Black/Asian carp model selection

Eddie Wu

2023-07-20

Introduction

This .Rmd file is to show the progress on asian carp temperature and condition analyses. Since sub-sampling from spatial autocorrelation does not give significantly different results from normal analysis, we present the results without sub-sampling here.

For each species, we have four different models:

1. Simple linear model (same slope, same intercept)
2. Linear additive model (same slope, different intercept)
3. Interaction model (different slope, same intercept)
4. Group-specific model (different slope, different intercept)

And we consider two temperature metrics:

1. Annual temperature
2. Winter temperature (temperature from the coldest quarter)

```
library(ggplot2)
library(ggfortify)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 4.1.3
```

```
## Import data
asian.carp <- read.csv("asian carp final.csv")
asian.carp$Condition <- as.factor(asian.carp$Condition)

Black <- read.csv("eddie_carp_new.csv")
Black$condition <- as.factor(Black$condition)

## Separate by species
Grass <- asian.carp[asian.carp$Species=="Grass",]
Bighead <- asian.carp[asian.carp$Species=="Bighead",]
```

```r
Silver <- asian.carp[asian.carp$Species=="Silver",]
Big.sil <- rbind(Bighead, Silver) # combine the two groups
```

Asian carp

Using annual temperature

```r
# Clean data
asian.carp.clean <- asian.carp %>%
  filter(Condition %in% c("natural", "artificial"))

# Build the models
asian.simple <- lm(log(AAM)~AnnualTemp, data = asian.carp.clean)
asian.linear <- lm(log(AAM)~AnnualTemp+Condition, data = asian.carp.clean)
asian.int <- lm(log(AAM)~AnnualTemp:Condition, data = asian.carp.clean)
asian.group <- lm(log(AAM)~AnnualTemp*Condition, data = asian.carp.clean)

# Compare the AICs
AIC(asian.simple, asian.linear, asian.int, asian.group)
```

```
##              df      AIC
## asian.simple  3 17.86841
## asian.linear  4 19.86757
## asian.int     4 19.84202
## asian.group   5 21.80296
```

```r
# R^2 value for the four models
r_2 <- data.frame(
  Model = c("Simple linear", "Linear additive", "Interaction", "Grouped"),
  R2 = c(summary(asian.simple)$r.squared, summary(asian.linear)$r.squared,
         summary(asian.int)$r.squared, summary(asian.group)$r.squared)
)
kable(r_2)
```

Model	R2
Simple linear	0.5595246
Linear additive	0.5595301
Interaction	0.5596980
Grouped	0.5599546

- For Asian carp using annual air temperature, the simple linear model works the best.

Using cold temperature

```r
# Build the models
asian.simple <- lm(log(AAM)~ColdTemp, data = asian.carp.clean)
asian.linear <- lm(log(AAM)~ColdTemp+Condition, data = asian.carp.clean)
asian.int <- lm(log(AAM)~ColdTemp:Condition, data = asian.carp.clean)
asian.group <- lm(log(AAM)~ColdTemp*Condition, data = asian.carp.clean)

# Compare the AICs
AIC(asian.simple, asian.linear, asian.int, asian.group)
```

```
##              df      AIC
## asian.simple  3 24.36562
## asian.linear  4 26.20637
## asian.int     4 26.34968
## asian.group   5 28.17110
```

```
# R^2 value for the four models
r_2 <- data.frame(
  Model = c("Simple linear", "Linear additive", "Interaction", "Grouped"),
  R2 = c(summary(asian.simple)$r.squared, summary(asian.linear)$r.squared,
         summary(asian.int)$r.squared, summary(asian.group)$r.squared)
)
kable(r_2)
```

Model	R2
Simple linear	0.5146705
Linear additive	0.5158227
Interaction	0.5147860
Grouped	0.5160775

- Same conclusion when using the cold temperature. But using annual temperature seems to have a better model as explained by lower AIC values.

Grass carp

Using annual temperature

```
# Clean data
grass.clean <- Grass %>%
  filter(Condition %in% c("natural", "artificial"))

# Build the models
grass.simple <- lm(log(AAM)~AnnualTemp, data = grass.clean)
grass.linear <- lm(log(AAM)~AnnualTemp+Condition, data = grass.clean)
grass.int <- lm(log(AAM)~AnnualTemp:Condition, data = grass.clean)
grass.group <- lm(log(AAM)~AnnualTemp*Condition, data = grass.clean)

# Compare the AICs
AIC(grass.simple, grass.linear, grass.int, grass.group)
```

```
##              df       AIC
## grass.simple  3  9.848057
## grass.linear  4 10.801064
## grass.int     4 10.759725
## grass.group   5 12.663058
```

```
# R^2 value for the four models
r_2 <- data.frame(
  Model = c("Simple linear", "Linear additive", "Interaction", "Grouped"),
  R2 = c(summary(grass.simple)$r.squared, summary(grass.linear)$r.squared,
         summary(grass.int)$r.squared, summary(grass.group)$r.squared)
)
kable(r_2)
```

Model	R2
Simple linear	0.6474882
Linear additive	0.6578771
Interaction	0.6582810
Grouped	0.6592235

- For grass carp using annual air temperature, the simple linear model works the best. Although there does not seem to be much difference among the models.

Using cold temperature

```
# Build the models
grass.simple <- lm(log(AAM)~ColdTemp, data = grass.clean)
grass.linear <- lm(log(AAM)~ColdTemp+Condition, data = grass.clean)
grass.int <- lm(log(AAM)~ColdTemp:Condition, data = grass.clean)
grass.group <- lm(log(AAM)~ColdTemp*Condition, data = grass.clean)

# Compare the AICs
AIC(grass.simple, grass.linear, grass.int, grass.group)
```

```
##              df       AIC
## grass.simple  3  9.573623
## grass.linear  4 11.500450
## grass.int     4 11.407066
## grass.group   5 13.293477
```

```
# R^2 value for the four models
r_2 <- data.frame(
  Model = c("Simple linear", "Linear additive", "Interaction", "Grouped"),
  R2 = c(summary(grass.simple)$r.squared, summary(grass.linear)$r.squared,
         summary(grass.int)$r.squared, summary(grass.group)$r.squared)
)
kable(r_2)
```

Model	R2
Simple linear	0.6502415
Linear additive	0.6509719
Interaction	0.6519019
Grouped	0.6530298

- Same conclusion as using annual temperature: the simple linear model works the best.
- While models using cold temperature does not show any significant differences compared to using annual temperature (AICs <= 2).

Bighead and silver carp

Using annual temperature

```
# Clean data
big.sil.clean <- Big.sil %>%
  filter(Condition %in% c("natural", "artificial"))
```

```r
# Build the models
big.sil.simple <- lm(log(AAM)~AnnualTemp, data = big.sil.clean)
big.sil.linear <- lm(log(AAM)~AnnualTemp+Condition, data = big.sil.clean)
big.sil.int <- lm(log(AAM)~AnnualTemp:Condition, data = big.sil.clean)
big.sil.group <- lm(log(AAM)~AnnualTemp*Condition, data = big.sil.clean)

# Compare the AICs
AIC(big.sil.simple, big.sil.linear, big.sil.int, big.sil.group)
```

```
##                df      AIC
## big.sil.simple  3 12.16249
## big.sil.linear  4 12.52862
## big.sil.int     4 12.65871
## big.sil.group   5 14.49020
```

```r
# R^2 value for the four models
r_2 <- data.frame(
  Model = c("Simple linear", "Linear additive", "Interaction", "Grouped"),
  R2 = c(summary(big.sil.simple)$r.squared, summary(big.sil.linear)$r.squared,
         summary(big.sil.int)$r.squared, summary(big.sil.group)$r.squared)
)
kable(r_2)
```

Model	R2
Simple linear	0.4299145
Linear additive	0.4582916
Interaction	0.4560849
Grouped	0.4589415

- For bighead and silver carp using annual air temperature, the simple linear model works the best. Although there does not seem to be much difference among the models.

Using cold temperature

```r
# Build the models
big.sil.simple <- lm(log(AAM)~ColdTemp, data = big.sil.clean)
big.sil.linear <- lm(log(AAM)~ColdTemp+Condition, data = big.sil.clean)
big.sil.int <- lm(log(AAM)~ColdTemp:Condition, data = big.sil.clean)
big.sil.group <- lm(log(AAM)~ColdTemp*Condition, data = big.sil.clean)

# Compare the AICs
AIC(big.sil.simple, big.sil.linear, big.sil.int, big.sil.group)
```

```
##                df      AIC
## big.sil.simple  3 17.48189
## big.sil.linear  4 18.24104
## big.sil.int     4 19.47980
## big.sil.group   5 20.23999
```

```r
# R^2 value for the four models
r_2 <- data.frame(
  Model = c("Simple linear", "Linear additive", "Interaction", "Grouped"),
  R2 = c(summary(big.sil.simple)$r.squared, summary(big.sil.linear)$r.squared,
         summary(big.sil.int)$r.squared, summary(big.sil.group)$r.squared)
```

```
)
kable(r_2)
```

Model	R2
Simple linear	0.3268168
Linear additive	0.3524208
Interaction	0.3268606
Grouped	0.3524420

- Same conclusion as annual temperature. Annual temperature fits better models (significant differences compared to using cold temperature).
- R^2 seems to be much lower.

Black carp

Using annual temperature

```
# Clean data
black.clean <- Black %>% filter(!row_number() == 15) %>% filter(sex != "male")

# Build the models
black.simple <- lm(log(AAM)~AnnualTemp, data = black.clean)
black.linear <- lm(log(AAM)~AnnualTemp+condition, data = black.clean)
black.int <- lm(log(AAM)~AnnualTemp:condition, data = black.clean)
black.group <- lm(log(AAM)~AnnualTemp*condition, data = black.clean)

# Compare the AICs
AIC(black.simple, black.linear, black.int, black.group)
```

```
##              df       AIC
## black.simple  3 -3.487813
## black.linear  4 -2.617656
## black.int     4 -3.407818
## black.group   5 -1.473474
```

```
# R^2 value for the four models
r_2 <- data.frame(
  Model = c("Simple linear", "Linear additive", "Interaction", "Grouped"),
  R2 = c(summary(black.simple)$r.squared, summary(black.linear)$r.squared,
         summary(black.int)$r.squared, summary(black.group)$r.squared)
)
kable(r_2)
```

Model	R2
Simple linear	0.3500632
Linear additive	0.3812190
Interaction	0.4021161
Grouped	0.4038204

- The interaction model seems to perform slightly better than the others, although the difference is not significant.

