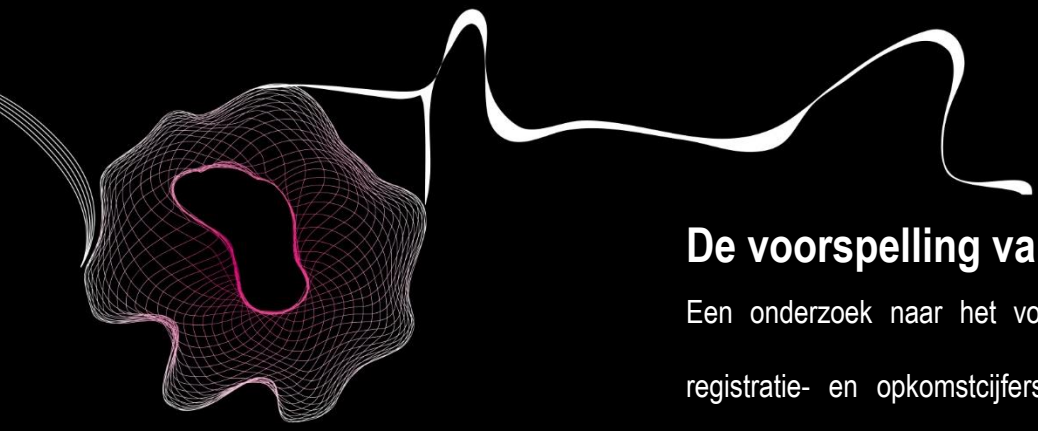
BACHELOROPDRACHT



De voorspelling van de Bachelor Open Dagen

Een onderzoek naar het vooraf inzichtelijk maken van de verwachte registratie- en opkomstcijfers van de Bachelor Open Dagen van de Universiteit Twente

STUDENT

T.J. Bemthuis (Thijs)

FACULTY OF BEHAVIOURAL, MANAGEMENT AND SOCIAL SCIENCES (BMS)

TECHNISCHE BEDRIJFSKUNDE

INTERNE BEGELEIDERS

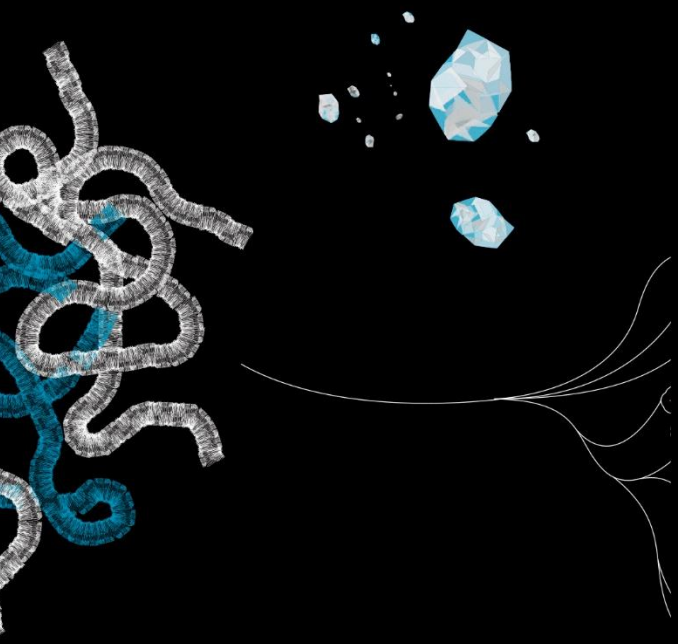
Dr. ir. M.R.K. Mes (Martijn)

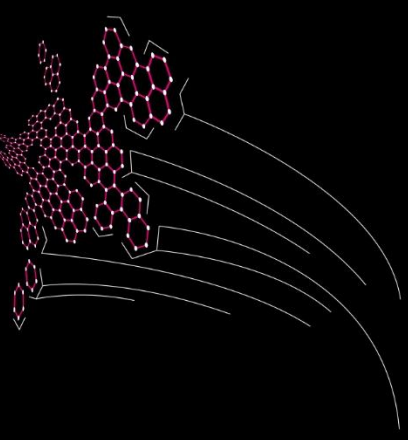
Dr. P.C. Schuur (Peter)

EXTERNE BEGELEIDERS

E.M. Overbeek (Marlies)

R.A. Schepers (Rob)





PUBLICATIE DATUM

02-10-2016

UNIVERSITY OF TWENTE.

"Prediction is very difficult, especially if it's about the future."

Nils Bohr (1885 – 1962)

Deens natuurkundige

UNIVERSITEIT TWENTE.

Student

T.J. Bemthuis (Thijs)
Student Technische Bedrijfskunde
T.J.Bemthuis@student.utwente.nl
Studentnummer:

Begeleiders

Universiteit Twente (intern)

Begeleider
Dr. ir. M.R.K. Mes (Martijn)
Faculty of Behavioural, Management and Social Sciences
M.R.K.Mes@utwente.nl

Meelezer
Dr. P.C. Schuur (Peter)
Faculty of Behavioural, Management and Social Sciences
P.C.Schuur@utwente.nl

Universiteit Twente (extern), afdeling Marketing & Communicatie

M.E. Overbeek (Marlies)
M.E.Overbeek@utwente.nl

R.A. Schepers (Rob)
R.A.Schepers@utwente.nl

MANAGEMENTSAMENVATTING

Dit verslag bevat een onderzoek dat verricht is naar het vooraf inzichtelijk maken van de verwachte registratie- en opkomstcijfers van de Bachelor Open Dagen van de Universiteit Twente. Dit onderzoek is uitgevoerd als bacheloropdracht van de studie Technische Bedrijfskunde, in opdracht van de afdeling Marketing & Communicatie van de Universiteit Twente. De afgelopen jaren heeft het organiserende team binnen de afdeling Marketing & Communicatie gemerkt dat er een groeiende groep bezoekers afkomt op de Open Dagen. Deze groeiende belangstelling onder studiekiezers en hun ouders/begeleiders brengt organisatorische uitdagingen met zich mee. De organisatie loopt tegen het probleem aan dat het aantal beschikbare zalen met passende omvang niet in lijn ligt met de benodigde capaciteit. Door dit capaciteitsprobleem is de organisatie genoodzaakt geweest om maximum capaciteiten in te stellen voor iedere voorlichtingsronde, met betrekking op de zaal waarin de voorlichting plaatsvindt. De zalen worden momenteel naar het beste inzicht verdeeld over de verschillende opleidingen, op basis van locatie, smaak en historie. Momenteel ontbreekt er de ratio en is er geen ondersteuning voor de zaalplanning op basis van voorspellingen van het verwachte aantal registraties en de verwachte opkomst. Naast de missende ratio voor de zaalplanning, mist de ratio momenteel ook voor de personeelsplanning en de planning van andere middelen. Het is zodoende mogelijk dat er inefficiënte zit in planning van deze aspecten en er dus verspillingen kunnen voorkomen. Voor de zaalplanning kan dit ervoor zorgen dat de zaalplanning niet in lijn ligt met de instroomdoelstellingen van de Universiteit Twente en de bacheloropleidingen daarbinnen. Gezien de zaalplanning wordt verricht voordat het registratieproces geopend is, hebben we de volgende onderzoeksdoelstelling opgesteld: '*Hoe kunnen we vooraf het verwachte aantal registraties en de verwachte opkomst op de Bachelor Open Dagen nauwkeurig inzichtelijk maken, zodoende we met dat inzicht een efficiënte bijdrage kunnen leveren aan de instroomdoelstellingen van de Universiteit Twente en de bacheloropleidingen daarbinnen.'*

Voorspellen kijkt hoe (verborgen) trends in het heden, signalen afgeven voor veranderingen in de toekomst (*Saffo, 2007*). Voorspellen maakt zodoende gebruik van een causaal verband tussen het verleden en de toekomst (*Makridakis & Wheelwright, 1989*). Voor het vooraf inzichtelijk maken van het verwachte aantal registraties en de verwachte opkomst op de Open Dagen hebben we gebruik gemaakt van de verschillende eigenschappen die de data van de aanmeldingen uit het verleden bevatten en deze geëxtrapoleerd naar de toekomst. Voor het nauwkeurig opstellen van de voorspelling hebben we gebruik gemaakt van het gecomponeerd voorspellen. Dit hebben we gedaan doormiddel van het combineren van de voorspellingen van de verschillende doelgroepen die we geïdentificeerd hebben in de data, namelijk aanmeldingen uit 4 VWO, 5 VWO, 6 VWO, overige aanmeldingen uit Nederland, Duitsland en de rest van de wereld. De voorspelling van het aantal aanmeldingen van iedere doelgroep hebben we verder gecomponeerd uit een voorspelling van het aantal aanmeldingen van studiezoekers en daarbij behorende aanmeldingen van extra personen. Doordat we gebruik hebben gemaakt van het onderverdelen van de data in verschillende datasets, was het mogelijk om van iedere specifieke doelgroep de eigenschappen van de data te bepalen. Zodoende was het mogelijk om voor iedere doelgroep een eigen specifieke voorspelling te maken, die we samen hebben kunnen voegen tot één nauwkeurige voorspelling van het verwachte aantal aanmeldingen voor de Open Dagen, die hogere nauwkeurigheid laat zien ten opzichte van het niet gecomponeerd voorspellen.

Uit de analyse van de verschillende doelgroepen hebben we kunnen concluderen dat iedere doelgroep voorkeuren vertoont voor bepaalde voorlichtingsdagen en voorlichtingsprogramma's binnen de Open Dagen. We kunnen zodoende de globale voorspelling van iedere doelgroep statistisch onderverdelen naar dag- en voorlichtingsniveau, waardoor het mogelijk is om een voorspelling te geven voor het aantal aanmeldingen per dag- en voorlichtingsronde. Voor de bepaling van de opkomstcijfers hebben we per voorlichtingsronde en per doelgroep de geregistreerde percentages no-show bepaald.

Doormiddel van deze gegevens kunnen we van het voorspelde aantal aanmeldingen de verwachte opkomst voorspellen. Deze verwachte opkomst geeft een voorspelling weer van het verwachte aantal personen dat aanwezig zal zijn bij een bepaald voorlichtingsprogramma. Met dit inzicht kunnen we bepalen hoe groot een zaal moet zijn per voorlichtingsprogramma om de verwachte opkomst op te kunnen vangen. Door op deze manier per voorlichtingsronde de verwachte opkomsten te bepalen, kunnen per voorlichtingsronde de zalen verdeeld worden, waarbij rekening gehouden kan worden met de instroomdoelstellingen van de verschillende bacheloropleidingen.

Naar aanleiding van het onderzoek hebben we een vijftal aanbevelingen opgesteld, die in onze ogen noodzakelijk zijn voor een nauwkeurige voorspelling óf kunnen bijdragen aan het verbeteren van de nauwkeurigheid van toekomstige voorspellingen:

- De opgestelde voorspelling is opgebouwd uit historische gegevens. Er is gebleken dat gegevens uit een nabijer verleden van grotere waarde zijn voor de nauwkeurigheid van de voorspelling ten opzichte van gegevens uit een ververleden. Het is daarom voor de nauwkeurigheid van de voorspelling aan te raden het voorspellingsmodel telkens aan te passen naar aanleiding van de laatste geobserveerde gegevens, zodat afwijkingen in de trends snel geobserveerd kunnen worden en het voorspellingsmodel kan worden aangepast om de nauwkeurigheid voor toekomstige voorspellingen hoog te houden.
- Voorspellingen kunnen altijd een mate van onnauwkeurigheid met zich meebrengen en zijn dus geen waarzeggerij. Het is daarom van belang om altijd rekening te houden met deze mogelijke onnauwkeurigheid als er beslissingen worden genomen op basis van de voorspellingen.
- In de verkenning van de data is gebleken dat bepaalde doelgroepen bij bepaalde voorlichtingsprogramma's een relatief hoog percentage no-show laten zien, terwijl verklaringen hiervoor momenteel niet onderzocht zijn. We verwachten dat vervuiling in de data een belangrijke invloed hierop heeft, zoals dubbele aanmeldingen en spookaanmeldingen. Voor toekomstig onderzoek raden we aan om de mogelijke factoren voor de percentages no-show te onderzoeken om zo dit percentage eventueel te kunnen verlagen.
- In de opstelling van het voorspellingsmodel was het voor ons niet mogelijk om causale verbanden te verifiëren, terwijl we wel aanwijzingen vonden voor mogelijke causale verbanden met betrekking op het aantal VWO scholieren in Nederland. Voor toekomstig onderzoek raden we dan ook aan om onderzoek te verrichten naar de mogelijke relaties tussen het aantal aanmeldingen/opkomst en factoren in het omliggende milieu, zoals aantal VWO scholieren, marketingactiviteiten en weersinvloeden.
- Het huidige model hebben we gebaseerd op het inzichtelijk maken van de verwachte registratie- en opkomstcijfers voordat het registratieproces geopend is. Dit is gedaan omdat de zaalplanningen voor de opening van het registratieproces globaal vast staan. Voor toekomstig onderzoek raden we aan om naast dit model, ook een onderzoek te verrichten naar de mogelijkheden van een continue voorspellingmodel, waarbij het mogelijk zou zijn om iedere dag gedurende het registratieproces de verwachte aantal aanmeldingen op de Open Dagen te voorspellen. Een model gebaseerd op het "Nearest Neighbour Principle" lijkt ons geschikt daarvoor. We verwachten dat hoe dichter bij de Open Dagen, hoe nauwkeuriger het aantal aanmeldingen voorspeld kunnen worden. Het organiserende team kan zo waar nodig gepaste maatregelen nemen voor eventuele achterblijvers of uitschieters.

VOORWOORD

In april 2016 ben ik aan het einde beland van mijn bacheloropleiding Technische Bedrijfskunde aan de Universiteit Twente, wat nog restte was de bacheloropdracht. Een bacheloropdracht vormt de afsluiting van de bachelor opleiding Technische Bedrijfskunde aan de Universiteit Twente en heeft tot doel de geleerde theorie van de bachelor in de praktijk te brengen, doormiddel van een onderzoeks- en/of ontwerpopdracht. Na verschillende mogelijkheden onderzocht te hebben, ben ik uiteindelijk via dhr. Mes in contact gekomen met mevr. Overbeek van de afdeling Marketing & Communicatie, van de universiteit Twente. Na het eerste contact en een grondig vooronderzoek, ben ik uiteindelijk medio mei begonnen met het werkelijke onderzoek. Voor u ligt het eindverslag van september 2016.

Na de oriënterende eerste gesprekken, werd de doelstelling en de scope van het onderzoek snel duidelijk. Hoewel ik algemeen geïnteresseerd was in de theorie van voorspellingen en het statistisch bewerken van gegevens, was de theorie redelijk nieuw voor mij. Net als bij ieder zelfstandige opdracht, komen er ups en down voor gedurende het proces. Het inlezen van de theorie, om zodoende een goede onderbouwing te krijgen van mogelijke oplossingen, was een veelal tijdrovende bezigheid. Het in praktijk brengen van deze theorie, met in het oog houdend de doelstelling van het onderzoek, kon op verschillende manieren. Deze verschillende manieren zorgde ervoor dat ik wel eens de verkeerde gang in het doolhof van de theorie had genomen, waardoor er weer een stuk terug gelopen moest worden. Maar, gelukkig ben ik niet de enige die problemen ondervond bij het doen van voorspellingen. *"Prediction is very difficult, especially if it's about the future."*, is een beroemde uitspraak van Nils Bohr, die ook op dit onderzoek toepasbaar is. Uiteindelijk viel de theorie op de plaats en heb ik het einde van het doolhof bereikt.

Dit onderzoek, alsmede dit verslag, kon ik natuurlijk niet compleet zelfstandig bewerkstelligen. Als eerste wil ik de betrokken medewerkers vanuit de afdeling Marketing & Communicatie, Marlies Overbeek en Rob Schepers, bedanken voor de begeleiding vanuit de organisatie. De openheid en de duidelijke beantwoording van vragen was zeer bruikbaar en kon ik erg op prijs stellen. Daarnaast wil ik Martijn Mes bedanken voor zijn ondersteuning en kritische blik als eerste begeleider. Zonder zijn medewerking was dit verslag niet mogelijk geweest. Daarnaast wil ik mijn dank uiten aan Peter Schuur voor zijn rol als medebeoordelaar van dit onderzoek. Zijn constructieve blik op het onderzoek en op het verslag was zeer bruikbaar om het verslag naar een hoger niveau te brengen. Als laatste wil ik de overige personen bedanken die direct, of indirect, mij hebben geholpen met het doorlopen van het onderzoek en de opstelling van dit verslag.

Enschede, september 2016

Thijs Bemthuis

Inhoudsopgave

In deze publieke versie zijn enkele gegevens onherkenbaar gemaakt of vervangen door fictieve cijfers.

Hoofdstuk 1

Inleiding

“I never think of the future, it comes soon enough”

Albert Einstein (1879 – 1955)

Duits - Amerikaans natuurkundige

1 INLEIDING

In het kader van de afronding van mijn bacheloropleiding Technische Bedrijfskunde aan de Universiteit Twente, heb ik onderzoek verricht bij de afdeling Marketing & Communicatie van de Universiteit Twente. Dit onderzoek is gefocust op het vooraf inzichtelijk maken van de verwachte registratie- en opkomstcijfers van de Bachelor Open Dagen. Momenteel worden de planning van verschillende aspecten binnen de Open Dagen gebaseerd op historie, locatie en smaak. Een wetenschappelijke onderbouwing op basis van voorspellingen mist, waardoor mogelijke inefficiënties zich kunnen bevinden in de planning van onder andere de zalen.

Hoofdstuk 1 geeft de inleiding tot het onderzoek weer. In paragraaf 1.1 gaan we in op de achtergrondinformatie van de opdrachtgever, om zo een helder beeld te verschaffen van de context waarin het onderzoek zich afspeelt. Paragraaf 1.2 beschrijft de problemen die we geobserveerd hebben en geeft zodoende de aanleiding en de doelstelling van het onderzoek weer. Opeenvolgend zijn in paragraaf 1.3 de onderzoeksvragen weergegeven. In paragraaf 1.4 sluiten we dit hoofdstuk af met een uiteenzetting van de structuur van het verslag.

1.1 DE UNIVERSITEIT TWENTE

De Universiteit Twente is een universiteit gevestigd in de gemeente Enschede. De universiteit biedt zowel bachelor, als masteropleidingen aan op het gebied van techniek en sociale wetenschappen. In 1961 is universiteit opgericht als de toenmalige Technische Hogeschool Twente (THT), als derde technische hogeschool van Nederland, naast Delft en Eindhoven. Met deze twee universiteiten werkt de Universiteit Twente samen onder het 3TU partnership. In 1986 is de naam van de Technische Hogeschool Twente (THT) veranderd naar de huidige naam Universiteit Twente (UT).

De Universiteit Twente onderscheidt zich al vanaf de oprichting in 1961 door de combinatie van techniek en sociale wetenschappen. Daarnaast is de Universiteit Twente de enige universiteit in Nederland die gevestigd is op een campusterrein, waarbij het voor studenten mogelijk is om zowel te studeren als te wonen op het terrein van de universiteit. De UT profileert zich als de ondernemende universiteit. Op de campus zijn ongeveer honderd spin-off bedrijven gevestigd en de universiteit heeft al meer dan 800 succesvolle spin-off bedrijven voortgebracht.

3.300 wetenschappers en professionals zijn werkzaam bij de universiteit op het gebied van onderzoek, innovatie en onderwijs voor de meer dan 9.000 studenten, waarvan ruim 5.000 bachelor studenten en 4.000 master studenten.

1.1.1 De afdeling Marketing & Communicatie

De afdeling Marketing & Communicatie (M&C) van de Universiteit Twente is een centrale afdeling tussen het College van Bestuur, faculteiten, instituten, andere diensten en de wereld buiten de universiteit. De afdeling is opgebouwd uit een drietal kleinere afdelingen, te weten Marketing, Communicatie en Medialab. De afdeling M&C helpt met het realiseren van het strategisch instellingsplan, doormiddel van het versterken van de reputatie van de UT, het creëren van positieve referenten, effectieve werving en zorg te dragen voor een optimale communicatie tussen de universiteit en de verschillende doelgroepen. Met het uiteindelijke doel de instroom van de universiteit te versterken. De marketing- en communicatiestrategie van de afdeling is gericht op hogere zichtbaarheid en profilering, instroom en werving en het kweken en binden van interne en externe ‘ambassadeurs’.

1.1.2 De subafdeling Marketing

De afdeling Marketing is een afdeling binnen de dienst Marketing & Communicatie en richt zich primair op de instroom van studenten voor University College Twente, bachelor- en masteropleidingen. De kerntaken van de afdeling zijn gericht op (verdere) ontwikkeling en realisatie van de marketing- en werving strategieën, beleid en campagnes, advisering College van Bestuur en faculteiten over de inzet

van (marketing)communicatie, opstellen en realisatie jaarplannen, werving & voorlichting (evenementen, activiteiten, middelen), redactie, vormgeving & productie van communicatiemiddelen, beheer en verdere ontwikkeling van CRM en de inzet en coördinatie van studentvoorlichters via het Studie Informatiecentrum.

1.1.3 De open dagen

De Open Dagen van de Universiteit Twente worden twee keer per jaar georganiseerd vanuit de afdeling M&C, in het voor- en in het najaar. De bezoekersaantallen groeien ieder jaar, waarbij de open dagen in het najaar beter bezocht worden dan de open dagen in het voorjaar. Bij de bachelor- en master open dagen van november 2015 waren ongeveer 10.000 bezoekers geregistreerd. Zowel studiekiezers, als hun ouders/begeleiders maken gebruik van de mogelijkheid om de universiteit te bezoeken tijdens de open dagen. Voor de bezoekers is de komst naar één of meerdere open dagen een belangrijke stap in de oriëntatie van de studiekeuze. Voor de marketingdoelstellingen van de afdeling M&C, en daarmee voor de gehele Universiteit Twente, is het van belang om een zo groot mogelijke groep studiekiezers en hun ouders/begeleiders te kunnen ontvangen. Voor aanvang van de open dagen kunnen studiekiezers zich registreren via de website van de universiteit. Hierbij kunnen ze aangeven of er ook begeleiders/ouders mee komen.

1.2 DE AANLEIDING

Voor de structurering van het onderzoek naar de aanleiding van de problematiek, maken we gebruik van fase 1, de probleemidentificatie, van de Algemene Bedrijfskundige Probleemaanpak (*Heerkens & van Winden, 2012).* Aan de hand van de vierstappen die deze fase voorschrijft (Tabel 1.1*)* hebben we deze paragraaf opgesteld. Subparagraaf 1.2.1 bevat de inventarisatie van de problemen die we geobserveerd hebben. Subparagraaf 1.2.2 bevat de probleemkluwen. In Subparagraaf 1.2.3 hebben we de het kernprobleem opgesteld, aan de hand van de probleemkluwen. Stap 4 behandelen we in subparagraaf 1.2.4, waarbij we ingaan op de operationalisatie van de gevonden problematiek.

Stap 1:	Inventariseer welke problemen er zijn.
Stap 2:	Geef aan wat oorzaken en gevolgen zijn en zet ze in een probleemkluwen.
Stap 3:	Kies het kernprobleem dat je gaat aanpakken.
Stap 4:	Maak je probleem meetbaar.

Tabel 1.1: Stappen van de probleemidentificatie (ABP, Heerkens & van Winden, 2012)

1.2.1 De probleeminventarisatie

De eerste stap die de probleemidentificatie voorschrijft is de inventarisatie van de problematiek. In dit onderzoek hebben we de definitie van een probleem aangenomen die de Algemene Bedrijfskundige Probleemaanpak, ABP, voorschrijft. Een probleem wordt hierbij gedefinieerd als een situatie waarbij de werkelijkheid afwijkt van de norm. In de ABP wordt deze definitie gedefinieerd als een handelingsprobleem[1].

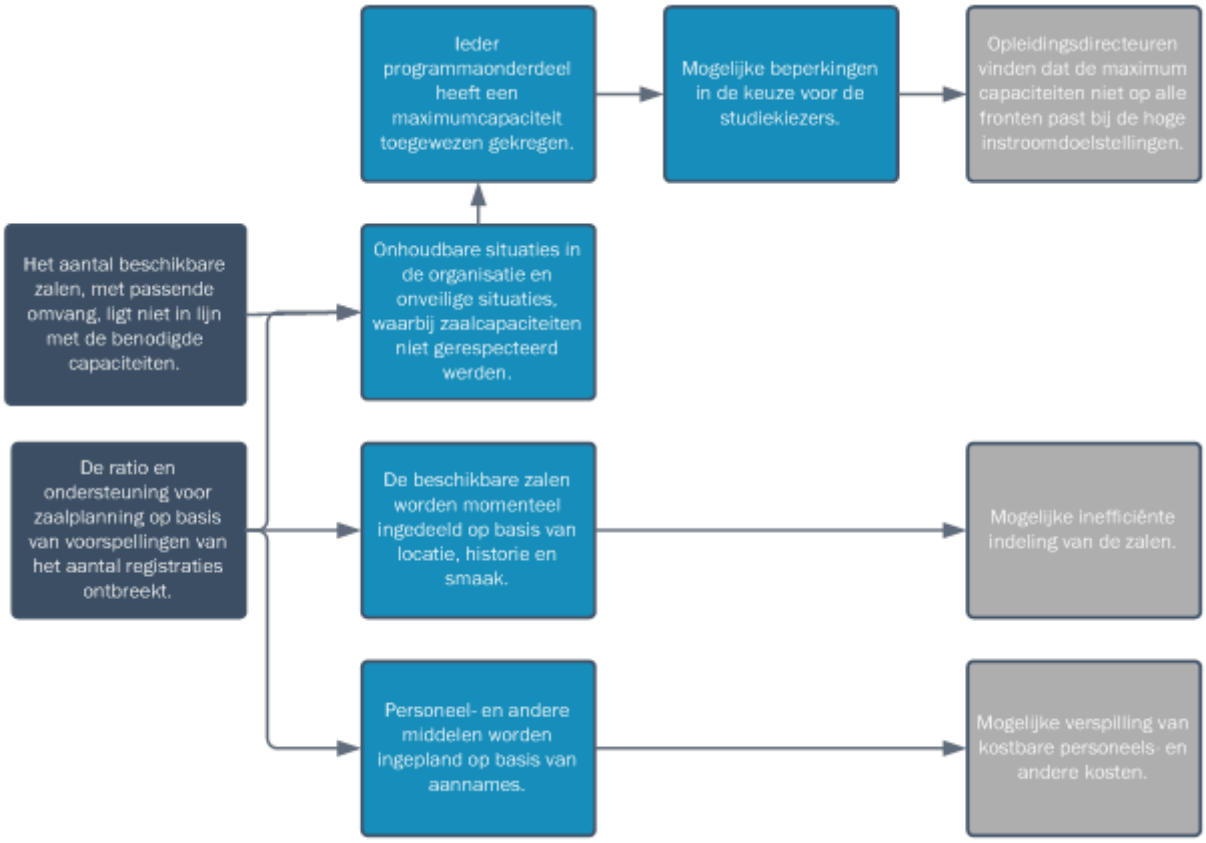
De problemen hebben we geïnventariseerd op basis van gesprekken met de betrokken medewerkers van de afdeling Marketing & Communicatie[2]. Hierbij hebben wij ook de voortijdig beschikbare informatie meegenomen, zoals weergegeven in de gepubliceerde opdrachtomschrijving. Voor de volledigheid hebben we in Appendix 1 ook oorzaken geïdentificeerd, die niet gekenmerkt worden als een probleem, maar wel invloed hebben op de ontstaande problematiek.

[1] Een handelingsprobleem is een door de probleemhebber waargenomen discrepantie tussen de norm en werkelijkheid (Heerkens & van Winden, 2012).

[2] De betrokken medewerkers zijn hierbij geïdentificeerd als de probleemhebber.

1.2.2 De probleemkluwen

De volgende stap die de probleemidentificatie voorschrijft is het controleren van oorzaak-gevolg problemen, doormiddel van een probleemkluwen. Probleemkluwen zorgt voor samenhang tussen de problemen (*Heerkens & van Winden, 2012*). In Figuur 1.1 is de probleemkluwen weergegeven van de problemen die we in de vorige sub paragraaf hebben geïdentificeerd. De pijlen geven de richting aan van oorzaak naar gevolg. De vierkante blokken geven representatief het oorzaak of het gevolg aan. In de kluwen hebben we enkele aannames gemaakt, die we op de volgende pagina verder zullen toelichten. In Appendix 2 is een uitvergrote probleemkluwen te vinden. Hierin zijn voor de volledigheid ook de oorzaken, die niet we niet als problemen geïdentificeerd hebben, ook meegenomen.



Figuur 1.1: De probleemkluwen

De probleemkluwen in Figuur 1.1 geeft systematisch weer wat de gevolgen en oorzaken zijn van de waargenomen problematiek. Het probleem dat de beschikbare zalen niet in lijn liggen met de benodigde capaciteit heeft echter verschillende oorzaken, die niet terug te vinden zijn in het figuur. Deze oorzaken hebben we niet meegenomen in de probleemkluwen, gezien een probleem pas in de probleemkluwen komt te staan als het ook daadwerkelijk een probleem is. De gevonden oorzaken hebben we niet als problemen geïdentificeerd, maar hebben we voor de volledigheid wel weergeven in Appendix 2.

Uit de probleemkluwen kunnen we afleiden dat er twee problemen zijn die aan de basis liggen van de kluwen. Als eerste hebben we het probleem dat het aantal beschikbare zalen niet in lijn ligt met de benodigde capaciteiten. De organisatie van de Open Dagen heeft voor de zaalindeling voorrang op alle beschikbare zalen van de universiteit. Met andere woorden kan het probleem dus omschreven worden dat er momenteel een te kort is aan grotere zalen die gebruikt kunnen worden voor de Open Dagen. Dit komt onder andere door een stijging in het aantal bezoekers, maar heeft nog meer oorzaken, zoals weergegeven in Appendix 2. Dit capaciteitsprobleem heeft in het verleden ervoor gezorgd dat er onveilige situaties zich hebben kunnen voordoen. In november 2014 leidde dit tot onhoudbare situaties in de organisatie en onveilige situaties waarbij de zaalcapaciteiten niet

gerespecteerd werden. Voor bepaalde programmaonderdelen was dus een zaal met een bepaalde capaciteit ter beschikking gesteld. Deze capaciteit was gebaseerd op het verwachte aantal bezoekers, maar tijdens de Open Dagen oversteeg het aantal bezoekers de zaalcapaciteit. Hierbij wordt ook meteen een ander probleem duidelijk, namelijk dat het werkelijke aantal bezoekers uitsteeg boven het verwachte aantal bezoekers. Immers, als de organisatie van te voren wist dat er meer bezoekers zouden komen, zouden ze bijbehorende maatregelen kunnen treffen om het probleem te voorkomen. Dit onderschrijft het tweede probleem dat aan de basis ligt van de kluwen: Het probleem dat de ratio en ondersteuning mist voor de zaalplanning op basis van voorspellingen.

Dit gebrek aan ratio en ondersteuning heeft tot gevolg dat momenteel de zalen worden ingedeeld op basis van historie, locatie en smaak. Met andere woorden betekent dit dat momenteel de zalen worden ingedeeld naar de voorkeuren van de opleidingsdirecteuren. Voor eventuele aanpassingen in de indeling is geen ondersteuning op basis van voorspellingen, waardoor dit op weerstand kan rekenen van enkele opleidingsdirecteuren en mogelijk ruimte inefficiënt gebruikt wordt. Daarnaast wordt het personeel en andere middelen ook ingepland op basis van het verwachte aantal bezoekers. Zonder een degelijke ondersteuning, is het dus mogelijk dat personeel inefficiënt wordt ingepland, wat leidt tot onnodige kosten.

De onveilige situaties hebben tot gevolg gehad dat de organisatie zich genoodzaakt heeft om maximum capaciteiten in te stellen per programmaonderdeel. Deze maximum capaciteiten moeten tot gevolg hebben dat de onveilige situaties tot het verleden behoren, maar daarnaast heeft het instellen van maximum capaciteiten ook tot gevolg dat studiekiezers minder keuze hebben in de dagdelen, indien een programmaonderdeel vol is. Hierbij moet wel vermeld worden dat de organisatie ervoor zorgt dat het altijd mogelijk is voor een studiekiezer om een voorlichting in een ander dagdeel te volgen. De mogelijkheid bestaat hierdoor dat potentiële studiekiezers niet naar een Open Dag komen, omdat het gewenste dagdeel vol zit. Voor opleidingsdirecteuren zou dit kunnen betekenen dat potentiele studenten wegblijven en dit is nadelig voor de opleidingen, als ook voor de instroomdoelstellingen van de organisatie. Het is daarbij van belang om de zaalindeling zo efficiënt mogelijk te plannen, zodat minimale ruimte wordt verspilt en maximale studiekiezers kunnen worden toegelaten.

1.2.3 Het kernprobleem

In de vorige subparagraaf hebben we de geïnventariseerde problemen uit paragraaf 1.2.1 in een probleemkluwen weergegeven. In Tabel 1.2 is het stappenplan weergegeven dat de ABP voorschrijft voor het identificeren van het kernprobleem uit de probleemkluwen. Het kernprobleem is het probleem dat we aanpakken in dit onderzoek.

Regel 1:	Het probleem moet ook daadwerkelijk een probleem zijn.
Regel 2:	In de keten van de problemen ga je terug naar de problemen die zelf geen oorzaak meer hebben.
Regel 3:	Wat je niet kan beïnvloeden, kan geen kernprobleem zijn.
Regel 4:	Kies als er nog problemen overblijven in de probleemkluwen, het belangrijkste probleem om op te lossen.

Tabel 1.2: Stappenplan kernprobleem identificatie (Heerkens & van Winden, 2012)

Naast de regels, zoals weergegeven in Tabel 1.2, is het voor het onderzoek van belang dat we het kernprobleem kiezen in overeenstemming met de opdrachtgever. Immers, als we een kernprobleem kiezen zonder overeenstemming met de opdrachtgever, kan het eindresultaat niet het gewenste resultaat bevatten en daarmee het hele onderzoek nutteloos maken. Deze restrictie nemen wij dan ook mee, als vijfde regel.

Alle problemen die weergegeven zijn in de probleemkluwen voldoen aan regel 1, immers een probleem komt pas in de probleemkluwen te staan als het ook een werkelijk een probleem is. Aan regel 2 voldoen twee verschillende problemen volgens de probleemkluwen, zoals weergeven in Figuur

1.1 Deze problemen zijn helemaal links in de probleemkluwen terug te vinden en hebben we donderblauw weergegeven.

Het eerste probleem dat we onderzoeken is het capaciteitsprobleem. Dit probleem heeft geen oorzaken meer volgens de probleemkluwen, zoals weergeven in Figuur 1.1. Echter, heeft dit probleem meerdere oorzaken, zoals te zien is in Appendix 2. Deze oorzaken zijn niet als problemen geïdentificeerd, omdat deze (voor een deel) bijdragen aan de doelstellingen van de afdeling M&C en daarmee van de universiteit. Kort door de bocht wordt dit probleem veroorzaakt door meer bezoekers en een stabiel aanbod van zalen met passende omvang. De oorzaak meer bezoekers ligt in lijn met de doelstelling van de universiteit en zal daarom niet moeten worden opgelost (het is immers geen probleem). Het aanbod van zalen met passende omvang kan worden opgehoogd, maar dit betekent dus dat er geschikte zalen bij moeten komen. Uit een gesprek met de betrokken medewerkers van de afdeling M&C blijkt dat een doel van de Open Dagen het laten zien van de campus is en daarmee de uitstraling van de universiteit laten zien. Het eventueel betrekken van zalen buiten het campusterrein heeft daarom niet de voorkeur en valt daarom af. Zodoende voldoet dit probleem niet aan regel 3 en valt dit probleem dus af als kernprobleem.

Het probleem dat overblijft als kernprobleem is het gebrek aan ratio en ondersteuning op basis van voorspellingen. Dit probleem voldoet aan regel 2, immers het heeft geen oorzaak meer. Daarnaast voldoet dit probleem ook aan regel 3, immers een ratio en ondersteuning kan gecreëerd worden en daarmee kan dit probleem beïnvloed worden. Gezien er geen problemen meer overblijven, hoeft regel 4 niet in werking te treden.

Naast dat dit probleem aan alle regels voldoet, is dit probleem ook aangedragen als kernprobleem door de probleemhebbers. Daarmee is het gebrek aan ratio en ondersteuning voor de zaalplanning van de Open Dagen op basis van voorspellingen van het aantal registraties gekozen als kernprobleem voor dit onderzoek.

1.2.4 Operationalisatie

In de vorige subparagraaf hebben we het kernprobleem geïdentificeerd vanuit de probleemkluwen. In deze subparagraaf maken we het geïdentificeerde kernprobleem meetbaar en daarmee hanteerbaar voor het verdere onderzoek. In paragraaf 1.2.1 hebben we alle problemen in de probleemidentificatie gedefinieerd als handelingsproblemen. Een handelingsprobleem is hierbij een verschil tussen de norm en de werkelijkheid. In het meetbaar maken van het kernprobleem is het noodzakelijk om de norm en werkelijk uit te drukken in concrete en meetbare variabelen (*Heerkens & van Winden, 2012).* Voor het meetbaar maken van het kernprobleem hebben we gekozen om één variabele te gebruiken. Hiervoor is gekozen om zo efficiënt mogelijk het probleem op te kunnen lossen. Uit het kernprobleem hebben we het volgende handelingsprobleem opgesteld:

Handelingsprobleem: *Er is een gebrek aan ondersteuning voor de zaalplanning van de bachelor open dagen op basis van nauwkeurige voorspellingen van het aantal registraties.*

De variabele is hetgene dat in de werkelijkheid afwijkt ten opzichte van de norm. In het handelingsprobleem hebben we de ondersteuning gekozen als de variabele. Deze variabele kan liggen tussen geen ondersteuning (werkelijkheid) en wel ondersteuning (norm). Het is duidelijk dat hierbij de werkelijkheid afwijkt van de norm.

Ondersteuning moet gecreëerd worden. We kunnen uit het handelingsprobleem afleiden dat er momenteel een gebrek aan ondersteuning of op basis van voorspellingen. Het verschil tussen de norm en werkelijkheid is de mate waarin nauwkeurige voorspellingen beschikbaar zijn voor de ondersteuning, immers zullen onnauwkeurige voorspellingen geen goede ondersteuning kunnen bieden.

Het doel van het onderzoek is de werkelijkheid en de norm gelijk maken, oftewel het handelingsprobleem op te lossen. Uit het handelingsprobleem hebben we de volgende onderzoeksdoelstelling geformuleerd:

Onderzoeksdoelstelling: *Het vooraf nauwkeurig inzichtelijk maken van het verwachte aantal registraties en de verwachte opkomst op de bachelor Open Dagen, zodoende dat inzicht een efficiënte bijdrage kan leveren aan de instroomdoelstellingen van de UT en de bacheloropleidingen daarbinnen.*

Aan het eind van het onderzoek is het te verwachten dan de doelstelling behaald is. Het is hierbij belangrijk om deze doelstelling meetbaar te maken. De variabele in deze doelstelling is het nauwkeurige inzicht in het verwachte aantal registraties en opkomst op de Open Dagen. Het operationaliseren van deze variabele geeft betekenis aan het onderzoek. Nauwkeurig inzicht is een breed begrip. We kunnen deze variabele meetbaar maken door te kijken naar de invloed die dit inzicht eventueel kan hebben. Uit de probleemkluwen (Figuur 1.1*)*, is af te leiden dat dit inzicht gebruikt kan gaan worden voor verschillende planningsaspecten. Het is hierbij van belang dat het inzicht nauwkeurig is, immers een inzicht met grote onnauwkeurigheid zal het probleem niet verhelpen. De variabele inzicht kan dus meetbaar gemaakt worden door de afwijking van het verkregen inzicht tegenover de realiteit. Het doel is een zo nauwkeurig mogelijk inzicht, met dus een zo min mogelijke onnauwkeurigheid. Gezien we de toekomst niet met zekerheid weten, kunnen we het inzicht meetbaar maken op basis van historische gegevens.

Het te verkrijgen inzicht moet een bijdrage leveren aan de instroomdoelstellingen van de UT en de bacheloropleidingen daarbinnen. Uit verschillende interviews met de betrokken medewerkers van M&C is gebleken dat deze bijdrage voornamelijk betrekking heeft op de verdeling van de zalen onder de verschillende voorlichtingsprogramma's. De instroomdoelstellingen per bacheloropleiding verschillen en daardoor hebben sommige kleinere studies meer ruimte nodig om potentiele studenten binnen te halen, tegenover opleidingen die normalerwijs altijd een hoog aantal studenten binnen halen. De zaalindeling wordt bepaald voordat het inschrijfproces open is voor studiezoekers. Dat betekend dus dat, dat het vooraf inzichtelijk maken betrekking heeft op het voorspellen voordat de zaalplanning gedaan wordt en dus voordat het registratieproces geopend wordt.

1.3 ONDERZOEKSVRAGEN

In de vorige paragraaf hebben we de problemen geïdentificeerd en het kernprobleem meetbaar gemaakt. Uit het kernprobleem hebben we de onderzoeksdoelstelling bepaald. In deze paragraaf gebruiken we de opgestelde onderzoeksdoelstelling om de onderzoeksvragen voor dit onderzoek op te stellen.

Als eerste zetten we de onderzoeksdoelstelling om in een onderzoeksvraag. Het antwoord op deze hoofdvraag representeert het eindresultaat van het onderzoek. Uit de onderzoeksdoelstelling hebben we de volgende hoofdvraag opgesteld:

Hoofdvraag: *Op welke wijze kan vooraf nauwkeurig inzicht worden gegeven in het verwachte aantal registraties en de verwachte opkomst op de Bachelor open dagen en op welke wijze kan dit inzicht leiden tot een efficiënte bijdrage aan de instroomdoelstellingen van de UT en de bacheloropleidingen daarbinnen?*

Voor de beantwoording van de hoofdvraag, hebben we de hoofdvraag onderverdeeld in verschillende deelvragen. Elk van deze deelvragen levert een bijdrage aan de beantwoording van de hoofdvraag en daarmee aan het behalen van de onderzoeksdoelstelling.

Deelvraag 1: Theoretisch kader
Centrale deelvraag: *Hoe kan efficiënt voorspeld worden?*

Deze deelvraag gaan we beantwoorden doormiddel van de volgende subvragen:

- *Wat is voorspellen?*
- *Welke voorspellingsmethoden en technieken bestaan er?*
- *Wat is het proces van efficiënt voorspellen?*

- *Welke data is noodzakelijk voor efficiënt voorspellen?*
- *Hoe kan de nauwkeurigheid van een voorspelling bepaald worden?*

In het theoretische kader scheppen we de wetenschappelijke basis voor de rest van het onderzoek. We gaan hierbij in op de verschillende aspecten die de literatuur voorschrijft voor het efficiënt doen van voorspellingen. Als eerste gaan we hierbij in op het begrip voorspellen en zetten we uiteen wat precies hieronder verstaan wordt, zodat er een wetenschappelijk kader zichtbaar wordt waarin het onderzoek zich afspeelt. Vervolgens gaan we ons verder verdiepen in de verschillende methoden en technieken die aanwezig zijn binnen de literatuur, met betrekking op het doen van voorspellingen. Om het onderzoek efficiënt te laten verlopen, besteden we in dit hoofdstuk ook aandacht aan het proces van voorspellen. Het proces gaan we hierbij uiteenzetten zodat we de noodzakelijke data kunnen identificeren voor verder onderzoek. Als laatste besteden we in dit hoofdstuk aandacht aan het nauwkeurigheidsprincipe, zodat we een onderbouwing geven voor het begrip nauwkeurigheid wat in de onderzoeksdoelstelling vermeld is.

