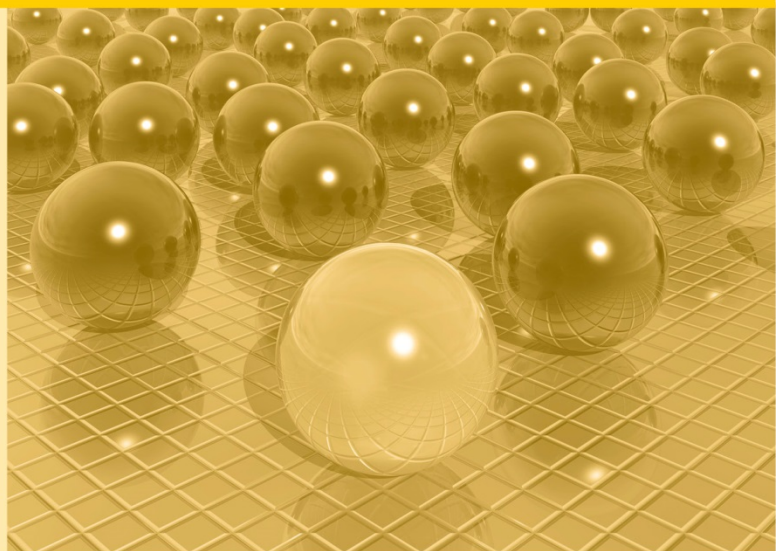


Metadatenreport



Teil II: Produktspezifische Informationen zur Nutzung der Einzeldaten des Verbraucherpreisindex 2018 (EVAS-Nummer: 61111) per Scientific Use File

DOI: 10.21242/61111.2018.00.00.3.1.0

Version 1

Impressum

Herausgeber: Statistische Ämter des Bundes und der Länder
Herstellung: Information und Technik Nordrhein-Westfalen
Telefon 0211 9449-01 • Telefax 0211 9449-8000
Internet: www.forschungsdatenzentrum.de
E-Mail: forschungsdatenzentrum@it.nrw.de

Fachliche Informationen

zu dieser Veröffentlichung:

Forschungsdatenzentrum der
Statistischen Ämter der Länder
– Standort Hessen –
Tel.: 0611 3802-822
Fax: 0611 3802-890
forschungsdatenzentrum@statistik.hessen.de

Informationen zum Datenangebot:

Statistisches Bundesamt
Forschungsdatenzentrum
Tel.: 0611 75-2420
Fax: 0611 72-3915
forschungsdatenzentrum@destatis.de

Forschungsdatenzentrum der
Statistischen Ämter der Länder
– Geschäftsstelle –
Tel.: 0211 9449-2883
Fax: 0211 9449-8087
forschungsdatenzentrum@it.nrw.de

Erscheinungsfolge: unregelmäßig
Erschienen im März 2022

Diese Publikation wird kostenlos als PDF-Datei zum Download unter www.forschungsdatenzentrum.de angeboten.

© Information und Technik Nordrhein-Westfalen, Düsseldorf, 2022
(im Auftrag der Herausbergemeinschaft)

Vervielfältigung und Verbreitung, nur auszugsweise, mit Quellenangabe gestattet. Alle übrigen Rechte bleiben vorbehalten.

Fotorechte Umschlag: ©artSILENCEcom – Fotolia.com

Empfohlene Zitierung:

Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder: Metadatenreport. Teil II: Produktspezifische Informationen zur Nutzung der Einzeldaten des Verbraucherpreisindex 2018 (EVAS-Nummer: 61111) per Scientific Use File. Version 1. DOI: 10.21242/61111.2018.00.00.3.1.0. Wiesbaden 2022.

Metadatenreport

Teil II: Produktspezifische Informationen zur Nutzung der Einzeldaten des Verbraucherpreisindex 2018 (EVAS-Nummer: 61111) per Scientific Use File

DOI: 10.21242/61111.2018.00.00.3.1.0

Version 1

Inhalt

1	Datenaufbereitung in den FDZ.....	2
1.1	Datenaufbereitung	2
1.2	Anonymisierungsmaßnahmen	3
1.3	Methodik der Verknüpfung.....	5
2	Produkt	6
2.1	Merkmale und Merkmalsbeschreibung	6
2.2	Vergleichbarkeit der Merkmale über die Zeit	23
2.3	Eckwerte relevanter Merkmale und Merkmalskombinationen	24
2.4	Auswertbare regionale Ebene.....	24
3	Praktische Hinweise	25
3.1	Hinweise zur Geheimhaltung	25
3.1.1	Gesetzliche Grundlagen der statistischen Geheimhaltung	25
3.1.2	Geheimhaltung von Ergebnissen	26
3.1.3	Praktische Tipps zur Vermeidung von Geheimhaltungsfällen ...	27
3.2	FAQ	28
3.3	Verfügbare Tools	28
4	Literaturverzeichnis	28
	Anhang 1 – Merkmalsübersicht	29
	Anhang 2 – Anonymisierungskonzept der Feinbeschreibungsmerkmale (Auspraegung1 bis Auspraegung10)	33

1 Datenaufbereitung in den FDZ

1.1 Datenaufbereitung

Scientific Use Files (SUF) sind standardisierte Einzeldatensätze, die von den Statistischen Ämtern für wissenschaftliche Vorhaben bereitgestellt werden. Im Gegensatz zu Gastwissenschaftlerarbeitsplätzen oder der kontrollierten Datenfernverarbeitung können SUF außerhalb der geschützten Räumlichkeiten der amtlichen Statistik genutzt werden.

§ 16 Abs. 6 Bundesstatistikgesetz (BStatG) ermöglicht den Statistischen Ämtern des Bundes und der Länder Einzelangaben an Hochschulen und andere Einrichtungen der unabhängigen wissenschaftlichen Forschung zu übermitteln. Voraussetzung hierfür ist jedoch, dass die Einzelangaben nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft zugeordnet werden können. „Unverhältnismäßig“ bedeutet, dass der Aufwand für eine Reidentifikation höher ist, als der Nutzen, der daraus gezogen werden kann.

SUF haben daher stets zwei sehr gegensätzliche Anforderungen zu erfüllen. Auf der einen Seite sollen sie für einen möglichst großen Teil an Forschungsvorhaben ein Höchstmaß an Analysepotenzial bieten, zum anderen müssen sie den strengen Maßgaben der faktischen Anonymität genügen.

Den Anforderungen des BStatG zur Sicherstellung der faktischen Anonymität wird bei einem SUF zum einen dadurch entsprochen, dass mit dem Datennutzer ein Vertrag geschlossen wird, in dem ein De-Anonymisierungsverbot festgeschrieben ist und Sanktionen bei Nichteinhaltung dieses Verbotes festgesetzt werden. Zum anderen wird die faktische Anonymität durch datenverändernde Maßnahmen erreicht. Im Folgenden werden die datenverändernden Maßnahmen beschrieben.

Für detaillierte Informationen zum Verbraucherpreisindex für Deutschland wird an dieser Stelle auf die Metadaten zur Erhebung (Teil I) verwiesen.

