

ENOB: dataNWG

Forschungsdatenbank Nichtwohngebäude

Förderkennzeichen 03 ET1315 A

**Institut Wohnen
und Umwelt GmbH**

Forschungseinrichtung
des Landes Hessen und
der Stadt Darmstadt

Rheinstraße 65
64295 Darmstadt
Germany

Tel: +49 (0)6151 / 2904-0
Fax: +49 (0)6151 / 2904-97

info@iwu.de
www.iwu.de



Leibniz-Institut
für ökologische
Raumentwicklung



**BERGISCHE
UNIVERSITÄT
WUPPERTAL**

Anleitung Fernrechnen (R-Skript)

Gefördert vom

Bundesministerium für Wirtschaft und Energie

im Förderbereich

Energieoptimierte Gebäude und Quartiere

der Fördermaßnahme *Anwendungsorientierte nichtnukleare FuE* des 6. Energie-
forschungsprogramms der Bundesregierung



**Bundesministerium
für Wirtschaft
und Energie**

8. August 2022

Impressum

Projekt	Forschungsdatenbank Nichtwohngebäude. Repräsentative Primärdatenerhebung zur statistisch validen Erfassung und Auswertung der Struktur und der energetischen Qualität des Nichtwohngebäudebestands in Deutschland.
Kurztitel	ENOB:dataNWG
Teilprojekt	Repräsentative Stichprobenerhebung und Auswertung typologischer, struktureller und energetischer Merkmale
Gefördert mit Mitteln von	Bundesministerium für Wirtschaft und Energie (BMWi) im Förderschwerpunkt Energieoptimierte Gebäude und Quartiere des 6. Energieforschungsprogramms der Bundesregierung
Förderkennzeichen	03 ET1315 A
Fördermittelnehmer	Institut Wohnen und Umwelt Rheinstraße 65 64295 Darmstadt Tel. +49 (0) 6151 / 2904 -0
Projektpartner	<ul style="list-style-type: none">• Leibniz-Institut für ökologische Raumentwicklung (IÖR)• Bergische Universität Wuppertal, Fachgebiet Ökonomie des Planens und Bauens (BUW-ÖPB)
Auftragnehmer	
Bericht	Anleitung Fernrechnen (R-Skript)
Verfasser	Dr. Holger Cischinsky, Hendrik Schäfer
Mitarbeit	Michael Hörner, Galina Nuss
Datum	8. August 2022

1 Einführung

Im Rahmen des Forschungsprojekts „Forschungsdatenbank Nichtwohngebäude“ (ENOB:dataNWG; gefördert durch das Bundesministerium für Wirtschaft und Energie; Förderkennzeichen: 03ET1315) wurde im Rahmen einer sog. Breitenerhebung eine deutschlandweite Repräsentativbefragung der Eigentümerinnen und Eigentümer von Nichtwohngebäuden durchgeführt. Die in dieser Erhebung gewonnenen Daten sollen auf datenschutzkonforme Weise auch Dritten für eigene Auswertungszwecke zugänglich gemacht werden, und zwar auf dem Wege des sog. Fernrechnens (auch kontrollierte Datenfernverarbeitung genannt). Eine der Möglichkeiten des Fernrechnens besteht darin, dass Dritte – im Folgenden „Anwender“ genannt – mithilfe der Statistik-Software R (<http://www.r-project.org>) nach vorgegebenem Muster Auswertungsdateien erstellen, die das IWU auf die Forschungsdatenbank anwendet, um die Auswertungsergebnisse dem Anwender im Anschluss daran zukommen zu lassen.

Die Rahmenbedingungen hierfür werden in einem mit dem IWU abzuschließenden Vertrag festgelegt. Der Vertrag steht unter https://www.datanwg.de/fileadmin/user/iwu/210429_datanwg_Vertrag_Fernrechnen.pdf zum Download zur Verfügung und ist vom Anwender in zweifacher Ausfertigung auszufüllen, zu unterschreiben und dem IWU auf postalischem Wege zuzuschicken. Nach der Gegenzeichnung des Vertrags durch das IWU ist der Vertrag geschlossen.

