



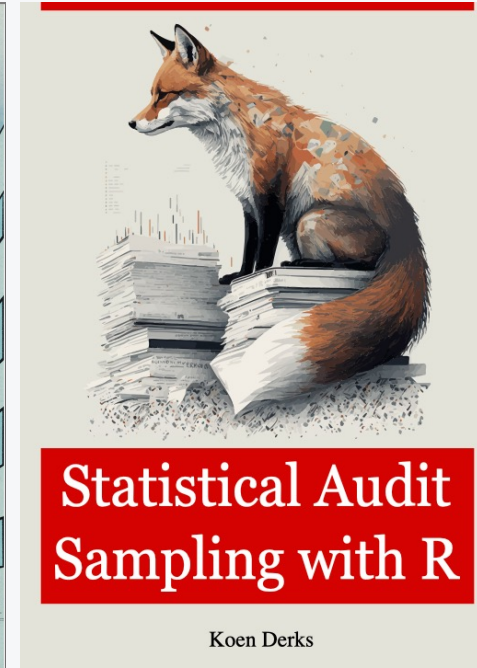
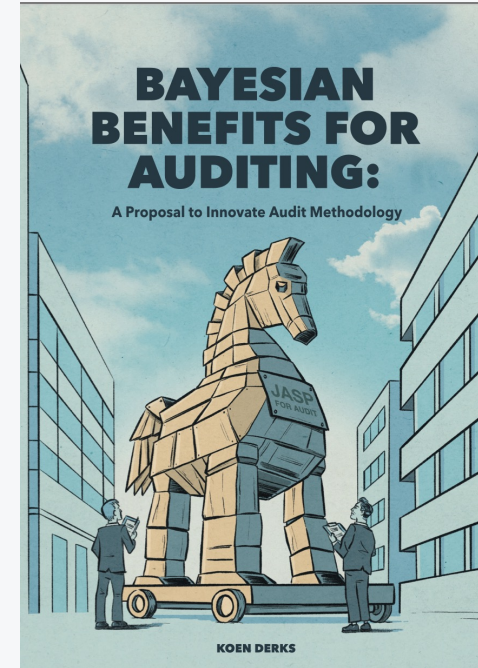
Een Bayesiaanse blik op steekproeven heeft voordelen voor de auditor

Koen Derks – K.Derks@nyenrode.nl

Biografie



dr. Koen Derks
Assistant Professor
Center of Accounting, Auditing & Control
www.statisticalauditing.com



JASP for Audit



Free | Friendly | Flexible | Download at www.jasp-stats.org

Wat is de boodschap van dit praatje?

- 1. Data-analyse maakt steekproeven efficiënter en leuk,**
- 2. op een statistisch verantwoorde manier,**
- 3. binnen het statistisch raamwerk gebruikt door de Belastingdienst.**

Boodschap is gebaseerd op dit MAB-artikel

MAB MAANDBLAD VOOR
ACCOUNTANCY EN
BEDRIJFSECONOMIE

Maandblad voor Accountancy en Bedrijfseconomie 98(4) (2024): 131–143
DOI 10.5117/mab.98.124167

Research Article



Toepassing van data-analyse om de steekproef te rationaliseren

Koen Derks, Lotte Mensink, Jacques de Swart, Ruud Wetzels

Received 28 March 2024 | **Accepted** 29 August 2024 | **Published** 18 September 2024

<http://mab-online.nl/article/124167/>

Voorbeeld: Juistheidscontrole op vennootschapsbelasting, omzetbelasting en loonheffingen

- **Belastingplichtige:** “De Fietsenmakerij”
- **Locatie:** Breukelen
- **Branche:** Fietsenwinkel en reparatieservice
- **Controleperiode:** Periode 2023
- **Methode:** Geïntegreerde statistische steekproef (§11.3)
- **Onderzoeksubject:** Alle uitgaven van belastingplichtige

Handboek Controle Belastingdienst: Steekproeven

Normaliter 4 fases in het doen van een steekproef:

1. Planning (§11.3)
2. Trekking (*Selection*) (§11.3.2)
3. Uitvoering (*Execution*) (§11.4)
4. Evaluatie (*Evaluation*) (§11.5)



Handboek Controle Belastingdienst: Planning

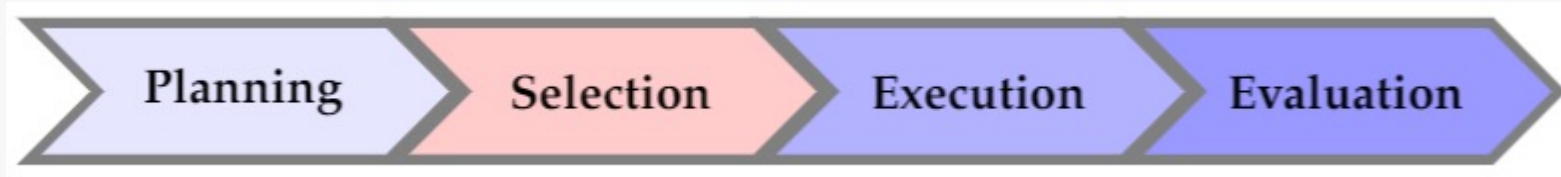


Hoofdstuk 11: Juistheidscontroles

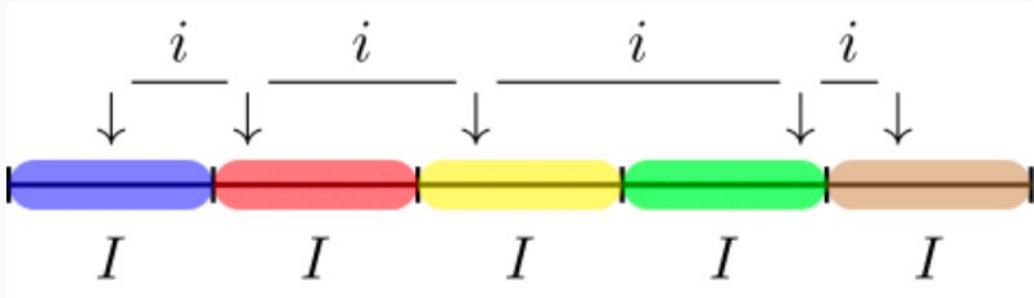
- Omzet organisatie = € 850.000
- Uitgaven belastingplichtige = € 800.000
- Materialiteit (§11.3.2.1) = € 30.000
- 95% betrouwbaarheid (§11.3.2.2)
- Nulfoutenverwachting (§11.3.2.3)

Omvang organisatie (in Nederlandse omzet)		Materialiteit
500.000		Omzet x 5%
500.000	- 1.000.000	30.000
1.000.000	- 2.000.000	60.000
2.000.000	- 4.000.000	120.000
4.000.000	- 8.000.000	240.000
8.000.000	- 15.000.000	330.000
15.000.000	- 30.000.000	600.000
30.000.000	- 60.000.000	1.200.000
60.000.000	- 120.000.000	1.800.000
120.000.000	- 240.000.000	3.000.000
240.000.000	-	6.000.000

Handboek Controle Belastingdienst: Trekking



Hoofdstuk 11: Juistheidscontroles

- *Cell sampling (§11.3.2.5)* → 
- Interval I = (controletolerantie = materialiteit) / 3 (§11.3.2.6)
= € 30.000 / 3 = € 10.000
- Steekproefgrootte = € 800.000 / € 10.000 = 80 te controleren geldeenheden

Handboek Controle Belastingdienst: Evaluatie



Hoofdstuk 11: Juistheidscontroles

- Uitkomsten steekproef: $n = 80$ geldeenheden, $k = 0$ fouten
- Maximale fout (= statistische bovengrens) $<$ materialiteit? → Goedkeuren (§ 11.5)

Om de vraag te kunnen beantwoorden of de populatie kan worden goedgekeurd, zet de controle-medewerker de berekende maximale fout af tegen de materialiteit. De materialiteit is immers de goedkeuringsgrens voor de gecontroleerde populatie.

De maximale fout geeft – met een betrouwbaarheid van 95% – aan hoe groot de fout in de populatie in het ongunstigste geval kan zijn.⁷³ De wijze waarop de maximale fout wordt berekend, gaat te ver voor dit handboek. Raadpleeg daarvoor een steekproefspecialist.

⁷³ Het betreft hier – in technische termen – de bovengrens van het 95%-betrouwbaarheidsinterval.

Hoe wordt de statistische bovengrens bepaald?

Populatie uitgaven (N): € 800.000

Materialiteit: € 30.000 (3.75%)

Steekproefgrootte (n): 80 euro's

Fouten gevonden (k): 0

Meest waarschijnlijke fout ($\frac{k}{n} = \frac{0}{80}$): € 0,00

95% Bovengrens: ?

Hoe wordt de statistische bovengrens bepaald?

Populatie uitgaven (N): € 800.000

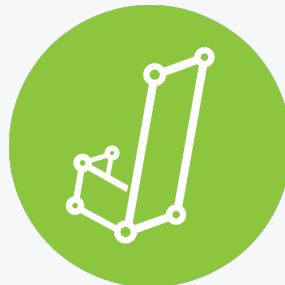
Materialiteit: € 30.000 (3.75%)

Steekproefgrootte (n): 80 euro's

Fouten gevonden (k): 0

Meest waarschijnlijke fout ($\frac{k}{n} = \frac{0}{80}$): € 0,00

95% Bovengrens:



JASP

Tabel 1. Evaluatiesamenvatting

	Waarde
Uitvoeringsmaterialiteit	30000.0000
Steekproefomvang	80
Fouten	0
Taint	0.0000
Meest waarschijnlijke fout	0.0000
95% Bovengrens	29957.3227
Nauwkeurigheid	29957.3227
BF ₊	∞

Opmerking. De resultaten zijn berekend door middel van de **gammaverdeling**.

Hoe wordt de statistische bovengrens bepaald?