1.2 Anonymisierungsmaßnahmen

Für den SUF des Verbraucherpreisindex für Deutschland (VPI) werden nur die Merkmale bereitgestellt, die sich für die Durchführung wissenschaftlicher Vorhaben auf Basis faktisch anonymer Daten eignen. Eine Übersicht über die im SUF enthaltenen Merkmale befindet sich in Anhang 2.

Neben der Löschung von Merkmalen sind weitere Maßnahmen erforderlich, um die faktische Anonymität der Daten sicherzustellen. Diese Maßnahmen stellen im Wesentlichen auf die Merkmale ab, deren Information sowohl in den vertraulichen Daten als auch in dem möglichen Zusatzwissen eines potentiellen Datenangreifers enthalten sein können (so genannte Überschneidungsmerkmale). Bei der Recherche des möglichen Zusatzwissens haben sich insbesondere die Regionalangabe, die Feinbeschreibungsmerkmale, die Geschäftstypengewichte und die Kombination aus den Merkmalen Monat, 10-Steller, Bundesland und Geschäftstyp als mögliche Überschneidungsmerkmale herausgestellt.

Im Folgenden werden die Maßnahmen beschrieben, die auf die einzelnen Merkmale angewendet wurden.

Gemeinde

Dieses Merkmal entfällt, sodass eine regionale Gliederung nur bis zur Ebene der Bundesländer möglich ist. Dadurch werden seltene und regional einmalig vorkommende Branchen vor Aufdeckung geschützt (z.B. Schornsteinfeger, Apotheken, Möbelhäuser, Instrumentenbauer etc.). Da für die Nachverfolgung

der erhobenen Produkte über die Zeit die Gemeinde ein zentrales Merkmal ist, wird eine ID erstellt, die die erhobenen Produkte im Zeitverlauf identifiziert.

Berichtsstelle

Wie im On-Site-Material wird für die Berichtsstelle eine neue Nummer vergeben, die allerdings die Systematik der Klassifikation – dezentral, zentral und Bund – widerspiegelt. Die neue Berichtsstellenummer wird konsistent für alle SUF Berichtsjahre vergeben.

Auspraegung1 bis Auspraegung10 (Feinbeschreibungsmerkmale)

Die Inhalte der Merkmale werden vollständig pseudonymisiert¹. Entgegen der On-Site-Nutzung, ist die Verwendung einer Whitelist, also die Darstellung bestimmter Begriffe im Klartext, nicht möglich. Die Pseudonyme werden konsistent für alle Berichtsjahre vergeben.

Entfernen von Erhebungspositionen

Durch eine gemeinsame Betrachtung der Merkmale Monat, 10-Steller, Bundesland und Geschäftstyp werden Erhebungspositionen identifiziert, die zusammen in weniger als drei verschiedenen dezentralen Berichtsstellen erhoben wurden. Da diese Kombination von Eigenschaften die Geheimhaltung der dezentralen Berichtsstellen gefährden kann, werden die zugrundeliegenden Erhebungspositionen (ca. 0,9 Prozent) aus dem Datensatz entfernt.

¹ Für detaillierte Informationen zum Verfahren der Pseudonymisierung und Vereinheitlichung siehe Kaukal und Peters 2019.

Geschäftstypengewichtung

Aufgrund der sensiblen Information in der Geschäftstypengewichtung, wird das Gewicht auf der 10-Steller-Ebene nur für Deutschland insgesamt (also durch die Bundesländer gewichtete Durchschnitte der Geschäftstypengewichte) zur Verfügung gestellt.

1.3 Methodik der Verknüpfung

Der verfügbare Datensatz stellt einen ‚gestapelten‘ Datensatz aller Berichtsmonate eines Jahres dar und bildet somit ein Berichtsjahr ab. Die einzelnen Waren und Dienstleistungen erhalten im Original keine eigene ID, die über die Berichtsmonate konstant bleibt, da Produktwechsel innerhalb einer Güterklasse vorkommen können. Für die Bundesländer erhobene Preisreihen eines Berichtsmonats können über den On-Site-Zugang jedoch mit Hilfe der Merkmale „Gemeinde“, „Berichtsstelle“, „COICOP“, „MB“ (Meldebogennummer) und „Produktvariante“ über die Berichtsmonate innerhalb eines Basisjahres hinweg identifiziert werden. Für das Bundesgebiet insgesamt gültige zentral erhobene Preisreihen (Bundeslandkennung „99“) sind durch die Merkmale „Gemeinde“, „Berichtsstelle“, „COICOP“ und „Produktvariante“ eindeutig identifizierbar.

Da im vorliegenden SUF die Gemeinde zur Identifikation der Preisreihe nicht zur Verfügung steht, wird durch das FDZ eine systemfreie ID erstellt, die über die Berichtsjahre eines Basisjahrs hinweg konsistent vergeben wird. Beobachtungen der selben Produkte bzw. der entsprechenden Ersatzprodukte nach einem Erzeugniswechsel sind somit auch im SUF über die Zeit hinweg möglich.

2 Produkt

2.1 Merkmale und Merkmalsbeschreibung

Genannt sind zuerst der Merkmalsname und anschließend die Merkmalsbeschreibung.

ID – ID einer Erhebungsposition im Basisjahr

Da das Merkmal „Gemeinde“, das für die Identifikation eindeutiger Preisreihen notwendig ist, im SUF nicht zur Verfügung steht, wird seitens des FDZ eine neue ID erstellt. Diese ID wird konsistent innerhalb eines Basisjahres vergeben und ermöglicht somit die Nachverfolgung einzelner Erhebungspositionen über die Zeit.

Die ID wird anhand der Merkmale „Gemeinde“, „Berichtsstelle“, „MB“, „COICOP“ und „Produktvariante“ erstellt. Dies gilt allerdings nur für den Zeitraum zwischen zwei Meldebogenreformen. Bei einer Meldebogenreform wird der Erhebungskatalog an das Konsumverhalten der Bevölkerung angepasst. Eine Anpassung des Erhebungskatalogs findet alle fünf Jahre statt (zuletzt zum Dezember 2014).

Monat – Erhebungsmonat

Monat, in dem die Preiserhebung des Produkts durchgeführt wurde.

Jahr – Erhebungsjahr

Jahr, in dem die Preiserhebung des Produkts durchgeführt wurde.

Bundesland – Erhebungsbundesland

Amtlicher Schlüssel des Bundeslandes, in dem der Preis für das Produkt erhoben wurde.

Ausprägungen:

01 = Schleswig-Holstein	10 = Saarland
02 = Hamburg	11 = Berlin
03 = Niedersachsen	12 = Brandenburg
04 = Bremen	13 = Mecklenburg-Vorpommern
05 = Nordrhein-Westfalen	14 = Sachsen
06 = Hessen	15 = Sachsen-Anhalt
07 = Rheinland-Pfalz	16 = Thüringen
08 = Baden-Württemberg	99 = Bund
09 = Bayern	

Preise, die dem Code 99 „Bund“ zugewiesen werden, werden durch das Statistische Bundesamt erhoben und besitzen keinen Bundeslandbezug. Dies betrifft bspw. die Angaben zu Pauschalreisen.

