



Getty Images (z.d.)

## De invloed van betrokkenheid bij de gemeenschap op het goed voorspellen van de verkiezingsuitslag

Bachelor werkstuk Sociologie

*“In hoeverre kunnen mensen die in een hogere mate betrokken zijn bij de gemeenschap beter de winnaar van de Amerikaanse verkiezingen voorspellen dan mensen die nauwelijks betrokken zijn bij de gemeenschap?”*

Anna Reusken (S4109473)

BAW Sociologie

Begeleider: Vincenz Frey

Datum: 08-06-2022

## Abstract

Verwachtingen over de uitkomst van verkiezingen kunnen een invloed hebben op het stemgedrag en zelfs op de overweging van individuen om te gaan stemmen. Dit werkstuk beschrijft de resultaten van een kwantitatief empirisch onderzoek naar factoren die mogelijk een invloed uitoefenen op de accuraatheid van voorspellingen over de winnaar van de Amerikaanse verkiezingen. Er wordt gekeken naar het effect van betrokkenheid bij de gemeenschap op het maken van een goede voorspelling. De betrokkenheid van individuen bij hun gemeenschap is namelijk dalende, terwijl een grotere betrokkenheid van individuen bij hun gemeenschap vaak positieve gevolgen heeft. Zodoende zou er ook verwacht kunnen worden dat een grotere betrokkenheid van een individu bij de gemeenschap ertoe leidt dat het individu beter weet hoe andere mensen op politiek gebied denken en daardoor beter kan voorspellen wie de winnaar van de verkiezingen zal zijn. De analyses worden uitgevoerd op data die verzameld is door de *American National Election Studies* (ANES). De steekproef bestaat uit 399 respondenten. De onderzoeksresultaten tonen aan dat de individuen die zich meer betrekken bij de gemeenschap niet beter of slechter zijn in het voorspellen van de winnaar van de verkiezingen dan individuen die minder betrokken zijn bij de gemeenschap.

# Inhoudsopgave

<b>1</b>	<b>Inleiding</b> .....	4
<b>2</b>	<b>Theorie</b> .....	7
2.1	Betrokkenheid bij de gemeenschap en het voorspellen van de winnaar .....	7
2.2	Discussiëren over politiek.....	8
2.3	Identificatie met een partij.....	9
2.4	Interesse in de verkiezingen, opleidingsniveau, leeftijd en geslacht .....	10
<b>3</b>	<b>Methoden</b> .....	12
3.1	Beschrijving databestand .....	12
3.2	Beschrijving onderzoeksdesign .....	13
3.3	Beschrijving operationalisaties.....	13
3.4	Analyse-opzet.....	16
<b>4</b>	<b>Resultaten</b> .....	18
4.1	Beschrijvende statistieken.....	18
4.2	Modevaluatie.....	20
4.3	Hypothesetoetsing .....	24
<b>5</b>	<b>Conclusie en discussie</b> .....	26
	<b>Literatuurlijst</b> .....	29
	<b>Bijlage 1</b> .....	33
1.1	Variabele Goed voorspellen van de winnaar: goed_voorspeld .....	33
1.2	Variabele Betrokkenheid bij de gemeenschap: betrokken_gem .....	39
1.3	Variabele Discussiëren over politiek: discuss_pol.....	44
1.4	Variabelen Identificatie met een partij: ident_W en ident_L .....	47
1.5	Controlevariabele Interesse in de verkiezingen: interesse_verk .....	53
1.6	Controlevariabele Opleidingsniveau: opleidingsniveau .....	56
1.7	Controlevariabele Leeftijd: leeftijd.....	59
1.8	Controlevariabele Geslacht: vrouw .....	62
	<b>Bijlage 2</b> .....	64
2.1	Analyse van de beschrijvende statistieken (tabel 1) .....	64
2.2	Associatiematen (tabel 2).....	65
2.3	Regressieanalyses (tabel 3 en 4) .....	76
	<b>Bijlage 3</b> .....	88
3.1	Assumptietoetsing.....	88
3.2	Outliers .....	88
3.3	Multicollineariteit.....	92

# 1 Inleiding

Wanneer men het stemgedrag in een samenleving wil kunnen verklaren, is het belangrijk om te begrijpen welke factoren invloed hebben op het vermogen van individuen om verkiezingsuitslagen te voorspellen. Verwachtingen over de uitkomst van verkiezingen kunnen namelijk invloed hebben op het stemgedrag of zelfs op de overweging om daadwerkelijk te gaan stemmen (Delavande & Manski, 2012). Daarnaast kan een discrepantie tussen individuele verwachtingen en de uiteindelijke verkiezingsuitslag ertoe leiden dat de tevredenheid van individuen met de democratie afneemt en er negatieve gevoelens over de werking van de politiek worden gevormd (Mongrain, 2021). Het vertrouwen van Amerikanen in de federale overheid is de afgelopen jaren dalende en op dit moment historisch laag (Brenan, 2021; Rainie & Perrin, 2020). Voormalig president en verliezend Republikeins presidentskandidaat van 2020 Donald Trump stelt daarnaast zelf dat de verkiezingen in 2020 waren “gestolen” en de verkiezingsuitslag dus onjuist is. Deze beweringen leidden er in 2021 zelfs toe dat een woedende menigte het Amerikaanse Capitool in Washington bestormde (NOS, 2021). Kortom, een tegenstrijdigheid tussen de verwachtingen van een deel van het electoraat met betrekking tot de verkiezingsuitslag en de daadwerkelijke verkiezingsuitslag heeft recent grote gevolgen gehad.

Vanwege bovengenoemde redenen is het relevant om onderzoek te doen naar de accuraatheid van verwachtingen over de verkiezingen en de factoren die daar invloed op hebben. Over het algemeen kunnen mensen vaak een goede inschatting maken van wie de winnende kandidaat zal zijn (Lewis-Beck & Tien, 1999). Met betrekking tot voorspellingen over de verkiezingsuitslag is het bekend dat het bezitten van veel politieke kennis mensen kan helpen om accurate voorspellingen te vormen (Mongrain, 2021). Echter, de uiteindelijke verwachtingen van een individu over de verkiezingen kunnen ook gekleurd worden door individuele voorkeuren van kiezers met betrekking tot een coalitie, partij of kandidaat (Mongrain, 2021). Verschillende karakteristieken van de sociale netwerken waarin mensen ingebed zijn, hebben ook een invloed op de verwachtingen die worden gevormd (Leiter et al., 2018).

Daarnaast toont Putnam (1995) aan dat de sociale betrokkenheid van mensen steeds verder afneemt de laatste jaren. Dit brengt verschillende gevolgen met zich mee. Zo neemt onder andere het sociaal kapitaal in de samenleving af: er zijn minder vormen van sociale organisatie zoals netwerken, normen en sociaal vertrouwen die het samenwerken onder mensen bevorderen. Dit lijkt nadelig, aangezien een grotere sociale betrokkenheid van individuen vaak positieve gevolgen blijkt te hebben (Ramachandra & Naha Abu Mansor, 2014; Attree et al., 2010). Dit maakt het interessant om te onderzoeken of een sterkere betrokkenheid bij de gemeenschap ertoe zou kunnen leiden dat mensen accuratere voorspellingen maken over de winnaar van de verkiezingen. Het hebben van discussies met andere mensen over politieke kwesties maakt het vormen van voorspellingen wellicht gemakkelijker (Leiter et al., 2018). Eerder werd al genoemd dat individuele voorkeuren van een

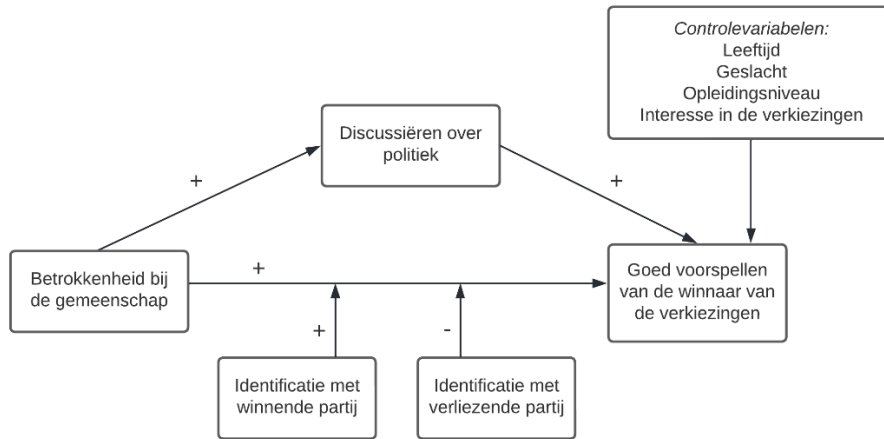
kiezer de verwachtingen van dit individu over de verkiezingsuitslag kan kleuren. Heeft een verbondenheid met een specifieke partijideologie wellicht een invloed op de sterkte van het verband tussen betrokken zijn bij de gemeenschap en beter de verkiezingsuitslag kunnen voorspellen? Deze speculaties leiden tot de volgende onderzoeksvraag en sub-vragen: *“In hoeverre kunnen mensen die in een hogere mate betrokken zijn bij de gemeenschap beter de winnaar van de Amerikaanse verkiezingen voorspellen dan mensen die nauwelijks betrokken zijn bij de gemeenschap? In hoeverre verloopt dit verband via het regelmatig discussiëren over politiek, en in hoeverre blijft dit verband bestaan wanneer mensen zich sterk identificeren met een politieke partij?”*.

‘Betrokkenheid bij de gemeenschap’ wordt hierbij gedefinieerd als het samenwerken met anderen uit de gemeenschap om kwesties van gemeenschappelijk belang te bespreken en aan te pakken (Christens & Speer, 2015). Een grotere mate van betrokkenheid onder de inwoners van de gemeenschap leidt tot regionale en maatschappelijke vooruitgang (Ramachandra & Naha Abu Mansor, 2014). Betrokken zijn bij de gemeenschap blijkt namelijk een effectieve wijze om veranderingen in beleid en systemen te bereiken op lokaal, regionaal en zelfs nationaal niveau (Christens & Speer, 2015). Gemeenschappen waarin mensen veel participeren worden daarnaast door sociale wetenschappers en politici beschouwd als “betere”, “democratischere” en “rechtvaardigere” gemeenschappen (Graycar, 1977). Individuen die zichzelf veel betrekken bij de gemeenschap profiteren hier zelf ook van doordat hun fysieke en psychologische gezondheid verbetert, hun zelfvertrouwen en gevoel van persoonlijke kracht stijgt en ze meer sociale relaties aangaan (Attree et al., 2010).

Er is sprake van goede verwachtingen van de verkiezingswinnaar wanneer de juiste winnende kandidaat wordt voorspeld; alle andere voorspelde kandidaten of uitkomsten zijn niet goed (Mongrain, 2021). Accurate voorspellingen over de verkiezingswinnaars zijn vaak het gevolg van partijpolitieke voorkeuren: individuen die een sterke voorkeur hebben voor de uiteindelijke winnaar kunnen, als gevolg van partijdige oordelen, betere voorspellingen maken dan individuen die een voorkeur hebben voor de uiteindelijke verliezer (Mongrain, 2021). Politieke kennis en aandacht voor het nieuws kunnen daarnaast ook de accuraatheid van voorspellingen verklaren.

De sterkte van identificatie met een politieke partij houdt de mate in waarin het individu zichzelf beschouwd als lid van een (politieke) groep (Brown & Hohman, 2022; Haas et al., 2019). Deze identificatie gaat vaak samen met het overnemen van bijbehorende ideologische waarden zodat een sterkere politieke identificatie leidt tot het sterker aanhouden van bepaalde ideologische waarden waarmee een groep individuen zichzelf kan onderscheiden van een andere (politieke) groep (Mason, 2014). Discussiëren over politiek houdt in dat een individu met een discussiepartner praat over politieke onderwerpen op basis waarvan het individu vervolgens zijn of haar politieke opinie vormt (Campus, 2012). De onderzoeksvraag zal middels een kwantitatief empirisch onderzoek beantwoord

worden. In dit onderzoek zal gebruik worden gemaakt van data uit de American National Election Studies (ANES). Deze data betreft vergaarde informatie over het stemgedrag van Amerikaanse burgers en de publieke opinie omtrent de presidentiële verkiezingen in de Verenigde Staten.



*Figuur 1: Grafische weergave van het onderzoeksmodel met 'Goed voorspellen van de winnaar van de verkiezingen' als afhankelijke variabele*

## 2 Theorie

### 2.1 Betrokkenheid bij de gemeenschap en het voorspellen van de winnaar

Het goed voorspellen van de winnaar van de verkiezingen houdt in dat een individu in staat is om voorafgaand aan de verkiezingen de juiste winnende kandidaat of partij aan te wijzen (Mongrain, 2021). Er is in aanloop naar de verkiezingen veel algemene kennis beschikbaar – in de vorm van polls – over de stemintenties van het electoraat en dus de vooruitzichten van de kandidaten in de verkiezingen, maar dit betekent niet dat ieder individu dezelfde verwachtingen heeft over de uiteindelijke verkiezingsuitslag (Delavande & Manski, 2012). Deze individuen maken namelijk hun eigen inschattingen die, naast polls en andere berichtgeving in de media, gebaseerd kunnen zijn op hun eigen voorkeuren, voorgaande ervaringen, de huidige campagnes en de sociale context van het individu (Leiter et al., 2020). De sociale context wordt mede-gevormd door (de interacties die plaatsvinden in) het sociale netwerk van een individu.

Dit zou kunnen betekenen dat het maken van voorspellingen over de verkiezingswinnaar wordt beïnvloed door de sterkte van betrokkenheid die een individu heeft bij zijn of haar gemeenschap. Betrokken zijn bij de gemeenschap betekent immers dat de bewoners van een gemeenschap samenwerken om kwesties die van gemeenschappelijk belang zijn te bespreken en actie te ondernemen om de kwestie te behandelen (Christens & Speer, 2015). Dit betekent dat, door te participeren in de gemeenschap, mensen interacteren en kennis delen met andere bewoners uit de gemeenschap (Speer et al., 2021). Het sociale netwerk van een individu breidt zich uit door deze sociale contacten.

Mensen baseren hun verwachtingen over verkiezingsuitslagen vervolgens onder andere op informatie die zij vanuit hun sociale netwerk ontvangen (Leiter et al., 2018). Naast dat het individu leert van voorgaande verkiezingsjaren, persoonlijke ervaring en de media, bieden netwerken dus informatie op basis waarvan een individu verwachtingen vormt (Meffert et al., 2011; Lewis-Beck & Tien, 1999). Hoe groter het sociale netwerk van het individu, hoe groter de kans dat het individu informatie ontvangt die representatief is voor de samenleving (Banerjee & Fudenberg, 2004). Door interacties aan te gaan met (nog onbekende) buurtgenoten, wordt het sociale netwerk meer divers en daaruit volgt contact waarvan een individu ook kan leren (Stevenson et al., 2020). Door dit diverse sociale netwerk zullen burgers hun verwachtingen bijvoorbeeld kunnen bijstellen wanneer zij in hun sociale netwerk observeren hoe andere leden zich gedragen en denken over politieke kwesties (Leiter et al., 2018). Een groter sociaal netwerk als gevolg van een grotere betrokkenheid bij de gemeenschap vergroot dan de kans dat het individu kennismaakt met meer diverse overtuigingen. Hierdoor kan (deels) worden voorkomen dat er onjuiste verwachtingen worden gevormd op basis van het geloof dat andere burgers dezelfde overtuigingen hebben met betrekking tot politieke

kwesties (Mongrain, 2021). In tegenstelling tot mensen die geïsoleerd zijn van de gemeenschap en weinig contacten hebben, worden mensen die zich centraler in de gemeenschap begeven dus vaker geconfronteerd met andere verwachtingen (Leiter et al., 2018). Mensen die ingebed zijn in grotere sociale netwerken, als gevolg van hun grotere betrokkenheid bij de gemeenschap, kunnen daardoor accuratere verwachtingen vormen over de verkiezingswinnaar.

Hieruit volgt hypothese 1: *Naarmate mensen meer betrokken zijn bij de gemeenschap kunnen ze beter de winnaar van de verkiezingen voorspellen.*

## 2.2 Discussiëren over politiek

Zoals eerder genoemd hebben mensen die meer betrokken zijn bij de gemeenschap meer sociale interacties en daardoor een groter sociaal netwerk (Speer et al., 2021). Een divers sociaal netwerk biedt mensen de kans om te uiten dat zij het oneens zijn met bepaalde informatie. Diverse sociale netwerken zorgen er namelijk voor dat mensen in aanraking komen met verschillende standpunten en hierdoor open staan voor nieuwe standpunten (McClurg, 2006). Verschillende leden uit het netwerk kunnen met elkaar discussiëren over deze verschillende standpunten en daar vervolgens van leren (McClurg, 2006; Leiter et al., 2018). Het leren van elkaars standpunten kan er vervolgens toe leiden dat mensen beter op de hoogte zijn van hoe andere mensen in de maatschappij denken op politiek gebied waardoor er juiste verwachtingen gevormd kunnen worden over de verkiezingsuitslag.

Er zou echter ook nog een andere verklaring kunnen zijn voor de positieve invloed die discussiëren over politiek heeft op het vormen van correcte verwachtingen over de verkiezingsuitslagen. Mensen die vaker discussiëren over politieke onderwerpen bezitten namelijk ook meer politieke kennis doordat zij politiek nieuws dat door massamedia naar buiten wordt gebracht beter kunnen ontvangen en verwerken (Scheufele, 2002). Ze zijn dan immers al bekend met het nieuws doordat zij over soortgelijke kwesties hebben gediscussieerd met hun sociale contacten. Hoe groter het sociale netwerk van een individu, hoe hoger het niveau van zijn politieke kennis, waardoor er vervolgens betere voorspellingen kunnen worden gevormd over de verkiezingen (Kwak et al., 2005).

Uit voorgaande redeneringen volgt hypothese 2: *Het positieve effect van betrokken zijn bij de gemeenschap op het correct voorspellen van de verkiezingsuitslag wordt gedeeltelijk verklaard doordat mensen die meer betrokken zijn bij de gemeenschap vaker over politiek discussiëren.*



## 2.3 Identificatie met een partij

Mensen die zichzelf sterk identificeren met een politieke partij beschouwen zichzelf in grote mate als lid van een (politieke) groep (Brown & Hohman, 2022; Haas et al., 2019). Er zijn twee wijzen waarop een sterke identificatie met een politieke partij gevolgen kan hebben voor de mate waarin een individu betrokken is bij zijn of haar gemeenschap en de betreffende verkiezingskandidaat die het individu verwacht te winnen. Ten eerste kunnen mensen die zich sterk identificeren met een politieke partij zich soms simpelweg niet bewust zijn van de meningsverschillen die zij hebben met anderen. Amerikanen blijken vooral hun politieke opinie achter te houden voor mensen met wie zij sterk van mening verschillen om op deze wijze mogelijke conflicten te vermijden (Cowan & Baldassarri, 2017). Dit wordt de *selective disclosure of attitudes* genoemd. Mensen die zich sterk identificeren met een politieke partij zullen hierdoor een sociale omgeving ervaren die, met betrekking tot (politieke) opvattingen, homogeen is (Cowan & Baldassarri, 2017). Wanneer zij zichzelf betrekken bij de gemeenschap en interacties aangaan met buurtgenoten zullen zij niet geconfronteerd worden met afwijkende meningen. Een grotere betrokkenheid bij de gemeenschap zal voor deze mensen betekenen dat zij enkel meer in aanraking komen met politieke visies die gelijk zijn aan die van henzelf. Individuen kunnen hierdoor onder andere ten onrechte denken dat iedereen zal denken en handelen zoals zij – het *false consensus effect* (Delavande & Manski, 2012). Voor hen die zich sterk identificeren met de verliezende partij heeft dit tot gevolg dat zij ten onrechte denken dat hun partij veel aanhangers heeft en een goede kans maakt op de verkiezingswinst (Mongrain, 2021). Voor de mensen die zich identificeren met de winnende partij zou dit gunstigere gevolgen hebben: zij blijven overtuigd van het winnen van de door hun geprefereerde verkiezingskandidaat en hebben uiteindelijk juiste verwachtingen gehad.

Ten tweede blijkt vanuit de homofilie-theorie dat mensen met sterke politieke overtuigingen zich vaak omgeven in gemeenschappen met gelijkgestemde mensen (Baccara & Yariv, 2012; Goel et al., 2010). Sterk partijdige mensen besteden hun tijd graag in gezelschap van andere leden van de politieke partij of politieke groep (Mason, 2015). Politiek gelijkdenkenden groeperen zich dus met elkaar en bevinden zich hierdoor vaker in zogenoemde “echo chambers” (Leiter et al., 2020; Goel et al., 2010). Doordat politiek andersdenkenden fysiek van elkaar gescheiden zijn, betekent een grotere betrokkenheid bij de gemeenschap dat er meer interacties plaatsvinden met mensen met dezelfde overtuigingen waardoor eigen overtuigingen worden bekrachtigd (Goel et al., 2010). De overtuigingen van een individu in een dergelijke *echo chamber* worden op deze manier niet in twijfel getrokken, en de meningen en opinies die het individu wel hoort zijn verre van representatief voor het electoraat in de samenleving (Leiter et al., 2020). Een mogelijk gevolg hiervan kan wederom het optreden van het *false consensus effect* zijn (Delavande & Manski, 2012). Mensen die zich

identificeren met de politieke partij die de verkiezingen uiteindelijk wint, hebben hierdoor de juiste verwachtingen gehad over de verkiezingswinnaar. Individuen die zich sterk identificeren met de verliezende partij zullen ten onrechte positieve verwachtingen hebben gehad van hun eigen partij en hebben hierdoor waarschijnlijk de verkeerde verkiezingswinnaar voorspeld.

