

## 2.5 Totaalbeeld

### Samenvatten

Je hebt nu alle theorie van het onderwerp **Data verwerken** doorgewerkt. Het is nu tijd om een overzicht over het geheel te krijgen.

### Begrippenlijst

- discrete, continue variabele — dotplot — klassenindeling, klassenbreedte, klassenmidden
- centrummaten: gemiddelde, modus en mediaan — spreidingsmaten: spreidingsbreedte en interkwartielafstand — boxplot
- symmetrische, scheve, meertoppige verdelingen — cumulatieve frequenties of somfrequenties — cumulatief frequentiepolygoon
- de standaardafwijking of standaarddeviatie — de normale verdeling

### Activiteitenlijst

- discrete en continue variabelen van elkaar onderscheiden — dotplots interpreteren — klassenindelingen maken
- frequentieverdelingen samenvatten met behulp van centrummaten en spreidingsmaten — een boxplot tekenen en interpreteren
- frequentieverdelingen typeren — een cumulatief frequentiepolygoon tekenen en interpreteren
- de standaardafwijking berekenen — gemiddelde en standaardafwijking schatten vanuit een gegeven normale verdeling

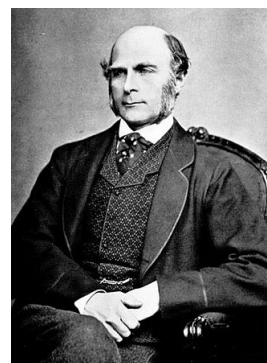
### Achtergronden

Met de toepassing van de kansrekening op het verzekeringswezen ontstond ook voor het eerst de behoefte tot het bijhouden van bevolkingsgegevens. En zo werd een klimaat geschapen voor de ontwikkeling van de statistiek. In 1662 publiceerde de Brit John Graunt zijn 'Natural and Political Observations', waarin een eerste statistische analyse voorkwam van de wekelijkse lijst van sterfgevallen in en om London, de zogenaamde 'Bills of mortality'. En in 1693 kwam de astronoom Edmond Halley met een levensverwachtingstabel gebaseerd op de sterftcijfers van de stad Breslau.

In de achttiende eeuw zette zich deze statistische traditie verder voort. Daarnaast vonden in de kansrekening de nodige ontwikkelingen plaats: de normale verdeling werd bedacht door de Britse wiskundige De Moivre en door de Belg Adolphe Quetelet in het begin van de negentiende eeuw toegepast op de sociale statistiek. De echte start van de **mathematische statistiek** is echter toe te schrijven aan Francis Galton (1822 - 1911) en vond plaats aan het einde van de negentiende eeuw.

**Francis Galton** (1822 - 1911) was een neef van Charles Darwin, de bioloog die de evolutietheorie opzette. Hij paste statistische methoden toe bij de analyse van sociale gegevens en erfelijke eigenschappen. Hij dacht dat de normale verdeling de mate van variatie van fysieke eigenschappen aangaf. Hij werkte met het begrip **standaarddeviatie** als maat voor de spreiding van de normale verdeling. In tegenstelling tot Quetelet dacht hij niet zozeer in termen van 'foute afwijkingen van het juiste gemiddelde', maar als noodzakelijke verscheidenheid in het licht van de evolutietheorie.

Galton paste zijn statistische theorieën vooral toe op de eugenetica, de studie van erfelijke eigenschappen.



