Close

Web of Science
Page 1 (Records 1 -- 23)
[1]

Print

Record 1 of 23

Title: zCompositions - R Package for multivariate imputation of left-censored data under a compositional approach

Author(s): Palarea-Albaladejo, J (Palarea-Albaladejo, Javier); Martin-Fernandez, JA (Antoni Martin-Fernandez, Josep)

Source: CHEMOMETRICS AND INTELLIGENT LABORATORY SYSTEMS **Volume:** 143 **Pages:** 85-96 **DOI:** 10.1016/j.chemolab.2015.02.019 **Published:** APR 15 2015

Abstract: zCompositions is an R package for the imputation of left-censored data under a compositional approach. It is pertinent when the analyst assumes that the relevant information is contained on the relative variation structure of the data. For instance, in cases where the experimental data are simultaneously measured in amounts related to a same total weight or volume. The approach is used in fields like geochemistry of waters or sedimentary rocks, environmental studies related to air pollution, physicochemical analysis of glass fragments in forensic science, and among many others. In these fields, rounded zeros and nondetects are usually regarded as left-censored data that hamper any subsequent data analysis. The implemented methods consider aspects of relevance for a compositional approach such as scale invariance, subcompositional coherence or preserving the multivariate relative structure of the data. Based on solid statistical frameworks, it comprises the ability to deal with single and varying censoring thresholds, consistent treatment of closed and non-closed data, exploratory tools, multiple imputation, MCMC, robust and non-parametric alternatives, and recent proposals for count data. Key methodological aspects, new contributions, computational implementation and the practical application of the approach are discussed. (C) 2015 Elsevier B.V. All rights reserved.

Accession Number: WOS:000353730300009

Author Identifiers:

Author	Web of Science ResearcherID	ORCID Number
Palarea-Albaladejo, Javier	J-5591-2013	0000-0003-0162-669X
Martin-Fernandez, Josep Antoni	B-9208-2011	0000-0003-2366-1592

ISSN: 0169-7439

eISSN: 1873-3239

Record 2 of 23

Title: MIMCA: multiple imputation for categorical variables with multiple correspondence analysis

Author(s): Audigier, V (Audigier, Vincent); Husson, F (Husson, Francois); Josse, J (Josse, Julie)

Source: STATISTICS AND COMPUTING **Volume:** 27 **Issue:** 2 **Pages:** 501-518 **DOI:** 10.1007/s11222-016-9635-4 **Published:** MAR 2017

Abstract: We propose a multiple imputation method to deal with incomplete categorical data. This method imputes the missing entries using the principal component method dedicated to categorical data: multiple correspondence analysis (MCA). The uncertainty concerning the parameters of the imputation model is reflected using a non-parametric bootstrap. Multiple imputation using MCA (MIMCA) requires estimating a small number of parameters due to the dimensionality reduction property of MCA. It allows the user to impute a large range of data sets. In particular, a high number of categories per variable, a high number of variables or a small number of individuals are not an issue for MIMCA. Through a simulation study based on real data sets, the method is assessed and compared to the reference methods (multiple imputation using the loglinear model, multiple imputation by logistic regressions) as well to the latest works on the topic (multiple imputation by random forests or by the Dirichlet process mixture of products of multinomial distributions model). The proposed method provides a good point estimate of the parameters of the analysis model considered, such as the coefficients of a main effects logistic regression model, and a reliable estimate of the variability of the estimators. In addition, MIMCA has the great advantage that it is substantially less time consuming on data sets of high dimensions than the other multiple imputation methods.

Accession Number: WOS:000395004300013

Author Identifiers:

Author	Web of Science ResearcherID	ORCID Number
josse, julie		0000-0001-9547-891X

ISSN: 0960-3174

eISSN: 1573-1375

Record 3 of 23

Title: Multiple imputation in the presence of non-normal data

Author(s): Lee, KJ (Lee, Katherine J.); Carlin, JB (Carlin, John B.)

Source: STATISTICS IN MEDICINE **Volume:** 36 **Issue:** 4 **Pages:** 606-617 **DOI:** 10.1002/sim.7173 **Published:** FEB 2017

Abstract: Multiple imputation (MI) is becoming increasingly popular for handling missing data. Standard approaches for MI assume normality for continuous variables (conditionally on the other variables in the imputation model). However, it is unclear how to impute non-normally distributed continuous variables. Using simulation and a case study, we compared various transformations applied prior to imputation, including a novel non-parametric transformation, to imputation on the raw scale and using predictive mean matching (PMM) when imputing non-normal data. We generated data from a range of non-normal distributions, and set 50% to missing completely at random or missing at random. We then imputed missing values on the raw scale, following a zero-skewness log, Box-Cox or non-parametric transformation and using PMM with both type 1 and 2 matching. We compared inferences regarding the marginal mean of the incomplete variable and the association with a fully observed outcome. We also compared results from these approaches in the analysis of depression and anxiety symptoms in parents of very preterm compared with term-born infants. The results provide novel empirical evidence that the decision regarding how to impute a non-normal variable should be based on the nature of the relationship between the variables of interest. If the relationship is linear in the untransformed scale, transformation can introduce bias irrespective of the transformation used. However, if the relationship is non-linear, it may be important to transform the variable to accurately capture this relationship. A useful alternative is to impute the variable using PMM with type 1 matching. Copyright (C) 2016 John Wiley & Sons, Ltd.

