

Spreadingsdiagram en trendlijn in de 2^{de} graad doorstroomfinaliteit

Dag van de wiskunde
Kortrijk, 27 november 2021

Kennismaking

Kennismaking – wie zijn jullie?

- educatieve bachelor?
 - met onderwijsvak wiskunde?
- master?
 - welke master?
- geef je les in de tweede graad?
 - derde jaar?
 - vierde jaar?
- in welk jaar plan je het onderwerp spreidingsdiagrammen en trendlijnen?
 - derde jaar?
 - vierde jaar?
 - beide?

Kennismaking – wie ben ik?

- van opleiding wiskundige
- verleden (25 jaar) als docent wiskunde in economisch hoger onderwijs op universitair niveau
- sinds 1996 betrokken bij universitaire lerarenopleiding wiskunde (praktijklector, docent UAntwerpen, docent KU Leuven)
- onderzoeker vakdidactiek wiskunde
- redactielid Uitwiskeling
- meegewerkt aan de ontwikkeling van (de vorige en) huidige eindtermen
- meegewerkt aan peilingsonderzoeken

Materiaal voor deze sessie

- eerder (denk- en ontwikkel)werk voor Uitwisseling
 - JD, Els Vanlommel (2020). Eerstegraadsfuncties. UW 36/1
 - JD, Regi Op de Beeck, Luc Van den Broeck (2011). Leren modelleren. UW 27/3
 - JD, Jan Roels, Pedro Tytgat (2005). Lineaire regressie. UW 21/2
 - zie website <https://perswww.kuleuven.be/~u0010098/>
(of website KU Leuven > Interne tools > wie-is-wie > ... > website)
- aangevuld voor deze sessie → zie slides
- nieuw artikel in volgende nummer van Uitwisseling, specifiek over spreidingsdiagrammen en trendlijnen in de 2de graad
- er is ook mooi materiaal van anderen dat niet in deze sessie verwerkt is: zie bv. <https://dwengo.org/> (zie 'data' bij spreidingsdiagram, lineaire regressie, klimaat, ...)

De eindterm over spreidingsdiagram en trendlijn

De eindterm

basiseindterm in de doorstroomfinaliteit

De leerlingen onderzoeken het verband tussen twee numerieke grootheden in een dataset met behulp van een spreidingsdiagram.

- Met inbegrip van kennis
 - Feitenkennis
 - Conceptuele kennis
 - Procedurele kennis
- Met inbegrip van context
- Met inbegrip van dimensies eindterm

De eindterm

De leerlingen onderzoeken het verband tussen twee numerieke grootheden in een dataset met behulp van een spreidingsdiagram.

- Feitenkennis

Spreidingsdiagram / Voorschrift en grafiek van ... / Recht evenredig verband, omgekeerd evenredig verband, lineair verband, kwadratisch verband

- Conceptuele kennis

- Spreidingsdiagram

- Informeel begrip van trendlijn

- Informeel begrip van de correlatiecoëfficiënt bij een lineair verband

- Voorschrift en grafiek van

- Eerstegraadsfuncties van de vorm $f(x) = ax + b$ met $a \in \mathbb{R}_0$ en $b \in \mathbb{R}$

- Kwadratische functies van de vorm $f(x) = ax^2$ met $a \in \mathbb{R}_0$

- Functies van de vorm $f(x) = c/x$ met $c \in \mathbb{R}_0$

- Verbanden tussen grootheden: recht evenredig, lineair, omgekeerd evenredig, kwadratisch

- Vaak voorkomende fouten, misconcepties, tekortkomingen en manipulaties bij het grafisch voorstellen, het numeriek samenvatten en het interpreteren van statistische informatie: samenhang versus causaliteit

De eindterm

De leerlingen onderzoeken het verband tussen twee numerieke grootheden in een dataset met behulp van een spreidingsdiagram.

- Met inbegrip van kennis
 - Procedurele kennis: met ICT
 - Opstellen en interpreteren van een spreidingsdiagram
 - Bepalen en interpreteren van de trendlijn met bijhorend voorschrift
 - Bepalen en interpreteren van de correlatiecoëfficiënt bij een lineair verband
- Met inbegrip van context

De eindterm wordt met context gerealiseerd.
- Met inbegrip van dimensies eindterm

Cognitieve dimensie: beheersingsniveau analyseren

De toelichting

- past in de leerlijn statistiek
 - eerste graad: kleinschalig statistisch onderzoek uitvoeren
 - tweede graad
 - beschrijvende statistiek van één variabele
 - nu: beschrijvende statistiek van twee variabelen
 - derde graad: beschrijvende statistiek van twee variabelen met kruistabellen
 - statistische berekeningen en grafische voorstellingen kaderen in het beantwoorden van een zinvolle vraag
 - aandacht voor misconcepties, fouten en manipulaties
- sluit ook aan bij de bouwsteen 'relaties en verandering'
- link met andere STEM-domeinen, maar ook met economie en humane wetenschappen

Stap voor stap door de leerinhoud

Spreidingsdiagram

Spreadingsdiagram: wat is het?

- twee variabelen
- algemeen: x en y
- hier: prijs (y) en grootte (x) van harde schijven
- gegeven bepaalt punt in het vlak
- Engels: scatterplot

Let op voor naamsverwarring

- bij één veranderlijke ...

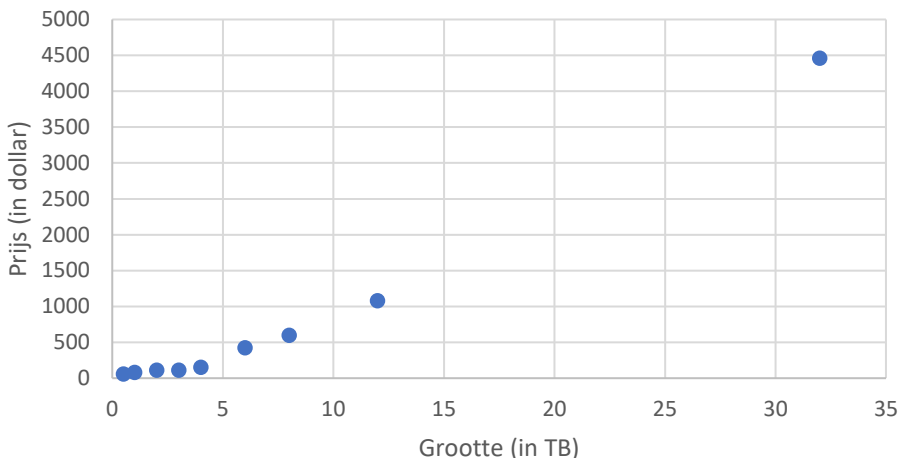
- ... begrip 'spreiding'

variatiebreedte,
standaardafwijking,
interkwartielafstand

- ... sommigen noemen histogram of staafdiagram onterecht spreadingsdiagram

'diagram dat toont hoe de gegevens gespreid zijn'

Prijs en grootte van harde schijven (0,5-32 TB)

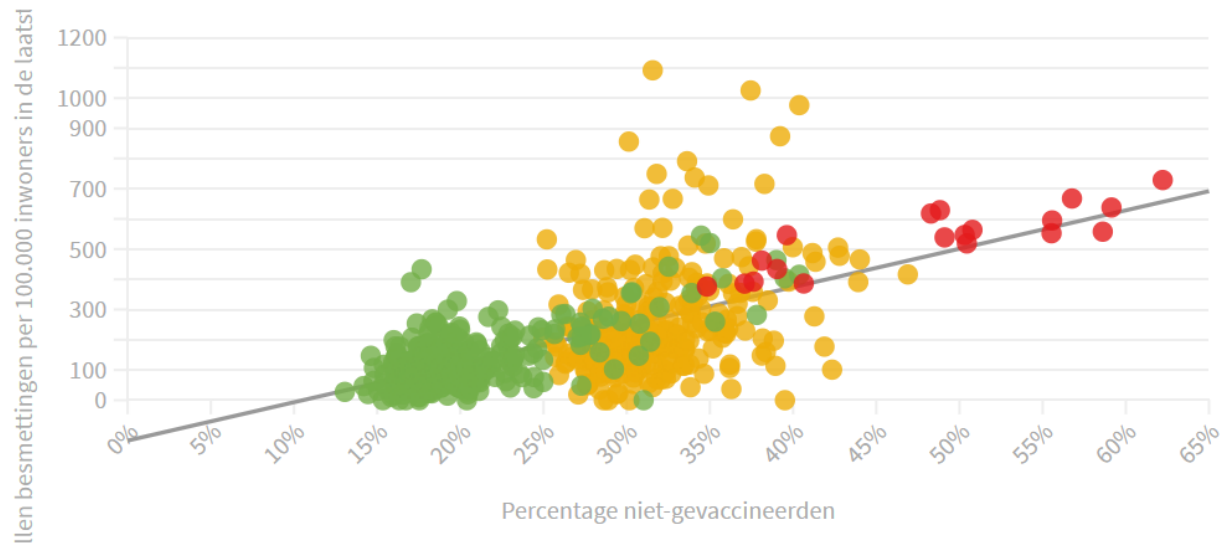


Spreadingsdiagram: in de media

Zijn er echt veel besmettingen waar er veel niet-gevaccineerden zijn?

