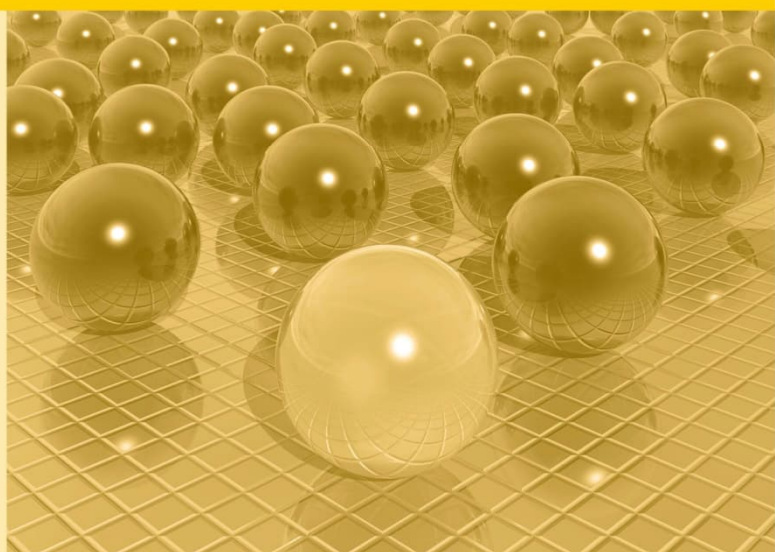


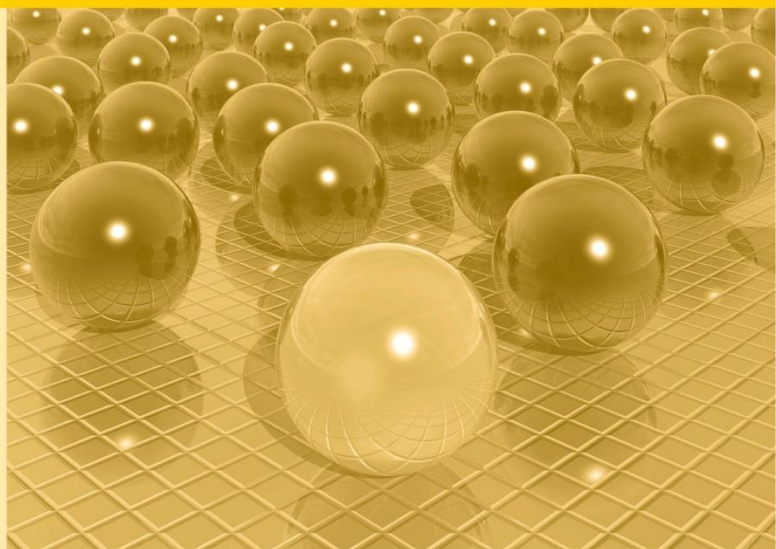
Metadatenreport



Lohn- und Einkommensteuerstatistik

Teil II: Produktspezifische Informationen zur Lohn- und Einkommensteuerstatistik 2021 für die On-Site-Nutzung

Metadatenreport



Teil II: Produktspezifische Informationen zur Lohn- und Einkommensteuerstatistik 2021 für die On-Site-Nutzung

DOI: 10.21242/73111.2021.00.00.1.1.0 (KDFV) 10.21242/73111.2021.00.00.2.1.0 (GWAP)

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:

Statistisches Bundesamt
Forschungsdatenzentrum

Tel.: 0611 75-2420
Fax: 0611 75-3915
forschungsdatenzentrum@destatis.de

Informationen zum Datenangebot:

Statistisches Bundesamt
Forschungsdatenzentrum

Tel.: 0611 75-2420
Fax: 0611 75-3915

Forschungsdatenzentrum der
Statistischen Ämter der Länder
– Geschäftsstelle –
Tel.: 0211 9449-2873
Fax: 0211 9449-8087
forschungsdatenzentrum@it.nrw.de

Erscheinungsfolge: unregelmäßig
Erschienen im Juni 2026

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

© Information und Technik Nordrhein-Westfalen, Düsseldorf, 2026
(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 Lohn- und Einkommensteuerstatistik (EVAS-Nummer: 73111). Version 1. DOI: 10.21242/73111.2021.00.00.1.1.0 (KDFV), 10.21242/73111.2021.00.00.2.1.0 (GWAP), Wiesbaden 2026.

