

# Allgemeine und methodische Hinweise zur repräsentativen Wahlstatistik der Bundestagswahl 2021

## 1 Allgemeines

Rechtsgrundlage für die Erstellung der repräsentativen Wahlstatistik ist das Gesetz über die allgemeine und die repräsentative Wahlstatistik bei der Wahl zum Deutschen Bundestag und bei der Wahl der Abgeordneten des Europäischen Parlaments aus der Bundesrepublik Deutschland (Wahlstatistikgesetz – WStatG) vom 21. Mai 1999 (Bundesgesetzblatt Teil I, Seite 1023), zuletzt geändert durch Artikel 1a des Gesetzes vom 27. April 2013 (Bundesgesetzblatt Teil I, Seite 962).

Das WStatG legt für die Durchführung der repräsentativen Wahlstatistik im Wesentlichen folgendes fest:

1. Aus dem Ergebnis der Bundestagswahlen sind unter Wahrung des Wahlgeheimnisses in ausgewählten Wahlbezirken repräsentative Wahlstatistiken über
  - die Wahlberechtigten, Wahlscheinvermerke und die Beteiligung an der Wahl nach Geschlecht und Geburtsjahresgruppen (Statistik der Wahlbeteiligung),
  - die Wählerinnen und Wähler und ihre Stimmabgabe für die einzelnen Wahlvorschläge nach Geschlecht und Geburtsjahresgruppen sowie die Gründe für die Ungültigkeit von Stimmen (Statistik der Stimmabgabe)

als Bundesstatistik zu erstellen (§ 2 WStatG).

Für die Statistik der Wahlbeteiligung dürfen höchstens zehn Geburtsjahresgruppen gebildet werden, in denen jeweils mindestens drei Geburtsjahrgänge zusammengefasst sind. Für die Statistik der Stimmabgabe sind höchstens sechs Geburtsjahresgruppen zulässig, in denen jeweils mindestens sieben Geburtsjahrgänge zusammengefasst sind (§ 4 WStatG).

2. Die Auswahl der Stichprobenwahlbezirke trifft der Bundeswahlleiter im Einvernehmen mit den Landeswahlleitungen und den Statistischen Ämtern der Länder. Es dürfen nicht mehr als 5 % der Wahlbezirke und der Briefwahlbezirke des Bundesgebietes sowie nicht mehr als 10 % der Wahlbezirke und der Briefwahlbezirke eines Landes an der repräsentativen Wahlstatistik teilnehmen. Ein für die repräsentative Wahlstatistik ausgewählter Wahlbezirk muss mindestens 400 Wahlberechtigte umfassen,

ein ausgewählter Briefwahlbezirk mindestens 400 Wählerinnen und Wähler bei der voran gegangenen Bundestagswahl umfasst haben. Die Wahlberechtigten sind in geeigneter Weise darauf hinzuweisen, dass der Wahlbezirk in die repräsentative Wahlstatistik einbezogen ist (§ 3 WStatG).

3. Die Statistik der Wahlbeteiligung wird von den Gemeinden, in denen ausgewählte Wahlbezirke liegen, unter Auszählung der Wählerverzeichnisse durchgeführt. Die Gemeinden teilen die Ergebnisse getrennt nach Wahlbezirken dem zuständigen Statistischen Amt des Landes mit (§ 5 Absatz 1 WStatG).

Die Statistik der Stimmabgabe wird unter Verwendung von amtlichen Stimmzetteln, die zudem Unterscheidungsmerkmale nach Geschlecht und Geburtsjahresgruppen enthalten, durchgeführt. Die Gemeindebehörden leiten die ihnen von den Wahlvorstehern übergebenen, verpackten und versiegelten Stimmzettel der für die Statistik ausgewählten Wahlbezirke ungeöffnet und getrennt nach Wahlbezirken zur Auswertung an das Statistische Amt des Landes weiter. Gemeinden mit einer abgeschotteten Statistikstelle (§ 16 Absatz 5 Satz 2 des Bundesstatistikgesetzes) können die Auswertung der Stimmzettel mit Zustimmung der Landeswahlleitung selbst in der Statistikstelle vornehmen; in diesem Fall teilen sie die Ergebnisse getrennt nach Wahlbezirken dem zuständigen Statistischen Amt des Landes mit. Wählerverzeichnisse und gekennzeichnete Stimmzettel oder Ergebnisaufzeichnungen von Wahlgeräten dürfen nicht zusammengeführt werden (§ 5 Absatz 2 WStatG).

