

Oplossingenblad — Proefexamen 2 Statistiek

voor de ouder: volledige uitwerking + socratische hulpvragen + veelgemaakte fouten

Hoe gebruik je dit oplossingenblad?

- **Tijdens het examen:** laat je dochter eerst zelfstandig werken (50 min).
- **Stokt ze?** Gebruik de **socratische hulpvragen** per vraag, van algemeen naar specifiek — zo verklar je het antwoord *niet*.
- **Bij correctie:** tussenstappen (formule, rangnummer, invullen) tellen volwaardig mee.
- **Dit proefexamen legt extra nadruk op:** soort variabele (nominaal/ordinaal), scheve verdelingen vanaf een grafiek, 'precies vs cumulatief' en 'aantal vs procent'.
- **Veelgemaakte fouten** staan onderaan elke oplossing — gebruik ze als checklist.

Score-overzicht

Vraag	Topic	Punten	Behaald
V1	Soort variabele	4	___ / 4
V2	Frequentietabel: precies vs cumulatief	5	___ / 5
V3	Centrummaten & kwartielen	6	___ / 6
V4	Boxplot tekenen	4	___ / 4
V5	Boxplot lezen: % vs aantal	4	___ / 4
V6	Scheve verdeling vanaf grafiek	4	___ / 4
V7	Misleiding, cluster & uitschieter	3	___ / 3
V8	Variantie & standaardafw.	4	___ / 4
V9	Spreidingsdiagram	3	___ / 3
V10	Correlatiecoëfficiënt	3	___ / 3
	TOTAAL	40	___ / 40

Vraag 1. [Soort variabele] (4 ptn)

Geef voor elke enquêtevraag het soort variabele **zo nauwkeurig mogelijk** (kwantitatief discreet/continu óf kwalitatief nominaal/ordinaal). Verklaar telkens kort.

- a) Uit hoeveel leden bestaat je gezin?
- b) In welke **mate** ben je tevreden over de schoolmaaltijden? (zeer ontevreden → zeer tevreden)
- c) Wat is je lievelingssport?
- d) Hoe groot ben je (in cm)?

Volledige uitwerking

- a) **Kwantitatief discreet** — een telling met gehele getallen waarmee je kunt rekenen.
- b) **Kwalitatief ordinaal** — een eigenschap **MÉT** rangorde (zeer ontevreden → zeer tevreden is geordend).
- c) **Kwalitatief nominaal** — een eigenschap **ZÓNDER** rangorde (sporten kun je niet zinvol ordenen van klein naar groot).
- d) **Kwantitatief continu** — een meting (lengte).

Socratische hulpvragen — gebruik in volgorde, stop zodra ze weer verder kan:

1. Kun je met het antwoord rekenen (een getal) of is het een eigenschap/naam?
2. Als het een eigenschap is: zit er een rangorde in de antwoorden, ja of nee?
3. Als het een getal is: tel je (discreet) of meet je (continu)?

Veelgemaakte fouten (instinkers)

- b) 'in welke mate / tevredenheid' nominaal noemen — het is ordinaal (er is rangorde).
- c) 'lievelingssport / voorkeur' ordinaal noemen — geen rangorde, dus nominaal.
- Maar één niveau geven ('discreet' i.p.v. 'kwantitatief discreet') → halve punt.

Vraag 2. [Frequentietabel: precies vs cumulatief] (5 ptn)

Het aantal huisdieren per gezin werd geteld ($n = 40$). Vul de tabel aan: rf_i (op 3 decimalen), $rf_i\%$, caf_i en crf_i (op 3 decimalen).

x_i	af_i	rf_i	$rf_i\%$	caf_i	crf_i
0	6				
1	18				
2	9				
3	4				
4	2				
5	1				
n	40	1	100 %		

- a) Hoeveel **procent** van de gezinnen heeft **1 of 2** huisdieren?
b) Hoeveel **gezinnen** hebben **ten hoogste 2** huisdieren?
c) Hoeveel **gezinnen** hebben **meer dan 2** huisdieren?

