



rijksuniversiteit
groningen

*LANGER DOORWERKEN IN NEDERLAND: HOE
VERSCHILLEN DE WENSEN EN VERWACHTINGEN VAN
OUDERE WERKENDEN?*

Bachelorscriptie

23-06-2022

Max van der Meide

S350457, m.y.van.der.meide@student.rug.nl

Begeleider: dr. R.M.C Smaniotto

Tweede lezer: dr. J. Dijkstra.

Abstract

Door een in rap tempo vergrijzende samenleving verandert de verhouding tussen werkenden en gepensioneerden sterk. Dit dreigde te leiden tot problemen op de arbeidsmarkt en een problematische betaalbaarheid van het socialezekerheidsstelsel. Beleidsmatige hervormingen waren noodzaak. Het Kabinet sloot samen met haar sociale partners een nieuw Pensioenakkoord (2019). De centrale boodschap is evident: Nederlanders moeten gradueel langer doorwerken. Dit heeft gevolgen voor de wijze waarop werkenden het proces van pensioneren voor zich zien. Met dit proces zijn twee concepten sterk verbonden: de verwachte en gewenste pensioenleeftijd. Eerder onderzoek toont aan dat het verschil tussen de verwachte en gewenste pensioenleeftijd -door het geheel aan beleidsmatige hervormingen- groter is geworden en tot uiting komt in ontevredenheid en frustraties over het pensioenbeleid. Het doel van deze thesis (N=1149) was te onderzoeken hoe groot dit verschil is en waar deze verschillen mee samenhangen. Hiervoor werd gekeken naar sociaaleconomische factoren en individuele baankenmerken. De data is afkomstig van de werk- en scholingsmethode van het LISS-Panel en was verzameld op basis van digitale gestructureerde vragenlijsten. In deze thesis bedroeg het verschil tussen de verwachte en gewenste pensioenleeftijd van oudere werkenden gemiddeld 3 jaar en 110 dagen. Uit een lineair hiërarchische regressieanalyse is gebleken dat dit verschil verklaard kon worden door de mate waarin het werk als fysiek belastend werd ervaren (H3) en de mogelijkheid deel te kunnen nemen aan trainingsactiviteiten (H4). Ondanks de brede ondersteuning in de literatuur kon het verschil niet worden verklaard door het inkomen (H1) en opleidingsniveau (H2). Door de beperkingen van deze thesis was voorzichtigheid gebaat bij het doen aanbevelingen. Niettemin is toekomstig onderzoek meer dan ooit aan de orde van dag, gezien het feit dat de pensioenleeftijd voor toekomstige generaties onzeker blijft. Tot slot is door het ontbreken van passende regelingen voor bepaalde groepen op de arbeidsmarkt de angel nog lang niet uit de pensioenenkwestie.

Inhoudsopgave

1. Inleiding	3
2. Theoretisch kader	6
2.1. <i>Sociaaleconomische factoren: inkomen en opleidingsniveau</i>	6
2.1.1. Inkomen	6
2.1.2. Opleidingsniveau	7
2.2. <i>Individuele baankenmerken: Fysieke belasting en trainingsmogelijkheden</i>	7
2.2.1. Fysieke belasting	7
2.2.2. Trainingsmogelijkheden	8
2.3. <i>Controlevariabelen</i>	9
3. Methoden & Dataverzameling	10
3.1. <i>Beschrijving van de deelnemers van het onderzoek</i>	10
3.1.1. Het databestand	10
3.1.2. Steekproeftrekking en non-respons	11
3.1.3. Selectie op het databestand en item non-respons	11
3.2. <i>Onderzoeksdesign en Procedure</i>	12
3.3. <i>Operationalisaties</i>	12
3.3.1. Onafhankelijke variabele	12
3.3.2. Verklarende variabelen	13
3.3.3. Controlevariabelen	14
3.4. <i>Analysebeschrijving</i>	15
4. Resultaten	16
4.1. <i>Beschrijvende statistieken</i>	16
4.1.1. Univariate beschrijvende statistieken	16
4.1.2. Bivariate beschrijvende statistieken	18
4.2. <i>Modevaluatie</i>	20
4.2.1. Modelfit	20
4.2.2. Multicollineariteit	21
4.2.3. Uitbijters	21
4.2.4. Assumptietoetsing	22
4.3. <i>Hypothesetoetsing</i>	22
4.3.1. Inkomen	22
4.3.2. Opleidingsniveau	22
4.3.3. Fysieke Belasting	23
4.3.4. Trainingsmogelijkheden	23
4.3.5. Controlevariabele	24
5. Conclusie en Discussie	24
6. Literatuurlijst	28
Bijlage 1	32
Bijlage 2	64
Bijlage 3	89

1. Inleiding

Als gevolg van de snelle vergrijzing van de samenleving staat de Nederlandse arbeidsmarkt aanhoudend voor grootschalige uitdagingen. Een dalend geboortecijfer en een toenemende levensverwachting dreigen te leiden tot een tekort in de (potentiële) beroepsbevolking. Indien de pensioengerechtigde leeftijd – momenteel: 66 jr., 7 mnd. - niet zou stijgen, zou het aantal pensioengerechtigden tot 2040 met 50% toenemen, terwijl de omvang van de potentiële beroepsbevolking met 5% zou dalen (De Beer & Van der Gaag, 2019). Evenzeer heeft dit gevolgen voor de Algemene Ouderdomswet: de verhouding tussen werkenden en gepensioneerden verandert sterk en dat heeft zijn weerslag op het socialezekerheidsstelsel. Dergelijke problemen vragen om beleidsmatige hervormingen: het Kabinet sloot samen met haar sociale partners een nieuw Pensioenakkoord (2019). De centrale boodschap van dit Akkoord is evident: Nederlanders moeten (gradueel) langer doorwerken. Dit heeft gevolgen voor de wijze waarop oudere werkenden het proces van pensionering voor zich zien (Becker, et al. 2021). Aan dit proces zijn een aantal vragen onlosmakelijk verbonden: wanneer verwachten oudere werkenden met pensioen te kunnen gaan? En verschilt deze verwachte pensioenleeftijd van de leeftijd waarop oudere werkenden wensen met pensioen te gaan? In deze thesis zal ik op die vragen ingaan en proberen mogelijke verklaringen te geven.

Voordat ik het hedendaagse proces van pensioneren voor oudere werkenden toelicht, baken ik kort twee concepten af die centraal staan binnen dit proces: *gewenste* en *verwachte* pensioenleeftijd. De *gewenste* pensioenleeftijd is de leeftijd waarop een individu zou willen pensioneren wanneer diegene dat zelf voor het zeggen zouden hebben. De *verwachte* pensioenleeftijd is een realistische schatting van de leeftijd waarop een individu verwacht met pensioen te kunnen gaan. Ondanks dat beide concepten zuiver subjectief zijn van aard, is de *verwachte* pensioenleeftijd - in tegenstelling tot de *gewenste* pensioenleeftijd - ingebed in contextuele macrofactoren zoals het pensioensysteem. Als gevolg van veranderingen aan het pensioensysteem van Nederland de afgelopen jaren is de leeftijd waarop oudere werkenden met pensioen verwachten te gaan gestegen, terwijl de wens om later met pensioen te gaan slechts beperkt meestijgt (de Beer & van der Gaag, 2019; Visser, 2020).

Het is aannemelijk dat oudere werkenden (>45 jr.) verwachten later met pensioen te kunnen gaan ten gevolge van het geheel aan beleidsmatige hervormingen. Na decennia van minimale veranderingen aan het pensioensysteem werd in 2006 afscheid genomen van pre-pensioenregelingen, waarvan de Vervroegde Uittreding (VUT) het meest bekend en gebruikt. Door het afschaffen van deze pre-pensioenregelingen nam de *verwachte* pensioenleeftijd al sterk toe (De Beer & Van der Gaag, 2019). Nadien volgden diverse (de-) regulaties met als summum het loslaten van de vaste AOW-leeftijd van 65 jaar. In termijnen – definitief vastgelegd in het Pensioenakkoord (2019) - wordt de AOW-leeftijd verhoogd van 65 naar 67 jaar in 2024, en daarna gekoppeld aan ontwikkelingen van de levensverwachting. Hierbij zal de AOW-leeftijd met 8 maanden stijgen voor elk vol jaar stijging in de

levensverwachting op de leeftijd van 65 jaar. Deze trend van pensioenhervormingen is zichtbaar in veel andere Europese landen (Hess, 2016).

Het is algemeen bekend en breed gedragen in de wetenschappelijk literatuur dat deze hervormingen van het pensioensysteem onvrede oproepen (De Beer, et al., 2017; Schippers, 2019; Visser, 2020). Een deel van de oudere werkenden – Visser spreekt over 65 procent - in Nederland is ontevreden met de hervormingen en wensen niet (gradueel) langer door te hoeven werken. Ten minste een deel van de beroepsbevolking is dus toornig over de stijging van de wettelijk pensioengerechtigde leeftijd, zij zouden voor het bereiken van die leeftijd willen stoppen met werken. Gebaseerd op deze onderzoeksresultaten lijken er discrepanties te zijn tussen de verwachte en gewenste pensioenleeftijd van de oudere werkenden in Nederland. Bij de eerste oogopslag zijn deze verschillen onschuldig.

Echter, er schuilt veel structurele ongelijkheid achter deze eerste oogopslag (Schippers, 2019). Verschillende onderzoeken (Henkens, et al. 2016; Hess, 2018) laten zien dat langer doorwerken bij een aanzienlijke groep, bijvoorbeeld bij werkenden met een lager opleidingsniveau of fysiek belastend werk, niet vanzelfsprekend is. Deze verschillen zijn onwenselijk, maar eveneens waardevol om onderzoek naar te verrichten. Bepaalde groepen werkenden kunnen hun wensen niet inwilligen en dit zorgt voor ontevredenheid en frustratie. In het uiterste geval leiden deze ontevredenheid en frustratie tot een afname van de levenskwaliteit (Oude Mulders, et al. 2021).

In eerder onderzoek naar het Nederlandse pensioensysteem werd veelal gebruik gemaakt van de wettelijk gerechtigde pensioenleeftijd als objectieve maatstaf. Deze leeftijd zal voortaan voor iedere 5 jaar worden vastgesteld (De Beer & van der Gaag, 2019). Aangezien het onduidelijk is hoe sterk de levensverwachting zal veranderen, is de hoogste van deze leeftijd voor toekomstige generaties onzeker. Deze thesis onderscheidt zich door toe te leggen op individuele wensen en verwachtingen die oudere werkenden hebben in plaats van op een - voor bepaalde leeftijdscohorten - mogelijk onjuiste wettelijk gerechtigde pensioenleeftijd.

Aangenomen dat er discrepanties bestaan is het van belang om te onderzoeken waar de discrepanties mee samenhangen. De concepten waar ik naar zal kijken zijn de sociaaleconomische factoren: *inkomen en opleidingsniveau* en de individuele baankenmerken: *fysieke belasting van het werk en de scholingsmogelijkheden*.

Inkomensverschillen hebben hun weerslag op het pensioengedrag. Mensen met een hoger inkomen kunnen meer invloed uitoefenen op hun proces van pensioneren (Hess, 2018). Over het algemeen hebben zij een hoger aanvullend pensioen en meer spaargeld. Hierdoor hebben zij financieel meer draagvlak om te kunnen pensioneren wanneer zij dat zouden willen (Steiber & Kohli, 2015). Eerder onderzoek heeft differente resultaten opgeleverd: met name bij hogere inkomens lopen de wensen en verwachtingen verder uiteen. De tweede factor is het opleidingsniveau. De kloof tussen hoog- en laagopgeleide mensen vertaalt zich in verschillend pensioengedrag tussen deze groepen (CPB, 2016). Omdat de duur van de onderwijsdeelname verschilt tussen een hoog- en laagopgeleiden, zit er snel ruim 5 jaar verschil tussen de leeftijd waarop beide groepen de arbeidsmarkt betreden (Schippers, 2019).

Recente onderzoeken (de Beer & Van der Gaag, 2019; Schippers, 2021) laten zien dat veel onvrede en frustratie over het pensioenbeleid volgt uit dit verschil.

Het eerste baankenmerk is de fysieke belasting van het werk. Jarenlang fysiek belastend werk heeft gevolgen voor de gezondheidstoestand. Hierdoor zijn oudere werkenden met fysiek zwaar werk minder goed in staat hun werkzaamheden te verrichten (De Beer, et al. 2018). Eerder onderzoek toont dat werkenden met fysiek belastend werk over het algemeen niet bereid zijn om langer door te werken; zij wensen eerder met pensioen te gaan (Sell, et al., 2009; Topa, et al., 2009). Tot slot is het tweede baankenmerk de mogelijkheid om scholing te volgen. Onder het conto van leven lang leren en ontwikkelen worden (vaardigheids-) trainingen in toenemende mate aangeboden aan oudere Nederlandse werkenden (Schippers, 2019). Oudere werkenden die kunnen deelnemen aan trainingsactiviteiten ontwikkelen een hogere arbeidsmotivatie en zijn meer toegewijd aan hun organisatie. Dit vertaalt zich in de wens om langer door te blijven werken (Rhoades & Eisenberger, 2002).

Al met al lijken discrepanties te bestaan tussen de gewenste en verwachte pensioenleeftijd van oudere werkenden in Nederland. Deze discrepanties lijken tot uiting te komen in frustratie en ontevredenheid over de pensioenhervormingen. Er lijken sociaaleconomische factoren en individuele baankenmerken samen te hangen met deze discrepanties. Hieruit volgt de volgende onderzoeksvraag:

- *Hoe groot is het verschil tussen gewenste en verwachte pensioenleeftijd van oudere werkenden in Nederland, en in hoeverre wordt dit verklaart door sociaaleconomische factoren en baankenmerken?*

2. Theoretisch kader

In dit onderzoek staat het verschil tussen leeftijd waarop een oudere werkende wenst en verwacht te pensioneren centraal. Binnen de huidige wetenschappelijke literatuur bestaat geen consensus over de definitie van pensioneren. Door de veranderende aard van pensioneren naar een steeds complexer en dynamischer proces dat sterk verschilt tussen individuen neemt de verdeeldheid over accurate definitie toe (Furunes, et al. 2015; Schultz & Olson, 2013). In deze thesis beschouw ik pensioneren als het volledig stoppen met betaalde arbeid (Oude Mulders, 2019).

2.1. Sociaaleconomische factoren: inkomen en opleidingsniveau

2.1.1. Inkomen

Mensen met een hoog inkomen hebben meer keuzevrijheid in het proces van pensioneren. Tijdens de werkzame periode hebben werkenden met een hoger inkomen over het algemeen meer aanvullend pensioen opgebouwd en zij hebben meer spaargeld (Steiber & Kohli, 2015). Deze bevinding volgt uit vergelijkende analyse van diverse Europese pensioenstelsels, maar gaat eveneens op voor Nederland in het specifiek. Nederland kent naast de overheidsvoorziening AOW - het basispensioen voor mensen die de pensioengerechtigde leeftijd hebben bereikt - een pensioenstelsel met aanvullende collectieve pensioenen en de individuele verzekering die mensen zelf treffen (SVB, 2022). Hierbij wordt de AOW ook wel de eerste pijler van het systeem genoemd en de anderen respectievelijk pijler twee en drie. Het Nederlandse systeem onderscheidt zich van dat van andere landen door een –vrijwel altijd- verplichte opbouw van pensioen op basis van een percentage van het looi. In andere gevallen is de opbouw van het pensioen vrijblijvender en in een mindere mate gereguleerd. Naarmate het inkomen van een werknemer hoger ligt tijdens zijn werkzame leven bouwt deze dus over het algemeen meer pensioeninkomen op uit de tweede en derde pijler (De Beer, 2016). Zodoende hebben werkenden met een hoog inkomen vooruitzicht op een hoger inkomen uit het pensioen. Daarnaast hebben werkenden in hoge inkomensgroepen meer financiële bezittingen - waaronder spaargeld - dan zij in lage inkomensgroepen (CPB, 2021).

Logischerwijs heeft dit gevolgen voor het verschil in de leeftijd waarop een oudere werkenden wensen en verwachten te pensioneren. Het is aannemelijk dat werkenden die een hoog inkomen genieten in bevredigendere mate hun wensen kunnen inwilligen en aan hun individuele verwachting voldoen dan werkenden met een lager inkomen, omdat bij die groep de financiële armslag ontbreekt. Dit blijkt dan ook het geval: individuen met een hoog inkomen verwachten en gaan eerder met pensioen dan zij met een lager inkomen (Hayward, 1986; Montizaan, 2017).

Al met al hebben oudere werkenden met een hoog inkomen meer keuzevrijheid in het proces van pensioneren, waardoor de discrepanties kleiner zullen zijn naarmate een werkende een hoger inkomen heeft. Hieruit volgt de volgende hypothese (H1):

- *Naarmate een oudere werkende een hoger inkomen heeft, zal de discrepantie tussen verwachte en gewenste pensioenleeftijd kleiner zijn.*

2.1.2. Opleidingsniveau

Allereerst verwachten mensen met een laag opleidingsniveau eerder te pensioneren, omdat zij op eerdere leeftijd toetreden tot de arbeidsmarkt. In deze thesis worden werkenden met een afgeronde Hbo of WO-opleiding gekwalificeerd als hoogopgeleid. De onderwijsduur van hoog- en laagopgeleiden verschilt sterk, waardoor hoogopgeleiden al snel pas vijf jaar later toetreden tot de arbeidsmarkt (Schipper, 2019).

Daarnaast hebben mensen met een hoog opleidingsniveau attractievere baankenmerken. Hoogopgeleid werk wordt gekenmerkt door een hogere mate van autonomie; hoogopgeleiden mogen vaker hun eigen werktijden bepalen en zijn onafhankelijker van leidinggevendenden waardoor zij in hogere mate eigen keuzes mogen maken (Wheatley, 2017). Mensen met autonoom werk hebben over het algemeen een hogere baantevredenheid. Deze baantevredenheid uit zich in de wens om later met pensioen te gaan dan mensen met een lagere baantevredenheid (Potocnik, et al., 2009; Pota, et al., 2009).

Bovendien hebben mensen met een hoog opleidingsniveau over het algemeen betere werkomstandigheden: hoogopgeleiden hoeven minder vaak fysiek belastend werk te verrichten. De invloed van fysiek belastend werk op wensen en verwachtingen omtrent de pensioenleeftijd wordt nader toegelicht in paragraaf 2.2.1. Vooralsnog lijken mensen met een hoog opleidingsniveau te wensen en verwachten later met pensioen te gaan. Hieruit volgt de volgende hypothese (H2):

- *Naarmate het opleidingsniveau van een oudere werkende hoger ligt, zal de discrepantie tussen verwachte en gewenste pensioenleeftijd kleiner zijn.*

2.2. **Individuele baankenmerken: Fysieke belasting en trainingsmogelijkheden**

2.2.1. Fysieke belasting

Fysieke belasting is de letterlijke belasting die het lichaam ondervindt aan het te verrichten werk. Mensen met fysiek belastend werk hebben de wens om eerder met pensioen te gaan dan verwacht (Grolsch & Pansky, 2009). Uit onderzoek van De Beer en Van der Gaag (2019) blijkt dat het voor werkenden met fysiek belastend werk moeilijker is om door te werken tot de pensioengerechtigde leeftijd, simpelweg door slijtage aan het lichaam. Door een verminderde gezondheid – als gevolg van

het werk- zijn mensen met fysiek belastend werk dus minder goed in staat om langer door te werken. Tevens gaan individuen met fysiek belastend werk sneller achteruit in hun functionele gezondheid in de jaren voor pensionering (Van Zon, 2016). De uitdaging om het fysiek belastende werk te verrichten wordt dus groter naarmate het pensioen nadert. Zodoende hebben mensen met fysiek belastend werk over het algemeen een mindere gezondheidstoestand dan mensen met werk dat fysiek niet belastend is, waarbij de gezondheidsverschillen toenemen naarmate de leeftijd stijgt.

Door een verminderde gezondheidstoestand – resulterend in een problematische werkbaarheid van hun baan - zullen mensen met fysiek belastend werk onvoldoende in staat zijn om door te werken. Hierdoor ligt hun gewenste pensioenleeftijd lager (Hess, 2016, 2019). Echter, door de uniform stijgende AOW-leeftijd – er wordt niet gedifferentieerd op basis van baankenmerken - stijgt de verwachte pensioenleeftijd van mensen met fysiek belastend werk. Dit resulteert in een grotere discrepantie tussen gewenste- en verwachte pensioenleeftijd naarmate het werk als fysiek meer belastend wordt ervaren. Hieruit volgt de volgende hypothese (H3):

- *Naarmate het werk als fysiek meer belastend wordt ervaren, zal de discrepantie tussen verwachte en gewenste pensioenleeftijd groter zijn.*

2.2.2. Trainingsmogelijkheden

Door het aanbieden van trainingen kunnen oudere werkenden hun kennis en vaardigheden ontwikkelen. Hun productiviteit zal niet dalen als het gevolg van verouderde kennis en vaardigheden. Naarmate mensen meer kennis hebben over en vaardigheden in hun specifieke vakgebied ervaren zij hun werk als waardevoller (Froidevaux & Hirschi, 2015). Zodoende ervaren oudere werkenden die deel kunnen nemen aan trainingsactiviteiten – door het ontwikkelen en bijtijds houden van kennis en vaardigheden - hun werk als waardevol(-ler).

Daarnaast zorgt het aanbieden van trainingsmogelijkheden voor een gevoel van erkenning bij oudere werkenden. De mogelijkheid om deel te nemen aan trainingen was voor oudere werkenden tot voor de eeuwwisseling verre van vanzelfsprekend. Vanuit economische overwegingen was – en incidenteel is – het niet rationeel om trainingsmogelijkheden aan te bieden aan oudere werkenden (Tsui, et al., 1997; Schippers, 2019). Door deze economische barrières ervaren oudere werkenden het aanbieden van trainingsmogelijkheden als een sterk signaal van organisationele erkenning (Armstrong-Stassen & Ursel, 2009). Immers, de keuze om trainingen aan te bieden aan oudere werkenden vervangt een ‘ingesleten patroon’ waarbinnen wordt gekozen voor jongere werkenden in plaats herscholing voor oudere werkenden.

Dit gevoel van erkenning vanuit de organisatie leidt tot een hogere arbeidsmotivatie en toegewijdsheid aan de organisatie (Rhoades & Eisenberger, 2002). De arbeidsmotivatie en toegewijdsheid van de oudere werkenden is hoger, omdat de ervaren erkenning het wederkerigheidsgevoel versterkt om zich in te blijven zetten voor de organisatie. Met andere woorden: in ruil voor de welwillendheid van

organisaties om trainingsmogelijkheden aan te bieden, zijn oudere werkenden – al dan niet bewust – bereid tot wederdiensten in de vorm van toegenomen arbeidsmotivatie en toegewijdheid. Zodoende hebben oudere werkenden die de mogelijkheid krijgen om deel te nemen aan trainingsactiviteiten – door een hogere arbeidsmotivatie en toegewijdheid aan de organisatie - over het algemeen de wens om langer door te werken.

In tegenstelde richting blijkt uit onderzoek van Damman en collega's (2013) dat oudere werkenden die niet de mogelijkheid hebben om aan trainingsactiviteiten deel te nemen de neiging hebben om hun toegewijdheid aan de organisatie te verliezen: zij zijn minder gemotiveerd en wensen eerder met pensioen te gaan.

Al met al hebben oudere werkenden die de mogelijkheid hebben om aan trainingsactiviteiten te volgen – als het gevolg van een groter menselijk kapitaal en een hogere arbeidsmotivatie en toegewijdheid aan de organisatie – de wens om langer door te werken. De gewenste pensioenleeftijd zal hoger zijn dan wanneer er vaker mogelijkheden worden geboden om deel te nemen aan trainingen. Hierdoor is de discrepantie voor deze groep kleiner. Hieruit volgt de volgende hypothese (H4):

- *Naarmate er vaker mogelijkheden bestaan om trainingsactiviteiten te volgen, zal de discrepantie tussen gewenste en verwachte pensioenleeftijd kleiner zijn.*

2.3. Controlevariabelen

Om de invloed van andere indicatoren te verkleinen zal ik controleren voor een vijftal indicatoren. Allereerst voor de gebruikelijke demografische kenmerken leeftijd en geslacht. Daarnaast voor de beroepssector. De wensen en verwachtingen van oudere werkenden zijn afhankelijk van de sector waar zij in werken (Schippers, 2021). Per sector kunnen de baankenmerken en financiële beloningen sterk verschillen. Tevens wordt gecontroleerd voor het hebben van een partner. Uit onderzoek van Van Dam en collega's (2009) blijkt dat de het hebben van een partner (gehuwd of ongehuwd) een invloedrijke rol speelt in het pensioengedrag. De wensen en verwachtingen van werkenden worden vaak afgestemd op die van de partner. Ten slot zal ik controleren voor het hebben van kinderen. Het opvoeden van kinderen is een kostbaar en tijdrovend proces. Om financiële redenen moeten oudere werkenden met kinderen dus langer doorwerken (Steiber & Kohli, 2015). Oudere werkenden zonder kinderen ervaren dit proces niet.

3. Methoden & Dataverzameling

In deze paragraaf heb ik de methoden toegelicht die worden gebruikt voor deze thesis. Allereerst heb ik een databeschrijving gegeven, inclusief informatie over de steekproeftrekking, werving en non-respons en items non-respons (§3.1). Vervolgens heb ik het onderzoeksdesign beschreven (§3.2), gevolgd door een beschrijving van de operationalisaties (§3.3). Tot slot de beschrijving van de statistische analyse (§3.4).

3.1. Beschrijving van de deelnemers van het onderzoek

3.1.1. Het databestand

In deze thesis zal gebruik worden gemaakt van secundaire data afkomstig uit de longitudinale werk- en scholingsmodule van het LISS-panel (Longitudinal Internet Study for Social Sciences). Het LISS-panel kent ongeveer 7500 leden, allen inwoner van Nederland. Dit panel is probability-based: het is gebaseerd op een probabilistische steekproef van huishoudens uit het bevolkingsregister van het Centraal Bureau voor de Statistiek. Het is niet mogelijk jezelf aan te melden als panellid. Zodoende wordt getracht de samenstelling en representativiteit van het panel te waarborgen.

Het doel van het LISS-panel is inzichten verschaffen in de veranderingen in het leven van Nederlanders, hun reactie op levensgebeurtenissen en het meten van effecten van maatschappelijke verandering en beleidsmaatregelen. Hierbij focust de werk- en scholing module – een van de acht LISS-modules – op arbeidsmarkt(-participatie), pensioen(-plannen), individuele baankenmerken en scholing. In deze thesis zal de dertiende wave van deze module worden gebruikt. De onderzoekspopulatie van deze module is alle mannelijke- en vrouwelijke inwoners van Nederland ouder dan 15 jaar.

De data werden verzameld aan de hand van een internet-survey. Benaderde panelleden konden digitaal een vragenlijst invullen. Panelleden die door een gebrek aan digitale infrastructuur anders niet mee zouden kunnen doen, werden hierin voorzien. Panelleden ontvingen een financiële vergoeding voor het invullen van de vragenlijsten.

Er werden voor deze module twee separate metingen verricht, zodat non-respondenten van de eerste meting nogmaals konden worden benaderd. De eerste meting vond plaats tussen 13 en 28 april en de tweede meting vond plaats tussen 4 en 26 mei. Beide in het jaar 2020. Naast de vragen uit de werk- en scholing module hebben respondenten vragen ingevuld aangaande achtergrondinformatie. Deze data bevat de meest recente (van dat moment) informatie over achtergrondkenmerken, zoals demografische- en sociaaleconomische kenmerken van de respondenten.