Aantal besmettingen per 100.000 inwoners afgelopen 14 dagen in verhouding tot het percentage niet-gevaccineerden

● Vlaanderen ● Brussel ● Wallonië

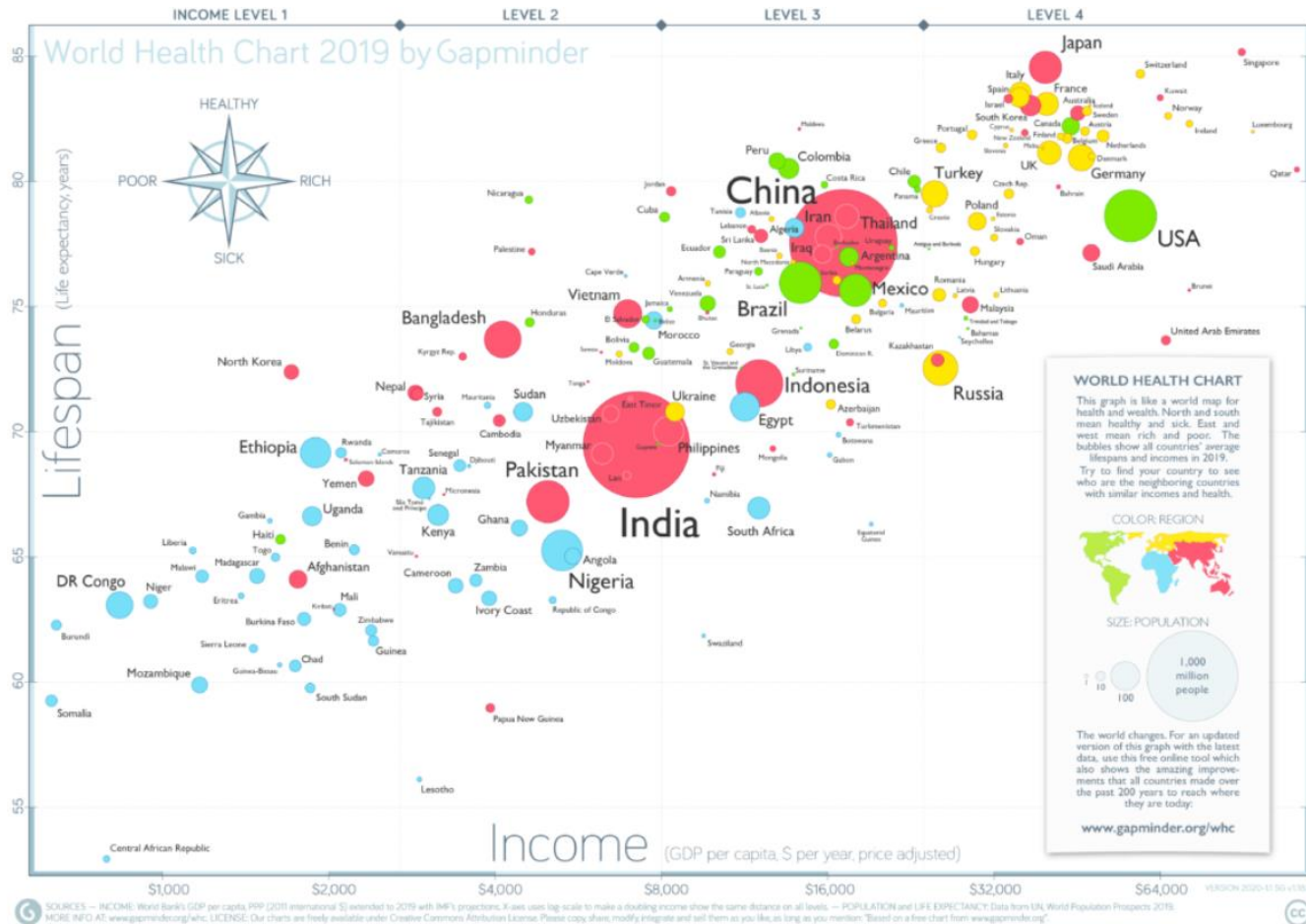


DS-infografiek | De Standaard op basis van gegevens Sciensano en Statbel

✳ A Flourish scatter chart

Bron: <https://www.standaard.be/krant/publicatie/20210922/ds/dn/alg/optimized>

Spreadingsdiagrammen in andere schoolvakken: wetenschappen, economie



Bron: [https://www.gapminder.org/tools/#\\$chart-type=bubbles&url=v1](https://www.gapminder.org/tools/#$chart-type=bubbles&url=v1)

Spreadingsdiagram

- tweedimensionale data leren voorstellen
- leer je leerlingen zo'n grafiek beschrijven
 - variabele die op de horizontale/verticale as staat, gebruikte eenheden, ...
 - welke patronen kun je eventueel herkennen?
- toon gebruik in media en in andere schoolvakken (niet alleen STEM-vakken!)
- toon diverse voorbeelden, bv. al dan niet 'tijd' als variabele op de horizontale as
- leg verbanden met andere delen van de wiskunde
 - beschrijven van punten in het vlak met coördinaten
 - (bij trendlijn) verband met functies

Rechte trendlijnen

Van 'exacte wiskunde' naar 'wiskundig model'

Startpunt: eerstegraadsfunctie met context: graden Celsius en Fahrenheit

- verband tussen temperatuur in graden Celsius en graden Fahrenheit: $T_F = 32 + 1,8 \cdot T_C$
- toepassing op eerstegraadsfuncties met veel mogelijke opdrachten voor de leerlingen
 - ...
- deze vergelijking beschrijft een ‘exact verband’
bv. ‘twee punten bepalen juist één rechte’ blijft van toepassing: ‘stel de vergelijking op als je weet dat water bevriest bij 32°F en kookt bij 212°F’
- in veel situaties in de realiteit geldt niet zo’n ‘exact verband’

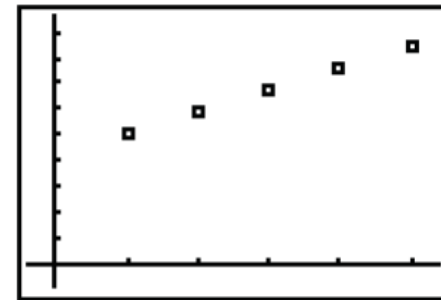
Voorbeeld: flyers drukken – eerst exact ...

aantal exemplaren	1000	2000	3000	4000	5000
prijs (in EUR)	125	146	167	188	209

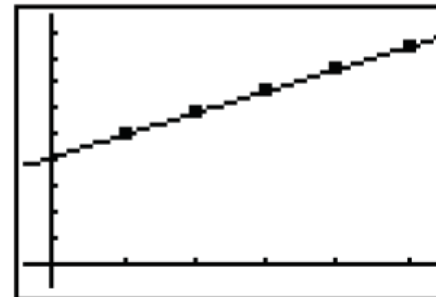
L1	L2	L3	Z
1	125	-----	
2	146		
3	167		
4	188		
5	209		

L2(6) =			

```
WINDOW
Xmin=-.4
Xmax=5.4
Xscl=1
Ymin=-20
Ymax=240
Yscl=25
↓Xres=■
```



```
Plot2 Plot3
Y1=21*X+104
Y2=
Y3=
Y4=
Y5=
Y6=
Y7=
```

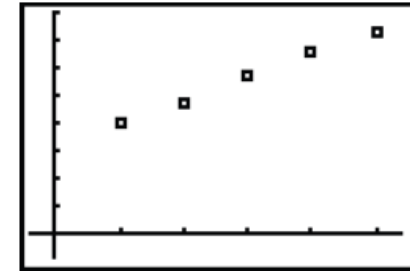


x in duizendtallen

bron: UW 27/3, p. 18

Voorbeeld: flyers drukken – ... vervolgens benaderend

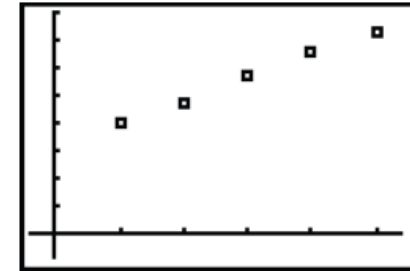
aantal exemplaren	1000	2000	3000	4000	5000
prijs (in EUR)	99	120	142	163	183