Deelvraag 2: Inventarisatie historische data
Centrale deelvraag: *Wat zijn de eigenschappen van de beschikbare historische data?*

In de vorige deelvraag scheppen we een theoretisch kader voor het onderzoek en gaan we onder andere in op het proces van efficiënt voorspellen. In deze deelvraag gaan we verder in op het proces van voorspellen en onderzoeken we welke historische data beschikbaar zijn voor de voorspelling en wat de eigenschappen van deze data zijn.

Deelvraag 3: Het verwachte aantal aanmeldingen
Centrale deelvraag: *Hoe kan het aantal verwachte registraties nauwkeurig voorspeld worden?*

In de vorige deelvraag verrichten we een onderzoek naar de data die beschikbaar zijn voor de voorspelling en zetten we de eigenschappen van deze data uiteen. In deze deelvraag gaan we hier verder op in en gaan we de verschillende eigenschappen van de data linken aan de theorie. Dit gaan we doen aan de hand van de verschillende voorspellingstechnieken- en methoden die naar voren zijn gekomen in de eerste deelvraag. Hieruit gaan we de meest geschikte voorspellingstechniek kiezen, aan de hand van de hoogste nauwkeurigheid en toepasbaarheid, waarmee we het werkelijke voorspellingsmodel van het verwachte aantal aanmeldingen gaan opstellen.

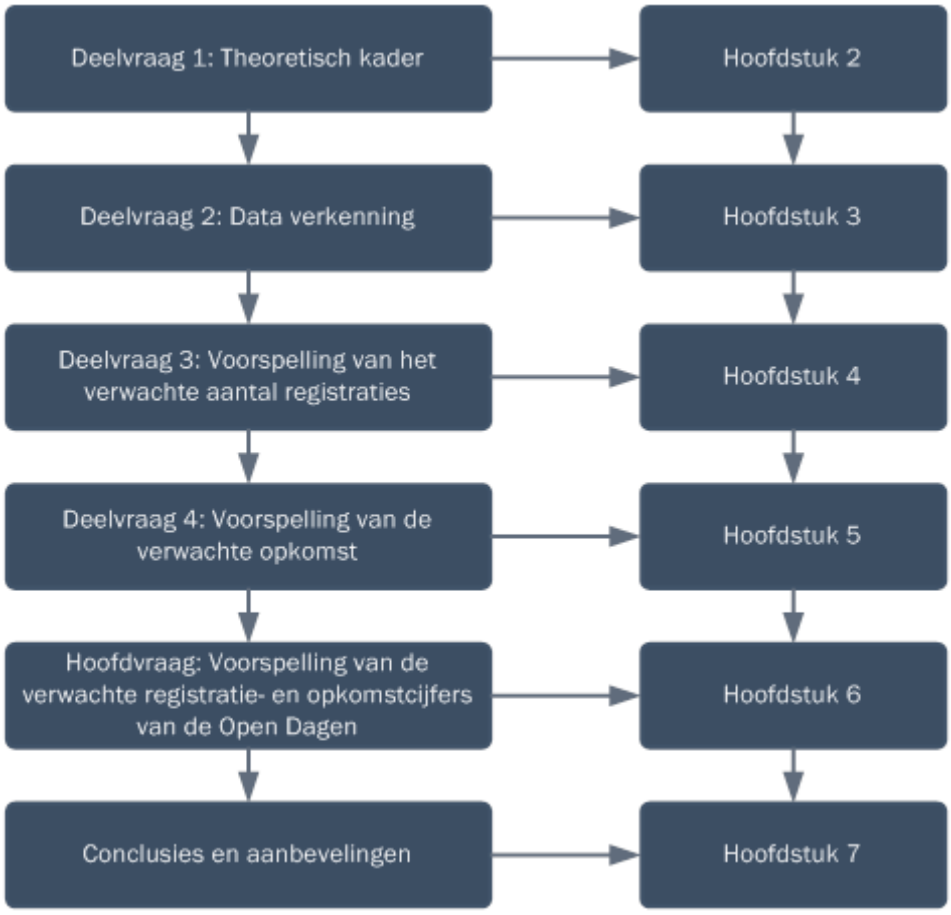
Deelvraag 4: De verwachte opkomst
Centrale deelvraag: *Hoe kan de opkomst nauwkeurig voorspeld worden?*

In de vorige deelvraag stellen we het voorspellingsmodel van het verwachte aantal aanmeldingen op. In deze deelvraag gaan we hier verder op in en gaan we het voorspellingsmodel van de werkelijke opkomst opstellen. Dit gaan we doen aan de hand van de data eigenschappen die naar voren zijn gekomen in de tweede deelvraag. We gaan hieruit de meest geschikte voorspellingstechniek kiezen, aan de hand van de hoogste nauwkeurigheid en toepasbaarheid.

1.4 DE STRUCTUUR

De deelvragen die zijn opgesteld in de vorige deelvraag gaan we in chronologische volgorde behandelen in dit verslag (Figuur 1.2). De deelvragen zijn onderling afhankelijke en borduren voort op kennis die vergaard is in de voorgaande deelvragen. We gaan beginnen met de beantwoording van deelvraag 1, zodat we een wetenschappelijke kader hebben voor de beantwoording van de deelvragen die daarop volgen. In hoofdstuk 3 gaan we de kennis uit hoofdstuk 2 gebruiken om de noodzakelijke data te vergaren en hiervan de eigenschappen te bepalen. Hiermee gaan we deelvraag 2 beantwoorden. Nadat de data verzameld is en we de eigenschappen ervan bepaald hebben, is het

mogelijk om de theorie te linken aan de praktijk en de meest geschikte voorspellingstechniek te kiezen. Dit gaan we doen in hoofdstuk 4 en daarbij gaan we het voorspellingsmodel opstellen voor het verwachte aantal aanmeldingen, waarmee we deelvraag 3 gaan beantwoorden. In hoofdstuk 5 borduren we hierop voort en gaan we de uitkomsten uit deelvraag 1, 2 en 3 gebruiken om een voorspellingsmodel op te stellen voor de verwachte opkomst op de Open Dagen. Nadat we de verschillende deelvragen beantwoord hebben, gaan we de hoofdvraag beantwoorden. Voor de beantwoording van de hoofdvraag gaan we in hoofdstuk 6 de verschillende voorspellen samenvoegen, om zodoende een voorspelling te maken van de toekomstige Open Dagen. Hierbij gaan we ook de nauwkeurigheid van voorspellingen uit het verleden uiteenzetten. Dit verslag sluiten we af in hoofdstuk 7 met de conclusies van dit onderzoek en de aanbevelingen voor toekomstig onderzoek.



Figuur 1.2: Opbouw van het verslag

Hoofdstuk 2

Theoretisch kader

(Deelvraag 1)

“In theory, theory and practice are the same. In practice, they are not.”

Albert Einstein (1879 – 1955)

Duits - Amerikaans natuurkundige

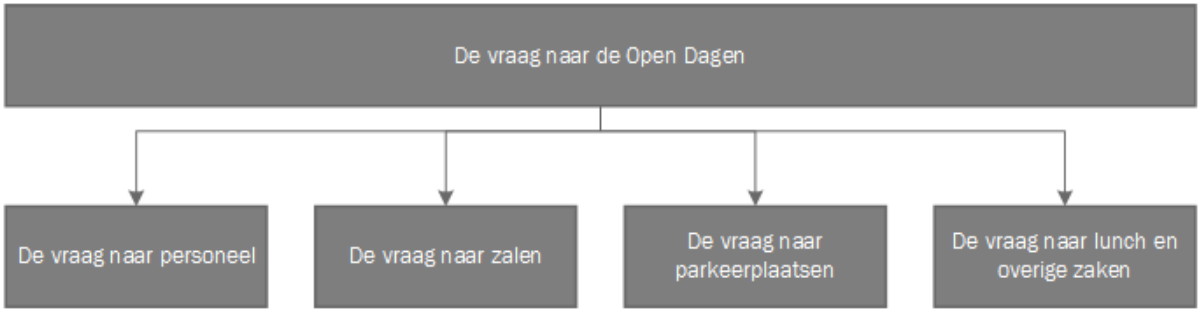
2 THEORETISCH KADER

Hoofdstuk 2 geeft een theoretisch kader weer voor de uitvoering van het onderzoek. De eerste paragraaf bevat een introductie tot het begrip voorspellen. De paragraaf die volgt gaat in op de verschillende voorspellingsmethoden die de literatuur voorschrijft en hoe deze te combineren vallen. In paragraaf 2.3 wordt het voorspellingsproces uiteengezet en aandacht besteed aan de noodzakelijke stappen voor een succesvolle en efficiënte voorspelling. Paragraaf 2.4 bevat een uiteenzetting van het begrip data en gaat in op de verschillende manieren hoe data vergaard kan worden. In paragraaf 2.5 wordt ingegaan op verschillende voorspellingstechnieken die de literatuur voorschrijft en tot slot bevat paragraaf 2.6 een uiteenzetting van verschillende methoden om de nauwkeurigheid van een voorspelling te bepalen.

2.1 INLEIDING

Het Nederlandse woordenboek definieert voorspellen als “het als verwachting uitspreken” (*van Dale, 2016*). In Wetenschappelijke termen kan voorspellen gedefinieerd worden als een projectie van de toekomst (*Heroman, Davis & Farmer, 2012*). In het dagelijks leven wordt veel gebruik gemaakt van voorspellingen. Voorspellen gebeurd op verschillende manieren en voor allerlei verschillende toepassingen, bijvoorbeeld aardbevingen, beurskoersen en voetbaluitslagen.

De doelstelling van dit onderzoek is het vooraf inzichtelijk maken van het verwachte aantal registraties en de verwachte opkomst op de Open Dagen. Het vooraf inzichtelijk maken van het verwachte aantal registraties en de verwachte opkomst kan gezien worden als het voorspellen van de vraag naar de Open Dagen. De vraag naar de Open Dagen zorgt voor een vraag naar verschillende aspecten, zoals lunch, ruimte en personeel (Figuur 2.1). Een voorspelling kan zodoende op meer aspecten van invloed zijn dan in eerste instantie gedacht kan worden.



Figuur 2.1: Mogelijke invloed van de vraag naar de Open Dagen

Niet alles kan voorspeld worden. Voorspellen kijkt hoe (verborgen) trends in het heden, signalen afgeven voor veranderingen in de toekomst (*Saffo, 2007*). Voorspellen maakt zodoende gebruik van een causaal verband tussen het verleden en de toekomst (*Makridakis & Wheelwright, 1989*) (Figuur 2.2).

Figuur 2.2: Systematische weergave causaal verband in voorspellen

Voorspellingen gaan altijd gepaard met een bepaalde onzekerheid en het is dus geen waarzeggerij. Voorspellen heeft als doel een breed scala aan mogelijkheden te ontdekken en dus niet een gelimiteerd scala aan zekerheden (*Saffo, 2007)*. Hoewel voorspellingen van grote waarde kunnen zijn voor veel doeleinden, heeft het ook beperkingen. Deze beperkingen kunnen kort samengevat worden in de drie wetten van voorspellen (Tabel 2.1*).*

Wet 1:	Voorspellingen zijn altijd verkeerd (er is altijd onnauwkeurigheid)
Wet 2:	Samengevoegde voorspellingen zijn nauwkeuriger dan op zichzelf staande voorspellingen
Wet 3:	De onnauwkeurigheid van de voorspelling neemt toe met de tijd

Tabel 2.1: De drie wetten van voorspellen (Hopp & Spearmann, 2001)

2.2 VOORSPELLINGMETHODEN

Voorspellingen zijn op een breed veld toepasbaar. Bijna alles waar we in het dagelijks leven mee in aanraking komen maakt op de een of andere manier gebruik van voorspellingen. Dit kan iets simpels zijn zoals de vraag van een lokaal restaurant, tot ingewikkelde voorspellingen van aardbevingen en beurskoersen. Gezien voorspellingen op zo'n breed veld toepasbaar zijn, zijn er verschillende methoden voor het uitvoeren van voorspellingen, met elk een specifieke toepassing. Globaal gesproken kunnen voorspellingen onderverdeeld worden in twee categorieën, namelijk kwalitatieve voorspellingen en kwantitatieve voorspellingen (*Hopp & Spearmann, 2001) (*Figuur 2.3*).* Volgens wet 2 van voorspellingen (Tabel 2.1) zorgt een combinatie van voorspellingen voor een hogere nauwkeurigheid.

Figuur 2.3: Soorten voorspellingsmethoden

2.2.1 Kwalitatieve voorspellingmethoden

Kwalitatief voorspellen maakt gebruikt van menselijke kennis voor het voorspellen van de toekomst. Deze methode kan gebruikt worden bij situaties waarbij weinig data uit het verleden beschikbaar is, zoals bij het voorspellen van de vraag van een nieuw product (*Chopra & Meindl, 2007*). Het voorspellen gebeurd dan op basis van de expertise van de voorspeller. Er zijn verschillende soorten kwalitatieve voorspellingen, zoals genius forecasting, scenario forecasting en consensus forecasting (*Joseph, 1992*).

Genius forecasting is gebaseerd op een combinatie van intuïtie en inzicht. Om een duidelijk beeld te krijgen van een extreme vorm van genius forecasting kan gedacht worden aan sciencefiction schrijvers. Een voorbeeld hiervan is de film “Back to the future 2” uit 1989, waarin verschillende technologieën uit 2015 voorspeld werden. Uiteindelijk bleken enkele aspecten juist voorspeld te zijn. De zwakte van deze vorm van voorspellen is dat het onmogelijk is om de juistheid van de voorspelling in te schatten, totdat uiteindelijk de voorspelling werkelijk plaatsvindt.

Scenario forecasting maakt gebruik van verschillende verhaallijnen die een potentiele gebeurtenis beschrijven. In de scenario’s kunnen causale verbanden worden meegenomen, maar ook invloeden op het systeem als geheel. Het kan gezien worden als een script voor het voorspellen van de toekomst, waarin bijvoorbeeld de invloed van nieuwe technologieën of het veranderen van consumentengedrag wordt meegenomen. De scenario’s zijn meestal geschreven voor het voorspellen van de lange termijn en worden meestal in een positieve versie, als een negatieve versie geschreven. Scenario voorspellingen hebben voornamelijk tot doel om de besluitvormers aan het denken te zetten wat mogelijke veranderingen voor invloed kunnen hebben. Een voorbeeld hiervan zijn de verschillende scenario’s die worden weergegeven bij een vertrek van Groot-Brittannië uit de Europese Unie (Brexit).

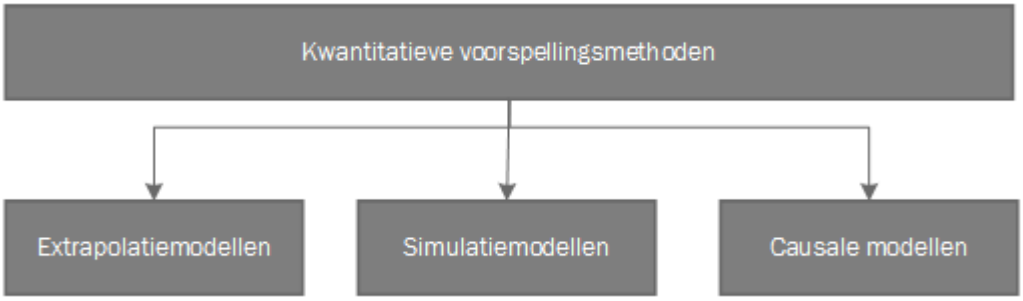
Consensus forecasting maakt gebruik van een combinatie van meningen van experts op bepaalde gebieden. Een overeenstemming van de opinies van deze experts wordt dan meegenomen als voorspelling. In de simpele consensus methode (bijvoorbeeld in een groepsdiscussie) kan het voorkomen dat bepaalde experts het voortouw nemen en hun mening proberen door te drukken, waardoor de voorspelling bevooroordeeld kan zijn. Hierbij kan gedacht worden aan het voorspellen van een voetbaluitslag bij Studio Sport. Een betrouwbaarder methode is de Delphi voorspellingsmethode. De Delphi Methode (Figuur 2.4*)* maakt gebruik van verschillende rondes met geanonimiseerde opinies van verschillende experts. Deze opinies worden anoniem voorgelegd aan de verschillende experts, waarbij ze hierop reageren en eventueel hun eigen opinie bijstellen. Dit gaat enkele rondes door tot de laatste ronde waarin gevraagd wordt aan de experts om hun eigen originele mening bij te stellen naar aanleiding van de mening van de andere experts. Het geanonimiseerde aspect vergroot de betrouwbaarheid ten opzichte van groepsdiscussies.

Figuur 2.4: Systematische weergave Delphi methode

Naast de bovengenoemde kwalitatieve voorspellingsmethoden, zijn er nog meer methoden die onder deze noemer geschaard kunnen worden, zoals voorspellen op basis van marktonderzoek via vragenlijsten. Deze methoden komen allemaal overeen dat ze voorspellen op basis van menselijke input. Voor het voorspellen van het verwachte aantal registraties en opkomst op de Open Dagen zullen we een wetenschappelijke aanpak hanteren, die voornamelijk betrekking heeft op het wiskundig bewerken van historische gegevens; het kwantitatief voorspellen. Zodoende zullen we niet verder ingaan op kwalitatieve voorspellingsmethoden.

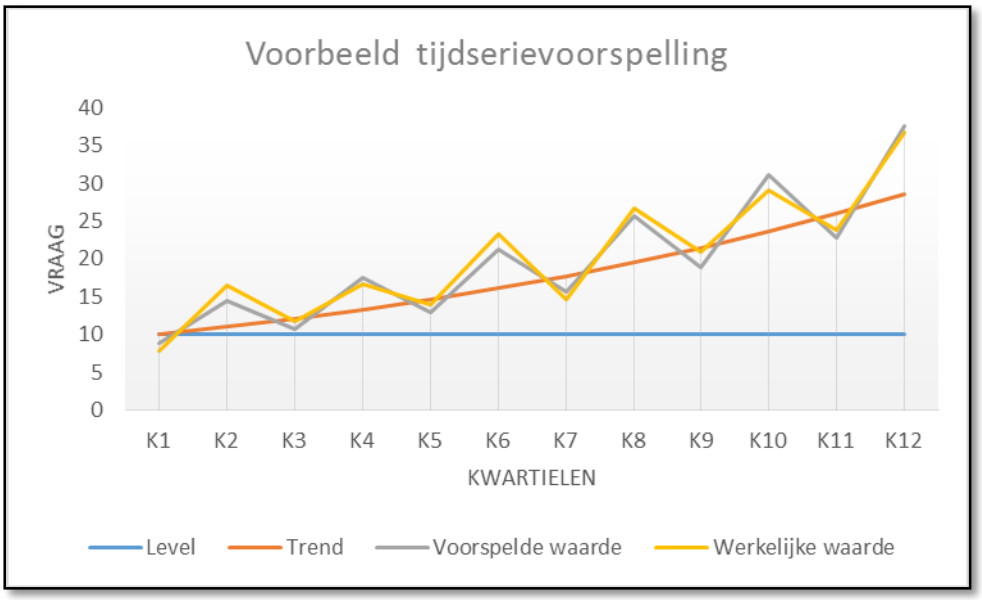
2.2.2 Kwantitatieve voorspellingmethoden

Kwantitatief voorspellen maakt, in tegenstelling tot kwalitatief voorspellen, gebruik van het wiskundig bewerken van gegevens uit het verleden, om zodoende een voorspelling te maken van de toekomst. Hierbij wordt er vanuit gegaan dat in het verleden behaalde resultaten een goede indicator zijn voor de toekomst. Er bestaan verschillende soorten kwantitatieve voorspellingen, die globaal onderverdeeld kunnen worden in extrapolatiemodellen, causale modellen en simulatiemodellen (*Chopra & Meindl, 2007)* (Figuur 2.5*)*.



Figuur 2.5: Kwantitatieve voorspellingsmodellen

Extrapolatievoorspellingsmodellen maken gebruik van extrapoleren van het in het verleden behaalde resultaten naar de toekomst. Hierbij wordt gebruik gemaakt van bepaalde resultaten per tijdspanne en worden ook wel tijdreeksen genoemd. Deze vorm van voorspellen is het meest bruikbaar bij situaties waarbij er weinig significante variatie is in de vraag over de jaren. Over het algemeen zijn tijdreeksvoorspellingsmethoden makkelijk te implementeren en zijn daarom een goed startpunt voor het doen van voorspellingen. In tijdreeksen wordt er veelal onderscheid gemaakt tussen verschillende patronen in de historische gegevens, namelijk een level, trend, seizoen- en onvoorspelbare factor (*Chopra & Meindl, 2007)*.

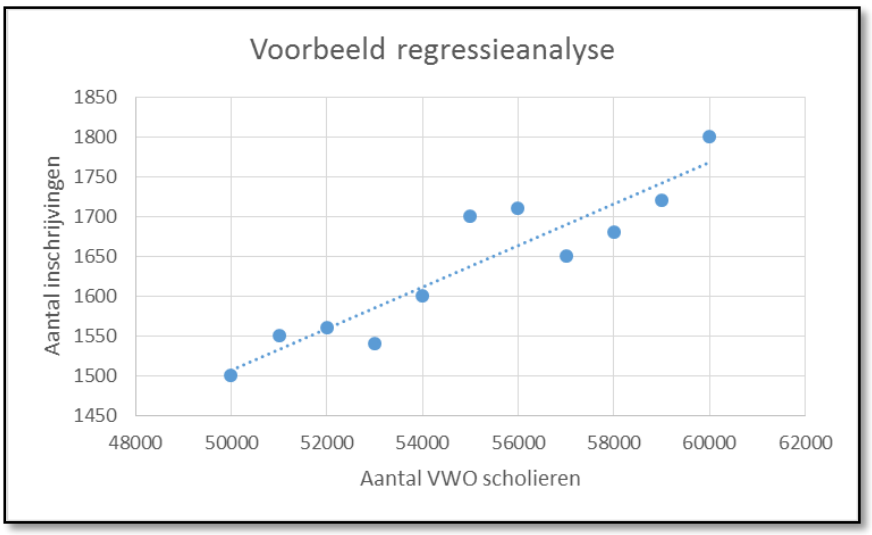


Figuur 2.6: Voorbeeld statistische tijdserievoorspelling in k0 voor de komende 3 jaar, gebaseerd op fictieve gegevens.

In Figuur 2.6 is een voorbeeld weergegeven van een statistische tijdserie voorspelling gemaakt in *k0*, voor de komende drie jaren. Het statisch voorspellen houdt in dat de gegevens voor de voorspelling niet worden aangepast gedurende de tijd en dus één keer berekend worden. Daarnaast kan ook op een adaptieve manier voorspeld worden, wat inhoudt dat het model telkens geüpdatet wordt op basis van de nieuwste gegevens. Het level is het gemiddelde uit de historische gegevens. De trend geeft een

bepaalde stijging weer die waar is genomen. De seizoen factor houdt rekening met een waargenomen fluctuatie per seizoen. De voorspelde waarde houdt rekening met het level, de trend en de seizoensinvloeden. Daarnaast zijn ook de werkelijke (gerealiseerde) waarden weergegeven. Het verschil tussen de werkelijke waarde en de voorspelde waarde is de onvoorspelbaarheid, oftewel de onnauwkeurigheid van de voorspelling. De bovenstaande figuur is gebaseerd op een voorspelling op basis van voorspelling = level * trend * seizoenfactor. Tijdserievoorspellingen kennen verschillende methodes en technieken, met elk hun eigen toepassing en invulling van de voorspelling (*Chopra & Meindl, 2007)*. In paragraaf 2.5.2 zullen we deze tijdseriemodellen verder behandelen.

Causale voorspellingmodellen maken gebruik van een afhankelijkheid tussen een afhankelijke, te voorspellen waarde *y* en een te bepalen onafhankelijke variabele *x.* Deze voorspellingsmodellen zijn dus toepasbaar wanneer er een verband bestaat tussen de te voorspellen waarde en een variabele in het omliggende milieu. Hierbij kan bijvoorbeeld gedacht worden dat de vraag naar de Open Dagen afhankelijk is van andere Open Dagen die plaatsvinden en/óf het aantal VWO inschrijvingen op de middelbare scholen in Nederland. De mate van afhankelijkheid kan bepaald worden via regressieanalyse.



Figuur 2.7: Voorbeeld regressieanalyse, gebaseerd op fictieve cijfers

In Figuur 2.7 is een voorbeeld weergegeven van een simpele vorm van enkelvoudige regressieanalyse en de afhankelijkheid tussen het aantal inschrijvingen op de y-as, tegenover het aantal Vwo scholieren op de x-as. Hieruit kan worden opgemaakt of er een verband bestaat tussen de twee variabelen en in wat voor mate. In paragraaf 2.5.1 zullen we verder ingaan op de verschillende causale voorspellingsmethoden.

Simulatie voorspellingsmodellen zijn gebaseerd op het imiteren (het simuleren) van het omliggende milieu waarin de voorspellingskwestie plaatsvindt. Doormiddel van het simuleren kan voorspeld worden wat de uitkomst zou zijn als er veranderingen plaatsvinden in het omliggende milieu. In de simulatie kunnen verschillende voorspellingsmodellen worden meegenomen, zoals causaal als tijdseries (*Chopra & Meindl, 2007)*. Simulatie kan bijvoorbeeld gebruikt worden om na te gaan wat het verwachte aantal inschrijvingen tijdens de Open Dagen zou zijn, als er tegelijktijdig ook andere universiteiten of onderwijsinstellingen open dag houdt. Het is hierbij van belang om de causale verbanden goed te onderzoeken en betrouwbaar weer te gegeven, anders zou de uitkomst hoge onnauwkeurigheid bevatten en dus weinig waard zijn.

2.2.3 Combineren van modellen

Naast de kwalitatieve en kwantitatieve voorspellingsmethoden, zijn er methoden beschikbaar die deze twee soorten methoden combineren tot één voorspelling. Uit meerdere studies is het naar voren gekomen dat samengevoegde voorspellingen een nauwkeuriger, en daarmee een betere, voorspelling opleveren ten opzichte van op zichzelf staande voorspellingen (*Chopra & Meindl, 2007).* Dit valt ook terug te zien in de drie wetten van voorspellen. Bij het samenvoegen van de voorspellingen worden ook de onnauwkeurigheden van de voorspellingen samengevoegd. Dit zorgt in theorie voor een kleinere onnauwkeurigheid. In paragraaf 2.6 zullen we verder ingaan op het onnauwkeurigheid principe.

Het combineren van voorspellingen kan op verschillende manieren. De meest simpele manier is het gemiddelde te nemen van de verschillende uitkomsten van de voorspellingen. Dit kan echter alleen als de verschillende uitkomsten even belangrijk/nauwkeurig worden geacht. Daarnaast kunnen gewichten aan de voorspellingen gekoppeld worden, om zodoende de voorspellingen samen te voegen tot één voorspelling. In de literatuur verschillen de manieren waarop de gewichten bepaald kunnen worden. Als de verschillende onderzoeken bekeken worden, blijkt dat de manier waarop de voorspellingen worden samengevoegd niet cruciaal te zijn voor de nauwkeurigheid. Het samenvoegen van de voorspellingen op basis van een simpel gewogen gemiddelde blijkt in de meeste gevallen een betere voorspelling op te leveren dan complexe methodes om de gewichten te bepalen (*Clemen, 1989).*
Het combineren van voorspellingen kan op verschillende manieren toegepast worden. Sommige voorspellingen kunnen onderverdeeld worden in sub voorspellingen, waarbij een samenvoeging hiervan de werkelijke voorspelling oplevert. Daarnaast kunnen voorspellingen met het zelfde doel worden gecombineerd, om zodoende de onnauwkeurigheid proberen te verminderen. Een voorbeeld hiervan is de “Pollyvote” voorspelling. Pollyvote voorspelt de uitkomsten van de verkiezingen in de Verenigde Staten. Ze maken hiervoor gebruik van de zogenaamde componenten methode, wat betekend dat hun voorspelling is opgebouwd uit verschillende componenten. Deze componenten zijn verschillende categorieën, die elk weer verschillende voorspellingen bevatten. Voor de samengestelde voorspelling, bepalen ze eerst het gemiddelde binnen elk component en daarna tussen de componenten. In 2012 voorspelde Pollyvote de uitslag van de Amerikaanse verkiezingen met 0,7% nauwkeurigheid, wat een van de nauwkeurigste voorspellingen was (*Graefe, Armstrong, Jones, Cuzán, 2013*). Het combineren van verschillende voorspellingen heeft in deze casus dus voor een duidelijke afname in onnauwkeurigheid gezorgd.

2.3 HET VOORSPELLINGSPROCES

De succes en nauwkeurigheid van een voorspelling is grotendeels te wijten of te danken aan het proces van de opstelling van de voorspelling (Ozcan, 2009). Een duidelijk gestructureerd proces vergroot de kans van succes en nauwkeurigheid. Chopra en Meindl beschrijven in hun boek “Supply Chain Management” een zes-stappen methode die gebruikt kan worden voor het efficiënt doen van voorspellingen en het succesvol implementeren ervan. Deze methode is toegespitst op voorspellingen in de supply chain, maar kan omgeschreven worden zodat de methode toepasbaar wordt voor dit onderzoek.

Stap 1: Begrijp het doel van de voorspelling

Voorspellingen kunnen gebruikt worden als leidraad voor bepaalde keuzes die gemaakt moeten worden. Keuzes die worden gemaakt op foutieve of niet relevante voorspellingen kunnen zodoende grote invloed hebben. Het is daarom van belang om duidelijk te hebben wat het doel van de voorspelling is en hier de voorspelling ook naar in te richten. Daarnaast is het ook relevant om iedereen, die in aanraking komt met keuzes die gemaakt worden op basis van voorspellingen, te informeren hierover. Als bijvoorbeeld de zaalcapaciteit wordt aangepast, op basis van een hoger aantal voorspelde bezoekers, dan is het ook handig om de betrokken personeelsleden hierover te informeren. Anders zou dit voor vervelende situaties kunnen zorgen.

Stap 2: Integreer de voorspelling door de gehele organisatie
Voorspellingen kunnen aan de basis staan voor bepaalde beslissingen die worden genomen in een organisatie. Echter, kan het voorkomen dat de voorspellingen relevant zijn voor meer onderdelen binnen de organisatie, dan waarvoor de voorspelling initieel bedoeld was. Als er bijvoorbeeld een grotere opkomst wordt verwacht, zal niet alleen dit door moeten worden gespeeld aan de zaalplanner, maar ook aan de parkeerwachten, personeelsplanning en dergelijke, zodat zij niet voor verassingen komen te staan en de capaciteit niet beschikbaar hebben (of juist overcapaciteit hebben, wat voor verspilling zorgt). Als dit niet het geval zal zijn, zou het kunnen voorkomen dat de klanttevredenheid van de bezoekers omlaag gaat en/of onveilige situaties zich kunnen voordoen.

Stap 3: Begrijp en identificeer de doelgroepen
De juiste klantsegmenten moeten worden meegenomen in de voorspellingen. Als de voorspelling op de verkeerde segmenten is focust, kan het voorkomen dat de voorspelling niks waard is. Het is hierbij van belang goed te identificeren wie de klanten zijn en wat de kenmerken van deze verschillende klanten zijn. Hierbij is het waarschijnlijk mogelijk om de klanten te categoriseren in verschillende doelgroepen, met elk hun eigen eigenschappen. Over het algemeen worden verschillende voorspellingsmethoden gebruikt voor verschillende doelgroepen. Een duidelijke identificatie van de doelgroepen maakt het mogelijk dat het voorspellingsmodel simpeler en nauwkeuriger wordt.

Stap 4: Identificeer de belangrijkste factoren die de vraag beïnvloeden
Voor een succesvolle voorspelling is het noodzakelijk om alle belangrijke factoren mee te nemen die van invloed kunnen zijn op de daadwerkelijke uitkomst. Hierbij moet bijvoorbeeld rekening gehouden worden met andere open dagen die gelijktijdig plaatsvinden, alsmede ook andere bijzondere omstandigheden (slecht weer kan ook grote invloed spelen, zoals bij de voorspelde opkomst bij verkiezingen (*Gatrell & Bierly, 2002*)). Uit de historische gegevens kunnen bepaalde eigenschappen afgelezen worden, zoals een stijgende of dalende opkomst (trends) en seizoensinvloeden. Ook kunnen promotieactiviteiten en activiteiten van concurrenten invloed hebben op de daadwerkelijke vraag. Deze factoren moeten worden meegenomen, zodoende de voorspelling niet op te positieve of negatieve gegevens wordt gebaseerd. Hierbij moet ook rekening gehouden worden met het uiteindelijke doel van de voorspelling. Als laatste moet gekeken worden naar verschillende factoren binnen de vraag die elkaar beïnvloeden. Hierbij kan gedacht worden aan opleidingen die erg op elkaar lijken en daardoor veel potentiele studiekiezers van elkaar weghouden. De voorspelling van deze opleidingen zou dan samen gedaan moeten worden.

Stap 5: Bepaal de meest geschikte voorspellingstechniek
De keuze voor de juiste voorspellingstechniek is belangrijk voor de betrouwbaarheid van de voorspelling. Het is hierbij van belang om eerst alle dimensies te begrijpen die relevant zijn voor de voorspelling. Deze dimensies kunnen bijvoorbeeld geografische data zijn, alsmede ook gesegmenteerde opleidingen, als specifieke doelgroepen. Iedere dimensie kan verschillende eigenschappen met zich meebrengen, zoals trends, bepaalde groei- en seizoensinvloeden. Voor de keuze voor de juiste techniek is het van belang om duidelijk te hebben wat de eigenschappen zijn van de verschillende vormen van vraag. Per dimensie kan een specifieke voorspellingstechniek gekozen worden, zoals beschreven in paragraaf 2.2. Een combinatie van technieken zorgt voor de hoogste nauwkeurigheid.

Stap 6: Bepaal de prestatie van de voorspelling en de onnauwkeurigheid
Voorspellingen kunnen altijd een mate van onnauwkeurigheid met zich meebrengen, zoals vermeld in de drie wetten van voorspellen. Voor de evaluatie van het voorspellingsmodel is het van belang om te kijken in welke mate deze onnauwkeurigheid voorkomt. Als het model ondermaats presteert, dus met een hoge onnauwkeurigheid, zou dit reden kunnen zijn om het model aan te passen.

2.4 DATA

Voorspellingen maken gebruik van een causaal verband tussen het verleden en de toekomst (*Makridakis & Wheelwright, 1989*). Gegevens uit het verleden worden zodoende gebruikt voor een projectie van de toekomst. Analyse van goede en correcte data uit het verleden is van belang om een nauwkeurig voorspellingsmodel op te stellen, zoals beschreven in stap 3 en 4 van het voorspellingsproces. In deze paragraaf zullen we de theoretische achtergrond geven voor het begrip data. Data kan op verschillende manieren worden weergegeven en kan globaal onderverdeeld worden in kwantitatieve data, kwalitatieve data en informatie (*Robinson, 2004*).

Kwantitatieve data heeft betrekking op data uitgedrukt in de vorm van getallen. In kwalitatieve data komen echter geen getallen voor, maar wordt data weergegeven in de vorm van afbeeldingen of woorden. Daarnaast is er nog data weergegeven in de vorm van informatie. Kort gezegd is informatie data met een betekenis. Informatie is de betekenis die de mens toekent aan gegevens (data), doormiddel van afspraken over de vorm waarin de gegevens worden gepresenteerd (*Robinson, 2004*).

2.4.1 Data functies

Naast de verschillende vormen waarop data weergegeven kan worden, heeft data binnen onderzoek ook verschillende functies. In het modelleren van (wiskunde) modellen kunnen deze data functies onderverdeeld worden in drie categorieën (*Pidd, 2003*). De eerste categorie is contextuele data. Deze data heeft tot doel een goede context te verlenen voor het onderzoek. Dit houdt in dat deze data onder andere de situatie en de problemen moet weergeven. Zodoende kan het onderzoeksterrein afgebakend worden en kan het onderzoek een duidelijke richting uit. Deze functie is meestal niet van toepassing op al te gedetailleerde data.

De volgende categorie datafunctie is realiserende data. Deze vorm van data heeft tot doel het werkelijke model op te stellen, voor het realiseren van de problemen die geïdentificeerd zijn. De data die dit tot doel heeft is meestal gedetailleerde data, zoals ratio's.

Als laatste categorie datafunctie is validerende data. Deze vorm van data heeft tot doel het opgestelde model te valideren en te controleren op nauwkeurigheid (*Robinson, 2004*). Data die dit tot doel heeft zijn bijvoorbeeld werkelijk behaalde waarden en de voorspelde waarden.

2.4.2 Data verzamelen

Nadat de verschillende functies van de data geïdentificeerd zijn, en alsmede ook geïdentificeerd is welke data noodzakelijk is voor het onderzoek, kan de data verzameld worden. Hierbij kan het voorkomen dat data direct beschikbaar is, of dat het verkregen moet worden. Globaal gesproken kan de beschikbaarheid van de data onderverdeeld worden in drie categorieën, namelijk (1) data die direct beschikbaar is, (2) data die niet direct beschikbaar is, maar wel te verzamelen en (3) data die niet direct beschikbaar is en ook niet te verzamelen (*Robinson, 2004*).

Categorie 1 data is direct beschikbaar, omdat deze data bekend is, of dat deze data al eerder verzameld is. Deze vorm van data verzamelen levert geen problemen op. Categorie 2 data moet verzameld worden. Hier kan bijvoorbeeld gedacht worden aan bepaalde opkomstcijfers. Het is hierbij wel van belang dat de data in de juiste vorm wordt verzameld en accuraat is. Als laatste is er categorie 3 data. Deze data is niet beschikbaar en valt ook niet te verzamelen. Dit kan voorkomen als iets onderzocht moet worden wat nog niet bestaat, of dat er simpelweg te weinig tijd is voor het verzamelen. Zo kan het voorkomen dat er wel data verzameld kan worden, maar de hoeveelheid data die verzameld kan worden te weinig is om een realistisch beeld te geven over de situatie en daardoor in categorie 3 terecht komt.

Er kan op verschillende manieren worden omgegaan met data die niet beschikbaar is en niet te verzamelen valt (categorie 3 data). De eerste mogelijkheid is om de data te schatten en de tweede mogelijkheid is om de data niet te behandelen als vaste waarden, maar als een experimentele factor. Data kan bijvoorbeeld geschat worden, door te kijken naar data uit vergelijkbare situaties. Ook kunnen interviews met experts geschatte data opleveren. Als laatste redmiddel kan er altijd nog een

intelligente gok gedaan worden. Het schatten van data verhoogd de onbetrouwbaarheid van de uitkomst en kan daardoor niet het gewenste resultaat opleveren. Deze onnauwkeurigheid kan gedeeltelijk weggenomen worden door verschillende bewerkingen toe te passen op de data. Een manier is een gevoeligheidsanalyse toepassen op de werkelijke uitkomst. Zo kan gekeken worden wat het effect is van veranderingen in de data.

Naast het schatten van data is het ook mogelijk om de data als een experimentele factor te behandelen. Bij het schatten van de data wordt afgevraagd wat de data zou zijn. Bij de experimentele factor wordt deze vraag omgedraaid en afgevraagd wat deze data zou moeten zijn. Zo kan vanuit een gewenste uitkomst gekeken worden wat de data zou moeten zijn.

Als het niet mogelijk is om via het schatten of de experimentele factor de data te bepalen, zijn er nog drie mogelijkheden hoe om te gaan met de gebrek aan data. De eerste mogelijkheid is het vooronderzoek te veranderen, zodat de noodzaak voor data uit het onderzoek wordt gehaald. De tweede mogelijkheid is de doelstellingen van het onderzoek aanpassen, zodat de data niet meer nodig is. De derde en laatste mogelijkheid is om niet door te gaan met het onderzoek.

2.4.3 Betrouwbaarheid van de data

Als er data verzameld kan worden óf al beschikbaar is, hoeft het niet meteen te betekenen dat de data ook juist is en meteen toepasbaar voor het onderzoek is. Hierbij moet onderzocht worden wat de betrouwbaarheid van de data is. Als eerste moet de bron van de data onderzocht worden. Hierbij moet gekeken worden of de bron betrouwbaar is. Ook zal gekeken moeten worden op welke manier de data verzameld is. Als dit op een onzorgvuldige manier gebeurd is, neemt de betrouwbaarheid van de data af. Ook is het interessant om af te vragen voor welke redenen de data verzameld was. Als de data voor hele andere reden verzameld is, dan het onderzoek in kwestie, kan het voorkomen dat de data helemaal niet toepasbaar is. Een manier om de data te controleren is het weergeven van de data in grafieken, om zodoende opvallende patronen te herkennen en te kunnen onderzoeken. In veel gevallen geeft de data een voorstelling weer van het verleden. Dit betekend niet meteen dat de data ook een goede indicator zou zijn voor de toekomst, doordat er andere factoren meespeelden of meegaan spelen.

Als de data te onnauwkeurig is, dan kan er op zoek gegaan worden naar een alternatieve databron of geprobeerd worden de data op te schonen. Als dat niet mogelijk is, dan kan de data dus niet gebruikt worden en valt deze data automatisch in categorie 3 (*Robinson, 2004*).

2.5 VOORSPELLINGSMODELLEN

In paragraaf 2.2 is een korte omschrijving gegeven van de verschillende voorspellingsmethoden en is globaal ingegaan op de verschillende technieken die deze methoden huisvesten. In deze paragraaf gaan we hier verderop in en gaan we de verschillende kwantitatieve voorspellingsmodellen uiteenzetten, waarmee we een selectie gegeven voor stap vijf van het voorspellingsproces; de keuze voor de juiste voorspellingstechniek.

2.5.1 Causale voorspellingsmodellen

Causale voorspellingsmodellen zijn gebaseerd op het idee dat de te voorspellen waarde *y* afhankelijk is van de onafhankelijke variabele *x*. Er zijn verschillende modellen beschikbaar die dit verband tussen twee (of meerdere) variabelen meenemen. In de basis zijn ze op hetzelfde principe gebaseerd, namelijk op regressie. Een manier om een lineaire afhankelijk tussen twee (of meerdere) variabelen te bepalen is via regressieanalyse. In een regressieanalyse wordt gekeken óf er een verband is tussen de verschillende variabelen. Als er een verband gevonden is, kan deze bijvoorbeeld worden omgezet in een lineaire formule in de vorm $y = a + bx$, waarbij *b* de afhankelijkheid van variabele *y* ten opzichte van variabele *x* weergeeft.

Een van de meest gebruikte voorspellingsmodellen voor de toepassing van de lineaire regressie, is een simpel wiskundig model in de vorm:

$$Y = b_0 + b_1X_1 + b_2X_2 + \cdots + b_mX_m \quad (2.1)$$

Hierbij is b_0 een constante en b_i de parameter die de mate van afhankelijkheid van variabele x_i weergeeft. Een sommatie van deze verschillende variabelen geeft uiteindelijk de voorspelling *Y* weer (*Hopp & Spearmann, 2001*).

De lineaire afhankelijkheid hoeft niet altijd perse een rechte lijn te zijn, zoals in het voorbeeld. Er zijn verschillende vormen van lineaire afhankelijkheid, zoals kwadratisch, exponentieel of volgens een inverse functie. De parameters zijn in deze vergelijkingen nog steeds lineair, waardoor nog steeds van een lineair model gesproken kan worden. Naast deze verschillende vormen van lineaire modellen, kan het ook voorkomen dat de afhankelijkheid niet lineair is. Over een niet lineaire afhankelijkheid kan gesproken worden als de parameters b_i niet lineair zijn. Om deze vorm van regressie te bepalen zijn verschillende methoden ontwikkeld, die we in dit onderzoek verder achterwege laten, tenzij we aanwijzingen tegenkomen die wijzen op een niet lineaire afhankelijkheid.

