Lecture 6

September 10, 2018

Regression Terminology

- In this lecture and the next, we shall go over some common regression technology:

Regression Terminology

- In this lecture and the next, we shall go over some common regression technology:
  1. Regression plane

Regression Terminology

- In this lecture and the next, we shall go over some common regression technology:
  1. Regression plane
  2. Fitted values and the hat matrix

Regression Terminology

- In this lecture and the next, we shall go over some common regression technology:
  1. Regression plane
  2. Fitted values and the hat matrix
  3. Residuals

Regression Terminology

- In this lecture and the next, we shall go over some common regression technology:
  1. Regression plane
  2. Fitted values and the hat matrix
  3. Residuals
  4. Leverages

Regression Terminology

- In this lecture and the next, we shall go over some common regression technology:
  1. Regression plane
  2. Fitted values and the hat matrix
  3. Residuals
  4. Leverages
  5. Residual Sum of Squares (RSS) and R^2

Regression Terminology

- In this lecture and the next, we shall go over some common regression technology:
  1. Regression plane
  2. Fitted values and the hat matrix
  3. Residuals
  4. Leverages
  5. Residual Sum of Squares (RSS) and R^2
  6. Residual Standard Error

Regression Terminology

- In this lecture and the next, we shall go over some common regression technology:
  1. Regression plane
  2. Fitted values and the hat matrix
  3. Residuals
  4. Leverages
  5. Residual Sum of Squares (RSS) and R^2
  6. Residual Standard Error
  7. Standardized or Studentized Residuals

The Regression Plane

- If we get a new subject whose explanatory variable values are $x_1, \ldots, x_p$, then our prediction for its response variable value is
$$y = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_p x_{ip}.$$

The Regression Plane

- If we get a new subject whose explanatory variable values are $x_1, \ldots, x_p$, then our prediction for its response variable value is
$$y = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_p x_{ip}.$$

- This equation represents a plane which we call the regression plane.

Fitted Values and the Hat Matrix

- These are the values predicted by the linear model for the n subjects.

Fitted Values and the Hat Matrix

- These are the values predicted by the linear model for the n subjects.
- The values of the explanatory variables are $x_{i1}, \ldots, x_{ip}$ for the ith subject.

Fitted Values and the Hat Matrix

- These are the values predicted by the linear model for the n subjects.
- The values of the explanatory variables are $x_{i1}, \ldots, x_{ip}$ for the ith subject.
- Thus the linear model prediction for the ith subject is

$$\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_p x_{ip}.$$

Fitted Values and the Hat Matrix

- These are the values predicted by the linear model for the n subjects.
- The values of the explanatory variables are $x_{i1}, \ldots, x_{ip}$ for the ith subject.
- Thus the linear model prediction for the ith subject is

$$\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_p x_{ip}.$$

- Because the value of the response variable for the ith subject is y_i, it makes sense to call the above prediction $\hat{y}_i$. Thus

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_p x_{ip}, \quad \text{for } i = 1, \ldots, p.$$

Fitted Values and the Hat Matrix

- These values $\hat{y}_1, \ldots, \hat{y}_n$ are called fitted values and the vector $\hat{Y} = (\hat{y}_1, \ldots, \hat{y}_n)^T$ is called the vector of fitted values.

Fitted Values and the Hat Matrix

- These values $\hat{y}_1, \ldots, \hat{y}_n$ are called fitted values and the vector $\hat{Y} = (\hat{y}_1, \ldots, \hat{y}_n)^T$ is called the vector of fitted values.
- This vector can be written succinctly as $\hat{Y} = X\hat{\beta}$.

Fitted Values and the Hat Matrix

- These values $\hat{y}_1, \ldots, \hat{y}_n$ are called fitted values and the vector $\hat{Y} = (\hat{y}_1, \ldots, \hat{y}_n)^T$ is called the vector of fitted values.
- This vector can be written succinctly as $\hat{Y} = X\hat{\beta}$.
- Because $\hat{Y} = X(X^TX)^{-1}X^TY$, we can write

$$\hat{Y} = X(X^TX)^{-1}X^TY$$

Fitted Values and the Hat Matrix

- The vector of fitted values $\hat{Y}$ is the (orthogonal) projection of Y onto the column space of X.

