Fachhochschule Südwestfalen
University of Applied Sciences

Statistical Modeling

CH.3 - Regression Diagnostics

2024 || Prof. Dr. Buchwitz

Wir geben Impulse

Outline

1 Organizational Information

2 Regression Diagnostics

Course Contents

Session	Topic
1	Simple Linear Regression
2	Multiple Linear Regression
3	Regression Diagnostics
4	Qualitative Variables as Predictors
5	Transformation of Variables
6	Weighted Least Squares
7	Correlated Errors
8	Analysis of Collinear Data
9	Working with Collinear Data
10	Variable Selection Procedures
11	Logistic Regression
12	Further Topics

Outline

1 Organizational Information

2 Regression Diagnostics

Introduction

- In this chapter we talk about the **standard regressions assumptions**, the consequences when violating them and how to detect violations so that we can focus on the remainder of the course on methods of how to correct or compensate for violations.
- When those assumptions are violated, the discussed and derived results for making inferences about the regression coefficients do not hold, which essentially means that **conclusions drawn on the corresponding models are wrong**.
- The majority of the discussed methods are **graphical methods** which means that they may be somewhat subjective here or there, which needs to be kept in mind when interpreting diagnostic plots.

Overview

1. Assumptions about the form of the model.
2. Assumptions about the errors.
3. Assumptions about the predictors.
4. Assumptions about the observations.

The properties of the least squares estimators (BLUE) are based on the discussed assumptions!

Assumption 1: Model

- The model that relates Y and $X_1, X_2, \ldots, X_p$ is assumed to be **linear in the regression parameters** $\beta_0, \beta_1, \ldots, \beta_p$ so that

$$\text{Model:} \quad Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \epsilon$$

$$\text{Observation:} \quad y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \epsilon_i, \;\; i = 1, 2, \ldots, n$$

- This assumption is called the **linearity assumption**.
- In simple linear regression checking can be done using a scatterplot of Y versus X. For multiple linear regression there are other plotting techniques which we will discuss.
- When the linearity assumption does not hold, transforming the data may lead to linearity (transformations are discussed at a later point).

Assumption 2: Errors

- The errors $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are assumed to be **independently and identically distributed** (iid) normal random variables each with mean zero and common variance σ^2. This implies:
  - **Normality Assumption:** The error $\epsilon_i, i = 1, 2, \ldots, n$ has a normal distribution.
  - The errors $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ have mean zero.
  - **Constant Variance Assumption:** The errors have the same (but unknown) variance σ^2. When this assumption does not hold we have the *heteroscedasticity problem*.
  - **Independent errors Assumption:** $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are independent of each other (pairwise covariances are zero). Violations lead to the *autocorrelation problem*.

Assumption 3: Predictors

- There are three assumptions for the predictor variables.
  - The predictor variables $X_1, X_2, \ldots, X_p$ are **nonrandom**. This means the values $x_{1j}, x_{2,j}, \ldots, x_{nj}$ with $j = 1, 2, \ldots, p$ are fixed (which is usually only fully satisfied under experimental conditions). In practice the results presented hold, but results are conditional on the data.
  - The values $x_{1j}, x_{2,j}, \ldots, x_{nj}$ are measured without error (which is hardly ever satisfied). In practice it is sufficient, when the measurement error is small compared to the random error ϵ_i.
  - The predictor variables $X_1, X_2, \ldots, X_p$ are assumed to be linearly independent of each other. This assumption guarantees the uniqueness of the least squares solution. If this assumption is violated this is referred to as the *collinearity problem*.

The first two assumptions cannot be checked and do not play a role in our analysis. They have to be kept in mind when collecting data.

Assumption 4: Observations

- All observations are equally reliable and have an approximately equal role in determining the regression results. This means that they are equally relied on when drawing conclusions.

Assumption 4: Observations

- All observations are equally reliable and have an approximately equal role in determining the regression results. This means that they are equally relied on when drawing conclusions.

Conclusion

Small or minor violations of the underlying assumptions do not invalidate the inferences or conclusions drawn from the analysis. Gross validations, however, can seriously distort conclusions. **It is essential to investigate all signs of assumption validations by *always* checking the structure of the residuals and the data patterns at least using graphs!**

Residuals

- Analysing residuals is a simple and effective method for detecting model deficiencies in regression analysis. In most analyses it is probably the **most important** part of an analysis.
- Residual plots may lead to suggestions for structure or point to information in the data that might be missed or overlooked. Those clues can lead to a better understanding (and possibly a better model) of the underlying process.
- Starting point for the analysis are the **ordinary** least squares residuals that can be calculated after obtaining the fitted values:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \ldots + \hat{\beta}_p x_{ip}$$
$$e_i = y_i - \hat{y}_i \quad \text{for} \quad i = 1, 2, \ldots, n$$

Residuals

- The fitted values can also be written as function of the predictor variables, where p_{ij} only depends on the predictor variables (essentially values from the hat matrix **P**).

$$\hat{y}_i = p_{i1}y_1 + p_{i2}y_2 + \ldots + p_{in}y_n$$

- When $i = j$ the value p_{ii} represents the weight (leverage) given to y_i in determining the i-th fitted value $\hat{y}_i$. The n **leverage values** $p_{11}, p_{22}, \ldots, p_{nn}$ can be obtained by

$$p_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum(x_i - \bar{x})^2}$$

A high leverage value indicates some "extremeness" in X.

Residuals

- The *ordinary least squares residuals* $e_1, e_2, \ldots, e_p$ do not have **unequal variances** $Var(e_i) = \sigma^2(1 - p_{ii})$. Analyzing requires **standardized residuals** by calculating

$$z_i = \frac{e_i}{\sigma\sqrt{1 - p_{ii}}}$$

- This requires an unbiased estimate for the unknown standard deviation σ of ϵ for which we have two unbiased estimates to choose from

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n - p - 1} = \frac{SSE}{n - p - 1} \quad \text{with} \quad \hat{\sigma}^2_{(i)} = \frac{SSE_{(i)}}{(n - 1) - p - 1} = \frac{SSE_{(i)}}{n - p - 2}$$

- $SSE_{(i)}$ is the sum of squared residuals when the i-th observation is left out so that the model is fitted using $n - 1$ observations.

Residuals

The choice of variance estimates results in two different types of residuals, although both are unbiased estimates.

Internally studentized residuals (using $\hat{\sigma}^2$)

$$r_i = \frac{e_i}{\hat{\sigma}\sqrt{1 - p_{ii}}} \quad \text{with} \quad \hat{\sigma}^2 = \frac{SSE}{n - p - 1}$$

Externally studentized residuals (using $\hat{\sigma}^2_{(i)}$)

$$r_i^* = \frac{e_i}{\hat{\sigma}_{(i)}\sqrt{1 - p_{ii}}} \quad \text{with} \quad \hat{\sigma}^2_{(i)} = \frac{SSE_{(i)}}{n - p - 2}$$

Called *externally studentized* because e_i is not involved in (external to) $\hat{\sigma}^2_{(i)}$.

In practice the difference between r_i and r_i^* is small and both could be used, so the difference is ignored in the following notation.

Graphical Methods

Dimensionality:

- One-dimensional graphs, indicate the distribution of a particular variable (e.g. symmetry, skewness) and allow identification of outliers.
- Two-dimensional graphs allow exploration of relationships (by pairing variables) and general patterns.

Step in Model Selection Process:

- Graphs **before** fitting a model, to e.g. correct data errors, select variables and preparation for model selection.
- Graphs **after** fitting a model to check assumptions and assessing the goodness of fit.