Populatie uitgaven (N): € 800.000

Materialiteit: € 30.000 (3.75%)

Steekproefgrootte (n): 80 euro's

Fouten gevonden (k): 0

Meest waarschijnlijke fout ($\frac{k}{n} = \frac{0}{80}$): € 0,00

95% Bovengrens:

In Excel: **`GAMMA.INV(0,95; 1 + k; 1 / n) * N`**

`GAMMA.INV(0,95; 1 + 0; 1 / 80) * 800000 = € 29957.3227`

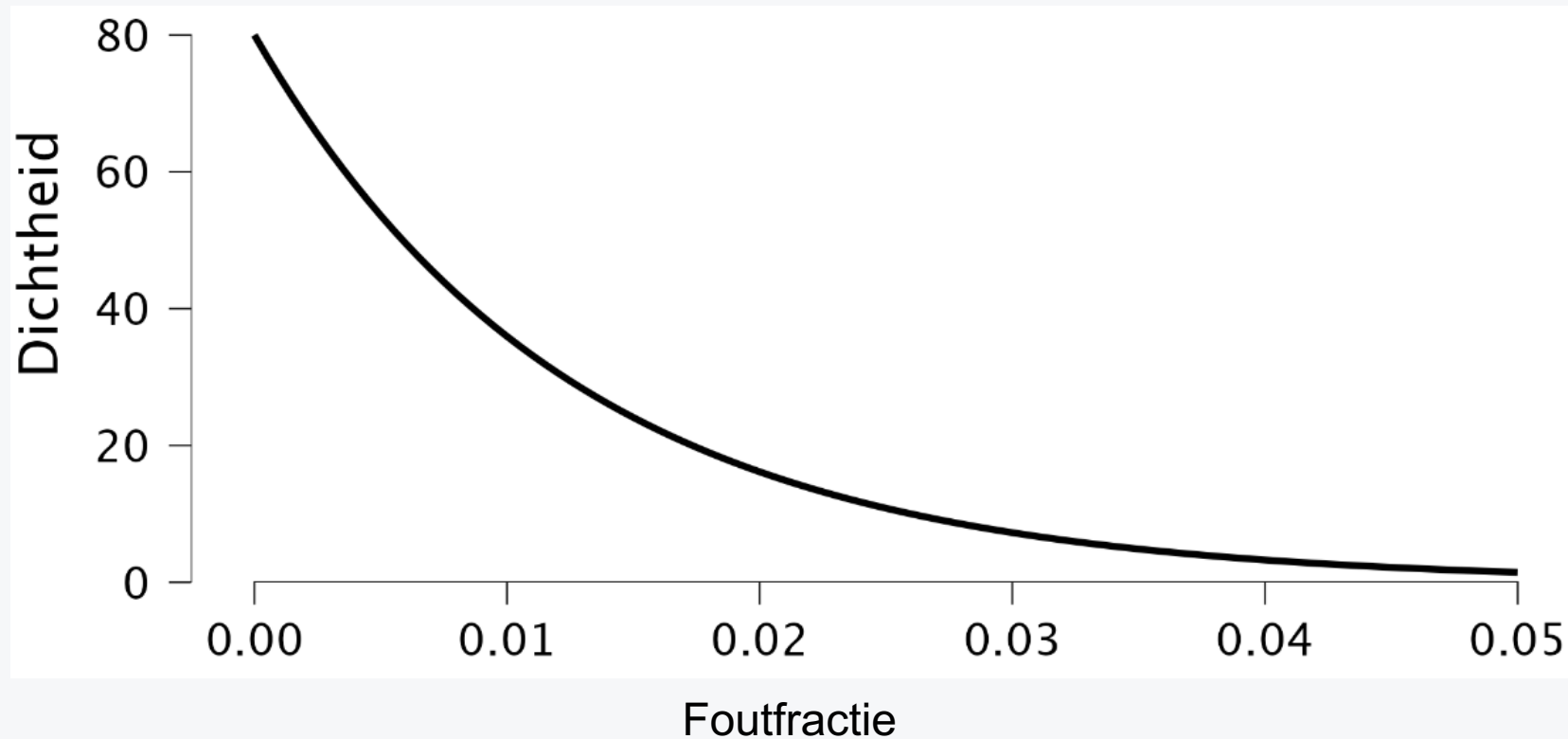
Tabel 1. Evaluatiesamenvatting

	Waarde
Uitvoeringsmaterialiteit	30000.0000
Steekproefomvang	80
Fouten	0
Taint	0.0000
Meest waarschijnlijke fout	0.0000
95% Bovengrens	29957.3227
Nauwkeurigheid	29957.3227
BF ₊	∞

Opmerking. De resultaten zijn berekend door middel van de **gammaverdeling**.

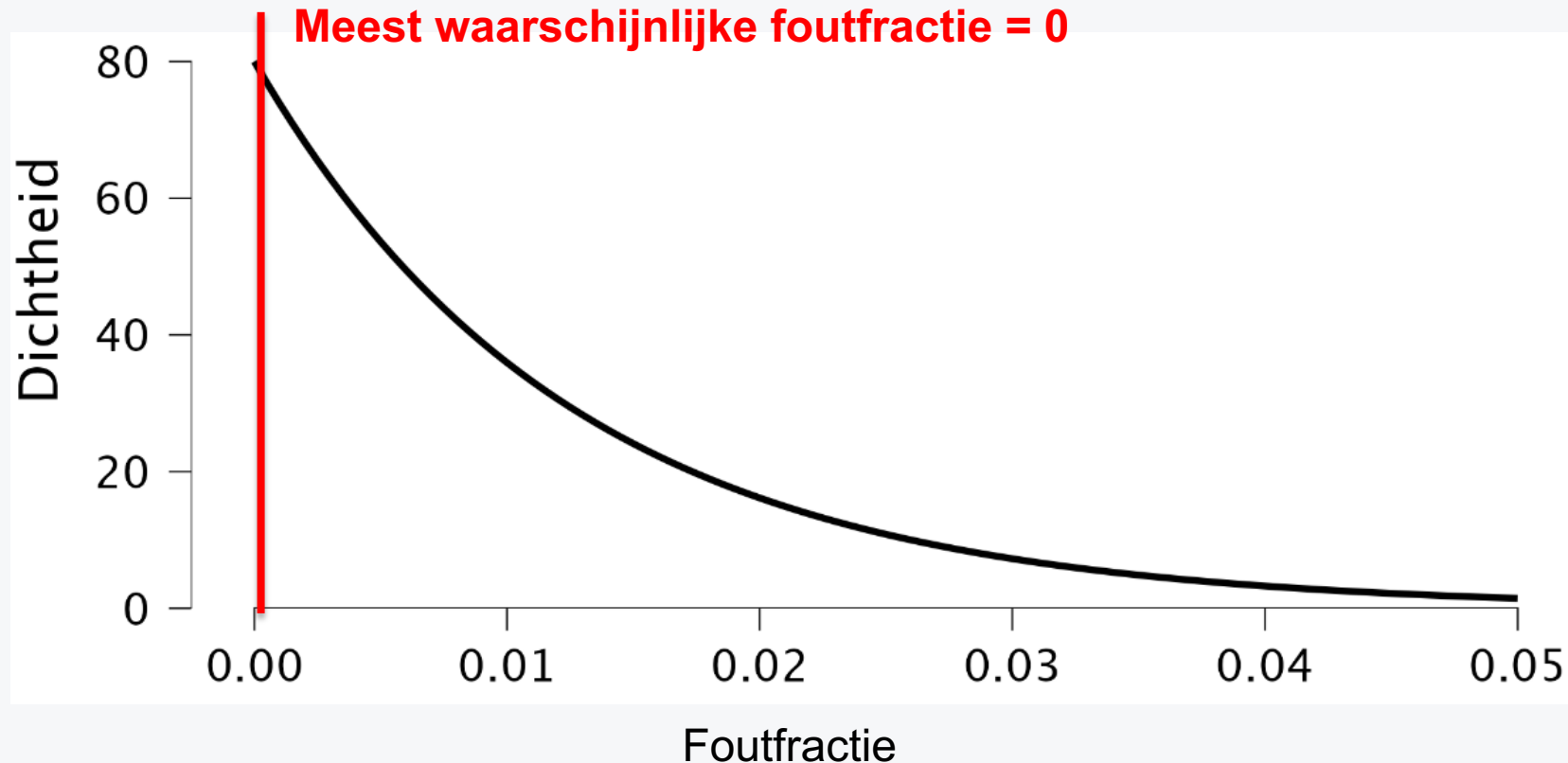
Wat ligt ten grondslag aan deze bovengrens?

Uitgangspunt: Foutfractie volgt een **gamma(1, 1 / 80) verdeling**



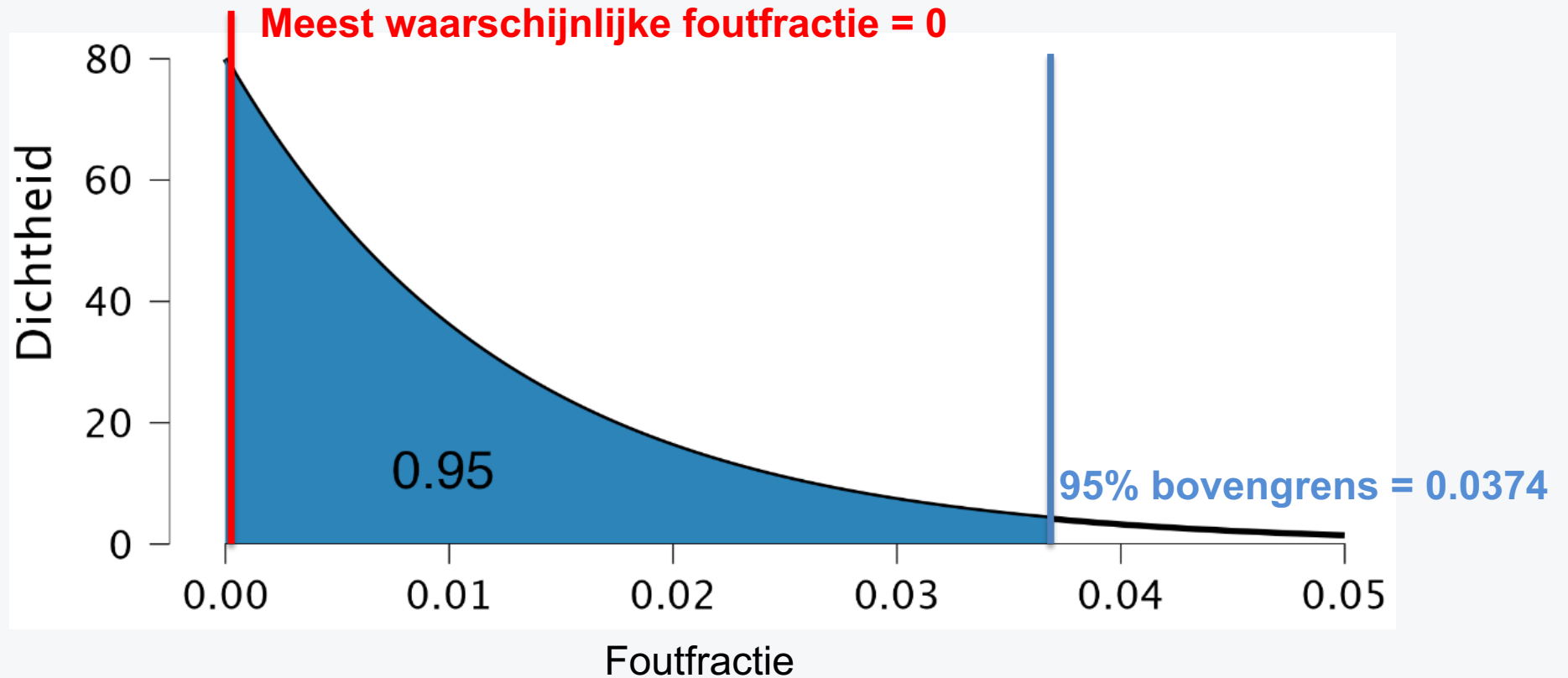
Wat ligt ten grondslag aan deze bovengrens?

