‘Oud maar niet versleten’

Een onderzoek naar beweegredenen van mensen om door te werken na het bereiken van de AOW-leeftijd.

Universiteit: Rijksuniversiteit Groningen

Faculteit: Gedrags- en Maatschappijwetenschappen

Opleiding: Sociologie (bachelor)

Vak: Bachelor Afsluitend werkstuk

Begeleider: Vincenz Frey

Student: Nikki van Kempen (S3997839)

Datum: 08-06-2021

Inhoudsopgave

Samenvatting 2
Inleiding 4
Theoretisch Kader 6
Fysieke Belasting van het werk 6
Partnerstatus 7
Verwachte Inkomsten 8
Controlevariabelen 8
Methoden 10
Dataset en Steekproef 10
Operationalisaties 11
Analyseplan 12
Resultaten 14
Beschrijvende statistieken 14
Assumptietoetsing 16
Hypothesetoetsing 18
Modelevaluatie 21
Conclusie en Discussie 23
Literatuurlijst 27
Bijlage 1 30
Operationalisaties 30
Bijlage 2 38
Descriptieve statistieken 38
Samenhang tussen de variabelen 44
Hypothesetoetsing 51
Modelfit 57
Bijlage 3 59
Assumpties 59
Multicollineariteit 61
Uitschieters 62

Samenvatting

De krapte op de arbeidsmarkt is anno 2022 groter dan ooit. Een van de mogelijke oplossingen voor dit probleem is het motiveren van de groep mensen die de AOW-leeftijd nadert, om langer door te blijven werken. Door de vergrijzing van Nederland wordt deze groep mensen in onze samenleving steeds groter. Als ook maar een klein deel van deze groep besluit om langer door te werken, kan dit al een groot verschil maken voor de krapte op de arbeidsmarkt. Hiervoor is het van belang om onderzoek te doen naar de beweegredenen van mensen die na het bereiken van AOW-leeftijd langer willen doorwerken. Dit onderzoek richt zich op drie mogelijke beweegredenen die mensen kunnen hebben om langer door te willen werken. De onderzoeksvraag is als volgt: *'Wat is de invloed van de fysieke belasting van het werk, het hebben van een partner en de verwachte inkomsten op het langer willen doorwerken na de AOW-leeftijd?'.* Om deze onderzoeksvraag te beantwoorden is gebruik gemaakt van data uit het LISS-panel, waarbij de steekproef (N=1311) mensen van een leeftijd tussen 45 en 67 uit Nederland bevat. De resultaten van dit onderzoek laten zien dat de mate van fysieke belasting van het werk een grote invloed heeft op de keuze om langer door te willen werken. Er is geen ondersteuning gevonden voor de invloed van het hebben van een partner of de verwachte inkomsten op de keuze om langer door te willen werken. Dit biedt nuttige inzichten voor te voeren beleid om mensen te motiveren tot langer doorwerken. In dit onderzoek wordt dan ook aangeraden om bij het opstellen van dergelijk beleid de focus te leggen op werknemers die geen, of af en toe, fysiek belastend werk verrichten. Daarnaast wordt ook aangeraden om door middel van beleid werkgevers die fysiek belastend werk faciliteren, te motiveren om dergelijk werk voor ouderen minder belastend te maken, zodat zij eerder langer willen (en kunnen) doorwerken.

Inleiding

Het Nederlandse pensioenstelsel is al jarenlang een van de beste ter wereld (ABP, 2019). De betrouwbaarheid, houdbaarheid en ook betaalbaarheid maken dat het stelsel erg goed functioneert. Ook financieel gezien hebben de pensioengerechtigden het in Nederland vrij ruim, doordat zij naast het ontvangen van een AOW-uitkering ook een mogelijkheid hebben tot het opbouwen van een aanvullend pensioen (ABP, 2019). Desondanks kiezen steeds meer mensen ervoor om door te werken na het bereiken van de AOW-leeftijd. Volgens het CBS is het aantal werkende 65-plussers toegenomen van 75.000 in 2003 tot 300.000 in 2020 (CBS, 2021). Een reden voor de stijging van het aantal werkende 65-plussers zou de vergrijzing van Nederland kunnen zijn. Het aantal 65-plussers is vanaf 2003 namelijk met 1,2 miljoen toegenomen (CBS, 2021). Ook is deze stijging logischerwijs het grootst vanaf 2013, omdat vanaf dit jaar de AOW-leeftijd stapsgewijs is verhoogd (Wet verhoging AOW- en pensioenrichtleeftijd 2012). Het is dus niet zo dat al deze werkende 65-plussers ervoor hebben gekozen om door te werken na het bereiken van de AOW-leeftijd. Echter geldt wel dat de mensen die door willen werken na het bereiken van de AOW-leeftijd hier een bewuste keuze in maken; de standaard is namelijk 'gewoon' met pensioen gaan. Vrijwel iedereen die in loondienst werkt wordt namelijk automatisch ontslagen bij het bereiken van de AOW-leeftijd. Niet iedereen die wil doorwerken krijgt daar dan ook de mogelijkheid toe. Daarvoor moeten er dus werkgevers zijn die met hen in zee willen gaan (Oude Mulders, 2019)

Alleen het stijgen van de AOW-leeftijd verklaart echter niet waarom het aantal werkende 68-jarigen bijna is verdubbeld (CBS, 2021). In 2020 ligt de AOW-leeftijd rond de 67 en desondanks waren er een dergelijk 40.000 mensen van 68 jaar aan het werk tegenover 20.000 in 2003. Ook de 75-jarigen werken vaker door, in 2020 waren dit er 30.000 tegenover 3.000 in 2003. Mensen hoeven op deze leeftijden niet nóg langer te werken, maar kiezen hier bewust voor.

In de wetenschappelijke literatuur is veel informatie te vinden over doorwerken na het bereiken van de AOW-leeftijd. Zo is er onderzocht dat doorwerken populairder is bij mannen dan bij vrouwen en dat het vaker voorkomt bij mensen die werkzaam zijn in hogere functies (Solinge & Henkens, 2016). Ook is er onderzoek gedaan naar de effecten van langer doorwerken op de gezondheid (Wind et al., 2020). Daarnaast is onderzoek gedaan naar redenen voor de werkgevers om mensen te laten doorwerken na de AOW-leeftijd (Oude Mulders, 2019). Onderzoek naar de beweegredenen van ouderen zelf om langer door te werken is ook te vinden (Sewdas, 2018; Steenbeek, 2017;). Deze onderzoeken zijn echter

voornamelijk gebaseerd op interviews en enquêtes en missen statistische onderbouwing. Dit onderzoek doet een poging om onderzochte theorie statistisch te onderbouwen en is daarom een aanvulling op de bestaande literatuur.

Empirisch onderzoek naar langer doorwerken is van groot belang voor de maatschappij. Zoals eerder genoemd vergrijst Nederland. Het aantal jongeren neemt af terwijl het aantal ouderen verhoudingsgewijs toeneemt CBS, 2022). Dit zorgt voor problemen op de arbeidsmarkt en daarbuiten. In een artikel van NOS (2021) wordt dit tekort aan mensen zelfs een soort vervolgcrisis op de coronacrisis genoemd, omdat het economisch herstel hierdoor wordt belemmerd. Wanneer er geen personeel is, kunnen de gewenste diensten en producten immers niet geleverd worden. Een ander probleem van vergrijzing komt naar voren in een rapport van het Centraal Planbureau (2019). Hier wordt genoemd dat de overheid door de toenemende vergrijzing meer aan ouderen betaalt dan dat zij belasting betalen, terwijl werkenden meer moeten betalen dan dat ze ontvangen. Eén oplossing voor de problemen omtrent vergrijzing is er niet, maar het vergroten van de arbeidsparticipatie onder ouderen kan een goed begin zijn. Ouderen als remedie in plaats van deel van het probleem. Onderzoek naar langer doorwerken is dus nodig om te begrijpen wat mensen ertoe drijft. Op basis van dit onderzoek kan vervolgens beleid worden gemaakt.

In dit onderzoek wordt verder ingegaan op het fenomeen 'langer doorwerken'. De onderzoeksvraag is als volgt: *'Wat is de invloed van de fysieke belasting van het werk, het hebben van een partner en de verwachte inkomsten op het langer willen doorwerken na de AOW-leeftijd?'.* Een belangrijke opmerking hierbij is dat bij dit onderzoek met 'mensen' volwassenen met een leeftijd tussen 45 en 67 jaar worden bedoeld. Drie mogelijke verklaringen voor het langer willen doorwerken van deze mensen zijn zowel statistisch als theoretisch onderzocht. Als eerste wordt naar de invloed van de fysieke belasting van het werk gekeken. Zo wordt er onder andere verwacht dat mensen die fysiek zwaar werk verrichten niet langer door willen werken, omdat ze dit simpelweg fysiek gezien niet meer kunnen. Ten tweede wordt er gekeken naar de partnerstatus op pensioengerechtigde leeftijd. Verwacht wordt dat het hebben van een partner het doorwerken na de AOW-leeftijd kan ontmoedigen en dat het niet hebben van een partner juist motiveert om dit te doen. Ten derde wordt gekeken naar de inkomsten die verwacht worden wanneer iemand de AOW-leeftijd bereikt. Zo wordt verwacht dat wanneer deze inkomsten laag zijn, men eerder door wil werken na het bereiken van de AOW-leeftijd, bijvoorbeeld uit financiële noodzaak.

Theoretisch Kader

Door gebruik te maken van bestaande literatuur, is onderzocht of verschillende factoren een antwoord kunnen bieden op de vraag wat mensen ertoe drijft om door te willen werken na het bereiken van de AOW-leeftijd. De onderzochte factoren zijn: de fysieke belasting van het werk, de partnerstatus en de verwachte inkomsten na het bereiken van de AOW-leeftijd.

Fysieke Belasting van het werk

Een eerste mogelijke verklaring voor het langer doorwerken na het bereiken van de AOW-leeftijd heeft te maken met de fysieke belasting van het werk dat een persoon doet. Met fysiek belastend werk wordt werk bedoeld dat een effect heeft op het lichaam (CZ, 2019). Dit kan op verschillende manieren voorkomen, bijvoorbeeld door tillen, duwen, trekken, veel lopen en langdurig staan. Het is goed voor te stellen dat mensen die fysiek belastend werk verrichten niet langer door willen werken, omdat ze dit simpelweg fysiek gezien niet meer kunnen. Zo blijkt uit een onderzoek van Oude Mulders & Damman (2019) dat bij mensen die werkzaam zijn in de bouw het gat tussen de leeftijd waarop ze met pensioen kunnen en wanneer ze dat zouden willen, het grootst is. Dit gat bedraagt zo'n 2 jaar en 10 maanden. Volgens de onderzoekers is dit gat zo groot, omdat de werknemers door de fysieke zwaarte van hun werk wel graag met pensioen zouden willen, maar dit om financiële redenen nog niet kunnen. Uit hetzelfde onderzoek blijkt ook dat zorgen over het überhaupt halen van de (gestegen) AOW-leeftijd het grootst zijn bij oudere werknemers die gezondheidsklachten hebben en lichamelijk zwaar werk verrichten. Wanneer men zich zorgen maakt over het halen van de AOW-leeftijd, lijkt willen doorwerken hierna dan ook helemaal onwaarschijnlijk. Ook is het goed mogelijk dat mensen die fysiek zwaar werk verrichten een andere sociaaleconomische positie hebben dan mensen die een kantoorbaan hebben. In een onderzoek van De Wind et al. (2020) wordt genoemd dat mensen die fysiek zwaar werk verrichten vaker een lage sociaaleconomische positie hebben. Deze mensen zijn doorgaans door een kortere opleidingsperiode op jongere leeftijd gestart met werken en hebben dus ook al een langer arbeidsleven. Hierdoor wordt het nog langer willen doorwerken onwaarschijnlijk. Uit bovenstaande volgt de volgende hypothese:

Wanneer mensen geen fysiek belastend werk verrichten zullen zij vaker langer door willen werken na het bereiken van de AOW-leeftijd dan mensen die wel fysiek belastend werk verrichten (H1)

Partnerstatus

Een tweede mogelijke verklaring voor het langer willen doorwerken na het bereiken van de AOW-leeftijd is de partnerstatus. Het gaat er hier om of de respondent een partner heeft, dit kan zowel een gehuwde als ongehuwde partner zijn. Aan de ene hand kan een partner het doorwerken na de AOW-leeftijd aanmoedigen of juist ontmoedigen. Onderzoek van Eismann et al. (2020) laat zien dat pensioenbeslissingen niet alleen worden beïnvloed door factoren op het werk, maar ook door de situatie binnen het huishouden. Werknemers passen hun pensioenpreferenties en gedrag aan naar de wensen en preferenties van hun partner. Volgens Henkens et al. (2009) is een (huwelijks)partner die doorwerken aanmoedigt een van de belangrijkste impulsen voor een later moment van pensioneren. Ook blijkt uit onderzoek dat de pensionering van de ene partner, de andere partner kan beïnvloeden. Wanneer een van de partners met pensioen gaat of juist beslist om langer door te werken, is de kans groot dat de tweede partner hetzelfde doet (Eismann et al., 2020). Het is dus duidelijk dat partners elkaar beïnvloeden bij de keuze om door te werken na het bereiken van de AOW-leeftijd. Hiernaast kan het niet hebben van een partner ook invloed hebben op deze keuze. Zo blijkt uit een onderzoek van Dingemans & Dingemans (2016) dat vrouwen die geen partner (meer) hebben een veel grotere kans hebben om door te willen werken, in vergelijking met vrouwen die wel een partner hebben. Voor mannen is dit 'weduwnaar effect' echter niet gevonden (Dingemans; Dingemans, 2016). Ook kan het niet hebben van een partner een rol spelen in de mate van eenzaamheid. Uit onderzoek van Sewdas (2018) blijkt dat niet eenzaam willen zijn vaak een voorwaarde is voor ouderen om door te blijven werken na het bereiken van de AOW-leeftijd, omdat deze ouderen niet alleen thuis willen zitten. Het is goed voor te stellen dat mensen zonder partner meer waarde hechten aan de sociale contacten op het werk, omdat ze deze thuis minder hebben. Wanneer een persoon een partner heeft zullen eenzaamheid en de sociale contacten op het werk een minder grote rol spelen op het door willen werken na de AOW-leeftijd. Uit de genoemde informatie is de volgende hypothese af te leiden:

(H2) Wanneer mensen een partner hebben zullen zij minder langer door willen werken na het bereiken van de AOW-leeftijd, in vergelijking met mensen die geen partner hebben.