Accession Number: WOS:000393304400004

PubMed ID: 27862164

Author Identifiers:

Author	Web of Science ResearcherID	ORCID Number
Carlin, John	B-3492-2012	0000-0002-2694-9463

ISSN: 0277-6715

eISSN: 1097-0258

Record 4 of 23

Title: Effects of Different Missing Data Imputation Techniques on the Performance of Undiagnosed Diabetes Risk Prediction Models in a Mixed-Ancestry Population of South Africa

Author(s): Masconi, KL (Masconi, Katya L.); Matsha, TE (Matsha, Tandi E.); Erasmus, RT (Erasmus, Rajiv T.); Kengne, AP (Kengne, Andre P.)

Source: PLOS ONE **Volume:** 10 **Issue:** 9 **Article Number:** e0139210 **DOI:** 10.1371/journal.pone.0139210 **Published:** SEP 25 2015

Abstract: Background
Imputation techniques used to handle missing data are based on the principle of replacement. It is widely advocated that multiple imputation is superior to other imputation methods, however studies have suggested that simple methods for filling missing data can be just as accurate as complex methods. The objective of this study was to implement a number of simple and more complex imputation methods, and assess the effect of these techniques on the performance of undiagnosed diabetes risk prediction models during external validation.
Methods
Data from the Cape Town Bellville-South cohort served as the basis for this study. Imputation methods and models were identified via recent systematic reviews. Models' discrimination was assessed and compared using C-statistic and non-parametric methods, before and after recalibration through simple intercept adjustment.
Results
The study sample consisted of 1256 individuals, of whom 173 were excluded due to previously diagnosed diabetes. Of the final 1083 individuals, 329 (30.4%) had missing data. Family history had the highest proportion of missing data (25%). Imputation of the outcome, undiagnosed diabetes, was highest in stochastic regression imputation (163 individuals). Overall, deletion resulted in the lowest model performances while simple imputation yielded the highest C-statistic for the Cambridge Diabetes Risk model, Kuwaiti Risk model, Omani Diabetes Risk model and Rotterdam Predictive model. Multiple imputation only yielded the highest C-statistic for the Rotterdam Predictive model, which were matched by simpler imputation methods.
Conclusions
Deletion was confirmed as a poor technique for handling missing data. However, despite the emphasized disadvantages of simpler imputation methods, this study showed that implementing these methods results in similar predictive utility for undiagnosed diabetes when compared to multiple imputation.

Accession Number: WOS:000361800700191

PubMed ID: 26406594

Author Identifiers:

Author	Web of Science ResearcherID	ORCID Number
Matsha, Tandi		0000-0001-5251-030X

ISSN: 1932-6203

Record 5 of 23

Title: A comparative analysis of the UK and Italian small businesses using Generalised Extreme Value models

Author(s): Andreeva, G (Andreeva, Galina); Calabrese, R (Calabrese, Raffaella); Osmetti, SA (Osmetti, Silvia Angela)

Source: EUROPEAN JOURNAL OF OPERATIONAL RESEARCH **Volume:** 249 **Issue:** 2 **Pages:** 506-516 **DOI:** 10.1016/j.ejor.2015.07.062 **Published:** MAR 1 2016

Abstract: This paper presents a cross-country comparison of significant predictors of small business failure between Italy and the UK. Financial measures of profitability, leverage, coverage, liquidity, scale and non-financial information are explored, some commonalities and differences are highlighted. Several models are considered, starting with the logistic regression which is a standard approach in credit risk modelling. Some important improvements are investigated. Generalised Extreme Value (GEV) regression is applied in contrast to the logistic regression in order to produce more conservative estimates of default probability. The assumption of non-linearity is relaxed through application of BGEVA, non-parametric additive model based on the GEV link function. Two methods of handling missing values are compared: multiple imputation and Weights of Evidence (WOE) transformation. The results suggest that the best predictive performance is obtained by BGEVA, thus implying the necessity of taking into account the low volume of defaults and non-linear patterns when modelling SME performance. WoE for the majority of models considered show better prediction as compared to multiple imputation, suggesting that missing values could be informative. (C) 2015 Elsevier B.V. and Association of European Operational Research Societies (EURO) within the International Federation of Operational Research Societies (IFORS). All rights reserved.

Accession Number: WOS:000366951100012

ISSN: 0377-2217

eISSN: 1872-6860

Record 6 of 23

Title: Multivariate Imputation of Unequally Sampled Geological Variables

Author(s): Barnett, RM (Barnett, Ryan M.); Deutsch, CV (Deutsch, Clayton V.)

Source: MATHEMATICAL GEOSCIENCES **Volume:** 47 **Issue:** 7 **Pages:** 791-817 **DOI:** 10.1007/s11004-014-9580-8 **Published:** OCT 2015

Abstract: Unequally sampled data pose a practical and significant problem for geostatistical modeling. Multivariate transformations are frequently applied in modeling workflows to reproduce the multivariate relationships of geological data. Unfortunately, these transformations may only be applied to data observations that sample all of the variables. In the case of unequal sampling, practitioners must decide between excluding incomplete observations and imputing (inferring) the missing values. While imputation is recommended by missing data theorists, the use of deterministic methods such as regression is generally discouraged. Instead, techniques such as multiple imputation (MI) are advocated to increase the accuracy, decrease the bias, and capture the uncertainty of imputed values. As missing data theory has received little attention within geostatistical literature and practice, MI has not been adapted from

its conventional form to be suitable for geological data. To address this, geostatistical algorithms are integrated within an MI framework to produce parametric and non-parametric methods. Synthetic and geometallurgical case studies are used to demonstrate the feasibility of each method, where techniques that use both spatial and colocated information are shown to outperform the alternatives.

