

Examen HAVO

2023

tijdvak 2
tijdsduur 3 uur

wiskunde A

Dit examen bestaat uit 23 vragen.

Voor dit examen zijn maximaal 80 punten te behalen.

Voor elk vraagnummer staat hoeveel punten met een goed antwoord behaald kunnen worden.

Als bij een vraag een verklaring, uitleg of berekening vereist is, worden aan het antwoord meestal geen punten toegekend als deze verklaring, uitleg of berekening ontbreekt.

Geef niet meer antwoorden (redenen, voorbeelden e.d.) dan er worden gevraagd. Als er bijvoorbeeld twee redenen worden gevraagd en je geeft meer dan twee redenen, dan worden alleen de eerste twee in de beoordeling meegeteld.

Vuistregels voor de grootte van het verschil van twee groepen

2×2 kruistabel $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, met $phi = \frac{ad - bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}$,

waarin a , b , c en d absolute aantallen zijn.

- als $phi < -0,4$ of $phi > 0,4$, dan zeggen we “het verschil is groot”,
- als $-0,4 \leq phi < -0,2$ of $0,2 < phi \leq 0,4$, dan zeggen we “het verschil is middelmatig”,
- als $-0,2 \leq phi \leq 0,2$, dan zeggen we “het verschil is gering”.

Maximaal verschil in cumulatief percentage ($\max V_{cp}$)

(met steekproefomvang $n > 100$)

- als $\max V_{cp} > 40$, dan zeggen we “het verschil is groot”,
- als $20 < \max V_{cp} \leq 40$, dan zeggen we “het verschil is middelmatig”,
- als $\max V_{cp} \leq 20$, dan zeggen we “het verschil is gering”.

Effectgrootte $E = \frac{\bar{X}_1 - \bar{X}_2}{\frac{1}{2}(S_1 + S_2)}$, met \bar{X}_1 en \bar{X}_2 de steekproefgemiddelden

($\bar{X}_1 \geq \bar{X}_2$), S_1 en S_2 de steekproefstandaardafwijkingen

- als $E > 0,8$, dan zeggen we “het verschil is groot”,
- als $0,4 < E \leq 0,8$, dan zeggen we “het verschil is middelmatig”,
- als $E \leq 0,4$, dan zeggen we “het verschil is gering”.

Twee boxplots vergelijken

- als de boxen¹⁾ elkaar niet overlappen, dan zeggen we “het verschil is groot”,
- als de boxen elkaar wel overlappen en een mediaan van een boxplot buiten de box van de andere boxplot ligt, dan zeggen we “het verschil is middelmatig”,
- in alle andere gevallen zeggen we “het verschil is gering”.

noot 1 De ‘box’ is het interval vanaf het eerste kwartiel tot en met het derde kwartiel.

Betrouwbaarheidsintervallen

Het 95%-betrouwbaarheidsinterval voor de populatieproportie is

$$p \pm 2 \cdot \sqrt{\frac{p(1-p)}{n}}, \text{ met } p \text{ de steekproefproportie en } n \text{ de steekproefomvang.}$$

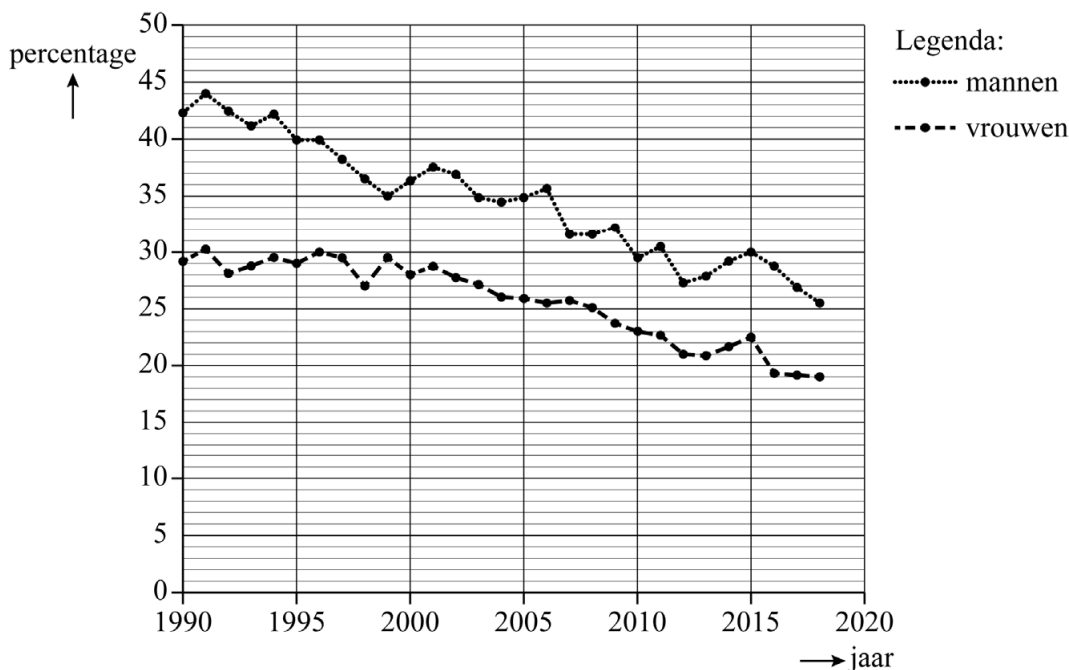
Het 95%-betrouwbaarheidsinterval voor het populatiegemiddelde is