Fitted Values and the Hat Matrix

- The vector of fitted values $\hat{Y}$ is the (orthogonal) projection of Y onto the column space of X.
- Let $H = X(X^T X)^{-1} X^T$ so that $\hat{Y} = HY$.

Fitted Values and the Hat Matrix

- The vector of fitted values $\hat{Y}$ is the (orthogonal) projection of Y onto the column space of X.
- Let $H = X(X^TX)^{-1}X^T$ so that $\hat{Y} = HY$.
- Because multiplication by H changes Y into $\hat{Y}$, the matrix H is called the Hat Matrix. It is very important in linear regression. It has the following three easily verifiable properties:

Fitted Values and the Hat Matrix

- The vector of fitted values $\hat{Y}$ is the (orthogonal) projection of Y onto the column space of X.
- Let $H = X(X^T X)^{-1} X^T$ so that $\hat{Y} = HY$.
- Because multiplication by H changes Y into $\hat{Y}$, the matrix H is called the Hat Matrix. It is very important in linear regression. It has the following three easily verifiable properties:
    1. It is a symmetric $n \times n$ matrix.

Fitted Values and the Hat Matrix

- The vector of fitted values $\hat{Y}$ is the (orthogonal) projection of Y onto the column space of X.
- Let $H = X(X^TX)^{-1}X^T$ so that $\hat{Y} = HY$.
- Because multiplication by H changes Y into $\hat{Y}$, the matrix H is called the Hat Matrix. It is very important in linear regression. It has the following three easily verifiable properties:
  1. It is a symmetric $n \times n$ matrix.
  2. It is idempotent i.e., $H^2 = H$.

Fitted Values and the Hat Matrix

- The vector of fitted values $\hat{Y}$ is the (orthogonal) projection of Y onto the column space of X.
- Let $H = X(X^TX)^{-1}X^T$ so that $\hat{Y} = HY$.
- Because multiplication by H changes Y into $\hat{Y}$, the matrix H is called the Hat Matrix. It is very important in linear regression. It has the following three easily verifiable properties:
  1. It is a symmetric $n \times n$ matrix.
  2. It is idempotent i.e., $H^2 = H$.
  3. $HX = X$.

Fitted Values and the Hat Matrix

- The vector of fitted values $\hat{Y}$ is the (orthogonal) projection of Y onto the column space of X.
- Let $H = X(X^T X)^{-1} X^T$ so that $\hat{Y} = HY$.
- Because multiplication by H changes Y into $\hat{Y}$, the matrix H is called the Hat Matrix. It is very important in linear regression. It has the following three easily verifiable properties:
    1. It is a symmetric $n \times n$ matrix.
    2. It is idempotent i.e., $H^2 = H$.
    3. $HX = X$.
- These can be easily derived from the definition $H = X(X^T X)^{-1} X^T$.

Fitted Values and the Hat Matrix

- The vector of fitted values $\hat{Y}$ is the (orthogonal) projection of Y onto the column space of X.
- Let $H = X(X^TX)^{-1}X^T$ so that $\hat{Y} = HY$.
- Because multiplication by H changes Y into $\hat{Y}$, the matrix H is called the Hat Matrix. It is very important in linear regression. It has the following three easily verifiable properties:
  1. It is a symmetric $n \times n$ matrix.
  2. It is idempotent i.e., $H^2 = H$.
  3. $HX = X$.
- These can be easily derived from the definition $H = X(X^TX)^{-1}X^T$.
- Because of these, we get

Fitted Values and the Hat Matrix

- The vector of fitted values $\hat{Y}$ is the (orthogonal) projection of Y onto the column space of X.
- Let $H = X(X^TX)^{-1}X^T$ so that $\hat{Y} = HY$.
- Because multiplication by H changes Y into $\hat{Y}$, the matrix H is called the Hat Matrix. It is very important in linear regression. It has the following three easily verifiable properties:
  1. It is a symmetric $n \times n$ matrix.
  2. It is idempotent i.e., $H^2 = H$.
  3. $HX = X$.
- These can be easily derived from the definition $H = X(X^TX)^{-1}X^T$.
- Because of these, we get

