Lecture 7

September 13, 2018

Regression Terminology

Today we shall cover

Regression Terminology

Today we shall cover

- Residual Sum of Squares (RSS) and R^2.

Regression Terminology

Today we shall cover

- Residual Sum of Squares (RSS) and R^2.
- Residual Standard Error

Regression Terminology

Today we shall cover

- Residual Sum of Squares (RSS) and R^2.
- Residual Standard Error
- Standardized or Studentized Residuals

The Residual Sum of Squares and R^2

- The sum of squares of the residuals is called RSS:

The Residual Sum of Squares and R^2

- The sum of squares of the residuals is called RSS:

$$RSS = \sum_{i=1}^{n} \hat{e}_i^2 =$$

The Residual Sum of Squares and R^2

- The sum of squares of the residuals is called RSS:

$$RSS = \sum_{i=1}^{n} \hat{e}_i^2 = \hat{e}^T \hat{e} = Y^T (I - H) Y.$$

The Residual Sum of Squares and R^2

- The sum of squares of the residuals is called RSS:

$$RSS = \sum_{i=1}^{n} \hat{e}_i^2 = \hat{e}^T \hat{e} = Y^T(I - H)Y.$$

- The RSS is measuring the discrepancy between y_i and $\hat{y}_i$ for $i = 1, \ldots, n$.

The Residual Sum of Squares and R^2

- The sum of squares of the residuals is called RSS:

$$RSS = \sum_{i=1}^{n} \hat{e}_i^2 = \hat{e}^T \hat{e} = Y^T (I - H) Y.$$

- The RSS is measuring the discrepancy between y_i and $\hat{y}_i$ for $i = 1, \dots, n$.
- It is therefore an in-sample measure of the prediction accuracy of the linear model.

The Residual Sum of Squares and R^2

- The sum of squares of the residuals is called RSS:

$$RSS = \sum_{i=1}^{n} \hat{e}_i^2 = \hat{e}^T \hat{e} = Y^T (I - H) Y.$$

- The RSS is measuring the discrepancy between y_i and $\hat{y}_i$ for $i = 1, \ldots, n$.
- It is therefore an in-sample measure of the prediction accuracy of the linear model.
- In general, RSS decreases (or remains the same) as we add more explanatory variables to the model.

The Residual Sum of Squares and R^2

- One problem with RSS as a measure of the prediction accuracy is that it is not scale-free. When we change the units in which we are measuring the response, the RSS changes.

The Residual Sum of Squares and R^2

- One problem with RSS as a measure of the prediction accuracy is that it is not scale-free. When we change the units in which we are measuring the response, the RSS changes.
- To fix this, one usually looks at a scaled version of RSS called the R^2 (aka multiple R-squared or coefficient of determination).

The Residual Sum of Squares and R^2

- One problem with RSS as a measure of the prediction accuracy is that it is not scale-free. When we change the units in which we are measuring the response, the RSS changes.
- To fix this, one usually looks at a scaled version of RSS called the R^2 (aka multiple R-squared or coefficient of determination).
- What is a reasonable way to scale RSS?

The Residual Sum of Squares and R^2

- One problem with RSS as a measure of the prediction accuracy is that it is not scale-free. When we change the units in which we are measuring the response, the RSS changes.
- To fix this, one usually looks at a scaled version of RSS called the R^2 (aka multiple R-squared or coefficient of determination).
- What is a reasonable way to scale RSS?
- For this, consider the following. Suppose we are told to predict the response of a future subject without using any of the data on the explanatory variables i.e., we are only supposed to use $y_1, \dots, y_n$.

The Residual Sum of Squares and R^2

- One problem with RSS as a measure of the prediction accuracy is that it is not scale-free. When we change the units in which we are measuring the response, the RSS changes.
- To fix this, one usually looks at a scaled version of RSS called the R^2 (aka multiple R-squared or coefficient of determination).
- What is a reasonable way to scale RSS?
- For this, consider the following. Suppose we are told to predict the response of a future subject without using any of the data on the explanatory variables i.e., we are only supposed to use $y_1, \ldots, y_n$.
- In this case, it is obvious that our prediction would be $\bar{y}$.