- niet meer exact volgens een eerstegraadsfunctie...
- maar wel nog bij benadering
- bv. rechte door eerste en laatste punt geeft $p = 21x + 78$
- drie keer juiste prijs, twee keer 1 EUR te laag
- we kunnen een eenvoudige formule gebruiken als we tevreden zijn met benaderende waarden
- de eerstegraadsfunctie is een wiskundig model
- de rechte is een trendlijn

Voorbeeld: flyers drukken – ... vervolgens benaderend

aantal exemplaren	1000	2000	3000	4000	5000
prijs (in EUR)	99	120	142	163	183



- rechte door eerste en laatste punt geeft $p = 21x + 78$
drie keer juiste prijs, twee keer 1 EUR te laag
- andere mogelijke rechten, bv. $p = 21x + 78,5$
maximale afwijking van 0,5 EUR
- laat leerlingen wennen aan
 - andere afspraken bij het gebruik van een wiskundig model dan bij 'exacte wiskunde'
 - meer dan één model mogelijk
- laat leerlingen enkele keren de trendlijn 'met gezond verstand' bepalen

Trendlijn met ICT bepalen

Trendlijn met behulp van ICT: waarom?

- trendlijn met ‘gezond verstand bepalen’
 - veel werk
 - iedereen vindt iets anders
- eindterm: trendlijn met behulp van ICT bepalen

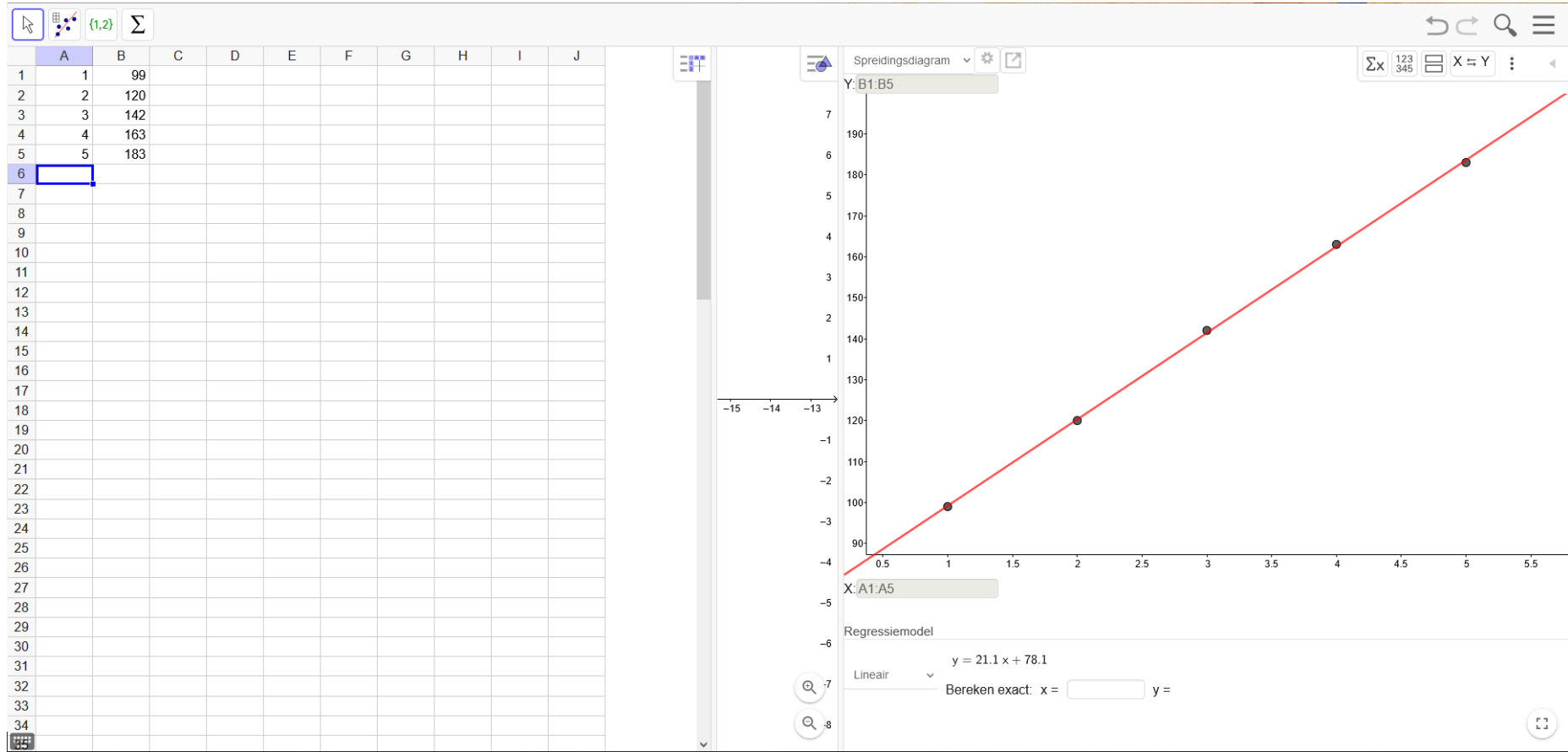
Trendlijn met behulp van TI-84

```
EDIT [2nd][MODE] TESTS
1:1-Var Stats
2:2-Var Stats
3:Med-Med
4:LinReg(ax+b)
5:QuadReg
6:CubicReg
7↓QuartReg
```

```
LinReg(ax+b) L1,
L2,Y1
```

```
LinReg
y=ax+b
a=21.1
b=78.1
```

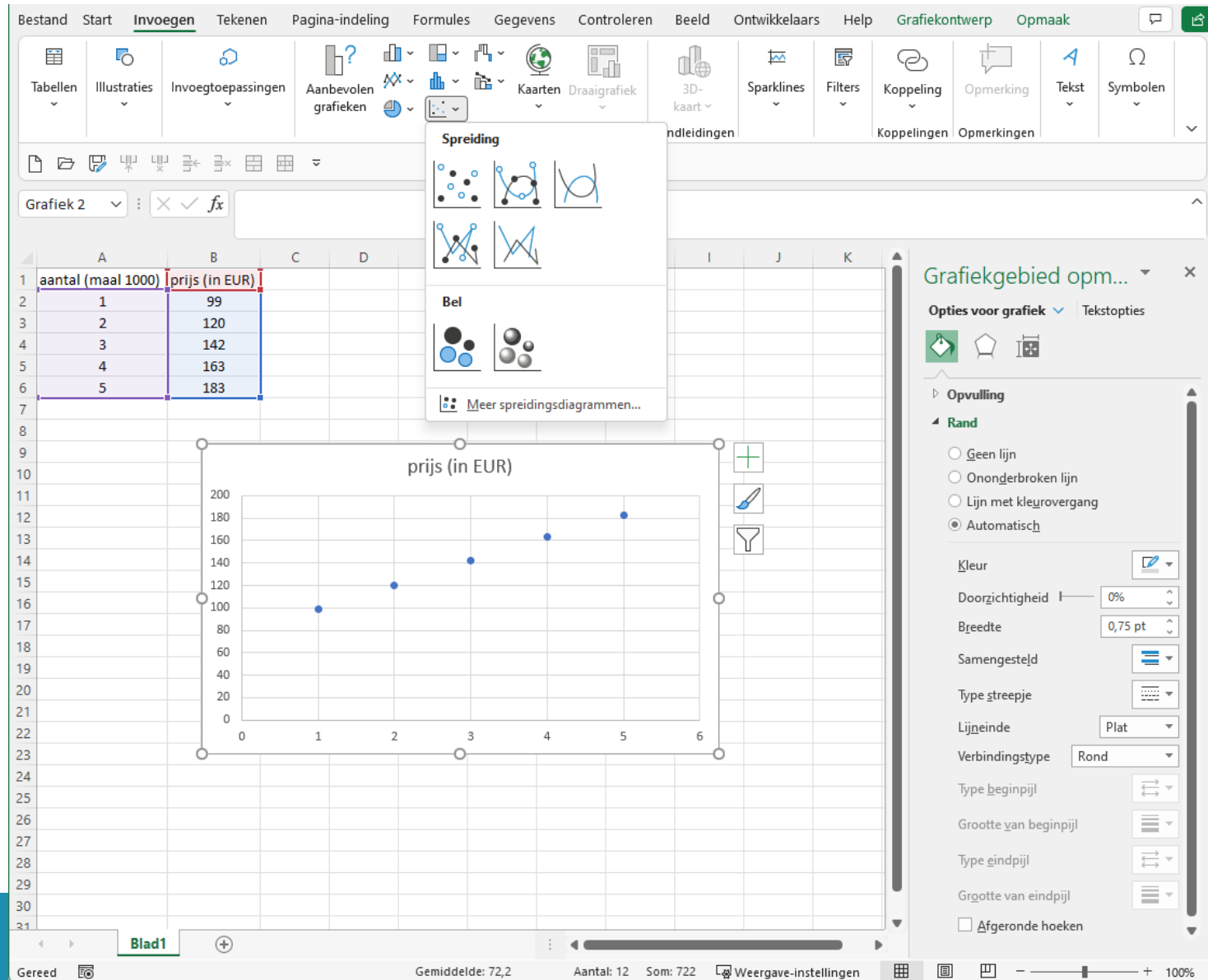
Trendlijn met behulp van GeoGebra



hier: GeoGebra Klassiek > Rekenblad > ... > Regressieanalyse > Regressiemodel: Lineair

