



# StaWi

Staat und Wirtschaft in Hessen

72. Jahrgang 1 | 2017



## Kleinräumige Statistik

Geoinformationen

Small-Area-Verfahren

# Impressum

ISSN 0344 – 5550 (Print)  
ISSN 1616 – 9867 (Digital)

**Copyright:**

Hessisches Statistisches  
Landesamt, Wiesbaden, 2017

Vervielfältigung und Verbreitung, auch  
auszugsweise, mit Quellenangabe  
gestattet.

**Herausgeber:**

Hessisches Statistisches Landesamt,  
Wiesbaden, Rheinstraße 35/37  
Telefon: 0611 3802-0,  
Telefax: 0611 3802-890  
E-Mail: [vertrieb@statistik.hessen.de](mailto:vertrieb@statistik.hessen.de)  
Internet: <https://statistik.hessen.de>

**Schriftleitung:**

Sanyel Arikan, Rheinstraße 35/37,  
Wiesbaden, Telefon: 0611 3802-825

**Haus-/Lieferanschrift:**

Hessisches Statistisches Landesamt,  
Rheinstraße 35/37, 65185 Wiesbaden

**Postanschrift:**

Hessisches Statistisches Landesamt,  
65175 Wiesbaden

**Erscheinungsweise:** vierteljährlich

**Bezugspreis:**

Print: 13,00 Euro  
Jahresabonnement: 44,20 Euro  
(jew. inkl. Versandkosten)  
PDF-Datei als kostenloser Download im Internet.

**Gesamtherstellung:**

Hessisches Statistisches Landesamt

**Auskünfte und Informationen**

aus allen Bereichen der hessischen  
Landesstatistik erteilt die zentrale  
Informationsstelle:  
Telefon: 0611 3802-802 oder -807,  
E-Mail: [info@statistik.hessen.de](mailto:info@statistik.hessen.de)

[GESUNDES *unternehmen*]



**Zeichenerklärung**

- = genau Null (nichts vorhanden) bzw. keine Veränderung eingetreten.
- 0 = Zahlenwert ungleich Null, aber weniger als die Hälfte der kleinsten in der Tabelle nachgewiesenen Einheit.
- . = Zahlenwert unbekannt oder geheim zu halten.
- ... = Zahlenwert lag bei Redaktionsschluss noch nicht vor.
- ( ) = Aussagewert eingeschränkt, da der Zahlenwert statistisch unsicher ist.
- / = keine Angabe, da Zahlenwert nicht sicher genug.
- X = Tabellenfach gesperrt, weil Aussage nicht sinnvoll (oder bei Veränderungsraten ist die Ausgangszahl kleiner als 100).
- D = Durchschnitt.
- s = geschätzte Zahl.
- p = vorläufige Zahl.
- r = berichtigte Zahl.

Aus Gründen der Übersichtlichkeit sind nur negative Veränderungsraten und Salden mit einem Vorzeichen versehen.

Positive Veränderungsraten und Salden sind ohne Vorzeichen.

Im Allgemeinen ist ohne Rücksicht auf die Endsumme auf- bzw. abgerundet worden.

Das Ergebnis der Summierung der Einzelzahlen kann deshalb geringfügig von der Endsumme abweichen.

# Editorial



Liebe Leserinnen und Leser,

**Staat und Wirtschaft in Hessen** will statistisches Wissen noch kompakter und nutzerfreundlicher für Sie aufbereiten. Ab sofort versorgen wir Sie deshalb quartalsweise mit Themenheften zur amtlichen Statistik. Im aktuellen Heft stellen wir Ihnen Facetten regionaler Datengewinnung und -aufbereitung vor. Sie lernen Projekte kennen, die im Rahmen der Kooperation des HSL mit der Universität Trier im Studiengang *European Master in Official Statistics (EMOS)* am Lehrstuhl von Prof. Dr. Ralf Münnich entstanden sind. Mit dieser Zusammenarbeit öffnen wir uns aktiv unserem Partner Wissenschaft und bleiben auf dem neuesten Stand zeitgemäßer Methodik. Natürlich finden Sie weiterhin vertraute Rubriken wie bspw. „Hessen in Europa“.

Wir starten das neue Konzept mit dem zukunftsweisenden Thema „Geoinformation“. Geographische Informationssysteme wie *GoogleMaps* sind für uns alle längst zu Alltagsbegleitern geworden. Auch in der Statistik gewinnen die geobasierten Daten und Darstellungsweisen immer mehr an Bedeutung: Hot-Spot-Analysen beispielsweise können darstellen, wie statistische Ergebnisse sich regional verteilen. Die Nachfrage nach kleingliedrigen Daten wächst stetig – Wirtschaft, Verwaltung und Politik sind nur drei Abnehmer, die sich für detaillierte raumbezogene Daten interessieren. Um den neuen Anforderungen nachzukommen, hat das HSL das Competence Center *Geoinformation* eingerichtet, das die moderne kartographische Darstellung raumbezogener Statistiken übernimmt. Lernen Sie das Portfolio unseres Competence Centers *Geoinformation* in dieser Ausgabe kennen!

Neben den technischen und gestalterischen Herausforderungen, die Geoinformationsdaten mit sich bringen, wächst auch der Anspruch an die statistische Methodik. Um up to date zu bleiben, loten die Autoren des EMOS-Kooperationsprojektes die Anwendung von Small-Area-Verfahren aus. Sie ermöglichen kleinräumige Schätzungen auch da, wo für klassische Hochrechnungsverfahren keine ausreichende Datenbasis vorhanden ist.

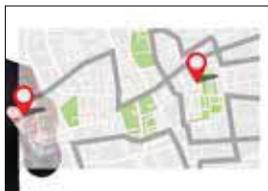
Ich wünsche Ihnen viel Freude beim Entdecken der neuen **Staat und Wirtschaft in Hessen!**

Ihre

A handwritten signature in blue ink that reads "Christel Figgner". The signature is written in a cursive style.

Dr. Christel Figgner

Präsidentin des Hessischen Statistischen Landesamts



© adiruch na chiangmai - Fotolia.com

## Geoinformationen

3

Geoinformationen – die „Veredelung“ räumlich-statistischer Daten im digitalen Zeitalter. *Von Sarah Scholze* |3

Das Competence Center Geoinformation im Hessischen Statistischen Landesamt. *Von Philip Graze* |13



© allvision - Fotolia.com

## Small-Area-Verfahren

17

Der Einsatz von Small-Area-Verfahren – erste Erfahrungen mit Area-Level-Schätzungen. *Von Dr. Peter Gottfried* |17

Die Regionalisierung der Hessischen Jahreserhebung im Einzelhandel anhand von Small-Area-Verfahren. *Von Julia Manecke* |29



© jorisvo - Fotolia.com

## Hessen in Europa

42

Erwerbstätigenquote in der EU-28, den EFTA-Staaten und den Beitrittskandidaten 2015. *Von Benedikt Kull* |42

## Ausgewählte Daten zur wirtschaftlichen Entwicklung in Hessen

44

## Veröffentlichungen des Hessischen Statistischen Landesamtes 12/16 bis 02/17

45

## Vorschau auf das nächste Heft 2/2017

48



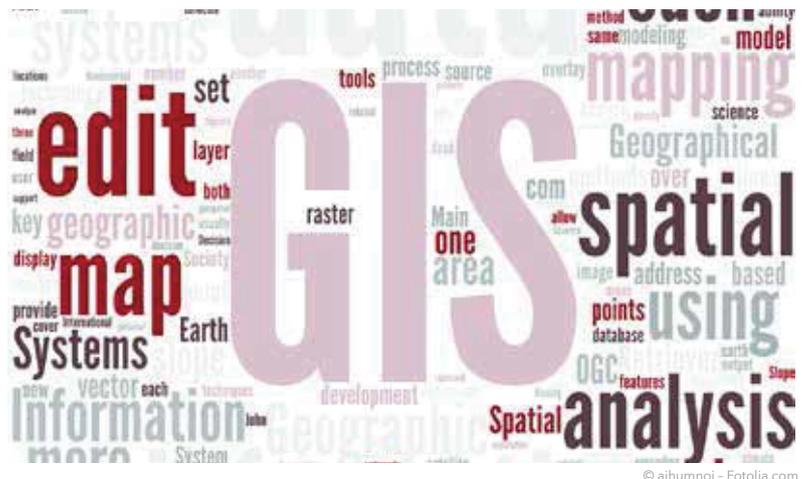
© gustavofraza - Fotolia.com

# Geoinformationen – die „Veredelung“ räumlich-statistischer Daten im digitalen Zeitalter

Moderne Informations- und Kommunikationstechnologien erfassen und nutzen oft Daten mit einem Raumbezug, wodurch relevante Geoinformationen für ökonomische, ökologische und gesellschaftliche Planungs- und Entscheidungsprozesse generiert werden. Geografische Informationssysteme (GIS) werden daher direkt oder indirekt in vielen Branchen und Disziplinen aus Wirtschaft, Verwaltung, Forschung und Lehre oder auch im privaten Bereich eingesetzt und sind fester Bestandteil des digitalen Zeitalters. Was verbirgt sich aber hinter GIS und Schlagworten wie „Geodaten“ oder „Georeferenz“ bzw. „Geokodierung“? Und wie werden aus Rohdaten maßgeschneiderte und nutzerorientierte Geoinformationen, die räumliche Muster und Beziehungen aufzeigen? In diesem Beitrag werden die wesentlichen Begriffe aus dem Umfeld der Geoinformationen erläutert. Anschließend werden Möglichkeiten skizziert, um verschiedene räumlich-statistische Datentypen aufzubereiten, zu analysieren und zu visualisieren. **Von Sarah Scholze**

## Digitale Geowelten und die Rolle des „Wo“

Mobile Endgeräte ausgestattet mit Global Positioning System (GPS)-Technologie, vernetzte Sensoren- und Fernerkundungstechnik oder das sogenannte GeoWeb (Google Maps, Open Street Map etc.) – all diese Informations- und Kommunikationstechnologien sind im digitalen Zeitalter stets präsent. Dabei werden permanent Daten erfasst und ausgetauscht, die oft über einen Raumbezug verfügen<sup>1)</sup> und somit auch Informationen zur Lage, Beschaffenheit und Nutzung eines Ortes bzw. Standortes liefern. Geodaten, also jene Daten mit Raumbezug, sind eine wesentliche Grundlage, um auftretende Phänomene, Prozesse und Entwicklungen zu beschreiben und nachzuvollziehen. Mit Hilfe von Geodaten können beispielsweise optimierte Routen in Bezug auf die Entfernung, Fahrzeit oder Kosten in einem Straßennetz ermittelt werden. Unternehmen optimieren dadurch ihre Logistik und damit die Versorgung ihrer Kundschaft mit Produkten und Dienstleistungen. Im öffentlichen Sektor gibt es ebenfalls vielfältige Einsatzmöglichkeiten von Geodaten, wie beispielsweise beim geplanten Bau einer



Umgehungsstraße. Mit einer Kombination aus kleinräumigen Bevölkerungsdaten und Kartierungen von Naturschutzflächen kann identifiziert werden, wie viele Anwohner am Ortsrand durch den Bau zusätzlich belastet werden und ob es naturschutzrechtliche Bedenken gibt.

Im Vordergrund steht stets das Ziel, anhand von Daten und Zahlen in zeitlicher und räumlicher Auflösung, die Welt zu verstehen, daraus Erkenntnisse zu gewinnen und entsprechende Entscheidungen für Handlungen und Maßnahmen zu treffen. Komplexe und vielschichtige Sachverhalte mittels statistischer Daten

1) In Publikationen wird oft unterstellt, dass 80 % aller Daten bzw. Informationen über einen Raumbezug verfügen. Einer Forschungsstudie von Hahmann und Burghardt (2012) zufolge kann der Anteil der raumbezogenen Daten nicht eindeutig quantifiziert werden.

zu beschreiben, räumlich aufzubereiten und dem Datenkonsumenten sowie einer breiten Öffentlichkeit verständlich zu kommunizieren, sind auch wichtige Aufgabenfelder der amtlichen Statistik und können mit geografischen Informationssystemen (GIS) effektiv umgesetzt werden. GIS ermöglichen es, räumliche Informationen („Wo ist etwas?“) mit erklärenden sachlichen Informationen („Was oder wie ist etwas?“ und ggf. „Wann ist etwas?“) digital zu verknüpfen, zu analysieren und in geeigneter Form zu visualisieren.

### GIS – die „Bühne“ der Geodaten

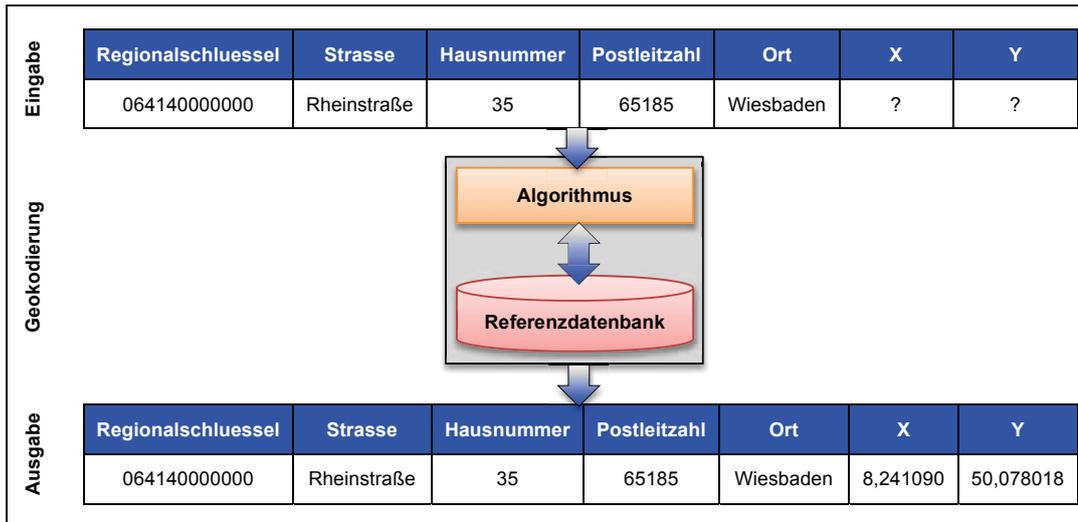
Eine große Anzahl von Daten der amtlichen Statistik basiert auf Einheiten mit einer räumlichen Komponente, wodurch sie mit einem Standort oder einem Gebiet verknüpft sind. Die statistischen Daten werden oft mittels einer Datenbank oder eines Statistikprogramms, wie SAS, verarbeitet und aufbereitet. Solche Systeme berücksichtigen jedoch nicht explizit die räumliche Dimension der Geodaten. GIS bieten dagegen Funktionalitäten, um diese Geodaten effizient zu erfassen, zu verarbeiten, zu analysieren und zu präsentieren (EVAP). Die grundlegenden Funktionen in einem GIS basieren auf dem Layer-Prinzip, das verschiedene Formate von Geodaten miteinander kombiniert, überlagert oder verrechnet. Dabei werden Geobasis- und Geofachdaten unterschieden. Geobasisdaten enthalten Informationen zur Lage auf der Erdoberfläche und repräsentieren Objekte aus der realen Welt, wie Straßen, Gebäude oder Grundstücke. Sie werden insbesondere durch das amtliche Vermessungswesen im Rahmen ihres gesetzlichen Auftrages erfasst und bereitgestellt. Zu den Produkten gehören beispielsweise das „Amtliche Topographisch-Kartographische Informationssystem (ATKIS)“ oder das „Amtliche Liegenschaftskatasterinformationssystem (ALKIS)“. Zudem existieren im GeoWeb auch nutzergenerierte Geobasisdaten, die unter dem Stichwort „Volunteered Geographic Information (VGI)“ zusammengefasst werden (Goodchild, 2007). Ein bekanntes VGI-Projekt ist die Open Street Map (OSM), bei der die Daten von Freiwilligen weltweit erhoben und für jedermann zur Verfügung gestellt werden.

Die Geobasisdaten liegen entweder als Raster- oder Vektordaten in digitaler Form vor. In einem Rasterdatenmodell werden die geografischen Objekte in Zellen oder Pixeln in einer gleichmäßigen Matrix angeordnet und z. B. in den Formaten GeoTiff oder JPEG gespeichert. Die Eigenschaften von geografischen Objekten (z. B. Hangneigung) sind in den Rasterzellen als Farb- oder Zahlenwerte kodiert. Klassische Beispiele für die Verwendung von Rasterdaten sind Luft- und Satellitenbilder. Vektordaten speichern die geografischen Objekte anhand der geometrischen Elemente Punkt, Linie und Fläche in GIS-spezifischen Formaten, wie Shapefile oder FileGeodatabase. Im Vektorformat werden beispielsweise die Verwaltungsgrenzen oder die geografischen Raster dargestellt. Mit Hilfe der Geobasisdaten können Informationen zu Entfernungen, Flächengrößen oder Nachbarschaftsbeziehungen ermittelt werden. Die thematischen Eigenschaften von geografischen Objekten werden anhand der Geofachdaten beschrieben, deren Inhalte beispielsweise aus den Bereichen der amtlichen Statistik, der Epidemiologie oder dem Marketing kommen können.

Interaktiv verbunden werden die Geobasis- und Geofachdaten anhand ihrer Georeferenz, bei der es sich entweder um einen direkten oder indirekten Raumbezug handelt. Der direkte Raumbezug beschreibt eine Position auf der Erdoberfläche und wird durch Geokoordinaten angegeben, die z. B. durch das GPS erfasst werden. Die meisten amtlichen Statistiken weisen jedoch einen indirekten Raumbezug auf, z. B. in Form eines Regionalschlüssels (06414000000) oder durch die Anschrift (Straße, Hausnummer, Postleitzahl und Ort). Der indirekte Raumbezug kann in einen direkten Raumbezug mit Hilfe der Geokodierung<sup>2)</sup> umgewandelt werden. Dadurch wird jede beschreibende Information zu einem räumlichen Punkt transformiert und kann auf einer Karte dargestellt werden (Goldberg, 2008: 5). In der einfachsten Form werden bei der Geokodierung 2 Datensätze, der zu geokodierende Datensatz und der Referenzdatensatz mit den Geokoordinaten, miteinander verglichen. Bei einem „Treffen“ werden die Geokoordinaten dem zu geokodierenden Datensatz zugewiesen (vgl. Abb. 1).

2) Die Begriffe Georeferenzierung und Geokodierung werden in der Literatur sowohl synonym als auch in unterschiedlicher Terminologie verwendet. Goldberg (2008: 3) bezeichnet die Georeferenzierung allgemein als die Zuweisung zu einem Raumbezug, während die Geokodierung als Methode der Georeferenzierung angesehen wird. Weitere Methoden der Georeferenzierung sind die Rektifizierung im Kontext von Luft- und Satellitenbildern sowie das Geotagging in Bezug auf das digitale Foto.

Abbildung 1: Schematischer Prozess der Geokodierung



Bei der Geokodierung sind korrekte Angaben der Adressen eine wesentliche Voraussetzung für ein gutes qualitatives Ergebnis. Fehlerhafte oder unterschiedliche Bezeichnungen müssen daher zuvor aufgedeckt und korrigiert werden. Die Geokodierung ist somit auch unter Qualitätsaspekten ein grundlegender Arbeitsschritt bei der Aufbereitung von adressenbasierten Statistiken. Anhand der georeferenzierten Daten können dann in einem GIS umfassende räumliche Analysen für viele Anwendungsbereiche durchgeführt werden.

### Räumliche (Daten-) Analyse – das Herzstück eines GIS

GIS beinhalten unterschiedliche Werkzeuge zur Datenerfassung, kartografischen Bearbeitung sowie zur räumlich-statistischen Analyse. Die räumliche Analyse gilt dabei als das Herzstück eines GIS, da neue Informationen für entscheidungsunterstützende Abläufe und Erkenntnisse aus den vorliegenden Geodaten generiert werden können (Longley et al., 2015: 291). Ein bekanntes historisches Beispiel für eine gewinnbringende räumliche Analyse ist die des Cholera-Ausbruchs im Jahr 1854 in London. Der englische Arzt Dr. John Snow kartierte sowohl die Wohnorte von Cholera-Opfern als auch die dort vorhandene örtliche Wasserversorgung. Dadurch konnte er die Kontamination einer öffentlichen Pumpe in der Broad Street als Ursache für den Ausbruch der Krankheit aufzufindig machen und unter Einbeziehung des räumlichen Kontexts die Epi-

demie eindämmen. Dies veranschaulicht, dass durch die Verknüpfung unterschiedlicher Daten und deren Visualisierung komplexe räumliche Zusammenhänge aufgedeckt werden können, die sonst nicht ersichtlich gewesen wären.

Die Beantwortung der Fragen „Wo befindet sich etwas?“, „Wie steht etwas miteinander in Zusammenhang?“, „Welche Bedeutung hat dies und welcher mögliche Handlungsbedarf ergibt sich daraus?“ basiert auf einer Kombination unterschiedlicher Methoden und Techniken (z. B. lagebezogene Abfragen und Selektion, Verschneidung oder Interpolationen), die sich auf die räumliche Dimension beziehen (Longley et al., 2015: 291). Bei der Auswahl räumlicher Analysewerkzeuge spielen die verschiedenen räumlich-statistischen Datentypen eine grundlegende Rolle. Diese werden nach Cressie (1993) in kontinuierliche Daten („Continuous Data“), Punktmusterdaten („Point Pattern Data“) und aggregierte Regionaldaten („Areal Data“) klassifiziert.

Kontinuierliche Daten beschreiben Phänomene, die an jedem Punkt eines Gebietes existieren, wie der Niederschlag oder die Bevölkerungsdichte. In der Realität liegen für solche Ereignisse allerdings nur Stichprobenwerte für bestimmte Standorte auf der Erdoberfläche vor. Mittels geostatistischer Verfahren, wie Kriging, können Werte an den unbekannt Positionen aus den erhobenen Daten geschätzt werden, um anschließend die räumliche Variation in einem Untersuchungsgebiet zu ermitteln. Das Ergebnis einer solchen räumlichen Interpolation ist eine

kontinuierliche thematische Oberfläche, deren Darstellung beispielsweise in Wetterkarten vorkommt.

Punktmusterdaten und aggregierte Regionaldaten repräsentieren sogenannte diskrete Phänomene, die räumlich deutlich voneinander abgegrenzt werden können. In der amtlichen Statistik werden Ereignisse oft auf Basis administrativer Einheiten aggregiert, wodurch die tatsächliche räumliche Verteilung der Werte im Raum entweder nur annähernd oder gar nicht wiedergegeben wird. In der Realität konzentrieren sich z. B. Straßenverkehrsunfälle auf bestimmte Gebiete innerhalb einer administrativen Einheit. Die Kerndichteschätzung (KDE: Kernel Density Estimation) ist eine Möglichkeit, die kleinräumige Verteilung von punktbasierten Ereignissen zu ermitteln. Dabei wird anhand der zugrundeliegenden Punktmusterdaten und den jeweiligen Distanzen zueinander eine kontinuierlich eingefärbte Rasteroberfläche erzeugt, die die Dichteverteilungen in einem Gebiet darstellt. Kerndichtekarten können z. B. für räumliche Analysen von Straftaten, Krankheitsfällen oder auch Straßenverkehrsunfällen eingesetzt werden, wobei solche Ereignisse stets nur lokal an einem bestimmten Ort auftreten und somit nicht explizit kontinuierliche Phänomene repräsentieren. Kerndichtekarten können jedoch erste Hinweise z. B. auf lokale Unfallschwerpunkte (sogenannte Hot Spots) oder räumliche Zusammenhänge zwischen den Unfällen und physischen Strukturmerkmalen eines Gebietes liefern, um daraus Präventionsmaßnahmen abzuleiten.

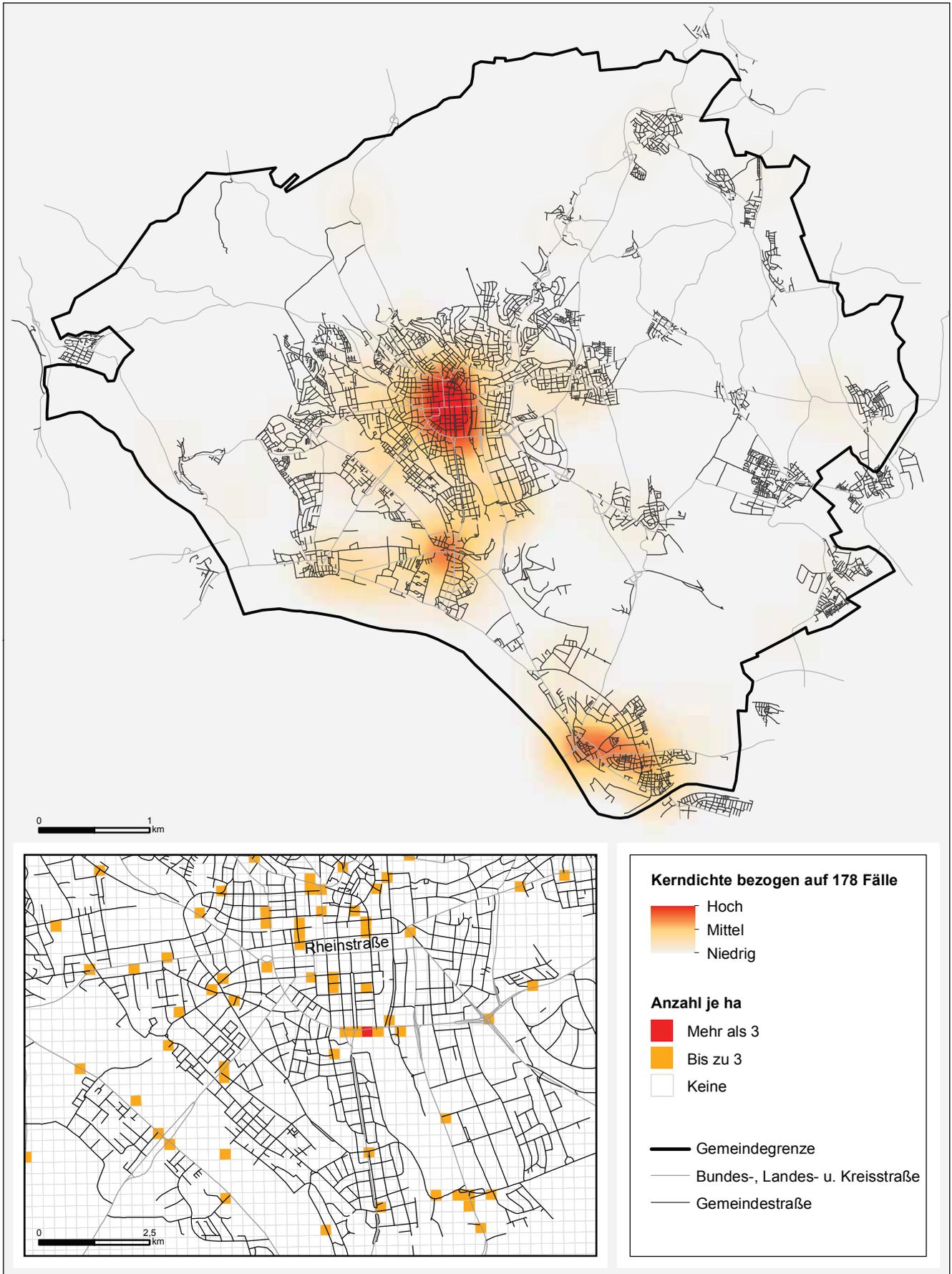
Abbildung 2 stellt die Straßenverkehrsunfälle mit Beteiligung von Fahrrad Fahrenden in der Stadt Wiesbaden auf Basis einer solchen Kerndichteberechnung dar. Der rote Bereich in der Innenstadt von Wiesbaden, im Einzugsgebiet der Biebricher Straße und im Kern von Mainz-Kastel zeigt eine hohe Konzentration von Unfällen, während der orangefarbene und graue Bereich eine geringe Dichte in den jeweiligen Gebieten aufweist. Im zugehörigen Kartenausschnitt wird der Unfallschwerpunkt in der Wiesbadener Innenstadt auf Basis des geografischen Rasters mit einer Seitenlänge von 100 m dargestellt, wodurch detaillierte Aussagen über das Auftreten von Unfällen vor Ort möglich sind.

Weitere Vorzüge von geografischen Rastern in der amtlichen Statistik werden bei Scholze (2015) diskutiert.

Räumliche Cluster auf Basis von aggregierten Regionaldaten können anhand von Hot-Spot-Analysen identifiziert werden. Dabei wird mittels räumlich-statistischer Verfahren (z. B. Gi\*-Statistik) überprüft, inwieweit benachbarte Gebiete in Bezug auf die zu untersuchenden statistischen Ereignisse miteinander korrelieren (Anselin, 1995; Getis und Ord, 1992). Abbildung 3 visualisiert den Anteil der Einpersonenhaushalte je km<sup>2</sup> in % im Regierungsbezirk Darmstadt. Die Darstellung im kleinen Kartenausschnitt vermittelt einen ersten Eindruck zur räumlichen Verteilung, welche jedoch abhängig von den definierten Klassengrenzen ist. Die resultierenden Ergebnisse der Hot-Spot-Analyse basieren auf der Berechnung der Gi\*-Statistiken und geben Auskunft darüber, ob sich die lokalen Werte benachbarter geografischer Rasterzellen statistisch signifikant vom Durchschnitt jener im Regierungsbezirk unterscheiden. Die roten Bereiche, wie etwa in den Städten und Umgebungen von Frankfurt am Main, Offenbach am Main oder Wiesbaden, kennzeichnen dabei die lokalen Hot Spots. Diese Gebiete weisen eine hohe Anzahl von Einpersonenhaushalten auf und sind gleichzeitig auch von Gebieten mit einer ähnlich hohen Anzahl von Einpersonenhaushalten umgeben. Umgekehrt handelt es sich bei den blauen Bereichen um lokale Gebiete mit niedriger Anzahl von Einpersonenhaushalten, die gleichzeitig auch von Gebieten mit einer ähnlich niedrigen Anzahl von Einpersonenhaushalten umgeben sind (sogenannte Cold Spots), wie etwa die Gemeinden Brensbach, Höchst im Odenwald, Fischbachtal oder Fränkisch-Crumbach. Mit einer Wahrscheinlichkeit von 90 % oder mehr sind diese auftretenden räumlichen Zusammenhänge zwischen dem Gebiet und dem Merkmal „Einpersonenhaushalte“ nicht zufälliger Natur. Gebiete mit hohen oder niedrigen Werten sind dabei nicht zwingend statistisch signifikante Hot oder Cold Spots, sondern können auch einer zufälligen räumlichen Verteilung unterliegen (siehe gelber Bereich in der Karte).

