Statistical Methods in Medical Research 2011; **20**: 299–330

A review of multivariate longitudinal data analysis

S Bandyopadhyay Indian Institute of Public Health, Hyderabad, India, **B Ganguli** Department of Statistics, University of Calcutta, India and **A Chatterjee** Department of Statistics, The University of Burdwan, India

Repeated observation of multiple outcomes is common in biomedical and public health research. Such experiments result in multivariate longitudinal data, which are unique in the sense that they allow the researcher to study the joint evolution of these outcomes over time. Special methods are required to analyse such data because repeated observations on any given response are likely to be correlated over time while multiple responses measured at a given time point will also be correlated. We review three approaches for analysing such data in the light of the associated theory, applications and software. The first method consists of the application of univariate longitudinal tools to a single summary outcome. The second method aims at estimating regression coefficients without explicitly modelling the underlying covariance structure of the data. The third method combines all the outcomes into a single joint multivariate model. We also introduce a multivariate longitudinal dataset and use it to illustrate some of the techniques discussed in the article.

1 Introduction

Experiments in medical and social science research are often complex and characterised by multiple observations on several outcomes measured repeatedly over time. Such experiments are unique in the sense that they allow the researcher to study the joint evolution of multiple outcomes over time. Special methods are required to analyse the resulting data as repeated observations on any given response are likely to be correlated over time while multiple responses measured at a given time point will also be correlated. If the issue is to understand the relationship between the responses, then one must account for the correlation between various outcomes measured at the same or at different time points. Statistical analysis becomes further complicated when outcomes are a combination of continuous and discrete responses.

In 2007, *Statistical Methods in Medical Research* published a special issue on methods for analysis of multivariate longitudinal data. The five articles in that issue primarily focused on modelling a longitudinal response with missing observations or joint modelling of longitudinal responses.[1] In Section 6 we have elaborated on the difference between the approaches discusses in the current article and those discussed in the special issue.

Address for correspondence: S Bandyopadhyay, Indian Institute of Public Health, C/O Indian Institute of Health and Family Welfare, Vengal Rao Nagar, Hyderabad-500038, India. E-mail: bansouvik@gmail.com

10.1177/0962280209340191

Soc Indic Res (2010) 99:285–299
DOI 10.1007/s11205-010-9580-0

Income Affluence in Poland

Michal Brzezinski

Accepted: 17 January 2010 / Published online: 29 January 2010

Abstract This paper examines the evolution of income affluence (richness) in Poland during 1998–2007. Using household survey data, the paper estimates several statistical indices of income affluence including income share of the top percentiles, population share of individuals receiving incomes higher than the richness line, and measures that take into account both the extent and the intensity of affluence. Results show that over the period under study there was a statistically significant and socio-economically sizable rise in income affluence by between 9 and 50%, depending on the index used. The overall income distribution in the period has shifted in favour of the rich as relative poverty and relative size and income share of the middle class have declined.

Keywords Affluence · Richness · Income distribution · Poland

1 Introduction

There is a voluminous theoretical and empirical literature on income distribution dealing with incomes of the poor and overall inequality of incomes in the society. Until very recently, a different distributive problem of measuring incomes at the top of the distribution (incomes of the rich) was rarely analysed. However, in a few recent years social scientists have been expressing a growing interest in the theoretical construction of affluence lines and indicators (Medeiros 2006; Peichl et al. 2006, 2008) and in the empirical measurement of top incomes (e.g., Piketty 2001; Atkinson and Piketty 2007).[1]

[1] Long run high-quality estimates of the income shares of the top $p\%$ of the population have been produced for at least fourteen developed countries and at least four developing economies (Atkinson and Piketty 2007; Leigh 2009). An almost exhaustive reference list of recent papers devoted to the empirical measurement of top incomes can be found in Leigh (2009). These papers usually rely on incomes reported for tax purposes, but the number of studies based on household survey data grows as well.

M. Brzezinski (✉)
University of Warsaw, Warsaw, Poland
e-mail: mbrzezinski@wne.uw.edu.pl

Changes in perceived effect of practice guidelines among primary care doctors

Lee Cheng MD MSc,[1] Linda Z. Nieman PhD[2] and James L. Becton MD[3]

[1]Assistant Professor, The Department of Family and Community Medicine, The University of Texas Health Science Center at Houston, Houston, TX, USA

[2]Professor and Vice Chair for Educational Affairs and Director, Joint Primary Care Fellowship, The Department of Family and Community Medicine, The University of Texas Health Science Center at Houston, Houston, TX, USA

[3]Pediatric Fellow, Joint Primary Care Fellowship, The University of Texas Health Science Center at Houston, Houston, TX, USA

Keywords
practice guidelines, primary care doctors

Correspondence
Lee Cheng
The Department of Family and Community Medicine
6431 Fannin Street
Suite JJL308
Houston
TX 77030
USA
E-mail: lee.cheng@uth.tmc.edu

Accepted for publication: 8 February 2006

Abstract

Rationale, aims and objectives Evidence suggests that when doctors use systematically developed clinical practice guidelines they have the potential to improve the safety, quality and value of health care. The purpose of this study was to evaluate recent changes in the perceptions of practice guidelines among US primary care doctors.

