Lecture 11

Finish Model Checking

Implementation of WLS and robust variance estimation in R

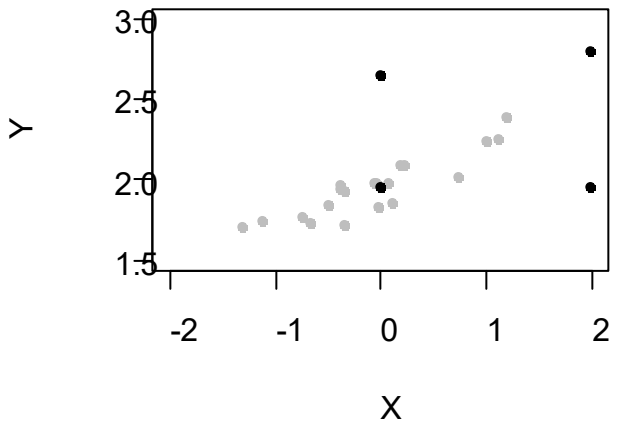
Key Assumptions by Order of Importance

- E(Y|X) = Xβ, i.e. the mean model is "correctly" specified
  - Misspecification of Xβ can lead to biased β / misinterpretations
  - Omitted variable Bias
  - Correct functional form for continuous X
- Residuals are independent
  - This assumption is violated due to the design of the study
  - Longitudinal study
  - Clustered design
  - Show today: ignoring the correlation will impact $Var(\hat{\beta})$ and derive weighted least squares
- Variance of residuals is constant
  - Often the variance is a function of some X
  - Show today: same impact and solution as violation of independence
- Residuals are normally distributed
  - CLT, bootstrap procedure
- There are not a small number of highly influencial observations
  - Sensitivity analyses

Leverage and Influence

- Leverage: A measure of how far an individual's predictors (X_i) are from the mean X_i
  - Hat matrix: $h_{ii} = \frac{(X_i - \bar{X})^2}{\sum(X_i - \bar{X})^2}$

- Influence: An observation (Y_i, X_i) such that including this value would greatly change the fitted values: $\hat{\beta}$ and $\hat{Y}$.

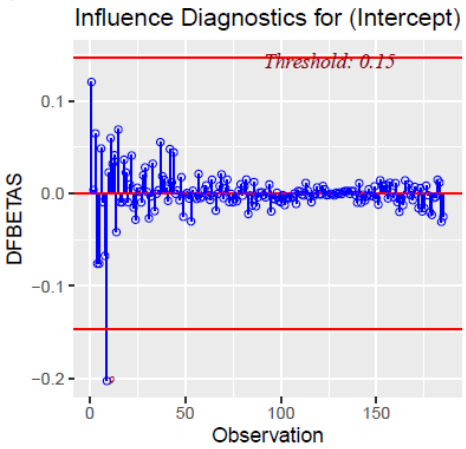


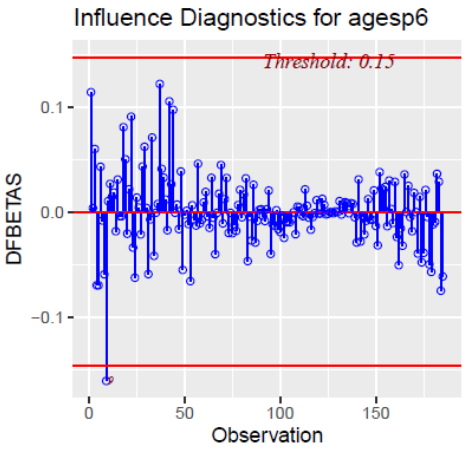
Influence statistics

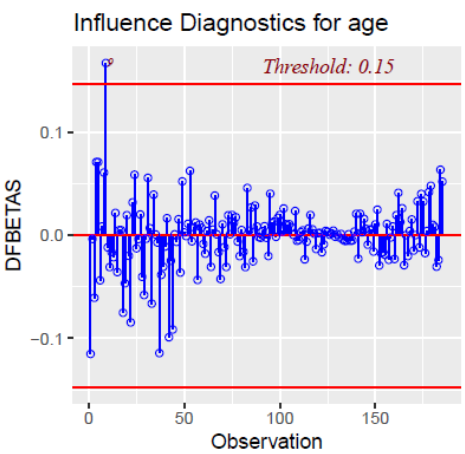
There are several influence statistics that are used in practice:

- $DBETA_{ij} = \hat{\beta}_j - \hat{\beta}_{j(-i)}$
- $DBETAS_{ij} = \frac{DBETA_{ij}}{\hat{se}(\hat{\beta}_{j(-i)})}$
- $DFIT_i = \hat{Y}_i - \hat{Y}_{i(-i)}$
- $DFITS_i = \frac{DFIT_i}{\hat{se}(\hat{Y}_{i(-i)})}$

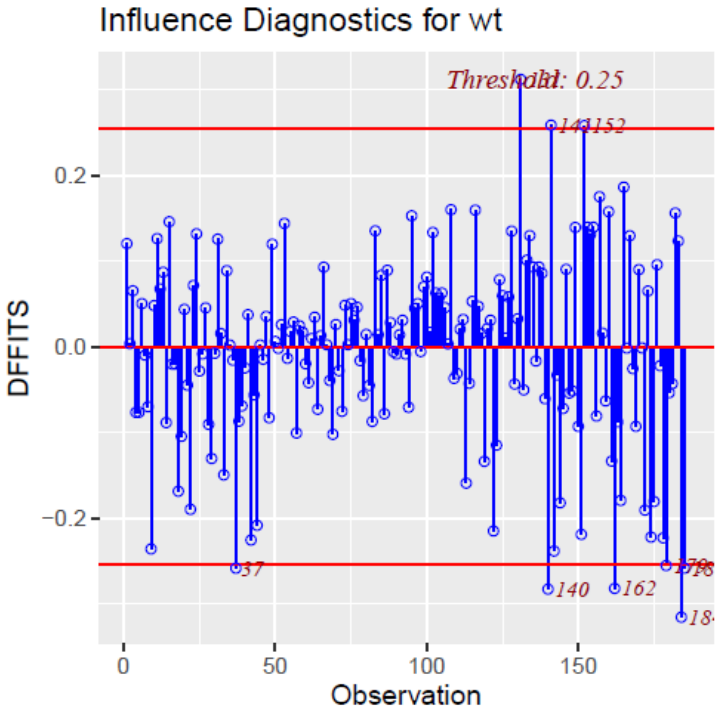
Example: Nepali Anthropometry Data





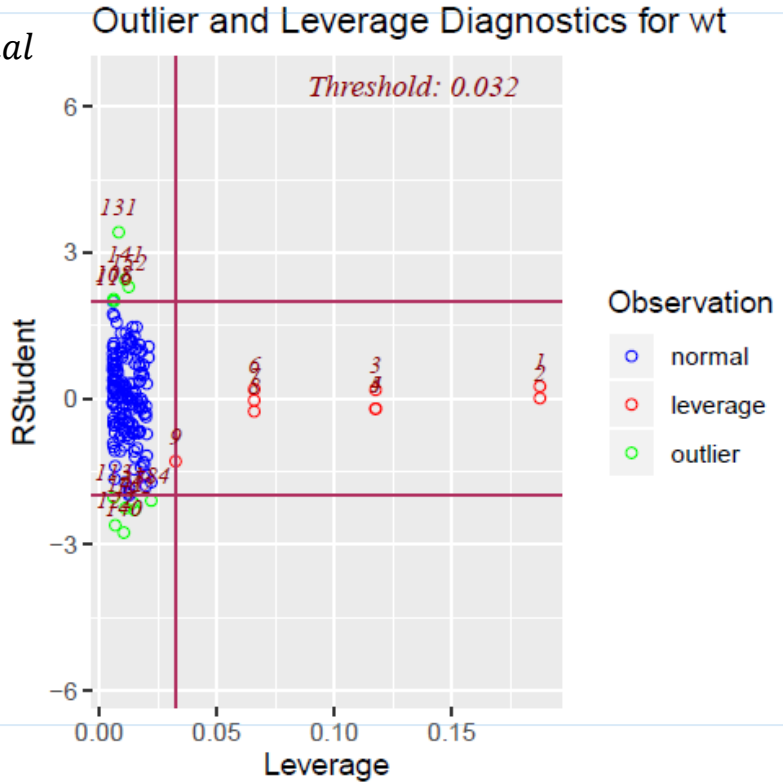


Example: Nepali Anthropometry Data



Example: Nepali Anthropometry Data

- *RStudent*: *studentized residual*
- $\frac{y_i - \hat{y}_{(i)}}{se(y_i - \hat{y}_{(i)})} = \frac{y_i - \hat{y}_i}{\sqrt{MSE_{(i)}(1 - h_{ii})}}$



Implementation of WLS in R

- For the remainder of the lecture, we will work through some analyses to demonstrate how to fit WLS in R
- In addition, I will show one approach for obtained robust variance estimates for different working correlation models. Here we will use the *gee* package in R.
- See Handout 11.