Einsatzmöglichkeiten für Hot-Spot-Analysen gibt es in vielen unterschiedlichen Bereichen, wie

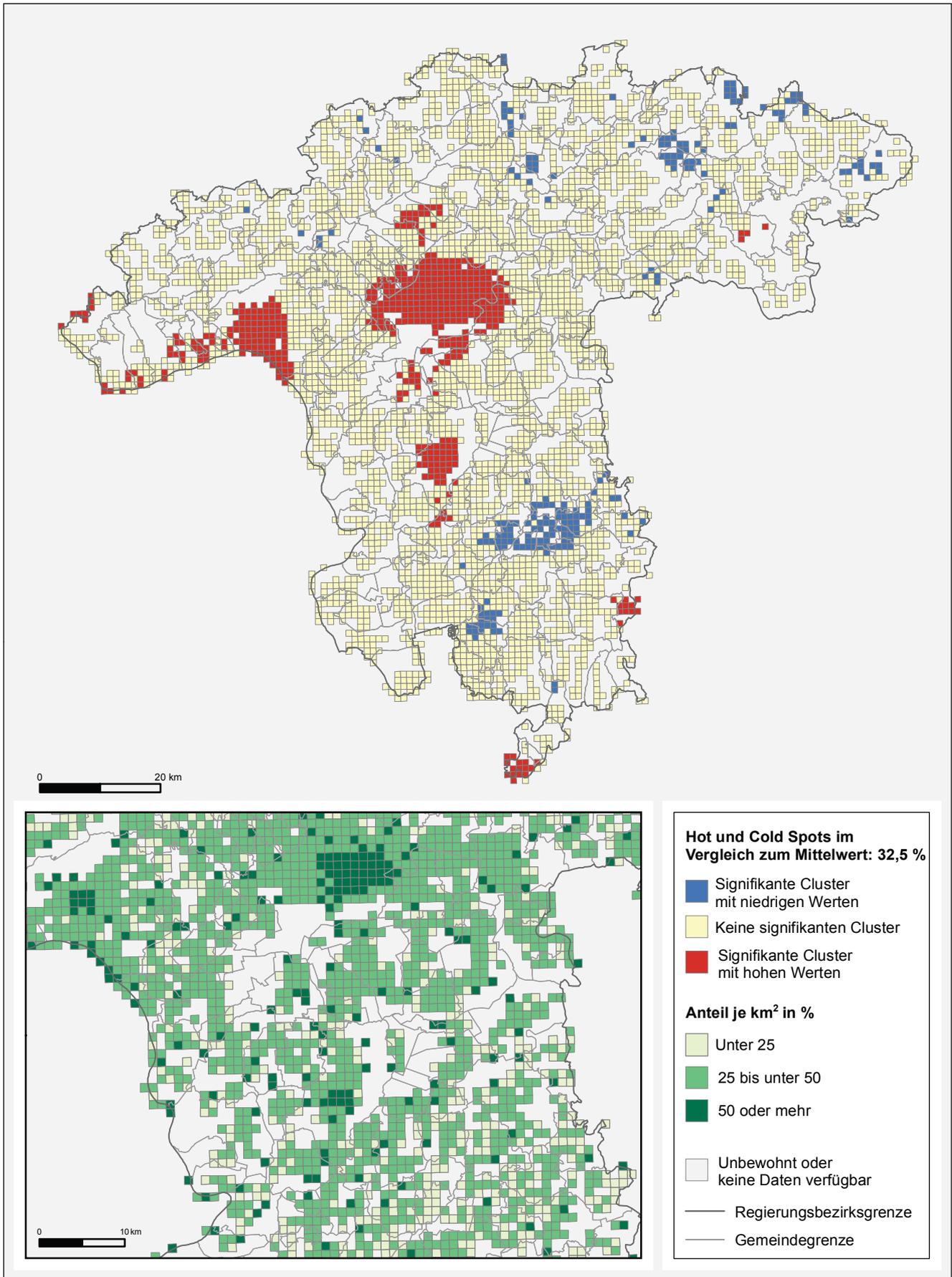
Abbildung 2: Straßenverkehrsunfälle mit Beteiligung von Fahrrad Fahrenden<sup>1)</sup> in Wiesbaden 2015



1) Fahrrad und Pedelec ohne Versicherungskennzeichen

© Hessische Verwaltung für Bodenmanagement und Geoinformation, 2016  
© Hessisches Statistisches Landesamt, Wiesbaden, 2017  
Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.

Abbildung 3: Einpersonenhaushalte im Regierungsbezirk Darmstadt zum 9.5.2011 auf Basis 1-km-Rasterzellen<sup>1)</sup>



1) Die statistische Signifikanz basiert auf den Konfidenzintervallen 90 %, 95 % und 99 %. Diese sind in der kartografischen Darstellung zusammengefasst.

beispielsweise der Kriminalitätsanalyse, Epidemiologie, Wirtschaft, Demografie oder auch bei Wahlen. Die räumliche Analyse soll Aufschluss darüber geben, wie sich statistische Ereignisse räumlich verteilen und ob es eventuell Cluster, d. h. Gebiete mit einer relativen Konzentration von Merkmalen, gibt. Eine entscheidende Rolle spielt dabei auch die Visualisierung, um die räumlichen Ergebnisse zu präsentieren und darüber hinaus mögliche Einflussfaktoren gezielt vor Ort zu identifizieren, also die Frage nach dem „Warum“ zu klären.

## Digitale Karten als „Gala“ der Geoinformationen

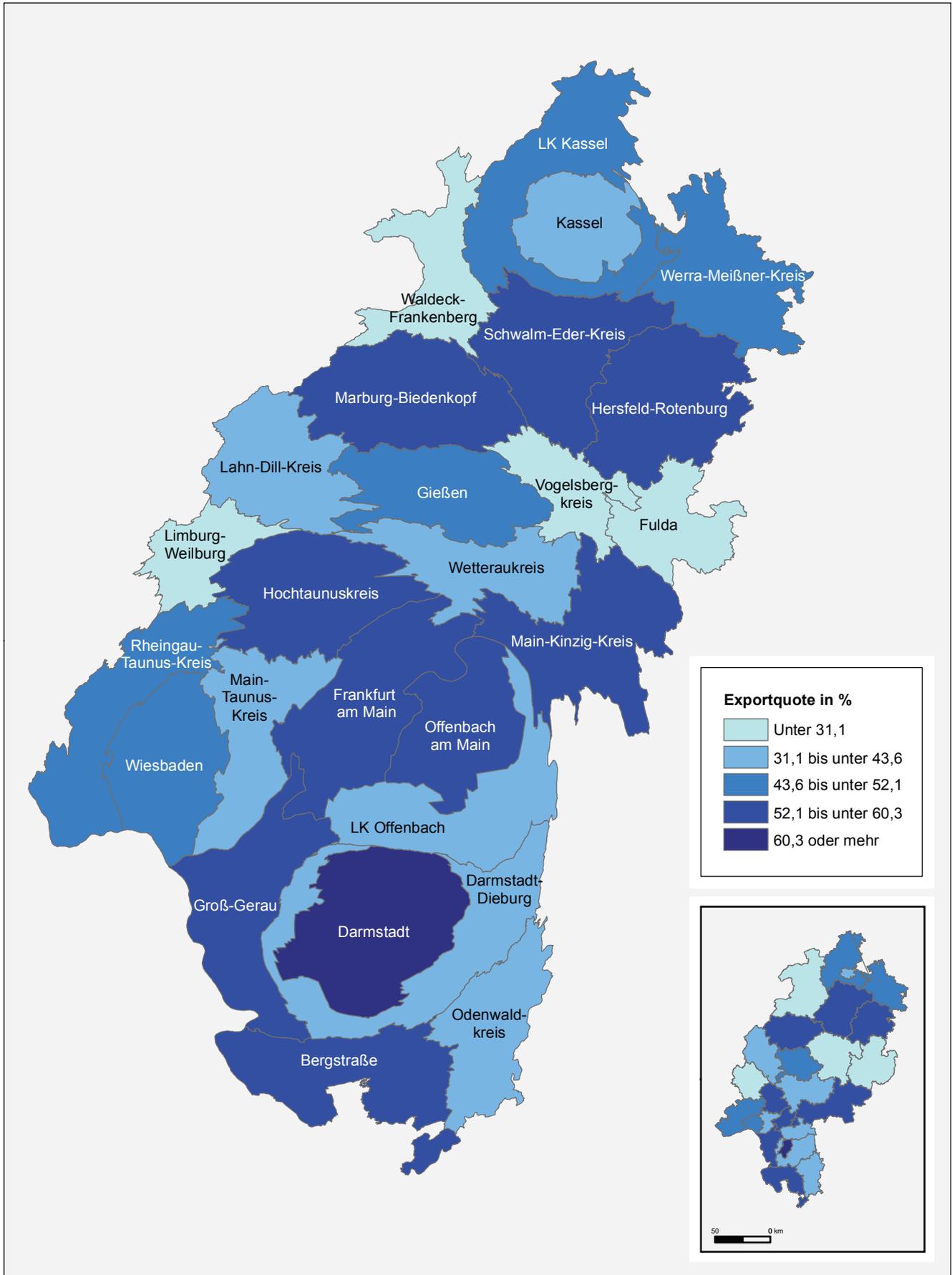
Geodaten weisen eine komplexe Datenstruktur auf und beinhalten abstrakte Sachverhalte, weshalb räumliche Muster und Trends in Form von regional aufgeschlüsselten Tabellen nicht bzw. nur schwer „sichtbar“ werden. Je größer und unstrukturierter dabei die Datenmengen sind, desto komplizierter ist die Vermittlung der relevanten Informationen. Die Aufbereitung des Datenmaterials muss daher so erfolgen, dass die räumlichen Informationen schnell und intuitiv interpretiert werden können. Eine effektive Möglichkeit, Geoinformationen zu vermitteln, ist die Visualisierung in Form von thematischen Karten. Dabei werden die statistischen Werte (z. B. Exportquote) mit den Lageinformationen (z. B. Landkreis) verlinkt, wodurch beispielsweise Vergleiche mit benachbarten Kreisen durchgeführt werden können. Am häufigsten wird die Choroplethenkarte eingesetzt, bei der die statistischen Werte klassifiziert und mit farblichen Abstufungen dargestellt werden. Die Klassifikation der Werte beeinflusst das Erscheinungsbild der resultierenden Karte und somit auch die zu vermittelnde Aussage, da unterschiedlich festgelegte Intervalle auch verschiedene räumliche Verteilungsmuster und Zusammenhänge suggerieren (Monmonier, 2005: 218 f.). Die Auswahl einer Klassifikationsmethode ist somit einerseits abhängig von der zugrundeliegenden Datenstruktur, andererseits vom Zweck sowie der zugehörigen Botschaft der Karte. Die Klassifikation mittels „Jenks Natürlicher Unterbrechungen“ ist beispielsweise eine gebräuchliche Methode für ungleichmäßig verteilte Datenwerte und geeig-

net, um Cluster aufzuzeigen. Die Werte werden dabei den Klassen so zugeordnet, dass die Unterschiede innerhalb der Klassen minimiert und zwischen den Klassen maximiert werden. Weitere gängige Klassifikationsmethoden sind „Gleiches Intervall“, „Quantil“ oder „Standardabweichung“ (Mitchell, 1999: 48 ff.). Die definierten Klassenintervalle werden in Farbskalen übersetzt, die die Aussagen der Karte unterstützen sollen (Harrower und Brewer, 2003: 27). Sequenzielle Farbverläufe von „hell“ bis „dunkel“ symbolisieren „niedrige“ bzw. „hohe“ metrische Werte und assoziieren eine logische Reihenfolge. Eine Datenreihe mit positiven und negativen Werten oder einem Schwellenwert wird anhand divergierender kontrastierender Farbschemata dargestellt, um entgegengesetzte Richtungen zu verdeutlichen. Qualitative Farbschemata suggerieren hingegen aufgrund unterschiedlicher Farben, dass kein Zusammenhang zwischen den Kategorien besteht und sind daher für Themen, wie z. B. Religion oder Landnutzungsregionen, geeignet<sup>3)</sup>.

Choroplethenkarten basieren i. d. R. auf einer Kartengrundlage mit flächentreuer Projektion und vermitteln ein räumliches Bild der dargestellten Phänomene in den entsprechenden Bezugsflächeneinheiten, wie beispielsweise die Bevölkerung in Gemeinden je km<sup>2</sup>. Viele sozioökonomische Indikatoren, wie Wahlergebnisse in den einzelnen Wahlbezirken oder die Exportquote, beziehen sich jedoch nicht direkt auf die Fläche, sondern auf die thematischen Inhalte selbst. Burgdorf (2009) empfiehlt daher eine Ergänzung flächentreuer Choroplethenkarten mit amorphen Kartogrammen, deren Bezugsflächen proportional zu einer thematischen Variablen skaliert werden. Bei Kartogrammen werden der genaue Lagebezug und die topografische Orientierung vernachlässigt und die Flächen, Formen oder Entfernungen verzerrt. Die Verzerrung wird oft in Abhängigkeit eines Wertes der Objekte vorgenommen. Abbildung 4 zeigt die Exportquote in Hessen als amorphes Kartogramm und zum Vergleich im kleinen Kartenausschnitt als klassische Choroplethenkarte. Bei einem amorphen Kartogramm, das auf dem Algorithmus von Gastner und Newman (2004) basiert, werden jene Flächen, die einen hohen Wert aufweisen

3) Bei der Auswahl von geeigneten Farbskalen unterstützt die interaktive Anwendung ColorBrewer (siehe <http://colorbrewer2.org/>).

Abbildung 4: Exportquote im Verarbeitenden Gewerbe im 1. Halbjahr 2015



© GeoBasis-DE / BKG 2016 (Daten verändert)

© Hessisches Statistisches Landesamt, Wiesbaden, 2017  
 Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.

im Verhältnis zu anderen Flächen in der Karte größer dargestellt und umgekehrt. Die Größe der Wissenschaftsstadt Darmstadt spiegelt somit hohe Datenwerte der Variable „Exportquote“ wider. Das Merkmal Größe wird hier verwendet, um gezielt den Blick auf die thematischen Indikatoren zu lenken und dadurch Aufmerksamkeit zu erreichen. Das Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR) im Bundesamt für Bauwesen und Raumordnung (BBR) veröffentlicht daher verschiedene Themen in Publikationen in Form von amorphen Kartogrammen, wie beispielsweise zur Entwicklung am Wohnungsmarkt (BBSR, 2016).

Die Visualisierung und Verbreitung von Geoinformationen erfolgt aufgrund zunehmender Vernetzung und Mobilität vermehrt in Web-Map-Applikationen, die von einfachen Kartendarstellungen mit Zoom- oder Pan-Funktionen bis hin zu komplexen interaktiven Karten reichen, bei denen der Nutzer direkt mit der Karte interagieren kann. Die Statistischen Ämter des Bundes und der Länder haben z. B. ausgewählte Ergebnisse des Zensus 2011 auf Basis von geografischen Rastern aufbereitet und bieten diese in einem interaktiven Atlas an<sup>4)</sup>. Dieser ermöglicht, neben einfachen WebGIS-Funktionen, auch die Bevölkerungszahl für nutzerdefinierte Gebiete anhand des implementierten Einwohnerrechners zu ermitteln. Eine weitere Kommunikationsform ist die sogenannte „Story Map“<sup>5)</sup>, bei der zu vermittelnde Themen und Botschaften in einer Kombination aus Karten, Grafiken, begleitendem Text und Multimedia-Elementen einer breiten Öffentlichkeit präsentiert werden. Dabei können auch komplexe und heterogene große Datenmengen klar und benutzerfreundlich aufbereitet werden, um räumliche Sachverhalte und relevante Informationen auf einen Blick zu erfassen.

### „Small is Smart“ – kleinräumige Informationen im „Rampenlicht“

Geografische Informationssysteme bieten mittels räumlicher Analysemethoden und Visualisierungstechniken effektive Möglichkeiten, sta-

tistische Daten zu veredeln und daraus „smarte“ Geoinformationen zu generieren. Dadurch können räumliche Variationen, Muster und Trends identifiziert, beschrieben und erläutert werden, um letztendlich erfolgsbringende Entscheidungen zu treffen. Voraussetzung dafür sind qualitativ hochwertige und relevante georeferenzierte Daten, die maßgeschneidert auf bestimmte Fragestellungen oder ein Projekt bereitgestellt werden. Dabei steht, neben dem Umfang, der Vielfalt und Aktualität der Daten, die verwendete räumliche Auflösung im Rampenlicht. In der amtlichen Statistik werden für ausgewählte Themenbereiche kleinräumige Daten auf Basis geografischer Raster ab einer Mindestgröße von 1 ha angeboten. Dazu regelt das E-Government-Gesetz (§ 14 E-GovG) die Georeferenzierung von statistischen Daten und Registern, wonach die Koordinaten ergänzt werden sollen. Dies ermöglicht einerseits detaillierte und präzise räumliche Analysen, andererseits spielt gerade bei räumlich hochaufgelösten statistischen Daten die Frage der Gewährleistung der Geheimhaltung eine zentrale Rolle. In diesem Kontext besteht die Möglichkeit, die kleinräumigen Daten mittels räumlich-statistischer Verfahren aufzubereiten und dem Datenkonsumenten ein auf dem Analyseergebnis basierendes kartografisches Endprodukt bereitzustellen. So veröffentlicht Statistics Canada in Kooperation mit der Canadian Community Health Survey (CCHS) beispielsweise kleinräumige Indikatoren zum Gesundheitswesen in thematischen Karten als Ergebnis einer Hot-Spot-Analyse<sup>6)</sup>, wodurch Rückschlüsse auf einzelne Personen nicht möglich sind. Eine Verschneidung von kleinräumigen Daten mit anderen Informationsinhalten aus öffentlichen oder privaten Quellen ist, im Hinblick auf maßgeschneiderte und „intelligente“ statistische Produkte für den Endnutzer, ebenfalls denkbar (Wonka, 2008: 116). Dies ermöglicht eine Erweiterung des Nutzerkreises für amtlich statistische Daten und erhöht somit auch den Marktwert von Geoinformationen.

**Sarah Scholze**; Tel: 0611 3802-282;  
E-Mail: sarah.scholze@statistik.hessen.de

4) Der interaktive Zensusatlas (inkl. Einwohnerrechner) ist online verfügbar unter <https://atlas.zensus2011.de>.

5) Beispiele für Story Maps können online unter <https://storymaps.arcgis.com/de/gallery/> eingesehen werden.

6) Thematische Karten mit Indikatoren aus dem Gesundheitswesen sind online verfügbar unter <http://www.statcan.gc.ca/pub/82-583-x/2011001/themat-eng.htm>.

## Literaturverzeichnis

- Anselin, L. (1995): Local Indicators of Spatial Association – LISA. In: *Geographical Analysis* 27(2), 93–115.
- BBSR – Bundesinstitut für Bau-, Stadt- und Raumforschung (Hrsg., 2016): *Neue Ansichten auf die Wohnungsmieten*, BBSR-Analysen KOMPAKT 08/2016, Bonn. Online unter: <http://www.bbsr.bund.de/BBSR/DE/Veroeffentlichungen/AnalysenKompakt/2016/ak-08-2016.html>, [Stand: 13.02.2017].
- Burgdorf, M. (2009): Kartogramme: aus der Form geraten oder auf den Punkt gebracht? In: *Informationen zur Raumentwicklung* (10/11), 689–699. Online unter: [http://www.bbsr.bund.de/BBSR/DE/Veroeffentlichungen/R/2009/10\\_11/Inhalt/DL\\_Burgdorf.pdf?\\_\\_blob=publicationFile&v=2](http://www.bbsr.bund.de/BBSR/DE/Veroeffentlichungen/R/2009/10_11/Inhalt/DL_Burgdorf.pdf?__blob=publicationFile&v=2), [Stand: 13.02.2017].
- Cressie, N. A. C. (1993): *Statistics for spatial data*, New York.
- Gastner, M. T. und Newman, M. E. J. (2004): Diffusion-based method for producing density-equalizing maps. In: *Proceeding of Academy for Science of the United States of America (PNAS)*, 101, 7499–7504. Online unter: <http://www.pnas.org/content/101/20/7499.full.pdf>, [Stand 03.02.2017].
- Getis, A. und Ord, J. K. (1992): The Analysis of Spatial Association by Use of Distance Statistics. In: *Geographical Analysis* 24(3), 189–206.
- Goldberg, D. (2008): *A Geocoding Best Practices Guide*, North American Association of Central Cancer Registries, Springfield.
- Goodchild, M. F. (2007): Citizens as Sensors: The World of Volunteered Geography. In: *GeoJournal* 69(4), 211–221.
- Hahmann, S. und Burghardt, D. (2012): Forschungsergebnisse zur Frage: Haben 80 % aller Daten einen Raumbezug? In: *gis.Science* (3), 101–108.
- Harrower, M. A. und Brewer, C. A. (2003): ColorBrewer.org: An Online Tool for Selecting Color Schemes for Maps. In: *The Cartographic Journal* 40(1), 27–37.
- Longley, P. et al. (2015): *Geographic Information Science and Systems*, New York.
- Mitchell, A. (1999): *The ESRI Guide to GIS Analysis, Volume 1: Geographic Patterns and Relationships*, Redlands.
- Monmonier, M. (2005): Lying with Maps. In: *Statistical Science* 20(3), 215–222.
- Scholze, S. (2015): Statistische Ergebnisse „im Quadrat“ – Geografische Raster am Beispiel des Zensus 2011. In: *Staat und Wirtschaft in Hessen* (5), 211–218.
- Wonka, E. (2008): Von der Amtsstatistik zu Geoinformationen. In: Dutter, R. (Hrsg.): *Festschrift – 50 Jahre Österreichische Statistische Gesellschaft*, Wien, 115–131.

# Das Competence Center Geoinformation im Hessischen Statistischen Landesamt

Bei dem Begriff Statistik denken die meisten an Tabellen, Schaubilder, Verteilungen, Regressionen, Stichproben usw. Den wenigsten kommen dabei Karten oder räumliche Auswertungen in den Sinn. In der amtlichen Statistik wurde zur Visualisierung regionaler Daten in der Vergangenheit selten die Kartenform verwendet und noch seltener führte man räumlich-statistische Auswertungen vorgenommen. Im Hessischen Statistischen Landesamt (HSL) wurde 2016 das Competence Center Geoinformation gegründet, welches sich mit der kartografischen Aufbereitung, Analyse und Darstellung von kleinräumigen Daten beschäftigt. **Von Philip Graze**

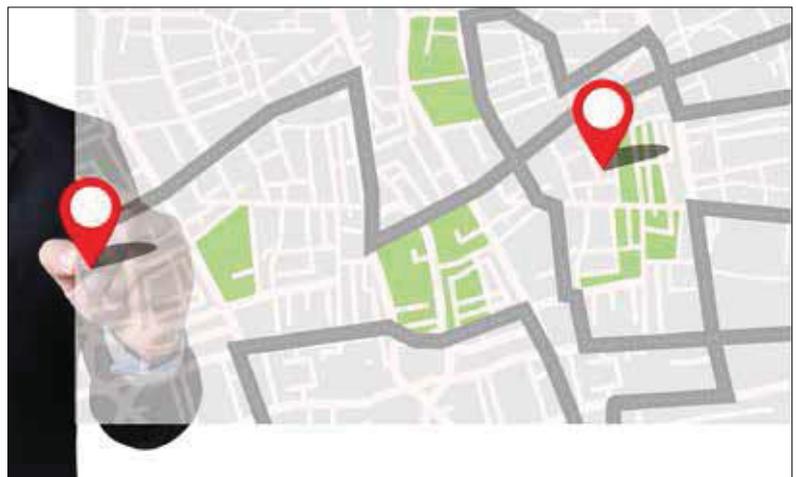
## Bedeutung von Geoinformation

In unserem Alltag haben wir ständig mit raumbezogenen Informationen, sogenannten Geoinformationen zu tun – sowohl bei Routenplanern, die ein müheloses Navigieren ermöglichen als auch bei Apps für Smartphones, die anhand von Standortinformationen unterschiedlichste Auskünfte und Hinweise bereitstellen. Über den privaten Bereich hinaus sind raumbezogene Daten eine unverzichtbare Grundlage der öffentlichen Daseinsvorsorge. Sie sind beispielsweise von besonderer Bedeutung für die öffentliche Bedarfsplanung von Krankenhäusern, Schulen und vielem mehr.

Eine Vielzahl von Aktivitäten im öffentlichen und privaten Bereich beruht auf Geoinformationen – also auf Daten, die z. B. die Umwelt, die Gesellschaft, den Verkehr oder die Wirtschaftsstruktur beschreiben.

## Das Competence Center Geoinformation stellt sich vor

Die Darstellung und Analyse von Geoinformationen werden in der amtlichen Statistik immer wichtiger. Auch von Kundenseite steigen die Erwartungen an die zeitgemäße Aufbereitung regionaler Informationen. Das Thema Geoinfor-



© adiruch na chiangmai - Fotolia.com

mation soll im HSL deshalb gezielt ausgebaut werden. Hierfür wurde im vergangenen Jahr das Competence Center Geoinformation gegründet. Es besteht derzeit aus insgesamt drei Personen. Eine Mitarbeiterin bearbeitet regionalstatistische Fragestellungen und erstellt thematische Karten für Publikationen, Pressemeldungen und die Homepage. Eine weitere Mitarbeiterin befasst sich mit komplexen räumlichen Datenverarbeitungs- und Analyseprozessen. Der Koordinator ist für konzeptionelle Fragestellungen verantwortlich und ist hausintern der erste Ansprechpartner der Fachbereiche (z. B. Wirtschaftsstatistiken, Bevölkerung etc.) für Auftragsarbeiten des Competence Centers.

Das Competence Center hat die Aufgabe, kartografische Arbeiten und räumliche Auswertungen durchzuführen, geeignete räumliche Fragestellungen zu identifizieren und qualitätssichernd im Statistikproduktionsprozess mitzuwirken.



© Robert Kneschke - Fotolia.com

#### *Zentrale Ausführung kartografischer Arbeiten und raumbasierter Auswertungen*

Im HSL werden alle geodatenverarbeitenden Prozesse – Thematische Karten, raumbasierte Auswertungen, Geokodierungen etc. – zentral vom Competence Center übernommen. Das Competence Center ist Dienstleister für die Fachbereiche. Für die Realisierung ihrer Ideen richten sich die Fachbereiche an das Competence Center. Dieses prüft und bewertet das Vorhaben hinsichtlich der Machbarkeit, der zeitlichen Umsetzung sowie des Aufwandes und hält anschließend wieder Rücksprache mit dem Fachbereich. Der Fachbereich führt dann eine Wirtschaftlichkeitsprüfung durch und entscheidet, ob das Projekt umgesetzt werden soll.

#### *Identifizieren von räumlichen Fragestellungen*

Das Competence Center ist nicht nur Dienstleister der Fachbereiche. Es geht auch auf die Fachbereiche zu, um geeignete räumliche Fragestellungen zu identifizieren und aufzubereiten. Die Fachbereiche kennen die Besonderheiten der statistischen Daten und beurteilen in Kooperation mit dem Competence Center, inwiefern eine kleinräumige Datenaufbereitung möglich sowie inhaltlich und wirtschaftlich sinnvoll ist. Weitere

zu klärende Fragen sind die Geheimhaltung und die Darstellungsform. Die Darstellungsform kann z. B. einzelne digitale Karten für eine Veröffentlichung, eine Pressemeldung oder die Homepage umfassen. Darüber hinaus können die Karten auch als eigene Anwendung auf der Homepage angeboten werden, die ein Erkunden der Daten in großem Maßstab ermöglicht.

#### *Qualitätssicherung*

Das Thema Qualitätssicherung ist sehr weitläufig und vielschichtig. Die Qualitätssicherung durch das Competence Center bezieht sich in erster Linie auf Prozesse der Datenaufbereitung und -verarbeitung. Hierbei handelt es sich um ein Themenfeld, das noch stärker ausgebaut und systematisch in bestehende Prozesse integriert werden soll. Einzelne qualitätssichernde Maßnahmen fanden beispielsweise im Bereich der Landwirtschaftsstatistik statt. In einem ersten Schritt wurden primär erfasste Anschriften anhand eines Standardverfahrens mit einer geografischen Koordinate versehen. Für einen Teil der Anschriften konnten keine Koordinaten zugewiesen werden. Das Competence Center entwickelte daher ein auf die nicht eindeutigen Anschriften der Landwirtschaftsstatistik abgestimmtes Programm, wodurch auch die restlichen Anschriften geokodiert werden konnten.