Verwachte Inkomsten

De laatste onderzochte verklaring voor het langer door willen werken betreft de financiële situatie. In dit onderzoek wordt gekeken naar de inkomsten die de respondenten verwachten na het bereiken van de AOW-leeftijd. Uit onderzoek van Bijlsma et al. (2017) blijkt dat vooral mensen met een laag pensioen vaker doorwerken na de AOW-gerechtigde leeftijd. Volgens Henkens et al. (2009) vergroten financiële prikkels de kans op langer doorwerken. In het onderzoek van Henkens is onderzocht dat wanneer de pensioenuitkering als percentage van het laatstverdiende loon lager is, men ook later met pensioen gaat. Ook stellen de onderzoekers dat het niet achteruit willen gaan in inkomen voor werknemers een grotere motivator is om langer door te werken dan bijvoorbeeld een financiële bonus die het doorwerken beloont. Uit een rapport van de grootste seniorenbond van Nederland (KBO-PCOB) blijkt dat een op de zes doorwerkende AOW-eenpersoonshuishoudens een relatief laag aanvullend overig inkomen heeft (zoals ouderdoms- of partnerpensioen) en hun inkomsten daarom aanvullen met een betaalde baan (Straatmeijer, 2020). Wanneer de verwachte inkomsten na het bereiken van de AOW-leeftijd laag zijn, kan doorwerken dus als financiële noodzaak worden gezien, omdat mensen hun huidige levensstandaard graag willen behouden of bijvoorbeeld nog schulden of hypotheek moeten aflossen. Bij het huidig pensioenbeleid wordt ook rekening gehouden met het willen behouden van de levensstandaard van de burgers. In de Perspectiefnota Toekomst Pensioenstelsel uit 2016 wordt genoemd dat het kabinet het van belang acht dat alle werkenden een toereikend pensioen opbouwen om bij pensionering geen grote terugval te hebben in bestedingsmogelijkheden (Knoef et al., 2017). Uit het bovenstaande is de volgende hypothese opgesteld:

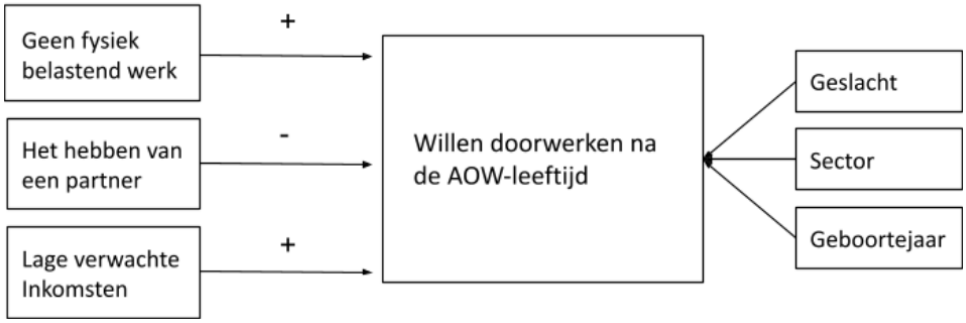
(H3) Wanneer mensen lage inkomsten verwachten na het bereiken van de AOW-leeftijd zullen zij vaker langer willen doorwerken dan wanneer hoge inkomsten worden verwacht.

Controlevariabelen

Om te onderzoeken of er daadwerkelijk een verband is tussen het langer willen doorwerken na de AOW-leeftijd en de verklaringsfactoren, worden controlevariabelen in het model opgenomen. Door deze controlevariabelen mee te wegen is het mogelijk om een zuiverder effect van de verklarende variabelen op het willen doorwerken na de AOW-leeftijd te schatten. In dit onderzoek worden *geslacht, sector* en *geboortejaar* als controlevariabelen gebruikt. Het is verstandig om te controleren voor geslacht, omdat

uit de literatuur bekend is dat mannen vaker doorwerken na de AOW-leeftijd dan vrouwen. Ook is het verstandig om te controleren voor sector, omdat men zich kan afvragen of het doorwerken niet alleen in een bepaalde arbeidssector voorkomt. Ten slotte is het ook verstandig om te controleren voor geboortejaar, omdat het zo kan zijn dat oudere generaties anders tegen doorwerken aankijken dan jongere generaties. Zo kan worden uitgesloten dat een verschil in houding tegenover langer doorwerken na de AOW-leeftijd met de jaren bestaat.

Het bovenstaande kan overzichtelijk worden weergegeven in één conceptueel model:



Figuur 1: Conceptueel model

Methoden

Dataset en Steekproef

Er is voor dit onderzoek gebruikt gemaakt van data uit het LISS-panel (Longitudinal Internet studies for the Social Sciences). Deze data bevat een steekproef uit de Nederlandse samenleving en betreft ongeveer 5000 huishoudens, met in totaal 7500 individuen. Deze huishoudens zijn aselect geselecteerd op basis van data van het Centraal Bureau voor de Statistiek. Respondenten kunnen zichzelf dus niet opgeven om mee te doen, waardoor een 'zelfselectie bias' wordt voorkomen. Volgens de website van het LISS-panel zorgen voornoemde eigenschappen ervoor dat de samenstelling van de data representatief is voor de Nederlandse samenleving (CentERdata, 2021). De respondenten vullen regelmatig vragenlijsten in over vaste onderwerpen en krijgen hier €15,00 per uur voor betaald.

Voor het doen van dit onderzoek zijn drie datasets uit het LISS-panel samengevoegd. De eerste is genaamd: 'Work & Schooling' en de data hierin is verzameld in april-mei 2020. Deze dataset bevat een reeks jaarlijks herhaalde vragen omtrent scholing, werk en pensioen(plannen). De tweede dataset is genaamd: 'langer doorwerken en pensioenen', deze data is verzameld in februari 2020. Er wordt hier dieper ingegaan op enkele vragen rondom langer doorwerken en pensioenplanning. Ook is er een dataset met achtergrondvariabelen toegevoegd. Deze dataset bevat de meest recente informatie over achtergrondkenmerken van de respondenten. Hierbij gaat het vooral om demografische en sociaaleconomische variabelen. De data hiervan is verkregen in mei 2020.

Voor de dataset 'langer doorwerken en pensioenen' zijn alleen mensen van boven de 40 jaar, die nog niet met pensioen waren, ondervraagd. Ook zijn bepaalde vragen omtrent gewenste en verwachte pensioenleeftijd alleen gesteld aan mensen vanaf leeftijd 45. Dit zorgt ervoor dat de uiteindelijke steekproef een stuk minder respondenten bevat dan het totaal aantal in de dataset (N = 5758). Voor dit onderzoek zijn ook de respondenten die geboren zijn vóór 1953 uit de data verwijderd. Dit omdat wordt verwacht dat mensen die de AOW-leeftijd al hebben bereikt, op een andere manier uitspraken doen over het langer willen doorwerken dan mensen die deze leeftijd nog niet hebben bereikt. Ook is een aantal cases met missende waarden uit de dataset verwijderd. De uiteindelijke steekproef bedraagt N=1311.

Operationalisaties

Voordat met de logistische regressie begonnen kon worden, dienen de variabelen eerst geoptimaliseerd te worden, de details hiervan staan in bijlage 1. De afhankelijke variabele van dit onderzoek over het willen doorwerken na het bereiken van de AOW-leeftijd is een zelfgeconstrueerde dummyvariabele. Deze variabele is geconstrueerd door de gewenste pensioenleeftijd - 67 (de AOW gerechtigde leeftijd in 2020) te doen. Wanneer dit verschil 1 of groter was is dit gecodeerd als 1 en de rest is gecodeerd als 0, waardoor 1 staat voor willen doorwerken en 0 voor dit niet willen. Dit betekent dat alleen mensen die minstens een jaar door willen werken na het bereiken van de AOW-leeftijd als zodanig worden geïdentificeerd. Dit is zo gedaan omdat in de gebruikte dataset alleen antwoord op de vraag over gewenste pensioenleeftijd is gegeven in hele jaren. Het is met de gebruikte data dan ook niet mogelijk om de grens van 'langer doorwerken' te leggen op bijvoorbeeld 67 jaar en 1 maand. Uiteindelijk hebben 1311 mensen met een leeftijd tussen 45 en 67 jaar, deze vraag beantwoord, waarvan 76 mensen hebben gezegd door te willen werken na het bereiken van de AOW-leeftijd (5,8%).

De variabele over de mate van fysieke belasting van het werk bestaat uit vier samengevoegde schaalvariabelen. De variabelen die zijn samengevoegd gaan over of het werk fysiek veeleisend is, er zware objecten getild moeten worden, er geknield of gebukt moet worden en of het werk vermoeiend is. De respondenten konden kiezen uit drie antwoordcategorieën: 1 'vaak', 2 'soms' en 3 'nooit'. Voordat de variabelen zijn samengevoegd is gekeken naar de samenhang tussen de verschillende variabelen door gebruik te maken van Cronbach's Alpha. Deze bleek 0,85 te zijn, wat zich vertaalt in een goede mate van samenhang. Vervolgens zijn de vier variabelen samen tot één nieuwe variabele gecodeerd die het gemiddelde van de antwoorden neemt. De schaal van de variabele is vervolgens gespiegeld zodat men deze intuïtiever kan interpreteren (1: 'nooit', 2: 'soms', 3 'vaak'). Een hogere score op deze variabele betekent dus dat mensen vaker fysiek belastend werk verrichten.

De tweede verklarende variabele kijkt naar de partnerstatus van de respondent. Het is een dummyvariabele met 0 als 'nee' (geen partner) en 1 als 'ja' (wel een partner). De variabele is verder niet geoperationaliseerd.

De laatste verklarende variabele gaat over de verwachte netto inkomsten per maand (in hele getallen) van de respondent wanneer deze de AOW-leeftijd bereikt. Omdat de verwachte inkomsten op

een erg brede schaal liggen, zijn alle inkomsten door 10.000 gedeeld, zodat het interpreteren van de resultaten uit de regressieanalyse makkelijker is.

Er zijn voor dit onderzoek ook drie controlevariabelen gebruikt. De eerste hiervan is Geslacht en heeft de antwoordcategorieën 1 'Man' en 2 'Vrouw'. De tweede controlevariabele is de sector waarin iemand werkzaam is. Voor deze variabele kon gekozen worden uit 14 verschillende sectoren of de categorie 'overig'. Voor dit onderzoek is deze variabele verder verdeeld in vier groepen: 'Bouw en Industrie', ' Diensten en Handel', 'Publieke 'Sector' en 'Overig'. Deze variabelen zijn vervolgens tot 3 dummyvariabelen gemaakt, waarbij 0 'niet binnen de categorie' en 1 'wel binnen de categorie' aangeeft. De groep 'Bouw en Industrie' is als referentiecategorie gebruikt. De laatste controlevariabele is die van geboortejaar. Deze variabele is verder niet geoperationaliseerd.

Analyseplan

De opgestelde hypothesen in dit onderzoek worden aan de hand van een multipele logistische regressieanalyse getoetst. Dit omdat de afhankelijke variabele over het willen doorwerken een dichotome variabele is (wel of niet willen doorwerken) en er meerdere verklarende factoren onderzocht worden. Ten eerste worden de beschrijvende statistieken van de variabelen gegeven, ten tweede worden de assumpties die horen bij een logistische regressie behandeld en tenslotte worden de resultaten van de analyse gepresenteerd. Als eerste stap in deze analyse wordt een model geschat met de afhankelijke variabelen en alle controlevariabelen. Als tweede stap zal gekeken worden naar het effect van elke 'losse' onafhankelijke variabele op de afhankelijke variabele, zonder controlevariabelen op te nemen in het model. Vervolgens wordt in het laatste model het effect van de drie onafhankelijke variabelen op de afhankelijke variabele geschat, waarbij wel gecontroleerd is voor de controle variabelen. Voor de onafhankelijke variabelen zal gekeken worden of het effect op de afhankelijke variabele verandert wanneer wordt gecontroleerd voor de andere (controle) variabelen. Voor het toetsen van de vooropgestelde hypothesen wordt zowel naar de modellen met de 'losse' onafhankelijke variabelen als het 'complete' model gekeken. Er zal gekeken worden naar het verschil in effect van de onafhankelijke variabelen op de afhankelijke variabele wanneer wel en wanneer niet gecontroleerd is voor andere (controle) variabelen. Om vervolgens te kijken of het toevoegen van de variabelen zorgt voor meer verklaarde variantie in het model, wordt gekeken naar de -2 log likelihood en het

bijbehorende significantieniveau. Ook zal de Hosmer and Lemeshow Goodness-of-Fit Test gebruikt worden om te kijken of de gebruikte data bij het uiteindelijke model past.

Resultaten

Beschrijvende statistieken

Tabel 1: Beschrijving van de in de analyse opgenomen variabelen: gemiddelde (standaarddeviatie), minimum- en maximumwaarde, scheefheid, kurtosis en totaal aantal respondenten

Variabele	Gemiddelde (standaarddeviatie)[a]	Minimum	Maximum	Scheefheid	Kurtosis	N totaal
Langer willen doorwerken	94,2% nee					1311
(0 = nee, 1= ja)	5,8% ja					
Fysieke Belasting Werk	52,5% nooit					1311
(3=nooit, 2= soms, 1= vaak)	32,6% soms					
	17,4% vaak					
Partner	23,8% nee					1311
(0=nee, 1= ja)	76,2% ja					
Verwachte Inkomsten	3690,17	0	102.000	7,40	66,69	1311
	(8120,76)					
Geslacht	52% man					1311
(0= man, 1 vrouw)	48% vrouw					
Sector	17,8% Bouw & Industrie					1311
	25,7% Diensten & Handel					
	45,3% Publieke Sector					
	11,1% Overig					
Geboortejaar	1964	1953	1975	-,015	-1,056	1311
	(5,861)					

a Bij nominale variabelen is de frequentieverdeling vermeld in percentages.

In tabel 1 zijn de beschrijvende statistieken van alle variabelen te zien. In deze tabel valt een paar dingen op. Zo is te zien dat maar een klein deel van de respondenten aangeeft door te willen werken na het bereiken van de AOW-leeftijd (5,8%) ten opzichte van het deel dat aangeeft dit niet te willen (94,2%). Ook valt op dat het grootste deel van de respondenten (52,2%) nooit fysiek belastend werk verricht. Respondenten die vaak fysiek belastend werk verrichten komen het minst vaak voor, maar het betreft toch meer dan een kwart (17,4%). De meeste respondenten geven aan een partner te hebben (76,2%). De gemiddelde verwachte inkomsten van de respondenten na het bereiken van de AOW-leeftijd, bedraagt €3690,17 per persoon per maand. Echter, is ook een grote standaarddeviatie te zien, omdat er een aantal hogere inkomsten gerapporteerd zijn. Zo zijn de maximum verwachte inkomsten €102.000. In bijlage 2 is te zien dat de mediaan van deze variabele 2200 bedraagt en het dus om een rechtsscheve verdeling gaat. Verder is de man/vrouw verhouding onder de respondenten vrij gelijk verdeeld: met 52,2% zijn er net iets meer mannen aanwezig in de steekproef. De meeste respondenten zijn werkzaam in de Publieke Sector (44,7%) en zijn gemiddeld ongeveer 57 jaar oud tijdens het invullen van de vragenlijst. De oudste respondent is 67 jaar oud en de jongste respondent 45.

