

Model checking

- Checking the fit of your model(s)
- Not much advice (in literature) for general likelihood approaches
- Not commonly done for non-standard models (i.e. not standard regression)
- Examining diagnostics: gold-standard practice for linear Normal models
- It should be gold-standard for all models!

What are the issues?

- Systematic departures of the process (biological) model from the data
 - e.g. proposed model is linear whereas the true model is nonlinear
- A poor error distribution
 - e.g. proposed error distribution is Normal whereas the true error distribution is lognormal

What are the issues?

- Mistakes in the data
 - Incorrect entries, broken instruments etc
 - Rule of thumb: 1-5% of data points
- Outliers
 - Data points that lie unexpectedly far from the prediction
 - Could be mistakes (oops)
 - Could be biological exceptions (interesting)

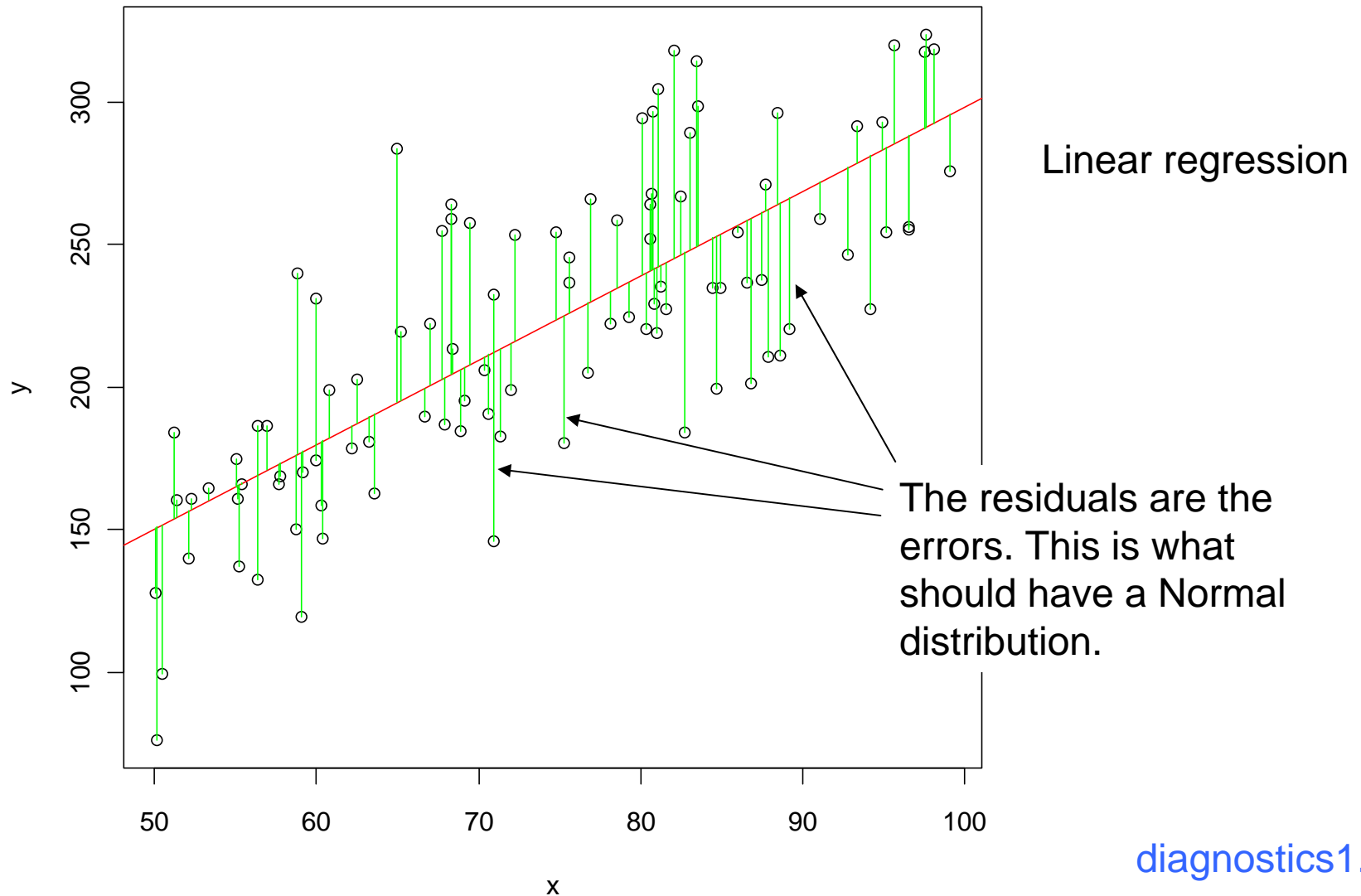
What are the issues?

- Influential data points
 - Data points that cause radically different parameter estimates.

Tools so far

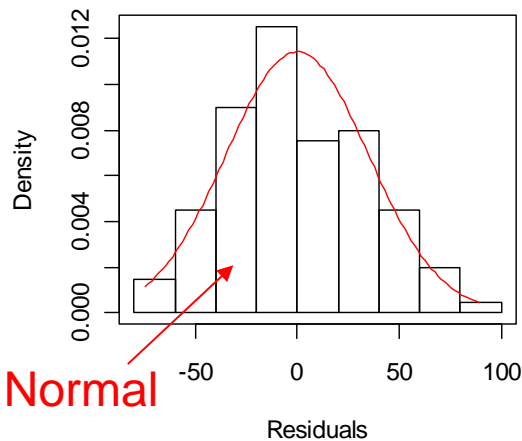
- Plotting the fitted model with the data
- Checking likelihood profiles
 - have we found the minimum?

