

# Docentenhandleiding

# Statistiek en computationeel denken

Data onderzoeken met Excel

Lesmateriaal Computationeel en wiskundig denken



Universiteit Utrecht



## Inhoudsopgave

Inhoudsopgave.....	2
Vooraf .....	4
Vorbereiding .....	4
Uitvoering .....	5
Toetsing en data.....	5
Uitgangspunten van het lesmateriaal.....	6
Leerdoel en opbouw van de lessenserie.....	7
Werkwijze .....	7
Lesvoorbereiding .....	7
Uitwerkingen, tips en suggesties voor vragen .....	8
Hoofdstuk 1. Probleemoplossen met computationeel denken.....	8
Hoofdstuk 2. Titanic.....	8
2.1 Vrouwen en kinderen eerst? .....	8
2.2 Stappenplan uitwerken in Excel.....	10
2.3 Andere factoren .....	12
Hoofdstuk 3. Van 'Cijfers? Boeien!' naar 'Cijfers boeien!' .....	14
3.1 Kies een dataset .....	14
3.2 Stappenplan uitwerken in Excel.....	15
3.3 Verslag.....	16
Docentenlogboek.....	17

## **Colofon**

Dit lesmateriaal is ontwikkeld in het kader van het NRO langlopend onderzoek “Computationeel en wiskundig denken”, projectnummer 40.5.18540.130.

Doelgroep: vwo-5 wiskunde A

Auteurs: Sylvia van Borkulo, Marcel Daems, Maria Kallia, John Val, Tim Leenders, Christ van den Brand

Status: versie 5, 15 november 2021

© Freudenthal Instituut, Universiteit Utrecht

## Vooraf

Dit document is een docentenhandleiding bij het lesmateriaal rond computationeel denken in vwo-5 wiskunde A, met als titel *Statistiek en computationeel denken. Data onderzoeken met Excel*. Een kanttekening hierbij is dat deze handleiding voor een deel geschreven is “vanachter het bureau”. We vragen de docenten die het gebruiken dan ook om dit document zoveel mogelijk aan te vullen, zodat ook docenten die minder bekend zijn met de gedachten achter de lessenserie met het materiaal uit de voeten kunnen. Naast deze handleiding is er een Engelstalige Rationale, die de theoretische en onderzoeksmatige achtergrond verder belicht.

## Voorbereiding

Voor de lessenserie van start kan gaan is een aantal voorbereidende stappen nodig.

- Toestemming  
Omdat de lessen worden geobserveerd en data van leerlingen worden verzameld, is het nodig dat de leerlingen (en hun ouders, afhankelijk van de leeftijd) vooraf een toestemmingsformulier ondertekenen en inleveren. Een format voor dit formulier wordt aangeleverd. Het is prettig om dit voorafgaand aan de eerste les te regelen.
- Repro lesmateriaal  
Reproductie van het lesmateriaal kan op de school zelf plaatsvinden of worden aangeleverd vanuit de UU.
- Informatie leerlingen  
Het is goed om de leerlingen in te lichten over de doelen van de lessenserie en het achterliggende onderzoek. Daarbij is het wellicht belangrijk om bij de leerlingen te benadrukken dat de wiskunde die aan de orde komt van belang is voor het examen, dus dat het niet gaat om extra onderwerpen of verloren tijd. Daarnaast is het goed om duidelijk te maken of hun werk op een of andere manier meetelt of wordt beloond. Of dit gebeurt en hoe is aan de docent zelf.
- ICT-faciliteiten  
De lessen maken gebruik van Excel, dus het is van belang dat leerlingen dat in de les kunnen gebruiken. Dat kan via een computerlokaal, of via laptops, chromebooks of tablets in de klas. Met het oog op klassengesprekken en nabesprekingen is het ook goed als een Excelscherm kan worden geprojecteerd.
- File management  
Van tevoren moet worden afgesproken waar de leerlingen hun resultaten opslaan zodat de docent (en de onderzoekers uiteindelijk ook) er toegang toe hebben. Ook dit wordt aan de docent zelf overgelaten. Het is denkbaar dat leerlingen per tweetal Excel-bestanden maken. Een eigen digitaal werkboek waarin de vragen kunnen worden ingevuld is wel wenselijk voor elke leerling.
- Voorkennis  
De wiskundige voorkennis van de leerlingen betreft kennis van statistiek en kennis van Excel. De voorkennis van leerlingen wat betreft Excel kan worden geactualiseerd met een introductiepracticum van 1 lesuur, zie bijvoorbeeld Getal & Ruimte 11e editie Hoofdstuk 9, 9.6 Werken met Excel, en 9.7 Werken met datasets. Statistiekennis zoals het gebruik van draaitabellen en odds ratio kan worden opgefrist met Getal & Ruimte 11e editie Hoofdstuk 11, 11.6 Draaitabellen en draaigrafieken, en 11.7 Data analyseren. De tijd die dit kost, is niet ingecalculeerd in de vijf lessen die dit materiaal zal kosten.

