Methoden zur Analyse von Gebäudebeständen, Alterungsmodelle und Prognose von Stoffströmen mit Anwendung auf den Gebäudebestand der Stadt Ettlingen[1]

Guido Bader
Institut für
Mathematische Stochastik
Universität Karlsruhe

Bärbel Schwaiger
Institut für
Industrielle Bauproduktion
Universität Karlsruhe

Claudio Ferrara
Institut für
Industrielle Bauproduktion
Universität Karlsruhe

Arbeitsbericht in der Version vom 15. August 2001

[1]Entstanden aus dem DFG-Projekt: „Validierung eines integrierten dynamischen Modells des Deutschen Gebäudebestandes“, KO 1488/4.

Inhaltsverzeichnis

Einleitung **2**

1 Zielsetzung und Vorgehen **5**
1.1 Zielsetzung 5
1.2 Vorgehen 6

2 Datenerhebung und Stichprobentheorie **9**
2.1 Beschreibung der Datenquellen 10
2.1.1 Statistisches Bundesamt - Statistisches Landesamt 11
2.1.2 Liegenschaftskataster 12
2.1.3 Bebauungsplan 16
2.1.4 Feuerversicherungsakten 16
2.1.5 Bauanträge 17
2.1.6 Ettlingen-spezifische Daten 17
2.2 Prinzipien der Datenerhebung - Stichproben 18
2.2.1 Das ALB als Grundlage für Stichproben 18
2.2.2 Nutzungsklassen 20
2.2.3 Stichproben 22
2.3 Praktisches Vorgehen bei der Datenerhebung 26
2.3.1 Übernahme von Daten aus ALB und ALK 26
2.3.2 Daten der Feuerversicherung 27
2.3.3 Bebauungsplan 28
2.3.4 Bauanträge 28
2.3.5 Probleme bei der Datenerhebung 29
2.4 Datenorganisation und Datenbank 30

3 Analyse des Gebäudebestandes **34**
3.1 Bestandsanalyse der Wohngebäude 34
3.1.1 Gebäudezahl in den Klassen K_1, K_2, K_3 35
3.1.2 Analyse des Bruttorauminhalts 38
3.1.3 Vergleich der Ergebnisse mit den Berechnungen aus Abschnitt 2.2.3 43

3.1.4 Der prozentuale Anteil der Klassen $K_1 - K_3$ am BRI von Wohngebäuden . . . 45
3.1.5 Die Verteilung des BRIes von Einfamilienhäusern . . . 47
3.1.6 Altersklassen von Einfamilienhäusern . . . 50
3.1.7 Die BRIe in den Altersklassen der Einfamilienhäuser . . . 51
3.2 Bestandsanalyse der nichtlandwirtschaftlichen Betriebsgebäude . . . 52
3.2.1 Analyse des Bruttorauminhaltes . . . 53
3.2.2 Altersklassen von nichtlandwirtschaftlichen Betriebsgebäuden 55
3.2.3 Die BRIe in den Altersklassen der nichtlandwirtschaftlichen Betriebsgebäude . . . 57
3.3 Die Alters-Nutzungs-Matrix . . . 57
3.3.1 Die Altersklassen der Nutzungsklassen K_2, K_4 und K_6 . . . 58
3.3.2 Die Matrix . . . 59

4 Dynamik von Gebäuden 61

5 Prognosen zu Bestand und Materialfluss 62

A Mathematischer Anhang 63
A.1 Stichproben . . . 63
A.1.1 Mittelwertschätzung eines Merkmals . . . 63
A.1.2 Verhältnisschätzung . . . 68
A.2 Die logarithmische Normalverteilung . . . 70

B Zuordnung von ALB-Nutzungsklassen 72

C Die Variablen von Kapitel 3 74

Literaturverzeichnis 76

Einleitung

Eine Vielzahl von Untersuchungen über den deutschen Gebäudebestand und die Entwicklung der Bautätigkeit zeigt, dass der Nutzung, Erneuerung und dem Umbau des Gebäudebestandes künftig eine immer größere Bedeutung zukommt. Ein vom Antragsteller im Rahmen des von der DFG geförderten Projektes „Dynamisches Gebäudebestands-Modell zu Energiehaushalt, Stoffflüssen und Finanzbedarf für Investitionen und Betrieb“ unter der Nummer RI 816/1-1 und RI 816/1-2, vgl. Barth und Schwaiger (1998), entwickelter Modellansatz erlaubt mit Hilfe von berechneten Szenarien Abschätzungen über ökonomische und ökologische Auswirkungen von Bewirtschaftungsstrategien, gesetzlichen Regelungen oder dem Einsatz neuer Technologien.
Der Modellansatz ermöglicht Berechnungen von künftigen Stoffströmen, Ressourcenverbrauch, Zusammensetzung des Bauschutts, Energieverbrauch und Emissionen sowie dadurch entstehenden Kosten. Die neutrale Modellgrundlage ist nicht durch eine spezifische Sicht auf den Bestand definiert und kann daher als Verbindung verschiedener Untersuchungsansätze dienen.
Wesentliche Faktoren für die Zuverlässigkeit der Berechnungen sind einerseits die Güte des Modells andererseits aber auch die Qualität der in das Modell eingehenden Daten über die Struktur, den Aufbau und die Dynamik des zu untersuchenden Bestandes. Das entwickelte Modell ist für beliebige Gebäudebestände einsetzbar - dies kann der deutsche Gebäudebestand sein, aber auch der einzelner Bundesländer oder Städte.

In diesem Forschungsprojekt steht der Thematik der Datengrundlage im Mittelpunkt. Erst hinreichend genaue Daten ermöglichen zuverlässige Berechnungen mit dem oben beschriebenen Modell. Im Idealfall wären

- die Struktur (z.B. Nutzung, Alter, Größe),
- der Aufbau nach Konstruktionsart und Baustoffen und
- der Gebäudezustand sowie das Alterungsverhalten

des zu untersuchenden Bestandes vollständig bekannt. Eine solche „totale Information“ wird man für größere Bestände jedoch nie erhalten, da es auf Grund der Datenmenge und der fehlenden Digitalisierung unmöglich ist, all diese Daten zu

erheben. Darüber hinaus können dynamische, zukünftige Prozesse wie das Alterungsverhalten nicht bestimmt sondern nur prognostiziert werden. Folglich muss eine Methodik entwickelt werden, die es erlaubt Kennzahlen zur Struktur eines großen Bestandes sowie dessen Dynamik möglichst einfach und mit hinreichender Präzision zu ermitteln.

Kapitel 1

Zielsetzung und Vorgehen

1.1 Zielsetzung

Ziel des Projektes ist die Entwicklung einer allgemein anwendbaren Methodik, um möglichst umfangreiche und präzise Kenntnisse über einen Gebäudebestand zu erlangen. Dies bezieht sich einerseits auf seine aktuelle Struktur hinsichtlich Gebäudenutzung, Alter, Konstruktionsart und Größe, sowie andererseits auf die Dynamik, d.h. das Veränderungsverhalten bezüglich dieser Merkmale. Untersuchungen über die Dynamik sind wichtig zur Verbesserung der Eingangsparameter für die „Alterungsmodelle“, mit deren Hilfe Szenarien berechnet werden, die die Entwicklung des Bestandes und damit seinen Zustand abbilden.
Darüber hinaus soll eine Methodik entwickelt werden, Daten unterschiedlicher Qualität, Herkunft und Struktur miteinander in einer Datenbank zu verknüpfen, um zu besseren Grundlagenkenntnissen zu gelangen.

Stichpunktartig lassen sich die Ziele wie folgt formulieren:

1. Analyse von verfügbaren Datenquellen in Bezug auf die Verwendung für die quantitative Bestandsanalyse.
2. Aufbau einer relationalen Datenbank zur Verknüpfung unterschiedlicher Datenquellen.
3. Entwicklung einer grundlegenden Methode für eine zuverlässige Analyse von Beständen hinsichtlich ihrer Struktur und Größe.
4. Entwicklung einer grundlegenden Methode für eine zuverlässige Analyse der Dynamik eines Gebäudebestandes.
5. Entwicklung von Prognosetechniken für die Bestandsentwicklung.
6. Beispielhafte Veranschaulichung und Validierung der Methodik anhand des Gebäudebestandes der Stadt Ettlingen.

1.2 Vorgehen

Die Durchführung der ursprünglichen Idee, ein validiertes Gebäudebestandsmodell in unterschiedlichen Schritten und Genauigkeitsstufen anhand verschiedener Datenquellen und verschiedener Teilbestände zu erarbeiten, war auf Grund der schlechten Datengrundlage nicht zweckmäßig. So wurde der geplante erste Schritt, anhand von Daten des statistischen Bundesamtes einen „großen“ Bestand (z.B. den gesamtdeutschen Gebäudebestand) in seiner Struktur zu untersuchen, verworfen, da es bereits unmöglich war, die vor dem Krieg erstellten Gebäude auch nur annähernd in ihrem Umfang zu ermitteln. Auch auf Landesebene, wo zum Teil umfangreichere Daten vorhanden sind als auf Bundesebene, musste der naheliegende Ansatz mittels Daten der statistischen Landesämter Bestandsanalysen durchzuführen verworfen werden, da auch diese Daten nur unzureichend sind. Andere Bestände, über die umfangreicheres Datenmaterial verfügbar ist, wie z.B. größere Bestände einer Firma oder einer Wohnungsbaugesellschaft, sind in ihrer Struktur nicht repräsentativ und erlauben keinerlei Rückschlüsse auf regionale (Städte, Landkreise) oder überregionale (Bundesländer, Deutschland) Bestände.

Bereits vorhandene Daten waren damit für statistische Bestandsanalysen und zuverlässige Hochrechnungen unbrauchbar, so dass im Projekt neu mit der gezielten und geplanten Erhebung repräsentativer Daten begonnen werden musste. Zum neuen Ziel wurde dabei nicht die explizite Analyse großer Bestände (wie den gesamtdeutschen), sondern die Erarbeitung einer generellen Methodik der statistisch abgesicherten Bestandsanalyse, die anhand eines Beispiels erläutert, durchgeführt und somit validiert werden sollte. Besonderer Wert wurde dabei auf die Generalität des Vorgehens gelegt, so dass eine Übertragbarkeit der Ansätze jederzeit auch auf andere, größere Bestände möglich ist.

Es galt zunächst einen „Beispiel-Bestand“ zu finden, der möglichst viele gebäudetypologischen Phänomene aufweist und der mit den bereitstehenden Mitteln gut zu erfassen war. Es war naheliegend, auf den räumlich abgegrenzten Bestand einer Stadt zurückzugreifen, da hier sowohl eine gemischte Altersstruktur wie auch eine gemischte Nutzungsstruktur in einer breiten Vielfalt zu erwarten ist. In einer Stadt lässt sich der vorhandene Gebäudebestand in seiner Gesamtheit bestimmen, und seine Absolutzahl (Anzahl der Gebäude) - nicht aber sein Alter, seine Größe oder seine materielle Struktur - kann als vollständig bekannt vorrausgesetzt werden. Dies ist vor allem für die statistischen Auswertungen eine wichtige Voraussetzung. Anhand des Bestandes einer Stadt lassen sich alle grundsätzlichen Untersuchungen durchführen, die dann analog für andere Bestände übertragbar sind.

Mit dem Ziel, dass einige der Ergebnisse Rückschlüsse auf die Bestandsstruktur

und Dynamik der Bundesrepublik zulassen, wurde eine möglichst „durchschnittlich“ Stadt ausgewählt. Solche Rückschlüsse sind jedoch aus statistischer Sicht in keiner Weise zu belegen, dennoch können die für den Beispiel-Bestand berechneten Zahlen den Experten einige Anhaltspunkte für die Struktur des gesamtdeutschen Gebäudebestandes liefern. Bei der Auswahl der „Beispielstadt“ wurden daher folgende Kriterien zu Grunde gelegt:

- Nutzungsvielfalt (keine reiner Wohnstandort),
- Vorhandensein einer historischen Struktur,
- Nähe zu Karlsruhe (um den finanziellen und zeitlichen Aufwand für die Datenermittlung gering zu halten),
- Im Rahmen der vorzufindenden Möglichkeiten guten Zugang zu den Datenquellen.

Nach einigen Voruntersuchungen wurde in der Stadt Ettlingen ein kooperationsbereiter Partner gefunden. Ohne die Unterstützung von der Stadtverwaltung wäre dieses Projekt nicht möglich gewesen. Da auch schon zu einem frühen Zeitpunkt klar war, dass der Datenbestand der ehemaligen Landesfeuerversicherung von Interesse sein würde, lag ein weiterer Vorteil darin, dass die entsprechenden Akten in Karlsruhe selbst archiviert wurden und Kontakte aus vorherigen Projekten bereits vorhanden waren.

Bei der Stadt Ettlingen handelt sich um eine Kleinstadt in Baden-Württemberg mit ca. 25.000 Einwohnern, die sowohl hinsichtlich der Gebäudealtersstruktur eine gute Mischung bietet, wie auch in Bezug auf die Nutzungsstruktur. Es ist kein reiner Wohnstandort, sondern es sind sowohl alte wie auch neue Gewerbezweige angesiedelt.

Im folgenden werden kurz die einzelnen Kapitel des vorliegenden Arbeitsberichtes beschrieben.

1. In Kapitel 2 werden zunächst alle verfügbaren Datenquellen gesichtet und auf ihren Aussagegehalt hin überprüft. Danach wird mittels des mathematischen Instrumentariums der Stichprobentheorie festgelegt, welche Daten in welchem Umfang sinnvollerweise erhoben werden. Hierbei spielen vor allem Katasterunterlagen eine wesentliche Rolle, die in früheren Projekten noch nicht für Bestandsuntersuchungen herangezogen wurden. Zuletzt werden die erhobenen Daten in einer Datenbank strukturiert.
2. Nach der Erhebung umfangreichen, repräsentativen Datenmaterials folgt in Kapitel 3 die Analyse des aktuellen Gebäudebestandes mittels statistischer Methoden. Dies erfolgt vor allem in Hinblick auf die Gebäudestruktur der

Stadt, d.h. es wird untersucht, welche Gebäude (Nutzungen) in welcher Anzahl bzw. Größe vorhanden sind und wie alt sie sind.

3. Kapitel 4 ist noch im entstehen.

4. Kapitel 5 ist noch im entstehen.

Kapitel 2

Datenerhebung und Stichprobentheorie

Das Wissen über den Deutschen Gebäudebestand ist, wie in vorangegangenen Studien bereits festgestellt wurde, äußerst gering. Die statistischen Bundes- und Landesämter liefern zwar Daten über die Bautätigkeit, aber eine Aussage über die Anzahl oder Größe der vorhandenen Gebäude ist nicht möglich. Lediglich der Bestand an Wohnungen ist einigermaßen zuverlässig bekannt, wobei Aussagen über die Wohngebäude schon dürftiger sind. Über den Bestand der Nichtwohngebäude, die vermutlich 50 % des Gesamtbestandes in Deutschland ausmachen, gibt es für die Nachkriegsgebäude nur geringfügiges und für die Gebäude, die vor dem Krieg errichtet wurden gar kein Datenmaterial bei den statistischen Ämtern. So enthalten deren Statistiken z.B. keinerlei Informationen über den nicht unerheblichen Bestand der vor 1950 errichteten Nichtwohngebäude.
Neben den statistischen Ämtern gibt es andere Datenquellen (meist nur für kleinere Gebäudebestände, z.B. von Städten, Firmen, Wohnungsbaugesellschaften, usw.), in denen sich sehr heterogenes Datenmaterial findet, das oft nur wenig Informationen liefert. So beinhalten diese Datenquellen z.B. entweder gute Angaben zur Größe von Gebäuden und keine zur Alterstruktur, oder es fehlen Angaben über die Gebäudenutzung etc..
Im Rahmen der klassischen Gebäudebewirtschaftung und hinsichtlich eines Facility Managements werden von einzelnen Besitzern größerer Gebäudebestände wie Firmen, Wohnungsbaugesellschaften, etc. viele Daten von einzelnen Gebäuden erhoben und verwaltet, die sich aber überwiegend auf den Gebäudebetrieb und nicht auf Gebäudestruktur und Zustand beziehen. Auch hier sind die Datenstruktur und der Umfang der verwalteten Daten pro Gebäude sehr unterschiedlich, auf einen speziellen Zweck zugeschnitten und liefern somit kaum globale Angaben über den jeweiligen Gesamtbestand.

Für dieses Projekt ist es wichtig, Datenquellen zu erschliessen, die möglichst um-

fangreiche und gut zugängliche Informationen über Gebäudebestände liefern. Dieses widmet sich der Problematik der Erhebung von geeigneten Daten zur Analyse eines beliebigen Gebäudebestandes. Die Methodik wird anhand der Stadt Ettlingen beschrieben, ist aber auf beliebige andere Bestände übertragbar.

Im ersten Abschnitt werden für das Projekt relevante Datenquellen allgemein analysiert. Wesentlich sind dabei Informationsgehalt, Datenstruktur und Nutzungsmöglichkeit für dieses Projekt.

Die zugänglichen Datenquellen sind oft sehr umfangreich und liegen in der Regel nicht digitalisiert vor, was eine vollständige Auswertung unmöglich macht. Um Aussagen über die Alters- oder Größenstruktur von Gebäuden eines Bestandes zu erhalten, werden folglich nur Teile der Datenquellen ausgewertet, um damit eine „Hochrechnung“ durchzuführen. Dies soll auf mathematisch fundierte Weise geschehen, wozu bereits die Bestimmung eines „geeigneten“ Stichprobenumfangs vor der Datenerhebung gehört. Diese Untersuchungen sind Gegenstand des zweiten Abschnitts.

Im dritten Abschnitt werden die tatsächlich augewerteten Datenquellen detailliert bezüglich der erhobenen Informationen beschrieben, und im vierten Abschnitt wird kurz auf die EDV-Organisation der erhobenen Daten eingegangen.

2.1 Beschreibung der Datenquellen

Zunächst soll ein Überblick über vorhandene Datenquellen gegeben werden. Wesentliche Kriterien für deren Nutzbarkeit sind Inhalt, die Art der Archivierung, die Vollständigkeit, die Verfügbarkeit sowie der Aufwand einer gezielten Datenerhebung. Ebenfalls ist die Möglichkeit der Ergänzung und Verknüpfung der Datenquellen untereinander zu prüfen. Untersuchungsgegenstand bildet die Stadt Ettlingen zunächst mit Ortsteilen; die detailliertere Datenerhebung bezieht sich dann nur noch auf die Kernstadt. Da die entwickelte Methodik sowohl auf andere Städte als auch andere beliebig definierte Bestände übertragbar sein soll, werden zunächst nur die Daten berücksichtigt, die allgemein verfügbar sind und nicht nur speziell für die Stadt Ettlingen vorliegen, vgl. Abschnitt 2.1.6. In einer zweiten Stufe werden diese Daten durch spezifische Informationen ergänzt, um genauere Aussagen über den Gebäudebestand Ettlingens zu erhalten.

Im Folgenden werden Datenquellen allgemein hinsichtlich ihres Inhalts und ihrer Struktur beschrieben. Welche Daten im Detail ausgewertet werden, wird im Abschnitt 2.3 beschrieben.

2.1.1 Statistisches Bundesamt - Statistisches Landesamt

Bei den statistischen Ämtern werden die für unsere Zwecke relevanten Daten in dem Bereich „Bautätigkeit und Wohnungen“ erfasst. Auf Bundesebene werden die Erfassungen in folgende Bereiche gegliedert:

- Baugenehmigungen,
- Baufertigstellungen,
- Bauüberhang,
- Bewilligungen im sozialen Wohnungsbau und
- Bestand an Wohngebäuden und Wohnungen.

In der Bautätigkeitsstatistik werden Baugenehmigungen und Baufertigstellungen in unterschiedlichen Kategorien geführt, nach Gebäudeart und Bauherren, nach Art der Konstruktion und nach Art der Beheizung. Weiterhin gibt es eine Unterteilung nach städtebaulichen Festsetzungen und im Wohnungsbau wird noch nach Gebäudeart und Haustyp unterschieden. Diese Daten sind meist in Zeitreihen und auf jeden Fall jährlich zu erhalten. Sie erlauben aber nur den Zuwachs zum Bestand zu ermitteln und lassen keinerlei Rückschlüsse auf den vorhandenen Bestand zu. Will man den Bestand durch Aufaddieren der Zeitreihen aufbauen, so stößt man auf folgende Probleme:

- es gibt keinen zuverlässigen Ausgangspunkt,
- ältere Statistiken sind zum Teil anders gegliedert bzw. nicht so detailliert,
- Statistiken über Gebäudeabgänge sind nur in geringem Umfang vorhanden.

Maßnahmen an bestehenden Gebäuden werden bisweilen auch in der Bautätigkeitsstatistik erfasst. Hierbei handelt es sich jedoch nur um genehmigungspflichtige Maßnahmen. Reine Instandhaltungsmaßnahmen werden satistisch nicht erfasst.

Lediglich bei den Wohngebäuden wird eine Bestandsstatistik geführt. Diese ergibt sich durch eine Fortschreibung mittels der Bautätigkeitsstatistik auf Basis der „Gebäude und Wohnungszählung“ vom 25.5.1987. Für die Neuen Länder und Berlin-Ost basiert sie auf der „Wohnraum- und Gebäudezählung“ vom 31.12.1981. Hierbei werden zwar Anzahl der Wohnungen in den Gebäuden und auf Länderebene verschiedene Ausstattungsstandards berücksichtigt, es werden jedoch keine Angaben zu der Altersstruktur des Baubestandes gemacht.
Im Rahmen dieses Projektes wurde beim Statistischen Landesamt Baden-Württemberg angefragt, inwieweit noch Daten über den Bestand der Gebäude, die

vor 1945 gebaut wurden, existieren. Es stellte sich jedoch heraus, dass nur noch sehr wenige Informationen vorhanden sind, die in keinster Weise eine zuverlässige Grundlage für die Validierung des Gesamtbestandes liefern können. Daraufhin wurde von einer weiteren Auswertung auf Landesebene abgesehen.

Weitere Statistiken, die im Rahmen der Bauwirtschaft erhoben werden, wurden zunächst nicht genauer untersucht. Generell besteht hier jedoch das Problem, dass sie meist die gesamte Bauwirtschaft (Hoch- und Tiefbau) umfassen und keine seperate Betrachtung mehr zulassen, oder bei der Erhebung über Betriebe, die in der Bauwirtschaft tätig sind, werden die kleinen Betriebe (unter 20 Beschäftigte), die vor allem in den Ausbaugewerken häufig vorkommen, nicht miterfasst.

2.1.2 Liegenschaftskataster

Die Arbeitsgemeinschaft der Vermessungsverwaltungen der Bundesrepublik Deutschland (AdV) definiert das Liegenschaftskataster (LK) als ein „ ... von Katasterbehörden geführtes öffentliches Register, in dem die Liegenschaften eines Landes (Flurstücke und Gebäude) dargestellt und beschrieben sind“.

Das Liegenschaftskataster liegt in der Regel auf Gemeindeebene vor und hat zwei grundlegende Zweckbestimmungen:

1. Als amtliches Verzeichnis der Grundstücke für den Eigentumsnachweis im Grundbuch wird die Lage des Grundeigentums in Form von Flurstücken dargestellt und beschrieben. Bei Bedarf können die Grenzen der Flurstücke örtlich mit rechtlicher Wirkung aufgezeigt werden. Außerdem weist das Liegenschaftskataster die Ergebnisse der amtlichen Bodenschätzung nach (klassische Hauptfunktion).

2. Es bildet eine Basis für weitere Bereiche. So soll es den Anforderungen des Rechtsverkehrs, der Verwaltung und der Wirtschaft gerecht werden und insbesondere die Bedürfnisse der Landesplanung, der Bauleitplanung, der Bodenordnung, der Ermittlung von Grundstückswerten sowie des Umwelt- und des Naturschutzes angemessen berücksichtigen. Landesinformationssysteme identifizieren die Grundstücke ebenfalls über das Liegenschaftskataster.

Das Liegenschaftskataster besteht aus

- dem Liegenschaftsbuch, dem beschreibenden textlichen Teil,

- der Liegenschaftskarte, dem darstellenden graphischen Teil.