Uitgangspunt: Foutfractie volgt een **gamma(1, 1 / 80)** verdeling



Wat ligt ten grondslag aan deze bovengrens?

Uitgangspunt: Foutfractie volgt een **gamma(1, 1 / 80)** verdeling



Frequentisme vs. Bayesianisme

Frequentistisch

Bayesiaans

Frequentisme vs. Bayesianisme

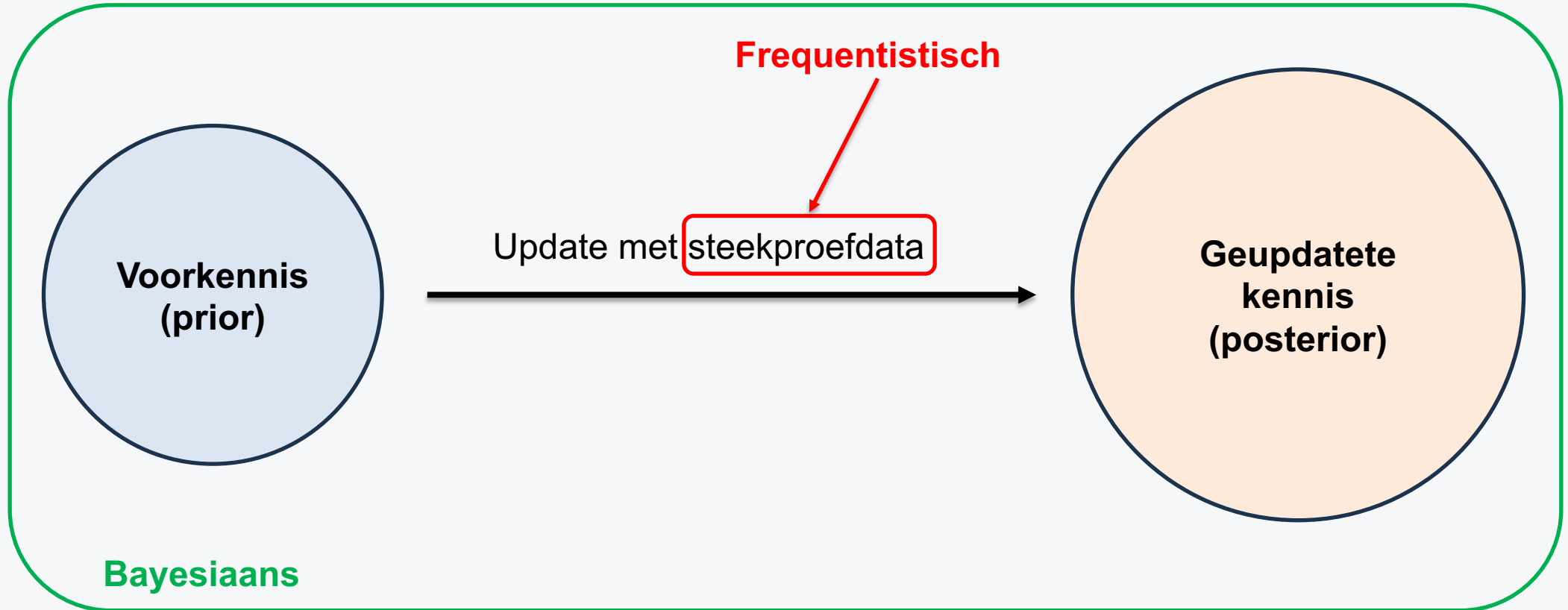
Frequentistisch



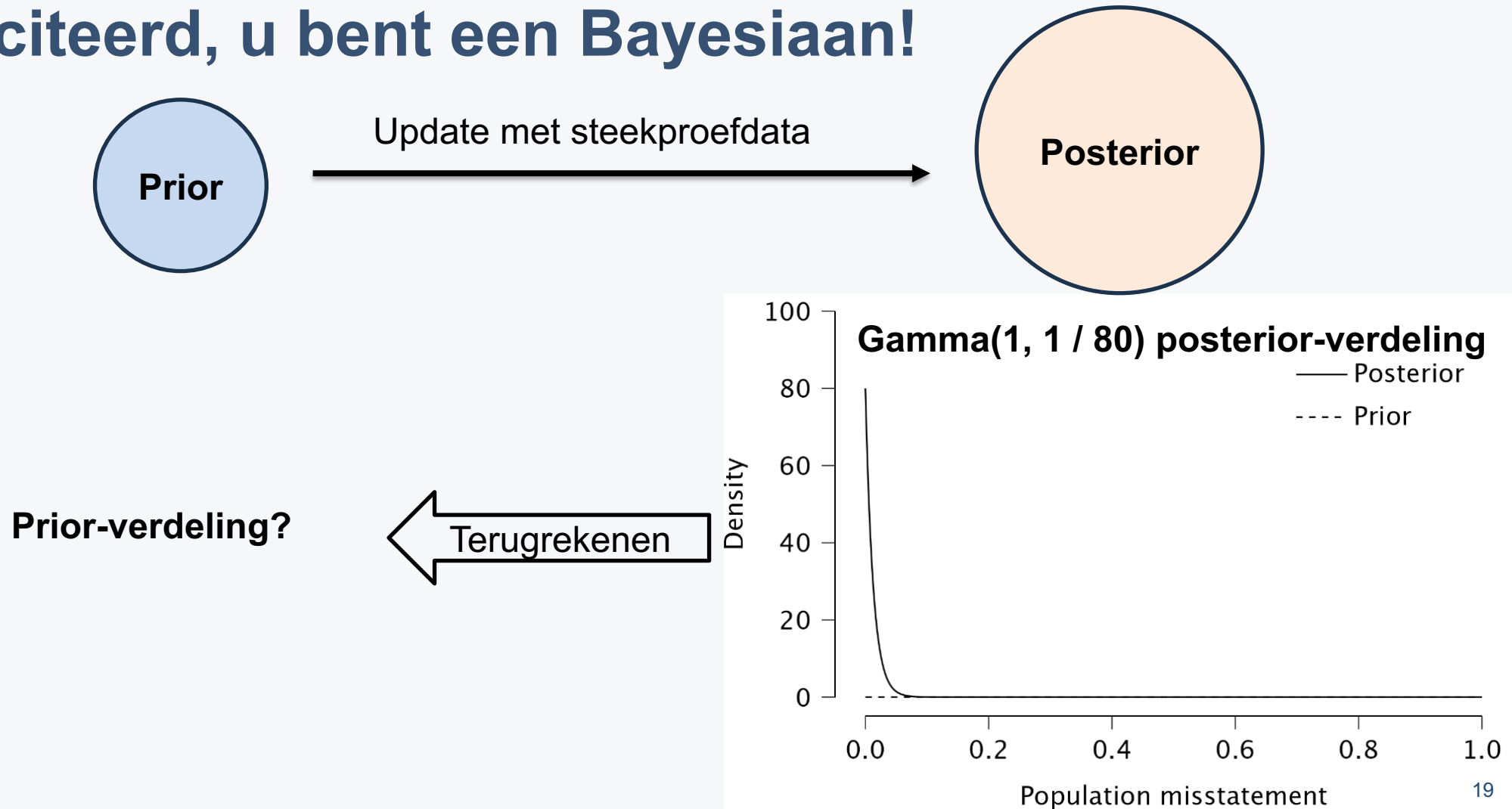
steekproefdata

Bayesiaans

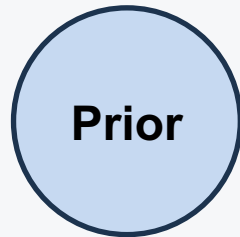
Frequentisme vs. Bayesianisme



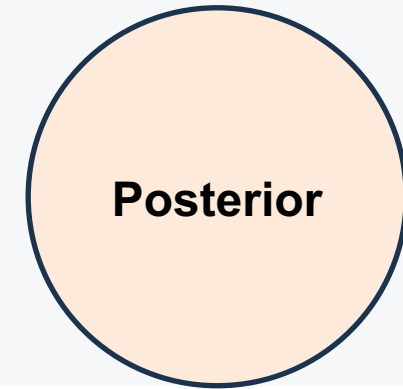
Gefeliciteerd, u bent een Bayesiaan!



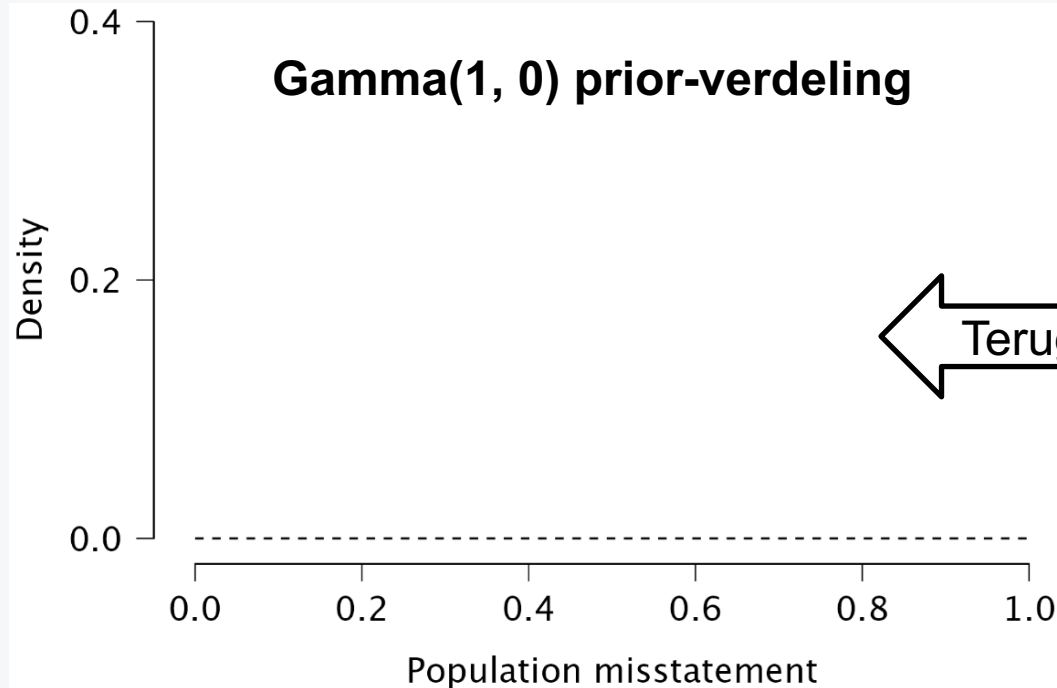
Gefeliciteerd, u bent een Bayesiaan!