#### - Tijdsplanning

We weten nog niet precies hoe snel leerlingen van vwo-5 wiskunde A door dit materiaal gaan. De opdrachten laten veel ruimte voor eigen inbreng. Wellicht is differentiatie mogelijk wat betreft het aantal vragen of het soort vragen dat leerlingen gaan onderzoeken.

### Uitvoering

Tijdens de uitvoering zijn de volgende aandachtspunten van belang.

- De lessenserie gaat over computationeel denken. Het moet de leerlingen duidelijk worden dat ze m.b.v. computationeel denken vragen bij datasets gaan beantwoorden. Het is dus van belang om tijdens de lessenserie op gezette momenten (bijv. in klassengesprekken) de aspecten van computationeel denken te benoemen die aan de orde zijn gekomen, zoals gegevens verkennen, ordenen, bewerken, probleem opdelen in kleinere deelproblemen, precies formuleren, formules gebruiken en daarbij parameters precies invullen, en evalueren of het juiste antwoord gevonden is. Als voorbeeld kan gebruik gemaakt worden van een video ter illustratie van computationeel denken en het belang van het exact beschrijven van je stappen (peanutbutter-jelly video): <https://fb.watch/45ZvJfAjpp/>
- Benadruk bij leerlingen dat ze hun **antwoorden invullen** in hun werkboekjes en dat de werkboekjes en Excel-bestanden worden ingenomen, en zorg ervoor dat dit aan het einde van de lessenserie ook daadwerkelijk gebeurt.
- Bespreek svp in de les ook klassikaal de kernopgaven en vraag daarbij aan leerlingen of het is gelukt, wat ze hebben gedaan, om hun aanpak uit te leggen, om de moeilijkheden die ze hadden te beschrijven, om de Excel-technieken uit te leggen, etc. In de lesvoorbereidingen aan het einde van dit document wordt aangegeven welke opgaven kernopgaven zijn en staat een aantal suggesties voor vragen die bij deze klassengesprekken aan de orde kunnen komen.
- Vul na elke les svp zelf een logboekblad in; zie hiervoor het template, maar voel je vrij daarvan af te wijken.
- Geef wijzigingen in roostergegevens (tijdstippen, lokalen, lesduur, etc.) en door aan de beoogde observanten.

### Toetsing en data

Het toetspakket bij deze lessenserie bestaat uit drie elementen:

1. Een voortoets voorafgaand aan de lessenserie, die wordt aangeleverd vanuit het onderzoek. Het moet de leerlingen duidelijk zijn dat het om computationeel denken gaat. Deze papier-en-pen toets duurt naar schatting 25 minuten. Bij deze toets is een beoordelingsmodel/correctievoorschrift beschikbaar. Nakijken ervan kan door leden van het onderzoeksteam worden gedaan.
2. Een eindopdracht die leerlingen in drietallen maken en inleveren. Voorstel is om hiervoor hoofdstuk Cijfers uit het lesmateriaal te gebruiken. Beoordeling hiervan kan door de docent i.s.m. met het onderzoeksteam worden gedaan. De docent beslist of en op welke manier het resultaat voor leerlingen meetelt.

3. Natoets na afloop: deze bestaat uit twee delen, een eerste algemene deel, vergelijkbaar met de voortoets, en een tweede deel dat specifiek op de lessenserie ingaat. Het moet de leerlingen duidelijk zijn dat het om computationeel denken gaat. Het is een pen-en-papier toets van in totaal 50 minuten. Bij deze toets is een beoordelingsmodel/correctievoorschrift beschikbaar. Nakijken ervan kan door leden van het onderzoeksteam worden gedaan. De docent beslist of en op welke manier het resultaat voor leerlingen meetelt.