Example: Hamiltons Data

```
P103
```

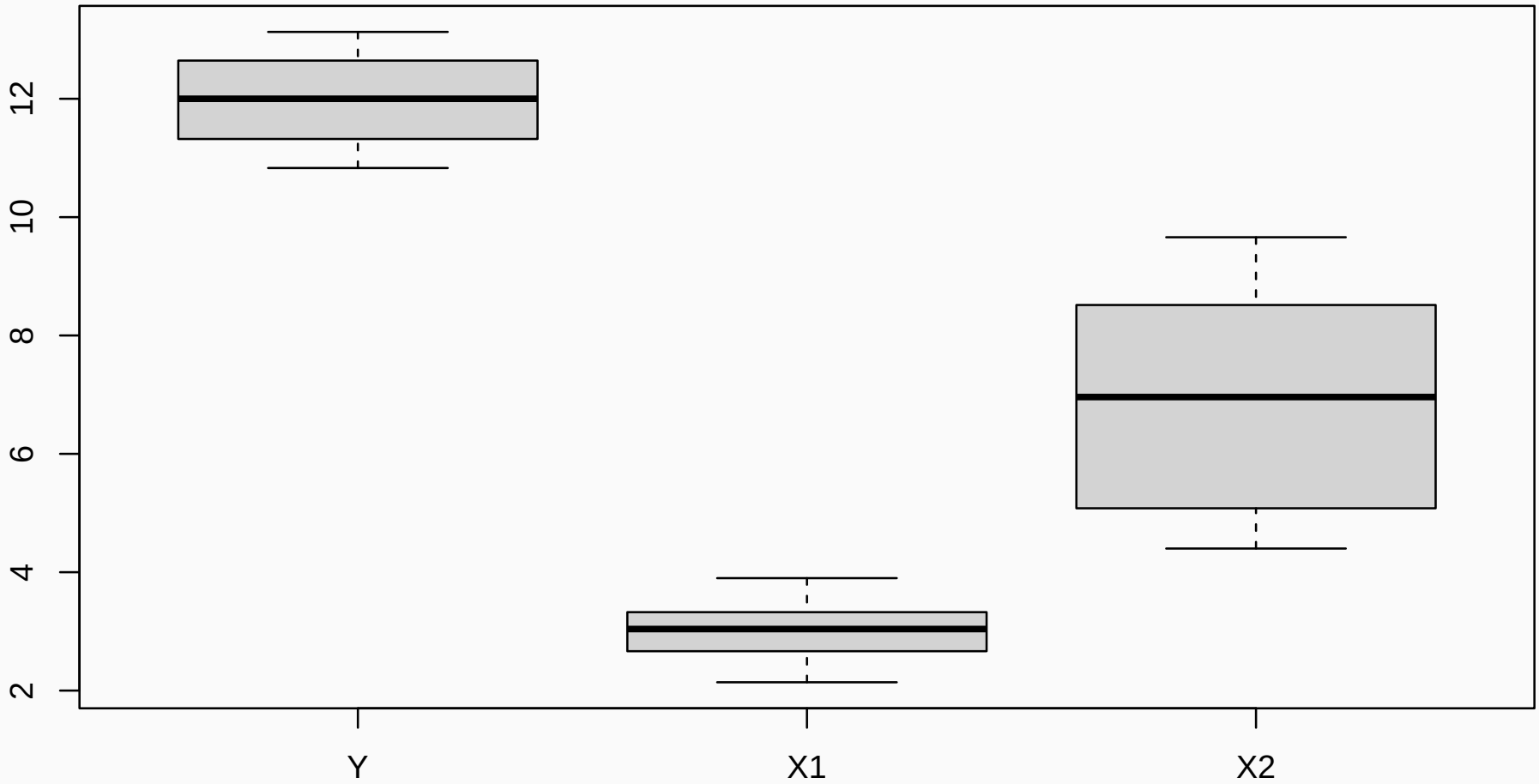
```
##         Y   X1   X2
## 1  12.37 2.23 9.66
## 2  12.66 2.57 8.94
## 3  12.00 3.87 4.40
## 4  11.93 3.10 6.64
## 5  11.06 3.39 4.91
## 6  13.03 2.83 8.52
## 7  13.13 3.02 8.04
## 8  11.44 2.14 9.05
## 9  12.86 3.04 7.71
## 10 10.84 3.26 5.11
## 11 11.20 3.39 5.05
## 12 11.56 2.35 8.51
## 13 10.83 2.76 6.59
## 14 12.63 3.90 4.90
## 15 12.46 3.16 6.96
```

Graphs before fitting a Model

1. Boxplot
2. Histogram
3. Pairsplot

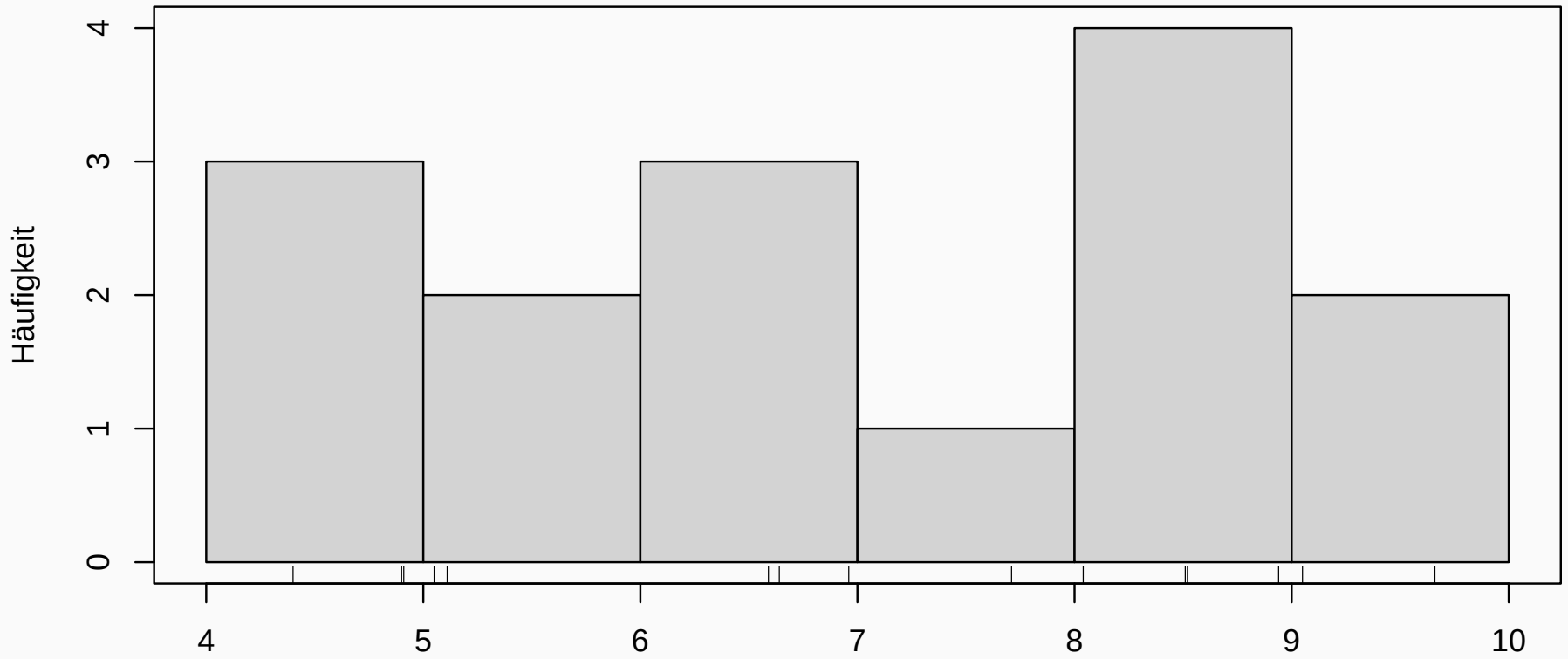
Boxplot

```
boxplot(P103)
```



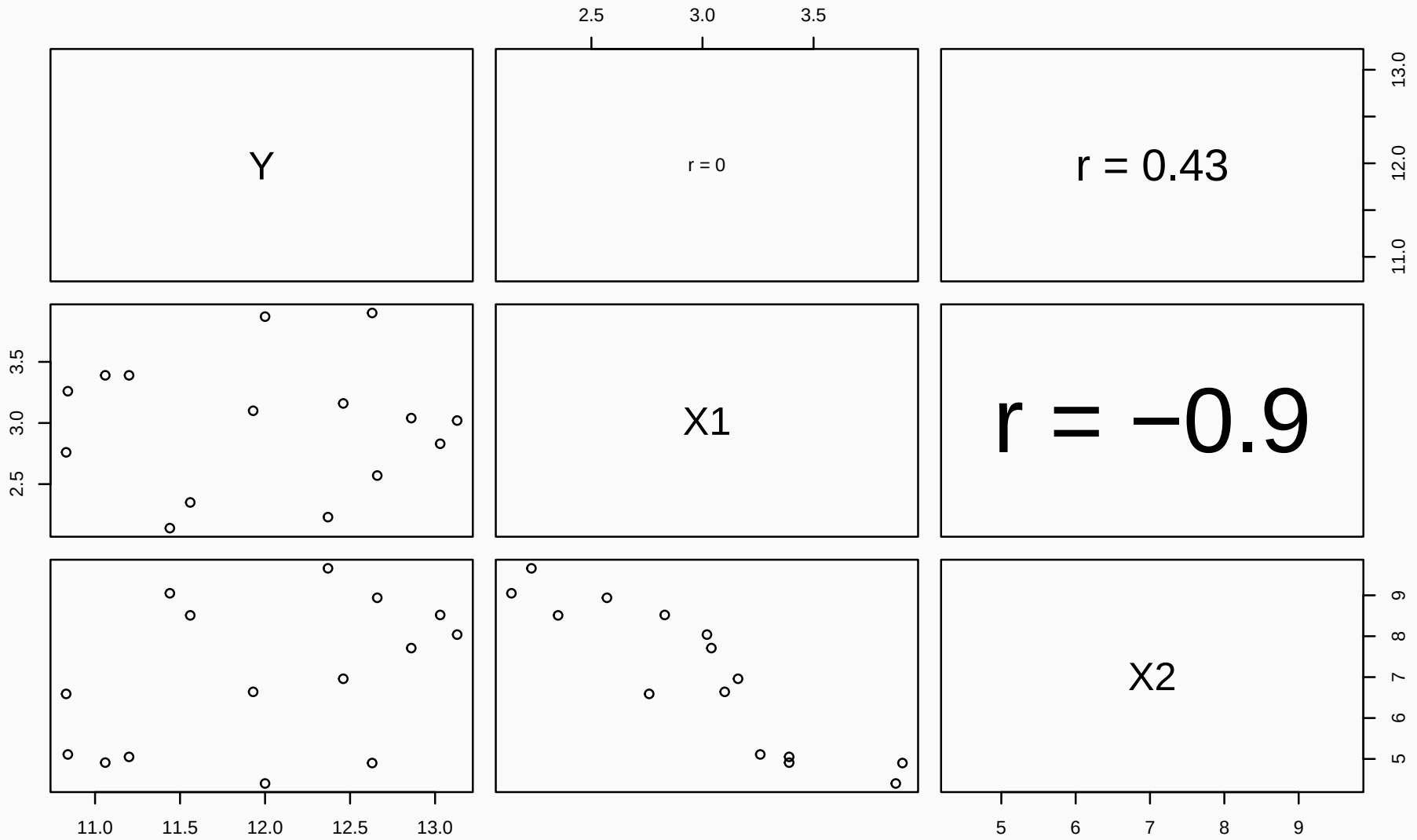
Histogram

```
hist(P103$X2, main = "", ylab = "Häufigkeit", xlab = "")
rug(P103$X2)
box()
```