Nach Abschluss der Aufbereitung durch die Statistischen Ämter der Länder sind die Wahlunterlagen unverzüglich den Gemeindebehörden zurückzugeben und von diesen entsprechend den wahlrechtlichen Bestimmungen zu behandeln (§ 7 Absatz 3 WStatG).

4. Die Ergebnisse der repräsentativen Wahlstatistik dürfen nur für die Bundes- und Landesebene und ausschließlich durch das Statistische Bundesamt und die Statistischen Ämter der Länder veröffentlicht werden. Die Bekanntgabe von Ergebnissen für einzelne Wahlbezirke ist unzulässig.

Bei Bundestagswahlen ist die repräsentative Wahlstatistik erstmals im Jahre 1953 durchgeführt worden, hier allerdings in Rheinland-Pfalz und Bayern ohne Differenzierung nach Geburtsjahresgruppen und ohne das Saarland, das noch nicht politisch der Bundesrepublik Deutschland beigetreten war. 1957 bis 1990 sowie seit 2002 ist die Statistik unter Beteiligung aller Bundesländer vollständig erhoben worden. Für die Bundestagswahlen 1994 und 1998 hat der Bundesgesetzgeber die Durchführung der repräsentativen Wahlstatistik ausgesetzt. Ergebnisse der repräsentativen Wahlstatistik liegen für jede der seit 1979 stattfindenden Europawahlen vor.

Im Gegensatz zur allgemeinen Wahlstatistik, bei der es sich um eine Dokumentation und Auswertung der bei den Wahlorganen angefallenen Wahlergebnisse handelt, wird in der repräsentativen Wahlstatistik das Wahlverhalten, das heißt die Wahlbeteiligung und die Stimmabgabe verschiedener Bevölkerungsgruppen, untersucht. Einfachere Untersuchungen dieser Art waren schon immer in der Weise möglich, dass man Wahlergebnisse für bestimmte regionale Einheiten mit der Zusammensetzung ihrer Bevölkerung verglich (zum Beispiel Gruppierung der Kreise oder Gemeinden nach dem Anteil der Religionszugehörigkeit, der sozial-ökonomischen Struktur und dergleichen). Ein solches Verfahren erfordert aber einen großen Aufwand und vermittelt nur gröbere Aussagen. Feststellungen über die Stimmabgabe nach dem Alter konnten überhaupt nicht getroffen werden.

Die repräsentative Wahlstatistik ist mit ihrem Stichprobenumfang deutlich breiter aufgestellt, als bei entsprechenden Untersuchungen nichtamtlicher Stellen, die sich auf deutlich weniger Personen stützen. Außerdem handelt es nicht um die Auswertung dessen, was Personen über ihr Wahlverhalten vor oder nach der Wahl aussagen, sondern um die Auswertung des tatsächlichen Wahlverhaltens nach den Wählerverzeichnissen und Stimmzetteln.

Für die Feststellung der Wahlberechtigten wurden die Wählerverzeichnisse der Auswahlbezirke herangezogen, die auch Angaben über das Geschlecht und den Geburtstag enthalten. Die Feststellungen über die Stimmabgabe wurden durch Ausgabe von Stimmzetteln mit Unterscheidungsaufdruck nach Geschlecht und jeweils sechs Geburtsjahresgruppen in den Stichprobenbezirken ermöglicht. Eine Verletzung des Wahlgeheimnisses war hierbei nicht zu befürchten. In die zur Feststellung der Wahlbeteiligung herangezogenen Wählerverzeichnisse können die Gemeindebehörden,

die Wahlvorstände und die Öffentlichkeit (während der Auslegungsfrist der Wählerverzeichnisse) ohnehin Einblick nehmen.