Gewicht_BL – Gewicht des Bundeslandes

Gewichte, mit denen im Rahmen der Berechnung des Verbraucherpreisindex die Teilindizes der Bundesländer pro 10-Steller zu Teilindizes pro 10-Steller der Bundesebene zusammengefasst werden.

Berichtsstelle – Identifizierungsnummer der Berichtsstelle

Identifizierungsnummer der Berichtsstelle für die Preiserhebung. Die Vergabe der Nummern entspricht folgender Systematik:

- 1 – 4999 = Diese Nummern sind den dezentralen Berichtstellen in den Ländern vorbehalten. Sie werden einmalig in Gemeinden, aber mehrfach im Bundesland vergeben.
- 5000 – 9999 = Hierbei handelt es sich um zentral erhobene Preise (siehe auch „Erfassungsart“). Das bedeutet, dass ein Statistisches Landesamt oder das Statistische Bundesamt stellvertretend den Preis bei einem überregionalen Anbieter erhebt (z. B. einer Handelskette) und dieser auf die übrigen Bundesländer angewendet wird. Des Weiteren ist es möglich, dass innerhalb eines Landes nur eine Berichtsstelle besucht wird und deren Preise auf die übrigen Gemeinden mit Filialen übertragen wird (z. B. eine Discounterkette).
- >= 10 000 = Hierbei handelt es sich um Preise, die das Statistische Bundesamt bei zentralen Anbietern (bspw. Reisebüros, E-Commerce) erhebt und für das gesamte Bundesgebiet gültig sind.

Eine Berichtseinheit (Einzelhandels- und Dienstleistungseinheiten, einschließlich öffentlich-rechtlicher und staatlicher Anbieter) kann bei mehreren Erhebungspositionen verschiedene Berichtsstellennummern aufweisen.

Für die Nutzung im FDZ wird die Berichtsstellennummer in eine systemfreie Nummer überführt. Die Bereiche der drei Kategorien (0-4999 = dezentral; 5000-9999 = zentral; >=10000 = zentral, Statistisches Bundesamt) bleiben erhalten.

Berichtsstellenmultiplikator – Anzahl der Filialen

Gibt an, mit welcher expliziten Gewichtung alle Preise einer Berichtsstelle (Erhebungseinheit) in die Indexberechnung einfließen. Das Feld „Anzahl der Filialen“ ist in der Regel mit dem Wert „1“ belegt.

Multiplikatoren (auch: Vervielfacher) bieten die Möglichkeit, unterhalb der Geschäftstypengewichtung eine Wägungskomponente zu schaffen. Dadurch kann der Einfluss der einzelnen Preisreihe innerhalb des Durchschnittspreises der Elementarindexabgrenzung variiert und beeinflusst werden.

GKat – Geschäftskategorie

Bei der Geschäftskategorie wird zwischen folgenden Geschäftstypen unterschieden:

- 1 = Kaufhaus/Warenhaus
- 2 = Verbrauchermarkt/SB-Warenhaus
- 3 = Supermarkt
- 4 = Discounter/Fachmarkt
- 5 = Fachgeschäft
- 6 = sonstiger Einzelhandel
- 7 = Dienstleistungen/Miete
- 8 = Versandhandel/Internethandel

Gewicht_GKat – Gewicht der Geschäftstypen

Gewicht, mit denen die Elementarindizes jeder Elementarindexabgrenzung (pro Geschäftstyp) zu Teilindizes pro COICOP-10-Steller auf Bundeslandebene zusammengefasst werden.

Das Merkmal „Gewicht_GKat“ enthält nur für die einzelnen Positionen („COICOP-Typ“ = 10) ein Geschäftstypengewicht, bei Varianten eines 10-Stellers ist das Feld leer. Das Gewicht eines Geschäftstyps für einen 10-Steller mit Varianten kann stattdessen dem Merkmal „Gewicht_GKat_Oberposition“ entnommen werden.

Aufgrund der sensiblen Angaben des Geschäftstypengewichts, wird das Gewicht auf der 10-Steller-Ebene nur bezogen auf Deutschland insgesamt (also durch die Bundesländer gewichtete Durchschnitte der Geschäftstypengewichte) zur Verfügung gestellt.

Gewicht_GKat_Oberposition – Gewicht der Geschäftstypen eines 10-Stellers mit Varianten (übergeordnete Erhebungsposition)

Das Merkmal gibt das Gewicht des Geschäftstyps der Teilindizes für 10-Steller an, bei denen es sich um solche mit Varianten (Unterpositionen) handelt („COICOP_Typ“ = 12). Das bedeutet, dass nicht die einzelnen Varianten explizit gewichtet werden, sondern der aus den einzelnen Varianten zusammengesetzte 10-Steller. Die Varianten sind über das Merkmal „COICOP“ und die Meldebogennummer „MB“ identifizierbar.

Wie bei „Gewicht_GKat“ wird auch hier eine gewichtete Durchschnittsberechnung auf 10-Steller-Ebene vorgenommen.

Vermietertyp – Nummer des Vermietertyps

Es wird zwischen drei Vermietertypen unterschieden:

1 = Privater Kleinvermieter

2 = Öffentliche Trägerschaft, Wohnungsgenossenschaft

3 = Wohnungsunternehmen

Aufgrund von versehentlichen Eingaben bestehen teilweise auch Nennungen von Vermietertypen in COICOP Bereichen, die nicht der Vermietung zuzuordnen sind.

Erhebungsart – Bezeichnung der Erhebung

Hier wird die Form der Erhebung in Textform genannt. Mögliche Ausprägungen sind:

- „Begehung“,
- „Elektronisch“
- „Papier und elektronisch“.
- „Selbstaufüller“
- „Telefonisch“

MB – Identifizierungsnummer des Meldebogens

Beinhaltet die Identifizierungsnummer des Meldebogens für Preiserfassungen und ist geeignet, den COICOP weiter auszudifferenzieren. Zentral erfasste

Preise, die ausschließlich durch das Statistische Bundesamt erhoben werden, weisen hier einen fehlenden Wert auf².

COICOP – Klassifikation der Verwendungszwecke des Individualverbrauchs

Nummer der internationalen Klassifikation der Verwendungszwecke des Individualverbrauchs (COICOP = Classification of individual consumption by purpose) auf 10-Steller Ebene.

https://www.destatis.de/DE/Methoden/Klassifikationen/Private-Haushalte/sea-2013.pdf?__blob=publicationFile&v=3

[https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Classification_of_individual_consumption_by_purpose_\(COICOP\)/de](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Classification_of_individual_consumption_by_purpose_(COICOP)/de)

In den Veröffentlichungen der Statistischen Ämter werden einzelne 10-Steller zu einem neuen 10-Steller zusammengefasst (erkennbar anhand von drei Nullen am Ende des 10-Stellers). Diese zusammengefassten 10-Steller sind nicht im FDZ-Datensatz enthalten, deren zugrundeliegenden 10-Steller hingegen schon.