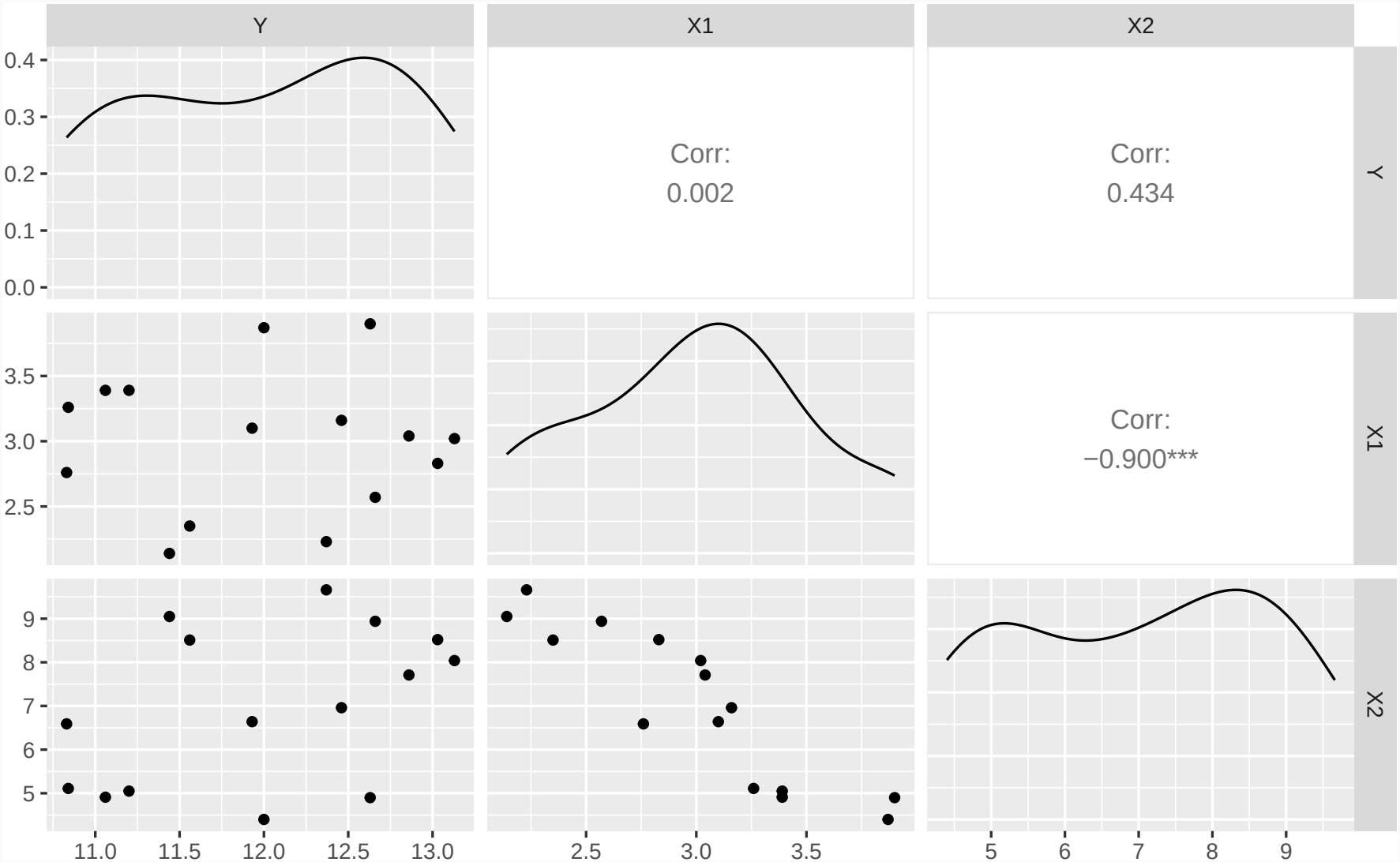
Pairsplot

```
pairs(P103, upper.panel=panel.cor)
```



Pairsplot

```
GGally::ggpairs(P103)
```



Pairsplot

Interpretation:

- The pairwise correlation should always be interpreted in conjunction with the scatter plots.
  - The correlation coefficient only measures **linear** dependence.
  - The correlation coefficient is **non-robust** and may be substantially influenced by few data points.
- The appearence of the scatter plot only serves as an indication of the results to be expected.
  - In *simple linear regression* the plot of Y and X is expected to **show a linear pattern**.
  - In mulitple linear regression the scatter plots between Y and each X **may or may not** show a linear pattern.

The absence of a linear pattern does not invalidate the linear model!

Graphs after fitting a Model

1. Graphs for checking the linearity and normality assumptions
2. Graphs for the detection of outliers and influential observations
3. Diagnostic plots for the effect of variables

Model Fitting `lm()`

```
mod <- lm(Y ~ 1 + X1 + X2, data=P103)
summary(mod)
```

```
##
## Call:
## lm(formula = Y ~ 1 + X1 + X2, data = P103)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.01363 -0.00945 -0.00228  0.00863  0.01632
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.51541    0.06114   -73.8   <2e-16 ***
## X1           3.09701    0.01227   252.3   <2e-16 ***
## X2           1.03186    0.00368   280.1   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0107 on 12 degrees of freedom
## Multiple R-squared:     1,   Adjusted R-squared:     1
## F-statistic: 3.92e+04 on 2 and 12 DF,  p-value: <2e-16
```

Model Fitting `olsrr::ols_regress()`

```
mod <- olsrr::ols_regress(Y ~ 1 + X1 + X2, data=P103)
mod
```

```
##                          Model Summary
## ----------------------------------------------------------------
## R                       1.000       RMSE                 0.010
## R-Squared               1.000       MSE                  0.000
## Adj. R-Squared          1.000       Coef. Var            0.089
## Pred R-Squared          1.000       AIC                -88.861
## MAE                     0.009       SBC                -86.029
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##  AIC: Akaike Information Criteria
##  SBC: Schwarz Bayesian Criteria
##
##                                 ANOVA
## --------------------------------------------------------------------
##                Sum of
##               Squares        DF    Mean Square        F          Sig.
## --------------------------------------------------------------------
## Regression      9.007         2          4.504    39222.343    0.0000
## Residual        0.001        12          0.000
## Total           9.009        14
## --------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t        Sig      lower      upper
## ----------------------------------------------------------------------------------------
## (Intercept)    -4.515         0.061                -73.851    0.000    -4.649    -4.382
##          X1     3.097         0.012        2.064   252.314    0.000     3.070     3.124
##          X2     1.032         0.004        2.292   280.079    0.000     1.024     1.040
```

Model Presentation: Stargazer

```
mod <- lm(Y ~ 1 + X1 + X2, data=P103)
stargazer::stargazer(mod, header=F, single.row = T)
```

Table 2

	Dependent variable:
	Y
X1	3.097^{***} (0.012)
X2	1.032^{***} (0.004)
Constant	-4.515^{***} (0.061)
Observations	15
R^2	1.000
Adjusted R^2	1.000
Residual Std. Error	0.011 (df = 12)
F Statistic	$39{,}222.000^{***}$ (df = 2; 12)
Note:	*p<0.1; **p<0.05; ***p<0.01

Model Presentation: Texreg

```
mod <- lm(Y ~ 1 + X1 + X2, data=P103)
texreg::texreg(mod)
```

	Model 1
(Intercept)	-4.52^{***}
	(0.06)
X1	3.10^{***}
	(0.01)
X2	1.03^{***}
	(0.00)
R^2	1.00
Adj. R^2	1.00
Num. obs.	15
$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$	