Onderzoeksdata die worden verzameld omvatten:

1. Leerlingenwerk voortoets: graag scannen en als pdf aan onderzoeksteam ter beschikking stellen.
2. Leerlingwerkboeken: in geval van papier graag scannen en als pdf aan onderzoeksteam ter beschikking stellen.
3. Leerlingenwerk eindopdracht: graag scannen en als pdf aan onderzoeksteam ter beschikking stellen.
4. Leerlingenwerk natoets: graag scannen en als pdf aan onderzoeksteam ter beschikking stellen.
5. Logboek van uitvoerende docent.
6. Online interviews die onderzoeker afneemt via een laptop tijdens de lessenserie, in overleg en met medewerking van de docent.
7. Het laatste rapportcijfer voor wiskunde van de leerlingen.

### **Uitgangspunten van het lesmateriaal**

Bij het ontwerpen van het lesmateriaal is een aantal uitgangspunten leidend geweest. Ten eerste moet de wiskundige inhoud aansluiten bij het programma van wiskunde A in vwo-5. We richten ons in de lessenserie op statistiek en het onderzoeken van data. Leerlingen hebben hier in principe al kennis van.

Een tweede uitgangspunt is dat computationeel denken een rol speelt. Wat nieuw is in de behandeling van het onderwerp statistiek, is de expliciete link met computationele denkvaardigheden, zoals dataverkenning, datarepresentatie en algoritmisch denken. Een systematische aanpak waarbij steeds wordt gereflecteerd op de aanpak en het doel is hier van belang. Leerlingen ontwikkelen inzicht in het gebruik van een digitale tool om vragen bij een dataset te beantwoorden. De tool staat in dienst van de vraag en het onderzoek naar het antwoord.

Ten derde hebben we gekozen voor een algemeen bekende tool Excel, dat ook van nut is in de toekomst van de leerling.

Een vierde uitgangspunt is dat er een balans is tussen 'unplugged' en 'plugged' activiteiten. Unplugged activiteiten zijn erop gericht om voldoende aandacht te besteden aan het inzetten van computationele vaardigheden zoals het verkennen en begrijpen van data en om te voorkomen dat leerlingen verzanden in 'Excel-klikken'. Bij iedere data-vraag wordt de leerling eerst gevraagd om een stappenplan op te stellen waarmee ze de vraag willen beantwoorden. Plugged activiteiten bestaan uit het bewerken van de data met Excel-commando's zoals AANTALLEN.ALS en het uitvoeren van geschikte data-representaties met de juiste variabelen.

Ten vijfde willen we dat iedere leerling uit de voeten kan met de opgaven. Er is ruimte om de vragen bij de data aan te passen aan de leerling, er kan meer of minder ondersteuning en tips worden gegeven.

## Leerdoel en opbouw van de lessenserie

	Leerling activiteit	Leerdoelen wiskundig denken	Leerdoelen computationeel denken	Excel-technieken
Hoofdstuk 1 Probleemoplossen met computationeel denken	- Leerling maakt kennis met computationeel denken	-	-	-
Hoofdstuk 2 Titanic	- Leerling maakt begeleid opdrachten om de data van de Titanic te verkennen. - Leerling onderzoekt begeleid de vraag 'vrouwen en kinderen eerst?'	- Gemiddelden vergelijken - Groepen vergelijken	- Data verkennen - Data bewerken -	- Sorteren - Rijen verwijderen - Gebruik ALS - Gebruik AANTALLEN.ALS
Hoofdstuk 3 Van 'Cijfers? Boeien!' naar 'Cijfers boeien!'	- Leerling formuleert zelf een vraag bij de dataset.	- Gemiddelden vergelijken - Groepen vergelijken - Mogelijk regressie, odds ratio, normale verdeling	- Data verkennen - Data bewerken - Probleem formuleren - Probleem opdelen	- Idem - Mogelijk regressie

## Werkwijze

De leerlingen werken in kleine groepjes van twee of drie leerlingen zodat ze kunnen overleggen en verschillende ideeën kunnen bespreken.

Zo nodig kunnen opdrachten individueel gedaan worden om meer vaart in de les te krijgen.

Hieronder staan op bepaalde momenten in de lessenserie vragen die gebruikt kunnen worden om te reflecteren op het leerproces van de leerling en om te discussiëren over het werk van de leerlingen. De docent kan hier uiteraard een keuze in maken, al naar gelang de voortgang en de behoefte in de les. De vragen worden in het onderzoek ook gebruikt voor de leerling-interviews.