$$\bar{X} \pm 2 \cdot \frac{S}{\sqrt{n}}, \text{ met } \bar{X} \text{ het steekproefgemiddelde, } n \text{ de steekproefomvang en } S \text{ de steekproefstandaardafwijking.}$$

Minder rokers

Figuur 1 komt uit het rapport Volksgezondheid Toekomst Verkenning 2018 van het RIVM¹⁾. In deze figuur zijn voor Nederland het percentage van de mannen dat rookt en het percentage van de vrouwen dat rookt op 1 juli voor de jaren 1990 tot en met 2018 weergegeven.

figuur 1 Percentage rokers 18 jaar of ouder



Deze hele opgave gaat alleen over mannen en vrouwen van 18 jaar of ouder.

Op 1 juli 2015 telde Nederland 6 618 000 mannen en 6 853 000 vrouwen van 18 jaar of ouder.

- 4p 1 Bereken met behulp van figuur 1 hoeveel procent van de rokers op 1 juli 2015 man was. Geef je antwoord in hele procenten.

Voor de periode vanaf 2018 heeft het RIVM een formule opgesteld om de ontwikkeling van het percentage van de mannen dat rookt te voorspellen. Men gaat uit van een lineaire afname volgens de volgende formule:

$$P_m = -0,61t + 25,93 \quad (\text{formule 1})$$

Hierin is P_m het percentage van de mannen dat rookt en t het aantal jaren vanaf 1 juli 2018.

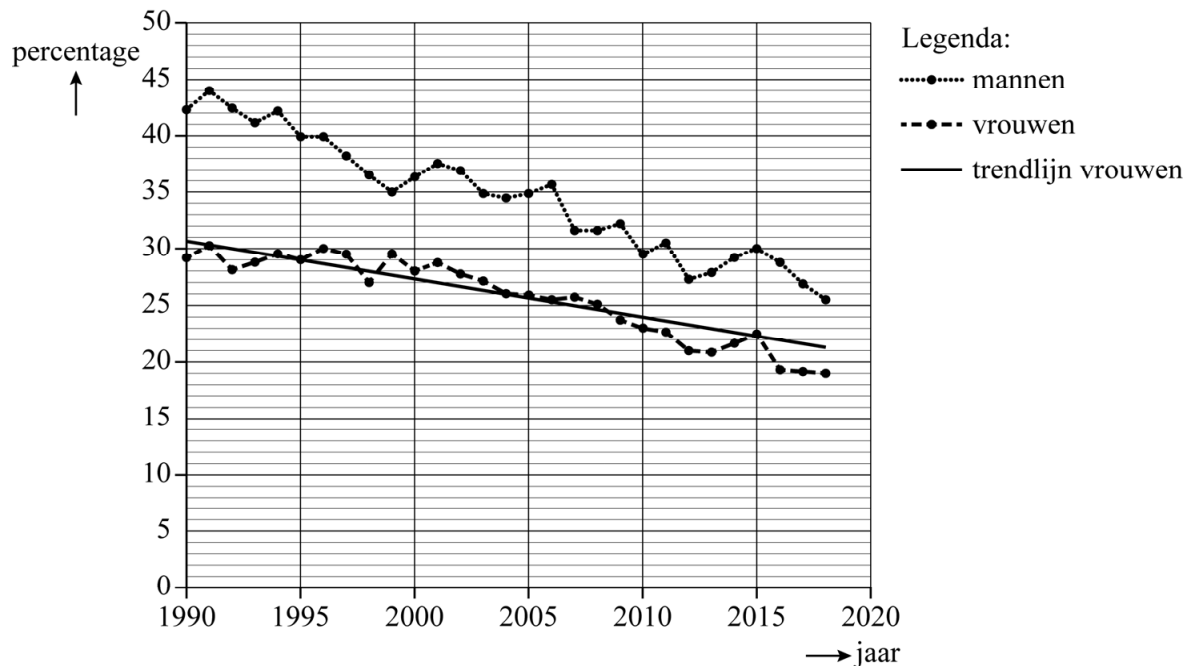
noot 1 RIVM = Rijksinstituut voor Volksgezondheid en Milieu

Het RIVM beweert dat volgens formule 1 het percentage van de mannen dat rookt voor de start van het jaar 2040 gehalveerd is ten opzichte van 1 juli 2018.

