asian carp first part

Eddie Wu

2023-07-10

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",]
```

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

Asian carp

Using annual temperature

```
# 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 19.08105
## asian.linear  4 21.05438
## asian.int     4 21.05838
## asian.group   5 23.05379
```

```
# 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	0.5629096
Linear additive	0.5631005
Interaction	0.5630719
Grouped	0.5631048

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

Using cold temperature

```
# 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 25.05163
## asian.linear  4 26.72525
## asian.int     4 27.03839
## asian.group   5 28.72257
```

```
# 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	0.5179641
Linear additive	0.5205364
Interaction	0.5180687
Grouped	0.5205574

- 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 13.82776
## big.sil.linear  4 14.04192
## big.sil.int     4 14.42420
## big.sil.group   5 16.04189
```

```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	0.4104460
Linear additive	0.4495808
Interaction	0.4414281
Grouped	0.4495815

- 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 18.40785
## big.sil.linear  4 19.10025
## big.sil.int     4 20.39045
## big.sil.group   5 21.09175
```

```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	0.2968828
Linear additive	0.3313697
Interaction	0.2973531
Grouped	0.3315882

- 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  0.03963796
## black.linear  4 -0.93938475
## black.int     4 -1.07498134
## black.group   5  0.76248019
```

```
# 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	0.2775793
Linear additive	0.3775542
Interaction	0.3817600
Grouped	0.3867641

- 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 -0.53666717
## black.linear  4 -1.73129584
## black.int     4  1.35778275
## black.group   5 -0.07443596
```

```
# 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	0.2980990
Linear additive	0.4017188
Interaction	0.3017935
Grouped	0.4118959

- 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 -7.042199
## black.linear.r  4 -6.407973
## black.int.r     4 -7.389406
```

```
## black.group.r   5 -5.487496
```

```
# 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	0.2980990	0.3400199
Linear additive	0.4017188	0.3857962
Interaction	0.3017935	0.4167170
Grouped	0.4118959	0.4197206

- 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 -9.115725
## black.linear.r  4 -8.629428
## black.int.r     4 -7.607280
## black.group.r   5 -7.385258
```

```
# 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	0.2980990	0.4082545
Linear additive	0.4017188	0.4535690
Interaction	0.3017935	0.4233674

Model	R2_before	R2_after
Grouped	0.4118959	0.4748796

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

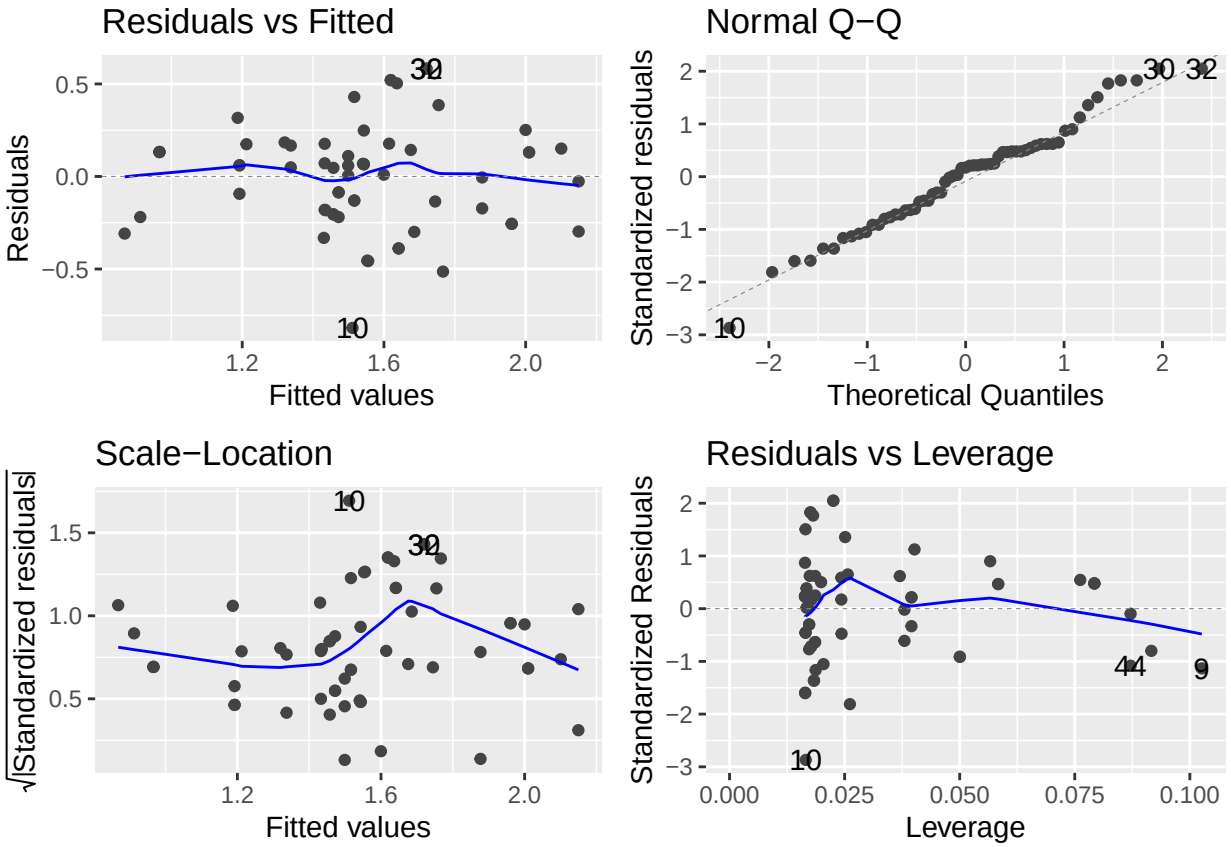
Concluding points

- 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.

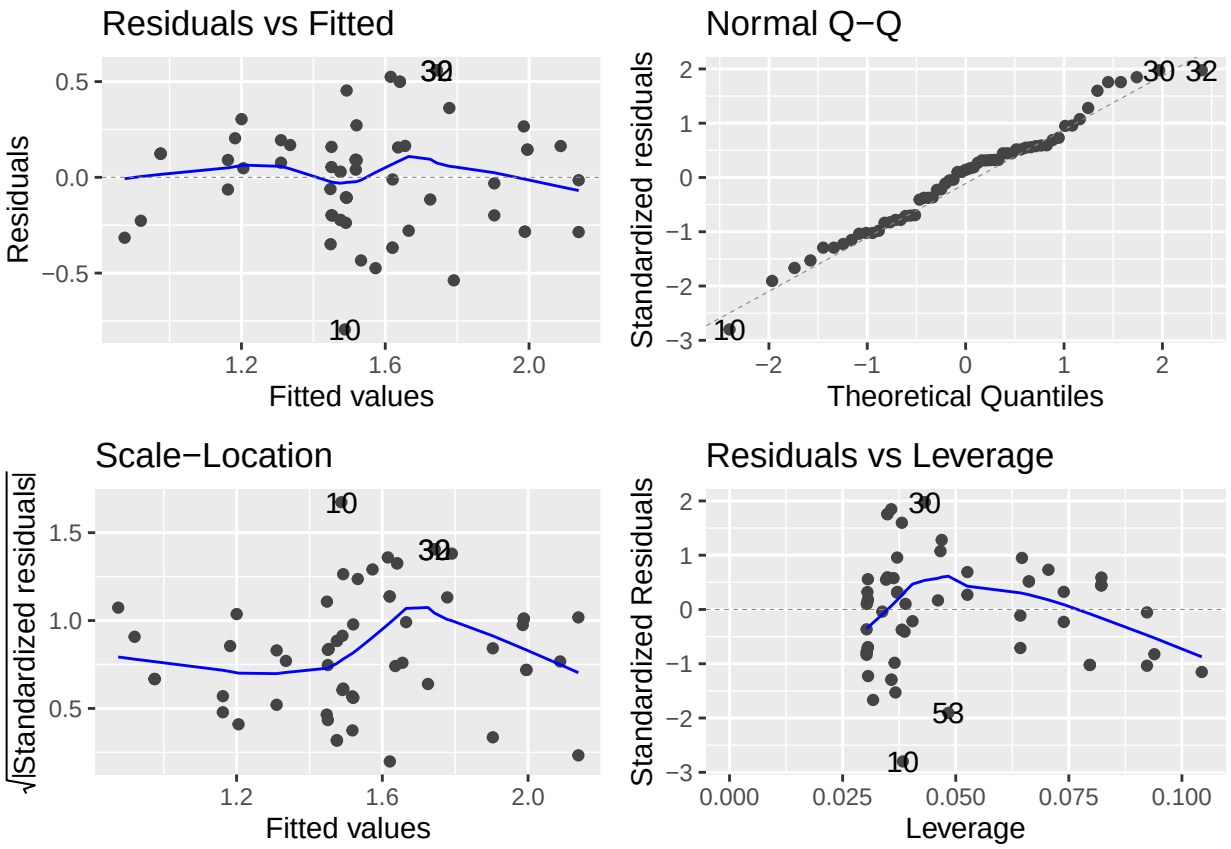
Diagnostic plots

Only the diagnostic plots with cold temperature

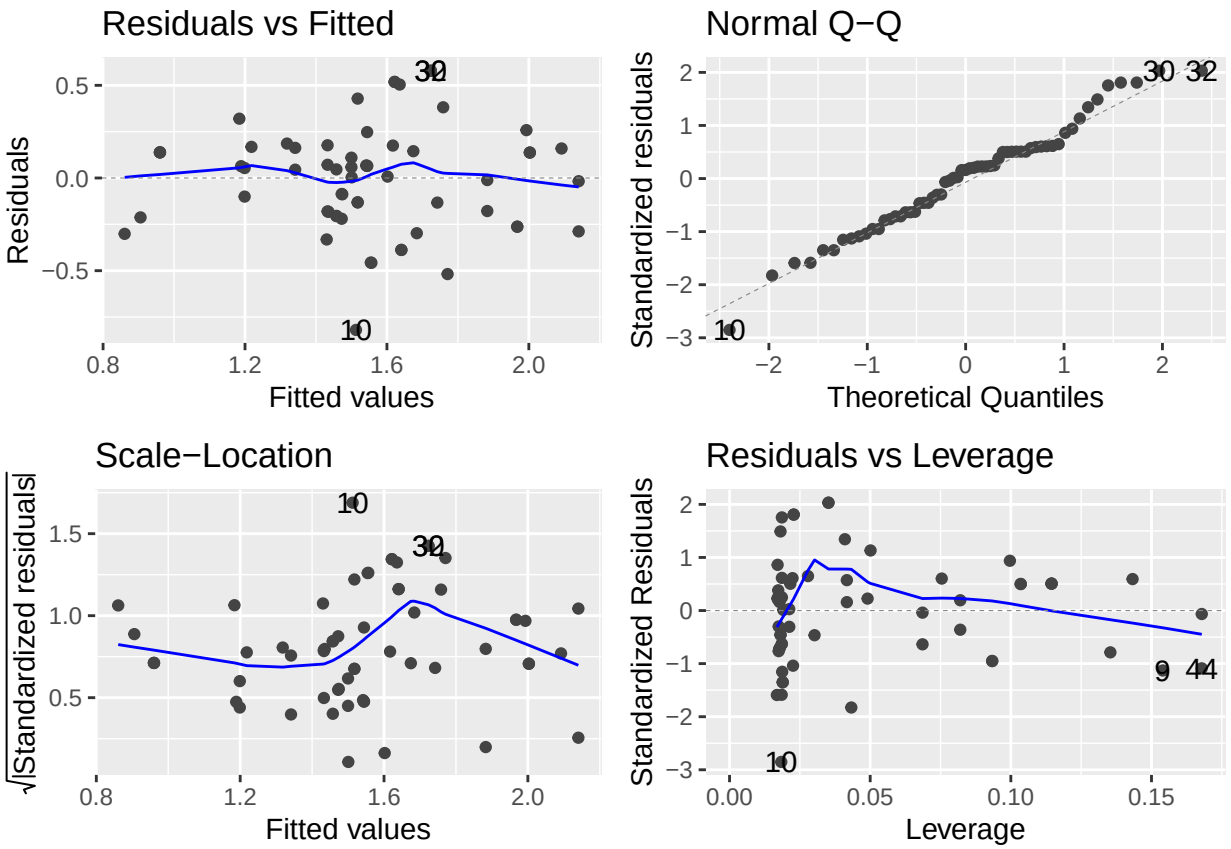
```
# Asian carp

