Residual-fitted value plots

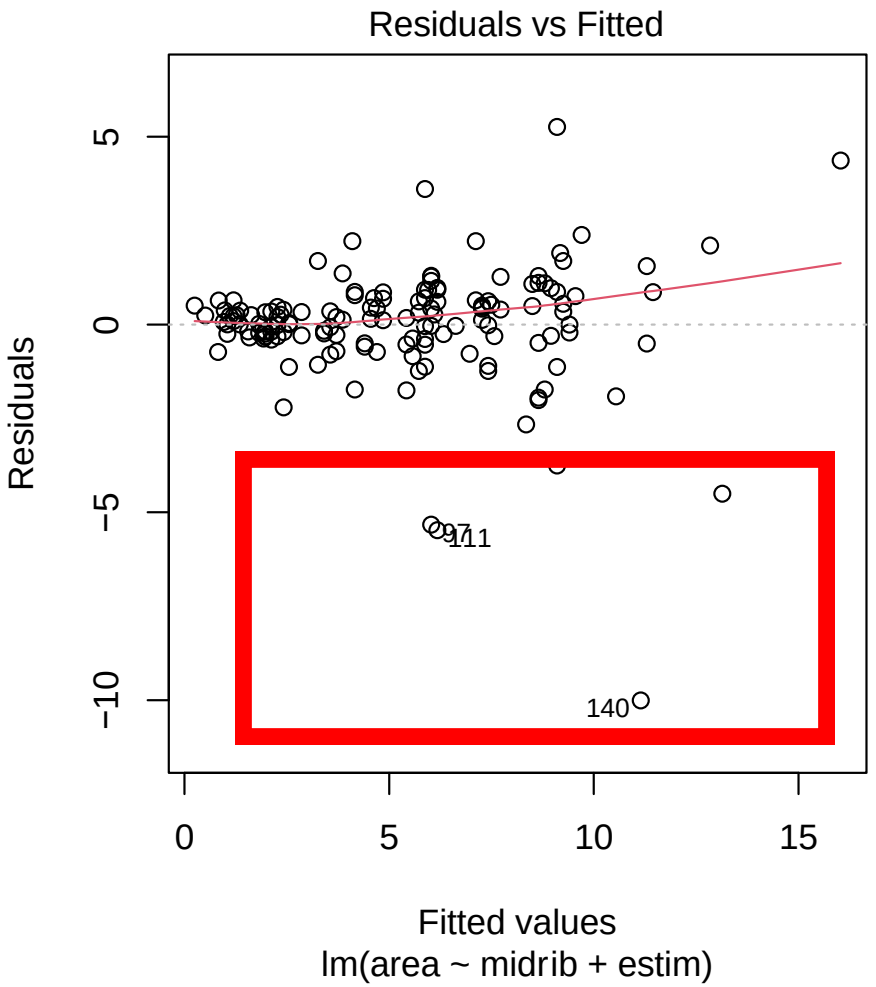
`plot(model, which=1)`

- What it tells us:
  - Points with large residual
    - Problem: Potential outlier (need to verify with Cook's distance)
  - A non-horizontal trend in residuals (red trend line isn't horizontal near residual=0)
    - Problem: missing intercept or Y not linearly related to X (violation of assumption I).
  - A pattern in the residuals
    - Problem: Correlated random error. Violation of assumption (IV).