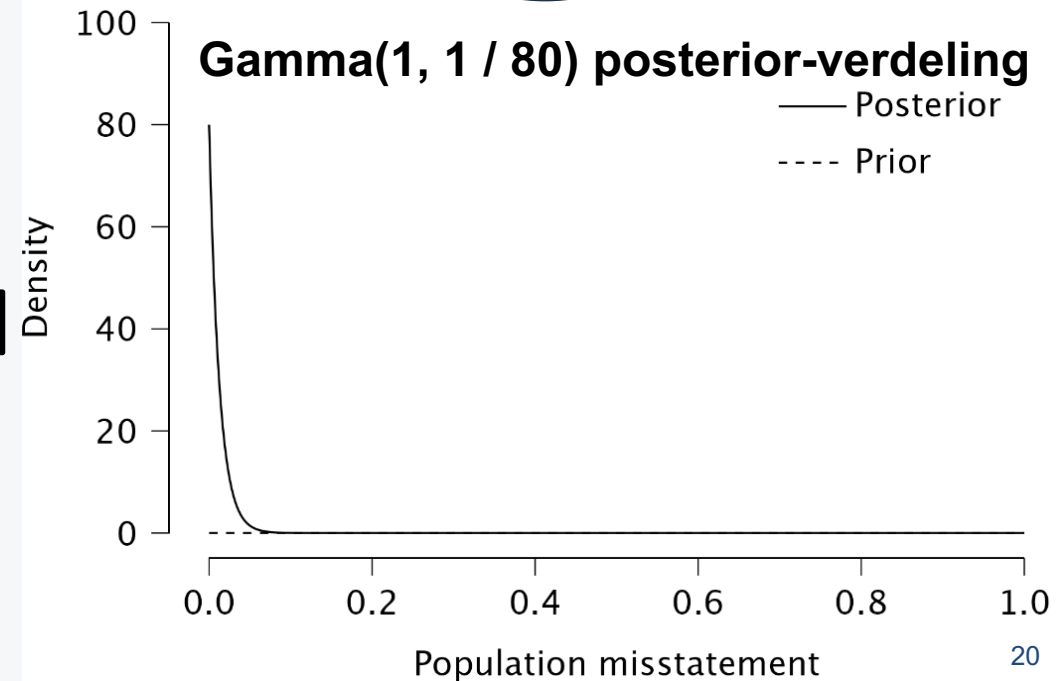
Update met steekproefdata



Gamma(1, 0) prior-verdeling

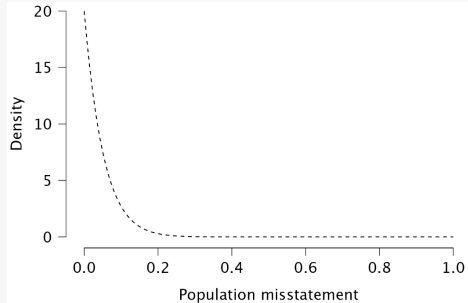


Gamma(1, 1 / 80) posterior-verdeling

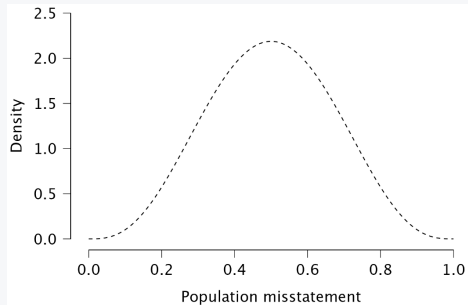


Hoe wordt een prior-verdeling geïnterpreteerd?

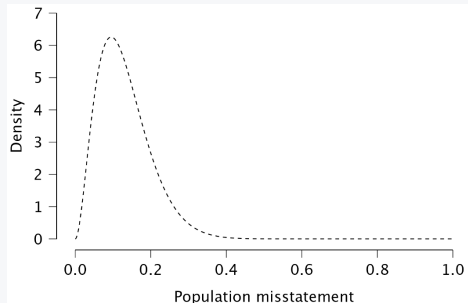
“Denken in plaatjes” – Paul van Batenburg



- Hoogste punt op 0
- Veel kansmassa bij lage foutfracties
- Weinig kansmassa bij hoge foutfracties



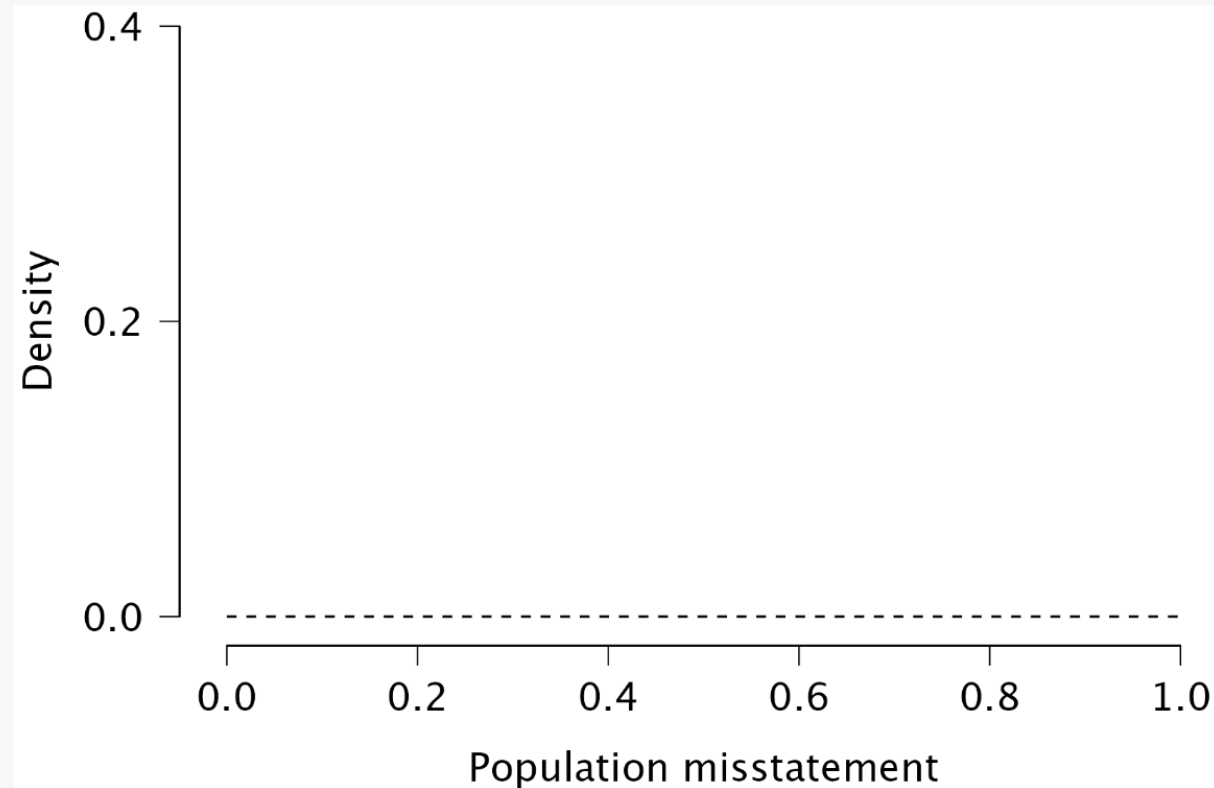
- Hoogste punt op 0.5
- Veel kansmassa bij foutfracties rond 0.5
- Weinig kansmassa bij foutfracties 0 en 1



- Hoogste punt op 0.1
- Veel kansmassa bij relatief lage foutfracties
- Weinig kansmassa bij hoge foutfracties

Hoe wordt een prior-verdeling geïnterpreteerd?

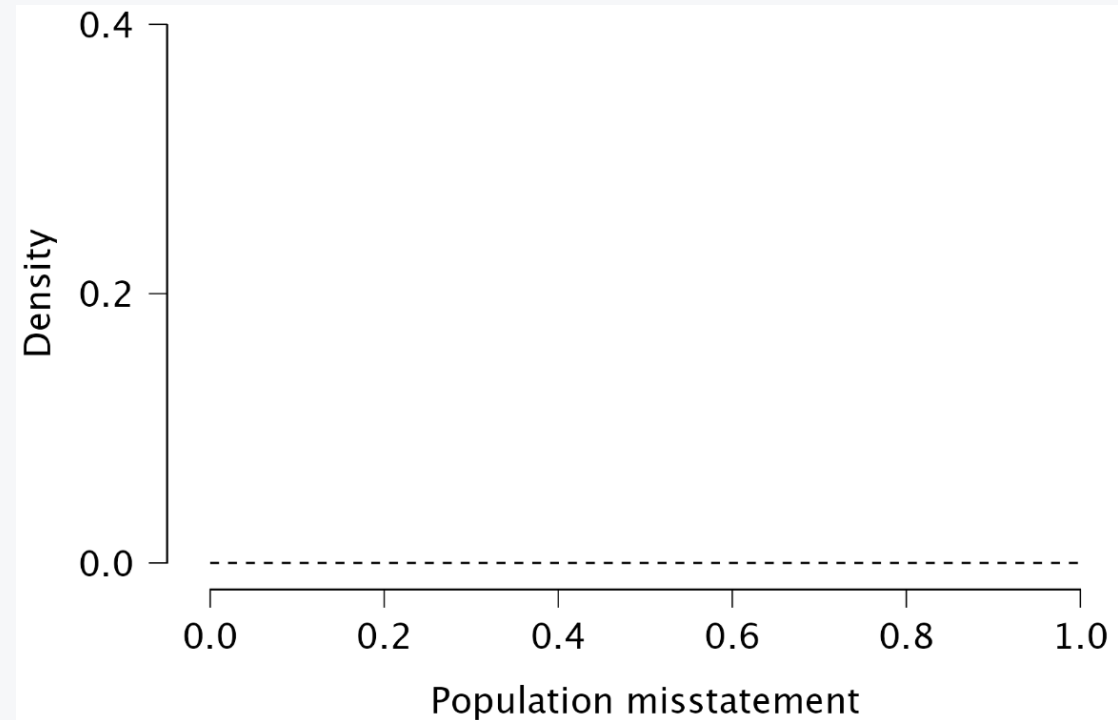
Hoe zou u de gebruikte **gamma(1, 0)** prior-verdeling interpreteren?



