university of groningen

faculty of behavioural and social sciences

Wie voorspelt de verkiezingen beter?

De invloed van gender, politieke kennis en opleidingsniveau op het voorspellen van verkiezingen

Bachelorwerkstuk

Rosina Ubbens S3207455
r.m.r.ubbens@student.rug.nl
Studiejaar: 2021-2022
Begeleider: Vincenz Frey
Datum: 7-6-2022

Inhoud

Inhoud

Wie voorspelt de verkiezingen beter?....1
Abstract....3
1 Inleiding....4
2 Theorie....7
Het juist voorspellen van de verkiezingen 7
Gender en politieke kennis 7
Opleidingsniveau 9
Controlevariabelen....10
3 Methoden 11
Beschrijving steekproef 11
Beschrijvende statistieken 15
Univariate statistieken 15
Bivariate statistieken 17
Statistisch analyseplan 19
4 Resultaten....21
Gender en de mediatie van politieke kennis....21
Opleidingsniveau 25
Model evaluatie 26
5 Conclusie en discussie....28
6 Literatuurlijst....30
7 Bijlagen 33

Abstract

Dit is een onderzoek draagt bij aan het relatief nieuwe veld van *citizen forecasting*. Dit is een methode waarmee de verkiezingsuitslag voorspeld wordt door burgers te vragen wie zij denken dat de verkiezingen gaat winnen. In dit onderzoek wordt onderzocht of er sociaal demografische factoren zijn die deze voorspelling van de verkiezingsuitslag beïnvloeden. Hiervoor is een binaire logistische regressie uitgevoerd met behulp van American National Election Survey data uit 1996, 2004 en 2008. Ten eerste is onderzocht of vrouwen de verkiezingen slechter voorspellen dan mannen. Daarna is gekeken of een verschil tussen mannen en vrouwen deels verklaard kan worden doordat vrouwen gemiddeld minder politieke kennis hebben dan mannen. De resultaten laten zien dat er geen significant verschil is tussen mannen en vrouwen in het juist voorspellen van de verkiezingsuitslag. Daardoor was er ook geen invloed van politieke kennis op dit verschil gevonden. Ten tweede is onderzocht of mensen met een hoger opleidingsniveau de verkiezingsuitslag beter voorspellen dan mensen met een lager opleidingsniveau. Uit de analyse blijkt dat mensen met een hoger opleidingsniveau de verkiezingsuitslag significant beter voorspellen dan het laagste opleidingsniveau, maar tussen de andere mensen met andere opleidingsniveaus waren geen significante verschillen. Uit de resultaten blijkt dat er mogelijk geen effecten bestaan van demografische kenmerken op het juist voorspellen van de verkiezingen.

1 Inleiding

Tijdens de verkiezingen in de Verenigde Staten in 2020 bleek hoe gevaarlijk het is als mensen de uitslag van een verkiezing niet geloven. Veel aanhangers van de toenmalige president Donald Trump waren overtuigd van grootschalige verkiezingsfraude, daarbij waren zij niet bereid om de uitkomst van de hertelling of het fraudeonderzoek te geloven (NOS, 2021a). Dit leidde tot een bestorming van het Capitool, het gebouw waarin het Amerikaanse congres gevestigd is (NOS, 2021b).

Deze gebeurtenissen maken het relevant om te kijken naar wat factoren zijn die het voorspellen van de verkiezingsuitslag beïnvloeden. De fanatieke aanhangers van president Trump maakten deel uit van een specifieke demografische groep. Zij zijn vooral witte conservatieve mannen met een lage opleiding (RTL Nieuws, 2021). Mogelijk hebben deze demografische kenmerken een rol gespeeld in waarom de Trump aanhangers de verkiezingsuitslag niet geloofden.

Een manier om naar het voorspellen van verkiezingsuitslagen te kijken is om te bedenken dat het te maken heeft met een juiste inschatting maken van het gedrag van anderen. Onderzoek wijst uit dat mensen vaak proberen in te schatten wat anderen doen en daar hun gedrag op aanpassen (Graupensperger et al., 2020). Daarvoor moeten zij goed in kunnen schatten wat anderen gaan doen. De verkiezingsuitslag juist voorspellen heeft hier mee te maken omdat zij dan juist ingeschat hebben wat een meerderheid van de kiezers zal gaan doen. Uit onderzoek blijkt dat een merendeel van de mensen goed voorspelt wie de verkiezingen gaat winnen (Mongrain, 2021; Delavande & Manski, 2012; Leiter et al., 2018).

Meer informatie over factoren die ervoor zorgen dat mensen in het geval van verkiezingsuitslagen een goede inschatting maken zou ook kunnen zorgen voor inzicht in andere gevallen waarin mensen een inschatting doen van het gedrag van anderen.

Daarom zal dit onderzoek kijken of sociaaldemografische factoren de juistheid van de voorspelling van een verkiezingsuitslag beïnvloeden. Dat burgers gevraagd wordt om de verkiezingsuitslag te voorspellen staat in de wetenschappelijke literatuur bekend als *citizen forecasting* of burger prognose. Dit is een relatief nieuwe manier om voorspellingen te doen vergeleken met bijvoorbeeld opiniepeilingen of statistische modellen. In een wetenschappelijke survey wordt respondenten gevraagd wie zij denken dat de verkiezingen gaat winnen, vaak gaat het om verkiezingen die binnen een aantal maanden of weken plaatsvinden. Dit kan gaan om plaatselijke of om landelijke verkiezingen, in beide gevallen

voorspelt de meerderheid van de mensen de uitslag juist (Badad et al.,1993; Delavande & Manski, 2012; Mongrain, 2021).

In dit onderzoek zal gekeken worden of de sociaal demografische kenmerken gender en opleidingsniveau invloed hebben op of mensen de verkiezingsuitslag juist voorspellen. Er is binnen dit onderzoeksveld weinig onderzoek gedaan naar genders buiten mannen en vrouwen, daarom zal dit onderzoek uitgaan van mannen en vrouwen. Er is nog niet veel onderzoek gedaan naar de invloed van gender op het juist voorspellen van de verkiezingsuitslag. In de onderzoeken waarin gender gebruikt wordt is dit vaak een controlevariabele. Hier zijn nog geen consistente trends vastgesteld (Mongrain, 2021). Daarom is het interessant om te kijken of er in dit onderzoek wel een effect te vinden gaat zijn. Vooral omdat er gekeken zal worden naar een combinatie van factoren. De verwachting van dit onderzoek is dat vrouwen minder vaak de verkiezingsuitslag juist voorspellen dan mannen. Het is mogelijk dat het verschil in politieke kennis tussen mannen en vrouwen invloed heeft op hoe goed zij de verkiezingen kunnen voorspellen. Vrouwen hebben namelijk over het algemeen minder politieke kennis dan mannen (Dolan & Hansen, 2020). In dit onderzoek zal gender dus gebruikt worden om te kijken of mannen of vrouwen de verkiezingsuitslag beter voorspellen. Bovendien zal onderzocht worden of een verschil in de juistheid van de voorspelde verkiezingsuitslag verklaard kan worden door een verschil in politieke kennis tussen mannen en vrouwen.

Ook zal onderzocht worden of opleidingsniveau invloed heeft op de het juist voorspellen van de verkiezingsuitslag. Naar dit aspect van *citizen forecasting* is meer onderzoek gedaan en zijn meerdere theorieën ontwikkeld. Uit eerder onderzoek blijkt dat mensen met een hoger opleidingsniveau vaker een juiste verkiezingsuitslag voorspellen dan mensen met een lager opleidingsniveau (Lewis-Beck & Tien,1999). Vaak wordt verondersteld dat dit komt doordat mensen met een hoger opleidingsniveau meer politieke kennis hebben (Lewis-Beck & Tien, 1999). Aangezien er in dit onderzoek rekening gehouden wordt met de mogelijke invloed van politieke kennis, kan getest worden of alternatieve verklaringen ook een deel van het verschil kunnen verklaren. Een voorbeeld van een theorie draait om de sociale netwerken van hoger- en lageropgeleiden. Die stelt dat hogeropgeleiden een groter en diverser sociaal netwerk hebben waardoor zij het stemgedrag van anderen beter in kunnen schatten (Leiter et al., 2018). Als blijkt dat het verschil verklaard kan worden door politieke kennis dan kan de oplossing voor het verschil liggen bij voorlichting voor lager opgeleiden. Als echter blijkt dat er meer mogelijke verklaringen zijn dan kan er meer onderzoek in die richting gedaan worden.

In dit onderzoek wordt gebruikt gemaakt van de American National Election Studies uit de jaren 1996, 2004 en 2008 (ANES, 2022). Dit zijn surveys die vanaf 1948 tijdens alle Amerikaanse presidentsverkiezingen afgenomen worden. Voor het onderzoek wordt dus een verkiezing gebruikt die plaatsvindt in een tweepartijenstelsel. De presidentsverkiezingen in de Verenigde Staten een districtenstelsel en kiesmannen (Coleman et al., 2001). Dit betekent dat voor elke staat een aantal kiesmannen beschikbaar is en die kiesmannen stemmen uiteindelijk op de president (Coleman et al., 2001).

Voor het onderzoek zal een hiërarchische binaire logistische regressie uitgevoerd worden met als afhankelijke variabele de juist voorspelde verkiezingsuitslag. Daarin zal een mediatie analyse uitgevoerd worden om te analyseren of politieke kennis een deel van het verschil in juist voorspelde verkiezingsuitslag kan verklaren. Daarnaast zal de invloed van opleidingsniveau op de juist voorspelde verkiezingsuitslag bekeken worden.

In dit onderzoek zal de hoofdvraag zijn wat de invloed van de sociale demografische kenmerken gender en opleidingsniveau op het juist voorspellen van de verkiezingsuitslag zijn. Voor het beantwoorden van deze vraag zijn er drie deelvragen opgesteld. De eerste twee over gender zijn: voorspellen vrouwen de verkiezingen slechter dan mannen? Kan een deel van dit verschil verklaard worden doordat vrouwen minder politieke kennis hebben dan mannen? De laatste over opleidingsniveau is: voorspellen mensen met een hoger opleidingsniveau de verkiezingen beter dan mensen met een lager opleidingsniveau?

2 Theorie

Het juist voorspellen van de verkiezingen

Op basis van eerder onderzoek valt te verwachten dat de kans dat iemand de presidentsverkiezingen juist voorspelt groter is dan vijftig procent en dat de meerderheid van de mensen de verkiezingsuitslag juist zal voorspellen (Mongrain, 2021). Een probleem met het gebruiken van data over verkiezingen in de Verenigde Staten is dat er gestemd wordt via een districtenstelsel en niet via welke kandidaat de meeste stemmen krijgt, de *popular vote* (Coleman et al., 2001). Een persoon kan dus goed inschatten op wie het meest gestemd wordt, maar toch de president verkeerd voorspellen. In de gekozen verkiezingsjaren, 1996, 2004 en 2008, heeft echter de presidentskandidaat met de meeste stemmen ook de verkiezingen gewonnen.

Zoals in de inleiding uitgelegd is, zijn sociaal demografische kenmerken mogelijk van invloed op of iemand de verkiezingsuitslag goed of fout voorspelt. Eerst zal uiteengezet worden wat de invloed van gender op de juist voorspelde verkiezingsuitslag is en hoe dat mogelijk verklaard kan worden door een verschil in politieke kennis. Daarna zal uitgelegd worden wat de verwachte invloed van opleidingsniveau is op het juist voorspellen van de verkiezingsuitslag.

Gender en politieke kennis

De eerste verwachting die besproken zal worden, is dat vrouwen de verkiezingsuitslag minder vaak goed zullen voorspellen dan mannen. Uit eerder onderzoek wordt niet duidelijk of dit verschil er wel of niet is (Mongrain, 2021). In *citizen forecasting* wordt gender vaak als controlevariabele gebruikt, zonder een specifieke motivatie. Onderzoek waarin gender als controlevariabele gebruikt is, laat kleine effecten van gender zien maar niet altijd in dezelfde richting (Delavande & Manski, 2012). Soms was er geen significant verschil tussen mannen en vrouwen, soms voorspelden mannen beter en soms vrouwen. Het is dus interessant om te onderzoeken of de theorie die in dit deze paragraaf geïntroduceerd wordt wel een duidelijk effect van gender op de verkiezingsuitslag laat zien.

Een reden om een verschil tussen mannen en vrouwen te verwachten in het juist voorspellen van de verkiezingsuitslagen, is dat politieke kennis tussen mannen en vrouwen vaak verschilt (Dolan & Hansen, 2020). Meer kennis over de politiek zorgt voor een grotere

kans om de verkiezingsuitslag juist te voorspellen (Mongrain, 2021). Politieke kennis is dus belangrijk voor het juist voorspellen van de juiste verkiezingsuitslag. Dit betekent dat een verschil in kennis over politiek zal leiden tot een verschil in het juist voorspellen van de verkiezingsuitslag.

Vrouwen hebben over het algemeen minder kennis over politiek dan mannen (Dolan & Hansen, 2020). Dit gaat voornamelijk over algemene politieke kennis, zoals kennis over welke partij de meerderheid in het parlement heeft of welke partijen onderdeel uitmaken van de coalitie. Als het gaat over politieke kennis die specifiek over vrouwen gaat, is de kennis van mannen en vrouwen gelijk (Dolan & Hansen, 2020). Het gemiddelde van mannen en vrouwen is voor deze onderwerpen even laag.

Een verklaring voor het gevonden verschil is de manier waarop vragen over politiek door vrouwen beantwoord worden. Vrouwen kiezen vaker voor een antwoord als 'weet ik niet zeker' of 'geen antwoord', dan mannen. Dit komt doordat vrouwen vaak minder risico nemen dan mannen(Pereira, 2019). In dit geval vinden vrouwen het erg als ze de vraag fout beantwoorden, om dit te voorkomen gaan ze voor de veiligere keuze om helemaal geen antwoord te geven (Pereira, 2019). Dit zou ook kunnen komen doordat vrouwen een bescheidenere houding aannemen in survey settings (Pereira, 2019). Onderzoek heeft uitgewezen dat deze verklaring inderdaad een deel van het verschil in politieke kennis tussen mannen en vrouwen kan verklaren (Dolan & Hansen, 2020). Onderzoek waarbij rekening gehouden werd met risicovermijdend gedrag en de bescheidenheid van vrouwen vonden veel minder grote verschillen in politieke kennis tussen mannen en vrouwen (Dolan & Hansen, 2020). Het is dus waarschijnlijk dat in dit onderzoek gemiddeld een lager kennisniveau over politieke gevonden zal worden dan bij mannen. Een andere verklaring voor het verschil in politieke kennis is dat vrouwen sterk beïnvloed worden door vrouwelijke representatie op het hoogste niveau van de politiek (Pereira, 2019). Wanneer er veel vrouwelijke politici betrokken zijn in de hoogste niveaus van de politiek dan is de kennis van vrouwen over politiek groter. Dit zou komen doordat er een stigma ligt op vrouwen en politiek (Pereira, 2019). Door deze negatieve associatie tussen vrouwen en politiek slaan vrouwen kennis over politiek slecht op (Pereira, 2019). Dit stigma wordt verkleind door vrouwelijke voorbeelden in de politiek. Hierdoor worden vrouwen minder herinnerd aan negatieve stereotypen. Aangezien de eerste vrouwelijke presidentskandidaat in de Verenigde staten pas in 2016 deelnam aan de verkiezingen, is de kans groot dat dit stigma bestaat onder vrouwen in de Verenigde Staten. Daaruit volgt ook dat de kennis van vrouwen over politiek minder zal zijn bij vrouwen dan bij mannen.

Beide verklaringen geven dus een goede reden om te verwachten dat vrouwen waarschijnlijk slechtere politieke kennis hebben dan mannen en dat ze de verkiezingsuitslag slechter kunnen voorspellen.

Uit het voorgaande volgen de volgende hypothesen:

H1: *Vrouwen voorspellen vaker de verkiezingsuitslag onjuist dan mannen.*

H2: *Vrouwen voorspellen de verkiezingsuitslag minder vaak goed dan mannen, dit kan grotendeels verklaard worden doordat mannen gemiddeld meer kennis over politiek hebben dan vrouwen.*

Opleidingsniveau

De tweede verwachting die besproken zal worden heeft te maken met de sociaal demografische factor opleidingsniveau. De verwachting is dat mensen met een hoger opleidingsniveau vaker een juiste verkiezingsuitslag voorspellen dan mensen met een lager opleidingsniveau (Lewis-Beck & Tien, 1999). Het is logisch om aan te nemen dat een deel van dat effect verklaard kan worden doordat hoger opgeleiden meer politieke kennis hebben dan laagopgeleiden (Lewis-Beck & Tien, 1999). Zoals hierboven uitgelegd is, zorgt meer politieke kennis voor een betere voorspelling van de verkiezingsuitslag.

Het is echter mogelijk dat dit niet de gehele verklaring voor dit verschil is. Het is namelijk zo dat hoogopgeleiden vaak een groter en diverser sociaal netwerk hebben dan laagopgeleiden (Leiter et al., 2018). Zij komen in contact met meer mensen en met een grotere diversiteit aan mensen. Daardoor zijn zij beter in staat om in te schatten wat anderen gaan doen en daardoor kunnen zij een betere voorspelling maken (Leiter et al., 2018). Dit komt doordat diverse sociale netwerken ervoor zorgen dat mensen vaker iemand tegenkomen die niet dezelfde informatie of mening heeft als zij (Leiter et al., 2018). Vanwege deze ontmoetingen zullen hoger opgeleiden in gesprek gaan over de informatie die zij hebben en zal hun kennis verbeteren. Ze zullen ook een betere inschatting kunnen maken van de meningen en kennis die de mensen om hen heen hebben (Leiter et al., 2018). Want als zij het debat aangaan, kunnen zij ook worden gecorrigeerd als zij verkeerde informatie hebben, zowel over politieke als over wat anderen beslissen (Leiter et al., 2018). Dit heeft meer invloed dan een grote hoeveelheid aan verschillende passieve bronnen zoals de televisie en de krant omdat die eenzijdig zijn en discussie daarmee dus niet mogelijk is(Leiter et al., 2018).

Uit deze redenatie volgt dat de informatie die hoogopgeleiden hebben beter is en zij beter op de hoogte zijn van wat de mensen in hun netwerk gaan doen. In het geval van politiek betekent dit dat zij beter op de hoogte zijn van wat anderen gaan stemmen dit zal hen helpen om een betere voorspelling van de verkiezingsuitslag te doen.

Hieruit volgt de volgende hypothese:

H3 *Mensen met een hoger opleidingsniveau zullen de verkiezingsuitslag beter voorspellen dan mensen met een lager opleidingsniveau.*

Controlevariabelen

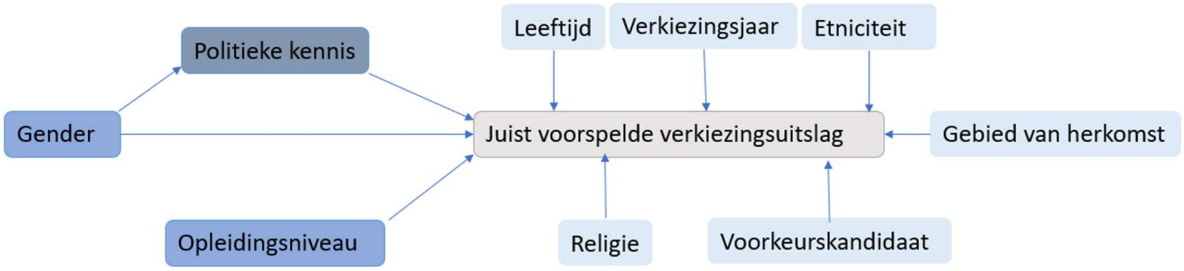
Om het effect van andere factoren uit te sluiten worden ook de volgende factoren meegenomen in dit onderzoek: voorkeurskandidaat, leeftijd, verkiezingsjaar, etniciteit, gebied van herkomst, en religie.

De reden dat iemands voorkeurskandidaat meegenomen wordt in het onderzoek is dat het grote invloed heeft voor wie men denkt dat de verkiezingen gaat winnen (Badad et al., 1993; Delavande & Manski, 2012; Mongrain, 2021). De reden dat de voorkeurskandidaat zo'n grote invloed heeft, is dat mensen een voorkeur voor een bepaalde kandidaat hebben en vaak aannemen dat de meeste anderen deze voorkeur delen. Een verklaring is dat zij vaak omgeven worden door mensen met dezelfde voorkeurskandidaat en daardoor het stemmen van mensen buiten hun netwerk niet goed inschatten (Leiter et al., 2018).

Ook zal er gecontroleerd worden voor leeftijd net als gender heeft dit in eerdere onderzoeken geen consistente invloed gehad (Mongrain, 2021). Maar het is toch goed om te controleren voor een mogelijke invloed. Het kan bijvoorbeeld zo zijn dat oudere mensen zij grotere en diverse netwerken hebben waardoor zij een beter inschatting kunnen maken van de verkiezingsuitslag. Of misschien zijn hun netwerken juist kleiner en minder divers en voorspellen zij de verkiezingsuitslag daarom vaker fout.

Het is mogelijk dat andere demografische factoren naast gender en opleidingsniveau ook een rol spelen bij het juist voorspellen van de verkiezingen. Ook deze factoren hebben net als gender onduidelijke resultaten gegeven in eerder onderzoek (Mongrain, 2021). Om de rol van deze factoren uit te sluiten worden ook de factoren etniciteit, gebied van herkomst en religie meegenomen in het onderzoek.

Ook zal er gecontroleerd worden voor verkiezingsjaar, want hoe moeilijk het is om de uitslag van de verkiezingen te voorspellen verschilt per jaar. In het jaar 2016 toen de kandidaten Donald Trump en Hillary Clinton was het verschil in stemmen bijvoorbeeld heel klein. Veel officiële exitpolls hadden de uitslag ook verkeerd voorspeld (CNN, 2016).



Figuur 3.1: Grafische weergave van het onderzoeksmodel

3 Methoden

Beschrijving steekproef

Voor dit onderzoek wordt data uit het American National Election Survey gebruikt (ANES, 2021). Deze organisatie verzamelt sinds 1948 data over de meningen van Amerikanen over de nationale presidentsverkiezingen en politiek in de Verenigde Staten. Deze survey wordt afgenomen bij mensen die stemgerechtigd zijn tijdens de eerst volgende verkiezingen. Voor dit onderzoek zal data gebruikt worden uit 1996, 2004 en 2008. Dit zijn de meest recente verkiezingen waarbij één van de presidentskandidaten duidelijk de winnaar was. Dit betekent dat meer mensen de uitslag juist voorspeld hebben en dat is geschikter voor dit onderzoek. De samenstelling van de vragenlijsten verschilt per jaar maar de data die nodig is voor het beantwoorden van de onderzoeksvraag is in elk jaar verzameld. Er wordt zowel voor als na de verkiezingen een vragenlijst afgenomen. Aangezien het onderzoek gaat over het voorspellen van de verkiezingen zal alleen de data uit de vragenlijsten *vóór* de presidentsverkiezingen relevant zijn. De data die voor dit onderzoek gebruikt wordt komt uit het cumulatieve databestand waarin alle jaren verwerkt zijn. Voor het onderzoek zijn alleen de relevante jaren geselecteerd daarnaast is een versie van de dataset gebruikt waarin geen van de respondenten missende waarden hebben op de gebruikte variabelen.

