Featured on this page

Hat Matrix

Featured on this page

Chapters and Articles

You might find these chapters and articles relevant to this topic.

Linear Models, Problems

John Fox, in Encyclopedia of Social Measurement, 2005

Leverage: Hat-Values

The fitted values $\hat{\mathbf{y}}$ in linear least-squares regression are a linear transformation of the observed response variable: $\hat{\mathbf{y}}=\mathbf{Xb}=\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}=\mathbf{Hy}$, where $\mathbf{H}= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is called the hat-matrix (because it transforms $\mathbf{y}$ to $\hat{\mathbf{y}}$). The matrix $\mathbf{H}$ is symmetric ($\mathbf{H}=\mathbf{H}^T$) and idempotent ($\mathbf{H}=\mathbf{H}^2$), and thus its ith diagonal entry, h_{ii}, gives the sum of squared entries in its ith row or column. Because the ith row of $\mathbf{H}$ represents the weight associated with the ith observed response in determining each of the fitted values, $h_i=h_{ii}$, called the hat-value, summarizes the potential contribution of observation i to the fitted values collectively, and is a suitable measure of the leverage of this observation in the least-squares fit.

Because $\mathbf{H}$ is a projection matrix (projecting $\mathbf{y}$ orthogonally onto the subspace spanned by the columns of $\mathbf{X}$), the average hat-value is p/n. By rule of thumb, hat-values are considered noteworthy when they exceed twice (or in small samples, three times) the average value. It is best, however, to examine hat-values graphically—for example, in an index plot (a scatterplot with hat-values on the vertical axis and

Read full chapter

URL: https://www.sciencedirect.com/science/article/pii/B0123693985001651

Linear Regression Models

Milan Meloun, Jiří Militký, in Statistical Data Analysis, 2011

6.5.2.2 Analysis of Projection Matrix Elements

Analysis of elements of the projection hat matrix plays an important role in regression diagnostics because the diagonal elements of this matrix $H_{ii} = \mathbf{x}_i(\mathbf{X}^T\mathbf{X})^{-1}\,\mathbf{x}^T_i$ indicate the presence of leverage points which are not detected by analysis of residuals. Diagonal elements (denoted in literature as "leverage") have some properties which come from the symmetry and idempotency of matrix **H**. Among the properties of matrix **H** are:

(1) The condition for the diagonal elements of a projection matrix is $0 < H_{ii} < 1$ and for nondiagonal elements $-1 < H_{ij} < 1$. When a model also contains an intercept term and the rank of matrix X is m, another condition for diagonal elements is valid, $1/n < H_{ii} < 1/C$, where C is the number of replicate measurements at each value of the controllable variable.

(2) For a model with an intercept term and the full rank of matrix, **X**: $\sum_{i=1}^{n} H_{\mathrm{ii}} = m$ and $\sum_{i=1}^{n} H_{\mathrm{ij}} = 1$. The mean value of the diagonal element $H_{ii} = m / n$.

Read more

Read full chapter

URL: https://www.sciencedirect.com/science/article/pii/B9780857091093500066

Volume 3

J.H. Kalivas, in Comprehensive Chemometrics, 2009

3.01.7.1 Effective Rank

To date, the author is aware of three approaches to estimate the ER. One uses a common basis set and the other two are independent of the basis set. The common basis set approach defines ER as

$$^{f}\mathrm{ER} = \sum_{i=1}^{\mathrm{k}} f_i = \mathrm{trace}(XX^{+}) = \mathrm{trace}(H) \qquad (15)$$

where **H** (the **H** matrix is known as the hat matrix because $\hat{y} = \mathrm{Hy}$) and $\mathbf{X}^+$ are for respective models in the **V** basis set, and f_i are filter values from Equation (13).[24]

perturbations to **y**, albeit results are rather invariant to the actual tuning parameter value, that is, a large range of values produce the same results.[36,41] This generalized degrees of freedom (GDF) ER is defined as the sum of sensitivities ($\hat{h}_i$) for the fitted values ($\hat{y}$) based on a multivariate calibration model to perturbations in respective observed values (**y**). The algorithm used in this chapter adds normally distributed noise (δ) to **y** N times, obtaining respective $\hat{y}$ vectors from the models formed using the corresponding perturbed **y**, calculating $\hat{h}_i$ for the ith sample as the regression slope to $\hat{y}_i = \alpha + \hat{h}_i \delta_{ji}$ for $j=1,\ldots,N$ with intercept α and estimating ER by