3p **2** Onderzoek of deze bewering juist is.

Figuur 2 is ontstaan door in figuur 1 de trendlijn te tekenen voor het percentage van de vrouwen dat rookt.

figuur 2 Percentage rokers 18 jaar of ouder



De percentages van de vrouwen die roken op 1 juli in 1995 en 2004 liggen precies op de trendlijn.

Met deze gegevens kan een formule voor de trendlijn worden opgesteld:

$$P_v = -0,33t + 30,67 \quad (\text{formule 2})$$

Hierin is P_v het percentage van de vrouwen dat rookt en t het aantal jaren vanaf 1 juli 1990.

De percentages van de vrouwen die roken in 1995 en 2004 zijn gehele getallen.

3p **3** Bereken met deze percentages de getallen in formule 2 in drie decimalen.

Het percentage van de mannen dat rookt, kan vanaf 1 juli 2018 worden berekend met formule 1. Het percentage van de vrouwen dat rookt, kan vanaf 1 juli 1990 worden berekend met formule 2.

5p **4** Onderzoek in welk jaar het percentage van de mannen dat rookt en het percentage van de vrouwen dat rookt gelijk zullen zijn.

Voedselconsumptie

In de periode 2012-2014 is onderzocht wat kinderen en volwassenen aten en dronken. Deze voedselconsumptiepeiling is uitgevoerd bij personen die toen tussen de 1 en 79 jaar oud waren en in Nederland woonden.

Om een goed beeld te krijgen van wat er gedurende het jaar gegeten en gedronken werd, is de gegevensverzameling gelijkmatig verdeeld over de dagen van het jaar.

- 2p 5 Geef een voorbeeld van welk vertekend beeld je zou kunnen krijgen wanneer de gegevensverzameling niet gelijkmatig over de dagen van het jaar verdeeld is.

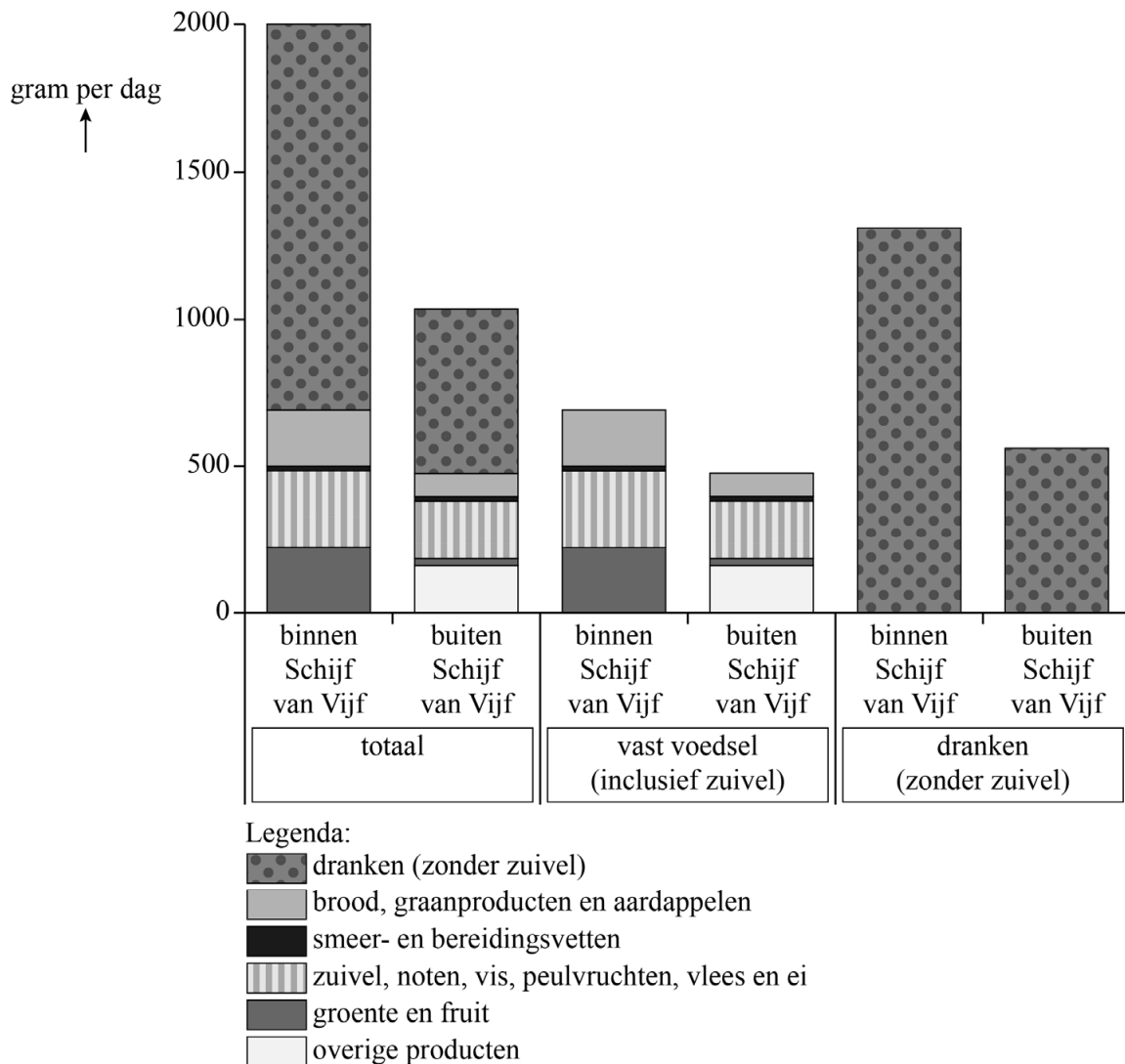
In de voedselconsumptiepeiling werd onderscheid gemaakt tussen vast voedsel (inclusief zuivel) en dranken (zonder zuivel). Daarnaast werd ook gekeken of de voedingsmiddelen binnen of buiten de **Schijf van Vijf** vallen. De Schijf van Vijf is een hulpmiddel dat wordt gebruikt in de voorlichting over gezonde voeding en bevat vijf categorieën met aanbevolen voedingsmiddelen:

- dranken (zonder zuivel)
- brood, graanproducten en aardappelen
- smeer- en bereidingsvetten
- zuivel, noten, vis, peulvruchten, vlees en ei
- groente en fruit

Buiten de Schijf van Vijf vallen bijvoorbeeld koekjes, cola, chips, witbrood en vla, maar ook groenten met toegevoegd zout in pot of fruit op siroop in blik.

Op de volgende bladzijde staat op de verticale as van figuur 1 de gemiddelde voedselconsumptie per persoon per dag (in gram).

figuur 1 Gemiddelde voedselconsumptie per persoon



4p **6** Bereken met behulp van figuur 1 welk percentage van de totale voedselconsumptie per persoon per dag bestond uit dranken (zonder zuivel). Geef je antwoord in hele procenten.

Figuur 1 kwam tot stand op basis van een steekproef van 2237 personen. Ga uit van de volgende aannames:

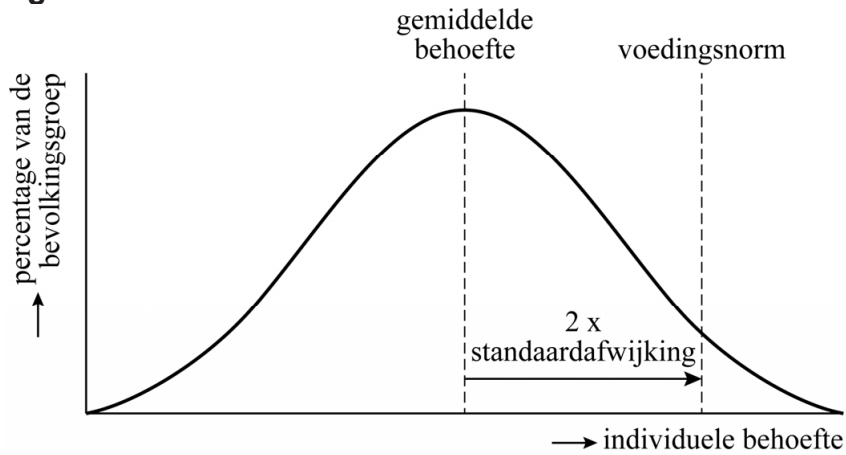
- de steekproef is aselekt en representatief
- de populatie bestaat uit in Nederland wonende personen met een leeftijd tussen de 1 en 79 jaar
- de gemiddelde voedselconsumptie per persoon is normaal verdeeld
- het bijbehorende 95%-betrouwbaarheidsinterval van het populatiegemiddelde van de gemiddelde voedselconsumptie per persoon is 25 gram breed

We kunnen dan de standaardafwijking berekenen die bij de steekproef hoort.

4p **7** Bereken deze standaardafwijking. Geef je antwoord in gehele grammen.

De behoefte aan voedingsstoffen verschilt per persoon en is normaal verdeeld. Een **voedingsnorm** is de aanbevolen hoeveelheid voedingsstof die genoeg is voor bijna alle mensen. Zie figuur 2.

figuur 2



Omdat figuur 2 gebaseerd is op de aanname dat de individuele behoefte normaal verdeeld is, kan er concreter worden aangegeven wat er bedoeld wordt met 'bijna alle mensen'.

