University of St.Gallen

Penalized Regression

University of St. Gallen
School of Management, Economics, Law,
Social Sciences, International Affairs
and Computer Science

Assignment 2

Data Analytics I: Predictive Econometrics
Prof. Jana Mareckova

submitted by

Cyril Janak, 16-611-287
Jonas Husmann, 16-610-917
Niklas Kampe, 16-611-618
Robin Scherrer, 18-617-969

08.12.2021

Contents

Requirements

To solve the following tasks, the required libraries and the data sets are loaded first.

```
library(glmnet)
library(corrplot)
library(ggplot2)

load("GHA/student-mat-train.RData")
load("GHA/student-mat-test.RData")
```

Exercise 1

There are 214 observations in the training data set and 143 observations in the test data set.

```
(n_obs_train <- nrow(train))
```

```
## [1] 214
```

```
(n_obs_test <- nrow(test))
```

```
## [1] 143
```

Exercise 2

The average grade is ~11.64, the minimum grade is 4 and the maximum grade is 19. All numbers were calculated using the training data.

```
(avg_grade <- mean(train$G3))
```

```
## [1] 11.64019
```

```
(min_grade <- min(train$G3))
```

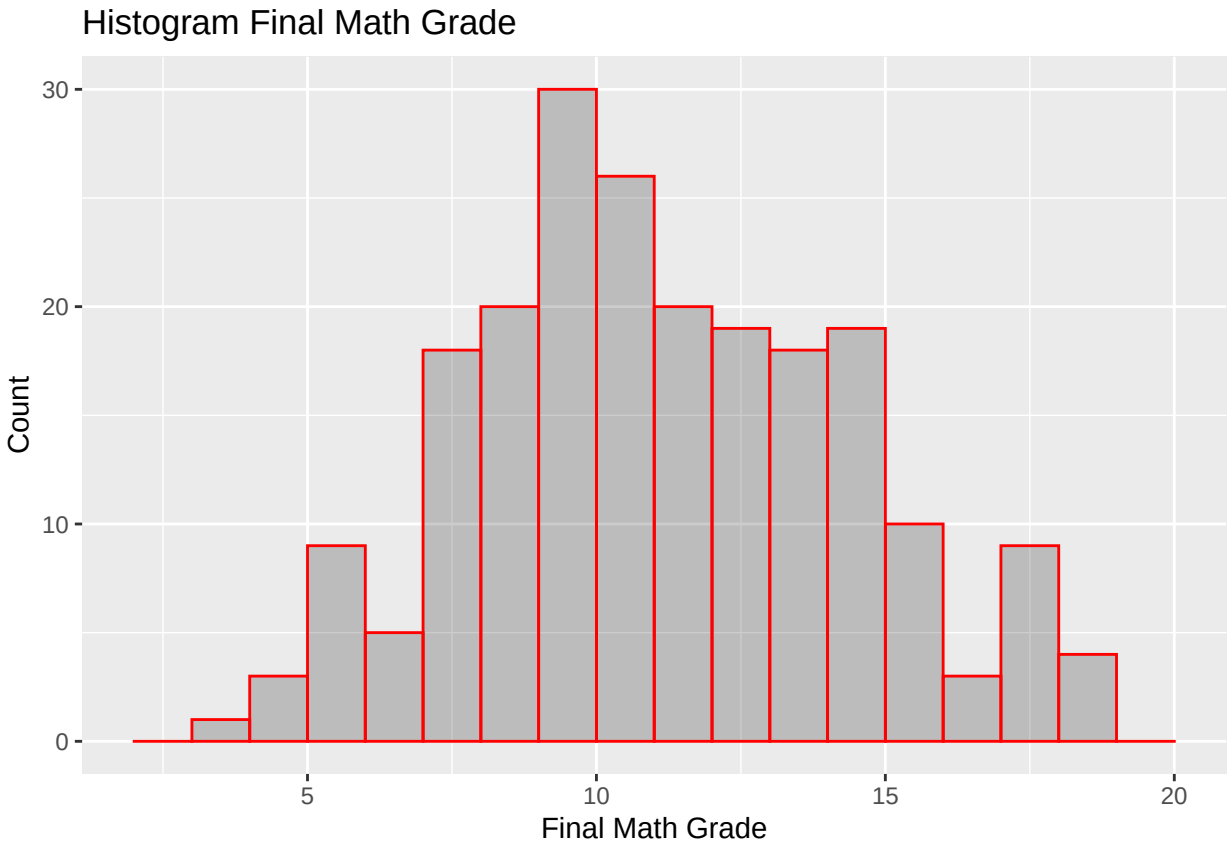
```
## [1] 4
```

```
(max_grade <- max(train$G3))
```

```
## [1] 19
```

Exercise 3

```
(final_grade_hist <- ggplot(data=train, aes(G3)) +
    geom_histogram(breaks=seq(2,20, by=1),
                   col="red",
                   fill="black",
                   alpha = 0.2)+
    labs(title="Histogram Final Math Grade", x="Final Math Grade", y="Count"))
```



Exercise 4

When doing causal modeling there are independent variables (x_1,..,x_n) which are considered as the cause of the dependent variable (y), therefore one would expect a direct impact of the independent variables on the dependent variable. For predictive modelling the goal is to establish a method that allows to make predictions of the dependent variable (y) based on the known independent variables (x_1,..,x_n).