Kurztext – Kurztext der COICOP-Klassifikation

Kurze Beschreibung der Erhebungsposition (10-Steller) in Worten, der das einzelne erhobene Gut zugeordnet ist.

² Dies ist nicht gleichzusetzen mit der Bundeslandkennung „99“ aus dem Merkmal „Bundesland“. Vereinzelt erhebt das Statistische Bundesamt für die Bundesländer zentral Preise, die dann auf die einzelnen Bundesländer übertragen werden. Fehlende Meldebogennummern kommen daher auch bei Erhebungspositionen vor, die den Bundesländern zugeordnet sind.

Langtext – Langtext der COICOP-Klassifikation

Umfangreiche Güterbeschreibung der Erhebungsposition (10-Steller) in Worten, der das einzelne erhobene Gut zugeordnet ist. Die Güterbeschreibung ist meist weiter gefasst, um die Auswahl der passenden Güter zu erleichtern.

COICOP_Typ – Typ des 10-Stellers

Dieses Merkmal gibt an, ob es sich bei einem Produkt um das eines klassischen 10-Steller handelt oder um eine Variante (Unterposition) eines 10-Stellers unter dem zwei oder mehrere Varianten zusammengefasst sind:

- 10 = 10-Steller ohne Varianten
- 12 = Variante eines 10-Stellers

Die Gewichte des Wägungsschemas zur Berechnung des VPI liegen auf der untersten Ebene auf dem 10-Steller. Für bestimmte Erhebungspositionen existieren unterhalb dieser 10-Steller-Ebene zwei oder mehr Varianten (sog. Unterpositionen mit jeweils eigenen Meldebogennummern „MB“), die allein zur besseren Strukturierung der Erhebung dienen, z. B. zum Abbilden bestimmter Saisonstrukturen. Im Rahmen der Berechnung werden diese zu einem Elementarindex (je Geschäftskategorie) auf 10-Steller-Ebene zusammengefasst.

Kurztext_Oberposition – Kurztext des 10-Stellers mit Varianten (übergeordnete Erhebungsposition)

Wenn "COICOP_Typ" = 12 für den aktuellen Datensatz gesetzt ist, dann ist hier der Kurztext der übergeordneten Erhebungsposition eingetragen. Ansonsten ist das Feld leer.

Produktvariante – Nummer der Produktvariante

Sofern in derselben Berichtsstelle mehr als eine Variante eines Produkts innerhalb einer COICOP-Klassifikationsnummer erhoben wird, gibt dieses Merkmal die entsprechende Produktvariante als fortlaufende Nummer an.

Merkmal1 bis Merkmal10 – COICOP spezifische Benennung des Merkmals

Die sogenannten Feinbeschreibungsmerkmale enthalten weitere Details zu den Produkten, deren Preise erhoben werden. Sie dienen dazu, während der Preisermittlung sicherzustellen, dass monatlich immer dasselbe Produkt erfasst wird. Merkmal1 bis Merkmal10 benennt, welches Produktdetail analog in Auspraegung1 bis Auspraegung10 dokumentiert werden soll.

Auspraegung1 bis Auspraegung10 – Feinbeschreibungsmerkmale nach COICOP-Klassifikation

Jeder 10-Steller verfügt über bis zu zehn verschiedene Feinbeschreibungsmerkmale, die bei der ersten Preisermittlung von den Preisermittlerinnen und Preisermittlern ausgefüllt und bei Bedarf, z. B. im Rahmen von Produktwechseln, angepasst werden (siehe „Merkmal1“ bis „Merkmal10“).

In der Bereitstellung für die wissenschaftliche Nutzung in Form des SUF werden die Feinbeschreibungsmerkmale vollständig pseudonymisiert, sodass gewährleistet werden kann, dass sie keine sensiblen Informationen enthalten (siehe Anhang).

PreisErhoben – Erhobener Preis des Produkts

Erhobener Preis des jeweiligen Produkts. Da aus technischen Gründen die Eingabe des Wertes Null im Verbundprogramm des Verbraucherpreisindex nur begrenzt zulässig ist, wird von den Statistischen Landesämtern für Waren und Dienstleistungen, die kostenfrei sind, ein Preis von 0,01 € eingetragen. Wurde kein Preis erhoben, liegt ein systemfehlender Wert bzw. in Einzelfällen eine Null oder negativer Wert vor. Die Gründe für eine fehlende Preiseingabe lassen sich den Signierungen in „SigE“ und „SigB“ entnehmen.

Bestimmte Angaben stehen teilweise lediglich als bundes- oder länderspezifisch aggregierte Messzahlen zur Verfügung, die durch das Statistische Bundesamt berechnet werden. Die zugrundeliegenden Einzelpreise stehen zum jetzigen Zeitpunkt nicht zur Verfügung.

Wenn es sich bei einem Einzelwert um eine Messzahl statt eines Einzelpreises handelt, ist in dem Merkmal „Auspraegung1“ das Wort „Messzahl“ enthalten. Dieses ist von der Pseudonymisierung ausgenommen und damit lesbar (siehe auch „Auspraegung1-10“).

PreisBearbeitet – Bearbeiteter Preis des erhobenen Produkts

Der bearbeitete Preis eines Produkts wird auf Basis des erhobenen Preises und der Menge durch das Verbundprogramm berechnet. Veränderungen der ursprünglich festgelegten Menge fließen in die Berechnung ein, um auch indirekte Preisänderungen zu berücksichtigen und die Vergleichbarkeit der Preise zu erhalten. Der erhobene Preis für die neue Menge des aktuellen Monats wird auf die Basismenge umgerechnet, sodass die implizite Gewichtung der in der Basisperiode festgelegten Stichprobe erhalten bleibt.

Bei Erzeugnis- oder Berichtsstellenausfällen und Zurückweisungen von Er-satzerzeugnissen oder -berichtsstellen (Signierung EV, E4, E5, BV, B3 oder B4) wird der bearbeitete Preis durch Fortschreibung des Normalpreises des Vormonats ermittelt, da kein erhobener Preis vorliegt.

Normalpreis – Normalpreis des erhobenen Produkts

Der Normalpreis spiegelt die Preisentwicklung ohne Sonderangebote wider. Im Normalfall entspricht der Normalpreis dem bearbeiteten Preis des aktuellen Berichtsmonats.

Qualitätsänderung – Qualitätsveränderung des Produktes seit dem Vormonat

Qualitätsveränderung des Produktes seit dem Vormonat. Hat sich die Qualität der Ware bei gleichem oder verändertem Preis verringert oder erhöht, ist das Ausmaß der Verteuerung / Verbilligung in Euro angegeben.

Menge – Menge des erhobenen Produkts

Menge des erhobenen Produkts. Die dazugehörige Maßeinheit wird unter „Mass“ angegeben.

Mass – Maßeinheit der Menge des erhobenen Produkts

Einheit, in der die Mengenangabe des erhobenen Produkts erfolgt ist.

