

Statistiek en computationeel denken

Data onderzoeken met Excel

Lesmateriaal Computationeel en wiskundig denken



Universiteit Utrecht



Colofon

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

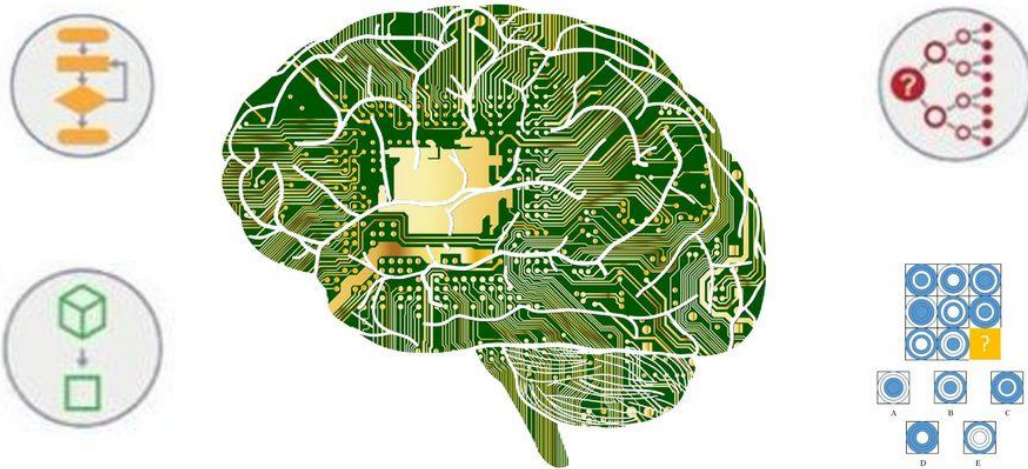
Doelgroep: vwo-5 wiskunde A

Auteurs: Sylvia van Borkulo, Marcel Daems, John Val, Christos Chytas

Status: versie 3.0, 20 oktober 2021 (herontwerp gebaseerd op versie 1.0 31 maart 2021)

© Freudenthal Instituut, Universiteit Utrecht

computationeel denken



bron: plaatje uit presentatie <https://slideplayer.nl/slide/17062405/>

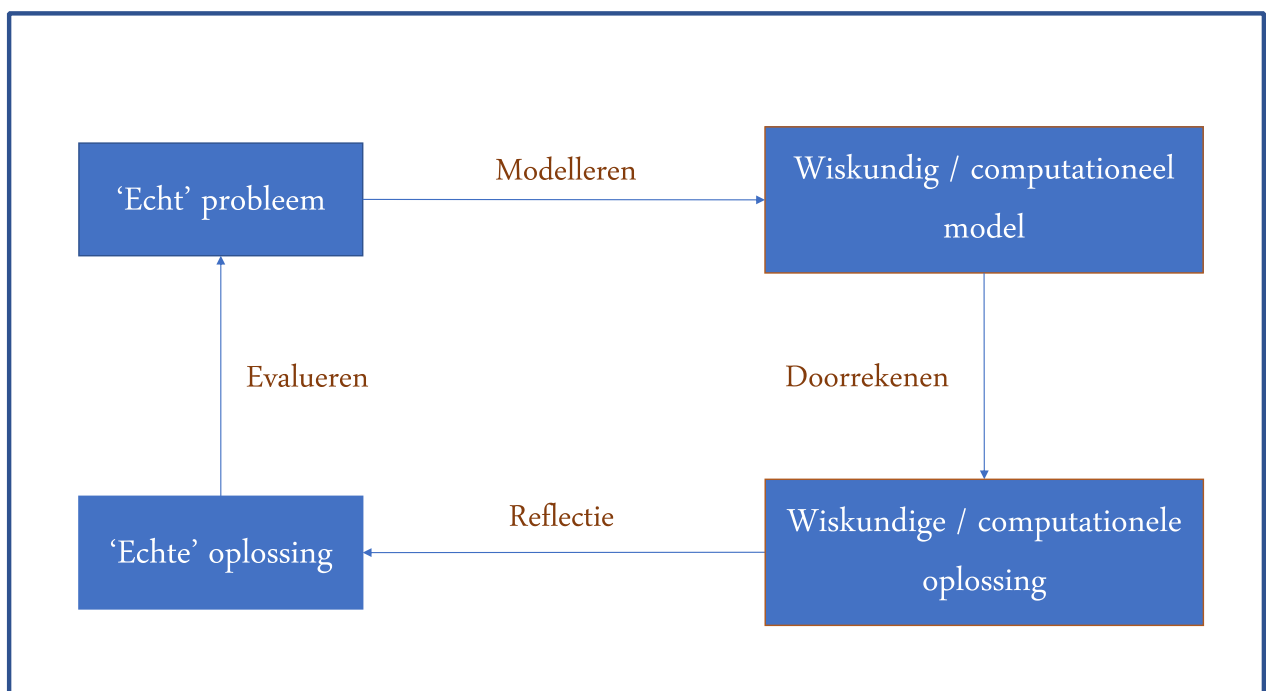
Het ontwikkelen van computationele denkvaardigheden kan helpen bij het oplossen van wiskundige problemen. Computationeel denken wordt soms samengevat als ‘denken als een computer’. Het is echter meer dan dat: het gaat vooral over het leren denken op een logische manier bij het oplossen van problemen. Denk bijvoorbeeld aan het opsplitsen van een probleem in deelproblemen, of aan het opstellen van methodes en algoritmes om (typen van) opgaven aan te pakken. Hierbij kun je de computer als hulpmiddel gebruiken. Computationeel denken kan gezien worden als een vaardigheid die helpt om complexe problemen aan te pakken.

