THE STATA JOURNAL

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go "beyond the Stata manual" in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

The *Stata Journal* is indexed and abstracted by *CompuMath Citation Index*, *Current Contents/Social and Behavioral Sciences*, *RePEc: Research Papers in Economics*, *Science Citation Index Expanded* (also known as *SciSearch*, *Scopus*, and *Social Sciences Citation Index*.

For more information on the *Stata Journal*, including information for authors, see the webpage

http://www.stata-journal.com

Subscriptions are available from StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, telephone 979-696-4600 or 800-STATA-PC, fax 979-696-4601, or online at

http://www.stata.com/bookstore/sj.html

Subscription rates listed below include both a printed and an electronic copy unless otherwise mentioned.

U.S. and Canada		Elsewhere	
1-year subscription	$ 79	1-year subscription	$115
2-year subscription	$155	2-year subscription	$225
3-year subscription	$225	3-year subscription	$329
3-year subscription (electronic only)	$210	3-year subscription (electronic only)	$210
1-year student subscription	$ 48	1-year student subscription	$ 79
1-year university library subscription	$ 99	1-year university library subscription	$135
2-year university library subscription	$195	2-year university library subscription	$265
3-year university library subscription	$289	3-year university library subscription	$395
1-year institutional subscription	$225	1-year institutional subscription	$259
2-year institutional subscription	$445	2-year institutional subscription	$510
3-year institutional subscription	$650	3-year institutional subscription	$750

Back issues of the *Stata Journal* may be ordered online at

http://www.stata.com/bookstore/sjj.html

Individual articles three or more years old may be accessed online without charge. More recent articles may be ordered online.

http://www.stata-journal.com/archives.html

The *Stata Journal* is published quarterly by the Stata Press, College Station, Texas, USA.

Address changes should be sent to the *Stata Journal*, StataCorp, 4905 Lakeway Drive, College Station, TX 77845, USA, or emailed to sj@stata.com.

The Stata Journal (2012)
12, Number 3, pp. 375–392

Apportionment methods

Ulrich Kohler
Wissenschaftszentrum Berlin
Berlin, Germany
kohler@wzb.eu

Janina Zeh
Universität Leipzig
Leipzig, Germany
janina.zeh@wiso.uni-hamburg.de

Abstract. Apportionment methods are used to translate a set of positive natural numbers into a set of smaller natural numbers while keeping the proportions between the numbers very similar. The methods are used to allocate seats in a chamber proportionally to the number of votes for a party in an election or proportionally to regional populations. In this article, we describe six apportionment methods and the user-written `egen` function `apport()`, which implements these methods.

Keywords: st0265, apport(), egen, apportionment method, elections, Hamilton's method, Jefferson's method, Webster's method, Hill's method, Dean's method, Adams's method

1 Introduction

Democracies throughout the world face the problem of apportionment: the allocation of seats in a parliament such that each state or province receives seats in proportion to its population, or each party wins seats according to the number of votes it receives (Balinski and Young 1982, 1). Although the principle of proportionality seems fairly simple, difficulties arise because proportions can have fractions, whereas seats cannot.

The problem of apportionment has vexed mathematicians and politicians for hundreds of years. Many solutions for this problem, referred to here as "apportionment methods", have been formulated over the years. In this article, we present six apportionment methods, as well as the new `egen` function `apport()`, which implements them for Stata.

The apportionment methods are Jefferson's method, Hamilton's method, Webster's method, Hill's method, Dean's method, and Adams's method. These methods are some of the most frequently used apportionment methods, although readers might know them by different names. The Jefferson method is also known as the greatest divisor method, the d'Hondt method, and the Hagenbach-Bischoff method. The Hamilton method is elsewhere called the Hare–Niemeyer method, the method of largest remainder, or Vinton's method. Other names for Webster's method are the method of Saint-Laguë/Schaepers and the method of major fraction. Hill's method is also known as Huntington's method and the method of equal proportions. Dean's method is otherwise known as the harmonic mean method, and Adams's method is sometimes called the method of smallest divisors (Malkevitch 2002a).

 st0265

We first introduce these apportionment methods in a formal way. We then describe the new `egen` function `apport()` and conclude with examples for using that function.

2 Apportionment methods and fairness

The problem of apportionment comes in two forms. The first is expressed, for example, in Article I, Section 2 of the U.S. Constitution, which sets forth that each state is to be represented in the U.S. House of Representatives proportionally to the number of persons living in each state. The second form is common in countries having a proportional electoral system. Here seats in a parliament are allocated to political parties in proportion to the votes they receive in an election. Both forms share, however, the same idea of fairness: the distribution of seats in a parliament is considered fair if the distribution of seats is proportional to the election results or to the distribution of the population.