Volledige uitwerking

rf: 0,150 ; 0,450 ; 0,225 ; 0,100 ; 0,050 ; 0,025.

rf%: 15% ; 45% ; 22,5% ; 10% ; 5% ; 2,5%.

caf: 6 ; 24 ; 33 ; 37 ; 39 ; 40. **crf:** 0,150 ; 0,600 ; 0,825 ; 0,925 ; 0,975 ; 1,000.

a) '1 of 2' = enkel die categorieën: 45% + 22,5% = **67,5%**. (Niet 82,5%: dat is crf t.e.m. 2 en telt de 0-categorie mee!)

b) 'ten hoogste 2' = cumulatief: caf bij 2 = **33 gezinnen**.

c) 'meer dan 2' = 40 - 33 = **7 gezinnen** (of 4 + 2 + 1).

Socratische hulpvragen — gebruik in volgorde, stop zodra ze weer verder kan:

1. Hoe bereken je rf en crf uit af en n? Tel je de caf/crf cumulatief op?
2. a) Staat er 'precies 1 of 2' of 'ten hoogste 2'? Welke tel je op, welke lees je cumulatief af?
3. c) Hoe gebruik je n en de caf om 'meer dan 2' te vinden?

Veelgemaakte fouten (instinkers)

- a) de cumulatieve 82,5% gebruiken (telt de 0-categorie mee) i.p.v. 67,5% op te tellen.
- crf niet strikt laten stijgen (elke crf > de vorige; de laatste = 1,000).
- 'ten hoogste 2' en 'meer dan 2' verwisselen.

Vraag 3. [Centrumsmaten & kwartielen] (6 ptn)

Voor dezelfde gegevens (aantal huisdieren, $n = 40$) hieronder: bereken het **gemiddelde**, de **modus**, de **mediaan**, **Q1** en **Q3**, en daarna **R** en **IKA**. Toon telkens je rangnummer en werkwijze.

x_i	0	1	2	3	4	5
af_i	6	18	9	4	2	1

Volledige uitwerking

$n = 6+18+9+4+2+1 = 40$.

Gemiddelde: $\bar{x} = (0 \cdot 6 + 1 \cdot 18 + 2 \cdot 9 + 3 \cdot 4 + 4 \cdot 2 + 5 \cdot 1) / 40 = 61 / 40 \approx 1,53$.

Modus: grootste af = 18 $\rightarrow$ **Mo** = 1.

Mediaan: rangnummer = $(40+1)/2 = 20,5 \rightarrow$ gemiddelde van de 20ste en 21ste waarde. caf: 0 $\rightarrow$ 6, 1 $\rightarrow$ 24, dus beide vallen bij waarde 1 $\rightarrow$ **Me** = 1.

Q1: mediaan van de onderste helft (plaats 1–20), rangnummer = 10,5 $\rightarrow$ 10de & 11de waarde = 1 $\rightarrow$ **Q1** = 1.

Q3: mediaan van de bovenste helft (plaats 21–40) = de 30ste & 31ste waarde. caf: 1 $\rightarrow$ 24, 2 $\rightarrow$ 33, dus beide = 2 $\rightarrow$ **Q3** = 2.

R = max – min = 5 – 0 = 5. **IKA** = Q3 – Q1 = 2 – 1 = 1.

Socratische hulpvragen — gebruik in volgorde, stop zodra ze weer verder kan:

1. Gemiddelde: deel je door 6 (aantal verschillende waarden) of door $n = 40$?
2. Mediaan: bereken eerst $(n+1)/2$. Is dat een kommagetal? Wat doe je dan? Gebruik de caf!
3. Q1 = mediaan van de onderste helft, Q3 van de bovenste. Lees de waarde af met de caf.

Veelgemaakte fouten (instinkers)

- Gemiddelde delen door 6 i.p.v. door 40.
- Mediaan/Q1/Q3 als kaal antwoord schrijven ZÓNDER rangnummer/werkwijze (kost punten, ook al klopt het eindantwoord!).
- R en IKA verwisselen.

Vraag 4. [Boxplot tekenen] (4 ptn)

Tien leerlingen behaalden volgende punten (op 20): **7 9 10 11 12 12 14 15 16 18**. Bepaal de 5 kengetallen (min, Q1, Me, Q3, max) en **teken de boxplot op schaal**.

Volledige uitwerking

$n = 10$ (even). Min = 7 ; Max = 18.

Me: rangnummer = $(10+1)/2 = 5,5 \rightarrow$ Me = $(12+12)/2 = 12$.

Q1: onderste helft (plaats 1–5): 7, 9, 10, 11, 12 $\rightarrow$ mediaan = 3de = 10.