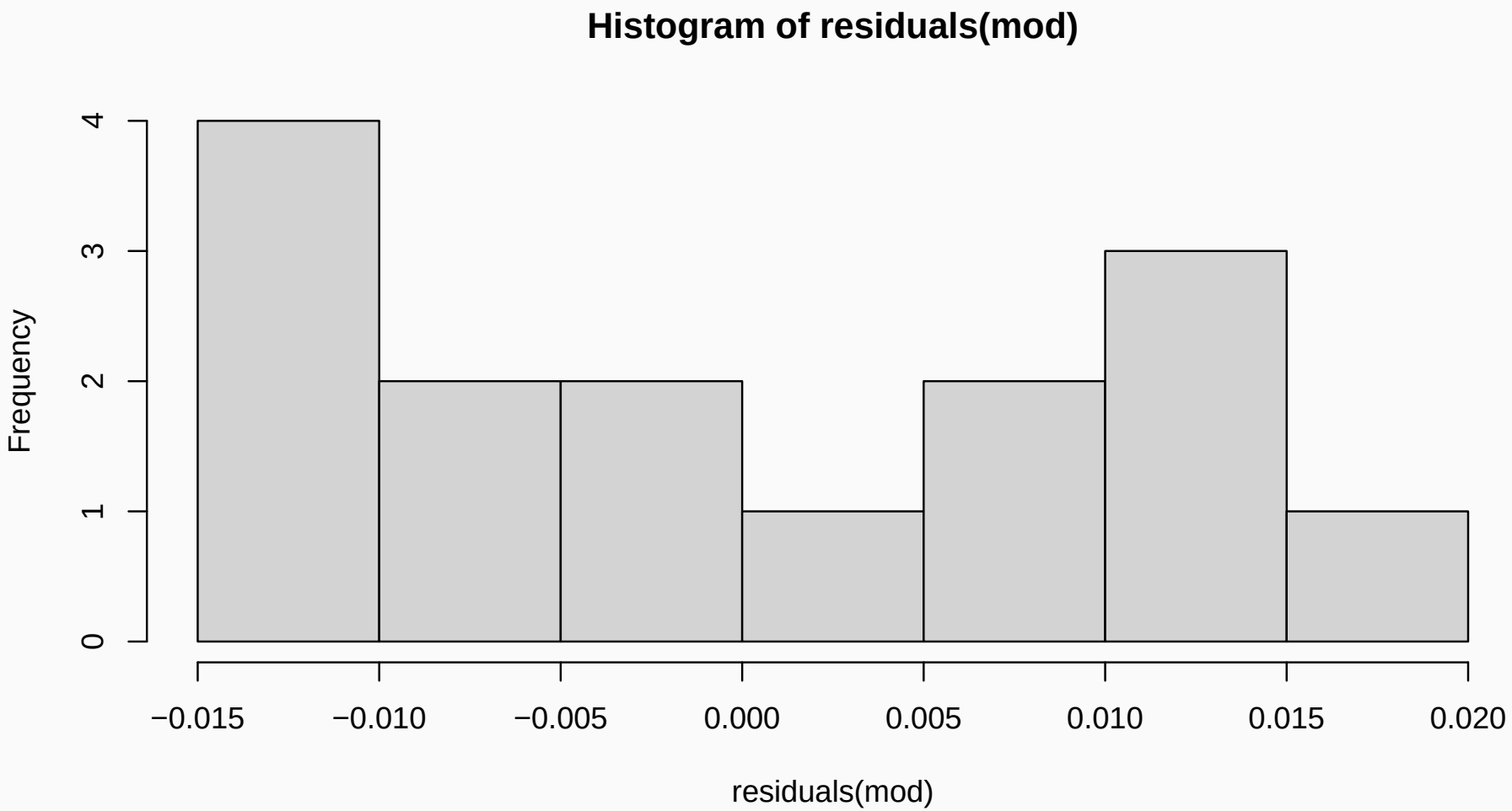
Table 3: Statistical models

Linearity and Normality

- **Residual Histogram:** Under the assumptions the histogram should resemble a normal distribution with symmetric shape and most observations around the center and few observations in the tails.
- **QQ-Plot:** Plot of the residuals versus the normals scores, which are what we would expect to obtain if the residuals were taken from a normal distribution. Under normality this plot should resemble a (nearly) straight line.
- **Residuals vs. Predictors:** The residuals should be uncorrelated with each of the predictors. If the assumptions hold, the plot should be a random scatter of points. Any pattern indicates a violation of an assumption, which often can be fixed using transformations.
- **Residuals vs. Fitted:** The residuals should also be uncorrelated with the fitted values, therefore this plot should also be a random scatter of points.

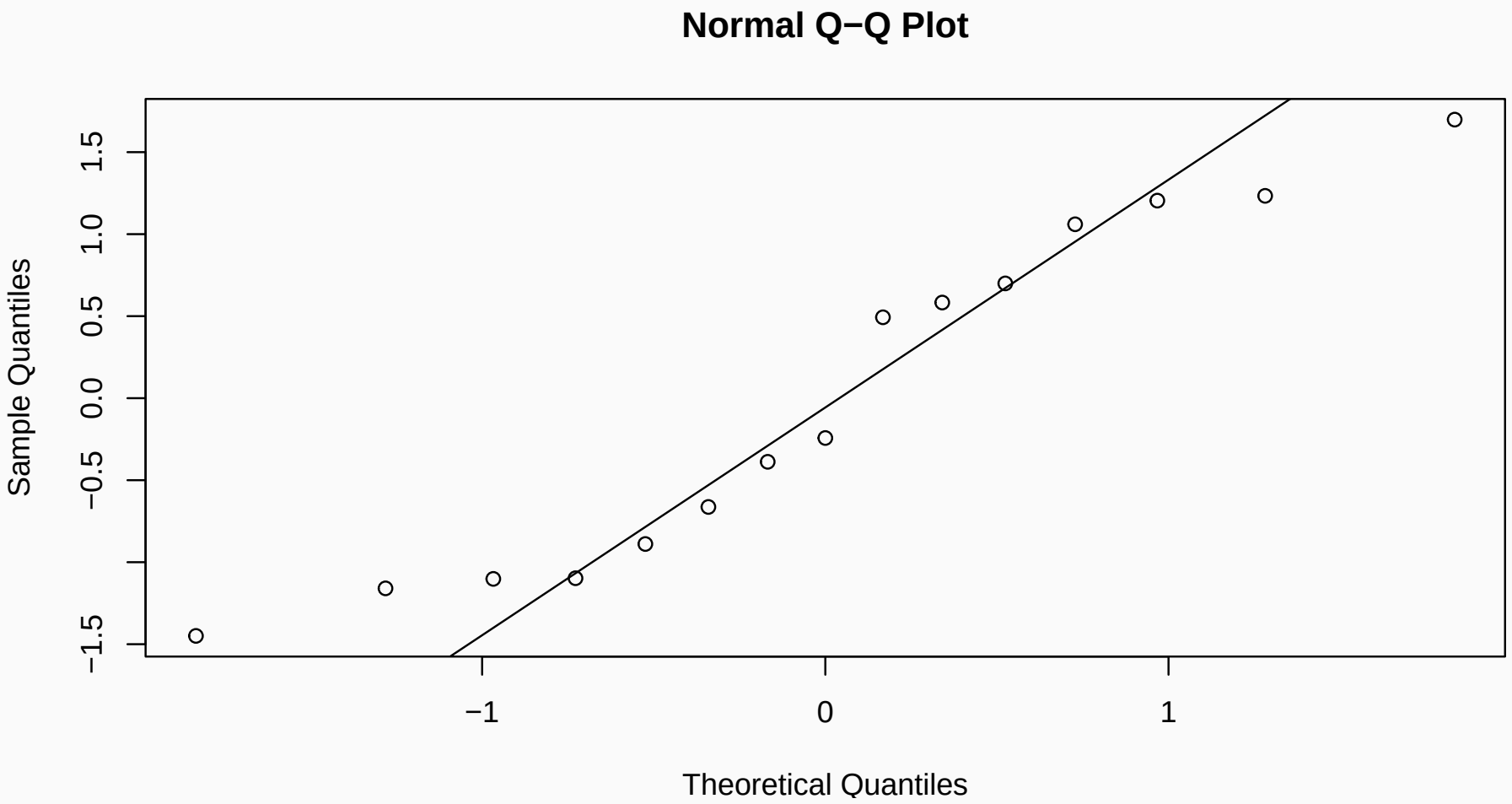
Histogram

```
hist(residuals(mod), density = )
```



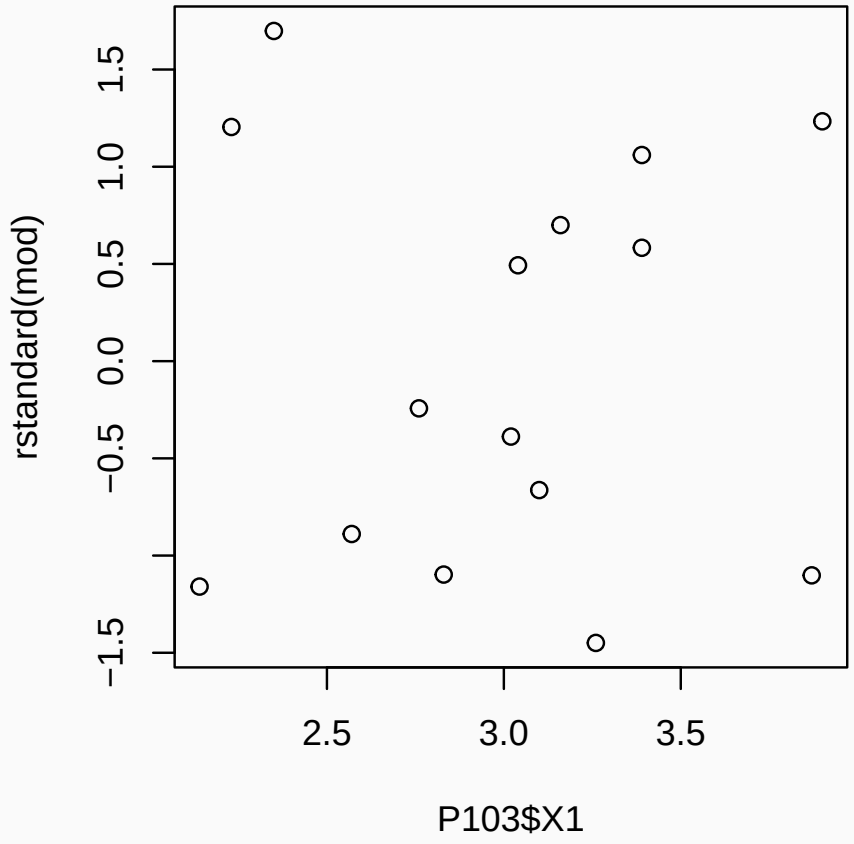
QQ-Plot (Standardized Residuals)

