Fakultät für Informatik

Studiengang Studiengang-einsetzen

Titel

Master Thesis/Bachelor Thesis

von

Emma C. Reinecke

Datum der Abgabe: tt.mm.jjjj
Erstprüfer: Prof. Dr.
Zweitprüfer: Prof. Dr.

Inhaltsverzeichnis

1 Quarto **2**

2 Running Code **4**

3 Introduction **4**

4 Exploratory data analysis **5**

4.1 Data visualization 5

4.2 Summary statistics 7

5 Modeling **7**

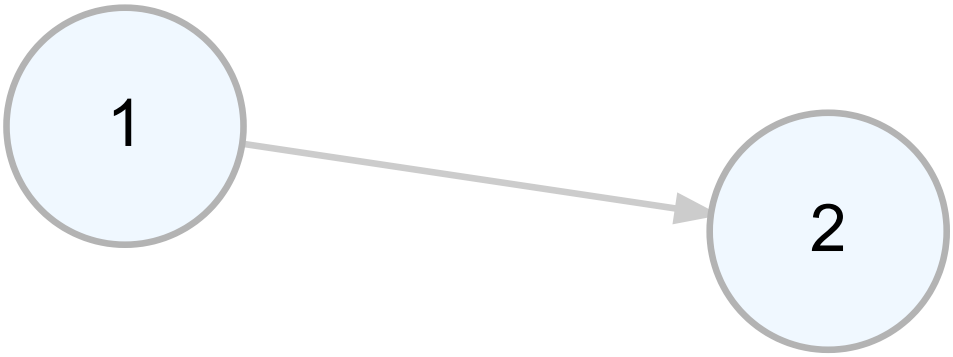
Literaturverzeichnis **9**

1 Quarto

```
library(openintro)  # for data
library(tidyverse)  # for data wrangling and visualization
library(knitr)      # for tables
library(broom)      # for model summary
library(DiagrammeR)
```

The average fuel efficiency for cars from 1974 was 20.1 miles per gallon.

```
#requires: library(DiagrammeR)
create_graph() %>%
  add_node() %>%
  add_node() %>%
  add_edge(from = 1, to = 2) %>%
  render_graph()
```



Quarto enables you to weave together content and executable code into a finished document. To learn more about Quarto see https://quarto.org. Grundlage für dieses qmd Quarto Authoring: https://quarto.org/docs/get-started/authoring/rstudio.html Ein einfaches Diagramm aus diesem Tutorial:

2 Running Code

When you click the **Render** button a document will be generated that includes both content and the output of embedded code. You can embed code like this:

```
1 + 1
```

```
[1] 2
```

You can add options to executable code like this

```
[1] 4
```

The `echo: false` option disables the printing of code (only output is displayed).

3 Introduction

Let's start by loading the packages we'll use for the analysis. See this analysis using literate programming.[1]

Grundlage für dieses qmd

$$price = \hat{\beta}_0 + \hat{\beta}_1 \times area + \epsilon \tag{1}$$

Let's start by loading the packages we'll use for the analysis. See this analysis using literate programming.[2]

We present the results of exploratory data analysis in Kapitel 4 and the regression model in Kapitel 5.

[1] Vgl. D. E. Knuth, „Literate Programming“, *The Computer Journal* 27, Nr. 2 (1. Februar 1984): 77, https://doi.org/10.1093/comjnl/27.2.97.

[2] Vgl. Knuth, „Literate Programming“, 77.

4 Exploratory data analysis

The data contains 98 houses. As part of the exploratory analysis let's visualize and summarize the relationship between areas and prices of these houses.