The Residual Sum of Squares and R^2

- What is the in-sample prediction accuracy of this prediction method?

The Residual Sum of Squares and R^2

- What is the in-sample prediction accuracy of this prediction method?
- The error of this method of prediction on the ith subject is $y_i - \bar{y}_i$ and thus the in-sample prediction accuracy is:

$$\sum_{i=1}^{n} (\bar{y}_i - y_i)^2.$$

The Residual Sum of Squares and R^2

- What is the in-sample prediction accuracy of this prediction method?
- The error of this method of prediction on the ith subject is $y_i - \bar{y}_i$ and thus the in-sample prediction accuracy is:

$$\sum_{i=1}^{n} (\bar{y}_i - y_i)^2.$$

- This quantity is called Total Sum of Squares (TSS):

$$TSS = \sum_{i=1}^{n} (\bar{y}_i - y_i)^2.$$

The Residual Sum of Squares and R^2

- What is the in-sample prediction accuracy of this prediction method?
- The error of this method of prediction on the ith subject is $y_i - \bar{y}_i$ and thus the in-sample prediction accuracy is:

$$\sum_{i=1}^{n} (\bar{y}_i - y_i)^2.$$

- This quantity is called Total Sum of Squares (TSS):

$$TSS = \sum_{i=1}^{n} (\bar{y}_i - y_i)^2.$$

- On the other hand, the RSS is the in-sample prediction accuracy of the linear model which uses the data on the explanatory variables.

The Residual Sum of Squares and R^2

- What is the in-sample prediction accuracy of this prediction method?
- The error of this method of prediction on the ith subject is $y_i - \bar{y}_i$ and thus the in-sample prediction accuracy is:

$$\sum_{i=1}^{n} (\bar{y}_i - y_i)^2.$$

- This quantity is called Total Sum of Squares (TSS):

$$TSS = \sum_{i=1}^{n} (\bar{y}_i - y_i)^2.$$

- On the other hand, the RSS is the in-sample prediction accuracy of the linear model which uses the data on the explanatory variables.
- It should therefore be clear that RSS is always smaller than or equal to TSS (this fact is crucially reliant on the fact that there is an intercept in our model).

The Residual Sum of Squares and R^2

- Therefore TSS is a reasonable yardstick for comparing RSS and hence we can scale RSS by TSS to obtain a reasonable scale-free measure of in-sample prediction accuracy.

The Residual Sum of Squares and R^2

- Therefore TSS is a reasonable yardstick for comparing RSS and hence we can scale RSS by TSS to obtain a reasonable scale-free measure of in-sample prediction accuracy.
- This motivates the multiple R^2. Formally, R^2 is defined as

$$R^2 = 1 - \frac{RSS}{TSS}.$$

The Residual Sum of Squares and R^2

- Therefore TSS is a reasonable yardstick for comparing RSS and hence we can scale RSS by TSS to obtain a reasonable scale-free measure of in-sample prediction accuracy.
- This motivates the multiple R^2. Formally, R^2 is defined as

$$R^2 = 1 - \frac{RSS}{TSS}.$$

- R^2 always lies between 0 and 1 and is often reported as a percentage. Note the following:

The Residual Sum of Squares and R^2

- Therefore TSS is a reasonable yardstick for comparing RSS and hence we can scale RSS by TSS to obtain a reasonable scale-free measure of in-sample prediction accuracy.
- This motivates the multiple R^2. Formally, R^2 is defined as

$$R^2 = 1 - \frac{RSS}{TSS}.$$

- R^2 always lies between 0 and 1 and is often reported as a percentage. Note the following:
  1. If R^2 is high, it means that RSS is much smaller compared to TSS and hence the explanatory variables are really useful in predicting the response.