Die beiden Bestandteile sind zunehmend in digitaler Form verfügbar.

2.1.2.1 Automatisiertes Liegenschaftsbuch - ALB

Ab 1973 wurde von der AdV die Entwicklung eines rechnergestützten Liegenschaftsbuches als erster Schritt zu einem automatisierten Liegenschaftskataster vorangetrieben. 1983 erfolgte die Freigabe des Systems, der die Einführung folgte. Heute liegt es bundesweit nahezu flächendeckend vor. Auch die neuen Bundesländer haben ihre bereits in elektronischer Form vorliegenden Liegenschaftsdaten (COLIDO) in das ALB-Format konvertiert. Da das ALB die Basisdaten für dieses Projektes liefert, werden Aufbau, Inhalt und Einheiten im Folgenden ausführlich beschrieben.

Die Bodenfläche des Landes ist politisch in Landkreise, kreisfreie Städte sowie Gemeinden und katastermäßig in Gemarkungen, Fluren sowie in Flurstücke geteilt. Die Gemarkung als übergeordnete Einheit ist eine katastertechnische Gebietseinheit, die eine zusammenhängende Gruppe von Flurstücken umfasst. Sie wird im Allgemeinen zur Nummerierung der Flurstücke in Fluren unterteilt, vgl. Abbildung 2.1.

Das Gemarkungskennzeichen identifiziert die Gemarkung bundesweit eindeutig über

- den Schlüssel für das Land und
- den Schlüssel für die Gemarkung.

Zur Gemarkung werden folgende Informationen geführt:

- der Gemarkungsname,
- die Gemarkungsfläche,
- die Anzahl der Flurstücke und
- die Zugehörigkeit zum Katasterbezirk, zum Amtsgericht und zur Gemeinde.

Die nächstkleinere Einheit wird als Flur bezeichnet und umfaßt eine Gruppe von Flurstücken, die in einem topographischen Zusammenhang stehen. Flurstücke (auch Parzellen genannt) sind die kleinste im Liegenschaftskataster aufgeführte Einheit. Das Flurstückskennzeichen identifiziert ein Flurstück bundesweit eindeutig über:

- den Schlüssel für das Land,
- den Schlüssel für die Gemarkung,
- die Flurnummer und

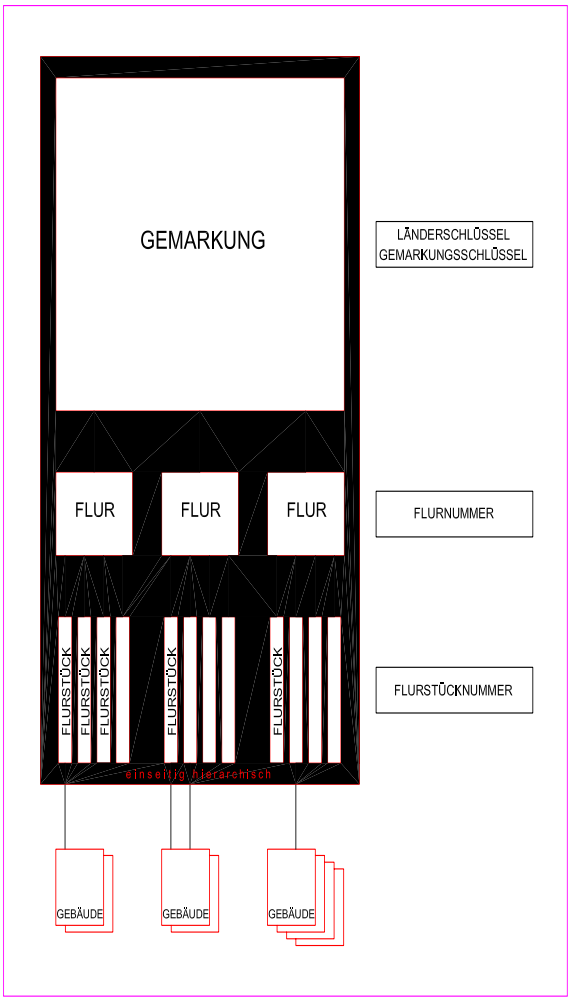


Abbildung 2.1: Aufbau des ALB

- die Flurstücksnummer.

Zum Flurstück werden i.A. die folgenden Informationen geführt:

- die Flurstücksfläche,
- die Nutzung,
- die Lagebezeichnung (z.B. Straße, Hausnummer),
- die Flurstückskoordinate (nur in Gebieten, in denen auch die ALK, vgl. Abschnitt 2.1.2.2, vorliegt),
- Hinweise auf die zugehörige Flurkarte und Vermessungsschrift,
- öffentlich-rechtliche Festlegungen (z.B. Bodenschätzungsergebnisse, Naturschutzgebiete),

- Hinweise auf Baulasten,
- Hinweise auf öffentlich-rechtliche Verfahren (z.B. Hinweise auf Bodenordnungsverfahren),
- das zuständige Finanzamt,
- die Entstehung, letzte Fortführung und Fortführungen des laufenden Jahres,
- das Vorgänger-/Nachfolgerflurstück und
- der Status (im Grundbuch gebucht, noch nicht im Grundbuch gebucht, historisch).

Da das ALB meistens noch mit dem Grundbuch verknüpft ist, werden auch Informationen über Eigentümer (Name, Adresse) und die sogenannte Buchungsart geführt. Dies wird hier nicht näher erläutert, da die Eigentümerinformationen dem Datenschutz unterliegen und für die Auswertung in diesem Projekt nicht relevant sind.

2.1.2.2 Automatisierte Liegenschaftskarte - ALK

Seit 1979 wird die Automatisierte Liegenschaftskarte (ALK) von der AdV als herstellerneutrales Programmsystem zur Führung des Zahlen- und Kartennachweises des Liegenschaftskatasters entwickelt und in der ALK-Verfahrensdokumentation beschrieben.

Die ALK dient als Grundlage für „parzellenscharfe“ Planungen und Bestandsnachweise in Wirtschaft und Verwaltung. Die Attribute „Flurstücksnummer“, „Flurnummer“ und „Gemarkung“ als Flurstückskennzeichen stellen den Bezug zum Liegenschaftsbuch her. In der ALK werden die Objekte der Liegenschaftskarte geführt (z.B. Flurstücke, Gebäude, Gewässer, etc.). Jede Grundrissinformation der Liegenschaftskarte wird Bestandteil eines Objektes. Die Objekte werden inhaltlich klassifiziert und entsprechend definierten Ebenen (Layern, Folien) zugeordnet. Sie sind grundsätzlich mit Geometrie und Fachbedeutung dokumentiert. Die Einteilung der Objektbereiche, -klassen und -arten erfolgt nach der Verwaltungsvorschrift der AdV oder dem im Amtlichen Topographischen-Kartographischen Informationssystem (ATKIS) beschrieben Objektartenkatalog (ATKIS-OK). Diese sind als zentrale Verschlüsselungskataloge verbindlich. Die im neueren Katalog ATKIS-OK aufgeführten Verschlüsselungen berücksichtigen die im Nutzungsartenverzeichnis der AdV definierten flächenbezogenen Nutzungsarten, das Verzeichnis der Gebäudenutzungen im Liegenschaftskataster sowie die Darstellungsinhalte der Liegenschaftskarte.
Die Verschlüsselungssystematik enthält Freiräume für länderspezifische Anwendungen. Nicht alle Ebenen- und Objektschlüsselkombinationen werden in den

länderspezifischen Katalogen übernommen. Die Flurstücksnummer als Attribut dient in vielen Anwendungsfällen als Fremdschlüssel zu Sachdatenbeständen, jedoch verfügen nur Objekte der Objektart „Flurstück“ über dieses Attribut, Gebäude sind nicht eindeutig identifizierbar. Mit Hilfe der ALK lassen sich unter anderem auch flächenbezogene Auswertungen hinsichtlich der Bodennutzung erstellen. Die Klassifizierung der Flächen innerhalb der ALK erfolgt gemäß der „Verwaltungsvorschrift für die Führung des Liegenschaftsbuches mit automatischer Datenverarbeitung“, Anlage 2. In Anlage 6 sind die Bezeichnungen für die Gebäudenutzung festgelegt.

2.1.3 Bebauungsplan

Der Bebauungsplan enthält als verbindlicher Bauleitplan die rechtsverbindlichen Festsetzungungen für die städtebauliche Ordnung und gilt jeweils für einen räumlich abgegrenzten Teilbereich der Stadt oder Gemeinde. Hier werden parzellenscharf die zulässigen Nutzungen der Bebauung festgelegt. Dabei wird gemäß § 17 BauNVO unterschieden zwischen Kleinsiedlungsgebiet, reines Wohngebiet, allgemeines Wohngebiet, besonderes Wohngebiet, Dorfgebiet, Mischgebiet, Kerngebiet, Gewerbegebiet, Industriegebiet und Sondergebiet. Den einzelnen Gebieten ist ein „Maß der Baulichen Nutzung“ zugeordnet, das mittels Grundflächenzahl (GRZ), Geschossflächenzahl (GFZ) und Baumassenzahl (BMZ) Rückschlüsse auf die Bebauungsdichte erlaubt. Weiterhin können im Bebauungsplan Festsetzungen zur Bauweise (offen, geschlossen), Anzahl der Vollgeschosse sowie Dachform und Ausrichtung der Gebäude festgesetzt werden. In Einzelfällen liegt auch eine Gestaltungssatzung vor, die Materialien und Farben vorschreibt.

2.1.4 Feuerversicherungsakten

Die Länderkammern der Feuerversicherungen hatten bis 1994 eine Monopolstellung, d.h. jedes Gebäude war bei den jeweiligen Landesversicherungsanstalten erfasst. Auf Grund der Liberalisierung des Versicherungsmarktes können mittlerweile aber auch andere Versicherer Brandschadensversicherungen anbieten, so dass nicht mehr der gesamte Gebäudebestand bei den Landesversicherungsanstalten geführt wird. Diese verfügen aber nach wie vor über den umfassensten Datenbestand, der im einzelnen weit detailliertere Informationen über Gebäude beinhaltet, als die bisher aufgeführten Datenquellen. So kann z.B. nur in den Akten der Feuerversicherungen das Baujahr von älteren Gebäuden festgestellt werden. Neben dem Baujahr werden Angaben zur Nutzung und das ungefähre Volumen der Gebäude erfasst. Kenntnisse über Ausstattungsstandards, Heizsysteme und Bauweise (Hauptkonstruktionsmaterial sowie Dachdeckung) sind nicht in jedem Fall vorhanden. Hinzu kommt natürlich der sogenannte Wiederbeschaffungswert, der für die Versicherung wesentlich ist.

Die Feuerversicherungsakten beinhalten auch Informtionen über die Veränderung von Gebäuden im Laufe der Zeit, da diese - soweit versicherungsrelevant und gemeldet - dokumentiert sind nach Maßnahme, Volumen und Veränderungsgrund. Da die Feuerversicherungsakten die Hauptdatenquelle über die Gebäudedaten darstellen, wird auf deren genauen Informationsgehalt in Abschnitt 2.3.2 detailliert eingegangen.

Die Erfassung und Datenorganisation ist von Länderkammer zu Länderkammer unterschiedlich. EDV-Unterlagen über die Gebäude selbst sind in der Regel nicht vorhanden und somit ist eine Datenauswertung äußerst mühsam. Die für Ettlingen zuständige Versicherungskammer führt neuere Daten zwar in ihrer EDV, für die alten Verträge wurden nachträglich aber nur wenige Daten in die EDV eingegeben. Vor allem das Baujahr des Gebäudes wird nicht mitgeführt, sondern nur das letzte Änderungsdatum des Vertrages. Damit waren diese Dateien für die Auswertung nicht zu verwenden. Die Informationen lagen ausschließlich in einer oder mehreren Akten pro Liegenschaft vor, die nach Adressen der Liegenschaften sortiert sind.

2.1.5 Bauanträge

Der Bauantrag als amtliches Dokument auf Stadt- bzw. Gemeindeebene ist heute die umfangreichste Zusammenstellung von Daten und Informationen, die über ein einzelnes Gebäude verfügbar ist. Er besteht aus einem Formularsatz, einer Statik sowie zugehörigen Bauplänen. Diese können Grundlage für die detaillierte Erfassung der Gebäude sein, da mit ihrer Hilfe eine Massenermittlung möglich ist sowie die dazugehörigen konstruktiven und materiellen Angaben zu ermitteln sind. Für eine Analyse des Gebäudebestandes sind vor allem die Angaben, die in den statistischen Erhebungsbögen zusammengestellt werden, von Interesse. Der Inhalt der Unterlagen hat sich im Laufe der Jahre immer wieder geringfügig geändert, meist auf Grund geänderter Baugesetzgebung. Statistische Erhebungsbögen gibt es erst seit den siebziger Jahren und liegen auch nicht in jedem Fall vor.
Die Anträge werden in Aktenform archiviert und sind nicht in EDV-Form vorhanden. Der Bestand an Bauantragsakten ist, vor allem für die Vorkriegsbestände, meist nicht so komplett vorhanden wie die Akten der Feuerversicherungen.

2.1.6 Ettlingen-spezifische Daten

Neben denen in den vorangegangenen Abschnitten vorgestellten Datenquellen, die Deutschlandweit zur Verfügung stehen (ALK und ALB für die neuen Bundesländer noch nicht flächendeckend vorhanden) wurden im weiteren einige Datenquellen untersucht, die sicherlich in vielen Städten und Gemeinden vorhanden sind, jedoch unterschiedlich detailliert und strukturiert. Da für diese Datenquellen keine

übergeordneten Gliederungen oder Konventionen bestehen (wie z.B. bei ALB oder Bauanträgen), können Inhalte und Detaillierungsgrad sehr unterschiedlich sein. Im einzelnen handelt es sich bei diesen Datenquellen um Planunterlagen zur Bestandsaufnahme für die innerstädtische Rahmenplanung (liegt oft in Verbindung mit einer Ortskernsanierung vor) sowie einer damit erarbeiteten Liste der zur Aufnahme in das Denkmalkataster vorgeschlagenen Gebäude. Weiterhin wurde der Datenbestand der städtischen Wohnungsbaugesellschaft sowie anderer Baugenossenschaften oder ähnlicher Einrichtungen über deren verwalteten Baubestand untersucht. Hierin sind Daten zu Gebäudealter, Nutzung und Nutzflächen (nach DIN277) sowie Daten über Gebäudeunterhalt und Erneuerung enthalten.

2.2 Prinzipien der Datenerhebung - Stichproben

Wie in der Einleitung zu diesem Kapitel bereits erwähnt, ist es unmöglich, alle in Abschnitt 2.1 aufgeführen Datenquellen komplett auszuwerten. Hauptgrund dafür ist, dass viele Daten nicht als digitalisierte Dateien, sondern nur als Akten in Papierform vorliegen und die Auswertung zu zeitaufwendig wäre. Folglich werden je nach Datenquelle nur einzelne, zufällig ausgewählte Gebäude des zu untersuchenden Bestandes analysiert. Um die Größe der Stichprobe bestimmen zu können, muss eine „Liste aller Gebäude“, die den kompletten Bestand umfasst, als Grundlage vorhanden sein. Ein solche Liste kann dem ALB entnommen werden.

2.2.1 Das ALB als Grundlage für Stichproben

Das Liegenschaftsbuch liegt im Fall der Stadt Ettlingen automatisiert, d.h. in EDV-Form vor und enthält alle Liegenschaften der Stadt, wobei eine Liegenschaft nicht gleichbedeutend mit einem Gebäude ist. Im ALB werden aber alle auf einem Flurstück befindlichen Gebäude aufgeführt. In einem ersten Arbeitsschritt muss folglich aus dem ALB eine *Liste aller Gebäude* gewonnen werden. Ein Gebäude wird als physikalisch zusammenhängendes Bauwerk verstanden, das einer überwiegenden Nutzung zugeordnet werden kann. Es wird hier die im ALB verwendete Definition und Nutzungsklassifizierung übernommen. Dies erleichtert auch die Erstellung der Liste für den Gesamtbestand. Gebäude im Sinne obiger Definition sind weder über die Flurstücke noch über die Adresse eindeutig identifizierbar. Insgesamt treten folgende Fälle auf:

- Auf einer Parzelle stehen mehrere Gebäude, vgl. Abbildung 2.2.

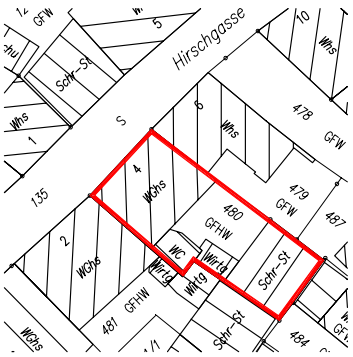


Abbildung 2.2: Parzelle mit mehreren Gebäuden

- Gebäude stehen teilweise auf mehreren Parzellen und werden bei beiden Parzellen aufgeführt, vgl. Abbildung 2.3. Dies wurde aber immer durch einen Zusatz „Teil auf Flurstück xyz“ kenntlich gemacht.

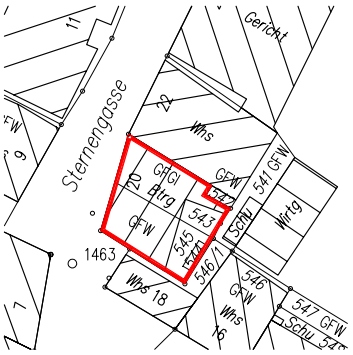


Abbildung 2.3: Gebäude auf mehreren Parzellen

- Gebäude liegen auf einem Flurstück und haben unterschiedliche Adressen, vgl. Abbildung 2.4.

- Gebäude liegen auf mehreren Flurstücken und haben die gleiche Adresse.

- Gebäude stehen auf einem eigenen Flurstück und haben keine Adresse.

Da weder Flurstück noch Adresse oder Eigentümer zu einer eindeutigen Gebäudeliste führen, werden mittels SQL aus der ALB-Datenbank die Gebäude direkt nach den einzelnen Nutzungen abgefragt. Dies ist nur über unsere Definition eines Gebäudes möglich. Damit sind aber immer noch Doppelnennungen möglich (z.B. durch die Führung auf verschiedenen Grundstücken). Auch ist die Konvention der auf einem Grundstück geführten Gebäude unterschiedlich. So kann

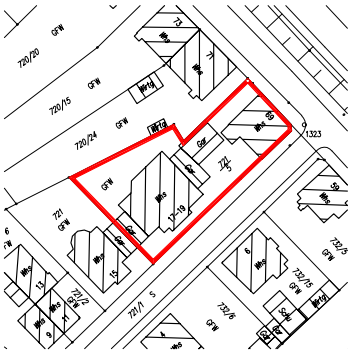


Abbildung 2.4: Gebäude auf einem Flurstück mit verschiedenen Adressen

z.B. eine Gebäudenutzungstyp drei mal als Einzelgebäude aufgeführt werden „Betriebsgebäude, Betriebsgebäude, Betriebsgebäude“ oder das Datenbankfeld enthält den Eintrag „3 Betriebsgebäude“. Somit lassen sich die jeweiligen Gebäude nicht durch eine einfache Zählung einer SQL Abfrage ermittlen.
Um vorhandene Doppelnennungen und sonstige Sonderfälle zu untersuchen wurden alle nicht eindeutigen Gebäude auf der ALK überpüft. Dieses Verfahren erscheint zunächst sehr aufwendig, lässt sich aber mit Hilfe eines Geoinformationssystems in vertretbarem Aufwand bewerkstelligen. Es war im ersten Schritt sehr wichtig, einen möglichst gut überprüften Bestand zu erhalten, um dann für das „schnellere“ Verfahren (reine SQL Abfrage aus dem ALB) die Abweichungen mit dem genau überprüften Bestand bestimmen zu können.

Es ist davon auszugehen, daß der im ALB geführte Bestand nicht exakt der Realität entspricht. Dies ist bedingt durch aktuelle Neuerungen, die noch nicht nachgetragen sind. Da der Anteil noch nicht erfasster Gebäude aber als sehr gering angesehen werden kann, wollen wir die dem ALB entnommene Gebäudeliste, die die Basis für dieses Projekt bildet, als vollständig ansehen und künftig immer vom Gesamtbestand Ettlingens sprechen.

2.2.2 Nutzungsklassen

Die aus dem ALB ermittelten Gebäude werden in Bezug auf ihre Nutzung in Klassen unterteilt, die die Basis für eine Stichprobennahme bilden. Dies ist in Hinblick auf spätere statistische Auswertungen wie auch für eine geeignete Stichprobenauswahl sinnvoll. Ziel der Klassenbildung ist es, möglichst homogene Klassen zu erhalten, so dass bereits wenige Vertreter der einzelnen Klassen Auskunft über bestimmte Merkmale der gesamten Klasse geben.

Die im Folgenden verwendeten Nutzungsklassen entsprechen denen des Stati-

stischen Bundesamtes, wobei die weitere Unterteilung der Wohngebäude auf gebäudetypischen und für spätere Berechnungen sinnvollen Überlegungen beruht. Folgende Klassen werden unterschieden:

1. Wohngebäude

   K_1: Einfamilienhäuser und kleine Mehrfamilienhäuser (bis zu drei Wohneinheiten)

   K_2: Mehrfamilienhäuser (ab vier Wohneinheiten bis zu sechs Vollgeschossen)

   K_3: Hochhäuser (ab 7 Vollgeschossen).

2. Nichtwohngebäude

   K_4: Büro und Verwaltungsgebäude

   K_5: Anstaltsgebäude

   K_6: Landwirtschaftliche Betriebsgebäude

   K_7: Nichtlandwirtschaftliche Betriebsgebäude

   K_8: Sonstige Nichtwohngebäude.

Die Zuordnung der Gebäude zu den einzelnen Klassen geschieht unter Zuhilfenahme der Nutzungsangaben im ALB. Die Nichtwohngebäude werden dabei den Klassen $K_4 - K_8$ nach dem offiziellen Signierschlüssel des Statistischen Bundesamtes zugeordnet. Was dies im Fall Ettlingens bedeutet, wird aus Anhang B ersichtlich. Bei Wohngebäuden ist aus den Angaben im ALB allerdings keine Zuteilung zu den Klassen $K_1 - K_3$ möglich. Genauer gesagt ist im ALB die folgende Information enthalten:

Seien

N_j - Anzahl der Gebäude in Klasse j, $j = 1, \ldots, 8$,

N - Gesamtzahl aller Gebäude, $N = \sum_{j=1}^{8} N_j$,

p_j - prozentualer Anteil der Klasse K_j am Gesamtbestand, $p_j = \frac{N_j}{N}$.

Dann können wir dem ALB die folgenden Daten entnehmen:

$N_4 - N_8$ bzw. $p_4 - p_8$, d.h. die Anzahl der Gebäude in den Klassen $K_4 - K_8$,

$N_{wohn} := \sum_{j=1}^{3} N_i$, d.h. die Anzahl der Wohngebäude.

Insbesondere erhalten wir keine Auskunft über die Größen $p_1 - p_3$.

2.2.3 Stichproben

Prinzipiell interessieren alle Daten, die aus den in Abschnitt 2.1 beschriebenen Datenquellen zu entnehmen sind. Mit diesen Daten können präzise Aussagen über Größe, Alter, Zustand, Materialien, usw. des aktuellen Gebäudebestandes der zu untersuchenden Gesamtheit getroffen werden. Ebenso enthalten die Daten Informationen über die Dynamik des Gebäudebestandes.
Gewinnt man die vollständigen Daten des Gesamtbestandes aus den in Abschnitt 2.1 beschriebenen Datenquellen, so spricht man von einer **Vollerhebung**, die aber sehr zeitaufwendig und teuer ist, da die Daten nur zum Teil digitalisiert vorliegen. Die eigentlich relevanten Informationen müssen aus Feuerversicherungs- bzw. Bauordnungsamtsakten „von Hand“ gewonnen werden. Aus diesem Grund ist eine Vollerhebung unmöglich. Es muss also eine zufällige, möglichst kleine Stichprobe gezogen werden, eine sogenannte **Teilerhebung**, die ihrerseits möglichst viel Information über den Gesamtbestand enthalten soll. Diese „Information“ wird durch die verschiedenen Größen wie Baujahr, Volumen, usw. der Datensätze charakterisiert, die auch als **Merkmale** bezeichnet werden.