2.5.2 Tijdserievoorspellingsmodellen

Tijdseries hebben betrekking op verschillende datapunten op een continue tijdsinterval. Het voorspellen op basis van tijdseries kan op verschillende manieren. In het algemeen proberen tijdserievoorspellingen trends en afwijkingen uit het verleden vast te leggen, om deze te extrapoleren naar de toekomst (*Hopp & Spearmann, 2001).* Er zijn verschillende technieken beschikbaar voor het voorspellen op basis van tijdreeksen. Welke technieken toepasbaar zijn, hangt af van de eigenschappen van de tijdreeks. Deze eigenschappen kunnen onderverdeeld worden in de volgende punten:

- Het level (*L*);
- De trend (*T*);
- De seizoeninvloeden (*S*);
- Conjuncturele factor (*C*);
- Onregelmatigheid (*E*).

Het level geeft een bepaalde waarde aan die de tijdreeks heeft in periode *t*. De trend geeft een tendens weer van stijging of daling van dit level over een bepaald aantal perioden. De seizoeninvloeden hebben betrekking op een bepaalde cyclische afwijking in de vraag in het jaar. Hierbij kan gedacht worden aan een hogere vraag in de zomer naar ijsjes, en in de winter een kleinere vraag (*Chopra & Meindl, 2007)*. De conjuncturele factoren zijn vergelijkbaar met de seizoeninvloeden, echter hebben de conjuncturele factoren betrekking op cyclische fluctuaties op de lange termijn. Hierbij kan gedacht worden aan fluctuaties in de economie, zoals dat (meestal) na een recessie weer een periode van economische bloei aanvangt. De onregelmatigheden zijn de waarden in de tijdserie die niet verklaard kunnen worden door de bovengenoemde factoren. Dit kan ook gezien worden als de error of onvoorspelbaarheid van de tijdreeks.

Een tijdreeks kan enkele van de bovengenoemde eigenschappen bevatten, of allemaal. De manier hoe de tijdreeks is opgebouwd uit deze componenten kan verschillen. Op een multiplicatieve manier zijn de waarden uit de tijdreeks opgebouwd volgens het vermenigvuldigen van de verschillende componenten. Een adaptieve manier maakt niet gebruik van het vermenigvuldigen van de componenten, maar het simpel bij elkaar optellen ervan. Daarnaast kan de tijdreekswaarden ook opgebouwd zijn uit een mixed model. Hierbij kan bijvoorbeeld het level en de trend bij elkaar opgeteld worden en dan vermenigvuldigd worden met de seizoenfactor en eventueel met de conjuncturele factor (*Chopra & Meindl, 2007)*.

Het al of dan niet bevatten van verschillende eigenschappen, maakt het mogelijk dat sommige technieken wel toepasbaar zijn voor de voorspelling op basis van tijdreeksen en sommige technieken

niet. Een simpele manier om na te gaan of een tijdreeks een trend bevat is het bepalen van de zogenaamde eerste verschillen, in de vorm van:

$$W_t = f_t - f_{t-1} \tag{2.2}$$

Het enigste wat de formule doet is het verschil weergeven tussen de waarde in periode *t* minus de waarde in periode *t-1*. Als de som van de eerste verschillen in de tijdreeks een gemiddelde van 0 heeft én een horizontale reeks vormt, dan is er geen trend aanwezig in de tijdreeks. Mocht het gemiddelde niet gelijk zijn aan 0, maar wel een horizontale reeks vormen, dan is er sprake van een trend in de tijdreeks. Als het gemiddelde groter dan 0 is, dan kan gesproken worden over een positieve trend en visa versa. Als er geen trend aanwezig is in de tijdreeks, dan kan gesproken worden over een stationaire tijdreeks. De tijdreeks moet dan standaardnormaal verdeeld zijn, wat betekend dat de waarden in de tijdreeks rond een gemiddelde bevinden, met een bepaalde standaardafwijking. Bij het aanwezig zijn van een trend verplaatst dit gemiddelde, waardoor dus niet over een stationaire tijdreeks gesproken kan worden. Sommige voorspellingstechnieken zijn alleen toepasbaar voor stationaire tijdreeksen. Echter, kunnen niet stationaire tijdreeksen in sommige gevallen stationair gemaakt worden. In de eerste verschillen is gekeken of er een trend aanwezig is in de tijdreeks. Mocht dit het geval zijn, dan kunnen deze eerste verschillen gebruikt worden om de tweede verschillen van de tijdreeks te bepalen, via de volgende formule:

$$Z_t = W_t - W_{t-1} \tag{2.3}$$

Als de eerste verschillen van periode *t* minus de eerste verschillen van periode *t-1* wordt gedaan, kan een reeks verkregen worden die de tweede verschillen genoemd wordt. In veel gevallen is de tijdreeks nu wel stationair (*Doorn, 2006*). Het stationair proberen te maken van een tijdreeks op deze manier wordt ook wel het differentiëren van de tijdreeks genoemd (*Wiener, 1949*).

Het ARIMA voorspellingsmodel maakt gebruik van deze manier van differentiëren om de tijdreeks stationair te maken. Het AutoRegressive Integrated Moving Average model, zoals het ARIMA model voluit heet, is vergelijkbaar met een ARMA model. Echter, is het ARMA model alleen toepasbaar voor stationaire tijdreeksen. Het ARIMA(p,d,q) model maakt gebruik van het d-keer differentiëren van de tijdreeks, om de tijdreeks stationair te maken. Zodra de tijdreeks stationair is, wordt precies dezelfde techniek toegepast als het ARMA(p,q) model. Het ARMA(p,q) en het ARIMA (p,d,q) model maken gebruik van het voorspellen op basis van twee verschillende componenten, namelijk een autoregressief (AR) gedeelte én een moving average (MA) gedeelte (*Zhang, 2001*).

Het AR(p) model is gebaseerd op het idee dat de afhankelijk variabele *Y* (de te voorspellen variabele) regressief is met dezelfde afhankelijke variabele *Y* p-perioden eerder. In andere woorden betekend dit dat de variabele samenhang heeft met zichzelf. Dit model kan dus worden toegepast als er een verband is tussen de afhankelijke variabele én dezelfde afhankelijke variabele enkele perioden eerder. De formule voor het AR(*p*) model is als volgt:

$$(Y(t) - \mu) = \alpha 1(Y(t-1) - \delta) + \alpha 2(Y(t-2) + \cdots + \alpha p(Y(t-p) - \delta) + \varepsilon \tag{2.4}$$

Met:
$Y(t) =$ De te voorspellen waarde Y in periode t
$\mu =$ Gemiddelde van Y (constant)
$\varepsilon =$ Gemiddelde afwijking met constante variatie (constant)
$\alpha =$ Wegingsfactor

In een AR(p) model wordt dus de variabele *Y* voorspeld in periode *t* door de afhankelijkheid van deze variabele met de waarden van *Y* over de afgelopen p-perioden.

Het MA(q) model is gebaseerd op het idee dat de te voorspellen waarde Y de som is van een constante waarde en een bewegend gemiddelde, volgens de volgende formule:

$$Y(t) = \mu + \epsilon(t) + \beta(0)\varepsilon(t) + \beta(1)\varepsilon(t-1) + \cdots + \beta(q)\varepsilon(t-q) \quad (2.5)$$

Met;
$\mu =$ constante waarde
$\varepsilon =$ Gemiddelde afwijking met constante variatie
$\beta =$ Wegingsfactor

De te voorspellen waarde *Y* in periode *t* is dus gelijk aan een constante factor en een constante afwijking plus de gewogen afwijkingen in q-perioden.

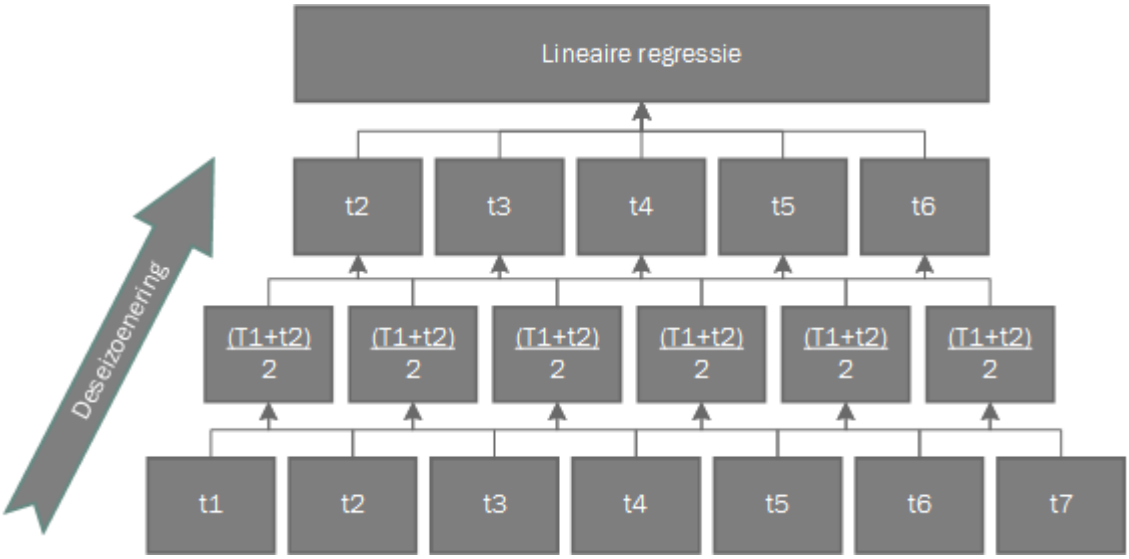
Het ARMA model kan toegepast worden als een tijdreeks stationair is (of *d-keer* gedifferentieerd is om de reeks stationair te maken) én zowel autoregressief is áls voorspeld kan worden op basis van moving average. Het ARMA(p,q) model is een samenvoeging van het AR(p) model en het MA(q) model. Een ARMA (1,1) model kan dus volgens de volgende formule worden weergegeven.

$$Y(t) = \theta + \alpha_1 Y_{t-1} + \beta_0 \varepsilon_t + \beta_1 \varepsilon_{t-1} \quad (2.6)$$

In de bovenstaande formule wordt de te voorspellen waarde *Y* in periode *t* voorspeld door een autoregressie met periode *t-1* en een moving average met periode *t-1.*

Naast het voorspellen op basis van stationaire tijdreeksen, zijn er ook voorspellingstechnieken beschikbaar die gebruik maken van andere eigenschappen van de tijdreeks. Aan het begin van deze paragraaf is beschreven dat een tijdreeks verschillende eigenschappen kan bevatten, namelijk een level, trend, seizoeninvloed, conjuncturele factor en een onnauwkeurigheid. Deze factoren kunnen op verschillende manieren bepaald worden.

Het level heeft betrekking op de waarde die de tijdreeks heeft in een bepaalde periode *t*. De trend geeft weer of er een stijgende of dalende tendens is in dit level. Om de trend van de tijdreeks te bepalen, moet eerst bekeken worden of er conjuncturele- of seizoeninvloeden aanwezig zijn in de tijdreeks. Als de tijdreeks grafisch wordt weergegeven, kan aangenomen worden dat er seizoeninvloeden aanwezig zijn als er cyclische pieken en dalen voorkomen in de grafiek. Deze invloeden kunnen uit de data gehaald worden, door het zogenaamd deseizoeneren van de data. Het deseizoneren van data kan gedaan worden door de gemiddelde waarden te nemen tussen twee datapunten. Als er een even aantal seizoenen zijn, moet het gemiddelde genomen worden van twee berekende gemiddelden om de seizoeninvloed uit de vraag te halen, anders zou de "gedeseizoeneerde" vraag betrekking hebben op de vraag tussen twee datapunten (*Chopra & Meindl, 2007)* (Figuur 2.8*)*.



Figuur 2.8: Deseizoenering van datapunten

Als de data gedeseizoeneerd is, kan een lineaire regressie worden toegepast om de trend in de tijdreeks te bepalen. De seizoeninvloed per seizoen kan dan bepaald worden door de ratio te nemen van de werkelijke waarde per seizoen en de gedeseizoeneerde waarde per seizoen, volgens de volgende formule:

$$\overline{S_t} = \frac{D_i}{\overline{D_t}} \tag{2.7}$$

Deze verschillende eigenschappen van de tijdreeks kunnen op verschillende manieren gebruikt worden voor het doen van voorspellingen. De eerste mogelijkheid is een statistische manier van voorspellen. Hierbij wordt één keer de eigenschappen van de tijdreeks bepaald en deze waarden worden dan gebruikt voor het voorspellen van alle toekomstige waarden. Daarnaast kan gebruik worden gemaakt van een adaptieve manier van voorspellen. In deze manier van voorspellen worden de eigenschappen van de tijdreeks telkens aangepast, op basis van de gerealiseerde waarden. Er zijn hier een aantal methoden voor ontwikkeld, die elk specifieke eigenschappen van de tijdreeks meenemen in de adaptieve manier van voorspellen (*Chopra & Meindl, 2007)* (Tabel 2.2).

Soort extrapolatiemodel:	**Heeft betrekking op:**
Simple, moving and weighted average	Level
Simple exponential smoothing	Level
Holt's model	Level en trend
Winter's model	Level, trend en seizoeninvloed

Tabel 2.2: Adaptieve tijdserievoorspellingsmodellen

Simple, weighted and moving average methoden zijn een van de meest gebruikte voorspellingsmethoden (*Winston, 2004)*. Deze methoden voorspellen op basis van een bepaald level, namelijk het gemiddelde van de afgelopen perioden.

De simple average methode voorspelt de toekomstige vraag in de opvolgende periode *t* als het gemiddelde van de afgelopen perioden *m*. Bij deze methode van voorspellen wordt alleen rekening gehouden met het gemiddelde van de afgelopen perioden en dus niet met eventuele trends of seizoeninvloeden. De simple average kan volgens de volgende wiskundige formule voorspeld worden:

$$F(t+1) = \frac{\sum_{i=1}^{t} A(i)}{t} \tag{2.8}$$

Waarbij,

F(t+1) = De voorspelde waarde voor periode t+1;
A(i) = De feitelijke waarde uit periode i;

De simple average methode gaat ervanuit dat alle waarden uit het verleden even representatief zijn voor de toekomst. Data uit een ververleden heeft zodoende evenveel invloed, als data van één periode verleden. In de weighted average methode kan rekening gehouden worden met het verschuiven van belangrijkheid van de data. Dit kan gedaan worden door gewichten *w* te hangen aan de uitkomsten uit periode *i.* Dit kan worden door formule 2.8 aan te passen naar de vorm:

$$F(t+1) = \frac{\sum_{i=t}^{t} A(i) * W(i)}{\sum W(i)} \tag{2.9}$$

Waarbij,

F(t+1) = De voorspelde waarde voor periode t;
A(i) = De feitelijke waarde uit periode i;
W(i) = Weging w voor periode i

In de formule van de weighted average wordt ervanuit gegaan dat alle data nog een bepaald gewicht bijdragen aan de te voorspellen waarde. Data kan ook uitgesloten worden uit het voorspellingsproces. Het gewicht dat dan aan de data wordt gegeven is gelijk aan 0, waardoor de data dus geen invloed meer heeft in de voorspelling. Als de gewichten van *m* periodes uit een ververleden op 0 wordt gezet en de rest van de waarden op 1, dan komt de weighted average methode overeen met de moving average methode, met *t-m* perioden. Het uitsluiten van veel data uit het verleden, betekend dat data uit het heden een grotere invloed heeft op de voorspelling. De voorspelling wordt hierdoor minder stabiel (*Hopp & Spearmann, 2001*).

(simple) Expontential smoothing methode is een uitbereiding van de formule van de simple average methode. Deze methode baseert de voorspelling ook alleen nog maar op het level en gaat er dus vanuit dat er geen trend of seizoeninvloeden aanwezig zijn. In de simple exponential smoothing methode worden naast de gerealiseerde waarden, ook de voorspelde waarden voor de bijbehorende perioden meegenomen in de voorspellen van periode *t+1,* volgens de volgende formule:

$$F(t+1) = \alpha D(t) + (1-\alpha)F(t+1) \tag{2.10}$$

Waarbij,

F(t+1) = De voorspelde waarde voor periode *t+1*;
D(t) = De werkelijke waarde in periode t
α = Smoothing constante

De weging α wordt de smoothing constante genoemd en is een getal tussen de 0 en 1. Deze α kan gezien worden als het gewicht dat gegeven wordt aan de gerealiseerde waarde, ten opzichte van de voorspelde waarde. Een hoge α betekend dat het model sneller reageert op veranderingen in de werkelijke waarde, immers een hoge α betekend dat de werkelijke waarde een grotere invloed heeft op de te voorspellen waarde ten opzichte van de voorspelde waarde. Een lage α betekend dat werkelijke vraag weinig invloed heeft op de te voorspellen waarde, waardoor het model stabieler wordt. Doordat in het model geen trends en seizoeninvloeden worden meegenomen, kan een eventuele trend in de historische waarden ervoor zorgen dat de α te veel wordt meegenomen, of juist te weinig. Het is daarom van belang goed de historische data te onderzoeken op trends, voordat deze methode nauwkeurig gebruikt kan worden (*Hopp & Spearmann, 2001*).

Holt's Methode: Expotentiele smoothing met een lineaire trend is vergelijkbaar met de simple exponential smoothing methode, echter wordt in deze methode wel rekening gehouden met de aanwezigheid van trends in de historische data. De te voorspellen waarde wordt hierbij opgebouwd uit de sommatie van de voorspelde waarde en de voorspelde trend van een periode eerder, volgens de volgende formules:

$$F(t+1) = L(t) + T(t) \quad (2.11)$$
$$F(t+n) = L(t) + nT(t) \quad (2.12)$$
$$L(t+1) = \alpha D(t+1) + (1-\alpha)(L(t) + T(t)) \quad (2.13)$$
$$T(t+1) = \beta(L(t+1) - L(t)) + (1-\beta)T(t) \quad (2.14)$$

Waarbij,

F(t+1) = De voorspelde waarde voor periode *t+1*;
D(t) = De werkelijke waarde in periode t;
L(t) = Het level in periode *t;*
T(t) = De trend in periode *t*;
α = Smoothing constante voor het level
β = Smoothing constante voor de trend

Zowel bij het bepalen van het level en de trend voor de nieuwe periode (*t+1*) wordt iedere keer het gewogen gemiddelde genomen van de werkelijke waarde en de oude voorspelling. Een hoge β betekend dat een kleine toename of afname in het level (trend) zwaar wordt meewogen in het bepalen van de trend voor de nieuwe periode (*Chopra & Meindl, 2007*).

Winters' methode is vergelijkbaar met Holt's methode, maar naast de trends worden in deze methode ook seizoensinvloeden meegenomen in de te voorspellen waarde. Winters (1960) heeft de seizoensinvloeden in de voorspellingtechniek weten in te bouwen op de volgende manier:

$$F(t+1) = \big(L(t) + T(t)\big)S(t+1) \quad (2.15)$$

$$F(T+l) = \big(L(t) + lT(t)\big)S(t+l) \quad (2.16)$$

$$L(t+1) = \alpha\frac{D(t+1)}{S(t+1)} + (1-\alpha)(L(t) + T(t)) \quad (2.17)$$

$$T(t+1) = \beta(L(t+1) - L(t)) + (1-\beta)T(t) \quad (2.18)$$

$$S(t+p+1) = \gamma\frac{D(t+1)}{L(t+1)} + (1-\gamma)S(t+1) \quad (2.19)$$

Waarbij,

F(t+1) = De voorspelde waarde voor periode *t+1*;
D(t) = De werkelijke waarde in periode t;
L(t) = Het level in periode *t;*
T(t) = De trend in periode *t*;
S(t) = De seizoeninvloed in periode t;
α = Smoothing constante voor het level;
β = Smoothing constante voor de trend;
γ = Smoothing constante voor de seizoeninvloed;

Deze methode werkt verder op dezelfde manier als de Holt's methode (*Chopra & Meindl, 2007).*

De hierboven uiteengezette voorspellingstechnieken zijn slechts een klein gedeelte van de voorspellingstechnieken die beschikbaar zijn in de literatuur. De onderzochte voorspellingstechnieken hebben we afgestemd met de onderzoeksdoelstelling en we achten het dan ook waarschijnlijk dat de onderzochte technieken voldoende zijn voor het behalen van de onderzoeksdoelstelling.

2.6 NAUWKEURIGHEID

Voorspellingen zijn nooit correct en hebben altijd een mate van onnauwkeurigheid. Deze onnauwkeurigheid kan niet verklaard worden door historische gegevens en wordt daarom ook wel het "random" onderdeel van de voorspelling genoemd. Naast het random onderdeel, bestaat de voorspelling uit een component wat wel te herleiden valt naar historische gegevens. Deze component wordt het systematische onderdeel van de voorspelling genoemd. Een voorspelling is zodoende opgebouwd uit de volgende onderdelen:

Gerealiseerde waarde = Systematisch onderdeel (de voorspelling) + Random onderdeel (de error).

Het systematisch onderdeel, oftewel de voorspelling, geeft de projectie van het verleden weer voor de aankomende periode. Dit onderdeel kan bijvoorbeeld rekening houden met groeipercentages uit het verleden, alsmede ook seizoensinvloeden. De samenstelling van het systematisch onderdeel kan verschillen en valt te herleiden uit de historische gegevens. Voor een goede voorspelling moet dit systematische onderdeel onderzocht worden en de verschillende invloedrijke factoren geïdentificeerd worden, om zodoende het juiste voorspellingsmodel hiervoor te kiezen. Het voorspellingsmodel heeft veel invloed op de nauwkeurigheid van de voorspelling en daarmee op de bruikbaarheid ervan.

Het random onderdeel van de voorspelling is de mate waarin het systematische onderdeel (de voorspelling) afwijkt van de gerealiseerde waarde. Dit random onderdeel wordt ook wel de afwijking of error genoemd. Deze afwijking is lastig te bepalen en wil je in een voorspelling zo klein mogelijk hebben (je wilt immers de voorspelling zo correct mogelijk hebben). Deze mate van afwijking kan bepaald worden doormiddel van de standaardafwijking van de voorspelling. Een goede voorspelling

heeft een standaardafwijking die gelijk is aan het random onderdeel van de voorspelling. Het doel van voorspellen is de random component, de afwijking, uit de historische gegevens te filteren, zodat de voorspellingen alleen gebaseerd kunnen worden op het systematische onderdeel en niet op de afwijking (*Chopra & Meindl, 2007).* Naast de bovengenoemde reden waarom het bepalen van de onnauwkeurigheid uit historische gegevens belangrijk is, zegt de onnauwkeurigheid ook wat over de voorspelling zelf:

1. Op basis van de onnauwkeurigheid kan bekeken worden of de nauwkeurigheid van de voorspelling binnen de marges valt. Als de onnauwkeurigheid sterk afwijkt van de behaalde resultaten uit het verleden met het voorspellingsmodel, dan kan het een aanwijzing zijn dat de methode moet worden bijgesteld.
2. Voorspellingen kunnen de basis zijn voor veel verschillende keuzes die worden gemaakt, door een gehele organisatie. Het is daarom van belang de onnauwkeurigheid goed te bepalen en dit ook bij de voorspelling te melden, zodat iedereen die hierop keuzes baseert ook hier rekening mee kan houden. Een eventuele plotselinge stijging van de onnauwkeurigheid kan zodoende voor grote problemen zorgen. Daarom is het van belang dat de onnauwkeurigheid binnen de marges blijft.

Het bepalen van de nauwkeurigheid, de random component van de voorspelling, is daarom belangrijk. Dit kan gedaan worden door het verschil te bepalen tussen de voorspelde waarde en de werkelijke waarde, zoals weergegeven in de volgende formule:

$$E_t = F_t - D_t \tag{2.20}$$

Waarbij,

$E_t = Voorspellingserror$
$F_t = Voorspelde\ waarde$
$D_t = Werkelijke\ waarde$

Naast het bepalen van de nauwkeurigheid van een voorspelling, kan de nauwkeurigheid van de voorspellingsmethode op verschillende manieren gecontroleerd worden. Een van die manieren is via de Mean Absolute Deviation (MAD), zoals weergegeven in de volgende formule:

$$MAD_n = \frac{1}{n}\sum_{t=1}^{n} A_t \tag{2.21}$$

De MAD geeft de gemiddelde absolute afwijkingen weer van een reeks van voorspellingen. De absolute onnauwkeurigheden per voorspellingen kunnen worden bepaald doormiddel van de volgende formule:

$$A_t = |E_t| \tag{2.22}$$

Daarnaast kan de MAD gebruikt worden om de standaard verdeling van de afwijking te bepalen, hierbij wordt wel uitgegaan dat de afwijking standaard normaal verdeeld is, met een gemiddelde afwijking van 0. De deviatie kan volgens de volgende formule berekend worden:

$$\sigma = 1.25 * MAD \tag{2.23}$$

Volgens de Mean Absolute Percentage Error (MAPE) kan de gemiddelde afwijking als het percentage van de gerealiseerde waarde berekend worden. Hierbij wordt uitgegaan dat de afwijking standaardnormaal verdeeld is. De MAPE kan volgens de volgende formule berekend worden:

$$MAPE_n = \frac{\sum_{t=1}^{n}\left|\frac{E_t}{D_t}\right| * 100}{n} \tag{2.24}$$

Om de onnauwkeurigheid beter in beeld te brengen en te controleren of de voorspellingsmethode boven- of ondermaats presteert, kan de bias uitgerekend worden. In de bias wordt de som genomen van alle afwijkingen, zoals weergegeven in de volgende formule:

$$Bias_n = \sum_{t=1}^{n} E_t \tag{2.25}$$

In de ideale situatie zou de uitkomst van de bias rond de 0 moeten zijn. Dit betekend dat de afwijking compleet random en niet vooringenomen is. De Tracking Signal (TS) is de ratio van de bias ten opzichte van de MAD, zoals weergegeven in de volgend formule:

$$TS_t = \frac{bias_t}{MAD_t} \tag{2.26}$$

Als de TS groter is dan 6 of kleiner is dan -6, betekend dat de voorspellingsmethode een bias heeft. Dit kan een positieve bias zijn (TS > 6), of een negatieve bias (TS < -6). Als de voorspellingsmethode een bias heeft, kan er dus gekozen worden om een andere voorspellingsmethode te kiezen of de methode grondig te onderzoeken. Dit kan bijvoorbeeld aangeven dat er een nieuwe trend aanwezig is in historische gegevens, waardoor er telkens te laag wordt voorspeld, of te hoog. De trend moet dan geïdentificeerd worden en het voorspellingsmodel aangepast worden (*Chopra & Meindl, 2007*).

2.7 CONCLUSIE

Met behulp van de verschillende sub deelvragen hebben we in dit hoofdstuk bepaald hoe efficiënt voorspeld kan worden. In paragraaf 2.1 hebben we onderzocht wat voorspellen precies is. Hieruit blijkt dat voorspellen samengevat kan worden als een projectie van de toekomst (*Heroman, Davis & Farmer, 2012*). In paragraaf 2.2 en 2.5 zijn we ingegaan op de verschillende technieken en methoden die toepasbaar zijn voor het doen van voorspellingen. Hieruit blijkt dat er verschillende methoden beschikbaar zijn, met elk een specifieke toepassing. Globaal kunnen de methoden onderverdeeld worden in kwalitatieve en kwantitatieve modellen. Kwalitatieve modellen maken gebruik van menselijke input en deze zullen we verder achterwege laten in het verslag. Kwantitatieve modellen maken gebruik van het wiskundig bewerken van gegevens. Hier bestaan verschillende technieken voor, die elk een specifieke toepassing hebben, gebaseerd op de eigenschappen die een tijdreeks heeft. In paragraaf 2.3 zijn we ingegaan op het proces van voorspellen. Hierin hebben we een zes stappen model beschreven (*Chopra & Meindl, 2007)*, die gebruikt kan worden voor het efficiënt doen van voorspellingen. In de paragraaf die daarop volgt hebben we de invloed van betrouwbare data onderzocht, met betrekking op het doen van voorspellingen. In paragraaf 2.6 hebben we de verschillende technieken uiteengezet die bruikbaar zijn voor het bepalen van de nauwkeurigheid van voorspellingen en de noodzaak hiervan.

We kunnen uit de bovenstaande paragrafen concluderen dat efficiënt voorspeld kan worden doormiddel van zes stappen proces (paragraaf 2.3). De eerste twee stappen hebben betrekking op het doel van de voorspelling. Stap 1 hebben we in hoofdstuk 1 opgesteld en stap 2 zullen we in dit onderzoek verder achterwege laten, gezien dit voornamelijk betrekking heeft op de implementatie binnen de organisatie. Stappen 3 en 4 van het proces hebben betrekking op de verschillende eigenschappen van de data. Deze gaan we in het volgende hoofdstuk behandelen. Stap 5 en 6 hebben betrekking op de keuze voor de juiste voorspellingtechnieken en het bepalen van de betrouwbaarheid hiervan. In paragraaf 2.2, 2.5 en 2.6 zijn we hierop ingegaan en deze theorie gaan we toepassen in hoofdstuk 4 en 5.

Hoofdstuk 3

Inventarisatie historische data

(Deelvraag 2)

“Nobody should try to use data unless he has collected data.”

William Edwards Deming (1900 – 1993)

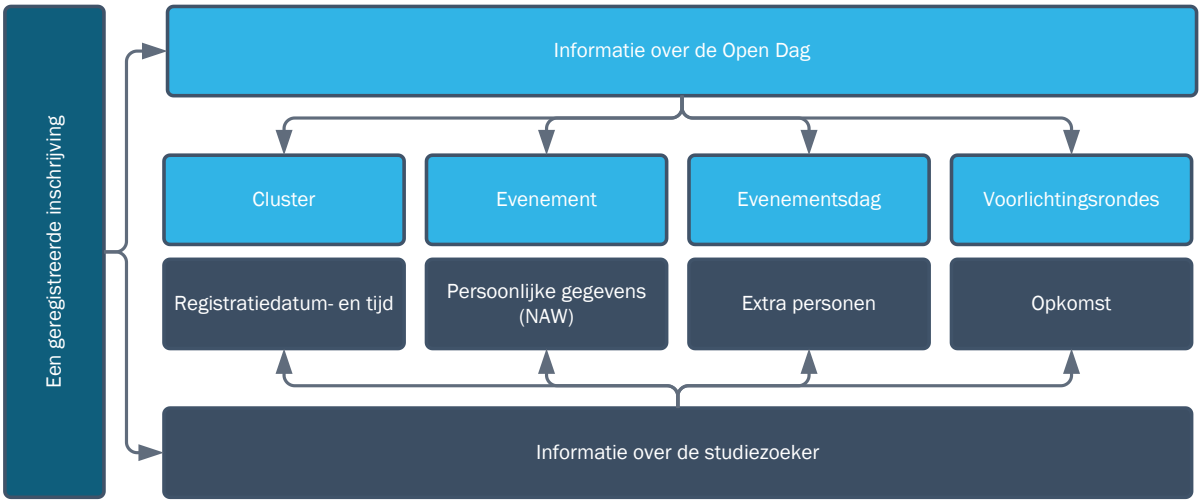
Amerikaans statisticus

3 INVENTARISATIE HISTORISCHE DATA

Dit hoofdstuk geeft de inventarisatie weer van historische data die beschikbaar zijn voor dit onderzoek. In het vorige hoofdstuk is een zes stappenplan opgesteld voor het efficiënt doen van voorspellingen. Stap drie en vier van dit proces heeft betrekking op het verzamelen en analyseren van de noodzakelijke data. De eerste paragraaf van dit hoofdstuk gaat in op de data die noodzakelijk zijn voor het behalen van de onderzoeksdoelstelling. De paragrafen die daarop volgen gaan opeenvolgend in op het analyseren van de data, aan de hand van stap drie en vier van het voorspellingsproces. Als eerst zal een analyse worden toegepast met betrekking op het aantal aanmeldingen, waarna vervolgens de data met betrekking tot de werkelijke opkomst geanalyseerd gaat worden. Dit hoofdstuk wordt afgesloten met de beantwoording van deelvraag 2.

3.1 DATA VOOR HET AANTAL AANMELDINGEN

Betrouwbare en volledige data zijn van belang voor de betrouwbaarheid en nauwkeurigheid van de voorspelling. Wanneer een studiezoeker zich registreert voor de Open Dagen, wordt er relevante informatie doorgespeeld naar de organiserende afdeling binnen de universiteit. Deze informatie heeft betrekking op onder andere persoonlijke gegevens, maar ook op de specifieke dag waarvoor ingeschreven is en de geselecteerde voorlichtingsrondes. Globaal kan deze data worden onderverdeeld in algemene informatie met betrekking tot de Open Dag waarvoor aangemeld is, en specifieke informatie met betrekking op de studiekiezer die zich aangemeld heeft (Figuur 3.1).



Figuur 3.1: Beschikbare gegevens over de Open Dagen

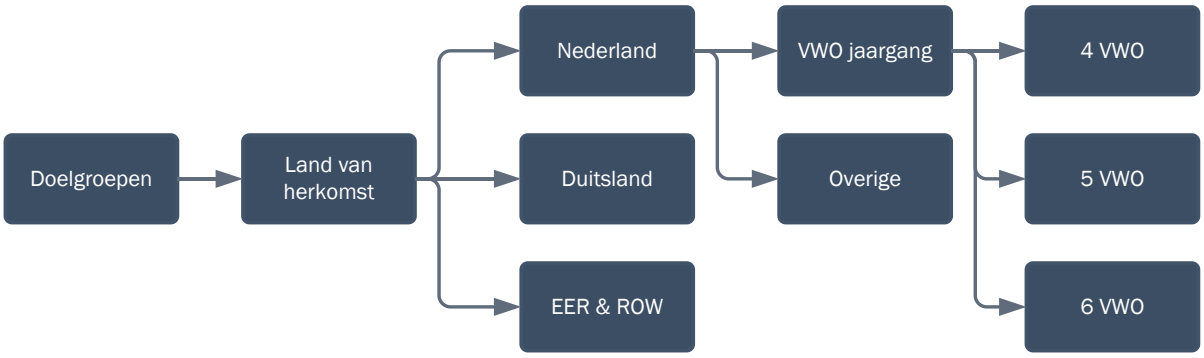
De informatie weergegeven in Figuur 3.1 wordt vanaf het najaar van 2012 bijgehouden in een centrale database binnen de organisatie. Deze gegevens worden momenteel gebruikt voor onder andere de evaluatie van de Open Dagen en het waarnemen van specifieke trends, uitschieters en achterblijvers. Naast deze dataset met uitgebreide informatie over het registratieverloop van de Open Dagen vanaf het najaar 2012, hebben we voor dit onderzoek ook toegang tot minder uitgebreide datasets over de jaren 2009, 2010 en 2011. Deze data heeft voornamelijk betrekking op een sommatie van het aantal registraties.

De beschreven datasets zijn afkomstig van M&C. We achten deze informatie dan ook betrouwbaar genoeg om te gebruiken voor het onderzoek. Ons onderzoek tot doel het verwachte aantal registraties en opkomst te voorspellen. Gezien de data die in de datasets beschikbaar zijn, achten we deze datasets een goede uitvalsbasis voor het onderzoek.

3.2 BELANGRIJKSTE DOELGROEPEN BINNEN DE OPEN DAGEN

Het van Dale woordenboek beschrijft een doelgroep als “degenen die je bij een bepaalde bezigheid op het oog hebt.”(*van Dale, 2016)*. Met andere woorden kan een doelgroep vertaald worden naar een specifieke groep mensen die de organisatie van de Open Dagen wil bereiken. De organisatie heeft het uiteindelijk doel met het bereiken van deze doelgroepen de instroom van de universiteit te versterken. Een doelgroep bestaat meestal uit een aantal gemeenschappelijke kenmerken, zoals leeftijd, geslacht, opleidingsniveau en/of een combinatie hiervan. Naast dat het identificeren van de doelgroepen belangrijk is voor de marketing van de Open Dagen, is het ook belangrijk voor het voorspellingsproces. In stap drie van het voorspellingsproces wordt beschreven dat de juiste doelgroepen cruciaal zijn voor een betrouwbare voorspelling. Door het identificeren van de juiste doelgroepen, kunnen de juiste voorspellingstechnieken gekozen worden, waardoor de betrouwbaarheid van de voorspelling kan toenemen.

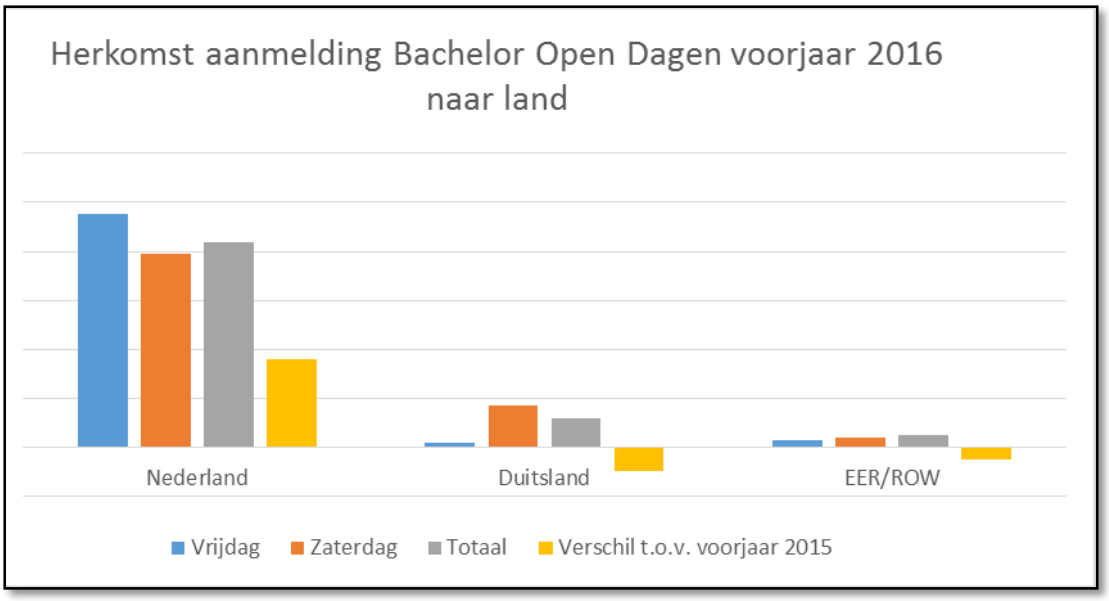
Om de juiste doelgroepen voor het voorspellingsproces te identificeren, hebben we gekeken naar de doelgroepen die de organisatie van de Open Dagen voor ogen heeft. In de evaluaties van de Open Dagen wordt qua doelgroepen onderscheid gemaakt tussen de verschillende landen waaruit de aanmeldingen binnenkomen. Er wordt hierbij onderscheid gemaakt in aanmeldingen uit Nederland, Duitsland en EER/ROW (Europese Economische Ruimte/Rest van de wereld). De aanmeldingen uit Duitsland en EER/ROW worden niet verder onderverdeeld in specifieke doelgroepen. De aanmeldingen uit Nederland worden wel verder onderverdeeld, namelijk in aanmeldingen uit specifieke vooropleidingsjaren. Hierbij wordt voornamelijk aandacht besteed aan aanmeldingen die vanuit de verschillende jaren van het VWO komen (Figuur 3.2).



Figuur 3.2: Onderverdeling van de doelgroepen

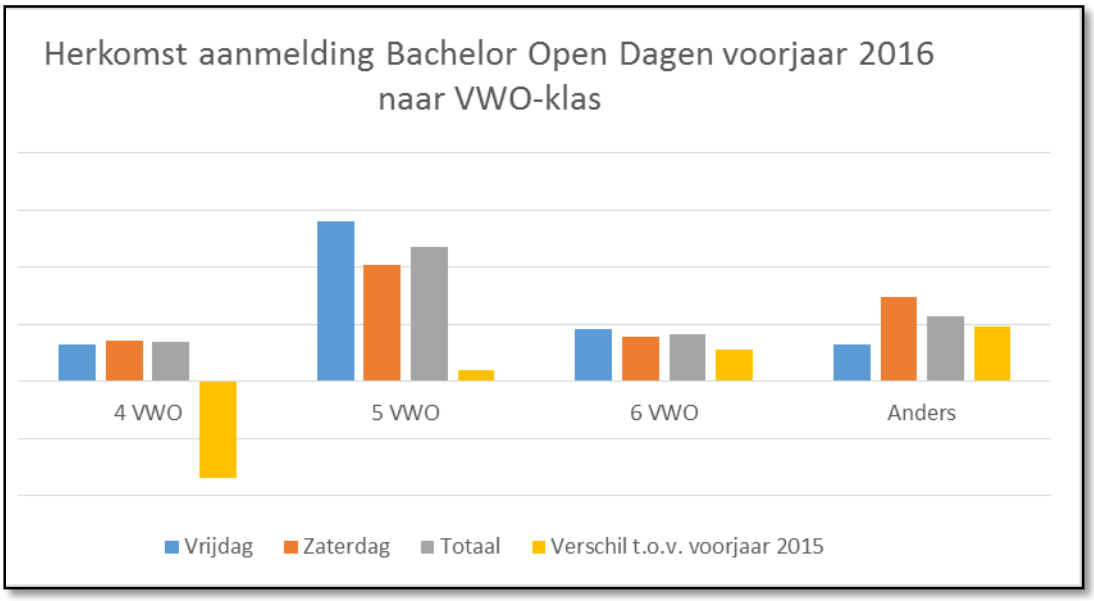
De data om de verschillende doelgroepen te identificeren zijn ook aanwezig in de aangeleverde dataset vanaf het najaar van 2012 (Figuur 3.1). Het land van herkomst wordt in de dataset geregistreerd doormiddel van het IP-adres van de aanmelder. De klas en school van de aanmelder kan de aanmelder handmatig invullen. Als een aanmelder vanuit het HBO komt, of vanuit het buitenland, wordt de studiezoeker niet onder de VWO doelgroepen geschaard. Uit de datasets van de jaren ervoor is de specifieke informatie om de doelgroepen te identificeren in mindere mate aanwezig.

Uit de evaluaties van de bachelor Open Dagen van het voorjaar van 2016 blijkt dat ieder land van herkomst een specifiek aandeel heeft in het aantal aanmeldingen voor de Bachelor Open Dagen. Dit aandeel verschilt per vrijdag en zaterdag en is daarnaast niet stabiel ten opzichte van het voorjaar van 2015 (Figuur 3.3).



Figuur 3.3: Herkomst aanmeldingen naar land van herkomst

Als we kijken naar het aantal aanmeldingen op basis van de VWO klassen, zien we hetzelfde verschijnsel als bij het land van herkomst. In Figuur 3.4 zijn de percentuele aanmeldingen te zien van het totaal aantal aanmeldingen. Percentueel gezien zijn er dus (ongeveer) evenveel aanmeldingen uit 4 VWO, maar minder aanmeldingen dan vorig jaar. Absoluut gezien zijn de aanmeldingen vanuit 4 VWO gestegen, maar door de toename van de andere klassen is dit verschil percentueel niet terug te vinden. Onder het “anders” worden alle aanmeldingen verzameld, waarbij niks is ingevuld bij de klas. Dus hier worden zowel buitenlandse aanmeldingen onder verzameld, als aanmeldingen vanuit het HBO.

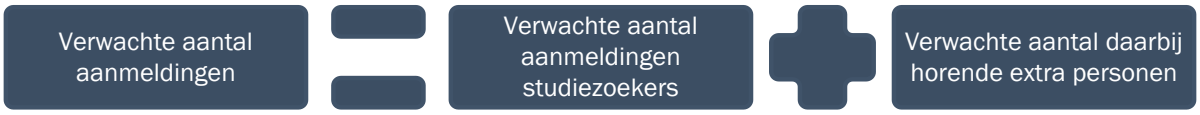


Figuur 3.4: Herkomst aanmeldingen naar VWO-klas

3.3 ANALYSE VAN HET AANTAL AANMELDINGEN

In de vorige paragraaf zijn de belangrijkste doelgroepen voor de voorspelling geïdentificeerd. Hieruit is gebleken dat het aandeel aanmeldingen van iedere doelgroep fluctueert over de jaren. Stap vier van het voorspellingsproces beschrijft dat het voor een succesvolle voorspelling noodzakelijk is om de factoren te identificeren die deze fluctuaties veroorzaken.