Inhalt

1 Datenaufbereitung in den FDZ.....	4
1.1 Datenaufbereitung	4
1.2 Anonymisierungsmaßnahmen	6
1.3 Methodik der Verknüpfung.....	6
2 Produkt	6
2.1 Merkmale und Merkmalsbeschreibung	6
2.2 Vergleichbarkeit der Merkmale über die Zeit	6
2.3 Eckwerte relevanter Merkmale	7
2.4 Auswertbare regionale Ebene.....	7
3 Praktische Hinweise	7
3.1 Hinweise zur Geheimhaltung	7
3.1.1 Gesetzliche Grundlagen der statistischen Geheimhaltung.....	7
3.1.2 Geheimhaltung von Ergebnissen.....	8
3.1.3 Praktische Tipps zur Vermeidung von Geheimhaltungsfällen...	9
3.2 FAQ	9
3.3 Methodische Hinweise.....	9
Anhang	12
Abkürzungsverzeichnis.....	12
Tabellenverzeichnis.....	13
Anlage 1 – Berechnungsschema.....	14

1 Datenaufbereitung in den FDZ

1.1 Datenaufbereitung

Die Lohn- und Einkommensteuerstatistik (LESt) wird zur Onsite-Nutzung in zwei Versionen aufbereitet. Version 1 enthält die Statistik als formal anonymisiertes Vollmaterial. Diese kann uneingeschränkt per kontrollierter Datenfernverarbeitung (KDFV) genutzt werden. Aufgrund des großen Datenumfangs der LESt kann Version 1 aus technischen Gründen nicht an jedem Gastwissenschaftlerarbeitsplatz (GWAP) bereitgestellt werden. Zusätzlich zum Vollmaterial wird auch eine 10%-Stichprobe¹ der LESt formal anonymisiert und mit pseudonymisierten Gemeindeangaben für Bayern aufbereitet. Diese Version 2 kann an allen GWAPs der FDZ der Statistischen Ämter des Bundes und der Länder genutzt werden.

Folgende Schritte der Datenaufbereitung wurden in beiden Versionen vorgenommen und dienen insbesondere der Minimierung vorhersehbarer Geheimhaltungsprobleme durch unvollständige Angaben oder unplausible Merkmalsausprägungen sowie der komfortableren Handhabung der Daten.

¹ Die 10%-Stichprobe entspricht der Stichprobe, anhand der auch die faktisch anonymisierten Steuerdaten (FAST) erstellt werden. Für weitere Informationen zur Stichprobe können die Metadaten zu FAST herangezogen werden.

Tabelle 1: Übersicht der Maßnahmen zur Datenaufbereitung

Variable	Maßnahme
Land	Die Angabe zum Bundesland wurde einheitlich zu den bereits vorliegenden Wellen der LEST an den Anfang der Datei gestellt und von EF84 in „Land“ umbenannt.
EF7	Zu manchen Fällen lag im AGS nur die Landesangabe (Zweisteller) vor. Hier wurden die Angaben zum Land mit Nullen auf den Achtsteller aufgefüllt, um Geheimhaltungsprobleme bei der Auswertung des AGS zu vermeiden.
EF10a, EF10b	Die Angaben zur Steuerklasse wurden von EF10u1 in EF10a und EF10u2 in EF10b umbenannt.
EF22, EF23	Die von der Finanzverwaltung gelieferten Angaben zur Gewerbekennzahl (GKZ) bzw. dem Wirtschaftszweig (WZ) (EF22, EF23) werden nur bis zum Fünfsteller plausibilisiert, weshalb die Angaben bei der Aufbereitung auf den Fünfsteller gekürzt wurden. In Fällen, bei denen die Angaben nur auf höherer Ebene vorhanden waren, wurde die GKZ mit Nullen bis zum Fünfsteller aufgefüllt, um unnötige Geheimhaltungsfälle zu vermeiden.
EF30a, EF30b	Die Angaben zum Geburtsjahr der veranlagten Personen wurden von EF30u1 in EF30a und von EF30u2 in EF30b umbenannt. Nicht-numerische Werte wurden durch Missings ersetzt.
EF44_grund, EF44_splitting	Die Grund- und Splittingtabelle wurde gemäß den Angaben in EF19 aus der kombinierten EF44 in die zwei Variablen EF44_grund und EF44_splitting umgeschrieben.
C65100, C65120, C65140, C65160, C65220, C65240, C65260, C65309	Die Merkmale C65100, C65120, C65140, C65160, C65220, C65240, C65260, C65309 wurden als Summen aus den Angaben zu A und B der jeweiligen Einkunftsarten berechnet und ergänzt, analog zur Aufbereitung der faktisch anonymisierten Steuerdaten (FAST).
EF26, EF31, EF49, EF55, EF63, EF70, EF71, diverse C-Variablen	Freie Felder und leere bzw. kaum besetzte Kennzahlen wurden gelöscht. Die nicht auswertbaren C-Variablen sind in der Datensatzbeschreibung unter „Bemerkungen“ dokumentiert.