Waegungsanteil – Gewicht des jeweiligen 10-Stellers

Das Wägungsschema beschreibt die Anteile der einzelnen Güterarten (COICOP-Positionen) an den gesamten Konsumausgaben für Waren und Dienstleistungen privater Haushalte. Unter Verwendung dieses Wägungsschemas werden die Teilindizes für die einzelnen Güterarten zum Gesamtindex aggregiert.

Die Basisinformationen für die Berechnung der Gewichtung der Güterarten stammen aus der Einkommens- und Verbrauchsstichprobe (EVS). Diese werden anhand der Ergebnisse der Laufenden Wirtschaftsrechnungen (LWR) aktualisiert und ergänzt. Das Wägungsschema ist für alle Bundesländer gleich.

Die Gewichte des Wägungsschemas zur Berechnung des VPI, die sogenannten „Wägungsanteile“, liegen auf der untersten Ebene auf dem 10-Steller. Für bestimmte Erhebungspositionen existieren unterhalb dieser 10-Steller-Ebene zwei oder mehr Varianten (sog. Unterpositionen mit jeweils eigenen Meldebogennummern „MB“), die allein zur besseren Strukturierung der Erhebung dienen, z. B. zum Abbilden bestimmter Saisonstrukturen. Im Rahmen der Berechnung werden diese zu einem Elementarindex (je Geschäftskategorie) auf 10-Steller-Ebene zusammengefasst.

Das Merkmal „Waegungsanteil“ enthält nur für die einzelnen Positionen („COICOP-Typ“ = 10) einen Wägungsanteil, bei Varianten eines 10-Stellers ist das Feld leer. Der Wägungsanteil für einen 10-Steller mit Varianten kann stattdessen dem Merkmal „Waegungsanteil_Oberposition“ entnommen werden.

Für mehr Details zur Bedeutung für den Verbraucherpreisindex siehe Teil I des Metadatenreports Abschnitt 2.4.

Waegungsanteil_Oberposition – Gewicht eines 10-Stellers mit Varianten (übergeordnete Erhebungsposition)

Das Merkmal gibt den Wägungsanteil für 10-Steller an, bei denen es sich um solche mit Varianten (Unterpositionen) handelt („COICOP_Typ“ = 12). Das bedeutet, dass nicht die einzelnen Varianten explizit gewichtet werden, sondern der aus den einzelnen Varianten zusammengesetzte 10-Steller. Die Varianten sind über das Merkmal „COICOP“ und die Meldebogennummer „MB“ identifizierbar.

SigP – Preissignatur

Bei der Preissignatur wird zwischen folgenden Signierungen unterschieden:

- 00 = Preis ist gegenüber dem Vormonat unverändert
- PA = Preisänderung gegenüber dem Vormonat
- PA/PS = Preisänderung gegenüber dem Vormonat und Sonderangebot
- PK = Korrigierter Preis
- PS = Preis ist Sonderangebot

SigE – Erzeugnissignatur

Bei der Erzeugnissignatur wird zwischen folgenden Signierungen unterschieden:

- 00 = Erzeugnis gegenüber dem Vormonat unverändert
- E1 = Ersetzung durch gleichwertiges Erzeugnis
- E2 = Ersatzerzeugnis mit abweichender Qualität
- E3 = Ersetzung durch unvergleichbares Erzeugnis

E4 = Auswahl eines Ersatzerzeugnisses nicht möglich

E5 = Ersatzerzeugnis wurde zurückgewiesen

EM = Veränderte Menge bei sonst unverändertem Erzeugnis

EN = Aufnahme eines neuen Erzeugnisses

ES = Saisonartikel nicht mehr im Angebot

EV = Erzeugnis vorübergehend nicht verfügbar

ED = Erzeugnis dauerhaft nicht verfügbar

SigB – Berichtsstellensignatur

Bei der Berichtsstellensignatur wird zwischen folgenden Signierungen unterschieden:

00 = Berichtsstelle gegenüber dem Vormonat unverändert

BV = Berichtsstelle vorübergehend geschlossen

BD = Berichtsstelle dauerhaft geschlossen

B1 = Ersatzberichtsstelle innerhalb des Geschäftstyps

B2 = Ersatzberichtsstelle außerhalb des Geschäftstyps

B3 = Auswahl einer Ersatzberichtsstelle nicht möglich

B4 = Ersatzberichtsstelle wurde zurückgewiesen

BN = Aufnahme einer neuen Berichtsstelle

Durchschnittspreis – Durchschnittspreis des 10-Stellers in der jeweiligen Geschäftskategorie und pro Bundesland (Elementarindexabgrenzung), nicht gerundet

Der ermittelte Durchschnittspreis bezieht sich auf den arithmetisch gemittelten Wert aller bearbeiteten Preise innerhalb einer Elementarindexabgrenzung.

Preissätze mit B2-Signierung fließen in die Berechnung der neuen Elementarindexabgrenzung ein. Sätze mit EN-Signierungen werden nicht berücksichtigt. Für Preissätze mit E4- oder B3-Signierungen wird der fortgeschriebene Normalpreis berücksichtigt.

Abweichung Durchschnitt – Prozentuale Abweichung des bearbeiteten Preises vom Durchschnittspreis

Rev – Revidierter Wert

Der Eintrag bezieht sich auf den jeweiligen Durchschnittspreis eines 10-Stellers pro Geschäftskategorie, nicht auf den einzelnen korrigierten Preis. Erfolgt eine Fehlerkorrektur mit anschließender Neuberechnung in bereits abgeschlossenen Berichtsmonaten werden die gegenüber den ursprünglichen Ergebnissen abweichenden Durchschnittspreise mit einem "r" gekennzeichnet. Bitte beachten: Bei den genannten Korrekturen handelt es sich nicht um solche, die im Rahmen der regulären Revision des Verbraucherpreisindex vorgenommen werden!

DatPreisermittlung – Datum der Preiserhebung

DatSonderangebot – Datum, seitdem das Produkt im Sonderangebot ist

DatNichtVerfuegbar – Datum, seitdem das Produkt nicht verfügbar ist

Ursprüngliches angedachtes Erhebungsdatum, an dem ein Produkt jedoch nicht verfügbar war.

DatPreisUnveraendert – Letzte Änderung des Preises

Datum, seit dem der Preis des erhobenen Produkts unverändert geblieben ist.

Saisongut

Sofern es sich bei einer 10-Steller-Position um ein Saisongut handelt, ist das Merkmal mit 1 codiert. Dies betrifft allerdings nur Produkte der Abteilung 1 und 3 des COICOP Systems.

Als saisonale Güter werden Waren und Dienstleistungen bezeichnet, die nur saisonal verfügbar sind, d. h. die zu bestimmten Zeiten des Jahres nicht oder nur in geringen (vernachlässigbaren) Mengen angeboten werden: frischer Fisch, frisches Obst, frisches Gemüse, Bekleidungsartikel (Sommer- und Winterbekleidung).

Berechnung – Indexrelevanz des Preissatzes im gewählten Berichtsmo- nat

In der Erhebung werden u. U. nicht nur die Preise erhoben, die direkt in den Verbraucherpreisindex eingehen, sondern auch Waren und Dienstleistungen, die erst zukünftig zur Indexberechnung herangezogen werden sollen. Dies kann z. B. der Fall sein, wenn eine neue Erhebungsmethode getestet wird.