Uit voorgaande redeneringen volgt hypothese 3: *Het positieve effect van betrokken zijn bij de gemeenschap op het correct voorspellen van de verkiezingsuitslag is sterker naarmate mensen zich sterker identificeren met de winnende politieke partij.* Tevens volgt uit de redeneringen hypothese 4: *Het positieve effect van betrokken zijn bij de gemeenschap op het correct voorspellen van de verkiezingsuitslag is minder sterk naarmate mensen zich sterker identificeren met de verliezende politieke partij.*

## 2.4 Interesse in de verkiezingen, opleidingsniveau, leeftijd en geslacht

Er zijn een aantal concepten die mogelijk een invloed kunnen hebben op de mate van betrokkenheid die een individu heeft bij de gemeenschap of op de mate waarin een individu de verkiezingswinnaar goed kan voorspellen. Bij het onderzoeken van bovenstaande verwachte verbanden in de statistische analyses zal er voor deze concepten gecontroleerd moeten worden. Allereerst wordt er gecontroleerd voor de mate van interesse in de verkiezingen vanwege het aanvullende effect dat de mate van interesse in de verkiezingen mogelijk heeft op het goed voorspellen van de verkiezingswinnaar. Een mogelijke verklaring voor het goed kunnen voorspellen van de verkiezingswinnaar betreft het hebben van een grote betrokkenheid bij de politiek (Lewis-Beck & Tien, 1999). Zo kan er van mensen die de politiek (uit interesse) nauw volgen via het nieuws verwacht worden dat zij op de hoogte zijn van de strijd in aanloop naar de verkiezingen en dat zij op basis van deze kennis een goed doordachte voorspelling kunnen maken (Lewis-Beck & Tien, 1999; Leiter et al., 2020). De verwachting is daarom dat kiezers die zeer geïnteresseerd zijn in de verkiezingen, betere voorspellingen kunnen maken met betrekking tot de verkiezingsuitslagen. Om te controleren voor dit aanvullende effect wordt de variabele ‘interesse in de verkiezingen’ in de regressiemodellen opgenomen als controlevariabele.

De tweede variabele waarvoor wordt gecontroleerd is opleidingsniveau. Opleidingsniveau zou een schijnverband tussen ‘betrokkenheid bij de gemeenschap’ en ‘goed voorspellen van de verkiezingswinnaar’ kunnen blootleggen. Mensen die hoger opgeleid zijn, zijn namelijk vaker betrokken bij de gemeenschap (Speer et al., 2021). Hoger opgeleiden hebben daarnaast vaker accurate verwachtingen over verkiezingsuitkomsten dan lager opgeleiden (Mongrain, 2021; Lewis-Beck & Tien, 1999). Hoger opgeleiden hebben ten eerste namelijk vaak een omvangrijker netwerk

waaruit zij informatie ontvangen, maar bezitten daarnaast ook vaak vaardigheden die hen helpen in het omzetten van deze informatie in juiste verwachtingen (Beck and Tien, 1999). Vanwege het effect dat opleidingsniveau heeft op zowel betrokkenheid bij de gemeenschap als het vormen van juiste verwachtingen over de verkiezingswinnaar wordt er gecontroleerd voor opleidingsniveau.

De derde variabele waarvoor wordt gecontroleerd is leeftijd. Mensen met een hogere leeftijd zijn namelijk meer betrokken bij de gemeenschap (Speer et al., 2021). Tot slot wordt er gecontroleerd voor geslacht. Geslacht kan namelijk invloed hebben op de mate van betrokkenheid bij de gemeenschap (Speer et al., 2021). Evenals leeftijd wordt ook geslacht vaak meegenomen in voorgaande onderzoeken als kenmerk dat mogelijk een effect kan hebben op het vermogen om verkiezingsuitslagen te voorspellen (Mongrain, 2021). Om deze reden wordt dit in dit onderzoek ook gedaan.

## 3 Methoden

### 3.1 Beschrijving databestand

In dit onderzoek wordt gebruik gemaakt van een databestand van de *American National Election Study* (American National Election Study, 2021) van de jaren 2004, 2008 en 2012. De ANES is een onderzoek dat sinds 1948 in alle verkiezingsjaren wordt uitgevoerd om het Amerikaanse electoraat te kunnen beschrijven en hypothesen te kunnen toetsen betreffende het stemgedrag en de publieke opinie omtrent de presidentiële verkiezingen in de Verenigde Staten. Het onderzoek wordt uitgevoerd onder Amerikaanse burgers met een leeftijd van 18 jaar of ouder vóór of op de dag van de verkiezingen in het desbetreffende jaar. In 2004 werd de data enkel face-to-face verzameld aan de hand van *Computer Assisted Personal Interviews* (CAPI). De data is verzameld in twee waves: voor de verkiezingen en na de verkiezingen. De respons van de pre-verkiezingsinterviews bedroeg 66,1 procent: 1.211 respondenten zijn geïnterviewd. Een aantal van 1.066 respondenten heeft ook na de presidentsverkiezingen weer een interview laten afnemen. In 2008 werd de data, eveneens in twee waves en tevens face-to-face verzameld middels CAPI, maar ook middels *Audio-Computer Assisted Self-Interviews* (ACASI). De response-rate bedroeg in dit jaar 63,7 procent. Van de 2.322 respondenten die voorafgaand aan de verkiezingen een interview hebben afgevoerd, deden 2.102 dit na de verkiezingen opnieuw. In 2012 werd de data verzameld middels een dual-mode survey: zowel face-to-face in CAPI- en CASI-vorm (*Computer Aided Self-Evaluation*) als aan de hand van online-interviews via het internet. De face-to-face-steekproef is weer uitgevoerd in twee waves, en de internet-steekproef in vier waves (waarvan twee vóór de verkiezingen, en twee na de verkiezingen). De respons binnen de face-to-face-steekproef bedroeg een percentage van 49 procent. Voor de verkiezingen werden er 2.054 respondenten benaderd, en hiervan werden na de verkiezingen 1.929 opnieuw bereikt. Bij de internet-steekproef was er een respons van 2 procent. Voor de verkiezingen werden middels deze wijze 3,860 respondenten benaderd, waarvan er na de verkiezingen 3.581 opnieuw werden bereikt.

Voor dit onderzoek werd een selectie gemaakt uit de dataset. Voor elk van de 3 jaren werd een willekeurige steekproef van 300 respondenten getrokken. De data van de overige respondenten is uit het databestand verwijderd. Dit betekent dat er hierna een aantal van 900 respondenten resteerde. Vervolgens is er missende data verwijderd. De steekproef werd hierdoor opnieuw kleiner in omvang. Dit betekent dat er uiteindelijk met data van 399 respondenten is gewerkt. Met dit kleinere aantal respondenten kunnen bijzonderheden in de data, zoals uitbijters of invloedrijke punten, gemakkelijker worden herkend.

## 3.2 Beschrijving onderzoeksdesign

De data die gebruikt zal worden in dit onderzoek is verzameld middels gestructureerde vragenlijsten die opgesteld zijn door de ANES. In een vragenlijst kwamen verschillende onderwerpen aan de orde. Participanten werden onder andere bevraagd naar hun partijdigheid en houdingen ten aanzien van politieke partijen; hun individuele evaluaties van presidentskandidaten; hun opvattingen ten aanzien van verschillende kwesties in de VS zoals de sociale welvaart, het raciale beleid, het economische beleid, buitenlandse betrekkingen en het klimaat; hun ideologie en waarden; hun burgerlijke betrokkenheid en hun (politieke) kennis. De duur van een individueel interview lag over het algemeen tussen de 65 en 90 minuten (American National Election Study, 2021).

## 3.3 Beschrijving operationalisaties

In deze paragraaf worden de operationalisaties van de variabelen die in dit onderzoek gebruikt worden, beschreven. Onderstaande beschrijvingen zijn van beknopte aard. In bijlage 1 is een uitgebreide toelichting van elke operationalisatie te vinden.

### 3.3.1 Correct voorspellen van de verkiezingsuitslag

Het goed voorspellen van de verkiezingsuitslag is gemeten op basis van de vraag “Who do you think will be elected President in November?”. De variabele *goed\_voorspeld* is een dummy met de waarden 1=goed voorspeld en 0=niet goed voorspeld. Voor het jaar 2004 betekent dit dat verwachtingen dat de Republikeinse kandidaat zou winnen onder de goed voorspelde antwoorden worden geschaard. Alle andere antwoorden worden geschaard onder de niet-goed voorspelde antwoorden. Voor de jaartallen 2008 en 2012 betekent dit dat verwachtingen over een winnende Democratische kandidaat als ‘goed voorspeld’ werden gerekend, en de overige antwoorden als ‘niet goed voorspeld’. Dit betekent dat ook antwoorden die aangaven dat de respondent nog niet wist wie de winnende presidentskandidaat zou zijn, zijn geschaard onder de niet-goed voorspelde antwoorden. Deze keuze valt te betwisten. Ik zal hier in het discussiehoofdstuk op terugkomen.

### 3.3.2 Betrokkenheid bij de gemeenschap

Betrokkenheid bij de gemeenschap is gemeten door respondenten te vragen of zij de afgelopen 12 maanden op een manier betrokken zijn geweest bij hun gemeenschap via de volgende twee items: (1) “During the past 12 months, have you worked with other people to deal with some issue facing your community?”, en (2) “During the past twelve months, did you attend a meeting about an issue

facing your community or schools?”. De itemscores zijn bij elkaar opgeteld tot een schaal (Cronbachs  $\alpha=0,65$ ). De score op deze schaalvariabele, *betrokken\_gem*, representeert het aantal keer dat een respondent ‘yes’ heeft geantwoord op eerdergenoemde twee vragen. Een hogere score duidt zodoende op een grotere betrokkenheid bij de gemeenschap. De variabele zal als continue variabele worden beschouwd. De gerapporteerde Cronbach’s alpha-waarde duidt erop dat de betrouwbaarheid van de variabele twijfelachtig is. Naast bovenstaande twee items was er nog één andere variabele in de dataset beschikbaar die respondenten bevroeg naar het wel of niet hebben gedaan van vrijwilligerswerk in de afgelopen 12 maanden. Vanwege inhoudelijke overwegingen is deze echter niet als derde item meegenomen in de schaalvariabele. In het discussiehoofdstuk kom ik terug op deze keuze. Het uitblijven van andere geschikte items betekent dat er niet een andere variabele aan de schaalvariabele kan worden toegevoegd om de betrouwbaarheid te vergroten.

### **3.3.3 Discussiëren over politiek**

Discussiëren over politiek (*discuss\_pol*) is gemeten met de vraag “How many days in the past week did you talk about politics with your family or friend?”. Men kon antwoorden op een 8-puntsschaal waarbij 0=None or never discuss politics; 1=one day; 2=two days; 3=three days; 4=four days; 5=five days; 6=six days, en 7=seven days. Er is één kanttekening te plaatsen bij het gebruiken van deze variabele in dit onderzoek. In het theoriehoofdstuk is gesteld dat het bij ‘discussiëren over politiek’ met name gaat over het discussiëren met buurtgenoten. De vraag die de respondenten gesteld is verwijst echter naar het discussiëren met een familielid of vriend. Er is niet met zekerheid te zeggen dat respondenten hun discussies met buurtgenoten ook hebben meegenomen in hun antwoord op bovenstaande vraag. De gevolgen van deze mogelijke discrepantie bespreek ik nader in het discussiehoofdstuk.

### **3.3.4 Partij-identificatie**

Partij-identificatie is gemeten aan de hand van meerdere vragen. De eerste vraag was: “Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent, or what?”. Als een respondent op deze eerste vraag antwoorde dat hij of zij zichzelf als een Republikein of Democraat zag, werd gevraagd: “Would you call yourself a strong (REP/DEM) or a not very strong (REP/DEM)?”. Als een respondent op de eerste vraag had geantwoord dat hij of zichzelf als een onafhankelijk persoon zag, werd gevraagd: “Do you think of yourself as closer to the Republican or Democratic party?”. Aan de hand van deze gegevens zijn twee variabelen geconstrueerd die de

verwachte (afzonderlijke) effecten van partij-identificatie zullen meten: identificatie met de winnende partij en identificatie met de verliezende partij. Op beide variabelen kan worden geantwoord aan de hand van 4 antwoordcategorieën. Voor identificatie met de winnende partij (variabelenaam *ident\_W*) zijn deze: 0=independent/identification with losing party; 1=weak identification; 2=medium identification; 3=strong identification. Voor identificatie met de verliezende partij (variabelenaam *ident\_L*) zijn deze: 0=independent/identification with winning party; 1=weak identification; 2=medium identification; 3=strong identification. De variabelen worden in dit onderzoek als continue beschouwd. De variabelen *ident\_W* en *ident\_L* zijn tevens onderdeel van twee interactievariabelen die in de regressieanalyses zullen worden meegenomen. Interactievariabele 1 (*BETRxDENT\_W*) is een vermenigvuldiging van de variabele *betrokken\_gem* met *ident\_W*. Interactievariabele 2 (*BETRxDENT\_L*) is een vermenigvuldiging van de variabelen *betrokken\_gem* en *ident\_L*.

### **3.3.5 Interesse in de verkiezingen**

Interesse in de verkiezingen (*interesse\_verk*) is gemeten door de vraag “Some people don’t pay much attention to political campaigns. How about you, would you say that you have been/were very much interested, somewhat interested, or not much interested in the political campaigns (so far) this year?”. Men kon antwoorden op een 3-puntsschaal waarbij 1=not much interested; 2=somewhat interested en 3=very much interested. Deze variabele wordt ook als continue beschouwd in dit onderzoek.

### **3.3.6 Opleidingsniveau**

Opleidingsniveau (*opleidingsniveau*) is in 2004 en 2008 gemeten middels de volgende vragen: “What is the highest grade of school or year of college you have completed? Did you get a high school diploma or pass a high school equivalency test? What is the highest degree that you have earned?”. In 2012 werd opleidingsniveau gemeten met de vraag “What is the highest level of school you have completed or the highest degree you have received?”. Men kon in alle jaren op deze vragen antwoorden met de waarden 1=grade school or less (0-8 grades); 2=high school (12 grades or fewer); 3=some college (13 grades or more but no degree) en 4=college or advanced degree. *Opleidingsniveau* zal tevens als continue variabele beschouwd worden.

### 3.3.7 Leeftijd en geslacht

Leeftijd (*leeftijd*) is gemeten aan de hand van de vraag “What is the month, day and year of your birth?”. De leeftijd werd voor elke respondent op basis van de opgegeven geboortemaand, -dag en -jaar berekend. Geslacht is gemeten door respondenten te vragen wat hun geslacht is. De antwoordcategorieën waren 1=male en 2=female. De variabelenaam die voor deze variabele gebruikt zal worden is *vrouw*.

## 3.4 Analyse-opzet

Voor de beantwoording van de onderzoeksvraag, en het toetsen van de hypothesen, zullen 8 regressiemodellen geschat worden. Model 3 en 4 zijn lineaire regressiemodellen. Modellen 1, 2, 5, 6, 7 en 8 zijn logistische regressiemodellen. Het eerste model zal bestaan uit de controlevariabelen leeftijd, geslacht, opleidingsniveau en interesse in de verkiezingen, en de afhankelijke variabele ‘goed voorspellen van de winnaar’. Hierdoor kan het effect van de controlevariabelen op de afhankelijke variabele worden geschat. De variabelen leeftijd, opleidingsniveau en interesse in de verkiezingen worden als continue variabelen beschouwd.

In het tweede model wordt de variabele ‘betrokkenheid bij de gemeenschap’ toegevoegd. Dit betreft het toetsen van het positieve (hoofd-)effect van betrokkenheid bij de gemeenschap op de kans op een goede voorspelling van de verkiezingswinnaar dat wordt voorspeld in hypothese 1.

Het derde model is een lineair regressiemodel waarin ‘discussiëren over politiek’ de afhankelijke variabele is. Dit model bevat de controlevariabelen leeftijd, geslacht, opleidingsniveau en interesse in de verkiezingen. Het vierde model is ook een lineair regressiemodel met ‘discussiëren over politiek’ als afhankelijke variabele. Dit model bevat niet enkel de controlevariabelen maar ook de onafhankelijke variabele ‘betrokkenheid bij de gemeenschap’. Hiermee wordt het eerste onderdeel van het verwachte mediatie-effect uit hypothese 2 getoetst: heeft betrokkenheid bij de gemeenschap een positief effect op het discussiëren over politiek?

Het vijfde model is een logistisch regressiemodel waarin ‘goed voorspellen van de winnaar’ de afhankelijke variabele is. Dit model bevat de controlevariabelen en de variabelen ‘betrokkenheid bij de gemeenschap’ en ‘discussiëren over politiek’. Hiermee wordt het tweede onderdeel van het verwachte mediatie-effect uit hypothese 2 getoetst: heeft het discussiëren over politiek een positief effect op het goed voorspellen van de verkiezingswinnaar?

Het zesde model bevat de controlevariabelen, de variabele ‘betrokkenheid bij de gemeenschap’ en de variabelen ‘identificatie met winnende partij’ en ‘identificatie met verliezende partij’. De afhankelijke variabele van dit model is ‘goed voorspellen van de winnaar’. Hiermee



worden de verwachte moderatie-effecten uit hypothese 3 en hypothese 4 getoetst.

Het zevende model bevat de controlevariabelen, de onafhankelijke variabele 'betrokkenheid bij de gemeenschap', 'identificatie met winnende partij', 'identificatie met verliezende partij' en de interactie  $BETR \times IDENT\_W$  om daarmee hypothese 3 te toetsen. Hiermee wordt bekeken of het hoofdverband tussen betrokkenheid bij de gemeenschap en het goed voorspellen van de winnaar verschilt voor verschillende maten van sterkte van identificatie met de winnende partij.

Het achtste model bevat de controlevariabelen, de onafhankelijke variabele 'betrokkenheid bij de gemeenschap', 'identificatie met winnende partij', 'identificatie met verliezende partij' en de interactie  $BETR \times IDENT\_L$  om daarmee hypothese 4 te toetsen. Hiermee wordt bekeken of het hoofdverband tussen betrokkenheid bij de gemeenschap en het goed voorspellen van de verkiezingswinnaar verschilt voor verschillende maten van sterkte van identificatie met de verliezende partij.

## 4 Resultaten

### 4.1 Beschrijvende statistieken

Tabel 1: Univariate beschrijving van de statistieken

	Gem. (SD) <sup>a</sup>	Range	Min.	Max.	N
<b>goed_voorspeld<sup>a</sup></b>					399
0 = niet goed voorspeld	37,60%				
1 = goed voorspeld	62,40%				
<b>betrokken_gem</b>	0,52 (0,76)	2,00	0,00	2,00	399
<b>discuss_pol</b>	2,35 (2,47)	7,00	0,00	7,00	399
<b>ident_W</b>	0,93 (1,21)	3,00	0,00	3,00	399
<b>ident_L</b>	0,91 (1,14)	3,00	0,00	3,00	399
<b>interesse_verk</b>	2,22 (0,68)	2,00	1,00	3,00	399
<b>opleidingsniveau</b>	2,84 (0,88)	3,00	1,00	4,00	399
<b>leeftijd</b>	47,54 (17,28)	72,00	18,00	90,00	399
<b>vrouw<sup>a</sup></b>					399
1 = man	44,40%				
2 = vrouw	55,60%				

<sup>a</sup> De beschrijvende statistieken van categorische variabelen zijn in percentages weergegeven.

In bovenstaande tabel 1 worden de beschrijvende statistieken van de variabelen uit dit onderzoek weergegeven. Een uitgebreide toelichting van de wijze waarop deze gegevens zijn verzameld wordt weergegeven in bijlage 2. De in tabel 1 weergegeven percentages niet-goed voorspelde en goed voorspelde winnaars tonen aan dat de meeste mensen de verkiezingswinnaar goed hebben voorspeld. Dit komt overeen met de eerdergenoemde bevindingen in de inleiding en het theoriehoofdstuk. Betrokkenheid bij de gemeenschap is een schaalvariabele bestaande uit twee items. Het gemiddelde van deze variabele is laag ( $\mu = 0,52$ ): de meeste respondenten zijn niet erg sterk betrokken bij de gemeenschap. De variabele Discussiëren over politiek is een variabele van ordinaal meetniveau. Het gemiddelde ( $\mu = 2,35$ ) en de standaarddeviatie ( $SD = 2,47$ ) wijzen uit dat relatief de meeste respondenten zich aan de onderkant van de schaal bevinden: weinig respondenten discussiëren zes of zeven keer per week over politiek. Tot slot wordt nog de man-vrouw verdeling onder de respondenten weergegeven: ongeveer 44 tegen 56 procent. Hieruit blijkt dat er meer vrouwelijke dan mannelijke respondenten zijn opgenomen in dit onderzoek.

Tabel 2: Tabel met samenhangende maten van alle variabelen in het onderzoeksmodel

	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. goed_voorspeld	-								
2. betrokken_gem	0,032 <sup>c</sup>	-							
3. discuss_pol	0,071 <sup>c</sup>	**0,149 <sup>a</sup>	-						
4. ident_W	**0,402 <sup>c</sup>	0,014 <sup>a</sup>	0,096 <sup>a</sup>	-					
5. ident_L	**0,428 <sup>c</sup>	0,033 <sup>a</sup>	0,038 <sup>a</sup>	** -0,614 <sup>a</sup>	-				
6. interesse_verk	0,000 <sup>c</sup>	**0,132 <sup>a</sup>	**0,405 <sup>a</sup>	**0,173 <sup>a</sup>	0,079 <sup>a</sup>	-			
7. opleidingsniveau	0,000 <sup>c</sup>	**0,241 <sup>a</sup>	**0,190 <sup>a</sup>	0,055 <sup>a</sup>	0,007 <sup>a</sup>	**0,149 <sup>a</sup>	-		
8. leeftijd	0,032 <sup>c</sup>	0,039 <sup>a</sup>	**0,129 <sup>a</sup>	*0,127 <sup>a</sup>	0,060 <sup>a</sup>	**0,168 <sup>a</sup>	*-0,105 <sup>a</sup>	-	
9. vrouw	0,016 <sup>b</sup>	0,032 <sup>c</sup>	0,032 <sup>c</sup>	0,000 <sup>c</sup>	*0,122 <sup>c</sup>	0,045 <sup>c</sup>	0,000 <sup>c</sup>	0,032 <sup>c</sup>	-

\* Significant op 0,05; \*\* Significant op 0,01; <sup>a</sup> correlatie op basis van t-toets voor verschil in gemiddelden; <sup>b</sup> Cramer's V; <sup>c</sup> correlatie op basis van ANOVA

In tabel 2 zijn de correlaties tussen alle variabelen uit het onderzoeksmodel weergegeven. Een uitgebreide toelichting van de wijze waarop deze gegevens zijn verzameld is wederom terug te vinden in bijlage 2. Uit bovenstaande tabel blijkt ten eerste dat er een sterke significante samenhang bestaat tussen zowel Goed voorspellen van de winnaar en Identificatie met winnende partij ( $r = 0,402$ ) als Goed voorspellen van de winnaar en Identificatie met verliezende partij ( $r = 0,428$ ). Ten tweede is er een opvallende correlatie tussen Discussiëren over politiek en Interesse in de verkiezingen ( $r = 0,405$ ). Deze suggereert dat er een sterke, positieve (en tevens significante) samenhang bestaat tussen deze variabelen: vaker discussiëren over politiek gaat samen met een grotere interesse in de verkiezingen. Ten derde valt de sterke en negatieve correlatie tussen Identificatie met winnende partij en Identificatie met verliezende partij op ( $r = -0,614$ ). Een grotere identificatie met de winnende partij gaat gepaard met een lagere identificatie met de verliezende partij, en vice versa.

