FIRST LATIN AMERICAN ISI SATELLITE MEETING ON SMALL AREA ESTIMATION

(REVIEWED SEPT. 2015)

AUGUST 3-5, 2015, SANTIAGO, CHILE
INTERNATIONAL STATISTICAL INSTITUTE (ISI) SATELLITE MEETING AT PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
WWW.ENCUESTAS.UC.CL/SAE2015
First Latin American ISI Satellite Meeting on Small Area Estimation
3-5 August, Santiago, Chile

Book of Abstracts
# Table of Contents

- 1. Aims and Scope 5
- 2. The Advisory Board 6
- 3. The Program Committee 6
- 4. The Local Organizing Committee 6
- 5. The Program at a Glance 7
- 6. The Program in Detail 8
- 7. The Abstracts 13
- 8. The Sponsors & Organizers 39
1. AIMS AND SCOPE

Welcome to the First Latin American ISI Satellite Meeting on Small Area Estimation (SAE 2015), held at the Pontificia Universidad Católica de Chile on August 3-5 of 2015.

The main purpose of organizing a conference on small area estimation in Chile is to assess the current state of development and usage of small area methodology. We hope that this meeting serves as a bridge between statisticians and practitioners working on small area estimation in academia, private and government agencies in Chile. Holding the meeting in Santiago will give researchers in Latin American countries an opportunity to learn about state-of-the-art small area estimation techniques from the experts in the field.

The SAE 2015 conference is been co-sponsored by the International Statistical Institute (ISI), the International Association of Survey Statisticians (IASS), the Sociedad Chilena de Estadística (SOCHE), the Instituto Nacional de Estadísticas (INE), the Ministerio de Desarrollo Social (MDS), the Comisión Nacional de Investigación Científica y Tecnológica (CONICYT) and the Universidad Católica de Chile (Instituto de Sociología, Departamento de Estadística, Departamento de Salud Pública, Centro de Encuestas y Estudios Longitudinales and Vicerrectoría de Investigación).

We are excited to have you here in Santiago!

Dr. Carolina Casas-Cordero, Universidad Católica de Chile, Chile
Chair, Local Organizing Committee

Dr. Partha Lahiri, University of Maryland, College Park, USA.
Chair, Advisory Board
2. THE ADVISORY BOARD

Partha Lahiri, JPSM, University of Maryland, U.S.A. - Chairperson
Raymond Chambers, University of Wollongong, Australia
Manuel Galea, Departamento de Estadística, Univ. Católica de Chile, Chile
Malay Ghosh, University of Florida, U.S.A.
Gloria Icaza, Sociedad Chilena de Estadística (SOCHE), Chile
Graham Kalton, Westat, U.S.A.
Peter Lanjouw, The World Bank, USA.
Osvaldo Larrañaga, United Nations Development Program (UNDP), Chile.
Danny Pfeffermann, Chief Government Statistician of Israel, Israel
J.N.K. Rao, Carleton University, Canada

3. THE PROGRAM COMMITTEE

Raymond Chambers, University of Wollongong, Australia
Gloria Icaza, Sociedad Chilena de Estadística (SOCHE), Chile
J.N.K. Rao, Carleton University, Canada

4. THE LOCAL ORGANIZING COMMITTEE

Carolina Casas-Cordero, Instituto de Sociología, Univ. Católica de Chile - Chairperson
David Bravo, Centro de Encuestas y Estudios Longitudinales, Univ. Católica de Chile
Jaime Espina, Instituto Nacional de Estadísticas (INE)
Manuel Galea, Departamento de Estadística, Univ. Católica de Chile
Gloria Icaza, Sociedad Chilena de Estadística (SOCHE)
Patricia Medrano, Centro de Encuestas y Estudios Longitudinales, Univ. Católica de Chile
Isabel Millán, Ministerio de Desarrollo Social (MDS)
David Niculcar, Instituto Nacional de Estadísticas (INE)
Luis Villarroel, Departamento de Salud Pública, Univ. Católica de Chile
## 5. The Program at a Glance

<table>
<thead>
<tr>
<th>DAY</th>
<th>TIME</th>
<th>ACTIVITY</th>
<th>SPEAKER(S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday, August 3rd</td>
<td>08:00 - 09:00</td>
<td>Registration</td>
<td></td>
</tr>
<tr>
<td></td>
<td>09:00 - 10:00</td>
<td>Opening Ceremony</td>
<td>Local Authorities; G. Kalton</td>
</tr>
<tr>
<td></td>
<td>10:00 - 10:15</td>
<td>Coffee Break</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10:15 - 11:45</td>
<td>Invited Talk 1</td>
<td>A. Luna-Hernandez</td>
</tr>
<tr>
<td></td>
<td>11:45 - 12:00</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12:00 - 13:00</td>
<td>Invited Talk 2</td>
<td>R. Tiller + D. Pfeffermann; W. Bell (Discussant)</td>
</tr>
<tr>
<td></td>
<td>13:00 - 14:15</td>
<td>Lunch Break</td>
<td>Gatopardo Restaurant (1st floor)</td>
</tr>
<tr>
<td></td>
<td>14:15 - 16:15</td>
<td>Invited Talk 3</td>
<td>Panel – Recent Interesting App. of SAE Methods</td>
</tr>
<tr>
<td></td>
<td>16:15 - 16:30</td>
<td>Coffee Break</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16:30 - 18:30</td>
<td>Special Topic Session 1</td>
<td>New Perspectives on SAE</td>
</tr>
<tr>
<td></td>
<td>18:30 - 19:30</td>
<td>Reception</td>
<td>Poster Session (Hallway, 2nd floor)</td>
</tr>
<tr>
<td>Tuesday, August 4th</td>
<td>08:00 - 09:00</td>
<td>Registration</td>
<td></td>
</tr>
<tr>
<td></td>
<td>09:00 - 10:30</td>
<td>Special Topic Session 2</td>
<td>Small Area Applications</td>
</tr>
<tr>
<td></td>
<td>10:30 - 10:45</td>
<td>Coffee Break</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10:45 - 11:45</td>
<td>Invited Talk 4</td>
<td>M. Ghosh</td>
</tr>
<tr>
<td></td>
<td>11:45 - 12:00</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12:00 - 13:00</td>
<td>Invited Talk 5</td>
<td>D. Rubin</td>
</tr>
<tr>
<td></td>
<td>13:00 - 14:15</td>
<td>Lunch Break</td>
<td>Gatopardo Restaurant (1st floor)</td>
</tr>
<tr>
<td></td>
<td>14:15 - 15:45</td>
<td>Special Topic Session 3</td>
<td>SAE Foundation Issues</td>
</tr>
<tr>
<td></td>
<td>15:45 - 16:00</td>
<td>Coffee Break</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16:00 - 17:00</td>
<td>Invited Talk 6</td>
<td>P. Lahiri</td>
</tr>
<tr>
<td></td>
<td>17:00 - 17:15</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td></td>
<td>17:15 - 18:30</td>
<td>Contributed Session 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>18:30 - 19:45</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td></td>
<td>19:45 - 22:00</td>
<td>Banquet</td>
<td>Los Adobes de Argomedo Restaurant</td>
</tr>
<tr>
<td>Wednesday, August 5th</td>
<td>08:00 - 09:00</td>
<td>Registration</td>
<td></td>
</tr>
<tr>
<td></td>
<td>09:00 - 10:30</td>
<td>Special Topic Session 4</td>
<td>Non-Standard SAE Modeling</td>
</tr>
<tr>
<td></td>
<td>10:30 - 10:45</td>
<td>Coffee Break</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10:45 - 11:45</td>
<td>Invited Talk 7</td>
<td>J.N.K. Rao</td>
</tr>
<tr>
<td></td>
<td>11:45 - 12:00</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12:00 - 13:00</td>
<td>Invited Talk 8</td>
<td>W. Fuller</td>
</tr>
<tr>
<td></td>
<td>13:00 - 14:15</td>
<td>Lunch Break</td>
<td>Gatopardo Restaurant (1st floor)</td>
</tr>
<tr>
<td></td>
<td>14:15 - 16:15</td>
<td>Invited Talk 9</td>
<td>Panel – SAE on Government Programs</td>
</tr>
<tr>
<td></td>
<td>16:15 - 16:30</td>
<td>Coffee Break</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16:30 - 18:00</td>
<td>Contributed Session 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>18:00 - 18:15</td>
<td>Closing Ceremony</td>
<td></td>
</tr>
</tbody>
</table>
# 6. The Program in Detail

## Monday August 3rd, 2015

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:00 – 09:00</td>
<td>Registration</td>
<td>(Hallway, 2nd floor)</td>
</tr>
<tr>
<td>09:00 – 10:00</td>
<td>Opening Ceremony</td>
<td>• Authorities:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Dr. Ignacio Sánchez, Rector of Universidad Católica de Chile</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Mrs. Ximena Clark, Director of Instituto Nacional de Estadísticas de Chile</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Mr. Marcos Barraza, Minister of Ministerio de Desarrollo Social</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Dr. Gloria Icaza, President of Sociedad Chilena de Estadísticas</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Dr. Carolina Casas-Cordero, Local Organizing Committee SAE 2015</td>
</tr>
<tr>
<td></td>
<td>Key Note Speech: <strong>Meeting the Demand for Small Area Estimates: Direct and Indirect Estimation Methods</strong></td>
<td>• Dr. Graham Kalton, Westat (USA)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Chair: David Bravo (Chile)</td>
</tr>
<tr>
<td>10:00 – 10:15</td>
<td>Coffee Break</td>
<td>(Hallway, 2nd floor)</td>
</tr>
<tr>
<td>10:15 – 11:45</td>
<td>Invited Talk 1: <strong>From Start to Finish: Specify – Analyse / Adapt – Evaluate (SAE)</strong></td>
<td>• Speaker: Angela Luna-Hernandez (UK)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Chair: Luis Villarroel (Chile)</td>
</tr>
<tr>
<td>11:45 – 12:00</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>12:00 – 13:00</td>
<td>Invited Talk 2: <strong>Small Area Labor Force Statistics using Time Series Models</strong></td>
<td>• Speakers: Richard Tiller (USA), Danny Pfeffermann (Israel)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Discussant: William Bell (USA)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Chair: Raymond Chambers (Australia)</td>
</tr>
<tr>
<td>13:00 – 14:15</td>
<td>Lunch Break</td>
<td>(Gatopardo Restaurant, 1st floor)</td>
</tr>
<tr>
<td>14:15 – 16:15</td>
<td>Invited Talk 3: <strong>Panel on Recent Interesting Applications of Small Area Methods</strong></td>
<td>Organizers: Carolina Franco (USA), Junni Zhang (China)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Speakers:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• C. Franco (USA): <em>Combining Estimates from Related Surveys via Bivariate Area-Level Models without Covariates.</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• J. Zhang (China): <em>Bayesian Small Area Forecasting.</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• T. Fujii (Singapore): <em>Application of Small Area Estimation to Geographic Targeting</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• S. Haslett (New Zealand): <em>Diagnostic Techniques to Identify Anomalies in Small Area Estimation Models.</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• J. Sunthornchost (Thailand): <em>Bayesian Models for Poverty Mapping.</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Chair: Ralf Münich (Germany)</td>
</tr>
<tr>
<td>16:15 – 16:30</td>
<td>Coffee Break</td>
<td>(Hallway, 2nd floor)</td>
</tr>
</tbody>
</table>
MONDAY AUGUST 3RD, 2015 (CONTINUED)