autoplot(asian.simple)
```



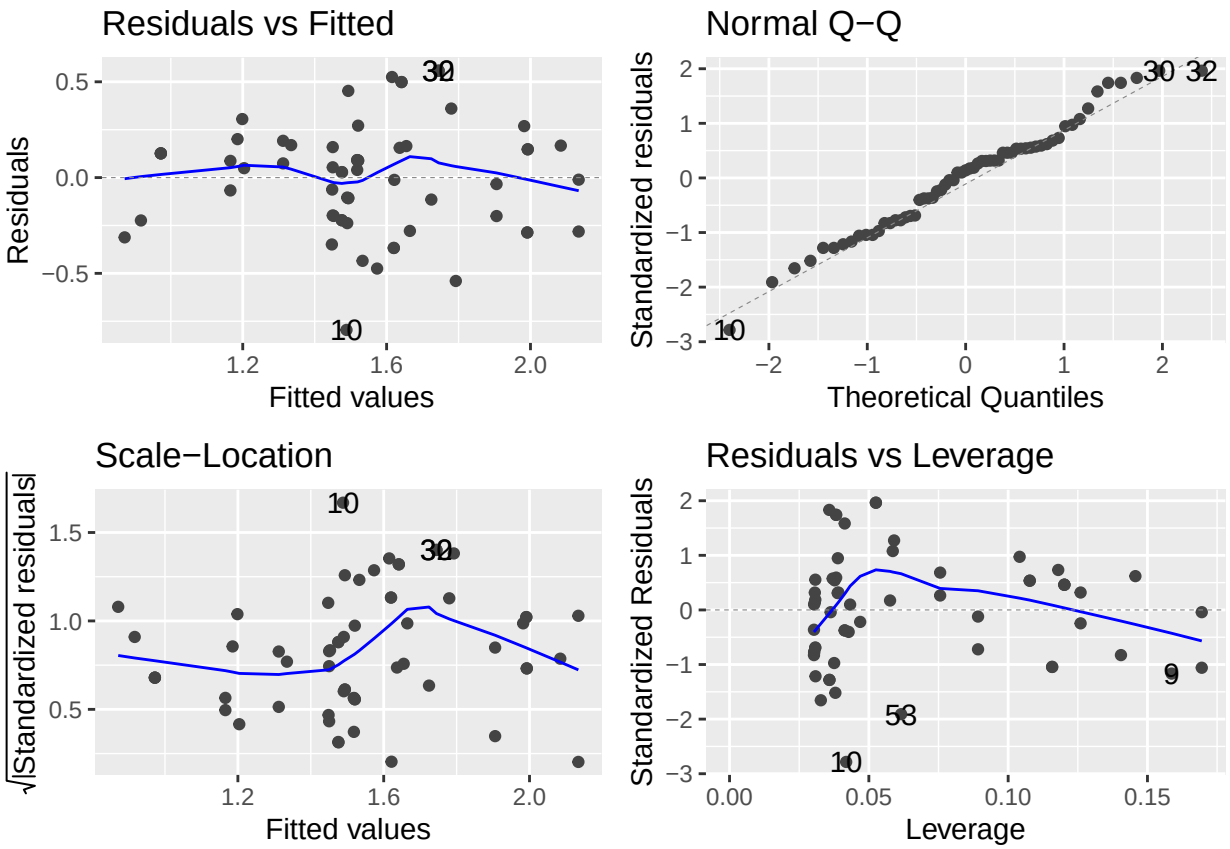
```
autoplot(asian.linear)
```



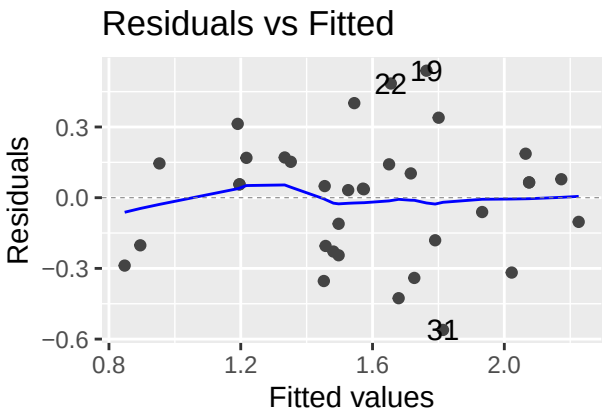