Op basis van hypothese 1 werd een sterke samenhang verwacht tussen Goed voorspellen van de winnaar en Betrokkenheid bij de gemeenschap. Uit bovenstaande tabel blijkt echter dat er vrijwel geen samenhang bestaat tussen deze twee concepten ( $r = 0,032$ ). Daarnaast kan op basis van hypothese 2 gesteld worden dat er een sterke samenhang werd verwacht tussen Betrokkenheid bij de gemeenschap en Discussiëren over politiek, en ook tussen Discussiëren over politiek en Goed voorspellen van de winnaar. Betrokkenheid bij de gemeenschap en Discussiëren over politiek blijken echter niet in sterke mate samen te hangen ( $r = 0,149$ ). De samenhang is wel significant. De samenhang tussen Discussiëren over politiek en Goed voorspellen van de winnaar is nog zwakker en de correlatie is ook niet significant ( $r = 0,071$ ).

## 4.2 Modevaluatie

Tabel 3: Parameters uit de logistische regressieanalyse met 'Goed voorspellen van de verkiezingswinnaar' als afhankelijke variabele

	Model 1 <sup>a</sup>			Model 2 <sup>a</sup>			Model 3 <sup>b</sup>		Model 4 <sup>b</sup>		Model 5 <sup>a</sup>		
	<i>b</i> (SE)	Odds-ratio	<i>p</i>	<i>b</i> (SE)	Odds-ratio	<i>p</i>	<i>b</i> (SE)	<i>p</i>	<i>b</i> (SE)	<i>p</i>	<i>b</i> (SE)	Odds-ratio	<i>p</i>
Constante	0,584 (0,622)	1,793	0,348	0,542 (0,625)	1,720	0,386	-2,665 (0,673)	<0,001	-2,569 (0,675)	<0,001	0,368 (0,637)	1,445	0,563
leeftijd	0,004 (0,006)	1,004	0,496	0,004 (0,006)	1,004	0,476	0,011 (0,007)	0,085	0,011 (0,007)	0,098	0,005 (0,006)	1,005	0,404
vrouw	-0,074 (0,209)	0,928	0,722	-0,070 (0,209)	0,932	0,737	0,201 (0,226)	0,375	0,191 (0,226)	0,399	-0,058 (0,210)	0,944	0,784
opleidingsniveau	-0,019 (0,121)	0,981	0,873	0,000 (0,124)	1,000	0,998	0,399 (0,130)	0,002	0,356 (0,134)	0,008	0,024 (0,125)	1,024	0,851
interesse_verk	-0,047 (0,158)	0,954	0,765	-0,038 (0,159)	0,963	0,812	1,362 (0,171)	<0,001	1,340 (0,171)	<0,001	0,053 (0,172)	1,055	0,757
betrokken_gem				-0,094 (0,141)	0,910	0,503			0,215 (0,154)	0,162	-0,081 (0,142)	0,922	0,569
discuss_pol											-0,067 (0,046)	0,935	0,150
Deviance	527,632			527,185							525,111		
X2-toets	0,677		0,954	0,447		0,504					2,075		0,150
Hosmer-Lemeshowtoets	12,128		0,146	6,454		0,597					5,320		0,723
R <sup>2</sup> <sub>adjusted</sub>							0,181		0,183				
F-change / <i>p</i>							23,015 <0,001		18,850 <0,001				
<i>n</i>	399			399			399		399		399		

<sup>a</sup> Logistisch regressiemodel met goed\_voorspeld als afhankelijke variabele; <sup>b</sup> Lineair regressiemodel met discuss\_pol als afhankelijke variabele

Tabel 4: Parameters uit de logistische regressieanalyse met 'Goed voorspellen van de verkiezingswinnaar' als afhankelijke variabele

	Model 6			Model 7			Model 8		
	<i>b</i> (SE)	Odds-ratio	<i>p</i>	<i>b</i> (SE)	Odds-ratio	<i>p</i>	<i>b</i> (SE)	Odds-ratio	<i>p</i>
Constante	0,616 (0,693)	1,851	0,374	0,618 (0,693)	1,855	0,373	0,654 (0,710)	1,924	0,357
leeftijd	0,004 (0,007)	1,004	0,568	0,004 (0,007)	1,004	0,564	0,004 (0,007)	1,004	0,580
vrouw	0,103 (0,240)	1,108	0,669	0,106 (0,241)	1,111	0,661	0,093 (0,243)	1,098	0,701
opleidingsniveau	-0,041 (0,141)	0,960	0,773	-0,041 (0,141)	0,960	0,770	-0,040 (0,141)	0,961	0,775
interesse_verk	-0,103 (0,188)	0,902	0,585	-0,102 (0,188)	0,903	0,587	-0,103 (0,188)	0,902	0,583
betrokken_gem	-0,084 (0,160)	0,919	0,600	-0,096 (0,183)	0,909	0,601	-0,121 (0,216)	0,886	0,575
ident_W	0,594 (0,153)	1,811	<0,001	0,581 (0,180)	1,789	0,001	0,595 (0,152)	1,813	<0,001
ident_L	-0,497 (0,130)	0,608	<0,001	-0,498 (0,130)	0,608	<0,001	-0,514 (0,148)	0,598	<0,001
BETRxDIDENT_W				0,020 (0,155)	1,021	0,895			
BETRxDIDENT_L							0,034 (0,133)	1,034	0,800
Deviance	434,382			434,365			434,318		
X2-toets	93,927		<0,001	0,017		0,895	0,064		0,800
Hosmer-Lemeshowtoets	6,759		0,563	7,330		0,502	5,379		0,716
<i>n</i>	399			399			399		

In bovenstaande tabellen 3 en 4 worden de resultaten van de uitgevoerde regressieanalyses weergegeven. Tabel 3 bevat de resultaten van modellen 1 tot en met 5. De tabel bevat zowel logistische regressiemodellen met de afhankelijke variabele 'Goede voorspelling' als lineaire modellen met de afhankelijke variabele 'Discussiëren politiek'. Tabel 4 bevat de resultaten van de logistische regressiemodellen 6 tot en met 8. In bijlage 2 wordt uitgebreid toegelicht op welke wijze de resultaten, die in deze tabellen worden weergegeven, zijn verkregen. In deze paragraaf zal de kwaliteit van de getoetste modellen geanalyseerd worden.

Model 1 – te zien in tabel 3 – bevat enkel de controlevariabelen leeftijd, vrouw, opleidingsniveau en interesse verkiezingen. De Hosmer-Lemeshowtoets van dit model (*Hosmer-Lemeshow*  $X^2 = 12,128$ ;  $p = 0,146$ ) geeft geen indicatie dat het model een slechte fit zou hebben: de observaties die gedaan worden zijn vermoedelijk redelijk gelijk aan de schattingen die worden gemaakt aan de hand van dit model. Zodra de odds op het goed voorspellen van de winnaar niet enkel worden geschat op basis van de controlevariabelen, maar ook op basis van de mate van betrokkenheid die een respondent heeft bij de gemeenschap, blijft het aantal fouten dat het model maakt vrijwel gelijk. Het aantal fouten dat het model maakt wordt aangegeven door de deviance. Het aantal fouten dat model 2 maakt ( $D = 527,185$ ) is vrijwel gelijk aan het aantal fouten dat wordt gemaakt door model 1 ( $D = 527,632$ ). De deviance daalt heel weinig dus de modelfit wordt eigenlijk nauwelijks beter wanneer Betrokkenheid bij gemeenschap wordt toegevoegd aan het model.

Modellen 3 en 4 in tabel 3 zijn lineaire regressiemodellen. De proportie verklaarde variantie van model 3 is redelijk goed ( $R^2_a = 0,181$ ). De modelfit wordt nog lichtelijk beter wanneer Betrokkenheid bij gemeenschap wordt toegevoegd aan het model ( $R^2_a = 0,183$ ), maar dit verschil is erg klein. Aan de hand van model 5 kan het tweede pad van het mediatie-effect worden bekeken. In dit model wordt namelijk het effect van discussiëren over politiek op het goed voorspellen van de winnaar berekend, gecontroleerd voor de controlevariabelen en de betrokkenheid bij de gemeenschap. De deviance van dit vijfde model ( $D = 525,111$ ) is lichtelijk afgenomen in vergelijking met model 2. Wanneer Discussiëren over politiek wordt toegevoegd aan het model kan de kans op het correct voorspellen van de verkiezingsuitslag dus niet significant beter worden geschat.

Voor het onderzoeken van de moderatie-effecten wordt gekeken naar de modellen 6, 7 en 8 in tabel 4. Model 6 bevat de controlevariabelen, Betrokkenheid bij gemeenschap, en de modererende variabelen Identificatie winnende partij en Identificatie verliezende partij. De Hosmer-Lemeshowtoets geeft geen indicatie voor het niet goed passen van dit model (*Hosmer-Lemeshow*  $X^2 = 6,759$ ;  $p = 0,563$ ). Model 6 is als het ware een uitbreiding van model 2: de twee modererende variabelen worden toegevoegd in vergelijking met model 2. Om deze reden kan de deviance van model 6 ( $D = 434,382$ ) vergeleken worden met die van model 2 ( $D = 527,185$ ). Op basis van deze

vergelijking kan worden geconcludeerd dat er minder fouten worden gemaakt in het schatten van de odds op Goede voorspelling als de twee modererende variabelen aan het model worden toegevoegd.

De deviance daalt lichtelijk wanneer de interactieterm  $BETR \times IDENT\_W$  (het product van de variabelen Betrokkenheid bij gemeenschap en Identificatie winnende partij) aan het model wordt toegevoegd in model 7 ( $D = 434,365$ ). Het model wordt dus iets beter wanneer het deze interactieterm bevat. Model 8 bevat de interactieterm  $BETR \times IDENT\_L$  (het product van de variabelen Betrokkenheid bij de gemeenschap en Identificatie met verliezende partij). Het aantal fouten dat dit model maakt ( $D = 434,318$ ) is wederom iets lager dan het aantal fouten dat model 6 ( $D = 434,382$ ) maakt, maar de daling is wederom wel erg klein. Mogelijk wordt de modelfit nauwelijks beter wanneer de interactieterm aan het model wordt toegevoegd.

Tot slot zal ik nog kort ingaan op de controle van assumpties, multicollineariteit en uitbijters. Een uitgebreide bespreking van deze controle is terug te vinden in bijlage 3. Ten eerste geldt dat er moet worden voldaan aan de assumptie van onafhankelijke waarnemingen. Bij de wijze van dataverzameling moet hiermee rekening gehouden worden. De data die wordt gebruikt in dit onderzoek is afkomstig van de American National Election Studies. Zij verzamelen de data door respondenten op basis van willekeurige steekproeven te benaderen (American National Election Study, 2021). Op deze wijze wordt ervoor gecontroleerd dat de verzamelde cases onafhankelijk van elkaar zijn: er wordt voldaan aan deze assumptie.

Voor het controleren op outliers in de data wordt er gekeken naar de leverage-waarden en DFBETA-waarden. Op basis van de DFBETA worden er geen mogelijke uitbijters aangewezen: de waarden van alle parameters liggen ruim onder de gehanteerde criteriumwaarde voor DFBETA. Hetzelfde geldt voor de analyse op outliers op basis van de leverage. Alle cases in de dataset hebben leverage-waarden die onder de gehanteerde criteriumwaarden vallen. Er kan worden geconcludeerd dat er geen sprake is van outliers in de dataset. Het is dus niet nodig om cases uit het databestand te verwijderen.

Om de multicollineariteit in het databestand te onderzoeken worden de VIF-scores van de eindmodellen geanalyseerd. Uit deze analyses blijkt dat de VIF-scores in de eindmodellen allemaal ruim onder de gehanteerde criteriumwaarde vallen. Dit betekent dat er in deze regressieanalyses geen sprake is van problematische multicollineariteit.

### 4.3 Hypothesetoetsing

Hypotheses 1 tot 4 worden getoetst aan de hand van de regressiemodellen die in tabel 3 en tabel 4 weergegeven worden. Een uitgebreide toelichting van de effecten die in de tabellen zijn weergegeven is te vinden in bijlage 2.

De controlevariabelen zijn allereerst opgenomen in model 1. In tegenstelling tot leeftijd en geslacht hebben opleidingsniveau en interesse in de verkiezingen niet het verwachte effect op de odds op een goede voorspelling. Om de hypothese te toetsen dat een grotere betrokkenheid bij de gemeenschap zou leiden tot het goed voorspellen van de winnaar werd de variabele Betrokkenheid bij gemeenschap toegevoegd in model 2. Uit tabel 3 blijkt echter dat een grotere betrokkenheid leidt tot lagere odds op een goede voorspelling ( $b = -0,094$ ;  $p = 0,503$ ;  $OR = 0,910$ ). Een grotere betrokkenheid bij de gemeenschap hangt dus negatief samen met het maken van een goede voorspelling. Echter, de  $p$ -waarde ( $p = 0,503$ ) duidt aan dat het effect niet significant is. Er kan gesteld worden dat hypothese 1 niet wordt ondersteund door de resultaten.

Het negatieve hoofdeffect van betrokkenheid bij de gemeenschap dat in model 2 wordt gesuggereerd is daarnaast ook in strijd met hypothese 2: de verwachting dat het positieve verband tussen betrokkenheid en een goede voorspelling zou verlopen via het discussiëren over politiek. Model 4 en model 5 – weergegeven in tabel 3 – zijn de andere twee modellen aan de hand waarvan de hypothese kan worden beantwoord. Model 4 in tabel 3 bevestigt dat een hogere betrokkenheid bij de gemeenschap leidt tot het vaker discussiëren over politiek ( $b = 0,340$ ;  $p = 0,006$ ). De resultaten uit model 5 ondersteunen de verwachtingen uit hypothese 2 daarentegen niet. De odds op het correct voorspellen van de verkiezingsuitslag dalen namelijk wanneer er vaker wordt gediscussieerd over politiek, gecontroleerd voor de overige variabelen in het model ( $b = -0,060$ ;  $p = 0,194$ ;  $OR = 0,942$ ). Dit negatieve effect blijkt niet significant ( $p = 0,194$ ). De resultaten die vergaard zijn uit de modellen zijn uiteindelijk niet in overeenstemming met hypothese 2.

Zoals middels model 2 aangetoond is, bestaat er een negatief (niet-significant) effect tussen een sterkere betrokkenheid en een goede voorspelling. In hypothese 3 en 4 werden twee modererende effecten op dit hoofdeffect voorspeld. Het hoofdeffect blijkt negatief en uit model 7 volgt dat de interactie positief is ( $b = 0,020$ ;  $p = 0,895$ ;  $OR = 1,021$ ). Voor mensen die zich sterker identificeren met de winnende partij is het negatieve effect van betrokkenheid op het goed voorspellen dus een beetje kleiner, maar dit effect is niet significant. Dit betekent dat de resultaten hypothese 3 niet ondersteunen. Alhoewel dit interactie-effect niet significant is, blijkt (uit model 6) dat de moderator (identificatie met de winnende partij) wél een significant positief effect heeft op de odds op het goed voorspellen van de winnaar ( $b = 0,594$ ;  $p < 0,001$ ;  $OR = 1,811$ ): de odds op het maken van een goede voorspelling worden 1,811 keer groter als de identificatie met de winnende partij één eenheid



sterker wordt.

Uit model 8 blijkt vervolgens dat de tweede interactie ook positief is ( $b = 0,034$ ;  $p = 0,800$ ;  $OR = 1,034$ ). Ook voor mensen die zich sterker identificeren met de verliezende partij is het negatieve effect van betrokkenheid op het goed voorspellen een klein beetje zwakker. Echter, dit effect is wederom niet significant. De resultaten zijn daarmee niet in overeenstemming met hypothese 4. Alhoewel dit interactie-effect niet significant is, is het effect van de moderator (identificatie met de verliezende partij) op de odds op het goed voorspellen van de winnaar wederom wél significant ( $b = -0,497$ ;  $p < 0,001$ ;  $OR = 0,608$ ). Er is een negatief effect: hoe sterker de identificatie met de verliezende partij, hoe kleiner de odds op het maken van een goede voorspelling.

## 5 Conclusie en discussie

In de inleiding werd de aanleiding beschreven tot het onderzoeken of mensen die in een hogere mate betrokken zijn bij de gemeenschap beter de Amerikaanse verkiezingsuitslag kunnen voorspellen dan mensen die nauwelijks betrokken zijn bij de gemeenschap. Hierbij bleek er tevens aanleiding tot onderzoeken in hoeverre dit verband verklaard kan worden doordat mensen die sterker betrokken zijn bij hun gemeenschap vaker over politiek discussiëren en daardoor beter de winnaar van de verkiezingen kunnen aanwijzen. Tot slot werd er de vraag gesteld of het verband zou blijven bestaan wanneer mensen zich sterk identificeren met een politieke partij.

Het onderzoek heeft ten eerste laten zien dat mensen die meer betrokken zijn bij de gemeenschap niet beter de winnaar van de verkiezingen kunnen voorspellen dan mensen die nauwelijks betrokken zijn bij de gemeenschap. Er werd verwacht dat mensen hun verwachtingen onder andere baseren op informatie die zij vanuit hun sociale netwerk ontvangen. Een groter en meer divers netwerk als gevolg van een grotere betrokkenheid bij de gemeenschap zou leiden tot representatievere informatie op basis waarvan mensen accuratere verwachtingen zouden kunnen vormen – onder meer over de winnaar van de verkiezingen (Banerjee & Fudenberg, 2004; Leiter et al., 2018; Mongrain, 2021; Speer et al., 2021). De discrepantie tussen de verwachtingen en de resultaten van dit onderzoek kunnen mogelijk te wijten zijn aan de onbetrouwbaarheid van de geconstrueerde schaal voor betrokkenheid bij de gemeenschap. Dit betekent dat er veel ruis is bij het meten van dit concept. De metingen in dit onderzoek zijn hierdoor niet erg betrouwbaar en de resultaten daardoor minder goed te generaliseren. De onbetrouwbare schaal vormt hiermee een mogelijke verklaring voor het gevonden tegengestelde, niet-significante resultaat. Daarnaast werd reeds bij de beschrijving van de operationalisaties benoemd dat niet enkel het aanwijzen van de verkeerde winnende kandidaat, maar ook het aangeven door de respondent dat hij of zij nog niet wist welke kandidaat zou winnen (“Dont know/Depends”), werd geschaard onder ‘Niet goed voorspeld’. Ter vergelijking, bij andere variabelen werd deze antwoordcategorie onder de missende waarden geschaard. Deze keuze kan de gevonden resultaten hebben vertekend. Echter, dit antwoord is wegens inhoudelijke overwegingen uiteindelijk wel onder de niet-goed voorspelde antwoorden geschaard: de respondent kon vanuit de informatie die hij of zij had verzameld (vanuit de media of het sociale netwerk) immers geen goede voorspelling maken. Het antwoord “don’t know” krijgt daardoor inhoudelijk een andere lading.

De resultaten van dit onderzoek zijn daarnaast ook niet in overeenstemming met de tweede hypothese. Er werd verwacht dat mensen met een grotere betrokkenheid bij de gemeenschap vaker zouden discussiëren over politieke onderwerpen doordat een divers sociaal netwerk tot gevolg heeft dat mensen met nieuwe standpunten in aanraking komen en ervoor open staan om hier van te leren

(McClurg, 2006; Leiter et al., 2018; Scheufele, 2002; Kwak et al., 2005). De resultaten ondersteunen de verwachting dat een grotere betrokkenheid leidt tot het vaker discussiëren over politieke onderwerpen. De verwachting dat het discussiëren over politieke onderwerpen vervolgens leidt tot het beter kunnen voorspellen van de verkiezingswinnaar wordt daarentegen niet ondersteund door de resultaten. Er wordt in het onderzoek immers een tegengesteld, niet-significant resultaat gevonden voor deze verwachting. Er zijn echter kanttekeningen te plaatsen bij de wijze waarop het concept van discussiëren over politiek is gemeten in dit onderzoek. Vanuit de gebruikte theorieën werd verondersteld dat mensen zouden discussiëren over politiek met hun buurtgenoten als gevolg van de betrokkenheid bij hun gemeenschap. De vraag die aan respondenten gesteld is verwijst echter naar het discussiëren met een familielid of vriend, en niet naar het discussiëren met een buurtgenoot. De metingen zijn hierdoor minder betrouwbaar en het gevonden resultaat mogelijk minder goed te generaliseren naar de gehele populatie.

De verwachting uit de derde hypothese luidde als volgt: naarmate mensen zich sterker identificeren met de winnende partij, is het positieve effect van betrokkenheid op een goede voorspelling sterker. Op basis van de *selective disclosure of attitudes* en de homofilie-theorie werd namelijk gesteld dat mensen die zich sterk identificeren met een politieke partij vrijwel niet geconfronteerd worden met andere meningen (Cowan & Baldassarri, 2017; Baccara & Yariv, 2012; Goel et al., 2010). Hierdoor zullen zij door een grotere betrokkenheid enkel meer in aanraking komen met politieke visies die gelijk zijn aan die van henzelf (Delavande & Manski, 2012; Mongrain, 2021; Leiter et al., 2020). Mensen die zich identificeren met de winnende partij blijven hierdoor overtuigd van het winnen door de door hen geprefereerde verkiezingskandidaat en hebben de winnaar uiteindelijk goed voorspeld. Zoals eerder benoemd is uit dit onderzoek gebleken dat er geen sprake is van het verwachte verband van een betrokkenheid bij de gemeenschap op het goed voorspellen van de winnaar. Uit de onderzoeksresultaten volgt vervolgens dat dit niet-bestaande verband ook niet verschilt voor verschillende sterktes van identificatie met de winnende partij. Dit betekent dat de hypothese niet wordt ondersteund door de bevindingen in het onderzoek.

In de vierde hypothese werd verwacht dat naarmate mensen zich sterker identificeren met de verliezende partij, het positieve effect van betrokkenheid op een goede voorspelling zwakker zou worden. De verwachting wordt hier ook gevormd door de *selective disclosure of attitudes* en homofilie-theorie (Cowan & Baldassarri, 2017; Baccara & Yariv, 2012; Goel et al., 2010). Mensen die zich sterk identificeren met de verliezende partij denken hierdoor ten onrechte dat hun partij veel aanhangers heeft en een goede kans maakt op de verkiezingswinst. Zij maken hierdoor een voorspelling over de verkiezingswinnaar die uiteindelijk niet goed blijkt te zijn. Wederom volgen uit het onderzoek niet-significante resultaten. Deze duiden erop dat naarmate mensen zich sterker

identificeren met de verliezende partij, het negatieve verband tussen de betrokkenheid en het maken van een goede voorspelling zwakker wordt. De vierde hypothese wordt hiermee niet ondersteund.