16:30 – 18:30 Special Topic Session 1: New Perspectives on SAE
- Speaker:
  - Snigdhansu Chatterjee (USA): Small Area Inference using Resampling in the Presence of Transformations and Constraints.
  - Sumonkanti Das (Australia): Estimation of Small Area Distribution Functions under Heteroskedasticity
  - Trijya Singh (USA): A Robust Technique for Small Area Estimation in the Case of Missing Information on the Auxiliary
- Chair: Nicolás Libuy (Chile)

18:30 – 19:30 Reception + Poster session (Hallway, 2nd floor)
- Poster presenters:
  - Carmen Armas (Peru): Combining Agricultural Surveys and Censuses to Develop a Typology of Small and Medium Family Farms in Peru.
  - José Luis Avila (Chile): Using the EM algorithm in Small Area Estimation by the Fay-Herriot Model.
  - Rohosen Bandyopadhyay (USA): Benchmarking the Observed Best Predictor.
  - Isabelle Beaudry (USA): Inference for Respondent-Driven Sampling with Misclassification.
  - Andreea Erciulescu (USA): Bootstrap Confidence Intervals for Small Area Means.
  - Judith Law (USA): The Effect of Nonstationarity on the Rao-Yu EBLUP Derived under the Assumption of Stationarity.
  - Carla Leal (Chile): Evaluating Concordance between Chilean Comuna-level Poverty Rates Estimated using a Design-based Method and a SAE Method.
  - Jiyoun Myung (USA): Robust Small Area Estimation Modelling both the Mean and the Variance.
Tuesday August 4th, 2015

08:00 – 09:00 | Registration (Hallway, 2nd floor)
09:00 – 10:30 | Special Topic Session: Small Area Applications
   • Speakers:
     • Isabel Molina (Spain): Small Area Estimation of General Parameters under Complex Sampling Designs.
     • Ralf Munnich (Germany): Small Area Estimation of Spruce Timber using Penalized B-Splines.
     • Bac Tran (USA): Estimation of Employment for Small Areas Using the U.S. Census Bureau’s Annual Survey of Public Employment and Payroll.
     • William R. Waldron (USA): Empirical Bayes Estimation of Radio Ratings for Small Geographic Areas
   • Chair: David Niculcar (Chile)

10:30 – 10:45 | Coffee Break (Hallway, 2nd floor)
10:45 – 11:45 | Invited Talk 4: Small Area Estimation with Health Applications
   • Speaker: Malay Ghosh (USA)
   • Chair: Gloria Icaza (Chile)

11:45 – 12:00 | Break
12:00 – 13:00 | Invited Talk 5: On the Conditionally Calibrated Bayesian in Design and Analysis
   • Speaker: Donald Rubin (USA)
   • Chair: David Bravo (Chile)
13:00 – 14:15 | Lunch Break (Gatopardo Restaurant, 1st floor)
14:15 – 15:45 | Special Topic Session 3: SAE Foundation Issues
   • Speakers:
     • David Haziza (Canada): A Unified Framework for Robust Small Area Estimation.
     • Alan H. Dorfman (USA): Valid Confidence Intervals for Small Area Estimates.
     • Jae-Kwang Kim (USA): Measurement Error Model Approach to Small or Large Area Estimation Incorporating Several Sources of Information.
     • Marcin Szymkowiak (Poland): Robust Model Based Approach to Assess Accuracy of Internet Data Sources.
   • Chair: Wayne Fuller (USA)
15:45 – 16:00 | Coffee Break (Hallway, 2nd floor)
16:00 – 17:00 | Invited Talk 6: Small Area Model Selection and Jackknife
   • Speaker: Partha Lahiri (USA)
   • Chair: Carolina Casas-Cordero (Chile)
17:00 – 17:15 | Break
Tuesday August 4th, 2015 (continued)

17:15 – 18:30
Contributed Session 1
• Alvaro Castillo (Chile): Geographical Distribution of Alcohol-attributable Mortality in Chile: a Bayesian Spatial Analysis.
• Gloria Icaza (Chile): Lung Cancer Mortality in Chile: Spatio-Temporal Analysis.
• Jan Pablo Burgard (Germany): Small Area Estimation of Biodiversity Measures.
• Chair: Malay Ghosh (USA)

18:30 – 19:30 Break

19:30 – 22:00 Banquet
• Bus Pick-Up: from 19:30 to 19:45, Centro de Extensión UC (main entrance).
• Dinner: from 20:00 to 23:00, Los Adobes de Argomedo Restaurant.
• Bus Drop-Off: from 23:00 to 23:10 at
  ➢ Centro de Extensión UC (main entrance)
  ➢ Crowne Plaza Hotel
  ➢ El Fundador Hotel
## Wednesday August 5th, 2015

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:00 – 09:00</td>
<td><strong>Registration</strong> (Hallway, 2nd floor)</td>
</tr>
<tr>
<td>09:00 – 10:30</td>
<td><strong>Special Topic Session 4: Non-Standard SAE Modeling</strong></td>
</tr>
<tr>
<td></td>
<td>• <strong>Speakers:</strong></td>
</tr>
<tr>
<td></td>
<td>• Mamadou S. Diallo (USA): <em>Bayesian Monte Carlo Method for Estimating Small Area Complex Parameters under Unit-level Models with Skew-Normal Errors.</em></td>
</tr>
<tr>
<td></td>
<td>• Michail Sverchkov (USA): <em>Small Area Estimation under Informative Sampling and Nonresponse.</em></td>
</tr>
<tr>
<td></td>
<td>• Monica Pratesi (Italy): <em>Small Area Estimation Models with Outliers in Covariates.</em></td>
</tr>
<tr>
<td></td>
<td>• <strong>Chair:</strong> Isabel Molina (Spain)</td>
</tr>
<tr>
<td>10:30 – 10:45</td>
<td><strong>Coffee Break</strong> (Hallway, 2nd floor)</td>
</tr>
<tr>
<td>10:45 – 11:45</td>
<td><strong>Invited Talk 7: Measuring uncertainty of Small Area Estimators</strong></td>
</tr>
<tr>
<td></td>
<td>• <strong>Speaker:</strong> J.N.K. Rao (Canada)</td>
</tr>
<tr>
<td></td>
<td>• <strong>Chair:</strong> Carolina Casas-Cordero (Chile)</td>
</tr>
<tr>
<td>11:45 – 12:00</td>
<td><strong>Break</strong></td>
</tr>
<tr>
<td>12:00 – 13:00</td>
<td><strong>Invited Talk 8: Bootstrap Methods for Small Area Predictions</strong></td>
</tr>
<tr>
<td></td>
<td>• <strong>Speaker:</strong> Wayne Fuller (USA)</td>
</tr>
<tr>
<td></td>
<td>• <strong>Chair:</strong> Gloria Icaza (Chile)</td>
</tr>
<tr>
<td>13:00 – 14:15</td>
<td><strong>Lunch Break</strong> (Gatopardo Restaurant, 1st floor)</td>
</tr>
<tr>
<td>14:15 – 16:15</td>
<td><strong>Invited Talk 9: Panel on Small Area Methodology in Government Programs</strong></td>
</tr>
<tr>
<td></td>
<td>• <strong>Organizers:</strong> Ray Chambers (Australia), Domingo Morales (Spain), Denisse Silva (Brazil)</td>
</tr>
<tr>
<td></td>
<td>• <strong>Speakers:</strong></td>
</tr>
<tr>
<td></td>
<td>• Christine Bycroft (New Zealand), Statistics New Zealand.</td>
</tr>
<tr>
<td></td>
<td>• Débora Souza (Brazil), Brazilian Institute of Geography and Statistics</td>
</tr>
<tr>
<td></td>
<td>• Carol Crawford (USA), National Agricultural Statistics Service.</td>
</tr>
<tr>
<td></td>
<td>• Alan Dorfman (USA), National Center for Health Statistics.</td>
</tr>
<tr>
<td></td>
<td>• Frauke Kreuter (Germany), IAB Statistical Methods group, University of Mannheim, University of Maryland.</td>
</tr>
<tr>
<td></td>
<td>• Isabel Millán (Chile), Ministerio de Desarrollo Social.</td>
</tr>
<tr>
<td></td>
<td>• <strong>Chair:</strong> Ray Chambers (Australia)</td>
</tr>
<tr>
<td>16:15 – 16:30</td>
<td><strong>Coffee Break</strong> (Hallway, 2nd floor)</td>
</tr>
<tr>
<td>16:30 – 18:00</td>
<td><strong>Contributed Session 2</strong></td>
</tr>
<tr>
<td></td>
<td>• <strong>Speakers:</strong></td>
</tr>
<tr>
<td></td>
<td>• Tomasz Józefowski (Poland): <em>Small Area Estimation in the Light of the Europe 2020 Strategy.</em></td>
</tr>
<tr>
<td></td>
<td>• Luis Fringuetti (Chile): <em>A Simulation Study of Properties of Partial Least Squares and Ridge Regression.</em></td>
</tr>
<tr>
<td></td>
<td>• Marcin Szymkowiak (Poland): <em>At Risk of Poverty Mapping in Poland at NUTS 4 using Small Area Estimation Methods.</em></td>
</tr>
<tr>
<td>18:00 – 18:15</td>
<td><strong>Closing Ceremony</strong></td>
</tr>
</tbody>
</table>
7. The Abstracts

A

Combining Agricultural Surveys and Censuses to Develop a Typology of Small and Medium Family Farms in Peru

Armas Carmen (Group for the Analysis of Development, GRADE), Escobal Javier
carmas@grade.org.pe

This study combines the information of the 2012 Peruvian Agriculture Census with most recent information provided by the 2014 Peruvian Agricultural Survey on production levels and net income for the small and medium family farms. It estimates a mixed effect model including random effects at the province level using a two equation model that corrects for selectivity. The model corrects for selection bias after excluding twenty percent of observations which showed a negative net income. After including the inverse Mills ratio in the income equation, we predict for census observations (1) the probability of having a positive income and (2) the level of income given that the probability is positive for the census data, after which we aggregate the data at the district and provincial level. Using the simulated data we calculate four categories of agriculture producers. In the extremes we have a subsistence farmer, whose farm net income is insufficient to cover a minimum subsistent level, proxied by the official extreme poverty line; and a consolidated farmer, whose farm net income is enough to reduce the probability of been poor below 10%. The study also uses small area estimates to calculate the relative importance of production for self-consumption and for the market at the provincial level.

Key Words: family farms, agricultural producer, small areas, Peru.

Using the EM Algorithm in Small Area Estimation by the Fay-Herriot Model

Avila, José Luis (Universidad de Valparaíso), Riquelme Marco Antonio, Leiva Víctor
jose.avila@postgrado.uv.cl

In the Fay-Herriot small area model, standard variance component estimation methods can produce zero estimates of the strictly positive model variance. This problem is due to the empirical best linear unbiased predictor of a small area mean not take into account the variance of the random effect of small areas, reducing it only to a regression estimator. In this work, we consider the expectation-maximization (EM) algorithm to solve the mentioned problem. We show through Monte Carlo (MC) simulations that the EM algorithm always produces strictly positive model variance estimates. We compare the performance of the proposed results to two recently proposed methods using MC simulations. We apply the our results to poverty and food security data.