Using cold temperature

```
# Build the models
black.simple <- lm(log(AAM)~ColdTemp, data = black.clean)
black.linear <- lm(log(AAM)~ColdTemp+condition, data = black.clean)
black.int <- lm(log(AAM)~ColdTemp:condition, data = black.clean)
black.group <- lm(log(AAM)~ColdTemp*condition, data = black.clean)

# Compare the AICs
AIC(black.simple, black.linear, black.int, black.group)
```

```
##              df       AIC
## black.simple  3 -3.887665
## black.linear  4 -3.063173
## black.int     4 -2.784515
## black.group   5 -2.229267
```

```
# R^2 value for the four models
r_2 <- data.frame(
  Model = c("Simple linear", "Linear additive", "Interaction", "Grouped"),
  R2 = c(summary(black.simple)$r.squared, summary(black.linear)$r.squared,
         summary(black.int)$r.squared, summary(black.group)$r.squared)
)
kable(r_2)
```

Model	R2
Simple linear	0.3612646
Linear additive	0.3930896
Interaction	0.3856918
Grouped	0.4230928

- The linear additive model seems to be the best one, while the interaction model is the worst fit.
- Using both temperature metrics give very low R^2 value.

For black carp, if we remove the Southern Ukraine data (AAM = 11) point:

Using annual temperature

```
# Remove the Southern Ukraine data point
black.clean.r <- black.clean[black.clean$AAM != 11,]

# Build the models
black.simple.r <- lm(log(AAM)~AnnualTemp, data = black.clean.r)
black.linear.r <- lm(log(AAM)~AnnualTemp+condition, data = black.clean.r)
black.int.r <- lm(log(AAM)~AnnualTemp:condition, data = black.clean.r)
black.group.r <- lm(log(AAM)~AnnualTemp*condition, data = black.clean.r)

# Compare the AICs
AIC(black.simple.r, black.linear.r, black.int.r, black.group.r)
```

```
##                df       AIC
## black.simple.r  3 -12.35690
## black.linear.r  4 -10.64275
## black.int.r     4 -11.66646
```

```
## black.group.r   5 -10.47263
```

```
# R^2 value for the four models
r_2 <- data.frame(
  Model = c("Simple linear", "Linear additive", "Interaction", "Grouped"),
  R2_before = c(summary(black.simple)$r.squared, summary(black.linear)$r.squared,
          summary(black.int)$r.squared, summary(black.group)$r.squared),
  R2_after = c(summary(black.simple.r)$r.squared, summary(black.linear.r)$r.squared,
          summary(black.int.r)$r.squared, summary(black.group.r)$r.squared)
)
kable(r_2)
```

Model	R2_before	R2_after
Simple linear	0.3612646	0.4251995
Linear additive	0.3930896	0.4326196
Interaction	0.3856918	0.4584163
Grouped	0.4230928	0.4779029

- The linear addictive and interaction model.
- Improved R^2 value compared to before removing the data points.

Using cold temperature

```
# Build the models
black.simple.r <- lm(log(AAM)~ColdTemp, data = black.clean.r)
black.linear.r <- lm(log(AAM)~ColdTemp+condition, data = black.clean.r)
black.int.r <- lm(log(AAM)~ColdTemp:condition, data = black.clean.r)
black.group.r <- lm(log(AAM)~ColdTemp*condition, data = black.clean.r)

# Compare the AICs
AIC(black.simple.r, black.linear.r, black.int.r, black.group.r)
```

```
##                df       AIC
## black.simple.r  3 -14.64766
## black.linear.r  4 -12.90250
## black.int.r     4 -14.42409
## black.group.r   5 -12.84532
```

```
# R^2 value for the four models
r_2 <- data.frame(
  Model = c("Simple linear", "Linear additive", "Interaction", "Grouped"),
    R2_before = c(summary(black.simple)$r.squared, summary(black.linear)$r.squared,
          summary(black.int)$r.squared, summary(black.group)$r.squared),
  R2_after = c(summary(black.simple.r)$r.squared, summary(black.linear.r)$r.squared,
          summary(black.int.r)$r.squared, summary(black.group.r)$r.squared)
)
kable(r_2)
```

Model	R2_before	R2_after
Simple linear	0.3612646	0.4820404
Linear additive	0.3930896	0.4880055
Interaction	0.3856918	0.5222198
Grouped	0.4230928	0.5312809

- Linear additive model.
- Improved R^2 values.

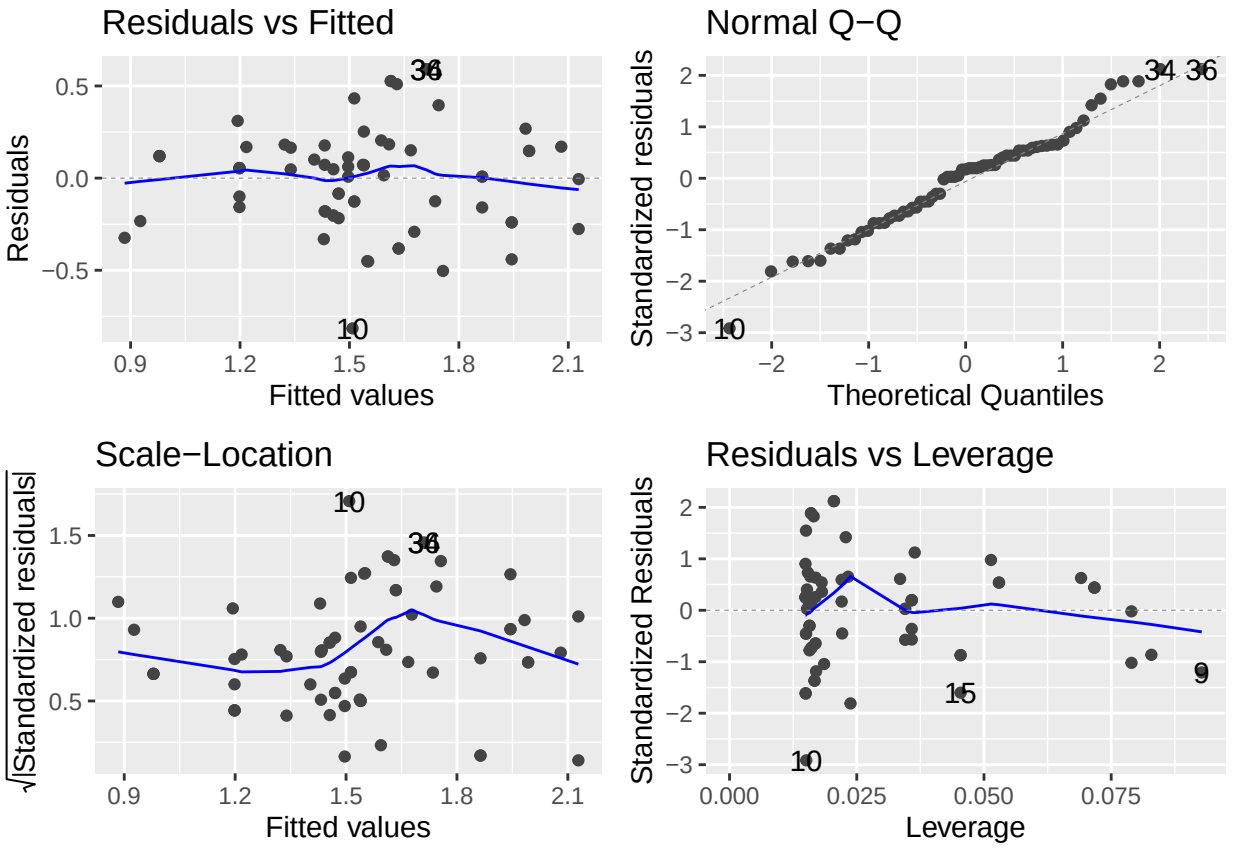
Concluding points

- For all the other asian carp species, the simple linear model seems to be the best fit. Besides, using cold temperature and annual temperature does not produce different results.
- For black carp, the linear additive model (*same slope, different intercept*) seems to be a better fit. Using cold temperature on the original data suggests that the interaction model is the best (probably because the winter temperature at these two locations are very similar.)
- After removing the two data points, we see that the linear additive model still seems to be a better fit. Besides, greatly improved R^2 value suggests better fit.