An algebraic way to define this overarching idea of fairness can be expressed using the so-called ideal share (see Schwingenschlögl and Pukelsheim [2006, 190]) π_i with

$$\pi_i = \frac{v_i}{V} \times S$$

where S is the size of the parliament, v_i is the number of valid votes for party i (or the population of state i), and V is the total number of valid votes cast in the election (or a country's total population). The value π_i represents the number of seats that a party (or a state) ideally should get to maintain a "fair" distribution of seats.

The problem of apportionment is that the ideal share is normally not an integer. Because fractional seats are not possible, the condition that the seats for each party are always equal to π_i cannot be met for parliaments of a fixed size.[1] A political decision must be made with respect to what kinds of deviations from the ideal share are acceptable. The following concepts of fairness predominate in these considerations:

Quota condition. According to the quota condition, an apportionment method is considered fair if the difference between the ideal share and the realized number of seats remains within narrow bounds. A formal definition of the quota condition is that each party should get at least the rounded-down number of the ideal share and at most the rounded-up number of the ideal share; that is,

$$\lfloor \pi_i \rfloor \leq s_i \leq \lceil \pi_i \rceil$$

with s_i being the number of seats allocated to party i, and $\lfloor \pi_i \rfloor$ and $\lceil \pi_i \rceil$ denoting the floor and ceiling of the ideal share, respectively.

Consistency condition. An apportionment method is considered consistent if additional votes for party C cannot lead to party A losing seats to party B. Apportion-

1. In what follows, we use the terms "parties", "votes", and "total votes cast" for i, v, and V, respectively. However, all statements made are equally true for states, state populations, and the total population.

ment methods that fulfill the consistency condition prevent paradoxes regarded as unfair. Specifically, consistent apportionment methods guarantee that

- no party loses seats if the parliament grows in size (*house monotony*); and that
- no party loses seats if it wins additional votes, and vice versa (*vote monotony*, *population monotony*).

Majority condition. The majority condition states that if a party obtains the absolute majority of the votes, then it should also get an absolute majority of seats. There is often a distinction between a weak majority condition and a strict majority condition. The weak majority condition is fulfilled if a party with at least 50% of the votes gets at least 50% of the seats; the strict majority condition is fulfilled if a party with more than 50% of the votes gets at least 50% of the seats. Similarly to the majority condition, the minority condition states that if a party does not reach the absolute majority of votes, it should not get the absolute majority of seats.

Global optimization. More recently, so-called global optimization approaches have been proposed for fairness evaluation. These approaches search for apportionment methods that minimize the sum of a given fairness measure over all the parties (Malkevitch 2002b). One example of such an attempt is to assess the fairness of an apportionment method according to the success value of votes, which is

$$e_i = \frac{s_i/S}{v_i/V} = \frac{s_i/S}{\pi_i}$$

If all parties received the ideal share π_i, the success value would become 1 for all parties. It is therefore possible to define the sum of squared differences between 1 and the success value,

$$\text{ESS} = \sum v_i \left(e_i - 1\right)^2 \tag{1}$$

and to consider as fair those apportionment methods that minimize the error sum of squares (ESS) (Pukelsheim 2003). However, assessing the fairness of an apportionment method by means of ESS is just one example of a global optimization method. Another approach is based on a comparison of the so-called "representative weight of one seat" with the ideal representative weight.[2] The representative weight of one seat is $w_i = v_i/s_i$, the number of voters for party i that is represented by one seat for that party. If each party had received the ideal share, the representative weight would have been $w_i = v_i/\{(v_i \times S)/V\} = V/S$ for all parties. It can be said that the fairness of an apportionment increases as w_i approximates V/S. This view of fairness has been promoted in several decisions made by the German federal constitutional court, which are based on pairwise comparisons of

2. The term "representative weight of one seat" is our translation of the German term *Vertretungsgewicht eines Mandats*, used by Pukelsheim (2000).

representative weights of parties (for details, see Pukelsheim [2000]). A global optimization criterion that directly follows from this is the sum of squared differences between the representative weight and the ideal representative weight:

$$\mathrm{WSS} = \sum s_i \left(w_i - \frac{V}{S} \right)^2$$

Other global optimization criteria can likewise be developed.

A scientific answer to the question of which apportionment method is the fairest could be provided if one method beat the others on any of the above criteria. Unfortunately, this cannot be the case. In a very important contribution, Balinski and Young (1982) proved that apportionment methods fulfill either the consistency condition or the quota condition, but never both. Moreover, no single apportionment method optimizes all the global optimization criteria proposed thus far. The decision to use a particular method therefore depends on the policy aims. Different countries have made different decisions and quite often without the knowledge that we have today. Presumably, the historical reality was that the decisions were driven not only by issues of fairness but also by presumptions about the relative advantages for the political actors at the time of the decisions. Finally, algebraically equivalent apportionment methods have been developed independently from one another in different countries. All of this helps explain the existence of the many different solutions for the same problem and the variations therein.