Accession Number: WOS:000360080000003

ISSN: 1874-8961

eISSN: 1874-8953

Record 7 of 23

Title: Identification of predicted individual treatment effects in randomized clinical trials

Author(s): Lamont, A (Lamont, Andrea); Lyons, MD (Lyons, Michael D.); Jaki, T (Jaki, Thomas); Stuart, E (Stuart, Elizabeth); Feaster, DJ (Feaster, Daniel J.); Tharmaratnam, K (Tharmaratnam, Kukatharmini); Oberski, D (Oberski, Daniel); Ishwaran, H (Ishwaran, Hemant); Wilson, DK (Wilson, Dawn K.); Van Horn, ML (Van Horn, M. Lee)

Source: STATISTICAL METHODS IN MEDICAL RESEARCH **Volume:** 27 **Issue:** 1 **Pages:** 142-157 **DOI:** 10.1177/0962280215623981 **Published:** JAN 2018

Abstract: In most medical research, treatment effectiveness is assessed using the average treatment effect or some version of subgroup analysis. The practice of individualized or precision medicine, however, requires new approaches that predict how an individual will respond to treatment, rather than relying on aggregate measures of effect. In this study, we present a conceptual framework for estimating individual treatment effects, referred to as predicted individual treatment effects. We first apply the predicted individual treatment effect approach to a randomized controlled trial designed to improve behavioral and physical symptoms. Despite trivial average effects of the intervention, we show substantial heterogeneity in predicted individual treatment response using the predicted individual treatment effect approach. The predicted individual treatment effects can be used to predict individuals for whom the intervention may be most effective (or harmful). Next, we conduct a Monte Carlo simulation study to evaluate the accuracy of predicted individual treatment effects. We compare the performance of two methods used to obtain predictions: multiple imputation and non-parametric random decision trees. Results showed that, on average, both predictive methods produced accurate estimates at the individual level; however, the random decision trees tended to underestimate the predicted individual treatment effect for people at the extreme and showed more variability in predictions across repetitions compared to the imputation approach. Limitations and future directions are discussed.

Accession Number: WOS:000419874400010

PubMed ID: 26988928

ISSN: 0962-2802

eISSN: 1477-0334

Record 8 of 23

Title: Analysing Mark-Recapture-Recovery Data in the Presence of Missing Covariate Data Via Multiple Imputation

Author(s): Worthington, H (Worthington, Hannah); King, R (King, Ruth); Buckland, ST (Buckland, Stephen T.)

Source: JOURNAL OF AGRICULTURAL BIOLOGICAL AND ENVIRONMENTAL STATISTICS **Volume:** 20 **Issue:** 1 **Pages:** 28-46 **DOI:** 10.1007/s13253-014-0184-z **Published:** MAR 2015

Abstract: We consider mark-recapture-recovery data with additional individual time-varying continuous covariate data. For such data it is common to specify the model parameters, and in particular the survival probabilities, as a function of these covariates to incorporate individual heterogeneity. However, an issue arises in relation to missing covariate values, for (at least) the times when an individual is not observed, leading to an analytically intractable likelihood. We propose a two-step multiple imputation approach to obtain estimates of the demographic parameters. Firstly, a model is fitted to only the observed covariate values. Conditional on the fitted covariate model, multiple "complete" datasets are generated (i.e. all missing covariate values are imputed). Secondly, for each complete dataset, a closed form complete data likelihood can be maximised to obtain estimates of the model parameters which are subsequently combined to obtain an overall estimate of the parameters. Associated standard errors and 95 % confidence intervals are obtained using a non-parametric bootstrap. A simulation study is undertaken to assess the performance of the proposed two-step approach. We apply the method to data collected on a well-studied population of Soay sheep and compare the results with a Bayesian data augmentation approach. Supplementary materials accompanying this paper appear on-line.

Accession Number: WOS:000352698600002

Author Identifiers:

Author	Web of Science ResearcherID	ORCID Number
Buckland, Stephen	A-1998-2012	
King, Ruth	K-8297-2019	

ISSN: 1085-7117

eISSN: 1537-2693

Record 9 of 23

Title: Mixture models for undiagnosed prevalent disease and interval-censored incident disease: applications to a cohort assembled from electronic health records

Author(s): Cheung, LC (Cheung, Li C.); Pan, Q (Pan, Qing); Hyun, N (Hyun, Noorie); Schiffman, M (Schiffman, Mark); Fetterman, B (Fetterman, Barbara); Castle, PE (Castle, Philip E.); Lorey, T (Lorey, Thomas); Katki, HA (Katki, Hormuzd A.)

Source: STATISTICS IN MEDICINE **Volume:** 36 **Issue:** 22 **Pages:** 3583-3595 **DOI:** 10.1002/sim.7380 **Published:** SEP 30 2017

Abstract: For cost-effectiveness and efficiency, many large-scale general-purpose cohort studies are being assembled within large health-care providers who use electronic health records. Two key features of such data are that incident disease is interval-censored between irregular visits and there can be pre-existing (prevalent) disease. Because prevalent disease is not always immediately diagnosed, some disease diagnosed at later visits are actually undiagnosed prevalent disease. We consider prevalent disease as a point mass at time zero for clinical applications where there is no interest in time of prevalent disease onset. We demonstrate that the naive Kaplan-Meier cumulative risk estimator underestimates risks at early time points and overestimates later risks. We propose a general family of mixture models for undiagnosed prevalent disease and interval-censored incident disease that we call prevalence-incidence models. Parameters for parametric prevalence-incidence models, such as the logistic regression and Weibull survival (logistic-Weibull) model, are estimated by direct likelihood maximization or by EM algorithm. Non-parametric methods are proposed to calculate cumulative risks for cases without covariates. We compare naive Kaplan-Meier, logistic-Weibull, and non-parametric estimates of cumulative risk in the cervical cancer screening program at

Kaiser Permanente Northern California. Kaplan-Meier provided poor estimates while the logistic-Weibull model was a close fit to the non-parametric. Our findings support our use of logistic-Weibull models to develop the risk estimates that underlie current US risk-based cervical cancer screening guidelines. Published 2017. This article has been contributed to by US Government employees and their work is in the public domain in the USA.