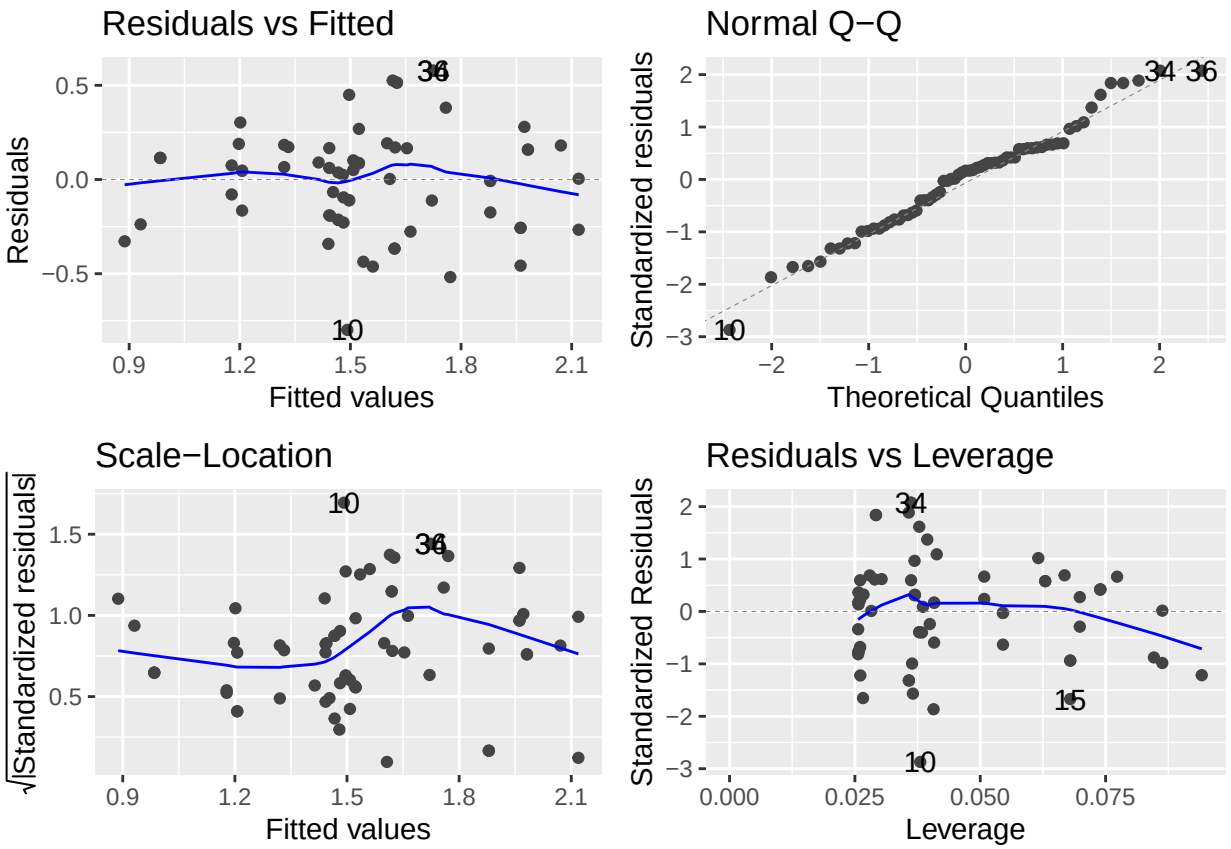
Diagnostic plots

Only the diagnostic plots with cold temperature

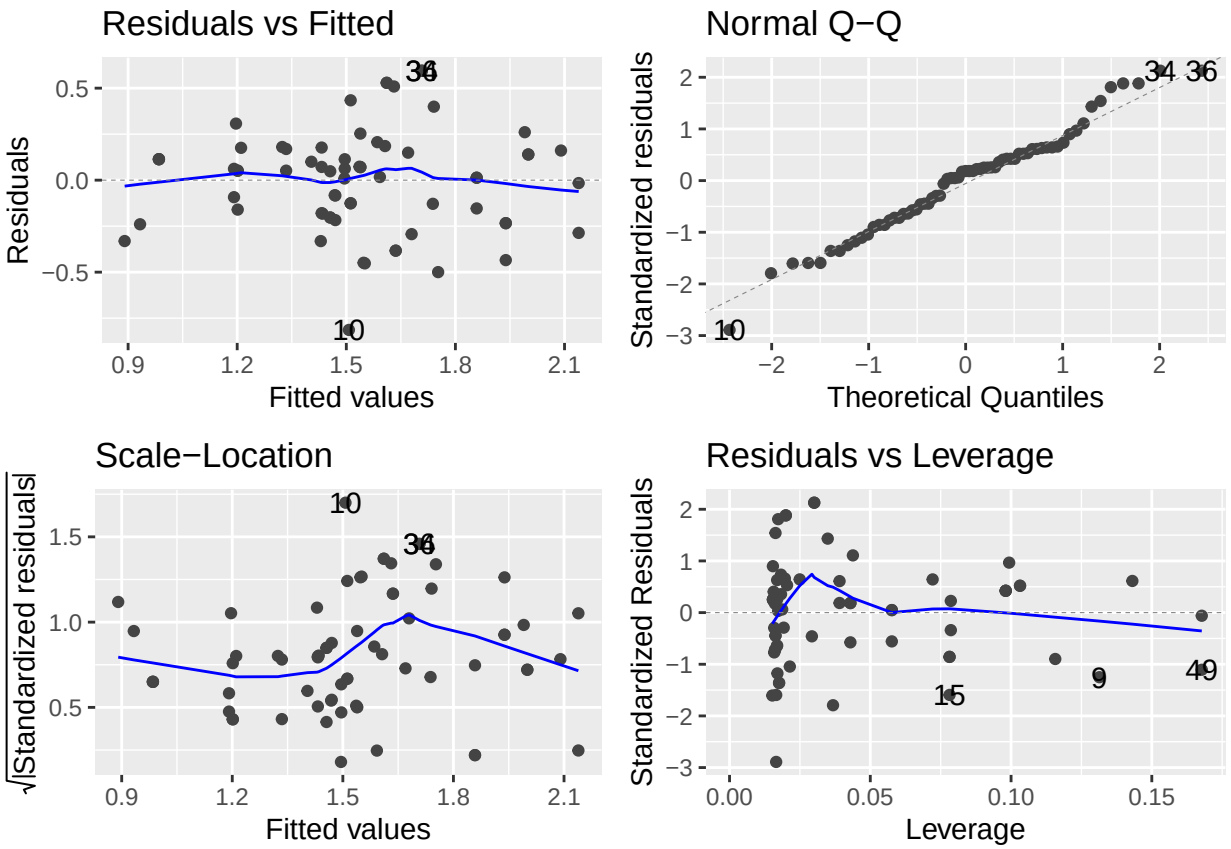
```
# Asian carp

autoplot(asian.simple)
```



