

Meten en Interpreteren



Gecertificeerde
NLT module
voor vwo

Colofon



De module Meten en interpreteren is bestemd voor de lessen Natuur, Leven en Technologie (NLT). De module is op 10 december 2009 gecertificeerd door de Stuurgroep NLT voor gebruik op het vwo in domein B, Fundament van wetenschap en technologie. Het certificeringsnummer van de module is 4125-047-VB.

De originele gecertificeerde module is in pdf-formaat downloadbaar via ► <http://www.betavak-nlt.nl>.

Op deze website staat uitgelegd welke aanpassingen docenten aan de module mogen maken, voor gebruik in de les, zonder daardoor de certificering teniet te doen.

De module is gemaakt in opdracht van het Landelijk Ontwikkelpunt NLT.

Deze module is ontwikkeld door:

- Willem Lodewijk Gymnasium, Groningen, J.M. van Kammen
- Stellingwerfcollege, Oosterwolde, M.J. Lemstra

Expert:

- Rijksuniversiteit Groningen, Groningen, prof.dr. H.A.J. Meijer

Aangepaste versies van deze module mogen alleen verspreid worden, indien in dit colofon vermeld wordt dat het een aangepaste versie betreft, onder vermelding van de naam van de auteur van de wijzigingen.

Materialen die leerlingen nodig hebben bij deze module zijn beschikbaar via het vaklokaal NLT:

► <http://www.vaklokaal-nlt.nl/>

© 2009. Versie 1.0

Het auteursrecht op de module berust bij Stichting Leerplan Ontwikkeling (SLO). SLO is derhalve de rechthebbende zoals bedoeld in de hieronder vermelde creative commons licentie.

De auteurs hebben bij de ontwikkeling van de module gebruik gemaakt van materiaal van derden en daarvoor toestemming verkregen. Bij het achterhalen en voldoen van de rechten op teksten, illustraties, enz. is de grootst mogelijke zorgvuldigheid betracht. Mochten er desondanks personen of instanties zijn die rechten menen te kunnen doen gelden op tekstgedeeltes, illustraties, enz. van een module, dan worden zij verzocht zich in verbinding te stellen met SLO.

De module is met zorg samengesteld en getest. Landelijk Ontwikkelpunt NLT, Stuurgroep NLT, SLO en auteurs aanvaarden geen enkele

aansprakelijkheid voor onjuistheden en/of onvolledigheden in de module. Ook aanvaarden Landelijk Ontwikkelpunt NLT, Stuurgroep NLT, SLO en auteurs geen enkele aansprakelijkheid voor enige schade, voortkomend uit (het gebruik van) deze module.

Voor deze module geldt een



Creative Commons Naamsvermelding-Niet-commercieel-Gelijk delen 3.0 Nederland Licentie

► <http://creativecommons.org/licenses/by-nc-sa/3.0/nl>

Foto voorpagina: <http://www.meridiaan-hsl.nl/wiskunde/rekenen/meten/meten.gif>

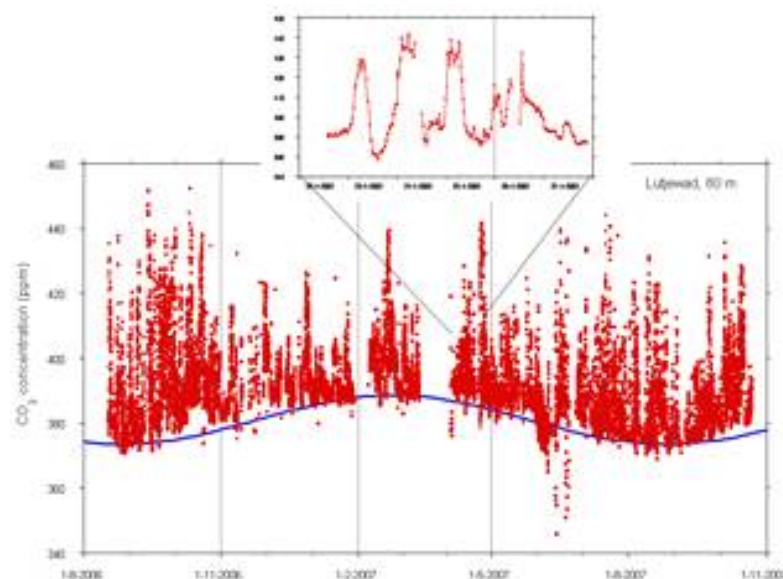
Inhoud

Inleiding.....	1
1 Meten	5
1.1 Meten aan CO ₂ , de koolstofkringloop.....	5
1.2 Meetschalen	6
2 Meetfouten.....	12
2.1 Classificatie van fouten.....	14
2.2 Toevallige fouten verkleinen	19
2.3 Meten van CO ₂ -concentraties.....	21
2.4 Weergave van fouten	22
2.5 Rekenen met meetfouten	24
3 Experimenten	28
3.1 Experiment dichtheid	28
3.3 Experiment valversnelling.....	31
3.4 Experiment veerconstante	32
4 Interpreteren	34
4.1 Seizoensinvloed op CO ₂ -concentratie	34
4.2 Combineren en filteren van meetgegevens	35
4.3 Correlaties	40
4.4 Lineaire regressie.....	52
5 Eigen onderzoek CO ₂ -metingen	57
5.1 Meten van CO ₂	57
5.2 Opbouw van de atmosfeer	62
5.3 Eigen onderzoek	66
Bijlage 1 Rekenen met meetfouten.....	68
Bijlage 2 Werken met lijsten in de GRM.....	74
Bijlage 3 Importeren van de gegevens van de meetstations in Excel	76
Bijlage 4 Excel	77
Bijlage 5 URL lijst	84

Inleiding

Klimaatverandering ten gevolge van de uitstoot van CO_2 : je hoeft de krant maar open te slaan of de tv maar aan te zetten en je wordt overspoeld met informatie over smeltende ijskappen, een stijgende zeespiegel, extreme droogtes of juist enorme overstromingen. Politici doen er alles aan om te laten merken dat ze dit probleem hoog op de agenda hebben staan en menige internationale top is eraan gewijd. Welke invloed de toename van CO_2 in de atmosfeer precies heeft op het klimaat, is nog niet duidelijk. Wordt het natter, droger, warmer of zit er juist een nieuwe ijstijd aan te komen? De meningen van deskundigen lopen uiteen.

In deze module zullen we het nu eens niet hebben over de klimaatverandering. We zullen ons concentreren op de CO_2 -metingen die ten grondslag liggen aan deze discussie. Het lijkt zo eenvoudig: je plaatst een meetsensor, meet een paar jaar de CO_2 -concentratie en kijkt of er sprake is van een toename. De werkelijkheid is echter niet zo simpel. Het goed uitvoeren van metingen en er vervolgens ook nog de juiste conclusies aan ontleunen, is een tak van sport waarin je nooit uitgeleerd bent. Veel wetenschappers voelen zich wat onzeker worden bij het onderwerp foutenanalyse en die onzekerheid neemt toe wanneer statistiek op de meetwaarden moet worden toegepast. Bijkomend probleem bij de CO_2 -metingen is dat de concentratie enorm varieert door menselijke activiteiten, weersomstandigheden en plantengroei, zoals te zien is in figuur 1. Het is dan ook lastig om er duidelijke trends in te herkennen. Daarnaast heb je natuurlijk ook nog te maken met de (on)nauwkeurigheid van de gebruikte meetapparatuur.



Figuur 1: gemeten CO_2 concentratie in de lucht over een periode van ruim een jaar.

Wereldwijd houden vele wetenschappers zich bezig met de metingen aan CO₂. Enkele jaren geleden is ook de Rijksuniversiteit Groningen met een project begonnen waarbij CO₂-meters bij middelbare scholen worden geplaatst. Eerst alleen in Nederland, maar inmiddels is er een netwerk van deelnemende scholen in Europa ontstaan: het school CO₂-net. De meetgegevens van dit netwerk zijn voor iedereen toegankelijk en in deze module zul je uiteindelijk een eigen onderzoek uitvoeren waarbij je zinnige conclusies probeert te trekken uit de brij van meetgegevens van het school CO₂-net.

Uiteraard is dit niet mogelijk zonder kennis op het gebied van meten en interpreteren. Aan het eind van de module heb je dan ook geleerd:

- wat wordt verstaan onder een meting en op welke niveaus er gemeten kan worden
- hoe onnauwkeurigheden in metingen optreden en hoe je deze overzichtelijk kunt weergeven
- wat het onderscheid is tussen systematische en toevallige fouten
- een meting zo nauwkeurig mogelijk uit te voeren en bij het uitdenken van een experiment al rekening te houden met de nauwkeurigheid van de metingen
- hoe meetonnauwkeurigheden doorwerken in het verwerken van de meetgegevens
- wanneer je met een deel van de meetgegevens verder mag rekenen zonder dat er sprake is van vervalsing van de gegevens
- hoe je een (al dan niet rechtlijnige) correlatie tussen twee variabelen kunt aantonen
- hoe je de formule van een rechtlijnige regressielijn opstelt
- waardoor de CO₂-concentratie in de atmosfeer erg varieert.

De opbouw is hierbij als volgt:

In hoofdstuk 1 kijken we naar wat allemaal onder een meting wordt verstaan. De uitkomst van een meting kan zijn dat de afstand van school naar huis 9,87 km bedraagt, maar de uitkomst van een meting kan ook zijn dat je een ijsco met aardbeiensmaak lekkerder vindt dan één met bananensmaak. We zeggen dat er op verschillende schalen wordt gemeten. De schaal waarop je meet is van invloed op de conclusies die je later kunt trekken.

Hoofdstuk 2 is getiteld 'Meetfouten'. Het woord 'meetfouten' is misleidend, maar het wordt in vrijwel alle wetenschappelijke literatuur gebruikt. Het suggereert namelijk dat er iets verkeerd is gegaan bij de meting, waardoor de uitkomst niet helemaal (of helemaal niet!) klopt. Soms is dat ook zo, maar meestal worden meetonnauwkeurigheden bedoeld: door de beperkte nauwkeurigheid van de gebruikte apparatuur en door menselijke factoren zitten er in bijna alle metingen kleine

onnauwkeurigheden. In hoofdstuk 2 leer je waar deze onnauwkeurigheden vandaan komen, hoe je ze netjes kunt weergeven en hoe je ermee om moet gaan bij de verwerking van meetgegevens.

In hoofdstuk 3 zul je een aantal bekende natuurkunde-experimenten uitvoeren. We benaderen de experimenten op een andere manier dan je waarschijnlijk gewend bent. We richten ons namelijk op de vraag: hoe groot zijn de onnauwkeurigheden in de gemeten waarden en wat kunnen we doen om de metingen zo precies mogelijk uit te voeren? Je past hierbij ook de theorie uit hoofdstuk 2 toe. Daar heb je immers geleerd om meetresultaten met een zekere onnauwkeurigheid erin te verwerken.

In hoofdstuk 4 is het combineren en filteren van meetgegevens het eerste onderdeel dat we bespreken. Beide technieken kunnen, mits goed toegepast, orde scheppen in een brij aan meetgegevens. Worden ze niet goed toegepast, dan leidt dat al snel tot een onjuiste weergave van de werkelijkheid. Je kunt immers de meetwaarden die je niet zo goed uitkomen, weglaten en zo toewerken naar het resultaat dat je graag ziet. Dat is natuurlijk niet de bedoeling!

Daarna komen correlatie en regressie aan bod. We spreken van correlatie wanneer twee of meer variabelen van elkaar afhankelijk zijn. Je kunt je voorstellen dat de eigenaar van een café bij warm weer meer omzet draait op z'n terras dan bij koud weer. We zeggen dan dat de variabelen temperatuur en omzet een correlatie vertonen. Hoe kun je dat met meetwaarden netjes aantonen? En als je dat hebt aangetoond, is deze beïnvloeding dan ook in een formule te vatten? Het opstellen van een dergelijke formule wordt regressie genoemd.

Metingen aan CO₂ vormen zoals gezegd de rode draad binnen de module. Dergelijke metingen zijn echter zo complex (denk maar aan alle factoren die van invloed zijn op de concentratie van CO₂ in de atmosfeer), dat we vaak heel andere voorbeelden gebruiken om de theorie uit hoofdstuk 1, 2 en 4 uit te leggen. Waar mogelijk speelt CO₂ of het klimaat wel een rol.

Hoofdstuk 5 ten slotte omvat het eigen onderzoek, waarin je aan de slag gaat met de meetgegevens van het school CO₂-net. Je krijgt de beschikking over de ruwe meetgegevens van de meters van één of meerdere scholen. Het is aan jou om deze gegevens op de juiste wijze te verwerken, waarbij je zoveel mogelijk van de opgedane kennis toepast.

Je zult in de module op een aantal verschillende manieren worden beoordeeld:

- De uitwerkingen van (een deel van) de vragen in de module lever je bij jouw docent in. (25%)

- Je levert een verslag in van de experimenten die je in deze module uitvoert. (15%)
- Je maakt een schriftelijke toets. (30%)
- Als eindopdracht voer je een eigen onderzoek uit waarbij je een ruwe serie CO₂ -meetgegevens verwerkt en interpreteert. Alle opgedane kennis breng je hierbij in de praktijk. (30%)

Dan rest ons niets anders dan je veel plezier en succes te wensen bij het doorwerken van de module!

1 Meten

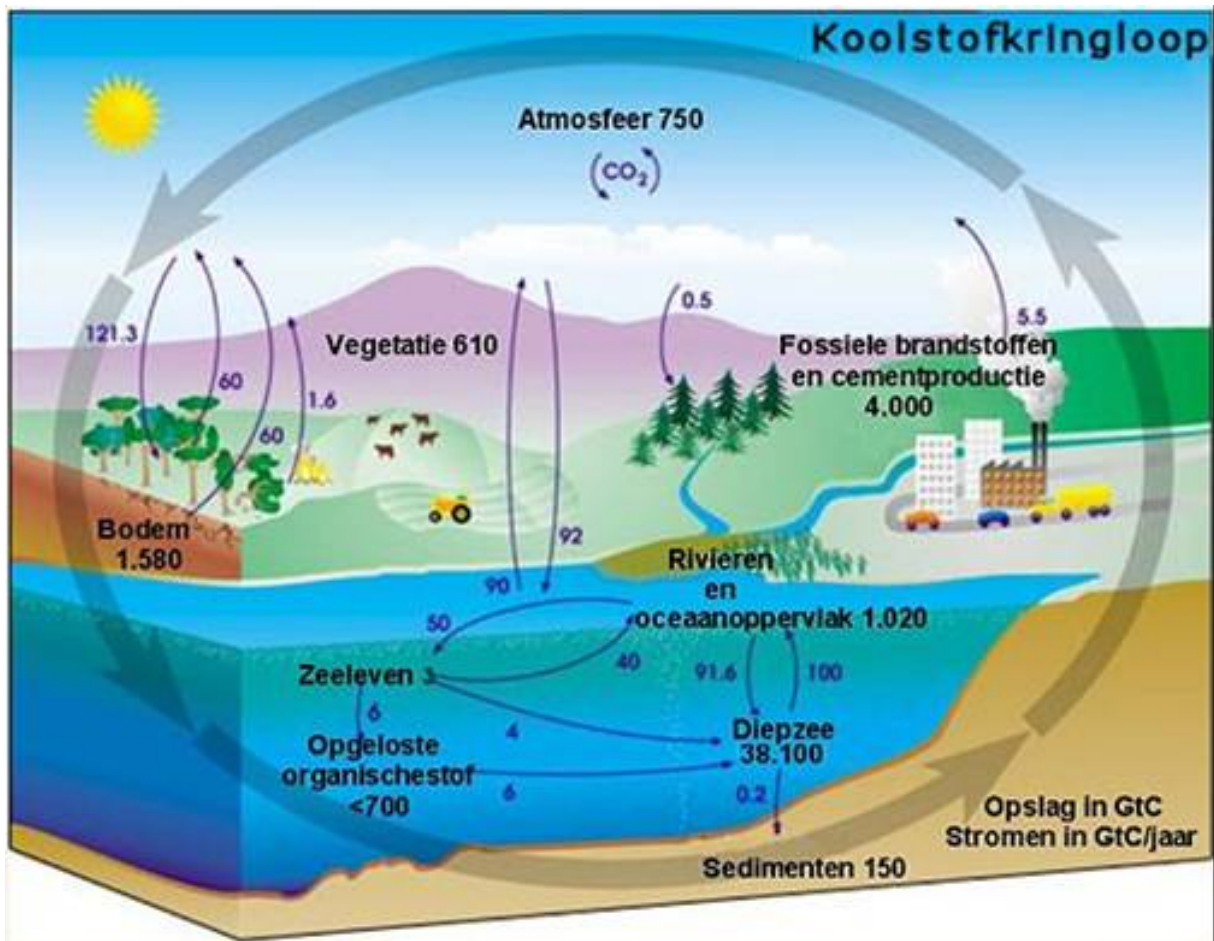
“Meten: de lengte, inhoud, temperatuur, oppervlakte enz. bepalen van”
(Dikke van Dale).

Het lijkt zo duidelijk als je de beschrijving leest van het woord ‘meten’ in de Dikke van Dale. En waarschijnlijk heb je er nog nooit bij stilgestaan wat eigenlijk wordt bedoeld met het woord ‘meten’. De afstand tussen school en huis, het gewicht van een olifant, de snelheid van een vliegtuig, allemaal grootheden die te meten zijn. Logisch toch?

1.1 Meten aan CO₂, de koolstofkringloop

Het meten van CO₂ gebeurt op veel plaatsen in de wereld, bijvoorbeeld op Hawaï, op Antarctica en in de Verenigde Staten. Maar ook in Lutjewad bij de Waddenzee en op de daken van een aantal scholen in Nederland en Europa. De verzamelde gegevens worden vervolgens door de wetenschap gebruikt. Maar om bruikbare gegevens te krijgen, moet je goed weten wat je meet.

CO₂ komt niet alleen voor in de **atmosfeer** (dampkring), maar ook in de **hydrosfeer** (het water op aarde) en de **lithosfeer** (het door de planten- en dierenwereld bezette deel van de aarde). In elke levende cel zit koolstof in de moleculen. Al deze koolstof is uiteindelijk via fotosynthese afkomstig uit CO₂. Uiteindelijk is bijna al het leven op aarde afhankelijk van fotosynthese. Koolstof van levende wezens komt door middel van verbranding in de lucht; daarna leggen de planten die koolstof weer vast door fotosynthese. Dit proces heet de **koolstofkringloop**. Zie figuur 2.



Figuur 2: koolstofkringloop. De getallen tonen hoeveel miljarden tonnen koolstof (GtC) in de verschillende opslagplaatsen aanwezig zijn. De getallen bij de pijlen tonen hoeveel koolstof tussen de verschillende opslagplaatsen per jaar uitgewisseld wordt. N.B.:vegetatie omvat alle organismen.

Je zou verwachten dat er een vaste hoeveelheid koolstof aanwezig is in de koolstofkringloop. Toch lijkt er steeds meer koolstof op aarde te komen door menselijk handelen waardoor het versterkte broeikaseffect ontstaat. Maar is dat ook zo?

1. Opdracht

- Bekijk figuur 2 van de koolstofkringloop. Waar komt de koolstof vandaan die door menselijk handelen in de atmosfeer terecht komt?
- Vind je de uitspraak dat er steeds meer koolstof in de koolstofkringloop komt, terecht? Gebruik in jouw antwoord de begrippen 'kleine en grote koolstofkringloop'. Zoek eventueel de betekenis van deze begrippen op.

1.2 Meetschalen

In het begin van dit Hoofdstuk heb je de definitie van meten gezien die in de Dikke van Dale staat. ("Meten: de lengte, inhoud, temperatuur, oppervlakte enz. bepalen van").

Maar meten kan in een veel ruimere zin van het woord worden gebruikt. Kijk maar eens naar figuur 3. Elke afbeelding heeft te maken met een meting.



a. Afrikaanse olifant

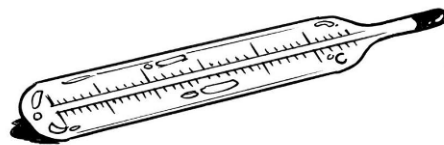


c. Matterhorn (4478 m)

Vind u het belangrijk om de verwarming lager te zetten als u daar positieve invloed mee heeft op het versterkte broeikaseffect?

- ☐ Heel belangrijk
- ☐ Belangrijk
- ☐ Redelijk belangrijk
- ☐ Niet zo belangrijk
- ☐ Helemaal niet belangrijk

b. Enquêtevraag



d. Thermometer

Figuur 3: bij elke afbeelding kan er sprake zijn van een meting.

Bij de afbeelding van de thermometer is het direct duidelijk dat het om een meetinstrument gaat. De hoogte van de Matterhorn is verschillende keren gemeten en men heeft de hoogte bepaald op 4478 m boven zeeniveau. Bij de enquêtevraag is het misschien minder duidelijk, maar ook daar wordt gemeten, namelijk de mening van een respondent. Maar hoe zit het bij de afbeelding van de olifant?

Je moet je realiseren dat je op verschillende niveaus kunt meten. Daarbij kun je onderscheid maken tussen de volgende meetschalen:

1. Nominale schaal
2. Ordinale schaal
3. Interval schaal
4. Ratio schaal.

Nominale schaal

Wanneer je op het **nominale niveau** meet, gaat het slechts om het benoemen van datgene wat je meet. De variabele is dan altijd discreet. Voorbeelden zijn namen van landen, haarkleuren, bloedgroepen en diersoorten. Dit verklaart ook het plaatje van de Afrikaanse olifant.

Bedenk:

- dat een **discrete variabele** slechts gehele getallen of klassen als mogelijke uitkomsten heeft
- dat nominaal is afgeleid van het Latijnse 'nomen' dat 'naam' betekent
- dat er **kwalitatieve en kwantitatieve variabelen** bestaan
- dat je alleen met de waarden van kwantitatieve variabelen kunt rekenen.

Variabelen in de nominale schaal zijn altijd kwalitatief, zelfs als er wel gebruik wordt gemaakt van cijfers. Voorbeelden hiervan zijn de toekenning van getallen aan bepaalde categorieën (1: rood haar, 2: blond haar, 3: bruin haar) en een telefoonnummer of bankrekeningnummer. Bij de laatste twee is de cijferreeks immers alleen bedoeld om een individu (de specifieke telefoonaansluiting) herkenbaar te maken. Rekenen met zulke getallen is zinloos.

2. Vraag

Geef nog een nominale variabele waarbij cijfers worden gebruikt.

Ordinale schaal

Bij de **ordinale schaal** is sprake van ordening. Een hogere waarde op de schaal betekent dat een eigenschap groter is, of hoger, langer, beter, mooier, enz. De enquêtevraag over het lager zetten van de verwarming is hier een voorbeeld van. De onderverdeling is, net als bij de nominale schaal, kwalitatief en discreet. Rekenen met de uitkomsten is dus niet mogelijk!



Figuur 4: is rekenen met een ordinale schaal betrouwbaar?

Voorbeelden van een ordinale schaal zijn het onderwijssysteem (met een onderverdeling in bijvoorbeeld vmbo, havo en vwo), de rangen in het leger en de diploma's bij het zwemmen.

Ordinale schalen worden veel gebruikt om meningen of percepties van mensen in een enquête te peilen.

3. Vraag

Geef zelf een ander voorbeeld van een ordinale schaal.

Intervalschaal

De *intervalschaal* is de laagste schaal waarbij wel kwantitatief wordt gemeten. Ook hierbij is er sprake van een rangschikking, maar nu heeft het verschil tussen twee waarden wel een betekenis. Het bekendste voorbeeld is de temperatuursmeting in graden Celsius. Het verschil tussen 15°C en 20°C is even groot als het verschil tussen 35°C en 40°C. Er kan alleen niet worden gezegd dat het bij 40°C twee keer zo warm is als bij 20°C. Dat komt doordat het nulpunt willekeurig is gekozen. Dit ontbreken van een absoluut nulpunt is kenmerkend voor de intervalschaal.

Een ander bekend voorbeeld van de intervalschaal is de tijdsaanduiding gedurende de dag (dus niet de tijdsduur!).

Het verschil tussen een ordinale schaal en een intervalschaal is soms lastig te bepalen. Neem de enquêtevragen in figuur 5.



Figuur 5: voorbeelden van enquêtevragen.

Bij de vraag over cola (figuur 5) is sprake van een ordening, waarbij het verschil tussen 0 en 1 niet even groot hoeft te zijn als het verschil tussen 1 en 2. Daar is dus sprake van een ordinale schaal.

Ook bij de vraag over sportiviteit (figuur 5) is sprake van een ordening, maar doordat er nu een groter aantal keuzemogelijkheden is, kun je stellen dat het verschil tussen 2 en 3 ongeveer even groot is als het verschil tussen 5 en 6. Deze vaste stapgrootte betekent dat we nu eigenlijk op intervalschaal aan het meten zijn. Door meer gedetailleerde antwoordmogelijkheden te gebruiken kan een ordinale schaal dus veranderen in een intervalschaal.

