



## Advies steekproefomvang controletellingen verkiezingsuitslagen

### 1. Inleiding en vraagstelling

Naar verwachting zal bij de Tweede Kamerverkiezingen op 17 maart 2021 een controleprotocol worden gehanteerd dat ziet op de werking van de optellingen in de software OSV2020. Deze software telt onder andere 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 de overwogen steekproef van drie, vijf of zeven lijsten per gemeente kunnen worden verkregen, indien de controletellingen geen fouten aan het licht brengen. Dit rapport beschrijft het gevraagde advies. Het uitgangspunt is dat er sprake is van 355 gemeenten en ongeveer 30 lijsten waarop gestemd kan worden.

### 2. Steekproefomvang en betrouwbaarheid

Zoals in de inleiding aangegeven, wordt er een aantal tellingen gecontroleerd waarbij het resultaat in principe steeds goed of fout kan zijn. We gaan ervan uit dat er bij deze controles geen fouten worden gevonden. Omdat er slechts een steekproef van de uitslagen gecontroleerd wordt, betekent dit niet dat er geen fouten voorkomen onder alle optellingen. Daarom willen we op grond van het aantal gecontroleerde uitslagen iets zeggen over de aannemelijke waarden van de proportie fouten in de hele populatie van uitslagen. In het bijzonder willen we weten hoe groot het aantal foutloze controles moet zijn (de steekproefgrootte) om te concluderen dat de proportie fouten in de populatie klein is.

Een antwoord op deze vraag vanuit statistisch perspectief is te formuleren in termen van een betrouwbaarheidsinterval voor de waarden van de onbekende populatieproportie, zeg  $P$ . Dit interval bevat de waarden van  $P$  die betrouwbaar zijn in de zin dat het voor die waarden waarschijnlijk is om een steekproef zonder fouten aan te treffen. Dit zijn de kleine waarden van de proportie  $P$ , bij grotere waarden van  $P$  zal het steeds onwaarschijnlijker worden om steekproeven zonder fouten aan te treffen. Voor waarden van  $P$  buiten het betrouwbaarheidsinterval is de kans klein dat er in een steekproef geen fouten voorkomen. Bij een 95% betrouwbaarheidsinterval is de kans op een steekproef zonder fouten voor populatieproporties *buiten het interval* kleiner dan 5% (en analoog voor andere percentages). Waarden voor  $P$  buiten het interval verwerpen we in dat geval als niet plausibel. Het betrouwbaarheidsinterval voor  $P$  loopt van 0 tot  $p_0$ , waarbij  $p_0$  de bovengrens is.

Er bestaat een eenvoudige methode om een betrouwbaarheidsinterval te berekenen voor de situatie dat we een steekproef hebben met  $n$  twee-waardige uitkomsten, zoals hier: *fout vs niet-fout*, en we



treffen in de steekproef geen enkele keer de uitkomst fout aan. Die methode staat bekend als de "Rule of three". Deze regel zegt dat we 95%, 97% en 99% betrouwbaarheidsintervallen voor P kunnen vinden door de bovengrens voor het interval,  $p_0$  te bepalen zoals weergegeven in tabel 1. De methode resulteert in een conservatief interval waardoor de exacte bovengrens iets lager zal zijn, maar voor steekproeven met  $n$  groter dan 100 is het verschil verwaarloosbaar.

*Tabel 1. Bovengrenzen voor betrouwbaarheidsintervallen voor de populatie proportie fouten op basis van een steekproef ter grootte van  $n$  waarin geen fouten zijn aangetroffen*

<b>Interval</b>	<b>bovengrens</b>
95%	3/n
97%	3,51/n
99%	4,61/n

De ondergrens van het betrouwbaarheidsinterval is altijd 0 en, zoals te zien in tabel 1, hangt de bovengrens af van de steekproefomvang. Naarmate de steekproef groter wordt, en we zien nog steeds geen fouten, wordt het betrouwbaarheidsinterval kleiner. De range van plausible populatieproporties van foute uitkomsten wordt dus steeds verder beperkt als de steekproefomvang toeneemt.

In tabel 2 zijn de grenzen toegepast op een drietal steekproefomvangen die corresponderen met de aantallen controletellingen die worden overwogen.

*Tabel 2. Bovengrenzen voor betrouwbaarheidsintervallen voor de populatieproportie fouten op basis van drie verschillende steekproefomvangen.*

<b>Steekproefomvang <math>n</math></b>	<b>Interval</b>		
	95%	97%	99%
1065	0,0028	0,0033	0,0043
1775	0,0017	0,0020	0,0026
2485	0,0012	0,0014	0,0019

### 3. Conclusie

Op basis van tabel 2 kan het volgende gesteld worden ten aanzien van de overwogen aantallen controletellingen:

- Bij drie controletellingen per gemeente, dus een steekproefomvang van 1065 en nul gevonden fouten, ligt het 99% betrouwbaarheidsinterval voor de proportie foute uitslagen in de hele populatie van tellingen tussen de 0 en 0,43%. Het is dan, vrij verwoord, 99% zeker dat minder dan 0,43% van alle tellingen fout kan zijn. Het 95% betrouwbaarheidsinterval ligt tussen de 0 en 0,28%.
- Bij vijf controletellingen per gemeente, dus een steekproefomvang van 1775 en nul gevonden fouten, ligt het 99% betrouwbaarheidsinterval voor de proportie foute uitslagen in de hele populatie van tellingen tussen de 0 en 0,26%. Het is dan, vrij verwoord, 99% zeker dat minder dan 0,26% van alle tellingen fout kan zijn. Het 95% betrouwbaarheidsinterval ligt tussen de 0 en 0,17%.



- Bij zeven controletellingen per gemeente, dus een steekproefomvang van 2485 en nul gevonden fouten, ligt het 99% betrouwbaarheidsinterval voor de proportie foute uitslagen in de hele populatie van tellingen tussen de 0 en 0,19%. Het is dan, vrij verwoord, 99% zeker dat minder dan 0,19% van alle tellingen fout kan zijn. Het 95% betrouwbaarheidsinterval ligt tussen de 0 en 0,12%.