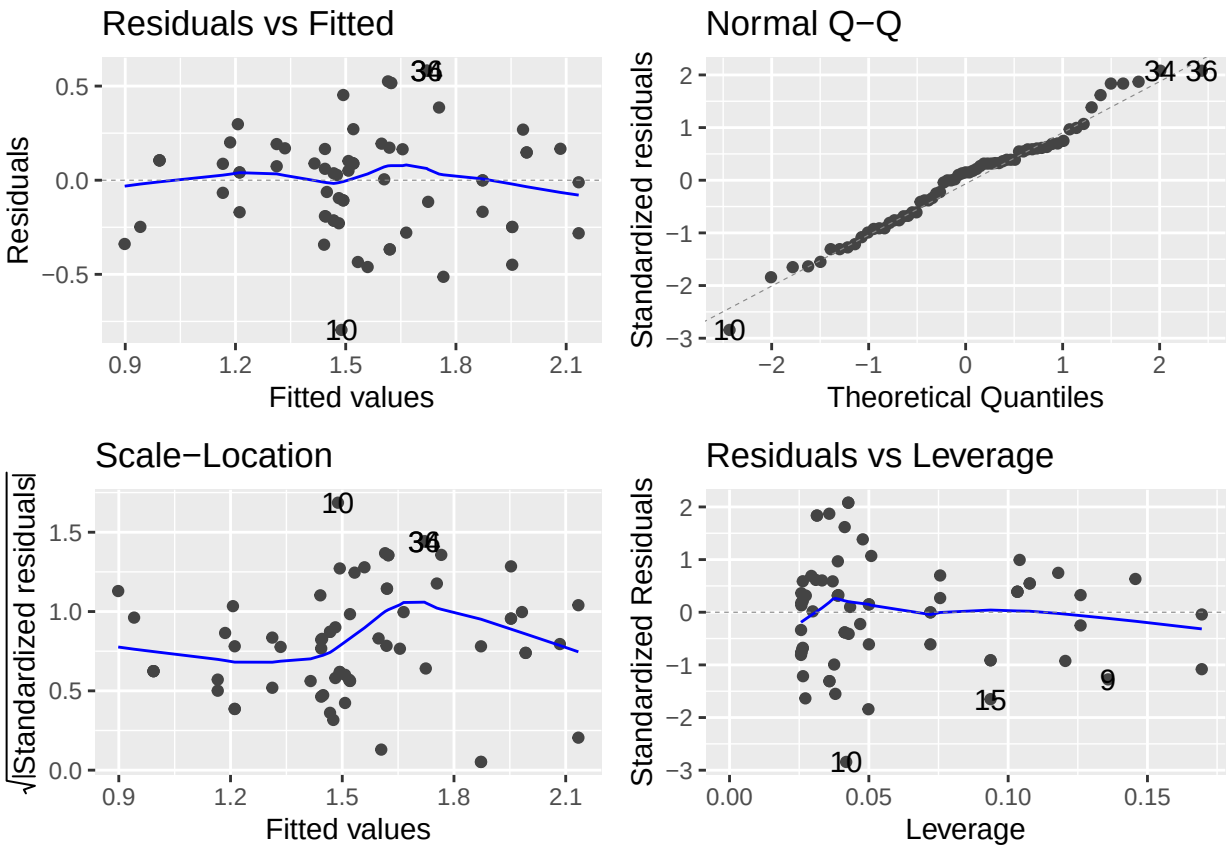

```
autoplot(asian.linear)
```



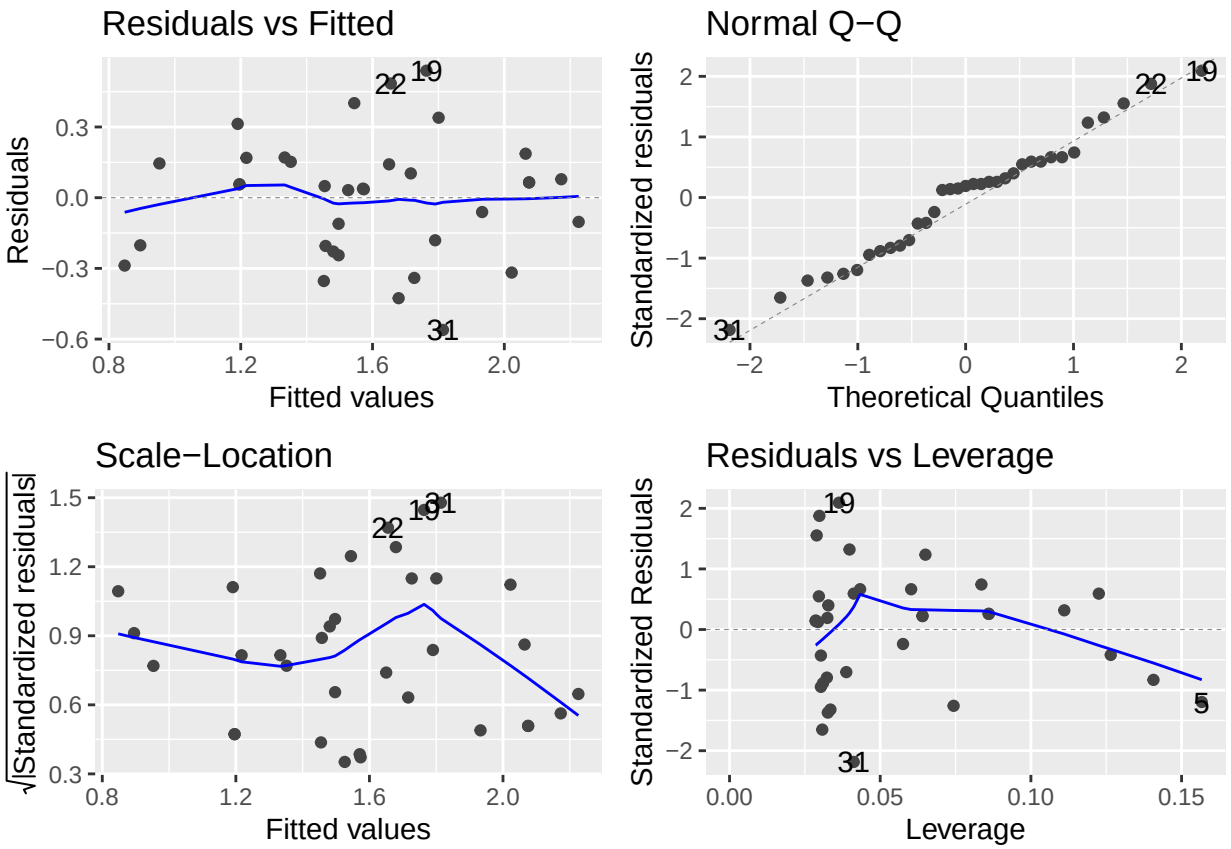

```
autoplot(asian.int)
```



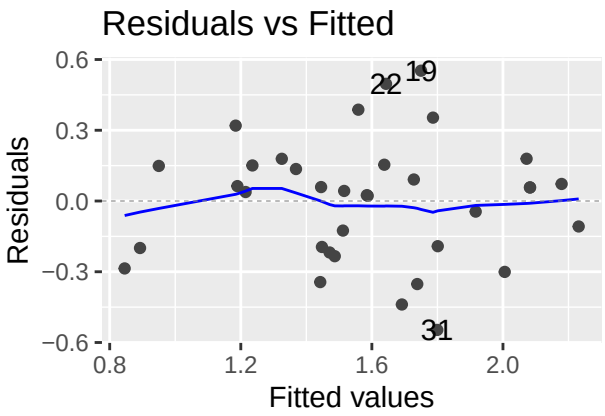

```
autoplot(asian.group)
```

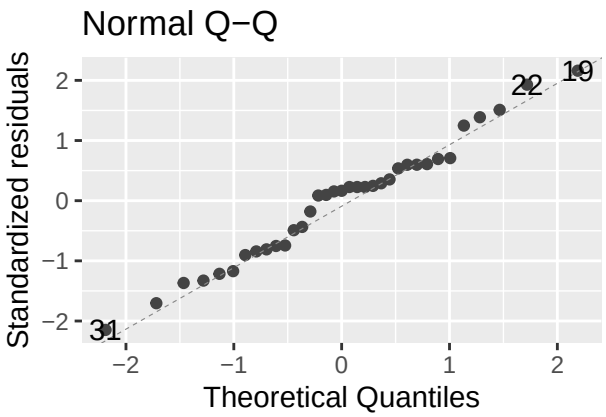


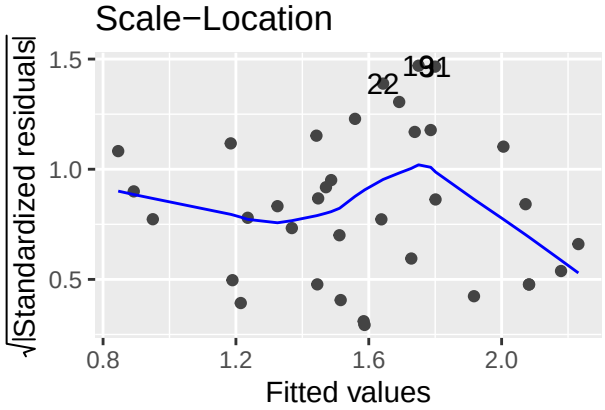

```
# Grass carp
autoplot(grass.simple)
```