Normal likelihood diagnostics



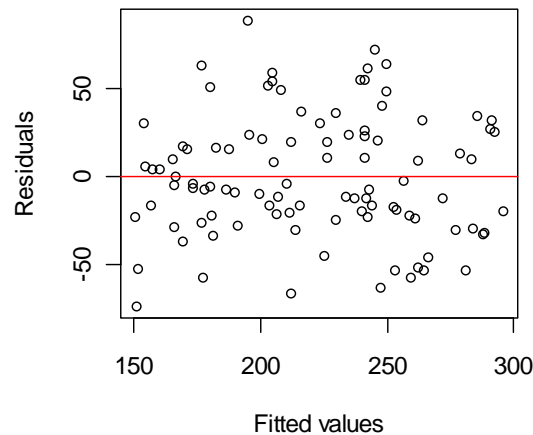
Normal likelihood diagnostics

Histogram of residuals



Looks Normal

Residuals vs fitted

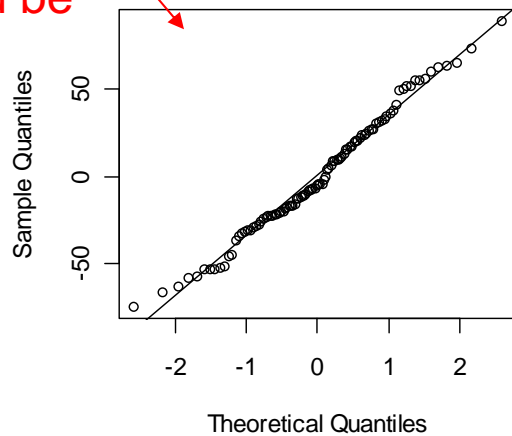


Linear regression

This is what it should look like

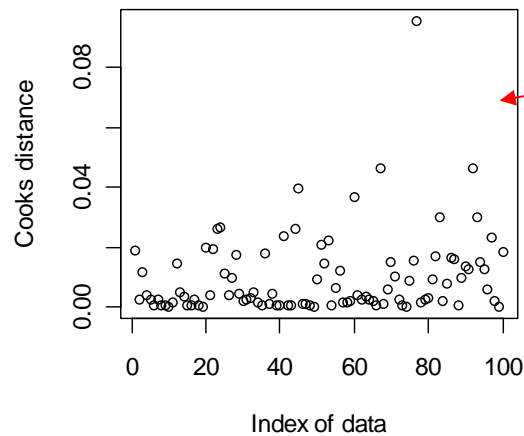
A special plot for the Normal distribution

Normal Q-Q Plot



Should be linear

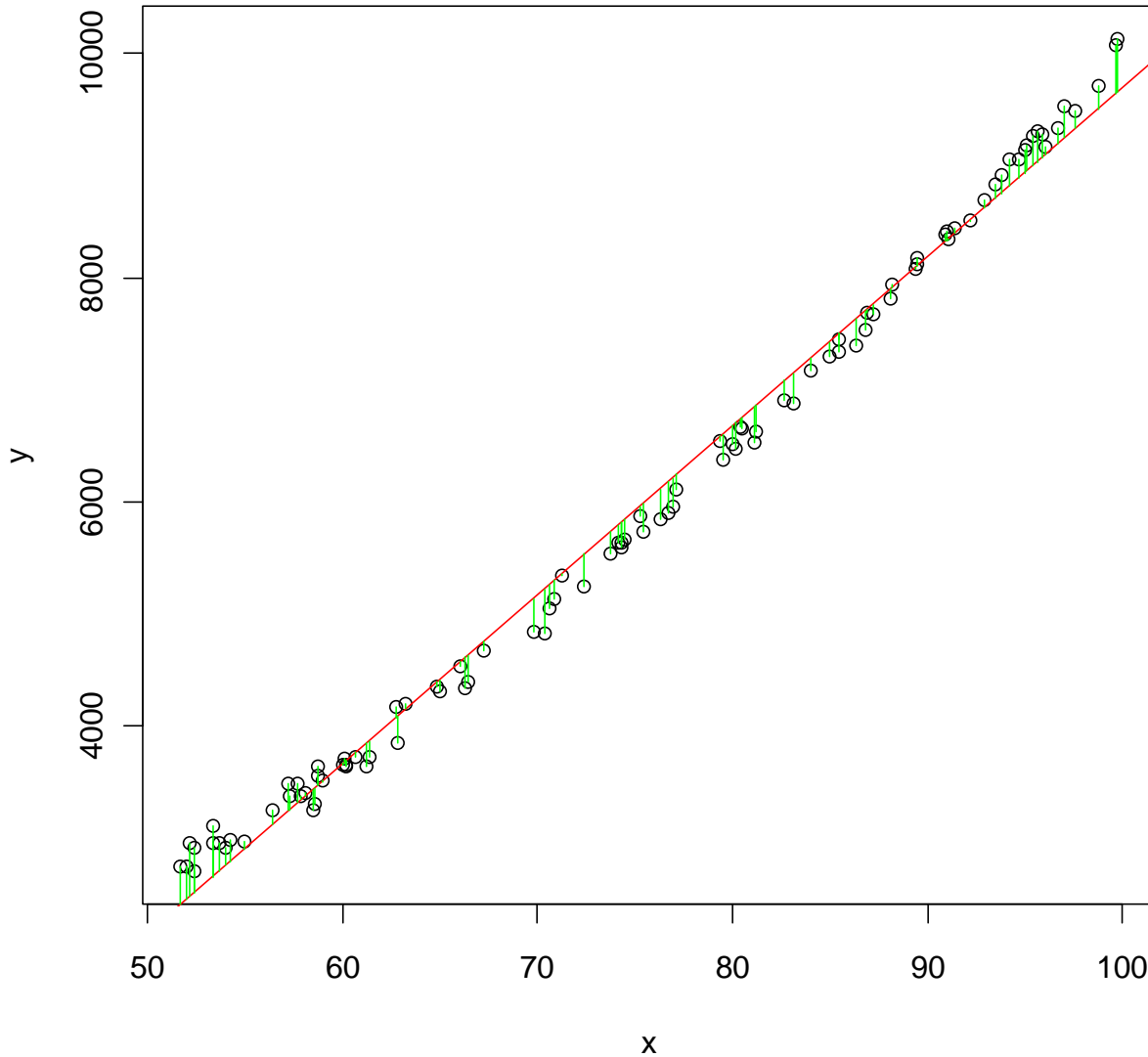
Influence



A special plot for linear models incl. GLM.

