avoir différents risques selon qu'un certain dommage s'est produit durant leur vie foetale.*

*Des données du registre norvégien du cancer sont analysées à l'aide de ce modèle.*

Odd. A. AALEN, Section of Medical Statistics University of Oslo

P.O. Box 1122 Blindern N-0317 Oslo NORWAY

o.o.aalen@basalmed.uio.no

#### [MS-4]

#### **Goodness-of-Fit Tests for the Multiplicative Intensity Model**

#### *Tests d'adéquation pour des modèles à intensité multiplicative*

David HARRINGTON\*, Dana-Faber Cancer Institute

Cheryl JONES, Harvard School of Public Health

The last few years have seen increased use of semi-parametric regression models for repeated event data. Many of the methods used generalize the Cox proportional hazards model; the most popular of these models are the multiplicative intensity (MI) model due to Andersen and Gill, the marginal model for correlated survival data, due to Wei, Lin and Weissfeld, and the renewal-type model studied by Prentice, Williams and Petersen. The multiplicative intensity model is relatively easy to estimate using standard software that is widely available, but has the strongest assumptions of these three approaches. This talk will examine numerical and graphical goodness-of-fit methods for the MI model based on the generalized residuals for right censored data studied in the papers by Barlow and Prentice and by Therneau, Grambsch and Fleming. More specifically, we examine test statistics designed to detect violations from the mean zero, independent increment assumptions of the martingales that are basis for the MI model. We examine both asymptotic and small sample sampling distributions of the test statistics under a variety of assumptions.

*Les dernières années ont connu une utilisation grandissante des modèles de régression semi-paramétriques pour les données à événements répétés. Plusieurs méthodes utilisées généralisent le modèle de hasard proportionnel de Cox ; les plus populaires de ces modèles sont le modèle d'intensité multiplicatif (IM) de Anderson et Gill, le modèle marginal pour les données corrélées des durées de survie de Wei, Lin et Weissfeld, et le modèle de renouvellement étudié par Prentice, Williams et Petersen. Le modèle d'intensité multiplicatif est relativement simple à estimer par les logiciels standards d'usage courant, mais c'est celui qui repose sur les hypothèses les plus contraignantes. Cette présentation examine les méthodes numériques et graphiques des tests d'adéquation pour les modèles IM basés sur les résidus généralisés pour les données censurées à droite étudiées dans les papiers de Barlow et Prentice et de Therneau, Grambsch et Fleming. Plus spécifiquement, nous examinons les tests statistiques faits pour détecter la non nullité de la moyenne, l'indépendance des accroissements des martingales qui sont à la base des modèles IM. Nous examinons les distributions pour les petits échantillons et les distributions asymptotiques des tests statistiques sous des hypothèses assez variées.*

David HARRINGTON, Division of Biostatistics Dana-Faber Cancer Institute 44 Binney Street Boston, MA 02115 U.S.A.

dph@jimmy.harvard.edu

#### [MS-5]



## Analysis of Event History Data Through Rate Functions Modeling

### *Analyse des données historiques d'événements par la modélisation des fonctions de taux*

Claude NADEAU\*, Université Laval

In the last few years, various new approaches for the analysis of event history data (or point process data) have been proposed. These methods, usually based on models that don't involve intensity functions, provide alternatives to likelihood inference.

The methodology we present consists in modeling the marginal means (also called rates) of point processes. The models proposed may be parametric or semiparametric, easily handle covariates and may involve the specification of higher moments. Unbiased estimating functions are used to perform inference (parameter and robust variance estimates). This approach mimics what Liang and Zeger (1986) (and others afterwards) did for the analysis of longitudinal data. We'll discuss the merits of our approach relative to standard analysis based on intensity functions or other alternatives. We'll do the same for the limitations with an emphasis on censoring issues and what our approach (and others') can and cannot handle well.

*Au cours des dernières années, plusieurs nouvelles approches pour l'analyse des données d'événements historiques (ou données de processus de points) ont été proposées. Ces méthodes, usuellement basées sur des modèles ne faisant pas intervenir les fonctions d'intensité fournissent des alternatives à l'inférence par la fonction de vraisemblance.*

*La méthodologie que nous présentons consiste à modéliser les moyennes marginales (aussi appelées taux) de processus de points. Les modèles proposés peuvent être paramétriques ou semi-paramétriques, facilement inclure des covariables et incorporer des spécifications de moments d'ordres supérieurs. Les fonctions d'estimation sans biais sont utilisées pour des fins d'inférence (Estimation des paramètres et des variances robuste). Cette approche mime ce que Liang and Zeger (1986) ( et d'autres par la suite) ont fait pour l'analyse des données longitudinales. Nous discuterons des mérites de notre approche en comparaison avec les analyses standards basées sur les fonctions d'intensité ou d'autres alternatives. Nous discuterons également des limitations, avec un accent sur les problèmes de censure, et sur ce que l'approche suggérée et ses concurrentes peuvent traiter.*

Claude NADEAU, École d'actuariat Université Laval Québec QC G1K 7P4  
cnadeau@act.ulaval.ca

## Session 4

### Probability : Stochastic Processes

#### *Probabilité : Processus stochastiques*

[MS-6]

## Empirical Processes Based on Pseudo-Observations

### *Processus empiriques basés sur des pseudo-observations*

Bruno RÉMILLARD\*, Université du Québec à Trois-Rivières

Usually, empirical distribution functions are constructed to estimate the theoretical distribution function of known functions  $\theta(X)$  of the observable random variable  $X$ .

In practice, many statisticians are using empirical distribution functions constructed from residuals, which are estimations of the non-observable error terms in linear models. A more general problem is to approximate the distribution function of a non-observable random variable  $\theta(P, X)$  depending on an observable random variable  $X$  together with its unknown law  $P$ . When  $P$  is estimated by  $P_n$ , the quantities  $\theta(P_n, X_i)$  are called pseudo-observations. Some work has been done recently when the pseudo-observations are the so-called residuals of linear models.

The goal of this talk is to provide some tools to study the asymptotic behaviour of empirical processes constructed from general pseudo-observations. Interesting examples of pseudo-observations will be given together with applications to copulas and other statistical concepts.

*Les fonctions de distribution empiriques sont habituellement construites afin d'estimer la fonction de distribution théorique d'une fonction connue  $\theta(X)$  d'une variable aléatoire  $X$ .*

*Toutefois, en pratique, plusieurs statisticiens utilisent des fonctions de distribution empiriques construites à partir de résidus, qui sont des estimations de termes d'erreur non-observables dans les modèles linéaires. Un problème encore plus général est l'estimation de la fonction de distribution d'une variable aléatoire non-observable  $\theta(P, X)$ , dépendant d'une variable aléatoire observable  $X$  et de sa loi inconnue  $P$ . Lorsque  $P$  est estimée par  $P_n$ , les quantités  $\theta(P_n, X_i)$  sont dites pseudo-observations. Quelques travaux ont été réalisés récemment lorsque les pseudo-observations sont les résidus de modèles linéaires.*

*Le but de cette conférence est de donner des outils pour l'étude du comportement asymptotique de processus empiriques basés sur des pseudo-observations générales. Des exemples intéressants de pseudo-observations seront aussi donnés ainsi que des applications pour les copules et autres concepts statistiques.*

Bruno RÉMILLARD, Dép. de mathématiques et d'informatique UQTR  
C.P. 500 Trois-Rivières QC G9A 5H7  
bruno\_remillard@uqtr.quebec.ca

[MS-7]

## A Probability Model for the Continuity Laws

### *Un modèle probabiliste pour les lois de continuité*

René FERLAND\*, Université du Québec à Montréal

Brigitte CHAUVIN, Université du Québec à Montréal

Gaston GIROUX, Université du Québec à Montréal

Fluid mechanic equations are difficult to solve in general. For a fluid whose velocity is function of its density, we can study the problem using a "kinetic" formulation due to (Lions, Perthame et Tadmor, 1994). We present here a probabilistic model relaxing this formula-

tion. The non linear resulting process is described by the equation:

$$\begin{cases} \partial_t f(t, x, \theta) + v(\theta) \cdot \nabla_x f(t, x, \theta) = [\mathbf{1}_{[0, G \star p(r, t)]}(\theta) - f(r, \theta, t)] \\ f(0, x, \theta) = f_0(x, \theta). \end{cases}$$

Our purpose is to approximate uniformly the equation's weak solutions and to control the errors. We describe our results for chaos propagation and for tight bound fluctuations.

*Les équations de la mécanique des fluides sont difficiles à résoudre en général. Pour un fluide dont la vitesse est une fonction de la densité, on peut étudier le problème par le biais d'une formulation "cinétique" (Lions, Perthame et Tadmor, 1994). Nous présentons ici un modèle probabiliste pour la relaxation de cette formulation. Le processus non linéaire résultant est décrit par l'équation :*

$$\begin{cases} \partial_t f(t, x, \theta) + v(\theta) \cdot \nabla_x f(t, x, \theta) = [\mathbf{1}_{[0, G \star p(r, t)]}(\theta) - f(r, \theta, t)] \\ f(0, x, \theta) = f_0(x, \theta). \end{cases}$$

*Notre objectif est d'approcher uniformément les solutions faibles de cette équation et de contrôler les erreurs. Nous décrivons les résultats de propagation du chaos et de fluctuation ("tight bounds") obtenus.*

René FERLAND, Département de mathématiques Université du Québec à Montréal C.P. 8888 Montréal QC H3C 3P8  
ferland@math.uqam.ca

## [MS-8]

### Super-Tree Random Measures

#### Construction de mesures aléatoires par super branchement

Hassan ALLOUBA\*, Duke University

We use supercritical branching processes with random walk steps of geometrically decreasing size to construct random measures. Special cases of our construction give close relatives of the super-(spherically symmetric stable) processes. However, other cases can produce measures with very smooth densities in any dimension.

*Nous utilisons un processus de branchement super critique avec des pas de marche aléatoire de taille géométriquement décroissante pour construire des mesures aléatoires. Des cas spéciaux de notre construction fournissent des éléments de la famille des super processus stables sphériquement symétriques, d'autres cas peuvent produire des mesures à densités très lisses pour n'importe quelle dimension.*

Hassan ALLOUBA, Department of Mathematics P.O. Box 90320 Duke University Durham, N.C. 27708 U.S.A.  
alloba@math.duke.edu

## Session 5

### Survey Methods : Health Surveys

#### Méthodologie d'enquête : Enquêtes de santé

## [MS-9]

### Global Overview of the Data Quality of the National Population Health Survey

#### Un aperçu global de la qualité des données de l'Enquête nationale sur la santé de la population

Yves BÉLAND\*, Statistique Canada

Janine BOUTROS\*, Statistique Canada

The National Population Health Survey is a longitudinal survey which also produces cross-sectional estimates. This double objective creates important methodological challenges. For the 1996-97 wave, a huge sample of additional units was bought by three provinces for the cross-sectional component of the survey (more than twice the sample size of the longitudinal sample). This allows us to look at data quality under a different angle. The following aspects will be discussed:

- How to keep updated the sampled unit address register in order to trace longitudinal respondents.
- Differences between proxy and non-proxy responses in a longitudinal context.
- Response bias caused by telephone collection mode versus personal interviews.
- Data inconsistencies over time.

*L'Enquête nationale sur la santé de la population (ENSP) est une enquête longitudinale qui produit aussi des estimés transversaux. Ce double objectif crée des défis méthodologiques importants. En 1996-97, un ajout important d'unités échantillonales pour le volet transversal (plus de 2 fois l'échantillon longitudinal) nous permet de nous pencher, sous un nouvel angle, sur la qualité des données. Les sujets suivants seront étudiés :*

- La qualité de l'information reliée à la localisation et le dépistage des répondants longitudinaux pour le maintien du registre d'adresses.
- L'impact de la réponse par procuration dans le cadre d'une étude longitudinale.
- Le biais de réponse dû à la collecte par téléphone versus une collecte par personne.
- L'incohérence des données dans le temps.

Yves BÉLAND & Janine BOUTROS, 20-M Édifice R.H. Coats Statistique Canada Ottawa, Ontario K1A 0T6  
belayve@Statcan.ca, Bustros@Statcan.ca

## [MS-10]

### Quality of Data from Direct and by Telephone Interviews in a Survey on Health and Social Security

#### Qualité des données recueillies par entretiens téléphoniques et en face à face dans une enquête sur la santé et la protection sociale

Arié MIZRAHI\*, ARGSES, France

To implement the annual survey on health, social protection, and medicine consumption, names and addresses of households to be interviewed are available, with, in 75% of the households, a phone number. In the absence of his phone number, the household is interviewed directly. To get a reasonable sample size, we consider a survey over

three consecutive years. This involves 7,725 phone interviews (tallying 22,271 individuals) and 2,400 direct interviews (7,508 individuals).

This communication compares the responses of individuals interviewed by phone with those interviewed directly. Many items are investigated, for instance the number of refusals, the number of non-responses, quality of information registered about the individual from a questionnaire administered by a person in charge of the survey (socio-demography, protection when illness occurs, risks,...) and self-administered ("morbidity", medicine consumption—drug store, prescription—). Moreover, 278 households (7,885 individuals) have been discarded from the comparison since they do not have a telephone and exhibit special social characteristics.

*Pour effectuer l'enquête annuelle sur la santé, la protection maladie et la consommation médicale, on dispose des noms et des adresses de ménages à enquêter, et, pour les trois quarts d'entre eux, de leur numéro de téléphone. Lorsqu'on n'a pas le numéro de téléphone, l'enquête est faite en face à face. Pour avoir un échantillon de ménages suffisant, on regroupe trois années d'enquête, on obtient ainsi 7725 enquêtes téléphoniques (22 271 personnes) et 2 400 enquêtes en face à face (7 508 personnes).*

*La communication présente une comparaison des déclarations des personnes selon qu'elles ont été interrogées par téléphone ou en face à face : nombre de refus, de non réponses, quantité d'information relevée par personne dans les questionnaires administrés par enquêteur (socio-démographie, couverture maladie, risques, ...) et auto-administrés (morbidity, consommation médicale — pharmacie, soins de médecin —). Par ailleurs, 278 ménages (7885 personnes) n'ont pas le téléphone et ont des caractéristiques sociales très spécifiques et sont, de ce fait, exclues de cette comparaison.*

Arié MIZRAHI, 34, avenue de la République 92340 Bourg la Reine France  
mizrahi@cnam.fr

## Session 6

### Smoothing, Quality Control and Designs of Experiment

#### *Lissage, contrôle de qualité et schémas expérimentaux*

[MS-11]

#### **Adventures in Functional Linear Modeling** *Aventures dans la modélisation linéaire fonctionnelle*

J. O. RAMSAY\*, McGill University

When each observation is a function or a curve, the possibilities for regression-type models explodes relative to what we are familiar with for multivariate data. This talk will summarize these possibilities, and discuss the role of roughness penalties to force smoothness on the functional parameters and the model values. Some functional linear models for Canadian climate, speech, and handwriting will be offered by way of illustration.

*Quand chaque réponse est une fonction ou une courbe, les possibilités des modèles de type regression abondent relativement à ce que l'on est habitué avec des données multivariées. Cette présentation résumera ces possibilités et discutera le rôle des pénalités pour l'irrégularité, et ce pour forcer le lissage des paramètres fonctionnels et des valeurs du modèle. Quelques modèles linéaires pour le climat canadien, le discours, et l'écriture manuscrite seront offerts pour fins d'illustrations.*

J. O. RAMSAY, Dept. of Mathematics and Statistics Burnside Hall  
McGill University Montreal QC H3A 2T5  
ramsay@ramsay2.psych.mcgill.ca

[MS-12]

#### **Bump Hunting in Regression Revisited** *Recherche de bosses en régression*

Jarek HAREZLAK\*, University of British Columbia

Suppose bivariate data  $\{(t_i, Y_i), i = 1, \dots, n\}$  are observed at times  $a \leq t_1 \leq t_2 \leq \dots \leq t_n \leq b$ . Given a nonparametric regression model  $Y_i = m(t_i) + \varepsilon_i$  with i.i.d. residual with mean 0, we want to estimate the number of modes of the underlying regression function  $m(\cdot)$ . We use the penalized least square technique to get an estimate of  $m(\cdot)$ , i.e. the function minimizing

$$\sum_{i=1}^n (Y_i - m(t_i))^2 + \lambda \int_a^b (D(m(t))) dt$$

where  $\int_a^b D(m(t)) dt$  is a penalty function and  $\lambda$  is a smoothing parameter.

A new test of multimodality is introduced and its performance is studied. Our idea is motivated by the test proposed by Silverman (1981) concerning the number of modes in the density function. He used a "critical bandwidth" as a test statistic in a kernel smoothing context. He noted that if the data are strongly bimodal, we would need a large value of a bandwidth to obtain a unimodal estimate. In our case we define the "critical smoothing parameter"  $\lambda_{crit}$  as the smallest  $\lambda$  giving an estimate with the specified number of modes.

We assess the properties of our test using bootstrap techniques. The simulations are performed using several penalty functions and different regression functions. Comparison is done with Bowman's monotonicity test (1996) in the case of unimodality. The test is also applied to the children's growth data in studying the number of bumps in the derivatives of a growth function.

*Supposons que l'on a observé les données bivariées  $\{(t_i, Y_i), i = 1, \dots, n\}$  où  $a \leq t_1 \leq t_2 \leq \dots \leq t_n \leq b$ . Pour un modèle non-paramétrique de régression  $Y_i = m(t_i) + \varepsilon_i$  où les résidus sont i.i.d. de moyenne nulle, nous désirons estimer le nombre de modes de la fonction de régression  $m(\cdot)$ . Nous utilisons la technique des moindres carrés pénalisés pour obtenir une estimation de  $m(\cdot)$ , i.e. la fonction minimisant*

$$\sum_{i=1}^n (Y_i - m(t_i))^2 + \lambda \int_a^b (D(m(t))) dt$$

où  $\int_a^b D(m(t)) dt$  est une fonction de pénalité et  $\lambda$  est un paramètre de lissage.

*Un nouveau test pour l'unimodalité est introduit et sa performance est étudiée. Le nouveau test s'inspire de celui de Silverman*

(1981) concernant le nombre de modes de la fonction de densité. Il a utilisé une largeur de bande critique comme test statistique dans le contexte de lissage par la méthode du noyau. Il a noté que si les données sont fortement bimodales on a besoin d'une bande assez large pour obtenir une estimation unique du mode. Dans notre cas nous définissons le paramètre critique de lissage  $\lambda_{crit}$  comme le plus petit  $\lambda$  fournissant un estimé avec un nombre spécifié de modes.

Nous évaluons les propriétés de notre test en utilisant les techniques du bootstrap. Des simulations sont réalisées en considérant plusieurs fonctions de pénalité ainsi que de différentes fonctions de régression. La comparaison est effectuée avec le test de monotonie de Bowman (1996) dans le cas de l'unimodalité. Le test est aussi appliqué à des données de croissance d'enfants dans une étude du nombre de bosses de la dérivée de la fonction de croissance.

Jarek HAREZLAK, Department of Statistics, University of British Columbia, Vancouver, B.C.  
jarek@stat.ubc.ca

### [MS-13]

#### Monitoring Process Mean and Variability With One EWMA Chart

*Contrôle de la moyenne et de la variabilité d'un processus à l'aide d'un seul diagramme MMEP*

Gemai CHEN\*, University of Regina

Smiley W. CHENG, University of Manitoba

Hansheng XIE, University of Manitoba

Exponentially weighted moving average (EWMA) control charts are known to be sensitive in detecting small changes in process mean and/or variability. However, if any one of the existing EWMA charts is used and a change is detected, one usually cannot tell whether the change is due to the mean, or due to the variability, or due to both. One solution to this problem is to use a second EWMA chart, usually for the process variability. In this talk, we propose a new EWMA chart which can monitor both the mean and the variability at the same time, and can indicate the source and direction of a change whenever a change is detected.

*Les diagrammes de contrôle de la moyenne mobile exponentiellement pondérée (MMEP) sont reconnus comme étant sensibles à des petits changements de la moyenne d'un processus et/ou de sa variabilité. Par ailleurs, si n'importe lequel des diagrammes de la MMEP existants est utilisé et un certain changement est détecté, on ne peut affirmer avec certitude que ce changement est dû à la moyenne ou à la variabilité ou aux deux. Une solution à ce problème est d'utiliser un second diagramme de la MMEP pour la variabilité du processus. Dans cet exposé, nous proposons un nouveau diagramme de la MMEP qui peut contrôler simultanément la moyenne et la variabilité et indiquer la source et la direction du changement détecté.*

Gemai CHEN, Dept. of Mathematics and Statistics, University of Regina, Regina, Saskatchewan, Canada S4S 0A2  
gchen@math.uregina.ca

### [MS-14]

#### Defect Rate Estimation Using Imperfect Zero-Defect Sampling with Rectification

*Estimation des taux d'articles défectueux par utilisation de l'échantillonnage zéro-un avec rectification*

Pavan WADHWA\*, Mitchell Madison Group

An important aspect of any quality control program is estimation of the quality of outgoing products. This paper applies Acceptance Sampling with rectification to the problem of quality assurance when the inspection procedure is imperfect. The objective is to develop effective rectification sampling plans and estimators based on these plans without making the assumption of a perfect inspection procedure. We develop estimators, under two different sampling plans, for the number of undetected defects remaining after a set of lots has been passed. We compare, by extensive simulation, the proposed estimators with existing ones in terms of Root Mean Squared Error (RMSE). One of our estimators, an empirical Bayes estimator, is seen to consistently obtain substantially lower RMSE overall.

Key Words: Acceptance Sampling, Rectification, Mean Squared Error, Empirical Bayes Estimator, Imperfection Errors

*Un aspect important des programmes de contrôle de qualité est l'estimation de la qualité des produits finis. Ce papier utilise l'acceptation échantillonnale (Acceptance Sampling) avec rectification au problème de l'assurance de la qualité quand les procédures d'inspection sont imparfaites. Le but est de développer des plans de rectification*

*d'échantillonnage et des estimateurs basés sur ces plans sans admettre l'hypothèse de procédure d'inspection parfaite. Nous développons des estimateurs, sous deux plans d'échantillonnage, pour le nombre d'articles défectueux non détectés après qu'un ensemble de lots soit passé. Nous comparons, par des simulations extensives, les estimateurs proposés à ceux existants en terme de racine de la moyenne carrée d'erreur (RMCE). Un de nos estimateurs, un estimateur empirique de Bayes, a une RMCE substantiellement faible relativement à tous les autres estimateurs.*

Pavan WADHWA, 17 Oak Park Ave. Darien CT 06820  
pwadhwa@sprintmail.com

### [MS-15]

#### Friedman Type Tests for Cyclic and Other Incomplete Designs

*Tests de type Friedman pour les plans cycliques et pour d'autres plans incomplets*

Paul CABILLO\*, Acadia University

Mayer ALVO, University of Ottawa

A cyclic design is an incomplete block design which is easily obtained by the cyclic development of an initial block. Rank based methods for the analysis of block designs originate with the test of Friedman for the randomized block design, and were extended by Durbin to the case of balanced incomplete block designs. In this talk we present a unified approach to such analyses which relies on the concept of compatibility to define measures of similarity between incomplete rankings. This approach provides a natural extension of the Friedman-Durbin test to cyclic designs.

*Un plan expérimental cyclique représente un exemple d'un plan d'échantillonnage en blocs incomplets. On peut définir un plan cyclique suite à un développement cyclique d'un bloc initial. Les méthodes d'analyse de données en blocs basées sur les rangs tracent leur origine au test de Friedman pour des blocs complets et au test de Durbin pour des blocs incomplets mais balancés. Dans cette communication, on présente une approche générale qui utilise la notion de compatibilité pour définir des mesures de similarité entre les rangs incomplets. Cette approche qui généralise les tests de Friedman et de Durbin s'applique en particulier aux plans cycliques.*

Paul CABILIO, Dept. of Mathematics and Statistics Acadia University Wolfville, N.S., CANADA B0P 1X0  
paul.cabilio@acadiau.ca

### [MS-16]

#### **Experimental Designs for Orientation Models** *Plans d'expérience pour les modèles d'orientation*

Hwa Shin H. SHIN\*, Queen's University

Chang (Ann Stat, 1986) and Prentice (JRSS B, 1989) developed spherical regression models for matched pairs of direction vectors and matched pairs of spatial orientations respectively. We consider designs for these models in 3 dimensional space and show that the number of matched pairs may be reduced from the values stated by Chang and Prentice for unique estimates of parameters in the models. Optimal designs for the two models based on the minimum number of matched pairs are presented. The designs are based on minimizing the volume of confidence regions and stability of estimates of the parameters in each model.

*Chang (Ann. Stat., 1986) et Prentice (JRSS B, 1989) ont développé des modèles sphériques de régression pour des vecteurs de direction et pour des matrices d'orientation spatiales. Nous considérons des plans d'expérience pour ces modèles dans l'espace à trois dimensions et montrons que le nombre de paires jumelées, nécessaire pour obtenir des estimations bien définies, peut être réduit par rapport aux valeurs données par Chang et Prentice. Des plans d'expérience optimaux pour ces deux modèles basés sur le nombre minimum de paires jumelées sont présentés. Les plans sont basés sur la minimisation de volume des régions de confiance et sur la stabilité des estimations des paramètres dans chaque modèle.*

Hwa Shin H. SHIN, Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada, K7L 3N6  
hwashin@mast.queensu.ca

## Session 7

### **Biostatistics : Statistical Genetics New ideas, New challenges**

*Biostatistique : La statistique génétique, nouvelles idées et nouveaux défis*

### [MS-17]

#### **Mapping Minor Histocompatibility Genes Involved in Graft-versus-Host Disease: A Novel Application of Discordant Sibpairs Methodology**

*Images des gènes presque compatibles intervenant dans la maladie greffe-contre-hôte : une nouvelle application de la méthodologie des paires discordantes*

Kathy LUNETTA\*, Dana-Farber Cancer Institute and Harvard School of Public Health

We introduce a novel application for linkage analysis: using bone marrow donor-recipient sib pairs to search for genes influential in graft-versus-host disease (GVHD), a major complication of allogeneic bone marrow transplantation. In particular, we show that transplant sib pairs in which the recipient developed severe GVHD can be used to map genes in the same way as discordant (affected/unaffected) sib pairs (DSPs). For a plausible GVHD model, we demonstrate that the transplant/DSP analog of the "possible triangle test" (Holmans, 1993) has similar power to that of the simpler "restricted test" proposed by Risch (1990, 1992). Moreover, we show that the restricted test has superior power in much of the DSP possible triangle and significantly inferior power in only a small region. We conclude that the restricted test is preferable for localizing genes with transplant/DSPs. Finally, we examine the effects of heterogeneity on the power to detect GVHD loci.

*Nous introduisons une nouvelle application de l'analyse de linkage génétique. Nous utilisons, lors de greffes de moelle osseuse, des couples donneur-receveur apparentés pour chercher les gènes responsables de la maladie "graft-versus-host" (greffe-contre-hôte ou GCH), qui est une complication grave des greffes de moelle osseuse. En particulier, nous montrons que les paires donneurs-receveurs où le receveur a développé la maladie GCH fournissent les mêmes informations génétiques que les paires discordantes pour la maladie. Pour un modèle plausible de la GCH, nous donnons une version du "possible triangle test" (Holmans ; 1993) qui a une puissance similaire à celle du test restreint proposé par Risch (1990, 1992). En plus, nous montrons que le test restreint est plus puissant dans la plupart du triangle des paires discordantes possibles, et significativement moins puissant sur uniquement une petite région. Nous concluons que le test restreint est préférable pour localiser les gènes pour l'incompatibilité des membres d'une paire. Finalement, nous examinons les effets de l'hétérogénéité sur la puissance pour détecter les sites de la maladie GCH.*

Kathy LUNETTA, Dept. of Biostatistical Sciences Dana-Farber Cancer Institute and Dept. of Biostatistics Harvard School of Public Health Boston, MA 02115 U.S.A.  
lunetta@jimmy.harvard.edu

### [MS-18]

#### **Triangle Constraints in Triangle Affected Sib Pairs Models with Covariates**

*Contraintes triangulaires dans le modèle de fratries affectées avec covariables*

Celia GREENWOOD\*, University of Toronto

For complex diseases, studying affected sibling pairs can lead to simple tests for genetic linkage. If both sibs inherit the same copy of a gene from one parent, that gene is said to be identical by descent (IBD) in the two sibs. Tests for linkage can be obtained by comparing the proportions of sibling pairs who share zero, one or two marker variants (alleles) IBD to the proportions expected when there is no disease gene close by. The power of such tests for linkage is normally improved by constraining the sharing proportions to values which are plausible under many genetic models, a small triangle within the parameter space.

However, the evidence for linkage at one gene location can vary with environmental factors or other genes, and this can be explicitly modelled by using multinomial logistic models for the expected allele sharing proportions. These models can give predicted allele sharing estimates which fall outside the plausible triangle. Methods for choosing and fitting constrained models in the presence of covariates will be described, and the potential effects on power evaluated.

*Pour les maladies complexes, étudier des paires de frères (soeurs) affecté(es) peut aboutir à des tests simples de linkage génétique. Si les deux frères (soeurs) héritent la même copie d'un parent, ce gène est dit identique par descendance (IPD) chez les deux frères (soeurs). Les tests de linkage peuvent être obtenus en comparant les proportions de frères (soeurs) ayant zéro, un ou deux allèles IPD aux proportions estimées en l'absence de maladie génétique. La puissance de tels tests de linkage est normalement améliorée lorsque les proportions d'allèles en commun sont limitées aux valeurs usuellement obtenues par plusieurs modèles génétiques. Ces valeurs forment un petit triangle dans l'espace des paramètres.*

*Par ailleurs, l'évidence de linkage à un site sur un gène peut varier avec les facteurs environnementaux ou avec celle pour d'autres gènes ; ceci peut être explicitement modélisé par des modèles logistiques multinômiaux pour les proportions estimées d'allèles en commun. Ces modèles peuvent fournir des prédictions des estimés des allèles communs lorsque ces estimés tombent à l'extérieur du triangle plausible. Les méthodes pour sélectionner et ajuster les modèles sous contraintes en présence de covariables seront décrites, et les effets potentiels sur la puissance seront évalués.*

Celia GREENWOOD, Preventive Medicine and Biostatistics University of Toronto, 14th floor, McMurrich Building 12 Queen's Park Cres. West Toronto, Ontario M5S 1A8  
celia@genome.ri.mgh.mcgill.ca

## [MS-19]

### **How Genetics Challenges the Ethics of Research** **Comment la génétique défie l'éthique de la recherche**

Charles WEIJER\*, Mount Sinai Hospital, Toronto

In March 1998 researchers in Canada will be subject to novel guidelines found in the Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans. The document is a considerable departure from earlier Canadian guidelines in terms of length, prescriptiveness and attention to the ethical challenges posed by genetic research. The Policy Statement recognizes that the impact of genetic information may extend beyond the individual research subject to family members or the community and recommends that "informed

choice shall also invite those social structures, as far as is practical and possible." The practical implications of this requirement have yet to be determined. The conventional framework for the analysis of ethical problems in research, as set out in the Belmont Report, focuses on individual research subjects (and groups only insofar as they may be wrongfully included in or excluded from research). Genetic studies involving multi-generational families bring into question the generalizability of norms established for individual research subjects. For example, the Policy Statement expresses concern about the fact that "members within a family may be coerced by other members to join the study." The appropriateness of applying the language of coercion to relations among family members can be questioned. Genetic studies involving entire communities, e.g., Ashkenazi Jews and mutations in the APC gene, raise the possibility of harm to the community as a whole. Are additional protections, above and beyond those afforded to individual research participants, required for communities in research? The Policy Statement suggests a variety of possible additional protections, but considerable obstacles lie in the way of their implementation.

*En Mars 1998, les chercheurs au Canada seront assujettis à de nouvelles lignes de conduite apparaissant dans la déclaration politique du conseil tripartite sur l'éthique de la recherche comportant des humains. Ce document est un changement considérable par rapport aux directives précédentes, que ce soit en terme de longueur, de présentation ou d'attention portée aux défis posés par la recherche génétique. La déclaration reconnaît que non seulement les sujets, mais également leurs familles et la communauté en général sont concernés par l'information génétique ; elle recommande que "les responsables de la recherche fassent participer ces intervenants sociaux dans la mesure du possible". Les implications pratiques de ces recommandations restent à déterminer.*

*Le cadre de travail conventionnel pour l'analyse des problèmes d'éthique dans la recherche, tel que rapporté dans le rapport de Belmont, s'intéresse aux individus servant de sujets à la recherche (et aux groupes seulement dans la mesure où ils sont inclus ou exclus de la recherche pour des mauvaises raisons). Les études génétiques portant sur des générations multiples de familles soulèvent la question de la généralisation des normes établies pour les individus. Par exemple, la déclaration s'inquiète du fait que certains membres de la famille pourraient être contraints de participer à l'étude. La légitimité d'utiliser un langage de menace parmi les membres de la famille est à discuter. Les études génétiques portant sur toute la communauté, e.g., les juifs Ashkenazi et les mutations des gènes "APC", soulève la possibilité de causer du tort à toute une communauté. Est-ce que des protections supplémentaires, au delà de celles garanties pour les individus participant à la recherche, sont nécessaires pour les communautés ? La déclaration politique suggère une multitude de protections supplémentaires possibles, mais des obstacles considérables entravent leurs mise en place.*

Charles WEIJER, Executive Offices, Mount Sinai Hospital 600 University Avenue Toronto, ON M5G 1X5  
cweijer@mtsinai.on.ca

## Session 8

### Biostatistics : Risk Analysis and Related -

## Topics

### *Biostatistique : Analyse de risques et sujets connexes*

#### [MS-20]

#### **The Statistics Function in a Canadian Health Protection Agency**

#### *Le rôle de la statistique au sein d'un organisme canadien voué à la protection de la santé*

S. BARTLETT\*, Health Canada

D. MACLEOD, Health Canada

B. MORI, Health Canada

E. ORMSBY, Health Canada

L. PAPATHANASOPOULOU, Health Canada

M. RAHMAN, Health Canada

W.H. ROSS, Health Canada

Statistics plays an important role in the quantitative design, analysis and interpretation of information in health protection programs. The Health Protection Branch (HPB) of Health Canada has a number of statisticians who work in the food, environmental health and therapeutics programs and in the Laboratory Centre for Disease Control. A team of HPB statisticians and nonstatisticians recently reviewed their roles and functions to determine if statistical resources are effectively utilized in supporting program activities and if they are adequate for the future needs of the agency. This paper highlights some of the findings.

*Les statistiques jouent un rôle important dans la façon d'établir les plans d'étude, d'analyser et d'interpréter quantitativement l'information concernant les programmes de protection de la santé. La Direction générale de la protection de la santé (DGPS) de Santé Canada emploie un certain nombre de statisticiens qui travaillent dans le domaine des aliments, de l'hygiène du milieu et des programmes thérapeutiques de même qu'au Laboratoire de lutte contre la maladie. Une équipe de statisticiens et de non-statisticiens de la DGPS ont examiné récemment leur rôle et leurs fonctions afin de déterminer si les ressources statistiques sont utilisées efficacement pour appuyer les activités de programmes et si elles sont suffisantes pour répondre aux besoins futurs de l'organisme. Le présent article expose certaines de leurs conclusions les plus importantes.*

Sheryl BARTLETT, A/Chief, Biostatistics Division Environmental Health Directorate Health Canada PL 0801B2 Tunney's Pasture Ottawa, Ontario K1A 0L2

Sheryl\_bartlett@hc-sc.gc.ca

#### [MS-21]

#### **Determination of Toxicity Health States for Interferon Maintenance Treatment in Multiple Myeloma and 4-Health-State Q-TWiST Evaluation of Interferon Effectiveness**

#### *Détermination des niveaux d'états de santé dûs à la*

#### *toxicité pour le traitement par interféron dans l'évaluation de l'efficacité de l'interféron pour des mélanomes multiples selon 4 états de santé et par le TAQSST*

Tong LI\*, Queen's University

Benny ZEE, Queen's University

Keith JAMES, Queen's University

Michael BRUNDAGE, Queen's University

Background: A previous study on the effect of interferon (IFN) maintenance therapy using quality-adjusted time without symptoms and toxicity (Q-TWiST) method was performed (Zee et al., ASCO 1997). In order to apply the results of randomized trials to the individual patient thus enabling an informed treatment decision, a clear description of toxicity (Tox) health states is critical. Methods: In this study, two methods were used to classify the IFN toxicity health states for patients: 1) a separate survey of clinicians who treated myeloma patients; 2) correspondence analysis (CA) of the toxicity data from patients' diaries of a randomized trial. The toxicities considered in both the clinicians' survey and the CA included fever, muscle aches, headache, nausea, vomiting, chills, diarrhea, and fatigue. A 4-health-state Q-TWiST model, which includes two Tox health states, TWiST, and duration after relapse, was then used to compare the results of the two approaches and to evaluate the effectiveness of IFN maintenance treatment. Results: Similar toxicity classification results were obtained from the survey and the CA. For the clinicians' survey, among the 43 clinicians who had participated in the previous NCIC CTG multiple myeloma study, 29 replied to the survey. The results of the survey gave two toxicity groupings. The first group includes all mild toxicities except for vomiting, and the second group includes all other moderate and severe toxicities plus any grade of vomiting. Alternatively, two health states, similar to the results of the survey with only one minor difference, were identified by CA. The first group contains all mild toxicities except vomiting, plus moderate fatigue; the second group contains any grade of vomiting plus the rest of moderate or worse toxicities. The average time spent in the Tox health states defined by the clinicians' survey are 6.31 and 3.81 months respectively. The average time spent in the two Tox health states for the CA are 8.13 and 1.99 months respectively. A Q-TWiST analysis using a four health states model gave a slightly more conservative result in the clinicians' survey than that of the CA. However, both methods produce a similar conclusion that IFN is preferable with a wide range of utility coefficients. Conclusion: The division of the toxicity duration into two shorter health states with clearly different utilities results in a 4-health-state model which reduces the complication and the potential heterogeneous utilities of the Tox health state. The Q-TWiST analyses support the use of IFN in multiple myeloma patients over a wide range of plausible utility values assigned to periods with toxicity or disease relapse.