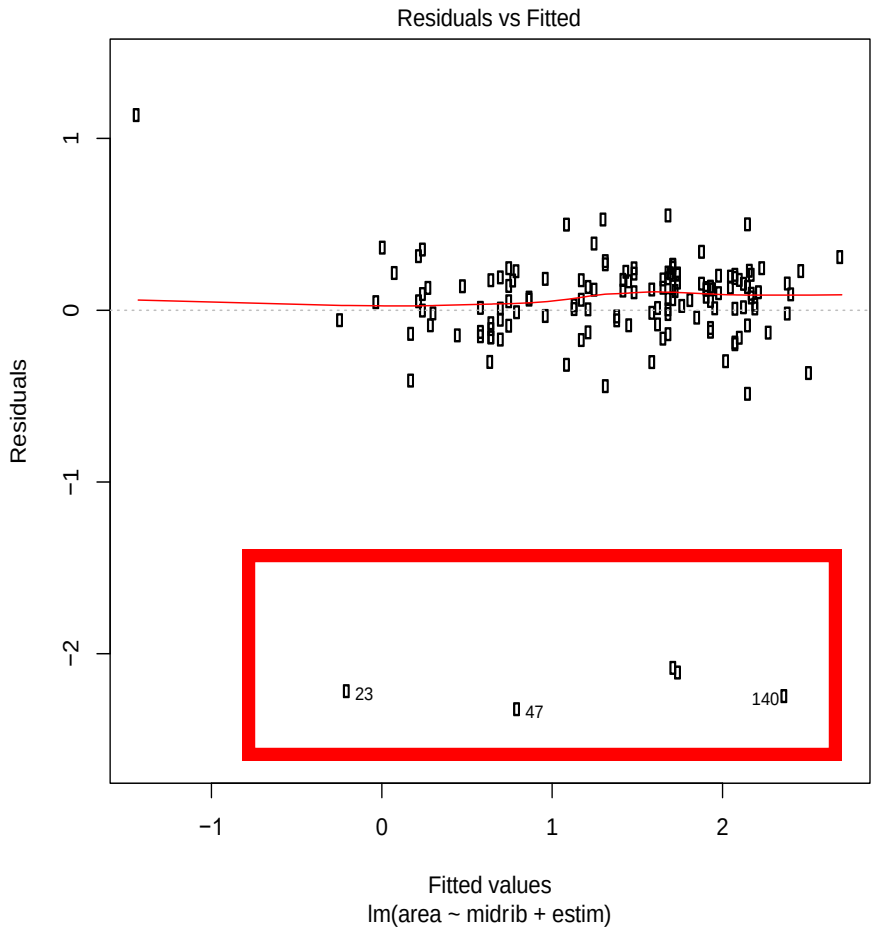
Comparing residual-fitted value plots

`plot(model, which=1)`

Before log-transform



After log-transform

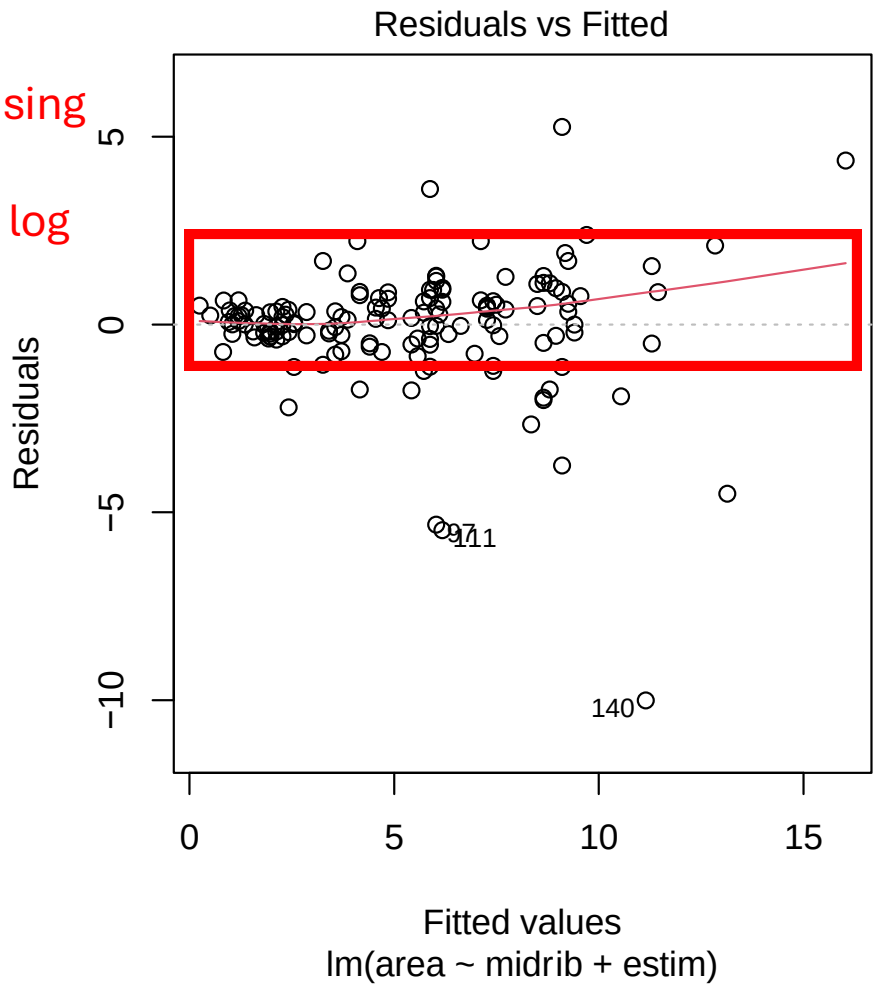


Large residuals occur before and after log-transform!