Methods Data were collected from the Community Tracking Survey 1996–97 and 2000–01. All results were weighted and adjusted to reflect the complex survey design.

Results Over the 5 years, the proportion of primary care doctors who said that practice guidelines had at least a moderate effect on their practice of medicine increased from 45.8% to 60.7%. This increase was nearly equal among primary care doctors of family medicine, internal medicine and paediatrics. In the 2001 survey, a higher perceived effect of practice guidelines was described by female doctors (OR = 1.39, 95% CI 1.19–1.63) and doctors who were practising in a large model group (OR = 1.73; 95% CI 1.04–2.89). Doctors who graduated from medical school within 10 years of the survey were more likely to report that practice guidelines had a positive effect on their practice of medicine than doctors who graduated 10 or more years before the survey.

Conclusion The perceived effect of practice guidelines on primary care doctors increased over time. Improved dissemination of guidelines and curriculum changes may have led recent primary care graduates to view practice guidelines as more important.

Introduction

Although evidence suggests that the use of systematically developed evidence-based clinical practice guidelines has the potential to improve the quality of patient care as well as the satisfaction of patients [1–3], doctor compliance with these guidelines remains low [4–7]. In 1997, the Community Tracking Study (CTS), a national doctor survey, showed that 46% of primary care doctors perceived that practice guidelines had a moderate to very large effect on their medical practices [8].

The successful adoption of practice guidelines has been shown to positively affect doctor awareness, self-efficacy, outcome expectancy, practice habits, and other patient- and system-related factors [4,5]. Assessing doctors' attitudes towards guidelines is the first critical step towards improving practice guidelines dissemination and implementation as well as refining their role in evaluating health care quality [9–11]. In addition, health care delivery systems have experienced unparalleled economic and competitive challenges. They have adopted quality improvement principles to improve their economic situation, with a particular focus on patient-centred care and consumer satisfaction. Patient satisfaction, along with outcomes and costs, has become an important measure of health system performance [12,13,14].

Many evidence-based guidelines have been developed by medical schools, specialty medical groups, government agencies, and health care companies. These guidelines range from how to treat common ailments such as asthma and hypertension to how to perform surgeries and how to tackle serious diseases such as cancer [3–5]. Most evidence-based guidelines for clinical practice have predominantly focused on inpatient care, chronic disease management, and preventive medicine [3]. We believed that the primary care doctors' office was an ideal setting for the adoption of practice guidelines for patient care. However, change in the perceived effect of practice guidelines among primary care doctors

Empir Econ (2012) 43:499–524
DOI 10.1007/s00181-011-0499-3

On statistical inference for inequality measures calculated from complex survey data

Judith A. Clarke · Nilanjana Roy

Received: 12 February 2011 / Accepted: 23 May 2011 / Published online: 13 August 2011

Abstract We examine inference for Generalized Entropy and Atkinson inequality measures with complex survey data, using Wald statistics with variance–covariance matrices estimated from a linearization approximation method. Testing the equality of two or more inequality measures, including sub-group decomposition indices and group shares, are covered. We illustrate with Indian data from three surveys, examining pre-school children's height, an anthropometric measure that can indicate long-term malnutrition. Sampling involved an urban/rural stratification with clustering before selection of households. We compare the linearization complex survey outcomes with those from an incorrect independently and identically distributed (iid) assumption and a bootstrap that accounts for the survey design. For our samples, the results from the easy to implement linearization method and the more computationally burdensome bootstrap are typically quite similar. This finding is of interest to applied researchers, as bootstrapping is currently the method that is most commonly used for undertaking statistical inference in this literature.