2p 8 Geef het percentage dat hier bedoeld wordt met 'bijna alle mensen'.

Ga verder op de volgende pagina.

Aantal stemmen per liedje

De Top 2000 is een jaarlijks terugkerend radioprogramma. Mensen kunnen vooraf stemmen op liedjes. De 2000 populairste liedjes worden achter elkaar gedraaid. Alleen de aantallen stemmen op het liedje op positie 1 en het liedje op positie 2000 worden openbaar gemaakt. Men vermoedt dat de volgende formule een goede schatting geeft voor het aantal stemmen op een liedje in de ranglijst:

$$S = \frac{E}{1 + \frac{E(N-1)}{2000 \cdot L}} \text{ voor } N = 1, 2, 3, \dots, 2000 \quad (\text{formule 1})$$

Hierin is:

- N de positie in de ranglijst;
- S het geschatte aantal stemmen op het liedje dat op positie N in de ranglijst staat;
- E het werkelijke aantal stemmen op het liedje dat op positie 1 in de ranglijst staat;
- L het werkelijke aantal stemmen op het liedje dat op positie 2000 in de ranglijst staat.

Met formule 1 kun je een schatting maken van het aantal stemmen op het liedje dat op positie 1 in de ranglijst staat.

- 2p **9** Laat zien dat dit geschatte aantal overeenkomt met het werkelijke aantal stemmen op dit liedje.

Het liedje met de meeste stemmen komt op positie 1, het liedje dat daarna de meeste stemmen heeft, komt op positie 2, enzovoort. Logischerwijs zou dus het geschatte aantal stemmen S op een liedje lager moeten worden naarmate N groter wordt.

- 3p **10** Laat met een redenering zien dat formule 1 deze eigenschap heeft.

Voor de Top 2000 van 2006 geldt:

- het liedje op positie 1 had 10 000 stemmen;
- het liedje op positie 2000 had 150 stemmen.

Hiermee kan formule 1 herleid worden tot:

$$S = \frac{300\,000}{N + 29} \quad (\text{formule 2})$$

4p **11** Geef deze herleiding.

Over de Top 2000 van 2006 is meer bekend dan gebruikelijk: er zijn in totaal 1,5 miljoen stemmen uitgebracht.

3p **12** Bereken hoeveel procent van de stemmen in totaal op de top 3 van deze Top 2000 is uitgebracht volgens formule 2. Geef je antwoord in hele procenten.

Omdat het aantal stemmen per liedje normaal gesproken niet openbaar gemaakt wordt, vraagt men zich af wat men bij de Top 2000 doet als een liedje precies evenveel stemmen krijgt als een ander liedje.

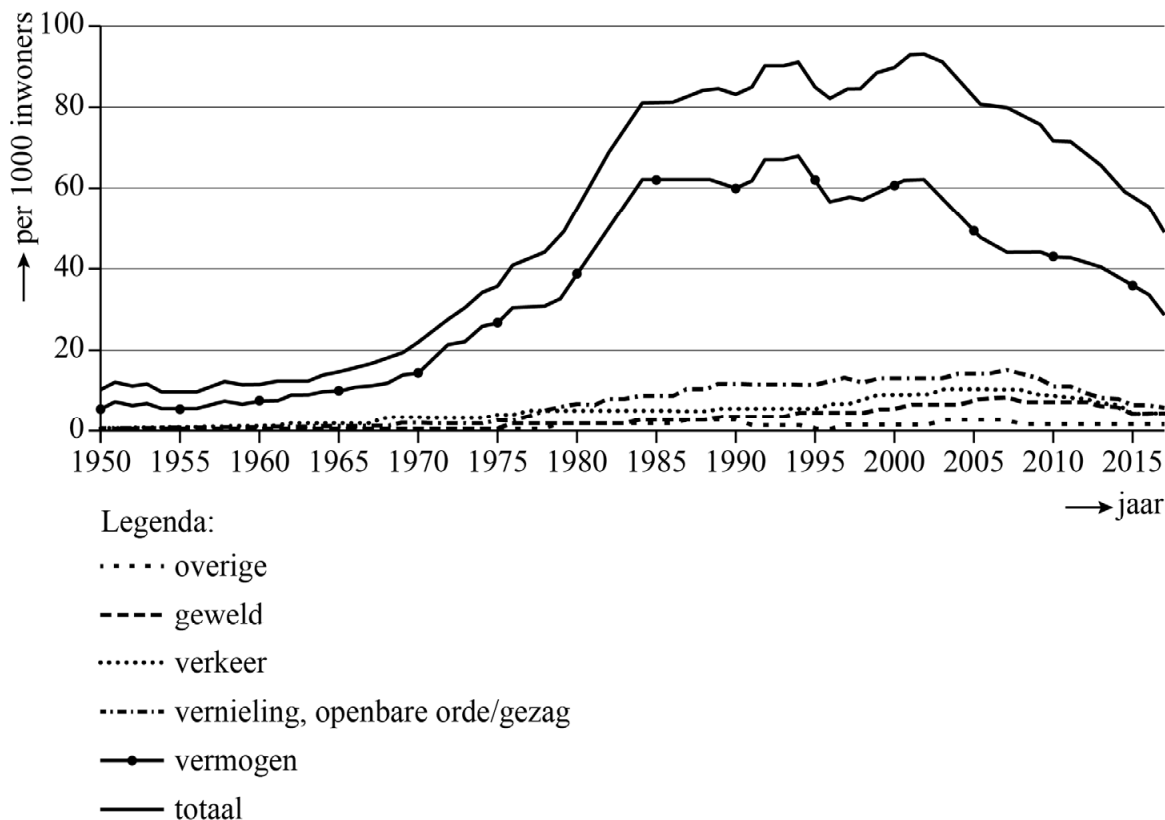
Als we met formule 2 het geschatte aantal stemmen per liedje berekenen en dat vervolgens afronden op een geheel getal, dan kan het zijn dat er na afronding liedjes zijn die evenveel stemmen hebben gekregen.