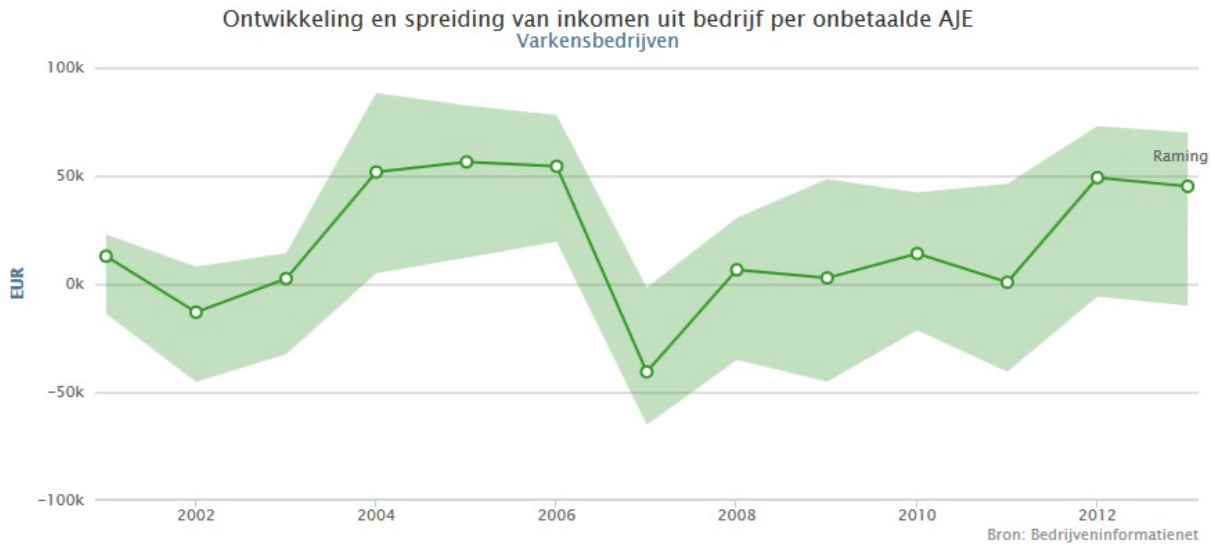
Figuur 1

## Testen

### Opgave 1

Bekijk de grafiek over ontwikkeling en spreiding van inkomen uit varkensbedrijven.

De groene lijn stelt het gemiddelde inkomen voor. De groene strook stelt de groep van 30% bedrijven voor die onder of boven het gemiddelde zitten. Een arbeidskracht die per jaar 2000 uur of meer werkt, wordt gezien als 1 arbeidsjaareenheid (AJE).

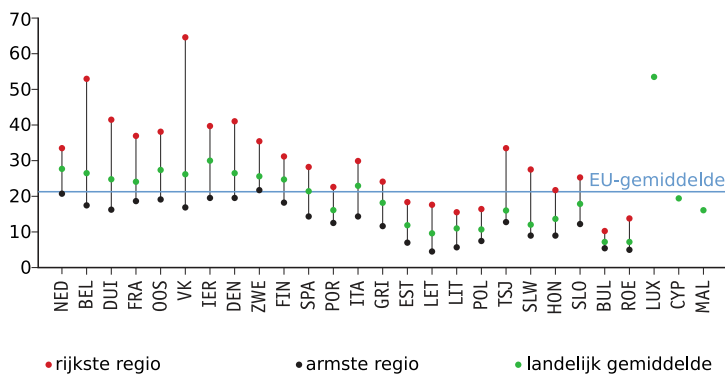


**Figuur 2**

- Bekijk de grafiek in het jaar 2012. Schat hoeveel procent van de varkensbedrijven een negatief inkomen had. Verwacht men voor 2013 dat dit percentage toe- of afneemt?
- Schat hoeveel procent van de bedrijven in 2007 een negatief inkomen had.
- Het verschil in inkomen tussen de 30%-bovengrens en de 30%-ondergrens is hier de spreidingsmaat. Schat de spreiding in 2000, 2005 en 2010.

### Opgave 2

Bekijk het diagram over de regionale spreiding van het bruto regionaal product per inwoner. Het betreft regio's in Europa. Het rode (bovenste) bolletje van de drie bolletjes geeft de rijkste regio per land, het groene (middelste) bolletje de gemiddelde regio, het zwarte (onderste) bolletje de armste regio. Op de verticale as vind je bedragen in duizenden euro/jaar.



**Figuur 3**

- Welke spreidingsmaat zie je hier weergegeven?
- In welk land is de spreiding het grootst?