The Residual Sum of Squares and R^2

- Therefore TSS is a reasonable yardstick for comparing RSS and hence we can scale RSS by TSS to obtain a reasonable scale-free measure of in-sample prediction accuracy.
- This motivates the multiple R^2. Formally, R^2 is defined as

$$R^2 = 1 - \frac{RSS}{TSS}.$$

- R^2 always lies between 0 and 1 and is often reported as a percentage. Note the following:
  1. If R^2 is high, it means that RSS is much smaller compared to TSS and hence the explanatory variables are really useful in predicting the response.
  2. If R^2 is low, it means that RSS is only a little bit smaller than TSS and hence the explanatory variables are not useful in predicting the response.

The Residual Sum of Squares and R^2

- Let us emphasize once again that R^2 is an in-sample measure of prediction accuracy.

The Residual Sum of Squares and R^2

- Let us emphasize once again that R^2 is an in-sample measure of prediction accuracy.
- In other words, the predictions are checked on the subjects already present in the sample (as opposed to checking them on new subjects).

The Residual Sum of Squares and R^2

- Let us emphasize once again that R^2 is an in-sample measure of prediction accuracy.
- In other words, the predictions are checked on the subjects already present in the sample (as opposed to checking them on new subjects).
- In particular, these are the same subjects on whom the model is fitted (or trained), so R^2 can be made to look very good by fitting models with lots of parameters.

The Residual Sum of Squares and R^2

- Let us emphasize once again that R^2 is an in-sample measure of prediction accuracy.
- In other words, the predictions are checked on the subjects already present in the sample (as opposed to checking them on new subjects).
- In particular, these are the same subjects on whom the model is fitted (or trained), so R^2 can be made to look very good by fitting models with lots of parameters.
- When more parameters are added to the model, RSS decreases and R^2 increases.

Estimating σ: The Residual Standard Error

- In addition to $\beta_0, \ldots, \beta_p$, the quantity σ is also a parameter in the linear model and it needs to be estimated from data.

Estimating σ: The Residual Standard Error

- In addition to $\beta_0, \ldots, \beta_p$, the quantity σ is also a parameter in the linear model and it needs to be estimated from data.
- Estimating σ is important for assessing the accuracies of $\beta_0, \ldots, \beta_p$.

Estimating σ: The Residual Standard Error

- In addition to $\beta_0, \ldots, \beta_p$, the quantity σ is also a parameter in the linear model and it needs to be estimated from data.
- Estimating σ is important for assessing the accuracies of $\beta_0, \ldots, \beta_p$.
- Indeed, recall that $Cov(\hat{\beta}) = \sigma^2(X^T X)^{-1}$ and hence we need to estimate σ^2 in order to estimate the accuracy of $\hat{\beta}$.

Estimating σ: The Residual Standard Error

- In addition to $\beta_0, \ldots, \beta_p$, the quantity σ is also a parameter in the linear model and it needs to be estimated from data.
- Estimating σ is important for assessing the accuracies of $\beta_0, \ldots, \beta_p$.
- Indeed, recall that $Cov(\hat{\beta}) = \sigma^2(X^T X)^{-1}$ and hence we need to estimate σ^2 in order to estimate the accuracy of $\hat{\beta}$.
- The estimate of σ that we shall use is

$$\hat{\sigma} = \sqrt{\frac{RSS}{Residual.df}} = \sqrt{\frac{RSS}{n-p-1}}.$$

Estimating σ: The Residual Standard Error

- In addition to $\beta_0, \ldots, \beta_p$, the quantity σ is also a parameter in the linear model and it needs to be estimated from data.
- Estimating σ is important for assessing the accuracies of $\beta_0, \ldots, \beta_p$.
- Indeed, recall that $Cov(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$ and hence we need to estimate σ^2 in order to estimate the accuracy of $\hat{\beta}$.
- The estimate of σ that we shall use is

$$\hat{\sigma} = \sqrt{\frac{RSS}{Residual.df}} = \sqrt{\frac{RSS}{n-p-1}}.$$

- Recall that the Residual:df term above equals $n - p - 1$. This $\hat{\sigma}$ is called the Residual Standard Error.