3p **13** Bereken hoeveel liedjes er in de Top 2000 van 2006 zijn die volgens formule 2 na afronding 200 stemmen zouden hebben gekregen.

Geregistreerde misdrijven

In de onderstaande figuur staat voor Nederland het aantal geregistreerde misdrijven per duizend inwoners per jaar. Er is ook een uitsplitsing gemaakt naar verschillende types misdrijven, waaronder vermogensmisdrijven en geweldsmisdrijven.

figuur



Van 1960 tot 1985 steeg het totaal aantal geregistreerde misdrijven van 13 naar 81 per duizend inwoners. Neem aan dat deze stijging exponentieel verliep.

- 4p **14** Bereken met deze gegevens het jaarlijkse groeipercentage.
Geef je antwoord in één decimaal.

In 1985 was bijna 80% van het totaal aantal geregistreerde misdrijven een vermogensmisdrijf.

- 3p **15** Beredeneer met behulp van de figuur, zonder een berekening te maken, of dit percentage in 2002 hoger of lager was dan in 1985.

Sinds het begin van deze eeuw daalt het totaal aantal geregistreerde misdrijven, met name door een daling van het aantal geregistreerde vermogensmisdrijven.

Een voorbeeld van een vermogensmisdrijf is een woninginbraak. Iemand vermoedt dat er in het begin van deze eeuw bij een kleiner percentage van de woningen werd ingebroken dan aan het eind van de vorige eeuw. Om hier zicht op te krijgen is in 2004 in een aselechte steekproef uit de woningen gekeken bij welk percentage er in dat jaar was ingebroken. De resultaten kunnen worden vergeleken met die van een aselechte steekproef uit de woningen in 1998. Zie de tabel.

tabel

	1998	2004
aantal woningen	3437	1456
percentage woningen waarbij is ingebroken	2,15	1,58

De tabel laat zien dat het steekproefpercentage woningen waarbij is ingebroken in 2004 lager is dan in 1998. We gaan deze gegevens op twee manieren met elkaar vergelijken.

- 3p **16** Onderzoek of het steekproefpercentage uit 2004 in het 95%-betrouwbaarheidsinterval ligt van het populatiepercentage in 1998.

In 2004 is het steekproefpercentage woningen waarbij is ingebroken lager dan in 1998.

- 4p **17** Onderzoek met behulp van het formuleblad of het verschil in steekproefpercentage woningen waarbij is ingebroken in de twee steekproeven groot, middelmatig of gering is.