## Lesvoorbereiding

Hieronder staat een overzicht van de hoofdstukken en de indeling van de opgaven binnen de hoofdstukken. De kernopgaven zijn: 2.1, 2.2, 3.1, 3.2.

Hoofdstuk	Probleem-/data-verkenning	Uitwerken in Excel	Verdieping
-----------	---------------------------	--------------------	------------

Hoofdstuk 1 Probleemoplossen met computationeel denken			
Hoofdstuk 2 Titanic	2.1	2.2	2.3
Hoofdstuk 3 Van 'Cijfers? Boeien!' naar 'Cijfers boeien!'	3.1	3.2	3.3 verslag

Een suggestie voor de les-/tijdsindeling is:

Voortoets 45 minuten

Hoofdstuk 1 t/m Hoofdstuk 2.1: 30 minuten

Hoofdstuk 2.2: 60 minuten

Hoofdstuk 2.3: 30 minuten

Hoofdstuk 3.1: 30 minuten

Hoofdstuk 3.2: 60 minuten

Hoofdstuk 3.3 (verslag): 30 minuten

Natoets 50 minuten

Vragenlijst 15 minuten

### Uitwerkingen, tips en suggesties voor vragen

Hieronder volgt een verdere uitwerking met tips en vragen die kunnen helpen bij een klassikale discussie. De docent kan hieruit een keuze maken, al naar gelang de voortgang en de behoefte in de les.

### Hoofdstuk 1. Probleemoplossen met computationeel denken

Hele klas:

Introductie onderwerp computationeel denken, Hoofdstuk 1. Probleem oplossen met computationeel denken, introductie Hoofdstuk 2. Titanic

Link naar video (3:26) ter illustratie van computationeel denken en het belang van het exact beschrijven van je stappen (peanutbutter-jelly video):

<https://fb.watch/45ZvJfAjpp/>

### Hoofdstuk 2. Titanic

#### 2.1 Vrouwen en kinderen eerst?

Deze paragraaf draait om het verkennen van de data.

Op het tweede tabblad van het Titanic-bestand staat de beschrijving van de kolommen.

Mogelijke vragen om te bespreken:

- Hoe zijn de gegevens in Excel weergegeven?
- Welke informatie kun je uit de gegevens halen?
- Waarom wordt de variabele 'survived' in binaire gehele getallen weer gegeven en niet bijvoorbeeld in tekst, zoals 'overleefd' en 'niet-overleefd'? Wat kan het voordeel zijn van het



gebruik van binaire getallen? Met binaire getallen is het eenvoudiger om berekeningen en bewerkingen uit te voeren.

- Zijn er ontbrekende gegevens in de dataset? D.w.z., zijn er bijv. passagiers van wie de passagiersklasse of de leeftijd niet ingevuld is?

*VRAAG A: Is bij de ramp van de Titanic de volgende uitspraak waar? 'Vrouwen en kinderen werden eerst van boord gehaald.'*

Met deze vraag wordt het te onderzoeken probleem weergegeven. Was dit principe geldig bij de ramp van de Titanic?

Om deze vraag te kunnen beantwoorden moet de juiste wiskunde/Excel-techniek met de juiste variabelen worden gekozen. Belangrijke variabelen bij deze onderzoeksvraag zijn geslacht, leeftijd en overleefd. De dataset geeft echter niet aan wanneer iemand als kind of volwassene wordt beschouwd. Deze informatie kan worden gehaald uit de variabele leeftijd. Je kunt een nieuwe kolom in Excel aanmaken met **ALS (E2 < 14, "kind", "volwassene")**. In deze formule is gekozen voor de leeftijdsgrens van 14 jaar tussen volwassen en kind. Dit is iets om over na te denken en zelf een beslissing over te nemen.

Vragen die met de klas besproken kunnen worden:

- Wat heb je gedaan om de data te verkennen en waarom?
- Heb je alle informatie om de vraag te kunnen beantwoorden?
- Hoe ga je het aanpakken?
- Welke stappen zijn er nodig om de vraag te kunnen beantwoorden?
- Welke formules zou je in Excel kunnen gebruiken om een nieuwe kolom aan te maken?

Leerlingen kunnen op het gebruik van de Excel spiekbrieven gewezen worden.