Een van de oorzaken die gegeven wordt door M&C voor de ontstaande problematiek is een stijging in het aantal aanmeldingen. Ze hebben hierbij geobserveerd dat het aandeel extra personen dat meekomt naar de Open Dag een hogere stijging met zich meebrengt, dan het aantal studiezoekers. Het verwachte aantal aanmeldingen kan zodoende worden onderverdeeld in twee componenten, met elk eigen specifieke eigenschappen (Figuur 3.5).

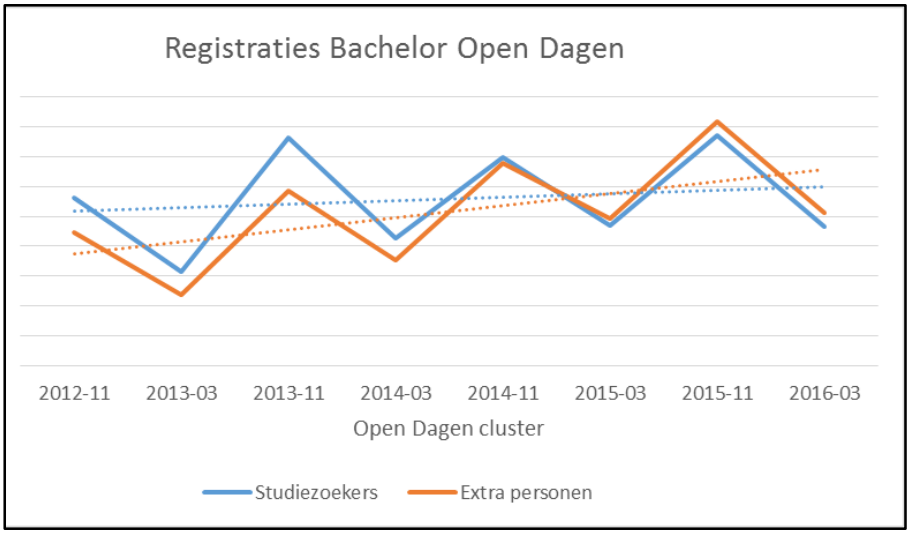


Figuur 3.5: Opbouw verwachte aantal aanmeldingen

Een stijging in één van de twee (of beiden) van de componenten zorgt voor een toename in het verwachte aantal aanmeldingen. Elk van deze twee componenten kunnen verschillende factoren bevatten die voor een stijging, daling of het stabiel houden van het aantal aanmeldingen kunnen zorgen. In de volgende sub paragrafen gaan we per component de beschikbare data analyseren en waar mogelijk specifieke trends waarnemen. We gaan dit stapsgewijs aanpakken, door eerst de data van het totaal aantal aanmeldingen te analyseren en daarna in te gaan de data van het aantal studiezoekers en extra personen. Deze data gaan we verder onderzoeken aan de hand van de geïdentificeerde doelgroepen. Data over het continue verloop van de inschrijvingen laten we verder achterwege, gezien de voorspellingsdoelstelling betrekking heeft op het voorspellen voordat het inschrijfproces geopend is. Het analyseren van het continue inschrijvingsproces heeft zodoende geen meerwaarde voor dit onderzoek.

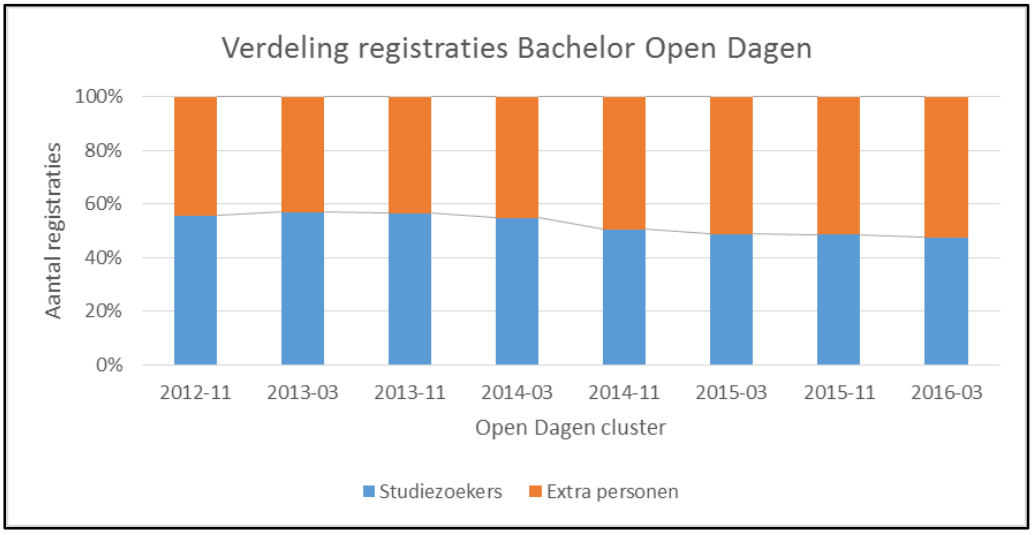
3.3.1 Analyse van het totaal aantal aanmeldingen

Als we naar het verloop van het totaal aantal aanmeldingen over de jaren kijken (Figuur 3.6), zien we een duidelijke trend in beide verschillende componenten. Het aantal aanmeldingen van de studiezoekers laat de afgelopen jaren een positieve trend zien, net zoals de aanmeldingen van het aantal extra personen. Echter, geeft het aantal aangemelde extra personen een positievere trend weer. Met andere woorden betekend dit dus dat het aantal extra personen harder is toegenomen over de afgelopen jaren ten opzichte van het aantal studiezoekers.



Figuur 3.6: Registraties Bachelor Open Dagen

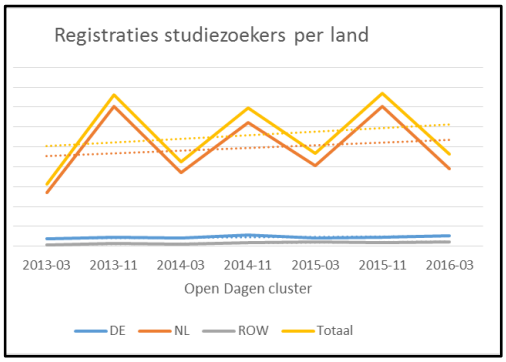
Dit valt ook terug te zien in Figuur 3.7. Het aantal aangemelde extra personen is de afgelopen jaren gestegen tot meer dan het werkelijke aantal aangemelde studiezoekers. Er melden momenteel dus meer extra personen zich aan voor de Open Dagen, dan werkelijke potentiële studenten.



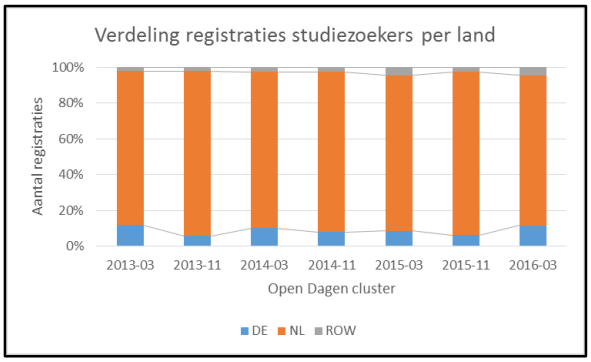
Figuur 3.7: Verdeling registraties Bachelor Open Dagen

3.3.2 Analyse van de aangemelde studiezoekers

In de vorige sub paragraaf hebben we het totaal aantal aanmeldingen onderverdeeld in aanmeldingen van studiezoekers en de daarbij behorende extra personen. Beide componenten bevatten eigen eigenschappen. In Figuur 3.8 is het verloop van het aantal aanmeldingen weergegeven van de studiezoekers over de afgelopen jaren, onderverdeeld naar de verschillende doelgroepen, zoals geïdentificeerd in de vorige paragraaf. Hieruit valt af te leiden dat iedere doelgroep een eigen aandeel heeft in het totaal aantal aanmeldingen, en dat iedere doelgroep eigen specifieke trends bevat. Uit het figuur is duidelijk af te leiden dat het percentage aanmeldingen van studiezoekers uit Nederland een stijging heeft doorgemaakt over de afgelopen jaren, terwijl de andere doelgroepen een stabieler niveau laten zien.

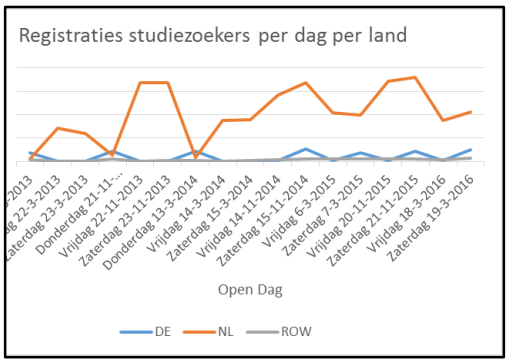


Figuur 3.8: Registraties studiezoekers per land

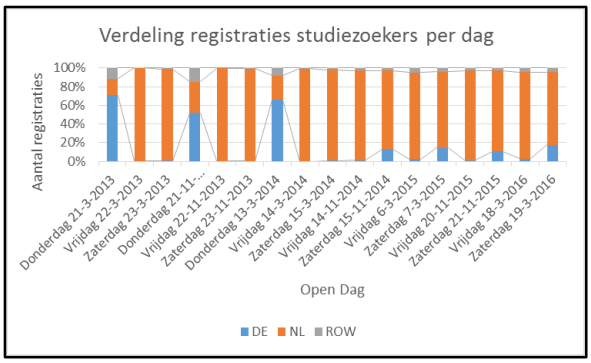


Figuur 3.9: Verdeling registraties studiezoekers per land

Dit valt ook terug te zien als we het totaal aantal aanmeldingen van de studiezoekers onderverdelen in het percentuele aandeel van iedere doelgroep (Figuur 3.9). Hieruit valt op te maken dat over de afgelopen jaren het percentuele aandeel aanmeldingen per land van herkomst fluctueert. Hieruit valt af te leiden dat over de afgelopen jaren het percentage aanmeldingen van studiezoekers uit Duitsland en de EER/ROW zijn toegenomen, en aanmeldingen uit Nederland percentueel dus zijn afgenomen.
Als we het aantal aanmeldingen van de studiezoekers onderverdelen per dag zien we een duidelijk patroon vanaf de Open Dagen van najaar 2014 (Figuur 3.10). Dit komt mede door dat vanaf het najaar van 2014 de Open Dagen enkel nog op de vrijdag en zaterdag gehouden worden. De internationale Open Dag op de donderdag is toen komen te vervallen.

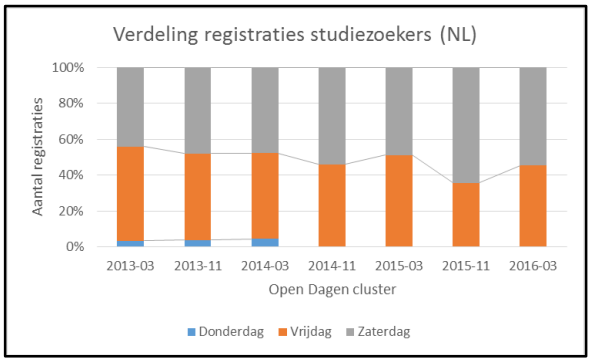


Figuur 3.10: Registraties studiezoekers per dag per land

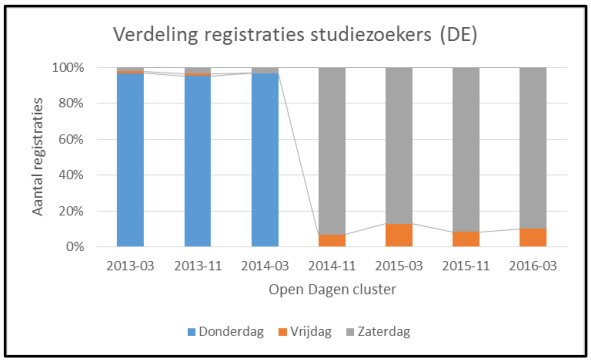


Figuur 3.11: Verdeling registraties studiezoekers per dag

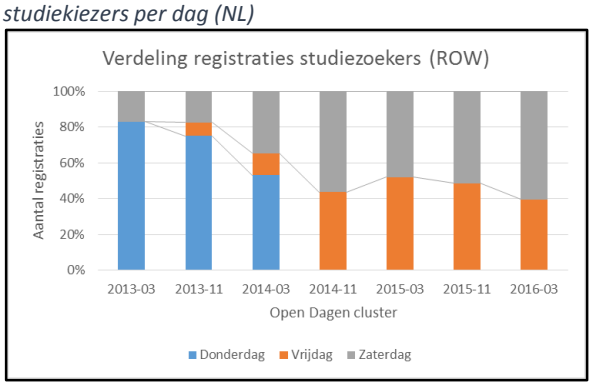
Het effect van het vervallen van de donderdag als internationale Open Dag is ook duidelijk terug te zien in het percentueel aantal aanmeldingen per doelgroep (Figuur 3.11). Hieruit blijkt dat iedere dag verschillende samenstellingen heeft van de doelgroepen die op de Open Dag afkomt. In Figuur 3.12, Figuur 3.13 en Figuur 3.14 zijn de verdelingen weergegeven van de verschillende doelgroepen over de beschikbare Open Dagen.



Figuur 3.12: Percentuele verdeling aanmeldingen studiekiezers per dag (NL)

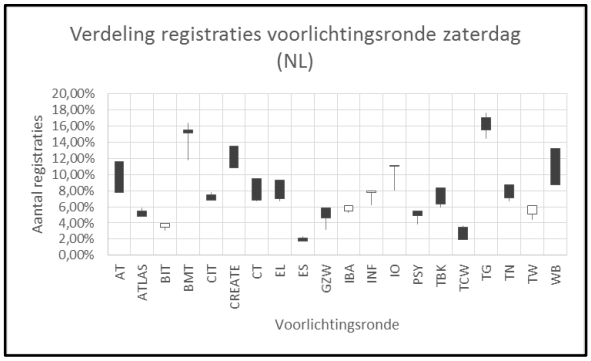


Figuur 3.13: Percentuele verdeling aanmeldingen studiekiezers per dag (DE)

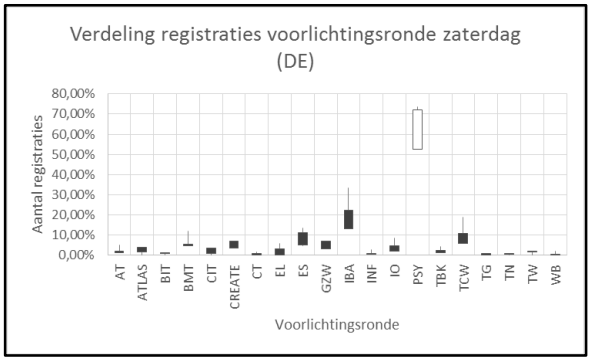


Figuur 3.14: Percentuele verdeling aanmeldingen studiekiezers per dag (ROW)

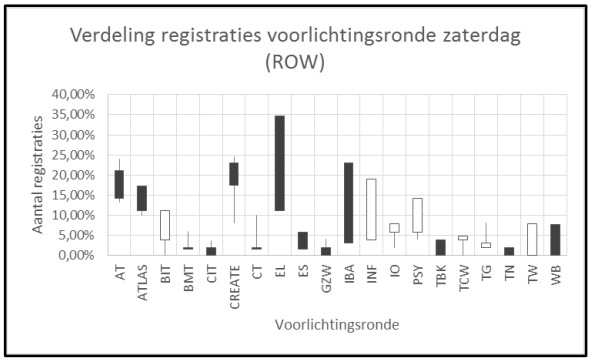
Naast de specifieke aandelen van de studiezoekers per dag, laten de verschillende doelgroepen ook voorkeuren zien voor bepaalde voorlichtingsrondes. De voorkeuren voor de voorlichtingsprogramma's op zaterdag zijn per doelgroep terug te vinden in opeenvolgend Figuur 3.15, Figuur 3.16 en Figuur 3.17. De gegevens zijn gebaseerd op het percentage registraties per voorlichtingsronde ten opzichte van het totaal aantal registraties per voorlichtingsdag. De organisatie van de Open Dagen verdeelt momenteel het aantal registraties over de ochtend en de middag. Voor de voorspelling is het zodoende niet relevant om het aantal registraties verder onder te verdelen naar ochtend- en middagniveau.



Figuur 3.15: Verdeling registraties studiezoekers per voorlichtingsronde zaterdag (NL)



Figuur 3.16: Verdeling registraties studiezoekers per voorlichtingsronde zaterdag (DE)

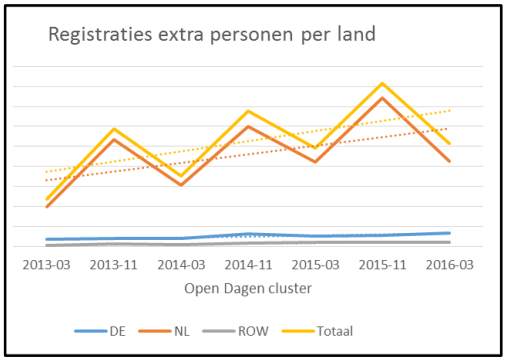


Figuur 3.17: Verdeling registraties studiezoekers per voorlichtingsronde (ROW)

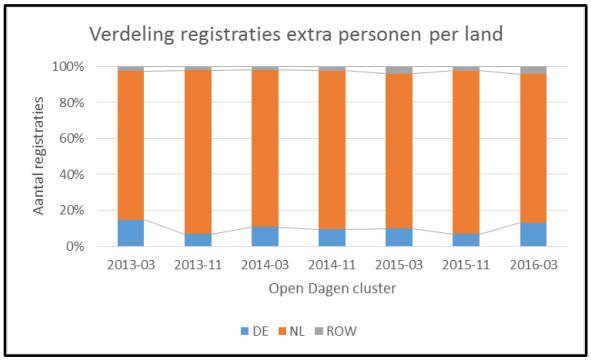
Uit de bovenstaande figuren valt af te leiden dat iedere doelgroep studiezoekers een eigen specifieke voorkeur hebben voor bepaalde voorlichtingsrondes op de zaterdag. De balken geven de spreiding weer binnen de percentage aanmeldingen over de afgelopen jaren, vanaf najaar 2014. De verdelingen van de studiezoekers over de vrijdag zijn terug te vinden in Appendix 3 en 5. Een verdere analyse van aanmeldingen uit Nederland, onderverdeeld in de verschillende VWO jaargangen, is terug te vinden in Appendix 7.

3.3.3 Analyse van het aantal aangemelde extra personen

Naast de aanmeldingen van de studiezoekers, heeft het totaal aantal aanmeldingen ook betrekking op het aantal extra personen dat de studiemelders aanmelden. In Figuur 3.18 is het verloop te zien van het aantal aangemelde extra personen over de afgelopen jaren. Bij de Nederlandse extra personen is een duidelijke trend zichtbaar, terwijl bij de andere doelgroepen een stabieler beeld gegeven wordt.



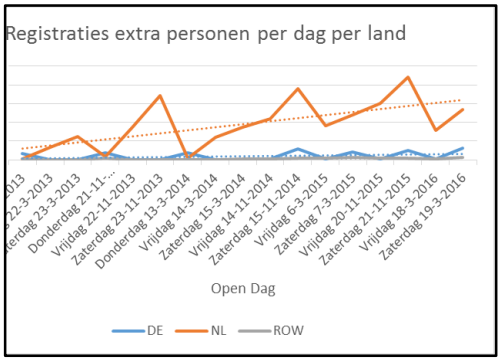
Figuur 3.18: Absolute verdeling aanmeldingen extra personen per doelgroep



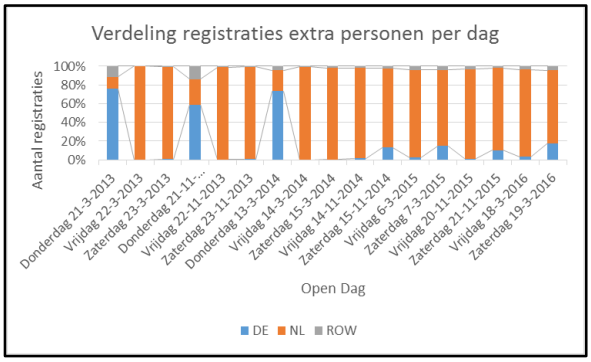
Figuur 3.19: Percentuele verdeling aanmeldingen extra personen per doelgroep

In Figuur 3.19 zijn de verdelingen tussen de doelgroepen percentueel weergegeven. Hieruit valt op dat de Nederlandse extra personen momenteel rond de 80% van het totaal uitmaken, gevolgd door aanmeldingen uit Duitsland en de rest van de wereld.

Als we deze gegevens verder uitsplitsen per dag, zien we een vergelijkbaar beeld (Figuur 3.20 *en* Figuur 3.21*).* Uit de figuren valt af te leiden dat voornamelijk de aanmeldingen van extra personen uit Duitsland een voorkeur hebben voor de zaterdag.

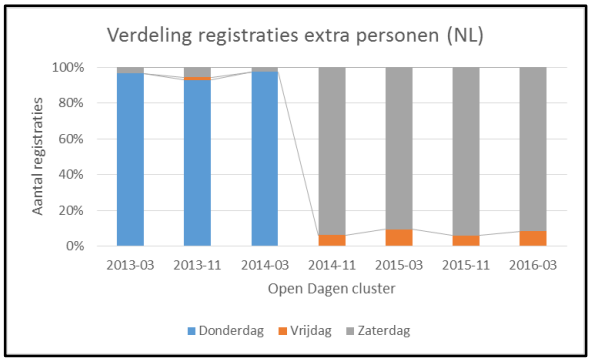


Figuur 3.20: Absolute verdeling aanmeldingen extra personen per dag, per doelgroep

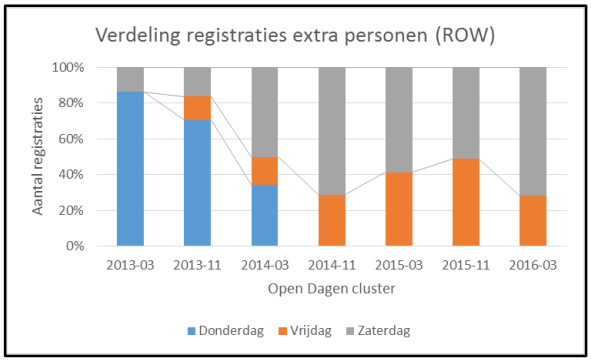


Figuur 3.21: Percentuele verdeling aanmeldingen extra personen per dag, per doelgroep

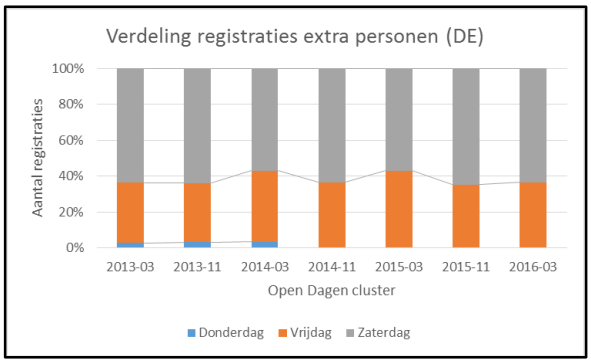
Figuur 3.20 laat zien dat de waargenomen trend van een stijgend aantal aanmeldingen per Open Dag zowel voor de vrijdag, zoals de zaterdag geldt. Net zoals in Figuur 3.18 valt in Figuur 3.20 op dat de stijging voornamelijk te danken is door een toename van extra personen uit Nederlandse aanmeldingen over de afgelopen jaren. In Figuur 3.22, Figuur 3.23 en Figuur 3.24 zijn per doelgroep het totaal aantal aanmeldingen per Open Dagen als cluster percentueel onderverdeeld per open dag.



Figuur 3.22: Verdeling registraties extra personen (NL)



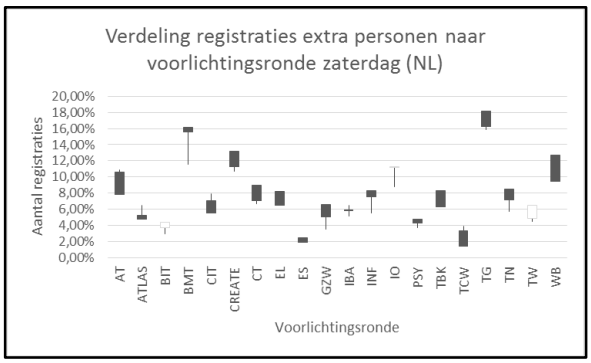
Figuur 3.23: Verdeling registraties extra personen (ROW)



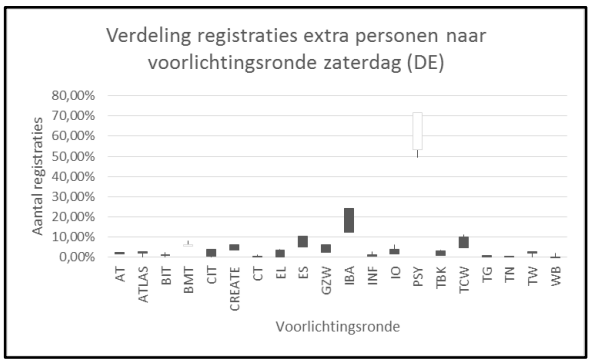
Figuur 3.24: Verdeling registraties extra personen (DE)

De bovenstaande figuren 3.22, 3.23 en 3.24 geven een vergelijkbaar beeld weer als de verdeling van de studiekiezers per doelgroep. Bij de aanmeldingen van extra personen uit Nederland zien we een (ongeveer) stabiele verdeling, waarbij het merendeel van de aanmeldingen voor de zaterdag is. Bij de aanmeldingen uit Duitsland zien we dit verschijnsel ook. De aanmeldingen uit de EER/ROW laten een minder stabiel beeld zien.

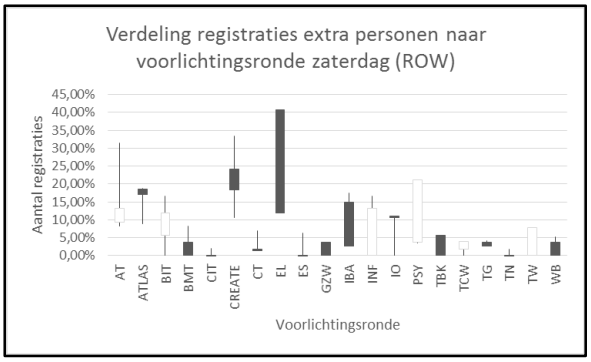
Naast de specifieke aandelen van de extra personen per dag, laten de verschillende doelgroepen ook voorkeuren zien voor bepaalde voorlichtingsrondes. De verdelingen van de registraties naar voorlichtingsrondes op zaterdag zijn terug te vinden in Figuur 3.25, Figuur 3.26 en Figuur 3.27.



Figuur 3.25: Verdeling registraties extra personen naar voorlichtingsronde zaterdag (NL)



Figuur 3.26: Verdeling registraties extra personen naar voorlichtingsronde zaterdag (DE)

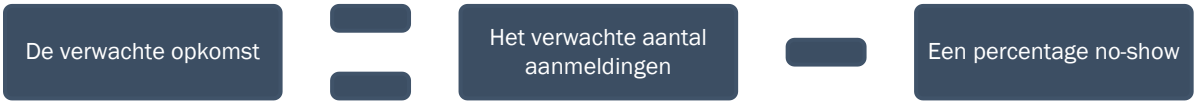


Figuur 3.27: Verdeling registraties extra personen naar voorlichtingsronde zaterdag (ROW)

De bovenstaande verdelingen van aanmeldingen van extra personen op de zaterdag laten overeenkomsten zien met de verdeling van aanmeldingen van studiekiezers. Echter, laten de grafieken ook verschillen zien in voorlichtingsrondes. In Appendices 4 en 6 zijn de percentuele verdelingen te vinden per doelgroep op de vrijdag. Een uitgebreidere analyse van aanmeldingen van extra personen uit Nederland, onderverdeeld naar de VWO doelgroepen, is terug te vinden in Appendix 8.

3.4 ANALYSE VAN DE VERWACHTE OPKOMST

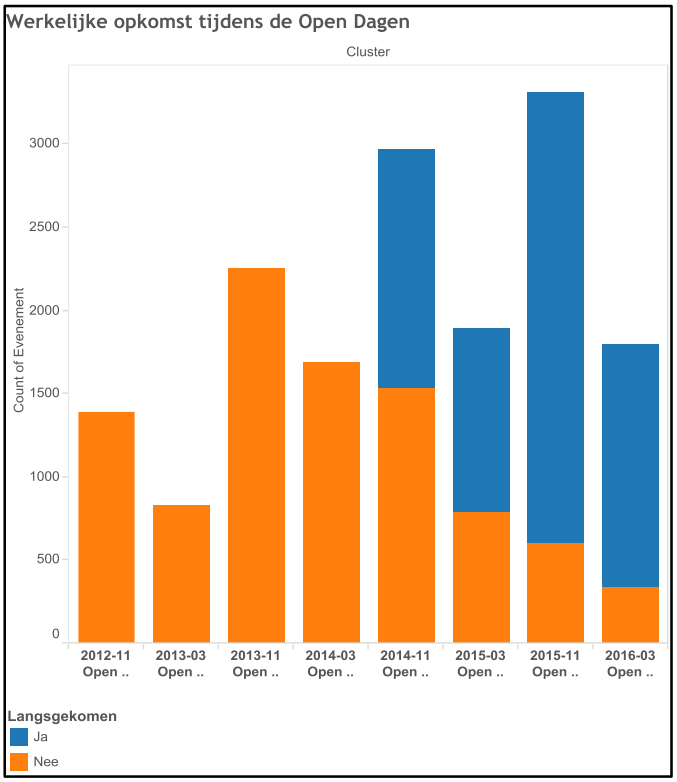
In de onderzoeksdoelstelling maken we onderscheid tussen het verwachte aantal aanmeldingen en de verwachte opkomst op de Open Dagen. Uit interviews met de betrokken medewerkers van M&C is gebleken dat tijdens de Open Dagen het aantal personen dat geteld wordt veelal afwijkt van het aantal aanmeldingen die geregistreerd zijn (Figuur 3.28).



Figuur 3.28: De werkelijke opkomst

De organisatie van de Open Dagen maakt momenteel al gebruik van een percentage no-show om de verwachte opkomst te bepalen. Dit percentage no-show wordt momenteel gebaseerd op handtellingen over de afgelopen jaren. Hier wordt momenteel gebruik gemaakt van een simple-average voorspellingstechniek, waarbij het gemiddelde percentage over de afgelopen Open Dagen gebruikt wordt voor het bepalen van het percentage no-show voor de komende Open Dagen. Deze manier van voorspellen brengt onnauwkeurigheden met zich mee, gezien er handmatig geteld wordt en dat er geen onderscheid gemaakt kan worden tussen verschillende componenten en doelgroepen. Immers kan iedere doelgroep eigen specifieke kenmerken bevatten die nodig zijn voor het nauwkeurig bepalen van het percentage no-show.

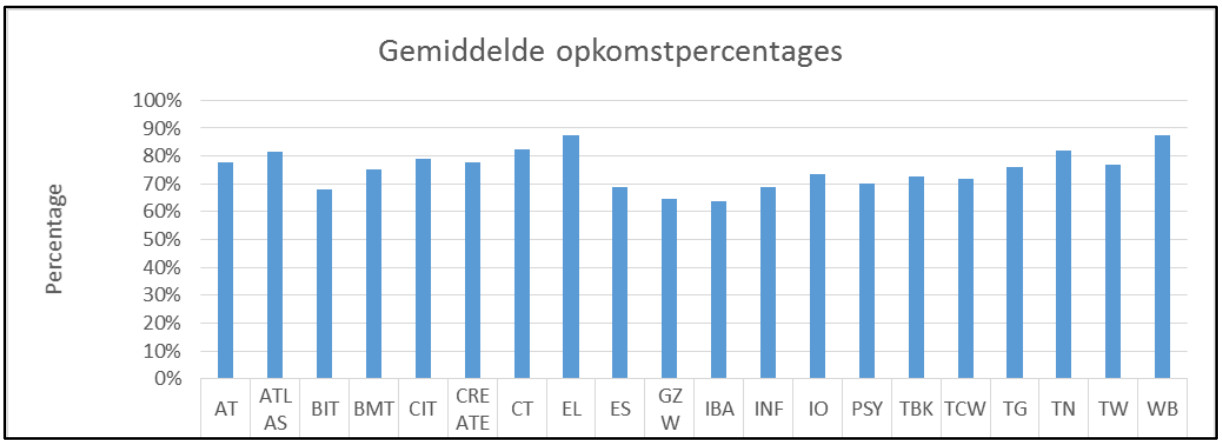
Om het bovengenoemde probleem van onnauwkeurigheid op te lossen en de opkomstcijfers beter in kaart te kunnen brengen, is vanaf het najaar van 2014 een digitaal systeem geïntroduceerd. Als een studiezoeker zich heeft geregistreerd voor de Open Dagen krijgt hij/zij hiervan een email met daarop een barcode. Deze barcode is gelinkt aan de specifieke studiezoeker en wordt gescand bij iedere voorlichtingsronde. Zodoende kan gekeken worden op persoonsniveau of iemand aanwezig was en met hoeveel extra personen. Vanaf het najaar van 2014 is deze data beschikbaar (*figuur 39)*. Echter, zijn tijdens de eerste twee Open Dagen onnauwkeurigheden in de data aangetroffen, waardoor deze niet beschikbaar zijn voor analyse. We zullen de data analyse dus uitvoeren op basis van gegevens van het najaar van 2015 en het voorjaar van 2016. In de volgende paragraaf zullen we de opkomstcijfers van deze Open Dagen analyseren per doelgroep, zoals geïdentificeerd in paragraaf 3.2.



Figuur 3.29: Beschikbare no-show data

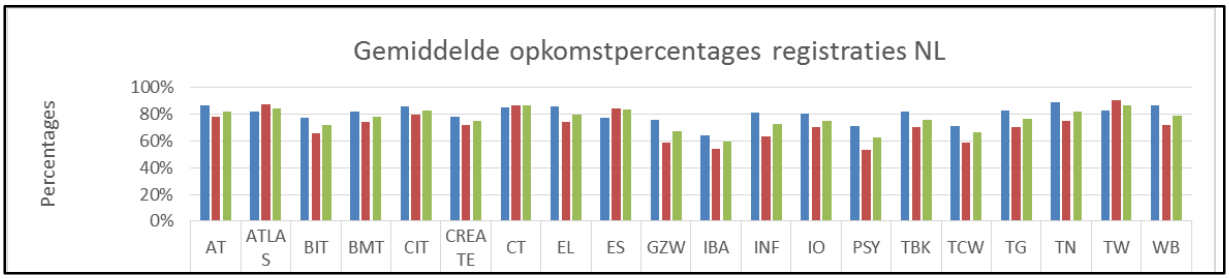
3.4.1 Opkomstcijfers per doelgroep

In Figuur 3.30 zijn de gemiddelde opkomstpercentages weergegeven over de afgelopen twee Open Dagen. De opkomstpercentages per doelgroep zijn hierbij uitgesplist in de opkomstpercentages van de aangemelde studiekiezers, extra personen en de totale opkomst, gebaseerd op absolute getallen. Het kan voorkomen dat een opkomstpercentage op 0% staat. Dit betekend dat er geen personen gescand waren, óf geen extra personen geregistreerd waren en wel extra personen gescand waren. Als er geen extra personen geregistreerd waren en wel gescand waren (dus niet opgegeven, maar wel meekomen), is dit terug te zien in het opkomstpercentage van het totaal aantal aanmeldingen. In Appendix 9 zijn de uitgesplitste opkomstpercentages weergegeven van opeenvolgend 4, 5 en 6 VWO aanmeldingen. Hierbij moet rekening gehouden worden met de percentuele opkomstweergave in de grafieken, niet de absolute. Als een bepaalde voorlichtingsronde een zeer klein aantal aanmeldingen heeft, zou één persoon no-show percentueel gezien een veel grotere invloed hebben ten opzichte van druk bezochte voorlichtingsprogramma's en zal daarom een minder stabiel beeld geven.



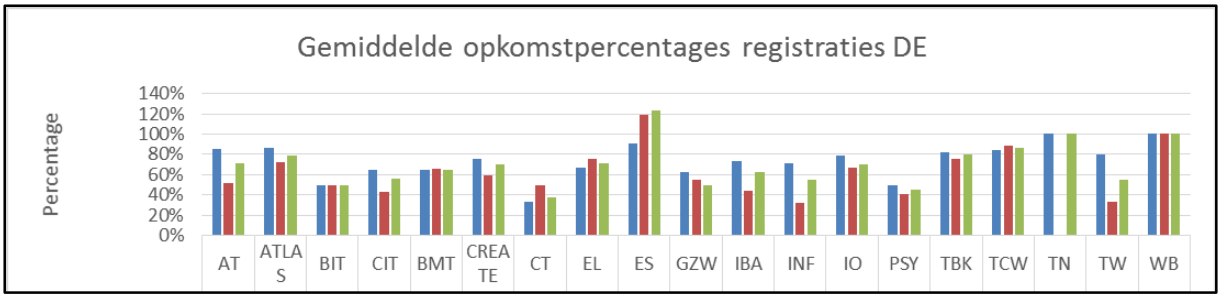
Figuur 3.30: Gemiddelde opkomstpercentages Open Dagen 2015-11 en 2016-03

De gemiddelde opkomstpercentages kunnen uitgesplitst worden in de opkomstpercentages van de verschillende doelgroepen die geïdentificeerd zijn. In Figuur 3.31 zijn de gemiddelde opkomstpercentages te vinden voor de registraties uit Nederland, uitgesplitst tegen de studiezoekers, extra personen en de totale opkomst.



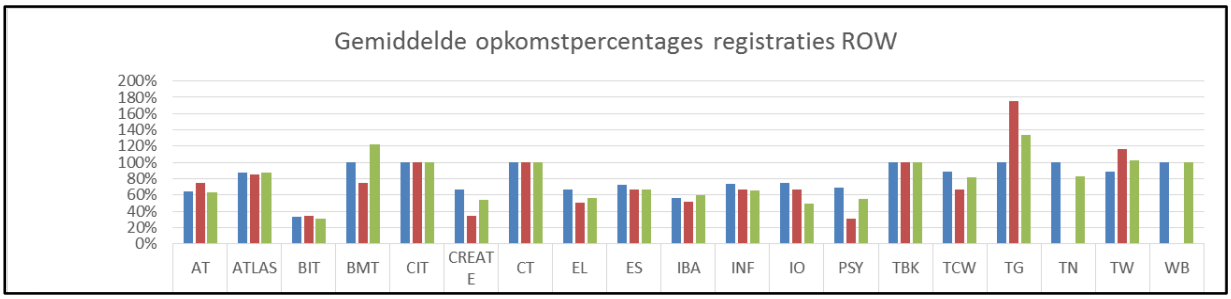
Figuur 3.31: Gemiddelde opkomstpercentages registraties uit NL over de Open Dagen 2015-11 en 2016-03

Naast de aanmeldingen uit Nederland, hebben we ook de registraties uit Duitsland als doelgroep geïdentificeerd. In Figuur 3.32 zijn de gemiddelde opkomstpercentages weergegeven van registraties afkomstig uit Duitsland, weergegeven als de opkomstpercentages van studiezoekers, extra personen en de totale opkomst. In Appendix 10 zijn de opkomstpercentages per voorlichtingsronde weergegeven.



Figuur 3.32: Gemiddelde opkomstpercentages registraties uit DE over de Open Dagen 2015-11 en 2016-03

De laatste doelgroep die we hebben geïdentificeerd voor de voorspelling zijn aanmeldingen uit de rest van de wereld. Deze groep aanmeldingen is percentueel gezien het kleinste aandeel van de Open Dagen. Doordat deze groep een klein aandeel heeft in de aanmeldingen van de Open Dagen, heeft één aanmelder extra veel invloed op het percentage no-show van deze doelgroep. Hierdoor geeft de opkomstpercentages een minder stabiel beeld weer, ten opzichte van aanmeldingen uit Nederland. In Figuur 3.33 zijn de gemiddelde opkomstpercentages weergegeven. In Appendix 12 zijn deze gemiddelde opkomstpercentages uiteengezet naar de opkomstpercentages per dag.



Figuur 3.33: Gemiddelde opkomstpercentages registraties uit ROW

3.5 CONCLUSIE

In dit hoofdstuk hebben we een inventarisatie gemaakt van de benodigde en beschikbare historische data. We hebben hierbij de beschikbare en benodigde data verzameld en geanalyseerd, aan de hand van stap 3 en 4 van het voorspellingsproces. De data is onderzocht op het bevatten van verschillende doelgroepen. Hieruit is naar voren gekomen dat de registraties voor de Open Dagen onderverdeeld kunnen worden in registraties uit de verschillende landen van herkomst, namelijk aanmeldingen uit Nederland, Duitsland en de rest van de wereld. De registraties uit Nederland kunnen verder onderverdeeld worden in registraties uit de verschillende vooropleidingsjaren uit het VWO.

Elk van de registraties uit de verschillende doelgroepen hebben we kunnen onderverdelen in aanmeldingen van studiezoekers en de daarbij behorende extra personen. Uit de data analyse is gebleken dat elke doelgroep eigen eigenschappen bevat, op het gebied van een level, trend en seizoeninvloed.

De verschillende doelgroepen laten ook voorkeuren zien voor bepaalde voorlichtingsdagen en voorlichtingsrondes. Zo laten de Duitse aanmeldingen een voorkeur zien voor Engelstalige bacheloropleidingen en laten de Nederlandse aanmeldingen een voorkeur zien voor voorlichtingen als Biomedische Technologie.

Uit interviews met de betrokken medewerkers van M&C is gebleken dat er een verschil bestaat tussen het aantal aanmeldingen en de werkelijke opkomst op de Open Dagen. Voor de data analyse van de werkelijke opkomstpercentages hebben we gebruik gemaakt van een nieuw registratiesysteem dat sinds de Open Dagen van het najaar van 2015 betrouwbaar in gebruik is. Hieruit is naar voren gekomen dat de opkomstpercentages verschillenen per doelgroep en per voorlichtingsronde.

Door middel van de uitgevoerde data analyse in dit hoofdstuk hebben we de eigenschappen van de beschikbare data bepaald en daarmee hebben we deelvraag 2 beantwoord.

Hoofdstuk 4

De voorspelling van het verwachte aantal registraties

(Deelvraag 3)

“The goal of forecasting is not to predict the future, but to tell you what you need to know to take meaningful action in the present.”

Paul Saffo

Amerikaans (technologisch) voorspeller

4 VOORSPELLING VAN HET VERWACHTE AANTAL REGISTRATIES

Dit hoofdstuk gaat in op de opstelling van de voorspelling van het verwachte aantal registraties van de bachelor Open Dagen. In hoofdstuk drie is naar voren gekomen dat de historische data onderverdeeld kan worden in verschillende doelgroepen met elk eigen specifieke kenmerken. Deze datakenmerken gaan we in dit hoofdstuk gebruiken om een betrouwbaar voorspellingsmodel op te stellen. Paragraaf 1 zet de keuze voor de juiste voorspellingstechniek uiteen. De paragraaf die daarop volgt bevat de opstelling van het werkelijke voorspellingsmodel. In paragraaf 3 gaan we de verschillende opgestelde componenten samenvoegen tot de werkelijke voorspelling van het aantal registraties. We sluiten dit hoofdstuk af met een uiteenzetting van de betrouwbaarheid van de voorspelling.