Het is uiteraard lastig aan te geven waar de grens tussen deze beide niveaus precies gelegd moet worden, deze onderverdeling blijft een beetje arbitrair.

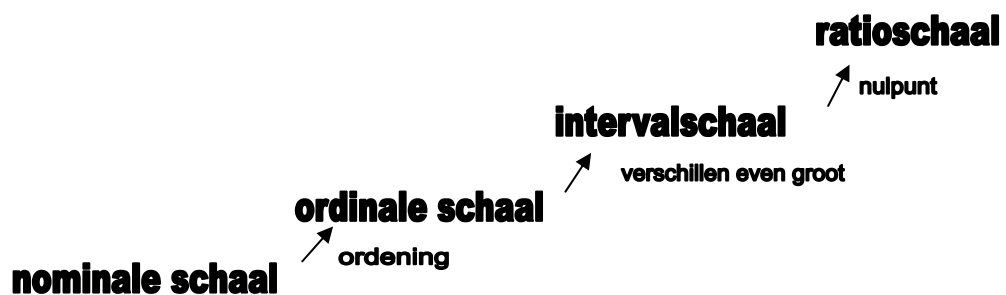
4. Vraag

Zowel discrete als continue variabelen zijn op de intervalschaal te meten. Geef een voorbeeld van elk van beide metingen.

Ratioschaal

Het hoogste meetniveau is de **ratioschaal**. Deze schaal is wel kwantitatief. Net als bij de intervalschaal heeft het verschil tussen twee waarden een betekenis, maar daarnaast heeft de ratioschaal een absoluut nulpunt. Hierbij kun je denken aan de temperatuur uitgedrukt in Kelvin, afstand, lengte, of de tijdsduur. Anders dan bij de tijdsaanduiding (intervalschaal) is er bij tijdsduur wel sprake van een nulpunt. Daardoor kun je ook zeggen dat een klus van 6 uur twee keer zo lang duurt als een klus van 3 uur.

Bij elke schaal hoger komt er dus één extra eigenschap bij (zie figuur 6).



Figuur 6: meet schalen met de bijbehorende eigenschappen.

Je kunt je afvragen waarom we zo nadrukkelijk onderscheid maken tussen de verschillende schalen waarop wordt gemeten. Bij de verwerking en interpretatie van meetgegevens (**statistische analyse**) spelen de verschillende schalen weer een rol. Welke manier van verwerken (**statistiek**) je op de meetgegevens mag toepassen hangt af van de schaal waarop de variabele is gemeten. Het zou bijvoorbeeld onzin zijn om een gemiddelde te bepalen van een variabele die op nominale schaal is gemeten.

Uiteindelijk is het natuurlijk de bedoeling om vooruit te denken, zodat je de meetmethodes al aanpast op de verwerking die je later op de meetgegevens wilt toepassen.

Je hebt gezien dat 'meten' een erg brede betekenis kan hebben. Wanneer we het in het vervolg van de module over 'meten' hebben, zullen we daarmee een kwantitatieve meting bedoelen. Binnen de bètawetenschappen maken we immers zoveel mogelijk gebruik van dit soort metingen, waar we aan kunnen rekenen.

5. Opdracht

5.1 Meetschalen

Hieronder is een aantal situaties geschetst of vragen gesteld waarbij sprake is van een meting. Probeer in elk van de gevallen aan te geven op welke schaal wordt gemeten. Leg je antwoord uit.

- a. Sharon vult een enquête in met allemaal ja-nee vragen.
- b. De Olympische ringen hebben de volgende kleuren: blauw, geel, rood, groen, zwart.
- c. Dit T-shirt is te koop in de maten XXS t/m XXL.
- d. Wat is je sofinummer?
- e. Op deze cd staan 15 nummers.
- f. Bij meten maken we onderscheid tussen de volgende schalen: nominaal, ordinaal, interval en ratio.
- g. Er zijn verschillende redenen waarom iemand een ansichtkaartje koopt (verjaardag, huwelijk, ...).
- h. Het aantal bezoekers van Pinkpop.
- i. Het behaalde cijfer op jouw laatste proefwerk Nederlands.
- j. Elke voetballer heeft een eigen rugnummer.
- k. Mijn opa is 80 jaar.
- l. Bij een enquête wordt een gewichtsindeling in klassen gebruikt: < 40 kg, 40-50 kg, 50-60 kg, 60-80 kg, > 80 kg.
- m. Bij een enquête wordt een gewichtsindeling in klassen gebruikt: 0-20 kg, 20-40 kg, 40-60 kg, 60-80 kg, 80-100 kg, 100-120 kg, 120-140 kg.
- n. Bij de reactie kwam 5,0 mol CO₂ vrij.
- o. De weerstand van dit lampje is 2 ohm.
- p. Wat zijn de cijfers van je postcode?
- q. Een computer gebruikt het binaire telsysteem.

5.2 Rekenkundige formules

Welke rekenkundige bewerkingen je kunt uitvoeren met een variabele, hangt af van het niveau waarop de variabele is gemeten.

Geef in het schema in figuur 7 aan welke rekenkundige formules op welk niveau zinvol zijn.

Rekenkundige formule	Nominale schaal	Ordinale schaal	Interval schaal	Ratio schaal
$3 < 4$				
$3 + 4 = 7$				
$2 * 3 = 6$				

Figuur 7: tabel rekenkundige formules.

5.3 Enquêtevragen

Afhankelijk van de wijze waarop een enquêtevraag wordt gesteld, kan de leeftijd van een respondent op een ordinale of op een ratioschaal worden gemeten.

Bedenk hiervoor twee verschillende enquêtevragen.

2 Meetfouten

"To err is human; to describe the error properly is sublime."
-- Cliff Swartz, Physics Today 37 (1999), 388.

In de loop der jaren is enorm veel gepubliceerd over foutenanalyse. Eén van de redenen dat er zoveel is geschreven, ligt in het feit dat er nog geen eenduidige richtlijnen zijn om met fouten om te gaan.

In dit hoofdstuk maak je eerst kennis met het begrip ‘meetfouten’. We kijken wat voor soorten fouten er kunnen optreden, vervolgens hoe je die netjes kunt weergeven en tot slot hoe je bij de verwerking van meetresultaten doorrekent met deze meetfouten.

Eigenlijk is ‘meetfouten’ een misleidend woord, want het wekt de suggestie dat er bij een meting iets fout is gegaan, waardoor het antwoord niet klopt. Dat hoeft lang niet altijd zo te zijn. Meetapparatuur en meetmethodes hebben vaak een beperkte nauwkeurigheid, waardoor de gemeten waarde iets kan afwijken van de werkelijkheid. Meetonzekerheid of meetonnauwkeurigheid is dan eigenlijk een beter woord. Soms wordt er ook op een verkeerde manier gemeten of is de meetapparatuur niet goed afgesteld. In dat geval treden er wel echte ‘meetfouten’ op.

Lees het artikel in bron 1 maar eens door.

1. Bron: plotselinge afkoeling klimaat in 1945 gevolg van meetfout.
Gepubliceerd: 1 juni 2008 16:10 | Gewijzigd: 1 juni 2008 17:21
Door Karel Knip

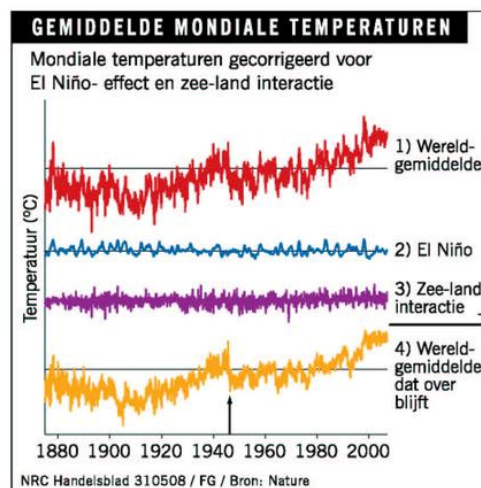
Rotterdam, 1 juni. Niet alle temperatuurverandering die klimatologen de afgelopen 140 jaar waarnamen, was klimaatverandering. Amerikaanse en Britse onderzoekers laten zien dat de plotselinge mondiale afkoeling die aan het eind van de Tweede Wereldoorlog optrad een artefact is (Nature, 29 mei).

Toen de oorlog in 1945 afliep, nam het aantal Amerikaanse schepen dat in vaste routine warmtemetingen deed aan zeewater snel af. Tegelijk vermeerderde het aantal Britse schepen. Dat had zijn weerslag op de berekening van de gemiddelde mondiale zeewatertemperatuur, omdat Britse zeelieden de zeewatertemperatuur meten aan water dat ze in een emmer omhoog halen. Amerikanen meten aan het koelwater dat de

machinekamer in gaat. De Amerikanen komen te hoog uit, de Britten te laag.

In het klimaatdebat speelt het verloop van de gemiddelde mondiale temperatuur zoals die voor de afgelopen 140 jaar valt te berekenen uit directe metingen (met thermometers) een grote rol. In 1986 publiceerden Britse onderzoekers van de University of East Anglia geleid door Phil Jones een reconstructie vanaf 1861 gebaseerd op historische metingen.

In ruwe lijn blijkt de mondiale temperatuur na 1900 gestaag te stijgen tot in 1945 een abrupte afkoeling inzet. Die houdt aan tot ongeveer 1970, dan begint de huidige sterke opwarming. De vreemde daling na 1945, toen het broeikaseffect zich al manifesteerde, is toegeschreven aan het koelende effect van zwaveldioxide uit zwavelhoudende brandstof.



Figuur 8: de abrupte temperatuursdaling in 1945 is goed te zien in een grafiek die is 'opgeschoond' voor het El Niño-verschijnsel in de Grote Oceaan en de invloed van sterke westelijke winden.

Voor het recente artikel in Nature, waaraan Phil Jones meewerkte, is de grillige reeks 'opgeschoond'. De opgeschoonde grafiek toont de abrupte daling rond 1945 extra duidelijk (zie hierboven). De plotselinge daling blijkt dus niet, zoals andere dalingen, toe te schrijven aan een vulkaanuitbarsting en ook niet aan de atoomaanval op Japan, maar aan een artefact. Omdat het aardoppervlak voor tweederde uit water bestaat is de invloed van de Britse emmermetingen op de berekende mondiale luchttemperatuur erg groot.

Bron:

http://www.nrc.nl/wetenschap/article1109109.ece/Plotselinge_afkoeling_klimaat_in_1945_gevolg_van_meetfout

6. Opdracht

In het artikel in bron 1 wordt geconcludeerd dat de abrupte temperatuursdaling rond 1945 is toe te schrijven aan een 'artefact'. Leg uit wat hiermee wordt bedoeld.

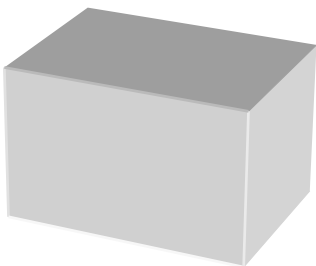
Het artikel maakt in elk geval het belang van goede meetwaarden duidelijk. Men was al bezig allemaal logische verklaringen te vinden voor de wereldwijde temperatuursdaling na 1945, terwijl de daling eigenlijk helemaal niet bleek te bestaan. Het heeft geen zin met interpreteren te beginnen voordat je zeker bent van je meetwaarden!

Hetzelfde geldt voor de metingen aan CO₂. We kunnen wel proberen voorspellingen te doen over de veranderingen van het klimaat door de toename van CO₂ in de atmosfeer, maar dan moeten we wel zeker weten dat de CO₂-concentratie inderdaad aan het toenemen is.

Omdat het zo belangrijk is om nauwkeurige metingen te verrichten en om te weten hoe groot de eventuele fouten in de meetwaarden zijn, zullen we ons daar in dit hoofdstuk in verdiepen. Hoe je doorrekent met waarden waar een zekere onzekerheid in zit, zul je zien in paragraaf 2.5. Soms is de onnauwkeurigheid in het uiteindelijke antwoord veel groter dan in de oorspronkelijke meetwaarde! In hoofdstuk 3 krijg je de kans om het geleerde in de praktijk te brengen, door experimenten uit te voeren waarbij alles draait om nauwkeurig meten.

2.1 Classificatie van fouten

Stel je krijgt voor een experiment van je docent een blokje metaal. Je probeert zo nauwkeurig mogelijk de dichtheid van dit blokje te bepalen en vergelijkt jouw uitkomst met die van je klasgenoten, die hetzelfde blokje gebruiken.



Figuur 9: blokje metaal.

Je zult verschillen in de uitkomsten zien. De meeste dichtheden zullen wel enigszins bij elkaar in de buurt liggen, maar wellicht dat een enkele ook heel sterk afwijkt.

7. Vraag

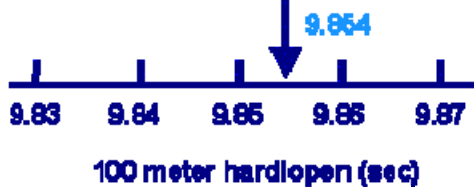
Geef zoveel mogelijk verklaringen waarom de uitkomsten van elkaar verschillen.

Waar gemeten wordt, treden meetfouten en/of meetonnauwkeurigheden op. Stel dat de lengte van het blokje dat je gemeten hebt bij benadering 2,65574 cm is, dan is dat met de meetinstrumenten die je op school tot je beschikking hebt, niet zo nauwkeurig te meten. Natuurlijk zijn er instrumenten waarmee je veel preciezer kunt meten, maar daarmee kom je slechts tot een betere benadering.

Wist je nog:.....

dat lengte net als bijvoorbeeld tijd en massa een continue variabele is? Alleen al daarom is de precieze lengte nooit te meten.

Hoe dicht 2 tijden ook bij elkaar liggen, er past altijd een tijd tussen.



Figuur 10: continue variabele.

Meeton nauwkeurigheden zijn dan ook onvermijdelijk en dat is niet erg. Het is alleen wel zaak dat je weet dat ze optreden en dat je weet hoe je er mee moet omgaan. Vooral wanneer je verder rekt met meetgegevens en daarbij geen rekening houdt met kleine onzekerheden, kunnen je uitkomsten klinkklare onzin zijn. Het is dus zaak op de juiste wijze met onnauwkeurigheden om te gaan.

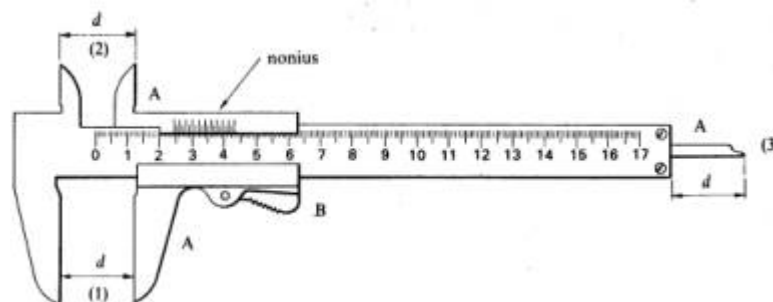
Bij het uitvoeren van metingen treden twee soorten fouten (onnauwkeurigheden) op: systematische fouten en toevallige fouten.

Toevallige fouten

Toevallige fouten kunnen door verschillende oorzaken optreden.

Meetapparatuur heeft maar een beperkte nauwkeurigheid. Stel dat je met een liniaal met een onderverdeling in millimeters de lengte van een staafje van ongeveer 6,812 cm probeert te meten. De meting kan slechts op 1 mm nauwkeurig worden gedaan en afgezien van andere fouten zul je uitkomen op 6,8 cm, waarbij je met enige zekerheid kunt zeggen dat de werkelijke lengte ergens ligt tussen 6,75 cm en 6,85 cm.

Gebruik je nauwkeuriger meetapparatuur, als een schuifmaat met nonius (zie figuur 11), dan kun je de werkelijke lengte nauwkeuriger benaderen, maar wees je ervan bewust dat de uitkomst een benadering blijft.



Figuur 11: schuifmaat met nonius.

Het is in deze gevallen eigenlijk beter om van meeton nauwkeurigheden te spreken dan van meetfouten, omdat degene die de meting uitvoert niets te verwijten valt. De onnauwkeurigheid die optreedt, was vooraf te voorspellen.

Elk meetinstrument waarmee je continue variabelen meet, brengt dergelijke onnauwkeurigheden met zich mee. Op veel glaswerk dat je bij scheikunde gebruikt, staan de mogelijke afwijkingen vermeld. Bij veel elektronische apparatuur, zoals een elektronische weegschaal, een

voltmeter of een ampèremeter, is in de handleiding aangegeven hoe groot de onnauwkeurigheid is.



Figuur 12: stopwatch.

8. Opdracht

Zoek uit hoe groot de meetonnauwkeurigheid is bij de volgende meetapparatuur die je op school gebruikt:

- een elektronische weegschaal
- een maatcilinder van 50 ml
- een maatcilinder van 10 ml
- een voltmeter
- een ampèremeter
- een buret van 25 ml
- een stopwatch.

Het optreden van andere toevallige fouten kunnen we toelichten aan de hand van een slinger waarvan je de slingerduur probeert te bepalen. Stel dat je met een chronometer timet wanneer de slinger in de hoogste positie aan de linkerzijde is. Elke maal dat de slinger in dat punt komt, druk je op de chronometer, die op 0,01 seconde nauwkeurig kan worden afgelezen. De menselijke onnauwkeurigheid bij het indrukken van de chronometer is echter veel groter dan die 0,01 seconde. Je mag al heel blij zijn als de onderlinge metingen niet meer dan 0,2 seconden van elkaar afwijken. Hier is dus de reactietijd de beperkende factor in de nauwkeurigheid. Tijdens de meting kunnen ook kleine schommelingen in de omstandigheden optreden. Misschien is er in de ruimte waarin gemeten wordt een kleine luchtstroom aanwezig, waardoor de slingertijd steeds iets varieert. Of misschien probeer je wel de stroomsterkte te meten van een kleine stroom die een klein beetje fluctueert. Dan zit de meetonnauwkeurigheid ook in het fysische verschijnsel zelf.

9. Vraag

Geef een aantal manieren waarop je de slingertijd van de slinger nauwkeuriger kunt bepalen dan op de hierboven beschreven wijze.



Figuur 13: toevallige fouten.

Kenmerk van toevallige fouten is dat ze even vaak een positieve als een negatieve afwijking laten zien ten opzichte van de werkelijke waarde (er vanuit gaande dat die werkelijke, constante waarde al bestaat...). Dit is te zien in figuur 13. Als je je voorstelt dat de cirkel een dartbord is en iemand heeft geprobeerd de bulls eye te raken, dan zie je hier de

toevallige fout. De werkelijke waarde is de bulls eye, de stippen zijn de 'gemeten' waarden.

Door het gemiddelde te nemen van veel metingen, is de toevallige fout behoorlijk te verkleinen, maar daarover later meer.

Systematische fouten

Anders dan toevallige fouten, treden **systematische fouten** categorisch op. Dat betekent dat er geen min of meer toevallig verschil in de meetwaarden te zien is, maar dat de meetwaarden stelselmatig te hoog of te laag uitvallen. Een mogelijke oorzaak hiervan ligt in het ijken van de meetapparatuur. Veel elektronische apparatuur moet om de zoveel tijd opnieuw geijkt worden, omdat de metingen anders teveel gaan afwijken van de werkelijke waarde. Wordt dit ijken niet goed of niet tijdig uitgevoerd, dan kan de meetapparatuur verkeerde waarden geven.

2. Bron: meer dan duizend verkeersboetes verscheurd.

DEN HAAG (GPD) - De verkeerspolitie moet honderden bonnen van geflitste automobilisten verscheuren. De agenten stonden afgelopen zaterdag en zondag op de A1 tussen Muiden en Diemen te controleren met radarapparatuur, die niet op tijd was geijkt. In totaal krijgen 1556 bestuurders geen boete thuis.

Dat bevestigt het Korps Landelijke Politie Diensten (KLPD). De radarapparatuur van de dienst moet jaarlijks worden geijkt. Als dat niet gebeurt, mag de flitsapparatuur niet worden gebruikt om snelheidsovertreders te bekeuren. Er is een fout gemaakt, stelt woordvoerder Ed Kraszewski van het KLPD. „De set was geijkt tot vrijdag 30 juni. Daar kwamen we pas na het weekeinde achter. De bonnen zijn inmiddels verscheurd.”

De fout werd bij toeval ontdekt door vrijwilligers van de website flitsservice.nl, die dagelijks door het land trekken om snelheidscontroles op de foto te zetten. De makers van de site vinden dat politie en justitie voortaan ijkrapporten openbaar moeten maken. „Veel mensen is een domper op een mooie dag bespaard gebleven door oplettendheid van onze fotograaf”, aldus Martijn Hurks. „Die maakte een foto van de apparatuur en zag op de sticker dat de ijkdatum was verlopen.

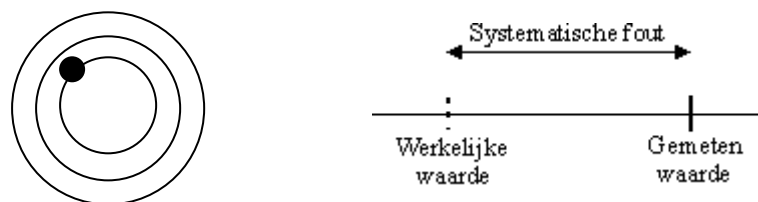
Bron: Noord Hollands Dagblad

10. Opdracht

Een voorbeeld van een meetinstrument op school dat regelmatig geijkt moet worden, is een pH-meter. Zoek uit (vraag bijvoorbeeld aan de TOA) hoe deze ijking wordt uitgevoerd.

Systematische fouten kunnen ook optreden door foutief handelen van degene die de meting uitvoert. Een bekend voorbeeld is het optreden van **parallaxfouten**, die ontstaan doordat de kijkrichting bij het aflezen van een meetinstrument niet loodrecht staat op de schaalverdeling. Ook kiezen voor een verkeerde meetmethode kan tot systematische afwijkingen leiden.

Alle systematische fouten hebben als overeenkomst dat de gemeten waarde stelselmatig of te hoog, of te laag is ten opzichte van de werkelijke waarde. Als je hier weer de vergelijking met een dartbord maakt dan heeft de werper weer geprobeerd het midden te raken (de werkelijke waarde). Deze persoon heeft echter een constante afwijking. Hij raakt steeds hetzelfde punt, maar dat is niet de bulls eye. Systematische fouten zijn hiermee goed te onderscheiden van toevallige fouten.



Figuur 14: systematische fouten.

11. Opdracht

- Je kunt je voorstellen dat toevallige fouten en systematische fouten in veel gevallen ook tegelijkertijd optreden. Geef aan hoe beide bovenstaande figuren er in dat geval uitzien.
- Verklaar waarom fouten die ontstaan bij het aflezen van niet digitale meetapparatuur, soms toevallige fouten zijn en soms systematische fouten.
- Wat voor soort fouten zijn opgetreden in de meting van de wereldwijde zeewatertemperaturen volgens het artikel uit de NRC in bron 1? Geef een toelichting.

Naast de al genoemde toevallige en systematische fouten, zijn er ook ‘domme’ fouten die door onhandigheid van de onderzoeker kunnen ontstaan. De gevolgen hiervan zijn soms enorm...

3. Bron: meetfout kost Nieuwegein 3 ton

zaterdag 26 juli 2008

NIEUWEGEIN - Een meetfout bij de bouw van basisschool De Toonladder in Nieuwegein kost de gemeente 3 ton.



Figuur 15: heipalen. ARCHIEFFOTO ANP

Pas toen vorige maand de heipalen waren geslagen, werd ontdekt dat er een meetfout van acht meter was gemaakt.

Daardoor zou de school op de verkeerde plek worden gebouwd. De fout ontstond doordat er geen goed kaartmateriaal voorhanden was.

Het heien moet nu opnieuw. De al geslagen palen worden boven de grond afgezaagd.

Bron: <http://www.ad.nl>

2.2 Toevallige fouten verkleinen

Door verschillende metingen uit te voeren, is de invloed van toevallige fouten sterk te verkleinen. Toevallige fouten hebben immers de eigenschap dat de meting de ene keer te hoog uitvalt, de andere keer te laag. Laat je vader of moeder maar eens met de auto op cruisecontrol (als je die hebt) met constante snelheid over de snelweg rijden. Als je met een stopwatch de tussentijd tussen twee hectometerpaaltjes probeert te meten, zullen de metingen een zekere spreiding vertonen. Als je er vanuit gaat dat de cruisecontrol heel precies werkt (en er dáárdor dus geen toevallige fouten ontstaan), zul je net zo vaak een iets te korte als een iets te lange tijd meten ten

$t(s)$	$(t - \bar{t})$	$(t - \bar{t})^2$
4,3	0,1	0,01
4,4	0,2	0,04
4,0	-0,2	0,04
4,1	-0,1	0,01
4,2	0	0
4,2	0	0
4,3	0,1	0,01
4,3	0,1	0,01
4,1	-0,1	0,01
3,9	-0,3	0,09
4,4	0,2	0,04
4,2	0	0
3,9	-0,3	0,09
4,3	0,1	0,01
4,4	0,2	0,04 +
	$\Sigma(t - \bar{t})^2$	0,40

Figuur 16: berekening standaardafwijking met:

t : uitkomst individuele tijdsmeting

$\bar{t}$: gemiddelde uitkomst tijdsmetingen (= 4,2 s).

opzichte van de werkelijke tijd. Deze afwijkingen zijn weer te geven als **standaardafwijking** of **standaarddeviatie** (σ). Waarschijnlijk heb je hier wel eens van gehoord.