Im Folgenden werden die im mathematischen Anhang A.1 zu Stichproben eingeführten Begriffe und Bezeichnungen verwendet, ohne dass sie hier explizit wiederholt werden.

2.2.3.1 Kriterien für Stichproben

Dem Datenbestand sind sehr viele Informationen zu einem einzelnen Gebäude zu entnehmen. Wollte man in Hinblick auf *all diese Merkmale* eine „repräsentative“ Stichprobe nehmen, so müsste man, um die Repräsentativität sicherzustellen, nahezu den gesamten Datenbestand analysieren, d.h. fast eine Vollerhebung durchführen. Man wird sich also auf wenige, wichtige Merkmale konzentrieren, bezüglich denen man die Stichprobe so groß wählt, dass Repräsentativität gewährleistet ist. Das heißt jedoch nicht, dass die Stichprobe über andere, hier nicht berücksichtigte Merkmale unzureichende Information liefert - es wird lediglich im Vorfeld nicht sichergestellt, dass sie das tut.

Das für dieses Projekt entscheidende Merkmal bei der Suche nach einem sinnvollen Stichprobenumfang ist der

Bruttorauminhalt (BRI) eines Gebäudes.

Es gilt nun zu klären, was man unter der Repräsentativität einer Stichprobe zu verstehen hat.

Fixiert man eine der Nutzungsklassen aus Abschnitt 2.2.2, so kennt man nach Abschnitt 2.2.1 die genaue Anzahl der Gebäude Ettlingens innerhalb dieser Klasse mit

Ausnahme von $K_1 - K_3$ (Wohngebäude). Im Folgenden werden aus den Gebäuden jeder Klasse eine möglichst kleine Anzahl (Stichprobe) zufällig ausgewählt und deren Merkmalsausprägungen aus den verschiedenen Akten ermittelt. Insbesondere wird der Bruttorauminhalt bestimmt. Diese zufällige Stichprobe soll aber so groß sein, dass schon im Vorfeld sichergestellt ist, dass man den *mittleren BRI* der Klasse anhand der gemessenen BRI der Stichprobe sehr gut und mit großer Sicherheit prognostizieren kann. Im Detail heißt das:
Man gibt sich eine Intervalllänge ℓ und ein Niveau $\alpha \in (0, 1)$ vor und bestimmt dann den Stichprobenumfang n so, dass anhand der Stichprobe ein Intervall angeben werden kann, das maximal die Länge ℓ hat und in dem der wahre mittlere BRI mit einer Wahrscheinlichkeit von mindestens $1 - \alpha$ liegt. Wie das genau geschieht, ist im mathematischen Anhang A.1 beschrieben.

Auf eine Besonderheit sei nochmals ausdrücklich hingewiesen. Anhand des ALB können die Klassen $K_1 - K_3$ nicht unterschieden werden. Die genaue Anzahl der z.B. zu Klasse K_1 gehörigen Gebäude ist also unbekannt und somit kann auch nicht eine bestimmt Anzahl dieser Gebäude ausgewählt werden. Folglich wird man die Wohngebäude zu einer Klasse $K_{wohn} = K_1 + K_2 + K_3$ zusammenfügen. Die Stichprobe soll jetzt aber nicht nur Auskunft über den mittleren BRI der Klasse K_{wohn} geben, sie soll vielmehr auch hinreichend genau Auskunft über die prozentualen Anteile der Klassen K_1, K_2, K_3 an K_{wohn} geben und auch über den mittleren BRI der einzelnen Klassen.

Um nicht zu große Stichproben nehmen zu müssen, ist es hilfreich im Vorfeld bereits einige Informationen über die BRI der einzelnen Klassen zu besitzen, vgl. Bemerkung A.1.3. Besonders wichtig ist eine grobe Vorstellung, in welcher Größenordnung sich der wahre mittlere BRI bewegt, um den Parameter ℓ sinnvoll festlegen zu können. Ebenso ist es wichtig etwas über die Streuung der BRI innerhalb der Klassen zu wissen. Dies ermöglicht eine nicht zu grobe Abschätzung für die Varianz der Merkmale innerhalb der Klasse, vgl. auch hierzu Bemerkung A.1.3.

2.2.3.2 Vorinformationen

Auf Grund älterer Untersuchungen stehen Daten zum BRI von verschiedenen Gebäuden, die unseren Nutzungsklassen zugeteilt werden können, zur Verfügung. Für diese Beispieldaten wurden der mittlere BRI sowie die Varianz der vorliegenden Daten innerhalb der Nutzungsklassen berechnet und in Tabelle 2.1 aufgeführt. Zur Klasse K_6 der landwirtschaftlichen Betriebsgebäude lagen keine Beispieldaten vor. Einige der 387 Gebäude der Klasse K_{wohn} konnten als Einfamilienhäuser identifiziert werden und wurden in der Spalte mit der Überschrift K_1 nochmals gesondert aufgeführt.

	K_{wohn}	K_1	K_4	K_5	K_7	K_8
Anzahl	387	27	99	38	197	639
Mittlerer BRI	5.183	1.327	18.297	19.818	12.118	12.064
Varianz in Mio.	75	0,28	322	214	363	459
Max. BRI	110.706	2.855	104.865	52.675	193.018	282.204
Min. BRI	344	762	412	1.051	340	384

Tabelle 2.1: Vorinformationen

2.2.3.3 Berechnungen

Hier werden kurz die Berechnungen angeführt, die zu den ausgewählten Stichprobenumfängen führen. Zu allen Details und zu den Bezeichnungen vergleiche Anhang A.1.

In einem ersten Schritt müssen die Intervalllänge ℓ des Konfidenzbereichs sowie die Varianzabschätzung σ_o^2 festgelegt werden. Ausgehend von den Beispieldaten in Tabelle 2.1, schienen die in Tabelle 2.2 angegebenen Festlegungen sinnvoll. Für die Klasse K_6, für die uns keine Beispieldaten vorlagen, wurden plausibel erscheinende Werte gewählt.

	K_{wohn}	K_4	K_5	K_6	K_7	K_8
ℓ	1.200	4.000	3.500	2.500	3.300	3.200
σ_o^2 in Mio.	75	322	214	150	363	459

Tabelle 2.2: Festlegung von ℓ und σ_o^2

Mittels dieser Werte liefert Formel A.4 aus Anhang A.1 wie in Beispiel A.1.4 die in Tabelle 2.3 aufgeführten Stichprobenumfänge. Die Anzahl der Gebäude pro Klasse wurden dabei aus dem ALB der Stadt Ettlingen ermittelt. Als Niveau ist $\alpha = 0.25$ zu Grunde gelegt. (Vgl. die Ausführungen im Anhang A.1, denen zufolge die aus der Stichprobe resultierenden Konfidenzintervalle ein deutlich höheres Niveau als $1 - \alpha$ besitzen, nämlich in etwa 95 %.)

	K_{wohn}	K_4	K_5	K_6	K_7	K_8
Gebäudebestand laut ALB	3084	198	8	697	1186	85
Stichprobenumfang $\tilde{n}_{l,\alpha}$	635	123	8	248	368	77

Tabelle 2.3: Stichprobenumfänge

Bei diesen Stichprobenumfängen kann somit eine „gute“ Schätzung des mittleren BRI in den einzelnen Klassen garantiert werden. Es bleibt jetzt noch zu analysieren, inwieweit eine Stichprobe von 635 Wohngebäuden auch gute Informationen zum prozentualen Anteil der Klassen $K_1 - K_3$ an K_{wohn} liefert. Ebenso bleibt zu klären, ob die mittleren BRI der Klassen $K_1 - K_3$ ebenfalls gut geschätzt werden können.

Um die Frage nach dem prozentualen Anteil zu beantworten, wird man analog zu Beispiel A.1.5 vorgehen. Bei gleichem Niveau $\alpha = 0.25$ und einer Intervalllänge von $\ell = 0.08$ ergibt Formel A.4 einen Stichprobenumfang von 520 Gebäuden. Es müssen also mindestens 520 Gebäude analysiert werden, um für den wahren prozentualen Anteil einer der Klassen $K_1 - K_3$ an K_{wohn} einen Bereich von höchstens 8% Länge angeben zu können, in dem der wahre Anteil mit einer Wahrscheinlichkeit von mindestens 75% liegt. Somit erhält man mit dem Stichprobenumfang von 635 Wohngebäuden auch eine gute Näherung für die prozentualen Anteile der einzelnen Klassen.
Die mittleren BRI innerhalb der Klassen $K_1 - K_3$ können natürlich nur hinreichend genau bestimmt werden, wenn die Stichprobe auch genügend Vertreter der jeweiligen Klasse liefert. Dies wird bei Klassen mit geringem prozentualen Anteil an K_{wohn} natürlich nicht der Fall sein. Insbesondere werden wir über die Klasse K_3 der Hochhäuser nur sehr unzureichende Informationen erhalten.
Anders verhält es sich z.B. mit der Klasse K_1 der Einfamilienhäuser. Hier ist mit einem wahren Anteil von mindestens 40% an K_{wohn} zu rechnen. Folglich wird eine Stichprobe vom Umfang 520 nach obigen Überlegungen und nach der Konstruktion des Konfidenzintervalls mindestens 32% Gebäude der Klasse K_1 enthalten (mit einer Wahrscheinlichkeit von mindestens 75%). Das bedeutet, dass in der Stichprobe vom Umfang 635 mit mindestens 203 Gebäuden der Klasse K_1 gerechnet werden kann.
Andererseits kann man sich fragen, wieviele Gebäude der Klasse K_1 untersucht werden müssten, um bei einem Niveau von $\alpha = 0.25$ für den mittleren BRI der Klasse K_1 ein Konfidenzintervall von maximal der Länge $\ell = 200$ zu erhalten. Da nur sehr wenige Beipieldaten zur Klasse K_1 vorhanden sind, wählen wir als Varianzschranke $\sigma_o^2 = 500.000$, also fast den doppelten Wert der Beispieldaten. Da der Mindeststichprobenumfang $\tilde{n}_{\ell,\alpha}$ aus Formel A.4 mit steigender Grundgesamtheit N monoton wächst, wird man hier für N einen beliebig großen Wert annehmen und sogar den Grenzwert von $\tilde{n}_{\ell,\alpha}$ für $N \to \infty$ betrachten. Dies liefert einen Mindeststichprobenumfang von 200 Gebäuden der Klasse K_1. Da man aus dieser Klasse nach obigen Überlegungen mit großer Wahrscheinlichkeit sogar mehr als 203 Gebäude erwarten kann, ist der Gesamtstichprobenumfang von 635 Wohngebäuden auch bezüglich der Frage nach dem mittleren BRI der Klasse K_1 ausreichend.

2.2.3.4 Die Zufallsauswahl

Für die zufällige Auswahl der Gebäude wurden gemäß dem Stichprobenumfang und der Gesamtzahl der Gebäude Zufallszahlen mit einem FORTRAN-Zufallszahlengenerator generiert. Z.B. wurden für die Stichprobe der Wohngebäude zufällig 635 Zahlen (ohne Wiederholungen) aus den Zahlen 1-3084 ausgewählt.

2.3 Praktisches Vorgehen bei der Datenerhebung

2.3.1 Übernahme von Daten aus ALB und ALK

Aus dem ALB werden für die Erfassung in der Gebäudedatenbank folgende Daten für den gesamten Gebäudebestand der Kernstadt Ettlingen direkt übernommen:

- Straße und Hausnummer,
- Flurstücksnummer,
- Größe des Flurstücks [m^2],
- Nutzung des Flurstücks,
- Gebäudenutzung.

Aus der ALK können mit Hilfe eines Geoinformationssystems noch die „Überbauten Flächen“ [m^2] erhoben werden. Die Auswertung des ALB und der ALK liefert das in Tabelle 2.4 angeführte Ergebnis für den Gesamtbestand.

Klasse	ALB überprüft	sql aus ALB	obj aus ALK	txt aus ALK
$K_1 - K_3$	3084	3086		2935
K_4	198	176		184
K_5	8	8		8
K_6	697	679		631
K_7	1186	1047		1087
K_8	85	75		76
Garagen (G)	1543	1433		1185
Summe ohne (G)	5258	5068		4921
Summe mit (G)	6801	6501	6289	6106

Tabelle 2.4: Gesamtzahl aller Gebäude gemäß Datenquellen

Die Unterschiede zwischen der überprüften Gebäudeanzahl (Spalte „ALB überprüft“) und den Ergebnissen aus der SQL-Abfrage (Spalte „sql aus ALB“) erklären sich aus der Struktur der Datenbank, die für die Benennung der Gebäude nicht im Sinne einer relationalen Datenbank ausgeführt sind. Das Textfeld liefert im Sinne der eigentlichen Nutzung nur einen ergänzenden Text und wird „frei ausgefüllt“. Dies bedeutet, dass jeder Tippfehler nicht im Abfrageraster erscheint. Die Spalten „obj aus ALK“ und „txt aus ALK“ kommen aus Abfragen, die mittels des Geoinformationssystems auf Basis der ALK erfolgten. Bei der Objektabfrage wurden nur als Gebäude definierte Objekte erfasst, welche über eine Objektnummer und nicht über die Beschriftung der Nutzung zu identifizieren sind. Da die Objektnummer oftmals die Hausnummer ist, kann diese nicht als eindeutiger Identifikator dienen. Es müssen in jedem Fall die Objektkoordinaten zur eindeutigen Lokalisierung mitgeführt werden, was die Datenmenge erheblich wachsen lässt. Auch aus diesem Grund gab es später Probleme bei der Stichprobenzuordnung zu den Gebäuden und der Ermittlung der jeweiligen bebauten Flächen, vgl. Abschnitt 2.3.5.
Jedes als Gebäude definierte Objekt verfügt darüber hinaus über einen Text, der die Gebäudenutzung beschreibt. Damit kann die Nutzungsverteilung gemäß ALK festgestellt werden. Die Differenzen zu den Zahlen aus dem ALB sind vor allem durch fehlende Aktualisierungen begründet. Das ALB ist meist aktueller als die ALK, was auch die insgesamt niedrigere Anzahl von Gebäuden in der ALK erklärt. Wenn in einzelnen Klassen die ALK eine größere Anzahl Gebäude liefert, so ist dies mit Sicherheit die genauere Zahl, da jedes Gebäude einen Text erhält und im ALB das schon zu Beginn beschriebene Problem mit der Bezeichnung der Gebäude vorliegt.

2.3.2 Daten der Feuerversicherung

Zunächst wurden die Erhebungsbögen der Feuerversicherungen analysiert, die sich im Laufe der Jahre auch geringfügig verändert haben. Dann wurden die zu erhebenden Merkmale festgelegt und für die Erfassung eine spezielle Eingabemaske erstellt, die eine reibungsfreie Übernahme der Daten in die Datenbank ermöglicht.

Die durch die oben beschriebene Stichprobennahme ermittelten Gebäude wurden dann bei der Feuerversicherung aus den Akten erhoben. Aus Zeit- und Kostengründen wurden nur vier Nutzungsklassen im kompletten Stichprobenumfang bearbeitet, um die Methodik und die möglichen Auswertungen aus den Daten zu veranschaulichen. Da die Akten nach Liegenschaften organisiert sind, wurden jeweils alle auf der Liegenschaft befindlichen Gebäude miterfasst, d.h. es wurden auch Gebäude, die nicht in der Stichprobe der jeweiligen Nutzungsklasse erhalten sind, miterfasst. Diese werden dann aber bei der Auswertung getrennt aufgeführt. Erhoben wurden die Klassen K_1, K_2, K_3 der Wohnegebäude und die Klasse der Nichtlandwirschaftlichen Betriebsgebäude K_7. Die Angaben aus dem ALB dienten

in der Datenbank als Grundlage. Folgende Merkmale wurden bei der Feuerversicherung erhoben:

- Nutzung,
- Baujahr,
- Rauminhalt Keller,
- Rauminhalt Normalgeschosse,
- Rauminhalt Dachraum,
- Rauminhalt Zusatzräume,
- Materialien Außenwand,
- Dachform (Sattel-, Pult-, Flachdach etc.),
- Material Dachdeckung,
- Veränderung (Beschreibung der Maßnahme und Auswirkungen auf Material, Nutzung und Volumen),
- Veränderungsgrund.

Der Rauminhalt, der bei der Feuerversicherung beschrieben wird, entspricht nicht einer exakten Ermittlung des Bruttorauminhaltes gemäß DIN 277.

2.3.3 Bebauungsplan

Da für die Kernstadt Ettlingen außer für Teile der Gewerbegebiete keine im heutigen Sinne qualifizierten Bebauungspläne vorlagen, konnten hier keine weiteren Auswertungen gemacht werden. Es wurden jedoch die Gebäude der jeweiligen Gebiete der alten Baufluchtenpläne und Bebauungspläne mittels des GIS selektiert und ihnen das jeweilige Jahr des Inkrafttretens zugeordnet. Dies ermöglicht flächendeckend einen ersten Überblick der Altersverteilung der Bausubstanz.

2.3.4 Bauanträge

Da nicht für alle Gebäude, die in der Stichprobe ermittelt wurden, die Akte bei der Feuerversicherung auffindbar war, wurden für die entsprechenden Gebäude die Bauantragsakten ausgewertet. Diese stellten sich jedoch als wesentlich heterogener als die Feuerversicherungsakten heraus. Hier sind zwar in den meisten Fällen Baupläne hinterlegt, aber die Daten, die für die Ermittlung der Baustruktur benötigt werden, waren vor allem in den älteren Akten nicht vollständig. In jedem Fall war aber das Baujahr des Gebäudes feststellbar. Angaben über Rauminhalte waren nur teilweise vorhanden.

2.3.5 Probleme bei der Datenerhebung

Zunächst muss festgestellt werden, dass ALB und ALK wie vermutet nicht auf dem aktuellsten Stand sind. Die ALB-Daten wurden trotzdem als gültiger Bestand für die Berechnungsgrundlage festgelegt. Diesem kann damit kein bestimmtes Jahr zugeordnet werden, in dem der Bestand aktuell war oder ist. In den Stichproben wurden keine nach 1996 erstellten Gebäude gefunden, nach Auskunft der Baubehörde heißt das jedoch nicht, dass der Bestand bis 1996 komplett dokumentiert ist.

Eine zweite Schwierigkeit lag darin, dass für Gebäude, die in der Stichprobe untersucht werden sollten, keine Akte bei der Feuerversicherung auffindbar war. Gründe hierfür waren:

1. Die Akte des entsprechenden Gebäudes war entliehen.
2. Die Akte befindet sich bei dem Versicherungsnehmer. Dies war überwiegend bei gewerblich genutzten Gebäuden der Fall.

Für die Gebäude aus der Stichprobe, die nicht in der Feuerversicherung (FVS) aufzufinden waren, wurden die Akten aus dem städtischen Bauaktenarchiv herangezogen. Hierbei lag das große Problem in den mangelnden Angaben zu Rauminhalten, was vor allem bei alten Gebäuden der Fall war. Die Statistik der Erhebung aus den Datenquellen ist in den Tabellen 2.5 und 2.6 wiedergegeben.

Stichprobenumfang laut Tabelle 2.3		635
Davon in den FVS Akten gefunden		586
Stichprobengebäude ohne FVS Akte aber mit Bauakte		26
davon mit BRI	13	
Summe der erhobenen Gebäude		612
davon K_1	539	
davon K_2	68	
davon K_3	5	
Summe der erhobenen Gebäude mit BRI		599
davon K_1	533	
davon K_2	61	
davon K_3	5	
Keinerlei Akten vorhanden		23

Tabelle 2.5: Ergebnis der Datenerhebung der Wohngebäude

Die Schwierigkeiten hinsichtlich der Datenorganisation ergaben sich hauptsächlich aus der Zuordnung der Gebäude. Im ALB wurden Einzelgebäude bestimmt, die

eine Adresse und eine Flurstücknummer hatten. Die Feuerversicherung organisiert ihre Akten nach Adressen. Bei der Stichprobe für Wohngebäude war eine Zuordnung meist unproblematisch, da die Gebäude (meist sogenannte Nebengebäude), die mit den Wohngebäuden auf dem Grundstück stehen, unterschiedliche Nutzungen haben, z.B. Wohngebäude, Stall, Scheune, Garage. Bei Gewerbebetrieben sind jedoch meist mehrere Gebäude gleicher Nutzung unter einer Adresse archiviert. Daher wurden jeweils alle in der Akte enthaltenen Gebäude erhoben. Dies erklärt vor allem bei der Klasse K_7 die hohe Anzahl der „Zusatzgebäude“, vgl. Tabelle 2.6.

Stichprobenumfang laut Tabelle 2.3		368
Davon in den FVS Akten gefunden		206
Zusätzlich erhobene Gebäude (unter derselben Adresse)		771
Stichprobengebäude ohne FVS Akte aber mit Bauakte		34
davon mit Baujahr	32	
davon mit BRI	29	
davon mit BRI und Baujahr	27	
Gesamtzahl der erhobenen Gebäude		1021
davon $K_1 - K_3$	26	
davon K_4	64	
davon K_6	23	
davon K_7	903	
davon K_8	3	
davon ohne Nutzungsangabe	2	

Tabelle 2.6: Ergebnis der Datenerhebung der Klasse K_7

Eine Besonderheit, die bei den Wohngebäuden auftrat, ergab sich aus der Definition eines Gebäudes. So wird ein Mehrfamilienhaus, das drei Eingänge hat und somit drei Adressen, im ALB als ein Gebäude „Wohnhaus“ geführt. In der Feuerversicherung können, je nach Besitzverhältnissen, daraus drei Gebäude werden.

2.4 Datenorganisation und Datenbank

Um die erhobenen Daten optimal nutzen und auswerten zu können, ist die Anlage einer Datenbank unerlässlich. Es wird hier nicht der komplette Aufbau der Datenbank beschrieben sondern es werden lediglich einige dafür erforderliche Grundsätze und Problemstellungen erläutert.

Für dieses Projekt wurde eine eigene Access Datenbank angelegt. Diese „externe“ Datenbank war notwendig, da sich die ALB-Datenbank in ihrer Struktur nicht für

die Untersuchungszwecke erweitern ließ. Dies liegt im Wesentlichen daran, dass die für das Projekt erforderliche Grundeinheit „Gebäude“ nicht die Grundeinheit im ALB bildet, wo die kleinste Identifikationseinheit das Flurstück ist, auf dem sich, wie bereits erwähnt, mehrere Gebäude befinden können. Außerdem hätten wegen Datenschutzrechtlicher Belange die Besitzerdaten gesperrt werden müssen. Um für dieses Projekt zunächst auf die Vorteile einer gemeinsamen Nutzung von ALB und ALK mittels eines Geoinformationssystems (GIS) zurückgreifen zu können, wurde die neue Access Gebäudedatenbank über eine ODBC Schnittstelle mit der ALB Datenbank verbunden. Dies hat neben der räumlichen Darstellung der Ergebnisse den Vorteil, dass die auf räumlicher Ebene zu erfassenden Informationen (Angaben zu Altersklassen aufgrund der Erstellung der Bebauungspläne, Räumliche Untersuchungen zur Altstadtsanierung, etc.) über Blockweise Selektion in der ALK gebündelt in der Datenbank ergänzt werden können. Die Selektion kann wegen des oben beschriebenen Problems jedoch nur Grundstücksweise erfolgen. Die neu angelegte Datenbank kann als reine Gebäudedatenbank auch unabhängig von ALB und ALK weiter genutzt werden.