„Ja“ = geht aktuell in die Berechnung ein

„Zukunft“ = geht in Zukunft in die Berechnung ein

„Januar“ = geht ab kommendem Januar in die Berechnung ein

„E4“ = geht ab kommendem Januar nicht mehr in die Berechnung ein

Produktmultiplikator

Multiplikatoren (auch: Vervielfacher) bieten die Möglichkeit, unterhalb der Geschäftstypengewichtung eine Wägungskomponente zu schaffen. Dadurch kann der Einfluss der einzelnen Preisreihe innerhalb des Durchschnittspreises der Elementarindexabgrenzung variiert und beeinflusst werden.

Regionenvervielfacher

Multiplikatoren (auch: Vervielfacher) bieten die Möglichkeit, unterhalb der Geschäftstypengewichtung eine Wägungskomponente zu schaffen. Dadurch kann der Einfluss der einzelnen Preisreihe innerhalb des Durchschnittspreises der Elementarindexabgrenzung variiert und beeinflusst werden.

2.2 Vergleichbarkeit der Merkmale über die Zeit

Die zeitliche Vergleichbarkeit der einzelnen Positionen über die Berichtsmo-
nate hinweg ist grundsätzlich gegeben. Durch Erzeugnis- und Berichtsstellen-
wegfälle kann es jedoch zu Erzeugnisänderungen und Berichtsstellenwech-
seln kommen. Die entsprechenden Signierungen (SigP, SigE, SigB) sind daher
zu beachten. Über die Zeit vergleichbar ist letztlich der bearbeitete Preis (Preis-
Bearbeitet).

Darüber hinaus wird der Erhebungskatalog alle fünf Jahre angepasst, sodass
er das Konsumverhalten der Bevölkerung möglichst repräsentativ abbildet. Für
die Berichtsjahre 2015 bis 2019 war dies zuletzt der Dezember 2014. Dies ist
bei der Bildung von Zeitreihen zu berücksichtigen.

Die Berichtsjahre 2015 bis 2019 entsprechen dem Warenkorb des Basisjahres
2015.

2.3 Eckwerte relevanter Merkmale und Merkmalskombinationen

	2018
Januar	686 191
Februar	687 975
März	690 388
April	697 127
Mai	704 895
Juni	708 954
Juli	707 792
August	707 792
September	707 792
Oktober	707 792
November	707 792
Dezember	707 792
Gesamt	8 422 282

2.4 Auswertbare regionale Ebene

Im SUF steht eine Gliederung bis zur Bundeslandebene zur Verfügung.

3 Praktische Hinweise

3.1 Hinweise zur Geheimhaltung

3.1.1 Gesetzliche Grundlagen der statistischen Geheimhaltung

Unter Geheimhaltung versteht man das Herstellen der absoluten Anonymität der Ergebnisse statistischer Analysen. Konkret bedeutet das, dass im Rahmen der Geheimhaltung sichergestellt wird, dass mit den veröffentlichten Ergebnissen keine Rückschlüsse auf einen Einzelfall (z.B. Person, Betrieb, Einrichtung) gezogen werden können. Statistische Geheimhaltung wird überall dort angewendet, wo statistische Ergebnisse oder Einzeldaten die geschützten Räume der amtlichen Statistik verlassen.

Die Geheimhaltung in der amtlichen Statistik ist in § 16 Bundesstatistikgesetz (BStatG) geregelt und beinhaltet, dass Einzelangaben über persönliche und sachliche Verhältnisse, die für eine Bundesstatistik angegeben werden, von den jeweils durchführenden statistischen Stellen geheim zu halten sind, soweit es keine anderslautenden Bestimmungen gibt. Dies wird auch als Statistikgeheimnis bezeichnet. Das Statistikgeheimnis verpflichtet die amtliche Statistik, die erhaltenen Informationen zu schützen, d.h. sie in einer Form zu anonymisieren, die keine Rückschlüsse mehr auf die betreffende Person und den dargelegten Sachverhalt enthält. Die Geheimhaltung ist auch im Hinblick auf die informationelle Selbstbestimmung von besonderem Interesse: Viele Erhebungen der amtlichen Statistik unterliegen der Auskunftspflicht, somit steht es den Befragten nicht frei, selbst zu entscheiden, ob sie eine Information weitergeben möchten. Die amtliche Statistik muss deshalb sicherstellen, dass die erhobenen Daten keinem Befragten zugeordnet werden können.

Das BStatG sieht jedoch auch Fälle vor, in denen das Statistikgeheimnis nicht gilt. In § 16 BStatG sind die Ausnahmen von der Geheimhaltungspflicht dargestellt. Unter anderem wird dort festgelegt, unter welchen Umständen die Daten der amtlichen Statistik für die Wissenschaft zugänglich gemacht werden dürfen und welche Regeln dabei einzuhalten sind.

Die Ausführungen in den folgenden Abschnitten 3.1.2 und 3.1.3 beziehen sich auf Ergebnisse, die am Gastwissenschaftlerarbeitsplatz oder per Kontrollierter Datenfernverarbeitung erzeugt wurden (On-Site-Nutzung). Dort werden lediglich formal-anonyme Einzeldaten ausgewertet, so dass die Ergebnisse vor Herausgabe einer sorgfältigen Geheimhaltungsprüfung unterzogen werden müssen.

Die Einzeldaten eines Scientific Use Files (SUF) sind hingegen stärker anonymisiert (faktisch-anonym). Eine Geheimhaltungsprüfung der Ergebnisse durch das FDZ-Personal erfolgt hier nicht. Die Nutzerinnen und Nutzer sind allerdings ihrerseits durch entsprechende Regelungen in den Nutzungsverträgen zur Einhaltung der statistischen Geheimhaltung verpflichtet.

3.1.2 Geheimhaltung von Ergebnissen

Um die gesetzlich vorgeschriebene Geheimhaltung von Einzelfällen in den Daten sicherzustellen, müssen alle Ergebnisse, die am Gastwissenschaftlerarbeitsplatz oder per Kontrollierter Datenfernverarbeitung erzeugt werden, vor ihrer Freigabe an den Nutzer von den FDZ einer Geheimhaltungsprüfung unterzogen werden. Dabei stellen die FDZ sicher, dass die Ergebnisse absolut anonym sind und eine Reidentifikation einzelner Befragter nach menschlichem Ermessen ausgeschlossen werden kann. Entsprechend handeln auch die Fachabteilungen der Statistischen Ämter vor der Veröffentlichung von Ergebnissen.

Zur Sicherstellung der Geheimhaltung wenden die FDZ verschiedene Geheimhaltungsregeln an, die jeweils individuell auf die jeweilige Statistik zugeschnitten sind. In der Broschüre „Regelungen zur Auswertung von Mikrodaten in den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder“ werden die gebräuchlichsten Regeln zur primären Geheimhaltung dargestellt. Diese Regeln werden in den FDZ im Grunde auf alle Statistiken angewendet. Die Anlage dieser Broschüre enthält Informationen darüber, welche Geheimhaltungsregeln auf welche Statistiken anzuwenden sind.