4.1 GESCHIKTE VOORSPELLINGSTECHNIEKEN

In hoofdstuk 2 hebben we verschillende mogelijke geschikte voorspellingstechnieken onderzocht, met elk een specifieke toepassing. In het vorige hoofdstuk is naar voren gekomen dat de beschikbare data verschillende doelgroepen bevat, met elk specifieke eigenschappen (Tabel 4.1).

Onderzochte doelgroep:	**Waargenomen eigenschappen:**
Studiezoekers/extra personen uit 4/5/6 VWO en overig NL	(Level) Trend en seizoeninvloed
Studiezoekers/extra personen uit DE	(Level) Trend en seizoeninvloed
Studiezoekers/extra personen uit ROW	(Level) Trend en seizoeninvloed

Tabel 4.1: Data eigenschappen

In Tabel 4.2 zijn de onderzochte extrapolatiemodellen weergegeven. Hieruit valt op dat de waargenomen eigenschappen uit de data analyse overeen komen met de eigenschappen die Winter's method gebruikt voor het doen van voorspellingen. Op basis van de data analyse achten we Winter's method de meest bruikbare methode om de aanmeldingen van de verschillende doelgroepen te kunnen voorspellen. In Appendix 13 hebben we een enkelvoudige voorspelling uitgevoerd voor het verwachte aantal registraties op clusterniveau, met de verschillende voorspellingstechnieken. Winters' Method is ook hiervoor het nauwkeurigst gebleken. We zullen zodoende de voorspelling op gaan stellen door gebruik te maken van Winters' method.

	Level	**Trend**	**Seizoeninvloed**
Simple, moving and Weighter average			
Simple exponential smoothing			
Holt's Method			
Winters' Method			

Tabel 4.2: Extrapolatiemodellen

In hoofdstuk 2 hebben we naast de verschillende soorten extrapolatiemodellen, ook causale modellen onderzocht. In de literatuur wordt gesproken over de invloed van marketing en andere Open Dagen van vergelijkbare onderwijsinstellingen op het aantal mogelijke aanmeldingen. De data om een verband tussen deze factoren en het aantal aanmeldingen van de Open Dagen te onderzoeken is momenteel niet aanwezig. Gezien de tijd die voor dit onderzoek staat, achten we het niet mogelijk om de data hiervoor te verkrijgen. We zullen deze factoren dan ook achterwege laten.

Naast de bovengenoemde factoren, achten we het ook mogelijk dat het aantal VWO scholieren in Nederland van invloed is op het aantal aanmeldingen van de Open Dagen. De aanmeldingen van VWO scholieren als studiezoeker hebben we als een aparte doelgroep meegenomen in de data analyse. Vanaf het najaar van 2012 zijn eenduidige gegevens hierover beschikbaar (Tabel 4.3*)*.

	2012-11	2013-03	2013-11	2014-03	2014-11	2015-03	2015-11	2016-03
4 Vwo	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx
5 Vwo	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx
6 Vwo	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx
Totaal	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 4.3: Aanmeldingen uit het VWO

Het Centraal Bureau voor de Statistiek (CBS) houdt gegevens bij over het aantal VWO scholieren in Nederland, zoals weergegeven in Tabel 4.4.

Klas	2012/'13	2013/'14	2014/'15	2015/'16	Verschil max/min
4 VWO	41628	41678	42220	42797	2,73%
5 VWO	40527	40206	40733	42820	6,10%
6 VWO	36689	36393	36965	37148	2,03%
Totaal:	118844	118277	119918	122765	3,66%

Tabel 4.4: Aantal VWO scholieren in Nederland (CBS, 2016)

In Appendix 14 zijn verschillende plots weergegeven die het verband tussen de VWO aanmeldingen en VWO scholieren in Nederland illustreren. In deze plots lijkt het erop dat er enigszins een correlatie is tussen het aantal aangemelde VWO scholieren en het aantal VWO scholieren in Nederland, op basis van een enkelvoudige lineaire regressie doormiddel van de kleinste kwadratische som. In Appendix 14 zijn de regressie analyses weergegeven van de verschillende correlaties en de daar bijhorende t-toetsen. Uit deze t-toetsen komt naar voren dat, mede door het geringe aantal datapunten, niet met betrouwbaarheid nagegaan kan worden óf er een correlatie bestaat tussen de twee variabelen. Dit valt ook te herleiden uit de kwadraatsom van de gemiddelde afwijking van iedere correlatie, die een zwak tot zeer zwakke correlatie laat zien. We zullen deze correlaties dan ook niet meenemen in het voorspellingsmodel.

Gebruik maken van verschillende voorspellingstechnieken verhoogd de betrouwbaarheid van de voorspelling, zoals weergegeven in de drie regels van voorspellen (Hoofdstuk 2). Doordat momenteel het niet mogelijk is om de causale verbanden te verifiëren tussen het aantal registraties en factoren in het omliggende milieu, kan deze causale voorspellingstechniek niet gebruikt worden. Volgens de drie regels van voorspellen betekend dit dat we de kwaliteit en betrouwbaarheid van de voorspelling niet kunnen verhogen door gebruik te maken van een extra voorspellingsmethode.

4.2 TIJDREEKSVOORSPELLING

In het vorige hoofdstuk is de data geanalyseerd met betrekking op de verschillende doelgroepen die we geïdentificeerd hebben. Er is gebleken dat er sprake is van verschillende trends in de doelgroepen en verschillen tussen de doelgroepen. Voor de voorspelling op basis van de tijdreeksen van deze doelgroepen, maken we gebruik van het individueel voorspellen van elk van de componenten binnen de doelgroepen en voegen deze samen voor een gecombineerde voorspelling van het aantal aanmeldingen als cluster (Appendix 16). In de data hebben we trends aangetroffen, alsook seizoeninvloed. In de literatuur wordt hiervoor het Winters model aangeraden, voor het adaptief voorspellen van het toekomstige aantal aanmeldingen.

4.2.1 Winters' methode

In hoofdstuk drie hebben we zes verschillende doelgroepen geïdentificeerd, namelijk aanmeldingen uit Nederland (onderverdeeld in VWO jaargangen en overige aanmeldingen), Duitsland en de rest van de wereld. Elk van deze aanmeldingen hebben we onderverdeeld in aanmeldingen van studiezoekers en daarbij behorende extra personen. Van elk van deze tijdreeksen zullen we een voorspelling gaan opstellen.

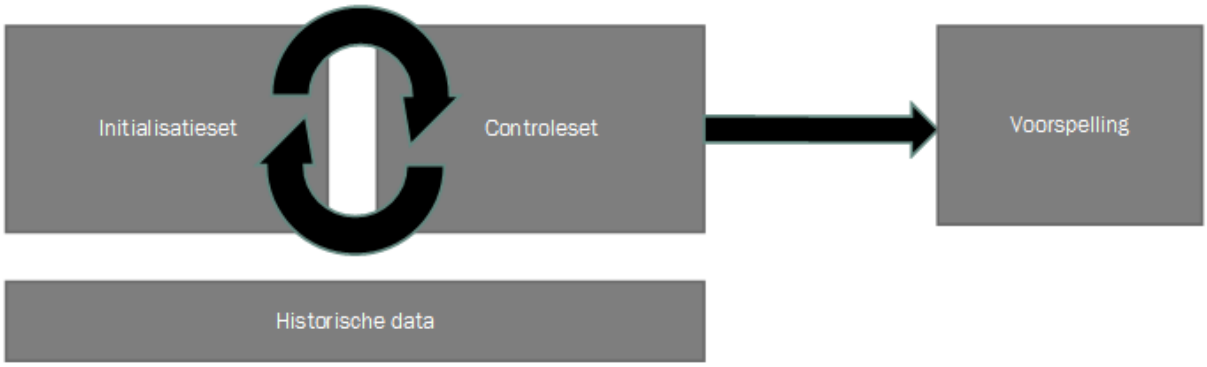
Winters' methode maakt gebruik van het adaptief voorspellen, waarbij de voorspellingsvariabelen telkens worden geüpdatet aan de hand van de nieuwste gerealiseerde waarden. Voor de opstelling van de voorspelling maken we gebruik van een vierstappen methode (*Chopra & Meindl, 2007)* om het systematische onderdeel van de data te gaan bepalen en te voorspellen.

Stap 1: Het initialiseren
De eerste stap heeft betrekking op het analyseren van het systematische onderdeel. Hiervoor maken we gebruik van een initalisatieset (Figuur 4.1). Deze initalisatieset bestaat uit historische gegevens, die we gebruiken voor het bepalen van het level, de trend en de seizoenfactor in de data. We hebben acht betrouwbare datapunten voor de opstelling van ons voorspellingsmodel. Dit betekend dat we de eerste vier datapunten gebruiken voor het initialiseren van het model.

Stap 2: Het voorspellen
In de eerste stap zijn uit de initalisatieset de gegevens geanalyseerd die nodig zijn voor de opstelling van Winters' methode. Met deze gegevens kunnen we een voorspelling maken voor data uit de controleset. Deze controleset bestaat, net als de initialisatieset, ook uit historische data. We kunnen zodoende een voorspelling maken van datapunten vijf tot en met acht. Deze controleset wordt gebruikt om de prestatie van de voorspelling te bepalen en waar nodig bij te sturen.

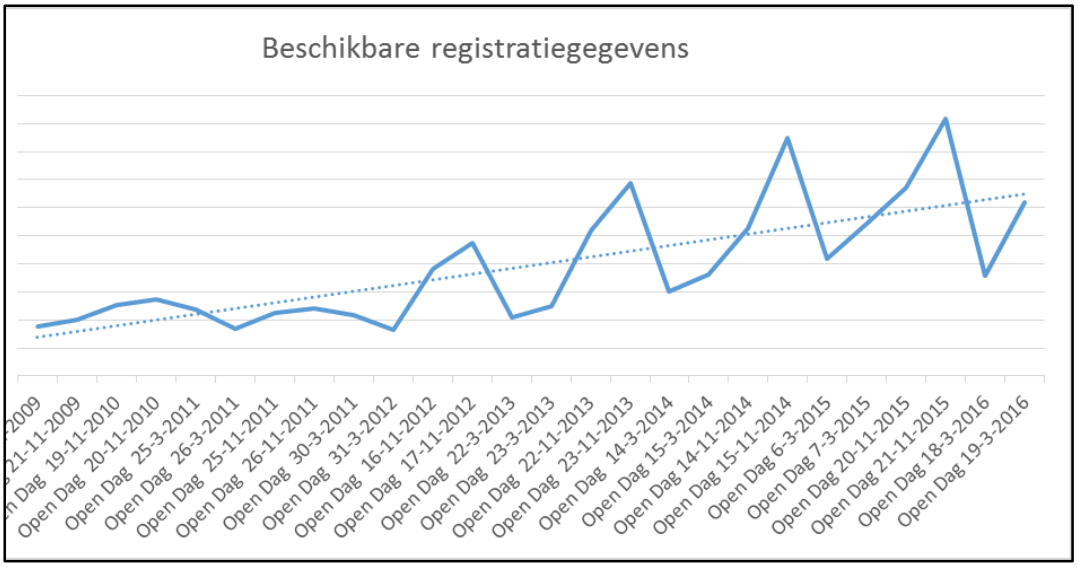
Stap 3: De onnauwkeurigheid bepalen
Doordat we gebruik maken van een initialisatieset en een controleset van historische data, is het mogelijk om de voorspelling te controleren op betrouwbaarheid op basis van historische gegevens. Dit doen we door het verschil te bepalen tussen de voorspelde waarde en de werkelijke waarde.

Stap 4: Model aanpassen
In de derde stap hebben we de nauwkeurigheid van de voorspelling bepaald op basis van historische gegevens. Als de voorspelling erg afwijkt van de werkelijke waarde, passen we het model aan en doorlopen we stap 2 tot en met 4 nog een keer totdat het model goed presteert.



Figuur 4.1: De opstelling van de voorspelling

Voor de initialisatie- en controleset maken we gebruik van 8 datapunten, vanaf het voorjaar van 2013. We hebben gekozen om de data uit het verdere verleden niet mee te nemen in de initialisatie- en controleset, omdat de data afwijkende patronen vertoond (Figuur 4.2) en uit deze data niet de eigenschappen van de verschillende doelgroepen geanalyseerd konden worden, die wel aanwezig zijn in de data vanaf het maart 2013.



Figuur 4.2: Beschikbare registratiedata

4.2.1.1 De voorspelling van het verwachte aantal aanmeldingen uit Duitsland

De aanmeldingen uit Duitsland hebben we onderverdeeld in aanmeldingen van studiezoekers en daarbij behorende extra personen. Elk van de deze componenten bevatten een eigen dataset, met eigen eigenschappen (Hoofdstuk 3). De eerste stap van de opstelling van het voorspellingsmodel is het initialiseren van de data uit de initialisatieset (Tabel 4.5).

Open Dagen	Studiezoekers	Extra personen
2013-03	xxxxx	xxxxx
2013-11	xxxxx	xxxxx
2014-03	xxxxx	xxxxx
2014-11	xxxxx	xxxxx

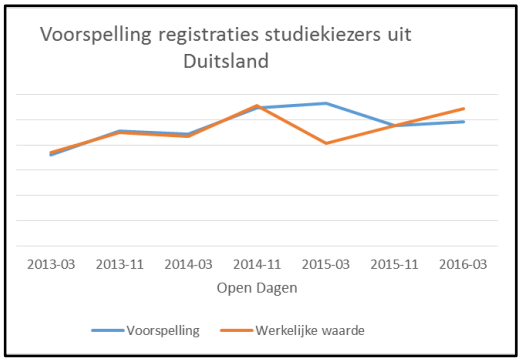
Tabel 4.5: Initialisatieset Duitse aanmeldingen

Winters' model schrijft voor om de data uit de initialisatie eerst te deseizoeneren (Hoofdstuk 2). Door het deseizoeneren van de data worden de seizoeninvloeden uit de data gehaald. Deze seizoeninvloeden komen later terug in de voorspelling doormiddel van een seizoenfactor. Voor de voorspelling van het aantal aanmeldingen uit Duitsland betekend dit dat er twee seizoenen zijn; een laag seizoen (maart) en een hoog seizoen (november). Nadat de data gedeseizoeneerd is, bepalen we doormiddel van regressie analyse het level in T0 en de trend. Waarna we de seizoenfactoren bepalen, zoals in hoofdstuk 2 beschreven is.

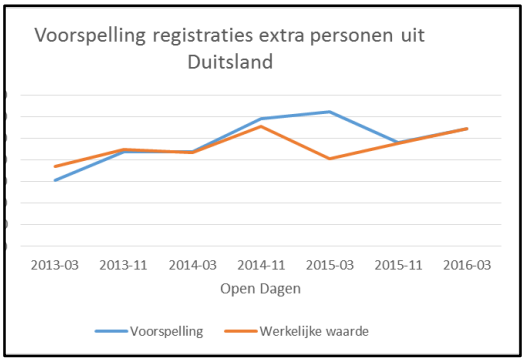
Aanmeldingen uit Duitsland	Leven in T0	Trend	Seizoenfactor hoog	Seizoenfactor laag
Studiezoekers	xxxxx	xxxxx	xxxxx	xxxxx
Extra personen	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 4.6: Gegevens uit de initialisatieset (Aanmeldingen uit Duitsland)

In Tabel 4.6 zijn de gegevens weergegeven uit de initialisatieset. Met deze gegevens kan het voorspellingsmodel voor het verwachte aantal registraties uit Duitsland opgesteld worden. Hiervoor maken we gebruik van Winters' method, zoals beschreven is in hoofdstuk 2. Jaarpunten van de Open Dagen van 2015-03 tot en met 2016-03 gebruiken we als controleset voor de voorspelling.



Figuur 4.3: Voorspelling registraties studiekiezers (DE)



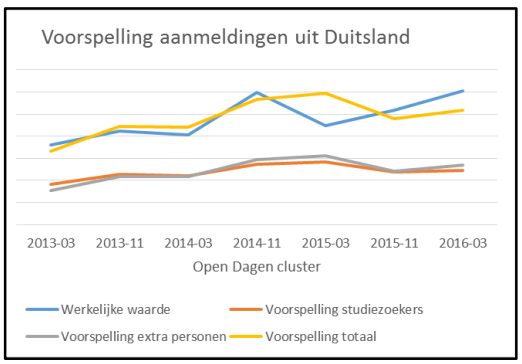
Figuur 4.4: Voorspelling registraties extra personen (DE)

In Figuur 4.3 is de voorspelling te vinden van het verwachte aantal aanmeldingen van studiezoekers uit Duitsland. In Figuur 4.4 is de voorspelling te vinden van het aantal daarbij behorende extra personen.

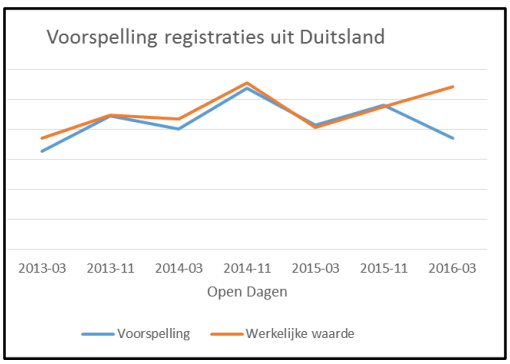
Voorspelling	Studiekiezers	Extra personen	Totaal gecombineerd	Totaal enkelvoudig
Alpha	xxxxx	xxxxx	xxxxx	xxxxx
Beta	xxxxx	xxxxx	xxxxx	xxxxx
Gamma	xxxxx	xxxxx	xxxxx	xxxxx
MAD	xxxxx	xxxxx	xxxxx	xxxxx
st	xxxxx	xxxxx	xxxxx	xxxxx
Mape	xxxxx	xxxxx	xxxxx	xxxxx
Bias	xxxxx	xxxxx	xxxxx	xxxxx
TS	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 4.7: Voorspelling aanmeldingen uit Duitsland

In Tabel 4.7 zijn de gegevens te vinden van de voorspelling van het aantal aanmeldingen uit Duitsland. De onnauwkeurigheden zijn bepaald op basis van de controleset. Uit de figuren valt duidelijk op te maken wat het effect is van het adaptief voorspellen van Winters' methode. Het model leert en wordt steeds nauwkeuriger. De alpha, beta en gamma zorgen ervoor dat het model bijgesteld wordt naar de meest optimale waarde van de MAPE, gebaseerd op de jongste gerealiseerde waarde. In Figuur 4.5 is de samengestelde voorspelling weergegeven van de aanmeldingen uit Duitsland en in Figuur 4.6 de enkelvoudige voorspelling. Als we de onnauwkeurigheden van deze twee voorspellingstechnieken vergelijken (Tabel 4.7) zien we dat er een duidelijk verschil in nauwkeurigheid tussen de twee methoden, dit valt ook te zien in de figuren hieronder. De afname in onnauwkeurigheid door het gecombineerd voorspellen valt voornamelijk terug te zien in de bias van de voorspelling. De gecombineerde voorspelling laat een beduidend lagere bias zien, wat betekend dat de voorspelling meer in evenwicht is en minder uitschieters heeft.



Figuur 4.5: Geclusterde voorspelling aanmeldingen DE



Figuur 4.6: Enkelvoudige voorspelling aanmeldingen DE

4.2.1.2 De voorspelling van het verwachte aantal aanmeldingen uit de ROW

De aanmeldingen uit de ROW hebben we onderverdeeld in aanmeldingen van studiezoekers en daarbij behorende extra personen. Elk van de deze componenten bevatten een eigen dataset, met eigen eigenschappen (Hoofdstuk 3). De eerste stap van de opstelling van het voorspellingsmodel is het initialiseren van de data uit de initialisatieset (Tabel 4.8).

Open Dagen	Studiezoekers	Extra personen
2013-03	xxxxx	xxxxx
2013-11	xxxxx	xxxxx
2014-03	xxxxx	xxxxx
2014-11	xxxxx	xxxxx

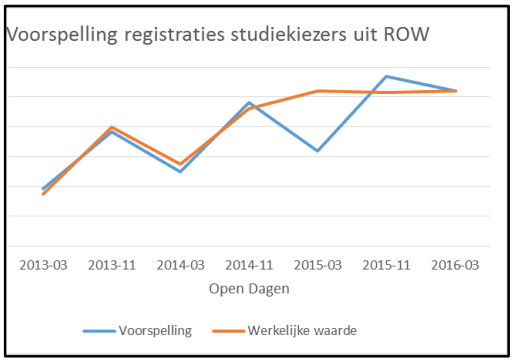
Tabel 4.8: Initialisatieset ROW aanmeldingen

Winters' model schrijft voor om de data uit de initialisatie eerst te deseizoeneren (Hoofdstuk 2). Door het deseizoeneren van de data worden de seizoeninvloeden uit de data gehaald. Deze seizoeninvloeden komen later terug in de voorspelling doormiddel van een seizoenfactor. Voor de voorspelling van het aantal aanmeldingen uit de ROW betekend dit dat er twee seizoenen zijn; een laag seizoen (maart) en een hoog seizoen (november). Nadat de data gedeseizoeneerd is, bepalen we doormiddel van regressie analyse het level in T0 en de trend. Waarna we de seizoenfactoren bepalen, zoals in hoofdstuk 2 beschreven is.

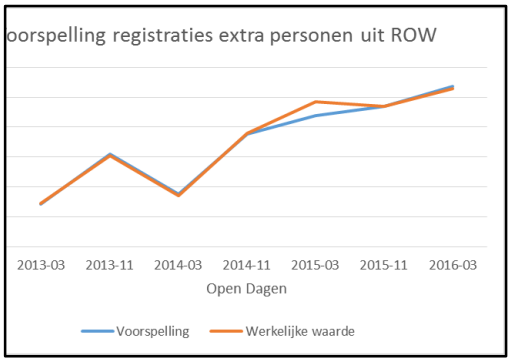
Aanmeldingen uit ROW	Leven in T0	Trend	Seizoenfactor hoog	Seizoenfactor laag
Studiezoekers	xxxxx	xxxxx	xxxxx	xxxxx
Extra personen	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 4.9: Gegevens uit de initialisatieset (Aanmeldingen uit ROW)

In Tabel 4.9 zijn de gegevens weergegeven uit de initialisatieset. Met deze gegevens kan het voorspellingsmodel voor het verwachte aantal registraties uit ROW opgesteld worden. Hiervoor maken we gebruik van Winters' method, zoals beschreven is in hoofdstuk 2. Jaarpunten van de Open Dagen van 2015-03 tot en met 2016-03 gebruiken we als controleset voor de voorspelling.



Figuur 4.7: Voorspelling registraties studiekiezers (ROW)



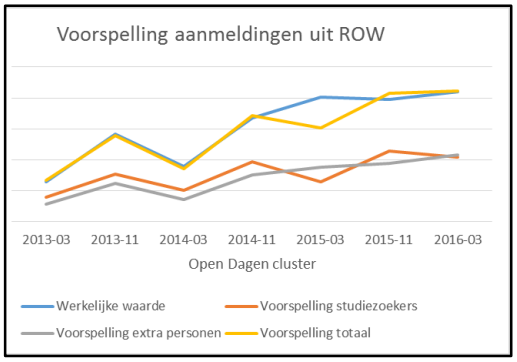
Figuur 4.8: Voorspelling registraties extra personen (ROW)

In Figuur 4.7 is de voorspelling te vinden van het verwachte aantal aanmeldingen van studiezoekers uit ROW. In Figuur 4.8 is de voorspelling te vinden van het aantal daarbij behorende extra personen.

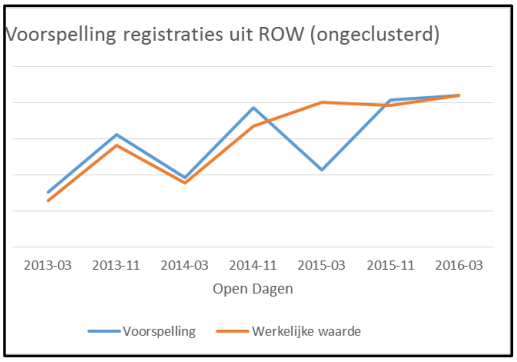
Voorspelling	Studiekiezers	Extra personen	Totaal gecombineerd	Totaal enkelvoudig
Alpha	xxxxx	xxxxx	xxxxx	xxxxx
Beta	xxxxx	xxxxx	xxxxx	xxxxx
Gamma	xxxxx	xxxxx	xxxxx	xxxxx
MAD	xxxxx	xxxxx	xxxxx	xxxxx
st	xxxxx	xxxxx	xxxxx	xxxxx
Mape	xxxxx	xxxxx	xxxxx	xxxxx
Bias	xxxxx	xxxxx	xxxxx	xxxxx
TS	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 4.10: Voorspelling aanmeldingen uit ROW

In Tabel 4.10 zijn de gegevens te vinden van de voorspelling van het aantal aanmeldingen uit ROW. De onnauwkeurigheden zijn bepaald op basis van de controleset. Uit de figuren valt duidelijk op te maken wat het effect is van het adaptief voorspellen van Winters' methode. De alpha, beta en gamma zorgen ervoor dat het model bijgesteld wordt naar de meest optimale waarde van de MAPE, gebaseerd op de jongste gerealiseerde waarde. Het model kan zodoende een periode uitschieten, terwijl de periode daarna deze uitschieter gecorrigeerd wordt. In Figuur 4.9 is de samengestelde voorspelling weergegeven van de aanmeldingen uit ROW en in Figuur 4.10 de enkelvoudige voorspelling. Als we de onnauwkeurigheden van deze twee voorspellingstechnieken vergelijken (Tabel 4.10) zien we dat er een duidelijk verschil in nauwkeurigheid tussen de twee methoden, dit valt ook te zien in de figuren hieronder. De afname in onnauwkeurigheid door het gecombineerd voorspellen valt voornamelijk terug te zien in de bias van de voorspelling, maar ook in de MAPE. De gecombineerde voorspelling laat een beduidend lagere bias zien, wat betekend dat de voorspelling meer in evenwicht is en minder uitschieters heeft. Ook de MAPE is beduidend lager, wat duidt op een minder afwijking van de voorspelling. Dit valt ook terug te zien in Figuur 4.9 en Figuur 4.10.



Figuur 4.9: Geclusterde voorspelling aanmeldingen ROW

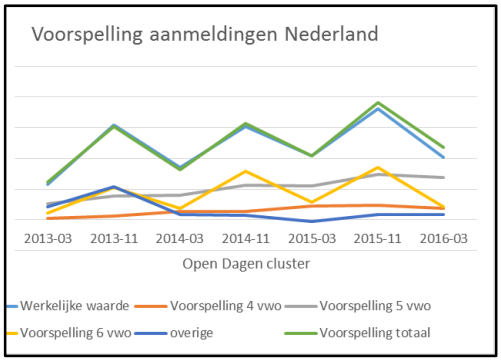


Figuur 4.10: Enkelvoudige voorspelling aanmeldingen ROW

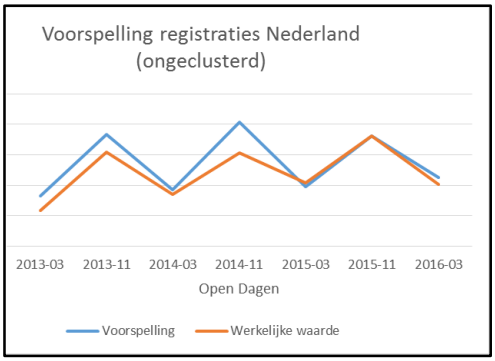
4.2.1.3 De voorspelling van het verwachte aantal aanmeldingen uit Nederland

De aanmeldingen uit Nederland hebben we onderverdeeld in aanmeldingen uit de verschillende jaren uit het VWO en overige aanmeldingen uit Nederland. Deze aanmeldingen hebben we weer verder onderverdeeld in aanmeldingen van studiezoekers en daarbij behorende extra personen. In Appendix 17 zijn de voorspellingen opgesteld van deze verschillende deelcomponenten.

In Figuur 4.11 is de geclusterde voorspelling weergegeven van het aantal aanmeldingen uit Nederland, gebaseerd op een voorspelling van het aantal aanmeldingen van 4, 5, 6 VWO en overige aanmeldingen uit Nederland. In Figuur 4.12 is een enkelvoudige voorspelling weergegeven van het aantal aanmeldingen uit Nederland, gebaseerd op de toepassing van Winters' methode op de cluster gegevens van de aanmeldingen uit Nederland.



Figuur 4.11: Geclusterde voorspelling aanmeldingen uit NL



Figuur 4.12: Enkelvoudige voorspelling aanmeldingen uit NL

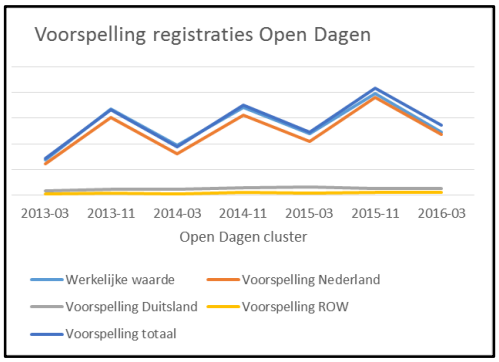
In Tabel 4.11 zijn de onnauwkeurigheden weergegeven van de voorspellingen van de aanmeldingen uit Nederland. Deze onnauwkeurigheden zijn, net zoals bij de voorspelling van aanmeldingen uit Duitsland en de ROW, gebaseerd op de controleset. De gecombineerde voorspelling wijkt gemiddeld gesproken 7,46% af van de werkelijke waarde, terwijl de ongeclusterde voorspelling slechts gemiddeld 5,47% afwijkt. Daarnaast laat de gecombineerde voorspelling een grote positieve afwijking zien (bias) ten opzichte van de enkelvoudige voorspelling. Volgens de gegevens in Tabel 4.11 lijkt het erop dat de enkelvoudige voorspelling nauwkeuriger is, echter moet hier een kanttekening bij geplaatst worden dat de onnauwkeurigheden gebaseerd zijn op de controleset van slechts 3 datapunten. We achten het dan ook te voorbarig om te kunnen concluderen dat een enkelvoudige voorspelling nauwkeuriger is, ten opzichte van een gecombineerde voorspelling. Dit zou in de toekomst moeten uitwijzen, zodra er meer datapunten beschikbaar komen. Daarnaast laten beide modellen duidelijk de werking van Winters' methode zien, waarbij het model leert en zich aanpast aan gemaakte fouten.

Voorspelling	Totaal gecombineerd	Totaal enkelvoudig
Alpha	xxxxx	xxxxx
Beta	xxxxx	xxxxx
Gamma	xxxxx	xxxxx
MAD	xxxxx	xxxxx
st	xxxxx	xxxxx
Mape	xxxxx	xxxxx
Bias	xxxxx	xxxxx
TS	xxxxx	xxxxx

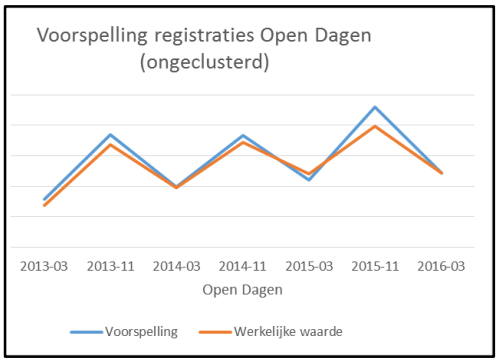
Tabel 4.11: Voorspelling aanmeldingen uit Nederland

4.3 GECOMBINEERDE VOORSPELLING VAN DE OPEN DAGEN

In de bovenstaande paragrafen hebben we de deel componenten voorspeld die de basis vormen voor de voorspelling van het verwachte aantal registraties op de Open Dagen. In Appendix 16 is te zien dat het samenvoegen van deze verschillende componenten een voorspelling oplevert van het verwachte aantal registraties op de Open Dagen. In Figuur 4.13 is de gecombineerde voorspelling weergegeven van het verwachte aantal registraties. Deze voorspelling is opgebouwd uit de voorspelling van het aantal aanmeldingen uit Nederland, Duitsland en de rest van de wereld. In Figuur 4.14 is een enkelvoudige voorspelling weergegeven van het verwachte aantal registraties op de Open Dagen, gebaseerd op de toepassing van Winters' Methode op de gegevens op clusterniveau.



Figuur 4.13: Gecombineerde voorspelling Open Dagen



Figuur 4.14: Enkelvoudige voorspelling Open Dagen

Beide voorspellingstechnieken laten een mate van onnauwkeurigheid zien, zoals te vinden in Tabel 4.12. De gecombineerde voorspelling wijkt over de afgelopen drie Open Dagen gemiddeld 6,35% positief af ten opzichte van de werkelijke waarde. De gecombineerde voorspelling voorspeld dus gemiddeld gesproken 6,35% te veel aanmeldingen voor de Open Dagen, wat neerkomt op gemiddeld 362 aanmeldingen te veel (MAD). De enkelvoudige voorspelling laat een gemiddelde afwijking (MAPE) zien van 8,17%, wat neerkomt op gemiddeld 561 aanmeldingen te veel voorspeld.

Voorspelling	Totaal gecombineerd	Totaal enkelvoudig
MAD	xxxxx	xxxxx
st	xxxxx	xxxxx
Mape	xxxxx	xxxxx
Bias	xxxxx	xxxxx
TS	xxxxx	xxxxx

Tabel 4.12: Gegevens voorspelling Open Dagen

4.4 CONCLUSIE

In dit hoofdstuk hebben we de voorspelling opgesteld voor het verwachte aantal registraties op de bachelor Open Dagen. De voorspelling hebben we opgesteld op basis van de verschillende doelgroepen die we geïdentificeerd hebben in hoofdstuk 3. In hoofdstuk 3 is naar voren gekomen dat iedere doelgroep een bepaald level, trend en seizoeninvloeden bevatten. Zodoende hebben we gekozen om Winter's methode te gebruiken voor de opstelling van de voorspelling. Daarnaast is gebleken dat Winters' methode de nauwkeurigste voorspelling opleverde ten opzichte van de overige onderzochte voorspellingstechnieken. Naast de extrapolatiemodellen, hebben we ook onderzoek verricht naar de mogelijkheden om de voorspelling te baseren op casuele verbanden in het omliggende milieu. Hierbij hebben we onderzoek gedaan naar de verbanden tussen VWO scholieren in Nederland en het aantal aanmeldingen van deze VWO scholieren voor de Open Dagen. Echter, is het niet mogelijk geweest om deze causale verbanden te verifiëren op betrouwbaarheid en daarom hebben we ze niet meegenomen in het werkelijke voorspellingsmodel. Naast de mogelijke correlaties tussen VWO aanmeldingen en VWO scholieren, achten we het ook mogelijk dat andere factoren, zoals andere Open Dagen en marketinguitingen van invloed zijn op het aantal aanmeldingen, en daarmee op de voorspelling. Momenteel hebben we de mogelijke invloed van deze factoren niet onderzocht en daarom hebben we ze niet meegenomen in de opstelling van het werkelijke voorspellingsmodel.

Het aantal aanmeldingen van iedere doelgroep binnen de Open Dagen hebben we apart voorspeld, op basis van het aantal aangemelde studiezoekers en daarbij behorende extra personen. De voorspellingen hebben we opgebouwd aan de hand van een vierstappen methode (*Chopra & Meindl, 2007)*, waarbij we gebruik hebben gemaakt van een initialisatieset en een controleset voor de data. Deze verschillende sets bevatten historische data, waardoor het mogelijk is om het voorspellingsmodel op nauwkeurigheid te controleren en waar nodig bij te sturen.

Het samenvoegen van de verschillende voorspellingen levert betrouwbare voorspelling op die gemiddeld 6,35% afwijkt over de afgelopen drie jaren. Als er gemiddeld gesproken 5.000 aanmeldingen verwacht worden op de Open Dagen, dan wijkt de voorspelling gemiddeld 317 aanmeldingen af van de werkelijke waarde. De ongeclusterde voorspelling laat een hogere onnauwkeurigheid zien van gemiddeld 8,17% afwijking. Dit komt neer op een afwijking van ongeveer 408 aanmeldingen per voorspelling.

In dit hoofdstuk is een voorspellingsmodel opgesteld voor het verwachte aantal aanmeldingen van de Bachelor Open Dagen, met een bepaalde nauwkeurigheid voor de afgelopen drie jaren, en daarmee is deelvraag 3 beantwoord.

Hoofdstuk 5

De voorspelling van de verwachte opkomst

(Deelvraag 4)

“One need only think of the weather, in which case the prediction even for a few days ahead is impossible.”

Albert Einstein (1879-1955)

Duits - Amerikaans natuurkundige

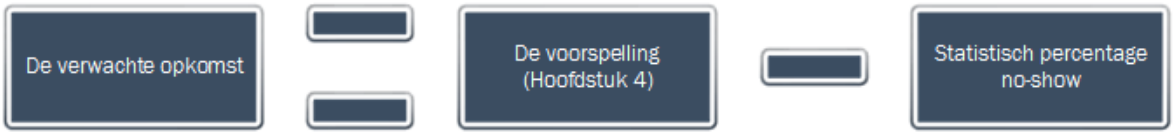
5 DE VERWACHTE OPKOMST

In hoofdstuk 1 hebben we de onderzoeksdoelstelling onderverdeeld in een voorspelling voor het verwachte aantal registraties en een voorspelling voor de verwachte opkomst op de bachelor Open Dagen. In het vorige hoofdstuk hebben we de voorspelling opgesteld voor het verwachte aantal registraties voor de Open Dagen. In hoofdstuk 3 is gebleken dat er een verschil bestaat tussen het aantal registraties en de werkelijke opkomst op de Open Dagen. In dit hoofdstuk gaan we deze percentages no-show bepalen voor de opgestelde voorspelling uit hoofdstuk 4. In de eerste paragraaf zetten we de noodzaak van het percentage no-show uiteen. In paragraaf 2 gaan we de percentages no-show bepalen voor de voorspelling uit het vorige hoofdstuk, zodoende dat we een voorspelling kunnen geven van de verwachte opkomst. We sluiten dit hoofdstuk af met de beantwoording van deelvraag 4.

5.1 DOEL

In hoofdstuk 3 hebben we een inventarisatie gemaakt van de historische data die beschikbaar is van de Open Dagen. Hieruit is gebleken dat de data onderverdeeld kan worden in verschillende doelgroepen, met elk eigen specifieke eigenschappen voor de opstelling van het voorspellingsmodel. Naast de eigenschappen van iedere doelgroep op het gebied van level, trend en seizoenfactoren, blijkt uit de analyse dat iedere doelgroep specifieke percentages no-show met zich meebrengen, die tevens verschillen per voorlichtingsprogramma.

In het vorige hoofdstuk hebben we een voorspelling opgesteld van het verwachte aantal registraties. In hoofdstuk 3 is uiteengezet dat het aantal registraties minus een percentage no-show de opkomst is. Dit verband kunnen we gebruiken voor de bepaling van de verwachte opkomst op de Open Dagen (Figuur 5.1). Door de voorspellingen te minimaliseren met een statistisch percentage no-show, kunnen we de voorspelling van het aantal registraties omzetten in een voorspelling van de verwachte opkomst.



Figuur 5.1: Bepaling van de verwachte opkomst

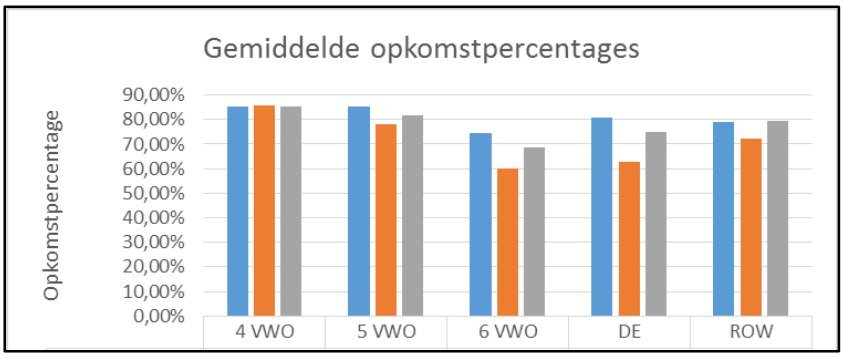
5.2 VERKLARENDE FACTOREN

In de literatuur worden verschillende mogelijkheden gegeven die van invloed kunnen zijn op opkomstpercentages, zoals weersinvloeden of persoonlijke omstandigheden. Voor dit onderzoek hebben we momenteel geen data beschikbaar om deze correlaties te kunnen verifiëren. Gezien de tijd waarvoor dit onderzoek staat, achten we het niet mogelijk om betrouwbare data hiervoor te verzamelen. We zullen zodoende de mogelijke verklarende factoren voor de percentages no-show niet meenemen in de bepaling ervan en we focussen ons dan ook op een statistisch percentage no-show. Naast de mogelijke verklarende factoren in het omliggende milieu, achten we het ook mogelijk dat een bepaald percentage no-show bepaald wordt door spookaanmeldingen en dubbele aanmeldingen. Voor de analyse van de opkomstpercentages hebben we de dataset schoon gemaakt en waar mogelijk dubbele aanmeldingen verwijderd. Dit geeft echter geen zekerheid dat er geen dubbele aanmeldingen of spookaanmeldingen meer in de dataset aanwezig zijn. Zodoende is het mogelijk dat het percentage no-show lager uitvalt dat de werkelijkheid.

5.3 STATISTISCH PERCENTAGE NO-SHOW

In hoofdstuk drie hebben we verschillende doelgroepen geïdentificeerd. Voor elk van deze doelgroepen hebben we een voorspelling opgesteld in het vorige hoofdstuk. Zoals in paragraaf 5.1 beschreven is, bestaat er een verschil tussen het aantal registraties en de werkelijke opkomst op de Open Dagen. Uit de data analyse is gebleken dat iedere doelgroep specifieke eigenschappen laat zien op basis van opkomst. In hoofdstuk drie hebben we deze opkomstpercentages uitgesplitst naar doelgroep en naar voorlichtingsprogramma. Doormiddel van deze gegevens, is het mogelijk om van de voorspellingen van het verwachte aantal registraties, de verwachte opkomst te voorspellen.

In Figuur 5.2 zijn de gemiddelde opkomstpercentages weergegeven van de doelgroepen waarvoor een voorspelling opgesteld is. Deze gemiddelde opkomstpercentages kunnen gebruikt worden om van de doelgroep voorspelling uit hoofdstuk vier de werkelijke opkomst te voorspellen.



Figuur 5.2: Gemiddelde opkomstpercentages

De bovenstaande opkomstpercentages zijn bepaald, zoals vermeld in hoofdstuk 3, doormiddel van een simple-average voorspellingstechniek. Hierbij hebben we de gemiddelde opkomsten genomen per doelgroep over de afgelopen twee Open Dagen, omdat vanaf het najaar van 2015 de gegevens digitaal bijgehouden worden. In de jaren daarvoor werd gebruik gemaakt van handtellingen tijdens de Open Dagen om de opkomst te bepalen. Deze handtellingen hebben we niet meegenomen in de bepaling van het percentage no-show, omdat hier geen onderscheid wordt gemaakt tussen de verschillende componenten, die wel aanwezig zijn in de digitale gegevens.

5.4 CONCLUSIE

In dit hoofdstuk hebben we een model opgesteld die gebruikt kan worden voor de bepaling van de verwachte opkomst op de Open Dagen. Dit hebben we gedaan door gebruik te maken van de voorspellingen uit hoofdstuk 4 en hiervan een bepaald percentage no-show af te halen. Dit percentage no-show hebben we bepaald aan de hand van een simple-average voorspellingstechniek, waarbij we de gemiddelde opkomstpercentages per doelgroep over de afgelopen twee Open Dagen hebben genomen. Het is niet mogelijk geweest om causale verbanden tussen de opkomst en variabelen in het omliggende milieu te kunnen verifiëren, hoewel we wel aannemen dat deze er kunnen zijn. Daarnaast verwachten we dat spookaanmeldingen en dubbele aanmeldingen ook van invloed zijn op de percentages no-show. In de data analyse hebben we de data schoon gemaakt, maar dit sluit niet uit dat in de berekening van het percentage no-show nog spookaanmeldingen en dubbele aanmeldingen zijn meegenomen.