In figuur 16 is het voorbeeld van de hectometerpaaltjes verder uitgewerkt en daar kun je ook zien hoe de standaardafwijking berekend is.

Voor de standaardafwijking geldt nu:

$$\sigma = \sqrt{\frac{\sum(t - \bar{t})^2}{n}} = \sqrt{\frac{0,40}{15}} \approx 0,163 \quad (1)$$

waarbij n het aantal meetwaarden is (in dit voorbeeld is dus $n = 15$).

De metingen voor tijdsduur tussen twee hectometerpaaltjes hebben dus een gemiddelde van 4,2 s met een standaardafwijking van ongeveer 0,163 s.

Reken je echter verder met het gemiddelde van de meetwaarden, dan mag je de standaardafwijking nog delen door de wortel van het aantal metingen. Doe je veel meer metingen, dan zal de standaardafwijking van een individuele meting immers niet veranderen, maar je bepaalt het gemiddelde wel steeds preciezer, omdat een individuele meting steeds minder invloed heeft. Daardoor neemt de standaardafwijking af.

Voor de standaardafwijking $\sigma_{\bar{t}}$ van het gemiddelde van de 15 meetwaarden geldt:

$$\sigma_{\bar{t}} = \frac{\sigma}{\sqrt{n}} \approx \frac{0,163}{\sqrt{15}} \approx 0,042 \quad (2)$$

Als je tien keer 15 waarnemingen doet en telkens het gemiddelde berekent, is de standaardafwijking van die tien gemiddelden dus 0,042. Herhaal je een meting maar vaak genoeg, dan is de toevallige fout hiermee sterk te verkleinen. Het herhalen van metingen is echter tijdrovend en soms ook kostbaar.

Jaar	Gem. temperatuur in °C
2008	6,5
2007	7,1
2006	1,5
2005	5,3
2004	3,6
2003	2,5
2002	4,4
2001	2,6
2000	4,3
1999	5,2

Figuur 17: gemiddelde temperatuur in januari gedurende de afgelopen tien jaar.

12. Opdracht

12.1 Meetwaarden cruisecontrol

De meetwaarden in formule (1) en (2) zijn verkregen met een auto waarvan de cruisecontrol stond ingesteld op 80 km/u. De uitkomsten van de metingen zijn hier niet helemaal mee in overeenstemming. Geef een mogelijke verklaring voor dit verschil.

12.2 Temperatuur

In de tabel van figuur 17 staan de gemiddelde temperaturen van de maand januari van de afgelopen 10 jaar.

Bereken de standaardafwijking.

	1981	1982
jan	338,4	339,2
feb	338,4	339,3
mrt	338,3	339,4
apr	338,3	339,2
mei	338,5	339,0
jun	338,2	339,1
jul	338,4	339,2
aug	339,0	340,0
sep	338,7	339,9
okt	338,8	339,7
nov	338,9	339,7
dec	339,1	339,6

Figuur 18: meetgegevens CO₂-concentratie in de atmosfeer in ppm.

12.3 CO₂-concentratie

In de tabel van figuur 18 staan de meetgegevens van de CO₂-concentratie in de atmosfeer op het eiland Amsterdam in de Indische Oceaan.

Bereken de standaardafwijking van de gemiddelde CO₂-concentratie voor de periode 1981-1982.

2.3 Meten van CO₂-concentraties

De gemiddelde januaritemperatuur in Nederland is over een lange periode 2,8 °C. Als je dat vergelijkt met je uitkomsten over de gemiddelde januaritemperatuur in opdracht 12.2 zou je kunnen concluderen dat het een stuk warmer is geworden in Nederland. Het versterkte broeikaseffect moet wel waar zijn!

Jammer genoeg is het niet zo gemakkelijk. Er zijn veel, heel veel meetgegevens nodig om die conclusie te kunnen trekken.

Wist je:.....

dat ppm 'parts per million betekent? Een CO₂-concentratie van 300 ppm betekent dat 300 van elke miljoen moleculen in de lucht, CO₂-moleculen zijn.

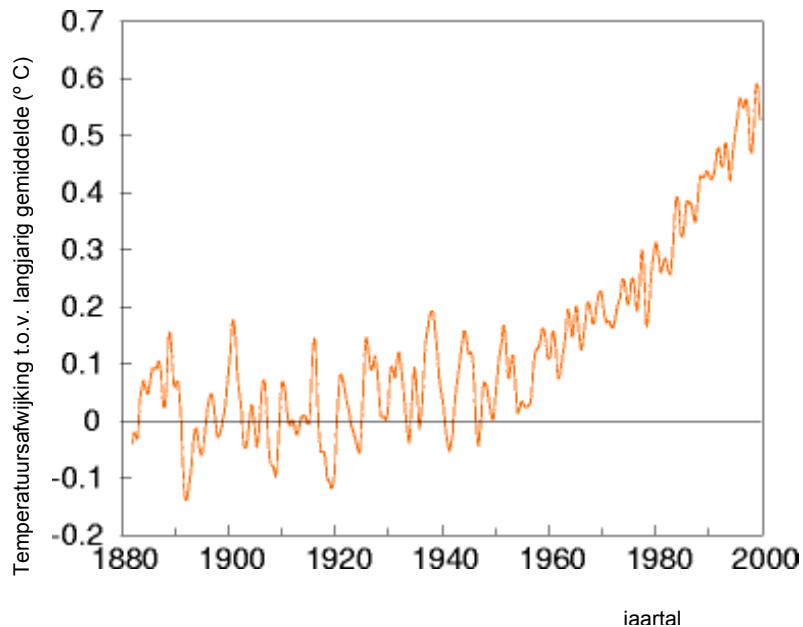
Met het oprichten van het school CO₂-net wordt een deel van de meetgegevens verzameld door middelbare scholen in Europa. Met behulp van een Vaisalameter wordt de koolstofdioxideconcentratie in de lucht gemeten (in ppm). De CO₂-waarden die gemeten worden, variëren per uur, per dag, per seconde en per jaar. Het is hierdoor moeilijk om goede gemiddelden te bepalen en daarmee te kijken of de hoeveelheid CO₂ in de atmosfeer echt stijgt. Er kan dus sprake zijn van meetfouten of meeton nauwkeurigheden.

13. Vraag

- Wat kan van invloed zijn op het variëren van de CO₂-waarden per uur of per dag? Licht je antwoord toe.
- Op welke manier kan je het verloop van de CO₂-concentratie in de atmosfeer op de langere termijn wel goed meten?

Maar waarom is het meten van de CO₂-concentratie zo belangrijk? Dat heeft te maken met het klimaat en klimaatveranderingen. De hoeveelheid CO₂ in de atmosfeer bepaald mede hoeveel warmte er wordt vastgehouden in de atmosfeer. Hoe meer CO₂, hoe meer warmte wordt vastgehouden. Dat wordt ook wel het broeikaseffect genoemd. De belangrijkste stoffen in onze atmosfeer die de warmte vasthouden zijn water (H₂O), methaan (CH₄), ozon (O₃), distikstofoxide (N₂O), stikstofoxide (NO₂) en dus koolstofdioxide (CO₂). Zonder het broeikaseffect zou het gemiddeld -18 °C zijn op aarde, met broeikaseffect is dat gemiddelde 18 °C.

Tot nu toe lijkt er geen probleem te zijn. Maar onder andere door menselijk handelen neemt de hoeveelheid CO_2 in de atmosfeer sterk toe. En daarmee zou de gemiddelde temperatuurstijging te verklaren moeten zijn. In figuur 19 zie je de afwijking van de gemiddelde temperatuur in de afgelopen 120 jaar.



Figuur 19: temperatuursverloop gedurende de afgelopen 120 jaar.

Voor welk deel van de temperatuurstijging de toename van CO_2 verantwoordelijk is, blijft onderwerp van discussie. Het belang van goede CO_2 -metingen is echter duidelijk.

2.4 Weergave van fouten

Meetfouten (veelal onnauwkeurigheden in metingen) kunnen we zowel absoluut als relatief weergeven. We zullen dit verder uitleggen aan de hand van een voorbeeld over de luchtkwaliteit in het schoollokaal.



Figuur 20: mufte schoollokalen.

Voorbeeld

De luchtkwaliteit in schoollokalen laat vaak te wensen over, doordat er te weinig ventilatie is. Je besluit een onderzoekje op school te doen waarbij je de CO_2 -concentratie in het klaslokaal meet en ook buiten de school. Je gebruikt hiervoor een CO_2 -meter met een afleesnauwkeurigheid van 1 ppm. In het klaslokaal geeft de meter 1417 ppm aan, buiten de school 392 ppm.

Allereerst kun je je afvragen wat de nauwkeurigheid van de CO_2 -meter precies inhoudt. Betekent de afleesnauwkeurigheid dat de werkelijke concentraties 1 ppm hoger of lager kunnen liggen?

Als de afleesnauwkeurigheid van een meetinstrument 1 ppm is, mag je er vanuit gaan dat de gevoeligheid van de meter hier ook minimaal mee overeenkomt. Het zou natuurlijk geen zin hebben als een CO₂-meter de concentratie op 1 ppm nauwkeurig geeft, terwijl de apparatuur niet nauwkeurig genoeg is om het verschil tussen 100 en 110 ppm te kunnen detecteren. Het is echter niet vanzelfsprekend dat de meter op dit moment ook echt zo nauwkeurig is. Onderhoud en ijking zijn van wezenlijk belang om schijnnaauwkeurigheid te voorkomen.

Ervan uitgaande dat de meter naar behoren functioneert, weten we nu dat de werkelijke CO₂-concentratie in het klaslokaal niet 1416 of 1418 ppm is. De werkelijke concentratie ligt dus ergens tussen 1416,5 en 1417,5 ppm. De **absolute fout** is gedefinieerd als de grootst mogelijke afwijking tussen de gemeten waarde en de werkelijke waarde en bedraagt hier dus 0,5 ppm.

In formuletaal betekent dit:

$$a - AF < A < a + AF \quad (3)$$

waarin:

- A = de werkelijke waarde
- a = de gemeten waarde
- AF = de absolute fout.

De gemeten concentratie noteer je als $1417 \pm 0,5$ ppm, maar je kunt ook kortweg 1417 ppm noteren. Uit de gangbare regels voor afronding blijkt immers dat er bij deze notatie een afwijking van maximaal 0,5 ppm kan zijn.

Je weet niet altijd hoe nauwkeurig een meetinstrument is. In dat geval werk je met een zo goed mogelijke schatting van de absolute fout.

Bij jouw metingen is de absolute fout zowel binnen als buiten 0,5 ppm. Maar in dit geval zegt de **relatieve fout** veel meer over de nauwkeurigheid dan de absolute fout.

Voor de relatieve fout geldt:

$$RF = \frac{AF}{A} \quad (4)$$

waarin:

- RF = de relatieve fout
- AF = de absolute fout
- A = de werkelijke waarde.

Uit praktisch oogpunt wordt echter vaak de volgende formule gehanteerd:

$$RF = \frac{AF}{a} \quad (5)$$

De uitkomst van formule (5) wijkt maar heel weinig af van de echte formule voor de relatieve fout in vergelijking (4).

14. Vraag

- Waarom wordt volgens jou in de praktijk altijd formule (5) voor de relatieve fout gebruikt?
- Bewijs dat de relatieve fout voor de meting in het lokaal ongeveer 0,00035 is en die van de meting buiten 0,0013.

Relatieve fouten worden meestal als percentage gegeven, dus CO₂-concentratie in het lokaal is 1417 ppm ± 0,035 % en de concentratie buiten 392 ppm ± 0,13 %.

Absoluut gezien is concentratie binnen en buiten even nauwkeurig bepaald, maar relatief is de concentratie binnen veel nauwkeuriger gemeten dan dat de concentratie buiten. Het is zaak om de te gebruiken meetinstrumenten af te stemmen op wat je wilt meten om te grote relatieve fouten te voorkomen.

Er zijn echter ook genoeg voorbeelden van zinloze pogingen om een variabele zeer nauwkeurig te bepalen. Zinloos, omdat de onzekerheid van een andere relevante variabele in het experiment vele malen groter is.

2.5 Rekenen met meetfouten

Nadat je een meting hebt verricht en de bijbehorende onzekerheid hebt bepaald, is het experiment meestal nog niet afgelopen. Met de gemeten waarde reken je verder om de waarde van andere variabelen te vinden. Maar hoe ga je dan om met de onzekerheid in de meetwaarde? Hiermee zul je in hoofdstuk 3 te maken krijgen, bij de verwerking van de experimenten.

Omdat de theorie over het doorrekenen met meetfouten nogal uitgebreid is, verwijzen we hiervoor naar bijlage 1. In de huidige paragraaf staan alleen de benodigde formules, zonder het bijbehorende bewijs. Je leert uiteraard wel hoe je de formules moet toepassen.

Stel je hebt twee meetwaarden:

waarde 1: $x \pm a$

waarde 2: $y \pm b$

Met a en b worden de absolute fouten van x , respectievelijk y bedoeld.

Wat gebeurt er met deze fouten wanneer je verder rekt met x en y ?

Optellen of aftrekken

We sommeren waarde 1 en waarde 2. De grootst mogelijke waarde die de uitkomst kan hebben, is:

$$x + a + y + b = x + y + a + b$$

Onthoud dat je bij optellen en aftrekken de absolute fouten dus moet optellen om de absolute fout in het antwoord te bepalen.

15. Vraag

We tellen de volgende (gewogen) gewichten bij elkaar op:

0,6 g ; 1,6 g ; 0,9 g ; 2,3 g ; 0,2 g ; 0,9 g

- Hoe groot is de absolute fout in elke meetwaarde?
- Voer de berekening uit en bereken de absolute fout in het antwoord.

Vermenigvuldigen met een constante

Stel je wilt meetwaarde 1 vermenigvuldigen met de constante c . Wat wordt nu de onzekerheid in de uitkomst?

De hoogst mogelijke uitkomst is:

$$c(x + a) = cx + ca$$

Onthoud dat je bij vermenigvuldigen met een constante de absolute fout met dezelfde constante vermenigvuldigt.

16. Vraag

Stel je hebt een meetlat van 100 cm. Hiermee bedoelen we dus een meetlat met $l = 100 \pm 0,5$ cm. De onnauwkeurigheid in de meetlat is niet het gevolg van problemen bij het aflezen (dat kan véél nauwkeuriger dan 0,5 cm), maar is het gevolg van een grove fabricatie. Voor de sportdag op school heb je een atletiekbaan van 60 m nodig. Deze atletiekbaan zet je met behulp van de meetlat uit. Afleesfouten laten we dus buiten beschouwing.

- Bereken de absolute fout in de lengte van de atletiekbaan.
- Hoe groot is deze fout volgens de normale regels voor significantie die je tot dusverre op school hebt geleerd? Is deze waarde reëel? Licht je antwoord toe.
- Als je twee even lange banen wilt uitzetten met dezelfde meetlat, zullen die op deze wijze veel in onderlinge lengte verschillen? Licht je antwoord toe.

Vermenigvuldigen

Stel je wilt meetwaarde 1 vermenigvuldigen met meetwaarde 2. Telde je bij optellen en aftrekken de beide absolute fouten bij elkaar op, nu moet je beide relatieve fouten optellen om de relatieve fout in het antwoord te bepalen:

$$\text{relatieve fout} = \frac{a}{x} + \frac{b}{y}$$

Wil je de absolute fout in het antwoord weten, dan is die vervolgens eenvoudig uit te rekenen.

Onthoud dat je bij vermenigvuldigen de relatieve fouten moet optellen om de relatieve fout in het antwoord te bepalen.

17. Vraag

Gegeven is een grasveld met $l = 23,0$ m en $b = 6,4$ m.

- Hoe groot zijn de absolute fouten in lengte en breedte?
- Bereken de oppervlakte met bijbehorende relatieve fout.
- Bereken de absolute fout in de oppervlakte.

Delen

Ook wanneer je twee meetwaarden door elkaar wilt delen, kun je beide relatieve fouten optellen om de relatieve fout in het antwoord te bepalen.

18. Vraag

Stel je hebt een blokje metaal waarvan je hebt bepaald dat het volume 8 cm^3 is (dus $8 \pm 0,5 \text{ cm}^3$) en waarvan je de massa hebt vastgesteld op 22 g (dus $22 \pm 0,5 \text{ g}$).

- Geef de formule waarmee je de dichtheid kunt berekenen.
- Bereken de dichtheid van het metaal.
- Bereken de relatieve fout in de dichtheid.
- Bereken ook de absolute fout in de dichtheid.

Onthoud dat je ook bij delen de relatieve fouten kunt optellen om de relatieve fout in het antwoord te bepalen.

Kwadrateren

19. Vraag

Stel er is een cirkel waarvan je de straal hebt gemeten: $2,3$ m. Om de oppervlakte te berekenen gebruik je de formule $O = \pi \cdot r^2$.

- Bedenk dat je de formule voor de oppervlakte kunt schrijven als $O = \pi \cdot r \cdot r$. Welke regel voor het doorrekenen met fouten mag je dus gebruiken?
- Hoe groot is de absolute fout in de straal?

- c. Bereken de oppervlakte van de cirkel met de bijbehorende relatieve fout.

Onthoud dat kwadrateren een bijzondere vorm van vermenigvuldigen is en dat daarvoor dus dezelfde regels gelden voor het doorrekenen met fouten.

3 Experimenten

Hoe voer je een meting goed uit? Wat verstaan we eigenlijk onder het goed uitvoeren van een meting? Je moet natuurlijk goed weten wat je meet en die meting dan ook nog met een zo klein mogelijke afwijking van de werkelijkheid doen.

Dat betekent nauwkeurig werken, geschikte meetapparatuur gebruiken en je bewust zijn van de onnauwkeurigheden die de meting met zich meebrengt. In dit hoofdstuk voer je een aantal experimenten uit die je misschien al wel kent. Maar als je de experimenten dit keer uitvoert, moet je steeds blijven letten op meetfouten en de theorie uit hoofdstuk 2 in de praktijk brengen. Kies de meest geschikte meetapparatuur, werk precies en noteer welke onnauwkeurigheid er volgens jou in jouw metingen zit.

Het hoofdstuk beschrijft vier experimenten. Afhankelijk van de beschikbare tijd en de beschikbare materialen geeft de docent aan welke proeven je wel doet en welke niet. De laatste twee experimenten in paragraaf 3.3 (valversnelling) en paragraaf 3.4 (veerconstante) lijken op elkaar, dus het is logisch een keuze voor één van beide te maken. Voer alles in tweetallen uit.

In hoofdstuk 2 heb je geleerd hoe je moet omgaan met meetfouten. Gebruik deze theorie om de vragen bij de experimenten te beantwoorden. Van elke proef maak je een verslag, waarin je kort omschrijft wat je hebt gedaan, waarin je de resultaten van de metingen geeft (mét onnauwkeurigheid) en waarin je antwoord geeft op de gestelde vragen. Je kunt hiervoor de ► werkinstructie practicumverslag in de NLT Toolbox gebruiken. De verslagen lever je per tweetal in bij de docent, die bij de beoordeling vooral zal kijken of je de theorie uit hoofdstuk 2 goed hebt toegepast.

3.1 Experiment dichtheid

Inleiding

Daar waar gemeten wordt, worden fouten gemaakt. Deze fouten zijn niet te voorkomen, maar je moet er wel rekening mee houden als je de waarden gaat gebruiken. Hoe je dat doet is al uitgebreid uitgelegd in hoofdstuk 2 (je hebt bij deze proef de formules nodig van paragraaf 2.5).

Je gaat werken met meetfouten door de dichtheid van verschillende materialen te meten. De dichtheid kan berekend worden met behulp van

de formule:

$$\rho = \frac{m}{V} \quad (6)$$

waarin:

- ρ = de dichtheid in kilogram per kubieke meter (kg/m^3)
- m = de massa in kilogram (kg)
- V = het volume in kubieke meter (m^3).

Benodigdheden

Houten blokje, aluminium cilindertje, digitale weegschaal, geodriehoek, schuifmaat.

Experiment 1

- Plaats het blokje op een digitale weegschaal en noteer de massa.
- Meet van het blokje, met behulp van een geodriehoek, de lengte, de breedte en de hoogte op. Bedenk hierbij dat metingen met een geodriehoek afgerond worden op een waarde van 0,001 m (= 1 mm).

20. Opdracht: houten blokje

- a. Noteer de massa van het blokje met de bijbehorende meetfout. Bereken de minimale en de maximale waarde van de massa.
- b. Noteer de lengte, breedte en hoogte van het blokje met de bijbehorende meetfout
- c. Bereken de minimale waarde voor het volume van het blokje.
- d. Bereken de maximale waarde voor het volume van het blokje.
- e. Bereken de minimale waarde voor de dichtheid van het hout. Bedenk hiertoe welke waarden (minimaal of maximaal) van de massa en het volume op elkaar gedeeld moeten worden.
- f. Bereken ook de maximale waarde voor de dichtheid van het hout.
- g. Noteer de uitkomst voor de dichtheid met de bijbehorende meetfout.

Experiment 2

- Plaats de cilinder op een digitale weegschaal en noteer de massa.
- Meet van de cilinder, met behulp van een schuifmaat, de diameter en de hoogte. Bedenk hierbij dat metingen met een schuifmaat afgerond worden op een waarde van 0,0001 m (= 0,1 mm).

21. Opdracht: aluminium cilinder

- a. Noteer de massa van de cilinder met de bijbehorende meetfout.
- b. Noteer de hoogte van het blokje met de bijbehorende meetfout.
- c. Noteer de diameter van de cilinder met de bijbehorende meetfout.
- d. Bereken de minimale waarde voor het volume van de cilinder.
- e. Bereken de maximale waarde voor het volume van de cilinder.

- f. Bereken de minimale waarde voor de dichtheid van het aluminium.
- g. Bereken ook de maximale waarde voor de dichtheid van het aluminium.
- h. Noteer de uitkomst voor de dichtheid met de bijbehorende meetfout.

3.2 Experiment dikte van aluminiumfolie



Figuur 21: rol aluminiumfolie.

Inleiding

Om een bepaalde grootte te meten, zijn vaak verschillende meetmethoden beschikbaar. Maar welke methode kies je? Dat is afhankelijk van zaken als: de gewenste nauwkeurigheid en de beschikbare meetapparatuur. In deze proef bepaal je de dikte van aluminiumfolie op drie verschillende manieren.

Een grootte die hierbij zal worden toegepast is de dichtheid. Deze kan berekend worden met behulp van de formule (6).

Benodigheden

Rol aluminiumfolie, micrometer, schuifmaat, geodriehoek.

Let in onderstaande experimenten en in de berekeningen steeds goed op de meetonnauwkeurigheden!

Experiment 1

Meet met een micrometerschroef of met een schuifmaat de dikte van het aluminiumfolie. Controleer goed de nulinstelling. Gebruik de slip bij de micrometer. Geef je antwoord zo nauwkeurig mogelijk en schrijf op wat volgens jou de onzekerheid in het antwoord is.

Experiment 2

Neem een stuk aluminiumfolie ter grootte van een vel A3. Vouw het vel dubbel en dubbel en dubbel totdat je niet verder kan. Het aluminiumfolie is dan dik genoeg om met een geodriehoek de dikte te meten. Meet de dikte van het dubbelgevouwen aluminiumfolie en noteer ook weer de onzekerheid in dit antwoord. Bereken hieruit de dikte van een enkel vel aluminiumfolie en de onzekerheid die hierbij hoort.

Experiment 3

Op de verpakking van een rol aluminiumfolie staat (onder meer) “20 m aluminiumfolie”. Het hele pak is gewogen toen het nog niet aangebroken was. De massa bleek 254,81 gram te zijn.

Geef commentaar op deze meting (gebruik in je antwoord de woorden bruto en netto). Voer eventueel nog extra metingen uit en maak hiermee een schatting van de dikte van het aluminiumfolie met de bijbehorende onzekerheid.

Maak een vergelijking van de uitkomsten van experiment 1, 2 en 3. Wat kun je zeggen over de nauwkeurigheid van de experimenten?

Experiment 4

Ontwerp een methode om veel nauwkeuriger dan in experiment 3 de dikte van het aluminiumfolie van een rol aluminiumfolie te bepalen zonder de rol aluminiumfolie geheel te demonteren. Voer de meting uit en bereken opnieuw de dikte van een enkel vel aluminiumfolie met de bijbehorende onzekerheid.

3.3 Experiment valversnelling

Inleiding

Omdat de periode van een eenvoudige slinger op een bekende manier afhangt van de valversnelling g kunnen we een proef met een slinger gebruiken om de waarde van de valversnelling te bepalen.

Om die bepaling nauwkeurig te kunnen uitvoeren, onderzoeken we eerst hoe de trillingstijd (= periode) T afhangt van:

- de gebruikte massa m ,
- de lengte van de slinger ℓ
- de beginhoek α .