In deze lessenserie ga je computationele denkvaardigheden oefenen in statistiekopdrachten waarbij je Excel gebruikt. In de laatste les maak je een verslag over een eigen onderzoek.

1. Probleemoplossen met computationeel denken

Een probleem oplossen kun je zien als een klein onderzoek waarbij je antwoord op een vraag probeert te vinden. Als we daarbij computationeel denken betrekken, gaat het er om het probleem op te lossen met hulp van de computer en de tools die je daarbij beschikbaar hebt. Daar is niet altijd een vast stappenplan voor. Het is zelfs niet altijd zeker dat je het antwoord zult vinden of welke tool je moet gebruiken. Je kunt wel een algemene aanpak volgen. Dit noemen we hier de *computationele onderzoekscyclus* (zie figuur 1).

Met behulp van de computationele onderzoekscyclus ga je op zoek naar wat je nodig hebt om antwoord op je vraag te krijgen. Je volgt de verschillende onderdelen, maar de weg naar het antwoord ligt niet vast. Soms moet je terug naar een vorig onderdeel, bijv. je probleem verder opdelen, om het vervolgens in stukjes te kunnen oplossen. Soms moet je je model aanpassen, omdat het geen goede oplossing geeft.

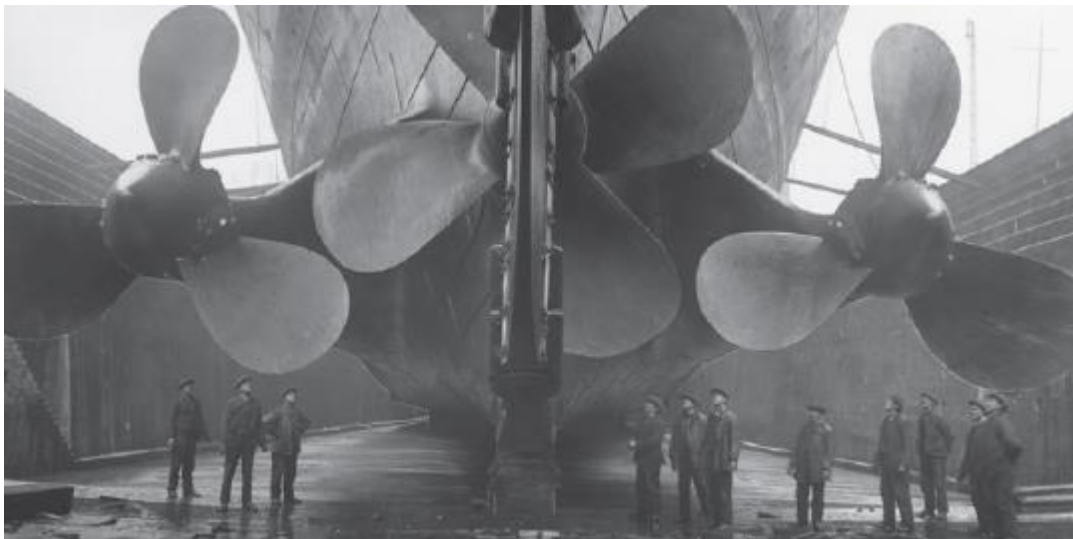


Figuur 1. De computationele onderzoekscyclus

In deze lessenserie gaan we de computationele onderzoekscyclus allereerst bekijken met een gegeven probleem bij de scheepsramp van de Titanic. Later ga je ook zelf je eigen probleem of vraag formuleren en daarbij het antwoord zoeken.

Denk als een onderzoeker!