In what follows, we introduce one quota method and five different divisor methods. The major difference between these two families is that the quota method fulfills the quota condition, whereas all divisor methods are consistent. The formulas we give are used in the implementation of the `apport()` function described below. We have primarily developed them following Balinski and Young (1982, 96–102), with additional helpful hints found in Malkevitch (2002a) and on the fine German-language website Wahlrecht.de, maintained by Wilko Zicht, Martin Fehndrich, and Matthias Cantow.[3] We wrote the formulas in a notation that simplifies the implementation in Stata.

2.1 Hamilton's method

The Hamilton method was introduced by the treasury secretary Alexander Hamilton in 1791. That year, it was approved by Congress as the method to be used for the first American election, but subsequently was vetoed by President Washington—in the very first exercise of the veto power by a president of the United States.[4] The method was used for the House of Representatives between 1852 and 1911, when it was replaced by Webster's method. It is still used in Russia, Ukraine, Namibia, and Hong Kong. Until recently, it was used in Germany.

3. See http://www.wahlrecht.de/verfahren/index.html.
4. See Balinski and Young (1982, chap. 3) for a comprehensive description of the controversy over the apportionment method for the U.S. Congress.

The Hamilton method distributes the seats in two steps. First, the integer value of the ideal share is directly allocated to each party as its number of seats:

$$s_i^{\text{Ham, Init}} = \left\lfloor \frac{v_i}{V} \times S \right\rfloor = \lfloor \pi_i \rfloor$$

Second, the remaining seats are allocated to the parties in the order of their remainder r_i:

$$r_i = |\pi_i - \lfloor \pi_i \rfloor|$$

If several parties have the same decimal parts, either the seats get raffled (Germany) or they are allocated in sequence of the π_i strength of the party (Russia). By default, the `apport()` function does the latter, but this can be changed with the `raffle` option.

2.2 Divisor methods

The aim of divisor methods is to divide the number of votes by a common number D and to allocate to each party the rounded number of this fraction. Divisor methods differ in the rule used for rounding, and the main task is to find a value for D such that the sum of the allocated seats is equal to the size of the parliament. A formal way to describe the general idea of divisor methods is

$$s_i^{\text{Div}} = \left\lceil \frac{v_i}{D} - \tau \right\rceil \quad \text{and} \quad \sum s_i^{\text{Div}} = S \quad \text{for some } D \tag{2}$$

Note how $\lceil v_i/D - \tau \rceil$ is used to express rounding of the values for v_i/D. For standard rounding, $\tau = 0.5$, but in the general case, τ can be any breakpoint value between 0 and 1. The various divisor methods described below are defined by setting a value for τ.

Once τ is fixed, the task is to find a value for D such that $\sum s_i = S$. There is generally an interval of possible values for D that all produce the same seat allocation.

Finding the value for D involves some trial and error, but the general procedure is as follows. Choose as a starting value the ideal representative weight (or average district size),

$$D^{\text{Init}} = \frac{V}{S}$$

and use this value to produce an initial allocation of seats using the first rule of (2). If this initial allocation of seats already sums up to S, we keep this allocation. If

$$\sum \left\lceil \frac{v_i}{D^{\text{Init}}} - \tau \right\rceil < S$$

we choose another value for D somewhat below D^{Init}. Otherwise, we choose a value slightly above D^{Init}. There is a technique to ascertain how slight "slightly" should be; it is described below, under the heading *Finding the modified district size*.

Before presenting the various divisor methods one by one, we should point out the similarity between divisor methods and the Hamilton method. As shown above, the initial step of the Hamilton method allocates the integer part of the ideal share π_i to the parties. Because

$$\left\lceil \frac{v_i}{D^{\mathrm{Init}}} - \tau \right\rceil = \left\lceil \frac{v_i}{\frac{V}{S}} - \tau \right\rceil = \left\lceil \frac{v_i}{V} \times S - \tau \right\rceil = \lceil \pi_i - \tau \rceil$$

this resembles the initial allocation for a divisor method with $\tau = 1$ (the Jefferson method).

Jefferson's method

The Jefferson method is defined by $\tau = 1$, which means using the integer value of v_i/D for the allocation of seats. The method was developed by the future president Thomas Jefferson and enforced for the allocation of seats in the House of Representatives after President George Washington's veto of the Hamilton method. It remained the apportionment method until 1842 and is still used in countries such as Japan, Slovenia, the Netherlands, and Israel.

Like any divisor method, the Jefferson method violates the quota condition. The violation of the quota condition is such that it favors big parties or states (Marshall, Olkin, and Pukelsheim 2002; Schuster et al. 2003).