Estimating σ: The Residual Standard Error

- It turns out that this estimator satisfies the following:

$$\mathbb{E}\hat{\sigma}^2 = \sigma^2.$$

Estimating σ: The Residual Standard Error

- It turns out that this estimator satisfies the following:
$$\mathbb{E}\hat{\sigma}^2 = \sigma^2.$$
- i.e, $\hat{\sigma}^2$ is an unbiased estimate of σ^2(this does not mean that $\hat{\sigma}$ is an unbiased estimator of σ).

Estimating σ: The Residual Standard Error

- It turns out that this estimator satisfies the following:

$$\mathbb{E}\hat{\sigma}^2 = \sigma^2.$$

- i.e, $\hat{\sigma}^2$ is an unbiased estimate of σ^2(this does not mean that $\hat{\sigma}$ is an unbiased estimator of σ).
- Let us now prove (2). For this we essentially need to show that

$$\mathbb{E}(RSS) = \sigma^2(n-p-1).$$

Estimating σ: The Residual Standard Error

- It turns out that this estimator satisfies the following:
$$\mathbb{E}\hat{\sigma}^2 = \sigma^2.$$
- i.e, $\hat{\sigma}^2$ is an unbiased estimate of σ^2(this does not mean that $\hat{\sigma}$ is an unbiased estimator of σ).
- Let us now prove (2). For this we essentially need to show that
$$\mathbb{E}(RSS) = \sigma^2(n - p - 1).$$
- To show this, first note that $RSS = Y^T(I - H)Y$.

Estimating σ: The Residual Standard Error

- It turns out that this estimator satisfies the following:
$$\mathbb{E}\hat{\sigma}^2 = \sigma^2.$$
- i.e, $\hat{\sigma}^2$ is an unbiased estimate of σ^2(this does not mean that $\hat{\sigma}$ is an unbiased estimator of σ).
- Let us now prove (2). For this we essentially need to show that
$$\mathbb{E}(RSS) = \sigma^2(n-p-1).$$
- To show this, first note that $RSS = Y^T(I-H)Y$.
- Because $Y = X + e$ and $HX = 0$, it follows that $RSS = e^T(I-H)e$.

Estimating σ: The Residual Standard Error

- It turns out that this estimator satisfies the following:

$$\mathbb{E}\hat{\sigma}^2 = \sigma^2.$$

- i.e, $\hat{\sigma}^2$ is an unbiased estimate of σ^2(this does not mean that $\hat{\sigma}$ is an unbiased estimator of σ).
- Let us now prove (2). For this we essentially need to show that

$$\mathbb{E}(RSS) = \sigma^2(n - p - 1).$$

- To show this, first note that $RSS = Y^T(I - H)Y$.
- Because $Y = X\beta + e$ and $HX = \mathrm{X}$, it follows that $RSS = e^T(I - H)e$.
- Therefore,

Estimating σ: The Residual Standard Error

- It turns out that this estimator satisfies the following:
$$\mathbb{E}\hat{\sigma}^2 = \sigma^2.$$
- i.e, $\hat{\sigma}^2$ is an unbiased estimate of σ^2(this does not mean that $\hat{\sigma}$ is an unbiased estimator of σ).
- Let us now prove (2). For this we essentially need to show that
$$\mathbb{E}(RSS) = \sigma^2(n - p - 1).$$
- To show this, first note that $RSS = Y^T(I - H)Y$.
- Because $Y = X + e$ and $HX = 0$, it follows that $RSS = e^T(I - H)e$.
- Therefore,

$$\mathbb{E}(RSS) = \mathbb{E}(Y^T(I - H)Y) =$$

Estimating σ: The Residual Standard Error

- It turns out that this estimator satisfies the following:
$$\mathbb{E}\hat{\sigma}^2 = \sigma^2.$$
- i.e, $\hat{\sigma}^2$ is an unbiased estimate of σ^2(this does not mean that $\hat{\sigma}$ is an unbiased estimator of σ).
- Let us now prove (2). For this we essentially need to show that
$$\mathbb{E}(RSS) = \sigma^2(n-p-1).$$
- To show this, first note that $RSS = Y^T(I-H)Y$.
- Because $Y = X + e$ and $HX = 0$, it follows that $RSS = e^T(I-H)e$.
- Therefore,
$$\mathbb{E}(RSS) = \mathbb{E}(Y^T(I-H)Y) = \mathbb{E}\left(\sum_{i,j}(I-H)_{i,j}e_ie_j\right) =$$