2. Titanic



De Titanic was een luxeschip, dat een groot deel van het trans-Atlantisch passagiersverkeer moest verwerken. In de nacht van 14 op 15 april 1912 kwam de Titanic op haar eerste reis kort voor middernacht in aanvaring met een ijsberg. Een deel van de stuurboordzijde werd op verscheidene plaatsen doorboord en binnen drie uur was het schip gezonken.

Een animatie in 'real time' van het zinken van de Titanic (niet helemaal afkijken! ;-)):

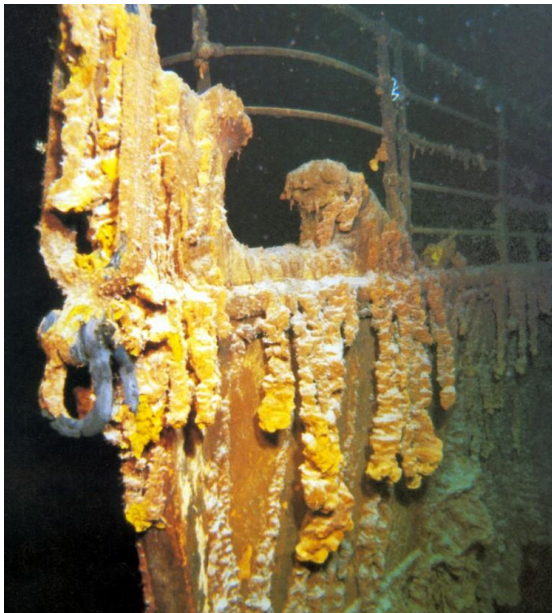
<https://www.youtube.com/watch?v=rs9w5bgtJC8>

Van de naar schatting in totaal 2223 passagiers kwamen 1522 om het leven. Hoewel dit naar aantal slachtoffers niet de grootste scheepsramp in de geschiedenis is, is het wel de bekendste, vooral omdat het schip destijds als "praktisch onzinkbaar" bekend stond.

De scheepsramp heeft veel filmmakers geïnspireerd. De verfilming van de ramp *Titanic* uit 1997 is de duurste en succesvolste films ooit, en won 11 Academy Awards. Fragment van de crash (duur 2:54):

<https://www.youtube.com/watch?v=DWvM3mOXWDO>

(bron: wikipedia)



In de volgende opdracht krijg je (een deel van) de passagierslijst van de Titanic en ga je onderzoeken hoe je aan de hand van deze gegevens vragen kunt beantwoorden over wat er is gebeurd.

2.1 Vrouwen en kinderen eerst?

'Vrouwen en kinderen eerst' is een regel uit de scheepvaart die aangeeft dat vrouwen en kinderen eerst gered moeten worden in geval van nood. De regel dateert van ongeveer 1860. We noemen de uitspraak 'vrouwen en kinderen werden eerst van boord gehaald' onze **aanname**.

VRAAG A

Is bij de ramp van de Titanic de volgende uitspraak waar?

'Vrouwen en kinderen werden eerst van boord gehaald.'

Werk in kleine groepjes om vraag A te beantwoorden. Hieronder staan vragen om je te helpen hierbij. Sla dit bestand op onder de naam van je groepje en vul de antwoorden hieronder in.

Namen groepsleden:

1 Welke informatie heb je nodig om vraag A te beantwoorden?

2. Is deze informatie beschikbaar in de dataset?

Bekijk de dataset en verken de gegevens die je hebt.

3. Welke stappen moet je doen om vraag A te beantwoorden?

Gebruik je wiskundeboek om op te zoeken welke wiskundige gereedschappen je kan gebruiken. Kijk zonodig in de Excel spiekbrieven.

4. Hoe kun je deze stappen vertalen in Excel?

En nu: trial & error of eerst nadenken...?

Beschrijf het stappenplan zo precies mogelijk, maar voer het nog niet uit in Excel.

2.2 Stappenplan uitwerken in Excel

In dit deel van de opdracht ga je het eerder gemaakte stappenplan uitwerken in Excel. Sla je werk in Excel op in een bestand met je eigen naam.

We hadden de volgende vraag:

VRAAG A

Is bij de ramp van de Titanic de volgende uitspraak waar?

'Vrouwen en kinderen werden eerst van boord gehaald.'

5. Voer je stappenplan uit in Excel en vind een antwoord op vraag A.

Was het in een keer goed? Of moest je je stappen herzien? Dit hoort bij onderzoek doen!

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?

7. Herken je stappen die je gemaakt hebt in de cyclus?

Zou je deze stappen ook voor een ander probleem kunnen gebruiken? Helpt de computationele onderzoekscyclus je om het grotere geheel te zien i.p.v. de losse stappen?

