



Advies steekproefomvang herindelingsverkiezingen 2021 en gemeenteraadsverkiezingen 2022

1. Inleiding en vraagstelling

Per 1 januari 2022 zijn enkele gemeenten samengevoegd. Heerhugowaard en Langedijk worden Dijk en Waard; Uden en Landerd worden Maashorst; Boxmeer, Cuijk, Grave, Mill en Sint Hubert en Sint Anthonis worden Land van Cuijk; Purmerend en Beemster worden Purmerend. Zo ontstaan vier nieuwe gemeenten, waarvoor herindelingsverkiezingen zijn gehouden op woensdag 24 november 2021. In deze nieuwe gemeenten waren er respectievelijk 12, 13, 13, 15 lijsten bij de herindelingsverkiezingen.

Net als bij de Tweede Kamerverkiezingen op 17 maart 2021 werd er bij de herindelingsverkiezingen een controleprotocol gehanteerd dat ziet op de werking van de optellingen in de software OSV2020. Hetzelfde zal het geval zijn bij de gemeenteraadsverkiezingen van 16 maart 2022. De software telt in elke gemeente het aantal stemmen per lijst van alle stembureaus op tot een lijsttotaal van de gemeente. Als onderdeel van het controleprotocol wordt in elke gemeente op basis van een steekproef het gemeentelijk totaal van een aantal lijsten nageteld en vergeleken met de uitkomsten van de software.

De Kiesraad heeft aan het Centraal Bureau voor de Statistiek gevraagd om een advies ten aanzien van de betrouwbaarheidsuitspraken die met behulp van een steekproef van lijsten per gemeente kunnen worden verkregen, indien de controletellingen geen fouten aan het licht brengen, zowel ten aanzien van de herindelingsverkiezingen van 24 november 2021 als de gemeenteraadsverkiezingen van 16 maart 2022. Dit rapport beschrijft het gevraagde advies, en is een aanvulling op het [eerdere advies](#) dat is gegeven voor de Tweede Kamerverkiezingen van 17 maart 2021.

In tegenstelling tot bij de Tweede Kamerverkiezingen, wil de Kiesraad bij de herindelingsverkiezingen en de gemeenteraadsverkiezingen niet alleen weten welke betrouwbaarheidsuitspraken er mogelijk zijn op basis van een steekproef in de controletelling voor alle lijstoptellingen binnen Nederland, maar ook per gemeente. Dat wil zeggen dat naast een uitspraak over de waarschijnlijkheid van softwarematige optelfouten of manipulatie op *landelijk* niveau, de interesse ook uitgaat naar de waarschijnlijkheid van *lokale* invoerfouten of lokale manipulatie van het systeem waar de software draait.

2. Betrouwbaarheid voor kleine populaties

Beschouw een van de gemeenten. Laat N het aantal lijsten in deze gemeente zijn, en n de steekproefomvang voor controletelling. Noem m het aantal fouten dat wordt gevonden als alle lijsten zouden worden gecontroleerd. Deze parameter m is in de praktijk onbekend. We zoeken de grootste integer waarde m_0 voor m waarvoor het nog aannemelijk is dat bij een steekproef met omvang n nul fouten worden gevonden.

Om m_0 te kunnen berekenen moet worden afgesproken wat als 'aannemelijk' wordt beschouwd. Dat doen we met behulp van kansrekening. Definieer de kansvariabele k als het aantal fouten bij een



steekproef van n lijsten. De kans op nul fouten bij controle van een steekproef met omvang n wordt genoteerd als $P(\underline{k} = 0)$. Hoe groter deze kans, hoe aannemelijker het is dat bij een aselekt getrokken steekproef met omvang n daadwerkelijk nul fouten worden gevonden. We spreken af dat deze kans ten minste α moet zijn, waarbij de waarde van het significantieniveau α nog te bepalen is:

$$(1) \quad P(\underline{k} = 0) \geq \alpha.$$

De kans in deze formule kan worden uitgedrukt in termen van Binomiaalcoëfficiënten. Als bij m lijsten de telling door de software fout is, dan is bij $N - m$ lijsten de telling door de software goed. Het is dus alleen mogelijk om nul fouten te vinden in een steekproef met omvang n als $n \leq N - m$, en er geldt

$$(2) \quad P(\underline{k} = 0) = \frac{\binom{m}{0} \binom{N-m}{n}}{\binom{N}{n}} = \frac{(N-m)!}{(N-m-n)!} \cdot \frac{(N-n)!}{N!}.$$

Combinatie van (1) en (2) levert de ongelijkheid

$$(3) \quad \frac{(N-m)!}{(N-m-n)!} \cdot \frac{(N-n)!}{N!} \geq \alpha.$$

De gezochte m_0 is de grootste integer waarde van m waarvoor deze ongelijkheid waar is. Tabel 1 bevat de waarden van m_0 naar steekproefomvang n en populatieomvang N voor $\alpha \in \{0,05; 0,1; 0,2\}$. De populatieomvang N is hierin weergegeven voor de lijsten bij de herindelingsgemeenten en voor een hypothetische gemeente met 25 lijsten.