Auch die Methode zur Feststellung der Stimmabgabe nach Geschlecht und dem Geburtsjahr lässt keine Verletzung des Wahlgeheimnisses zu. Bei der Auswertung der Stimmzettel mit Unterscheidungsbezeichnungen wird festgestellt, wie viele Frauen oder Männer einer jeden der sechs Geburtsjahresgruppen eine bestimmte Partei gewählt haben. Da aber zu jeder der zwölf Untersuchungsgruppen zahlreiche Personen gehören, können daraus keine Anhaltspunkte für die Stimmabgabe einer bestimmten Person gewonnen werden.

Die Wahlberechtigten werden durch amtliche Bekanntmachungen der Gemeindebehörden, in denen ausgewählte Wahlbezirke liegen, sowie in den betroffenen Wahllokalen durch Aushänge und Faltblätter über die Durchführung der repräsentativen Wahlstatistik informiert. In den ausgewählten Briefwahlbezirken wurden Faltblätter mit den Briefwahlunterlagen versandt.

## 2 Grundlagen des Stichprobenplans

Die Auswahl der Stichprobenwahlbezirke erfolgte durch eine Ziehung nach dem Zufallsprinzip auf Basis mathematischer Grundsätze. Dadurch sind zum einen die Objektivität des Auswahlvorgangs und die daraus resultierende Akzeptanz der Stichprobe sichergestellt. Zum anderen kann nur so die Präzision der Ergebnisse zuverlässig abgeschätzt werden.

Als Auswahlgrundlage für die Ziehung der Stichprobenwahlbezirke für die repräsentative Wahlstatistik diente die allgemeine Wahlbezirksstatistik der Bundestagswahl 2017.

Die Auswahl erfolgte durch eine geschichtete Zufallsstichprobe. Die erste Schichtung war durch die Länder gegeben, welche die sogenannten Schichtgruppen bildeten. Innerhalb dieser Schichtgruppen wurden in einem zweiten Schritt mit einem mathematischen Verfahren der Clusteranalyse<sup>1</sup> Gruppierungen (= Schichten) aller Urnenwahlbezirke gebildet. Diese fassten ähnliche Bezirke nach den Ergebnissen der Bundestagswahl 2017 zusammen. Das heißt es wurden in einem Land Wahlbezirke einer Schicht zugewiesen, in denen die Verteilung der Zweitstimmen für die Parteien CDU/CSU, SPD, AfD, FDP, DIE LINKE und GRÜNE möglichst ähnlich waren. Die Hinzunahme einer Schicht vergrößert den Anteil der durch die Schichtung erklärten Varianz, diese Zunahme nimmt jedoch mit jeder zusätzlichen Schicht immer weiter ab. Es wurden so lange Schichten hinzugefügt, bis der Anteil der zusätzlich erklärten Varianz zu klein wurde. Für die Briefwahlbezirke wurden je Land eine eigene Schicht gebildet.

Die Verteilung der Stichprobe (Allokation) zwischen den Ländern und den Schichten geschah proportional zur Anzahl der Wahlbezirke je Schicht. Dieses Vorgehen wurde getrennt nach Brief- und Urnenwahlbezirken durchgeführt. Jede Schicht bildete für sich eine eigene Auswahlgrundlage, aus der eine einfache Zufallsstichprobe gezogen wurde. Hierbei galt die Bedingung, dass nur Bezirke mit der oben beschriebenen Mindestgröße berücksichtigt werden durften und jede Schicht mit mindestens drei Wahlbezirken vertreten war. Damit wurde gegenüber einer einfachen Zufallsauswahl sichergestellt, dass möglichst alle Ergebniskonstellationen (zum Beispiel bedingt durch regionale Hochburgen einer Partei oder regionale demografische Strukturen)

berücksichtigt waren und die Hochrechnung entsprechend präzise möglich war.

