

# Wat zijn de overlevingskansen? Omgaan met gecensureerde data



**Wiskunde A-lympiade voorronde-opdracht  
28 november 2003**

De Wiskunde A-lympiade wordt gesponsord door Texas Instruments



# Werkwijzer bij de voorronde opdracht van de Wiskunde A-lympiade 2003/2004

## Werkwijzer

Deze Wiskunde A-lympiade opdracht bestaat uit 2 inleidende opdrachten en een slotopdracht.

### Algemene adviezen bij het werken aan deze opdracht

1. Lees eerst de volledige tekst van de opgave door zodat je weet wat jullie allemaal te doen staat.
2. Bewaak de tijd die jullie besteden aan de inleidende opgaven. Verdeel de taken waar mogelijk, en overleg als nodig.
3. Als je taken hebt verdeeld, bespreek dan voor je aan de slotopdracht begint de resultaten van de inleidende opdrachten met elkaar.
4. Bij de slotopdracht is het belangrijk dat je je advies goed onderbouwt. Dit is een belangrijk aspect in de beoordeling!
5. Zorg dat het resultaat een echt advies is met alles wat gevraagd wordt erin. Zorg dat alle ideeën en uitwerkingen die je van opdracht 1 en 2 hebt op één of andere manier verwerkt zijn in het advies bij de slotopdracht.

### Inleveren

Een volledig advies – product van de slotopdracht - dat als zelfstandig geheel los van de opdracht en los van het werk op de inleidende opgaven door de ‘klant’ te lezen is. De antwoorden van de inleidende opdrachten kun je als bijlage opnemen.

**Schrijf met zwarte pen of print A4 blaadjes in verband met de kopieerbaarheid! De jury kijkt kopiën na!**

### Beoordeling

Bij de beoordeling wordt onder andere gelet op:

- de helderheid van het beschreven advies, de onderbouwing en het uitgewerkte voorbeeld
- de volledigheid van het werk (zit alles erbij is alles gedaan?)
- het gebruik van wiskunde
- de gebruikte argumentatie en de verantwoording van gemaakte keuzes (hierbij kan de realiteitswaarde een rol spelen)
- de diepgang waarmee een en ander is gedaan
- manier van presenteren o.a de vorm - is het een echt advies? -, leesbaarheid, structuur - gebruik en functie van bijlagen enzovoort.

veel plezier en succes

## Wat zijn de kansen?

Arno Smit is een medisch specialist die veel met heroïneverslaafden werkt. Hij doet onderzoek naar het gebruik van een chemische vervanger voor heroïne. Hij wil nagaan wat het effect hiervan is op de levensverwachting van de verslaafden. Hiertoe volgt hij gedurende 10 jaar een groep zwaar verslaafden, die de chemische vervanger gebruiken.

De groep uit het onderzoek is niet constant gedurende die tien jaar: er overlijden verslaafden, er zijn er die niet meer op komen dagen, anderen verhuizen of haken om andere redenen af....

### *Methadonbus*



In de medische wetenschap is het gebruikelijk om de effectiviteit van een behandeling aan te geven met de *overlevingskans* na verloop van tijd.

In de tabel vind je de resultaten uit het onderzoek van dokter Smit.

Jaar sinds de start van het onderzoek	Aantal levende verslaafden in het onderzoek (aan het begin van het jaar)	Aantal overleden verslaafden in het onderzoek (gedurende het jaar)	Aantal verslaafden dat uit het onderzoek is verdwenen (gedurende het jaar)
[0,1> = het eerste jaar	138	9	14
[1,2>	115	11	10
[2,3>	94	15	5
[3,4>	74	10	7
[4,5>	57	9	9
[5,6>	39	7	11
[6,7>	21	3	5
[7,8>	13	1	7
[8,9>	5	2	1
[9,10>	2	1	1

Doordat er tussentijds verslaafden uit het onderzoek zijn verdwenen, zijn de gegevens niet volledig beschikbaar. In de statistiek noemt men een niet-volledige verzameling gegevens '*gecensureerde data*'. Hoewel dokter Smit dus met gecensureerde data moet werken, wil hij toch een zo betrouwbaar mogelijke uitspraak doen over het percentage verslaafden dat na vijf jaar nog in leven is.