Accession Number: WOS:000408991500010

PubMed ID: 28660629

Author Identifiers:

Author	Web of Science ResearcherID	ORCID Number
Cheung, Li		0000-0003-1625-4331

ISSN: 0277-6715

eISSN: 1097-0258

Record 10 of 23

Title: Sensitivity to imputation models and assumptions in receiver operating characteristic analysis with incomplete data

Author(s): Karakaya, J (Karakaya, Jale); Karabulut, E (Karabulut, Erdem); Yucel, RM (Yucel, Recai M.)

Source: JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION **Volume:** 85 **Issue:** 17 **Pages:** 3498-3511 **DOI:** 10.1080/00949655.2014.983111 **Published:** 2015

Abstract: Modern statistical methods using incomplete data have been increasingly applied in a wide variety of substantive problems. Similarly, receiver operating characteristic (ROC) analysis, a method used in evaluating diagnostic tests or biomarkers in medical research, has also been increasingly popular problem in both its development and application. While missing-data methods have been applied in ROC analysis, the impact of model mis-specification and/or assumptions (e.g. missing at random) underlying the missing data has not been thoroughly studied. In this work, we study the performance of multiple imputation (MI) inference in ROC analysis. Particularly, we investigate parametric and non-parametric techniques for MI inference under common missingness mechanisms. Depending on the coherency of the imputation model with the underlying data generation mechanism, our results show that MI generally leads to well-calibrated inferences under ignorable missingness mechanisms.

Accession Number: WOS:000371315300007

PubMed ID: 26379316

ISSN: 0094-9655

eISSN: 1563-5163

Record 11 of 23

Title: Correcting bias due to missing stage data in the non-parametric estimation of stage-specific net survival for colorectal cancer using multiple imputation

Author(s): Falcaro, M (Falcaro, Milena); Carpenter, JR (Carpenter, James R.)

Source: CANCER EPIDEMIOLOGY **Volume:** 48 **Pages:** 16-21 **DOI:** 10.1016/j.canep.2017.02.005 **Published:** JUN 2017

Abstract: Background: Population-based net survival by tumour stage at diagnosis is a key measure in cancer surveillance. Unfortunately, data on tumour stage are often missing for a non-negligible proportion of patients and the mechanism giving rise to the missingness is usually anything but completely at random. In this setting, restricting analysis to the subset of complete records gives typically biased results. Multiple imputation is a promising practical approach to the issues raised by the missing data, but its use in conjunction with the Pohar-Perme method for estimating net survival has not been formally evaluated.

Methods: We performed a resampling study using colorectal cancer population-based registry data to evaluate the ability of multiple imputation, used along with the Pohar-Perme method, to deliver unbiased estimates of stage-specific net survival and recover missing stage information. We created 1000 independent data sets, each containing 5000 patients. Stage data were then made missing at random under two scenarios (30% and 50% missingness).

Results: Complete records analysis showed substantial bias and poor confidence interval coverage. Across both scenarios our multiple imputation strategy virtually eliminated the bias and greatly improved confidence interval coverage.

Conclusions: In the presence of missing stage data complete records analysis often gives severely biased results. We showed that combining multiple imputation with the Pohar-Perme estimator provides a valid practical approach for the estimation of stage-specific colorectal cancer net survival. As usual, when the percentage of missing data is high the results should be interpreted cautiously and sensitivity analyses are recommended. (C) 2017 Elsevier Ltd. All rights reserved.

Accession Number: WOS:000405151500003

PubMed ID: 28315607

Author Identifiers:

Author	Web of Science ResearcherID	ORCID Number
Carpenter, James		0000-0003-3890-6206

ISSN: 1877-7821

eISSN: 1877-783X

Record 12 of 23

Title: A robust imputation method for missing responses and covariates in sample selection models

Author(s): Ogundimu, EO (Ogundimu, Emmanuel O.); Collins, GS (Collins, Gary S.)

Source: STATISTICAL METHODS IN MEDICAL RESEARCH **Volume:** 28 **Issue:** 1 **Pages:** 102-116 **DOI:** 10.1177/0962280217715663 **Published:** JAN 2019

Abstract: Sample selection arises when the outcome of interest is partially observed in a study. Although sophisticated statistical methods in the parametric and non-parametric framework have been proposed to solve this problem, it is yet unclear how to deal with selectively missing covariate data using simple multiple imputation techniques, especially in the absence of exclusion restrictions and deviation from normality. Motivated by the 2003-2004 NHANES data, where previous authors have studied the effect of socio-economic status on blood pressure with missing data on income variable, we proposed the use of a robust imputation technique based on the selection-t sample selection model. The imputation method, which is developed within the frequentist framework, is compared with competing alternatives in a simulation study. The results indicate that the robust alternative is not susceptible to the absence

of exclusion restrictions - a property inherited from the parent selection-t model - and performs better than models based on the normal assumption even when the data is generated from the normal distribution. Applications to missing outcome and covariate data further corroborate the robustness properties of the proposed method. We implemented the proposed approach within the MICE environment in R Statistical Software.

