

1 Introduction to Residual Analysis

Residual analysis is a widely used model validation technique. A residual is simply the difference between an observed value and the corresponding fitted value, as predicted by the model. The rationale is that, if the model is properly fitted to the model, then the residuals would approximate the random errors that one should expect. that is to say, if the residuals behave randomly, with no discernible trend, the model has fitted the data well. If some sort of non-random trend is evident in the model, then the model can be considered to be poorly fitted.

Statistical software environments, such as the R Programming language, provides a suite of tests and graphical procedure sfor appraising a fitted linear model, with several of these procedures analysing the model residuals.

2 Framework for Model Validation using Residual Diagnostics

In statistical modelling, the process of model validation is a critical step, but also a step that is too often overlooked. A very simple procedure is to examine commonly encountered metrics, such as the R^2 value. However, using a small handful of simple measures and methods is insufficient to properly assess the quality of a fitted model. To do so properly, a full and comprehensive analysis that tests of all of the assumptions, as far as possible, must be carried out. A statistical model, whether of the fixed-effects or mixed-effects variety, represents how you think your data were generated. Following model specification and estimation, it is of interest to explore the model-data agreement by raising questions such as

- Does the model-data agreement support the model assumptions?
- Should model components be refined, and if so, which components? For example, should regressors be added or removed, and is the covariation of the observations modeled properly?
- Are the results sensitive to model and/or data? Are individual data points or groups of cases particularly influential on the analysis?

2.1 Residual Analysis

A residual is the difference between an observed quantity and its estimated or predicted value. Residual analysis is a widely used model validation technique. A residual is simply the difference between an observed value and the corresponding fitted value, as predicted by the model. The rationale is that, if the model is properly fitted to the model, then the residuals would approximate the random errors that one should expect. that is to say, if the residuals behave randomly, with no discernible trend, the model has fitted the data well. If some sort of non-random trend is evident in the model, then the model can be considered to be poorly fitted. Statistical software environments, such as the R Programming language, provides a suite of tests and graphical procedure sfor appraising a fitted linear model, with several of these procedures analysing the model residuals.

In classical linear models, an examination of model-data agreement has traditionally revolved around

The second part of the chapter looks at diagnostics techniques for LME models, firstly covering the theory, then proceeding to a discussion on implementing these using R code.

While a substantial body of work has been developed in this area, there is still areas worth exploring. In particular the development of graphical techniques pertinent to LME models should be looked at.

2.2 Outliers and Leverage

The question of whether or not a point should be considered an outlier must also be addressed. An outlier is an observation whose true value is unusual given its value on the predictor variables. The leverage of an observation is a further consideration. Leverage describes an observation with an extreme value on a predictor variable is a point with high leverage. High leverage points can have a great amount of effect on the estimate of regression coefficients.

Influence can be thought of as the product of leverage and outlierness. An observation is said to be influential if removing the observation substantially changes the estimate of the regression coefficients. The R programming language has a variety of methods used to study each of the aspects for a linear model. While linear models and GLMS can be studied with a wide range of well-established diagnostic techniques, the choice of methodology is much more restricted for the case of LMEs.

For classical linear models, residual diagnostics are typically conducted using a plot of the observed residuals and the predicted values. A visual inspection for the presence of trends inform the analyst on the validity of distributional assumptions, and to detect outliers and influential observations.

2.3 Matrix Notation for Case Deletion

For notational simplicity, $\mathbf{A}(i)$ denotes an $n \times m$ matrix \mathbf{A} with the i -th row removed, a_i denotes the i -th row of \mathbf{A} , and a_{ij} denotes the (i, j) -th element of \mathbf{A} .

2.4 Extension of Diagnostic Methods to LME models

iiiiiii HEAD

When similar notions of statistical influence are applied to mixed models, things are more complicated. Removing data points affects fixed effects and covariance parameter estimates. Update formulas for *leave-one-out* estimates typically fail to account for changes in covariance parameters.

? noted the case deletion diagnostics techniques have not been applied to linear mixed effects models and seeks to develop methodologies in that respect. ? develops these techniques in the context of REML. ===== Christensen et al. (1992) noted the case deletion diagnostics techniques had not been applied to linear mixed effects models and seeks to develop methodologies in that respect. Christensen et al. (1992) develops these techniques in the context of REML. llllllll origin/master

Demidenko (2004) extends several regression diagnostic techniques commonly used in linear regression, such as leverage, infinitesimal influence, case deletion diagnostics, Cook's distance, and local influence to the linear mixed-effects model. In each case, the proposed new measure has a direct interpretation in terms of the effects on a parameter of interest, and reduces to the familiar linear regression measure when there are no random effects.

The new measures that are proposed by Demidenko (2004) are explicitly defined functions and do not require re-estimation of the model, especially for cluster deletion diagnostics. The basis for both the cluster deletion diagnostics and Cook's distance is a generalization of Miller's simple update formula for case deletion for linear models. Furthermore Demidenko (2004) shows how Pregibon's infinitesimal case deletion diagnostics is adapted to the linear mixed-effects model.

