Hat matrix

From Wikipedia, the free encyclopedia

In statistics, the **hat matrix**, ***H***, sometimes also called **projection matrix**, maps the vector of observed values to the vector of fitted values. It describes the influence each observed value has on each fitted value.[1] The diagonal elements of the hat matrix are the leverages, which describe the influence each observed value has on the fitted value for that same observation.

If the vector of observed values is denoted by $\mathbf{y}$ and the vector of fitted values by $\hat{\mathbf{y}}$,

$$\hat{\mathbf{y}} = H\mathbf{y}.$$

As $\hat{\mathbf{y}}$ is usually pronounced "y-hat", the hat matrix is so named as it "puts a hat on $\mathbf{y}$".

Suppose that we wish to solve a linear model using linear least squares. The model can be written as

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where X is a matrix of explanatory variables (the design matrix), $\boldsymbol{\beta}$ is a vector of unknown parameters to be estimated, and $\boldsymbol{\varepsilon}$ is the error vector.

Contents

Uncorrelated errors

For uncorrelated errors, the estimated parameters are

$$\hat{\boldsymbol{\beta}} = \left(X^\top X\right)^{-1} X^\top \mathbf{y},$$

so the fitted values are

$$\hat{\mathbf{y}} = X\hat{\boldsymbol{\beta}} = X\left(X^\top X\right)^{-1} X^\top \mathbf{y}.$$

Therefore the hat matrix is given by

$$H = X\left(X^\top X\right)^{-1} X^\top.$$

In the language of linear algebra, the hat matrix is the orthogonal projection onto the column space of the design matrix X. (Note that $\left(X^\top X\right)^{-1} X^\top$ is the pseudoinverse of X.)

The hat matrix corresponding to a linear model is symmetric and idempotent, that is, $H^2 = H$. However, this is not always the case; in locally weighted scatterplot smoothing (LOESS), for example, the hat matrix is in general neither symmetric nor idempotent.

The formula for the vector of residuals **r** can be expressed compactly using the hat matrix:

$$\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - H\mathbf{y} = (I - H)\mathbf{y}.$$

The covariance matrix of the residuals is therefore, by error propagation, equal to $\left(I - H\right)^\top \Sigma \left(I - H\right)$, where Σ is the covariance matrix of the errors (and by extension, the observations as well). For the case of linear models with independent and identically distributed errors in which $\Sigma = \sigma^2 I$, this reduces to $(I - H)\sigma^2$.[1]

For linear models, the trace of the hat matrix is equal to the rank of X, which is the number of independent parameters of the linear model. For other models such as LOESS that are still linear in the observations $\mathbf{y}$, the hat matrix can be used to define the effective degrees of freedom of the model.

The hat matrix has a number of useful algebraic properties.[2][3] Practical applications of the hat matrix in regression analysis include leverage and Cook's distance, which are concerned with identifying observations which have a large effect on the results of a regression.

Correlated errors

The above may be generalized to the case of correlated errors. Suppose that the covariance matrix of the errors is Σ. Then since

$$\hat{\boldsymbol\beta} = \left(X^\top \Sigma^{-1} X\right)^{-1} X^\top \Sigma^{-1} \mathbf{y},$$

the hat matrix is thus

$$H = X\left(X^\top \Sigma^{-1} X\right)^{-1} X^\top \Sigma^{-1},$$

and again it may be seen that $H^2 = H$

Blockwise formula

Suppose the design matrix C can be decomposed by columns as $C = [A, B]$. Define the Hat operator as $H(X) = X\left(X^\top X\right)^{-1} X^\top$. Similarly, define the residual operator as $M(X) = I - H(X)$. Then the Hat matrix of C can be decomposed as follows:

$H(C) = H(A) + H(M(A)B)$ [4]

There are a number of applications of such a partitioning. The classical application has A a column of all ones, which allows one to analyze the effects of adding an intercept term to a regression. Another use is in the fixed effects model, where A is a large sparse matrix of the dummy variables for the fixed effect terms. One can use this partition to compute the hat matrix of C without explicitly forming the matrix C, which might be too large to fit into computer memory.

See also

- Moore–Penrose pseudoinverse
- Studentized residuals
- Effective degrees of freedom
- Idempotent matrix

References

1. ^ ***a*** ***b*** Hoaglin, David C.; Welsch, Roy E. (February 1978), "The Hat Matrix in Regression and ANOVA", *The American Statistician* **32** (1): 17–22, doi:10.2307/2683469 (http://dx.doi.org/10.2307%2F2683469), JSTOR 2683469 (https://www.jstor.org/stable/2683469)
2. ^ Gans, P. (1992) *Data Fitting in the Chemical Sciences,*, Wiley. ISBN 978-0-471-93412-7
3. ^ Draper, N.R., Smith, H. (1998) *Applied Regression Analysis*, Wiley. ISBN 0-471-17082-6
4. ^ Rao, C. Radhakrishna; Toutenburg, Shalabh, Heumann (2008). *Linear Models and Generalizations* (3rd ed.). Berlin: Springer. p. 323. ISBN 978-3-540-74226-5.

Categories: Statistical terminology | Regression analysis | Matrices