Lectures 25 & 26 – The Hat Matrix

Definition: For design matrix $\mathbf{X}$, the Hat Matrix $\mathbf{H}$ is defined as $\mathbf{H} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T$. The name follows from the fact that in regression, simple or multiple, $\mathbf{H}\mathrm{Y} = \hat{\mathrm{Y}}$, i.e., $\mathbf{H}$ puts the "hat" on Y.

Facts and Properties of the Hat matrix $\mathbf{H}$

1. $\mathbf{H}$ is a Projection Matrix: $\mathbf{H}$ "projects" vectors, such as Y, onto the column space of the design matrix $\mathbf{X}$, $Col\left(\mathbf{X}\right)$. In Linear Algebra, $\mathbf{H}$ would be called a Projection Matrix.

2. $\mathbf{H}$ is an $n \times n$ symmetric matrix, i.e., $\mathbf{H}^T = \mathbf{H}$. This is really a two-part statement.

a. The hat matrix is a square $n \times n$ matrix because the first and last matrices in its definition, $\mathbf{X}$ and $\mathbf{X}^T$, have n rows and n columns, respectively.

b. $\mathbf{H}^T = \left(\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\right)^T = \left(\mathbf{X}^T\right)^T\left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right)^T(\mathbf{X})^T = (\mathbf{X})\left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right)\left(\mathbf{X}^T\right) = \mathbf{H}$, where we made repeated use of the properties of transposes. **Note:** The middle matrix $\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$ is itself a symmetric 2×2 matrix.

3. $\mathbf{H}$ is Idempotent, i.e., $\mathbf{HH} = \mathbf{H}$. *Proof*,

$$\mathbf{HH} = \left(\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\right)\left(\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\right) = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\left(\mathbf{X}^T\mathbf{X}\right)\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T = \mathbf{H}$$

4. The matrix $\left(\mathbf{I} - \mathbf{H}\right)$, where $\mathbf{I}$ is the $n \times n$ identity matrix, is both symmetric and idempotent as well. The proof will make a nice quiz question.