A measure of the influence of individual points

Normal likelihood diagnostics

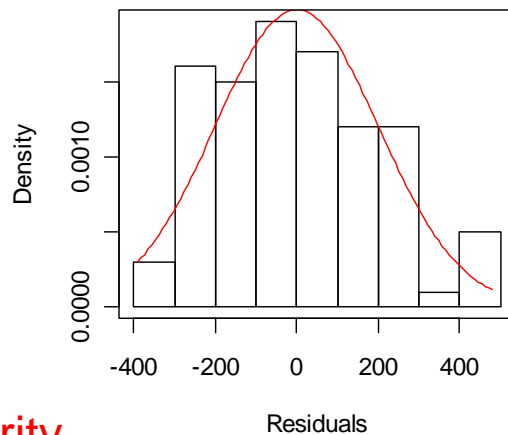


Linear regression

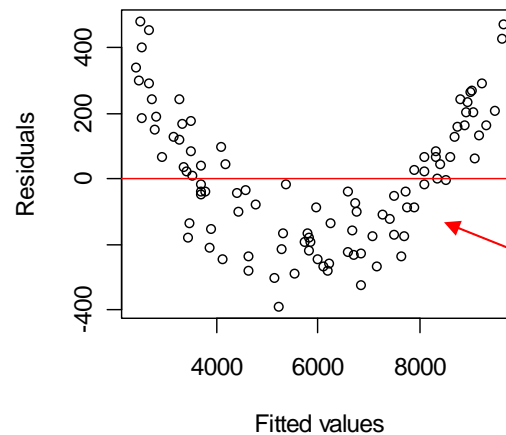
diagnostics1.R

Normal likelihood diagnostics

Histogram of residuals



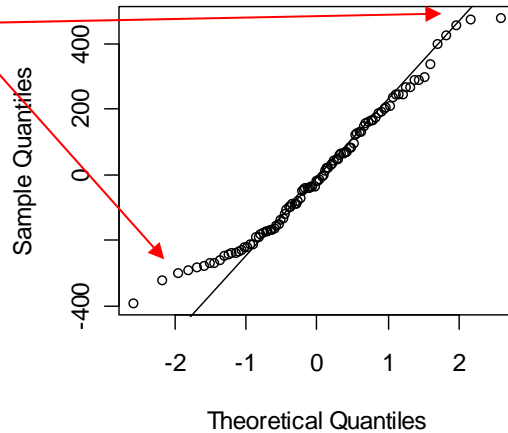
Residuals vs fitted



Linear regression

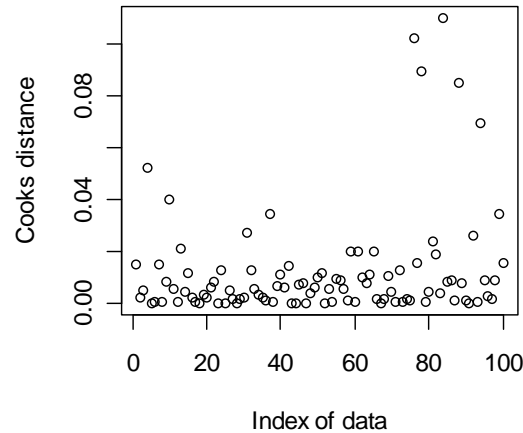
Nonlinearity
clearly seen

Normal Q-Q Plot

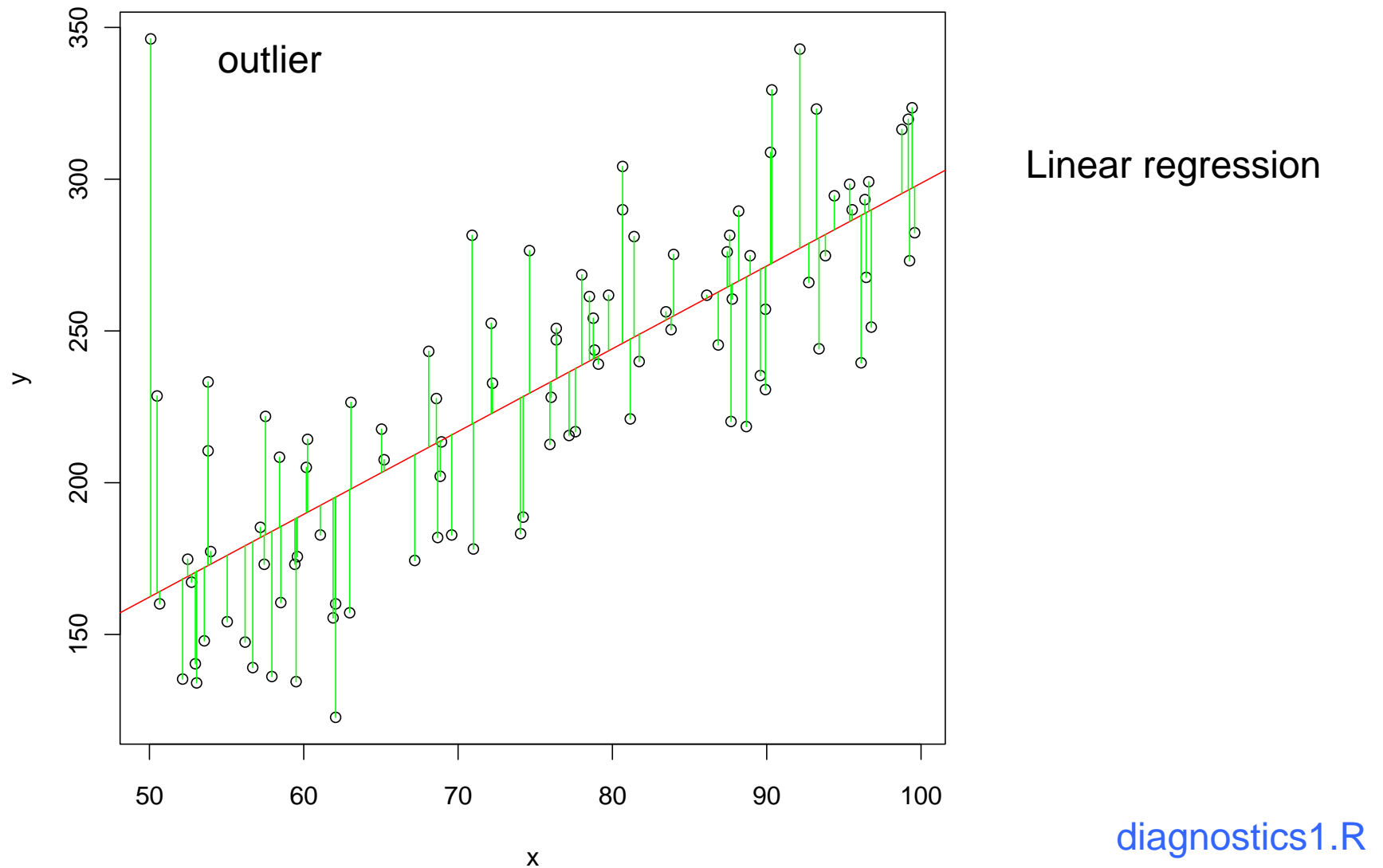


Nonlinearity
clearly seen.
Curves off the
ends.