Accession Number: WOS:000454598800007

PubMed ID: 28679340

Author Identifiers:

Author	Web of Science ResearcherID	ORCID Number
Collins, Gary	A-2258-2014	0000-0002-2772-2316

ISSN: 0962-2802

eISSN: 1477-0334

Record 13 of 23

Title: Sensitivity to censored-at-random assumption in the analysis of time-to-event endpoints

Author(s): Lipkovich, I (Lipkovich, Ilya); Ratitch, B (Ratitch, Bohdana); O'Kelly, M (O'Kelly, Michael)

Source: PHARMACEUTICAL STATISTICS **Volume:** 15 **Issue:** 3 **Pages:** 216-229 **DOI:** 10.1002/pst.1738 **Published:** MAY-JUN 2016

Abstract: Over the past years, significant progress has been made in developing statistically rigorous methods to implement clinically interpretable sensitivity analyses for assumptions about the missingness mechanism in clinical trials for continuous and (to a lesser extent) for binary or categorical endpoints. Studies with time-to-event outcomes have received much less attention. However, such studies can be similarly challenged with respect to the robustness and integrity of primary analysis conclusions when a substantial number of subjects withdraw from treatment prematurely prior to experiencing an event of interest. We discuss how the methods that are widely used for primary analyses of time-to-event outcomes could be extended in a clinically meaningful and interpretable way to stress-test the assumption of ignorable censoring. We focus on a 'tipping point' approach, the objective of which is to postulate sensitivity parameters with a clear clinical interpretation and to identify a setting of these parameters unfavorable enough towards the experimental treatment to nullify a conclusion that was favorable to that treatment. Robustness of primary analysis results can then be assessed based on clinical plausibility of the scenario represented by the tipping point. We study several approaches for conducting such analyses based on multiple imputation using parametric, semi-parametric, and non-parametric imputation models and evaluate their operating characteristics via simulation. We argue that these methods are valuable tools for sensitivity analyses of time-to-event data and conclude that the method based on piecewise exponential imputation model of survival has some advantages over other methods studied here. Copyright (C) 2016 John Wiley & Sons, Ltd.

Accession Number: WOS:000379925300003

PubMed ID: 26997353

ISSN: 1539-1604

eISSN: 1539-1612

Record 14 of 23

Title: Evaluation of modelling approaches in predicting forest volume and stand age for small-scale plantation forests in New Zealand with RapidEye and LiDAR

Author(s): Xu, C (Xu, Cong); Manley, B (Manley, Bruce); Morgenroth, J (Morgenroth, Justin)

Source: INTERNATIONAL JOURNAL OF APPLIED EARTH OBSERVATION AND GEOINFORMATION **Volume:** 73 **Pages:** 386-396 **DOI:** 10.1016/j.jag.2018.06.021 **Published:** DEC 2018

Abstract: In New Zealand, 30% of plantation forests are small-scale (< 1000 ha) and knowledge of these forests, especially those less than 100 ha, is limited. These forests are expected to comprise more than 40% of the total harvest volume by 2020, so it is critical to understand the small-scale forest resource in order to plan effectively for marketing, harvesting, logistics and transport capacity. A remote sensing solution to small-scale forest description is necessary because conducting a comprehensive ground-based survey of those patchy forests is impractical. However, the utility of remote sensing prediction techniques for application in small-scale forests is unknown. This research evaluated two parametric models (multiple linear regression and seemingly unrelated regression) and two non-parametric models (k-Nearest Neighbour and Random Forest) models to predict stand variables (mean top height, basal area, volume and stand age) using model inputs including RapidEye-derived metrics and LiDAR-derived metrics. LiDAR-derived metrics were better at predicting all forest stand variables relative to RapidEye metrics. Combining LiDAR metrics with RapidEye metrics did not improve variable prediction results (on average 0.2% reduction in RMSE). Non-parametric models and parametric models performed similarly. Of all approaches tested in this study, multiple linear regression (MLR) using LiDAR-derived metrics was deemed to be the best performing modelling approach for predicting stand variables for small-scale plantation forests in New Zealand. MLR predicted mean top height (MTh) with a root-mean-square-error (RMSE) of 1.81 m, basal area (BA) with an RMSE of 9.92 m(2) ha(-1), stand volume with an RMSE of 94.93 m(3) ha(-1) and age with an RMSE of 2.17 years.

Accession Number: WOS:000446291100033

ISSN: 0303-2434

Record 15 of 23

Title: Mean response estimation with missing response in the presence of high-dimensional covariates

Author(s): Li, YJ (Li, Yongjin); Wang, QH (Wang, Qihua); Zhu, LP (Zhu, Liping); Ding, XB (Ding, Xiaobo)

Source: COMMUNICATIONS IN STATISTICS-THEORY AND METHODS **Volume:** 46 **Issue:** 2 **Pages:** 628-643 **DOI:** 10.1080/03610926.2014.1002935 **Published:** 2017

Abstract: This paper studies the problem of mean response estimation where missingness occurs to the response but multiple-dimensional covariates are observable. Two main challenges occur in this situation: curse of dimensionality and model specification. The non parametric imputation method relieves model specification but suffers curse of dimensionality, while some model-based methods such as inverse probability weighting (IPW) and augmented inverse probability weighting (AIPW) methods are the opposite. We propose a unified non parametric method to overcome the two challenges with the aiding of sufficient dimension reduction. It imposes no parametric structure on propensity score or conditional mean response, and thus retains the non parametric flavor. Moreover, the estimator achieves the optimal efficiency that a double robust estimator can attain. Simulations were conducted and it demonstrates the excellent performances of our method in various situations.