Die Vereinigung aller so erhaltenen Teilstichproben bildete dann die Stichprobe der Wahlbezirke für die repräsentative Wahlstatistik. Sofern ein Stichprobenwahlbezirk gegenüber der Bundestagswahl 2017 nicht mehr in einem ähnlichen Umfang abgegrenzt wurde oder/und dieser nun weniger als 400 Wahlberechtigte beziehungsweise 400 Briefwählerinnen und Briefwähler abdeckte, wurde aus derselben Schicht ein Ersatzbezirk gezogen. Eine nachträgliche Ergänzungsstichprobe von neuen Wahlbezirken aus Gebieten, die in der Auswahlgrundlage noch zu keinem Wahlbezirk gehörten, war nicht erforderlich.

Die repräsentative Bundestagswahlstatistik 2021 beruht auf den Wahlergebnissen in 2.559 ausgewählten Wahlbezirken (davon 1.844 allgemeine und 715 Briefwahlbezirke), die für die insgesamt 94.668 Wahlbezirke (einschließlich Briefwahlbezirke) als repräsentativ angesehen werden können. Sie erstreckt sich dadurch auf 1,9 Millionen der 61,2 Millionen Wahlberechtigten und 1,4 Millionen der 46,9 Millionen Wählerinnen und Wähler.

Im Einzelnen setzte sich die Stichprobe wie folgt zusammen:

| <u>Land</u>            | <u>Auswahlsatz</u> |
|------------------------|--------------------|
| Deutschland            | 2,7 %              |
| Schleswig-Holstein     | 2,9 %              |
| Mecklenburg-Vorpommern | 4,9 %              |
| Hamburg                | 3,0 %              |
| Niedersachsen          | 2,2 %              |
| Bremen                 | 9,0 %              |
| Brandenburg            | 2,9 %              |
| Sachsen-Anhalt         | 3,8 %              |
| Berlin                 | 2,6 %              |
| Nordrhein-Westfalen    | 2,4 %              |
| Sachsen                | 2,9 %              |
| Hessen                 | 2,3 %              |
| Thüringen              | 4,4 %              |
| Rheinland-Pfalz        | 2,7 %              |
| Bayern                 | 2,4 %              |
| Baden-Württemberg      | 2,4 %              |
| Saarland               | 6,0 %              |

Die dazugehörigen Stichprobenumfänge sind in der Übersicht 2.1 dargestellt.

<sup>1</sup> Angewendet wurde die k-means-Methode, die die Daten in k Gruppen teilt, sodass die Summe der quadrierten Abweichungen vom Cluster-schwerpunkt minimal ist.

### 3 Auszählung und Aufbereitung der Ergebnisse

Für genauere Feststellungen über die Wahlbeteiligung wurden bei der Bundestagswahl 2021 zehn Geburtsjahresgruppen gebildet, die ungefähr folgenden Altersgruppen entsprechen:

| <u>Geburtsjahresgruppe</u> | <u>entspricht etwa Altersgruppe</u> |
|----------------------------|-------------------------------------|
| 2001 – 2003                | unter 21 Jahre                      |
| 1997 – 2000                | 21 – 24 Jahre                       |
| 1992 – 1996                | 25 – 29 Jahre                       |
| 1987 – 1991                | 30 – 34 Jahre                       |
| 1982 – 1986                | 35 – 39 Jahre                       |
| 1977 – 1981                | 40 – 44 Jahre                       |
| 1972 – 1976                | 45 – 49 Jahre                       |
| 1962 – 1971                | 50 – 59 Jahre                       |
| 1952 – 1961                | 60 – 69 Jahre                       |
| 1951 und früher            | 70 Jahre und älter                  |

Für die Feststellungen über die Stimmabgabe wurden sechs Geburtsjahresgruppen gebildet:

| <u>Geburtsjahresgruppe</u> | <u>entspricht etwa Altersgruppe</u> |
|----------------------------|-------------------------------------|
| 1997 – 2003                | unter 25 Jahre                      |
| 1987 – 1996                | 25 – 34 Jahre                       |
| 1977 – 1986                | 35 – 44 Jahre                       |
| 1962 – 1976                | 45 – 59 Jahre                       |
| 1952 – 1961                | 60 – 69 Jahre                       |
| 1951 und früher            | 70 Jahre und älter                  |