$$^{\mathrm{GDF}}\mathrm{ER} = \sum_{i=1}^{m} \hat{h}_i \qquad (16)$$

The second independent basis set ER measure is based on leave-one-out cross-validation.[42] The pseudo-degrees of freedom (PDF) ER for a particular model is computed from

$$^{\mathrm{PDF}}\mathrm{ER} = m\left(1 - \left[\frac{\sum_{i=1}^{m} (y_i - \hat{y}_i)^2_{\mathrm{calib}}}{\sum_{i=1}^{m} (y_i - \hat{y}_i)^2_{\mathrm{CV}}}\right]^{1/2}\right) \qquad (17)$$

where $\sum_{i=1}^{m} (y_i - \hat{y}_i)^2_{\mathrm{CV}}$ is from a leave-one-out cross-validation and $\sum_{i=1}^{m} (y_i - \hat{y}_i)^2_{\mathrm{calib}}$ is for the calibration samples. If the data are mean centered, then $^{\mathrm{PDF}}\mathrm{ER}$ needs to be adjusted by subtracting 1. ER computed from the three methods have been found to be essentially equivalent.[38,43]

As with harmonious curves, the three ER measures provide mechanisms for an impartial graphical comparison of PCR, RR, and PLS.

Read full chapter

URL: https://www.sciencedirect.com/science/article/pii/B9780444527011000727

Model Complexity (and How Ensembles Help)

Robert Nisbet, ... Gary Miner, in
Handbook of Statistical Analysis and Data Mining Applications, 2009

Generalized Degrees of Freedom

For LR, the degrees of freedom, K, equal the number of terms, though this does not extrapolate to nonlinear regression. But there exists another definition that does:

$$K = \mathrm{trace}(\,) = \Sigma \;\delta\, Y\_\mathrm{hat} \,/\, \delta\, Y \qquad (1)$$

$$\delta\, Y = \mathrm{Ye} - Y, \text{and } \delta\, Y\_\mathrm{hat} = \mathrm{Ye\_hat} - Y\_\mathrm{hat} \tag{2}$$

$$Y\_\mathrm{hat} = f\left(Y, X\right) \text{for model } f\left(\;\right), \text{output } Y, \text{and input vectors}, X; \mathrm{Y} \tag{3}$$

$$\mathrm{Ye} = Y + N(0,\; \sigma_\varepsilon\;). \tag{4}$$

Read more

Read full chapter

URL: https://www.sciencedirect.com/science/article/pii/B9780123747655000188

Residual and influence diagnostics

Xian Liu, in
Methods and Applications of Longitudinal Data Analysis, 2016

6.2.2 Leverage

In linear regression models, leverage is used to assess outliers with respect to the independent variables by identifying the observations that are distant from the average predictor values. While potentially impactful on the parameter estimates and the model fit, a higher leverage point does not necessarily indicate strong influence on the regression coefficient estimates because a far distance for a subject's predictor values from those of others can be situated in the same regression line as other observations (Fox, 1991). Therefore, checking a substantial influence must combine high leverage with discrepancy of the case from the rest of the data.

The basic measurement of leverage is the so-called hat-value, denoted by h_i. In general linear models, the hat-value is specified as a weight variable in the expression of the fitted value of the predicted response $\hat{y}_j$, given by

$$\hat{y}_{\tilde{j}} = \sum_{i=1}^{N} h_{i\tilde{j}} y_i, \tag{6.21}$$

Read more

URL: https://www.sciencedirect.com/science/article/pii/B978012801342700006X

Volume 3

J. Ferré, in Comprehensive Chemometrics, 2009

3.02.2.5 Fitted Model and Hat Matrix

After solving Equation (6), the estimated (fitted) regression equation is

$$\hat{y} = b_0 + b_1 x_1 + \cdots + b_J x_J \tag{8}$$

which is also known as the sample regression function to emphasize that it is an estimate of the population regression function calculated using the actual statistical sample. The predicted (estimated, fitted) value of y_i at the ith data point is

$$\hat{y}_i = x_i^{\mathrm{T}} b \tag{9}$$

where $\hat{y}_i$ is the estimated mean of y at the chosen levels of the x-variables and $\mathbf{x}_i^{\mathrm{T}}$ the ith row of matrix $\mathbf{X}$. The estimated responses for the complete set of regression data, i=1, …, I are

$$\hat{y} = \mathrm{Xb} \tag{10}$$

By combining Equations (7) and (10), the prediction is also given by

$$\hat{y} = X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y = Hy \tag{11}$$

Read more

Read full chapter

URL: https://www.sciencedirect.com/science/article/pii/B9780444527011000764

Review of preprocessing methods for univariate volatile time-series in power system applications