*Vraag 1: Welke informatie heb je nodig om vraag A te kunnen beantwoorden?*

Het gaat hier om welke informatie er allemaal in de dataset aanwezig is. Het is handig om een nieuwe kolom aan te maken. Sorteren kan helpen om te bepalen welke rijen 'volwassen' en welke 'kind' zijn. Op die manier ontstaat er snel een overzicht van het aantal aanwezige kinderen en volwassenen op de Titanic. Let op dat een andere sortering een bepaalde berekening misschien in de war gooit. Het kan makkelijk rekenen als deze nieuwe variabele binaire waarden heeft.

Voorkomende problemen bij het berekenen van de nieuwe kolom 'volwassenheid' (is iemand kind of volwassene), afhankelijk van de instellingen in Excel:

- De leeftijd 0.1 is geen getal. Dit heeft te maken met Nederlandse versus Engelse versie van Excel. Volgens het Nederlandse formaat is 0,1 een getal en 0.1 niet.

*Vraag 2: Is deze informatie beschikbaar in de dataset?*

Vervolgens kan er nog gekeken worden of alle informatie compleet is. Er zijn in de dataset regels met ontbrekende gegevens. Van sommige passagiers ontbreekt de naam en van sommige passagiers de leeftijd. De variabele leeftijd geeft belangrijke informatie voor het beantwoorden van de vraag, terwijl

de variabele naam geen belangrijke informatie oplevert. De regels met ontbrekende gegevens over de variabele leeftijd zouden dan ook verwijderd kunnen worden. Hiervoor moet de dataset op leeftijd gesorteerd worden en vervolgens moeten de regels met ontbrekende gegevens verwijderd worden.

*Vraag 3: Welke stappen moet je doen om vraag A te beantwoorden?*

Je zou eerst alle informatie kunnen verzamelen en gegevens sorteren. Vervolgens kun je ontbrekende gegevens die relevant zijn voor de vraag verwijderen uit de dataset. Hierna kan er over worden gegaan op de berekeningen. Om een antwoord te vinden op de onderzoeksvraag, moet er vast worden gesteld hoeveel mensen de ramp hebben overleefd. Hierna kan er gekeken worden naar hoeveel mannelijke volwassenen, vrouwelijke volwassenen en kinderen het hebben overleefd, en hoeveel niet, waarbij het makkelijker is om dit in percentages uit te drukken.

*Vraag 4: Hoe kun je deze stappen vertalen in Excel?*

Vragen die je hier kunt stellen zijn: Welke formules heb je nodig om het antwoord te kunnen berekenen? Welke informatie neem je allemaal mee in je formule? Hoe kun je de leeftijd van een persoon meenemen in de formule voor Excel? Voordat leerlingen in Excel aan het rekenen slaan is het goed om stil te staan bij de stappen die genomen gaan worden. De relevante kolommen zijn geslacht, overleefd, volwassene.

Om het percentages overlevenden per geslacht te vinden, is de formule AANTALLEN.ALS van belang. Deze formule geeft weer hoeveel percent van de variabele de ramp heeft overleefd aan de hand van meerdere criteria. De criteria zijn in dit geval het geslacht (male/female) en de leeftijd (volwassen/kind). Deze formule kan voor de mannelijke overlevenden worden toegepast, en vervolgens op de vrouwelijke overlevenden, en de kinderen.

## 2.2 Stappenplan uitwerken in Excel

In Titanic-Excel-bestand "20201126-Titanic\_ANTWOORDEN.xlsx" staat een voorbeeld-uitwerking.

*Vraag 5: Voer je stappenplan uit en vind een antwoord op vraag A.*

Om een antwoord te vinden op de vraag, kunnen de percentages mannen, vrouwen en kinderen die het hebben overleefd bekeken worden. Hiervoor kan de formule AANTALLEN.ALS gebruikt worden.

Tip: voor het uitvoeren van deze stappen is het handig om de gegevens zo te sorteren dat de passagiers die het hebben overleefd bovenaan staan. Op die manier is het eenvoudiger om rijen te vermelden in de formule AANTALLEN.ALS.

Voorbeeld-oplossing: Nadat je de gegevens zo hebt gesorteerd dat de overlevenden bovenaan staan, klik je op cel I4 en typ je de volgende formule om het aantal overlevende mannen te tellen: AANTALLEN.ALS (D2: D428, "male", F2: F428, "volwassene" ). Let op dat de uitkomst van deze formule verandert, als de sortering verandert.