Die Angaben über die Wahlberechtigten, Wählerinnen und Wähler beziehungsweise Nichtwählerinnen und -wähler nach Geschlecht und Altersgruppen wurden nach der Wahl anhand der Wählerverzeichnisse gewonnen. In den Wählerverzeichnissen sind die Wahlberechtigten eingetragen und es ist bei jedem Wahlberechtigten ohne Sperrvermerk „W“ für ausgegebene Wahlscheine vermerkt, wenn er seine Stimme abgegeben hat.

Dem Inhalt der Wählerverzeichnisse entsprechend wurden folgende Zahlen unterschieden:

#### A. Wahlberechtigte insgesamt,

davon

- Wahlberechtigte ohne Sperrvermerk „W“ (= A1)
- Wahlberechtigte mit Sperrvermerk „W“ für ausgegebene Wahlscheine (= A2)

#### B. Wählerinnen und Wähler ohne Wahlschein, die in dem jeweiligen Urnenwahlbezirk ihre Stimme abgegeben haben

#### C. Nichtwählerinnen und -wähler ohne Wahlschein, das heißt Wahlberechtigte ohne Sperrvermerk „W“, die in dem jeweiligen Urnenwahlbezirk ihre Stimme nicht abgegeben haben

Die Anzahl der Wählerinnen und Wähler beziehungsweise Nichtwählerinnen und -wähler unter den Wahlscheininhabern (im Allgemeinen sind dies Briefwählerinnen und -wähler) konnten im Wahlbezirk nicht festgestellt werden, da die Personen mit Wahlschein ohne Unterlagen für die Briefwahl in einem beliebigen Wahlbezirk des Wahlkreises wählen können und die Personen mit Wahlschein und Unterlagen für die Briefwahl ihren Wahlbrief an die dafür zuständigen Stellen senden.

Bei der Anzahl der **Wahlberechtigten** wird unterschieden zwischen den Wahlberechtigten ohne und mit Sperrvermerk für ausgegebene Wahlscheine. In der Tabelle 1 werden folgende Werte ausgewiesen:

- a) Die Wahlberechtigten insgesamt (A), das heißt Wahlberechtigte mit und ohne Sperrvermerk „W“ für ausgegebene Wahlscheine.
- b) Nur Wahlberechtigte mit Sperrvermerk für ausgegebene Wahlscheine (A2 und A3).

Die **Wahlbeteiligung** wurde anhand der Anzahl der Wählerinnen und Wähler ohne Wahlschein zuzüglich der Anzahl der Wahlberechtigten mit Wahlschein gemessen an der Anzahl der Wahlberechtigten insgesamt berechnet.

Bei den Wahlberechtigten mit Wahlschein wurde berücksichtigt, dass nicht alle Wahlscheininhaberinnen und -inhaber an der Wahl teilnahmen. Dazu wurde je Land der aus der allgemeinen Wahlstatistik bekannte Anteil der tatsächlichen Wahlscheinwählerinnen und -wähler (siehe folgende Tabelle) mit der Anzahl der Wahlberechtigten mit Sperrvermerk über alle Bevölkerungsgruppen gleichermaßen multipliziert. Dadurch wurde eine Annäherung der Wahlbeteiligung an die amtlich festgestellte Wahlbeteiligung erreicht.

| Land                   | Wahlbeteiligung unter Wahlscheininhaber/-innen |
|------------------------|--|
| Deutschland            | 95,4 %   |
| Schleswig-Holstein     | 95,3 %   |
| Mecklenburg-Vorpommern | 92,6 %   |
| Hamburg                | 96,1 %   |
| Niedersachsen          | 95,3 %   |
| Bremen                 | 95,4 %   |
| Brandenburg            | 95,4 %   |
| Sachsen-Anhalt         | 95,4 %   |
| Berlin                 | 92,1 %   |
| Nordrhein-Westfalen    | 95,5 %   |
| Sachsen                | 95,9 %   |
| Hessen                 | 95,6 %   |
| Thüringen              | 95,8 %   |
| Rheinland-Pfalz        | 95,7 %   |
| Bayern                 | 95,8 %   |
| Baden-Württemberg      | 95,7 %   |
| Saarland               | 95,8 %   |

das Bundesgebiet liegen jedoch Ergebnisse nach den Urnen- und Briefwahlbezirken vor.