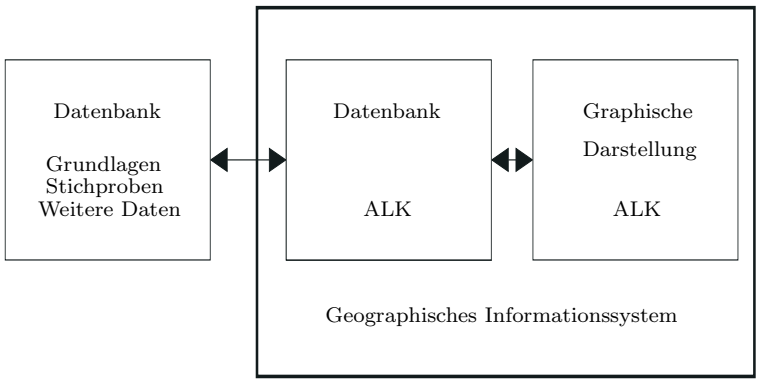


Abbildung 2.5: Datenbankstruktur

Die Festlegung des Bestandes wurde bereits erläutert. Alle im ALB vorhandenen Gebäude wurden festgestellt und bekamen in der Datenbank einen numerischen Identifikator (db_id). Einige weitere Angaben wurden mit aus dem ALB übernommen und zugeordnet (Nutzung, Flurstücknummer, Adresse, etc. vgl. Abschnitt 2.3.1). Diese werden im weiteren Verlauf als „Attribute“ bezeichnet. Alle Daten, die aus weiteren Datenquellen stammen, sollen nun den Gebäuden zugeordnet werden. Je nach Datenquelle muss ein Identifikator gefunden werden, der das Gebäude eindeutig bezeichnet und damit dem Basisdatensatz aus dem ALB zuzuordnen ist. Jedes bereits vorhandene Attribut kann zu Identifikationszwecken in beliebiger Kombination verwendet werden. Im Wesentlichen werden die identifizierenden Attribute durch die Inhalte und Organisation der Datenquellen bestimmt. Die in der Stichprobe aus dem ALB ausgewählten Gebäude, die durch Da-

ten der Feuerversicherungs- und Bauordnungsamtakten zu ergänzen waren, wurden zunächst über die Adresse identifiziert. Solange nur ein Gebäude einer Adresse zugeordnet wird, ist dieses Attribut als Identifikator ausreichend. Dies ist jedoch selten der Fall. Teilweise genügt als zweiter Identifikator die Nutzung des Gebäudes. So sind vor allem im Ortskern bei alten Hofreiten meist „Wohngebäude“, „Stall“ oder „Betriebsgebäude“ und „Scheune“ mit der gleichen Adresse belegt. Da in diesem Beispiel die Gebäudenutzung nur jeweils einmal pro Flurstück oder Adresse vorkommt, ist über das Attribut „Adresse“ und „Nutzung“ eine eindeutige Zuordnung möglich. Voraussetzung ist natürlich, dass beide Attribute jeweils in den Datenquellen vorhanden sind. Schwierig wird die Zuordnung vor allem bei Gewerbegeländen. Hier befinden sich oftmals mehrere Fabrik- oder Betriebsgebäude auf einem Gelände. Teilweise ist es möglich, über die Größe der Gebäude eine Zuordnung zu schaffen.

Bei der maschinellen Verknüpfung innerhalb der Datenbank müssen bestimmte Konventionen beachtet werden. Während die Vereinheitlichung der Schreibweise (Flurstücknummern, Adressen etc.) ein geringes Problem darstellt, erfordert die Vereinheitlichung der Bezeichnungen unter Umständen die Erstellung von Zuordnungstabellen. Dies ist z.B. notwendig für Nutzung, Flächenangaben oder Massnahmen an Gebäuden. Hinsichtlich der Gebäudenutzung gibt es Bezeichnungen gemäß der „Verwaltungsvorschrift für die Führung des Liegenschaftsbuches mit automatischer Datenverarbeitung“, wie sie im ALB verwendet wird, dann den „Signierschlüssel für Nichtwohngebäude“, vgl. Statistisches Bundesamt (1977), oder den Bauwerkszuordnungskatalog, um hier nur einige zu nennen. In diesem Projekt wurden zunächst die Nutzungen im ALB den Nutzungsklassen, wie sie vom Statistischen Bundesamt verwendet werden, zugeordnet. Die Nutzungszuordnungstabelle kann für neue Datenquellen jeweils erweitert werden.

Im Bereich der Materialien und Baumaßnahmen gibt es keine vorgegebenen Gruppierungen. Hier wurden die Angaben aus den jeweiligen Datenquellen übernommen und nachträglich in einer Zuordnungstabelle überarbeitet und teilweise auch in eine für die Auswertung sinnvolle Klasseneinteilung überführt. Sämtliche Grunddaten sind jedoch vorhanden und die Klasseneinteilung lässt sich, je nach Untersuchungszweck, wieder neu gestalten. Das gleiche gilt auch für die Angaben zu den baulichen Veränderungen an Gebäuden.

Was die Einheiten hinsichtlich Raum- und Flächeninhalten anbelangt, so trifft man hier ebenfalls auf Unterschiede. Selbst wenn die Berechnungen nach DIN 277 erfolgt sind, können geringfügige Unterschiede aufgrund in verschiedenen Jahren leicht geänderter Berechnungsvorschriften auftreten. Oder es wird - wie beispielsweise in den Feuerversicherungsakten - ein Rauminhalt berechnet, der nicht mit dem Bruttorauminhalt nach DIN 277 exakt gleichzusetzen ist. Hier gilt es, die ge-

naue Berechnungsgrundlage zu dokumentieren und Fallweise zu untersuchen, welche Werte vergleichbar sind. Sollen Kostenkennwerte den Gebäuden zugeordnet werden, so ist auch hier genau die Basis, auf der die Kennwerte erhoben wurden (incl. Jahr) zu dokumentieren, um spätere Ergänzungen auf der richtigen Ebene zuordnen zu können.

Kapitel 3

Analyse des Gebäudebestandes

Dieses Kapitel behandelt die *statistische Analyse von Gebäudebeständen* am Beispiel der Stadt Ettlingen. Ausgangspunkt sind die Datenbestände, die aus den Akten der Feuerversicherung und des Bauordnungsamtes gewonnen wurden, was in Kapitel 2 ausführlich beschrieben ist. Anhand dieser Daten werden *statistisch fundierte* Aussagen über das Alter, die Größe, die Zusammensetzung u.v.m. des Gebäudebestandes von Ettlingen getroffen. Viele der verwendeten statistischen Methoden sind auf andere Bestände (andere Städte, Regionen, o.ä.) übertragbar, vorausgesetzt es liegen vergleichbare Datensätze vor.
Zu beachten ist zudem, dass es sich hier um eine *statische* Analyse des Gebäudebestandes handelt, d.h. die getroffenen Aussagen beschreiben den Gebäudebstand zu einem festen Zeitpunkt, wobei der Bestand des ALB im Jahr 2000 nicht dem realen Bestand dieses Jahres entspricht, vgl. Kapitel 2. Auf die Veränderung von Gebäudebeständen im Laufe der Zeit wird in Kapitel 4 eingegangen, Prognosen für den zukünftigen Bestand sind Thema von Kapitel 5.
Grundlage für die Anwendung statistischer Methoden sind hinreichend große Datensätze. Da die Datenerhebung sehr zeitaufwendig und somit sehr teuer ist und in der vorliegenden Arbeit vor allem die Methodik unseres Vorgehens an einem Beispiel erläutert werden soll, werden nicht alle der verschiedenen Nutzungsklassen $K_1 - K_8$ untersucht, sondern lediglich die Wohngebäude $K_1 - K_3$ sowie die nichtlandwirtschaftlichen Betriebsgebäude K_7.

3.1 Bestandsanalyse der Wohngebäude

Zunächst werden die Wohngebäude Ettlingens betrachtet. Im Folgenden treten viele verschiedene Variablen auf, deshalb sei schon hier darauf hingewiesen, dass in Anhang C eine tabellarische Übersicht über alle Variablen mit der jeweiligen Bedeutung gegeben ist. Diese Tabelle ist nach den einzelnen Abschnitten geordnet und dient als kleine Nachschlagehilfe.

3.1.1 Gebäudezahl in den Klassen K_1, K_2, K_3

Bei den Wohngebäuden gab es eine Besonderheit der Datenerhebung, die in einem ersten Schritt näher beleuchtet werden soll. Zwar konnten dem ALB die

Gesamtzahl aller Wohngebäude Ettlingens $N_{wohn} = 3084$

entnommen werden. Diese schlüsselte sich aber nach Klasseneinteilung aus Abschnitt 2.2.2 auf in die Klasse K_1 der *Einfamilienhäuser* mit N_1 Gebäuden, die Klasse K_2 der *Mehrfamilienhäuser* mit N_2 Gebäuden und die Klasse K_3 der *Wohnhochhäuser* mit N_3 Gebäuden. Die Zahlen N_1, N_2, N_3 waren dabei unbekannt. Bezeichnet man mit $q_i = \frac{N_i}{N_{wohn}}$, $i = 1, 2, 3$ den prozentualen Anteil der einzelnen Klassen am Gesamtbestand der Wohngebäude, so werden in diesem Abschnitt *Konfidenzintervalle für die* q_i angeben, also Intervalle, in denen die unbekannten q_i mit einer großen festgelegten Sicherheit liegen.
Zunächst werden die dafür relevanten Daten der Stichprobe aufgelistet. Es sei n_i, $i = 1, 2, 3$, *die Anzahl der Wohngebäude aus Klasse* K_i, die in unserer Stichprobe enthalten ist. Weiter sei $n_{wohn} = n_1 + n_2 + n_3$ die *Gesamtzahl der in der Stichprobe enthaltenen Wohngebäude*. Konkret ergab die Stichprobe die in Tabelle 3.1 abgebildeten Zahlen.

	Gesamtzahl	K_1	K_2	K_3
Anzahl in der Stichprobe	$n_{wohn} = 612$	$n_1 = 539$	$n_2 = 68$	$n_3 = 5$

Tabelle 3.1: Aufschlüsselung der Stichprobe auf K_1, K_2, K_3

Man beachte, dass nach Tabelle 2.3 eine Stichprobe von 635 Wohngebäuden vorgesehen war. Per Zufallszahlengenerator wurden 635 Wohngebäude aus dem ALK ermittelt, zu denen aber bei den Recherchen in den Feuerversicherungen und dem Bauordnungsamt nicht immer eine Akte gefunden werden konnte. Zu dieser Problematik vergleiche man Abschnitt 2.3.5.

Es werden nun die Ausführungen aus Anhang A.1 angewandt und die Modellierung (A.5) aus Beispiel A.1.5 für die Klassen K_1, K_2, K_3 vorausgesetzt. Alle auftretenden Größen seien bezüglich dieser Modellierung zu verstehen.

Für die unbekannten prozentualen Anteile q_1, q_2, q_3 erhält man mit dem Stichprobenmittel (A.1) die *Schätzungen*

$$\overline{q}_1 = \frac{539}{612} \approx 88,07\,\%, \quad \overline{q}_2 = \frac{68}{612} \approx 11,11\,\%, \quad \overline{q}_3 = \frac{5}{612} \approx 0,82\,\%. \tag{3.1}$$

Natürlich sind diese Schätzungen wenig aussagekräftig, so lange man nichts über deren Qualität weiss. Dazu benötigt man die zugehörigen *Stichprobenvarianzen*.

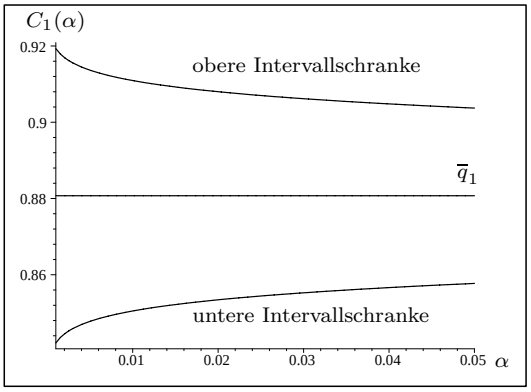
Diese berechnen sich mittels (A.2) zu

$$\hat{\sigma}_1^2 = \frac{39347}{374544} \approx 0,1051, \quad \hat{\sigma}_2^2 = \frac{8}{81} \approx 0,0988, \quad \hat{\sigma}_3^2 = \frac{3035}{374544} \approx 0,0081.$$

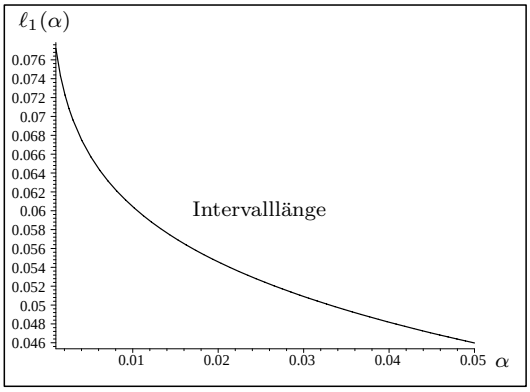
Jetzt können zu jedem $\alpha \in (0,1)$ asymptotische Konfidenzintervalle zum Niveau $1-\alpha$ gemäß (A.6) berechnet werden. Der Stichprobenumfang von $n_{wohn} = 612$ ist groß genug, um dies Asymptotik zu rechtfertigen. Ohne weiteres kann man davon sprechen, dass die tatsächliche Überdeckungswahrscheinlichkeit der asymptotischen Konfidenzintervalle „genau“ $1-\alpha$ ist. Damit erhält man für die drei unbekannten Werte q_i, $i = 1,2,3$, die folgenden Konfidenzintervalle $C_i(\alpha)$, wobei das Argument α die Abhängigkeit vom Niveau zum Ausdruck bringen soll.

$$\begin{aligned} C_i(\alpha) &= \\ &= \left[\overline{q}_i - \sqrt{\frac{\hat{\sigma}_i^2}{n_{wohn}} \left(1 - \frac{n_{wohn}}{N_{wohn}}\right)} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \right. \\ &\qquad \left. \overline{q}_i + \sqrt{\frac{\hat{\sigma}_i^2}{n_{wohn}} \left(1 - \frac{n_{wohn}}{N_{wohn}}\right)} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \right] \\ &= \left[\overline{q}_i - \sqrt{\frac{\hat{\sigma}_i^2 \cdot 103}{78642}}\, \Phi^{-1}\left(1 - \frac{\alpha}{2}\right),\ \overline{q}_i + \sqrt{\frac{\hat{\sigma}_i^2 \cdot 103}{78642}}\, \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \right]. \end{aligned} \tag{3.2}$$

Dieses Ergebnis soll graphisch sowie an Beispielen verdeutlicht werden. Die obere und die untere Schranke der Intervalle werden als Funktionen von α in ein Schaubild eingezeichnet. In einem separaten Schaubild wird die Länge der Konfidenzintervalle in Abhängigkeit von α veranschaulicht.



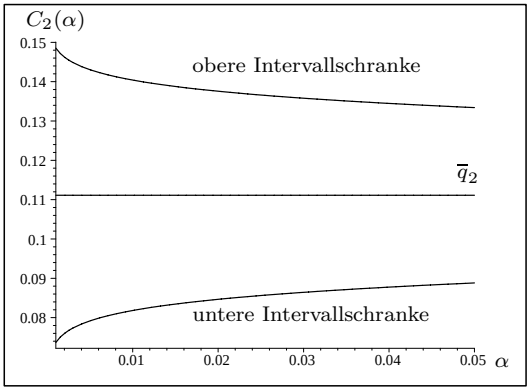
Konfidenzintervalle $C_1(\alpha)$



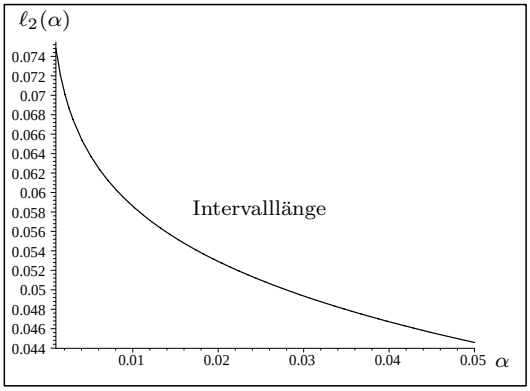
Intervalllänge $|C_1(\alpha)|$

Abbildung 3.1: Konfidenzintervalle $C_1(\alpha)$ für q_1

Für kleine Werte α sind zu $C_3(\alpha)$ gehörigen unteren Intervallschranken gemäß 3.2

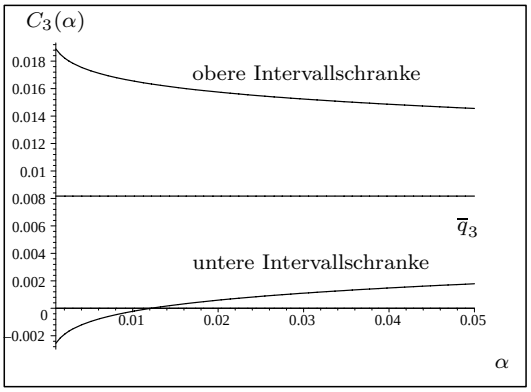


Konfidenzintervalle $C_2(\alpha)$

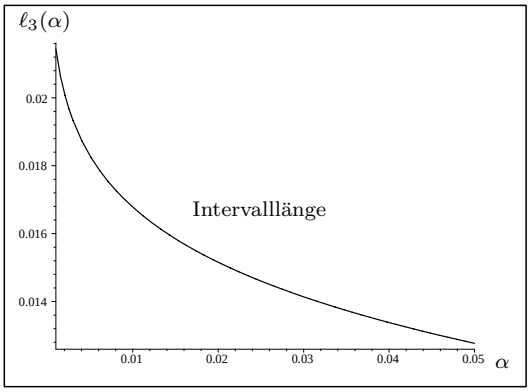


Intervalllänge $|C_2(\alpha)|$

Abbildung 3.2: Konfidenzintervalle $C_2(\alpha)$ für q_2



Konfidenzintervalle $C_3(\alpha)$



Intervalllänge $|C_3(\alpha)|$

Abbildung 3.3: Konfidenzintervalle $C_3(\alpha)$ für q_3

alle negativ. Deshalb können sie im Unterschied zu Abbildung 3.3 auch durch die natürliche Schranke 0 ersetzt werden. Die Intervalllänge entspricht dann dem Wert der oberen Intervallschranke.

Abschließend werden noch einige explizite Beispiele angeführt.

Entscheidet man sich beispielsweise für das Niveau $\alpha = 0,03$, so erhält man die Konfidenzintervalle

$$C_1(0,03) = [0,8553;\ 0,9062], \quad C_2(0,03) = [0,0864;\ 0,1358],$$

$$C_3(0,03) = [0,0011;\ 0,0152].$$

Der wahre Anteil q_2 der Mehrfamilienhäuser an den gesamten Wohngebäuden liegt also mit einer 97%–igen Sicherheit in dem Intervall $C_2(0,03) = [8,64\%;\ 13,58\%]$. Ebenso ist mit einer 97%–igen Sicherheit der Anteil der Wohnhochhäuser geringer

als $1,52\%$.

Entscheidet man sich für eine große Sicherheit der Aussage, d.h. für ein hohes Niveau der Konfidenzintervalle, so muss α sehr klein gewählt werden, z.B. $\alpha = 0,001$. Das ergibt die Intervalle

$$C_1(0,001) = [0,8421;\ 0,9193], \quad C_2(0,001) = [0,0737;\ 0,1485],$$

$$C_3(0,001) = [0;\ 0,0189].$$

Der wahre Anteil q_1 der Einfamilienhäuser an den gesamten Wohngebäuden liegt damit mit einer $99,9\%$–igen Sicherheit im Intervall $C_1(0,001) = [84,21\%;\ 91,93\%]$. Abbildung 3.4 soll das Ergebnis nochmals verdeutlichen. Der obere Bereich jedes Balkens stellt dabei den Bereich (das Konfidenzintervall), in dem der wahre Prozentsatz mit 95%–iger Sicherheit liegt, dar.

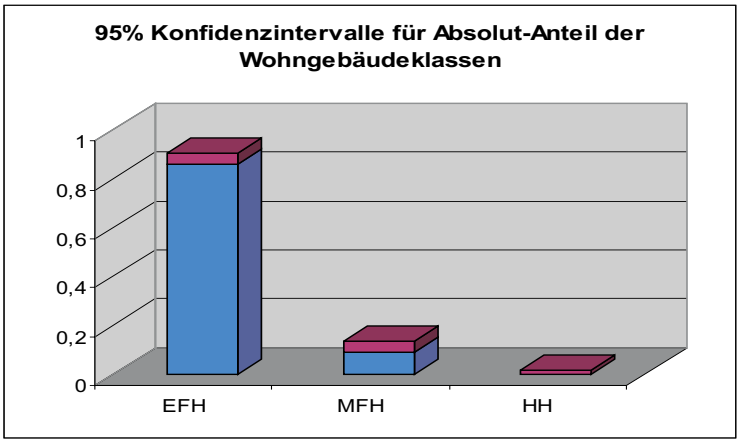


Abbildung 3.4: Konfidenzintervalle zum 95%-Niveau

3.1.2 Analyse des Bruttorauminhalts

Der nächste wichtige Punkt der Gebäudebestandsanalyse ist die Untersuchung des Bruttorauminhaltes, dessen Berechnung in der DIN 277, Deutsches Institut für Normung e.V. (1988), festgelegt ist. Das komplette Volumen eines Gebäudes wird damit beschrieben, sowohl unterirdisch (Keller und Fundamente) wie auch der sichtbare Gebäudeteil. Maßgeblich sind dabei die äusseren Begrenzungsflächen. Der Bruttorauminhalt in den Akten der Feuerversicherungen ist nur in Anlehnung an die DIN 277 grob ermittelt. Die Berechnung erfolgt überwiegend über die Gebäudeaußenmaße, d.h. dass z.B. nicht jeder Erker detailliert erfasst wurde.

Im Gegensatz zu Abschnitt 3.1.1, wo zu allen 612 Gebäuden des vorliegenden Datensatzes eine Zuordnung zu einer der Klassen $K_1 - K_3$ möglich war, liegt hier nicht

bei jedem Gebäude der Bruttorauminhalt vor. Genauer gesagt war eine Ermittlung des BRI aus den Akten des Bauordnungsamtes nicht in allen Fällen möglich. Insgesamt ergaben sich die in Tabelle 3.2 dargestellten Datenumfänge.

	Gesamtzahl	K_1	K_2	K_3
Anzahl	$n_{BRI,wohn} = 599$	$n_{BRI,1} = 533$	$n_{BRI,2} = 61$	$n_{BRI,3} = 5$

Tabelle 3.2: Stichprobenumfänge mit bekannter BRI

Zunächst sollen ganz analog zu Abschnitt 3.1.1 einfache Untersuchungen zum *mittleren BRI* angestellt werden. Dabei wird zum einen der Bestand der Wohngebäude gesamtheitlich, d.h. nicht nach Klassen aufgeschlüsselt untersucht. Der unbekannte mittlere BRI aller 3084 Wohngebäude wird mit $\overline{y}_{BRI,wohn}$ bezeichnet. Für diesen uns interessierenden Wert werden wiederum Konfidenzintervalle angegeben, in denen der Wert mit großer Sicherheit liegt.
Zum anderen werden die beiden Klassen K_1 und K_2 separat betrachtet. Die Klasse K_3 der Wohnhochhäuser bietet mit nur fünf Vertretern in der Stichprobe ($n_{BRI,3} = 5$) keine Grundlage für sinnvolle statistische Untersuchungen. Mit $\overline{y}_{BRI,1}$ bzw. $\overline{y}_{BRI,2}$ werden die unbekannten mittleren BRIe in den Klassen K_1 bzw. K_2 bezeichnet.
Als *Schätzungen* für diese unbekannten mittleren BRIe ergeben sich aus der Stichprobe die folgenden Mittelwerte

$$\overline{Y}_{BRI,wohn} = 1.325\ [m^3], \qquad \overline{Y}_{BRI,1} = 1.036\ [m^3], \qquad \overline{Y}_{BRI,2} = 3.210\ [m^3].$$

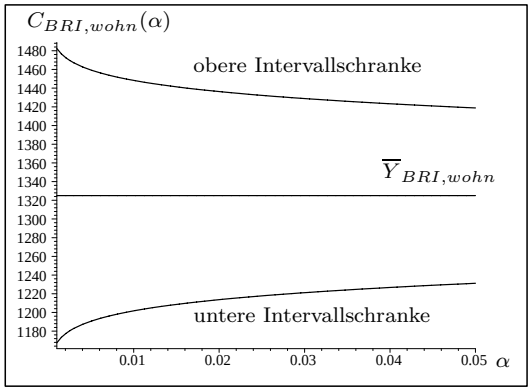
Für die Konstruktion von Kofidenzintervallen benötigt man natürlich auch wieder die zugehörigen *Stichprobenvarianzen*

$$\hat{\sigma}^2_{BRI,wohn} = 1.702.544, \qquad \hat{\sigma}^2_{BRI,1} = 325.614, \qquad \hat{\sigma}^2_{BRI,2} = 3.186.646.$$

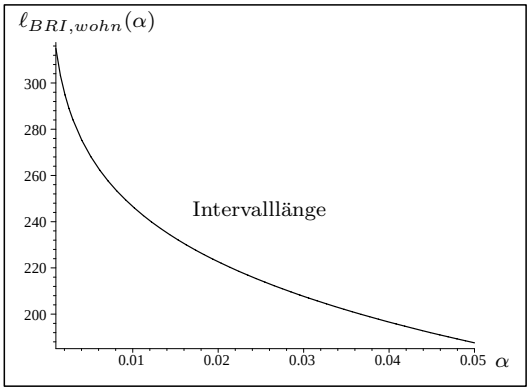
Zunächst wird nun ein asymptotisches Konfidenzintervall für $\overline{y}_{BRI,wohn}$ angegeben. Man beachte dabei, dass auf Grund des großen Stichprobenumfangs $n_{BRI,wohn} = 599$ asymptotische Ergebnisse ohne Bedenken angewandt werden können. In Abhängigkeit von $\alpha \in (0,1)$ ergibt sich aus Formel (A.6)

$$\begin{aligned} C_{BRI,wohn}(\alpha) = & \\ = & \left[\overline{Y}_{BRI,wohn} - \sqrt{\frac{\hat{\sigma}^2_{BRI,wohn}}{n_{BRI,wohn}}\left(1 - \frac{n_{BRI,wohn}}{N_{wohn}}\right)}\Phi^{-1}\left(1-\frac{\alpha}{2}\right),\right. \\ & \left. \overline{Y}_{BRI,wohn} + \sqrt{\frac{\hat{\sigma}^2_{BRI,wohn}}{n_{BRI,wohn}}\left(1 - \frac{n_{BRI,wohn}}{N_{wohn}}\right)}\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right] \\ = & \left[1.325 - 47,86 \cdot \Phi^{-1}\left(1-\frac{\alpha}{2}\right), 1.325 + 47,86 \cdot \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right]. \end{aligned}$$

Die folgende Abbildung veranschaulicht dieses Ergebnis.