Key Words: EM algorithm, The Fay-Herriot model, Mean squared prediction error.
**Benchmarking the Observed Best Predictor**

Bandyopadhyay Rohosen (University of California, Davis), Jiang Jiming
rbandyopadhyay@ucdavis.edu

Jiang, Nguyen and Rao in their 2011 paper derived the Observed Best Predictor (OBP) in order to predict small area mean in a linear mixed effect model. In contrast to the traditional Empirical Best Linear Unbiased Estimator (EBLUP), which can be considered as a hybrid of estimation and prediction procedures, OBP was developed from a complete predictive point of view under possible model misspecification.

Benchmarking is a very important and required property of the predictors in small area estimation which requires the weighted sum of the small area predictors to be equal to the design consistent estimator. The original OBP does not satisfy the benchmarking property. In this paper we have developed two improved versions of OBP which also satisfies benchmarking property, one using an adjustment method and the other using an augmented model method. We have also developed a method of estimating the MSPE of the benchmarked OBP using a Monte Carlo Jackknife method.

*Key Words: Small Area Estimation, Observed Best Predictor, Benchmarking, MSPE, Jackknife, Monte Carlo Jackknife*

**Inference for Respondent-Driven Sampling with Misclassification**

Beaudry Isabelle S. (University of Massachusetts Amherst), Gile Krista J., Mehta Shruti H.
beaudry@math.umass.edu

RDS is a sampling method designed to sample from hard-to-reach human populations connected through social networks. The populations typically addressed by RDS are of interest to small area estimation research in that membership in the target population is often stigmatized, and therefore difficult to detect on general population surveys. RDS begins with a convenience sample, each participant receives a small number of coupons, which they distribute to their contacts who become eligible. RDS studies ask participants to report on the number of contacts they share with the studied population. Also, a set of characteristics is observed for each participant. The accuracy of these attributes is not considered in prevalence estimators. However, ignoring misclassification may lead to biased estimates. The main contribution of this study is to propose an analytical correction and to apply the Misclassification Simulation Extrapolation (SIMEX MC) procedure to correct for the bias introduced by misclassification. These two methods are assessed under varying levels of misclassification across simulated social networks of varying features. Also, the methods are illustrated with data from 26 studies performed in India. Finally, we conclude with our proposed extension of two bootstrap procedures to estimate the variability of the adjusted estimators.

*Key Words: Respondent-Driven Sampling; Misclassification (Measurement Error); Network Inference; Misclassification Simulation Extrapolation (SIMEX MC); Social Networks.*
**Multilevel Hierarchical Bayesian vs. State-Space Approach in Time Series Small Area Estimation: the Case of the Dutch Travel Survey**

Bollineni-Balabay Oksana (Statistics Netherlands, Quality Maastricht University), Van den Brakel Jan A., Palm Franz, Jan Boonstra Harm

o.balabay@maastrichtuniversity.nl

This study compares two time series techniques in a small area application: a multilevel time series model analyzed with a hierarchical Bayesian (HB) approach and a structural time series (STS) model, analyzed with the Kalman filter. The application chosen is the Dutch Travel Survey featuring level breaks caused by the survey redesigns, as well as small sample sizes at the provincial level. This paper also shows how STS models can be exploited to alleviate excessive volatility and possibly a bias in design variance estimates used as input in both models.

The two approaches deliver comparable results. Slight differences in variance estimates appear mostly in small-scaled domains and are due to neglecting uncertainty around the hyperparameter estimates in STS models, and to a lesser extent due to skewness in the posterior distributions. The results suggest that the reduction in design standard errors with the HB approach is above 50% and 30% at the provincial and national level, respectively. In other words, in order to reduce the true variance within the design-based framework to the extent the time series techniques do, one would have to increase the sample size more than four-fold (twice) for the provincial (national) level (conditional on the point-estimates).

**Key Words:** Gibbs sampling; Hierarchical Bayes; hyperparameter uncertainty; multilevel model; state space model.

**Small Area Estimation of Biodiversity Measures**

Burgard Jan Pablo (Trier University, Germany), Feldmeier Stephan, Münnich Ralf, Rosenthal Philip, Veith Michael

burgardj@uni-trier.de

Sustainability management and the protection of wildlife, plants and biodiversity, in general, are of increasing importance for policy makers. Therefore, good estimators for biodiversity measures that include information about species richness and abundance on regional level are needed. Hill-Numbers provide a framework for such measures and have recently been further developed (see [2] and [1]). Because sampling of wildlife is very costly, cumbersome and often restricted to certain regions, sampling sizes may vary and even be insufficient for some regions. In these small areas, classical estimation techniques are likely to yield inaccurate outcomes of biodiversity measures. Hence, model assisted small area methods are expected to lead to more reliable estimates assuming that relevant prior information is available for the areas of interest.

A multinomial mixed logit model (see [3]) is applied to estimate the relative species abundance as a major input for several nonlinear biodiversity measures based on Hill-Numbers. Further, above estimator is compared to the results of direct estimators using an animal data set. Finally, different set-ups are used to indicate the behaviour of design-based versus model-based estimators of Hill numbers.

**References:**


**Geographical Distribution of Alcohol-attributable Mortality in Chile: a Bayesian Spatial Analysis**

Castillo-Carniglia Álvaro (National Service for Prevention and Rehabilitation of Drug and Alcohol Consumption, SENDA), Kaufman Jay S., Pino Paulina
alvacast@gmail.com

Objective: To describe the distribution of alcohol-attributable mortality (AAM) at the local level (345 municipalities) in Chile, including fully and partially attributable causes of death in 2009. Methods: Attributable fractions to estimate AAM was calculated for the population 15 years of age and older using per capita alcohol consumption combined with survey estimates. The effect of alcohol on each cause of death was extracted from the published scientific literature. We used Bayesian hierarchical models to smooth the standardized mortality ratio for each municipality for six groups of causes related to alcohol consumption (total causes, neuro-psychiatric, cardiovascular, cancer, injuries and other causes). Results: The percentage of municipalities with high risk of AAM for any group of causes in each region ranges from 0% to 87.0%. Municipalities with high risk were concentrated in south-central and southern Chile for all groups of causes related to alcohol. Conclusions: AAM risk shows marked geographic concentrations, mainly in south-central and southern regions of Chile. This combination of methods for small-area estimates of AAM is a powerful tool to identify high risk regions and associated factors, and may be used to inform local policies and programs.

**Small Area Inference using Resampling in the Presence of Transformations and Constraints**

Chatterjee Snigdha (University of Minnesota), Lahiri Partha
chatterjee@stat.umn.edu

In many examples of small area applications to real data problems, a number of additional transformations, benchmarking, and robustness inducing steps are performed. One or more of these steps may be non-linear and non-smooth. Consequently, the actual probability model underlying such applications are very complicated, and routine small area estimation procedures may not be adequate for accurate and precise inference. We discuss the use of parametric bootstrap as an alternative when such transformations and constraints are in place in a small area framework. We present some theoretical and some numeric results.
Estimation of Small Area Distribution Functions under Heteroskedasticity

Das Sumonkanti (University of Wollongong), Chambers Ray
sd535@uowmail.edu.au

Multi-level models with homoskedastic nested errors are widely used for estimation of small area means, and can also be used to estimate the corresponding small area distribution functions. However, this type of model may not be adequate if these nested errors are heteroskedastic. In this study we consider estimation of small area distribution functions for a finite population defined by a two-level superpopulation model with unknown heteroskedasticity at level-one. In particular, we assume that the level-one variance for this model corresponds to a smooth function of the auxiliary variables. The proposed method uses a moment approach to first estimate the variance of homoskedastic level-two errors, and then a non-parametric smoothing technique is used to estimate level-one variances. This fit is combined with the Chambers and Dunstan (1986) type smearing approach to estimate small area distribution functions. Finally, a non-parametric bootstrap procedure based on this estimated distribution function is used to estimate small area quantities of interest with their mean squared errors. We compare this methodology with a parametric bootstrap procedure proposed by Elbers, Lanjouw and Lanjouw (2003) via simulation studies considering linear and log-linear two-level superpopulation models. Simulations results indicate that the proposed method shows some promise in a poverty mapping application.

Key Words: Auxiliary information; small area distribution function; hierarchical superpopulation model; non-parametric bootstrap procedure; smoothing technique.

Bayesian Monte Carlo Method for Estimating Small Area Complex Parameters under Unit-level Models with Skew-Normal Errors

Diallo Mamadou S. (Westat)
MamadouDiallo@westat.com

Small Area Estimation (SAE) methods are increasingly used to provide local estimates in support of public policy decisions. Under normality assumption, Molina et al. (2014) developed a hierarchical Bayesian (HB) approach to estimate small area complex parameters, in particular poverty indicators. When the distribution of the variable of interest is asymmetrical, normality-based estimators may be inefficient in terms of MSE especially for complex parameters. In this paper, we relax the normality assumption and consider a larger family called skew-normal which includes the normal distribution as a special case. The resulting HB method for the skew-normal model only uses Monte Carlo techniques and does not require Markov chain Monte Carlo (MCMC) methods. Avoiding MCMC is important since Monte Carlo methods do not have mixing chains issues. The posterior density has closed-form expression hence only the grid method and sampling importance resampling (SIR) technique are used to draw samples from the posterior distribution. Simulation results and application to survey data are presented to compare this HB method to the empirical Bayes (EB) method proposed by Diallo and Rao (2014).

Key Words: Hierarchical Bayes (HB), Monte Carlo, best prediction (EB), complex parameters, unit-level, skew-normal.
Validating Confidence Intervals for Small Area Estimates

Dorfman Alan H. (National Center for Health Statistics)
xws5@cdc.gov

We suggest an approach to validating confidence intervals for estimates derived from small area estimation procedures. The approach requires a small auxiliary survey complementing the principal survey and relies on results that relate (a) the rate at which confidence intervals from one survey include point estimates from another independent survey to (b) the ratio of mean square errors of point estimates from the two surveys (Dorfman, 2011) [http://www.pakjs.com/pjslib.php].

Key Words: bias, mean square error estimation, variance, working model.

Bootstrap Confidence Intervals for Small Area Means

Erciulescu Andreea L. (Iowa State University), Fuller Wayne A. andreeae@iastate.edu

Most small area studies focus on constructing predictors for the area means and on estimating the variance of the prediction errors. However, agencies and policy makers are often interested in confidence intervals for the small area predictors. We present two sided confidence intervals for the small area means of a binary response variable. We consider unit level data and stochastic covariates. The estimation of the prediction error variance and the estimation of the cutoff points are key components in the construction of confidence intervals for the small area means. A linear approximation of the model is considered and a Taylor variance approximation is presented for the prediction error variance. We compare different bootstrap estimation methods for the cutoff points using a simulation study.

Key Words: Unit level model; Auxiliary information; Parametric bootstrap; Fast double bootstrap; Mean squared error; Confidence interval.
Firinguetti Luis (University of Bio-Bio), Araya Rodrigo
lfiringu@ubiobio.cl

In calibration and the estimation of linear regression equations it is common to be confronted with the multicollinearity problem. A number of estimators have been proposed to tackle this problem, among them Ridge Regression (RR) and Partial Least Squares (PLS). In fact, RR has been proposed as an estimation method to deal with multicollinearity in the context of Small Area Estimation (SAE). Although, to the best of our knowledge, no such proposal has been made with regard to the PLS estimator, we are convinced that PLS may also be useful in SAE.