In plaats van sorteren, kun je ook meer voorwaarden gebruiken in de *AANTALLEN.ALS* -formule die je hierboven hebt gebruikt. Als je bijvoorbeeld het aantal de overlevende mannen wilt berekenen, kun je de volgende formule gebruiken als je de gegevens niet sorteert: *AANTALLEN.ALS* (D2: D1047, "male", F2: F1047, "volwassene", A2: A1047, "= 1" ). Op dezelfde manier kun je het aantal vrouwen en kinderen berekenen dat het heeft overleefd

Voorkomende problemen:

- Hoe sorteer je? Zie menu Data en dan Sort. Kies de kolom waarop je wilt sorteren.
- Leerling gebruikt *AANTAL.ALS* i.p.v. *AANTALLEN.ALS*. Dan kun je maar één criterium gebruiken. Met *AANTALLEN.ALS* kun je meerdere criteria gebruiken.
- Als je *AANTALLEN.ALS* gebruikt, verschijnt een volgend bereik/criterium, als je klikt in het laatste criterium. Soms moet je even scrollen omdat het venster te klein is.
- Waar moet je die formule intypen? Ga in de cel staan waar je het aantal wil hebben, en typ '=' of klik op het formule teken boven in de balk.
- Percentage mannen overleefd. Percentage waarvan? De berekende percentages zijn t.o.v. het aantal overlevenden. Maar hoeveel mannen zaten er überhaupt aan boord? Als dat aantal klein is, verwacht je onder de overlevenden ook weinig mannen... Je kunt dus beter kijken naar het totaal aantal mannen.
- Maak een kruistabel om de percentages te berekenen.

*Vraag 6: Wat maakt je, uitgaande van je berekeningen, zeker dat je antwoord op vraag 5 waar is? Kun je met zekerheid zeggen dat de uitspraak waar of niet waar is?*

Wanneer trek je een conclusie? Hoe weet je zeker of het 'waar' is? Wanneer is het verschil groot genoeg?

In plaats van naar de overlevingspercentages te kijken, kan er hier ook gekeken worden naar de overlevingskansen voor de verschillende categorieën. De overlevingskansen voor mannen, vrouwen en kinderen kunnen berekend worden aan de hand van de formule *ODDS RATIO*.

Uitleg odds-ratio: <https://www.youtube.com/watch?v=yLtOkUmo4i8>.

Hoe groot is de kans voor vrouwen en kinderen om de ramp te overleven?

*Vraag 7: Herken je de stappen die je gemaakt hebt in de cyclus?*

Hele klas:

Discussie over het werk van de groepjes. Onderstaande vragen kunnen hierbij helpen.

Hoe ging het? Waar liep je tegenaan? Hoe kon je dat oplossen?
Heb je de data gesorteerd? Waarom wel/niet? Hoe heeft dat jou geholpen?

Hoe ging je om met ontbrekende gegevens? Heb je data verwijderd? Waarom?
Wat heb je gedaan om te bepalen hoeveel mensen de ramp overleefd hebben?
Begrijp je wat de commando's ALS, AANTALLEN.ALS betekenen en hoe het werkt?
Heb je de (odds-ratio) tabel gebruikt? Wat betekenen de getallen in je (odds-ratio) tabel?
Kun je met zekerheid zeggen dat de uitspraak waar of niet waar is?
Heb je de computationele onderzoekscyclus gebruikt? Hoe?

### 2.3 Andere factoren

*VRAAG B: Zijn er andere factoren die de overlevingskans beïnvloeden?*

*8. Gebruik de computationele onderzoekscyclus om vraag B te onderzoeken en beantwoorden.*

Mogelijke factoren die bekeken kunnen worden zijn leeftijd en klasse. Een handige manier om te onderzoeken of andere factoren de overlevingskans beïnvloedde, is door gebruik te maken van beschrijvende statistiek. Elke factor (leeftijd/klasse) kan beschreven worden aan de hand van het gemiddelde, de mediaan, de standaard deviatie en de modus.