Keywords Atkinson · Complex survey · Decomposition · Generalized entropy · Inequality · Linearization

JEL Classification C12 · C42 · D31

J. A. Clarke · N. Roy (✉)
Department of Economics, University of Victoria, P.O. Box 1700, STN CSC, Victoria, BC V8W 2Y2, Canada
e-mail: nroy@uvic.ca

J. A. Clarke
e-mail: jaclarke@uvic.ca

Stat. Meth. & Appl. (2006) 15:243–265
DOI 10.1007/s10260-006-0021-9

ORIGINAL ARTICLE

The new strategy for the concise presentation of sampling errors in the Italian Structural Business Statistics Survey

Piero Demetrio Falorsi · Salvatore Filiberti · Antonio Pavone

Accepted: 31 May 2006 / Published online: 1 August 2006

Abstract Reporting sampling errors of survey estimates is a problem that is commonly addressed when compiling a survey report. Because of the vast number of study variables or population characteristics and of interest domains in a survey, it is almost impossible to calculate and to publish the standard errors for each statistic. A way of overcoming such problem would be to estimate indirectly the sampling errors by using *generalized variance functions*, which define a statistical relationship between the sampling errors and the corresponding estimates. One of the problems with this approach is that the model specification has to be consistent with a roughly constant design effect. If the design effects vary greatly across estimates, as in the case of the Business Surveys, the prediction model is not correctly specified and the least-square estimation is biased. In this paper, we show an extension of the *generalized variance functions*, which address the above problems, which could be used in contexts similar to those encountered in Business Surveys. The proposed method has been applied to the Italian Structural Business Statistics Survey case.

Keywords Generalized variance functions · Design effect · Business survey

P. D. Falorsi (✉) · A. Pavone
Servizio progettazione e supporto metodologico nei processi di produzione statistica, Istat,
via Magenta, 2, Roma, Italy
e-mail: falorsi@istat.it

S. Filiberti
Servizio delle Statistiche strutturali sulle imprese industriali e dei servizi, Istat,
via Tuscolana, 1788, Roma, Italy
e-mail: filibert@istat.it

A. Pavone
e-mail: pavone@istat.it

Soc Indic Res (2007) 84:97–116
DOI 10.1007/s11205-006-9077-z

ORIGINAL PAPER

Fathers' participation in the domestic activities of everyday life

Maria Clelia Romano · Dario Bruzzese

Received: 8 July 2006 / Accepted: 21 November 2006 / Published online: 16 January 2007

Abstract In this paper, the data from the multi-purpose survey on household ''Time Use'' conducted by Istat (the Italian National Statistical Institute) in 2002–2003 and the data from this same survey conducted in 1988–1989 will be analysed with the purpose of describing the fathers' daily participation in the domestic activities and of highlighting the changes that have taken place during the 14 years elapsed between the two survey editions. The analysis will be carried out using standard time-use data analysis' tool, time budget tables and by applying a multivariate regression model with the objective of separating the relative contribution of the behavioural and structural factors to explain the variation observed.

Keywords Families and work · Time use

1 Introduction

In past years, the international literature has increasingly focused on the study of paternity and, in particular, of the father's role in child care' activities[1] (Eggebeen & Knoester, 2001; Lamb, 1987; Parke, 1995). Various reasons have spurred such growing attention. First, the higher number of separations and divorces, and thus the consequent growth of single-parent families, has led to a growing interest for the

[1] Numerous studies conducted in the United States show how a *new fatherhood* is currently emerging, characterised by an emotionally more intense father–son relationship, by a greater involvement of fathers in their children's life, by more equal prospects as regards the gender roles and by a more frequent involvement in taking care of the children. See, among others: Lamb (1987), Parke (1995), Eggebeen and Knoester (2001).

M. C. Romano · D. Bruzzese (✉)
ISTAT Italian National Statistical Institute, Via Ravà 150, 00142 Rome, Italy
e-mail: bruzzese@istat.it

Metrika (2007) 65: 29–42
DOI 10.1007/s00184-006-0057-8

ORIGINAL ARTICLE

Horst Stenger · Siegfried Gabler

Optimal strategies in 2-stage sampling

Received: 9 September 2005 / Published online: 26 April 2006

Abstract In many survey situations simple random sampling of units and estimation of a total of interest by the expansion estimator are attractive methods, at least at first sight. Considering cost aspects suggests rather to use multiple stage sampling which, in general, is cheaper, but less effective. The design effect is an adequate criterion of the decrease of efficiency. We discuss this criterion for clusters (primary units) of equal size and derive exact conditions for a decrease of efficiency. The equality condition for cluster sizes seems not to be very restrictive, because in many cases one will be interested in clusters of approximately the same size, or, if sizes differ essentially, the clusters are partitioned into strata according to their sizes and the procedures for different strata are independent, each dealing with clusters of equal size or nearly so. In the context considered the use of the Horvitz–Thompson estimator is quite general. We examine a class of estimators with the Horvitz–Thompson estimator and a straight forward modification of it as special elements. As for the design effect all elements of the class are very similar, as for other aspects such as admissibility there are remarkable differences.