Demidenko (2004) proposes two likelihood-based diagnostics for identifying individuals that can influence the choice between two competing models.

3 Model Validation Framework

In statistical modelling, the process of model validation is a critical step of model fitting process, but also a step that is too often overlooked. A very simple procedure is to examine commonly-used metrics, such as the R^2 value. However, using a small handful of simple measures and methods is insufficient to properly assess the quality of a fitted model. To do so properly, a full and comprehensive analysis that tests of all of the assumptions, as far as possible, must be carried out.

Schabenberger (2005) describes the model validation framework as comprised of the following tasks

- overall measures of goodness-of-fit
- the informal, graphical examination of estimates of model errors to assess the quality of distributional assumptions: residual analysis
- the quantitative assessment of the inter-relationship of model components; for example, collinearity diagnostics
- the qualitative and quantitative assessment of influence of cases on the analysis, i.e. influence analysis.

The sensitivity of a model is studied through measures that express its stability under perturbations. You are not interested in a model that is either overly stable or overly sensitive. Changes in the data or model components should produce commensurate changes in the model output. The difficulty is to determine when the changes are substantive enough to warrant further investigation, possibly leading to a reformulation of the model or changes in the data (such as dropping outliers). This paper is primarily concerned with stability of linear mixed models to perturbations of the data; that is, with influence analysis.

3.1 Residual Analysis

Residual analysis is a widely used model validation technique. A residual is simply the difference between an observed value and the corresponding fitted value, as predicted by the model. The rationale is that, if the model is properly fitted to the model, then the residuals would approximate the random errors that one should expect. that is to say, if the residuals behave randomly, with no discernible trend, the model has fitted the data well. If some sort of non-random trend is evident in the model, then the model can be considered to be poorly fitted.

Statistical software environments, such as the R Programming language, provides a suite of tests and graphical procedure sfor appraising a fitted linear model, with several of these procedures analysing the model residuals.

3.2 Outliers and Leverage

The question of whether or not a point should be considered an outlier must also be addressed. An outlier is an observation whose true value is unusual given its value on the predictor variables. The leverage of an observation is a further consideration. Leverage describes an observation with an extreme value on a predictor variable is a point with high leverage. High leverage points can have a great amount of effect on the estimate of regression coefficients.

Influence can be thought of as the product of leverage and outlierness. An observation is said to be influential if removing the observation substantially changes the estimate of the regression coefficients. The R programming language has a variety of methods used to study each of the aspects for a linear model. While linear models and GLMS can be studied with a wide range of well-established diagnostic technqiues, the choice of methodology is much more restricted for the case of LMEs.

4 Regression Of Differences On Averages

Further to Carstensen, we can formulate the two measurements y_1 and y_2 as follows:

$$y_1 = \alpha + \beta\mu + \epsilon_1$$

$$y_2 = \alpha + \beta\mu + \epsilon_2$$

4.1 Remarks on the Multivariate Normal Distribution

Diligence is required when considering the models. Carstensen specifies his models in terms of the univariate normal distribution. Roy's model is specified using the bivariate normal distribution. This gives rise to a key difference between the two models, in that a bivariate model accounts for covariance between the variables of interest. The multivariate normal distribution of a k -dimensional random vector $X = [X_1, X_2, \dots, X_k]$ can be written in the following notation:

$$X \sim \mathcal{N}(\mu, \Sigma),$$

or to make it explicitly known that X is k -dimensional,

$$X \sim \mathcal{N}_k(\mu, \Sigma).$$

with k -dimensional mean vector

$$\mu = [E[X_1], E[X_2], \dots, E[X_k]]$$

and $k \times k$ covariance matrix

$$\Sigma = [\text{Cov}[X_i, X_j]], \quad i = 1, 2, \dots, k; \quad j = 1, 2, \dots, k$$

1. Univariate Normal Distribution

$$X \sim \mathcal{N}(\mu, \sigma^2),$$

2. Bivariate Normal Distribution

(a)

$$X \sim \mathcal{N}_2(\mu, \Sigma),$$

(b)

$$\mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}.$$

4.2 Note 1: Coefficient of Repeatability

The coefficient of repeatability is a measure of how well a measurement method agrees with itself over replicate measurements (Bland and Altman, 1999). Once the within-item variability is known, the computation of the coefficients of repeatability for both methods is straightforward.

4.3 Note 2: Carstensen model in the single measurement case

Carstensen (2004) presents a model to describe the relationship between a value of measurement and its real value. The non-replicate case is considered first, as it is the context of the Bland-Altman plots. This model assumes that inter-method bias is the only difference between the two methods.

$$y_{mi} = \alpha_m + \mu_i + e_{mi} \quad e_{mi} \sim \mathcal{N}(0, \sigma_m^2) \quad (1)$$

The differences are expressed as $d_i = y_{1i} - y_{2i}$.

For the replicate case, an interaction term c is added to the model, with an associated variance component.

4.4 Note 3: Model terms

It is important to note the following characteristics of this model.