1.2 Anonymisierungsmaßnahmen

Aus den Daten wurden alle Hilfsmerkmale² und direkten Identifikatoren gelöscht, da diese aus Anonymisierungsgründen nicht bereitgestellt werden dürfen. Darüber hinaus wurde die Zeilennummer durch eine systemfreie Nummer ersetzt.

In Version 2 (10%-Stichprobe) wurden für die Bereitstellung der Daten am GWAP zusätzlich die bayerischen Gemeindeangaben des AGS pseudonymisiert. Auswertungen auf Gemeindeebene für Bayern sind damit am GWAP zwar möglich, die Ergebnisse können jedoch nicht den jeweiligen Gemeinden zugeordnet werden.

1.3 Methodik der Verknüpfung

Bei der Erstellung dieses Produktes fand keine Verknüpfung statt. Verknüpfungen sind jedoch gemäß § 13a BStatG³ und § 7a StStatG⁴ möglich und können bei Bedarf projektspezifisch angefragt werden.

2 Produkt

2.1 Merkmale und Merkmalsbeschreibung

Die Definitionen der nachgewiesenen Merkmale entsprechen der steuerlichen Abgrenzung der im Rahmen der Einkommensteuerveranlagung erfassten Angaben. Eine detaillierte Merkmalsliste wird aufgrund der Vielzahl an Kennzahlen der LEST in der Excel-Tabelle „ML_LESt_2021“ geführt. Unterschiede zwischen den beiden Versionen für KDFV und GWAP sind in der Merkmalsbeschreibung gekennzeichnet.

2.2 Vergleichbarkeit der Merkmale über die Zeit

Durch häufige Änderungen des Steuerrechts ist eine zeitliche Vergleichbarkeit nur eingeschränkt möglich. Zudem hat sich die Anzahl der in der Statistik nachgewiesenen nicht

² Gemäß § 5 StStatG.

³ https://www.gesetze-im-internet.de/bstatg_1987/_13a.html

⁴ https://www.gesetze-im-internet.de/ststatg_1995/_7a.html

veranlagten Lohnsteuerfälle durch die Umstellung auf die elektronischen Lohnsteuerbescheinigungen seit 2001 deutlich erhöht.⁵

Die Aufbereitungs- und Anonymisierungsmaßnahmen unterscheiden sich zwischen den verfügbaren Wellen nur in Abhängigkeit zur jeweiligen Besonderheit der Originaldaten, wobei hiervon insbesondere die älteren Wellen betroffen sind. Beispielsweise liegen für die Daten der LEST 1992 keine tiefgegliederten regionalen Angaben vor. Auf die jeweiligen Besonderheiten wird in den zugehörigen Metadatenreports verwiesen.

2.3 Eckwerte relevanter Merkmale

Auf die Ausweisung von Eckwerttabellen wird ab dem Veranlagungsjahr 2021 der LEST verzichtet. Angaben zu fehlenden und besetzten Beobachtungen, Summe, Mittelwert und Median können der Excel-Tabelle „ML_LESt_2021“ entnommen werden. Bei kategorialen Variablen, Datumsangaben und Merkmalen mit weniger als 100 besetzten Beobachtungen sind Summe, Mittelwert und Median nicht ausgewiesen. Variablen mit weniger als drei Beobachtungen sind zwar in der Datensatzbeschreibung aufgeführt, können allerdings nicht von den Datennutzenden ausgewertet werden.

2.4 Auswertbare regionale Ebene

Die tiefste auswertbare regionale Ebene ist die Gemeinde. Die Daten der LEST 2021 wurden mit dem Gebietsstand 31.12.2024 erstellt.

Am GWAP steht die Gemeindekennziffer für das Bundesland Bayern lediglich pseudonymisiert zur Verfügung.⁶

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

⁵ Für weitere Hinweise zur allgemeinen Vergleichbarkeit bitte auch den „Metadatenreport Teil I - Statistik“ zur LEST berücksichtigen.

⁶ Siehe „1.2 Anonymisierungsmaßnahmen“.

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 Merkmalsträger 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 dargelegt. 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.

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 Merkmalsträger 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 prinzipiell 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 zusammenzufassen 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 Methodische Hinweise

Der Fachbereich empfiehlt bei der Auswertung der LEST sich auf veranlagte unbeschränkt steuerpflichtige zu konzentrieren. Hierfür werden folgende Filterungen nötig:

Tabelle 2: Empfohlene Filterungen

Bis 2010	EF17 = 1, 2 oder 3 und EF1 = 01, 02, 03, 04, 05 oder 06
Ab 2012	EF47 = 1, 2 oder 3 und EF54 = 01, 02, 03, 04, 05 oder 06

Folgende Fälle werden durch diese Filterung von den Analysen ausgeschlossen:

Tabelle 3: Durch Filterung ausgeschlossene Fälle

EF54 = 09	Verlustfeststellungen, diese gab es bis 2011 nicht und es sind nur wenige EF-Variablen gefüllt.
EF54 = 11	Nur Sparzulagenfälle, diese gab es ebenfalls bis 2011 noch nicht und es sind auch hierzu nur wenige EF-Variablen und die Kennzahlen C66515/C65565 gefüllt.
EF54 = 18	Offene Elster-Lohn-Fälle = nichtveranlagte Fälle. Hier liegen nur die vom Arbeitgeber gemeldeten Angaben vor.
EF47 = 4, 5 oder 6	Beschränkt und eingeschränkt unbeschränkt Steuerpflichtige, diese haben keinen Wohnsitz im Inland und sind daher für regionale Auswertungen auszuschließen.

Zudem wurden aus Speicherplatzgründen die Nullen und Missings in den Mikrodaten miteinander vertauscht. In den betroffenen Datenfiles gibt es daher zusätzlich eine Variable `_vertauscht_`, welche angibt, ob eine Vertauschung erfolgt ist. Mit folgendem Code kann die Vertauschung rückgängig gemacht werden:

Tabelle 4: Syntaxvorlagen in SAS und Stata zu `_vertauscht_`

SAS	<pre>%macro tauschen(inputdatei, outputdatei, variablen); data &outputdatei (drop=_vertauscht_) / view = &outputdatei ; set &inputdatei (keep= &variablen _vertauscht_); array _c{*} c;; if _vertauscht_ = "1" then do _i=1 to dim(_c); if _c[_i]=0 then _c[_i]=.; else if _c[_i]=. then _c[_i]=0; end;</pre>
-----	---

	<pre> drop _i; run; %mend; </pre>
Stata	<pre> foreach x of varlist c...-c... { recode `x' (0=.) (.=0) if vertauscht==1 } </pre>

Anhang

Abkürzungsverzeichnis

AGS	= Amtlicher Gemeindeschlüssel
BStatG	= Bundesstatistikgesetz
FAST	= Faktisch anonymisierte Steuerstatistik
GKZ	= Gewerbekennzahl
GWAP	= Gastwissenschaftlerarbeitsplatz
KDFV	= Kontrollierte Datenfernverarbeitung
LESt	= Lohn- und Einkommensteuerstatistik
StStatG	= Gesetz über Steuerstatistiken
WZ	= Wirtschaftszweig

Tabellenverzeichnis

Tabelle 1: Übersicht der Maßnahmen zur Datenaufbereitung	5
Tabelle 2: Empfohlene Filterungen	10
Tabelle 3: Durch Filterung ausgeschlossene Fälle	10
Tabelle 4: Syntaxvorlagen in SAS und Stata zu <code>_vertauscht_</code>	10
Tabelle 5: Berechnungsschema zur Einkommensteuer aus der Lohn- und Einkommensteuerstatistik 2021	14

Anlage 1 – Berechnungsschema

Tabelle 5: Berechnungsschema zur Einkommensteuer aus der Lohn- und Einkommensteuerstatistik 2021

Merkmal	Milliarden Euro
Einkünfte aus:	
Land- und Forstwirtschaft	11,3
+ Gewerbebetrieb	196,3
+ selbstständige Arbeit	99,4
+ nichtselbstständige Arbeit	1.487,2
+ Kapitalvermögen	7,4
+ Vermietung und Verpachtung	43,1
+ Sonstige	134,3
= Summe der Einkünfte	1.979,0
- Altersentlastungsbetrag	4,0
- Entlastungsbetrag für Alleinerziehende	4,8
= Gesamtbetrag der Einkünfte	1.969,9
- Verlustabzug	4,4
- Sonderausgaben	287,4
- außergewöhnliche Belastungen	22,7
- Altersvorsorgebeiträge	6,3
- Steuerbegünstigungen	0,5
= Einkommen	1.652,3
- Kinderfreibetrag	32,0
= zu versteuerndes Einkommen	1.620,1
= tarifliche Einkommensteuer	353,3
- Steuerermäßigungen	20,2
+ hinzuzurechnendes Kindergeld	10,6
+ Anspruch auf Altersvorsorgezulage	0,9
+ hinzuzurechnende Steuer nach § 32d EstG	12,6
= festzusetzende Einkommensteuer*	357,1

Quelle: Statistischer Bericht - Lohn- und Einkommensteuer 2021, Tabelle 73111-07

(<https://www.destatis.de/DE/Themen/Staat/Steuern/Lohnsteuer-Einkommensteuer/Publikationen/Downloads-Lohn-und-Einkommenssteuern/statistischer-bericht-lohn-einkommensteuer-2140710217005>)

* Für Fälle ohne Einkommensteuer-Veranlagung: Einbehaltene Lohnsteuer