Keywords 2-stage designs · Design effect · Horvitz-Thompson estimator · Risk points · Self-weighting designs

H. Stenger (✉)
L7, 3-5, University of Mannheim,
68131 Mannheim, Germany
E-mail: stenger@rumms.uni-mannheim.de

S. Gabler
B2, 1, Centre for Survey Research and Methodology,
Postfach 12 21 55, 68072 Mannheim, Germany
E-mail: gabler@zuma-mannheim.de

Population Health Metrics

BioMed Central

Research

Open Access

Assessment of the health of Americans: the average health-related quality of life and its inequality across individuals and groups

Yukiko Asada*

Address: Department of Community Health and Epidemiology, Faculty of Medicine, Dalhousie, University, 5790 University Avenue, Halifax, Nova Scotia, B3H 1V7, Canada

Email: Yukiko Asada* - yukiko.asada@dal.ca

* Corresponding author

Published: 13 July 2005

Population Health Metrics 2005, **3**:7 doi:10.1186/1478-7954-3-7

Received: 03 January 2005

Accepted: 13 July 2005

This article is available from: http://www.pophealthmetrics.com/content/3/1/7

Abstract

Background: The assessment of population health has traditionally relied on the population's average health measured by mortality related indicators. Researchers have increasingly recognized the importance of including information on health inequality and health-related quality of life (HRQL) in the assessment of population health. The objective of this study is to assess the health of Americans in the 1990s by describing the average HRQL and its inequality across individuals and groups.

Methods: This study uses the 1990 and 1995 National Health Interview Survey from the United States. The measure of HRQL is the Health and Activity Limitation Index (HALex). The measure of health inequality across individuals is the Gini coefficient. This study provides confidence intervals (CI) for the Gini coefficient by a bootstrap method. To describe health inequality by group, this study decomposes the overall Gini coefficient into the between-group, within-group, and overlap Gini coefficient using race (White, Black, and other) as an example. This study looks at how much contribution the overlap Gini coefficient makes to the overall Gini coefficient, in addition to the absolute mean differences between groups.

Results: The average HALex was the same in 1990 (0.87, 95% CI: 0.87, 0.88) and 1995 (0.87, 95% CI: 0.86, 0.87). The Gini coefficient for the HALex distribution across individuals was greater in 1995 (0.097, 95% CI: 0.096, 0.099) than 1990 (0.092, 95% CI: 0.091, 0.094). Differences in the average HALex between all racial groups were the same in 1995 as 1990. The contribution of the overlap to the overall Gini coefficient was greater in 1995 than in 1990 by 2.4%. In both years, inequality between racial groups accounted only for 4–5% of overall inequality.

Conclusion: The average HRQL of Americans was the same in 1990 and 1995, but inequality in HRQL across individuals was greater in 1995 than 1990. Inequality in HRQL by race was smaller in 1995 than 1990 because race had smaller effect on the way health was distributed in 1995 than 1990. Analysis of the average HRQL and its inequality provides information on the health of a population invisible in the traditional analysis of population health.

Qual Quant (2012) 46:545–552
DOI 10.1007/s11135-010-9385-x

Estimating latent class model parameters for filter questions with skip patterns

Ting Hsiang Lin

Published online: 21 November 2010

Abstract Filter questions with skip patterns have been widely used in survey research, and latent class models (LCM) are often used to analyze this type of categorical data. The LCM parameters are usually estimated by means of an EM (expectation maximization) algorithm. When the pattern is present, the non-response of the skip pattern cannot be treated as random missingness. We thus propose a modified algorithm to estimate the latent class parameters when non-response is present, and the approach is attractive for two reasons. First, the latent class model with the algorithm is very flexible in the sense that it can model the association of variables with the skip patterns under study. Secondly, the algorithm can be easily implemented using any computer language. An empirical example is used to demonstrate the usefulness of the algorithm. The algorithm may also be flexibly generalized to more complex surveys, for example, polytomous responses.

Keywords Latent class model · Skip patterns · EM algorithm · Non-response

1 Introduction

Many constructs of interest cannot be observed directly. Examples include health attitudes, medical knowledge, and disease perceived risk, among others. Such constructs can only be measured indirectly by means of some observable indicators, like a set of items in questionnaires or psychological scales. The Short Form 36-item Health Survey (SF-36) (Ware et al. 1993), for example, has been used to measure the functional status of the respondents, and the National Health Interview Survey (NHIS 2004) is intended to provide estimates for health conditions, health behaviors, and the usage of medical resources. Variables that can be observed directly are manifest variables, while, on the contrary, variables that cannot be observed directly are latent variables. We expect the latent variables to explain the relationships among the manifest variables as much as possible.