Hoe wordt een prior-verdeling geïnterpreteerd?

Hoe zou u de gebruikte **gamma(1, 0)** prior-verdeling interpreteren?

- Hoogste punt: Geen
- Kansmassa: Evenredig verdeeld
- Dus: Vooraf wordt aangenomen dat alle foutfracties **even waarschijnlijk** zijn
- *Realistisch?*
- *Conservatief?*



Samenvatting - I

1. Steekproeven worden bij de Belastingdienst **Bayesiaans** geëvalueerd
2. De 'klassieke steekproef' gebruikt een **conservatieve prior-verdeling**

Wat is de rol van positieve voorinformatie?

Positieve voorinformatie verantwoord een reductie van de steekproef (§ 11.3.3)

Positieve voorinformatie, verkregen tijdens de fase van de analyse van de administratie, kan worden vertaald in een vermindering van het gegevensgerichte controlewerk. In het geval van een steekproef spreken we dan over de reductie van de steekproefomvang.

Dat kan omdat het de aanname dat lage foutfracties waarschijnlijker zijn dan grote foutfracties aannemelijk maakt.

Wat is de rol van positieve voorinformatie?

Positieve voorinformatie verantwoord een reductie van de steekproef (§ 11.3.3)

Positieve voorinformatie, verkregen tijdens de fase van de analyse van de administratie, kan worden vertaald in een vermindering van het gegevensgerichte controlewerk. In het geval van een steekproef spreken we dan over de reductie van de steekproefomvang.

Dat kan omdat het de aanname dat lage foutfracties waarschijnlijker zijn dan grote foutfracties aannemelijk maakt.

In ieder geval **twee** manieren om positieve voorinformatie te verkrijgen:

1. D.m.v. testen van werking interne beheersing (§ 11.3.3)
2. D.m.v. toepassen van data-analyse (§ 6.5.2.1)

Twée soorten positieve voorinformatie

1. Opzet en werking van de interne beheersing (§ 11.3.3)

Positieve voorinformatie ⁷²	Reductie van de steekproef met	Te verrichten meerwerk
Opzet en bestaan van de interne beheersing	75%	25 'dual purpose tests'

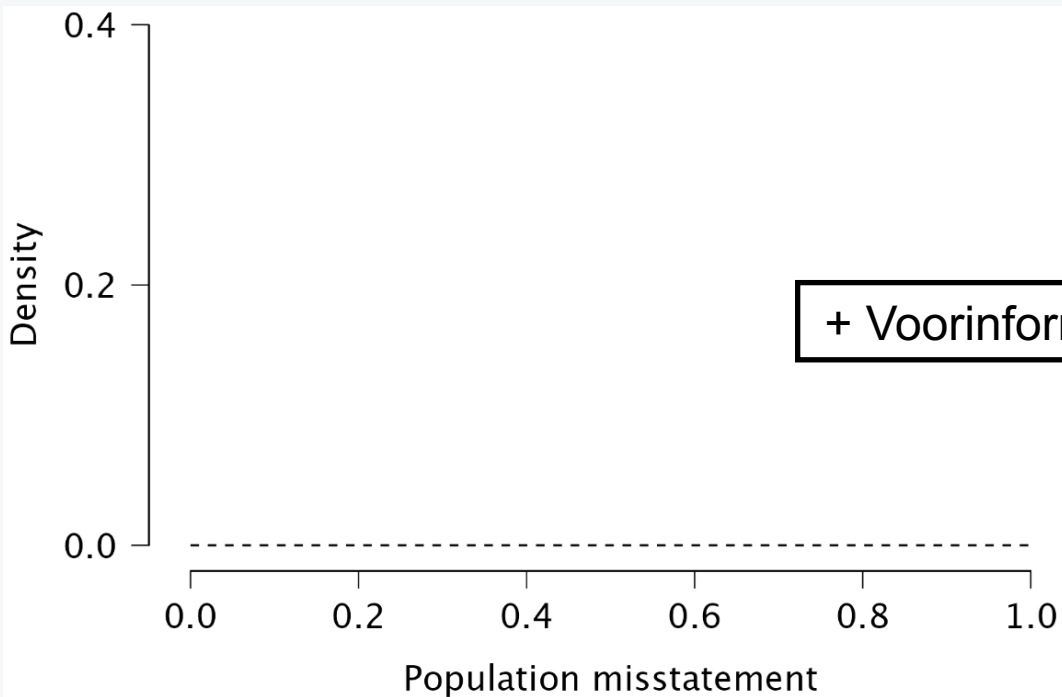
Voorwaarde: 25 dual purpose tests

Reductie: $80 * 75\% = 60$ euro's

Nieuwe steekproefomvang: $80 - 60 = 20$ euro's

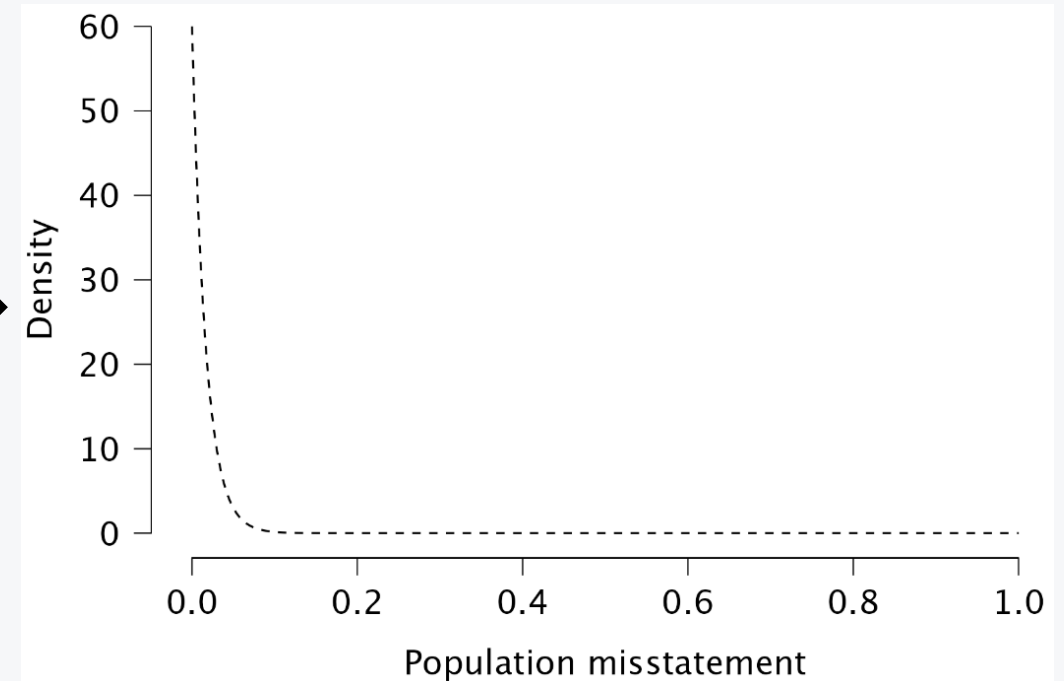
Wat is de prior-verdeling op basis van de IB?

Gamma(1, 0) prior-verdeling



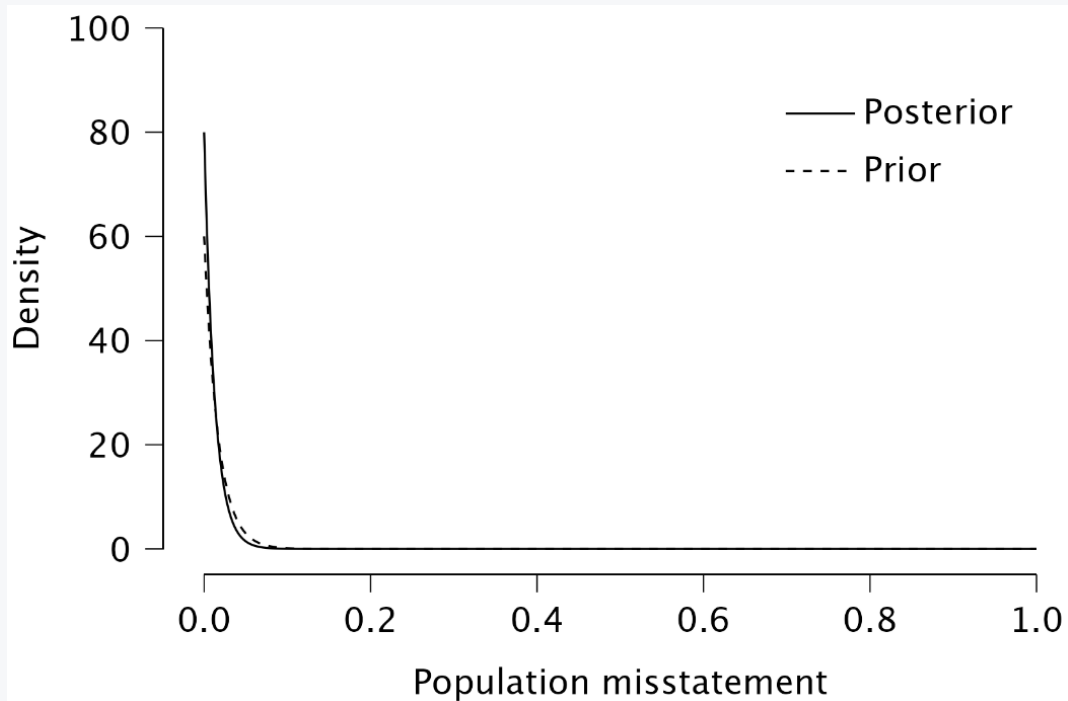
+ Voorinformatie IB

Gamma(1, 1 / 60) prior-verdeling



Wat is de prior-verdeling op basis van de IB?