Accession Number: WOS:000386396500010

ISSN: 0361-0926
eISSN: 1532-415X

Record 16 of 23

Title: Multiple imputation of a randomly censored covariate improves logistic regression analysis

Author(s): Atem, FD (Atem, Folefac D.); Qian, J (Qian, Jing); Maye, JE (Maye, Jacqueline E.); Johnson, KA (Johnson, Keith A.); Betensky, RA (Betensky, Rebecca A.)

Source: JOURNAL OF APPLIED STATISTICS **Volume:** 43 **Issue:** 15 **Pages:** 2886-2896 **DOI:** 10.1080/02664763.2016.1155110 **Published:** DEC 2016

Abstract: Randomly censored covariates arise frequently in epidemiologic studies. The most commonly used methods, including complete case and single imputation or substitution, suffer from inefficiency and bias. They make strong parametric assumptions or they consider limit of detection censoring only. We employ multiple imputation, in conjunction with semi-parametric modeling of the censored covariate, to overcome these shortcomings and to facilitate robust estimation. We develop a multiple imputation approach for randomly censored covariates within the framework of a logistic regression model. We use the non-parametric estimate of the covariate distribution or the semi-parametric Cox model estimate in the presence of additional covariates in the model. We evaluate this procedure in simulations, and compare its operating characteristics to those from the complete case analysis and a survival regression approach. We apply the procedures to an Alzheimer's study of the association between amyloid positivity and maternal age of onset of dementia. Multiple imputation achieves lower standard errors and higher power than the complete case approach under heavy and moderate censoring and is comparable under light censoring. The survival regression approach achieves the highest power among all procedures, but does not produce interpretable estimates of association. Multiple imputation offers a favorable alternative to complete case analysis and ad hoc substitution methods in the presence of randomly censored covariates within the framework of logistic regression.

Accession Number: WOS:000384263000012
PubMed ID: 27713593
ISSN: 0266-4763
eISSN: 1360-0532

Record 17 of 23

Title: Bayesian non-parametric generation of fully synthetic multivariate categorical data in the presence of structural zeros

Author(s): Manrique-Vallier, D (Manrique-Vallier, Daniel); Hu, JC (Hu, Jingchen)

Source: JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES A-STATISTICS IN SOCIETY **Volume:** 181 **Issue:** 3 **Pages:** 635-647 **DOI:** 10.1111/rssa.12352 **Published:** JUN 2018

Abstract: Statistical agencies are increasingly adopting synthetic data methods for disseminating microdata without compromising the privacy of respondents. Crucial to the implementation of these approaches are flexible models, able to capture the nuances of the multivariate structure in the original data. In the case of multivariate categorical data, preserving this multivariate structure also often involves satisfying constraints in the form of combinations of responses that cannot logically be present in any data set-like married toddlers or pregnant men-also known as structural zeros. Ignoring structural zeros can result in both logically inconsistent synthetic data and biased estimates. Here we propose the use of a Bayesian non-parametric method for generating discrete multivariate synthetic data subject to structural zeros. This method can preserve complex multivariate relationships between variables, can be applied to high dimensional data sets with massive collections of structural zeros, requires minimal tuning from the user and is computationally efficient. We demonstrate our approach by synthesizing an extract of 17 variables from the 2000 US census. Our method produces synthetic samples with high analytic utility and low disclosure risk.

Accession Number: WOS:000434143700005
ISSN: 0964-1998
eISSN: 1467-985X

Record 18 of 23

Title: A heuristic approach to handling missing data in biologics manufacturing databases

Author(s): Mante, J (Mante, Jeanet); Gangadharan, N (Gangadharan, Nishanthi); Sewell, DJ (Sewell, David J.); Turner, R (Turner, Richard); Field, R (Field, Ray); Oliver, SG (Oliver, Stephen G.); Slater, N (Slater, Nigel); Dikicioglu, D (Dikicioglu, Duygu)

Source: BIOPROCESS AND BIOSYSTEMS ENGINEERING **Volume:** 42 **Issue:** 4 **Pages:** 657-663 **DOI:** 10.1007/s00449-018-02059-5 **Published:** APR 2019

Abstract: The biologics sector has amassed a wealth of data in the past three decades, in line with the bioprocess development and manufacturing guidelines, and analysis of these data with precision is expected to reveal behavioural patterns in cell populations that can be used for making predictions on how future culture processes might behave. The historical bioprocessing data likely comprise experiments conducted using different cell lines, to produce different products and may be years apart; the situation causing inter-batch variability and missing data points to human- and instrument-associated technical oversights. These unavoidable complications necessitate the introduction of a pre-processing step prior to data mining. This study investigated the efficiency of mean imputation and multivariate regression for filling in the missing information in historical bio-manufacturing datasets, and evaluated their performance by symbolic regression models and Bayesian non-parametric models in subsequent data processing. Mean substitution was shown to be a simple and efficient imputation method for relatively smooth, non-dynamical datasets, and regression imputation was effective whilst maintaining the existing standard deviation and shape of the distribution in dynamical datasets with less than 30% missing data. The nature of the missing information, whether Missing Completely At Random, Missing At Random or Missing Not At Random, emerged as the key feature for selecting the imputation method.