De vragenlijsten zijn afgenomen in 1996 tussen 3 september en 4 november afgenomen bij 1714 respondenten, in 2004 tussen 7 september bij 1212 respondenten en 1 november en in 2008 tussen 2 september en 4 november bij 2322. In 1996 werden een deel van de respondenten (1316) geselecteerd uit eerdere panels over de verkiezingen uit 1992 en 1994. Een kleiner deel werd geselecteerd met cross-section (398). In 2004 en 2008 werden

alle respondenten geselecteerd door middel van de cross-section methode. In 1996 werden de vragenlijsten zowel *face-to-face* (1646) en via de telefoon afgenomen (67). In 2004 en 2008 werden alle interviews *face-to-face* afgenomen waarbij in 2008 ook gebruik gemaakt werd van een computer. De interviews werden in het algemeen in het Engels afgenomen behalve in 2008 toen ook enkele interviews in het Spaans afgenomen zijn. De respons in 1996 was 71%, de respons in 2004 was 66,1% en de respons in 2008 was 59,5%. De respons was in 1996 waarschijnlijk hoger omdat toen ook respondenten uit eerdere jaren uitgenodigd zijn.

Operationalisatie

De afhankelijke variabele voor dit onderzoek is de juist voorspelde verkiezingsuitslag. De respondenten is gevraagd: 'wie denkt u dat de verkiezingen in november gaat winnen?'. Daarop konden zij antwoorden: de Democratische kandidaat, de Republikeinse kandidaat, een andere kandidaat en weet ik niet. Om weer te geven of de voorspelling juist was is een dummy variabele gemaakt. De respondenten die in 1996 voor de Democratische kandidaat kozen kregen een 1, de respondenten die in 2000 en in 2004 voor de Republikeinse kandidaat kozen kregen een 1 en de respondenten die in 2008 voor de Democratische kandidaat kozen kregen een 1. De overige antwoorden kregen een 0. Dat betekent dat nu 1 een juiste voorspelling is en 0 niet in staat om een juist voorspelling te maken.

Gender is gemeten als een observatie van de interviewer. De oorspronkelijke variabele is gehercodeerd naar een dummyvariabele met 0=man en 1=vrouw.

Om politieke kennis te meten is een schaal van drie verschillende vragen gemaakt. De vragen zijn drie open vragen waarbij de interviewer de naam van een ambtenaar noemt en de respondent het beroep van de ambtenaar moet noemen. De vraag is: 'welke baan of ambt heeft [*naam]*? Waarbij de naam genoemd werd van (1) *de Speaker of the U.S. house, (2) de Vice-President*, (3) *de Chief Justice of the U.S. Supreme Court*. De respondent konden een goed antwoord geven, een verkeerd antwoord of aangeven dat ze het antwoord niet wisten. De vragen zijn eerst gehercodeerd zodat een juist antwoord 1 is en een fout antwoord of een niet beantwoorde vraag 0. Daarna zijn de scores op elke vraag bij elkaar opgeteld zo ontstond een ordinale somschaal. Met 0=geen vraag goed, 1=één vraag goed, 2=twee vragen goed en 3=is drie vragen goed. De betrouwbaarheid van de schaal is een Cornbach's Alpha van 0.499. Dit is relatief laag maar aangezien er geen andere manier is om politieke kennis met deze dataset te meten zal toch deze schaal gebruikt worden.

Het opleidingsniveau is gemeten door mensen te vragen naar hun hoogst behaalde opleiding. In de cumulatieve dataset zijn de antwoordmogelijkheden teruggebracht naar vier categorieën. De vier categorieën van de oorspronkelijke variabele zijn *grade school or less* (basisschool of minder), *high school* (middelbare school), *some college* (een deel van het hoger onderwijs) en *college of advanced degree* (universiteit of een gevorderde diploma). Aangezien dit een categorische variabele is, zijn voor de regressie analyse dummyvariabelen nodig. De referentiegroep zal grade school zijn (de groep is met de laagste opleiding). De dummy variabelen zullen zijn: *grade school* met 1= *grade school* of lager als hoogst behaalde opleiding en 0= niet *grade school* als hoogst behaalde opleiding, high school met 1=*high school* als hoogst behaalde opleiding en 0=*high school* niet als hoogst behaalde opleiding, deel van *college* met 1=deel van *college* als hoogst behaalde opleiding en 0=deel van *college* niet als hoogst behaalde opleiding en *college* of hoger met 1=*college* of hoger als hoogst behaalde opleiding en 0=*college* of hoger niet als hoogst behaalde opleiding.

De controlevariabele leeftijd is de leeftijd van de respondent in jaren op het moment van het interview. De controlevariabele etniciteit is gemeten door de respondent naar zijn of haar etniciteit te vragen. De antwoorden zijn de eerst genoemde etniciteit van de respondent. In de cumulatieve dataset zijn de antwoordmogelijkheden teruggebracht naar vier categorieën. De categorieën zijn Wit (niet-hispanic), Zwart (niet Hispanic), Hispanic en een andere etniciteit of meerdere. Aangezien etniciteit een nominale variabele is moeten er dummies gemaakt worden om de variabele te gebruiken in de regressieanalyse. Als referentiegroep is de witte etniciteit gekozen aangezien dat de grootste groep is. Dan zijn de dummy variabelen: Zwart met 1=Zwart (niet Hispanic) en 0=niet-zwart (niet Hispanic), Hispanic met 1=Hispanic en 0=niet-Hispanic en Anders met 1=Anders en 0=niet-Anders.

De controlevariabele verkiezingsjaar is ordinaal en bestaat uit de variabelen 1996, 2004 en 2008. Daarvoor zijn uit het cumulatieve databestand alleen de relevante verkiezingsjaren geselecteerd. Omdat de afstand tussen alle variabelen niet gelijk is zijn er dummyvariabelen nodig voor de regressieanalyse. Het jaar 1996 is gekozen als referentie groep. De dummy variabelen zijn: 2004 met 1=2004 en 0=niet-2004 en 2008 met 1=2008 en 0=niet-2008.

Voor de controlevariabele gebied van herkomst is de herkomst van de kandidaat genoteerd zoals de Amerikaanse census dat indeelt. De variabele bestaat uit Noordoost, Noord Centraal, Zuid en West. Aangezien dit een nominale variabele is zijn er dummy variabelen gecodeerd. Er is gekozen voor het Zuiden als referentiegroep aangezien het grootste deel van de respondenten daar vandaan komt. Dan zijn de dummyvariabelen:

Noordoost met 1=Noordoost en 0=niet-Noordoost, Noord Centraal met 1=Noord Centraal en 0=niet-Noord Centraal en West met 1=West en 0=niet-West.

De controlevariabele religie is gemeten op basis van verschillende vragen bijvoorbeeld over het soort erediensten dat de respondent bezoekt en of zij zich deel voelen van een bepaalde religie. In de cumulatieve dataset is zijn de vragen over religie teruggebracht naar vier antwoordmogelijkheden met Protestant, (Rooms) Katholiek, Joods, andere religie, geen religie of weet ik niet. Aangezien dit een nominale variabele is zijn er dummy variabelen gecodeerd. Er is gekozen voor Protestant als de referentiegroep omdat dit de grootste groep is. Dan worden de dummy variabelen: (Rooms) Katholiek met 1=(Rooms) Katholiek en 0=niet-(Rooms) Katholiek, Joods met 1=Joods en 0=niet-Joods en anders met 1=anders/geen/weet ik niet en 0=niet- anders/geen/weet ik niet.

De laatste controlevariabele is voorkeurskandidaat. Dit is gemeten door de affiliatie van de respondenten met de twee partijen of hun onafhankelijkheid te meten. Deze variabele bestaat uit meerdere vragen die gecombineerd de affiliatie van de respondent vast stellen. De mogelijke affiliaties zijn: Democraten (*including leaners*), Independents, Republikeinen (*including leaners*) en 0= missing. In dit geval betekent *leaners* dat ook respondenten met een lichte voorkeur ingedeeld zijn bij de Democraten of de Republikeinen. Aangezien dit een nominale variabele is zijn er dummy variabelen gecodeerd. Als referentiegroep is gekozen voor de Independents zodat de regressie het verschil tussen mensen met en zonder een voorkeur weergeeft. De dummyvariabelen zijn Democraat met 1=Democraat en 0=niet-Democraat en Republikein met 1=Republikein en 0= niet-Republikein.

Beschrijvende statistieken

Univariate statistieken

Tabel 3.1: Univariate beschrijving van de statistieken

Variabele	(n=4538) Gemiddelde (SD) /percentages	Min.	Max.
Voorspelling			
Juist	64,5%		
Onjuist	35,5%		
Gender			
Man	45,0%		
Vrouw	55,0%		
Politieke kennis	1,35 (0,90)	0	3
Opleidingsniveau			
Grade school	3,7%		
High school	39,8%		
Deel van college	30,3%		
College of hoger	26,1%		
Leeftijd	47,63 (17,18)	18	93
Verkiezingsjaar			
1996	33,2%		
2004	22,7%		
2008	44,1%		
Etniciteit			
Wit (niet Hispanic)	64,9%		
Zwart (niet Hispanic)	17,3%		
Hispanic	14,5%		
Anders of meerdere	3,3%		
Gebied van herkomst			
Noordoosten	13,6%		
Noord centraal	22,6%		
Zuiden	41,1%		
Westen	22,7%		
Partij affiliatie			
Democraten	54,4%		
Republikeinen	35,5%		
Onafhankelijken	10,0%		
Religie			
Protestant	52,3%		
Katholiek	24,1%		
Joods	1,6%		
Anders/atheïst	21,9%		

In tabel 3.1 zijn de beschrijvende statistieken van de variabelen weergegeven. Wat belangrijk is voor het onderzoek is dat de verkiezingsuitslag voorspelling ongelijk verdeeld is tussen juiste en onjuiste antwoorden. Een grote groep respondenten (64,5%) heeft de president juist voorspeld, dit komt overeen met de verwachtingen op basis van de literatuur die beschreven is in het theoriehoofdstuk (Mongrain, 2021). Het valt ook op dat het aantal mannen (45%) in de steekproef lager is dan het aantal vrouwen (55%). Uitgaande van de populatie zou het aantal mannen en vrouwen ongeveer gelijk moeten zijn. Dat betekent dat ergens in het nemen van de steekproef er iets gebeurd is waardoor mannen in de uiteindelijke dataset onder gerepresenteerd zijn. Aangezien er gewerkt wordt met een grote dataset (n=4538) is het aantal mannen in de steekproef (n=2042) nog steeds groot en zal dat geen grote invloed op de uitkomst van de regressie analyse moeten hebben.

Het valt op dat het politieke kennis gemiddelde met 1,35 redelijk laag is, want de schaal bestaat uit een minimum van nul en een maximum van 3. Dit betekent dat de vragen lastig waren en dat veel respondenten waarschijnlijk één of twee vragen goed hadden.

Het aantal mensen in de referentiegroep van opleidingsniveau, *grade school*, is ook laag (3,7%) vergeleken met de andere opleidingsniveaus. Dit komt ongeveer overeen met de populatie in 2020 had 88.5% van de inwoners van Verenigde Staten van boven de 25 een high school diploma of hoger behaald (American Census Bureau, 2022). Dat betekent dat er in de steekproef minder mensen met het laagst mogelijke opleidingsniveau zitten dan in de populatie.

Bivariate statistieken

Tabel 3.2: Bivariate statistieken

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1 Voorspelling	-									
2 Gender	0,01[b]	-								
3 Politieke kennis	*0,04[b]	*0,16[b]	-							
4 Opleidingsniveau	0,05[b]	*0,06[b]	*0,23[b]	-						
5 Leeftijd	*0,05[c]	0,03[c]	*0,14[c]	**0,24[c]	-					
6 Verkiezingsjaar	*0,24[b]	0,03[b]	*0,21[b]	*0,08[b]	0,03[c]	-				
7 Etniciteit	*0,07[b]	0,04[b]	*0,19[b]	*0,13[b]	*0,16[c]	*0,20[b]	-			
8 Herkomst gebied	0,03[b]	0,03[b]	*0,07[b]	*0,08[b]	*0,07[c]	*0,18[b]	*0,03[b]	-		
9 Voorkeur-kandidaat	**0,13[b]	**0,10[b]	**0,13[b]	**0,09[b]	**0,05[c]	**0,09[b]	**0,24[b]	0,04[b]	-	
10 Religie	0,02[b]	*0,09[b]	*0,06[b]	*0,06[b]	*0,13[c]	*0,06[b]	*0,19[b]	0,03[b]	0,03[b]	-

** significant bij p<0,05 **significant bij* p*<0,01 [a] Pearson Correlatie; [b] Cramers V; [c] correlatie op basis van ANOVA;*

In tabel 3.2, worden de samenhangende maten tussen de variabelen in het onderzoek weergegeven. Een groot deel van de variabelen zijn categorische variabelen. Eerst zal de samenhang tussen de onderzoeksvariabelen besproken worden en daarna andere opvallende correlaties.

Wat opvalt is dat de samenhang tussen de afhankelijke variabele, de voorspelling van de verkiezingsuitslag, en de onafhankelijke variabelen die gebruikt worden om de hypothesen te toetsen gender, politieke kennis en opleidingsniveau laag is. De correlatie tussen gender en de juist voorspelde verkiezingsuitslag is 0,01, wat ontzettend klein en niet significant is. Er lijkt dus geen samenhang tussen gender en het juist voorspellen van de verkiezingsuitslag te zijn. Ook de correlatie tussen politieke kennis en het juist voorspellen van de verkiezingsuitslag is heel laag maar significant (r = 0,04; p<0,05). Het verband tussen politieke kennis en een het juist voorspellen van de verkiezingsuitslag is dus erg klein. De correlatie tussen opleidingsniveau en het juist voorspellen van de verkiezingsuitslag is ook zeer klein met en niet significant (r = 0,05; p>0,05).

Het verband tussen politieke kennis en gender is klein en significant (r = 0,16; p<0,05). Aangezien mannen gecodeerd zijn als 0 en vrouwen als 1 zou dit impliceren dat vrouwen meer politieke kennis hebben dan mannen. Voor het interpreteren van dit verschil is een tabel nodig met de verdelingen tussen mannen en vrouwen op politieke kennis. Dat

vrouwen slechtere politieke kennis hebben dan mannen zou namelijk tegen de verwachtingen op basis van de literatuur ingaan (Dolan & Hansen, 2020; Pereira, 2019).

Tabel 3.3: Kruistabel van gender en politieke kennis

Politieke kennisvragen		0 juiste antwoorden	1 juist antwoord	2 juiste antwoorden	3 juiste antwoorden	Totaal
Gender	Man	304	668	763	288	2041
		(16,7%)	(32,7%)	(37,4%)	(14,1%)	
	Vrouw	565	967	807	158	2497
		(22,6%)	(38,7%)	(32,3%)	(6,3%)	
Totaal		869	1653	1570	446	4538
Pearson Chikwadraattoets					120681	(p<0,001)

Uit tabel 3.3 valt af te lezen dat gender inderdaad een verschil maakt voor de hoe mensen scoren op de politieke kennis schaal. Het is echter duidelijk dat meer vrouwen (61,4%) dan mannen (49,4%) geen of één politieke kennis vraag goed hadden. Ook had bijna een kwart van de vrouwen niet één politieke kennis vraag goed (22,6%). Dit komt overeen met de literatuur, vrouwen hebben minder politieke kennis dan mannen.

De literatuur gaf al aanwijzing dat er mogelijk geen verband tussen demografische kenmerken en het juist voorspellen van de verkiezingsuitslag zouden zijn. Het is mogelijk dat de verbanden zullen veranderen als er voor de andere variabelen gecontroleerd wordt in een regressiemodel. Daarom is het toch nuttig om de analyse voort te zetten ondanks het gebrek aan samenhang in de data.

Daarnaast is het gebruikelijk om de correlaties tussen de controlevariabelen te bespreken die matig zwak tot sterk zijn. Voor deze dataset betekent dat een samenhang van 0,30 of groter. Ook correlaties die kleiner maar opvallend zijn worden besproken.

Tabel 3.4: Kruistabel van juist voorspelde verkiezingsuitslag en verkiezingsjaar

	Voorspelling				
Verkiezingsjaar		1996	2004	2008	Totaal
	Onjuist	189	361	697	1247
		(12,5%)	(35,0%)	(34,8%)	
	Juist	1317	670	1304	3291
		(87,5%)	(65,0%)	(65,2%)	
Totaal		1506	1031	2001	4538

Het verband tussen verkiezingsjaar en het juist voorspellen van de verkiezingsuitslag 0,24 dat is zwak maar aanwezig en bovendien significant. In tabel 3.4 op basis waarvan de correlatie door middel van de Cramer's V berekend is, is te zien dat in 1996 maar een heel klein deel (N=189) van het totaal aantal respondenten (N=1505) de voorspelling fout heeft. In 2008 heeft ongeveer een vierde (N=697) dan de respondenten een foute voorspelling gedaan (N=2001). Daardoor kan de samenhang verklaard worden. Het is duidelijk dat in het ene jaar de verkiezingsuitslag makkelijker te voorspellen dan het andere.

Ook is er een kleine significante samenhang tussen de voorkeurskandidaat en juist voorspelde verkiezingsuitslag (r = 0,13; p<0,05). Uit de literatuur bleek dat de voorspelling sterk samenhangt met de voorkeurskandidaat van de respondent (Badad et al., 1993; Delavande & Manski, 2012; Mongrain, 2021).

Statistisch analyseplan

De onderzoeksvraag zal beantwoord worden door het schatten van een hiërarchische binaire logistische regressie.

Hiervoor zal eerst een model geschat worden met de controlevariabelen, leeftijd, etniciteit, verkiezingsjaar, gebied van herkomst en voorkeurskandidaat. Van dat model zal de afhankelijke variabele de juist geschatte verkiezingsuitslag zijn.

Vervolgens zal het tweede model geschat worden waaraan opleidingsniveau toegevoegd wordt. Dit zal helpen om hypothese 1 te toetsen dat een hoger opleidingsniveau leidt tot een beter kans om de verkiezingsuitslag juist te voorspellen.

In het derde model wordt de variabele gender aan de eerder genoemde variabelen toegevoegd. Dit zal helpen om de hypothese te toetsen dat mannen de verkiezingsuitslag vaker juist voorspellen dan vrouwen (H1).

In het vierde model model worden gender en de controlevariabelen opgenomen en is politieke kennis de afhankelijke variabele. Met dit model wordt getoetst of er een verband is tussen politieke kennis en gender zoals gesteld in hypothese 2.

Vervolgens wordt het vijfde model geschat, dit is model drie waaraan politieke kennis toegevoegd is. Nu kunnen de hellingen van gender in model drie en vijf vergeleken worden. Hiermee kan het mediatie effect getoetst worden en er kan antwoord gegeven worden op de vraag of het verschil in de verkiezingsuitslag juist voorspellen tussen mannen en vrouwen deels verklaard kan worden door hun verschil in politieke kennis (H2). Daarnaast kan worden gekeken of de helling van opleidingsniveau positief is voor de tweede hypothese. Dan kan antwoord gegeven worden op de vraag of een hoger opleidingsniveau inderdaad leidt tot een beter voorspelde verkiezingsuitslag.

Omdat in dit onderzoek veel gebruikt gemaakt wordt van nominale variabelen zal alleen iets gezegd kunnen worden over het verschil in kans vergeleken met de gekozen referentiegroepen. Dit zal niet heel veel uitmaken aangezien de meeste van deze variabelen controlevariabelen zijn.

Ook is het goed om te bedenken dat een waardevolle uitleg van de kans alleen gegeven kan worden op basis van het uitrekenen van de kans voor bepaalde groepen.

4 Resultaten

In dit gedeelte van het onderzoek zullen de resultaten van de binaire logistische regressie weergegeven en kritisch besproken worden. Aangezien dit een logistische regressie is hoeft alleen gecontroleerd te worden voor de assumptie van onafhankelijke waarnemingen. Uit het methode hoofdstuk blijkt dat daar aan voldaan is, het was mogelijk voor één persoon per huishouden om deel te nemen aan het onderzoek (ANES, 2021). Het is daarom onwaarschijnlijk dat er een samenhang is tussen de waarnemingen.

In tabel 4.1 is de hiërarchische binaire logistische regressie analyse weergegeven, inclusief model vier die een lineaire regressie bevat voor het analyseren van de mediatie. In tabel 4.1 is af te lezen dat eerst de controlevariabelen toegevoegd zijn. Dit is zodat in elke volgende stap gecontroleerd wordt voor deze variabelen. De interpretatie van deze hellingen is handig om te kijken of de bevindingen overeenkomen met eerdere onderzoeken. Uit model vijf blijkt dat de juist voorspelde verkiezingsuitslag sterk samenhangt met het jaar waarin het onderzoek uitgevoerd is. Dit is logisch, uit het methoden hoofdstuk tabel 3.4 blijkt dat in 1996, de referentiegroep, relatief de meeste mensen de verkiezingsuitslag juist voorspeld hadden. Het is dan logisch dat een voorspelling uit 2004 (β= -1,34, p=<0,01) of 2008 (β= -1,43, p=<0,01) zorgt voor een kleinere kans om de verkiezingsuitslag juist te voorspellen. Uit eerder onderzoek (Badad et al., 1993; Delavande & Manski, 2012; Mongrain, 2021) werd verwacht dat of mensen de verkiezingsuitslag juist voorspellen samenhangt met wie zij willen dat de verkiezingen wint. Uit model vijf blijkt dat mensen met een voorkeur voor de Democraten een groter kans hebben om de verkieizingen juist te voorspellen (β= 0,54, p=<0,01). Aangezien twee van de drie verkiezingsjaren gewonnen zijn door de Democratische kandidaat, komt dit overeen met eerder onderzoek.

Gender en de mediatie van politieke kennis

De eerste hypothesen die getoetst zullen worden zijn de eerste en de tweede: of vrouwen de verkiezingen slechter voorspellen dan mannen, en of een deel van dit verschil verklaard kan worden door een verschil in politieke kennis.

Ten eerste moet gekeken worden of er een verschil is tussen mannen en vrouwen in het juist voorspellen van de verkiezingsuitslag. Het eerste model waarin dit afgelezen kan worden is model 3. De helling van gender in model 3 is negatief maar niet significant (β= -0,09, p=0,20). Dat de helling negatief is betekent dat de kans dat een vrouw de

verkiezingsuitslag juist voorspelt kleiner is dan dat een man de verkiezingsuitslag juist voorspelt. Het verschil in de odds-ratio tussen mannen en vrouwen in model drie is 0,91, de odds-ratio zijn niet opgenomen in de tabel. Dit betekent dat de odds 9 procent dalen om de verkiezingen juist te van een vrouw ten opzichte van een man. Dat de helling niet significant is betekent dat er in de steekproef wel een verschil gevonden is maar dat het onwaarschijnlijk is dat dit verschil ook gevonden wordt in de populatie.

Het interpreteren van de helling is lastig daarom wordt het verschil in de kans berekend. Aangezien er veel nominale variabelen in de analyse gebruikt worden zal de kans voor een specifieke groep berekend worden. Deze groep bestaat uit de gemiddelde leeftijd (47,63) en voor de demografische kenmerken is gekozen voor de referentiegroepen. De kans dat een man uit deze groep de verkiezingsuitslag juist voorspelt is 0,78, voor een vrouw uit deze groep is de kans 0,76. Vrouwen voorspellen de verkiezingsuitslag dus iets slechter dan mannen maar het verschil is ontzettend klein. Ondanks dit kleine verschil is besloten om de mediatie analyse verder uit te voeren.

De eerste stap daarvoor is om te kijken of er een verband is tussen gender en politieke kennis. Voor het berekenen van dit model is een lineaire regressie gebruikt. Aangezien politieke kennis een schaal met vier punten is worden de voorwaarden voor een lineaire regressie geschonden. De afhankelijke variabele is namelijk niet continue. Een lineaire regressie maakt het wel mogelijk om het verband tussen gender en politieke kennis te onderzoeken terwijl er ook rekening gehouden wordt met de controlevariabelen. Omdat de voorwaarden voor een lineaire regressie geschonden worden is het moeilijk om de eventuele significante resultaten te generaliseren naar de populatie. De helling van gender in het vierde model is negatief en significant (β= -0,26, p=<0,001). In deze lineaire regressie betekent een helling van -0,26 dat vrouwen -0,26 lager scoren dan mannen op de geschatte score gegeven dat alle variabelen constant blijven. De schaal van politieke kennis bestaat uit vier punten dus een daling van -0,26 is redelijk groot. Er is dus een verband gevonden tussen gender en de mediator variabele politieke kennis.