#### *Externe Auftragsarbeiten*

Bei den kartografischen Arbeiten und räumlichen Auswertungen, der Identifizierung räumlicher Fragestellung und der Qualitätssicherung handelt es sich um interne Aufträge. Entsprechende Angebote an Externe werden hingegen nicht aktiv beworben. Dennoch werden auch Anfragen nach kartografischer Aufbereitung statistischer Daten beispielsweise im Rahmen von Sonderauswertungen von außerhalb an das HSL herangetragen. Das Competence Center Geoinformation hat auch den Landeswahlleiter unterstützt und ein Shapefile<sup>1)</sup> mit Landtagswahlkreisen für die Landtagswahlen erstellt. Diese sogenannte Basisgeometrie ist unverzichtbar zur kartografischen Darstellung der hessischen Landtagswahlergebnisse. Außerdem hat das Competence Center Geoinformation dem Landeswahlleiter kartografische Übersichten bereitgestellt, um

1) „Als ein einfaches, nicht topologisches und vektorbasiertes Binär-Format für Geodaten dienen Shapefiles zur Darstellung geographischer Daten und zum Speichern der geographischen Position und der Attributinformationen von geographischen Objekten. [...] In einem Shapefile können geographische Objekte als Punkte, Linien oder Polygone (Flächen) – jedoch jeweils immer nur ein Elementtyp (z. B. entweder Punkte oder Flächen) – dargestellt werden. Durch topologische Verknüpfungen stehen alle drei Elemente in geometrischen bzw. mathematischen Bezug zueinander. Die Topologie ist dabei dank Georeferenzierung absolut messbar und innerhalb des Systems vom Nutzer bestimmbar“ (GISwiki: Shapefile. <http://gis.wiki.fau.de/mediawiki/index.php/Shapefile> (abgerufen am 9.3.2017)).

mögliche Änderungen von Wahlkreisabschnitten zu identifizieren. Diese und ähnliche externe Anfragen werden ebenfalls vom Competence Center bearbeitet.

## Neue Handlungsfelder

Die im vorherigen Abschnitt aufgeführten Aufgaben werden kontinuierlich und dauerhaft fortgeführt. Daneben sind, mit einem eher pilothaften Charakter, die Einführung von digitalen Anwendungen, Web-Map-Services (WMS) und Werkzeugen zur Geokodierung angedacht.

Im Bereich der digitalen Anwendungen befasst sich das Competence Center Geoinformation aktuell mit der Aufbereitung der Straßenverkehrsunfälle. Die Straßenverkehrsunfälle sind hierzu zum einen besonders geeignet, weil die Daten der einzelnen Unfälle bereits Geokoordinaten enthalten. D. h. die Geokoordinaten müssen nicht aufwändig ermittelt werden. Zum anderen handelt es sich um ein Thema zu dem viele Kunden einen Bezug haben - ob als Autofahrer, Fußgänger oder Radfahrer. Geplant ist eine digitale Anwendung auf der Homepage des HSL, die dem Nutzer ein individuelles und kleinräumiges Erkunden des Unfallgeschehens in Hessen ermöglicht.

Im Anschluss sollen weitere, für digitale Anwendungen geeignete Themenfelder identifiziert werden. Aus heutiger Sicht vielversprechend sind die Bereiche Gesundheit und Bautätigkeit.

Die in den digitalen Anwendungen enthaltenen Geoinformationen sollen nicht nur auf der Homepage in einer proprietären Umgebung mit eingeschränkten Möglichkeiten der Weiterverwendung durch die Anwender nutzbar sein. Vielmehr ist uns daran gelegen, dass die Nutzer die Informationen mit ihren Fachdaten kombinieren und weiterverarbeiten können. D. h. es sollen Möglichkeiten geschaffen werden, die Geoinformationen am eigenen Arbeitsplatz weiterverarbeiten zu können oder die Geoinformationen in vorhandenen Portalen und Viewern, wie z. B. dem Geoportal Hessen, mit anderen Fachdaten zusammen zu spielen. Die Realisierung erfolgt über sogenannte WMS. Der WMS ist ein web-basierter Kartendienst, der auf der Basis von Geoinformationen einen Kartenausschnitt gene-

riert und über das Web zur Verfügung stellt. Der WMS liefert nicht die Geoinformationen selbst, sondern lediglich ein visuelles Bild, beispielsweise im Format PNG, GIF oder JPEG. Die Funktionalität eines WMS beschränkt sich demnach auf die visuelle Darstellung von Geoinformationen in Form statischer Karten oder Bilder.



© Dreaming Andy - Fotolia.com

Egal, ob es um Fragen der Qualität oder der Analyse geht, bei der Arbeit mit kleinräumigen Daten werden immer wieder Geokoordinaten benötigt. Geokoordinaten ermöglichen die punktgenaue Verortung von Ereignissen anhand von X- und Y-Koordinaten im Raum. Gerade bei Arbeiten abseits der administrativen Einheiten, wie Landkreise oder Gemeinden, sind die Geokoordinaten essentiell. In den Daten der amtlichen Statistik sind die Geokoordinaten meist nicht standardmäßig enthalten. Jedoch enthalten viele Fachdaten zu Erhebungszwecken bereits Adressen. Eine Anschrift ermöglicht ebenfalls eine punktgenaue Verortung von Ereignissen oder Sachverhalten im Raum. Und hier setzt die so genannte Geokodierung an: Vorhandene Adressen werden mittels Referenzdatenbeständen um die Geokoordinaten ergänzt. Dabei handelt es sich um ein mehrstufiges, hochkomplexes Vorgehen, weshalb eine Eigenentwicklung durch das HSL unter wirtschaftlichen Aspekten nicht zielführend ist. Eine anerkannte Anwendung ist der BKG-Geocoder, der von der Arbeitsgemeinschaft der Vermessungsverwaltungen entwickelt und vom Bundesamt für Kartographie und Geodäsie (BKG) bereitgestellt wird. Die Nutzung dieses Werk-

zeugs durch das HSL ist seit Kurzem möglich. Dafür ist eine Registrierung beim Hessischen Landesamt für Bodenmanagement und Geoinformation notwendig. Dadurch kann das Competence Center ein erprobtes Werkzeug nutzen, um automatisiert geeignete Datenbestände zu georeferenzieren.

### Fazit

Vor dem Hintergrund eines stetig steigenden Interesses an kleinräumigen Analysen sowie kartografischen Darstellungen, ist der stärkeren Nutzung von Geoinformationen im HSL durch die Einrichtung des Competence Centers Geoinformation Rechnung getragen worden.

Von einer intensiveren Nutzung der Geoinformationen profitieren in erster Linie die Kunden des HSL. Die Produktpalette des Statistischen

Landesamtes wird zunehmend um Karten und räumliche Informationen erweitert, die es dem Nutzer ermöglichen, statistische Informationen räumlich zu sehen, zu analysieren und weiterzuverarbeiten.

Es gibt jedoch noch viel Gestaltungsspielraum für den Einsatz von Geoinformationen, sowohl für externe Interessenten als auch für Projekte des HSL selbst. Denkbar wäre z. B. die Einteilung von Interviewerbezirken für Haushalbefragungen, der Aufbau und die Pflege von Anschriftenbeständen und Wohnraumprüfungen im Rahmen des Zensus 2021 oder weiterführende räumlich-statistische Analysen.

**Philip Graze;** Tel: 0611 3802-810;  
E-Mail: philip.graze@statistik.hessen.de

# Hessische Kreiszahlen



## Ausgewählte Daten für Landkreise und kreisfreie Städte

■ Für Strukturanalysen und Kreisvergleiche bringt diese Veröffentlichung zweimal jährlich reichhaltiges Material

über jeden Landkreis aus allen Bereichen der amtlichen Statistik. Einiges Kreismaterial wird speziell für diese Veröffentlichung aufbereitet.

Jeder Band enthält zusätzlich Daten für die kreisangehörigen Gemeinden mit 50 000 oder mehr Einwohnern und für den Regionalverband FrankfurtRhein-Main. Ein Anhang bietet ausgewählte Daten im Zeitvergleich.

Format DIN A 4, Umfang 70 bis 90 Seiten, kartoniert.

Mit Schaubildern zu ausgewählten Themenbereichen des Inhalts.

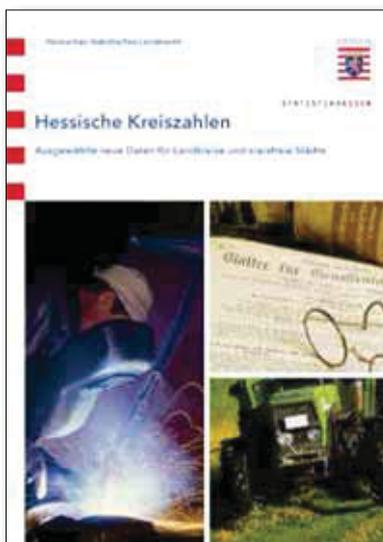
Die Hessischen Kreiszahlen können als PDF-Datei mit Excel-Anhang im Internet unter <https://statistik.hessen.de> kostenlos heruntergeladen werden.

STATISTIKHESSEN



W-160-17

Aktuelle Angaben über die neuesten Ausgaben sowie Preise finden Sie im Internet oder nehmen Sie direkt Kontakt mit uns auf. Bestellen Sie bei Ihrer Buchhandlung oder direkt bei uns. Hessisches Statistisches Landesamt, 65175 Wiesbaden  
Telefon: 0611 3802-950 · Fax: 0611 3802-992  
E-Mail: [vertrieb@statistik.hessen.de](mailto:vertrieb@statistik.hessen.de)  
Internet: <https://statistik.hessen.de>



# Der Einsatz von Small-Area-Verfahren – erste Erfahrungen mit Area-Level-Schätzungen

*Unter dem Begriff Small-Area-Verfahren versteht man gemeinhin Methoden, die darauf abzielen, aus stichprobenbasierten Erhebungen belastbare statistische Ergebnisse in tiefer fachlicher oder räumlicher Gliederung abzuleiten. Eines dieser Verfahren, das Area-Level-Modell, wird im Rahmen einer Monte-Carlo-Simulation zur Schätzung der Einzelhandelsumsätze auf Ebene der Landkreise und kreisfreien Städte eingesetzt. Die Schätzergebnisse werden den Ergebnissen der klassischen Schätzmethode, der direkten Schätzung, gegenübergestellt. Die Simulation zeigt, dass Area-Level-Schätzungen sehr häufig, aber nicht immer, genauere Schätzergebnisse liefern als die direkte Schätzung. Da beim Area-Level-Ansatz auch der Varianz der direkten Schätzer eine bedeutsame Rolle zukommt, werden zudem unterschiedliche Ansätze zur Schätzung der Varianz des direkten Schätzers untersucht und miteinander verglichen. Von Dr. Peter Gottfried*

In jüngerer Zeit häufen sich im Hessischen Statistischen Landesamt (HSL) die Nutzeranfragen nach regional und fachlich sehr tief gegliederten Daten. Die Gründe für diese Entwicklung sind vielfältig. Regional- und wirtschaftspolitische Maßnahmen werden vermehrt Evaluationen unterzogen, welche sehr detailliertes Datenmaterial erfordern. Auch wirtschaftliche Interessenverbände sind zunehmend bestrebt, ihre wirtschaftliche und gesellschaftliche Bedeutung für Land und Regionen in „harten Zahlen“ darzustellen. Dies sind nur zwei mögliche Motive für ein anhaltend hohes bzw. stetig steigendes Nutzerinteresse an tief gegliederten Daten. Diese Entwicklung stellt das HSL allerdings vor eine neue Herausforderung, da derart detaillierte statistische Angaben, sofern sie auf Stichproben-Erhebungen basieren, mit der herkömmlichen Methodik nicht in der gewohnten Zuverlässigkeit der Daten bereitgestellt werden können.

Small-Area-Verfahren bieten hier die Möglichkeit statistisch belastbarere Ergebnisse zu gewinnen. In der Wissenschaft sind die Verfahren mittlerweile fest etabliert und werden stetig weiterentwickelt. Sie fanden, international betrachtet, auch Einzug in die of-

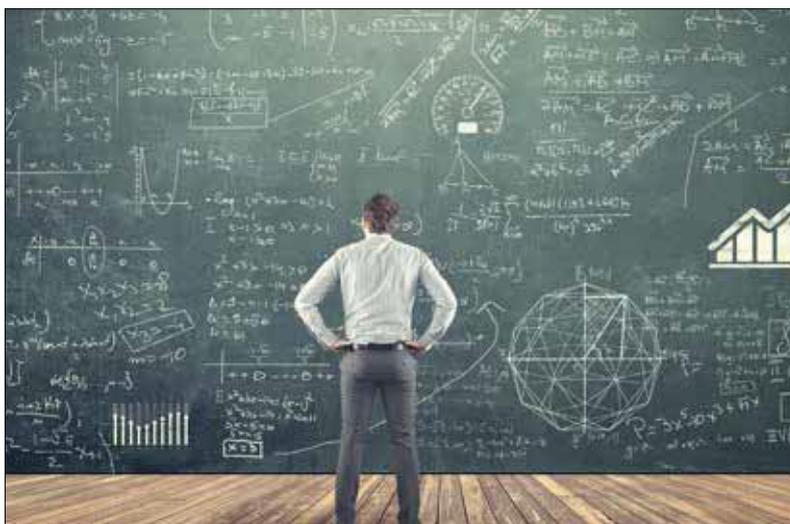
fizielle Statistik<sup>1)</sup>, werden aber bislang in der deutschen amtlichen Statistik eher zurückhaltend eingesetzt. Zu den wenigen Ausnahmen zählt die Arbeit von Dieterle (2011)<sup>2)</sup>, die zur Schätzung des Schweinebestandes auf Kreisebene neben dem klassischen Vorgehen auch einen synthetischen und einen zusammengesetzten Schätzer einsetzte und die Ergebnisse einander gegenüberstellte. In ähnlicher Weise befasst sich der vorliegende Beitrag mit Area-Level-Schätzungen, einem weiteren Small-Area-Verfahren<sup>3)</sup>. Als Testanwendung sollen die Umsätze des Einzelhandels auf Ebene der Landkreise und kreisfreien Städte geschätzt werden. Im Kern steht dabei ebenfalls die Frage, inwieweit sich mit Area-Level-Schätzungen belastbare kleinräumige Ergebnisse erzielen lassen. Zur Untersuchung der Frage wird auf Simulationstechnik zurückgegriffen. Da stichprobenbasierte Ergebnisse zufälligen Schwankungen unterliegen, wird das Vorgehen – Stichprobenziehen nebst Auswertung derselben – mehrfach wiederholt, um so ein umfassenderes Bild von der Leistungsfähigkeit des Verfahrens (im vorliegenden Anwendungsfall) zu erhalten. Der Beitrag ist somit eher methodischer Natur und richtet sich vornehmlich an Leser, die

1) Siehe beispielsweise Gonzales, M. E., Placek, P.J. und Scott, C. (1996): Synthetic Estimation of Followback Surveys at the National Center for Health Statistics. In Schaible, W. L. (Herausgeber): Indirect Estimators in U.S. Federal Programs.

2) Dieterle, M. (2011): „Die Schätzung regionaler Daten mithilfe von Small Area-Schätzmethoden“, Statistisches Bundesamt (Herausgeber): Wirtschaft und Statistik, Dezember 2011. Seite 1212-1218.

3) Eine kurze Übersicht zu Small-Area-Verfahren findet sich ebenfalls in Dieterle (2011).

vielleicht ihrerseits erwägen, detaillierte statistische Ergebnisse bereitzustellen und deshalb an Erfahrungsberichten mit einschlägigen Schätzverfahren interessiert sind, oder einfach Interesse dafür haben, was mit dem betrachteten Verfahren erreicht werden kann. Aufgrund dieser Intention sind im Textverlauf an einigen Stellen auch kurze Anmerkungen dazu eingefügt, wie das für die Berechnungen verwendete statistische Softwarepaket SAS eingesetzt wurde, um die jeweiligen Ergebnisse zu erzielen.



© allvision - Fotolia.com

Im nächsten Abschnitt wird das Testszenario etwas eingehender erläutert. Danach folgen die Ergebnisse der direkten Schätzung nebst einer eingehenderen Betrachtung der Varianzschätzung als Basis für die Beurteilung der Zuverlässigkeit der zuvor gewonnenen Schätzergebnisse. In Abschnitt 4 wird das Area-Level-Verfahren vorgestellt. Dem folgen die Präsentation der Schätzergebnisse sowie eine kurze Zusammenfassung und ein Fazit.

### Das Testszenario

Ein naheliegender Ansatzpunkt für das Vorhaben, die Umsätze der hessischen Einzelhandelsunternehmen auf Kreisebene herunter zu brechen, wäre die Jahrerhebung im Handel. Aus dieser Stichprobe ließe sich eine direkte Schätzung der gesuchten Umsätze als gewichtete

Summe der Umsätze aller der WZ-Abteilung 47 zugeordneten hessischen Unternehmen vornehmen, mit den in der Stichprobe angegebenen Hochrechnungsfaktoren als Gewichte. Dieses nicht nur in der amtlichen Statistik sehr weit verbreitete Vorgehen ist, nebenbei bemerkt, mit der eingangs erwähnten herkömmlichen Methodik gemeint. Die formellere Bezeichnung des soeben beschriebenen Schätzverfahrens lautet „Horvitz-Thomson-Schätzer“ oder „direkter Schätzer“. Des Weiteren böte es sich an, zur Durchführung der Area-Level-Schätzung zusätzliche Informationen aus dem Statistischen Unternehmensregister (URS) heranzuziehen. Schließlich stellt letzteres die Auswahlgrundlage für die Jahrerhebung im Handel dar. Ein Vergleich der beiden Schätzergebnisse wäre allerdings von einer Reihe von Faktoren beeinträchtigt, die nicht zwangsläufig die Schätzgüte der beiden Verfahren reflektieren<sup>4</sup>. Ferner wäre einer Wiederholung der Schätzungen mit einer anderen Zufallsauswahl von Einheiten nicht möglich und vor allem hätte man keine Anhaltspunkte dafür, wie stark die jeweiligen Ergebnisse vom tatsächlichen Wert abweichen.

Aus diesen Gründen wurde ein etwas anderer Weg beschritten: Als Ausgangspunkt diente die Registerkopie des URS 2014 für Hessen. Alle darin enthaltenen Unternehmen, die der WZ-Abteilung 47 zugeordnet sind und einen steuerbaren Umsatz größer 0 aufweisen, bilden die Grundgesamtheit oder Auswahlgrundlage für den Test. Sie umfasst 26 955 Einzelhandelsunternehmen. Für diese Grundgesamtheit lassen sich die nach den 26 hessischen Landkreisen und kreisfreien Städten zusammengefassten Umsätze unmittelbar errechnen. Diese Umsätze repräsentieren die tatsächlichen Werte, die es im Folgenden über eine Stichprobe aus der Grundgesamtheit zu schätzen gilt.

Für die Stichprobe wurde die Grundgesamtheit nach den Merkmalen *Branche* und *Umsatzgrößenklasse* geschichtet. Die Brancheneinteilung erfolgte nach WZ-Gruppen (Dreisteller), wobei die Gruppe 47.7 (Einzelhandel mit sonstigen Gütern) zusätzlich in Apotheken und Nicht-Apotheken untergliedert wurde, sodass insgesamt 10 Branchenschichten entstanden. Jede davon wurde in 5 Umsatzgrößenklassen unterteilt. Die

4) Z. B. die Frage bezüglich der Konsistenz beider Quellen: Unternehmensregister auf der einen und Jahrerhebung im Handel auf der anderen Seite.

### Box 1: Stichprobenfehler, Schichtung, relativer Standardfehler und Mean Squared Error (MSE)

Statistische Ergebnisse, die auf der Auswertung von Stichproben basieren, sind im Grundsatz mit Fehlern behaftete Schätzungen des tatsächlichen Werts. Schließt man systematische Fehler aus, die über die Auswahl der Einheiten in die Stichprobe oder durch die Verwendung eines bestimmten Schätzverfahrens entstehen können, so sind die Schätzwerte immer noch mit zufallsbedingten Fehlern behaftet. Je nachdem, welche Einheiten zufällig in die Stichprobe gezogen werden, fällt der Schätzwert von Stichprobe zu Stichprobe unterschiedlich aus. Diese zufallsbedingten Fehler werden auch als **Stichprobenfehler** bezeichnet.

Bei einfachen Zufallsstichproben hängt die Höhe der zufälligen Schwankungen in den Schätzergebnissen im Wesentlichen von zwei Faktoren ab: der Heterogenität der Einzelwerte in der Grundgesamtheit und dem Stichprobenumfang. Je unterschiedlicher die Einzelwerte in der Grundgesamtheit sind, desto höher fällt tendenziell auch der Stichprobenfehler aus. Je höher dagegen der Stichprobenumfang gewählt wird, desto geringer schwanken in der Regel die Schätzergebnisse. Eine Erhöhung des Stichprobenumfangs ist aber kostspielig, da z. B. mehr Einheiten zu befragen sind, die Ergebnisse zu plausibilisieren sind etc. Daher sucht man günstigere Alternativen, um den Stichprobenfehler möglichst klein ausfallen zu lassen. Eine davon ist die Verwendung von **geschichteten Stichproben**. Dabei wird die Grundgesamtheit in homogenere Teilmengen unterteilt und aus jeder dieser Teilmengen eine Teilstichprobe gezogen. Nach welchen Kriterien die Teilmengen gebildet werden, hängt von den jeweiligen Untersuchungszielen ab. Die bundesweite Jahrerhebung im Handel ist beispielsweise nach Bundesländern, Branchen (Wirtschaftszweige (WZ)-Gruppen) und Umsatzgrößenklassen geschichtet. Stabilere Ergebnisse im Vergleich zur einfachen Zufallsstichprobe ließen sich bereits dadurch erzielen, dass aus allen Schichten derselbe Anteil von Einheiten in die Stichprobe gezogen wird, d. h. für alle Teilstichproben derselbe **Auswahlsatz** verwendet wird. Durch Variation der Auswahlsätze, Totalerfassung der Einheiten in der höchsten Umsatzgrößenklasse und kontinuierlich abnehmenden Auswahlsätzen für die darunterliegenden Umsatzschichten, ist ein noch stabileres Schätzergebnis (bei insgesamt gleichem Stichprobenumfang) erzielbar. Der unterschiedlichen Besetzungszahl der einzelnen Schichten wird bei der Auswertung der Stichprobe durch entsprechende **Gewichtung** der Einzelwerte Rechnung getragen. In der Regel entspricht das jeweilige Gewicht dem Kehrwert des Auswahlsatzes.

Als Maß für den Stichprobenfehler dient die **Fehlervarianz** der Schätzwerte  $\hat{\theta}$ . Sie beschreibt die erwartete quadratische Abweichung der Schätzwerte vom Erwartungswert der Schätzung  $E(\hat{\theta})$ :

$$V(\hat{\theta}) = E\left(\left(\hat{\theta} - E(\hat{\theta})\right)^2\right).$$

Da die Fehlervarianz sehr hohe Werte annehmen kann und aufgrund ihrer quadratischen Dimension sperrig zu interpretieren ist, wird an ihrer Stelle häufig der **Standardfehler** als Maßgröße herangezogen. Er errechnet sich einfach als Wurzel aus der Fehlervarianz und beschreibt damit die mittlere Abweichung der einzelnen Schätzwerte von ihrem Erwartungswert. Noch etwas griffiger und in der amtlichen Statistik am häufigsten verwendet ist der **relative Standardfehler**, bei dem der Standardfehler ins Verhältnis zum Schätzwert gesetzt ist:

$$RS(\hat{\theta}) = \frac{\sqrt{V(\hat{\theta})}}{\hat{\theta}}.$$

Er stellt ein dimensionsloses und einfach zu interpretierendes Maß dar. Ein relativer Standardfehler von beispielsweise 3 % gibt an, dass die einzelnen Schätzwerte im Durchschnitt um 3 % (des mittleren Schätzwerts) vom mittleren Schätzwert abweichen. Bei **erwartungstreuen** Schätzungen entspricht der Erwartungswert des Schätzers dem tatsächlichen Wert. Dann ließe sich – auf das Beispiel bezogen – sagen, dass die Einzelschätzungen durchschnittlich um 3 % vom tatsächlichen Wert abweichen.

Dem Qualitätsbericht der Jahrerhebung im Handel zufolge lag der relative Standardfehler für das Merkmal Umsatz im Handel insgesamt bei 0,4 %<sup>5)</sup>. Je tiefer die Ergebnisse gegliedert sind, desto höher fällt jedoch der relative Standardfehler trotz Schichtung aus. In der amtlichen Statistik ist es gängige Praxis, Ergebnisse in Klammern zu setzen, wenn der relative Standardfehler zwischen 10 % und 20 % liegt und auf den Ergebnisausweis zu verzichten, wenn der relative Standardfehler 20 % übersteigt<sup>6)</sup>.

Bei nicht erwartungstreuen Schätzfunktionen, wie etwa der hier betrachteten Area-Level-Schätzung, wird üblicherweise ein umfassenderes Maß – der **MSE** – verwendet. Er kennzeichnet die mittlere quadratische Abweichung der Schätzwerte vom tatsächlichen Wert:

$$MSE(\hat{\theta}) = E((\hat{\theta} - \theta)^2) = (B(\hat{\theta}))^2 + V(\hat{\theta}).$$

Er lässt sich in die Verzerrung  $B(\hat{\theta})$ , die Abweichung des mittleren Schätzwerts vom tatsächlichen Wert und die Fehlervarianz zerlegen. Das Maß umfasst somit neben der Streuung der einzelnen Schätzwerte ebenso die Treffsicherheit bzw. Genauigkeit des Schätzers. Bei erwartungstreuen Schätzern reduziert sich der MSE auf die Fehlervarianz. Analog zu den vorangehenden Maßen ist auch hier die Verwendung des **Root Mean Squared Error (RMSE)** – die Wurzel aus dem MSE – gebräuchlicher. Als relative Größe ergibt sich der **Relative Root Mean Squared Error (RRMSE)** nach Division des RMSE durch den tatsächlichen Wert:

$$RRMSE(\hat{\theta}) = \frac{\sqrt{MSE(\hat{\theta})}}{\theta}.$$

Einteilung erfolgte branchenspezifisch, wobei die Klassengrenzen aus der Jahrerhebung im Handel 2015 übernommen wurden. Gleiches gilt für die Stichproben-Auswahlsätze in den einzelnen Schichten<sup>7)</sup>. Insgesamt ist die Vorgehensweise damit sehr eng an das Vorgehen bei der Jahrerhebung angelehnt, ist aber kontrollierbarer und transparenter, da u. a. der Einfluss des Konzepts der Stichprobenrotation und die Problematik rund um echte und unechte Antwortausfälle ausgeblendet sind. Das Szenario entspricht gewissermaßen der Vorstellung, dass die im Unternehmensregister ausgewiesenen Größen alle zeitnah und valide sind und die Ergebnisse der Jahrerhebung ebenfalls unmittelbar ohne zeitliche Verzögerung vorliegen. Ferner kann die Stichprobenziehung und -auswertung beliebig oft wiederholt werden, wodurch ein umfassenderes Bild von der Qualität der Schätzverfahren entsteht.

### Direkte Schätzung

Um die Güte der klassischen Vorgehensweise in diesem Szenario beurteilen zu können, wurde zunächst eine direkte Schätzung der Einzelhandelsumsätze auf Ebene der Landkreise und kreisfreien Städte vorgenommen. Tabelle 1

zeigt die Ergebnisse. Als Vergleichsgröße sind in der ersten Tabellenspalte die „tatsächlichen Werte“ aufgelistet. Sie ergeben sich durch Aufsummieren der Umsatzwerte nach Landkreisen und kreisfreien Städten über alle Einheiten der Grundgesamtheit hinweg. Die Spalten 2 und 3 stellen die Schätzergebnisse aus einer einzelnen Stichprobenziehung dar. Für die Ziehung der Stichprobe wurde die SAS-Prozedur SURVEYSELECT verwendet, für die Auswertung die Prozedur SURVEYMEANS. Die Schätzwerte errechnen sich als gewichtete Summe der Umsätze der Unternehmen (des/der jeweils betrachteten Landkreises bzw. kreisfreien Stadt) in der Stichprobe, wobei die aus dem Stichprobendesign gewonnenen Hochrechnungsfaktoren<sup>8)</sup> als Gewichte dienen. Nähere Hinweise zur Rolle der Gewichte wie auch zur Berechnung der im weiteren Verlauf des Beitrags herangezogenen Maße zur Beurteilung der Güte der Schätzung finden sich in der Box 1 „Stichprobenfehler, Schichtung, relativer Standardfehler und Mean Squared Error (MSE)“.

Die in Spalte 3 aufgelisteten relativen Standardfehler werden ermittelt, indem man die geschätzte Standardabweichung zu den Umsatz-Schätzwerten ins Verhältnis setzt. Die pro-

5) Statistisches Bundesamt: Qualitätsbericht, Jahrerhebung im Handel sowie in der Instandhaltung und Reparatur von Kfz 2014, Wiesbaden 2016.

6) Siehe beispielsweise: Statistisches Bundesamt: Qualitätsbericht, Einkommens- und Verbrauchsstichprobe EVS 2013, Wiesbaden 2016, Seite 8.

7) Die hierzu notwendigen Informationen wurden dankenswerterweise vom Statistischen Bundesamt zur Verfügung gestellt.

8) Die Hochrechnungsfaktoren sind im vorliegenden Fall einfach der Kehrwert der jeweiligen Stichproben-Auswahlsätze.

