PARAMETRIC MODELLING OF LONGITUDINAL DATA: MODEL SELECTION AND GODDNESS OF FIT

$\underline{\text{V.A. N}\acute{u}\widetilde{n}\text{ez-Ant}\acute{o}n}$

Universidad del País vasco (UPV/EHU), Bilbao, Spain

[†]E-mail: etpnuanv@bs.ehu.es or vicente.nunezanton@ehu.es

In the past two decades a parametric multivariate regression modelling approach for analyzing growth curve data has achieved prominence. The approach, which has several advantages over classical analysis-of-variance and general multivariate approaches, consists of postulating, fitting, evaluating, and comparing parametric models for the data's mean structure and covariance structure. We provide an overview of the approach, using unified terminology and notation. Well-established models and some developed more recently are described, with emphasis given to those models that allow for nonstationarity and for measurement times that differ across subjects and are unequally spaced. Graphical diagnostics that can assist with model postulation and evaluation are discussed, as are more formal methods for fitting and comparing models. Our recommendations are along the lines of first constructing profile and covariance diagnostic plots (such as the OSM and PRISM, with a saturated mean model). Second, and based on the examination of these plots, determine a set of plausible covariance structures, being somewhat liberal in what is considered plausible. Fit these and select the model(s) that fit(s) best on the basis of AIC, BIC and/or LRT (or their REML analogues). Finally, test for a parsimonious mean model (e.g. linear, quadratic, or cubic) if the profile plot suggests that it may be worthwhile. Three examples, two of them from Biology (i.e., the cattle data and the race data) and one from Medicine (i.e., the speech recognition data), serve to illustrate the methodology and to reveal the relative strengths and weaknesses of the various parametric models.