Unter https://www.datanwg.de/fileadmin/user/iwu/210504_Vorlage_Auswertung.txt wird eine Vorlage für eine Auswertungsdatei zum Download zur Verfügung gestellt. Diese hat das Format einer Textdatei (.txt). Der Anwender erstellt daraus und unter Beachtung der in der vorliegenden Anleitung dokumentierten Syntax seine Auswertungsdatei und sendet sie per E-Mail an das IWU (fernrechnen@iwu.de). Die Datei muss weiterhin das Format einer Textdatei haben. Andere Formate (z.B. Word, Excel) sind nicht zulässig.

Das IWU verwendet die Auswertungsdatei als Eingabedatei für eine mit der Statistik-Software R durchgeführte Auswertung der Forschungsdatenbank. Anschließend wird dem Anwender eine Ergebnisdatei zurückgesendet. Diese hat ebenfalls das Format einer Textdatei.

Die Methodik des Forschungsprojekts und detaillierte Informationen zu seiner Durchführung sowie zahlreiche Auswertungsergebnisse finden sich auf der Projektwebsite <http://www.datanwg.de>. Demgegenüber liefert dieser Bericht Informationen, die für eigenständige Auswertungen relevant sind, und stellt daher für den Anwender eine wichtige Arbeitsgrundlage zur Erstellung seiner Auswertungsroutinen dar. Fundierte Kenntnisse und Erfahrungen bei der statistischen Auswertung von Befragungsdaten werden dabei vorausgesetzt. Kenntnisse in R sind hilfreich.

2 Zur Verfügung stehende Auswertungsmerkmale

Die Forschungsdatenbank enthält 5.630 Datensätze von funktional relevanten Nichtwohngebäuden, zu denen aus der Breiterehebung Daten vorliegen. Bei den erhobenen Daten handelt es sich zum einen um (aufbereitete) Angaben der Befragungspersonen zu diversen Fragebogenfragen, zum anderen um nachträglich zugespielte Gebäudeeigenschaften. Die entsprechenden Variablen (auch „Merkmale“ genannt), auf die im Rahmen des Fernrechnens zurückgegriffen werden kann, sind im Dokument „Variablendokumentation Fernrechnen“¹ in systematischer Form aufgelistet.

Zu beachten ist, dass nahezu alle Variablen fehlende Werte (sog. Missings) enthalten. Das Verfassen von Auswertungsdateien setzt voraus, dass der Anwender Kenntnisse in der Behandlung von Missings hat. Darüber hinaus wird unterstellt, dass der Anwender mit Kap. 3.3.3. des Berichts „E.1.4.5 – Stichprobe: Modellierung und Ziehung“² vertraut ist, wo die Vorgehensweise zur Schätzung sog. Merkmalssummen bei Anwesenheit von fehlenden Werten erläutert wird.

¹ Vgl. https://www.datanwg.de/fileadmin/user/iwu/210422_IWU_Variablendokumentation.pdf

² Vgl. https://www.datanwg.de/fileadmin/user/iwu/210210_IWU_E1-4-5_Stichprobe_Modellierung_und_Ziehung.pdf

3 Aufbau der Auswertungsdatei und Durchführung der Auswertungen

Zur Erstellung der Auswertungsdatei ist die Syntax der Statistik-Software R (<http://www.r-project.org>) zu verwenden. Auswertungsbefehle werden zeilenweise eingegeben. Zu beachten ist, dass R zwischen Groß- und Kleinschreibung unterscheidet. Kommentarzeilen, die beim Programmablauf ignoriert werden sollen, müssen mit dem Rautezeichen (#) eingeleitet werden.