```
autoplot(asian.int)
```

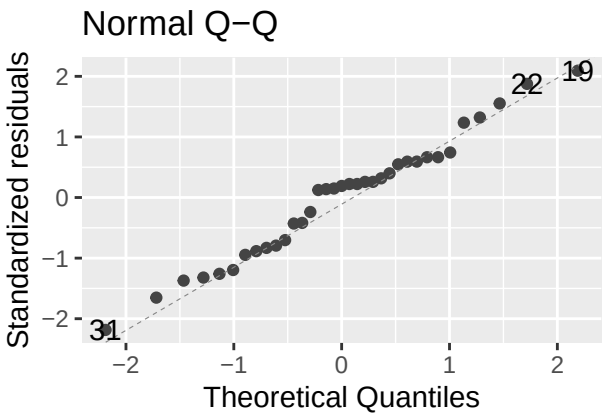


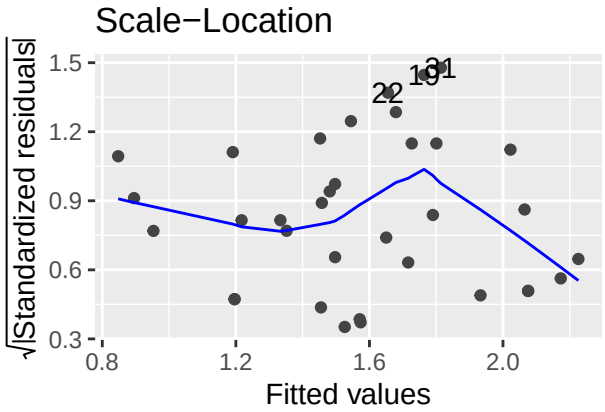

```
autoplot(asian.group)
```

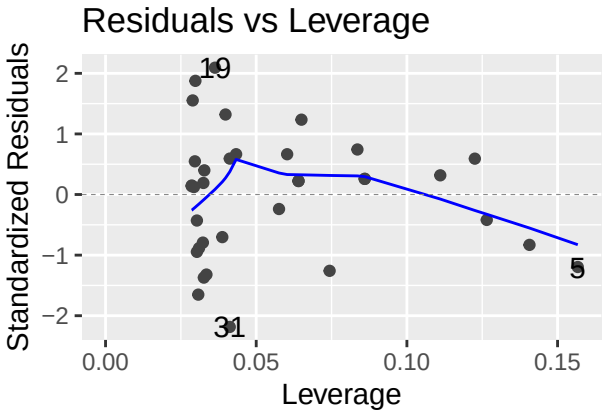



```
# Grass carp
autoplot(grass.simple)
```

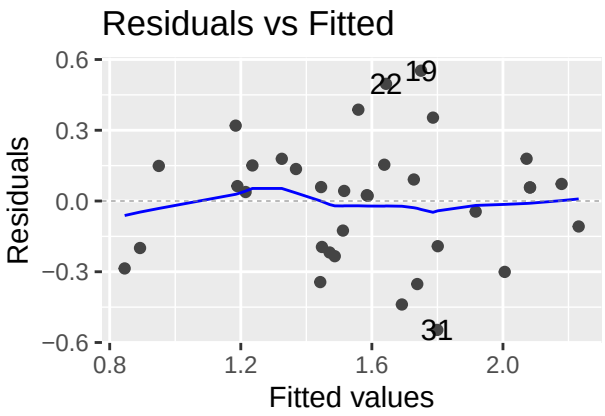


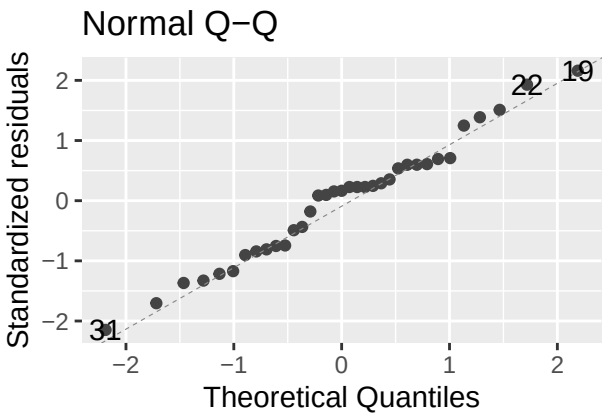


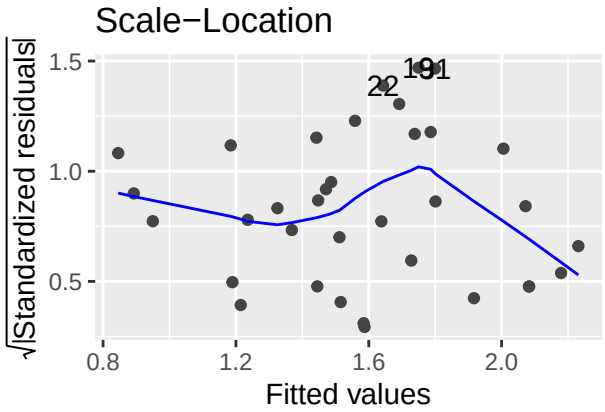


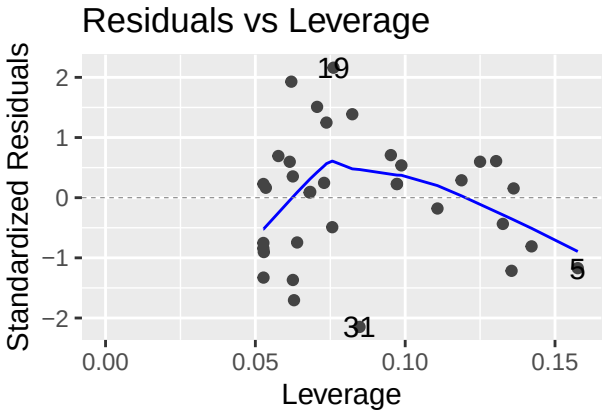



```
autoplot(grass.linear)
```

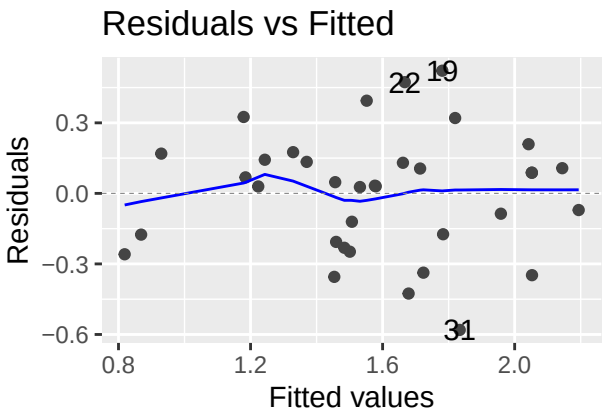


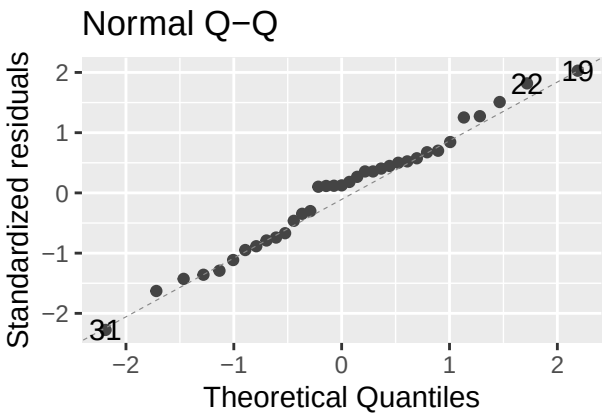


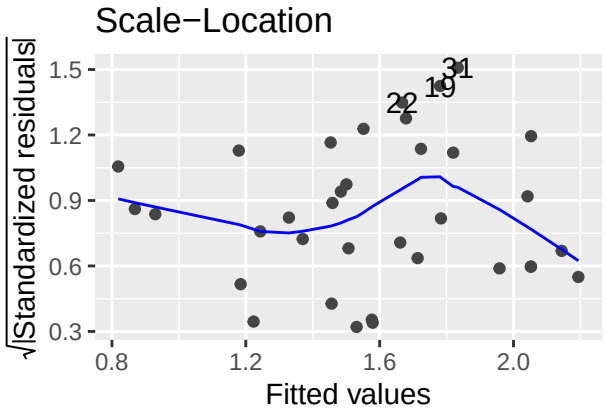


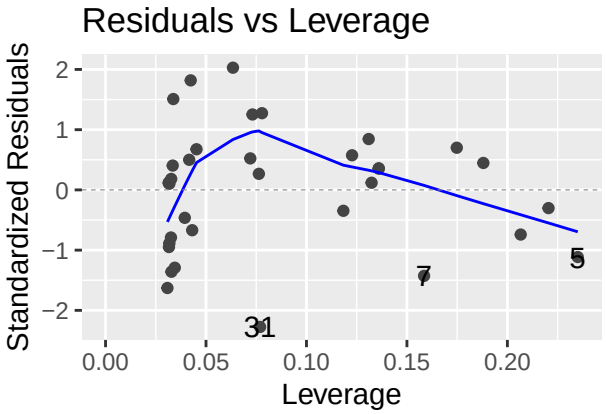



