JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

752 → grading
Quiz 2 →
753 → posted

Lab 6
Lab 7 and 8

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

WLS

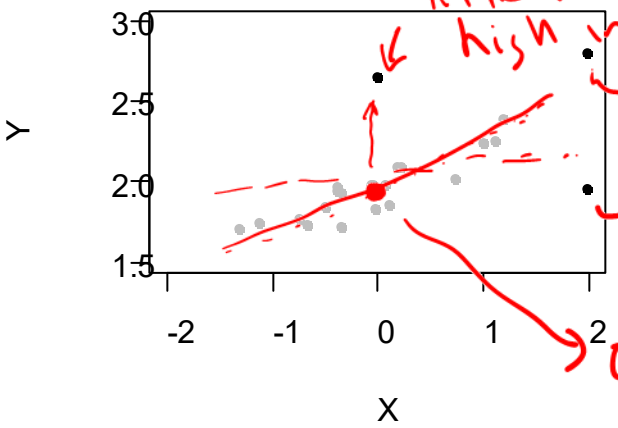
WLS

correlated data ⇒ resample level 1 units

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 $\hat{Y}$.



little to no leverage
high influence → $\hat{\beta}$.
high leverage
no/little influence
high leverage point
high influence
0 leverage / no leverage
little to no influence

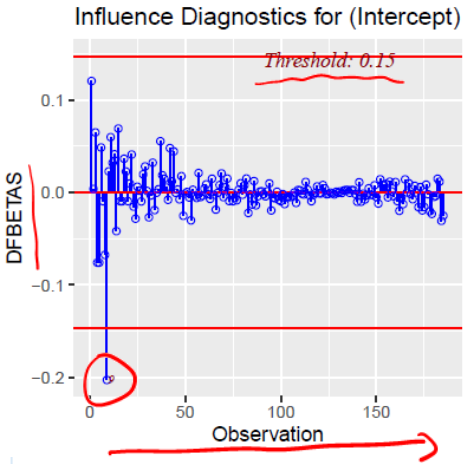
Influence statistics

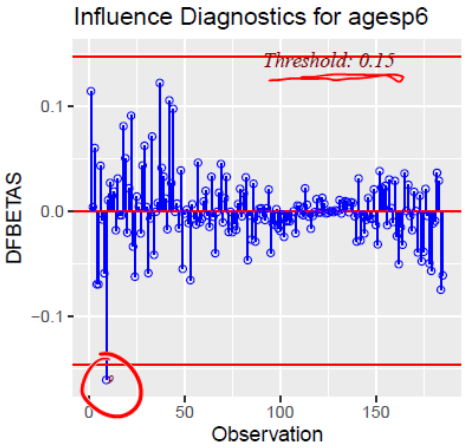
There are several influence statistics that are used in practice:

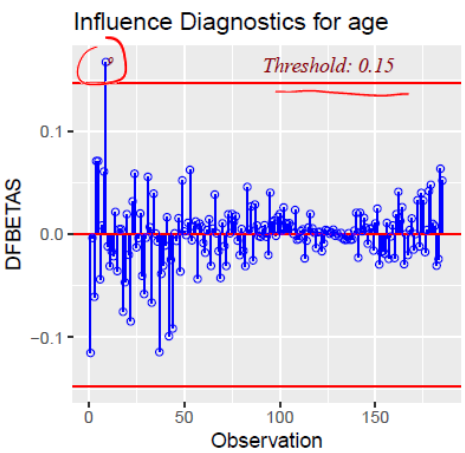
- $DBETA_{ij} = \hat{\beta}_j - \hat{\beta}_{j(-i)}$
- $DBETAS_{ij} = \frac{DBETA_{ij}}{\hat{se}(\hat{\beta}_{j(-i)})}$
- $DFIT_i = \hat{Y}_i - \hat{Y}_{i(-i)}$
- $DFITS_i = \frac{DFIT_i}{\hat{se}(\hat{Y}_{i(-i)})}$