Q3: bovenste helft (plaats 6–10): 12, 14, 15, 16, 18 $\rightarrow$ mediaan = 3de = 15.

Boxplot: rechthoek van 10 tot 15, streepje op 12, snorharen tot 7 en 18 (teken een schaal van 0 tot 20).

Socratische hulpvragen — gebruik in volgorde, stop zodra ze weer verder kan:

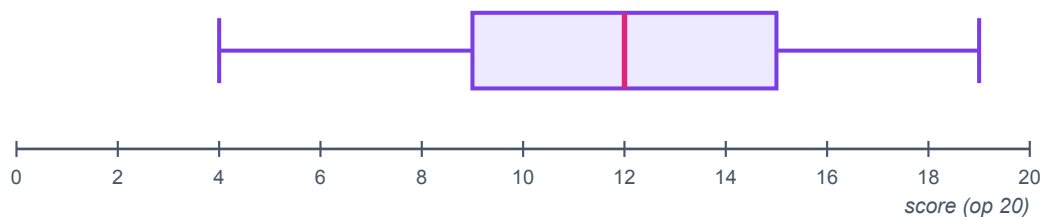
1. Is n even of oneven? Hoe bepaal je dan de mediaan?
2. Bij even n : hoe verdeel je de rij in twee helften voor Q1 en Q3?
3. Welke 5 kengetallen heb je nodig, en wat stelt de rechthoek voor?

Veelgemaakte fouten (instinkers)

- Bij even n de twee helften fout afbakenen.
- Het mediaanstreepje vergeten in de box.
- Niet netjes op schaal tekenen (meet- of tekenfout).

Vraag 5. [Boxplot lezen: % vs aantal] (4 ptn)

Onderstaande boxplot toont de toetscores (op 20) van een klas met **24 leerlingen**.



- Bepaal de variatiebreedte R en de interkwartielafstand IKA.
- Hoeveel **procent** van de leerlingen scoort hoger dan 15 (= Q3)?
- Hoeveel **leerlingen** scoren lager dan 9 (= Q1)?

Volledige uitwerking

Afgelezen: min = 4 ; Q1 = 9 ; Me = 12 ; Q3 = 15 ; max = 19.

- $R = 19 - 4 = 15$. IKA = $15 - 9 = 6$.
- Boven Q3 ligt **25%** van de leerlingen.
- Onder Q1 ligt 25% → 25% van 24 = **6 leerlingen** (let op: een aantal, geen %).

Socratische hulpvragen — gebruik in volgorde, stop zodra ze weer verder kan:

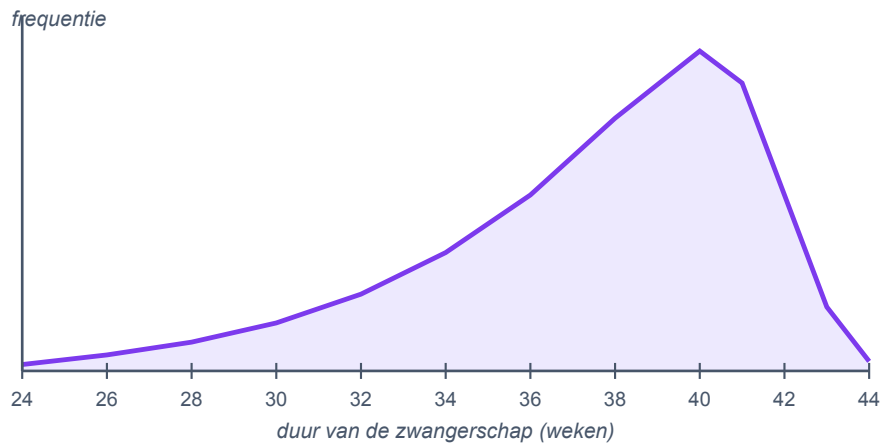
- Welke 5 kengetallen lees je af op een boxplot? Let op de schaal.
- Elk stuk van de boxplot bevat hoeveel procent van de gegevens?
- c) De vraag vraagt leerlingen, niet %: hoeveel is 25% van 24?

Veelgemaakte fouten (instinkers)

- Bij c) '25%' antwoorden i.p.v. een aantal leerlingen (aantal = % · n).
- Verkeerde kengetallen aflezen (schaal!).
- R en IKA verwisselen.