Für die Ermittlung der nach Geschlecht und Geburtsjahresgruppen differenzierten **Stimmabgabe** wurden alle in den ausgewählten Wahlbezirken abgegebenen Stimmzettel herangezogen. Das heißt sowohl die Stimmzettel der im Wählerverzeichnis eingetragenen Wahlberechtigten ohne Sperrvermerk „W“ als auch die Stimmzettel der Personen mit Wahlschein, die nicht per Briefwahl, sondern in dem jeweiligen Urnenwahlbezirken gewählt haben. Im Allgemeinen ist daher die Gesamtanzahl der auf Grund der Stimmzettel ermittelten abgegebenen Stimmen höher als die Anzahl der Wählerinnen und Wähler ohne Wahlschein (Wählerinnen und Wähler mit Stimmabgabevermerk im Wählerverzeichnis).

Briefwahlbezirke wurden erstmals bei der repräsentativen Bundestagswahlstatistik 2002 berücksichtigt. Briefwählerinnen und -wähler werden in die Statistik der Stimmabgabe einbezogen, um systematische Verzerrungen durch das Fehlen der Briefwahlstimmen zu vermeiden. Seit Einführung der Briefwahl bei der Bundestagswahl 1957 ist der Anteil der Briefwählerinnen und -wähler an den Wählerinnen und Wählern insgesamt von 4,9 % auf 47,3 % bei der Bundestagswahl 2021 angestiegen.

Das Ergebnis der repräsentativen Bundestagswahlstatistik 2021 wurde in den Ländern aus den Urnen- und Briefwahlbezirken zusammengefasst. Bedingt durch den jeweiligen begrenzten Auswahlsatz in den Ländern sind hochgerechnete Ergebnisse der Stimmabgabe nach Geburtsjahresgruppen und Geschlecht beschränkt auf die Briefwählerinnen und -wähler nicht möglich. Für

#### 4 Hochrechnung der Ergebnisse

Um die Ergebnisse zur repräsentativen Wahlstatistik zu erstellen, kam ein Kalibrierungsverfahren auf Basis einer Regressionsschätzung zum Einsatz. Die Regressionsschätzung hat den Charakter einer Hochrechnung mit Anpassung an „Eckwerte“ (gebundene Hochrechnung), nämlich hier das amtliche Endergebnis der Wahl. Das heißt es wurden die Hochrechnungsgewichte so kalibriert (GREG-Schätzung<sup>2</sup>), dass die Summe aller Ausprägungen eines Merkmals mit dem entsprechenden amtlichen Endergebnis übereinstimmte.

Der verallgemeinerte Regressionsschätzer für den Totalwert  $t_Y$  eines interessierenden Merkmals  $Y$  lässt sich darstellen als:

$$\hat{t}_{Y,GREG} = \sum_{k=1}^n w_k y_k,$$

wobei  $y_k$  die Merkmalsausprägung im  $k$ -ten Stichprobenwahlbezirk ist und das Hochrechnungsgewicht für den  $k$ -ten Stichprobenwahlbezirk ( $w_k$ ) bestimmt wird durch:

$$w_k = \left[ 1 + (t_X - \hat{t}_{X,HT}) \left( \sum_{k=1}^n d_k \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \right] \mathbf{x}_k d_k.$$

In obiger Formel bezeichnet  $\mathbf{x}_k$  den Vektor der Ausprägungen der sogenannten Hilfsvariablen im  $k$ -ten Stichprobenwahlbezirk,  $t_X$  sind die aus dem amtlichen Endergebnis bekannten Totalwerte der Hilfsvariablen (Eckwerte) und  $\hat{t}_{X,HT}$  ein durch „freie Hochrechnung“ ermittelter Schätzwert für den Totalwert  $t_X$ :

$$\hat{t}_{X,HT} = \sum_{k=1}^n d_k \mathbf{x}_k.$$

Ferner bezeichnet  $d_k$  die reziproke (schichtspezifische) Auswahlwahrscheinlichkeit des  $k$ -ten Stichprobenwahlbezirkes.