- Let the number of replicate measurements on each item i for both methods be n_i , hence $2 \times n_i$ responses. However, it is assumed that there may be a different number of replicates made for different items. Let the maximum number of replicates be p . An item will have up to $2p$ measurements, i.e. $\max(n_i) = 2p$.
- Later on \mathbf{X}_i will be reduced to a 2×1 matrix, to allow estimation of terms. This is due to a shortage of rank. The fixed effects vector can be modified accordingly.
- \mathbf{Z}_i is the $2n_i \times 2$ model matrix for the random effects for measurement methods on item i .
- \mathbf{b}_i is the 2×1 vector of random-effect coefficients on item i , one for each method.
- $\boldsymbol{\epsilon}$ is the $2n_i \times 1$ vector of residuals for measurements on item i .
- \mathbf{G} is the 2×2 covariance matrix for the random effects.
- \mathbf{R}_i is the $2n_i \times 2n_i$ covariance matrix for the residuals on item i .
- The expected value is given as $E(\mathbf{y}_i) = \mathbf{X}_i\boldsymbol{\beta}$. (Hamlett et al., 2004)
- The variance of the response vector is given by $\text{Var}(\mathbf{y}_i) = \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i' + \mathbf{R}_i$ (Hamlett et al., 2004).

Roy uses and LME model approach to provide a set of formal tests for method comparison studies.

Four candidates models are fitted to the data.

These models are similar to one another, but for the imposition of equality constraints.

These tests are the pairwise comparison of candidate models, one formulated without constraints, the other with a constraint.

Roy's model uses fixed effects $\beta_0 + \beta_1$ and $\beta_0 + \beta_1$ to specify the mean of all observations by methods 1 and 2 respectively.

Roy adheres to Random Effect ideas in ANOVA

Roy treats items as a sample from a population.

Allocation of fixed effects and random effects are very different in each model

Carstensen's interest lies in the difference between the population from which they were drawn.

Carstensen's model is a mixed effects ANOVA.

$$Y_{mir} = \alpha_m + \mu_i + c_{mi} + e_{mir}, \quad c_{mi} \sim \tau_{\Downarrow}^{\epsilon}, \quad e_{mir} \sim \sigma_{\Downarrow}^{\epsilon},$$

This model includes a method by item interaction term.

Carstensen presents two models. One for the case where the replicates, and a second for when they are linked.

Carstensen's model does not take into account either between-item or within-item covariance between methods.

In the presented example, it is shown that Roy's LoAs are lower than those of Carstensen. Carstensen makes some interesting remarks in this regard.

The only slightly non-standard (meaning "not often used") feature is the differing residual variances between methods.

4.5 Note 3: Model terms

It is important to note the following characteristics of this model.

- Let the number of replicate measurements on each item i for both methods be n_i , hence $2 \times n_i$ responses. However, it is assumed that there may be a different number of replicates made for different items. Let the maximum number of replicates be p . An item will have up to $2p$ measurements, i.e. $\max(n_i) = 2p$.
- Later on \mathbf{X}_i will be reduced to a 2×1 matrix, to allow estimation of terms. This is due to a shortage of rank. The fixed effects vector can be modified accordingly.
- \mathbf{Z}_i is the $2n_i \times 2$ model matrix for the random effects for measurement methods on item i .
- \mathbf{b}_i is the 2×1 vector of random-effect coefficients on item i , one for each method.
- $\boldsymbol{\epsilon}$ is the $2n_i \times 1$ vector of residuals for measurements on item i .
- \mathbf{G} is the 2×2 covariance matrix for the random effects.
- \mathbf{R}_i is the $2n_i \times 2n_i$ covariance matrix for the residuals on item i .
- The expected value is given as $E(\mathbf{y}_i) = \mathbf{X}_i\boldsymbol{\beta}$. (Hamlett et al., 2004)
- The variance of the response vector is given by $\text{Var}(\mathbf{y}_i) = \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i' + \mathbf{R}_i$ (Hamlett et al., 2004).

5 Regression Of Differences On Averages

Further to Carstensen, we can formulate the two measurements y_1 and y_2 as follows:

$$y_1 = \alpha + \beta\mu + \epsilon_1$$

$$y_2 = \alpha + \beta\mu + \epsilon_2$$

References

- Bland, J. and D. Altman (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research* 8(2), 135–160.
- Carstensen, B. (2004). Comparing and predicting between several methods of measurement. *Biostatistics* 5(3), 399–413.
- Christensen, R., L. M. Pearson, and W. Johnson (1992). Case-deletion diagnostics for mixed models. *Technometrics* 34(1), 38–45.
- Demidenko, E. (2004). *Mixed Models: Theory And Application*. Dartmouth College: Wiley Interscience.
- Hamlett, A., L. Ryan, and R. Wolfinger (2004). On the use of PROC MIXED to estimate correlation in the presence of repeated measures. *Proceedings of the Statistics and Data Analysis Section, SAS Users Group International 198-229*, 1–7.
- Schabenberger, O. (2005). Mixed model influence diagnostics. In *SUGI*, Volume 29, pp. 189–29.