Comparing residual-fitted value plots

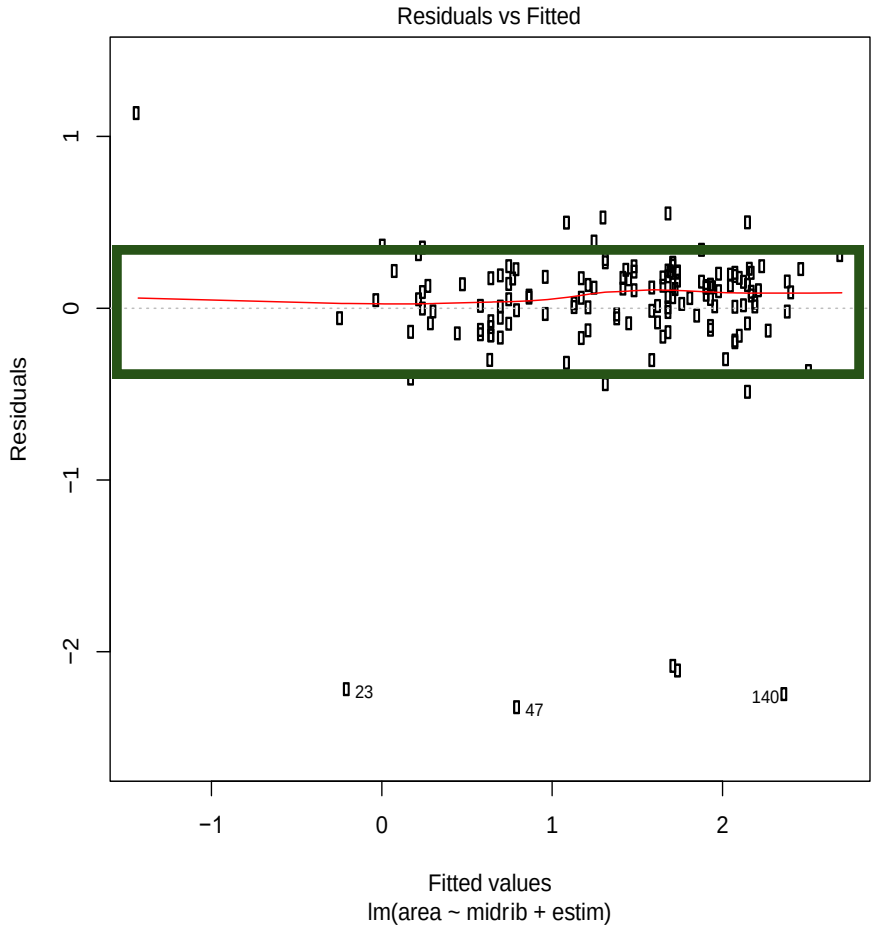
`plot(model, which=1)`

Before log-transform



Slight increasing exponential trend before log transform

After log-transform



Trendline is almost horizontal around residual = 0 after log-transform

QQ-normal plot for standardised residuals

`plot(model, which=2)`

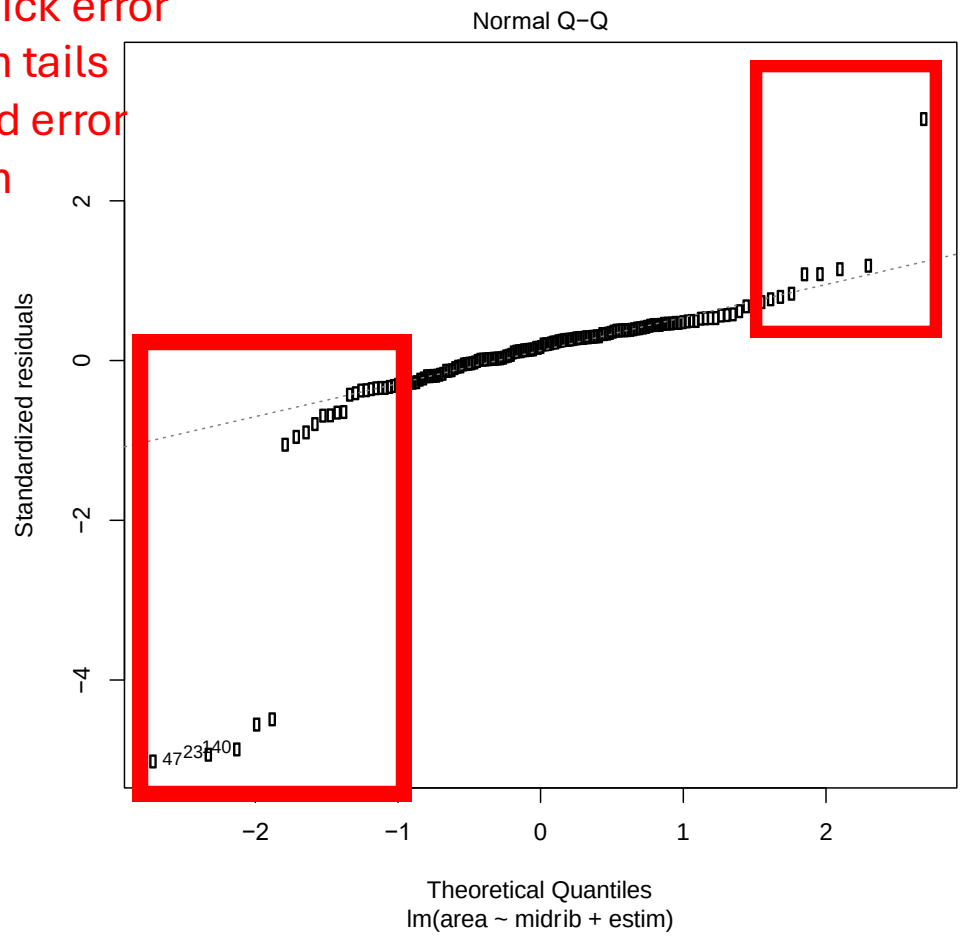
- What it tells us:
  - A single extreme quantile is far from the dotted line
    - Problem: Potential outlier (need to verify with Cook's distance)
  - Highest and lowest empirical quantiles are far from dotted line
    - Problem: Over- or under- estimation of error distribution tails.
  - Points below the dotted line are further away from line than points above dotted line.
    - Random error follows a skewed distribution.