Es kann gezeigt werden, dass der Schätzer  $\hat{t}_{Y,GREG}$  die Eigenschaft

$$\hat{t}_{X,GREG} = \sum_{k=1}^n w_k \mathbf{x}_k = t_X$$

besitzt, das heißt bei Schätzung der Totalwerte der Hilfsvariablen durch Hochrechnung der Stichprobe

mittels der Gewichte  $w_k$  werden diese exakt getroffen. In anderen Worten,  $\hat{t}_{Y,GREG}$  liefert eine an die Eckwerte  $t_X$  gebundene Schätzung.

Die Bestimmung der Hochrechnungsgewichte  $w_k$  wird mit Hilfe einer statistischen Analysesoftware durchgeführt.

Wie bereits erwähnt, wurden im Fall der repräsentativen Wahlstatistik die Eckwerte, bei denen es sich um Stimmzahlen auf Landesebene handelt (Details siehe unten), aus der allgemeinen Wahlstatistik herangezogen. Dies hat den angenehmen Effekt, dass die Randsummen aus der repräsentativen Wahlstatistik mit den entsprechenden amtlichen Endergebnissen übereinstimmen, was bei einer freien Hochrechnung nicht gegeben wäre. Ferner wirkt sich das Schätzverfahren positiv auf die Präzision der Schätzungen aus. Insbesondere wirkt das Hochrechnungsverfahren dem Mangel entgegen, dass Wahlbezirke mit weniger als 400 Wahlberechtigten (Urne) beziehungsweise Wählerinnen und Wählern (Brief) nicht in die Stichprobe gelangen dürfen.

Die Ergebnisse der repräsentativen Wahlstatistik lassen sich drei thematischen Gruppen zuordnen: Wahlbeteiligung, Wahlverhalten und Analyse der Ungültigkeit von Wahlzetteln. Die repräsentative Wahlstatistik vertieft die Ergebnisse der allgemeinen Wahlstatistik zu diesen Themen, indem sie darstellt, wie die Ausprägungen von den Merkmalen Geburtsjahresgruppe und Geschlecht der Wählerinnen und Wähler abhängen. Die Ziele lassen sich genauer formulieren als:

1. Schätzung der Anzahl der Wahlberechtigten, der Wahrscheinlichkeiten und der Beteiligung an der Wahl in Abhängigkeit von Geschlecht und zehn Geburtsjahresgruppen.
2. Ermittlung des Wahlverhaltens in Erst- und Zweitstimme in Abhängigkeit von Geschlecht und sechs Geburtsjahresgruppen.
3. Aufschlüsselung der Häufigkeit ungültiger Stimmen nach Art und Geschlecht.

Für jede der drei thematischen Gruppen wurden, für jedes Bundesland getrennt, die Hochrechnungsgewichte  $w_k$  bestimmt, wobei stets der oben beschriebene Regressionsschätzer benutzt wird. Neben den Ergebnissen aus der Stichprobe gingen hier die Eckwerte  $t_X$  ein, die in den drei Fällen jeweils unterschiedlich sind:

2 GREG: „Generalized Regression“, siehe Deville, Jean-Claude/Särndal, Carl-Erik. Calibration Estimators in Survey Sampling. In: Journal of the

American Statistical Association. Jahrgang 87. Band 418. Juni 1992, Seite 376 ff.