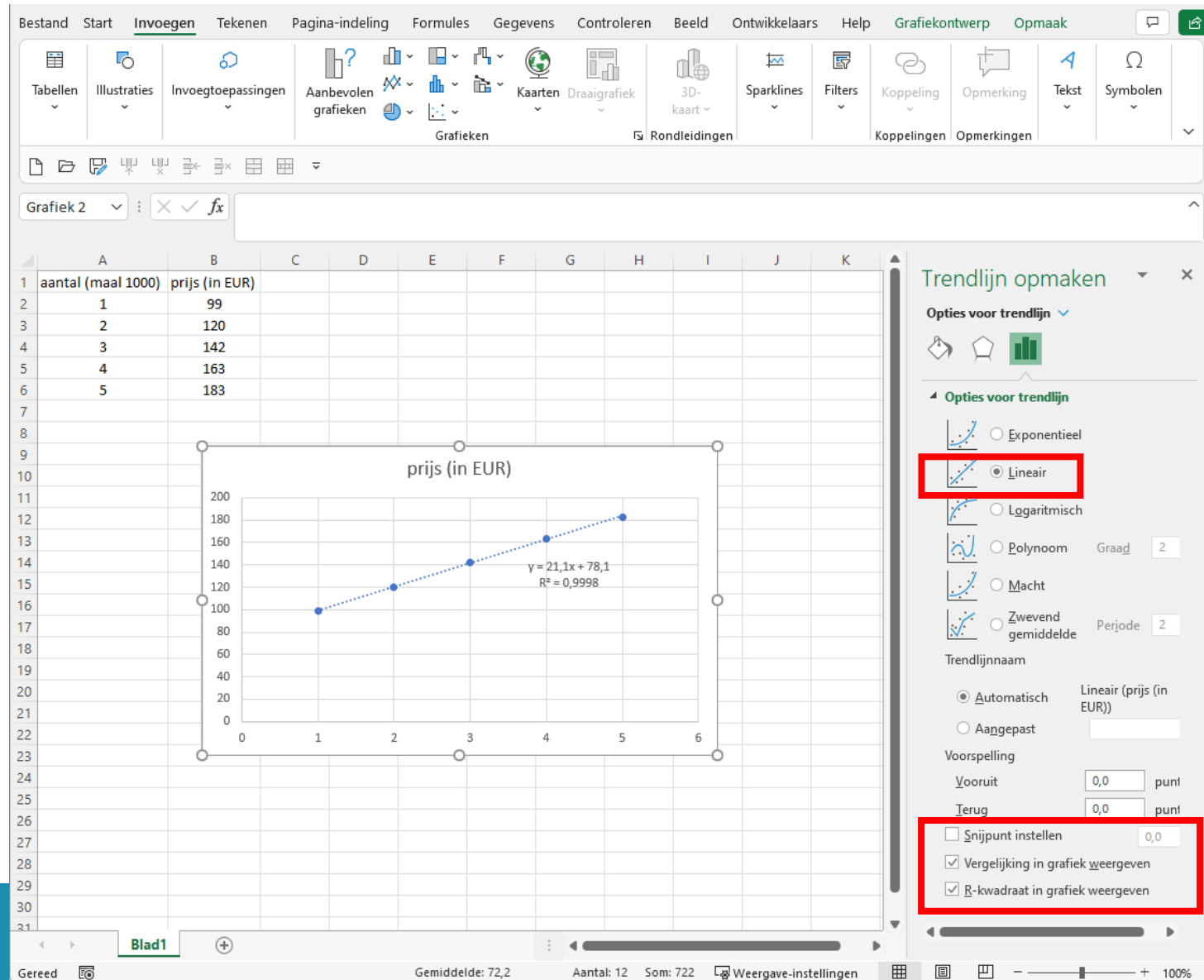
Trendlijn met behulp van Excel

Invoegen > ...



Trendlijn met behulp van Excel

- selecteer punten
- rechter-muisknop
- > Trendlijn toevoegen
- ...
- duid gepaste rondjes en vakjes aan in venster 'Trendlijn opmaken'



Trendlijn met ICT bepalen

- al deze ICT-tools geven dezelfde trendlijn
- verschillende benamingen: lineaire regressie, trendlijn, kleinste kwadraten rechte

Trendlijn als 'black box'

- niet overdrijven met 'black box' in wiskundeles!
- maar hier is het verantwoord
 - het kan moeilijk anders...
 - ... maar toch nuttig om reeds in het secundair onderwijs trendlijnen te bestuderen in de wiskundeles
- de eindterm vraagt niet dat je de wiskundige achtergrond aan de leerlingen uitlegt
- UW 36/1: mogelijkheden om een tip van de sluier te lichten
- eventueel liggen hier mogelijkheden om de brug te maken met computationeel denken
- concentreren op 'wat is de betekenis van deze rechte?', 'wat kun je er mee doen?', 'is het gerechtvaardigd om het verband met een rechte weer te geven? ...

De rechte is niet het eindpunt

Voorbeeld: tijden winnaars 100 meter sprint op Olympische Spelen

- y : tijd in seconden
- x : tijd in jaar sinds 1948
- trendlijn $y = -0,0129x + 11,486$
- betekenis van de coëfficiënten

- rico

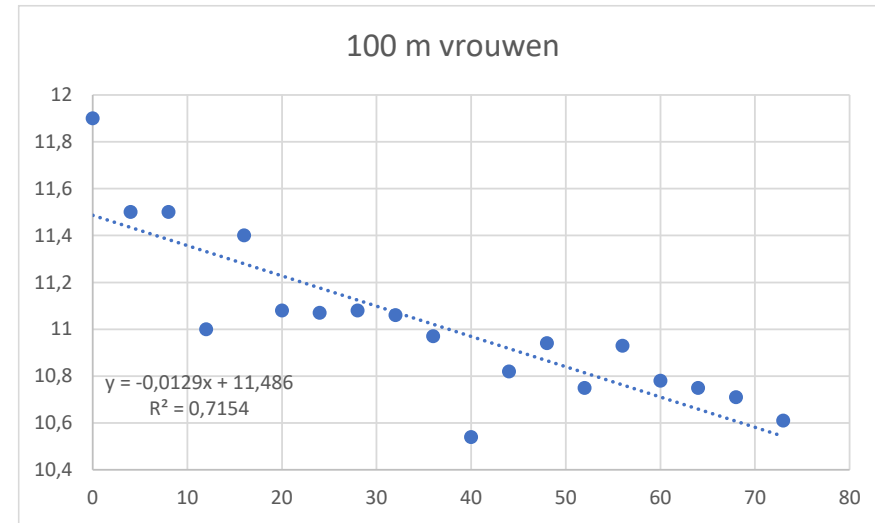


- constante term is niet gelijk aan de echte tijd in 1948!

- voorspelling tijd winnaar 2024?



- wanneer onder de 10 seconden?