$$\mathbb{E}(\hat{Y}) =$$

Fitted Values and the Hat Matrix

- The vector of fitted values $\hat{Y}$ is the (orthogonal) projection of Y onto the column space of X.
- Let $H = X(X^T X)^{-1} X^T$ so that $\hat{Y} = HY$.
- Because multiplication by H changes Y into $\hat{Y}$, the matrix H is called the Hat Matrix. It is very important in linear regression. It has the following three easily verifiable properties:
  1. It is a symmetric $n \times n$ matrix.
  2. It is idempotent i.e., $H^2 = H$.
  3. $HX = X$.
- These can be easily derived from the definition $H = X(X^T X)^{-1} X^T$.
- Because of these, we get

$$\mathbb{E}(\hat{Y}) = \mathbb{E}(HY) =$$

Fitted Values and the Hat Matrix

- The vector of fitted values $\hat{Y}$ is the (orthogonal) projection of Y onto the column space of X.
- Let $H = X(X^TX)^{-1}X^T$ so that $\hat{Y} = HY$.
- Because multiplication by H changes Y into $\hat{Y}$, the matrix H is called the Hat Matrix. It is very important in linear regression. It has the following three easily verifiable properties:
  1. It is a symmetric $n \times n$ matrix.
  2. It is idempotent i.e., $H^2 = H$.
  3. $HX = X$.
- These can be easily derived from the definition $H = X(X^TX)^{-1}X^T$.
- Because of these, we get

$$\mathbb{E}(\hat{Y}) = \mathbb{E}(HY) = H\mathbb{E}(Y) = HX\beta =$$

Fitted Values and the Hat Matrix

- The vector of fitted values $\hat{Y}$ is the (orthogonal) projection of Y onto the column space of X.
- Let $H = X(X^T X)^{-1} X^T$ so that $\hat{Y} = HY$.
- Because multiplication by H changes Y into $\hat{Y}$, the matrix H is called the Hat Matrix. It is very important in linear regression. It has the following three easily verifiable properties:
  1. It is a symmetric $n \times n$ matrix.
  2. It is idempotent i.e., $H^2 = H$.
  3. $HX = X$.
- These can be easily derived from the definition $H = X(X^T X)^{-1} X^T$.
- Because of these, we get

$$\mathbb{E}(\hat{Y}) = \mathbb{E}(HY) = H\mathbb{E}(Y) = HX\beta =$$

Fitted Values and the Hat Matrix

- The vector of fitted values $\hat{Y}$ is the (orthogonal) projection of Y onto the column space of X.
- Let $H = X(X^T X)^{-1} X^T$ so that $\hat{Y} = HY$.
- Because multiplication by H changes Y into $\hat{Y}$, the matrix H is called the Hat Matrix. It is very important in linear regression. It has the following three easily verifiable properties:
  1. It is a symmetric $n \times n$ matrix.
  2. It is idempotent i.e., $H^2 = H$.
  3. $HX = X$.
- These can be easily derived from the definition $H = X(X^T X)^{-1} X^T$.
- Because of these, we get

$$\mathbb{E}(\hat{Y}) = \mathbb{E}(HY) = H\mathbb{E}(Y) = HX\beta = X\beta =$$

Fitted Values and the Hat Matrix

- The vector of fitted values $\hat{Y}$ is the (orthogonal) projection of Y onto the column space of X.
- Let $H = X(X^T X)^{-1} X^T$ so that $\hat{Y} = HY$.
- Because multiplication by H changes Y into $\hat{Y}$, the matrix H is called the Hat Matrix. It is very important in linear regression. It has the following three easily verifiable properties:
  1. It is a symmetric $n \times n$ matrix.
  2. It is idempotent i.e., $H^2 = H$.
  3. $HX = X$.
- These can be easily derived from the definition $H = X(X^T X)^{-1} X^T$.
- Because of these, we get

$$\mathbb{E}(\hat{Y}) = \mathbb{E}(HY) = H\mathbb{E}(Y) = HX\beta = X\beta = \mathbb{E}(Y).$$

Fitted Values and the Hat Matrix

- Thus $\hat{Y}$ and Y have the same expectation.