Die Broschüre finden Sie hier:

<https://www.forschungsdatenzentrum.de/de/geheimhaltung>.

3.1.3 Praktische Tipps zur Vermeidung von Geheimhaltungsfällen

Treten in den erstellten Analysen Geheimhaltungsfälle auf, werden diese Werte von den FDZ zur Sicherstellung der Geheimhaltung durch ein Sperrmuster ersetzt. Gerade in Kreuztabellen entstehen so durch die notwendige Sekundärsperrung schnell viele „Löcher“ in den Auswertungen. Da eine einmal zur Sekundärsperrung herangezogene Tabellenzelle auch in allen folgenden Analysen gesperrt werden muss (tabellenübergreifende Geheimhaltung) – auch, wenn es in der neu erstellten Tabelle nicht nötig wäre – ist es sinnvoll, bei jeder Ergebniserstellung darauf zu achten, dass möglichst keine Geheimhaltungsfälle erzeugt werden. Treten in einem Output Geheimhaltungsfälle auf, steht es dem betreuenden FDZ frei, die Prüfung und Freigabe des Outputs abzulehnen.

Um Geheimhaltungsfälle in den Analysen zu vermeiden, sollte immer darauf geachtet werden, dass die erstellten Analysen auf ausreichend großen Fallzahlen beruhen. Bei geringen Fallzahlen empfiehlt es sich, Variablenausprägungen zusammen zu fassen und damit größere Fallzahlen zu erzielen.

3.2 FAQ

Bitte wenden Sie sich bei auftretenden Fragen an den im Impressum für fachliche Informationen genannten FDZ-Standort.

3.3 Verfügbare Tools

Für dieses Produkt werden seitens der Forschungsdatenzentren keine weiterführenden Tools angeboten.

4 Literaturverzeichnis

Kaukal, Malte und Peters, Normen, 2019: Vom Wort zur Zahl: Wie mit Hilfe automatisierter Verfahren Produktbeschreibungen in der Verbraucherpreisstatistik für das Forschungsdatenzentrum effizient bereitgestellt werden können – Ein Werkstattbericht. Hessisches Statistisches Landesamt [Zugriff am 25.03.2022]. Verfügbar unter <https://statistikhessen-blog.de/?p=977>

Anhang 1 – Merkmalsübersicht

Merkmal	Bezeichnung	Basisjahr 2010					Basisjahr 2015					
		2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
ID	ID einer Erhebungsposition im Basisjahr	+	+	+	+	+	+	+	+	+	+	+
Monat	Erhebungsmonat	+	+	+	+	+	+	+	+	+	+	+
Jahr	Erhebungsjahr	+	+	+	+	+	+	+	+	+	+	+
Bundesland	Erhebungsbundesland	+	+	+	+	+	+	+	+	+	+	+
Gewicht_BL	Bundeslandgewichtung	+	+	+	+	+	+	+	+	+	+	+
Berichtsstelle	Identifizierungsnummer der Berichtsstelle	+	+	+	+	+	+	+	+	+	+	+
Berichtsstellenmultiplikator	Anzahl Filialen	+	+	+	+	+	+	+	+	+	+	+
GKat	Geschäftskategorie	+	+	+	+	+	+	+	+	+	+	+
Gewicht_GKat	Gewicht der Geschäftstypen	+	+	+	+	+	+	+	+	+	+	+
Gewicht_GKat_Oberposition	Gewicht der Geschäftstypen eines 10-Stellers mit Varianten (übergeordnete Erhebungsposition)	+	+	+	+	+	+	+	+	+	+	+
Vermietertyp	Nummer des Vermietertyps	+	+	+	+	+	+	+	+	+	+	+
Erhebungsart	Bezeichnung der Erhebung	+	+	+	+	+	+	+	+	+	+	+
MB	Identifizierungsnummer des Meldebogens	+	+	+	+	+	+	+	+	+	+	+
COICOP	Klassifikation der Verwendungszwecke des Individualverbrauchs	+	+	+	+	+	+	+	+	+	+	+
Kurztext	Kurztext der COICOP Klassifikation	+	+	+	+	+	+	+	+	+	+	+
Langtext	Langtext der COICOP Klassifikation	+	+	+	+	+	+	+	+	+	+	+
COICOP_Typ	Typ des 10-Stellers	+	+	+	+	+	+	+	+	+	+	+

Merkmal	Bezeichnung	Basisjahr 2010					Basisjahr 2015					
		2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Kurztext_Oberposition	Kurztext des 10-Stellers mit Varianten (übergeordnete Erhebungsebene)	+	+	+	+	+	+	+	+	+	+	+
Produktvariante	Nummer der Produktvariante	+	+	+	+	+	+	+	+	+	+	+
Merkmal1	COICOP spezifische Benennung des Merkmals	+	+	+	+	+	+	+	+	+	+	+
Auspraegung1	Feinbeschreibungsmerkmal	+	+	+	+	+	+	+	+	+	+	+
Merkmal2	COICOP spezifische Benennung des Merkmals	+	+	+	+	+	+	+	+	+	+	+
Auspraegung2	Feinbeschreibungsmerkmal	+	+	+	+	+	+	+	+	+	+	+
Merkmal3	COICOP spezifische Benennung des Merkmals	+	+	+	+	+	+	+	+	+	+	+
Auspraegung3	Feinbeschreibungsmerkmal	+	+	+	+	+	+	+	+	+	+	+
Merkmal4	COICOP spezifische Benennung des Merkmals	+	+	+	+	+	+	+	+	+	+	+
Auspraegung4	Feinbeschreibungsmerkmal	+	+	+	+	+	+	+	+	+	+	+
Merkmal5	COICOP spezifische Benennung des Merkmals	+	+	+	+	+	+	+	+	+	+	+
Auspraegung5	Feinbeschreibungsmerkmal	+	+	+	+	+	+	+	+	+	+	+
Merkmal6	COICOP spezifische Benennung des Merkmals	+	+	+	+	+	+	+	+	+	+	+
Auspraegung6	Feinbeschreibungsmerkmal	+	+	+	+	+	+	+	+	+	+	+
Merkmal7	COICOP spezifische Benennung des Merkmals	+	+	+	+	+	+	+	+	+	+	+
Auspraegung7	Feinbeschreibungsmerkmal	+	+	+	+	+	+	+	+	+	+	+
Merkmal8	COICOP spezifische Benennung des Merkmals	+	+	+	+	+	+	+	+	+	+	+
Auspraegung8	Feinbeschreibungsmerkmal	+	+	+	+	+	+	+	+	+	+	+
Merkmal9	COICOP spezifische Benennung des Merkmals	+	+	+	+	+	+	+	+	+	+	+
Auspraegung9	Feinbeschreibungsmerkmal	+	+	+	+	+	+	+	+	+	+	+