In het vijfde model is af te lezen of dit verband invloed heeft op het eerder gevonden verband tussen gender en de juist voorspelde verkiezingsuitslag. In het vijfde model valt af te lezen dat de helling van gender -0,08 en niet significant is (p=0,25). De odds-ratio die bij deze helling hoort is 0,92. Dit betekent dat de odds met 8 procent dalen om de verkiezingen juist te voorspellen voor een vrouw vergeleken met een man. De helling in model vijf is kleiner dan de helling in model drie. Dit wijst er op dat politieke kennis inderdaad een deel van het verschil tussen mannen en vrouwen in het voorspellen van de verkiezingsuitslag

kan verklaren. Het verschil is echter heel klein, het verschil in de odds-ratio is bijvoorbeeld maar één procent en niet significant.

De data ondersteunt de hypothese niet dat vrouwen de verkiezingsuitslag slechter voorspellen dan mannen. De data ondersteunt ook de tweede hypothese niet, dat een deel van dat verschil verklaard kan worden een verschil politieke kennis tussen mannen en vrouwen. De effecten die gevonden zijn gaan wel in de voorspelde richting maar de effecten zijn erg klein. Bovendien zijn de hellingen niet significant dus is het onwaarschijnlijk dat deze resultaten ook in de populatie gevonden zullen worden.

Tabel 4.1:Hiërarchische binaire logistische regressie analyse met de afhankelijke variabele juist voorspelde verkiezingsuitslag en in model 4 een lineaire regressie met politieke kennis als afhankelijke variabele

	Model 1[a]		Model 2[a]		Model 3[a]		Model 4[b]		Model 5[a]		VIF	*N*
	b	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*		
Constante	1,98**	0,18	1,44**	0,26	1,49**	0,26	0,26**	0,09	1,49**	0,27		
Politieke kennis									0,03	0,05	1,38	
Gender (Man= 0; Vrouw=1)					-0,09	0,07	-0,26**	0,02	-0,08	0,07	1,05	
Opleidingsniveau												
Grade school of minder (ref)												
High school			0,55**	0,19	0,55**	0,19	0,51**	0,06	0,53**	0,19	7,69	
Deel van college			0,43*	0,19	0,43*	0,19	0,85**	0,07	0,41*	0,20	7,33	
College of hoger			0,57**	0,20	0,57**	0,20	1,20**	0,07	0,53**	0,20	7,09	
Leeftijd	-0,01**	0,01	-0,01*	0,01	-,01*	0,00	0,01**	0,00	0,01*	0,00	1,17	
Verkiezingsjaar												
1996 (ref)												
2004	-1,34**	0,10	-1,34**	0,10	-1,34**	0,10	-0,28**	0,03	-1,34**	0,10	1,34	
2008	-1,43**	0,10	-1,44**	0,10	-1,44**	0,10	-0,17**	0,03	-1,43**	0,10	1,45	
Etniciteit												
Wit, niet-Hispanic (ref)												
Zwart niet-Hispanic	0,34**	0,11	0,35**	0,11	0,35**	0,11	-0,40**	0,04	-0,01*	0,00	1,43	
Hispanic	-0,03	0,11	0,03	0,11	0,02	0,11	-0,28**	0,04	0,36**	0,11	1,36	
Meerdere, andere	-0,30	0,19	-0,29	0,19	-0,29	0,19	-0,16*	0,07	0,03	0,11	1,04	
Gebied van herkomst												
Zuiden (ref)												
Noordoost	0,01	0,12	0,01	0,12	0,01	0,12	0,10	0,04	0,01	0,12	1,24	
Noord Centraal	-0,08	0,09	-0,09	0,10	-0,09	0,10	0,03	0,03	-0,09	0,10	1,30	
West	-0,07	0,09	-0,08	0,66	-0,08	0,09	0,03	0,03	-0,08	0,10	1,32	
Voorkeurskandidaat												
Onafhankelijken (ref)											3,15	
Republikein	-0,04	0,12	-0,05	0,12	-0,06	0,12	0,35**	0,04	-0,07	0,12		
Democraat	0,56**	0,12	0,55**	0,12	0,55**	0,12	0,31**	0,04	0,54**	0,12	3,05	
Religie												
Protestants (ref)												
Katholiek	0,06	0,09	0,05	0,10	0,05	0,09	-0,02	0,03	0,05	0,09	1,35	
Joods	0,36	0,32	0,31	0,32	0,31	0,32	0,14	0,09	0,30	0,32	1,06	
Andere en geen religie	0,07	0,10	0,07	0,09	0,06	0,06	-0,05	0,03	0,06	0,09	1,26	
Hosmer-Lemeshowtoets	69.028**		74.610**		57.721**				49.722**			
Deviance	4941,661		4930,974		4929,326				4928,971			4538
X^2-toets	69,028**		10,687*		1,648				0,355			

**significant bij $p<0,01$; *significant bij $p<0,05$

[a] afhankelijke variabele is juist voorspelde verkiezingsuitslag (0=onjuist; 1=juist); [b] afhankelijke variabele is politieke kennis

Opleidingsniveau

Voor het toetsen van de derde hypothese, mensen met een hoger opleidingsniveau voorspellen de verkiezingsuitslag beter dan mensen met een lager opleidingsniveau, worden de hellingen uit het vijfde model bekeken omdat daarin gecontroleerd wordt voor alle variabelen.

Omdat opleidingsniveau nominaal is wordt er gebruikt gemaakt van een referentiegroep. In dit geval is dat de groep mensen die een opleiding heeft van '*grade school*' of lager. Alle hellingen geven het verschil tussen de betreffende groep en deze referentiegroep. De helling van alle opleidingsniveaus zijn in model vijf positief en significant. Dat betekent dat iemand met een '*high school*' diploma (β= 0,53, p= <0,05), die een deel van '*college*' afgerond hebben (β= 0,40, p=<0,05) of college of een hogere diploma behaald hebben (β= 0,53, p=<0,01), meer kans heeft om de verkiezingen juist te voorspellen, gegeven dat alle andere variabelen constant blijven.

Hoe groot deze verschillen zijn is alleen uit te rekenen voor een specifieke groep omdat een groot deel van de variabelen bestaat uit nominale variabelen. Toch is zo meer inzicht te krijgen in het verschil tussen de groepen. Deze groep bestaat uit mannen van de gemiddelde leeftijd (47,63) en de referentiegroepen. Er is gekozen voor twee juiste kennisvragen omdat het merendeel van de respondenten één of twee vragen goed had. Dat betekent dat een man met een *grade school* opleiding of minder, een kans van 0,79 heeft om de verkiezingsuitslag juist te voorspellen. Voor een man een *high school* opleidingsniveau uit dezelfde groep geldt een kans van 0,86 om de verkiezingsuitslag juist te voorspellen. Uit die kansen en odds ratio's blijkt dat de kans om de verkiezingen juist te voorspellen met bijna 0,10 verschilt voor twee mensen die in de hiervoor genoemde categorieën vallen. Dat is een redelijk groot verschil. Voor iemand uit dezelfde groep maar die een deel van *college* afgerond heeft een kans van 0,49 om de verkiezingsuitslag juist te voorspellen. Iemand die in plaats daarvan *college* afgerond heeft of een hoger heeft een kans van 0,86 om de verkiezingen juist te voorspellen. Hieruit blijkt dat de kans een stuk hoger is om de verkiezingen juist te voorspellen als het opleidingsniveau hoger is dan grade school.

Tabel 4.2: resultaten logistische regressie waarbij de referentiegroep in model 5a veranderd is en de vergeleken groepen weergegeven zijn

Variabele in Model 5[a]	
Deel van college= high school	*p*=0,137
College of hoger= high school	*p*=0,976
College of hoger= deel van college	*p*=0,139

Om te kijken wat het verschil is tussen de andere opleidingsniveaus is de analyse nogmaals uitgevoerd met steeds een ander opleidingsniveau als referentiegroep In tabel 4.2 is steeds de significantie van de helling weergegeven van het verschil tussen beide groepen. Uit de resultaten in tabel 4.2 blijkt dat geen van de andere hellingen significant zijn. Dit betekent dat het gevonden tussen de andere opleidingsniveaus waarschijnlijk niet gevonden zal worden in de populatie. De derde hypothese, mensen met een hoger opleidingsniveau zullen de verkiezingsuitslag beter voorspellen dan mensen met een lager opleidingsniveau, wordt dus alleen ondersteund als het gaat om het laagste opleidingsniveau vergeleken met de hogere opleidingsniveaus. De hypothese wordt niet ondersteund voor de hogere opleidingsniveaus.

Model evaluatie

Als laatste wordt het gebruikte model geëvalueerd. Er wordt gekeken hoe goed de variabelen bij elkaar passen en of het model bij de data past en of het toevoegen van de variabelen ervoor zorgt dat de modellen verbeteren.

Ten eerste is hiervoor de VIF berekend een maat die nagaat hoeveel multicollineariteit er is, dit betekent dat de variabelen te sterk met elkaar samenhangen. Vaak wordt uitgegaan van te veel multicollineariteit als de VIF groter is dan 4. Voor dit model zijn dat de variabelen die opleidingsniveau meten, die zijn allen groter dan 7. Dit komt waarschijnlijk omdat één van de opleiding dummies de anderen voorspelt. Het aanpassen van deze variabele is niet mogelijk maar waarschijnlijk heeft dit geen invloed op de regressieanalyse.

Als laatste kan er iets gezegd worden over de *modelfit*, hoe goed past het gekozen model bij de data. De Hosmer-Lemeshow test van het laatste model is significant, wat bij deze toets betekent dat het model niet goed bij de data past (X^2=49.722, *p*=<0,001). Hier moet rekening mee gehouden worden tijdens het bespreken van de uiteindelijke resultaten. De Chi-kwadraat toets in tabel 4.1 geeft de verbetering van het model ten opzichte van het vorige model aan. Het eerste model geeft een significante verbetering van het model ten opzichte van het lege model

weer (X^2=69,028, p=<0,001). Het tweede model waaraan opleidingsniveau is toegevoegd, is ook een significante verbetering te opzichte van model één (X^2=10,687, p=<0,05). Het derde model waaraan gender toegevoegd is, is geen verbetering ten opzichte van het tweede model (X^2=1,648, p=0,12). Ook het vijfde model waaraan politieke kennis toegevoegd is, is geen verbetering ten opzichte van het vierde model (X^2=0,335, p=0,55). Dit is geen verrassing aangezien de hellingen van opleidingsniveau significant waren en de interpretatie van de kansen ook een groot verschil liet zien. Die variabele kan dus een deel van de verschillen in voorspellingen tussen respondenten verklaren. De hellingen van gender en politieke kennis waren echter klein en bovendien was die van gender niet significant. Het is logisch dat het toevoegen van deze variabelen niet tot betere modellen leidden.

De deviance is ook een maat voor hoe goed het model bij de data past. Als er een verbetering zou zijn met elk model dan zou de deviance duidelijk moeten dalen. Per model daalt de deviance wel iets maar niet in grote stappen.

De conclusie over de *model fit* is dus dat de geschatte modellen niet goed bij de data passen. Vooral het toevoegen van de variabelen gender en politieke kennis lijken niet tot nauwelijks verbetering van het model te geven. Deze factoren lijken dus weinig effect te hebben op het juist voorspellen van de verkiezingsuitslag. Het kan zijn dat dit ligt aan de manier waarop politieke kennis geoperationaliseerd is. Mogelijk zou een andere soort analyse beter bij de data passen. Het kan ook zijn dat de demografische kenmerken die gebruikt zijn in dit onderzoek bijna geen invloed hebben op het juist voorspellen van de verkiezingen.

5 Conclusie en discussie

In dit onderzoek werd onderzocht wat de invloed van demografische factoren op het juist voorspellen van de verkiezingsuitslag is met drie verschillende hypothesen. De eerste hypothese was dat mannen de verkiezingsuitslag beter kunnen voorspellen dan vrouwen. Het onderzoek heeft laten zien dat in de steekproef mannen inderdaad de verkiezingen vaker juist voorspellen dan vrouwen. Het verschil tussen mannen en vrouwen was echter klein en niet statisch significant. De analyse ondersteunt de hypothese dus niet.

De tweede hypothese was dat een deel van het verschil tussen mannen en vrouwen in het juist voorspellen van de verkiezingsuitslag, verklaard kan worden doordat mannen gemiddeld meer kennis hebben over politiek dan vrouwen. Uit dit onderzoek blijkt dat mannen inderdaad meer politieke kennis hebben dan vrouwen, zoals eerder onderzoek ook heeft laten zien (Pereira, 2019). Ook verklaart dit een deel van het gevonden verschil tussen mannen en vrouwen in het juist voorspellen van de verkiezingsuitslag. Maar aangezien de verschillen ontzettend klein waren en niet significant wordt de hypothese niet gesteund door de analyse.

De derde en laatste hypothese was dat mensen met een hoger opleidingsniveau de verkiezingen beter kunnen voorspellen dan mensen met een lager opleidingsniveau. Het onderzoek laat zien dat mensen met een hoger opleidingsniveau dan *grade school* inderdaad een betere voorspelling doen dan mensen met een *grade school* opleiding of lager. Dit geldt echter alleen voor een *grade school* opleidingsniveau. Er waren geen verschillen tussen de andere opleidingsniveaus. Dus alleen een hoger opleidingsniveau dan het laagste opleidingsniveau zorgt voor een betere voorspelling. De derde hypothese wordt dus deels ondersteund door de analyse.

De theorie voorspelde dat er vooral een verschil zou zijn tussen mensen met en zonder een universitaire of *college* opleiding (Lewis-Beck & Tien, 1999). Deze theorie wordt in dit onderzoek echter niet ondersteund. Een groter deel van de theorie die leidde tot de verwachting dat hoger opgeleiden een betere voorspelling zouden doen is gebaseerd op de samenstelling van netwerken (Leiter et al., 2018). De daadwerkelijke samenstelling van de netwerken is echter niet meegenomen in dit onderzoek. Het kan zo zijn dat de aanname over de samenhang tussen opleidingsniveau en de diversiteit van het persoonlijke netwerk niet klopte. Het kan ook zijn dat opleidingsniveau niet de juiste manier is om de diversiteit van een netwerk te meten. Mogelijk zal een onderzoek dat specifiek kijkt naar de diversiteit in netwerken van mensen wel vinden dat een netwerk met meer diversiteit zorgt voor een betere voorspelling van de verkiezingsuitslag.

Een voordeel van dit onderzoek is dat er veel respondenten in de steekproef zaten. Bovendien zijn er meerdere verkiezingsjaren meegenomen in de analyse. De steekproef is ook redelijk representatief voor de Amerikaanse bevolking in het algemeen. Dit betekent dat het mogelijk is om de resultaten te generaliseren.

Een probleem met het onderzoek is dat politieke kennis niet op een ideale manier gemeten is. De variabele waarmee politieke kennis gemeten is bestond maar uit drie vragen. Deze vragen waren *face-to-face* afgenomen en uit eerder onderzoek blijkt al dat dit kan leiden tot risico vermijdend gedrag (Pereira, 2019). De vragen gingen ook voornamelijk over mannelijke bewindslieden en alleen over de binnenlandse federale politiek. Het is vreemd dat met deze niet ideale vragen geen groter verschil in politieke kennis tussen mannen en vrouwen gevonden is. Een onderzoek waarbij uitgebreidere politieke kennis vragen gesteld worden op een andere manier dan *face-to-face* zou moeten kunnen verklaren wat de precieze invloed van politieke kennis op het juist voorspellen van de verkiezingen is.

In de theorieparagraaf werd beargumenteerd dat betere vrouwelijke representatie in de politiek een oplossing zou kunnen zijn voor een verschil in politieke kennis tussen mannen en vrouwen (Pereira, 2019). Zoals in de vorige paragraaf werd uitgelegd is het vreemd dat er geen verschil tussen mannen en vrouwen gevonden is. Vrouwelijke representatie in de politiek blijft belangrijk maar mogelijk is het verschil in politieke kennis daar geen goed argument voor. Beleidsmakers die willen argumenteren voor meer vrouwelijke participatie in de hoogste politiek zouden daarom opzoek kunnen gaan naar wetenschappelijke argumenten waar meer bewijs voor gevonden is.

Dit onderzoek draagt bij aan de kennis over *citizen forecasting* en vooral aan de rol van demografische kenmerken daarin. Uit dit onderzoek blijkt dat demografische kenmerken bijna geen of geen invloed hebben op het juist voorspellen van de verkiezingen. Wat vooral uit lijkt te maken is wie de ondervraagde wil dat gaat winnen en het jaar waarin de voorspelling gedaan wordt.

6 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/

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.

CNN. (2016, 23 november). *Exit Polls 2016.* Geraadpleegd op 3 april 2022, van https://edition.cnn.com/election/2016/results/exit-polls

Coleman, K. J., Neale, T. H., & Cantor, J. E. (2001). Presidential Elections in the United States: A Primer. Nova Publishers.

Delavande, A., & Manski, C. F. (2012). Candidate preferences and expectations of election outcomes. *Proceedings of the National Academy of Sciences*, *109*(10), 3711-3715.

Dolan, K., & Hansen, M. A. (2020). The variable nature of the gender gap in political knowledge. *Journal of Women*, Politics & Policy, 41(2), 127-143.

Leiter, D., Murr, A., Ramírez, E. R., & Stegmaier, M. (2018). Social networks and citizen election forecasting: The more friends the better. *International Journal of Forecasting*, *34*(2), 235-248.

Lewis-Beck, M. S., & Tien, C. (1999). Voters as forecasters: a micromodel of election prediction. *International Journal of Forecasting*, *15*(2), 175-184.

Graupensperger, S., Turrisi, R., Jones, D., & Evans, M. B. (2020). Longitudinal associations between perceptions of peer group drinking norms and students' alcohol use frequency within college sport teams. *Alcoholism: Clinical and Experimental Research, 44*(2), 541-552.

Murr, A. E. (2016). The wisdom of crowds: what do citizens forecast for the 2015 British general election?. *Electoral Studies*, *41*, 283-288.

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, 102317.

NOS. (2021a, 11 november) *Handmatige hertelling van de stemmen in Amerikaanse staat Georgia*. Geraadpleegd op 3 april 2022, van https://nos.nl/collectie/13849/artikel/2356183-handmatige- hertelling-van-de-stemmen-in-amerikaanse-staat-georgia

NOS. (2021b, 7 januari). *In beeld: zo verliep de bestorming van het Capitool in Washington*. Geraadpleegd op 3 april 2022, van https://nos.nl/collectie/13854/artikel/2363360-in-beeld -zo-verliep-de-bestorming-van-het-capitool-in-washington

Pereira, F. B. (2019). Gendered political contexts: The gender gap in political knowledge. *The Journal of Politics*, *81*(4), 1480-1493.

United States Census Bureau. (z.d.). Quick Facts United States. Geraadpleegd op 26 mei 2022, van https://www.census.gov/quickfacts/fact/table/US/PST04522

RTL Nieuws. (2021, 7 januari). *De gezichten achter de bestoriming: 'Trump moet weten dat we hem steunen*. Geraadpleegd op 3 april 2021, van https://www.rtlnieuws.nl/nieuws/buitenland/ artikel/5207228/bestorming-capitool-rellen-geweld-trump-aanhangers-washington-vs

Bijlage 1

1 Voorspelde verkiezingsuitslag

De vraag die voor deze variabele gebruikt is, is VCF0700: Who Will Be Elected President in November?

De antwoordmogelijkheden voor deze vraag waren de volgende:

1. Democratic candidate 2. Republican candidate 7. Other candidate 8. DK; depends 9. NA; refused; no Pre IW INAP. question not used

```
FREQUENCIES VARIABLES=VCF0700
  /ORDER=ANALYSIS.
```

Who Will Be Elected President in November

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1. Democratic candidate	4166	59,1	59,2	59,2
	2. Republican candidate	2317	32,8	32,9	92,2
	7. Other candidate	89	1,3	1,3	93,4
	8. DK; depends	461	6,5	6,6	100,0
	Total	7033	99,7	100,0	
Missing	9. DK; NA; no Pre IW	22	,3		
Total		7055	100,0		

Het is duidelijk dat meer mensen voorspellen dat de Democratische kandidaat wint dan dat de Republikeinse kandidaat wint.

Van deze variabele is een variabele gemaakt die aangeeft of de respondent de verkiezingen juist voorspeld heeft. Dit is gedaan door per jaar te berekenen wat de juiste verkiezingsuitslag was. Mensen die geantwoord hebben dat ze het niet wisten of de verkeerde kandidaat voorspeld hebben krijgen een score van 0. Mensen die de juiste kandidaat voorspeld hebben krijgen een score van 1.

```
COMPUTE Goed_geschat=((VCF0004 = 2008) & (VCF0700 = 1)) | ((VCF0004 = 2004) &
   (VCF0700 = 2)) | ((VCF0004 = 2000) & (VCF0700 = 2)) | ((VCF0004 = 1996) & (VCF0700=1))
EXECUTE.
```

De nieuwe variabele ziet er nu uit als volgt:

```
FREQUENCIES VARIABLES=Goed_geschat
  /ORDER=ANALYSIS.
```

Goed_geschat

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Fout	2498	35,4	35,5	35,5
	Goed	4535	64,3	64,5	100,0
	Total	7033	99,7	100,0	
Missing	System	22	,3		
Total		7055	100,0		

Om te controleren of het goed gegaan is hebben we een kruistabel met de voorspelling en het verkiezingsjaar nodig.

Year of Study * Who Will Be Elected President in November Crosstabulation

Count

		Who Will Be Elected President in November				
		1. Democratic candidate	2. Republican candidate	7. Other candidate	8. DK; depends	Total
Year of Study	1996	1490	165	22	33	1710
	2000	848	783	47	126	1804
	2004	341	775	1	93	1210
	2008	1487	594	19	209	2309
Total		4166	2317	89	461	7033

```
CROSSTABS
  /TABLES=VCF0004 BY VCF0700
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Het berekenen van de categorieën leidt tot: 1996 Democraat (1490). 2000 Republikein (783). 2004 Republikein (775). 2008 Democraat (1487). Dat is 4.535 en dat klopt met de frequentietabel van de gemaakte variabele.

3) Hoe ziet de uiteindelijke variabele eruit?

Uit de uiteindelijke dataset zijn alle missende waarden gehaald wat leidt tot de volgende frequenties.

```
FREQUENCIES VARIABLES=Gen_Dum, Pol_Ken, VCF0101, VCF0105b, Goed_geschat, VCF0110, VCF0128, VCF0112,
VCF0303, Voorkeur
  /ORDER=ANALYSIS.
```

Goed_geschat

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Fout	1247	27,5	27,5	27,5
	Goed	3291	72,5	72,5	100,0
	Total	4538	100,0	100,0	

2 Gender

De vraag die voor deze variabele gebruikt is, is VCF0104 Respondent - Gender. De gender van de respondent is (in de voor deze analyse gebruikte jaren) gebaseerd op een observatie van de interviewer.

Dit leidt in de dataset tot de volgende antwoordmogelijkheden:

1 Male 2 Female en 3 Other 0 geen antwoord

1) Hoe ziet de oorspronkelijke variabele eruit?