3.1.2. Steekproeftrekking en non-respons

De vragenlijst van de werk- en scholingsmodule werd voorgelegd aan 6.890 panelleden. Eerder beschreef ik dat het LISS-panel bestaat uit ongeveer 7.500 leden. Er is geen informatie beschikbaar over de wijze waarop en waarom door het Panel dit specifieke aantal leden is gekozen. Na de twee meetmomenten hadden 5646 panelleden (81,9%) de vragenlijst ingevuld. De non-respons bedraagt 1244 panelleden (18,1%). De onderzoekers geven geen additionele informatie over de groep non-respondenten. Uiteindelijk bestaat het databestand uit 5578 respondenten (N=5578), de ingevulde vragenlijsten van 68 respondenten werd als incompleet beschouwd (1.2%). De onderzoekers lichten niet toe wanneer een vragenlijst incompleet is ingevuld.

3.1.3. Selectie op het databestand en item non-respons

Voor de resultatenanalyse was er een selectie gemaakt op de oorspronkelijke dataset (N=5578). Deze subset (N=1149) werd gebruikt voor de statistische analyse. In deze subset bevonden zich alleen respondenten zonder missende waarden, d.w.z. zij hebben op alle variabelen in de thesis een valide score, die ouder zijn dan 44 jaar en met een positieve score op de afhankelijke variabele. Er werden 1949 (34.9%) respondenten weggelaten, omdat zij jonger zijn dan 45. Dit is inherent aan de vragenlijst van de werk- en scholing module: de vragen omtrent gewenste en verwachte pensioenleeftijd waren alleen gesteld aan werkende respondenten ouder dan 44 jaar. Deze 'oudere generaties' werkenden zitten in arbeidsjaren het kortst op de pensioengerechtigde leeftijd en zijn de eerste generatie(s) die onderhevig zijn aan de beleidsmatige hervormingen van pensioensysteem. Dit maakt het interessant om inzichten te krijgen in de verwachtingen en wensen omtrent pensioneren van deze groep oudere werkenden. Zodoende bestaat de onderzoekspopulatie van deze thesis uit de werkende Nederlandse bevolking van 44 jaar of ouder. De steekproefpopulatie van deze thesis omvat alle werkende – c.q. niet gepensioneerde- werk- en scholing panelleden ouder dan 44 jaar.

De subset bevat alleen respondenten zonder missende waarden, d.w.z. zij hebben op alle variabelen in de thesis een valide score. Dit waren 1219 respondenten. Kijkend naar de item non-respons was er een patroon zichtbaar. Veel respondenten hadden een missende waarde op tenminste één ondersteunde variabele van de discrepantie in pensioenleeftijd. De verwachte en gewenste pensioenleeftijd hadden respectievelijk 2498 en 2454 missende waarden. Respondenten met een item non-respons op deze variabelen scoorden significant verschillend op leeftijd en inkomen. Mensen met een item non-respons op de verwachte pensioenleeftijd waren gemiddeld 67 jaar en hadden een maandelijks netto-inkomen van 1775 euro, terwijl respondenten gemiddeld 56 jaar zijn met een netto-inkomen 2283 euro. Mensen met een item-non respons op de gewenste pensioenleeftijd waren gemiddeld 67.5 jaar en hun maandelijks netto-inkomen bedroeg 1773 euro, terwijl respondenten gemiddeld 55 jaar waren met een netto-inkomen van 2283 euro. Mensen scoorden niet significant verschillend op het opleidingsniveau, de ervaren fysieke belasting van het werk en de mogelijkheid deel

te nemen aan trainingsactiviteiten. Al met al waren er aanzienlijke verschillen in gemiddelde scores tussen de respondenten en non-respondenten op bepaalde variabelen. Dit zal het uiteindelijk problematischer maken om in de conclusieparagraaf uitspraken te doen over de gehele populatie.

Tot slot, in de subset worden alleen respondenten meegenomen met een positieve verschillscore op de discrepantie in pensioenleeftijd. Er waren 70 respondenten met een negatieve score. Nadat de negatieve score uit subset zijn weggelaten, bevatte de subset 1149 respondenten (N=1149). In de operationalisatie van deze variabele heb ik toegelicht waarom alleen de positieve scores worden meegenomen in de statistische analyse.

3.2. Onderzoeksdesign en Procedure

De data van zowel de werk- en scholing module als de achtergrondinformatie zijn verzameld aan de hand van een vragenlijstonderzoek. De vragenlijsten waren gestructureerd en zijn digitaal afgenomen. Respondenten werden benaderd middels een e-mail. Indien zij na zeven dagen de vragenlijst niet hadden ingevuld ontvingen zij één (zevende dag na start meting) of twee (veertiende dag na start meting) herinneringsmails. Ondanks dat het codeboek in beschikbaar is in het Engels, waren de vragenlijsten voor respondenten alleen beschikbaar in het Nederlands.

De gestructureerde vragenlijst van de werk- en scholing module bestond uit vragen over arbeidsmarktparticipatie, individuele baankenmerken, pensioenen, scholing en cursussen. Respondenten deden er gemiddeld 30 tot 50 minuten over om de vragenlijst in te vullen. De gestructureerde vragenlijst over de achtergrondinformatie bestond uit vragen over demografische- en sociaaleconomische kenmerken van de respondent. Er is geen informatie beschikbaar over tijd die het invullen van deze vragenlijst in beslag nam. Aangezien panelleden herhaaldelijk vragenlijsten invullen voor het Panel bestond de mogelijkheid aan te geven dat de achtergrondinformatie niet was veranderd. Voor het databestand werd dan reeds verkregen achtergrondinformatie over de respondenten gebruikt.

3.3. Operationalisaties

In deze deelparagraaf volgt een beschrijving van alle meetinstrumenten die zijn gebruikt om de hypothesen te toetsen, inclusief de controlevariabelen. Beginnend met de afhankelijke variabele, gevolgd door de verklarende variabelen en afgesloten met de controlevariabelen. In deze deelparagraaf zijn bewerkingen aan meetinstrumenten beknopt beschreven. Een uitgebreide toelichting op bewerkingen is te vinden in de eerste bijlage.

3.3.1. Onafhankelijke variabele

1. Discrepantie in pensioenleeftijd

De discrepantie in pensioenleeftijd is de onafhankelijke variabele in het onderzoeksmodel. De variabele is de gemeten door een verschillscore tussen de verwachte pensioenleeftijd en de gewenste pensioenleeftijd: 'Discrepantie in pensioenleeftijd = verwachte pensioenleeftijd – gewenste pensioenleeftijd'. Het meetniveau is ratio. Des te hoger de verschillscore van een respondent is, des te

groter is het verschil tussen de leeftijd waarop de respondent verwacht de pensioneren en de leeftijd waarop wordt gewenst te pensioneren.

In de analyse worden alleen de positieve verschillen – d.w.z. respondenten waarvan de verwachte pensioenleeftijd hoger ligt dan de gewenste pensioenleeftijd - meegenomen. In deze thesis staan de hervormingen van pensioensysteem, waardoor Nederlanders (gradueel) langer moet doorwerken, centraal. Uit eerder onderzoek (De Beer, et al., 2017; Schippers, 2019; Visser, 2020) blijkt dat veel Nederlanders ontevreden over deze hervormingen. In deze thesis geef ik hier mogelijke verklaringen voor aan de hand van sociaaleconomische factoren en individuele baankenmerken. Respondenten met een negatieve verschilscore (= 70 respondenten) zouden evengoed ontevreden kunnen zijn, maar zijn dat niet over de recente hervormingen waarbij Nederlanders langer door moeten werken. Immers, zij wensen langer door te werken dan dat zij verwachten. Zodoende worden deze respondenten niet meegenomen in de analyse.

2. Gewenste pensioenleeftijd

De gewenste pensioenleeftijd is gemeten door de vraag: “Als u zelf zou (of had) mogen kiezen, op welke leeftijd zou u dan willen stoppen met werken?” Respondenten konden hierop antwoorden met een leeftijd tussen de 0 en 100 jaar. Het meetniveau is ratio. Deze variabele wordt niet – direct - meegenomen in de analyse, maar gebruikt om de onafhankelijke variabele te construeren. Aan de variabele zijn geen bewerkingen uitgevoerd. Daarentegen is de vraag alleen gesteld aan respondenten ouder dan 44 jaar.

3. Verwachte pensioenleeftijd

De verwachte pensioenleeftijd is gemeten door de vraag: “Op welke leeftijd verwacht u te stoppen met werken?” Respondenten konden hierop antwoorden met een leeftijd tussen de 0 en 100 jaar. Het meetniveau is ratio. Deze variabele wordt niet – direct - meegenomen in de analyse, maar gebruikt om de onafhankelijke variabele te construeren. Aan de variabele zijn geen bewerkingen uitgevoerd. Daarentegen is de vraag alleen gesteld aan respondenten ouder dan 44 jaar.

3.3.2. Verklarende variabelen

4. Inkomen (in dzd. Euro's)

Het maandelijks netto-inkomen is gemeten door de vraag “Persoonlijk netto maandinkomen in euro's?” Respondenten konden deze vraag beantwoorden door exact hun maandelijks netto-inkomen in te vullen. Het meetniveau is ratio. Het inkomen is een geïmputeerde variabele. Indien een respondent het maandelijks netto-inkomen niet heeft ingevuld, is het bijgeschat op basis van maandelijks bruto-inkomen. Om de invloed van het maandelijks netto-inkomen in de analyse eenvoudiger te kunnen interpreteren, is het gehercodeerd tot het inkomen in duizenden euro's. $\text{Inkomen (in dzd. Euro's)} = \text{netto-maandelijks inkomen in Euro's} / 1000$.

5. Opleidingsniveau

Het opleidingsniveau van de respondenten is gemeten door de vraag “Wat is uw hoogst afgeronde opleiding?” (In CBS-categorieën). Respondenten konden hierop antwoorden met: basisonderwijs (1), vmbo (2), havo/vwo (3), Mbo (4), Hbo (5) en Wo (6). Het meetniveau is nominaal. Deze variabele is gehercodeerd tot: 1: laagopgeleid (=basisonderwijs en vmbo), 2: middelbaar opgeleid (=Mbo) en 3 hoogopgeleid (= Hbo en Wo). Van ieder van deze drie opleidingsgroepen – laag, middelbaar en hoog- zijn dummy-variabelen gemaakt met antwoordcategorieën: 1= ja en 0 = nee. In de statistische analyse werd de groep laagopgeleide respondenten gebruikt als referentiegroep.

6. Fysieke belasting

De mate waarin respondenten hun werk als fysiek belastend ervaren is gemeten door de vraag “Is uw werk lichamelijk zwaar?” Respondenten konden hierop antwoorden met: vaak (1), soms (2) en nooit (3). Het meetniveau is ordinaal Om een variabele te krijgen waarbij een hoge score duidt op het ervaren van het werk als fysiek belastend zijn de scores vaak (1) en nooit (3) gespiegeld.

7. Trainingsmogelijkheden

De mate waarin respondenten de mogelijkheid werd geboden deel te nemen aan vaardigheidstrainingen is gemeten door de vraag “Ik kreeg gelegenheden om nieuwe vaardigheden aan te leren?” Respondenten konden hierop antwoorden met: helemaal niet mee eens (1), niet mee eens (2), mee eens (3) en helemaal mee eens (4). Het meetniveau is ordinaal. Aan de variabele zijn verder geen bewerkingen verricht.

3.3.3. Controlevariabelen

8. Leeftijd

De leeftijd van respondenten is gemeten door de vraag “Wat is uw leeftijd?” Respondenten konden hierop antwoorden met hun leeftijd in jaren. Het meetniveau is ratio. Er zijn geen bewerkingen uitgevoerd op de variabele.

9. Geslacht

Het geslacht van respondenten is gemeten door de vraag “Wat is uw geslacht?” Respondenten konden hierop antwoorden met man (1) of vrouw (2). Het meetniveau is nominaal. De variabele is gehercodeerd tot een dummy- variabele met antwoordcategorieën 1=ja en 0=nee.

10. Partnerstatus

Of respondenten een partner hebben is gemeten door de vraag “Heeft u een partner? (Gehuwd of ongehuwd)”. Respondenten konden hierop antwoorden met nee (0) of ja (1). Het meetniveau is nominaal. De variabele is gehercodeerd tot dummy- variabele Partner met antwoordcategorieën 1=ja en 0=nee

\

11. Ouderschap

Of respondenten kinderen hebben is gemeten door de vraag “Hebt u kinderen?”. Respondenten konden hierop antwoorden met nee (0) en ja (1). Het meetniveau is nominaal. De variabele is gehercodeerd tot dummy- variabele Ouderschap met antwoordcategorieën 1=ja en 0=nee.

12. Beroepssector

De sector waarin respondenten werken is gemeten door de vraag “In welke branche bent u werkzaam?” Respondenten moesten aangeven welke categorie het meest bij hun beroep in de buurt kwam. Zij konden hierop antwoorden met: landbouw, jacht, bosbouw en visserij (1) winning van delfstoffen (2), industrie (3), productie en distributie van en handel in elektriciteit, aardgas, stoom en water (4), bouwnijverheid (5), handel, inclusief reparatie van consumentenartikelen (6), horeca (7), vervoer, opslag en communicatie (8), financiële instellingen (9), zakelijke dienstverlening, inclusief onroerend goed en verhuur van roerende goederen (10), overheidsdiensten, openbaar bestuur en sociale verzekeringen (11) en onderwijs (12), gezondheids- en welzijnszorg (13), milieudienstverlening, cultuur, recreatie en overige dienstverlening. Het meetniveau is nominaal. De variabele is gehercodeerd tot 4 beroepssectoren: primaire sector (1), zakelijke dienstverlening (2), 3 publieke dienstverlening en zorg en welzijn (4). In bijlage 1 staat nader beschreven tot welke van deze beroepssectoren en branche behoort. Van ieder van deze 4 beroepssectoren is een dummy-variabele gemaakt met antwoordcategorieën 1=ja en 0=nee. In de statistische analyse werd de primaire sector gebruikt als referentiegroep.

3.4. Analysebeschrijving

De analyse heb ik verricht op basis van een lineaire regressieanalyse. In de analyse heb ik een hiërarchisch model geschat. Tijdens het schatten voegde ik telkens handmatig variabelen toe aan het model. Met behulp van VIF-scores heb ik de indicatoren in het model op multicollineariteit gecontroleerd. In het eerste model heb ik de controlevariabelen toegevoegd. Vervolgens heb ik in ieder volgend model een verklarende variabele toegevoegd. Voor ieder model heb ik de helling, standaardfout, p-waarde, aangepaste R^2 en (partiele) F-toets geïnterpreteerd. Deze resultaten van deze analyse zijn te vinden in Bijlage 2.

Na het schatten van de modellen heb ik het uiteindelijke model gecontroleerd op uitbijters. Op basis van deze controle heb ik 6 uitbijters gevonden. Nadien heb ik de analyse opnieuw uitgevoerd zonder deze invloedrijke punten. De modevaluatie en de hypothesetoetsing in de resultatenparagraaf heb ik dan ook beschreven op basis van de analyse – en de modellen - zonder deze uitbijters.

Om te controleren of er geldige uitspraken kunnen worden gedaan over het (uiteindelijke) model heb ik tot slot de assumpties voor een regressieanalyse getoetst. Deze assumpties heb ik getoetst voor zowel het model met de invloedrijke punten als het model zonder deze punten.

4. Resultaten

In deze sectie heb ik de resultaten van de thesis besproken. Allereerst werden de uni- en bivariate statistieken van de variabelen in het model beschreven. Vervolgens heb ik de (regressie-) modellen geëvalueerd om zo de kwaliteit van de modellen te kunnen beoordelen. Tot slot heb ik de hypothesen getoetst.

4.1. Beschrijvende statistieken

4.1.1. Univariante beschrijvende statistieken

Tabel 1: Gemiddelden, standaarddeviatie, minimum en maximum van discrepantie in pensioenleeftijd, gewenste pensioenleeftijd, verwachte pensioenleeftijd, inkomen (in dzd. Euro's) en leeftijd (N=1149)

	Gemiddelde	SD	Minimum	Maximum
1. Discrepantie pensioenleeftijd*	3.30	3.41	0	50
2. Gewenste pensioenleeftijd	63.20	4.20	20	99
3. Verwachte pensioenleeftijd	66.50	3.45	55	100
4. Inkomen (in dzd. Euro's)	2.28	1.15	0.00	10.00
5. Leeftijd	55.46	6.55	45	82

*De discrepantie in pensioenleeftijd is de verschilscore tussen verwachte gewenste pensioenleeftijd

In tabel 1 staan de beschrijvende statistieken van de continue variabelen met een ratio schaal. De eerste continue en tevens de afhankelijke variabele is de discrepantie in pensioenleeftijd. Respondenten wensen gemiddeld 3 jaar en 110 dagen eerder met pensioen te gaan dan dat zij verwachten. Een simpele rekensom bevestigt een correcte constructie van de gemiddelde discrepantie: $66.50 - 63.20 = 3.30$. Het minimum bedraagt 0 en het maximum 50 jaar. De variantie in de discrepantie is relatief hoog: 3 jaar en 150 dagen, mogelijk omdat de variabele is geconstrueerd als verschilscore.

Uit tabel 1 volgt dat respondenten gemiddeld met pensioen wensen te gaan met 63 jaar en 73 dagen bedraagt met een standaarddeviatie van 4 jaar en 73 dagen. Deze variabele wordt niet meegenomen in het model, maar wordt bondig toegelicht in inzicht te krijgen in de berekening van de discrepantie in pensioenleeftijd. Het minimum bedraagt 20 en het maximum 100.

De derde variabele in tabel 1 is de verwachte pensioenleeftijd. Respondenten verwachten gemiddeld gezien te pensioneren 66 jaar en 183 dagen. Deze variabele wordt eveneens niet meegenomen in het model, maar wordt bondig toegelicht in inzicht te krijgen in de berekening van de discrepantie in pensioenleeftijd. De variantie bedraagt 3 jaar en 223 dagen. Respondenten verwachten op zijn vroegst te pensioneren in de leeftijd van 55 en uiterlijk in de leeftijd van 100.

Uit tabel 1 blijkt dat respondenten in de complete dataset maandelijks gemiddeld 2280 euro netto verdienen met een standaarddeviatie van 1147 euro. Het minimum bedraagt 0 euro – of terwijl geen inkomen- en het maximum 10 duizend euro.

Tot slot volgt uit tabel 1 dat de respondenten in de complete dataset gemiddeld 55 jaar en 168 dagen oud zijn met een standaarddeviatie van 3 jaar en 223 dagen. Het minimum bedraagt 45 jaar en het maximum 102.

Tabel 2: Relatieve frequentieverdelingen fysieke belasting, trainingsmogelijkheden, geslacht, partnerstatus ouderschap en beroepssector (N=1149)

	Percentage	Cumulatief Percentage
Opleiding		
<i>Laag</i>	18.1	18.1
<i>Middelbaar</i>	36.9	55.0
<i>Hoog</i>	45.0	100.0
Fysieke Belasting		
<i>Nooit</i>	53.2	53.2
<i>Soms</i>	33.9	87.0
<i>Vaak</i>	13.0	100.0
Trainingsmogelijkheden		
<i>Helemaal niet mee eens</i>	3.7	3.7
<i>Niet mee eens</i>	18.1	21.8
<i>Mee eens</i>	67.0	88.8
<i>Helemaal mee eens</i>	13.0	100.0
Geslacht		
<i>Man</i>	53.2	53.2
<i>Vrouw</i>	46.8	100.0
Partnerstatus		
<i>Zonder partner</i>	23.6	23.6
<i>Met partner</i>	76.4	100.0
Ouderschap		
<i>Zonder kinderen</i>	22.4	22.4
<i>Met kinderen</i>	77.6	100.0
Beroepssector		
<i>Primaire sector</i>	28.5	28.5
<i>Zakelijke dienstverlening</i>	25.3	53.9
<i>Publieke dienstverlening</i>	24.5	78.3
<i>Zorg en Welzijn</i>	21.7	100.0

In tabel 2 staan de relatieve frequentieverdelingen van het opleidingsniveau, de ervaren fysieke belasting van het werk, de mogelijkheid om deel te nemen aan trainingen, geslacht, partnerstatus, ouderschap en beroepssector. In deze thesis zijn minder laagopgeleiden (18.1%) dan hoogopgeleiden (45%) en meer mensen die hun werk nooit als fysiek belastend ervaren (53.2%) dan als vaak fysiek belastend (13%). De proportie mannen (53.2%) is ongeveer gelijk aan het aantal vrouwen (46.8%). Er zijn meer respondenten met partner (76.4%) en kinderen (77.6%) dan zonder (respectievelijk 23.6% en 22.4%). Ongeveer evenveel respondenten zijn werkzaam in de verschillende beroepssectoren: in de primaire sector (28.5) iets meer respondenten dan in de zorgsector (21.7%). In de thesis zijn wel aanzienlijk meer respondenten die eens zijn (67.0%) met de mogelijkheid trainingsactiviteit te krijgen aangeboden. Dit is in overeenstemming met de theorie: deze mogelijkheden worden in toenemende mate aangeboden aan oudere werkenden.

4.1.2. Bivariate beschrijvende statistieken

Tabel 3: Associaties tussen alle variabelen in het onderzoek (n=1149)

	1	2	3	4	5	6	7	8	9	10	11	12
1. Discrepanctie pensioenleeftijd	-	-.647*	.369*	-.015	.043 ^a	.100 ^a	.041 ^a	-.019	-.276*	.073 ^a	-.038	0.01
2. Gewenste pensioenleeftijd		-	.469*	.093*	.073	.100 ^a	.046 ^a	-.040	.336*	.066 ^b	.011	-.021
3. Verwachte pensioenleeftijd			-	.096*	.038 ^a	.006 ^a	.009 ^a	-.065	.087*	.059 ^a	-.035	-.038
4. Inkomen in dzd. Euro's				-	.388 ^a	.274 ^a	.138 ^a	-.385 ^a	.029*	.142 ^a	-.014	.068
5. Opleiding					-	.272 ^c	.099 ^c	.074 ^c	.023 ^a	.204 ^c	.040 ^c	.043 ^c
6. Fysieke belasting						-	.112 ^c	.105 ^c	.025 ^a	.290 ^c	.034 ^c	.046 ^c
7. Trainingsmogelijkheden							-	.052 ^c	.001 ^a	.174 ^c	.041 ^c	.075 ^c
8. Geslacht								-	.031 ^a	.449 ^c	.054 ^d	.035 ^d
9. Leeftijd									-	.111 ^a	.006	.055
10. Beroepssector										-	.140 ^c	.106 ^c
11. Partnerstatus											-	.272 ^d
12. Ouderschap												-

^{4.} * = associatie is significant bij tweezijdige $p < 0.01$; a= op basis van F-toets voor verschil in gemiddelden; c= Cramers' V; d= Phi/ ϕ

In tabel 2 worden de associaties tussen alle variabelen in het onderzoeksmodel gegeven. De associaties geven een eerste indruk van de verbanden in het model. De gewenste en verwachte pensioenleeftijd zitten niet in dit model, maar worden gegeven om de interpretatie van de associaties van de discrepantie in pensioenleeftijd te vergemakkelijken. Deze onafhankelijke variabele is de verschilscore tussen de verwachte en gewenste pensioenleeftijd. In deze deelsectie heb ik de opvallende associaties en associaties tussen de onafhankelijke variabele en de verklarende indicatoren toegelicht. De associaties in het model worden nader toegelicht in bijlage 2.

De associaties tussen continue variabelen en continue en dichotome variabelen zijn berekend aan de hand van de Pearson's correlatie. Dit getal drukt de sterkte van de associatie tussen de twee variabelen uit. De associatie tussen continue en niet-dichotome categorische variabelen heb ik berekend aan de hand van de proportie verklaarde variantie ($r = \sqrt{R^2}$). Deze maat zegt niets over de richting van de associatie. Voor de associaties tussen twee categorische variabelen wordt de Cramer's V gebruikt. Dit getal is een relatieve maat voor de sterkte van de samenhang. Deze maat zegt eveneens niets over de richting van de associatie. De Phi-coëfficiënt wordt gebruikt als maat voor de associatie tussen dichotome variabelen. Deze is gelijk aan het Pearson's correlatiecoëfficiënt.

Uit de tabel 2 volgt een significante en sterk negatieve associatie tussen de discrepantie in pensioenleeftijd en de leeftijd waarop respondenten wensen te pensioneren ($r = -.647; p < 0.01$). Een negatieve correlatie duidt erop dat een hoge score op de ene variabele samenhangt met een lage score op de andere variabele. In dit geval is er dus een tegengesteld verband tussen een score op de discrepantie in pensioenleeftijd en verwachte pensioenleeftijd. Indien een respondent vroeg wenst te pensioneren, hangt dit samen met een hoge score op de discrepantie en wanneer een respondenten laat wenst pensioneren hangt dit samen met een lage discrepantie.

Daarnaast is er een significant positieve correlatie tussen de discrepantie en de verwachte pensioenleeftijd ($r=.369; p<0.01$). Een score op de verwachte pensioenleeftijd hangt dus in positieve mate samen met een score op de discrepantie in pensioenleeftijd. Dit verband is – ondanks de significantie - aanzienlijk minder sterk de correlatie tussen gewenste pensioenleeftijd en de discrepantie. Daarnaast is de associatie positief in plaats van negatief.

Aan de hand van de theorie voorspelde ik een negatieve associatie tussen de discrepantie in het maandelijks netto-inkomen van een respondent. In tabel 2 is er een negatieve samenhang tussen discrepantie en inkomen zichtbaar, maar deze is zwak en niet significant ($r= -.015$ en $p=.614$). Daarnaast is er een zwakke niet-significante associatie zichtbaar tussen de discrepantie in pensioenleeftijd en het opleidingsniveau van respondenten ($r=.052^a$ en $p=.142$) en de mogelijkheid om deel te nemen aan trainingsactiviteiten ($r=.041^a$ en $p=.166$). Bovendien zegt deze maat op basis van de proportie verklaarde variantie niets over de richting van het verband: dit getal is altijd positief. Respondenten met een laag opleidingsniveau een ervaren gemiddeld geen hoger of lagere discrepantie dan respondenten met een middelbaar of hoog opleidingsniveau. De gemiddelde discrepantie verschilt niet tussen de opleidingsniveaus.

Daarentegen is er een significant- en positieve associatie tussen de discrepantie in pensioenleeftijd en de ervaren fysieke belasting van het werk ($r = .100$ en $p<0.01$). Het vaak ervaren van het werk als fysiek belastend hangt samen met een hogere scores op de discrepantie in pensioenleeftijd. Er kan worden gesteld dat respondenten die hun werk als fysiek belastend ervaren gemiddeld meer verschil hebben tussen hun gewenste en verwachte pensioenleeftijd dan respondenten die hun werk niet zo ervaren.

Verder volgt uit tabel 2 een negatief en significant verband tussen de discrepantie en de leeftijd van een respondent ($r=-.276$ en $p<0.01$). Naarmate respondenten ouder worden lijken zij dus een kleiner verschil te percipiëren in hun gewenste en verwachte pensioenleeftijd. Aangezien respondenten niet met terugwerkende kracht kunnen pensioneren zal hun gewenste pensioenleeftijd waarschijnlijk meestijgen met hun leeftijd. Deze waarschijnlijkheid is terug te zien in het significante en positieve verband tussen gewenste pensioenleeftijd en leeftijd ($r=.332$ en $p<0.01$).

Tot slot volgen er uit tabel 2 geen sterke associaties die mogelijk op multicollineariteit zouden kunnen duiden. Een sterk verband tussen voorspellers is onwenselijk, omdat de voorspellers elkaar dan voorspellen en er geen extra variantie wordt verklaard in de discrepantie in pensioenleeftijd. Of de verbanden tussen de variabelen daadwerkelijk problematisch zijn zal ik beoordelen tijdens de modevaluatie aan de hand van VIF-scores.

4.2. Modevaluatie

4.2.1. Modelfit

Tabel 4: Resultaten van de hiërarchische regressieanalyse met discrepantie in pensioenleeftijd als onafhankelijke variabele, zonder outliers. (N=1143).