In this presentation, in the context of the Classical Linear Regression Model (CLRM) we compare, through simulations, the properties of PLS and RR with Ordinary Least Squares (OLS). We consider varying degrees of multicollinearity, different numbers of observations and explanatory variables and different magnitudes of the error variance. It is found that both, PLS and RR, estimators produce significant reductions in the Prediction Mean Square Error (PMSE) and in the Total Mean Square Error (TMSE) over OLS. When comparing RR and PLS, the last estimator tend to produce better results when there are more variables in the model and they present greater collinearity. On the other hand RR tend to perform better when the level of noise is greater.

Franco Carolina (US Census Bureau), Bell William R.
Carolina.Franco@Census.gov

We explore use of bivariate area-level models to improve small area estimates from one survey by borrowing strength from another, possibly larger survey. Through simple theoretical calculations and with several real data examples, we show that substantial reductions in standard errors may be achieved by borrowing strength in such a manner even without covariates obtained from auxiliary information. The theoretical calculations show how the extent of standard error reduction depends on the characteristics of the underlying data. In our applications, we borrow strength from estimates from the American Community Survey, the largest U.S. household survey, to improve estimates from smaller surveys, such as the National Health Interview Survey, the Survey of Income and Program Participation, and the Current Population Survey. We apply both a bivariate Fay-Herriot model, and, for proportions, a bivariate Binomial-Logit Normal (BLN) model.

Key Words: Small Area Estimation, Fay-Herriot, Binomial Logit Normal Model, shrinkage estimation, borrowing information, government survey.
APPLICATION OF SMALL AREA ESTIMATION TO GEOGRAPHIC TARGETING

Fujii Tomoki (Singapore Management University)
tfujii@smu.edu.sg

Application of small-area estimation has become increasingly popular over the last decade among policy-makers. Today, poverty maps based on small-area estimation have already been produced in dozens of countries. Despite this progress, how much we can gain from small-area estimates is not obvious or well known. In this study, we offer an analytical tool to highlight this issue and discuss empirical application to Cambodia.

BOOTSTRAP METHODS FOR SMALL AREA PREDICTIONS

Fuller Wayne A. (Iowa State University), Erciulescu Andreea L.
waf@iastate.edu

The unit level nested error model is an important model for prediction of small area means. We study bootstrap confidence intervals for the small area mean of data satisfying the logistic model. We give a Taylor variance approximation that is used to construct pivot-like statistics. Statistics that enable the user to construct an alpha-level confidence interval for a range of alphas are given.

Key Words: Confidence interval, Taylor approximation, small area estimation, double bootstrap.

DISEASE MAPPING: THE OTHER HALF OF SMALL AREA ESTIMATION

Ghosh Malay (University of Florida)
ghoshm@stat.ufl.edu

The paper reviews Bayesian and empirical Bayesian methods in the context of disease mapping. It traces the history and touches also on some of the more recent developments.
A Unified Framework for Robust Small Area Estimation

Haziza David (Université de Montréal), Cyril Favre Martinoz, Crest-Ensaı Irma, Tzavidis Nikos
haziza@DMS.UMontreal.CA

Influential units occur frequently in surveys, especially in the context of business surveys that collect economic variables whose distribution are highly skewed. A unit is said to be influential when its inclusion or exclusion from the sample has an important impact on the magnitude of survey statistics. Robust small area estimation has received a lot of attention in recent years; see Gosh et al. (2008), Sinha and Rao (2009), Dongmo Jiongo et al. (2013), Chambers et al. (2013) and Fabrizi et al. (2014), among others. So far, researchers have mainly focused on unit level models and continuous characteristics of interest. Several robust versions of the empirical best linear unbiased predictor based on linear mixed models (LMM) have been proposed in the literature, including an M-quantile regression approach and an approach based on the concept of conditional bias of a unit. In practice, one must often face binary and count data. In this case, methods based on LMMs are not suited. We propose a unified framework for robust small area estimation in the context of generalized LMMs. We construct a general robust estimator based on the concept of conditional bias. Results from a simulation study assessing the performance of the proposed method in terms of bias and efficiency will be presented.

Diagnostic Techniques to Identify Anomalies in Small Area Estimation Models

Haslett Stephen (Massey University), Sefton Alison
s.j.haslett@massey.ac.nz

An important application of small area estimation is its use in poverty mapping, to provide estimates which allow organizations to implement policies and target aid to the areas in the greatest deprivation. A useful extension is to identifying the prevalence of under-nutrition in children less than five years of age. Generally in SAE, diagnostics and model checking techniques are important tools to ensure the predicted small area statistics are unbiased and to identify any anomalies in the data which may be causing some small area estimates to be unreliable. By identifying the effect of each variable in the model in each the small areas, it is possible to identify which variables in the model are causing a particular small area to be unusual. Applying the diagnostic technique to malnutrition data in Nepal we were able to identify which combinations of variables were driving a particular small area’s estimated under-nutrition rate to be exceptionally high. As a result, an error was identified in one of the auxiliary variables. After correcting this error the under-nutrition map could be updated and a potential misallocation of funding was avoided.
Simple Parametric Bootstrap Estimator of Mean Squared Error for EBLUP under the Fay-Herriot Small Area Model

Hirose Masayo (The Institute of Statistical Mathematics), Lahiri Partha
masayo@ism.ac.jp

It is well-known that existing simple parametric bootstrap estimator of the mean squared error (MSE) estimator of empirical best linear unbiased predictor (EBLUP) of a small area mean is not second-order unbiased. The MSE estimator can be biased corrected using single or more computationally double bootstrap procedure to produce second-order unbiased MSE estimator. Such bias correction, especially double bootstrap, is computer-intensive and, more importantly, does not ensure strictly positive MSE estimator. Adjustments to ensure positivity have been proposed in the literature (Hall and Maiti, 2006), but a rigorous proof of such second-order unbiasedness of adjusted parametric bootstrap method is missing in the literature. The paper presents a simple single parametric bootstrap MSE estimator that does not require any bias adjustment procedure and ensure the important dual properties: second-order unbiasedness and strictly positivity. The procedure is explained using the well-celebrated Fay-Herriot small area model.

Key Words: MSE of EBLUP, Parametric bootstrap

Lung Cancer Mortality in Chile: Spatio-Temporal Analysis

Icaza Gloria (Universidad de Talca), Buzeta Andrés, Torres-Avilés Francisco, Núñez Loreto
gicaza@utalca.cl

Lung cancer is the first cause of cancer death in the world. In Chile is the second cause of cancer death after stomach cancer. Chile has one of the highest prevalence of tobacco smoking in the world, in addition, between 1958 and 1970 the northern region of Antofagasta suffered from arsenic drinking water contamination. Both tobacco and arsenic, well known lung cancer risk factors. This scenario offers the opportunity to study trends in space and time to know how are the risks (measured by means of standardized mortality rates, SMR) of mortality from lung cancer counties in the period 2000-2010. Previous studies show that mortality risks are high in the regions of Antofagasta and the Metropolitan Region, and no significant changes in overall country trends in both sexes. Therefore, it is of interest to analyze the risks of mortality in a disaggregated manner, taking into account the space and time at community level.

Mortality and population data was obtained by the Chilean Ministry of Health. Intercensal population projections come from 2002 Census. Knorr-Held and Besag (2001) and autoregressive Martinez-Beneito et al. (2007) models were compared. Both models have different structures for risk estimation: the Knorr-Held and Besag model is very similar to convolution Besag, York and Mollie model (1991), with the difference that a temporal convolution component, which does not generate interaction, is added to the spatial component. Moreover, Martinez-Beneito et al. propose to incorporate an autoregressive component into the model, therefore the risks estimates depend on what happened in the previous period, in order to generate a space-time dependency in the estimates. As prior information on the variability, various informative flat distributions were tested and the best selection
Small Area Estimation in the Light of the Europe 2020 Strategy

Józefowski Tomasz (Poznan University of Economics), Beręsewicz Maciej, Szymkowiak Marcin, Wawrowski Łukasz
t.jozefowski@stat.gov.pl

The Europe 2020 Strategy is a 10-year plan of the European Union initiated in 2010 and aiming to improve economic growth and the employment rate. The objective of the strategy is not only to solve problems related to the economic crisis that is gradually being overcome by EU member states but also to correct the shortcomings of the European model of economic growth and create conditions for smart, sustainable and inclusive growth. The Europe 2020 Strategy addresses 5 main areas, which the European Union is planning to improve by 2020. These include employment, research and development, climate and energy, education, social integration and eradication of poverty.

One of the main targets of the Europe 2020 Strategy is to reduce the number of Europeans living below national poverty lines from 120 million to 100 million. To enable poverty mapping and estimation the Europe 2020 Strategy has specified a set of indicators based on data from the EU-SILC survey.

The task of monitoring the implementation of the Europe 2020 Strategy in Poland is not possible at lower levels of spatial aggregation owing to a small sample size of the EU-SILC survey. For this reason in Poland selected indicators from the Europe 2020 Strategy are only published at the national level. This situation can be remedied by applying small area estimation. More detailed information at a lower level of spatial aggregation (e.g. NUTS 2) would improve diagnosis of the socio-economic situation and would allow a more adequate planning of social policy and intervention.

The main aim of the paper is to analyse how small area estimation (SAE) could be used to estimate the indicator of low work intensity – one of the headline indicators of the Europe 2020 Strategy – at NUTS 2 level, i.e. two levels lower than the national estimates published to date.
The presentation will review selected SAE estimators which account for correlations in time and space [1], [2], [3], [4]. Theoretical considerations will be illustrated by exploring the possibilities of applying indirect estimation to estimate the indicator of low work intensity in Poland at NUTS 2 level. The study and its results will contribute to implementing the Europe 2020 Strategy, which obliges all UE member states to respect its objectives.

*Key Words: small area estimation, strategy Europe 2020, indicator of low work intensity*

**References:**


---

**MEETING THE DEMAND FOR SMALL AREA ESTIMATES: DIRECT AND INDIRECT ESTIMATION METHODS**

Kalton Graham (Westat)
grahamkalton@westat.com

Recent years have seen increasing demands by policy makers and others for survey estimates for small geographical areas. The demands span a range of subject-matter and a variety of types of small area, from the regions of a country down to local administrative districts. This paper presents an overview of methods for producing small area estimates. Survey sample designs may be fashioned to produce standard direct survey estimators that have adequate levels of precision for fairly large geographic areas. However, survey samples are rarely large enough to produce direct estimates of acceptable precision for smaller areas. To respond to the demands for estimates for smaller areas, a wide range of statistical models have been developed for producing indirect, model-dependent, estimates. This paper outlines the components of a program for producing indirect small area estimates, with an emphasis on the need to find auxiliary data that are good predictors for the particular small area estimate under study. The paper also comments on the quality of indirect small area estimates.
A Unified Monte-Carlo Jackknife for Small Area Estimation after Model Selection

Lahiri Partha (JPSM, University of Maryland), Jiang Jiming, Nguyen Thuan
plahiri@umd.edu

We consider estimation of mean squared prediction error (MSPE) in small area estimation (SAE) when a procedure of model selection is involved prior to the estimation. We discuss the difficulty of achieving both second-order unbiasedness and positivity at the same time, the so-called double-goal, in MSPE estimation, and propose a simple alternative by estimating the logarithm of the MSPE. A unified Monte-Carlo jackknife method, called McJack, is proposed for estimating the log-MSPE. We demonstrate the performance of McJack in assessing uncertainty in SAE after model selection through theoretical and empirical studies.