- c Waaron zeggen deze verschillen in dit diagram niets over de verschillen tussen de inwoners in een land?
- d Waaron is van Luxemburg, Malta en Cyprus alleen het landelijk gemiddelde gegeven?

**Opgave 3**

Uit onderzoek van het gemengde boerenbedrijf bleek het houden van kippen een belangrijke rol te spelen bij het tot stand komen van het inkomen van deze boeren. Daarom werd de boeren gevraagd het aantal kippen op hun bedrijf op te geven.

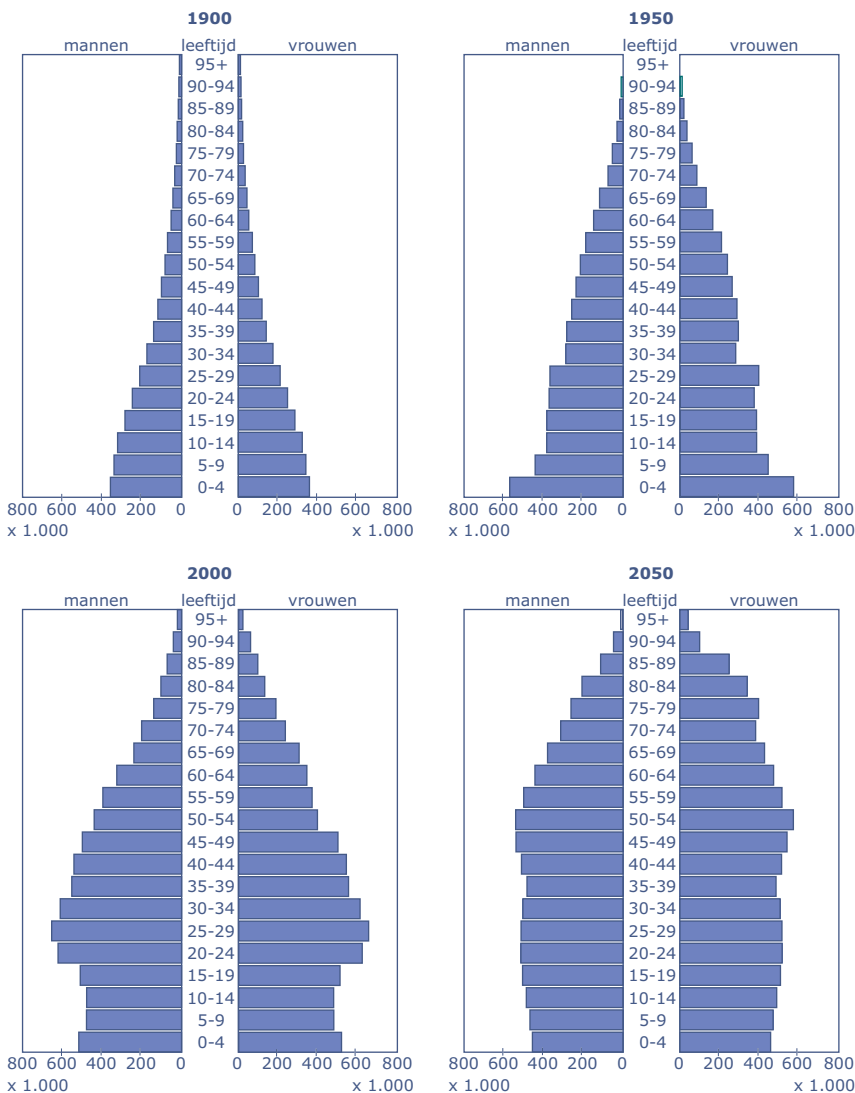
- a Waaron kun je geen dotplot maken van deze gegevens?
- b Schat de centrummaten mediaan en gemiddelde. Onderbouw je schatting.
- c Schat de drie spreidingsmaten. Onderbouw je schatting.
- d Hoeveel procent van de bedrijven heeft meer dan 160 kippen?
- e Hoeveel procent van de bedrijven heeft minder dan 41 kippen?
- f Hoe kun je de verdeling karakteriseren zonder er een diagram van te maken?

