MANIPAL
ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

Linear Regression Models

Segment 3 – Other Considerations in the MLRM
Topic 2 – Nonlinearity of Data: Residual Plots, Heteroskedasticity: Non-constant Variance of Error & Weighted Least Squares

Sudarsan N.S. Acharya (sudarsan.acharya@manipal.edu)

Topics

Residual plots: Introduction

Residual plots: Introduction

- Recall that the residual $R = Y - \hat{Y} = Y - \mathbf{X}\beta$ is a catch-all quantifier for everything that the linear model does not capture.

Residual plots: Introduction

- Recall that the residual $R = Y - \hat{Y} = Y - \mathbf{X}\beta$ is a catch-all quantifier for everything that the linear model does not capture.
- Residual is uncorrelated with the fitted response value:

Residual plots: Introduction

- Recall that the residual $R = Y - \hat{Y} = Y - \mathbf{X}\beta$ is a catch-all quantifier for everything that the linear model does not capture.
- Residual is uncorrelated with the fitted response value:

$$\text{cov}(R, \hat{Y})$$

Residual plots: Introduction

- Recall that the residual $R = Y - \hat{Y} = Y - \mathbf{X}\beta$ is a catch-all quantifier for everything that the linear model does not capture.
- Residual is <u>uncorrelated</u> with the fitted response value:

$$\text{cov}(R, \hat{Y}) = E\left[\left(R - \underbrace{E[R]}_{\approx \frac{1}{n}\mathbf{1}^{\mathrm{T}}\mathbf{r}=0}\right)\left(\hat{Y} - E\left[\hat{Y}\right]\right)\right]$$

Residual plots: Introduction

- Recall that the residual $R = Y - \hat{Y} = Y - \mathbf{X}\beta$ is a catch-all quantifier for everything that the linear model does not capture.
- Residual is <u>uncorrelated</u> with the fitted response value:

$$\text{cov}(R, \hat{Y}) = E\left[\left(R - \underbrace{E[R]}_{\approx \frac{1}{n}\mathbf{1}^{\mathrm{T}}\mathbf{r}=0}\right)\left(\hat{Y} - E\left[\hat{Y}\right]\right)\right] = E\left[R\hat{Y} - RE\left[\hat{Y}\right]\right]$$

Residual plots: Introduction

- Recall that the residual $R = Y - \hat{Y} = Y - \mathbf{X}\beta$ is a catch-all quantifier for everything that the linear model does not capture.
- Residual is <u>uncorrelated</u> with the fitted response value:

$$\text{cov}(R, \hat{Y}) = E\left[\left(R - \underbrace{E[R]}_{\approx \frac{1}{n}\mathbf{1}^{\mathrm{T}}\mathbf{r}=0}\right)\left(\hat{Y} - E\left[\hat{Y}\right]\right)\right] = E\left[R\hat{Y} - RE\left[\hat{Y}\right]\right]$$

$$= \underbrace{E\left[R\hat{Y}\right]}_{\approx \frac{1}{n}\mathbf{r}^{\mathrm{T}}\hat{\mathbf{y}}} - \underbrace{E\left[R\right]}_{=0} E\left[\hat{Y}\right]$$

Residual plots: Introduction

- Recall that the residual $R = Y - \hat{Y} = Y - \mathbf{X}\beta$ is a catch-all quantifier for everything that the linear model does not capture.
- Residual is <u>uncorrelated</u> with the fitted response value:

$$\text{cov}(R, \hat{Y}) = E\left[\left(R - \underbrace{E[R]}_{\approx \frac{1}{n}\mathbf{1}^{\mathrm{T}}\mathbf{r}=0}\right)\left(\hat{Y} - E\left[\hat{Y}\right]\right)\right] = E\left[R\hat{Y} - RE\left[\hat{Y}\right]\right]$$

$$= \underbrace{E\left[R\hat{Y}\right]}_{\approx \frac{1}{n}\mathbf{r}^{\mathrm{T}}\hat{\mathbf{y}}} - \underbrace{E\left[R\right]}_{=0} E\left[\hat{Y}\right] \approx \frac{1}{n}\mathbf{r}^{\mathrm{T}}\left(\underbrace{\mathbf{X}\hat{\beta}}_{=\hat{\mathbf{y}}}\right)$$