Je zult merken dat de gemiddelde leeftijd van de overlevenden nagenoeg hetzelfde is bij de niet-overlevenden als bij de overlevenden, en hetzelfde geldt voor de mediaan en de standaarddeviatie. Het lijkt er dus op dat alleen de leeftijd de overlevingskans niet heeft beïnvloed (in de praktijk hebben we meer statistieken nodig om deze vraag te beantwoorden, maar deze vallen buiten het doel van deze oefening).

Een andere aanpak om dit te onderzoeken is door de leeftijd op te splitsen in categorieën met gelijke bereiken en de leeftijdscategorie te identificeren met het hoogste percentage overlevenden en degene met de minste en de overlevingskansen per categorie te vergelijken.

Je kunt zien dat de leeftijdsgroep met het laagste percentage overlevenden de leeftijdsgroep is met leeftijden tussen 61-70 jaar oud, en de leeftijdsklasse met het hoogste percentage overlevenden die met passagiers van 1-10 jaar oud. In feite was de kans om te overleven 3,9 groter als een passagier tussen de 1 en 10 jaar oud was dan voor passagiers tussen 61 en 70 jaar. Voor bepaalde leeftijdscategorieën was er dus wel degelijk een hogere of lagere kans om te overleven. De vraag is echter of de variabele leeftijd als een factor gezien kan worden. De factor leeftijd kan niet als een factor beschouwd worden, doordat deze niet normaal verdeeld is.

Als je kijkt naar de passagiersklasse, zie je dat klasse A het hoogste percentage overlevenden heeft vergeleken met de andere twee klassen, waarbij klasse C het laagste percentage heeft. In feite had een passagier 4,96 meer kans om te overleven als hij/zij in de A-klasse zat vergeleken met een passagier in

de C-klasse. Het lijkt er dus op dat de klasse van een passagier een mogelijke factor is die de overlevingskansen beïnvloedt.

Zijn er combinaties van factoren mogelijk die de overlevingskans hebben beïnvloed?  
Bijv., was leeftijd een mogelijke overlevingsfactor per passagiersklasse?

Mogelijke berekeningen: percentage en kansen per klasse. Wanneer er naar de overlevingskans per klasse gekeken wordt, is het handig om deze percentages uit te drukken in een grafiek. Maak grafieken/diagrammen om het percentage overlevenden per klasse weer te geven.

Welke soort grafiek of diagram is het handigst? Wat zijn de voor- en nadelen van een grafiek of diagram i.p.v. een tabel?

## Hoofdstuk 3. Van 'Cijfers? Boeien!' naar 'Cijfers boeien!'

### 3.1 Kies een dataset

In deze opdracht draait het om het opstellen en onderzoeken van een eigen vraag bij een van de drie gegeven datasets. Correlatie en regressie zou hierbij aan bod kunnen komen. Daar kan dan bij het oproepen en bespreken van de vragen op aangestuurd worden.

Als voorbeeld is hier de dataset "Leerlingen\_17-18\_anoniem.xlsx" genomen.

- Klassendiscussie

Wat voor vragen roept de beschrijving van de dataset op?

We kunnen kijken naar de onderlinge verbanden tussen de schoolvakken. Wat kun je zeggen over de samenhang van het cijfer voor bijvoorbeeld Nederlands met andere vakken? Of met profielkeuze? Hoe is de relatie tussen het cijfer voor Nederlands en het cijfer voor wiskunde, Engels, economie, ... . Verwacht je dat leerlingen die goed zijn in talen minder goed zijn in wiskunde, juist beter? Zijn leerlingen met een EM-profiel beter in Nederlands dan leerlingen met een NG-profiel? Is er een verschil tussen Havo en VWO?

De vragen worden door de docent op een rij gezet en al naar gelang hoeveel er zijn, kan ieder groepje enkele vragen onderzoeken of worden de vragen verdeeld over de groepjes. Leerlingen worden in kleine groepjes ingedeeld om te werken aan deze opdracht.

Onderstaande vragen kunnen helpen bij de klassikale discussie.

Wat kun je zeggen over de data als je alleen naar bestand 'beschrijving dataset cijfers.xlsx' kijkt?
--

Welk vragen of verbanden komen in je op? Waarom denk je dat deze interessant zijn om te onderzoeken?
--

#### **Suggesties voor verdieping**

Sommige mensen beweren dat meisjes beter zijn in sociale vakken dan jongens, en dat jongens op hun beurt beter zijn in wiskunde en bèta-vakken. Onderzoek m.b.v. de dataset(s) of dit waar is voor de betreffende school.