**Tabelle 1: Ergebnisse der direkten Schätzung der Einzelhandelsumsätze**

Kreisfreie Stadt (St.) Landkreis	Tatsächlicher Umsatz	Einzelschätzung		10000 Schätzungen	
		Schätzwert	Geschätzter relativer Standard- fehler	Mittlerer Schätzwert	Relativer Standard- fehler
	1000 €	1000 €	%	1000 €	%
	1	2	3	4	5
Darmstadt, Wissenschaftsst.	605 790	529 384	18,1	605 112	8,6
Frankfurt am Main, St.	2 952 605	2 936 499	15,5	2 952 720	3,5
Offenbach am Main, St.	296 135	426 115	17,5	296 509	14,3
Wiesbaden, Landeshauptst.	936 084	1 043 615	14,9	937 009	7,2
Bergstraße	838 543	826 490	16,7	838 883	7,8
Darmstadt-Dieburg	608 199	633 462	12,7	608 434	10,5
Groß-Gerau	1 163 169	1 175 765	41,8	1 162 882	5,6
Hochtaunuskreis	642 328	702 669	14,6	642 083	10,6
Main-Kinzig-Kreis	1 974 214	2 090 032	25,4	1 973 569	4,3
Main-Taunus-Kreis	774 718	706 918	20,2	774 151	7,7
Odenwaldkreis	275 050	327 274	24,2	275 194	14,5
Offenbach	921 122	711 237	21,7	921 981	8,0
Rheingau-Taunus-Kreis	322 264	283 851	15,6	322 878	15,6
Wetteraukreis	1 769 164	1 803 192	33,0	1 769 441	4,3
Gießen	1 120 724	1 012 148	20,6	1 121 076	6,8
Lahn-Dill-Kreis	762 324	825 895	15,2	761 299	9,6
Limburg-Weilburg	459 798	395 998	16,2	459 312	12,4
Marburg-Biedenkopf	729 466	770 354	13,5	729 461	9,8
Vogelsbergkreis	310 176	314 263	24,0	310 355	15,3
Kassel, documenta-St.	725 974	734 871	16,6	726 451	9,2
Fulda	2 085 578	2 122 960	47,0	2 085 811	3,4
Hersfeld-Rotenburg	411 452	497 515	23,2	410 576	14,0
Kassel	630 557	773 808	18,8	630 951	11,4
Schwalm-Eder-Kreis	1 689 450	1 600 101	45,3	1 688 645	3,9
Waldeck-Frankenberg	495 914	428 535	15,6	496 192	12,2
Werra-Meißner-Kreis	286 819	324 544	19,7	286 089	15,5

zentualen Angaben beziehen sich somit auf die geschätzten Umsatzwerte. Selbst wenn man berücksichtigt, dass Kreiszugehörigkeit kein Schichtungsmerkmal darstellt und die Standardfehler für ungeschichtete Auswertungsvariablen tendenziell größer ausfallen als für geschichtete, erscheinen die aufgelisteten Werte überraschend hoch. Legt man die in Box 1 angesprochene gängige Praxis bezüglich des Ausweises von unsicheren Werten auf die Tabellenwerte der Einzelschätzung an, so würde man 9 von 26 Schätzwerten durch einen Schrägstrich ersetzen, da der relative Standardfehler 20 % übersteigt und die übrigen Werte, deren Standardfehler zwischen 10 % und 20 % liegt, in Klammern setzen. Ferner überrascht, dass diejenigen Schätzwerte, deren relative Standardfehler mit über 40 % am

höchsten eingeschätzt werden, doch relativ nahe an den tatsächlichen Werten liegen, mitunter um weniger als 2 % von diesen abweichen.

Die beiden letzten Tabellenspalten zeigen die Ergebnisse der Monte-Carlo-Simulation. Simuliert wurden 10 000 Stichprobenziehungen mit anschließender direkter Schätzung der Umsätze. Der in Spalte 4 ausgewiesene mittlere Schätzwert ist das arithmetische Mittel aus allen Einzelschätzungen. Durch Vergleich der Mittelwerte mit den tatsächlichen Werten erkennt man unschwer die der direkten Schätzung inhärente Eigenschaft der Erwartungstreue. Die mittleren Schätzwerte weichen um weniger als 0,2 % vom Vorgabewert ab. Damit ist auch die aus den Einzelschätzungen errechnete Standardabweichung

chung eine relativ gute Approximation des tatsächlichen Standardfehlers. Nach Division der Standardfehler durch den jeweiligen mittleren Schätzwert ergeben sich wiederum die in Spalte 5 ausgewiesenen relativen Standardfehler. Offensichtlich fallen diese deutlich kleiner aus als die entsprechenden Schätzungen in Spalte 3. Nur in 15 von 26 Fällen liegt der relative Standardfehler tatsächlich über 10 % und in allen Fällen unter 20 %.

Dieses Ergebnis legt zwei Überlegungen nahe. Zum einen befindet man sich typischerweise in der Situation, die durch die beiden Spalten der Einzelschätzung reflektiert wird: man beurteilt die Güte der ermittelten Schätzwerte mit Blick auf die relativen Standardfehler. Dabei sollte man aber nicht aus den Augen verlieren, dass die Standardfehler ihrerseits nur Schätzwerte darstellen, die aus der einen verfügbaren Stichprobe abgeleitet sind. Zum anderen existieren in der Literatur eine Reihe von alternativen Schätzern für die Varianz des direkten Summenschätzers, die auf Basis unterschiedlicher vereinfachender Annahmen abgeleitet wurden. Da die Varianzschätzungen für den direkten Schätzer auch unmittelbar in die Small-Area-Schätzung eingehen und ihnen somit eine doppelte Bedeutung zukommt, erscheint es angesichts der Resultate von Tabelle 1 ratsam zu prüfen, ob im vorliegenden Fall – bei welchem Aggregate zu schätzen sind, die nicht zu der Schichtkonzeption passen – nicht geeignetere Schätzer existieren. Dies wird im nächsten Abschnitt untersucht.

### Alternative Varianzschätzer

Getestet wurden 5 verschiedene Alternativen. Eine formale Darstellung der einzelnen Schätzgleichungen befindet sich im Anhang. Die ersten beiden Alternativen gehen unmittelbar auf die allgemeine Darstellung<sup>9)</sup> der Varianz des direkten Summenschätzers unter Verwendung von Inklusionswahrscheinlichkeiten<sup>10)</sup> zurück. Variante I berücksichtigt dabei, dass die gemeinsame Inklusionswahrscheinlichkeit davon abhängt, ob die beiden betrachteten Einheiten zur selben Schicht gehören oder nicht, Variante II unterstellt hingegen durchweg Unabhängigkeit der Inklusionswahrscheinlichkeiten.

Da eine Schätzung der Einzelhandelsumsätze nach Landkreisen und kreisfreien Städten beim Stichprobendesign durch entsprechende Schichtung nicht einbezogen wurde, wird die Zahl der Stichprobeneinheiten je Verwaltungsbezirk von Stichprobe zu Stichprobe unterschiedlich ausfallen. Schätzvariante III trägt dem Umstand Rechnung, dass der Stichprobenumfang je Landkreis bzw. kreisfreier Stadt eine Zufallsvariable darstellt. Bei den Varianten IV und V handelt es sich um die in der Prozedur SURVEYMEANS implementierten Varianzschätzer. Sie kommen standardmäßig zum Einsatz, wenn kein explizites Schätzverfahren über die Option VARMETHOD aufgerufen wird. Bei Variante III wurden der Prozedur die Besetzungszahlen der Schichten mitgegeben, bei Variante IV nicht. Varianzschätzungen auf der Basis von Bootstrap-Methoden, die ebenfalls in SURVEYMEANS implementiert sind, werden in Hinblick auf die längere Rechenzeit, die diese Verfahren benötigen, nicht getestet.

Als Beurteilungsmaßstab wird der RRMSE verwendet. Je kleiner der RRMSE ausfällt, desto effizienter und damit zuverlässiger ist die Schätzung. Dabei dient die durch Simulation ermittelte Varianz des Schätzers als Näherungslösung für den tatsächlichen Wert. Die RRMSE's der einzelnen Schätzvarianten sind in Tabelle 2 aufgeführt. Augenscheinlich entsprechen sich die Varianten I und IV sowie die Varianten III und V. Dies gilt nicht nur in Bezug auf den RRMSE, die Varianten führen trotz unterschiedlicher Ansätze auch zu identischen Schätzwerten.

Hinsichtlich der Schätzgüte produzieren die Varianten I und IV noch die verlässlichsten Schätzungen, wobei auch hier die Schätzwerte für die Varianz bei einigen Landkreisen, wie etwa dem Odenwaldkreis und dem Landkreis Hersfeld-Rotenburg, um durchschnittlich mehr als 40 % von den tatsächlichen Werten abweichen können. Recht ähnliche Ergebnisse werden mit Variante II erzielt, wobei die Annahme der Unabhängigkeit zwischen den Inklusionswahrscheinlichkeiten (insbesondere auch dann, wenn zwei Einheiten aus derselben Schicht vorliegen) zu tendenziell leicht höheren Schätzwerten führt. Bei Verwendung der explizit auf ungeplante Domains ausgerichteten Variante III wie auch bei der Varianz-

9) Die formale Darstellung findet sich beispielsweise in Münnich, R. (2008): Varianzschätzung in komplexen Erhebungen. Austrian Journal of Statistics, Vol. 37, Seite 321.

10) Die Inklusionswahrscheinlichkeit kennzeichnet die Wahrscheinlichkeit dafür, dass die entsprechenden Einheiten in die Stichprobe gezogen werden.

schätzung über SURVEYMEANS ohne Angabe der Schichtenbesetzung tritt hingegen relativ häufig der Fall auf, dass die Varianz sehr stark überschätzt wird. Die in Tabelle 1 ausgewiesenen Standardfehler wurden über die Variante V bestimmt und sind hierfür ein Beispiel. Wird SAS SURVEYMEANS zur direkten Schätzung von Summen oder Mittelwerten verwendet, dann sollten, soweit möglich, der Prozedur auch Angaben zur Schichtgröße in der Grundgesamtheit und die Schichtungsmerkmale übergeben werden.

### Das Area-Level-Modell

Area-Level-Schätzer stellen eine Möglichkeit dar, gegenüber einer direkten Schätzung insbesondere dann verlässlichere Schätzwerte zu liefern, wenn die Areas in der Stichprobe nur sehr schwach (oder gar nicht) vertreten sind. Die Stabilisierung wird durch die Einbeziehung von Hilfsinformationen bzw. Hilfsvariablen erzielt. Das grundlegende Area-Level-Modell unterstellt dabei einen linearen Zusammenhang zwischen den tatsächlichen Umsatzmittelwerten  $\bar{u}_d$  und den Mittelwerten der  $p$  Hilfsvariablen

$$Z_d = (z_{d1}, \dots, z_{dp})^T, \quad \bar{u}_d = Z_d^T \beta + v_d,$$

wobei der lineare Zusammenhang durch einen kreissspezifischen Random Effekt  $v_d$  überlagert wird<sup>11)</sup>. Die  $v_d$  sind hierbei identisch und unabhängig verteilt mit Erwartungswert 0 und Varianz  $\sigma_v^2$ .

Weiterhin würde eine direkte Schätzung die Umsatzmittelwerte erwartungstreu schätzen

$$\hat{u}_d = \bar{u}_d + e_d,$$

wobei die Stichprobenfehler  $e_d$  unabhängig mit Erwartungswert 0 und Varianz  $D_d$  verteilt sind. Beides zusammengenommen ergibt das Modell

$$\hat{u}_d = Z_d^T \beta + v_d + e_d.$$

Aus diesem Zusammenhang lassen sich die Random Effekte abschätzen und damit auch die Area-Level-Schätzer bestimmen. Letztere lassen sich mit

$$\hat{u}_d^{FH} = \hat{\gamma}_d \hat{u}_d + (1 - \hat{\gamma}_d) Z_d^T \hat{\beta} \quad \text{und} \quad \hat{\gamma}_d = (\hat{\sigma}_v^2 + \hat{D}_d)^{-1} \hat{\sigma}_v^2$$

als Linearkombination der direkten Schätzer und dem Regressionsteil darstellen. Dabei werden die  $\hat{u}_d$  und  $\hat{D}_d$  üblicherweise von der direkten Schätzung übernommen, die übrigen Größen im

**Tabelle 2: Effizienz der getesteten Varianzschätzer – bewertet über Relative Root Mean Squared Errors (RRMSE)**

Kreisfreie Stadt (St.) Landkreis	Schätzvarianten				
	I	II	III	IV	V
Darmstadt, Wissenschaftsst.	40,3	41,8	312,1	40,3	312,1
Frankfurt am Main, St.	19,2	24,9	1 924,3	19,2	1 924,3
Offenbach am Main, St.	45,3	46,4	113,9	45,3	113,9
Wiesbaden, Landeshauptst.	28,6	31,2	360,7	28,6	360,7
Bergstraße	28,0	30,2	338,8	28,0	338,8
Darmstadt-Dieburg	30,9	33,1	81,7	30,9	81,7
Groß-Gerau	31,7	33,8	5 504,0	31,7	5 504,0
Hochtaunuskreis	30,9	32,8	85,1	30,9	85,1
Main-Kinzig-Kreis	22,0	25,0	3 753,9	22,0	3 753,9
Main-Taunus-Kreis	31,3	33,0	502,0	31,3	502,0
Odenwaldkreis	48,2	49,1	275,5	48,2	275,5
Offenbach	26,1	29,0	409,5	26,1	409,5
Rheingau-Taunus-Kreis	36,4	38,2	50,8	36,4	50,8
Wetteraukreis	28,4	30,7	6 587,4	28,4	6 587,4
Gießen	26,1	27,9	702,2	26,1	702,2
Lahn-Dill-Kreis	24,7	27,0	179,8	24,7	179,8
Limburg-Weilburg	35,0	36,4	67,8	35,0	67,8
Marburg-Biedenkopf	29,0	31,4	108,3	29,0	108,3
Vogelsbergkreis	42,7	43,8	158,9	42,7	158,9
Kassel, documenta-St.	32,1	34,0	182,7	32,1	182,7
Fulda	29,5	31,3	20 400,9	29,5	20 400,9
Hersfeld-Rotenburg	44,4	45,7	210,4	44,4	210,4
Kassel	29,0	30,6	229,9	29,0	229,9
Schwalm-Eder-Kreis	34,6	36,1	13 819,2	34,6	13 819,2
Waldeck-Frankenberg	32,1	33,6	72,3	32,1	72,3
Werra-Meißner-Kreis	43,0	44,1	103,4	43,0	103,4

I Näherungslösung für Standardformel mit Korrektur bei gleicher Schicht  
 II Näherungslösung für Standardformel ohne Korrektur  
 III Näherungsformel für zufällige Domains  
 IV SAS SURVEYMEANS mit Angabe der Besetzungszahlen der Schichten  
 V SAS SURVEYMEANS ohne Angabe der Besetzungszahlen

Modell geschätzt. Dieses Modell geht auf Fay und Herriot (1979)<sup>12)</sup> zurück, weshalb die Schätzer üblicherweise als FH-Schätzer  $\hat{u}_d^{FH}$  bezeichnet werden.

### Anmerkungen zum Vorgehen bei der Schätzung

Für die nachfolgenden Schätzungen wurde nur eine Hilfsvariable verwendet. Wie bereits die kurze Modellbeschreibung nahe legt, scheidet der Fay-Herriot-Schätzer wohl umso besser ab, je enger der Zusammenhang zwischen der Hilfsvariablen und der Schätzvariablen ist, d. h. je höher sie miteinander korreliert sind. Wollte man die

11) Eine detailliertere Darstellung des grundlegenden Area-Level-Modells findet sich beispielsweise in Rao, J. N. K. und Molina, I. (2015): Small Area Estimation, 2nd Edition, Seite 76ff.

12) Fay, R. E. und Herriot, R. A. (1979): Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. Journal of the American Statistical Association, Bd. 74, S. 269-277.



© rosifan19 - Fotolia.com

Einzelhandelsumsätze in den Verwaltungsbezirken mit diesem Schätzverfahren aus der Jahrerhebung schätzen, so böten sich die in der Registerkopie des URS hinterlegten steuerbaren Umsätze, entsprechend zusammengefasst, als Hilfsvariable an. Bereits auf der Ebene der Unternehmen liegt eine hohe Korrelation zwischen den jeweiligen Umsatzwerten vor. In dem hier zu Grunde gelegten Testszenario entfällt diese Möglichkeit. Deshalb wird aus den im URS ausgewiesenen steuerbaren Umsätzen eine künstliche Hilfsgröße erzeugt<sup>13)</sup>. Dieses Vorgehen hat den Vorteil, dass über die Konstruktion der Hilfsvariablen auch die Korrelation zur Schätzgröße „gesteuert“ werden kann. Der Korrelationskoeffizient zwischen den Kreismittelwerten der konstruierten Hilfsvariablen und der Zielvariablen beträgt rund 0,96. Ein ähnlich hoher Wert ist für die einschlägigen Größen aus dem URS und der Jahrerhebung zu verzeichnen.

Wie bereits zuvor wurden die Schätzungen unter Verwendung von SAS durchgeführt. Allerdings stellt SAS keine spezielle Prozedur für Area-Level-Modelle bereit. Für die Varianzschätzungen wurde die Prozedur MIXED eingesetzt, die Berechnung der Schätzwerte erfolgte im Anschluss daran in der IML (Interactive Matrix Language)-Umgebung von SAS gewissermaßen manuell. Auf eine explizite Beschreibung des Vorgehens wird verzichtet und auf den Beitrag von Mukhopadhyay und McDowell (2011)<sup>14)</sup> im SAS Global Forum verwiesen. Er beschreibt sehr

detailliert anhand von Beispielen die Umsetzung einer Reihe von Small-Area-Modellen in SAS Programmcode, unter anderem auch das hier verwendete Fay-Herriot-Modell.

### Schätzergebnisse

Die eingangs aufgeworfene und eher allgemein gehaltene Frage hinsichtlich der Belastbarkeit von Area-Level-Schätzungen gegenüber dem klassischen Verfahren wird hier durch zwei mehr pragmatisch orientierte Teilfragen konkretisiert: Wie gut werden die vorgegebenen tatsächlichen Werte geschätzt? Wie stark variieren die Schätzwerte von Stichprobe zu Stichprobe?

In Tabelle 3 sind die Ergebnisse zweier Einzelschätzungen – einer direkten Mittelwertschätzung und einer Area-Level-Schätzung – einander gegenübergestellt. Die geschätzten Umsatz-Mittelwerte wurden anschließend über die Anzahl der Unternehmen in den einzelnen Landkreisen und kreisfreien Städten in Gesamtbeträge umgerechnet. Vergleicht man die jeweiligen Schätzwerte mit dem tatsächlichen Umsatz, so zeigt sich, dass die FH-Schätzwerte bei 19 von 26 Landkreisen und kreisfreien Städten oder in rund 73 % aller Fälle näher am Vorgabewert liegen als die direkten Schätzwerte. Ausgehend von den in diesen Fällen größeren Abweichungen der direkten Schätzwerte reduziert die FH-Schätzung diese Abweichungen um durchschnittlich 50 %. Für 7 Landkreise bzw. kreisfreie Städte (23 % der Fälle) fällt die FH-Schätzung dagegen schlechter aus. Die Verschlechterung macht dabei 70 % der jetzt größeren Abweichung aus. Würde man Verbesserungen und Verschlechterungen der Schätzergebnisse miteinander verrechnen, um zu einer (zugegebenermaßen) kruden Gesamtschätzung zu gelangen, so fielen die Abweichungen der FH-Schätzwerte von den entsprechenden Vorgabewerten in dieser Stichprobe um durchschnittlich 20 % geringer aus als die Abweichungen der direkten Schätzwerte von den Vorgabewerten. Da diese durchschnittliche Verbesserung rein rechnerischer Natur ist und sich nicht gleichmäßig über die Schätzwerte erstreckt, sei die Einschätzung der Sinnhaftigkeit dieser Zusammenfassung der eigenen Beurteilung überlassen.

13) Rund 15 % der Umsatzwerte werden zufällig ausgewählt und durch die Werte einer gleichverteilten Zufallsvariablen ersetzt. Die Variable nimmt dabei Werte zwischen 20 und 600 an. Die nicht ersetzten Umsatzwerte werden um den Faktor 10 gekürzt und mit einer normalverteilten Zufallsvariablen überlagert. Die normalverteilte Zufallsvariable hat den Erwartungswert 0 und die Standardabweichung 100.

14) Mukhopadhyay, P. K. und McDowell, A. (2011): Small Area Estimation for Survey Data Analysis Using SAS Software, SAS Global Forum 2011, Paper 336-2011.

**Tabelle 3: Ergebnisse einer direkten Schätzung und einer Area-Level-Schätzung im Vergleich**

Kreisfreie Stadt (St.) Landkreis	Tatsächlicher Umsatz	Direkte Schätzung		Area-Level-Schätzung	
		Direkter Schätzwert	Geschätzter relativer Standard- fehler	FH-Schätzwert	Geschätzter RRMSE
	1000 €	1000 €	%	1000 €	%
	1	2	3	4	5
Darmstadt, Wissenschaftsst.	605 790	558 165	15,6	595 755	10,8
Frankfurt am Main, St.	2 952 605	2 924 090	7,9	2 909 474	7,1
Offenbach am Main, St.	296 135	498 477	23,1	373 230	21,2
Wiesbaden, Landeshauptst.	936 084	891 238	13,1	942 293	10,9
Bergstraße	838 543	801 097	13,2	820 773	10,9
Darmstadt-Dieburg	608 199	696 312	16,7	681 739	16,0
Groß-Gerau	1 163 169	1 060 720	15,3	1 164 562	9,5
Hochtaunuskreis	642 328	627 707	13,2	613 814	11,6
Main-Kinzig-Kreis	1 974 214	2 217 350	11,7	2 123 433	9,7
Main-Taunus-Kreis	774 718	774 199	15,3	757 028	11,3
Odenwaldkreis	275 050	166 569	19,8	215 920	10,8
Offenbach	921 122	695 411	10,1	725 102	7,4
Rheingau-Taunus-Kreis	322 264	422 758	20,6	391 253	20,8
Wetteraukreis	1 769 164	1 809 175	13,6	1 550 378	8,6
Gießen	1 120 724	1 060 180	13,8	1 097 227	9,9
Lahn-Dill-Kreis	762 324	793 258	16,6	780 463	12,6
Limburg-Weilburg	459 798	628 208	19,8	509 598	17,8
Marburg-Biedenkopf	729 466	728 825	13,2	711 992	11,0
Vogelsbergkreis	310 176	274 502	21,1	280 152	13,7
Kassel, documenta-St.	725 974	644 924	13,6	664 220	10,1
Fulda	2 085 578	2 171 946	15,5	2 151 320	10,3
Hersfeld-Rotenburg	411 452	418 768	18,8	377 803	12,0
Kassel	630 557	841 135	18,9	737 402	15,0
Schwalm-Eder-Kreis	1 689 450	1 726 630	18,6	1 322 889	7,9
Waldeck-Frankenberg	495 914	577 334	16,0	506 122	13,8
Werra-Meißner-Kreis	286 819	380 000	21,9	327 590	17,1

Für eine erste Einschätzung der Stabilität der Schätzwerte sind in den Spalten 3 und 5 von Tabelle 3 die relativen Standardfehler der direkten Schätzwerte und die RRMSE der FH-Schätzwerte aufgelistet. Beim Vergleich der beiden Größen ist allerdings zu beachten, dass der RRMSE nicht nur die Streuung der FH-Schätzwerte sondern auch deren Verzerrung umfasst<sup>15)</sup>. Da die direkte Schätzung erwartungstreu ist, kann der relative Standardfehler auch als RRMSE interpretiert werden. Insofern sind die Werte der Spalten 3 und 5 durchaus vergleichbar, reflektieren aber eher die Effizienz der Schätzer insgesamt. Bedenkt man, dass die Verzerrung quadratisch in den MSE eingeht, kann die relative Streuung der FH-Schätzwerte nicht größer ausfallen als der

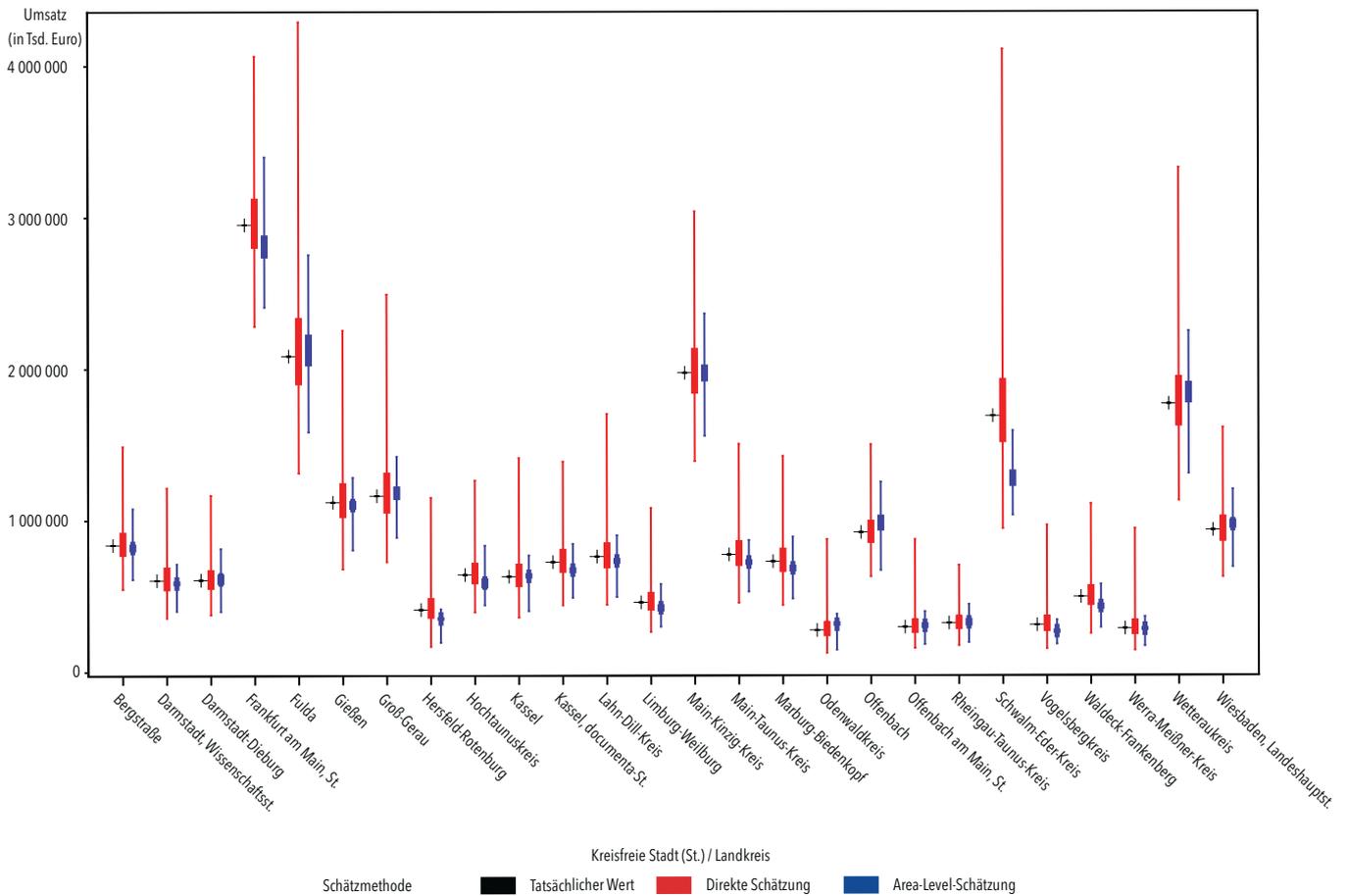
RRMSE. Da die RRMSE-Schätzwerte bereits kleiner sind als die entsprechenden Schätzwerte für den relativen Standardfehler der direkten Schätzung, lässt sich unmittelbar folgern, dass auch die Streuung des FH-Schätzers deutlich kleiner ist als diejenige des direkten Schätzers.

Einen umfassenderen Eindruck bezüglich des Abschneidens der beiden Schätzverfahren vermitteln die in Abbildung 1 dargestellten Ergebnisse der Simulationsrechnung. Simuliert wurden wiederum jeweils 10 000 Einzelschätzungen. Die Verteilungen der Schätzwerte sind in der Abbildung 1 in Form von Box-Plots dargestellt<sup>16)</sup>. Zusätzlich zu den Verteilungen sind in der Figur auch die tatsächlichen Werte in Form von Kreuzen gekennzeichnet. Bereits auf den ersten Blick

15) Siehe Box 1 „Stichprobenfehler, Schichtung, relativer Standardfehler und Mean Squared Error (MSE)“.

16) Bei jedem der 52 dargestellten Plots kennzeichnet die dickere Linie die mittleren 50 % der jeweiligen Schätzwerte. Die an das Rechteck oben und unten anschließenden dünnen Linien umfassen die höchsten und die niedrigsten 25 % der Schätzwerte.

Abbildung 1: Direkte Schätzung und Area-Level-Schätzung im Vergleich – Ergebnisse der Simulationsrechnung



lässt sich erkennen, dass die blau dargestellten FH-Schätzwerte deutlich weniger streuen als die direkten Schätzwerte. Aus den Simulationsergebnissen errechnet, liegen die relativen Standardfehler der FH-Schätzungen je nach betrachtetem Landkreis bzw. betrachteter kreisfreier Stadt zwischen 3,3 % und 9,5 %. Area-Level-Schätzungen führen somit in der Tat zu stabileren Ergebnissen.

Im Idealfall verteilen sich dabei die geringer streuenden FH-Schätzwerte mehr oder minder symmetrisch um den tatsächlichen Wert. Diese Situation ist in Abbildung 1 beispielsweise bei den Schätzungen für den Landkreis Bergstraße, für Darmstadt-Dieburg oder für den Main-Kinzig-Kreis zu beobachten. Es treten jedoch auch Fälle auf, bei denen die Stabilisierung ober- oder unterhalb des tatsächlichen Wertes erfolgt. In den allermeisten Fällen halten sich diese „Verschiebungen“ jedoch in Grenzen, sodass die FH-Schätzung den tatsächlichen Wert sehr wahrscheinlich nicht genau trifft, aber mit hoher Wahrscheinlichkeit näher am tatsächlichen Wert liegt als die direkte Schätzung. Andererseits zeigen die Ergebnisse für den Landkreis Hersfeld-Rotenburg und vor allem diejenigen für den Schwalm-Eder-Kreis, dass die FH-Schätzungen hier den tatsächlichen Wert allesamt unterschätzen. In Einzelfällen kann dies bei direkten Schätzungen ebenfalls auftreten, die Wahrscheinlichkeit dafür ist aber wesentlich geringer.