Influence

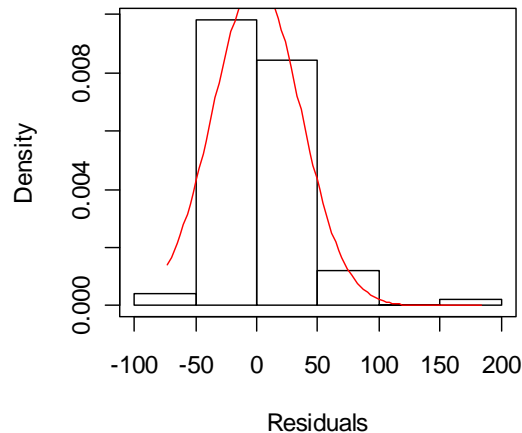


Normal likelihood diagnostics

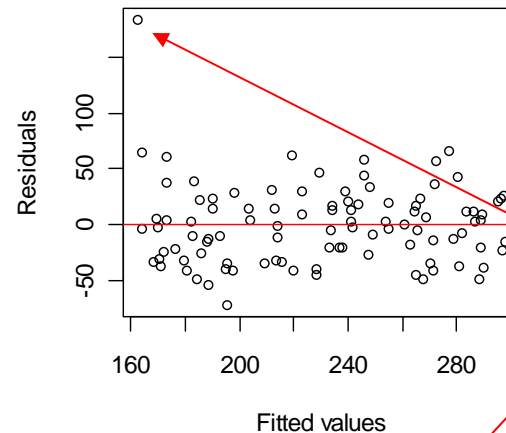


Normal likelihood diagnostics

Histogram of residuals



Residuals vs fitted

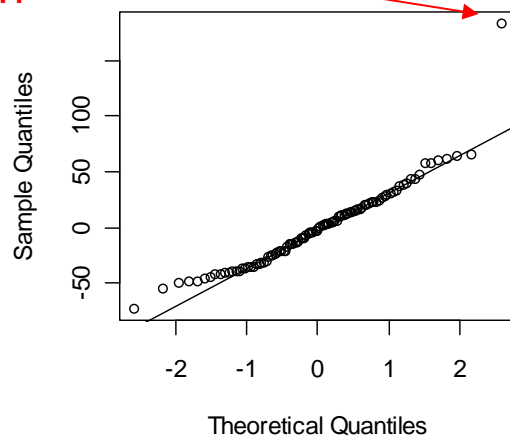


Linear regression

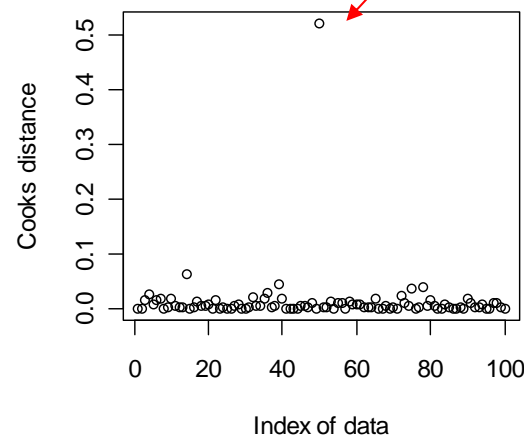
Outlier
clearly seen

Outlier
clearly seen

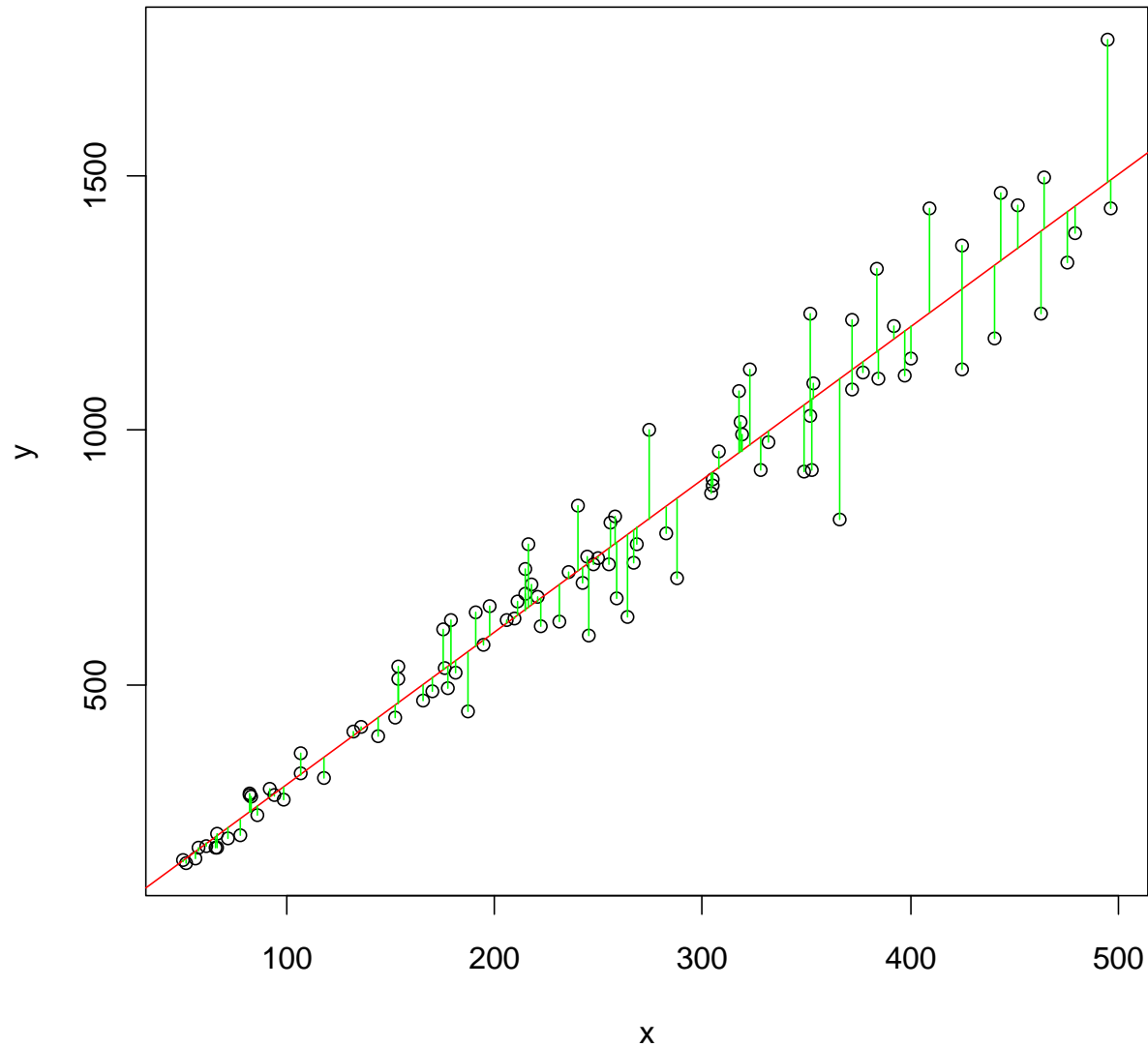
Normal Q-Q Plot



Influence



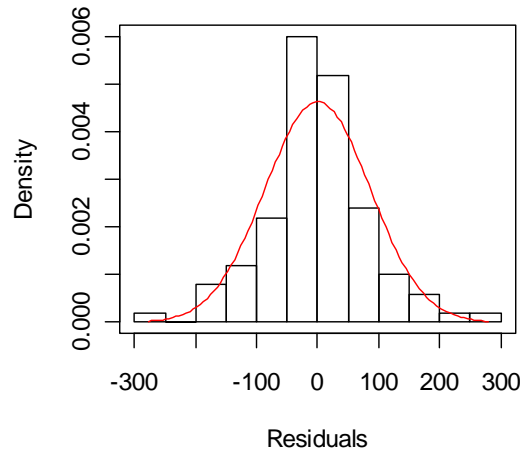
Normal likelihood diagnostics



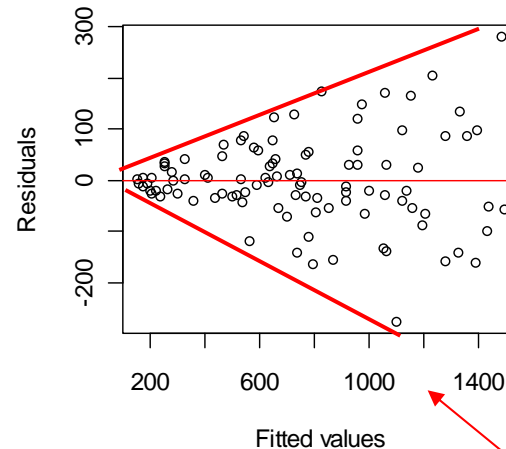
Linear regression