Webster's method

Webster's method is defined by $\tau = 0.5$, which means using the standard rounding of v_i/D for the allocation of seats. The method was developed by leading American statesman Daniel Webster in 1832 and used as an apportionment method for the House of Representatives between 1842 and 1852 and between 1911 and 1941. Other countries using Webster's method are Bosnia, Norway, and (since 2009) Germany.

The Webster method also violates the quota condition. However, it favors big parties less than the Jefferson method does (Marshall, Olkin, and Pukelsheim 2002; Schuster et al. 2003) and minimizes ESS as defined by (1) above (Pukelsheim 2000).

Hill's method

Hill's method is defined by using the geometric mean of the floor and ceiling of v_i/D as the threshold value for rounding. The geometric mean of two arbitrary numbers x_1 and x_2 is defined as $\sqrt{x_1 \times x_2}$. If x_1 and x_2 are neighboring integer values, the breaking point τ is the fractional remainder of the geometric between two integers. In our case, it can be obtained by subtracting the floor of v_i/D from the geometric mean of the floor and ceiling of v_i/D:

$$\tau_i = \sqrt{\left\lfloor \frac{v_i}{D} \right\rfloor \times \left\lceil \frac{v_i}{D} \right\rceil} - \left\lfloor \frac{v_i}{D} \right\rfloor$$

Trying out this formulation with some arbitrary values helps to understand the logic. If v_i/D is below 1, the breaking point becomes 0,

```
. display sqrt(floor(0.5) * ceil(0.5)) - floor(0.5)
0
```

and if v_i/D becomes larger, the breaking point approximates 0.5 when v_i/D becomes large:

```
. display sqrt(floor(1.5) * ceil(1.5)) - floor(1.5)
.41421356
. display sqrt(floor(42.5) * ceil(42.5)) - floor(42.5)
.49705872
```

Hence, unlike the divisor methods discussed so far, τ is not constant for all parties. The characteristic that τ_i becomes 0 for small numbers of v_i/D guarantees that all parties with at least one vote receive a seat. This characteristic makes the method particularly attractive for the apportionment problem of the House of Representatives, where each state must receive at least one seat. Consequently, the method, which was proposed in 1911 by the former director of the U.S. Census Bureau, Joseph A. Hill, has been used for the seat apportionment of the U.S. Congress since 1941.

Hill's method also violates the quota condition. It favors big parties less than both the Jefferson method and the Webster method do (Marshall, Olkin, and Pukelsheim 2002; Schuster et al. 2003), primarily because of its characteristic of giving seats to parties that otherwise would not receive one.

Dean's method

Dean's method is defined by using the harmonic mean as a threshold value for rounding v_i/D. The harmonic mean of two arbitrary numbers x_1 and x_2 is $2x_1x_2/(x_1+x_2)$, so that the breaking point τ of Dean's method is defined as

$$\tau_i = \frac{2 \times \left\lfloor \frac{v_i}{D} \right\rfloor \times \left\lceil \frac{v_i}{D} \right\rceil}{\left\lfloor \frac{v_i}{D} \right\rfloor + \left\lceil \frac{v_i}{D} \right\rceil} - \left\lfloor \frac{v_i}{D} \right\rfloor$$

Like Hill's method, τ is not constant for all parties. It is again 0 for $v_i/D = 0$ and exponentially approximates 0.5 with v_i/D, although at a smaller rate.

The American mathematician James Dean proposed this method in 1832 as a method for allocating seats for the House of Representatives. However, it has never been used for this task or for allocating seats anywhere else, perhaps in part because Dean's method favors small parties more than the other methods discussed so far do (Marshall et al. 2002; Schuster et al. 2003).

Adams's method

Adams's method is defined by $\tau = 0$, which means using the ceiling value of v_i/D for the allocation of seats. The apportionment method was suggested in 1822 by former President John Quincy Adams but was never used.

Like all divisor methods, the Adams method violates the quota condition. Of all the methods discussed, it mostly favors small parties or states (Marshall et al. 2002; Schuster et al. 2003), which again may be one reason why it is so rarely used.

Finding the modified district size

As mentioned above, all divisor methods use the average district size $D^{\text{Init}} = V/S$ for the initial allocation of seats. If this initial allocation of seats does not sum up to S, one must find the modified district size D^{Mod} for which

$$\sum \left\lceil \frac{v_i}{D^{\text{Mod}}} - \tau \right\rceil = S$$

for a given τ. The remainder of this subsection shows how D^{Mod} can be calculated from the results of the initial step.