Fitted Values and the Hat Matrix

- Thus $\hat{Y}$ and Y have the same expectation.
- Also

Fitted Values and the Hat Matrix

- Thus $\hat{Y}$ and Y have the same expectation.
- Also

$$Cov(\hat{Y}) = HCov(Y)H^T =$$

Fitted Values and the Hat Matrix

- Thus $\hat{Y}$ and Y have the same expectation.
- Also

$$Cov(\hat{Y}) = HCov(Y)H^T = H(\sigma^2 I_n)H =$$

Fitted Values and the Hat Matrix

- Thus $\hat{Y}$ and Y have the same expectation.
- Also

$$Cov(\hat{Y}) = HCov(Y)H^T = H(\sigma^2 I_n)H = \sigma^2 H.$$

Residuals

- The difference between y_i and $\hat{y}_i$ is called the residual for the ith subject: $\hat{e}_i := y_i - \hat{y}_i$.

Residuals

- The difference between y_i and $\hat{y}_i$ is called the residual for the ith subject: $\hat{e}_i := y_i - \hat{y}_i$.
- The vector $\hat{e} = (\hat{e}_1, \ldots \hat{e}_n)^T$ is called the vector of residuals

Residuals

- The difference between y_i and $\hat{y}_i$ is called the residual for the ith subject: $\hat{e}_i := y_i - \hat{y}_i$.
- The vector $\hat{e} = (\hat{e}_1, \ldots \hat{e}_n)^T$ is called the vector of residuals

Residuals

- The difference between y_i and $\hat{y}_i$ is called the residual for the ith subject: $\hat{e}_i := y_i - \hat{y}_i$.
- The vector $\hat{e} = (\hat{e}_1, \ldots \hat{e}_n)^T$ is called the vector of residuals

$$\hat{e} = \hat{Y} - Y$$

Residuals

- The difference between y_i and $\hat{y}_i$ is called the residual for the ith subject: $\hat{e}_i := y_i - \hat{y}_i$.
- The vector $\hat{e} = (\hat{e}_1, \ldots \hat{e}_n)^T$ is called the vector of residuals

$$\hat{e} = Y - \textit{Y-hat} = (I - H)Y.$$

Residuals

- The difference between y_i and $\hat{y}_i$ is called the residual for the ith subject: $\hat{e}_i := y_i - \hat{y}_i$.
- The vector $\hat{e} = (\hat{e}_1, \ldots \hat{e}_n)^T$ is called the vector of residuals

$$\hat{e} \;=\; \hat{Y} - Y \;=\; (I - H)Y.$$

- The vector of residuals $\hat{e}$ acts as a proxy for the unobserved error vector e.

Residuals

- The difference between y_i and $\hat{y}_i$ is called the residual for the ith subject: $\hat{e}_i := y_i - \hat{y}_i$.
- The vector $\hat{e} = (\hat{e}_1, \ldots \hat{e}_n)^T$ is called the vector of residuals

$$\hat{e} \;=\; \hat{Y} - Y \;=\; (I - H)Y.$$

- The vector of residuals $\hat{e}$ acts as a proxy for the unobserved error vector e.
- The most important fact about the residuals in the linear model is that they are orthogonal to the column space of X.

Residuals

- The difference between y_i and $\hat{y}_i$ is called the residual for the ith subject: $\hat{e}_i := y_i - \hat{y}_i$.
- The vector $\hat{e} = (\hat{e}_1, \ldots \hat{e}_n)^T$ is called the vector of residuals

$$\hat{e} = \hat{Y} - Y = (I - H)Y.$$

- The vector of residuals $\hat{e}$ acts as a proxy for the unobserved error vector e.
- The most important fact about the residuals in the linear model is that they are orthogonal to the column space of X.
- This happens because $HX = X$ so that

Residuals

- The difference between y_i and $\hat{y}_i$ is called the residual for the ith subject: $\hat{e}_i := y_i - \hat{y}_i$.
- The vector $\hat{e} = (\hat{e}_1, \ldots \hat{e}_n)^T$ is called the vector of residuals

$$\hat{e} = \hat{Y} - Y = (I - H)Y.$$

- The vector of residuals $\hat{e}$ acts as a proxy for the unobserved error vector e.
- The most important fact about the residuals in the linear model is that they are orthogonal to the column space of X.
- This happens because $HX = X$ so that