QQ-normal plot for standardised residuals

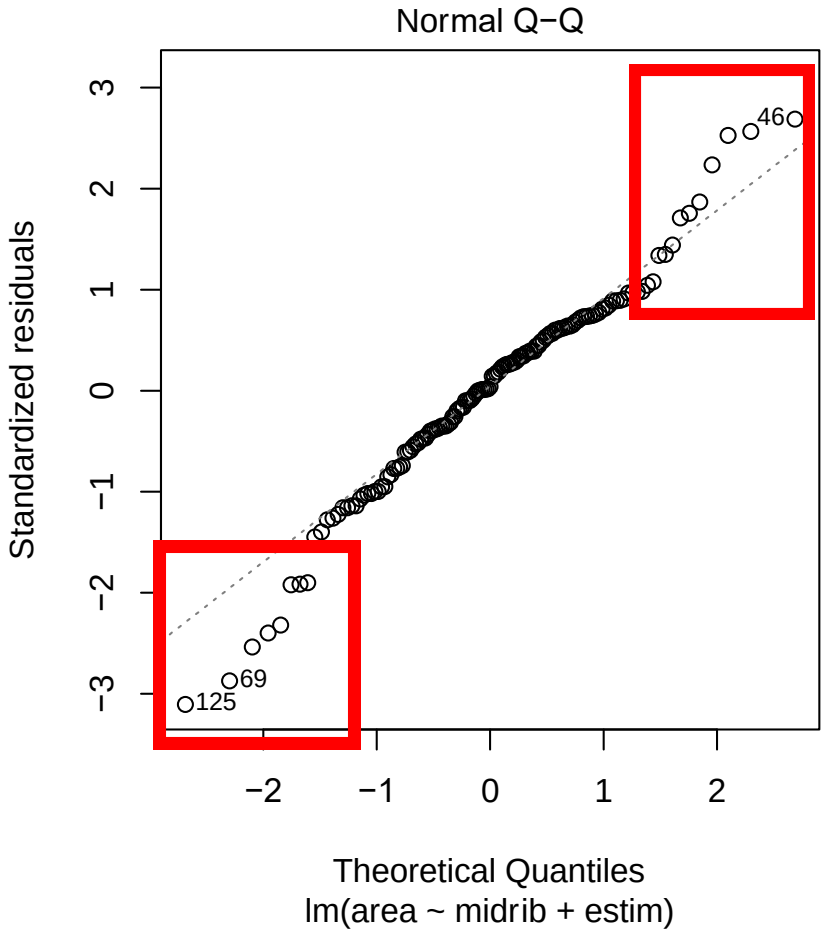
`plot(model, which=2)`

Before removing outliers, we observe two issues: unsually thick error distribution tails and skewed error distribution

Before removing outlier



After removing outliers



After deleting outliers, there is still some evidence of unusually thick error distribution tails!

But evidence of skewness is no longer present.

Scale-location

```
plot(model, which=3)
```

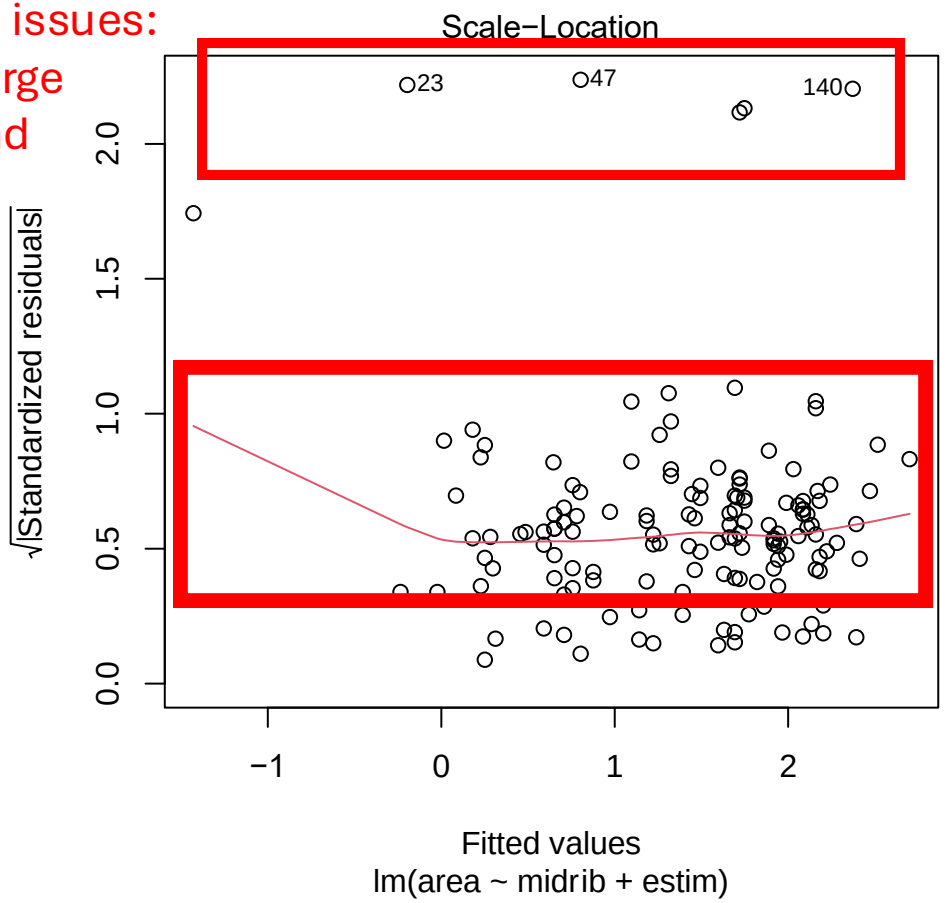
- What it tells us:
  - A single unusually high sqrt-standardised residual
    - Problem: Potential outlier (need to verify with Cook's distance)
  - Non-horizontal trendline (trendline isn't horizontal)
    - Problem: non-constant error variance. Violation of assumption (IV).