Konfidenzintervalle $C_{BRI,wohn}(\alpha)$



Intervalllänge

Abbildung 3.5: Konfidenzintervall $C_{BRI,wohn}(\alpha)$ für $\overline{y}_{BRI,wohn}$

Auch an Zahlenbeispielen soll das Ergebnis verdeutlich werden.
So liegt der unbekannte mittlere BRI $\overline{y}_{BRI,wohn}$ mit einer Wahrscheinlichkeit von 95% ($\alpha = 0,05$) im Intervall

$$C_{BRI,wohn}(0,05) = [1.231\ m^3,\ 1.419\ m^3].$$

Die Intervalllänge beträgt 188 m^3.
Möchte man eine Sicherheit von 99%, so ergibt sich als Konfidenzintervall

$$C_{BRI,wohn}(0,01) = [1.202\ m^3,\ 1.448\ m^3].$$

Ist sogar eine Sicherheit von $99,9\%$ erwünscht, so erhält man als zugehöriges Intervall

$$C_{BRI,wohn}(0,001) = [1.168\ m^3,\ 1.482\ m^3],$$

mit einer Länge von 314 m^3.

Schwieriger ist die Angabe von Konfidenzintervallen für $\overline{y}_{BRI,1}$ und $\overline{y}_{BRI,2}$. Zwar können in beiden Fällen die asymptotischen Konfidenzintervalle betrachtet werden (die Stichprobenumfänge von $n_{BRI,1} = 533$ bzw. $n_{BRI,2} = 61$ sind hierfür groß genug), diese sind nach (A.6) aber von der Form

$$C_{BRI,i}(\alpha) = \left[\overline{Y}_{BRI,i} - \sqrt{\frac{\hat{\sigma}^2_{BRI,i}}{n_{BRI,i}}\left(1 - \frac{n_{BRI,i}}{N_i}\right)}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right),\right.$$
$$\left.\overline{Y}_{BRI,i} + \sqrt{\frac{\hat{\sigma}^2_{BRI,i}}{n_{BRI,i}}\left(1 - \frac{n_{BRI,i}}{N_i}\right)}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right],$$

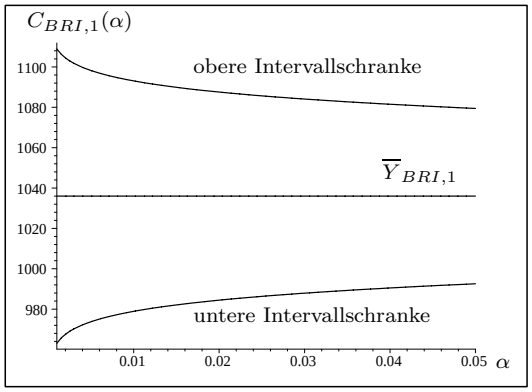
$i = 1, 2$, mit unbekannten N_1 bzw. N_2. Diese Problematik lässt sich dadurch lösen, dass man die Quotienten $\frac{n_{BRI,i}}{N_i} = \frac{n_{BRI,i}}{q_i N_{wohn}}$ durch die (konsistenten) Schätzer $\frac{n_{BRI,i}}{\overline{q}_i N_{wohn}}$, $i = 1, 2$, aus (3.1) ersetzt. Dies ändert nichts an der Asymptotik! Dadurch ergeben sich die asymptotischen Konfidenzintervalle

$$C_{BRI,1}(\alpha) = \left[1.036 - 22,16 \cdot \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), 1.036 + 22,16 \cdot \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right]$$

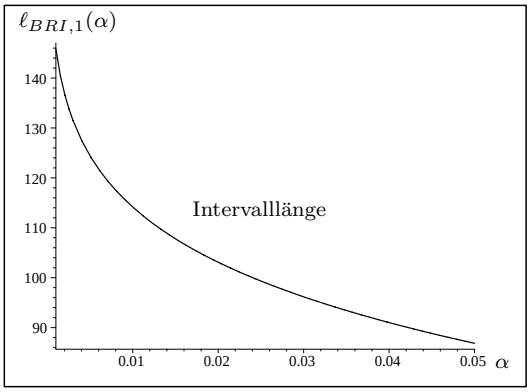
bzw.

$$C_{BRI,2}(\alpha) = \left[3.210 - 207,22 \cdot \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), 3.210 + 207,22 \cdot \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right].$$

Im Fall der Klasse K_1 der Einfamilienhäuser ergibt das die folgenden Schaubilder.



Konfidenzintervalle $C_{BRI,1}(\alpha)$



Intervalllänge

Abbildung 3.6: Konfidenzintervall $C_{BRI,1}(\alpha)$ für $\overline{y}_{BRI,1}$

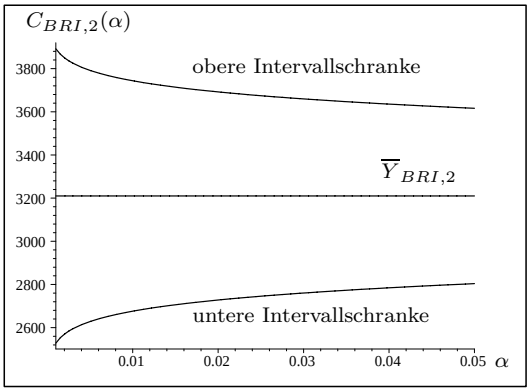
Auch hier ein kurzes Beispiel: Bei einem Niveau von 98%, d.h. $\alpha = 0.02$ ergibt sich als Konfidenzintervall

$$C_{BRI,1}(0,02) = [984\ m^3,\ 1.086\ m^3]$$

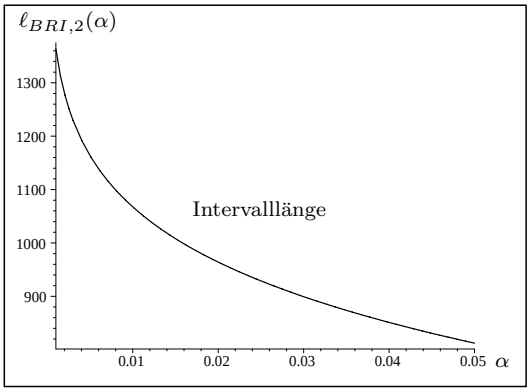
mit einer Intervalllänge von 102 m^3. Diese relativ kurze Intervalllänge (im Verhältnis zum Mittelwert von 1.036 m^3) ist ein sehr gutes Ergebnis und ist natürlich auf den großen Stichprobenumfang zurückzuführen.

Deutlich schwieriger ist da der Fall der Mehrfamilienhäuser (Klasse K_2). Hier haben wir nicht nur eine verhältnismäßig große Varianz, sondern auch einen deutlich geringeren Stichprobenumfang von $n_{BRI,2} = 61$. Das hat zum einen zur Folge, dass wir beim Niveau der Konfidenzintervalle nicht mehr von „genau“ $1 - \alpha$ sprechen wollen, sondern nur noch von „in etwa“ $1 - \alpha$. Der Stichprobenumfang ist zwar schon so groß, dass man die asymptotischen Konfidenzintervalle anwenden

kann (man erinnere sich an die Faustregel: „Stichprobenumfang größer 30“), etwas Vorsicht ist dennoch geboten. Andererseits werden wir nur sehr große Intervalle erhalten. Dies verdeutlichen die folgenden Graphiken.



Konfidenzintervalle $C_{BRI,2}(\alpha)$



Intervalllänge

Abbildung 3.7: Konfidenzintervall $C_{BRI,2}(\alpha)$ für $\overline{y}_{BRI,2}$

Wählt man beispielsweise $\alpha = 0,05$, so erhält man das zugehörige Konfidenzintervall

$$C_{BRI,2}(0,05) = [2.804\ m^3,\ 3.616\ m^3]$$

mit einer Intervalllänge von 812 m^3. Der wahre mittlere BRI der Mehrfamilienhäuser Ettlingens liegt also mit einer Wahrscheinlichkeit von „in etwa“ 95% im Intervall $[2.804\ m^3,\ 3616\ m^3]$.
Wählt man dagegen $\alpha = 0,01$, d.h. man möchte eine Überdeckungswahrscheinlichkeit von in etwa 99%, so ergibt sich das Intervall

$$C_{BRI,2}(0,01) = [2.676\ m^3,\ 3.744\ m^3]$$

mit der Länge 1.068 m^3.
In beiden Fällen erhält man eine (im Verhältnis zum geschätzten Mittelwert $\overline{Y}_{BRI,2} = 3.210\ m^3$) große Intervalllänge.
In Abbildung 3.8 werden die 95% Konfidenzintervalle dargestellt. Der Endbereich eines jeden Balkens zeigt den Bereich an, in dem der wahre mittlere BRI mit 95%–iger Sicherheit liegt. Deutlich erkennt man den sehr großen Bereich bei den Mehrfamilienhäusern, der neben der großen Varianz auch auf den geringen Stichprobenumfang zurückzuführen ist. Der Balken für die gesamten Wohngebäude K_{wohn} unterscheidet sich nicht so stark von dem der Einfamilienhäuser K_1, wie das vielleicht zu erwarten gewesen wäre. Dies liegt daran, dass die Einfamilienhäuser mit geschätzten 88,07% (vgl. Abschnitt 3.1.1) einen extrem großen Anteil an K_{wohn} haben.

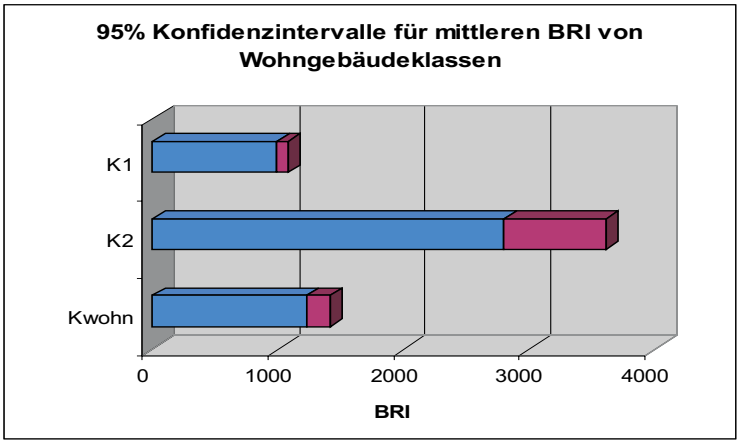


Abbildung 3.8: Konfidenzintervalle zum 95%-Niveau

3.1.3 Vergleich der Ergebnisse mit den Berechnungen aus Abschnitt 2.2.3

Nachdem in den vorangegangenen beiden Abschnitten die Wohngebäude Ettlingens in Hinblick auf den prozentualen Anteil der einzelnen Klassen sowie den mittleren BRI analysiert wurden, sollen jetzt die Ergebnisse mit den Vorüberlegungen, die im Rahmen der Stichprobentheorie in Abschnitt 2.2.3 angestellt wurden, verglichen werden. Dazu sind die Zahlen nochmals in den Tabellen 3.3 und 3.4 aufgelistet.

	K_{wohn}	K_1
Vorinformationen, vgl. Tabelle 2.1		
Anzahl der Beispieldaten	387	27
Mittlerer BRI der Beispieldaten	5.183	1.327
Varianz der Beispieldaten	75.000.000	280.000
Stichprobentheorie, vgl. Abschnitt 2.2.3.3		
Vorgabe Intervalllänge ℓ	1.200	200
Vorgabe Varianzoberschranke	75.000.000	500.000
Geforderter Stichprobenumfang	635	200
Ergebnis der Datenauswertung, vgl. Abschnitt 3.1.2		
Anzahl der Datensätze	599	533
Mittlerer BRI	1.325	1.036
Varianz	1.702.544	325.614
Intervalllänge (95%-.Niveau)	188	87

Tabelle 3.3: Vergleich der Analysedaten aus Abschnitt 3.1.2 zum BRI von Wohngebäuden mit den Zahlen aus Abschnitt 2.2.3 zur Stichprobentheorie

	K_1
Stichprobentheorie, vgl. Abschnitt 2.2.3.3	
Vorgabe Intervalllänge ℓ	0,08=8%
Vorgabe Varianzoberschranke	0,25
Geforderter Stichprobenumfang	520
Ergebnis der Datenauswertung, vgl. Abschnitt 3.1.1	
Anzahl der Datensätze	612
Prozentualer Anteil an K_{wohn}	0,8807=88,07%
Varianz	0,1052
Intervalllänge (95%-.Niveau)	0,046=4,6%

Tabelle 3.4: Vergleich der Analysedaten aus Abschnitt 3.1.1 zum prozentualen Anteil der Klasse K_1 an K_{wohn} mit den Zahlen aus Abschnitt 2.2.3 zur Stichprobentheorie

Es sind mehrere Punkte anzumerken:

1. Am Auffälligsten ist die große Diskrepanz zwischen dem mittleren BRI der Klasse K_{wohn} der Beispieldaten (5.183 m^3) und dem entsprechenden Schätzwert für Ettlingen (1.325 m^3), die in Tabelle 3.3 aufgelistet sind. Für die Vorüberlegungen zur Stichprobentheorie wurden alte Daten herangezogen, deren Zusammensetzung (EFH, MFH, HH) und Größe nicht repräsentativ für Ettlingen waren. Daher muss eindringlich vor Schätzwerten für den mittleren BRI gewarnt werden, die nicht auf einer *zufälligen Stichprobe* aus der Grundgesamtheit beruhen. Auch die Ergebnisse für Ettlingen können nicht für andere Gebäudebestände übernommen werden.
   Ein direkter Vergleich der tatsächlichen Intervallänge (für das 95%–Niveau) mit der Stichprobenvorgabe ℓ ist auf Grund der unterschiedlichen BRI-Werte nicht sinnvoll. Daher betrachtet man die Quotienten aus Intervalllänge und BRI-Mittelwerten. Für Ettlingen ergibt sich dabei mit $\frac{188}{1.325} = 14,19\%$ ein deutlich kleinerer Wert als die Vorgaben aus der Stichprobentheorie, wo der Quotient $\frac{1.200}{5.183} = 23,36\%$ beträgt, was auf die kleinere Varianz der Daten Ettlingens zurückzuführen ist.

2. Etwas weniger gravierend ist der Unterschied der mittleren BRI für die Klasse K_1 der Einfamilienhäuser. Während die (nur) 27 Gebäude der Beispieldaten einen mittleren BRI von 1.327 m^3 ergaben, erhielt man für Ettlingen einen Schätzwert von 1.036 m^3, vgl. Tabelle 3.3. Auch hier kann weder auf Grundlage der Beispieldaten noch auf der der Daten Ettlingens eine statistisch vertretbare Prognose für andere Gebäudebestände getroffen werden.
   Die etwas größere Stichprobenvarianz der EFH Ettlingens (im Vergleich zur

Varianz der Beispieldaten) stellte auf Grund der großzügig vorgegebenen Varianzoberschranke von 500.000 kein Problem dar, so dass (auch auf Grund des großen Stichprobenumfangs von 533) die tatsächliche Intervalllänge von 87 deutlich kleiner ist als die Vorgabe von 200.

3. Betrachtet man den prozentualen Anteil der Klasse K_1 an K_{wohn}, so ist auch hier die tatsächliche Intervalllänge von $4,6\%$ deutlich kleiner als die Vorgabe von 8%, s. Tabelle 3.4. Dies ist vor allem auf die im Vergleich zur Varianzoberschranke geringe Stichprobenvarianz von 0,1052 zurückzuführen.

Zusammenfassend ist zu sagen, dass die vorgegebenen Intervalllängen der Konfidenzintervalle klar unterschritten werden konnten. Sehr deutlich ist zu erkennen, dass Rückschlüsse von *nicht rein zufällig ausgewählten* Beispieldaten auf eine „fremde“ Grundgesamtheit zu völlig falschen Ergebnissen führen würden. Um Ergebnisse für andere Städte oder Bundesländer zu erhalten, muss man den in der vorliegenden Arbeit aufgezeigten Weg neu durchlaufen. Die Ergebnisse für die Stadt Ettlingen sind *nicht übertragbar* - wohl aber die Methodik.

3.1.4 Der prozentuale Anteil der Klassen $K_1 - K_3$ am BRI von Wohngebäuden

Nachdem in Abschnitt 3.1.1 der prozentualen Anteil der Klassen $K_1 - K_3$ an der Gesamtzahl der Wohngebäude untersucht wurde und nachdem in Abschnitt 3.1.2 auch der BRI innerhalb der Klassen K_1 und K_2 analysiert wurde, wird sich dieser Abschnitt mit dem prozentualen Anteil von $K_1 - K_3$ am gesamten BRI der Wohngebäude beschäftigen. Im Folgenden sind also die Variablen

$$p_{BRI,i} = \frac{\text{Summe der BRIe der Gebäude in Klasse } K_i}{\text{Summe der BRIe aller Wohngebäude}}, \qquad i = 1, 2, 3,$$

von Interesse. Um Schätzungen und Vertrauensbereiche für die unbekannten Prozentsätze $p_{BRI,i}$ zu errechnen, werden die Resultate aus Abschnitt A.1.2, v.a. Beispiel A.1.6 in Anhang A.1 angewandt. Man beachte, dass die Stichprobe den Umfang von $n_{BRI,wohn} = 599$ Gebäuden mit bekanntem BRI enthielt. In Tabelle 3.5 sind die Schätzungen $\hat{p}_{BRI,i}$ sowie Konfidenzintervalle zu den Niveaus 95% und 99% für die unbekannten $p_{BRI,i}$ angegeben.

Abbildung 3.9 veranschaulicht die Resultate von Tabelle 3.5. Die Konfidenzintervalle für den BRI-Anteil sind deutlich größer als die für den Absolut-Anteil, vgl. Abbildung 3.4.

Interessant ist auch der Vergleich der (Schätzwerte der) mittleren BRI-Anteile mit den mittleren Absolut-Anteilen aus Abschnitt 3.1.1. Natürlich ist der BRI-Anteil

Klasse	$\hat{p}_{BRI,i}$	Konfidenzintervall (95%)	Konfidenzintervall (99%)
K_1	69,61%	[63,70%, 75,52%]	[61,84%, 77,38%]
K_2	24,68%	[19,51%, 29,84%]	[17,89%, 31,47%]
K_3	5,71%	[1,00%, 10,42%]	[0%, 11,90%]

Tabelle 3.5: Prozentuale Anteile von $K_1 - K_3$ am BRI der EFH

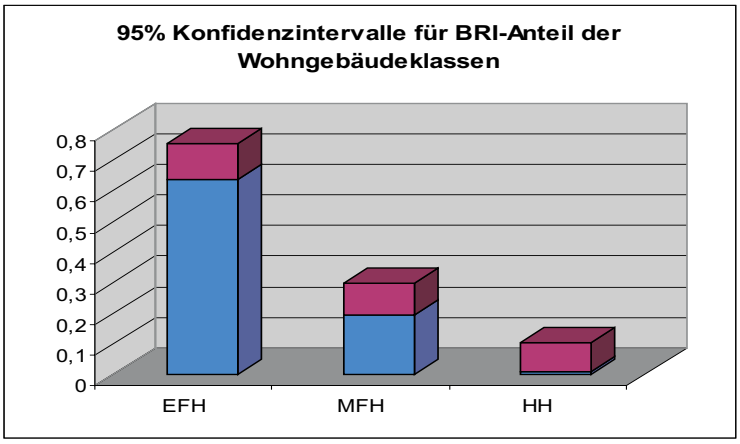


Abbildung 3.9: Konfidenzintervalle zum 95%-Niveau

von Einfamilienhäusern deutlich geringer als deren Absolut-Anteil. Dennoch machen die EFH auch bezüglich Ihres Volumens den mit Abstand größten Teil der Wohngebäude aus. Abbildung 3.10 verdeutlicht die Unterschiede grafisch.

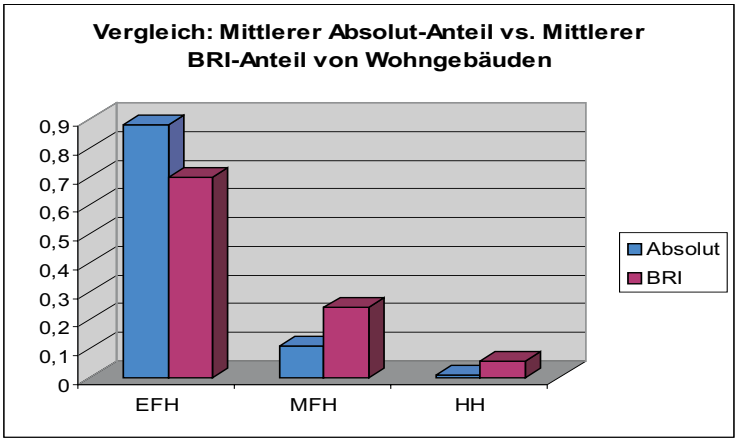


Abbildung 3.10: Vergleich der Schätzwerte für den mittleren Absolut-Anteil mit dem mittleren BRI-Anteil der einzelnen Wohngebäudeklassen

3.1.5 Die Verteilung des BRIes von Einfamilienhäusern

Die bisherigen Untersuchungen haben sich auf den mittleren Bruttorauminhalt von Gebäuden der einzelnen Nutzungsklassen konzentriert. Ausgehend von den Daten wurden Schätzwerte für diesen Mittelwert sowie für die Varianz des Merkmals „BRI“ bestimmt und darauf aufbauend Intervalle konstruiert, von denen man mit großer Sicherheit annehmen kann, dass sie den wahren aber unbekannten mittleren BRI enthalten. In diesem Abschnitt wird der Bruttorauminhalt für Einfamilienhäuser (Klasse K_1) mit einem anderen Modellansatz untersucht. Da die Klasse der Einfamilienhäuser sehr homogen ist, kann der im Folgenden beschriebene Ansatz nicht ohne Weiteres auf andere Nutzungsklassen übertragen werden. Andererseits stellen die Einfamilienhäuser die mit Abstand größte Klasse aller Gebäude dar und rechtfertigen somit eine gesonderte Untersuchung.