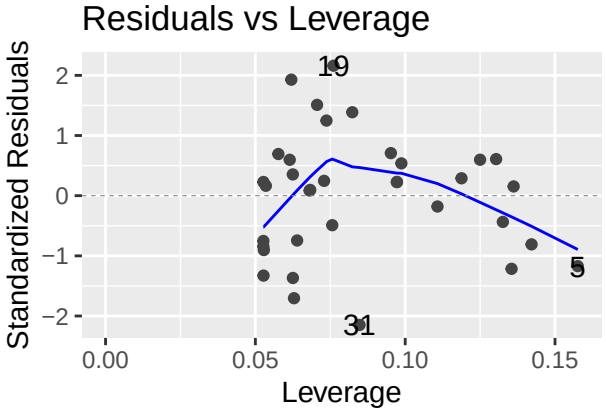



```
autoplot(grass.linear)
```

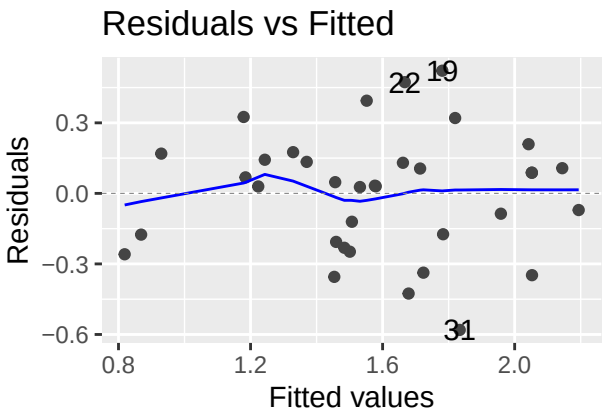


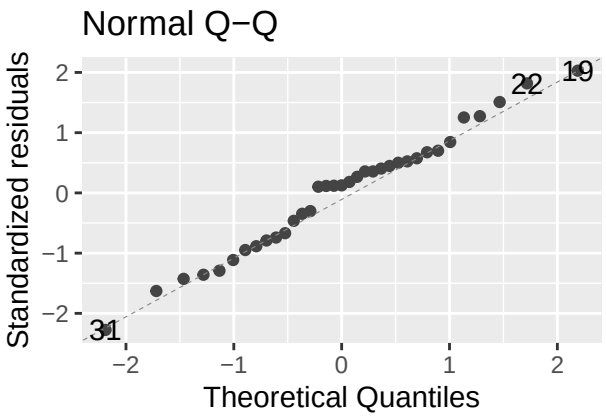


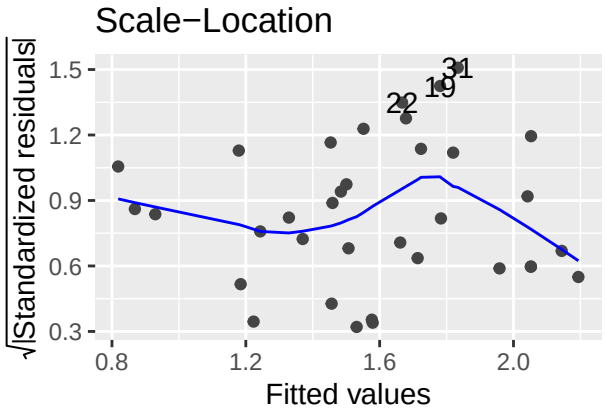


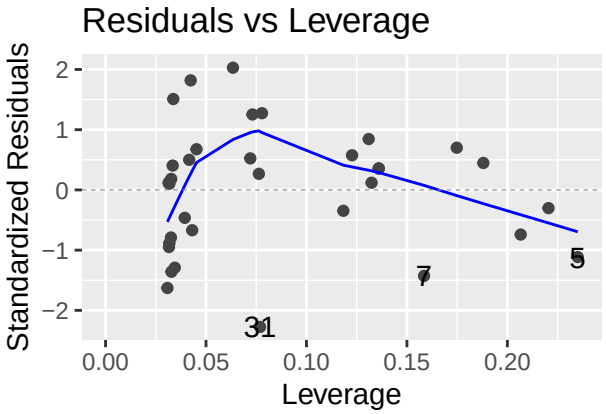



```
autoplot(grass.int)
```

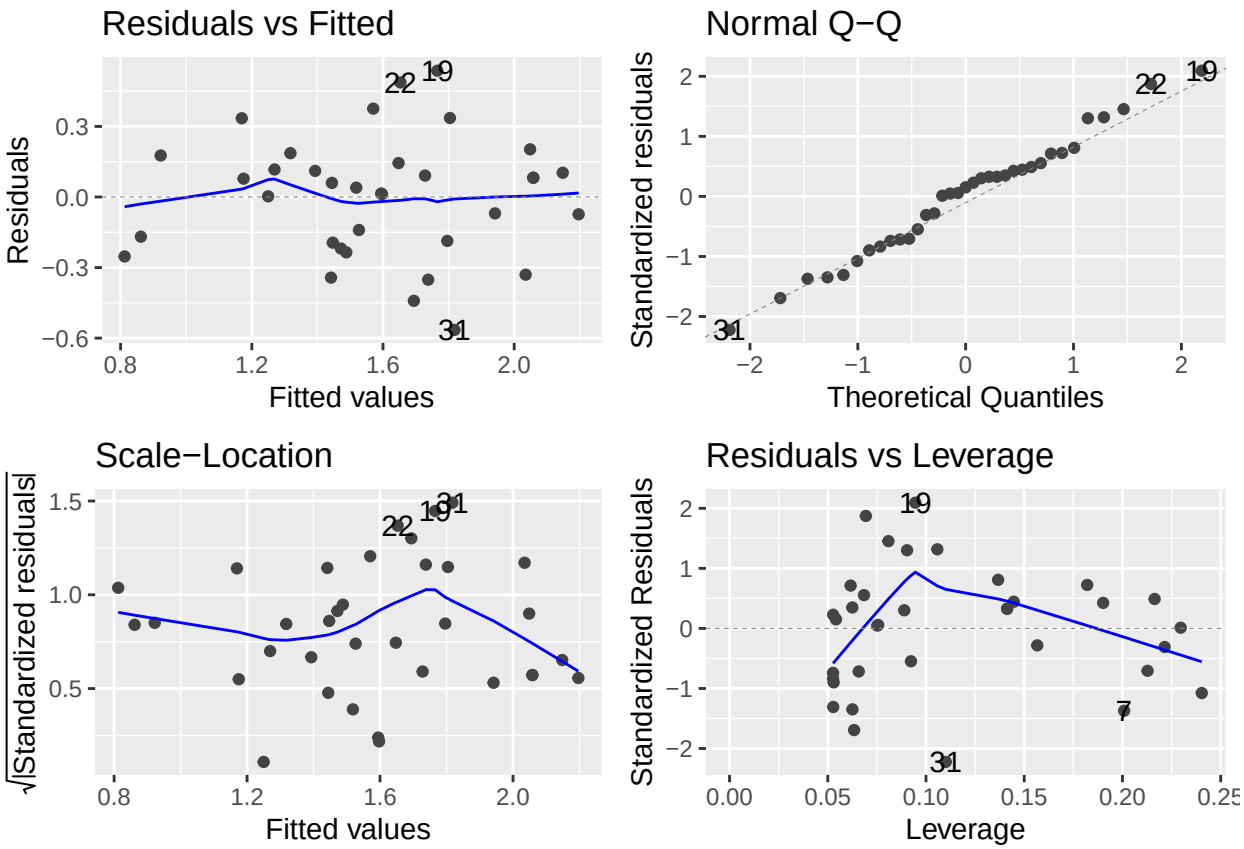




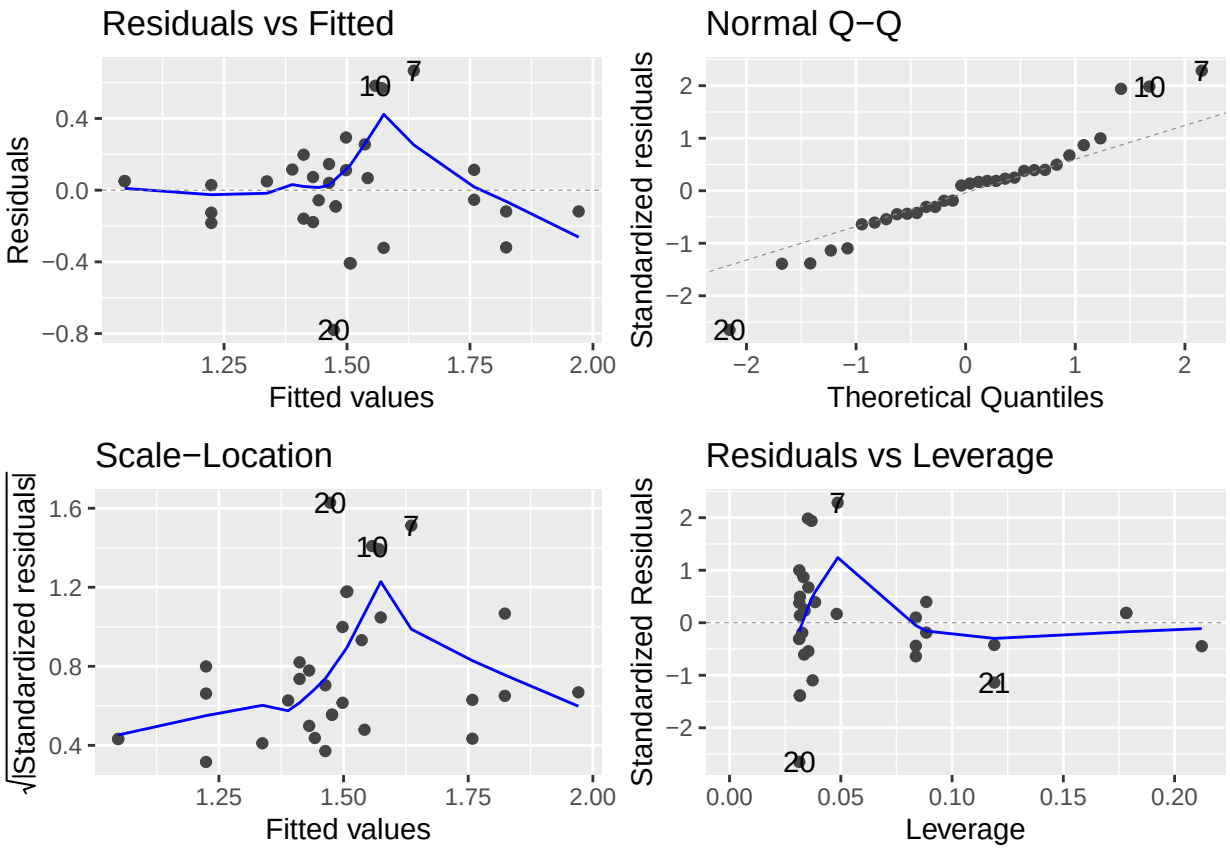




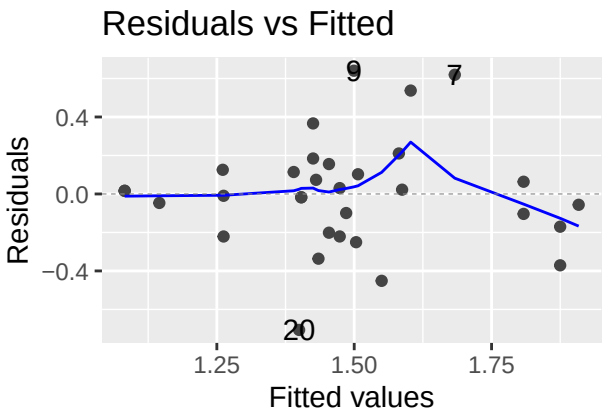

```
autoplot(grass.group)
```

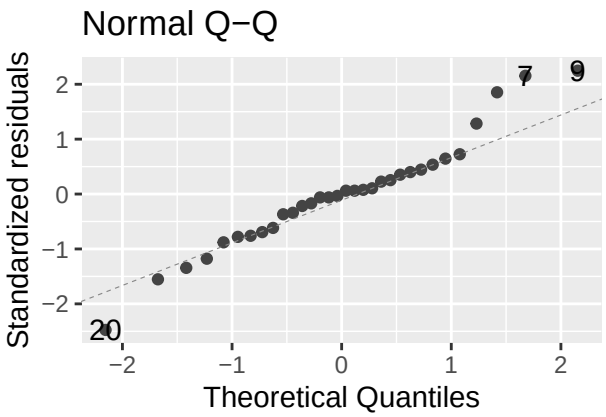


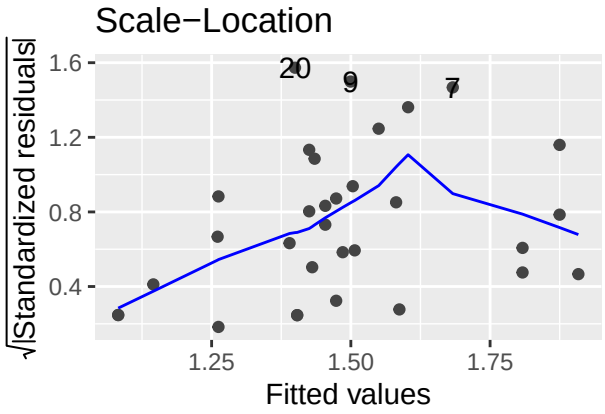

```
# Bighead and silver carp
autoplot(big.sil.simple)
```