Bij trillingen van een massa m , die hangt aan een touw met een lengte ℓ , blijkt (onder bepaalde voorwaarden) voor de trillingstijd T te gelden:

$$T = 2\pi\sqrt{\frac{\ell}{g}} \quad (7)$$

waarin:

- T = de trillingstijd in seconde (s)
- ℓ = de lengte van de slinger in meter (m)
- g = de valversnelling in meter per seconde kwadraat (m/s^2).

De trillingstijd blijkt dus onafhankelijk te zijn van de hoek waaronder je het gewichtje loslaat (α) en van de massa van het gebruikte gewichtje (m). Als je ℓ en g kent, is de trillingstijd te berekenen. Of andersom, als je de trillingstijd en ℓ kent, kun je de waarde van de valversnelling g berekenen. Dat is wat we in dit experiment gaan doen.

Benodigheden

Touw, gewichtjes, geodriehoek, digitale weegschaal en een statief.

Experiment

Ga steeds uit van een standaardsituatie (bijvoorbeeld $\ell = 1,00 \text{ m}$, $m = 50 \text{ g}$ en $\alpha = 20^\circ$).

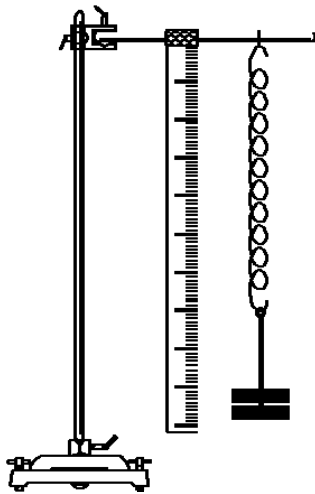
Ga in afzonderlijke proeven na wat de invloed is van een verandering van m , van ℓ en van α op de waarde van T .

Bepaal zo nauwkeurig mogelijk het verband tussen T en ℓ om daaruit de waarde van g af te leiden. Bepaal ook de onnauwkeurigheid in de waarde van g .

Hint: Maak gebruik van een diagram, waarin T^2 is uitgezet tegen de lengte ℓ .

Vergelijk de uitkomst met de waarde uit BINAS en geef commentaar.

3.4 Experiment veerconstante



Figuur 22: opstelling bepaling veerconstante.

Inleiding

De "veerconstante" van een veer kun je bepalen op een manier, die direct verband houdt met de definitie: je oefent (door middel van gewichtjes) een kracht uit en meet de uitrekking die daarvan het gevolg is.

Er bestaat ook een andere manier om de veerconstante te bepalen. Deze berust op de invloed van de veerconstante op de trillingstijd bij trillingen van een "**massa-veer systeem**".

Voor de uitrekking van een veer geldt, dat de veerkracht F_v evenredig is met de uitrekking u :

$$F_v = -C \cdot u \quad (8)$$

waarin:

- F_v = de veerkracht in Newton (N)
- C = de veerconstante in Newton per meter (N/m)
- u = de uitrekking in meter (m).

Voor trillingen van een massa m , die hangt aan een veer met veerconstante C , blijkt voor de trillingstijd T te gelden:

$$T = 2\pi \sqrt{\frac{m}{C}} \quad (9)$$

waarin:

- T = de trillingstijd in seconde (s)
- m = de massa in kilogram (kg)
- C = de veerconstante in Newton per meter (N/m).

Voor massa m moet je de massa invullen van het gewicht, dat aan de veer hangt, plus (een deel van) de massa van de veer. Daarom kunnen we m ook schrijven als $m_v + m$.

m_v is dan de massa van de veer, voor zover die invloed heeft op de trilling en m is de massa van het gewicht.

Ga na, dat we de uitdrukking voor de trillingstijd in formule (9) ook kunnen schrijven als:

$$T^2 = 4\pi^2 \frac{m + m_v}{C}$$

Benodigheden

Veer, statief, 5 verschillende gewichtjes, stopwatch, digitale weegschaal.

Experiment 1

Bepaal voor minstens vijf verschillende gewichten de uitrekking van een veer.

Vat de uitkomsten van de metingen samen in een diagram en bepaal de veerconstante met behulp van dat diagram (en laat zien hoe je dat doet). Probeer ook de onnauwkeurigheid in jouw antwoord aan te geven.

Experiment 2

Bepaal voor minstens vijf verschillende gewichten de trillingstijd T door voor 10 of 20 trillingen de tijd te meten (waarom doe je dat zó?).

Zet de resultaten uit in een diagram : verticaal T^2 en horizontaal m .

Bepaal de veerconstante met de bijbehorende onnauwkeurigheid uit de helling van de rechte lijn door de meetpunten.

Tip : de uitdrukking $T^2 = 4\pi^2 \frac{m}{C} + 4\pi^2 \frac{m_v}{C}$ lijkt op de formule $y = ax + b$.

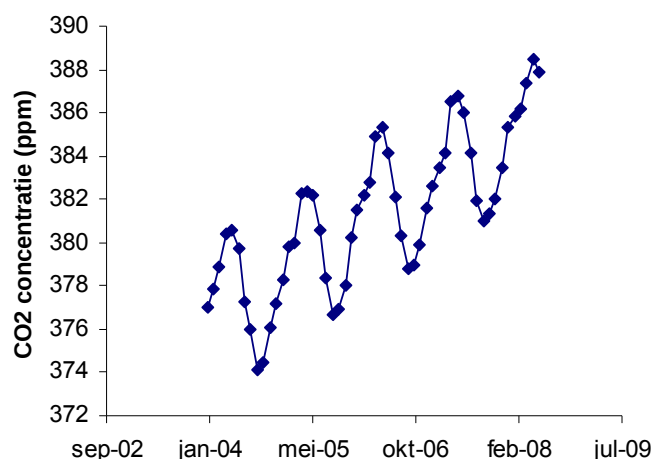
Vergelijk ten slotte de verschillende waarden voor de veerconstante C en geef je commentaar!

4 Interpreteren

Dit hoofdstuk gaat in z'n geheel over de verwerking van meetgegevens. Als je een mooie meetreeks hebt, wat mag je er dan wel uit concluderen en wat niet? Dit is een zeer uitgebreid onderwerp, waar boekenkasten vol wetenschappelijke literatuur over geschreven zijn. We zullen ons hier beperken en je enkele methodes leren die je goed kunt gebruiken bij jouw eigen onderzoek in hoofdstuk 5.

4.1 Seizoensinvloed op CO₂-concentratie

De CO₂-concentratie is de afgelopen eeuw gestegen. Maar hoe komen we aan die conclusie? Als we een reeks van een paar jaar metingen uitzetten, blijkt de concentratie nogal te variëren (zie figuur 23).



Figuur 23: verloop CO₂-concentratie.

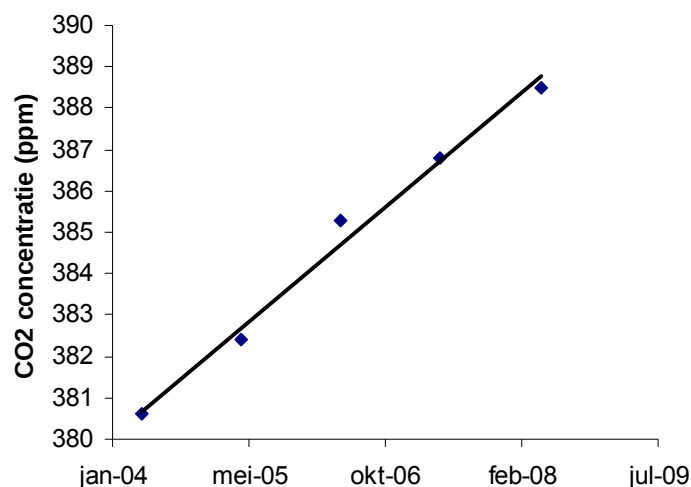
We zien wel dat de concentratie stijgt, maar het verloop is grillig. De gegevens zijn afkomstig van het noordelijk halfrond, waar de CO₂-concentratie aan het eind van de zomer het laagst is en aan het eind van de winter het hoogst.

Wist je:.....

dat de jaarlijkse variatie in de CO₂-concentratie wordt veroorzaakt door plantengroei? In het groeiseizoen neemt vegetatie netto CO₂ op, in de winter niet.

Als we willen kijken naar de meerjarige trend, dan is deze seizoensinvloed helemaal niet handig. We kunnen dan beter alleen kijken naar de gegevens van een bepaalde maand. In figuur 24 staat dezelfde grafiek nogmaals,

maar nu met alleen de metingen van de maand mei, wanneer de concentratie het hoogste punt in het jaar bereikt.



Figuur 24: gefilterde CO₂-concentratie.

Deze methode, waarbij we maar met een deel van de meetgegevens verder werken, wordt in deze module '**filteren van meetgegevens**' genoemd. Zoals je ziet, kan dit een veel beter beeld opleveren, maar je moet er erg mee uitkijken. Als je dat deel van de gegevens weglaat dat jou niet zo goed uitkomt, is het geen filteren, maar vervalsen! In paragraaf 4.2 gaan we verder in op het filteren van meetgegevens.

Wist je:.....

dat de CO₂-concentratie in werkelijkheid niet lineair stijgt? De afgelopen eeuw is de concentratie steeds sneller toegenomen. In ons voorbeeld hebben we te weinig meetwaarden om die trend te kunnen herkennen.

In figuur 24 is ook de lineaire trendlijn toegevoegd. Je ziet dat de 5 meetwaarden min of meer op deze rechte lijn liggen. Kun je daarmee voorspellingen doen over de toekomst? En wanneer mogen we eigenlijk concluderen dat meetwaarden op een rechte lijn liggen? Dit zijn de onderwerpen 'correlatie' en 'regressie' die we in de paragrafen 4.3 en 4.4 behandelen.

4.2 Combineren en filteren van meetgegevens

Combineren van meetgegevens kan leiden tot inzichten die niet uit de afzonderlijke meetreeksen zijn af te leiden. Vaak bevat één enkele meetreeks, die betrekking heeft op één variabele, te weinig informatie om de gewenste conclusies te kunnen trekken. Combineren we de gegevens over deze variabele met die van een andere variabele, dan zal dat tot



Figuur 25: strandtent.

verrassende inzichten kunnen leiden. We zullen hier zo dadelijk een voorbeeld van zien.

Filteren van meetgegevens is een lastiger verhaal. Met filteren wordt bedoeld dat je maar een deel van de meetgegevens gebruikt voor de analyse. Je kunt natuurlijk niet domweg de meetwaarden die jou wat minder goed uitkomen, weglaten. Je moet een goede reden hebben om meetgegevens weg te laten, anders leidt deze procedure al snel tot vervalsing van je meting. Dan zijn de meetwaarden altijd wel aan te passen aan de vooraf opgestelde hypothese en heeft het onderzoek weinig waarde meer.

Wat wel mag met combineren en filteren van meetgegevens, kunnen we het best illustreren aan de hand van een voorbeeld.

Voorbeeld

Een eigenaar van een strandtent heeft gedurende vier weken bijgehouden hoeveel drankjes hij dagelijks op het terras schenkt. Hij hoopt hiermee de hoeveelheid personeel goed te kunnen afstemmen op de vraag.

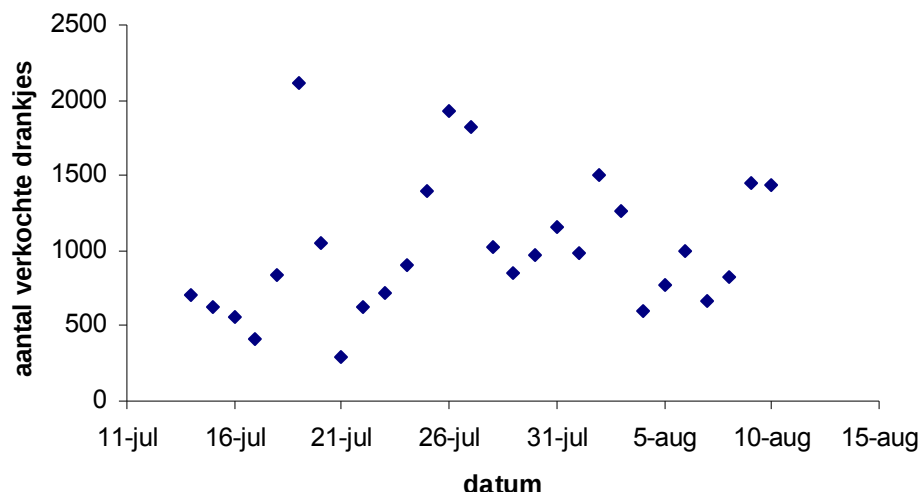
Personeelsleden die geen werk hebben, maar wel uitbetaald worden, kan hij zich eigenlijk niet veroorloven. De vier weken waarin de strandtenthouder zijn metingen uitvoert, vallen alle vier in de schoolvakanties.

Zijn meetwaarden staan in figuur 26.

datum	verkochte drankjes	datum	verkochte drankjes
14-07	708	28-07	1028
15-07	628	29-07	851
16-07	552	30-07	977
17-07	418	31-07	1155
18-07	834	01-08	983
19-07	2110	02-08	1505
20-07	1057	03-08	1266
21-07	295	04-08	602
22-07	627	05-08	777
23-07	718	06-08	993
24-07	908	07-08	663
25-07	1400	08-08	830
26-07	1933	09-08	1447
27-07	1828	10-08	1437

Figuur 26: drankverkoop strandtent.

Om een overzicht van de gegevens te krijgen, zet de strandtenthouder de waarden uit in een spreidingsdiagram (zie figuur 27).



Figuur 27: drankverkoop strandtent spreidingsdiagram.

Nog steeds kan de eigenaar van de strandtent weinig met deze gegevens. Hij weet dat hij per 400 verkochte drankjes per dag één personeelslid moet inplannen. Maar hij kan moeilijk elke dag rekening houden met het grootste aantal drankjes dat hij in de afgelopen 4 weken op één dag (2110) verkocht heeft.

De strandtenthouder vermoedt een verband tussen de temperatuur en het aantal verkochte drankjes. Als hij dat verband weet te vinden, kan hij aan de hand van de weersverwachting voor de komende dagen zijn personeel inplannen. Op de site van het KNMI zoekt hij de maximale dagtemperaturen van de afgelopen 4 weken op. Hij gaat dus de gegevens van de drankverkoop combineren met de gegevens van de temperatuur. De maximale dagtemperatuur voegt hij in de tabel in (zie figuur 28).

datum	max. temperatuur (°C)	verkochte drankjes	datum	max. temperatuur (°C)	Verkochte Drankjes
14-07	23	708	28-07	28	1028
15-07	21	628	29-07	25	851
16-07	20	552	30-07	27	977
17-07	18	418	31-07	30	1155
18-07	21	834	01-08	23	983
19-07	20	2110	02-08	24	1505
20-07	18	1057	03-08	21	1266
21-07	16	295	04-08	21	602
22-07	21	627	05-08	24	777
23-07	23	718	06-08	27	993
24-07	26	908	07-08	22	663
25-07	28	1400	08-08	21	830
26-07	29	1933	09-08	23	1447
27-07	28	1828	10-08	23	1437

Figuur 28: drankverkoop en maximale dagtemperatuur.

22. Opdracht

- Zet in een spreidingsdiagram het aantal verkochte drankjes uit tegen de temperatuur.
- Wijst het spreidingsdiagram op (lineair) verband?

Waarschijnlijk heb je nog geen erg duidelijk verband gevonden tussen de temperatuur en de drankverkoop. Dit is natuurlijk lastig voor de eigenaar van de strandtent, want nu kan hij nog steeds niet aan de hand van de weersverwachting zijn personeel inplannen.

Na enig piekeren beseft hij dat hij waarschijnlijk ook rekening moet houden met de dag van de week. Uit ervaring weet hij dat het in het weekend een stuk drukker is dan door de week, ook al is de temperatuur wat lager.

23. Opdracht

De eerste dag van de metingen, 14 juli, viel op een maandag. Markeer in het spreidingsdiagram uit opdracht 22 de meetwaarden die op een zaterdag of zondag vielen. Wat valt je op?

Als het goed is heb je bij opdracht 23 gezien dat vrijwel alle meetwaarden van zaterdag en zondag op een rechte lijn liggen. Dat betekent dat je door filtering van je meetgegevens (je laat immers de waarden van door de week weg) een steeds duidelijker verband ziet tussen de temperatuur en het aantal verkochte drankjes. Deze filtering is in sommige situaties te verdedigen, in andere niet. Als de strandtenthouder wil weten hoeveel personeelsleden hij in het weekend moet inplannen, dan is het logisch dat hij de metingen van maandag tot vrijdag weglaat. In dat geval is de filtering dus terecht. Maar wil hij weten of zijn zaak rendabel is, dan moet hij er wel rekening mee houden dat zijn omzet door de week een stuk lager is dan in het weekend. In dat geval zal hij van alle meetwaarden moeten uitgaan, anders zal hij van veel te hoge omzetten uitgaan en zichzelf ten onrechte rijk rekenen.

De eigenaar zit nog een beetje met de ene uitschieter in de meetwaarden van het weekend in zijn maag. Hij bekijkt de gegevens nog eens goed en ziet dat de uitschieter afkomstig is van een erg hoge omzet op zaterdag 19 juli. Hij pakt z'n agenda erbij en herinnert zich dan dat op die dag een grote beachparty op het strand werd georganiseerd. Het was die dag erg druk en ondanks de vrij lage temperatuur was de omzet in z'n zaak erg groot.

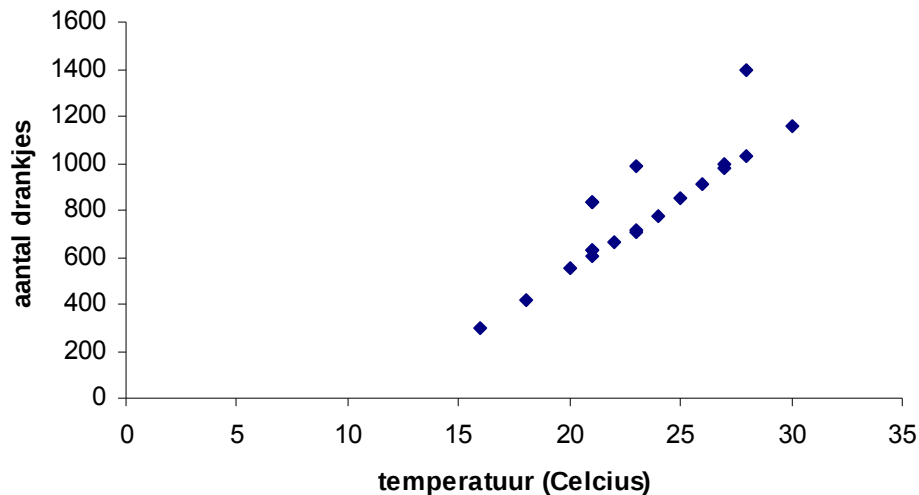
24. Vraag

- Is het gerechtvaardigd dat je de uitschieter van 19 juli wegstreept uit de meetgegevens als:
 - Je de benodigde inzet van personeel in het weekend wilt plannen?
 - Je een jaar later de drankinkoop voor de maand juli wilt regelen?

Licht je antwoorden toe.

- b. Hoeveel personeelsleden zou jij inplannen voor een zaterdag in begin augustus, wanneer de weersvoorspelling luidt dat het 25°C wordt? Reken weer op één personeelslid per 400 geschonken drankjes.

We houden dan nog de gegevens van de doordeweekse dagen over (zie figuur 29).



Figuur 29: drankverkoop op doordeweekse dagen.

Binnen deze gegevens lijkt er nog sprake te zijn van twee subgroepen. De gegevens lijken immers op twee verschillende rechte lijnen te liggen. Is er misschien sprake van nog een verantwoorde manier van filteren?

25. Vraag

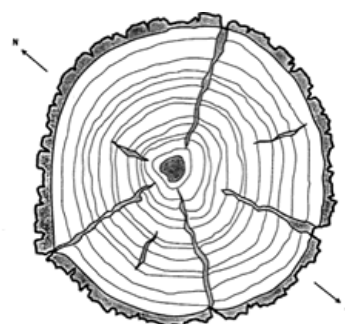
- a. Zoek uit welke meetwaarden op de bovenste rechte lijn lijken te liggen in figuur 22 (let op, dit zijn er vier!).
- b. Is er een reden waarom deze meetwaarden weggefilterd zouden mogen worden uit de volledige meetreeks? Zo ja, wat is die reden? Op 15 augustus 2008 (nog steeds in de vakantie) werd het liefst 30°C.
- c. Geef een schatting van het aantal drankjes dat op die dag in de strandtent verkocht is (bijzondere omstandigheden buiten beschouwing gelaten).

Je hebt nu geleerd dat combinatie van verschillende meetreeksen een serie gegevens waar in eerste instantie geen logica in lijkt te zitten, ineens erg logisch kan maken. Daarnaast is het van belang dat je een selectie durft te maken van de meetgegevens die je meeneemt in jouw onderzoek. Het selectie criterium mag echter nooit zijn dat sommige gegevens wel goed passen in het plaatje en andere niet. Weglaten van meetgegevens of verder werken met een deel van de meetgegevens mag alleen wanneer je een goede reden kunt geven waarom je dit doet!

4.3 Correlaties

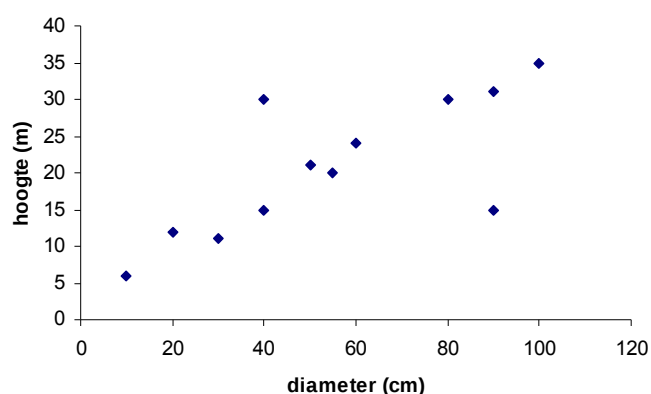
Er is sprake van een **correlatie** wanneer er een verband blijkt te zijn tussen twee gemeten grootheden. Stel dat in een dennenbos van een aantal bomen de hoogte en de stamdiameter op 1 m hoogte worden gemeten. De resultaten staan in figuur 30.

stamdiameter (cm)	hoogte (m)
10	6
20	12
30	11
40	15
40	30
50	21
55	20
60	24
80	30
90	15
90	31
100	35



Figuur 30: stamdiameter is vaak afhankelijk van de hoogte van een boom.

De stamdiameter en de hoogte zijn zogenaamde **'gepaarde' variabelen**. Dat betekent dat je niet zomaar van één van beide lijsten de volgorde kunt veranderen. De hoogte van 6 meter is immers gemeten bij de boom met een stamdiameter van 10 cm. Deze twee getallen zijn dus gepaard. In figuur 31 zijn deze gegevens uitgezet in een spreidingsdiagram.



Figuur 31: boomhoogte uitgezet tegen stamdiameter.

Je ziet in figuur 31 dat de bomen met een grote stamdiameter over het algemeen ook hoog zijn, terwijl de bomen met een kleine stamdiameter veelal een stuk minder hoog zijn. Met andere woorden: er is sprake van een correlatie tussen de stamdiameter en de hoogte van dennenbomen. In dit geval is sprake van een **positieve correlatie**: een grotere diameter betekent meestal ook een hogere boom. Ook is het een min of meer een rechte lijnige correlatie, omdat de punten allemaal dicht bij een denkbeeldige rechte lijn liggen.

Om de mate van correlatie aan te geven, wordt een **correlatiecoëfficiënt** berekend. Hiervan zijn verschillende varianten bekend, maar om die te begrijpen, moet je eerst weten wat covariantie betekent.

Bij een enkele variabele, zoals de lengte van leerlingen in een klas, wordt meestal de standaarddeviatie (de wortel van de variantie) gebruikt als spreidingsmaat. Is er, zoals in het voorbeeld van de stamdiameter, sprake van twee gekoppelde variabelen (de hoogte en de diameter), dan is de **covariantie** een goede spreidingsmaat.

Laten we weer kijken naar de dennenbomen. We noemen de diameter van de boom (in cm) x en de hoogte (in m) y . Voor de covariantie geldt:

$$\text{cov}(x, y) = \sum_{i=1}^n \frac{(x_i - \bar{x}) \cdot (y_i - \bar{y})}{n} \quad (10)$$

waarin:

- $\bar{x}$ = de gemiddelde stamdiameter in centimeter (cm)
- $\bar{y}$ = de gemiddelde hoogte in meter (m)
- n = het aantal meetwaarden.

Het punt met de coördinaten $(\bar{x}, \bar{y})$ wordt het **centrale punt** genoemd.