Zoals in de inleiding reeds werd gesteld, kunnen mensen over het algemeen vaak de juiste winnende presidentskandidaat aanwijzen voorafgaand aan de verkiezingen. Vanwege de gevonden nulresultaten is het lastig om op basis van dit onderzoek factoren aan te wijzen die invloed hebben op de accuraatheid van verwachtingen over de verkiezingen. De resultaten uit dit onderzoek kunnen namelijk niet goed gegeneraliseerd worden naar de populatie, het Amerikaanse electoraat. De onbetrouwbaarheid van enkele metingen zouden daar een mogelijke verklaring voor zijn. Vervolgonderzoek zou de wijze van meten van de in dit onderzoek gebruikte concepten kunnen verbeteren en opnieuw de verwachte verbanden kunnen analyseren.

In de inleiding werd gesteld dat een grotere betrokkenheid van inwoners bij hun gemeenschap veel positieve consequenties heeft: regionale en maatschappelijke vooruitgang (Ramachandra & Naha Abu Mansor, 2014), veranderingen in (lokaal, regionaal en nationaal) beleid (Christens & Speer, 2015), democratischere en rechtvaardigere gemeenschappen (Graycar, 1977), een verbeterde individuele gezondheid en een verhoogd individueel zelfvertrouwen (Attree et al., 2010). Echter, dit onderzoek toont aan dat een grotere gemeenschappelijke betrokkenheid onder individuen niet altijd duidelijke positieve gevolgen heeft. Beleid dat de betrokkenheid van inwoners bij hun gemeenschap stimuleert, leidt er dus niet noodzakelijkerwijs toe dat mensen beter weten hoe anderen politiek denken. Daar zijn mogelijk specifiekere interventies voor nodig.

## Literatuurlijst

- American National Election Study. (2021, 19 november). *Time series cumulative data file (1948–2020)* [Dataset]. <https://electionstudies.org/data-center/anes-time-series-cumulative-data-file/>
- Attree, P., French, B., Milton, B., Povall, S., Whitehead, M., & Popay, J. (2010). The experience of community engagement for individuals: A rapid review of evidence. *Health and Social Care in the Community*, 19(3), 250–260. <https://doi.org/10.1111/j.1365-2524.2010.00976.x>
- Babad, E., Hills, M., & O’Driscoll, M. (1992). Factors influencing wishful thinking and predictions of election outcomes. *Basic and Applied Social Psychology*, 13(4), 461–476. [https://doi.org/10.1207/s15324834basp1304\\_6](https://doi.org/10.1207/s15324834basp1304_6)
- Baccara, M., & Yariv, L. (2012). Homophily in peer groups. *SSRN Electronic Journal*, 5(3), 69–96. <https://doi.org/10.2139/ssrn.2045156>
- Banerjee, A., & Fudenberg, D. (2004). Word-of-mouth learning. *Games and Economic Behavior*, 46(1), 1–22. [https://doi.org/10.1016/s0899-8256\(03\)00048-4](https://doi.org/10.1016/s0899-8256(03)00048-4)
- Brenan, B. M. (2021, 20 november). *Americans’ trust in government remains low*. Gallup.Com. <https://news.gallup.com/poll/355124/americans-trust-government-remains-low.aspx>
- Brown, J. K., & Hohman, Z. P. (2022). Extreme party animals: Effects of political identification and ideological extremity. *Journal of Applied Social Psychology*. <https://doi.org/10.1111/jasp.12863>
- Campus, D. (2012). Political discussion, opinion leadership and trust. *European Journal of Communication*, 27(1), 46–55. <https://doi.org/10.1177/0267323111434580>
- Christens, B. D., & Speer, P. W. (2015). Community organizing: Practice, research, and policy implications. *Social Issues and Policy Review*, 9(1), 193–222. <https://doi.org/10.1111/sipr.12014>
- Cowan, S. K., & Baldassarri, D. (2017). “It could turn ugly”: Selective disclosure of attitudes in political discussion networks. *Social Networks*, 52, 1–17. <https://doi.org/10.1016/j.socnet.2017.04.002>
- Delavande, A., & Manski, C. F. (2012). Candidate preferences and expectations of election outcomes. *Proceedings of the National Academy of Sciences*, 109(10), 3711–3715. <https://doi.org/10.1073/pnas.1200861109>

- Ezrow, L., & Xezonakis, G. (2014). Satisfaction with democracy and voter turnout. *Party Politics*, 22(1), 3–14. <https://doi.org/10.1177/1354068814549335>
- Getty Images. (z.d.). [Night terrors: As the night wore on, and the possibility of a Trump win became more real, the fear of Clinton fans became more pronounced]. Mail Online. <https://www.dailymail.co.uk/news/article-3918838/Dejected-Clinton-supporters-party-goes-flat-result-result-turns-against-cries-lock-louder-Trump-party.html>
- Goel, S., Mason, W., & Watts, D. J. (2010). Real and perceived attitude agreement in social networks. *Journal of Personality and Social Psychology*, 99(4), 611–621. <https://doi.org/10.1037/a0020697>
- Graycar, A. (1977). The relevance of community involvement to social welfare and public administration. *Australian Journal of Public Administration*, 36(3), 238–248. <https://doi.org/10.1111/j.1467-8500.1977.tb02503.x>
- Haas, I. J., Jones, C. R., & Fazio, R. H. (2019). Social identity and the use of ideological categorization in political evaluation. *Journal of Social and Political Psychology*, 7(1), 335–353. <https://doi.org/10.5964/jspp.v7i1.790>
- Kwak, N., Williams, A. E., Wang, X., & Lee, H. (2005). Talking politics and engaging politics: *Communication Research*, 32(1), 87–111. <https://doi.org/10.1177/0093650204271400>
- Leiter, D., Murr, A., Rascón Ramírez, E., & Stegmaier, M. (2018). Social networks and citizen election forecasting: The more friends the better. *International Journal of Forecasting*, 34(2), 235–248. <https://doi.org/10.1016/j.ijforecast.2017.11.006>
- Leiter, D., Reilly, J. L., & Stegmaier, M. (2020). Echoing certainty in uncertain times: Network partisan agreement and the quality of citizen forecasts in the 2015 Canadian election. *Electoral Studies*, 63, 1–6. <https://doi.org/10.1016/j.electstud.2019.102115>
- Lewis-Beck, M. S., & Tien, C. (1999). Voters as forecasters: A micromodel of election prediction. *International Journal of Forecasting*, 15(2), 175–184. [https://doi.org/10.1016/s0169-2070\(98\)00063-6](https://doi.org/10.1016/s0169-2070(98)00063-6)
- Madson, G. J., & Hillygus, D. S. (2019). All the best polls agree with me: Bias in evaluations of political polling. *Political Behavior*, 42(4), 1055–1072. <https://doi.org/10.1007/s11109-019-09532-1>
- Mason, L. (2015). “I disrespectfully agree”: The differential effects of partisan sorting on social and issue polarization. *American Journal of Political Science*, 59(1), 128–145. <https://doi.org/10.1111/ajps.12089>

- McClurg, S. D. (2006). The electoral relevance of political talk: Examining disagreement and expertise effects in social networks on political participation. *American Journal of Political Science*, 50(3), 737–754. <https://doi.org/10.1111/j.1540-5907.2006.00213.x>
- Meffert, M. F., Huber, S., Gschwend, T., & Pappi, F. U. (2011). More than wishful thinking: Causes and consequences of voters' electoral expectations about parties and coalitions. *Electoral Studies*, 30(4), 804–815. <https://doi.org/10.1016/j.electstud.2011.08.001>
- Miller, M. K., Wang, G., Kulkarni, S. R., Poor, H. V., & Osherson, D. N. (2012). Citizen forecasts of the 2008 U.S. presidential election. *Politics & Policy*, 40(6), 1019–1052. <https://doi.org/10.1111/j.1747-1346.2012.00394.x>
- Mongrain, P. (2021). Did you see it coming? Explaining the accuracy of voter expectations for district and (sub)national election outcomes in multi-party systems. *Electoral Studies*, 71, 1–12. <https://doi.org/10.1016/j.electstud.2021.102317>
- NOS. (2021, 7 januari). *Zo kwam het hart van de Amerikaanse democratie even tot stilstand*. <https://nos.nl/collectie/13854/artikel/2363358-zo-kwam-het-hart-van-de-amerikaanse-democratie-even-tot-stilstand>
- Putnam, R.D. (1995). Bowling Alone: America's Declining Social Capital. *Journal of Democracy* 6(1), 65-78. [doi:10.1353/jod.1995.0002](https://doi.org/10.1353/jod.1995.0002)
- Rainie, L., & Perrin, A. (2020, 30 mei). *Key findings about Americans' declining trust in government and each other*. Pew Research Center. <https://www.pewresearch.org/fact-tank/2019/07/22/key-findings-about-americans-declining-trust-in-government-and-each-other/>
- Scheufele, D. A. (2002). Examining differential gains from mass media and their implications for participatory behavior. *Communication Research*, 29(1), 46–65. <https://doi.org/10.1177/0093650202029001003>
- Ramachandra, A., & Naha Abu Mansor, N. (2014). Sustainability of community engagement – in the hands of stakeholders? *Education + Training*, 56(7), 588–598. <https://doi.org/10.1108/et-07-2014-0084>
- Speer, P. W., Christens, B. D., & Peterson, N. A. (2021). Participation in community organizing: Cross-sectional and longitudinal analyses of impacts on sociopolitical development. *Journal of Community Psychology*, 49(8), 3194–3214. <https://doi.org/10.1002/jcop.22578>

Stevenson, C., Turner, R., & Costa, S. (2020). "Welcome to our neighbourhood": Collective confidence in contact facilitates successful mixing in residential settings. *Group Processes & Intergroup Relations*, 24(8), 1448–1466. <https://doi.org/10.1177/1368430220961151>



## Bijlage 1

### 1.1 Variabele Goed voorspellen van de winnaar: goed\_voorspeld

#### Onderdeel 1: Oorspronkelijke variabele

*Syntax:*

\*Oorspronkelijke variabele VCF0700.

```
FREQUENCIES VARIABLES=VCF0700
```

```
/ORDER=ANALYSIS.
```

```
DESCRIPTIVES VARIABLES=VCF0700
```

```
/STATISTICS=MEAN STDDEV MIN MAX.
```

\*Oorspronkelijke variabele voorspellen.

```
FREQUENCIES VARIABLES=voorspellen
```

```
/ORDER=ANALYSIS.
```

```
DESCRIPTIVES VARIABLES=voorspellen
```

```
/STATISTICS=MEAN STDDEV MIN MAX.
```

*Output:*

#### VCF0700 Who Will Be Elected President in November

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 1. Democratic candidate	270	30,0	45,1	45,1
	2 2. Republican candidate	272	30,2	45,4	90,5
	7 7. Other candidate	2	,2	,3	90,8
	8 8. DK; depends	55	6,1	9,2	100,0
	Total	599	66,6	100,0	
Missing	9 9. DK; NA; no Pre IW	1	,1		
	System	300	33,3		
	Total	301	33,4		
Total		900	100,0		

#### Descriptive Statistics

N	Minimum	Maximum	Mean	Std. Deviation
---	---------	---------	------	----------------

VCF0700 Who Will Be Elected President in November	599	1	8	2,12	1,958
Valid N (listwise)	599				

*Toelichting:* In de gebruikte dataset (een cumulatieve file) worden de antwoorden die in het jaar 2012 zijn vergaard niet meegenomen in deze variabele *VCF0700*. De variabele die in 2012 is gebruikt om dit concept te meten, is iets ander gecodeerd. Echter, het is wel mogelijk om de variabele van 2012 samen te voegen met de variabele die in de cumulatieve file is gebruikt voor de jaren 2004 en 2008. De belangrijkste antwoordopties (1=Democratic candidate; 2=Republican candidate) zijn namelijk wel op dezelfde wijze gecodeerd. Daarom heb ik de cumulatieve file gemerged met de aparte dataset van 2012 en vervolgens een nieuwe variabele gemaakt – genaamd *voorspellen* – die de antwoorden in deze jaren combineert. In bovenstaande variabele *VCF0700* zijn de antwoorden van 2012 dus niet meegenomen. Daarom geef ik hieronder ook de frequentietabel en beschrijvende statistieken van de “complete” variabele *voorspellen*. Op basis daarvan zal ik conclusies trekken over de verdeling van de antwoordcategorieën bij deze variabele.

### voorspellen Who does R think will be elected President

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	-9,00 Refused	1	,1	,1	,1
	-8,00 Don't know	1	,1	,1	,3
	1,00 Democratic candidate	330	36,7	47,4	47,7
	2,00 Republican candidate	306	34,0	44,0	91,7
	5,00 Other	1	,1	,1	91,8
	7,00 Other candidate	2	,2	,3	92,1
	8,00 DK; depends	55	6,1	7,9	100,0
	Total	696	77,3	100,0	
Missing	System	204	22,7		
Total		900	100,0		

### Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
voorspellen	696	-9,00	8,00	1,9885	1,93041
Valid N (listwise)	696				

*Toelichting:* Er zijn binnen deze variabele ongeveer evenveel respondenten die verwachten dat de democratische presidentskandidaat zal winnen als respondenten die verwachten dat de republikeinse presidentskandidaat zal winnen.

## **Onderdeel 2: Bewerkingen**

*Toelichting:* In de eerste 2 commando's wordt duidelijk hoe bovenstaande variabele *voorspellen* tot stand is gekomen. Zoals eerder aangegeven zijn er eerst twee verschillende datasets gemerged tot één bestand. Om de scores op de twee afzonderlijke variabelen uit de twee verschillende datasets binnen één variabele opgenomen te krijgen, is de variabele *voorspellen* gemaakt door een gemiddelde te berekenen van de twee afzonderlijke variabelen.

*Syntax:*

\*Dataset 1c: mergen met variabele VCF0006 en preswin\_win van de dataset uit 2012.

```
SORT CASES BY VCF0006.
```

```
DATASET ACTIVATE DataSet6.
```

```
SORT CASES BY VCF0006.
```

```
DATASET ACTIVATE DataSet3.
```

```
MATCH FILES /FILE=*
```

```
  /FILE='DataSet6'
```

```
  /BY VCF0006.
```

```
EXECUTE.
```

\*Variabele *voorspellen* maken van variabele uit totale dataset en variabele uit dataset 2012.

```
COMPUTE voorspellen=mean.1(VCF0700,preswin_win).
```

```
EXECUTE.
```

*Toelichting:* Binnen de oorspronkelijke variabele *voorspellen* werd categorie -9 nog niet herkend als missende waarde. Daarom heb ik deze variabele gehercodeerd waarbij ik -9 hercodeer naar system missing. De overige categorieën worden niet gewijzigd. Deze variabele wordt vervolgens gekopieerd naar de variabele *goed\_voorspeld1* om problemen op te lossen die ervaren werden bij het maken van de uiteindelijke variabele.

De uiteindelijke variabele *goed\_voorspeld* wordt namelijk als nieuwe variabele gemaakt door per verkiezingsjaar (gemeten aan de hand van variabele *VCF0004*) te kijken naar welke presidentskandidaat de verkiezingen heeft gewonnen: de Republikein of de Democraat. De nieuwe variabele *goed\_voorspeld* wordt een dummyvariabele waarbij 0=niet goed voorspeld; 1=goed voorspeld. Aangezien in 2004 de Republikeinse kandidaat de verkiezingen won, krijgt de waarde 2 (=Republican candidate) in de nieuwe variabele de waarde 1. In 2008 en 2012 won de Democratische kandidaat, dus de waarde 1 in de oude variabele *goed\_voorspeld1* krijgt in de nieuwe variabele *goed\_voorspeld* de waarde 1. De overige 4 categorieën (te zien in de frequentietabel van de oorspronkelijke variabele *voorspellen*) worden ook geschaard onder waarde 0 en worden dus gezien als een niet-goede voorspelling. Respondenten die voor deze categorieën kozen gaven aan niet te weten wie de verkiezingen zou winnen of gaven aan dat een andere (derde) kandidaat de verkiezingen zou winnen. Deze derde kandidaat heeft in geen van de verkiezingsjaren gewonnen en wordt daarom telkens onder een niet-goede voorspelling geschaard.

De laatste stap binnen de bewerking van elk van de variabelen is dat de missing data binnen die variabele uit de dataset wordt gefilterd. De steekproefgrootte wijzigt hierdoor van n=900 naar n=399. Voor elke variabele worden de beschrijvende statistieken van de uiteindelijke variabele gegeven nadat de missing data voor elke variabele is verwijderd. Dit zijn immers de variabelen die uiteindelijk zullen worden gebruikt in de regressieanalyses en geven daardoor het meest complete beeld van de variabelen in de dataset.

*Syntax:*

\*Categorie -9.00 (refused) van variabele *voorspellen* hercoderen naar system missing.

```
RECODE voorspellen (1.00=1.00) (2.00=2.00) (5.00=5.00) (7.00=7.00) (8.00=8.00) (-9.00=SYSMIS)
(-8.00=-8.00) INTO voorspellen1.
```

```
EXECUTE.
```

\*Variabele *voorspellen1* kopiëren.

```
COMPUTE goed_voorspeld1=voorspellen1.
```

```
EXECUTE.
```

\*Variabele maken: *goed\_voorspeld*.

```

COMPUTE goed_voorspeld=((VCF0004=2004)&(goed_voorspeld1=2)) |
((VCF0004=2008)&(goed_voorspeld1=1)) | ((VCF0004=2012)&(goed_voorspeld1=1)).

EXECUTE.

```

\*Verwijderen missing data van goed\_voorspeld.

```

DATASET ACTIVATE DataSet1.

```

```

USE ALL.

```

```

COMPUTE filter_$=(goed_voorspeld = 0 | goed_voorspeld = 1).

```

```

VARIABLE LABELS filter_$ 'goed_voorspeld = 0 | goed_voorspeld = 1 (FILTER)'.

```

```

VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.

```

```

FORMATS filter_$ (f1.0).

```

```

SELECT IF filter_$=1.

```

```

EXECUTE.

```

### Onderdeel 3: Uiteindelijke variabele

*Syntax:*

\*Uiteindelijke variabele goed\_voorspeld.

```

FREQUENCIES VARIABLES=goed_voorspeld

```

```

/ORDER=ANALYSIS.

```

```

DESCRIPTIVES VARIABLES=goed_voorspeld

```

```

/STATISTICS=MEAN STDDEV MIN MAX.

```

*Output:*

\*Uiteindelijke variabele goed\_voorspeld.

#### goed\_voorspeld Is de winnaar van de verkiezingen goed voorspeld?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00 Niet goed voorspeld	150	37,6	37,6	37,6
	1,00 Goed voorspeld	249	62,4	62,4	100,0
	Total	399	100,0	100,0	

### Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
goed_voorspeld Is de winnaar van de verkiezingen goed voorspeld?	399	,00	1,00	,6241	,48497
Valid N (listwise)	399				

*Toelichting:* De frequentietabel laat zien dat er meer respondenten zijn die een goede voorspelling hebben gemaakt dan respondenten die een niet-goede voorspelling hebben gemaakt. Dit komt overeen met bevindingen uit bestaande theorie die beschreven zijn in de inleiding.

## 1.2 Variabele Betrokkenheid bij de gemeenschap: betrokken\_gem

### Onderdeel 1: Oorspronkelijke variabele

*Syntax:*

\*Oorspronkelijke variabelen VCF9257 en VCF9258.

FREQUENCIES VARIABLES=VCF9257 VCF9258

/ORDER=ANALYSIS.

DESCRIPTIVES VARIABLES=VCF9257

/STATISTICS=MEAN STDDEV MIN MAX.

DESCRIPTIVES VARIABLES=VCF9258

/STATISTICS=MEAN STDDEV MIN MAX.

*Output:*

\*Oorspronkelijke variabele (item 1) VCF9257.

#### VCF9257 Has R worked with others to deal with issue in community during the past 12 months

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 1. Yes	223	24,8	27,3	27,3
	2 2. No	593	65,9	72,7	100,0
	Total	816	90,7	100,0	
Missing	-9 -9. RF; NA; Inap, no post data	84	9,3		
Total		900	100,0		

#### Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
VCF9257 Has R worked with others to deal with issue in community during the past 12 months	816	1	2	1,73	,446
Valid N (listwise)	816				

\*Oorspronkelijke variabele (item 2) VCF9258.

**VCF9258 Did R attend a meeting about an issue in community or schools during the past 12 months**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 1. Yes	194	21,6	23,8	23,8
	2 2. No	621	69,0	76,2	100,0
	Total	815	90,6	100,0	
Missing	-9 -9. RF; NA; Inap, no post data	84	9,3		
	-8 -8. DK	1	,1		
	Total	85	9,4		
Total		900	100,0		

**Descriptive Statistics**

	N	Minimum	Maximum	Mean	Std. Deviation
VCF9258 Did R attend a meeting about an issue in community or schools during the past 12 months	815	1	2	1,76	,426
Valid N (listwise)	815				

*Toelichting:* Beide items zijn scheef verdeeld. Relatief meer respondenten geven aan niet met anderen te hebben samengewerkt om een kwestie in de gemeenschap te bespreken, of een bijeenkomst bijgewoond te hebben om een kwestie in de gemeenschap of school te bespreken. De gemiddelden (respectievelijk 1,73 en 1,76) liggen hierdoor ook dichterbij de waarde 2 (=no) dan bij de waarde 1 (=yes).

**Onderdeel 2: Bewerkingen**

*Toelichting:* Ik voer een schaalconstructie uit die bestaat uit het maken van een somscore van bovenstaande variabelen. In deze schaalvariabele zal een hogere score een grotere betrokkenheid bij de gemeenschap moeten representeren. Om dit te bereiken zullen eerst de twee aparte items worden gehercodeerd. De betekenis achter de waarde 2 (no) wordt gehercodeerd, zodat 0=no; 1=yes. De waarde 0 staat binnen deze items dan voor 'geen betrokkenheid', terwijl de waarde 1 zal duiden op een betrokkenheid bij de gemeenschap.

Deze gehercodeerde variabelen worden opgenomen in de somscore van de schaalvariabele. De schaalvariabele Betrokkenheid bij de gemeenschap wordt dus gevormd door de twee afzonderlijke, gehercodeerde items bij elkaar op te tellen. De uiteindelijke schaalvariabele bevat dan de mogelijke



waarden 0, 1, 2. Elke score telt het aantal keer dat een respondent 'yes' heeft geantwoord op een van de items in deze schaal.