Gamma(1, 1 / 80) posterior-verdeling



Tabel 1. Planningssamenvatting

	Waarde
Uitvoeringsmaterialiteit	0.0375
Auditrisico	0.0500
Toelaatbare fouten	0.0000
Minimale steekproefomvang ^a	20

Opmerking. De minimale steekproefomvang is gebaseerd op de gamma verdeling ($\alpha = 1$, $\beta = 60$)

Samenvatting - II

1. Steekproeven worden bij de Belastingdienst **Bayesiaans** geëvalueerd
2. De 'klassieke steekproef' gebruikt een **conservatieve prior-verdeling**
3. **Positieve voorinformatie** maakt de aanname dat lage foutfracties waarschijnlijker zijn dan grote foutfracties aannemelijk
4. Alle positieve voorinformatie (bijv. IB) komt in de **prior-verdeling** terecht
5. Dat zorgt voor een **reductie** in de steekproef

Twée soorten positieve voorinformatie

1. Opzet en werking van de interne beheersing (§ 11.3.3)

Positieve voorinformatie ⁷²	Reductie van de steekproef met	Te verrichten meerwerk
Opzet en bestaan van de interne beheersing	75%	25 'dual purpose tests'

2. Resultaten van data analyse (§ 6.5.2.1)

Data-analyse geeft op zich geen redelijke zekerheid over de mate waarin de populatie vrij is van materiële fouten.

Data-analyse kan ook positieve voorinformatie leveren die aanleiding geeft om de gegevensgerichte werkzaamheden in de statistische steekproef te reduceren.

Positieve voorinformatie uit data-analyse kan wiskundig vertaald worden naar een prior-verdeling

Vereist: Analyse geeft kansuitspraak over de fout(fractie) in de populatie.



Wat voor soorten data-analyse bestaan er?

3 Hoofdtoepassingen van machine learning onderverdeeld in 2 categorieën:



1. *Regressie (begeleid)*: Voorspellen van een numerieke variabele



2. *Classificatie (begeleid)*: Voorspellen van een categorische variabele



3. *Clusteren (onbegeleid)*: Structuur in de gegevens vinden (zie *MAB-artikel*)

Wat voor soorten data-analyse bestaan er?

3 Hoofdtoepassingen van machine learning onderverdeeld in 2 categorieën:



1. *Regressie (begeleid)*: Voorspellen van een numerieke variabele



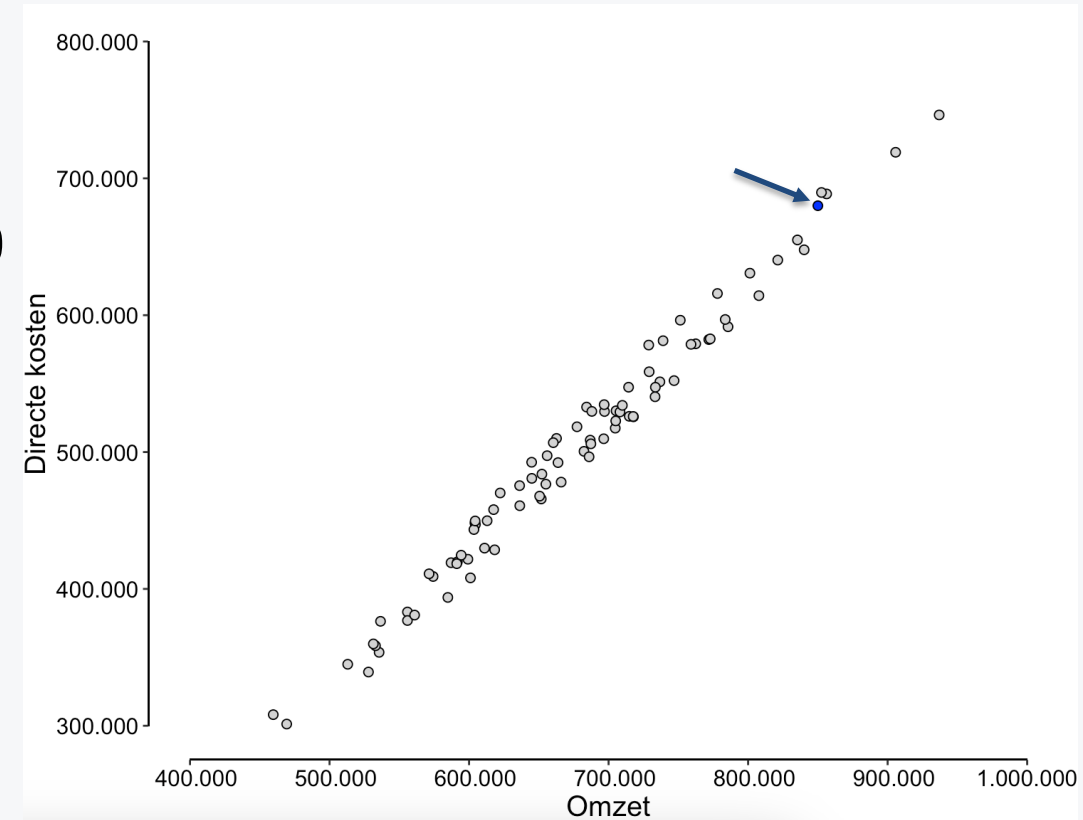
2. *Classificatie (begeleid)*: Voorspellen van een categorische variabele



3. *Clusteren (onbegeleid)*: Structuur in de gegevens vinden (zie *MAB-artikel*)

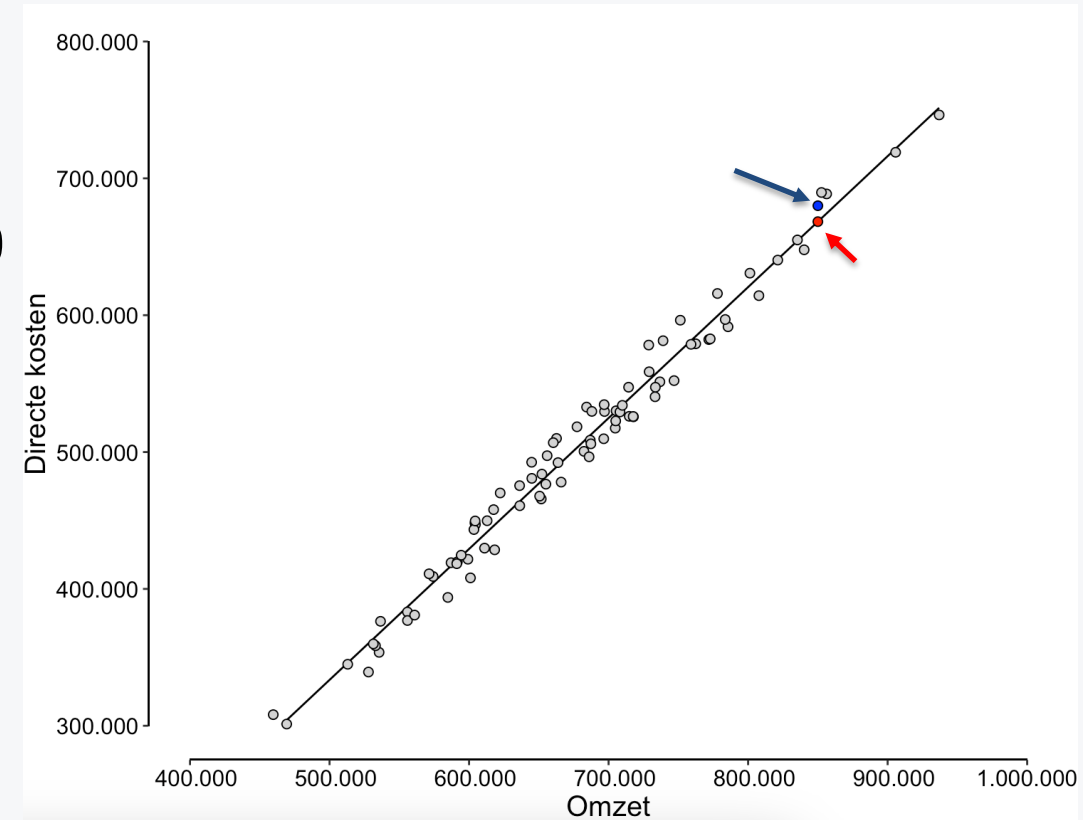
Regressie: Voorspellen van directe kosten

- Auditor kent directe kosten ● en omzet van belastingplichtige
- Kent ook omzet en directe kosten van 100 fietsenmakers van zelfde omvang in NL ○