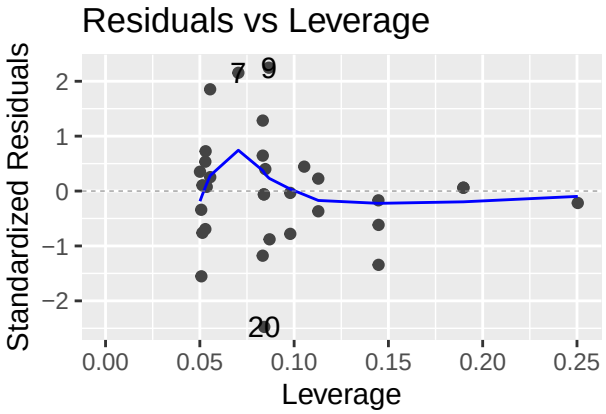



```
autoplot(big.sil.linear)
```

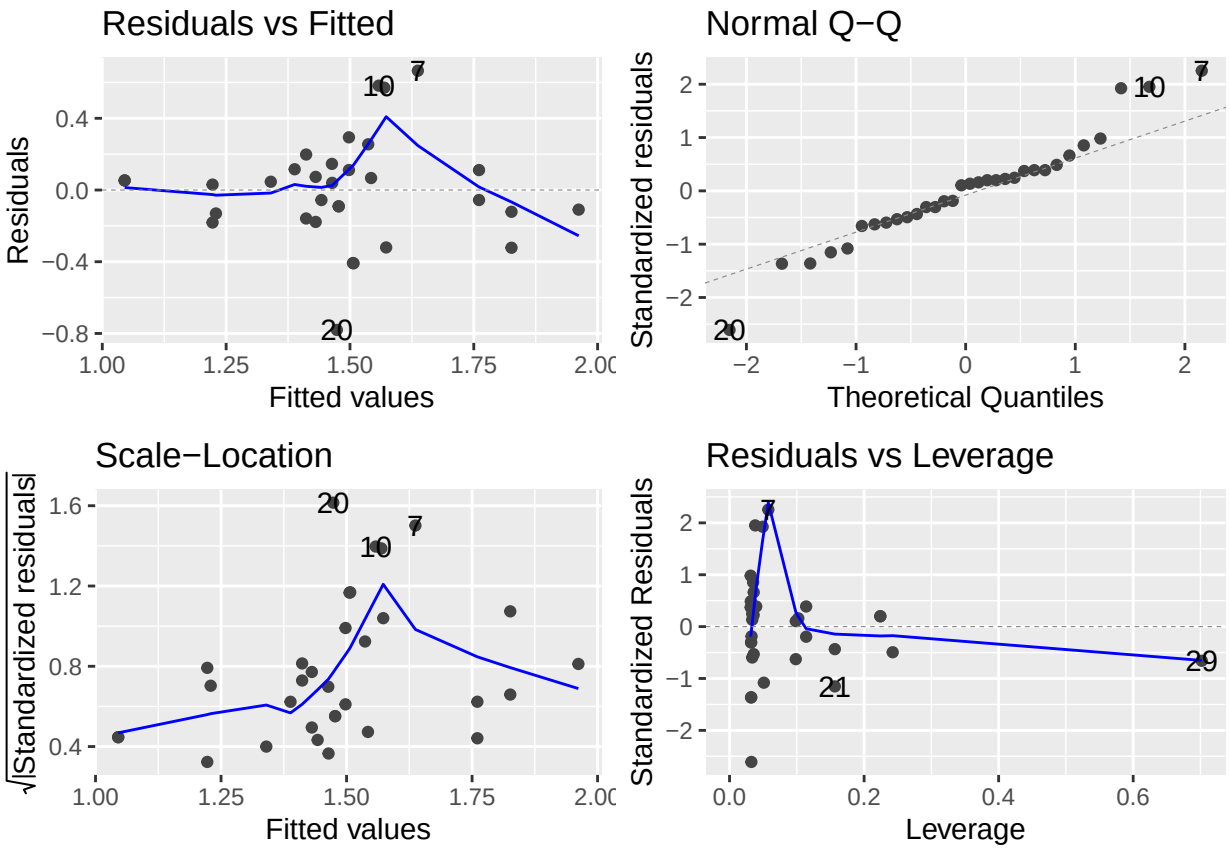




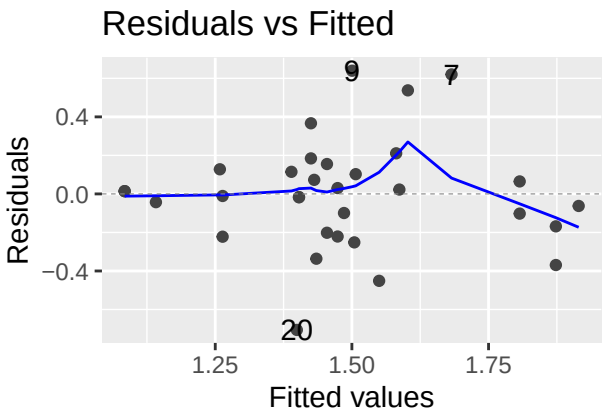


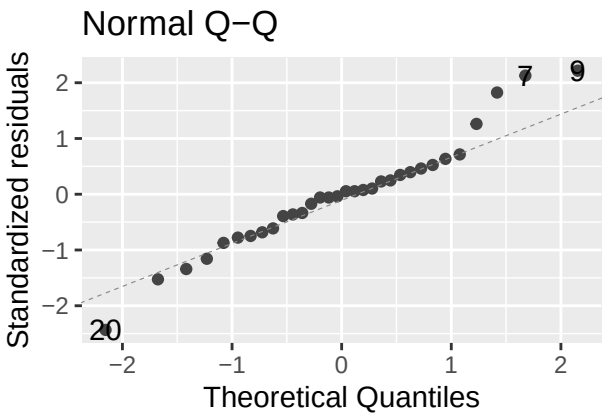


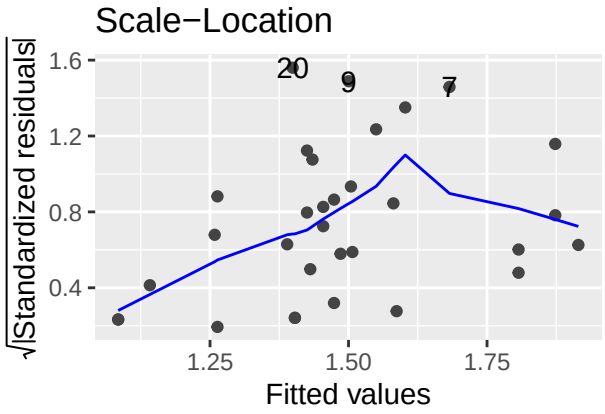

```
autoplot(big.sil.int)
```

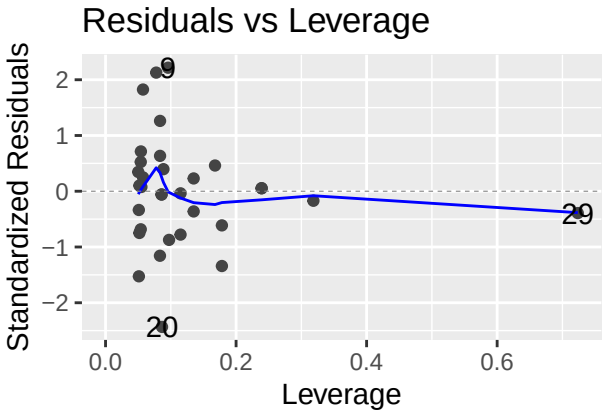



```
autoplot(big.sil.group)
```

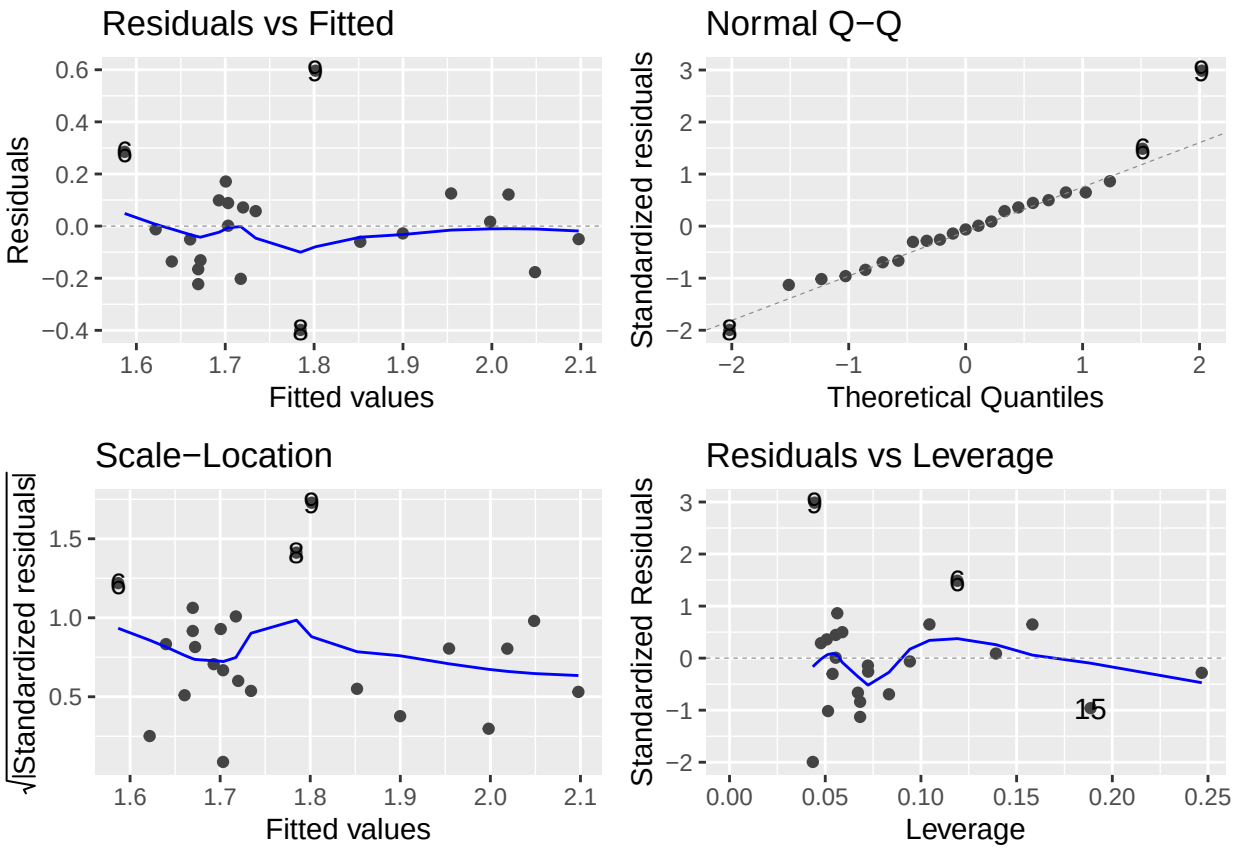




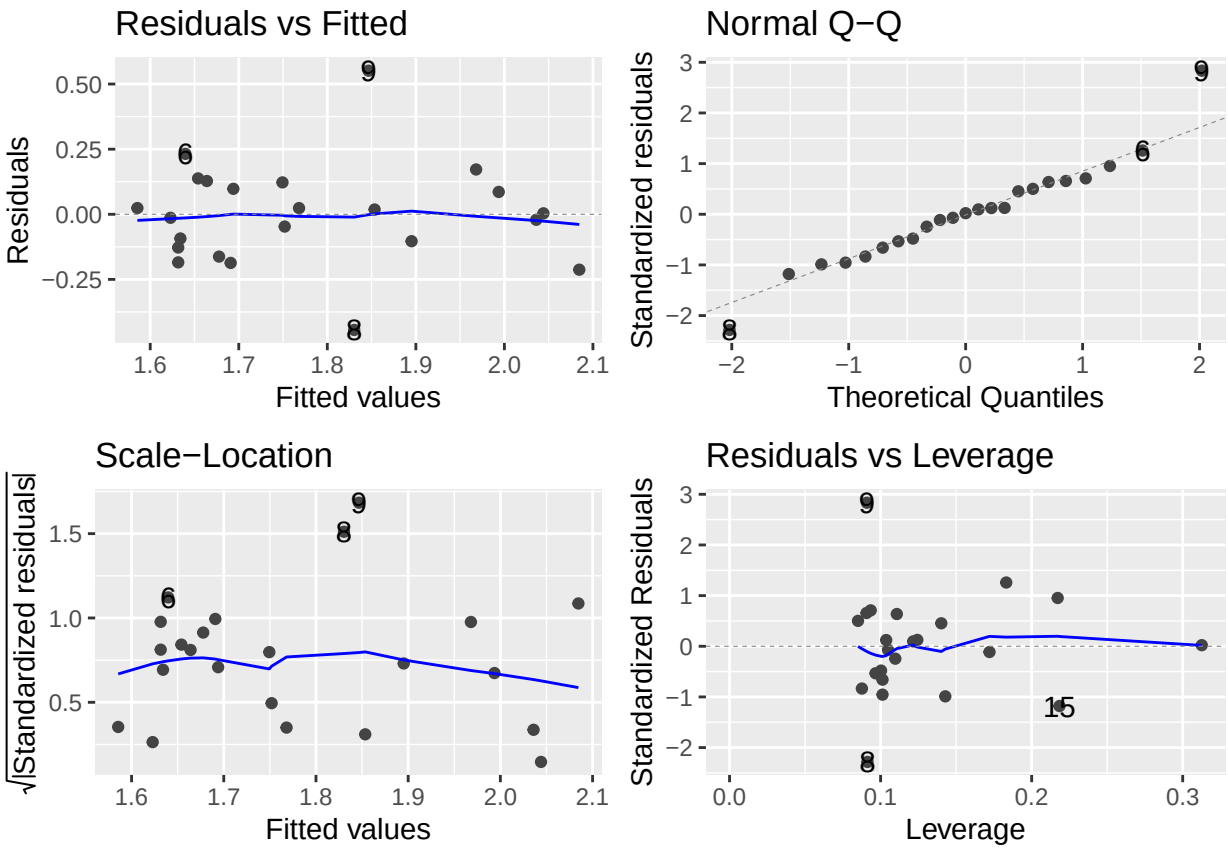




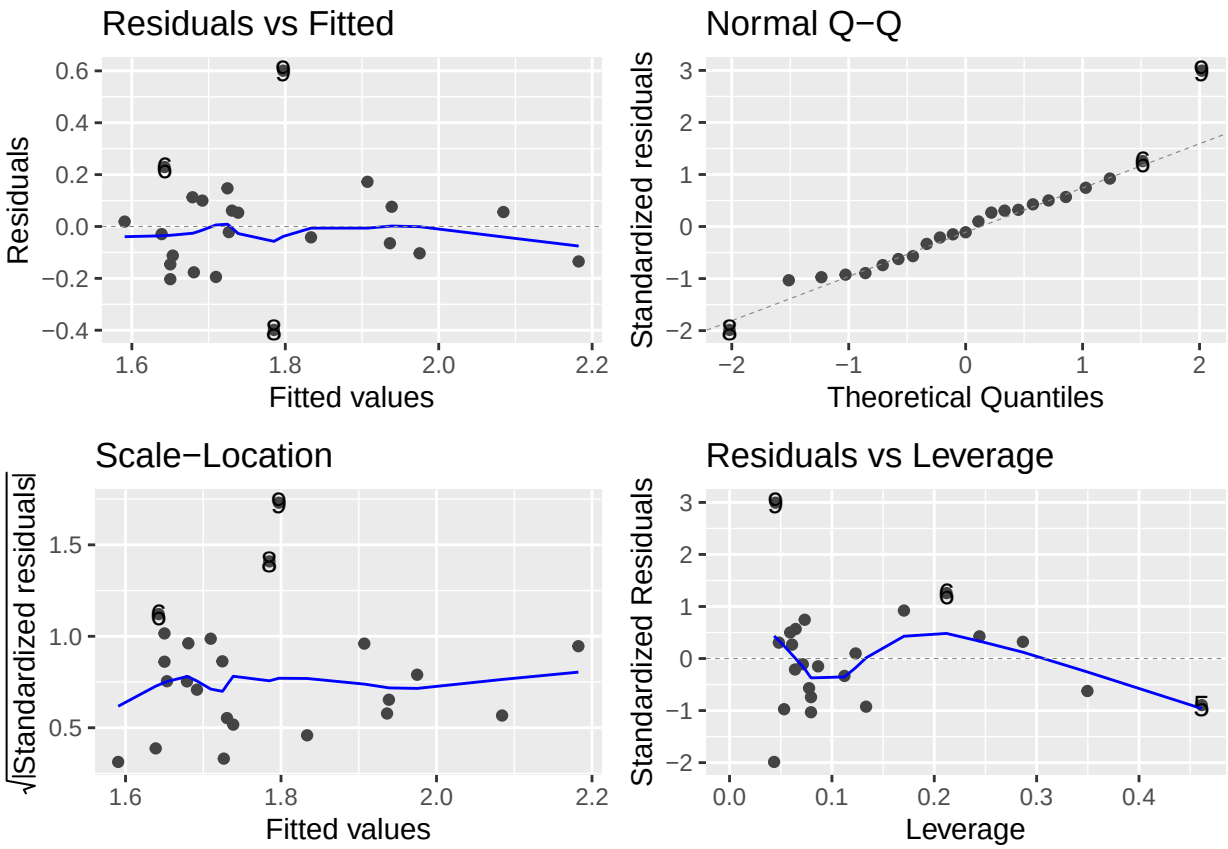

```
# Black carp
autoplot(black.simple)