Om de betrouwbaarheid van deze schaalvariabele te controleren, doe ik ook een betrouwbaarheidsanalyse van de twee items. De cronbach's alpha geeft een waarde van  $\alpha = 0,654$ . Dit betekent dat de mate van interne consistentie twijfelachtig is, maar niet slecht of onacceptabel.

Tot slot verwijder ik in de uiteindelijke schaalvariabele de missing data.

*Syntax:*

\*Hercoderen variabelen VCF9257 en VCF9258 voor schaalvariabele betrokkenheid bij gemeenschap.

```
RECODE VCF9257 VCF9258 (2=0) (1=1) INTO VCF9257_new VCF9258_new.
```

```
EXECUTE.
```

\*Construeren van de schaalvariabele betrokken\_gem (betrokkenheid bij de gemeenschap).

```
COMPUTE betrokken_gem=SUM(VCF9257_new,VCF9258_new).
```

```
EXECUTE.
```

\*Betrouwbaarheidsanalyse met Cronbach's alpha.

```
RELIABILITY
```

```
  /VARIABLES=VCF9257 VCF9258
```

```
  /SCALE('ALL VARIABLES') ALL
```

```
  /MODEL=ALPHA
```

```
  /SUMMARY=TOTAL.
```

\*Verwijderen missing data van betrokken\_gem.

```
USE ALL.
```

```
COMPUTE filter_$=(0 <= betrokken_gem <= 2).
```

```
VARIABLE LABELS filter_$ '0 <= betrokken_gem <= 2 (FILTER)'.  
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.  
FORMATS filter_$ (f1.0).  
SELECT IF filter_$=1.  
EXECUTE.
```

Output:

\*Betrouwbaarheidsanalyse met Cronbach's alpha.

### Reliability Statistics

Cronbach's Alpha	N of Items
,654	2

### Onderdeel 3: Uiteindelijke variabele

Syntax:

\*Uiteindelijke variabele betrokken\_gem.

FREQUENCIES VARIABLES=betrokken\_gem

/ORDER=ANALYSIS.

DESCRIPTIVES VARIABLES=betrokken\_gem

/STATISTICS=MEAN STDDEV MIN MAX.

Output:

\*Uiteindelijke variabele betrokken\_gem.

### betrokken\_gem Betrokkenheid bij de gemeenschap

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00 0x betrokken	256	64,2	64,2	64,2
	1,00 1x betrokken	79	19,8	19,8	84,0
	2,00 2x betrokken	64	16,0	16,0	100,0
	Total	399	100,0	100,0	

### Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
betrokken_gem Betrokkenheid bij de gemeenschap	399	,00	2,00	,5188	,75623
Valid N (listwise)	399				

*Toelichting:* De schaalvariabele is scheef verdeeld. Dit is te verklaren vanuit de scheef verdeelde items waaruit deze schaalvariabele bestaat. Het gemiddelde (0,52) ligt ietwat aan de lage kant op een schaal van 0 tot 2. De standaarddeviatie (0,76) is wel hoog, dus er zit wel redelijk wat spreiding in de scores op deze variabele. In dit onderzoek wordt de variabele vanwege de ordinale categorieën als continue beschouwd.

### 1.3 Variabele Discussiëren over politiek: discuss\_pol

#### Onderdeel 1: Oorspronkelijke variabele

Syntax:

\*Oorspronkelijke variabele discuss\_pol.

FREQUENCIES VARIABLES=VCF0733

/ORDER=ANALYSIS.

DESCRIPTIVES VARIABLES=VCF0733

/STATISTICS=MEAN STDDEV MIN MAX.

Output:

\*Oorspronkelijke variabele discuss\_pol.

#### VCF0733 How Often in the Last Week Respondent Discussed Politics

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 0. None or never discuss politics ('no' to lead-in)	208	23,1	31,6	31,6
	1 1. One day	115	12,8	17,5	49,0
	2 2. Two days	97	10,8	14,7	63,7
	3 3. Three days	67	7,4	10,2	73,9
	4 4. Four days	42	4,7	6,4	80,3
	5 5. Five days	31	3,4	4,7	85,0
	6 6. Six days	12	1,3	1,8	86,8
	7 7. Every day	87	9,7	13,2	100,0
	Total	659	73,2	100,0	
Missing	9 9. DK/NA how often; DK/NA if ever discuss; NA how often	241	26,8		
Total		900	100,0		

#### Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
VCF0733 How Often in the Last Week Respondent Discussed Politics	659	0	7	2,30	2,394
Valid N (listwise)	659				

*Toelichting:* De variabele is lichtelijk scheef verdeeld. Overwegend de meeste respondenten geven aan nooit of enkel één dag per week te discussiëren over politiek. Er zijn desondanks ook nog een redelijk aantal respondenten die aangeven twee, drie of zeven dagen per week te discussiëren over politiek.

## **Onderdeel 2: Bewerkingen**

De missing data van de variabele `discuss_pol` wordt verwijderd. Er blijven 399 respondenten over.

*Syntax:*

\*Verwijderen missing data van `discuss_pol`.

USE ALL.

COMPUTE filter\_\$=(0 <= discuss\_pol <= 7).

VARIABLE LABELS filter\_\$ '0 <= discuss\_pol <= 7 (FILTER)'.  
VALUE LABELS filter\_\$ 0 'Not Selected' 1 'Selected'.

FORMATS filter\_\$ (f1.0).

SELECT IF filter\_\$=1.

EXECUTE.

## **Onderdeel 3: Uiteindelijke variabele**

*Syntax:*

\*Uiteindelijke variabele `discuss_pol`.

FREQUENCIES VARIABLES=`discuss_pol`

/ORDER=ANALYSIS.

DESCRIPTIVES VARIABLES=`discuss_pol`

/STATISTICS=MEAN STDDEV MIN MAX.

*Output:*

\*Uiteindelijke variabele `discuss_pol`.

**`discuss_pol` Aantal dagen discussiëren over politiek in de afgelopen week**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 0. None or never discuss politics ('no' to lead-in)	132	33,1	33,1	33,1
	1 1. One day	66	16,5	16,5	49,6
	2 2. Two days	50	12,5	12,5	62,2
	3 3. Three days	41	10,3	10,3	72,4
	4 4. Four days	27	6,8	6,8	79,2
	5 5. Five days	16	4,0	4,0	83,2
	6 6. Six days	9	2,3	2,3	85,5
	7 7. Every day	58	14,5	14,5	100,0
	Total	399	100,0	100,0	

### Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
discuss_pol Aantal dagen discussiëren over politiek in de afgelopen week	399	0	7	2,35	2,472
Valid N (listwise)	399				

*Toelichting:* Het gemiddelde ( $\mu = 2,35$ ) ligt ietwat aan de lage kant. Overwegend het grootste deel van de data valt aan de onderkant van de schaal. De variabele is een continue variabele.

## 1.4 Variabelen Identificatie met een partij: ident\_W en ident\_L

### Onderdeel 1: Oorspronkelijke variabele

*Syntax:*

\*Oorspronkelijke variabele VCF0301.

FREQUENCIES VARIABLES=VCF0301

/ORDER=ANALYSIS.

DESCRIPTIVES VARIABLES=VCF0301

/STATISTICS=MEAN STDDEV MIN MAX.

*Output:*

\*Oorspronkelijke variabele VCF0301.

#### VCF0301 Party Identification of Respondent- 7-point Scale

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 1. Strong Democrat	195	21,7	22,0	22,0
	2 2. Weak Democrat	146	16,2	16,5	38,5
	3 3. Independent - Democrat	134	14,9	15,1	53,6
	4 4. Independent - Independent	109	12,1	12,3	65,9
	5 5. Independent - Republican	109	12,1	12,3	78,2
	6 6. Weak Republican	90	10,0	10,2	88,4
	7 7. Strong Republican	103	11,4	11,6	100,0
	Total	886	98,4	100,0	
Missing	0 0. DK; NA; other; refused to answer; no Pre IW	14	1,6		
Total		900	100,0		

#### Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
VCF0301 Party Identification of Respondent- 7-point Scale	886	1	7	3,53	2,038
Valid N (listwise)	886				

*Toelichting:* De verdeling over de verschillende categorieën binnen deze variabele is redelijk gelijk. Op een schaal van 1 tot 7 ligt het gemiddelde (3,53) daarom ook in het midden van deze schaal.

## Onderdeel 2: Bewerkingen

*Toelichting:* Het meten van de sterkte van identificatie met een partij van een respondent wordt uitgesplitst naar 'identificatie met de winnende partij' en 'identificatie met de verliezende partij'. Vanuit het theoriehoofdstuk en de opgestelde hypothesen 3 en 4 volgt namelijk dat er verschillende effecten verwacht worden voor deze twee verschillende concepten. Per jaar (2004, 2008 en 2012) wordt er eerst gekeken wie de winnende presidentskandidaat was tijdens de verkiezingen in dat jaar. Hiervoor wordt een variabele geconstrueerd (*WinY*) waarbij 1=republican candidate; en 0=democratic candidate. Vervolgens worden de variabelen *ident\_dem* en *ident\_rep* geconstrueerd. *Ident\_dem* meet de sterkte van identificatie met de democratische partij voor elke respondent. De antwoordcategorieën van de oorspronkelijke variabele *VCF0301* worden gehercodeerd en als volgt ingedeeld: 0=independent/republican; 1=weak democrat; 2=medium democrat; 3=strong democrat. De oorspronkelijke categorieën 4 (independent-independent), 5 (independent-republican), 6 (weak republican) en 7 (strong republican) vallen hierbij onder de nieuwe categorie 0 (independent/republican). De oorspronkelijke categorieën 1 (strong democrat), 2 (weak democrat) en 3 (independent-democrat) worden respectievelijk 3 (strong democrat), 2 (medium democrat) en 1 (weak democrat). Op dezelfde wijze wordt de variabele *ident\_rep* geconstrueerd die de sterkte van identificatie met de republikeinse partij meet.

Aan de hand van de variabelen *WinY*, *ident\_dem* en *ident\_rep* worden vervolgens de uiteindelijke variabelen *ident\_W* (identificatie met de winnende partij) en *ident\_L* (identificatie met de verliezende partij) geconstrueerd. In de jaren dat de democratische kandidaat de verkiezingen won (*WinY=0*), kan de sterkte van identificatie met de winnende partij uitgedrukt worden in de sterkte van identificatie met de democratische partij. De sterkte van identificatie met de verliezende partij staat dan gelijk aan de sterkte van identificatie met de republikeinse partij. Daarom:  $ident\_W = ident\_dem$ ; en  $ident\_L = ident\_rep$ . In het jaar dat de republikeinse kandidaat won (*WinY=1*), wordt de sterkte van identificatie met de winnende partij uitgedrukt in de sterkte van identificatie met de republikeinse partij en de sterkte van identificatie met de verliezende partij in de mate van identificatie met democratische partij.

Daarnaast worden de missende waarden verwijderd bij *ident\_W* en *ident\_L*. Nadat dit voor elke variabele is gedaan blijft er een aantal respondenten van  $n=399$  over in de dataset.



Tot slot worden ook twee interactievariabelen geconstrueerd. Interactie 1 bestaat uit een vermenigvuldiging van *betrokken\_gem* en *ident\_W*. Interactie 2 bestaat uit een vermenigvuldiging van *betrokken\_gem* en *ident\_L*. Deze interacties zullen ook worden meegenomen in de regressieanalyses. In onderstaande syntax wordt weergegeven hoe deze variabelen zijn gemaakt in SPSS.

*Syntax:*

\*Construeren van de variabele WinY.

```
RECODE VCF0004 (2004=1) (2008=0) (2012=0) INTO WinY.
```

```
VARIABLE LABELS WinY 'Winner in year'.
```

```
EXECUTE.
```

\*Construeren van variabele ident\_dem.

```
RECODE VCF0301 (4=0) (5=0) (6=0) (7=0) (3=1) (2=2) (1=3) INTO ident_dem.
```

```
VARIABLE LABELS ident_dem 'Democratische identificatie sterkte'.
```

```
EXECUTE.
```

\*Construeren van variabele ident\_rep.

```
RECODE VCF0301 (4=0) (1=0) (2=0) (3=0) (5=1) (6=2) (7=3) INTO ident_rep.
```

```
VARIABLE LABELS ident_rep 'Republikeinse identificatie sterkte'.
```

```
EXECUTE.
```

\*Construeren van variabelen ident\_W en ident\_L.

```
DO IF WinY=0.
```

```
  COMPUTE ident_W=ident_dem.
```

```
  COMPUTE ident_L=ident_rep.
```

```
END IF.
```

```
EXECUTE.
```

```
DO IF WinY=1.
```

COMPUTE ident\_W=ident\_rep.

COMPUTE ident\_L=ident\_dem.

END IF.

EXECUTE.

\*Verwijderen missing data van ident\_W.

USE ALL.

COMPUTE filter\_\$=(0 <= ident\_W <= 3).

VARIABLE LABELS filter\_\$ '0 <= ident\_W <= 3 (FILTER)'.  
'

VALUE LABELS filter\_\$ 0 'Not Selected' 1 'Selected'.  
'

FORMATS filter\_\$ (f1.0).

SELECT IF filter\_\$=1.

EXECUTE.

\*Verwijderen missing data van ident\_L.

USE ALL.

COMPUTE filter\_\$=(0 <= ident\_L <= 3).

VARIABLE LABELS filter\_\$ '0 <= ident\_L <= 3 (FILTER)'.  
'

VALUE LABELS filter\_\$ 0 'Not Selected' 1 'Selected'.  
'

FORMATS filter\_\$ (f1.0).

SELECT IF filter\_\$=1.

EXECUTE.

\*Construeren van interactievariabelen BETRxIDENT\_W en BETRxIDENT\_L.

COMPUTE BETRxIDENT\_W=betrokken\_gem \* ident\_W.

EXECUTE.

COMPUTE BETRxIDENT\_L=betrokken\_gem \* ident\_L.

EXECUTE.

### Onderdeel 3: Uiteindelijke variabelen

*Syntax:*

\*Uiteindelijke variabele ident\_W.

```
FREQUENCIES VARIABLES=ident_W
```

```
/ORDER=ANALYSIS.
```

```
DESCRIPTIVES VARIABLES=ident_W
```

```
/STATISTICS=MEAN STDDEV MIN MAX.
```

\*Uiteindelijke variabele ident\_L.

```
FREQUENCIES VARIABLES=ident_L
```

```
/ORDER=ANALYSIS.
```

```
DESCRIPTIVES VARIABLES=ident_L
```

```
/STATISTICS=MEAN STDDEV MIN MAX.
```

*Output:*

\*Uiteindelijke variabele ident\_W.

#### ident\_W Sterkte identificatie met winnende partij

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00 Independent / Losing party	228	57,1	57,1	57,1
	1,00 Weak identification	51	12,8	12,8	69,9
	2,00 Medium identification	41	10,3	10,3	80,2
	3,00 Strong identification	79	19,8	19,8	100,0
	Total	399	100,0	100,0	

#### Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
ident_W Sterkte identificatie met winnende partij	399	,00	3,00	,9273	1,21019
Valid N (listwise)	399				

\*Uiteindelijke variabele ident\_L.

### ident\_L Sterkte identificatie met verliezende partij

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00 Independent / Winning party	218	54,6	54,6	54,6
	1,00 Weak identification	59	14,8	14,8	69,4
	2,00 Medium identification	63	15,8	15,8	85,2
	3,00 Strong identification	59	14,8	14,8	100,0
	Total	399	100,0	100,0	

### Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
ident_L Sterkte identificatie met verliezende partij	399	,00	3,00	,9073	1,13595
Valid N (listwise)	399				

*Toelichting:* Zowel de variabele *ident\_W* als *ident\_L* is scheef verdeeld. Dit is gemakkelijk te verklaren, aangezien er onder de categorie 0 steeds 4 oorspronkelijke categorieën vallen. De categorie 0 van beide variabelen bevat ongeveer evenveel respondenten: 228 en 218. Van deze twee aantallen zijn 109 respondenten onafhankelijk. Dit volgt uit de frequentietabel van de oorspronkelijke variabele *VCF0301*, waarin 109 respondenten aangaven onafhankelijk te zijn (categorie 4). Er kan vervolgens ook geconcludeerd worden dat er in de dataset ongeveer evenveel respondenten zijn die zich met de winnende partij identificeren als respondenten die zich met de verliezende partij identificeren. Zowel *ident\_W* als *ident\_L* worden als continue variabelen beschouwd.

## 1.5 Controlevariabele Interesse in de verkiezingen: interesse\_verk

### Onderdeel 1: Oorspronkelijke variabele

*Syntax:*

\*Oorspronkelijke variabele VCF0310.

```
FREQUENCIES VARIABLES=VCF0310
```

```
/ORDER=ANALYSIS.
```

```
DESCRIPTIVES VARIABLES=VCF0310
```

```
/STATISTICS=MEAN STDDEV MIN MAX.
```

*Output:*

\*Oorspronkelijke variabele VCF0310.

#### VCF0310 Interest in the Elections

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 1. Not much interested	111	12,3	14,7	14,7
	2 2. Somewhat interested	354	39,3	47,0	61,8
	3 3. Very much interested	287	31,9	38,1	99,9
	9 9. DK	1	,1	,1	100,0
	Total	753	83,7	100,0	
Missing	0 0. NA; no Pre IW; version B (2008)	147	16,3		
Total		900	100,0		

#### Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
VCF0310 Interest in the Elections	753	1	9	2,24	,732
Valid N (listwise)	753				

*Toelichting:* De controlevariabele Interesse in de verkiezingen is redelijk normaal verdeeld.

Gemiddeld gezien zijn de respondenten in deze steekproef redelijk geïnteresseerd in de verkiezingen: het gemiddelde is 2,24 op een schaal van 1 tot 3.

## Onderdeel 2: Bewerkingen

*Toelichting:* Er is één missende waarde (categorie met waarde 9: DK) die nog niet als zodanig is gecodeerd in de oorspronkelijke variabele. Deze variabele wordt daarom gehercodeerd naar een variabele (*interesse\_verk*) die deze ene missende waarde wel als system missing opneemt. De overige categorieën worden niet veranderd. Vervolgens verwijder ik de missing data van deze variabele.

*Syntax:*

\*Construeren van de nieuwe variabele *interesse\_verk*.

```
RECODE VCF0310 (9=SYSMIS) (1=1) (2=2) (3=3) INTO interesse_verk.
```

```
EXECUTE.
```

\*Verwijderen missing data van *interesse\_verk*.

```
USE ALL.
```

```
COMPUTE filter_$=(1 <= interesse_verk <= 3).
```

```
VARIABLE LABELS filter_$ '1 <= interesse_verk <= 3 (FILTER)'.  
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.  
FORMATS filter_$ (f1.0).  
SELECT IF filter_$=1.  
EXECUTE.
```

## Onderdeel 3: Uiteindelijke variabele

*Syntax:*

\*Uiteindelijke variabele *interesse\_verk*.

```
FREQUENCIES VARIABLES=interesse_verk
```

```
 /ORDER=ANALYSIS.
```

```
DESCRIPTIVES VARIABLES=interesse_verk
```

```
 /STATISTICS=MEAN STDDEV MIN MAX.
```

*Output:*

\*Uiteindelijke variabele *interesse\_verk*.

**interesse\_verk Interesse in de verkiezingen**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00 Not much interested	57	14,3	14,3	14,3
	2,00 Somewhat interested	198	49,6	49,6	63,9
	3,00 Very much interested	144	36,1	36,1	100,0
	Total	399	100,0	100,0	

**Descriptive Statistics**

	N	Minimum	Maximum	Mean	Std. Deviation
interesse_verk Interesse in de verkiezingen	399	1,00	3,00	2,2180	,67629
Valid N (listwise)	399				

*Toelichting:* Het gemiddelde van deze uiteindelijke variabele ( $\mu = 2,22$ ) ligt iets lager dan het gemiddelde van de oorspronkelijke variabele. Echter, er geldt nog steeds dat respondenten over het algemeen redelijke geïnteresseerd zijn in de verkiezingen. Het gemiddelde is immers hoog op een schaal van 1 tot 3. De uiteindelijke variabele wordt als continue variabele beschouwd.

## 1.6 Controlevariabele Opleidingsniveau: opleidingsniveau

### Onderdeel 1: Oorspronkelijke variabele

*Syntax:*

\*Oorspronkelijke variabele VCF0110.

FREQUENCIES VARIABLES=VCF0110

/ORDER=ANALYSIS.

DESCRIPTIVES VARIABLES=VCF0110

/STATISTICS=MEAN STDDEV MIN MAX.

*Output:*

\*Oorspronkelijke variabele VCF0110.

#### VCF0110 Respondent - Education 4-category

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 1. Grade school or less (0-8 grades); 2020: less than high school credential	24	2,7	2,7	2,7
	2 2. High school (12 grades or fewer, incl. non-college	344	38,2	38,5	41,2
	3 3. Some college (13 grades or more but no degree;	281	31,2	31,5	72,7
	4 4. College or advanced degree (no cases 1948)	244	27,1	27,3	100,0
	Total	893	99,2	100,0	
Missing	0 0. DK; NA; no Pre IW; short-form 'new' Cross Section	7	,8		
	Total	900	100,0		

#### Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
VCF0110 Respondent - Education 4-category	893	1	4	2,83	,860
Valid N (listwise)	893				



*Toelichting:* Het valt in de frequentietabel op dat er enkel 24 respondenten vallen onder categorie 1 van deze controlevariabele: weinig respondenten hebben enkel onderwijs gevolgd aan een 'grade school'. Vrijwel alle respondenten hebben minstens onderwijs gevolgd aan een 'high school'. De respondenten zijn daarnaast redelijk gelijk verdeeld over de categorieën 2, 3 en 4.

## **Onderdeel 2: Bewerkingen**

*Toelichting:* Voor deze variabele is enkel de missing data verwijderd als bewerking.

*Syntax:*

\*Verwijderen missing data van opleidingsniveau.

USE ALL.

COMPUTE filter\_\$=(1 <= opleidingsniveau <= 4).

VARIABLE LABELS filter\_\$ '1 <= opleidingsniveau <= 4 (FILTER)'.  
'

VALUE LABELS filter\_\$ 0 'Not Selected' 1 'Selected'.  
'

FORMATS filter\_\$ (f1.0).

SELECT IF filter\_\$=1.

EXECUTE.

## **Onderdeel 3: Uiteindelijke variabele**

*Syntax:*

\*Uiteindelijke variabele opleidingsniveau.

FREQUENCIES VARIABLES=opleidingsniveau

/ORDER=ANALYSIS.

DESCRIPTIVES VARIABLES=opleidingsniveau

/STATISTICS=MEAN STDDEV MIN MAX.

*Output:*

\*Uiteindelijke variabele opleidingsniveau.