```
FREQUENCIES VARIABLES=VCF0104
  /ORDER=ANALYSIS.
```

Respondent - Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1. Male	3123	44,3	44,3	44,3
	2. Female	3932	55,7	55,7	100,0
	Total	7055	100,0	100,0	

Vrouwen zijn licht over gerepresenteerd in deze dataset. Het verschil is echter niet heel groot en zal waarschijnlijk geen grote invloed hebben.

2) Welke bewerkingen heb je op de variabele/ items uitgevoerd?

Ik heb de variabele geherocodeerd naar een dummyvariabele zodat de nominale variabele bruikbaar is in een regressie:

```
RECODE VCF0104 (1=0) (2=1) (3=SYSMIS) INTO Gen_Dum.
VARIABLE LABELS  Gen_Dum 'Gen_Dum'.
EXECUTE.
FREQUENCIES VARIABLES=Gen_Dum
  /ORDER=ANALYSIS.
```

Gen_Dum

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Man	3123	44,3	44,3	44,3
	Vrouw	3932	55,7	55,7	100,0
	Total	7055	100,0	100,0	

Toelichting: alles is goed gegaan ze zien er hetzelfde uit.

3) Hoe ziet de uiteindelijke variabele eruit?

```
FREQUENCIES VARIABLES=Gen_Dum, Pol_Ken, VCF0101, VCF0105b, Goed_geschat, VCF0110, VCF0128, VCF0112,
VCF0303, Voorkeur
```

Respondent - Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1. Male	2041	45,0	45,0	45,0
	2. Female	2497	55,0	55,0	100,0
	Total	4538	100,0	100,0	

3 Politieke kennis

Voor het maken van deze schaal is gebruik gemaakt van de volgende vragen:
VCF9260 What job or office is held by named Speaker of the U.S. House,
VCF9261 What job or office is held by named Vice-President,
VCF9262 What job or office is held by named Chief Justice of the U.S. Supreme Court
De interviewer vroeg welke baan of welk ambt degene had die bij naam genoemd werd. Bijvoorbeeld: What job or office is held by Kamala Harris? De respondent moet dan antwoorden vice president.

De antwoordmogelijkheden uit de dataset waren: 1 correct 2 No identification, identification is incomplete or wrong en -9 is system missing

1) Hoe ziet de oorspronkelijke variabele eruit?

```
FREQUENCIES VARIABLES=VCF9260
  /ORDER=ANALYSIS.
```

What job or office is held by named Speaker of the U.S. House

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1. Correctly identifies	1883	26,7	40,0	40,0
	2. No identification; identification is incomplete or wrong	2819	40,0	60,0	100,0
	Total	4702	66,6	100,0	
Missing	-9. Inap, no post data	546	7,7		
	System	1807	25,6		
	Total	2353	33,4		
Total		7055	100,0		

Iets meer dan de helft van de respondenten had deze vraag fout.

```
FREQUENCIES VARIABLES=VCF9261
  /ORDER=ANALYSIS.
```

What job or office is held by named Vice-President

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1. Correctly identifies	3678	52,1	78,2	78,2
	2. No identification; identification is incomplete or wrong	1024	14,5	21,8	100,0
	Total	4702	66,6	100,0	
Missing	-9. Inap, no post data	546	7,7		
	System	1807	25,6		
	Total	2353	33,4		
Total		7055	100,0		

Een groot deel van de respondenten hebben deze vraag goed.

```
FREQUENCIES VARIABLES=VCF9261
  /ORDER=ANALYSIS.
```

What job or office is held by named Chief Justice of the U.S. Supreme Court

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1. Correctly identifies	896	12,7	14,3	14,3
	2. No identification; identification is incomplete or wrong	5361	76,0	85,7	100,0
	Total	6257	88,7	100,0	
Missing	-9. Inap, no post data	798	11,3		
Total		7055	100,0		

Dit lijkt de lastigste vraag uit de dataset want een overgroot deel van de respondenten heeft deze vraag fout.

2) Welke bewerkingen heb je op de variabele/ items uitgevoerd?

De variabele is zo gehercodeerd dat een juist antwoord op de vraag een score van 1 representeert en een onjuist of onvolledig antwoord een 0.

```
RECODE VCF9260 (1=1) (2=0) (-9=SYSMIS) INTO Speaker_Dum.
EXECUTE.
RECODE VCF9261 (1=1) (2=0) (-9=SYSMIS) INTO VicePres_Dum.
EXECUTE.
RECODE VCF9262 (1=1) (2=0) (-9=SYSMIS) INTO ChiefJus_Dum.
EXECUTE.
FREQUENCIES VARIABLES= Speaker_Dum VicePres_Dum ChiefJus_Dum
  /ORDER=ANALYSIS.
```

Speaker_Dum

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Fout	2819	40,0	60,0	60,0
	Goed	1883	26,7	40,0	100,0
	Total	4702	66,6	100,0	
Missing	System	2353	33,4		
Total		7055	100,0		

VicePres_Dum

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Fout	1024	14,5	21,8	21,8
	Goed	3678	52,1	78,2	100,0
	Total	4702	66,6	100,0	
Missing	System	2353	33,4		
Total		7055	100,0		

ChiefJus_Dum

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Fout	5361	76,0	85,7	85,7
	Goed	896	12,7	14,3	100,0
	Total	6257	88,7	100,0	
Missing	System	798	11,3		
Total		7055	100,0		

De verdelingen zien er hetzelfde uit als bij de originele variabelen dus het hercoderen van de variabelen is gelukt.

Betrouwbaarheid

Om te kijken of deze variabelen samen een schaal mogen vormen moet de betrouwbaarheid van de schaal gecontroleerd worden.

```
RELIABILITY
  /VARIABLES=VCF9260  VCF9261  VCF9262
  /SCALE('ALL VARIABLES') ALL
  /MODEL=ALPHA.
```

Reliability Statistics

Cronbach's Alpha	N of Items
,499	3

Een Cornbach's alpha van .499 is eigenlijk te laag voor het maken van een schaal. In de ideale omstandigheden zou een ander manier gevonden kunnen worden om politieke kennis te meten. Dit is in dit geval niet mogelijk dus zal toch deze schaal gebruikt moeten worden.

Hiervoor is een nieuwe variabele gecodeerd die de somschaal is van de dummy variabelen:

```
COMPUTE Pol_Ken=Speaker_Dum + VicePres_Dum+ ChiefJus_Dum.
EXECUTE.
FREQUENCIES VARIABLES= Pol_Ken
  /ORDER=ANALYSIS.
```

Pol_Ken

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Geen vragen goed	927	13,1	19,7	19,7
	1 vraag goed	1712	24,3	36,4	56,1
	2 vragen goed	1608	22,8	34,2	90,3
	Alle vragen goed	455	6,4	9,7	100,0
	Total	4702	66,6	100,0	
Missing	System	2353	33,4		
Total		7055	100,0		

Deze verdeling lijkt overeen te komen met de voorgaande variabelen. Het coderen van de variabelen is dus geslaagd.

3) Hoe ziet de uiteindelijke variabele eruit?

FREQUENCIES VARIABLES=Gen_Dum, Pol_Ken, VCF0101, VCF0105b, Goed_geschat, VCF0110, VCF0128, VCF0112, VCF0303, Voorkeur

Pol_Ken

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Geen vragen goed	869	19,1	19,1	19,1
	1 vraag goed	1653	36,4	36,4	55,6
	2 vragen goed	1570	34,6	34,6	90,2
	Alle vragen goed	446	9,8	9,8	100,0
	Total	4538	100,0	100,0	

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Pol_Ken	4538	,00	3,00	1,3510	,89863
Valid N (listwise)	4538				

```
DESCRIPTIVES VARIABLES=Pol_Ken
  /STATISTICS=MEAN STDDEV MIN MAX.
```

4 Opleidingsniveau

De variabele die voor deze variabele gebruikt is, is VCF0110 Respondent - Education 4-category dit is een samengestelde variabele die andere vragen over opleidingsniveau samenvoegt tot vier antwoordmogelijkheden.

De antwoordmogelijkheden zijn:

1. Grade school or less (lager dan middelbare school)
2. High school (middelbare school)
3. Some college (Een deel van hoger onderwijs)
4. College or advanced degree (universiteit of gevorderd diploma)

0. missing values

1) Hoe ziet de oorspronkelijke variabele eruit?

```
FREQUENCIES VARIABLES=VCF0110
  /ORDER=ANALYSIS
```

Respondent - Education 4-category

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1. Grade school or less (0-8 grades); 2020: less than high school credential	274	3,9	3,9	3,9
	2. High school (12 grades or fewer, incl. non-college	2772	39,3	39,4	43,3
	3. Some college (13 grades or more but no degree;	2104	29,8	29,9	73,2
	4. College or advanced degree (no cases 1948)	1881	26,7	26,8	100,0
	Total	7031	99,7	100,0	
Missing	0. DK; NA; no Pre IW; short-form 'new' Cross Section	24	,3		
Total		7055	100,0		

.

Het aantal mensen dat alleen een grade school diploma heeft is ontzettend klein vergeleken met de andere opleidingsniveaus. Ook blijkt dat de meeste respondenten een high school diploma als hoogste opleiding hebben. Daarna ‘some college’ en daarna college of advanced degree. Maar deze categorieën zijn ongeveer vergelijkbaar.

2) Welke bewerkingen heb je op de variabele/ items uitgevoerd?

Voor de regressie analyse moet deze nominale variabele een dummy worden. Als referentiegroep is gekozen voor Grade School.

```
RECODE VCF0110 (1=0) (2=1) (3=0) (4=0) (0=SYSMIS) INTO High_School.
VARIABLE LABELS  High_School 'High_School'.
EXECUTE.
```

High_School

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	niet-High School	4259	60,4	60,6	60,6
	High School	2772	39,3	39,4	100,0
	Total	7031	99,7	100,0	
Missing	System	24	,3		
Total		7055	100,0		

De frequenties komen overeen dus het is gelukt

```
RECODE VCF0110 (1=0) (4=0) (0=SYSMIS) (2=0) (3=1) INTO Some_College.
VARIABLE LABELS  Some_College 'Some_College'.
EXECUTE.
```

Some_College

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Niet Some College	4927	69,8	70,1	70,1
	Some College	2104	29,8	29,9	100,0
	Total	7031	99,7	100,0	
Missing	System	24	,3		
Total		7055	100,0		

De frequenties kloppen.

```
RECODE VCF0110 (1=0) (0=SYSMIS) (2=0) (4=1) (3=0) INTO College_advanced_degree.
VARIABLE LABELS  College_advanced_degree 'College_Advanced_Degree'.
EXECUTE.
```

College_Advanced_Degree

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Geen college of advanced degree	5150	73,0	73,2	73,2
	College of advanced degree	1881	26,7	26,8	100,0
	Total	7031	99,7	100,0	
Missing	System	24	,3		
Total		7055	100,0		

Het ziet er hetzelfde uit als de onbewerkte variabele dus het hercoderen van de variabele is gelukt.

```
RECODE VCF0110 (1=1) (2=0) (3=0) (4=0) (0=SYSMIS) INTO Grade_School.
VARIABLE LABELS  Grade_School 'Grade_School'.
EXECUTE.
```

Grade_School

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Niet GRade School	6757	95,8	96,1	96,1
	Grade School	274	3,9	3,9	100,0
	Total	7031	99,7	100,0	
Missing	System	24	,3		
Total		7055	100,0		

Het ziet er hetzelfde uit als de onbewerkte variabele dus het hercoderen van de variabele is gelukt.

3) Hoe ziet de uiteindelijke variabele eruit?

```
FREQUENCIES VARIABLES=Gen_Dum, Pol_Ken, VCF0101, VCF0105b, Goed_geschat, VCF0110, VCF0128, VCF0112,
VCF0303, Voorkeur
```

Respondent - Education 4-category

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1. Grade school or less (0-8 grades); 2020: less than high school credential	168	3,7	3,7	3,7
	2. High school (12 grades or fewer, incl. non-college	1808	39,8	39,8	43,5
	3. Some college (13 grades or more but no degree;	1377	30,3	30,3	73,9
	4. College or advanced degree (no cases 1948)	1185	26,1	26,1	100,0
	Total	4538	100,0	100,0	

5 Leeftijd

De vraag die voor deze variabele gebruikt is, is VCF0101 Respondent Age de leeftijd van de respondent in jaren op het moment van het eerste interview.

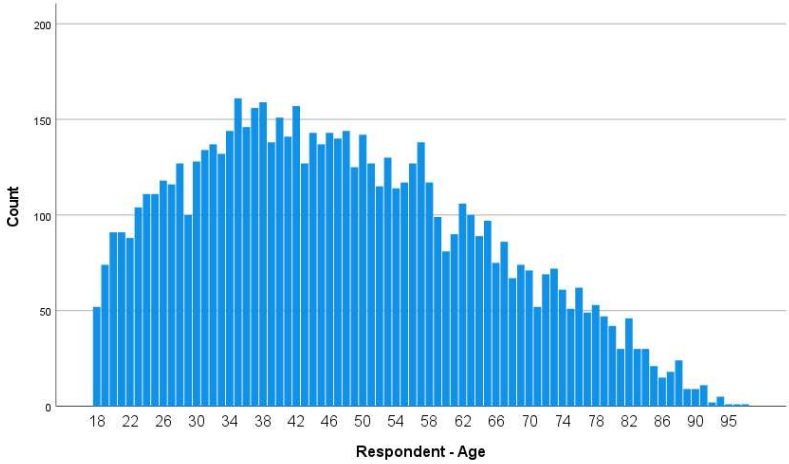
1) Hoe ziet de oorspronkelijke variabele eruit?

```
DESCRIPTIVES VARIABLES=VCF0101
  /STATISTICS=MEAN STDDEV MIN MAX KURTOSIS SKEWNESS.
```

Descriptive Statistics

	N Statistic	Minimum Statistic	Maximum Statistic	Mean Statistic	Std. Deviation Statistic	Skewness Statistic	Skewness Std. Error	Kurtosis Statistic	Kurtosis Std. Error
Respondent - Age	6999	18	97	47,35	17,237	,361	,029	-,674	,059
Valid N (listwise)	6999								

```
GRAPH
  /BAR(SIMPLE)=COUNT BY VCF0101.
```



Toelichting: meer jongere dan oudere respondenten maar als het goed is lijkt dat op de populatieverdeling.

3) Hoe ziet de uiteindelijke variabele eruit?

```
DESCRIPTIVES VARIABLES=VCF0101
  /STATISTICS=MEAN STDDEV MIN MAX.
```

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Respondent - Age	4538	18	93	47,63	17,185
Valid N (listwise)	4538				

6 Verkiezingsjaar

Jaar van verkiezing:

VCF0004 Year of Study

1) Hoe ziet de oorspronkelijke variabele eruit?

```
FREQUENCIES VARIABLES=VCF0004
  /ORDER=ANALYSIS.
```

Year of Study

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1996	1714	24,3	24,3	24,3
	2000	1807	25,6	25,6	49,9
	2004	1212	17,2	17,2	67,1
	2008	2322	32,9	32,9	100,0
	Total	7055	100,0	100,0	

Toelichting: het aantal respondenten uit 2004 is klein en het aantal respondenten uit 2008 is groot (bijna het dubbele).

2) Welke bewerkingen heb je op de variabele/ items uitgevoerd?

De variabelen hebben wel een duidelijke volgorde maar ze hebben geen gelijke afstand tussen elkaar. Daarom zijn er dummies gemaakt om de variabele in de regressie te kunnen gebruiken.

De variabele is hergecodeerd van de jaartallen naar twee dummies voor het jaar 2004 en 2008 en het jaar 1996 is het referentiejaar.

```
RECODE VCF0004 (2004=1) (1996=0) (2008=0) INTO Verkiezing_2004.
VARIABLE LABELS  Verkiezing_2004 'Verkiezingsjaar 2004'.
EXECUTE.
RECODE VCF0004 (1996=0) (2004=0) (2008=1) INTO Verkiezing_2008.
VARIABLE LABELS  Verkiezing_2008 'Verkiezingsjaar 2008'.
EXECUTE.
```

3) Hoe ziet de uiteindelijke variabelen eruit?

```
FREQUENCIES VARIABLES=Verkiezing_2004 Verkiezing_2008
  /ORDER=ANALYSIS.
```

Verkiezingsjaar 2004

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00	3507	77,3	77,3	77,3
	1,00	1031	22,7	22,7	100,0
	Total	4538	100,0	100,0	

Verkiezingsjaar 2008

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00	2537	55,9	55,9	55,9
	1,00	2001	44,1	44,1	100,0
	Total	4538	100,0	100,0	

Year of Study

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1996	1506	33,2	33,2	33,2
	2004	1031	22,7	22,7	55,9
	2008	2001	44,1	44,1	100,0
	Total	4538	100,0	100,0	

7 Etniciteit

De vraag die voor deze variabele gebruikt is, is VCF0105b. Dit is een samengestelde variabele van meerdere vragen uit voorgaande jaren. In 1996 was de etniciteit van een respondent gebaseerd op een observatie van de interviewer en de vraag of de respondent zichzelf als Hispanic ziet. In 2004 en 2008 werd de respond gevraagd tot welke etniciteit zij behoren.

Dit geeft de volgende antwoordmogelijkheden:
1. White non-Hispanic (wit niet hispanic) 2. Black non-Hispanic (Zwart niet hispanic) 3. Hispanic 4. Other or multiple races, non-Hispanic (anders, niet hispanic) en 9. Missing

1) Hoe ziet de oorspronkelijke variabele eruit?

```
FREQUENCIES VARIABLES=VCF0105b
  /ORDER=ANALYSIS.
```

Race-ethnicity summary, 4 categories

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1. White non-Hispanic	4659	66,0	66,7	66,7
	2. Black non-Hispanic	1125	15,9	16,1	82,9
	3. Hispanic	924	13,1	13,2	96,1
	4. Other or multiple races, non-Hispanic	273	3,9	3,9	100,0
	Total	6981	99,0	100,0	
Missing	9. Missing, DK/REF/NA	74	1,0		
Total		7055	100,0		

Er is een overrepresentatie van witte mensen in deze dataset. Dit kan komen omdat pas in 2008 Spaanse vragenlijsten afgenomen zijn. Ook zijn is toen pas moeite gedaan om bewust niet-witte in de dataset mee te nemen.

2) Welke bewerkingen heb je op de variabele/ items uitgevoerd?

Voor de regressie analyse moet deze nominale variabele een dummy worden. Als referentiegroep is gekozen voor White non-Hispanic.

```
RECODE VCF0105b (1=0) (2=1) (3=0) (4=0) (9=SYSMIS) INTO Zwart_niet_Hispanic.
VARIABLE LABELS  Zwart_niet_Hispanic 'Zwart niet-Hispanic'.
EXECUTE.
RECODE VCF0105b (1=0) (4=0) (9=SYSMIS) (2=0) (3=1) INTO Hispanic.
VARIABLE LABELS  Hispanic 'Hispanic'.
EXECUTE.
FREQUENCIES VARIABLES=Zwart_niet_Hispanic Hispanic Other_multiple
  /ORDER=ANALYSIS.
```

Zwart niet-Hispanic

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Niet Zwart niet Hispanic	5856	83,0	83,9	83,9
	Zwart niet Hispanic	1125	15,9	16,1	100,0
	Total	6981	99,0	100,0	
Missing	System	74	1,0		
Total		7055	100,0		

Hispanic

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	niet-Hispanic	6057	85,9	86,8	86,8
	Hispanic	924	13,1	13,2	100,0
	Total	6981	99,0	100,0	
Missing	System	74	1,0		
Total		7055	100,0		

Anders, meerdere

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	niet Anders of meerdere	6708	95,1	96,1	96,1
	Anders of meerdere	273	3,9	3,9	100,0
	Total	6981	99,0	100,0	
Missing	System	74	1,0		
Total		7055	100,0		

Toelichting: de verhoudingen kloppen vergeleken met de originele variabele dus de bewerking klopt.

3) Hoe ziet de uiteindelijke variabele eruit?
FREQUENCIES VARIABLES=Gen_Dum, Pol_Ken, VCF0101, VCF0105b, Goed_geschat, VCF0110, VCF0128, VCF0112, VCF0303, Voorkeur

Race-ethnicity summary, 4 categories

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1. White non-Hispanic	2944	64,9	64,9	64,9
	2. Black non-Hispanic	785	17,3	17,3	82,2
	3. Hispanic	660	14,5	14,5	96,7
	4. Other or multiple races, non-Hispanic	149	3,3	3,3	100,0
	Total	4538	100,0	100,0	

8 Gebied van herkomst

De vraag die voor deze variabele gebruikt is, is VCF0112 Census Region. De respondent werd gevraagd naar de staat waarin die nu woont. Deze antwoorden zijn samengebracht in een schaal waar de staten die tot een bepaalde census regio behoren samengevoegd zijn.

1. Northeast (CT, ME, MA, NH, NJ, NY, PA, RI, VT)
2. North Central (IL, IN, IA, KS, MI, MN, MO, NE, ND, OH, SD, WI)
3. South (AL, AR, DE, D.C., FL, GA, KY, LA, MD, MS, NC, OK, SC,TN, TX, VA, WV)
4. West (AK, AZ, CA, CO, HI, ID, MT, NV, NM, OR, UT, WA, WY)

0 (Missing)

1) Hoe ziet de oorspronkelijke variabele eruit?

```
FREQUENCIES VARIABLES=VCF0112
  /ORDER=ANALYSIS.
```

Census Region

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1. Northeast (CT, ME, MA, NH, NJ, NY, PA, RI, VT)	1048	14,9	14,9	14,9
	2. North Central (IL, IN, IA, KS, MI, MN, MO, NE, ND,	1619	22,9	22,9	37,8
	3. South (AL, AR, DE, D.C., FL, GA, KY, LA, MD, MS, NC	2815	39,9	39,9	77,7
	4. West (AK, AZ, CA, CO, HI, ID, MT, NV, NM, OR, UT, WA,	1573	22,3	22,3	100,0
	Total	7055	100,0	100,0	

Een groot deel van de respondenten komt uit het Zuiden. Ook zijn er geen missende waarden voor deze variabele.

2) Welke bewerkingen heb je op de variabele/ items uitgevoerd?

Voor de regressie analyse moet deze nominale variabele een dummy worden. Als referentiegroep is gekozen voor het Zuiden.