	Model 1		Model 2		Model 3		Model 4		Model 5	
	<i>b</i>	SE	<i>b</i>	SE	<i>b</i>	SE	<i>b</i>	SE	<i>b</i>	SE
<i>Intercept</i>	13.211**	.848	12.160**	.860	12.269**	.883	11.237**	.934	12.447**	1.025
<i>Leeftijd</i>	-.151**	.015	-.151**	.015	-.152**	.015	-.151**	.015	-.153**	.015
<i>Geslacht (1= Man)</i>	.016	.199	.144	.214	.111	.215	.081	.241	.088	.214
<i>Partnerstatus (1= met partner)</i>	-.550	.227	-.569	.227	-.553	.227	-.550	.226	-.537	.225
<i>Ouderschap (1= met kinderen)</i>	.237	.231	.274	.232	.278	.232	.274	.231	.298	.230
<i>Sector (ref = primair)</i>										
<i>Zakelijke dienstverlening</i>	-.052	.109	-.055	.109	-.058	.109	-.064	.109	-.076	.109
<i>Publieke dienstverlening</i>	-.148	.239	-.121	.239	-.024	.247	.019	.246	-.081	.247
<i>Zorg en Welzijn</i>	-.393	.266	-.409	.266	-.353	.268	-.479	.270	-.426	.269
<i>Inkomen in dzd. Euro's</i>			-.090	-.051	-.083	.099	-.032	.100	-.005	.100
<i>Opleiding (ref= laag)</i>										
<i>Midden^a</i>					-.170	.268	-.038	.100	-.066	.269
<i>Hoog</i>					-.425	.287	-.174	.296	-.185	.295
<i>Fysieke belasting</i>							.469**	.144	.431*	.144
<i>Trainingsmogelijkheden</i>									-.409*	.145
<i>R² Adjusted</i>		0.87		.088		.089		.096		.102
<i>F-Change</i>		16.522**		2.748		1.221		10.559*		7.959

*significant bij $p < 0.05$; ** significant bij < 0.01 ; referentiegroep sector: primaire sector; referentiegroep opleiding: laag opleidingsniveau

Tabel 4 laat zien in welke mate de verschillende modellen in staat zijn om de variantie in de discrepantie tussen de verwachte en gewenste pensioenleeftijd te verklaren. De statistische maat voor de proportie verklaarde variantie is R^2 . In tabel 3 wordt ook de $R^2_{adjusted}$ gegeven, dit is de proportie verklaarde variantie in de discrepantie, gecorrigeerd voor het aantal voorspellers. Ter evaluatie van de modellen gebruik ik deze maat. Op basis van de F-change kan ik beoordelen of een model een significant betere voorspeller is dan het vorige model.

In tabel 4 is te zien dat het eerste model 8.7 procent ($R^2 = .087$) van de variantie in het verschil tussen gewenste- en verwachte pensioenleeftijd verklaard gecorrigeerd voor het aantal voorspellers. Dit model bevat alleen de controlevariabelen: geslacht, leeftijd, beroepssector, partnerstatus en ouderschap. Het model is een significant betere voorspeller van de discrepantie in pensioenleeftijd dan een leeg model ($F_{change}(7,1135) = 16.522, p < 0.01$).

In het tweede model wordt het netto-maandelijks inkomen van de respondent toegevoegd en in het derde model de opleiding. Gecorrigeerd voor het aantal voorspellers, verklaren deze modellen iets meer variantie in het verschil tussen gewenste en verwachte pensioenleeftijd respectievelijk 8.8% ($R^2 a = .088$) en 8.9% ($R^2 a = .089$). Het model met het inkomen verklaart niet significant meer variantie dan het model met alleen de controlevariabelen ($Fchange (1,1134) = 2.748, p < 0.098$) en het model met opleiding verklaart niet significant meer variantie dan het model met inkomen en de controlevariabelen ($Fchange (1, 1132) = p = 0.295$).

Daarentegen is het vierde model met de ervaren fysieke belasting van het werk een significant betere voorspeller van de discrepantie in pensioenleeftijd ($Fchange (1, 1131) = p = 0.01$) dan het derde model. Gecorrigeerd voor het aantal voorspellers, verklaart dit model 9.6% ($R^2 a = .096$) van de variantie in de discrepantie. In het laatste model wordt de mogelijkheid om trainingsactiviteiten te volgen toegevoegd. Dit vijfde en laatste model verklaart 10.2% ($R^2 a = .096$) van de variantie in de discrepantie in pensioenleeftijd. Dit model verklaart niet significant meer variantie in de discrepantie in de pensioenleeftijd van respondenten ($Fchange (1, 1130) = p = 0.05$), ondanks dat de aangepaste proportie verklaarde variantie aanzienlijk toeneemt.

Samenvattend: het eerste model met alleen de controlevariabelen is een significant betere voorspeller van variantie in de discrepantie in pensioenleeftijd dan een leeg model. Dit resultaat ondersteunt de keuze voor deze controlevariabelen. Vervolgens neemt in elk geschat model de aangepaste proportie verklaarde variantie toe. Echter, van alle modellen is alleen het vierde model - waarin de ervaren fysieke belasting van het werk door de respondent wordt toegevoegd - een significant betere voorspeller dan het daarvoor geschatte model.

4.2.2. Multicollineariteit

Om te controleren of er sprake van multicollineariteit heb ik VIF-scores berekend. Indien hier sprake van is voorspellen de voorspellers elkaar en wordt er geen extra variantie verklaard in de discrepantie in pensioenleeftijd. Dit is onwenselijk. Een vuistregel is dat er vanaf een VIF van 4 een probleem is voor het schatten van de regressiecoëfficiënt van de betreffende voorspeller. Ik heb geen problematische VIF-scores gevonden. De VIF-scores van het uiteindelijke model zijn te vinden in Bijlage 3

4.2.3. Uitbijters

In de data zijn 6 uitbijters gevonden. Deze respondenten heb ik beschouwd als uitbijter, omdat zij te hoog hebben gescoord op verschillende maten voor uitbijters. De scores van de respondenten lagen op zowel de Cook's Distance als de DFFit en tenminste één DFBeta boven de kritieke waarde. Deze uitbijters hebben een grote negatieve invloed op de kwaliteit van het geschatte model, daarom heb ik deze respondenten uit de data verwijderd. In Bijlage 3 wordt een nadere toelichting geven op deze uitbijters.

4.2.4. Assumptietoetsing

Het doel van de statistische analyse is het kunnen doen van generaliserende uitspraken over de onderzoekspopulatie. Om deze uitspraken te doen moet de data die wordt gebruikt in de analyse voldoen aan een viertal assumpties: onafhankelijkheid, lineariteit, homoscedasticiteit en normaliteit. Mochten assumpties worden geschonden, heeft dit gevolgen voor de conclusies die kunnen worden getrokken op basis van de analyse. In de data van de analyse werd de assumptie van normaliteit geschonden. Aan de andere assumpties werd voldaan. In Bijlage 3 wordt de toetsing van assumpties nader toegelicht.

4.3. Hypothesetoetsing

In deze deelsectie bespreek ik de resultaten gecentreerd rond de vier hypothesen (H1 t/m 4) van deze thesis. Ik heb een hiërarchische regressieanalyse gebruikt om de hypothesen te toetsen. In tabel 4 zijn de uitkomsten van deze analyse te zien. Telkens is de significantie van een regressiecoëfficiënt zichtbaar alsmede de standaarddeviatie.

4.3.1. Inkomen

De hypothese aangaande het maandelijks netto-inkomen luidt als volgt: (H1) naarmate een oudere werkende een hoger inkomen heeft zal de discrepantie in pensioenleeftijd kleiner zijn. De resultaten ondersteunen deze hypothese niet. Het maandelijks netto-inkomen heeft een niet-significant negatief effect op de discrepantie in pensioenleeftijd ($b = -.005$, $p = .957$). Dit effect van netto-inkomen is robuust: het is negatief en niet-significant in alle modellen. Indien het inkomen met duizend euro stijgt, zal het verschil tussen de verwachte en gewenste pensioenleeftijd 18 dagen kleiner worden. Bij een gemiddeld inkomen van 2280 euro zorgt het netto maandelijks inkomen voor een discrepantie van 41 dagen, gecontroleerd voor de andere indicatoren in het model. Ondanks dat het effect van het inkomen met de hypothese mee gaat wordt H2 op basis van de resultaten verworpen.

4.3.2. Opleidingsniveau

De tweede hypothese luidt als volgt (H2): *naarmate het opleidingsniveau van een oudere werkende hoger ligt, zal de discrepantie in pensioenleeftijd kleiner zijn*. Om deze hypothese accuraat te kunnen toetsen is de referentiegroep voor respondenten met een laag opleidingsniveau. Zo kan het verschil tussen deze groep en de twee groepen met hogere opleidingsniveaus - het middelbare en hoge opleidingsniveau - worden geschat.

De resultaten ondersteunen de hypothese niet. Er is geen significant verschil tussen de opleidingsniveaus in de discrepantie in pensioenleeftijd. Dit niet-significante verschil is robuust over de verschillende modellen. Het verschil in discrepanties in pensioenleeftijd tussen opleidingsniveaus gaat wel met de richting van de hypothese mee. Model 5 in tabel 4 laat zien dat zowel middelbaaropgeleiden

($b=-.066$, $p=.805$) als hoogopgeleiden ($b=-.185$, $p=.531$) verschillen in hun discrepantie in pensioenleeftijd ten opzichte van laagopgeleiden. Voor middelbaaropgeleiden is het verschil tussen hun verwachte en gewenste pensioenleeftijd 24 dagen kleiner dan voor laagopgeleiden. Voor hoogopgeleiden is het verschil tussen hun verwachte en gewenste pensioenleeftijd 68 dagen kleiner dan voor de groep laagopgeleiden.

Ondanks dat de verschillen in discrepanties in pensioenleeftijd tussen de verschillende opleidingsniveaus met de richting van de hypothese meegaan, zijn deze verschillen niet significant. Op basis van de resultaten wordt H2 verworpen.

4.3.3. Fysieke Belasting

De derde hypothese luidt (H3): *naarmate het werk als fysiek meer belastend wordt ervaren, zal de discrepantie tussen verwachte en gewenste pensioenleeftijd groter zijn*. De resultaten ondersteunen deze hypothese. Model 5 in tabel 5 laat zien dat het ervaren van werk als fysiek belastend een significant positieve invloed ($b = .431$, $p<0.05$) heeft. Zodra oudere werkenden hun werk als fysiek meer belastend ervaren, neemt het verschil tussen hun verwachte en gewenste pensioenleeftijd toe. Dit effect is behoorlijk robuust: in alle modellen met fysieke belasting is het effect significant en positief. Indien een respondent het werk nooit als fysiek belastend ervaart is het verschil tussen de verwachte en gewenste pensioenleeftijd – gecontroleerd voor de andere indicatoren - 157 dagen, terwijl dit verschil voor respondenten die hun werk vaak als fysiek belastend ervaren 1 jaar en 106 dagen bedraagt. Zodoende is de discrepantie voor het ervaren van het werk als fysiek belastend 314 dagen groter dan wanneer het werk nooit als fysiek belastend wordt ervaren. Het effect van fysieke belasting gaat dus mee in de richting van de hypothese en is robuust significant. Zodoende zijn de resultaten in overeenstemming met de hypothese.

4.3.4. Trainingsmogelijkheden

De vierde hypothese luidt (H4): *Naarmate er vaker mogelijkheden bestaan om trainingsactiviteiten te volgen, zal de discrepantie tussen gewenste en verwachte pensioenleeftijd kleiner zijn*. De resultaten ondersteunen deze hypothese. Het aanbieden van trainingsmogelijkheden heeft een significant negatieve invloed op de discrepantie in pensioenleeftijd ($b = .409$, $p<0.05$). Naarmate een oudere werkende er in toenemende mate mee instemt de mogelijkheid te hebben gekregen om deel te nemen aan trainingsactiviteiten, neemt wordt verschil tussen de verwachte en gewenste pensioenleeftijd kleiner. Indien een respondent heeft ervaren geregeld nieuwe vaardigheden kunnen ontwikkelen is de discrepantie in pensioenleeftijd 1 jaar en 83 dagen kleiner dan die van respondenten die nooit hebben ervaren nieuwe vaardigheden te kunnen ontwikkelen. Het effect van fysieke belasting gaat dus mee in de richting van de hypothese en is significant. Zodoende zijn de resultaten in overeenstemming met de hypothese.

4.3.5. Controlevariabele

In de analyse werd het model voor een vijftal variabelen gecontroleerd. Hiervan heeft de leeftijd van een respondent een significant negatief effect op de discrepantie tussen de verwachte en gewenste pensioenleeftijd ($b = -.153, p < 0.01$). Het effect van leeftijd is robuust: het is significant in alle modellen. Indien een respondent één jaar ouder wordt, zal de discrepantie tussen de verwachte en gewenste pensioenleeftijd 56 dagen kleiner worden. Er zijn geen hypothesen opgesteld, maar het effect van leeftijd lag in de lijn der verwachting. Respondenten kunnen niet met terugwerkende kracht pensioneren, zodoende zal hun gewenste pensioenleeftijd waarschijnlijk meestijgen met hun leeftijd. De effecten van de andere controlevariabelen geslacht ($b = 0.88, p = 0.680$), partnerstatus ($b = -.537, p = .017$), ouderschap ($b = .298, p = .198$) en de beroepssectoren: zakelijke dienstverlening ($b = -.076, p = 0.485$), publieke dienstverlening ($b = .081, p = .741$) en zorg en welzijn ($b = -.426, p = .114$) zijn niet significant. De effecten van deze controlevariabelen zijn niet-significant in alle modellen. De mate waarin respondenten verwachten dat hun pensioenwensen uitkomen verschilt niet tussen mannen en vrouwen, mensen met of zonder partner, met of zonder kinderen of de sector waarin een respondent werkzaam is.

5. Conclusie en Discussie

Door de snelle vergrijzing van de samenleving staat de Nederlandse arbeidsmarkt voor grootschalige uitdagingen. Het Kabinet kwam in actie en sloot samen met haar sociale partners een nieuw Pensioenakkoord (2019). De centrale boodschap in het Akkoord is evident: Nederlanders moeten gradueel langer doorwerken. Dit leidt bij veel werkenden, met name zij die binnen afzienbare tijd gaan pensioneren, tot ontevredenheid en frustraties. Het gradueel langer door moeten werken heeft gevolgen voor de wijze waarop oudere werkenden het proces van pensioneren voor zich zien. Als gevolg van de beleidsmatige hervormingen – met als summum het loslaten van de vaste AOW-leeftijd van 65 jaar - is de leeftijd waarop werkenden met pensioen verwachten te gaan gestegen, terwijl de wens om later met pensioen te gaan slechts beperkt meestijgt (de Beer & van der Gaag, 2019). Er lijken dus fermere discrepanties – verschillen tussen verwachte en gewenste pensioenleeftijd- te bestaan, die tot uiting komen in ontevredenheid en frustraties over het gevoerde pensioenbeleid. Zodoende is in deze thesis gezocht naar een antwoord op de vraag: *Hoe groot is het verschil tussen gewenste en verwachte pensioenleeftijd van oudere werkenden in Nederland, en in hoeverre wordt dit verklaart door sociaaleconomische factoren en baankenmerken?*

Uit de literatuur volgt veelvuldig dat deze discrepanties samenhangen met de sociaaleconomische factoren: inkomen en opleidingsniveau. Mensen met een hoger inkomen kunnen in bevredigendere mate hun wensen inwilligen en aan hun individuele verwachting voldoen dan werkenden met een lager inkomen, omdat bij die groep de financiële armslag ontbreekt. H1 stelde dat naarmate een oudere werkende een hoger inkomen heeft, de discrepantie tussen verwachte en gewenste

pensioenleeftijd kleiner zal zijn. Deze thesis heeft laten zien dat naarmate het inkomen van oudere werkenden hoger ligt, het verschil tussen de verwachte pensioenleeftijd en gewenste pensioenleeftijd niet kleiner is. Op basis van het inkomen zijn er geen verschillen tussen oudere werkenden in de mate waarin zij verwachten te kunnen pensioneren op de door hun gewenste leeftijd.

De genoten opleiding is bepalend voor de leeftijd waarop een werkende verwacht met pensioen te gaan. Laagopgeleiden treden door een kortere onderwijsduur al snel vijf jaar eerder toe tot de arbeidsmarkt. Uit eerder onderzoek blijkt dat dit resulteert in een verwachting eerder met pensioen te kunnen gaan. Daarentegen wensen hoogopgeleiden door een attractievere baankenmerken en gunstigere werkomstandigheden, zoals een hogere mate van autonomie en fysiek minder belastend werk, later met pensioen te gaan. H2 stelde dat naarmate de oudere werkende een hoger opleidingsniveau heeft, het verschil tussen de verwachte en gewenste pensioenleeftijd kleiner zal zijn. Deze thesis heeft laten zien dat er geen verschillen zijn tussen de opleidingsniveaus en de mate waarin oudere werkenden verwachten te kunnen pensioneren op de door hun gewenste leeftijd. Hoogopgeleiden hebben geen kleiner verschil tussen verwachte en gewenste pensioenleeftijd dan laagopgeleiden.

Naast sociaaleconomische factoren volgt uit de literatuur dat baankenmerken een rol spelen in de discrepantie tussen verwachte en gewenste pensioenleeftijd. Hiervan is de fysieke belasting zonder twijfel het meest besproken. Als gevolg van fysiek belastend werk hebben oudere werkende te kampen met een problematische uitvoerbaarheid van het werk. Met andere woorden: door fysieke slijtage zijn zij simpelweg niet in staat om langer door te werken. H3 stelde zodoende dat naarmate een oudere werkende het werk als meer fysiek belastend ervaart, het verschil tussen de verwachte en gewenste pensioenleeftijd groter zal zijn. Deze thesis heeft laten zien dat de mate waarin oudere werkenden verwachten dat hun pensioenwensen uitkomen verschillen tussen mensen met fysiek belastend werk en mensen zonder. De discrepantie tussen verwachte en gewenste pensioenleeftijd is ruim 10 maanden groter voor oudere werkenden met fysiek belastend werk.

Verder worden er in toenemende mate trainingsmogelijkheden geboden aan oudere werkenden. Hierdoor kunnen zij hun vaardigheden ontwikkelen en hun kennis bijtijds houden. Vanuit economische overwegingen was – en incidenteel is – het niet rationeel om trainingsmogelijkheden aan te bieden. Door deze economische barrières ervaren oudere werkenden het aanbieden van trainingsmogelijkheden als een sterk signaal van organisationele erkenning. Eerder onderzoek toont dat dit erkenningsgevoel zorgt voor een wens om langer door te werken. H4 stelde dat naarmate er vaker mogelijkheden bestaan om trainingsactiviteiten te volgen, de discrepantie tussen gewenste en verwachte pensioenleeftijd kleiner zal zijn. Deze thesis heeft laten zien dat de mate waarin oudere werkenden verwachten dat hun pensioenwensen uitkomen verschillen tussen oudere werkenden die geregeld de mogelijkheid ervaren deel te nemen aan trainingsactiviteiten en zij die dat nooit ervaren. De discrepantie tussen verwachte en gewenste pensioenleeftijd is ruim 1 jaar en 2 maanden groter voor oudere werkenden die nooit de mogelijkheid hebben deel te nemen aan trainingsactiviteiten.

Alles overziend bestaan er discrepanties tussen gewenste en verwachte pensioenleeftijd van oudere werkenden in Nederland. In deze thesis bedroeg de gemiddelde discrepantie 3 jaar en 110 dagen. Het verschil tussen de gewenste en verwachte pensioenleeftijd kon niet worden verklaard door het inkomen of opleidingsniveau van een oudere werkende. Daarentegen kon de mate waarin oudere werkenden verwachten dat hun pensioenwensen ook uitkomen wel worden verklaard door verschillen in de ervaren fysieke belasting van het werk en de mate waarin het mogelijk is om deel te nemen aan trainingsmogelijkheden. Het is opmerkelijk dat - ondanks de brede ondersteuning in de literatuur - de discrepantie tussen verwachte en gewenste pensioenleeftijd niet verschilt voor inkomen en opleidingsniveau. In het onderzoek heb ik nadrukkelijk gecontroleerd of dit werd veroorzaakt door de aanwezigheid van de baankenmerken; hier is geen sprake van.

Niettemin heeft deze thesis ook beperkingen. Ten eerste heeft een aanzienlijk deel van respondenten geen antwoord gegeven op de vragen over de verwachte en gewenste pensioenleeftijd. Deze groep non-respondenten zijn jonger en hebben een lager inkomen dan de respondenten die de vraag wel hebben beantwoord. Hierdoor zijn oudere werkenden die relatief gezien verder van hun pensioen af zitten en een lager inkomen hebben minder vertegenwoordigd in deze thesis. Dit heeft de representativiteit van het onderzoek als geheel geschaad. Bovendien bevestigt het hoge aantal non-respondenten op deze vragen de complexiteit van het pensioneringproces. Zelfs mensen die binnen afzienbare tijd pensioneren vinden het lastig hun wensen en verwachtingen kenbaar te maken. Ten tweede is de assumptie van normaliteit geschonden. Dit maakt het problematischer om op basis van de statistische analyse generaliserende uitspraken te doen. Tot slot zijn er twee zuiver methodologische beperkingen. Het netto-maandelijkse inkomen van respondenten is een geïmputeerde variabele: missende waarden zijn bijgeschat. Het inkomen werd dan voorspeld op basis van het bruto-inkomen. Met voorspellingen komen voorspelfouten. Dit schaadt de betrouwbaarheid van het inkomen als voorspeller. Daarnaast bevat het meetinstrument van de mogelijkheid deel te nemen aan trainingsactiviteiten een dubbele ontkenning. Dit schaadt de betrouwbaarheid en validiteit van de antwoorden van respondenten. Door deze statistische en methodologische beperkingen is voorzichtigheid gebaat bij het doen van aanbevelingen.

Onderzoek naar de verwachtingen en wensen aangaande het proces van pensioneren is meer dan ooit aan de orde van de dag. Door de koppeling van de pensioengerechtigde leeftijd aan ontwikkeling van de levensverwachting, blijft deze leeftijd onzeker voor toekomstige generaties. De hoogte van de pensioengerechtigde leeftijd blijft hierdoor een (politiek) gevoelige discussie. Factoren die samenhangen met de mate waarin de verwachtingen van oudere werkenden overeenkomen met hun pensioenwensen kunnen deze discussie voeden. Toekomstig onderzoek naar deze wensen en verwachtingen kan hier meer inzichten in geven. Hierbij is het minstens zo interessant om de zgn. uitbijters of mensen een negatieve discrepantie in dit onderzoek onder de loep te leggen. Zij geven aan grote verschillen te hebben tussen hun verwachtingen en wensen of de wens te hebben later te pensioneren dan dat zij verwachten. Echter, dient er hoe dan ook een lastig probleem onderuit te worden

gehaald. Het proces van pensioneren wordt steeds individueler, complexer en dynamischer. Zelfs tot kort voor het pensioneren vinden mensen het lastig om hun verwachtingen en wensen kenbaar te maken. Ondanks de nadrukkelijke aanwezigheid in de wetenschappelijke literatuur zijn concepten verwachte en pensioenleeftijd – mijns inziens – zeer gevoelig voor zowel random als systematische fouten. Toekomstig onderzoek moet hier zonder twijfel rekening mee houden.

Tot slot concludeer ik dat er discrepanties zijn tussen de verwachte en gewenste pensioenleeftijd zijn van oudere werkenden in Nederland. Deze discrepanties zijn verschillend voor de kenmerken van het werk dat wordt verricht. Dit plaatst vraagtekens bij wenselijkheid van een *one size fits all* bepaling van de AOW-leeftijd voor toekomstige generaties die –naar alle verwachting- aanzienlijk langer door moeten werken. Het differentiëren van de pensioenleeftijd naar deze baankenmerken is een politiek veelbesproken kwestie, nochtans schort het aan een concrete uitwerking. Met name het -nog steeds-ontbreken van regelingen voor mensen met fysiek werk is heikel punt. De angel is nog lang niet uit de pensioenenkwestie.

\

6. Literatuurlijst

- Armstrong-Stassen, M., & Ursel, N. D. (2009). Perceived organizational support, career satisfaction, and the retention of older workers. *Journal of occupational and organizational psychology*, 82(1), 201-220.
- Blekesaune, M., & Skirbekk, V. (2012). Can personality predict retirement behaviour? A longitudinal analysis combining survey and register data from Norway. *European Journal of Ageing*, 9, 199–206. doi:10.1007/s10433-011-0212-6
- Boxall, P., & Macky, K. (2009). Research and theory on high-performance work systems: progressing the high-involvement stream. *Human resource management journal*, 19(1), 3-23.
- Damman, M., Henkens, K., & Kalmijn, M. (2013). Late-career work disengagement: The role of proximity to retirement and career experiences. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, 68, 455–463
- De Beer, J., & Van der Gaag, N. (2019). Stijging van de AOW-leeftijd: Nieuw akkoord, einde discussie? *Tijdschrift voor Arbeidsvraagstukken*, 35(4), 370-381.
- De Beer, J., Van Dalen, H., & Henkens, K. (2017). Wanneer voelt de verhoging van de AOW-leeftijd als diefstal? *Me Judice*, (9), 1-4.
- Dingemans, E., Henkens, K., & Van Solinge, H. (2017). Working retirees in Europe: Individual and societal determinants. *Work, Employment and Society*, 31, 972-991.
- Fisher, G., Chaffee, D.S., & Sonnega, A. (2016). Retirement timing: A review and recommendations for future research. *Work, Aging and Retirement*, 2(2), 230-261.
- Froidevaux, A., & Hirschi, A. (2015). Managing the transition to retirement: From meaningful work to meaning in life at retirement. In *Handbook of research on sustainable careers*. Edward Elgar Publishing.
- Furunes, T., Mykletun, R. J., Solem, P. E., de Lange, A. H., Syse, A., Schaufeli, W. B., & Ilmarinen, J. (2015). Late career decisionmaking: A qualitative panel study. *Work, Aging and Retirement*, 1, 284–295. doi:10.1093/workar/wav011

- Grosch, J. W., & Pransky, G. (2009) Safety and health issues for an aging workforce. In S. Czaja & J. Sharit (Eds.), *Aging and work: Issues and implications in a changing landscape* (pp. 334–358). Baltimore, MD: The Johns Hopkins University Press.
- Hayward, M. (1986). The influence of occupational characteristics on men's early retirement. *Social Forces*, 64(4), 1032–45.
- Henkens, K., Van Solinge, H., Damman, M., & Dingemans, E. (2016). Langer doorwerken valt nog niet mee. *Demos Bulletin over Bevolking en Samenleving*, 32(2), 1-4.
- Hess, M. (2016). Expected and preferred retirement age in Germany. *Zeitschrift Fur Gerontologie Und Geriatrie*, 51(1), 98–104.
- Hess, M. (2018). Retirement Expectations in Germany-Towards Rising Social Inequality? *Societies*(8), 50-54. doi:10.3390/soc8030050
- Hess, M., Naegele, L., Becker, L., Mäcken, J., & De Tavernier, W. (2021). Planned Retirement Timing in Europe: Are Europeans Adapting to the Policy of Extending Working Lives. *Frontiers in sociology*, 6, 691066. <https://doi.org/10.3389/fsoc.2021.691066>
- Luginbuhl, R., Smid, B. (2021, 20 mei). De verscheidenheid van vermogens van Nederlandse huishoudens: update. Centraal Bureau voor de Statistiek. Geraadpleegd op 15 april 2022, van <https://www.cpb.nl/sites/default/files/omnidownload/CPB-Achtergronddocument-verscheidenheid-vermogens-Nederlandse-huishoudens-update.pdf>.
- Montizaan, R. (2017) Statistiek: Lageropgeleiden Werken Langer Door Dan Hogeropgeleiden. *ESB Statistiek*, 4750(1), 294-295.
- Örestig, J., Strandh, M., & Stattin, M. (2013). A Wish Come True? A Longitudinal Analysis of the Relationship between Retirement Preferences and the Timing of Retirement. *Population and Ageing*, 6 (2), 99–118. doi:10.1007/s12062-012-9075-7
- Oude Mulders, J., Henkens, K., Van Dalen, P. (2021). Employee's Emotional, Cognitive and Behavioral Responses to Increasing Statutory Retirement Ages. *BIOMed Research International*, 4 (2), 2-14.
- Oude Mulder, J. (2019). Hoe gaan werkgevers om met langer doorwerken? *Tijdschrift voor Pensioenvraagstukken*, 33 (3), 5-9.