**opleidingsniveau Opleidingsniveau**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 1. Grade school or less (0-8 grades); 2020: less than high school credential	11	2,8	2,8	2,8
	2 2. High school (12 grades or fewer, incl. non-college)	158	39,6	39,6	42,4
	3 3. Some college (13 grades or more but no degree;	115	28,8	28,8	71,2
	4 4. College or advanced degree (no cases 1948)	115	28,8	28,8	100,0
	Total	399	100,0	100,0	

### Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
opleidingsniveau	399	1	4	2,84	,877
Opleidingsniveau					
Valid N (listwise)	399				

*Toelichting:* Na het verwijderen van alle missing data ligt het gemiddelde 0,01 hoger in de uiteindelijke variabele. De meeste respondenten hebben een redelijk hoog opleidingsniveau. Enkel 11 respondenten vallen onder de laagste categorie. De variabele *opleidingsniveau* zal in dit onderzoek ook als continue worden beschouwd.

## 1.7 Controlevariabele Leeftijd: leeftijd

### Onderdeel 1: Oorspronkelijke variabele

*Syntax:*

\*Oorspronkelijke variabele VCF0101.

DESCRIPTIVES VARIABLES=VCF0101

/STATISTICS=MEAN STDDEV MIN MAX.

GRAPH

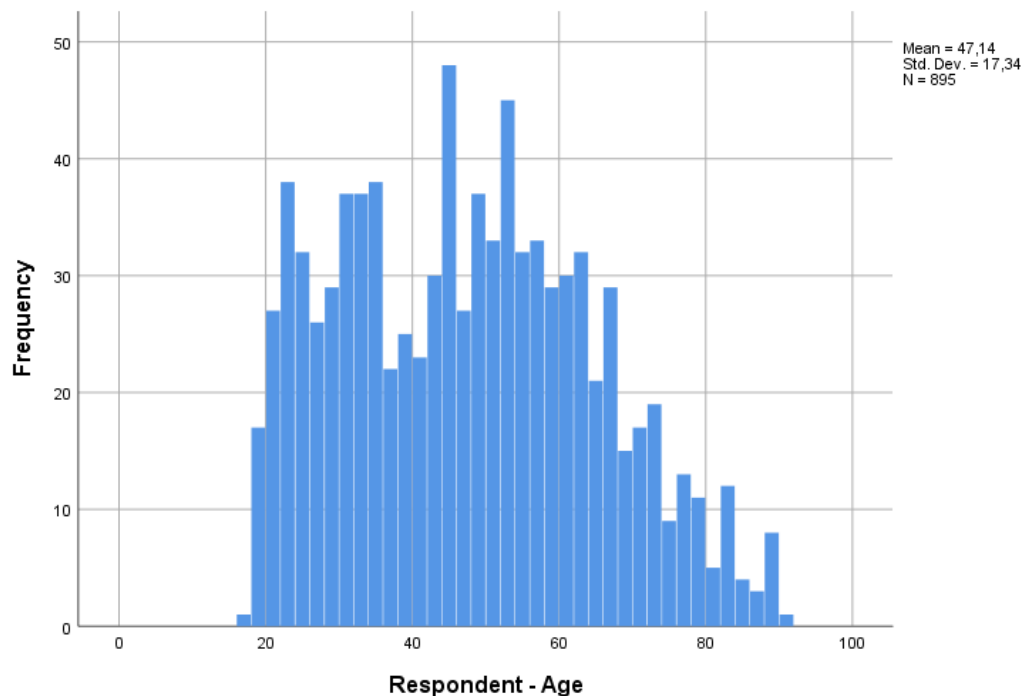
/HISTOGRAM=VCF0101.

*Output:*

\*Oorspronkelijke variabele VCF0101.

#### Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
VCF0101 Respondent - Age	895	17	90	47,14	17,340
Valid N (listwise)	895				



*Toelichting:* De variabele is lichtelijk scheef verdeeld. Dit is te verklaren aan de hand van het feit dat Amerikaanse burgers onder de 18 jaar geen kiesgerechtigden zijn en daarom niet onder de onderzoekspopulatie vallen en dus ook niet opgenomen in de steekproef opgenomen zijn. De

linkerkant van de histogram maakt hierdoor de klokvormige vorm van een normale verdeling niet compleet. Het minimum van 17 en de zeer kleine staaf uiterst links in de histogram laten zien dat er waarschijnlijk een enkele respondent is geweest die 17 jaar was ten tijde van het pre-verkiezingsinterview maar nog 18 jaar werd vóór of op de dag van de verkiezingen.

## Onderdeel 2: Bewerkingen

*Toelichting:* De missing data wordt verwijderd.

*Syntax:*

\*Verwijderen missing data van leeftijd.

```
DATASET ACTIVATE DataSet1.
```

```
USE ALL.
```

```
COMPUTE filter_$=(18 <= leeftijd <= 90).
```

```
VARIABLE LABELS filter_$ '18 <= leeftijd <= 90 (FILTER)'.  
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.  
FORMATS filter_$ (f1.0).  
SELECT IF filter_$=1.  
EXECUTE.
```

## Onderdeel 3: Uiteindelijke variabele

*Syntax:*

\*Uiteindelijke variabele leeftijd.

```
DESCRIPTIVES VARIABLES=leeftijd
```

```
  /STATISTICS=MEAN STDDEV MIN MAX.
```

```
GRAPH
```

```
  /HISTOGRAM=leeftijd.
```

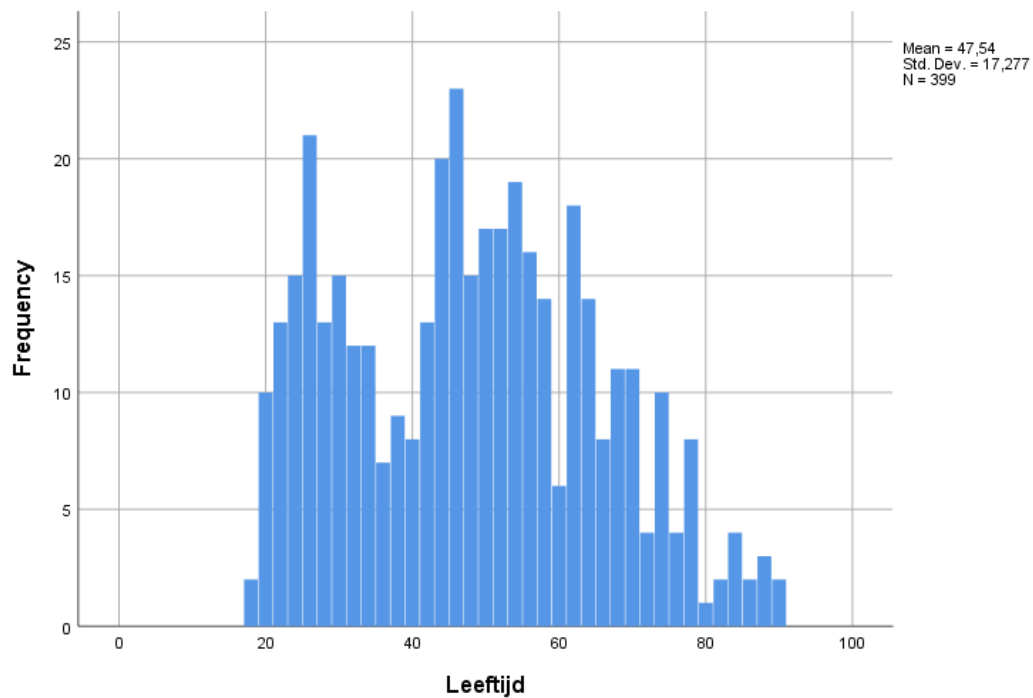
*Output:*

\*Uiteindelijke variabele leeftijd.

### Descriptive Statistics

N	Minimum	Maximum	Mean	Std. Deviation
---	---------	---------	------	----------------

leeftijd Leeftijd	399	18	90	47,54	17,277
Valid N (listwise)	399				



*Toelichting:* De laagst gescoorde waarde is in de uiteindelijke variabele 18. Het gemiddelde blijft vrijwel gelijk. Het verwijderen van de missing data van alle variabelen heeft dus vrijwel geen gevolgen gehad voor de verdeling van de variabele leeftijd.

## 1.8 Controlevariabele Geslacht: vrouw

### Onderdeel 1: Oorspronkelijke variabele

*Syntax:*

\*Oorspronkelijke variabele VCF0104.

FREQUENCIES VARIABLES=VCF0104

/ORDER=ANALYSIS.

DESCRIPTIVES VARIABLES=VCF0104

/STATISTICS=MEAN STDDEV MIN MAX.

*Output:*

\*Oorspronkelijke variabele VCF0104

#### VCF0104 Respondent - Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 1. Male	391	43,4	43,4	43,4
	2 2. Female	509	56,6	56,6	100,0
	Total	900	100,0	100,0	

#### Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
VCF0104 Respondent - Gender	900	1	2	1,57	,496
Valid N (listwise)	900				

*Toelichting:* Aan de hand van de frequentietabel kan worden geconcludeerd dat de variabele een beetje scheef verdeeld is: het percentage vrouwen dat is opgenomen in deze steekproef is hoger dan het percentage mannen.

### Onderdeel 2: Bewerkingen

*Toelichting:* Er zijn geen bewerkingen uitgevoerd op deze variabele. De variabele heeft enkel een andere naam gekregen. De beschrijvende statistieken zijn veranderd door het verwijderen van missing data van andere variabelen. Deze variabelen bevatte zelf geen missing data.

### Onderdeel 3: Uiteindelijke variabele

*Syntax:*

\*Uiteindelijke variabele vrouw.

```
FREQUENCIES VARIABLES=vrouw
```

```
/ORDER=ANALYSIS.
```

```
DESCRIPTIVES VARIABLES=vrouw
```

```
/STATISTICS=MEAN STDDEV MIN MAX.
```

*Output:*

\*Uiteindelijke variabele vrouw.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 1. Male	177	44,4	44,4	44,4
	2 2. Female	222	55,6	55,6	100,0
	Total	399	100,0	100,0	

	N	Minimum	Maximum	Mean	Std. Deviation
vrouw Geslacht	399	1	2	1,56	,497
Valid N (listwise)	399				

*Toelichting:* Het verwijderen van missing data van andere variabelen heeft vrijwel geen gevolgen gehad voor de man-vrouw verdeling in de dataset.

## Bijlage 2

### 2.1 Analyse van de beschrijvende statistieken (tabel 1)

Syntax:

\*Tabellen met beschrijvende statistieken.

```
FREQUENCIES VARIABLES=goed_voorspeld vrouw
```

```
/ORDER=ANALYSIS.
```

```
DESCRIPTIVES VARIABLES=betrokken_gem discuss_pol ident_W ident_L interesse_verk  
opleidingsniveau leeftijd
```

```
/STATISTICS=MEAN STDDEV RANGE MIN MAX.
```

Output:

\*Tabellen met beschrijvende statistieken.

#### goed\_voorspeld Is de winnaar van de verkiezingen goed voorspeld?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00 Niet goed voorspeld	150	37,6	37,6	37,6
	1,00 Goed voorspeld	249	62,4	62,4	100,0
	Total	399	100,0	100,0	

#### vrouw Geslacht

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 1. Male	177	44,4	44,4	44,4
	2 2. Female	222	55,6	55,6	100,0
	Total	399	100,0	100,0	

#### Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation
betrokken_gem Betrokkenheid bij de gemeenschap	399	2,00	,00	2,00	,5188	,75623
discuss_pol Aantal dagen discussiëren over politiek in de afgelopen week	399	7	0	7	2,35	2,472



ident_W Sterkte identificatie met winnende partij	399	3,00	,00	3,00	,9273	1,21019
ident_L Sterkte identificatie met verliezende partij	399	3,00	,00	3,00	,9073	1,13595
interesse_verk Interesse in de verkiezingen	399	2,00	1,00	3,00	2,2180	,67629
opleidingsniveau Opleidingsniveau	399	3	1	4	2,84	,877
leeftijd Leeftijd	399	72	18	90	47,54	17,277
Valid N (listwise)	399					

*Toelichting:* Voor de variabelen die ik als continue variabelen beschouw, heb ik de volgende beschrijvende statistieken berekend: gemiddelde, standaarddeviatie, range, minimum, maximum en het aantal respondenten binnen de betreffende variabelen. Voor de overige (binaire) variabelen – de afhankelijke variabele en de controlevariabele leeftijd – heb ik de percentages berekend behorende bij de mogelijke scores en wederom het aantal respondenten binnen de variabelen. Deze statistieken heb ik in tabel 1 verwerkt die te vinden is in hoofdstuk 4.1 Beschrijvende statistieken.

In de frequentietabel van goed\_voorspeld is te zien dat het grootste deel van de respondenten in de dataset de verkiezingswinnaar goed konden voorspellen: 62,4% goed voorspeld en 37,6% niet goed voorspeld. De gemiddelde betrokkenheid bij de gemeenschap van de respondenten ligt ietwat aan de lage kant ( $\mu = 0,52$ ). De schaal varieerde immers van 0 tot 2. Het gemiddelde aantal dagen dat de respondenten over politiek discussiëren per week is 2,35. De spreiding in deze variabele (*discuss\_pol*) is hoog ( $SD = 2,47$ ). De mate van identificatie die respondenten aangeven te hebben met een politieke partij ligt aan de zwakke kant. De gemiddeldes van *ident\_W* en *ident\_L* zijn respectievelijk  $\mu = 0,93$  en  $\mu = 0,91$ . Op een schaal van 0 tot 3 zijn deze redelijk laag. De interesse in de verkiezingen van de respondenten is gemiddeld redelijk hoog ( $\mu = 2,22$ ). Het gemiddelde opleidingsniveau van de respondenten is ook redelijk hoog ( $\mu = 2,84$ ). De gemiddelde leeftijd is 47,54 jaar. De jongste respondent is 18 jaar, de oudste respondent is 90 jaar. Het percentage vrouwen (55,60%) is iets hoger dan het percentage mannen (44,40%).

## 2.2 Associatiematen (tabel 2)

*Syntax:*

\*Kolom Correct voorspellen van de verkiezingsuitslag, ANOVA.

UNIANOVA betrokken\_gem BY goed\_voorspeld

/METHOD=SSTYPE(3)

```
/INTERCEPT=INCLUDE  
/CRITERIA=ALPHA(0.05)  
/DESIGN=goed_voorspeld.
```

UNIANOVA discuss\_pol BY goed\_voorspeld

```
/METHOD=SSTYPE(3)  
/INTERCEPT=INCLUDE  
/CRITERIA=ALPHA(0.05)  
/DESIGN=goed_voorspeld.
```

UNIANOVA ident\_W BY goed\_voorspeld

```
/METHOD=SSTYPE(3)  
/INTERCEPT=INCLUDE  
/CRITERIA=ALPHA(0.05)  
/DESIGN=goed_voorspeld.
```

UNIANOVA ident\_L BY goed\_voorspeld

```
/METHOD=SSTYPE(3)  
/INTERCEPT=INCLUDE  
/CRITERIA=ALPHA(0.05)  
/DESIGN=goed_voorspeld.
```

UNIANOVA interesse\_verk BY goed\_voorspeld

```
/METHOD=SSTYPE(3)  
/INTERCEPT=INCLUDE  
/CRITERIA=ALPHA(0.05)  
/DESIGN=goed_voorspeld.
```

UNIANOVA opleidingsniveau BY goed\_voorspeld

```
/METHOD=SSTYPE(3)  
/INTERCEPT=INCLUDE
```

```
/CRITERIA=ALPHA(0.05)  
/DESIGN=goed_voorspeld.
```

UNIANOVA leeftijd BY goed\_voorspeld

```
/METHOD=SSTYPE(3)  
/INTERCEPT=INCLUDE  
/CRITERIA=ALPHA(0.05)  
/DESIGN=goed_voorspeld.
```

\*Rij Vrouw, ANOVA.

UNIANOVA betrokken\_gem BY vrouw

```
/METHOD=SSTYPE(3)  
/INTERCEPT=INCLUDE  
/CRITERIA=ALPHA(0.05)  
/DESIGN=vrouw.
```

UNIANOVA discuss\_pol BY vrouw

```
/METHOD=SSTYPE(3)  
/INTERCEPT=INCLUDE  
/CRITERIA=ALPHA(0.05)  
/DESIGN=vrouw.
```

UNIANOVA ident\_W BY vrouw

```
/METHOD=SSTYPE(3)  
/INTERCEPT=INCLUDE  
/CRITERIA=ALPHA(0.05)  
/DESIGN=vrouw.
```

UNIANOVA ident\_L BY vrouw

```
/METHOD=SSTYPE(3)
```

```
/INTERCEPT=INCLUDE  
/CRITERIA=ALPHA(0.05)  
/DESIGN=vrouw.
```

UNIANOVA interesse\_verk BY vrouw

```
/METHOD=SSTYPE(3)  
/INTERCEPT=INCLUDE  
/CRITERIA=ALPHA(0.05)  
/DESIGN=vrouw.
```

UNIANOVA opleidingsniveau BY vrouw

```
/METHOD=SSTYPE(3)  
/INTERCEPT=INCLUDE  
/CRITERIA=ALPHA(0.05)  
/DESIGN=vrouw.
```

UNIANOVA leeftijd BY vrouw

```
/METHOD=SSTYPE(3)  
/INTERCEPT=INCLUDE  
/CRITERIA=ALPHA(0.05)  
/DESIGN=vrouw.
```

\*Correlaties tussen continue variabelen, t-toets voor gemiddelden.

CORRELATIONS

```
/VARIABLES=betrokken_gem discuss_pol ident_W ident_L interesse_verk opleidingsniveau leeftijd  
/PRINT=TWOTAIL NOSIG  
/STATISTICS DESCRIPTIVES  
/MISSING=PAIRWISE.
```

\*Correlatie tussen twee categorische variabelen, Cramer's V.

CROSSTABS

/TABLES=goed\_voorspeld BY vrouw  
 /FORMAT=AVALUE TABLES  
 /STATISTICS=PHI  
 /CELLS=COUNT  
 /COUNT ROUND CELL.

Output:

### Tests of Between-Subjects Effects

Dependent Variable: betrokken\_gem Betrokkenheid bij de gemeenschap

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	,287 <sup>a</sup>	1	,287	,501	,480
Intercept	103,465	1	103,465	180,693	,000
goed_voorspeld	,287	1	,287	,501	,480
Error	227,322	397	,573		
Total	335,000	399			
Corrected Total	227,609	398			

a. R Squared = ,001 (Adjusted R Squared = -,001)

### Tests of Between-Subjects Effects

Dependent Variable: discuss\_pol Aantal dagen discussiëren over politiek in de afgelopen week

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	12,164 <sup>a</sup>	1	12,164	1,995	,159
Intercept	2144,355	1	2144,355	351,721	,000
goed_voorspeld	12,164	1	12,164	1,995	,159
Error	2420,412	397	6,097		
Total	4633,000	399			
Corrected Total	2432,576	398			

a. R Squared = ,005 (Adjusted R Squared = ,002)

### Tests of Between-Subjects Effects

Dependent Variable: ident\_W Sterkte identificatie met winnende partij

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	94,589 <sup>a</sup>	1	94,589	76,903	,000

Intercept	241,206	1	241,206	196,105	,000
goed_voorspeld	94,589	1	94,589	76,903	,000
Error	488,303	397	1,230		
Total	926,000	399			
Corrected Total	582,892	398			

a. R Squared = ,162 (Adjusted R Squared = ,160)

### Tests of Between-Subjects Effects

Dependent Variable: ident\_L Sterkte identificatie met verliezende partij

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	94,211 <sup>a</sup>	1	94,211	89,189	,000
Intercept	398,572	1	398,572	377,323	,000
goed_voorspeld	94,211	1	94,211	89,189	,000
Error	419,357	397	1,056		
Total	842,000	399			
Corrected Total	513,569	398			

a. R Squared = ,183 (Adjusted R Squared = ,181)

### Tests of Between-Subjects Effects

Dependent Variable: interesse\_verk Interesse in de verkiezingen

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	,018 <sup>a</sup>	1	,018	,039	,844
Intercept	1844,970	1	1844,970	4024,198	,000
goed_voorspeld	,018	1	,018	,039	,844
Error	182,012	397	,458		
Total	2145,000	399			
Corrected Total	182,030	398			

a. R Squared = ,000 (Adjusted R Squared = -,002)

### Tests of Between-Subjects Effects

Dependent Variable: opleidingsniveau Opleidingsniveau

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	,063 <sup>a</sup>	1	,063	,082	,775
Intercept	3020,735	1	3020,735	3914,611	,000
goed_voorspeld	,063	1	,063	,082	,775

Error	306,348	397	,772		
Total	3518,000	399			
Corrected Total	306,411	398			

a. R Squared = ,000 (Adjusted R Squared = -,002)

### Tests of Between-Subjects Effects

Dependent Variable: leeftijd Leeftijd

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	125,943 <sup>a</sup>	1	125,943	,421	,517
Intercept	841266,605	1	841266,605	2814,358	,000
goed_voorspeld	125,943	1	125,943	,421	,517
Error	118671,040	397	298,919		
Total	1020704,000	399			
Corrected Total	118796,982	398			

a. R Squared = ,001 (Adjusted R Squared = -,001)

### Tests of Between-Subjects Effects

Dependent Variable: betrokken\_gem Betrokkenheid bij de gemeenschap

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	,237 <sup>a</sup>	1	,237	,413	,521
Intercept	104,898	1	104,898	183,156	,000
vrouw	,237	1	,237	,413	,521
Error	227,372	397	,573		
Total	335,000	399			
Corrected Total	227,609	398			

a. R Squared = ,001 (Adjusted R Squared = -,001)

### Tests of Between-Subjects Effects

Dependent Variable: discuss\_pol Aantal dagen discussiëren over politiek in de afgelopen week

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1,895 <sup>a</sup>	1	1,895	,310	,578
Intercept	2157,985	1	2157,985	352,461	,000
vrouw	1,895	1	1,895	,310	,578
Error	2430,681	397	6,123		
Total	4633,000	399			
Corrected Total	2432,576	398			

a. R Squared = ,001 (Adjusted R Squared = -,002)

### Tests of Between-Subjects Effects

Dependent Variable: ident\_W Sterkte identificatie met winnende partij

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	,013 <sup>a</sup>	1	,013	,009	,925
Intercept	338,269	1	338,269	230,395	,000
vrouw	,013	1	,013	,009	,925
Error	582,879	397	1,468		
Total	926,000	399			
Corrected Total	582,892	398			

a. R Squared = ,000 (Adjusted R Squared = -,002)

