



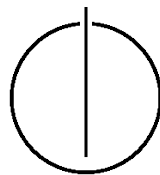
FAKULTÄT FÜR INFORMATIK

DER TECHNISCHEN UNIVERSITÄT MÜNCHEN

Abschlussarbeit in Informatik

**Effiziente statistische Methoden für
Datenbanksysteme**

Thomas Heyenbrock





FAKULTÄT FÜR INFORMATIK

DER TECHNISCHEN UNIVERSITÄT MÜNCHEN

Abschlussarbeit in Informatik

Effiziente statistische Methoden für Datenbanksysteme

Efficient statistical methods for database systems

Autor: Thomas Heyenbrock
Supervisor: Prof. Alfons Kemper
Advisor: Maximilian E. Schüle, M.Sc.
Datum: 15.01.2017



Ich versichere, dass ich diese Diplomarbeit selbständig verfasst und nur die angegebenen Quellen und Hilfsmittel verwendet habe.

München, den 1. November 2017

Thomas Heyenbrock

Abstract

An abstracts abstracts the thesis!

Contents

Abstract	vii
Outline of the Thesis	xi
1. Einführung und typische statistische Problemstellungen	1
1.1. Latex Introduction	1
2. Grundlagen statistischer Methoden	3
2.1. Lineare Regression	3
3. Anwendung statistischer Methoden	5
3.1. Latex Introduction	5
4. Statistische Methoden in Datenbanken	7
4.1. Latex Introduction	7
5. Erweiterungspotenzial in Datenbanksystemen	9
5.1. Latex Introduction	9
6. Fazit	11
6.1. Latex Introduction	11
Appendix	15
A. Detailed Descriptions	15
Bibliography	17

Outline of the Thesis

Teil I: Introduction and Theory

CHAPTER 1: INTRODUCTION

This chapter presents an overview of the thesis and its purpose. Furthermore, it will discuss the sense of life in a very general approach.

CHAPTER 2: THEORY

No thesis without theory.

Teil II: The Real Work

CHAPTER 3: OVERVIEW

This chapter presents the requirements for the process.

1. Einführung und typische statistische Problemstellungen

Here starts the thesis with an introduction. Please use nice latex and bibtex entries [1]. Do not spend time on formating your thesis, but on its content.

1.1. Latex Introduction

There is no need for a latex introduction since there is plenty of literature out there.

2. Grundlagen statistischer Methoden

Bei der Regressionsanalyse geht es allgemein darum, das Verhalten einer Größe Y in Abhängigkeit einer oder mehrerer anderer Größen X_1, X_2, \dots, X_n zu prognostizieren. Die Größe Y wird abhängig genannt, die Größen X_i nennt man unabhängig. Die Prognose entspricht einem funktionalen Zusammenhang der folgenden Art:

$$Y = f(X_1, \dots, X_n)$$

Als Grundlage für den Findungsprozess dieser Funktion f besitzt man eine Menge von N Datenpunkten:

$$(x_{1,j}, x_{2,j}, \dots, x_{n,j}, y_j) \quad \text{für } j = 1, \dots, N$$

Im Allgemeinen ist es nicht möglich eine einfache Funktion f zu finden, die alle Datenpunkte exakt prognostiziert. Stattdessen über- und unterschätzt die Funktion f die realen Datenpunkte. Die konkrete Abweichung für den j -ten Datenpunkt wird in der Statistik als Fehler bezeichnet:

$$e_j = y_j - f(x_{1,j}, \dots, x_{n,j})$$

2.1. Lineare Regression

Bei der linearen Regression geht man von einem linearen Zusammenhang zwischen der abhängigen und den unabhängigen Variablen aus. Die Funktion f ist also von folgender Form:

$$f(x_1, \dots, x_n) = \beta_0 + \sum_{i=1}^n \beta_i \cdot x_i \quad \text{mit } \beta_i \in \mathbb{R}$$

Man möchte diejenigen Parameter β_i bestimmen, so dass die entstehende Funktion die vorgegebenen Datenpunkte möglichst gut annähert. Das Maß für die Qualität einer Funktion f definiert durch die Parameter β_0, \dots, β_n ist die Summe der quadrierten Fehlerterme:

$$E(\beta_0, \dots, \beta_n) = \sum_{j=1}^N e_j^2 = \sum_{j=1}^N \left(y_j - \beta_0 - \sum_{i=1}^n \beta_i \cdot x_{i,j} \right)^2$$

Ziel ist es nun die Parameter so zu wählen, dass die Funktion E minimiert wird. Dieses Vorgehen ist als Methode der kleinsten Quadrate bekannt.

3. Anwendung statistischer Methoden

Here starts the thesis with an introduction. Please use nice latex and bibtex entries [1]. Do not spend time on formating your thesis, but on its content.

3.1. Latex Introduction

There is no need for a latex introduction since there is plenty of literature out there.

4. Statistische Methoden in Datenbanken

Here starts the thesis with an introduction. Please use nice latex and bibtex entries [1]. Do not spend time on formating your thesis, but on its content.

4.1. Latex Introduction

There is no need for a latex introduction since there is plenty of literature out there.

5. Erweiterungspotenzial in Datenbanksystemen

Here starts the thesis with an introduction. Please use nice latex and bibtex entries [1]. Do not spend time on formating your thesis, but on its content.

5.1. Latex Introduction

There is no need for a latex introduction since there is plenty of literature out there.

6. Fazit

Here starts the thesis with an introduction. Please use nice latex and bibtex entries [1]. Do not spend time on formating your thesis, but on its content.

6.1. Latex Introduction

There is no need for a latex introduction since there is plenty of literature out there.

Appendix

A. Detailed Descriptions

Here come the details that are not supposed to be in the regular text.

Bibliography

- [1] Leslie Lamport. *LaTeX : A Documentation Preparation System User's Guide and Reference Manual*. Addison-Wesley Professional, 1994.