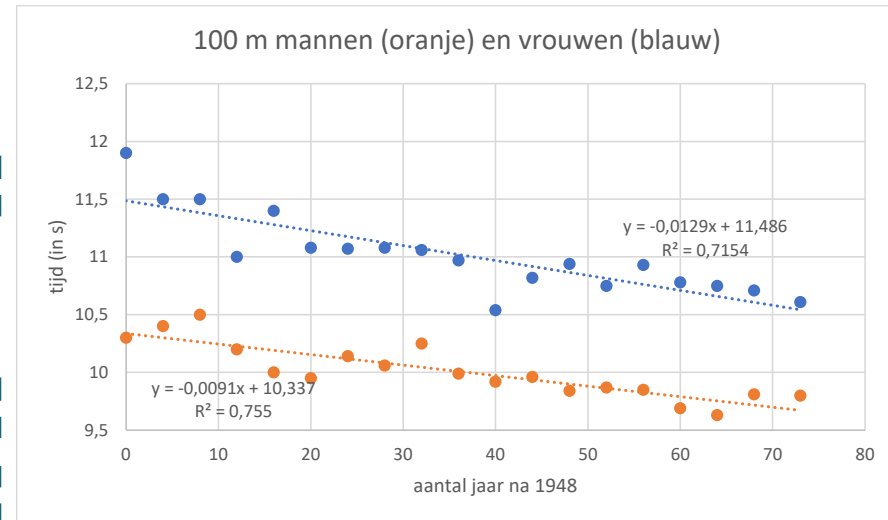


Bron: UW 27/3

Voorbeeld: tijden winnaars 100 meter sprint op Olympische Spelen

- vrouwen: $y = -0,0129x + 11,486$
- mannen: $y = -0,0091x + 10,337$
- wanneer lopen de vrouwen sneller dan de mannen?

- wanneer lopen de vrouwen de 100 meter in 0 s?



De trendlijn is niet het eindpunt

... doe er ook iets mee!

- interpreteer coëfficiënten
- bereken y -waarden door in te vullen
(interpolatie en – met mate – voorspelling/extrapolatie)
- bereken x -waarden via vergelijkingen en ongelijkheden
- ook voorbeelden en oefeningen met twee trendlijnen

Nadenken over relatie tussen model en realiteit

Nadenken over relatie tussen model en realiteit

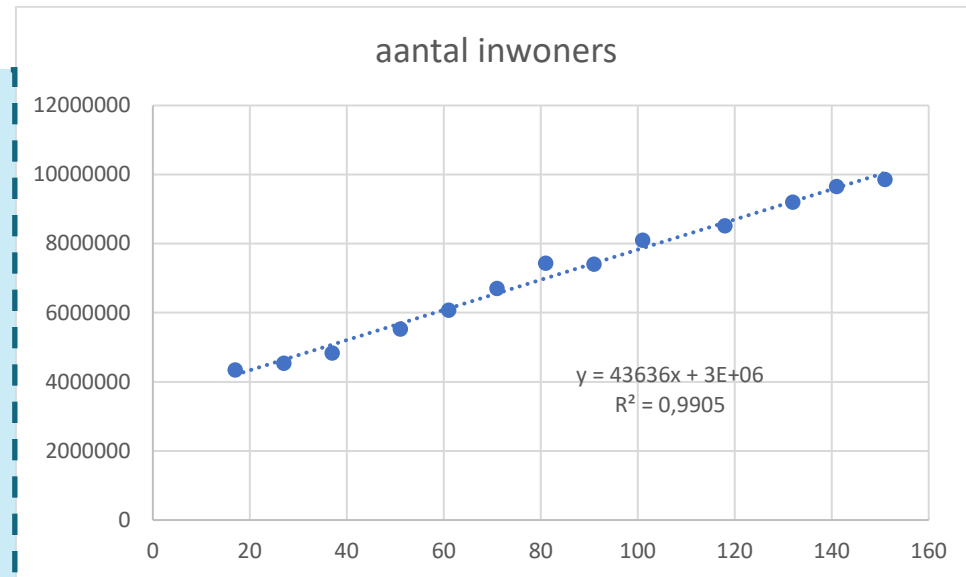
- wiskunde als model gebruiken veronderstelt dat je nadenkt over de relatie tussen model en realiteit
- leer dit ook aan je leerlingen en laat hen dat ook oefenen!
- via goed gekozen voorbeelden en oefeningen
- via een aangepaste didactiek
 - leerlingen die in tweetallen hun keuze aan elkaar verantwoorden
 - klasdiscussie
 - ...
- overdrijf niet
 - dit is nieuw voor leerlingen
 - dit is niet altijd gemakkelijk voor leerlingen
 - beperk je bv. tot enkele thema's

Nadenken over relatie tussen model en realiteit: voorbeelden van thema's

- overdreven extrapolatie: zie hoger
- is een rechte een goed model?
zie ook: correlatiecoëfficiënt
- een patroon in de afwijkingen
- geeft de context een verklaring voor de afwijkingen?
- een zwak maar zinvol verband: zie bij correlatiecoëfficiënt en in UW 36/1
- een invloedrijk punt
- niet één maar meerdere trendlijnen: zie UW 36/1
- ...

Voorbeeld: volkstellingen in België (context verklaart afwijkingen)

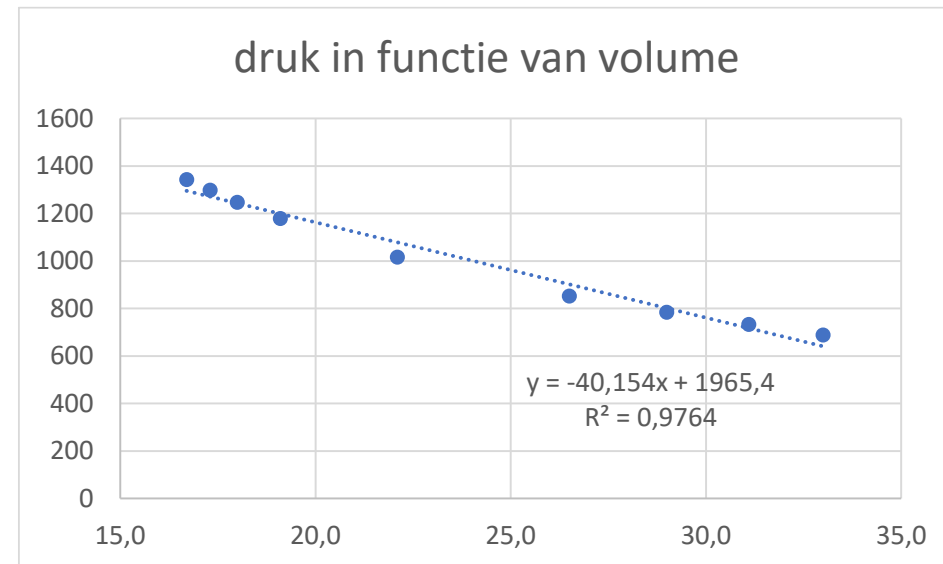
- x = tijd in jaar sinds 1830
- y = aantal inwoners



Bron: UW 27/3

Voorbeeld: druk en volume gas bij constante temperatuur (patroon in afwijkingen)

- leerlingenproef
- links en rechts: punten liggen boven de trendlijn
- midden: punten liggen onder de trendlijn
- hier past een gebogen trendlijn beter
- zie verder!

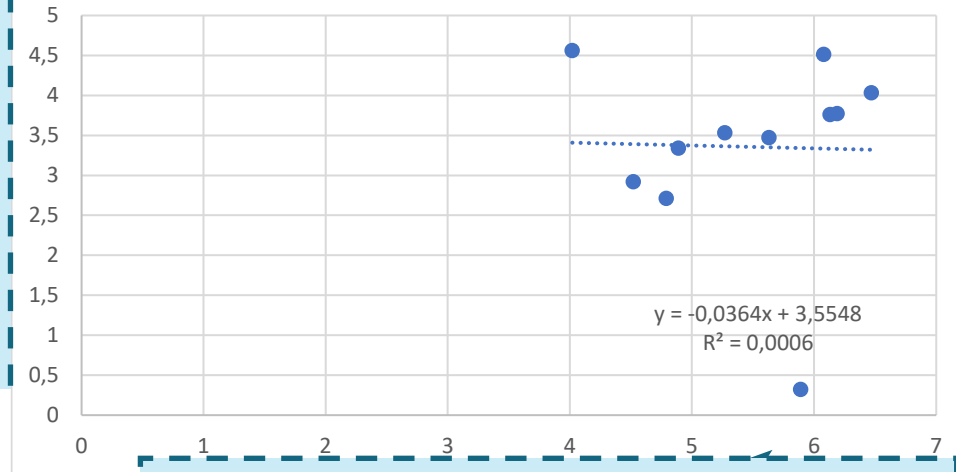


Bron: UW 21/2

Voorbeeld: uitgaven alcohol en tabak in Groot-Brittannië (invloedrijk punt)

- x = uitgaven alcohol (in pond)
- y = uitgaven tabak (in pond)

Uitgaven voor alcohol (horizontaal) en tabak (verticaal)