```
RECODE VCF0112 (1=1) (2=0) (3=0) (4=0) INTO Noordoost.
VARIABLE LABELS  Noordoost 'Noordoost'.
EXECUTE.
RECODE VCF0112 (3=0) (4=0) (1=0) (2=1) INTO Noord_centr.
VARIABLE LABELS  Noord_centr 'Noord Centraal'.
EXECUTE.
RECODE VCF0112 (3=0) (1=0) (2=0) (4=1) INTO West.
VARIABLE LABELS  West 'West'.
EXECUTE.
FREQUENCIES VARIABLES= Noordoost Noord_centr West
  /ORDER=ANALYSIS.
```

Noordoost

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	niet-Noordoost	6007	85,1	85,1	85,1
	Noordoost	1048	14,9	14,9	100,0
	Total	7055	100,0	100,0	

Noord Centraal

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	niet-Noord Centraal	5436	77,1	77,1	77,1
	Noord Centraal	1619	22,9	22,9	100,0
	Total	7055	100,0	100,0	

West

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	niet-West	5482	77,7	77,7	77,7
	West	1573	22,3	22,3	100,0
	Total	7055	100,0	100,0	

De verhoudingen zijn hetzelfde dus het coderen van de variabelen is gelukt.
3) Hoe ziet de uiteindelijke variabele eruit?
FREQUENCIES VARIABLES=Gen_Dum, Pol_Ken, VCF0101, VCF0105b, Goed_geschat, VCF0110, VCF0128, VCF0112, VCF0303, Voorkeur

Census Region

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1. Northeast (CT, ME, MA, NH, NJ, NY, PA, RI, VT)	618	13,6	13,6	13,6
	2. North Central (IL, IN, IA, KS, MI, MN, MO, NE, ND,	1027	22,6	22,6	36,2
	3. South (AL, AR, DE, D.C., FL, GA, KY, LA, MD, MS, NC	1865	41,1	41,1	77,3
	4. West (AK, AZ, CA, CO, HI, ID, MT, NV, NM, OR, UT, WA,	1028	22,7	22,7	100,0
	Total	4538	100,0	100,0	

9 Voorkeurskandidaat

De vraag die voor deze variabele gebruikt is, is VCF0303 Party Identification of Respondent-Summary 3-Category. Uit meerdere vragen over de politieke voorkeuren van de respondenten is een variabele gemaakt.

Die bestaat uit de volgende antwoordmogelijkheden:
1. Democrats (including leaners)
2. Independents
3. Republicans (including leaners)
0 (missing)
1) Hoe ziet de oorspronkelijke variabele eruit?
FREQUENCIES VARIABLES=VCF0303
/ORDER=ANALYSIS.

Party Identification of Respondent- Summary 3-Category

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1. Democrats (including leaners)	3739	53,0	53,6	53,6
	2. Independents	766	10,9	11,0	64,6
	3. Republicans (including leaners)	2470	35,0	35,4	100,0
	Total	6975	98,9	100,0	
Missing	0. DK; NA; other; refused to answer; no Pre IW	80	1,1		
Total		7055	100,0		

2) Welke bewerkingen heb je op de variabele/ items uitgevoerd?
Ik heb een nieuwe variabele berekend waarbij iedereen wiens voorkeurspartij gewonnen heeft een 1 score krijgt en iedereen wiens voorkeurspartij verloren heeft een 0.

```
RECODE VCF0303 (1=0) (2=0) (3=1) INTO Rep_Voorkeur.
VARIABLE LABELS  Rep_Voorkeur 'Voorkeur Republikein'.
EXECUTE.
```

Voorkeur Democraat

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00	2066	45,5	45,5	45,5
	1,00	2472	54,5	54,5	100,0
	Total	4538	100,0	100,0	

```
RECODE VCF0303 (1=1) (2=0) (3=0) INTO Dem_Voorkeur.
VARIABLE LABELS  Dem_Voorkeur 'Voorkeur Democraat'.
EXECUTE.
```

Voorkeur Republikein

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00	2928	64,5	64,5	64,5
	1,00	1610	35,5	35,5	100,0
	Total	4538	100,0	100,0	

3) Hoe ziet de uiteindelijke variabele eruit?

```
FREQUENCIES VARIABLES=Gen_Dum, Pol_Ken, VCF0101, VCF0105b, Goed_geschat, VCF0110, VCF0128, VCF0112,
VCF0303, Voorkeur
```

Party Identification of Respondent- Summary 3-Category

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1. Democrats (including leaners)	2472	54,5	54,5	54,5
	2. Independents	456	10,0	10,0	64,5
	3. Republicans (including leaners)	1610	35,5	35,5	100,0
	Total	4538	100,0	100,0	

10 Religie
De vraag die voor deze variabele gebruikt is, is VCF0128 Respondent - Religion Major Group. Respondenten is gevraagd of zij erediensten bijwonen. Zo ja welke protestants, katholiek, joods of een andere soort? Zo niet dan werd gevraagd of zij zichzelf beschouwen als protestants, katholiek, joods of een andere soort religieus.
1. Protestant

2. Catholic [Roman Catholic]
3. Jewish
4. Other and none (also includes DK preference)
0. NA (missing)
1) Hoe ziet de oorspronkelijke variabele eruit?

```
FREQUENCIES VARIABLES=VCF0128
  /ORDER=ANALYSIS.
```

Respondent - Religion Major Group

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1. Protestant	3663	51,9	52,1	52,1
	2. Catholic [Roman Catholic]	1703	24,1	24,2	76,4
	3. Jewish	133	1,9	1,9	78,3
	4. Other and none (also includes DK preference)	1527	21,6	21,7	100,0
	Total	7026	99,6	100,0	
Missing	0. DK; NA; refused to answer; no Pre IW; no Post IW;	29	,4		
Total		7055	100,0		

Heel veel Protestanten en deze variabele laat niet veel ruimte voor variatie binnen religies.
2) Welke bewerkingen heb je op de variabele/ items uitgevoerd?
Voor de regressie analyse moet deze nominale variabele een dummy worden. Als referentiegroep is gekozen voor Protestanten.

```
RECODE VCF0128 (3=0) (1=0) (2=0) (4=1) INTO Anders_geen.
VARIABLE LABELS  Anders_geen 'Anders en geen'.
EXECUTE.
RECODE VCF0128 (3=0) (1=0) (4=0) (2=1) INTO Katholiek.
VARIABLE LABELS  Katholiek 'Katholiek'.
EXECUTE.
RECODE VCF0128 (1=0) (4=0) (2=0) (3=1) INTO Joods.
VARIABLE LABELS  Joods 'Joods'.
EXECUTE.
```

Anders en geen

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	niet Anders Geen	5499	77,9	78,3	78,3
	Anders en Geen	1527	21,6	21,7	100,0
	Total	7026	99,6	100,0	
Missing	System	29	,4		
Total		7055	100,0		

Katholiek

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	niet-Katholiek	5323	75,5	75,8	75,8
	Katholiek	1703	24,1	24,2	100,0
	Total	7026	99,6	100,0	
Missing	System	29	,4		
Total		7055	100,0		

Joods

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	niet-Joods	6893	97,7	98,1	98,1
	Joods	133	1,9	1,9	100,0
	Total	7026	99,6	100,0	
Missing	System	29	,4		
Total		7055	100,0		

3) Hoe ziet de uiteindelijke variabele eruit?

FREQUENCIES VARIABLES=Gen_Dum, Pol_Ken, VCF0101, VCF0105b, Goed_geschat, VCF0110, VCF0128, VCF0112, VCF0303, Voorkeur

Respondent - Religion Major Group

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1. Protestant	2375	52,3	52,3	52,3
	2. Catholic [Roman Catholic]	1094	24,1	24,1	76,4
	3. Jewish	73	1,6	1,6	78,1
	4. Other and none (also includes DK preference)	996	21,9	21,9	100,0
	Total	4538	100,0	100,0	

1 Bewerking van de gehele dataset

Voor deze dataset zijn er best veel missende waarden vooral voor politieke kennis. Daarom is besloten om de missende waarden uit de dataset te halen.

```
COMPUTE m1=0.
IF MISSING(Gen_Dum) m1=1.

COMPUTE m2=0.
IF MISSING(Pol_Ken) m2=1.

COMPUTE m3=0.
IF MISSING(VCF0101) m3=1.

COMPUTE m4=0.
IF MISSING(VCF0105b) m4=1.

COMPUTE m5=0.
IF MISSING(Goed_geschat) m5=1.

COMPUTE m6=0.
IF MISSING(VCF0110) m6=1.

COMPUTE m7=0.
IF MISSING(VCF0128) m7=1.

COMPUTE m8=0.
IF MISSING(VCF0112) m8=1.

COMPUTE m9=0.
IF MISSING(VCF0303) m9=1.

COMPUTE m10=0.
IF MISSING(Voorkeur) m10=1

COMPUTE miss = m1+m2+m3+m4+m5+m6+m7+m8+m9+m10.
EXECUTE.
```

```
FREQUENCIES VARIABLES=miss
  /ORDER=ANALYSIS.

DATASET COPY  Dataset3.
DATASET ACTIVATE  Dataset3.
FILTER OFF.
USE ALL.
SELECT IF (miss<9).
EXECUTE.
DATASET ACTIVATE  Dataset3.

FREQUENCIES VARIABLES=miss
  /ORDER=ANALYSIS.

DELETE VARIABLES m1 m2 m3 m4 m5 m6 m7 m8 m9 m10 miss.

*Analyze Patterns of Missing Values.
MULTIPLE IMPUTATION  Gen_Dum Pol_Ken VCF0101 VCF0105b Goed_geschat VCF0110 VCF0128 VCF0112 VCF0303
Voorkeur
   /IMPUTE METHOD=NONE
   /MISSINGSUMMARIES  OVERALL VARIABLES (MAXVARS=25 MINPCTMISSING=10) PATTERNS.

MVA VARIABLES=Gen_Dum Pol_Ken VCF0101 VCF0105b Goed_geschat VCF0110 VCF0128 VCF0112 VCF0303 Voorkeur
  /MAXCAT=25
  /CATEGORICAL=Gen_Dum Pol_Ken VCF0105b Goed_geschat VCF0110 VCF0128 VCF0112 VCF0303 Voorkeur
  /TTEST NOPROB PERCENT=5
  /TPATTERN PERCENT=1.

DATASET COPY  CompleteDataset.
DATASET ACTIVATE  CompleteDataset.
FILTER OFF.
USE ALL.
SELECT IF (NMISS(Gen_Dum, Pol_Ken, VCF0101, VCF0105b, Goed_geschat, VCF0110, VCF0128, VCF0112, VCF0303,
Voorkeur)  < 1). EXECUTE.
```

Bijlage 2

2.1 Bivariate statistieken

Syntax en output voor alle correlaties. Voor de correlaties zijn de originele variabelen uit de dataset van ANES gebruikt.

2.1.1 Correlatie gender juist voorspelde verkiezingsuitslag

```
CROSSTABS
  /TABLES=Goed_geschat BY Gen_Dum
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Goed_geschat * Gen_Dum Crosstabulation

Count

		Gen_Dum		
		Man	Vrouw	Total
Goed_geschat	Fout	552	695	1247
	Goed	1489	1802	3291
Total		2041	2497	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	-,009	,554
	Cramer's V	,009	,554
N of Valid Cases		4538	

2.1.2 Correlatie politieke kennis en juist voorspelde verkiezingsuitslag

```
CROSSTABS
  /TABLES=Goed_geschat BY Pol_Ken
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Goed_geschat * Pol_Ken Crosstabulation

Count

		Pol_Ken				
		Geen vragen goed	1 vraag goed	2 vragen goed	Alle vragen goed	Total
Goed_geschat	Fout	262	464	394	127	1247
	Goed	607	1189	1176	319	3291
Total		869	1653	1570	446	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,042	,044
	Cramer's V	,042	,044
N of Valid Cases		4538	

2.1.3 Correlatie opleidingsniveau en juist voorspelde verkiezingsuitslag

```
CROSSTABS
  /TABLES=Goed_geschat BY VCF0110
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Goed_geschat * Respondent - Education 4-category Crosstabulation

Count

		Respondent - Education 4-category				
		1. Grade school or less (0-8 grades); 2020: less than high school credential	2. High school (12 grades or fewer, incl. non-college	3. Some college (13 grades or more but no degree;	4. College or advanced degree (no cases 1948)	Total
Goed_geschat	Fout	60	467	409	311	1247
	Goed	108	1341	968	874	3291
Total		168	1808	1377	1185	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,052	,006
	Cramer's V	,052	,006
N of Valid Cases		4538	

2.1.4 Correlatie verkiezingsjaar en juist voorspelde verkiezingsuitslag

```
CROSSTABS
 /TABLES=Goed_geschat BY VCF0004
 /FORMAT=AVALUE TABLES
 /STATISTICS=CHISQ PHI
 /CELLS=COUNT
 /COUNT ROUND CELL.
```

Goed_geschat * Year of Study Crosstabulation

Count

		Year of Study			
		1996	2004	2008	Total
Goed_geschat	Fout	189	361	697	1247
	Goed	1317	670	1304	3291
Total		1506	1031	2001	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,236	<,001
	Cramer's V	,236	<,001
N of Valid Cases		4538	

2.1.5 Correlatie etniciteit en juist voorspelde verkiezingsuitslag

```
CROSSTABS
 /TABLES=Goed_geschat BY VCF0105b
 /FORMAT=AVALUE TABLES
 /STATISTICS=CHISQ PHI
 /CELLS=COUNT
 /COUNT ROUND CELL.
```

Goed_geschat * Race-ethnicity summary, 4 categories Crosstabulation

Count

		Race-ethnicity summary, 4 categories				
		1. White non-Hispanic	2. Black non-Hispanic	3. Hispanic	4. Other or multiple races, non-Hispanic	Total
Goed_geschat	Fout	821	171	202	53	1247
	Goed	2123	614	458	96	3291
Total		2944	785	660	149	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,068	<,001
	Cramer's V	,068	<,001
N of Valid Cases		4538	

2.1.6 Correlatie herkomstgebied en juist voorspelde verkiezingsuitslag

```
CROSSTABS
 /TABLES=Goed_geschat BY VCF0112
 /FORMAT=AVALUE TABLES
 /STATISTICS=CHISQ PHI
 /CELLS=COUNT
 /COUNT ROUND CELL.
```

Goed_geschat * Census Region Crosstabulation

Count

		Census Region				
		1. Northeast (CT, ME, MA, NH, NJ, NY, PA, RI, VT)	2. North Central (IL, IN, IA, KS, MI, MN, MO, NE, ND,	3. South (AL, AR, DE, D.C., FL, GA, KY, LA, MD, MS, NC	4. West (AK, AZ, CA, CO, HI, ID, MT, NV, NM, OR, UT, WA,	Total
Goed_geschat	Fout	157	280	509	301	1247
	Goed	461	747	1356	727	3291
Total		618	1027	1865	1028	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,026	,382
	Cramer's V	,026	,382
N of Valid Cases		4538	

2.1.7 Correlatie politieke kennis en juist voorspelde verkiezingsuitslag

```
CROSSTABS
 /TABLES=Goed_geschat BY VCF0303
 /FORMAT=AVALUE TABLES
 /STATISTICS=CHISQ PHI
 /CELLS=COUNT
 /COUNT ROUND CELL.
```

Goed_geschat * Party Identification of Respondent- Summary 3-Category Crosstabulation

Count

		Party Identification of Respondent- Summary 3-Category			
		1. Democrats (including leaners)	2. Independents	3. Republicans (including leaners)	Total
Goed_geschat	Fout	548	155	544	1247
	Goed	1924	301	1066	3291
Total		2472	456	1610	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,130	<,001
	Cramer's V	,130	<,001
N of Valid Cases		4538	

2.1.8 Correlatie religie en juist voorspelde verkiezingsuitslag

```
CROSSTABS
 /TABLES=Goed_geschat BY VCF0128
 /FORMAT=AVALUE TABLES
 /STATISTICS=CHISQ PHI
 /CELLS=COUNT
 /COUNT ROUND CELL.
```

Goed_geschat * Respondent - Religion Major Group Crosstabulation

Count

		Respondent - Religion Major Group				
		1. Protestant	2. Catholic [Roman Catholic]	3. Jewish	4. Other and none (also includes DK preference)	Total
Goed_geschat	Fout	655	300	14	278	1247
	Goed	1720	794	59	718	3291
Total		2375	1094	73	996	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,024	,452
	Cramer's V	,024	,452
N of Valid Cases		4538	

2.2.1 Correlatie gender en politieke kennis

```
CROSSTABS
 /TABLES=Gen_Dum BY Pol_Ken
 /FORMAT=AVALUE TABLES
 /STATISTICS=CHISQ PHI
 /CELLS=COUNT
 /COUNT ROUND CELL.
```

Gen_Dum * Pol_Ken Crosstabulation

Count

		Pol_Ken				
		Geen vragen goed	1 vraag goed	2 vragen goed	Alle vragen goed	Total
Gen_Dum	Man	304	686	763	288	2041
	Vrouw	565	967	807	158	2497
Total		869	1653	1570	446	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,163	<,001
	Cramer's V	,163	<,001
N of Valid Cases		4538	

2.1.9 Correlatie voorkeurskandidaat en juist voorspelde verkiezingsuitslag

```
CROSSTABS
  /TABLES=Goed_geschat BY VCF0303
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Goed_geschat * Party Identification of Respondent- Summary 3-Category Crosstabulation

Count

		Party Identification of Respondent- Summary 3-Category			
		1. Democrats (including leaners)	2. Independents	3. Republicans (including leaners)	Total
Goed_geschat	Fout	548	155	544	1247
	Goed	1924	301	1066	3291
Total		2472	456	1610	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,130	<,001
	Cramer's V	,130	<,001
N of Valid Cases		4538	

2.2.3 Correlatie gender en opleidingsniveau

```
CROSSTABS
  /TABLES=Gen_Dum BY VCF0110
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Gen_Dum * Respondent - Education 4-category Crosstabulation

Count

		Respondent - Education 4-category				
		1. Grade school or less (0-8 grades); 2020: less than high school credential	2. High school (12 grades or fewer, incl. non-college	3. Some college (13 grades or more but no degree;	4. College or advanced degree (no cases 1948)	Total
Gen_Dum	Man	71	805	575	590	2041
	Vrouw	97	1003	802	595	2497
Total		168	1808	1377	1185	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,062	<,001
	Cramer's V	,062	<,001
N of Valid Cases		4538	

2.2.4 Correlatie gender en verkiezingsjaar

```
CROSSTABS
  /TABLES=Gen_Dum BY VCF0004
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Gen_Dum * Year of Study Crosstabulation

Count

		Year of Study			
		1996	2004	2008	Total
Gen_Dum	Man	679	490	872	2041
	Vrouw	827	541	1129	2497
Total		1506	1031	2001	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,031	,117
	Cramer's V	,031	,117
N of Valid Cases		4538	

2.2.5 Correlatie gender en herkomstgebied

```
CROSSTABS
  /TABLES=Gen_Dum BY VCF0112
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Gen_Dum * Census Region Crosstabulation

Count

		Census Region				
		1. Northeast (CT, ME, MA, NH, NJ, NY, PA, RI, VT)	2. North Central (IL, IN, IA, KS, MI, MN, MO, NE, ND,	3. South (AL, AR, DE, D.C., FL, GA, KY, LA, MD, MS, NC	4. West (AK, AZ, CA, CO, HI, ID, MT, NV, NM, OR, UT, WA,	Total
Gen_Dum	Man	292	471	802	476	2041
	Vrouw	326	556	1063	552	2497
Total		618	1027	1865	1028	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,034	,152
	Cramer's V	,034	,152
N of Valid Cases		4538	

2.2.5 Correlatie gender en etniciteit

```
CROSSTABS
  /TABLES=Gen_Dum BY VCF0105b
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,041	,053
	Cramer's V	,041	,053
N of Valid Cases		4538	

Gen_Dum * Race-ethnicity summary, 4 categories Crosstabulation

Count

		Race-ethnicity summary, 4 categories				
		1. White non-Hispanic	2. Black non-Hispanic	3. Hispanic	4. Other or multiple races, non-Hispanic	Total
Gen_Dum	Man	1349	326	289	77	2041
	Vrouw	1595	459	371	72	2497
Total		2944	785	660	149	4538

2.2.6 Correlatie gender en voorkeurskandidaat

```
CROSSTABS
  /TABLES=Gen_Dum BY VCF0303
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Gen_Dum * Party Identification of Respondent- Summary 3-Category Crosstabulation

Count

		Party Identification of Respondent- Summary 3-Category			
		1. Democrats (including leaners)	2. Independents	3. Republicans (including leaners)	Total
Gen_Dum	,00	1003	215	823	2041
	1,00	1469	241	787	2497
Total		2472	456	1610	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,099	<,001
	Cramer's V	,099	<,001
N of Valid Cases		4538	

2.2.7 Correlatie gender en religie

```
CROSSTABS
 /TABLES=Pol_Ken BY VCF0128
 /FORMAT=AVALUE TABLES
 /STATISTICS=CHISQ PHI
 /CELLS=COUNT
 /COUNT ROUND CELL.
```

Gen_Dum * Respondent - Religion Major Group Crosstabulation

Count

		Respondent - Religion Major Group				
		1. Protestant	2. Catholic [Roman Catholic]	3. Jewish	4. Other and none (also includes DK preference)	Total
Gen_Dum	,00	984	500	31	526	2041
	1,00	1391	594	42	470	2497
Total		2375	1094	73	996	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,091	<,001
	Cramer's V	,091	<,001
N of Valid Cases		4538	

2.3.1 Correlatie politieke kennis en religie

```
CROSSTABS
 /TABLES=Pol_Ken BY VCF0128
 /FORMAT=AVALUE TABLES
 /STATISTICS=CHISQ PHI
 /CELLS=COUNT
 /COUNT ROUND CELL.
```

Pol_Ken * Respondent - Religion Major Group Crosstabulation

Count

		Respondent - Religion Major Group				
		1. Protestant	2. Catholic [Roman Catholic]	3. Jewish	4. Other and none (also includes DK preference)	Total
Pol_Ken	,00	441	220	3	205	869
	1,00	888	366	17	382	1653
	2,00	838	387	38	307	1570
	3,00	208	121	15	102	446
Total		2375	1094	73	996	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,097	<,001
	Cramer's V	,056	<,001
N of Valid Cases		4538	

2.3.2 Correlatie politieke kennis en voorkeurskandidaat

```
CROSSTABS
 /TABLES=Pol_Ken BY VCF0303
 /FORMAT=AVALUE TABLES
 /STATISTICS=CHISQ PHI
 /CELLS=COUNT
 /COUNT ROUND CELL.
```

Pol_Ken * Party Identification of Respondent- Summary 3-Category Crosstabulation

Count

		Party Identification of Respondent- Summary 3-Category			
		1. Democrats (including leaners)	2. Independents	3. Republicans (including leaners)	Total
Pol_Ken	,00	525	178	166	869
	1,00	926	146	581	1653
	2,00	809	118	643	1570
	3,00	212	14	220	446
Total		2472	456	1610	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,232	<,001
	Cramer's V	,164	<,001
N of Valid Cases		4538	

2.3.3 Correlatie politieke kennis en religie

```
CROSSTABS
 /TABLES=Pol_Ken BY VCF0105b
 /FORMAT=AVALUE TABLES
 /STATISTICS=CHISQ PHI
 /CELLS=COUNT
 /COUNT ROUND CELL.
```

Pol_Ken * Race-ethnicity summary, 4 categories Crosstabulation

Count

		Race-ethnicity summary, 4 categories				
		1. White non-Hispanic	2. Black non-Hispanic	3. Hispanic	4. Other or multiple races, non-Hispanic	Total
Pol_Ken	,00	320	287	236	26	869
	1,00	1071	298	226	58	1653
	2,00	1184	174	160	52	1570
	3,00	369	26	38	13	446
Total		2944	785	660	149	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,324	<,001
	Cramer's V	,187	<,001
N of Valid Cases		4538	

2.3.4 Correlatie politieke kennis en herkomstgebied

```
CROSSTABS
 /TABLES=Pol_Ken BY VCF0112
 /FORMAT=AVALUE TABLES
 /STATISTICS=CHISQ PHI
 /CELLS=COUNT
 /COUNT ROUND CELL.
```

Pol_Ken * Census Region Crosstabulation

Count

		Census Region				
		1. Northeast (CT, ME, MA, NH, NJ, NY, PA, RI, VT)	2. North Central (IL, IN, IA, KS, MI, MN, MO, NE, ND,	3. South (AL, AR, DE, D.C., FL, GA, KY, LA, MD, MS, NC	4. West (AK, AZ, CA, CO, HI, ID, MT, NV, NM, OR, UT, WA,	Total
Pol_Ken	,00	88	156	449	176	869
	1,00	220	419	648	366	1653
	2,00	229	354	618	369	1570
	3,00	81	98	150	117	446
Total		618	1027	1865	1028	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,122	<,001
	Cramer's V	,070	<,001
N of Valid Cases		4538	

2.3.3 Correlatie politieke kennis en verkiezingsjaar

```
CROSSTABS
 /TABLES=Pol_Ken BY VCF0004
 /FORMAT=AVALUE TABLES
 /STATISTICS=CHISQ PHI
 /CELLS=COUNT
 /COUNT ROUND CELL.
```

Pol_Ken * Year of Study Crosstabulation

Count

		Year of Study			
		1996	2004	2008	Total
Pol_Ken	,00	174	136	559	869
	1,00	451	552	650	1653
	2,00	752	243	575	1570
	3,00	129	100	217	446
Total		1506	1031	2001	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,303	<,001
	Cramer's V	,214	<,001
N of Valid Cases		4538	

2.3.3 Correlatie politieke kennis en opleidingsniveau

```
CROSSTABS
  /TABLES=Pol_Ken BY VCF0110
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Pol_Ken * Respondent - Education 4-category Crosstabulation

Count

		Respondent - Education 4-category				
		1. Grade school or less (0-8 grades); 2020: less than high school credential	2. High school (12 grades or fewer, incl. non-college	3. Some college (13 grades or more but no degree;	4. College or advanced degree (no cases 1948)	Total
Pol_Ken	,00	87	518	203	61	869
	1,00	56	736	554	307	1653
	2,00	24	481	501	564	1570
	3,00	1	73	119	253	446
Total		168	1808	1377	1185	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,397	<,001
	Cramer's V	,229	<,001
N of Valid Cases		4538	

2.4.1 Correlatie opleidingsniveau en religie

```
CROSSTABS
  /TABLES=VCF0110 BY VCF0128
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Respondent - Education 4-category * Respondent - Religion Major Group Crosstabulation

Count

		Respondent - Religion Major Group				
		1. Protestant	2. Catholic [Roman Catholic]	3. Jewish	4. Other and none (also includes DK preference)	Total
Respondent - Education 4-category	1. Grade school or less (0-8 grades); 2020: less than high school credential	91	52	0	25	168
	2. High school (12 grades or fewer, incl. non-college	972	425	10	401	1808
	3. Some college (13 grades or more but no degree;	762	307	12	296	1377
	4. College or advanced degree (no cases 1948)	550	310	51	274	1185
Total		2375	1094	73	996	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,146	<,001
	Cramer's V	,084	<,001
N of Valid Cases		4538	

2.4.2 Correlatie opleidingsniveau en voorkeurskandidaat

```
CROSSTABS
 /TABLES=VCF0110 BY VCF0303
 /FORMAT=AVALUE TABLES
 /STATISTICS=CHISQ PHI
 /CELLS=COUNT
 /COUNT ROUND CELL.
```

Respondent - Education 4-category * Party Identification of Respondent- Summary 3-Category Crosstabulation

Count

		Party Identification of Respondent- Summary 3-Category			
		1. Democrats (including leaners)	2. Independents	3. Republicans (including leaners)	Total
Respondent - Education 4-category	1. Grade school or less (0-8 grades); 2020: less than high school credential	113	26	29	168
	2. High school (12 grades or fewer, incl. non-college	1052	225	531	1808
	3. Some college (13 grades or more but no degree;	720	135	522	1377
	4. College or advanced degree (no cases 1948)	587	70	528	1185
Total		2472	456	1610	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,160	<,001
	Cramer's V	,113	<,001
N of Valid Cases		4538	

2.4.3 Correlatie opleidingsniveau en etniciteit

```
CROSSTABS
 /TABLES=VCF0110 BY VCF0105b
 /FORMAT=AVALUE TABLES
 /STATISTICS=CHISQ PHI
 /CELLS=COUNT
```

/COUNT ROUND CELL.