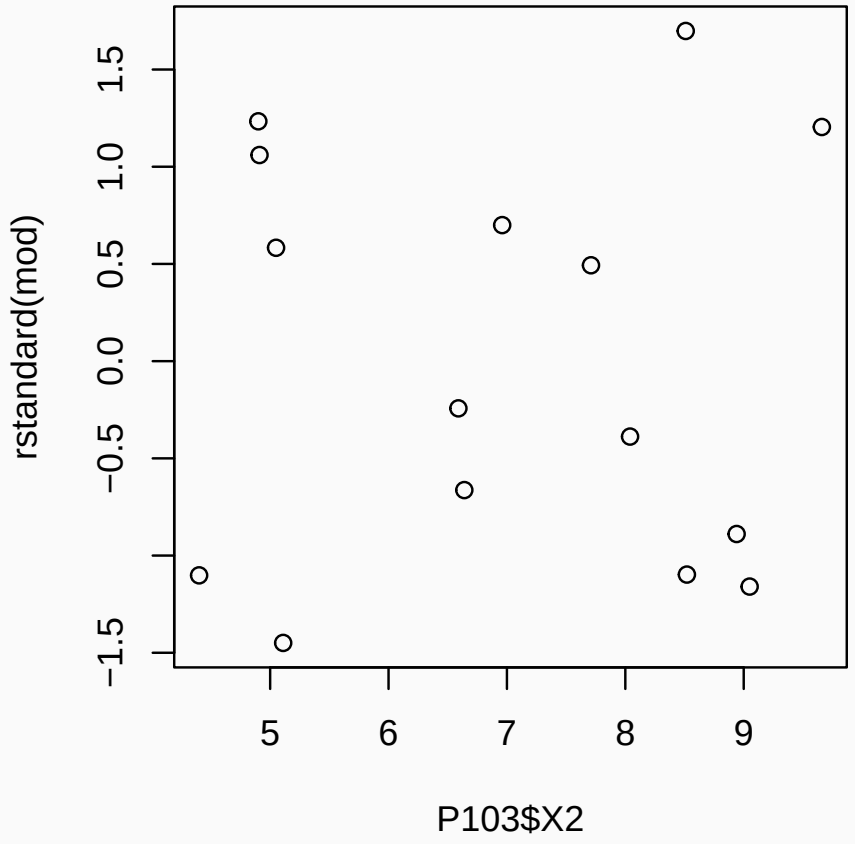
```
qqnorm(rstandard(mod))
qqline(rstandard(mod))
```



Standardized Residuals vs. Predictors

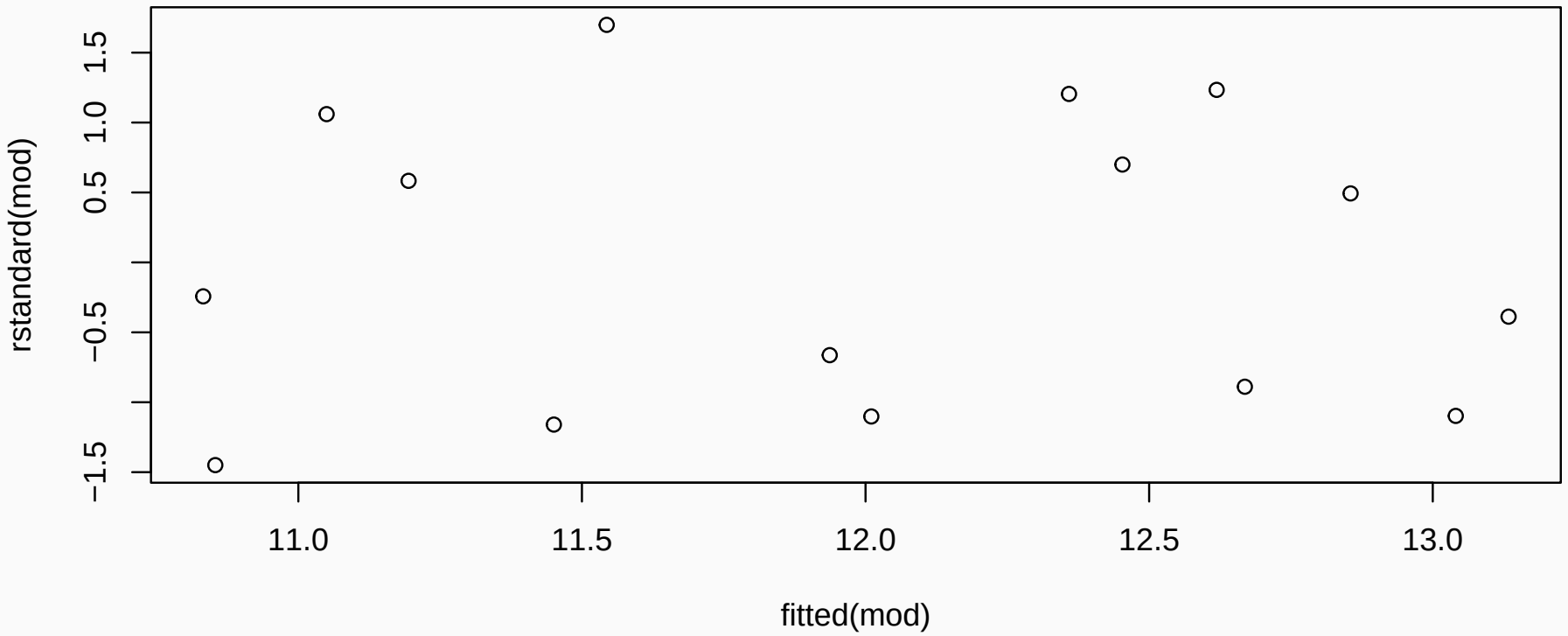
```
par(mfrow=c(1,2))
plot(y=rstandard(mod),x=P103$X1)
plot(y=rstandard(mod),x=P103$X2)
```





Standardized Residuals vs. fitted Values

```
par(mfrow=c(1,1))
plot(y=rstandard(mod), x=fitted(mod))
```



Leverage, Influence and Outliers

- We want to ensure that the model is not overly determined by one or a few observations. In multiple linear regression this cannot be simply detected graphically.
- When identifiying influential data points (points that drag the regression line in their direction) looking at residuals does not necessarily help.
- Influential points can be identified, when the regression coefficients (fitted values, *t*-Tests, etc) change heavily, when we **omit these points** while estimating the model.

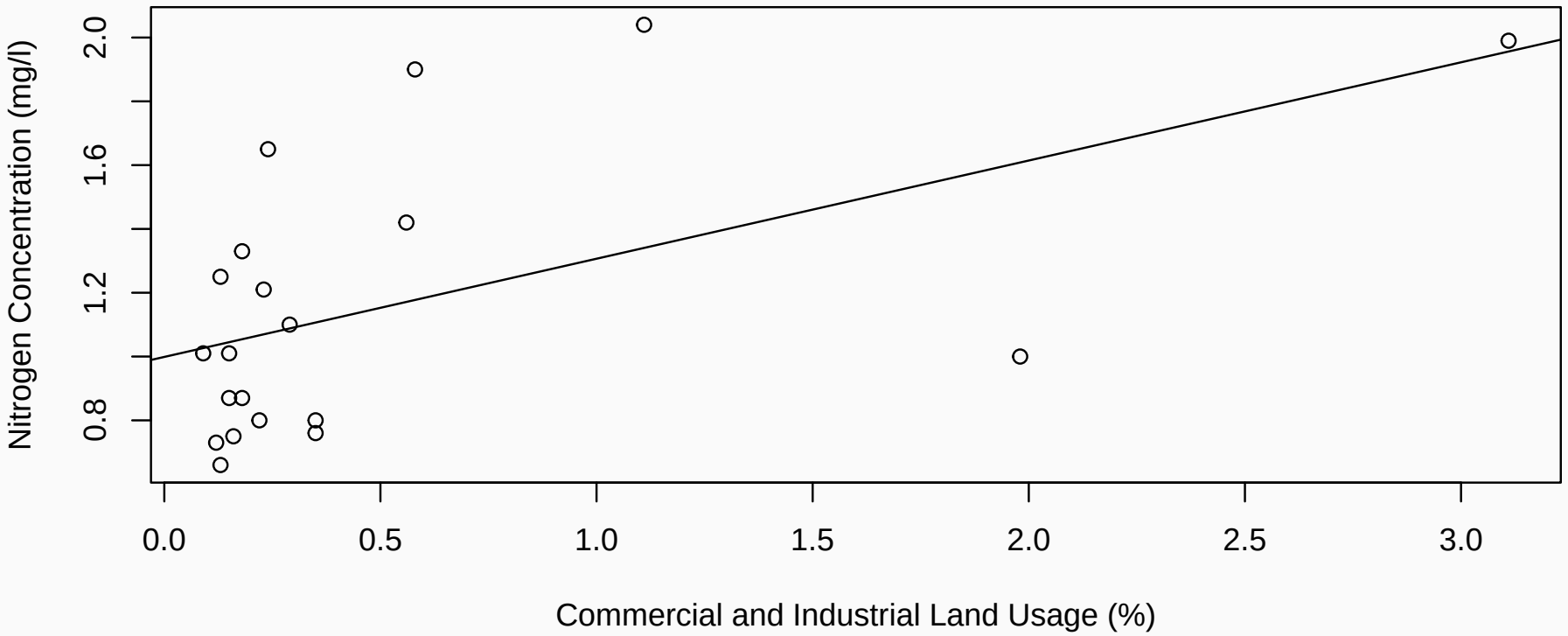
Outliers

- **Outliers in the Response Variable:** Observations with large standardized residuals are outliers in the response variable as they lie far from the fitted line (in Y direction). Outliers indicate a model failure for these observations.
- **Outliers in the Predictors:** Outliers can also occur in the X-Space. The leverage values p_{ii} allow to measure these discrepancies (based on distance from $\bar{x}$) and are called **high-leverage** points. It should be checked if those points are also **influential** before they are treated.

Inspecting the residuals is necessary but not sufficient as high-leverage points (usually twice the average size of $(p + 1)/n$) usually have small residuals.