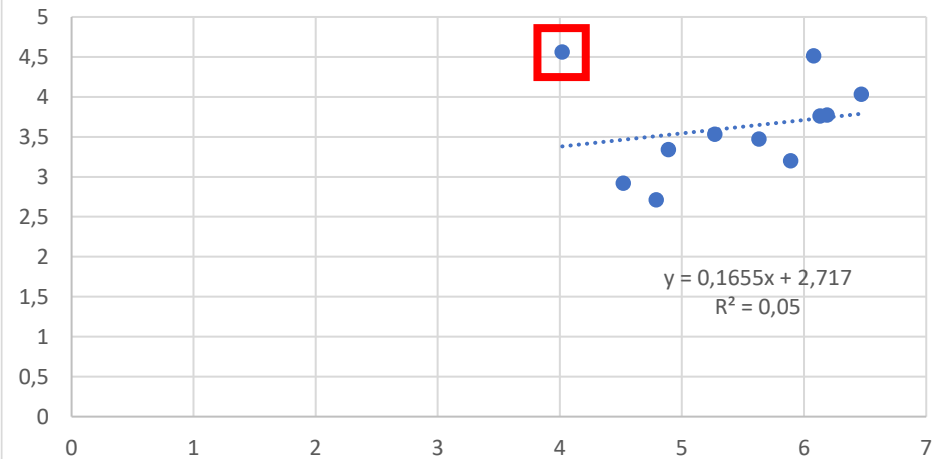
Bron:

<https://dasl.datadescription.com/datafile/tobacco-and-alcohol/>

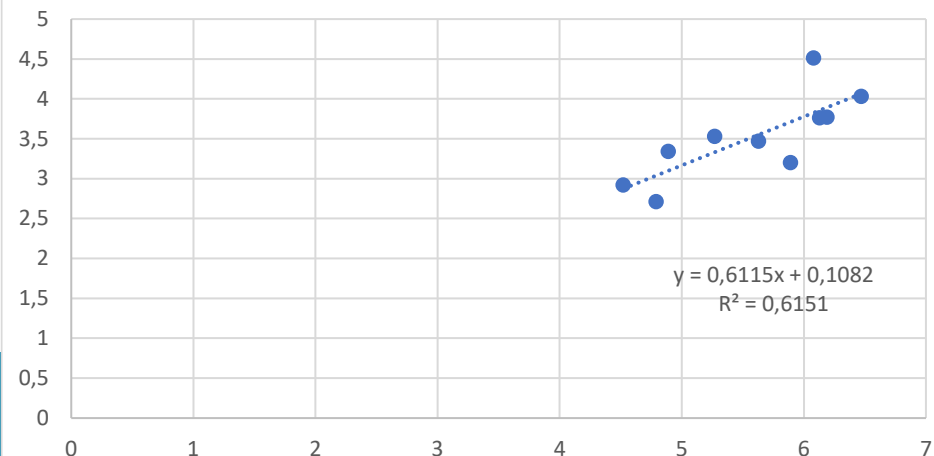
Voorbeeld: uitgaven voor alcohol en tabak... (invloedrijk punt)

- ook nu nog: trendlijn niet OK
- punt linksboven wijkt af van het patroon
- punt linksboven weglaten: trendlijn verandert van
 - van $y = 0,1655x + 2,717$
 - naar $y = 0,6115x + 0,1082$
- weglaten van dit punt geeft een sterke verandering in de rico van de trendlijn
- invloedrijk punt

Uitgaven voor alcohol (horizontaal) en tabak (verticaal)



Uitgaven voor alcohol (horizontaal) en tabak (verticaal), zonder Northern Ireland

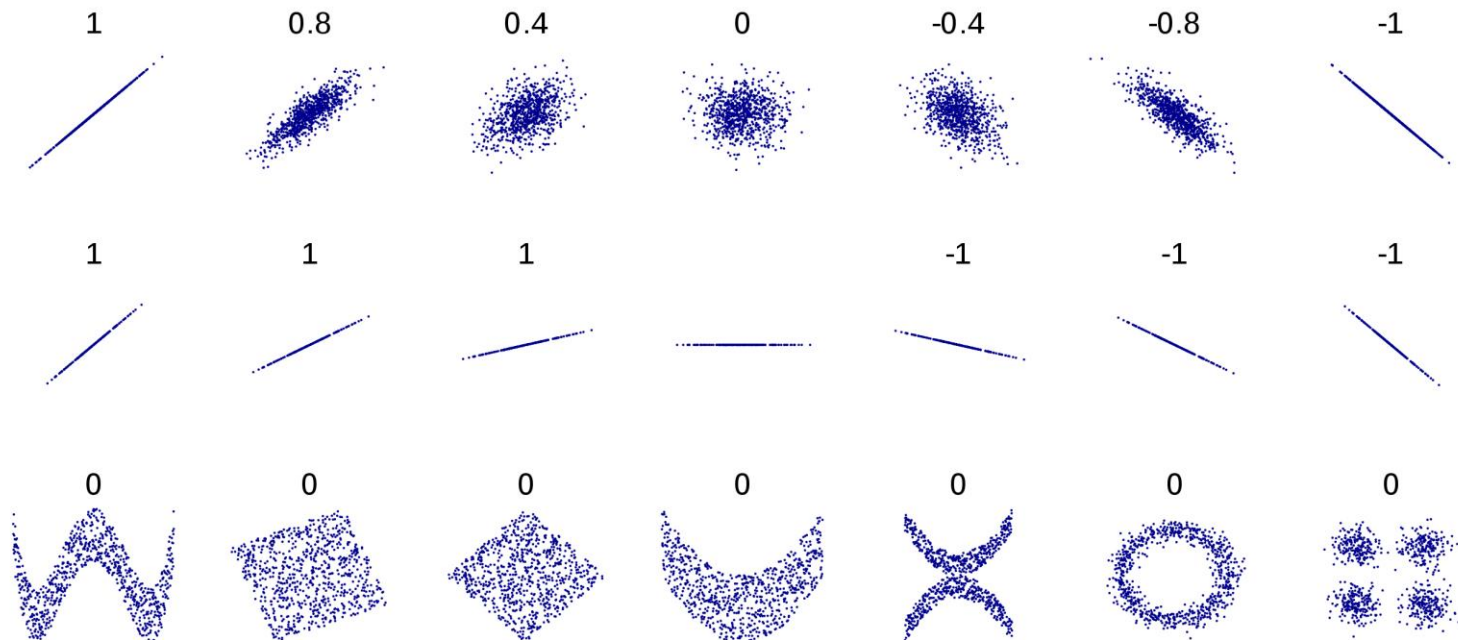


Correlatie en causaliteit

Correlatiecoëfficiënt

Correlatiecoëfficiënt

- getal tussen -1 en 1
- geeft sterkte van lineaire samenhang tussen 2 variabelen
- dus: geeft weer hoe sterk punten bij rechte aansluiten
- gelijk aan 1 : perfecte aansluiting bij stijgende rechte, ongeacht de rico
- gelijk aan -1 : perfecte aansluiting bij dalende rechte, ongeacht de rico
- https://commons.wikimedia.org/wiki/File:Correlation_examples2.svg

