

# Forschungsdatenbank Nichtwohngebäude (ENOB:dataNWG)

Forschungsprojekt im Förderbereich  
**Energieoptimierte Gebäude und Quartiere im  
6. Energieforschungsprogramm der Bundesregierung**  
Förderkennzeichen 03ET1315

## Motivation der methodischen Vorgehensweise

Dr. Holger Cischinsky, Michael Hörner  
Institut Wohnen und Umwelt  
04. Mai 2020

**KfW**  
Bank für Sozialwirtschaft

Zur Energieanalyse von  
Gebäuden wird das Werkzeug  
VSA 2.0 verwendet. VSA 2.0  
wird vom IWU mit Mitteln der  
KfW Bankengruppe erstellt.

Gefördert durch:

Bundesministerium  
für Wirtschaft  
und Energie

aufgrund eines Beschlusses  
des Deutschen Bundestages

Quelle: Adobe Stock

## Warum erfolgen alle Stichprobenziehungen zufallsgesteuert?

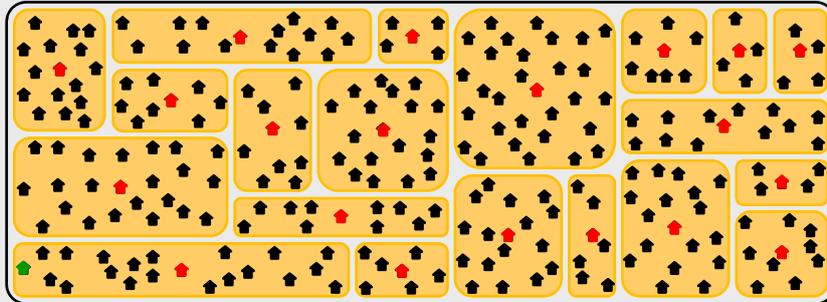
Gründe:

- Vermeidung von Selektionseffekten ⇒ **Objektivität**
- Nutzung des Instrumentariums der Wahrscheinlichkeitstheorie, um
  - Erwartungstreue und damit **Repräsentativität** sicherzustellen
  - die stichprobenbedingte Ergebnisunsicherheit angeben zu können (→ **Standardfehler**)⇒ wahrscheinlichkeitstheoretische Fundierung der Stichprobentheorie

## Können im Nachgang weitere Nichtwohngebäude in die Forschungsdatenbank eingefügt werden?



Ja, aber....



- ♦ : zufalls-gesteuert ausgewähltes NWG
- ◆ : nicht-zufalls-gesteuert ausgewähltes NWG

- Konsequenzen:
- Der Hochrechnungsfaktor (HRF) des zufalls-gesteuert ausgewählten „Repräsentativ-NWG“ ♦ müsste um 1 reduziert werden
  - Der HRF des nicht-zufalls-gesteuert ausgewählte NWG ◆ muss auf 1 festgelegt werden – mit der Folge, dass dieses NWG bei der Auswertung (fast) keine Bedeutung hat

3

## Warum wurde eine Stichprobe von Hausumringen und nicht gleich von NWG gezogen?



### Ideale Auswahlgrundlage:

- Liste aller Nichtwohngebäude in Deutschland (und nur dieser) mitsamt:
- Objektschrift
  - kompetenten Ansprechpartnern und
  - fundamentalen Gebäudeeigenschaften

Objektschrift	Ansprechpartner	Gebäudebeschreibung
Musterstraße 4 (Musterstadt) 12345 Musterhausen	Max Mustermann (Verwalter) Tel.: 0123456789 12345 Musterhausen	Bürogebäude 4 Vollgeschosse Baujahr 2000
Musterweg 17 54321 Musterdorf	Max Muster AG Frau Muster 99999 Musterstadt	Werkstatt 1 Vollgeschoss Baujahr 1960
Musterpfad 6 99999 Musterstadt	Max Mustermann Herr Mustermann 99999 Musterstadt	Autohaus 1 Vollgeschoss Baujahr 1995
(-)	(-)	(-)

- Abhilfe: Datenbank (amtlicher) Hausumringe (**HU-DE**) (= 2D-Gebäudegrundriss (Teilauszug aus der amtlichen Liegenschaftskarte) als Auswahlgrundlage für die Stichprobenziehung