Residual plots: Introduction

- Recall that the residual $R = Y - \hat{Y} = Y - \mathbf{X}\beta$ is a catch-all quantifier for everything that the linear model does not capture.
- Residual is <u>uncorrelated</u> with the fitted response value:

$$\text{cov}(R, \hat{Y}) = E\left[\left(R - \underbrace{E[R]}_{\approx \frac{1}{n}\mathbf{1}^{\mathrm{T}}\mathbf{r}=0}\right)\left(\hat{Y} - E\left[\hat{Y}\right]\right)\right] = E\left[R\hat{Y} - RE\left[\hat{Y}\right]\right]$$

$$= \underbrace{E\left[R\hat{Y}\right]}_{\approx \frac{1}{n}\mathbf{r}^{\mathrm{T}}\hat{\mathbf{y}}} - \underbrace{E\left[R\right]}_{=0} E\left[\hat{Y}\right] \approx \frac{1}{n}\mathbf{r}^{\mathrm{T}}\left(\underbrace{\mathbf{X}\hat{\beta}}_{=\hat{\mathbf{y}}}\right) = \frac{1}{n}\left(\underbrace{\mathbf{X}^{\mathrm{T}}\mathbf{r}}_{=0}\right)^{\mathrm{T}}\hat{\beta} = 0.$$

Residual plots: Introduction

- Recall that the residual $R = Y - \hat{Y} = Y - \mathbf{X}\beta$ is a catch-all quantifier for everything that the linear model does not capture.
- Residual is uncorrelated with the fitted response value:

$$\text{cov}(R, \hat{Y}) = E\left[\left(R - \underbrace{E[R]}_{\approx \frac{1}{n}\mathbf{1}^{\mathrm{T}}\mathbf{r}=0}\right)\left(\hat{Y} - E\left[\hat{Y}\right]\right)\right] = E\left[R\hat{Y} - RE\left[\hat{Y}\right]\right]$$

$$= \underbrace{E\left[R\hat{Y}\right]}_{\approx \frac{1}{n}\mathbf{r}^{\mathrm{T}}\hat{\mathbf{y}}} - \underbrace{E\left[R\right]}_{=0} E\left[\hat{Y}\right] \approx \frac{1}{n}\mathbf{r}^{\mathrm{T}}\left(\underbrace{\mathbf{X}\hat{\beta}}_{=\hat{\mathbf{y}}}\right) = \frac{1}{n}\left(\underbrace{\mathbf{X}^{\mathrm{T}}\mathbf{r}}_{=0}\right)^{\mathrm{T}}\hat{\beta} = 0.$$

- Residual is +vely correlated with true response value:

Residual plots: Introduction

- Recall that the residual $R = Y - \hat{Y} = Y - \mathbf{X}\beta$ is a catch-all quantifier for everything that the linear model does not capture.
- Residual is uncorrelated with the fitted response value:

$$\text{cov}(R, \hat{Y}) = E\left[\left(R - \underbrace{E[R]}_{\approx \frac{1}{n}\mathbf{1}^{\mathrm{T}}\mathbf{r}=0}\right)\left(\hat{Y} - E\left[\hat{Y}\right]\right)\right] = E\left[R\hat{Y} - RE\left[\hat{Y}\right]\right]$$

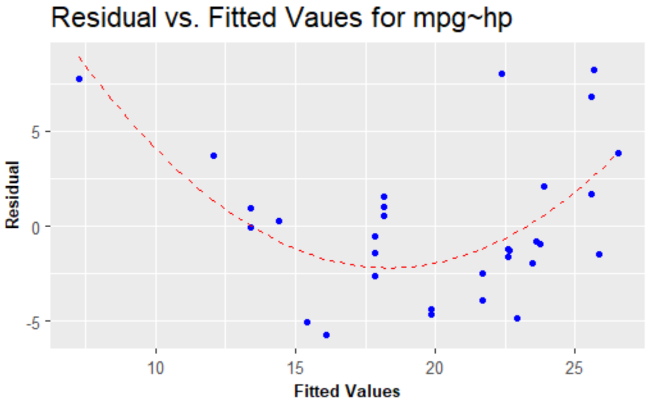
$$= \underbrace{E\left[R\hat{Y}\right]}_{\approx \frac{1}{n}\mathbf{r}^{\mathrm{T}}\hat{\mathbf{y}}} - \underbrace{E\left[R\right]}_{=0} E\left[\hat{Y}\right] \approx \frac{1}{n}\mathbf{r}^{\mathrm{T}}\left(\underbrace{\mathbf{X}\hat{\beta}}_{=\hat{\mathbf{y}}}\right) = \frac{1}{n}\left(\underbrace{\mathbf{X}^{\mathrm{T}}\mathbf{r}}_{=0}\right)^{\mathrm{T}}\hat{\beta} = 0.$$

- Residual is +vely correlated with true response value: $\text{cov}(R, Y) = \sigma^2$.