The Hat Matrix and the Vector of Residuals

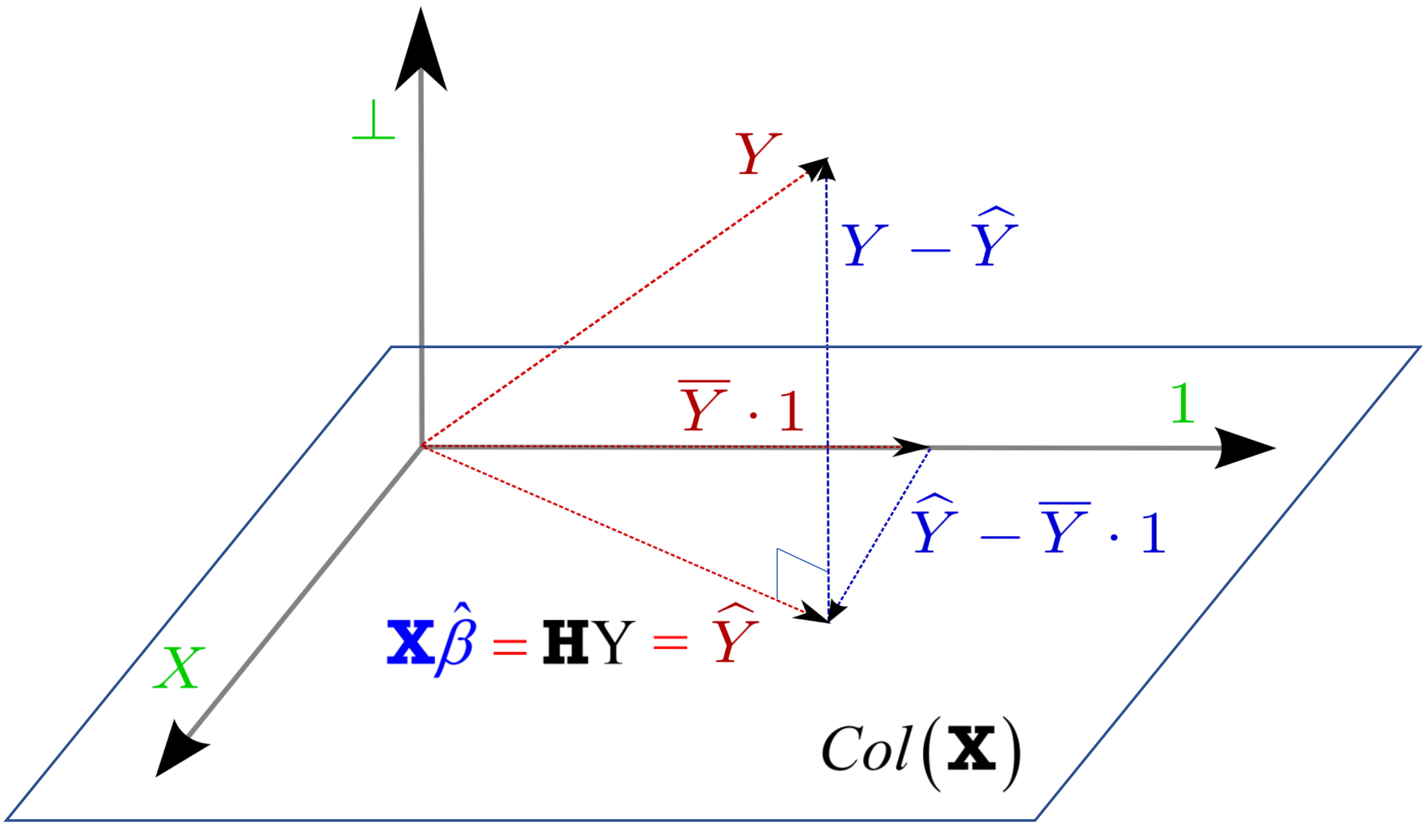
From the equation $\mathrm{Y} = \hat{\mathrm{Y}} + \mathrm{e}$, the vector of residuals becomes $\mathrm{e} = \mathrm{Y} - \hat{\mathrm{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} - \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix}$. Using the hat matrix, this can be rewritten $\mathrm{e} = \mathrm{Y} - \mathbf{H}\mathrm{Y} = \left(\mathbf{I} - \mathbf{H}\right)\mathrm{Y}$ .We can use this representation of the vector of residuals to prove a number of results in regression. For example,

The residual vector is orthogonal to the vector of fitted values. *Proof*,

$\mathrm{e}^T\hat{\mathbf{Y}} = \left(\left(\mathbf{I} - \mathbf{H}\right)\mathrm{Y}\right)^T\mathbf{H}\mathrm{Y} = \mathrm{Y}^T\left(\mathbf{I} - \mathbf{H}\right)\mathbf{H}\mathrm{Y} = \mathrm{Y}^T\left(\mathbf{H} - \mathbf{H}\right)\mathrm{Y} = 0$, where the properties of $\mathbf{H}$ and $\left(\mathbf{I} - \mathbf{H}\right)$ were used.

The Hat Matrix and the Analysis of Variance Table

First, consider the picture below.



The $n \times 1$ vectors Y, $\hat{\mathrm{Y}}$, and e are vectors in the vector space $\mathbb{R}^n$, (an *n*-dimensional space whose components are real numbers, as opposed to complex numbers). From Lecture 24, we know that the vectors $\hat{\mathrm{Y}}$ and e sum to Y, but we can show a much more surprising result that will eventually lead to powerful conclusions about the three sums of squares *SSE*, *SSR*, and *SST* that appear in the Analysis of Variance table in regression output.