Tabel 2: samenhangende maten van de afhankelijke en onafhankelijke variabelen in het onderzoeksmodel

	1.	2.	3.	4.
1. Willen doorwerken	-			
2. Fysieke Belasting	*,116[b]	-		
3. Partner	,007[b]	,093[b]	-	
4. Verwachte Inkomsten	**,123[a]	,140[c]	,025[a]	-

* significant op 0,05; ** significant op 0,01; [a] Pearson correlatie; [b] Cramer's V; [c] correlatie op basis van ANOVA

In tabel 2 is de mate van samenhang gegeven tussen de afhankelijke en onafhankelijke variabelen in het model. Over het algemeen is de mate van samenhang vrij zwak. De voor dit onderzoek relevante verbanden worden hier verder besproken. Zoals in tabel 2 te zien zijn twee van de drie correlaties tussen de afhankelijke en onafhankelijke variabelen significant getoetst. Zo is de samenhang tussen Fysieke Belasting en Willen Doorwerken significant, maar zwak (χ^2 (8) = 17,06; p= ,024). Fysieke Belasting en willen doorwerken lijken dus enigszins positief samen te hangen. Mensen die aangeven door te willen werken verrichten het vaakste werk dat niet fysiek belastend is. Verder kan er gesproken worden van een significante samenhang tussen de variabelen 'Verwachte Inkomsten' en 'Willen Doorwerken'. Deze samenhang tussen Verwacht Inkomsten en Willen Doorwerken is positief en maar ook weer erg zwak (r=,123; p= ,002). Inhoudelijk betekent dit dat mensen die willen doorwerken hogere inkomsten verwachten. Tabel 2 laat ook zien dat de variabelen 'Willen Doorwerken' en 'Partner' niet significant samenhangen (r=,064; p= ,800). Meer details betreffende de samenhang tussen de variabelen is te vinden in bijlage 2.

Assumptietoetsing

Bij het doen van een logistische regressieanalyse hoort een aantal assumpties. Één assumptie is dat de waarnemingen in de steekproef onafhankelijk van elkaar zijn. Dit houdt in dat de steekproefeenheden (respondenten) onafhankelijk van elkaar zijn getrokken en er geen samenhang tussen de geobserveerde scores van de respondenten bestaat. Wanneer er geen sprake is van onafhankelijkheid zullen de geschatte standaardfouten in het algemeen kleiner zijn dan ze in werkelijkheid moeten zijn. Hierdoor kan het zo zijn dat de betrouwbaarheidsintervallen smaller worden geschat dan dat ze eigenlijk zijn. Dit kan op zijn beurt weer zorgen voor het concluderen van een significant effect, terwijl het effect eigenlijk niet significant is (Type 2 fout). De steekproefeenheden van dit onderzoek zijn aselect geselecteerd uit data van het Centraal Bureau voor de Statistiek (CentERdata, 2021). De steekproefeenheden zijn dus onafhankelijk van elkaar geselecteerd, maar over of er geen samenhang tussen de scores van de respondenten bestaat valt te twisten. Er is voor dit onderzoek namelijk gebruik gemaakt van data afkomstig uit een huishoudenregister. Het is niet uit te sluiten dat respondenten die samenwonen elkaar beïnvloeden in de antwoorden die ze geven op vragenlijsten. Data die afkomstig is van mensen uit hetzelfde huishouden is echter niet vergelijkbaar met data uit bijvoorbeeld dezelfde klas, een klassiek voorbeeld van (perfect) afhankelijke waarnemingen. Voor mensen uit hetzelfde huishouden hoeft niet te

gelden dat ze de vragenlijst op dezelfde plek en dezelfde tijd hebben ingevuld. Verdere details over het aantal respondenten dat hetzelfde huishouden deelt ontbrak helaas in de gebruikte dataset [1]. Er kan voor dit onderzoek niet met zekerheid gezegd worden dat de steekproefeenheden perfect onafhankelijk zijn, maar ook niet dat deze per se afhankelijk van elkaar zijn. De assumptie van onafhankelijke waarnemingen kan dan ook als gedeeltelijk geschonden worden beschouwd. Met de conclusies van dit onderzoek moet dan ook voorzichtig worden omgegaan.

Naast assumpties is ook gekeken naar de samenhang tussen de onafhankelijke variabelen door middel van VIF-scores. Een te grote samenhang, ook wel multicollineariteit genoemd, zorgt ervoor dat het verschil in impact van de variabelen op de afhankelijke variabele niet goed meer te zien is. Het is door een grotere standaardfout erg lastig om te schatten wat de variabelen afzonderlijk van elkaar betekenen. In B3.2 is te zien dat er geen sprake is van te grote samenhang tussen de variabelen. Multicollineariteit is dus geen probleem.

Ten slotte is het van belang om te kijken naar mogelijke uitschieters, omdat deze de data erg kunnen vertekenen. Hiervoor is gekeken naar de leverage waarden. De leverage geeft aan in welke mate een bepaald punt bij wijze van spreken 'aan de lijn trekt' en hoe groot het effect van dat punt op de schattingen is. In B3.3 zijn door naar de leverage waarden te kijken, duidelijk twee uitbijters te zien. Er is gekeken naar het effect van deze twee punten op de resultaten van dit onderzoek. Hier is geen significante invloed geconstateerd. Het gaat immers ook om twee 'gevallen' uit een steekproef van 1311 respondenten. Meer informatie over de gegeven VIF en Leverage waarden is te vinden in bijlage 3.

[1] Een aantal medestudenten bleken wel variabelen in hun dataset te hebben met meer details over de huishoudens. Deze zijn door mijn (eerste) begeleider echter uit de gegeven dataset verwijderd.

Hypothesetoetsing

Tabel 3: Parameterschattingen voor onafhankelijke en controlevariabelen

	Model 1			Model 2			Model 3			Model 4			Model 5		
	b(SE)	*P*	*OR*	*b(SE)*	*p*	*OR*	*b(SE)*	*p*	*OR*	*b(SE)*	*p*	*OR*	*b(SE)*	*p*	*OR*
Constante	90,653 (68,48)	,186	2,34	-1,674 (,51)	,000	,188	-2,395 (,20)	<0,001	,091	-2,690 (,89)	<0,001	0,068	98,81 (70,3)	,160	8,18
Geslacht (0=M; 1=V)	-,698 (,46)	,129	,498										-,649 (,467)	,165	,523
Overig	-,195 (,87)	,823	,823										,150 (,890)	,866	,861
Diensten & Handel	,618 (,55)	,258	1,885										,464 (,563)	,410	1,590
Publieke Sector	-,543 (,59)	,364	,581										-,561 (,608)	,356	,571
Geboorte-jaar	-,047 (,04)	,178	,954										-,050 (,0,36)	,163	,951
Fysieke Belasting (1=nooit, 2= soms & 3= vaak)				-,680 (,24)	,005	,506							-1,533 (,579)	0,008	,216216
Partner (0=Nee; 1= Ja)							-,069 (,27)	0,800	,933				-,595 (,438)	,175	,552
Verwachte Inkomsten										0,251 (,141)	,075	1,285	,138 (,152)	,365	1,148
Deviance	211,173			571,42			680,31			221,39			199,59		
X^2	12,623	<0,05		8,954	0,003		0,064	0,801		2,544	0,111		24,348	,002	
Df	5			1			1			1			8		

In tabel 3 zijn vijf modellen te zien. Hoe de data hiervan precies is verkregen is te vinden in bijlage 2. In het eerste model zijn de verwachte effecten van de controlevariabelen: Geslacht, Geboortejaar en de variabelen die horen bij Sector op het langer willen doorwerken na de AOW-leeftijd, te zien. Geen van de controlevariabelen heeft een significant effect heeft op langer willen doorwerken. Wel zijn er relatief sterke effecten te zien zoals de variabele Geslacht die een vrij sterk negatief effect op langer willen

doorwerken laat zien (b=-,698; p= ,129). Vrouwen zijn minder geneigd om door te willen werken na het bereiken van de AOW-leeftijd in vergelijking met mannen. Dit komt overeen met de vooropgestelde verwachting over het effect van Geslacht op het langer willen doorwerken, maar is niet significant getoetst. Ook de dummy variabele Diensten & Handel laat een vrij sterk positief effect (b=,681; p-,258) zien. Werken binnen deze sector gaat vaker samen met langer willen doorwerken na de AOW-leeftijd in vergelijking met de referentiecategorie 'Bouw en Industrie'. Werken binnen de Publieke Sector gaat niet vaker samen met het langer willen doorwerken in vergelijking met de referentiecategorie(b=-,543; p=,364). Dat er tussen de sectoren verschillen zijn in wel of niet willen doorwerken, komt overeen met de vooropgestelde verwachting over het effect van Sector op het langer willen doorwerken. Echter zijn de effecten van de verschillende sectoren zoals eerder genoemd niet significant en kunnen dus niet worden gegeneraliseerd naar de populatie. In welk jaar iemand geboren is blijkt een verwaarloosbaar effect te hebben op het willen doorwerken na het bereiken van de AOW-leeftijd. Dit spreekt de vooropgestelde verwachting over het effect van Geboortejaar op het langer willen doorwerken tegen, maar ook dit effect is niet significant getoetst.

In model 2 is het effect van de variabele Fysieke Belasting op het door willen werken na het bereiken van de AOW-leeftijd te zien. Uit de resultaten blijkt dat vaker fysiek belastend werk verrichten een sterk negatief effect heeft op het willen doorwerken na het bereiken van de AOW-leeftijd (b=,680; p=0,05). De odds-ratio van ,506 laat zien dat de odds op het willen doorwerken na de AOW-leeftijd met ongeveer de helft kleiner wordt wanneer werk fysiek zwaarder wordt op de schaal. Deze gegevens ondersteunen Hypothese 1.

In model 3 is het effect van de variabele Partner op het door willen werken na het bereiken van de AOW-leeftijd te zien. Er is een negatief effect van het hebben van een partner op het willen doorwerken te zien (b=-,069; p= ,800). Een odds-ratio van ,933 laat zien dat de odds op het willen met doorwerken na de AOW-leeftijd met 6,7% afneemt wanneer mensen een partner hebben. De odds op het willen doorwerken wordt dus kleiner wanneer iemand een partner heeft. Deze gegevens laten zien dat dit effect overeenstemt met Hypothese 2, maar dat dit statistisch gezien niet significant is getoetst. Ook neemt de Deviance bij model 3 erg weinig af en is niet significant (p=0,801; df= 1). Dit laat zien dat het toevoegen van de variabele Partner er niet voor zorgt dat een model beter bij de data past.

In model 4 is het effect van de variabele Verwachte Inkomsten op het door willen werken na het bereiken van de AOW-leeftijd te zien. Er is een niet significant, redelijk sterk positief effect van de

inkomsten die iemand verwacht na het bereiken van de AOW-leeftijd op het willen doorwerken na deze leeftijd te zien (b=,251; p= 0,075). Een odds-ratio van 1,285 laat zien dat bij mensen met hogere verwachte inkomsten de odds op het willen werken na het bereiken van de AOW-leeftijd met meer dan een kwart toeneemt. Dit effect stemt niet overeen met vooropgestelde hypothese 3. Het effect blijkt andersom te werken, hogere verwachte inkomsten vergroten de kans dat iemand door wil werken na het bereiken van de AOW-leeftijd. Dit 'omgekeerde' effect is echter, zoals eerder genoemd, niet significant getoetst. De afname in Deviance (p=0,111; df=1), laat vervolgens ook zien dat het toevoegen van de variabele Verwacht Inkomsten er niet voor zorgt dat het model beter bij de data past.

In de modellen 2 tot en met 4 is gekeken naar het effect van de onafhankelijke variabelen op de afhankelijke variabelen zonder te controleren voor elkaar of voor de controlevariabelen. In model 5 is hier wel voor gecontroleerd. Wat meteen opvalt, is het effect van Fysieke Belasting van werk op het langer willen doorwerken na de AOW-leeftijd. Dit effect is sterk, negatief en significant (b= -1,533; p-0,008). Hypothese 1 wordt dus ondersteund door de resultaten, zelfs wanneer er wordt gecontroleerd voor de controle en andere variabelen. De andere twee onafhankelijke variabelen hebben net als in de 'losse' modellen geen significant effect op het langer willen doorwerken na de AOW-leeftijd.

Modelevaluatie

Een belangrijk onderdeel om te beoordelen hoe goed een model is, is onderaan de tabel te vinden. De waarde in deviance, met de bijbehorende chi-ckwadraat toets, geeft een maat voor hoe slecht een model is. Een afname in deviance geeft aan dat het model 'minder slecht' wordt en dus beter bij de data past. Bij model 1 is een significante afname van de deviance met 12,623 (p=<0,05; df=5) te zien wanneer de controlevariabelen aan het model worden toegevoegd. Het toevoegen van de controlevariabelen zorgt ervoor dat het model goed bij de data past en dit model betere voorspellingen kan maken. Ook bij model 2 neemt de deviance significant af met 8,954 (p=,003; df= 1) en laat dus zien dat het toevoegen van de variabele Fysieke Belasting er voor zorgt dat het model beter bij de data past in vergelijking met een model zonder variabelen. Bij model 3 neemt de deviance erg weinig af en is de afname niet significant (p=0,064; df= 1). Dit laat zien dat het toevoegen van de variabele Partner er niet voor zorgt dat het model beter bij de data past. Ditzelfde geldt voor model 4, de niet significante afname in deviance (p=0,111; df=1), laat zien dat het toevoegen van de variabele Verwacht Inkomsten er niet voor zorgt dat het model beter bij de data past. Uiteindelijk heeft het beste model de laagste deviance, omdat dit model de minste onverklaarde variantie bevat. De deviance is in model 5 het laagst van alle modellen (X^2= 24,348; p= ,002).