```
autoplot(grass.int)
```

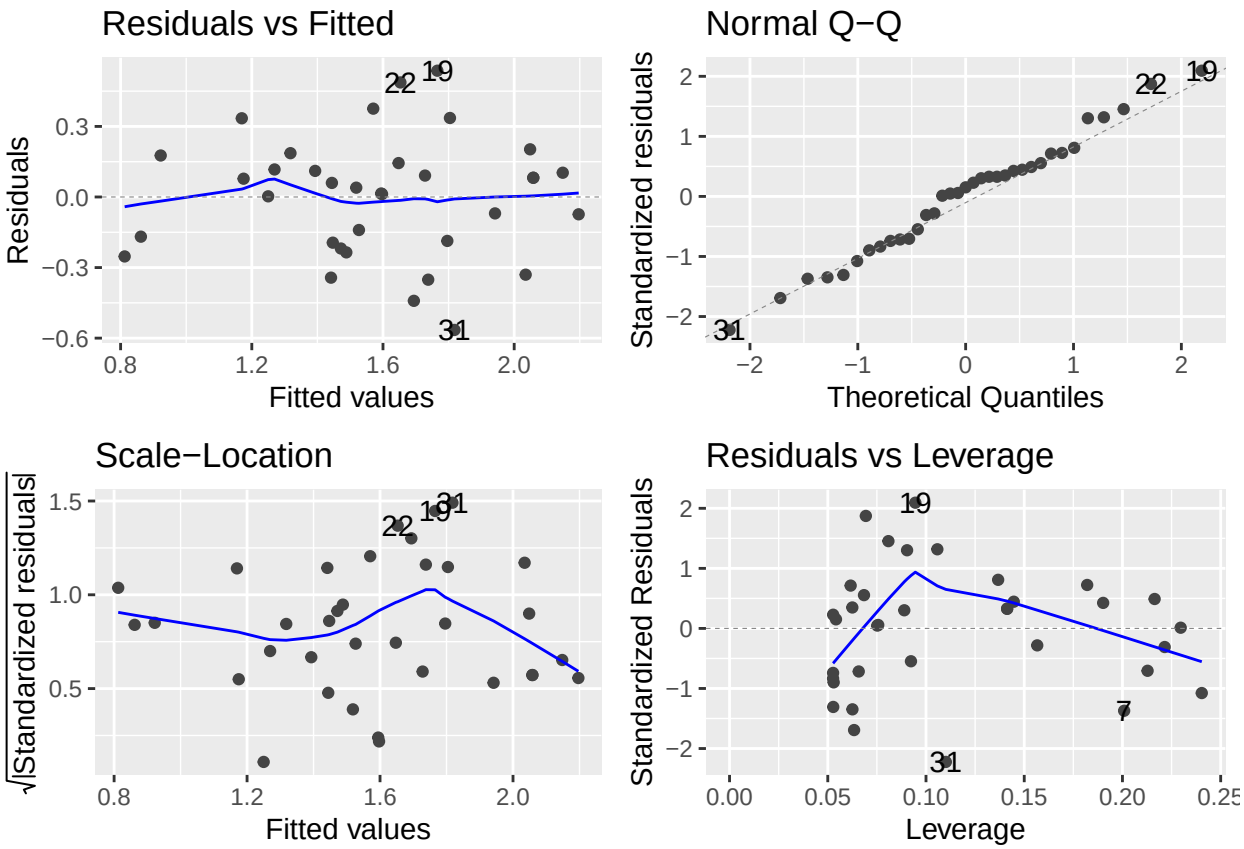




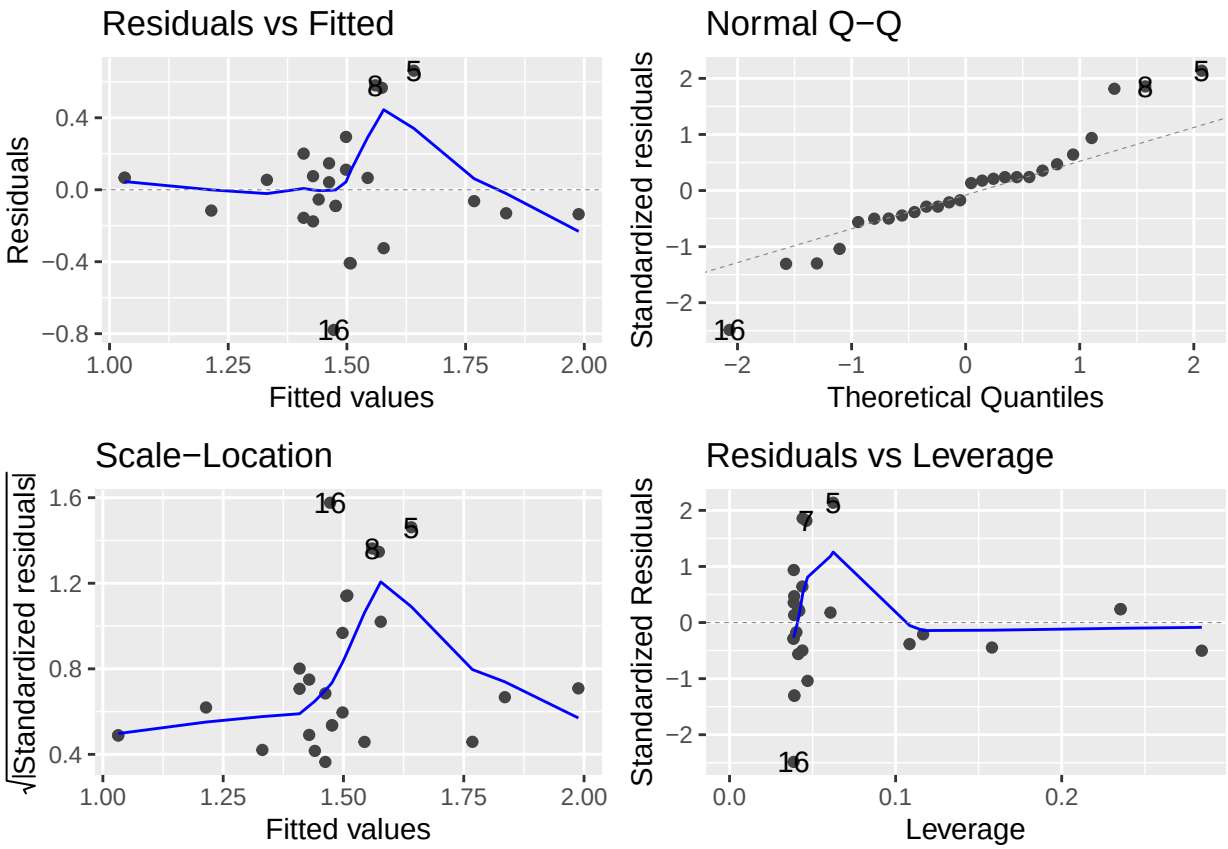




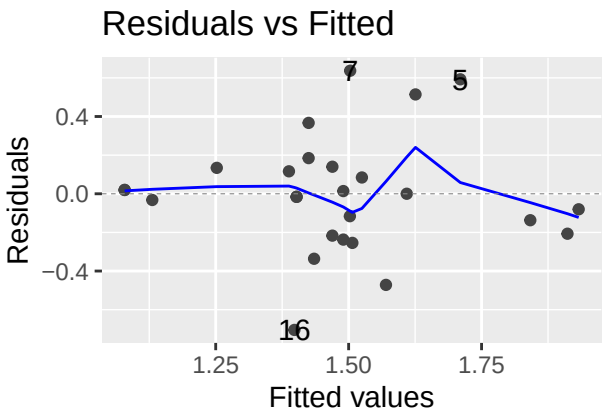

```
autoplot(grass.group)
```

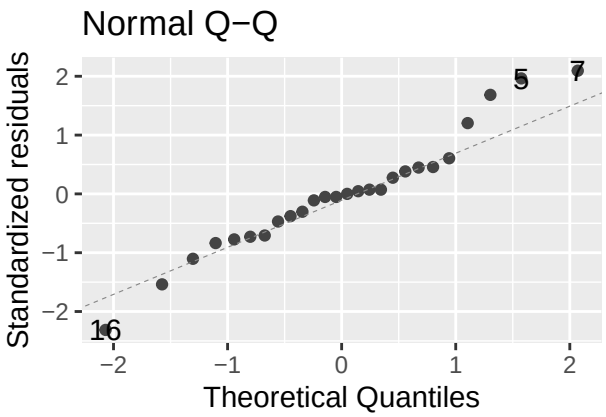


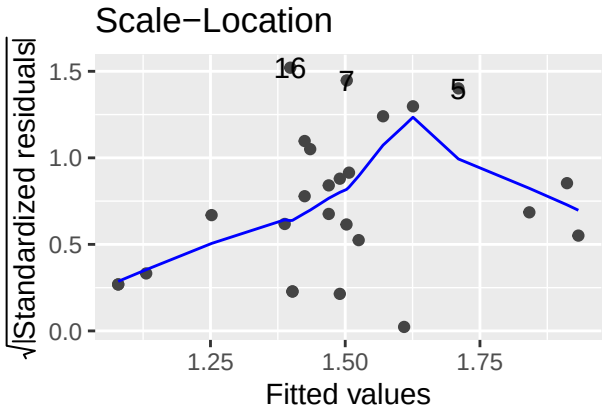