aantal kippen	aantal bedrijven
1 - 10	5
11 - 20	12
21 - 30	19
31 - 40	24
41 - 50	33
51 - 60	52
61 - 70	69
71 - 80	75
81 - 90	108
91 - 100	120
101 - 110	123
111 - 120	101
121 - 130	85
131 - 140	79
141 - 150	60
151 - 160	43
161 - 170	21
171 - 180	9
181 - 190	4
191 - 200	2

Tabel 1

**Opgave 4**

Bekijk de vier leeftijdsdiagrammen van Nederland.



Bron: Bevolkingsprognose Midden Variant

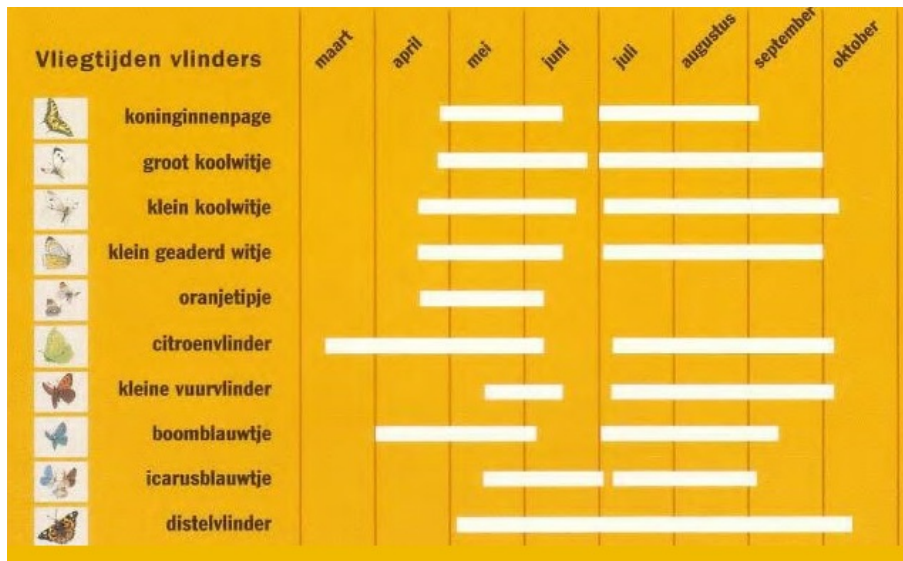
Figuur 4

- a Wat voor soort diagrammen zijn dit?

- b Hoeveel kinderen van 0 – 4 jaar waren er naar schatting in 1900, 1950 en 2000?
- c Waar vind je de kinderen die in 1950 in de klasse 0 – 4 jaar zaten terug in de piramides van 2000 en 2050?
- d Noem in elk van de vier diagrammen de modale klasse. Wat valt je op?
- e Hoe onderbouw je de uitspraak met getallen dat het percentage 85-plussers sinds 1900 flink is toegenomen en nog verder zal toenemen?