*Une étude antérieure de l'effet de l'interféron (IFN) comme agent de thérapie de maintenance utilisant le temps ajusté selon la qualité sans symptômes ni toxicité (TAQSST) a été appliquée (Zee et al., ASCO 1997). Pour appliquer les résultats des essais cliniques à un patient, lui permettant ainsi de prendre une décision éclairée, une description claire de l'état de santé liée à la toxicité (Tox) est pertinente. Les méthodes : Dans cette étude, deux méthodes sont utilisées pour classer l'état de santé dû à la toxicité par IFN chez les patients :*

1) un sondage auprès des cliniciens qui traitaient des malades ayant des mélanomes ; 2) une analyse des correspondances (AC) des données sur la toxicité tirées aléatoirement des agendas des patients. Les toxicités considérées tant dans le sondage auprès des cliniciens que dans l'AC comportent des données sur la fièvre, les douleurs musculaires, les maux de tête, les nausées, le vomissement, les frissons, la diarrhée et la fatigue. Un modèle pour les 4 états de santé et pour le TAQSST qui comporte deux états de santé dus à la Tox, le TAQSST, et la durée après rechute, a été utilisé pour comparer les résultats des deux approches et pour évaluer l'efficacité de la maintenance par traitement par l'IFN. Résultats : Des résultats de classification de toxicité similaires ont été obtenus par sondage et l'AC. Pour le sondage auprès des cliniciens, parmi les 43 participants à l'étude précédente du mélanome "NCIC CTG", 29 ont répondu au sondage. Les résultats du sondage permettent de former deux groupes ; le premier groupe comporte toutes les toxicités faibles à l'exception du vomissement, le second groupe comporte toutes les autres toxicités sévères et modérées et tous les degrés de vomissements. Alternative-ment, l'AC fournit des résultats similaires aux résultats du sondage avec deux états de santé, mais permet d'identifier une petite différence ; le premier groupe contient toutes les toxicités à l'exception du vomissement, en plus d'une fatigue modérée ; le second groupe contient tous les degrés de vomissement en plus de toxicité accentuée ou modérée. Les temps moyens, passés dans les états de santé dus à la Tox, définis par le sondage sur les cliniciens est de 6.1 et 3.81 mois respectivement. Les temps moyens par l'AC est de 8.31 et 1.99 mois respectivement. Une analyse par le TAQSST utilisant un modèle à 4 états de santé fournit des résultats un peu plus conservateurs pour le sondage sur les cliniciens que par l'AC. Par ailleurs, les deux méthodes produisent la même conclusion que l'IFN est préférable avec une plage assez large pour des coefficients d'utilité. Conclusion : La subdivision de la durée de toxicité en deux états de santé plus courts avec des utilités clairement distinctes résulte en un modèle à 4 états de santé qui réduit la difficulté et les utilités potentiellement hétérogènes de l'état de santé dû à la toxicité. Les analyses par TAQSST supportent l'utilisation de l'IFN pour les patients présentant des mélanomes multiples sur une plage assez large pour des valeurs plausibles d'utilité assignées aux périodes avec une toxicité ou une rechute due à la maladie.

Tong LI, Dept. Comm Hlth & Epid, National Cancer Institute of Canada Clinical Trials Group (NCIC-CTG), Queen's University, Kingston ON K7L 3N6  
4t13@q1ink.queensu.ca

#### [MS-22]

### The Statistical Modelling of Micronutrient Interactions in Nutrient Risk Assessments: The Effect of Calcium on Iron

*L'établissement de modèles statistiques des interactions des micronutriments dans l'évaluation du risque asso-*

### *cié aux nutriments : l'effet du calcium sur le fer*

W.H. ROSS\*, Health Canada  
K. COCKELL, Health Canada  
S. DUBOIS, Health Canada  
M. L'ABBE, Health Canada  
B. JUNKINS, Health Canada  
D. MILLER, Health Canada

As part of a nutrient risk assessment, an attempt is made to establish a dose response relationship between the level of exposure and identified adverse health effects. This model is then used to assist in the overall assessment of risks and benefits associated with the addition of micronutrients to foods. There are a number of challenges to statistical modelling in this context, related to the integration of widely diverse sources of information and data. These include mechanistic stochastic models for the underlying kinetics, meta-analysis of relevant published literature, and the statistical analysis of nutrition surveys. This process is illustrated by modelling the inhibitory effect of calcium on iron absorption.

*Dans le cadre d'une évaluation du risque nutritionnel, les auteurs ont tenté d'établir une relation dose-effet entre le niveau d'exposition et les effets indésirables sur la santé. Ce modèle est ensuite utilisé pour l'évaluation générale des risques et des avantages associés à l'ajout de micronutriments aux aliments. L'établissement de modèles statistiques dans ce contexte présente un certain nombre de difficultés, liées à l'intégration de sources très variées d'information et de données, dont des modèles mécanistes/stochastiques pour la cinétique sous-jacente, la méta-analyse des publications pertinentes et l'analyse statistique d'enquêtes sur la nutrition. Cette méthode est illustrée par une modélisation de l'effet inhibiteur exercé par le calcium sur l'absorption du fer.*

W.H. ROSS, Health Canada Banting Bldg. (AL# 2203B) Tunneys Pasture Ottawa, Ont. K1A 0L2  
William\_Ross@hc-sc.gc.ca

#### [MS-23]

### Statistical Issues in the Analysis of Parallel Time Series of Health and Environmental Variables

*Débat sur l'analyse statistique des séries chronologiques des variables décrivant des problèmes de santé et des variables environnementales*

Sabit ÇAKMAK\*, Health Canada  
R. BURNETT, Health Canada  
D. KREWSKI, Health Canada

Time series of daily administrative cardio-respiratory health and environmental information have been extensively used to assess the potential public health impact of ambient air pollution. Both series are subject to strong but unrelated temporal cycles. These cycles must be removed from the time series prior to examining the role air pollution plays in exacerbating cardio-respiratory disease. In this paper, we examine a number of methods of temporal filtering that



have been proposed to eliminate such temporal effects. The techniques are illustrated by linking the number of daily admissions to hospital for respiratory diseases in Toronto, Canada for the 11 year period 1981 to 1991 with daily concentrations of ambient ozone. The ozone-hospitalization relationship was found to be highly sensitive to the presence or absence of temporal adjustments for seasonal and day of the week effects, although the statistical method of adjustment was not a critical factor. The association between ambient ozone levels and hospital admissions for respiratory diseases was also sensitive to the season of examination, with minimal effects observed outside the summer months. It is recommended that both the health and environmental time series be adjusted for long term trends, seasonal and subseasonal patterns, and day of the week effects. Season specific analyses should also be considered.

*Les séries chronologiques journalières de données administratives concernant la santé cardio-respiratoire et l'information environnementale ont été utilisées de manière extensive pour évaluer l'impact de la pollution dans l'air ambiant sur la santé publique. Ces deux séries chronologiques sont assujetties à des cycles temporaires forts mais indépendants. Ces cycles doivent être éliminés de la série temporelle avant d'examiner le rôle que joue la pollution de l'air ambiant dans l'exacerbation des maladies cardio-respiratoires. Dans ce papier, nous examinons un certain nombre de méthodes de filtrage temporel, proposées auparavant, pour éliminer ces effets "temporels". Les techniques sont illustrées en reliant journalièrement le nombre d'entrées à l'hôpital pour traitement de maladies respiratoires à Toronto, Canada, pour une période des 11 ans, s'étalant de 1981 à 1991, aux concentrations d'ozone dans l'air ambiant. La relation entre l'hospitalisation et la concentration d'ozone s'est avérée très sensible à la présence d'ajustements temporels des effets des saisons et des jours de la semaine, alors que la méthode statistique d'ajustement n'est pas un facteur critique. L'association des niveaux d'ozone ambiant aux admissions aux hôpitaux pour des maladies respiratoires à été aussi sensible à la saison où a lieu l'examen médical; les effets observés hors des d'été sont minimes. Il est recommandé que les séries temporelles de la santé et celles de l'environnement soient ajustées pour les tendances à long terme, pour les composantes saisonnières et sous saisonnières, et pour les effets journaliers. Des analyses spécifiques aux saisons devraient aussi être envisagées.*

Sabit CAKMAK, Research Scientist Environment Health Directorate Health Protection Branch Health Canada 203 Environmental Health Center Tunney's Pasture Ottawa, Ont., K1A 0L2  
scakmak@ehd.hwc.ca

[MS-24]

### **A Quantal Isobologram Model to Identify Joint Action for Chemical Mixtures**

*Un modèle quantitatif isobologramme pour identifier l'action conjointe de mélanges chimiques*

Ding G. CHEN\*, Department of Fisheries and Oceans

The isobologram model is a commonly used and powerful graphical and statistical tool for analyzing the joint action for simple chemical mixtures. The typical experiment will involve to administer a

set of doses from two or more substances (such as drugs, chemicals and toxins) to some experimental subjects and the responses (such as mortalities, LDH releases) are measured. Substantial research has been done for the quantitative response and the amount of research in the qualitative framework is minuscule compared to the quantitative framework. In the paper, isobologram model is proposed to analyze the joint action of chemical mixtures for quantal dose-response relationship based on the generalized linear model technique. A *link function* is used to associate the systematic component of dose-response relationship, defined by *linear predictor* as

$$\eta = \beta_0 + \beta_1 d_1 + \beta_2 d_2 + \beta_{12} d_1 d_2$$

to the random component classified by the distribution of the response, such as binomial distribution. Model parameters can be estimated by the maximum likelihood estimation and then to be used to construct the isobologram so that the joint action from the chemicals can be identified both by the isobologram and the statistical inference for interaction parameter  $\beta_{12}$ .

**Key Words:** Chemical Mixtures; Risk Assessment; Toxicology; Maximum Likelihood Estimation; Quantal Dose-Response Modeling; ED<sub>100p</sub> Isobols.

*Le modèle isobologramme est un moyen graphique et statistique puissant souvent utilisé pour analyser l'action commune de mélanges chimiques simples. L'expérience type consiste à administrer un certain nombre de doses fabriquées à partir de deux ou plusieurs substances (telles que des médicaments, des produits chimiques et des toxines) à des sujets expérimentaux et les réponses (telles que la mortalité, la dose "LDH") sont mesurées. La recherche concernant les modèles pour réponses quantitatives est substantielle, par contre celle menée dans le cadre des réponses qualitatives paraît minuscule. Dans ce papier, le modèle isobologramme est proposé pour analyser l'action conjointe des mélanges chimiques en reliant la quantité de dose à la réponse via la technique des modèles linéaires généralisés. Une fonction de lien est utilisée pour associer la composante systématique de la relation dose-réponse, définie par le prédicteur linéaire*

$$\eta = \beta_0 + \beta_1 d_1 + \beta_2 d_2 + \beta_{12} d_1 d_2$$

à la composante aléatoire classifiée par la distribution de la réponse, e.g. une distribution binômiale. Les paramètres du modèle peuvent être estimés par la méthode du maximum de vraisemblance et ensuite utilisés pour construire un isobologramme pour que l'action conjointe des produits chimiques puisse être identifiée par l'isobologramme et par le paramètre d'interaction  $\beta_{12}$  du modèle linéaire considéré.

Ding G. CHEN, Department of Fisheries and Oceans Pacific Biological Station Nanaimo, B.C. V9R 5K6, Canada  
chend@china.pbs.dfo.ca

[MS-25]

### **Stochastic Modelling of HIV/AIDS Patients Migration: Application to Atlantic Canada HIV/AIDS Migration Pilot Study**

*Modélisation stochastique de la migration d'individus atteints du HIV/SIDA : Application à une étude pilote dans les provinces atlantiques concernant la migration*

***des individus atteints du HIV/SIDA***

Swarna WEERASINGHE\*, Dalhousie University

Lamont SWEET, Dalhousie University

Diane MACDONALD, Dalhousie University

Andrea SCOTT, Dalhousie University

Jerrold YATMAN, Dalhousie University

Lynn JOHNSTON, Dalhousie University

Characterizing migration patterns of people with infected disease is crucial for planning health, community and social services needs. In this paper we consider two types of models; probabilistic and conventional regression type models. Probabilistic models developed to explain migration pattern takes the form of Markov or Semi-Markov processes (Ginsberg, 1971). Extension to static migration models under random utility theory, developed by social scientist, to incorporate causal structure of HIV/AIDS patients is discussed. Possible regression and time series type models to explain individual patient and regional parameters are also reviewed. Merits and limitations of both methods are discussed. Methods presented in this paper incorporate disease progression, health care needs and high risk sexual behavior in both spatial interaction models (Stillwell, 1991) and discrete choice models (Maier and Weiss, 1991). Atlantic Canada HIV/AIDS migration pilot study data will be used in the analysis.

*La caractérisation de la forme de migration des personnes infectées de maladies est cruciale pour la planification concernant la santé et les services sociaux dans les communautés. Dans ce papier, nous considérons deux types de modèles; des modèles probabilistes et des modèles de type régression conventionnelle. Les modèles probabilistes développés pour expliquer les formes de migration prennent la forme de processus de Markov ou de Semi-Markov (Ginsberg, 1971). L'extension aux modèles migratoires statiques dans le cadre de la théorie de fonction d'utilité aléatoire, développée en sciences sociales, pour incorporer les structures causales des individus infectés par l'HIV/SIDA est discutée. Des modèles de régression et de séries chronologiques pour expliquer les paramètres spécifiques tant aux individus qu'aux régions sont aussi présentés. Les mérites et les limitations de ces deux méthodes sont discutés. Les méthodes présentées dans ce papier incorporent la progression de la maladie, les soins de santé, et les risques forts dus au comportement sexuel dans les modèles d'interaction spatiale (Stillwell, 1991) et dans les modèles de choix discrets (Maier and Weiss, 1991). Des données de l'étude pilote de migration d'individus atteints du HIV/SIDA au Canada Atlantique sera utilisée dans cette analyse.*

Swarna WEERASINGHE, Dalhousie University P.O. Box 1000, Halifax, NS B3J 2X4

sweerasi@tupdean1.med.dal.ca

**Session 9****Probability : Limit Theorems for Multiparameter and Set-Indexed Processes*****Probabilité : Théorèmes limites pour des******processus multiparamètres indicés par des ensembles*****[MS-26]****Functional LIL for a Multiparameter Martingale*****La loi du logarithme itéré pour une martingale à plusieurs paramètres***

André DABROWSKI\*, University of Ottawa

We consider lattice-indexed random variables where the variables form strong martingale difference array. We demonstrate that a functional law of the iterated logarithm holds under appropriate conditions on the sum over rectangles of the differences between the unconditional and conditional second moments of the variables. The key tool is a strong approximation of the difference array by an array of independent normal variables.

*Nous considérons une suite de variables aléatoires indicée par un treillis où les variables sont des différences de martingale, au sens fort. Nous démontrons que la loi du logarithme itéré reste valide sous certaines conditions concernant les sommes, sur les rectangles, des différences entre les moments d'ordre deux inconditionnels et conditionnels des variables. L'idée clé est une approximation forte des différences par des variables aléatoires indépendantes normalement distribuées.*

André DABROWSKI, Mathematics and Statistics University of Ottawa 585 King Edward Station A, P.O. Box 450 Ottawa, Ontario K1N 6N5

ards@uottawa.ca

**[MS-27]****Poisson Convergence for a Set-Indexed Empirical Process*****La convergence de Poisson pour un processus empirique indicé par des ensembles***

Gail IVANOFF\*, University of Ottawa

Martingale methods may be extended to set-indexed processes, and in particular a compensator may be defined for a set-indexed submartingale. Just as in the classical case of point processes on the positive real numbers, convergence of the compensators of a sequence of (set-indexed) point processes to a continuous deterministic limit ensures weak convergence of the sequence to a Poisson process. This result may be applied to the set-indexed empirical process on  $\mathbf{R}^n$ .

*Les méthodes utilisées pour les martingales peuvent être étendues aux processus indicés par des ensembles, plus particulièrement un compensateur peut être défini pour une sous-martingale indicée par des ensembles. Exactement comme dans le cas classique des processus de points sur la demi droite réelle positive, la convergence des compensateurs d'une suite de processus de points (indicés par des ensembles) vers une limite déterministe assure la convergence faible de la suite vers un processus de Poisson. Ce résultat peut être appliqué aux processus empiriques indicés par des ensembles sur  $\mathbf{R}^n$ .*

Gail IVANOFF, Mathematics and Statistics University of Ottawa 585 King Edward Station A, P.O. Box 450 Ottawa, Ontario K1N 6N5  
bgisg@matrix.cc.uottawa.ca

**[MS-28]**

**A Central Limit Theorem for Set-Indexed Martingales**

*Un Théorème de la Limite Centrale pour les martingales indicées par des ensembles*

Dean SLONOWSKY\*, University of Ottawa

After presenting the necessary set-indexed framework, we give conditions under which a sequence of set-indexed strong martingales converges in finite dimensional distribution to a set-indexed Brownian motion. Conditions ensuring functional CLT's are also discussed. Applications include a CLT for a Gaussian bootstrap.

*Après la présentation du cadre de travail, nous donnons les conditions sous lesquelles une suite de martingales fortes, indicée par des ensembles, converge, en loi de dimension finie, vers un mouvement Brownien indicé par des ensembles. Les conditions assurant le Théorème de la Limite Centrale fonctionnel sont aussi discutées. Les applications portent sur un Théorème de la Limite Centrale pour un bootstrap Gaussien.*

Dean SLONOWSKY, Mathematics and Statistics University of Ottawa 585 King Edward Station A, P.O. Box 450 Ottawa, Ontario K1N 6N5

Session 10

Survey Methods : Administrative Data

*Méthodologie d'enquête : Utilisation de données administratives*

**[MS-29]**

**Use of Administrative Records in the Canadian Survey of Employment, Payrolls and Hours**

*L'utilisation des rapports administratifs dans l'enquête sur l'emploi canadien, la rémunération et les heures*

Éric RANCOURT\*, Statistique Canada

Mike HIDIROGLOU, Statistique Canada

The Survey of Employment, Payrolls and Hours is a monthly establishment survey covering most industrial activities in Canada. Starting in 1993, two additional data items (payrolls and employment) for Payroll Deduction accounts from Revenue Canada have become available. A major redesign was undertaken to use this wealth of information in order to reduce the sample size, the associated costs and the response burden, while maintaining the quality of the estimates. To achieve these goals, administrative records are used in a variety of ways. These include the creation of a survey frame for part

of the survey; the building of an improved design; the use in regression of the independent variables obtained from the payrolls source of Revenue Canada and quality assessment of the surveyed portion. This paper gives an overview of the redesign and presents the methodological aspects of the use of administrative records in the Survey of Employment, Payrolls and Hours.

*L'enquête sur l'emploi, la rémunération et Les heures est une enquête mensuelle des établissements couvrant la majorité des activités industrielles au Canada. Commencant en 1993, deux items de données additionnels (la rémunération et l'emploi) sur les comptes des déductions sur la rémunération de Revenu Canada sont devenus disponibles. Une remaniement majeur de l'enquête a été entrepris afin d'utiliser cette source d'information pour réduire la taille de l'échantillon, le coût associé et la charge des réponses, tout en préservant la qualité des estimations. Pour réaliser cet objectif, les rapports administratifs sont utilisés de plusieurs manières : La création d'un plan d'enquête pour une partie de la population sondée, ensuite la construction d'un plan amélioré, l'utilisation dans la régression des variables indépendantes obtenues de la source de rémunération de Revenu Canada ainsi que l'évaluation de la qualité de la portion enquêtée. Ce papier revoit cette replanification et présente les aspects méthodologiques concernant l'utilisation des rapports administratifs pour l'enquête sur l'emploi, la rémunération et les heures.*

Éric RANCOURT, Division des méthodes d'enquêtes sociales Statistique Canada Tunney's Pasture Ottawa, Ontario K1A 0T6  
rancour@statcan.ca

**[MS-30]**

**Administrative Data and Non-Observed Heterogeneity: A Permanent Union for an Efficient Longitudinal Study**

*Données administratives et hétérogénéité non observée : Un mariage sans divorce pour une analyse longitudinale efficace*

Abdellatif DEMNATI\*, Statistique Canada

J.-F. BEAUMONT, Statistique Canada

Longitudinal analysis, and in particular duration analysis, require a long study period, a high number of responses and little censoring. Because of these requirements, administrative data offer, in relation to surveys, an interesting alternative at a moderate cost. This document first describes how the Small Area and Administrative Data Division of Statistics Canada was able to successfully use administrative data in order to create a longitudinal administrative database. This database, which is updated annually, is increasingly used to produce longitudinal analyses on the income dynamics of the Canadian population. The use of administrative data depends on a number of issues, content being one. Some variables may be missing or simply unobservable. The absence of predictive variables may be compensated for, in part or entirely, by the introduction of random variables into the model. Modelling unobserved or missing variables into the sample is the second objective of this document. The document initially offers an introduction to the problem of unobserved heterogeneity and a presentation of the statistical tools needed for modelling and estimation. Discussions and comments on the methodology are also provided.

The authors would like to dedicate this work to families and childhood families of those who contributed to the creation and/or the development of the Longitudinal Administrative Databank. This work was made possible thanks to the Small Area and Administrative Data Division.

*L'analyse longitudinale, et particulièrement l'analyse des durées, nécessite à la fois une longue période d'étude, un nombre élevé de réponses et peu de censures. À cause de ces exigences, les données administratives offrent une alternative intéressante à un coût modeste par rapport aux enquêtes. Ce document montre en premier lieu comment la division des données régionales et administratives de Statistique Canada a réussi à recycler avec succès des données administratives pour créer une base de données administratives longitudinales. Cette base de données, dont la mise à jour est dynamique dans le temps, sert de plus en plus à produire des analyses longitudinales importantes portant sur la dynamique du revenu de la population canadienne.*

*L'utilisation des données administratives dépend d'un certain nombre de facteurs dont le contenu. Certaines variables peuvent être manquante ou tout simplement non observable. L'absence de variables prédictives importantes peut être compensée en totalité ou en partie par l'introduction de variables aléatoires dans le modèle. La modélisation des variables non observables ou manquantes dans l'échantillon est l'objectif du présent document. Le document comporte d'abord une motivation du sujet par des exemples et une présentation des outils statistiques nécessaires pour la modélisation et l'estimation. Des discussions et commentaires sur la méthodologie utilisée sont données.*

*Les auteurs tiennent à dédier ce travail aux familles et aux familles de l'enfance de ceux et celles qui ont contribué à la création et/ou au développement de la Banque de données administratives longitudinales. Ce travail est rendu possible grâce à la Division des données régionales et administratives de Statistique Canada.*

Abdellatif DEMNATI, Division des méthodes d'enquêtes sociales Statistique Canada Tunney's Pasture Ottawa, Ontario K1A 0T6  
DEMNA BD@statcan.ca

### [MS-31]

#### **Impacts of the Use of a Mixture Two Way Response on Transversal and Longitudinal Estimation of Income**

#### *Impacts de l'utilisation d'un mode de collecte mixte sur les estimations transversales et longitudinales du revenu*

Martin RENAUD\*, Statistique Canada

Michel LATOUCHE, Statistique Canada

The survey on the dynamics of labor and income is a longitudinal survey of households. It aims to measure level changes of the economic welfare of individuals and the factors that may influence these changes. The sample is made of 30 000 households interviewed twice a year for six years. In order to reduce the response burden and errors, the survey uses a mixture of two way response in which the members of the sample have to choose to give information on their income themselves or allow Statistics Canada to get this information directly from Revenue Canada. This study aims to assess the impacts of using two sources of data on the estimation of the levels, the ten-

dencies, and the measures of the lowest income as well as making comparisons among some targeted groups of people.

*L'enquête sur la dynamique du travail et du revenu (EDTR) est une enquête longitudinale menée auprès des ménages. L'EDTR vise à mesurer les changements au niveau du bien-être économique des individus et les facteurs qui peuvent influencer ces changements.*

*L'échantillon est composé de 30 000 ménages interviewés deux fois par année pendant six ans. Afin de réduire le fardeau et les erreurs de réponses, l'enquête utilise un mode de collecte mixte dans lequel les membres de l'échantillon ont le choix entre donner eux-mêmes l'information sur leurs revenus ou permettre à Statistique Canada d'obtenir cette information directement de Revenu Canada. Cette étude vise à évaluer les impacts de l'utilisation de deux sources de données sur les estimations de niveaux, de tendances, les mesures de faible revenu ainsi que sur les comparaisons entre certains groupes cibles de gens.*

Martin RENAUD, Division des méthodes d'enquêtes entreprises Statistique Canada Tunney's Pasture Ottawa, Ontario K1A 0T6  
renamar@statcan.ca

## Session 11

### Extreme Values and Time Series

#### *Valeurs extrêmes et séries chronologiques*

### [MS-32]

#### **Exceedance over High Thresholds: A Guide to Threshold Selection**

#### *Modélisation des grandes valeurs : Un guide pour la sélection du seuil*

D.J. DUPUIS\*, Dalhousie University

The modeling of exceedances over high thresholds is considered. The natural distribution for such exceedances, the generalized Pareto distribution (GPD), is used and the problematic issue of threshold selection is addressed. We fit the GPD robustly to the data using techniques based on optimal bias-robust estimates. The robust procedure will assign weights between 0 and 1 to each data point. These weights are used to assess the validity of the GPD model for exceedances of the proposed threshold and thus can guide threshold selection. That is, we can initially consider a low threshold and increase it (thus reducing the number of data points) until all weights are close to one. The new approach is used to analyze two of the NERC data sets.

*La modélisation des valeurs au-delà d'un seuil est considérée. La distribution naturelle des écarts au seuil, la distribution de Pareto généralisée (DPG), est utilisée ; on traite du problème de la sélection du seuil. Nous ajustons la DPG de façon robuste utilisant des techniques basées sur les estimateurs à biais robustes optimaux. La procédure robuste donnera des poids entre 0 et 1 à chaque donnée. Ces poids sont utilisés pour évaluer la validité du modèle DPG pour les excédents du seuil proposé et peut ainsi guider la sélection du seuil.*

*C'est-à-dire, nous pouvons considérer un premier seuil assez bas et nous l'augmentons graduellement (réduisant le nombre de données dans l'analyse) jusqu'à ce que tous les poids soient près de 1. Cette nouvelle approche est illustrée sur deux jeux de données NERC.*

D.J. DUPUIS, Dept. of Engineering Mathematics DalTech, Dalhousie University P.O. Box 1000, Halifax, NS B3J 2X4 Ph. 902-420-7825; Fax 902-423-1801  
dupuis@tuns.ca

### [MS-33]

#### **Estimation Procedure for the Dependence Function of Extreme Value Copulas**

#### *Méthode d'estimation de la fonction de dépendance des copules de valeurs extrêmes*

Julie JACQUES\*, Université Laval

The distribution of bivariate extreme values are characterised by their marginals and by a dependence function of the extreme value copula. Many parametric and nonparametric procedures have been proposed to estimate this function. In this presentation, we will talk about three of these nonparametric methods. We will present comparisons of these estimators based on simulations and data analysis.

*Pour comprendre le comportement des valeurs extrêmes, il est nécessaire de s'intéresser à leur modélisation. Lorsque les valeurs extrêmes sont unidimensionnelles, l'ensemble de leurs lois possède une structure paramétrique. Par contre, lorsque les valeurs extrêmes sont bidimensionnelles, ces lois sont caractérisées par leurs marginales et par une fonction appelée fonction de dépendance de la loi bivariable des valeurs extrêmes. Plusieurs méthodes paramétriques et non paramétriques ont été proposées pour estimer cette fonction. Dans cette présentation, nous nous intéressons principalement à trois de ces méthodes non paramétriques. Nous comparerons ces estimateurs grâce à différentes simulations qui seront effectuées ainsi qu'à l'analyse d'un ensemble de données.*

Julie JACQUES, Dép. mathématiques et statistique Université Laval  
Sainte-Foy QC G1K 7P4  
jjacques@mat.ulaval.ca

### [MS-34]

#### **On a Lag-Window Estimator of the Covariances of Sample Autocovariances for a Strongly Mixing Process**

#### *Sur un estimateur à délai pondéré des covariances des autocovariances échantillonnées pour un processus stationnaire fortement mélangeant*

Daniel MOREAULT\*, Université de Montréal

Under some regularity conditions for strongly mixing stationary processes, Berlinet and Francq (1998) have proposed recently a consistent lag-window estimator of the covariance between two sample autocovariances. The main problem with their estimate lies in the

choice of a smoothing parameter. In this communication, under some regularity assumptions, we give expression for the asymptotic bias and variance of this estimator and we propose a consistent iterative estimator of the smoothing parameter minimizing the mean square error. The finite sample behavior of our procedure is illustrated in a simulation study.

#### Reference

Berlinet, A. and Francq, C. (1998). On Bartlett's formula for nonlinear processes. *J. Time Ser. Anal.* In press.

*Sous des conditions de régularité pour des processus stationnaires fortement mélangeant, Berlinet et Francq (1998) ont proposé récemment un estimateur à délai pondéré convergent de la covariance entre deux autocovariances échantillonnées. Le problème central de leurs méthode d'estimation réside dans le choix d'un paramètre de lissage. Dans cette communication, sous des conditions de régularité, nous donnons des expressions pour le biais et la variance asymptotique de cet estimateur et nous proposons un estimateur itératif convergent du paramètre de lissage minimisant l'erreur quadratique moyen. La performance de notre procédure pour des séries finies est ensuite illustrée avec une étude par simulation.*

Daniel MOREAULT, Université de Montréal, Département de mathématique et statistique C.P. 6128 Succursale Centre-Ville Montréal (Québec) H3C 3J7  
moreault@dms.umontreal.ca

### [MS-35]

#### **Animation of the Storm in Osaka, Japan**

#### *Animation de l'orage à Osaka au Japon*

Ryan LAM\*, University of Western Ontario

A.I. MCLEOD, University of Western Ontario

In this research we are visualizing the storm pattern through animation. The flexible piecewise-planar variogram model (R. Barry & J. Hoef, 1996) is applied as a step in kriging the two-dimensional storm data. This non-parametric approach minimizes the problems of misspecifying the variogram model. The two assumptions that we make in our model are temporal and spatial intrinsic stationarity. These two terms will be discussed in more detail. To optimize the kriging region, we compute the convex hull region that contains all the stations. Within the convex hull, we create a fine grid in order to perform the planar interpolation of the storm data. We also discuss how to krig the storm data when we encounter many zero values that were recorded at the stations.

*Dans cette recherche nous visualisons la forme d'un orage par animation. Le modèle flexible du variogramme linéaire par morceaux (R. Barry & J. Hoef, 1996) est appliqué comme une étape du krigeage des données bidimensionnelles concernant l'orage. Cette approche non paramétrique minimise les problèmes de spécification du modèle du variogramme. Les deux hypothèses que nous adoptons (dans notre modèle) sont temporelle et spatiale, stationnairement intrinsèques. Ces deux notions seront discutées plus en détail. Pour optimiser la région de krigeage, nous calculons l'enveloppe convexe qui contient toutes les stations. À l'intérieur de l'enveloppe, nous créons un maillage assez fin pour l'interpolation des données sur l'orage.*

*Nous discutons également comment kriger les données (sur le orages) quand plusieurs zéros sont enregistrés par les stations.*

Ryan LAM, Department of Statistical & Actuarial Sciences University of Western Ontario, London, Ontario, Canada, N6A 5B7  
lam@fisher.stats.uwo.ca

## Session 12

### Biostatistics : Cancer Research

#### *Biostatistique : Recherche sur le cancer*

#### [MS-36]

#### **Mechanistic Models of Carcinogenesis: What is their Use for Cancer Research and Risk Assessment ?**

***Modèles mécanistes de cancérogénèse : quels sont leurs utilisations dans la recherche sur le cancer et dans l'évaluation des risques ?***

Annette KOPP-SCHNEIDER\*, Deutsches Krebsforschungszentrum

For the past 50 years, mechanistic models of carcinogenesis have been used to describe the formation of benign and malignant tumors. An overview will be presented for the models of carcinogenesis as they have been proposed by numerous researchers. The class of multistage models of carcinogenesis has been of particular interest since these models incorporate biologically interpretable parameters and can be treated with standard methods for stochastic processes. Multistage models describe the process of carcinogenesis on the level of cellular events such as mutation, cell replication and cell death. They can be formulated in the framework of stochastic processes. An overview will be provided for the mathematical techniques used to treat multistage models.

There are two main reasons for investigating models of carcinogenesis. One is to elucidate the biological process of carcinogenesis by formulating biological hypotheses as mathematical models, applying the models to appropriate data and interpreting the model fit and the estimated parameters. This procedure will be exemplified in the case of mouse skin carcinoma formation where hypotheses will be tested about the number of steps necessary for carcinoma formation and the precursor populations for carcinomas.

The other reason for investigating models of carcinogenesis is to provide a rational basis for risk assessment, e.g. for the extrapolation of dose-response curves of environmental agents down to a range of interest for regulating agencies. The use and the limitations of carcinogenesis models for the task of low-dose extrapolation will be discussed. Examples will be given for the incorporation of dose into multistage models, especially into the two-stage model with clonal expansion.

*Durant les 50 dernières années, les modèles mécanistes de la cancérogénèse ont été utilisés pour décrire la formation des tumeurs bénignes et malignes. Une vue globale des modèles pour la cancérogénèse sera présentée telle qu'elle a été proposée par plusieurs chercheurs. La classe des modèles multi-stages de la cancérogénèse ont*

*reçu un intérêt particulier car ils incorporent des paramètres biologiquement interprétables et ils peuvent être traités par les méthodes standards des processus stochastiques. Les modèles multi-stages décrivent le processus de cancérogénèse au niveau des événements cellulaires telles que la mutation, la division et la mort cellulaires. Ces événements trouvent leur place dans le champ d'application des processus stochastiques. Une vue globale des techniques mathématiques utilisées pour le traitement des modèles multi-stages sera présentée.*

*Il y a deux raisons essentielles pour examiner les modèles de cancérogénèse. La première est d'élucider le processus biologique de cancérogénèse en formulant des hypothèses biologiques sous forme de modèles mathématiques, d'appliquer ces modèles à des données appropriées et d'interpréter l'ajustement du modèle ainsi que les paramètres estimés. Cette procédure sera appliquée, par exemple, à la formation de cancers de la peau chez les rats où des hypothèses sur le nombre d'étapes nécessaire pour la formation de cancer et sur les stades précédant le cancer seront testées.*

*Les modèles de cancérogénèse permettent en outre de fournir une base rationnelle pour l'évaluation du risque, e.g. l'extrapolation de la courbe de dose-réponse des agents environnementaux jusqu'à des valeurs intéressantes pour les agences de régulation. L'utilisation et les limitations des modèles de cancérogénèse pour des tâches d'extrapolation à faibles doses seront discutées. Des exemples seront donnés où les doses sont incorporées dans les modèles multi-stages, spécialement dans le modèle à stage double avec expansion clonale.*

Annette KOPP-SCHNEIDER, Abteilung Biostatistik Deutsches Krebsforschungszentrum D-69009, Heidelberg GERMANY  
kopp@dkfz-heidelberg.de

#### [MS-37]

#### **Establishing Guidelines for Exposures to Carcinogens**

***Etablissement de lignes directrices pour l'exposition aux produits cancérogènes***

David GAYLOR\*, National Center of Toxicological Research

One of the roles of health oriented regulatory agencies is to set guidelines for exposures to toxic substances. This paper with focus on carcinogenic (cancer causing) contaminants that appear in very low concentration in consumer products. Adequate human data seldom are available to evaluate cancer risks from such exposures. Exposure guidelines generally are based on data from animal bioassays. In the absence of biologically-based dose-response models to estimate cancer risk (incidence), some form of linear extrapolation from high dose animal bioassays to low-dose human exposures generally is used. This procedure is believed to provide conservative overestimates of low-dose risk.

For carcinogens that do not react directly with the DNA (nongenotoxic carcinogens), nonlinear dose-response curves are generally expected. Perhaps, threshold doses below which no cancer occurs may exist. In such cases for noncancer endpoints, exposure guidelines are usually based on a no-observed-adverse-effect-level (NOAEL) or lowest-observed-adverse-effect-level (LOAEL). The exposure guideline or reference dose (RfD) is established by dividing the NOAEL or LOAEL by safety (uncertainty) factors. Uncertainty factors are typically employed to account for extrapolation from results in animals to humans, variation of sensitivity among individuals, and the ratio of the LOAEL to NOAEL where a NOAEL is not available. Default

values of 10 are typically used for each of the uncertainty factors. The health risk at the RfD is presumed to be zero, or at most, negligible. Because of the lack of a precise definition of risk at the NOAEL, various authors have suggested that the NOAEL or LOAEL be replaced by a benchmark dose. The benchmark dose is defined as a dose associated with a low incidence, e.g. 1 to 10%, of an adverse effect that is estimable from bioassay data.

Using a database of 20 chemicals that provide cancer risk estimates based on both human and animal data, the ratio of human to animal cancer potency appears to be lognormally distributed. Under some general assumptions, tolerance limits can be estimated for exposures to carcinogens. These tolerance limits are compared to the results obtained with linear extrapolation and with RfDs resulting from the division of benchmark doses by default values of uncertainty factors.

*Un des rôles des agences régularisant les normes de santé est d'établir des seuils à l'exposition aux produits toxiques. Ce papier met l'accent sur les contaminants cancérigènes (causant la cancer) qui apparaissent en faible concentration dans les produits de consommation. Des données adéquates sur les humains sont rarement disponibles pour évaluer les risques de cancer. Les seuils sont généralement évalués à partir de données provenant des expériences sur des cobayes "amiaux au laboratoire". En absence de modèles basés sur des arguments biologiques reliant la dose à l'incidence du cancer on extrapole linéairement la réponse pour des faibles doses sur les humains à partir des résultats obtenus avec de fortes doses sur des cobayes. Cette méthode fournit généralement des estimations conservatrices des risques associés aux doses faibles.*

*Pour les produits cancérigènes qui ne réagissent pas directement avec l'ADN (cancérigènes non génétiques), les courbes reliant la dose à la réponse sont souvent non linéaires. Il existe sans doute des seuils en dessous desquels le cancer ne peut se produire. Dans ces situations de latence, les seuils à l'exposition au cancer sont souvent basées sur "no-observed-adverse-effect-level" la non observation d'effet adverse (NOEA) ou sur l'observation du niveau adverse le plus faible (ONAPF) "lowest-observed-adverse-effect-level". La dose de référence (DDR) est établie en divisant le NOEA ou le ONAPF par des facteurs de sécurité (ou d'incertitude). Les facteurs d'incertitude sont typiquement employés pour compenser pour l'extrapolation des résultats sur les animaux aux résultats sur les humains, pour la variation de la sensibilité parmi les individus, et pour le quotient du ONAPF à la NOEA quand la NOEA n'est pas disponible. Par défaut la valeur 10 est utilisée pour chacun des facteurs d'incertitude. Les risques pour la santé à la DDF sont supposés être zéro, ou au plus, négligeable. A cause du manque d'une définition précise du risque associé à la NOEA, plusieurs auteurs ont suggéré de remplacer la NOEA et/ou la ONAPF par une dose de base. La dose de base est définie comme étant la dose provoquant une faible incidence d'effets nocifs, e.g. de 1 à 10%, estimée à partir des données sur des cobayes.*

*Dans une banque de données de 20 produits chimiques qui fournissent des estimations de risque de cancer à partir de données sur les humains et sur les animaux, le quotient d'incidence du cancer chez les humains à celle des cobayes semble suivre une loi lognormale. Sous des hypothèses générales, les limites de tolérance peuvent être estimées. Ces limites de tolérance sont comparées aux résultats obtenus suite à l'extrapolation linéaire et aux résultats obtenus par la DDR en divisant des doses de référence par les valeurs par défaut des facteurs d'incertitude.*

David GAYLOR, Division of Biometry National Center of Toxicological Research Food and Drug Administration Jefferson, AR 72079-

9502 U.S.A.

dgaylor@NCTR.FDA.GOV

## Session 13

### Multidimensional Statistics

#### *Statistique multidimensionnelle*

[MS-38]

**Some Comments on Some Selected Matrix Inequalities and Equalities, with Some Statistical Applications—dedicated to the memory of Geoffrey Stuart Watson (1921-1998)**

*Commentaires concernant un florilège d'égalités et d'inégalités de matrices, avec des applications à la statistique - à la mémoire de Geoffrey Stuart Watson (1921-1998)*

George P. H. STYAN\*, McGill University

We will consider some selected matrix inequalities and equalities, with some statistical applications. We will select results from those attributed to Bergström, Bloomfield-Watson, Bouniakowsky, Brauer-Mewborn, Cassels, Cauchy, Euclid, Fischer, Frucht-Kantorovich, Hadamard, Haynsworth, Knott, Kürschák, Pólya-Szegő, Schur, Schwarz, Schweitzer, Styan-Jia, and Wielandt. Some historical and biographical information will also be mentioned. We will end the talk with a solution to a question recently posed in the research journal *Economic Theory*—and a further question (and solution?).

