
Predictive Models and Survey Methods

Chair: Christian Léger (Université de Montréal)

CHRISTOPHER VAN BOMMEL, St. Francis Xavier University

Keeping Pace: Physical Activity and Dietary Intake of Nova Scotia Students

Keeping Pace is a study of physical activity and dietary intake of students in Grades 3, 7, and 11 in Nova Scotia conducted in 2009. The goals of this project are to determine the physical activity levels and usual food intakes of these boys and girls, and the factors that influence their physical activity and dietary intake. In this presentation, I discuss the method used for sample selection which involved a stratification and two-stage cluster sampling, the methods used to analyze the data collected, and some key results.

MOSHE FEDER, University of Southampton, U.K.

Empirical Likelihood Approach to Non-Response

Surveys are often affected by informative nonresponse (INR), which is a challenging problem when fitting models to survey data. Methods involving weighting-class adjustments and calibration were developed years ago. Recently, methods that assume parametric models for the response probabilities and calibration equations were proposed (Chang & Kott, 2008). Qin et al. (2002) consider an empirical likelihood (EL) method for INR assuming a parametric model for the response probabilities and known population means of the covariates. In this presentation we study EL methods for model estimation that account for informative sampling and INR, without assuming a PM for the response probabilities.

ISABEL MOLINA, Universidad Carlos III de Madrid, Spain

Empirical Best Estimation under a Nested Error Linear Regression Model with Log Transformation

Log transformation of the response is often needed when fitting models to socio-economic variables. The reason is that monetary welfare variables such as income are markedly skewed and this transformation might make their distribution approximately normal. However, typical parameters of interest are the area means of the untransformed welfare variables, which will be means of the exponentials. For these particular non-linear parameters an exact analytical expression for best predictor is obtained. An analytical expression for the mean squared error of the new empirical best predictor that is second order correct for large number of areas is also derived.

FATEMEH DORRI, University of Waterloo

Covariate Shift Adaptation by Metric Learning

A common assumption in most predictive models is that training data and test data are drawn from the same underlying distribution. In many applications, however, this assumption may not be correct. We propose a novel algorithm to address this problem. The proposed method computes a transformation that projects the data to a new space with two properties. First, the distribution of training and test data are as close as possible in the transformed space, and second, dependency between predictors and response variable is maximized. This method can also reduce the dimensionality of the data while it preserves the aforementioned properties.

YAN YUAN, University of Alberta

Prediction Error Estimation with a Changing Covariates Distribution

When statistical models are used to predict the values of unobserved random variables, the prediction accuracy is often quantified via loss functions. The expected loss over the covariates distribution is prediction error. Current estimators of prediction error

assume that the distribution of covariates is unchanged in the new data as in the “training” data used to specify the predictive model. We relax this condition on covariate distribution and propose an estimator that weighs the prediction losses on the training data to estimate the average loss over the covariate distribution in the new data. Different methods of choosing weights are compared.