Über alle getätigten Schätzungen hinweg betrachtet, liegt der FH-Schätzwert in 69 % der Fälle näher am Vorgabewert als der direkte Schätzwert. Dabei fallen bei der FH-Schätzung die Abweichungen des Schätzwerts zum tatsächlichen Wert um durchschnittlich 59 % geringer aus als bei der direkten Schätzung. In 31 % der Fälle stellt sich das umgekehrte Resultat ein und die FH-Schätzung trägt durchschnittlich 51 % zur größeren Abweichung bei. Verrechnet man wiederum positive wie negative Veränderungen so führt die FH-Schätzung gegenüber der direkten Schätzung zu durchschnittlich 25 % geringeren Abweichungen.

Das insgesamt gute Abschneiden der Area-Level-Schätzung wird nicht zuletzt von der Güte der Hilfsvariablen, genauer gesagt, der hohen Korrelation zwischen Hilfs- und Zielvariablen getragen. In einem weiteren Simulationslauf wurde deshalb der Einfluss der Korrelation auf die Schätzergebnisse geprüft. Hierzu wurde die Hilfsvariable so konstruiert, dass der Korrelationskoeffizient von ursprünglich 0,96 auf nunmehr 0,68 absinkt<sup>17)</sup>. Ohne auf die Ergebnisse im Einzelnen einzugehen, führte die Area-Level-Schätzung unter diesen Voraussetzungen in immerhin 61 % der Fälle zu im Vergleich zur direkten Schätzung näher am tatsächlichen Wert liegenden Schätzwerten. Gegenüber der direkten Schätzung sind die Abweichungen um durchschnittlich 36 % kleiner. Bei 39 % der Schätzungen ergab hingegen die Area-Level-Schätzung die größeren Abweichungen, wobei die Vergrößerung im Schnitt 44 % der (nun größeren) Abweichung ausmachte. Neben der abgesunkenen Präzision der Area-Level-Schätzung steigt auch die Streuung der einzelnen Schätzwerte. Sie liegt aber auch unter diesen Voraussetzungen deutlich unterhalb der Streuung der direkten Schätzwerte.

## Zusammenfassung und Fazit

Der Beitrag befasste sich mit der Frage, wie gut eine Area-Level-Schätzung gegenüber dem klassischen Schätzverfahren – einer direkten Schätzung – abschneidet. Insbesondere wenn es darum geht, tief gegliederte Ergebnisse zu schätzen, bei denen das Gliederungsmerkmal nicht bei der Konzeption der Stichprobe berücksichtigt wurde. Als Anwendungsfall sollten die im Statistischen Unternehmensregister für Hessen ausgewiesenen, auf Ebene der Landkreise und kreisfreien Städte aggregierten Einzelhandelsumsätze auf der Basis einer Stichprobe geschätzt werden. Dieser Ansatz hat den Vorteil, dass die „wahren Werte“ vorab bekannt sind und mit den jeweiligen Schätzergebnissen verglichen werden können. Für die Area-Level-Schätzung wurde eine einzige, aus den Daten erzeugte Hilfsvariable verwendet. Stichprobenziehung und Schätzung wurden mehrfach wiederholt, um eine umfassendere Einschätzung der Ergebnisse zu erhalten.

Wenn die Korrelation zwischen der zu schätzenden Größe und der Hilfsvariablen sehr hoch ist, dann führt die Area-Level-Schätzung sowohl zu insgesamt präziseren als auch zu stabileren Schätzungen. Gegenüber der direkten Schätzung verbessern sich die Schätzergebnisse jedoch nicht gleichmäßig. Für jede Stichprobe sind einige Schätzwerte zu verzeichnen, bei denen die direkte Schätzung den Vorgabewert besser trifft als die Area-Level-Schätzung. In einzelnen Fällen unterschätzt der Area-Level-Schätzer bei mehrfacher Wiederholung der Stichprobenziehung konsequent den tatsächlichen Wert. Dieses Ergebnis ist zum Teil dem etwas mechanischen Vorgehen in der Simulation geschuldet. Im konkreten Anwendungsfall würde man auf Basis der (einen) vorliegenden Stichprobe mehr Sorgfalt auf die Passgenauigkeit des Modells legen und gegebenenfalls im Regressionsteil Ausreißer aus der Parameterschätzung ausschließen.

Insgesamt legen die erzielten Resultate die Schlussfolgerung nahe, dass Area-Level-Schätzungen unter den hier vorliegenden Voraussetzungen das geeignetere Schätzverfahren zur Erzielung von kleinräumigen Ergebnissen sind.

## Anhang: In der Simulation getestete Schätzgleichungen zur Bestimmung der Varianz des direkten Summenschätzers

Notation:

$\hat{V}$	Geschätzte Varianz	$\hat{U}_d$	Geschätzte Umsatzsumme von Landkreis bzw. kreisfreier Stadt $d$
$\pi_i$	Inklusionswahrscheinlichkeit erster Ordnung für Einheit $i$	$s_d$	Indexmenge der Stichprobeneinheiten von Landkreis bzw. kreisfreier Stadt $d$
$\pi_{ij}$	Inklusionswahrscheinlichkeit zweiter Ordnung für die Einheiten $i$ und $j$	$y_i$	Umsatz von Unternehmen $i$
$\pi_{l,i}$	Inklusionswahrscheinlichkeit erster Ordnung für Einheit $i$ aus Schicht $l$	$y_{l,i}$	Umsatz von Unternehmen $i$ in Schicht $l$
$\hat{V}_l$	Varianzschätzung für Schicht $l$	$n$	Stichprobenumfang
$f_l$	Auswahlsatz für Schicht $l$	$n_l$	Umfang der Teilstichprobe für Schicht $l$

Schätzgleichungen I und II:

Aus der Varianz des HT-Summenschätzers ergibt sich unmittelbar die erste getestete Schätzgleichung

$$\hat{V}^{HT}(\hat{U}_d) = \sum_{i \in s_d} (1 - \pi_i) \left( \frac{y_i}{\pi_i} \right)^2 + 2 \sum_{i \in s_d} \sum_{\substack{j \in s_d \\ j > i}} \left( 1 - \frac{\pi_i \pi_j}{\pi_{ij}} \right) \frac{y_i y_j}{\pi_i \pi_j}$$

17) Die geringere Korrelation wurde durch Ersetzung eines höheren Anteils von Einzelwerten mit den Werten einer gleichverteilten Zufallsvariablen erzielt.

Hierbei kennzeichnen die  $\pi_i$  die Inklusionswahrscheinlichkeiten erster Ordnung – also die Wahrscheinlichkeit, dass Einheit  $i$  in die Stichprobe gezogen wird. In einfachen Modellen entspricht diese in der Regel dem Auswahlatz für die Schicht zu der Einheit  $i$  gehört. Entsprechend kennzeichnet  $\pi_{ij}$  die Wahrscheinlichkeit, dass die Einheiten  $i$  und  $j$  gemeinsam in die Stichprobe gezogen werden. Diese Inklusionswahrscheinlichkeiten zweiter Ordnung wurden auf 2 unterschiedliche Arten bestimmt: In Variante I wurde  $\pi_{ij} = \frac{n_{li} \cdot n_{lj} - 1}{N_{li} \cdot N_{lj} - 1}$  gewählt, falls die Einheiten  $i$  und  $j$  aus derselben Schicht stammen, ansonsten  $\pi_{ij} = \pi_i \pi_j$ ; in Variante B wurde unabhängig von der Schichtzugehörigkeit der Einheiten vereinfacht durchweg  $\pi_{ij} = \pi_i \pi_j$  unterstellt. Somit entfällt in Variante II der zweite Term auf der rechten Seite der obigen Schätzgleichung. Variante B ist beispielsweise in der einschlägigen Schätzgleichung des R-packages „sae“<sup>18)</sup> umgesetzt.

### Schätzgleichung III

Als dritte Alternative wurde die in Lehtonen und Veijanen (2009)<sup>19)</sup> angegebene Schätzgleichung für Domain-Analysen getestet, die für den Fall vorgeschlagen wird, dass die betrachteten Domains – hier die Landkreise und kreisfreien Städte – nicht in der Stichprobenkonzeption berücksichtigt wurden. Die Schätzgleichung lautet

$$\hat{v}^{IV}(\hat{U}_d) = \frac{n}{n-1} \sum_{i \in S} \left( I_{d,i} \frac{y_i}{\pi_i} - \frac{\hat{\tau}_d}{n} \right)^2 \quad \text{mit} \quad \hat{\tau}_d = \sum_{i \in S} I_{d,i} \frac{y_i}{\pi_i}.$$

Demnach wird die Varianz des Schätzwerts  $\hat{U}_d$  unter Einbeziehung aller Stichprobenwerte bestimmt, wobei die Indikatorvariable  $I_{d,i}$  die Kreiszugehörigkeit kennzeichnet. Sie nimmt den Wert 1 an, wenn die Einheit  $i$  zu Landkreis bzw. kreisfreier Stadt  $d$  gehört und den Wert 0 wenn nicht.

18) R ist ein kostenfrei verfügbares und im wissenschaftlichen Umfeld weit verbreitetes statistisches Softwarepaket, welches für zahlreiche statistische Aufgaben spezielle Prozeduren bzw. Packages – so auch das Package „sae“ (Small-Area-Estimation) – bereitstellt.

19) Lehtonen, R. und Veijanen, A. (2009): Design-based Methods of Estimation for Domains and Small Areas. Sample Surveys: Inference and Analysis, Vol. 29B, S. 227.

Schätzgleichungen IV und V:

Die in der SAS Prozedur SURVEYMEANS implementierte Schätzgleichung setzt die Varianzschätzung aus den Teilvarianzen in den einzelnen Schichten zusammen:

$$\hat{v}^{SAS}(\hat{U}_d) = \sum_{l=1}^L \hat{v}_l(\hat{U}_d),$$

wobei der Index  $l = 1, \dots, L$  die Schichten kennzeichnet. Erfolgt die Varianzschätzung für Domains bzw. Landkreise und kreisfreie Städte, dann kennzeichnet auch hier eine Indikatorvariable  $I_{l,i}$  die Domain- bzw. Kreiszugehörigkeit. Die Variable nimmt den Wert 1 an, wenn die betrachtete Einheit  $i$  von Schicht  $l$  zu Domain  $d$  gehört, ansonsten den Wert 0. Die Teilvarianz errechnet sich über

$$\hat{v}_l(\hat{U}_d) = \frac{n_l(1-f_l)}{n_l-1} \sum_{i=1}^{n_l} \left( I_{l,i} \frac{y_{li}}{\pi_{li}} - \bar{y}_l \right)^2 \quad \text{mit} \quad \bar{y}_l = \frac{1}{n_l} \sum_{i=1}^{n_l} I_{l,i} \frac{y_{li}}{\pi_{li}}.$$

Wird in SURVEYMEANS eine Varianzschätzung gemäß Taylor-Approximation für eine Domain-Analyse gewählt, greift die Prozedur auf diese Schätzformel zurück. Bei Angabe der Besetzungszahlen der Schichten in der Grundgesamtheit (Verwendung der Option TOTAL = in Verbindung mit dem Statement STRATA in der Prozedur) werden die Schätzgleichungen direkt umgesetzt. Ohne die Angabe geht die Prozedur davon aus, dass es nur eine Schicht gibt. Man bemerkt die ähnliche Struktur zwischen der obigen Schätzgleichung und Schätzgleichung III. In der Tat führen beide zum selben Ergebnis, wenn in Gleichung IV/V sowohl auf den Korrekturfaktor  $(1-f_l)$  als auch auf eine schichtweise Berechnung der Teilvarianzen verzichtet wird.

**Peter Gottfried;** Tel: 0611 3802-868;  
E-Mail: peter.gottfried@statistik.hessen.de

# Die Regionalisierung der Hessischen Jahresherhebung im Einzelhandel anhand von Small-Area-Verfahren

*Der Einzelhandel stellt als Bindeglied zwischen Produzenten und Konsumenten einen wichtigen Indikator zur Beurteilung der aktuellen Konjunktur dar. Insbesondere auf regionaler Ebene sind entsprechende Kennzahlen aussagekräftig zur Beurteilung der einzelhandelsbezogenen Attraktivität des Standortes. Eine Bewertung der Struktur, der Rentabilität und der Produktivität hessischer Einzelhandelsunternehmen wird durch die bundesweit durchgeführte Jahresherhebung im Einzelhandel ermöglicht. Im Folgenden wird untersucht, inwieweit die Anwendung sogenannter Small-Area-Verfahren bei der Auswertung der Jahresherhebung im Einzelhandel zu zuverlässigen Schätzungen auf Ebene der Landkreise und kreisfreien Städte führen. Von **Julia Manecke***

## 1. Die Hessische Jahresherhebung im Einzelhandel

Die Jahresherhebung im Einzelhandel erfasst als Teil der Handelsstatistik aussagekräftige Kennzahlen rechtlich selbstständiger Unternehmen, die ihren Sitz in Deutschland haben und gemäß der aktuellen Klassifikation der Wirtschaftszweige (WZ) der Abteilung 47 „Einzelhandel (ohne Handel mit Kraftfahrzeugen)“ angehören (siehe Statistisches Bundesamt, 2008). Da die Erhebungseinheit das rechtlich selbstständige Unternehmen ist, umfassen die erhobenen Kennzahlen einer rechtlichen Einheit alle dazugehörigen rechtlich unselbstständigen inländischen Verkaufsfilialen sowie Hilfs- und Nebenbetriebe (siehe Statistisches Bundesamt, 2016a). Im Jahr 2014 belief sich die Anzahl der rechtlich eigenständigen Einzelhandelsunternehmen in Hessen auf insgesamt 23 291 Einheiten (siehe Hessisches Statistisches Landesamt, 2016).

Nach Abzug der echten und unechten Antwortausfälle haben für das Jahr 2014 insgesamt 2274 Einheiten an der hessischen Jahresherhebung im Einzelhandel teilgenommen. Die Erhebung erfolgt anhand einer dreifach geschichteten Zufallsstichprobe mit den Schichtungsmerkmalen „Bundesland“, „Umsatzgrößenklasse“ und „Branchengruppe“. Die Branchengruppen der WZ-Ab-



© Eisenhans - Fotolia.com

teilung 47 (Einzelhandel (ohne Handel mit Kraftfahrzeugen)) sind die WZ-Gruppen 47.1 bis 47.9, im Folgenden auch WZ-Dreisteller genannt.

Der im Statistischen Bericht der Hessischen Jahresherhebung 2014 veröffentlichte geschätzte Gesamtumsatz aller hessischen Einzelhandelsunternehmen beträgt 23,68 Mrd. Euro (siehe Hessisches Statistisches Landesamt, 2016). Zudem wurde die Zahl der beschäftigten Personen auf 177 213 geschätzt. Die Höhe der Bruttoentgelte beträgt schätzungsweise 2,40 Mrd. Euro. Durch das Schichtungsverfahren ist zudem gewährleistet, dass je Bundesland zuverlässige

**Tabelle 1: Kennzahlen hessischer Einzelhandelsunternehmen (ohne Handel mit Kraftfahrzeugen) 2014 nach Wirtschaftszweigen**

WZ-Gruppe/ WZ-Abteilung	Wirtschaftszweig	Unternehmen	Beschäftigte	Umsatz	Bruttoentgelte
				Mill. Euro	
47	Einzelhandel (ohne Handel mit Kraftfahrzeugen)	23 291	177 213	23 684	2 396
47.1	Einzelhandel mit Waren verschiedener Art (in Verkaufsräumen)	3 076	52 753	7 507	639
47.2	Einzelhandel mit Nahrungs- und Genussmitteln, Getränken und Tabakwaren (in Verkaufsräumen)	2 883	15 832	1 210	145
47.3	Einzelhandel mit Motorenkraftstoffen (Tankstellen)	350	2 999	517	26
47.4	Einzelhandel mit Geräten der Informations- und Kommunikationstechnik (in Verkaufsräumen)	920	6 655	1 452	165
47.5	Einzelhandel mit sonstigen Haushaltsgeräten, Textilien, Heimwerker- und Einrichtungsbedarf (in Verkaufsräumen)	3 095	18 390	2 065	297
47.6	Einzelhandel mit Verlagsprodukten, Sportausrüstungen und Spielwaren (in Verkaufsräumen)	1 765	9 595	842	101
47.7	Einzelhandel mit sonstigen Gütern (in Verkaufsräumen)	8 696	58 753	7 172	826
47.8	Einzelhandel an Verkaufsständen und auf Märkten	343	1 153	65	5
47.9	Einzelhandel, nicht in Verkaufsräumen, an Verkaufsständen oder auf Märkten	2 162	11 083	2 853	194

Schätzungen für die einzelnen Einzelhandelsgruppen separat durchgeführt werden können. Dementsprechend veröffentlicht das Hessische Statistische Landesamt (HSL) jährlich Schätzwerte für einschlägige Kennzahlen der hessischen Einzelhandelsunternehmen, gegliedert nach den WZ-Dreistellern der WZ-Abteilung 47. Tabelle 1 gibt einen Überblick über die anhand der Jahreserhebung im Einzelhandel 2014 berechneten Schätzwerte der Variablen „Umsatz“, „Beschäftigte“ sowie „Bruttoentgelte“ für alle hessischen Einzelhandelsunternehmen auf Ebene der jeweiligen WZ-Dreisteller.

Die Ermittlung der obigen Schätzwerte erfolgt durch eine Hochrechnung der erhobenen Unternehmenskennzahlen auf die Grundgesamtheit. Hierfür wird jedem befragten Unternehmen ein Hochrechnungsfaktor zugewiesen, der der Kehrwert seiner jeweiligen Auswahlwahrscheinlichkeit ist. Die hochgerechneten Werte aller Unternehmen, die zu der Teilgesamtheit gehören für die man Schätzwerte ermitteln möchte, werden aufsummiert. Da bei diesem auf Horvitz und Thompson (1952) zurückgehenden Schätzer lediglich Informationen aus der interessierenden Teilgesamtheit (hier: der jeweiligen Branchengruppe) verwendet werden, wird das Schätzverfahren auch direkte Schätzung genannt.

## 2. Die Disaggregation der Ergebnisse der Jahreserhebung im Einzelhandel am Beispiel des Merkmals „Beschäftigte“

### 2.1 Direkter Schätzer

Im Hinblick auf eine Beurteilung der regionalen Konjunktur sind zusätzlich zu den obigen Informationen, die das gesamte Bundesland Hessen umfassen, auch kleinräumige Schätzwerte von Interesse. Im Folgenden wird dies anhand des Merkmals „Beschäftigte“ verdeutlicht. Zunächst soll die geschätzte Zahl der 177 213 Beschäftigten aller hessischen Einzelhandelsunternehmen – d. h. unabhängig von ihrer jeweiligen Einzelhandels-Branche – auf die Ebene der Landkreise bzw. kreisfreien Städte disaggregiert werden. Die Grundgesamtheit der berücksichtigten Unternehmen  $U$  mit  $N = 23\,291$  Elementen mit dem Laufindex  $j = 1, \dots, N$  wird also entsprechend der 26 hessischen Landkreise bzw. kreisfreien Städte in  $D = 26$  überschneidungsfreie Teilgesamtheiten eingeteilt. Für jede Teilgesamtheit mit dem Laufindex  $d = 1, \dots, D$  soll der Totalwert  $\tau_d$ , d. h. die insgesamt Beschäftigten, geschätzt werden. Der direkte Horvitz-Thompson-Schätzer ist eine unverzerrte Schätzfunktion für  $\tau_d$ . Er ist gegeben durch

$$\hat{\tau}_d^{HT} = \sum_{j \in S_d} \pi_j^{-1} y_j$$

Dabei bezeichnet  $S_d$  den Teil der Stichprobe mit den befragten Unternehmen, die zu Teilge-

samtheit  $d$  gehören. Zudem ist  $y_j$  die Anzahl der Beschäftigten in Unternehmen  $j$  und  $\pi_j$  die jeweilige Auswahlwahrscheinlichkeit des Unternehmens.

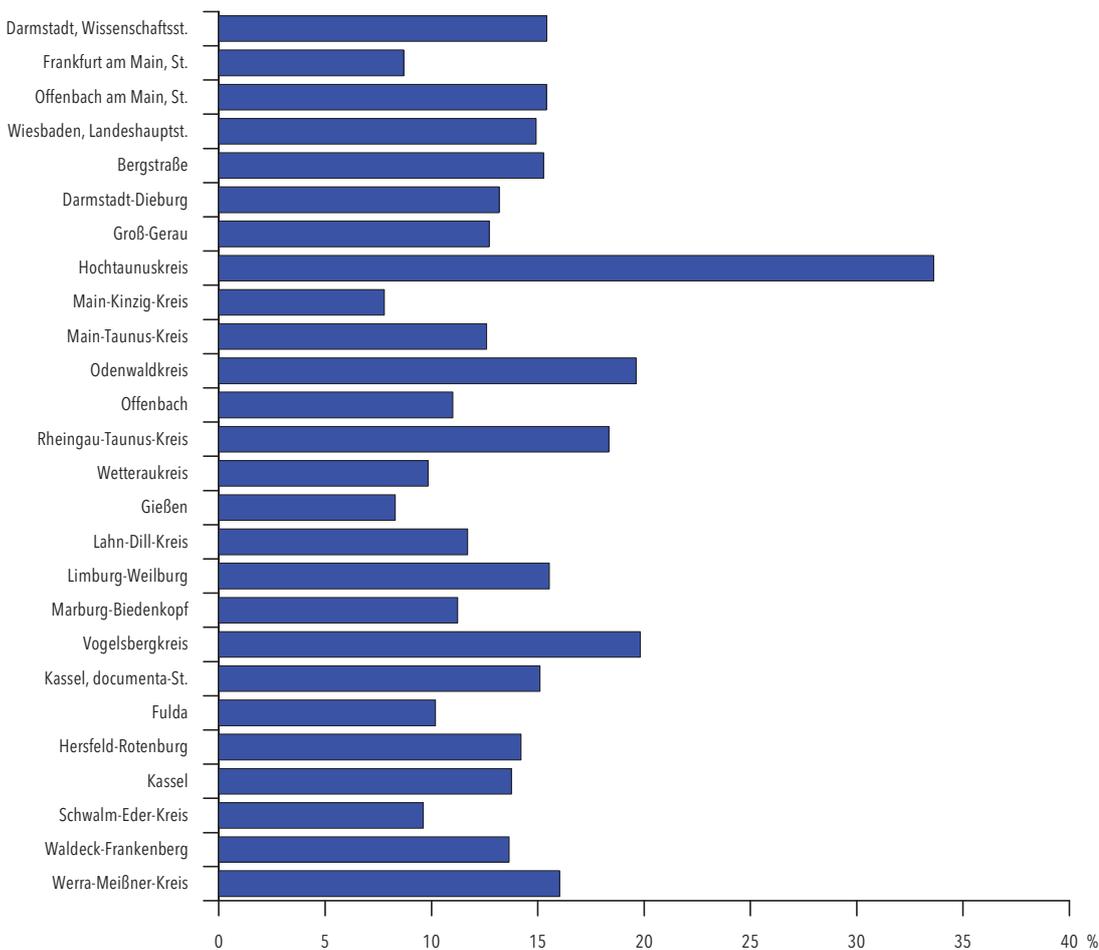
Die Jahrerhebung ist jedoch nicht für eine direkte Schätzung auf Ebene der Landkreise und kreisfreien Städte ausgelegt. Aufgrund der Tatsache, dass die Ebene der Landkreise und kreisfreien Städte nicht bei dem Design der Stichprobe berücksichtigt worden ist, ist der jeweilige kreisspezifische Stichprobenumfang zufällig, was zu einer erhöhten Variabilität der Schätzungen führt. Eine verbreitete Kennzahl zur Beurteilung der Genauigkeit eines Schätzwertes ist der relative Standardfehler (RS) bzw. die relative Wurzel der mittleren quadratischen Abweichung (**Relative Root Mean Squared Error (RRMSE)**). Für die Herleitung bzw. Erläuterung der entsprechenden Kenngrößen sei an dieser Stelle auf den Abschnitt „Stichprobenfehler,

Schichtung, relativer Standardfehler und Mean Squared Error (MSE)“ des vorherigen Artikels (siehe Gottfried, 2017) verwiesen.

In der amtlichen Statistik wird der Aussagewert von Schätzwerten mit einem relativen Standardfehler von 10 % bis unter 20 % als eingeschränkt eingeschätzt. Spätestens aber ab einem relativen Standardfehler von 20 % werden Schätzungen aufgrund einer zu geringen Schätzgüte in der Regel nicht mehr veröffentlicht (siehe Statistisches Bundesamt, 2016b: 8).

Abbildung 1 stellt die Höhe des geschätzten relativen Standardfehlers einer direkten Schätzung der Anzahl der Beschäftigten auf Ebene der hessischen Landkreise und kreisfreien Städte dar. Es ist zu erkennen, dass der Großteil der Schätzwerte lediglich mit eingeschränktem Aussagewert veröffentlicht werden könnte. Lediglich die Schätzwerte von 5 Landkreisen bzw. kreisfreien Städten können die Anforderung an

**Abbildung 1: Geschätzter relativer Standardfehler einer direkten Schätzung der Beschäftigten insgesamt im Einzelhandel<sup>1)</sup> 2014 nach Landkreisen und kreisfreien Städten (St.)**



1) Wirtschaftszweig 47 – Einzelhandel (ohne Handel mit Kraftfahrzeugen).

einen relativen Standardfehler von unter 10 % erfüllen. Das Ergebnis des Hochtaunuskreises liegt mit einem Wert von 33,6 % sogar deutlich über der Obergrenze einer noch akzeptierbaren Schätzgüte.

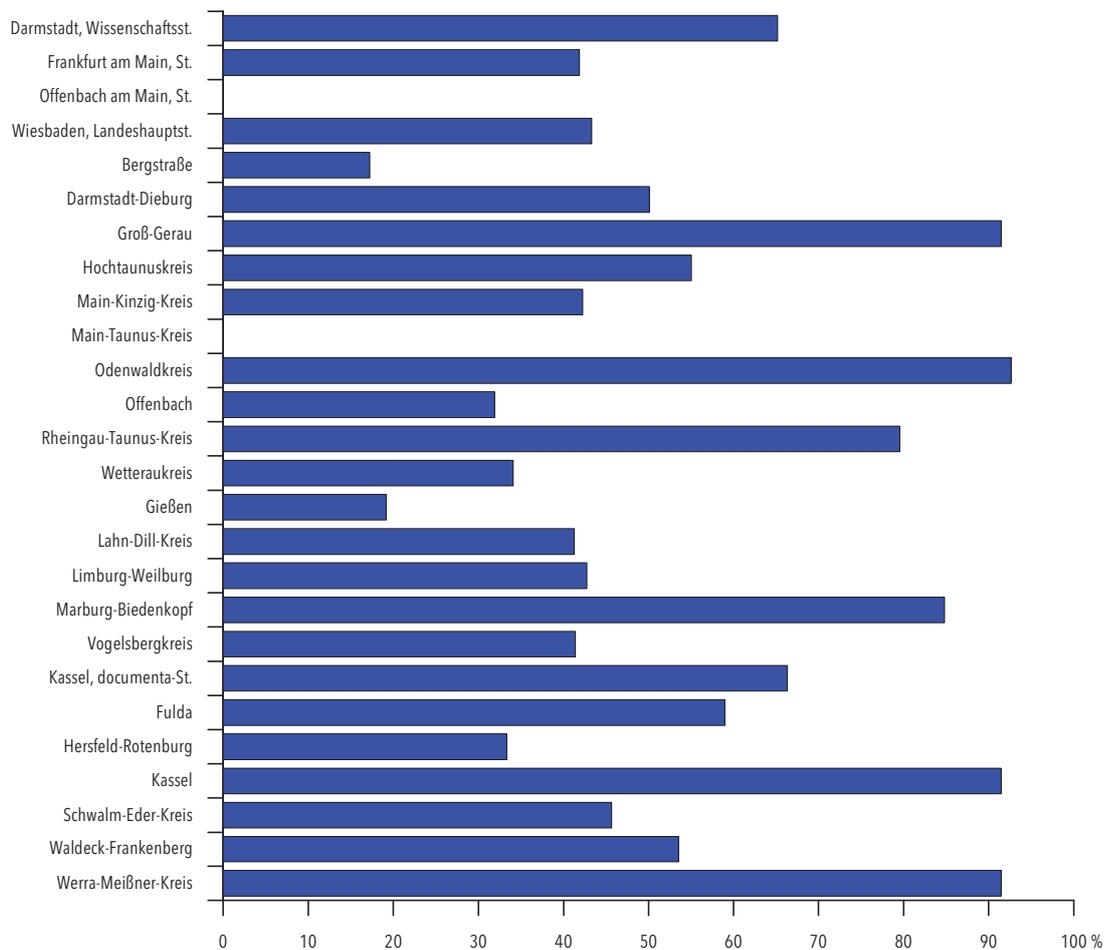
Darüber hinaus soll eine Disaggregation der 9 branchenspezifischen Schätzwerte auf die einzelnen Landkreise und kreisfreien Städte erfolgen. Die Grundgesamtheit  $U$  mit  $N = 23\,291$  Unternehmen wird nun also in  $D = 234$  überschneidungsfreie Teilgesamtheiten<sup>1)</sup> eingeteilt. Für jede Teilgesamtheit wird nun wieder der Totalwert  $\tau_d$  mit dem obig dargestellten Horvitz-Thompson-Schätzer geschätzt. Eine Disaggregation der Schätzwerte auf derart kleine Teilgesamtheiten hat allerdings aufgrund zu geringer Teilstichprobenumfänge untragbar hohe Schätzfehler zur Folge. Abbildung 2 zeigt den geschätzten relativen Standardfehler einer direkten Schätzung der Anzahl der Beschäftig-

ten in WZ 47.3 (Einzelhandel mit Motorenkraftstoffen (Tankstellen))<sup>2)</sup> auf Ebene der hessischen Landkreise und kreisfreien Städte.