Some joint work with Gülhan Alpargu, Simo Puntanen, Geoffrey S. Watson (1921-1998), and Hans Joachim Werneris included.

*Nous considérons quelques égalités et inégalités sélectionnées de matrices, avec des applications en statistique. Nous considérons des résultats attribués à Bergström, Bloomfield-Watson, Bouniakowsky, Brauer-Mewborn, Cassels, Cauchy, Euclid, Fischer, Frucht-Kantorovich, Hadamard, Haynsworth, Knott, Kürschák, Pólya-Szegő, Schur, Schwarz, Schweitzer, Styan-Jia, and Wielandt. De l'information historique et bibliographique sera aussi mentionnée. Nous terminerons cet exposé par une solution à un problème récemment posé dans le périodique "Economic Theory", par un problème supplémentaire (et sa solution ?)*

*Ces travaux sont faits en commun avec Gülhan Alpargu, Simo Puntanen, Geoffrey S. Watson (1921-1998), et Hans Joachim Werneris.*

George P. H. STYAN, Department of Mathematics and Statistics McGill University, Montreal QC H3A 2T5

styan@Math.McGill.CA

[MS-39]

**Point Estimation of a Coefficient of Determination**  
*Estimation ponctuelle d'un coefficient de détermination*

Eric MARCHAND\*, Université du Nouveau-Brunswick

We consider a decision-theoretic approach to the problem of estimating a coefficient of determination based on a random sample generated by a multivariate normal distribution. We present some recent results. First, we evaluate the performance of the mle and adjusted mle, which are both widely used. We aim to propose "good" estimators that perform favourably in comparison to the mle and adjusted mle. Secondly, we study the posterior distributions associated with Beta priors, which can be expressed as mixtures of Betas, with Generalized Hypergeometric distributions as the mixing distributions.

*Nous considérons une approche décisionnelle pour l'estimation ponctuelle d'un coefficient de détermination, ou l'échantillon provient d'une loi normale multivariée. Nous présentons quelques résultats récents. D'une part, nous évaluons la performance de deux estimateurs couramment utilisés, soit l'evm et l'evm modifié, et recherchons des estimateurs plus performants. D'autre part, nous étudions les distributions a posteriori associées aux lois a priori Beta, qui s'expriment comme des mélanges de lois Beta ou les paramètres de mélange sont des lois Hypergéométriques généralisées.*

Eric MARCHAND, Dept. of Mathematics and Statistics University of New Brunswick P.O. Box 4400 Fredericton NB E3B 5A3  
marchane@dms.umontreal.ca

#### [MS-40]

### A Robust Test for a Given Variance

#### *Un test robuste pour une variance donnée*

Martin BILODEAU\*, Université de Montréal

A robust test statistic  $W$  for testing the hypothesis  $H_0 : \Sigma = I$  against  $H_1 : \Sigma \neq I$  is proposed. Its asymptotic null distribution is  $\chi_f^2$ ,  $f = p(p+1)/2$ , for all underlying elliptical distributions with finite fourth moments. The asymptotic nonnull distribution under a sequence of local alternatives is shown to be noncentral chi-square. The test  $W$  is asymptotically fully efficient at the normal distribution as compared to the modified likelihood ratio test.

*Nous proposons un nouveau test robuste  $W$  pour confronter les hypothèses  $H_0 : \Sigma = I$  et  $H_1 : \Sigma \neq I$ . Sa distribution asymptotique, sous  $H_0$ , est  $\chi_f^2$ , où  $f = p(p+1)/2$ , et ce quelque soit la distribution elliptique des observations en autant qu'elle possède des moments du quatrième ordre. La distribution asymptotique sous une suite d'alternatives contiguës est la loi du khi-deux décentrée. Lorsque la distribution des observations est normale, le test  $W$  est complètement efficace relativement au test modifié du rapport de vraisemblance.*

Martin BILODEAU, Université de Montréal C.P. 6128, Succursale Centre-ville Montréal QC H3C 3J7  
bilodeau@dms.umontreal.ca

#### [MS-41]

### Bivariate Sign Tests Based on the Sup, $L_1$ and $L_2$ Norms

#### *Tests de signe basés sur les normes Sup, $L_1$ et $L_2$ pour des observations bivariées*

Denis LAROCQUE\*, Université de Montréal

The bivariate location problem is considered. The sup,  $L_1$  and  $L_2$  norms are used to construct bivariate sign tests from the univariate sign statistics computed on the projected observations on all lines passing through the origin. The tests so obtained are distribution-free under the null hypothesis. The sup-norm gives rise to Hodges' test. A class of tests derived from the  $L_2$ -norm, with Blumen's test as a member, is seen to be related to a class proposed by Oja and Nyblom (*J. Amer. Statist. Assoc.*, 1989). The  $L_1$ -norm gives rise to a new test. Its asymptotic null distribution is seen to be the same as that of the  $L_1$ -norm of a certain normal process related to the standard Wiener process. An explicit expression of its cumulative distribution function is given.

*Nous allons considérer le problème d'un échantillon bivarié. Nous construisons des tests de signe bivariés à l'aide des normes sup,  $L_1$  et  $L_2$  à partir des statistiques de signe univariées calculées au moyen des projections des observations sur toutes les droites passant par l'origine. Les tests ainsi obtenus sont libres (distribution-free) sous l'hypothèse nulle. Le test de Hodges (1955) découle de l'utilisation de la norme sup. Nous examinons le lien entre une classe de tests, dont le test de Blumen (1958) fait partie, découlant de l'utilisation de la norme  $L_2$  et une classe de tests proposée par Oja et Nyblom (1989). De l'utilisation de la norme  $L_1$  découle un nouveau test. Sa loi asymptotique sous l'hypothèse nulle est la même que celle de la norme  $L_1$  d'un certain processus gaussien qui est lié au processus de Wiener. Une expression explicite de sa fonction de répartition est également donnée.*

Denis LAROCQUE, GRIP Université de Montréal C.P. 6128, Succursale Centre-ville Montréal QC H3C 3J7  
larocque@dms.umontreal.ca

#### [MS-42]

### The Multivariate Linear Model with Matric-T Error Variables

#### *Le modèle linéaire multivarié où les variables erreurs suivent la loi $t$ matricielle*

B. M. Golam KIBRIA\*, The University of Western Ontario

M. Safiul HAQ, The University of Western Ontario

The predictive inference for the multivariate linear models with Matric-T error distribution has been considered in this paper. First we have derived the marginal likelihood function of unknown covariance parameters. Then we have derived the prediction distribution for known covariance parameters by using the structural relation of the model. It is observed that for known covariance parameters the prediction distribution for a set of future responses of the model has an  $mn_f$  dimensional Matric-T distribution with appropriate degrees of freedom. The prediction distribution has been modified by replacing the estimates of covariance parameters obtained from the marginal likelihood function. The results for some special cases have been obtained from the general results. As an application, the  $\beta$ -expectation tolerance region for the future responses has been constructed.

*Ce papier traite de la prédiction pour les modèles linéaires multivariés où les variables erreurs suivent la loi  $t$  matricielle. Premièrement nous calculons la fonction de vraisemblance marginale des*



éléments inconnus de la matrice de covariance. Ensuite nous déterminons la loi de prédiction, lorsque la covariance est connue, en utilisant la relation structurelle du modèle. Lorsque la covariance est connue, la loi de prédiction d'observations futures du modèle a une distribution  $t$  matricielle, de dimension  $mn_f$  et de degré de liberté approprié. La loi de prédiction a été modifiée lorsqu'on remplace les paramètres par leurs estimateurs provenant de la fonction de vraisemblance marginale. Les résultats pour des cas particuliers ont été retrouvés. En guise d'application, la région de tolérance de l'espérance de la loi Béta pour des observations futures est construite.

B. M. Golam KIBRIA, Department of Statistical & Actuarial Sciences, The University of Western Ontario, London, Canada N6A 5B7

[MS-43]

## Two Spaces of Projection for Regression

### Deux espaces de projection pour la régression

Marylène TROUPÉ\*, Université des Antilles-Guyane

Regression is generally presented in space of variables. Particularly, in this context, multivariate multiple regression with  $q$  response variables is equivalent to  $q$  multiple regressions (Johnson et al ;1992).

We propose an original definition of multivariate multiple regression in the space of individuals in order to use possible links between response variables (Troupé 1994, 1995). In both cases (space of variables or space of individuals), operators of projection are used. Thus, for this purpose, a (semi-) distance of weights is used in the space of variables, and in the space of individuals, semi-distances with relationship effects are used (Alfaro et al; 1997, Schektman; 1989, Troupé; 1998). Furthermore, two criteria (Schektman; 1989) are used : the first one is the variability effect, and the second one, the association effect.

For  $q$  response variables, the classical model of multidimensional multiple regression (which is equivalent to  $q$  multiple regressions) is obtained when the semi-distance with relationship effects is equal to the algebraic generalised inverse of the quadratic form of inertia (Troupé; 1994). The association effect is then equal to zero.

The aim of the proposed definition in the space of individuals is to protect (Cazes; 1975) multivariate multiple regression in having, for two predictor variables, a correlation coefficient which increases towards the (positive) correlation coefficient of these variables.

- **Keywords:** association effect, correlation coefficient, multivariate multiple regression, relational semi-distances, variability effect.

*La régression est généralement présentée dans l'espace des variables. Particulièrement, dans ce contexte, la régression multiple multivariée à  $q$  variables réponses est équivalente à la  $q$ -régression multiple (Johnson et al ; 1992).*

*Nous proposons une définition originale pour la régression multiple multivariée dans l'espace des individus afin d'utiliser les relations existantes entre les variables réponses (Troupé ; 1994, 1995). Dans les deux cas (espace des variables ou espace des individus) les opérateurs de projection sont utilisés. La (semi-) distance des poids est utilisée dans l'espace des variables, les semi-distances des effets sont utilisées dans l'espace des individus (Alfaro et al ; 1997, Schektman ; 1989, Troupé ; 1998). En plus deux critères (Schektman ; 1989)*

*sont utilisées : le premier est l'effet de la variabilité, le second est l'effet d'association.*

*Pour les  $q$  variables réponse, le modèle classique de la régression multiple multivariée (qui est équivalent à la  $q$ -régression multiple) est obtenu quand la semi-distance des effets est égale à l'inverse généralisé (algébrique) de la forme quadratique de l'inertie (Troupé ; 1994). L'effet d'association est alors nul.*

*Le but de cette définition, dans l'espace des individus, est de protéger (Cazes ; 1975) la régression multiple multivariée d'avoir, pour les deux variables prédictives, un coefficient de corrélation qui croît vers le coefficient de corrélation (positif) de ces variables.*

#### REFERENCES :

- Alfaro, M., Villalobos, M. (1997) Implementación del ACP usando distancias relacionales en presencia de 2 grupos de variables. X<sup>e</sup> Simposio Métodos Aplicados a las Ciencias, Costa Rica.
- Cazes P. (1975) Protection de la régression par utilisation de contraintes linéaires et non linéaires. *Revue de Statistique Appliquée*, vol. XXIII, n<sup>o</sup> 3, pp.37-57.
- Johnson, R.A., Wichern, D.W. (1992) Applied multivariate statistical analysis. Prentice-Hall International Editions.
- Schektman Y. (1989) Inner products and association indices useful for analysing some multiway tables. In : *Multiway Data Analysis*, Coppi R. & Bolasco S. (eds.), North-Holland, Amsterdam, pp.203-212.
- Troupé, M. (1994) Contribution à la régression multiple multidimensionnelle et à la génération de règles. *Thèse de Doctorat, Université Paul Sabatier, Toulouse.*
- Troupé M. (1995) Una presentación de la regresión múltiple multidimensional en el espacio de los individuos. IX<sup>e</sup> Simposio Métodos Aplicados a las Ciencias, Costa Rica, pp.183-192.
- Troupé M. (à paraître) Application of generation of statistical rules to medical data on arterial hypertension. IX<sup>th</sup> Meeting of the Caribbean Academy of Science, Mai 1998.

Marylène TROUPÉ, Département de Mathématiques-Informatique, Université des Antilles-Guyane, Campus de Fouillole 97139 Pointe-à-Pitre cedex (FWI)

Marylène.Troupe@univ-ag.fr

## Session 14

### Statistical Education : Panel Discussion on Probability and Statistics in Canadian Schools

#### Éducation statistique : Atelier sur l'enseignement de la probabilité et de la statistique dans les écoles canadiennes

[MS-44]

### Marginalized by the instructors and quickly forgotten by the students

#### Marginalisés par les enseignants et vite oubliés par les élèves

Gordon FICK\*, University of Calgary

**[MS-45]****Skills for Daily Living - talking back to numbers*****Un atout pour la vie quotidienne - savoir répondre aux chiffres***

Shirley MILLS\*, Carleton University

**[MS-46]****Painless statistics for non-statisticians - how can it be done?*****Des statistiques faciles pour les non-statisticiens - Est-ce possible ?***

Hew GOUGH\*, Statistics Canada

## Session 15

**Survey Methods: The Project to Improve Provincial Economic Statistics (PIPES)*****Méthodologie d'enquête : Le projet d'amélioration des statistiques économiques provinciales (PASEP)*****[MS-47]****Project to Improve Provincial Economic Statistics: Objectives and Pilot Surveys*****Projet d'amélioration des statistiques économiques provinciales : Objectifs et enquête pilote***

Normand LANIEL\*, Statistique Canada

D. ROYCE, Statistique Canada

In October 1996, the Government of Canada and the governments of Nova Scotia, New Brunswick and Newfoundland signed an agreement to harmonize their federal and provincial taxes effective April 1, 1997. Statistics Canada will be the primary source of the data to be used in the formula for allocating the pooled revenues among the participating governments. In order to meet this new requirement, Statistics Canada needs to substantially improve the quality, reliability and detail of its provincial economic statistics. Therefore, over the next three years, Statistics Canada will be undertaking a major project, known as the Project to Improve Provincial Economic Statistics (PIPES). Its main objective is to measure final sales of goods and services accurately, on an annual basis by province, in sufficient detail and in a timely manner.

The PIPES implementation plan is based on a major transformation of the current model used for the collection and compilation of business surveys. During the session, speakers will illustrate some of the major changes that will be implemented over the next three years. All aspects of the survey process will be affected. The Business Register will become the single frame for all business surveys,

the enterprise will be the primary unit for data collection and analysis, questionnaires will be redesigned to meet the requirements for allocating taxes; the sampling, estimation and survey processing for all annual surveys will be integrated in a coherent manner and data based on the fiscal year will be adjusted to the calendar year. In the introduction to the session, the focus will be on the major objectives of PIPES and the key elements of the design of the Pilot Survey for reference year 1997.

*Au mois d'octobre 1996, le gouvernement du Canada et les gouvernements de la Nouvelle-Écosse, du Nouveau-Brunswick et de Terre-Neuve ont signé une entente pour harmoniser les taxes fédérale et provinciale et ce, à compter du 1er avril 1997. Statistique Canada est l'organisme qui fournira les données utilisées dans la formule de péréquation qui servira à répartir les revenus entre les quatre gouvernements. Afin de pouvoir satisfaire à cette nouvelle exigence, Statistique Canada a besoin d'améliorer de façon substantielle la qualité, la fiabilité et le niveau de détail de ses statistiques économiques actuelles. Donc, au cours des trois prochaines années, Statistique Canada entreprendra un projet majeur connu sous le nom de Projet d'amélioration des statistiques économiques provinciales (PASEP). L'objectif principal de ce projet est de mesurer avec exactitude sur une base annuelle, pour chacune des provinces la taxe finale sur les produits et services et cela, avec suffisamment de détails et dans des délais de production rapides.*

*La mise en oeuvre du PASEP sera fondée sur une transformation majeure du modèle utilisé présentement pour la collecte et la compilation des enquêtes économiques. Durant la session, les conférenciers exposeront certains des changements majeurs qui surviendront au cours des trois prochaines années. Tous les aspects intervenants dans le traitement d'une enquête seront touchés. Le Registre des entreprises deviendra la seule base de sondage pour toutes les enquêtes-entreprises; l'entreprise sera l'unité principale pour la collecte et l'analyse; les questionnaires seront remaniés afin de satisfaire les besoins de la péréquation; les processus utilisés pour l'échantillonnage, l'estimation et le traitement de toutes les enquêtes annuelles économiques seront intégrés d'une manière cohérente et les données selon l'année fiscale seront ajustées selon le calendrier. Dans l'introduction à la session, on se concentrera sur les objectifs majeurs du PASEP ainsi que les faits saillants quant à la conception de l'enquête pilote pour l'année de référence 1997.*

Normand LANIEL, Division des méthodes d'enquêtes entreprises, Statistique Canada Tunney's Pasture Ottawa, Ontario K1A 0T6  
laninor@statcan.ca

**[MS-48]****The Business Register*****Le Registre des entreprises***

Élaine CASTONGUAY\*, Statistique Canada

Since 1988, Statistics Canada has maintained a Business Register. This register was developed based on administrative files of payroll deductions sent to Revenue Canada by Canadian employers. Over time, many economic surveys at Statistics Canada have moved over to the Business Register.

With the advent of the Project to Improve Provincial Economic Statistics (PIPES), the Business Register will become the only samp-

ling frame for all economic surveys. In order to meet the requirements of these surveys, the Register must improve its coverage. As well, several size measures and the structures of businesses must be added to the Register.

*Depuis 1988, Statistique Canada met à jour un Registre des entreprises. Ce registre a été développé à partir des fichiers administratifs de retenues à la source envoyés à Revenu Canada par les employeurs canadiens. Au cours du temps, plusieurs enquêtes économiques de Statistique Canada se sont greffées au Registre.*

*Avec la venue du projet d'amélioration des statistiques provinciales (PASEP), il a été décidé que le registre des entreprises devienne la seule base de sondage pour toutes les enquêtes économiques. Afin de rencontrer les exigences de ces enquêtes, le registre doit améliorer sa couverture. De plus, de nombreuses variables de taille et de structure des entreprises seront ajoutées au registre.*

Élaine CASTONGUAY, Division des méthodes d'enquêtes entreprises, Statistique Canada Tunney's Pasture Ottawa, Ontario K1A 0T6  
castela@statcan.ca

#### [MS-49]

### Questionnaire Development in PIPES

#### *La conception de questionnaires à travers la vision du PASEP*

Mike SIROIS\*, Statistique Canada

Sylvain PERRON, Statistique Canada

Jean DROUIN, Statistique Canada

Sylvain POIRIER, Statistique Canada

This presentation will underline the recent developments in business survey questionnaires at Statistics Canada, specifically in the framework of the Project to Improve Provincial Economic Statistics (PIPES). The addition of questions related to the sale of services has proven to be essential in order to calculate the variable "value added", a concept required by PIPES. Given the gathering of many surveys into one main project that has its own objectives, it was necessary to establish a strategy for the harmonisation of the questionnaires (existing or new questionnaires according to the industrial sector). Different types of questionnaires were created in order to take into account the size, the geographical situation and the industrial activities of the enterprises. Field tests of the questionnaires were an important stage in this large-scale project. Finally, a policy of administrative data utilisation was introduced to reduce the response burden of the enterprises. The latest challenges that we are facing will be discussed in the conclusion.

*Cette présentation soulignera les principaux développements survenus lors de la conception de questionnaires pour les entreprises à Statistique Canada et ce, dans le cadre du Projet d'amélioration des statistiques économiques provinciales (PASEP). Tout d'abord, l'ajout de questions reliées à la vente de services s'est avéré essentiel pour pouvoir calculer la valeur ajoutée, concept requis pour le PASEP. Ensuite, étant donné le regroupement de plusieurs enquêtes à l'intérieur d'un projet ayant ses propres objectifs, il a fallu établir une stratégie d'harmonisation des questionnaires (existants ou nouveaux selon le secteur industriel). De plus, différents types de questionnaires ont dû*

*être créés en tenant compte de la taille, de la situation géographique et des activités industrielles des entreprises. Par la suite, les tests des questionnaires effectués sur le terrain ont constitué une étape importante dans ce projet d'envergure. Enfin, une politique d'utilisation de données administratives a été introduite pour réduire le fardeau de réponse des entreprises. Nous verrons en conclusion les défis que nous devons affronter dans les années à venir.*

Mike SIROIS, Division des méthodes d'enquêtes entreprises, Statistique Canada Tunney's Pasture Ottawa, Ontario K1A 0T6  
siromik@statcan.ca

#### [MS-50]

### Sampling and Estimation for the Unified Enterprise Survey

#### *Échantillonnage et estimation dans le cadre de l'enquête unifiée sur les entreprises*

Michelle SIMARD\*, Statistique Canada

Guylaine DUBREUIL, Statistique Canada

This new project which consist of integrating different Statistics Canada annual programs with a unified approach is in full development in this first pilot year. One of the biggest challenges of this survey has been to develop a sampling and estimation strategy using standard methodologies, but with a more complex and global application than traditional business surveys. Furthermore, the production schedule is very tight. The most important aspects which were considered in the sampling design will be presented, among them: i) a common sampling unit for all of the different structure of Canadian business enterprises, ii) the sampling allocation and selection which were strongly related to the data acquisition strategy, iii) the use of one survey frame, iv) a two-phase approach, v) the users requirements in terms of precision and details. Those aspects are often different and contrasting with the importance of response burden reduction.

Also presented will be the proposed estimation method: the generalized regression estimator with a intensive use of fiscal data, selected in the first phase, as auxiliary information in the calibration of the estimator.

*Ce nouveau projet qui consiste à intégrer les différents programmes annuels à Statistique Canada avec une approche unifiée est en plein développement dans cette première année pilote. Un des plus grands défis de cette enquête a été de développer une stratégie d'échantillonnage et d'estimation basée sur des méthodes d'enquête standard, mais avec une application beaucoup plus vaste et complexe qu'une enquête entreprise traditionnelle. De plus, les échéanciers de production sont très serrés.*

*Les principaux aspects qui ont été considérés dans l'élaboration du plan d'échantillonnage seront présentés, entre autres : i) une unité d'échantillonnage commune pour les différentes structures d'entreprise au Canada, ii) la répartition et la sélection de l'échantillon qui sont fortement liées avec la stratégie d'acquisition des données, iii) l'utilisation d'une seule base de sondage, iv) une approche à deux-phases, v) les besoins en précision et en détails des utilisateurs, souvent différents et contrastant avec l'importance de réduire le fardeau de réponse.*

*Également présenté, la méthode d'estimation proposée : les estimateurs de régression généralisés avec l'utilisation de données fis-*

*cales, sélectionnés dans la première phase, utilisées comme information auxiliaire dans le calage des estimateurs.*

Michelle SIMARD, Division des méthodes d'enquêtes-entreprise, Statistique Canada Tunney's Pasture Ottawa, Ontario K1A 0T6  
simasyl@statcan.ca

## [MS-51]

### **Data collection for the Unified Enterprise Survey**

#### ***La collecte des données pour l'enquête unifiée sur les entreprises***

Sylvain PERRON\*, Statistique Canada

Mike SIROIS, Statistique Canada

Jean DROUIN, Statistique Canada

Sylvain POIRIER, Statistique Canada

Within the framework of the Project to Improve Provincial Economic Statistics (PIPES) a new Unified Enterprise Survey (UES) is being developed. This survey will eventually cover all the industrial sectors of the Canadian economy. However, for the first year, the UES will be sent to enterprises from 7 industrial sectors never surveyed before by Statistics Canada, or not covered by a survey recently. A data collection strategy has been developed for the UES. In this presentation, we will first describe the context in which the strategy was elaborated. We will then describe the strategy and the systems it requires. This will be followed by a description of the follow-up and editing strategy that will be used in the data collection process. We will finish the presentation by giving an overview of the work that will be done in the coming years, as the UES comes to encompass all industrial sectors.

*Dans le cadre du Projet d'amélioration des statistiques économiques provinciales (PASEP), une nouvelle enquête unifiée sur les entreprises (UES) est mise sur pied. À long terme, cette enquête couvrira tous les secteurs industriels de l'économie canadienne. Par contre, pour la première année, l'enquête sera menée auprès de répondants de sept secteurs industriels qui n'ont jamais été couverts par Statistique Canada auparavant, ou qui n'ont pas été couverts récemment. Une stratégie de collecte a donc été développée pour recueillir les données de cette enquête unifiée. Dans cette présentation on verra tout d'abord le contexte dans lequel cette nouvelle stratégie a été élaboré. On poursuivra avec une description de la stratégie de collecte et des systèmes qu'elle implique. On verra ensuite la stratégie de vérification et de suivie adoptée. On terminera la présentation avec un bref aperçu du travail qui sera entrepris dans les années futures lorsque l'enquête unifiée inclura d'autres secteurs économiques.*

Sylvain PERRON, Division des méthodes d'enquêtes-entreprises, Statistique Canada Tunney's Pasture Ottawa, Ontario K1A 0T6  
sperron@statcan.ca

## [MS-52]

### **Edit and Imputation of Data for the Unified Enterprise Survey**

#### ***La vérification et l'imputation des données de l'enquête***

### ***unifiée sur les entreprises***

Marie-Claude DUVAL\*, Statistique Canada

Julie BERNIER, Statistique Canada

The edit and imputation system for the Unified Enterprise Statistics (UES) allows the edit and imputation of quantitative data. The system must produce a complete microdata file such that the data are coherent at the record level. The record is defined here by the collection entity (CE), representing an establishment or a group of establishments from which the survey data was gathered. The Generalized Edit and Imputation System (GEIS) produced by Statistics Canada is used. Due to the lack of historical data, donor imputation was recommended; specifically, the nearest neighbour method. The mandatory variables common to the seven pilot projects of the UES (total revenue, total expenses, salary and wages, and the number of employees), as well as tax data will be used to choose the nearest neighbour. It is selected within a donor group defined generally by industrial activity and by geographical region.

Certain constraints of the survey design, the data collection, and GEIS have to be considered at the time of the implementation of the system. The impact of those constraints on the strategy used for the edit and imputation of data will be explained.

*Le système de vérification et d'imputation pour l'enquête unifiée sur les entreprises (EUE) permet la vérification et l'imputation de données quantitatives. Le système doit produire un fichier de micro-données complet de sorte que les données soient cohérentes au niveau de l'enregistrement. L'enregistrement est défini ici par l'unité de collecte(UC), représentant un établissement ou un groupe d'établissements, pour laquelle les données d'enquêtes sont recueillies. Le système généralisé de vérification et d'imputation produit par Statistique Canada, le SGVI, est utilisé. Faute de données historiques, l'imputation par donneur a été préconisée; spécifiquement la méthode du plus proche voisin. Les variables importantes et communes aux sept volets de l'EUE, telles le revenu total, les dépenses totales, les salaires et le nombre d'employées, ainsi que des données de taxes servent à choisir le plus proche voisin. Ce dernier est sélectionné à l'intérieur d'un groupe de données défini généralement par l'activité industrielle et par région géographique.*

*Certaines contraintes du plan de sondage, de collecte et du SGVI ont dû être prises en considération lors de la mise en oeuvre du système et seront élaborées pour expliquer la stratégie utilisée pour la vérification et l'imputation des données.*

Marie-Claude DUVAL, Division des méthodes d'enquêtes entreprises, Statistique Canada Tunney's Pasture Ottawa, Ontario K1A 0T6  
duvamar@statcan.ca

## [MS-53]

### **Calendarization in the Unified Enterprise Survey**

#### ***La calendrialisation de l'enquête unifiée sur les entreprises***

Yanick BEAUCAGE\*, Statistique Canada

Peter TARASSOFF, Statistique Canada

Many respondents to business surveys provide data for their respective fiscal years, which often differ both from one respondent to another and from the calendar year. Despite the mix of periods covered, business survey managers commonly aggregate all such data

and publish so-called "annual" estimates. However, several problems ensue from the fact that these estimates and any related growth rates are calculated from data with different fiscal year-end dates. These results would have been different if all of the data had been based on true calendar years. Calendarization, the transformation of fiscal year data to a calendar year basis, is a remedy to this problem. Statistics Canada's new Unified Enterprise Survey (UES) represents an opportunity for a more widespread application of calendarization. Under the aegis of the Project to Improve Provincial Economic Statistics (PIPES), the UES Pilot will target seven industries, and all respondent microdata are to be calendarized.

This presentation will show the importance of calendarization, explain how it is performed and what the necessary inputs are, and finally discuss certain aspects of the development of a calendarization system. These aspects include: the construction of seasonal profiles, the types of variable that are to be calendarized, the place of calendarization in a processing sequence, the constraints among variables and/or among units (enterprise/establishment) that calendarization ought to preserve, and finally the evaluation of calendarized values. This discussion will be made in an exploratory way since the UES calendarization system will not be ready until Fall 1998. Data from other surveys will be used to illustrate each point.

*Dans les enquêtes auprès des entreprises, la plupart des répondants rapportent leurs données selon leur année fiscale respective. Celle-ci diffère fréquemment d'un répondant à l'autre tout en étant différente de l'année civile. Malgré la diversité des périodes couvertes, les gestionnaires de ces enquêtes regroupent habituellement ces données et publient des estimations qualifiées d'années annuelles. Cependant, le fait de calculer des estimations ou des taux de croissance à partir de données provenant de différentes périodes fiscales engendre plusieurs problèmes. On obtiendrait des résultats différents si les données étaient basées sur l'année civile. La transformation de données fiscales en données civiles aussi appelée calendrialisation vient remédier au problème. Avec la nouvelle enquête unifiée sur les entreprises (EUE) de Statistique Canada, on a maintenant l'opportunité d'appliquer plus généralement la calendrialisation. Sous l'égide du projet d'amélioration des statistiques économiques provinciales (le PASEP), l'enquête pilote de l'EUE se concentrera sur sept industries cibles dont les données rapportées seront calendrialisées.*

*Cette présentation montrera l'importance de procéder à la calendrialisation, expliquera la méthode employée ainsi que les entrées nécessaires à son application et finalement présentera les différents aspects reliés à la mise en place d'un tel système. Parmi ces aspects, nous discuterons de l'élaboration des profils saisonniers, du type de variable à calendrialiser, du moment propice pour procéder à la calendrialisation, des contraintes à respecter entre les variables et/ou entre les différentes unités (entreprises/établissements) lors de la calendrialisation, et finalement d'une méthode de vérification des données calendrialisées. Ceci sera fait de façon exploratoire puisque le système ne sera mis en application qu'à l'automne 1998. Des chiffres provenant d'autres enquêtes seront utilisés pour illustrer les différents points.*

Yanick BEAUCAGE, Division des méthodes d'enquêtes entreprises, Statistique Canada Tunney's Pasture Ottawa, Ontario K1A 0T6  
beauyan@statcan.ca

## Session 16

### SSC Presidential Invited Address

#### *Allocution de l'invité du président*

[MS-54]

#### **Numbers and News: How Newsmakers- and the Media - Use and Abuse Economic Data**

#### ***Comment les gens qui font l'évènement- et les médias- utilisent et abusent des données économiques***

Bruce LITTLE\*, Globe and Mail

Every day, economic data fill the newspapers. Some stories are generated by the release of new statistics, others by politicians or lobbyists who try to put their own spin on old figures. For journalists, the challenge can be daunting because many are handicapped by their own innumeracy. After all, they became journalists because they prefer words to numbers. Faced with figures, their eyes glaze over and meaningless data finds its way into print -even if they contain internal inconsistencies. Readers often lose out. In some cases, reporters accept as fact statements that cannot be supported by readily-available data. In other cases, the biased interpretations of interest groups are passed on with too few questions raised. Paradoxically, there's a growing appetite for data even when hard measures are simply not available. Journalists like the certainty of yes-no answers and find it hard to comprehend the idea that many economic indicators (like productivity) are far more ambiguous than they look. If the public debate on some issues seems less than satisfying, the media's love-hate affair with numbers may be one reason.

*Chaque jour, des données économiques remplissent les journaux. L'émission de nouvelles statistiques est toujours génératrice de nouvelles, parfois ce sont les politiciens ou les lobbyistes qui tentent d'interpréter en leur faveur de vieux chiffres. Pour les journalistes, le défi peut être décourageant car plusieurs d'entre eux sont handicapés par leur incapacité à traiter les nombres. Après tout, ils deviennent journalistes car ils préfèrent les mots aux chiffres. En présence de chiffres, leurs yeux glissent rapidement et des données incompréhensibles trouvent souvent un chemin dans le journal- même si elles contiennent des contradictions. Les lecteurs sont souvent dérouterés. Dans certains cas, les reporters acceptent, comme des faits, des déclarations qui ne sont pas supportées par les données disponibles. Dans d'autres cas, les interprétations biaisées des groupes d'intérêt passent presque inaperçues. Paradoxalement, il y a un appétit grandissant pour des données même si des mesures strictes ne sont "tout simplement" pas disponibles. Les journalistes aiment la certitude des réponses oui-non et ont peine à comprendre l'idée que plusieurs indicateurs économiques, comme par exemple la productivité, sont plus ambigus qu'ils ne semblent à un premier coup d'oeil. Si le débat public sur certains sujets semble moins que satisfaisant, la relation trouble des médias avec les chiffres y est sans doute pour quelque chose.*

Bruce LITTLE, Globe and Mail 444 Front Street West Toronto, Ontario M5V 2S9  
BLittle@GlobeAndMail.ca

## Session 17

**Biostatistics: Issues in the Pharmaceutical Industry, Multiple Endpoints in Clinical Trials***Biostatistique : Débat dans l'industrie pharmaceutique, les essais cliniques à terminaison variable*

[MS-55]

**Handling Multiple Outcome Measures in Prospective Pharmacoeconomic Studies***Traitement de réponses multiples dans les études prospectives pharmaco-économiques*

Jayanti MUKHERJEE\*, Bristol-Mayers Squibb

In the field of clinical trials methodology there has been much debate about how to handle multiple outcome measures. In pharmacoeconomic studies this problem is compounded since sample size is calculated based on clinical outcome measures and not economic measures. The reason for collecting economic data is to convince payers (eg. provincial ministries of health) that they should provide reimbursement for a particular drug because it offers good "value for money" and therefore, this decision is based on the strength of the economic data. Although p-values may not be that useful in this context there has to be some form of guidance on how convincing the data actually is. This presentation will focus on the lessons learned from the study of an antibiotic in the treatment of acute exacerbation of chronic bronchitis. After a brief presentation of the study design, there will be a discussion of the following:- 1) handling of the multiple outcomes data, 2) interpretation of this data and 3) lessons learned; this will help in the planning of future trials.

*La méthodologie des essais cliniques a connu un grand débat sur la manière de traiter les réponses multiples. Dans les études pharmaco-économiques, ce problème est complexe car la taille de l'échantillon est calculée en fonction des mesures cliniques obtenues et non pas selon des mesures économiques. La raison de recueillir des données économiques est de réussir à convaincre les commanditaires (e.g. les ministères provinciaux de santé) des intérêts qu'ils ont à tirer en finançant un médicament particulier, car il présente une plus-value pour leur argent. Cette décision est alors basée sur des données économiques. Bien que les seuils observés ne soient pas très utiles dans ce contexte, ils indiquent certainement combien ces données économiques sont convaincantes. Cette présentation met le point sur les leçons acquises lors de l'étude d'un antibiotique dans le traitement d'une exacerbation d'une bronchite chronique. Après une présentation succincte du plan d'étude, il y aura une discussion portant sur : 1) le traitement des données à réponses multiples, 2) l'interprétation de ces données et 3) les leçons acquises ; ceci aidera aux planifications d'essais cliniques ultérieurs.*

Jayanti MUKHERJEE, Director Health Economics & Outcomes Research Bristol-Myers Squibb 2365, Côte-de-Liesse Montréal QC H4N 2M7

[jmukherj@uscmail.bms.com](mailto:jmukherj@uscmail.bms.com)

[MS-56]

**Design Issue of Phase II Cancer Clinical Trials with Multiple Endpoints***Débat sur la planification des essais cliniques à terminaison variable pour le cancer en phase II*

Benny ZEE\*, NCIC Clinical Trials Group

E. EISENHAUER, NCIC Clinical Trials Group

The objective of a phase II clinical trial is to assess the anti-tumour activity of a specific treatment regimen. In the past, multiple testing procedure and various optimal designs have been proposed. The problem of these methods based on response alone is that they may not be sensitive to effectively screen regimens that produce low response but high early progression rates. Zee et al. (1996) proposed a two-stage phase II stopping rule for both response and early progression, and demonstrated a gain in efficiency. In this paper, we would discuss applications of this method to various types of phase II studies. For example, compounds such as matrix metalloproteinase inhibitor or other anti-proliferative agents may play a minor role in anti-tumour activity but may be more effective in prolonging time to disease progression. Similarly, patients who were refractory to the first line treatment are not likely to respond to treatment but may prolong survival. Other example includes the design issues of an advanced prostate cancer trial using early treatment failure as an ancillary endpoint to the primary endpoint of change in PSA levels. Serious adverse event will also be considered. Simulation results will be shown to illustrate this method.

*L'objectif des essais cliniques en phase II est d'évaluer l'activité d'un agent anti-tumeur chez des sujets recevant un traitement spécifique. Auparavant, plusieurs tests multiples et des plans optimaux ont été proposés. Le problème de ces méthodes, basés uniquement sur les réponses, est qu'il ne peuvent pas être assez fiables pour montrer d'une manière précise les sujets qui produisent une réponse faible mais présentant précocement des taux de progression forts. Zee et al. (1996) a proposé une règle d'arrêt pour la phase II en deux étapes (pour la réponse et pour la progression précoce). Il a démontré qu'un gain en efficacité en résulte. Dans ce papier, nous discuterons des applications de cette méthode pour différents types d'études de phase II. Par exemple, des composés tels que l'inhibiteur "metalloprotéinase" de la matrice ou autres agents anti-prolifératifs peuvent jouer un faible rôle dans l'activité de l'agent anti-tumeur mais peuvent être plus efficaces dans le prolongement de la progression de la maladie. Similairement, les patients réfractaires au premier traitement ne sont pas susceptibles de répondre au traitement mais pourraient voir leurs durées de vie augmentées. Un autre exemple porte sur les problèmes de planification des essais cliniques du cancer avancé de la prostate. On utilise l'échec du traitement précoce comme point de terminaison "ancillaire" à la terminaison primaire du changement de niveaux du "PSA". Les événements adverses sérieux seront aussi considérés. Les résultats de simulation seront présentés pour illustrer cette méthode.*

Benny ZEE, Senior Biostatistician NCIC Clinical Trials Group 82-84, Barrie Street Queen's University Kingston, Ontario K7L 3N6  
[zeeb@ncic.ctg.queensu.ca](mailto:zeeb@ncic.ctg.queensu.ca)

**[MS-57]****Testing Hypotheses with Multiple Endpoints in the Regulatory Context*****Tests d'hypothèses à terminaison variable dans le contexte réglementé***

Peter A. LACHENBRUCH\*, Food and Drug Administration

In the drug licensure process, many possible endpoints (aka variables) may be proposed as demonstrating the efficacy of the product. The question of how to handle such multivariate problems has concerned licensing agencies for some time. Many tests have been proposed, and each has its strengths and weaknesses. The first test that comes to mind is Hotellings  $T^2$ , but its power can be limited. Other alternatives are to adjust for multiplicity by the Bonferroni inequality, modifying the Hotelling test for one-sided alternatives, developing a rank procedure. These will be reviewed. Recent articles by Pocock, Sankoh et al. will be summarized.