T. H. Lin (✉)
Department of Statistics, National Taipei University, Taipei, Taiwan, ROC
e-mail: tinghlin@mail.ntpu.edu.tw

Soc Indic Res
DOI 10.1007/s11205-011-9918-2

How much Confidence can we have in EU-SILC? Complex Sample Designs and the Standard Error of the Europe 2020 Poverty Indicators

Tim Goedemé

Accepted: 7 August 2011

Abstract If estimates are based on samples, they should be accompanied by appropriate standard errors and confidence intervals. This is true for scientific research in general, and is even more important if estimates are used to inform and evaluate policy measures such as those aimed at attaining the Europe 2020 poverty reduction target. In this article I pay explicit attention to the calculation of standard errors and confidence intervals, with an application to the European Union Statistics on Income and Living Conditions (EU-SILC). The estimation of accurate standard errors requires among others good documentation and proper sample design variables in the dataset. However, this information is not always available. Therefore, I complement the existing documentation on the sample design of EU-SILC and test the effect on estimated standard errors of various simplifying assumptions with regard to the sample design. It is shown that accounting for clustering within households is of paramount importance. Although this results in many cases in a good approximation of the standard error, taking as much as possible account of the entire sample design generally leads to more accurate estimates, even if sample design variables are partially lacking. The effect is illustrated for the official Europe 2020 indicators of poverty and social exclusion and for all European countries included in the EU-SILC 2008 dataset. The findings are not only relevant for EU-SILC users, but also for users of other surveys on income and living conditions which lack accurate sample design variables.

Keywords Europe 2020 poverty reduction target · Complex sample design · Incomplete sample design variables · Standard error · Confidence interval · EU-SILC · Clustering within households · Variance estimation

T. Goedemé (✉)
Herman Deleeck Centre for Social Policy, University of Antwerp, St. Jacobstraat 2 (M479), 2000 Antwerp, Belgium
e-mail: tim.goedeme@ua.ac.be
URL: http://www.ua.ac.be/tim.goedeme; http://www.centreforsocialpolicy.eu

T. Goedemé
Research Foundation—Flanders, Egmontstraat 5, 1000 Brussel, Belgium

Published online: 19 August 2011

Dose-Response (Prepress)
Formerly Nonlinearity in Biology, Toxicology, and Medicine

ISSN: 1559-3258
DOI: 10.2203/dose-response.11-038.Cox

International Dose-Response Society
www.Dose-Response.org

LOW-DOSE NONLINEAR EFFECTS OF SMOKING ON CORONARY HEART DISEASE RISK

Louis Anthony (Tony) Cox, Jr. □ Cox Associates

□ Some recent discussions of adverse human health effects of active and passive smoking have suggested that low levels of exposure are disproportionately dangerous, so that "The effects of even brief (minutes to hours) passive smoking are often nearly as large (averaging 80% to 90%) as chronic active smoking" (Barnoya and Glantz, 2005). Recent epidemiological evidence (Teo *et al.*, 2006) suggests a more linear relation. This paper reexamines the empirical relation between self-reported low levels of active smoking and risk of coronary heart disease (CHD) in public-domain data from the National Health and Nutrition Examination Survey (NHANES). Consistent with biological evidence on J-shaped and U-shaped relations between smoking-associated risk factors and CHD risks, we find that low levels of active smoking do not appear to be associated with increased CHD risk. Several methodological challenges in epidemiology may explain how model-derived estimates can predict low-dose linear or concave dose-response estimates, even if the empirical (i.e., data-based) relation does not show a clear increased risk at the lowest doses.

Keywords: Coronary Heart Disease (CHD), hormesis, U-shaped, J-shaped, empirical dose-response model, confounding, modeling bias, classification tree analysis

INTRODUCTION: DOES HORMESIS FAIL FOR SMOKING AND CORONARY HEART DISEASE?

An emerging working hypothesis for some toxicologists and risk assessors is that many – perhaps most – biological dose-response relations exhibit J-shaped or U-shaped regions at low doses. That is, probability of harm (or, more generally, of exposure-related departures of variables from their "normal" levels) decreases with increasing dose at sufficiently small exposure levels, even if it increases with increasing doses at higher exposure levels. When this pattern holds, responses to low levels of exposures cannot necessarily be extrapolated from observed dose-response relations at higher doses.