Tabel 1 Waarden van m_0 naar steekproefomvang n en populatieomvang N

N	12			13			15			25		
α	0,05	0,1	0,2	0,05	0,1	0,2	0,05	0,1	0,2	0,05	0,1	0,2
n												
3	6	5	4	7	6	4	8	7	5	15	12	9
4	5	4	3	5	4	3	7	5	4	12	10	7
5	4	3	2	4	3	2	5	4	3	10	8	6
6	3	2	2	3	3	2	4	3	2	8	7	5

Uit Tabel 1 volgt bijvoorbeeld dat bij een (nieuwe) gemeente met $N = 12$ lijsten, significantieniveau $\alpha = 0,2$ en steekproefomvang $n = 3$ de waarde van m_0 gelijk is aan 4, zie gearceerde cel. Dit betekent dat als bij controle van een steekproef van omvang 3 in een gemeenten met 12 lijsten blijkt dat nul fouten worden gevonden, dat het toch nog aannemelijk is dat er van de 12 lijsten 4 foutief zijn geteld door de software. Iets preciezer, als in werkelijkheid 4 lijsten foutief zijn geteld door de software, dan volgt uit formule (2) dat de kans op nul fouten bij een steekproef van 3 uit 12 lijsten gelijk is aan 0,2546.



3. Vergelijking met de betrouwbaarheidsuitspraken Tweede Kamerverkiezingen

Uit het voorgaande wordt duidelijk dat met kleine steekproeven op het niveau van gemeenten sprake is van grote betrouwbaarheidsintervallen. Bij de Tweede Kamerverkiezingen van 17 maart 2021 is een steekproef van 1.065 lijsten getrokken: in elk van de 355 gemeenten 3 lijsten. Dit leidde met nul gevonden fouten bij $\alpha = 0,05$ tot een aannemelijkheid van een proportie fouten van minder dan 0,0028 over alle uitslagen. Het verschil is dat de vraagstelling bij de Tweede Kamerverkiezingen zag op de waarschijnlijkheid van softwarematige fouten in de hele populatie van lijstopstellingen in Nederland.

Het is daarbij van belang om te realiseren dat betrouwbaarheidsuitspraken vanuit een steekproef voornamelijk worden bepaald door de omvang van de steekproef, en niet door de omvang van de populatie waaruit de steekproef wordt getrokken. Als men voor de herindelingsverkiezingen zou zien op een uitspraak over het totaal van de 53 lijstopstellingen, dan zouden 41 lijsten gecontroleerd moeten worden om met 95% betrouwbaarheid te kunnen concluderen dat bij nul foute optellingen in de steekproef, maximaal 2 van de 53 lijsten fout zijn opgeteld. Die laatste verhouding vertegenwoordigt nog steeds een duidelijk grotere mate van onzekerheid dan het landelijke betrouwbaarheidsinterval bij de Tweede Kamerverkiezingen. Indien men ook in kleine populaties een uitspraak wil doen met een betrouwbaarheidsinterval dat even klein is als bij de Tweede Kamerverkiezingen, dan resteert geen andere mogelijkheid dan het controleren van (vrijwel) alle lijstopstellingen.

4. Conclusie en advies

Vooruitlopend op de gemeenteraadsverkiezingen van 16 maart 2022, geldt dat een steekproef van 3 lijsten per gemeente ter controle bij die verkiezingen, tot een vergelijkbaar betrouwbaarheidsinterval leidt als bij de Tweede Kamerverkiezingen¹. Dat wil zeggen: Als men een uitspraak wil doen over de waarschijnlijkheid van softwarematige optelfouten of manipulatie op *landelijk* niveau. Het betrouwbaarheidsinterval is onafhankelijk van de aantallen lijsten per gemeente die, anders dan bij de Tweede Kamerverkiezingen, bij gemeenteraadsverkiezingen sterk kunnen variëren.

Indien men bij de gemeenteraadsverkiezingen een uitspraak wil over de waarschijnlijkheid van *lokale* invoerfouten of manipulatie én die uitspraak wil doen met betrouwbaarheidsintervallen die niet groter zijn dan die op landelijk niveau, dan is dat niet mogelijk met een beperkte steekproef. En dient men praktisch gesproken (vrijwel) alle lijstopstellingen in een betreffende gemeente te controleren. Indien men voor die lokale uitspraak een (veel) grotere mate van onzekerheid acceptabel vindt, dan kan men een keuze maken voor een bepaald significantieniveau α en een bepaalde steekproefomvang n met resultaten zoals uiteengezet in paragraaf 2.

¹ Het aantal deelnemende gemeenten kan iets lager liggen dan bij de Tweede Kamerverkiezingen en daarmee kan de totale steekproef (n) dus ook iets kleiner zijn. Zie de 'rule of three' regels uit het eerdere advies voor de exacte betrouwbaarheidsintervallen bij een bepaalde n .