Inhoudelijk betekent dit dat het toevoegen van de onafhankelijke variabelen en de controlevariabelen dus helpt om de afhankelijke variabele te verklaren. Onderaan model 5 is de afname in deviance te zien als je model 5 toetst tegen model 1 (X^2=24,348; p=,002). Er is sprake van een significante afname van 23,348 in Deviance bij model 5 (p=,002; df=8). Met andere woorden betekent dit dat het toevoegen van de onafhankelijke variabelen aan het model waar eerst alleen de controlevariabelen in zaten, ervoor zorgt dat het model nog beter is geworden. Het toevoegen van alle variabelen en daarvoor controleren zorgt voor een model met de minste onverklaarde variantie en is dus het meest geschikt om de afhankelijke variabele te verklaren.

Een andere manier om te kijken of het model goed bij de data past is de Hosmer-Lemeshow test. Hierbij wordt een chi-kwadraat waarde berekend, gebaseerd op de geobserveerde en verwachte frequenties. Een niet-statistische significante uitkomst van deze test geeft aan dat het model goed past, omdat resultaten uit het model dan niet statistisch significant verschillen van de geobserveerde waardes. Bij model 5, dus het complete model, is er geen bewijs gevonden voor het significant verschillen van de verwachte kansen op willen doorwerken ten opzichte van de geobserveerde kansen (χ^2 (8) = 3,769; p =

,877). Door te kijken naar de Hosmer Leweshow toets wordt nogmaals bevestigd dat model 5 goed bij de data past.

Conclusie en Discussie

In dit onderzoek is gezocht naar het antwoord op de vraag *'Wat is de invloed van de fysieke belasting van het werk, het hebben van een partner en de verwachte inkomsten op het langer willen doorwerken na de AOW-leeftijd?'* Meer inzicht in mogelijke beweegredenen om door te willen werken na het bereiken van de AOW-leeftijd is van maatschappelijk belang. Tekorten op de arbeidsmarkt door onder andere de vergrijzing van Nederland vormen een actueel probleem. De steeds groter wordende groep ouderen stimuleren om langer door te willen werken na het bereiken van de AOW-leeftijd kan een manier zijn om dit probleem te verkleinen. Het is van belang dat de beweegredenen achter het langer willen doorwerken bekend worden, zodat de werkgevers maar ook de overheid hierop kunnen inspelen. Dit onderzoek vult bestaande wetenschappelijke literatuur over het langer willen doorwerken aan, dat grotendeels is gebaseerd op kwalitatief onderzoek, door middel van statistische onderbouwing.

Om de onderzoeksvraag te beantwoorden is een drietal hypothesen geformuleerd. De eerste hypothese stelt dat wanneer iemand werk verricht dat niet fysiek belastend is, dit ervoor zorgt dat mensen vaker langer door willen werken na het bereiken van de AOW-leeftijd. De tweede hypothese gaat over het wel of niet hebben van een partner en stelt dat mensen die een partner hebben minder vaak langer willen doorwerken na het bereiken van de AOW-leeftijd, in vergelijking met mensen zonder partner. Tenslotte stelt de derde hypothese dat wanneer mensen lage inkomsten verwachten na het bereiken van de AOW-leeftijd, zij vaker willen doorwerken na het bereiken van deze leeftijd.

Uit de analyse volgt een sterke ondersteuning voor de eerste hypothese. Er is in eerste instantie gekeken naar de invloed van fysieke belasting van werk op het langer willen doorwerken na de AOW-leeftijd zonder te controleren voor andere variabelen. Hieruit bleek dat de mate van fysieke belasting een vrij grote invloed heeft op het langer willen doorwerken na de AOW-leeftijd. Ook wanneer werd gecontroleerd voor andere variabelen kwam een significant en nog sterkere invloed naar voren. Er kan dus gesteld worden dat wanneer iemand fysiek minder zwaar werk verricht de kans groter wordt dat deze persoon langer door wil werken na het bereiken van de AOW-leeftijd.

Voor de tweede hypothese is geen ondersteuning gevonden. Er is een te verwaarlozen verband gevonden tussen het hebben van een partner en het door willen werken na het bereiken van de AOW-leeftijd, dat overeenstemt met de vooropgestelde hypothese. Dit verband werd weliswaar sterker wanneer gecontroleerd werd voor andere variabelen, maar is erg zwak en niet significant en kan dus niet

naar de populatie kan worden gegeneraliseerd.

Voor de derde hypothese is ook geen ondersteuning gevonden. Er kan wel gesproken worden van een verband tussen verwachte inkomsten en langer willen doorwerken, maar dit verband stemt niet overeen met de hypothese. Mensen die hogere inkomsten verwachten na het bereiken van de AOW-leeftijd willen langer doorwerken en niet andersom. Dit effect was aanwezig in het model zonder én met controle op andere variabelen, maar in beide gevallen niet significant getoetst.

Een verklaring voor het 'omgekeerde' effect van de verwachte inkomsten zit hem mogelijk in de werktevredenheid. In een artikel van Judge et al. (2010) is onderzoek gedaan naar het verband tussen een hoger inkomen en de mate van werktevredenheid. Uit het onderzoek blijkt een positief verband tussen het inkomen en de mate van werktevredenheid. Het is goed mogelijk dat mensen met een hogere werktevredenheid langer willen doorwerken dan mensen met een lage werktevredenheid. Het effect van werktevredenheid op langer door willen werken na het bereiken van de AOW-leeftijd kan een interessante factor zijn voor vervolgonderzoek.

Een verklaring voor het niet significante effect van de verwachte inkomsten kan mogelijk komen door een response bias. Het kan goed zijn dat respondenten vragen over hun inkomsten als 'ongemakkelijk' ervaren en deze niet naar waarheid invullen. Ook kan het goed zijn dat de respondenten nog geen goed beeld hebben van hun inkomsten en dus een 'vage' schatting hebben gegeven. Door deze vage schattingen is er mogelijk 'ruis' ontstaan die ervoor zorgt dat significante verbanden moeilijker te vinden zijn. Voor vervolgonderzoek is het dan ook aan te raden om gebruik te maken van een variabele over de verwachte inkomsten die gebruikt maakt van een schaal. Een dergelijke variabele is aanwezig in de gebruikte dataset, maar deze betreft slechts 20 valide cases en is daarom niet gebruikt.

Bovenstaande informatie kan kort worden samengevat: Soms of niet fysiek belastend werk verrichten drijft mensen ertoe om langer door te werken na het bereiken van de AOW-leeftijd. Het hebben van een partner en het verwachten van lage inkomsten blijkt geen significant effect te hebben op het langer willen doorwerken na het bereiken van de AOW-leeftijd.

Over dit onderzoek is een aantal dingen op te merken. Zo is de variabele over door willen werken na de AOW-leeftijd gecreëerd met een aangenomen pensioenleeftijd van 67 jaar. Deze pensioengerechtigde leeftijd geldt echter niet voor iedereen. Zo kan het van de geboortemaand afhangen wanneer iemand pensioengerechtigd is. Iemand die geboren is na 31 augustus 1953 en voor 1 september 1954, bereikt bijvoorbeeld vanaf 66 jaar en vier maanden de AOW-leeftijd (Ministerie van

Algemene Zaken, 2021). In dit onderzoek is hier geen rekening mee gehouden, omdat de dataset geen informatie over geboortemaanden bevat. Voor vervolgonderzoek is het wellicht interessant om dit verschil in pensioengerechtigde leeftijd per geboortemaand wel mee te nemen. Door dit verschil niet mee te nemen zou het namelijk kunnen dat 'ruis' ontstaat bij de resultaten die ervoor zorgt dat significante verbanden moeilijker te vinden zijn. Een ander punt betreffende de variabele over het willen doorwerken zit hem in het feit dat respondenten alleen als 'doorwerkers' zijn gecodeerd wanneer is aangegeven dat deze minimaal een jaar of langer door willen werken. Het is echter goed mogelijk dat respondenten slechts een paar maanden of driekwart jaar langer door willen werken. Dit is zo gedaan omdat in de gebruikte dataset alleen antwoord op de vraag over gewenste pensioenleeftijd is gegeven in hele jaren. Het is met de gebruikte data dan ook niet mogelijk om de grens van 'langer doorwerken' te leggen op bijvoorbeeld 67 jaar en 1 en 5 maanden. Voor vervolgonderzoek wordt aangeraden om deze respondenten wel mee te nemen, omdat dit een meer realistisch beeld geeft van het aantal mensen dat door wil werken na het bereiken van de AOW-leeftijd in Nederland.

Ook moet niet vergeten worden dat voor dit onderzoek aan een van de belangrijkste assumpties die horen bij een logistische regressie, namelijk die van onafhankelijke waarnemingen, slechts gedeeltelijk wordt voldaan. Aan de ene kant is de steekproef die grotendeels bestaat uit huishoudens willekeurig geselecteerd, maar aan de andere kant kunnen respondenten binnen één huishouden elkaar beïnvloeden in de gegeven antwoorden. Er kan voor dit onderzoek niet met zekerheid gezegd worden dat de steekproefeenheden perfect onafhankelijk zijn, maar ook niet dat deze per se afhankelijk van elkaar zijn. Met de conclusies van dit onderzoek moet dan ook voorzichtig worden omgegaan.

Voor het doen van vervolgonderzoek is een aantal suggesties gedaan. Zo dient werktevredenheid meegenomen te worden in verder onderzoek naar het effect van verwachte inkomsten op het langer willen doorwerken en wordt ook aangeraden om gebruik te maken van een schaal voor het meten van de variabele over de verwachte inkomsten. Ook wordt voor toekomstig onderzoek aangeraden om naar verschil in pensioengerechtigde leeftijd per geboortemaand te kijken en ook de respondenten die niet een heel jaar, maar een aantal maanden door willen werken mee te nemen. Ten slotte is het om de resultaten van vervolgonderzoek te kunnen generaliseren naar de hele populatie van groot belang dat aan de assumptie van onafhankelijke waarnemingen wordt voldaan.

Dit onderzoek kan gezien worden als een opstap voor verder statistisch onderzoek naar de beweegredenen achter het willen doorwerken na de AOW-leeftijd. Kwalitatief onderzoek naar dit

onderwerp bestaat al in grote mate, maar statistische resultaten dienen ook in acht genomen te worden. Deze resultaten kunnen gebruikt worden om beleid te vormen dat poogt mensen rond de AOW-leeftijd te stimuleren om langer door te werken. Ook dit onderzoek biedt nuttige informatie dat gebruikt kan worden om beleidsaanbevelingen te doen. Zo blijkt duidelijk dat mensen die fysiek belastend werk doen, minder bereid zijn om langer door te willen werken. Dit is ook logisch, dergelijk werk put het menselijk lichaam uit. Voor toekomstig beleid dat als doel heeft mensen te motiveren om langer door te werken wordt dan ook aangeraden om hierbij te focussen op mensen die geen of soms fysiek belastend werk verrichten. Echter dient de toch grote groep mensen die fysiek belastend werk verricht niet vergeten te worden. Er moet aandacht worden besteed aan mogelijkheden om het werk voor ouderen minder fysiek belastend te maken. Een goed voorbeeld hiervan is te zien bij Afvalverwerking Roermond dat vanaf 2003 een zelfstandig beleid voert om de inzetbaarheid van oudere werknemers te verhogen. Ergonomische aanpassingen, herontwerp van functies, flexibele werkinhoud en herplaatsing vormen de kernelementen van het beleid. Door bijvoorbeeld gebruik te maken van ondergrondse containers hoeven de afvalzakken niet meer getild worden, maar door middel van een afstandsbediening geleegd (Ecorys, 2009). Werkgevers moeten door de overheid gemotiveerd worden om dit soort investeringen te doen zodat ouderen langer door kunnen werken bij fysiek belastende banen. Hierbij kan bijvoorbeeld gedacht worden aan subsidies, maar ook kan voorlichting aangeboden worden aan de werkgever die de voordelen van het langer doorwerken van ouder personeel belicht.

Literatuurlijst

ABP. (2019, 21 oktober). *Nederlands pensioenstelsel scoort wereldwijd het hoogst*. Geraadpleegd op 13 maart 2022, van https://www.abp.nl/over-abp/actueel/nieuws/nederlands-pensioenstelsel-krijgt-opnieuw-hoogste-waardering.asp

Bijlsma, I., Fouarge, D., & Montizaan, R. (2017). Werken met een AOW. Omvang, beweegredenen en risico's van verdringing. *Netspar Industry Paper 69*.

CentERdata. (2021, 9 juni). *LISS panel*. Centerdata EN. Geraadpleegd op 3 mei 2022, van https://www.centerdata.nl/en/liss-panel

Centraal Bureau voor de Statistiek. (2021, 25 augustus). *300 duizend werkende 65-plussers in 2020*. Geraadpleegd op 13 maart 2022, van https://www.cbs.nl/nl-nl/nieuws/2021/34/300-duizend-werkende-65-plussers-in-2020

Centraal Bureau voor de Statistiek. (2022, February 25). *Leeftijdsverdeling*. https://www.cbs.nl/nl- nl/visualisaties/dashboard-bevolking/leeftijd/bevolking

Centraal Planbureau, Adema, Y., & Van Tilburg, I. (2019, december). *CPB Vergrijzingsstudie.* https://www.cpb.nl/sites/default/files/omnidownload/CPB-Vergrijzingsstudie-2019-Zorgen-om-morgen.pdf

CZ. (2019). *Omgaan met fysieke belasting tijdens werk*. https://www.cz.nl/-/media/zakelijk/informatiekit-fysieke-belasting-fysiek-zwaar-werk.pdf

Dingemans, E. A. A., & Dingemans, E. A. A. (2016). *Working after retirement: determinants and consequences of bridge employment* (dissertation). Rijksuniversiteit Groningen.

ECORYS, Cuelenaere, B., Deckers, K., Siegert, J., & De Bruin, G. (2009, oktober). *Langer doorwerken met beleid*.https://www.eumonitor.nl/9353000/1/j4nvgs5kjg27kof_j9vvik7m1c3gyxp/vib98o6umgzo/f=/blg22177.pdf

Eismann, M., Henkens, K., & Kalmijn, M. (2020). Vervroegd met pensioen: de invloed van een partner [in de etalage]. *Mens En Maatschappij,* 95(1), 58–61.