```



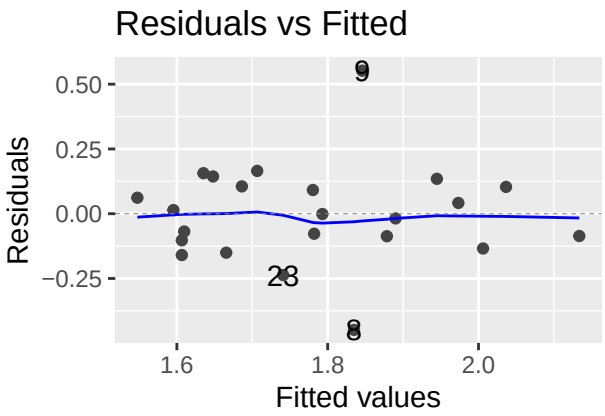

```
autoplot(black.linear)
```

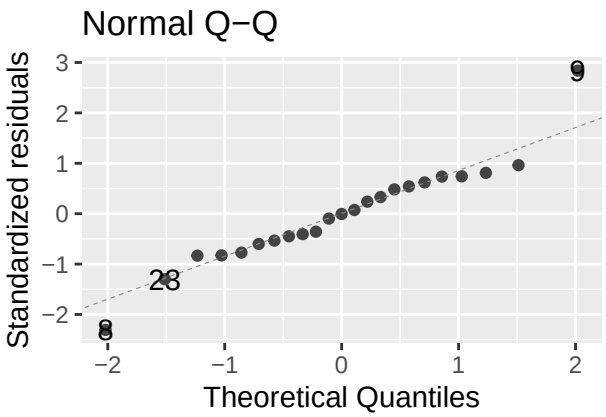


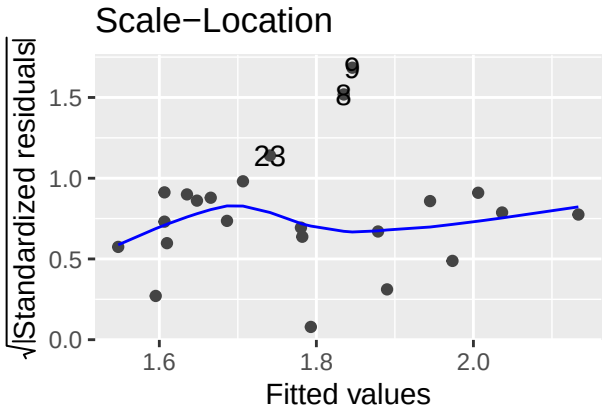

```
autoplot(black.int)
```

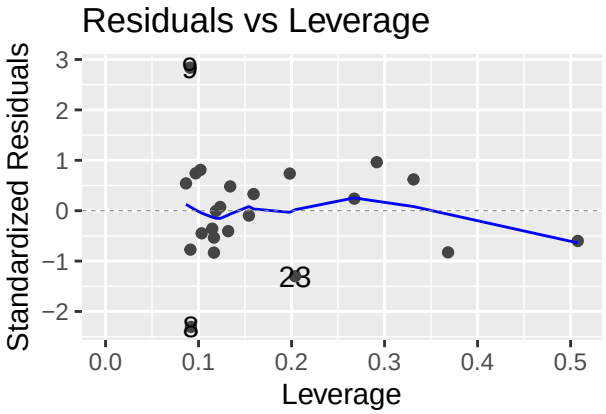



```
autoplot(black.group)
```

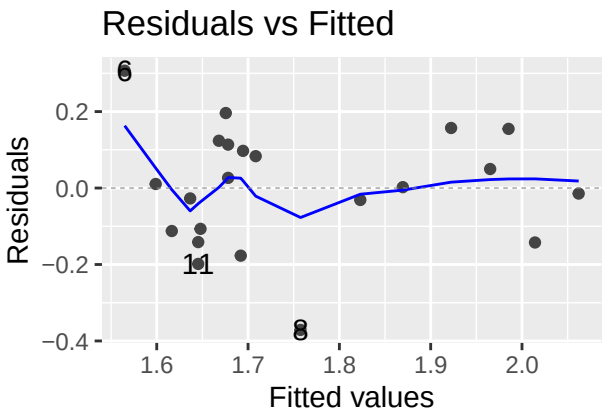


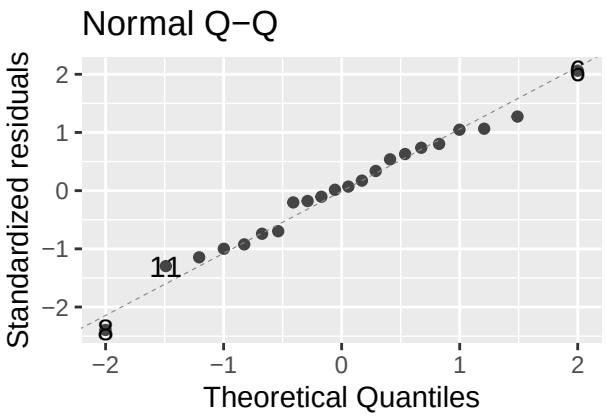


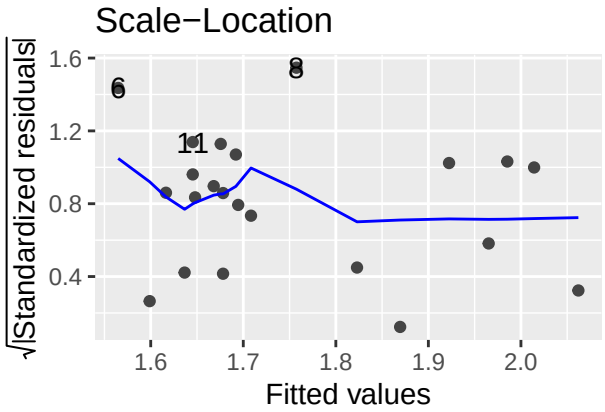


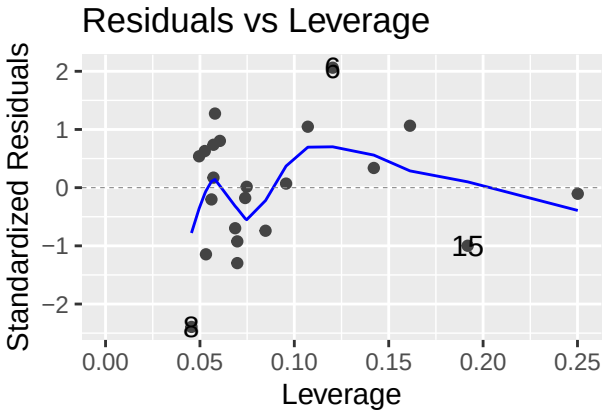



```
# Black carp modified
autoplot(black.simple.r)
```