- Potočnik, K., Tordera, N., & Peiró, J. M. (2010). The Influence of the Early Retirement Process on Satisfaction with Early Retirement and Psychological Well-Being. *The International Journal of Aging and Human Development*, 70(3), 251–273. <https://doi.org/10.2190/AG.70.3.e>
- Rhoades, L. & Eisenberger, R. (2002). Perceived organizational support: A review of the literature. *Journal of Applied Psychology*, 87(2), 698-714.
- Schippers, J. (2019). Langer doorwerken in Nederland: Een succesverhaal met rafelranden. *Tijdschrift voor Arbeidsvraagstukken*, 35(4), 354-366.
- Sell, L., Bultmann, U., Rugulies, R., Villadsen, E., Faber, A., & Sogaard, K. (2009). Predicting long-term sickness absence and early retirement pension from self-reported work ability. *International Archives of Occupational and Environmental Health*, 82, 1133–1138. doi:10.1007/s00420-009-0417-6
- Shultz, K. S., & Olson, D. A. (2013). The changing nature of work and retirement. In M. Wang (Ed.), *The Oxford handbook of retirement* (pp. 543–558). New York, NY: Oxford University Press
- Steiber, N. & Kohli, M. (2017). You can't always get what you want. Actual and preferred ages of retirement in Europe. *Ageing and Society*, 37(2), 352-385.
- Sociale Verzekeringsbank. (2022). Overzicht wetten en regelingen, geraadpleegd op 14 april 2022, van <https://www.svb.nl/nl/over-de-svb/wie-zijn-we/wetten-en-regelingen>
- Topa, G., Moriano, J.A., Depolo, M., Alcovover, C. & Morales, J. (2009). Antecedents and consequences of retirement planning and decision-making: A meta-analysis and model. *Journal of Vocational Behaviour*, 75(1), 38-55.
- Tsui, A. S., Pearce, J. L., Porter, L. W., & Tripoli, A. M. (1997). Alternative approaches to the employee-organization relationship: does investment in employees pay off?. *Academy of Management journal*, 40(5), 1089-1121.
- Van der Horst, A. C., Klehe, U. C., & Van der Heijden, B. I. (2017). Adapting to a looming career transition: How age and core individual differences interact. *Journal of Vocational Behavior*, 99, 132-145.

Van Solinge, H. & Henkens, K. (2010). Living longer, working longer? The impact of subjective life expectancy on retirement intentions and behaviour. *European Journal of Public Health*, 20 (1), 47–51.

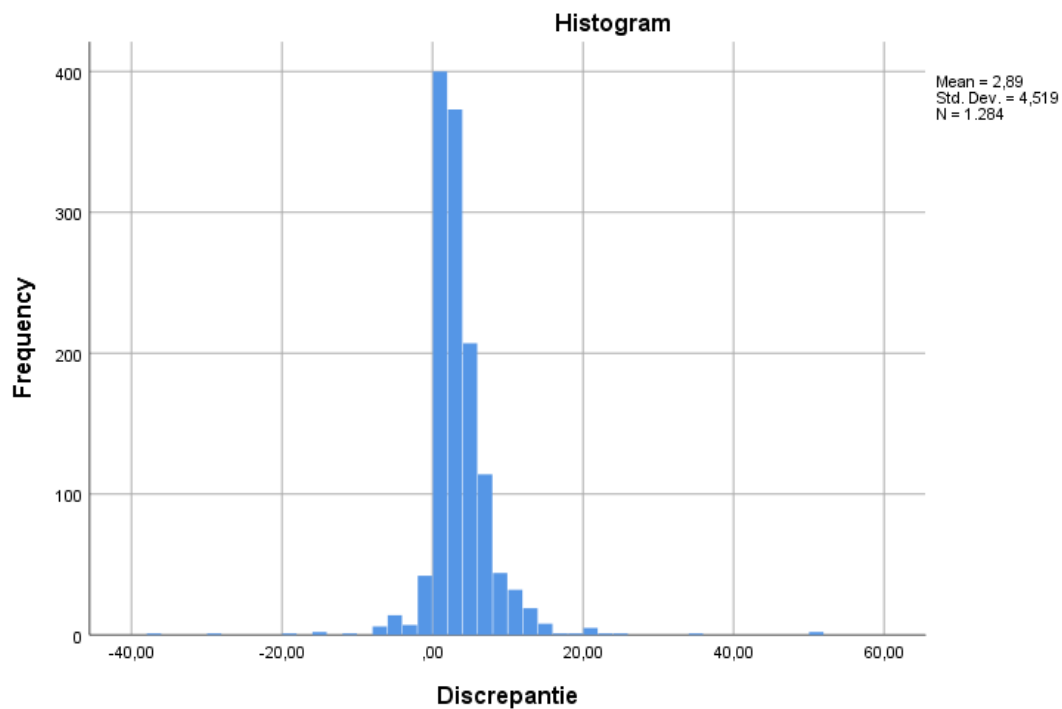
van Zon, S. (2016). Functional health decline before and after retirement: A longitudinal analysis of the Health and Retirement Study. *Social Science & Medicine*, 170, 26 – 34. doi: 10.1016/j.socscimed.2016.10.002.

Bijlage 1

In deze eerste bijlage wordt een volledige en overzichtelijke rechtvaardiging gegeven voor de totstandkoming van variabelen in het onderzoeksmodel. Per variabele wordt een frequentieverdeling van de oorspronkelijke variabele gegeven, de (statistische) bewerking op deze variabele en de frequentieverdeling van de uiteindelijke variabele – zoals deze- die in de analyse wordt opgenomen. De wijze waarop deze subset voor de analyse is geconstrueerd staat beschreven in de methodeparagraaf

1. Discrepantie in pensioenleeftijd

1.1 Histogram van de oorspronkelijke variabele discrepantie in pensioenleeftijd



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

1.2 Bewerkingen

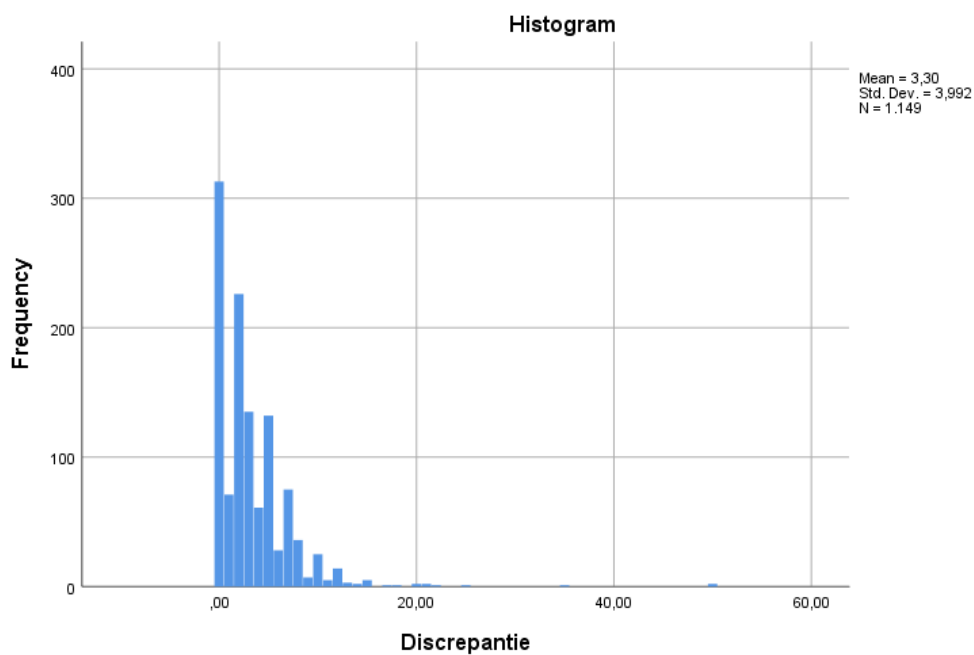
De variabele discrepantie is geconstrueerd op basis van een verschilscore tussen verwachte en gewenste pensioenleeftijd: $discrepantie = verwachte\ pensioenleeftijd - gewenste\ pensioenleeftijd$.

```
COMPUTE Discrepantie=cw20m289 - cw20m288.  
EXECUTE.
```

In de complete dataset zijn de negatieve scores uit de dataset gefilterd. In de methodenparagraaf heb ik toegelicht waarom ik dit heb gedaan.

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

1.3 Frequentieverdeling van de uiteindelijke variabele in de analyse, inclusief histogram.



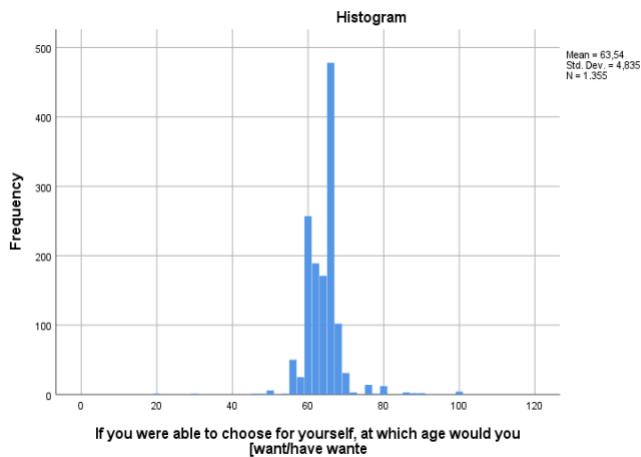
		Discrepantie			Cumulative
		Frequency	Percent	Valid Percent	Percent
Valid	,00	313	27,2	27,2	27,2
	1,00	71	6,2	6,2	33,4
	2,00	226	19,7	19,7	53,1
	3,00	135	11,7	11,7	64,8
	4,00	61	5,3	5,3	70,1
	5,00	132	11,5	11,5	81,6
	6,00	28	2,4	2,4	84,1
	7,00	75	6,5	6,5	90,6
	8,00	36	3,1	3,1	93,7
	9,00	7	,6	,6	94,3
	10,00	25	2,2	2,2	96,5

11,00	5	,4	,4	97,0
12,00	14	1,2	1,2	98,2
13,00	3	,3	,3	98,4
14,00	2	,2	,2	98,6
15,00	5	,4	,4	99,0
17,00	1	,1	,1	99,1
18,00	1	,1	,1	99,2
20,00	2	,2	,2	99,4
21,00	2	,2	,2	99,6
22,00	1	,1	,1	99,7
25,00	1	,1	,1	99,7
35,00	1	,1	,1	99,8
50,00	2	,2	,2	100,0
Total	1149	100,0	100,0	

```
FREQUENCIES VARIABLES=Discrepantie
/HISTOGRAM
/ORDER=ANALYSIS
```

2. Gewenste pensioenleeftijd

2.1 Histogram van de oorspronkelijke verdeling van gewenste pensioenleeftijd



FREQUENCIES

```
VARIABLES=cw20m288
/NTILES=4
/STATISTICS=STDDEV MINIMUM MAXIMUM MEAN MEDIAN
/HISTOGRAM
/ORDER=ANALYSIS.
```

2.2 Bewerkingen

In het oorspronkelijke databestand van de werk- en scholing module waren codering -9 en -8 toegekend aan respondenten die hadden geantwoord op een vraag met: -9 = ik weet het niet, -8 = ik wil het niet zeggen. Deze waarden zijn gehercodeerd als system-missing.

Ter constructie van de initiële dataset - een dataset geconstrueerd met alleen respondenten die ten minste één valide score hebben- had ik een dummy-variabelen gemaakt van de variabele gewenste pensioenleeftijd met antwoordcategorieën: 0 = een valide score, 1= system missing, -9 through -8. Er is voor iedere variabele in het model op dezelfde wijze een indicator-dummy gemaakt. Respondenten met een somscore van deze indicatoren gelijk aan het aantal variabelen in het model zijn uit het databestand verwijderd. De bovenstaande bewerkingen op deze wijze uitgevoerd in SPSS:

```
COMPUTE Indicator_gewenst=cw20m288.
EXECUTE.
RECODE Indicator_gewenst (-8=1) (-9=1) (SYSMIS=1) (1 thru Highest=0).
EXECUTE.

COMPUTE SOMindicatoren=Indicator_gewenst + Indicator_verwacht +
Indicator_inkomen +
Indicator_opleidingsniveau + Indicator_Fbelasting +
Indicator_vaardigheden +
Indicator_beroepssector + Indicator_kinderen.
EXECUTE.

FILTER OFF.
USE ALL.
SELECT IF (SOMindicatoren < 8).
EXECUTE.
```

Ter constructie van de complete dataset – waaruit de frequentieverdeling van de variabele zoals gebruikt in de analyse volgt- had ik (opnieuw) een indicator_dummy gemaakt van de variabele gewenste pensioenleeftijd: m1 met 0= valide score, 1= system missing. Er is voor iedere variabele in het model op dezelfde wijze een indicator-dummy gemaakt. Respondenten met een somscore van meer dan 0 op deze indicatoren – respondenten met dus ten minste één missende score - zijn uit het databestand verwijderd.

```
COMPUTE m1=0.
IF MISSING(cw20m288) m1=1.
EXECUTE.

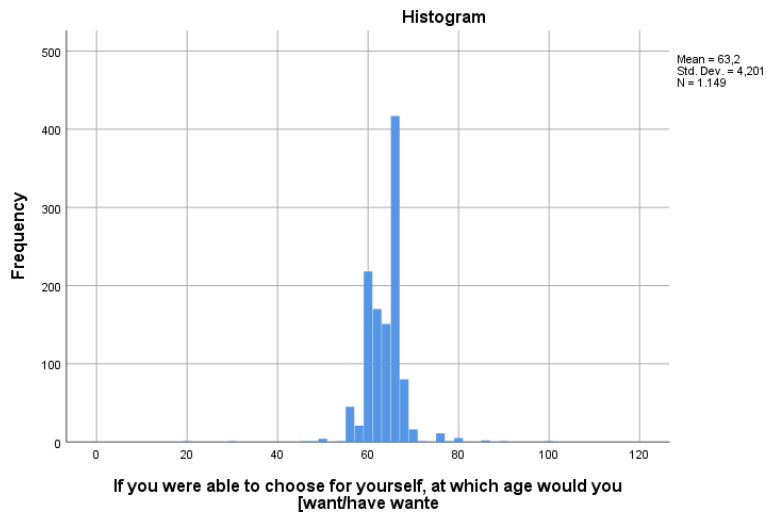
COMPUTE miss = m1+m2+m3+m4+m5+m6+m7+m8.
EXECUTE.

RECODE missendewaarden (0=0) (1 thru 7=1).
EXECUTE.

USE ALL.
COMPUTE filter_$=(miss = 0).
```

```
VARIABLE LABELS filter_$ 'miss = 0 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.
```

2.3 Frequentieverdeling van de uiteindelijke variabele in de analyse, inclusief histogram.



cw20m288 If you were able to choose for yourself, at which age would you [want/have wante

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	20	1	,1	,1	,1
	30	1	,1	,1	,2
	45	1	,1	,1	,3
	48	1	,1	,1	,3
	50	4	,3	,3	,7
	53	1	,1	,1	,8
	55	42	3,7	3,7	4,4
	56	3	,3	,3	4,7
	57	5	,4	,4	5,1
	58	16	1,4	1,4	6,5
	59	8	,7	,7	7,2
	60	210	18,3	18,3	25,5
	61	6	,5	,5	26,0
	62	164	14,3	14,3	40,3
	63	105	9,1	9,1	49,4
	64	46	4,0	4,0	53,4
	65	388	33,8	33,8	87,2
	66	29	2,5	2,5	89,7
	67	58	5,0	5,0	94,8
68	22	1,9	1,9	96,7	
70	16	1,4	1,4	98,1	

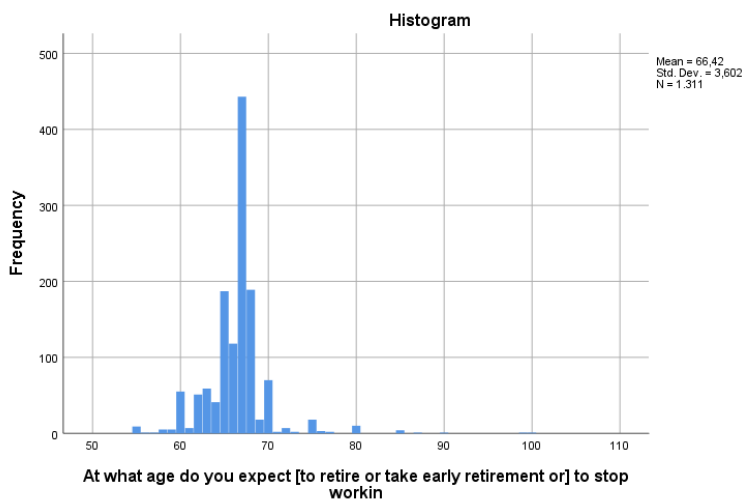
cw20m289 At what age do you expect [to retire or take early retirement or] to stop working

	Frequency	Percent	Valid Percent	Cumulative Percent
72	1	,1	,1	98,2
75	11	1,0	1,0	99,1
77	1	,1	,1	99,2
80	5	,4	,4	99,7
85	2	,2	,2	99,8
90	1	,1	,1	99,9
99	1	,1	,1	100,0
Total	1149	100,0	100,0	

```
FREQUENCIES VARIABLES=cw20m288
  /NTILES=4
  /STATISTICS=STDDEV MINIMUM MAXIMUM MEAN MEDIAN
  /HISTOGRAM
  /ORDER=ANALYSIS.
```

3. Verwachte pensioenleeftijd

3.1 Histogram van de oorspronkelijke verdeling van verwachte pensioenleeftijd



FREQUENCIES

```
VARIABLES=cw20m289
  /HISTOGRAM
  /ORDER=ANALYSIS.
```

3.2 Bewerkingen

In het oorspronkelijke databestand van de werk- en scholing module waren codering -9 en -8 toegekend aan respondenten die hadden geantwoord op een vraag met: -9 = ik weet het niet, -8 = ik wil het niet zeggen. Deze waarden zijn gehercodeerd als system-missing.

Ter constructie van de initiële dataset - een dataset geconstrueerd met alleen respondenten die ten minste één valide score hebben- had ik een dummy-variabelen gemaakt van de variabele verwachte pensioenleeftijd met antwoordcategorieën: 0 = een valide score, 1= system missing, -9 through -8. Er is voor iedere variabele in het model op dezelfde wijze een indicator-dummy gemaakt. Respondenten met een somscore van deze indicatoren gelijk aan het aantal variabelen in het model zijn uit het databestand verwijderd. De bovenstaande bewerkingen op deze wijze uitgevoerd in SPSS:

```

COMPUTE Indicator_verwacht=cw20m289.
EXECUTE.
RECODE Indicator_gewenst (-8=1) (-9=1) (SYSMIS=1) (1 thru Highest=0).
EXECUTE.

COMPUTE      SOMindicatoren=Indicator_gewenst      +      Indicator_verwacht      +
Indicator_inkomen +
      indicator_opleidingsniveau      +      Indicator_Fbelasting      +
Indicator_vaardigheden +
      Indicator_beroepssector + Indicator_kinderen.
EXECUTE.

FILTER OFF.
USE ALL.
SELECT IF (SOMindicatoren < 8).
EXECUTE.

```

Ter constructie van de complete dataset – waaruit de frequentieverdeling van de variabele zoals gebruikt in de analyse volgt- had ik (opnieuw) een indicator_dummy gemaakt van de variabele verwachte pensioenleeftijd: m1 met 0= valide score, 1= system missing. Er is voor iedere variabele in het model op dezelfde wijze een indicator-dummy gemaakt. Respondenten met een somscore van meer dan 0 op deze indicatoren – respondenten met dus ten minste één missende score - zijn uit het databestand verwijderd.

```

COMPUTE m2=0.
IF MISSING(cw20m289) m2=1.
EXECUTE.

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

RECODE missendewaarden (0=0) (1 thru 7=1).
EXECUTE.

USE ALL.
COMPUTE filter_$=(miss = 0).
VARIABLE LABELS filter_$ 'miss = 0 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

```

3.3 Frequentieverdeling van de uiteindelijke variabele in de analyse, inclusief histogram

cw20m289 At what age do you expect [to retire or take early retirement or] to stop workin

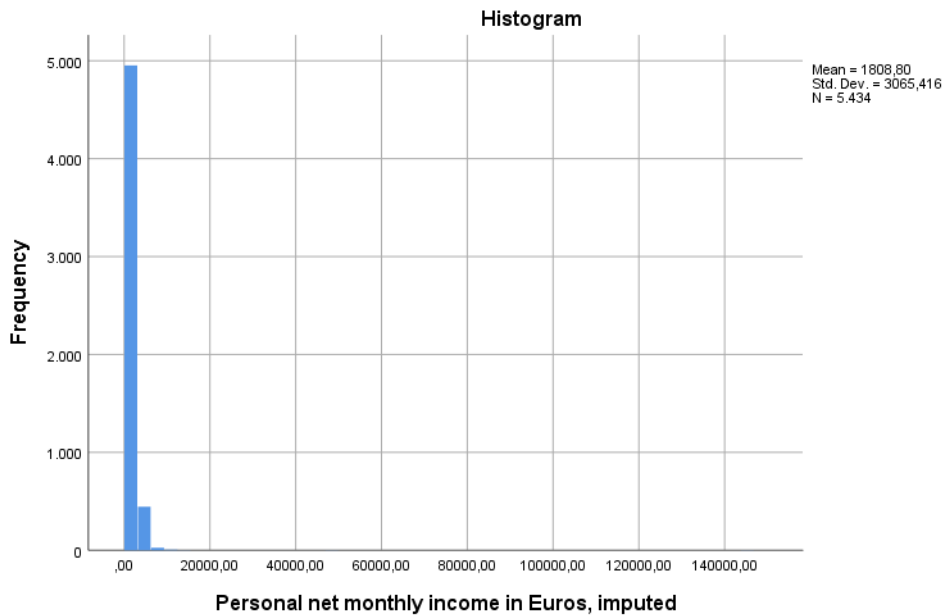
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	55	4	,3	,3	,3
	57	1	,1	,1	,4
	58	3	,3	,3	,7
	59	5	,4	,4	1,1
	60	41	3,6	3,6	4,7
	61	5	,4	,4	5,1
	62	40	3,5	3,5	8,6
	63	54	4,7	4,7	13,3
	64	34	3,0	3,0	16,3
	65	175	15,2	15,2	31,5
	66	98	8,5	8,5	40,0
	67	399	34,7	34,7	74,8
	68	171	14,9	14,9	89,6
	69	16	1,4	1,4	91,0
	70	63	5,5	5,5	96,5
	71	2	,2	,2	96,7
	72	6	,5	,5	97,2
	73	1	,1	,1	97,3
	75	12	1,0	1,0	98,3
	76	2	,2	,2	98,5
	77	2	,2	,2	98,7
	80	8	,7	,7	99,4
	85	3	,3	,3	99,7
87	1	,1	,1	99,7	
90	1	,1	,1	99,8	
99	1	,1	,1	99,9	
100	1	,1	,1	100,0	
Total		1149	100,0	100,0	



```
FREQUENCIES VARIABLES=cw20m289
  /HISTOGRAM
  /ORDER=ANALYSIS.
```

4. Inkomen (in dzd. Euro's)

4.1 Frequentieverdeling van de oorspronkelijke variabele, inclusief histogram.



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

4.2 Bewerkingen

In het oorspronkelijke databestand van de werk- en scholing module waren codering -9 en -8 toegekend aan respondenten die hadden geantwoord op een vraag met: -9 = ik weet het niet, -8= ik wil het niet zeggen. Deze waarden zijn gehercodeerd als system-missing.

Ter constructie van de initiële dataset - een dataset geconstrueerd met alleen respondenten die ten minste één valide score hebben- had ik een dummy-variabelen gemaakt van de variabele inkomen met antwoordcategorieën: 0 = een valide score, 1= system missing, -9 through -8. Er is voor iedere variabele in het model op dezelfde wijze een indicator-dummy gemaakt. Respondenten met een somscore van deze indicatoren gelijk aan het aantal variabelen in het model zijn uit het databestand verwijderd. De bovenstaande bewerkingen op deze wijze uitgevoerd in SPSS:

```
COMPUTE Indicator_inkomen=nettoink_f.
EXECUTE.
RECODE Indicator_inkomen (-8=1) (-9=1) (SYSMIS=1) (1 thru Highest=0).
EXECUTE.

COMPUTE SOMindicatoren=Indicator_gewenst + Indicator_verwacht +
Indicator_inkomen +
indicator_opleidingsniveau + Indicator_Fbelasting +
Indicator_vaardigheden +
Indicator_beroepssector + Indicator_kinderen.
EXECUTE.

FILTER OFF.
USE ALL.
SELECT IF (SOMindicatoren < 8).
EXECUTE.
```

Ter constructie van de complete dataset – waaruit de frequentieverdeling van de variabele zoals gebruikt in de analyse volgt- had ik (opnieuw) een indicator_dummy gemaakt van de variabele inkomen: met 0= valide score, 1= system missing. Er is voor iedere variabele in het model op dezelfde wijze een indicator-dummy gemaakt. Respondenten met een somscore van meer dan 0 op deze indicatoren – respondenten met dus ten minste één missende score - zijn uit het databestand verwijderd.

```
COMPUTE m3=0.
IF MISSING(nettoink_f) m3=1.
EXECUTE.

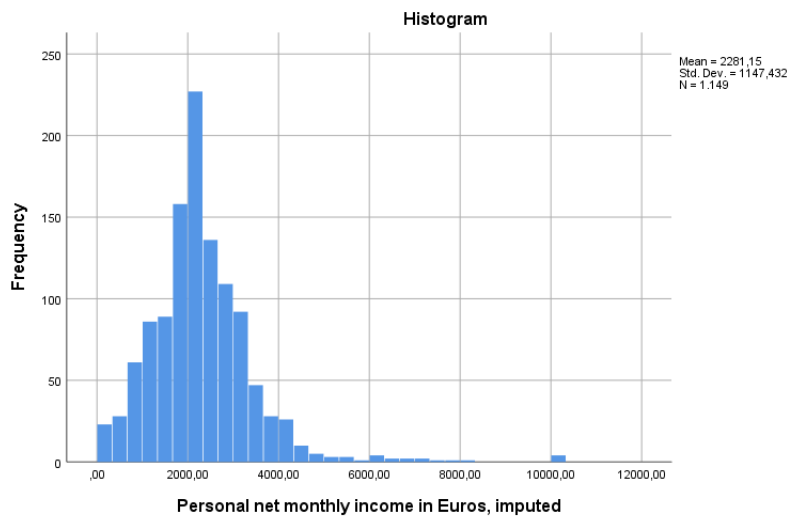
COMPUTE miss = m1+m2+m3+m4+m5+m6+m7+m8.
EXECUTE.

RECODE missendewaarden (0=0) (1 thru 7=1).
EXECUTE.

USE ALL.
COMPUTE filter_$=(miss = 0).
VARIABLE LABELS filter_$ 'miss = 0 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
```

```
FILTER BY filter_$.  
EXECUTE.
```

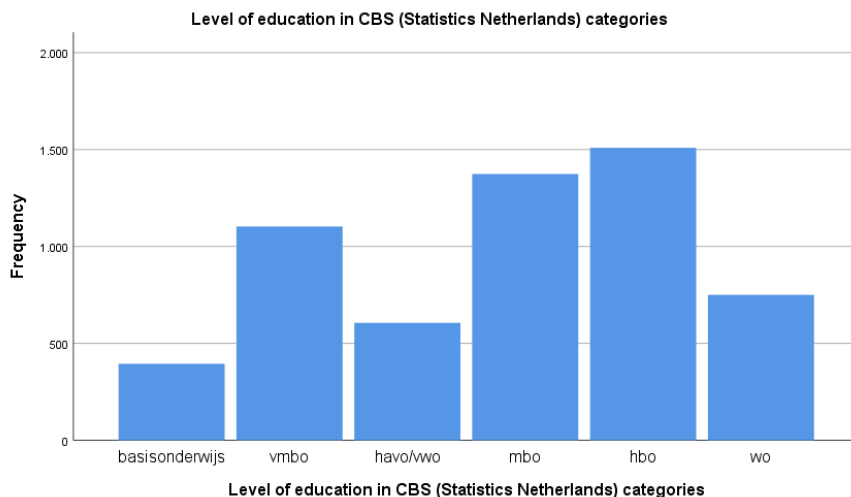
4.3 Frequentieverdeling van de uiteindelijke variabele in de analyse, inclusief histogram



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

5. Opleidingsniveau

5.1 Frequentieverdeling van de oorspronkelijke variabele, inclusief staafdiagram



oplc at Level of education in CBS (Statistics Netherlands) categories

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 basisonderwijs	395	6,9	6,9	6,9
	2 vmbo	1103	19,2	19,2	26,1
	3 havo/vwo	606	10,5	10,6	36,7
	4 mbo	1374	23,9	23,9	60,6
	5 hbo	1509	26,2	26,3	86,9
	6 wo	750	13,0	13,1	100,0
	Total	5737	99,7	100,0	
Missing	System	17	,3		
	Total	5754	100,0		

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

5.2 Bewerkingen

In het oorspronkelijke databestand van de werk- en scholing module waren codering -9 en -8 toegekend aan respondenten die hadden geantwoord op een vraag met: -9 = ik weet het niet, -8 = ik wil het niet zeggen. Deze waarden zijn gehercodeerd als system-missing.