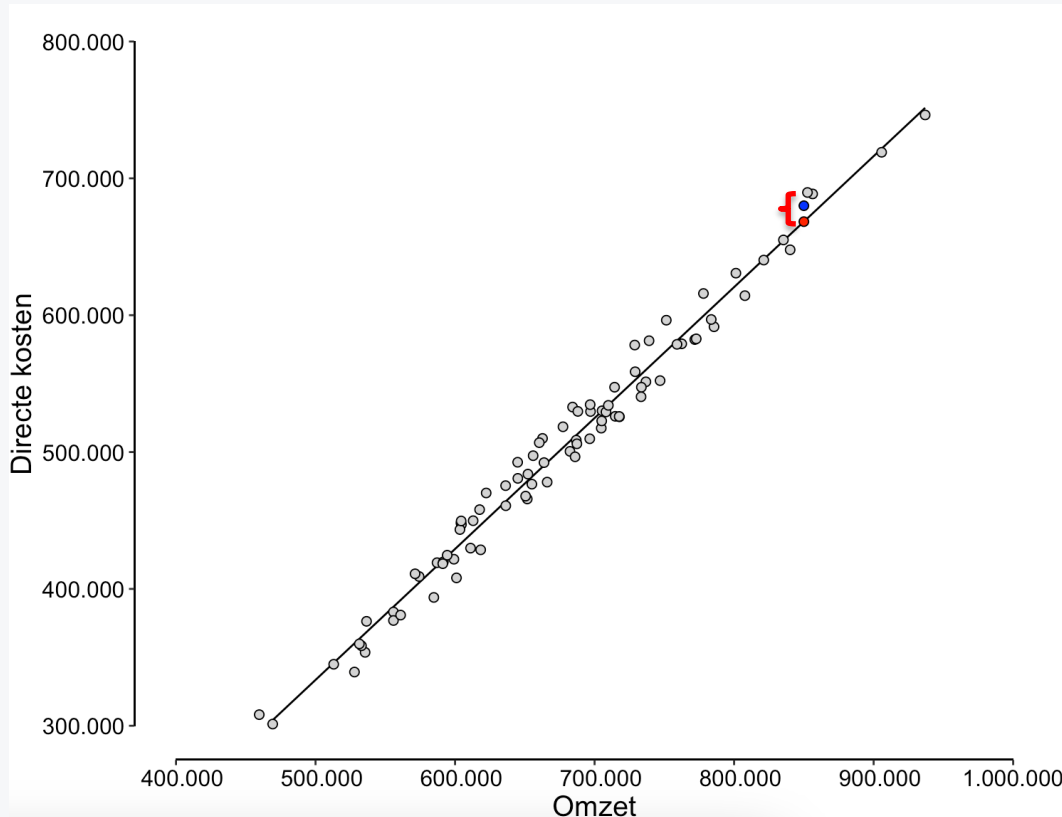


Regressie: Voorspellen van directe kosten

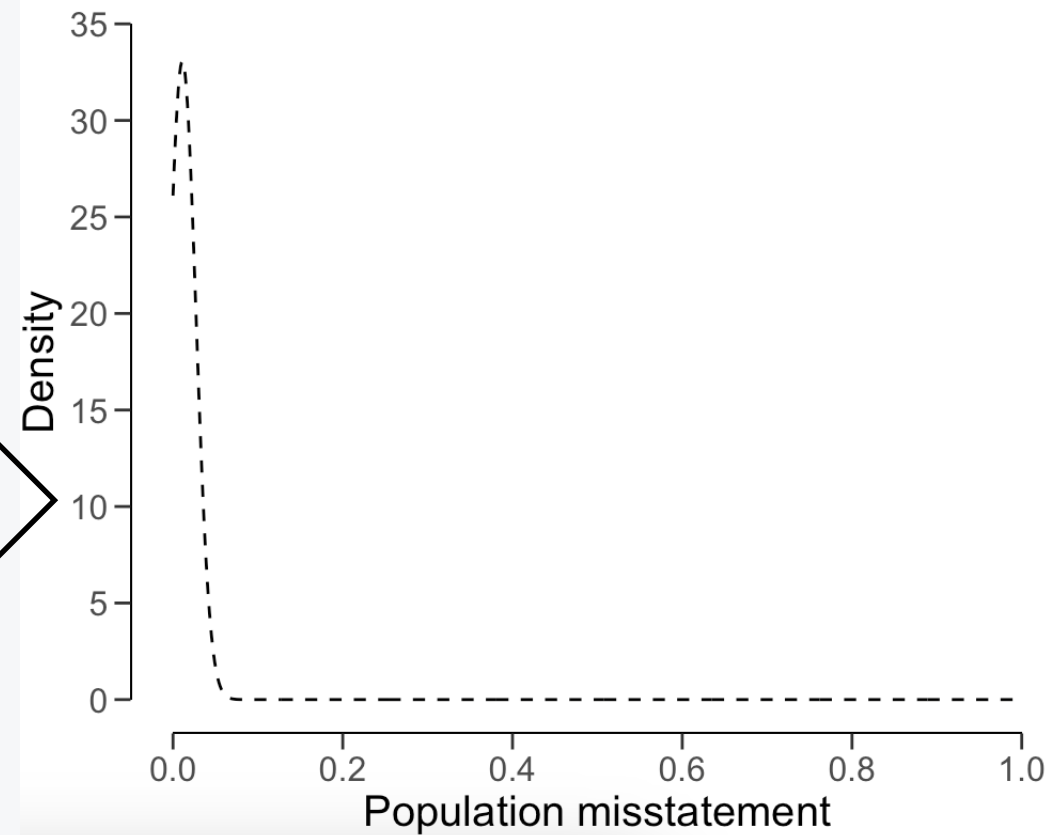
- Auditor kent directe kosten ● en omzet van belastingplichtige
- Kent ook omzet en directe kosten van 100 fietsenmakers van zelfde omvang in NL ○
- Past lineaire regressie toe om verband tussen directe kosten en omzet te modelleren
- Wil vervolgens afwijking tussen opgevoerde ● en voorspelde ● directe kosten bepalen en meenemen in analyse



Een relatief kleine afwijking van het voorspelde bedrag kan gezien worden als positieve voorinformatie

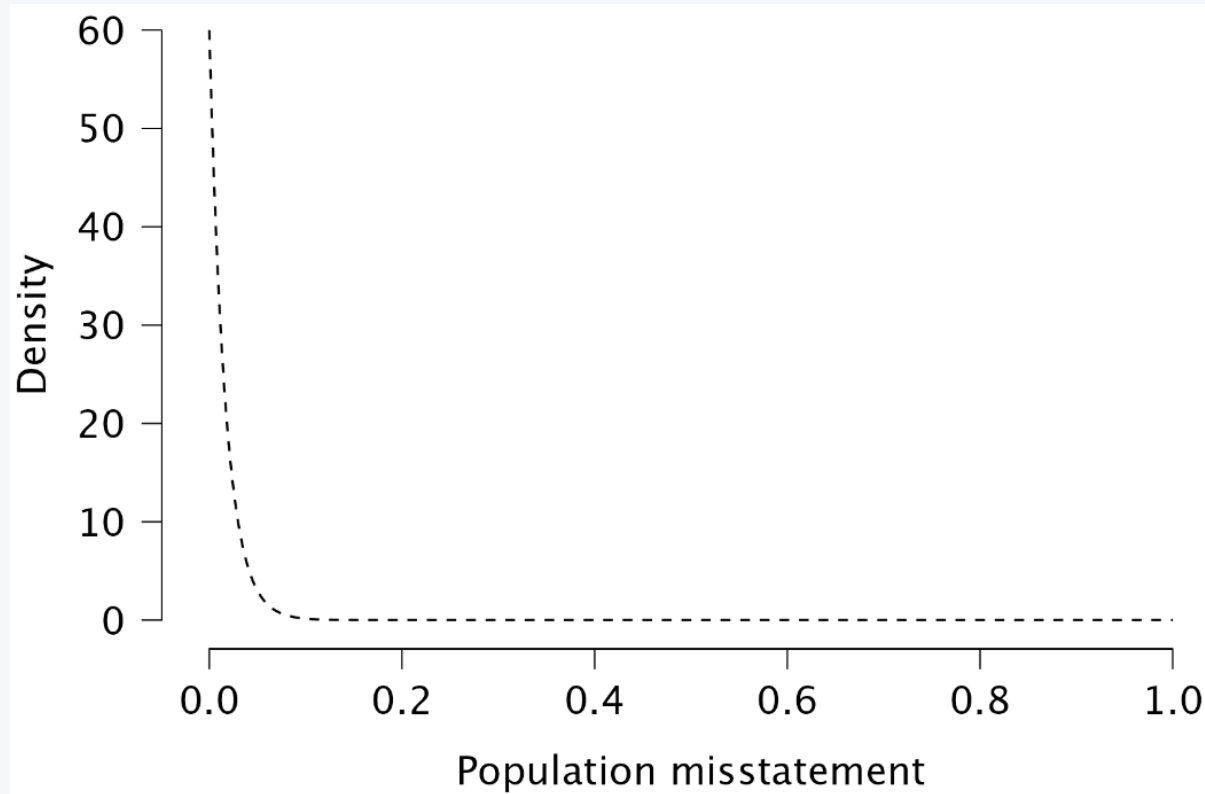


Vertaling

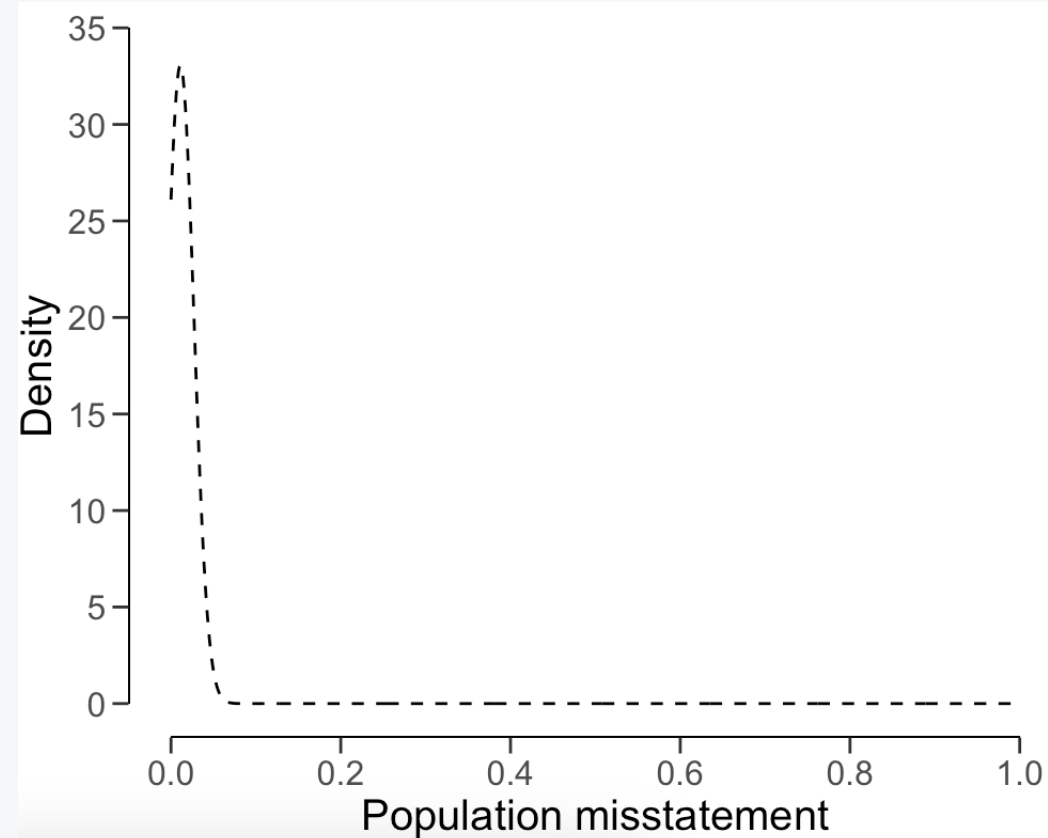


Prior-verdeling: normaal(0,011; 0,016)T(0, 1)