Measurement Error Model Approach to Small or Large Area Estimation Incorporating Several Source of Information

Kim Jae-kwang (Iowa State University), Wang Zhonglei, Park Sangun, Young Linda
jkim@iastate.edu

Combining information from different source is an important practical problem. Using hierarchical measurement error model for area level models, we establish a framework for combining information from different source to get improved prediction for small or large area estimation. The best prediction is obtained by the conditional expectation of the observable latent variable given all available observation. The model parameters are estimated by two-level EM algorithm. Estimation of the mean squared prediction error is discussed. Sponsored by National Agricultural Statistical Agency (NASS) of US department of Agriculture, the proposed method was applied to the crop acreage prediction problem combining information from three sources: The first source is the June Area Survey (JAS), which is obtained by the probability sampling. The second source is from the Farm Service Agency (FSA) data, which is obtained from a voluntary participation of certain programs. The third source is from the classification of the Cropland Data Layer (CDL).
The Effect of Nonstationarity on the Rao-Yu EBLUP Derived Under the Assumption of Stationarity

Law Judith (University of Maryland)
jlaw@math.umd.edu

In time series and cross-sectional small area models, both stationary and random walk models have been used to borrow strength across past data and areas. The standard practice has been to choose one of the two options for modeling the time series component and develop small area estimators and the corresponding mean squared error estimators for the small areas. We have designed a Monte Carlo simulation experiment to examine the effect on performances of the small area estimators and the corresponding mean squared error estimators developed under a stationary time series cross-sectional model, e.g., Rao and Yu (1994), when a random walk model or a stationary model with autocorrelation very near one is used to generate data. We use estimates from the Current Population Survey (CPS) to design a Monte Carlo simulation experiment to examine the effect of nonstationarity on the mean squared error of the Rao-Yu EBLUP and their MSE estimator as the first order autocorrelation coefficient becomes large and actually equal to one, and when it exceeds one, under varying levels of between-small-area variation.

Key Words: autocorrelated errors, best linear unbiased prediction, cross-sectional data, Fay-Herriot model, mean-squared error estimation, Rao-Yu model, small area estimation, nonstationarity, time-series data.

Evaluating Concordance Between Chilean Comuna-level Poverty Rates Estimated Using a Design-based Method and a SAE Method

Leal Carla (Universidad Católica de Chile), Galea Manuel
cgleal@uc.cl

When moving from design-based estimation methods to model-based estimation methods, evaluation of model assumptions are important. Evaluating the degree of agreement between estimates derived from these alternative methods also seems like a desirable property to assess. Specially when new estimation methods are to be introduced in replacement of methods that are already in place for deriving the target estimates.

The degree of agreement between measurements generated by different methods can be evaluated by the Concordance Correlation Coefficient (CCC) [1]. In practice it is common to assume that the data are normally distributed, however in the presence of outliers, the normality assumption is no larger valid, leading to biased estimates of the CCC, thus affecting the conclusions about the agreement between the measurements.

To overcome the above problem, we propose to estimate the CCC based on the t-Student distribution. Further, to detect the sensitivity at the estimator of the CCC, different perturbations scheme are applied and local influence is studied.

We apply our methodology to estimates of comuna-level poverty rates derived using two alternative estimation methods: (1) a traditional design-based method for large areas (“direct method”), and (2) a small area estimation method (“SAE method”) developed by the Ministry of Social Development for comuna-level poverty rates estimated from the National Socioeconomic Characterization Survey (Casen) for the years 2009 and 2011 [2][3].
The objective of this presentation is to outline some of the key steps in applying Small Area Estimation (SAE). This is a vast topic and inevitably we will not cover every aspect of SAE. To put things into context and provide some focus, the presentation is built around a recent SAE case study in Mexico. This case study offers to us the opportunity to explore the practical steps in using a range of SAE methods -for discrete and continuous outcomes- that aim at estimating linear (averages and proportions) and non-linear indicators (percentiles of income distributions, inequality and complex deprivation indicators).

The presentation will start with the definition of the problem and the target parameters. The clear definition of the target parameters is of paramount importance as this will determine the data required for implementing different SAE methods. The next step discusses the definition of the domains of interest and the study of the sample size and coverage for these domains. The target parameters, the data availability and the sample size and coverage are key aspects that will determine the types of SAE methods to be used i.e. direct vs. indirect and unit level vs. area level methods.

A key target in the Mexico case study is the estimation of income-based linear and non-linear indicators. Other key targets involve SAE with multi-category, binary and count outcomes. Small area estimates must be produced at the level of municipalities with over 50% of municipalities being out of sample and the remaining consisting of small to moderate sample sizes. Hence, the use of indirect methods seems appropriate in this case. The presentation proceeds by exploring the use of industry standard EBLUP and synthetic-type estimators under the unit level nested error regression model (Battese et al., 1988). Extensions of this model for discrete outcomes are also considered. For the case of a continuous outcome we focus on the use of Empirical Best Prediction (Molina & Rao, 2010). The process of building the model and the use of residual diagnostics for checking the model assumptions are described. The presentation then discusses in some detail the use of transformations for meeting the model assumptions. The practical choice of transformations and the sensitivity of SAE to different types of transformations are also discussed. Approaches to elaborating the model are discussed. This includes the use of robust/non-parametric methods, methods with more that two levels.

A substantial part of the presentation is devoted to SAE evaluation. This includes the estimation of the Mean Squared Error and of Coefficients of Variation, issues in designing model-based and design-based simulations and the use of benchmarking. Finally, the presentation discusses the availability of SAE software and the use of parallel computing for handling the use of computer intensive methods with large datasets.
**Small Area Estimation of General Parameters under Complex Sampling Designs**

Molina Isabel (*Universidad Carlos III de Madrid*), Guadarrama M., Rao J.N.K., Tillé Y.  
isabel.molina@uc3m.es

The empirical best (EB) method for estimation of general parameters (e.g. poverty indicators) in small areas assumes that the population model holds for the sample and it is basically designed only for simple random sampling, since it does not incorporate the sampling weights in the estimation procedure. As a consequence, it can deliver biased estimates under severely informative sampling. We propose a Pseudo EB method designed for estimation of general parameters in small areas when the sample is drawn by complex sampling designs. We analyze the properties of this new method under simple and complex designs, including informative sampling. Keywords: Empirical best estimator; Nested-error model; Poverty mapping; Unit level models.

**An Adjusted Estimator for the Fay-Herriot Log-Level Model with Measurement Error in Covariates**

Mosaferi Sepideh (*University of Maryland*)  
smosafer@umd.edu

The Fay-Herriot model in a logarithmic scale has been increasingly used for the last few years to produce reliable estimates for fund allocation and regional planning. This model can borrow strength from relevant auxiliary information to provide more accurate estimates of population parameters with reduced mean squared error compared to the direct estimator. However, if covariates suffer from measurement error, the best predictor derived under the assumption of no measurement errors in the covariate may result in larger mean squared error than simple direct estimator. In this paper, we rectify the problem by proposing an adjusted estimator under the Fay-Herriot log-level model with measurement errors in covariates and illustrate its performance using real life data analysis.  

*Key Words*: Adjusted back transformed estimator; Adjusted likelihood; Scoring algorithm; Fay-Herriot log-level model; Measurement error.

**Small Area Estimation of Spruce Timber using Penalized B-Splines**

Munnich Ralf (*University of Trier*), Wagner Julian, Hill Joachim, Lamprecht Sebastian, Stoffels Johannes, Udelhoven Thomas  
muennich@uni-trier.de

Recently, SAE techniques were applied to estimate mean forest biomass in small areas in the Norwegian National Forest (see [1]) and to estimate spruce timber reserves in individual forestry restricts of Rhineland-Palatinate, a federal state of Germany (see [3]). Considering the data in [3], a non-linear relationship between the response and the auxiliary variable, which is the mean canopy height derived from airborne laser scanning, is recognized. Hence, penalized B-splines, a powerful tool in...
non-parametric modelling (see [2]), are used to model the non-linear relationship. One of the main advantages of penalized B-splines is the mixed model representation which allows a direct link to small area models (cf. [4] and [5]). However, this link is only possible under specific assumptions which may be violated in practice. Therefore, an alternative approach, also based on penalized B-splines, is proposed that does not rely on normal assumptions. The before mentioned methods are compared to each other in a model-context and at finally applied to estimate spruce timber reserves in individual forestry restricts of Rhineland-Palatinate.

Key Words: Small Area Estimation; Penalized B-splines; Forest Resource Estimation

References:

ROBUST SMALL AREA ESTIMATION MODELLING BOTH THE MEAN AND THE VARIANCE

Myung Jiyoun (University of Florida), Ghosh Malay, Moura F.A.S. 
jiyoun@ufl.edu

A typical area model assumes the variance \( V \) of the error term is known in order to avoid non-identifiability. The assumptions of known \( V \) almost becomes mandatory for secondary users of survey data who do not have access to any micro data for estimation of the \( V \). In reality, however, the \( V \) are random based on sampled data. Thus, in situations when one has additional data to estimate the \( V \), they can be used efficiently for modelling the \( V \).

We address small area estimation problems where we have additional data to model the variance \( V \) of the error term in the area level model. Also, for robustification, we assume t-distribution of the random effects. For a full Bayesian analysis, we find prior distributions for all the hyper-parameters. One of the key features in our analysis is including the degrees freedom \( \nu \). A modified version of Jeffreys’ prior leads to posterior propriety. For MCMC implementation, we use the Metropolis-Hasting algorithm to generate samples for analyzing data.

Key Words: Modelling the variance, Robustification, t-distribution, Modified Jeffreys’ prior, Bayesian analysis
Vaccination Coverage in India: A Small Area Estimation Approach

Santanu Pramanik (Public Health Foundation of India), Muthusamy Nithiyananthan, Gera Rajeev, Laxminarayan Ramanan
santanu.pramanik@phfi.org

Information on population health indicators in India come from a number of surveys that vary in frequency, scope and detail. In the case of immunization, the most recent coverage indicators are derived from Annual Health Survey (AHS) which was conducted only in 9 of 35 states and union territories. The most recent national survey of immunization coverage was conducted in 2009 (Coverage Evaluation Survey) by UNICEF. Therefore, reliable immunization coverage data for the entire country since 2009 is lacking. We used an established approach of small area estimation to predict coverage rates of several vaccinations for the remaining 26 states. We considered a linear mixed model that combines data from five cross sectional surveys representing five different time points. Our model encompasses sampling error of the survey estimates, area specific random effects, autocorrelated area by time random effects and hence, borrows strength across areas and time points both. Model-based estimates are almost identical to the AHS estimates for the nine states, suggesting that our model provides reliable prediction of vaccination coverage as AHS estimates are highly precise because of their large sample size. Results indicate that coverage inequality between rural and urban areas has been reduced significantly for most states in India.

Small Area Estimation Models with Outliers in Covariates

Pratesi Monica (University of Pisa), Giusti Caterina, Marchetti Stefano, Salvati Nicola
monica.pratesi@unipi.it

The presence of outliers is a common feature in real data. Chambers [1] classifies outliers into two groups. Representative outliers are correctly measured sample values that are outlying relative to the rest of the sample data and for which there is no reason to believe that similar values do not exist in the non-sampled part of the survey population. Non-representative outliers are gross errors in the sample data, which have nothing to do with the values in the non-sampled part of the survey population. Either type of outlier can have a substantial impact on the estimates, if ignored.