*Dans le processus d'autorisation de médicaments, plusieurs terminaisons possibles (variables "aka") peuvent être proposées selon l'efficacité du produit. La façon de traiter de tels problèmes a intéressé les agences agréées responsables pour un certain temps. Plusieurs tests d'hypothèses ont été proposés et chacun a ses forces et ses faiblesses. Le premier test qui vient à l'esprit est le  $T^2$  de Hotelling, mais sa puissance peut être faible. D'autres alternatives sont ajustées pour la multiplicité par l'inégalité de Bonferroni, modifiant le test de Hotelling pour des alternatives unilatérales, en développant une procédure de rang. Ceci sera revu. De récents articles de Pocock, Sankoh et al. seront résumés.*

Peter A. LACHENBRUCH, Epidemiology and Biostatistics Food and Drug Administration 1401 Rockville Pike Rockville, MD 20852 U.S.A. LACHENBRUCH@cber.fda.gov

**Session 18****Industrial Statistics*****Statistique industrielle*****[MS-58]****A Review of Response Surface Methodology: 1988-1998*****La méthodologie par les surfaces de réponses de 1988 à 1998***

G. Geoffrey VINING\*, University of Florida

Much has happened in the response surface methodology (RSM) literature since 1988. Recent important areas of research include: alternatives to robust parameter design, variance dispersion graphs, restrictions on randomization, multiple responses, generalized linear models, and new software. This talk highlights and explains the basic developments of these current trends in RSM.

*Beaucoup de développements de la méthodologie par les surfaces de réponses (MSR) sont apparus dans la littérature depuis 1988. Les recherches récentes portent sur : des alternatives au plan d'expérience robuste pour les paramètres, les graphiques de dispersion des variances, les restrictions à la randomisation, les réponses multiples, les modèles linéaires généralisés, et des nouveaux programmes informatiques. Cet exposé souligne et explique le développement de cette nouvelle tendance dans la MSR.*

G. Geoffrey VINING, Department of Statistics 103 Griffin-Floyd Hall, Box 118545 University of Florida Gainesville, Florida 32611-8545 U.S.A.

vining@Stat.Ufl.Edu

**[MS-59]****Fractional Factorial Split-Plot Designs for Industrial Experiments*****Fraction de plan d'expériences factorielles de type split-plot pour des expériences industrielles***

Randy R. SITTER\*, Simon Fraser University

It is often impractical to perform experimental runs of a fractional factorial in a completely random order. In these cases, restrictions on the randomization of experimental runs are imposed and the design is said to have a split-plot structure. With emphasis on two-level experiments, we discuss design and analysis issues of fractional factorial split-plot designs and their application to industrial experiments.

*Il n'est souvent pas pratique de réaliser des expériences d'une fraction d'un plan factoriel de manière complètement aléatoire. Dans ces cas, des restrictions sur la randomisation sont imposées. Le plan d'expérience qui en résulte est dit avoir une structure split-plot. En mettant l'accent sur les expériences à deux niveaux, nous discutons des difficultés qui apparaissent lors de la planification et de l'analyse de fractions d'expériences factorielles de type split-plot ainsi que de leur application dans les expériences industrielles.*

Randy R. SITTER, Dept. of Mathematics and Statistics Simon Fraser University Burnaby, B.-C. V5A 1S6

sitter@math.sfu.ca

**[MS-60]****Current Research Directions in Statistical Process Control*****Les directions de recherche actuelles dans le contrôle statistique de la qualité***

William H. WOODALL\*, University of Alabama

An overview is given of current research in control charting methods for quality monitoring and improvement. The discussion includes an historical perspective along with ideas for future research topics. Research streams include, for example, variable sample size and sampling interval methods, economic designs, attribute data methods, charts based on autocorrelated observations, multivariate methods, and methods using resampling. Recommendations and references are

provided to those interested in pursuing research ideas in statistical process control.

*Un aperçu de la recherche actuelle sur les méthodes graphiques de contrôle pour l'amélioration de la qualité est présenté. La discussion a une perspective historique et suggère des sujets de recherche. Les domaines de recherche comportent, par exemple, les échantillons de tailles différentes possiblement pris à intervalles de temps variables, les plans d'économie, les méthodes pour données catégorisées, les graphiques basés sur des observations autocorrélées, les méthodes multivariées, et les méthodes utilisant le rééchantillonnage. Des recommandations et des références bibliographiques sont fournies à ceux qui sont intéressés à poursuivre l'étude de ces idées de recherche sur le contrôle statistique de la qualité.*

William H. WOODALL, Dept. of Management Science and Statistics  
University of Alabama Tuscaloosa, AL 35487-0226 U.S.A.  
WWoodall@CBA.UA.EDU

## Session 19

### Biostatistics : Linear Models

#### *Biostatistique : Modèles linéaires*

#### [MS-61]

#### **Change Point Model for Asthma Mortality**

#### ***Modèle avec point de rupture pour la mortalité par l'asthme***

Hong YANG\*, McGill University

Robert PLATT, McGill University

Sammy SUISSA, McGill University

MDI beta-2 agonists have been suspected of causing excess mortality in asthmatics. A dose-response curve was fitted by a maximum likelihood change-point model to understand the relationship between the medication and fatal asthma attacks, and to identify the point at which the rate increased. The maximum likelihood estimate of the change-point, obtained by the profile likelihood method, is 1.8 canisters/month (95bootstrap CI: 0.17-4.75), slightly larger than the maximum recommended dose (1.2 canisters/month). The excessive rate of fatal asthma attack for one additional canister per month before the change-point is 3.70 (95CI:3.53-10.40) after. Maximum profile likelihood methods and bootstrap confidence intervals are useful in making inference on the change-point for these problems; however, improvements which would give narrower confidence intervals would be useful.

*Les "beta-2 MDI" ont été suspectés d'augmenter la mortalité des asthmatiques. Une courbe de dose-réponse, avec point de rupture, a été ajustée par la méthode du maximum de vraisemblance pour comprendre la relation entre la médication et les crises d'asthme fatales et pour identifier le point à partir duquel le taux augmente. L'estimé du maximum de vraisemblance du point de rupture est de 1.8 vaporisateurs par mois (l' intervalle de confiance (IC) par la méthode du*

*bootstrap est 0.17-4.75), un peu supérieur à la dose maximale recommandée (1.2 vaporisateurs par mois). L'augmentation du taux de mortalité suite à une attaque d'asthme, liée à l'utilisation d'un vaporisateur additionnel par mois, est de 3.70 (l'IC à 95% est 2.73-4.67) pour 1000 asthmatiques par an avant le point de rupture et de 6.98 (l'IC à 95% est 3.53-10.40) après. Les constructions des IC par la méthode du bootstrap et par la méthode du maximum de vraisemblance sont utiles pour étudier le point de rupture pour ces problèmes. Des méthodes améliorées, fournissant des intervalles de confiance plus courts, seraient les bienvenues.*

Hong YANG, McGill University Departments of Pediatrics and Epidemiology and Biostatistics McGill University – Montreal Children's Hospital Research Institute 2300 Tupper St. Montreal, Quebec, Canada H3H 1P3  
hongy@epid.lan.mcgill.ca

#### [MS-62]

#### **Bootstrap Methods for the Sensitivity of a Diagnostic Test**

#### ***Des méthodes bootstrap pour la sensibilité d'un test de diagnostic***

Robert PLATT\*, McGill University

The sensitivity at a fixed specificity is often a parameter of interest for diagnostic tests. Estimates of this parameter based only on the distribution of scores in the abnormal test subjects underestimate the variance and give confidence intervals that are too narrow. Parametric and nonparametric methods to approximate the variance in the distribution of scores in the normal test subjects have been proposed. I propose a bootstrap approach to the problem which accurately captures the variance in both groups of subjects and provides confidence intervals with satisfactory properties. Bootstrap estimates and  $BC_a$  confidence intervals for the sensitivity for a given specificity are developed. Simulation studies demonstrate that for certain types of diagnostic test data confidence intervals based on the bootstrap perform much better than uncorrected methods and at least as well as other methods. The bootstrap methods perform better in terms of confidence interval coverage probability and standard error estimation than other methods when there is significant overlap between the normal and abnormal scores and the distributions of the scores are non-normal.

*La sensibilité pour une spécificité fixée est souvent un paramètre d'intérêt pour les tests de diagnostic. Les estimés de ces paramètres, basés uniquement sur la distribution des scores chez les individus anormaux, sous-estime la variance, ce qui fournit des intervalles de confiance très courts. Des méthodes paramétriques et non paramétriques pour approximer la variance de la distribution des scores des individus anormaux ont été proposées. Je propose une approche de bootstrap qui capture de manière précise la variance des deux groupes d'individus et fournit des intervalles de confiance avec des propriétés satisfaisantes. Les estimés par la méthode du bootstrap et les intervalles de confiances  $BC_a$  pour la sensibilité pour une spécificité donnée sont développés. Des études par simulation démontrent que les intervalles de confiance basés sur la méthode du bootstrap*



performent mieux que les méthodes non corrigées et aussi bien que les autres méthodes. Les méthodes du bootstrap donnent de meilleurs résultats pour les taux de couverture des intervalles de confiance et pour l'estimation de l'écart-type échantillonnal que les autres méthodes quand il y a un chevauchement significatif entre les scores normaux et non-normaux et que la distribution des scores n'est pas normale.

Robert PLATT, McGill University Departments of Pediatrics and of Epidemiology and Biostatistics McGill University – Montreal Children's Hospital Research Institute 2300 Tupper St. Montreal, Quebec, Canada H3H 1P3  
robertp@epid.lan.mcgill.ca

### [MS-63]

#### **Simultaneous Modeling of Operative and Long-Term Mortality after Coronary Artery Bypass Surgery, with Application to British Columbia Cardiac Registries Data**

*Modélisation simultanée de la mortalité opératoire et de la mortalité à long terme suite à la chirurgie coronarienne par contournement d'artères et application aux "British Columbia Cardiac Registries Data"*

M. GHAMRANI\*, Simon Fraser University

C. DEAN, University of British Columbia

J. SPINELLI, Centre for Health Evaluation and Outcome Sciences

Coronary Artery Bypass (CAB) Surgery is the most commonly performed open heart procedure. Operative mortality is defined as death within 30 days after surgery. We develop a parametric model for investigating the effect of covariates both on operative and long-term mortality simultaneously. This model is an analogue of that proposed by Lambert (1992; Technometrics) for zero-inflated Poisson regression. It identifies those factors which affect operative and long-term mortality differently, and those which affect them similarly. We illustrate the model and its development using data on all CAB surgeries performed in British Columbia from 1991 to 1994.

*La chirurgie coronarienne par contournement d'artères (CCCA) est la chirurgie à cœur ouvert la plus souvent pratiquée. La mortalité opératoire est définie par la mort dans un intervalle de 30 jours après la date de l'opération. Nous développons un modèle paramétrique pour étudier les effets de covariables sur la mortalité opératoire et sur la mortalité à long terme. Ce modèle est analogue à celui présenté par Lambert (1992; Technometrics) pour une régression de Poisson avec zéro-augmenté. Il identifie les facteurs qui affectent la mortalité opératoire et la mortalité à long terme différemment, et ceux qui affectent de la même manière. Nous illustrons ce modèle et son développement en utilisant les données de toutes les CCCA pratiquées en Colombie Britannique entre 1991 et 1994.*

Melody GHAMRANI, Dept. of Math & Stats, SFU, Burnaby BC V5A 1S6  
mghahraa@cs.sfu.ca

### [MS-64]

#### **Power Comparisons of Several Tests in the Analysis of Cross-Over Clinical Trials**

*Comparaison de la puissance de plusieurs tests pour l'analyse de plans croisés*

José Andrés CORREA\*, McGill University

François BELLAVANCE, McGill University

The power of the nonparametric approach proposed by Bellavance and Tardif (1995), to test the hypotheses concerning direct treatment and carryover effects for the three treatment three period and six sequences crossover design, is compared via simulations to the power of 1) the usual analysis of variance model based on ordinary least squares, 2) a modified F-test approximation, and 3) the test based on the generalized estimating equations method. The above four tests are compared under the null hypotheses of no direct treatment effect and no carry over effect as well as under some alternative hypotheses. Furthermore, different covariance structures, to take into account the within subjects correlation arising from the repeated measurements, and different probability distributions for the responses, namely normal and gamma, are used in the simulations.

*La puissance de l'approche nonparamétrique proposée par Bellavance et Tardif (1995), pour confronter les hypothèses concernant les effets des traitements et les effets rémanents dans le plan croisé à trois traitements trois périodes et six séquences, est comparée, à l'aide de simulations, à la puissance 1) des tests de l'analyse de variance basés sur les moindres carrés ordinaires, 2) d'un test approximatif F modifié, et 3) des tests issus de la méthode d'estimation des équations généralisées. Ces quatre tests sont comparés sous l'hypothèse nulle d'absence d'effets des traitements et d'effets rémanents ainsi que sous quelques hypothèses alternatives. Aussi, différentes structures de la matrice de covariances, pour tenir compte des corrélations introduites par les répétitions des mesures chez un même sujet, et différentes lois de probabilité pour la variable dépendante, à savoir la loi normale et la loi gamma, sont utilisées dans les simulations.*

José Andrés CORREA, McGill University McGill University Purvis Hall 1020 Pine Avenue West Montreal QC H3A 1A2

### [MS-65]

#### **Modeling Effect Size Estimates in Meta-Analysis Using Fixed and Mixed Effects Linear Models**

*Modélisation de la variable dépendante dans une étude de méta-analyse à l'aide de modèles linéaires comprenant des effets fixes et aléatoires*

Asmaâ MANSOUR\*, McGill University

François BELLAVANCE, McGill University

Martin COLE, McGill University

The main objective of this presentation is to focus on a quantitative method, called meta-analysis, to model data abstracted from different studies on a same research topic. Two models will be presented : a fixed and a mixed effects linear model. Under the fixed effects specification, the study characteristics are presumed to account

completely for variation in the true effect sizes. In contrast, the random effects specification in the mixed effects model assumes that part of the variability in these true effects is unexplainable by the model. Both, the method of moments and the method of maximum likelihood will be used to estimate the parameters of the model. An example dealing with the prognosis of depression in the elderly primary care and community based patients will be presented to illustrate the use of these models.

*L'objectif de cet exposé est de présenter une méthode quantitative, appelée méta-analyse, qui permet de modéliser les données provenant de différentes études sur un même sujet de recherche. Deux modèles seront présentés : un modèle linéaire à effets fixes et un autre à effets mixtes. En présence du modèle à effets fixes, nous supposons que les caractéristiques de l'étude explique entièrement la variabilité de la variable dépendante. Par contre, la spécification d'effets aléatoires dans le modèle à effets mixtes présuppose qu'une partie importante de la variabilité de la variable indépendante demeure inexplicable par le modèle. Deux méthodes seront utilisées pour estimer les paramètres du modèle : la méthode des moments et la méthode du maximum de vraisemblance. Un exemple portant sur le pronostic de la dépression chez les personnes âgées sera présenté afin d'illustrer l'utilisation de ces modèles.*

Asmaâ MANSOUR, Dept. of Mathematics and Statistics Burnside Hall McGill University Montreal QC H3A 2T5

#### [MS-66]

### Rasch Models for Polychotomous Scales

#### *Modèles de Rasch pour échelles polychotomiques*

Jean-Marie TRICOT\*, Université de Bretagne Sud

Mounir MESBAH , Université de Bretagne-Sud

Within the framework of Item Response Theory (IRT), measurement of an unidimensional latent trait will be described in two cases : the dichotomous item case and the polytomous item case as an extension of the first case. In the polytomous case, a Bernoulli model can be defined as a conditional probability of passing threshold  $h$  (item category  $h$ ) given either response  $h$  or response  $h-1$ , to compare this case with the dichotomous case. Parameter estimates and tests are also discussed.

Key words : IRT Models, Rasch Models, Latent trait, Sufficiency, Score, Predictors, LR Statistic.

*Nous nous plaçons dans le cadre de la Théorie des Réponses aux Items (IRT), et mesurons un trait latent individuel unidimensionnel dans les deux cas suivants : le cas où les items sont dichotomiques et le cas où ils sont polytomiques, ce dernier cas étant présenté comme une extension du cas précédent. Il est possible, dans le contexte polytomique, d'introduire un modèle de Bernoulli défini par la probabilité conditionnelle d'atteindre le seuil de réponse  $h$  (réponse positive à la catégorie  $h$  de litem considéré) sachant que soit  $h$  soit  $h-1$  a été atteint. Ceci permet de se ramener à un modèle de type dichotomique. Les estimations et les tests sur les paramètres sont abordés.*

Mots-clés : Modèles IRT, Modèles de Rasch, Trait latent, Exhaustivité, Score, Variables prédictives, Statistique LR.

Jean-Marie TRICOT, Laboratoire de Statistique Appliquée Bretagne Sud S.A.B.R.E.S.; Université de Bretagne Sud Rue Yves Mainguy,

Tohannic; 56 000 Vannes, France. Tel 33 2 97 68 32 27/32, Fax 33 2 97 68 32 26

jean-marie.tricot@univ-ubs.fr

## Session 20

### Special Address on Survey Methods

#### *Conférence spéciale sur la méthodologie d'enquête*

#### [MS-67]

### Dealing with Non-Response Correction by Benchmarking or by Constrained Sampling

#### *La correction de la non-réponse par calage ou par échantillonnage équilibré*

Jean-Claude DEVILLE\*, I.N.S.E.E.

The complete non response correction is generally made by re-weighting the elementary units. It results naturally from the estimation (using the responses) of a parametric response model. The weight corrections appear to be like the inverse of the estimated response probabilities. In the estimation procedure the use of the benchmarking principle works well since it allows for a variance reduction due to the randomness of the non response, its evaluation and its estimation. In doing so, we are led to develop new benchmarking techniques more general than those provided by the general theory. This new development may present an interest of its own. The partial non response correction is usually made by imputation. Among the different possible methods (hot-deck methods as we call them in France which use donors) are particularly appealing because of their simplicity and their non parametric aspect. They are also astonishingly related to the sampling theory. Particularly the use of constrained sampling techniques allows a very substantial reduction of the noisy and useless variance introduced by the random choice of the donors.

*La correction de la non-réponse totale se fait généralement par une ré pondération des unités élémentaires. Celle ci résulte logiquement de l'estimation à partir des données d'un modèle paramétrique de réponse. Les corrections de poids apparaissent alors comme les inverses des probabilités de réponses estimées. Cette façon de voir les choses amène à se poser des questions sur les méthodes d'estimation à mettre en oeuvre pour parvenir à ces fins. Il apparaît que l'utilisation d'un principe de calage arrange singulièrement bien les choses en autorisant une réduction de la variance causée par l'aléa de non-réponse, en permettant aussi de la calculer et de l'estimer. Au passage, on est amené à développer une forme de calage un peu plus générale que celle de la théorie habituelle et qui peut présenter un intérêt autonome. La correction de la non-réponse partielle se fait généralement par imputation. Parmi les différentes méthodes possibles celles qui font appel à des donneurs (hot-deck comme on dit en France) sont particulièrement appelantes à cause de leur simplicité et de leur caractère non-paramétrique. Elles entretiennent aussi des relations assez étonnantes avec la théorie de l'échantillonnage. En particulier le recours à des techniques d'échantillonnage contraint (ou équilibré) permet de diminuer très fortement la variance parasite et inutile introduite par le choix aléatoire des donneurs.*

Jean-Claude DEVILLE, I.N.S.E.E., Unité méthodes statistiques 18,  
Boulevard Adolphe PINARD 75675-Paris Cedex 14  
JEAN-CLAUDE.DEVILLE@DG75-F401.insee.atlas.fr

## Session 21

### Poster Session

#### *Séance par affichage*

#### [MS-68]

#### **A Categorical Constraint Guided Matching Algorithm** *Un algorithme de jumelage guidé par des contraintes catégorielles*

Tzen-Ping LIU, Statistics Canada

Statistical matching is a technique for integration of data from multiple sources (i.e., the matching files) at the micro level by identifying and linking records that correspond to similar individuals. In a real situation, the matching files may contain survey weights and the resulting matched file has to fulfill additional outside requirements on its size and the use of all records from all matching files. In this article, considering pairs of matching files, a new categorical constraints-guided, minimum distance and maximum weight matching algorithm is introduced. The algorithm iteratively synthesizes matched records from a pair of matching files in a manner that the nearest and heaviest weight records match first, and the farthest and lightest weight records match last, while satisfying the full auxiliary categorical constraints. The resulting matched file has an architecture which preserves the categorical association of the variables and the weights, and also satisfies the outside requirements.

*La méthode statistique de jumelage est une technique d'intégration de données provenant de différentes sources (i.e. les fichiers de jumelage) au niveau le plus fin par identification et linkage des données enregistrées qui correspondent aux individus similaires. En situations réelles, le fichier de jumelage peut contenir des poids calculés à partir de sondages, le fichier jumelé résultant doit obéir à des exigences additionnelles externes concernant sa taille et l'utilisation de toutes les données sur tous les fichiers jumelés. Dans cet article, considérant les paires de fichiers jumelés, un nouvel algorithme guidé par des contraintes de distance minimum et de poids maximums est introduit. L'algorithme synthétise itérativement les données jumelées à partir d'une paire de deux fichiers jumelés de sorte que les deux plus proches et les plus lourds poids enregistrés sont jumelés les premiers, les plus loins et les plus légers sont jumelés les derniers, tout en satisfaisant complètement les contraintes catégorielles. Le fichier jumelé qui en résulte a une architecture qui préserve l'association catégorielle des variables et des poids, et aussi satisfait les exigences externes.*

Tzen-Ping LIU, Statistique Canada Immeuble R.H. Coats Ottawa,  
Ontario K1A 0T6  
tzenliu@statcan.ca

#### [MS-69]

#### **Semi-parametric Estimation of the Distribution of Incubation Time for AIDS Using Cohort Data**

#### *Estimation semi-paramétrique de la distribution du temps d'incubation du SIDA à l'aide des données de cohorte*

David E. MATTHEWS, University of Waterloo

Jianguo SUN, University of Missouri

Estimation of the distribution of incubation times for AIDS has been extensively discussed in the literature. In particular, various parametric and nonparametric methods have been proposed. However, no single parametric model seems to provide a satisfactory fit to both short and long incubation times. The problem with nonparametric estimation is that the data for the tail of the incubation time distribution are usually sparse. To overcome these difficulties, we propose a semi-parametric method of estimation — a compromise between the wholly parametric and nonparametric approaches — which is based on a change-point model. An EM algorithm for implementing this semi-parametric approach is outlined. The proposed technique is used to analyze data from the Toronto Sexual Contact Study.

*L'estimation de la distribution du temps d'incubation du SIDA a été extensivement discutée dans la littérature. En particulier, plusieurs méthodes paramétriques et non paramétriques ont été proposées. Par ailleurs, aucun modèle paramétrique ne semble s'ajuster, de manière satisfaisante, aux temps d'incubation courts et longs. Le problème de l'estimation non paramétrique est que les données tombant sur la queue de la distribution, du temps d'incubation, sont en nombre très réduit. Pour surmonter ces difficultés, nous proposons une méthode semi paramétrique d'estimation— un compromis entre les approches purement paramétriques et non paramétriques— basée sur un modèle avec point de rupture. Pour mettre en oeuvre cette approche semi paramétrique un algorithme d'estimation-maximisation est indiqué. La technique proposée est utilisée pour analyser des données d'une étude de contacts sexuels à Toronto.*

David E. MATTHEWS, Statistics & Actuarial Science, University of Waterloo, Waterloo, Ontario Canada N2L 3G1  
dematthews@uwaterloo.ca

#### [MS-70]

#### **Physician Intervention with Respect to their Smoking Patients**

#### *Les interventions des médecins face aux patients fumeurs*

Niyonsenga T., Université de Sherbrooke

Courteau J., Université de Sherbrooke

Xhignesse M., Université de Sherbrooke

Vanasse N., Université de Sherbrooke

Laplante P., Université de Sherbrooke

Grant A., Université de Sherbrooke

A 1985 survey of primary care physicians in Massachusetts found that 90% of physicians routinely asked patients about smoking while

58% of them were prepared to counsel patients regarding cessation. Only 3% of surveyed physicians expressed confidence about the success of their efforts. Nevertheless, over 90% of smokers actually wanted to quit. Of those who did and achieved long term abstinence, the vast majority did so without the help of any formal program, material, or clinical intervention. Using the FAMUS (Family Medicine, Université de Sherbrooke) data register, we focused on responding to the following questions. Did physician actions regarding smoking cessation as well as their beliefs towards the efficacy of interventions change? Is there any decrease in the trends of cigarette consumption over time, and if so, is there a gender difference? Does the quit rate increase with the number of cardiovascular (CVS) risk factors, the presence of CVS disease or other health problems? Is there any effect of physician interventions on the quit rate? Three dependent variables will be used: the number of cigarettes smoked per day; the cumulative smoking index; the smoking cessation indicator. The covariates will include: gender and age groups; time between consecutive observations; length of follow up time; physician counselling interventions related to cigarettes; CVS risk; personal and family history of CVS disease. Mixed linear and non linear models will be used to point out average rates of change in the response variables and the effects of given covariates on them.

*Un sondage effectué en 1985 auprès des médecins du Massachusetts a montré que 90% d'entre eux s'informaient régulièrement du statut des patients en ce qui concerne l'usage du tabac et que 58% étaient disposés à conseiller aux patients fumeurs de cesser de fumer. Cependant, seulement 3% des médecins interrogés étaient confiants quant au succès de leurs efforts. Or, plus de 90% des fumeurs désirent cesser de fumer. Parmi ceux qui ont cessé de fumer, la grande majorité l'a fait sans l'aide de programmes ou d'interventions cliniques. En utilisant le registre de données FAMUS (Family Medicine, Université de Sherbrooke), nous proposons de répondre aux questions suivantes. Est-ce que les interventions des médecins vis-à-vis le tabac ainsi que leurs croyances sur l'efficacité de ces interventions changent ? Y a-t-il une décroissance du nombre de cigarettes consommées par jour et si oui, y a-t-il une différence entre les deux sexes ? Est-ce que le taux de cessation croît avec le nombre de facteurs de risque associés aux maladies cardiovasculaires (MCV), la présence de MCV ou d'autres problèmes de santé ? Y a-t-il un effet des interventions médicales sur le taux de cessation ? Trois variables dépendantes seront utilisées : le nombre de cigarettes consommées par jour, l'indice cumulatif (cumulative smoking index) et l'indicateur de cessation. Les covariables incluront : le sexe et le groupe d'âge ; le temps entre deux visites consécutives ; la durée du suivi ; les conseils du médecin ; le risque associé aux MCV et enfin, les histoires personnelle et familiale des MCV. Des modèles linéaires mixtes et des modèles non-linéaires seront utilisés pour évaluer les taux moyens de changements des variables dépendantes et les effets des covariables sur ceux-ci.*

T. NIYONSENGA, Centre de recherche clinique Centre hospitalier de l'U. de Sherbrooke 3001, 12 ième Avenue Nord Sherbrooke QC J1H 5N4  
tniyonse@aix1.si.usherb.ca

## Biostatistics: Applications *Biostatistique : Applications*

[MS-71]

### Survival Prediction for Terminally Ill Cancer Patients *Prédiction des durées de vie de patients en phase terminale de cancer*

Benoît MÂSSE\*, Université Laval

Kamal DESAI, Université Laval

Pierre ALLARD, Université Laval

The Maison Michel-Sarrazin, located in Sillery, Quebec, is a palliative care center with 15 beds specializing in the treatment of terminally ill cancer patients. In recent years, the median survival of patients admitted to the center has been around 10 days. Since 1985, data on patients has been collected at admission and during follow-up. To date, a computerized data set of more than 2500 patients has been compiled.

One of the main objectives of the study is the identification of potential prognostic factors of patient survival. Most of the factors observed, coded as dichotomous variables, indicate presence or absence of symptoms at admission. The eventual goal is the development of a predictive model of patient survival in order to assist physicians in deciding whether to admit a patient based on clinical examination at admission. A patient is not admitted if the physician judges the expected survival to be longer than 60 days.

Three approaches were used to develop predictive survival models. Specifically, we explored techniques based on (1) the traditional Cox model (2) the logspline model (HARE) where the conditional hazard function is estimated by a linear combination of B-spline functions and (3) survival and regression trees. The last two approaches possess similarities in the manner in which covariables are chosen: the knot addition procedure of HARE is comparable to tree growing and knot deletion to tree pruning. Both approaches employ a criterion (e.g., AIC or BIC) for the selection of the final model.

The large sample size enables the use of part of the sample for estimation while the other part is reserved for validation of the predictive models. The three approaches are compared and their respective advantages and disadvantages are discussed. Furthermore, the validity of the predictive models are compared to physician's estimation of patient survival in the last year.

*La Maison Michel-Sarrazin située à Sillery est un centre hospitalier de 15 lits se spécialisant dans les soins palliatifs pour des patients en phase terminale de cancer. La durée médiane de survie y est d'environ 10 jours (pour les dernières années). Depuis 1985, les informations obtenues lors de l'admission des patients et sur leurs survies sont recueillies et informatisées. Une banque de données avec plus de 2500 patients a ainsi été constituée.*

*Un des objectifs principaux de cette étude est d'identifier les facteurs potentiellement pronostiques de la survie des patients. La plupart des facteurs sont des covariables indiquant la présence ou l'absence d'un symptôme lors de l'admission. Le but étant de développer un modèle prédictif de la survie qui pourrait assister les cliniciens sur la décision d'admettre un patient lors de l'examen clinique à l'admission. Un patient n'est pas admis si le clinicien juge qu'il possède un pronostique de survie trop long (plus de 60 jours).*

*Trois approches ont été utilisées afin de développer des modèles prédictifs de survie. Plus spécifiquement nous avons exploré les ap-*

proches basées sur (1) le modèle (traditionnel) de Cox, (2) la méthodologie logspline (HARE) où le logarithme de la fonction de risques conditionnels est estimé par une combinaison linéaire de fonctions B-splines, et (3) les arbres de régression. Les deux dernières approches possèdent certaines similitudes aux niveaux du choix des covariables : la procédure d'addition de noeuds de HARE est comparable à la pousse de l'arbre ("tree growing") tandis que la procédure d'élimination de noeuds est comparable à l'émondage de l'arbre ("tree pruning"). Pour les deux approches, le modèle final est choisi à l'aide d'un critère (e.g., AIC ou BIC).

La grande taille d'échantillon permet d'utiliser une partie pour l'estimation et l'autre partie pour la validation des modèles prédictifs. Les trois approches sont comparées et les avantages et désavantages de chacune sont discutés. De plus, en utilisant les prédictions sur la survie des patients admis effectuées par les cliniciens au cours de la dernière année, la performance des modèles prédictifs est comparée aux prédictions des cliniciens.

Benoît MÂSSE, Centre de recherche, CHAUQ Pavillon  
Saint-Sacrement 1050, chemin Sainte-Foy Québec QC G1S 4L8  
Benoit.Masse@gre.ulaval.ca

## [MS-72]

### Family History Assessment: Development of a Family Risk Scale in the Evaluation of the Risk of Cardiovascular Diseases

#### *Évaluation de l'histoire familiale : développement d'une échelle de risque familial pour les maladies cardiovasculaires*

Théophile NIYONSENGA\*, Centre de recherche clinique (CRC-CUSE)

J. COURTEAU, Centre de recherche clinique (CRC-CUSE)

A. CIAMPI, Dept. of Family Medicine of McGill University

S. LUSSIER-CACAN, Institut de Recherche Clinique de Montréal (IRCM)

M. ROY, Institut de Recherche Clinique de Montréal (IRCM)

M. XHIGNESSE, CRC-CUSE, Univ. de Sherbrooke, IRCM

Evaluation of health status, disease risk and necessary screening procedures are largely dependent on an adequate family history risk assessment. The clinician's aim is to find out risk factors or pathologies which could affect his/her patient's health by targeting the presence of disease, and causes of death if any, among the patient's parents, siblings, children as well as grandparents. In the case of cardiovascular disease (CVD), clinician usually considers a patient to have a positive family history if at least one member of his/her first degree relatives (father, mother, brothers, sisters or children) have experienced CVD before the age of 60 years without further specification.

Intuitively, we can identify two potential problems when using this definition which combines the available amount of information concerning a patient's family into a single dummy variable. First, an unknown family history (adopted patient) does not necessarily mean

"a negative" family history. Secondly, a positive or negative family history does not hold the same amount of information for two different patients. Indeed, consider for instance two patients of the same age whose parents passed away by accident at age 40. One of the patients has a sibship of 6 brothers, aged 45 to 60, of which one had a myocardial infarction (MI) at age 55. The other has a sibship of 6 brothers among which 5 of the 6 had MI between age 50 and 60. Do these patients have the same positive family history? Is there a way to weight their CVD risk according to the available information on their respective sibships?

The purpose of this study is to introduce a more specific methodology for assessing family history of CVD using the available clinical information. The general approach is as follows: to each subject we associate family history covariates, say  $Z$ , one for each parent, each grandparent and one for the sibship (for the subject and his/her parents). Each covariate takes value between 0 and 1. For parents and grandparents, the covariate  $Z$  takes the value 1 or 0 depending whether the family member had CVD before the age of 60 or not. For those deceased without CVD,  $Z$  equals the probability of experiencing CVD problems if they had survived to age 60. This probability is evaluated assuming a hypothetical survival function and estimating its parameters using the available data and the maximum likelihood (ML) method. For the sibship covariates, we need to estimate the proportion of members of the sibship that develop a major cardiovascular problem at a given age. The covariate  $Z$  is defined by means of the empirical bayesian approach using specified prior distributions. We used data from the 1990 Santé Québec survey as well as data from the IRCM lipid clinic where family histories were more exclusively documented to validate this methodology. Two statistical methods (logistic regression model and prediction trees) were used to assess the effect of these familial covariates as well as other known risk factors for CVD on the overall CVD risk.

*L'histoire familiale est un élément clé de l'examen clinique d'un patient. Le clinicien cherche alors à cibler les facteurs de risque susceptibles d'influencer la santé de son patient ou les pathologies auxquelles il pourrait être plus sujet, en recherchant les maladies, et la cause du décès s'il y a lieu, des parents, des frères et soeurs, des enfants et des grands-parents de son patient. Pour les maladies cardiovasculaires (MCV), le cliniciens parlera d'une histoire positive si un parent au premier degré (père, mère, frères, soeurs ou enfants) a présenté des troubles cardiovasculaires avant l'âge de 60 ans. Ce concept de positivité (ou de négativité) représente donc, de par sa nature, un condensé de données qui se trouve capté dans la variable "histoire familiale positive (ou négative)".*

*Intuitivement, il est certain qu'une histoire familiale dite positive ou négative de deux patients donnés ne porte pas toujours le même "poids" d'information. Par exemple, considérons deux individus du même âge, dont les parents sont décédés accidentellement à 40 ans. Un de ces individus aurait 6 frères âgés de 45 à 60 ans dont un qui aurait fait un infarctus à 55 ans. L'autre individu aurait également 6 frères mais 5 sur 6 auraient fait un infarctus entre 50 et 60 ans. Ces deux individus ont-ils simplement une histoire familiale positive ou y a-t-il moyen de "pondérer" leurs risques respectifs par rapport à leurs histoires familiales ?*

*Ce projet vise à construire une échelle de risque familial basée sur l'information familiale disponible pour un patient. Nous utilisons deux méthodologies statistiques, l'une basée sur l'approche de régression (le modèle de Cox, la régression logistique) et l'autre basée sur la construction d'arbres de prédiction. Ces deux approches sont évaluées en utilisant les données issues de l'Enquête québécoise*

coise sur la santé cardiovasculaire et d'autres sources de données cliniques, telles que celles recueillies dans le contexte d'études familiales sur les hypercholestérolémies d'ordre génétique à l'Institut de recherches cliniques de Montréal. L'objectif final est de s'assurer d'une échelle de risque familial pour les maladies cardiovasculaires qui serait à la fois simple et efficace, utilisable ultérieurement dans différents contextes cliniques et épidémiologiques. Nous définissons, pour chaque sujet, un certain nombre de variables  $z$  prenant des valeurs comprises entre 0 et 1, et qui résument l'histoire cardiovasculaire familiale disponible.

Pour le père, la mère et les grands-parents du sujet : nous définissons une question du genre "Est-ce que le père de ce sujet a eu un problème cardiovasculaire majeur avant l'âge de 60 ans ?" La variable  $z$  prend la valeur 1 pour une réponse affirmative à la question et la valeur 0 dans le cas d'une réponse négative ; pour tous les autres cas, dans lesquels la réponse n'est pas connue, la valeur de la variable est donnée par la probabilité de réponse affirmative, calculée selon un modèle statistique simple (modèle de survie exponentielle, Weibull, ...) dont les paramètres sont estimés à partir des données.

Pour les fratries du sujet et de ses parents : supposons que le nombre d'occurrences de la maladie cardiovasculaire avant l'âge de 60 ans, dans une fratrie de taille  $n$  ( $n$  est variable), suit une loi de Poisson tronquée, de telle sorte que la probabilité soit concentrée entre 0 et  $n$ , avec une moyenne  $n\lambda$  et qu'on y observe  $m$  événement cardiaques. En spécifiant une distribution a priori de  $\lambda$ , une définition intéressante pour la variable  $z$  pourrait être l'estimateur de Bayes de  $\lambda$ , et cet estimateur est calculé par une approche bayésienne empirique à partir de la distribution a posteriori de  $\lambda$  sachant  $n$  et  $m$ .

Théophile NYOUSENGA, Centre de recherche clinique Centre hospitalier de l'U. de Sherbrooke 3001, 12 ième Avenue Nord Sherbrooke QC J1H 5N4

tnyonse@aix1.si.usherb.ca

[MS-73]

## From Trees to Neural Nets

### Des arbres aux réseaux de neurones

Antonio CIAMPI\*, Université McGill

Prediction trees and artificial neural nets (ANNs) are two powerful tools to construct non-linear predictors from data. Both of them can be formulated as statistical models and the construction process can be seen, in both cases, as an example of adaptive, data-driven modeling. This point of view is useful : on the one hand, it places both approaches in the same framework; on the other, it facilitates the development of 'hybrid' techniques that take advantage of the best features of traditional statistical models and ANNs. In this talk I will start by showing the advantages of prediction trees over traditional statistical models, the main one being the interpretability of the results. In spite of these advantages, some intrinsic limitations emerge, and one of them has prompted the development of a new approach called a tree with soft nodes. A soft node is based on a continuous variable, like age, and sends an individual to the right and to the left with data-specified probability : thus if I wish to predict, say, heart attacks from age, an ordinary, 'hard' tree will discover a threshold at which the probability of a heart attack suddenly jumps; by contrast, a

soft tree will identify a soft threshold around a certain age, and...I can continue celebrating my birthdays! Other limitations of trees have prompted generalizations like RECPAM, in which a tree can have a 'global' and a 'local' component, in other words, a RECPAM tree is a mixture of linear models and traditional trees. In one version, the leaves of the trees are associated to linear prediction models in certain specified variables, instead of constant predictors. I will discuss the general RECPAM approach to emphasize its potential for constructing such models from data. As a third attempt to improve on trees, Y. Lechevallier and I have worked on building an ANN 'around a tree, and then in combining such an ANN with another based on an additive model. Finally, a general approach is emerging, called the generalized soft tree. This can be seen as a hybrid neural net containing a hierarchy of statistical models of increasing complexity. General principles of statistical modeling permit a gradual ascent in the hierarchy if and only if enough data are available to warrant the increasing complexity. The generalized soft tree is inspired by the multi-level hierarchy of experts model of Hinton and Jordan, but, unlike the latter, is a constructive method. This aspect facilitates interpretability, a highly desirable feature in certain application, like biomedicine.

