



Steekproef voor controletellingen bij Gemeenteraadsverkiezingen 2022

datum 17 maart 2022

samenvatting Deze notitie beschrijft het ontwerp en de trekking van de steekproef ten behoeve van controles van de software die wordt gebruikt bij de Gemeenteraadsverkiezingen van 2022.

trefwoorden Enkelvoudig aselechte steekproef, uniforme verdeling

1. Inleiding

Bij de Gemeenteraadsverkiezingen op 16 maart 2022 worden optellingen in de software OSV2020 steekproefsgewijs gecontroleerd. Deze software telt onder andere in elke gemeente per lijst de aantallen stemmen per stembureau op tot stemtotalen per lijst in de betreffende gemeente. In elke gemeente worden voor een steekproef van drie lijsten de stemtotalen per lijst nageteld en vergeleken met de uitkomsten van de software.

De Kiesraad heeft aan het Centraal Bureau voor de Statistiek (CBS) gevraagd om de steekproef ten behoeve van deze controles te trekken. Voor elke gemeente is een steekproef met omvang drie gewenst uit de lijsten die in de betreffende gemeente deelnemen aan de Gemeenteraadsverkiezingen van 2022. De steekproeven per gemeente moeten onafhankelijk van elkaar worden getrokken. Ten behoeve van deze steekproef heeft het CBS een bestand van de Kiesraad ontvangen met de 333 gemeenten waar op 16 maart 2022 verkiezingen plaatsvinden en de lijsten die in die gemeente deelnemen. Deze notitie bevat informatie over het toegepaste steekproefontwerp en het resultaat van de trekking volgens dat ontwerp.

2. Ontwerp en trekking van de steekproef

Het meest eenvoudige en meest bekende steekproefontwerp is het enkelvoudig aselechte steekproefontwerp. Dit ontwerp staat het dichtst bij wat intuïtief onder het trekken van een steekproef wordt verstaan. Ieder element heeft dezelfde kans om in de steekproef te komen en er wordt geen gebruik gemaakt van hulpinformatie. Het ontwerp kan zowel met als zonder terugleggen worden toegepast. In de praktijk hebben steekproeven zonder terugleggen de voorkeur omdat deze de meeste informatie geven.

Een voorbeeld van een enkelvoudig aselechte steekproef zonder terugleggen is de lottomachine. Een voor een worden balletjes getrokken, waarbij alle balletjes die in de machine zitten dezelfde kans hebben om te worden geselecteerd en getrokken balletjes worden niet in de machine teruggelegd.

De formele definitie van het enkelvoudig aselechte steekproefontwerp zonder terugleggen voor steekproeven met vaste omvang n uit een populatie is: elke deelverzameling van n verschillende elementen uit de populatie heeft dezelfde kans om als steekproef te worden gerealiseerd. Als N de populatieomvang is, dan heeft elk element uit de populatie kans n/N om te worden geselecteerd. De kans dat een vooraf aangewezen tweetal elementen in de steekproef terecht komt is $n(n-1)/(N(N-1))$.

Voor de steekproef van te controleren lijsten per gemeente is gekozen om in elke gemeente, onafhankelijk van elkaar, het enkelvoudig aselechte steekproefontwerp zonder terugleggen toe te passen. Per gemeente is volgens dit ontwerp een steekproef met omvang drie getrokken uit de deelnemende lijsten in de betreffende gemeente¹. Het aantal deelnemende lijsten per gemeente ziet er uit zoals weergegeven in tabel 1.

Tabel 1 Aantal lijsten Gemeenteraadsverkiezingen maart 2022

Aantal deelnemende lijsten in een gemeente	Aantal gemeenten met het genoemde aantal lijsten
3	3
4	7
5	29
6	49
7	61
8	57
9	28
10	28
11	20
12	13
13	4
14	11
15	7
16	6
17	3
18	2
20	2
21	1
23	1
26	1

Bij de trekking van de steekproef is softwarepakket R, versie 4.0.3, gebruikt. Door de standaardfunctie 'sample' aan te roepen met argumenten x = de vector met te kiezen lijsten in een gemeente, en $n = 3$, wordt een enkelvoudig aselechte steekproef zonder terugleggen getrokken met omvang 3 uit de te kiezen lijsten in de betreffende gemeente.

De sample functie maakt gebruik van een pseudorandomgenerator die bij benadering een uniform verdeelde output produceert. Ten behoeve van de reproduceerbaarheid van de output is gebruik gemaakt van een random seed. De getallenreeks van de pseudorandomgenerator wordt volledig bepaald door de keuze van de random seed. Als de pseudorandomgenerator opnieuw wordt geïnitieerd met dezelfde random seed, zal hij dezelfde getallenreeks produceren. De waarde van de random seed hoeft niet willekeurig te zijn. Vanwege de aard van het algoritme zullen de getallen

¹ Er zijn drie gemeenten waar slechts drie lijsten meedoen. De lijsten van deze gemeenten komen alle drie in de steekproef.

die het algoritme genereert de uniforme verdeling volgen, dit met uitzondering van het eerste getal. Door de random seed willekeurig te kiezen verdwijnt deze uitzondering. Dat is gedaan door de seconden, nauwkeurig tot op de milliseconde, van de op een willekeurig moment gekozen systeemtijd van de computer te vermenigvuldigen met duizend. Voor de reproduceerbaarheid is het ook van belang te melden dat de gemeentesteekproeven in volgorde van het invoerbestand van de Kiesraad zijn getrokken en geleverd.

Het resultaat van de steekproeftrekking is een steekproef van te controleren lijsten per gemeente. De steekproef is opgenomen in een bestand dat door het CBS is geleverd aan de Kiesraad. Dit bestand bevat vijf kolommen die gescheiden zijn met een puntkomma. De kolommen bevatten achtereenvolgens gemeentecode, gemeentenaam en de drie te controleren lijsten.