Die Vorlage für die Auswertungsdatei steht als txt-Datei unter https://www.dat-anwg.de/fileadmin/user/iwu/210504_Vorlage_Auswertung.txt zum Download zur Verfügung. Zudem gibt es eine Beispieldatei unter https://www.dat-anwg.de/fileadmin/user/iwu/210504_Beiispiel_Auswertung.txt. Beide txt-Dateien dienen als Grundlage für die nachfolgende Erläuterung der Vorgehensweise.

Die Auswertungsdatei ist in mehrere Abschnitte unterteilt:

Titel

Die Auswertungsdatei trägt folgenden Titel, der vom Anwender nicht verändert werden darf:

```
#####
#
# Auswertungsdatei für die Durchführung eines Auswertungsgangs mit der Forschungsdatenbank
# Nichtwohngebäude (Projekt "Forschungsdatenbank Nichtwohngebäude (ENOB:dataNWG)",
# gefördert durch das Bundesministerium für Wirtschaft und Energie;Förderkennzeichen:
# 03ET1315)
#
#####
```

Header

Im auf den Titel folgenden Header macht der Anwender die geforderten Angaben:

```
#####
#
#Bitte folgende Angaben vervollständigen:
#
#Dateiname des Programmcodes: <Name>.txt
#erstellt: <Datum>
#von: <Name>
#E-Mail: <E-Mail-Adresse>
#Tel.: <Telefonnummer>
#
#####
```

Kopfzeilen – Schritte 1 bis 3

Die Kopfzeilen dürfen vom Anwender nicht verändert werden. Am Ende dieses Abschnitts steht die Gesamtheit aller 5.630 Nichtwohngebäude in der Variablen *gbaeude* zu Auswertungszwecken zur Weiterverarbeitung zur Verfügung.

```
#####
# 1. Laden des Headers
#       Lädt Gebäudedaten, stichprobenrelevante Daten
#       inkl. der Schätzfunktionen in den Arbeitsspeicher
#####
source("//Dateiserver/energie/1075_ENOBdat-
aNWG/02_work_IWU/AP1_Konzeptphase/AP1.4_Erhebungsmethodik/AP1.4.1_Datenbank/Umsetzung_Auswertung/05_Schaetzer_BRE/20210509_header_bre.r")
# Verzeichnis für Exportdateien
pfad <- "C:/Users/Nutzer/Desktop/"

#####
# 2. Festlegen des Arbeitsverzeichnisses
#####
setwd("//Dateiserver/energie/1075_ENOBdataNWG/02_work_IWU/AP4_Auswertung/")
getwd()

#####
# 3. Aufbereiten von dichotomen Merkmalen und
#   Aufbereiten der Ausprägungen von uk_geb
#   NICHT ÄNDERN!
#####
gebaeude <- dichot_merkm_aufbereiten()
```

Variablen und Teilmengen definieren – Schritte 4 und 5

Der Anwender hat die Möglichkeit, auf Basis der in der „Variablendokumentation Fernrechnen“ bereitgestellten Variablen eigene Variablen zu generieren. Darüber hinaus legt der Anwender fest, für welche Teilmenge(n) des Gebäudebestandes er Auswertungen vornehmen möchte.

```
#####
# 4. Ggf. neue Merkmale definieren
# Hinweis: Nur fehlende Werte dürfen negativ codiert sein.
#####

#####
# 5. Definition von Teilmengen
#####
```

Zur Bildung eigener Variablen und zur Definition der Teilmengen stehen die folgenden logischen Ausdrücke zur Verfügung, die den Wert TRUE (wahr) oder FALSE (falsch) annehmen:

<u>Symbol</u>	<u>Bedeutung</u>
==	gleich (Das doppelte Gleichheitszeichen ist zu verwenden!)
!=	ungleich

>	größer als
>=	größer gleich
<	kleiner als
<=	kleiner gleich
!	nicht (Negation)
&	und
	oder (inklusive „oder“, d. h. nicht-ausschließende Disjunktion)

Bei komplizierteren Ausdrücken sind entsprechende Klammern zu setzen. Dabei werden runde Klammern verwendet: ().