$$\hat{e}^T X =$$

Residuals

- The difference between y_i and $\hat{y}_i$ is called the residual for the ith subject: $\hat{e}_i := y_i - \hat{y}_i$.
- The vector $\hat{e} = (\hat{e}_1, \dots \hat{e}_n)^T$ is called the vector of residuals

$$\hat{e} = \hat{Y} - Y = (I - H)Y.$$

- The vector of residuals $\hat{e}$ acts as a proxy for the unobserved error vector e.
- The most important fact about the residuals in the linear model is that they are orthogonal to the column space of X.
- This happens because $HX = X$ so that

$$\hat{e}^T X = ((I - H)Y)^T X =$$

Residuals

- The difference between y_i and $\hat{y}_i$ is called the residual for the ith subject: $\hat{e}_i := y_i - \hat{y}_i$.
- The vector $\hat{e} = (\hat{e}_1, \ldots \hat{e}_n)^T$ is called the vector of residuals

$$\hat{e} = \hat{Y} - Y = (I - H)Y.$$

- The vector of residuals $\hat{e}$ acts as a proxy for the unobserved error vector e.
- The most important fact about the residuals in the linear model is that they are orthogonal to the column space of X.
- This happens because $HX = X$ so that

$$\hat{e}^T X = ((I - H)Y)^T X = Y^T (I - H) X =$$

Residuals

- The difference between y_i and $\hat{y}_i$ is called the residual for the ith subject: $\hat{e}_i := y_i - \hat{y}_i$.
- The vector $\hat{e} = (\hat{e}_1, \ldots \hat{e}_n)^T$ is called the vector of residuals

$$\hat{e} = \hat{Y} - Y = (I - H)Y.$$

- The vector of residuals $\hat{e}$ acts as a proxy for the unobserved error vector e.
- The most important fact about the residuals in the linear model is that they are orthogonal to the column space of X.
- This happens because $HX = X$ so that

$$\hat{e}^T X = ((I-H)Y)^T X = Y^T(I-H)X = Y^T(X - HX) = 0.$$

Residuals

- The difference between y_i and $\hat{y}_i$ is called the residual for the ith subject: $\hat{e}_i := y_i - \hat{y}_i$.
- The vector $\hat{e} = (\hat{e}_1, \ldots \hat{e}_n)^T$ is called the vector of residuals

$$\hat{e} = \hat{Y} - Y = (I - H)Y.$$

- The vector of residuals $\hat{e}$ acts as a proxy for the unobserved error vector e.
- The most important fact about the residuals in the linear model is that they are orthogonal to the column space of X.
- This happens because $HX = X$ so that

$$\hat{e}^T X = ((I - H)Y)^T X = Y^T (I - H) X = Y^T (X - HX) = 0.$$

- As a result $\hat{e}^T Xu = 0$ for every vector u which means that $\hat{e}$ is orthogonal to the column space of X.

Residuals

- The first column of X consists of ones.

Residuals

- The first column of X consists of ones.
- Because $\hat{e}$ is orthogonal to everything in the column space of X, it must therefore be orthogonal to the vector of ones which means that
$$\sum_{i=1}^{n} \hat{e}_i = 0.$$

Residuals

- The first column of X consists of ones.
- Because $\hat{e}$ is orthogonal to everything in the column space of X, it must therefore be orthogonal to the vector of ones which means that

$$\sum_{i=1}^{n} \hat{e}_i = 0.$$

- $\hat{e}$ is also orthogonal to every column of X:

$$\sum_{i=1}^{n} \hat{e}_i x_{ij} = 0 \quad \text{for every} \quad j.$$

Residuals

- The vector of fitted values belongs to the column space of X because $\hat{Y} = X\hat{\beta}$.

Residuals

- The vector of fitted values belongs to the column space of X because $\hat{Y} = X\hat{\beta}$.
- Therefore $\hat{e}$ is also orthogonal to $\hat{Y}$.

Residuals

- The vector of fitted values belongs to the column space of X because $\hat{Y} = X\hat{\beta}$.
- Therefore $\hat{e}$ is also orthogonal to $\hat{Y}$.
- Because $X^T\hat{e} = 0$, the residuals satisfy $p + 1$ linear equalities.