Normal likelihood diagnostics

Histogram of residuals



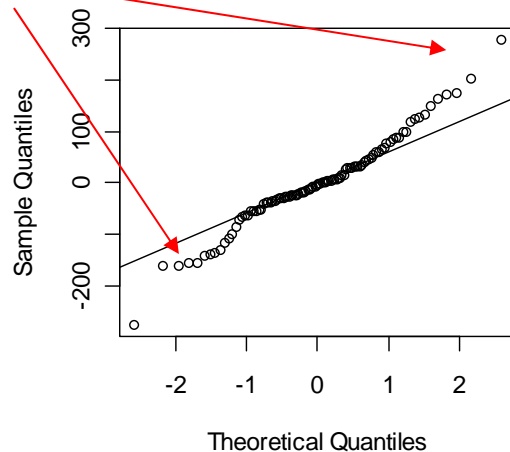
Residuals vs fitted



Linear regression

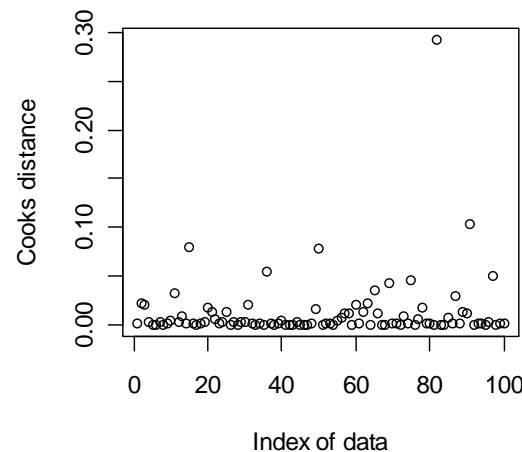
Heteroscedasticity:
v-shaped residuals

Normal Q-Q Plot



Rolls off
the ends

Influence



General likelihood diagnostics

	Normal, linear	Normal, nonlinear	Not Normal
Residual vs fitted	X	X	X standardized
Histogram of residuals	X	X	
Q-Q plot	X	X	some
Case deletion	X Cook's	X	X

Standardized residuals

$$r^* = \frac{r}{\sigma}$$

Raw residual