Henkens Kène, Dalen, H. P. van, & Solinge, H. van. (2009*). De vervagende grens tussen werk en pensioen* :over doorwerkers, doorstarters en herintreders (Ser. Nidi-rapport, no. 78). KNAW Press.Jansen,

N., Stynen, D., de Grip, A., & Kant, I. J. (2018). Bevordering arbeidsparticipatie oudere werknemers. *Tbv -Tijdschrift Voor Bedrijfs- En Verzekeringsgeneeskunde*, 26(10), 526–532. https://doi.org/10.1007/s12498-018-0315-5

Judge, T. A., Piccolo, R. F., Podsakoff, N. P., Shaw, J. C., & Rich, B. L. (2010). The relationship between pay and job satisfaction: a meta-analysis of the literature. *Journal of Vocational Behavior,* 77(2), 157–167. https://doi.org/10.1016/j.jvb.2010.04.002

Knoef, M., Goudswaard, K., Rhuggenaath, J., Caminada, K., & Been, J. (2017, 10 november). *De toereikendheid van pensioenopbouw na de crisis en pensioenhervormingen*. Netspar. Geraadpleegd op 4 juni 2022, van https://www.netspar.nl/publicatie/toereikendheid-pensioenopbouw-crisis-en-pensioenhervormingen/

Netspar, Oude Mulders, J., & Damman, M. (2019, april). *Doorwerken tot op hogere leeftijd. Wat vinden werkgevers, werknemers en zzp'ers?* (Netspar Brief 16). https://www.netspar.nl/home/netspar-brief-16-doorwerken-tot-op-hogere-leeftijd-wat-vinden-werkgevers-werknemers-en-zzpers.

Ministerie van Algemene Zaken. (2021, December 21). *Wanneer gaat mijn AOW in?* Rijksoverheid.nl. https://www.rijksoverheid.nl/onderwerpen/algemene-ouderdomswet-aow/vraag-en-antwoord/wanneer-gaat-mijn-aow-in

NOS. (2021, 3 juli). *Overal personeel gezocht, "tekort aan mensen is de nieuwe crisis*". NOS. https://nos.nl/artikel/2387727-overal-personeel-gezocht-tekort-aan-mensen-is-de-nieuwe-crisis

Oude Mulders, J. (2019). Welke werkgevers laten mensen doorwerken na de aow-leeftijd? *Tijdschrift Voor Arbeidsvraagstukken*, 35(4), 395–406.

Oude Mulders, J., & Dingemans, E. (2017). Werken na de pensionering in Nederland. *Gedrag & Organisatie*, 30(2). https://doi.org/10.5117/2017.030.002.002

Sewdas, R. (2018). Doorwerken na 65 jaar. *Tbv - Tijdschrift Voor Bedrijfs- En Verzekeringsgeneeskunde*, 26(2), 68–69. https://doi.org/10.1007/s12498-018-0027-x

Sewdas, R., De Wind, A., Van der Zwaan, L.G.L., Van der Borg, W.E., Steenbeek, R., Van der Beek, A.J., & Boot, C.R.L. (2017). Why older workers work beyond the retirement age: A qualitative study. *BMC Public Health,* 17, 672.

Solinge, H., & Henkens, K. (2016). Doorwerken na pensioen vooral populair in hogere functies. *Demos: Bulletin Over Bevolking En Samenleving*, *32*(5)

Steenbeek, R. S., Bijlsma, I. B., Fouarge, D. F., Hooftman, W. H., & Montizaan, R. M. (2017). *Doorwerken na de AOW-gerechtigde leeftijd: Van impact en drijfveren naar implicaties voor beleid. TNO ; ROA, Leiden.* http://resolver.tudelft.nl/uuid:a6ed77fd-5798-4de2-aceb- 4605059344a5

Straatmeijer, J. S. (2020, september). *Doorwerken na de AOW-leeftijd* (Nr. 20086). Regioplan. https://www.kbo-pcob.nl/wp-content/uploads/2020/09/20086-Eindrapport-Doorwerken-na-de-AOW-leeftijd-Regioplan-7sept2020-1.pdf

Wet verhoging AOW- en pensioenrichtleeftijd (33.290). (2012). Eerste Kamer der Staten-Generaal. Geraadpleegd op 21 maart 2022, van https://www.eerstekamer.nl/wetsvoorstel/33290_wet_verhoging_aow_en

Wind, A. de, Beek, A. J. van der, Hulshof, C. T. J., Gulden, J. W. J. van der, & Boot, C. R. L. (2020). Doorwerken na je pensioen: wel of niet goed voor je gezondheid? *Tijdschrift Voor Bedrijfs- En Verzekeringsgeneeskunde*, 28, 38–41.

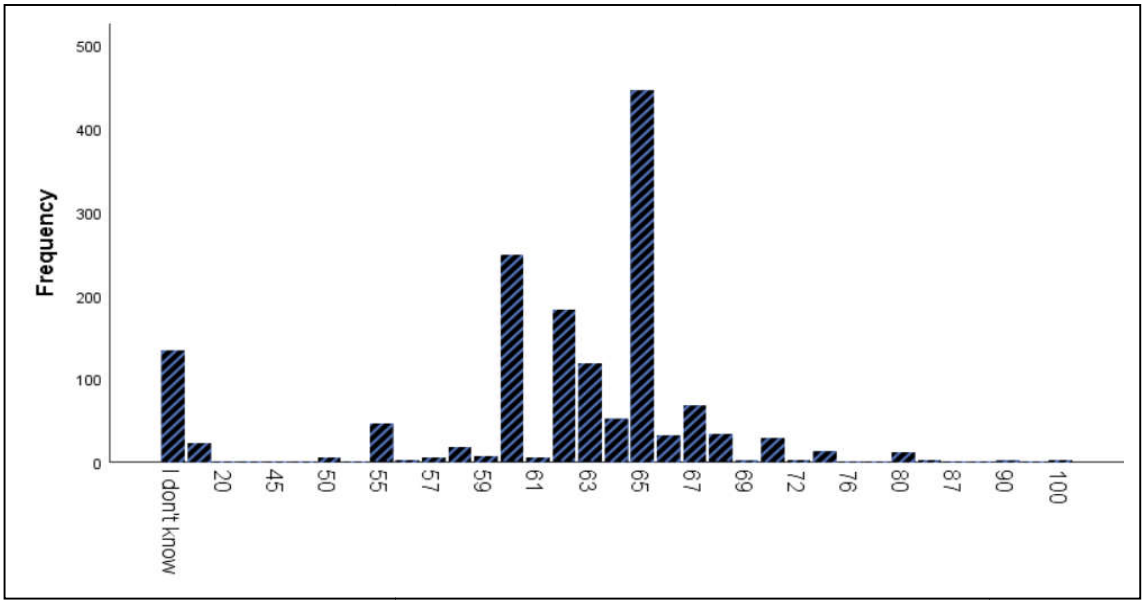
Bijlage 1

In deze bijlage is beschreven bij welke variabelen verdere operationalisaties zijn toegepast en hoe dat is gedaan. Eerst wordt gekeken naar de oorspronkelijke variabele Gewenste Pensioenleeftijd en vervolgens naar de hiermee zelf geconstrueerde variabele Doorwerken. Daarna wordt gekeken naar de uit vier variabelen samengevoegde variabele Fysieke Belasting Werk. Ook wordt de operationalisatie van de variabele Verwachte Inkomsten belicht en tot slot die van Sector.

Operationalisaties

Doorwerken & Gewenste pensioenleeftijd

De variabele doorwerken is zelf geconstrueerd door de variabele 'gewenste pensioenleeftijd' - 67 te doen. Hier is eerst een beschrijving van de variabele gewenste pensioenleeftijd te zien:



B1.1: verdeling van variabele over gewenste pensioenleeftijd

De syntax voor bovenstaande verdeling is als volgt:

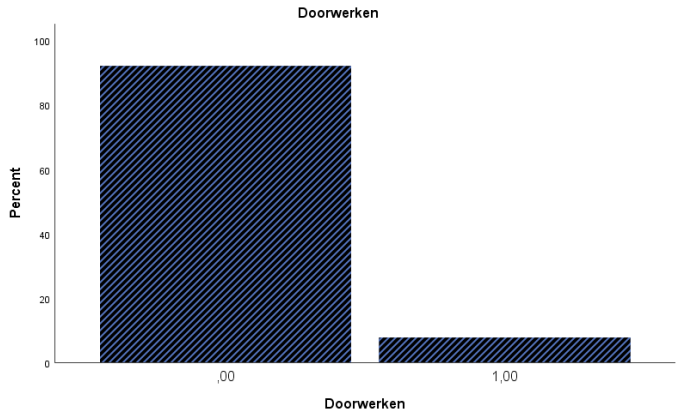
```
FREQUENCIES VARIABLES=cw20m288
 /BARCHART FREQ
 /ORDER=ANALYSIS.
```

Hieronder is de Syntax te zien van het construeren van de variabele **doorwerken**. Wanneer het verschil tussen 67 en de gewenste pensioenleeftijd 1 of groter (maximaal 33) was is dit gecodeerd als 1 en alles daaronder is gecodeerd als 0.

```
*Doorwerken.
DATASET ACTIVATE DataSet8.
COMPUTE Doorwerken=GewenstePensioenlftd - 67.
EXECUTE.
RECODE Doorwerken (1 thru 33=1) (-47 thru 0=0) INTO Doorwerken.
EXECUTE.
```

Vervolgens is de steekproef voor dit onderzoek geselecteerd door alleen de cases met Doorwerken = 0 of Doorwerken = 1 te selecteren. Op deze manier heb je alleen de cases die informatie bieden over het langer doorwerken.

```
*steekproef.
USE ALL.
COMPUTE filter_$=(Doorwerken >= 0).
VARIABLE LABELS filter_$ 'Doorwerken >= 0 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.
```



B1.2: Histogram van de variabele doorwerken

Hierboven is een histogram te zien met het percentage mensen dat door wil werken na het bereiken van de AOW-leeftijd (5,8%) en het percentage mensen dat niet door wil werken na het bereiken van de AOW-leeftijd (94,2).

Fysieke Belasting werk

Voorafgaand aan het samenvoegen van variabelen uit de dataset tot de variabele Fysieke Belasting Werk is gekeken naar de Cronbach's Alpha. De Syntax hiervoor is als volgt:

```
DATASET ACTIVATE

RELIABILITY

 /VARIABLES=cw20m416 cw20m417 cw20m418 cw20m419

 /SCALE('ALL VARIABLES') ALL

 /MODEL=ALPHA

 /STATISTICS=DESCRIPTIVE SCALE CORR

 /SUMMARY=TOTAL.
```

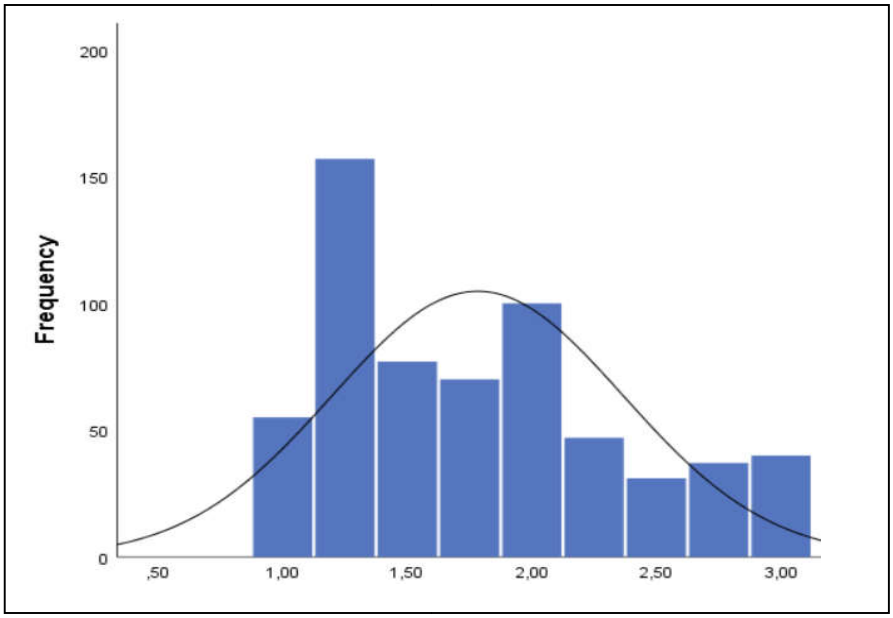
Nadat een Chonbach’s Alpha van 0,847 een indicatie gaf van een goede mate van samenhang tussen de variabelen, zijn deze samengevoegd. De bijbehorende syntax is als volgt:

COMPUTE FysiekeBelasting=MEAN(cw20m416,cw20m417,cw20m418,cw20m419).

EXECUTE.

Ook is de schaal omgedraaid zodat 1 staat voor 'nooit', 2 voor 'soms' en 3 voor 'vaak'. Syntax hiervoor is:

```
RECODE FysiekeBelasting (1=3) (3=1) (1.25=2.75) (2.75=1.25) (1.5=2.5) (2.5=1.5) (1.75=2.25)

 (2.25=1.75) (2=2) INTO FysiekeBelastingR.

EXECUTE.
```



B1.3: Verdeling van de variabele Fysieke Belasting Werk

Hierboven is een verdeling te zien van de variabele Fysieke Belasting Werk. De variabele is zoals te zien niet normaal verdeeld en neigt meer naar een rechtsscheef scheve verdeling. Respondenten die aangeven niet fysiek zwaar werk verrichten komen vaker voor dan respondenten die wel fysiek zwaar werk verrichten.

B1.4: Frequentietabel van de variabele Fysieke Belasting

Fysieke Belasting Werk	*Frequentie*	*%*
1 'nooit' 2 'soms' 3 'vaak'		
1	133	10,1
1,25	372	28,4
1,50	184	14,0
1,75	112	8,5
2	210	16,0
2,25	105	8,0
2,5	59	4,5
2,75	63	4,8
3	73	5,6
Totaal	1311	100
Missing	0	0

Hierboven is een Frequentietabel te zien van de variabel Fysieke Belasting Werk. Door een schaalscore te construeren uit 4 items krijg je een score die meer op een continue score lijkt dan de losse items zelf, en ook meer informatie geeft. Zoals te zien komt een gemiddelde score van 1,25 het vaakste voor.