Um die gegebenen Daten (die Bruttorauminhalte der Einfamilienhäuser der vorliegenden Stichprobe) zu beschreiben, soll jetzt davon ausgegangen werden, dass alle 499 Daten unabhängige Realisierungen einer Zufallsvariablen sind. Derartige „idealisierte“ Modellannahmen beschreiben die Realität nie exakt; ist die Beschreibung jedoch hinreichend genau, so kann man trotzdem versuchen, weitergehende Überlegungen auf dem jeweiligen Modell aufzubauen. Im vorliegenden Fall soll also die Wahrscheinlichkeitsverteilung des BRIes von Einfamilienhäusern bestimmt werden. Dazu wird man zunächst ein Histogramm betrachten, das die Anzahl der in fest vorgegebenen Bereichen liegenden Daten in Form von Balkendiagrammen wiedergibt. In Abbildung 3.11 ist dies für die BRIe der Einfamilienhäuser geschehen. Die x–Achse wurde in Schritte von je 100 m^3 unterteilt. Deutlich ist dabei

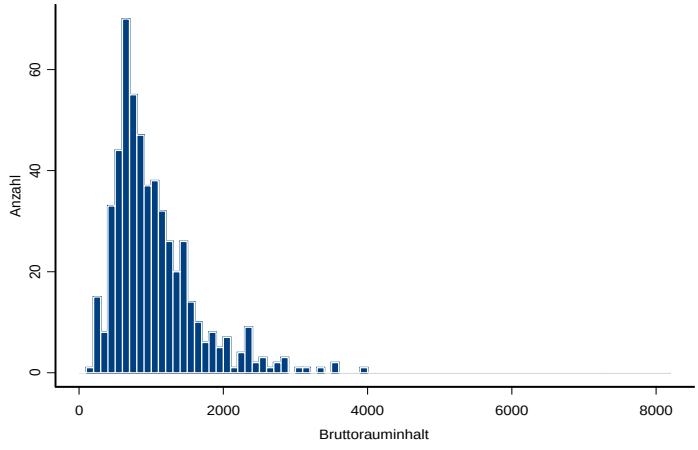


Abbildung 3.11: Histogramm für die Bruttorauminhalte der Klasse K_1

eine Struktur zu erkennen, die der Dichte einer logarithmischen Normalverteilung,

wie sie in Abbildung A.1 auf Seite 71 wiedergegeben ist, ähnelt. Diese Verteilung tritt in vielen Anwendungsbereichen (Ökonomie, Biologie, Ökologie, Geologie, u.v.m.) auf, vgl. Crow und Shimizu (1988). Ein wesentlicher Unterschied muss aber berücksichtigt werden: Während die logarithmische Normalverteilung ihre „Wahrscheinlichkeitsmasse“ im Intervall $(0, \infty)$ besitzt, muss bei der Verteilung des Bruttorauminhaltes davon ausgegangen werden, dass er mit Wahrscheinlichkeit Null kleiner als 100 m^3 ist. Das führt uns zu dem folgenden statistischen Modellansatz: Beschreibt die Zufallsvariable $Y_{BRI,1}$ den Bruttorauminhalt eines Einfamilienhauses, so wollen wir davon ausgehen, dass $Y_{BRI,1} - 100$ logarithmisch normalverteilt ist mit unbekannten Parametern $\mu_{BRI,1}$ und $\tau^2_{BRI,1}$, also

$$Y_{BRI,1} - 100 \sim LN(\mu_{BRI,1}, \tau^2_{BRI,1}), \qquad \mu_{BRI,1} \in \mathbb{R}, \tau^2_{BRI,1} > 0 \text{ unbekannt.}$$

Auf Grund der beobachteten Bruttorauminhalte unserer Stichprobe, die als unabhängige Realisierungen der Zufallsvariablen $Y_{BRI,1}$ aufgefasst werden, kann man nun die unbekannten Parameter schätzen. Für $\mu_{BRI,1}$ und $\tau^2_{BRI,1}$ ergeben sich gemäß Formel (A.9) und (A.10) in Anhang A.2 die Schätzungen

$$\hat{\mu}_{BRI,1} = 6,675288 \qquad \hat{\tau}^2_{BRI,1} = 0,3449612.$$

Zeichnen wir die Dichte der $LN(\hat{\mu}_{BRI,1}, \hat{\tau}^2_{BRI,1})$–Verteilung in obiges Histogramm (entsprechend um 100 m^3 nach links verschoben) ein, so erhalten wir Abbildung 3.12. Hier wird schon rein optisch ersichtlich, dass die verschobenen Daten durch die geschätzte Verteilung „gut“ beschrieben werden. Da jedoch die Qua-

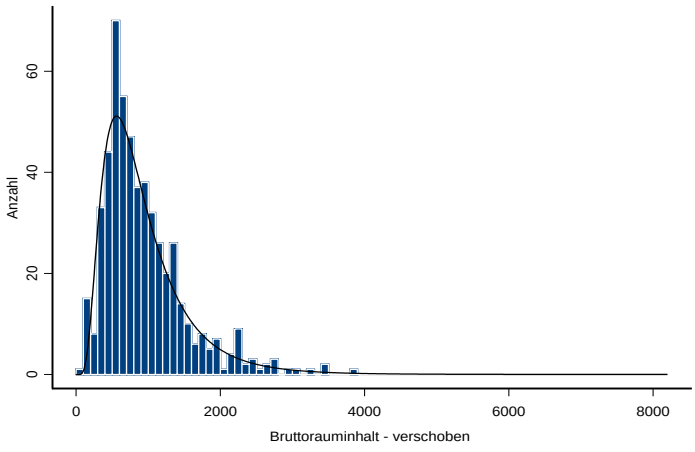


Abbildung 3.12: Histogramm für den BRI der Klasse K_1 mit angepasster Dichte

lität der Anpassung der $LN(\hat{\mu}_{BRI,1}, \hat{\tau}^2_{BRI,1})$–Verteilung an die verschobenen Daten anhand von Histogrammen nur schlecht zu beurteilen ist, müssen weitergehende

Überlegungen diesbezüglich gemacht werden. Ein klassischer Ansatz dazu ist der Vergleich der Verteilungsfunktion der $LN(\hat{\mu}_{BRI,1}, \hat{\tau}^2_{BRI,1})$–Verteilung mit der empirischen Verteilungsfunktion der verschobenen Daten, die beide in Abbildung 3.13 eingezeichnet sind. Da die Funktionen nahezu deckungsgleich sind, kann man

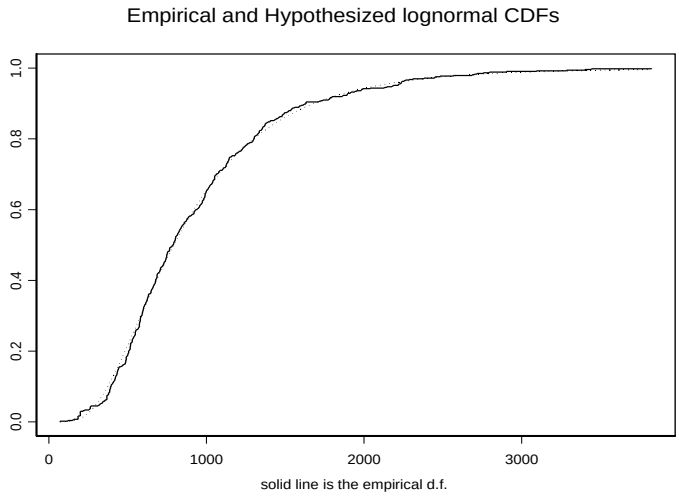


Abbildung 3.13: Verteilungsfunktion von $LN(\hat{\mu}_{BRI,1}, \hat{\tau}^2_{BRI,1})$ und empirische Verteilungsfunktion der Daten

schon anhand des Schaubildes von einer „guten Anpassung“ sprechen. Vergleichen wir die verschobenen Daten mit der $LN(\hat{\mu}_{BRI,1}, \hat{\tau}^2_{BRI,1})$–Verteilung mittels eines Kolmogoroff-Smirnov-Tests, so ergibt sich ein p–Wert von $41,52\%$, d.h. der Test würde die Hypothese, dass die verschobenen Daten $LN(\hat{\mu}_{BRI,1}, \hat{\tau}^2_{BRI,1})$–verteilt sind, erst ab einen Niveau von $41,52\%$ ablehnen (übliche Niveaus sind $5 - 10\%$), ein weiteres Indiz für die Brauchbarkeit der Anpassung. Die Verwendung eines Kolmogoroff-Smirnov Tests, bei dem die Hypothese aus der „geschätzten“ logarithmischen Normalverteilung besteht, ist natürlich nur ein sehr vereinfachtes Verfahren. Es gibt wesentlich bessere Testverfahren, wo die Hypothese aus der ganzen Klasse der logarithmischen Normalverteilungen besteht. Da im Folgenden jedoch nicht mit der Anpassung weitergerechnet wird, ist der Kolmogoroff ausreichend, um eine grobe Vorstellung von deren Güte zu erhalten.
Nach Anhang A.2 berechnet sich für $Y_{BRI,1}$ mit $Y_{BRI,1} - 100 \sim LN(\hat{\mu}_{BRI,1}, \hat{\tau}^2_{BRI,1})$ der Erwartungswert zu $\mathrm{E}(Y_{BRI,1}) = 1042\ m^3$ und die Varianz zu $\mathrm{Var}(Y_{BRI,1}) = 365.363$. Beide Werte liegen nahe an dem empirischen Mittelwert der Daten $\overline{Y}_{BRI,1} = 1.036\ m^3$ bzw. der empirischen Varianz $\hat{\sigma}^2_{BRI,1} = 325.614$.

3.1.6 Altersklassen von Einfamilienhäusern

Neben der Einteilung von Gebäuden in die Nutzungsklassen $K_1, \ldots, K_8$ ist auch eine die Unterteilung in Altersklassen bei der Analyse von Gebäudebeständen sinnvoll. Ziel bei der Klasseneinteilung ist es, möglichst homogene Gruppen zu schaffen in Hinblick auf BRI, Materialien oder ähnlich Charakteristika. Es werden die folgenden Perioden unterschieden:

$A1$: Gebäude bis 1870

$A2$: Gebäude von 1871 bis 1914

$A3$: Gebäude von 1915 bis 1945

$A4$: Gebäude von 1946 bis 1965

$A5$: Gebäude von 1966 bis 1978

$A6$: Gebäude ab 1979.

Die Einteilung der Altersklassen richtet sich überwiegend nach baukonstruktiven Perioden. Dies ist sowohl in Hinblick auf die Verwendung bestimmter Baustoffe wie auch auf Energieverbrauch hervorgerufen durch Wärmeschutzverordnung zu verstehen. Bestimmte „Bautypen“ werden ebenfalls in einzelne Altersklassen unterschieden.

So ergab z.B. die Stichprobe von $n_{wohn} = 612$ Wohngebäuden nach Tabelle 3.1 insgesamt $n_1 = 539$ Einfamilienhäuser, die sich wie in Tabelle 3.6 aufgeschlüsselt auf die Altersklassen verteilen.

	Gesamtzahl	A_1	A_2	A_3	A_4	A_5	A_6
Anzahl EFH	$n_1 = 539$	30	67	100	192	63	87

Tabelle 3.6: Einfamilienhäuser der Stichprobe in den Altersklassen $A_1 - A_6$

Es wird für den gesamten Gebäudebestand durch die Einteilung in Alters- sowie Nutzungsklassen eine sogenannte Alters-Nutzungs-Matrix definiert. Man interessiert sich dabei für den prozentualen Anteil

$$p_{A_i,K_j} = \frac{\text{Anzahl der Gebäude in Altersklasse } A_i \text{ und Nutzungsklasse } K_j}{\text{Gesamtzahl aller Gebäude}},$$

$i = 1, \ldots, 6$, $j = 1, \ldots, 8$ jeder Alters-Nutzungsklasse (z.B. $K_1 \cap A_5$) am Gesamtbestand. Für die unbekannten Prozentsätze der zur Nutzungsklasse K_1 der Einfamilienhäuser gehörigen Alters-Nutzungsklassen sind in Tabelle 3.7 Schätzwerte

Altersklasse	$\hat{p}_{A_i,K_1}$	Konfidenzintervall (95%)	Konfidenzintervall (99%)
A_1	2,86%	[1,98%, 3,77%]	[1,69%, 4,06%]
A_2	6,42%	[5,12%, 7,72%]	[4,71%, 8,13%]
A_3	9,58%	[8,05%, 11,12%]	[7,56%, 11,61%]
A_4	18,40%	[16,47%, 20,33%]	[15,86%, 20,94%]
A_5	6,04%	[4,77%, 7,30%]	[4,38%, 7,70%]
A_6	8,34%	[6,89%, 9,79%]	[6,43%, 10,25%]

Tabelle 3.7: Prozentuale Anteile der Altersklassen am gesamten Gebäudebestand

$\hat{p}_{A_i,K_1}$ sowie Konfidenzintervalle zu den Niveaus 95% und 99% angegeben.

Dabei wurden zunächst mit denselben Methoden wie in Abschnitt 3.1.1 Schätzer für die prozentualen Anteile der Alters-Nutzungsklassen an der Gesamtzahl $N_{wohn} = 3084$ aller Wohngebäude berechnet. Danach wurden diese Prozentsätze mit dem Quotienten

$$\frac{\text{Anzahl aller Wohngebäude}}{\text{Gesamtzahl aller Gebäude}} = \frac{3084}{5258}$$

multipliziert, wobei sich die Gesamtzahl $N_{gesamt} = 5258$ aller Gebäude Ettlingens aus Tabelle 2.3 ergibt.

3.1.7 Die BRIe in den Altersklassen der Einfamilienhäuser

Analog zu den Untersuchungen des Bruttorauminhaltes der gesamten Wohngebäude bzw. der Wohngebäude der Nutzungsklassen K_1 und K_2 in Abschnitt 3.1.2 werden hier die BRIe innerhalb der verschiedenen Altersklassen von Einfamilienhäusern (Klasse K_1) analysiert. Auch hier waren nicht von allen Einfamilienhäusern der Stichprobe die BRIe bekannt, was der Vergleich der Tabellen 3.6 und 3.8 verdeutlicht.

	Gesamtzahl	A_1	A_2	A_3	A_4	A_5	A_6
Anzahl EFH	$n_1 = 533$	30	67	99	190	63	84

Tabelle 3.8: EFH mit BRI der Stichprobe in den Altersklassen $A_1 - A_6$

Die Ergebnisse der Analyse sind in Tabelle 3.9 angeführt.

Abbildung 3.14 veranschaulicht die Ergebnisse von Tabelle 3.9 graphisch. Jeder Horizontalbalken steht dabei für eine Altersklasse. Der dunklere Endbereich jedes

Altersklasse	Mittlerer BRI	Varianz	Konfidenzintervall (95%)
A_1	944 m^3	314.164	[762 m^3, 1126 m^3]
A_2	1304 m^3	495.658	[1151 m^3, 1457 m^3]
A_3	929 m^3	319.483	[828 m^3, 1030 m^3]
A_4	1035 m^3	280.136	[967 m^3, 1103 m^3]
A_5	1136 m^3	367.212	[1000 m^3, 1272 m^3]
A_6	909 m^3	175.691	[827 m^3, 991 m^3]

Tabelle 3.9: BRIe in den Altersklassen der Einfamilienhäuser

Balkens stellt das 95% Konfindenzintervall dar, also den Bereich, in dem der wahre, unbekannte mittlere BRI der Altersklasse mit 95%–iger Sicherheit liegt. Deutlich sind die Größenunterschiede in Bezug auf den BRI in den einzelnen Altersklassen zu erkennen.

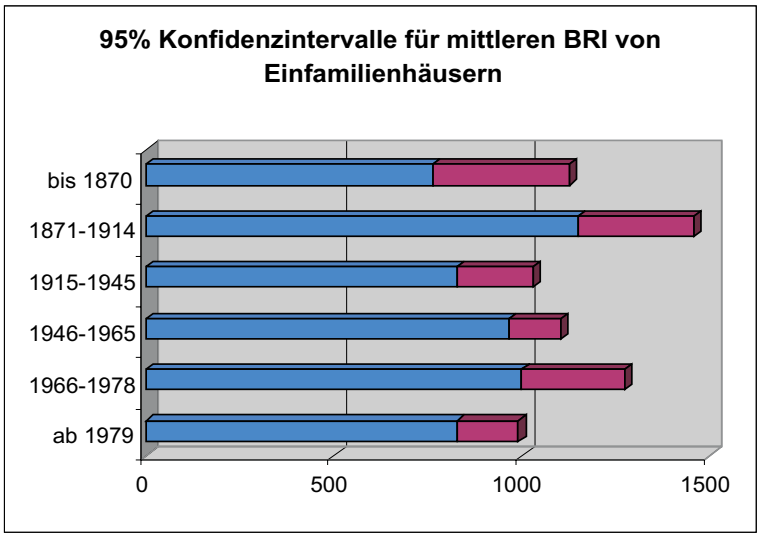


Abbildung 3.14: Konfidenzintervalle für mittleren BRI zum 95%–Niveau

3.2 Bestandsanalyse der nichtlandwirtschaftlichen Betriebsgebäude

Exemplarisch für die Bestände an Nichtwohngebäuden wird die Klasse K_7 der nichtlandwirtschaftlichen Betriebsgebäude analysiert. Im Zentrum stehen dabei wieder der Bruttorauminhalt sowie das Baujahr. Nach Tabelle 2.3 war aus den insgesamt $N_7 = 1186$ Gebäuden der Klasse K_7 eine Stichprobe von 368 zufällig

ermittelten Vertretern vorgesehen. Bei den Recherchen in den Akten der Feuerversicherung sowie des Bauordnungsamtes wurden allerdings alle auf dem jeweiligen Grundstück befindlichen Gebäude miterhoben, so dass die tatsächliche Anzahl erhobener Gebäude $n_1 = 779$ beträgt. Davon konnte bei $n_{BRI,7} = 766$ Gebäuden der Bruttorauminhalt ermittelt werden, von $n_{BJ,7} = 775$ das Baujahr und von $n_{BRI,BJ,7} = 765$ sowohl der BRI als auch das Baujahr. Ein Überblick über diese Anzahlen an Gebäuden gibt Tabelle 3.10.

Gesamtzahl der Gebäude Ettlingens	$N_7 = 1186$
Vorgesehener Stichprobenumfang	368
Tatsächlich erhobene Gebäude	$n_1 = 779$
Erhobene Gebäude mit Bruttorauminhalt	$n_{BRI,7} = 766$
Erhobene Gebäude mit Baujahr	$n_{BJ,7} = 775$
Erhobene Gebäude mit BRI und Baujahr	$n_{BRI,BJ,7} = 765$

Tabelle 3.10: Anzahl an Gebäuden der Klasse K_7

3.2.1 Analyse des Bruttorauminhaltes

Ziel der Stichprobenermittlung war es, die mittleren Bruttorauminhalte in den einzelnen Gebäudeklassen möglichst genau bestimmen zu können. Als Schätzung für den unbekannten mittleren Bruttorauminhalt $\overline{y}_{BRI,7}$ der $N_7 = 1186$ Gebäude der Klasse K_7 wird (analog zum Vorgehen für die Wohngebäude) der Mittelwert $\overline{Y}_{BRI,7}$ der $n_{BRI,7} = 766$ erhobenen nichtlandwirtschaftlichen Betriebsgebäude mit BRI genommen. Dieser berechnet sich aus den Daten zu

$$\overline{Y}_{BRI,7} = 4118 \ [m^3].$$

Als Stichprobenvarianz ergibt sich

$$\hat{\sigma}^2_{BRI,7} = 58.292.033.$$

Bei der Ermittlung einer geeigneten Stichprobe wurden Vorinformationen (Daten zu nichtlandwirtschaftlichen Betriebsgebäuden aus älteren Untersuchungen, die nicht von Ettlingen stammen, vgl. Tabelle 2.1) herangezogen. Dieser Datensatz umfasste 197 Gebäude und ergab einen mittleren BRI von 12.118 $[m^3]$. Auch hier zeigt sich deutlich, dass Daten aus älteren Untersuchungen, die nicht rein zufällig erhoben wurden, keinerlei Rückschlüsse auf andere Gesamtheiten (wie die Stadt Ettlingen) zulassen. Der für Ettlingen ermittelte mittlere BRI beträgt kaum mehr als ein Drittel vom mittleren BRI der alten Daten. Hauptgrund hierfür ist der Folgende: Während man bei einer gezielten Untersuchung von Gebäuden vor allem Wert auf die Analyse von größeren und aus architektonisch und energetischer

Sicht interessanten Gebäuden legt, wird bei einer rein zufälligen Auswahl auch die Vielzahl kleiner nahezu identischer Gebäude wie kleinen Lagern, Schöpfen und Umformern berücksichtigt. Die Stichprobenvarianz $\hat{\sigma}^2_{BRI,7}$ ist infolge dessen mit ca. 58 Mio. ebenfalls deutlich kleiner als die 363 Mio. der alten Daten. Dies sowie der große Stichprobenumfang (im Bezug zur Gesamtzahl N_7) führt im Folgenden zu relativ kleinen Konfidenzintervallen, d.h. guten Prognosen für den unbekannten Wert $\overline{y}_{BRI,7}$, vgl. Tabelle 3.11.

Niveau	Konfidenzintervall	Intervalllänge
99%	$[3.695\ m^3,\ 4.541\ m^3]$	846
98%	$[3.736\ m^3,\ 4.500\ m^3]$	764
97%	$[3.762\ m^3,\ 4.474\ m^3]$	712
96%	$[3.781\ m^3,\ 4.455\ m^3]$	674
95%	$[3.796\ m^3,\ 4.440\ m^3]$	644

Tabelle 3.11: Konfidenzintervalle für den mittleren BRI der Klasse K_7

Die Ergebnisse von Tabelle 3.11 sind in Abbildung 3.15 verdeutlicht.

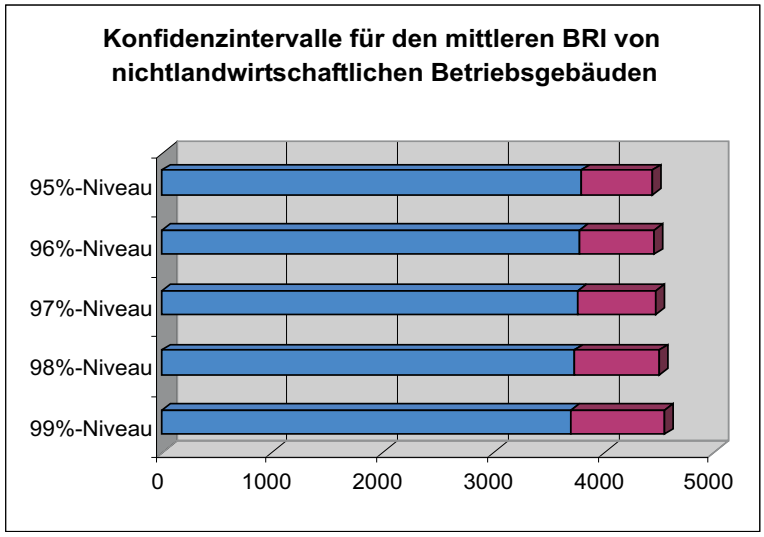


Abbildung 3.15: Konfidenzintervalle für den mittleren BRI der Klassen K_7 zu verschiedenen Niveaus

3.2.2 Altersklassen von nichtlandwirtschaftlichen Betriebsgebäuden

Das Histogramm 3.16 spiegelt die Altersverteilung der Stichprobe wieder.