```
# Bighead and silver carp
autoplot(big.sil.simple)
```

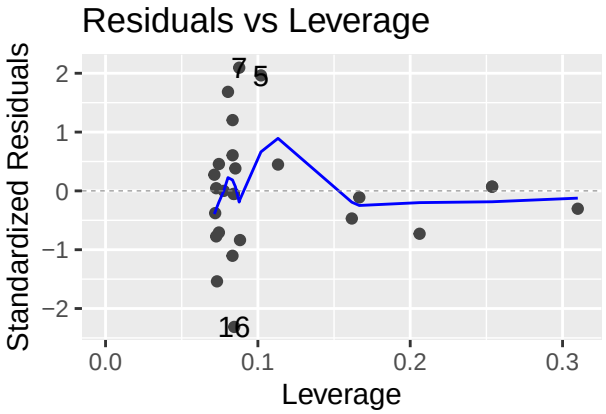



```
autoplot(big.sil.linear)
```

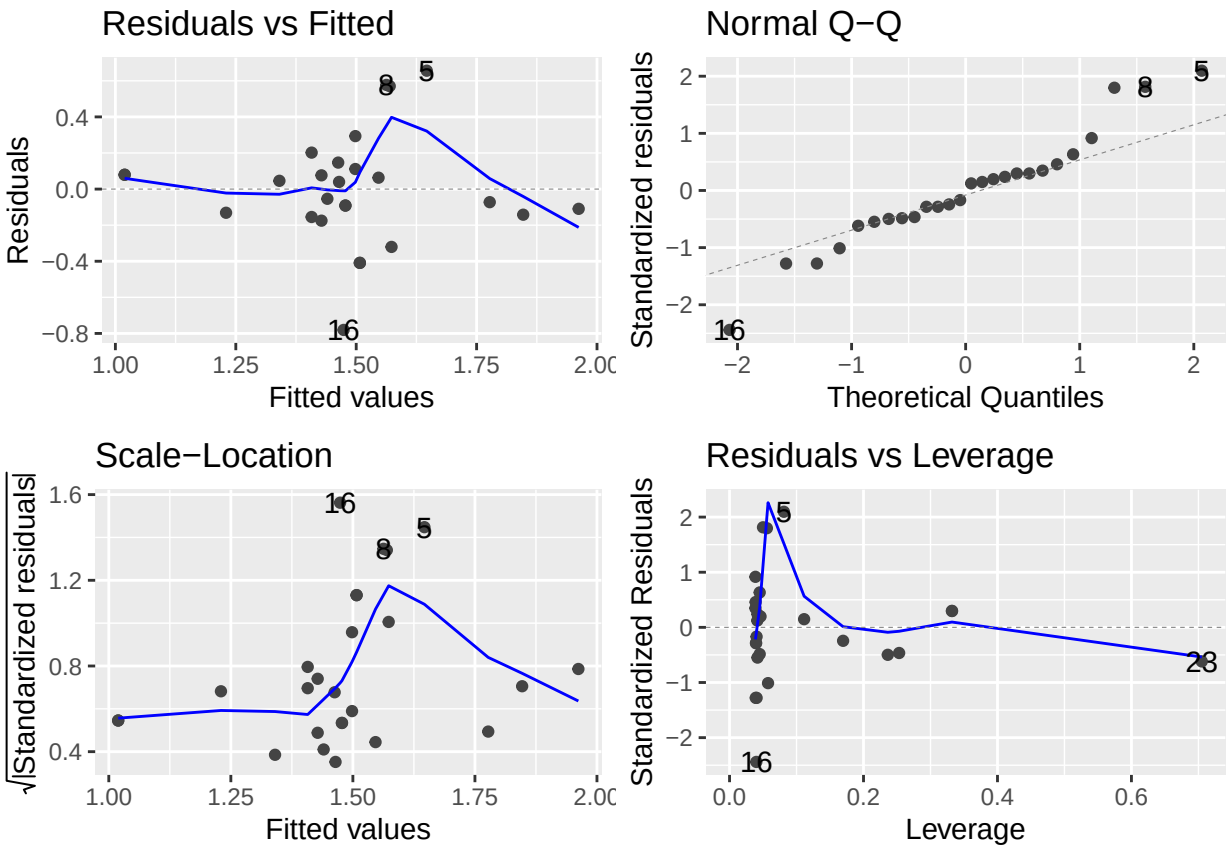




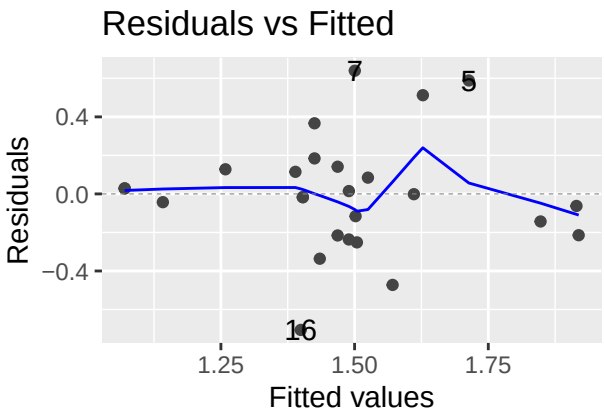


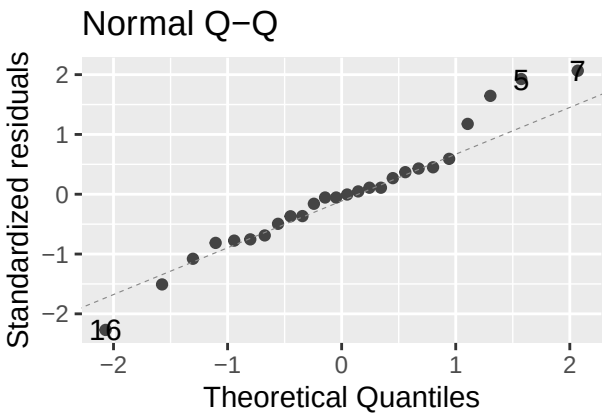


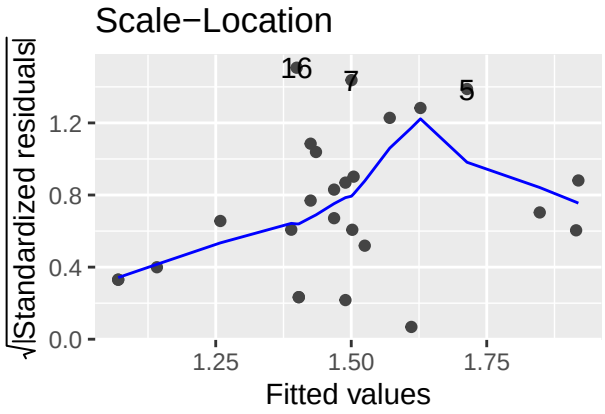