De syntax voor bovenstaande frequentietabel en histogram is als volgt:

```
FREQUENCIES VARIABLES=FysiekeBelastingR

 /STATISTICS=STDDEV MINIMUM MAXIMUM MEAN MEDIAN SKEWNESS SESKEW

 /HISTOGRAM NORMAL

 /ORDER=ANALYSIS.
```

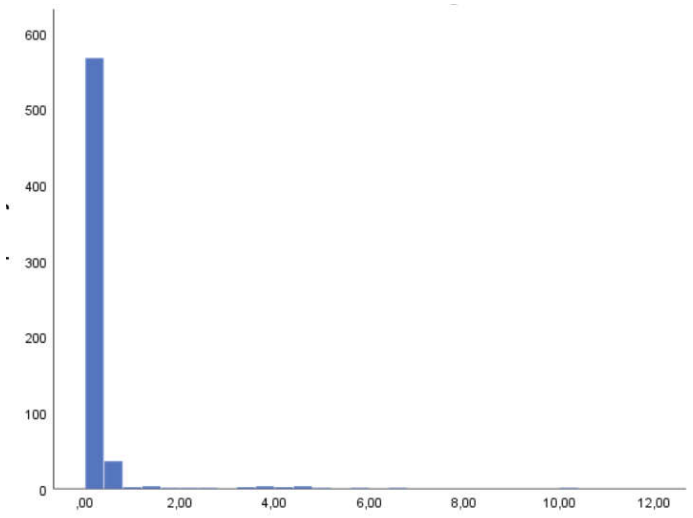
Verwachte Inkomsten

De variabele Verwachte Inkomsten betreft een erg brede schaal en grote verschillen in inkomsten. Om de resultaten van deze variabele beter interpreteerbaar te maken is er voor gekozen om het verwachte inkomsten door 10.000 te delen. De syntax hiervoor is als volgt:

```
COMPUTE VerwachtInkomstenREC=Verwachtinkomsten / 10000.

EXECUTE.
```

Vervolgens is naar de verdeling van de variabele gekeken:

```
FREQUENCIES VARIABLES=VerwachteInkomstenREC

 /HISTOGRAM

 /ORDER=ANALYSIS.
```



B1:5 Histogram van geoperationaliseerde variabele Verwachte Inkomsten

Hierboven is te zien dat de variabele over Verwachte inkomsten niet normaal, maar rechtsscheef verdeeld is.

Sector

Aangezien de controlevariabele ‘sector’ uit 15 categorieën bestaat en deze variabele een nominale schaal betreft, was het voor deze analyse van belang om dummy variabelen te creëren. Eerst is er voor gekozen om de 15 categorieën in 4 groepen onder te verdelen. Categorieën 6 tot en met 10 onder ‘Diensten & Handel’ , 11 tot en met 14 onder ‘Publieke Sector’ en categorie 15 ten slotte onder ‘Overig’. Vervolgens zijn van deze groepen 3 dummyvariabelen gemaakt en is de groep ‘Bouw en Industrie’ als referentiegroep gebruikt. De syntax hiervoor is als volgt:

```
*Dummy'ssector.

RECODE Sector (15=1) (ELSE=0) INTO Overig.

RECODE Sector (11 thru 14=1)(ELSE=0) INTO PubliekeSector.

RECODE Sector (6 thru 10=1)(ELSE=0) INTO DienstenHandel.
```

Bijlage 2

In deze bijlage wordt eerst gekeken naar de descriptieve statistieken voor alle variabelen. Daarna is gekeken naar de mate van samenhang tussen afhankelijke en onafhankelijke variabelen. De Pearson's correlatie is gebruikt voor de samenhang tussen continue variabelen (en dummy's), Cramer's v en chi-kwadraattoetsen voor de samenhang van categorische variabelen onderling gebruikt en zijn de correlaties van een continue met een categorische variabelen (kleiner dan 2) berekend aan de hand van de R2 uit ANOVA analyses. Een overzicht hiervan is in tabel 2 te vinden. Vervolgens wordt de Hypothesetoetsing verder belicht en wordt ten slotte gekeken naar de module evaluatie.

Descriptieve statistieken

Hieronder volgt de Syntax en ouput voor de descriptieve statistieken van de variabele Doorwerken:

```
FREQUENCIES VARIABLES=Doorwerken
 /HISTOGRAM
 /ORDER=ANALYSIS.
```

Statistics

Doorwerken

N	Valid	1311
	Missing	0
Minimum		,00
Maximum		1,00

Doorwerken

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Nee	1235	94,2	94,2	94,2
	Ja	76	5,8	5,8	100,0
	Total	1311	100,0	100,0	

B2:1 Output van de descriptieve statistieken van Doorwerken

Hieronder volgt de Syntax en output voor de descriptieve statistieken van de variabele Fysieke Belasting:

```
FREQUENCIES VARIABLES=FysiekeBelastingR
 /HISTOGRAM
 /ORDER=ANALYSIS.
```

Statistics

FysiekeBelastingR

N	Valid	1311
	Missing	0
Minimum		1,00
Maximum		3,00

FysiekeBelastingR

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	133	10,1	10,1	10,1
	1,25	372	28,4	28,4	38,5
	1,50	184	14,0	14,0	52,6
	1,75	112	8,5	8,5	61,1
	2,00	210	16,0	16,0	77,1
	2,25	105	8,0	8,0	85,1
	2,50	59	4,5	4,5	89,6
	2,75	63	4,8	4,8	94,4
	3,00	73	5,6	5,6	100,0
	Total	1311	100,0	100,0	

B2:2 Output van de descriptieve statistieken van Fysieke Belasting

De variabele Fysieke Belasting bestaat uit 5 samengevoegde variabelen waar vervolgens het gemiddelde van is genomen. Om deze reden zijn er ook getallen ontstaan als 1,25 en 1,75. Voor de weergave in tabel 1 in het resultatenhoofdstuk is ervoor gekozen om respondenten met antwoordcategorieën als deze samen te voegen. Zo zijn bijvoorbeeld respondenten met antwoordcategorieën 1, 1,25 en 1,5 samen weergegeven als 1 en respondenten met antwoordcategorieën 1,75, 2 en 2,25 samen weergegeven als 2.

B2.3 : tabel met verdeling van de variabele Fysieke Belasting werk na creëren van dummy's

Fysieke Belasting Werk 1 'nooit' 2 'soms' 3 'vaak'	*Frequentie*	*%*
1	689	52,5%
2	427	32,6%
3	195	14,4%
Totaal	1311	100

Hieronder volgt de Syntax en output voor de descriptieve statistieken van de variabele Partner:

```
FREQUENCIES VARIABLES=Partner
 /HISTOGRAM
 /ORDER=ANALYSIS.
```

Statistics

The household head lives together with a partner (wedded or unwedded)

N	Valid	1311
	Missing	0
Minimum		0
Maximum		1

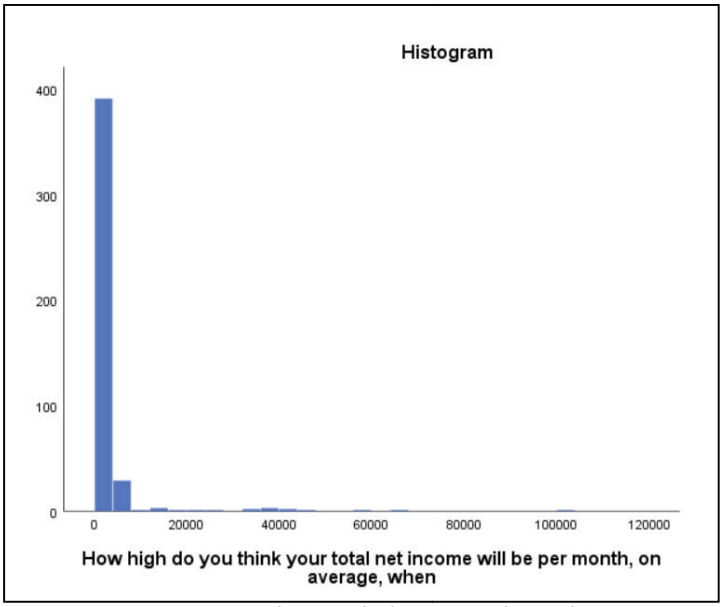
The household head lives together with a partner (wedded or unwedded)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Nee	312	23,8	23,8	23,8
	Ja	999	76,2	76,2	100,0
	Total	1311	100,0	100,0	

B2.5 output van de descriptieve statistieken van de variabele Partner

Hieronder volgt de Syntax en output voor de descriptieve statistieken van de variabele Verwachte Inkomsten:

```
FREQUENCIES VARIABLES=Verwachteinkomsten
 /STATISTICS=STDDEV MINIMUM MAXIMUM MEAN MEDIAN MODE SUM SKEWNESS SESKEW KURTOSIS
SEKURT
 /HISTOGRAM
 /ORDER=ANALYSIS.
```



B2.5: Histogram van de variabele Verwachte Inkomsten

Statistics

How high do you think your total net income will be per month, on average, when

N	Valid	438
	Missing	873
Mean		3690,17
Median		2200,00
Mode		2000
Std. Deviation		8120,762
Variance		65946783,27
Skewness		7,403
Std. Error of Skewness		,117
Kurtosis		66,693
Std. Error of Kurtosis		,233
Minimum		0
Maximum		102000

B2.6 Descriptieve statistieken van de variabele Verwachte Inkomsten

Zoals te zien is de variabele over het verwachte inkomen rechtsscheef verdeeld. Ook is een mediaan van 2200 en een grote skewness te zien.

Hieronder volgt de Syntax en output voor de descriptieve statistieken van de variabele Geslacht:

Statistics

Gender

N	Valid	1311
	Missing	0
Minimum		1
Maximum		2

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Man	682	52,0	52,0	52,0
	Vrouw	629	48,0	48,0	100,0
	Total	1311	100,0	100,0	

B2.7 output van de descriptieve statistieken van Geslacht

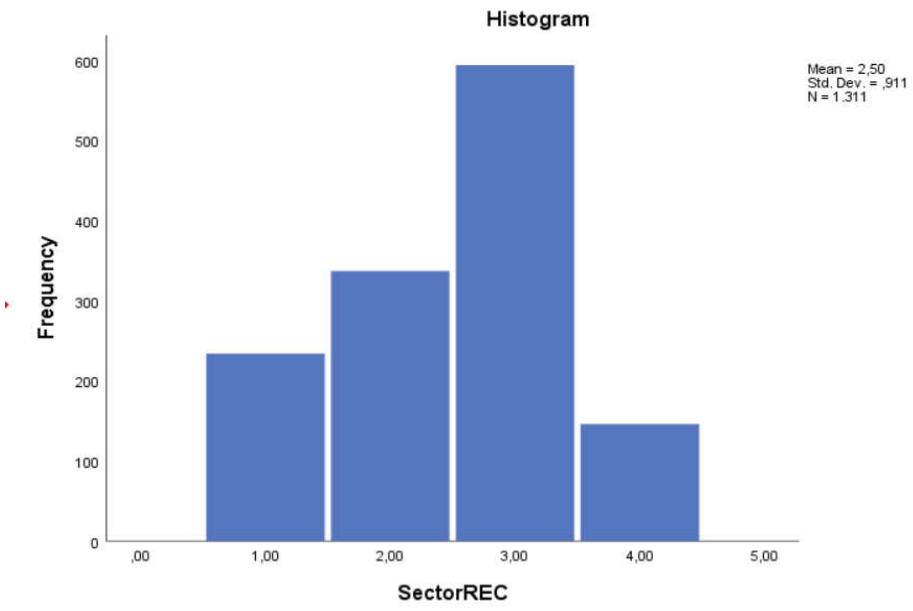
Zoals te zien zitten er iets meer mannen de vrouwen in de dataset.

Voor de variabele Sector in tabel 1 is de oorspronkelijke variabele Sector gehercodeerd in 4 categorieën. Deze syntax is als volgt:

```
RECODE Sector (15=4) (1 thru 5=1) (6 thru 10=2) (11 thru 14=3) INTO SectorREC.
EXECUTE.
```

Later zijn hier voor het uitvoeren van de regressieanalyse dummy variabelen van gemaakt. Hieronder volgt de Syntax en output voor de descriptieve statistieken en een histogram van de gehercodeerde variabele Sector:

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Bouw & Industrie	234	17,8	17,8	17,8
	Diensten & Handel	337	25,7	25,7	43,6
	Publieke Sector	594	45,3	45,3	88,9
	Overig	146	11,1	11,1	100,0
	Total	1311	100,0	100,0	



B2.7 Output van de descriptieve statistieken + histogram van de variabele Sector

Zoals te zien in bovenstaande tabel en histogram zijn de meeste respondenten werkzaam in de publieke sector. Het kleinste deel van de mensen geeft aan in de overige sector werkzaam te zijn.

Hieronder volgt de Syntax en output voor de descriptieve statistieken van de variabele Geboortejaar:

```
FREQUENCIES VARIABLES=Geboortejaar
 /STATISTICS=STDDEV MINIMUM MAXIMUM MEAN MEDIAN MODE SUM SKEWNESS SESKEW KURTOSIS
SEKURT
 /ORDER=ANALYSIS.
```

Statistics

Respondent's year of birth

N	Valid	1311
	Missing	0
Mean		1964,26
Median		1964,00
Mode		1963
Std. Deviation		5,861
Variance		34,356
Skewness		-,015
Std. Error of Skewness		,068
Kurtosis		-1,056
Std. Error of Kurtosis		,135
Minimum		1953
Maximum		1975

B2.8 output van de statistische statistieken van de variabele geboortejaar

Zoals te zien komen de jongste respondenten uit het geboortejaar 1975 en de oudste uit 1953. Het vaakste komen mensen uit 1964.

Samenhang tussen de variabelen

Cramer's V

Om te kijken naar de samenhang tussen categorische variabelen te kijken is Cramers V gebruikt.

Voor de variabelen Partner en Fysiekebelasting volgt de syntax + output hieronder:

```
CROSSTABS
 /TABLES=Partner BY FysiekebelastingR
 /FORMAT=AVALUE TABLES
 /STATISTICS=PHI
 /CELLS=COUNT
 /COUNT ROUND CELL.
```

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
The household head lives together with a partner (wedded or unwedded) * FysiekeBelastingR	1311	100,0%	0	0,0%	1311	100,0%

The household head lives together with a partner (wedded or unwedded) * FysiekeBelastingR Crosstabulation

Count

		FysiekeBelastingR									
		1,00	1,25	1,50	1,75	2,00	2,25	2,50	2,75	3,00	Total
The household head lives together with a partner (wedded or unwedded)	Nee	27	83	40	29	54	37	11	13	18	312
	Ja	106	289	144	83	156	68	48	50	55	999
Total		133	372	184	112	210	105	59	63	73	1311

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,093	,186
	Cramer's V	,093	,186
N of Valid Cases		1311	

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	11,296[a]	8	,186
Likelihood Ratio	10,738	8	,217
Linear-by-Linear Association	1,479	1	,224
N of Valid Cases	1311		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 14,04.

B2.9: Output voor de samenhang tussen de variabelen Partner en Fysiekebelasting op basis van Cramer's V

In bovenstaande output is te zien dat en een erg zwakke en niet significante samenhang tussen Partner en Fysieke Belasting bestaat. Er is te zien dat mensen met en zonder een partner het vaakst in categorie 1,23 te zitten en dus geen fysiek werk verrichten.