26. Vraag

- Bereken de gemiddelde stamdiameter en de gemiddelde hoogte van de dennenbomen uit het voorbeeld. Wat zijn dus de coördinaten van het centrale punt?
- Neem de tabel in figuur 32 over en vul deze verder in.

diameter x_i (cm)	hoogte y_i (m)	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$
10	6			
20	12			
30	11			
40	15			
40	30			
50	21			
55	20			
60	24			
80	30			
90	15			
90	31			
100	35			
totaal				+

Figuur 32: tabel voor berekening covariantie.

- Bereken met behulp van de tabel de covariantie.

Wist je:.....

dat je de berekeningen in een tabel, zoals bij opgave 26, heel gemakkelijk en snel in de grafische rekenmachine kunt uitvoeren? Kijk in bijlage 2 hoe dat werkt!

Je hebt bij opgave 26 waarschijnlijk gevonden dat de covariantie ongeveer 184 bedraagt. Dat deze waarde positief is, komt doordat de meeste uitkomsten van $(x_i - \bar{x}) \cdot (y_i - \bar{y})$ ook positief zijn. In opdracht 27 zul je zien dat je met dit gegeven een uitspraak kunt doen over het soort correlatie.

27. Opdracht

- Neem de grafiek in figuur 31 met daarin de hoogte en de diameter over.
- Zet een + bij de punten waarvan de uitkomst van $(x_i - \bar{x}) \cdot (y_i - \bar{y})$ positief is en zet een - bij de punten waarvan de uitkomst van $(x_i - \bar{x}) \cdot (y_i - \bar{y})$ negatief is.
- Wat valt je op? Waar liggen de punten met een + en waar de punten met een -? Kun je dit verklaren?
- In ons voorbeeld zijn er vooral punten waar je een + bij hebt gezet. Van wat voor correlatie is sprake?

In opdracht 27 heb je gezien dat een positief teken van de covariantie duidt op een positieve correlatie. Dat betekent dat we ook met de covariantie hebben bewezen dat voor de bomen uit ons voorbeeld geldt, dat een grotere diameter over het algemeen ook een grotere hoogte betekent.

Maar we willen graag meer zeggen over het verband tussen diameter en hoogte. Is het een sterk verband? Is het zinvol om aan de hand van dit verband een uitspraak te doen over de waarschijnlijke hoogte van een boom met een diameter van 70 cm? En over de hoogte van een boom met een diameter van 120 cm?

We weten al dat het teken van de covariantie (positief of negatief) aangeeft of er sprake is van positieve of negatieve correlatie. Geeft de absolute waarde van de covariantie dan misschien de mate van correlatie aan?

We bekijken nogmaals het voorbeeld van de bomen. Stel dat de diameter niet in centimeters is uitgedrukt, maar net als de hoogte in meters.

28. Vraag

- Bereken nogmaals de covariantie van de bomen uit ons voorbeeld, maar ga nu uit van de diameter die is uitgedrukt in meters.

- b. Zegt de absolute waarde van de covariantie iets over de mate van correlatie?

Uit vraag 28 is gebleken dat je met de covariantie nog geen uitspraak kunt doen over de mate van correlatie tussen twee variabelen. Daarvoor moet de covariantie eerst gestandaardiseerd worden:

$$R(x, y) = \frac{\text{cov}(x, y)}{\sigma_x \cdot \sigma_y} \quad (11)$$

waarin:

- $R(x, y)$ = de Pearson correlatiecoëfficiënt
- σ_x = de standaardafwijking van x
- σ_y = de standaardafwijking van y .

$R(x, y)$ wordt ook wel de **product-momentcorrelatiecoëfficiënt (pmcc)** genoemd. Door te delen door het product van de standaarddeviaties van x en y geldt altijd:

$$-1 \leq R(x, y) \leq 1$$

Liggen alle gemeten punten in een spreidingsdiagram op een rechte, stijgende lijn, dan geldt $R(x, y) = 1$ en is er sprake van een volledige, positieve correlatie. Liggen alle punten op een dalende rechte lijn, dan geldt $R(x, y) = -1$ en dan is er een volledige, negatieve correlatie tussen beide gemeten grootheden.

29. Vraag

We gaan weer uit van het voorbeeld van de dennenbomen. We nemen weer de diameter in centimeters en de hoogte in meters.

- Bereken de Pearson correlatiecoëfficiënt.
Ga nu uit van de diameter in meters.
- Bereken ook nu de Pearson correlatiecoëfficiënt.
- Wat valt je op?
- Is er volgens jou, op basis van de correlatiecoëfficiënt, sprake van een sterke correlatie tussen boomdiameter en boomhoogte?

We hebben nu een methode geleerd om te bepalen in hoeverre er sprake is van een lineaire samenhang tussen twee gepaarde variabelen x en y . Samengevat voer je daarvoor de volgende stappen uit:

1. Bereken $\bar{x}$ en $\bar{y}$. Het punt $(\bar{x}, \bar{y})$ noemen we het centrale punt.
2. Bereken de covariantie: $\text{cov}(x, y)$
3. Bereken de standaardafwijking van beide variabelen: σ_x en σ_y .
4. Bereken de Pearson correlatiecoëfficiënt.

De uitkomst van de Pearson correlatiecoëfficiënt leert ons in hoeverre er sprake is van een lineair verband.

Omdat we van beide gepaarde variabelen gemeten waarden moeten hebben, dienen beide minimaal op intervalschaal gemeten te zijn.

30. Opdracht

30.1 Pearson correlatiecoëfficiënt

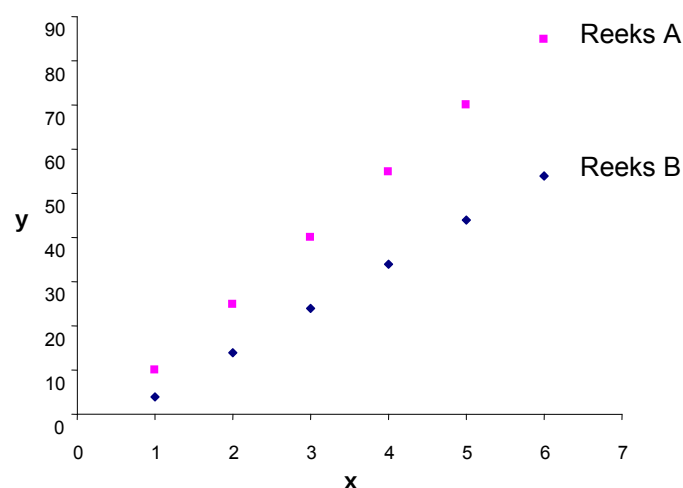
Tussen de variabele 'hoe vaak eet iemand vis per jaar' en de variabele 'hoe vaak eet iemand worteltjes per jaar' wordt als Pearson correlatiecoëfficiënt gevonden $R(x,y) = 0,00$.

Welke bewering is waar? Licht je antwoord toe!

- A. In de steekproef is zeker geen relatie tussen beide variabelen.
- B. In de steekproef is waarschijnlijk geen relatie tussen beide variabelen.
- C. Er is aangetoond dat beide variabelen onafhankelijk zijn.
- D. Er is sprake van een perfecte correlatie, die positief noch negatief is.

30.2 Spreidingsdiagram

Bekijk het spreidingsdiagram in figuur 33.



Figuur 33: spreidingsdiagram.

Welke bewering is waar? Licht je antwoord toe!

- A. Reeks A: $R(x, y) = 1$ en reeks B: $R(x, y) < 1$
- B. Reeks A: $R(x, y) < 1$ en reeks B: $R(x, y) < 1$
- C. Reeks A: $R(x, y) > 1$ en reeks B: $R(x, y) = 1$
- D. Reeks A: $R(x, y) = 1$ en reeks B: $R(x, y) = 1$
- E. Reeks A: $R(x, y) > 1$ en reeks B: $R(x, y) < 1$

30.3 Avondmaaltijd

Er wordt onderzoek gedaan naar het gewicht van een proefpersoon en de gemiddelde grootte van de avondmaaltijd die de persoon tot zich neemt. Aan zes mensen wordt gevraagd om een week lang het gewicht van hun avondmaaltijd bij te houden. Per proefpersoon wordt hieruit het

gemiddelde gewicht van de avondmaaltijd berekend. Daarna worden de proefpersonen gewogen. Dit levert de gegevens in figuur 34 op.

Gemiddeld gewicht avondmaaltijd (g)	Gewicht proefpersoon (kg)
850	83
720	65
600	53
850	88
1000	95
700	70

Figuur 34: gegevens gewicht avondmaaltijd en gewicht proefpersoon.

- Zet de gegevens uit in een spreidingsdiagram met het gewicht van de proefpersonen op de horizontale as.
- Bereken met de Pearson correlatiecoëfficiënt in hoeverre er sprake is van een lineaire correlatie tussen het gewicht van de proefpersonen en het gemiddelde gewicht van hun avondmaaltijden.
- Heb je met de berekening bewezen dat mensen die meer eten dikker worden?

30.4 Koffie en thee

Een onderzoeker vraagt zich af, of er verband bestaat tussen het aantal koppen koffie en het aantal koppen thee dat dagelijks in een huishouden wordt gedronken. Hij doet onderzoek naar het verbruik in 13 huishoudens. Dit levert de gegevens in figuur 35 op.

koffiegebruik x	3	6	2	4	5	3	4	2	6	2	4	5	6
theegebruik y	6	2	4	6	5	3	2	6	4	5	4	2	3

Figuur 35: gegevens koffie- en theegebruik.

- Teken het bijbehorende spreidingsdiagram met het aantal koppen koffie op de horizontale as.
- Van wat voor correlatie lijkt er sprake te zijn? Is dat logisch te verklaren?
- Bereken de correlatiecoëfficiënt $R(x, y)$.

30.5 Stroomkring

Op een weerstand wordt een variabele spanningsbron aangebracht. In de stroomkring wordt in serie geschakeld een stroommeter opgenomen. Bij verschillende spanningen wordt vervolgens de stroomsterkte gemeten. De uitkomsten staan in de tabel van figuur 36.

spanning x (V)	5	20	30	50	70	90	100
stroomsterkte y (A)	0,26	1,07	1,58	2,61	3,68	4,75	5,26

Figuur 36: gegevens spanning en stroomsterkte.

- Bepaal aan de hand van de Pearson correlatiecoëfficiënt of er sprake is van correlatie.
- Verklaar het gevonden resultaat met een formule uit de Binas.
- Welke correlatiecoëfficiënt verwacht je eigenlijk te vinden? Hoe komt het dat jouw antwoord daar iets van afwijkt?

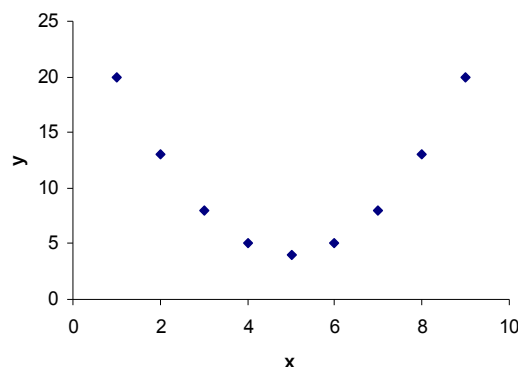
Tot nu toe hebben we alleen naar rechtlijnige correlaties gekeken. Uit het hierop volgende voorbeeld zal echter blijken dat er ook andere correlaties zijn. De Pearson correlatiecoëfficiënt is dan niet te gebruiken.

Stel er zijn twee gepaarde variabelen, x en y . De waarden staan in figuur 37.

x	1	2	3	4	5	6	7	8	9
y	20	13	8	5	4	5	8	13	20

Figuur 37: meetwaarden van de gepaarde variabelen x en y .

In figuur 38 zijn de variabelen x en y uitgezet in een spreidingsdiagram.



Figuur 38: spreidingsdiagram van de gepaarde variabelen x en y .

Er lijkt wel degelijk sprake te zijn van correlatie (je herkent hierin immers een kwadratisch verband), maar rechtlijnig is het niet.

31. Vraag

- Bereken voor dit voorbeeld de Pearson correlatiecoëfficiënt
- Had je deze uitkomst kunnen voorspellen?

Bij vraag 31 heb je aangetoond dat er geen sprake is van rechtlijnige correlatie. Toch zie je in figuur 38 duidelijk dat er sprake is van een zeker verband tussen x en y . Misschien had je zelfs al het kwadratisch verband

$$y = (x - 5)^2 + 4 \text{ herkend.}$$

De product-momentcorrelatiecoëfficiënt voldoet in dit geval niet en dus is er behoefte aan een andere methode om de mate van een niet rechtlijnige correlatie te beschrijven.

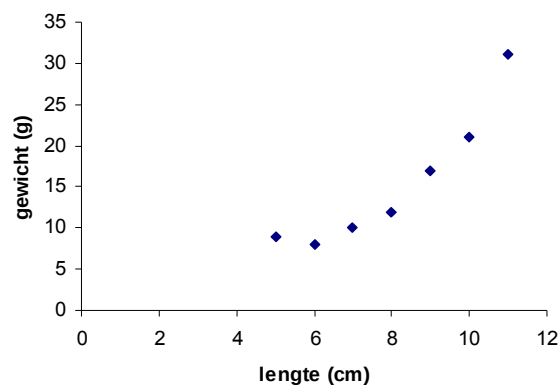
De **rangcorrelatiecoëfficiënt** biedt uitkomst. Zoals we straks zullen zien, kun je deze coëfficiënt gebruiken om een monotone, niet lineaire correlatie aan te tonen. Met monotoon wordt bedoeld dat bij groter wordende x de y ofwel steeds groter, ofwel steeds kleiner wordt, maar dat er niet noodzakelijk een rechtlijnig verband is.

We gaan uit van een voorbeeld waarbij Jessica een aquarium heeft met daarin 7 vissen. Op een dag meet zij van alle vissen de lengte en het gewicht. De resultaten staan in figuur 39.

lengte x (cm)	7	10	6	5	9	11	8
gewicht y (g)	10	21	8	9	17	31	12

Figuur 39: lengte en gewicht van de vissen in het aquarium van Jessica.

In figuur 40 zijn de gegevens uitgezet in een spreidingsdiagram.



Figuur 40: spreidingsdiagram van lengte en gewicht van de vissen in het aquarium.

Er is geen sprake van een duidelijk rechtlijnig verband. We zullen de rangcorrelatiecoëfficiënt gebruiken om het kromlijnige verband aan te tonen. Allereerst kennen we aan elke lengte en aan elk gewicht dat Jessica heeft gemeten een zogenaamd rangnummer toe. De kleinste lengte krijgt rangnummer 1, evenals het laagste gewicht. De resultaten staan in figuur 41.

lengte x_i (cm)	gewicht y_i (g)	rangnr. x_i	rangnr. y_i
7	10	3	3
10	21	6	6
6	8	2	1
5	9	1	2

9	17	5	5
11	31	7	7
8	12	4	4

Figuur 41: rangnummers van lengte en gewicht van de vissen in het aquarium.
 Merk op, dat een tabel als in figuur 41 al is op te stellen voor variabelen die op ordinaal niveau zijn gemeten. Dit in tegenstelling tot de meetwaarden die je nodig hebt om de Pearson correlatiecoëfficiënt te berekenen. Die moeten immers minimaal op intervalschaal zijn gemeten.

Nu je de rangnummers hebt, kun je kiezen uit twee mogelijkheden:

1. Je berekent met de rangnummers alsnog de Pearson correlatiecoëfficiënt. Door rangnummers te gebruiken kun je nu ook niet rechtlijnige correlaties met deze coëfficiënt aantonen.
2. Je berekent met de rangnummers de Spearman rangcorrelatiecoëfficiënt. Deze wordt gebruikt om monotone, niet rechtlijnige correlaties aan te tonen.

In de verdere uitwerking zul je zien dat beide mogelijkheden vaak leiden tot precies dezelfde correlatiecoëfficiënt.

32. Vraag

- a. Bereken op basis van de rangnummers de pmcc.
- b. Wat is jouw conclusie, uitgaande van de uitkomst bij a?

Voor de **Spearman correlatiecoëfficiënt** geldt de volgende formule:

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (12)$$

waarin:

- r_s = de Spearman rangcorrelatiecoëfficiënt (Spearman's rho)
- d_i = het verschil in rangnummer van twee gepaarde variabelen
- n = het aantal meetwaarden.

De interpretatie van de Spearman correlatiecoëfficiënt is hetzelfde als de Pearson correlatiecoëfficiënt: een waarde van 1 betekent een volledige, positieve correlatie en een waarde van -1 duidt op een volledige negatieve correlatie. Is de uitkomst 0, dan is er geen correlatie bewezen.

Met de tabel in figuur 42 kan Spearmans rho worden berekend.

lengte x_i (cm)	gewicht y_i (g)	rangnr. x_i	rangnr. y_i	d_i	d_i^2
7	10	3	3	0	0
10	21	6	6	0	0
6	8	2	1	1	1
5	9	1	2	1	1
9	17	5	5	0	0
11	31	7	7	0	0
8	12	4	4	0	0
totaal					2

Figuur 42: tabel waarmee Spearmans rho kan worden berekend.

De berekening gaat nu als volgt:

$$\begin{aligned}
 r_s &= 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \\
 &= 1 - \frac{12}{7(7^2 - 1)} \approx 0,96
 \end{aligned}$$

Wist je:.....

dat er sprake is van een 'knoop' wanneer er twee vissen zijn met dezelfde lengte? Als dat de lengtes zijn waaraan we rangnummer 3 en 4 willen toekennen, lossen we dat op door beide rangnummer 3,5 te geven.

Deze waarde is exact gelijk aan de pmcc over de rangnummers die je in vraag 32 hebt berekend. Spearmans rho is minder bewerkelijk om te berekenen en heeft dus de voorkeur.

33. Vraag

In het aquarium van Jessica wordt een nieuwe vis met een lengte van 10 cm en een gewicht van 22 g losgelaten.

- Bereken voor de 8 vissen die zich nu in het aquarium bevinden, de pmcc over de rangnummers.
- Bereken voor de 8 vissen ook Spearmans rho.
- Wat valt je op?

Wist je:.....

dat de pmcc over de rangnummers precies dezelfde uitkomst heeft als Spearmans rho? Maar dat dat alleen zo is wanneer er geen knopen zijn? Anders is er wel een (klein) verschil.

Mits het aantal knopen niet te groot is, hebben we nu ook een goede methode om een ander monotoon verband aan te tonen. Er zijn situaties waarbij je beter voor de rangcorrelatiecoëfficiënt kunt kiezen dan voor de Pearson correlatiecoëfficiënt. Zulke situaties zijn bijvoorbeeld:

- Je vermoedt dat het een niet-rechthoekig, monotoon verband betreft.
- De variabelen zijn niet beide op minimaal intervalschalen gemeten.
- Er is sprake van een paar sterke uitschieters in je meetwaarden. Die uitschieters hebben een erg grote invloed op de pmcc, wat meestal niet terecht is. De invloed van zulke uitschieters op de rangcorrelatiecoëfficiënt is veel kleiner.

34. Vraag

Ga uit van de meetwaarden in figuur 43.

x_i	1	2	3	4	5	6	7
y_i	1	2	3	15	5	6	7

Figuur 43: meetwaarden x_i en y_i .

- Bereken de Pearson correlatiecoëfficiënt.
- Bereken Spearmans rho.
- Welk van beide correlatiecoëfficiënten geeft volgens jou in deze situatie het beste beeld en waarom?

Nu je in staat bent om zowel een rechthoekig verband als een ander monotoon verband aan te tonen, is het van belang dat je je bewust wordt van de valkuilen bij het aantonen van correlaties.

Een correlatie staat niet gelijk aan een oorzakelijk verband. Vaak is er een derde, onderliggende oorzaak.

Vreemd??!!

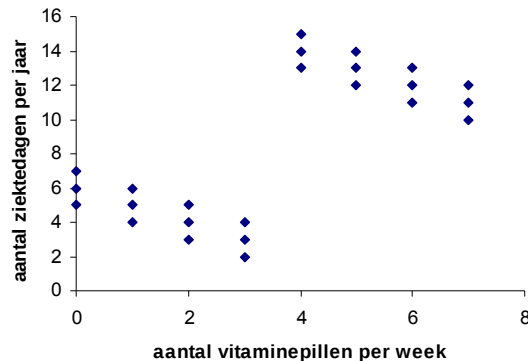
De afgelopen 50 jaar is het aantal auto's in Nederland snel toegenomen en het aantal weidevogels snel afgenomen. Dus weidevogels hebben een hekel aan uitlaatgassen of weidevogels zijn in het verleden erg schadelijk geweest voor de auto's.....

35. Vraag

Kun jij wel een logische verklaring bedenken voor het voorbeeld van de weidevogels en de auto's? Wat is de onderliggende oorzaak?

Goed onderzoek doen naar correlaties is lastig. Je moet je metingen zo uitvoeren, dat er geen storende invloed is van andere variabelen. In de praktijk is dat moeilijk, want veel dingen houden verband met elkaar.

Het is tevens van groot belang dat je goed nadenkt over je onderzoeksgroep. Stel dat je onderzoek doet naar het verband tussen de inname van vitaminepillen en het aantal dagen dat iemand per jaar ziek is. De meetwaarden staan in figuur 44.



Figuur 44: verband tussen inname vitaminepillen en het aantal ziekte-dagen.



Figuur 45: vitaminepillen.

Zou je met deze gegevens een correlatiecoëfficiënt uitrekenen, dan zou je een (weliswaar zwakke) positieve correlatie vinden. Dat zou de suggestie wekken dat de inname van vitaminepillen het aantal ziekte-dagen per jaar verhoogt. Dat is natuurlijk vreemd. Er zijn twee mogelijke verklaringen voor het optreden van dit fenomeen.

Ten eerste heb je net geleerd dat een correlatie nog geen causaal verband betekent. Als het al een causaal verband betreft, is er nog het beroemde 'de kip en het ei' probleem: wat is de oorzaak en wat het gevolg? Word je ziek van het innemen van vitaminepillen? Of zijn mensen die vaak ziek zijn sneller geneigd vitaminepillen te slikken?

Ten tweede is er in het spreidingsdiagram duidelijk sprake van twee subgroepen. Stel dat je het onderzoek hebt uitgevoerd bij jouw klasgenoten en in de verzorgingsflat waar je oma woont. Als je de gegevens beter bestudeert, zie je dat de punten linksonder in het diagram van figuur 44 afkomstig zijn van jouw klasgenoten en de punten rechtsboven van de kennissen van je oma. Er is dus sprake van twee subgroepen. Binnen een subgroep zie je wel de negatieve correlatie die je hoopte te vinden: mensen die meer vitaminepillen slikken, zijn minder ziek.

Blijkbaar zijn oudere mensen sneller geneigd vitaminepillen in te nemen. Maar omdat juist zij vaak een zwakke gezondheid hebben, geeft dat een erg vertekend beeld wanneer je de groepen niet afzonderlijk bekijkt.

Tot slot moet je uitkijken dat je jouw conclusies niet teveel veralgemeniseert: heb je in jouw klas een zekere correlatie gevonden tussen koffie- en theeconsumptie, dan hoeft die correlatie nog niet te gelden voor alle Nederlandse leerlingen. Daarvoor is jouw steekproef niet representatief.

36. Opdracht

Bedenk zelf een kort onderzoekje dat je in jouw klas gaat uitvoeren en waarmee je een correlatie hoopt te vinden. Zet de gegevens die je gemeten hebt uit in een spreidingsdiagram en bepaal welke correlatiecoëfficiënt je gaat berekenen. Is er inderdaad sprake van een correlatie?

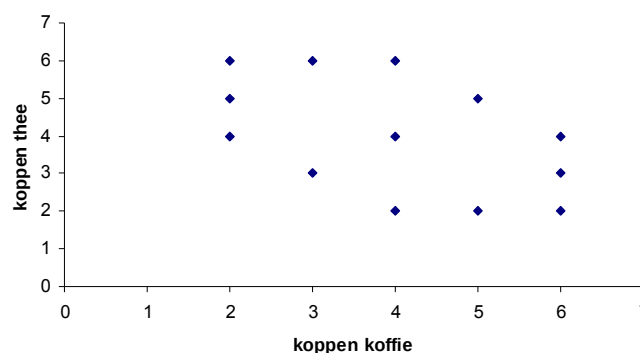
4.4 Lineaire regressie

In paragraaf 4.3 heb je technieken geleerd om na te gaan of er een verband bestaat tussen twee variabelen. Dat noemen we **correlatieanalyse**. Het logische vervolg hierop is **regressie**: kun je, als je eenmaal een verband hebt aangetoond, ook voorspellingen doen? Kun je, als je een verband hebt gevonden tussen de dikte van dennenbomen en hun hoogte, ook voorspellen hoe hoog een dennenboom met een diameter van 70 cm is? Als dat mogelijk is, zou je ook voorspellingen kunnen doen over situaties die nu niet te meten zijn.

Denk eens terug aan opdracht 30.5 over de stroomkring met daarin een variabele spanningsbron, een weerstand en een stroommeter. Als de spanningsbron een maximale uitgangsspanning heeft van 100 V, dan kun je niet meten hoe groot de stroomsterkte is bij een spanning van 120 V. Maar omdat je zo'n duidelijke, rechtlijnige correlatie hebt gevonden, kun je dat wel voorspellen. Als je dat doet, dan houd je je bezig met regressie.