```
autoplot(big.sil.int)
```

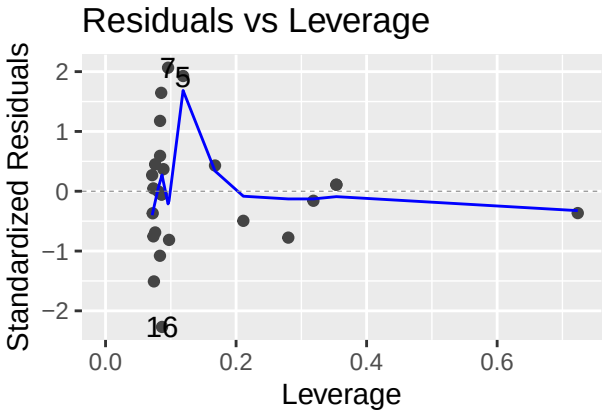



```
autoplot(big.sil.group)
```

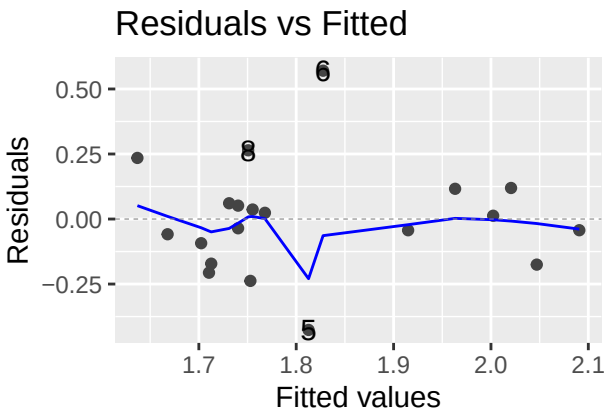


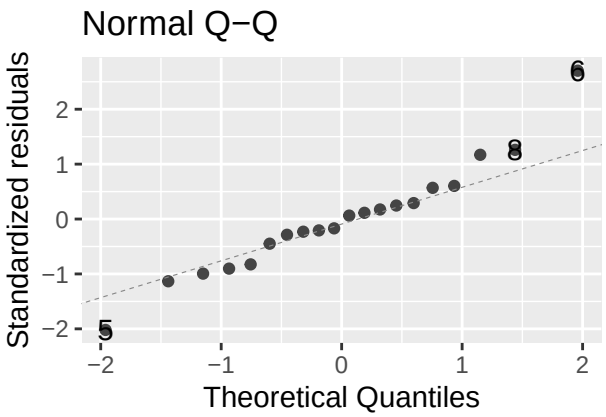


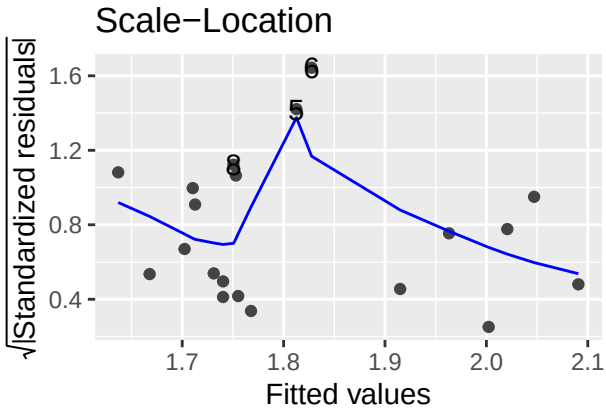


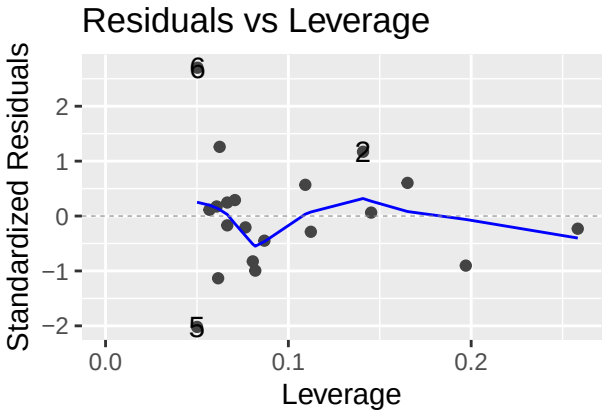



```
# Black carp
autoplot(black.simple)
```

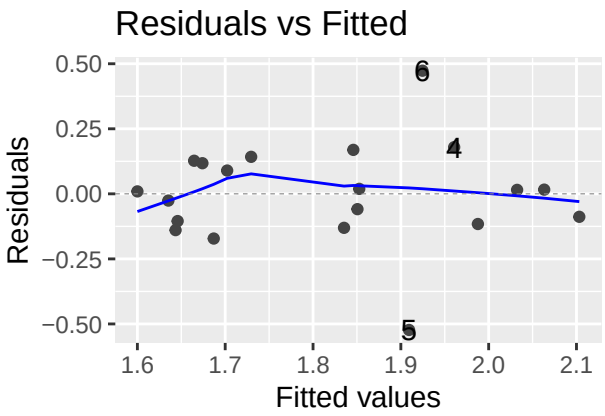


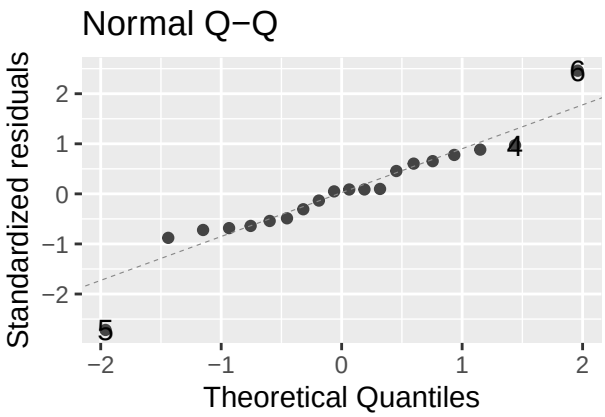


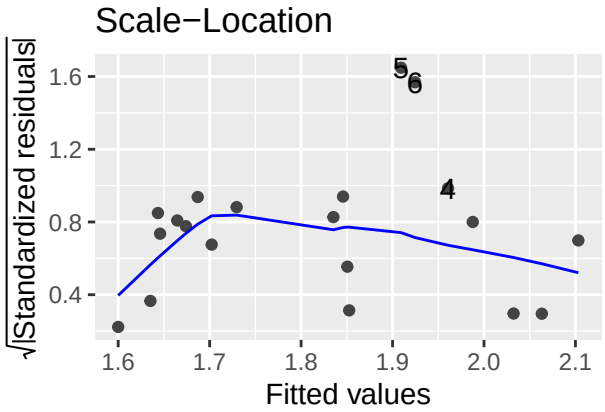


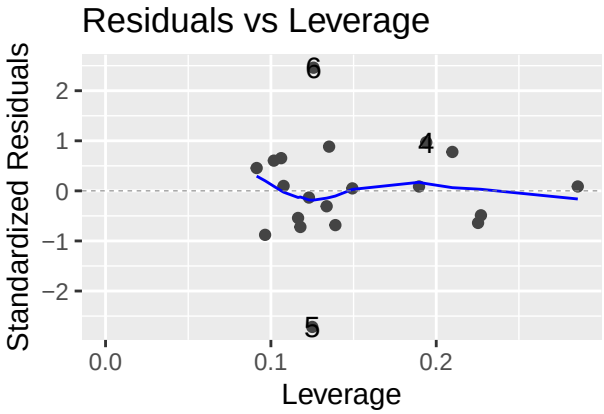



```
autoplot(black.linear)
```

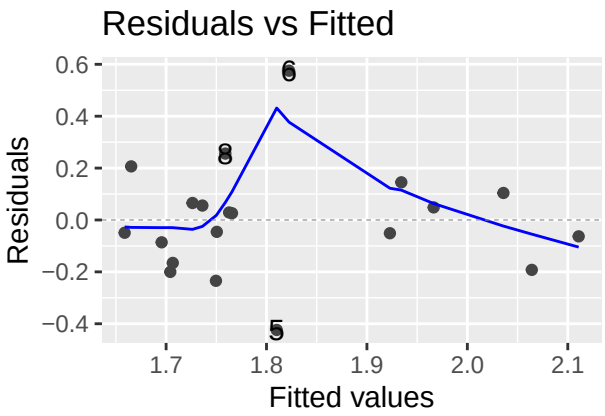


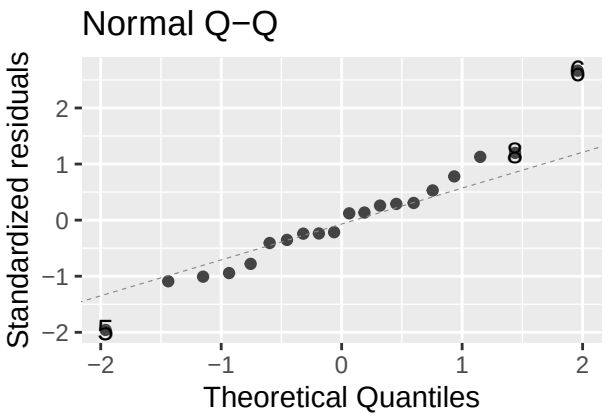


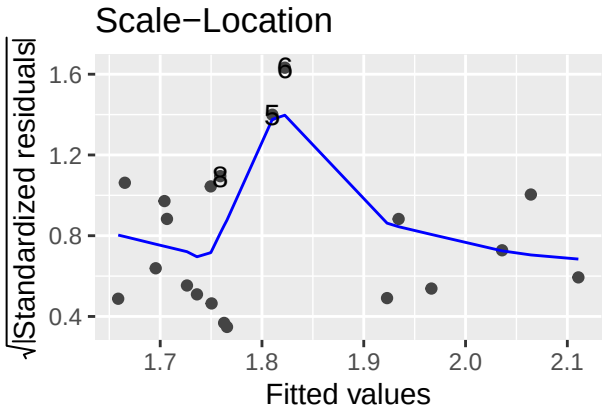


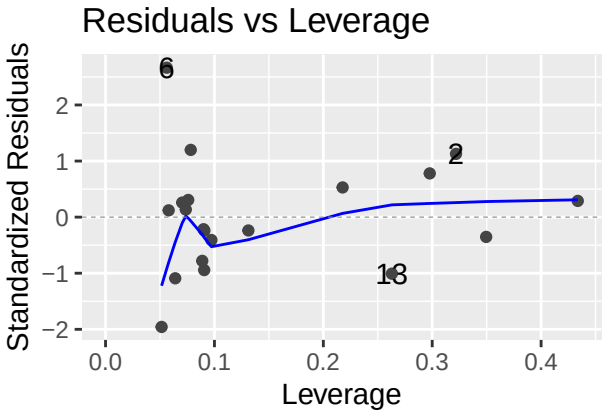