Estimating σ: The Residual Standard Error

Estimating σ: The Residual Standard Error

- Therefore,

$$\mathbb{E}(RSS) = \sum_{i,j}(I - H)_{i,j}\mathbb{E}\left(e_i e_j\right) =$$

Estimating σ: The Residual Standard Error

- Therefore,

$$\mathbb{E}(RSS) = \sum_{i,j} (I - H)_{i,j} \mathbb{E}\left(e_i e_j\right) =$$

Estimating σ: The Residual Standard Error

- Therefore,

$$\mathbb{E}(RSS) = \sum_{i,j}(I-H)_{i,j}\mathbb{E}\left(e_i e_j\right) = \sum_{i=1}^{n}(I-H)_{i,i}\sigma^2$$

Estimating σ: The Residual Standard Error

- Therefore,

$$\mathbb{E}(RSS) = \sum_{i,j} (I-H)_{i,j} \mathbb{E}\left(e_i e_j\right) = \sum_{i=1}^{n} (I-H)_{i,i} \sigma^2$$

and so

$$\mathbb{E}(RSS) = \sigma^2 (n - \sum_{i=1}^{n} H_{i,i}).$$

Estimating σ: The Residual Standard Error

- Therefore,

$$\mathbb{E}(RSS) = \sum_{i,j}(I-H)_{i,j}\mathbb{E}\left(e_i e_j\right) = \sum_{i=1}^{n}(I-H)_{i,i}\sigma^2$$

  and so

$$\mathbb{E}(RSS) = \sigma^2(n - \sum_{i=1}^{n} H_{i,i}).$$

- The sum of the diagonal entries of a square matrix is called its trace i.e, $tr(A) = \sum_{i=1}^{n} A_{i,i}$.

Estimating σ: The Residual Standard Error

- Therefore,

$$\mathbb{E}(RSS) = \sum_{i,j}(I-H)_{i,j}\mathbb{E}\left(e_i e_j\right) = \sum_{i=1}^{n}(I-H)_{i,i}\sigma^2$$

 and so

$$\mathbb{E}(RSS) = \sigma^2(n - \sum_{i=1}^{n} H_{i,i}).$$

- The sum of the diagonal entries of a square matrix is called its trace i.e, $tr(A) = \sum_{i=1}^{n} A_{i,i}$. We can therefore write

$$\mathbb{E}(RSS) = \sigma^2(n - tr(H)).$$

Estimating σ: The Residual Standard Error

- Therefore,

Ask

$$\mathbb{E}(RSS) = \sum_{i,j}(I-H)_{i,j}\mathbb{E}\left(e_i e_j\right) = \sum_{i=1}^{n}(I-H)_{i,i}\sigma^2$$

and so

$$\mathbb{E}(RSS) = \sigma^2(n - \sum_{i=1}^{n} H_{i,i}).$$

- The sum of the diagonal entries of a square matrix is called its trace i.e, $tr(A) = \sum_{i=1}^{n} A_{i,i}$. We can therefore write

$$\mathbb{E}(RSS) = \sigma^2(n - tr(H)).$$

- A very important fact about trace is $tr(AB) = tr(BA)$.

Estimating σ: The Residual Standard Error

- Therefore,

$$\mathbb{E}(RSS) = \sum_{i,j} (I-H)_{i,j}\mathbb{E}\left(e_i e_j\right) = \sum_{i=1}^{n} (I-H)_{i,i}\sigma^2$$

and so

$$\mathbb{E}(RSS) = \sigma^2(n - \sum_{i=1}^{n} H_{i,i}).$$

- The sum of the diagonal entries of a square matrix is called its trace i.e, $tr(A) = \sum_{i=1}^{n} A_{i,i}$. We can therefore write