In the context of model-based estimation outliers can affect both the response and the auxiliary variables. The special case of outliers in the response variable has been treated in the literature both under the linear mixed models and under the M-quantile approach to SAE ([2]; [3]; [4]). Sinha and Rao [2] suggest a way to extend their robust small area estimator to account for outliers in the auxiliary variables. In this paper we develop their estimator, and we propose a new M-quantile based small area estimator that account for outliers in the covariates.

Key Words: M-quantile, Robust estimation, Influence function.

References:
Measuring Uncertainty of Small Area Estimators

Rao J. N. K. (Carleton University)
jrao@math.carleton.ca

Model-based indirect estimators of small area means or totals, based on linking models that can borrow information across small areas and thus increase the efficiency of area estimators, are widely used. In particular, area level and unit level models have received a lot of attention in the literature. In this review talk, I will address issues related to measuring uncertainty associated with empirical best (EB) estimators associated with the basic area level and unit level models. We study different linearization based MSE estimators based on unconditional and conditional frameworks and also jackknife and bootstrap MSE estimators. Parametric confidence intervals on the area means will also be studied. Hierarchical Bayes (HB) estimators and associated posterior variances under assumed priors on the model parameters will also be studied, including the choice of matching priors that can provide frequentist validity to posterior variances.

Synthetic Data Sources in the Spatial Analysis of Poverty in Poland

Roszka Wojciech (Poznan University of Economics)
wojciech.roszka@ue.poznan.pl

One of the key elements of the state’s social policy is the poverty prevention. Information about its spatial diversity is very helpful in the context of allocation of resources and to take measures to prevent its growth. Characteristics linked to the poverty estimates are mainly based on European Union Statistics on Income and Living Conditions (EU - SILC), which is a sample survey conducted annually in all countries of the European Union. Limitations associated with the sample size mean that the at-risk-of-poverty rate are estimated at most to the level of the provinces (NUTS 2, voivodeships in Poland). The application of the small area estimation methods (SAE) allows better quality estimation without increasing the sample. The problem in the context of poverty mapping in Poland was, among others, raised by Wawrowski (2014), who used a Fay - Herriot model to estimate the at-risk-of-poverty rate at NUTS 3 level. Simultaneously with the model approach methods, techniques for creating synthetic

reconstruction methods will be presented. With the joint usage of statistical matching methods (Raessler 2002) and iterative proportional fitting models (Rahman 2008) it is possible to construct the synthetic micro-populations at a small area level in such a way that all known constraints at the small area level are reproduced.

The aim of the paper is to assess the possibility of obtaining multivariate estimates of acceptable quality for at-risk-of-poverty at NUTS 4 (powiats in Poland) using a synthetic data set created on the basis of EU - SILC 2011 dataset and information from the census. The resulting estimates will be evaluated by their reliability, consistency and quality, as well as a comparative analysis will be carried out with the results obtained by “classical” SAE models.

Key Words: small area estimation, synthetic reconstruction, data integration, poverty mapping

ON THE CONDITIONALLY CALIBRATED BAYESIAN IN DESIGN AND ANALYSIS

Rubin Don (Harvard University)
rubin@stat.harvard.edu

A variety of statisticians have argued for using Bayesian thinking to create procedures and frequentist operating characteristics to ensure that selected procedures are calibrated across a variety of realistic situations. This presentation supports that approach at the design stage, meaning before the actual data set has been collected and observed, but distinguishes that calibration from the more refined and conditional calibration that a statistician should employ after seeing the data set and selecting procedures to trust for drawing inferences from that specific data set. Such evaluations are particularly appropriate when considering Bayesian models for small area estimation.

S

A ROBUST TECHNIQUE FOR SMALL AREA ESTIMATION IN THE CASE OF MISSING AUXILIARY INFORMATION

Singh Trijya (Le Moyne College)
singht@lemoyne.edu

Small area methods typically combine direct estimates from a survey with predictions from a model to obtain estimates of population quantities such that mean squared error is minimized and auxiliary information is used in many models to develop such a predictor. However, in several cases, covariate information for some small areas may be missing due to non-availability of samples. In this scenario, traditional small area estimation methods will not be suitable because the distribution of the response variable may not be normal. Moreover, missing covariate values lead to increased bias and reduced efficiency of parameter estimates. In this paper, a robust method of estimation of parameters for small areas is proposed for this scenario. This technique does not rely on any assumptions related to the distribution of the missing covariate information. A simulation study is conducted to empirically investigate the performance of the proposed method in terms of robustness and efficiency.
A Robust Technique for Small Area Estimation in the Case of Missing Auxiliary Information

Singh Trijya (Le Moyne College)
singht@lemoyne.edu

Small area methods typically combine direct estimates from a survey with predictions from a model to obtain estimates of population quantities such that mean squared error is minimized and auxiliary information is used in many models to develop such a predictor. However, in several cases, covariate information for some small areas may be missing due to non-availability of samples. In this scenario, traditional small area estimation methods will not be suitable because the distribution of the response variable may not be normal. Moreover, missing covariate values lead to increased bias and reduced efficiency of parameter estimates. In this paper, a robust method of estimation of parameters for small areas is proposed for this scenario. This technique does not rely on any assumptions related to the distribution of the missing covariate information. A simulation study is conducted to empirically investigate the performance of the proposed method in terms of robustness and efficiency.

Robust Model Based Approach to Assess Accuracy of Internet Data Sources

Szymkowiak Marcin (Poznan University of Economics), Beręsewicz Maciej
m.szymkowiak@ue.poznan.pl

Estimation conducted in statistical studies relies heavily on data from different data sources, namely surveys, registers and censuses. Nowadays there is a growing interest in the possibility of using data from new sources of information, such as the Internet or what is popularly known as big data. Information from those sources is increasingly being used in the model-based approach as a source of auxiliary variables. This is also true of small area estimation, in particular in the field of area-level models, for example supplied with data from Google Trends [1].
The main aim of the paper is to present an attempt at estimating selected characteristics of the real estate market in Poland using information from Internet data sources. For this purpose, the authors plan to apply the model-based approach to estimate the bias of data from administrative registers proposed by [2].
The authors intend to present an extended concept of modelling bias described by [2] using robust linear mixed models (capable of handling outliers) with autocorrelation in the context of Internet data sources. The methods presented in the paper will be exemplified on real data from real estate portals in Poland.

Key Words: small area estimation, modelling uncertainty, linear mixed models, Internet Data Sources, big data, real estate market.

References:
At Risk of Poverty Mapping in Poland at Nuts 4 using Small Area Estimation Methods

Szymkowiak Marcin (Poznan University of Economics), Wawrowski Łukasz
m.szymkowiak@ue.poznan.pl

The European Survey on Income and Living Conditions (EU-SILC) is the basic source of information published by GUS (Central Statistical Office in Poland) about the relative poverty indicator both for the country as a whole and at the regional level. This also applies to other countries facing a growing demand for good poverty maps. In order to follow appropriate social, which is consistent with the guidelines of the cohesion policy, one needs to measure poverty and provide information about this phenomenon at lower levels of spatial aggregation. In this context poverty maps are used to support decisions concerning important political issues, such as allocation of development funds by governments, National Ministries of Infrastructure and Development or international organizations, such as the World Bank. Those decisions should be based on the most accurate poverty indicators, estimates or numbers and should be delivered at the lowest level of spatial aggregation.

However, given the small sample size in the relevant cross classifications of the EU-SILC survey, it is necessary to use the latest techniques of indirect estimation and draw on alternative data sources to estimate the parameters of interest at low levels of spatial aggregation with acceptable precision. Since the EU-ILC survey does not cover adequately all the specific areas or population subgroups, the required information can only be obtained using small area estimation techniques based on the idea of “borrowing strength”. In Poland, for instance, EU-SILC data are only sufficient to publish the at-risk-of-poverty rate at the level of the whole country and at the regional level (NUTS 1). Owing to small sample sizes and low precision of estimation, adequate estimates at lower level of spatial aggregation cannot be delivered.

In the field of poverty mapping there are different approaches to choose from. In particular, common SAE-based poverty mapping methods may include: (1) direct estimates, which are in generally inefficient, (2) Fay-Herriot estimates, which enable aggregation, specific modelling, specification of sampling variances, (3) ELL estimates, which are used by the World Bank and may be poorly efficient when auxiliary variables do not explain the entire between-area variation, (4) the EB approach based on a nested-error model, which is very efficient under normality, (5) the HB approach based on a nested-error model, which is similar to the EB approach but is less computationally demanding and (6) M-quantile methods, which are less sensitive to outliers.

All of these methods may be used to tackle the problem of estimating different poverty parameters. In particular, these methods enable poverty mapping at lower levels of spatial aggregation. In Poland SAE methodology has so far been used in the area of the labour market, agriculture and business statistics. In 2013 the Centre for Small Area Estimation, which is a special unit at the Statistical Office in Poznań, in cooperation with GUS and the World Bank prepared a poverty map of Poland at the level of subregions (NUTS 3) using the Fay-Herriot approach.

small area estimation techniques and data from different statistical sources (EU-SILC, census or administrative registers).

The main aim of the research is to show theoretical aspects of small area estimation methods in the context of poverty mapping. Theoretical considerations will be illustrated with practical applications of SAE techniques in Poland at lower levels of spatial aggregation – NUTS 4 i.e. lower level which has not been published by the Central Statistical Office to date.

Key Words: small area estimation, mapping poverty, at risk of poverty, EU-SILC, census, administrative registers
Bayesian Models for Poverty Mapping

Suntornchost Jiraphan (Chulalongkorn University), Lahiri Partha
Jiraphan.S@chula.c.th

Poverty mapping that displays spatial distribution of various poverty indices such as poverty ratio, poverty gap, and poverty severity are most useful to policymakers and researchers when they are disaggregated into small geographic units, such as cities, municipalities, regions, or other administrative partitions of a country. Typically, national household surveys that contain welfare variables such as income and expenditures provide limited or no data for small areas. In this paper, we develop a Bayesian methodology for producing poverty maps and the associated uncertainties.

Key Words: Hierarchical models, Poverty mapping, FGT indicator, Cross-Validation, Bayesian model, Monte Carlo simulations.

Small Area Estimation under Informative Sampling and Nonresponse

Sverchkov Michail (Bureau of Labor Statistics), Pfeffermann Danny
Sverchkov.Michael@bls.gov

Pfeffermann and Sverchkov [1] considered Small Area Estimation for the case where the selection of the sampled areas is with probabilities that are related to the true (unknown) area means, and the sampling of units within the selected areas is with probabilities that are related to the study variable values, in both cases after conditioning on the model covariates. In this paper we extend this approach to situations of incomplete response, and where the response is not missing at random (NMAR). For this, we define a response model and estimate the response probabilities following the Missing Information Principle: we define the likelihood for the whole sample, pretending that the outcome is also observed for the nonrespondents, we then integrate out the unobserved outcomes from this likelihood over the outcome distribution for the nonrespondents, employing the relationship between the sample and sample-complement distributions, and finally we solve the resulting likelihood equations following [2]. Once the response probabilities are estimated, we multiply them by the sample selection probabilities and apply the approach of [1] to the observed data, again employing the relationships between the sample, sample-complement and population distributions. We illustrate our approach by a small simulation study.

