

Nonparametric Regression Techniques for Inference in Highly Skewed Quality Of Life Data – A Case Study with LDS Diabetes Patients Using Sf-36 Scores

Sanjoy K. Paul, Rury Holman

International Society for Clinical Biostatistics (2006); 215

This study explores the applicability of various non-parametric regression techniques including the quantile regression (QR) technique to model negatively skewed quality of life (QOL) data. Selection of covariates (risk factors) is an important aspect to address in the non-parametric linear and non-linear modelling, while dealing with the clinical trial data. This aspect, along with the inference and model selection issues in the context of QOL data from the Lipids in Diabetes Study (LDS) is addressed.

The LDS was a prospective, randomised, placebo-controlled, clinical outcome trial. The principal objective of the trial was to determine whether lipid reduction with a statin (cerivastatin) or a fibrate (fenofibrate) could substantially reduce cardiovascular related morbidity and mortality in subjects with type 2 diabetes (non-insulin dependent diabetes). 4191 people with type 2 diabetes but not known to have coronary heart disease and who were not thought to require lipid lowering therapy were randomized to lipid-lowering therapy with cerivastatin and fenofibrate in a two-by-two factorial design in thirty UK clinical centres.

The analysis is based on 4051 people (65.2 % male, 90 % caucasian) with type 2 diabetes at enrolment to the LDS. The QOL measurements, based on eight domains of SF-36 scores, were explored for these diabetic patients by their Body Mass Index (BMI) categories (45.4 % obese) and diabetes related complications (3.4 %). The statistical and clinical significance of the effects of potential clinical and demographic risk factors on the quality of life were best explored using the first, second and third quantile regression^a. QR process based on asymmetrically weighted absolute residuals offers mechanism for estimating models for a full range of conditional quantile functions. For heteroscedastic error, the Koenker method underestimates the SE. We have obtained bootstrapped SE using the standard xy – pair bootstrap (replication = 2000). We have tested if all the conditional quantile functions have the same slope parameters using the Wald approach based on asymptotic covariance matrix^b. The demographic and clinical parameters: age, duration of diabetes, gender, blood pressure, HbA_{1C} and HDL level were highly statistically significantly affecting all the eight SF-36 domains. All the QOL scores were significantly lower for patients with obesity and diabetes related complications. Estimating models with conditional quantile functions clearly showed the advantage (compared to other nonparametric methods) of assessing the effects of risk factors over the whole distribution of the highly skewed QOL scores.

^a Koenker R. & Xiao Z.: Inference on quantile regression process. *Econometrica* (2002) **70**.

^b Koenker R. : Quantile Regression in R: A Vignette. <http://cran.r-project.org>