Let $s_i^{\text{Div,Init}}$ be the number of seats allocated using D^{Init}, and $\Delta = \sum s_i^{\text{Div,Init}} - S$ be the difference between the number of seats allocated and the number of seats to be allocated. If $\Delta \neq 0$, obtain a modified number of seats for each party by either increasing or decreasing the initial number of seats by 1:

$$s_i^* = \begin{cases} s_i^{\text{Div,Init}} - 1 & \text{if } \Delta > 0 \\ s_i^{\text{Div,Init}} + 1 & \text{if } \Delta < 0 \end{cases}$$

We now use this modified number of seats to obtain a modified breakup value τ_i^* according to the rounding rule of the respective apportionment method:

$$\tau_i^* = \begin{cases} 1 & \text{Jefferson} \\ 0.5 & \text{Webster} \\ \sqrt{s_i^* \times (s_i^* + 1)} - s_i^* & \text{Hill} \\ \frac{2 \times s_i^* \times (s_i^*+1)}{s_i^* + (s_i^*+1)} - s_i^* & \text{Dean} \\ 0 & \text{Adams} \end{cases}$$

The modified seat numbers s_i^* and the modified breakup values τ_i^* are then used to obtain i suggestions for D:

$$D_i^* = \frac{v_i}{s_i^* + \tau *_i}$$

We sort these suggestions by size and pick a value either slightly above the smallest one or slightly below the largest one; formally,

$$D^{\text{Mod}} = \begin{cases} D^*_{|\Delta|} + \delta & \text{if } \Delta > 0 \\ D^*_{K-|\Delta|+1} - \delta & \text{if } \Delta < 0 \end{cases}$$

with K being the number of parties or states in the respective apportionment problem and δ being a floating-point number close to 0.[5] In other words, D^{Mod} becomes a value just above (or below) the Δ smallest (or largest) value of D^*.

The technique works well as long as the initial allocation of seats does not differ from the optimal allocation by more than one seat for any party. This is usually the case for the Webster, Hill, and Dean methods, but often is not the case for the Jefferson and Adams methods. A modification is to also obtain all values $s_i^{*d} = s_i \pm d$, $d = 1, \ldots, \Delta$ and the corresponding values of D^{*d}. D^{Mod} is then the Δ smallest or Δ largest value of all the values D_i^{*d}. The `egen` function `apport()` uses this modification for all divisor methods.

2.3 Variations of apportionment methods

Besides the choice of the apportionment method, election laws throughout the world prescribe a number of variations for the application of the method. Two interrelated variations have been implemented in the `apport()` function.

Barring clauses. These clauses define a threshold for participation in the allocation of seats. In Germany and Latvia, for example, only parties with more than 5% of the valid votes participate in the allocation of seats. In Israel, this threshold is 2%, and in Turkey, it is 10% (Nohlen et al. 2000, 355ff).

Exceptions. Countries with barring clauses sometimes allow exceptions from it. For example, parties of ethnic minorities are sometimes exempted from the barring clause. Another example is the German *Grundmandate* rule. German election law gives each voter two votes: the first vote is for a candidate in a constituency, while the second vote is for a party list. The composition of parties in the parliament is decided (almost) exclusively by the distribution of the second votes, but parties that win the most first votes in at least three constituencies are exempted from the barring clause.

3 The egen function apport()

3.1 Syntax

`egen` [*type*] *newvar* `=` `apport(`*varname*`)` [*if*] [*in*] [`,` `method(`*keyword*`)`
`size(`*#* | *varname*`)` `threshold(`*#* | *varname*`)` `exception(`*exp*`)` `raffle`]

`by` is allowed; see [U] **11.1.10 Prefix commands**.

5. The `apport()` function uses the value of `epsfloat()`.

3.2 Description

The egen function `apport(`*varname*`)` creates a new variable holding the number of seats on the basis of the absolute number of valid votes in *varname*.

3.3 Options

`method(`*keyword*`)` is used to select the apportionment method. The apportionment methods described above can be specified using one of the following keywords:

Method	Keyword and synonyms
Hamilton	`hamilton`, `hare-niemeyer`, `remainder`, `vinton`
Jefferson	`jefferson`, `dhondt`, `hagenbach-bischoff`, `greatest`
Webster	`webster`, `stlague`, `majorfraction`
Hill	`hill`, `huntington`, `geometric`
Dean	`dean`, `harmonic`
Adams	`adam`, `smallest`

The default is `method(jefferson)`.

`size(`*#* | *varname*`)` is used to specify the number of seats to be allocated. Either use a positive integer or use the name of a variable holding the number of seats to be allocated. In the latter case, the variable should be constant within one apportionment problem (that is, for one election). The default is `size(100)`.

`threshold(`*#* | *varname*`)` is used to set the barring clause. Within the parentheses, put the size for the barring clause as a percentage or specify the name of a variable holding the value for the barring clause. The variable must be constant within one apportionment problem (that is, for one election).

`exception(`*exp*`)` is used to specify exceptions from the barring clause. Within the parentheses, specify an expression indicating the exempted observations (see `help exp`).