```
autoplot(black.int)
```

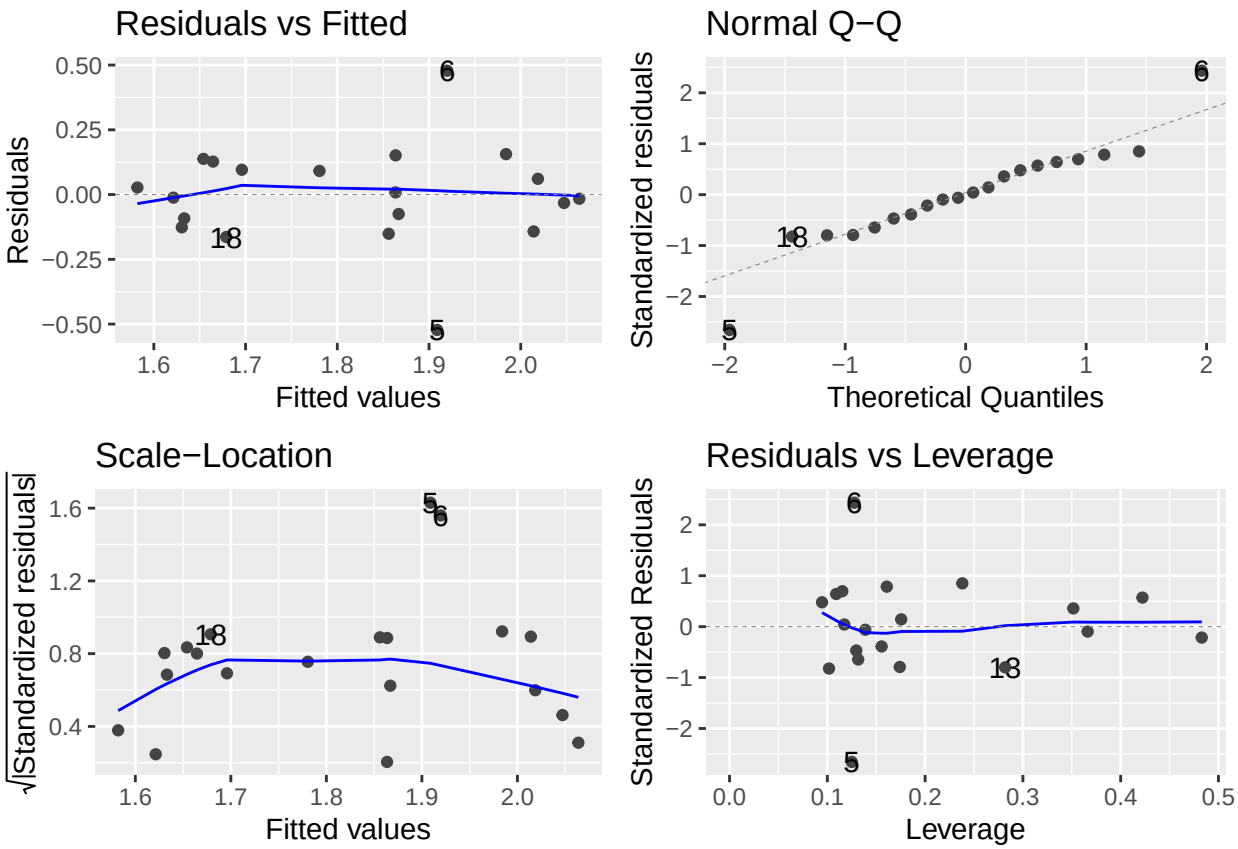




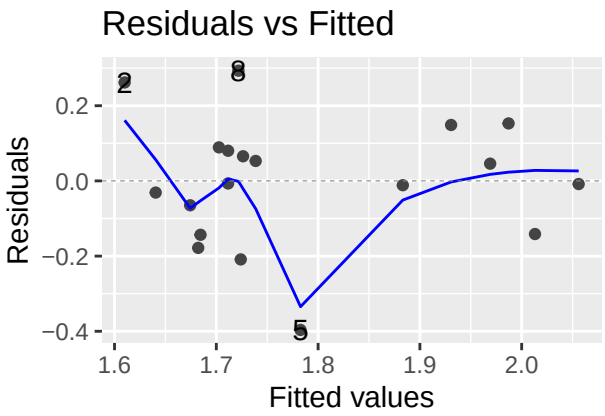


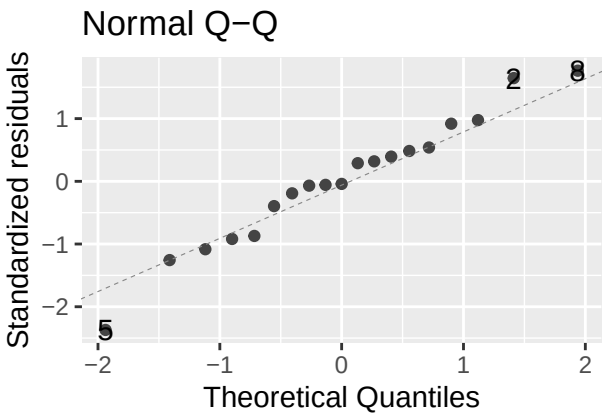


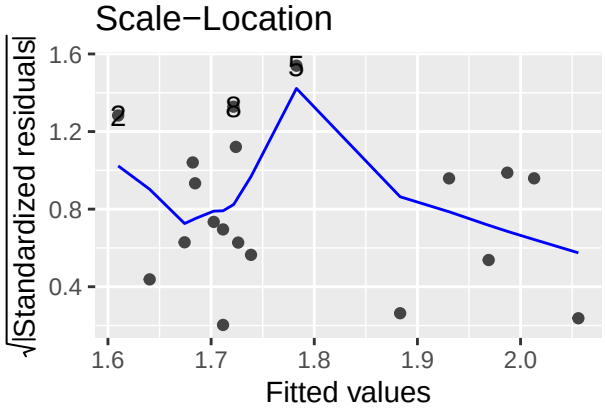