Example: River Data

```
mod <- lm(Nitrogen ~ 1 + ComIndl, data = P010)
plot(y=P010$Nitrogen, x=P010$ComIndl, ylab="Nitrogen Concentration (mg/l)",
     xlab="Commercial and Industrial Land Usage (%)")
abline(mod)
```



Example: River Data

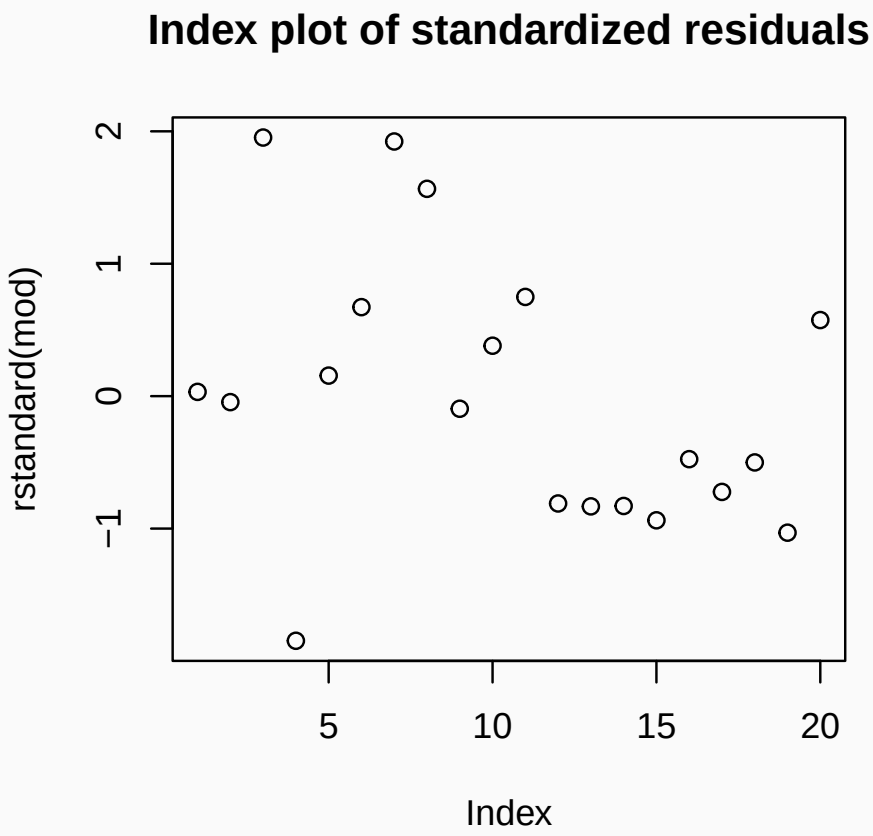
```
round(cbind(obs=P010$ComIndl,residuals=rstandard(mod), leverage=hatvalue
```

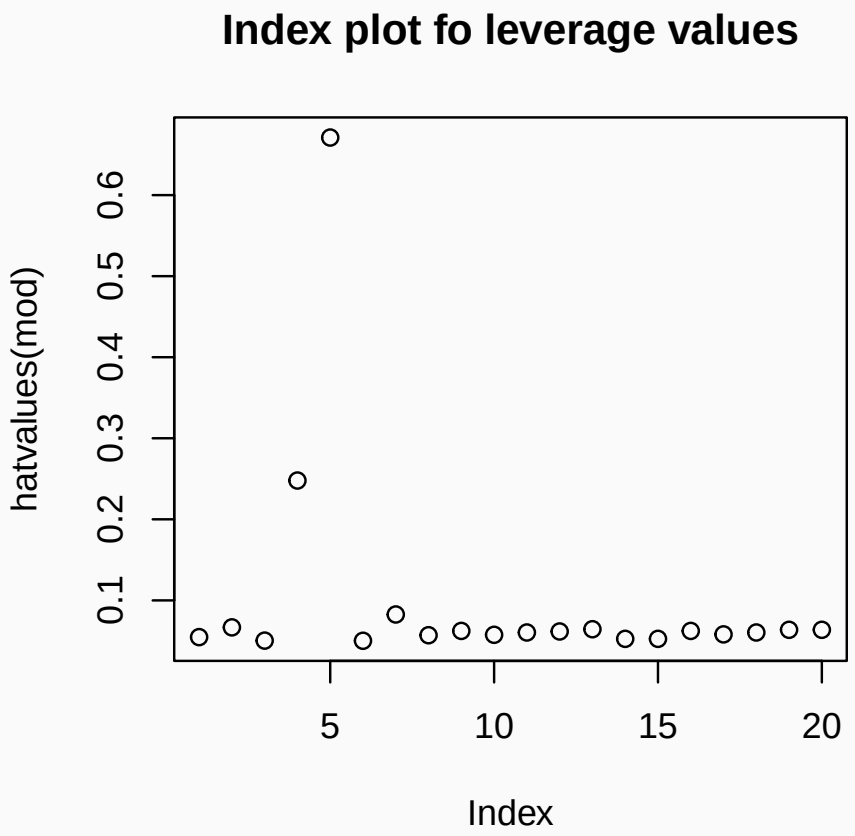
```
##                obs residuals leverage
## Olean         0.29      0.03     0.05
## Cassadaga     0.09     -0.05     0.07
## Oatka         0.58      1.95     0.05
## Neversink     1.98     -1.85     0.25
## Hackensack    3.11      0.16     0.67
## Wappinger     0.56      0.67     0.05
## Fishkill      1.11      1.92     0.08
## Honeoye       0.24      1.57     0.06
## Susquehanna   0.15     -0.10     0.06
## Chenango      0.23      0.38     0.06
## Tioughnioga   0.18      0.75     0.06
## West Canada   0.16     -0.81     0.06
## East Canada   0.12     -0.83     0.06
## Saranac       0.35     -0.83     0.05
## Ausable       0.35     -0.94     0.05
## Black         0.15     -0.48     0.06
## Schoharie     0.22     -0.72     0.06
## Raquette      0.18     -0.50     0.06
##  Oswegatchie  0.13     -1.03     0.06
## Cohocton      0.13      0.57     0.06
```

A small residual value is desireable and the standardized residuals show no outlier. However, the small residual is not due to a good fit, but due to high leverage of the observations for Neversink and Hackensack.

Example: River Data

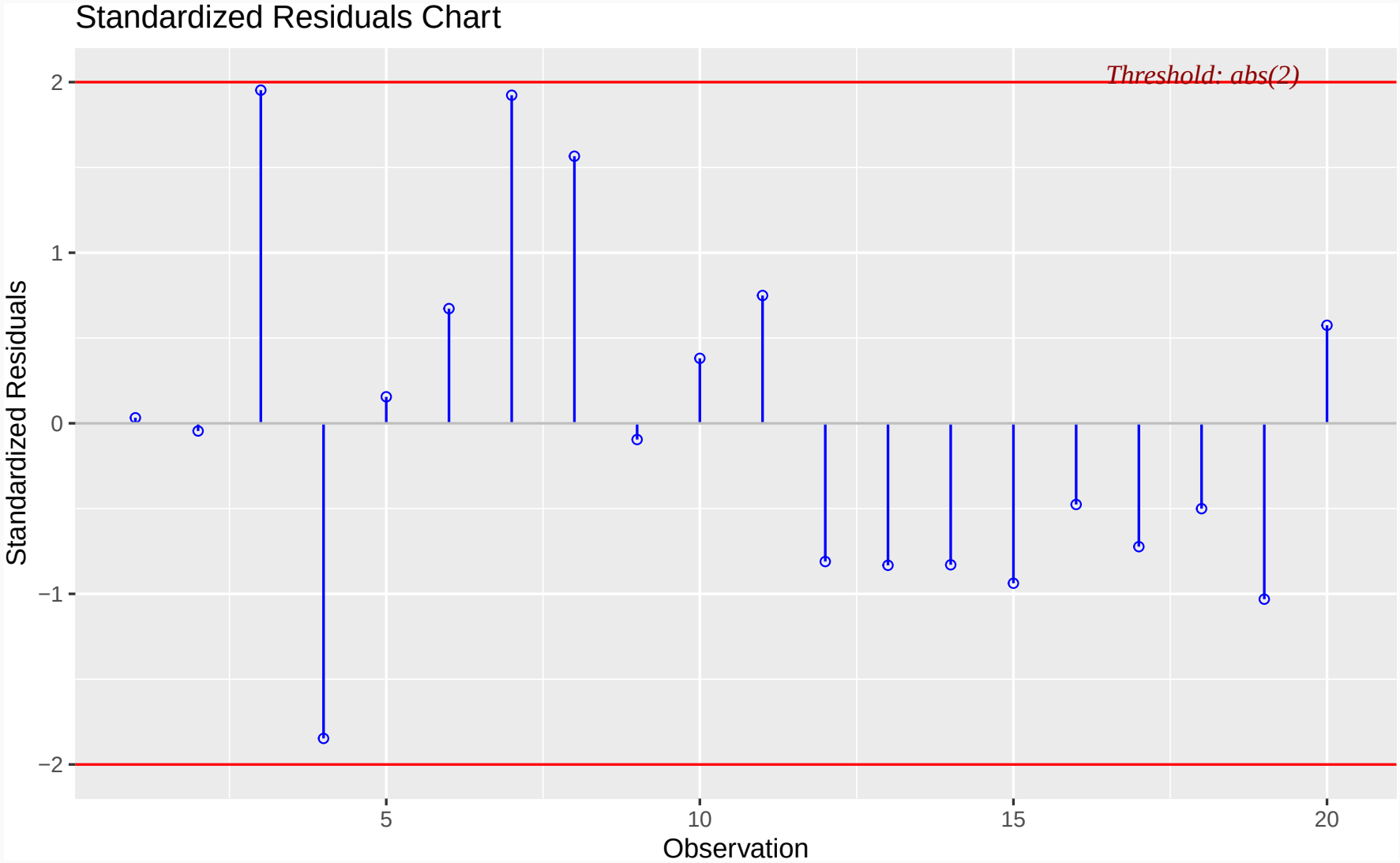
```
par(mfrow=c(1,2))
plot(rstandard(mod), main="Index plot of standardized residuals")
plot(hatvalues(mod), main="Index plot fo leverage values")
```





