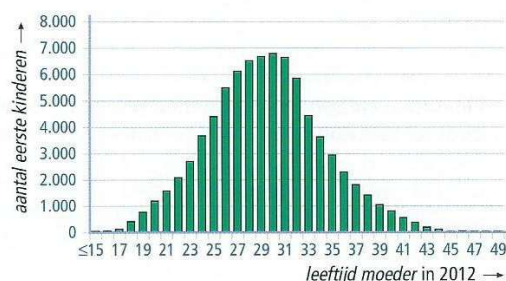
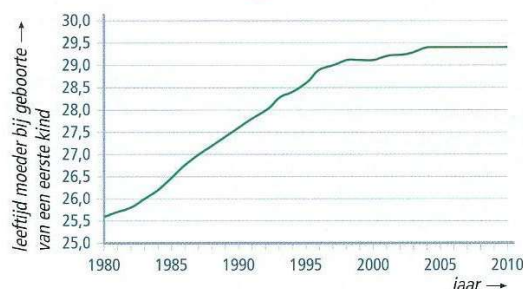


5-6

Data analyse



- 23 Hierboven zie je resultaten van een onderzoek naar de toename van de Nederlandse bevolking.
- Van welke populatie heeft de onderzoeker gegevens verzameld?
 - Om welke variabelen gaat het bij de data?
 - Zijn deze data continu of discreet?
 - Wat zou de probleemstelling geweest kunnen zijn bij dit onderzoek?

Theorie

Bij het doen van statistisch onderzoek kun je de volgende stappen onderscheiden:

- > Probleemstelling
- > Formuleren van vraag bij probleemstelling
- > Het verzamelen van data
 - Welke populatie, welke data, hoeveel data?
- > Data verwerken
 - Hoe ga je de data weergeven om de juiste conclusies te kunnen trekken.
- > Presenteren en conclusies trekken

- 24 Bij het bevolkingsonderzoek naar darmkanker bij Nederlanders tussen de 55 en 75 jaar oud wordt er gekeken of er bloed in de ontlasting zit. Als de uitslag positief (wel bloed aanwezig) is, komt er een vervolgonderzoek. Als de uitslag negatief (geen bloed aanwezig) is, is dit geen garantie dat je ook geen darmkanker hebt. In de tabel kun je zien dat van de 1000 onderzochte mensen er 76 ziek zijn, terwijl dat niet uit de uitslag van het onderzoek bleek.
- Hoe zullen de tegenstanders van het bevolkingsonderzoek deze data gebruiken?
 - En hoe zullen de voorstanders van het bevolkingsonderzoek deze data gebruiken?

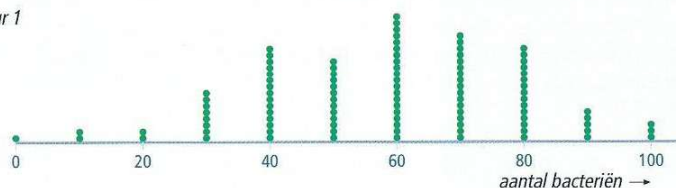
	uitslag	
	negatief	positief
ziek	76	37
niet ziek	874	13

- 25 East Coast Milk (ECM) is een bedrijf in de VS dat elke dag melk bij honderd boeren in de regio ophaalt. Voordat de melk in de transporttank gaat, wordt bij elke boer een monster van de melk genomen. In het laboratorium van ECM wordt de melk onderzocht op het voorkomen van bacteriële vervuiling. Daartoe wordt in elk monster het aantal bacteriën per centiliter (cL) geteld.
- De Public Health Service (PHS), verantwoordelijk voor de kwaliteitsbewaking, stelt als eis een maximum van 80 bacteriën per cL. Hieronder staat een frequentietabel van de gevonden waarden in de honderd monsters van een dag. Het gemiddelde is gelijk aan 59 bacteriën per cL en de standaardafwijking is 20,3 bacteriën per cL. Er zijn bij deze gegevens nog twee andere representaties gemaakt, zie figuur 1 en figuur 2.

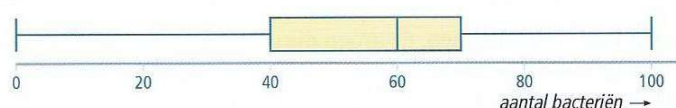
tabel

aantal bacteriën	frequentie
0	1
10	1
20	2
30	8
40	15
50	13
60	20
70	17
80	15
90	5
100	3
totaal	100

figuur 1



figuur 2



- Welke van deze drie representaties is/zijn niet geschikt om het aantal overschrijdingen van de PHS-norm vast te stellen? Licht je antwoord toe.
- De totale hoeveelheid melk die elke dag wordt opgehaald, heet een dagproductie. In de honderd monsters zitten gemiddeld 59 bacteriën per cL. Is het redelijk om aan te nemen dat dit ook het gemiddelde van de dagproductie is?
- Het laboratorium let goed op uitschieters omdat die op bijzonderheden kunnen wijzen. Vaak zijn dat meetfouten. Het laboratorium gebruikt de volgende vuistregel: Een uitschieter ligt buiten het interval mediaan $\pm 1,5 \times$ interkwartielafstand. Ga na hoeveel uitschieters er zijn. Bij welke zou het om een meetfout kunnen gaan?
- De twee monsters met de laagste waarde worden opnieuw gemeten. Dat levert in beide gevallen een waarde van 50 bacteriën per cL op. Beredeneer, zonder een berekening te maken, dat met deze nieuwe meetwaarden het gemiddelde groter wordt en de standaardafwijking kleiner wordt.
- Beredeneer wat er verandert aan de mediaan.
(ontleend aan de syllabus havo wiskunde A 2017)