Ter constructie van de initiële dataset - een dataset geconstrueerd met alleen respondenten die ten minste één valide score hebben - had ik een dummy-variabelen gemaakt van de variabele opleidingsniveau met antwoordcategorieën: 0 = een valide score, 1 = system missing, -9 through -8. Er is voor iedere variabele in het model op dezelfde wijze een indicator-dummy gemaakt. Respondenten met een somscore van deze indicatoren gelijk aan het aantal variabelen in het model zijn uit het databestand verwijderd. De bovenstaande bewerkingen op deze wijze uitgevoerd in SPSS:

```
COMPUTE Indicator_opleidingsniveau=oplcat.
EXECUTE.
RECODE Indicator_opleidingsniveau (-8=1) (-9=1) (SYSMIS=1) (1 thru
Highest=0).
EXECUTE.

COMPUTE SOMindicatoren=Indicator_gewenst + Indicator_verwacht +
Indicator_inkomen +
indicator_opleidingsniveau + Indicator_Fbelasting +
Indicator_vaardigheden +
Indicator_beroepssector + Indicator_kinderen.
EXECUTE.

FILTER OFF.
```

```
USE ALL.
SELECT IF (SOMindicatoren < 8).
EXECUTE.
```

Ter constructie van de complete dataset – waaruit de frequentieverdeling van de variabele zoals gebruikt in de analyse volgt- had ik (opnieuw) een indicator_dummy gemaakt van de variabele opleidingsniveau: met 0= valide score, 1= system missing. Er is voor iedere variabele in het model op dezelfde wijze een indicator-dummy gemaakt. Respondenten met een somscore van meer dan 0 op deze indicatoren – respondenten met dus ten minste één missende score - zijn uit het databestand verwijderd.

```
COMPUTE m4=0.
IF MISSING(oplcat) m4=1.
EXECUTE.
```

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

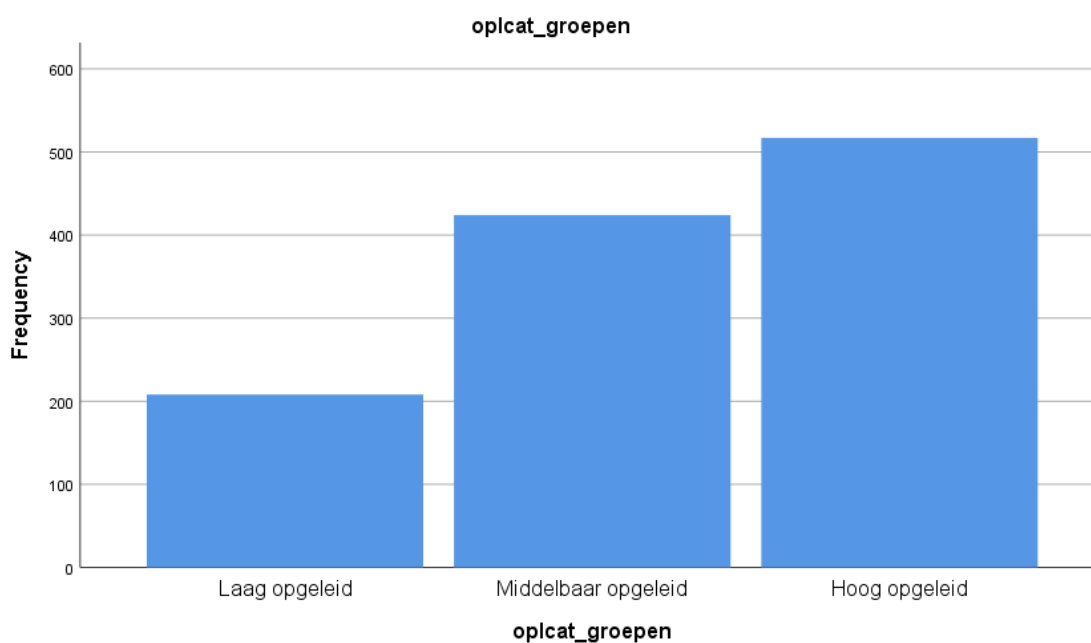
```
RECODE missendewaarden (0=0) (1 thru 7=1).
EXECUTE.
```

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

Tot slot heb ik de variabele opleidingsniveau gedummificeerd in drie dummy's gemaakt om uiteindelijk de verschillen tussen de groepen en de referentiegroep op de discrepantie in pensioenleeftijd te kunnen toelichten in de resultaten. De eerste dummy: respondenten met een laag opleidingsniveau – het basisonderwijs (1) en vmbo (2) - zijn het referentiekader. De tweede dummy is het opleidingsniveau middelbaar – havo/vwo (3) en mbo (4). De laatste dummy betreft de groep hooggeleide respondenten: hbo (5) en wo (6).

```
RECODE oplcat (1 thru 2=1) (ELSE=0) INTO D_REF_opleidingsniveau_laag.
EXECUTE.
RECODE oplcat (3 thru 4=1) (ELSE=0) INTO D_opleidingsniveau_middelbaar.
EXECUTE.
RECODE oplcat (5 thru 6=1) (ELSE=0) INTO D_opleidingsniveau_hoog.
EXECUTE.
```

5.3 Frequentieverdeling van de uiteindelijke variabele in de analyse, inclusief staafdiagram

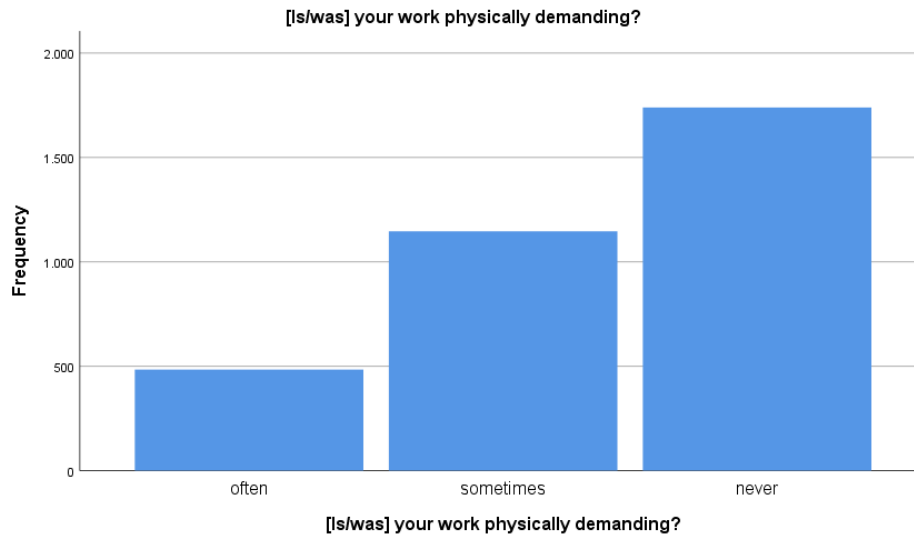


		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00 Laag opgeleid	208	18,1	18,1	18,1
	2,00 Middelbaar opgeleid	424	36,9	36,9	55,0
	3,00 Hoog opgeleid	517	45,0	45,0	100,0
Total		1149	100,0	100,0	

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

6. Fysieke Belasting

6.1 Frequentieverdeling van de oorspronkelijke variabele, inclusief staafdiagram



		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 often	484	8,4	14,4	14,4
	2 sometimes	1146	19,9	34,0	48,4
	3 never	1739	30,2	51,6	100,0
	Total	3369	58,6	100,0	
Missing	System	2385	41,4		
	Total	5754	100,0		

```
FREQUENCIES VARIABLES=cw20m416
  /BARCHART FREQ
  /ORDER=ANALYSIS.
```

6.2 Bewerkingen

In het oorspronkelijke databestand van de werk- en scholing module waren codering -9 en -8 toegekend aan respondenten die hadden geantwoord op een vraag met: -9 = ik weet het niet, -8 = ik wil het niet zeggen. Deze waarden zijn gehercodeerd als system-missing.

Ter constructie van de initiële dataset - een dataset geconstrueerd met alleen respondenten die ten minste één valide score hebben- had ik een dummy-variabelen gemaakt

van de variabele fysieke belasting met antwoordcategorieën: 0 = een valide score, 1= system missing, -9 through -8. Er is voor iedere variabele in het model op dezelfde wijze een indicator-dummy gemaakt. Respondenten met een somscore van deze indicatoren gelijk aan het aantal variabelen in het model zijn uit het databestand verwijderd. De bovenstaande bewerkingen op deze wijze uitgevoerd in SPSS:

```

COMPUTE Indicator_Fbelasting=cw20m416.
EXECUTE.
RECODE Indicator_Fbelasting (-8=1) (-9=1) (SYSMIS=1) (1 thru Highest=0).
EXECUTE.

COMPUTE      SOMindicatoren=Indicator_gewenst      +      Indicator_verwacht      +
Indicator_inkomen +
      indicator_opleidingsniveau      +      Indicator_Fbelasting      +
Indicator_vaardigheden +
      Indicator_beroepssector + Indicator_kinderen.
EXECUTE.

FILTER OFF.
USE ALL.
SELECT IF (SOMindicatoren < 8).
EXECUTE.

```

Ter constructie van de complete dataset – waaruit de frequentieverdeling van de variabele zoals gebruikt in de analyse volgt- had ik (opnieuw) een indicator_dummy gemaakt van de variabele fysieke belasting: met 0= valide score, 1= system missing. Er is voor iedere variabele in het model op dezelfde wijze een indicator-dummy gemaakt. Respondenten met een somscore van meer dan 0 op deze indicatoren – respondenten met dus ten minste één missende score - zijn uit het databestand verwijderd.

```

COMPUTE m5=0.
IF MISSING(cw20m416) m5=1.
EXECUTE.

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

RECODE missendewaarden (0=0) (1 thru 7=1).
EXECUTE.

USE ALL.
COMPUTE filter_$=(miss = 0).
VARIABLE LABELS filter_$ 'miss = 0 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

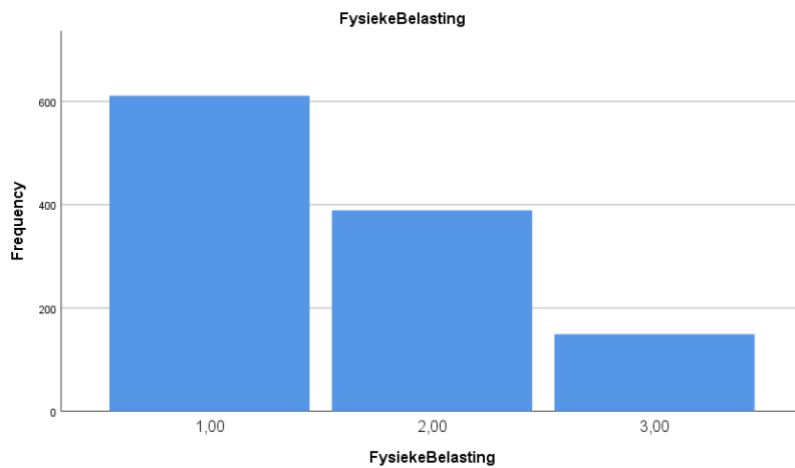
```

Tot slot heb ik- om de interpretatie van de variabele in de analyse te verduidelijken, de variabele gespiegeld (omgecodeerd). De oorspronkelijke gaaf score 3 aan het ‘nooit hebben van fysiek veeleisend werk’ Een score 1 aan het ‘‘vaak hebben van fysiek veeleisend werk’. Deze ordening is onlogisch, omdat fysieke belasting loopt van ‘vaak’ naar ‘nooit’. Een hoge

score op fysieke belasting duidt dan op fysiek niet belastend werk. Aan het hebben van fysiek niet-belastend werk wordt nu score 1 toegekend en aan het hebben van vaak fysiek belastend werk score 3.

```
RECODE cw20m416 (1=3) (2=2) (3=1) INTO FysiekeBelasting.
EXECUTE.
```

6.3 Frequentieverdeling van de uiteindelijke variabele in de analyse, inclusief staafdiagram

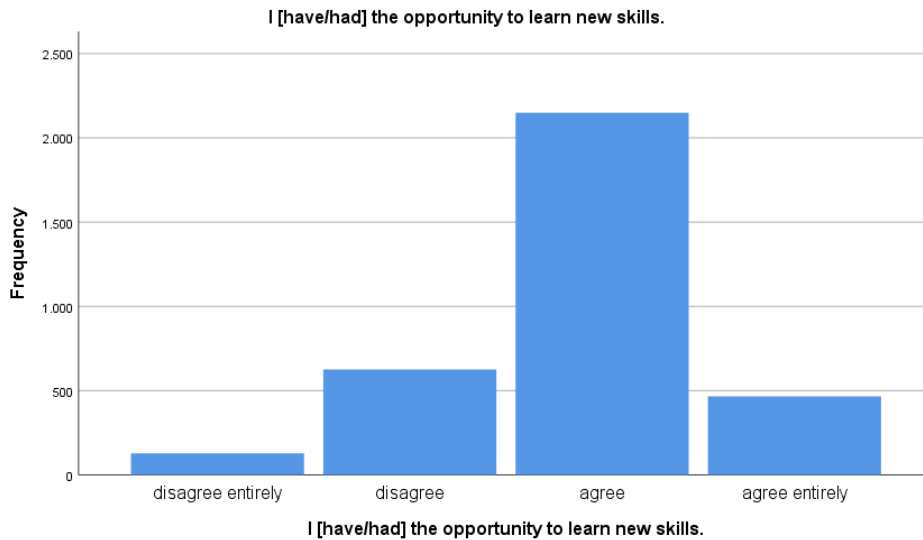


		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	611	53,2	53,2	53,2
	2,00	389	33,9	33,9	87,0
	3,00	149	13,0	13,0	100,0
Total		1149	100,0	100,0	

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

7. Trainingsmogelijkheden

7.1 Frequentieverdeling van de oorspronkelijke variabele, inclusief staafdiagram



cw20m430 I [have/had] the opportunity to learn new skills.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 disagree entirely	128	2,2	3,8	3,8
	2 disagree	626	10,9	18,6	22,4
	3 agree	2149	37,3	63,8	86,2
	4 agree entirely	466	8,1	13,8	100,0
	Total	3369	58,6	100,0	
Missing	System	2385	41,4		
Total		5754	100,0		

```
FREQUENCIES VARIABLES=cw20m430
  /BARChart FREQ
  /ORDER=ANALYSIS.
```

7.2 Bewerkingen

In het oorspronkelijke databestand van de werk- en scholing module waren codering -9 en -8 toegekend aan respondenten die hadden geantwoord op een vraag met: -9 = ik weet het niet, -8 = ik wil het niet zeggen. Deze waarden zijn gehercodeerd als system-missing.

Ter constructie van de initiële dataset - een dataset geconstrueerd met alleen respondenten die ten minste één valide score hebben- had ik een dummy-variabelen gemaakt van de variabele trainingsmogelijkheden met antwoordcategorieën: 0 = een valide score, 1=

system missing, -9 through -8. Er is voor iedere variabele in het model op dezelfde wijze een indicator-dummy gemaakt. Respondenten met een somscore van deze indicatoren gelijk aan het aantal variabelen in het model zijn uit het databestand verwijderd. De bovenstaande bewerkingen op deze wijze uitgevoerd in SPSS:

```

COMPUTE Indicator_vaardigheden=cw20m430.
EXECUTE.
RECODE Indicator_vaardigheden (-8=1) (-9=1) (SYSMIS=1) (1 thru Highest=0).
EXECUTE.

COMPUTE    SOMindicatoren=Indicator_gewenst    +    Indicator_verwacht    +
Indicator_inkomen +
    indicator_opleidingsniveau    +    Indicator_Fbelasting    +
Indicator_vaardigheden +
    Indicator_beroepssector + Indicator_kinderen.
EXECUTE.

FILTER OFF.
USE ALL.
SELECT IF (SOMindicatoren < 8).
EXECUTE.

```

Ter constructie van de complete dataset – waaruit de frequentieverdeling van de variabele zoals gebruikt in de analyse volgt- had ik (opnieuw) een indicator_dummy gemaakt van de variabele trainingsmogelijkheden: met 0= valide score, 1= system missing. Er is voor iedere variabele in het model op dezelfde wijze een indicator-dummy gemaakt. Respondenten met een somscore van meer dan 0 op deze indicatoren – respondenten met dus ten minste één missende score - zijn uit het databestand verwijderd.

```

COMPUTE m6=0.
IF MISSING(cw20m430) m6=1.
EXECUTE.

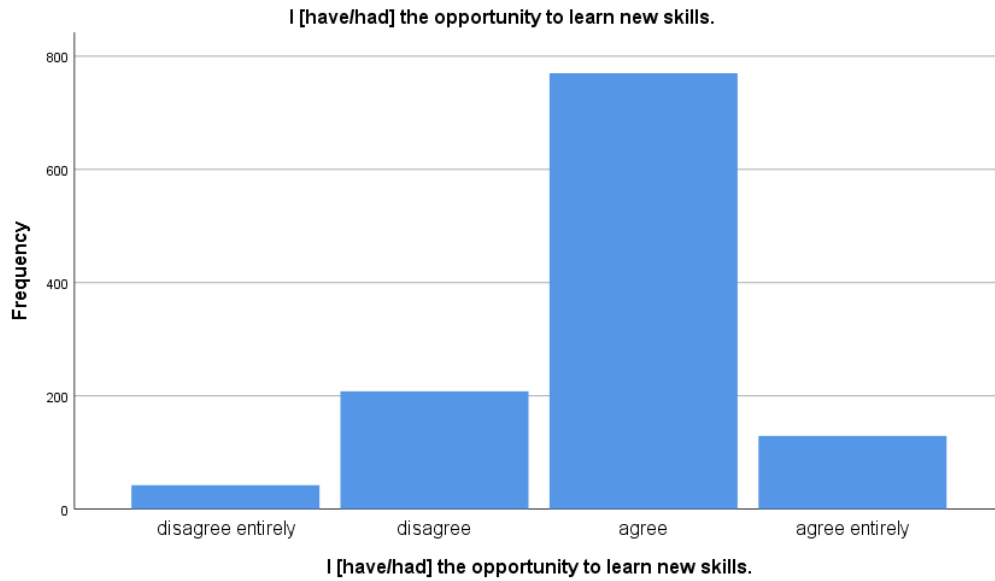
COMPUTE miss = m1+m2+m3+m4+m5+m6+m7+m8.
EXECUTE.

RECODE missendewaarden (0=0) (1 thru 7=1).
EXECUTE.

USE ALL.
COMPUTE filter_$=(miss = 0).
VARIABLE LABELS filter_$ 'miss = 0 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

```

7.3 Frequentieverdeling van de uiteindelijke variabele in de analyse, inclusief staafdiagram



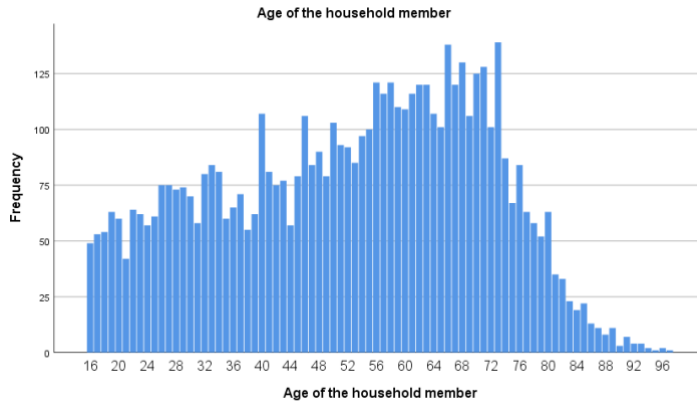
cw20m430 I [have/had] the opportunity to learn new skills.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 disagree entirely	42	3,7	3,7	3,7
	2 disagree	208	18,1	18,1	21,8
	3 agree	770	67,0	67,0	88,8
	4 agree entirely	129	11,2	11,2	100,0
	Total	1149	100,0	100,0	

FREQUENCIES VARIABLES=cw20m430
 /BARCHART FREQ
 /ORDER=ANALYSIS.

8. Leeftijd

8.1 Frequentieverdeling van de oorspronkelijke variabele, inclusief staafdiagram

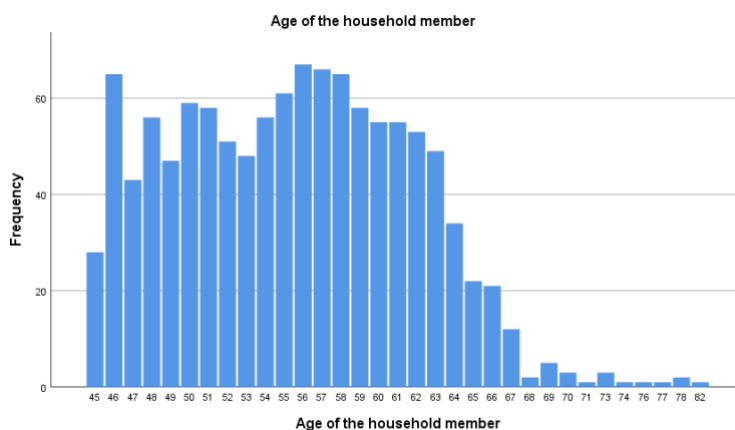


```
FREQUENCIES VARIABLES=leeftijd  
/BARChart FREQ  
/ORDER=ANALYSIS.
```

8.2 Bewerkingen

Er zijn geen bewerkingen uitgevoerd op deze variabele. Wel is de minimale leeftijd in de complete dataset ouder dan 44 jaar, omdat de vragen omtrent verwachte en gewenste pensioenleeftijd alleen zijn gesteld aan respondenten ouder dan 44. Dit is verder toegelicht in de methode paragraaf

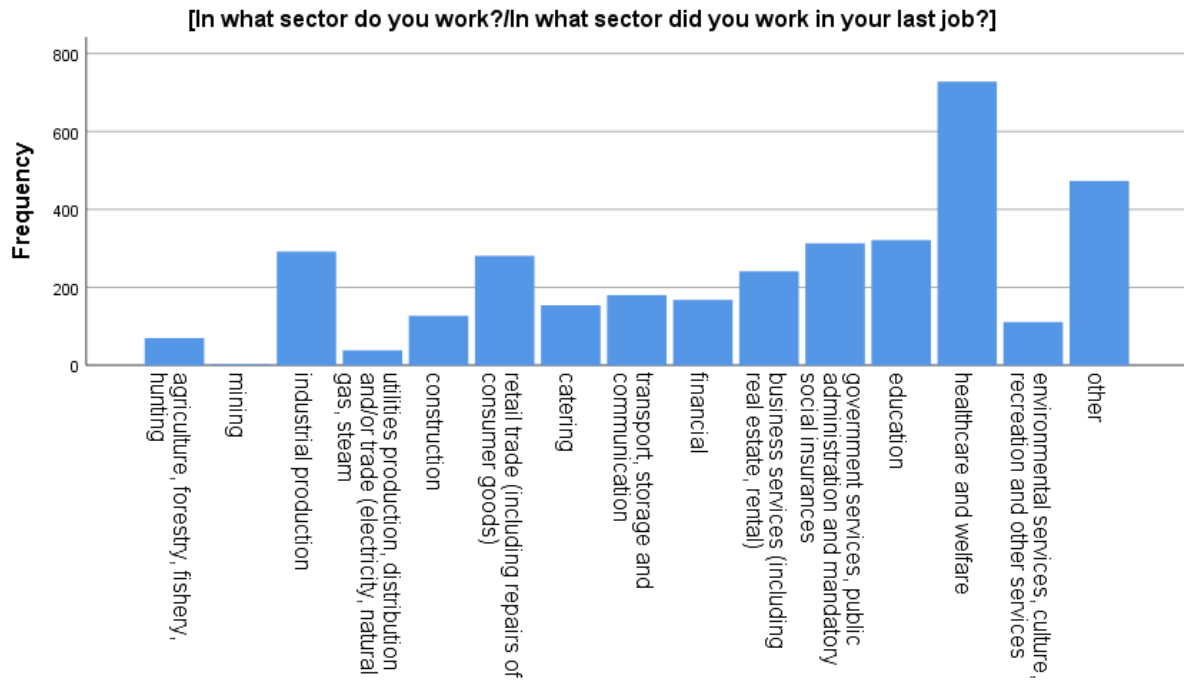
8.3 Frequentieverdeling van de uiteindelijke variabele in de analyse, inclusief staafdiagram



```
FREQUENCIES VARIABLES=leeftijd  
/BARChart FREQ  
/ORDER=ANALYSIS.
```

9. Beroepssector

9.1 Frequentieverdeling van de oorspronkelijke variabele, inclusief staafdiagram



cw20m402 [In what sector do you work?/In what sector did you work in your last job?]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 agriculture, forestry, fishery, hunting	70	1,2	2,0	2,0
	2 mining	2	,0	,1	2,1
	3 industrial production	292	5,1	8,3	10,4
	4 utilities production, distribution and/or trade (electricity, natural gas, steam)	38	,7	1,1	11,5
	5 construction	127	2,2	3,6	15,1
	6 retail trade (including repairs of consumer goods)	281	4,9	8,0	23,1
	7 catering	154	2,7	4,4	27,6
	8 transport, storage and communication	180	3,1	5,1	32,7
	9 financial	168	2,9	4,8	37,5
	10 business services (including real estate, rental)	241	4,2	6,9	44,4

	11 government services, public administration and mandatory social insurances	313	5,4	8,9	53,3
	12 education	321	5,6	9,2	62,5
	13 healthcare and welfare	728	12,7	20,8	83,3
	14 environmental services, culture, recreation and other services	111	1,9	3,2	86,5
	15 other	473	8,2	13,5	100,0
	Total	3499	60,8	100,0	
Missing	System	2255	39,2		
Total		5754	100,0		

```
FREQUENCIES VARIABLES=cw20m402
  /BARCHART FREQ
  /ORDER=ANALYSIS.
```

9.2 Bewerkingen

In het oorspronkelijke databestand van de werk- en scholing module waren codering -9 en -8 toegekend aan respondenten die hadden geantwoord op een vraag met: -9 = ik weet het niet, -8 = ik wil het niet zeggen. Deze waarden zijn gehercodeerd als system-missing.