Example: River Data

```
olsrr::ols_plot_resid_stand(mod)
```



Influence

- The influence of an observation may be measured by the effects on the fit when it is omitted from the data in the fitting process.
- $\hat{\beta}_{0(i)}, \hat{\beta}_{1(i)}, \ldots, \hat{\beta}_{p(i)}$ denote the regression coefficients obtained when the i-th observation is deleted. $\hat{y}_{1(i)}, \hat{y}_{2(i)}, \ldots, \hat{y}_{n(i)}$ and $\hat{\sigma}^2_{(i)}$ denote the predicted values and residual mean square error when dropping the i-th observation respectively. The resulting observation equation for the model follows by

$$\hat{y}_{m(i)} = \hat{\beta}_{0(i)} + \hat{\beta}_{1(i)} x_{m1} + \ldots + \hat{\beta}_{p(i)} x_{mp}$$

- Measures to assess influence usually look at differences produced in $\hat{\beta}_j - \hat{\beta}_{j(i)}$ or $\hat{y}_j - \hat{y}_{j(i)}$.

Cook's Distance

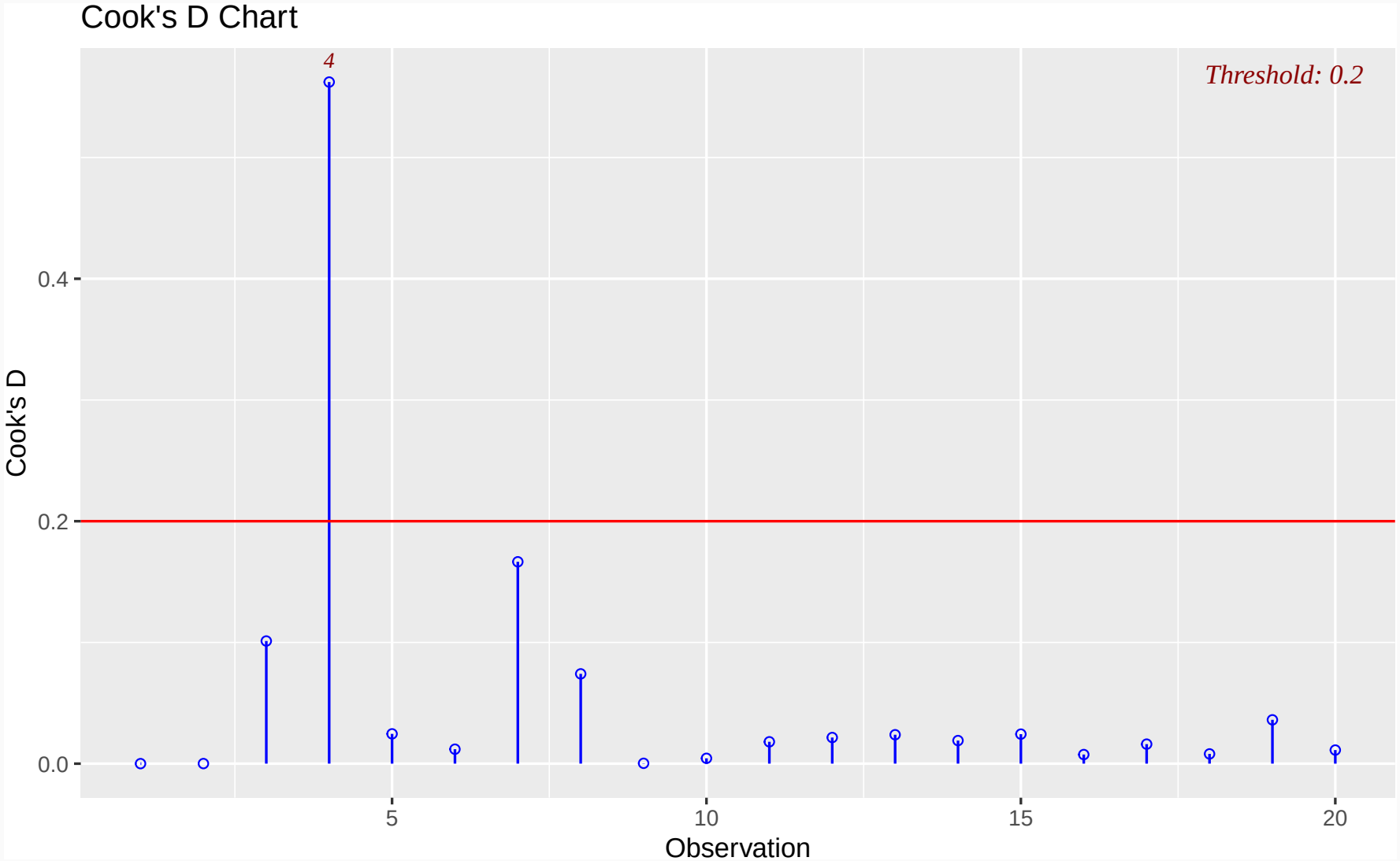
- *Cook's distance* measures the difference between the regression coefficients obtained from the full data and the regression coefficients obtained by deleting the i-th observation.
- The influence of the i-th observation for $i = 1, \ldots, n$ is given by

$$C_i = \frac{\sum_{j=1}^{n} (\hat{y}_j - \hat{y}_{j(i)})^2}{\hat{\sigma}^2 (p+1)} = \frac{r_i^2}{p+1} \cdot \frac{p_{ii}}{1 - p_{ii}}$$

- Cook's distance is a product of the squared residual and the so called **potential function** $p_{ii}/(1 - p_{ii})$.
- If C_i is large omitting a data point will cause large changes in the model. Points are said to be influential when their Cook's distance meets or exceeds the 50% point of the F-distribution with $p + 1$ and $n - p - 1$ degrees of freedom. A rule of thumb is do investigate points where $C_i \geq 1$.

Cook's Distance

```
olsrr::ols_plot_cooksd_chart(mod)
```



Welsch and Kuh Measure (DFITS)

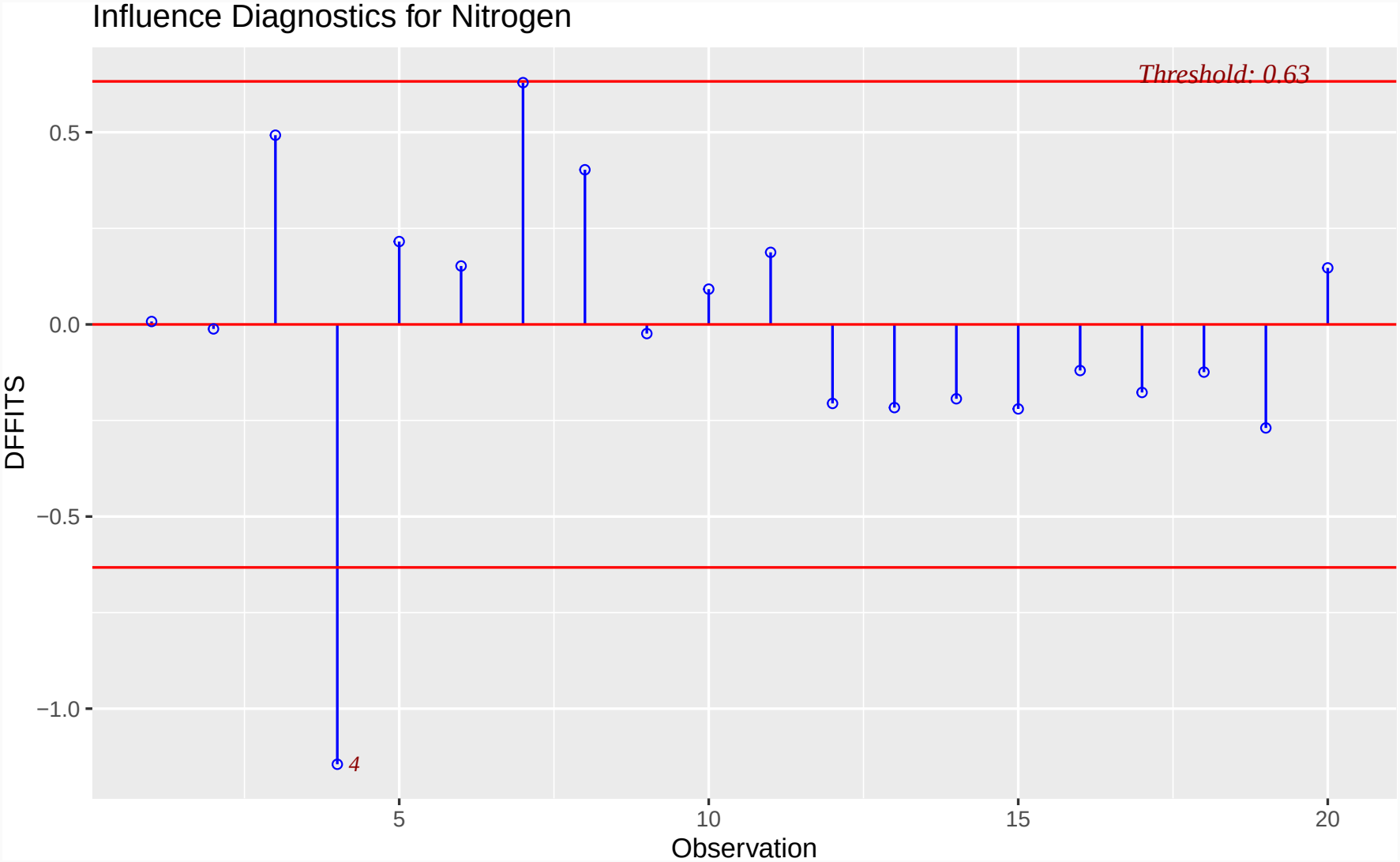
- The measure proposed by Welsch and Kuh is the scaled difference between the i-th fitted value obtained from the full data and the i-th fitted value obtained by deleting the respective observation.

$$DFITS_i = \frac{\hat{y}_i - \hat{y}_{i(i)}}{\sigma^2_{(i)}\sqrt{p_{ii}}} = r_i^* \sqrt{\frac{p_{ii}}{1 - p_{ii}}}$$

- Points with $|DFITS_i| > 2\sqrt{(p+1)/(n-p-1)}$ are usually classified as influential.
- C_i and $DFITS_i$ are functions of the residual and leverage values. It is often sufficient to inspect either the Cook's distance or *DFITS*.