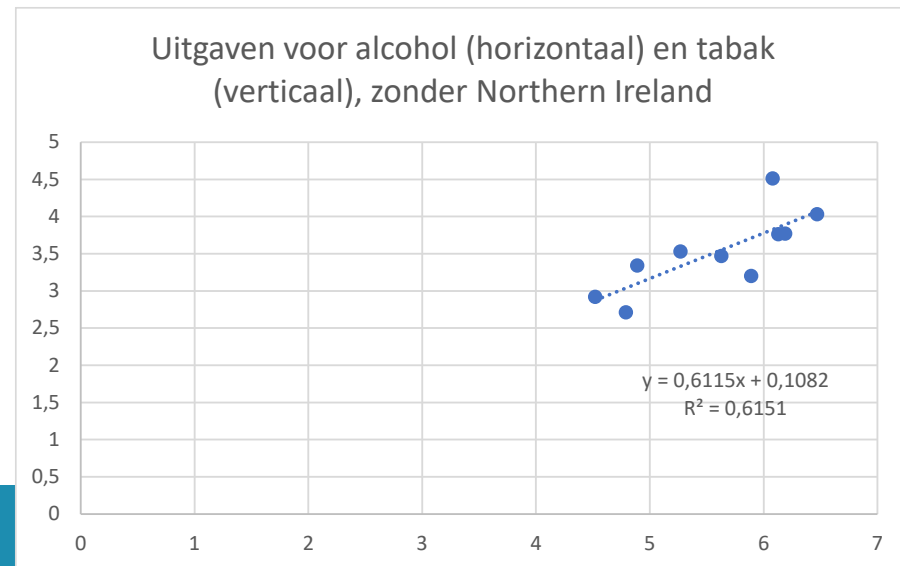
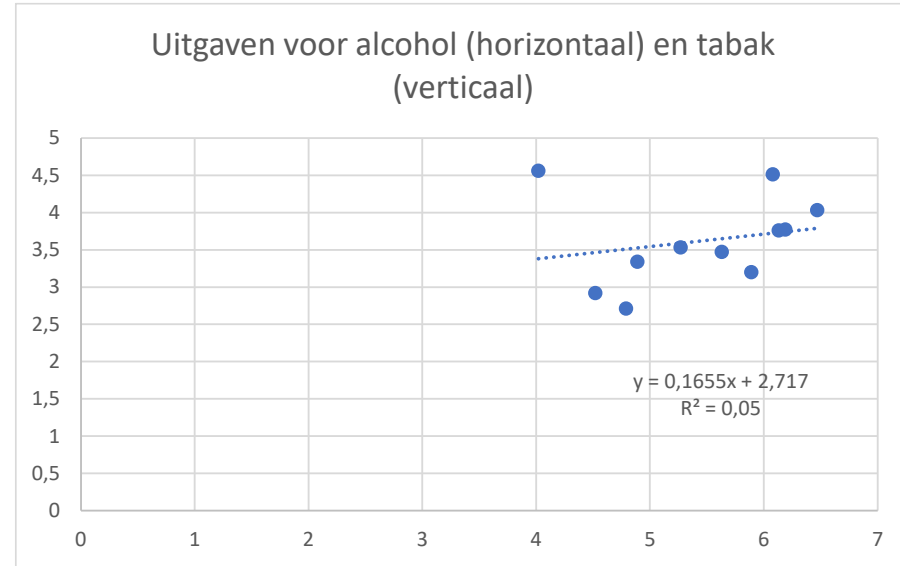


Correlatiecoëfficiënt

- black box
- berekening met ICT
- ICT geeft soms determinatiecoëfficiënt = kwadraat van correlatiecoëfficiënt
- helpt om te oordelen of een lineaire trendlijn zinvol is...
- ... maar vul aan met (zie ook vroeger)
 - visuele beoordeling
 - aandacht voor afwijkingen (patroon? verklaring uit de context?)
 - het doel dat de trendlijn moet dienen

Enkele voorbeelden

- Excel: determinatiecoëfficiënt i.p.v. correlatiecoëfficiënt
- determinatiecoëfficiënt ligt veel dichterbij 1 als afwijkende punt weggelaten wordt

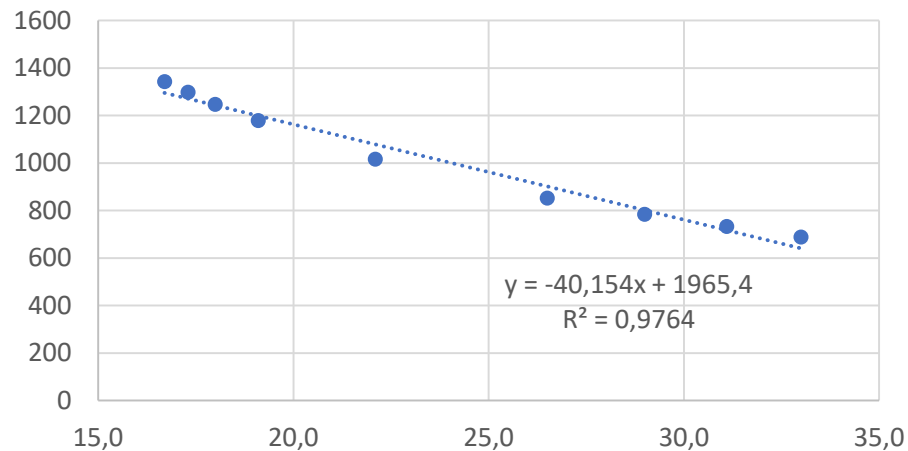


Enkele voorbeelden

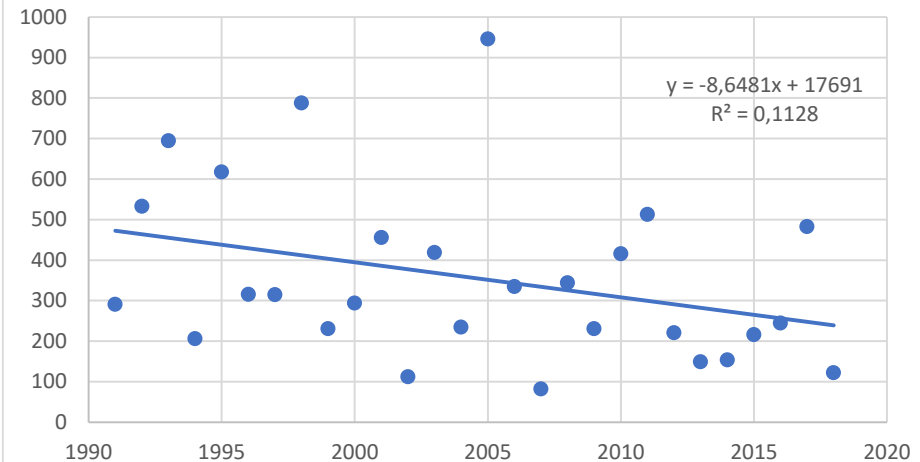
- rechte is geen goed model...
- ... ondanks hoge determinatiecoëfficiënt

- rechte is een zinvol model...
- ... ondanks lage determinatiecoëfficiënt

druk in functie van volume



Neerslag in Los Angeles (in mm)

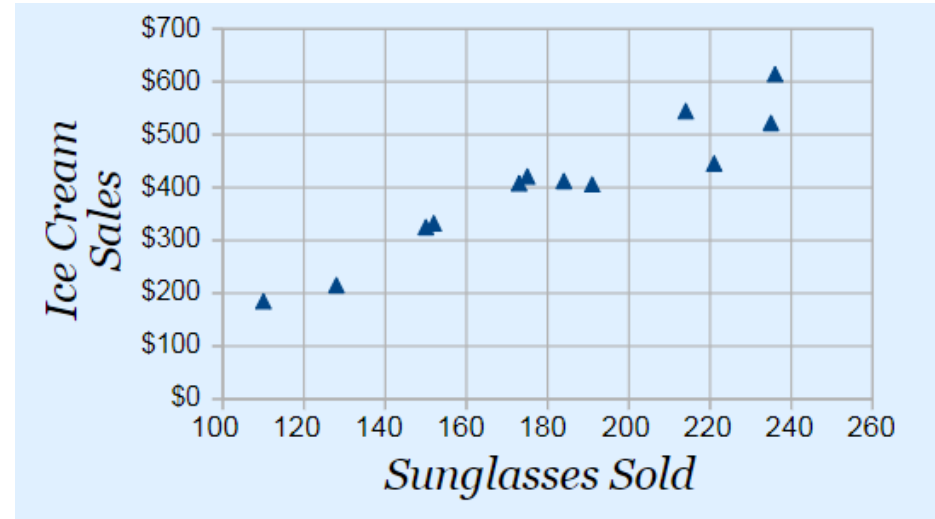


Bron: UW 36/1

Misvatting over correlatie en causaliteit

Voorbeeld

- x = verkoop zonnebrillen
- y = verkoop ijsjes
- duidelijk een hoge correlatiecoëfficiënt
- samenhang (= correlatie) tussen x en y
- hoge verkoop van ijsjes wordt niet veroorzaakt door de hoge verkoop van zonnebrillen
- beide hebben een gemeenschappelijke (andere) oorzaak, namelijk hoge temperatuur



Bron:
<https://medium.com/@seema.singh/why-correlation-does-not-imply-causation-5b99790df07e>

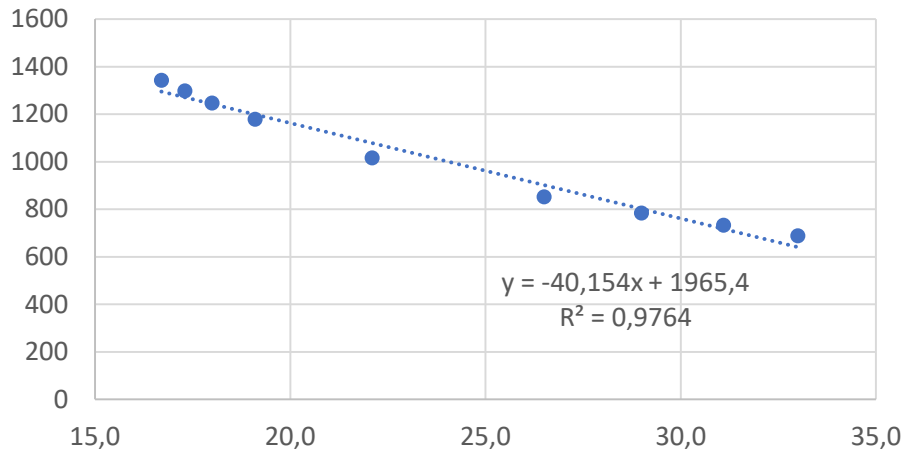
Misvatting over correlatie en causaliteit