Es ist deutlich zu erkennen, dass nur 2 Teilgesamtheiten einen akzeptierbaren relativen Standardfehler von weniger als 20 % aufweisen. Darüber hinaus ist in der Stadt Offenbach und im Main-Taunus-Kreis kein Unternehmen befragt worden, welches dem WZ 47.3 (Einzelhandel mit Motorenkraftstoffen (Tankstellen)) zuzuordnen ist. Für diese Teilgesamtheiten kann daher kein Schätzwert gebildet und dementsprechend auch kein relativer Standardfehler geschätzt werden. Für die anderen Branchengruppen des Einzelhandels ergibt sich ein ähnliches Bild.

Abbildung 3 stellt die Häufigkeiten des geschätzten relativen Standardfehlers der direkten Schätzer über alle 234 Teilgesamtheiten dar. Insgesamt haben die Beschäftigten-Schätzwerte

**Abbildung 2: Geschätzter relativer Standardfehler einer direkten Schätzung der Beschäftigten insgesamt im Einzelhandel mit Motorenkraftstoffen (Tankstellen)<sup>1)</sup> 2014 nach Landkreisen und kreisfreien Städten (St.)**



1) Die 234 Teilgesamtheiten entstehen durch die Kombination der 9 Branchengruppen (WZ-Dreisteller 47.1 bis 47.9) mit den 26 hessischen Landkreisen bzw. kreisfreien Städten.

2) Zur Verdeutlichung der Problematik kleiner Teilgesamtheiten sowie von Teilgesamtheiten, in denen kein Unternehmen Teil der Stichprobe ist, wurde der WZ 47.3 als Beispiel ausgewählt.

1) Wirtschaftszweig 47.3.

von 190 der 234 interessierenden Teilgesamtheiten einen geschätzten relativen Standardfehler von mindestens 20 %. Kein Schätzwert erfüllt die Anforderungen eines relativen Standardfehlers von unter 10 %. Die Schätzungen von 31 Teilgesamtheiten können immerhin mit einem eingeschränkten Aussagewert (relativer Standardfehler zwischen 10 % und 20 %) veröffentlicht werden. In den übrigen 13 Teilgesamtheiten ist kein Unternehmen befragt worden. Spätestens in diesen Fällen ist es daher nicht mehr möglich anhand von direkten Verfahren, die lediglich Informationen aus der jeweiligen Teilgesamtheit selbst berücksichtigen, Schätzwerte zu ermitteln.

## 2.2 Der Schätzer nach Battese-Harter-Fuller

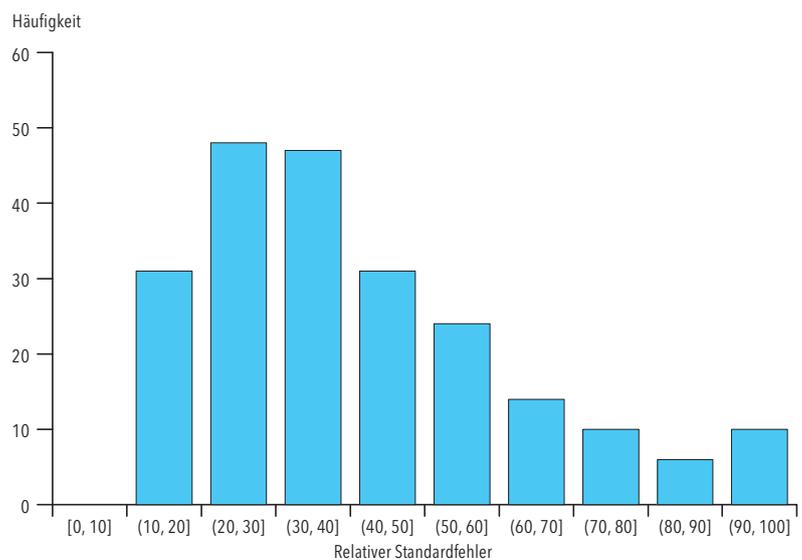
### 2.2.1 Allgemeine Methodik

Teilgesamtheiten, deren jeweiliger Stichprobenumfang nicht für eine hinreichend präzise Schätzung anhand eines direkten Schätzers ausreicht, werden als *Small Areas* bezeichnet (siehe Rao & Molina, 2015: 2). Hier bieten sogenannte Small-Area-Verfahren eine potenzielle Lösung. Einen umfassenden Überblick über verschiedenste Methoden und Anwendungsbeispiele im Kontext der Zensusforschung gibt ein von Münich et. al. (2013) veröffentlichter Artikel.

Small-Area-Verfahren zielen durch die Verwendung von zusätzlichen Hilfsinformationen generell darauf ab, die mittlere quadratische Abweichung der Schätzung deutlich zu minimieren. Darüber hinaus ermöglichen entsprechende modellbasierte Small-Area-Verfahren die Bildung von Schätzwerten für Teilgesamtheiten, in denen kein Unternehmen befragt wurde. Da die Zusatzinformationen, welche im Rahmen eines entsprechenden Modells berücksichtigt werden, auch aus anderen Quellen als der interessierenden Teilgesamtheit selbst stammen, spricht man hierbei von einer indirekten Schätzung.

Im Gegensatz zu den im vorherigen Artikel (siehe Gottfried, 2017) dargestellten Area-Level-Modellen verwenden sogenannte Unit-Level-Modelle Informationen auf Ebene der Erhebungseinheiten. Aufgrund der weit aussagekräftigeren Infor-

**Abbildung 3: Verteilung des geschätzten relativen Standardfehlers bei einer direkten Schätzung der Beschäftigten insgesamt auf der Ebene der Landkreise und kreisfreien Städte (St.) und Einzelhandelsbranchengruppen<sup>1)</sup>**



1) Wirtschaftszweiggruppen des Einzelhandels (ohne Handel mit Kraftfahrzeugen) 47.1 bis 47.9.

mationen, die in Mikrodaten enthalten sind, wird durch die Verwendung von Unit-Level-Modellen eine noch effizientere Schätzung ermöglicht. Das Standardverfahren ist der sogenannte Battese-Harter-Fuller-Schätzer (siehe Battese et. al., 1988). Im Hinblick auf die zu schätzende Anzahl der Beschäftigten je Teilgesamtheit sei  $y_{aj}$  die Anzahl der Beschäftigten in Unternehmen  $j$  in der Teilgesamtheit  $d$  und  $x_{aj}$  der Vektor mit den Hilfsvariablen dieser Einheit. Zudem ist  $N_d$  die Anzahl der Unternehmen, die zur Teilgesamtheit  $d$  gehören. Das dem Battese-Harter-Fuller-Schätzer zugrunde liegende und für die Grundgesamtheit angenommene Modell ist eine Sonderform des allgemeinen gemischten linearen Regressionsmodells und gegeben durch:

$$y_{aj} = x_{aj}^T \beta + v_d + e_{aj}, \quad d = 1, \dots, D, \quad j = 1, \dots, N_d$$

$$v_d \stackrel{iid}{\sim} (0, \sigma_v^2)$$

$$e_{aj} \stackrel{iid}{\sim} (0, \sigma_e^2).$$

Hierbei ist  $x_{aj}^T \beta$  der Regressionsterm mit dem Vektor der Regressionskoeffizienten  $\beta$ , welcher über alle Teilgesamtheiten und Einheiten die fixen Effekte, d. h. den Zusammenhang zwischen der interessierenden Variable  $y_{aj}$  und den Hilfsvariablen  $x_{aj}$  erfasst. Der sogenannte Zufallseffekt  $v_d$  berücksichtigt Schwankungen zwischen den Teilgesamtheiten, die nicht durch den fixen Effekt des Regressionsterms erklärt werden können. Zudem beschreibt die Störgröße  $e_{aj}$  die in-

dividuellen Stichprobenfehler. Dabei haben  $v_d$  und  $e_{dj}$  jeweils den Erwartungswert 0 und die Varianzen  $\sigma_v^2$  beziehungsweise  $\sigma_e^2$ , wobei angenommen wird, dass beide Größen unabhängig voneinander sind. Die Varianz der Zufallseffekte  $\sigma_v^2$  wird auch als Modellvarianz bezeichnet, da sie die Varianz zwischen den Teilgesamtheiten angibt, die nicht durch die fixe Komponente des Modells erklärt werden kann.

Unter der Annahme, dass das obige gemischte Regressionsmodell ebenfalls für die Stichprobe gilt, wird der Mittelwert der interessierenden Variable je Teilgesamtheit  $\mu_d$  geschätzt durch den sogenannten **Empirical Best Linear Unbiased Predictor (EBLUP)** nach Battese, Harter und Fuller:

$$\begin{aligned} \hat{\mu}_d^{BHF} &= \bar{X}_d^T \hat{\beta} + \hat{v}_d \quad \text{mit} \\ \hat{v}_d &= \hat{\gamma}_d (\bar{y}_d - \bar{x}_d^T \hat{\beta}) \quad \text{und} \\ \hat{\gamma}_d &= \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + \hat{\sigma}_e^2 / n_d} . \end{aligned}$$

Hierbei bezeichnet  $\bar{x}_d$  den bekannten wahren Populationsdurchschnitt der verwendeten Hilfsvariablen in Teilgesamtheit  $d$ . Zudem sind  $\bar{x}_d$  und  $\bar{y}_d$  die Stichprobendurchschnitte der Hilfsvariablen beziehungsweise der interessierenden Variablen je Teilgesamtheit  $d$ . Darüber hinaus ist  $n_d$  die entsprechende Größe der Stichprobe in  $d$ . Da die Modellvarianz  $\sigma_v^2$  und die Varianz des Stichprobenfehlers  $\sigma_e^2$  in der Realität unbekannt sind, werden diese zur Ermittlung des EBLUPs geschätzt<sup>3)</sup>. Der Battese-Harter-Fuller Schätzer für Teilgesamtheit  $d$  ist dementsprechend eine Kombination aus einem synthetischen Schätzer  $\bar{x}_d^T \hat{\beta}$ , der sich aus der fixen Modellkomponente bildet, sowie dem jeweiligen geschätzten Zufallseffekt  $\hat{v}_d$ . Durch die Erfassung der Zusammenhänge der Hilfsvariablen und der interessierenden Variable über die innerhalb der Stichprobe befragten Einheiten werden also die entsprechenden Parameter anhand des gemischten linearen Regressionsmodells geschätzt und ermöglichen so in Kombination mit den Populationsdurchschnitten der Hilfsvariable eine Berechnung von Schätzwerten für die Population.

Der EBLUP lässt sich umformen zu der zusammengesetzten Schätzfunktion

$$\hat{\mu}_d^{BHF} = \hat{\gamma}_d (\bar{y}_d + (\bar{X}_d - \bar{x}_d)^T \hat{\beta}) + (1 - \hat{\gamma}_d) \bar{x}_d^T \hat{\beta} .$$

Hier ist zu erkennen, dass der Battese-Harter-Fuller-Schätzer ein gewichtetes Mittel aus

dem direkten Stichprobenregressionsschätzer  $\bar{y}_d + (\bar{X}_d - \bar{x}_d)^T \hat{\beta}$  und dem synthetischen Schätzer  $\bar{x}_d^T \hat{\beta}$  ist. Der Gewichtungsfaktor  $\hat{\gamma}_d$  gibt je Teilgesamtheit den Anteil der Modellvarianz an der Gesamtvarianz an und legt fest, wie viel Gewicht auf die jeweiligen Komponenten gelegt wird. Bei einer sehr hohen Modellvarianz  $\hat{\sigma}_v^2$  oder einem großen Stichprobenumfang  $n_d$  wird entsprechend viel Vertrauen in den direkten Stichprobenregressionsschätzer gelegt. Im Gegenzug dazu nähert sich der EBLUP bei einer geringen Modellvarianz bzw. einem geringen Teilstichprobenumfang tendenziell der synthetischen Komponente an, welche die Variablenausprägungen der Stichprobeneinheiten aus Teilgesamtheit  $d$  nicht berücksichtigt. Entsprechend besteht der EBLUP bei Teilgesamtheiten, in denen keine Einheit befragt wurde ( $n_d = 0$ ) vollständig aus dem synthetischen Schätzer, sodass hier  $\hat{\mu}_d^{BHF} = \bar{x}_d^T \hat{\beta}$  gilt.

## 2.2.2 Anwendung auf die Jahreserhebung im Einzelhandel

Mit dem Statistischen Unternehmensregister des Landes Hessen liegt dem HSL eine umfangreiche Datenbank mit einer Vielzahl von Kennzahlen auf Ebene der Erhebungseinheiten vor. Entsprechend enthält die Informationsquelle sowohl für die im Rahmen der Einzelhandelserhebung befragten als auch für die nicht-befragten Unternehmen u. a. Angaben zu den Merkmalen „Umsatz“, „sozialversicherungspflichtig Beschäftigte“ in Voll- bzw. Teilzeit, „geringfügig Beschäftigte“ sowie zu der Anzahl und Größe der zugehörigen Niederlassungen. Diese Mikrodaten sind potenzielle Hilfsvariablen für das im vorherigen Abschnitt dargestellte Unit-Level-Modell. Auf dieser Grundlage wurde der Battese-Harter-Fuller-Schätzer herangezogen, um für die 234 Teilgesamtheiten, in die sich die hessischen Einzelhandelsunternehmen entsprechend ihrer Branchengruppe und ihrer Kreiszugehörigkeit unterteilen lassen, jeweils die Totalwerte  $\tau_d$  der Beschäftigten, des Umsatzes und der Bruttoentgelte zu schätzen. Im Folgenden wird die Vorgehensweise anhand der Totalwertschätzung des Merkmals „Beschäftigte“ dargestellt.

Die im Rahmen der Einzelhandelserhebung abgefragte Zahl der Beschäftigten  $y_{ij}$  umfasst alle im Unternehmen tätigen Personen. Dazu gehö-

3) Für die Schätzung der Varianzkomponenten gibt es verschiedene Verfahren. Für die im Folgenden dargestellte Unit-Level-Schätzung wurde die Restricted Maximum Likelihood-Methode verwendet.

ren neben sozialversicherungspflichtig und geringfügig Beschäftigten u. a. auch mitarbeitende Inhaber, unbezahlt mithelfende Familienangehörige, Gesellschafter sowie Vorstandsmitglieder (siehe Statistisches Bundesamt, 2016a). Das Statistische Unternehmensregister hingegen enthält lediglich Angaben zu der Zahl der sozialversicherungspflichtig Beschäftigten (gegliedert in Voll- und Teilzeitbeschäftigte) und der Zahl der geringfügig Beschäftigten.

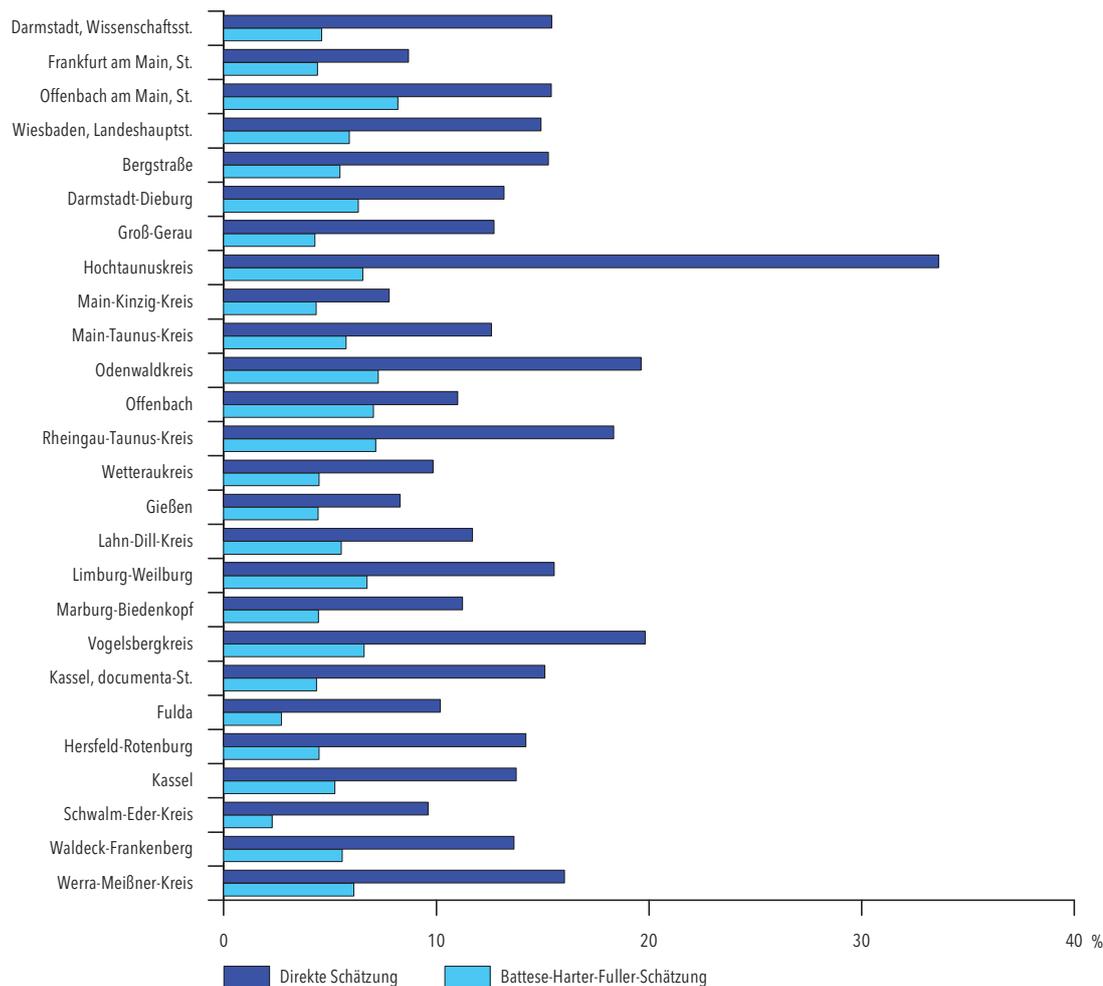
Die Modellbildung und Schätzung wird – wie bereits bei der direkten Schätzung – zunächst über alle Einzelhandelsunternehmen und dann schließlich zur weiteren Disaggregation separat für alle neun WZ-Dreisteller durchgeführt. Dementsprechend werden insgesamt 10 verschiedene Modelle zur Prädiktion des Totalwertes der Beschäftigten verschiedener Einzelhandelsbranchen in den  $D = 26$  hessischen Landkreisen bzw.

kreisfreien Städten geschätzt.

Die Auswahl der Registervariablen, die als Hilfsinformation in das Modell einbezogen werden, erfolgt anhand des „conditional AIC“, einem Modellselektionskriterium, das die prädiktive Güte des jeweiligen Modells misst (siehe Vaida & Blanchard, 2005). Des Weiteren wurde bei der Selektion überprüft, inwieweit die modellbasierte Schätzung anhand einer Kombination bestimmter Hilfsvariablen zu möglichst unverzerrten Schätzwerten in Verbindung mit einer möglichst geringen Schätzvarianz führt. Dabei können sich die letztendlich ausgewählten Variablenkombinationen für die einzelnen Modelle unterscheiden.

Die final ausgewählten Hilfsvariablen zur Schätzung der insgesamt Beschäftigten im gesamten WZ 47 sowie in den WZ-Dreistel-

**Abbildung 4: Geschätzter RRMSE<sup>1)</sup> einer direkten Schätzung und der Schätzung mit dem Battese-Harter-Fuller-Verfahren der Beschäftigten insgesamt im Einzelhandel<sup>2)</sup> 2014 nach Landkreisen und kreisfreien Städten (St.)**



1) Relative Root Mean Squared Error. – 2) Wirtschaftszweig 47 – Einzelhandel (ohne Handel mit Kraftfahrzeugen).

lern 47.1 bis 47.9 sind die im Unternehmensregister 2014 angegebene Anzahl der sozialversicherungspflichtig Beschäftigten (SVB) sowie die Anzahl der geringfügig Beschäftigten (GEB). Dementsprechend ist das gemischte lineare Regressionsmodell gegeben durch

$$y_{dj} = \beta_0 + \beta_1 * SVB_{dj} + \beta_2 * GEB_{dj} + v_d + e_{dj}$$

$$d = 1, \dots, 26, j = 1, \dots, N_d.$$

Im Anschluss an die Variablenselektion wird eine Ausreißerdiagnostik durchgeführt. Diese dient zur Identifikation einzelner Unternehmen, die das Modell aufgrund ungewöhnlicher Variablenausprägungen massiv beeinflussen oder deren Daten potenziell fehlerbehaftet sind. Entsprechende Einheiten werden zur Stabilisierung der Schätzung aus dem Modell genommen<sup>4)</sup>. In einer Monte-Carlo-Simulationsstudie, die im Rahmen der hier zugrundeliegenden Masterarbeit durchgeführt wurde, konnte beobachtet werden, dass der Ausschluss von entsprechenden Ausreißern aus dem Schätzmodell zu einer enormen Verbesserung der Schätzgüte des Battese-Harter-Fuller-Schätzers führt (siehe Manecke, 2017). Der endgültige Prädiktor der Gesamtzahl der Beschäftigten je Teilgesamtheit ist schließlich das Produkt aus dem ausreißerbereinigtem Mittelwertschätzer nach Battese-Harter-Fuller und der bekannten Anzahl der Unternehmen aus Teilgesamtheit  $d$  in der Grundgesamtheit:  $\hat{t}_d^{BHF} = \hat{\mu}_d^{BHF} * N_d$ .

Zur Beurteilung der Effizienz der Small-Area-Schätzung wird der MSE des Unit-Level-EBLUPs anhand eines sogenannten parametrischen Bootstrap-Verfahrens (siehe González-Manteiga et. al., 2008) geschätzt und anschließend mit dem in Abschnitt 2.1 ermittelten MSE der direkten Schätzung verglichen. In Anlehnung an Abbildung 1 zeigt Abbildung 4 neben dem geschätzten RRMSE der direkten Schätzung<sup>5)</sup> der in WZ 47 Beschäftigten je Landkreis bzw. kreisfreier Stadt nun zusätzlich den geschätzten RRMSE der Schätzung mit dem Battese-Harter-Fuller-Verfahren. Es ist zu sehen, dass die Verwendung zusätzlicher Hilfsinformationen aus dem Unternehmensregister im Rahmen des vorgestellten Unit-Level-Modells zu einer deutlichen Reduktion des geschätzten RRMSE führt.

Noch deutlicher wird das Potenzial von Small-Area-Verfahren bei einer weiteren Disaggregation der Grundgesamtheit entsprechend der Einzelhandelsbranchen. In Ergänzung zu Abbildung 2 zeigt Abbildung 5 nun die Schätzgüte der Schätzung der in WZ 47.3 (Einzelhandel mit Motorenkraftstoffen (Tankstellen)) insgesamt beschäftigten Personen je Landkreis bzw. kreisfreier Stadt anhand der beiden vorgestellten Verfahren. Der Präzisionsgewinn durch die Verwendung von Small-Area-Verfahren ist aufgrund der sehr unzuverlässigen direkten Schätzwerte auf dieser Disaggregationsebene sehr offensichtlich. Zudem ermöglicht die Verwendung von Hilfsinformationen nun die Berechnung von Schätzwerten für die Stadt Offenbach und den Main-Taunus-Kreis, obwohl in diesen Verwaltungsbezirken kein Unternehmen des WZ 47.3 Teil der Stichprobe der Einzelhandelserhebung war. Allerdings liegt der geschätzte RRMSE dieser zwei voll-synthetischen Schätzungen jeweils über der akzeptierbaren Obergrenze von 20 %.

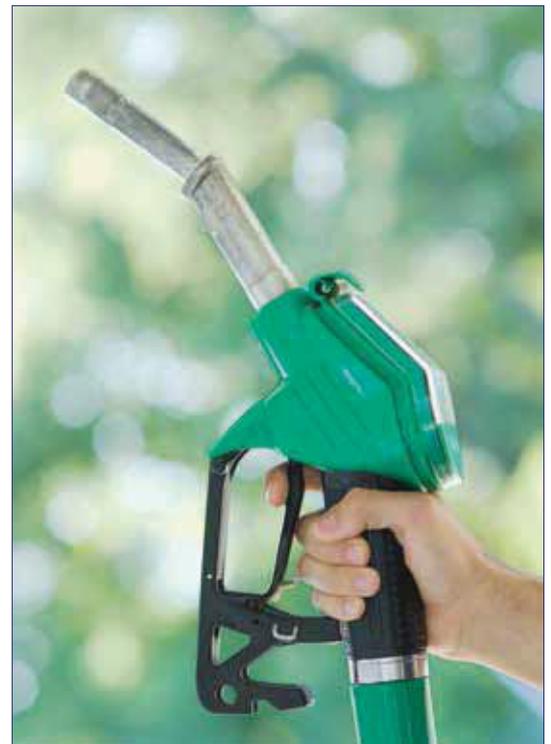
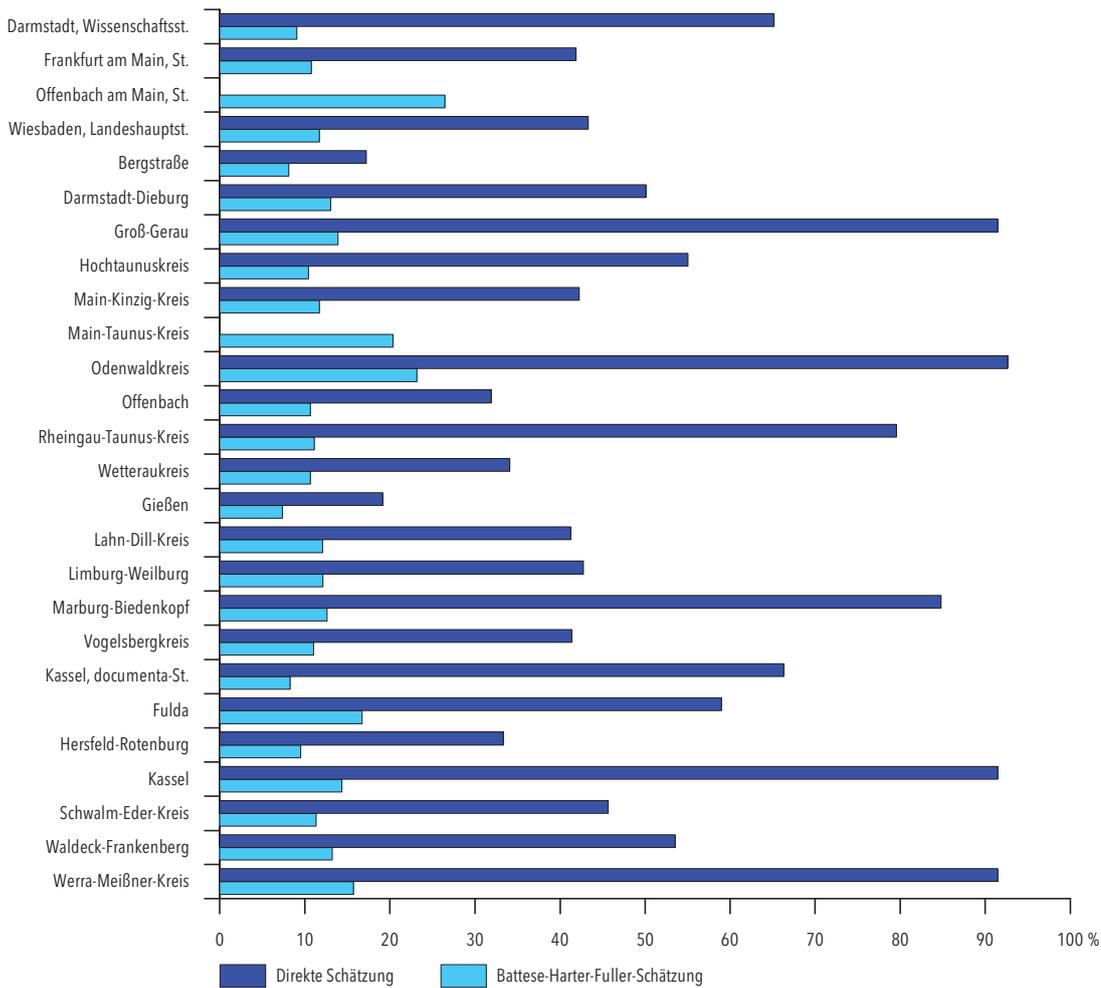


Abbildung 6 stellt für jede der 234 Teilgesamtheiten jeweils die Stichprobengröße in Relation zu dem geschätzten RRMSE der Schätzung mit dem Unit-Level-Modell dar. Es ist zu erkennen, dass das Potenzial der Small-Area-Schätzung nicht unwesentlich von der Stichprobengröße der Teilgesamtheit abhängt. Dies erklärt somit

4) Um Ausreißer geeignet identifizieren zu können, wurden die Cook-Distanz, die Leverage-Werte der Regression sowie die skalierten quadrierten Residuen als Kriterium herangezogen (siehe Rao & Molina, 2015: 114ff & 228ff; Zewotir & Galpin, 2007)

5) Da der direkte Horvitz-Thompson-Schätzer eine unverzerrte Schätzfunktion ist, entspricht sein RRMSE dem relativen Standardfehler.