Ter constructie van de initiële dataset - een dataset geconstrueerd met alleen respondenten die ten minste één valide score hebben- had ik een dummy-variabelen gemaakt van de variabele beroepssector met antwoordcategorieën: 0 = een valide score, 1= system missing, -9 through -8. Er is voor iedere variabele in het model op dezelfde wijze een indicator-dummy gemaakt. Respondenten met een somscore van deze indicatoren gelijk aan het aantal variabelen in het model zijn uit het databestand verwijderd. De bovenstaande bewerkingen op deze wijze uitgevoerd in SPSS:

```
COMPUTE Indicator_beroepssector=cw20m402.
EXECUTE.
RECODE Indicator_beroepssector (-8=1) (-9=1) (SYSMIS=1) (1 thru Highest=0).
EXECUTE.

COMPUTE SOMindicatoren=Indicator_gewenst + Indicator_verwacht +
Indicator_inkomen +
indicator_opleidingsniveau + Indicator_Fbelasting +
Indicator_vaardigheden +
Indicator_beroepssector + Indicator_kinderen.
EXECUTE.

FILTER OFF.
USE ALL.
SELECT IF (SOMindicatoren < 8).
EXECUTE.
```

Ter constructie van de complete dataset – waaruit de frequentieverdeling van de variabele zoals gebruikt in de analyse volgt- had ik (opnieuw) een indicator_dummy gemaakt van de variabele beroepssector: m7 met 0= valide score, 1= system missing. Er is voor iedere variabele in het model op dezelfde wijze een indicator-dummy gemaakt. Respondenten met een somscore van meer dan 0 op deze indicatoren – respondenten met dus ten minste één missende score - zijn uit het databestand verwijderd.

```
COMPUTE m7=0.
IF MISSING(cw20m402) m7=1.
EXECUTE.

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

RECODE missendewaarden (0=0) (1 thru 7=1).
EXECUTE.

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

Tot slot heb ik de variabele beroepssector gedummificeerd in vier dummy's om uiteindelijk de verschillen tussen de groepen en de referentiegroep op de discrepantie in pensioenleeftijd te kunnen analyseren. Hierbij heb ik het aantal antwoordcategorieën teruggebracht van 15 naar 4, waarbij de categorie 'anders' is toegevoegd aan de dummy met de minste respondenten. In de beschrijving van de dummyvariabelen staan de oorspronkelijke antwoordcategorieën weergegeven tussen de haakjes.

De eerste dummy, '*de primaire sector + anders*' omvat de respondenten die werken in de sectoren– land- en bosbouw, de jachtsector en de visserij (1), industrie (2), productie van nutsvoorzieningen (4), de bouwsector (5) en anders (15). Deze dummy is de referentiegroep.

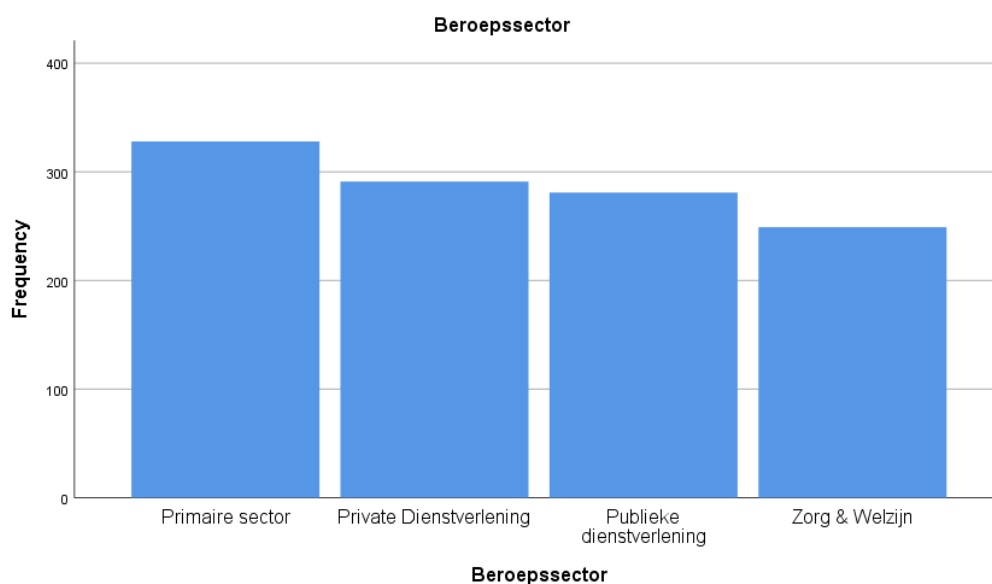
De tweede dummy - '*de private dienstverlening*' - omvat de respondenten die werken in de sectoren: Retail- en detailhandel (6), catering/horeca (7), transportsector (8), financiële dienstverlening (9) en commerciële dienstverlening (10).

De derde dummy – '*de publieke dienstverlening*' – omvat respondenten die werken in de sectoren: overheidsdiensten, openbaar bestuur en zelfstandige bestuursorganen (11), de onderwijssector (12) en milieudiensten, cultuur, recreatie en overige dienstverlening (14).

De vierde en laatste dummy – ‘zorg en welzijn’- omvat respondenten die werken in de zorgsector (13).

```
RECODE cw20m402 (1=1) (3=1) (4=1) (5=1) (15=1) (ELSE=0) INTO
Primare_sector_REF.
EXECUTE.
RECODE cw20m402 (6=1) (7=7) (8=1) (9=1) (10=1) (ELSE=0) INTO
Zakelijke_dienstverlenin.
EXECUTE.
RECODE cw20m402 (11=1) (12=1) (14=1) (ELSE=0) INTO Publieke_dienstverlenin.
EXECUTE.
RECODE cw20m402 (13=1) (ELSE=0) INTO Zorg_Welzijn.
EXECUTE.
```

9.3 Frequentieverdeling van de uiteindelijke variabele in de analyse, inclusief staafdiagram



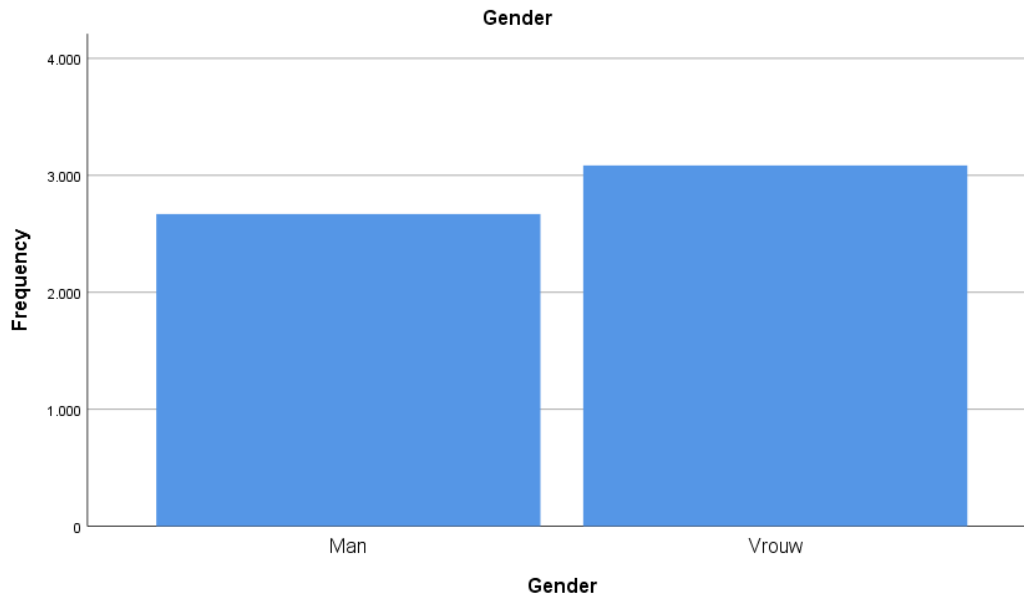
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00 Primaire sector	328	28,5	28,5	28,5
	2,00 Private Dienstverlening	291	25,3	25,3	53,9
	3,00 Publieke dienstverlening	281	24,5	24,5	78,3
	4,00 Zorg & Welzijn	249	21,7	21,7	100,0
	Total	1149	100,0	100,0	

FREQUENCIES VARIABLES=Beroepssector

```
/BARCHART FREQ
/ORDER=ANALYSIS.
```

10. Geslacht

10.1 Frequentieverdeling van de oorspronkelijke variabele, inclusief staafdiagram



		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Man	2669	46,4	46,4	46,4
	2 Vrouw	3085	53,6	53,6	100,0
	Total	5754	100,0	100,0	

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

10.2 Bewerkingen

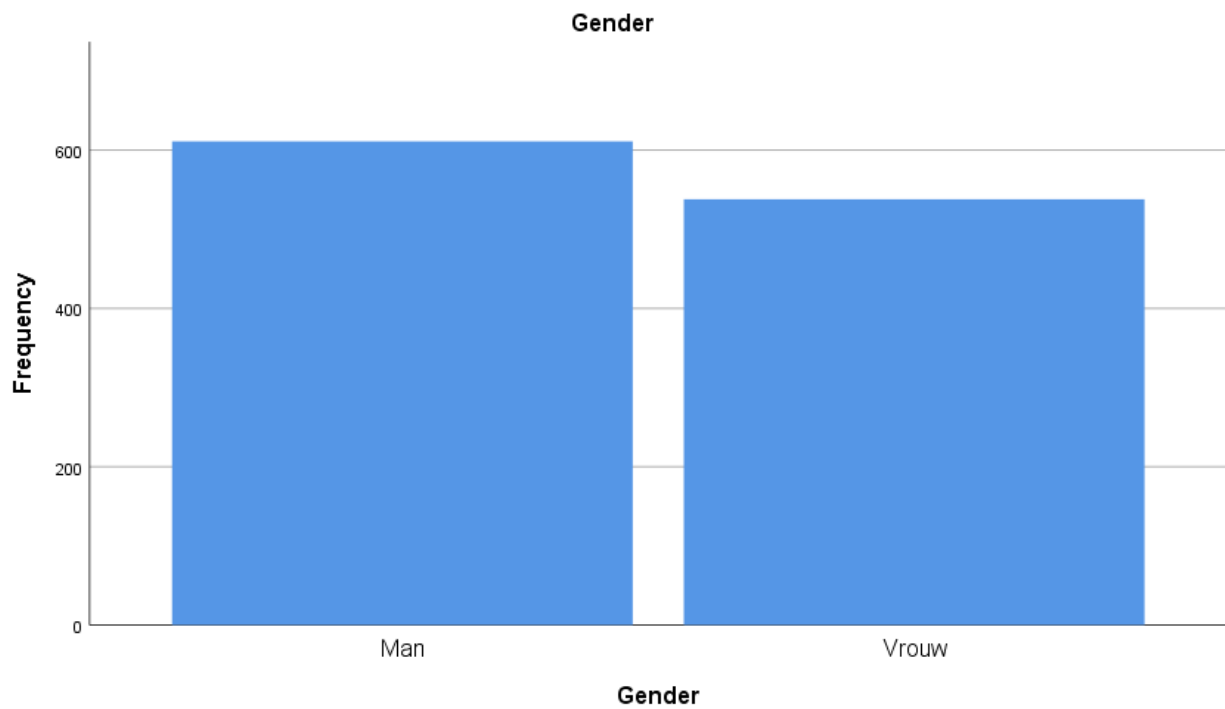
Van de variabele geslacht heb ik dummy variabelen gemaakt. De dummy man is de referentiegroep.

```
RECODE geslacht (1=1) (2=0) INTO D_man.
EXECUTE.
```

```
RECODE geslacht (2=1) (1=0) INTO D_vrouw.
```

EXECUTE .

10.3 Frequentieverdeling van de uiteindelijke variabele in de analyse, inclusief staafdiagram



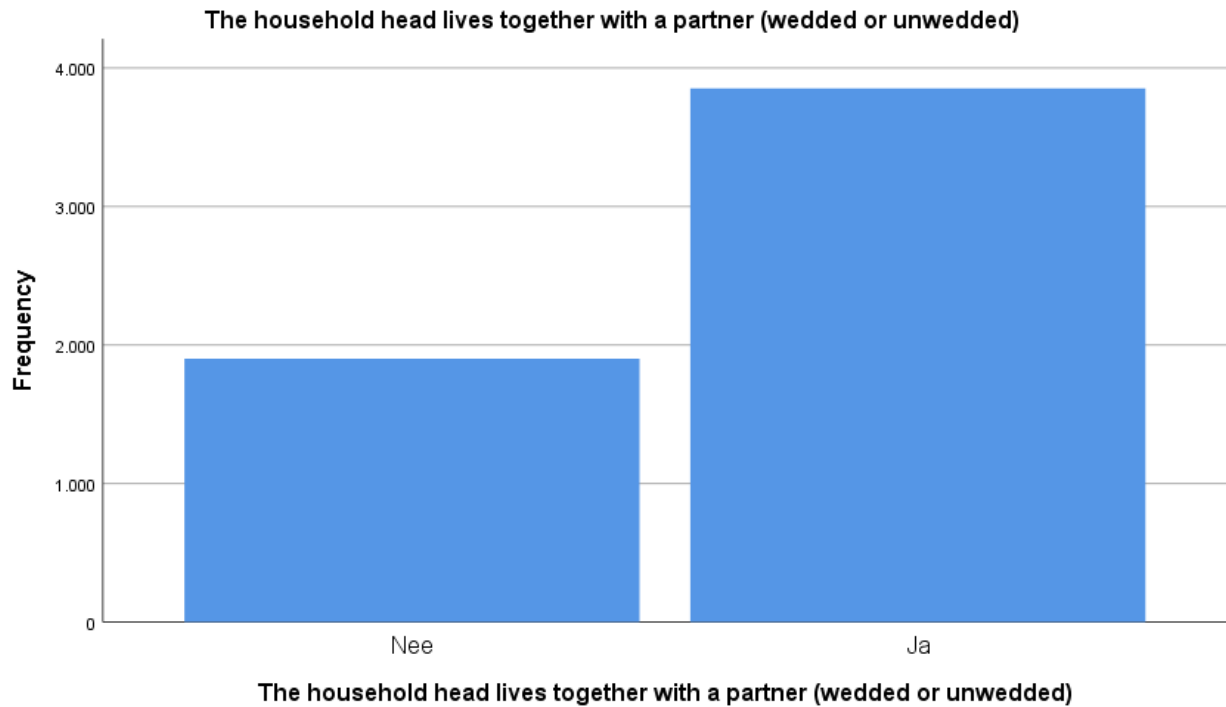
geslacht Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Man	611	53,2	53,2	53,2
	2 Vrouw	538	46,8	46,8	100,0
	Total	1149	100,0	100,0	

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

11. Partnerstatus

11.1 Frequentieverdeling van de oorspronkelijke variabele, inclusief staafdiagram



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

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 Nee	1901	33,0	33,0	33,0
	1 Ja	3853	67,0	67,0	100,0
	Total	5754	100,0	100,0	

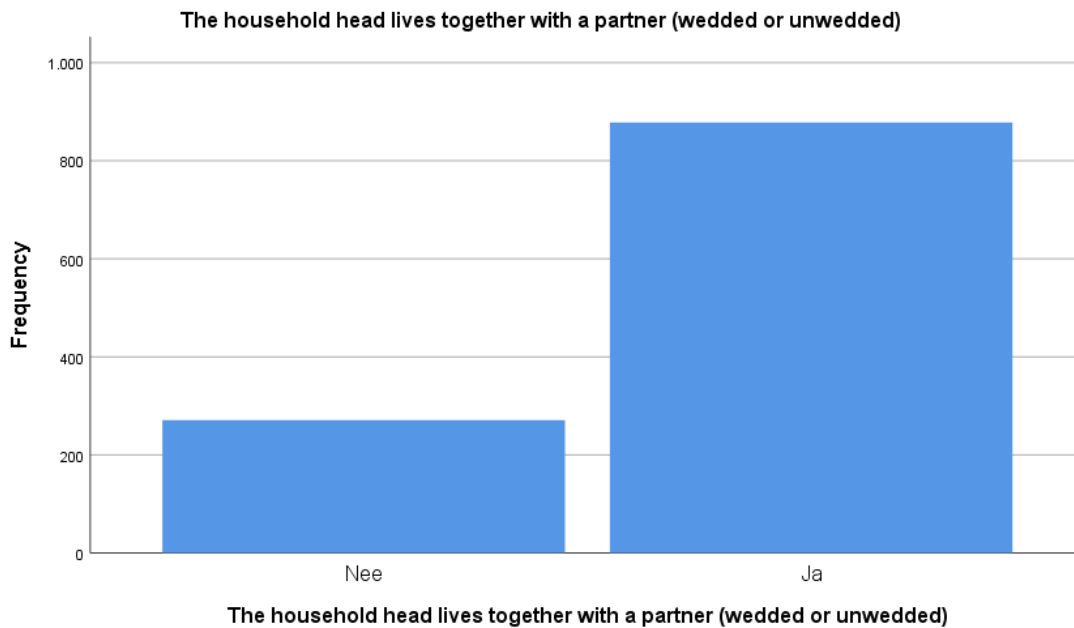
```
FREQUENCIES VARIABLES=partner
  /BARCHART FREQ
  /ORDER=ANALYSIS.
```

11.2 Bewerkingen

Van de variabele partnerstatus heb ik dummy variabelen gemaakt. De groep respondenten zonder partner (single) zijn de referentiegroep.

```
RECODE partner (0=1) (1=0) INTO D_Single.
EXECUTE.
RECODE partner (1=1) (0=0) INTO D_Partner.
EXECUTE
```

11.3 Frequentieverdeling van de uiteindelijke variabele in de analyse, inclusief staafdiagram



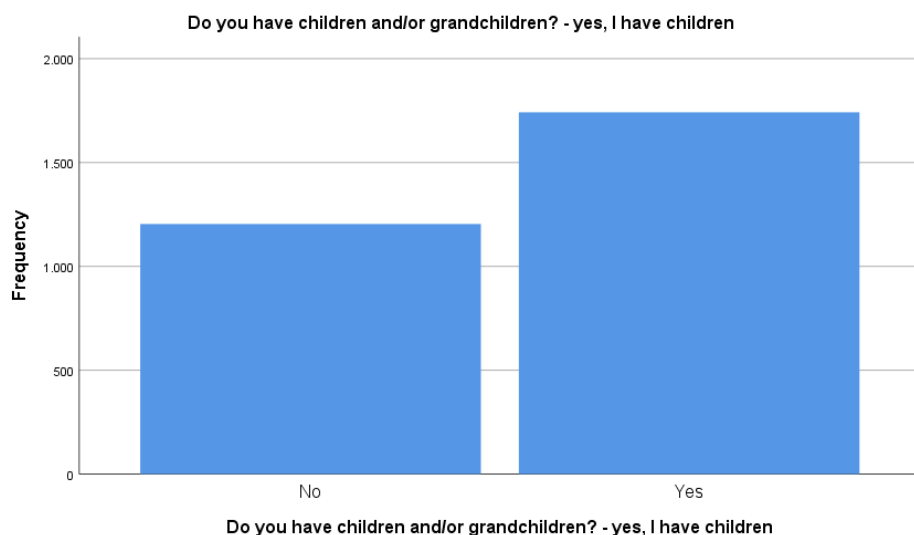
partner The household head lives together with a partner (wedded or unwedded)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 Nee	271	23,6	23,6	23,6
	1 Ja	878	76,4	76,4	100,0
	Total	1149	100,0	100,0	

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

12. Ouderschap

12.1 Frequentieverdeling van de oorspronkelijke variabele, inclusief staafdiagram



cw20m436 Do you have children and/or grandchildren? - yes, I have children

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 No	1204	20,9	40,9	40,9
	1 Yes	1742	30,3	59,1	100,0
	Total	2946	51,2	100,0	
Missing	System	2808	48,8		
Total		5754	100,0		

```
FREQUENCIES VARIABLES=cw20m436
/BARCHART FREQ
/ORDER=ANALYSIS.
```

12.2 Bewerkingen

In het oorspronkelijke databestand van de werk- en scholing module waren codering -9 en -8 toegekend aan respondenten die hadden geantwoord op een vraag met: -9 = ik weet het niet, -8 = ik wil het niet zeggen. Deze waarden zijn gehercodeerd als system-missing.

Ter constructie van de initiële dataset - een dataset geconstrueerd met alleen respondenten die ten minste één valide score hebben- had ik een dummy-variabelen gemaakt van de variabele kinderen met antwoordcategorieën: 0 = een valide score, 1= system missing, -9 through -8. Er is voor iedere variabele in het model op dezelfde wijze een indicator-dummy gemaakt. Respondenten met een somscore van deze indicatoren gelijk aan het aantal variabelen

in het model zijn uit het databestand verwijderd. De bovenstaande bewerkingen op deze wijze uitgevoerd in SPSS:

```
COMPUTE Indicator_kinderen=cw20m436.
EXECUTE.
RECODE Indicator_kinderen (-8=1) (-9=1) (SYSMIS=1) (1 thru Highest=0).
EXECUTE.

COMPUTE      SOMindicatoren=Indicator_gewenst      +      Indicator_verwacht      +
Indicator_inkomen +
      indicator_opleidingsniveau      +      Indicator_Fbelasting      +
Indicator_vaardigheden +
      Indicator_beroepssector + Indicator_kinderen.
EXECUTE.

FILTER OFF.
USE ALL.
SELECT IF (SOMindicatoren < 8).
EXECUTE.
```

Ter constructie van de complete dataset – waaruit de frequentieverdeling van de variabele zoals gebruikt in de analyse volgt- had ik (opnieuw) een indicator_dummy gemaakt van de variabele kinderen: m8 met 0= valide score, 1= system missing. Er is voor iedere variabele in het model op dezelfde wijze een indicator-dummy gemaakt. Respondenten met een somscore van meer dan 0 op deze indicatoren – respondenten met dus ten minste één missende score - zijn uit het databestand verwijderd.

```
COMPUTE m8=0.
IF MISSING(cw20m402) m8=1.
EXECUTE.

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

RECODE missendewaarden (0=0) (1 thru 7=1).
EXECUTE.

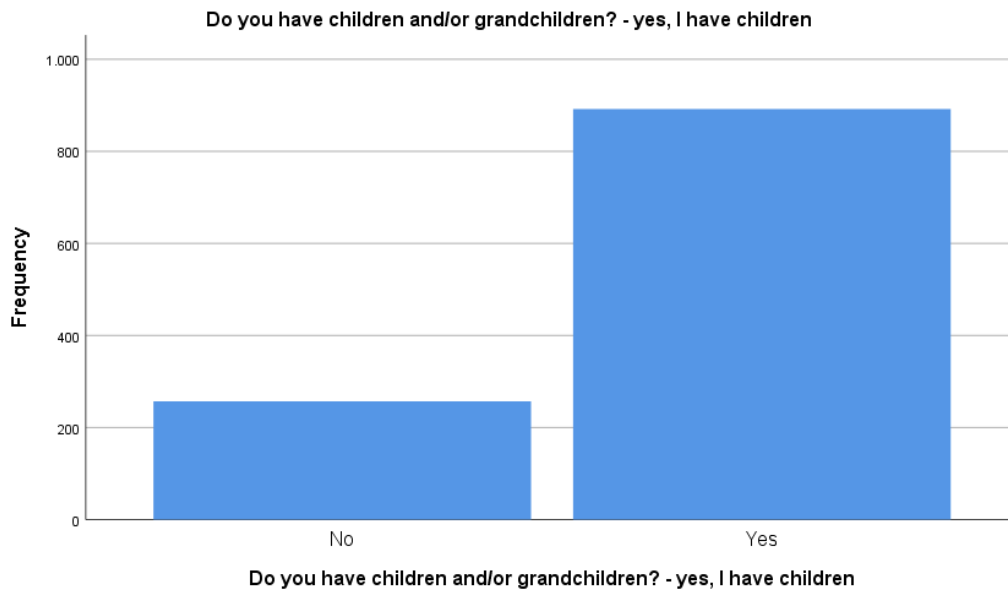
USE ALL.
COMPUTE filter_$=(miss = 0).
VARIABLE LABELS filter_$ 'miss = 0 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.
```

Tot slot heb ik de variabele kinderen gedummificeerd. De groep respondenten zonder kinderen (kindloos) is de referentiegroep.

```
RECODE cw20m436 (0=1) (1=0) INTO D_kindloos_ref.
```

```
EXECUTE.
RECODE cw20m436 (1=1) (0=0) INTO D_met_kinderen.
EXECUTE.
```

11.3 Frequentieverdeling van de uiteindelijke variabele in de analyse, inclusief staafdiagram



cw20m436 Do you have children and/or grandchildren? - yes, I have children

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 No	257	22,4	22,4	22,4
	1 Yes	892	77,6	77,6	100,0
	Total	1149	100,0	100,0	

```
FREQUENCIES VARIABLES=cw20m436
  /BARCHART FREQ
  /ORDER=ANALYSIS.
```


Bijlage 2

In deze tweede bijlage wordt de syntax en output van en toelichting gegeven op de analyses die ik heb uitgevoerd. Per deelsectie van de resultatenparagraaf geef ik deze toelichting. Te beginnen met de univariate statistieken (§1), gevolgd door de bivariate statistieken (§2) en daarna de modelevalautie (§3). Tot slot licht ik de hypothesetoetsing (§4).

1. Univariate beschrijvende statistieken

Tabel 1: Gemiddelden, standaarddeviatie, minimum en maximum van discrepantie in pensioenleeftijd, gewenste pensioenleeftijd, verwachte pensioenleeftijd, inkomen (in dzd. Euro 's) en leeftijd (N=1149)

	N	Minimum	Maximum	Mean	Std. Deviation
Discrepantie	1149	.00	50,00	3,2985	3,99178
cw20m288 If you were able to choose for yourself, at which age would you [want/have wante	1149	20	99	63,20	4,201
cw20m289 At what age do you expect [to retire or take early retirement or] to stop workin	1149	55	100	66,50	3,445
Inkomen_dzd_euros	1149	.00	10,00	2,2812	1,14743
leeftijd Age of the household member	1149	45	82	55,46	6,231
Valid N (listwise)	1149				

```
DESCRIPTIVES VARIABLES=Discrepantie cw20m288 cw20m289 Inkomen_dzd_euros
leeftijd
/STATISTICS=MEAN STDDEV MIN MAX.
```

In tabel 1 worden de univariate beschrijvende statistieken van de variabelen met een ratio (continu) meetniveau gegeven. De discrepantie in pensioenleeftijd is een verschilscore tussen de verwachte en gewenste pensioenleeftijd.

In de tabellen 2 tot en met 8 worden de relatieve frequentieverdelingen van de ordinale en nominale variabelen in het model gegeven. In de resultatenparagraaf zijn alleen de verdelingen van de voorspellende variabelen toegelicht. Onderstaand worden ook de verdelingen van controlevariabelen: geslacht, partnerstatus en ouderschap gegeven.

Tabel 2: de frequentieverdelingen van het opleidingsniveau (N=1149)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00 Laag opgeleid	208	18,1	18,1	18,1
	2,00 Middelbaar opgeleid	424	36,9	36,9	55,0
	3,00 Hoog opgeleid	517	45,0	45,0	100,0
	Total	1149	100,0	100,0	

Tabel 3: de frequentieverdelingen van de ervaren fysieke belasting (N=1149)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00 never	611	53,2	53,2	53,2
	2,00 sometimes	389	33,9	33,9	87,0
	3,00 often	149	13,0	13,0	100,0
	Total	1149	100,0	100,0	

Tabel 4: de frequentieverdelingen van de mogelijkheden om deel te nemen aan trainingsactiviteiten (N=1149)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 disagree entirely	42	3,7	3,7	3,7
	2 disagree	208	18,1	18,1	21,8
	3 agree	770	67,0	67,0	88,8
	4 agree entirely	129	11,2	11,2	100,0
	Total	1149	100,0	100,0	

Tabel 5: de frequentieverdelingen van geslacht (N=1149)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Man	611	53,2	53,2	53,2
	2 Vrouw	538	46,8	46,8	100,0
	Total	1149	100,0	100,0	

Tabel 6: de frequentieverdelingen van partnerstatus (N=1149, 1= met parter)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 Nee	271	23,6	23,6	23,6
	1 Ja	878	76,4	76,4	100,0
	Total	1149	100,0	100,0	

Tabel 7: de frequentieverdelingen van ouderschap (N=1149, 1= met kinderen)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 No	257	22,4	22,4	22,4
	1 Yes	892	77,6	77,6	100,0
	Total	1149	100,0	100,0	

Tabel 8: de frequentieverdelingen van beroepssector (N=1149)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00 Primaire sector	328	28,5	28,5	28,5
	2,00 Private Dienstverlening	291	25,3	25,3	53,9
	3,00 Publieke dienstverlening	281	24,5	24,5	78,3
	4,00 Zorg & Welzijn	249	21,7	21,7	100,0
	Total	1149	100,0	100,0	