Vraag 6. [Scheve verdeling vanaf grafiek] (4 ptn)

De grafiek toont de duur van de zwangerschap (in weken) bij een grote groep vrouwen.



- a) Beschrijf de vorm van de verdeling **zo nauwkeurig mogelijk**.
- b) Licht het gemiddelde links of rechts van de top (modus)? Verklaar.
- c) Wat is groter: het gemiddelde of de mediaan?

Volledige uitwerking

- a) Een **linksscheve** (asymmetrische) verdeling: de top ligt rechts (rond 40 weken) en er is een lange **staart naar links** (de lagere waarden). *De naam volgt de staart, niet de top!*
- b) Het gemiddelde ligt **links** van de top: het wordt naar de kant van de staart getrokken (de lage waarden in de staart trekken het gemiddelde omlaag).
- c) Bij een linksscheve verdeling is het **gemiddelde kleiner dan de mediaan** (gemiddelde < mediaan < modus).

Socratische hulpvragen — gebruik in volgorde, stop zodra ze weer verder kan:

1. Aan welke kant ligt de lange, dunne staart — links of rechts?
2. De staart bepaalt zowel de naam als de kant waar het gemiddelde naartoe schuift.
3. Als het gemiddelde naar de lage waarden schuift, is het dan groter of kleiner dan de mediaan?

Veelgemaakte fouten (instinkers)

- De grafiek 'rechtsscheef' noemen omdat de top rechts ligt (kijk naar de STAART, niet de top).
- Een los getal gokken voor het gemiddelde i.p.v. de positie (links/rechts t.o.v. de mediaan) te geven.

Vraag 7. [Misleiding, cluster & uitschieter] (3 ptn)

Waar of niet waar? Verklaar telkens kort.

- a) 'Een grafiek die lichaamslengtes vergelijkt met een verticale as die bij 1,50 m begint (i.p.v. bij 0), is een misleidende grafiek.'
- b) 'Een cluster is één extreem grote of extreem kleine waarde.'

Volledige uitwerking

- a) **Waar.** Doordat de as niet bij 0 begint, lijken kleine verschillen veel groter dan ze zijn. (Bij figuurtjes/pictogrammen vertekent bovendien de breedte/oppervlakte het beeld.)
- b) **Niet waar.** Één losse, extreme waarde is een **uitschieter**. Een **cluster** is een *groep* waarnemingen die dicht bij elkaar liggen (met een gemeenschappelijk kenmerk).

Socratische hulpvragen — gebruik in volgorde, stop zodra ze weer verder kan:

1. a) Waar hoort de verticale as eerlijk te beginnen om verschillen correct weer te geven?
2. b) Hoe noem je één losse waarde die ver van de rest ligt?
3. b) Is een cluster één waarde of een groep waarden?

Veelgemaakte fouten (instinkers)

- b) cluster en uitschieter verwisselen.
- Enkel 'waar/niet waar' antwoorden zonder verklaring (verklaar altijd kort).

Vraag 8. [Variantie & standaardafw.] (4 ptn)

Vijf metingen van de middagtemperatuur: **12°, 14°, 16°, 18°, 20°**. Bereken de variantie σ^2 en de standaardafwijking σ (op 2 decimalen).

Volledige uitwerking

Gemiddelde: $\bar{x} = (12+14+16+18+20)/5 = 80/5 = 16^\circ$.
Afwijkingen: -4, -2, 0, 2, 4. **Kwadraten:** 16, 4, 0, 4, 16.
Variantie: $\sigma^2 = (16+4+0+4+16)/5 = 40/5 = 8$.
Standaardafwijking: $\sigma = \sqrt{8} \approx 2,83^\circ$.

Socratische hulpvragen — gebruik in volgorde, stop zodra ze weer verder kan:

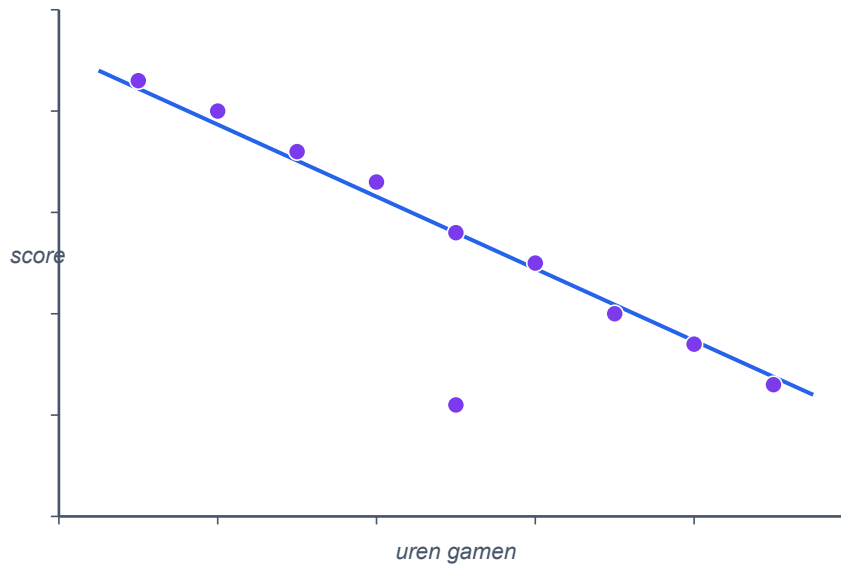
1. Welk getal heb je altijd eerst nodig voor de variantie?
2. Wat doe je met elke afwijking vóór je optelt? (denk aan 'kwadratische')
3. Hoe ga je van variantie naar standaardafwijking?

Veelgemaakte fouten (instinkers)

- Vergeten te kwadrateren (de afwijkingen optellen geeft 0!).
- De wortel vergeten bij de standaardafwijking.
- Delen door 4 i.p.v. door $n = 5$.

Vraag 9. [Spreidingsdiagram] (3 ptn)

Het spreidingsdiagram toont voor 10 leerlingen het aantal uren gamen per week (x) en hun gemiddelde toetsscore (y).



- Welk soort verband zie je? Beschrijf de richting.
- Is het verband sterk of zwak? Verklaar aan de hand van de puntenwolk.
- Welk punt zou je als een mogelijke uitschieter aanduiden?

Volledige uitwerking

- Een **negatief** verband: meer uren gamen gaat samen met een lagere score. De trendlijn **daalt**.
- De meeste punten liggen dicht bij de dalende rechte → een **vrij sterk** negatief lineair verband (op één punt na).
- Het punt dat duidelijk los onder de dalende trend ligt (veel lager dan verwacht).

Socratische hulpvragen — gebruik in volgorde, stop zodra ze weer verder kan:

- Gaan grote x-waarden samen met grote of met kleine y-waarden?
- Hoe dichters de punten bij de rechte, hoe ... het verband?
- Welk punt ligt duidelijk los van de algemene trend?

Veelgemaakte fouten (instinkers)

- Positief en negatief verband verwisselen (kijk naar de richting van de wolk).
- Sterkte verwarren met richting.
- De uitschieter op de trendlijn zoeken i.p.v. er ver vanaf.

Vraag 10. [Correlatiecoëfficiënt] (3 ptn)

Een leerling berekent met Excel de correlatiecoëfficiënt $r = -0,78$ voor een dataset.

- a) Welk soort lineair verband leid je hieruit af?
- b) Een tweede leerling gebruikt dezelfde gegevens, maar zet x en y omgekeerd op de assen. Welke waarde krijgt zij voor r? Verklaar.
- c) Na het verwijderen van één afwijkend meetpunt wordt $r = -0,91$. Wat besluit je over de invloed van uitschieters?

Volledige uitwerking

- a) $|r| = 0,78$ ligt dicht bij 1 en is negatief → een **vrij sterk negatief** lineair verband.
- b) r blijft **exact gelijk**: $-0,78$. Het omwisselen van x en y heeft geen invloed op de correlatiecoëfficiënt.
- c) Één afwijkend punt veranderde r van $-0,78$ naar $-0,91$. Besluit: **r is gevoelig voor uitschieters**.

Socratische hulpvragen — gebruik in volgorde, stop zodra ze weer verder kan:

- 1. Waar ligt r tussen, en wat betekent een waarde dicht bij 1 (of -1)?
- 2. b) Verandert de samenhang als je enkel de assen verwisselt?
- 3. c) Wat gebeurt er met r door één punt dat ver van de rest ligt?

Veelgemaakte fouten (instinkers)

- b) denken dat r van teken verandert of $1/r$ wordt.
- a) 'sterk' en 'zwak' verwisselen ($0,78$ is vrij sterk).
- Het minteken vergeten (het verband is negatief).