Scale-location

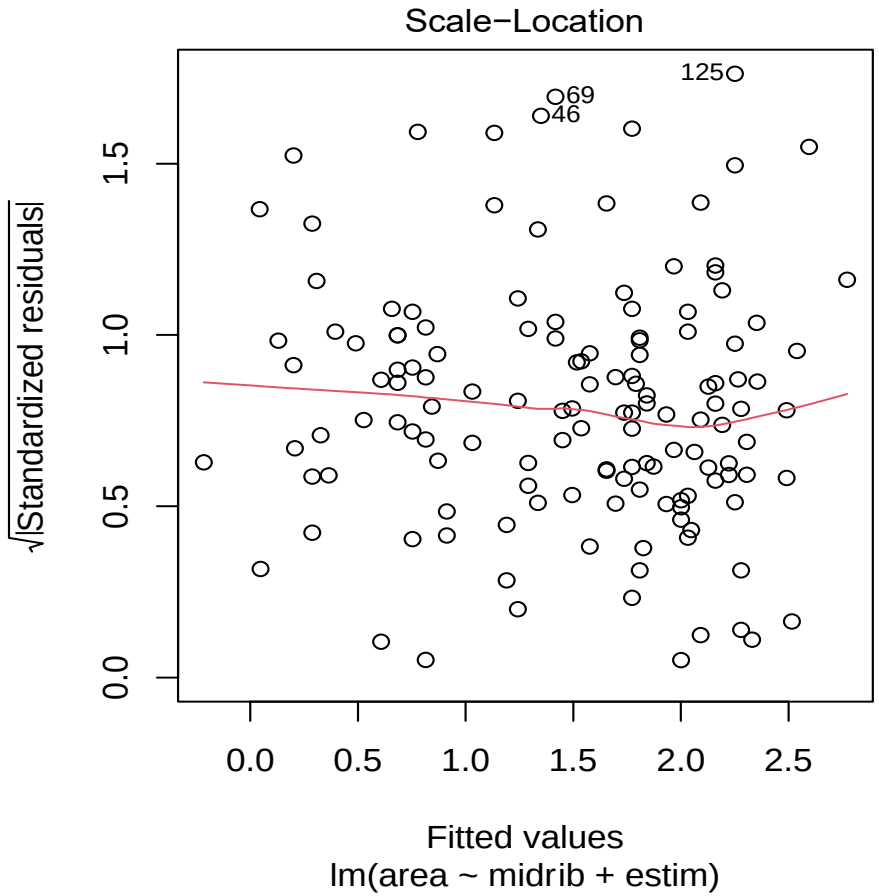
```
plot(model, which=3)
```

Before removing outliers, we observe two issues: unusually large residuals and

Before removing outlier



After removing outliers



After removing outliers, both issues are resolved

Cook's distance plot

`plot(model, which=4)`

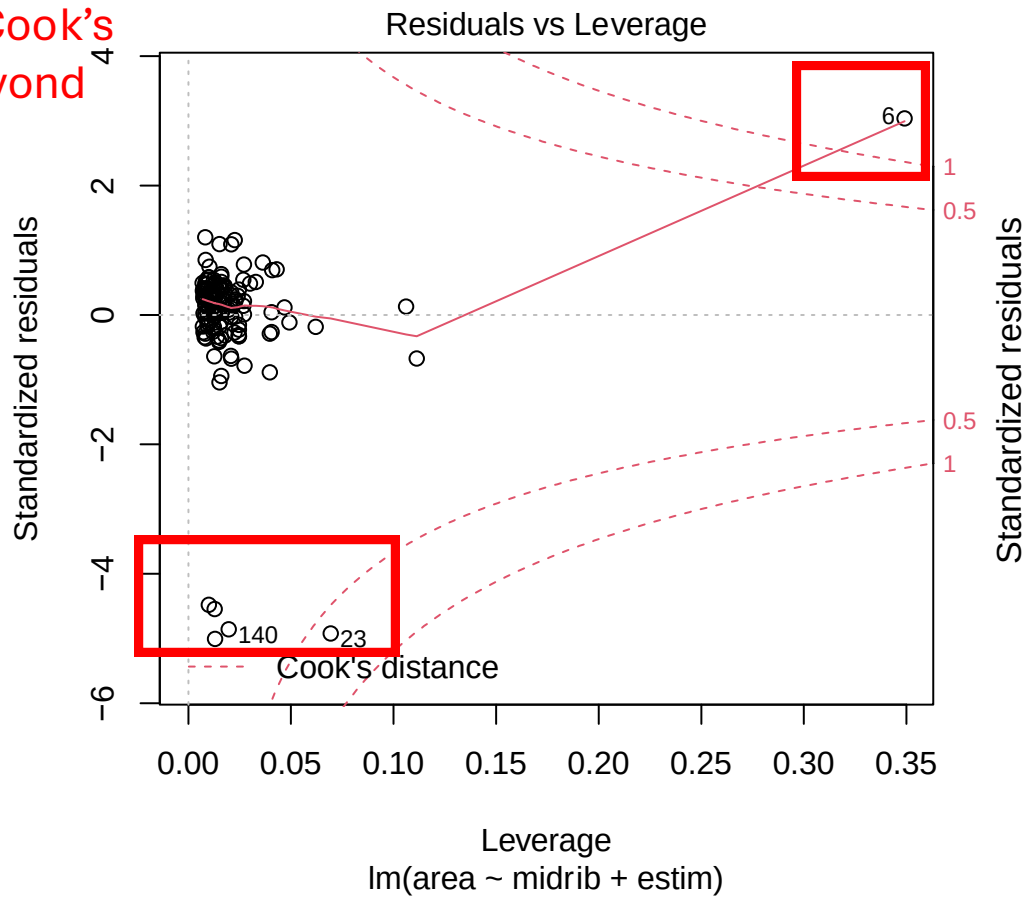
- Horizontal axis = leverage
- Vertical axis = standardised residual
- Contour = Cook's distance
- What it tells us:
  - Points lie beyond high Cook's distance contours
    - Problem: Strong evidence for an outlier