Zu beachten ist ferner, dass Variablennamen weder mit einer Zahl beginnen noch – außer dem Unterstrich („_“) – keine Sonderzeichen enthalten dürfen.

Im Folgenden werden zwei Beispiele zur Bildung eigener Variablen gegeben. Die beiden Beispiele sind der Beispieldatei entnommen, die unter https://www.dat-anwg.de/fileadmin/user/iwu/210504_Beispiel_Auswertung.txt zum Download zur Verfügung steht.

Im ersten Beispiel soll die neue Variable *hk_geb_agg* gebildet werden, bei der gegenüber der in der „Variablendokumentation Fernrechnen“ enthaltenen Variable *hk_geb* (Hauptkategorie des Gebäudes) Merkmalsausprägungen („Codes“) zusammengelegt werden. Mittels der Zuweisung

```
# Aggregierte Hauptgebäudefunktion hk_geb_agg
gebaeude$hk_geb_agg <- ((gebaeude$hk_geb < 0) * -1
+ ((gebaeude$hk_geb >= 1 & gebaeude$hk_geb <= 7) | gebaeude$hk_geb == 9) * 1
+ (gebaeude$hk_geb == 8 | gebaeude$hk_geb == 10 | gebaeude$hk_geb == 11) * 2)
```

wird dies erreicht. Die neue Variable *hk_geb_agg* hat demnach im Gegensatz zu *hk_geb* nur noch die drei Ausprägungen -1, 1 und 2, die jeweils rechts von „*“ stehen. Konkret bekommen die negativen Ausprägungen von *hk_geb* den Wert -1 und die Ausprägungen von 1 bis 7 sowie 9 den Wert 1 zugewiesen. Die restlichen Ausprägungen 8, 10 und 11 werden mit dem Wert 2 zusammengefasst. Indem vor *hk_geb* und *hk_geb_agg* jeweils *gebaeude\$* geschrieben wird, wird sichergestellt, dass alle 5.630 Gebäudedatensätze angesprochen werden.

Im nächsten Beispiel wird die metrische Variable *ober_bgf* (oberirdische Bruttogrundfläche (BGF) des Gebäudes) klassifiziert, d.h. in eine kategoriale Variable überführt, die mit *ober_bgf_kl* bezeichnet wird und sieben gültige Merkmalsausprägungen (von 1 bis 7) aufweist. Wiederum bekommen die negativen Ausprägungen den Wert -1 zugewiesen.

```
# Klassifizierte oberirdische BGF ober_bgf_kl
gebaeude$ober_bgf_kl <- ((gebaeude$ober_bgf < 0) * -1
+ (gebaeude$ober_bgf >= 0 & gebaeude$ober_bgf < 100) * 1
+ (gebaeude$ober_bgf >= 100 & gebaeude$ober_bgf < 1000) * 2
+ (gebaeude$ober_bgf >= 1000 & gebaeude$ober_bgf < 5000) * 3
+ (gebaeude$ober_bgf >= 5000 & gebaeude$ober_bgf < 10000) * 4)
```

```
+ (gebaeude$ober_bgf >=10000 & gebaeude$ober_bgf < 25000) * 5
+ (gebaeude$ober_bgf >=25000 & gebaeude$ober_bgf < 50000) * 6
+ (gebaeude$ober_bgf >=50000 ) * 7)
```