Respondent - Education 4-category * Race-ethnicity summary, 4 categories Crosstabulation

Count

		Race-ethnicity summary, 4 categories				
		1. White non-Hispanic	2. Black non-Hispanic	3. Hispanic	4. Other or multiple races, non-Hispanic	Total
Respondent - Education 4-category	1. Grade school or less (0-8 grades); 2020: less than high school credential	64	37	64	3	168
	2. High school (12 grades or fewer, incl. non-college	1083	400	289	36	1808
	3. Some college (13 grades or more but no degree;	879	241	199	58	1377
	4. College or advanced degree (no cases 1948)	918	107	108	52	1185
Total		2944	785	660	149	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,229	<,001
	Cramer's V	,132	<,001
N of Valid Cases		4538	

2.4.4 Correlatie opleidingsniveau en herkomstgebied

CROSSTABS
/TABLES=VCF0110 BY VCF0112
/FORMAT=AVALUE TABLES
/STATISTICS=CHISQ PHI
/CELLS=COUNT
/COUNT ROUND CELL.

Respondent - Education 4-category * Census Region Crosstabulation

Count

		Census Region				
		1. Northeast (CT, ME, MA, NH, NJ, NY, PA, RI, VT)	2. North Central (IL, IN, IA, KS, MI, MN, MO, NE, ND,	3. South (AL, AR, DE, D.C., FL, GA, KY, LA, MD, MS, NC	4. West (AK, AZ, CA, CO, HI, ID, MT, NV, NM, OR, UT, WA,	Total
Respondent - Education 4-category	1. Grade school or less (0-8 grades); 2020: less than high school credential	22	27	98	21	168
	2. High school (12 grades or fewer, incl. non-college	215	480	770	343	1808
	3. Some college (13 grades or more but no degree;	181	315	538	343	1377
	4. College or advanced degree (no cases 1948)	200	205	459	321	1185
Total		618	1027	1865	1028	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,142	<,001
	Cramer's V	,082	<,001
N of Valid Cases		4538	

2.4.5 Correlatie opleidingsniveau en verkiezingsjaar

CROSSTABS
/TABLES=VCF0110 BY VCF0004
/FORMAT=AVALUE TABLES
/STATISTICS=CHISQ PHI

```
/CELLS=COUNT
/COUNT ROUND CELL.
```

Respondent - Education 4-category * Year of Study Crosstabulation

Count

		Year of Study			
		1996	2004	2008	Total
Respondent - Education 4-category	1. Grade school or less (0-8 grades); 2020: less than high school credential	63	28	77	168
	2. High school (12 grades or fewer, incl. non-college	600	352	856	1808
	3. Some college (13 grades or more but no degree;	413	333	631	1377
	4. College or advanced degree (no cases 1948)	430	318	437	1185
Total		1506	1031	2001	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,104	<,001
	Cramer's V	,074	<,001
N of Valid Cases		4538	

2.5.1 Correlatie verkiezingsjaar en religie

```
CROSSTABS
 /TABLES=VCF0004 BY VCF0128
 /FORMAT=AVALUE TABLES
 /STATISTICS=CHISQ PHI
 /CELLS=COUNT
 /COUNT ROUND CELL.
```

Year of Study * Respondent - Religion Major Group Crosstabulation

Count

		Respondent - Religion Major Group				
		1. Protestant	2. Catholic [Roman Catholic]	3. Jewish	4. Other and none (also includes DK preference)	Total
Year of Study	1996	817	380	28	281	1506
	2004	521	258	26	226	1031
	2008	1037	456	19	489	2001
Total		2375	1094	73	996	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,080	<,001
	Cramer's V	,056	<,001
N of Valid Cases		4538	

2.5.2 Correlatie verkiezingsjaar en voorkeurskanidaat

```
CROSSTABS
 /TABLES=VCF0004 BY VCF0303
 /FORMAT=AVALUE TABLES
 /STATISTICS=CHISQ PHI
 /CELLS=COUNT
 /COUNT ROUND CELL.
```

Year of Study * Party Identification of Respondent- Summary 3-Category Crosstabulation

Count

		Party Identification of Respondent- Summary 3-Category			
		1. Democrats (including leaners)	2. Independents	3. Republicans (including leaners)	Total
Year of Study	1996	792	129	585	1506
	2004	493	99	439	1031
	2008	1187	228	586	2001
Total		2472	456	1610	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,121	<,001
	Cramer's V	,085	<,001
N of Valid Cases		4538	

2.5.3 Correlatie verkiezingsjaar en etniciteit

```
CROSSTABS
 /TABLES=VCF0004 BY VCF0105b
 /FORMAT=AVALUE TABLES
 /STATISTICS=CHISQ PHI
 /CELLS=COUNT
 /COUNT ROUND CELL.
```

Year of Study * Race-ethnicity summary, 4 categories Crosstabulation

Count

		Race-ethnicity summary, 4 categories				
		1. White non-Hispanic	2. Black non-Hispanic	3. Hispanic	4. Other or multiple races, non-Hispanic	Total
Year of Study	1996	1194	155	122	35	1506
	2004	746	152	90	43	1031
	2008	1004	478	448	71	2001
Total		2944	785	660	149	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,289	<,001
	Cramer's V	,204	<,001
N of Valid Cases		4538	

2.5.4 Correlatie verkiezingsjaar en herkomstgebied

```
CROSSTABS
 /TABLES=VCF0004 BY VCF0112
 /FORMAT=AVALUE TABLES
 /STATISTICS=CHISQ PHI
 /CELLS=COUNT
 /COUNT ROUND CELL.
```

Year of Study * Census Region Crosstabulation

Count

		Census Region				
		1. Northeast (CT, ME, MA, NH, NJ, NY, PA, RI, VT)	2. North Central (IL, IN, IA, KS, MI, MN, MO, NE, ND,	3. South (AL, AR, DE, D.C., FL, GA, KY, LA, MD, MS, NC	4. West (AK, AZ, CA, CO, HI, ID, MT, NV, NM, OR, UT, WA,	Total
Year of Study	1996	230	410	560	306	1506
	2004	180	275	352	224	1031
	2008	208	342	953	498	2001
Total		618	1027	1865	1028	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,166	<,001
	Cramer's V	,117	<,001
N of Valid Cases		4538	

2.6.1 Correlatie etniciteit en religie

```
CROSSTABS
 /TABLES=VCF0105b BY VCF0128
 /FORMAT=AVALUE TABLES
 /STATISTICS=CHISQ PHI
 /CELLS=COUNT
 /COUNT ROUND CELL.
```

Race-ethnicity summary, 4 categories * Respondent - Religion Major Group Crosstabulation

Count

		Respondent - Religion Major Group				
		1. Protestant	2. Catholic [Roman Catholic]	3. Jewish	4. Other and none (also includes DK preference)	Total
Race-ethnicity summary, 4 categories	1. White non-Hispanic	1519	681	64	680	2944
	2. Black non-Hispanic	633	31	2	119	785
	3. Hispanic	161	356	4	139	660
	4. Other or multiple races, non-Hispanic	62	26	3	58	149
Total		2375	1094	73	996	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,381	<,001
	Cramer's V	,220	<,001
N of Valid Cases		4538	

2.6.2 Correlatie etniciteit en voorkeurskandidaat

```
CROSSTABS
 /TABLES=VCF0105b BY VCF0303
 /FORMAT=AVALUE TABLES
 /STATISTICS=CHISQ PHI
 /CELLS=COUNT
 /COUNT ROUND CELL.
```

Race-ethnicity summary, 4 categories * Party Identification of Respondent- Summary 3-Category Crosstabulation

Count

		Party Identification of Respondent- Summary 3-Category			
		1. Democrats (including leaners)	2. Independents	3. Republicans (including leaners)	Total
Race-ethnicity summary, 4 categories	1. White non-Hispanic	1301	275	1368	2944
	2. Black non-Hispanic	669	76	40	785
	3. Hispanic	417	85	158	660
	4. Other or multiple races, non-Hispanic	85	20	44	149
Total		2472	456	1610	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,346	<,001
	Cramer's V	,244	<,001
N of Valid Cases		4538	

2.6.3 Correlatie etniciteit en herkomstgebied

```
CROSSTABS
 /TABLES=VCF0105b BY VCF0112
 /FORMAT=AVALUE TABLES
 /STATISTICS=CHISQ PHI
 /CELLS=COUNT
 /COUNT ROUND CELL.
```

Race-ethnicity summary, 4 categories * Census Region Crosstabulation

Count

		Census Region				
		1. Northeast (CT, ME, MA, NH, NJ, NY, PA, RI, VT)	2. North Central (IL, IN, IA, KS, MI, MN, MO, NE, ND,	3. South (AL, AR, DE, D.C., FL, GA, KY, LA, MD, MS, NC	4. West (AK, AZ, CA, CO, HI, ID, MT, NV, NM, OR, UT, WA,	Total
Race-ethnicity summary, 4 categories	1. White non-Hispanic	431	821	1004	688	2944
	2. Black non-Hispanic	95	144	491	55	785
	3. Hispanic	59	36	328	237	660
	4. Other or multiple races, non-Hispanic	33	26	42	48	149
Total		618	1027	1865	1028	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,310	<,001
	Cramer's V	,179	<,001
N of Valid Cases		4538	

2.7.1 Correlatie voorkeurskandidaat en religie

```
CROSSTABS
 /TABLES=VCF0303 BY VCF0128
 /FORMAT=AVALUE TABLES
 /STATISTICS=CHISQ PHI
 /CELLS=COUNT
 /COUNT ROUND CELL.
```

Party Identification of Respondent- Summary 3-Category * Respondent - Religion Major Group Crosstabulation

Count

		Respondent - Religion Major Group				
		1. Protestant	2. Catholic [Roman Catholic]	3. Jewish	4. Other and none (also includes DK preference)	Total
Party Identification of Respondent- Summary 3-Category	1. Democrats (including leaners)	1246	615	59	552	2472
	2. Independents	194	111	2	149	456
	3. Republicans (including leaners)	935	368	12	295	1610
Total		2375	1094	73	996	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,130	<,001
	Cramer's V	,092	<,001
N of Valid Cases		4538	

2.7.2 Correlatie voorkeurskandidaat en herkomstgebied

```
CROSSTABS
 /TABLES=VCF0303 BY VCF0112
 /FORMAT=AVALUE TABLES
 /STATISTICS=CHISQ PHI
 /CELLS=COUNT
 /COUNT ROUND CELL.
```

Party Identification of Respondent- Summary 3-Category * Census Region Crosstabulation

Count

		Census Region				
		1. Northeast (CT, ME, MA, NH, NJ, NY, PA, RI, VT)	2. North Central (IL, IN, IA, KS, MI, MN, MO, NE, ND,	3. South (AL, AR, DE, D.C., FL, GA, KY, LA, MD, MS, NC	4. West (AK, AZ, CA, CO, HI, ID, MT, NV, NM, OR, UT, WA,	Total
Party Identification of Respondent- Summary 3-Category	1. Democrats (including leaners)	341	527	1022	582	2472
	2. Independents	63	118	198	77	456
	3. Republicans (including leaners)	214	382	645	369	1610
Total		618	1027	1865	1028	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,055	,033
	Cramer's V	,039	,033
N of Valid Cases		4538	

2.8.1 Correlatie religie en herkomstgebied

```
CROSSTABS
 /TABLES=VCF0128 BY VCF0112
 /FORMAT=AVALUE TABLES
 /STATISTICS=CHISQ PHI
 /CELLS=COUNT
 /COUNT ROUND CELL.
```

Respondent - Religion Major Group * Census Region Crosstabulation

Count

		Census Region				
		1. Northeast (CT, ME, MA, NH, NJ, NY, PA, RI, VT)	2. North Central (IL, IN, IA, KS, MI, MN, MO, NE, ND,	3. South (AL, AR, DE, D.C., FL, GA, KY, LA, MD, MS, NC	4. West (AK, AZ, CA, CO, HI, ID, MT, NV, NM, OR, UT, WA,	Total
Respondent - Religion Major Group	1. Protestant	205	579	1201	390	2375
	2. Catholic [Roman Catholic]	232	253	340	269	1094
	3. Jewish	21	8	25	19	73
	4. Other and none (also includes DK preference)	160	187	299	350	996
Total		618	1027	1865	1028	4538

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,274	<,001
	Cramer's V	,158	<,001
N of Valid Cases		4538	

2.9.1 Correlatie leeftijd en politieke kennis

```
UNIANOVA VCF0101 BY Pol_Ken
 /METHOD=SSTYPE(3)
 /INTERCEPT=INCLUDE
 /PRINT F DESCRIPTIVE
 /CRITERIA=ALPHA(.05)
 /DESIGN=Pol_Ken.
```

Tests of Between-Subjects Effects

Dependent Variable: Respondent - Age

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	29056,398[a]	3	9685,466	33,499	<,001
Intercept	7932089,660	1	7932089,660	27434,930	,000
Pol_Ken	29056,398	3	9685,466	33,499	<,001
Error	1310887,036	4534	289,124		
Total	11635790,000	4538			
Corrected Total	1339943,434	4537			

a. R Squared = ,022 (Adjusted R Squared = ,021)

2.9.2 Correlatie leeftijd en verkiezingsjaar

```
UNIANOVA VCF0101 BY VCF0004
 /METHOD=SSTYPE(3)
 /INTERCEPT=INCLUDE
 /CRITERIA=ALPHA(0.05)
 /DESIGN=VCF0004.
```

Tests of Between-Subjects Effects

Dependent Variable: Respondent - Age

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	746,662[a]	2	373,331	1,264	,283
Intercept	9593031,388	1	9593031,388	32485,441	,000
VCF0004	746,662	2	373,331	1,264	,283
Error	1339196,772	4535	295,302		
Total	11635790,000	4538			
Corrected Total	1339943,434	4537			

a. R Squared = ,001 (Adjusted R Squared = ,000)

2.9.3 Correlatie leeftijd en etniciteit

Tests of Between-Subjects Effects

Dependent Variable: Respondent - Age

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	33055,087[a]	3	11018,362	38,226	<,001
Intercept	3339100,803	1	3339100,803	11584,374	,000
VCF0105b	33055,087	3	11018,362	38,226	<,001
Error	1306888,347	4534	288,242		
Total	11635790,000	4538			
Corrected Total	1339943,434	4537			

a. R Squared = ,025 (Adjusted R Squared = ,024)

2.9.3 Correlatie leeftijd en herkomstgebied

```
UNIANOVA VCF0101 BY VCF0112
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(0.05)
  /DESIGN=VCF0112.
```

Tests of Between-Subjects Effects

Dependent Variable: Respondent - Age

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6583,732[a]	3	2194,577	7,463	<,001
Intercept	8875175,796	1	8875175,796	30179,438	,000
VCF0112	6583,732	3	2194,577	7,463	<,001
Error	1333359,702	4534	294,080		
Total	11635790,000	4538			
Corrected Total	1339943,434	4537			

a. R Squared = ,005 (Adjusted R Squared = ,004)

2.9.3 Correlatie leeftijd en voorkeurskandidaat

```
UNIANOVA VCF0101 BY VCF0303
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(0.05)
  /DESIGN=VCF0303.
```

Tests of Between-Subjects Effects

Dependent Variable: Respondent - Age

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4045,041[a]	2	2022,521	6,866	,001
Intercept	6178461,821	1	6178461,821	20974,143	,000
VCF0303	4045,041	2	2022,521	6,866	,001
Error	1335898,393	4535	294,575		
Total	11635790,000	4538			
Corrected Total	1339943,434	4537			

a. R Squared = ,003 (Adjusted R Squared = ,003)

2.9.4 Correlatie leeftijd en religie

```
UNIANOVA VCF0101 BY VCF0128
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(0.05)
  /DESIGN=VCF0128.
```

Tests of Between-Subjects Effects

Dependent Variable: Respondent - Age

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	28803,472[a]	3	9601,157	33,201	<,001
Intercept	2263152,100	1	2263152,100	7826,115	,000
VCF0128	28803,472	3	9601,157	33,201	<,001
Error	1311139,962	4534	289,180		
Total	11635790,000	4538			
Corrected Total	1339943,434	4537			

a. R Squared = ,021 (Adjusted R Squared = ,021)

2.9.5 Correlatie leeftijd en religie

```
UNIANOVA VCF0101 BY Gen_Dum
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(0.05)
  /DESIGN=Gen_Dum.
```

Tests of Between-Subjects Effects

Dependent Variable: Respondent - Age

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	957,372[a]	1	957,372	3,243	,072
Intercept	10172045,358	1	10172045,358	34459,207	,000
Gen_Dum	957,372	1	957,372	3,243	,072
Error	1338986,062	4536	295,191		
Total	11635790,000	4538			
Corrected Total	1339943,434	4537			

a. R Squared = ,001 (Adjusted R Squared = ,000)

2.9.6 Correlatie leeftijd en gender

```
UNIANOVA VCF0101 BY Gen_Dum
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(0.05)
  /DESIGN=Gen_Dum.
```

Tests of Between-Subjects Effects

Dependent Variable: Respondent - Age

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	957,372[a]	1	957,372	3,243	,072
Intercept	10172045,358	1	10172045,358	34459,207	,000
Gen_Dum	957,372	1	957,372	3,243	,072
Error	1338986,062	4536	295,191		
Total	11635790,000	4538			
Corrected Total	1339943,434	4537			

a. R Squared = ,001 (Adjusted R Squared = ,000)

2.9.7 Correlatie leeftijd en voorspelling

```
UNIANOVA VCF0101 BY Goed_geschat
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(0.05)
  /DESIGN=Goed_geschat.
```

Tests of Between-Subjects Effects

Dependent Variable: Respondent - Age

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2151,575[a]	1	2151,575	7,295	,007
Intercept	8327199,353	1	8327199,353	28234,718	,000
Goed_geschat	2151,575	1	2151,575	7,295	,007
Error	1337791,859	4536	294,928		
Total	11635790,000	4538			
Corrected Total	1339943,434	4537			

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

2.9.8 Correlatie leeftijd en opleiding

```
UNIANOVA VCF0101 BY VCF0110
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(0.05)
  /DESIGN=VCF0110.
```

Tests of Between-Subjects Effects

Dependent Variable: Respondent - Age

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	73073,273[a]	3	24357,758	87,174	<,001
Intercept	5287579,785	1	5287579,785	18923,713	,000
VCF0110	73073,273	3	24357,758	87,174	<,001
Error	1266870,161	4534	279,416		
Total	11635790,000	4538			
Corrected Total	1339943,434	4537			

a. R Squared = ,055 (Adjusted R Squared = ,054)

3.1 Logistische regressie

De logistische regressie die uitgevoerd is, is een binaire hiërarchische logistische regressie in model 1, 2, 3 en 5. Juist voorspelde verkiezingsuitslag zal de afhankelijke variabele zijn. Het plan de regressie is het volgende:
Model 1: alle controlevariabelen worden aan het model toegevoegd, dit zijn leeftijd, verkiezingsjaar, religie, etniciteit, gebied van herkomst en voorkeurskandidaat. Juist voorspelde verkiezingsuitslag is de afhankelijke variabele.
Model 2: opleidingsniveau wordt toegevoegd aan de controlevariabelen. Juist voorspelde verkiezingsuitslag is de afhankelijke variabele.
Model 3: gender wordt toegevoegd aan opleidingsniveau en de controlevariabelen. Juist voorspelde verkiezingsuitslag is de afhankelijke variabele.
Model 4: een lineaire regressie wordt uitgevoerd met politieke kennis als afhankelijke variabele en gender, opleidingsniveau en de controlevariabelen als de onafhankelijke variabelen.
Model 5: politieke kennis wordt toegevoegd aan gender, opleidingsniveau en de controle variabelen. Juist voorspelde verkiezingsuitslag is de afhankelijke variabele.