**Abbildung 5: Geschätzter RRMSE<sup>1)</sup> einer direkten Schätzung und der Schätzung mit dem Battese-Harter-Fuller-Verfahren der Beschäftigten insgesamt im Einzelhandel mit Motorenkraftstoffen (Tankstellen)<sup>2)</sup> 2014 nach Landkreisen und kreisfreien Städten (St.)**



1) Relative Root Mean Squared Error. – 2) Wirtschaftszweig 47.3.

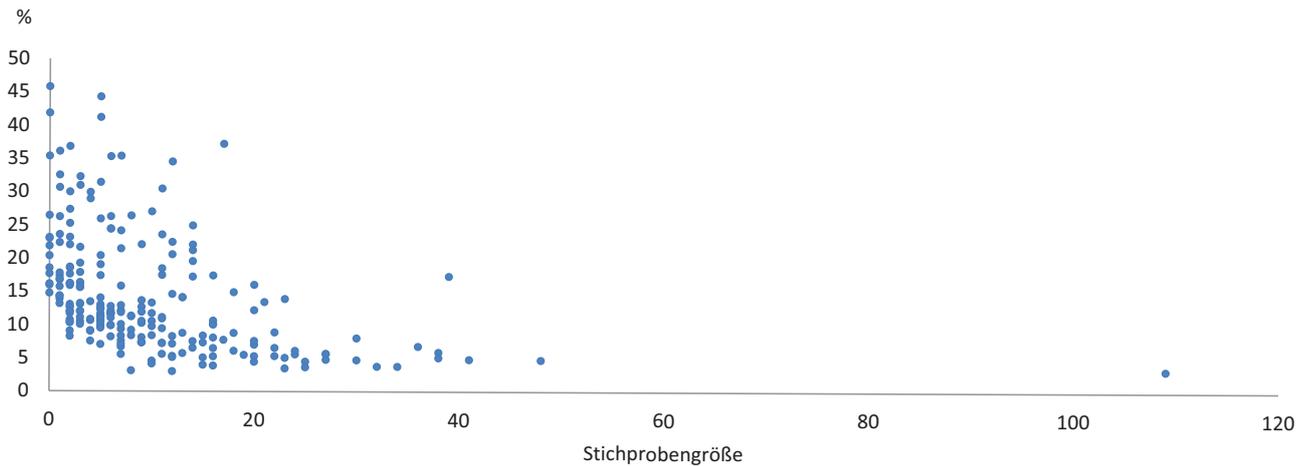
auch den vergleichsweise hohen RRMSE der zwei voll-synthetischen Schätzwerte für die Stadt Offenbach und den Main-Taunus-Kreis.

Allerdings kann der Grafik ebenfalls entnommen werden, dass die Schätzwerte der Beschäftigtenzahl einiger Teilgesamtheiten im Hinblick auf ihren RRMSE deutlich über 20 % liegen. Somit führen Small-Area-Verfahren zwar in der Regel zu einer deutlich präziseren und stabileren Schätzung, dennoch kann insbesondere bei besonders stark untergliederten Teilgesamtheiten nicht gewährleistet werden, dass die resultierenden Schätzwerte in jedem Fall einen RRMSE von unter 20 % aufweisen.

Allerdings sei an dieser Stelle erneut darauf hingewiesen, dass es sich bei den dargestellten RRMSE-Werten um Schätzungen handelt, die ebenfalls entsprechenden Schwankungen

unterliegen. Auch stützt sich das verwendete Battese-Harter-Fuller-Schätzverfahren auf sehr strenge Verteilungsannahmen, die – sofern sie erfüllt sind – einen deutlichen Präzisionsgewinn gegenüber direkten Schätzverfahren erbringen. Unternehmensdaten – so auch die Merkmale der hessischen Einzelhandelsunternehmen – weisen typischerweise sehr rechtsschiefe Strukturen auf. Diese äußern sich durch das Vorhandensein vieler kleiner und mittlerer Unternehmen sowie weniger sehr großer Unternehmen, die die Modelle des Unit-Level-EBLUPs stark dominieren und die Bedingungen streng genommen verletzen. Potenzielle Lösungsansätze, sogenannte robuste Schätzmethoden, werden in letzter Zeit intensiv diskutiert. Eine von Sinha und Rao (2009) entwickelte robuste Erweiterung des EBLUPs konnte im Rahmen der parallel durchgeführten

Abbildung 6: RRMSE<sup>1)</sup> des Battese-Harter-Fuller-Schätzers und Stichprobengröße je Teilgesamtheit in Relation



1) Relative Root Mean Squared Error.

Monte-Carlo-Simulationsstudie (siehe Manecke, 2017) allerdings nicht vollständig im Hinblick auf die relative Verzerrung und den RRMSE der ermittelten Schätzwerte überzeugen. Hier hat der vorgestellte ausreißerbereinigte Battese-Harter-Fuller-Schätzer größeres Potenzial gezeigt, weswegen er für die vorgestellte Untersuchung verwendet worden ist. Allerdings ist nicht vollständig geklärt, wie stark sich der Ausschluss von Ausreißern aus dem Modell auf die Schätzung des RRMSE auswirkt. Die Ergebnisse der Simulationsstudie zeigen jedoch, dass ein Präzisionsgewinn der obig dargestellten Größenordnung durchaus realistisch ist.

Aus den dargestellten Gründen dienen die im Rahmen der Untersuchung ermittelten und im Fol-

genden tabellierten Ergebnisse zunächst einem ersten Überblick und einer Verdeutlichung des Potenzials von Small-Area-Verfahren. Dennoch ist die Verletzung der Modellannahmen durch die vorliegenden rechtsschiefen Unternehmensdaten eine Problematik, die sich langfristig nicht ignorieren lässt. Um die Auswirkungen einer Nichterfüllung der Modellannahmen besser einschätzen zu können und um potenziell geeignetere robuste Small-Area-Schätzmethoden zu finden, sind weitere Simulationsstudien notwendig.

### 3. Ergebnisdarstellung

Die Verwendung von Small-Area-Methoden schafft die Voraussetzung für eine kleinräumige Auswertung der Jahreserhebung im Ein-

Tabelle 2: Kennzahlen der Einzelhandelsunternehmen (ohne Handel mit Kraftfahrzeugen) der Wissenschaftsstadt Darmstadt 2014 nach Wirtschaftszweigen

WZ-Gruppe/ WZ-Abteilung	Wirtschaftszweig	Unternehmen	Beschäftigte	Umsatz	Bruttoentgelte
				Mill. Euro	
47	Einzelhandel (ohne Handel mit Kraftfahrzeugen)	595	4 823	547	79
47.1	Einzelhandel mit Waren verschiedener Art (in Verkaufsräumen)	71	570	39	3
47.2	Einzelhandel mit Nahrungs- und Genussmitteln, Getränken und Tabakwaren (in Verkaufsräumen)	81	332	25	4
47.3	Einzelhandel mit Motorenkraftstoffen (Tankstellen)	9	119	25	1
47.4	Einzelhandel mit Geräten der Informations- und Kommunikationstechnik (in Verkaufsräumen)	21	160	35	4
47.5	Einzelhandel mit sonstigen Haushaltsgeräten, Textilien, Heimwerker- und Einrichtungsbedarf (in Verkaufsräumen)	67	385	34	8
47.6	Einzelhandel mit Verlagsprodukten, Sportausrüstungen und Spielwaren (in Verkaufsräumen)	46	463	44	6
47.7	Einzelhandel mit sonstigen Gütern (in Verkaufsräumen)	242	2 377	278	43
47.8	Einzelhandel an Verkaufsständen und auf Märkten	9	35	1	0
47.9	Einzelhandel, nicht in Verkaufsräumen, an Verkaufsständen oder auf Märkten	49	382	65	10

zelhandel, die anhand von herkömmlichen direkten Schätzverfahren nicht ohne weiteres möglich wäre. So lässt sich jede im Statistischen Bericht zur Einzelhandelserhebung (siehe Hessisches Statistisches Landesamt, 2016) veröffentlichte landesweite Kennzahl ebenfalls für die einzelnen Landkreise bzw. kreisfreien Städte schätzen. Zur Verdeutlichung zeigt Tabelle 2 – in Anlehnung an Tabelle 1 – die ermittelten branchenspezifischen Schätzwerte der Variablen „Umsatz“, „Beschäftigte“ sowie „Bruttoentgelte“ für die Einzelhandelsunternehmen der kreisfreien Stadt Darmstadt.

Die für den gesamten WZ 47 ausgewiesenen Werte sind die jeweilige Summe der für die einzelnen WZ-Dreisteller ermittelten Schätzwerte. Diese entsprechen allerdings nicht zwangsläufig den anhand des branchenübergreifenden Modells für Darmstadt geschätzten Werten des gesamten Einzelhandels. So beträgt der modellbasierte Schätzwert der in der Wissenschaftsstadt Darmstadt im Einzelhandel Beschäftigten 4787. Die Summe des Umsatzes wurde zudem auf 546 Mill. Euro und die Summe der Bruttoentgelte auf 75 Mill. Euro geschätzt. Diese sogenannte fehlende vertikale Kohärenz ist ein erkennbares Problem des Battese-Harter-Fuller-Schätzers und vieler geläufiger Small-Area-Verfahren generell. Konkret bedeutet das, dass die Aggregation der für Subgruppen ermittelten Schätzwerte (hier: WZ-Dreisteller) nicht zwangsläufig dem übergeordneten Schätzwert (hier: Einzelhandel insgesamt) entspricht. Dies ist darauf zurückzuführen, dass zur Schätzung der unter- und übergeordneten Werte aus Präzisionsgründen jeweils separate Modelle verwendet worden sind. Potenzielle Lösungen für diese Problematik bieten sogenannte Benchmarking-Ansätze, die sich aktuell stark in der Entwicklung befinden. Dies sind Verfahren, die auf eine Kohärenz auf den verschiedenen Schätzebenen abzielen.

Darüber hinaus lassen sich die Schätzwerte sämtlicher Einzelhandelskennzahlen – sowohl branchenabhängig als auch branchenunabhängig – anhand von Small-Area-Verfahren auf alle hessischen Landkreise bzw. kreisfreien Städte aufgliedern und ermöglichen so eine Bewertung der einzelhandelsbezogenen Attraktivität einzelner Regionen. Tabelle 3 gibt einen

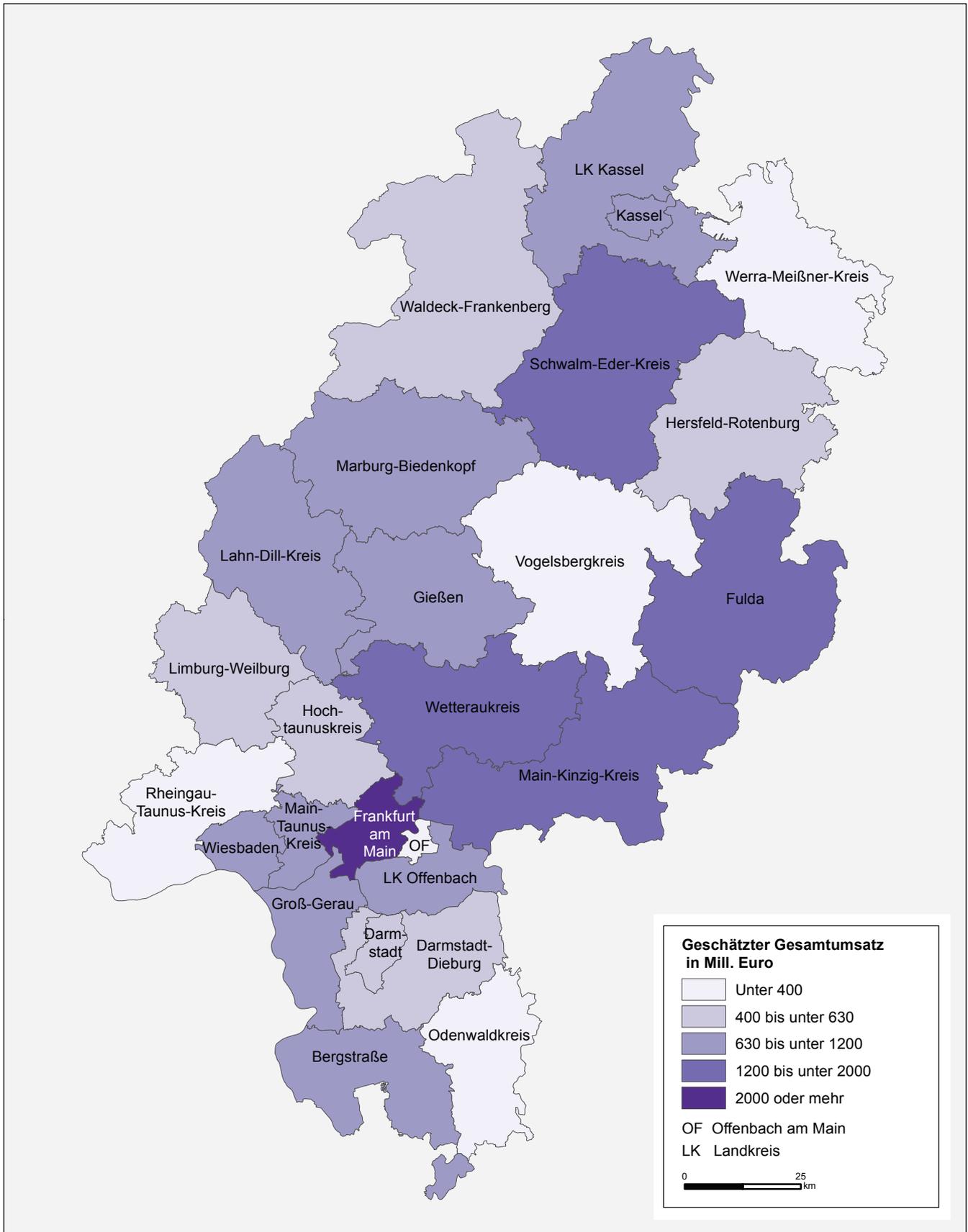
**Tabelle 3: Kennzahlen der Einzelhandelsunternehmen (ohne Handel mit Kraftfahrzeugen) 2014 nach Landkreisen und kreisfreien Städten (St.)**

Landkreis/kreisfreie Stadt (St.)	Unternehmen	Beschäftigte	Umsatz	Bruttoentgelte
			Mill. Euro	
Darmstadt, Wissenschaftsst.	595	4 787	546	75
Frankfurt am Main, St.	2 644	17 626	2 663	320
Offenbach am Main, St.	484	2 525	265	30
Wiesbaden, Landeshauptst.	1 167	7 020	876	81
Bergstraße	1 071	6 685	829	81
Darmstadt-Dieburg	1 024	5 729	562	61
Groß-Gerau	771	6 742	1 114	105
Hochtaunuskreis	946	5 532	586	71
Main-Kinzig-Kreis	1 633	12 409	1 795	176
Main-Taunus-Kreis	812	5 240	725	83
Odenwaldkreis	401	2 301	265	27
Offenbach	1 454	7 220	857	82
Rheingau-Taunus-Kreis	626	3 474	339	35
Wetteraukreis	1 194	8 642	1 652	123
Gießen	1 021	8 499	1 159	104
Lahn-Dill-Kreis	878	6 199	700	69
Limburg-Weilburg	676	4 034	414	46
Marburg-Biedenkopf	867	7 277	746	87
Vogelsbergkreis	394	2 520	306	29
Kassel, documenta-St.	807	6 654	730	79
Fulda	930	12 205	1 957	198
Hersfeld-Rotenburg	427	4 000	442	47
Kassel	767	5 814	632	59
Schwalm-Eder-Kreis	683	11 021	1 636	148
Waldeck-Frankenberg	619	4 441	459	49
Werra-Meißner-Kreis	402	2 852	307	36
Land Hessen	23 293	171 448	22 562	2 300

Überblick über die mit dem Battese-Harter-Fuller-Modell geschätzten Kennzahlen des gesamten Einzelhandels auf Ebene der Landkreise und kreisfreien Städte. Der RRMSE der Schätzung der Beschäftigten entspricht den in Abbildung 4 dargestellten Werten. Der RRMSE der Umsatzzschätzung ist jeweils von vergleichbarer Höhe. Lediglich bei der Schätzung der Summe der Bruttoentgelte weisen die Schätzwerte für 6 Landkreise<sup>6)</sup> jeweils einen geschätzten RRMSE auf, der zwischen 20 % und 26 %, d. h. leicht über dem Grenzwert, liegt. Dies ist auf einen weniger starken Zusammenhang zwischen den Hilfsvariablen und der zu schätzenden Variable zurückzuführen. Die Kennzahlen für das Land Hessen (siehe Tabelle 3) entstehen durch die Aggregation der einzelnen kreisspezifischen Schätzwerte. Auch

6) Hiervon sind die Schätzungen für den Landkreis Darmstadt-Dieburg, den Odenwaldkreis, den Landkreis Offenbach, den Rheingau-Taunus-Kreis, den Vogelsbergkreis und die Stadt Offenbach betroffen.

Abbildung 7: Mit dem Battese-Harter-Fuller-Verfahren geschätzter Gesamtumsatz im Einzelhandel (ohne Handel mit Kraftfahrzeugen)<sup>1)</sup> 2014



1) Wirtschaftszweig 47.

hier ist aus den oben genannten Gründen keine vertikale Kohärenz zu den anhand des direkten Horvitz-Thompson-Verfahrens geschätzten und im Statistischen Bericht bzw. Tabelle 1 angegebenen Schätzwerten gegeben. In Anlehnung an Tabelle 3 illustriert Abbildung 7 den branchenübergreifenden Einzelhandelsumsatz 2014 je Verwaltungsbezirk in Mill. Euro. Im Vergleich zu den anderen Landkreisen bzw. kreisfreien Städten wurde in Frankfurt am Main im Jahr 2014 mit schätzungsweise 2,66 Mrd. Euro der höchste Einzelhandelsumsatz erzielt, was insbesondere auf die hohe Zahl der dort ansässigen Einzelhandelsunternehmen zurückzuführen ist. Während der hohe Gesamtumsatz im Wetteraukreis sowie im Main-Kinzig-Kreis ebenfalls größtenteils durch überdurchschnittlich viele Einzelhandelsunternehmen zu begründen ist, scheint er im Landkreis Fulda sowie im Schwalm-Eder-Kreis vor allem durch sehr umsatzstarke Einzelhandelsunternehmen herbeigeführt worden zu sein.

#### 4. Fazit

Auf der gezeigten feinen Gliederungsebene ist es nicht möglich, mit herkömmlichen direk-

ten Schätzverfahren zuverlässige Prädiktionen durchzuführen. Durch die Verwendung von Hilfsinformationen aus dem Statistischen Unternehmensregister im Rahmen von Small-Area-Verfahren kann jedoch ein klarer Präzisionsgewinn erzielt werden. Die Registerinformationen unterstützen dabei insbesondere Schätzungen für Teilgesamtheiten, in denen nur sehr wenige bis gar keine Unternehmen befragt wurden. Allerdings muss stets geprüft werden, ob die resultierenden Schätzwerte hinsichtlich ihrer Präzision den Anforderungen gerecht werden können. Die Problematik, dass der vorliegende Unternehmensdatensatz streng genommen die Modellannahmen des Battese-Harter-Fuller-Modells verletzt, lässt sich zudem langfristig nicht ignorieren. Um das Ausmaß der Modellverletzung beurteilen und das Potenzial alternativer Schätzverfahren abschätzen zu können, sind weitere Untersuchungen sowie Simulationsstudien durchzuführen.

**Julia Manecke;** Tel: 0611 3802-868;  
E-Mail: [redaktion@statistik.hessen.de](mailto:redaktion@statistik.hessen.de)

#### Literaturverzeichnis

- Battese, G. E., Harter, R. M., & Fuller, W.A. (1988): An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data, *Journal of the American Statistical Association* 83(401), Seiten 28-36.
- González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D., & Santamaría, L. (2008): Bootstrap mean squared error of a small-area EBLUP. *Journal of Statistical Computation and Simulation*, 78(5), Seiten 443-462.
- Gottfried, P. (2017): Der Einsatz von Small-Area-Verfahren – erste Erfahrungen mit Area-Level-Schätzungen. *Staat und Wirtschaft* 2017(1).
- Hessisches Statistisches Landesamt (2016): Statistische Berichte: Strukturdaten des Einzelhandels in Hessen im Jahr 2014. Ergebnisse der Jahrerhebung. Kennziffer G 1 3 - j/14.
- Horvitz, D. G., & Thompson, D. J. (1952): A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260), Seiten 663-685.
- Manecke, J. (2017): Regionale Auswertungen von Unternehmensstatistiken: Methoden und Anwendungen im Kontext der Small Area-Statistik. Masterarbeit (unveröffentlicht).
- Münnich, R., Burgard, J.P., & Vogt, M. (2013). Small Area-Statistik: Methoden und Anwendungen. *ASTA Wirtschafts- und Sozialstatistisches Archiv*, 6(3-4), Seiten 149-191.
- Rao, J. N. K., & Molina, I. (2015): *Small area estimation*. John Wiley & Sons.
- Sinha, S. K., & Rao, J. N. K. (2009): Robust small area estimation. *Canadian Journal of Statistics*, 37(3), Seiten 381-399.
- Statistisches Bundesamt (2016a): Jahrerhebung im Handel sowie in der Instandhaltung und Reparatur von Kfz. 2014. *Qualitätsbericht*. Wiesbaden.
- Statistisches Bundesamt (2016b): *Qualitätsbericht, Einkommens- und Verbrauchsstichprobe EVS 2013*, Wiesbaden.
- Statistisches Bundesamt (2008): *Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008)*. Wiesbaden.
- Vaida, F., & Blanchard, S. (2005): Conditional Akaike Information for Mixed-Effect Models. *Biometrika*, 92(2), Seiten 351-370.
- Zewotir, T., & Galpin, J. S. (2007): A unified approach on residuals, leverages and outliers in the linear mixed model. *Test*, 16(1), Seiten 58-75.

## Erwerbstätigenquote

1) EFTA steht für „Europäische Freihandelsassoziation“. EFTA-Länder: Island, Liechtenstein, Norwegen und die Schweiz. Kandidatenländer: Albanien, Montenegro, die ehemalige jugoslawische Republik Mazedonien, Serbien und die Türkei. Die Datenverfügbarkeit für die Regionen variiert je nach Indikator. Datenabruf am 8.2.2017.

2) Das Erwerbspersonenpotenzial wird mit der Gruppe „15 Jahre und älter“ in der umfassendsten Abgrenzung nachgewiesen.

3) International vergleichbare Daten zur Beteiligung am Erwerbsleben werden im Rahmen der Arbeitskräftefeststichprobe der Europäischen Union ermittelt. Diese ist fester Bestandteil des Mikrozensus, einer jährlich in Deutschland durchgeführten Haushaltsbefragung bei rund 1 % der Bevölkerung. Erwerbstätige werden ebenfalls im Rahmen der Volkswirtschaftlichen Gesamtrechnungen ermittelt.

4) NUTS steht für „Nomenclature des Unités territoriales statistiques“ (Klassifikation der Gebiets-einheiten für die Statistik). In Deutschland entspricht die NUTS-1-Ebene den Bundesländern, die NUTS-2-Ebene in der Regel den Regierungsbezirken und die NUTS-3-Ebene den Landkreisen und kreisfreien Städten. Die Grundlage für die Datenauswahl bildet das Statistische Jahrbuch der Regionen: <http://ec.europa.eu/eurostat/publications/regional-yearbook>

5) Vgl. EUROPA 2020 - Eine Strategie für intelligentes, nachhaltiges und integratives Wachstum, Mitteilung der Kommission, Brüssel, 3.3.2010.

6) Vgl. Hessen in Europa zur „Erwerbslosenquote“ in Staat und Wirtschaft in Hessen, Ausgabe 2/2016, Februar 2016.

Nach der Geburtenziffer im sechsten Teil der Serie „Hessen in Europa“ steht im siebten Teil die Erwerbstätigenquote im Fokus<sup>1)</sup>. Die Erwerbstätigenquote ist der prozentuale Anteil der Erwerbstätigen an der erwerbsfähigen Bevölkerung. Unter Berücksichtigung längerer Ausbildungszeiten und dem damit verbundenen höheren Alter beim Eintritt in das Erwerbsleben, erfolgt der Nachweis in diesem Beitrag für die Gruppe der 20- bis 64-Jährigen. Die Erwerbspersonen umfassen die Erwerbstätigen und die Erwerbslosen. Zusammen bilden sie das Potenzial an Personen, die dem Arbeitsmarkt zur Verfügung stehen. Die Erwerbspersonen umfassten in der EU-28 im Jahr 2015 243,73 Mill. Personen (Hessen 3,18 Mill.)<sup>2)</sup>. Davon waren 220,85 Mill. erwerbstätig (Hessen 3,05 Mill.) und 22,88 Mill. erwerbslos (Hessen 0,13 Mill.)<sup>3)</sup>.

Im Zuge der Finanzkrise nahm die Erwerbstätigenquote der EU-28 von 70,2 % im Jahr 2008 kontinuierlich bis auf 68,3 % im Jahr 2013 ab. 2015 lag die Erwerbstätigenquote wieder bei 70 % und damit noch unter dem Vorkrisenniveau.

## Wie stellten sich die Erwerbstätigenquoten in den europäischen und insbesondere in den hessischen Regionen im Jahr 2015 dar<sup>4)</sup>?

Unter den 100 Regionen in der EU-28 mit einer Erwerbstätigenquote von 75 % oder mehr, auf der Karte dunkel dargestellt, liegen 33 in Deutschland und 30 im Vereinigten Königreich. Eine Erwerbstätigenquote in Höhe von 75 % ist eines der „Europa-2020-Ziele“<sup>5)</sup>. Die meisten der 35 Regionen mit der geringsten Erwerbstätigenquote, auf der Karte hell dargestellt, verteilen sich auf die von der Finanzkrise am stärksten betroffenen Länder Griechenland (11 Regionen), Spanien (8 Regionen) und Italien (8 Regionen). Im Gegenzug nahm in diesen Regionen nach der Finanzkrise die Erwerbslosigkeit zu<sup>6)</sup>. In der EU-28 reichte die Spanne zwischen der geringsten und höchsten Erwerbstätigenquote 2015 von

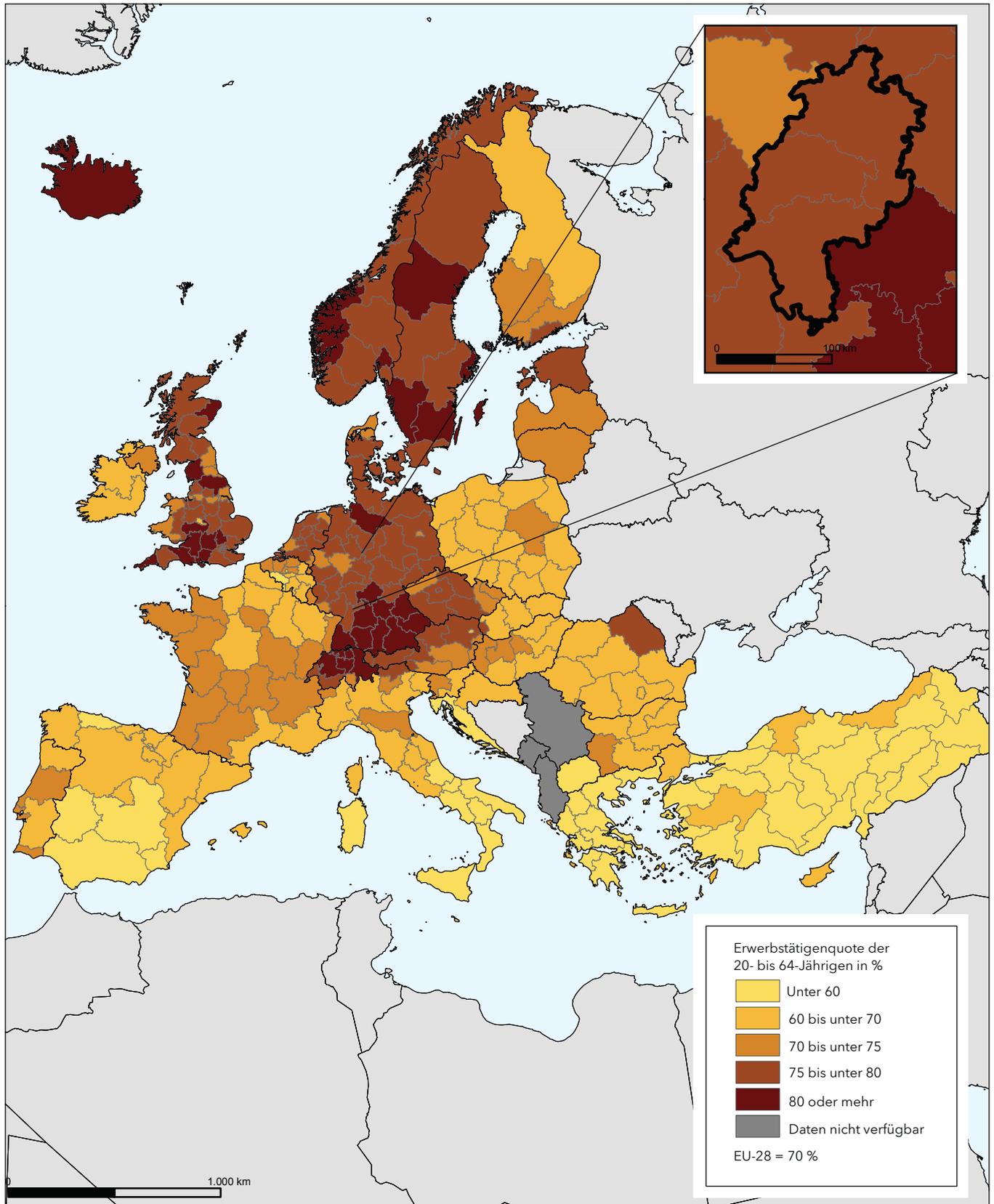
42,1 % in der italienischen Region Kalabrien bis zu 86,7 % in der finnischen Region Åland. Unter den 25 Regionen mit einer Erwerbstätigenquote von 80 % oder mehr befinden sich jeweils 10 Regionen in Deutschland und dem Vereinigten Königreich sowie 4 Regionen in Schweden.