Accession Number: WOS:000462198000015
PubMed ID: 30617419

Author Identifiers:

Author	Web of Science ResearcherID	ORCID Number
Dikicioglu, Duygu		0000-0002-3018-4790

ISSN: 1615-7591
eISSN: 1615-7605

Record 19 of 23

Title: Fitting Cox Models with Doubly Censored Data Using Spline-Based Sieve Marginal Likelihood

Author(s): Li, ZG (Li, Zhiguo); Owzar, K (Owzar, Kouros)

Source: SCANDINAVIAN JOURNAL OF STATISTICS **Volume:** 43 **Issue:** 2 **Pages:** 476-486 **DOI:** 10.1111/sjos.12186 **Published:** JUN 2016

Abstract: In some applications, the failure time of interest is the time from an originating event to a failure event while both event times are interval censored. We propose fitting Cox proportional hazards models to this type of data using a spline-based sieve maximum marginal likelihood, where the time to the originating event is integrated out in the empirical likelihood function of the failure time of interest. This greatly reduces the complexity of the objective function compared with the fully semiparametric likelihood. The dependence of the time of interest on time to the originating event is induced by including the latter as a covariate in the proportional hazards model for the failure time of interest. The use of splines results in a higher rate of convergence of the estimator of the baseline hazard function compared with the usual non-parametric estimator. The computation of the estimator is facilitated by a multiple imputation approach. Asymptotic theory is established and a simulation study is conducted to assess its finite sample performance. It is also applied to analyzing a real data set on AIDS incubation time.

Accession Number: WOS:000382553500015

PubMed ID: 27239090

ISSN: 0303-6898

eISSN: 1467-9469

Record 20 of 23

Title: Trends in the incidence and outcome of paediatric out-of-hospital cardiac arrest: A 17-year observational study

Author(s): Nehme, Z (Nehme, Ziad); Namachivayam, S (Namachivayam, Siva); Forrest, A (Forrest, Anri); Butt, W (Butt, Warwick); Bernard, S (Bernard, Stephen); Smith, K (Smith, Karen)

Source: RESUSCITATION **Volume:** 128 **Pages:** 43-50 **DOI:** 10.1016/j.resuscitation.2018.04.030 **Published:** JUL 2018

Abstract: Background: System-based improvements to the chain of survival have yielded increases in survival from out-of-hospital cardiac arrest (OHCA) in adults. Comparatively little is known about the long-term trends in incidence and survival following paediatric OHCA.
Methods: Between 2000 and 2016, we included children aged <= 16 years who suffered a non-traumatic OHCA in the state of Victoria, Australia. Trends in incidence and unadjusted outcomes were assessed using linear regression and a non-parametric test for trend. Multivariable logistic regression with multiple imputation was used to identify arrest factors associated with event survival and survival to hospital discharge.
Results: Of the 1301 paediatric OHCA events attended by emergency medical services (EMS), 948 (72.9%) received an attempted resuscitation. The overall incidence of EMS-attended and EMS-treated events was 6.7 and 4.9 cases per 100,000 person-years, with no significant changes in trend. Although the proportion of cases with OHCA identified in the call and receiving bystander CPR increased over time, EMS response times also increased. Unadjusted event survival rose from 23.3% in 2000 to 33.3% in 2016 (p trend = .007), and survival to hospital discharge rose from 9.4% to 17.7% over the same period (p trend = .04). Increases in survival to hospital discharge were largely driven by initial shockable arrests, which rose from 33.3% in 2000 to 60.0% in 2016 (p trend = .005). Survival after initial shockable arrests was higher if the first shock was delivered by either first responder or public AED compared with paramedics (83.3% vs. 40.0%, p = .04). After adjustment, the odds of event survival and survival to hospital discharge increased independent of baseline characteristics, by 7% (OR 1.07, 95% CI: 1.03, 1.11; p = .001) and 8% (OR 1.08, 95% CI: 1.01, 1.15; p = .02) per study year, respectively.
Conclusions: Survival following paediatric OHCA increased in our region over a 17 year period. This was driven, in part, by improving outcomes for initial shockable arrests.

Accession Number: WOS:000436411800014

PubMed ID: 29704520

Author Identifiers:

Author	Web of Science ResearcherID	ORCID Number
Nehme, Ziad		0000-0003-2432-1645

ISSN: 0300-9572

Record 21 of 23

Title: Acculturation, Depression, and Smoking Cessation: a trajectory pattern recognition approach

Author(s): Kim, SS (Kim, Sun S.); Fang, H (Fang, Hua); Bernstein, K (Bernstein, Kunsook); Zhang, ZY (Zhang, Zhaoyang); DiFranza, J (DiFranza, Joseph); Ziedonis, D (Ziedonis, Douglas); Allison, J (Allison, Jeroan)

Source: TOBACCO INDUCED DISEASES **Volume:** 15 **Article Number:** 33 **DOI:** 10.1186/s12971-017-0135-x **Published:** JUL 24 2017

Abstract: Background: Korean Americans are known for a high smoking prevalence within the Asian American population. This study examined the effects of acculturation and depression on Korean Americans' smoking cessation and abstinence.
Methods: This is a secondary data analysis of a smoking cessation study that implemented eight weekly individualized counseling sessions of a culturally adapted cessation intervention for the treatment arm and a standard cognitive behavioral therapy for the comparison arm. Both arms also received nicotine patches for 8 weeks. A newly developed non-parametric trajectory pattern recognition model (MI-Fuzzy) was used to identify cognitive and behavioral response patterns to a smoking cessation intervention among 97 Korean American smokers (81 men and 16 women).
Results: Three distinctive response patterns were revealed: (a) Culturally Adapted (CA), since all identified members received the culturally adapted intervention; (b) More Bicultural (MB), for having higher scores of bicultural acculturation; and (c) Less Bicultural (LB), for having lower scores of bicultural acculturation. The CA smokers were those from the treatment arm, while MB and LB groups were from the comparison arm. The LB group differed in depression from the CA and MB groups and no difference was found between the CA and MB groups. Although depression did not directly affect 12-month prolonged abstinence, the LB group was most depressed and achieved the lowest rate of abstinence (LB: 1.03%; MB: 5.15%; CA: 21.65%).
Conclusion: A culturally adaptive intervention should target Korean American smokers with a high level of depression and a low level of biculturalism to assist in their smoking cessation.