`raffle` is only allowed for `method(hamilton)`. It allows a seat to be raffled if several parties have the largest remainder. By default, the seat goes to the party with more voters. In practice, either decision rule is seldom necessary; however, in the exceptional case, a message is displayed.

3.4 Expected data structure

The `apport()` function requires as input apportionment problems in "long form", that is, the units of one apportionment problem should be the observations of the dataset. If the dataset holds more than one apportionment problem, then these should be organized with one below the other. As a very minimum, the dataset should contain a variable holding the absolute number of valid votes for a party or the population sizes of regional subdivisions of a country. This required data structure is exemplified in two supplementary datasets of this article.

The first dataset holds the results of the U.S. Censuses from 1790 to 1970 as reported by Balinski and Young (1982, 158–176). The dataset reports for each year the populations of the individual U.S. states and the size of the House of Representatives at that time. The list of the first 18 observations illustrates that the observations of one year form one apportionment problem.

```
. use uspop
(Balinsky (1982: 158-176))
. list in 1/18, sepby(year)
```

	state	pop	year	size
1.	Virginia	630560	1790	105
2.	Massachusetts	475327	1790	105
3.	Pennsylvania	432879	1790	105
4.	North Carolina	353523	1790	105
5.	New York	331589	1790	105
6.	Maryland	278514	1790	105
7.	Connecticut	236841	1790	105
8.	South Carolina	206236	1790	105
9.	New Jersey	179570	1790	105
10.	New Hampshire	141822	1790	105
11.	Vermont	85533	1790	105
12.	Georgia	70835	1790	105
13.	Kentucky	68705	1790	105
14.	Rhode Island	68446	1790	105
15.	Delaware	55540	1790	105
16.	Virginia	747362	1800	141
17.	Pennsylvania	601863	1800	141
18.	New York	577805	1800	141

The second dataset holds the results of all German federal elections since 1949. Similarly to the U.S. population data, this dataset reports for each election the absolute number of valid votes for all parties that participated in an election and the size of the German *Bundestag* at that time.[6] The following list illustrates that the structure of this election data differs from that of the U.S. population data only insofar as states are replaced by parties and population sizes are replaced by absolute numbers of valid votes.

```
. use electionsde
. list eldate party votes size if year(eldate)==1949, sepby(eldate)
```

	eldate	party	votes	size
1.	14 Aug 49	RWVP	21931	400
2.	14 Aug 49	EVD	26162	400
3.	14 Aug 49	SSW	75388	400
4.	14 Aug 49	RSF	216749	400
5.	14 Aug 49	DKP/DRP	429031	400
6.	14 Aug 49	WAV	681888	400
7.	14 Aug 49	Zentrum	727505	400
8.	14 Aug 49	DP	939934	400
9.	14 Aug 49	BP	986478	400
10.	14 Aug 49	Parteilose	1141647	400
11.	14 Aug 49	KPD	1361706	400
12.	14 Aug 49	FDP	2829920	400
13.	14 Aug 49	SPD	6934975	400
14.	14 Aug 49	CDU/CSU	7359084	400

The next section presents examples for applying the `apport()` function with the use of these datasets.

4 Using egen apport()

4.1 Introductory example

We will now allocate seats for the House of Representatives in the year 1790 by using the two methods discussed at that time. We will first use the Hamilton method and then the Jefferson method (the default), choosing option `size(105)` because 105 was the size of the House of Representatives in 1790.

6. This dataset was compiled from information printed in Statistisches Bundesamt (2005) and reported on the *Bundeswahlleiter* website (http://www.bundeswahlleiter.de/de/). It is part of a larger dataset of a research project on the potential influence of nonvoters on election outcomes; for details, see Kohler (2011).

```
. use uspop if year==1790
(Balinsky (1982: 158-176))
. egen ham = apport(pop), method(hamilton) size(105)
Seat goes to stronger party; consider option -raffle-
. egen jeff = apport(pop), size(105)
. list if ham != jeff
```

	state	pop	year	size	ham	jeff
1.	Virginia	630560	1790	105	18	19
15.	Delaware	55540	1790	105	2	1

The two methods differ in the number of seats allocated to Virginia and Delaware. The Jefferson method allocates one more seat to Virginia than the Hamilton method does, whereas Delaware gets one fewer seat. Hence, President Washington's veto of the Hamilton method not only hindered an "inconsistent" apportionment method but also brought his home state one additional seat.