Voor de variabelen Doorwerken en Fysieke belasting volgt de output en syntax hieronder:

```
CROSSTABS
 /TABLES=Doorwerken BY FysiekebelastingR
 /FORMAT=AVALUE TABLES
 /STATISTICS=PHI
 /CELLS=COUNT
```

/COUNT ROUND CELL.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Doorwerken * FysiekeBelastingR	1311	100,0%	0	0,0%	1311	100,0%

Doorwerken * FysiekeBelastingR Crosstabulation

Count

		FysiekeBelastingR									
		1,00	1,25	1,50	1,75	2,00	2,25	2,50	2,75	3,00	Total
Doorwerken	Nee	125	339	178	105	195	102	56	62	73	1235
	Ja	8	33	6	7	15	3	3	1	0	76
Total		133	372	184	112	210	105	59	63	73	1311

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,116	,024
	Cramer's V	,116	,024
N of Valid Cases		1311	

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	17,607[a]	8	,024
Likelihood Ratio	22,473	8	,004
Linear-by-Linear Association	8,209	1	,004
N of Valid Cases	1311		

a. 3 cells (16,7%) have expected count less than 5. The minimum expected count is 3,42.

B2.10: output voor de samenhang tussen de variabelen Doorwerken en Fysieke belasting op basis van Cramer's V

Er is een zwakke significante samenhang te zien tussen Doorwerken en Fysieke belasting. Het grootste deel van de respondenten dat door willen werken geeft aan in antwoordcategorie 1,25 te zitten. Wat opvalt is dat van de respondenten die aangeven door te willen werken, geen een vaak fysiek belastend werk verricht (antwoordcategorie 3).

Hieronder volgt de syntax en output voor de samenhang tussen Doorwerken en Partner:

```
CROSSTABS
 /TABLES=Doorwerken BY Partner
 /FORMAT=AVALUE TABLES
 /STATISTICS=PHI
 /CELLS=COUNT
 /COUNT ROUND CELL.
```

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Doorwerken * The household head lives together with a partner (wedded or unwedded)	1311	100,0%	0	0,0%	1311	100,0%

Doorwerken * The household head lives together with a partner (wedded or unwedded) Crosstabulation

Count

		The household head lives together with a partner (wedded or unwedded)		
		Nee	Ja	Total
Doorwerken	Nee	293	942	1235
	Ja	19	57	76
Total		312	999	1311

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	-,007	,800
	Cramer's V	,007	,800
N of Valid Cases		1311	

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,064[a]	1	,800		
Continuity Correction[b]	,013	1	,909		
Likelihood Ratio	,064	1	,801		
Fisher's Exact Test				,782	,446
Linear-by-Linear Association	,064	1	,800		
N of Valid Cases	1311				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 18,09.

b. Computed only for a 2x2 table

B2.11: Output voor de samenhang tussen Doorwerken en Partner op basis van Cramer's V

Er is een verwaarloosbare en niet significante samenhang tussen Doorwerken en Partner te zien. Het grootste gedeelte van de respondenten dat aangeeft door te willen werken heeft een partner, maar dit geldt ook voor de respondenten die aangeven niet door te willen werken.

<u>Pearson's correlatie</u>

Syntax en output voor de samenhang tussen Verwachte inkomsten en Doorwerken:

```
NONPAR CORR
 /VARIABLES=VerwachtInkomenREC Doorwerken
 /PRINT=BOTH TWOTAIL NOSIG
 /MISSING=PAIRWISE.
```

Correlations

			VerwachtInkomenREC	Doorwerken
Kendall's tau_b	VerwachtInkomenREC	Correlation Coefficient	1,000	,123**
		Sig. (2-tailed)	.	,002
		N	438	438
	Doorwerken	Correlation Coefficient	,123**	1,000
		Sig. (2-tailed)	,002	.
		N	438	1311
Spearman's rho	VerwachtInkomenREC	Correlation Coefficient	1,000	,148**
		Sig. (2-tailed)	.	,002
		N	438	438
	Doorwerken	Correlation Coefficient	,148**	1,000
		Sig. (2-tailed)	,002	.
		N	438	1311

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

B2.12: Output voor de samenhang tussen Verwachte Inkomsten en Doorwerken op basis van Pearson's correlatie

Er is een zwakke maar wel significante samenhang tussen Doorwerken en Verwachte inkomsten te zien.

Hieronder volgt de syntax en output voor de samenhang tussen Verwachte Inkomsten en Partner:

```
NONPAR CORR
 /VARIABLES=VerwachtInkomenREC Partner
 /PRINT=BOTH TWOTAIL NOSIG
 /MISSING=PAIRWISE.
```

Correlations

			VerwachtInkomenREC	The household head lives together with a partner (wedded or unwedded)
Kendall's tau_b	VerwachtInkomenREC	Correlation Coefficient	1,000	,025
		Sig. (2-tailed)	.	,527
		N	438	438
	The household head lives together with a partner (wedded or unwedded)	Correlation Coefficient	,025	1,000
		Sig. (2-tailed)	,527	.
		N	438	1311
Spearman's rho	VerwachtInkomenREC	Correlation Coefficient	1,000	,030
		Sig. (2-tailed)	.	,528
		N	438	438
	The household head lives together with a partner (wedded or unwedded)	Correlation Coefficient	,030	1,000
		Sig. (2-tailed)	,528	.
		N	438	1311

B2.13: Output voor de samenhang tussen Verwachte Inkomsten en Partner op basis van Pearson's Correlatie

Er is een verwaarloosbare mate van samenhang tussen de variabelen te zien die bovendien niet significant is.

Correlatie op basis van ANOVA

Hieronder volgt de syntax en output voor de samenhang tussen Verwachte Inkomsten en Fysiekebelasting:

```
UNIANOVA VerwachteInkomenREC BY Fysiekebelastingher
 /METHOD=SSTYPE(3)
 /INTERCEPT=INCLUDE
 /PLOT=PROFILE(Fysiekebelastingher) TYPE=LINE ERRORBAR=CI MEANREFERENCE=NO YAXIS=AUTO
 /PRINT DESCRIPTIVE
 /CRITERIA=ALPHA(0.05)
 /DESIGN=Fysiekebelastingher.
```

Descriptive Statistics

Dependent Variable: How high do you think your total net income will be per month, on average, when

Fysiekebelastingher	Mean	Std. Deviation	N
1,00	4864,86	10779,176	51
1,25	2273,72	965,949	39
1,50	4524,67	8236,979	57
1,75	4232,70	10010,467	159
2,00	2279,87	856,082	61
2,25	1658,27	479,025	11
2,50	4281,93	8543,021	15
2,75	1928,61	586,795	33
3,00	5285,17	12512,156	12
Total	3690,17	8120,762	438

Tests of Between-Subjects Effects

Dependent Variable: How high do you think your total net income will be per month, on average, when

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	540039705[a]	8	67504963,14	1,024	,417
Intercept	2751864476	1	2751864476	41,747	,000
Fysiekebelastingher	540039705,1	8	67504963,14	1,024	,417
Error	2,828E+10	429	65917726,30		
Total	3,478E+10	438			
Corrected Total	2,882E+10	437			

a. R Squared = ,019 (Adjusted R Squared = ,000)

B2.13: Output voor de samenhang tussen Verwachte Inkomsten en Fysieke Belasting op basis van ANOVA

Door de wortel van ,019 (=,140) te nemen komt de uiteindelijke correlatiemaat die is weergegeven in tabel 2 tot stand. Dit geeft een zwakke niet significante mate van samenhang tussen Verwachte Inkomsten en Fysieke Belasting weer.

Hypothesetoetsing

Model 1 & 5

In model 1 is het effect van de controlevariabelen op de afhankelijke variabelen te zien. De controlevariabelen zijn: 'Geboortejaar', 'Geslacht' en de drie dummy variabelen die horen bij 'Sector'. De uitleg en interpretatie hiervan is te vinden in het resultatenhoofdstuk bij tabel 2. De bijbehorende syntax is als volgt:

```
LOGISTIC REGRESSION VARIABLES Doorwerken
 /METHOD=ENTER Geslacht DienstenHandel PubliekeSector Geboortejaar Overig
 /METHOD=ENTER Partner VerwachtInkomenREC FysiekeBelastingR
 /SAVE=LEVER
 /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

Wat meteen opvalt is dat in bovenstaande syntax ook de onafhankelijke variabelen te zien zijn. De controlevariabelen zijn toegevoegd in blok 1 en de onafhankelijke variabelen in blok 2. Zo was het bijvoorbeeld mogelijk om het verschil in Deviance tussen model 1 en model 5 te toetsen. De syntax voor model 1 & 5 is dus hetzelfde, maar voor model 1 is alleen naar de output van blok 1 gekeken en voor model 5 naar de output van blok 2. Blok 2 betreft dan niet alleen de onafhankelijke variabelen, maar ook de al eerder toegevoegde controlevariabelen uit blok 1. In model 5 is dus het 'complete' model te zien met alle controle en onafhankelijke variabelen.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	12,636	5	,027
	Block	12,636	5	,027
	Model	12,636	5	,027

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	211,307[a]	,028	,071

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than ,001.

Classification Table[a]

	Observed		Predicted: Doorwerken: Nee	Predicted: Doorwerken: Ja	Percentage Correct
Step 1	Doorwerken	Nee	407	0	100,0
		Ja	31	0	,0
	Overall Percentage				92,9

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Gender	-,698	,460	2,300	1	,129	,498
	DienstenHandel	,618	,546	1,279	1	,258	1,855
	PubliekeSector	-,543	,598	,823	1	,364	,581
	Respondent's year of birth	-,047	,035	1,817	1	,178	,954
	Overig	-,195	,873	,050	1	,823	,823
	Constant	90,653	68,482	1,752	1	,186	2,345E+39

a. Variable(s) entered on step 1: Gender, DienstenHandel, PubliekeSector, Respondent's year of birth, Overig.

B2.14: Output van de logistische regressie analyse met alleen de controlevariabelen

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	11,690	3	,009
	Block	11,690	3	,009
	Model	24,351	8	,002

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	199,592[a]	,054	,135

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than ,001.

Classification Table[a]

	Observed		Predicted: Doorwerken: Nee	Predicted: Doorwerken: Ja	Percentage Correct
Step 1	Doorwerken	Nee	407	0	100,0
		Ja	31	0	,0
	Overall Percentage				92,9

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Gender	-,648	,467	1,923	1	,166	,523
	BouwIndustrie	,158	,890	,031	1	,859	1,171
	DienstenHandel	,619	,799	,601	1	,438	1,858
	PubliekeSector	-,406	,819	,245	1	,620	,667
	Respondent's year of birth	-,050	,036	1,942	1	,163	,951
	The household head lives together with a partner (wedded or unwedded)	-,595	,438	1,840	1	,175	,552
	VerwachtInkomenREC	,138	,152	,822	1	,365	1,148
	FysiekeBelastingR	-1,533	,579	7,007	1	,008	,216
	Constant	98,707	70,310	1,971	1	,160	7,377E+42

a. Variable(s) entered on step 1: The household head lives together with a partner (wedded or unwedded), VerwachtInkomenREC, FysiekeBelastingR.

B2.15: Output van de logistische regressie analyse van het complete model

Model 2

Met model 2 wordt het effect van de onafhankelijke variabele 'Fysieke Belasting' op op het langer willen doorwerken na de AOW-leeftijd getoetst. De resultaten hiervan zijn te vinden in tabel 2 en de interpretatie hiervan is te vinden in het resultatenhoofdstuk. De syntax voor model 2 is als volgt:

```
LOGISTIC REGRESSION VARIABLES Doorwerken
 /METHOD=ENTER FysiekeBelastingR
 /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	8,954	1	,003
	Block	8,954	1	,003
	Model	8,954	1	,003

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	571,420[a]	,007	,019

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than ,001.

Classification Table[a]

	Observed		Predicted: Doorwerken: Nee	Predicted: Doorwerken: Ja	Percentage Correct
Step 1	Doorwerken	Nee	1235	0	100,0
		Ja	76	0	,0
	Overall Percentage				94,2

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	FysiekeBelastingR	-,680	,241	7,993	1	,005	,506
	Constant	-1,674	,392	18,191	1	,000	,188

a. Variable(s) entered on step 1: FysiekeBelastingR.

B2.16: Output van de logistische regressie analyse met alleen de variabele over Fysieke Belasting

Model 3

Model 3 laat het effect van het hebben van een partner op langer door willen werken na het bereiken van de AOW-leeftijd zien. De resultaten hiervan zijn te vinden in tabel 2 en verdere uitleg hierover is te vinden in het resultatenhoofdstuk. De syntax voor model 3 is als volgt:

```
LOGISTIC REGRESSION VARIABLES Doorwerken
 /METHOD=ENTER Partner
 /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

		Chi-square	df	Sig.
Step 1	Step	,064	1	,801
	Block	,064	1	,801
	Model	,064	1	,801

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	580,310[a]	,000	,000

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

Classification Table[a]

	Observed		Predicted Doorwerken Nee	Predicted Doorwerken Ja	Percentage Correct
Step 1	Doorwerken	Nee	1235	0	100,0
		Ja	76	0	,0
	Overall Percentage				94,2

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	The household head lives together with a partner (wedded or unwedded)	-,069	,273	,064	1	,800	,933
	Constant	-2,736	,237	133,541	1	,000	,065

a. Variable(s) entered on step 1: The household head lives together with a partner (wedded or unwedded).

B2.17: Output van de logistische regressie analyse met alleen de variabel Partner

Model 4

In model 4 is het effect van de inkomsten die verwacht worden wanneer men de AOW-leeftijd bereikt op het langer willen doorwerken na deze leeftijd te zien. De resultaten hiervan zijn te vinden in tabel 2 en verdere uitleg hierover is te vinden in het resultatenhoofdstuk. De syntax voor model 4 is als volgt:

```
LOGISTIC REGRESSION VARIABLES Doorwerken
 /METHOD=ENTER VerwachteInkomstenREC
 /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.50)
```

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	2,544	1	,111
	Block	2,544	1	,111
	Model	2,544	1	,111

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	221,398[a]	,006	,014

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

Classification Table[a]

	Observed		Predicted: Doorwerken: Nee	Predicted: Doorwerken: Ja	Percentage Correct
Step 1	Doorwerken	Nee	407	0	100,0
		Ja	31	0	,0
	Overall Percentage				92,9

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	VerwachtInkomenREC	,251	,141	3,175	1	,075	1,285
	Constant	-2,690	,203	175,539	1	,000	,068

a. Variable(s) entered on step 1: VerwachtInkomenREC.

B2.18: Output van de logistische regressie analyse met alleen de variabele over Verwacht Inkomen

Modelfit

Hieronder is de syntax en output voor de Hosmer and Lemeshow toets te zien:

```
LOGISTIC REGRESSION VARIABLES Doorwerken
 /METHOD=ENTER Geboortejaar Overig PubliekeSector DienstenHandel Sector
 /METHOD=ENTER FysiekeBelastingR Partner VerwachtInkomenREC
 /PRINT=GOODFIT
 /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	3,769	8	,877

Contingency Table for Hosmer and Lemeshow Test

		Doorwerken = Nee		Doorwerken = Ja		
		Observed	Expected	Observed	Expected	Total
Step 1	1	44	43,574	0	,426	44
	2	42	43,206	2	,794	44
	3	43	42,776	1	1,224	44
	4	42	42,400	2	1,600	44
	5	43	41,992	1	2,008	44
	6	42	41,469	2	2,531	44
	7	42	40,848	2	3,152	44
	8	39	39,728	5	4,272	44
	9	38	38,293	6	5,707	44
	10	31	31,714	10	9,286	41

B2.19: Output van de resultaten van de Hosmer & Lemeshow toets

Bijlage 3

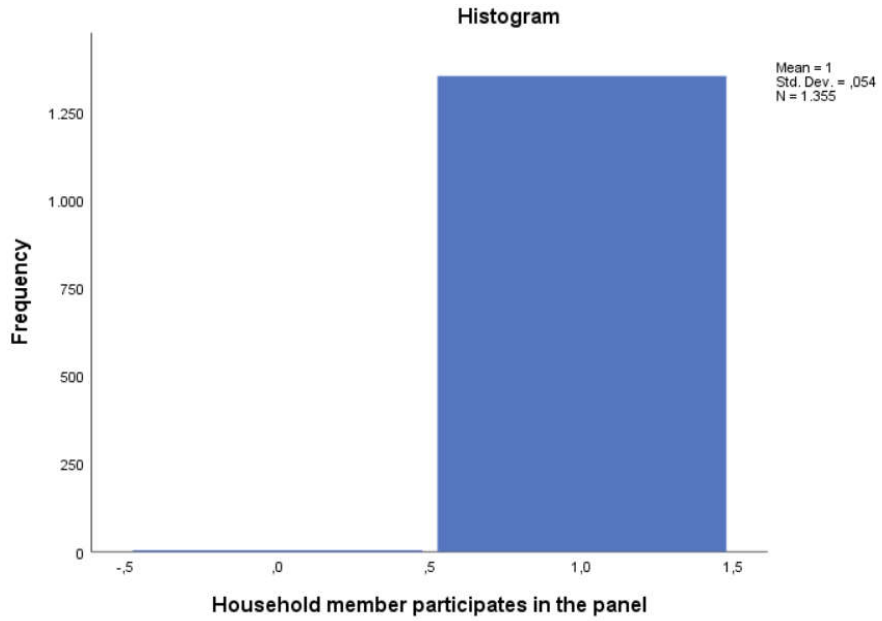
In deze bijlage is informatie te vinden over de assumpties die horen bij een logistische regressie. Ook wordt gekeken naar de aanwezigheid van multicollineariteit door middel van VIF-scores. Ten slotte wordt er gekeken naar uitschieters door te kijken naar Leverage waarden.

Assumpties

Bij het doen van een logistische regressie analyse wordt uitgegaan van twee assumpties. Één assumptie is dat de waarnemingen in de steekproef onafhankelijk van elkaar zijn. Dit houdt in dat de steekproefeenheden (respondenten) onafhankelijk van elkaar zijn getrokken en er geen samenhang tussen de geobserveerde scores van de respondenten bestaat. De steekproefeenheden die voor dit onderzoek gebruikt worden zijn aselect geselecteerd uit data van het Centraal Bureau voor de Statistiek (CentERdata, 2021). De steekproefeenheden zijn dus onafhankelijk van elkaar geselecteerd, maar over of er geen samenhang tussen de scores van de respondenten bestaat valt te twisten. Er is voor dit onderzoek namelijk gebruik gemaakt van data afkomstig uit een huishoudenregister. Het is niet uit te sluiten dat respondenten die samenwonen elkaar niet beïnvloeden in de antwoorden die ze geven op vragenlijsten. Er is om deze reden gekeken naar de variabele ‘doet mee’ waar respondenten antwoord gaven op de vraag of een huisgenoot ook participeert in het LISS-panel. Hieronder volgt de output en van de descriptieve statistieken van deze variabele:

Household member participates in the panel

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	no	4	,3	,3	,3
	yes	1351	99,7	99,7	100,0
	Total	1355	100,0	100,0	



B3.1: Descriptieve statistieken en histogram van de variabele Doetmee

In bovenstaande output is te zien dat 99,7% van alle respondenten een huisgenoot heeft die ook participeert in het LISS-panel. Echter betekent dit niet dat deze respondenten dezelfde vragenlijsten hebben ingevuld. Ze participeren aan het LISS-panel, maar deze doet veel verschillende onderzoeken. Cijfers over het aantal respondenten dat aan dezelfde vragenlijst heeft meegewerkt en in hetzelfde huishouden woont, ontbreken echter in de gebruikte dataset. De aanname van onafhankelijke waarnemingen voor dit onderzoek dan ook als gedeeltelijk geschonden beschouwt. De tweede assumptie van logistische regressie betreft een lineair verband tussen de onafhankelijke variabelen en de logit van y. Er wordt vanuit gegaan dat deze assumptie niet geschonden is.

Multicollineariteit

Een van de assumpties die hoort bij logistische regressie stelt dat er geen of nauwelijks sprake mag zijn van multicollineariteit tussen de variabelen. Multicollineariteit houdt een te sterke samenhang tussen de variabelen in, waardoor de betrouwbaarheid afneemt. Om hierop te controleren wordt gebruik gemaakt van VIF scores. Wanneer de VIF waarde >4 is er sprake van multicollineariteit. Zoals te zien in onderstaande tabel is hier geen sprake van.

B3.2 Tabel met VIF-scores

Variabele	VIF-score
FysiekeBelasting	1,030
Partner	1,018
VerwachtInkomsten	1,029
Geboortejaar	1,037
Geslacht	1,102
Sector	1,103

De syntax en output voor de informatie in bovenstaande tabel is als volgt:

```
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA COLLIN TOL
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
```

/DEPENDENT Doorwerken
/METHOD=ENTER Geboortejaar Sector FysiekeBelastingR Partner VerwachtInkomenREC.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	5,294	4,361		1,214	,225		
	Respondent's year of birth	-,003	,002	-,055	-1,153	,250	,964	1,037
	[In what sector do you work?/In what sector did you work in your last job?]	-,002	,004	-,033	-,659	,510	,906	1,103
	FysiekeBelastingR	-,072	,024	-,141	-2,940	,003	,971	1,030
	The household head lives together with a partner (wedded or unwedded)	-,030	,028	-,051	-1,068	,286	,983	1,018
	VerwachtInkomenREC	,024	,015	,076	1,595	,111	,971	1,029
	Gender	-,039	,026	-,075	-1,521	,129	,907	1,102

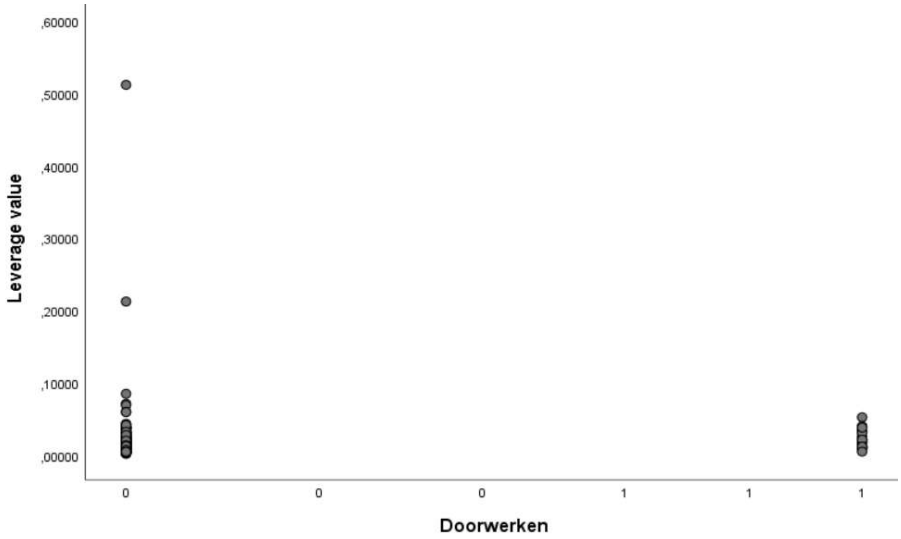
a. Dependent Variable: Doorwerken

Uitschieters

Ook is het van belang om te kijken naar mogelijke uitschieters, omdat deze de data erg kunnen vertekenen. Hiervoor is gekeken naar de leverage waarden. De leverage geeft aan in welke mate een bepaald punt bij wijze van spreken 'aan de lijn' trekt en hoe groot het effect van dat punt op de schattingen is. De syntax voor het verkrijgen van de Leverage waarden is als volgt:

```
LOGISTIC REGRESSION VARIABLES Doorwerken
 /METHOD=ENTER VerwachteInkomstenREC FysiekeBelastingR Partner DienstenHandel PubliekeSector
Overig Geslacht Geboortejaar
 /SAVE=LEVERAGE
 /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

Met deze waarden is vervolgens een ‘scatterplot’ gemaakt om de eventuele uitschieters te kunnen bekijken. Er zijn duidelijk twee uitschieters te zien (case ID 180 & 36)



B3.3 Scatterlot voor uitschieters door te kijken naar de Leverage waarden

De syntax hiervoor voor bovenstaande grafiek is als volgt:

```
/SCATTERPLOT(BIVAR)=Doorwerken WITH LEV_1
/MISSING=LISTWISE.
```

Vervolgens zijn deze 2 uitschieters uit de dataset gehaald en is de regressie analyse nog een keer gedaan om te kijken naar de invloed van deze twee punten op de analyse.

```
LOGISTIC REGRESSION VARIABLES Doorwerken
  /METHOD=ENTER Geboortejaar Geslacht DienstenHandel PubliekeSector Overig
  /METHOD=ENTER VerwachtInkomstenREC FysiekeBelastingR Partner
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

BLOK 1

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	11,746	5	,038
	Block	11,746	5	,038
	Model	11,746	5	,038

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	212,050[a]	,027	,066

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	9,652	8	,290

Contingency Table for Hosmer and Lemeshow Test

		Doorwerken = Nee		Doorwerken = Ja		
		Observed	Expected	Observed	Expected	Total
Step 1	1	43	44,821	3	1,179	46
	2	40	40,591	2	1,409	42
	3	42	41,321	1	1,679	43
	4	42	40,077	0	1,923	42
	5	43	43,681	3	2,319	46
	6	40	40,467	3	2,533	43
	7	44	40,842	0	3,158	44
	8	40	39,893	4	4,107	44
	9	38	38,857	6	5,143	44

10	34	35,451	9	7,549	43

Classification Table[a]

	Observed		Predicted: Doorwerken: Nee	Predicted: Doorwerken: Ja	Percentage Correct
Step 1	Doorwerken	Nee	406	0	100,0
		Ja	31	0	,0
	Overall Percentage				92,9

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	
Step 1[a]	Respondent's year of birth	-,048	,035	1,917	1	,166	
	Overig	-2,582	2,029	1,620	1	,203	
	PubliekeSector	-2,527	1,533	2,716	1	,099	
	DienstenHandel	-,476	,993	,230	1	,631	
	[In what sector do you work?/In what sector did you work in your last job?]	,189	,157	1,451	1	,228	
	Constant	91,591	68,341	1,796	1	,180	

BLOK 2

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	11,251	3	,010
	Block	11,251	3	,010
	Model	22,997	8	,003

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	200,799[a]	,051	,128

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	3,769	8	,877

Contingency Table for Hosmer and Lemeshow Test

		Doorwerken = Nee		Doorwerken = Ja		
		Observed	Expected	Observed	Expected	Total
Step 1	1	44	43,574	0	,426	44
	2	42	43,206	2	,794	44
	3	43	42,776	1	1,224	44
	4	42	42,400	2	1,600	44
	5	43	41,992	1	2,008	44
	6	42	41,469	2	2,531	44
	7	42	40,848	2	3,152	44
	8	39	39,728	5	4,272	44
	9	38	38,293	6	5,707	44
	10	31	31,714	10	9,286	41

Classification Table[a]

			Predicted		
			Doorwerken		Percentage
	Observed		Nee	Ja	Correct
Step 1	Doorwerken	Nee	406	0	100,0
		Ja	31	0	,0
	Overall Percentage				92,9

a. The cut value is ,500

		B	S.E.	Wald	df	Sig.
Step 1[a]	Respondent's year of birth	-,051	,036	2,008	1	,157
	Overig	-2,087	2,162	,932	1	,334
	PubliekeSector	-2,084	1,607	1,682	1	,195
	DienstenHandel	-,444	1,091	,166	1	,684
	[In what sector do you work?/In what sector did you work in your last job?]	,151	,168	,807	1	,369
	FysiekeBelastingR	-1,466	,571	6,596	1	,010
	The household head lives together with a partner (wedded or unwedded)	-,515	,434	1,408	1	,235
	VerwachtInkomenREC	,166	,150	1,234	1	,267
	Constant	98,671	70,038	1,985	1	,159

Wat opvalt aan bovenstaande informatie is voornamelijk het effect van de sectoren Overig en Publieke Sector. Dit effect is namelijk veel sterker geworden na het verwijderen van de twee cases, echter blijft dit effect niet significant. Ook het effect van partner lijkt veel groter na het verwijderen van de twee cases, maar is weer niet significant getoetst. Het effect van Fysieke Belasting blijft ook in het model zonder de twee uitbuiters significant en vrij sterk. Qua modelfit heeft model 5 nog steeds de laagste 2 log likelihood (Deviance) waarde en blijft dus het beste model ,omdat dit model de minste onverklaarde variantie bevat. Ook de Hosmer-Lemeshow test, waarbij een chi-kwadraat waarde berekend, gebaseerd op de geobserveerde en verwachte frequenties, laat een niet-statistische significante uitkomst van deze test zien dat het model goed past. Uit bovenstaande informatie kan geconcludeerd worden dat de twee uitbuiters een invloed hebben op de effecten van sommige variabelen. Echter wordt deze invloed voor dit onderzoek als verwaarloosbaar beschouwt, aangezien er niks verandert aan de significantie van deze effecten en de modelfit van het uiteindelijke model ook bijna niet verandert