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.

Kijk goed of je een vraag opgesteld hebt die je kunt beantwoorden. Mogelijk moet je de vraag preciezer maken of opdelen in kleinere deelvragen.

Bedenk welke wiskunde je nodig hebt om je vraag te beantwoorden. Ga je gemiddeldes vergelijken of percentages? Hoe test je het verschil? Gebruik je odds-ratio (voor uitleg zie <https://www.youtube.com/watch?v=yLt0kUmo4i8>) of regressie (zie kader hieronder)?

Welke technieken in Excel ga je gebruiken om je vraag te beantwoorden? Ga je een frequentietabel maken of een spreidingsdiagram? Hoe kun je de samenhang tussen twee variabelen onderzoeken in Excel?

Mogelijk onderzoek je de factor leeftijd. Vaak is het interessant om te onderzoeken of een populatie voor een kenmerk normaal verdeeld is. Bij leeftijd is dat meestal niet zo. Kijk in je wiskundeboek naar de eigenschappen van een normale verdeling en onderzoek of de populatie aan boord voor wat betreft leeftijd normaal verdeeld is.

Informatie over regressie

Informatie over regressie en hoe je dit in Excel onderzoekt kun je vinden in het document **AppendixLineaireRegressie.docx**. Een voorbeeld in Excel vind je in **VoorbeeldRegressieData.xlsm**.

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



Foto iStock

Leerlingen hebben steeds minder interesse in schoolvakken, zo was laatst in het nieuws:

<https://nos.nl/artikel/2312927-steeds-minder-interesse-in-schoolvakken-cijfergemiddelde-daalt.html> (NOS nieuws, 2 december 2019)

Naast een gebrek aan interesse is er ook een dalende trend in de cijfers vastgesteld, met name bij jongens. En volgens recent internationaal PISA-onderzoek lezen Nederlandse leerlingen steeds minder goed:

<https://www.nro.nl/pisa-2018-uitkomsten/> (Persbericht NRO, 3 december 2019)

Deze trend is van belang, want er wordt gezegd dat uiteindelijk alle schoolvakken profiteren van leerlingen die beter kunnen lezen, en dat je prestaties bij alle vakken kunnen lijden onder gebrekkige leesvaardigheid.

<https://www.knack.be/nieuws/belgie/uiteindelijk-profiteren-alle-schoolvakken-van-leerlingen-die-beter-kunnen-lezen/article-opinion-1540883.html>

Er lijkt dus een samenhang te zijn tussen de cijfers voor Nederlands en de andere schoolvakken.

3.1 Kies een dataset

Hoe zit het eigenlijk met de cijfers op een doorsnee middelbare school? En wat kun je zeggen over schoolverlaters? En speelt vrije tijdsbesteding een rol bij de cijfers?

We hebben drie sets bestanden die je kunt onderzoeken:

1. Gegevens van een HAVO-VWO school in Zuid-Holland met de cijfers voor de verschillende vakken van de schooljaren 2017/2018 ("Leerlingen_17-18_anoniem.xlsx") en 2018/2019 ("Leerlingen_18-19_anoniem.xlsx"). De beschrijving van de gegevens vind je in "beschrijving dataset cijfers.xlsx".
2. Gegevens over schoolverlaters ("Dataset Schoolverlaters.xlsx") met beschrijving in "Beschrijving Schoolverlaters.docx".
3. Gegevens over schoolcijfers en vrije tijdsbesteding ("Dataset COOL.xlsx") met beschrijving in "Beschrijving COOL.docx".

KLASSENDISCUSSIE

Voor elk van de datasets:

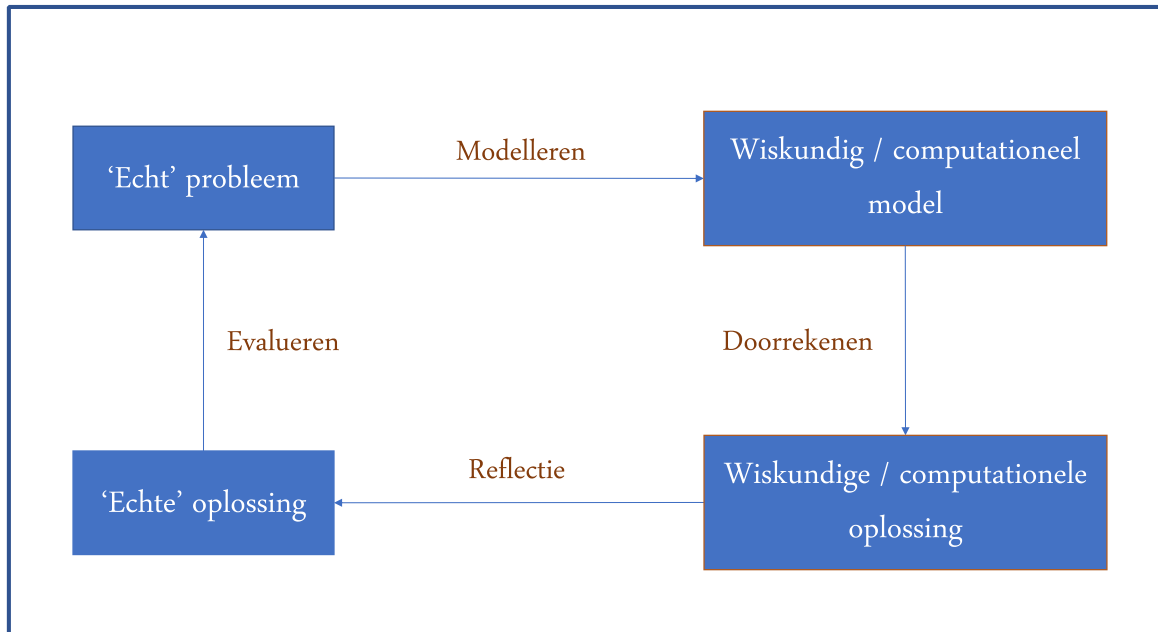
Bekijk de omschrijving van de dataset.

Welke gegevens zijn beschikbaar?

Welke vragen komen er bij je op?

Welke vragen zou je willen onderzoeken?

Werk in kleine groepjes om één of meerdere vragen te onderzoeken bij de gekozen dataset. Doorloop de stappen van de computationele onderzoekscyclus die je eerder hebt geleerd. Dit keer begin je bij het begin: het formuleren van het 'probleem'. Bedenk dat je soms mogelijk weer terug moet naar een vorige stap, omdat bijv. je probleem niet precies genoeg is of omdat je toch een andere berekening nodig hebt.



VRAAG A

Welke vragen zou je willen onderzoeken?

Werk je antwoord uit.

Volg de computationele onderzoekscyclus die je eerder hebt geleerd in de Titanic-opdracht, d.w.z. deel je vraag of probleem zonedig op in behapbare blokken zodat je vragen hebt die je kunt beantwoorden, bepaal welke wiskunde je hierbij nodig hebt, bepaal hoe je dat in Excel kunt uitvoeren en bepaal hoe je de resultaten gaat interpreteren. Bedenk hoe je de gegevens kunt groeperen, ombouwen of filteren om vergelijkingen tussen de bestanden te kunnen maken of antwoord te kunnen geven op de door jou gewenste vragen. Specificeer de stappen van je aanpak zo precies mogelijk, **voordat** je ze in Excel uit gaat voeren.

Deze opdracht wordt afgesloten met een **verslag**. Zorg dus dat je de stappen die je neemt goed beschrijft en vast legt in een document.

Hieronder staan vragen om je te helpen hierbij.

1. Welke informatie heb je nodig?

2. Is deze informatie beschikbaar in de dataset?

Bekijk de dataset en verken de gegevens die je hebt.

3. Hoe krijg ik deze informatie beschikbaar in de dataset?

Maak een plan hoe je de gegevens die je hebt kunt groeperen, ombouwen of filteren om de informatie die je nodig hebt te kunnen verkrijgen..

4. Welke stappen moet je doen om je vraag te toetsen?

Gebruik je wiskundeboek om op te zoeken welke wiskundige gereedschappen je kan gebruiken. Denk bijv. aan odds-ratio, normale verdeling of regressie.

5. Hoe kun je deze stappen vertalen in Excel?

Beschrijf het stappenplan zo precies mogelijk, maar voer het nog niet uit in Excel.

3.2 Stappenplan uitwerken in Excel

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?

KLASSENDISCUSSIE

Was je vraag makkelijk te beantwoorden of moeilijk?

Wat was er nodig om de vraag te kunnen beantwoorden?

Moest er nog data worden bewerkt?

Voldeed het aan je tijdsverwachtingen?

3.3 Verslag

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

Maak een document en sla het op onder je eigen naam.

In het verslag beschrijf je wat je hebt gedaan en waarom. Beschrijf hoe je je vraag hebt onderzocht en wat je conclusie is. Een manier om het verslag te structureren is door de stappen te beschrijven van de computationele onderzoekscyclus die je hebt gevolgd, inclusief eventuele onsuccesvolle wegen.