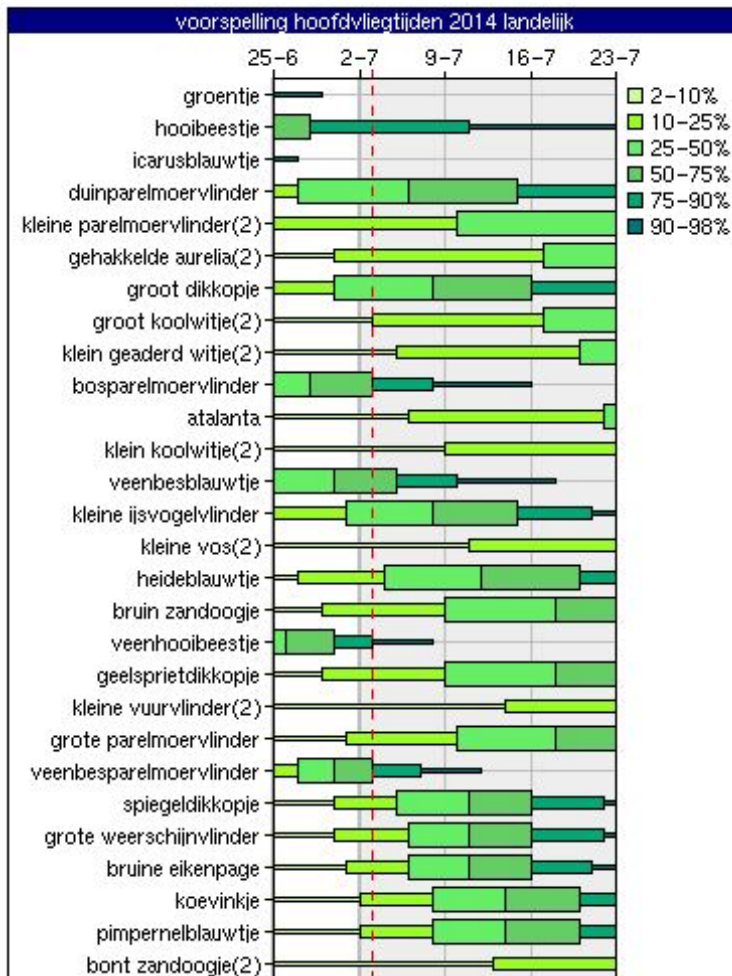
### Opgave 5

De Vlinderstichting geeft op verschillende manieren informatie over vlinders. Bekijk de tabel met vliegtijden van de meest voorkomende vlinders.



Tabel 2

De Vlinderstichting doet op haar site een voorspelling over de vliegtijd van vlinders die binnenkort moeten gaan vliegen. Per soort (en per generatie) wordt middels een tijdbalk gepresenteerd wanneer de hoofdvliegtijd waarschijnlijk zal beginnen. Het eerste (smalste) deel van de balk start op de dag waarop wordt verwacht dat 2% van de totale generatie gaat vliegen en eindigt op 10%. De zes volgende balken staan voor respectievelijk 2 - 10%, 10 - 25%, 25 - 50%, 50 - 75%, 75 - 90% en 90 - 98% van de populatie.



Figuur 5

- a Voor wie is, denk je, de tabel bedoeld? En voor wie de tijdbalk?
- b De vorm van de voorspellingen lijkt op die van boxplots. Welk verschil is er met boxplots?
- c De tabel geeft soms twee periodes aan bij de meest voorkomende vlinders. Wat is de betekenis daarvan? En hoe kun je dat terugzien in de tijdbalk?
- d Wat is de betekenis van de box in de tijdbalk van de duinparelmoervlinder?
- e Vergelijk de vliegtijden van het icarusblauwtje in de tabel met de tijdbalk. Wat valt je op? Wat zou dat kunnen betekenen?

## Toepassen

### Opgave 6: Reactiesnelheden

Verzin een manier om iemands reactiesnelheid te meten. Maak vervolgens een histogram van de verdeling van zijn reactiesnelheden. Doe dit voor meerdere personen en zet de gegevens in de computer.

- a Maak voor een aantal personen een histogram van de reactiesnelheden.
- b Beschrijf de vorm van deze histogrammen.
- c Bereken de centrummaten en spreidingsmaten die zinvol zijn en leg uit waarom ze dat zijn.



© 2021

Deze paragraaf is een onderdeel van het Math4All wiskundemateriaal.

Math4All stelt het op prijs als onvolkomenheden in het materiaal worden gemeld en ideeën voor verbeteringen in de content of dienstverlening kenbaar worden gemaakt.

Email: [f.spijkers@math4all.nl](mailto:f.spijkers@math4all.nl)

Met de Math4All maatwerkdienst kunnen complete readers worden samengesteld en toetsen worden gegenereerd. Docenten kunnen bij [a.f.otten@xs4all.nl](mailto:a.f.otten@xs4all.nl) een gratis inlog voor de maatwerkdienst aanvragen.

---