Residuals

- The vector of fitted values belongs to the column space of X because $\hat{Y} = X\hat{\beta}$.
- Therefore $\hat{e}$ is also orthogonal to $\hat{Y}$.
- Because $X^T\hat{e} = 0$, the residuals satisfy $p + 1$ linear equalities.
- Hence, although there are n residuals, there are effectively $n - p - 1$ of them. The number $n - p - 1$ is therefore referred to as the degrees of freedom of the residuals $\hat{e}_1, \ldots, \hat{e}_n$.

Residuals

- The vector of fitted values belongs to the column space of X because $\hat{Y} = X\hat{\beta}$.
- Therefore $\hat{e}$ is also orthogonal to $\hat{Y}$.
- Because $X^T\hat{e} = 0$, the residuals satisfy $p + 1$ linear equalities.
- Hence, although there are n residuals, there are effectively $n - p - 1$ of them. The number $n - p - 1$ is therefore referred to as the degrees of freedom of the residuals $\hat{e}_1, \ldots, \hat{e}_n$.
- The expectation of $\hat{e}$ is

Residuals

- The vector of fitted values belongs to the column space of X because $\hat{Y} = X\hat{\beta}$.
- Therefore $\hat{e}$ is also orthogonal to $\hat{Y}$.
- Because $X^T\hat{e} = 0$, the residuals satisfy $p + 1$ linear equalities.
- Hence, although there are n residuals, there are effectively $n - p - 1$ of them. The number $n - p - 1$ is therefore referred to as the degrees of freedom of the residuals $\hat{e}_1, \ldots, \hat{e}_n$.
- The expectation of $\hat{e}$ is

$$\mathbb{E}(\hat{e}) = \mathbb{E}((I - H)Y) =$$

Residuals

- The vector of fitted values belongs to the column space of X because $\hat{Y} = X\hat{\beta}$.
- Therefore $\hat{e}$ is also orthogonal to $\hat{Y}$.
- Because $X^T\hat{e} = 0$, the residuals satisfy $p + 1$ linear equalities.
- Hence, although there are n residuals, there are effectively $n - p - 1$ of them. The number $n - p - 1$ is therefore referred to as the degrees of freedom of the residuals $\hat{e}_1, \ldots, \hat{e}_n$.
- The expectation of $\hat{e}$ is

$$\mathbb{E}(\hat{e}) = \mathbb{E}((I - H)Y) = (I - H)\mathbb{E}(Y) =$$

Residuals

- The vector of fitted values belongs to the column space of X because $\hat{Y} = X\hat{\beta}$.
- Therefore $\hat{e}$ is also orthogonal to $\hat{Y}$.
- Because $X^T\hat{e} = 0$, the residuals satisfy $p + 1$ linear equalities.
- Hence, although there are n residuals, there are effectively $n - p - 1$ of them. The number $n - p - 1$ is therefore referred to as the degrees of freedom of the residuals $\hat{e}_1, \ldots, \hat{e}_n$.
- The expectation of $\hat{e}$ is

$$\mathbb{E}(\hat{e}) = \mathbb{E}((I - H)Y) = (I - H)\mathbb{E}(Y) = (I - H)X\beta = 0.$$

Residuals

- The vector of fitted values belongs to the column space of X because $\hat{Y} = X\hat{\beta}$.
- Therefore $\hat{e}$ is also orthogonal to $\hat{Y}$.
- Because $X^T\hat{e} = 0$, the residuals satisfy $p + 1$ linear equalities.
- Hence, although there are n residuals, there are effectively $n - p - 1$ of them. The number $n - p - 1$ is therefore referred to as the degrees of freedom of the residuals $\hat{e}_1, \ldots, \hat{e}_n$.
- The expectation of $\hat{e}$ is

$$\mathbb{E}(\hat{e}) = \mathbb{E}((I - H)Y) = (I - H)\mathbb{E}(Y) = (I - H)X\beta = 0.$$

- The Covariance matrix of $\hat{e}$ is

Residuals

- The vector of fitted values belongs to the column space of X because $\hat{Y} = X\hat{\beta}$.
- Therefore $\hat{e}$ is also orthogonal to $\hat{Y}$.
- Because $X^T\hat{e} = 0$, the residuals satisfy $p + 1$ linear equalities.
- Hence, although there are n residuals, there are effectively $n - p - 1$ of them. The number $n - p - 1$ is therefore referred to as the degrees of freedom of the residuals $\hat{e}_1, \dots, \hat{e}_n$.
- The expectation of $\hat{e}$ is

$$\mathbb{E}(\hat{e}) = \mathbb{E}((I - H)Y) = (I - H)\mathbb{E}(Y) = (I - H)X\beta = 0.$$