Welsch and Kuh Measure (DFITS)

```
olsrr::ols_plot_dffits(mod)
```



What to do with outliers?

- Outliers and influential observations should not routinely be deleted or automatically down-weighted.
- **Points with high leverage and high influence are not necessarily bad observations!** They can be an indication of model misspecification (e.g. non-linearity in the data) or show that the data did not come from a normal polulation. **In those cases they may be the most informative data points.**
- Each *outlier* should be inspected with care and checked individually. If data points are removed this should be documented in the research including reasons for the decision.
- Another option is to use *robust regression* where less weight is given to data points with high leverage.

Role of Variables in Regression Models

- Predictors are usually **sequentially introduced** into a regression equation.
- Given a model that contains p predictors, what is the effect of deleting (or adding) one of the variables from (or to) the model?
- One indication can be obtained by the t-test. If the t-value is large in absolute terms, the variables will be retained, otherwise omitted.
- **The results of the t-test will only be valid if the underlying assumptions hold, therefore additional graphs should be inspected when deciding whether or not to include a variable in the regression model.

Example: Scottish Hills Races Data

P120

```
##                            Time Distance Climb
## Greenmantle New Year Dash   965      2.5   650
## Carnethy                   2901      6.0  2500
## Craig Dunain               2019      6.0   900
## Ben Rha                    2736      7.5   800
## Ben Lomond                 3736      8.0  3070
## Goatfell                   4393      8.0  2866
## Bens of Jura              12277     16.0  7500
## Cairnpapple                2182      6.0   800
## Scolty                     1785      5.0   800
## Traprain Law               2385      6.0   650
## Lairig Ghru               11560     28.0  2100
## Dollar                     2583      5.0  2000
## Lomonds of Fife            3900      9.5  2200
## Cairn Table                2648      6.0   500
## Eildon Two                 1616      4.5  1500
## Cairngorm                  4335     10.0  3000
## Seven Hills of Edinburgh   5905     14.0  2200
## Knock Hill                 4719      3.0   350
## Black Hill                 1045      4.5  1000
## Creag Beag                 1954      5.5   600
## Kildoon                     957      3.0   300
## Meall Ant-Suiche           1674      3.5  1500
## Half Ben Nevis             2859      6.0  2200
## Cow Hill                   1076      2.0   900
## North Berwick Law          1121      3.0   600
## Creag Dubh                 1573      4.0  2000
## Burnswark                  2066      6.0   800
## Largo                      1714      5.0   950
## Criffel                    3030      6.5  1750
## Achmony                    1257      5.0   500
## Ben Nevis                  5135     10.0  4400
```

Example: Scottish Hills Races Data

```
lapply(P120, summary)
```

```
## $Time
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     957    1680    2385    3473    4118   12277
##
## $Distance
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2.00    4.50    6.00    7.53    8.00   28.00
##
## $Climb
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     300     725    1000    1815    2200    7500
```

Example: Scottish Hills Races Data

```
mod <- lm(Time ~ 1 + Distance + Climb, data=P120)
summary(mod)
```

```
##
## Call:
## lm(formula = Time ~ 1 + Distance + Climb, data = P120)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##   -973   -428    -71    142   3907
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -539.483    258.161   -2.09    0.045 *
## Distance     373.073     36.068   10.34  9.9e-12 ***
## Climb          0.663      0.123    5.39  6.4e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 881 on 32 degrees of freedom
## Multiple R-squared:  0.919,  Adjusted R-squared:  0.914
## F-statistic:  182 on 2 and 32 DF,  p-value: <2e-16
```

Added-Variable Plot

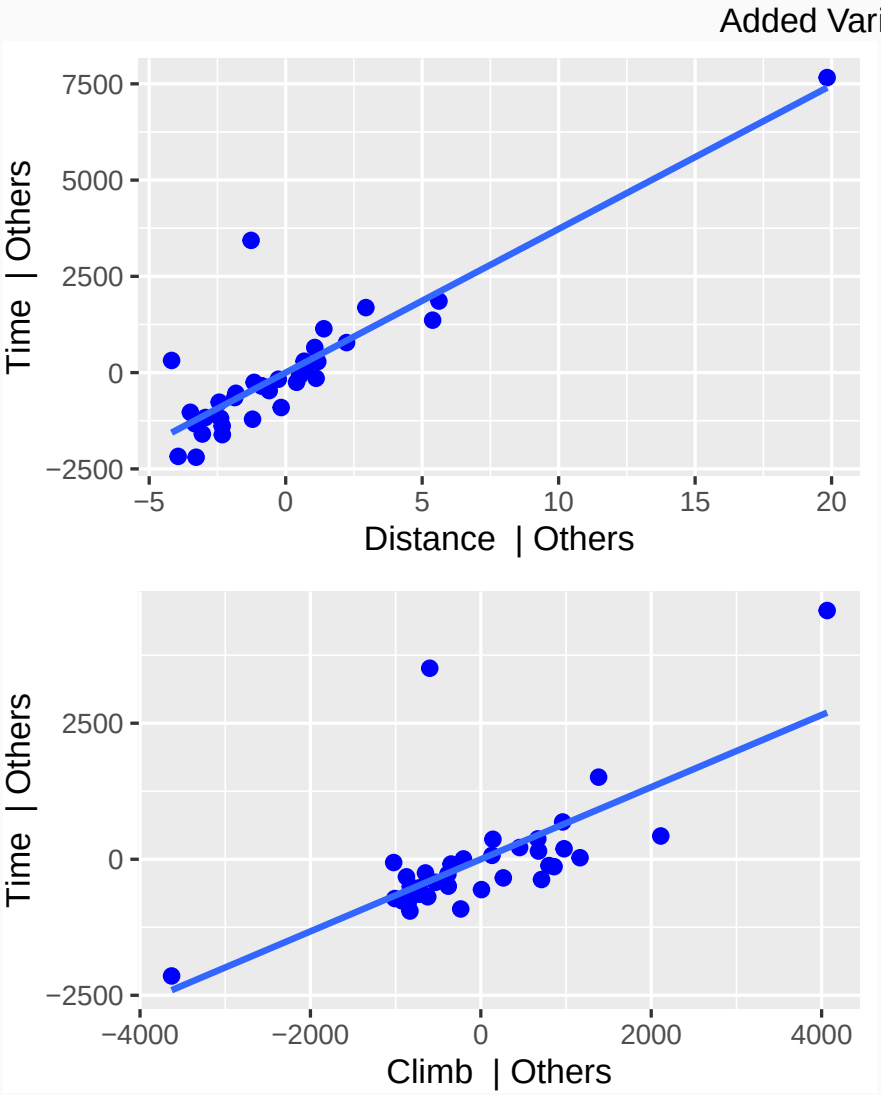
- The added-variable plot enables us to see the magnitude of the regression coefficient of the new variable that is being considered for inclusion.
- The slope of the least squares line representing the points in the plot is equal to the estimated regression coefficient of the new variable. Additionally, the plot shows data points that may be influential for this magnitude.

Added-Variable Plot

- The added variable plot is a plot between two pairs of residuals: the residuals when Y is regressed on predictors except X_j versus the residuals when regressing X_j on all other predictors.
  - First set of residuals corresponds to the remainder of Y that cannot be explained by the other regressors.
  - Second set of residuals corresponds to the part of X_j that cannot be explained by the other predictors.
  - The resulting slope is equal to $\hat{\beta}_j$ and essentially a visualization that is equivalent to the interpretation as partial regression coefficient.

Added-Variable Plot

```
olsrr::ols_plot_added_variable(mod)
```



Your turn

Do both variables contribute significantly to the model? Is there anything unusual that should be investigated further?

Effects of an additional Predictor

- When a new regressor is introduced two questions should be answered:
  - Is the regression coefficient of the new variable significant (different from zero)?
  - Does the introduction of the new variable substantially change the regression coefficients that are already in the model.

Effects of an additional Predictor

- **Option A:** insignificant and almost no change in coefficients $\rightarrow$ Should **not** be included unless theory dictates inclusion.
- **Option B:** significant and subtantial changes in coefficients. $\rightarrow$ Should be included, but needs to be checked for collinearity.
- **Option C:** significant but no substantial coefficient changes $\rightarrow$ Ideal condition as variable is uncorrelated with other regressors, variable should be retained.
- **Option D:** insignificant but substantial coefficient changes $\rightarrow$ indicates collinearity, corrective actions are required before discussion.