Arbres de prédiction et réseaux de neurones artificiels (RNA) sont deux puissantes techniques pour construire des prédicteurs à partir d'un ensemble de données. Dans les deux cas, le processus de construction peut être vu comme un exemple de modélisation adaptative et le prédicteur ainsi obtenu peut être formulé comme un modèle statistique. Ce point de vue est très utile : d'une part, il place les deux approches, apparemment distinctes, dans le même cadre théorique, de l'autre, il facilite le développement de techniques hybrides consentant l'exploitation des aspects plus avantageux de la modélisation statistique et des RNAs. Un premier exemple de développement facilité par le point de vue adopté ici, c'est le nouveau modèle appelé arbre à noeuds doux. Ce modèle a été développé afin de remédier à une des limites de la méthode de construction d'arbre classique, notamment le traitement des prédicteurs continus. Un noeud doux est un noeud, basé sur une variable continue, qui envoie un individu à droite ou à gauche avec une probabilité déterminé par les données. Par exemple, si l'on veut prédire une crise cardiaque à partir de la variable âge, un arbre ordinaire déterminera un seuil au dessus duquel la probabilité d'une crise cardiaque change de façon discontinue. Cela peut paraître contraire à l'intuition biomédicale. Par contre, un arbre doux identifiera un seuil doux, centré sur une certaine valeur d'âge, autour duquel la probabilité d'aller à droite évolue de 0 à 1 suivant une sigmoïde, ce qui est beaucoup plus conforme à l'intuition. Le centre et la forme de la sigmoïde seront déterminés par les données. D'autres faiblesses des arbres classiques ont suggéré des modèles plus souples, tels RECPAM, qui peut comprendre une composante à arbre et une composante linéaire 'globale'. Un troisième exemple de tentatives d'amélioration de la technique classique, c'est la construction de RNAs à partir d'un arbre, et, plus généralement, à partir d'un ou de plusieurs modèles statistiques facilement interprétables, les arbres et les modèles additifs étant deux choix naturels. Basé sur ces exemples, une approche générale est en train d'émerger, l'arbre doux généralisé. C'est un hybride RNA contenant une hiérarchie de modèles statistiques de complexité croissante. Des principes généraux de modélisation statistique permettent une montée graduelle dans la hiérarchie si et seulement si l'on dispose d'une quantité de données suffisante pour justifier la complexité de la

représentation. L'arbre doux généralisé est inspiré par la hiérarchie d'experts proposé par Hinton et Jordan, mais, à la différence de celle-ci, il se base sur une approche constructive 'par le bas'. Cet aspect facilite l'interprétation du prédicteur, une qualité très désirable dans certains domaines d'applications, tel le domaine biomédical.

Antonio CIAMPI, Epidemiology and Biostatistics McGill University  
Purvis Hall 1020 Pine Avenue West Montreal QC H3A 1A2  
ciampi@icm.umontreal.ca

## Session 23

### Special Session of the Centre de Recherches Mathématiques: Analysis of Shape

#### *Session spéciale du Centre de Recherches Mathématiques : Analyse de la forme*

#### [MS-74]

#### **Detecting Differences in Brain Shapes Using Random Fields**

#### *Détection des différences de forme du cerveau en utilisant les champs aléatoires*

Keith WORSLEY\*, McGill University

*Time* magazine recently reported that among hereditary depressives, the subgenual prefrontal cortex was 40% smaller than normal. There is great interest in the general question of whether external behavioural or cognitive measures are related to brain shape. Brain shape is measured from MRI images by three methods: 1) structure masks, taking the value 1 at points inside the structure and 0 outside, 2) surfaces, formed by shrink-wrapping a triangular mesh to the outer surface of the structure, and 3) non-linear 3D deformations required to warp the structure to an atlas standard. For each type of data, a test statistic for detecting changes in shape at each point is modeled as a random field, and localised shape changes are detected using some new results on the geometry of excursion sets of random fields.

*Time magazine a récemment rapporté que parmi les dépressifs héréditaires, le cortex préfrontal "subgenual" est 40% plus petit que la normale. Il y a un grand intérêt à déterminer si le comportement ou les mesures cognitives sont liés à la forme du cerveau. La forme du cerveau est mesurée à partir des images RMI par trois méthodes différentes : 1) les structures de masques, prenant la valeur 1 à l'intérieur de la structure et 0 à l'extérieur, 2) les surfaces, formées en enroulant un treillis triangulaire à la surface externe à la structure, et 3) les déformations 3D non linéaires nécessaires pour reporter la structure à un atlas standard. Pour chaque type de données, un test statistique pour détecter les changements de forme en chaque point est modélisé comme un champs aléatoire, et les changements de forme localisés sont détectés en utilisant quelques nouveaux résultats de la géométrie des ensembles d'excursions des champs aléatoires.*

Keith WORSLEY, Dept. of Mathematics and Statistics Burnside Hall  
McGill University Montreal QC H3A 2T5  
worsley@math.mcgill.ca

#### [MS-75]

#### **The Representation and Measurement of Shape and Applications**

#### *La représentation et la mesure des formes : théorie et applications*

Huiling LE\*, University of Nottingham

The study of shape has become increasingly important in a wide range of subjects. The basic problem which arises is that it is not possible to place 'shapes' on the real line in any logical manner, not even in a flat Euclidean space, without distorting to some extent the natural relation between them. The consequence of this is that classical statistical analysis, which is based on the metric structure of the real line or Euclidean space, cannot be applied directly to the statistical analysis of shape without, to some greater or lesser extent, misrepresenting the shapes themselves. In practice several different solutions have been developed to suit the intended applications.

This talk will discuss a few basic geometric properties of shape spaces and then demonstrate by examples how this can be applied to suggest some sensible statistics for analysis of shape and to tackle some mathematical issues arising with them.

*L'étude des formes est devenue importante dans une multitude de sujets. Le problème fondamental est qu'il n'est pas possible de placer, d'aucune manière logique, les formes sur la droite réelle, ni même sur un espace euclidien, sans violer à un certain degré la relation naturelle existante entre elles. Ainsi, l'analyse statistique classique, qui est basée sur la structure de la distance de la droite réelle ou de l'espace euclidien, ne peut s'appliquer directement à l'analyse statistique des formes sans, à un degré plus ou moins grand, modifier les formes elles-mêmes. En pratique, plusieurs solutions différentes ont été développées pour satisfaire des applications spécifiques.*

*Cet exposé discutera quelques propriétés géométriques élémentaires des espaces de formes. On montrera ensuite, par des exemples, comment appliquer celles-ci pour suggérer quelques statistiques naturelles pour l'analyse des formes et pour résoudre quelques problèmes mathématiques découlant de leur utilisation.*

Huiling LE, Department of Mathematics University of Nottingham  
University Park Nottingham NG7 2RD UNITED KINGDOM  
lhl@maths.nott.ac.uk

#### [MS-76]

#### **Statistical Shape Analysis of Landmarks and Images**

#### *Analyse statistique des formes codées à l'aide de repères et d'images*

Ian DRYDEN\*, University of Leeds

Shape analysis involves the study of the geometrical information of objects where location, rotation and scale effects are filtered out. Typical aims are to estimate population mean shapes, to describe shape variability and to test whether or not two populations have different mean shapes, where random samples of objects are available. We describe various methods for statistical shape analysis and a simple approach for planar data is to use the complex Watson distribution. Shape analysis of image data is also considered. In particular we describe a recent application where a shape model is combined with

multi-scale image information to assist in the automatic recognition of photographs of faces.

*L'analyse des formes porte sur l'étude de l'information concernant la géométrie des objets lorsque les effets de la position, de la rotation et de l'échelle ont été éliminés par filtrage. Les objectifs principaux sont l'estimation de la forme moyenne d'une population, la description de la variabilité des formes dans la population et la construction de tests d'homogénéité de plusieurs populations à partir d'échantillons aléatoires d'objets. Nous décrivons plusieurs méthodes statistiques pour l'analyse des formes. Une approche simple pour les données planes est d'utiliser la distribution complexe de Watson. L'analyse des formes pour les données d'image est aussi considérée. En particulier nous décrivons une application récente où un modèle de forme est combiné avec une information sur l'image à échelle multiple pour faciliter la reconnaissance automatique de photographies de visages.*

Ian DRYDEN, Department of Statistics University of Leeds Leeds LS2 9JT UNITED KINGDOM  
iand@amsta.leeds.ac.uk

## Session 24

### ASSQ Session : Quality Control in Industry

#### *Session organisée par l'ASSQ : Contrôle de la qualité en milieu industriel*

[MS-77]

#### **Coordinators "EAQ" and Statistical Methods**

#### **Coordonnateur EAQ et méthodes statistiques**

Marc DUCHESNE\*, Via Systems

Printed Circuit Boards (PCB) fabrication comprises numerous production steps. The copper plating process, an important part of the PCB fabrication, includes various sources of variation : lots, sample units positioning, electrical current distribution, etc. In this regard, multiple measurements of copper thickness were taken during a planned experiment. The exploratory analysis used for determining the similar variation patterns on the units measured will be described. Then, the DOE analysis will be presented briefly and the actions taken regarding the studied factors will serve as a conclusion.

*La fabrication des circuits imprimés comprend de nombreuses étapes de production. Le processus de placage du cuivre, étape importante de cette fabrication, comporte plusieurs sources de variation : lots, position des échantillons, distribution du courant électrique, etc. A cet effet, de multiples mesures d'épaisseur de cuivre ont été prises dans le cadre d'une expérience planifiée. L'analyse exploratoire employée pour déterminer les zones de variation semblables par échantillon mesurée sera décrite. Par la suite, l'analyse du plan d'expérience sera présentée succinctement. Les actions entreprises au niveau des facteurs étudiés conclueront cette présentation.*

Marc DUCHESNE, Via Systems 205, boul. Brunswick Pointe-Claire

QC H9R 1A5

marc.duchesne@circo.com

[MS-78]

#### **Confidence Limits for Capabilities Indices**

#### ***Intervalles de confiance pour les indices de capacités***

Daniel PROTEAU\*, C.M.P. Metal Products

Capability indices are very popular in quality departments of many industries. They are used to determine whether a production process is capable of producing items within the specified tolerance. Unfortunately, many industries use these indices in a wrong way. They forget these indices are computed from a sample and are therefore estimators.

No industry can inspect 100 % of its products for a simple reason of cost. The sample is often very small and does not represent the general production.

This communication will show some lower and upper confidence limits for the most commonly used capability indices, that is  $C_p$ ,  $k$ , CPU, CPL and CPk.

We will define U the upper specification limit and L the lower specification limit of a characteristic that we will call X.

We will assume that X is normally distributed. The process must be stable in order to produce reliable estimates of the mean and of the variance of X. We also assume that the process has reached a state of statistical control.

*Les indices de capacité sont très populaires auprès des départements de qualité de plusieurs industries. Ils sont utilisés pour déterminer si un processus de production est capable de produire des pièces à l'intérieur de limites spécifiées. Malheureusement, trop d'industries utilisent ces indices d'une façon incorrecte. Elles oublient que ces indices sont calculés à partir d'échantillons et sont donc des estimateurs.*

*En effet, aucune industrie ne peut inspecter 100 % de ses produits pour des raisons très évidentes de coût. De plus, les échantillons sont très souvent petits et non représentatifs de la production en générale.*

*Cette communication visera la présentation de quelques intervalles de confiance pour les indices de capacité les plus répandus dans les industries, c'est-à-dire,  $C_p$ ,  $k$ , CPU, CPL et Cpk.*

*Nous définirons alors U la limite de spécification supérieure et L est la limite de spécification inférieure d'une caractéristique qu'on appellera X.*

*Nous assumerons que X est normalement distribuée. Nous ferons l'hypothèse que le procédé est relativement stable dans le but d'obtenir des estimateurs fiables de la moyenne et de l'écart-type de X. Nous supposons aussi que le procédé a atteint un état de contrôle statistique.*

#### *Bibliographie*

*Kane, V.E. (1986). " Process Capability Indices ", Journal of Quality Technology, 18, pp 41-52*

*Sidney S. Lewis (1991), " Process Capability Estimates From Small Samples ", Quality Engineering(3) pp381-394*

*Rado, L.E. (1989). " Enhance Product Development by Using Capability Indexes ", Quality Progress 22,4, pp38-41*

*Youn-Min Chou, D.B. Owen, Salvador A. Borrego A.(1990), "Lower Confidence Limits on Process Capability Indices ", Journal of Quality Technology 22, pp 223-229*

*Youn-Min Chou, D.B. Owen (1984). " One-Sided Confidence Regions*



*on the Upper and Lower Tail Areas of the Normal Distribution. "*  
*Journal of Quality Technology 16, P150-158*

Daniel PROTEAU, C.M.P. Metal Products 1241, rue des Cascades  
Châteauguay QC J6J 4Z2  
cmpmetalproducts.quality@sympatico.ca

#### [MS-79]

### **Prevention of Premature Tool Wear in Metal Forming Using a Control Chart Adapted to the Nonnormality of the Data**

*Prévention de l'usure prématurée d'outils de formage du métal par une carte de contrôle adaptée à la non-normalité des données*

François PAGEAU\*, Les Technologies Industrielles SNC

The properties of the average control chart developed by Shewhart are well known when the measurements come from a normal distribution. In many applications, the effect of nonnormality is attenuated by the use of a subgroup size sufficiently large and the consequences of the central limit theorem. In some situations however, economic factors restricting the subgroup size, coupled with a non-normal distribution preclude the use of normal quantiles as a basis for control limits.

A study of the manufacturing process leads to a reasonable model for the distribution of the wall thickness uniformity of metal cylinders. We propose the sum of squares of the uniformity measurements as the statistic to plot on the control chart. The performance of the resulting control chart is compared to the usual average chart in terms of sensitivity and false alarm rate. Some practical aspects of the implementation on the shop floor will be mentioned.

*Les propriétés de la carte des moyennes développée par Shewhart sont bien connues lorsque les mesures proviennent d'une loi normale. Souvent, l'effet de la non-normalité de ces mesures est atténué par une taille de sous-groupe suffisamment grande accompagnée des conséquences du théorème de la limite centrale. Il arrive cependant que des facteurs économiques restreignent la taille des sous-groupes. Dans une telle situation, la non-normalité des mesures peut proscrire l'utilisation des quantiles de la loi normale pour établir la position des limites de contrôle.*

*Une étude du procédé manufacturier nous amène à postuler un modèle raisonnable pour la distribution des mesures d'uniformité de l'épaisseur de paroi de cylindres métalliques. Nous proposons la somme des mesures d'uniformité élevées au carré comme statistique à représenter sur la carte de contrôle. La performance de la carte de contrôle résultante sera discutée en termes de sensibilité et de taux de fausses alarmes. Certains aspects de la mise en service de la carte sur le plancher de l'usine seront mentionnés.*

François PAGEAU, Les Technologies Industrielles SNC 5, Montée des Arsenaux Le Gardeur QC J5Z 2P4  
pageauf@odyssee.net

## Survey Methods: Modeling Survey Data

### *Méthodologie d'enquête : Modélisation des données d'enquête*

#### [MS-80]

### **A General Framework for Estimating a Regression Model with or without Survey Data**

*Un cadre de travail pour estimer un modèle de régression avec ou sans données d'enquête*

Phillip S. KOTT\*, U.S. National Agricultural Statistics Service

In statistics we often estimate models that we know are not true representations of reality. A sensible strategy in such a situation is to loosen the assumptions underpinning the model. This note relaxes the strong assumption that the error term in a not-necessarily-linear regression model has mean zero given any set of realized values for the independent variables. In its place we make the much weaker assumption that the error term has mean zero unconditionally and is simply uncorrelated with the independent variables. This general regression framework is then extended to the estimation of a regression model with survey data. In so doing, a unified approach to estimating a regression model emerges.

*En statistique, nous estimons souvent des modèles tout en sachant qu'ils ne représentent pas vraiment la réalité. Une stratégie sensée lors d'une telle situation est de relâcher les hypothèses soutenant le modèle. Dans cette note, on relâche la forte hypothèse voulant que le terme d'erreur dans un modèle de régression - qui n'est pas nécessairement linéaire - ait zéro pour moyenne peu importe l'ensemble de valeurs obtenues pour les variables indépendantes. Au lieu, nous émettons l'hypothèse beaucoup plus faible que le terme d'erreur a inconditionnellement une moyenne égale à zéro et qu'il n'est pas corrélé avec les variables indépendantes. Ce cadre général de travail pour la régression est ensuite élargi à l'estimation d'un modèle de régression avec des données provenant d'enquêtes. Ce faisant, une approche unifiée pour estimer un modèle de régression est obtenue.*

Phillip S. KOTT, U.S. National Agricultural Statistics Service, Room 305, 3251 Old Lee Highway, Fairfax, VA 22030  
phil\_kott\_at\_nass-fx@nass.usda.gov

#### [MS-81]

### **Estimation of Populations Totals/Mean in Multiple Characteristic Surveys when Study Variable has a - Weaker Relationship with the Size Variable**

*Estimation de totaux/moyennes dans les sondages à caractéristiques multiples quand la variable étudiée est faiblement liée à la variable taille*

Pranesh KUMAR\*, University of Transkei

Sample surveys primarily deal with the estimation of population totals/ means of several characteristics simultaneously. It is well established that the estimates, using probability proportional to size

sampling methods, for those characteristics which are positively and linearly related with the size variable provide considerable gain in the efficiency. However, some of the characteristics under study may have weak relationship with the size variable. For such type of characteristics, Rao, J.N.K.(1966a,b) was the first to suggest alternative estimators which performed better than the usual estimators. Because of the practical importance of this problem, many researchers [ Lanke (1974), Bansal and Singh (1985), Kumar and Gupta (1985), Kumar and Herzel (1988), Amahia, Chaubey and Rao T.J. (1989), Bansal (1989,90), Rao, T. J. (1993), Kumar and Agarwal (1997)] became interested and made significant contributions to deal with the estimation in this situation.

In this paper, we present briefly the work carried out by various researchers and discuss comparisons among the available alternative estimators. The results of the study may help the survey practitioners to choose the alternative estimator on the basis of a priori information on certain population characteristics.

*Les sondages traitent, en premier abord, de l'estimation des totaux /moyennes de population quand plusieurs caractéristiques sont considérées simultanément. Il est reconnu que les estimés, par les méthodes d'échantillonnage utilisant la probabilité proportionnelle à la taille pour ces caractéristiques, qui sont positivement et linéairement reliées à la variable taille, fournit un gain substantiel en efficacité. Par ailleurs, quelques caractéristiques sous étude peuvent être faiblement liées à la variable taille. Pour un tel type de caractéristique, Rao, J.N.K.(1966a,b) est le premier à suggérer un estimateur alternatif plus performant que les estimateurs usuels. À cause de l'importance pratique de ce problème, plusieurs chercheurs [Lanke (1974), Bansal and Singh (1985), Kumar and Gupta (1985), Kumar and Herzel (1988), Amahia, Chaubey and Rao T.J. (1989), Bansal (1989,90), Rao, T. J. (1993), Kumar and Agarwal (1997)] se sont intéressés à ce type d'estimation, à laquelle ils ont contribué significativement.*

*Dans ce papier, nous présentons brièvement le travail fait par plusieurs chercheurs et comparons entre les estimateurs alternatifs disponibles. Les résultats de cette étude peuvent aider les personnes en charge des sondages à choisir l'estimateur alternatif convenable sur la base de l'information a priori sur certaines caractéristiques de la population.*

Pranesh KUMAR, Department of Statistics, University of Transkei,  
5117 Umtata, E. Cape, South Africa  
KUMAR.acd.U.TT.UTR@getafix.utr.ac.za

[MS-82]

### **Outlier Detection: A Comparison of Inter-quartile - Range versus a Variation of Sigma Gap Using Asymmetric Samples**

*Détection de données aberrantes sur des échantillons de distributions asymétriques. Comparaison de 2 méthodes : intervalle inter-quartile et règle de l'écart-sigma modifié*

K. NOBREGA\*, Statistics Canada

J. BERNIER, Statistics Canada

This presentation will show the results of a simulation that compared the number of observations flagged as outlying by an inter-quartile range and a variation of the sigma gap method. This study

has been undertaken for the new Unified Enterprise Survey. The survey will use donor imputation to impute missing data, excluding outliers (units that would not be good potential donors) from the donor pool. The outlier detection method used should be robust against asymmetry and small sample size and detect only those observations that are very dissimilar to other observations in the sample.

Results from simulated data and administrative data have shown that the sigma gap method consistently does not produce too many outlying observations, less than 1 percent. Also, unlike other traditional outlier detection methods, it is possible that no observations at all are selected as outliers if the distribution of gaps in the data is consistent across all observations. A similar outlier detection method will be used to identify units to review after imputation, using less restrictive parameters.

*Cette présentation exposera les résultats de simulations comparant le nombre d'observations aberrantes détectées selon la méthode de l'écart inter-quartile et une variante de la règle de l'écart-sigma. Cette étude a été élaborée dans le cadre de la nouvelle enquête unifiée sur les entreprises. Cette enquête utilisera l'imputation par donneur pour imputer les données manquantes, en excluant au préalable les valeurs aberrantes i.e. celles qui ne seraient pas de bons donneurs potentiels. On recherche une méthode de détection des données aberrantes robuste à un échantillon de petite taille et de distribution asymétrique et capable de détecter les unités différentes du reste des unités de l'échantillon.*

*Les résultats provenant de données simulées et de données administratives ont montré que la règle de l'écart-sigma ne détecte pas trop de données aberrantes soit moins de un percent. De plus, il est possible qu'aucune observation ne soit identifiée aberrante si les écarts entre les données ordonnées sont plutôt constants, contrairement aux méthodes traditionnelles de détection de données aberrantes. Une méthode similaire de détection de données aberrantes sera utilisée pour identifier des unités à réviser après imputation, en utilisant des paramètres moins restrictifs.*

K. NOBREGA, Statistics Canada Tunney's Pasture Ottawa, Ontario  
K1A 0T6  
nobrkar@statcan.ca

[MS-83]

### **A Statistical Approach for Disaggregating Mixed-Frequency Economic Time Series**

*Une approche statistique pour la désagrégation de fréquences mixtes dans les séries chronologiques*

Zhao-Guo CHEN\*, Statistics Canada

The problem of mixed-frequency time series data arises from changing the observation frequency. For example, we may have a time series with quarterly observations in the first portion and annual figures in the remainder. We shall call that quarter-year mixed-frequency data. In this paper we suggest a method to disaggregate the annual observations to quarterly values. The proposed method can easily be generalised to the year-quarter, quarter-month, year-month and other mixed-frequency situations; it may avoid difficulties of time series modelling and is easy to implement. A step-by-step algorithm of the method is given so that econometricians not expert in this area can still perform the procedure. The proposed method is illustrated

through two real examples. We also conduct a small scale Monte Carlo experiment to compare the proposed procedure with two existing alternative methods. Finally, some concluding remarks are given.

*Le problème suscité par la présence de fréquences mixtes dans les séries chronologiques apparaît lorsque la fréquence de la collecte des données change. Par exemple, les observations sont trimestrielles dans la première partie de la série et annuelles par après. Ce sont des données de fréquences mixtes trimestrielles- annuelles. Dans cet article, nous suggérons une méthode de désagrégation des observations annuelles en données trimestrielles. La méthode proposée peut facilement être généralisée pour traiter les cas de données annuelles-trimestrielles, trimestrielles- mensuelles, annuelles-mensuelles et autres cas; de plus elle évite les difficultés reliées à la modélisation des séries chronologiques et sa mise en pratique est simple. L'algorithme de la méthode est conçu de façon à permettre aux économétriciens qui ne sont pas experts dans ce domaine de suivre chaque étape et d'exécuter la procédure. Deux exemples réels sont utilisés pour expliquer la méthode proposée. Nous avons également effectué une petite expérience à l'aide de la méthode de Monte-Carlo pour comparer la procédure proposée à deux autres méthodes. Puis nous concluons.*

Zhao-Guo CHEN, Statistics Canada Tunney's Pasture Ottawa, Ontario K1A 0T6  
chenzha@statcan.ca

#### [MS-84]

### **Should Stores be Open on Sunday? The Impact of Sunday Opening on the Retail Trade Sector in New Brunswick**

*Les conséquences sur le commerce de détail au Nouveau Brunswick de l'ouverture des commerces le dimanche*

Benoît QUENNEVILLE\*, Statistics Canada

Pierre CHOLETTE, Statistics Canada

Marietta MORRY, Statistics Canada

Should stores be open on Sunday? This is a current issue for Canadians as the de-regulation of shopping hours is, depending on the province, either a recent phenomenon or being considered. This presentation focuses on the case of Sunday opening in the retail trade sector in New Brunswick and highlights some important aspects of that matter. One issue is whether Sunday opening of retail outlets increases total sales. A second issue is whether Sunday opening causes a redistribution of sales between trade groups in the retail trade sector. A third is whether Sunday opening causes a redistribution of the sales among the days of the week. These questions are answered using intervention analysis and trading-day regression. For the modeling of the latter, we present an alternative technique.

Key words: Easter effect; trading-day variation; intervention analysis; Fourier transform.

*Depuis 1991, les commerces de détail du Nouveau Brunswick peuvent ouvrir les dimanches pendant la période automnale. Lors de cette communication, nous tenterons de répondre à trois questions.*

*Premièrement, est-ce que l'ouverture des commerces le dimanche augmente le niveau provincial des ventes de l'ensemble des commerces de détail? Deuxièmement, est-ce qu'il y a une redistribution du total des ventes parmi les groupes de commerces lorsque les commerces ouvrent le dimanche? Finalement, quel est l'effet de l'ouverture des commerces le dimanche sur la distribution hebdomadaire des ventes? Nous utilisons un modèle d'intervention et un modèle de régression sur les jours ouvrables pour répondre à ces trois questions. Nous présentons aussi une nouvelle approche pour la modélisation des effets des jours ouvrables.*

Benoît QUENNEVILLE, Statistics Canada Tunney's Pasture Ottawa, Ontario K1A 0T6  
quenne@statcan.ca

## Session 26

### Biostatistics : Mixed Topics

### *Biostatistique : Sujets variés*

#### [MS-85]

### **Statistics Canada's Environmental Statistics Program: Future Directions**

*Le programme statistique sur l'environnement de Statistique Canada : Directions futures*

Cynthia BAUMGARTEN\*, Statistics Canada

Claude SIMARD, Statistics Canada

The Environment Statistics Program at Statistics Canada has been in place for close to 20 years. During this time, the program content has grown in response to increasing demands for environmental information. However, the demands for more information increase as new environmental issues arise on the horizon. With limited resources and in anticipation of ever increasing demands for data, the Environmental Statistics Program has mapped a program for data development that will meet information demands as we move into the 21 Century.

This paper will review the Program's progress to date and outline the conceptual framework that will serve as the blueprint for future data development for the Environment Statistics Program.

*Le programme statistique sur l'environnement de Statistique Canada a été mis en place il y a environ 20 ans. Pendant cette période, le contenu du programme a grandi suite à la demande croissante d'information environnementale. Par ailleurs, la demande pour plus d'information croît avec l'appréhension de nouveaux problèmes environnementaux. Avec des ressources limitées et une demande croissante de ces données, le programme statistique sur l'environnement a tracé un programme de développement de données qui répond à cette demande d'information, alors que le 21<sup>ème</sup> siècle approche.*

*Ce papier revoit la progression du programme jusqu'à nos jours et esquisse le cadre de travail conceptuel qui servira de plan directeur pour le développement de données pour le programme statistique sur l'environnement.*

Cynthia BAUMGARTEN, Statistique Canada Tunney's Pasture Ottawa, Ontario K1A 0T6  
baumcyn@statcan.ca

**[MS-86]****Simultaneous Modelling of Longitudinal and Cross-Sectional Dependence for Binary Outcomes****Modélisation simultanée de la dépendance longitudinale et transversale de réponses binaires**

Patrick J. FARRELL\*, Acadia University

Andreas I. SASHEGYI, University of Waterloo

K. Stephen BROWN, University of Waterloo

Some studies give rise to data which can be logically grouped into clusters in more than one way. The Waterloo Smoking Prevention Project (WSPP3), for instance, is a school-based smoking prevention trial in which the smoking status of a cohort of students was ascertained on an annual basis for seven years (beginning in grade 6), thus providing longitudinal data on students. At the same time the fact that these students were drawn from a variety of different schools induced cross-sectional clustering in the data, due to school effects. We present a model for analyzing this type of data, developed here specifically for dichotomous responses. It combines generalized estimating equations to account for the longitudinal dependence with an empirical Bayes approach for estimating random school effects, to account for the cross-sectional clustering. This facilitates a straightforward composite analysis. The estimation procedure for this model is discussed, as are the results of a simulation study used to investigate the properties of its estimates. An illustration using the WSPP3 data is given.

*Quelques études fournissent des données qui peuvent être naturellement groupées en classes de plusieurs manières. Le "Waterloo Smoking Prevention Project (WSPP3)", par exemple, est un programme de prévention contre la cigarette dans les écoles. Un suivi annuel des habitudes de fumer d'un groupe d'élèves est réalisé et ce pour une période de sept ans (commençant en sixième année), fournissant ainsi des données longitudinales sur les élèves. En même temps, le fait que ces élèves proviennent de différentes écoles induit de la dépendance dans les données. Nous présentons un modèle, développé spécialement pour les réponses dichotomiques, pour analyser ce type de données. Il combine les équations d'estimations généralisées qui tiennent compte de la dépendance longitudinale à une approche empirique de Bayes pour estimer l'effet aléatoire des écoles qui tiennent compte de la classification croisée par section. Ceci facilite une analyse composite directe. La procédure d'estimation pour ce modèle est discutée. Aussi des résultats d'une étude par simulation utilisée pour étudier les propriétés de ces estimateurs est discutée. Une illustration utilisant les données WSPP3 est donnée.*

Patrick J. FARRELL, Dep. of Mathematics Acadia University WOLFVILLE, N.S. CANADA B0P 1X0  
pfarrell@relay.acadiau.ca

**[MS-87]****Cooking with Statistics: Recipes for Calculating the Probability of Illness from Food Borne Microbial Pathogens****Cuisine statistique : Recettes pour calculer la probabilité de tomber malade à cause de microbes pathogènes****dans les aliments**

W. H. ROSS\*, Health Canada

S. HAYWARD, Health Canada

There are a number of statistical modelling issues related to quantifying and managing health risks associated with food borne microbial pathogens. Among these is the development of a dose response model relating the probability of illness to the number of organisms ingested. The challenges to developing these models is discussed and a generic template is constructed based on a meta-analysis of available human feeding studies.

*Il y a un bon nombre de problèmes reliés à la modélisation statistique de la façon de qualifier et de gérer les risques de santé associés aux microbes pathogènes issus des aliments. Parmi ceux-ci, on retrouve un modèle de dose-réponse reliant la probabilité de maladie au nombre d'organismes absorbés. Les défis rencontrés lors du développement d'un tel modèle sont discutés et une cadre générique, basé sur une meta-analyse des études disponibles sur les habitudes alimentaires des humains, est construit.*

W.H. ROSS, Health Canada Banting Bldg. (AL#2203B) Tunneys Pasture Ottawa, Ont. K1A 0L2

william\_ross@hc-sc.gc.ca

**[MS-88]****Local Likelihood and Profile Likelihood in Semiparametric Hazard Regression Model****Vraisemblance locale et vraisemblance de profil dans les modèles semi paramétriques de régression de la fonction de risque**

Xuewen LU\*, Agriculture and Agri-Food Canada

R.S. SINGH, University of Guelph

In the proportional hazards regression model, suppose the covariate effect may be written  $\exp\{Z^T\beta + \lambda(X)\}$ , where  $Z$  and  $X$  are covariates,  $Z$  may be vector-valued,  $X$  is a univariate.  $\beta$  is an unknown parameter vector,  $\lambda$  takes value in a real line and is an unknown smooth function. Suppose that baseline hazard function is parametrized as  $h_0(t) = h_0(t; \sigma)$ . In this paper, we discuss estimation of the risk function  $\lambda$  and the parameters  $\beta$  and  $\sigma$  of this type of semiparametric model. Two types of estimators are considered: maximum local likelihood estimators and maximum profile likelihood estimators. The asymptotic distribution theory for the estimators is developed and algorithms for computing the estimates are given. A Monte Carlo study and a data set are presented to illustrate the methodology

*Dans le modèle de régression à hasard proportionnel, supposons que l'effet de la covariable puisse être écrit  $\exp\{Z^T\beta + \lambda(X)\}$ , où  $Z$  et  $X$  sont des covariables,  $Z$  peut être vectorielle,  $X$  est univariée,  $\beta$  est un vecteur de paramètres inconnus,  $\lambda$  prend ses valeurs sur la droite réelle et est une fonction lisse inconnue. Supposons que la fonction hasard de base est paramétrisée comme  $h_0(t) = h_0(t; \sigma)$ . Dans ce papier, nous discutons l'estimation de la fonction de risque  $\lambda$  et l'estimation des paramètres  $\beta$  et  $\sigma$  pour ce type de modèle semi paramétrique. Deux types d'estimateurs sont considérés : Les estimateurs*

*du maximum de vraisemblance locale et les estimateurs du maximum de vraisemblance de profil. La théorie asymptotique de ces estimateurs est développée et les algorithmes pour calculer ces estimateurs sont donnés. Une étude par la méthode de Monte Carlo ainsi qu'un ensemble de données sont fournis pour illustrer cette méthodologie.*

Xuewen LU, Agriculture and Agri-Food Canada Kentvill Research Station Kentvill, NS B4N 1J5 Canada

[MS-89]

### **Least Squares By Completing The Square** *Moindres carrés en complétant les carrés*

Gordon H. FICK\*, University of Calgary

Most introductions to applied regression analysis avoid the use of matrix algebra and calculus. Few of these students have this sort of mathematics background and even if they have some of this background, it has typically been many years since they took the courses or used the tools. This then means that most of the key regression results and concepts that we teach are taken on faith. Regression analysis is then magic that appears from a 'black box'. I have been attempting a middle ground of exposition that provides most of the important results and concepts. I use only high school math that is typically seen in grade 11 in Alberta. After briefly illustrating this method of exposition, I will offer some observations and suggestions.

*La plupart des introductions à la regression appliquée évite l'utilisation des matrices et du calcul différentiel. Peu d'étudiants ont la formation mathématique requise et, même s'ils l'ont, ça fait plusieurs années qu'ils ont pris ces cours ou qu'ils les ont utilisés. Ceci veut dire que plusieurs des résultats clés de la régression et des concepts que nous enseignons font l'objet d'un acte de foi. L'analyse par régression semble alors magique, comme sortant d'une 'boite noire'. J'ai essayé une solution de compromis qui fournit la plupart des concepts et la plupart des résultats importants. J'ai utilisé uniquement les mathématiques élémentaires enseignées à la 11 ième année en Alberta. Après avoir illustré brièvement la méthode de l'exposé, j'offrirai des observations et des suggestions.*

Gordon H. FICK, Department of Community Health Sciences Faculty Of Medicine, University of Calgary, Calgary, Alberta, T2N 4N1  
gordon@health.ucalgary.ca

[MS-90]

### **Actuarial Adjustment for a Mortality Deficient Table** *Ajustements actuariels pour une table de mortalité déficiente*

Remy, Konga KONGA\*, Compagnie National d'Assurance, Cameroun

This paper aims to show the inadequation of a mortality table TD-TV 60/64 used by the African insurance companies. The resulting laws of mortality are therefore compared to those describing the

reality experienced by insurance companies. We propose some adjustments that takes into account these abnormalities.

We made use of some data taken from portfolios of the three largest insurance companies of Cameroun. These data stretch over the last ten years. By the laws of MAKEHAM and KEYFITZ, we constructed mortality tables. With MAKEHAM's law, we retained the generalized forms brought by QUIQUET, and used the KING and HARDY parameter estimations method. The quality of these adjustments have been approved by the KOLMOGOROFF-SMIRNOFF's goodness-of-fit test; (based on the maximum absolute difference between the empirical cumulative distribution function and the theoretical one for the survival of an individual).

We assessed the mathematical provisions of portfolios for the retained companies contracts with the mortality tables using different distributions. It results in a predicted deficit which explains some of the difficulties experienced by insurance companies.

*L'objet de notre papier est de montrer l'inadéquation de la table de mortalité TD-TV 60/64 utilisée par les compagnies d'assurance africaines et par conséquent les lois de mortalité qui en découlent par rapport à la réalité telle que vécue de nos jours dans nos compagnies d'assurances. Par la suite, nous proposons des ajustements pour tenir compte de ces anomalies.*

*Nous avons travaillé sur des données recueillies à partir des portefeuilles de trois importantes compagnies d'assurance du marché camerounais, données étalées sur les dix dernières années. Nous avons par les lois de MAKEHAM et de KEYFITZ, construit des tables de mortalité. Pour les lois de MAKEHAM, nous avons retenu les formes généralisées apportées par QUIQUET et avons utilisé pour l'estimation des paramètres la méthode de KING et HARDY. La qualité de ces ajustements a été confirmée par la mesure de l'écart maximum entre les fonctions de répartition empiriques et théoriques de la variable aléatoire de survie d'un individu (test de KOLMOGOROFF-SMIRNOFF).*

*Nous avons évalué enfin les provisions mathématiques des portefeuilles des différents contrats des compagnies retenues avec les tables de mortalité des différentes lois, il ressort un déficit de provision qui pourrait justifier les difficultés que connaissent certaines compagnies à faire face à leurs engagements.*

Remy, Konga KONGA, Chef du Département Vie, Compagnie National d'Assurance (CAN), B.P. 13038 Douala, Cameroun, Tel : (237) 42 32 29, Fax : (237) 43 19 30.

## Session 27

### Biostatistics : Estimation

#### *Biostatistique : Estimation*

[MS-91]

### **Weakly Parametric Models for Doubly Censored Data** *Modèles faiblement paramétriques pour des données doublement censurées*

Min ZHAN\*, University of Waterloo

We consider a form of doubly-censored data where the time of the initiating event is interval-censored and the time of the subsequent event is subject to right censoring. One example of doubly-censored data is the induction time between the infection time of HIV and the onset of the AIDS, since the infection time of HIV can be interval-censored and the onset of the AIDS can be right-censored. Our interests are mainly on estimating the survival function of the length of time between the two events. We discuss a piecewise constant hazard model for this type of data. The approach is also extended to deal with covariates by adopting a proportional hazard model. This method is illustrated in an example from AIDS studies.