4

## Welche Nachteile haben Hausumringe als Erhebungseinheiten?



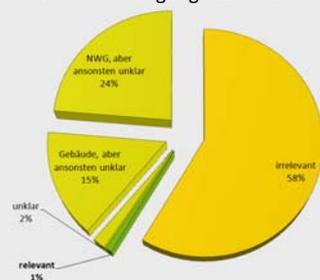
- **Vorhandensein irrelevanter Hausumringe**, ohne diese mit hinreichender Sicherheit identifizieren zu können.
- Häufig **keine Äquivalenz** zwischen Hausumringen als Erhebungs- und Gebäuden als Untersuchungseinheiten.
- **Keine (verlässlichen) Eigentümerinformationen**.

5

## Was sind Relevanzwahrscheinlichkeiten, wofür benötigen wir sie und wie werden sie bestimmt?



Problem: Nur ein kleiner Teil der knapp 49 Mio. Hausumringe der Auswahlgrundlage ist relevant in dem Sinne, dass die Hausumringe zu EnEV-relevanten NWG gehören. Die **Relevanz vieler Hausumringe** in Bezug auf die NWG-Eigenschaft und die EnEV-Relevanz ist **häufig** jedoch **unsicher** oder sogar gänzlich **unbekannt**.



vom Projektteam auf Basis verfügbarer Nutzungsattribute vorgenommene Relevanzeinstufung der Hausumringe

⇒ „einfache“ Zufallsstichprobe von Hausumringen liefert erwartungsgemäß zu viel „Ausschuss“

Lösung: Vornahme einer **Relevanzeinstufung** und Ziehung der Hausumring-Stichprobe so, dass relevante Hausumringe eher in die Stichprobe gelangen als irrelevante

⇒ **Ziehung einer „disproportional geschichteten Stichprobe“**

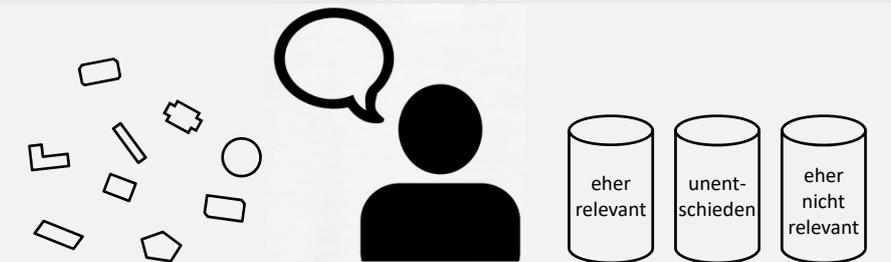
6

## Was sind Relevanzwahrscheinlichkeiten, wofür benötigen wir sie und wie werden sie bestimmt?

Mögliche Verfahren der Relevanzeinstufung:

Alternative 1:

Heuristische Relevanzeinstufung



**Vorteil:** Gewährleistung von i.d.R. **plausiblen** und **nachvollziehbaren Einstufungen**

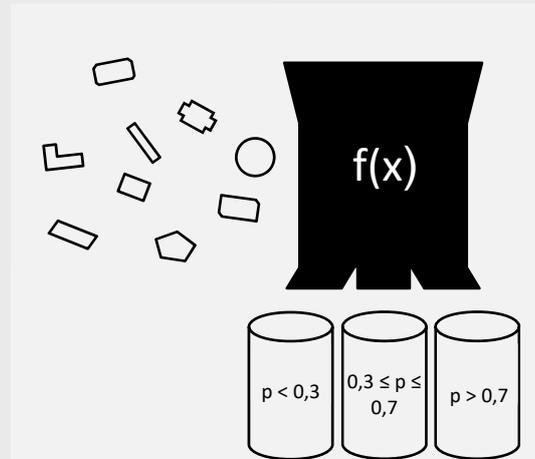
**Nachteile:** - **extrem aufwändig** (Einzelfallbehandlung von knapp 20 Mio. (!!!) Hausumringen)  
- **nur ordinale Einstufung** in max. fünf Gruppen möglich  
- subjektive Einstufung  $\Rightarrow$  **fehlende Objektivität**

7

## Was sind Relevanzwahrscheinlichkeiten, wofür benötigen wir sie und wie werden sie bestimmt?

Alternative 2 (hier gewählte Alternative):

Algorithmische Relevanzeinstufung der Hausumringe über **binär-logistischen Regressionsansatz**