Cook’s distance

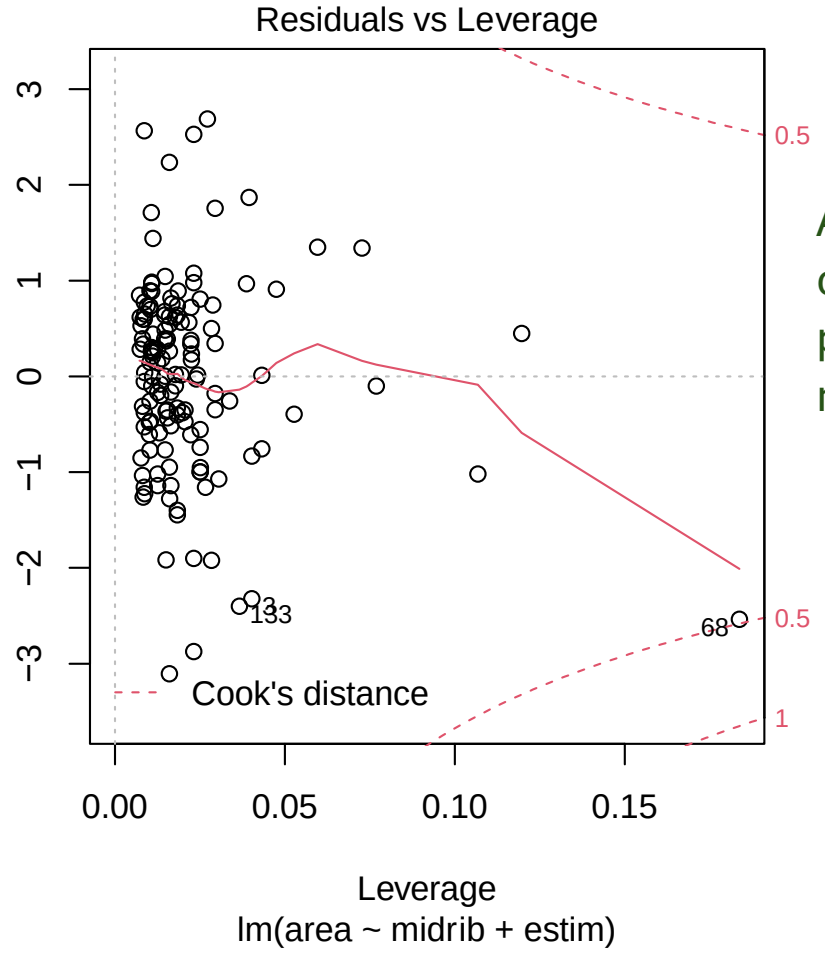
```
plot(model, which=4)
```

Before removing outliers, several points with Cook’s D near or beyond 0.5-contour

Before removing outlier



After removing outliers



After removing outliers, only one point with Cook’s D near 0.5-contour