*Nous considérons une forme de données doublement censurées où le temps de l'événement initial est censuré par intervalle alors que le temps de l'événement suivant est censuré à droite. Un exemple de données doublement censurées est l'intervalle de temps entre le temps d'infection par le HIV et le développement du SIDA. Le temps d'infection par le HIV peut être censuré par intervalle alors que le temps du développement complet du SIDA peut être censuré à droite. Note but est d'estimer la fonction de survie du temps entre les deux événements. Nous discutons un modèle à risque constant pour ce type de données. L'approche est aussi généralisée pour inclure les covariables en adoptant un modèle de risque proportionnel. Cette méthode est illustrée par un exemple tiré des études sur le SIDA.*

Min ZHAN, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1  
mzhan@barrow.uwaterloo.ca

#### [MS-92]

### **Bootstrap Variable Selection and Model Validation for Cox's Regression Model**

#### *Sélection de variables pour le modèle de Cox par la méthode du Bootstrap*

Feng ZHAO\*, Queen's University  
Dongsheng TU, Queen's University  
Joe PATER, Queen's University

Cox's proportional hazard regression model has been used extensively in medical studies to identify potential prognostic factors for diseases. In this presentation, a bootstrap method based on resampling from residuals is applied to a clinical trial comparing two chemotherapies in patients with epithelial ovarian cancer, as a tool to select relevant prognostic factors and validate the final model selected. This method is compared with the stepwise selection method and bootstrap resampling from original observations through data from a similar clinical trial on different patients. The validation accuracy of two bootstrap methods is also evaluated with this external dataset.

*Le modèle de Cox de régression à risque proportionnel a été extensivement utilisé dans les études médicales pour identifier les facteurs potentiellement pronostiques des maladies. Dans cette présentation, une méthode du bootstrap basée sur le rééchantillonnage des résidus est appliquée à un essai clinique comparant deux traitements par chiméotherapie chez les patients atteints de cancer "epithelial" de l'ovaire. Cette méthode est utilisée pour sélectionner les facteurs*

*pronostiques et valider le modèle final adopté. Cette méthode est comparée à la méthode de sélection pas à pas et à la méthode de rééchantillonnage du bootstrap de données originales d'un essai clinique similaire sur d'autres patients. La validation de la précision des deux méthodes du bootstrap est aussi évaluée sur ces données externes.*

Feng ZHAO, Department of Community Health and Epidemiology and NCIC Clinical Trials Group, Queen's University, Kingston, Ontario K7L 3N6  
6fz@qmlink.queensu.ca

#### [MS-93]

### **Choice of a Time Scale in Survival Analysis**

#### *Choix d'une échelle de temps en analyse de durées de vie*

T. DUCHESNE\*, University of Waterloo

In many survival analysis applications, there may not be a unique plausible scale in which to measure lifetime or assess performance. This especially occurs when several measures of usage or exposure are available on each unit. Some examples of such situations are the survival time of human beings exposed to a particular pollutant for many years, or the time to failure of automobile engines, which can depend on usage measures such as age, cumulative mileage, total number of cycles, etc. This talk will consider how one can define a "good" time scale, along with methods and benefits of combining different measures of usage and exposure into a single scale.

*Dans plusieurs applications en analyse de durées de vie, il est possible que plus d'une échelle soient plausibles pour mesurer la durée de vie ou la performance. Ceci se produit plus particulièrement lorsque plusieurs mesures d'utilisation ou d'exposition sont disponibles pour chaque unité. La durée de vie d'êtres humains exposés à un polluant particulier pendant plusieurs années, ou le temps de panne de moteurs d'automobiles, qui peut dépendre de mesures d'utilisation telles que l'âge, le kilométrage cumulatif, le nombre total de cycles, etc., sont des exemples de telles situations. Dans cette conférence, nous considèrerons comment définir une "bonne" échelle de temps, ainsi que les bénéfices et méthodes de combiner différentes mesures d'usage et d'exposition en une seule échelle.*

Thierry DUCHESNE, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1  
tduchesn@barrow.uwaterloo.ca

#### [MS-94]

### **Estimation of the Sterilization Time for Dental Instruments**

#### *Estimation du temps requis pour stériliser les instruments d'examen dentaire*

Marie Claude LAMBERT\*, Université Laval

The goal of this study is to find the time needed to sterilize dental instruments so that bacteria are eliminated and the instruments can be used safely on another patient.

The experimental procedure is to put bacteria on a number of instruments ( $N$ ). The instruments are placed in a sterilizer for  $X$  unit of time. The response is the number of sterilize instruments ( $Y$ ) among the  $N$  instruments. Several experiments are carried out, for various values of  $X$ . The data to be analyzed can be written as  $\{(Y_i, X_i, N_i)\}$ .

The distribution of the time needed for the instrument to become sterile belongs to the extreme value family. That is so because this time is the maximum survival time for all the bacteria on the instrument. Thus  $\pi(X_i)$ , the probability of success depends on three parameters. Estimation of these parameters is considered in this talk. Numerical algorithm for the estimation will be discussed and sample sizes required to get reliable estimators will be investigated.

*Cet exposé cherche à estimer le temps d'immersion nécessaire pour rendre stériles des instruments de dentisterie mis dans un stérilisateur à sel. La stérilisation est essentielle pour éviter la propagation de maladies d'un patient à un autre.*

*La démarche expérimentale est la suivante : on implante des bactéries sur un nombre d'instruments ( $N$ ). Les instruments sont mis dans le stérilisateur à sel pendant un certain temps ( $X$ ). Par la suite, chacun de ces instruments est examiné afin de déterminer la présence de bactéries. La variable réponse de cette étude est le nombre d'instruments stériles ( $Y$ ) parmi les  $N$  instruments de départ. Plusieurs expériences sont faites, à différentes valeurs de  $X$ . Les données à traiter sont du type  $\{(Y_i, X_i, N_i)\}$*

*La distribution du temps que les instruments doivent tremper pour devenir stériles sera du type valeur extrême. Ainsi  $\pi(X_i)$ , la probabilité de succès dépend de trois paramètres ( $w, \sigma, \gamma$ ) qu'il faut estimer. Cet exposé traitera de l'estimation des paramètres, des difficultés numériques rencontrées lors de la maximisation de la vraisemblance ainsi que des tailles d'échantillon requises pour avoir des estimations fiables.*

Marie Claude LAMBERT, Dép. mathématiques et statistique Pav. Vachon, bureau 1056 Université Laval Sainte-Foy QC G1K 7P4  
odelie@mat.ulaval.ca

#### [MS-95]

### Estimation of Small Mammals' Biomass Available for Coyotes Living in Forest and Farmland of Eastern Quebec

*Estimation de la biomasse de petits mammifères disponible aux coyotes dans les milieux forestier et agro-forestier de l'est du Québec*

Nancy CARON\*, Université Laval

In forest and mixed, forest-farmland, locations small mammals are the most important animal prey for east coyotes (*Canis latrans*) living in Eastern Quebec. Coyotes living in forest areas seem to have a smaller body mass than those living in mixed locations. To investigate the hypothesis that there are less preys available in forest site, we develop an estimator of biomass of the small mammals available for the coyote, and also an estimator for his variance. The talk will describe the sampling methods and the capture-recapture model used for the estimation.

*Dans les milieux forestier et agro-forestier, les petits mammifères représentent les proies animales les plus importantes dans le régime*

*alimentaire du coyote de l'est (Canis latrans) du Québec. Il semble cependant que les coyotes vivant en milieu forestier présentent une taille et une masse corporelle inférieure à celle des coyotes qui vivent en milieu agro-forestier. Pour nous permettre de vérifier l'hypothèse selon laquelle il y aurait une disponibilité de proies animales plus faible en milieu forestier qu'en milieu agro-forestier, nous présentons un estimateur de la biomasse de petits mammifères disponible pour le coyote ainsi qu'une estimation de sa variance. La méthode d'échantillonnage ainsi que le modèle de capture-recapture utilisé seront brièvement décrits.*

Nancy CARON, Dép. mathématiques et statistique Pav. Vachon, bureau 1056 Université Laval Sainte-Foy QC G1K 7P4  
ncaron @math.ulaval.ca

#### [MS-96]

### Comparison of Generalized Procrustree Analysis and the Intra-class Correlation Coefficient to Estimate Inter-Rater Reliability

*Comparaison de l'analyse Procustéenne généralisée au coefficient de corrélation intra-classe pour estimer la concordance inter-observateurs*

Natalie RODRIGUE\*, McGill University

François BELLAVANCE, McGill University

A problem often encountered in epidemiology is the evaluation of the validity of a questionnaire administered to subjects. This is usually performed with the help of the intra-class correlation coefficient (ICC). The ICC, however, has several drawbacks. This talk addresses questions on the use of Generalized Procrustes analysis (GPA) to estimate inter-rater reliability. GPA is one of the family of methods that are concerned with the analysis of data arising from several subjects. It is a powerful multivariate statistical technique developed in the framework of sensory evaluation to primarily assess inter-judge reliability and furthermore to measure the effect of treatments (usually food products). Because of the empirical nature of GPA, informal graphical methods that allow multidimensional spaces to be explored to reveal structure in the data are commonly used, hence facilitating the interpretation of results. Therefore, it is of interest to investigate how GPA could be adapted to epidemiology, more specifically, in the development of scales and in the evaluation of inter-rater reliability.

*Un problème courant en épidémiologie est l'estimation de la concordance entre observateurs. Dans ce contexte, le coefficient de corrélation intra-classes est un outil fréquemment utilisé qui présente cependant plusieurs lacunes. Cette présentation traite de l'utilisation de l'analyse Procustéenne généralisée (AGP) afin d'estimer la concordance entre les observateurs dans des études épidémiologiques. L'AGP fait partie d'une famille de méthodes pour l'analyse de données provenant de différents individus. L'AGP est une technique statistique multidimensionnelle développée dans le cadre de l'évaluation sensorielle afin de mesurer la consistance entre les juges. De plus, elle permet de mesurer l'effet des traitements, habituellement des aliments. L'AGP fournit des résumés graphiques qui permettent l'exploration d'espaces multidimensionnels afin de découvrir une structure dans les données facilitant ainsi l'interprétation des résultats. Toutes ces raisons justifient l'intérêt d'étudier comment l'AGP peut être adaptée*

*aux données épidémiologiques afin d'estimer la concordance inter-observateurs.*

Natalie RODRIGUE, McGill University Purvis Hall 1020 Pine Avenue West Montreal QC H3A 1A2

### [MS-97]

#### Joint Analysis of Survival and Quality of Life in Cancer Disease

##### *Analyse conjointe de la survie et de la qualité de vie dans le traitement du cancer*

Mounir MESBAH\*, Université de Bretagne Sud

Lucile AWAD, Rhône-Poulenc Rorer

Quality of life (QOL) analysis has been increasingly considered as essential for the evaluation of new anti-cancer therapies. An increase of survival must not be obtained at the price of the deterioration of the patients QOL. On the contrary, if the advantage in survival cannot be established, it may be important to show a benefit on the QOL. QOL is often considered as secondary end point in a clinical trial where the duration of life is the first end point. Clinician may be interested to know the effect of the drug on the patient QOL conditionally on its life duration.

In this communication statistical strategies for analyzing jointly survival and QOL in a longitudinal trial, will be presented.

Key words and phrases: Quality of Life, Survival Analysis, Longitudinal Data, Mixed Models, Summary Statistics, Latent Variable.

*L'analyse de la qualité de vie (QDV) prend une importance croissante dans l'évaluation de nouveaux médicaments en cancérologie. Lorsque le nouveau traitement allonge la survie, ce ne doit pas être au détriment de la QDV du patient. Si au contraire, l'avantage de survie n'est pas évident il peut être important de montrer un bénéfice en terme de QDV. La QDV intervient souvent comme critère secondaire dans un essai, où le critère principal est la durée de vie. Le clinicien souhaiterait connaître, l'effet du traitement sur la QDV du patient sachant sa durée de vie.*

*Dans cette communication, nous présentons différentes stratégies pour analyser conjointement les processus de durées de vie et des scores de QDV mesurés dans un essai longitudinal.*

Mounir MESBAH, Laboratoire de Statistique Appliquée Bretagne Sud S.A.B.R.E.S.; Université de Bretagne Sud Rue Yves Mainguy, Tohannic; 56 000 Vannes, France. Tel 33 2 97 68 32 27/32, Fax 33 2 97 68 32 26

Mounir.Mesbah@univ-ubs.fr

## Session 28

### ASSQ Session: Panel Discussion on Statistical Consulting

*Session organisée par l'ASSQ : Atelier sur la consultation statistique*

### [MS-98]

#### Panelists

##### *Panélistes*

Diane LEROUX, Ministère des transports du Québec

Michel FLUET, SOM recherches et sondages

Hélène CRÉPEAU, Université Laval

Many statisticians are working in domains of applications where frequent communications with non statisticians are the rule : they are known as our " clients ", for whom statistics is at most a tool among others and have only kept a vague recollection (sometimes painful !) of the statistics courses they had taken in the past. This type of client brings statisticians to act as popularizers, finding techniques, examples and analogies in order to state rather complex statistical concepts in common language, without relying too heavily on mathematics.

This panel will discuss the challenges, responsibilities and constraints as experienced by practitioners in statistical consulting and provide support to the idea that statisticians, whatever the job they hold, could capitalize on techniques (active listening, negotiation, etc.) and skills (human relations, curiosity, flexibility, etc.) that are traditionally observed in the consulting field.

Session format : starts with a short presentation by each participant, questions and discussion lead by the moderator, and questions or comments from the audience. French language will be used.

*Plusieurs d'entre nous évoluent dans des domaines d'application de la statistique qui impliquent des contacts fréquents avec des non-statisticiens : ce sont souvent nos "clients", ceux pour qui notre discipline est avant tout un outil parmi d'autres et qui, bien malgré eux, n'ont conservé que de vagues (et parfois pénibles !) souvenirs de leurs cours de statistique. Ce type de clientèle nous amène à devoir agir comme vulgarisateurs, en trouvant des trucs, des images et des approches qui permettent d'exposer en des termes simples des concepts statistiques parfois complexes sans référer trop intensivement aux formulations mathématiques.*

*Tout en élaborant sur les défis, les responsabilités et les contraintes tels que vécus par des praticiens de la consultation statistique, les discussions tenteront d'appuyer l'idée voulant que les statisticiens, ou qu'ils soient sur le marché du travail, pourraient mettre à profit les techniques (écoute active, négociation, etc.) et les habiletés (relations interpersonnelles, curiosité, flexibilité, etc.) que l'on retrouve traditionnellement dans le domaine de la consultation.*

*Formule utilisée : présentation d'ouverture de chaque participant, questions et discussion animée par le modérateur, et questions ou commentaires venant de l'auditoire. La langue utilisée sera le français.*

Diane LEROUX, Service économie, statistique et encadrement des plans de transport Ministère des transports du Québec 700, bd René-Lévesque Est, 22e ét. Québec QC G1R 5H1

Michel FLUET, Développement des affaires SOM recherches et sondages 969, Route de l'Église Bureau 300 Sainte-Foy QC G1V 3V4 Email: som@som.i

Hélène CRÉPEAU, Dép. de mathématiques et de statistique Pavillon Alexandre-Vachon Université Laval Sainte-Foy QC G1K 7P4 Email: hc@mat.ulaval.ca

## Session 29

### Survey Methods: Variance Estimation, Cal-



## ibration, and Related Topics

### *Méthodologie d'enquête : Estimation de la variance, calage et sujets connexes*

[MS-99]

#### **Enhancements in the Series of Monthly Estimates for the Retail Sales by a Calibration Estimator**

#### *Amélioration de la série d'estimations mensuelles des ventes au détail au moyen d'un estimateur par calage*

Joël BISSONNETTE\*, Statistique Canada

Isabelle MARCHAND, Statistique Canada

Martin ST-PIERRE, Statistique Canada

Julie TRÉPANIÉ, Statistique Canada

The Monthly Retail Trade Survey is a major survey conducted by Statistics Canada. The design uses stratified simple random sampling without replacement. The estimates of sales for the domains of interest are derived from the expansion estimator. The survey was last redesigned in 1988. The quality of some estimates has deteriorated since that time due to the aging stratification. A complete re-stratification occurred in 1998 and should fix some of the current problems. The calibration estimator has been studied in order to revise the estimates of previous years and to smooth the disruption with the old series of estimates. Such an estimator might replace the expansion estimator. This paper presents the development of this estimator subject to the survey constraints and the impact on the trends and the quality of the estimates.

*L'enquête mensuelle du commerce de détail est une enquête importante que mène Statistique Canada. Elle recourt à un plan de sondage stratifié aléatoire simple sans remise. Les estimations des ventes pour les domaines d'intérêt découlent de l'estimateur par dilatation. Le dernier remaniement important de l'enquête remonte à 1988. La qualité de certaines estimations s'est depuis détériorée peu à peu, en raison de la stratification vieillissante. Une re-stratification complète de la base de sondage a eu lieu en 1998 et devrait corriger certains des problèmes actuels. L'estimateur par calage a été étudié en vue de réviser les estimations des années antérieures et de les raccorder plus facilement aux nouvelles estimations. Un tel estimateur pourrait éventuellement remplacer l'estimateur par dilatation. Nous discutons dans cet article du développement d'un tel estimateur sous les diverses contraintes inhérentes à l'enquête et des effets sur les tendances et la qualité des estimations.*

Joël BISSONNETTE, Statistique Canada Immeuble R.H. Coats Ottawa, Ontario K1A 0T6  
bissjoe@statcan.ca

[MS-100]

#### **Parameter Specification for the Model Variance Structure in the Generalized Regression Estimator**

#### *Spécification du paramètre de la structure de variance*

#### *du modèle dans l'estimateur de régression généralisée*

Éric PELLETIER\*, Statistique Canada

Éric RANCOURT, Statistique Canada

In sample surveys, estimation is often performed with the aid of auxiliary information. A commonly used estimator which can utilize this information is the generalized regression (GREG) estimator. This estimator provides good flexibility and can be used in many situations; it is one of the main options available in Statistics Canada's Generalized Estimation System (GES). Within GES, it is possible to specify different variance structures for the model. However, the characteristics of the GREG estimator are not completely known for some specifications. This presentation is interested in the impact of the parameter specification for the model variance structure. Also, different outlier treatments which could reduce the effect on point estimation and variance estimation are presented.

*Dans les enquêtes par sondage on effectue souvent l'estimation à l'aide d'information auxiliaire. Un estimateur de choix permettant de tirer profit de cette information est l'estimateur de régression généralisée (GREG).*

*Cet estimateur fournit un cadre de travail flexible et assez vaste ; c'est d'ailleurs une des principales options disponibles dans le Système généralisé d'estimation (SGE) de Statistique Canada. Entre autres, il est possible de spécifier plusieurs structures de variance pour le modèle. Cependant les caractéristiques de GREG ne sont pas complètement connues pour ces spécifications. Cette présentation s'intéresse à l'impact de la spécification du paramètre de la structure de variance du modèle. De plus, on présente différents traitements de données aberrantes permettant d'atténuer les effets qu'elles ont lors du calcul de l'estimation et de l'estimation de la variance.*

Éric PELLETIER, Statistique Canada Immeuble R.H. Coats Ottawa, Ontario K1A 0T6  
pelleri@statcan.ca

[MS-101]

#### **Small Area Estimation Using Empirical Bayes Methods**

#### *Estimation dans les petites régions par la méthode de Bayes empirique*

Ève BELMONTE\*, Université Laval

In sampling theory, serious problems occur when estimating parameters using a small sample size, which is often the case in small areas. Henceforth the direct-survey estimators are not reliable anymore.

By fitting a model for the survey estimates, we are able to improve the efficiency of the direct estimates. The Empirical Bayes method offers an interesting alternative by proposing a combined estimator, i.e. a weighted average of a survey and a synthetic estimator. Such new estimators bring the problem of accurately estimating their mean squared error (MSE). Although such estimators already exist in the literature, we will suggest a new conditional model independent MSE estimator.

Finally, we will discuss its characteristics and its relationship with the Prasad-Rao estimator (1990). In conclusion, a numerical

example excerpted from the 1991 census will be presented.

*En échantillonnage, l'estimation de totaux ou de moyennes dans de petites régions est un sérieux problème lorsque la taille de l'échantillon observé est si petite que les estimateurs directs ne conviennent plus. Or, en supposant un modèle pour les estimés des petites régions, on arrive à contourner les problèmes qui surgissent avec les estimés directs. La méthode bayésienne empirique offre une alternative intéressante en proposant un estimateur combiné, soit une moyenne pondérée d'un estimateur direct et d'un estimateur synthétique.*

*Un autre problème est de trouver un bon estimateur pour l'erreur quadratique moyenne (EQM) de tels estimés. Plusieurs estimateurs existent dans la littérature mais on suggère ici un nouvel estimateur conditionnel de l'EQM qui mesure la variabilité par rapport au plan d'échantillonnage. On discute ensuite de ses caractéristiques et de sa relation avec l'estimateur de l'EQM proposé par Prasad et Rao (1990). On présente aussi un exemple numérique avec des données du recensement de 1991.*

Ève BELMONTE, Dép. mathématiques et statistique Pav. Vachon,  
Université Laval Sainte-Foy QC G1K 7P4  
belmonte@math.ulaval.ca

#### [MS-102]

### Robust Estimators by Benchmarking with Constrained Weights

#### *Estimateurs de calage robustes avec poids contraints*

Pierre DUCHESNE\*, Université de Montréal

Outlier issues are discussed in Lee (1995) in the framework of sampling theory. We consider the use of estimators by benchmarking for such a situation. Extension of the estimators by benchmarking of Deville and Särndal (1992) based on the  $QR$  Wright (1983) estimators are obtained. As an application, this class of estimators allows construction of robust estimators by benchmarking. Choosing an adequate metric, we obtain robust weights  $w_k$  which are completely bounded in a prespecified interval. All the robust estimators considered are based on estimators with high breaking points. Particularly, when the metric is quadratic, our proposed estimator extends one proposed by Lee (1991). A brief simulation study illustrates the new methodology.

*Le problème des valeurs aberrantes en théorie des sondages est discuté dans Lee (1995). Nous considérons l'utilisation d'estimateurs de calage dans de telles situations. Une extension de la classe des estimateurs de calage de Deville et Särndal (1992) reposant sur les estimateurs  $QR$  de Wright (1983) est obtenue. Comme application, cette classe d'estimateurs nous permet de considérer des estimateurs de calage robustes. En choisissant une métrique adéquate, nous obtenons des poids robustes  $w_k$  qui sont complètement bornés à un intervalle spécifié à l'avance. Tous les estimateurs robustes considérés reposent sur des estimateurs avec hauts points de rupture. Dans le cas particulier où la métrique choisie est la métrique quadratique, l'estimateur que nous suggérons est une généralisation d'une proposition de Lee (1991). Une brève étude de simulation illustre la nouvelle méthodologie.*

Pierre DUCHESNE, Département de mathématiques et de statistique  
Université de Montréal C.P. 6128, Succursale Centre-ville Montréal

QC H3C 3J7

duchesne@DMS.UMontreal.CA

#### [MS-103]

### Matching Data from the Labour Force Survey with the Census

#### *Appariement de données de l'Enquête sur la population active canadienne à celles du recensement de la population*

Isabelle LÉVESQUE\*, Statistique Canada

Benoit ALLARD, Statistique Canada

Johane DUFOUR, Statistique Canada

A matching project of data from the Labour Force Survey (LFS) and the data from the 1991 population census has been undertaken by Statistics Canada. The main purpose of this project is to carry out some studies on the quality of the LFS's data by using information from the 1991 census. The increasing use of the LFS as a general vehicle for surveys on households, the important number of additional surveys attached monthly to the LFS, and new surveys in Statistics Canada are among the reasons that motivate the implementation of this important project. Three studies have been carried out: i) profile comparison of those who answered to that of those who did not ii) a coverage study, and iii) a comparison of answers from the LFS and the census. For the first study, the matching of the two surveys had been performed at the households level. A data bank, containing characteristics of the households who did not answer the LFS and the matched to the census file, has been constructed. As to the two other studies, the matching data procedure has been applied to the individual level and a data bank containing information from the LFS and from the census has been created for matched and non matched individuals. The present paper summarizes the principal results of these three studies.

*Un projet d'appariement entre les données de l'Enquête sur la population active canadienne (EPA) et celle du Recensement de la population canadienne de 1991 a été entrepris à Statistique Canada. Le but principal de ce projet consistait en la réalisation d'un certain nombre d'études sur la qualité des données de l'EPA en utilisant l'information provenant du Recensement de la population canadienne de 1991. L'utilisation grandissante de l'EPA comme véhicule général pour les enquêtes-ménages, le nombre important d'enquêtes supplémentaires qui se greffent à tous les mois à l'EPA ainsi que la réalisation de nouvelles enquêtes à Statistique Canada s'inscrivent au nombre des raisons qui ont motivé la réalisation de ce projet d'envergure. Trois études ont été réalisées : i) une comparaison du profil des répondants et des non-répondants, ii) une étude de la couverture et iii) une comparaison des réponses recueillies par l'EPA et par le Recensement. Pour la première étude, l'appariement entre les deux enquêtes a été effectué au niveau des ménages et une base de données, contenant les caractéristiques des ménages non-répondants de l'EPA qui étaient appariés au fichier du Recensement, a été construite. Pour ce qui est des deux autres études, la procédure d'appariement des données a été appliquée au niveau des personnes et une base de données contenant l'information provenant de l'EPA et du recensement a été créée pour les individus appariés et non appariés. Le*

présent article résume les principaux résultats de ces trois études.

Isabelle LÉVESQUE, Statistique Canada Immeuble R.H. Coats Ottawa, Ontario K1A 0T6  
leveisa@statcan.ca

#### [MS-104]

### Model to Predict the Probability of a Person Having Employment Activity Limitation Using Data from the Health and Activity Limitation Survey (HALS)

*Modèle pour prédire la probabilité d'avoir une limitation d'activité reliée à l'emploi à partir de l'enquête sur la santé et les limitations d'activité*

Daniel HURTUBISE\*, Statistique Canada  
Éric LANGLET, Statistique Canada

Health and Activity Limitation Survey (HALS) was held in 1986 and 1991, but not in 1996 due to budget constraints. Human Resources Development Canada would like to produce tables of persons limited in their activity for their Employment Equity Program. Estimates of the probability of being limited will be based on a model connecting the 1991 HALS employment equity disability indicator with the 1991 Census variables. This model will then be applied to the 1996 Census data to obtain a probability of being disabled for each person in the 2B sample. Different methods to predict probabilities of being disabled are proposed and compared, namely the classification tree analysis, the logistic regression analysis, the discriminant analysis, and a proportion method.

*L'Enquête sur la Santé et les Limitations d'Activités (ESLA) a eu lieu en 1986 et 1991, mais n'a pas pu être menée en 1996 par manque de budget. Développement des ressources humaines Canada voudrait quand même produire des tableaux portant sur les limitations d'activités pour leur programme d'Équité en matière d'emploi. Les estimations des probabilités d'être limité seront basées sur un modèle reliant l'indicateur d'équité en matière d'emploi de limitation d'activité contenu dans l'ESLA 1991 aux variables du recensement de 1991. Ce modèle sera alors appliqué sur les données du recensement de 1996 afin d'obtenir une probabilité d'être limité pour chaque individu de l'échantillon 2B. Différentes méthodes pour prédire les probabilités d'être limité sont proposées et comparées, soit l'analyse par arbre de classification, l'analyse par régression logistique, l'analyse discriminante, et une méthode de proportion.*

Daniel HURTUBISE, Statistique Canada Immeuble R.H. Coats Ottawa, Ontario K1A 0T6  
hurtdan@statcan.ca

## Session 30

### Survey Methods and Biostatistics: New Approaches to Data Analysis

*Méthodologie d'enquête et biostatistique : Nouvelles approches pour l'analyse des données*

#### [MS-105]

### A New Look at Confidence Intervals in Survey Sampling

*Un nouveau regard sur les intervalles de confiance en échantillonnage*

V. P. GODAMBE\*, University of Waterloo

In survey sampling, as in other areas of statistics conventionally, confidence intervals for a parameter are often obtained by inverting the distribution of some approximate pivotal quantity,  $\{(\text{estimate} - \text{parameter}) / (\text{estimated variance})^{1/2}\}$ . Alternatively, estimating function theory suggests a more direct method of constructing pivotal quantity and hence confidence intervals. These alternatives confidence intervals perform much better than the conventional ones, in many simulation studies.

*En théorie de sondage, comme dans d'autres branches de la statistique, les intervalles de confiance pour les paramètres sont souvent obtenus en inversant les distributions de certaines statistiques pivotales, au moins approximativement,  $\{(\text{estimé} - \text{paramètre}) / (\text{variance estimée})^{1/2}\}$ . Alternativement, la théorie des fonctions d'estimation suggère une méthode plus directe pour construire les quantités pivotales et en suite les intervalles de confiance. Plusieurs études par simulation ont montrées que ces intervalles de confiance, obtenus par cette méthode alternative, sont plus performants que ceux obtenus par les méthodes conventionnelles.*

V. P. GODAMBE, Department of Statistics University of Waterloo Waterloo, Ontario N2L 3G1

#### [MS-106]

### Multistage Designs and Estimation in Biostatistics

*Plans à plusieurs degrés et estimation en biostatistique*

Jerry LAWLESS\*, University of Waterloo

Multistage sampling schemes are often used in biostatistics and epidemiology for reasons of economy or convenience. We consider contexts where the relationship between a response variable  $Y$  and a vector of covariates  $x$  is of interest. At the first stage, individuals are selected and certain measurements are taken. The initial sample is then subsampled in subsequent stages and additional measurements are taken on selected individuals. As a result, different amounts of data are missing on different individuals. We consider approaches to estimation in such settings and discuss efficiency.

*Les plans d'échantillonnage à plusieurs degrés sont souvent utilisés en biostatistique et en épidémiologie, et ce pour des raisons d'économie et de commodité. Dans ce travail notre intérêt porte sur la relation entre la variable réponse  $Y$  et un vecteur de covariables  $x$ . À la première étape, les individus sont sélectionnés et certaines mesures sont prises sur eux. L'échantillon initial subit ensuite une séquence de ré-échantillonnages fournissant des mesures additionnelles sur les individus sélectionnés. En conséquence, différentes quantités de données sont manquantes sur plusieurs individus. Nous considérons des approches d'estimation pour de telles situations et nous discutons des problèmes d'efficacité.*

Jerry LAWLESS, Department of Statistics University of Waterloo Waterloo, Ontario N2L 3G1  
jlawless@uwaterloo.ca

[MS-107]

### Design-Free Analysis

#### *Analyse libre du plan d'échantillonnage*

Fritz SCHEUREN\*, Ernst and Young

This paper describes the use of inverse sampling as a tool to free the analyst from the annoying aspects of the sample design; in this example, the analysis of interest is regression. A highly skewed population is generated, similar to populations describing economic or financial information. Three sample designs are simulated: one conventional design and two designs which are, in varying degrees, unconventional. Simple random samples are drawn from these complex samples using the inverse algorithm, and the analysis using the original samples and the analysis using the inverse samples are described and compared. New inverse theoretical results and also new sample design results will be given.

*Ce papier décrit l'utilisation de l'échantillonnage inverse comme moyen de libérer l'analyste des aspects ennuyeux que présente le plan d'échantillonnage ; dans cet exemple, l'analyse qui nous intéresse est la régression. Une population très asymétrique, similaire aux populations décrivant l'information sur l'économie et les finances, est générée. Nous avons utilisé des plans d'échantillonnage pour simuler trois échantillons : un plan d'échantillonnage conventionnel, et deux autres plans non conventionnels, à des degrés différents. Des échantillons aléatoires simples sont tirés de ces échantillons "complexes" utilisant l'algorithme inverse. L'analyse utilisant l'échantillon original et celle utilisant l'échantillon inverse sont décrites et comparées. De nouveaux résultats théoriques ainsi que de nouveaux résultats de l'échantillonnage inverse sont donnés.*

Fritz SCHEUREN, Ernst and Young 1225 Connecticut Avenue, N.W. Washington, DC 20036 U.S.A.  
fritz.scheuren@ey.com

## Session 33

### Topics in Environmental Statistics

#### *Statistique de l'environnement*

[MS-108]

#### **Use of Ratio Estimators to Estimate the Size of Common Eider Populations in Winter**

*L'estimation par le quotient de la taille des populations*

#### ***hivernantes d'Eider à duvet***

Nathalie PLANTE\*, Service canadien de la Faune

Daniel BORDAGE, Service canadien de la Faune

André BOURGET, Service canadien de la Faune

Sylvain PARADIS, Service canadien de la Faune

Monitoring of Common Eider numbers is essential for good population management. Before 1989, only visual estimates were used to assess the size of flocks located during aerial surveys. We now recognize that it is necessary to obtain estimates of population size for which we can evaluate the precision. In the winter of 1989, biologists from the Canadian Wildlife Service and Parks Canada tested a new survey method. The visual estimates were supplemented by photographs of eider groups which were later counted under a binocular stereoscope. However, good quality prints of whole flocks were often difficult to obtain. Therefore, we estimated eider numbers using a ratio estimator that combined photo counts and visual estimates. In addition, the use of a capture-recapture technique allowed an estimate of the number of eiders missed by the survey.

This survey method and its difficulties are discussed. One application of the ratio estimator is presented using two situations where sample selection is not controlled. The question of estimating the number of unobserved eiders is also addressed.

*Un suivi des effectifs d'Eider à duvet est essentiel à la bonne gestion des populations. Avant 1989, on estimait visuellement la taille des bandes observées lors d'inventaires aériens. On reconnaît depuis la nécessité d'obtenir des estimations de la taille de la population dont on puisse évaluer la précision. À l'hiver 1989, des biologistes du Service canadien de la faune et de Parcs Canada expérimentent une nouvelle méthode d'inventaire. Aux estimations visuelles s'ajoutent des photos des bandes dont on prévoit compter les eiders au binoculaire. Or il est difficile d'obtenir des photos de qualité pour toutes les bandes. On a par conséquent recours à la méthode d'estimation par le quotient combinant photos et estimations visuelles. En complément, une méthode de capture-recapture est utilisée pour estimer le nombre d'eiders étant passés inaperçus lors de l'inventaire.*

*La méthode d'inventaire et les difficultés qu'elle comporte seront exposées. Une application de la méthode d'estimation par le quotient sera présentée dans deux situations où la sélection de l'échantillon n'est pas contrôlée. La question de l'estimation du nombre d'eiders non observés sera finalement abordée.*

Nathalie PLANTE, Service canadien de la faune, 1141 route de l'Église, C.P. 10100, Sainte-Foy, Québec G1V 4H5  
Nathalie.Plante@ec.gc.ca

[MS-109]

#### **ANOVA Involving Qualitative and Quantitative Factors: Environment Impact on Mussel Growth**

*Croisement de facteurs qualitatif et quantitatif en ANOVA : Impact du milieu sur la croissance des moules*

Gaétan DAIGLE\*, Université Laval

E. BOURGET, Université Laval

L.-P. RIVEST, Université Laval

We examined the potential influence of coastal heterogeneity on mussel growth. Three piers of different lengths located on the south shore of the St. Lawrence Estuary were selected for the study. At each station, cages were placed with mussels at six experimental sites: upstream, apex, and downstream of the piers (heterogeneous zone), and in analogous positions along a nearby homogeneous control zone. Mussels growth were obtained after a 3 month immersion period. In the control zone, the upstream and downstream positions or not really different. They are potentially different only in heterogeneous zones. This means that to compare the downstream with the upstream position only data collected in a heterogeneous zone should be used. New sums of squares for the ANOVA table, based on meaningful contrasts, are proposed for this experimental setup.

*Nous examinons l'effet potentiel de l'hétérogénéité côtière sur la croissance des moules. Trois quais de différentes longueurs situés sur la rive sud de l'Estuaire du St-Laurent ont été sélectionnés pour l'étude. A chacune des stations, des cages remplies de moules ont été placées à 6 sites expérimentaux : en amont, à la pointe, et en aval des quais (zone hétérogène), puis à des positions analogues le long d'une zone contrôle relativement homogène. La croissance des moules a été observée après une période d'immersion de 3 mois. Dans la zone contrôle, les positions "amont" et "aval" ne sont pas réellement différentes. Elles sont potentiellement différentes seulement dans les zones hétérogènes. Ainsi, pour comparer la position "amont" avec "aval", seules les données provenant de zones hétérogènes doivent être utilisées. De nouvelles sommes de carrés dans la table d'ANOVA, construites à partir de contrastes orthogonaux, sont proposées pour analyser les données conformément à cette structure expérimentale.*

Gaétan DAIGLE, Dép. de mathématiques et de statistique Université Laval Sainte-Foy QC G1K 7P4  
gd@mat.ulaval.ca

#### [MS-110]

### Network of Monitoring Stations for Detecting Trends in the Quality of Water in Lakes in the Presence of Spatial Correlation

#### *Rationalisation d'un réseau d'échantillonnage pour la détection de tendance dans la qualité de l'eau des lacs en présence d'autocorrélation spatiale*

Claude LABERGE\*, STATEX Québec

While studying the impairing or the recovery of lacustrine water quality at the regional scale, a network of monitoring stations located on several lakes of a given region can be established to accelerate the detection of trends using the compounded (or combined) information. However, the use of several lakes in a region may produce spatially correlated observations if lakes are not chosen according to a probability-based sampling scheme. In a single-stage cluster sampling, the use of lakes too close from one another induces spatially correlated observations resulting in an underestimation of the variance causing too many trends to be detected. The effect on trend detection power is studied in a network rationalization point of view. The rationalization process is illustrated using a real network where

we show that a reduction of more than 50% only a slight decrease in power.

*Dans l'étude de l'amélioration ou de la détérioration de la qualité de l'eau des lacs au niveau régional, un réseau recueillant de l'information sur plusieurs lacs à l'intérieur de la région en question peut être établi afin d'accélérer la détection de tendance. Cependant, l'utilisation de plusieurs lacs dans la même région peut introduire de l'autocorrélation spatiale lorsque les lacs ne sont pas sélectionnés aléatoirement. Dans le cas d'un échantillonnage par grappe, l'utilisation de lacs très rapprochés amène la présence d'autocorrélation spatiale, ce qui a pour effet de sous-estimer la variance de la population et de détecter un trop grand nombre de tendances régionales. L'effet de l'autocorrélation spatiale sur la puissance est étudié dans le but de rationaliser un réseau de détection de tendance. La rationalisation est présentée pour un réseau réel pour lequel une diminution de 50% qu'une faible diminution de puissance de détection de tendance.*

Claude LABERGE, STATEX, 3332 de la Paix Sainte-Foy QC G1X 3W6

claudelaberge@sympatico.ca

## Session 35

### Theory and Applications of Classification

#### *Théorie et applications en classification*

#### [MS-111]

### Phylogenetics and Genomics

#### *Phylogénétique et Génomique*

David SANKOV\*, Université de Montréal

During biological evolution, inter- and intrachromosomal exchanges of chromosomal fragments disrupt the order of genes on a chromosome and, for multichromosomal genomes, the partition of genes among these chromosomes.

When comparing two evolutionarily diverging species, any (maximal) contiguous region of the genome in which gene content and order have been conserved in both species is called a *conserved segment*. Between any two adjacent conserved segments is a *breakpoint*. The number of conserved segments increases as they are disrupted by new events, so that they tend to become shorter over time. The number of chromosomal segments conserved during the divergence of two species, or equivalently, the number of breakpoints, can be used as a rough measure of their genomic distance.

Two approaches, the algorithmic and the statistical, have been taken to the reconstruction of genomic history based on the comparison of chromosomal gene content and order in two or more genomes. The first attempts to infer a most economical sequence of rearrangement events to account for the differences among the genomes, based only on the breakpoints, and neglects the contents of conserved segments. The second approach ignores the details of rearrangement history and assumes that a random model (the Nadeau-Taylor model) accounts for the differences in chromosomal gene content and order.