Standard deviation of the distribution evaluated at that point

$$\sigma = \sqrt{\text{variance}}$$

Case deletion diagnostics

- Simple idea: leave out each data point in turn and refit the model
- Likelihood displacement

$$LD = 2(\text{nll}(\theta, y_{-i}) - \text{nll}(\theta, y))$$

full dataset

deleting observation i

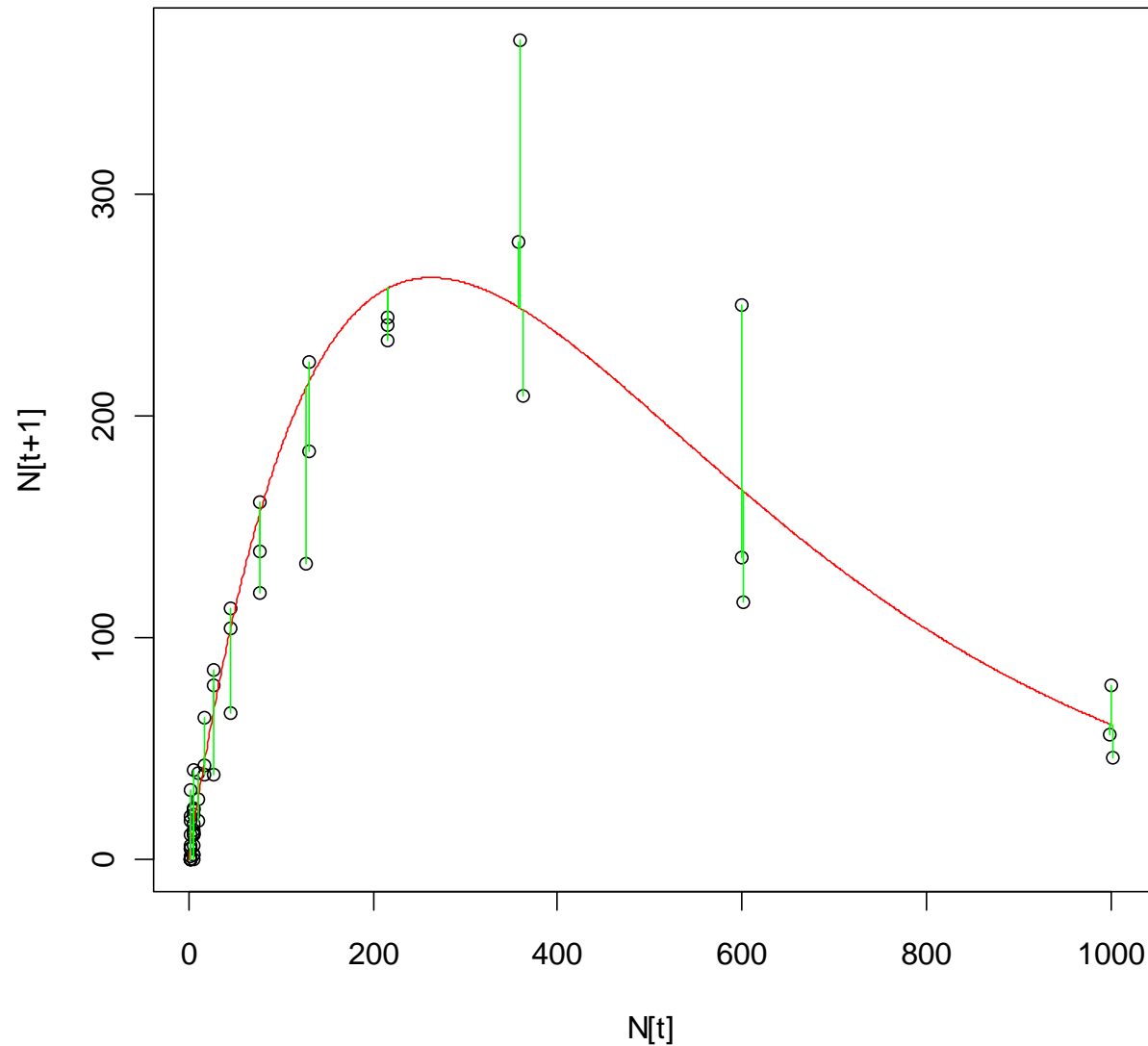
Notice LD is a
likelihood ratio

- Parameter sensitivity
 - percent change in parameter estimate by leaving out the data point

Cook and Weisberg (1982). *Residuals and influence in regression*. Chapman & Hall.