Residual plots: Continued

Residual plots: Continued

A residual plot shows the relationship between the residuals and the fitted values.



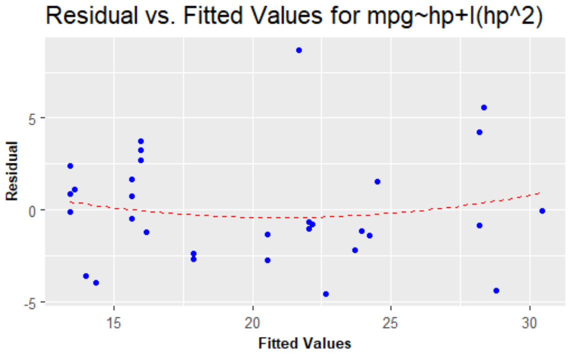
Interpreting Residual Plots

Interpreting Residual Plots

A residual plot with a discernible pattern is an indication of *nonlinearity*;

Interpreting Residual Plots

A residual plot with a discernible pattern is an indication of *nonlinearity*; apply nonlinear transformation for the predictor such as $X^2, \sqrt{X}$, etc.



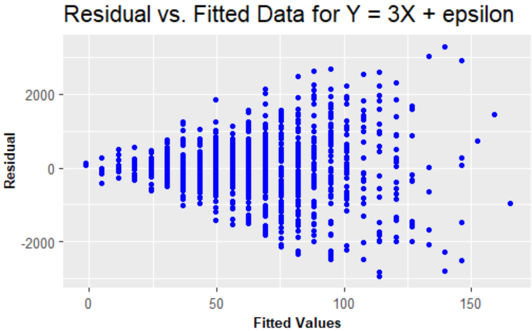
Interpreting Residual Plots: Continued

Interpreting Residual Plots: Continued

A funnel-shaped residual plot is an indication of heteroskedasticity,

Interpreting Residual Plots: Continued

A funnel-shaped residual plot is an indication of heteroskedasticity, which means that the random error term does not have a constant variance impacting the *standard error*, *confidence interval*, and hypothesis test calculations.



Heteroskedasticity: Non-constant Variance of Error & Weighted Least Squares

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

Heteroskedasticity: Non-constant Variance of Error & Weighted Least Squares

- In weighted least squares, we minimize the the sum of the squares of the *weighted* residuals for all samples in the dataset:

Heteroskedasticity: Non-constant Variance of Error & Weighted Least Squares

- In weighted least squares, we minimize the the sum of the squares of the *weighted* residuals for all samples in the dataset:

$$\min \sum_{i=1}^{n} \left(w_i r^{(i)}\right)^2 = \sum_{i=1}^{n} w_i^2 \left(y^{(i)} - \left(\beta_0 + \beta_1 x_1^{(i)} + \cdots + \beta_p x_p^{(i)}\right)\right)^2.$$

Heteroskedasticity: Non-constant Variance of Error & Weighted Least Squares

- In weighted least squares, we minimize the the sum of the squares of the *weighted* residuals for all samples in the dataset:
  $\min \sum_{i=1}^{n} \left(w_i r^{(i)}\right)^2 = \sum_{i=1}^{n} w_i^2 \left(y^{(i)} - \left(\beta_0 + \beta_1 x_1^{(i)} + \cdots + \beta_p x_p^{(i)}\right)\right)^2.$
- The weights are chosen such that samples with large error variances contribute less to the summation above,

Heteroskedasticity: Non-constant Variance of Error & Weighted Least Squares

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- In weighted least squares, we minimize the the sum of the squares of the *weighted* residuals for all samples in the dataset:
$$\min \sum_{i=1}^{n} \left(w_i r^{(i)}\right)^2 = \sum_{i=1}^{n} w_i^2 \left(y^{(i)} - \left(\beta_0 + \beta_1 x_1^{(i)} + \cdots + \beta_p x_p^{(i)}\right)\right)^2 .$$
- The weights are chosen such that samples with large error variances contribute less to the summation above, thus to the linear model.