- The Covariance matrix of $\hat{e}$ is

$$Cov(\hat{e}) = Cov((I - H)Y) =$$

Residuals

- The vector of fitted values belongs to the column space of X because $\hat{Y} = X\hat{\beta}$.
- Therefore $\hat{e}$ is also orthogonal to $\hat{Y}$.
- Because $X^T\hat{e} = 0$, the residuals satisfy $p + 1$ linear equalities.
- Hence, although there are n residuals, there are effectively $n - p - 1$ of them. The number $n - p - 1$ is therefore referred to as the degrees of freedom of the residuals $\hat{e}_1, \ldots, \hat{e}_n$.
- The expectation of $\hat{e}$ is

$$\mathbb{E}(\hat{e}) = \mathbb{E}((I - H)Y) = (I - H)\mathbb{E}(Y) = (I - H)X\beta = 0.$$

- The Covariance matrix of $\hat{e}$ is

$$Cov(\hat{e}) = Cov((I-H)Y) = (I-H)Cov(Y)(I-H)^T =$$

Residuals

- The vector of fitted values belongs to the column space of X because $\hat{Y} = X\hat{\beta}$.
- Therefore $\hat{e}$ is also orthogonal to $\hat{Y}$.
- Because $X^T\hat{e} = 0$, the residuals satisfy $p + 1$ linear equalities.
- Hence, although there are n residuals, there are effectively $n - p - 1$ of them. The number $n - p - 1$ is therefore referred to as the degrees of freedom of the residuals $\hat{e}_1, \ldots, \hat{e}_n$.
- The expectation of $\hat{e}$ is

$$\mathbb{E}(\hat{e}) = \mathbb{E}((I - H)Y) = (I - H)\mathbb{E}(Y) = (I - H)X\beta = 0.$$

- The Covariance matrix of $\hat{e}$ is

$$Cov(\hat{e}) = Cov((I-H)Y) = (I-H)Cov(Y)(I-H)^T = \sigma^2(I-H).$$

Residuals

- The vector of fitted values belongs to the column space of X because $\hat{Y} = X\hat{\beta}$.
- Therefore $\hat{e}$ is also orthogonal to $\hat{Y}$.
- Because $X^T\hat{e} = 0$, the residuals satisfy $p + 1$ linear equalities.
- Hence, although there are n residuals, there are effectively $n - p - 1$ of them. The number $n - p - 1$ is therefore referred to as the degrees of freedom of the residuals $\hat{e}_1, \dots, \hat{e}_n$.
- The expectation of $\hat{e}$ is

$$\mathbb{E}(\hat{e}) = \mathbb{E}((I - H)Y) = (I - H)\mathbb{E}(Y) = (I - H)X\beta = 0.$$

- The Covariance matrix of $\hat{e}$ is

$$Cov(\hat{e}) = Cov((I - H)Y) = (I - H)Cov(Y)(I - H)^T = \sigma^2(I - H).$$

- Note that the residuals have different variances.

Residuals

- Note that the residuals have different variances. (even though $e_1, \ldots, e_n$ are assumed to have the same variance σ^2).

Levarages

- The diagonal entries of the hat matrix are called leverages and these are denoted by $h_{ii}, i = 1, \ldots, n$.

Levarages

- The diagonal entries of the hat matrix are called leverages and these are denoted by $h_{ii}, i = 1, \ldots, n$.
- We will study them carefully later on under the topic of regression diagnostics. For now, let us just remark that the leverages are always between 0 and 1.

Levarages

- The diagonal entries of the hat matrix are called leverages and these are denoted by $h_{ii}, i = 1, \ldots, n$.
- We will study them carefully later on under the topic of regression diagnostics. For now, let us just remark that the leverages are always between 0 and 1.
- This can be proved in many ways but a simple way is to observe that $h_{ii} = var(y_i) \geq 0$, $1 - h_{ii} = var(\hat{e}_i) \geq 0$.