Om bovenstaande in meer detail te onderzoeken, kun je overwegen hoe het profiel van de leerling al dan niet hiermee samenhangt. Hoe presteren meisjes en jongens in sociale vakken, wiskunde en bètavakken, als we hun profiel in overweging nemen?

- Onderbouw je antwoord met berekeningen en grafieken, die samenhang laten zien.
- Let erop dat je iedere stap die je in Excel doet om de vraag te beantwoorden beschrijft in je verslag.

Kleine groepjes:

3.1 vraag 1-4 (Vraag specificeren en data verkennen, nadenken hoe Excel te gebruiken om vraag te beantwoorden.)

1. Welke informatie heb je nodig?  
Verken de dataset door te kijken naar missende waarden, type data, maximum/minimum waarden, gemiddelde, welke groepen leerlingen zijn er?
2. Is deze informatie beschikbaar in de dataset?
3. Hoe krijg ik deze informatie beschikbaar in de dataset?
4. Welke stappen moet je doen om je vraag te toetsen?
5. Hoe kun je deze stappen vertalen in Excel?  
Hier kan regressie en correlatie worden gebruikt, zoals beschreven in de bijlage.

Hele klas:

Discussie over het werk van de groepjes. Onderstaande vragen kunnen hierbij helpen.

Heb je genoeg data om je vraag te beantwoorden? Welke data ga je gebruiken en waarom? Moet je de data nog bewerken? Waarom?
Hoe ga je het aanpakken? Welke stappen die nodig zijn om je probleem op te lossen? Waarom deze stappen?
Welke wiskundige concepten of statistische begrippen heb je nodig?
Hoe denk je dat je Excel kunt gebruiken om je probleem op te lossen?

### 3.2 Stappenplan uitwerken in Excel

Kleine groepjes:

3.2 vraag 6-7 (Stappenplan uitwerken in Excel en daarop reflecteren.)

Ook hier geldt: leerlingen moeten niet zomaar gaan freewheelen in Excel, maar werken volgens een plan.

6. Voer je stappenplan uit in Excel en vind een antwoord op je vraag.
7. Ben je afgeweken van je oorspronkelijke stappenplan? Voldeed het aan je verwachtingen wat betreft de tijdsduur?  
Welke vragen waren wel/niet geschikt om te onderzoeken?  
Welke vragen waren makkelijk te beantwoorden, welke moeilijker?  
Vergelijk de verschillende aanpakken van de leerlingen. Welke waren succesvol, welke minder?  
Kostte het meer tijd dan verwacht?

Hele klas:

Discussie over het werk van de groepjes. Onderstaande vragen kunnen hierbij helpen.

Wat heb je gedaan om je vraag te beantwoorden?
Heb je de data bewerkt? Waarom wel/niet? Hoe heeft dat jou geholpen?
Welke commando's heb je gebruikt? Begrijp je wat ze betekenen en hoe het werkt? (Bijv. COUNT, AVERAGE, IF)
Wat betekenen de getallen in je (odds-ratio) tabel?
Was de vraag makkelijk te beantwoorden of moeilijk?
Voldeed het aan je tijdsverwachtingen?
Heb je de computationele onderzoekscyclus gebruikt en zo ja, hoe?
Hoe kan een tool zoals Excel gebruikt worden om een probleem in de echte wereld op te lossen?

### 3.3 Verslag

Kleine groepjes:

8. Schrijf een verslag over je eigen vraag/vragen.

Hints om het verslag vorm te geven:

- Volg computationele onderzoekscyclus.
- Voeg ook een beschrijving van een onsuccesvolle aanpak toe en leg uit wat je daarvan geleerd hebt.

Suggesties voor de inhoud:

- inleiding
- de onderzochte vraag
- de aanpak om de vraag te beantwoorden
- waarom nam je deze stappen?
- wat ben je tegengekomen?
- het antwoord op de vraag
- conclusie
- bronvermelding



## Docentenlogboek

Er is ook een los document met het docentenlogboek.

### Les #

Lesduur: ...

Datum: ...

Klas, docent: ...

#### Globale lesopbouw en inhoud:

#### Opvallende reacties van leerlingen:

#### Specifieke opmerkingen bij opgaven:

#### Terugblik: Waarover ben je tevreden en waarover minder tevreden?

#### Terugblik: Tip voor een collega die deze les nog gaat uitvoeren?