3.1.1 Model 1 Binaire logistische regressie met juist voorspelde verkiezingsuitslag al de afhankelijke variabele

```
LOGISTIC REGRESSION VARIABLES Goed_geschat
  /METHOD=ENTER VCF0101 Zwart_niet_Hispanic Hispanic Other_multiple Verkiezing_2004 Verkiezing_2008
    Noordoost Noord_centr West Rep_Voorkeur Dem_Voorkeur Katholiek Joods Anders_geen
  /METHOD=ENTER VCF0101 Zwart_niet_Hispanic Hispanic Other_multiple Noordoost Noord_centr West
    Anders_geen Katholiek Joods Verkiezing_2004 Verkiezing_2008 Rep_Voorkeur Dem_Voorkeur High_School
    College_advanced_degree Some_College
  /METHOD=ENTER VCF0101 Gen_Dum High_School Some_College College_advanced_degree
```

```
  Zwart_niet_Hispanic Hispanic Other_multiple Noordoost Noord_centr West Anders_geen Katholiek Joods
  Verkiezing_2004 Verkiezing_2008 Rep_Voorkeur Dem_Voorkeur
/METHOD=ENTER VCF0101 Pol_Ken Gen_Dum High_School Some_College College_advanced_degree
  Zwart_niet_Hispanic Hispanic Other_multiple Noordoost Noord_centr West Anders_geen
  Katholiek Joods Verkiezing_2004 Verkiezing_2008 Rep_Voorkeur Dem_Voorkeur
/SAVE=PRED PGROUP COOK LEVER DFBETA RESID ZRESID
/PRINT=GOODFIT CORR
/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Respondent - Age	-,006	,002	8,506	1	,004	,994
	Zwart niet-Hispanic	,337	,111	9,168	1	,002	1,401
	Hispanic	-,033	,109	,092	1	,762	,968
	Anders, meerdere	-,303	,185	2,691	1	,101	,739
	Verkiezingsjaar 2004	-1,337	,103	168,185	1	<,001	,263
	Verkiezingsjaar 2008	-1,429	,095	224,512	1	<,001	,240
	Noordoost	,014	,115	,016	1	,901	1,014
	Noord Centraal	-,082	,094	,749	1	,387	,922
	West	-,067	,093	,524	1	,469	,935
	Voorkeur Republikein	-,036	,119	,089	1	,766	,965
	Voorkeur Democraat	,557	,115	23,283	1	<,001	1,745
	Katholiek	,063	,094	,446	1	,504	1,065
	Joods	,362	,315	1,323	1	,250	1,436
	Anders en geen	,073	,093	,622	1	,430	1,076
	Constant	1,975	,175	127,414	1	<,001	7,205

a. Variable(s) entered on step 1: Respondent - Age, Zwart niet-Hispanic, Hispanic, Anders, meerdere, Verkiezingsjaar 2004, Verkiezingsjaar 2008, Noordoost, Noord Centraal, West, Voorkeur Republikein, Voorkeur Democraat, Katholiek, Joods, Anders en geen.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	4941,661[a]	,083	,120

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	69,028	8	<,001

3.1.2 Model 2 Binaire logistische regressie met juist voorspelde verkiezingsuitslag al de afhankelijke variabele

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Respondent - Age	-,005	,002	5,245	1	,022	,995
	Zwart niet-Hispanic	,352	,112	9,892	1	,002	1,422
	Hispanic	,026	,112	,053	1	,817	1,026
	Anders, meerdere	-,285	,185	2,368	1	,124	,752
	Verkiezingsjaar 2004	-1,342	,103	168,530	1	<,001	,261
	Verkiezingsjaar 2008	-1,438	,096	225,707	1	<,001	,237
	Noordoost	,010	,115	,008	1	,928	1,010
	Noord Centraal	-,088	,095	,858	1	,354	,916
	West	-,076	,094	,658	1	,417	,927
	Voorkeur Republikein	-,054	,120	,200	1	,655	,948
	Voorkeur Democraat	,547	,116	22,277	1	<,001	1,728
	Katholiek	,053	,094	,318	1	,573	1,054
	Joods	,312	,317	,971	1	,324	1,366
	Anders en geen	,066	,093	,499	1	,480	1,068
	High_School	,548	,187	8,623	1	,003	1,730
	College_Advanced_Degree	,566	,195	8,461	1	,004	1,761
	Some_College	,426	,191	4,956	1	,026	1,531
	Constant	1,442	,262	30,391	1	<,001	4,228

a. Variable(s) entered on step 1: Respondent - Age, Zwart niet-Hispanic, Hispanic, Anders, meerdere, Verkiezingsjaar 2004, Verkiezingsjaar 2008, Noordoost, Noord Centraal, West, Voorkeur Republikein, Voorkeur Democraat, Katholiek, Joods, Anders en geen, High_School, College_Advanced_Degree, Some_College.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	4930,974[a]	,085	,124

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	74,610	8	<,001

3.1.3 Model 3 Binaire logistische regressie met juist voorspelde verkiezingsuitslag al de afhankelijke variabele

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	4929,326[a]	,086	,124

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	57,721	8	<,001

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Respondent - Age	-,005	,002	5,117	1	,024	,995
	Zwart niet-Hispanic	,349	,112	9,682	1	,002	1,417
	Hispanic	,025	,112	,052	1	,820	1,026
	Anders, meerdere	-,291	,185	2,463	1	,117	,748
	Verkiezingsjaar 2004	-1,343	,103	168,774	1	<,001	,261
	Verkiezingsjaar 2008	-1,437	,096	225,395	1	<,001	,238
	Noordoost	,009	,115	,006	1	,938	1,009
	Noord Centraal	-,089	,095	,882	1	,348	,915
	West	-,078	,094	,693	1	,405	,925
	Voorkeur Republikein	-,059	,120	,238	1	,626	,943
	Voorkeur Democraat	,552	,116	22,701	1	<,001	1,737
	Katholiek	,050	,094	,281	1	,596	1,051
	Joods	,307	,317	,938	1	,333	1,359
	Anders en geen	,056	,093	,354	1	,552	1,057
	High_School	,549	,187	8,643	1	,003	1,731
	College_Advanced_Degree	,565	,195	8,412	1	,004	1,759
	Some_College	,430	,191	5,056	1	,025	1,537
	Gen_Dum	-,091	,071	1,645	1	,200	,913
	Constant	1,491	,264	31,805	1	<,001	4,443

a. Variable(s) entered on step 1: Respondent - Age, Zwart niet-Hispanic, Hispanic, Anders, meerdere, Verkiezingsjaar 2004, Verkiezingsjaar 2008, Noordoost, Noord Centraal, West, Voorkeur Republikein, Voorkeur Democraat, Katholiek, Joods, Anders en geen, High_School, College_Advanced_Degree, Some_College, Gen_Dum.

3.1.4 Model 4 lineaire regressie met politieke kennis als afhankelijke variabele

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA COLLIN TOL
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT Pol_Ken
  /METHOD=ENTER VCF0101 Gen_Dum High_School Some_College College_advanced_degree
    Zwart_niet_Hispanic Hispanic Other_multiple Noordoost Noord_centr West Anders_geen Katholiek Joods
    Verkiezing_2004 Verkiezing_2008 Rep_Voorkeur Dem_Voorkeur.
```

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	,255	,088		2,891	,004		
	Respondent - Age	,009	,001	,166	12,324	<,001	,885	1,130
	Gen_Dum	-,261	,023	-,145	-11,278	<,001	,976	1,024
	High_School	,506	,064	,275	7,900	<,001	,132	7,584
	Some_College	,852	,066	,436	12,957	<,001	,142	7,064
	College_Advanced_Degree	1,195	,067	,584	17,938	<,001	,151	6,615
	Zwart niet-Hispanic	-,396	,035	-,167	-11,165	<,001	,721	1,388
	Hispanic	-,278	,037	-,109	-7,432	<,001	,743	1,346
	Anders, meerdere	-,164	,065	-,033	-2,525	,012	,964	1,038
	Noordoost	,103	,037	,039	2,784	,005	,806	1,240
	Noord Centraal	,027	,031	,012	,857	,391	,772	1,296
	West	,027	,031	,013	,880	,379	,761	1,315
	Anders en geen	-,046	,031	-,021	-1,505	,132	,797	1,254
	Katholiek	-,015	,031	-,007	-,484	,628	,743	1,346
	Joods	,137	,093	,019	1,472	,141	,946	1,057
	Verkiezingsjaar 2004	-,278	,031	-,130	-8,922	<,001	,757	1,320
	Verkiezingsjaar 2008	-,169	,027	-,094	-6,175	<,001	,698	1,434
	Voorkeur Republikein	,352	,042	,187	8,396	<,001	,323	3,100
	Voorkeur Democraat	,307	,040	,170	7,764	<,001	,333	3,005

a. Dependent Variable: Pol_Ken

3.1.5 Model 5 Binaire logistische regressie met juist voorspelde verkiezingsuitslag al de afhankelijke variabele

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	4928,971[a]	,086	,124

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	49,722	8	<,001

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Respondent - Age	-,005	,002	5,455	1	,020	,995
	Zwart niet-Hispanic	,359	,113	10,019	1	,002	1,431
	Hispanic	,033	,112	,087	1	,769	1,034
	Anders, meerdere	-,287	,185	2,392	1	,122	,751
	Verkiezingsjaar 2004	-1,336	,104	164,698	1	<,001	,263
	Verkiezingsjaar 2008	-1,433	,096	222,878	1	<,001	,239
	Noordoost	,006	,115	,003	1	,956	1,006
	Noord Centraal	-,089	,095	,889	1	,346	,914
	West	-,078	,094	,698	1	,403	,925
	Voorkeur Republikein	-,069	,122	,319	1	,572	,934
	Voorkeur Democraat	,544	,117	21,732	1	<,001	1,723
	Katholiek	,051	,094	,289	1	,591	1,052
	Joods	,302	,317	,912	1	,340	1,353
	Anders en geen	,057	,093	,369	1	,543	1,058
	High_School	,534	,188	8,074	1	,004	1,706
	College_Advanced_Degree	,531	,202	6,899	1	,009	1,701
	Some_College	,407	,195	4,346	1	,037	1,502
	Gen_Dum	-,083	,072	1,351	1	,245	,920
	Pol_Ken	,027	,045	,355	1	,551	1,027
	Constant	1,487	,265	31,569	1	<,001	4,422

a. Variable(s) entered on step 1: Respondent - Age, Zwart niet-Hispanic, Hispanic, Anders, meerdere, Verkiezingsjaar 2004, Verkiezingsjaar 2008, Noordoost, Noord Centraal, West, Voorkeur Republikein, Voorkeur Democraat, Katholiek, Joods, Anders en geen, High_School, College_Advanced_Degree, Some_College, Gen_Dum, Pol_Ken.

3.1.6 Binaire logistische regressie met juist voorspelde verkiezingsuitslag al de afhankelijke variabele met grade school toegevoegd en high school weggelaten

```
LOGISTIC REGRESSION VARIABLES Goed_geschat
  /METHOD=ENTER VCF0101 Pol_Ken Gen_Dum Grade_School Some_College
College_advanced_degree
    Zwart_niet_Hispanic Hispanic Other_multiple Noordoost Noord_centr West Anders_geen
    Katholiek Joods Verkiezing_2004 Verkiezing_2008 Rep_Voorkeur Dem_Voorkeur
  /SAVE=PRED PGROUP COOK LEVER DFBETA RESID ZRESID
  /PRINT=GOODFIT CORR
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Respondent - Age	-,005	,002	5,455	1	,020	,995
	Pol_Ken	,027	,045	,355	1	,551	1,027
	Gen_Dum	-,083	,072	1,351	1	,245	,920
	Grade_School	-,534	,188	8,074	1	,004	,586
	Some_College	-,128	,086	2,206	1	,137	,880
	College_Advanced_Degree	-,003	,097	,001	1	,976	,997
	Zwart niet-Hispanic	,359	,113	10,019	1	,002	1,431
	Hispanic	,033	,112	,087	1	,769	1,034
	Anders, meerdere etniciteiten	-,287	,185	2,392	1	,122	,751
	Noordoost	,006	,115	,003	1	,956	1,006
	Noord Centraal	-,089	,095	,889	1	,346	,914
	West	-,078	,094	,698	1	,403	,925
	Anders en geen religie	,057	,093	,369	1	,543	1,058
	Katholiek	,051	,094	,289	1	,591	1,052
	Joods	,302	,317	,912	1	,340	1,353
	Verkiezingsjaar 2004	-1,336	,104	164,698	1	<,001	,263
	Verkiezingsjaar 2008	-1,433	,096	222,878	1	<,001	,239
	Voorkeur Republikein	-,069	,122	,319	1	,572	,934
	Voorkeur Democraat	,544	,117	21,732	1	<,001	1,723
	Constant	2,021	,188	114,949	1	<,001	7,546

a. Variable(s) entered on step 1: Respondent - Age, Pol_Ken, Gen_Dum, Grade_School, Some_College, College_Advanced_Degree, Zwart niet-Hispanic, Hispanic, Anders, meerdere etniciteiten, Noordoost, Noord Centraal, West, Anders en geen religie, Katholiek, Joods, Verkiezingsjaar 2004, Verkiezingsjaar 2008, Voorkeur Republikein, Voorkeur Democraat.

3.1.7 Binaire logistische regressie met juist voorspelde verkiezingsuitslag al de afhankelijke variabele met grade school toegevoegd en een deel van college weggelaten

```
LOGISTIC REGRESSION VARIABLES Goed_geschat
  /METHOD=ENTER VCF0101 Pol_Ken Gen_Dum Grade_School High_School
College_advanced_degree
    Zwart_niet_Hispanic Hispanic Other_multiple Noordoost Noord_centr West Anders_geen
    Katholiek Joods Verkiezing_2004 Verkiezing_2008 Rep_Voorkeur Dem_Voorkeur
  /SAVE=PRED PGROUP COOK LEVER DFBETA RESID ZRESID
  /PRINT=GOODFIT CORR
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Respondent - Age	-,005	,002	5,455	1	,020	,995
	Pol_Ken	,027	,045	,355	1	,551	1,027
	Gen_Dum	-,083	,072	1,351	1	,245	,920
	Grade_School	-,407	,195	4,346	1	,037	,666
	High_School	,128	,086	2,206	1	,137	1,136
	College_Advanced_Degree	,125	,096	1,696	1	,193	1,133
	Zwart niet-Hispanic	,359	,113	10,019	1	,002	1,431
	Hispanic	,033	,112	,087	1	,769	1,034
	Anders, meerdere etniciteiten	-,287	,185	2,392	1	,122	,751
	Noordoost	,006	,115	,003	1	,956	1,006
	Noord Centraal	-,089	,095	,889	1	,346	,914
	West	-,078	,094	,698	1	,403	,925
	Anders en geen religie	,057	,093	,369	1	,543	1,058
	Katholiek	,051	,094	,289	1	,591	1,052
	Joods	,302	,317	,912	1	,340	1,353
	Verkiezingsjaar 2004	-1,336	,104	164,698	1	<,001	,263
	Verkiezingsjaar 2008	-1,433	,096	222,878	1	<,001	,239
	Voorkeur Republikein	-,069	,122	,319	1	,572	,934
	Voorkeur Democraat	,544	,117	21,732	1	<,001	1,723
	Constant	1,893	,192	97,552	1	<,001	6,642

a. Variable(s) entered on step 1: Respondent - Age, Pol_Ken, Gen_Dum, Grade_School, High_School, College_Advanced_Degree, Zwart niet-Hispanic, Hispanic, Anders, meerdere etniciteiten, Noordoost, Noord Centraal, West, Anders en geen religie, Katholiek, Joods, Verkiezingsjaar 2004, Verkiezingsjaar 2008, Voorkeur Republikein, Voorkeur Democraat.