Although considerable empirical support has been advanced in support of this "hormesis" hypothesis (Calabrese and Baldwin, 2001), the universality of its application is still being assessed. The shape of dose-response functions for complex mixtures, such as diesel exhaust or cigarette smoke, can potentially be especially valuable in either supporting the hormesis hypothesis or in understanding how it breaks down.

Address correspondence to Dr. Tony Cox, Cox Associates, 503 Franklin Street, Denver, Colorado, 80218; Email: tcoxdenver@aol.com

Child Ind Res (2008) 1:350–371
DOI 10.1007/s12187-008-9014-6

Economic Resources, Relative Socioeconomic Position and Social Relationships: Correlates of the Happiness of Young Canadian Teens

Peter Burton · Shelley Phipps

Accepted: 26 March 2008 / Published online: 29 April 2008

Abstract This paper uses a large, nationally representative microdata survey to conduct a multivariate analysis of the correlates of self-assessed happiness for Canadian 12 to 15 year olds living in two-parent families. We ask whether the same factors matter for the happiness of young teens as for adults. And, we ask whether the correlates of being at the bottom of the young teen happiness distribution are the same as the correlates of being at the top. Results suggest that the level of family income correlates with the probability of being at the bottom of the young teen happiness distribution but not with the probability of being at the top. Relative socioeconomic position and peer social relationships, on the other hand, correlate with being at the top but not at the bottom of the young teen happiness distribution. Relationships involving significant adults (parents, teachers) are the most important correlates of young teen happiness.

Keywords Happiness · Subjective well-being · Children · Teens · Relative income · Socioeconomic status · Social relationships

1 Introduction

Much recent work by economists has studied the correlates of self-assessed happiness for adults (see Frey and Stutzer 2002 or Layard 2005 for overviews); children and youth have received considerably less attention in this literature. Using a large nationally representative survey (the Statistics Canada National Longitudinal Survey of

P. Burton
Department of Economics, Dalhousie University, Halifax, NS, Canada B3H 3J5
e-mail: Peter.Burton@dal.ca

S. Phipps (✉)
Canadian Institute for Advanced Research and Department of Economics,
Dalhousie University, Halifax, NS, Canada B3H 3J5
e-mail: Shelley.Phipps@dal.ca

Soc Indic Res (2010) 96:313–337
DOI 10.1007/s11205-009-9480-3

The Social Patterning of Work-Related Insecurity and its Health Consequences

Heather Scott-Marshall

Accepted: 18 April 2009 / Published online: 8 May 2009

Abstract This study examines the association between work-related insecurity and health, with a focus on how this relationship is moderated by social location (gender, age and race). Drawing on longitudinal data from a Canadian labour market survey (1999–2004) the findings show that certain groups have a higher prevalence of exposure to certain types of work-related insecurity including (among others) low earnings, poor job mobility and the absence of union protection. Results from regression analyses indicate that the negative health impact of work-related insecurity is also unevenly distributed across different social locations. In some cases, older age and visible minority status significantly elevated the health risk posed by work-related insecurity. The implications of these findings are discussed in terms of major shifts in the demographic composition of the labour market due to workforce ageing and the increased participation of women and visible minorities.

Keywords Work insecurity · Work organization · Social inequality · Employment contract · Health

1 Introduction

Decades of structural change in the economies of industrialised countries have given rise to fundamental changes in labour markets, work systems, firm structures and hence, individual work-related experiences. The forces of globalisation have restructured employment relations such that workers are expected to bear more of the risks of doing business. A growing proportion of workers lack job security, sufficient earnings, income security benefits, and opportunities for job and career advancement (Burke and Shields 1999; Grimshaw et al. 2002; Osterman et al. 2002; Vosko 2006). Research into the health effects of exposure to several structural changes to work indicates that workers are at risk of

H. Scott-Marshall (✉)
Institute for Work & Health, 481 University Avenue, Suite 800, Toronto, ON M5G 2E9, Canada
e-mail: hscott-marshall@iwh.on.ca

H. Scott-Marshall
Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

AStA Adv Stat Anal (2012) 96:69–98
DOI 10.1007/s10182-011-0159-0

ORIGINAL PAPER

Inference on finite population categorical response: nonparametric regression-based predictive approach

Sumanta Adhya · Tathagata Banerjee · Gaurangadeb Chattopadhyay

Received: 24 September 2009 / Accepted: 2 May 2011 / Published online: 27 May 2011