### Tests of Between-Subjects Effects

Dependent Variable: ident\_L Sterkte identificatie met verliezende partij

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7,727 <sup>a</sup>	1	7,727	6,065	,014
Intercept	313,061	1	313,061	245,700	,000
vrouw	7,727	1	7,727	6,065	,014
Error	505,841	397	1,274		
Total	842,000	399			
Corrected Total	513,569	398			

a. R Squared = ,015 (Adjusted R Squared = ,013)

### Tests of Between-Subjects Effects

Dependent Variable: interesse\_verk Interesse in de verkiezingen

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	,417 <sup>a</sup>	1	,417	,911	,340
Intercept	1944,417	1	1944,417	4250,422	,000
vrouw	,417	1	,417	,911	,340
Error	181,613	397	,457		
Total	2145,000	399			
Corrected Total	182,030	398			

a. R Squared = ,002 (Adjusted R Squared = ,000)

### Tests of Between-Subjects Effects



Dependent Variable: opleidingsniveau Opleidingsniveau

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	,102 <sup>a</sup>	1	,102	,132	,717
Intercept	3166,688	1	3166,688	4104,267	,000
vrouw	,102	1	,102	,132	,717
Error	306,309	397	,772		
Total	3518,000	399			
Corrected Total	306,411	398			

a. R Squared = ,000 (Adjusted R Squared = -,002)

### Tests of Between-Subjects Effects

Dependent Variable: leeftijd Leeftijd

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	139,627 <sup>a</sup>	1	139,627	,467	,495
Intercept	887921,642	1	887921,642	2970,780	,000
vrouw	139,627	1	139,627	,467	,495
Error	118657,355	397	298,885		
Total	1020704,000	399			
Corrected Total	118796,982	398			

a. R Squared = ,001 (Adjusted R Squared = -,001)

### Correlations

		discuss_p ol Aantal dagen	ident_W Sterkte	ident_L Sterkte	interesse_v erk	opleidingsnive au	leeftij d
betrokken_g em Betrokkenhe id bij de gemeenscha p	Pearson Correlati on	1	,149**	,014	,033	,132**	,241**
	Sig. (2- tailed)		,003	,783	,514	,008	,434
	N	399	399	399	399	399	399

discuss_pol	Pearson	,149**	1	,096	,038	,405**	,190**	,129**
Aantal dagen discussiëren over politiek in de afgelopen week	Correlatie							
	Sig. (2-tailed)	,003		,056	,445	,000	,000	,010
	N	399	399	399	399	399	399	399
ident_W	Pearson	,014	,096	1	-,614**	,173**	,055	,127*
Sterkte identificatie met winnende partij	Correlatie							
	Sig. (2-tailed)	,783	,056		,000	,001	,272	,011
	N	399	399	399	399	399	399	399
ident_L	Pearson	,033	,038	-,614**	1	,079	,007	,060
Sterkte identificatie met verliezende partij	Correlatie							
	Sig. (2-tailed)	,514	,445	,000		,116	,881	,234
	N	399	399	399	399	399	399	399
interesse_verk	Pearson	,132**	,405**	,173**	,079	1	,149**	,168**
Interesse in de verkiezingen	Correlatie							
	Sig. (2-tailed)	,008	,000	,001	,116		,003	,001
	N	399	399	399	399	399	399	399
opleidingsniveau	Pearson	,241**	,190**	,055	,007	,149**	1	-,105*
Opleidingsniveau	Correlatie							
	Sig. (2-tailed)	,000	,000	,272	,881	,003		,036
	N	399	399	399	399	399	399	399
leeftijd	Pearson	,039	,129**	,127*	,060	,168**	-,105*	1
Leeftijd	Correlatie							
	Sig. (2-tailed)	,434	,010	,011	,234	,001	,036	
	N	399	399	399	399	399	399	399

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

### Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	-,016	,748
	Cramer's V	,016	,748
N of Valid Cases		399	

*Toelichting:* tussen de variabelen die als continu worden beschouwd wordt een correlatie op basis van een t-toets voor verschil in gemiddelden berekend. Tussen de categorische afhankelijke variabele *goed\_voorspeld* en categorische controlevariabele *vrouw* wordt een correlatie op basis van Cramer's V berekend. De correlaties tussen een continue variabele en een van de categorische variabelen (*goed\_voorspeld* of *vrouw*) worden berekend op basis van ANOVA. De correlatie  $r$  kan in de ANOVA-tabellen worden afgeleid van de  $R^2$ .

De associatiematen bieden inzicht in de sterkte van de verbanden tussen de variabelen. De samenhang tussen Goed voorspellen van de winnaar en Betrokkenheid bij de gemeenschap blijkt zwak:  $r = \sqrt{0,001} = 0,032$ . De correlatie is niet significant ( $p = 0,480$ ). De correlatie tussen Goed voorspellen van de winnaar en Discussiëren over politiek is ook zwak en niet-significant ( $r = \sqrt{0,005} = 0,071$ ;  $p = 0,159$ ). De correlaties van Goed voorspellen van de winnaar met Identificatie met de winnende partij ( $r = 0,402$ ;  $p < 0,001$ ) en Identificatie met de verliezende partij ( $r = 0,428$ ;  $p < 0,001$ ) zijn wel sterk en significant. De samenhang van Goed voorspellen van de winnaar met de controlevariabelen blijken allen zwak en niet-significant.

De samenhang tussen Betrokkenheid bij de gemeenschap en Discussiëren over politiek is lichtelijk zwak ( $r = 0,149$ ), maar significant op 0,01. Betrokkenheid bij de gemeenschap blijkt redelijk sterk (en significant) gecorreleerd te zijn met Opleidingsniveau ( $r = 0,241$ ). Discussiëren over politiek hangt ook sterk samen met Interesse in de verkiezingen ( $r = 0,405$ ). De hoge correlatie tussen Identificatie met de winnende partij en Identificatie met de verliezende partij is daarnaast ook opvallend ( $r = -0,614$ ). Deze is gemakkelijk te verklaren: mensen die zich sterk identificeren met de verliezende partij scoren automatisch laag op Identificatie met de winnende partij. De correlatie is echter niet 1: mensen die zichzelf als onafhankelijk zijn scoren op beide variabelen (zowel *ident\_W* als *ident\_L*) laag (*ident\_W* = 0; *ident\_L* = 0). Tot slot is de samenhang tussen Identificatie met de winnende partij en Interesse in de verkiezingen opvallend. De correlatie is lichtelijk sterk en significant ( $r = 0,173$ ;  $p < 0,01$ ).

## 2.3 Regressieanalyses (tabel 3 en 4)

*Syntax:*

\*Model 1 + Model 2 + Model 5.

```
LOGISTIC REGRESSION VARIABLES goed_voorspeld  
/METHOD=ENTER leeftijd vrouw opleidingsniveau interesse_verk  
/METHOD=ENTER betrokken_gem  
/METHOD=ENTER discuss_pol  
/SAVE=LEVER DFBETA  
/PRINT=GOODFIT  
/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

\*Model 3.

```
REGRESSION  
/DESCRIPTIVES MEAN STDDEV CORR SIG N  
/MISSING LISTWISE  
/STATISTICS COEFF OUTS CI(95) R ANOVA CHANGE  
/CRITERIA=PIN(.05) POUT(.10)  
/NOORIGIN  
/DEPENDENT discuss_pol  
/METHOD=ENTER leeftijd vrouw opleidingsniveau interesse_verk.
```

\*Model 4.

```
REGRESSION  
/DESCRIPTIVES MEAN STDDEV CORR SIG N  
/MISSING LISTWISE  
/STATISTICS COEFF OUTS CI(95) R ANOVA CHANGE  
/CRITERIA=PIN(.05) POUT(.10)  
/NOORIGIN  
/DEPENDENT discuss_pol  
/METHOD=ENTER leeftijd vrouw opleidingsniveau interesse_verk betrokken_gem.
```

\*Model 6 + Model 7.

LOGISTIC REGRESSION VARIABLES goed\_voorspeld

/METHOD=ENTER leeftijd vrouw opleidingsniveau interesse\_verk betrokken\_gem ident\_W ident\_L

/METHOD=ENTER BETRxIDENT\_W

/SAVE=LEVER DFBETA

/PRINT=GOODFIT

/CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

\*(Model 6 +) Model 8.

LOGISTIC REGRESSION VARIABLES goed\_voorspeld

/METHOD=ENTER leeftijd vrouw opleidingsniveau interesse\_verk betrokken\_gem ident\_W ident\_L

/METHOD=ENTER BETRxIDENT\_L

/SAVE=LEVER DFBETA

/PRINT=GOODFIT

/CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

*Output:*

\*Model 1 + Model 2 + Model 5.

\*Model 1.

### Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	,677	4	,954
	Block	,677	4	,954
	Model	,677	4	,954

### Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	527,632 <sup>a</sup>	,002	,002

a. Estimation terminated at iteration number 3 because parameter estimates changed by less than ,001.

### Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	12,128	8	,146

### Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup>	Leeftijd	,004	,006	,463	1	,496	1,004
	Geslacht	-,074	,209	,126	1	,722	,928
	Opleidingsniveau	-,019	,121	,025	1	,873	,981
	Interesse in de verkiezingen	-,047	,158	,090	1	,765	,954
	Constant	,584	,622	,880	1	,348	1,793

a. Variable(s) entered on step 1: Leeftijd , Geslacht , Opleidingsniveau, Interesse in de verkiezingen.

\*Model 2.

### Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	,447	1	,504
	Block	,447	1	,504
	Model	1,124	5	,952

### Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	527,185 <sup>a</sup>	,003	,004

a. Estimation terminated at iteration number 3 because parameter estimates changed by less than ,001.

### Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	6,454	8	,597

### Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup>	Leeftijd	,004	,006	,508	1	,476	1,004
	Geslacht	-,070	,209	,112	1	,737	,932
	Opleidingsniveau	,000	,124	,000	1	,998	1,000
	Interesse in de verkiezingen	-,038	,159	,056	1	,812	,963

Betrokkenheid bij de gemeenschap	-,094	,141	,449	1	,503	,910
Constant	,542	,625	,752	1	,386	1,720

a. Variable(s) entered on step 1: Betrokkenheid bij de gemeenschap.

\*Model 5.

### Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	2,075	1	,150
	Block	2,075	1	,150
	Model	3,198	6	,784

### Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	525,111 <sup>a</sup>	,008	,011

a. Estimation terminated at iteration number 3 because parameter estimates changed by less than ,001.

### Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	5,320	8	,723

### Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup>	Leeftijd	,005	,006	,696	1	,404	1,005
	Geslacht	-,058	,210	,075	1	,784	,944
	Opleidingsniveau	,024	,125	,035	1	,851	1,024
	Interesse in de verkiezingen	,053	,172	,096	1	,757	1,055
	Betrokkenheid bij de gemeenschap	-,081	,142	,325	1	,569	,922
	Aantal dagen discussiëren over politiek in de afgelopen week	-,067	,046	2,077	1	,150	,935
	Constant	,368	,637	,334	1	,563	1,445

a. Variable(s) entered on step 1: Aantal dagen discussiëren over politiek in de afgelopen week.

\*Model 3.

### Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics	
						F Change	df1
1	,435 <sup>a</sup>	,189	,181	2,237	,189	23,015	4

a. Predictors: (Constant), interesse\_verk Interesse in de verkiezingen, vrouw Geslacht , opleidingsniveau Opleidingsniveau, leeftijd

### ANOVA<sup>a</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	460,736	4	115,184	23,015	,000 <sup>b</sup>
	Residual	1971,841	394	5,005		
	Total	2432,576	398			

a. Dependent Variable: discuss\_pol Aantal dagen discussiëren over politiek in de afgelopen week

b. Predictors: (Constant), interesse\_verk Interesse in de verkiezingen, vrouw Geslacht , opleidingsniveau Opleidingsniveau, leeftijd Leeftijd

### Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B	
		B	Std. Error				Lower Bound	Upper Bound
1	(Constant)	-2,665	,673		-3,963	,000	-3,987	-1,343
	leeftijd Leeftijd	,011	,007	,080	1,727	,085	-,002	,025
	vrouw Geslacht	,201	,226	,040	,889	,375	-,244	,645
	opleidingsniveau Opleidingsniveau	,399	,130	,142	3,061	,002	,143	,656
	interesse_verk Interesse in de verkiezingen	1,362	,171	,373	7,968	,000	1,026	1,698

a. Dependent Variable: discuss\_pol Aantal dagen discussiëren over politiek in de afgelopen week

\*Model 4.

### Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics	
						F Change	df1
1	,440 <sup>a</sup>	,193	,183	2,234	,193	18,850	5



a. Predictors: (Constant), betrokken\_gem Betrokkenheid bij de gemeenschap, vrouw Geslacht , leeftijd Leeftijd , interesse\_verk Interesse in de verkiezingen, opleidingsniveau Opleidingsniveau

### ANOVA<sup>a</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	470,535	5	94,107	18,850	,000 <sup>b</sup>
	Residual	1962,041	393	4,992		
	Total	2432,576	398			

a. Dependent Variable: discuss\_pol Aantal dagen discussiëren over politiek in de afgelopen week

b. Predictors: (Constant), betrokken\_gem Betrokkenheid bij de gemeenschap, vrouw Geslacht , leeftijd Leeftijd , interesse\_verk Interesse in de verkiezingen, opleidingsniveau Opleidingsniveau

### Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-2,569	,675		-3,805	,000	-3,897	-1,242
	leeftijd Leeftijd	,011	,007	,077	1,660	,098	-,002	,024
	vrouw Geslacht	,191	,226	,038	,845	,399	-,253	,635
	opleidingsniveau Opleidingsniveau	,356	,134	,126	2,661	,008	,093	,620
	interesse_verk Interesse in de verkiezingen	1,340	,171	,367	7,817	,000	1,003	1,678
	betrokken_gem Betrokkenheid bij de gemeenschap	,215	,154	,066	1,401	,162	-,087	,517

a. Dependent Variable: discuss\_pol Aantal dagen discussiëren over politiek in de afgelopen week

\*Model 6 + Model 7.

\*Model 6.

### Omnibus Tests of Model Coefficients

Step 1		Chi-square	df	Sig.
Step 1	Step	93,927	7	,000
	Block	93,927	7	,000
	Model	93,927	7	,000

### Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	434,382 <sup>a</sup>	,210	,286

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than ,001.

### Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	6,759	8	,563

### Variables in the Equation

Step 1 <sup>a</sup>		B	S.E.	Wald	df	Sig.	Exp(B)
	Leeftijd	,004	,007	,326	1	,568	1,004
	Geslacht	,103	,240	,183	1	,669	1,108
	Opleidingsniveau	-,041	,141	,083	1	,773	,960
	Interesse in de verkiezingen	-,103	,188	,298	1	,585	,902
	Betrokkenheid bij de gemeenschap	-,084	,160	,275	1	,600	,919
	Sterkte identificatie met winnende partij	,594	,153	15,178	1	,000	1,811
	Sterkte identificatie met verliezende partij	-,497	,130	14,674	1	,000	,608
	Constant	,616	,693	,789	1	,374	1,851

a. Variable(s) entered on step 1: Leeftijd , Geslacht , Opleidingsniveau, Interesse in de verkiezingen, Betrokkenheid bij de gemeenschap, Sterkte identificatie met winnende partij, Sterkte identificatie met verliezende partij.

\*Model 7.

### Omnibus Tests of Model Coefficients

Step 1		Chi-square	df	Sig.
	Step	,017	1	,895
	Block	,017	1	,895
	Model	93,944	8	,000

### Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	434,365 <sup>a</sup>	,210	,286

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than ,001.

### Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	7,330	8	,502

### Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup>	Leeftijd	,004	,007	,333	1	,564	1,004
	Geslacht	,106	,241	,192	1	,661	1,111
	Opleidingsniveau	-,041	,141	,086	1	,770	,960
	Interesse in de verkiezingen	-,102	,188	,295	1	,587	,903
	Betrokkenheid bij de gemeenschap	-,096	,183	,274	1	,601	,909
	Sterkte identificatie met winnende partij	,581	,180	10,485	1	,001	1,789
	Sterkte identificatie met verliezende partij	-,498	,130	14,679	1	,000	,608
	Interactie Betrokkenheid en Identificatie winnend	,020	,155	,017	1	,895	1,021
	Constant	,618	,693	,794	1	,373	1,855

a. Variable(s) entered on step 1: Interactie Betrokkenheid en Identificatie winnend.

\*(Model 6 +) Model 8.

\*Model 8.

### Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	,064	1	,800
	Block	,064	1	,800
	Model	93,991	8	,000

### Model Summary

Step	-2 Log likelihood	Cox & Snell R	Nagelkerke R
		Square	Square
1	434,318 <sup>a</sup>	,210	,286

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than ,001.

### Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	5,379	8	,716

### Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup>	Leeftijd	,004	,007	,307	1	,580	1,004
	Geslacht	,093	,243	,147	1	,701	1,098
	Opleidingsniveau	-,040	,141	,082	1	,775	,961
	Interesse in de verkiezingen	-,103	,188	,301	1	,583	,902
	Betrokkenheid bij de gemeenschap	-,121	,216	,315	1	,575	,886
	Sterkte identificatie met winnende partij	,595	,152	15,233	1	,000	1,813
	Sterkte identificatie met verliezende partij	-,514	,148	12,164	1	,000	,598
	Interactie Betrokkenheid en Identificatie verliezend	,034	,133	,065	1	,800	1,034
	Constant	,654	,710	,849	1	,357	1,924

a. Variable(s) entered on step 1: Interactie Betrokkenheid en Identificatie verliezend.

**Toelichting:** Model 1, 2, 5, 6, 7 en 8 zijn logistische regressiemodellen. Modellen 3 en 4 zijn lineaire regressiemodellen en worden daarom apart geschat van de andere modellen.

Model 1 bevat enkel de controlevariabelen: *leeftijd*, *vrouw*, *opleidingsniveau* en *interesse\_verk*. Aan de helling en odds-ratio van *leeftijd* te zien ( $b = 0,004$ ;  $p = 0,496$ ;  $OR = 1,004$ ) blijkt dat het effect van leeftijd op het maken van een goede voorspelling vrijwel nihil is. Vrouwen blijken in vergelijking met mannen een lagere odds te hebben op het maken van een goede voorspelling ( $b = -0,074$ ;  $p = 0,722$ ;  $OR = 0,928$ ). Als het opleidingsniveau van een respondent met één niveau toeneemt, worden de odds op een goede voorspelling 0,981 groter, gecontroleerd voor de overige variabelen in het model ( $b = -0,019$ ;  $p = 0,873$ ;  $OR = 0,981$ ). Oftewel, een hoger opleidingsniveau gaat samen met een kleinere odds op het maken van een goede voorspelling. Daarnaast blijkt dat als de mate van interesse in de verkiezingen met één eenheid toeneemt, de odds op een goede voorspelling 0,954 groter worden, gecontroleerd voor de overige controlevariabelen. Oftewel, hoe groter de interesse in de verkiezingen, hoe kleiner de odds op het maken van een goede voorspelling. Echter, de effecten van de controlevariabelen zijn allen vrij klein en niet significant.

Aan model 2 wordt ook de variabele *betrokken\_gem* toegevoegd. Betrokkenheid bij de gemeenschap blijkt te leiden tot een kleinere odds op het goed voorspellen van de verkiezingswinnaar ( $b = -0,094$ ;

$p = 0,503$ ;  $OR = 0,910$ ). Als de mate van betrokkenheid bij de gemeenschap met één eenheid toeneemt (dus een respondent heeft één keer vaker 'yes' geantwoord op een van de items waaruit de schaal bestaat), worden de odds op het maken van een goede voorspelling 0,910 groter, gecontroleerd voor de controlevariabelen. De effecten van de controlevariabelen veranderen blijven vrijwel gelijk. Enkel het effect van *opleidingsniveau* is enigszins opvallend. Deze verandert van een odds-ratio van  $OR = 0,981$  in model 1, naar een odds-ratio van  $OR = 1,000$  in model 2. Een hoger opleidingsniveau heeft nu geen effect meer op de odds op het maken van een goede voorspelling.

Model 3 en 4 zijn lineaire modellen en dienen voor het analyseren van de paden die worden gevormd door het verwachte mediatie-effect. In model 3 worden enkel de controlevariabelen gebruikt om *discuss\_pol* te schatten. Aan model 4 wordt de variabele *betrokken\_gem* toegevoegd. Door de modellen te vergelijken kan er gekeken worden of de modelfit verbetert als *betrokken\_gem* wordt toegevoegd en dus of de mate van betrokkenheid bij de gemeenschap volgens het model een goede voorspeller is voor de mate van discussiëren over politiek.

Uit model 3 blijkt dat een grotere interesse in de verkiezingen samen gaat met het vaker discussiëren over politiek ( $b = 1,362$ ;  $p < 0,001$ ). Op een schaal van 1 tot 7 (van *discuss\_pol*) is een helling van  $b = 1,362$  redelijk groot. De effecten van de controlevariabelen op *discuss\_pol* blijven vrijwel gelijk wanneer *betrokken\_gem* aan het model wordt toegevoegd. De modelfit wordt nauwelijks beter:  $R^2_{\sigma} = 0,181$  in model 3 en  $R^2_{\sigma} = 0,183$  in model 4. *Discuss\_pol* stijgt met 0,215 wanneer de mate van betrokkenheid met één eenheid stijgt, gecontroleerd voor de controlevariabelen ( $b = 0,215$ ;  $p = 0,162$ ).

Model 5 is een logistisch regressiemodel en bevat de controlevariabelen, *betrokken\_gem* en *discuss\_pol*. Dit model kan vergeleken worden met model 2, het model met enkel de controlevariabelen en *betrokken\_gem*. Aan de verandering in deviance is te zien dat het model nauwelijks beter wordt wanneer *discuss\_pol* wordt toegevoegd aan het model: het maakt nog steeds vrijwel evenveel fouten. Discussiëren over politiek blijkt een negatief effect te hebben op de odds op een goede voorspelling ( $b = -0,067$ ;  $p = 0,150$ ;  $OR = 0,935$ ). De odds op het maken van een goede voorspelling worden namelijk 0,935 keer groter wanneer iemand één dag meer over politiek discussieert, gecontroleerd voor de overige variabelen in het model. Aan de kleinere helling van *betrokken\_gem* (in vergelijking met model 2) is te zien dat het negatieve effect van betrokkenheid bij de gemeenschap op de odds op het maken van een goede voorspelling lichtelijk verzwakt. Opvallend is dat het effect van opleidingsniveau in plaats van "niet bestaand" ( $OR = 1,000$  in model 2) in model 5 positief is ( $b = 0,024$ ;  $p = 0,851$ ;  $OR = 1,024$ ). Wanneer het opleidingsniveau met één niveau stijgt,

worden de odds op het maken van een goede voorspelling 1,024 keer groter, gecontroleerd voor de overige variabelen in het model.

Logistische regressiemodellen 6, 7 en 8 dienen om de verwachte moderatie-effecten te analyseren. Model 6 bevat nog geen interactievariabelen maar enkel de afzonderlijke moderatoren: identificatie met de winnende partij en identificatie met de verliezende partij. Hoe sterker de identificatie met de winnende partij, hoe groter de odds op het maken van een goede voorspelling ( $b = 0,594$ ;  $p < 0,001$ ;  $OR = 1,811$ ). Identificatie met de verliezende partij heeft een negatief effect op de odds op een goede voorspelling: wanneer de sterkte van identificatie met de verliezende partij één stapje stijgt, wordt de odds op het maken van een goede voorspelling 0,608 groter, gecontroleerd voor de overige variabelen ( $b = -0,497$ ;  $p < 0,001$ ;  $a = 0,608$ ). Het effect van opleidingsniveau op de odds op een goede voorspelling, wanneer *ident\_W* en *ident\_L* in het model worden opgenomen, is negatief ( $b = -0,041$ ;  $p = 0,773$ ;  $OR = 0,960$ ). Dit is tegengesteld aan het effect dat opleidingsniveau in model 5, het model met *discuss\_pol*, had. Daarnaast is het effect van *vrouw* op de odds op een goede voorspelling opvallend. In de logistische modellen 1, 2 en 5 was dit namelijk een negatief effect. In dit model, en in de modellen 7 en 8, is het effect van *vrouw* positief. Als iemand een vrouw is, in plaats van een man, worden de odds op het maken van een goede voorspelling 1,108 keer groter ( $b = 0,103$ ;  $p = 0,669$ ;  $OR = 1,108$ ).

In model 7 wordt de interactievariabele *BETRxDIDENT\_W* toegevoegd. In het geval van een zwakke identificatie met de winnende partij (*ident\_W* = 1) en één stapje stijging in de mate van betrokkenheid bij de gemeenschap stijgen de odds op het maken van een goede voorspelling lichtelijk ( $b = 0,020$ ;  $p = 0,895$ ;  $OR = 1,021$ ). Dit effect is erg klein en niet significant. Model 8 bevat de interactievariabele *BETRxDIDENT\_L*. Het effect van de interactievariabele is redelijk gelijk aan het effect van de eerste interactievariabele. In het geval van een zwakke identificatie met de verliezende partij (*ident\_L* = 1) en één stapje stijging in de mate van betrokkenheid bij de gemeenschap stijgen de odds op het maken van een goede voorspelling lichtelijk ( $b = 0,034$ ;  $p = 0,800$ ;  $OR = 1,034$ ). Dit effect is echter ook niet significant.

Modellen 5, 7 en 8 zijn de eindmodellen. Aangezien het logistische regressiemodellen zijn, zijn de hellingen niet eenvoudig te interpreteren. Het is daarom nuttig om voor een bepaalde verzameling aan waarden op de gegeven variabelen de kans te berekenen op het maken van een goede voorspelling ( $y=1$ ). Ik bereken hieronder de kans op het maken van een goede voorspelling van de verkiezingswinnaar op basis van model 7. Ik kies voor de variabelen in het model waarden die dichtbij de gemiddelden van de betreffende variabelen liggen. De wijze waarop de gemiddelden verkregen zijn wordt in bijlage 2.1 beschreven. De gemiddelden zijn tevens terug te vinden in tabel 1.

Het berekenen van een kans in een logistisch regressiemodel gebeurt aan de hand van onderstaande formule:

$$P^{\wedge}(Y = 1) = \frac{e^{a+bx}}{1 + e^{a+bx}}$$

De formule (a + bx) van model 7 ziet er als volgt uit:

$$Y = 0,618 + 0,004leeftijd + 0,106vrouw - 0,041opleidingsniveau - 0,102interesse\_verk - 0,096betrokken\_gem + 0,581ident\_W - 0,498ident\_L + 0,020BETRxIDENT\_W$$

Een "gemiddelde" respondent scoort de volgende waarden op de variabelen:

- Leeftijd = 47
- Vrouw = 2 (vrouw)
- Opleiding = 3 ("some college but no degree")
- Interesse\\_verk = 2 ("somewhat interested")
- Betrokken\\_gem = 1 (1x "yes, betrokken")
- Ident\\_W = 1 ("weak identification with winning party")
- Ident\\_L = 0 ("identification with winning party")
- BETRxIDENT\\_W = 1x1 = 1

Wanneer deze gegevens worden ingevuld in de formule ziet het er als volgt uit:

$$Y = 0,618 + 0,004*47 + 0,106*2 - 0,041*3 - 0,102*2 - 0,096*1 + 0,581*1 - 0,498*0 + 0,020*1$$

$$Y = 1,196$$

De kans op het maken van een goede voorspelling van de verkiezingswinnaar is vervolgens:

$$P^{\wedge}(Y = 1) = \frac{e^{1,196}}{1 + e^{1,196}} = 0,767812....$$

Dit betekent dat een gemiddelde respondent (een vrouw van 47 jaar oud die enige leerervaring heeft op de universiteit (maar geen diploma), enigszins geïnteresseerd is in de verkiezingen, een beetje betrokken is bij haar gemeenschap en zich in zwakke mate identificeert met de winnende partij) een kans heeft van  $p = 0,768$  heeft om een goede voorspelling te maken van de winnaar van de verkiezingen. Dit is een redelijk grote kans.

## Bijlage 3

### 3.1 Assumptietoetsing

*Toelichting:* In tegenstelling tot een lineaire regressie, moet er bij een logistische regressie niet op veel assumpties worden gecontroleerd. Aangezien dit onderzoek 3 logistische eindmodellen betreft, worden er dus niet veel assumpties getoetst. De enkele assumptie waarmee wel rekening mee gehouden moet worden is de assumptie van onafhankelijke waarnemingen. Dit betekent dat er geen samenhang mag bestaan tussen de cases die in dit onderzoek gebruikt worden. De American National Election Studies verzamelen data door respondenten op basis van willekeurige steekproeven te benaderen (American National Election Study, 2021). Hiermee wordt ervoor gecontroleerd dat de verzamelde cases onafhankelijk van elkaar zijn. In mijn onderzoek gebruik ik data uit drie verkiezingsjaren: 2004, 2008 en 2012. Voorafgaand aan het uitvoeren van de analyses heb ik uit elk jaar nog een willekeurige steekproef van 300 cases getrokken. Daarna zijn er missende waarden uit de dataset verwijderd. Uiteindelijk werden de analyses uitgevoerd op een steekproef van 399 respondenten. Vanwege de onafhankelijk verzamelde data, en de willekeurige selecties die daarna zijn uitgevoerd op de data, kan er worden geconcludeerd dat er wordt voldaan aan de assumptie van onafhankelijke waarnemingen.

### 3.2 Outliers

*Syntax:*

\*Berekenen van de leverage- en DFBETA-waarden.

\*Eindmodel 5.