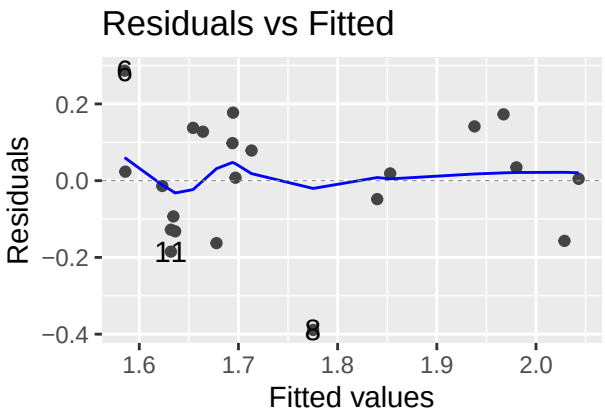


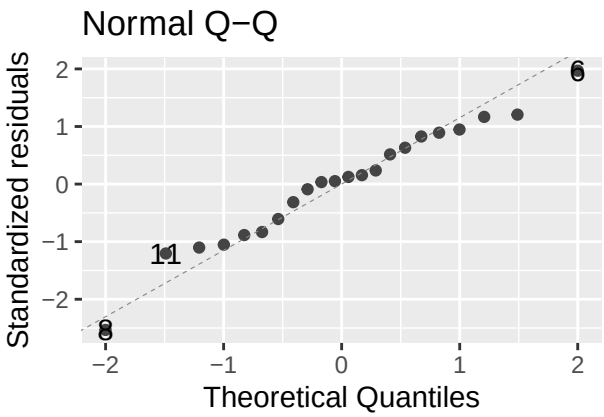


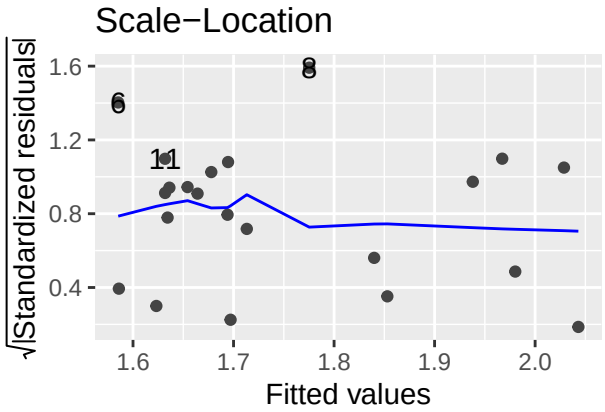


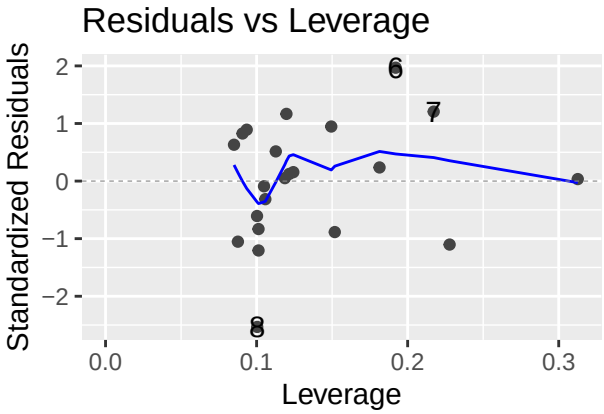



```
autoplot(black.linear.r)
```

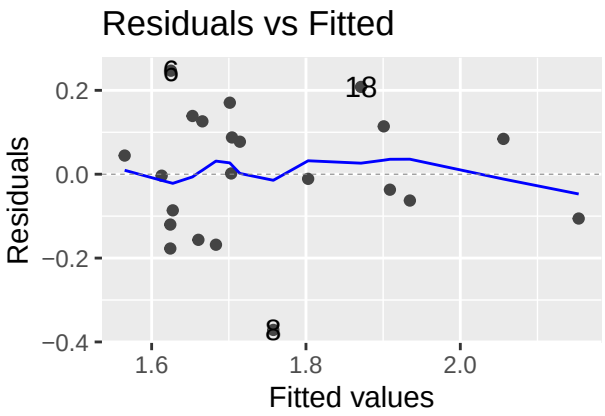


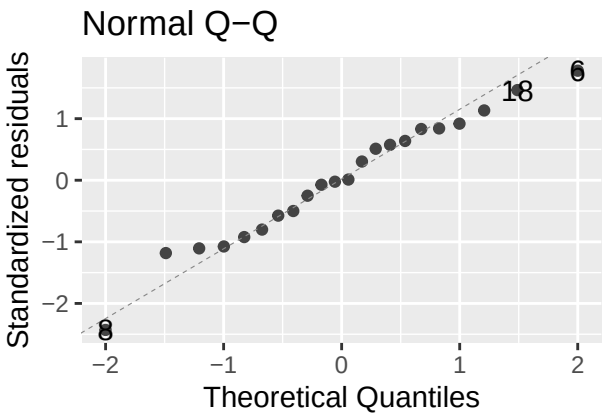


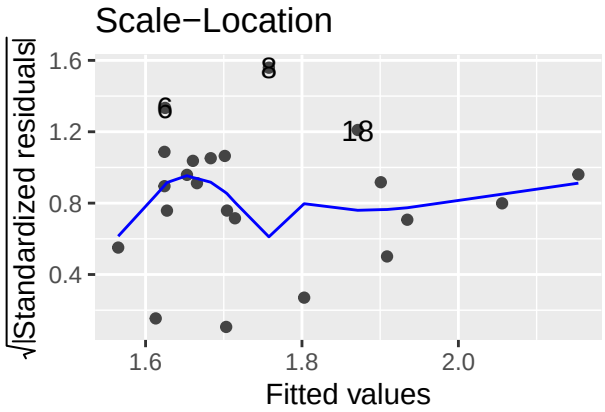


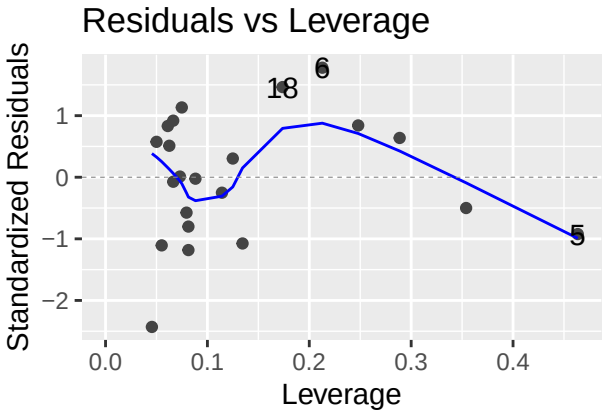



```
autoplot(black.int.r)
```

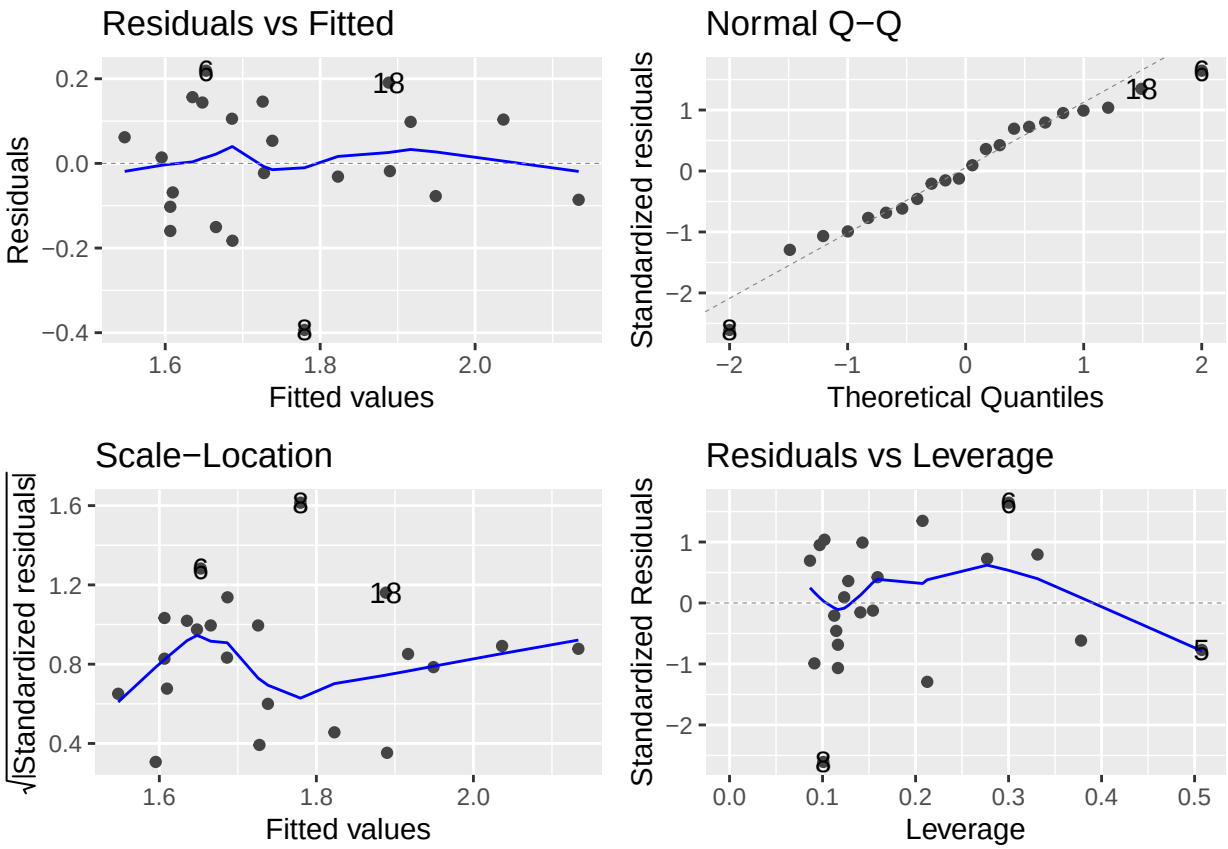









```
autoplot(black.group.r)
```


Residuals vs Fitted
Residuals
Fitted values
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized Residuals
Leverage