4.2 Solving more than one problem

The by prefix can be used for datasets with more than one apportionment problem. We will use the by prefix to produce the data for all appendix tables presented by Balinski and Young (1982), although we will present only the table for 1790 to save space.

```
. use uspop, clear
(Balinsky (1982: 158-176))
. foreach m in adam dean hill webster jefferson hamilton {
  2.         by year: egen `m´ = apport(pop), method(`m´) size(size)
  3. }
Seat goes to stronger party; consider option -raffle-
. list state adam-hamilton if year==1790, noobs sum sep(15)
```

	state	adam	dean	hill	webster	jeffer~n	hamilton
	Virginia	18	18	18	18	19	18
	Massachusetts	14	14	14	14	14	14
	Pennsylvania	12	12	12	13	13	13
	North Carolina	10	10	10	10	10	10
	New York	10	10	10	10	10	10
	Maryland	8	8	8	8	8	8
	Connecticut	7	7	7	7	7	7
	South Carolina	6	6	6	6	6	6
	New Jersey	5	5	5	5	5	5
	New Hampshire	4	4	4	4	4	4
	Vermont	3	3	3	2	2	2
	Georgia	2	2	2	2	2	2
	Kentucky	2	2	2	2	2	2
	Rhode Island	2	2	2	2	2	2
	Delaware	2	2	2	2	1	2
Sum		105	105	105	105	105	105

This time, we have specified a variable name with the `size()` option. The variable holds the number of seats to be allocated for each census year. It must be constant within each apportionment problem, though it may vary between them. An error message is issued if this is not the case.

4.3 Securing at least one seat

The example in the previous subsection does not fully reproduce the appendix tables of Balinski and Young (1982), because the U.S. Constitution gives at least one seat to each state, a feature not guaranteed by some of the apportionment methods discussed here. For example, the Jefferson method regularly does not allocate seats to all states:

```
. tabulate year if jefferson==0
Census Year |      Freq.     Percent        Cum.
------------+-----------------------------------
       1850 |          1        3.03        3.03
       1860 |          3        9.09       12.12
       1870 |          2        6.06       18.18
       1880 |          1        3.03       21.21
       1890 |          4       12.12       33.33
       1900 |          3        9.09       42.42
       1910 |          2        6.06       48.48
       1920 |          3        9.09       57.58
       1930 |          3        9.09       66.67
       1940 |          3        9.09       75.76
       1950 |          3        9.09       84.85
       1960 |          3        9.09       93.94
       1970 |          2        6.06      100.00
------------+-----------------------------------
      Total |         33      100.00
```

Two techniques seem sensible for guaranteeing that each state receives at least one seat. The first technique starts by giving one seat to all states that do not receive one through a “plain” application of the method. It then subtracts the number of these seats from the number of seats to be allocated. Finally, the apportionment method is applied to the reduced House size by using all other states. An application of this technique shows the differences between the initial and the modified solutions:

```
. by year, sort: egen null = sum(jefferson==0)
. generate resize = size - null
. by year: egen jeff1 = apport(pop) if jefferson, size(resize)
(33 missing values generated)
. replace jeff1 = 1 if !jefferson
(33 real changes made)
```

```
. list state jefferson jeff1 if year == 1890 & jefferson != jeff1

      +-----------------------------------+
      |          state   jeffer~n   jeff1 |
      |-----------------------------------|
277.  |       Virginia         10       9 |
285.  | South Carolina          7       6 |
292.  |          Maine          4       3 |
298.  |        Vermont          2       1 |
303.  |        Montana          0       1 |
      |-----------------------------------|
304.  |          Idaho          0       1 |
305.  |        Wyoming          0       1 |
306.  |         Nevada          0       1 |
      +-----------------------------------+
```

This technique leads to the allocations presented by Balinski and Young (1982, 157–177).

The second technique first allocates one seat to all states and then uses the apportionment method for the consequently reduced House size. Applying this procedure with the Jefferson method reproduces the results of Adams's method:

```
. by year, sort: replace resize = size - _N
(689 real changes made)
. by year: egen jeff2 = apport(pop), size(resize)
. replace jeff2 = jeff2 + 1
(689 real changes made)
. assert jeff2 == adam
```

Applying the second procedure by using the other methods would give even more seats to smaller states than would Adams's method. In the United States, the problem of allocations of zero seats emerged when Hamilton's method and Webster's method were used (1852–1911 and 1911–1941, respectively); during these periods, the first procedure was used to solve the problem. In 1942, the decision was made to use Hill's method, which by definition assigns one seat to each state.