Abstract Suppose that a finite population consists of N distinct units. Associated with the ith unit is a polychotomous response vector, d_i, and a vector of auxiliary variable x_i. The values x_i's are known for the entire population but d_i's are known only for the units selected in the sample. The problem is to estimate the finite population proportion vector P. One of the fundamental questions in finite population sampling is how to make use of the complete auxiliary information effectively at the estimation stage. In this article a predictive estimator is proposed which incorporates the auxiliary information at the estimation stage by invoking a superpopulation model. However, the use of such estimators is often criticized since the working superpopulation model may not be correct. To protect the predictive estimator from the possible model failure, a nonparametric regression model is considered in the superpopulation. The asymptotic properties of the proposed estimator are derived and also a bootstrap-based hybrid re-sampling method for estimating the variance of the proposed estimator is developed. Results of a simulation study are reported on the performances of the predictive estimator and its re-sampling-based variance estimator from the model-based viewpoint. Finally, a data survey related to the opinions of 686 individuals on the cause of addiction is used for an empirical study to investigate the performance of the nonparametric predictive estimator from the design-based viewpoint.

Keywords Predictive approach · Random coefficients splines model · Laplace approximation · EM algorithm

S. Adhya
Department of Statistics, West Bengal State University, North 24 Parganas, Barasat 700126, India

T. Banerjee (✉)
Production and Quantitative Methods, Indian Institute of Management, Ahmedabad, Vastrapura, Ahmedabad 380 015, India
e-mail: tathagata@iimahd.ernet.in

G. Chattopadhyay
Department of Statistics, University of Calcutta, 35 B.C. Road, Kolkata 700 019, India

PSYCHOMETRIKA—VOL. 72, NO. 2, 123–140
JUNE 2007
DOI: 10.1007/s11336-006-1453-8

MULTILEVEL AND LATENT VARIABLE MODELING WITH COMPOSITE LINKS AND EXPLODED LIKELIHOODS

SOPHIA RABE-HESKETH

UNIVERSITY OF CALIFORNIA AT BERKELEY AND UNIVERSITY OF LONDON

ANDERS SKRONDAL

LONDON SCHOOL OF ECONOMICS AND NORWEGIAN INSTITUTE OF PUBLIC HEALTH

Composite links and exploded likelihoods are powerful yet simple tools for specifying a wide range of latent variable models. Applications considered include survival or duration models, models for rankings, small area estimation with census information, models for ordinal responses, item response models with guessing, randomized response models, unfolding models, latent class models with random effects, multilevel latent class models, models with log-normal latent variables, and zero-inflated Poisson models with random effects. Some of the ideas are illustrated by estimating an unfolding model for attitudes to female work participation.

Key words: composite link, exploded likelihood, unfolding, multilevel model, generalized linear mixed model, latent variable model, item response model, factor model, frailty, zero-inflated Poisson model, gllamm.

Introduction

Latent variable models are becoming increasingly general. An important advance is the accommodation of a wide range of response processes using a generalized linear model formulation (e.g., Bartholomew & Knott, 1999; Skrondal & Rabe-Hesketh, 2004b). Other recent developments include combining continuous and discrete latent variables (e.g., Muthén, 2002; Vermunt, 2003), allowing for interactions and nonlinear effects of latent variables (e.g., Klein & Moosbrugger, 2000; Lee & Song, 2004) and including latent variables varying at different levels (e.g., Goldstein & McDonald, 1988; Muthén, 1989; Fox & Glas, 2001; Vermunt, 2003; Rabe-Hesketh, Skrondal, & Pickles, 2004a).

Instead of treating different model types as separate it is conceptually appealing to consider a unifying model framework. This encourages specification of models tailormade to research problems by making it easy to combine features of different model types. Furthermore, a framework facilitates a unified approach to estimation that can be implemented in a single software program.

When responses are noncontinuous, all main approaches to latent variable modeling such as item response modeling, structural equation modeling (including factor analysis), and multilevel modeling use either a generalized linear ('response function') formulation or a latent response formulation for the relationship between latent variables and responses. Models defined via a latent response formulation, as well as linear models, can equivalently be expressed using the generalized linear model formulation (e.g., Takane & de Leeuw, 1987; Bartholomew & Knott, 1999; Skrondal & Rabe-Hesketh, 2004b).

In this paper we show that frameworks based on a generalized linear model formulation can be extended considerably by using *composite links* (e.g., Thompson & Baker, 1981; Cox, 1984;

We wish to thank The Research Council of Norway for a grant supporting our collaboration.