Je hebt gezien dat er rechtlijnige correlaties zijn, maar ook niet-rechtlijnige. Handmatige regressie van niet-rechtlijnige correlaties is erg moeilijk. Meestal worden hiervoor computerprogramma's gebruikt. In deze module zullen we ons beperken tot regressie van min of meer rechtlijnige verbanden.

Bij lineaire regressie zoeken we een rechte lijn, die zo goed mogelijk aansluit bij alle meetwaarden. Laten we eens uitgaan van de gegevens uit opdracht 30.4, waarbij onderzoek was gedaan naar koffie- en theegebruik. In figuur 46 zijn deze gegevens weergegeven in een spreidingsdiagram.



Figuur 46: verband tussen koffie- en theeconsumptie.

Duidelijk is, dat als we in figuur 46 een rechte tekenen die zo goed mogelijk aansluit bij alle punten, dit een dalende lijn zal zijn. Dat hebben we bij de beantwoording van opdracht 30.4 ook al gezien: de Pearson correlatiecoëfficiënt was ongeveer -0,5, wat duidt op een vrij zwakke negatieve correlatie.

De best passende lijn wordt ook wel **regressielijn** genoemd. Maar hoe loopt deze rechte precies? We gaan hierbij uit van het kleinste-kwadratenprincipe.

De regressielijn gaat door het centrale punt $(\bar{x}, \bar{y})$. In ons voorbeeld is dat het punt (4,4). Hoe de helling van de lijn ook gekozen wordt, niet alle punten zullen op de lijn liggen. Sommige punten zullen een stukje boven de lijn liggen, andere punten een stukje eronder. Deze afstand van een willekeurig punt (x_i, y_i) tot de regressielijn noemen we het **residu** en dit geven we aan met d_i . Er geldt:

$$d_i = y_i - y_r \quad (13)$$

waarin:

- d_i = het residu van punt i tot aan de regressielijn
- y_i = de y -waarde van punt i
- y_r = de y -waarde van de regressielijn.

Wist je dat.....

residu een term is die ook in de scheikunde wordt gebruikt? Dan wordt er de stof mee bedoeld die na filtreren achterblijft in het filter.

Volgens de **kleinste-kwadratenmethode** moeten we de regressielijn zó tekenen dat de som van de kwadraten van alle residuen zo klein mogelijk is:

$$\sum_{i=1}^{i=k} d_i^2 \quad \text{zo klein mogelijk.}$$

Hierbij wordt gebruik gemaakt van de kwadraten van de residuen, omdat voor de afwijking ten opzichte van de regressielijn natuurlijk niet uitmaakt of een punt onder of boven de regressielijn ligt.

De algemene vergelijking van elke rechte, dus ook van de regressielijn, luidt:

$$y = ax + b \quad (14)$$

waarbij a de richtingscoëfficiënt is en b het startgetal. Om te voldoen aan de kleinste kwadratenmethode nemen we als richtingscoëfficiënt:

$$a = R \cdot \frac{\sigma_y}{\sigma_x} \quad (15)$$

waarin:

- R = de Pearson correlatiecoëfficiënt
- σ_x = de standaardafwijking van x
- σ_y = de standaardafwijking van y .

In het voorbeeld van de koppen koffie en thee hebben we berekend $R(x, y) \approx -0,50$ en

$\sigma_x \approx 1,47$ en $\sigma_y \approx 1,47$. Dit geeft:

$$a = R \cdot \frac{\sigma_y}{\sigma_x} \approx -0,50 \cdot \frac{1,47}{1,47} \approx -0,50$$

Voor de regressielijn geldt dan:

$$y = -0,50x + b$$

Met het gegeven dat de regressielijn door het centrale punt (4,4) loopt, kunnen we tot slot b berekenen:

$$y = -0,5x + b$$

$$4 = -0,5 \cdot 4 + b$$

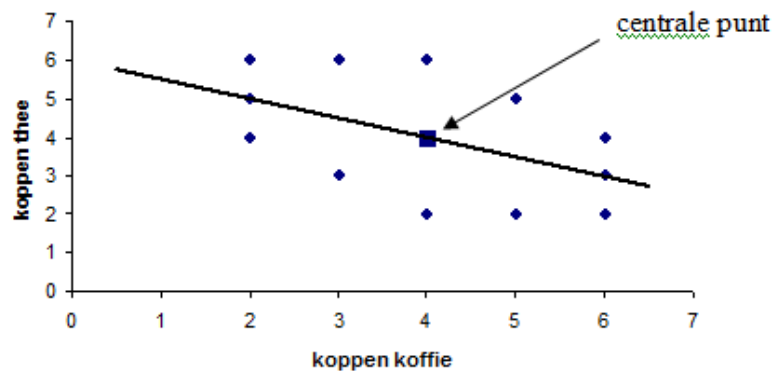
$$4 = -2 + b$$

$$b = 6$$

Voor de regressielijn geldt dan:

$$y = -0,5x + 6$$

In figuur 47 is de regressielijn samen met het centrale punt ingetekend in het spreidingsdiagram.



Figuur 47: spreidingsdiagram van koffie- en theeconsumptie, voorzien van het centrale punt en de regressielijn.

37. Opdracht

37.1 Kleinste-kwadratenmethode

- a. Bereken voor het voorbeeld over de koffie- en thee consumptie de

waarde van $\sum_{i=1}^{i=k} d_i^2$

- b. Teken een willekeurige andere lijn door het centrale punt en laat zien

dat de uitkomst van $\sum_{i=1}^{i=k} d_i^2$ nu inderdaad kleiner is geworden.

37.2 Schatting aantal koppen thee

- a. Geef met behulp van de regressielijn een schatting van het aantal koppen thee dat wordt gedronken in een huishouden waar dagelijks 8 koppen koffie worden gedronken.
- b. Hoe betrouwbaar is deze schatting?

37.3 Recepten voor pijnstillers

Een apotheker houdt bij hoeveel recepten hij dagelijks krijgt aangeleverd. Hierbij noteert hij ook bij hoeveel van deze recepten het om pijnstillers gaat. De resultaten zijn weergegeven in figuur 48.

Dag	Totaal aantal recepten	Recepten voor pijnstillers
1	301	34
2	230	30
3	420	51
4	253	30
5	278	33
6	309	38
7	301	37
8	360	46
9	391	48
10	304	38
11	403	48

Figuur 48: gegevens aantal recepten en aantal recepten voor pijnstillers.

- Zet de gegevens uit in een spreidingsdiagram. Lijkt er sprake te zijn van een lineaire correlatie?
- Bereken de Pearson correlatiecoëfficiënt.
- Geef de vergelijking van de lineaire regressielijn.
- Als het goed is kom je bij de regressielijn uit op een startwaarde van (bijna) 0. Waarom is dat logisch?
- Op een dag krijgt de apotheker 180 recepten binnen. Hoeveel recepten voor pijnstillers zitten hier naar verwachting bij?

37.4 Proefwerkcijfer

In klas 3A is x_i het cijfer dat leerlingen halen op het eerste proefwerk natuurkunde en y_i het cijfer dat leerlingen halen op het tweede proefwerk. Hierbij geldt:

$$\bar{x} = 6,4$$

$$\bar{y} = 4,5$$

$$\sigma_x = 0,8$$

$$\sigma_y = 0,5$$

Adriaan mist helaas het tweede proefwerk. De docent heeft geen zin om een inhaalversie te maken en besluit Adriaans cijfer uit te rekenen met regressieanalyse. Op het eerste proefwerk had Adriaan een 8,8 gehaald, op het tweede proefwerk kreeg hij van zijn docent prompt een 5,5.

- Bereken de Pearson correlatiecoëfficiënt.
- Vind je dit een eerlijke werkwijze van de docent? Ondersteun jouw antwoord met argumenten.

5 Eigen onderzoek CO₂-metingen

In dit hoofdstuk zul je een onderzoek uitvoeren waarbij je uitgaat van de meetgegevens die afkomstig zijn van het school CO₂-net. Om dit onderzoek goed uit te kunnen voeren, moet je eerst wat meer weten over de CO₂-metingen. Daarnaast heb je enige kennis nodig van atmosferische processen, zodat je weet welke andere variabelen mogelijk invloed hebben op de gemeten CO₂-concentratie. Deze informatie krijg je in paragraaf 1 en 2, waarna in paragraaf 3 het eigen onderzoek volgt.



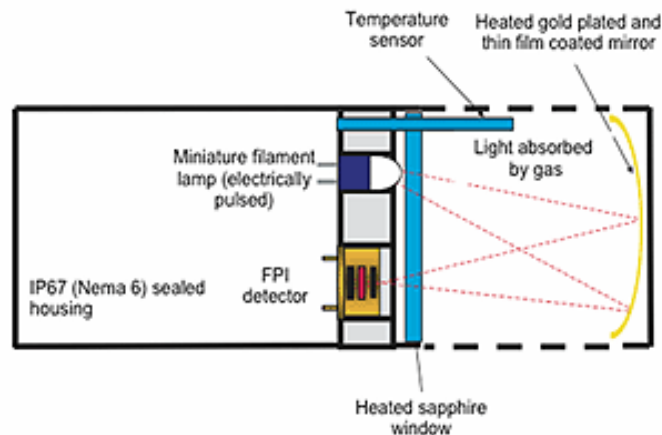
Figuur 49: weerstation met Vaisala Carbocap op het dak van het Willem Lodewijk Gymnasium te Groningen.

5.1 Meten van CO₂

Het meten van CO₂ moet zo nauwkeurig mogelijk gebeuren, zoals jullie nu duidelijk zal zijn. Maar net als bij andere metingen is het nauwkeurig meten van CO₂ niet makkelijk. Je moet goed weten hoe je meet, waar je meet en natuurlijk ook wat je meet. Daarnaast moet je een goede inschatting maken van de meetonnauwkeurigheden waar je mee te maken krijgt en alle storende factoren bij het meten. Hieronder staat omschreven hoe de meetapparatuur werkt die door het school CO₂-netwerk wordt gebruikt.

De CO₂-concentratie wordt gemeten met de zogenaamde Vaisala Carbocap GMP343. Deze detector wordt veel gebruikt in kassen om de lichtsamenstelling in de gaten te houden. Bij de scholen is de meter op het dak bevestigd. De meter kan ook gebruikt worden om metingen op andere plaatsen te doen, zoals in een klaslokaal of in een weiland.

De Vaisala is een zogenaamde '*non-dispersive infra-red*' sensor. In de Vaisala zit een lamp die infrarood licht uitzendt. Dit licht heeft een eigen golflengte en wordt door een spiegel teruggekaatst naar een detector voor infrarood licht. Onderweg komt het licht de CO₂-moleculen in de lucht tegen, die een deel van dit licht, met dezelfde golflengte, absorberen. De infrarooddetector zal dus een lagere lichtsterkte meten dan de lamp heeft uitgezonden. Dit verschil in intensiteit is een maat voor het aantal CO₂-moleculen dat het licht tegen is gekomen en dus voor de CO₂-concentratie in de lucht.



Figuur 50: doorsnede Vaisala Carbocap GMP 343.

Op elke school is ook een Davis Vantage PRO weerstation geplaatst. Dit weerstation stuurt waarden van de luchtdruk, temperatuur, en de luchtvochtigheid door naar de Vaisala. De Vaisala heeft deze gegevens nodig om de CO₂-metingen juist uit te voeren. Dit is het makkelijkste uit te leggen aan de hand van een voorbeeld. Als de luchtdruk in de Vaisala stijgt, dan betekent dat dat er meer luchtmoleculen in dezelfde ruimte zitten. Dus ook meer CO₂-moleculen. Hierdoor meet de Vaisala meer CO₂-moleculen bij een hogere luchtdruk. Omdat je de CO₂-concentratie onafhankelijk van de luchtdruk wilt weten, stuurt het weerstation de waarde voor de luchtdruk naar de Vaisala. De Vaisala rekent vervolgens de CO₂-concentratie om naar een vaste waarde voor de luchtdruk. Het weerstation houdt ook andere factoren bij, zoals de hoeveelheid neerslag, de windrichting en de windsnelheid.

38. Vraag

Waarom is het meten van de hoeveelheid neerslag, de windrichting en de windsnelheid bij de Vaisala zo belangrijk?

Fouten treden niet alleen op omdat mensen het meetwerk doen of omdat de omstandigheden kunnen veranderen, maar ook omdat onnauwkeurigheden kunnen ontstaan in de meter. Bij metingen moet je altijd beginnen met je af te vragen of de meetapparatuur wel goed genoeg is. En of de waarden die worden aangegeven door de apparatuur wel overeenstemmen met de definitie volgens het S.I.-eenhedenstelsel. Elk meetinstrument heeft een zekere afwijking. Om de meetonnauwkeurigheid zoveel mogelijk te verkleinen, kalibreren en justeren we meetinstrumenten. **Kalibreren** is het bepalen van de afwijking van het meetinstrument ten opzichte van de relevante standaard. Dat doe je door het meetinstrument te refereren aan een ander instrument of met een berekend model. **Justeren** is het verrichten van handelingen om een meetinstrument zo nauwkeurig mogelijk te laten functioneren. Vaak wordt voor deze handeling ook wel het woord 'afregelen' gebruikt. Bij kalibreren bepaal je dus de afwijking van het instrument en bij justeren pas je het

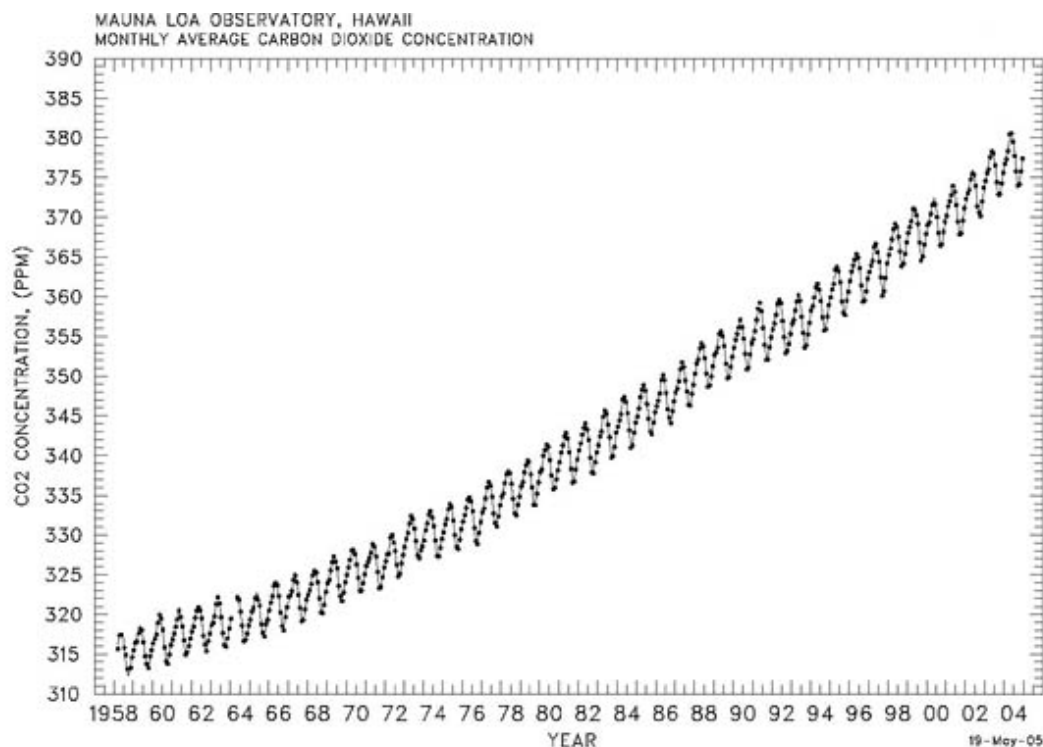
meetinstrument zo aan dat de afwijking verdwijnt. Ook wordt vaak het begrip ijken gebruikt.

Pas als dit hele traject is doorlopen weet je zeker dat je meetapparatuur nauwkeurig genoeg is.

Zoals in het begin al werd omschreven, is niet alleen hoe je meet van belang, maar ook waar je meet. Dat geldt natuurlijk ook voor CO₂-metingen. Een meting naast de schoorsteen van een fabriek zal een andere waarde geven dan een meting midden in het bos in de zomer. En geen van beide metingen geeft het antwoord op de vraag hoe groot de concentratie CO₂ gemiddeld in Nederland is. Datzelfde geldt natuurlijk ook op mondiale schaal. Weten waar je meet en welke invloed dat heeft op je meetgegevens is dus van groot belang bij het interpreteren van je meetgegevens.

De eerste onderzoeker die eind jaren '50 begon met het doen van nauwkeurige metingen aan CO₂-concentraties was een chemicus, Charles David Keeling. Hij werkte op het Scripps Institution of Oceanography te San Diego Verenigde Staten. Een instituut waar men voornamelijk onderzoek doet naar oceanen en de relatie tussen oceanen en de atmosfeer. Zijn onderzoek vond voornamelijk plaats op twee plekken op de wereld. De eerste plek was op Mauna Loa, Hawaï.

In figuur 51 zie je de resultaten van zijn metingen op Mauna Loa vanaf 1958 tot en met 2004. Deze grafiek staat ook wel bekend als de 'Keeling curve'.



Figuur 51: CO₂-concentraties gemeten op Mauna Loa (Hawai).

39. Vraag

- a. Waarom werden de CO₂-metingen verricht op Manua Loa op Hawaï?
- b. Wat zou de tweede plek kunnen zijn waar Keeling onderzoek deed?

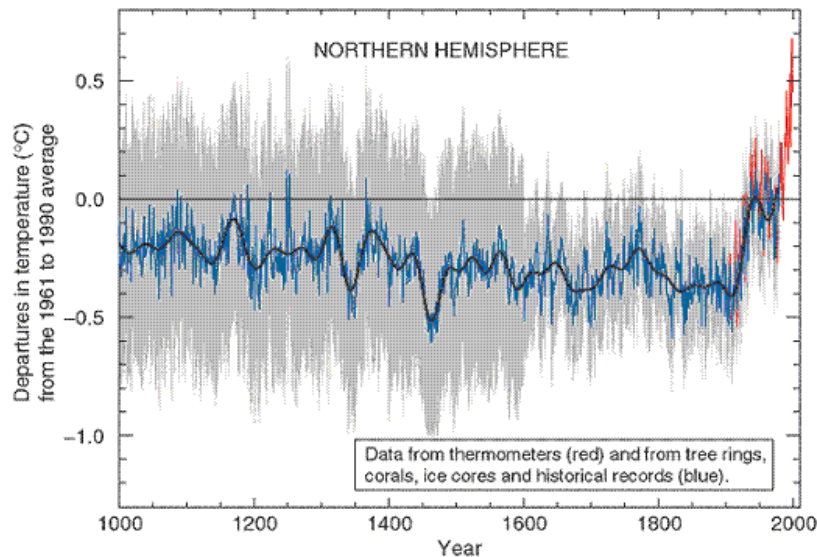
In Nederland worden de metingen gedaan op Ludjewad, bij de Waddenzee in Groningen op 60 meter hoogte. Deze hoogte is van groot belang omdat de luchtlaag hier relatief stabiel is en zo min mogelijk verstoord wordt. Naast Ludjewad worden er nog veel meer metingen gedaan via het netwerk. Zo staan bij scholen in de provincie Groningen en rondom Amsterdam meters. En niet alleen in Nederland, maar ook verder in Europa vinden metingen plaats. Geen van deze plekken zal ideaal zijn voor de metingen, maar samen kunnen ze waardevolle gegevens leveren.

Ga voor informatie over het meetstation te Ludjewad naar ► vaklokaal NLT en bekijk de informatiefilm.

Met zoveel verschillende gegevens uit een relatief klein gebied kan een model gemaakt worden. Dit model zou een aanvulling kunnen zijn op de reeds bestaande CO₂-modellen, vaak klimaatmodellen genoemd. Een model geeft de essentie weer van een systeem en laat daarbij alle overbodige zaken weg. Dus alle aandacht wordt gericht op het onderwerp van het model en alle andere, vaak storende, informatie wordt weggelaten. Een voorbeeld van een model is de koolstofkringloop zoals afgebeeld in figuur 2 in hoofdstuk 1. Hier worden alleen de belangrijkste uitwisselingen van koolstof getoond. Te kleine uitwisselingen of andere stoffen worden als overbodige zaken gezien en dus weggelaten. Het model is zo een versimpeling van de werkelijkheid. Het bepalen welke onderdelen essentieel zijn en welke overbodig is erg moeilijk. Dat maakt het maken van een model ook erg moeilijk. Want wat is essentieel of juist overbodig binnen een systeem? Vaak zal dat afhangen van de vraag die je wilt beantwoorden.

Klimaatmodellen worden vaak gebruikt omdat de werkelijkheid erg gecompliceerd is. En als je duidelijkheid wilt over een onderdeel van die werkelijkheid dan is een versimpeling van de werkelijkheid echt nodig. Zonder deze modellen zou het niet mogelijk zijn om goede voorspellingen te doen over de toekomst. Zelfs met goede modellen blijkt dat heel moeilijk.

Eén van de bekendste voorbeelden van onenigheid over het gebruik van klimaatmodellen is de kritiek die McIntyre en McKittrick hadden op de hockeystick grafiek van Michael Mann (zie figuur 52).



Figuur 52: “hockeystick-curve van Mann”.

De grafiek in figuur 52 is de uitkomst van een gebruikt model. De gegevens om deze grafiek te maken, zijn gepubliceerd in 1998. De gegevens zijn verkregen door een reconstructie van het klimaat aan de hand van jaarringen van bomen, ijskernen van de Noordpool en historische bronnen waarin temperaturen en neerslag beschreven staan. Deze laatste gegevens werden gebruikt om de andere gegevens te kalibreren.

De kritiek van McIntyre en McKittrick had te maken met de manier van verwerking van de gegevens. Maar misschien nog wel belangrijker was de kritiek die ze hadden op de gebruikte data. Zo zouden er bij de metingen van jaarringen van bomen te weinig metingen zijn gedaan; voor de periode 1400 -1450 zouden maar twee bomen gebruikt zijn. Ook is er maar naar één boomsoort onderzoek gedaan, een boom die maar heel plaatselijk voorkomt in Amerika en niet wereldwijd.

40. Vraag

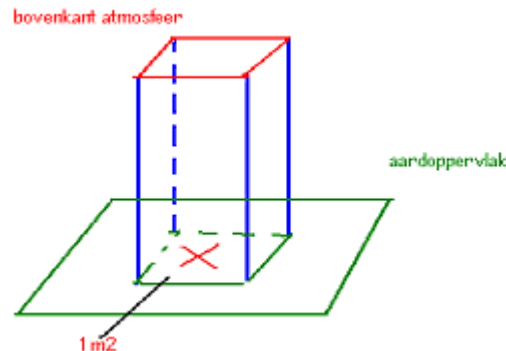
- Waarom is het van belang zoveel mogelijk meetgegevens te hebben als je onderzoek doet naar klimaatverandering?
- Waarom zou het bij het maken van een klimaatmodel van belang kunnen zijn om een boomsoort te kiezen die overal ter wereld groeit?

Het juiste model en de goede meetgegevens gebruiken is van groot belang. Vooral omdat het wereldwijde beleid op het gebied van milieumaatregelen wordt gemaakt aan de hand van de uitkomsten van het gebruikte klimaatmodel. Zo is de hockeystick grafiek van Mann gebruikt bij het schrijven van het Kyoto-protocol. In dit protocol hebben 37 industrielanden en de EU-lidstaten afgesproken dat ze tussen 2008 en 2012 een vermindering van 5% CO₂-uitstoot nastreven in vergelijking tot 1990.

5.2 Opbouw van de atmosfeer

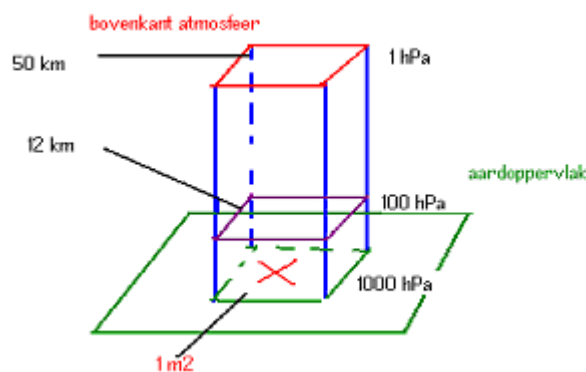
Luchtdruk

Lucht is opgebouwd uit gassen en gassen kunnen worden samengedrukt. Dichter bij het aardoppervlak wordt de lucht het meeste samengeperst door het gewicht van de bovenliggende luchtlagen. Hierdoor bevinden zich dichter bij het aardoppervlak meer deeltjes in één liter lucht en dat zorgt voor een hogere dichtheid. De kolom lucht zorgt voor een bepaalde druk op het aardoppervlak. Deze druk van de lucht noemen we **luchtdruk**.



Figuur 53: ontstaan van luchtdruk.