Kumar Gaurav Ranjan, ... Debashisha Jena, in Electric Power Systems Research, 2021

3 Well-established preprocessing methods

Amidst the numerous preprocessing methods, each category's well-established techniques are analyzed in detail. SWP and portrait their decent preprocessing capabilities [34-36,38]. The SWP-based method is widely used because of its ability to trace the trend and seasonality of time-series better than other algorithms. The portrait dataset-based method gives a new way of visualizing data, portraying the data's hidden characteristics like trend and seasonality, thus enhancing the approach's performance. B-spline smoothing-based method [47] is a nonparametric statistical method used by many researchers as a benchmark to compare their methods. This approach can preprocess any volatile time-series without making any

Read more

Read full article

URL: https://www.sciencedirect.com/science/article/pii/S0378779620306830

Statistical analysis of multivariate data

Milan Meloun, Jiří Militký, in Statistical Data Analysis, 2011

4.8.9 Receiver Operating Characteristic Plot (ROC)

Sometimes the investigator would want to know how much to trust such predictions. In other words, can the equation predict correctly a high proportion of the time? This question is different from asking about the statistical significance, as it is possible to obtain statistically significant results that do not predict very well. If the investigator wishes to classify cases, a cutoff point on the probability of being depressed, for example, must be found. This cutoff point is denoted by P_c. The user can quickly try a set of cutoff points and zero in on a good one.

It is possible to create an *ROC* (*receiver operating characteristic*) curve. *ROC* was originally proposed for signal-detection work when the signals were not always correctly received. This is often called the *true positive fraction*, or *sensitivity* in medical research. On the horizontal axis is the proportion of nondepressed persons classified as depressed (called *false positive fraction* or one minus *specificity* in medical studies). Three *ROC* curves are drawn in Fig 4.49. The top one represents a

Read full chapter

URL: https://www.sciencedirect.com/science/article/pii/B9780857091093500042

Computational Statistics with R

Hrishikesh D. Vinod, in Handbook of Statistics, 2014

6 Matrix Algebra in Regression Models

Consider the familiar regression model

$$y = X\beta + \epsilon, \tag{9}$$

in matrix notation, where y is a $T \times 1$ vector, X is $T \times p$ matrix, β is a $p \times 1$ vector, and ε is $T \times 1$ vector. In statistics, it is well known that

$$b = (X'X)^{-1}X'y \tag{10}$$

is the ordinary least squares (OLS) regression coefficient vector minimizing $\varepsilon'\varepsilon$, the error sum of squares.

For the cars example, consider the regression of fuel economy measured by mpg on weight and horsepower by the R commands:

```
reg1=lm(mpg~wt+hp); summary(reg1)
require(xtable) #create a Latex table of
regression results
```

Read more

Read full chapter

URL: https://www.sciencedirect.com/science/article/pii/B9780444634313000048

Financial, Macro and Micro Econometrics Using R

Giancarlo Ferrara, in Handbook of Statistics, 2020

2.3 P-splines: Computational aspects

P-splines, as introduced by Eilers and Marx (1996), are a parametric approach defined by bases of spline with penalties. They have many frontier models their main feature is making possible to impose additional monotonicity and concavity constraints on the fitted function.

The remainder outlines the P-spline framework as described by Eilers and Marx (1996) and highlights how it is possible to impose additional constraints to the fitted function in order to respect monotonicity and concavity properties separately for each covariate.

Let (x_i, Y_i) be the i-th observation $(i = 1, \ldots, n)$ associated to the response variable Y and to the covariate X. The interest here is aimed to the estimate of a smooth function $\psi(\cdot)$, such that $E[Y_i|x_i] = \mu_i = \psi(x_i)$. The key concept of this approach, fully parameterized, it is to express the

Read more

[Read full chapter](https://www.sciencedirect.com/science/article/pii/S016971611830097X)

URL: https://www.sciencedirect.com/science/article/pii/S016971611830097X

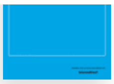
Related terms:

Confidence Interval, Phonological Awareness, Idempotent, Least-Squares Estimate, Leverage Point, Regression Outlier, Diagonal Element, Projection Matrix, Regression Parameter, Schweppe Weight.

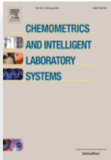
Recommended publications

Computational Statistics & Data Analysis

Journal

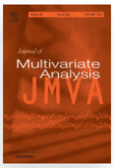


Journal



Chemometrics and Intelligent Laboratory Systems

Journal



Journal of Multivariate Analysis

Journal