```
FREQUENCIES VARIABLES=oplcat_groepen FysiekeBelasting cw20m430 geslacht
partner cw20m436 Beroepssector
/STATISTICS=MEDIAN
/ORDER=ANALYSIS
```

2. Bivariate beschrijvende statistieken

In de resultatenparagraaf wordt een tabel gegeven met de associaties tussen alle variabelen in het onderzoek. Het doel van deze tabel is een eerste indruk krijgen van de verbanden in het onderzoeksmodel. In deze deelsectie zal ik de associaties tussen de onafhankelijke variabele en alle andere variabelen in het model toelichten, alsmede de opvallende associatie. Hierbij is het belangrijk om te noemen dat ik gebruik maak van verschillende associatiematen en dat deze verschillend moeten worden geïnterpreteerd.

De associaties tussen continue variabelen en continue en dichotome variabelen zijn berekend aan de hand van de Pearson's correlatie. Dit getal drukt de sterkte van de associatie tussen de twee variabelen uit. Voor de associaties tussen twee categorische variabelen wordt de Cramer's V gebruikt. Dit getal is een relatieve maat voor de sterkte van de samenhang. Deze maat zegt niets over de richting van het verband. De Phi-coëfficiënt wordt gebruikt als maat voor de associatie tussen dichotome variabelen. Deze is gelijk aan het Pearson's correlatiecoëfficiënt

1. Correlaties tussen de continue en dichotome variabelen.

Tabel 9: Associaties tussen alle continue en dichotome variabelen in het model (N=1149).

Discrepancie	Pearson Correlation	1	-,647**	,369**	-,015	-,276**	-,019	-,038	,000
	Sig. (2-tailed)		,000	,000	,614	,000	,519	,197	,990
Gewenste pensioenleeftijd	Pearson Correlation	-,647**	1	,469**	,093**	,336**	-,040	,011	-,021
	Sig. (2-tailed)	,000		,000	,002	,000	,179	,701	,483
Verwachte pensioenleeftijd	Pearson Correlation	,369**	,469**	1	,096**	,090**	-,070*	-,030	-,025
	Sig. (2-tailed)	,000	,000		,001	,002	,017	,305	,400
Inkomen_dzd_euros	Pearson Correlation	-,015	,093**	,096**	1	,029	-,384**	-,011	,073*
	Sig. (2-tailed)	,614	,002	,001		,322	,000	,703	,013
leeftijd	Pearson Correlation	-,276**	,336**	,090**	,029	1	-,044	,006	,055
	Sig. (2-tailed)	,000	,000	,002	,322		,138	,840	,063
geslacht Gender	Pearson Correlation	-,019	-,040	-,070*	-,384**	-,044	1	-,054	,035
	Sig. (2-tailed)	,519	,179	,017	,000	,138		,068	,237
partner	Pearson Correlation	-,038	,011	-,030	-,011	,006	-,054	1	,272**
	Sig. (2-tailed)	,197	,701	,305	,703	,840	,068		,000
ouderschap	Pearson Correlation	,000	-,021	-,025	,073*	,055	,035	,272**	1
	Sig. (2-tailed)	,990	,483	,400	,013	,063	,237	,000	

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

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

Uit de tabel 9 volgt een significante en sterk negatieve associatie tussen de discrepantie in pensioenleeftijd en de leeftijd waarop respondenten wensen te pensioneren ($r = -.647; p < 0.01$). Een negatieve correlatie duidt erop dat een hoge score op de ene variabele samenhangt met een lage score op de andere variabele. In dit geval is er dus een tegengesteld verband tussen een score op de discrepantie in pensioenleeftijd en verwachte pensioenleeftijd. Indien een respondent vroeg wenst te pensioneren, hangt dit samen met een hoge score op de discrepantie en wanneer een respondenten laat wenst pensioneren hangt dit samen met een lage discrepantie.

Daarnaast is er een significant positieve correlatie tussen de discrepantie en de verwachte pensioenleeftijd ($r = .369; p < 0.01$). Een score op de verwachte pensioenleeftijd hangt dus in positieve mate samen met een score op de discrepantie in pensioenleeftijd. Dit verband is – ondanks de significantie - aanzienlijk minder sterk de correlatie tussen gewenste pensioenleeftijd en de discrepantie. Daarnaast is de associatie positief in plaats van negatief.

Aan de hand van de theorie voorspelde ik een negatieve associatie tussen de discrepantie in het maandelijks netto-inkomen van een respondent. In tabel 2 is er een negatieve samenhang tussen discrepantie en inkomen zichtbaar, maar deze is zwak en niet significant ($r = -.015$ en $p = .614$). Daarnaast is er een zwakke niet-significante associatie zichtbaar tussen de discrepantie in pensioenleeftijd en het opleidingsniveau van respondenten ($r = .052^a$ en $p = .142$) en de mogelijkheid om deel te nemen aan trainingsactiviteiten ($r = .041^a$ en $p = .166$). Bovendien zegt deze maat op basis van de proportie verklaarde variantie niets over de richting van het verband: dit getal is altijd positief. Respondenten met een laag opleidingsniveau een ervaren gemiddeld geen hoger of lagere discrepantie dan respondenten met een middelbaar of hoog opleidingsniveau. De gemiddelde discrepantie verschilt niet tussen de opleidingsniveaus.

Daarentegen is er een significant- en positieve associatie tussen de discrepantie in pensioenleeftijd en de ervaren fysieke belasting van het werk ($r = .100$ en $p < 0.01$). Het vaak ervaren van het werk als fysiek belastend hangt samen met een hogere scores op de discrepantie in pensioenleeftijd. Er kan worden gesteld dat respondenten die hun werk als fysiek belastend ervaren gemiddeld meer verschil hebben tussen hun gewenste en verwachte pensioenleeftijd dan respondenten die hun werk niet zo ervaren.

Verder volgt uit tabel 2 een negatief en significant verband tussen de discrepantie en de leeftijd van een respondent ($r = -.276$ en $p < 0.01$). Naarmate respondenten ouder worden lijken zij dus een kleiner verschil te percipiëren in hun gewenste en verwachte pensioenleeftijd. Aangezien respondenten niet met terugwerkende kracht kunnen pensioneren zal hun gewenste pensioenleeftijd waarschijnlijk meestijgen met hun leeftijd. Deze waarschijnlijkheid is terug te zien in het significante en positieve verband tussen gewenste pensioenleeftijd en leeftijd ($r = .332$ en $p < 0.01$).

Tot slot volgen er uit tabel 2 geen sterke associaties die mogelijk op multicollineariteit zouden kunnen duiden. Een sterk verband tussen voorspellers is onwenselijk, omdat de voorspellers elkaar dan voorspellen en er geen extra variantie wordt verklaard in de discrepantie in pensioenleeftijd. Of de verbanden tussen de variabelen daadwerkelijk problematisch zijn zal ik beoordelen tijdens de modevaluatie aan de hand van VIF-scores en partiële correlaties.

2. Assicuatues tussen de dichotome variabelen (Phi-coëfficiënt)

Tabel 10: associatie tussen geslacht en partnerstatus (N=1149).

		Symmetric Measures	
		Value	Approximate Significance
Nominal by Nominal	Phi	-,054	,068
	Cramer's V	,054	,068
N of Valid Cases		1149	

Deze correlatie is niet opvallend en licht ik daarin niet nader toe.

CROSSTABS

```
/TABLES=geslacht BY partner
/FORMAT=AVALUE TABLES
/STATISTICS=PHI
/CELLS=COUNT
/COUNT ROUND CELL.
```

Tabel 11: associatie tussen geslacht en ouderschap (N=1149)

		Symmetric Measures	
		Value	Approximate Significance
Nominal by Nominal	Phi	,035	,237
	Cramer's V	,035	,237
N of Valid Cases		1149	

Deze associatie is niet opvallend en licht ik daarom niet nader toe

CROSSTABS

```
/TABLES=geslacht BY cw20m436
/FORMAT=AVALUE TABLES
/STATISTICS=PHI
/CELLS=COUNT
/COUNT ROUND CELL.
```

Tabel 12: associatie tussen geslacht en partnerstatus (N=1149)

		Symmetric Measures	
		Value	Approximate Significance
Nominal by Nominal	Phi	,272	,000
	Cramer's V	,272	,000
N of Valid Cases		1149	

Er is een gematigde associatie tussen geslacht en partnerstatus. Naar mijns inziens logisch, omdat mensen – in de meeste gevallen – pas kinderen nemen als zij een partner nemen.

CROSSTABS

```
/TABLES=partner BY cw20m436
/FORMAT=AVALUE TABLES
/STATISTICS=PHI
/CELLS=COUNT
/COUNT ROUND CELL.
```

3. Associaties tussen de categorische variabelen met meerdere antwoordcategorieën en dichotome variabelen (Cramers' V)

Tabel 13: associatie tussen opleidingsniveau en geslacht (N=1149)

		Symmetric Measures	
		Value	Approximate Significance
Nominal by Nominal	Phi	,074	,280
	Cramer's V	,074	,280
N of Valid Cases		1149	

Deze associatie is niet opvallend en licht ik niet nader toe.

```
CROSSTABS
  /TABLES=oplcat BY geslacht
  /FORMAT=AVALUE TABLES
  /STATISTICS=PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Tabel 14: associatie tussen opleidingsniveau en partnerstatus (N=1149)

		Symmetric Measures	
		Value	Approximate Significance
Nominal by Nominal	Phi	,063	,481
	Cramer's V	,063	,481
N of Valid Cases		1149	

Deze associatie is niet opvallend en licht ik niet nader toe.

```
CROSSTABS
  /TABLES=oplcat BY partner
  /FORMAT=AVALUE TABLES
  /STATISTICS=PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Tabel 15: associatie tussen opleidingsniveau en ouderschap (N=1149)

		Symmetric Measures	
		Value	Approximate Significance
Nominal by Nominal	Phi	,080	,192
	Cramer's V	,080	,192
N of Valid Cases		1149	

Deze associatie is niet opvallend en licht ik niet nader toe.

```
CROSSTABS
  /TABLES=oplcat BY cw20m436
  /FORMAT=AVALUE TABLES
  /STATISTICS=PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Tabel 16: associatie tussen fysieke belasting en geslacht (N=1149)

		Symmetric Measures	
		Value	Approximate Significance
Nominal by Nominal	Phi	,096	,005
	Cramer's V	,096	,005
N of Valid Cases		1149	

Deze associatie is niet opvallend en licht ik niet nader toe.

```
CROSSTABS
  /TABLES=cw20m416 BY geslacht
  /FORMAT=AVALUE TABLES
  /STATISTICS=PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Tabel 17: associatie tussen fysieke belasting en partnerstatus (N=1149)

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,040	,393
	Cramer's V	,040	,393
N of Valid Cases		1149	

Deze associatie is niet opvallend en licht ik niet nader toe.

```
CROSSTABS
  /TABLES=cw20m416 BY partner
  /FORMAT=AVALUE TABLES
  /STATISTICS=PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Tabel 18: associatie tussen fysieke belasting en ouderschap (N=1149)

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,043	,351
	Cramer's V	,043	,351
N of Valid Cases		1149	

Deze associatie is niet opvallend en licht ik niet nader toe.

```
CROSSTABS
  /TABLES=cw20m416 BY cw20m436
  /FORMAT=AVALUE TABLES
  /STATISTICS=PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Tabel 18: associatie tussen trainingsmogelijkheden en geslacht (N=1149)

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,053	,365
	Cramer's V	,053	,365
N of Valid Cases		1149	

Deze associatie is niet opvallend en licht ik niet nader toe.

```
CROSSTABS
  /TABLES=cw20m430 BY geslacht
```



```

/FORMAT=AVALUE TABLES
/STATISTICS=PHI
/CELLS=COUNT
/COUNT ROUND CELL.

```

Tabel 19: associatie tussen trainingsmogelijkheden en partnerstatus (N=1149)

		Symmetric Measures	
		Value	Approximate Significance
Nominal by Nominal	Phi	,041	,588
	Cramer's V	,041	,588
N of Valid Cases		1149	

Deze associatie is niet opvallend en licht ik niet nader toe.

```

CROSSTABS
/TABLES=cw20m430 BY partner
/FORMAT=AVALUE TABLES
/STATISTICS=PHI
/CELLS=COUNT
/COUNT ROUND CELL.

```

Tabel 20: associatie tussen trainingsmogelijkheden en ouderschap (N=1149)

		Symmetric Measures	
		Value	Approximate Significance
Nominal by Nominal	Phi	,075	,092
	Cramer's V	,075	,092
N of Valid Cases		1149	

Deze associatie is niet opvallend en licht ik niet nader toe.

```

CROSSTABS
/TABLES=cw20m430 BY cw20m436
/FORMAT=AVALUE TABLES
/STATISTICS=PHI
/CELLS=COUNT
/COUNT ROUND CELL.

```

Tabel 21: associatie tussen beroepssector en geslacht (N=1149)

		Symmetric Measures	
		Value	Approximate Significance
Nominal by Nominal	Phi	,446	,000
	Cramer's V	,446	,000
N of Valid Cases		1149	

```
CROSSTABS
  /TABLES=cw20m402 BY geslacht
  /FORMAT=AVALUE TABLES
  /STATISTICS=PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Tabel 22: associatie tussen beroepssector en partnerstatus (N=1149)

		Symmetric Measures	
		Value	Approximate Significance
Nominal by Nominal	Phi	,140	,049
	Cramer's V	,140	,049
N of Valid Cases		1149	

```
CROSSTABS
  /TABLES=cw20m402 BY partner
  /FORMAT=AVALUE TABLES
  /STATISTICS=PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Tabel 23: associatie tussen beroepssector en partnerstatus (N=1149)

		Symmetric Measures	
		Value	Approximate Significance
Nominal by Nominal	Phi	,106	,453
	Cramer's V	,106	,453
N of Valid Cases		1149	

```
CROSSTABS
  /TABLES=cw20m402 BY cw20m436
  /FORMAT=AVALUE TABLES
  /STATISTICS=PHI
  /CELLS=COUNT
```

/COUNT ROUND CELL.

3. Associaties tussen de categorische variabelen met meer dan twee antwoordcategorieën
Tabel 24: associatie tussen opleidingsniveau en fysieke belasting (N=1149)

		Value	Approximate Significance
Nominal by Nominal	Phi	,384	,000
	Cramer's V	,272	,000
N of Valid Cases		1149	

Op basis van de Cramers'v zijn er een significante associatie tussen de groepen van opleidingsniveau en fysieke belasting. Deze twee variabelen hangen met elkaar samen. De maat zegt niets over de richting van het verband.

```
CROSSTABS
  /TABLES=oplcat BY cw20m416
  /FORMAT=AVALUE TABLES
  /STATISTICS=PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Tabel 25: associatie tussen opleidingsniveau en trainingsmogelijkheden (N=1149)

		Value	Approximate Significance
Nominal by Nominal	Phi	,171	,004
	Cramer's V	,099	,004
N of Valid Cases		1149	

Op basis van de Cramers'v zijn er een significante associatie tussen de groepen van opleidingsniveau en de mogelijkheid deel te nemen aan trainingsactiviteiten. Deze twee variabelen hangen met elkaar samen. De maat zegt niets over de richting van het verband.

```
CROSSTABS
  /TABLES=oplcat BY cw20m416
  /FORMAT=AVALUE TABLES
  /STATISTICS=PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Tabel 26: associatie tussen opleidingsniveau en beroepssector(N=1149)

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,457	,000
	Cramer's V	,204	,000
N of Valid Cases		1149	

Op basis van de Cramers'v zijn er een significante associatie tussen de groepen van opleidingsniveau en beroepssector. Deze twee variabelen hangen met elkaar samen. De maat zegt niets over de richting van het verband

Tabel 27: associatie tussen fysieke belasting en trainingsmogelijkheden(N=1149)

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,158	,000
	Cramer's V	,112	,000
N of Valid Cases		1149	

```
CROSSTABS
  /TABLES=cw20m416 BY cw20m430
  /FORMAT=AVALUE TABLES
  /STATISTICS=PHI
  /CELLS=COUNT
  /COUNT ROUND CELL
```

Tabel 28: associatie tussen fysieke belasting en beroepssector(N=1149)

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,410	,000
	Cramer's V	,290	,000
N of Valid Cases		1149	

```
CROSSTABS
  /TABLES=cw20m416 BY cw20m402
  /FORMAT=AVALUE TABLES
  /STATISTICS=PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

Tabel 30: associatie tussen trainingsmogelijkheden en beroepssector(N=1149)

		Symmetric Measures	
		Value	Approximate Significance
Nominal by Nominal	Phi	,302	,000
	Cramer's V	,174	,000
N of Valid Cases		1149	

```
CROSSTABS
  /TABLES=cw20m430 BY cw20m402
  /FORMAT=AVALUE TABLES
  /STATISTICS=PHI
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

4. Associaties tussen de continue en categorische variabelen met meer dan twee antwoord categorieën (R- waarde + F-toets voor verschil in gemiddelde score.

Tabel 31: associatie tussen discrepantie en opleidingsniveau (N=1149)

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,043 ^a	,002	,001	3,98977

a. Predictors: (Constant), oplcat Level of education in CBS (Statistics Netherlands) categories

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT Discrepantie
  /METHOD=ENTER oplcat.
```

Tabel 32: F-toets voor verschil in gemiddelden op de discrepantie tussen opleidingsniveaus (N=1149)

ANOVA					
Discrepantie					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	84,006	5	16,801	1,055	,384
Within Groups	18208,601	1143	15,931		
Total	18292,607	1148			

```
ONEWAY Discrepantie BY oplcat
  /MISSING ANALYSIS.
```

Er is geen significant verschil in de gemiddelde discrepantie voor de verschillende opleidingsniveaus.

Tabel 33: associatie tussen discrepantie en fysieke belasting (N=1149)

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,100 ^a	,010	,009	3,97340

a. Predictors: (Constant), cw20m416 [Is/was] your work physically demanding?

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT Discrepantie
  /METHOD=ENTER cw20m416
```

Tabel 34: F-toets voor verschil in gemiddelden op de discrepantie tussen ervaren fysieke belasting (N=1149)

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4,658	,415		11,218	,000
	cw20m416 [Is/was] your work physically demanding?	-,566	,166	-,100	-3,412	,001

a. Dependent Variable: Discrepantie
 ONEWAY Discrepantie BY cw20m416
 /MISSING ANALYSIS.

Er is een significant verschil in de gemiddelde discrepantie voor de verschillende ervaren fysieke belasting. Respondenten die hun werk als fysiek belastend ervaren scoren dus significant anders op de discrepantie dan respondenten die dit niet zo ervaren.

Tabel 35: associatie tussen discrepantie en trainingsmogelijkheden (N=1149)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,041 ^a	,002	,001	3,99019

a. Predictors: (Constant), cw20m430 I [have/had] the opportunity to

learn new skills.

REGRESSION

```

/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT Discrepantie
/METHOD=ENTER cw20m430

```

Tabel 36: F-toets voor verschil in gemiddelden op de discrepantie tussen trainingsmogelijkheden (N=1149)

ANOVA

Discrepantie

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	85,773	3	28,591	1,798	,146
Within Groups	18206,834	1145	15,901		
Total	18292,607	1148			

ONEWAY Discrepantie BY cw20m430

/MISSING ANALYSIS.

Er is geen significant verschil in de gemiddelde discrepantie voor de mate waarin wordt ervaren deel te kunnen nemen aan trainingsmogelijkheden

Tabel 38: associatie tussen discrepantie en beroepssector (N=1149)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,073 ^a	,005	,004	3,98297

a. Predictors: (Constant), cw20m402 [In what sector do you work?/In

what sector did you work in your last job?]

REGRESSION

```

/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT Discrepantie
/METHOD=ENTER cw20m402

```

Tabel 39.: F-toets voor verschil in gemiddelden op de discrepantie tussen trainingsmogelijkheden (N=1149)

ANOVA

Discrepantie	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	228,175	13	17,552	1,103	,352
Within Groups	18064,432	1135	15,916		
Total	18292,607	1148			

ONEWAY Discrepantie BY cw20m402
/MISSING ANALYSIS.

Er is geen significant verschil in de gemiddelde discrepantie voor de verschillende beroepssectoren

NB: ik heb geen f-toetsen voor verschil in gemiddelde gedaan op de ondersteunende variabelen: verwachte en gewenste pensioenleeftijd

Tabel 39: associatie tussen gewenste pensioenleeftijd en opleidingsniveau (N=1149)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,073 ^a	,005	,004	4,191

a. Predictors: (Constant), oplcat Level of education in CBS (Statistics

Netherlands) categories

REGRESSION

/MISSING LISTWISE

/STATISTICS COEFF OUTS R ANOVA

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

/NOORIGIN

/DEPENDENT cw20m288

/METHOD=ENTER oplcat

Tabel 40: associatie tussen gewenste pensioenleeftijd en fysieke belasting (N=1149)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,100 ^a	,010	,009	4,181

a. Predictors: (Constant), cw20m416 [Is/was] your work physically demanding?

REGRESSION

```

/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT cw20m288
/METHOD=ENTER cw20m416

```

Tabel 41: associatie tussen gewenste pensioenleeftijd en trainingsmogelijkheden (N=1149)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,046 ^a	,002	,001	4,198

a. Predictors: (Constant), cw20m430 I [have/had] the opportunity to learn new skills.

REGRESSION

```

/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT cw20m288
/METHOD=ENTER cw20m430

```

Tabel 42: associatie tussen gewenste pensioenleeftijd en beroepssector (N=1149)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,066 ^a	,004	,003	4,193

a. Predictors: (Constant), cw20m402 [In what sector do you work?/In what sector did you work in your last job?]

REGRESSION

```

/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT cw20m288
/METHOD=ENTER cw20m402

```

Tabel 43: associatie tussen verwachte pensioenleeftijd en opleidingsniveau (N=1149)

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,038 ^a	,001	,001	3,444

a. Predictors: (Constant), oplcat Level of education in CBS (Statistics Netherlands) categories

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT cw20m289
  /METHOD=ENTER oplcat
```

Tabel 44.: associatie tussen verwachte pensioenleeftijd en fysieke belasting (N=1149)

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,006 ^a	,000	-,000	3,447

a. Predictors: (Constant), cw20m416 [Is/was] your work physically demanding?

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT cw20m289
  /METHOD=ENTER cw20m41
```

Tabel 45: associatie tussen verwachte pensioenleeftijd en trainingsmogelijkheden (N=1149)

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,009 ^a	,000	-,001	3,447

a. Predictors: (Constant), cw20m430 I [have/had] the opportunity to learn new skills.

b.

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

/DEPENDENT cw20m289
/METHOD=ENTER cw20m430

Tabel 46: associatie tussen verwachte pensioenleeftijd en beroepssector (N=1149)

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,004 ^a	,000	-,001	3,447

a. Predictors: (Constant), cw20m402 [In what sector do you work?/In what sector did you work in your last job?]

```
REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT cw20m289
/METHOD=ENTER cw20m402
```

Modevaluatie met outliers

Tabel 47: modelstatistieken van de hiërarchische regressieanalyse met outliers (n=1149)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics			
						F Change	df1	df2	Sig. F Change
1	,283 ^a	,080	,075	3,83989	,080	14,231	7	1141	,000
2	,284 ^b	,081	,074	3,84039	,001	,706	1	1140	,401
3	,288 ^c	,083	,075	3,83987	,002	1,154	2	1138	,316
4	,302 ^d	,091	,082	3,82416	,008	10,373	1	1137	,001
5	,303 ^e	,092	,083	3,82358	,001	1,345	1	1136	,246

```
REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA COLLIN TOL CHANGE
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT Discrepantie
/METHOD=ENTER leeftijd D_man D_Partner D_met_kinderen Zakelijke_dienstverlenin
Publieke_dienstverlenin Zorg_Welzijn
/METHOD=ENTER Inkomen_dzd_euros
/METHOD=ENTER D_opleidingsniveau_middelbaar D_opleidingsniveau_hoog
/METHOD=ENTER FysiekeBelasting
/METHOD=ENTER cw20m430
```

Het model dat is geëvalueerd in de resultatenparagraaf is het model zonder de uitbijters. Ik zal nu bondig het model evalueren zonder de 6 uitbijters. In tabel 47 is te zien dat het eerste model 8 procent ($R^2=.080$) van de variantie in het verschil tussen gewenste- en verwachte pensioenleeftijd verklaard. Dit model bevat alleen de controlevariabelen: geslacht, leeftijd, beroepssector, partnerstatus en ouderschap.

Het model is een significant betere voorspeller van de discrepantie in pensioenleeftijd dan het lege model ($F_{change}(7,1141)=14.23, p<0.01$).

In het tweede model wordt het netto-maandelijks inkomen van de respondent toegevoegd en in het derde model het opleidingsniveau. Gecorrigeerd voor het aantal voorspellers, verklaren modellen niet meer variantie in het verschil tussen gewenste en verwachte pensioenleeftijd (beide; $R^2_{adjusted}: .075$). Daarentegen levert het vierde model inclusief fysieke belasting van het werk, wel een significant betere voorspelling van de discrepantie in pensioenleeftijd ($F_{change}(1, 1137) = 10.373, p < 0.02$). Dit model verklaard 0.8 procent ($R^2 = .082$) meer variantie in de discrepantie dan het derde model. In het laatste model wordt de mogelijkheid om trainingsactiviteiten te volgen toegevoegd. Dit vijfde en laatste model verklaart slechts 0.1% ($R^2 = .083$) meer variantie in de discrepantie dan het vierde model, maar in zijn geheel is het model wel een significante voorspeller van de variantie in het verschil tussen gewenste en verwachte pensioenleeftijd van respondenten ($F(12, 1136) = 9.602, p < 0.01$).

Samenvattend: het eerste model met alleen de controlevariabele een significante voorspeller van de variantie in de discrepantie. Dit resultaat ondersteunt de keuze voor deze controlevariabelen. Vervolgens neemt in elk geschat model de proportie verklaarde variantie toe. Indien gecorrigeerd voor het aantal voorspellers is hier louter sprake van in het vierde model. Van alle modellen is alleen het vierde model - waarin de ervaren fysieke belasting van het werk door de respondent wordt toegevoegd - een significant betere voorspeller dan het model daarvoor.

Modevaluatie zonder outliers

Tabel 48: : modelstatistieken van de hierarchische regressieanalyse met outliers (n=1143)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics			
						F Change	df1	df2	Sig. F Change
1	,304 ^a	,092	,087	3,11115	,092	16,522	7	1135	,000
2	,308 ^b	,095	,088	3,10876	,002	2,748	1	1134	,098
3	,311 ^c	,097	,089	3,10815	,002	1,221	2	1132	,295
4	,324 ^d	,105	,096	3,09506	,008	10,599	1	1131	,001
5	,334 ^e	,111	,102	3,08558	,006	7,959	1	1130	,005