4.4 Threshold and exceptions

We will now use the German election dataset to illustrate variations of apportionment methods. In Germany, the barrier clause has always been 5%. To reproduce the distribution of seats in the German *Bundestag*, we therefore specify `threshold(5)`. Using data for the elections of 1965, 1969, and 1972, we reproduce the distribution of seats as reported by Statistisches Bundesamt (2005, 54) and stored as variable `mandates` in the German election data.

```
. use electionsde if inrange(year(eldate),1965,1972), clear
. by eldate, sort: egen seats = apport(votes), size(size) method(dhondt)
> threshold(5)
(19 missing values generated)
. list eldate party mandates seats if mandates != ., sepby(eldate) noobs
```

eldate	party	mandates	seats
19 Sep 65	FDP	49	49
19 Sep 65	SPD	202	202
19 Sep 65	CDU/CSU	245	245
28 Sep 69	FDP	30	30
28 Sep 69	SPD	224	224
28 Sep 69	CDU/CSU	242	242
19 Nov 72	FDP	41	41
19 Nov 72	CDU/CSU	225	225
19 Nov 72	SPD	230	230

In Germany, the Jefferson method is known as the d'Hondt method; thus we used `method(dhondt)` for `method(jefferson)`.

One particularity of the German election system is that under certain conditions, a party participates in the allocation of seats even if it has not passed the election threshold if that party has won the majority of first votes in at least three constituencies. In 1994, for example, the Party of Democratic Socialism won 4.39% of valid second votes, but participated in the allocation of seats because it won four direct mandates. Exceptions from the barrier clause are specified by using an expression that identifies the respective observations. Here is an easy example of how to include the Party of Democratic Socialism in the allocation of seats when using the 1994 election results:

```
. use electionsde if year(eldate) == 1994, clear
. egen seats = apport(votes), size(656) method(hamilton) threshold(5)
> exception(party=="PDS")
Seat goes to stronger party; consider option -raffle-
(16 missing values generated)
. list eldate party mandates seats if mandates != ., noobs
```

eldate	party	mandates	seats
16 Oct 94	PDS	30	30
16 Oct 94	FDP	47	47
16 Oct 94	Gruene	49	49
16 Oct 94	SPD	252	248
16 Oct 94	CDU/CSU	294	282

In this last example, we do not fully reproduce the real distribution of seats because of a further particularity of the German electoral system by which the legal size of parliament may be increased because of so-called overhang seats. Overhang seats arise when the national proportion of votes for a party entitles the party to fewer seats than the number of constituencies it has won.[7] These overhang seats cannot be reproduced with the `apport()` function.

5 References

Balinski, M. L., and H. P. Young. 1982. *Fair Representation: Meeting the Ideal of One Man, One Vote*. New Haven, London: Yale University Press.

Kohler, U. 2011. Estimating the potential impact of nonvoters on outcomes of parliamentary elections in proportional systems with an application to German national elections from 1949 to 2009. *Electoral Studies* 30: 497–509.

Malkevitch, J. 2002a. Apportionment I. American Mathematical Society. http://www.ams.org/featurecolumn/archive/apportion1.html.

———. 2002b. Apportionment II. American Mathematical Society. http://www.ams.org/featurecolumn/archive/apportionII1.html.

Marshall, A. W., I. Olkin, and F. Pukelsheim. 2002. A majorization comparison of apportionment methods in proportional representation. *Social Choice and Welfare* 19: 885–900.

Nohlen, D., F. Grotz, M. Krennerich, and B. Thibaut. 2000. Appendix: Electoral systems in independent countries. In *International Encyclopedia of Elections*, ed. R. Rose, 353–379. Washington, DC: Congressional Quarterly.

Pukelsheim, F. 2000. Mandatszuteilungen bei Verhältniswahlen: Vertretungsgewichte der Mandate. *Kritische Vierteljahresschrift für Gesetzgebung und Rechtswissenschaft* 83: 76–103. http://www.uni-augsburg.de/pukelsheim/2000b.html.

———. 2003. Erfolgswertgleichheit der Wählerstimmen? *Stadtforschung und Statistik* 1/2003: 56–61. http://www.math.uni-augsburg.de/stochastik/pukelsheim/2003a.html.

Schuster, K., F. Pukelsheim, M. Drton, and N. R. Draper. 2003. Seat biases of apportionment methods for proportional representation. *Electoral Studies* 22: 651–676.

Schwingenschlögl, U., and F. Pukelsheim. 2006. Seat biases in proportional representation systems with thresholds. *Social Choice and Welfare* 27: 189–193.

Statistisches Bundesamt. 2005. *Wahl zum 16. Deutschen Bundestag am 18. September 2005. Heft 1: Ergebnisse und Vergleichszahlen früherer Bundestags-, Europa- und Landtagswahlen sowie Strukturdaten für die Bundestagswahlkreise*. Stuttgart: Metzler-Poeschel.

7. For further information, see http://www.wahlrecht.de/english/overhang.html.

About the authors

Ulrich Kohler is a sociologist at the Wissenschaftszentrum Berlin (Social Science Research Center) who has used Stata for several years. His research interests include social inequality and political sociology. With Frauke Kreuter, he is author of the textbook *Data Analysis Using Stata*.

Janina Zeh studied sociology and education at the University of Leipzig. She is a PhD student in a project on "Social Capital during the Life Course" at the University of Hamburg.