- wat is correlatie (wel) en wat is het niet?
- correlatie is samenhang
 - positieve correlatie: hoge waarden van y komen vaak samen met hoge waarden van x en lage waarden ...
 - negatieve correlatie: hoge waarden van y komen vaak samen met lage waarden van x en ...
- goed passende trendlijn of correlatie dicht bij 1 of -1 betekent niet automatisch dat er een oorzakelijk verband is tussen de twee variabelen
- nog voorbeelden
 - zie UW21/2
 - <https://www.tylervigen.com/spurious-correlations>

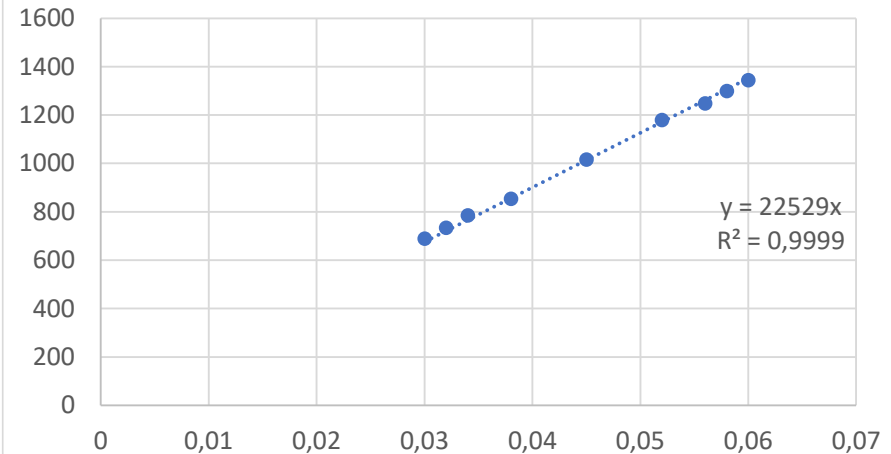
Gebogen trendlijnen

Voorbeeld: druk en volume gas bij constante temperatuur ($y = c/x$)

druk in functie van volume



druk in functie van 1/volume

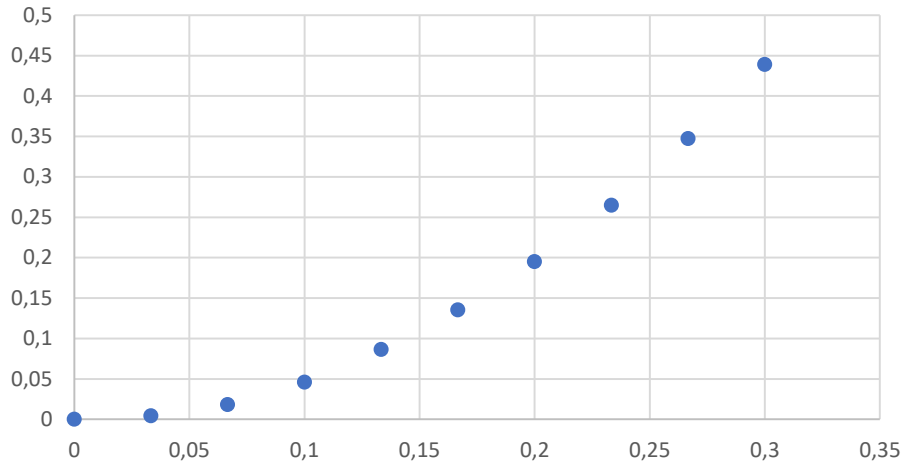


- hier past gebogen trendlijn beter
- fysica: druk \times volume constant
- druk = $c \times 1/\text{volume}$
- dus: maak extra kolom met waarden voor $1/\text{volume}$

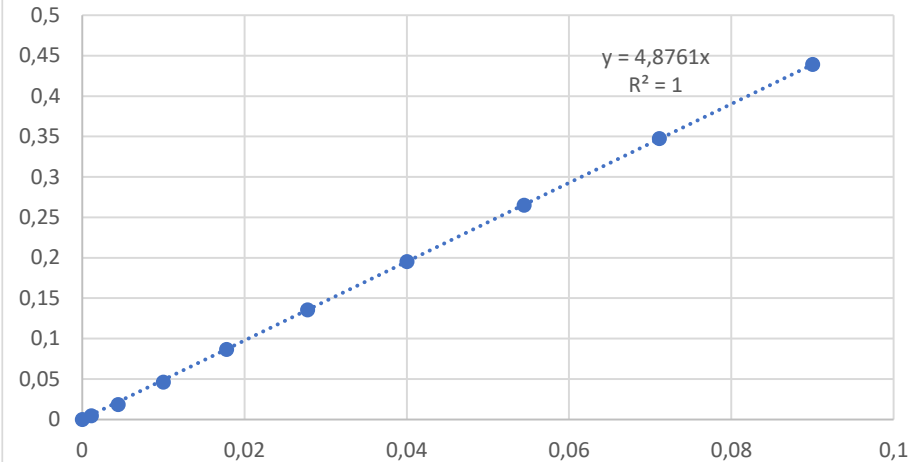
- zoek lineaire trendlijn door de oorsprong (vink 'snijpunt' aan in 'Trendlijn opmaken')
- druk = $22529/\text{volume}$
- alternatief: machtsregressie (maar dan is de exponent niet precies -1)

Voorbeeld: verplaatsing bij vrije val ($y = cx^2$)

afgelegde weg in functie van tijd



afgelegde weg in functie van tijd²



- Bron: data studentenproef vakdidactiek fysica
- verplaatsing = $c \times \text{tijd}^2$
- maak extra kolom met tijd^2
- zoek lineaire trendlijn door de oorsprong (vink 'snijpunt' aan in 'Trendlijn opmaken')

- verplaatsing = $4,871 \times \text{tijd}^2$
- we herkennen $g/2$!
- alternatief: machtsregressie of kwadratische regressie (maar ...)

Niet alleen lineaire trendlijnen

- modelleren in twee stappen
 - welke type van functie is geschikt?
dit bepaal je zelf, dat doet de tool niet voor jou
 - binnen het gekozen type de goede coëfficiënten zoeken
dat laat je aan de tool over
- beperkt tot zuiver kwadratische en omgekeerd evenredige verbanden

De wiskundige achtergrond: een tip van de sluier

Een tip van de sluier: hoe wordt de trendlijn bepaald?

- kleinstekwadratenrechte is niet de enige 'officiële' trendlijn, er bestaan ook trendlijnen die op een andere manier bepaald worden
 - bv. methodes die robuuster zijn (bv. minder gevoelig voor uitschieters),
 - SD-rechte: zie UW 21/2, p. 33
- ook een tip van de sluier oplichten over de correlatiecoëfficiënt?
 - zie einde werktekst 1 uit UW 27/3
- UW 36/1
 - GeoGebra-applet: vereenvoudiging die het probleem reduceert tot enkel het bepalen van de optimale waarde voor a , kleinstekwadratenrechte gaat door het zwaartepunt van de datapunten (wordt gebruikt als eigenschap zonder bewijs)
 - Excel-werkblad: numeriek bepalen van de optimale waarden van a en b

Beschrijvende statistiek en verklarende statistiek (niet voor leerlingen bedoeld)

- soms vormen de datapunten de hele populatie
 - en dan dient de kleinstekwadratenrechte als een benaderende beschrijving van deze populatie
- maar vaak zijn de datapunten afkomstig uit een steekproef en wil je een uitspraak doen over de populatie
 - zoals steeds in verklarende statistiek: vraag je af wat er zou gebeuren als je dit heel veel keren doet
 - nieuwe steekproef geeft nieuwe datapunten geeft nieuwe trendlijn met andere coëfficiënten, enz.
 - je werkt dus niet met één waarde voor een coëfficiënt, maar met een betrouwbaarheidsinterval, en met significantietesten (bv. is de coëfficiënt significant verschillend van 0?)
- zie §8 in UW 21/2

Bedankt voor je aandacht!

Vragen en opmerkingen
welkom!