Die Festlegung von Teilmengen ist nicht erforderlich, wenn der Anwender stets auf alle 5.630 Gebäudedatensätze zurückgreifen und daher nur Auswertungen für die Gesamtheit aller funktional relevanter Nichtwohngebäude durchführen möchte. Andernfalls sind je nach Auswertungsinteresse Teilmengen zu definieren. Nach der Zuweisung des Variablennamens steht dabei der Befehl *subset*. In der sich anschließenden Klammer muss zunächst die Obermenge spezifiziert werden. Danach folgt der logische Ausdruck, durch den die betreffende Teilmenge festgelegt wird. Was die Obermenge angeht, muss bei der als erstes festgelegten Teilmenge zwingend *gebaeude* stehen, wodurch die Obermenge aller für Auswertungszwecke zur Verfügung stehender 5.630 Gebäudedatensätze angesprochen wird. Bei danach festgelegten Teilmengen kann anstelle von *gebaeude* auch eine andere, zuvor definierte Teilmenge als Obermenge herangezogen werden. Das nachfolgende, aus der Beispieldatei entnommene Beispiel verdeutlicht die Vorgehensweise.

```
grundlage1 <- subset(gebaeude,enev_rel_nwg_be == 1)
grundlage2 <- subset(grundlage1,energietraeger >= 0 & hk_geb_agg >= 0)
grundlage3 <- subset(gebaeude,enev_rel_nwg_be == 1 &
  uk_geb %in% c(100,101,102,103,104,105,202,205,306,307,308))
```

grundlage1 steht demnach für die Teilmenge aller GEG-relevanten Nichtwohngebäude (5.107 Gebäudedatensätze). *grundlage2* ist ihrerseits eine Teilmenge von *grundlage1*, konkret eine Teilmenge aller GEG-relevanten Nichtwohngebäude, für die die Ausprägungen der Variablen *energietraeger* und *hk_geb_agg* nichtnegativ sind. *grundlage3* schließlich bildet die Teilmenge aller GEG-relevanten Nichtwohngebäude, deren Unterkategorie eine der elf in Klammern aufgeführten Ausprägungen annimmt. Der Passus „%in% c()“ ermöglicht dabei das Selektieren mehrerer Merkmalsausprägungen und reduziert dadurch den Schreibaufwand.

Auswertungen – Schritt 6

Nachdem in den Schritten 4 und 5 ggf. eigene Variablen und Teilmengen erstellt worden sind, sind nun Auswertungsroutinen festzulegen. Hierfür gibt es die Funktionen *Mittelwert*, *Spaltenprozente*, *Zeilenprozente* und *Zellenprozente*. Die Signatur, d.h. der Name und die zu übergebenden Parameter sowie deren Reihenfolge, ist wie folgt definiert:

- Mittelwert(Tabelle, Pfad, Name des metrischen Merkmals, Name vertikales Merkmal, Name horizontales Merkmal)
- Spaltenprozente(Tabelle, Pfad, Name vertikales Merkmal, Name horizontales Merkmal)
- Zeilenprozente(Tabelle, Pfad, Name vertikales Merkmal, Name horizontales Merkmal)
- Zellenprozente(Tabelle, Pfad, Name vertikales Merkmal, Name horizontales Merkmal)

„Tabelle“ ist der Platzhalter für die Datengrundlage der Auswertung. Soll sich die Auswertung auf alle 5.630 Gebäudedatensätze stützen, ist „Tabelle“ durch

„*gebäude*“ zu ersetzen. Interessiert dagegen nur eine Teilmenge des Nichtwohngebäudebestands, ist die unter Schritt 5 definierte Teilmenge einzusetzen (z.B. *grundlage1*, falls es sich um die Teilmenge aller GEG-relevanten Nichtwohngebäude handelt (s.o.)).

Mit „Pfad“ ist die in der Kopfzeile (Schritt 1) angeführte Variable *pfad* gemeint. Diese ist unveränderlich.