$$\mathbb{E}(RSS) = \sigma^2(n - tr(H)).$$

- A very important fact about trace is $tr(AB) = tr(BA)$.
- Thus

$$tr(H) = tr(X(X^TX)^{-1}X^T) =$$

Estimating σ: The Residual Standard Error

- Therefore,

$$\mathbb{E}(RSS) = \sum_{i,j}(I-H)_{i,j}\mathbb{E}\left(e_i e_j\right) = \sum_{i=1}^{n}(I-H)_{i,i}\sigma^2$$

  and so

$$\mathbb{E}(RSS) = \sigma^2(n - \sum_{i=1}^{n} H_{i,i}).$$

- The sum of the diagonal entries of a square matrix is called its trace i.e, $tr(A) = \sum_{i=1}^{n} A_{i,i}$. We can therefore write

$$\mathbb{E}(RSS) = \sigma^2(n - tr(H)).$$

- A very important fact about trace is $tr(AB) = tr(BA)$.
- Thus

$$tr(H) = tr(X(X^TX)^{-1}X^T) = tr((X^TX)^{-1}X^TX) = tr(I_{p+1}).$$

Estimating σ: The Residual Standard Error

- We proved
$$\mathbb{E}(RSS) = \sigma^2(n - p - 1).$$
which proves that $\hat{\sigma}^2$ is an unbiased estimator of

σ^2.

Standard Errors of $\hat{\beta}$

- We have seen that $\mathbb{E}\hat{\beta} = \beta$ and that $Cov(\hat{\beta}) = \sigma^2(X^T X)^{-1}$.

Standard Errors of $\hat{\beta}$

- We have seen that $\mathbb{E}\hat{\beta} = \beta$ and that $Cov(\hat{\beta}) = \sigma^2(X^TX)^{-1}$.
- The standard error of $\hat{\beta}_i$ is therefore defined as $\hat{\sigma}$ multiplied by the square root of the ith diagonal entry of $(X^TX)^{-1}$.

Standard Errors of $\hat{\beta}$

- We have seen that $\mathbb{E}\hat{\beta} = \beta$ and that $Cov(\hat{\beta}) = \sigma^2(X^T X)^{-1}$.
- The standard error of $\hat{\beta}_i$ is therefore defined as $\hat{\sigma}$ multiplied by the square root of the ith diagonal entry of $(X^T X)^{-1}$.
- The standard error gives an idea of the accuracy of $\hat{\beta}_i$ as an estimator of β_i. These standard errors are part of the R output for the summary of the linear model.

Standardized or Studentized Residuals

- Under the assumptions $\mathbb{E}Y = X\beta$ and $Cov(X) = \sigma^2 I_n$, what are the means and the variances of the residuals $\hat{e}_1, \ldots, \hat{e}_n$?

Standardized or Studentized Residuals

- Under the assumptions $\mathbb{E}Y = X\beta$ and $Cov(X) = \sigma^2 I_n$, what are the means and the variances of the residuals $\hat{e}_1, \ldots, \hat{e}_n$?
- The expectation of $\hat{e}$ is

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

Standardized or Studentized Residuals

- Under the assumptions $\mathbb{E}Y = X\beta$ and $Cov(X) = \sigma^2 I_n$, what are the means and the variances of the residuals $\hat{e}_1, \ldots, \hat{e}_n$?
- The expectation of $\hat{e}$ is

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

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

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

Standardized or Studentized Residuals

- Under the assumptions $\mathbb{E}Y = X\beta$ and $Cov(X) = \sigma^2 I_n$, what are the means and the variances of the residuals $\hat{e}_1, \ldots, \hat{e}_n$?
- The expectation of $\hat{e}$ is

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

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

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

- This implies therefore that

  $$Var(\hat{e}_i) = \sigma^2(1-h_{ii}).$$

  where h_{ii} denotes the ith diagonal entry of H which is also known as the ith leverage (the above variance formula together with the formula that $Cov(\hat{Y}) = \sigma^2 H$ both imply that all leverages lie between 0 and 1).

Standardized or Studentized Residuals

- The variance of $\hat{e}_i$ can be standardized to 1 if we divide the residual by $\sigma\sqrt{1 - h_{ii}}$.