```

REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA COLLIN TOL CHANGE
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT Discrepantie
  /METHOD=ENTER leeftijd D_man D_Partner D_met_kinderen Zakelijke_dienstverlenin
    Publieke_dienstverlenin Zorg_Welzijn
  /METHOD=ENTER Inkomen_dzd_euros
  /METHOD=ENTER D_opleidingsniveau_middelbaar D_opleidingsniveau_hoog
  /METHOD=ENTER FysiekeBelasting
  /METHOD=ENTER cw20m430

```

Aangezien in de resultatenparagraaf de modevaluatie van het model zonder uitbijters al is gegeven, zal ik nu alleen ingaan op het verschil met de modevaluatie met uitbijters (tabel 47). Ten opzichte van de analyse met uitbijters hebben alle modellen in de analyse zonder de uitbijters een hogere proportie verklaarde variantie. Dit ondersteunt de keuze om de uitbijters uit de data te halen. In het uiteindelijke model is de aangepaste proportie variantie zonder uitbijters bijna 2 procent meer. Het model zonder de uitbijters verklaart dus – helemaal gezien het geringe aantal voorspellers – aanzienlijk meer variantie in het verschil tussen de verwachte en gewenste pensioenleeftijd.

Hypothesetoetsing met outliers

Tabel 49: resultaten van de hiërarchische regressieanalyse met discrepantie in pensioenleeftijd als onafhankelijke, met outliers (n=1149)

Model		Unstandardized Coefficients		Standardized	t	Sig.
		B	Std. Error	Coefficients Beta		
1	(Constant)	13,211	1,044		12,658	,000
	leeftijd Age of the household member	-,176	,018	-,275	-9,546	,000
	D_man	,204	,246	,026	,832	,405
	D_Partner	-,459	,279	-,049	-1,644	,100
	D_met_kinderen	,294	,284	,031	1,035	,301
	Zakelijke_dienstverlenin	-,049	,132	-,011	-,374	,708
	Publieke_dienstverlenin	-,273	,294	-,029	-,927	,354
	Zorg_Welzijn	-,259	,326	-,027	-,793	,428
2	(Constant)	13,360	1,059		12,617	,000
	leeftijd Age of the household member	-,176	,018	-,275	-9,541	,000
	D_man	,285	,264	,036	1,080	,280
	D_Partner	-,473	,280	-,050	-1,693	,091
	D_met_kinderen	,319	,285	,033	1,117	,264
	Zakelijke_dienstverlenin	-,051	,132	-,012	-,383	,702
	Publieke_dienstverlenin	-,255	,295	-,028	-,866	,387
	Zorg_Welzijn	-,265	,327	-,027	-,811	,417
3	Inkomen_dzd_euros	-,091	,108	-,026	-,840	,401
	(Constant)	13,608	1,087		12,521	,000
	leeftijd Age of the household member	-,177	,018	-,277	-9,594	,000
	D_man	,255	,265	,032	,964	,335
	D_Partner	-,450	,280	-,048	-1,608	,108
	D_met_kinderen	,327	,286	,034	1,145	,253
	Zakelijke_dienstverlenin	-,055	,132	-,013	-,416	,677
	Publieke_dienstverlenin	-,149	,305	-,016	-,490	,624
	Zorg_Welzijn	-,202	,329	-,021	-,612	,541
	Inkomen_dzd_euros	-,024	,119	-,007	-,205	,838
D_opleidingsniveau_middelbaar	-,348	,330	-,042	-1,055	,292	
	D_opleidingsniveau_hoog	-,535	,352	-,067	-1,518	,129

4	(Constant)	12,361	1,150		10,753	,000
	leeftijd Age of the household member	-,176	,018	-,275	-9,568	,000
	D_man	,224	,264	,028	,849	,396
	D_Partner	-,444	,279	-,047	-1,593	,111
	D_met_kinderen	,322	,284	,034	1,134	,257
	Zakelijke_dienstverlenin	-,067	,131	-,015	-,510	,610
	Publieke_dienstverlenin	-,101	,304	-,011	-,334	,739
	Zorg_Welzijn	-,368	,332	-,038	-1,107	,268
	Inkomen_dzd_euros	,027	,120	,008	,224	,823
	D_opleidingsniveau_middelbaar	-,188	,332	-,023	-,567	,571
	D_opleidingsniveau_hoog	-,222	,364	-,028	-,609	,543
	FysiekeBelasting	,569	,177	,101	3,221	,001
5	(Constant)	12,980	1,267		10,243	,000
	leeftijd Age of the household member	-,177	,018	-,276	-9,607	,000
	D_man	,229	,264	,029	,866	,386
	D_Partner	-,437	,279	-,047	-1,568	,117
	D_met_kinderen	,335	,285	,035	1,178	,239
	Zakelijke_dienstverlenin	-,074	,131	-,017	-,562	,574
	Publieke_dienstverlenin	-,071	,305	-,008	-,232	,817
	Zorg_Welzijn	-,341	,333	-,035	-1,024	,306
	Inkomen_dzd_euros	,040	,121	,011	,331	,741
	D_opleidingsniveau_middelbaar	-,202	,332	-,024	-,609	,543
	D_opleidingsniveau_hoog	-,226	,364	-,028	-,621	,535
	FysiekeBelasting	,550	,177	,097	3,097	,002
	cw20m430 I [have/had] the opportunity to learn new skills.	-,208	,179	-,034	-1,160	,246

a. Dependent Variable: Discrepantie

```

REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA CHANGE
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT Discrepantie
  /METHOD=ENTER leeftijd D_man D_Partner D_met_kinderen
  Zakelijke_dienstverlenin
  Publieke_dienstverlenin Zorg_Welzijn
  /METHOD=ENTER Inkomen_dzd_euros
  /METHOD=ENTER D_opleidingsniveau_middelbaar D_opleidingsniveau_hoog
  /METHOD=ENTER FysiekeBelasting
  /METHOD=ENTER cw20m430.

```

In de resultatenparagraaf zijn de hypothesen van deze thesis getoetst op basis van de regressieanalyse zonder outliers. Ik zal nu bondig de hypothesen toetsen op basis van de als eerst verrichte analyse, die met de outliers.

De hypothese aangaande het maandelijks netto-inkomen luidt als volgt: (H1) naarmate een oudere werkende een hoger inkomen heeft zal de discrepantie in pensioenleeftijd kleiner zijn. Model 5 in tabel 5 laat zien dat de resultaten deze hypothese niet ondersteunen. Het maandelijks netto-inkomen

heeft namelijk een niet significant effect op de discrepantie in pensioenleeftijd ($b=.040$, $p=.741$). Dit effect is robuust: het is significant in geen enkel model. Het effect is telkens nihil in in de laatste twee modellen negatief in plaats van positief. Uit het laatste model volgt dat indien het inkomen met duizend euro stijgt zal het verschil tussen de verwachte en gewenste pensioenleeftijd met 15 dagen toenemen. . Naarmate het inkomen van een oudere werkende in de data stijgt, zal zijn discrepantie in pensioenleeftijd dus niet kleiner worden, maar groter. Op basis van de resultaten moet hypothese 1 worden verworpen.

De tweede hypothese luidt als volgt (H2): *naarmate het opleidingsniveau van een oudere werkende hoger ligt, zal de discrepantie in pensioenleeftijd kleiner zijn*. Om deze hypothese accuraat te kunnen toetsen is de referentiegroep voor respondenten met een laag opleidingsniveau. Zo kan het verschil tussen deze groep en de twee groepen met hogere opleidingsniveaus - het middelbare en hoge opleidingsniveau - worden geschat.

Model 5 in tabel 5 laat zien dat wel zowel middelbaaropgeleiden ($b=-.202$, $p=.543$) als hoogopgeleiden ($b=-.226$, $p=.535$) een niet significant verschillen in de discrepantie in pensioenleeftijd ten opzichte van laagopgeleiden. Voor oudere werkenden met een middelbaar opleidingsniveau is het verschil tussen hun verwachte en gewenste pensioenleeftijd 73 dagen kleiner dan voor oudere werkenden met een laag opleidingsniveau. Voor hoogopgeleide oudere werkenden is het verschil tussen hun verwachte en gewenste pensioenleeftijd 82 dagen kleiner dan voor de groep laagopgeleiden. Ondanks dat de resultaten meegaan in de richting van de hypothese, zijn deze in geen enkel model significant. Op basis van de resultaten moet de tweede hypothese worden verworpen.

De derde hypothese luidt (H3): *naarmate een oudere werkende het werk als fysiek meer belastend wordt ervaren, zal het verschil tussen gewenste en verwachte pensioenleeftijd groter zijn*. Model 5 in tabel 49 laat zien dat het ervaren van werk als fysiek belastend een significant positieve invloed ($b = .550$, $p<0.01$) heeft. Zodra oudere werkenden hun werk als fysiek meer belastend ervaren, neemt het verschil tussen hun verwachte en gewenste pensioenleeftijd toe. Dit effect is behoorlijk robuust: in alle modellen met fysieke belasting is het effect significant en positief. Het effect van fysieke belasting gaat dus mee in de richting van de hypothese en is robuust significant. Zodoende zijn de resultaten in overeenstemming met de hypothese.

De vierde en laatste hypothese luidt (H4): *Naarmate er vaker mogelijkheden bestaan om trainingsactiviteiten te volgen, zal de discrepantie tussen gewenste en verwachte pensioenleeftijd kleiner zijn*. De resultaten ondersteunen deze hypothese niet. Het aanbieden van trainingsmogelijkheden heeft een niet-significant negatieve invloed op de discrepantie in pensioenleeftijd ($b = .208$, $p=.246$). Naarmate een oudere werkende er in toenemende mate mee instemt de mogelijkheid te hebben gekregen om deel te nemen aan trainingsactiviteiten, neemt het verschil tussen de verwachte en gewenste pensioenleeftijd niet af. Het effect van fysieke belasting gaat mee wel mee in de richting van de hypothese, maar is niet significant. Zodoende worden de hypothese verworpen.

In de analyse werd het model voor een vijftal variabelen gecontroleerd. Hiervan heeft de leeftijd van een respondent een significant negatief effect op de discrepantie tussen de verwachte en gewenste pensioenleeftijd ($b = -.177, p < 0.01$). Het effect van leeftijd is robuust: het is significant in alle modellen. Indien een respondent één jaar ouder wordt, zal de discrepantie tussen de verwachte en gewenste pensioenleeftijd 56 dagen kleiner worden. Er zijn geen hypothesen opgesteld, maar het effect van leeftijd lag in de lijn der verwachting. Respondenten kunnen niet met terugwerkende kracht pensioneren, zodoende zal hun gewenste pensioenleeftijd waarschijnlijk meestijgen met hun leeftijd. De effecten van de andere controlevariabelen geslacht ($b = 0.88, p = 0.680$), partnerstatus ($b = -.537, p = .017$), ouderschap ($b = .298, p = .198$) en de beroepssectoren: zakelijke dienstverlening ($b = -.076, p = 0.485$), publieke dienstverlening ($b = .081, p = .741$) en zorg en welzijn ($b = -.426, p = .114$) zijn niet significant. De effecten van deze controlevariabelen zijn niet-significant in alle modellen. De mate waarin respondenten verwachten dat hun pensioenwensen uitkomen verschilt niet tussen mannen en vrouwen, mensen met of zonder partner, met of zonder kinderen en de sector waarin een respondent werkzaam is.

Hypothesetoetsing zonder outliers

Tabel 49: resultaten van de hiërarchische regressieanalyse met discrepantie in pensioenleeftijd als onafhankelijke, zonder outliers ($n = 1143$)

Model		Unstandardized Coefficients		Standardized	t	Sig.
		B	Std. Error	Coefficients Beta		
1	(Constant)	11,914	,848		14,050	,000
	leeftijd Age of the household member	-,151	,015	-,290	-10,101	,000
	D_man	,016	,199	,003	,082	,935
	D_Partner	-,550	,227	-,072	-2,427	,015
	D_met_kinderen	,237	,231	,030	1,026	,305
	Zakelijke_dienstverlenin	-,053	,109	-,015	-,486	,627
	Publieke_dienstverlenin	-,148	,239	-,020	-,621	,535
	Zorg_Welzijn	-,393	,266	-,050	-1,480	,139
2	(Constant)	12,160	,860		14,136	,000
	leeftijd Age of the household member	-,151	,015	-,289	-10,099	,000
	D_man	,144	,214	,022	,676	,499
	D_Partner	-,569	,227	-,074	-2,508	,012
	D_met_kinderen	,274	,232	,035	1,182	,237
	Zakelijke_dienstverlenin	-,055	,109	-,015	-,507	,612
	Publieke_dienstverlenin	-,121	,239	-,016	-,505	,613
	Zorg_Welzijn	-,409	,266	-,052	-1,541	,124
Inkomen_dzd_euros	-,149	,090	-,051	-1,658	,098	
3	(Constant)	12,269	,883		13,901	,000
	leeftijd Age of the household member	-,152	,015	-,291	-10,122	,000
	D_man	,111	,215	,017	,517	,605
	D_Partner	-,553	,227	-,072	-2,436	,015
	D_met_kinderen	,278	,232	,036	1,196	,232
	Zakelijke_dienstverlenin	-,058	,109	-,016	-,534	,593

	Publieke_dienstverlenin	-,024	,247	-,003	-,097	,923
	Zorg_Welzijn	-,353	,268	-,045	-1,319	,188
	Inkomen_dzd_euros	-,083	,099	-,029	-,831	,406
	D_opleidingsniveau_middelbaar	-,170	,268	-,025	-,635	,525
	D_opleidingsniveau_hoog	-,425	,287	-,065	-1,482	,139
4	(Constant)	11,237	,934		12,028	,000
	leeftijd Age of the household member	-,151	,015	-,289	-10,113	,000
	D_man	,081	,214	,012	,376	,707
	D_Partner	-,550	,226	-,072	-2,430	,015
	D_met_kinderen	,274	,231	,035	1,184	,237
	Zakelijke_dienstverlenin	-,064	,109	-,018	-,586	,558
	Publieke_dienstverlenin	,019	,246	,003	,079	,937
	Zorg_Welzijn	-,479	,270	-,061	-1,778	,076
	Inkomen_dzd_euros	-,032	,100	-,011	-,321	,748
	D_opleidingsniveau_middelbaar	-,038	,270	-,006	-,141	,888
	D_opleidingsniveau_hoog	-,174	,296	-,027	-,589	,556
	FysiekeBelasting	,469	,144	,102	3,256	,001
5	(Constant)	12,447	1,025		12,139	,000
	leeftijd Age of the household member	-,153	,015	-,292	-10,236	,000
	D_man	,088	,214	,014	,412	,680
	D_Partner	-,537	,225	-,070	-2,382	,017
	D_met_kinderen	,298	,230	,038	1,292	,196
	Zakelijke_dienstverlenin	-,076	,109	-,021	-,699	,485
	Publieke_dienstverlenin	,081	,247	,011	,330	,741
	Zorg_Welzijn	-,426	,269	-,054	-1,583	,114
	Inkomen_dzd_euros	-,005	,100	-,002	-,054	,957
	D_opleidingsniveau_middelbaar	-,066	,269	-,010	-,246	,805
	D_opleidingsniveau_hoog	-,185	,295	-,028	-,626	,531
	FysiekeBelasting	,431	,144	,093	2,987	,003
	cw20m430 I [have/had] the opportunity to learn new skills.	-,409	,145	-,081	-2,821	,004

a. Dependent Variable: Discrepantie

Uit tabel 50 volgen de resultaten van de hiërarchische regressieanalyse zonder outliers. Aangezien de hypothesetoetsing in de resultatenparagraaf aan de hand van deze analyse is verricht. Zal ik in deze sectie alleen de opvallende verschillen tussen de modellen met en zonder outliers. In het uiteindelijke model zonder outliers is het effect van de mogelijkheid deel te nemen aan trainingsactiviteiten significant geworden. Naarmate een oudere werkende er in toenemende mate mee instemt de mogelijkheid te hebben gekregen om deel te nemen aan trainingsactiviteiten, neemt wordt verschil tussen de verwachte en gewenste pensioenleeftijd kleiner wanneer er geen uitbijters in het model zijn. De robuustheid van de effecten van alle andere variabelen blijft hetzelfde in de regressie met of zonder uitbijters. De effecten van leeftijd en het ervaren van het werk als fysiek belastend zijn significant in zowel de analyse met als zonder uitbijters en alle andere effecten zijn niet-significant.

Bijlage 3

Deze derde bijlage bevat de uitgebreide bespreking van de multicollineariteit (1), outliers (2) en assumptietoetsing (3). Of er sprake is van multicollineariteit evenals mogelijke schending van assumpties zal worden getoetst voor het hiërarchisch geschatte model zonder outliers.

1. Multicollineariteit

Tabel 1: Resultaten van het uiteindelijke model van de hiërarchische regressieanalyse met discrepantie in pensioenleeftijd als onafhankelijke variabele, zonder outliers met VIF-Scores. (N=1143)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	12,447	1,025		12,139	,000		
	leeftijd Age of the household member	-,153	,015	-,292	-10,236	,000	,967	1,034
	D_man	,088	,214	,014	,412	,680	,732	1,365
	D_Partner	-,537	,225	-,070	-2,382	,017	,908	1,101
	D_met_kinderen	,298	,230	,038	1,292	,196	,905	1,105
	Zakelijke_dienstverlenin	-,076	,109	-,021	-,699	,485	,870	1,150
	Publieke_dienstverlenin	,081	,247	,011	,330	,741	,739	1,354
	Zorg_Welzijn	-,426	,269	-,054	-1,583	,114	,678	1,476
	Inkomen_dzd_euros	-,005	,100	-,002	-,054	,957	,654	1,529
	D_opleidingsniveau_middelbaar	-,066	,269	-,010	-,246	,805	,494	2,022
	D_opleidingsniveau_hoog	-,185	,295	-,028	-,626	,531	,387	2,582
	FysiekeBelasting	,431	,144	,093	2,987	,003	,805	1,242
	cw20m430 I [have/had] the opportunity to learn new skills.	-,409	,145	-,081	-2,821	,004	,948	1,055

a. Dependent Variable: Discrepantie

Tabel 1 laat de geschatte regressiecoëfficiënten van het uiteindelijke model zonder outliers zien. Om te controleren of er sprake van multicollineariteit staan worden in de tabel VIF (*Variance Inflation Factor*)-scores gegeven. Indien er sprake is van multicollineariteit voorspellen de voorspellers elkaar en wordt er geen extra variantie verklaard in de discrepantie in pensioenleeftijd. Dit is onwenselijk. Een vuistregel is dat er vanaf een VIF van 4 een probleem is voor het schatten van de regressiecoëfficiënt van de betreffende voorspeller. Tabel 1 laat zien dat er geen problematische VIF-scores zijn in het uiteindelijke model zonder outliers. De voorspellers met de hoogste score zijn het middelbare (VIF= 2.020) - en hoge opleidingsniveau (VIF=2.576). Dat deze twee voorspellers een hogere score hebben volgt uit het feit dat het dummy variabelen zijn. Zodoende wordt de relatief hoge VIF-score wordt

verklaard door de overlap in het opleidingssysteem: iemand met een middelbare opleiding beschikt ook over de kennis van iemand met een lager opleidingsniveau.

2. Outliers

Al in de analysebeschrijving beschreef ik dat er na de eerste analyse uitbijters waren gevonden. In totaal waren dit er 6. Deze uitbijters hebben een grote negatieve invloed op de kwaliteit van het geschatte model, daarom heb ik deze respondenten uit de data verwijderd. Om te beoordelen of een respondent een uitbijters is heb ik de scores van respondenten op een aantal statistische maten voor uitbijters vergeleken. Er zijn dus geen respondenten uit de data verwijderd, omdat zij een hoge score hadden op slechts één statistische maat.

Ten eerste heb ik gekeken naar de Cook's Distance. Dit is het product van de studentized residual en de leverage. Een uitbijter op de CD is hierdoor een uitbijter in zowel de x- richting: de geschatte hellingen als de y richting. Een CD hoger dan 1 is echt problematisch. De vuistregel is dat een CD groter dan 4 gedeeld door steekproefpopulatie een mogelijk probleem vormt. Voor deze thesis is die kritieke waarde: $4/1149 = 0.0035$. Er waren in totaal 39 respondenten met een score hoger dan deze kritieke waarde.

Ten tweede heb ik gekeken naar de DFFit. Deze score geeft de verandering in de gestandaardiseerde voorspelde y. De vuistregel is dat een DFFit-score invloedrijk is boven de 3 keer de wortel van het aantal parameters in het model gedeeld door het aantal respondenten in de steekproef. In deze thesis is die kritieke waarde 0.280. Er waren in totaal 6 respondenten met een score hoger dan deze waarde.

Tot slot heb ik gekeken naar de DFBeta. Deze score geeft de verandering in het gestandaardiseerde regressiecoëfficiënt van een indicator. De kritieke waarde bedroeg 0.059. In totaal waren op 4 indicatoren in het model invloedrijke DFBeta scores van in totaal 10 respondenten.

De respondenten die uiteindelijk uit de data zijn verwijderd hadden een score boven de kritieke waarde op de Cooks' Distance, de DFFIT en tenminste één DFBeta. In Bijlage 2; modevaluatie heb ik al toegelicht dat het model zonder deze respondenten ($R^2 = .102$) ook daadwerkelijk meer variantie kon verklaren in het verschil tussen de gewenste en verwachte pensioenleeftijd dan het model met uitbijters ($R^2 = .083$).

Tabel 2: scores van de uitbijters op discrepantie, geslacht, leeftijd, inkomen, fysieke belasting, inkomen, opleidingsniveau en trainingsmogelijkheden

	N	Minimum	Maximum	Mean	Std. Deviation
Discrepantie	6	10,00	50,00	31,6667	16,32993
geslacht Gender	6	1	2	1,33	,516
leeftijd Age of the household member	6	47	55	49,33	2,875
Inkomen_dzd_euros	6	,25	10,00	3,1515	3,46206
FysiekeBelasting	6	1,00	3,00	2,1667	,98319
oplcacat_groepen	6	1,00	3,00	2,3333	,81650
cw20m430 I [have/had] the opportunity to learn new skills.	6	2	4	3,17	,753

In tabel 2 worden de scores van deze uitbijters op een aantal variabelen gegeven. Kijkend naar de scores van deze uitbijters blijkt dat zij een stuk jonger zijn dan het steekproefgemiddelde en een enorm groot verschil hebben tussen hun gewenste en verwachte pensioenleeftijd. Het inkomen van deze uitbijters lijkt een stuk hoger te liggen, dit is niet zo. Een respondent met een netto-maandelijks inkomen van 10.000 trekt zwaar aan dit gemiddelde. Enerzijds zou het voor vervolgonderzoek interessant zijn om naar deze uitbijters te kijken: hoe kan het dat zij zo'n enorm grote discrepantie hebben? Anderzijds is het aantal respondenten met een dermate hoge verschilscore zo klein dat deze scores neigen naar ruis.

3. Assumptietoetsing

Het doel van de statistische analyse is het kunnen doen van generaliserende uitspraken over de onderzoekspopulatie. Om deze uitspraken te doen moet de data die wordt gebruikt in de analyse voldoen aan een viertal assumpties. In de deelsectie die volgt heb ik besproken of aan deze assumpties wordt voldaan. Mochten assumpties worden geschonden heeft dit gevolgen voor de conclusies die kunnen worden getrokken op basis van de statistische analyse.

Onafhankelijkheid

De eerste assumptie waaraan moet worden voldaan is die van onafhankelijke waarnemingen. Alle respondenten in het onderzoek moeten onafhankelijk van elkaar zijn in het geven van antwoorden. Er mag geen samenhang zijn tussen respondenten, bijvoorbeeld als gevolg van huishoudens met meerdere benaderde personen. Om onafhankelijkheid binnen onderzoek te garanderen moeten respondenten aselekt worden geselecteerd. Met andere woorden: iedereen in de onderzoekspopulatie moet dezelfde kans hebben om deel te kunnen nemen aan het onderzoek.

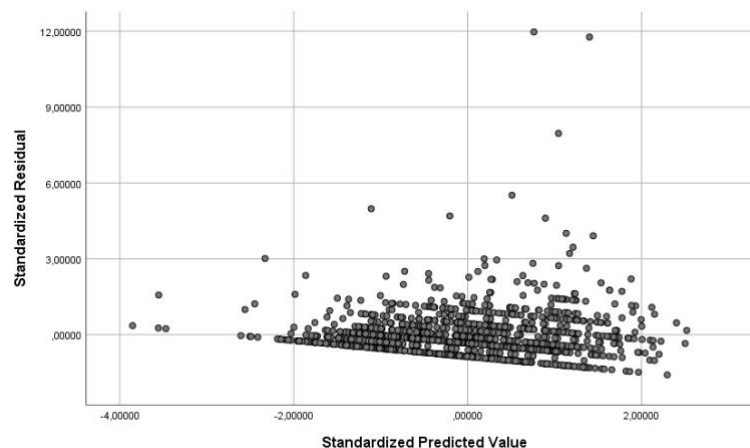
De respondenten in dit onderzoek zijn aselekt geselecteerd op basis van een probabilistische steekproef van huishoudens uit het bevolkingsregister van het Centraal Bureau voor de Statistiek. Ondanks dat de kans bestaat dat respondenten een huisgenoot heeft die deelneemt aan het panel, is er

geen aanwijzing dat deze huisgenoten dezelfde vragenlijst invullen. Zodoende wordt aan de assumptie van onafhankelijke waarnemingen voldaan.

Lineariteit

De tweede assumptie waaraan moet worden voldoen is die van lineariteit: er moet een lineair verband zijn tussen de onafhankelijke variabele en de voorspellers. Indien deze assumptie wordt geschonden zal er sprake zijn van misspecificatie: door een slechte fit van het model zullen uiteindelijk inaccurate conclusies worden getrokken. Om aan deze assumptie te voldoen moet het gemiddelde van de residuen – of voorspelfouten- nul zijn voor elke set van waarden van de voorspellers. Deze assumptie heb ik gecontroleerd aan de hand van een zgn. residual plot van de steekproefresiduen. Indien in deze plot geen systematische afwijking van de ‘nullijn’ zichtbaar is, wordt aan de assumptie voldaan. In figuur 1 is te zien dat er geen sprake is van een systematische afwijking. Er is geen afwijkend patroon in de scores. Zodoende wordt aan de assumptie van lineariteit voldaan.

Figuur 1: Spreidingsdiagram gestandariseerde voorspellingen (x-as) en gestandaardiseerde residuen (y-as)



Homoscedasticiteit

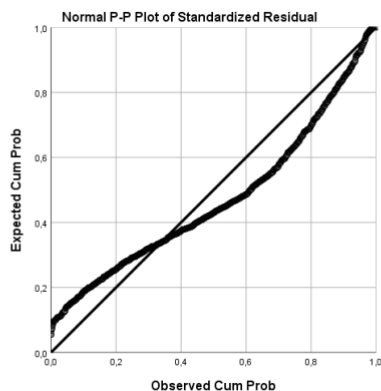
De derde assumptie waaraan moet worden voldaan is die van homoscedaciteit: er moet constante variantie zijn in de standaarddeviatie van de residuen. Voor elke set van warden van de voorspellers moet de standaarddeviatie van de residuen dus gelijk zijn. Indien deze assumptie wordt geschonden zullen regressiecoëfficiënten onzuiver worden geschat. Op basis van deze onzuivere schattingen worden dan inaccurate conclusies getrokken. Om deze assumptie te controleren heb ik wederom gebruik maken van de zgn. residual plot van de steekproefresiduen (figuur 1). Om aan de assumptie te voldoen mag op de plot geen systematische afwijking zijn in de spreiding rond nul. In figuur 1 is op de uitersten van de

as de spreiding wat groter, maar er zijn daar ook een stuk minder cases. Er is geen sprake van een systematische afwijking. Zodoende wordt aan de assumptie van homoscedasticiteit voldaan.

Normaliteit

De vierde assumptie voor een regressieanalyse is normaliteit: de residuen van afhankelijke variabele moeten normaal verdeeld zijn. In deze thesis moeten de residuen van de discrepantie in pensioenleeftijd dus normaal verdeeld zijn, gegeven de voorspellers. Mocht er niet aan de assumptie van normaliteit worden voldoen zullen de toetsen en betrouwbaarheidsintervallen van de voorspelers niet kloppen, waardoor er onterechte conclusies kunnen worden getrokken. Ik zal controleren of er aan de assumptie van normaliteit wordt voldaan aan de hand van een PP-plot en een histogram van de gestandaardiseerde residuen. Op de PP-plot mag de puntenwolk niet afwijken van de lineaire lijn en in het histogram moet vanzelfsprekend een normale verdeling zichtbaar zijn. In figuur 2 is duidelijk zichtbaar dat de puntenwolk afwijkt van de lijn. Zodoende wordt de assumptie van normaliteit geschonden.

Figuur 2: PP-plot van de gestand. Residuen van de discrepantie tussen verwachte en gewenste pensioenleeftijd.



Figuur 3: histogram van de gestand. Residuen van de discrepantie tussen verwachte en gewenste pensioenleeftijd.