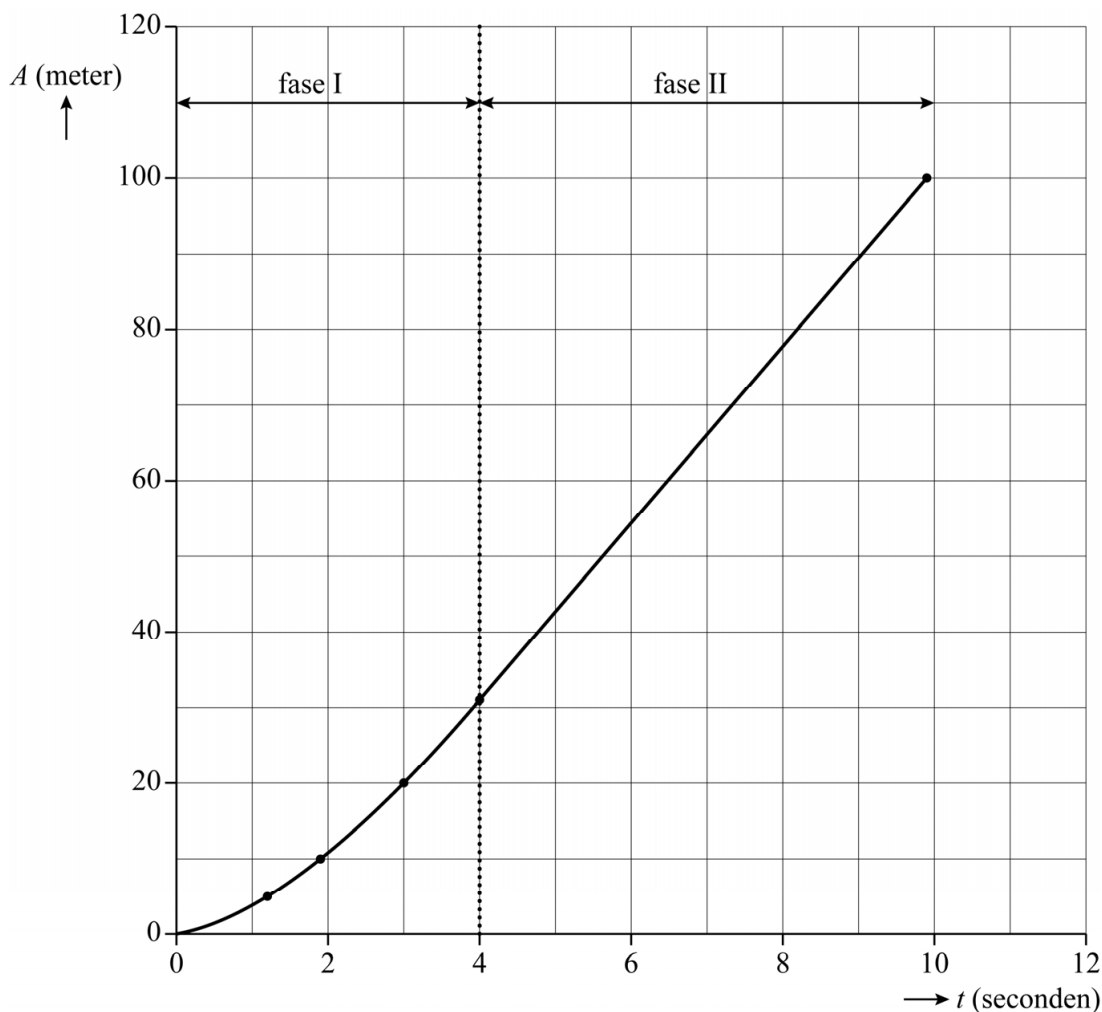
De 100 meter

In deze opgave bekijken we een wiskundig model van een sprint over 100 meter hardlopen. In dit model gaan we uit van 2 fases:

- fase I: zo snel mogelijk op topsnelheid komen;
- fase II: de topsnelheid vasthouden tot het einde van de sprint.

Van een bepaalde topatleet staat in figuur 1 de afgelegde afstand A in meter uitgezet tegen de tijd t in seconden.

figuur 1



Deze atleet legt zijn sprint in 9,9 seconden af. Na 4,0 seconden bereikt hij zijn topsnelheid; dan heeft hij al 31 meter afgelegd.

Je kunt met deze gegevens zijn topsnelheid berekenen; deze is afgerond 11,7 m/s.

3p 18 Bereken deze topsnelheid in twee decimalen.

Het eerste deel van de grafiek is (toenemend) stijgend. Je kunt je afvragen of er sprake is van exponentiële groei.

In de tabel staan een aantal waarden uit fase I van die sprint.

tabel

t (s)	1,2	1,9	3,0
A (m)	5	10	20

- 3p **19** Onderzoek met behulp van de tabelwaarden of er sprake kan zijn van exponentiële groei.

Het verband tussen de afgelegde afstand A in meter en de tijd t in seconden in fase I kan voor de atleet benaderd worden door de formule:

$$A = 3,8 \cdot t^{1,5}$$

De atleet moet in het begin nog op snelheid komen. Hij zal daarom ook langer doen over de eerste 25 meter dan over de laatste 25 meter van de 100 meter. Bij de laatste 25 meter is hij immers al op topsnelheid.

- 5p **20** Bereken hoeveel seconden hij hier langer over doet. Geef je antwoord in één decimaal.