References:
Conditional Probabilities for Small Subdomains Predictions in Panel Surveys

Thibaudeau Yves (U.S. Census Bureau), Slud Eric
yves.thibaudeau@census.gov

We show how to purposefully model conditional probability for the accurate prediction of small subdomains in the context of longitudinal survey measurements. We discuss methods that lead to “hybrid” predictors and estimators integrating conditional probabilities derived from log-linear models, extending the logistic models of Pfeffermann et al. (1998), and weights derived from the longitudinal design of survey. We present an example associated with the Survey of Income and Program Participation (SIPP). Because of its full-panel structure, SIPP can be seen as the ideal survey for longitudinal analyses. We present our method in the context of Markov modeling discussed by several authors (Fienberg 1980, Conaway Lohr 1994). We suggest a test for its validity based the generic test of Pfeffermann (1993).

Small Area Labor Force Statistics using Time Series Models

Tiller Richard (U.S. Bureau of Labor Statistics), Pfeffermann Danny
tiller_r@bls.gov

Periodic Labor Force Surveys (LFS) are conducted by government agencies to collect information on levels and change in employment and unemployment. These sets of statistics are the most important real time indicators of current economic activity. The purpose of this presentation is to review past and current developments in time series modeling of small area labor force statistics, with special attention to time series benchmarking. The first part of the presentation reviews research papers that formed the basis for the development of time series modeling of LFS data. The second part considers extension of the time series models by benchmark equations which guarantee consistency of publication and robustifies the inference against possible model failures. We compare time series model benchmarking with cross-sectional model benchmarking and extend the time series benchmarking to two-stage benchmarking. We conclude with discussion of the use of time series models for the production of official small area Labor Force estimates.

Key Words: Benchmarking, State Space Models, Correlated sampling error
Estimation of Employment for Small Areas Using the U.S. Census Bureau’s Annual Survey of Public Employment and Payroll

Tran Bac (U.S. Census Bureau)
bac.tran@census.gov

We illustrate a small area estimation methodology to estimate employment by combining the U.S. Census Bureau’s Annual Survey of Public Employment and Payroll (ASPEP) with the previous census records using an empirical best prediction (EBP) methodology. The employment data are usually subject to skewness and heteroscedasticity and thus the well-known EBP methodology based on unit level linear mixed normal model does not fit well. In order to get around the problem, we apply a unit level linear mixed normal model on the log-transformed employment. We compare different estimators by a design-based simulation using a finite population constructed from the 2012 Census of Government units. We also consider different mean squared error estimators of our proposed EBP of employment for small areas and compare them using design-based simulation.

Empirical Bayes Estimation of Radio Ratings for Small Geographic Areas

Waldron William R. (University of Maryland)
William.Waldron@nielsen.com

Nielsen Media Research surveys households in the US media markets to ascertain audience information on terrestrial radio stations. Radio ratings are compiled for the average and cumulative number of persons in the population that listened to a station during a specified time period. Sample sizes are maintained to guarantee precision ratings estimates at the overall market level. Applications arise where Nielsen would like to produce ratings estimates for geographies that lie below the market level. These small geographic areas include counties, zip codes, and census tracts that may be used as the building blocks for Political Districts, Retail Zones, and other areas of interest. At such granular levels, there is insufficient sample to produce satisfactory direct estimates. Using a nested error regression model, we combine survey data with census demographic data to produce empirical Bayes ratings estimates for geographical areas with at least one respondent from the sample survey and regression synthetic estimates for areas for which no information is available from the survey data. Parametric bootstrap is used to the measure uncertainty of empirical Bayes estimates.
Forecasts of means, rates or probabilities for small areas defined by social, demographic and geographic variables are an important planning tool. Standard forecasting methods do not, however, perform well with the noisy data that are typical of small areas. We develop a general class of Bayesian small area forecasting models for utilizing area-level cross-classified series. These models allow us to derive estimates and forecasts for small areas in a coherent way, to properly measure uncertainty, and to address some common practical challenges, such as changes in geographic boundary. In order to avoid over-extrapolation of historical series, we also tweak the commonly used benchmarking technique to incorporate expert judgement into the forecasting framework. We illustrate our approach with several real examples of forecasting demographic rates.

Key Words: Time Series, Bayesian Hierarchical Model, Area Level model, Expert Judgement, Benchmarking, Demography
8. THE SPONSORS & ORGANIZERS

- **Universidad Católica de Chile**
  Instituto de Sociología (ISUC), Universidad Católica de Chile
  http://sociologia.uc.cl/
  Departamento de Estadística, Universidad Católica de Chile
  http://www.mat.uc.cl/articulo/departamento-de-estadistica.html
  Departamento de Salud Pública, Universidad Católica de Chile
  http://medicina.uc.cl/salud-publica/
  Centro UC de Encuestas y Estudios Longitudinales, Universidad Católica de Chile
  http://www.encuestas.uc.cl/
  Vicerrectoría de Investigación, Universidad Católica de Chile
  http://investigacion.uc.cl/

- **Statistical Associations**
  International Statistical Institute (ISI)
  http://www.isi-web.org/
  International Association of Survey Statisticians (IASS)
  http://isi-iass.org/home/
  Sociedad Chilena de Estadística (SOCHE)
  http://www.soche.cl/

- **Chilean Government**
  Instituto Nacional de Estadísticas de Chile (INE)
  http://www.ine.cl/
  Ministerio de Desarrollo Social de Chile (MDS)
  http://www.ministeriodesarrollosocial.gob.cl/
  Comisión Nacional de Investigación Científica y Tecnológica (CONICYT)
  http://www.conicyt.cl/
Brief Summary

The Institute of Sociology (ISUC) was created in 1959 as the unit of research and teaching in the Faculty of Social Sciences of the Pontifical Catholic University of Chile. At the undergraduate level, a degree in Sociology is offered, and since 2013, a degree in Anthropology. At the postgraduate level, the Institute has the Master in Sociology, a full-time (daytime) program focused on training skills in basic research in the discipline; and a Doctoral program in Sociology, dedicated to the formation of advanced skills in research and teaching in sociology. In Continuing Education, the Institute has two programs: the Diploma in Data Processing and Analysis of Social Data and the Diploma on Advanced Techniques on Analysis of Social Data.

In the last years, the Institute has established itself as a leader in the discipline in the country, developing initiatives to strengthen its institutional identity, promoting teaching and generating knowledge through research and publications of its academic staff. The Institute has three Research Centers: the Department of Social Studies (DESUC), conducts studies and surveys on topics of public and private interest; the Center of Longitudinal Surveys and Studies (CEEL), which conducts complex surveys, especially focused on public policy matters, and the Center for the Study of Aging, aimed at the promotion of the research in this topic.

Breve Reseña

El Instituto de Sociología (ISUC) se creó en 1959 como la unidad de investigación y docencia en la disciplina de la Facultad de Ciencias Sociales de la Pontificia Universidad Católica de Chile. A nivel de pregrado, se imparte la carrera de Sociología y, desde 2013, se imparte la carrera de Antropología. A nivel de postgrado, el Instituto mantiene el Magíster en Sociología, un programa académico diurno y de dedicación exclusiva destinado a la formación de habilidades en investigación básica en la disciplina; y el Doctorado en Sociología, dedicado a la formación de competencias avanzadas en la investigación y docencia en sociología. En educación continua, el Instituto cuenta con dos programas: el Diplomado en Procesamiento y Análisis de Datos Sociales y el Diplomado en Técnicas avanzadas en Análisis de Datos Sociales.

En los últimos años, el Instituto se ha consolidado como un referente en la disciplina en el país, desarrollando iniciativas para fortalecer la identidad institucional, promoviendo la docencia y generando conocimientos a través de diversas investigaciones y publicaciones de su cuerpo académico. Además, el Instituto cuenta con 3 Centros de investigación aplicada: Dirección de Estudios Sociales (DESUC), realiza estudios y encuestas en temáticas de interés público y privado; el Centro de Encuestas y Estudios Longitudinales (CEEL), que realiza encuestas a gran escala, especialmente, dirigidas a las políticas públicas y, el Centro de Estudios de Vejez y Envejecimiento, dirigido a promover la investigación en la temática.
Brief Summary

The Faculty of Mathematics of the Pontificia Universidad Católica de Chile aims cultivation and development of the Mathematics and Statistics, conducting research, teaching and extension. The Faculty has two departments; the Department of Mathematics and the Department of Statistics.

At the undergraduate level a bachelor degree in Mathematics and Statistics is offered, as well as a Professional degree in Statistics. At the graduate level both Master and Doctoral degrees are offered in Mathematics and Statistics, and a master in Actuarial Science. The Ph.D. program in Mathematics at Pontificia Universidad Católica de Chile was created in 1972, and it is the oldest such program in the country. The Ph.D. program in Statistics was launched in 1998, and was the second of its kind in Latin America, where we have many by foreign students enrolled.

The Faculty also offers three non-degree seeking programs: Diplomado en Estadística Mención Métodos Estadísticos, Diplomado en Metodologías Estadísticas en Mejoramiento de Procesos, and Diplomado en Matemática para Profesores de Enseñanza Básica. The Faculty also has DATA UC – a service unit dependent from the faculty dedicated to advice and consultancy on the field of statistics.

Breve Reseña

La Facultad de Matemáticas de la Pontificia Universidad Católica de Chile tiene por objeto el cultivo y desarrollo de las Ciencias Matemáticas y Estadística, realizando investigación, impartiendo docencia y efectuando labores de extensión. La Facultad está constituida por un Departamento de Matemática y un Departamento de Estadística.


La Facultad también ofrece tres programas no conducentes a grado académico: Diplomado en Estadística Mención Métodos Estadísticos, Diplomado en Metodologías Estadísticas en Mejoramiento de Procesos, y Diplomado en Matemática para Profesores de Enseñanza Básica. La Facultad cuenta además con DATA UC - una unidad de servicios dependiente de la Facultad dedicada a la asesoría y consultoría en el área estadística.
Brief Summary

The Department of Public Health is part of the Faculty of Medicine of the Pontifical Universidad Católica de Chile. Its mission is to develop disciplines related to Public Health within the mission and the scope of the Catholic University, creating and communicating knowledge to improve health and wellbeing of the community.

Since its inception in the 60’s, the Department has developed strong research areas of local interest and relevance, achieving worldwide recognition as well, from the pioneering studies of the epidemiology of the streptococcal disease to the current Mauco project, a cohort study of chronic diseases collecting data on 10,000 inhabitants of the Maule commune in Chile’s VII region. The two earlier versions of the Chilean National Health Survey (2004, 2010), including the recently awarded 2015 version, have been designed and implemented by academics of the Department under contract by the Chilean Ministry of Health.

With a strong vocation for teaching, the Department is responsible for all Biostatistics courses offered at the Faculty of Medicine, along the courses on Public Health and Health Administration at the undergraduate level. At the graduate level, the Department offers a Master in Epidemiology and a Master in Health Administration. Starting August 2015, it will offer the first on-line Diploma in Biostatistics in Chile.

Breve Reseña

El Departamento de Salud Pública es parte de la Facultad de Medicina de la Pontificia Universidad Católica de Chile. Su misión es desarrollar las disciplinas relacionadas con la Salud Pública dentro del marco y la misión de la Universidad Católica, creando y comunicando el conocimiento que permita mejorar las condiciones de salud de la comunidad.