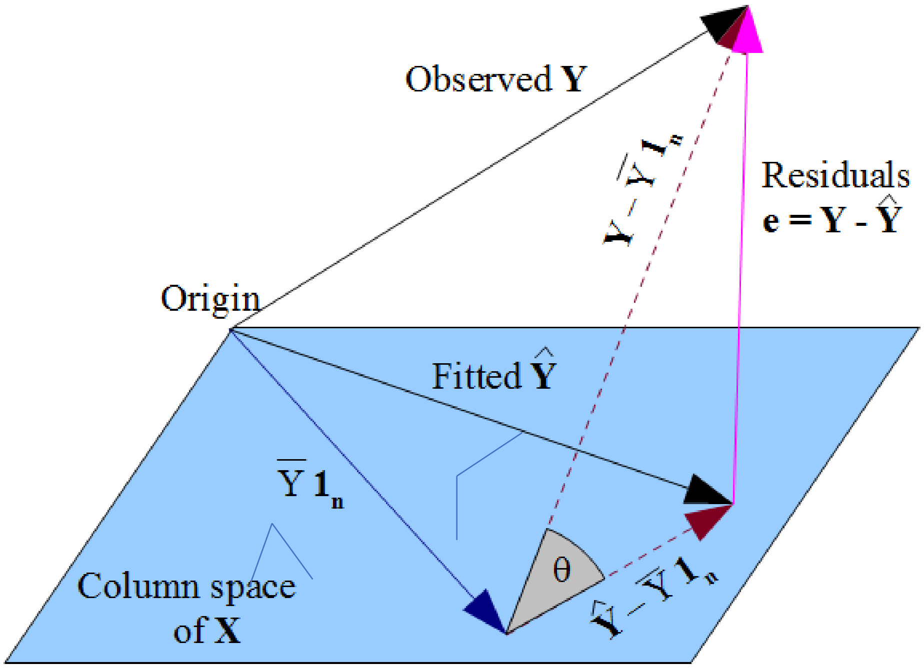
First, we have to know where each of the vectors Y, $\hat{\mathrm{Y}}$, and e "live":

- The *n*-vector of observations Y has no restrictions placed on it and therefore can lie anywhere in $\mathbb{R}^n$.

- $\hat{\mathrm{Y}} = \mathbf{X}\hat{\beta} = \hat{\beta}_0 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \beta_1 \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$ is restricted to the two-dimensional subspace of $\mathbb{R}^n$ spanned by the columns of the design matrix $\mathbf{X}$, called (appropriately) the **column space** of $\mathbf{X}$, or $Col(\mathbf{X})$.

- We saw in the video portion of Lecture 24 that proving $\sum e_i = 0$ and $\sum X_i e_i = 0$ is equivalent to showing that the residual vector e is orthogonal to the column vectors of the design matrix, $\mathsf{1}$ and X, respectively. This implies that e lives in a subspace of $\mathbb{R}^n$ **orthogonal** to the column space of $\mathbf{X}$.

Next, to develop a relationship between the three sums of squares *SSE*, *SSR*, and *SST* in regression, we need a picture with two new vectors, $\mathrm{Y} - \bar{\mathrm{Y}}$, and $\hat{\mathrm{Y}} - \bar{\mathrm{Y}}$, shown in the picture below. First, note that $\hat{\mathrm{Y}} - \bar{\mathrm{Y}}$ lies in the column space of the design matrix (its plane). Thus, the residual vector e must be orthogonal to $\hat{\mathrm{Y}} - \bar{\mathrm{Y}}$, and the triangle facing us with sides $\mathrm{Y} - \bar{\mathrm{Y}}$, $\hat{\mathrm{Y}} - \bar{\mathrm{Y}}$, and e is a right-triangle!