In this paper, we discuss the strengths and weaknesses of the two approaches.

In the phylogenetic context, a compromise approach can be adopted, algorithmic, but not attempting to infer precise details of hypothesized evolutionary events. This leads to a new, tractable, problem, the *median problem for breakpoints*. We give a solution to the version of this problem where all genomes have the same gene content, and extend it to the case where the median and other genomes involved may have partially different gene sets.

*Au cours de l'évolution biologique, des échanges, au niveau des chromosomes, de fragments chromosomiques perturbent l'ordre des gènes dans un chromosome et, pour un génome multichromosomal, il perturbe également la répartition des gènes parmi ces chromosomes.*

*Lorsqu'on compare deux espèces à évolution divergente, toute région (maximale) contiguë du génome dans laquelle le contenu et l'ordre des gènes ont été conservés dans les deux espèces est appelée segment conservé. Entre deux segments conservés adjacents se trouve un point de rupture. Le nombre de segments conservés augmente dans le temps lorsque de nouveaux événements viennent briser les segments existants. Les segments conservés ont alors tendance à se rétrécir au fil du temps. Le nombre de segments chromosomiques conservés au cours de la divergence de deux espèces, ou de manière équivalente, le nombre de points de rupture, peut être utilisé comme une mesure grossière de leur distance génomique.*

*Deux approches, algorithmique et statistique, basées sur la comparaison de contenus et d'ordre de gènes chromosomiques dans deux ou plusieurs génomes sont utilisées pour la reconstruction de l'histoire génomique. La première tente de trouver la suite la plus économique des événements réarrangés pour expliquer les différences entre les génomes, basée uniquement sur les points de rupture, et néglige le contenu des segments conservés. La seconde approche ignore les détails des arrangements historiques et suppose un modèle aléatoire (le modèle de Nadeau-Taylor) qui tient compte des différences de contenus et des ordres dans les gènes chromosomiques. Dans ce papier, nous discutons les forces et les faiblesses des deux approches.*

*Dans le contexte phylogénétique, une approche de compromis peut être adoptée. Cette dernière est algorithmique; elle ne prétend pas donner des détails précis concernant des événements hypothétiques de l'évolution. Ceci amène à un nouveau problème que l'on peut résoudre, appelé le problème médian des points de rupture. Nous donnons une solution à une version de ce problème où tous les génomes ont le même contenu génétique, et ensuite nous l'étendons au cas de génome médian et d'autres génomes impliqués ayant des ensembles de gènes partiellement différents.*

David SANKOV, C.R.M. Université de Montréal C.P. 6128, Succursale Centre-ville Montréal QC H3C 3J7

[MS-112]

## Classification Tests with Simply, Doubly, and Triply Repeated Measures

*Tests de classifications avec mesures répétées simples, doubles et triples*

Pierre DUTILLEUL\*, McGill University

In this talk, I shall first recall the concept of repeated measures and the contexts and dimensions in which they may arise. This reminder will be followed by a brief review of the literature of repeated

measures analysis of variance under the model of the vector normal distribution (Box 1954a, b; Huynh & Feldt 1970; Rouanet & Lepine 1970; Crowder & Hand 1990). Extensions of the relevant theory to the matrix normal model (Byakagaba 1987; Dutilleul 1994; Dutilleul & Pinel-Alloul 1996) and a multi-dimensional normal distribution called the tensor normal model (Dutilleul 1990) will then be considered, discussed and compared with the vector normal model. The cases of simply, doubly and triply repeated measures, possibly in combination with other classification factors (i.e., treatment or block), will be developed under each model. The theory will be illustrated by biological examples.

*Dans cette communication, je parlerai tout d'abord du concept de mesures répétées et des contextes et dimensions dans lesquels elles peuvent se rencontrer. Ce rappel sera suivi d'une brève revue de la littérature de l'analyse de variance avec mesures répétées sous le modèle du vecteur aléatoire normal (Box 1954a, b; Huynh & Feldt 1970; Rouanet & Lepine 1970; Crowder & Hand 1990). Les extensions de la théorie aux modèles de la matrice aléatoire normale (Byakagaba 1987; Dutilleul 1994; Dutilleul & Pinel-Alloul 1996) et d'une distribution multi-dimensionnelle appelée tenseur normal (Dutilleul 1990) seront ensuite considérées, discutées et comparées au modèle du vecteur normal. Les cas de mesures répétées simples, doubles et triples, possiblement en combinaison avec d'autres facteurs de classification (i.e., traitement ou bloc), seront développés sous chaque modèle. La théorie sera illustrée par des exemples biologiques.*

### REFERENCES :

- Box, G.E.P. (1954a). Some theorems on quadratic forms applied in the study of analysis of variance problems. I. Effect of inequality of variance in the one-way classification. *Annals of Mathematical Statistics*, 25 :290-302.
- Box, G.E.P. (1954b). Some theorems on quadratic forms applied in the study of analysis of variance problems. II. Effects of inequality of variance and of correlation between errors in the two-way classification. *Annals of Mathematical Statistics*, 25 :84-498.
- Byakagaba, M. (1987). Apport de la matrice normale aux problèmes d'analyse de variance et des mesures répétées. Doctoral dissertation, Faculty of Science, Université catholique de Louvain, Louvain-la-Neuve, Belgium.
- Crowder, M.J. & Hand, D.J. (1990). *Analysis of Repeated Measures*. Chapman and Hall, London.
- Dutilleul, P. (1990). Apport en analyse spectrale d'un périodogramme modifié et modélisation des séries chronologiques avec répétitions en vue de leur comparaison en fréquence. D.Sc. thesis, Department of Mathematics, Université catholique de Louvain, Louvain-la-Neuve, Belgium.
- Dutilleul, P. (1994). Maximum likelihood estimation for the matrix normal distribution. Technical Report 94-09, Department of Mathematics and Statistics, McGill University, Montréal, Québec, Canada.
- Dutilleul, P. & Pinel-Alloul, B. 1996. A doubly multivariate model for statistical analysis of spatio-temporal environmental data. *Environmetrics*, 7 :551-566.
- Huynh, H. & Feldt, S. (1970). Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. *Journal of the American Statistical Association*, 65 :1582-1589.
- Rouanet, H. & Lepine, D. (1970). Comparison between treatments in a repeated-measures design : ANOVA and multivariate methods. *British Journal of Mathematical and Statistical Psychology*, 23 :147-

163.

Pierre DUTILLEUL, Department of Plant Science Agricultural and Environmental Sc. Macdonald Campus of McGill Univ. 21, 111 Chemin du Bord de l'Eau Ste-Anne-de-Bellevue QC H9X 3V9  
CYDP@Musica.McGill.CA

**[MS-113]****How to Validate our Classifications?*****Comment valider nos classifications ?***

François-Joseph LAPOINTE\*, Université de Montréal

Several methods are currently available to validate the dendrograms obtained with clustering algorithms and additive trees (cladograms) derived with phylogenetic reconstruction methods. First, I will present some of the commonly used techniques to compare a tree with the data it is derived from (internal validation), or compare a tree to another tree or to more than one (external validation). I will also discuss the advantages of performing combined (total evidence) versus separate analyses (consensus) of independent data sets for validation purposes. A stepwise validation procedure defined across all levels of comparison will be introduced, along with a corresponding statistical test. A tree will be said to be globally validated only if it satisfies all the tests. An application to the phylogeny of kangaroos will be presented to illustrate the stepwise procedure.

*Il existe toute une panoplie de méthodes permettant de tester et de valider les dendrogrammes obtenus à l'aide d'algorithmes de classification automatique, ainsi que les arbres additifs (cladogrammes) obtenus par des algorithmes de reconstructions phylogénétiques. Tout d'abord, j'aborderai les techniques qui servent à valider un arbre (ultramétrique ou non) en le comparant aux données à partir desquelles cet arbre fut construit (validation interne). Ensuite, je présenterai les différentes méthodes pour comparer des dendrogrammes ou cladogrammes (validation externe). Je discuterai des avantages et des inconvénients de "l'approche combinée" et de "l'approche consensus" dans une perspective de validation. Finalement, un méthode séquentielle sera proposée pour tenir compte de tous ces aspects à la fois. Un arbre sera défini comme étant "globalement valide" si tous les tests sont significatifs. Cette méthode globale sera utilisée pour valider la phylogénie des kangourous.*

François-Joseph LAPOINTE, Département de Sciences biologiques, Université de Montréal C.P. 6128, Succursale Centre-ville Montréal, Québec, Canada, H3C 3J7  
lapoinf@ere.umontreal.ca

**Session 36****Probability and Mathematical Statistics*****Probabilité et statistique mathématique*****[MS-114]****Some Quadratic Estimating Function For the Analysis of Proportions*****Quelques fonctions quadratiques d'estimation pour l'analyse des proportions***

S.R. PAUL\*, University of Windsor

The purpose of this paper is to discuss a number of quadratic estimating functions and to propose a new one for the joint estimation of regression and dispersion parameters in the Analysis of Proportions. Large sample efficiency results will be presented.

*L'objet de ce papier est de discuter un certain nombre de fonctions quadratiques d'estimation et de proposer une nouvelle fonction quadratique d'estimation pour l'estimation conjointe des paramètres de la régression et de dispersion dans l'analyse des proportions. Des résultats d'efficacité asymptotique seront présentés.*

S.R. PAUL, Department of Mathematics & Statistics University of Windsor Windsor, Ontario, N9B 3P4  
mthsta2@uwindsor.ca

**[MS-115]****On Approximated Moments of Order Statistics from Progressively Censored Samples*****Les moments approchés des statistiques d'ordre d'échantillons censurés progressivement***

Uditha BALASOORIYA\*, Memorial University of Newfoundland

Sutaip S. SAW, Memorial University of Newfoundland

In this paper, we examine approximate moments of order statistics from progressively censored samples from normal and extreme-value distributions. Different approximations to these moments and their effects on failure-censored reliability sampling plans and best linear unbiased estimates are discussed.

*Dans ce papier, nous examinons les moments approchés de statistiques d'ordre à partir d'échantillons censurés progressivement de lois normale et de valeurs extrêmes. Des approximations différentes de ces moments et leurs effets sur les plans d'échantillonnages pour la fiabilité échec-censure et les meilleures estimations linéaires sans biais sont discutés.*

Uditha BALASOORIYA, Dep. of Mathematics and Statistics Memorial University St John's Nfld CANADA A1C 5S7  
ecsub@nus.edu.sg

**[MS-116]****The Gibbs Sampler Converges Slowly !*****L'échantillonneur de Gibbs converge lentement !***

Claude BÉLISLE\*, Université Laval

We consider the Gibbs sampler as a tool for generating an absolutely continuous probability measure  $\pi$  on  $R^d$ . When an appropriate irreducibility condition is satisfied, the Gibbs Markov chain  $(X_n; n \geq 0)$  converges in total variation to its target distribution  $\pi$ . Sufficient conditions for geometric convergence have been given by various authors. Here, we illustrate, by means of simple examples, how slow the convergence can be. In particular, we show that given a sequence of positive numbers decreasing to zero, say  $(b_n; n \geq 1)$ , one can construct an absolutely continuous probability measure  $\pi$  on  $R^d$  which is such that the total variation distance between  $\pi$  and the distribution of  $X_n$  converges to 0 at a rate slower than that of the sequence  $(b_n; n \geq 1)$ . This can even be done in such a way that  $\pi$  is the uniform distribution over a bounded connected open subset of  $R^d$ . Our results extend to Hit-and-Run samplers with direction distributions having supports with symmetric gaps.

*Nous considérons l'échantillonneur de Gibbs comme outil pour engendrer une loi absolument continue sur  $R^d$ , disons  $\pi$ . Lorsqu'une certaine condition d'irréductibilité est satisfaite, la chaîne de Markov  $(X_n; n \geq 0)$  associée à l'échantillonneur de Gibbs converge en variation totale vers sa loi cible  $\pi$ . Plusieurs auteurs ont donné des conditions suffisantes pour assurer que le taux de convergence soit géométrique. Dans le présent travail, nous illustrons, au moyen d'exemples simples, la lenteur avec laquelle cette convergence peut avoir lieu. En particulier, nous montrons que pour toute suite de nombres positifs  $(b_n; n \geq 1)$  qui converge vers 0, il existe dans  $R^d$  un ensemble ouvert, borné et connexe, disons  $G$ , qui est tel que si  $\pi$  est la loi uniforme sur  $G$ , et si  $(X_n; n \geq 0)$  est la chaîne de Markov associée à l'échantillonneur de Gibbs pour  $\pi$ , alors la distribution de  $X_n$  converge vers 0 plus lentement que la suite  $(b_n; n \geq 1)$ . Nos résultats se généralisent à l'échantillonneur Hit-and-Run lorsque celui-ci possède une distribution de direction avec un trou symétrique.*

Claude BÉLISLE, Dép. mathématiques et statistique Pav. Vachon, bureau 1056 Université Laval Sainte-Foy QC G1K 7P4  
belisle@mat.ulaval.ca

[MS-117]

**Bootstrapping the Convex Hull**

***La méthode du bootstrap appliquée à l'enveloppe convexe***

Mahmoud ZAREPOUR\*, University of Ottawa

The convex hull of a multivariate sample has a number of statistical applications. If the distribution of observations satisfies a multivariate regular variation condition then the asymptotic distribution of convex hulls can be modeled by using point processes. The limiting distributions can not be simplified and are extremely difficult to use. We show the bootstrap is asymptotically valid if the resampling sample size is  $m = o(n)$ , where  $n$  is the original sample size. We use this to also provide an asymptotically valid bootstrap for the area and perimeter of the convex hulls. In this case, (with the additional assumption of a circularly symmetric distribution) an alternative Monte Carlo method is studied, and it is shown that this method based on randomly rotated points is consistent.

*L'enveloppe convexe d'un échantillon multivarié a plusieurs applications en statistique. Si la distribution des observations satisfait une condition de variation régulière alors la distribution asymptotique d'une enveloppe convexe peut être modélisée en utilisant des processus ponctuels. Les distributions limites ne peuvent pas être simplifiées et sont extrêmement difficiles à utiliser. Nous montrons que le bootstrap est asymptotiquement valide si la taille  $m$  du sous échantillon de l'échantillon d'origine, de taille  $n$ , satisfait  $m = o(n)$ . Nous utilisons ceci pour aussi fournir un bootstrap asymptotiquement valide pour l'aire et le périmètre d'une enveloppe convexe de hull. Dans ce cas, (avec des hypothèses supplémentaires d'une distribution circulaire symétrique) une méthode alternative de Monte Carlo est étudiée. Nous montrons aussi que cette méthode, qui est basée sur les points subissant des rotations aléatoires, est convergente.*

Mahmoud ZAREPOUR, University of Ottawa Mathematics and Statistics department 585 King Edward Ave. Ottawa, Ontario K1N 6N5 Canada.  
zarepour@notiid.mathstat.uottawa.ca

[MS-118]

**The Vervaat Process: Theory and Applications**  
***Le processus de Vervaat : théorie et applications***

Ričardas ZITIKIS\*, University of Ottawa

When investigating asymptotic properties of the so-called Lorenz process, M. Csörgö and R. Zitikis (1996) [Strassen's LIL for the Lorenz curve *J. Multivariate Analysis*, **58**, 1-12] introduced and used in a decisive way the (general) Vervaat process. Since then the Vervaat process has appeared of fundamental importance when constructing confidence bands for the Lorenz, Bonferroni, generalized Lorenz, total time on test, Goldie and other curves of importance in Econometrics, Empirical Reliability, Social and other Sciences. Various asymptotic properties of the Vervaat process have also been established. During the talk we shall give a survey of these developments that have been carried out by M. Csörgö and R. Zitikis at Carleton University (Ottawa) together with J.L. Gastwirth (George Washington University, Washington D.C.)

*Lors de la recherche des propriétés asymptotiques du processus de Lorenz, M. Csörgö et R. Zitikis (1996) [la loi du logarithme itéré de Strassen pour la courbe de Lorenz *J. Multivariate Analysis*, **58**, 1-12] ont introduit et utilisé de manière décisive le processus (général) de Vervaat. Depuis, le processus de Vervaat est d'une importance capitale dans la construction des bandes de confiance pour les courbes de Lorenz, de Bonferroni, de Lorenz généralisé, du temps total en test, de Goldie et pour d'autres courbes importantes en économétrie, en fiabilité empirique, en sciences sociales et autres. Plusieurs propriétés asymptotiques du processus de Vervaat ont été établies. Durant la présentation nous parlerons des développements réalisés par Csörgö et R. Zitikis à l'université de Carleton (Ottawa) conjointement avec J.L. Gastwirth (George Washington University, Washington D.C.).*

Ričardas ZITIKIS, University of Ottawa 585 King Edward Station A, P.O. Box 450 Ottawa, Ontario K1N 6N5



[MS-119]

**Non Parametric Estimation of Functions Associated to a Copula***Estimation non-paramétrique de fonctions associées à une copule*

Abderrahmane CHAKAK\*, Université Abdelmalek Essaadi

M. EZZERG, Université Abdelmalek Essaadi

L. IMLAHI, Université Abdelmalek Essaadi

To a copula are associated the quantile function  $\psi(w, u)$ , whose graph, for  $w$  fixed, is the section  $\{(u, v) : C(u, v) = w\}$  and the generating quantile function  $K(w)$  defined as the cdf of the random variable  $Z = C(U, V)$ , where  $U, V$  are uniform  $(0, 1)$  jointly distributed as  $C$ . Some structural properties of the quantile functions and their relation to the Pearson correlation coefficient are presented. When  $C$  is absolutely continuous, an expression for  $K(w)$  and for its pdf  $k(w)$  are given. Some convergence properties for both the quantile function and the generating quantile function are established. Algorithms for simulating pseudo-observations for the copula and for the generating quantile function are given. For random sample, we describe an algorithm that provides the quantile function associated to the empirical cdf. Genest et al (1993) develop a non parametric estimator for the generating quantile function and evaluate its asymptotic variance while they were mainly interested to Archimedian copulas. Using  $\psi(\cdot, \cdot)$  and  $K$  we are able to apply their procedure to other families of copulas. Moreover, we derive an estimator for the conditional expectation involved in the expression of its variance. We will also discuss some use of the generating quantile functions.

*A une copule  $C$  nous associons la fonction quantile  $\psi(w, u)$  dont le graphe, pour  $w$  fixe, est la section  $\{(u, v) : C(u, v) = w\}$ , et la fonction génératrice des quantiles  $K(w)$  définie comme étant la fonction de répartition de la variable aléatoire  $Z = C(U, V)$ , où  $U, V$  sont uniformes  $(0, 1)$  de distribution conjointe  $C$ . Quelques propriétés structurales des fonctions quantiles et leurs relations au coefficient de corrélation de Pearson sont présentées. Quand  $C$  est absolument continue, les expressions de  $K(w)$  et de sa densité  $k(w)$  sont données. Quelques propriétés de convergence pour les fonctions quantiles et pour la fonction génératrice des quantiles sont établies. Deux algorithmes pour simuler des pseudo-observations à partir de  $C$  et à partir de  $K$  sont donnés. Pour un échantillon aléatoire, de  $C$ , nous fournissons un algorithme qui produit la fonction quantile associée à la ecdf  $C_n$ . Genest et al (1993) ont développé un estimateur non paramétrique pour la fonction génératrice des quantiles et ont évalué sa distribution asymptotique alors qu'ils s'intéressaient aux copules Archimédiennes. Utilisant les fonctions  $\psi(\cdot, \cdot)$  et  $K$ , nous sommes capable d'étendre leur procédure à d'autres familles de copules. En plus, nous fournissons un estimateur de l'espérance conditionnelle figurant dans l'expression de sa variance. Nous discuterons de quelque applications de la fonction génératrice des quantiles.*

Abderrahmane CHAKAK, Dép. de mathématiques, Faculté des sciences B.P. 2121 Tetouan Maroc  
chakak@mat.ulaval.ca

## Session 37

## Survey Methods : Survey Design

*Méthodologie d'enquête : Conception d'enquêtes*

[MS-120]

**An Example of the Challenges of the Special Business Surveys***Un exemple des défis des enquêtes spéciales d'entreprises*

Tony LABILLOIS\*, Statistique Canada

A group of people in Statistics Canada works to implement special business survey projects and to integrate projects that are already existing in this domain. These projects are carried out for clients who wish to get specific information not available within the standard releases of the statistical agency. Some survey projects, integration of meta-data, and matching micro data are carried out over short periods, with limited resources, on a cost recovery basis. The results of these various activities that involves many participants must after all meet the clients' statistical needs while respecting the policies of the agency, specially in terms of quality, methodology, confidentiality, bilinguism, and availability to the general public. Surveys on salary and treatment by profession have been in increasing demand for some years. These surveys are used to illustrate the challenges related to cost, timing and quality to be met for various implementation procedure.

*Il existe à Statistique Canada un groupe qui travaille à réaliser essentiellement des projets d'enquêtes spéciales auprès des entreprises et des projets d'intégration de données déjà existantes dans ce domaine. Ces projets sont réalisés pour des clients qui désirent obtenir des renseignements spécifiques qui ne sont pas directement disponibles parmi la gamme de produits habituellement diffusée par l'agence statistique. Des projets d'enquêtes, d'intégration de méta-données et d'appariement de micro-données sont ainsi réalisés sur des périodes relativement courtes avec des ressources limitées et avec recouvrement des coûts. Les résultats de ces diverses activités qui nécessitent plusieurs intervenants doivent avant tout satisfaire aux besoins statistiques des clients tout en étant conformes aux diverses politiques de l'agence, notamment en termes de qualité, de méthodologie, de confidentialité, de bilinguisme et de disponibilité au public en général. Les enquêtes sur les salaires et traitements par profession sont en demande croissante depuis quelques années. Ceci permet de comparer diverses approches en relation avec les résultats obtenus. Ces enquêtes seront utilisées pour illustrer les défis de coûts, d'échéancier et de qualité auxquels il faut faire face selon les diverses approches choisies pour les réaliser.*

Tony LABILLOIS, Statistique Canada Tunney's Pasture Ottawa, Ontario K1A 0T6  
labiton@statcan.ca

[MS-121]

**Electronic Methods of Data Collection for a Survey**

## Studying Internet Users in Quebec

### *Méthode électronique de collecte des données pour une enquête auprès des internautes québécois*

Eric LACROIX\*, Bureau de la Statistique du Québec

In the spring of 1998, le Bureau de la statistique du Québec, (BSQ) in association with le Centre francophone de recherche en informatisation des organisations (CEFRIO) and le Réseau Interordinateurs scientifique québécois (RISQ) is conducting a survey studying the practices and characteristics of Internet users in Quebec. One can easily understand the complexity involved in creating a survey frame that would present extensive coverage of this target population : many Internet users have multiple email addresses; some don't have email addresses; the ISPs (Internet Service Providers) are sometimes reluctant to give away the names and addresses of their customers; some Internet users are customers of ISPs who operate outside Quebec; etc.

The presentation will focus on the characteristics of the survey methodology that was used to overcome this problem, a combination of traditional RDD telephone sampling and Web-based data collection. Among other things, the follow-up procedures will be discussed - to get all the selected Internet users to complete the questionnaire only once - as well as the confidentiality aspects surrounding the data collection.

*Au printemps 1998, le Bureau de la statistique du Québec, (BSQ) en collaboration avec le Centre francophone de recherche en informatisation des organisations (CEFRIO) et le Réseau Interordinateurs scientifique québécois (RISQ) mène une enquête qui a pour objet d'étudier les habitudes et les caractéristiques des internautes québécois. On peut concevoir la difficulté d'établir une base de sondage qui couvrirait de façon exhaustive la population visée : plusieurs internautes ont de multiples adresses de courrier électronique ou n'en ont aucune, les fournisseurs de connectivité sont réticents à donner les coordonnées de leur abonnés, certains québécois font affaire avec des fournisseurs situés à l'extérieur du Québec, etc.*

*La présentation va résumer les principales caractéristiques de la méthodologie de l'enquête qui a été retenue pour contourner ce problème, soit une combinaison du sondage téléphonique probabiliste traditionnel et de collecte des données sur le World Wide Web. On y discutera, entre autres, des efforts requis pour assurer le suivi des répondants - que tous les internautes sélectionnés répondent une seule fois -, des aspects de la confidentialité, etc.*

Eric LACROIX, Statisticien-méthodologiste Bureau de la statistique du Québec 200 Chemin Sainte-Foy, Québec G1R 5T4  
Eric.Lacroix@bsq.gouv.qc.ca

### [MS-122]

## Workplace and Employee Survey Methodology Overview

### *Enquête sur le milieu de travail et les employés : aperçu de la méthodologie*

B. LAPIERRE\*, Statistique Canada

Z. PATAK, Statistique Canada

J. MOREL, Statistique Canada

The Workplace and Employee Survey (WES) is a new Statistics Canada longitudinal survey. It has been designed to explore a broad

range of issues relating to employers and their employees. Relationships among competitiveness, innovation, technology use and human resource management represent the focal points on the employer side, while those among technology use, training, job stability and earnings are studied on the employee side. The longitudinal nature of the survey will allow to measure changes over time with respect to those different issues. During this presentation, we will present the methodology developed for WES. We will discuss the survey content, the survey frame and the collection methods. Moreover, we will describe the stratified multi-stage sample design as well as the estimation and imputation methods to be used.

*L'Enquête sur le milieu de travail et les employés (EMTE) est une nouvelle enquête longitudinale de Statistique Canada. Elle a pour objectif d'explorer un large éventail de questions reliées aux employeurs et à leurs employés. Les relations entre la compétitivité, les innovations, l'utilisation de la technologie et la gestion des ressources humaines représentent le centre d'intérêt du côté de l'employeur tandis que celles entre l'utilisation de la technologie, la formation, la stabilité d'emploi et les revenus sont étudiées du côté de l'employé. La nature longitudinale de l'enquête permettra de mesurer les changements au cours du temps de ces aspects de l'économie. Lors de cette présentation, on décrira la méthodologie développée pour l'EMTE. On discutera du contenu de l'enquête, de la base de sondage et de la méthode de collecte. On décrira, de plus, le plan d'échantillonnage à plusieurs degrés ainsi que les méthodes d'estimation et d'imputation à être utilisées.*

B. LAPIERRE, Statistique Canada Tunney's Pasture Ottawa, Ontario K1A 0T6  
lapibru@statcan.ca

### [MS-123]

## Treatment of Enterprise Restructuring in Business Surveys

### *Traitement des restructurations d'entreprises dans les panels économiques*

Pierre LAVALLÉE\*, Statistique Canada

Peter STRUIJS, Eurostat

For business panels, follow-up through time can be done at different levels of statistical units like, for example, the enterprise or the location (or local unit). The choice of the longitudinal unit then depends on factors such as the panel objectives. Whatever the chosen type of unit, there is unfortunately no unique definition of the longitudinal enterprise or the longitudinal location. One reason for this problem is that enterprises, for example, can merge (concentration) or split (deconcentration). This complicates the longitudinal analysis because it supposes generally that the units are indivisible and that they cannot merge (like individuals, for example). It is therefore important to obtain a definition for the longitudinal enterprise that treats the cases of concentrations and deconcentrations of enterprises as well as possible. During the presentation, we will first present the usual case where the follow-up of enterprises through time is based on the use of the Business Number (BN). Second, we will discuss the other case where the follow-up is not solely based on the BN, but

by using other criteria. This forces then the use of methods to handle concentrations and deconcentrations. Third, we will propose four approaches to solve this problem of concentrations and deconcentrations together with a comparison of their advantages and disadvantages. Finally, we will give an example related to the Workplace and Employment Survey conducted by Statistics Canada.

*Dans les panels économiques, le suivi au cours du temps peut se faire au niveau de différentes unités statistiques comme, par exemple, l'entreprise ou l'emplacement (ou unité locale). Le choix de l'unité longitudinale dépend alors de certains facteurs dont notamment les objectifs du panel. Quel que soit le type d'unité choisi, il n'existe malheureusement pas de définition unique de l'entreprise longitudinale ou de l'emplacement longitudinal. Une des raisons de ce problème est que les entreprises, par exemple, peuvent se fusionner (concentration) ou se scinder (déconcentration); ce qui complique l'analyse longitudinale parce que cette dernière suppose généralement que les unités sont indivisibles et ne peuvent se fusionner (comme les individus, par exemple). Il importe donc de chercher à obtenir une définition de l'entreprise longitudinale qui traite le mieux possible les cas de concentrations et de déconcentrations d'entreprises. Au cours de la présentation, on traitera premièrement du cas courant où le suivi des entreprises au cours du temps est basé sur l'utilisation du Numéro d'entreprise (NE). Dans un deuxième temps, on abordera le cas où le suivi ne s'effectue pas uniquement à partir du NE, mais en utilisant d'autres critères. Ces derniers obligent alors le choix de méthodes pour traiter les concentrations et déconcentrations. Troisièmement, on proposera quatre approches pour traiter ce problème des concentrations et déconcentrations avec une comparaison des avantages et désavantages de celles-ci. On donnera finalement un exemple tiré de l'Enquête sur les lieux de travail et les employés effectuée par Statistique Canada.*

Pierre LAVALLÉE, Statistique Canada Tunney's Pasture Ottawa, Ontario K1A 0T6  
plavall@statcan.ca

#### [MS-124]

### **Redesign of the Survey of Household Spending : Increase in Data Quality and Redistribution of Response Burden**

*Remaniement de l'enquête sur les dépenses des ménages : amélioration de la qualité de l'information retenue et redistribution du fardeau de réponse*

Sophie ARSENAULT\*, Statistique Canada

Johanne TREMBLAY, Statistique Canada

The Survey of Household Spending, formerly known as the Family Expenditure Survey, collects detailed information on household expenditures with a 12 month retrospective interview. In the past, due to the level of expenditure detail and the length of the reference period, the average interview took two hours and thirty minutes. This resulted in too much response burden for individual respondents and may in turn have affected the quality of data collected from the survey. In the context of a major redesign, necessitated by the requirements of a major project to improve provincial economic statistics, the collection methodology of this survey is being revised. The approach

being studied would incorporate the use of a diary and interviews with various reference periods that would distribute the response burden imposed to respondents and increase data quality by reducing the length of the reference period. This new collection methodology will be presented and the major methodological issues that it raises will be discussed.

*L'enquête sur les dépenses des ménages, autrefois connue sous le nom d'enquête sur les dépenses des familles, recueille des informations sur les dépenses annuelles des ménages à l'aide d'une entrevue rétrospective de 12 mois. Par le passé, le niveau de détail des dépenses ainsi que la longueur de la période de référence faisaient en sorte que les entrevues duraient en moyenne deux heures et demi. Ceci imposait un fardeau de réponse trop lourd pour le répondant et pouvait affecter la qualité des données recueillies par l'enquête. Dans le cadre d'un remaniement important motivé par les besoins d'un projet majeur d'amélioration des statistiques économiques provinciales, la méthodologie de collecte de cette enquête est révisée. L'approche étudiée incorporerait l'utilisation d'un journal et d'entrevues de période de référence variables qui distribueraient le fardeau de réponse imposé aux répondants et augmenteraient la qualité des données recueillies en réduisant la période de référence. Cette nouvelle méthodologie de collecte sera présentée et les défis méthodologiques qu'elle soulève seront brièvement discutés.*

Sophie ARSENAULT, Statistique Canada Tunney's Pasture Ottawa, Ontario K1A 0T6  
arsesop@statcan.ca

#### [MS-125]

### **Data Quality of Earning Questions in the Canadian Labour Force Survey**

*Qualité des données sur le revenu pour l'Enquête canadienne sur la population active*

François BRISEBOIS\*, Statistique Canada

C. WALKER, Statistique Canada

Following the 1991 Canadian Census of Population, the Canadian Labour Force Survey underwent its fifth redesign since its introduction in 1952. Among the changes in this redesign, was the implementation of a new questionnaire intended to better reflect today's labour market. The new questionnaire, implemented in January 1997, includes more questions about employment, including among others, questions about earnings. Due to the sensitivity of the topic, these questions are often subject to higher rates of non-sampling error (specifically response and non-response error) than less sensitive questions. This paper will assess the extent of the non-sampling error to the new earning questions.

To assess the magnitude of response errors for the earnings questions, we will analyse some of the new measures available with computer assisted interviewing such as the on-line edit system and any interviewer comments or feedback related to these questions. To assess the extent of non-response errors, we will explore the item non-response rates and refusal rates and compare the characteristics of respondents and non-respondents to search for indications of differential non-response.

Suite au recensement de la population canadienne de 1991, l'Enquête canadienne sur la population active a été soumise à un cinquième remaniement depuis son introduction en 1952. Au nombre des changements apportés lors de ce dernier remaniement, figurait la mise en place d'un nouveau questionnaire visant à mieux refléter les réalités du marché du travail d'aujourd'hui. Le nouveau questionnaire, introduit en janvier 1997, inclut maintenant plus de questions portant sur l'emploi, dont entre autres, des questions portant sur le revenu. Étant donné la nature plutôt délicate de ce sujet, ces questions sont souvent sujettes à des taux d'erreurs non dues à l'échantillonnage (plus spécifiquement des erreurs de réponse et erreurs dues à la non-réponse) plus élevés que ceux associés à d'autres types de questions. Cette étude évaluera l'ampleur de l'erreur non due à l'échantillonnage présente dans les questions portant sur le revenu.

Afin d'évaluer l'importance de l'erreur de réponse associée aux questions sur le revenu, une analyse de nouvelles mesures disponibles grâce à l'interview assistée par ordinateur sera effectuée. Ces mesures sont obtenues en utilisant l'information générée par le système de contrôle en ligne des données de même que par les commentaires et/ou les remarques des intervieweurs. Pour évaluer l'importance de l'erreur due à la non-réponse, quelques approches seront explorées. Par exemple, les taux de non-réponse et de refus à chaque question se rapportant au revenu seront examinés et les caractéristiques entre répondants et non-répondants seront comparées.

François BRISEBOIS, Statistique Canada Tunney's Pasture Ottawa, Ontario K1A 0T6  
brisfra@statcan.ca

#### [MS-126]

### Experience in the Bias Analysis in the Canadian International Travel Survey

#### *Analyse du biais de l'enquête canadienne sur les voyages internationaux*

Stéphane TREMBLAY\*, Statistique Canada

François LAFLAMME, Statistique Canada

The International Travel Survey (ITS) is an ongoing survey conducted by Statistics Canada to provide estimates for travel expenditures and characteristics of international travellers and their trips. It covers both Canadian residents travelling outside Canada and international visitors to Canada. In addition to distributing questionnaires to travellers, Canada Customs helps the ITS collect information on the volume of travellers passing through the Canadian border.

The upcoming changes at Customs, the relatively low return rates of ITS questionnaires and the increasing demands of ITS users have led to a project at Statistics Canada to redesign the ITS. As part of the redesign, a field test was conducted in the summer of 1997 to assess the impact of the various questionnaire distribution methods on the response rates and the quality of survey responses.

This paper will describe this field test, the design of the test and the analysis of the quality of survey responses pertaining to different questionnaire distribution strategies.

*L'Enquête sur les voyages internationaux (EVI) est une enquête continue menée par Statistique Canada à la fois sur les Canadiens voyageant à l'extérieur du pays et sur les étrangers visitant le Canada. L'EVI produit un grand éventail de statistiques sur les dépenses des voyages et sur plusieurs caractéristiques des voyageurs inter-*

*nationaux ainsi que leurs voyages. Les Douanes canadiennes fournissent à l'EVI le volume de voyageurs traversant la frontière canadienne en plus de remettre des questionnaires aux voyageurs.*

*Les changements futurs aux douanes, la proportion relativement faible de questionnaires de l'EVI retournée et le nombre croissant de requêtes par les utilisateurs des données de l'EVI ont conduit Statistique Canada à remanier l'EVI. Au cours de l'été de 1997, une partie de ce remaniement s'est concrétisée par un projet spécial à plusieurs volets. L'objectif principal de ce projet est d'évaluer l'impact de diverses méthodes de distribution de questionnaires sur les taux de réponses et sur la qualité des réponses de l'enquête.*

*Cette article décrira ce projet ainsi que ses principales caractéristiques et présentera l'analyse de la qualité des réponses de l'EVI selon les différentes méthodes de distribution de questionnaires.*

Stéphane TREMBLAY, Statistique Canada Tunney's Pasture Ottawa, Ontario K1A 0T6  
tremste@statcan.ca

## Session 38

### Biostatistics : Topics in Environmental Statistics

#### *Biostatistique : Statistique de l'environnement*

#### [MS-127]

### The Accuracy of Ranked Set Sampling

#### *Sur la précision d'échantillonnage d'ensembles rangés*

Abdel H. EL-SHAARAWI\*, National Water Research Institute

Ranked set sampling (RSS) was introduced by McIntyre (1952) as an efficient method for estimating pasture yield. Recently several authors discussed its potential use in environmental studies. The RSS procedure consists of two logically distinct stages: a sampling selection stage and an estimation stage. Using linear estimators and sampling from a variety of distributions, it has been demonstrated analytically and numerically that RSS always produces more accurate estimators than those produced by other sampling schemes. This paper presents a general approach for assessing the accuracy of RSS, which is neither restricted to a specific distribution, a single parameter, nor to the class of linear estimators. Several examples will be used for illustration.

*L'échantillonnage des ensembles rangés (EER) a été introduit par McIntyre (1952) comme une méthode efficace pour estimer "le rendement- production" de pâturages. Récemment plusieurs auteurs discutent son utilisation potentielle dans les études environnementales. La procédure EER se compose de deux étapes logiques distinctes : une étape de sélection de l'échantillon et une étape d'estimation. En utilisant les estimateurs linéaires et en échantillonnant à partir de plusieurs lois, il a été prouvé analytiquement et numériquement que la EER produit toujours des estimateurs plus précis que ceux produits par d'autres schémas d'échantillonnage. Ce papier présente une approche générale pour évaluer la précision de la*

*EER, qui n'est pas liée à une loi particulière, à un seul paramètre, ni à une classe d'estimateurs linéaires. Plusieurs exemples seront donnés pour des fins d'illustration.*

Abdel H. EL-SHAARAWI, National Water Research Institute P.O. Box 5050 Burlington, Ontario L7R 4A6  
u101@csx.cciw.ca

[MS-128]

**Public Perception of Environmental Risks and Pre-hoc Statistical Analysis**

*La perception du public des risques environnementaux et les analyses statistiques 'Pre-hoc'*

Lara J. WOLFSON\*, University of Waterloo

A common problem in dealing with situations where environmental contamination has occurred is dealing with the near hysteria caused by public misconceptions of the risks involved – both of the contamination actually occurring, and the health risks associated with the contamination. In this talk, I will look at how eliciting beliefs and utilities to create the framework for a Bayesian decision analysis can illustrate the consequences of the mistaken perceptions of risk, and how eliciting these utilities and beliefs before sampling can guide the process of both sampling and analyzing data on the extent and nature of environmental contamination. The issue of public communication of environmental (and health) risks through the media will also be addressed.