1. Als Eckwerte dienten hier die Gesamtzahlen der Wahlberechtigten mit und ohne Wahlschein und der Wählerinnen und Wähler mit und ohne Wahlschein auf Landesebene.
2. In diesem Fall wurde an die Gesamtzahlen der Stimmen nach Partei beziehungsweise Gültigkeit auf Landesebene angepasst. Bei der Bundestagswahl 2021 wurden hier je für die Erst- und Zweitstimmen die zehn Kategorien „ungültig“, CDU/CSU, SPD, AfD, FDP, DIE LINKE, GRÜNE, CSU, FREIE WÄHLER, dieBasis und „Sonstige“ betrachtet.

Zur weiteren Stabilisierung der Hochrechnung wurde mit Ausnahme von Berlin für die Erststimmen der Parteien FREIE WÄHLER und dieBasis eine Abweichung von maximal 5 % vom Eckwert zugelassen. Ferner wurden für das Land Berlin die ehemaligen Gebiete Berlin-West und -Ost getrennt betrachtet. Hier durften bei allen Kategorien die Eckwerte um bis zu 10 % abweichen. Es galt jedoch die Bedingung, dass die Summen mit dem amtlichen Landesergebnis Berlins übereinstimmen. Die daraus resultierenden Abweichungen in den Randsummen können der folgenden Anhangtabelle entnommen werden.

3. Hier wurden die Gesamtzahlen ungültiger Erst- und Zweitstimmen auf Landesebene als Eckwerte verwendet.

## 5 Weitere Hinweise

Seit 1. Januar 2019 kennt das Recht drei Geschlechter (weiblich, männlich, divers) sowie die Möglichkeit, den Geschlechtseintrag im Geburtenregister offenzulassen. Grundsätzlich sind daher auch bei der repräsentativen Wahlstatistik alle Geschlechtsprägungen zu erheben. Aufgrund der erwarteten geringen Fallzahl von Personen mit dem dritten Geschlecht und ohne Angabe eines Geschlechts im Geburtenregister, wurden zum Schutz des Wahlheimnisses die Ausprägungen „männlich“, „divers“ und „ohne Angabe im Geburtenregister“ gemeinsam erhoben. Die Ergebnisse der Männer beinhalten also auch die Ergebnisse der Personen mit dem Geschlechtsmerkmal „divers“ als auch derjenigen ohne Geschlechtseintrag im Geburtenregister.

Die Zuordnung des Geschlechts der Wahlberechtigten und Wählerinnen und Wähler erfolgte grundsätzlich aufgrund des im Melderegister registrierten Geschlechts.

Da es sich um Ergebnisse einer Stichprobe handelt, weisen die Ergebnisse mit einer bestimmten Wahrscheinlichkeit einen Zufallsfehler auf. Dieser Zufallsfehler wird umso größer, desto kleiner die Zahl der in der Stichprobe erfassten Merkmalsträger ist. Die Größe des Fehlerbereiches jedes einzelnen Stichprobenergebnisses hängt dabei nicht nur von der Zahl der Erhebungseinheiten und dem Auswahlsatz ab, sondern auch von der Merkmalshäufigkeit. Deshalb hat jedes einzelne Merkmal seinen eigenen Fehlerbereich.

In einigen Übersichten und Tabellen, die nach Geburtsjahresgruppen differenzieren, sind zur besseren Verständlichkeit die Ergebnisse nach Altersgruppen dargestellt. Da bei der repräsentativen Wahlstatistik jedoch Geburtsjahresgruppen erhoben werden, können die Altersgruppen nicht exakt abgegrenzt werden. Übersicht 1.2 zeigt eine Gegenüberstellung der Alters- und Geburtsjahresgruppen für die Bundestagswahl 2021.

Ergebnisse für das „frühere Bundesgebiet“ beziehen sich auf die Bundesrepublik Deutschland einschließlich Berlin-West nach dem Gebietsstand bis zum 3. Oktober 1990. Ergebnisse für die „neuen Ländern“ beziehen sich auf das Gebiet der ehemaligen DDR inklusive Berlin-Ost.

Die Ergebnisse für das Bundesgebiet sowie das „frühere Bundesgebiet“ und die „neuen Ländern“ wurden durch Kumulation der Länderergebnisse gebildet.