In Hessen lagen die Erwerbstätigenquoten bei 76,3 % im Regierungsbezirk (RB) Gießen, 78,1 % im RB Kassel und 78,2 % im RB Darmstadt. Alle hessischen RB lagen somit über dem EU-28-Durchschnitt von 70 %. Mit dem RB Gießen vergleichbare Regionen sind Niederösterreich (76,3 %) oder Midtjylland (76,2 %) in Dänemark. Mit dem RB Kassel vergleichbare Regionen sind die tschechische Hauptstadtregion Prag und Shropshire and Staffordshire im Vereinigten Königreich (jeweils 78,1 %). Die dänische Region Hovedstaden und die schwedische Region Norra Mellansverige (jeweils 78,2 %) wiesen eine mit dem RB Darmstadt vergleichbare Erwerbstätigenquote auf.

Von den EFTA-Staaten Island, Liechtenstein, Norwegen und Schweiz liegen mit Ausnahme von Liechtenstein Daten vor. Die Spanne reicht in dieser Gruppe von 74,7 % in der Region Tessin in der Schweiz bis 86,5 % auf Island. Somit lagen, mit Ausnahme der Region Tessin, die Quoten aller verfügbaren EFTA-Regionen über 75 %. In der Gruppe der EU-Kandidatenländer liegen Werte für die Türkei und Mazedonien vor. Alle Regionen lagen deutlich unter dem EU-28-Durchschnitt. Die Spanne reichte von 35,4 % in der Region Mardin bis 62,1 % in der Region Trabzon, die sich beide in der Türkei befinden. Die mazedonischen NUTS-2- und NUTS-1-Regionen sind identisch und entsprechen dem Land insgesamt. Hier lag der Wert 2015 bei 51,9 %.

**Benedikt Kull**; Tel: 0611 3802-824;  
E-Mail: [benedikt.kull@statistik.hessen.de](mailto:benedikt.kull@statistik.hessen.de)

Erwerbstätigenquote in der EU-28, den EFTA-Staaten und den Beitrittskandidaten 2015 nach NUTS-2-Regionen<sup>1)</sup>



1) Quelle: Eurostat, Online Daten Code [lfst\_r\_lfe2emprt].

© EuroGeographics bezüglich der Verwaltungsgrenzen  
 © Hessisches Statistisches Landesamt, Wiesbaden, 2017  
 Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.

## Ausgewählte Daten zur wirtschaftlichen Entwicklung in Hessen

Jahr Monat Vierteljahr	Verarbeitendes Gewerbe <sup>1)</sup>			Bauhauptgewerbe <sup>3)</sup>		Einzelhandel	Preise	Arbeitsmarkt <sup>7)</sup>				
	Volumenindex des Auftragseingangs <sup>2)</sup> 2010 = 100			Beschäftigte	Index des Auftragseingangs <sup>4)</sup> 2010 = 100	Beschäftigte	Index der Umsätze <sup>5)</sup> 2010 = 100	Verbraucherpreisindex <sup>6)</sup> 2010 = 100	Arbeitslose <sup>8)</sup>	Arbeitslosenquote <sup>9)</sup>	Gemeldete Arbeitsstellen <sup>10)</sup>	Kurzarbeit (Personen) <sup>8)</sup>
	insgesamt	Inland	Ausland									
Grundzahlen <sup>11)</sup>												
2014 D	99,1	90,5	107,4	360 297	128,8	25 235	104,1	106,1	184 378	5,7	35 734	5 643
2015 D	98,8	93,3	104,7	361 736	125,5	24 616	106,5	106,3	177 944	5,5	42 549	5 614
2016 D	95,4	90,3	100,9	361 809	149,3	26 725	108,6	106,7	172 826	5,3	48 688	...
2015 November	98,9	97,5	101,7	363 622	121,8	24 608	107,8	106,4	167 209	5,1	45 965	5 987
Dezember	87,3	84,4	90,9	361 638	127,0	24 199	126,3	106,3	168 165	5,1	44 186	7 000
4. Vj. D	94,7	91,8	98,5	363 036	121,3	24 555	115,3	106,4	.	.	.	.
2016 Januar	98,1	84,8	109,9	359 718	94,2	25 415	97,7	105,3	182 460	5,6	43 713	10 682
Februar	96,1	90,2	102,3	360 240	137,7	25 617	98,9	105,8	182 811	5,6	45 366	10 747
März	96,9	90,3	103,6	359 930	169,3	25 750	108,7	106,6	179 393	5,5	47 315	8 790
1. Vj. D	97,0	88,4	105,3	359 963	133,7	25 594	101,8	105,9	.	.	.	.
April	95,9	93,0	99,8	359 684	146,4	26 113	110,4	106,2	176 006	5,4	47 929	4 408
Mai	89,4	86,4	93,3	359 450	182,1	26 625	107,7	106,6	171 125	5,2	48 699	4 398
Juni	104,3	99,5	109,7	360 333	192,2	26 970	108,7	106,5	168 571	5,1	49 999	3 365
2. Vj. D	96,5	93,0	100,9	359 822	173,6	26 569	108,9	106,4	.	.	.	.
Juli	93,7	86,1	101,0	361 885	174,1	27 252	107,7	106,9	172 954	5,3	50 898	...
August	87,9	88,3	89,0	363 585	166,0	27 574	105,0	107,0	175 598	5,3	51 431	...
September	93,5	89,4	98,2	365 231	148,4	27 639	105,4	107,0	169 428	5,1	50 900	...
3. Vj. D	91,7	87,9	96,1	363 567	162,8	27 488	106,1	107,0	.	.	.	.
Oktober	95,7	91,9	100,3	364 693	110,6	27 518	110,2	107,3	165 656	5,0	50 976	...
November	100,9	95,8	106,6	364 338	124,3	27 407	114,4	107,3	163 791	5,0	49 969	...
Dezember	92,0	87,9	96,6	362 619	145,7	26 824	128,2	108,3	166 119	5,0	47 065	...
4. Vj. D	96,2	91,9	101,2	363 883	126,9	27 250	117,6	107,6	.	.	.	.
Zu- bzw. Abnahme (-) jeweils gegenüber dem Vorjahr bzw. dem gleichen Zeitraum des Vorjahres in % <sup>11)</sup>												
2014 D	0,8	- 4,2	4,4	1,1	- 3,5	- 2,1	1,3	0,8	- 0,7	.	4,6	- 26,2
2015 D	- 0,3	3,1	- 2,5	0,4	- 2,6	- 2,5	2,3	0,2	- 3,5	.	19,1	- 0,5
2016 D	- 3,4	- 3,2	- 3,6	0,0	8,6	18,9	2,0	0,4	- 2,9	.	14,4	...
2015 November	- 7,7	8,8	- 16,8	0,1	14,5	- 5,2	0,5	0,3	- 4,0	.	20,2	48,2
Dezember	- 12,7	1,8	- 20,8	0,1	- 13,9	- 4,0	4,0	0,3	- 3,5	.	21,4	- 5,8
4. Vj.	- 7,7	3,8	- 14,5	0,1	- 6,0	- 4,8	2,8	0,3	.	.	.	.
2016 Januar	- 2,6	- 7,0	0,3	- 0,1	- 18,6	3,7	- 1,2	0,3	- 3,8	.	22,2	- 6,4
Februar	- 2,4	- 1,5	- 2,9	0,0	53,4	5,2	5,9	- 0,3	- 3,1	.	17,3	- 8,2
März	- 11,9	- 9,9	- 13,2	- 0,4	29,1	5,5	0,6	0,1	- 2,6	.	15,4	- 6,8
1. Vj.	- 5,9	- 6,3	- 5,6	- 0,2	19,2	4,8	1,7	0,0	.	.	.	.
April	- 9,4	- 1,5	- 14,3	- 0,3	25,3	6,5	3,1	- 0,3	- 3,1	.	18,4	10,4
Mai	- 7,6	- 4,0	- 10,0	- 0,2	42,1	8,1	2,6	0,0	- 3,5	.	18,1	38,5
Juni	- 2,2	0,9	- 4,4	- 0,1	34,0	9,7	2,9	0,0	- 3,3	.	17,0	...
2. Vj.	- 6,4	- 1,4	- 9,6	- 0,2	34,1	8,1	2,8	- 0,1	.	.	.	.
Juli	- 8,0	- 13,6	- 4,2	0,3	31,0	10,2	- 2,0	0,2	- 2,8	.	15,6	...
August	8,5	3,4	12,5	0,2	33,0	10,2	5,2	0,3	- 3,0	.	13,2	...
September	- 6,6	- 4,1	- 8,3	0,2	- 6,8	10,6	0,4	0,6	- 2,9	.	12,6	...
3. Vj.	- 2,8	- 5,2	- 1,1	0,2	17,1	10,4	1,1	0,4	.	.	.	.
Oktober	- 2,3	- 1,7	- 2,6	0,2	- 3,8	10,7	- 1,5	0,8	- 2,9	.	11,1	...
November	2,0	- 1,7	4,8	0,2	2,0	11,4	6,1	0,8	- 2,0	.	8,7	...
Dezember	5,4	4,1	6,3	0,3	14,7	10,8	1,5	1,9	- 1,2	.	6,5	...
4. Vj.	1,6	0,1	2,7	0,2	4,6	11,0	2,0	1,1	.	.	.	.
Zu- bzw. Abnahme (-) jeweils gegenüber dem Vormonat bzw. dem Vorquartal in % <sup>11)</sup>												
2015 November	0,9	4,3	- 1,3	- 0,1	5,9	- 1,0	- 3,7	0,0	- 2,0	.	0,2	78,2
Dezember	- 11,7	- 13,4	- 10,6	- 0,5	4,2	- 1,7	17,2	- 0,1	0,6	.	- 3,9	16,9
4. Vj.	0,4	- 1,0	1,3	0,1	- 12,7	- 1,4	9,9	- 0,2	.	.	.	.
2016 Januar	12,4	0,5	20,9	- 0,5	- 25,8	5,0	- 22,6	- 0,9	8,5	.	- 1,1	52,6
Februar	- 2,0	6,4	- 6,9	0,1	46,2	0,8	1,2	0,5	0,2	.	3,8	0,6
März	0,8	0,1	1,3	- 0,1	23,0	5,5	9,9	0,8	- 1,9	.	4,3	18,2
1. Vj.	2,4	- 3,7	6,9	- 0,8	10,3	4,2	- 11,7	- 0,5	.	.	.	.
April	- 1,0	3,0	- 3,7	- 0,1	- 13,5	1,4	1,6	- 0,4	- 1,9	.	1,3	49,9
Mai	- 6,8	- 7,1	- 6,5	- 0,1	24,4	2,0	- 2,4	0,4	- 2,8	.	1,6	- 0,2
Juni	16,7	15,2	17,6	0,2	5,6	1,3	0,9	- 0,1	- 1,5	.	2,7	...
2. Vj.	- 0,5	5,2	- 4,2	- 0,0	29,8	3,8	7,0	0,5	.	.	.	.
Juli	- 10,2	- 13,5	- 7,9	0,4	- 9,4	1,0	- 0,9	0,4	2,6	.	1,8	...
August	- 6,2	2,6	- 11,9	0,5	- 4,7	1,2	- 2,5	0,1	1,5	.	1,0	...
September	6,4	1,2	10,3	0,5	- 10,6	0,2	0,4	0,0	- 3,5	.	1,0	...
3. Vj.	- 5,0	- 5,5	- 4,8	1,0	- 6,2	3,5	- 2,6	0,6	.	.	.	.
Oktober	2,4	2,8	2,1	- 0,1	- 25,5	- 0,4	4,6	0,3	- 2,2	.	0,1	...
November	5,4	4,2	6,3	- 0,1	12,4	- 0,4	3,8	0,0	- 1,1	.	- 2,0	...
Dezember	5,4	4,1	6,3	- 0,5	17,2	- 2,1	12,1	0,9	1,4	.	- 5,8	...
4. Vj.	4,9	4,6	5,3	0,1	- 22,1	- 0,9	10,8	0,6	.	.	.	.

1) Einsch. Bergbau sowie Gewinnung von Steinen und Erden. Betriebe mit 50 und mehr Beschäftigten. Ab Januar 2009 Erfassung nach WZ 2008. – 2) Ohne Bergbau und ohne Gewinnung von Steinen und Erden. – 3) Betriebe von Unternehmen mit 20 oder mehr Beschäftigten. – 4) Wertindex. – 5) Ohne Mehrwertsteuer. – 6) Alle privaten Haushalte. – 7) Quelle: Bundesagentur für Arbeit, Regionaldirektion Hessen (vorläufige Werte). Verfahrensbedingte Abweichungen zu den offiziellen Zahlen sind nicht auszuschließen; die aktuellen Werte sind im Internet bei der Bundesagentur für Arbeit als detaillierte Übersichten zu finden. – 8) Bei Monatswerten Stand: Monatsmitte. Änderung der methodischen Grundlagen. – 9) Arbeitslose in % aller zivilen Erwerbspersonen. – 10) Ohne geförderte Stellen des zweiten Arbeitsmarktes. – 11) Gegenüber der letzten Ausgabe teilweise berichtigte Ergebnisse.

Titel	Regionale Gliederung	Dateiformat	Kennziffer
<b>A Bevölkerung, Gesundheitswesen, Gebiet, Erwerbstätigkeit</b>			
Die Krankenhäuser in Hessen am 31. Dezember 2015	L	PDF <sup>1)</sup>	A IV 2
Die Diagnosen der stationär behandelten Patienten/Patientinnen in Hessen 2015	L	PDF <sup>1)</sup>	A IV 9
<b>B Bildung, Kultur, Rechtspflege, Wahlen</b>			
Die allgemeinbildenden Schulen in Hessen 2016, Stand: 01. November 2016 (Erste Ergebnisse)	K	PDF <sup>1)</sup>	B I 1
<b>C Land- und Forstwirtschaft, Fischerei</b>			
Die Gemüseerhebung in Hessen 2016 (Total)	K	PDF <sup>1)</sup>	C I 3 mit C II
Erhebung über den Anbau von Strauchbeeren 2016	L	PDF <sup>1)</sup>	C I 9 mit C II
Die Obsternte in Hessen 2016	L	PDF <sup>1)</sup>	C II 2
Weinbestände in Hessen im Jahr 2016	L	PDF <sup>1)</sup>	C II 3
Rinderbestände und Rinderhaltungen im November 2016	K	PDF <sup>1)</sup>	C III 1-6
Schlachtungen in Hessen im 4. Vierteljahr 2016 (Vorläufiges Ergebnis)	K	PDF <sup>1)</sup>	C III 2
<b>D Gewerbeanzeigen, Unternehmen und Arbeitsstätten, Insolvenzen</b>			
Unternehmen und Betriebe in Hessen (Ergebnisse des Unternehmensregisters, Stand 29.02.2016)	G	PDF <sup>1)</sup>	D II 1
<b>E Produzierendes Gewerbe, Handwerk</b>			
Betriebe, Beschäftigte und Umsatz im Verarbeitenden Gewerbe (einschl. Bergbau und Gewinnung von Steinen und Erden) in Hessen (Betriebe mit im Allgemeinen 50 und mehr Beschäftigten)			
- im Oktober 2016	K	PDF <sup>1)</sup>	E I 1
- im November 2016	K	PDF <sup>1)</sup>	E I 1
- im Dezember 2016	K	PDF <sup>1)</sup>	E I 1
Indizes des Auftragseingangs und des Umsatzes im Verarbeitenden Gewerbe in Hessen			
- im Oktober 2016	L	PDF	E I 3
- im November 2016	L	PDF	E I 3
- im Dezember 2016	L	PDF	E I 3
Investitionen im Verarbeitenden Gewerbe in Hessen im Jahr 2015	K	PDF <sup>1)</sup>	E I 6
Das Bauhauptgewerbe in Hessen			
- im Oktober 2016	K	PDF <sup>1)</sup>	E II 1
- im November 2016	K	PDF <sup>1)</sup>	E II 1
Ergänzungserhebung im hessischen Bauhauptgewerbe vom Juni 2016	K	PDF <sup>1)</sup>	E II 2
Investitionen im Bauhaupt- und Ausbaugewerbe in Hessen im Jahr 2015	L	PDF <sup>1)</sup>	E II 3 mit E III 3
Das Ausbaugewerbe in Hessen im 3. Vierteljahr 2016 (Bauinstallation und sonstiges Ausbaugewerbe)	K	PDF <sup>1)</sup>	E III 1
Jahreserhebung im Ausbaugewerbe in Hessen vom Juni 2016 (Bauinstallation und sonstiges Ausbaugewerbe)	K	PDF <sup>1)</sup>	E III 2
Energieversorgung in Hessen			
- im September 2016	L	PDF <sup>1)</sup>	E IV 1, E IV 2 mit E IV 3
- im Oktober 2016	L	PDF <sup>1)</sup>	E IV 1, E IV 2 mit E IV 3
- im November 2016	L	PDF <sup>1)</sup>	E IV 1, E IV 2 mit E IV 3
Strom- und Gasabsatz sowie Erlöse in Hessen im Jahr 2015	L	PDF <sup>1)</sup>	E IV 5
Das Handwerk in Hessen im 3. Vierteljahr 2016 (Zulassungspflichtiges Handwerk; vorläufige Ergebnisse)	L	PDF <sup>1)</sup>	EV 1
<b>F Wohnungswesen, Bautätigkeit</b>			
Baugenehmigungen in Hessen			
- im Oktober 2016	K	PDF <sup>1)</sup>	F II 1
- im November 2016	K	PDF <sup>1)</sup>	F II 1
- im Dezember 2016	K	PDF <sup>1)</sup>	F II 1
<b>G Handel, Tourismus, Gastgewerbe</b>			
Entwicklung von Umsatz und Beschäftigten im Kfz- und Einzelhandel sowie in der Instandhaltung und Reparatur von Kfz in Hessen (Vorläufige Ergebnisse)			
- im Oktober 2016	L	PDF <sup>1)</sup>	G I 1
- im November 2016	L	PDF <sup>1)</sup>	G I 1

1) Mit Excel-Anhang

Statistische Berichte und sonstige Veröffentlichungen - Kostenfreie Download-Dateien

L = Landesebene – K = Kreisebene – G = Gemeindeebene

Titel	Regionale Gliederung	Dateiformat	Kennziffer
<b>Noch: G Handel, Tourismus, Gastgewerbe</b>			
Entwicklung von Umsatz und Beschäftigten im Großhandel und in der Handelsvermittlung in Hessen (Vorläufige Ergebnisse)			
- im September 2016	L	PDF <sup>1)</sup>	G I 2
- im Oktober 2016	L	PDF <sup>1)</sup>	G I 2
- im November 2016	L	PDF <sup>1)</sup>	G I 2
Die Ausfuhr Hessens (Vorläufige Ergebnisse)			
- im Oktober 2016	L	PDF <sup>1)</sup>	G III 1
- im November 2016	L	PDF <sup>1)</sup>	G III 1
- im Dezember 2016	L	PDF <sup>1)</sup>	G III 1
Die Einfuhr nach Hessen (Vorläufige Ergebnisse)			
- im Oktober 2016	L	PDF <sup>1)</sup>	G III 3
- im November 2016	L	PDF <sup>1)</sup>	G III 3
- im Dezember 2016	L	PDF <sup>1)</sup>	G III 3
Gäste und Übernachtungen im hessischen Tourismus (Vorläufige Ergebnisse)			
- im Oktober 2016	G	PDF <sup>1)</sup>	G IV 1
- im November 2016	G	PDF <sup>1)</sup>	G IV 1
- im Dezember 2016	G	PDF <sup>1)</sup>	G IV 1
Entwicklung von Umsatz und Beschäftigten im Gastgewerbe in Hessen (Vorläufige Ergebnisse)			
- im Oktober 2016	L	PDF <sup>1)</sup>	G IV 3
- im November 2016	L	PDF <sup>1)</sup>	G IV 3
<b>H Verkehr</b>			
Straßenverkehrsunfälle in Hessen (Vorläufige Ergebnisse)			
- im September 2016	K	PDF <sup>1)</sup>	H I 1
- im Oktober 2016	K	PDF <sup>1)</sup>	H I 1
- im November 2016	K	PDF <sup>1)</sup>	H I 1
- im Dezember 2016	K	PDF <sup>1)</sup>	H I 1
Personenverkehr mit Bussen und Bahnen in Hessen im 3. Vierteljahr 2016 (Vorläufige Ergebnisse)			
	L	PDF <sup>1)</sup>	H I 4
Personenverkehr mit Bussen und Bahnen in Hessen im Jahr 2015			
	L	PDF <sup>1)</sup>	H I 5
Binnenschifffahrt in Hessen			
- im September 2016	Häfen	PDF <sup>1)</sup>	H II 1
- im Oktober 2016	Häfen	PDF <sup>1)</sup>	H II 1
- im November 2016	Häfen	PDF <sup>1)</sup>	H II 1
<b>K Sozialleistungen</b>			
Kinder- und Jugendhilfe in Hessen im Jahr 2015 (Adoptionen, Pflegschaften, Vormundschaften, Beistandschaften, Pflege- erlaubnis, Sorgeerklärungen, Maßnahmen des Familiengerichts, vorläufige Schutzmaßnahmen)			
	K	PDF <sup>1)</sup>	KV 5
Die Pflegeeinrichtungen in Hessen am 15. Dezember 2015			
	L	PDF <sup>1)</sup>	KVIII 1
<b>L Öffentliche Finanzen, Personal, Steuern</b>			
Realsteuervergleich in Hessen im Jahr 2015 (Indikatoren für die wirtschaftliche und finanzielle Lage der Gemeinden)			
	G	PDF <sup>1)</sup>	LII 7
Versorgungsempfänger in Hessen			
- am 01. Januar 2015	L	PDF <sup>1)</sup>	LIII 3
- am 01. Januar 2016	L	PDF <sup>1)</sup>	LIII 3
Die Erbschaft- und Schenkungssteuer 2015 in Hessen			
	L	PDF <sup>1)</sup>	LIV 5
<b>M Preise und Preisindizes</b>			
Verbraucherpreisindex in Hessen			
- im November 2016	L	PDF <sup>1)</sup>	MI 2
- im Dezember 2016	L	PDF <sup>1)</sup>	MI 2
- im Januar 2017	L	PDF <sup>1)</sup>	MI 2
Verbraucherpreisindex in Hessen 2013 bis 2016 (Basisjahr 2010 = 100)			
	L	PDF <sup>1)</sup>	MI 2
Messzahlen für Bauleistungspreise und Preisindizes für Bauwerke in Hessen im November 2016			
	L	PDF <sup>1)</sup>	MI 4

1) Mit Excel-Anhang

Statistische Berichte und sonstige Veröffentlichungen - Kostenfreie Download-Dateien  
L = Landesebene – K = Kreisebene – G = Gemeindeebene

Titel	Regionale Gliederung	Dateiformat	Kennziffer
<b>Q Umwelt</b>			
Wasser- und Abwasserentgelte in Hessen 2014 bis 2016	G	PDF <sup>1)</sup>	Q I 6
Investitionen für Umweltschutz im Verarbeitenden Gewerbe in Hessen im Jahr 2014	K	PDF <sup>1)</sup>	Q III 1
Klimawirksame Stoffe 2015	L	PDF <sup>1)</sup>	Q IV 1
<b>Z Querschnittsveröffentlichungen</b>			
Hessische Gemeindestatistik 2016; Ausgewählte Strukturdaten aus Bevölkerung und Wirtschaft 2015	G	PDF <sup>1)</sup>	
<b>Kostenfreie Verzeichnisse</b>			
Schlüsselverzeichnis der Verwaltungsbezirke, Städte und Gemeinden in Hessen 2016 (Gebietsstand 01.01.1981)	G	PDF <sup>1)</sup>	Verz-2

1) Mit Excel-Anhang

## Verzeichnis der Krankenhäuser, Vorsorge- und Rehabilitationseinrichtungen in Deutschland



### Adressverzeichnis

■ Die vorliegende Publikation enthält Informationen zu allen Einrichtungen,

die zur stationären Versorgung beigetragen haben (sofern ihre Zustimmung für eine Veröffentlichung vorlag); z. B. Anschriften (einschl. Telefon- und Faxnummern), Träger der Einrichtung, Fachabteilungen, jahresdurchschnittlich aufgestellte Betten. Der Nachweis erfolgt getrennt nach Krankenhäusern und nach Vorsorge- oder Rehabilitationseinrichtungen. Die Einrichtungen sind nach Ländern, Regierungsbezirken und Landkreisen sortiert.

Das Verzeichnis wird als ZIP-Datei über E-Mail versandt.



STATISTIK HESSEN



Bildnachweis:  
Symbol of medicine  
© 4designersart -  
Fotolia.com

W-158-17

Aktuelle Angaben über die neuesten Ausgaben sowie Preise finden Sie im Internet oder nehmen Sie direkt Kontakt mit uns auf.

Bestellen Sie bei Ihrer Buchhandlung oder direkt bei uns.  
Hessisches Statistisches Landesamt, 65175 Wiesbaden  
Telefon: 0611 3802-950 · Fax: 0611 3802-992  
E-Mail: [vertrieb@statistik.hessen.de](mailto:vertrieb@statistik.hessen.de)  
Internet: <https://statistik.hessen.de>



© magele-picture - Fotolia.com

... auf das nächste Heft 2/2017 –  
Themen: Bildung, Kultur und Soziales

Das Heft 2/2017 können Sie voraussichtlich ab Juli 2017 auf unserer Homepage: <https://statistik.hessen.de> abrufen oder über unseren Vertrieb unter: [vertrieb@statistik.hessen.de](mailto:vertrieb@statistik.hessen.de) bestellen.



© Robert - Fotolia.com

## Musikfestivals und -festspiele 2015

von Dr. Joachim Lutz

In Deutschland gibt es derzeit kein umfassendes und einheitliches Datenangebot für den Kulturbereich. Im Rahmen des Projekts „Aufbau einer bundeseinheitlichen Kulturstatistik“ hat das Hessische Statistische Landesamt als Pate für den Kulturbereich eine § 7-Erhebung zu „Musikfestivals und -festspielen in Deutschland“ konzipiert und durchgeführt. Die Ergebnisse dieser Erhebung, insbesondere zu wesentlichen Merkmalen wie der Besucherzahl, den Musikstilen, Veranstaltungsorten sowie Eintrittspreisen, werden in diesem Artikel dargestellt.



© denisismagilov - Fotolia.com

## Die Kultur- und Kreativwirtschaft in Hessen

von Dr. Bianka Dettmer

Werbung, Architektur, Fotografie oder auch Film – all diese Bereiche charakterisieren unsere Kultur- und Kreativwirtschaft. Ein breites Spektrum an kulturellen Angeboten vor Ort gilt als wichtiger Standortfaktor, der die Attraktivität und damit den Zuzug von qualifizierten Arbeitskräften fördert. Regionalen Entscheidern bzw. Förderern fehlen aber oft Informationen für eine weitergehende Analyse der Situation vor Ort, da die meisten Daten nur auf Landesebene publiziert werden. Im Auftrag der Wirtschaftsförderung Frankfurt GmbH hat das Hessische Statistische Landesamt sich daher mit dieser Problematik beschäftigt und eine Auswertung zur Kultur- und Kreativwirtschaft in Frankfurt am Main sowie weiteren Vergleichsstädten erstellt. Die Ergebnisse dieser Analyse werden in einem kurzen Beitrag dargestellt. Außerdem wird auf die Besonderheiten der Auswertung auf kleinräumiger Ebene eingegangen.

## Außerdem im Heft:



© kasto - Fotolia.com

### **Bildungsstand der hessischen Bevölkerung – Auswertung des Mikrozensus 2015**

---

von *Yvonne Lieber*

Welchen höchsten allgemeinen Schulabschluss hat die hessische Bevölkerung? Welche beruflichen Bildungsabschlüsse besitzen die Hessinnen und Hessen? Und gibt es Unterschiede zwischen Männern und Frauen hinsichtlich des Bildungsstands? Der Artikel gibt mithilfe von Datenanalysen aus dem Mikrozensus Antworten auf all diese Fragen. Darüber hinaus wird der Bildungsstand der hessischen Bevölkerung anhand von weiteren soziodemografischen Merkmalen beleuchtet.



© Stockfotos-MG - Fotolia.com

### **Wohngeld – Rechtliche Determinanten und Daten zur Entwicklung in Hessen**

---

von *Petra Gerisch*

Das Wohngeld ist ein staatlicher Zuschuss zu den Aufwendungen für Wohnraum. Es dient gemäß § 1 des Wohngeldgesetzes (WoGG) der wirtschaftlichen Sicherung zum angemessenen und familiengerechten Wohnen. Politische Reformen und damit verbundene Gesetzesänderungen machen die Darstellung des Wohngeldes in längeren Zeitreihen zu einem komplexen Thema. Der Beitrag beschäftigt sich mit der Bedeutung des Begriffs „Wohngeld“ und stellt chronologisch dar, welche Gesetzesänderungen es hierzu seit dem Zweiten Weltkrieg gab. Zudem werden die Wohngeldstatistik und das Problem der Zeitreihenanalyse erläutert. Im Anschluss werden Auswertungen zur Entwicklung des Wohngeldes präsentiert.



Alle Ausgaben von Staat und Wirtschaft in Hessen finden sich auch als PDF-Datei zum kostenlosen Download auf unserer Homepage unter <https://statistik.hessen.de/publikationen/staat-und-wirtschaft>