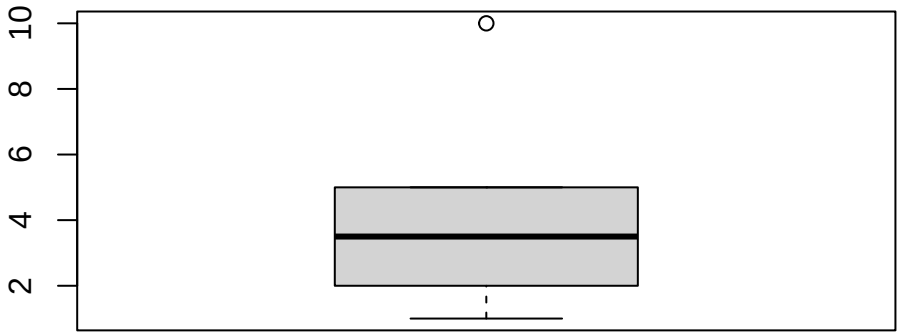


Abbildung 1: This dataset has 6 observations.

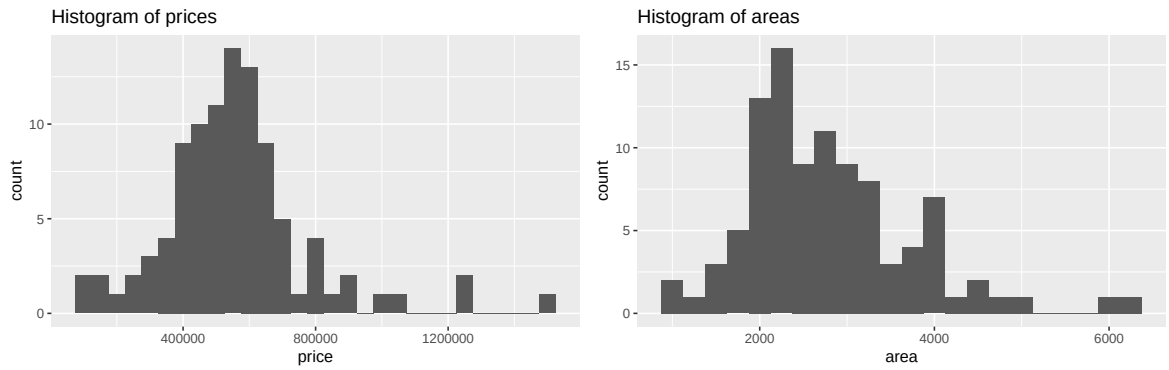
4.1 Data visualization

Abbildung 2 shows two histograms displaying the distributions of `price` and `area` individually.

```
ggplot(duke_forest, aes(x = price)) +
  geom_histogram(binwidth = 50000) +
  labs(title = "Histogram of prices")

ggplot(duke_forest, aes(x = area)) +
  geom_histogram(binwidth = 250) +
  labs(title = "Histogram of areas")
```

Abbildung 3 displays the relationship between these two variables in a scatterplot.



(a) Histogram of `prices`

(b) Histogram of `areas`

Abbildung 2: Histograms of individual variables

```
ggplot(duke_forest, aes(x = area, y = price)) +
  geom_point() +
  labs(title = "Price and area of houses in Duke Forest")
```