By the Pythagorean Theorem, $\left\|\hat{\mathrm{Y}} - \bar{\mathrm{Y}}\right\|^2 + \left\|\mathrm{e}\right\|^2 = \left\|\mathrm{Y} - \bar{\mathrm{Y}}\right\|^2$, where the squared vector lengths are evaluated as $\left(\hat{\mathrm{Y}} - \bar{\mathrm{Y}}\right)^T\left(\hat{\mathrm{Y}} - \bar{\mathrm{Y}}\right) + \mathrm{e}^T\mathrm{e} = \left(\mathrm{Y} - \bar{\mathrm{Y}}\right)^T\left(\mathrm{Y} - \bar{\mathrm{Y}}\right)$, and the sum is more recognizable as $SSR + SSE = SST$.

Lastly, the dimensions of the subspaces the three vectors "live" in determines their "degrees of freedom," so we've also shown that *SSE* has *n* - 2 degrees of freedom because the vector of residuals **e** is restricted to an *n* - 2 dimensional subspace of $\mathbb{R}^n$. (The term "degrees of freedom" is actually quite descriptive because the vector of residuals, e, is only "free" to assume values in this *n* - 2 dimensional subspace.)

The relationships in simple regression that we've derived from the vector-space approach are summarized below.

The Analysis of Variance (ANOVA) Table

Source	Degrees of Freedom (df)	Sums of Squares (SS)	Mean Square (MS) = SS/df
Regression Model	1	$SSR = \sum\left(\hat{Y}_i - \bar{Y}\right)^2 = \left(\hat{\mathbf{Y}} - \bar{\mathbf{Y}}\right)^{\mathbf{T}}\left(\hat{\mathbf{Y}} - \bar{\mathbf{Y}}\right)$	$MSR = \sum\left(\hat{Y}_i - \bar{Y}\right)^2$
Residual Error	*n* - 2	$SSE = \sum\left(Y_i - \hat{Y}_i\right)^2 = \mathbf{e}^{\mathbf{T}}\mathbf{e}$	$MSE = \dfrac{\sum\left(Y_i - \hat{Y}_i\right)^2}{n-2}$
Total	*n* - 1	$SST = \sum\left(Y_i - \bar{Y}\right)^2 = \left(\mathbf{Y} - \bar{\mathbf{Y}}\right)^{\mathbf{T}}\left(\mathbf{Y} - \bar{\mathbf{Y}}\right)$	

The Hat Matrix and Leverage

One of the less obvious uses of the hat matrix is the following: The diagonal elements of the hat matrix are the leverage values for the observations. (Because the hat matrix is not a function of the observations on the response variable ***Y***, the leverage of an observation depends only on the value of X chosen.) In the notes for lectures 9 & 10, the leverage of the i^{th} observation was given as $h_i = \frac{1}{n} + \frac{\left(X_i - \bar{X}\right)^2}{\sum\left(X_j - \bar{X}\right)^2}$.

Actually, the proper notation should have been $h_{ii} = \frac{1}{n} + \frac{\left(X_i - \bar{X}\right)^2}{\sum\left(X_j - \bar{X}\right)^2}$ to indicate that the leverage h_{ii} is the i^{th} diagonal element (i^{th} row, i^{th} column) of the hat matrix $\mathbf{H}$.

Example: For our stock data set

x	1	2	4	5
y	8	4	6	2

$$\mathbf{H} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \\ 1 & 5 \end{bmatrix}\left(\frac{1}{20}\begin{bmatrix} 23 & -6 \\ -6 & 2 \end{bmatrix}\right)\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 5 \end{bmatrix} = \frac{1}{20}\begin{bmatrix} 13 & 9 & 1 & -3 \\ 9 & 7 & 3 & 1 \\ 1 & 3 & 7 & 9 \\ -3 & 1 & 9 & 13 \end{bmatrix}$$

, so, reading off the diagonal elements of $\mathbf{H}$, the leverages of the four observations are

$$\begin{aligned} h_{11} &= 13/20 \\ h_{22} &= 7/20 \\ h_{33} &= 7/20 \\ h_{44} &= 13/20 \end{aligned}.$$

Below is Statgraphics output of a regression that includes the leverages:

Predictor X	Response Y	Leverages
1	8	0.65
2	4	0.35
4	6	0.35
5	2	0.65