Hoofdstuk 6

De voorspelling van de Open Dagen

(Hoofdvraag)

“The best way to predict the future is to create it.”

Peter Drucker (1909-2005)

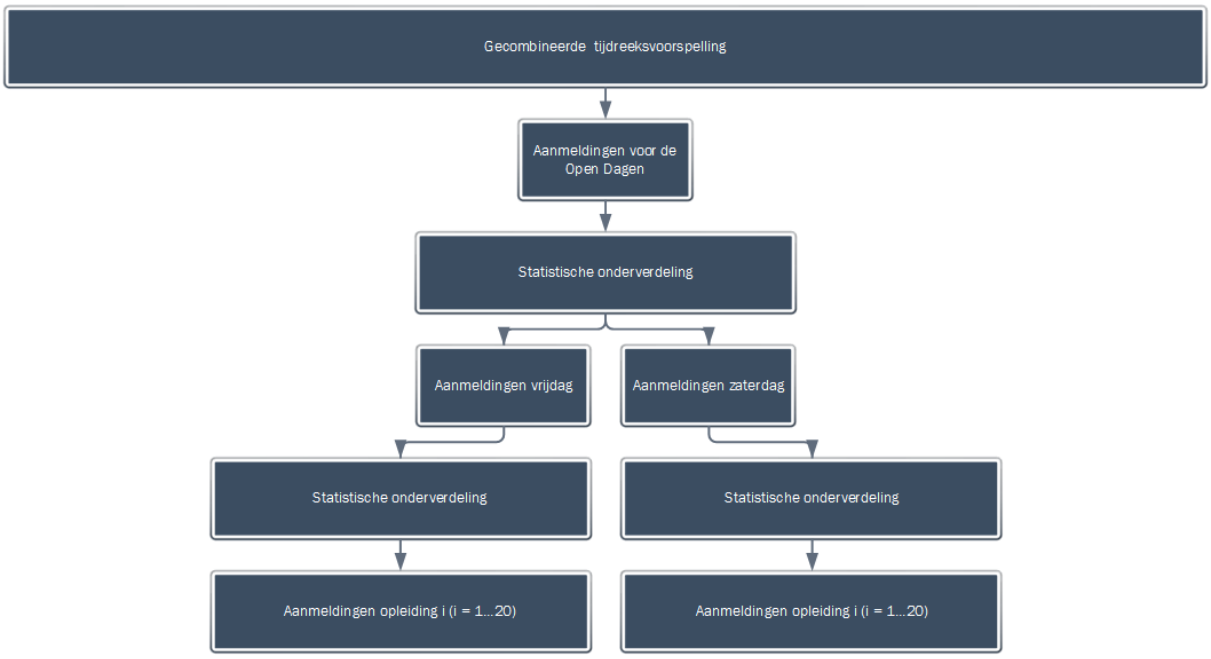
Oostenrijks - Amerikaans schrijver

6 DE VOORSPELLING VOOR DE OPEN DAGEN

In dit hoofdstuk gaan we de verschillende voorspellingsmodellen samenvoegen, om zodoende een voorspelling te maken van het verwachte aantal registraties en de verwachte opkomst op de Open Dagen. In de eerste paragraaf gaan we de methode beschrijven die we gebruiken voor het samenvoegen van de verschillende voorspellingen en de statistische onderverdelingen. In de paragraaf die daarop volgt gaan we dezelfde methode toepassen voor de toekomstige Open Dagen van november 2016, waarbij we de voorspelling voor deze Open Dag opstellen met bepaalde betrouwbaarheidsintervallen voor de opleidingen. We sluiten dit hoofdstuk af met een samenvattende conclusie.

6.1 DE METHODE

In hoofdstuk vier hebben we de methode toegelicht voor de opstelling van de voorspelling van het verwachte aantal registraties (Appendix 16). Hierbij hebben we ook aandacht besteed aan het statistisch onderverdelen van het verwachte aantal aanmeldingen naar dag- en voorlichtingsniveau. In hoofdstuk 3 hebben we de verdelingen geanalyseerd van de verschillende doelgroepen op zowel de vrijdag, als de zaterdag. Hieruit is gebleken dat voor iedere doelgroep de verhouding aanmeldingen voor vrijdag-zaterdag een (ongeveer) stabiel niveau laat zien. Daarnaast hebben we de aanmeldingen voor zowel de vrijdag, als de zaterdag verder uitgesplitst naar voorkeuren voor bepaalde voorlichtingsprogramma's. Hieruit is gebleken dat hier bepaalde trends zichtbaar zijn en dat iedere doelgroep voorkeuren heeft voor bepaalde voorlichtingsprogramma's. Deze statistische gegevens maakt het mogelijk dat we de voorspelling uit hoofdstuk 4 kunnen onderverdelen naar een voorspelling op zowel dag-, als voorlichtingsniveau (Figuur 6.1). Met dit inzicht op voorlichtingsniveau, kunnen de planners van de Open Dagen de verwachte aantal aanmeldingen verdelen over de verschillende dagdelen, waarmee een efficiënte zaalplanning bewerkstelligd kan worden.



Figuur 6.1: Statistische onderverdeling voorspelling

6.2 STATISTISCHE ONDERVERDELING

De eerste stap die we zetten in de onderverdeling van de voorspelling, is het onderverdelen naar dag niveau. In hoofdstuk 3 hebben we van de afgelopen jaren de verhoudingen bepaald tussen aanmeldingen voor de vrijdag en aanmeldingen voor de zaterdag. Door gebruik te maken van een simple-average voorspellingstechniek, kunnen we voor de toekomstige Open Dagen ook de verhoudingen voorspellen (Tabel 6.1).

	Vrijdag Studiezoekers	Zaterdag Studiezoekers	Deviatie	Vrijdag Extra personen	Zaterdag Extra personen	Deviatie
4 VWO	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx
5 VWO	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx
6 VWO	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx
Overige	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx
Duitsland	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx
EER/ROW	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 6.1: Verdeling doelgroepen naar dag

In Tabel 6.1 zijn de gemiddelde verhoudingen weergegeven tussen het aantal aanmeldingen voor de vrijdag en de zaterdag, als percentage van het totaal aantal aanmeldingen per doelgroep. Deze verhoudingen zijn gebaseerd op de gemiddelde verhouding over de afgelopen vier Open Dagen, omdat vanaf het najaar van 2014 de Open Dagen op de donderdag zijn komen te vervallen. Hierdoor geven de verhoudingen van de Open Dagen voor najaar 2014 een vertekend beeld weer voor de toekomst en nemen we deze dan ook niet mee in de voorspelling.

Voorlichtingsniveau

De volgende stap die we zetten in de onderverdeling van de voorspelling, is de onderverdeling van dagniveau naar voorlichtingsniveau. In hoofdstuk 3 zijn de voorkeuren bepaald van iedere doelgroep op het gebied van voorlichtingsprogramma's. Hieruit is naar voren gekomen dat iedere doelgroep eigen specifieke voorkeuren laat zien voor bepaalde voorlichtingsprogramma's. Deze voorkeuren verschillen voor zowel de vrijdag, als de zaterdag. In Appendix 18 zijn de gemiddelde voorkeuren weergeven over de afgelopen vier Open Dagen, met de daarbij behorende standaard deviatie. De percentages in de tabellen geven het percentage registraties van het totaal weer, dat zich heeft geregistreerd voor een bepaald voorlichtingsprogramma. De gesommeerde percentages komen dus boven de 100% uit, gezien studiezoekers zich voor meer dan één voorlichtingsprogramma kunnen registreren. Momenteel worden de studiezoekers door de organiserende afdeling onderverdeeld naar voorlichtingsronde. Het is zodoende dus niet relevant om de studiezoekers verder onder te verdelen naar voorlichtingsronde, hierdoor zal waarschijnlijk de onnauwkeurigheid alleen maar toenemen.

6.3 DE VOORSPELLING VAN TOEKOMSTIGE OPEN DAGEN

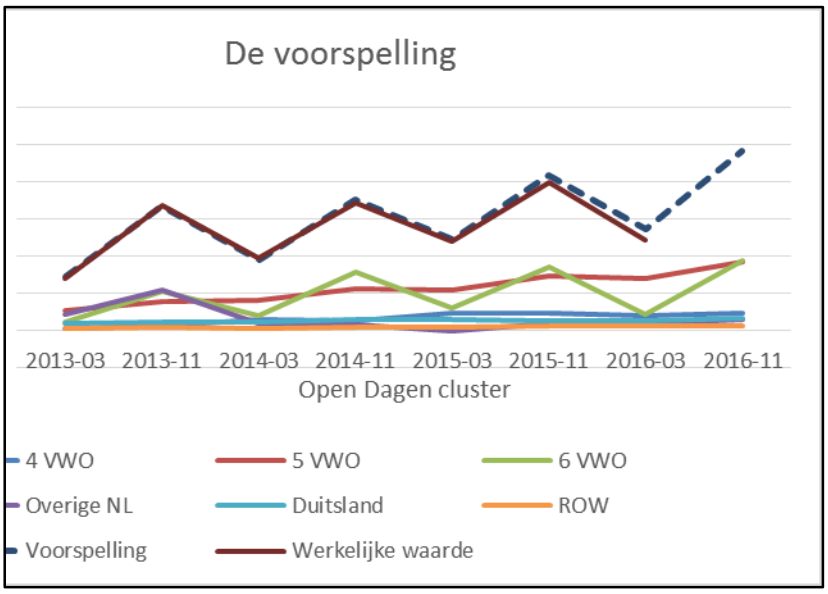
De voorspelling van het aantal aanmeldingen is opgebouwd uit verschillende componenten, de samenvoeging van deze componenten levert de voorspelling op van het verwachte aantal registraties op de Open Dagen. Voor de opstelling en de berekening van de voorspelling hebben we de verschillende voorspellingen samengevoegd doormiddel van Excel (Appendix 20).

De voorspelling van de Open Dagen van november 2016

In Tabel 6.2 zijn de voorspelde registratiecijfers weergegeven van de gecombineerde voorspelling van de Open Dagen van november 2016. De voorspellingen zijn in de tabel uitgesplitst naar de voorspelling van iedere doelgroep en de daarbij behorende studiezoekers en extra personen. In Figuur 6.2 is een grafische weergave weergeven van de voorspelling van het aantal registraties. De stippellijn geeft de voorspelling weer. Uit het figuur valt af te leiden dat voor de toekomstige Open Dagen van november 2016 een hoger aantal registraties wordt verwacht ten opzichte van vorig jaar november.

	4 VWO	5 VWO	6 VWO	Overige NL	Duitsland	ROW	Totaal
Studiezoekers	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx
Extra personen	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx
Totaal	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 6.2: Voorspelling voor de Open Dagen van 2016-11

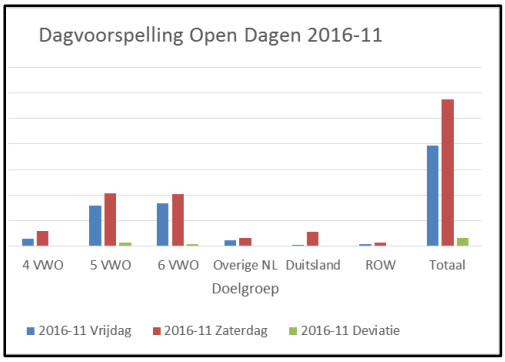


Figuur 6.2: De voorspelling van 2016-11

De gecombineerde voorspelling van het totaal aantal registraties voor de Open Dagen cluster kan verder worden onderverdeeld naar dagniveau (Tabel 6.3). De onderverdeling van het aantal aanmeldingen gebeurd op de manier zoals beschreven in paragraaf 6.2. De deviatie is de spreiding die is waargenomen over de afgelopen jaren in de verhoudingen tussen de twee dagen. Bij de registraties van 4 VWO worden XXX personen op de vrijdag verwacht, met een deviatie van XX personen. Dit betekend dat uit de historische gegevens blijkt dat er gemiddeld XX personen zich voor de vrijdag zullen registreren, met een marge van XX personen. Bij deze verdelingen moet wel rekening gehouden worden met de nauwkeurigheid van de voorspelling. De gecombineerde voorspelling laat over de afgelopen perioden een onnauwkeurigheid van 6,35% zien. Deze onnauwkeurigheid komt dus bovenop de verwachte aantal aanmeldingen per dag. In Figuur 6.3 is zijn de dagverdelingen grafisch weergegeven.

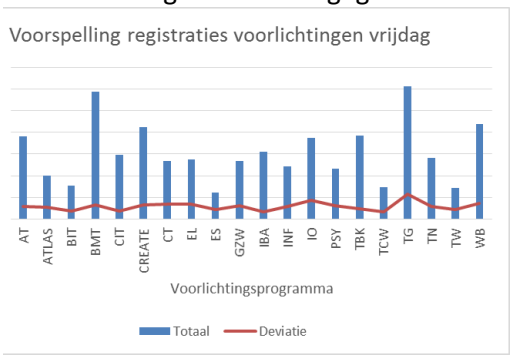
Dag:	4 VWO	5 VWO	6 VWO	Overige NL	Duitsland	ROW	Totaal
Vrijdag	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx
Zaterdag	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx
Deviatie	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 6.3: Dagverdeling voorspelling 2016-11

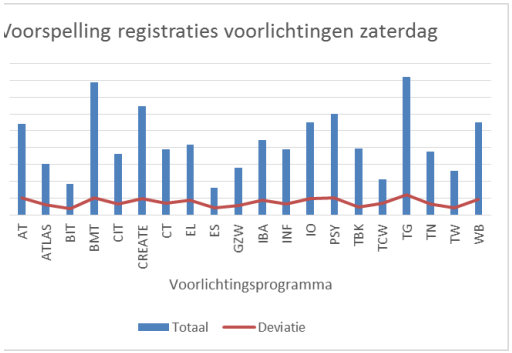


Figuur 6.3: Grafische weergave dagverdeling voorspelling 2016-11

De statistische onderverdeling van de gecombineerde voorspelling per dag kunnen we verder statistisch onderverdelen naar voorlichtingsniveau. In Figuur 6.4 is een grafische weergave weergegeven van de verwachte aantal aanmeldingen voor de verschillende voorlichtingsprogramma's op de vrijdag. In Figuur 6.5 is de weergave te vinden voor de zaterdag. De rode lijn geeft de deviatie, oftewel de statistische onnauwkeurigheid weer van de voorspelling. Deze deviatie is gebaseerd op historische gegevens. De voorspelling per voorlichtingsronde is het gemiddelde percentage studiezoekers en daarbij behorende extra personen dat zich heeft ingeschreven voor de betreffende voorlichting. De deviatie is de gemiddelde afwijking per Open Dag weer. Hierbij moet rekening gehouden met de onnauwkeurigheid van de voorspelling en de deviatie van de dagverdelingen. In Appendix 19 zijn zowel de registratiecijfers per voorlichtingsronde, als de verwachte opkomstcijfers per voorlichtingsronde weergegeven.



Figuur 6.4: Grafische weergave voorspellingen naar voorlichtingsniveau vrijdag 2016-11



Figuur 6.5: Grafische weergave voorspellingen naar voorlichtingsniveau zaterdag 2016-11

6.4 CONCLUSIE

In dit hoofdstuk hebben we de verschillende voorspellingen van de Open Dagen samengevoegd en hebben we de voorspelling uitgevoerd voor de eerstvolgende Open Dagen van november 2016. Dit hebben we aan de hand gedaan van het statistisch onderverdelen van de gecomponeerde voorspelling uit hoofdstuk 4. Uit de data analyse (hoofdstuk 3) is gebleken dat iedere doelgroep specifieke voorkeuren heeft voor bepaalde dagen en voorlichtingsprogramma's. Deze data eigenschappen hebben we gebruikt voor het statistisch onderverdelen van de gecomponeerde voorspelling naar een voorspelling op dagniveau en uiteindelijk naar een voorspelling voor de verschillende voorlichtingsprogramma's. Hoe verder we onder gaan verdelen, hoe meer deviatie in de voorspelling komt te zitten. De gecombineerde cluster voorspelling heeft over de afgelopen periodes gemiddeld 6,35% te positief gepresteerd. Deze geclusterde voorspelling met een bepaalde nauwkeurigheid hebben we verder onderverdeeld naar dagniveau. Deze onderverdeling zorgt ook voor extra onnauwkeurigheid, door variatie van de verdelingen over de afgelopen jaren. Hetzelfde geldt voor de verdere onderverdeling naar voorlichtingsniveau. Voor het gebruik van de gegevens is het dus noodzakelijk om hier rekening mee te houden.

Hoofdstuk 7

Conclusies en aanbevelingen

“With high hope for the future, no prediction is ventured.”

Abraham Lincoln (1809-1865)

16th president van de VS

7 CONCLUSIES EN AANBEVELINGEN

Dit hoofdstuk bevat de beantwoording van de hoofdvraag over hoe vooraf nauwkeurig inzicht gegeven kan worden in het verwachte aantal registraties en de verwachte opkomst op de Bachelor Open Dagen. De hoofdvraag beantwoorden we doormiddel van de vier opgestelde deelvragen, die in de voorgaande hoofdstukken beantwoord zijn. Daarnaast zijn we bij de beantwoording van de deelvragen aanbevelingen tegen gekomen voor eventueel toekomstige werkzaamheden en verbeteringen. Deze aanbevelingen gaan we in paragraaf twee toelichten en behandelen.

7.1 CONCLUSIES

In de afgelopen 4 hoofdstukken hebben we de verschillende deelvragen beantwoord. In hoofdstuk 2 zijn we begonnen met een theoretisch onderzoek, waarbij we onderzocht hebben hoe efficiënt en succesvol voorspeld kan worden. In het hoofdstuk dat volgde zijn we begonnen met het verkennen en verkrijgen van noodzakelijke data voor het onderzochte voorspellingsproces en daarmee voor de opstelling van de werkelijke voorspelling. In dit hoofdstuk hebben we de data geanalyseerd en hebben we zes verschillende doelgroepen geïdentificeerd in de data, namelijk aanmeldingen uit de verschillende VWO jaargangen en aanmeldingen uit Duitsland en de rest van de wereld. Deze doelgroepen hebben we verder onderverdeeld in aanmeldingen van studiezoekers en de daar bijhorende extra personen. Na analyse van deze componenten, zijn we er achter gekomen dat iedere component van iedere doelgroep eigen specifieke trends en seizoeninvloeden bevat. Deze data was noodzakelijk voor hoofdstuk 4, waarbij we het werkelijke voorspellingsmodel hebben opgesteld. Voor de opstelling van het voorspellingsmodel hebben we gebruik gemaakt van de data die geanalyseerd is in hoofdstuk 3. In dit hoofdstuk hebben we de theorie gelinkt met de data analyse en is naar voren gekomen dat Winter's methode het meest toepasbaar is voor ons voorspellingsmodel. Het voorspellingsmodel hebben we opgebouwd volgens verschillende componenten, bestaande uit de verschillende doelgroepen. Deze voorspelling leverde een inzicht op met betrekking op het verwachte aantal registraties voor de Open Dagen, met een bepaalde onnauwkeurigheid over de afgelopen jaren. In het hoofdstuk dat volgde hebben we een model opgesteld voor de voorspelling van de werkelijke opkomstcijfers. Hierbij hebben we gebruik gemaakt van een simple-average methode, waarbij we de gemiddelde opkomstpercentages hebben genomen over de afgelopen twee Open Dagen. Deze gemiddelde opkomstpercentages kunnen we, in combinatie met de voorspelling van het verwachte registraties, gebruiken om de verwachte opkomst te voorspellen. In hoofdstuk 3 hebben we de opkomstcijfers geanalyseerd tot aan voorlichtingsniveau, die we kunnen gebruiken voor het voorspellen van de verwachte opkomst per voorlichtingsronde. In hoofdstuk 6 hebben we deze opkomstcijfers gelinkt aan het voorspellingsmodel uit hoofdstuk 4 en hebben zodoende een voorspelling opgesteld voor de Open Dagen van november 2016.

Doormiddel van het beantwoorden van de verschillende deelvragen hebben we de hoofdvraag kunnen beantwoorden:

'Op welke wijze kan vooraf nauwkeurig inzicht worden gegeven in het verwachte aantal registraties en de verwachte opkomst op de Bachelor open dagen en op welke wijze kan dit inzicht leiden tot een efficiënte bijdrage aan de instroomdoelstellingen van de UT en de bacheloropleidingen daarbinnen?'

Door het voorspellen van verschillende doelgroepen, hebben we in hoofdstuk 6 een voorspelling kunnen opstellen van het verwachte aantal registraties en de verwachte opkomst voor de toekomstige Open Dagen, met een bepaalde onnauwkeurigheid. Door deze cluster voorspelling statistisch onder te verdelen naar individuele dagen, hebben we een voorspelling gemaakt voor het verwachte aantal registraties op zowel de vrijdag, als de zaterdag. Deze verwachte registratiecijfers per dag hebben we verder statistisch onderverdeeld naar voorlichtingsniveau, waarbij we een voorspelling hebben gemaakt voor de verwachte aantal registraties per voorlichtingsprogramma per dag. Door gebruik te

maken van een simple-average methode voor het bepalen van de opkomstpercentages, hebben we van deze opgestelde voorspelling de verwachte opkomst per voorlichtingsprogramma bepaald. Met dit inzicht is het mogelijk voor de organisatie van de Open Dagen om het verwachte aantal registraties per dag op te delen in dagdelen en hiervoor de geschikte zalen te selecteren, met in het oog houdend de instroomdoelstellingen van de universiteit.

7.2 AANBEVELINGEN

Naar aanleiding van het onderzoek hebben we een vijftal aanbevelingen opgesteld, die in onze ogen noodzakelijk zijn voor een nauwkeurige voorspelling óf kunnen bijdragen aan het verbeteren van de nauwkeurigheid van toekomstige voorspellingen:

- De opgestelde voorspelling is opgebouwd uit historische gegevens. Er is gebleken dat gegevens uit een nabijer verleden van grotere waarde zijn voor de nauwkeurigheid van de voorspellen ten opzichte van gegevens uit een ververleden. Het is daarom voor de nauwkeurigheid van de voorspelling aan te raden het voorspellingsmodel telkens aan te passen naar aanleiding van de laatste geobserveerde gegevens, zodat afwijkingen in de trends snel geobserveerd kunnen worden en het voorspellingsmodel kan worden aangepast om de nauwkeurigheid voor toekomstige voorspellingen hoog te houden.
- Voorspellingen kunnen altijd een mate van onnauwkeurigheid met zich meebrengen en zijn dus geen waarzeggerij. Het is daarom van belang om altijd rekening te houden met deze mogelijke onnauwkeurigheid als er beslissingen worden genomen op basis van de voorspellingen.
- In de verkenning van de data is gebleken dat bepaalde doelgroepen bij bepaalde voorlichtingsprogramma's een relatief hoog percentage no-show laten zien, terwijl verklaringen hiervoor momenteel niet onderzocht zijn. We verwachten dat vervuiling in de data een belangrijke invloed hierop heeft, zoals dubbele aanmeldingen en spookaanmeldingen. Voor toekomstig onderzoek raden we aan om de mogelijke factoren voor de percentages no-show te onderzoeken om zo dit percentage eventueel te kunnen verlagen.
- In de opstelling van het voorspellingsmodel was het voor ons niet mogelijk om causale verbanden te verifiëren, terwijl we wel aanwijzingen vonden voor mogelijke causale verbanden met betrekking op het aantal VWO scholieren in Nederland. Voor toekomstig onderzoek raden we dan ook aan om onderzoek te verrichten naar de mogelijke relaties tussen het aantal aanmeldingen/opkomst en factoren in het omliggende milieu, zoals aantal VWO scholieren, marketingactiviteiten en weersinvloeden.
- Het huidige model hebben we gebaseerd op het inzichtelijk maken van de verwachte registratie- en opkomstcijfers voordat het registratieproces geopend is. Dit is gedaan omdat de zaalplanningen voor de opening van het registratieproces globaal vast staan. Voor toekomstig onderzoek raden we aan om naast dit model, ook een onderzoek te verrichten naar de mogelijkheden van een continue voorspellingmodel, waarbij het mogelijk zou zijn om iedere dag gedurende het registratieproces de verwachte aantal aanmeldingen op de Open Dagen te voorspellen. Een model gebaseerd op het "Nearest Neighbour Principle" lijkt ons geschikt daarvoor. We verwachten dat hoe dichter bij de Open Dagen, hoe nauwkeuriger het aantal aanmeldingen voorspeld kunnen worden. Het organiserende team kan zo waar nodig gepaste maatregelen nemen voor eventuele achterblijvers of uitschieters.

REFERENTIES

Afdeling Marketing & Communicatie. (2016, 03 08). Opgehaald van Universiteit Twente: https://www.utwente.nl/mc/

Chopra, S., & Meindl, P. (2015). *Supply Chain Management: Strategy, Planning and Operations.*

Clemen, R. T. (1989). Combining forecasting: A review and annotated bibliography. *International journal of forecasting*, 559.

E. Joseph. (1992). "Quality Approaches to Long-Range Forecasts". *Futurics: A Quarterly Journal of Futures Research Vol. 16, Nos. 3 & 4,*, p. 14.

Feiten & Cijfers. (2016, 03 09). Opgehaald van Universiteit Twente: https://www.utwente.nl/organisatie/feiten-en-cijfers/

Feiten over de UT. (2016, 03 08). Opgehaald van Universiteit Twente: https://www.utwente.nl/organisatie/over-de-ut/

Forecasting Methods. (2016, 03 20). Opgehaald van Unité de gestion logisitique et modélisation (POMS): http://www.poms.ucl.ac.be/etudes/notes/prod2100/cours/Part%206-Forecast.pdf

Gatrell JD, B. G. (2002). Weather and voter turnout. Kentucky primary and general elections, 1990–2000. . *Southeast Geogr. 2002;42:114–134.*, 114–134.

Graefe, A. A. (2013). Combined forecasts of the 2012 election: The PollyVote. Foresight . *Foresight - The International Journal of Applied Forecasting*, 50-51.

Heerkens, H. &. (2012). *Geen probleem.*

Heroman, W. D. (2012). Demand Forecasting and Capacity Mangement in Primary Care. *American college of Physician Executives*.

Hopp, W. &. (2001). *Factory Physics: Foundations of Manufacturing Management.*

http://www.cbs.nl. (2016). Opgehaald van Centraal Bureau voor de Statistiek: http://www.cbs.nl

http://www.vandale.nl. (2016). Opgehaald van Van Dale.

Introduction to forecasting. (2016, 03 20). Opgehaald van Stockholm University: http://gauss.stat.su.se/gu/e/slides/MJK_ch1.pdf

Makridakis, S. &. (1989). Forecasting Methods for Management.

Ozcan, Y. (2009). *Quantitative Methods in Health Care Management.* San Francisco, USA: Jossey-Bass.

Pidd, M. (2003). Tools for Thinking: Modelling in Management Science.

Robinson, S. (2004). *Simulation: The Practice of Model Development and Use. .* Hoboken, USA: John Willy & Sons Inc .

Saffo, P. (2007). Six Rules for Effective Forecasting. *Harvard Business Review*.

Smith, J. C. (1971). How to Choose the Right Forecasting Technique. *Harvard Business review*.

Sunil Chopra, P. M. (2007). *Supply Chain Management: Strategy, Planning, and Operation.* Pearson Prentice Hall.

Wiener, N. (1949). *Extrapolation, Interpolation and Smoothing stationairy time series, with engineering application.* Cambridge, Massachusetts: The M.I.T. Press.

Winston, W. (2004). *Operations Research: Applications and Algorithms.* Belmont, USA: Brooks/Cole.

Zhang, G. (2001). Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. *Neurocomputing 50*, 159-175.

APPENDICES

APPENDIX 1: PROBLEEMINVENTARISATIE

De problemen hebben we geïdentificeerd aan de hand van gesprekken met de betrokken medewerkers van de afdeling Marketing & Communicatie (M&C). Hierbij hebben we gebruik gemaakt van de gepubliceerde opdrachtomschrijving als vooronderzoek. Voor de volledigheid hebben we naast de waargenomen problemen, ook oorzaken vermeld die niet als problemen waargenomen worden door de probleemhebbers.

Waargenomen probleem:	**Uitleg:**
Het aantal beschikbare zalen, met passende omvang, ligt niet in lijn met de benodigde capaciteiten.	Het aantal aanmeldingen voor sommige voorlichtingsrondes groeien. Deze groei brengt met zich mee dat er grotere zalen nodig zijn, om alle belangstellenden in te kunnen onderbrengen.
De ratio en ondersteuning voor zaalplanning op basis van voorspellingen van het aantal registraties ontbreekt.	Momenteel wordt de zaalplanning voornamelijk gebaseerd op historie en smaak, en wordt er geen gebruik gemaakt van wetenschappelijke onderbouwingen op basis van voorspellingen.
De beschikbare zalen worden momenteel ingedeeld op basis van locatie, historie en smaak.	Doordat er momenteel geen onderbouwing is op basis van voorspellingen, wordt de zaalplanning momenteel nog gebaseerd op locatie, historie (‘wij hebben hier altijd al gezeten’) en smaak (‘deze zaal past niet bij de uitstraling van de opleiding’).
Onhoudbare situaties in de organisatie en onveilige situaties, waarbij zaalcapaciteiten niet gerespecteerd werden.	Door onder andere het gebrek aan voorspellingen, is het in het verleden voorgekomen dat er meer mensen werkelijk aanwezig waren, dan aangemeld.
Ieder programmaonderdeel heeft een maximumcapaciteit toegewezen gekregen.	Doordat er onveilige situaties zich hebben kunnen voordoen, wordt momenteel gebruik gemaakt van een maximumcapaciteit per programmaonderdeel, gebaseerd op een bepaald percentage no-show.
Opleidingsdirecteuren vinden dat de maximumcapaciteiten niet op alle fronten passen bij de hoge instroomdoelstellingen.	Enkele opleidingsdirecteuren zijn van mening dat door de maximumcapaciteiten het mogelijk is dat studiezoekers niet terecht kunnen bij een bepaalde voorlichtingsronde en dat er dus eventueel een potentiele student niet de opleiding gaat volgen, wat niet in lijn ligt met de instroomdoelstellingen.
Mogelijke inefficiëntie in de indeling van de zalen	Doordat momenteel de zaalindeling wordt gebaseerd op basis van locatie, historie en smaak, wordt er niet gekeken naar of deze indelingen ook maximaal gebruik maken van de beschikbare zalen.
Mogelijke verspilling van kostbare personeels- en andere kosten.	Naast de zaalplanning die gebaseerd wordt op basis van locatie, historie en smaak, worden personeels- en andere middelen momenteel ook gebaseerd op een indicatie van het

	verwachte aantal aanmeldingen, terwijl hier geen basis voor is op basis van voorspellingen.

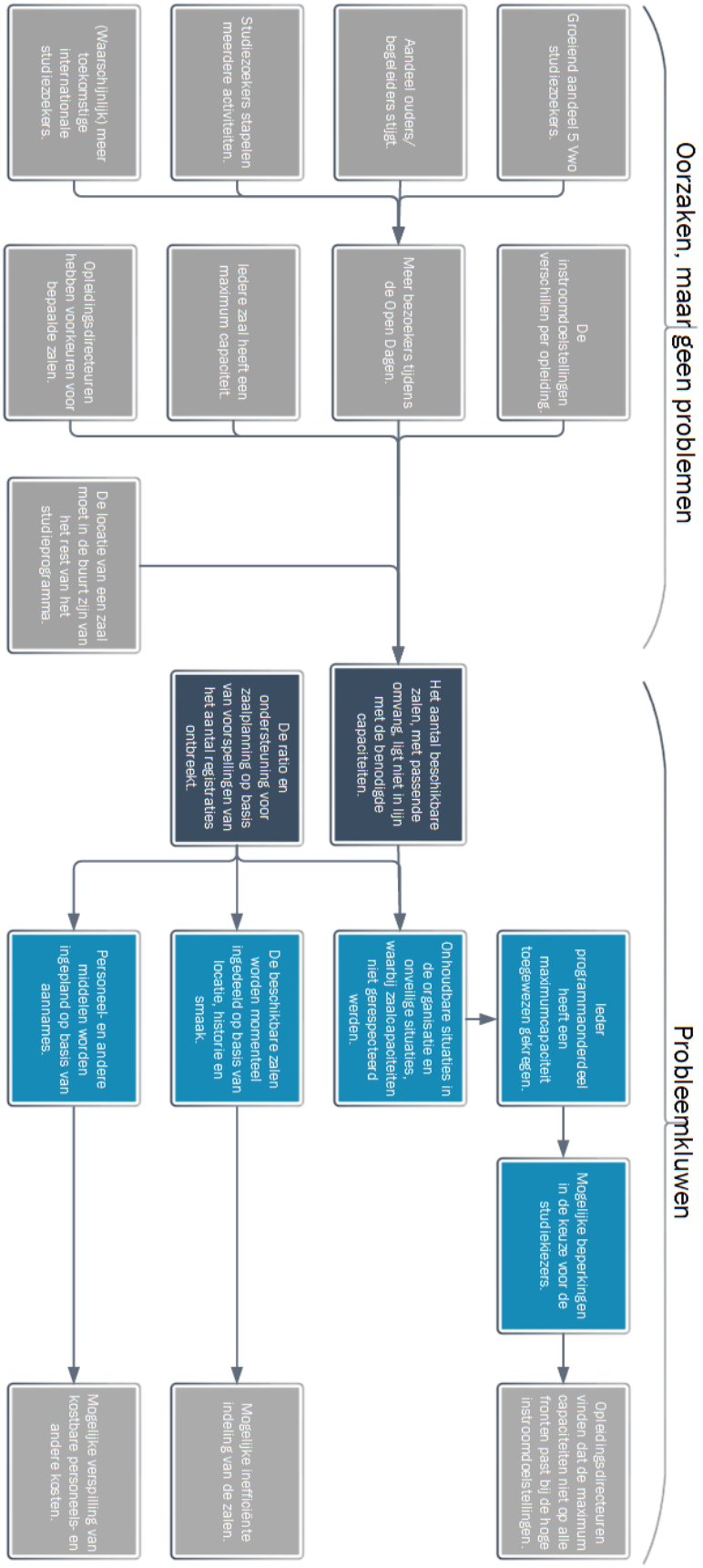
Tabel 0.1: Waargenomen problemen

Naast de bovenstaande geïdentificeerde problematiek, hebben we ook oorzaken waargenomen, die niet gezien worden als een probleem, maar wel invloed hebben op de problematiek.

Waargenomen oorzaken, maar geen problemen:
Groeiend aandeel 5 VWO studiezoekers.
Aandeel ouders/begeleiders stijgt.
Studiekiezers stapelen meerdere activiteiten.
Waarschijnlijk meer toekomstige internationale studiezoekers.
De instroomdoelstellingen verschillende per opleiding.
Meer bezoekers tijdens de Open Dagen.
Iedere zaal heeft een maximumcapaciteit.
Opleidingsdirecteuren hebben voorkeuren voor bepaalde zalen.
De locatie van een zaal moet in de buurt zijn van de rest van het studieprogramma.

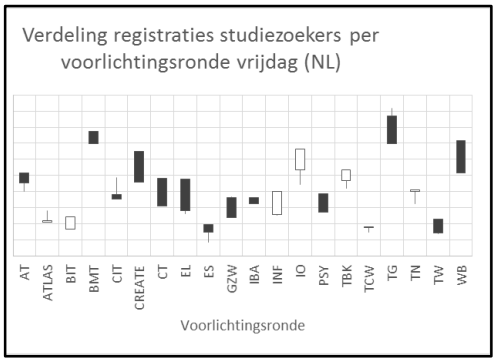
Tabel 0.2: Waargenomen oorzaken

APPENDIX 2: UITGEBREIDE PROBLEEMKLUWEN

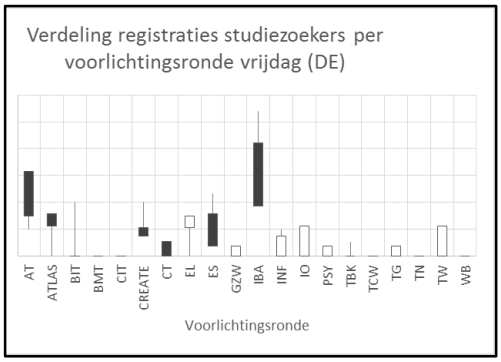


Figuur 0.1: Probleemkluwen

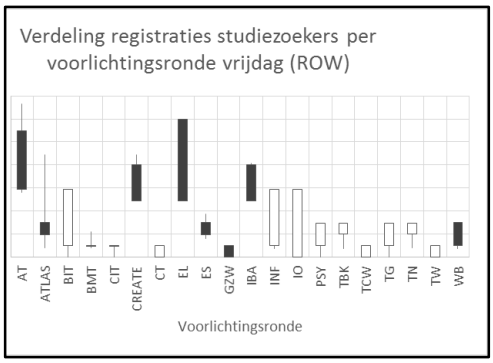
APPENDIX 3: VOORKEUREN STUDIEZOEKERS VRIJDAG (LAND VAN HERKOMST)



Figuur 0.2: Voorkeuren studiezoekers vrijdag (NL)

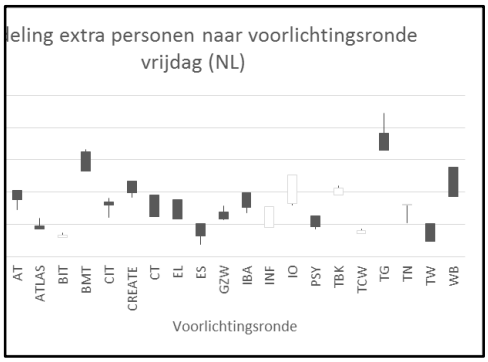


Figuur 0.3: Voorkeuren studiezoekers vrijdag (DE)

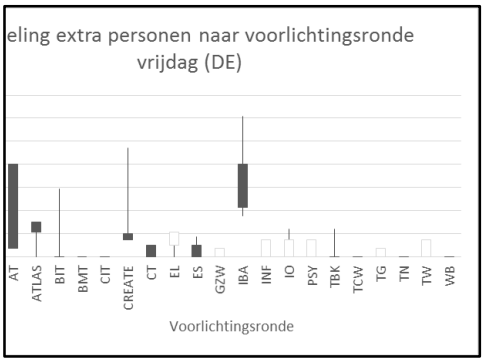


Figuur 0.4: Voorkeuren studiezoekers vrijdag (ROW)

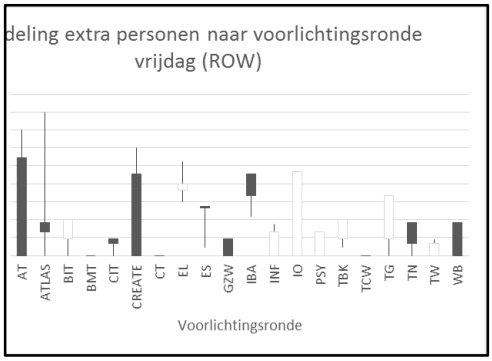
APPENDIX 4: VOORKEUREN EXTRA PERSONEN VRIJDAG (LAND VAN HERKOMST)



Figuur 0.5: Voorkeuren extra personen vrijdag (NL)

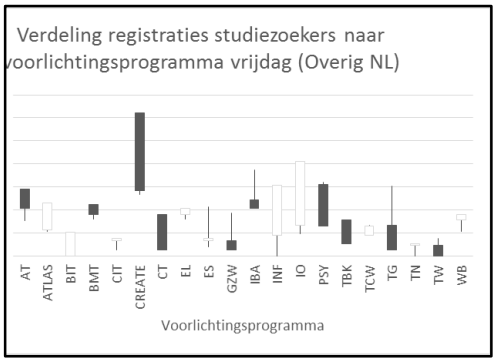


Figuur 0.6: Voorkeuren extra personen vrijdag (DE)

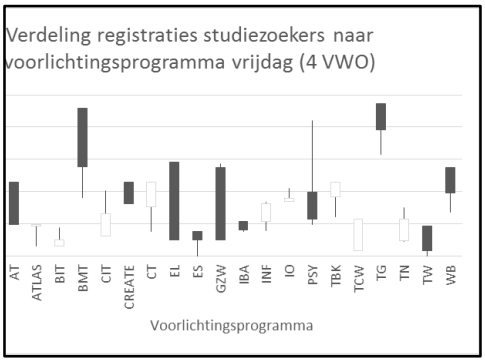


Figuur 0.7: Voorkeuren extra personen vrijdag (ROW)

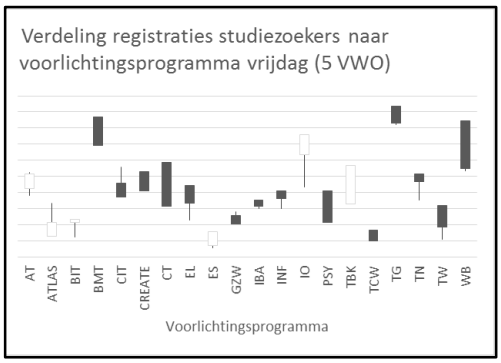
APPENDIX 5: VOORKEUREN STUDIEZOEKERS (VWO JAARGANGEN)



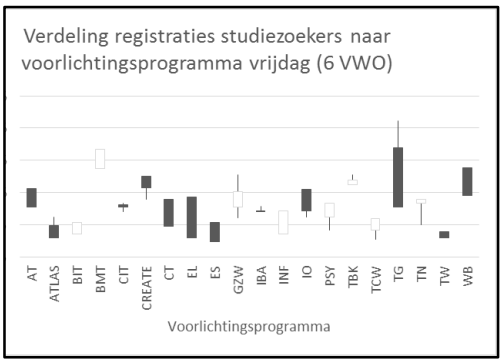
Figuur 0.8: Voorkeuren studiezoekers vrijdag (Overig NL)



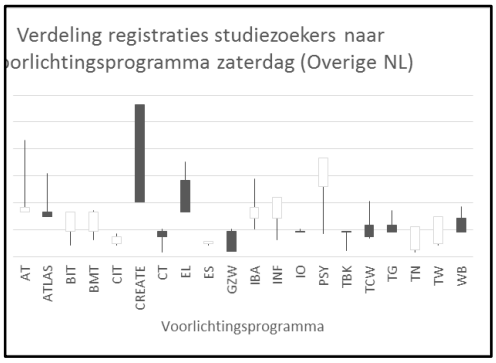
Figuur 0.9: Voorkeuren studiezoekers vrijdag (4 VWO)



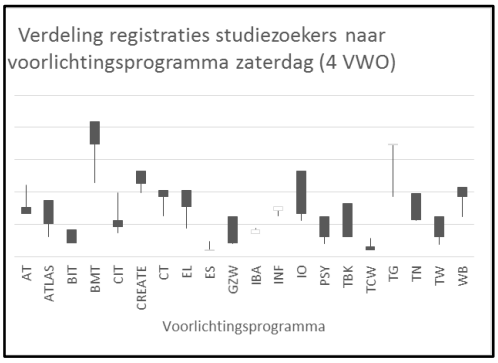
Figuur 0.10: Voorkeuren studiezoekers vrijdag (5 VWO)



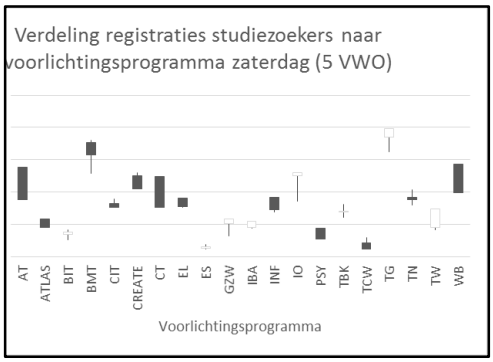
Figuur 0.11: Voorkeuren studiezoekers vrijdag (6 VWO)