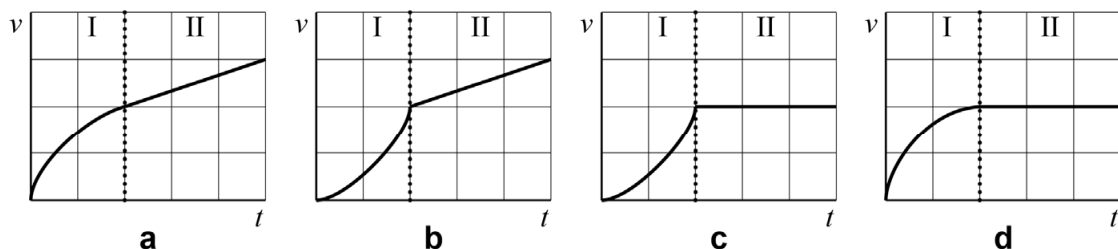
Er is voor fase I ook een formule die de snelheid v in m/s vrij goed beschrijft. Deze is van de vorm:

$$v = k \cdot \sqrt{t}$$

- 3p **21** Bereken met behulp van zijn topsnelheid de waarde van k voor deze atleet.

In de vragen hiervoor is de hele sprint van een topatleet over 100 meter beschreven. In figuur 2 staan vier globale grafieken (a, b, c, d), waarbij de snelheid v uitgezet is tegen de tijd t .

figuur 2



Er is één grafiek die het best past bij de sprint van de topatleet.

- 3p **22** Welke grafiek is dat? Licht je antwoord toe.

Testresultaten en afhakers

Wanneer artsen een nieuwe behandeling testen op patiënten, is het mogelijk dat er tijdens het onderzoek een aantal patiënten afhaakt (de **afhakers**). Dit komt bijvoorbeeld omdat ze de behandeling te belastend vinden.

Het succespercentage P van een nieuwe behandeling kan op twee verschillende manieren berekend worden:

exclusief afhakers

Hierbij worden patiënten die afhaken niet meegeteld.

In formulevorm:

$$P_{\text{exclusief}} = \frac{G}{G + N} \cdot 100$$

inclusief afhakers

Hierbij worden patiënten die afhaken wel meegeteld.

In formulevorm:

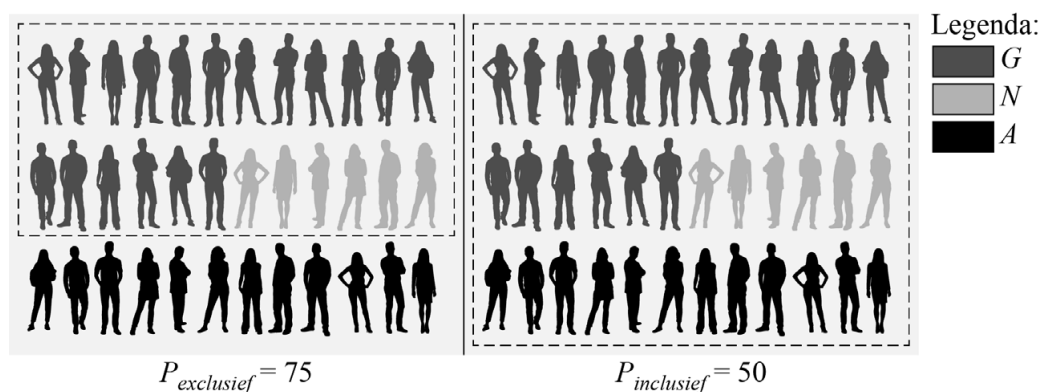
$$P_{\text{inclusief}} = \frac{G}{G + N + A} \cdot 100$$

Hierin geldt:

- G is het aantal patiënten dat de behandeling volhoudt en geneest
- N is het aantal patiënten dat de behandeling volhoudt en niet geneest
- A is het aantal patiënten dat afhaakt.

Het wel of niet meetellen van het aantal patiënten dat afhaakt in het testresultaat zal tot verschillende succespercentages leiden, zie figuur.

figuur



In het voorbeeld in de figuur is het absolute verschil tussen de twee succespercentages gelijk aan 25.

Het wel of niet meetellen van het aantal patiënten dat afhaakt bij de berekening van het succespercentage kan tot forse verschillen leiden. Men heeft afgesproken dat het succespercentage alleen berekend wordt als het percentage patiënten dat afhaakt niet hoger is dan 20%.

We kijken nu naar een onderzoek waarbij het percentage patiënten dat afhaakt gelijk was aan het maximale percentage van 20%. Bij dit onderzoek waren er uiteindelijk 240 patiënten die de behandeling hadden volgehouden en was het absolute verschil tussen de twee succespercentages 16.

Er kan nu onderzocht worden wat het aantal patiënten moet zijn geweest dat de behandeling heeft volgehouden, maar niet is genezen.

7p 23 Voer dit onderzoek uit en bepaal dit aantal.

Bronvermelding

Een opsomming van de in dit examen gebruikte bronnen, zoals teksten en afbeeldingen, is te vinden in het bij dit examen behorende correctievoorschrift.