Desde sus inicios en los años 60 el Departamento ha desarrollado sólidas líneas de investigación de interés e importancia para el país, logrando también reconocimiento mundial, desde los estudios pioneros en la epidemiología de la enfermedad estreptocócica, hasta el actual proyecto Mauco, un estudio de cohorte de enfermedades crónicas, que recoge datos de 10,000 habitantes de la comuna de Molina en la VII Región de Chile. Las dos versiones de la Encuesta Nacional de Salud (2004, 2010) fueron diseñadas e implementadas por académicos del departamento bajo contrato con el Ministerio de Salud de Chile.

Con una fuerte vocación hacia la docencia, el Departamento dicta todos los cursos de Bioestadística de la Escuela de Medicina, además de los cursos de Salud Pública y Administración en Salud a nivel de pregrado. A nivel de postgrado se ofrecen los programas de Magíster en Epidemiología y Magíster en Administración en Salud. A partir de Agosto 2015 se ofrecerá el primer Diplomado en Bioestadística on-line de Chile.
**Brief Summary**

The Centro de Encuestas y Estudios Longitudinales (Longitudinal Studies and Surveys Center) main objective is to develop empirical research based on surveys and longitudinal data, from an interdisciplinary perspective. For that purpose, the Center incorporates technology and innovation on the design of its research, on processes, methodologies of data recollection and on analysis, allowing an empirical perspective of social observation.

The Center is formed by a group of professionals of wide trajectory, specialist in different areas of knowledge, such as Education, Economics, Psychology and Sociology. Additionally, it relies on a permanent team of international academics who collaborate regularly with national researchers and who are also consulted on specific projects.

Longitudinal studies represent the international standard for the development of empirical research, as they can capture dynamic effects associated with public policy or with exogenous changes, allowing for a more rigorous evaluation of their impact. That is the reason why Impact Evaluations are a subject that is present transversely in all research areas that the Center focuses on.

Some of the principal areas of investigation are: Education, Early Childhood, Social Security, Labor Economics, Electoral behavior, Survey Methodology, Population and Family.

**Breve Reseña**

El Centro de Encuestas y Estudios Longitudinales tiene como objetivo desarrollar investigación empírica en base a encuestas y datos de carácter longitudinal, desde una perspectiva interdisciplinaria. Para ello incorpora tecnología e innovación en el diseño de la investigación, en los procesos, en las metodologías de recolección y en los análisis, permitiendo una perspectiva empírica en la observación social.

Está conformado por un grupo de profesionales de larga trayectoria, especialistas en distintas áreas del conocimiento, tales como educación, economía, psicología y sociología. Adicionalmente, cuenta con un equipo de académicos internacionales permanentes que colaboran de manera regular con los investigadores nacionales y que también asesoran en proyectos específicos.

Los estudios longitudinales representan el estándar internacional para el desarrollo de investigación empírica puesto que permiten capturar efectos dinámicos asociados a políticas o variaciones exógenas y posibilita la evaluación rigurosa de su impacto. Por lo tanto, las evaluaciones de impacto son un tópico que atraviesa de manera transversal cada una de las áreas de investigación que el Centro aborda.

Dentro de sus principales áreas de investigación se encuentran: educación, primera infancia, seguridad social, economía laboral, comportamiento electoral, metodología de encuestas, población y familia.
Brief Summary

The Chilean Statistical Society, SOCHE, was formed in 1977. It was one of the first statistical associations in Latin America, even though the field was quite underdeveloped at that time. Two main factors behind its creation were the large number of Chilean graduates from Interamerican Center for Teaching Statistics (CIENES) and the visit to Chile of a delegation of the American Statistical Association (ASA) in 1976.

The main activity of SOCHE has been the realization of yearly national statistical meetings, the Jornadas Nacionales de Estadística (41 to date). Since 1990, some of them are held jointly with the Sociedad Argentina de Estadística every two years, either in Chile or Argentina. In 2008 the Sociedad Uruguaya de Estadística joined these meeting (named CLATSE). In 2016, the XII CLATSE will be held in Lambayeque, Peru, with the participation of the Associação Brasileira de Estatística from Brazil. Following the example of ASA, SOCHE created special sections such as Biostatistics, Bayesian, Official Statistics, Statistical Education, among the most active.

Our Society publishes a Journal, the Chilean Journal of Statistics (http://chjs.soche.cl/), who took the place of Revista de la Sociedad Chilena de Estadística, which was published between 1984 and 2000. This year, in October, our XLII Jornadas Nacionales de Estadística, will be held in the city of Concepción, organized by Universidad del BíoBío (http://jne.ubiobio.cl/).

Breve Reseña

La Sociedad Chilena de Estadística, SOCHE, se formó en 1977. Fue una de las primeras asociaciones estadísticas en América Latina, a pesar que el campo estaba muy poco desarrollado en ese tiempo. Dos de los principales factores que hubo detrás de su creación fueron, el gran número de graduados chilenos del Centro Interamericano para la Enseñanza de la Estadística (CIENES) y la visita a Chile de una delegación de la American Statistical Association (ASA) en 1976.

La principal actividad de SOCHE ha sido la realización de congresos nacionales de estadística anuales, las Jornadas Chilenas de Estadística (41 a la fecha). Desde 1990, algunos de ellos se llevan a cabo en forma conjunta con la Sociedad Argentina de Estadística, cada dos años, ya sea en Chile o Argentina. En 2008, la Sociedad Uruguaya de Estadística unió a estos congresos (llamados CLATSE). En 2016, XII CLATSE se celebrará en Lambayeque, Perú, con la participación de la Associação Brasileira de Estatística de Brasil. Siguiendo el ejemplo de ASA, SOCHE creó secciones especiales como Bioestadística, Bayesiana, Estadística Oficial, Educación Estadística, entre las más activas.

Brief Summary

The National Statistical Institute (INE) is the body in charge of production of the official statistics of the country, assuring the technical quality and providing timely, reliable, comparable and accessible information for management of decision making at a national and regional level. Moreover, the Institute has a leading role in articulate the national statistical system. Its mission is to be a technically sound and reliable entity, well known for its transparency and the quality of processes aimed to generate statistical figures and the best internationally recognized technical standards.

The INE is one of the oldest public bodies in the country. It started up officially in 1843, date of its foundation. Since then, a number of censuses, surveys and studies have been developed in order to reflect the various aspects of the national reality. Currently, the INE provides the Chilean society with more than 70 statistical products, including indicators and other statistical figures in a range of fields such as employment, prices, population, culture, citizenship security, economics and other important issues for public policy decision-making.

The National Statistical Institute is planning a Census of Population and Housing to be conducted in April 2017. For the first time in Chile, this process will be carried out under an “abbreviated format”, given the urgent need to obtain relevant demographic data for the decision making in Chile. For this census operation, the INE will be summoning all Chilean citizens that are interested in being part of this great citizenship and republican event, from youngsters to elderly people. In this way, we expect to successfully achieve this great challenge for all the Chileans. For this very last reason, our motto for the 2017 Census is “WE ALL COUNT”.

Breve Reseña

El Instituto Nacional de Estadísticas de Chile (INE) es el organismo encargo de generar las estadísticas oficiales del país, garantizando la calidad técnica y la entrega de información confiable, oportuna, pertinente, comparable y de fácil acceso para la toma de decisiones a nivel nacional y territorial, liderando y articulando el Sistema Estadístico Nacional. Su misión es ser un organismo técnico sólido y confiable, reconocido por la transparencia y calidad de sus procesos de generación de estadísticas y su estándar técnico de clase mundial.

Es uno de los organismos públicos con mayor trayectoria en el país. Desde 1843, año de su creación oficial, cuenta en su quehacer numerosos censos, encuestas y estudios de la realidad nacional. En la actualidad, el INE entrega al país más de 70 indicadores de calidad en una diversidad temática que incluye empleo, precios, población, cultura, seguridad ciudadana, economía y otros aspectos relevantes para la toma de decisiones en políticas públicas.

El Instituto Nacional de Estadísticas (INE), realizará un censo de población y vivienda en abril de 2017, proceso que por primera vez en el país se desarrollará en formato “abreviado”, dada la necesidad urgente de obtener información demográfica relevante para la toma de decisiones en Chile. Para este censo, nuestra institución está convocando a todos los chilenos que quieran ser parte de este gran evento cívico y republicano, desde los jóvenes hasta los adultos mayores, para que juntos logremos el éxito en este inmenso desafío que nos pertenece a todos. Por eso, para el Censo 2017, “TODOS CONTAMOS”.
Brief Summary

The origins of the Ministry of Social Development go back to the decade of the 60, under the government of President Eduardo Frei Montalva (1964-1970), when the need for a technical organization dedicated to the national development and planning arises. The National Planning Office (ODEPLAN) is then created.

In 1990, ODEPLAN becomes the Ministry of Planning and Cooperation (MIDEPLAN), with the mission to collaborate with the President in the design and implementation of policies, plans and development of national programs, setting goals and evaluating public investment projects financed by the State, and coordinating different public sector initiatives aimed at eradicating poverty. In the mid 90 MIDEPLAN assumes the responsibility in the design and implementation of a National Program for Overcoming Poverty. This effort to optimize public policies to benefit the neediest population of the country led to the creation of the Chile Solidario System, which provides integral support to persons and families living under extreme poverty.

In October 2011, MIDEPLAN becomes the Ministry of Social Development (MDS), extending its powers and being positioned as the coordinating entity of all the social policies in Chile. MDS’s mission is to contribute to the design and implementation of policies, plans and programs on social development, especially those aimed at eradicating poverty and providing social protection to vulnerable persons or groups, promoting mobility and social integration.

Breve Reseña

Los orígenes del Ministerio de Desarrollo Social se remontan a la década de los 60’, en el gobierno del Presidente Eduardo Frei Montalva (1964-1970), cuando surge la necesidad de contar con un organismo técnico dedicado a la planificación del desarrollo a nivel nacional. En ese momento nace la Oficina de Planificación Nacional (ODEPLAN).

En 1990, ODEPLAN se transforma en el Ministerio de Planificación y Cooperación (MIDEPLAN), a quien se le asigna la misión de colaborar con el Presidente de la República en el diseño y aplicación de políticas, planes y programas de desarrollo nacional, proponer las metas de inversión pública y evaluar los proyectos de inversión financiados por el Estado, y coordinar las diferentes iniciativas del sector público dirigidas a erradicar la pobreza. Hacia mediados de los 90’ MIDEPLAN asume la responsabilidad del diseño e implementación del Programa Nacional de Superación de la Pobreza, cuyos ejes centrales eran la integralidad, la descentralización y la participación. Este esfuerzo por optimizar las políticas públicas en beneficio de los más necesitados del país motivó la creación del Sistema Chile Solidario, que entrega un apoyo integral a las personas y familias de extrema pobreza.

En octubre de 2011, MIDEPLAN pasa a ser Ministerio de Desarrollo Social (MDS), ampliando sus facultades y posicionándose como el ente coordinador de todas las políticas sociales de Chile. La misión del MDS es contribuir en el diseño y aplicación de políticas, planes y programas en materia de desarrollo social, especialmente aquellas destinadas a erradicar la pobreza y brindar protección social a las personas o grupos vulnerables, promoviendo la movilidad e integración social.