Als je verder van het aardoppervlak verwijderd bent, is de kolom lucht die zich boven je bevindt uiteraard ook kleiner. Dat zorgt voor een lagere luchtdruk op grotere hoogte (zie figuur 54).



Figuur 54: variatie van luchtdruk, afhankelijk van de hoogte boven het aardoppervlak.

Je kunt dus zeggen dat lucht gewicht heeft. Op zeeniveau weegt één liter lucht 1,3 gram. Luchtdruk wordt gemeten in millibaren of hectopascal.

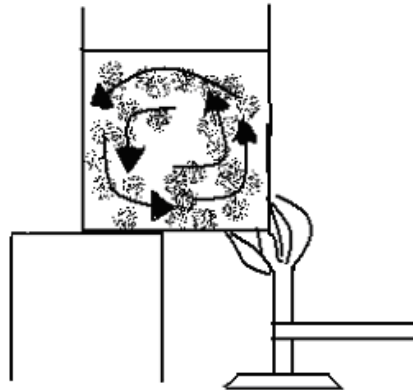
Ook de temperatuur heeft invloed op de luchtdichtheid. Hoe hoger de temperatuur hoe minder de luchtdichtheid zal zijn. De Franse natuurkundige Gay-Lussac heeft dat in de gaswet omschreven; deze gaswet luidt:

Als de temperatuur van een gas wordt verhoogd, zullen de moleculen van het gas sneller gaan bewegen, wat tot expansie van het gas leidt. Dit zorgt ervoor dat er minder moleculen in een volume van één liter aanwezig zijn en dus is de dichtheid lager.

Luchtdrukverdeling

Onder invloed van de temperatuur zal de luchtdruk dus veranderen. Zo is er een vaste luchtdrukverdeling over de wereld. Bij de evenaar (waar het gemiddeld het warmst is op het aardoppervlak) liggen lagedrukgebieden. Bij de Noord- en de Zuidpool liggen juist hogedrukgebieden omdat het daar koud is.

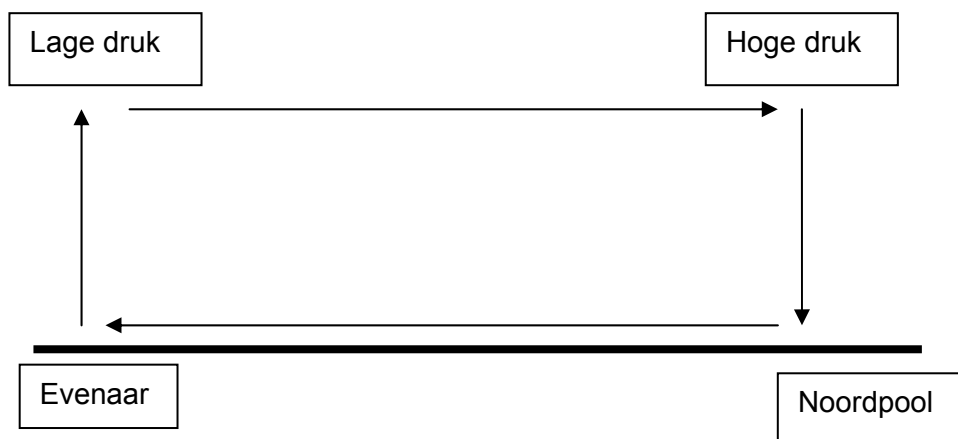
In figuur 55 is te zien dat hetzelfde principe optreedt in een vloeistof die plaatselijk verwarmd wordt.



Figuur 55: stroming in vloeistof.

Een gasvlam verwarmt het rechterdeel van een bak met water, waardoor de temperatuur rechts onderin hoger is dan elders in de bak. Net als lucht heeft warm water een lagere dichtheid dan koud water. Het warme water stijgt daardoor op; kouder en dus 'zwaarder' water, schuift onder het warmere water en zo ontstaat er een stroming. Wanneer we dit doortrekken naar de atmosfeer, dan verwachten we de volgende situatie (zie figuur 56):

- aan de evenaar opstijgende lucht
- aan de pool dalende lucht
- aan het aardoppervlak een luchtstroming rechtstreeks van pool naar evenaar
- in de bovenlucht een luchtbeweging van evenaar naar pool.

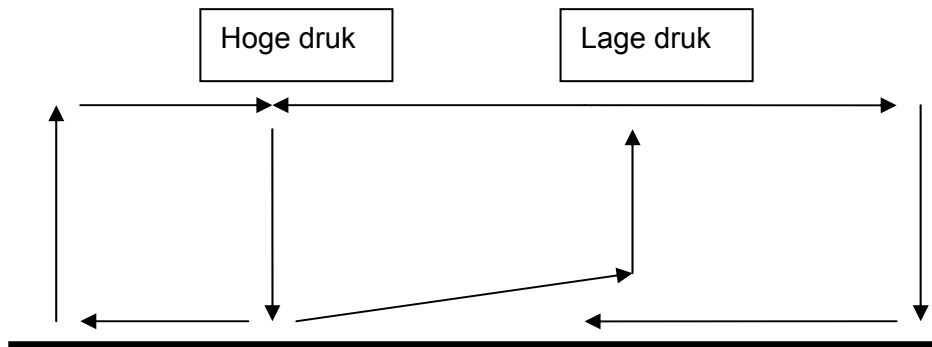


Figuur 56: ontstaan van luchtdruksystemen.

Figuur 56 is sterk vereenvoudigd. In het wereldsysteem van hoge- en lagedrukgebieden ligt er tussen evenaar en pool nog een gebied met relatief hoge luchtdruk en een gebied met relatief lage luchtdruk.

41. Vraag

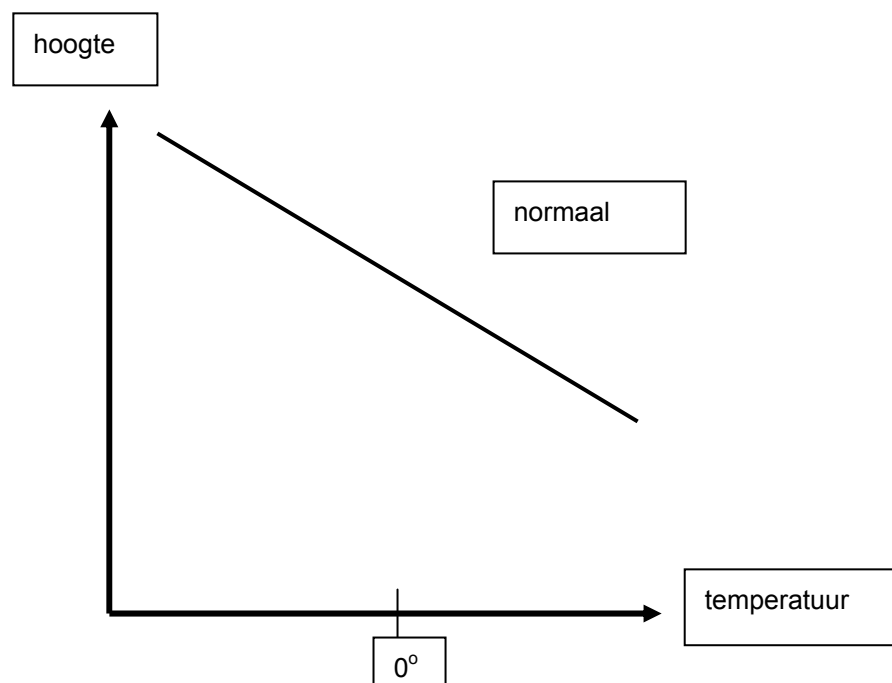
Leg aan de hand van onderstaande figuur uit hoe dit extra hogedrukgebied kan ontstaan en hoe het extra lagedrukgebied kan ontstaan.



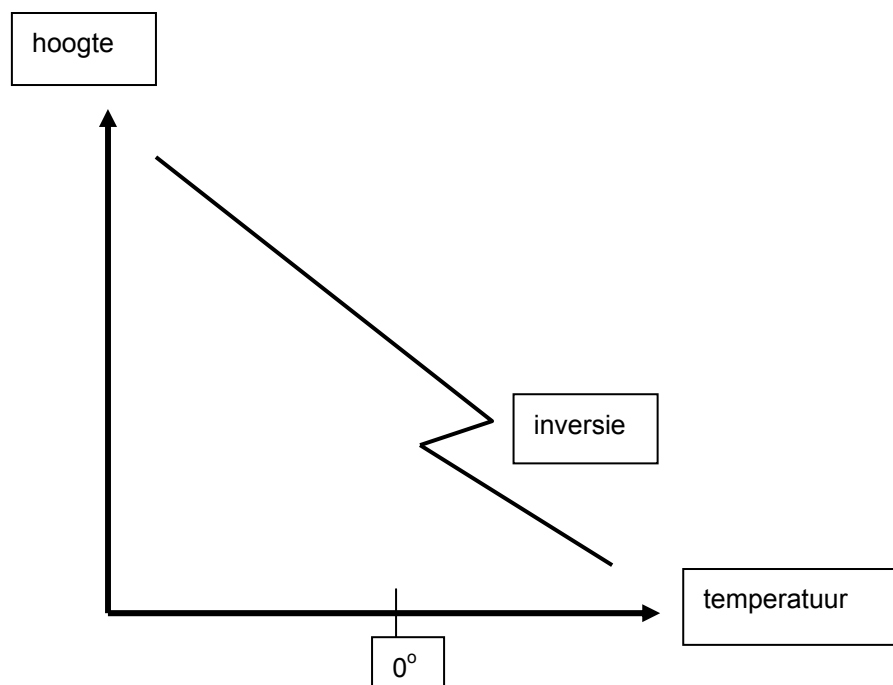
Figuur 57: gedetailleerder overzicht druksystemen.

Inversie

De onderste 11 kilometer van de atmosfeer wordt de troposfeer genoemd. In deze laag vindt het grootste deel van het weer plaats. Normaal is deze luchtlaag stabiel, hoe hoger je komt hoe kouder het wordt. Voor elke 100 meter hoger wordt het 0,6 °C kouder (zie figuur 58).



Figuur 58: afname van temperatuur bij toenemende hoogte.



Figuur 59: temperatuursverloop bij een inversielaag.

Maar de lucht kan ook onstabiel zijn, er treedt dan **inversie** op. Dit houdt in dat bij toenemende hoogte de lucht eerst afkoelt, daarna weer opwarmt om uiteindelijk weer verder af te koelen (zie figuur 59).

In Nederland treedt inversie meestal 's zomers in de nacht op, vooral als het windstil of onbewolkt is. Het aardoppervlak koelt dan sterk af terwijl de lucht nog warm is na een warme dag. Maar inversie kan ook in andere jaargetijden optreden. Denk maar aan een mooie zonnige dag in de winter terwijl er sneeuw ligt.

42. Vraag

Leg uit hoe inversie kan ontstaan op een mooie winterdag als er sneeuw ligt.

Inversie is soms goed te zien. In figuur 60 zie je goed dat de rook niet verder kan stijgen door de inversie.



Figuur 60: rook die door een inversielaag gehinderd wordt om verder op te stijgen.

In figuur 61 is duidelijk smog te zien. Ook smog ontstaat door inversie.



Figuur 61: smog.

43. Vraag

Leg uit hoe smog kan ontstaan door inversie.

5.3 Eigen onderzoek

Nu je iets weet over CO₂ in de atmosfeer, is het tijd de kennis die je hebt opgedaan in de eerste vier hoofdstukken in de praktijk te brengen. Als uitgangspunt heb je de gegevens die door het school CO₂-net beschikbaar worden gesteld. Het betreft niet alleen CO₂-metingen, maar ook de gegevens van de weerstations die bij de CO₂-meters geplaatst zijn. Je kunt zelf kiezen of je de gegevens van één enkele school wilt gebruiken of dat je juist gegevens van verschillende scholen wilt combineren. Hoe je aan deze gegevens kunt komen, vind je in bijlage 3.

Met de gegevens voer je de volgende opdrachten uit:

- Maak een keuze voor één of meer meetreeksen die je wilt bestuderen. Presenteer deze gegevens.
- Bedenk vooraf wat het doel is van jouw onderzoek. Formuleer onderzoeksvragen. Je kunt hierbij denken aan vergelijken van de meetgegevens van verschillende scholen, trends, correlaties en voorspellingen. Leg de onderzoeksvragen ter goedkeuring voor aan je docent.
- Voer het onderzoek uit. Maak hierbij gebruik van alle kennis die je hebt over de atmosfeer en over het combineren en filteren van meetgegevens en van je kennis over correlatie en regressie. Presenteer de berekeningen op een overzichtelijke manier. Bij de verwerking van de meetgegevens maak je gebruik van Excel. Kijk in bijlage 3 en 4 naar het gebruik van Excel. In bijlage 3 is uitgelegd hoe je de gegevens van

het CO₂-net kunt importeren in Excel, in bijlage 4 is uitgelegd hoe je Excel kunt gebruiken bij het verwerken van de meetgegevens.

- Sluit het onderzoek af met duidelijke conclusies. Is het je gelukt de onderzoeksvragen te beantwoorden?
- Het gehele onderzoek presenteert je als een verslag. Zorg ervoor dat het verslag goed leesbaar is en logisch is opgebouwd. Gebruik hiervoor de ►werkinstructie onderzoeksverslag in de NLT Toolbox.

Bijlage 1 Rekenen met meetfouten

"To err is human; to describe the error properly is sublime."
-- Cliff Swartz, Physics Today 37 (1999), 388.

In paragraaf 2.5 zijn formules gegeven die je kunt gebruiken wanneer je verder gaat rekenen met meetwaarden waar onnauwkeurigheden in zitten. In deze bijlage wordt uitgelegd hoe deze formules tot stand zijn gekomen. Wil je hier meer van weten, werk dan deze bijlage door.

In de loop der jaren is enorm veel gepubliceerd over foutenanalyse. Eén van de redenen dat er zoveel is geschreven, ligt in het feit dat er nog geen eenduidige richtlijnen zijn om met fouten om te gaan. Ook de hieronder beschreven theorie is een versimpelde weergave van de werkelijkheid. De uiteindelijke fouten worden hierbij eerder overschat dan onderschat, dus je zit met deze methode aan de 'veilige kant'.

Laten we bekijken wat er gebeurt met de nauwkeurigheid wanneer meetresultaten in formules worden gebruikt. Stel we hebben twee meetwaarden:

meetwaarde 1 : $x \pm a$

meetwaarde 2 : $y \pm b$

Met a en b worden de absolute fouten van x , respectievelijk y bedoeld.

Optellen en aftrekken

We sommeren meetwaarde 1 en meetwaarde 2.

De laagste mogelijke uitkomst is nu:

$$x - a + y - b = x + y - (a + b)$$

De hoogste mogelijke uitkomst is:

$$x + a + y + b = x + y + (a + b)$$

Zodat geldt:

$$x + y - (a + b) < X + Y < x + y + (a + b) \quad (16)$$

Waarbij X en Y de werkelijke waarden zijn.

Hieruit volgt dat bij optellen van meetwaarden de absolute meetfouten ook opgeteld moeten worden. Hierbij moet wel worden opgemerkt dat deze methode van doorrekenen van fouten alleen volstaat voor systematische fouten. Systematische fouten hebben immers als kenmerk dat ze 'dezelfde kant op werken'. Bij toevallige fouten is dit echter helemaal niet zeker en daarom mag je toevallige fouten kwadratisch optellen. De fout in het antwoord wordt dan $\sqrt{a^2 + b^2}$ en is iets kleiner dan de hierboven beschreven $a + b$. Eenzelfde redenering geldt wanneer we niet twee meetwaarden optellen, maar twee meetwaarden vermenigvuldigen of delen.

Het is alleen lang niet altijd duidelijk of je te maken hebt met systematische fouten, met toevallige fouten of met een combinatie van beide. Daarom werken we in deze module met de versimpelde vorm van doorrekenen, die als voordeel heeft dat de fouten die optreden hooguit worden overschat.

45. Vraag

Leid af dat dezelfde regel als bij optellen ook opgaat wanneer het verschil van twee meetwaarden wordt berekend.

Merk op, dat op de middelbare school veelal met significante cijfers wordt gewerkt. Bij de som en het verschil wordt dan als vuistregel gehanteerd dat het aantal cijfers achter de komma in het antwoord gelijk is aan het kleinste aantal cijfers achter de komma van de oorspronkelijke getallen. We kunnen nu echter bewijzen dat dit misschien een goede benadering is, maar dat de werkelijkheid iets anders ligt.

46. Vraag

We tellen de volgende (gewogen) gewichten bij elkaar op:

0,6 g ; 1,6 g ; 0,9 g ; 2,3 g ; 0,2 g ; 0,9 g

- Hoe groot is de absolute fout in elke meetwaarde?
- Voer de berekening uit volgens de methode van de significante cijfers. Hoe groot is de onzekerheid (absolute fout) volgens deze methode in het antwoord?
- Voer de berekening uit volgens de theorie die hierboven is uitgelegd. Hoe groot is de absolute fout volgens deze methode?
- Waarschijnlijk wijken jouw antwoorden bij vraag b en c van elkaar af. Is de methode met de significante cijfers nauwkeuriger bij toevallige meetfouten of bij systematische meetfouten?
- Als je er zeker van wilt zijn dat de werkelijke waarde niet meer dan de absolute fout afwijkt van de berekende waarde, voor welke methode kies je dan?

Je hebt gezien dat het werken met significante cijfers een eenvoudige manier is om meetwaarden met een bepaalde onnauwkeurigheid te kunnen optellen en aftrekken en toch iets over de onnauwkeurigheid in het eindantwoord te kunnen zeggen. Helemaal correct is deze methode echter niet, dus wanneer grote precisie is vereist, zul je bovenstaande theorie moeten volgen.

Vermenigvuldigen met een constante

Stel je wilt meetwaarde 1 (zie hierboven) vermenigvuldigen met de constante c . Wat wordt nu de onzekerheid in de uitkomst?

De laagst mogelijke uitkomst is:

$$c(x - a) = cx - ca$$

De hoogst mogelijke uitkomst is:

$$c(x + a) = cx + ca$$

Er geldt dus:

$$cx - ca < cX < cx + ca \quad (17)$$

Waarbij X weer de werkelijke waarde is.

Dit betekent dat de absolute fout ook met de constante is vermenigvuldigd.

Hoe heb je dit tot nog toe op school geleerd? Wanneer je werkt met significante cijfers wordt er onderscheid gemaakt tussen meet- en telwaarden. In het voorbeeld hierboven is x een **meetwaarde**, die dus een bepaalde onzekerheid heeft, die wordt uitgedrukt in het aantal significante cijfers waarmee je x weergeeft. De constante c daarentegen is een **telwaarde**, die geen onzekerheid kent. Is de waarde van c gelijk aan 6, dan bedoel je ook precies 6 en niet iets wat afgerond uitkomt op 6. In significante cijfers is dit niet weer te geven.

47. Vraag

Stel je hebt een meetlat van 100 cm. Hiermee bedoelen we dus een meetlat met een lengte $l = 100 \pm 0,5$ cm. De onnauwkeurigheid in de meetlat is niet het gevolg van problemen bij het aflezen (dat kan véél nauwkeuriger dan 0,5 cm), maar is het gevolg van een grove fabricatie. Voor de sportdag op school heb je een atletiekbaan van 60 m nodig. Deze meetbaan zet je met behulp van de meetlat uit. Afleesfouten laten we dus buiten beschouwing.

- Bereken de minimale en de maximale lengte die de atletiekbaan kan krijgen.
- Hoe groot is dus de absolute fout in de lengte van de atletiekbaan?
- Hoe groot is de onzekerheid wanneer je rekent met de normale regels voor significantie die je tot dusverre op school hebt geleerd?
- Als je twee even lange banen wilt uitzetten met dezelfde meetlat, zullen die op deze wijze veel in onderlinge lengte verschillen?
Stel dat de meetlat wél erg nauwkeurig is gefabriceerd, maar dat de absolute fout van 0,5 cm ontstaat door een toevallige fout bij het aflezen.
- Geef ook nu de minimale en de maximale lengte van de atletiekbaan.
- Als je nu twee even lange banen wilt uitzetten met dezelfde meetlat, zullen die op deze wijze veel in onderlinge lengte verschillen?
- Wat is het verschil met de situatie die bij vraag a t/m c is omschreven?

Je ziet dat ook in deze situatie de onzekerheden die kunnen optreden iets afwijken van de onzekerheden die je met de significantieregels bepaalt.

Vermenigvuldigen

Nu wil je meetwaarde 1 vermenigvuldigen met meetwaarde 2.

48. Vraag

Toon aan dat geldt:

$$xy - bx - ay + ab < XY < xy + bx + ay + ab$$

Bij bruikbare meetgegevens zullen de absolute fouten a en b veel kleiner zijn dan de meetwaarden x en y . De term ab in bovenstaande ongelijkheid is daarom verwaarloosbaar, zodat bij benadering geldt:

$$xy - bx - ay < XY < xy + bx + ay$$

$$xy - xy \left(\frac{b}{y} + \frac{a}{x} \right) < XY < xy + xy \left(\frac{b}{y} + \frac{a}{x} \right)$$

$$xy - xy(s + r) < XY < xy + xy(s + r) \quad (18)$$

met $r = \frac{a}{x}$ en $s = \frac{b}{y}$ dus r en s zijn de relatieve fouten in x respectievelijk y .

Bij vermenigvuldigen tel je dus de relatieve fouten bij elkaar op om de relatieve fout in het antwoord te bepalen. Die is eventueel weer om te rekenen naar de absolute fout in het antwoord.

49. Vraag

Gegeven is een grasveld met $l = 23,0$ m en $b = 6,4$ m.

- Hoe groot zijn de absolute fouten in lengte en breedte?
- Bereken de oppervlakte met bijbehorende relatieve fout.
- Hoe groot is de absolute fout in de oppervlakte?
- Ook voor vermenigvuldigen heb je eerder vuistregels geleerd voor de significantie. Bereken de oppervlakte van het grasveld opnieuw, maar bepaal de absolute en relatieve fout nu aan de hand van deze vuistregels. Zie je veel verschil?

Delen

Wat gebeurt er wanneer we meetwaarde 1 delen door meetwaarde 2?

$$\frac{x-a}{y+b} < \frac{X}{Y} < \frac{x+a}{y-b}$$

$$\frac{x\left(1-\frac{a}{x}\right)}{y\left(1+\frac{b}{y}\right)} < \frac{X}{Y} < \frac{x\left(1+\frac{a}{x}\right)}{y\left(1-\frac{b}{y}\right)}$$

$$\frac{x(1-r)}{y(1+s)} < \frac{X}{Y} < \frac{x(1+r)}{y(1-s)}$$

$$\frac{x}{y} \cdot \frac{1-r}{1+s} < \frac{X}{Y} < \frac{x}{y} \cdot \frac{1+r}{1-s}$$

met r en s de relatieve fouten in x respectievelijk y .

Zonder bewijs geven we:

$$\frac{1-r}{1+s} = 1 - r - s + rs + s^2 - rs^2 \dots \quad \text{voor } r < 1 \text{ en } s < 1$$

Bij kleine relatieve fouten kunnen we de termen vanaf rs verwaarlozen, zodat geldt:

$$\frac{x}{y} \cdot (1 - (r+s)) < \frac{X}{Y} < \frac{x}{y} \cdot (1 + (r+s)) \quad (19)$$

We zien nu dus dat ook bij delen de relatieve fouten bij elkaar opgeteld kunnen worden. Dit gaat op dezelfde wijze als bij vermenigvuldigen.

Experiment

Je krijgt van je docent een blokje onbekend metaal. Probeer zo nauwkeurig mogelijk de dichtheid van dit blokje uit te rekenen en bereken daarbij de absolute en de relatieve fout. Probeer de meetapparatuur te kiezen waarbij deze fouten het kleinst zijn. Vergelijk jouw uitkomsten met die van je klasgenoten en verklaar de resultaten.

Kwadrateren

50. Vraag

Meetwaarde 1 wordt gekwadrateerd. Leid af dat voor relatief kleine fouten bij benadering geldt:

$$x^2 - 2rx^2 < X^2 < x^2 + 2rx^2$$

De uitkomst bij vraag 50 betekent dat bij kwadrateren de relatieve fout verdubbelt, wat logisch is als je kwadrateren ziet als een speciale vorm van vermenigvuldigen.

51. Vraag

Stel dat we de zwaartekrachtsversnelling proberen te meten door een steentje te laten vallen van de galerij op de derde verdieping van een flatgebouw en met een stopwatch te meten hoe lang het duurt voordat het op de grond valt. Uit de bouwtekening van de flat weten we dat de vloer van de derde etage $H = 7.6 \pm 0.3$ m van de grond is. Verder hebben we een stellage gemaakt met een statief en een klem waaruit we het steentje loslaten. Ook hebben we gemeten dat het steentje dan op een hoogte van $h = 498.3 \pm 0.8$ mm van de vloer wordt losgelaten.

Vervolgens doen we de meting een aantal keer en meten met de stopwatch elke keer hoe lang het vallen duurt: 1.31, 1.39, 1.27, 1.43, 1.33, 1.46, 1.32, 1.36, 1.29, 1.41, 1.38 s.