```
LOGISTIC REGRESSION VARIABLES goed_voorspeld
  /METHOD=ENTER leeftijd vrouw opleidingsniveau interesse_verk
  /METHOD=ENTER betrokken_gem
  /METHOD=ENTER discuss_pol
  /SAVE=LEVER DFBETA
  /PRINT=GOODFIT
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

\*Eindmodel 7.

```
LOGISTIC REGRESSION VARIABLES goed_voorspeld
```



```

/METHOD=ENTER leeftijd vrouw opleidingsniveau interesse_verk betrokken_gem ident_W ident_L
/METHOD=ENTER BETRxIDENT_W
/SAVE=LEVER DFBETA
/PRINT=GOODFIT
/CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

```

\*Eindmodel 8.

LOGISTIC REGRESSION VARIABLES goed\_voorspeld

```

/METHOD=ENTER leeftijd vrouw opleidingsniveau interesse_verk betrokken_gem ident_W ident_L
/METHOD=ENTER BETRxIDENT_L
/SAVE=LEVER DFBETA
/PRINT=GOODFIT
/CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

```

\*Tabel met beschrijvende statistieken voor analyseren alle DFBETA-waarden.

```

DESCRIPTIVES VARIABLES=DFB1_1 DFB2_1 DFB3_1 DFB4_1 DFB5_1 DFB6_1 DFB1_2 DFB2_2 DFB3_2
  DFB4_2 DFB5_2 DFB6_2 DFB7_2 DFB8_2 DFB1_3 DFB2_3 DFB3_3 DFB4_3 DFB5_3 DFB6_3 DFB7_3
  DFB8_3
/STATISTICS=MEAN STDDEV MIN MAX.

```

\*Tabel met beschrijvende statistieken voor analyseren leverage-waarden.

```

DESCRIPTIVES VARIABLES=LEV_1 LEV_2 LEV_3
/STATISTICS=MEAN STDDEV MIN MAX.

```

*Output:*

\*Tabel met beschrijvende statistieken voor analyseren alle DFBETA-waarden.

<b>Descriptive Statistics</b>					
	N	Minimum	Maximum	Mean	Std. Deviation
DFB1_1 DFBETA for Leeftijd	399	-,00119	,00083	,0000000	,00031964
DFB2_1 DFBETA for Geslacht	399	-,01562	,02041	-,0000060	,01069762

DFB3_1 DFBETA for Opleidingsniveau	399	-,01596	,01544	-,0000028	,00641732
DFB4_1 DFBETA for Interesse in de verkiezingen	399	-,02262	,02943	-,0000024	,00856599
DFB5_1 DFBETA for Betrokkenheid bij de gemeenschap	399	-,02166	,01636	,0000030	,00716952
DFB6_1 DFBETA for Aantal dagen discussiëren over politiek in de afgelopen week	399	-,00666	,00636	,0000002	,00232542
DFB1_2 DFBETA for Leeftijd	399	-,00201	,00137	,0000001	,00037153
DFB2_2 DFBETA for Geslacht	399	-,03714	,03462	-,0000098	,01257462
DFB3_2 DFBETA for Opleidingsniveau	399	-,02376	,02982	-,0000007	,00730305
DFB4_2 DFBETA for Interesse in de verkiezingen	399	-,03087	,04782	-,0000104	,00980017
DFB5_2 DFBETA for Betrokkenheid bij de gemeenschap	399	-,03435	,04014	,0000063	,00941377
DFB6_2 DFBETA for Sterkte identificatie met winnende partij	399	-,06096	,02648	,0000015	,00891985
DFB7_2 DFBETA for Sterkte identificatie met verliezende partij	399	-,01282	,02258	,0000019	,00660582
DFB8_2 DFBETA for Interactie Betrokkenheid en Identificatie winnend	399	-,08045	,03177	,0000026	,00710339
DFB1_3 DFBETA for Leeftijd	399	-,00203	,00133	,0000002	,00037299
DFB2_3 DFBETA for Geslacht	399	-,03411	,03517	-,0000111	,01271202
DFB3_3 DFBETA for Opleidingsniveau	399	-,02273	,02662	-,0000030	,00726921
DFB4_3 DFBETA for Interesse in de verkiezingen	399	-,03156	,04778	-,0000119	,00982580
DFB5_3 DFBETA for Betrokkenheid bij de gemeenschap	399	-,06528	,03066	,0000067	,01074126
DFB6_3 DFBETA for Sterkte identificatie met winnende partij	399	-,04210	,02032	,0000004	,00739415

DFB7_3 DFBETA for Sterkte identificatie met verliezende partij	399	-,01773	,02840	,0000009	,00730713
DFB8_3 DFBETA for Interactie Betrokkenheid en Identificatie verliezend	399	-,01586	,03527	-,0000002	,00676030
Valid N (listwise)	399				

\*Tabel met beschrijvende statistieken voor analyseren leverage-waarden.

### Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
LEV_1 Leverage value Model 5	399	,00624	,04693	,0175439	,00654160
LEV_2 Leverage value Model 7	399	,00906	,05415	,0225564	,00895114
LEV_3 Leverage value Model 8	399	,00870	,06317	,0225564	,01040261
Valid N (listwise)	399				

*Toelichting:* Voor het controleren op outliers in de data wordt er gekeken naar de leverage-waarden en DFBETA-waarden. De DFBETA geeft de verandering in (gestandaardiseerde) regressiecoëfficiënten als een bepaalde case uit de dataset wordt weggelaten. Hoe groter de absolute waarde van DFBETA, hoe groter de invloed van de betreffende case op die parameterschatting. De vuistregel die in het geval van het analyseren van DFBETA-waarden wordt gehanteerd is dat een DFBETA die groter is dan  $\frac{2}{\sqrt{n}}$  of  $\frac{3}{\sqrt{n}}$  als invloedrijk kan worden beschouwd. De steekproefgrootte van dit onderzoek is  $n=399$ . Dit betekent dat cases met DFBETA-waarden die groter zijn dan 0,100 of 0,150 als invloedrijk kunnen worden beschouwd.

Zoals eerder genoemd worden ook de leverage-waarden bekeken. Een leverage-waarde geeft aan hoe sterk een bepaalde case afwijkt op de gemiddeldes van de onafhankelijke variabelen. Hoe groter de leverage-waarde, hoe meer een case afwijkt van deze gemiddeldes en hoe groter de invloed op de helling en daarmee de voorspelde waarde van de afhankelijke variabele. De vuistregel die in het geval van het analyseren van leverage-waarden wordt gehanteerd is dat er mogelijk sprake is van een probleem als een leverage-waarde hoger is dan  $\frac{3p}{n}$ . Het aantal parameters (inclusief constante) in eindmodel 5 is  $p = 7$ . Zoals in de output tabel met beschrijvende statistieken is te zien is  $n = 399$ . Dit betekent vervolgens dat de leverage-waarden niet groter mogen zijn dan 0,053. Het aantal

parameters in zowel eindmodel 7 als eindmodel 8 is  $p = 9$ . De steekproefgrootte is  $n = 399$ . Dit betekent dat de leverage-waarden in deze twee modellen onder de 0,068 moeten liggen.

Voor het berekenen van de DFBETA en leverage-waarden worden de regressieanalyses gedraaid waarbij de opties tot het bewaren van de DFBETA en leverage-waarden wordt aangeklikt. Van de DFBETA-waarden van alle parameters uit elk van de drie eindmodellen wordt een tabel met beschrijvende statistieken gemaakt. Er wordt gekeken in de kolom van maximale waarden of er daadwerkelijk parameters zijn met cases die boven de gehanteerde criteriumwaarde liggen. Zoals eerder benoemd mogen de DFBETA-waarden niet groter zijn dan 0,100 of 0,150. Uit de tabel kan worden geconcludeerd dat alle DFBETA-waarden in de dataset onder 0,048 liggen. Dit betekent dat de regressiecoëfficiënten uit de analyses niet veel zullen veranderen als er een respondent uit de analyse wordt gehaald. Op basis van de DFBETA worden er dus geen mogelijke uitbijters aangewezen.

Vervolgens wordt er ook een tabel met beschrijvende statistieken gemaakt van de 3 variabelen (voor de 3 eindmodellen) die de leverage-waarden aangeven voor elke case in de dataset. Voor eindmodellen 5, 7 en 8 zijn dit respectievelijk de variabelen LEV\_1, LEV\_2 en LEV\_3. Eerder is benoemd dat deze waarden binnen eindmodel 5 onder 0,053 moeten liggen, en in de eindmodellen 7 en 8 onder de 0,068. Uit de tabel met beschrijvende statistieken blijkt dat in model 5 het maximum qua leverage-waarden (afgerond) 0,047 is. In model 7 is dat 0,054 en in model 8 is dat maximum 0,063. Hieruit kan geconcludeerd worden dat alle cases in de modellen een leverage-waarde hebben die onder de eerdergenoemde criteriumwaarden valt. Aangezien er op basis van de DFBETA-waarden en de leverage-waarden geen mogelijke outliers worden gevonden, worden er geen cases uit de dataset verwijderd. Het databestand zal blijven bestaan uit 399 respondenten.

### 3.3 Multicollineariteit

*Syntax:*

\*Analyseren van de VIF-scores van eindmodellen 5, 7 en 8.

\*Eindmodel 5.

REGRESSION

/MISSING LISTWISE

/STATISTICS COEFF OUTS CI(95) R ANOVA COLLIN TOL CHANGE

/CRITERIA=PIN(.05) POUT(.10)

/NOORIGIN

/DEPENDENT goed\_voorspeld

/METHOD=ENTER leeftijd vrouw opleidingsniveau interesse\_verk betrokken\_gem discuss\_pol.

\*Eindmodel 7.

REGRESSION

/MISSING LISTWISE

/STATISTICS COEFF OUTS CI(95) R ANOVA COLLIN TOL CHANGE

/CRITERIA=PIN(.05) POUT(.10)

/NOORIGIN

/DEPENDENT goed\_voorspeld

/METHOD=ENTER leeftijd vrouw opleidingsniveau interesse\_verk betrokken\_gem ident\_W ident\_L  
BETRxIDENT\_W.

\*Eindmodel 8.

REGRESSION

/MISSING LISTWISE

/STATISTICS COEFF OUTS CI(95) R ANOVA COLLIN TOL CHANGE

/CRITERIA=PIN(.05) POUT(.10)

/NOORIGIN

/DEPENDENT goed\_voorspeld

/METHOD=ENTER leeftijd vrouw opleidingsniveau interesse\_verk betrokken\_gem ident\_W ident\_L  
BETRxIDENT\_L.

Output:

\*Eindmodel 5.

Model		Coefficients <sup>a</sup>					95,0% Confidence Interval for B		Collinearity Statistics	
		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Lower Bound	Upper Bound	Tolerance	VIF
		B	Std. Error	Beta						
1	(Constant)	,591	,150		3,948	,000	,297	,886		
	leeftijd Leeftijd	,001	,001	,043	,826	,410	-,002	,004	,944	1,060
	vrouw Geslacht	-,013	,049	-,014	-,271	,787	-,110	,083	,992	1,008

opleidingsniveau Opleidingsniveau	,006	,029	,010	,190	,849	-,052	,063	,893	1,120
interesse_verk Interesse in de verkiezingen	,012	,040	,017	,308	,759	-,067	,091	,807	1,239
betrokken_gem Betrokkenheid bij de gemeenschap	-,019	,034	-,029	-,563	,574	-,085	,047	,925	1,082
discuss_pol Aantal dagen discussiëren over politiek in de afgelopen week	-,016	,011	-,080	- 1,435	,152	-,037	,006	,807	1,240

a. Dependent Variable: goed\_voorspeld Is de winnaar van de verkiezingen goed voorspeld?

\*Eindmodel 7.

		Coefficients <sup>a</sup>								
		Unstandardized Coefficients		Standardized Coefficients		95,0% Confidence Interval for B		Collinearity Statistics		
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	,662	,131		5,047	,000	,404	,919		
	leeftijd Leeftijd	,001	,001	,025	,536	,592	-,002	,003	,916	1,091
	vrouw Geslacht	,018	,045	,019	,413	,680	-,069	,106	,953	1,049
	opleidingsniveau Opleidingsniveau	-,008	,026	-,014	-,291	,771	-,059	,044	,904	1,107
	interesse_verk Interesse in de verkiezingen	-,017	,034	-,024	-,501	,617	-,085	,051	,868	1,152
	betrokken_gem Betrokkenheid bij de gemeenschap	-,022	,037	-,035	-,602	,547	-,096	,051	,595	1,682
	ident_W Sterkte identificatie met winnende partij	,087	,028	,217	3,119	,002	,032	,142	,415	2,408
	ident_L Sterkte identificatie met verliezende partij	-,124	,026	-,290	- 4,838	,000	-,174	-,074	,557	1,794

BETRxDIDENT_W Interactie Betrokkenheid en Identificatie winnend	,007	,023	,020	,307	,759	-,038	,052	,461	2,171
---	------	------	------	------	------	-------	------	------	-------

a. Dependent Variable: goed\_voorspeld Is de winnaar van de verkiezingen goed voorspeld?

\*Eindmodel 8.

		Coefficients <sup>a</sup>								
		Unstandardized Coefficients		Standardized Coefficients		95,0% Confidence Interval for B		Collinearity Statistics		
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	,666	,134		4,985	,000	,403	,928		
	leeftijd Leeftijd	,001	,001	,023	,501	,617	-,002	,003	,915	1,093
	vrouw Geslacht	,016	,045	,016	,345	,730	-,073	,104	,940	1,063
	opleidingsniveau Opleidingsniveau	-,007	,026	-,013	-,272	,786	-,058	,044	,905	1,105
	interesse_verk Interesse in de verkiezingen	-,018	,034	-,025	-,513	,609	-,085	,050	,868	1,152
	betrokken_gem Betrokkenheid bij de gemeenschap	-,020	,038	-,031	-,538	,591	-,094	,054	,586	1,707
	ident_W Sterkte identificatie met winnende partij	,091	,024	,228	3,761	,000	,044	,139	,547	1,827
	ident_L Sterkte identificatie met verliezende partij	-,126	,029	-,296	- 4,405	,000	-,183	-,070	,445	2,246
	BETRxDIDENT_L Interactie Betrokkenheid en Identificatie verliezend	,005	,025	,014	,202	,840	-,044	,054	,447	2,238

a. Dependent Variable: goed\_voorspeld Is de winnaar van de verkiezingen goed voorspeld?

*Toelichting:* Het is, naast het controleren van de assumpties en de controle op outliers, belangrijk om eventuele multicollineariteit in het onderzoeksmodel op te sporen. Dit betreft het opsporen van problematisch sterke samenhang tussen de variabelen. Een mogelijk gevolg van een dergelijke sterke

samenhang kan immers zijn dat het belang van afzonderlijke variabelen in het onderzoeksmodel moeilijk te bepalen is.

Om de multicollineariteit te onderzoeken worden de VIF-scores van de eindmodellen bekeken. Dit wordt gedaan door lineaire regressies uit te voeren in SPSS voor deze eindmodellen en hierbij te kijken naar de VIF-waarden, die telkens worden gegeven in de uiterst rechtse kolom van de tabellen uit de output. Er wordt hierbij een grens gehanteerd van  $VIF < 4$ . Alle VIF-scores in van de variabelen in de eindmodellen dienen dus onder een waarde van 4 te vallen.

Het betreft in dit geval dus het analyseren van de VIF-scores van de variabelen in de modellen 5, 7 en 8. Ten eerste blijkt dat de VIF-scores van de variabelen in eindmodel 5 allemaal onder 1,240 liggen. De VIF-scores van de variabelen in eindmodel 7 liggen daarnaast allemaal onder 2,408. Tot slot liggen de VIF-scores van eindmodel 8 allemaal onder 2,246. De VIF-scores in de eindmodellen liggen allemaal ruim onder de gehanteerde grens van  $VIF < 4$ . Hieruit blijkt dat er in deze regressieanalyses geen sprake is van problematische multicollineariteit.