```
autoplot(black.group)
```

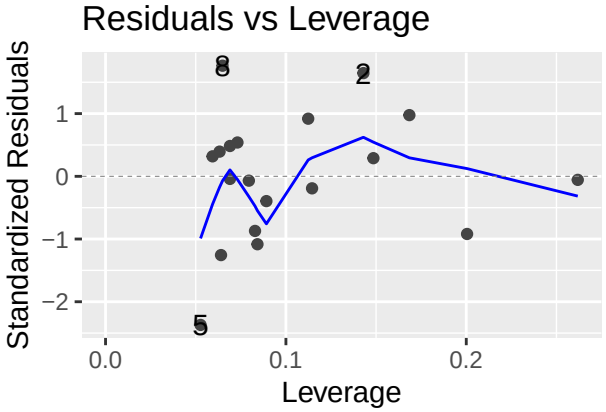



```
# Black carp modified
autoplot(black.simple.r)
```

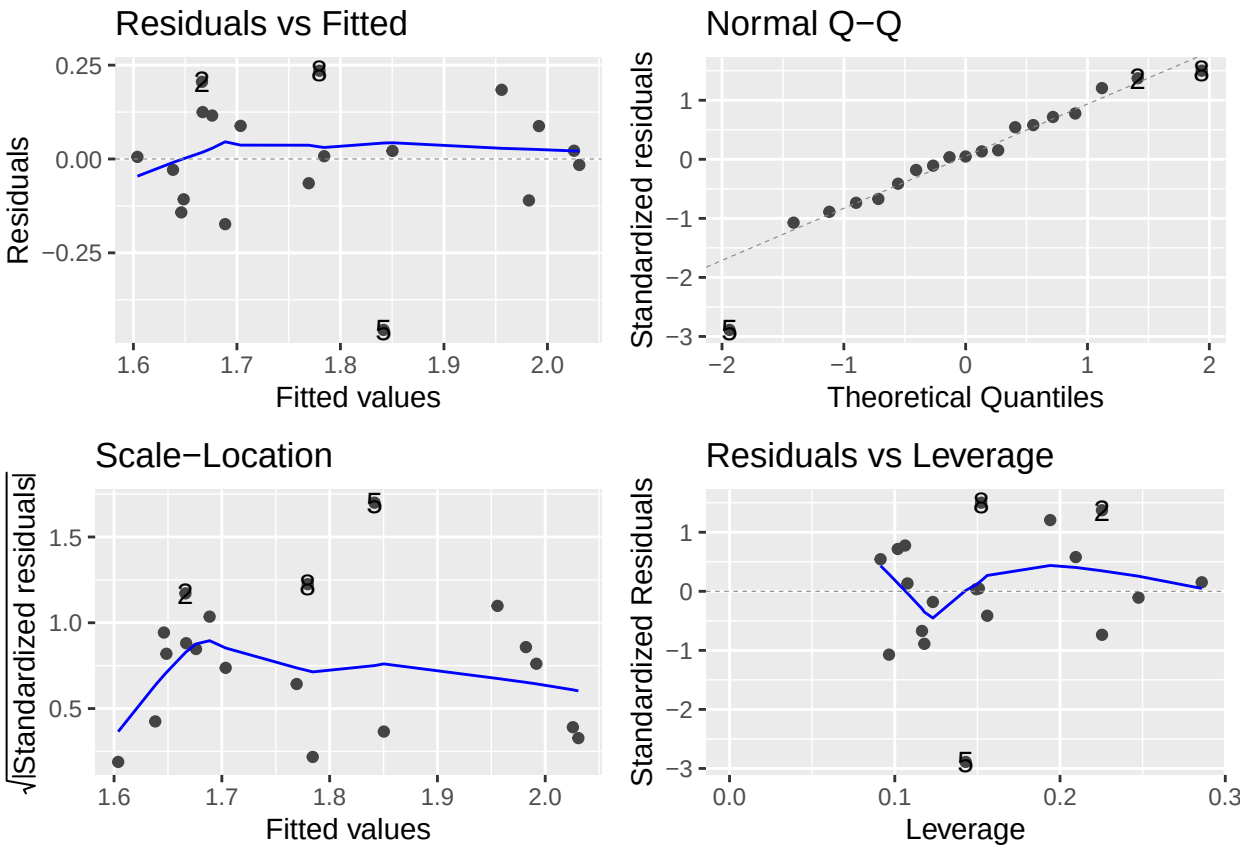




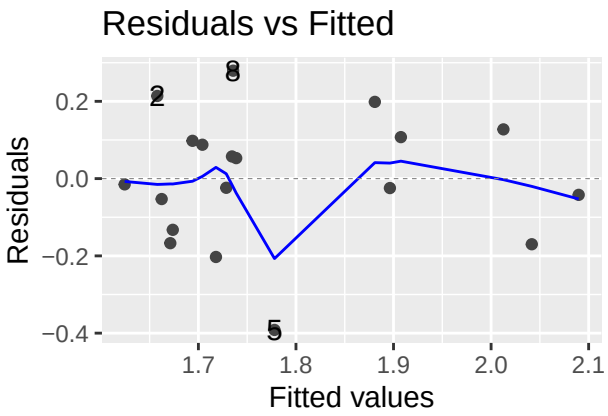


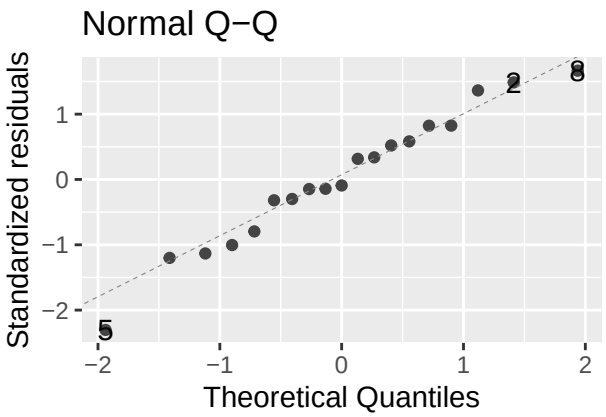


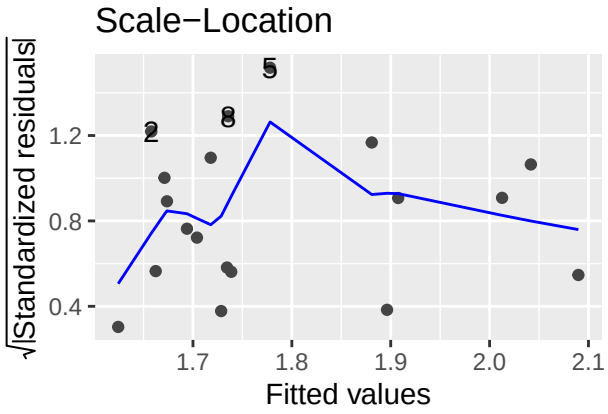

```
autoplot(black.linear.r)
```

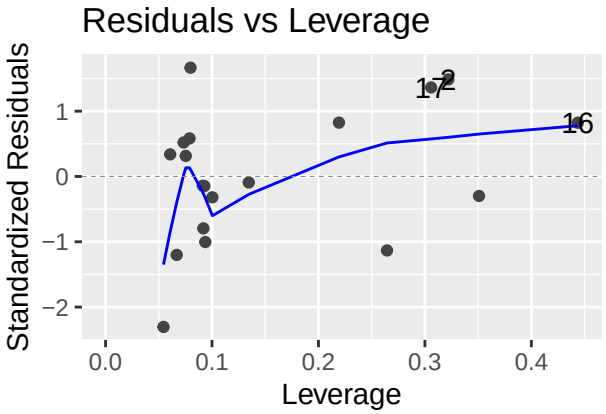



```
autoplot(black.int.r)
```

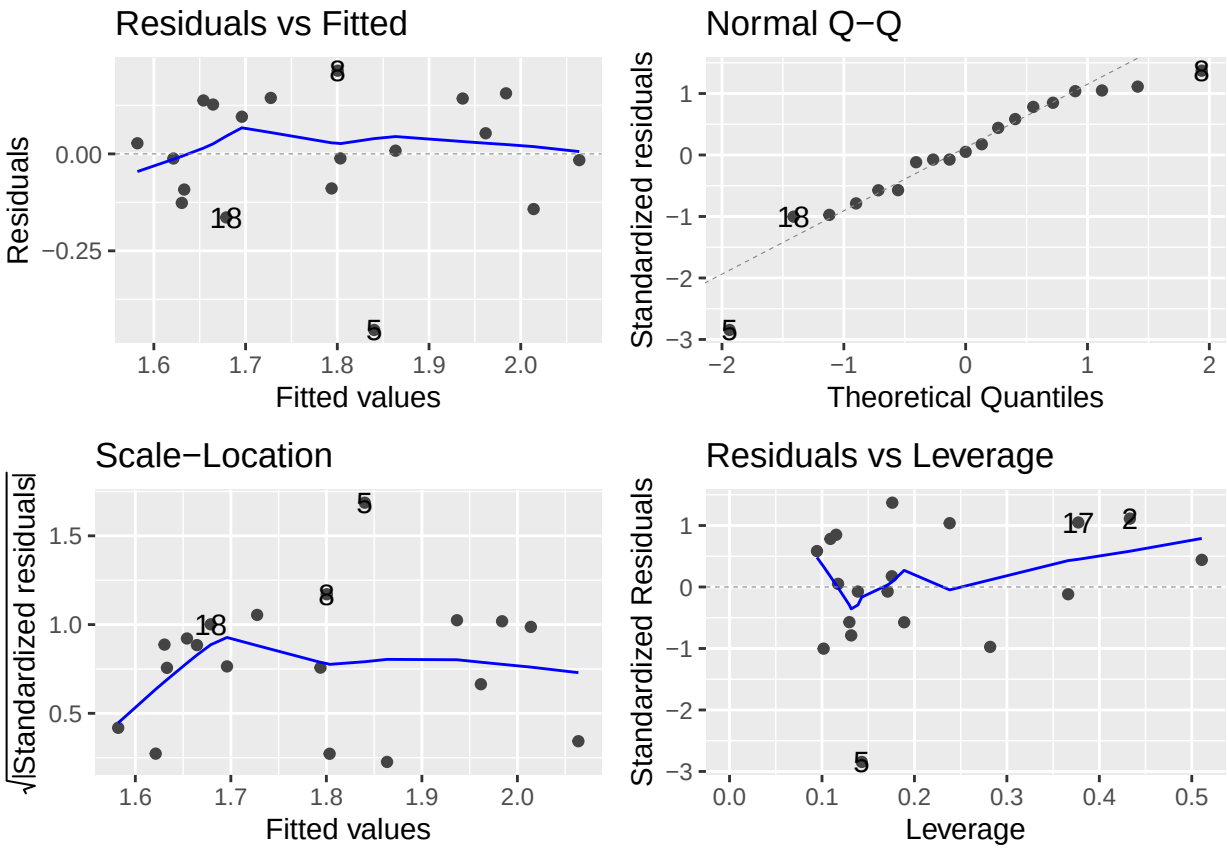









```
autoplot(black.group.r)
```


Residuals vs Fitted
Residuals
0.00
−0.25
1.6
1.7
1.8
1.9
2.0
Fitted values
Normal Q−Q
Standardized residuals
1
0
−1
−2
−3
−2
−1
0
1
2
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
1.5
1.0
0.5
1.6
1.7
1.8
1.9
2.0
Fitted values
Residuals vs Leverage
Standardized Residuals
1
0
−1
−2
−3
0.0
0.1
0.2
0.3
0.4
0.5
Leverage