**Vorteile:** - **relativ geringer Aufwand**  
- Zuordnung metrischer **Relevanzwahrscheinlichkeiten**  
- **objektive** Einstufung

**Nachteil:** **teilweise unplausible Einstufungen**

**Achtung:** Relevanzwahrscheinlichkeit  $\neq$  Ziehungswahrscheinlichkeit

**aber:** positive Korrelation zwischen Relevanz- und Ziehungswahrscheinlichkeit

8

## Warum befinden sich auch als irrelevant eingestufte Hausumringe in begrenzter Zahl in der Stichprobe?



**Problem:** Die **Einstufung als irrelevanter Hausumring** wurde vom Projektteam auf Basis verfügbarer Nutzungsattribute vorgenommen und ist mit einer **(unbekannten) Restunsicherheit** verbunden.

**Abhilfe:** Als irrelevant eingestufte Hausumringe werden **in begrenzter Zahl in die Stichprobe** aufgenommen, wodurch dem Postulat der Stichprobentheorie genügt wird, grundsätzlich jedem NWG in Deutschland eine Chance auf Aufnahme in die Stichprobe einzuräumen (**positive Auswahlwahrscheinlichkeiten**).

9

## Warum ist eine Screeningphase erforderlich?



**Hauptnutzen:**

- Einholen von **Informationen zu potentiellen Gebäudeansprechpartnern**
- **Eindeutige Relevanzfeststellung** in Bezug auf **alle** Stichproben-Hausumringe
- Abklärung des konkreten **Ziehungsgeflechts** zwischen Hausumring und NWG

**Zusatznutzen:**

- soweit möglich **Validierung der attribuierten Hausumring-Merkmale**
- **Erhebung weiterer Gebäudemerkmale** vor Ort

⇒ bereits das Screening liefert auf sehr breiter Stichprobenbasis **wesentliche Strukturmerkmale zum NWG-Bestand**

⇒ Screening als **groß angelegter Praxistest** zur (Weiter-) Entwicklung von Algorithmen zum geoinformatrischen Identifizieren von NWG auf Basis von Hausumringen

⇒ Identifizierung und **Korrektur etwaiger systematischer Gebäudeausfälle** bei der nachgelagerten Breiten- und Tiefenerhebung möglich

10

## Wie überführen wir hausumringspezifische in gebäudespezifische Ziehungswahrscheinlichkeiten?

Zwischen Hausumring- und Gebäude besteht nicht notwendigerweise eine 1:1-Beziehung!!!



Lösung: Entwicklung einer geeigneten, für alle denkbaren Hausumring-Gebäude-Beziehungsgeflechte anwendbare **Umrechnungsformel**:

$$\pi_{ij} = 1 - \prod_{h=1}^H \frac{(N_i(h) - M_{ij}(h))}{n_i(h)}$$

11

## Bedingen sich die drei Erhebungsphasen in methodischer Hinsicht gegenseitig?

Nicht gegenseitig, allerdings gibt es eine **Abhängigkeitskette**: Das Screening ist eine unabdingbare methodische Voraussetzung für die Breiten- und Tiefenerhebung, aber nicht umgekehrt. Und die Tiefenerhebung bedingt eine vorausgehende Breitenerhebung, wohingegen die Breitenerhebung ohne die Tiefenerhebung auskommt.

Insofern stellte das **Screening** auch eine (überwundene) **Sollbruchstelle** für die nachgelagerte Breitenerhebung und diese wiederum für die sich anschließende Tiefenerhebung dar.

12

## Warum gibt es Erhebungsbezirke und wie groß sind diese?

Die **Relevanzfeststellung** während des Screenings, das Einholen von **Informationen zu potentiellen Gebäudeansprechpartnern** und die Abklärung des konkreten **Ziehungsgeflechts** zwischen Hausumring und NWG bedingen eine Inaugenscheinnahme der in die Bruttostichprobe aufgenommenen Hausumringe vor Ort. Die Vorauswahl von Erhebungsbezirken **reduziert die damit einhergehenden Wegekosten**.

Die Erhebungsbezirke wurden vom IÖR unter Beachtung statistisch-methodischer Vorgaben (Reduzierung der späteren Ergebnisunschärfe) zugeschnitten, insbesondere dergestalt, dass die Erhebungsbezirke **erwartungsgemäß ähnlich große NWG-Bestände** haben. Dazu griff das IÖR auf die hausumringspezifischen Relevanzwahrscheinlichkeiten zurück.