=> defined for each parameter in the model, SLR Dbeta β_0, β_1 $j = 0, 1$

Example: Nepali Anthropometry Data



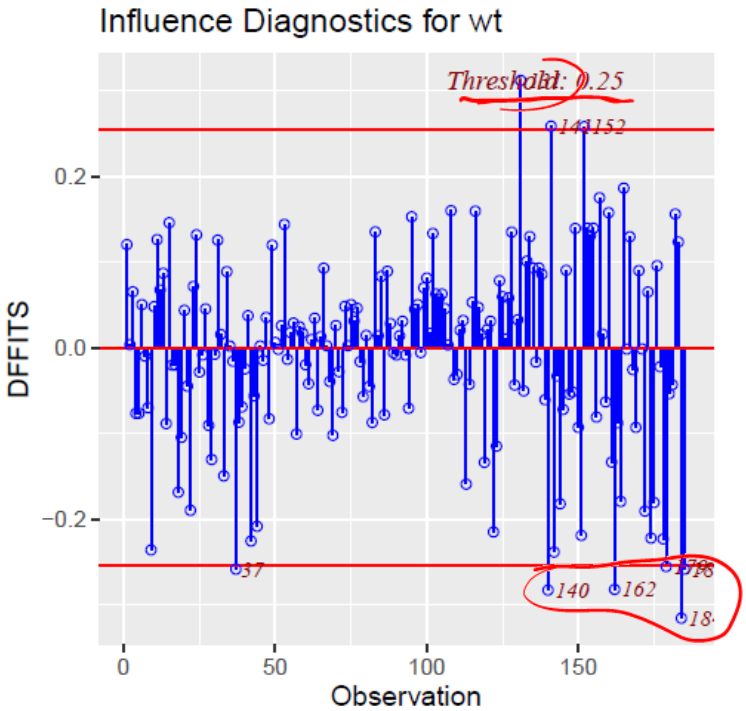




$WT = \beta_0 + \beta_1 age + \beta_2 (age - 6)^+ + \epsilon$

Obs 9

Example: Nepali Anthropometry Data

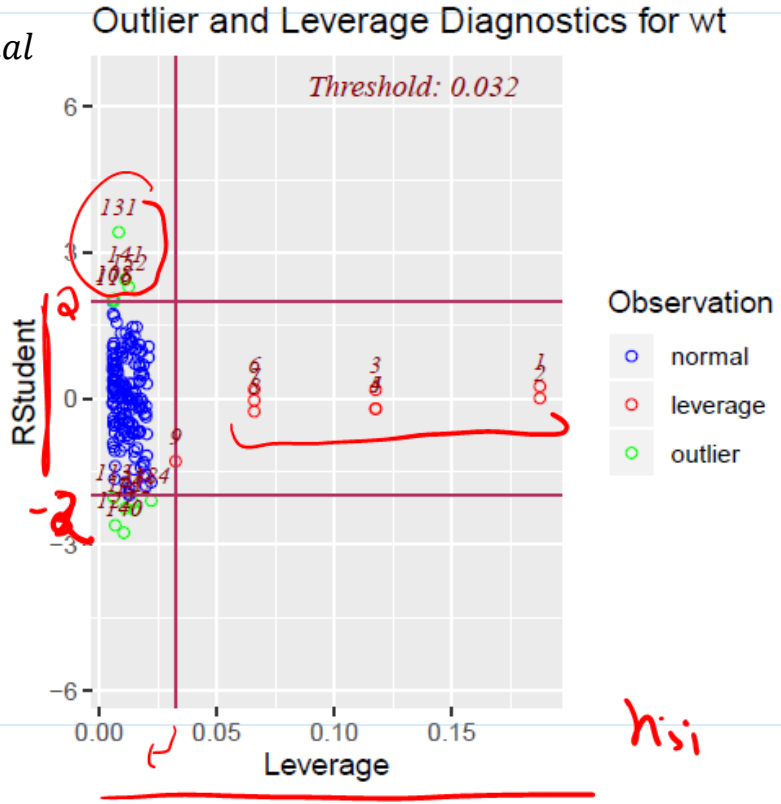


Example: Nepali Anthropometry Data

- *RStudent*: *studentized residual*

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

mean 0
SD 1



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.