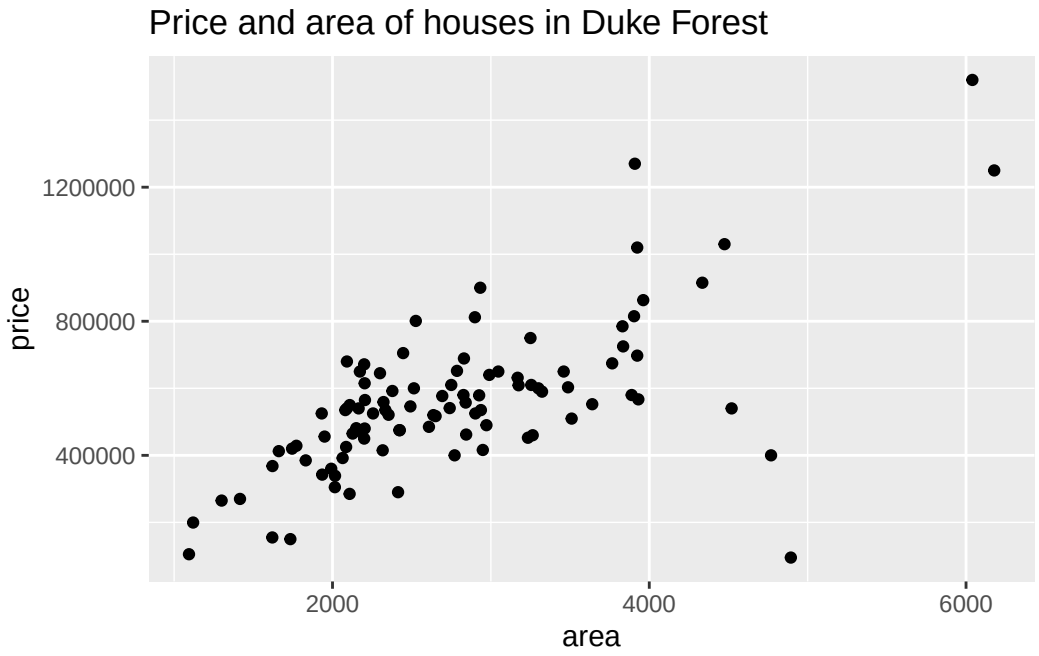


Abbildung 3: Scatterplot of price vs. area of houses in Duke Forest

4.2 Summary statistics

Tabelle 1 displays basic summary statistics for these two variables. Knuth[3] sagt blablah. Für weitere Zitierungen und Fußnoten: qmd citation. Ansonsten ist hier eine Fußnotenreference,[4] and another.[5]

This paragraph won't be part of the note, because it isn't indented. Here is an inline note.[6]

```
duke_forest %>%
  summarise(
    `Median price` = median(price),
    `IQR price` = IQR(price),
    `Median area` = median(area),
    `IQR area` = IQR(area),
    `Correlation, r` = cor(price, area)
    ) %>%
  kable(digits = c(0, 0, 0, 0, 2))
```

Tabelle 1: Summary statistics for price and area of houses in Duke Forest

Median price	IQR price	Median area	IQR area	Correlation, r
540000	193125	2623	1121	0.67

5 Modeling

We can fit a simple linear regression model of the form shown in Gleichung 2.

[3] „Literate Programming“, 33.

[4] Here is the footnote.

[5] Here's one with multiple blocks.

Subsequent paragraphs are indented to show that they belong to the previous footnote.

```
{ some.code }
```

The whole paragraph can be indented, or just the first line. In this way, multi-paragraph footnotes work like multi-paragraph list items.

[6] Inlines notes are easier to write, since you don't have to pick an identifier and move down to type the note.

$$price = \hat{\beta}_0 + \hat{\beta}_1 \times area + \epsilon \tag{2}$$

Tabelle 2 shows the regression output for this model.

```
price_fit <- lm(price ~ area, data = duke_forest)


price_fit %>%
  tidy() %>%
  kable(digits = c(0, 0, 2, 2, 2))
```

Tabelle 2: Linear regression model for predicting price from area

term	estimate	std.error	statistic	p.value
(Intercept)	116652	53302.46	2.19	0.03
area	159	18.17	8.78	0.00

We present the results of exploratory data analysis in Kapitel 4 and the regression model in Kapitel 5.

Abbildung 3 displays the relationship between these two variables in a scatterplot.

Tabelle 1 displays basic summary statistics for these two variables.

We can fit a simple linear regression model of the form shown in Gleichung 2.

Tabelle ohne R: *Felis catus* does not have nine lives. Tabelle 3 shows the number of lives of three known cats.

Tabelle 3: List of known cats and their number of lives

Cat	Number of lives
Hello Kitty	1
Felix the cat	1
Garfield	1

You can suppress generation of a bibliography by including suppress-bibliography: true option in your document yaml.

Schluss und Ende.

Literaturverzeichnis

Knuth, D. E. „Literate Programming“. *The Computer Journal* 27, Nr. 2 (1. Februar 1984): 97–111. https://doi.org/10.1093/comjnl/27.2.97.