Beispiel: Erhebungsbezirke in Schleswig-Holstein

13

## Wie erklärt sich die Festlegung auf 500 Erhebungsbezirke und wie wurden diese ausgewählt?

Festlegung auf **500 Stichproben-Erhebungsbezirke** ist ein Kompromiss aus zwei entgegengesetzten Anforderungen:

- 500 Erhebungsbezirke sind ausreichend viele, um sog. **Klumpeneffekte**, die sich negativ auf die Ergebnisschärfe auswirken, **einzudämmen** und um eine **befriedigende deutschlandweite Flächenabdeckung** sicherzustellen
- 500 Erhebungsbezirke sind **organisatorisch noch beherrschbar**.

Die Auswahl der Erhebungsbezirke erfolgte **zufallsgesteuert** und **geschichtet nach Landesteilen (Bundesland und Raumordnungsregion)**. Die Schichtung gewährleistet ex ante, dass aus allen Teilen Deutschlands am Ende Erhebungsdaten vorliegen.

14

## Wie gehen wir mit Gebäudeausfällen (Unit Nonresponse) während der Breiten- bzw. Tiefenerhebung um?

**Unit Nonresponse** in der Breiten- und Tiefenerhebung ist aufgrund der freiwilligen Teilnahmeoption **unvermeidbar**.

Da auch für „ausgefallene“ Gebäude dank vorausgegangenem Screening bzw. dank vorausgegangener Breitenenerhebung ausreichend viele Merkmale vorliegen, ist es möglich, **systematische und damit verzerrende Ausfälle zu identifizieren** und mittels entsprechender statistischer Gewichtungsverfahren **zu kompensieren**, um so die Repräsentativität der Gesamterhebung zu bewahren.

15

## Welche Möglichkeiten gibt es, die Forschungsdatenbank mit anderen Datenquellen zu verschneiden?

Grundsätzlich: Zur **Verschneidung** verschiedener Datenquellen benötigt man **Verknüpfungsvariablen**, die in **möglichst identischer Abgrenzung** in den zu verschneidenden Datenquellen vorhanden sind.

**Unproblematisch** „anreicherbar“ sind im vorliegenden Fall **regionalstatistische Merkmale**, sofern diese auf folgenden Raumabgrenzungen basieren:

- PLZ
- Stadt/Gemeinde
- Raumordnungsregion
- Bundesland

Eine **gebäude- bzw. adressscharfe Verschneidung** kommt im vorliegenden Fall nur in **Ausnahmefällen** in Betracht und scheitert häufig aus folgenden Gründen:

- fehlende Datenverfügbarkeit
- unterschiedliche Gebäudeabgrenzung
- Datenschutzbedenken

**aber:** Eine gebäudescharfe Anreicherung um Wirtschaftszeige war möglich

16

## Warum ist eine Nacherhebung durch das in das Projekt eingebundene Befragungsinstitut (IFAK) erforderlich?



Wie zu erwarten, kam nicht für alle an die Breiterehebung übergebenen Gebäudeadressen („breiterehebungsrelevante Gebäude“) ein Interview zustande. Dafür zeichneten grundsätzlich **zwei verschiedene Arten von Ausfallgründen** verantwortlich:

- **Irrelevanz**, d.h. fehlende Zielgruppenrelevanz
- **Nichterreichbarkeit bzw. Nonresponse** (insbes. Teilnahmeverweigerung)

Um sog. **Merkmalssummenschätzungen** zu ermöglichen (insbes. Hochrechnung der Anzahl EnEV-relevanter NWG aus der Stichprobe heraus), ist es methodisch **zwingend erforderlich, die Frage der Relevanz erst im Interview selbst zu klären**.

**aber:** Diese methodische Notwendigkeit wurde nicht beachtet.

⇒ **Erfordernis der Nachrecherche aller als irrelevant abgelegter Gebäude** dergestalt, dass versucht wird, ein Interview zu generieren und dort final die Frage der Relevanz zu klären.

Abgesehen davon waren die **Irrelevanz-Feststellungen bisweilen fragwürdig**.

17



**Vielen Dank für Ihre Aufmerksamkeit !!!**

18