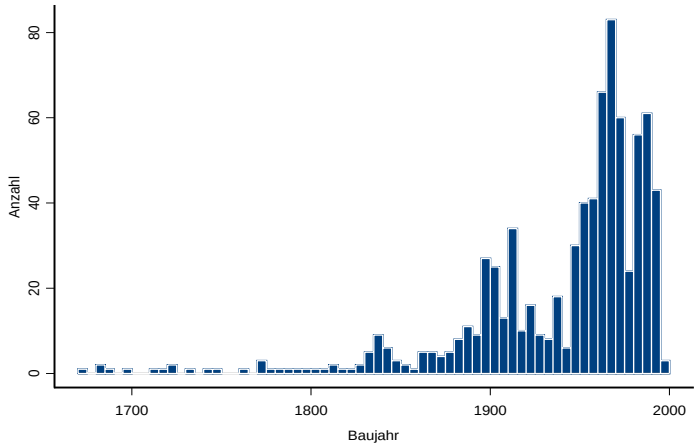


Abbildung 3.16: Histogramm für das Baujahr der Klasse K_7

Analog zum Abschnitt 3.1.6 wird nun die Verteilung der nichtlandwirtschaftlichen Betriebsgebäude auf die Altersklassen $A_1 - A_6$ untersucht. Bei insgesamt $n_{BJ,7} = 775$ der erhobenen Gebäude der Klasse K_7 konnte das Baujahr festgestellt werden. Diese teilen sich, wie in Tabelle 3.12 angegeben, auf die Altersklassen auf.

	Gesamtzahl	A_1	A_2	A_3	A_4	A_5	A_6
Anzahl	$n_{BJ,7} = 775$	65	135	68	177	158	172

Tabelle 3.12: Nichtlandwirtschaftliche Betriebsgebäude der Stichprobe in den Altersklassen $A_1 - A_6$

Zunächst werden für die prozentualen Anteile $q_{A_i,7}$ der einzelnen Altersklassen am Bestand von K_7, d.h.

$$q_{A_i} = \frac{\text{Anzahl der Gebäude in Altersklasse } A_i \text{ und Nutzungsklasse } K_7}{\text{Gesamtzahl aller Gebäude in } K_7},$$

$i = 1, \ldots, 6$, die kanonischen Schätzer $\overline{q}_{A_i,7}$ sowie Konfidenzintervalle bestimmt. Analog zu den Berechnungen in Abschnitt 3.1.1 erhält man die in Tabelle 3.13 wiedergegebenen Ergebnisse.

Altersklasse	$\overline{q}_{A_i,7}$	Konfidenzintervall (95%)	Konfidenzintervall (99%)
A_1	8,39%	[7,24%, 9,54%]	[6,88%, 9,90%]
A_2	17,42%	[15,85%, 18,99%]	[15,35%, 19,49%]
A_3	8,77%	[7,60%, 9,95%]	[7,23%, 10,32%]
A_4	22,84%	[21,10%, 24,58%]	[20,55%, 25,13%]
A_5	20,39%	[18,72%, 22,06%]	[18,19%, 22,58%]
A_6	22,19%	[20,47%, 23,92%]	[19,93%, 24,46%]

Tabelle 3.13: Prozentuale Anteile der Altersklassen von K_7 am Bestand der Nutzungsklasse K_7

In Hinblick auf die in Abschnitt 3.1.6 angesprochene Alters-Nutzungs-Matrix interessieren aber die prozentualen Anteile der Alters-Nutzungsklassen am Gesamtbestand aller Gebäude, also die Prozentsätze

$$p_{A_i,K_7} = \frac{\text{Anzahl der Gebäude in Altersklasse } A_i \text{ und Nutzungsklasse } K_7}{\text{Gesamtzahl aller Gebäude}}.$$

Schätzer $\hat{p}_{A_i,K_7}$ und Konfidenzintervalle ergeben sich, wenn man die Werte aus Tabelle 3.13 mit dem Quotienten

$$\frac{\text{Anzahl } N_7 \text{ aller Gebäude der Klasse } K_7}{\text{Gesamtzahl aller Gebäude}} = \frac{1186}{5258}$$

multipliziert, wobei sich die Gesamtzahl $N_{gesamt} = 5258$ aller Gebäude Ettlingens aus Tabelle 2.3 ergibt. Die Zahlen sind in Tabelle 3.14 eingetragen.

Altersklasse	$\hat{p}_{A_i,K_7}$	Konfidenzintervall (95%)	Konfidenzintervall (99%)
A_1	1,89%	[1,63%, 2,15%]	[1,55%, 2,23%]
A_2	3,93%	[3,57%, 4,28%]	[3,46%, 4,40%]
A_3	1,98%	[1,71%, 2,24%]	[1,63%, 2,33%]
A_4	5,15%	[4,76%, 5,54%]	[4,64%, 5,67%]
A_5	4,60%	[4,22%, 4,98%]	[4,10%, 5,09%]
A_6	5,01%	[4,62%, 5,39%]	[4,50%, 5,52%]

Tabelle 3.14: Prozentuale Anteile der Altersklassen von K_7 am gesamten Gebäudebestand Ettlingens

3.2.3 Die BRIe in den Altersklassen der nichtlandwirtschaftlichen Betriebsgebäude

Während in Abschnitt 3.2.1 der mittlere Bruttorauminhalt alle nichtlandwirtschaftlichen Betriebsgebäude analysiert wurde, werden hier analoge Überlegungen für die verschiedenen Altersklassen von K_7 durchgeführt. Die Gesamtzahl $n_{BRI,7}$ der Gebäude, für die der BRI bekannt ist, vgl. Tabelle 3.10, verteilt sich wie in Tabelle 3.15 aufgeführt auf die verschiedenen Altersklassen.

	Gesamtzahl	A_1	A_2	A_3	A_4	A_5	A_6
Anzahl	$n_{BRI,BJ,7} = 765$	65	134	62	176	157	171

Tabelle 3.15: Nichtlandwirtschaftliche Betriebsgebäude mit bekannter BRI der Stichprobe in den Altersklassen $A_1 - A_6$

Die Ergebnisse der Analyse sind in Tabelle 3.16 angeführt.

Altersklasse	Mittlerer BRI	Varianz	Konfidenzintervall (95%)
A_1	3.310 m^3	33.929.074	[2.466 m^3, 4.154 m^3]
A_2	2.729 m^3	16.406.231	[2.383 m^3, 3.201 m^3]
A_3	1.622 m^3	3.739.595	[1.355 m^3, 1.909 m^3]
A_4	4.663 m^3	84.350.334	[3.855 m^3, 5.471 m^3]
A_5	4.277 m^3	74.370.305	[3.473 m^3, 5.081 m^3]
A_6	5.683 m^3	71.845.199	[4.926 m^3, 6.440 m^3]

Tabelle 3.16: BRIe in den Altersklassen von K_7

In Abbildung 3.17 sind die 95%–Konfidenzintervalle visualisiert zusammen mit dem entsprechenden Intervall für die gesamte Klasse K_7.

3.3 Die Alters-Nutzungs-Matrix

Zum Abschluss dieses Kapitels wird die nahezu vollständige Alters-Nutzungs-Matrix angeführt. Nahezu vollständig bedeutet hierbei, dass die Matrix nur für die Nutzungsklassen K_1, K_2, K_4, K_6 und K_7 erstellt wird. Die übrigen Nutzungsklassen K_3, K_5 und K_8 machen lediglich einen geschätzten Anteil von 1,64 % aus, so dass sie in der vorliegenden Statistik getrost vernachlässigt werden können. Wesentlich für das Verständnis der Alters-Nutzungs-Matrix ist jedoch, dass alle eingetragenen Prozentsätze nur Schätzwerte für die tatsächlichen unbekannten Werte

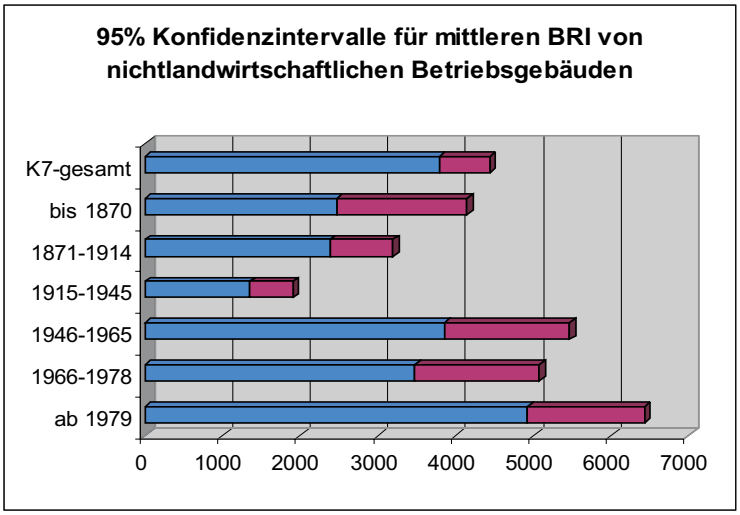


Abbildung 3.17: Konfidenzintervalle für mittleren BRI zum 95%–Niveau

sind. Die Matrix sagt selbst nichts über die Genauigkeit dieser Schätzungen aus. Um hierüber Auskunft zu erhalten müssen die in den Tabellen 3.7 und 3.14 für die Klassen K_1 und K_7 angeführten Konfidenzintervalle betrachtet werden. Solche Berechnungen fehlen aber noch für die Klassen K_2, K_4 und K_6 und werden im folgenden Abschnitt noch vor der eigentlichen Matrix angegeben.

3.3.1 Die Altersklassen der Nutzungsklassen K_2, K_4 und K_6

Zu den Klassen K_2, K_4 und K_6 liegen auf Grund der Stichprobennahme $n_2 = 68$, $n_4 = 60$ und $n_6 = 45$ Datensätze vor. Für die Klassen K_4 und K_6 wurden dabei nicht separat zufällige Stichproben gezogen, sondern die Daten ergaben sich als „Abfallprodukte“ aus der Datenerhebung für die Klasse K_7, wo jeweils alle auf den betreffenden Grundstücken befindlichen Gebäude miterhoben wurden. Es sei jedoch auch in diesem Fall für K_4 und K_6 vorausgesetzt, dass die erhobenen Daten einer rein zufälligen Auswahl unterliegen. Aufgrund des geringen Datenumfangs ergeben sich im Folgenden relativ große Konfidenzintervalle, so dass auch nur die 95% Intervalle und keine Intervalle zu höheren Niveaus berechnet werden - deutlich erkennt man hieran, wie sehr die Genauigkeit der Schätzungen von der Anzahl der Daten abhängt. Die folgenden Tabellen 3.17, 3.18 und 3.19 ergeben sich völlig analog zu den Ausführungen in den Abschnitten 3.1.6 bzw. 3.2.2. Sie bleiben deshalb hier ohne weitere Erläuterungen bzw. Herleitungen.

Altersklasse	$\hat{p}_{A_i,K_2}$	Konfidenzintervall (95%)
A_1	0%	[0%, 0%]
A_2	0,29%	[0%, 0,58%]
A_3	0,38%	[0,05%, 0,72%]
A_4	2,78%	[1,90%, 3,66%]
A_5	1,25%	[0,65%, 1,85%]
A_6	1,82%	[1,10%, 2,54%]

Tabelle 3.17: Prozentuale Anteile der Altersklassen von K_2 am gesamten Gebäudebestand Ettlingens

Altersklasse	$\hat{p}_{A_i,K_4}$	Konfidenzintervall (95%)
A_1	0,13%	[0%, 0,27%]
A_2	0,31%	[0,09%, 0,53%]
A_3	0,19%	[0,01%, 0,36%]
A_4	0,75%	[0,43%, 1,07%]
A_5	1,00%	[0,65%, 1,36%]
A_6	2,01%	[1,61%, 2,41%]

Tabelle 3.18: Prozentuale Anteile der Altersklassen von K_4 am gesamten Gebäudebestand Ettlingens

Altersklasse	$\hat{p}_{A_i,K_6}$	Konfidenzintervall (95%)
A_1	4,12%	[2,39%, 5,86%]
A_2	1,77%	[0,49%, 3,04%]
A_3	3,83%	[2,13%, 5,53%]
A_4	2,65%	[1,15%, 4,15%]
A_5	0,29%	[0%, 0,85%]
A_6	0,59%	[0%, 1,36%]

Tabelle 3.19: Prozentuale Anteile der Altersklassen von K_6 am gesamten Gebäudebestand Ettlingens

3.3.2 Die Matrix

Nachdem alle Schätzwerte $\hat{p}_{A_i,K_j}$ für die prozentualen Anteile der Alters-Nutzungsklassen $A_i \cap K_j$, $i = 1, \ldots, 6$, $j = 1, 2, 4, 6, 7$, berechnet und tabellarisch aufgeführt sind, können diese Werte in eine Matrix eingetragen werden. Dies soll hier in der visualisierten Form von Abbildung 3.18 geschehen. Es sei nochmals wiederholt, dass die nicht in der Abbildung enthaltenen Klassen insgesamt einen vernachlässigbaren geschätzen Anteil von 1,64 % ausmachen.

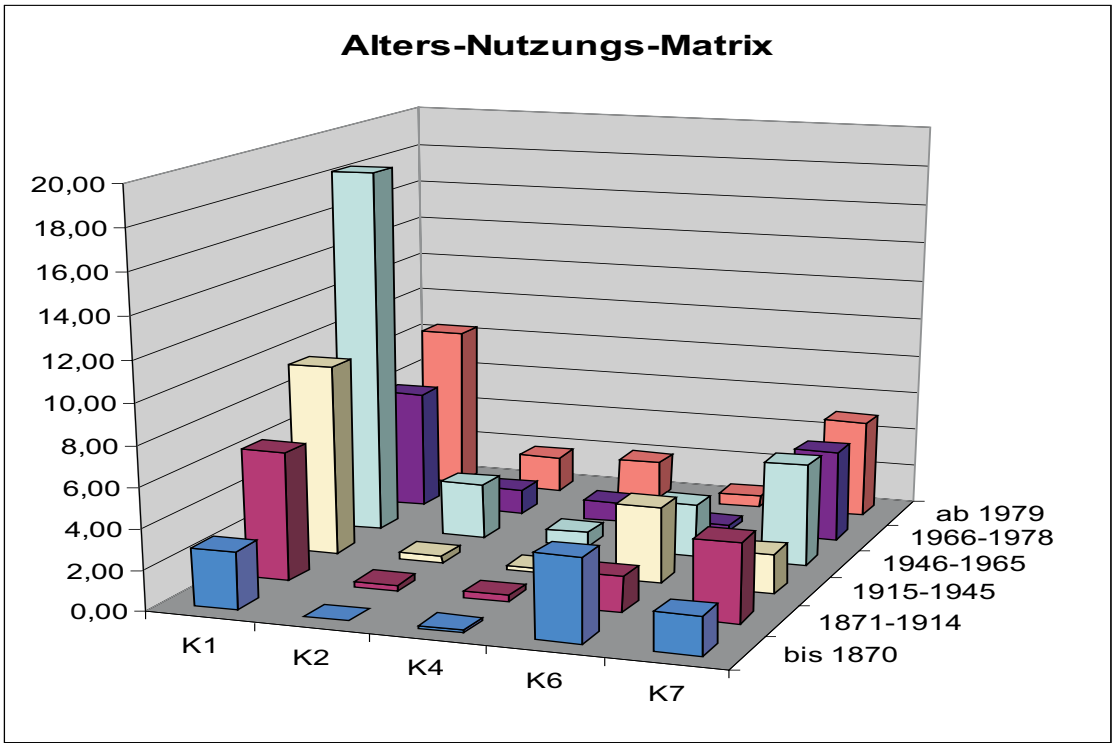


Abbildung 3.18: Alters-Nutzen-Matrix für die Klassen K_1, K_2, K_4, K_6, K_7

Kapitel 4

Dynamik von Gebäuden

Kapitel 5

Prognosen zu Bestand und Materialfluss

Anhang A

Mathematischer Anhang

A.1 Stichproben

Die folgenden Ausführungen orientieren sich an Stenger (1986).

A.1.1 Mittelwertschätzung eines Merkmals

Gegeben sei eine **Erhebungsgesamtheit** von N Objekten (z.B. Gebäude)

$$g_1, \ldots, g_N,$$

die auch **Merkmalsträger** genannt werden. Im Folgenden untersuchen wir ein **deterministisches** und **bestimmbares Merkmal**, eine reelle Zahl (z.B. die Bruttogeschossfläche eines Gebäudes). Wir gehen davon aus, dass wir dieses Merkmal ohne Fehler bestimmen können. Also haben wir eine Anzahl von unbekannten, aber feststellbaren deterministischen Größen

$$y_1, \ldots, y_N,$$

die **Merkmalsausprägungen** genannt werden. Würden wir bei jedem Objekt die Merkmalsausprägung bestimmen, so hätten wir die gesamte Information über den Bestand an Objekten bezüglich dieses Merkmals. Ein solches Vorgehen wird als **Vollerhebung** bezeichnet. Vollerhebungen sind bei großen Erhebungsgesamtheiten mitunter sehr aufwendig, sehr teuer oder gar nicht durchführbar. Letzteres kann z.B. in der Qualitätskontrolle auftreten, wenn die Bestimmung der Merkmalsausprägung das zu untersuchende Objekt zerstört. Aus diesen Gründen will man die Bestimmung bei wenigen **zufällig ausgesuchten** Objekten, einer sogenannten **Stichprobe** oder **Teilerhebung**, durchführen und auf Grund der Ergebnisse Rückschlüsse auf die Erhebungsgesamtheit ziehen.
Diese **Zufälligkeit der Auswahl** wollen wir im Folgenden stochastisch modellieren. Mit Hilfe der Modelle werden wir dann vor der Stichprobenziehung einen

„geeigneten“ Stichprobenumfang festlegen.

Zunächst müssen wir einige Begriffe einführen. Entscheidend für unsere Überlegungen sind

1. das **arithmetische Mittel**
$$\overline{y} := \frac{1}{N}\sum_{i=1}^{N} y_i$$
2. die **Varianz**
$$\sigma^2 := \frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{y})^2 = \frac{1}{N}\sum_{i=1}^{N} y_i^2 - \overline{y}^2$$

aller Merkmalsausprägungen. Da wir die Merkmalsausprägungen in der Regel nicht alle bestimmen können, können wir $\overline{y}$ bzw. σ^2 nicht explizit berechnen. Um diese Werte wenigstens schätzen zu können, nehmen wir eine *Stichprobe vom Umfang n* und bestimmen von jedem Element der Stichprobe die Merkmalsausprägung. Wir konzentrieren uns im Folgenden auf **uneingeschränkt zufällige Stichproben vom Umfang n**. Unter dem Ziehen einer solchen Stichprobe verstehen wir ein Laplace-Experiment mit Grundraum

$$\Omega := \{(a_1, \ldots, a_n) \mid a_1 < \ldots < a_n,\, \{a_1, \ldots, a_n\} \subset \{1, \ldots, N\}\}\,.$$

Jede Stichprobe wird mit der gleichen Wahrscheinlichkeit $\frac{1}{|\Omega|} = \binom{N}{n}^{-1}$ gezogen. Bestimmen wir die Merkmalsausprägungen der durch die Stichprobenziehung zufällig ausgewählten Merkmalsträger, so erhalten wir eine Zufallsvariable

$$Y : \left\{ \begin{array}{rcl} \Omega & \to & \mathbb{R}^n \\ (a_1, \ldots, a_n) & \mapsto & (y_{a_1}, \ldots, y_{a_n})^\top. \end{array} \right.$$

Analog zu obigen Größen definieren wir die Zufallsvariablen

1. das **Stichprobenmittel**
$$\overline{Y} := \frac{1}{n}\sum_{i=1}^{n} y_{a_i} \tag{A.1}$$
2. die **Stichprobenvarianz**
$$\hat{\sigma}^2 := \frac{1}{n}\sum_{i=1}^{n}(y_{a_i} - \overline{Y})^2. \tag{A.2}$$

Satz A.1.1 *Für die Schätzer $\overline{Y}$ bzw. $\hat{\sigma}^2$ gilt*

$$E(\overline{Y}) = \overline{y} \quad \textit{bzw.} \quad E(\hat{\sigma}^2) = \sigma^2 \cdot \frac{n-1}{n} \cdot \frac{N}{N-1}.$$

Weiter gilt

$$Var(\overline{Y}) = \frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right).$$

Oftmals will man anstelle eines Schätzwertes $\overline{Y}$ für $\overline{y}$ auch ein Intervall angeben, in dem $\overline{y}$ mit einer großen Wahrscheinlichkeit liegt, ein sogenanntes **Konfidenzintervall**.
Sei $\alpha \in (0,1)$ ein vorgegebenes Niveau, dann folgt mit der Ungleichung von Tschebyscheff, vgl. (Henze, 1995), S. 93

$$P\left(|\overline{Y} - \overline{y}| \leq \frac{1}{\sqrt{\alpha}}\sqrt{\mathrm{Var}(\overline{Y})}\right) \geq 1 - \alpha,$$

d.h. der Wert $\overline{y}$ liegt mit einer Wahrscheinlichkeit von mindestens $1-\alpha$ im Intervall

$$\begin{aligned} C &= \left[\overline{Y} - \frac{1}{\sqrt{\alpha}}\sqrt{\mathrm{Var}(\overline{Y})},\ \overline{Y} + \frac{1}{\sqrt{\alpha}}\sqrt{\mathrm{Var}(\overline{Y})}\right] \\ &= \left[\overline{Y} - \frac{1}{\sqrt{\alpha}}\sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right)},\ \overline{Y} + \frac{1}{\sqrt{\alpha}}\sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right)}\right]. \end{aligned}$$

Da jedoch die Varianz σ^2 unbekannt ist, muss sie entweder nach oben abgeschätzt werden, oder man ersetzt σ^2 durch den Schätzer $\hat{\sigma}^2$ und erhält als Konfidenzintervall

$$I = \left[\overline{Y} - \frac{1}{\sqrt{\alpha}}\sqrt{\frac{\hat{\sigma}^2}{n}\left(1 - \frac{n-1}{N-1}\right)},\ \overline{Y} + \frac{1}{\sqrt{\alpha}}\sqrt{\frac{\hat{\sigma}^2}{n}\left(1 - \frac{n-1}{N-1}\right)}\right].$$

Speziell für kleine Stichprobenumfänge kann jedoch die Überdeckungswahrscheinlichkeit dieses Intervalls I kleiner als $1-\alpha$ sein. Sind Grundgesamtheit und Stichprobenumfang sehr groß, so kann man mittels des zentralen Grenzwertsatzes nachweisen, dass die Überdeckungswahrscheinlichkeit des Intervalls I sehr viel größer als $1-\alpha$ ist. So kann man z.B. bei einem Niveau von $\alpha = 0.25$ und bei sehr großen N und n sagen, dass die Intervalle I den Mittelwert $\overline{y}$ mit einer Wahrscheinlichkeit von näherungsweise 0.9544 überdecken, vgl. (Stenger, 1986), S. 37.

Eine der wichtigsten Fragestellungen ist die **Suche nach einem geeigneten Stichprobenumfang**. Unter Vorgabe von

1. einer maximalen Intervalllänge ℓ und
2. einem Fehlerniveau α

sucht man den minimalen Stichprobenumfang $n_{l,\alpha}$, so dass man bei obigem Vorgehen immer ein Intervall C von maximal der Länge ℓ erhält, in dem $\overline{y}$ mit einer Wahrscheinlichkeit von mindestens $1-\alpha$ liegt.
Eine Antwort erhält man durch Lösen der Gleichung

$$|C| \stackrel{!}{=} \ell.$$

Dabei bezeichne $|C|$ die Intervalllänge von C, also

$$\frac{2}{\sqrt{\alpha}}\sqrt{\frac{\sigma^2}{n}\left(1-\frac{n-1}{N-1}\right)} \stackrel{!}{=} \ell.$$

Diese Gleichung hat die Lösung

$$n_{l,\alpha} = \frac{4N\sigma^2}{\alpha l^2(N-1)+4\sigma^2}. \tag{A.3}$$

Dieser Wert muss jetzt nur noch ganzzahlig nach oben gerundet werden.
Ist die Grundgesamtheit sehr groß und die Schranke für die Varianz klein ($\alpha l^2 N \gg 4\sigma^2$), so gilt annähernd

$$n_{l,\alpha} \approx \frac{4\sigma^2}{\alpha l^2},$$

die Größe N der Grundgesamtheit spielt also keine Rolle mehr.