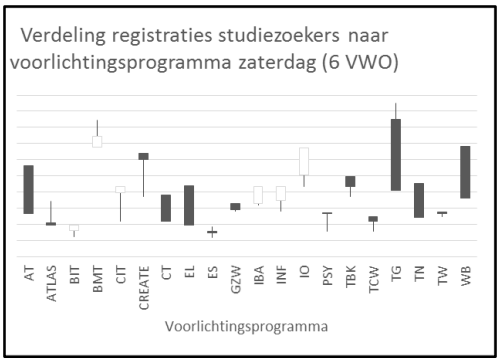
Figuur 0.12: Voorkeuren studiezoekers zaterdag (Overig NL)



Figuur 0.13: Voorkeuren studiezoekers zaterdag (4 VWO)

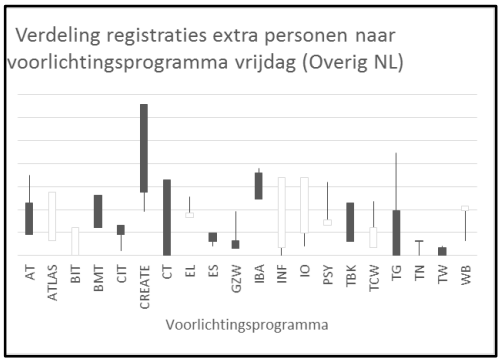


Figuur 0.14: Voorkeuren studiezoekers zaterdag (5 VWO)

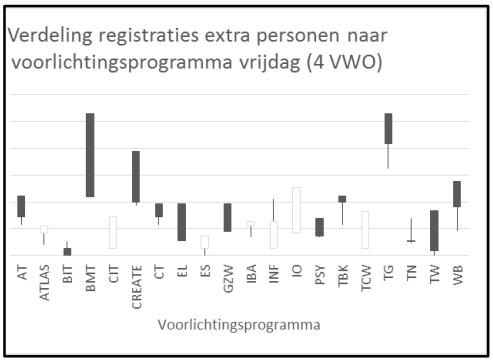


Figuur 0.15: Voorkeuren studiezoekers zaterdag (6 VWO)

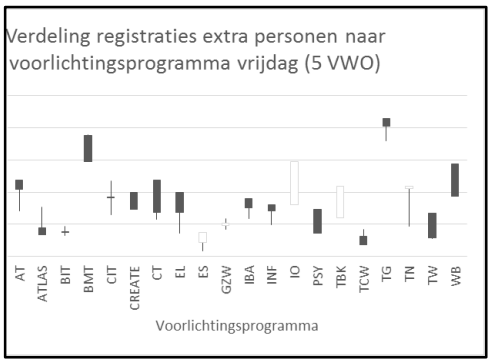
APPENDIX 6: VOORKEUREN EXTRA PERSONEN (VWO JAARGANGEN)



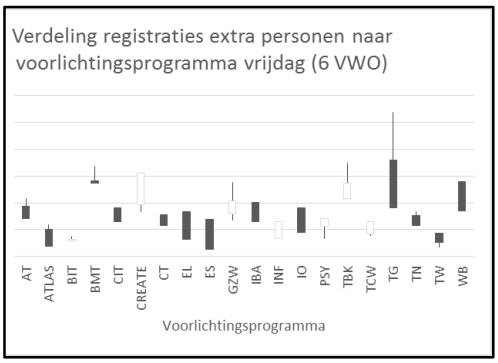
Figuur 0.16: Voorkeuren extra personen vrijdag (Overig NL)



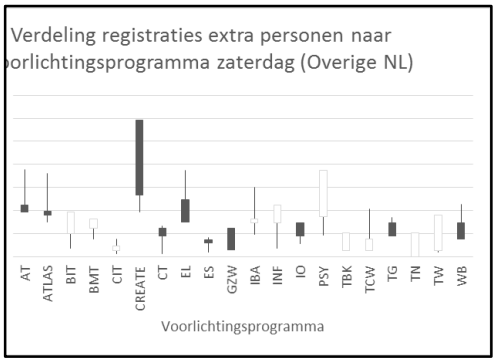
Figuur 0.17: Voorkeuren extra personen vrijdag (4 VWO)



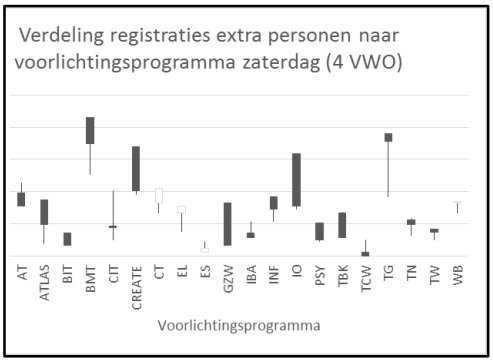
Figuur 0.18: Voorkeuren extra personen vrijdag (5 VWO)



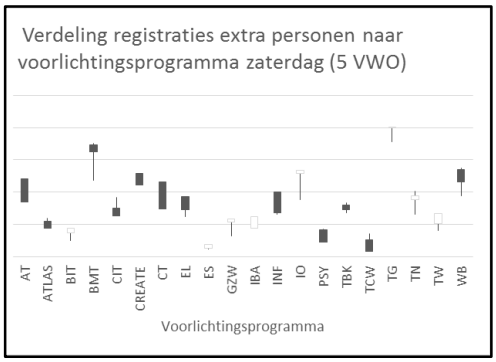
Figuur 0.19: Voorkeuren extra personen vrijdag (6 VWO)



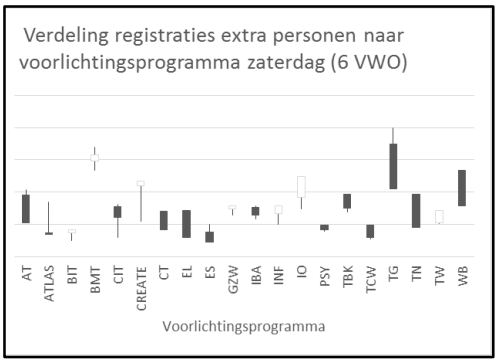
Figuur 0.20: Voorkeuren extra personen zaterdag (Overig NL)



Figuur 0.21: Voorkeuren extra personen zaterdag (4 VWO)

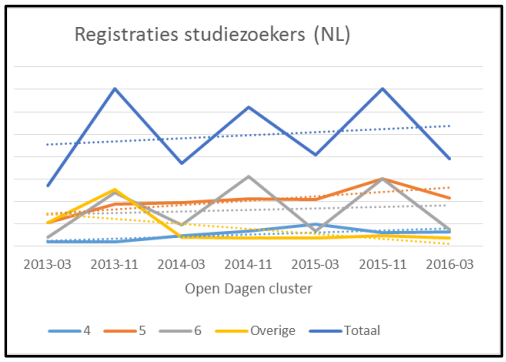


Figuur 0.22: Voorkeuren extra personen zaterdag (5 VWO)

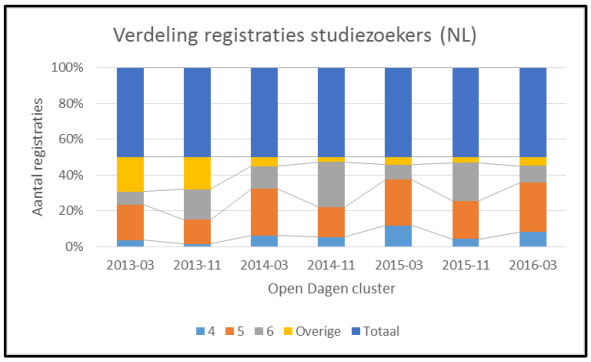


Figuur 0.23: Voorkeuren extra personen zaterdag (6 VWO)

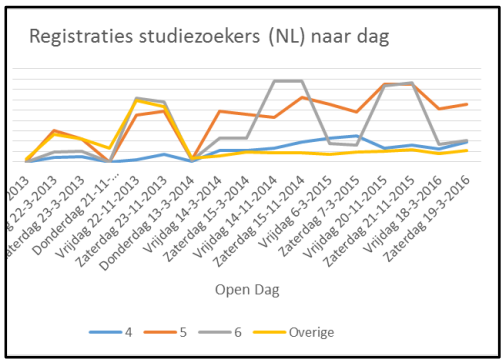
APPENDIX 7: ANALYSE REGISTRATIES STUDIEKIEZERS UIT NEDERLAND



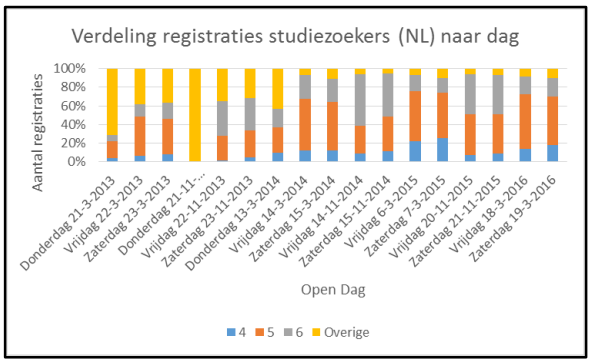
Figuur 0.24: Registraties studiezoekers (NL)



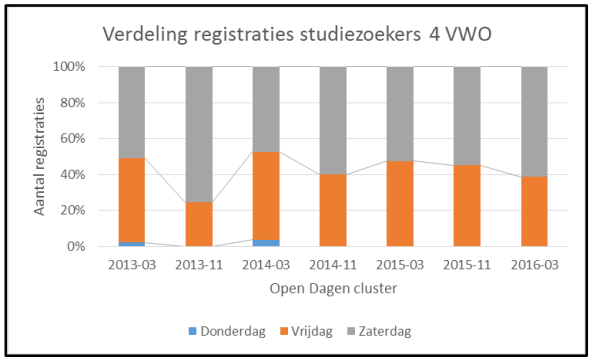
Figuur 0.25: Verdeling registraties studiezoekers (NL)



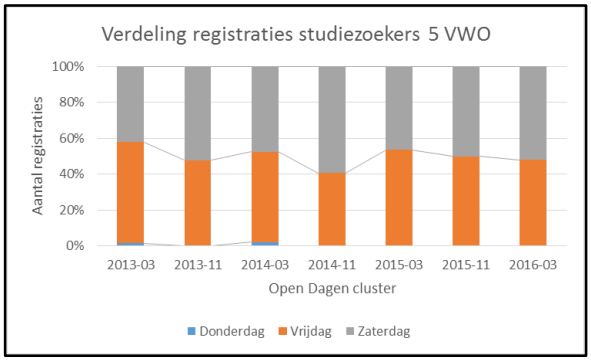
Figuur 0.26: Verdeling registraties studiezoekers naar dag (NL)



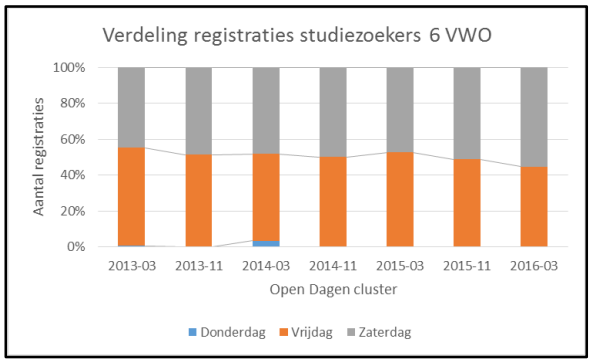
Figuur 0.27: Absolute verdeling registraties studiezoekers naar dag (NL)



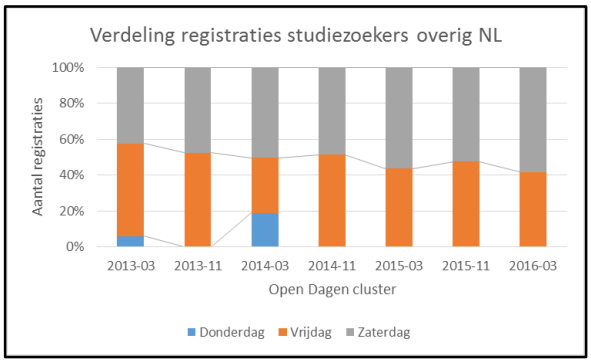
Figuur 0.28: Verdeling studiezoekers naar dag (4 VWO)



Figuur 0.29: Verdeling studiezoekers naar dag(5 VWO

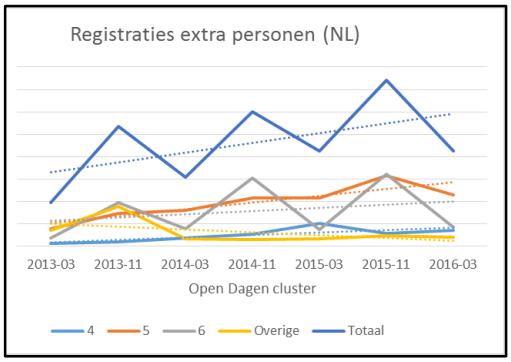


Figuur 0.30: Verdeling studiezoekers naar dag (6 VWO)

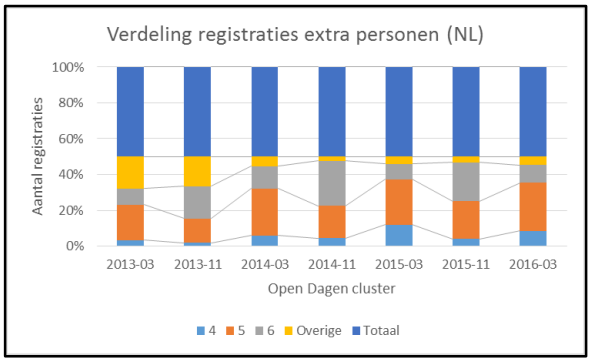


Figuur 0.31: Verdeling studiezoekers naar dag (Overig NL)

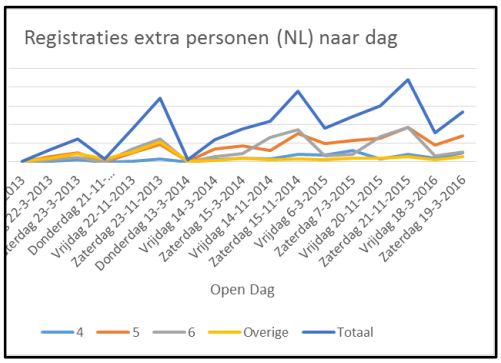
APPENDIX 8: ANALYSE REGISTRATIES EXTRA PERSONEN UIT NEDERLAND



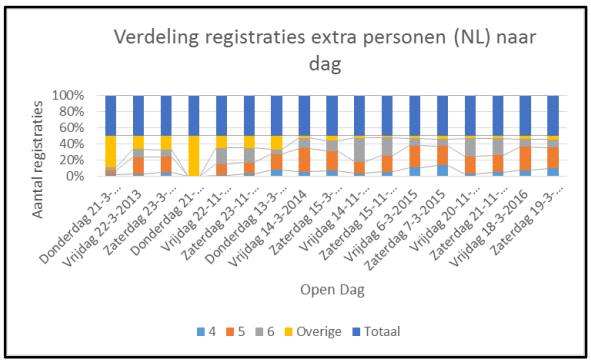
Figuur 0.32: Registraties extra personen (NL)



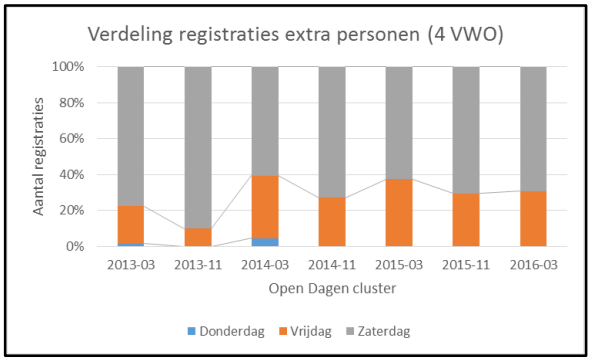
Figuur 0.33: Verdeling registraties extra personen (NL)



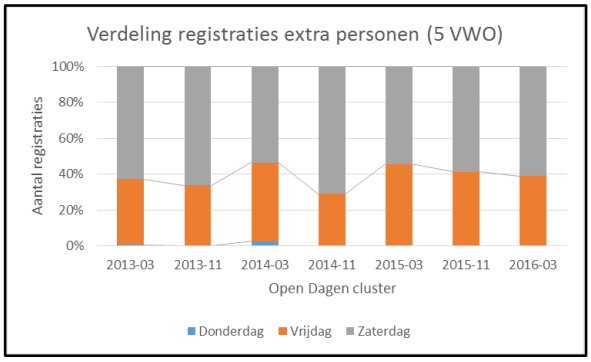
Figuur 0.34: Verdeling registraties extra personen naar dag (NL)



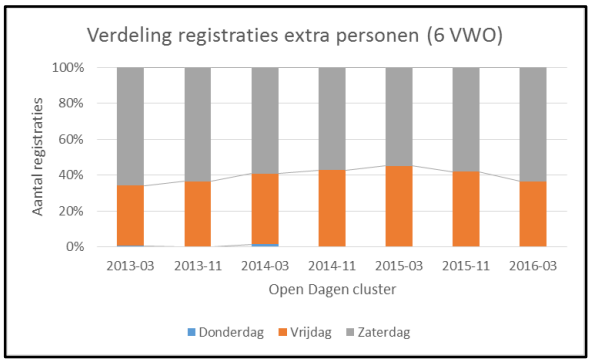
Figuur 0.35: Absolute verdeling registraties extra personen naar dag (NL)



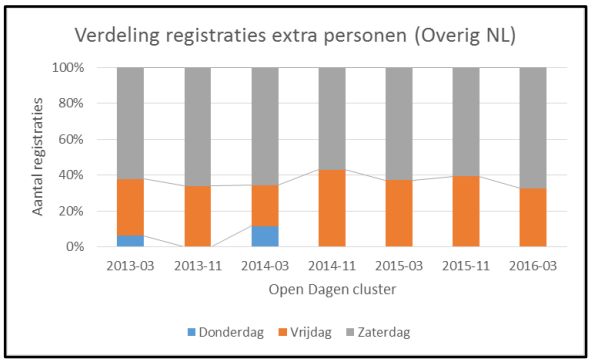
Figuur 0.36: Verdeling extra personen naar dag (4 VWO)



Figuur 0.37: Verdeling extra personen naar dag (5 VWO)

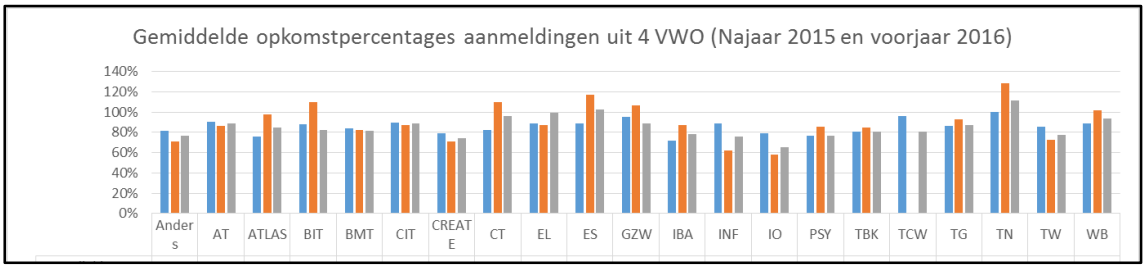


Figuur 0.38: Verdeling extra personen naar dag (6 VWO)

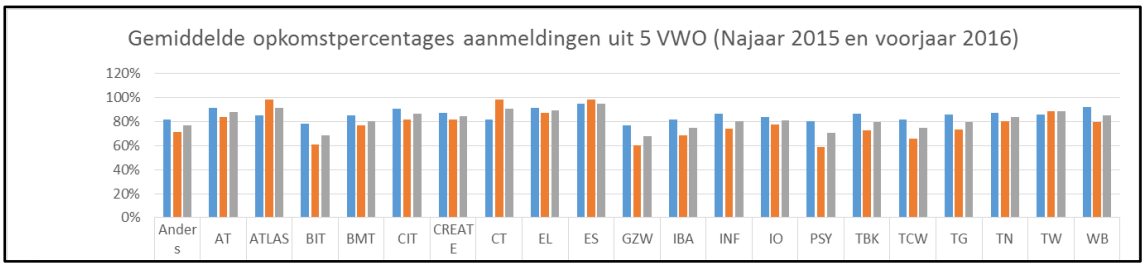


Figuur 0.39: Verdeling extra personen naar dag (Overig NL)

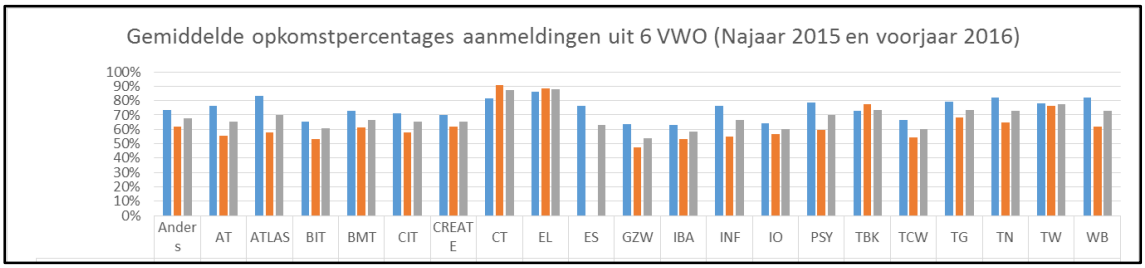
APPENDIX 9: OPKOMSTPERCENTAGES REGISTRATIES VWO JAARGANGEN



Figuur 0.40: Gemiddelde opkomstpercentages aanmeldingen (4 VWO)

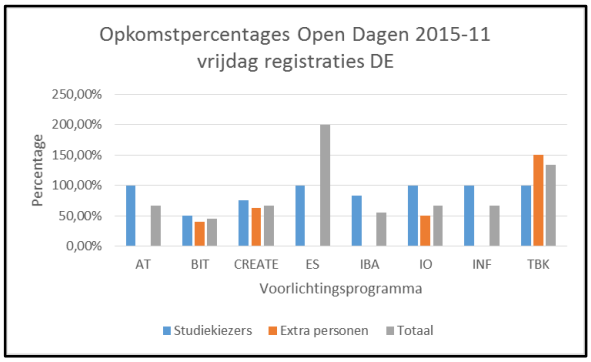


Figuur 0.41: Gemiddelde opkomstpercentages aanmeldingen (5 VWO)

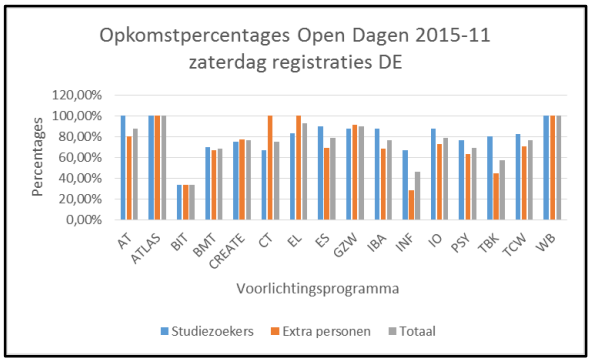


Figuur 0.42: Gemiddelde opkomstpercentages aanmeldingen (6 VWO)

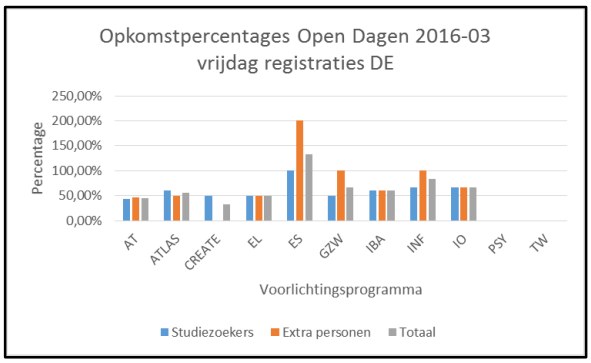
APPENDIX 10: OPKOMSTPERCENTAGES REGISTRATIES UIT DUITSLAND



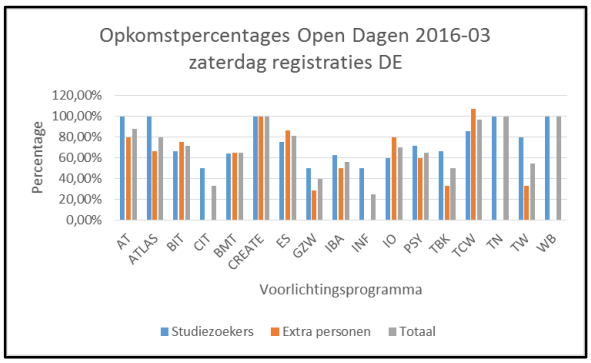
Figuur 0.43: Opkomstpercentages vrijdag 2015-11 (DE)



Figuur 0.44: Opkomstpercentages zaterdag 2015-11 (DE)

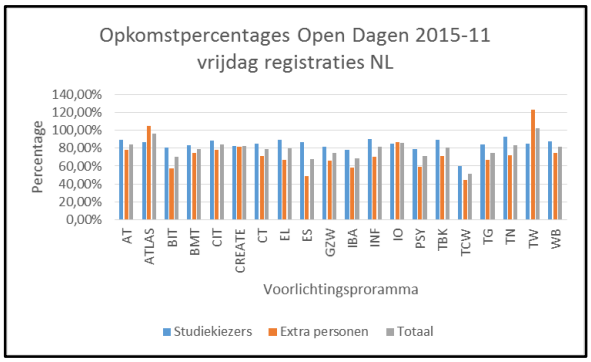


Figuur 0.45: Opkomstpercentages vrijdag 2016-03 (DE)

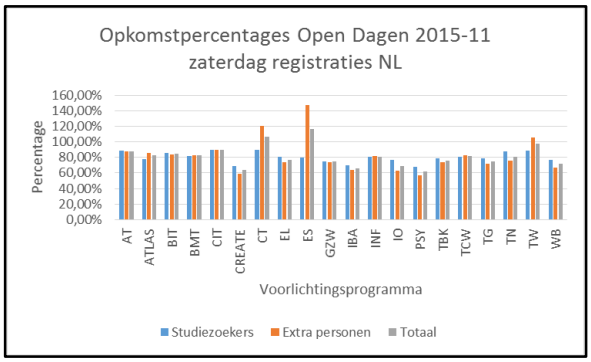


Figuur 0.46: Opkomstpercentages zaterdag 2016-03 (DE)

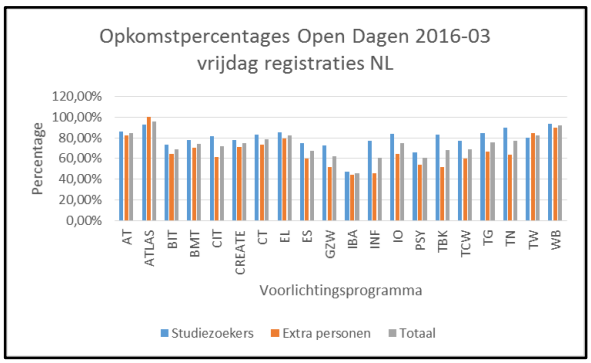
APPENDIX 11: OPKOMSTPERCENTAGES REGISTRATIES UIT NEDERLAND



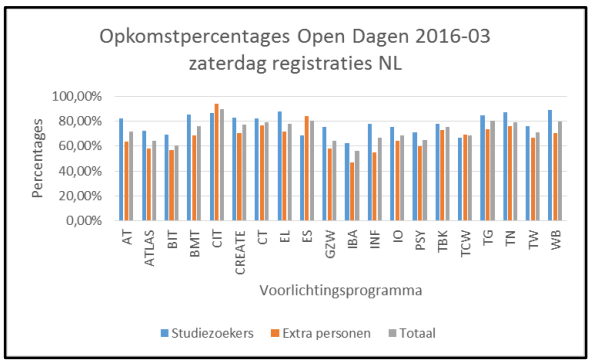
Figuur 0.47: Opkomstpercentages vrijdag 2015-03 (NL)



Figuur 0.48: Opkomstpercentages zaterdag 2015-11 (NL)

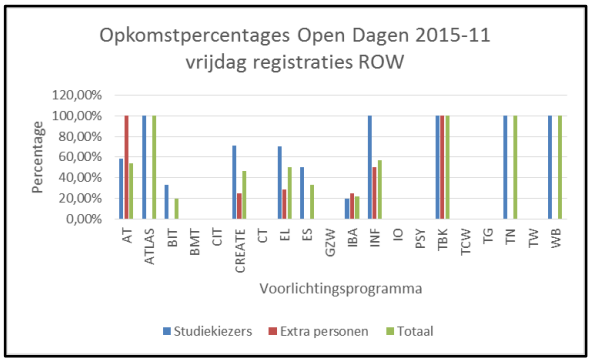


Figuur 0.49: Opkomstpercentages vrijdag 2016-03 (NL)

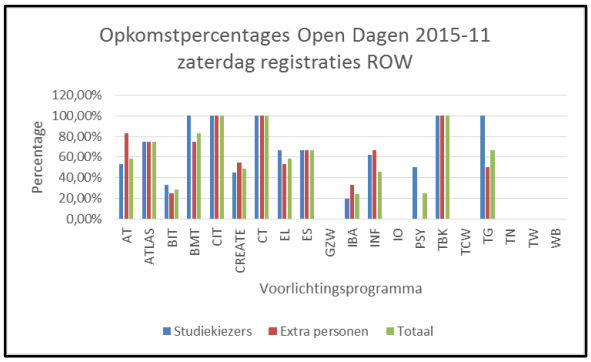


Figuur 0.50: Opkomstpercentages zaterdag 2016-03 (NL)

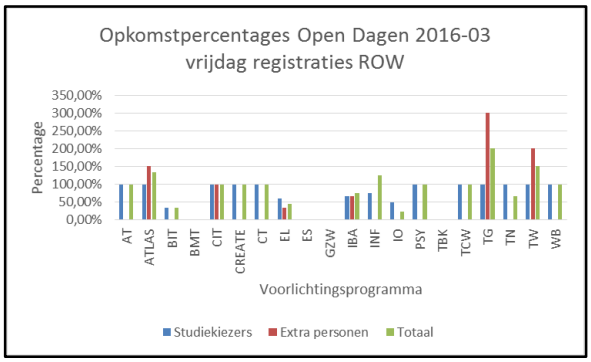
APPENDIX 12: OPKOMSTPERCENTAGES REGISTRATIES UIT ROW



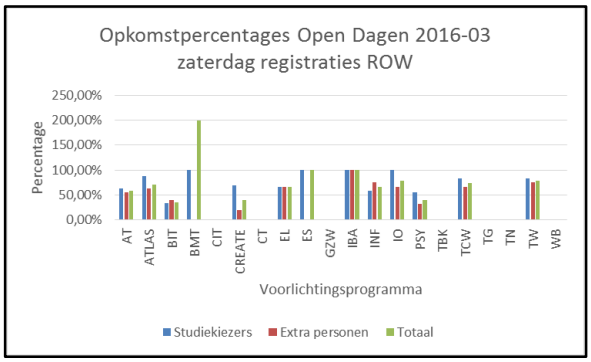
Figuur 0.51: Opkomstpercentages vrijdag 2015-11 (ROW)



Figuur 0.52: Opkomstpercentages zaterdag 2015-11 (ROW)



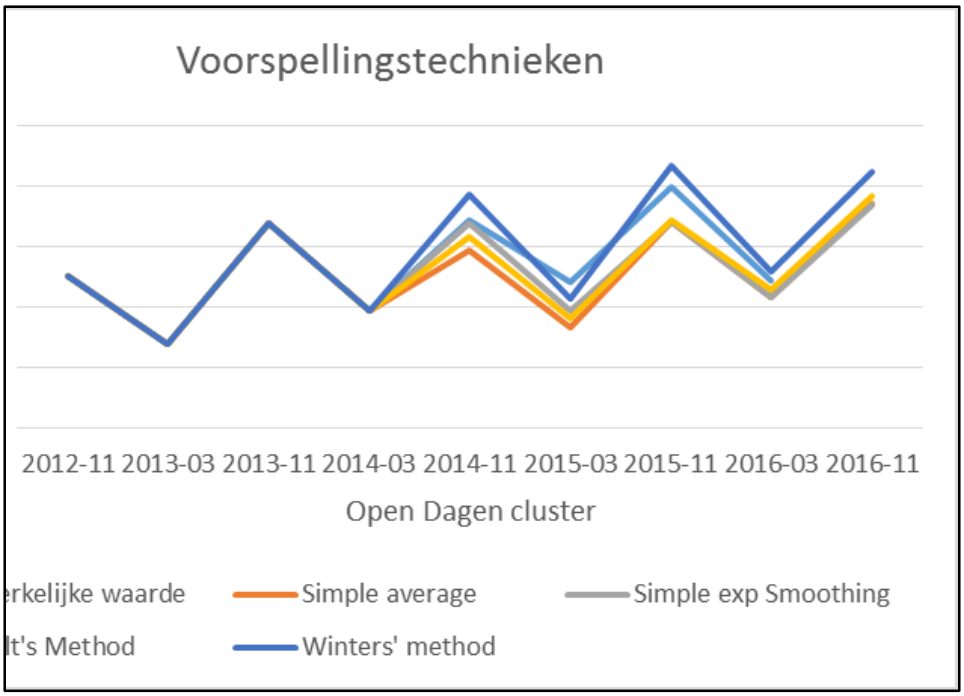
Figuur 0.53: Opkomstpercentages vrijdag 2016-03 (ROW)



Figuur 0.54: Opkomstpercentages zaterdag 2016-03 (ROW)

APPENDIX 13: BETROUWBAARHEID VOORSPELLINGSTECHNIEK

In Figuur 0.55 is een systematische weergave gegeven van de voorspelling van vier voorspellingstechnieken, namelijk Simple Average, Simple Exponential Smoothing, Holt's Method en Winters' Method. De voorspelling is gebaseerd op een enkelvoudige voorspelling van het aantal registraties voor de Open Dagen, zonder hierbij rekening te houden met verschillende doelgroepen binnen de data. Voor de Alpha, Beta en Gamma is het getal 0,6 genomen, omdat de voorspellingen met dit getal de minste onnauwkeurigheid vertonen.



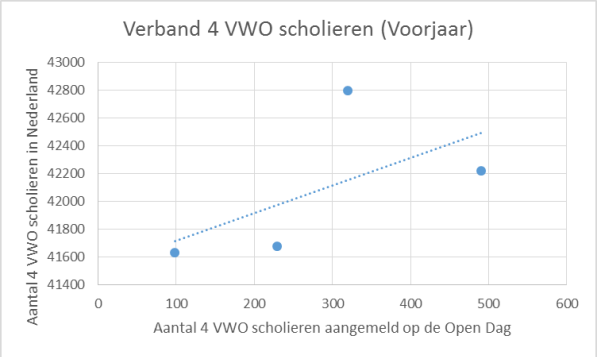
Figuur 0.55: Verschillende voorspellingstechnieken

Uit Figuur 0.55 valt af te leiden dat Winters' methode het nauwkeurigst lijkt. Dit valt ook op te maken uit de onnauwkeurigheidsgegevens, zoals weergeven in Tabel 0.3. De Simple Average voorspellingstechniek presteert het slechts op basis van de MAPE, gevolgd door de Holt's Method en Simple Exponential Smoothing. Winters' Method presteert het best met een gemiddelde afwijking (MAPE) van 3,9% positief.

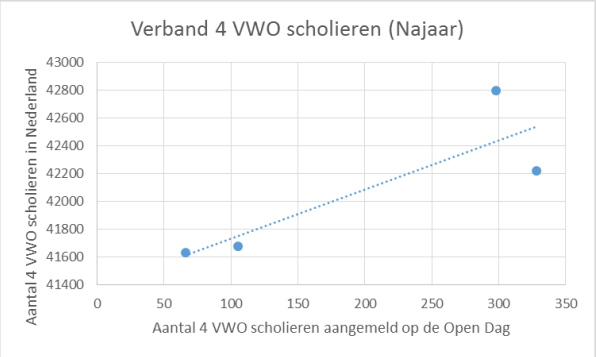
Errors	Simple average	Simple exp. Smoothing	Holt's Method	Winters' Method
MAD	xxxxx	xxxxx	xxxxx	xxxxx
St	xxxxx	xxxxx	xxxxx	xxxxx
MAPE	xxxxx	xxxxx	xxxxx	xxxxx
BIAS	xxxxx	xxxxx	xxxxx	xxxxx
TS	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 0.3: Onnauwkeurigheden voorspellingstechnieken

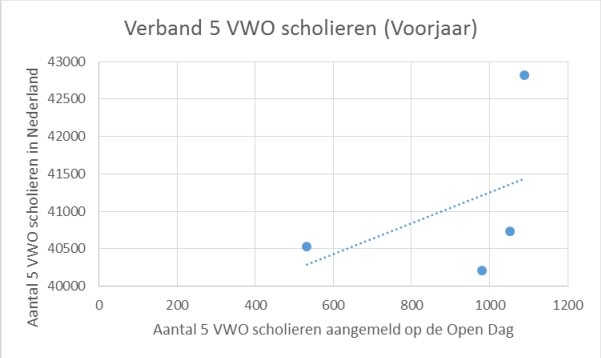
APPENDIX 14: SCATTERPLOTS VWO CORRELATIES



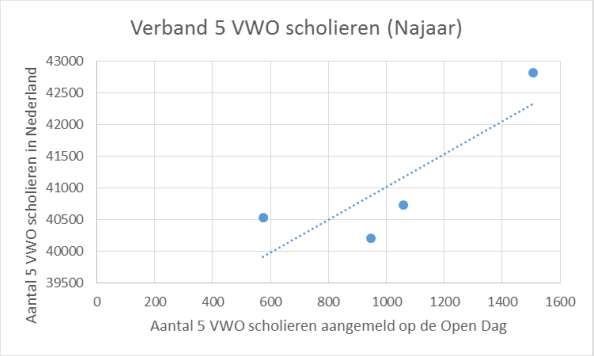
Figuur 0.56: Scatterplot 4 VWO voorjaar



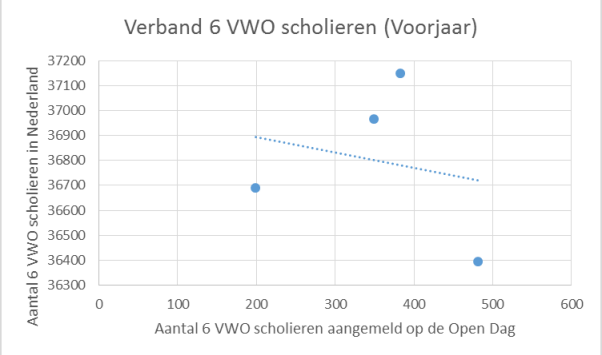
Figuur 0.57: Scatterplot 4 VWO najaar



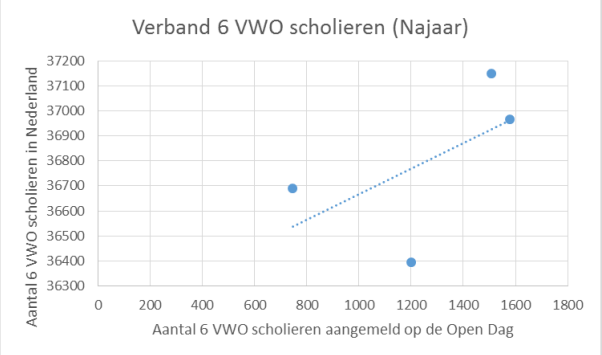
Figuur 0.58: Scatterplot 5 VWO voorjaar



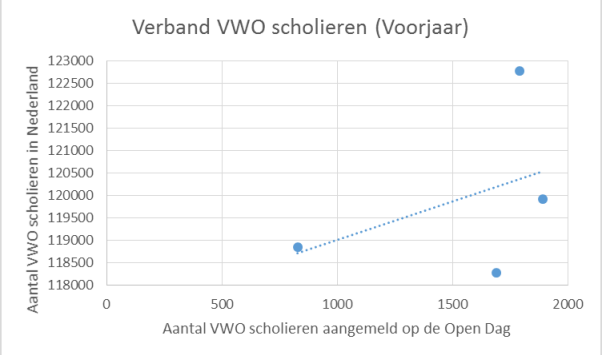
Figuur 0.59: Scatterplot 5 VWO najaar



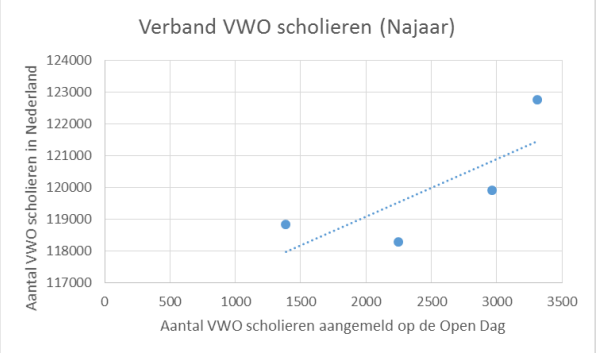
Figuur 0.60: Scatterplot 6 VWO voorjaar



Figuur 0.61: Scatterplot 6 VWO najaar



Figuur 0.62: Scatterplot VWO gezamenlijk voorjaar



Figuur 0.63: Scatterplot VWO gezamenlijk najaar

APPENDIX 15: REGRESSIE ANALYSES VWO CORRELATIES

We toetsen $H_0: \beta_1 = 0$ tegen $H_0: \beta_1 \neq 0$ (Er is geen correlatie, tegen er is wel een correlatie)

Toetsingsgrootheid $T = \frac{\widehat{\beta_1}}{se(\widehat{\beta_1})}$ onder nulhypothese $T \sim t_2$

Uitkomst van T:	Voorjaar	Najaar
4 VWO	-1,00607	-2,29244
5 VWO	-0,54319	-1,9674
6 VWO	0,355125	-0,97621

Tabel 0.4: Uitkomsten toetsingsgrootheid T (Uitkomsten regressieanalyses)

We verwerpen de nulhypothese als $T \leq -4{,}30 \; of \; T \geq 4{,}30 \; (df = 2, \alpha = 5\%)$

Uitkomsten van T liggen niet in het kritieke gebied, we verwerpen dus de nulhypothese niet.

We achten dus bewezen dat er geen correlatie is tussen het aantal VWO scholieren en het aantal aanmeldingen voor de Open Dagen.