diagnostics2.R

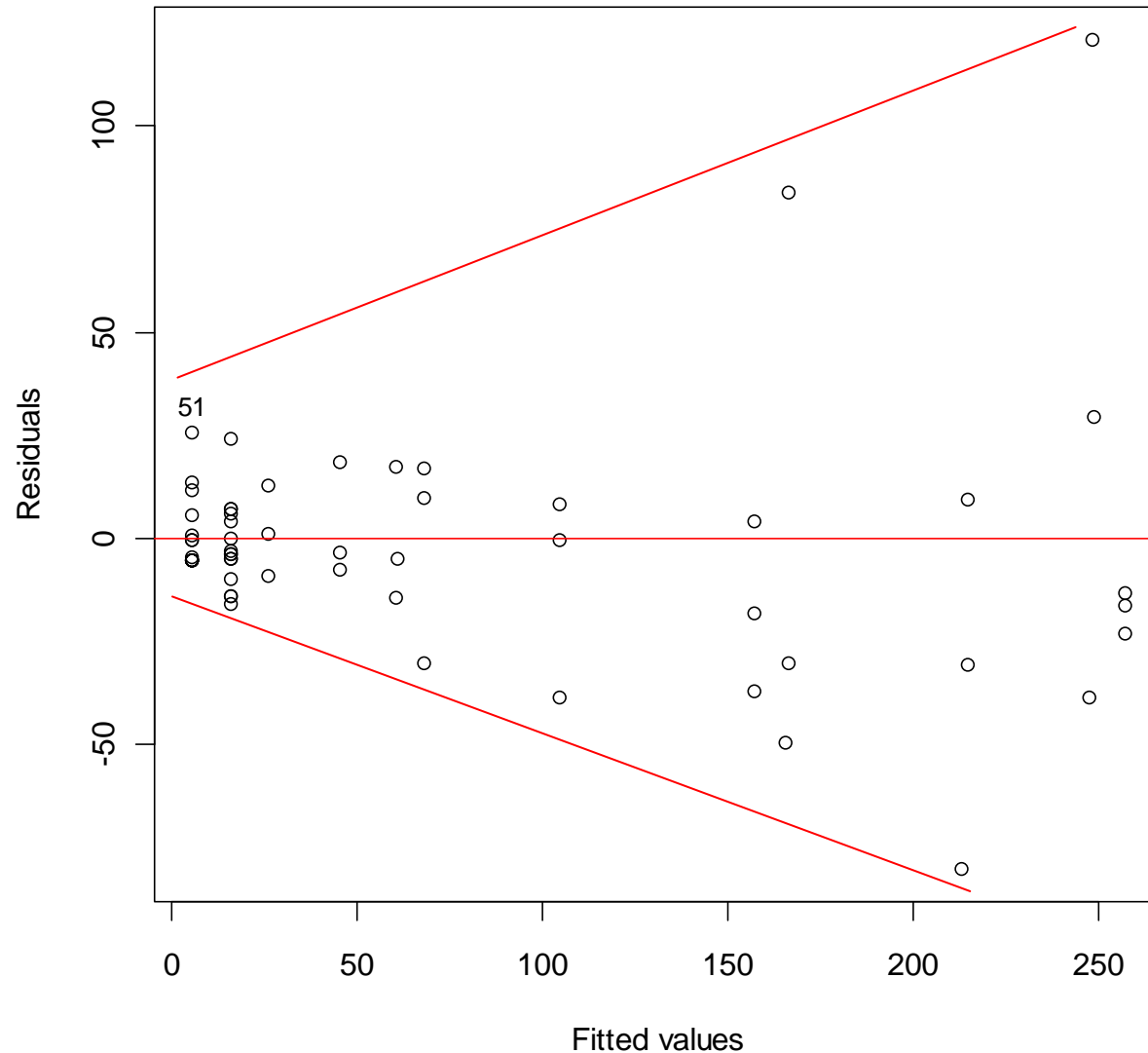
Nonlinear likelihood



Ricker model:

Negative
binomial
error

Model checking

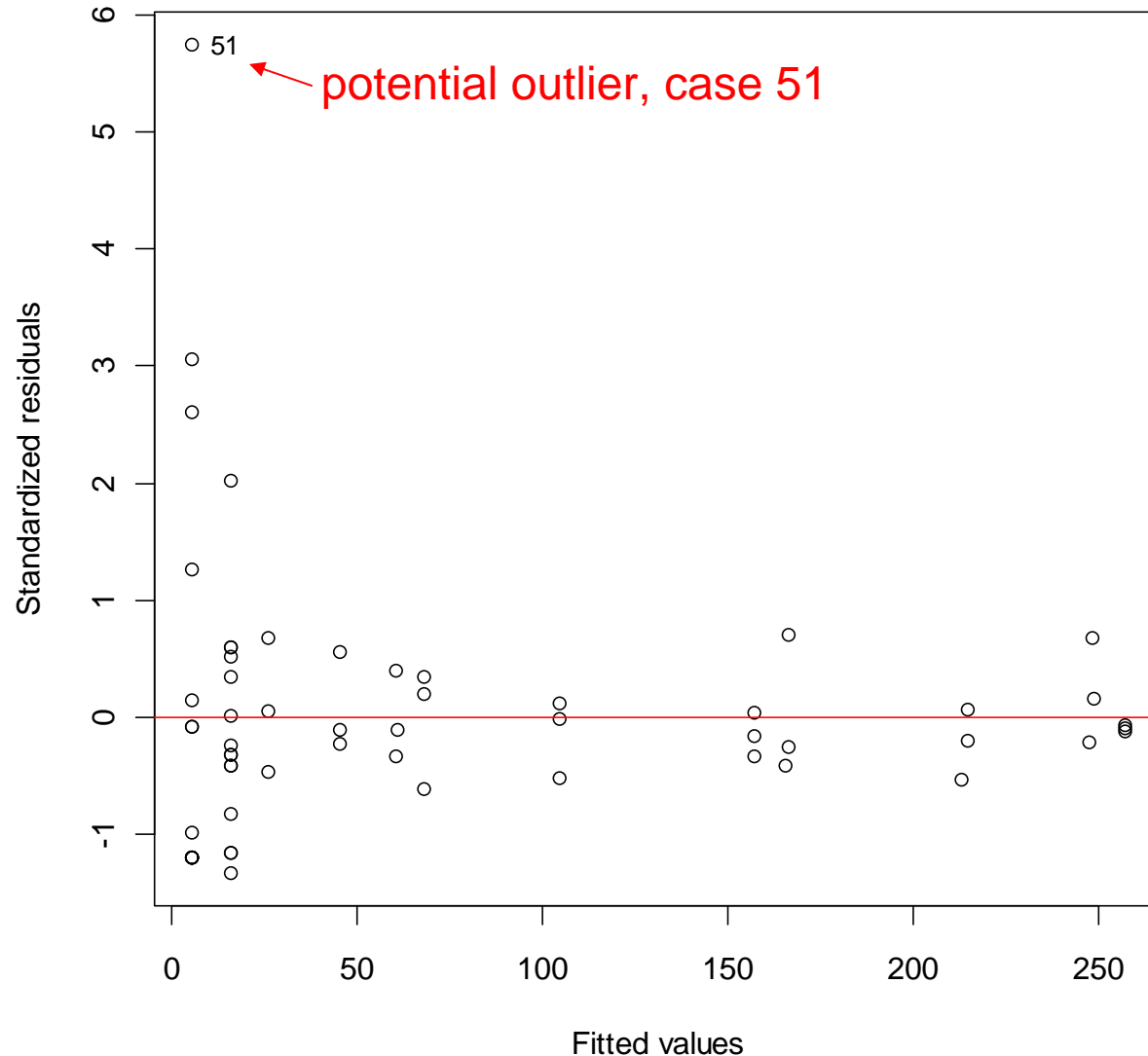


Residuals
versus
fitted values

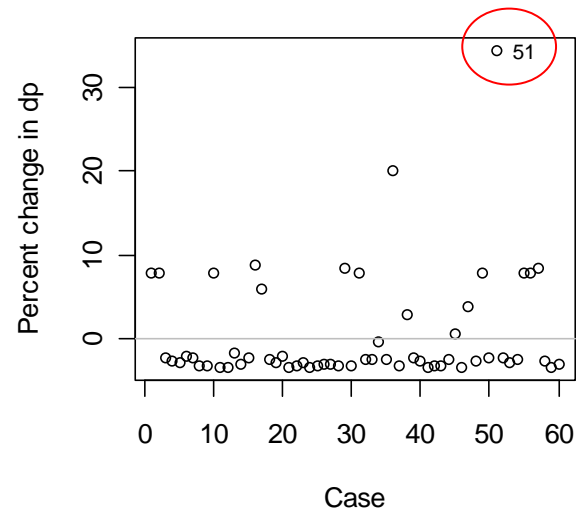
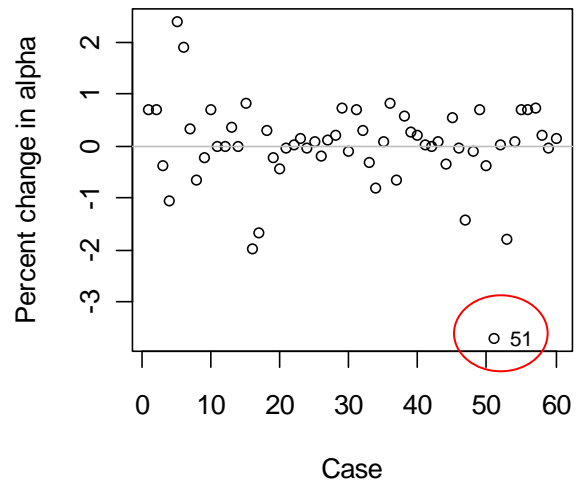
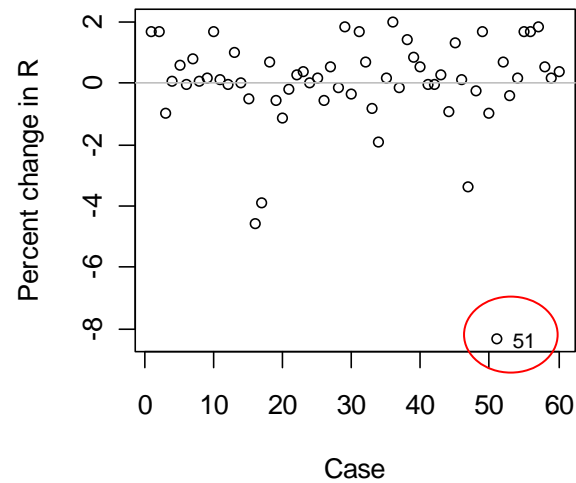
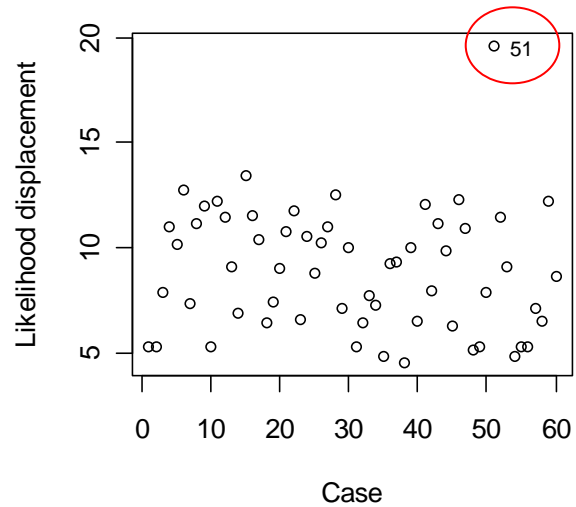
Shows increasing
variance with
increasing fv,
but this is what
you expect from
the neg binom

[diagnostics2.R](#)

Model checking



Influence



Influence:
Case
deletion
diagnostics

[diagnostics2.R](#)