- Geef de formule waarmee je de valversnelling kunt berekenen.
- Bereken de valversnelling.
- Bereken de relatieve fout in de totale hoogte van waar we het steentje laten vallen.
- We rekenen verder met het gemiddelde van de tijdsmetingen. Bereken de relatieve fout in dit gemiddelde.
- Bereken de absolute en de relatieve fout in de valversnelling die je berekend hebt.
- De berekende valversnelling wijkt nogal af van de uit de literatuur bekende waarde ($9,81 \text{ m/s}^2$). Geef hiervoor een verklaring.

Bijlage 2 Werken met lijsten in de GRM

Bij opgave 26 heb je de covariantie uitgerekend bij een voorbeeld met de stamdiameter en de hoogte van een boom. Daarvoor moest je de tabel invullen in figuur 62.

diameter x_i (cm)	hoogte y_i (m)	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$
10	6			
20	12			
30	11			
40	15			
40	30			
50	21			
55	20			
60	24			
80	30			
90	15			
90	31			
100	35			
totaal				+

Figuur 62: tabel voor de berekening van de covariantie.

Het invullen van de tabel kan natuurlijk handmatig, maar dat is erg veel werk. Zeker omdat je dergelijke berekeningen veel vaker moet uitvoeren. Daarom volgt hier uitleg hoe je de berekening ook met behulp van een grafische rekenmachine (TI-83 of TI-84) kunt uitvoeren:

1. Druk op **STAT**.
2. Selecteer **1:Edit...** en druk op **ENTER**.

Je bent nu in het menu waar je lijsten kunt invoeren. Het kan zijn dat er nog oude lijsten in jouw rekenmachine staan. Ga in dat geval met de cursor op L_1 staan en druk op **CLEAR** en vervolgens op **ENTER**. L_1 is nu leeg. Zo kun je eventueel ook de andere lijsten leegmaken.

We gaan er vanuit dat de lijsten nu leeg zijn. Je moet nu de gegevens uit de tabel in de GRM invullen:

3. Ga met de cursor in de kolom onder L_1 staan. Voer de waarde 10 in en druk op **ENTER**. Voer zo de eerste kolom in.
4. Voer op dezelfde wijze de kolom met de hoogtes in in L_2 .

We laten nu de GRM de berekeningen van kolom 3, 4 en 5 uitvoeren:

5. Ga met de cursor op L_3 staan.
6. Voer daar in: $L_1 - \text{mean}(L_1)$
 L_1 staat op het toetsenbord (2ND 1)
 Mean vind je onder 2ND STAT, MATH
7. Druk op ENTER.

In L_3 is nu een kolom met $x_i - \bar{x}$ ingevuld.

8. Voer op dezelfde wijze in: $L_4 = L_2 - \text{MEAN}(L_2)$
9. Voer tenslotte in L_5 in (ga daarvoor met de cursor weer op L_5 staan):
 $L_5 = L_3 \cdot L_4$

De tabel hierboven staat nu in z'n geheel in de GRM, zonder dat het je rekenwerk heeft opgeleverd. Het enige dat we nog nodig hebben, is het totaal van L_5 . Voer daarvoor de volgende stappen uit:

10. Druk op *Quit* (2ND MODE).
11. Druk op *LIST* (2ND STAT), selecteer het tabblad *MATH* en kies voor
 $5:\text{sum}($.
12. Druk op ENTER.
13. Zorg dat als invoer op je scherm komt te staan: $\text{sum}(L_5)$.
14. Druk op ENTER.

Als het goed is, verschijnt nu het totaal van de kolom $(x_i - \bar{x}) \cdot (y_i - \bar{y})$ in beeld (in ons voorbeeld is dat ongeveer 2206).

Gebruik de GRM in het vervolg van hoofdstuk 4. Dat bespaart je veel tijd en veel onnodig rekenwerk. Uiteraard zou je hiervoor ook een spreadsheet programma zoals MS Excel kunnen gebruiken.

Bijlage 3 Importeren van de gegevens van de meetstations in Excel

In principe bestaat het importeren uit de volgende stappen:

- Ga naar de website van het school CO₂-web: ►URL1.
 - o klik op meetgegevens
 - o selecteer een school en een begindatum en einddatum
 - o klik op Download
 - o sla het bestand op op een plaats waar je het gemakkelijk terug kunt vinden.
- Open MS Excel, zorg dat je een lege spreadsheet voor je hebt.
- Ga naar *Gegevens* en selecteer daar bij *externe gegevens importeren* de optie *van tekst*.
- Selecteer het bestand dat je hebt opgeslagen.
- De volgende wizard wordt automatisch geopend. Ga vervolgens te werk zoals hieronder is beschreven:
 - o klik op volgende
 - o selecteer als scheidingsteken de “komma”
 - o geef via de knop *Advanced* aan dat de decimalen gescheiden worden door een punt.
 - o klik op *voltooien*.
- De gegevens staan nu in Excel. Klik op *Start* en vervolgens op *Zoeken en selecteren* en dan op *Vervangen*.
- Geef aan dat je de punten door komma's wilt vervangen.
- Het bestand is nu te gebruiken in Excel.

Bijlage 4 Excel

Het is erg bewerkelijk om grote meetreeksen op deze manier met de hand door te rekenen. In de praktijk wordt daarom veel gebruik gemaakt van hiervoor geschikte computersoftware. Er zijn programma's beschikbaar waarin alle berekeningen al zijn voorgeprogrammeerd en je slechts de meetgegevens hoeft in te voeren om correlaties te bepalen. Ook de vergelijkingen van eventuele regressielijnen worden direct gegeven. Erg handig, maar er schuilt tevens een gevaar in deze programma's: het **'black-box'** gehalte is erg hoog. Dat betekent dat de gebruiker vaak weinig weet heeft van de berekeningen die worden uitgevoerd en dat de software ook driftig aan het rekenen gaat met meetresultaten waar de software eigenlijk niet voor bedoeld is. Eigenlijk is het overbodig om te zeggen dat de uitkomsten dan weinig zinvol zijn....

Bij het eigen onderzoek waarmee deze module wordt afgesloten, worden de berekeningen daarom in Microsoft Excel uitgevoerd. Microsoft Excel is een zogenaamd spreadsheetprogramma dat tal van berekeningen kan uitvoeren. De benodigde formules dien je zelf in te voeren. Om hierin enigszins thuis te raken, herhalen we de berekeningen uit paragraaf 4.1, maar nu voeren we ze in Excel uit. Werk deze bijlage door terwijl je zelf achter de computer zit. Alleen dan weet je zeker of je het goed doet!

We gaan weer uit van de gegevens in figuur 63 die de eigenaar van de strandtent ons heeft gegeven.

datum	dag	temperatuur (°C)	aantal verkochte drankjes
14-7	Ma	23	708
15-7	Di	21	628
16-7	Wo	20	552
17-7	Do	18	418
18-7	Vr	21	834
19-7	Za	20	2110
20-7	Zo	18	1057
21-7	Ma	16	295
22-7	Di	21	627
23-7	Wo	23	718
24-7	Do	26	908
25-7	Vr	28	1400
26-7	Za	29	1933
27-7	Zo	28	1828
28-7	Ma	28	1028

29-7	Di	25	851
30-7	Wo	27	977
31-7	Do	30	1155
01-8	Vr	23	983
02-8	Za	24	1505
03-8	Zo	21	1266
04-8	Ma	21	602
05-8	Di	24	777
06-8	Wo	27	993
07-8	Do	22	663
08-8	Vr	21	830
09-9	Za	23	1447
10-9	Zo	23	1437

Figuur 63: gegevens strandtent.

Invoer gegevens

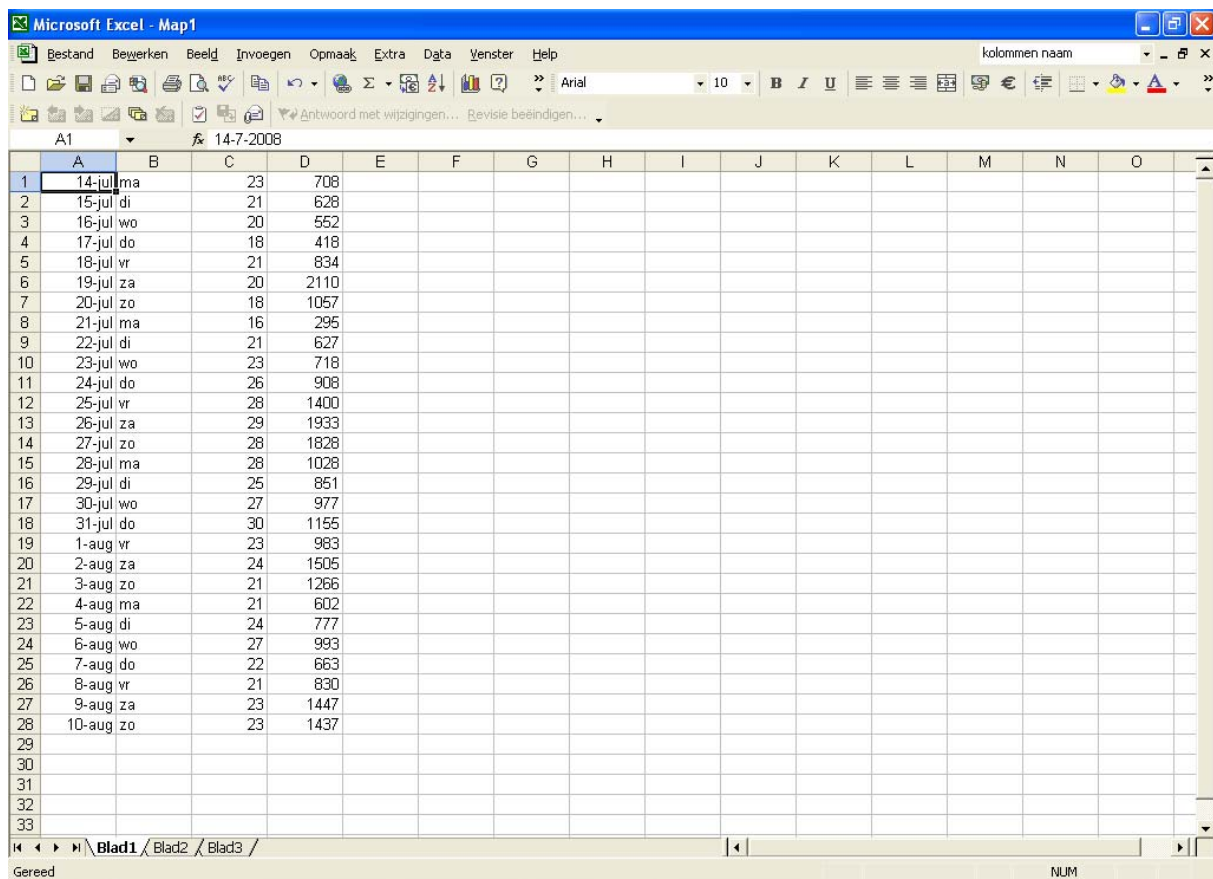
Nadat je Excel hebt geopend, verschijnt een leeg werkblad, waarbij de rijen genummerd zijn met de getallen 1, 2, 3etc. en de kolommen met letters worden aangegeven: A, B, C etc.

Voer om te beginnen de data in kolom A in. Om tekst of een getal in een cel in te voeren klik je met de linker muisknop op de cel of loop je er met de cursor heen. Staat er in een cel al iets ingevoerd en wil je de invoer iets aanpassen, dan selecteer je de cel en klik je met de linker muisknop op de werkbalk bovenin het scherm. De celinhoud staat ook in deze werkbalk en is hierin aan te passen. Met *enter* bevestig je de wijziging.

Invoer: 14-7 *enter*
 15-7 *enter*
 enz.

Als je deze notatie gebruikt, begrijpt Excel dat je data bedoelt en komt er 14-jul in de cel te staan.

Op dezelfde manier voer je in de kolommen B, C en D de dag, de temperatuur en het aantal verkochte drankjes in. Als het goed is, ziet je document er dan uit als in figuur 64



Figuur 64: gegevens strandtent in Excel.

Grafieken maken

Om in Excel een grafiek te maken met één variabele op de horizontale as en de andere op de verticale as, moet je eerst beide variabelen selecteren, waarbij je begint met de variabele voor de horizontale as. Klik hiervoor met de linker muisknop op de cel *14-jul* en houdt de linker muisknop ingedrukt. Sleep de muis vervolgens naar beneden tot de cel *10-aug* en laat pas daar de linker muisknop weer los. Als het goed is zie je dat alle gegevens in de eerste kolom gekleurd zijn. Houd vervolgens de *ctrl*-toets op het toetsenbord ingedrukt en selecteer op precies dezelfde manier de kolom met daarin de verkoopcijfers. Door de *ctrl*-toets ingedrukt te houden, blijft ook de selectie van de eerste kolom bestaan.



Figuur 65: wizard grafieken.

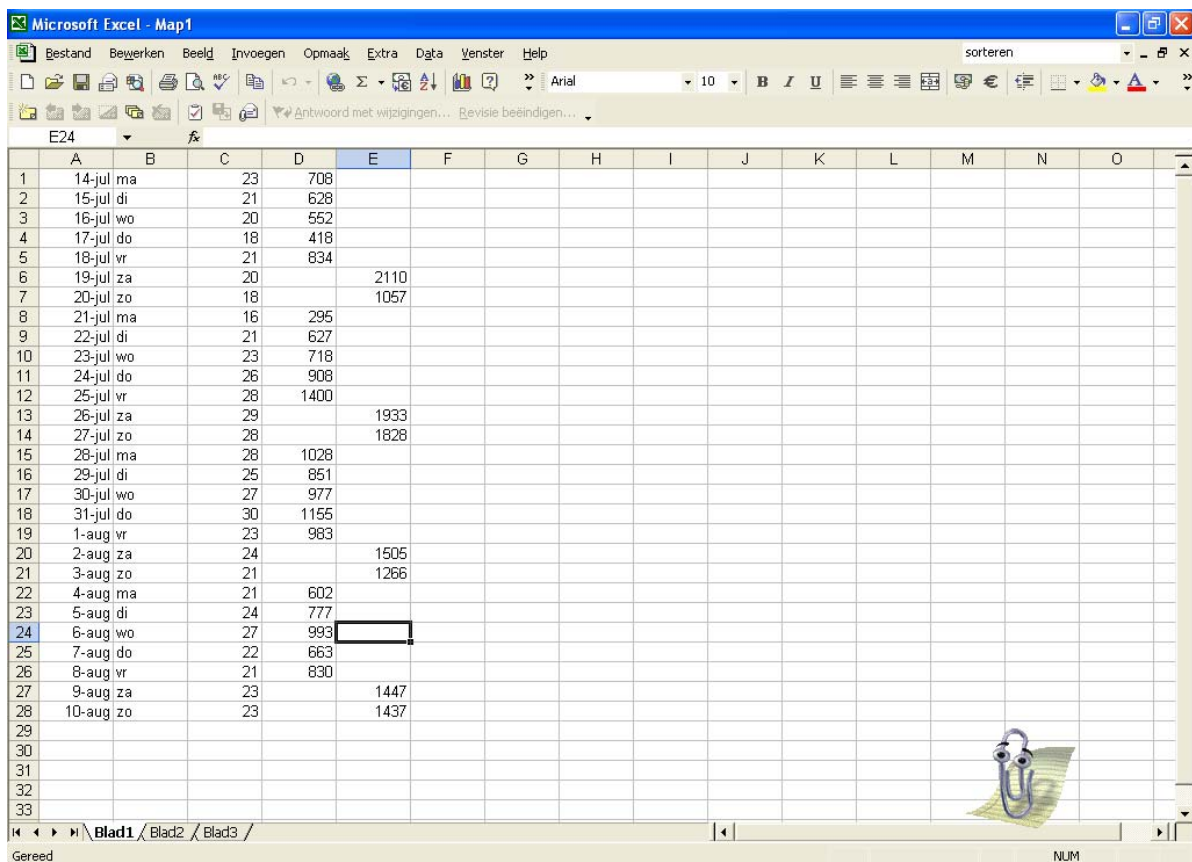
Druk vervolgens met de linker muisknop op *Wizard grafieken*:

Excel helpt je vervolgens in een aantal stappen bij het maken van een nette grafiek. Denk eraan dat je in de eerste stap kiest voor '*Spreiding*', anders krijg je een ander type grafiek dan een spreidingsdiagram. Als je in de 4e stap kiest voor '*Als nieuw blad in*', dan komt het spreidingsdiagram als apart tabblad in het Excelbestand te staan.

Als je het nog niet hebt gedaan, is het nu zeker tijd het gemaakte Excelbestand op te slaan!

Op precies dezelfde manier kun je een spreidingsdiagram maken met daarin de verkoop uitgezet tegen de temperatuur. Is het spreidingsdiagram hetzelfde als het spreidingsdiagram dat je gemaakt heb bij opdracht 22?

Bij opdracht 23 werd gevraagd om de meetwaarden uit het weekend te markeren. Ook dit is in Excel mogelijk, maar daarvoor moeten de gegevens van de doordeweekse dagen in een andere kolom staan dan die uit het weekend zoals is weergegeven in figuur 66.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	14-jul	ma	23	708											
2	15-jul	di	21	628											
3	16-jul	wo	20	552											
4	17-jul	do	18	418											
5	18-jul	vr	21	834											
6	19-jul	za	20		2110										
7	20-jul	zo	18		1057										
8	21-jul	ma	16	295											
9	22-jul	di	21	627											
10	23-jul	wo	23	718											
11	24-jul	do	26	908											
12	25-jul	vr	28	1400											
13	26-jul	za	29		1933										
14	27-jul	zo	28		1828										
15	28-jul	ma	28	1028											
16	29-jul	di	25	851											
17	30-jul	wo	27	977											
18	31-jul	do	30	1155											
19	1-aug	vr	23	983											
20	2-aug	za	24		1505										
21	3-aug	zo	21		1266										
22	4-aug	ma	21	602											
23	5-aug	di	24	777											
24	6-aug	wo	27	993											
25	7-aug	do	22	663											
26	8-aug	vr	21	830											
27	9-aug	za	23		1447										
28	10-aug	zo	23		1437										
29															
30															
31															
32															
33															

Figuur 66: opsplitsing gegevens weekend en week.

Je kunt de waarden die je wilt verplaatsen uit kolom D het beste selecteren en verslepen. Als je nu een grafiek wilt maken waarin je de meetwaarden uit het weekend kunt onderscheiden van die door de week, moet je achter elkaar kolom A, D en E selecteren waarbij je de *ctrl*-toets ingedrukt houdt. Bij kolom D en E moet je ook de lege cellen selecteren, zodat de drie geselecteerde kolommen even lang zijn. Via de grafiekenwizard krijg je nu een grafiek met twee gegevensreeksen, die door kleur en vorm van elkaar te onderscheiden zijn.

Berekenen pmcc

Om de pmcc te berekenen plaatsen we eerst alle waarden uit kolom E weer terug in kolom D.

Allereerst willen we de regressielijn voor het weekend te berekenen. Hiervoor moeten we eerst zorgen dat we de waarden van het weekend bijeen hebben staan.

Maak hiervoor eerst, ergens in het werkblad, dus bijvoorbeeld in kolom A onder de daar al ingevoerde gegevens, een kolom cellen met ma, di, wo, do, vr, za en zo aan en selecteer deze cellen:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	15-jul	di	21	628											
6	22-jul	di	21	627											
7	29-jul	di	25	851											
8	5-aug	di	24	777											
9	16-jul	wo	20	552											
10	23-jul	wo	23	718											
11	30-jul	wo	27	977											
12	6-aug	wo	27	993											
13	17-jul	do	18	418											
14	24-jul	do	26	908											
15	31-jul	do	30	1155											
16	7-aug	do	22	663											
17	18-jul	vr	21	834											
18	25-jul	vr	28	1400											
19	1-aug	vr	23	983											
20	8-aug	vr	21	830											
21	19-jul	za	20	2110											
22	26-jul	za	29	1933											
23	2-aug	za	24	1505											
24	9-aug	za	23	1447											
25	20-jul	zo	18	1057											
26	27-jul	zo	28	1828											
27	3-aug	zo	21	1266											
28	10-aug	zo	23	1437											
29															
30	ma														
31	di														
32	wo														
33	do														
34	vr														
35	za														
36	zo														
37															

Figuur 67: hulplijstje met dagen van de week.

Ga vervolgens naar *Extra, opties* en kies voor '*Aangepaste lijst*'. Klik hier op *importeren*, en druk op *ok*. Deze handeling zorgt ervoor dat Excel de lijst met de dagen van de week kent om daarop te sorteren, in de volgorde waarin jij ze hebt ingevoerd. Dit hulplijstje kun je nu van het werkblad verwijderen.

Selecteer vervolgens alle ingevoerde cellen en ga naar *Data, sorteren*. Klik hier op *opties* en kies als sorteervolgorde de door jou aangemaakte lijst. Druk vervolgens op *ok* en kies bij *Sorteren op* voor kolom B. Dat is immers de kolom met daarin de weekdays. Door nogmaals op *ok* te drukken, wordt de tabel in het werkblad aangepast: de maandagen staan bovenaan, gevolgd door de dinsdagen, enz. Wil je de sortering op dag van de week

weer ongedaan maken, selecteer dan eerst alle meetgegevens, ga dan opnieuw naar *Data*, *selecteren* en kies voor *Selectie op kolom A*. De waarden worden weer op volgorde van datum gezet. De gegevens die voor ons van belang zijn staan in de cellen vanaf rij 22. Dit zijn de gegevens van de zaterdagen en zondagen, behalve zaterdag de 19e, die in verband met de beachparty werd weggestreept.

Ga vervolgens met de cursor op cel E22 staan en voer daar in:
`=C22-GEMIDDELDE(C$22:C$28)`

Deze formule geeft aan dat je als uitkomst in cel E22 de waarde van cel C22 wilt hebben min het gemiddelde van alle waarden uit cel C22 t/m C28. De \$-tekens zijn toegevoegd om deze formule zo meteen naar de rest van kolom E te kunnen kopiëren. Als het goed is krijg je na op *enter* te hebben gedrukt, 5,285714 als uitkomst in cel E22.

Een dergelijke berekening wil je in de hele kolom E hebben staan, zonder deze berekening in elke cel in te hoeven voeren. Selecteer daarvoor om te beginnen cel E22. Ga vervolgens met de cursor precies in de rechter benedenhoek van deze cel staan. Let daarbij goed op de vorm van de cursor. Als je in de juiste positie staat, is dat een +. Houd vanaf dat moment de linker muistoets ingedrukt en sleep de cursor recht naar beneden, tot cel E22 t/m cel E28 geselecteerd is en laat dan de cursor los. Als het goed is komt in elke geselecteerde cel een getal te staan.

Om te zien wat er precies is gebeurd, selecteren we cel E24. In de werkbalk zien we de formule die in cel E24 terecht is gekomen:
`=C24-GEMIDDELDE(C$22:C$28)`

In cel E22 hebben we verwezen naar cel C22, nu wordt er keurig naar cel C24 verwezen. Toch wordt het gemiddelde nog steeds bepaald over cel C22 t/m C28. Dat deze nummering niet ook is veranderd komt door het gebruik van de \$-tekens.

Als we nu de waarden in kolom C zouden veranderen, rekent Excel dat automatisch door in kolom E. We zouden dus alle temperaturen kunnen vervangen door de temperaturen in september, die Excel vervolgens automatisch doorrekent. Probeer maar eens, maar zorg dat uiteindelijk de juiste temperaturen weer in de tabel staan.

Op precies dezelfde wijze berekenen we in kolom F de drankverkoop op een dag min de gemiddelde drankverkoop. In cel F22 voer je in:
`=D22-GEMIDDELDE(D$22:D$28)`

Deze formule breid je vervolgens weer uit over de hele kolom F. In kolom G berekenen we het product van de kolommen E en F. Hiervoor voer je in cel G22 in:

`=E22*F22`

Deze formule breid je weer over de hele kolom G uit.

Eerst laten we Excel de covariantie berekenen, waarvoor de formule is:

$$\text{cov}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n}$$

Zet bijvoorbeeld in cel B30 de afkorting cov. Je kunt dan altijd herkennen welke berekening erachter staat. In cel C30 zet je de volgende formule:

`=SOM(G22:G28)/7`

In deze cel komt de waarde van de covariantie te staan.

Dan laten we Excel de Pearson correlatiecoëfficiënt berekenen, waarvoor de formule is:

$$R = \frac{\text{cov}(x, y)}{\sigma_x \cdot \sigma_y}$$

In cel B31 zet je R , in cel C31 de bijbehorende formule:

`=C30/(STDEVP(C22:C28)*STDEVP(D22:D28))`

Zo staat in cel C31 de berekende waarde van de pmcc.

Je hebt hiermee geleerd hoe je in Excel voor grote aantallen meetwaarden de pmcc kunt berekenen. Handmatig is vervolgens de vergelijking van de regressielijn eenvoudig op te stellen.

Bijlage 5 URL lijst

URL1	Science Linx school CO2-web, Rijksuniversiteit Groningen http://www.rug.nl/sciencelinx/schoolco2web/index
------	---