SAMENVATTING UITVOER			Voorjaar 4 VWO					
gevens voor de regressie								
Meervouc	0,59452							
R-kwadraa	0,353454							
Aangepas	-2							
Standaard	162,1702							
Waarnem	1							
Variantie-analyse								
	Vrijheidsgrad	*adratensc*	*delde kwa*	*F*	*gnificantie F*			
Regressie	4	28754,43	7188,608	1,093359	#GETAL!			
Storing	2	52598,32	26299,16					
Totaal	6	81352,75						
	Coëfficiënte	*ndaardfo*	*istische ge*	*P-waarde*	*aagste 95%*	*oogste 95%*	*agste 95,0*	*ogste 95,0*
Snijpunt							-4E-306	6,6E-306
Variabele X 1							-1E-306	2,2E-306
Variabele X 2							0	0
Variabele	-7239,12	7195,459	-1,00607	0,420321	-38198,7	23720,44	-38198,7	23720,44
Variabele	0,178784	0,170981	1,045638	0,40548	-0,55689	0,914455	-0,55689	0,914455

Figuur 0.64: Regressieanalyse correlatie 4 VWO scholieren voorjaar

SAMENVATTING UITVOER			Voorjaar 5 VWO					
gevens voor de regressie								
Meervouc	0,445967							
R-kwadraa	0,198887							
Aangepas	-2							
Standaard	283,1986							
Waarnem	1							
Variantie-analyse								
	Vrijheidsgrad	*adratensc*	*delde kwa*	*F*	*gnificantie F*			
Regressie	4	39822,09	9955,522	0,496526	#GETAL!			
Storing	2	160402,9	80201,46					
Totaal	6	200225						
	Coëfficiënte	*ndaardfo*	*istische ge*	*P-waarde*	*aagste 95%*	*oogste 95%*	*agste 95,0*	*ogste 95,0*
Snijpunt							1,8E-304	1,8E-304
Variabele X 1							0	0
Variabele X 2							0	0
Variabele	-3077,48	5665,584	-0,54319	0,641446	-27454,5	21299,56	-27454,5	21299,56
Variabele	0,097172	0,137901	0,704646	0,554033	-0,49617	0,690513	-0,49617	0,690513

Figuur 0.65: Regressieanalyse correlatie 5 VWO scholieren voorjaar

SAMENVATTING UITVOER			Voorjaar 6 VWO					
gevens voor de regressie								
Meervouc	0,218021							
R-kwadraa	0,047533							
Aangepas	-2							
Standaard	139,7612							
Waarnem	1							
Variantie-analyse								
	Vrijheidsgrad	*adratensc*	*delde kwa*	*F*	*gnificantie F*			
Regressie	4	1949,627	487,4068	0,099811	#GETAL!			
Storing	2	39066,37	19533,19					
Totaal	6	41016						
	Coëfficiënte	*ndaardfo*	*istische ge*	*P-waarde*	*aagste 95%*	*oogste 95%*	*agste 95,0*	*ogste 95,0*
Snijpunt							0	0
Variabele X 1							0	0
Variabele X 2							0	0
Variabele	3197,495	9003,865	0,355125	0,75645	-35543	41938	-35543	41938
Variabele	-0,0773	0,244671	-0,31593	0,781979	-1,13003	0,975436	-1,13003	0,975436

Figuur 0.66: Regressieanalyse correlatie 6 VWO scholieren voorjaar

SAMENVATTING UITVOER			Najaar 4 VWO					
gevens voor de regressie								
Meervouc	0,85643							
R-kwadraa	0,733472							
Aangepas	-2							
Standaard	84,01531							
Waarnem	1							
Variantie-analyse								
	Vrijheidsgrad	*adratensc*	*delde kwa*	*F*	*gnificantie F*			
Regressie	4	38849,6	9712,401	5,50389	#GETAL!			
Storing	2	14117,15	7058,573					
Totaal	6	52966,75						
	Coëfficiënte	*ndaardfo*	*stische ge*	*P-waarde*	*aagste 95%*	*oogste 95%*	*agste 95,0*	*ogste 95,0*
Snijpunt							2,4E-307	2,4E-307
Variabele X 1							2,4E-307	2,4E-307
Variabele X 2							0	0
Variabele	-8545,62	3727,744	-2,29244	0,148919	-24584,8	7493,567	-24584,8	7493,567
Variabele	0,207812	0,08858	2,346037	0,14357	-0,17332	0,58894	-0,17332	0,58894

Figuur 0.67: Regressieanalyse correlatie 4 VWO scholieren najaar

SAMENVATTING UITVOER			Najaar 5 VWO					
gevens voor de regressie								
Meervouc	0,83741							
R-kwadraa	0,701255							
Aangepas	-2							
Standaard	256,9785							
Waarnem	1							
Variantie-analyse								
	Vrijheidsgrad	*adratensc*	*delde kwa*	*F*	*gnificantie F*			
Regressie	4	310026,8	77506,71	4,694677	#GETAL!			
Storing	2	132075,9	66037,95					
Totaal	6	442102,8						
	Coëfficiënte	*ndaardfo*	*stische ge*	*P-waarde*	*aagste 95%*	*oogste 95%*	*agste 95,0*	*ogste 95,0*
Snijpunt							0	0
Variabele X 1							0	0
Variabele X 2							0	0
Variabele	-10114,5	5141,033	-1,9674	0,188013	-32234,5	12005,63	-32234,5	12005,63
Variabele	0,27113	0,125134	2,16672	0,16259	-0,26728	0,809536	-0,26728	0,809536

Figuur 0.68: Regressieanalyse correlatie 5 VWO scholieren najaar

SAMENVATTING UITVOER			Najaar 6 VWO					
gevens voor de regressie								
Meervoud	0,588097							
R-kwadraa	0,345858							
Aangepas	-2							
Standaard	374,7524							
Waarnem	1							
Variantie-analyse								
	Vrijheidsgrad	*adratensc*	*delde kwa*	*F*	*gnificantie F*			
Regressie	4	148506	37126,51	1,057439	#GETAL!			
Storing	2	280878,7	140439,4					
Totaal	6	429384,8						
	Coëfficiënte	*ndaardfo*	*stische ge*	*P-waarde*	*aagste 95%*	*oogste 95%*	*agste 95,0*	*ogste 95,0*
Snijpunt							0	0
Variabele X 1							0	0
Variabele X 2							0	0
Variabele	-23568,4	24142,76	-0,97621	0,431915	-127446	80309,45	-127446	80309,45
Variabele	0,674634	0,656056	1,028319	0,411903	-2,14815	3,497414	-2,14815	3,497414

Figuur 0.69: Regressieanalyse correlatie 6 VWO scholieren najaar

APPENDIX 16: OPBOUW VOORSPELLING

Figuur 0.70: Opbouw voorspelling

APPENDIX 17: VOORSPELLINGEN AANMELDINGEN NEDERLAND

De aanmeldingen uit Nederland hebben we onderverdeeld in aanmeldingen uit 4, 5 en 6 VWO en overige aanmeldingen uit Nederland. Elk van deze doelgroepen hebben we verder onderverdeeld in aanmeldingen van studiezoekers en daarbij behorende extra personen.

4 VWO aanmeldingen

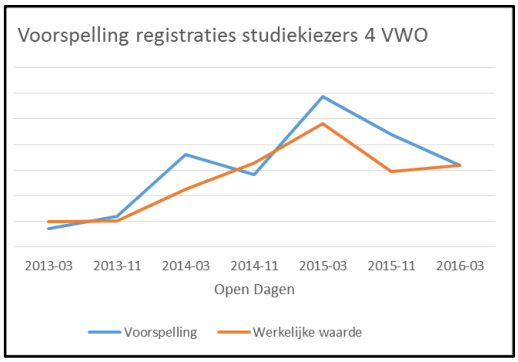
De voorspelling van het aantal verwachte aanmeldingen uit 4 VWO is op dezelfde manier berekend als de voorspelling van de verwachte aanmeldingen uit Duitsland en de ROW.

Open Dagen	Studiezoekers	Extra personen
2013-03	xxxxx	xxxxx
2013-11	xxxxx	xxxxx
2014-03	xxxxx	xxxxx
2014-11	xxxxx	xxxxx

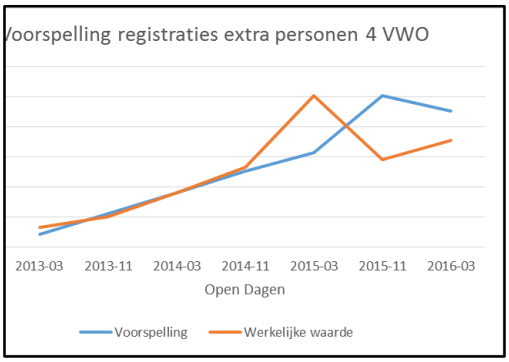
Tabel 0.5: Initialisatieset 4 VWO

Aanmeldingen uit 4 VWO	Leven in T0	Trend	Seizoenfactor hoog	Seizoenfactor laag
Studiezoekers	xxxxx	xxxxx	xxxxx	xxxxx
Extra personen	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 0.6: Opbouw voorspellen 4 VWO



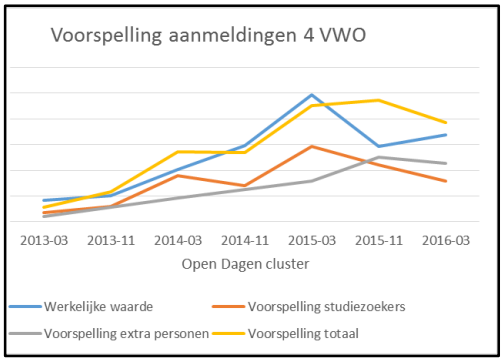
Figuur 0.71: Voorspelling studiekiezers (4 VWO)



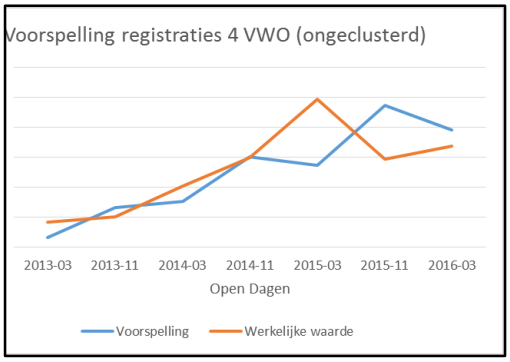
Figuur 0.72: Voorspelling extra personen (4 VWO)

Voorspelling	Studiekiezers	Extra personen	Totaal gecombineerd	Totaal enkelvoudig
Alpha	xxxxx	xxxxx	xxxxx	xxxxx
Beta	xxxxx	xxxxx	xxxxx	xxxxx
Gamma	xxxxx	xxxxx	xxxxx	xxxxx
MAD	xxxxx	xxxxx	xxxxx	xxxxx
st	xxxxx	xxxxx	xxxxx	xxxxx
Mape	xxxxx	xxxxx	xxxxx	xxxxx
Bias	xxxxx	xxxxx	xxxxx	xxxxx
TS	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 0.7: Voorspelling aanmeldingen uit 4 VWO



Figuur 0.73: Gecombineerde voorspelling aanmeldingen 4 VWO



Figuur 0.74: Enkelvoudige voorspelling aanmeldingen 4 VWO

5 VWO aanmeldingen

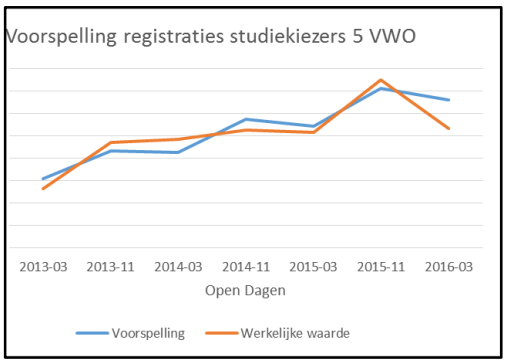
De voorspelling van het aantal verwachte aanmeldingen uit 5 VWO is op dezelfde manier berekend als de voorspelling van de verwachte aanmeldingen uit Duitsland en de ROW.

Open Dagen	Studiezoekers	Extra personen
2013-03	xxxxx	xxxxx
2013-11	xxxxx	xxxxx
2014-03	xxxxx	xxxxx
2014-11	xxxxx	xxxxx

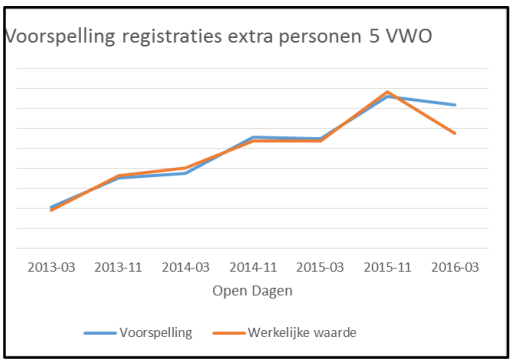
Tabel 0.8: Initialisatieset 5 VWO

Aanmeldingen uit 5 VWO	Leven in T0	Trend	Seizoenfactor hoog	Seizoenfactor laag
Studiezoekers	xxxxx	xxxxx	xxxxx	xxxxx
Extra personen	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 0.9: Opbouw voorspelling 5 VWO



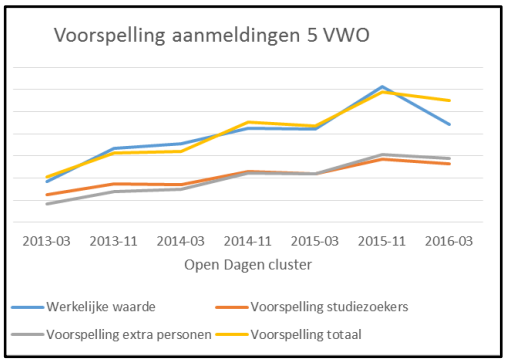
Figuur 0.75: Voorspelling studiekiezers 5 VWO



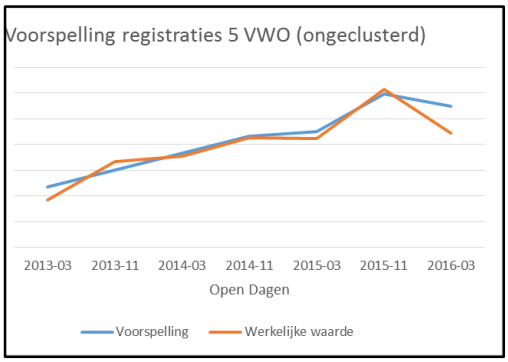
Figuur 0.76: Voorspelling extra personen 5 VWO

Voorspelling	Studiekiezers	Extra personen	Totaal gecombineerd	Totaal enkelvoudig
Alpha	xxxxx	xxxxx	xxxxx	xxxxx
Beta	xxxxx	xxxxx	xxxxx	xxxxx
Gamma	xxxxx	xxxxx	xxxxx	xxxxx
MAD	xxxxx	xxxxx	xxxxx	xxxxx
st	xxxxx	xxxxx	xxxxx	xxxxx
Mape	xxxxx	xxxxx	xxxxx	xxxxx
Bias	xxxxx	xxxxx	xxxxx	xxxxx
TS	xxxxx	xxxxx	xxxxx	xxxxx

Figuur 0.77: Gegevens voorspellingen 5 VWO



Figuur 0.78: Gecombineerde voorspelling 5 VWO



Figuur 0.79: Enkelvoudige voorspelling 5 VWO

6 VWO aanmeldingen

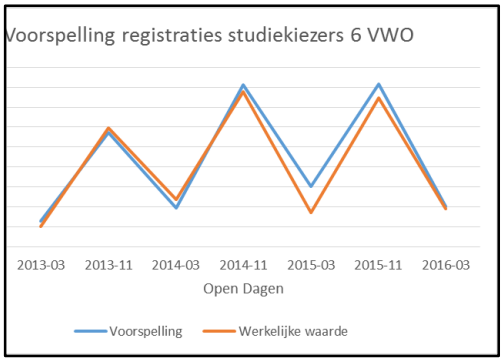
De voorspelling van het aantal verwachte aanmeldingen uit 6 VWO is op dezelfde manier berekend als de voorspelling van de verwachte aanmeldingen uit Duitsland en de ROW.

Open Dagen	Studiezoekers	Extra personen
2013-03	xxxxx	xxxxx
2013-11	xxxxx	xxxxx
2014-03	xxxxx	xxxxx
2014-11	xxxxx	xxxxx

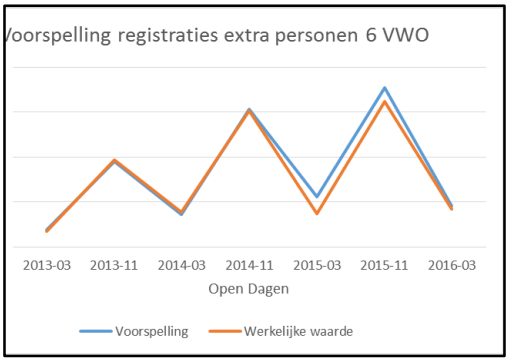
Tabel 0.10: Initialisatieset 6 VWO

Aanmeldingen uit 6 VWO	Leven in T0	Trend	Seizoenfactor hoog	Seizoenfactor laag
Studiezoekers	xxxxx	xxxxx	xxxxx	xxxxx
Extra personen	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 0.11: Opbouw voorspelling 6 VWO



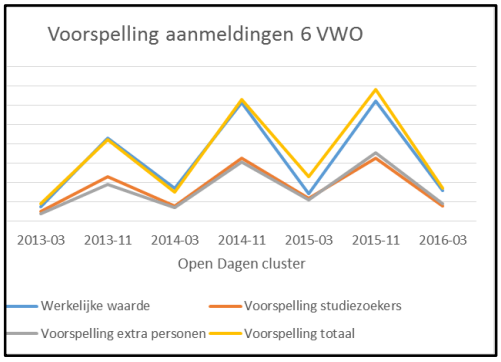
Figuur 0.80: Voorspelling studiezoekers 6 VWO



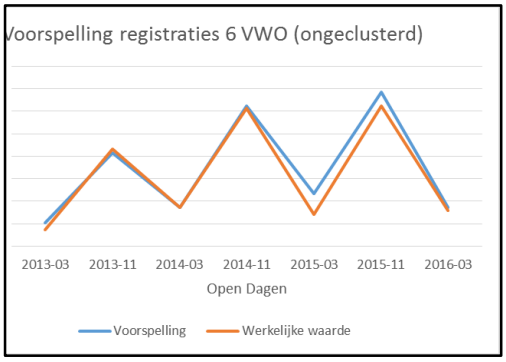
Figuur 0.81: Voorspelling extra personen 6 VWO

Voorspelling	Studiekiezers	Extra personen	Totaal gecombineerd	Totaal enkelvoudig
Alpha	xxxxx	xxxxx	xxxxx	xxxxx
Beta	xxxxx	xxxxx	xxxxx	xxxxx
Gamma	xxxxx	xxxxx	xxxxx	xxxxx
MAD	xxxxx	xxxxx	xxxxx	xxxxx
st	xxxxx	xxxxx	xxxxx	xxxxx
Mape	xxxxx	xxxxx	xxxxx	xxxxx
Bias	xxxxx	xxxxx	xxxxx	xxxxx
TS	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 0.12: Gegevens voorspellingen 6 VWO



Figuur 0.82: Gecombineerde voorspelling 6 VWO



Figuur 0.83: Enkelvoudige voorspelling 6 VWO

Overige NL aanmeldingen

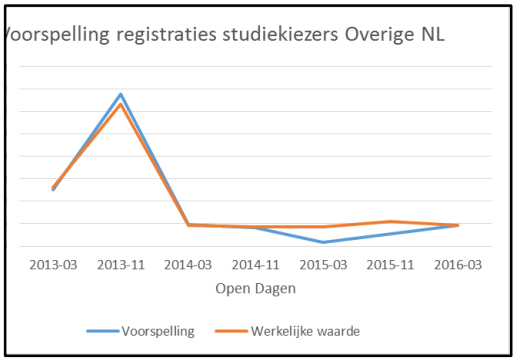
De voorspelling van het aantal verwachte aanmeldingen uit overige NL is op dezelfde manier berekend als de voorspelling van de verwachte aanmeldingen uit Duitsland en de ROW.

Open Dagen	Studiezoekers	Extra personen
2013-03	xxxxx	xxxxx
2013-11	xxxxx	xxxxx
2014-03	xxxxx	xxxxx
2014-11	xxxxx	xxxxx

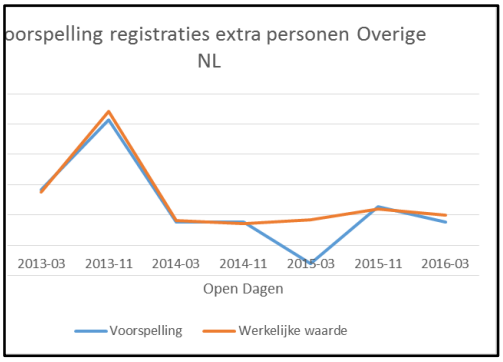
Tabel 0.13: Initialisatieset Overige NL

Aanmeldingen uit overige NL	Leven in T0	Trend	Seizoenfactor hoog	Seizoenfactor laag
Studiezoekers	xxxxx	xxxxx	xxxxx	xxxxx
Extra personen	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 0.14: Opbouw voorspelling Overige NL



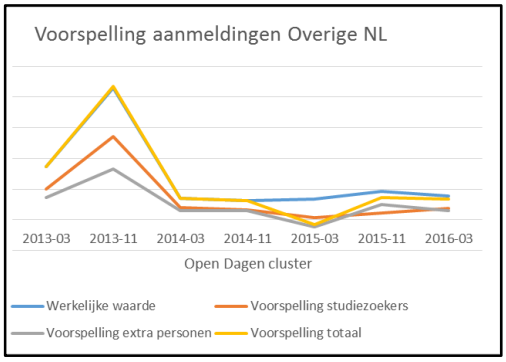
Figuur 0.84: voorspelling studiezoekers Overige NL



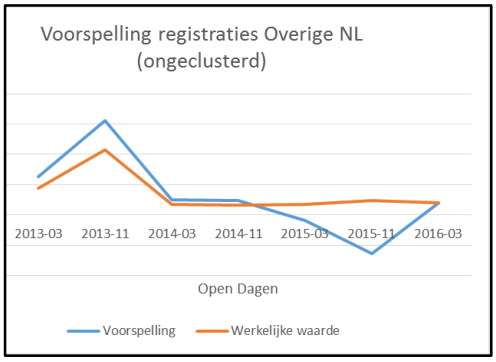
Figuur 0.85: Voorspelling extra personen Overige NL

Voorspelling	Studiekiezers	Extra personen	Totaal gecombineerd	Totaal enkelvoudig
Alpha	xxxxx	xxxxx	xxxxx	xxxxx
Beta	xxxxx	xxxxx	xxxxx	xxxxx
Gamma	xxxxx	xxxxx	xxxxx	xxxxx
MAD	xxxxx	xxxxx	xxxxx	xxxxx
st	xxxxx	xxxxx	xxxxx	xxxxx
Mape	xxxxx	xxxxx	xxxxx	xxxxx
Bias	xxxxx	xxxxx	xxxxx	xxxxx
TS	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 0.15: Gegevens voorspellingen Overige NL



Figuur 0.86: Gecombineerde voorspelling Overige NL



Figuur 0.87: Enkelvoudige voorspelling Overige NL

APPENDIX 18: STATISTISCHE VERDELINGEN PER VOORLICHTINGSPROGRAMMA

Gemiddelde voorkeuren per voorlichtingsprogramma vrijdag

Voorlichtingsprogramma Vrijdag ROW	Studiekiezers	Deviatie	Extra personen	Deviatie
AT	xxxxx	xxxxx	xxxxx	xxxxx
ATLAS	xxxxx	xxxxx	xxxxx	xxxxx
BIT	xxxxx	xxxxx	xxxxx	xxxxx
BMT	xxxxx	xxxxx	xxxxx	xxxxx
CIT	xxxxx	xxxxx	xxxxx	xxxxx
CREATE	xxxxx	xxxxx	xxxxx	xxxxx
CT	xxxxx	xxxxx	xxxxx	xxxxx
EL	xxxxx	xxxxx	xxxxx	xxxxx
ES	xxxxx	xxxxx	xxxxx	xxxxx
GZW	xxxxx	xxxxx	xxxxx	xxxxx
IBA	xxxxx	xxxxx	xxxxx	xxxxx
INF	xxxxx	xxxxx	xxxxx	xxxxx
IO	xxxxx	xxxxx	xxxxx	xxxxx
PSY	xxxxx	xxxxx	xxxxx	xxxxx
TBK	xxxxx	xxxxx	xxxxx	xxxxx
TCW	xxxxx	xxxxx	xxxxx	xxxxx
TG	xxxxx	xxxxx	xxxxx	xxxxx
TN	xxxxx	xxxxx	xxxxx	xxxxx
TW	xxxxx	xxxxx	xxxxx	xxxxx
WB	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 0.16: Statistische verdeling voorlichtingsprogramma's vrijdag (ROW)

Voorlichtingsprogramma Vrijdag DE	Studiekiezers	Deviatie	Extra personen	Deviatie
AT	xxxxx	xxxxx	xxxxx	xxxxx
ATLAS	xxxxx	xxxxx	xxxxx	xxxxx
BIT	xxxxx	xxxxx	xxxxx	xxxxx
BMT	xxxxx	xxxxx	xxxxx	xxxxx
CIT	xxxxx	xxxxx	xxxxx	xxxxx
CREATE	xxxxx	xxxxx	xxxxx	xxxxx
CT	xxxxx	xxxxx	xxxxx	xxxxx
EL	xxxxx	xxxxx	xxxxx	xxxxx
ES	xxxxx	xxxxx	xxxxx	xxxxx
GZW	xxxxx	xxxxx	xxxxx	xxxxx
IBA	xxxxx	xxxxx	xxxxx	xxxxx
INF	xxxxx	xxxxx	xxxxx	xxxxx
IO	xxxxx	xxxxx	xxxxx	xxxxx
PSY	xxxxx	xxxxx	xxxxx	xxxxx
TBK	xxxxx	xxxxx	xxxxx	xxxxx
TCW	xxxxx	xxxxx	xxxxx	xxxxx
TG	xxxxx	xxxxx	xxxxx	xxxxx
TN	xxxxx	xxxxx	xxxxx	xxxxx
TW	xxxxx	xxxxx	xxxxx	xxxxx
WB	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 0.17: Statistische verdeling voorlichtingsprogramma's vrijdag (DE)

Voorlichtingsprogramma Vrijdag 4 VWO	Studiekiezers	Deviatie	Extra personen	Deviatie
AT	xxxxx	xxxxx	xxxxx	xxxxx
ATLAS	xxxxx	xxxxx	xxxxx	xxxxx
BIT	xxxxx	xxxxx	xxxxx	xxxxx
BMT	xxxxx	xxxxx	xxxxx	xxxxx
CIT	xxxxx	xxxxx	xxxxx	xxxxx
CREATE	xxxxx	xxxxx	xxxxx	xxxxx
CT	xxxxx	xxxxx	xxxxx	xxxxx
EL	xxxxx	xxxxx	xxxxx	xxxxx
ES	xxxxx	xxxxx	xxxxx	xxxxx
GZW	xxxxx	xxxxx	xxxxx	xxxxx
IBA	xxxxx	xxxxx	xxxxx	xxxxx
INF	xxxxx	xxxxx	xxxxx	xxxxx
IO	xxxxx	xxxxx	xxxxx	xxxxx
PSY	xxxxx	xxxxx	xxxxx	xxxxx
TBK	xxxxx	xxxxx	xxxxx	xxxxx
TCW	xxxxx	xxxxx	xxxxx	xxxxx
TG	xxxxx	xxxxx	xxxxx	xxxxx
TN	xxxxx	xxxxx	xxxxx	xxxxx
TW	xxxxx	xxxxx	xxxxx	xxxxx
WB	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 0.18: Statistische verdeling voorlichtingsprogramma's vrijdag (4 VWO)

Voorlichtingsprogramma Vrijdag 5 VWO	Studiekiezers	Deviatie	Extra personen	Deviatie
AT	xxxxx	xxxxx	xxxxx	xxxxx
ATLAS	xxxxx	xxxxx	xxxxx	xxxxx
BIT	xxxxx	xxxxx	xxxxx	xxxxx
BMT	xxxxx	xxxxx	xxxxx	xxxxx
CIT	xxxxx	xxxxx	xxxxx	xxxxx
CREATE	xxxxx	xxxxx	xxxxx	xxxxx
CT	xxxxx	xxxxx	xxxxx	xxxxx
EL	xxxxx	xxxxx	xxxxx	xxxxx
ES	xxxxx	xxxxx	xxxxx	xxxxx
GZW	xxxxx	xxxxx	xxxxx	xxxxx
IBA	xxxxx	xxxxx	xxxxx	xxxxx
INF	xxxxx	xxxxx	xxxxx	xxxxx
IO	xxxxx	xxxxx	xxxxx	xxxxx
PSY	xxxxx	xxxxx	xxxxx	xxxxx
TBK	xxxxx	xxxxx	xxxxx	xxxxx
TCW	xxxxx	xxxxx	xxxxx	xxxxx
TG	xxxxx	xxxxx	xxxxx	xxxxx
TN	xxxxx	xxxxx	xxxxx	xxxxx
TW	xxxxx	xxxxx	xxxxx	xxxxx
WB	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 0.19: Statistische verdeling voorlichtingsprogramma's vrijdag (5 VWO)

Voorlichtingsprogramma Vrijdag 6 VWO	Studiekiezers	Deviatie	Extra personen	Deviatie
AT	xxxxx	xxxxx	xxxxx	xxxxx
ATLAS	xxxxx	xxxxx	xxxxx	xxxxx
BIT	xxxxx	xxxxx	xxxxx	xxxxx
BMT	xxxxx	xxxxx	xxxxx	xxxxx
CIT	xxxxx	xxxxx	xxxxx	xxxxx
CREATE	xxxxx	xxxxx	xxxxx	xxxxx
CT	xxxxx	xxxxx	xxxxx	xxxxx
EL	xxxxx	xxxxx	xxxxx	xxxxx
ES	xxxxx	xxxxx	xxxxx	xxxxx
GZW	xxxxx	xxxxx	xxxxx	xxxxx
IBA	xxxxx	xxxxx	xxxxx	xxxxx
INF	xxxxx	xxxxx	xxxxx	xxxxx
IO	xxxxx	xxxxx	xxxxx	xxxxx
PSY	xxxxx	xxxxx	xxxxx	xxxxx
TBK	xxxxx	xxxxx	xxxxx	xxxxx
TCW	xxxxx	xxxxx	xxxxx	xxxxx
TG	xxxxx	xxxxx	xxxxx	xxxxx
TN	xxxxx	xxxxx	xxxxx	xxxxx
TW	xxxxx	xxxxx	xxxxx	xxxxx
WB	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 0.20: Statistische verdeling voorlichtingsprogramma's vrijdag (6 VWO)

Voorlichtingsprogramma Vrijdag overige NL	Studiekiezers	Deviatie	Extra personen	Deviatie
AT	xxxxx	xxxxx	xxxxx	xxxxx
ATLAS	xxxxx	xxxxx	xxxxx	xxxxx
BIT	xxxxx	xxxxx	xxxxx	xxxxx
BMT	xxxxx	xxxxx	xxxxx	xxxxx
CIT	xxxxx	xxxxx	xxxxx	xxxxx
CREATE	xxxxx	xxxxx	xxxxx	xxxxx
CT	xxxxx	xxxxx	xxxxx	xxxxx
EL	xxxxx	xxxxx	xxxxx	xxxxx
ES	xxxxx	xxxxx	xxxxx	xxxxx
GZW	xxxxx	xxxxx	xxxxx	xxxxx
IBA	xxxxx	xxxxx	xxxxx	xxxxx
INF	xxxxx	xxxxx	xxxxx	xxxxx
IO	xxxxx	xxxxx	xxxxx	xxxxx
PSY	xxxxx	xxxxx	xxxxx	xxxxx
TBK	xxxxx	xxxxx	xxxxx	xxxxx
TCW	xxxxx	xxxxx	xxxxx	xxxxx
TG	xxxxx	xxxxx	xxxxx	xxxxx
TN	xxxxx	xxxxx	xxxxx	xxxxx
TW	xxxxx	xxxxx	xxxxx	xxxxx
WB	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 0.21: Statistische verdeling voorlichtingsprogramma's vrijdag (Overige NL)

Gemiddelde voorkeuren voor voorlichtingsprogramma's zaterdag

Voorlichtingsprogramma Zaterdag ROW	Studiekiezers	Deviatie	Extra personen	Deviatie
AT	xxxxx	xxxxx	xxxxx	xxxxx
ATLAS	xxxxx	xxxxx	xxxxx	xxxxx
BIT	xxxxx	xxxxx	xxxxx	xxxxx
BMT	xxxxx	xxxxx	xxxxx	xxxxx
CIT	xxxxx	xxxxx	xxxxx	xxxxx
CREATE	xxxxx	xxxxx	xxxxx	xxxxx
CT	xxxxx	xxxxx	xxxxx	xxxxx
EL	xxxxx	xxxxx	xxxxx	xxxxx
ES	xxxxx	xxxxx	xxxxx	xxxxx
GZW	xxxxx	xxxxx	xxxxx	xxxxx
IBA	xxxxx	xxxxx	xxxxx	xxxxx
INF	xxxxx	xxxxx	xxxxx	xxxxx
IO	xxxxx	xxxxx	xxxxx	xxxxx
PSY	xxxxx	xxxxx	xxxxx	xxxxx
TBK	xxxxx	xxxxx	xxxxx	xxxxx
TCW	xxxxx	xxxxx	xxxxx	xxxxx
TG	xxxxx	xxxxx	xxxxx	xxxxx
TN	xxxxx	xxxxx	xxxxx	xxxxx
TW	xxxxx	xxxxx	xxxxx	xxxxx
WB	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 0.22: Statistische verdeling voorlichtingsprogramma's zaterdag (ROW)

Voorlichtingsprogramma Zaterdag DE	Studiekiezers	Deviatie	Extra personen	Deviatie
AT	xxxxx	xxxxx	xxxxx	xxxxx
ATLAS	xxxxx	xxxxx	xxxxx	xxxxx
BIT	xxxxx	xxxxx	xxxxx	xxxxx
BMT	xxxxx	xxxxx	xxxxx	xxxxx
CIT	xxxxx	xxxxx	xxxxx	xxxxx
CREATE	xxxxx	xxxxx	xxxxx	xxxxx
CT	xxxxx	xxxxx	xxxxx	xxxxx
EL	xxxxx	xxxxx	xxxxx	xxxxx
ES	xxxxx	xxxxx	xxxxx	xxxxx
GZW	xxxxx	xxxxx	xxxxx	xxxxx
IBA	xxxxx	xxxxx	xxxxx	xxxxx
INF	xxxxx	xxxxx	xxxxx	xxxxx
IO	xxxxx	xxxxx	xxxxx	xxxxx
PSY	xxxxx	xxxxx	xxxxx	xxxxx
TBK	xxxxx	xxxxx	xxxxx	xxxxx
TCW	xxxxx	xxxxx	xxxxx	xxxxx
TG	xxxxx	xxxxx	xxxxx	xxxxx
TN	xxxxx	xxxxx	xxxxx	xxxxx
TW	xxxxx	xxxxx	xxxxx	xxxxx
WB	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 0.23: Statistische verdeling voorlichtingsprogramma's zaterdag (DE)

Voorlichtingsprogramma zaterdag 4 VWO	Studiekiezers	Deviatie	Extra personen	Deviatie
AT	xxxxx	xxxxx	xxxxx	xxxxx
ATLAS	xxxxx	xxxxx	xxxxx	xxxxx
BIT	xxxxx	xxxxx	xxxxx	xxxxx
BMT	xxxxx	xxxxx	xxxxx	xxxxx
CIT	xxxxx	xxxxx	xxxxx	xxxxx
CREATE	xxxxx	xxxxx	xxxxx	xxxxx
CT	xxxxx	xxxxx	xxxxx	xxxxx
EL	xxxxx	xxxxx	xxxxx	xxxxx
ES	xxxxx	xxxxx	xxxxx	xxxxx
GZW	xxxxx	xxxxx	xxxxx	xxxxx
IBA	xxxxx	xxxxx	xxxxx	xxxxx
INF	xxxxx	xxxxx	xxxxx	xxxxx
IO	xxxxx	xxxxx	xxxxx	xxxxx
PSY	xxxxx	xxxxx	xxxxx	xxxxx
TBK	xxxxx	xxxxx	xxxxx	xxxxx
TCW	xxxxx	xxxxx	xxxxx	xxxxx
TG	xxxxx	xxxxx	xxxxx	xxxxx
TN	xxxxx	xxxxx	xxxxx	xxxxx
TW	xxxxx	xxxxx	xxxxx	xxxxx
WB	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 0.24: Statistische verdeling voorlichtingsprogramma's zaterdag (4 VWO)

Voorlichtingsprogramma zaterdag 5 VWO	Studiekiezers	Deviatie	Extra personen	Deviatie
AT	xxxxx	xxxxx	xxxxx	xxxxx
ATLAS	xxxxx	xxxxx	xxxxx	xxxxx
BIT	xxxxx	xxxxx	xxxxx	xxxxx
BMT	xxxxx	xxxxx	xxxxx	xxxxx
CIT	xxxxx	xxxxx	xxxxx	xxxxx
CREATE	xxxxx	xxxxx	xxxxx	xxxxx
CT	xxxxx	xxxxx	xxxxx	xxxxx
EL	xxxxx	xxxxx	xxxxx	xxxxx
ES	xxxxx	xxxxx	xxxxx	xxxxx
GZW	xxxxx	xxxxx	xxxxx	xxxxx
IBA	xxxxx	xxxxx	xxxxx	xxxxx
INF	xxxxx	xxxxx	xxxxx	xxxxx
IO	xxxxx	xxxxx	xxxxx	xxxxx
PSY	xxxxx	xxxxx	xxxxx	xxxxx
TBK	xxxxx	xxxxx	xxxxx	xxxxx
TCW	xxxxx	xxxxx	xxxxx	xxxxx
TG	xxxxx	xxxxx	xxxxx	xxxxx
TN	xxxxx	xxxxx	xxxxx	xxxxx
TW	xxxxx	xxxxx	xxxxx	xxxxx
WB	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 0.25: Statistische verdeling voorlichtingsprogramma's zaterdag (5 VWO)

Voorlichtingsprogramma zaterdag 6 VWO	Studiekiezers	Deviatie	Extra personen	Deviatie
AT	xxxxx	xxxxx	xxxxx	xxxxx
ATLAS	xxxxx	xxxxx	xxxxx	xxxxx
BIT	xxxxx	xxxxx	xxxxx	xxxxx
BMT	xxxxx	xxxxx	xxxxx	xxxxx
CIT	xxxxx	xxxxx	xxxxx	xxxxx
CREATE	xxxxx	xxxxx	xxxxx	xxxxx
CT	xxxxx	xxxxx	xxxxx	xxxxx
EL	xxxxx	xxxxx	xxxxx	xxxxx
ES	xxxxx	xxxxx	xxxxx	xxxxx
GZW	xxxxx	xxxxx	xxxxx	xxxxx
IBA	xxxxx	xxxxx	xxxxx	xxxxx
INF	xxxxx	xxxxx	xxxxx	xxxxx
IO	xxxxx	xxxxx	xxxxx	xxxxx
PSY	xxxxx	xxxxx	xxxxx	xxxxx
TBK	xxxxx	xxxxx	xxxxx	xxxxx
TCW	xxxxx	xxxxx	xxxxx	xxxxx
TG	xxxxx	xxxxx	xxxxx	xxxxx
TN	xxxxx	xxxxx	xxxxx	xxxxx
TW	xxxxx	xxxxx	xxxxx	xxxxx
WB	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 0.26: Statistische verdeling voorlichtingsprogramma's zaterdag (6 VWO)

Voorlichtingsprogramma zaterdag overige NL	Studiekiezers	Deviatie	Extra personen	Deviatie
AT	xxxxx	xxxxx	xxxxx	xxxxx
ATLAS	xxxxx	xxxxx	xxxxx	xxxxx
BIT	xxxxx	xxxxx	xxxxx	xxxxx
BMT	xxxxx	xxxxx	xxxxx	xxxxx
CIT	xxxxx	xxxxx	xxxxx	xxxxx
CREATE	xxxxx	xxxxx	xxxxx	xxxxx
CT	xxxxx	xxxxx	xxxxx	xxxxx
EL	xxxxx	xxxxx	xxxxx	xxxxx
ES	xxxxx	xxxxx	xxxxx	xxxxx
GZW	xxxxx	xxxxx	xxxxx	xxxxx
IBA	xxxxx	xxxxx	xxxxx	xxxxx
INF	xxxxx	xxxxx	xxxxx	xxxxx
IO	xxxxx	xxxxx	xxxxx	xxxxx
PSY	xxxxx	xxxxx	xxxxx	xxxxx
TBK	xxxxx	xxxxx	xxxxx	xxxxx
TCW	xxxxx	xxxxx	xxxxx	xxxxx
TG	xxxxx	xxxxx	xxxxx	xxxxx
TN	xxxxx	xxxxx	xxxxx	xxxxx
TW	xxxxx	xxxxx	xxxxx	xxxxx
WB	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 0.27: Statistische verdeling voorlichtingsprogramma's zaterdag (Overige NL)

APPENDIX 19: VOORSPELLING VOORLICHTINGSPROGRAMMA'S 2016-11

Registratiecijfers	Totaal vrijdag	Deviatie	Totaal zaterdag	Deviatie
AT	xxxxx	xxxxx	xxxxx	xxxxx
ATLAS	xxxxx	xxxxx	xxxxx	xxxxx
BIT	xxxxx	xxxxx	xxxxx	xxxxx
BMT	xxxxx	xxxxx	xxxxx	xxxxx
CIT	xxxxx	xxxxx	xxxxx	xxxxx
CREATE	xxxxx	xxxxx	xxxxx	xxxxx
CT	xxxxx	xxxxx	xxxxx	xxxxx
EL	xxxxx	xxxxx	xxxxx	xxxxx
ES	xxxxx	xxxxx	xxxxx	xxxxx
GZW	xxxxx	xxxxx	xxxxx	xxxxx
IBA	xxxxx	xxxxx	xxxxx	xxxxx
INF	xxxxx	xxxxx	xxxxx	xxxxx
IO	xxxxx	xxxxx	xxxxx	xxxxx
PSY	xxxxx	xxxxx	xxxxx	xxxxx
TBK	xxxxx	xxxxx	xxxxx	xxxxx
TCW	xxxxx	xxxxx	xxxxx	xxxxx
TG	xxxxx	xxxxx	xxxxx	xxxxx
TN	xxxxx	xxxxx	xxxxx	xxxxx
TW	xxxxx	xxxxx	xxxxx	xxxxx
WB	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 0.28: Voorspelling registratiecijfers voorlichtingsprogramma's 2016-11

Opkomstcijfers	Totaal vrijdag	Deviatie	Totaal zaterdag	Deviatie
AT	xxxxx	xxxxx	xxxxx	xxxxx
ATLAS	xxxxx	xxxxx	xxxxx	xxxxx
BIT	xxxxx	xxxxx	xxxxx	xxxxx
BMT	xxxxx	xxxxx	xxxxx	xxxxx
CIT	xxxxx	xxxxx	xxxxx	xxxxx
CREATE	xxxxx	xxxxx	xxxxx	xxxxx
CT	xxxxx	xxxxx	xxxxx	xxxxx
EL	xxxxx	xxxxx	xxxxx	xxxxx
ES	xxxxx	xxxxx	xxxxx	xxxxx
GZW	xxxxx	xxxxx	xxxxx	xxxxx
IBA	xxxxx	xxxxx	xxxxx	xxxxx
INF	xxxxx	xxxxx	xxxxx	xxxxx
IO	xxxxx	xxxxx	xxxxx	xxxxx
PSY	xxxxx	xxxxx	xxxxx	xxxxx
TBK	xxxxx	xxxxx	xxxxx	xxxxx
TCW	xxxxx	xxxxx	xxxxx	xxxxx
TG	xxxxx	xxxxx	xxxxx	xxxxx
TN	xxxxx	xxxxx	xxxxx	xxxxx
TW	xxxxx	xxxxx	xxxxx	xxxxx
WB	xxxxx	xxxxx	xxxxx	xxxxx

Tabel 0.29: Voorspelling opkomstcijfers voorlichtingsprogramma's 2016-11

APPENDIX 20: EXCELSHEET

Figuur 0.88: Voorspelling verwachte aantal registraties per cluster

Figuur 0.89: Voorspelling verwachte aantal registraties per dag

Figuur 0.90: Voorspelling verwachte aantal registraties per voorlichtingsprogramma (vrijdag)

Figuur 0.91: Voorspelling verwachte aantal registraties per voorlichtingsronde (zaterdag)

Figuur 0.92: Statistische onderverdeling naar voorlichtingsronde

Figuur 0.93: Winters' Method voor 4 VWO studiezoekers