*Un problème usuel lorsqu'une contamination environnementale s'est produite est de gérer la quasi-hystérie causée par la méconnaissance du public des risques associés, que ce soit au sujet de la contamination réelle produite, ou des risques sanitaires qui y sont associés. Dans cet exposé, j'étudierai comment incorporer dans un système d'analyse de décisions bayésien la perception subjective des risques. Ceci permettra d'une part d'illustrer l'impact de perceptions erronées et, d'autre part, de moduler la procédure d'échantillonnage et l'analyse des données pour évaluer l'étendue de la contamination à la lumière des risques perçus. Le problème de communication avec le public au sujet des risques environnementaux (et de santé) à travers les médias sera aussi traité.*

Lara J. WOLFSON, Dept. Statistics and Actuarial Sc. University of Waterloo Waterloo, Ontario N2L 3G1  
ljwolfso@setosa.uwaterloo.ca

[MS-129]

**Statistical Issues in the Development of Regional and National Scale Environmental Monitoring Programs**

*Les aspects statistiques dans le développement de programmes de contrôles environnementaux à l'échelle régionale et nationale*

Anthony R. OLSEN\*, US EPA National Health and Environmental Effects Research Laboratory

Environmental monitoring has a rich history of over 50 years within the United States. However, environmental sector information

from monitoring has not received equal status with federal statistics generated for the economic, social, and health sectors. The latter have long-standing national recognition. What statistical issues must be addressed to aid the transformation of environmental monitoring into a co-equal environmental sector? Answering this question requires an understanding of the historical development of monitoring, recognizing how monitoring currently is operationally institutionalized, and defining the information needs (objectives) for the environmental sector. The paper briefly summarizes current national inventory and monitoring programs in the United States, including suggestions for combining existing efforts. In doing this, a number of statistical issues are identified that must be addressed. Issues include development of models for the ecological universe being studied, extension of finite population sampling to continuous population sampling, combining information from multiple monitoring programs, development of small area estimation procedures, among others. Examples will be given illustrating these issues as well as others. Results from recent statistical research will be given where available.

*Le contrôle de l'environnement a une histoire de plus de 50 ans aux Etats-Unis. Par contre, l'information sur l'environnement liée à ce contrôle n'a pas obtenu, au sein des statistiques fédérales, la même reconnaissance que dans les secteurs économique, social, et de la santé. Quels sont les problèmes statistiques à traiter pour faciliter la transformation du contrôle environnemental en un secteur de l'environnement? Répondre à cette question exige une compréhension du développement historique du contrôle, qui tient compte de son opérationnalisation institutionnelle et de ses objectifs. Le papier résume succinctement les programmes d'inventaire et de contrôle aux Etats-Unis et les suggestions faites pour optimiser les efforts existants. En opérant ainsi, un certain nombre de problèmes statistiques à traiter sont identifiés. Mentionnons, entre autres, le développement de modèles pour l'univers écologique en cours d'étude, l'extension de l'échantillonnage de populations finies à un échantillonnage de populations continues et le développement de procédures d'estimation dans des petits domaines. Des exemples seront présentés pour illustrer ces problèmes. Des résultats de recherches récentes seront fournis lorsque disponibles.*

Anthony R. OLSEN, US EPA National Health and Environmental Effects Res. Labo. Western Ecology Division 200 SW 35th Street Corvallis, Oregon 97333, U.S.A.  
tolsen@mail.cor.epa.gov

Session 39

Mathematical Statistics

*Statistique mathématique*

[MS-130]

**Two Barrier Wiener Process Model for Length of Stay in the Hospital**

*Un processus de Wiener à deux barrières pour le temps de séjour des patients à l'hôpital*

Julie HORROCKS\*, University of Waterloo

Length of stay in hospital is a widely used outcome measure in Health Services research. Patient characteristics, especially severity of illness at admission to hospital, should be accounted for, as far as is possible. Thus a modelling approach is necessary.

An interesting feature of this type of data is multiple destinations, as a hospital stay can end in healthy discharge, transfer to another institution or death. We consider a model for duration times derived from a Wiener process with drift with two absorbing barriers, one representing healthy discharge and the other death in hospital. Individuals who are transferred to another institution are accommodated through censoring. Population heterogeneity is incorporated through regression modelling of the barrier levels and the drift coefficient.

*La durée d'hospitalisation est souvent utilisée comme une résultante par les services de recherche sur la santé. Les caractéristiques des patients, spécialement la sévérité de la maladie lors de l'admission à l'hôpital, devrait être prise en compte, autant que possible. Une approche faisant intervenir un modèle est nécessaire.*

*Une caractéristique intéressante de ce type de données est la multiplicité de la destination, puisque l'hospitalisation peut déboucher en une de trois issues : i) en bonne santé, ii) un transfert dans une autre institution hospitalière ou iii) le décès. Nous considérons un modèle pour les durées obtenues à partir d'un processus de Wiener avec "drift" à deux barrières absorbantes, une représentant la sortie en bonne santé et l'autre le décès à l'hôpital. Les patients transférés à une autre institution sont accommodés par une procédure de censure. L'hétérogénéité est incorporée dans la modélisation par une régression des niveaux des seuils et de la dérive.*

Julie HORROCKS, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1  
jhorrock@barrow.uwaterloo.ca

#### [MS-131]

### Reducing Delays in Preemptive Priority Queues

#### *Réduction du retard pour les files répétées de priorité de préemption*

Steve DREKIC\*, University of Western Ontario

David A. STANFORD, University of Western Ontario

The preemptive repeat priority queue is used to model situations where work in progress is lost when a customer is preempted. Upon returning to the preempted customer, the server starts over from the beginning. The versions of this queue that have been previously studied have only addressed the case where the same service distribution is used on successive attempts. In today's customer oriented environment, it seems logical to ask what the impact would be of improving the service effort on successive attempts. This talk addresses that question, and presents several possible alternatives for improving service progressively to preempted customers. Numerical examples will be provided.

*La file d'attente avec priorité répétée et préemption est utilisée pour modéliser les situations où le travail en cours est perdu quand un client est préempté. Lorsque le serveur retourne au client préempté, le serveur recommence du début. Les études antérieures de cette file d'attente ont uniquement résolu les cas où la distribution*

*du temps de service est la même pour tous les essais successifs. Dans l'environnement actuel, où l'emphase est mise sur le service au client, il paraît naturel d'améliorer le temps de service des essais successifs. Cet exposé résout cette question, et présente plusieurs alternatives possibles pour améliorer le service progressivement pour les clients préemptés. Des exemples numériques sont donnés.*

Steve DREKIC, University of Western Ontario Department of Statistical and Actuarial Sciences London, Ontario, Canada N6A 5B7  
dreki@fisher.stats.uwo.ca

#### [MS-132]

### Shrinking the Polychotomous Logistic Regression via the Lasso

#### *Le rétrécissement de la régression logistique polychotomique avec le 'Lasso'*

Carmen MAK\*, University of Toronto

We propose a *Lasso* style shrinkage method for parameter estimation in (PLR). The *Lasso* PLR estimate maximizes the likelihood of the data subject to an upper limit on the sum of the absolute values of all pairwise logit differences (log-odds). The effect of *Lasso* is to shrink some of the pairwise logit differences and to set some others to zero, thus achieving shrinkage and variable selection simultaneously. The method serves as a unifying approach in dealing with usual problems associated with maximum likelihood parameter estimation such as bias in small sample, non-existence in data separation, and multicollinearity of predictor variables. This work is an extension of *Lasso* introduced by Tibshirani in 1994 in Generalized Linear Models. Real and hypothetical data will be used to illustrate the properties of *Lasso* PLR estimators in various data configurations.

*Nous proposons une méthode de contraction de style Lasso pour une estimation paramétrique en RLP. Les estimés de la RLP par le Lasso maximisent la vraisemblance des données sous une contrainte sur la limite supérieure de la somme des valeurs absolues des toutes différence des "logits (log-odds)". L'effet du Lasso est de contracter quelques différences de logits et d'annuler les autres, réalisant ainsi une contraction et une sélection de variables simultanément. La méthode sert d'approche unificatrice pour traiter les problèmes associés à l'estimation des paramètres par la méthode du maximum de vraisemblance tels que le biais dans les échantillons de faible taille, la non-existence due à la séparation de données, et la multicollinéarité des variables prédictives. Ce travail est une extension du lasso introduit par Tibshirani (1994) pour les modèles linéaires généralisés. Des données réelles et hypothétiques seront utilisées pour illustrer les propriétés des estimateurs de la RLP par le Lasso dans plusieurs configurations de données.*

Carmen MAK, University of Toronto Toronto, Ontario, Canada M5S 1A1  
mak@utstat.utoronto.ca

#### [MS-133]

### Some Practical Designs for Biased Bivariate Regres-

**sion****Quelques schémas faisables pour la régression bivariée biaisée**

Giseon HEO\*, University of Alberta

Huber (1975) constructed designs for simple linear regression that are robust against unspecified contamination of the response function. Wiens (1990, 1992) established an extension of Huber's approach to multiple regression. We present tractable designs for approximately quadratic or approximately cubic polynomial in bivariate regression models.

*Huber (1975) a construit des plans d'expérience pour la régression linéaire simple qui sont robustes à la contamination non spécifiée de la fonction réponse. Wiens (1990, 1992) a établi une extension de l'approche de Huber à la régression multiple. Nous présentons des plans d'expérience, à notre portée, pour les modèles polynomiaux approximativement quadratiques ou approximativement cubiques de régression bivariée.*

Giseon HEO, University of Alberta 434 Central Academic Building, Edmonton, AB, Canada T6G 2G1  
schmu@stat.ualberta.ca

**[MS-134]****Robust Confidence Intervals and the Bootstrap**  
**Intervalles de confiance robustes et le bootstrap**

Matias SALIBIAN-BARRERA\*, University of British Columbia

Many proposals for robust estimation have been made both for location and regression models. Robust testing and confidence intervals, on the other hand, have received less attention in the literature. The existing proposals for robust inference assume either that the error scale is known or that the error distribution is symmetric. The main difficulty lies in that the asymptotic distributions cannot be easily obtained under the more realistic assumptions of unknown variance and asymmetric error distribution.

Since the calculation of robust estimates requires a considerable amount of computing time, classical bootstrap methods are not feasible beyond low-dimensional regression problems. Moreover, they do not perform well when the underlying error distribution has heavy-tails.

I will discuss the problem of robust confidence intervals and propose a way to construct them using a new bootstrap method which I call "Robust Bootstrap" (RB). I will show that RB is both computationally feasible and resistant to heavy-tails distributions.

*Plusieurs estimations robustes des paramètres de localisation et de modèles de régression ont été proposées, alors que les tests d'hypothèses robustes et les intervalles de confiance robustes ont reçu une attention moindre dans la littérature. Les propositions existantes (pour l'inférence robuste) supposent la connaissance de l'échelle des erreurs ou la symétrie de la loi des erreurs. La difficulté majeure réside dans la dérivation de la loi asymptotique sous des hypothèses, plus réalistes, de variance inconnue et de loi asymétrique des erreurs.*

*Le calcul des estimés robustes nécessite un temps de calcul assez long, alors les méthodes du bootstrap classique n'est pas faisable*

*pour les problèmes de régression au delà de la dimension deux. En plus, leur performance est mauvaise quand la loi des erreurs présente des "ailes lourdes".*

*Je discuterai le problème des intervalles de confiance robustes et proposerai une procédure pour les construire utilisant une nouvelle méthode de bootstrap que j'appellerai "nouveau bootstrap"(NB). Je montrerai que le NB est faisable et est résistant aux distributions avec des "ailes lourdes".*

Matias SALIBIAN-BARRERA, Department of Statistics University of British Columbia 2021 West Mall Vancouver, BC Canada V5T 1Z2  
matias@stat.ubc.ca

**[MS-135]****Comparative Study of Somes Deconvolution Methods for Densities****Etude comparative de méthodes de déconvolution de densité**

Stéphanne GERMAIN\*, Université Laval

In many fields (imagery, teledetections, etc.) the available measurements are contaminated with noise. The noise generally come from the censor or some natural phenomenon. The problem is to extract the information from the perturbed data.

Formally, let  $X$  be the random variable of interest,  $\epsilon$  the noise, and  $Y$  the observed variable. These variables are linked by the relation  $Y = X + \epsilon$ . The aim is to estimate de probability density function of  $X$  from a random sample  $Y_1, \dots, Y_n$ . This problem is commonly call "deconvolution".

There exist many approaches to solve this problem. We study the kernel method, the regularisation method and the wavelets decomposition method. We treat the asymptotic properties and the practical implementations of these estimators. We also study the case where the error density is unknow and must be estimate from a random sample of noise  $\epsilon_1, \dots, \epsilon_N$ .

We applied these methods to the classification of galactics clouds.

*Dans plusieurs domaines (imagerie, télédétections, etc.) les mesures recueillies sont entachées d'erreurs. Ces erreurs sont généralement dues à l'appareillage utilisé ou à d'autres phénomènes naturels. Le problème consiste donc à extraire l'information pertinente à partir de données perturbées.*

*Formellement, désignons par  $X$  la variable aléatoire d'intérêt, par  $\epsilon$  le bruit et par  $Y$  la variable réelement observée. Ces valeurs sont reliées par la relation  $Y = X + \epsilon$ . Le but est d'estimer la densité de probabilité de  $X$  à partir d'une série d'observations  $Y_1, \dots, Y_n$ . L'estimation d'une telle densité est communément appelée "déconvolution".*

*Il existe de nombreuses techniques pour résoudre ce problème. Nous étudions, plus spécifiquement, les méthodes du noyau, de régularisation et de décomposition en ondelettes. Nous traitons les propriétés asymptotiques et la mise en oeuvre pratiques des estimateurs proposés. Nous distinguons également deux cas : le cas où la loi des erreurs ( $\epsilon$ ) est connue et celui où cette loi est inconnue mais en plus des observations  $Y_1, \dots, Y_n$  nous disposons d'un échantillon des erreurs  $\epsilon_1, \dots, \epsilon_N$ .*

*Comme application concrète de ces techniques nous utilisons la déconvolution pour classifier des nuages galactiques.*

Stéphane GERMAIN, Dép. de mathématiques et de statistique Université Laval Sainte-Foy QC G1K 7P4  
germain@math.ulaval.ca

## Session 40

### Survey Methods : Outliers in Survey Data

#### *Méthodologie d'enquête : Traitement des valeurs aberrantes*

[MS-136]

#### **Influential Observations : Identification and Treatment by M-estimators**

#### *Observations influentes : identification et traitement par les M-estimateurs*

Jean-Philippe GWET\*, Westat

Sample units with unduly large values or large weights or a combination of both often occur in sample surveys especially in business surveys. A wide variety of methods have been proposed in the literature for dealing with this problem. Each of the approaches known to date was developed for a specific design which limits its applicability. We will discuss a general approach for dealing with the robust estimation problem in survey sampling under a general sampling design. Some simple sampling designs such as simple random sampling and stratified sampling will be used for illustration purposes. A method of estimating the variance of the robust estimators using the jackknife methodology will be discussed as well.

An important area where this approach can prove useful is that of misclassified or mis-stratified units. In stratified sampling, the population is usually stratified on the basis of some auxiliary variables which are supposedly related to the variables of interest. Because of data entry errors, missing data or other technical difficulties some sample units will be mis-classified in wrong strata. If nothing is done the resulting estimates will have a high variance. It will be shown how the proposed techniques can handle this problem.

The potential of this general technique within the framework of model-based small area estimation will also be presented along with some simulation results.

*Les unités échantillonales présentant des valeurs excessivement grandes ou de larges pondérations ou une combinaison des deux apparaissent dans les sondages, spécialement dans les sondages sur les affaires. Un panorama de méthodes a été proposé dans la littérature pour traiter de ce problème. Chacune de ces approches connues jusqu'à date a été développée pour un plan d'échantillonnage spécifique, ce qui limite son applicabilité. Nous discuterons une approche générale pour traiter du problème d'estimation robuste dans la théorie du sondage sous un plan d'échantillonnage général. Quelques plans d'échantillonnage simples tels que l'échantillonnage aléatoire simple et l'échantillonnage stratifié seront utilisés pour des fins d'illustrations. Une méthode d'estimation de la variance des estimateurs robustes utilisant la méthodologie du Jackknife sera aussi discutée.*

*Cette approche devrait s'avérer utile pour des unités mal classées ou mal stratifiées. En échantillonnage stratifié, la population est souvent stratifiée sur la base de variables auxiliaires supposées être fortement reliées à la variable d'intérêt. A cause des erreurs lors de*

*la saisie des données, de données manquantes ou d'autres difficultés techniques, quelques unités de l'échantillon seront "mal" classées, dans de mauvaises strates. Si rien n'est fait, il en résulte une estimation avec une variance assez large. Nous montrerons comment les techniques proposées traitent de ce problème.*

*L'application de cette technique générale à l'estimation dans des petits domaines, à partir de modèles, sera aussi présentée avec quelques résultats de simulation.*

Jean-Philippe GWET, Westat Inc. 1650 Research Boulevard Rockville, MD 20850 U.S.A.

GwetJ1@Westat.com

[MS-137]

#### **On Winsorization in Business Surveys**

#### *Sur la winsorisation dans les enquêtes entreprises*

Phil KOKIC\*, University of Southampton

Ray CHAMBERS, University of Southampton

Paul SMITH, Office for National Statistics

Ratio and expansion estimation are common methods of estimating total in a stratified finite population. However, if by chance several unusually large (or small) observations fall in the sample, then either of these estimators may poorly estimate the true population total. One technique to deal with this problem is to reduce sampled observations greater than a cut-off to a value closer to that cutoff, and then re-estimate total using the adjusted values. The resulting estimator is called the one-sided Winsorized estimator of total. Its properties have been explored in a number of recent research articles including Kokic and Bell (1994), and Smith and Kokic (1997). In this paper we present theory for optimal two-sided cutoffs for Winsorizations in a stratified finite population when the residual error distributions are unimodal. In a repeated survey historical information may be used to estimate these cutoffs. Some of the practical advantages of two-sided compared to one-sided Winsorization are discussed. In particular, it has considerably smaller bias than its one-sided counterpart, and so it is far simpler to implement when derived survey variables are required. Its effectiveness in reducing mean squared error will be illustrated in a simulation analysis based on data collected in the UK monthly Retail Sales Inquiry.

*L'estimation par le quotient ou par expansion sont des méthodes souvent utilisées pour estimer le total dans une population finie et stratifiée. Par ailleurs, si par hasard plusieurs observations larges (ou faibles) apparaissent dans l'échantillon, alors ces deux estimateurs peuvent mal estimer le vrai total de la population. Une technique pour traiter de ce problème est de réduire les observations qui sont plus grandes qu'un seuil, fixé au préalable, à une valeur proche de ce seuil, et ensuite ré-estimer le total en utilisant ces valeurs ajustées. L'estimateur qui en résulte est appelé estimateur du total winsorisé à un seul côté. Ses propriétés ont été explorées dans des articles scientifiques récents de Kokic et Bell (1994), Smith and Kokic (1997) parmi d'autres. Dans ce papier nous présentons la théorie pour les seuils optimaux des deux côtés pour la winsorisation d'une population finie et stratifiée quand les distributions des erreurs résiduelles sont unimodales. Dans un sondage répété, l'information historique peut être*



utilisée pour estimer ces seuils. Quelques avantages pratiques de la winsorisation à deux cotés en comparaison avec celle à un seul coté sont discutés. En particulier, elle fournit un biais assez faible comparativement à son analogue à un seul coté, et elle est de loin facile à réaliser quand des variables dérivées de sondage sont nécessaires. Son efficacité à réduire l'erreur quadratique moyenne sera illustrée dans une analyse par simulation basée sur des données recueillies dans le "UK Monthly Retail Sales Inquiry".

Phil KOKIC, Department of Social Statistics University of Southampton Highfield, Southampton S017 1BJ UNITED KINGDOM  
pk@alcd.soton.ac.uk

[MS-138]

### Outlier Robust Generalized Regression Estimator

*Estimateur généralisé par la régression robuste à la présence de valeurs extrêmes*

Hyunshik LEE\*, Westat

Zdenek PATAK, Statistics Canada

Sarah FRANKLIN, Statistics Canada

The problem of outliers is common in sample surveys, particularly for business surveys. The Generalized Regression Estimator (GREG) is now well accepted and widely used in survey sampling when auxiliary data is available. However, the GREG estimator is vulnerable in the presence of outliers. In this paper we propose an outlier robust GREG estimator that is designed consistent. Some simulation results are also presented.

*Le problème des observations aberrantes est fréquent dans les sondages, et particulièrement dans les sondages sur les affaires. L'estimateur par la régression généralisée (ERG) est actuellement accepté et largement utilisé dans la théorie des sondages quand des données auxiliaires sont disponibles. Par ailleurs, le ERG est vulnérable aux valeurs aberrantes. Dans ce papier nous proposons un ERG robuste aux valeurs aberrantes qui est convergent par rapport au plan d'échantillonnage. Quelques résultats de simulation sont aussi présentés.*

Hyunshik LEE, Westat Inc. 1650 Research Boulevard Rockville, MD 20850 U.S.A.  
LEEHI@Westat.com

## Session 41

### Committee on Women in Statistics & The Caucus for Women in Statistics: Classification

*Comité des femmes en statistique & Caucus pour les femmes en statistique : Classification*

[MS-139]

### Clustering and characterization of HI Astronomical

### Objects Using Density Estimation

*Une approche basée sur l'estimation de la densité pour la caractérisation et la classification d'objets astronomiques*

Nadia GHAZZALI\*, Université Laval

Our study aims at the characterization of HI gas cloud images for classification purposes. We wish to identify the existing relations between the different clouds as a first step towards understanding the evolutionary and physical processes that form them. The HI images are characterized by their brightness temperature distribution. Our study starts with an exploratory analysis based on Principal Components Analysis (PCA) and on Cluster Analysis. As the inherent signal in each image is noise, the characterization is then studied by using a method of density estimation by deconvolution. The clustering of the images is done by considering single linkage and complete linkage criteria of cluster aggregation.

*Notre étude porte sur la caractérisation d'objets astronomiques extraits de nuages de gaz HI à des fins de classification. Le but est d'exhiber les relations qui existent entre ces différents nuages pour comprendre l'évolution et les processus physiques qui les forment. Chaque objet extrait de ses nuages est caractérisé par la distribution de ses intensités lumineuses. L'étude de tels objets est d'abord faite par des analyses exploratoires basées sur l'analyse en composantes principales (ACP) et sur la classification. Comme le signal inhérent à chaque objet est bruité, cette caractérisation est ensuite étudiée en s'appuyant sur une méthode d'estimation de densité par déconvolution. La classification des objets ainsi caractérisés se fait en considérant les critères classiques du saut minimal et du saut maximal.*

Nadia GHAZZALI, Dép. mathématiques et statistique Pav. Vachon, bureau 1056 Université Laval Sainte-Foy QC G1K 7P4  
ghazzali@mat.ulaval.ca.

[MS-140]

### Logistic Regression Trees, Reduced Logistic Regression and Generalized Individualized Regression

*Arbre de régression logistique, modèle logistique réduit et régression individuelle généralisée*

Pascale ROUSSEAU\*, Université de Québec à Montréal

Logistic regression models implying subsets of the categories of a polychotomous response variable have been studied by different authors. Begg and Collins (1984) proposed individualized regression models which used series of dichotomous models to estimate logistic models. Bull, Greenwood and Donner (1994) compared estimates of reduced logistic models and individualized-reduced estimates with full logistic model estimates. Rousseau and Sankoff (1989) proposed regression logistic trees which used series of polychotomous models organized hierarchically. The model that best fitted the data was then chosen as the one which corresponded to the tree with the maximum likelihood. Monte-Carlo simulation studies showed that these models or trees seemed to be mostly appropriate in cases of reduced models

. Comparing results obtained by these various authors, the selected regression logistic tree model seemed to find the subsets of categories that have the most efficient estimates. The tree selected with the maximum likelihood procedure represented the full model whenever the efficiency of the reduced estimates and of the individualized reduced estimates were the lowest. On the other hand, the selected tree that included subsets of the categories, had highest likelihood when the reduced model or individualized model was much more efficient than the full model. Another application of the regression logistic tree is a generalization of the individualized regression model to subsets of more than two categories of the response variable.

*Des modèles de régression logistique sur des sous-ensembles des  $m$  catégories d'une variable réponse polychotomique ont été étudiés par plusieurs auteurs. Begg et Collins (1984) ont proposé des modèles de régression logistique individuelle qui utilisent des séries de modèles dichotomiques pour estimer le modèle polychotomique avec  $m$  catégories. Bull, Greenwood et Donner (1994) ont comparé les estimateurs des modèles réduits et des modèles individuels avec les estimateurs du modèle polychotomique impliquant toutes les variables explicatives (modèle régulier). Rousseau et Sankoff (1989) ont proposé des modèles d'arbres de régression logistique qui utilisent des séries de modèles polychotomiques de  $k$  ( $i$ ) catégories,  $k$  ( $i$ ) <  $m$ , ordonnés selon une structure d'arbre. Le modèle d'arbre qui ajuste le mieux les données est estimé avec la méthode de vraisemblance maximale. Des études basées sur des simulations Monte-Carlo semblent indiquer que ces arbres de régression logistique sont davantage appropriés dans les cas de modèle réduit. Le modèle d'arbre sélectionné dans les différentes simulations est celui qui contient dans des sous-ensembles des catégories impliquant les modèles réduits ou individuels pour lesquelles les estimateurs sont les plus efficaces. En effet, l'arbre sélectionné représente le modèle régulier lorsque l'efficacité des autres modèles est la moins élevée. L'arbre incluant des sous-ensembles de catégories, a la plus grande vraisemblance lorsque les estimateurs des modèles réduits et individuels sont plus efficaces que ceux du modèle régulier. Les arbres de régression logistique permettent aussi la généralisation des modèles individuels dichotomiques à des modèles individuels polychotomiques avec  $2 < k < m$  catégories.*

Pascale ROUSSEAU, Département de mathématiques Université du Québec à Montréal Case postale 8888 Montréal QC H3C 3P8  
rousseau.pascale@uqam.ca

The translation team did its best. Please forgive us for our misinterpretations, we are neither professional translators nor specialists in all areas of statistics.

*L'équipe de traduction a fait de son mieux. Excusez-nous pour les erreurs de traduction, nous ne sommes ni des traducteurs professionnels ni des spécialistes dans tous les domaines de la statistique.*

## Author Index • Index des auteurs

- BERNIER, Julie ..... Session(s) 15
- AALLEN, Odd. A. .... Session(s) 2  
 ALLARD, Benoit ..... Session(s) 29  
 ALLARD, Pierre ..... Session(s) 22  
 ALLOUBA, Hassan ..... Session(s) 4  
 ALVO, Mayer ..... Session(s) 6  
 ARSENAULT, Sophie ..... Session(s) 37  
 AWAD, Lucile ..... Session(s) 27  
 AYERS, Dieter ..... Session(s) 34
- BALASOORIYA, Uditha ..... Session(s) 36  
 BARTLETT, S. .... Session(s) 8  
 BAUMGARTEN, Cynthia ..... Session(s) 26  
 BEAUCAGE, Yanick ..... Session(s) 15  
 BEAUMONT, J.-F. .... Session(s) 10  
 BELLAVANCE, François ..... Session(s) 19, 27  
 BELMONTE, Ève ..... Session(s) 29  
 BERNIER, J. .... Session(s) 25  
 BIERNACKA, Joanna ..... Session(s) 34  
 BILODEAU, Martin ..... Session(s) 13  
 BISSONNETTE, Joël ..... Session(s) 29  
 BORDAGE, Daniel ..... Session(s) 33  
 BOURGET, E. .... Session(s) 33  
 BOURGET, André ..... Session(s) 33  
 BOUTROS, Janine ..... Session(s) 5  
 BRISEBOIS, François ..... Session(s) 37  
 BROWN, K. Stephen ..... Session(s) 26  
 BRUNDAGE, Michael ..... Session(s) 8  
 BURNETT, R. .... Session(s) 8  
 BÉLAND, Yves ..... Session(s) 5  
 BÉLISLE, Claude ..... Session(s) 36
- CABILIO, Paul ..... Session(s) 6  
 CAKMAK, Sabit ..... Session(s) 8  
 CARON, Nancy ..... Session(s) 27  
 CASTONGUAY, Élaine ..... Session(s) 15  
 CHAKAK, Abderrahmane ..... Session(s) 36  
 CHAMBERS, Ray ..... Session(s) 40  
 CHAUVIN, Brigitte ..... Session(s) 4  
 CHEN, Ding G. .... Session(s) 8  
 CHEN, Gemai ..... Session(s) 6  
 CHEN, Zhao-Guo ..... Session(s) 25  
 CHENG, Smiley W. .... Session(s) 6  
 CHOLETTE, Pierre ..... Session(s) 25  
 CIAMPI, A. .... Session(s) 22  
 CIAMPI, Antonio ..... Session(s) 22  
 COCKELL, K. .... Session(s) 8
- COLE, Martin ..... Session(s) 19  
 CORREA, José Andrés ..... Session(s) 19  
 COURTEAU, J. .... Session(s) 21, 22  
 CRÉPEAU, Héléne ..... Session(s) 28
- DABROWSKY, André ..... Session(s) 9  
 DAIGLE, Gaétan ..... Session(s) 33  
 DEAN, C. .... Session(s) 19  
 DEMNATI, Abdellatif ..... Session(s) 10  
 DESAI, Kamal ..... Session(s) 22  
 DEVILLE, Jean-Claude ..... Session(s) 20  
 DREKIC, Steve ..... Session(s) 39  
 DROUIN, Jean ..... Session(s) 15  
 DRYDEN, Ian ..... Session(s) 23  
 DUBOIS, S. .... Session(s) 8  
 DUBREUIL, Guylaine ..... Session(s) 15  
 DUCHESNE, Marc ..... Session(s) 24  
 DUCHESNE, Pierre ..... Session(s) 29  
 DUCHESNE, T. .... Session(s) 27  
 DUFOUR, Johane ..... Session(s) 29  
 DUPUIS, D.J. .... Session(s) 11  
 DUTILLEUL, Pierre ..... Session(s) 35  
 DUVAL, Marie-Claude ..... Session(s) 15
- EISENHAUER, E. .... Session(s) 17  
 EL-SHAARAWI, Abdel H. .... Session(s) 38  
 EZZERG, M. .... Session(s) 36
- FARRELL, Patrick J. .... Session(s) 26  
 FELLEGI, Ivan P. .... Session(s) 1  
 FERLAND, René ..... Session(s) 4  
 FICK, Gordon ..... Session(s) 14  
 FICK, Gordon H. .... Session(s) 26  
 FLUET, Michel ..... Session(s) 28  
 FRANKLIN, Sarah ..... Session(s) 40
- GAYLOR, David ..... Session(s) 12  
 GERMAIN, Stéphanne ..... Session(s) 39  
 GHAHRAMANI, M. .... Session(s) 19  
 GHAZZALI, Nadia ..... Session(s) 41  
 GIROUX, Gaston ..... Session(s) 4  
 GODAMBE, V. P. .... Session(s) 30  
 GOUPH, Hew ..... Session(s) 14  
 GRANT, A. .... Session(s) 21  
 GREENWOOD, Celia ..... Session(s) 7  
 GWET, Jean-Philippe ..... Session(s) 40
- HAQ, M. Safiul ..... Session(s) 13  
 HAREZLAK, Jarek ..... Session(s) 6

- HARRINGTON, David ..... Session(s) 2  
 HAYWARD, S. .... Session(s) 26  
 HEO, Giseon ..... Session(s) 39  
 HIDIROGLOU, Mike ..... Session(s) 10  
 HORROCKS, Julie ..... Session(s) 39  
 HURTUBISE, Daniel ..... Session(s) 29
- IMLAHI, L. .... Session(s) 36  
 IVANOFF, Gail ..... Session(s) 9
- JACQUES, Julie ..... Session(s) 11  
 JAMES, Keith ..... Session(s) 8  
 JOHNSTON, Lynn ..... Session(s) 8  
 JONES, Cheryl ..... Session(s) 2  
 JUNKINS, B. .... Session(s) 8
- KIBRIA, B. M. Golam ..... Session(s) 13  
 textscKnopp-Schneider, Annette ..... Session(s) 12  
 KOKIC, Phil ..... Session(s) 40  
 KOKIOPOULOS, Bill ..... Session(s) 34  
 KONGA, Remy Konga ..... Session(s) 26  
 KOTT, Phillip S. .... Session(s) 25  
 KREWSKI, D. .... Session(s) 8  
 KUMAR, Pranesh ..... Session(s) 25  
 KUSTRA, Rafal ..... Session(s) 3
- L'ABBE, M. .... Session(s) 8  
 LABERGE, Claude ..... Session(s) 33  
 LABILLOIS, Tony ..... Session(s) 37  
 LACHENBRUCH, Peter A. .... Session(s) 17  
 LACROIX, Eric ..... Session(s) 37  
 LAFLAMME, François ..... Session(s) 37  
 LAM, Ryan ..... Session(s) 11  
 LAMBERT, Marie Claude ..... Session(s) 27  
 LANGLET, Éric ..... Session(s) 29  
 LANIEL, Normand ..... Session(s) 15  
 LAPIERRE, B. .... Session(s) 37  
 LAPLANTE, P. .... Session(s) 21  
 LAPOINTE, François-Joseph ..... Session(s) 35  
 LAROCQUE, Denis ..... Session(s) 13  
 LATOUCHE, Michel ..... Session(s) 10  
 LAVALLÉE, Pierre ..... Session(s) 37  
 LAWLESS, Jerry ..... Session(s) 30  
 LE, Huiling ..... Session(s) 23  
 LEE, Hyunshik ..... Session(s) 40  
 LEE, J. .... Session(s) 3  
 LEROUX, Diane ..... Session(s) 28  
 LEWIGER, Juan Pablo ..... Session(s) 3  
 LI, Tong ..... Session(s) 8  
 LITTLE, Bruce ..... Session(s) 16  
 LIU, C. .... Session(s) 3  
 LIU, Tzen-Ping ..... Session(s) 21  
 LU, Xuewen ..... Session(s) 26  
 LUNETTA, Kathy ..... Session(s) 7  
 LUSSIER-CACAN, S. .... Session(s) 22  
 LÉVESQUE, Isabelle ..... Session(s) 29
- MACDONALD, Diane ..... Session(s) 8
- MACDONALD, Peter ..... Session(s) 34  
 MACLEOD, D. .... Session(s) 8  
 MAK, Carmen ..... Session(s) 3, 39  
 MANSOUR, Asmaâ ..... Session(s) 19  
 MARCHAND, Eric ..... Session(s) 13  
 MARCHAND, Isabelle ..... Session(s) 29  
 MATTHEWS, David E. .... Session(s) 21  
 MCLEOD, A. Ian ..... Session(s) 3  
 MCLEOD, A.I. .... Session(s) 11  
 MESBAH, Mounir ..... Session(s) 19, 27  
 MILLER, D. .... Session(s) 8  
 MILLS, Shirley ..... Session(s) 14  
 MIZRAHI, Arié ..... Session(s) 5  
 MOREAULT, Daniel ..... Session(s) 11  
 MOREL, J. .... Session(s) 37  
 MORI, B. .... Session(s) 8  
 MORRY, Marietta ..... Session(s) 25  
 MUKHERJEE, Jayanti ..... Session(s) 17  
 MÂSSE, Benoît ..... Session(s) 22
- NADEAU, Claude ..... Session(s) 2  
 NIYONSENGA, T. .... Session(s) 21  
 NIYONSENGA, Théophile ..... Session(s) 22  
 NOBREGA, K. .... Session(s) 25
- OLSEN, Anthony R. .... Session(s) 38  
 ORMSBY, E. .... Session(s) 8
- PAGEAU, François ..... Session(s) 24  
 PAPATHANASOPOULOU, L. .... Session(s) 8  
 PARADIS, Sylvain ..... Session(s) 33  
 PATAK, Z. .... Session(s) 37  
 PATAK, Zdenek ..... Session(s) 40  
 PATER, Joe ..... Session(s) 27  
 PAUL, S.R. .... Session(s) 36  
 PELLETIER, Éric ..... Session(s) 29  
 PERRON, Sylvain ..... Session(s) 15  
 PLANTE, Nathalie ..... Session(s) 33  
 PLATT, Robert ..... Session(s) 19  
 POIRIER, Sylvain ..... Session(s) 15  
 PROTEAU, Daniel ..... Session(s) 24
- QUENNEVILLE, Benoît ..... Session(s) 25
- RAHMAN, M. .... Session(s) 8  
 RAJES, Jeevanantham ..... Session(s) 34  
 RAMSAY, J. O. .... Session(s) 6  
 RANCOURT, Éric ..... Session(s) 10, 29  
 RATHI, Keshavlal ..... Session(s) 34  
 RENAUD, Martin ..... Session(s) 10  
 RIVEST, L.-P. .... Session(s) 33  
 RODRIGUE, Natalie ..... Session(s) 27  
 ROSLIN, Nicky ..... Session(s) 34  
 ROSS, W.H. .... Session(s) 8, 26  
 ROUSSEAU, Pascale ..... Session(s) 41  
 ROY, M. .... Session(s) 22  
 ROYCE, D. .... Session(s) 15  
 RÉMILLARD, Bruno ..... Session(s) 4

SALIBIAN-BARRERA, Matias	Session(s) 39
SANKOV, David	Session(s) 35
SASHEGYI, Andreas I.	Session(s) 26
SAW, Sutaip S.	Session(s) 36
SCHEUREN, Fritz	Session(s) 30
SCOTT, Andrea	Session(s) 8
SHIN, Hwa Shin H.	Session(s) 6
SIMARD, Michelle	Session(s) 15
SIMARD, Claude	Session(s) 26
SINGH, R.S.	Session(s) 26
SIROIS, Mike	Session(s) 15
SITTER, Randy R.	Session(s) 18
SLONOWSKY, Dean	Session(s) 9
SMITH, Paul	Session(s) 40
SPIJAVCA, Ruxandra	Session(s) 3
SPINELLI, J.	Session(s) 19
SRINIVASAN, Tara	Session(s) 34
ST-PIERRE, Martin	Session(s) 29
STANFORD, David A.	Session(s) 39
STYAN, George P. H.	Session(s) 13
STRUIJS, Peter	Session(s) 37
SUISSA, Sammy	Session(s) 19
SUN, Jianguo	Session(s) 21
SWEET, Lamont	Session(s) 8
TARASSOFF, Peter	Session(s) 15
TREMBLAY, Johanne	Session(s) 37
TREMBLAY, Stéphane	Session(s) 37
TRICOT, Jean-Marie	Session(s) 19
TROUPÉ, Marylène	Session(s) 13
TRÉPANIÉ, Julie	Session(s) 29
TU, Dongsheng	Session(s) 27
VANASSE, N.	Session(s) 21
VINING, G. Geoffrey	Session(s) 18
VIVEROS, Roman	Session(s) 34
WADHWA, Pavan	Session(s) 6
WALKER, C.	Session(s) 37
WANG, Steven	Session(s) 34
WEERASINGHE, Swarna	Session(s) 8
WEIJER, Charles	Session(s) 7
WOLFSON, Lara J.	Session(s) 38
WOODALL, William H.	Session(s) 18
WORSLEY, Keith	Session(s) 23
XHIGNESSE, M.	Session(s) 21, 22
XIE, Hansheng	Session(s) 6
YANG, Hong	Session(s) 19
YATMAN, Jerrold	Session(s) 8
YOSHIO, Kajitani	Session(s) 3
ZAREPOUR, Mahmoud	Session(s) 36
ZEE, Benny	Session(s) 8, 17
ZHAN, Min	Session(s) 27
ZHAO, Feng	Session(s) 27
ZHOU, J.	Session(s) 3
ZITIKIS, Ričardas	Session(s) 36