Standardized or Studentized Residuals

- The variance of $\hat{e}_i$ can be standardized to 1 if we divide the residual by $\sigma\sqrt{1-h_{ii}}$.
- But because σ is unknown, one divides by $\hat{\sigma}\sqrt{1-h_{ii}}$ and we call the resulting quantities Stan- dardized Residuals or Studentized Residuals:

$$r_i = \frac{\hat{e}_i}{\sigma\sqrt{1-h_{ii}}}.$$

Standardized or Studentized Residuals

- The variance of $\hat{e}_i$ can be standardized to 1 if we divide the residual by $\sigma\sqrt{1 - h_{ii}}$.
- But because σ is unknown, one divides by $\hat{\sigma}\sqrt{1 - h_{ii}}$ and we call the resulting quantities Stan- dardized Residuals or Studentized Residuals:

$$r_i = \frac{\hat{e}_i}{\sigma\sqrt{1 - h_{ii}}}.$$

- The standardized residuals $r_1, \ldots, r_n$ are very important in regression diagnostics. Various assumptions on the unobserved errors $e_1, \ldots, e_n$ can be checked through them.

What we have done so far:

- We have data $Y_{n\times 1}$ and $X_{n\times p}$.

What we have done so far:

- We have data $Y_{n \times 1}$ and $X_{n \times p}$.
- Based on these, we computed $\hat{\beta} = (X^T X)^{-1} X^T Y$.

What we have done so far:

- We have data $Y_{n\times 1}$ and $X_{n\times p}$.
- Based on these, we computed $\hat{\beta} = (X^T X)^{-1} X^T Y$.
- If we assume that Y is random, X is deterministic and that $\mathbb{E}Y = X\beta$ for some vector β, then $\hat{\beta}$ becomes an unbased estimate of β.

What we have done so far:

- We have data $Y_{n\times 1}$ and $X_{n\times p}$.
- Based on these, we computed $\hat{\beta} = (X^T X)^{-1} X^T Y$.
- If we assume that Y is random, X is deterministic and that $\mathbb{E}Y = X\beta$ for some vector β, then $\hat{\beta}$ becomes an unbased estimate of β.
- If we also assume that $Cov(Y) = \sigma^2 I_n$, then $\hat{\beta}$ is the best linear unbiased estimator (BLUE) of β. We have also learned how to compute the standard errors of $\beta_0, \dots, \beta_p$.

What we have done so far:

- We have data $Y_{n\times 1}$ and $X_{n\times p}$.
- Based on these, we computed $\hat{\beta} = (X^T X)^{-1} X^T Y$.
- If we assume that Y is random, X is deterministic and that $\mathbb{E}Y = X\beta$ for some vector β, then $\hat{\beta}$ becomes an unbased estimate of β.
- If we also assume that $Cov(Y) = \sigma^2 I_n$, then $\hat{\beta}$ is the best linear unbiased estimator (BLUE) of β. We have also learned how to compute the standard errors of $\beta_0, \dots, \beta_p$.
- Next we want to test hypotheses of the form $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$ and also to obtain confidence intervals for β_1 etc. For these, we need more distributional assumptions on Y.

What we have done so far:

- We have data $Y_{n \times 1}$ and $X_{n \times p}$.
- Based on these, we computed $\hat{\beta} = (X^T X)^{-1} X^T Y$.
- If we assume that Y is random, X is deterministic and that $\mathbb{E} Y = X\beta$ for some vector β, then $\hat{\beta}$ becomes an unbased estimate of β.
- If we also assume that $Cov(Y) = \sigma^2 I_n$, then $\hat{\beta}$ is the best linear unbiased estimator (BLUE) of β. We have also learned how to compute the standard errors of $\beta_0, \ldots, \beta_p$.
- Next we want to test hypotheses of the form $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$ and also to obtain confidence intervals for β_1 etc. For these, we need more distributional assumptions on Y.
- The most standard distributional assumption is that of normality: We assume that $Y \sim N(X, \sigma I_n)$. Under this assumption, a very nice theory of hypothesis testing and confidence intervals is available. This is what we study next.