3.1.8 Binaire logistische regressie met juist voorspelde verkiezingsuitslag al de afhankelijke variabele met grade school toegevoegd en college of hoger weggelaten

```
LOGISTIC REGRESSION VARIABLES Goed_geschat
  /METHOD=ENTER VCF0101 Pol_Ken Gen_Dum Grade_School High_School Some_College
    Zwart_niet_Hispanic Hispanic Other_multiple Noordoost Noord_centr West Anders_geen
    Katholiek Joods Verkiezing_2004 Verkiezing_2008 Rep_Voorkeur Dem_Voorkeur
  /SAVE=PRED PGROUP COOK LEVER DFBETA RESID ZRESID
  /PRINT=GOODFIT CORR
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

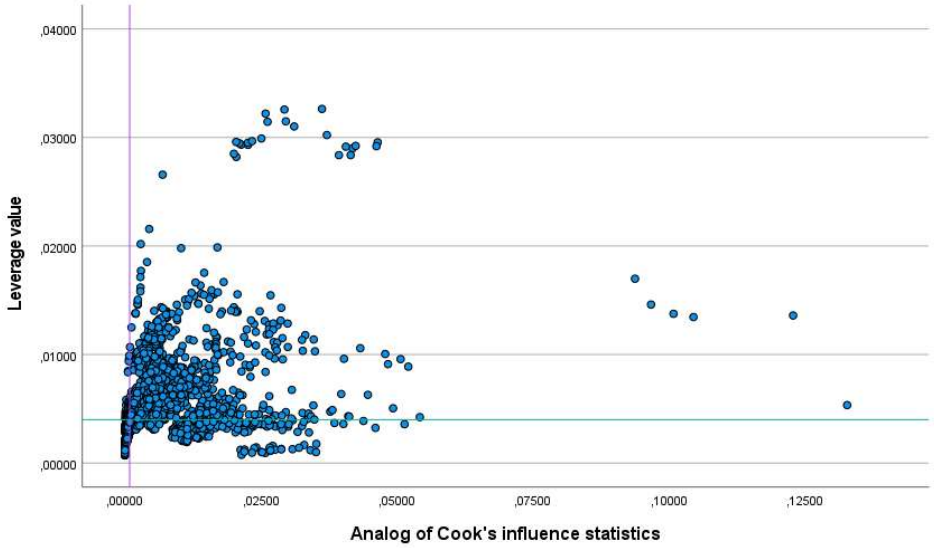
Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Respondent - Age	-,005	,002	5,455	1	,020	,995
	Pol_Ken	,027	,045	,355	1	,551	1,027
	Gen_Dum	-,083	,072	1,351	1	,245	,920
	Grade_School	-,531	,202	6,899	1	,009	,588
	High_School	,003	,097	,001	1	,976	1,003
	Some_College	-,125	,096	1,696	1	,193	,883
	Zwart niet-Hispanic	,359	,113	10,019	1	,002	1,431
	Hispanic	,033	,112	,087	1	,769	1,034
	Anders, meerdere etniciteiten	-,287	,185	2,392	1	,122	,751
	Noordoost	,006	,115	,003	1	,956	1,006
	Noord Centraal	-,089	,095	,889	1	,346	,914
	West	-,078	,094	,698	1	,403	,925
	Anders en geen religie	,057	,093	,369	1	,543	1,058
	Katholiek	,051	,094	,289	1	,591	1,052
	Joods	,302	,317	,912	1	,340	1,353
	Verkiezingsjaar 2004	-1,336	,104	164,698	1	<,001	,263
	Verkiezingsjaar 2008	-1,433	,096	222,878	1	<,001	,239
	Voorkeur Republikein	-,069	,122	,319	1	,572	,934
	Voorkeur Democraat	,544	,117	21,732	1	<,001	1,723
	Constant	2,018	,202	100,273	1	<,001	7,524

a. Variable(s) entered on step 1: Respondent - Age, Pol_Ken, Gen_Dum, Grade_School, High_School, Some_College, Zwart niet-Hispanic, Hispanic, Anders, meerdere etniciteiten, Noordoost, Noord Centraal, West, Anders en geen religie, Katholiek, Joods, Verkiezingsjaar 2004, Verkiezingsjaar 2008, Voorkeur Republikein, Voorkeur Democraat.

Bijlage 3 Assumpties, uitbijters en multicollineariteit

1 Assumpties
De enige assumptie voor een logistische regressie is die van onafhankelijke waarnemingen. In het methoden hoofdstuk is in de steekproef beschrijving uitgelegd hoe de steekproef getrokken is. Bij de ANES data is het maar mogelijk voor één persoon per huishouden om deel te nemen. Daarom kunnen we uitgaan van onafhankelijke waarnemingen (ANES, 2021).
2 Uitbijters
Daarnaast moet gekeken worden naar de uitbijters in de dataset. Aangezien de dataset erg groot is (n=4538) is het lastig om te controleren voor uitbijters. Daarvoor is gebruik gemaakt van de *Cook's distance* waarbij een score van 4/n=4/4538=0,0009 een mogelijke uitbijter is. En de leverage waarbij 2*p/n=2*10/4538=0,004 een mogelijke uitbijter is. Zoals verwacht zijn er veel mogelijke uitbijters in de dataset te vinden.



Figuur 3.2.1: Spreidingsdiagram met de Cook's distance op de x-as met de waarde 0,0009 gemarkeerd in groen en de leverage op de x-as met de waarde 0,004 gemarkeerd.

Daarnaast kunnen we ook de invloed van individuele punten op de helling bekijken. Daarvoor is naar de DFBETA van iedere helling gekeken. Aangezien het model zeer veel hellingen heeft is alleen gekeken naar de invloed op politieke kennis, gender, opleidingsniveau en juist voorspelde verkiezingen. Bovendien zijn hier de DFBETA's gebruikt die horen bij de regressie die high school als referentiegroep gebruikt (Bijlage 2, 3.1.6)

Figuur 2.2.2: De DFBETA van politieke kennis van laag naar hoog

	ZRE_1	DFB0_1	DFB1_1	DFB2_1	DFB3_1	DFB4_1	DFB5_1	DFB6_1	DFB7_1	DFB8
1	-1,30445	-,00437	,00006	-,00358	-,00266	,00088	,00300	,00363	-,00077	-,(
2	-1,89476	,01002	-,00005	-,00343	-,00265	,00247	,00322	,00474	-,00920	-,(
3	-1,58862	-,00085	,00007	-,00281	-,00205	-,00023	,00351	,00448	,00320	,(
4	-1,68690	,00220	,00001	-,00274	,00163	,00094	-,00224	,00173	-,00056	-,(
5	-2,04431	-,00317	,00011	-,00273	-,00243	-,00009	,00375	,00428	-,00626	-,(
6	-1,47261	,00347	-,00004	-,00253	-,00196	,00181	,00328	,00446	,00131	,(
7	-1,44611	,00264	,00001	-,00253	-,00186	,00052	,00287	,00347	,00236	,(
8	-1,54403	,00415	,00000	-,00251	-,00202	,00116	,00296	,00430	,00153	,(
9	-1,42999	,00796	-,00003	-,00250	-,00186	,00082	-,00190	,00136	-,00146	-,(
10	-1,18255	,00228	,00008	-,00250	-,00242	-,00169	,00291	,00367	-,00042	,(
11	-1,52159	,00238	,00004	-,00245	,00173	-,00030	-,00220	,00057	,00069	-,(
12	-1,43900	,00759	-,00005	-,00241	-,00217	,00135	,00248	,00337	,00111	,(
13	-2,74611	-,00501	,00004	-,00236	-,00304	-,00154	-,00346	-,00027	-,01158	-,(
14	-1,42729	,00650	-,00004	-,00234	-,00188	,00142	,00248	,00343	,00087	,(
15	-1,49695	,00159	,00000	-,00231	-,00173	-,00036	-,00184	,00183	,00253	,(
16	-2,18749	-,00285	,00002	-,00231	-,00237	-,00064	-,00225	,00205	,00005	-,(
17	-1,45476	,01133	-,00008	-,00230	-,00205	,00208	-,00241	,00087	-,00061	-,(

Figuur 2.2.3: De DFBETA van politieke kennis van hoog naar laag

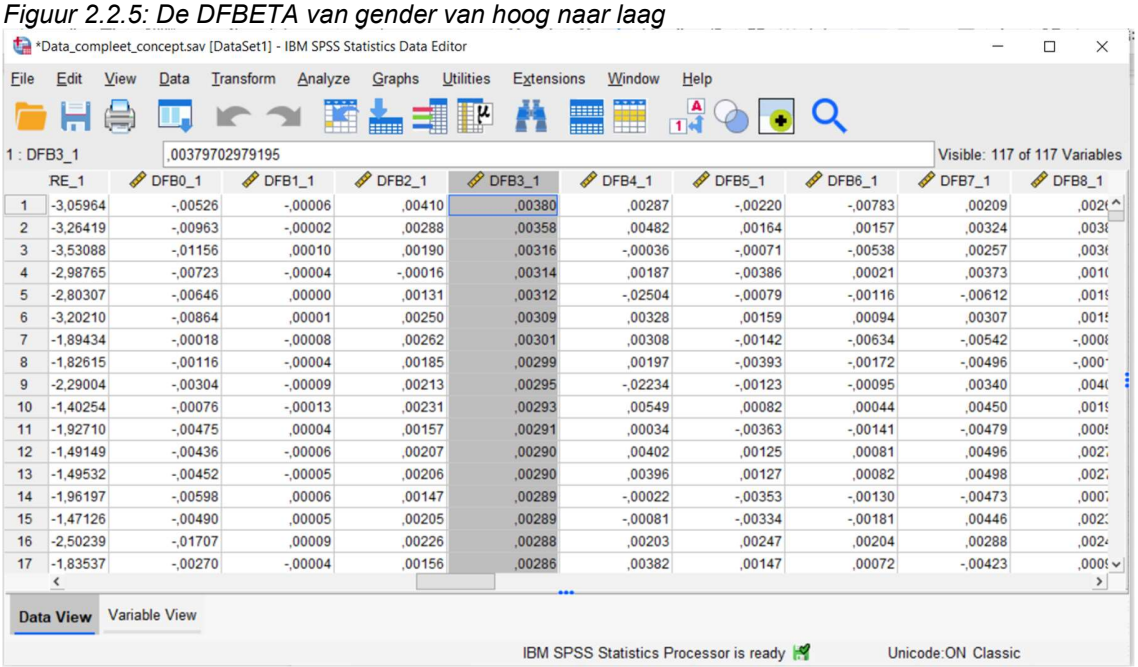
	ZRE_1	DFB0_1	DFB1_1	DFB2_1	DFB3_1	DFB4_1	DFB5_1	DFB6_1	DFB7_1	D
1	-3,05964	-,00526	-,00006	,00410	,00380	,00287	-,00220	-,00783	,00209	
2	-2,01233	,00509	-,00020	,00378	-,00138	,00456	-,00158	-,00627	-,00150	
3	-1,63502	,00068	-,00004	,00300	-,00082	,00086	-,00067	-,00355	,00158	
4	-2,81131	-,00738	-,00006	,00290	-,00054	,00316	-,00462	-,00204	,00471	
5	-3,26419	-,00963	-,00002	,00288	,00358	,00482	,00164	,00157	,00324	
6	-1,96897	-,00068	-,00015	,00264	-,00112	,00598	,00176	,00130	,00020	
7	-1,89434	-,00018	-,00008	,00262	,00301	,00308	-,00142	-,00634	-,00542	
8	-2,03017	-,00365	-,00001	,00261	-,00106	,00089	-,00391	-,00199	,00099	
9	-1,20902	,00205	-,00006	,00257	,00182	,00160	-,00091	-,00424	,00047	
10	-2,00974	-,00206	-,00012	,00253	-,00119	,00537	,00192	,00144	,00031	
11	-1,16048	-,00224	-,00001	,00251	,00214	,00059	-,00070	-,00435	,00202	
12	-3,20210	-,00864	,00001	,00250	,00309	,00328	,00159	,00094	,00307	
13	-2,19702	-,00329	-,00006	,00232	,00275	,00187	-,00115	-,00603	,00025	
14	-1,40254	-,00076	-,00013	,00231	,00293	,00549	,00082	,00044	,00450	
15	-1,59037	-,00435	,00001	,00229	-,00078	,00109	-,00060	-,00501	,00421	
16	-1,55444	-,00566	,00005	,00229	-,00070	,00012	-,00063	-,00522	,00488	
17	-2,09293	-,00392	-,00004	,00229	-,00130	,00386	,00181	,00091	,00095	

De invloed van ID 19942883 lijkt groot te zijn met 0,0041 zeker omdat het uiteindelijke effect slechts 0,027 is. De overige punten zijn lager en liggen bovendien dicht bij elkaar.

Figuur 2.2.4: De DFBETA van gender van laag naar hoog

	RE_1	DFB0_1	DFB1_1	DFB2_1	DFB3_1	DFB4_1	DFB5_1	DFB6_1	DFB7_1	DFB8_1
1	-2,74611	-,00501	,00004	-,00236	-,00304	-,00154	-,00346	-,00027	-,01158	-,002
2	-3,07650	-,01147	,00006	-,00116	-,00298	,00162	,00286	,00271	-,01032	-,001
3	-2,38167	-,00133	,00002	-,00177	-,00269	-,00040	,00051	-,00246	-,00187	,001
4	-3,38644	-,00606	,00009	-,00171	-,00269	,00089	,00363	,00468	,00211	,000
5	-2,36684	,00003	-,00002	-,00223	-,00269	,00223	,00059	-,00325	-,00286	-,011
6	-1,30445	-,00437	,00006	-,00358	-,00266	,00088	,00300	,00363	-,00077	-,005
7	-2,48504	-,00574	,00010	-,00145	-,00265	,00030	,00370	,00483	-,00097	,002
8	-2,33795	-,01146	,00006	-,00171	-,00265	-,00201	-,00257	,00113	-,00051	,001
9	-1,89476	,01002	-,00005	-,00343	-,00265	,00247	,00322	,00474	-,00920	-,001
10	-2,41151	-,00619	,00011	-,00007	-,00261	-,00030	,00306	,00301	-,00045	,000
11	-2,28693	-,01299	,00011	-,00035	-,00254	-,02697	-,00002	-,00042	-,00844	,002
12	-2,16835	,00065	,00000	-,00079	-,00253	-,00098	-,00255	,00135	-,00192	,000
13	-2,93141	-,00560	,00005	,00115	-,00250	,00223	,00248	,00201	-,01038	-,000
14	-2,33850	-,00395	,00006	,00013	-,00250	,00073	,00279	,00276	-,00062	,000
15	-2,50618	-,01491	,00011	-,00114	-,00249	-,00216	,00056	-,00351	-,00101	,003
16	-2,41523	-,01027	,00010	-,00156	-,00245	,00059	,00379	,00454	,00111	,000
17	-2,04431	-,00317	,00011	-,00273	-,00243	-,00009	,00375	,00428	-,00626	-,000

Figuur 2.2.5: De DFBETA van gender van hoog naar laag



	RE_1	DFB0_1	DFB1_1	DFB2_1	DFB3_1	DFB4_1	DFB5_1	DFB6_1	DFB7_1	DFB8_1
1	-3,05964	-,00526	-,00006	,00410	,00380	,00287	-,00220	-,00783	,00209	,002
2	-3,26419	-,00963	-,00002	,00288	,00358	,00482	,00164	,00157	,00324	,003
3	-3,53088	-,01156	,00010	,00190	,00316	-,00036	-,00071	-,00538	,00257	,003
4	-2,98765	-,00723	-,00004	-,00016	,00314	,00187	-,00386	,00021	,00373	,001
5	-2,80307	-,00646	,00000	,00131	,00312	-,02504	-,00079	-,00116	-,00612	,001
6	-3,20210	-,00864	,00001	,00250	,00309	,00328	,00159	,00094	,00307	,001
7	-1,89434	-,00018	-,00008	,00262	,00301	,00308	-,00142	-,00634	-,00542	-,000
8	-1,82615	-,00116	-,00004	,00185	,00299	,00197	-,00393	-,00172	-,00496	-,000
9	-2,29004	-,00304	-,00009	,00213	,00295	-,02234	-,00123	-,00095	,00340	,004
10	-1,40254	-,00076	-,00013	,00231	,00293	,00549	,00082	,00044	,00450	,001
11	-1,92710	-,00475	,00004	,00157	,00291	,00034	-,00363	-,00141	-,00479	,000
12	-1,49149	-,00436	-,00006	,00207	,00290	,00402	,00125	,00081	,00496	,002
13	-1,49532	-,00452	-,00005	,00206	,00290	,00396	,00127	,00082	,00498	,002
14	-1,96197	-,00598	,00006	,00147	,00289	-,00022	-,00353	-,00130	-,00473	,000
15	-1,47126	-,00490	,00005	,00205	,00289	-,00081	-,00334	-,00181	,00446	,002
16	-2,50239	-,01707	,00009	,00226	,00288	,00203	,00247	,00204	,00288	,002
17	-1,83537	-,00270	-,00004	,00156	,00286	,00382	,00147	,00072	-,00423	,000

Hier is geen waarde die er meer uitspringt dan een andere.

Figuur 2.2.6: De DFBETA van grade school van laag naar hoog

	_1	DFB2_1	DFB3_1	DFB4_1	DFB5_1	DFB6_1	DFB7_1	DFB8_1	DFB9_1	DFB10_1
1	00011	-,00035	-,00254	-,02697	-,00002	-,00042	-,00844	,00261	,00038	,00298
2	00000	,00131	,00312	-,02504	-,00079	-,00116	-,00612	,00196	-,00027	,00193
3	00007	,00003	,00274	-,02417	-,00054	-,00028	-,00747	-,00057	,00001	-,00969
4	00007	,00112	-,00130	-,02312	-,00082	-,00126	-,00662	,00111	-,00095	,00184
5	00002	,00076	-,00144	-,02280	-,00039	-,00064	,00058	-,00373	-,00066	,00275
6	00001	,00079	-,00142	-,02259	-,00043	-,00068	,00055	-,00380	-,00069	,00274
7	00009	,00213	,00295	-,02234	-,00123	-,00095	,00340	,00404	,00209	,00344
8	00005	-,00205	,00156	-,02189	,00060	,00120	-,00460	,00206	-,00003	,00135
9	00007	,00095	,00257	-,02165	-,00106	-,00157	-,00716	,00088	-,00059	,00162
10	00004	,00106	-,00118	-,02131	-,00067	-,00176	,00097	-,00798	-,00088	-,00194
11	00001	-,00187	,00146	-,02029	,00034	,00080	-,00601	-,00072	-,00058	-,00043
12	00005	,00156	-,00120	-,01973	-,00034	-,00043	,00157	,00238	,00021	,00166
13	00003	,00030	-,00114	-,01971	-,00006	-,00053	-,00367	,00259	,00010	,00101
14	00007	,00027	-,00152	-,01966	-,00015	-,00010	,00006	,00031	-,00032	-,00071
15	00007	,00109	-,00128	-,01962	-,00084	-,00079	,00117	,00480	,00076	,00330
16	00007	,00177	-,00117	-,01932	-,00040	-,00053	,00114	,00114	,00085	-,00787
17	00001	,00039	-,00109	-,01900	-,00017	-,00062	-,00370	,00238	,00001	,00099

Figuur 2.2.7: De DFBETA van grade school van hoog naar laag

	_1	DFB2_1	DFB3_1	DFB4_1	DFB5_1	DFB6_1	DFB7_1	DFB8_1	DFB9_1	DFB10_1
1	00004	,00019	,00157	,01887	-,00029	-,00043	-,00052	-,00060	-,00012	,00044
2	00000	-,00109	,00115	,01866	-,00009	,00015	-,00210	-,00252	-,00142	,00710
3	00001	-,00105	-,00174	,01804	-,00032	-,00033	-,00111	-,00263	-,00126	,00047
4	00007	-,00034	-,00158	,01781	-,00008	,00037	-,00228	-,00226	-,00158	,00035
5	00003	,00006	-,00137	,01749	-,00052	-,00095	-,00130	-,00142	-,00131	,00575
6	00003	,00016	-,00140	,01740	-,00047	-,00095	-,00156	-,00222	-,00074	-,00160
7	00005	,00015	,00139	,01713	-,00027	,00021	,00052	,00149	-,00038	,00056
8	00000	-,00008	,00127	,01699	-,00014	,00063	-,00043	,00370	-,00034	,00184
9	00003	-,00010	-,00158	,01695	-,00018	-,00071	-,00141	-,00180	-,00051	-,00165
10	00006	-,00034	-,00157	,01654	-,00013	,00000	-,00277	-,00308	-,00161	-,00150
11	00003	,00045	-,00137	,01654	-,00037	-,00036	-,00040	,00379	,00045	-,00030
12	00001	,00185	,00189	,01648	-,00089	-,00145	,00094	,00565	,00123	-,00090
13	00010	,00036	,00131	,01641	-,00027	,00014	-,00055	,00096	-,00026	-,00133
14	00000	,00043	-,00108	,01625	-,00010	-,00056	-,00109	-,00170	-,00210	,00469
15	00007	,00032	,00151	,01624	-,00016	,00029	-,00068	,00244	-,00066	-,00103
16	00001	-,00010	,00116	,01589	,00015	,00085	-,00134	,00368	,00004	-,00024
17	00004	,00067	-,00082	,01543	-,00046	-,00081	,00081	,00390	,00012	-,00148

Zoals te zien hebben de meeste punten vrij veel invloed op deze helling. De meeste van deze personen hebben ook een grade school opleiding afgerond. Dit is logisch aangezien er maar 168 mensen een grade school opleiding hadden. Tegenover de andere opleidingsniveaus waar meer dan 1000 respondenten in die categorie vielen. In deze dataset zijn mensen met zo'n lage opleiding dus eigenlijk outliers.

Figuur 2.2.8: De DFBETA van deel van college van laag naar hoog

	_1	DFB2_1	DFB3_1	DFB4_1	DFB5_1	DFB6_1	DFB7_1	DFB8_1	DFB9_1	DFB10_1
1	00006	,00290	-,00054	,00316	-,00462	-,00204	,00471	,00206	,00198	,00256
2	00002	,00194	-,00109	,00156	-,00450	-,00214	-,00665	,00003	,00008	,00226
3	00006	,00164	,00266	,00141	-,00417	-,00109	-,00108	,00128	,00046	,00110
4	00010	,00078	-,00126	,00249	-,00409	-,00179	,00118	-,00098	,00103	-,00049
5	00015	,00150	-,00133	,00372	-,00408	-,00065	-,00076	,00186	-,00006	,00383
6	00005	,00075	-,00140	-,00046	-,00397	-,00081	,00161	,00310	,00165	,00148
7	00005	,00162	,00277	,00129	-,00393	-,00127	,00054	-,00092	,00071	-,00056
8	00004	,00185	,00299	,00197	-,00393	-,00172	-,00496	-,00012	,00054	,00211
9	00001	,00261	-,00106	,00089	-,00391	-,00199	,00099	-,00020	,00081	-,00072
10	00004	-,00016	,00314	,00187	-,00386	,00021	,00373	,00109	,00177	,00296
11	00006	,00120	-,00153	,00128	-,00383	-,00089	-,00140	,00105	,00010	,00106
12	00008	,00142	-,00149	,00223	-,00378	-,00038	-,00093	,00125	,00104	-,00694
13	00006	,00177	-,00078	,00177	-,00374	-,00200	-,00481	-,00044	,00102	-,00842
14	00006	,00020	-,00164	,00196	-,00370	-,00053	-,00042	-,00764	,00069	-,00875
15	00010	,00057	,00264	,00259	-,00369	-,00092	-,00569	-,00113	,00001	,00220
16	00006	,00003	,00224	,00079	-,00368	,00003	-,00174	,00071	,00017	,00126
17	00004	,00157	,00291	,00034	-,00363	-,00141	-,00479	,00053	,00088	,00220

Figuur 2.2.9: De DFBETA van deel van college van hoog naar laag

	_1	DFB2_1	DFB3_1	DFB4_1	DFB5_1	DFB6_1	DFB7_1	DFB8_1	DFB9_1	DFB10_1
1	0014	-,00193	-,00217	,00048	,00396	,00469	,00414	,00234	,00127	,00302
2	0001	-,00131	,00169	,00264	,00394	,00499	-,00115	-,00040	-,02844	,00201
3	0010	-,00156	-,00245	,00059	,00379	,00454	,00111	,00099	,00003	,00255
4	0011	-,00273	-,00243	-,00009	,00375	,00428	-,00626	-,00030	-,00056	,00251
5	0010	-,00145	-,00265	,00030	,00370	,00483	-,00097	,00215	,00070	-,00713
6	0009	-,00171	-,00269	,00089	,00363	,00468	,00211	,00099	,00264	-,00788
7	0006	-,00140	-,00237	,00136	,00358	,00435	,00097	,00068	-,00012	,00252
8	0006	-,00110	,00141	,00040	,00356	,00469	-,00137	,00128	,00135	,00084
9	0006	-,00138	-,00236	,00145	,00356	,00432	,00095	,00065	-,00014	,00251
10	0010	-,00195	,00142	-,00010	,00355	,00439	,00001	,00220	,00251	,00099
11	0007	-,00281	-,00205	-,00023	,00351	,00448	,00320	,00110	,00178	-,00751
12	0012	-,00147	,00193	-,00031	,00346	,00423	,00387	,00117	,00255	-,00741
13	0007	-,00139	-,00240	,00041	,00343	,00371	,00002	-,00081	-,00027	-,00044
14	0006	-,00135	-,00239	,00058	,00339	,00367	-,00001	-,00087	-,00030	-,00045
15	0009	-,00097	,00183	,00069	,00338	,00463	-,00065	,00229	,00104	-,00719
16	0009	-,00053	-,00173	-,00030	,00337	,00349	,00046	,00351	-,02105	,00138
17	0005	,00030	-,00197	,00140	,00335	,00340	,00025	,00122	,00064	-,00060

Geen grote invloed rijke punten.

Figuur 2.2.10: De DFBETA van college of hoger van laag naar hoog

	_1	DFB2_1	DFB3_1	DFB4_1	DFB5_1	DFB6_1	DFB7_1	DFB8_1	DFB9_1	DFB10_1
1	0006	,00410	,00380	,00287	-,00220	-,00783	,00209	,00264	,00188	,00124
2	0007	,00172	,00272	,00230	-,00165	-,00648	-,00019	,00133	,00130	,00119
3	0008	,00262	,00301	,00308	-,00142	-,00634	-,00542	-,00086	,00032	,00208
4	0015	,00223	-,00093	,00498	-,00146	-,00631	,00337	,00011	,00115	,00263
5	0020	,00378	-,00138	,00456	-,00158	-,00627	-,00150	-,00024	,00121	,00035
6	0006	,00232	,00275	,00187	-,00115	-,00603	,00025	-,00148	,00059	-,00063
7	0007	,00048	-,00239	-,00020	-,00061	-,00580	-,00911	-,00035	,00050	,00382
8	0004	,00156	,00275	,00227	-,00133	-,00579	,00095	,00329	,00169	,00425
9	0001	,00145	,00257	,00308	-,00062	-,00572	-,00104	-,00996	-,00093	,00114
10	0002	,00194	-,00134	,00172	-,00078	-,00571	,00131	-,00675	,00188	-,00812
11	0001	,00054	-,00143	-,00025	-,00061	-,00559	-,00610	-,00119	,00001	-,00117
12	0000	,00081	,00243	,00026	-,00072	-,00556	-,00673	-,00227	,00016	-,00049
13	0000	,00095	,00262	,00064	-,00079	-,00542	-,00512	,00038	,00052	,00156
14	0002	,00224	-,00081	,00035	-,00069	-,00539	,00331	,00053	,00191	-,00025
15	0010	,00190	,00316	-,00036	-,00071	-,00538	,00257	,00369	,00338	-,00755
16	0000	,00038	-,00160	,00007	-,00043	-,00526	-,00691	-,00239	-,00017	-,00050
17	0003	,00183	-,00105	,00138	-,00049	-,00523	,00019	-,00376	,00084	-,00046

Figuur 2.2.11: De DFBETA van college of hoger van hoog naar laag

	_1	DFB2_1	DFB3_1	DFB4_1	DFB5_1	DFB6_1	DFB7_1	DFB8_1	DFB9_1	DFB10_1
1	0001	-,00131	,00169	,00264	,00394	,00499	-,00115	-,00040	-,02844	,00201
2	0010	-,00145	-,00265	,00030	,00370	,00483	-,00097	,00215	,00070	-,00713
3	0005	-,00343	-,00265	,00247	,00322	,00474	-,00920	-,00113	-,00090	-,00633
4	0014	-,00193	-,00217	,00048	,00396	,00469	,00414	,00234	,00127	,00302
5	0006	-,00110	,00141	,00040	,00356	,00469	-,00137	,00128	,00135	,00084
6	0009	-,00171	-,00269	,00089	,00363	,00468	,00211	,00099	,00264	-,00788
7	0009	-,00097	,00183	,00069	,00338	,00463	-,00065	,00229	,00104	-,00719
8	0017	,00163	-,00137	,00703	,00202	,00458	,00082	-,00091	,00121	-,00768
9	0010	-,00156	-,00245	,00059	,00379	,00454	,00111	,00099	,00003	,00255
10	0007	-,00281	-,00205	-,00023	,00351	,00448	,00320	,00110	,00178	-,00751
11	0004	-,00253	-,00196	,00181	,00328	,00446	,00131	,00018	,00080	,00014
12	0010	-,00195	,00142	-,00010	,00355	,00439	,00001	,00220	,00251	,00099
13	0006	-,00140	-,00237	,00136	,00358	,00435	,00097	,00068	-,00012	,00252
14	0006	-,00138	-,00236	,00145	,00356	,00432	,00095	,00065	-,00014	,00251
15	0000	-,00251	-,00202	,00116	,00296	,00430	,00153	,00248	,00128	-,00598
16	0011	-,00273	-,00243	-,00009	,00375	,00428	-,00626	-,00030	-,00056	,00251
17	0012	-,00147	,00193	-,00031	,00346	,00423	,00387	,00117	,00255	-,00741

Geen grote invloed rijke punten.

3 Multicollineariteit

Voor het controleren van de multicollineariteit is de regressie nogmaals uitgevoerd maar dan als lineaire regressie.

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA COLLIN TOL
```

/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT Goed_geschat
/METHOD=ENTER VCF0101 Pol_Ken Gen_Dum High_School Some_College College_advanced_degree
Zwart_niet_Hispanic Hispanic Other_multiple Noordoost Noord_centr West Anders_geen
Katholiek Joods Verkiezing_2004 Verkiezing_2008 Rep_Voorkeur Dem_Voorkeur .

Zoals in de tabel te zien is is er alleen een hoge multicollineariteit tussen de opleidingsniveau variabelen. Dit komt waarschijnlijk doordat ze een voorspeller voor elkaar zijn en is dus geen probleem voor de analyse.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	,773	,049		15,674	<,001		
	Respondent - Age	-,001	,000	-,036	-2,325	,020	,856	1,168
	Pol_Ken	,004	,008	,009	,540	,589	,725	1,380
	Gen_Dum	-,016	,013	-,018	-1,233	,217	,949	1,053
	High_School	,104	,036	,114	2,885	,004	,130	7,689
	Some_College	,079	,037	,082	2,116	,034	,136	7,326
	College_Advanced_Degree	,104	,039	,103	2,707	,007	,141	7,086
	Zwart niet-Hispanic	,068	,020	,058	3,410	<,001	,701	1,426
	Hispanic	,007	,021	,005	,328	,743	,734	1,363
	Anders, meerdere	-,057	,036	-,023	-1,558	,119	,962	1,039
	Noordoost	,002	,021	,002	,115	,909	,805	1,242
	Noord Centraal	-,013	,017	-,012	-,761	,447	,771	1,296
	West	-,014	,017	-,013	-,791	,429	,761	1,315
	Anders en geen	,011	,017	,010	,628	,530	,797	1,255
	Katholiek	,009	,017	,009	,529	,597	,743	1,346
	Joods	,047	,052	,013	,904	,366	,945	1,058
	Verkiezingsjaar 2004	-,222	,018	-,208	-12,612	<,001	,744	1,344
	Verkiezingsjaar 2008	-,239	,015	-,266	-15,542	<,001	,692	1,446
	Voorkeur Republikein	-,011	,024	-,011	-,446	,655	,318	3,148
	Voorkeur Democraat	,103	,022	,115	4,625	<,001	,328	3,045

a. Dependent Variable: Goed_geschat