## Opdracht 1

Veel specialisten redeneren als volgt:

*“Tel de mensen die uit het onderzoek verdwijnen (om een andere reden dan overlijden) tot en met het jaar dat je bestudeert niet mee in het onderzoek. Bekijk dan welk percentage van de overblijvende groep overleeft tot en met dat jaar. De overlevingskans na 5 jaar is dan in dit geval 42%”.*

Er zijn nog veel meer manieren om de overlevingskans te berekenen. Zo is er ook een methode die zegt:

*“De verdwenen groep personen uit het onderzoek mag je niet meetellen in het onderzoek. Die laat je dus helemaal weg. Van de overgebleven groep bepaal je het percentage dat binnen vijf jaar is overleden. De complementaire kans is dan de overlevingskans na 5 jaar en is dus in dit geval 21%.”*

Bedenk zelf nog minstens 2 andere methoden die gebruikt zouden kunnen worden. Geef de bijbehorende overlevingskansen, inclusief berekeningen. Geef ook aan welke methode je het meest betrouwbaar vindt, en waarom.

## Opdracht 2

Om meer greep te krijgen op de betrouwbaarheid van de methodes uit opdracht 1 bekijken we nu een dataset met *ongecensureerde data*, dat wil zeggen: er is *geen* sprake van een groep personen van wie we niet weten wat er mee gebeurt is.

Jaar sinds de start van het onderzoek	Aantal levende patiënten in het onderzoek (aan het begin van het jaar)	Aantal overleden patiënten in het onderzoek (gedurende het jaar)
[0,1>	120	27
[1,2>	93	18
[2,3>	75	21
[3,4>	54	9
[4,5>	45	4
[5,6>	41	5
[6,7>	36	3
[7,8>	33	3
[8,9>	30	2
[9,10>	28	2

- a) Bereken bij deze dataset de overlevingskans na één jaar van het onderzoek, na twee jaar, enzovoort. Kies een geschikte manier om de resultaten te presenteren.

Bij ongecensureerde data weten we dus exact wat de overlevingskansen zijn. Bij gecensureerde data kunnen we dit nooit te weten komen. Maar we kunnen wel een schatting doen. De betrouwbaarheid van de schatting hangt af van hoe je omgaat met de groep verdwenen patiënten. We gaan bij de gegeven ongecensureerde dataset voor verschillende manieren van omgaan met verdwenen patiënten onderzoeken wat de effecten op de overlevingskansen zijn als we met gecensureerde data te maken hebben.

- b) Maak van de hierboven gegeven dataset een gecensureerde dataset. Doe dat op minstens 3 verschillende manieren. Begin wel elke keer met 120 patiënten bij de start van het onderzoek. Verzin steeds zelf de aantallen ‘verdwenen’ patiënten. Licht daarbij toe hoe en waarom je dit zo doet.

- c) Pas op elk van de drie gecensureerde datasets die je hebt gemaakt de berekeningsmethoden toe van opdracht 1 om overlevingskansen na 5 jaar te bepalen en vergelijk de resultaten met de uitkomst van opdracht 2a. Bedenk oorzaken voor de gevonden verschillen.
- d) Kun je inmiddels meer zeggen over de betrouwbaarheid van de methoden uit opgave 1? Leg uit waarom, of waarom niet.

### Slotopdracht

Dokter Jozien Oosterhof is gynaecoloog in een academisch ziekenhuis. Ze doet onderzoek naar de vruchtbaarheid van vrouwen die langdurig de anticonceptiepil hebben geslikt. Ze wil een zo betrouwbaar mogelijke uitspraak doen over de kans op zwanger worden, afhankelijk van de tijdsduur vanaf het stoppen met de pil.

Ook zij heeft te maken met patiënten die in de loop van het onderzoek om wat voor reden dan ook afhaken. Ze beseft dat dit verschijnsel van invloed is op de resultaten, maar weet niet goed hoe ze met gecensureerde data om moet gaan.

*Anticonceptiepillen*



Gebruik wat je hebt geleerd van opdracht 1 en 2 om een uitgebreid advies aan dokter Oosterhof te schrijven. Zorg dus dat alle ideeën en uitwerkingen die je van opdracht 1 en 2 hebt, op één of andere manier verwerkt zijn in het advies voor deze slotopdracht.

Verwerk daarin ook minstens één uitgewerkt voorbeeld van een patiëntendataset waar Jozien Oosterhof in haar onderzoek mee te maken zou kunnen krijgen.

Bespreek in je advies verschillende manieren waarop zij de kans op zwanger worden binnen een bepaalde tijdsduur kan berekenen. Geef daarbij duidelijk de voordelen en de beperkingen van de gebruikte rekenmethodes aan.

Zorg dat dit advies als zelfstandig geheel te lezen is!