Die Konzepte von vertikalen und horizontalen Merkmalen sind im IWU-Tabellenkonfigurator optisch veranschaulicht. Dieser steht unter https://www.datanwg.de/fileadmin/user/iwu/210503_Tabellenkonfigurator.xlsm zum Download zur Verfügung. Die Namen der vertikalen und horizontalen Merkmale sind als in Anführungszeichen gesetzter Text zu übergeben. Ein vertikales Merkmal muss immer gewählt werden. Auf das horizontale Merkmal kann dagegen verzichtet werden, wobei dann „*leer*“ einzutragen ist. Die Mittelwert-Funktion benötigt zudem eine weitere Variable, die das metrische Merkmal beschreibt, über welches der Mittelwert gebildet werden soll. Die Werte, die ein metrisches Merkmal annimmt, werden als Zahlen interpretiert und nicht als Kategorien. Ein Beispiel für ein metrisches Merkmal ist die im Beispiel oben angeführte oberirdische Bruttogrundfläche (BGF) (*ober_bgf*).

Die nachfolgenden, wiederum aus der Beispieldatei entnommenen Beispiele veranschaulichen die Vorgehensweise.

```
Zellenprozente(grundlage2, pfad, "energetraeger", "hk_geb_agg")
Spaltenprozente(grundlage2, pfad, "energetraeger", "hk_geb_agg")
Mittelwert(grundlage3, pfad, "ober_bgf", "ober_bgf_kl", "leer")
```

Das erste Beispiel erzeugt eine relative Häufigkeitstabelle, in der GEG-relevante Nichtwohngebäude nach dem Energieträger und der aggregierten Hauptkategorie des Gebäudes aufgeschlüsselt sind. Das Beispiel greift dabei auf die zuvor vom Anwender definierte Teilmenge (*grundlage2*) und die ebenfalls vom Anwender neu gebildete Variable *hk_geb_agg* zurück. „Zellenprozente“ bedeutet, dass sich alle berechneten Anteile der Tabelle zu 100% aufaddieren.

Das zweite Beispiel unterscheidet sich vom ersten nur dadurch, dass anstelle von „Zellenprozente“ jetzt die Funktion „Spaltenprozente“ verwendet wird. Dies hat zur Folge, dass sich die berechneten Anteile nun spaltenweise (d.h. für jede Ausprägung der Variable *hk_geb_agg*) zu 100% aufaddieren.

Im dritten Beispiel wird für vom Anwender definierte Klassen der oberirdischen Bruttogrundfläche (BGF) die mittlere oberirdische Bruttogrundfläche (BGF) berechnet, und zwar unter Einbeziehung aller GEG-relevanten Nichtwohngebäude mit vorgegebener Gebäudeunterkategorie (*grundlage3*).

Als Ergebnis erhält der Anwender je eine Tabelle mit den hochgerechneten Auswertungsergebnissen sowie eine korrespondierende Tabelle mit den zugehörigen Standardfehlern. Aus technischen Gründen bzw. aufgrund von Datenschutzbelangen kann der Fall eintreten, dass Ergebnisse nicht oder nur partiell zur Verfügung gestellt werden können. Um die Wahrscheinlichkeit dafür gering zu halten, wird empfohlen, die Definition von Teilmengen unter Schritt 5 nicht zu restriktiv

vorzunehmen bzw. die Zahl der Ausprägungen der vertikalen bzw. horizontalen Merkmale durch die Zusammenlegung von Ausprägungen zu begrenzen.

Zu beachten ist, dass mittels der zur Verfügung gestellten Funktionen keine sog. Merkmalssummen berechnet werden können, wozu auch die Schätzung von Anzahlen gehört (z.B. hochgerechnete Zahl GEG-relevanter Nichtwohngebäude mit einer bestimmten Eigenschaft). Ist der Anwender an der Berechnung solcher Merkmalssummen interessiert, ist wie in Kap. 3.3.3. des Berichts „E.1.4.5 – Stichprobe: Modellierung und Ziehung“ (Download-Link siehe Fußnote 2) beschrieben vorzugehen. Ausgangspunkt für solche Merkmalssummenschätzungen sind gleichwohl mittels der Auswertungsdatei berechnete Anteile oder Mittelwerte.