Heteroskedasticity: Non-constant Variance of Error & Weighted Least Squares

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- In weighted least squares, we minimize the the sum of the squares of the *weighted* residuals for all samples in the dataset:
$$\min \sum_{i=1}^{n} \left(w_i r^{(i)}\right)^2 = \sum_{i=1}^{n} w_i^2 \left(y^{(i)} - \left(\beta_0 + \beta_1 x_1^{(i)} + \cdots + \beta_p x_p^{(i)}\right)\right)^2.$$
- The weights are chosen such that samples with large error variances contribute less to the summation above, thus to the linear model.
- Typically, the recorded response for the ith sample is calculated as an average of n_i raw observations.

Heteroskedasticity: Non-constant Variance of Error & Weighted Least Squares

- In weighted least squares, we minimize the the sum of the squares of the *weighted* residuals for all samples in the dataset:
$$\min \sum_{i=1}^{n} \left(w_i r^{(i)}\right)^2 = \sum_{i=1}^{n} w_i^2 \left(y^{(i)} - \left(\beta_0 + \beta_1 x_1^{(i)} + \cdots + \beta_p x_p^{(i)}\right)\right)^2.$$
- The weights are chosen such that samples with large error variances contribute less to the summation above, thus to the linear model.
- Typically, the recorded response for the ith sample is calculated as an average of n_i raw observations.
- If each of those raw observations are uncorrelated with constant variance σ^2,

Heteroskedasticity: Non-constant Variance of Error & Weighted Least Squares

- In weighted least squares, we minimize the the sum of the squares of the *weighted* residuals for all samples in the dataset:
$$\min \sum_{i=1}^{n} \left(w_i r^{(i)}\right)^2 = \sum_{i=1}^{n} w_i^2 \left(y^{(i)} - \left(\beta_0 + \beta_1 x_1^{(i)} + \cdots + \beta_p x_p^{(i)}\right)\right)^2.$$
- The weights are chosen such that samples with large error variances contribute less to the summation above, thus to the linear model.
- Typically, the recorded response for the ith sample is calculated as an average of n_i raw observations.
- If each of those raw observations are uncorrelated with constant variance σ^2, then the variance in the ith sample is σ^2/n_i.

Heteroskedasticity: Non-constant Variance of Error & Weighted Least Squares

- In weighted least squares, we minimize the the sum of the squares of the *weighted* residuals for all samples in the dataset:
$$\min \sum_{i=1}^{n} \left(w_i r^{(i)}\right)^2 = \sum_{i=1}^{n} w_i^2 \left(y^{(i)} - \left(\beta_0 + \beta_1 x_1^{(i)} + \cdots + \beta_p x_p^{(i)}\right)\right)^2.$$
- The weights are chosen such that samples with large error variances contribute less to the summation above, thus to the linear model.
- Typically, the recorded response for the ith sample is calculated as an average of n_i raw observations.
- If each of those raw observations are uncorrelated with constant variance σ^2, then the variance in the ith sample is σ^2/n_i.
- The weights for each sample can now be chosen to be proportional to the inverse of the associated variances;

Heteroskedasticity: Non-constant Variance of Error & Weighted Least Squares

- In weighted least squares, we minimize the the sum of the squares of the *weighted* residuals for all samples in the dataset:
$$\min \sum_{i=1}^{n} \left(w_i r^{(i)}\right)^2 = \sum_{i=1}^{n} w_i^2 \left(y^{(i)} - \left(\beta_0 + \beta_1 x_1^{(i)} + \cdots + \beta_p x_p^{(i)}\right)\right)^2.$$
- The weights are chosen such that samples with large error variances contribute less to the summation above, thus to the linear model.
- Typically, the recorded response for the ith sample is calculated as an average of n_i raw observations.
- If each of those raw observations are uncorrelated with constant variance σ^2, then the variance in the ith sample is σ^2/n_i.
- The weights for each sample can now be chosen to be proportional to the inverse of the associated variances; that is, $w_i = n_i$.

Summary

Summary

- Interpret residual plots for checking assumptions in ordinary least squares linear regression models

Summary

- Interpret residual plots for checking assumptions in ordinary least squares linear regression models
- Describe methods for dealing with non-constant variance random error in linear regression