Merkmal	Bezeichnung	Basisjahr 2010					Basisjahr 2015					
		2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Merkmal10	COICOP spezifische Benennung des Merkmals	+	+	+	+	+	+	+	+	+	+	+
Auspraegung10	Feinbeschreibungsmerkmal	+	+	+	+	+	+	+	+	+	+	+
PreisErhoben	Erhobener Preis des Produkts	+	+	+	+	+	+	+	+	+	+	+
PreisBearbeitet	Bearbeiteter Preis des erhobenen Produkts	+	+	+	+	+	+	+	+	+	+	+
Normalpreis	Normalpreis des erhobenen Produkts	+	+	+	+	+	+	+	+	+	+	+
Qualitaetsaenderung	Qualitätsveränderung des Produktes seit dem Vormonat	+	+	+	+	+	+	+	+	+	+	+
Menge	Menge des erhobenen Produkts	+	+	+	+	+	+	+	+	+	+	+
Mass	referenzierte Kurzbezeichnung der Maßeinheit	+	+	+	+	+	+	+	+	+	+	+
Waegungsanteil	Gewicht des jeweiligen 10-Stellers	+	+	+	+	+	+	+	+	+	+	+
Waegungsanteil_Oberposition	Gewicht eines 10-Stellers mit Varianten (übergeordnete Erhebungsposition)	+	+	+	+	+	+	+	+	+	+	+
SigP	Preissignatur	+	+	+	+	+	+	+	+	+	+	+
SigE	Erzeugnissignatur	+	+	+	+	+	+	+	+	+	+	+
SigB	Berichtsstellensignatur	+	+	+	+	+	+	+	+	+	+	+
Durchschnittspreis	Durchschnittspreis des 10-Stellers in der jeweiligen Geschäftskategorie und pro Bundesland (Elementarindexabgrenzung), nicht gerundet	+	+	+	+	+	+	+	+	+	+	+
AbweichungDurchschnitt	Prozentuale Abweichung des bearbeiteten Preises vom Durchschnittspreis	+	+	+	+	+	+	+	+	+	+	+
Rev	Kennzeichnung revidierter Werte mit „r“	+	+	+	+	+	+	+	+	+	+	+
DatPreisermittlung	Datum der Preiserhebung	+	+	+	+	+	+	+	+	+	+	+

Merkmal	Bezeichnung	Basisjahr 2010					Basisjahr 2015					
		2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
DatSonderangebot	Datum, seitdem das Produkt im Sonderangebot ist	+	+	+	+	+	+	+	+	+	+	+
DatNichtVerfuegbar	Datum, seitdem das Produkt nicht verfügbar ist	+	+	+	+	+	+	+	+	+	+	+
DatPreisUnveraendert	Letzte Änderung des Preises	+	+	+	+	+	+	+	+	+	+	+
Saisongut		+	+	+	+	+	+	+	+	+	+	+
Berechnung	Indexrelevanz des Preissatzes im gewählten Berichtsmonat	+	+	+	+	+	+	+	+	+	+	+
Produktmultiplikator		+	+	+	+	+	+	+	+	+	+	+
Regionenvervielfacher		+	+	+	+	+	+	+	+	+	+	+

Anhang 2 – Anonymisierungskonzept der Feinbeschreibungsmerkmale (Ausprägung1 bis Ausprägung10)

Die Feinbeschreibungsmerkmale stellen unstrukturierte Daten dar, die vereinzelt und unsystematisch sensible Daten enthalten können. Daher werden die u. U. vereinheitlichten Feinbeschreibungsmerkmale grundsätzlich vollständig pseudonymisiert. Die Pseudonyme werden über die zehn Feinbeschreibungsmerkmale und die beantragten Berichtsjahre konsistent vergeben. Sollten spezifische Begriffe für das Forschungsprojekt unerlässlich sein, besteht die Möglichkeit bei einer On-Site-Nutzung (Zugang über einen Gastwissenschaftsarbeitsplatz (GWAP) oder die Kontrollierte Datenfernverarbeitung (KDFV)) eine Liste von Begriffen zu definieren, die von der Pseudonymisierung ausgenommen werden. Die Liste wird durch die Mitarbeiter des FDZ Standort Hessen geprüft und freigegeben, sofern sie keine identifizierenden Inhalte aufweist. Die Klartexte der übrigen Feinbeschreibungsmerkmale werden durch die pseudonymisierten Begriffe ersetzt.

Für die Nutzung des SUF besteht die Möglichkeit der Verwendung einer Whiteliste nicht.

In Hinblick auf die vorzunehmende Anonymisierung durch Pseudonymisierung der Feinbeschreibungsmerkmale, soll durch eine Vereinheitlichung der Schreibweisen vermieden werden, dass unterschiedliche Schreibweisen zu mehreren Pseudonymen führen und der Nutzer fälschlicherweise von unterschiedlichen Feinbeschreibungen ausgehen muss.

Um zu entscheiden, welche Texteinträge zu einem gemeinsamen Eintrag vereinheitlicht werden sollen, beruht der Entscheidungsprozess auf einer Kombination aus deterministischen Entscheidungsregeln und der Anwendung maschinellen Lernens.

Die Entscheidungsregeln geben vor, welche Begriffe sich so sehr ähneln, dass sie potenziell zusammengefügt werden können. Sie wurden selbst aufgestellt, um unterschiedliche Begriffe mit fälschlicherweise hohen Ähnlichkeiten auszuschließen (Regel eins bis drei) bzw. sie wurden aus der Evaluation erster Probeläufe abgeleitet (Regel vier und fünf). Dies ist notwendig, um die grundsätzliche Wahrscheinlichkeit einer falschen Vereinheitlichung gering zu halten:

1. Begriffe werden nur innerhalb einer Produktkategorie (10-Steller) vereinheitlicht.
2. Die Begriffe müssen über denselben Anfangsbuchstaben verfügen.
3. Das Ähnlichkeitsmaß muss mindestens 90 Prozent betragen.
4. Texteinträge, in denen eine Zeichenfolge (bis zu drei Zeichen, bestehend aus Buchstaben, Ziffern oder Satzzeichen) durch Leerzeichen losgelöst vom übrigen Text steht, werden ignoriert (z. B. „Pflegestufe 1“).
5. Texteinträge, in denen zwei oder mehr Ziffernfolgen durch einen Punkt getrennt sind, werden ignoriert (z. B. „Version 1.1“, „Art-Nr. 1515.4546.5465.“ „04.07.2012“).

Die Arbeitsbelastung durch eine Prüfung jeder Vereinheitlichung soll möglichst geringgehalten werden und nur die Vereinheitlichungen geprüft werden, die vorher durch einen Algorithmus als falsch prognostiziert wurden. Zur Prognose mit möglichst hoher Genauigkeit wird hierfür ein Verfahren des vollüberwachten Maschinellen Lernens verwendet. Die als falsch prognostizierten Vereinheitlichungen werden geprüft und ggf. korrigiert.

Statistische Ämter des Bundes und der Länder,
Metadatenreport – Teil II: Produktspezifische Informationen zur Nutzung der Einzeldaten des
Verbraucherpreisindex 2018 (EVAS-Nummer: 61111) per Scientific Use File

Fotorechte Umschlag: ©artSILENCEcom – Fotolia.com