Accession Number: WOS:000406971600001

PubMed ID: 28747857

ISSN: 1617-9625

Record 22 of 23

Title: Cost-effectiveness of habit-based advice for weight control versus usual care in general practice in the Ten Top Tips (10TT) trial: economic evaluation based on a randomised controlled trial

Author(s): Patel, N (Patel, Nishma); Beeken, RJ (Beeken, Rebecca J.); Leurent, B (Leurent, Baptiste); Omar, RZ (Omar, Rumana Z.); Nazareth, I (Nazareth, Irwin); Morris, S (Morris, Stephen)

Source: BMJ OPEN **Volume:** 8 **Issue:** 8 **Article Number:** e017511 **DOI:** 10.1136/bmjopen-2017-017511 **Published:** AUG 2018

Abstract: Objective Ten Top Tips (10TT) is a primary care-led behavioural intervention which aims to help adults reduce and manage their weight by following 10 weight loss tips. The intervention promotes habit formation to encourage long-term behavioural changes. The aim of this study was to estimate the cost-effectiveness of 10TT in general practice from the perspective of the UK National Health Service.
Design An economic evaluation was conducted alongside an individually randomised controlled trial.
Setting 14 general practitioner practices in England.
Participants All patients were aged 18years, with body mass index 30kg/m(2). A total of 537 patients were recruited; 270 received the usual care offered by their practices and 267 received the 10TT intervention.
Outcomes measures Health service use and quality-adjusted life years (QALYs) were measured over 2years. Analysis was conducted in terms of incremental net monetary benefits (NMBs), using non-parametric bootstrapping and multiple imputation.
Results Over a 2-year time horizon, the mean costs and QALYs per patient in the 10TT group were 1889 (95% CI 1522 to 2566) pound and 1.51 (95% CI 1.44 to 1.58). The mean costs and QALYs for usual care were 1925 pound (95% CI 1599 pound to 2251) pound and 1.51 (95% CI 1.45 to 1.57), respectively. This generated a mean cost difference of -36 pound (95% CI -512 pound to 441) pound and a mean QALY difference of 0.001 (95% CI -0.080 to 0.082). The incremental NMB for 10TT versus usual care was 49 pound (95% CI -1709 pound to 1800) pound at a maximum willingness to pay for a QALY of 20000 pound. 10TT had a 52% probability of being cost-effective at this threshold.
Conclusions Costs and QALYs for 10TT were not significantly different from usual care and therefore 10TT is as cost-effective as usual care. There was no evidence to recommend nor advice against offering 10TT to obese patients in general practices based on cost-effectiveness considerations.
Trial registration number ISRCTN16347068; Post-results.

Accession Number: WOS:000446470200002
PubMed ID: 30104307
ISSN: 2044-6055

Record 23 of 23

Title: Burden of disease of people with epilepsy during an optimized diagnostic Check for trajectory: costs and quality of life

Author(s): Wijnen, BFM (Wijnen, Ben F. M.); Schat, SL (Schat, Scarlett L.); de Kinderen, RJA (de Kinderen, Reina J. A.); Colon, AJ (Colon, Albert J.); Ossenblok, PPW (Ossenblok, Pauly P. W.); Evers, SMAA (Evers, Silvia M. A. A.)

Source: EPILEPSY RESEARCH **Volume:** 146 **Pages:** 87-93 **DOI:** 10.1016/j.eplepsyres.2018.07.024 **Published:** OCT 2018

Abstract: Background: Diagnosing epilepsy can be lengthy and stressful, potentially leading to increased use of healthcare resources and a reduction in quality of life.
Aim: This study aims to determine cost and quality of life before and after an optimized diagnostic procedure for people suspected of having epilepsy from a societal perspective with a follow-up of 12 months. In addition, this study aims to differentiate between people diagnosed with epilepsy during the follow-up of the study and the people who are diagnosed as not having epilepsy or for whom diagnosis is still uncertain.
Methods: A questionnaire regarding the use of healthcare resources was used accompanied by the EQ-5D-3 L. Multiple imputations by chained equations with predictive mean matching was used to account for missing data. To investigate the uncertainty of the results, non-parametric bootstrapped (1000 times) was used.
Results: In total, 116 people were included in the study. Total average costs per patient made in the previous 3 months had decreased from (sic)4594 before the optimized diagnostic trajectory to (sic)2609 in the 12 months after the optimized diagnostic trajectory. Healthcare costs were the largest expense group (52-66%) and had decreased significantly from baseline measurement to 12 months after baseline ((sic)2395 vs (sic)1581). Productivity costs had decreased from (sic)1367 to (sic)442 per 3 months. Total annual costs were similar between people diagnosed with epilepsy during the follow-up of the study and the people who are diagnosed as not having epilepsy or for whom diagnosis is still uncertain. Quality of Life had significantly increased over the course of 12 months from 0.80 to 0.84 (Dutch tariff).
Discussion: This study indicates that an optimized diagnostic trajectory has positively influenced the use of healthcare resources and the quality of life in people with epilepsy. As chronic care patients make diverse costs, future research should identify the long-term costs after an optimized diagnostic trajectory for patients with epilepsy, possibly identifying patients who are at high risk of becoming high-cost users in the future for early intervention.

Accession Number: WOS:000445168400011
PubMed ID: 30086483
ISSN: 0920-1211
eISSN: 1872-6844