Requests for reprints should be sent to Sophia Rabe-Hesketh, 3659 Tolman Hall, Graduate School of Education, University of California, Berkeley, CA 94720-1670, USA. E-mail: sophiarh@berkeley.edu

BMC Medical Research Methodology

BioMed Central

Research article

Modeling of longitudinal polytomous outcome from complex survey data - application to investigate an association between mental distress and non-malignant respiratory diseases

Punam Pahwa*[1,2] and Chandima P Karunanayake[1]

Address: [1]Canadian Centre for Health and Safety in Agriculture, University of Saskatchewan, 103 Hospital Drive, Saskatoon, SK, S7N OW8, Canada and [2]Department of Community Health and Epidemiology, University of Saskatchewan, 103 Hospital drive, Saskatoon, SK, S7N OW8, Canada

Email: Punam Pahwa* - pup165@mail.usask.ca; Chandima P Karunanayake - cpk646@mail.usask.ca

* Corresponding author

Published: 17 December 2009

BMC Medical Research Methodology 2009, **9**:84 doi:10.1186/1471-2288-9-84

Received: 29 January 2009
Accepted: 17 December 2009

This article is available from: http://www.biomedcentral.com/1471-2288/9/84

Abstract

Background: The data from longitudinal complex surveys based on multi-stage sampling designs contain cross-sectional dependencies among units due to clustered nature of the data and within-subject dependencies due to repeated measurements. Special statistical methods are required to analyze longitudinal complex survey data.

Methods: Statistics Canada's longitudinal National Population Health Survey (NPHS) dataset from the first five cycles (1994/1995 to 2002/2003) was used to investigate the effects of demographic, social, life-style, and health-related factors on the longitudinal changes of mental distress scores among the NPHS participants who self-reported physician diagnosed respiratory diseases, specifically asthma and chronic bronchitis. The NPHS longitudinal sample includes 17,276 persons of all ages. In this report, participants 15 years and older (n = 14,713) were considered for statistical analysis. Mental distress, an ordinal outcome variable (categories: no/low, moderate, and high) was examined. Ordered logistic regression models based on the weighted generalized estimating equations approach were fitted to investigate the association between respiratory diseases and mental distress adjusting for other covariates of interest. Variance estimates of regression coefficients were computed by using bootstrap methods. The final model was used to predict the probabilities of prevalence of no/low, moderate or high mental distress scores.

Results: Accounting for design effects does not vary the significance of the coefficients of the model. Participants suffering with chronic bronchitis were significantly at a higher risk (OR_{adj} = 1.37; 95% CI: 1.12-1.66) of reporting high levels of mental distress compared to those who did not self-report chronic bronchitis. There was no significant association between asthma and mental distress. There was a significant interaction between sex and self-perceived general health status indicating a dose-response relationship. Among females, the risk of mental distress increases with increasing deteriorating (from excellent to very poor) self-perceived general health.

Conclusions: A positive association was observed between the physician diagnosed self-reported chronic bronchitis and an increased prevalence of mental distress when adjusted for important covariates. Variance estimates of regression coefficients obtained from the sandwich estimator (i.e. not accounting for design effects) were similar to bootstrap variance estimates (i.e. accounting for design effects). Even though these two sets of variance estimates are similar, it is more appropriate to use bootstrap variance estimates.

Metrika (2010) 72:111–138
DOI 10.1007/s00184-009-0244-5

On kernel nonparametric regression designed for complex survey data

Torsten Harms · Pierre Duchesne

Received: 8 September 2007 / Published online: 12 March 2009

Abstract In this article, we consider nonparametric regression analysis between two variables when data are sampled through a complex survey. While nonparametric regression analysis has been widely used with data that may be assumed to be generated from independently and identically distributed (iid) random variables, the methods and asymptotic analyses established for iid data need to be extended in the framework of complex survey designs. Local polynomial regression estimators are studied, which include as particular cases design-based versions of the Nadaraya–Watson estimator and of the local linear regression estimator. In this paper, special emphasis is given to the local linear regression estimator. Our estimators incorporate both the sampling weights and the kernel weights. We derive the asymptotic mean squared error (MSE) of the kernel estimators using a combined inference framework, and as a corollary consistency of the estimators is deduced. Selection of a bandwidth is necessary for the resulting estimators; an optimal bandwidth can be determined, according to the MSE criterion in the combined mode of inference. Simulation experiments are conducted to illustrate the proposed methodology and an application with the Canadian survey of labour and income dynamics is presented.

Keywords Bandwidth · Design-based inference · Local linear regression · Local polynomial regression · Model-based inference · Nonparametric regression · Sampling weights · Survey sampling

T. Harms (✉)
Freie Universität Berlin, Garystraße 21, 14195 Berlin, Germany
e-mail: Torsten.Harms@gmx.com

P. Duchesne
Département de mathématiques et statistique, Université de Montréal,
CP 6128 Succ. Centre-Ville, Montréal, QC H3C 3J7, Canada
e-mail: duchesne@dms.umontreal.ca