Natürlich können wir den minimalen Stichprobenumfang aus Formel A.3 nicht berechnen, da wir σ^2 nicht kennen. Da aber $n_{l,\alpha}$ als Funktion von σ^2 monoton wächst, können wir $n_{l,\alpha}$ nach oben abschätzen, wenn wir eine obere Schranke σ_o^2 von σ^2 kennen. Gilt also

$$\sigma^2 \leq \sigma_o^2,$$

so folgt

$$n_{l,\alpha} \leq \frac{4N\sigma_o^2}{\alpha l^2(N-1)+4\sigma_o^2} =: \tilde{n}_{l,\alpha}. \tag{A.4}$$

Problematisch ist natürlich die Suche nach einer solchen oberen Schranke σ_o^2 . Eine grobe, aber sehr einfache Abschätzung liefert das folgende Lemma, das direkt aus (Henze, 1995), Aufgabe 6.8 a folgt.

Lemma A.1.2 *Unter der Voraussetzung, dass eine obere und eine untere Schranke für die Merkmalsausprägungen existieren, also*

$$y^{min} \leq y_i \leq y^{max}, \qquad i = 1, \ldots, N,$$

d.h. alle Merkmalsausprägungen liegen im Intervall $[y^{min}, y^{max}]$ *mit der Länge*

$$\ell_y := y^{max} - y^{min},$$

gilt:

$$\sigma^2 \leq \frac{\ell_y^2}{4} =: \sigma_{max}^2.$$

Bemerkung A.1.3 *Die in Lemma A.1.2 angegebene Abschätzung ist für viele Anwendungen zu grob, d.h. die Stichprobenumfänge, die man dadurch erhält, sind zu groß. Dies belegt das folgende Beispiel A.1.4 nachhaltig. Für den Fall, dass die Schranken y^{min} bzw. y^{max} sehr nahe beisammen liegen, ist die Abschätzung aber durchaus brauchbar, vgl. Beispiel A.1.5.*
Ein anderes Problem beim Anwenden von Formel (A.3) bzw. (A.4) stellt die Vorgabe der Intervalllänge ℓ dar. Diese sollte in sinnvollem Zusammenhang zum wahren Mittelwert $\overline{y}$ stehen. Der Anwender benötigt also eine brauchbare Vorstellung vom wahren Mittelwert $\overline{y}$. Oftmals stehen im Vorfeld Beispieldaten zur Verfügung, die entweder von einer älteren Erhebung stammen oder von einer Erhebung vergleichbarer Merkmale. Aus diesen Daten erhält man in etwa die Größenordnung, in der sich $\overline{y}$ bewegen muss. Ebenso kann die Varianz dieses Datensatzes als Grundlage für eine obere Schranke σ_o^2 von σ^2 dienen. Dies ist extrem wichtig, um die angedeuteten Schwierigkeiten, die bei der Anwendung von Lemma A.1.2 auftreten, zu umgehen. Man vergleiche auch hierzu das folgende Beispiel A.1.4.

Beispiel A.1.4 *Vorgegeben sei eine Grundgesamtheit von 750 Objekten mit Merkmalsausprägungen $y_1, \ldots, y_{750}$. Für diese Merkmalsausprägungen kennen wir die Abschätzung*

$$100 \leq y_i \leq 30000, \qquad i = 1, \ldots, 750.$$

Den wahren Mittelwert $\overline{y}$ erwarten wir im Intervall $[2000, 3000]$. Für den Anwender sei in diesem Beispiel eine Intervalllänge $\ell = 500$ sinnvoll. Desweiteren geben wir uns ein Niveau von $\alpha = 0.2$ vor.

1. *Wenden wir Lemma A.1.2 an, so erhalten wir die Abschätzung*

$$\sigma^2 \leq \sigma_{max}^2 = \frac{(30000 - 100)^2}{4} = 223.502.500.$$

Jetzt liefert Formel (A.4) mit $\sigma_o^2 = \sigma_{max}^2$ den Mindeststichprobenumfang

$$\tilde{n}_{\ell,\alpha} = 720.$$

Hiernach ist nahezu eine Totalerhebung vonnöten. Die zu grobe Abschätzung der Varianz σ^2 führte zu diesem Ergebnis.

2. *Auf der anderen Seite liegt uns ein Beispieldatensatz vor. Der Mittelwert dieses Beispieldatensatzes liegt bei 2.173, die Varianz bei ca. 10.000.000. Auf Grund dieser Daten wählen wir als Abschätzung für die Varianz*

$$\sigma_o^2 = 12.000.000.$$

Jetzt liefert Formel (A.4) den Mindeststichprobenumfang

$$\tilde{n}_{\ell,\alpha} = 422.$$

Der große Unterschied zum Mindeststichprobenumfang aus 1. verdeutlicht die Bedeutung einer guten Abschätzung für die Varianz.

Beispiel A.1.5 *Wir untersuchen ein ausgewähltes Element der Grundgesamtheit auf Zugehörigkeit zu einer vorgegebenen Klasse K, also die Merkmalsausprägungen*

$$y_i = \begin{cases} 1 & \text{falls } g_i \in K \\ 0 & \text{falls } g_i \notin K \end{cases}, \qquad i = 1, \ldots, N. \tag{A.5}$$

Das arithmetische Mittel $\overline{y}$ ist damit gleich dem prozentualen Anteil p_K der Klasse K an der Gesamtheit. Für σ^2 erhalten wir mit Lemma A.1.2 die Abschätzung

$$\sigma^2 \leq \frac{1}{4} = \sigma^2_{max}.$$

Für einen maximalen Schätzfehler von 5% dieses prozentualen Anteils, d.h. einer Intervalllänge $\ell = 0.1$, bei einem Niveau von $\alpha = 0.2$ und einer Grundgesamtheit von $N = 3000$ liefert Formel (A.4) mit $\sigma_o^2 = \sigma^2_{max}$ einen Mindeststichprobenumfang

$$\tilde{n}_{\ell,\alpha} = 429.$$

Wie bereits angedeutet liefert die Ungleichung von Tschebyscheff nur eine sehr grobe Abschätzung. Die tatsächliche Überdeckungswahrscheinlichkeit der angegebenen Konfidenzintervalle ist in Wirklichkeit sehr viel größer als das vorgegebene Niveau. Mittels des zentralen Grenzwertsatzes kann man anstelle von C auch ein **asymptotisches Konfidenzintervall** C_a bestimmen (vgl. (Stenger, 1986), Anhang B3):

$$C_a = \left[\overline{Y} - \sqrt{\frac{\hat{\sigma}^2}{n}\left(1 - \frac{n}{N}\right)}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right),\ \overline{Y} + \sqrt{\frac{\hat{\sigma}^2}{n}\left(1 - \frac{n}{N}\right)}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \right]. \tag{A.6}$$

Hierbei bezeichnet Φ^{-1} die Quantilfunktion der Standardnormalverteilung. Für große Stichprobenumfänge n (als Faustformel diene $n \geq 30$) hat dieses Konfidenzintervall dann tatsächlich die Überdeckungswahrscheinlichkeit $1 - \alpha$. Das Niveau wird somit voll ausgeschöpft.

A.1.2 Verhältnisschätzung

Während wir bisher immer am Mittelwert *eines* Merkmals interessiert waren, betrachten wir nun eine Grundgesamtheit N bei der wir *zwei* Merkmale

$$y_1, \ldots, y_N$$

und

$$z_1, \ldots, z_N$$

bestimmen können. Wir setzen $z_i > 0$ für $i = 1, \ldots, N$ voraus und interessieren uns für das Verhältnis

$$\frac{\overline{y}}{\overline{z}}$$

der arithmetischen Mittel $\overline{y}$ bzw. $\overline{z}$ der beiden Merkmale. Wir ziehen eine uneingeschränkt zufällige Stichprobe vom Umfang n und erhalten die Stichprobenmerkmale $y_{a_1}, \ldots, y_{a_n}$ bzw. $z_{a_1}, \ldots, z_{a_n}$. Dies beschreiben wir wie in Abschnitt A.1.1 durch die *stochastisch abhängigen* Zufallsvariablen Y und Z, die die jeweiligen Merkmale der gezogenen Merkmalsträger beschreiben. Mit $\overline{Y}$ bzw. $\overline{Z}$ bezeichnen wir die Stichprobenmittel, vgl. (A.1), und mit $\hat{\sigma}_Y^2$ bzw. $\hat{\sigma}_Z^2$ die Stichprobenvarianzen, vgl. (A.2), zudem sei

$$\hat{\sigma}_{YZ} := \frac{1}{n}\sum_{i=1}^{n}(y_{a_i} - \overline{Y})(z_{a_i} - \overline{Z}) \tag{A.7}$$

die Stichprobencovarianz der beiden Zufallsvariablen. Dann ist nach Stenger (1986), S. 70, die Zufallsvariable

$$\frac{\frac{\overline{Y}}{\overline{Z}} - \frac{\overline{y}}{\overline{z}}}{\sqrt{\frac{1}{n\overline{Z}^2}\left(1 - \frac{n}{N}\right)\left(\hat{\sigma}_Y^2 - 2\frac{\overline{Y}}{\overline{Z}}\hat{\sigma}_{YZ} + \left(\frac{\overline{Y}}{\overline{Z}}\right)^2\hat{\sigma}_Z^2\right)}}$$

für große Werte von n annähernd $N(0,1)$–verteilt und wir können damit das asymptotisches Konfidenzintervall

$$C_\alpha = \left[\frac{\overline{Y}}{\overline{Z}} - \sqrt{\frac{\hat{\sigma}^2_{\frac{Y}{Z}}}{n}\left(1 - \frac{n}{N}\right)}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right),\ \frac{\overline{Y}}{\overline{Z}} + \sqrt{\frac{\hat{\sigma}^2_{\frac{Y}{Z}}}{n}\left(1 - \frac{n}{N}\right)}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right] \tag{A.8}$$

mit

$$\hat{\sigma}^2_{\frac{Y}{Z}} := \frac{1}{\overline{Z}^2}\left(\hat{\sigma}_Y^2 - 2\frac{\overline{Y}}{\overline{Z}}\hat{\sigma}_{YZ} + \left(\frac{\overline{Y}}{\overline{Z}}\right)^2\hat{\sigma}_Z^2\right)$$

für das unbekannte Verhältnis $\frac{\overline{y}}{\overline{z}}$ konstruieren, das für große Werte von n (als Faustformel dient wieder $n \geq 30$) die Überdeckungswahrscheinlichkeit $1 - \alpha$ hat.

Beispiel A.1.6 *Gegeben ist eine Grundgesamtheit von N Objekten $g_1, \ldots, g_N$ mit positiven Merkmalsausprägungen $z_1, \ldots, z_N$. Desweiteren sei eine Klasse K vorgegeben. Wir interessieren uns für den prozentualen Anteil der Klasse K am arithmetischen Mittel des gegebenen Merkmals, also an dem Quotienten*

$$\frac{\sum\limits_{i:g_i \in K} z_i}{\sum\limits_{i=1}^{N} z_i}.$$

Dazu definieren wir das Merkmal

$$y_i = \begin{cases} z_i & \text{falls } g_i \in K \\ 0 & \text{falls } g_i \notin K \end{cases}, \qquad i = 1, \ldots, N,$$

für das

$$\frac{\sum\limits_{i:g_i\in K} z_i}{\sum\limits_{i=1}^{N} z_i} = \frac{\overline{y}}{\overline{z}}$$

gilt. Damit ist durch Formel (A.8) ein asymptotisches Konfidenzintervall für den uns interessierenden Quotienten $\frac{\overline{y}}{\overline{z}}$ *gegeben.*

Beispiel A.1.7 *Gegeben ist eine Grundgesamtheit von* N *Objekten* $g_1, \ldots, g_N$ *sowie zwei Klassen* A *und* K*. Wir interessieren uns für den prozentualen Anteil der Elemete aus Klasse* K*, die zusätzlich Element der Klasse* A *sind, d.h. für*

$$\frac{\sum\limits_{i:g_i\in K\cap A} 1}{\sum\limits_{i:g_i\in K} 1}.$$

Dazu definieren wir die Merkmale

$$y_i = \begin{cases} 1 & \textit{falls } g_i \in K \cap A \\ 0 & \textit{falls } g_i \notin K \cap A \end{cases} \qquad \textit{und} \qquad z_i = \begin{cases} 1 & \textit{falls } g_i \in K \\ 0 & \textit{falls } g_i \notin K \end{cases},$$

$i = 1, \ldots, N$*. Dann gilt*

$$\frac{\sum\limits_{i:g_i\in K\cap A} 1}{\sum\limits_{i:g_i\in K} 1} = \frac{\overline{y}}{\overline{z}}$$

gilt. Damit ist durch Formel (A.8) ein asymptotisches Konfidenzintervall für den uns interessierenden Quotienten $\frac{\overline{y}}{\overline{z}}$ *gegeben.*

A.2 Die logarithmische Normalverteilung

In vielen Anwendungsbereichen treten Wahrscheinlichkeitsverteilungen auf, die nur positive Werte annehmen können. Ein wichtiger Vertreter dieser Klasse stellt die *logarithmische Normalverteilung* dar. Wir nennen eine Zufallsvariable X logarithmisch normalverteilt mit Parametern $\mu \in \mathbb{R}$ und $\tau^2 > 0$, in Zeichen $X \sim LN(\mu, \tau^2)$, falls $\ln(X)$ normalverteilt ist mit Erwartungswert μ und Varianz τ^2, also $\ln(X) \sim N(\mu, \tau^2)$. Erwartungswert und Varianz der logarithmisch normalverteilten Zufallsgröße X berechnen sich zu

$$\mathrm{E}(X) = \exp\left(\mu + \frac{1}{2}\tau^2\right)$$

bzw.

$$\mathrm{Var}(X) = \exp\left(2\mu + \tau^2\right) \cdot \left(\exp\left(\tau^2\right) - 1\right).$$

Desweiteren besitzt X eine Lebesgue-Dichte der Form

$$f(x) = \frac{1}{x\tau\sqrt{2\pi}} \cdot \exp\left(-\frac{(\ln x - \mu)^2}{2\tau^2}\right) \cdot 1_{(0,\infty)}(x), \qquad x \in \mathbb{R},$$

wie sie in Abbildung A.1 aufgezeigt ist, und die Verteilungsfunktion

$$F(t) = \Phi\left(\frac{\ln t - \mu}{\tau}\right), \qquad t > 0.$$

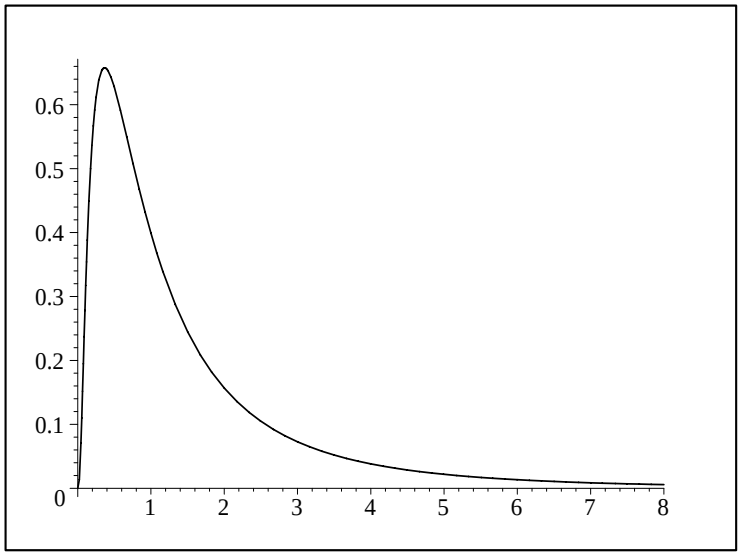


Abbildung A.1: Dichte der logarithmischen Normalverteilung $LN(0,1)$

Werden n unabhängige $LN(\mu,\tau^2)$–verteilte Zufallsvariablen beobachtet mit unbekannten Parametern μ und τ^2, und ergeben sich dabei die Beobachtungswerte $x_1,\ldots,x_n$, so sind erwartungstreue Schätzer für μ bzw. τ^2 gegeben durch

$$\hat{\mu}(x_1,\ldots,x_n) = \frac{1}{n}\sum_{i=1}^{n} \ln x_i \tag{A.9}$$

bzw.

$$\hat{\tau}^2(x_1,\ldots,x_n) = \frac{1}{n-1}\sum_{i=1}^{n} (\ln x_i - \hat{\mu})^2. \tag{A.10}$$

Die Schätzer $\hat{\mu}$ bzw. $\frac{n-1}{n}\cdot\hat{\tau}^2$ sind die Maximum-Likelihood-Schätzer für μ bzw. τ^2, vgl. Johnson, Kotz und Balakrishnan (1994), S. 220.

Abschließend sei auf das Buch von Crow und Shimizu (1988) hingewiesen, das ausschließlich die logarithmische Normalverteilung und ihre Anwendungen behandelt.

Anhang B

Zuordnung von ALB-Nutzungsklassen

Im Folgenden sind die Zuordnungen der Nutzungsangaben im ALB zu den von uns verwendeten Nutzungsklassen tabellarisch aufgelistet. Hierbei sind wir im Bereich der Nichtwohngebäude nach dem offiziellen Schlüssel des statistischen Bundesamtes vorgegangen, vgl. Statistisches Bundesamt (1977). Diese Methode ist somit auf ALB anderer Gemeinden problemlos übertragbar. Im Folgenden sind natürlich nur die Fälle aufgelistet, die im ALB von Ettlingen aufgetreten sind.

ALB	**Klasse**	**ALB**	**Klasse**
Altenheim	K_5	Badegebäude	K_8
Betriebsgebäude	K_7	Bibliotheksgebäude	K_8
Bürogebäude	K_4	Empfangsgebäude	K_7
Energieversorgungsgebäude	K_7	Fabrikgebäude	K_7
Feuerwehrgebäude	K_4	Friedhofsgebäude	K_8
Gaststätte	K_7	Gemeindehaus	K_8
Gerichtsgebäude	K_4	Geschäftshaus	K_7
Gewächshaus	K_6	Hallenbad	K_8
Heim	K_5	Hotel	K_7
Jugendherberge	K_5	Kapelle	K_8
Kindergarten	K_8	Kiosk	K_7
Kirche	K_8	Kläranlagengebäude	K_7
Kutschengebäude	K_8	Lagergebäude	K_7
Parkhaus	K_7	Polizeigebäude	K_4
Postamt	K_8	Rathaus	K_4
Scheuer	K_6	Scheuer und Stall	K_6
Schloß	K_8	Schloßgartenhalle	K_8
Schulgebäude	K_8	Schuppen	K_6
Schutzhütte	K_8	Sendeturm	K_7
Sporthalle	K_8	Stall	K_6
Tankstellengebäude	K_7	Tiefgarage	K_7
Toilettengebäude	K_8	Turm	K_6
Umformer	K_7	Veranstaltungsgebäude	K_8
Vergnügungsstätte	K_7	Verwaltungsgebäude	K_4
Wartehalle	K_7	Wasserbehälter	K_7
Werkstattgebäude	K_7	Wirtschaftsgebäude	K_6
Wochenendhaus	$K_1 - K_3$	Wohn- und Betriebsgeb.	K_7
Wohn- und Bürogebäude	K_4	Wohn- u. Geschäftshaus	K_7
Wohn- und Verwaltungsgeb.	K_4	Wohnhaus	$K_1 - K_3$

Anhang C

Die Variablen von Kapitel 3

Hier folgt eine tabellarische Übersicht der in Kapitel 3 verwendeten Variablen sowie ihrer Bedeutung. Die Tabellen sind chronologisch nach den Abschnitten geordnet. Eine Variable wird nur da aufgeführt, wo sie zum ersten Mal auftritt.

Variable	Bedeutung
K_1, K_2, K_3	Klassen der Wohngebäude (K_1 = EFH, K_2 = MFH, K_3 = HH)
N_{wohn}	Anzahl der Wohngebäude Ettlingens
N_1, N_2, N_3	Anzahl der Wohngebäude Ettlingens in den Klassen K_1, K_2, K_3
n_{wohn}	Anzahl der Wohngebäude der Stichprobe
n_1, n_2, n_3	Anzahl der Wohngebäude der Klassen K_1, K_2, K_3 der Stichprobe
q_1, q_2, q_3	Anteil der Klassen K_1, K_2, K_3 an N_{wohn}, $q_i = N_i/N_{wohn}$
$\overline{q_1}, \overline{q_2}, \overline{q_3}$	Stichprobenmittel, Schätzungen für q_1, q_2, q_3, $q_i = n_i/n_{wohn}$
$\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\sigma}_3^2$	Stichprobenvarianzen der prozentualen Anteile
$C_i(\alpha)$	Konfidenzintervalle für q_1, q_2, q_3

Tabelle C.1: Die Variablen aus Abschnitt 3.1.1

Variable	Bedeutung
$n_{BRI,wohn}$	Anzahl der Wohngebäude der Stichprobe mit BRI
$n_{BRI,i}$	Anzahl der Wohngebäude der Klassen K_i der Stichprobe mit BRI
$\overline{y}_{BRI,wohn}$	Unbekannte mittlerer BRI aller Wohngebäude
$\overline{Y}_{BRI,wohn}$	Mittlerer BRI der Stichprobe
$\hat{\sigma}^2_{BRI,wohn}$	Varianz der BRIe der Stichprobe
$\overline{y}_{BRI,1}, \overline{y}_{BRI,2}$	Unbekannte mittlerer BRI der Klassen K_1, K_2
$\overline{Y}_{BRI,1}, \overline{Y}_{BRI,2}$	Mittlerer BRI der Klassen K_1, K_2 der Stichprobe
$\hat{\sigma}^2_{BRI,1}, \hat{\sigma}^2_{BRI,2}$	Varianz der BRIe der Klassen K_1, K_2 der Stichprobe
$C_{BRI,wohn}(\alpha)$	Konfidenzintervall für $\overline{y}_{BRI,wohn}$
$C_{BRI,i}(\alpha)$	Konfidenzintervall für $\overline{y}_{BRI,i}$, $i = 1, 2$

Tabelle C.2: Die Variablen aus Abschnitt 3.1.2

Variable	Bedeutung
$p_{BRI,i}$	Prozentualer Anteil der Klasse K_i am BRI der Wohngebäude
$\hat{p}_{BRI,i}$	Schätzung für $p_{BRI,i}$

Tabelle C.3: Die Variablen aus Abschnitt 3.1.4

Variable	Bedeutung
$Y_{BRI,1}$	Zufallsvariable: BRI eines Einfamilienhauses
$\mu_{BRI,1}$	Wahrer Mittelwert von $\ln(Y_{BRI,1})$
$\hat{\mu}_{BRI,1}$	Schätzung für $\mu_{BRI,1}$
$\tau_{BRI,1}$	Wahre Varianz von $\ln(Y_{BRI,1})$
$\hat{\tau}_{BRI,1}$	Schätzung für $\tau_{BRI,1}$

Tabelle C.4: Die Variablen aus Abschnitt 3.1.5

Variable	Bedeutung
p_{A_i,K_j}	Prozentualer Anteil der Alters-Nutzungsklasse $A_i \cap K_j$ am gesamten Gebäudebestand
$\hat{p}_{A_i,K_j}$	Schätzung für p_{A_i,K_j}

Tabelle C.5: Die Variablen aus Abschnitt 3.1.6

Literaturverzeichnis

Barth, B. und Schwaiger, B. (1998). *Dynamisches Gebäudebestandsmodell zur Berechnung von Energiehaushalt, Stoffflüssen und Finanzbedarf für Investition und Betrieb* (DFG Schlussbericht RI 816/1-2). Institut für Industrielle Bauproduktion der Universität Karlsruhe. (Unveröffentlicht)

Crow, E. und Shimizu, K. (1988). *Lognormal Distributions: Theory and Applications.* New York, Basel: Marcel Dekker, Inc.

Deutsches Institut für Normung e.V. (1988). Kosten von Hochbauten, Flächen, Rauminhalte. In *DIN Taschenbuch* (Vol. 114). Beuth.

Henze, N. (1995). *Stochastik I.* Skriptum an der Universität Karlsruhe.

Johnson, N., Kotz, S. und Balakrishnan, N. (1994). *Continuous Univariate Distributions* (Vol. 1, 2. Aufl.). New York: Wiley.

Statistisches Bundesamt. (1977). *Signierschlüssel für Nichtwohngebäude.* Wiesbaden: Statistisches Bundesamt.

Stenger, H. (1986). *Stichproben.* Heidelberg, Wien: Physica-Verlag.