De prior-verdeling vertoont gelijkenis met de prior obv IB



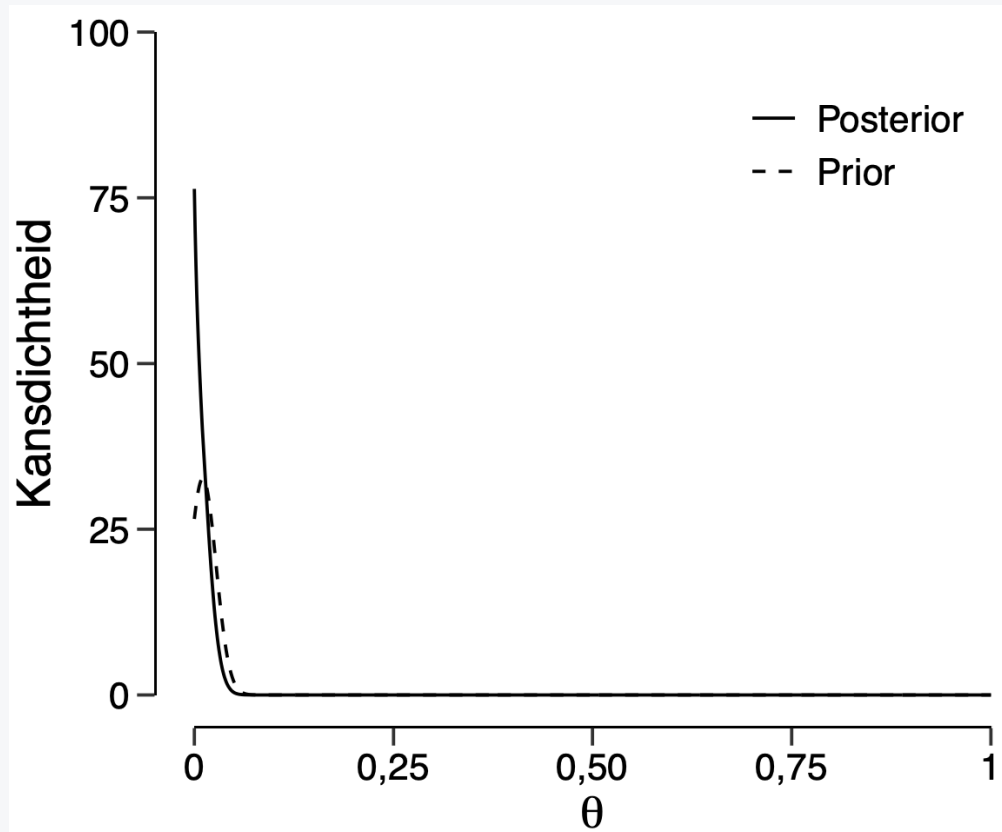
Prior-verdeling: $\text{gamma}(1, 1 / 60)$



Prior-verdeling: $\text{normaal}(0,011; 0,016)\text{T}(0, 1)$

38

De prior obv voorspelling reduceert de steekproef



Steekproef:

53 ipv 80 geldeenheden nodig om 95% bovengrens onder 3.75% te krijgen.

Tabel 1. Planningssamenvatting ▼

	Waarde
Uitvoeringsmaterialiteit	0.0375
Auditrisico	0.0500
Toelaatbare fouten	0.0000
Minimale steekproefomvang ^a	53

Hoe veel nodig als er ook nog gesteund kan worden op de interne beheersing?

Wat voor soorten data-analyse bestaan er?

3 Hoofdtoepassingen van machine learning onderverdeeld in 2 categorieën:



1. *Regressie (begeleid)*: Voorspellen van een numerieke variabele



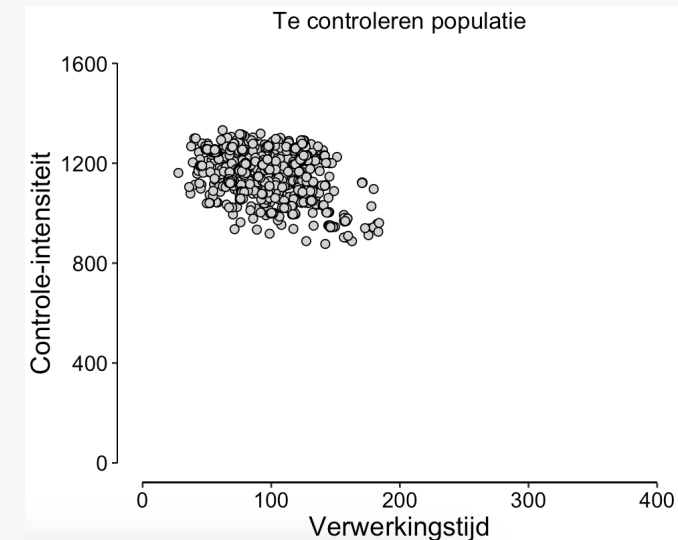
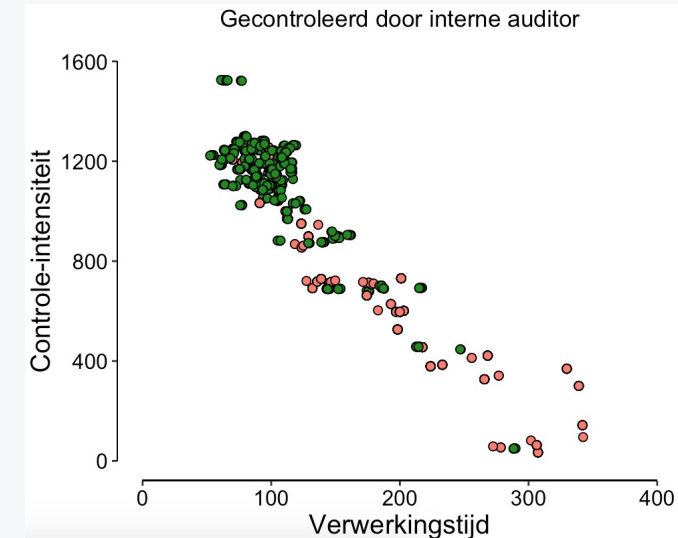
2. *Classificatie (begeleid)*: Voorspellen van een categorische variabele



3. *Clusteren (onbegeleid)*: Structuur in de gegevens vinden (zie *MAB-artikel*)

Classificatie: Voorspellen foute facturen

- Een interne auditor heeft wekelijks 10 inkoopfacturen gecontroleerd op fouten
- Gegevens over controle-intensiteit en verwerkingstijd zijn beschikbaar voor alle inkoopfacturen
- Auditor wil voortbouwen op het werk van interne auditor
- Beschikbare data: 520 al gecontroleerde inkoopfacturen
- Nog niet gecontroleerd: 1960 facturen

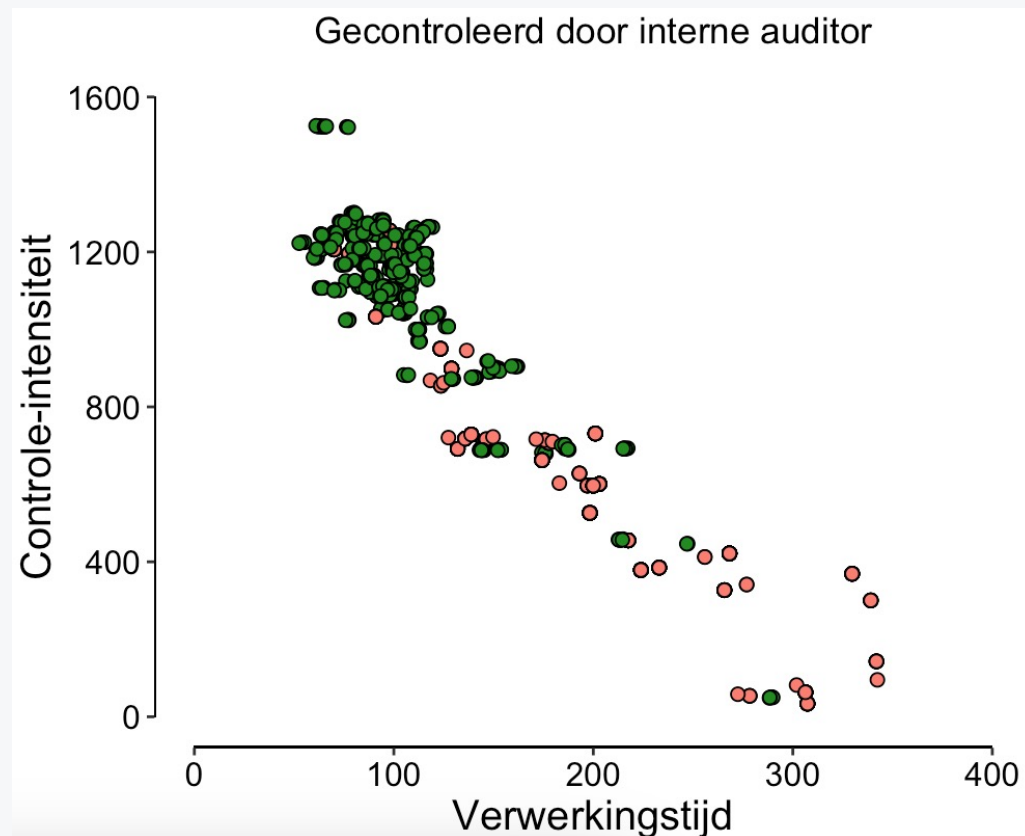


Data worden verdeeld in train- en testset

Train: 420

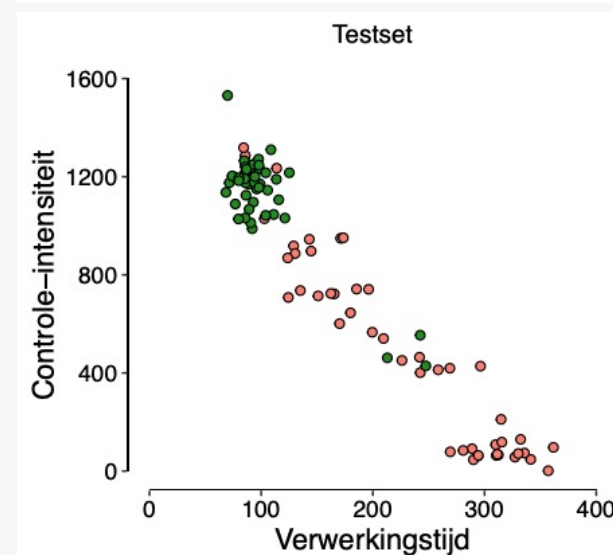
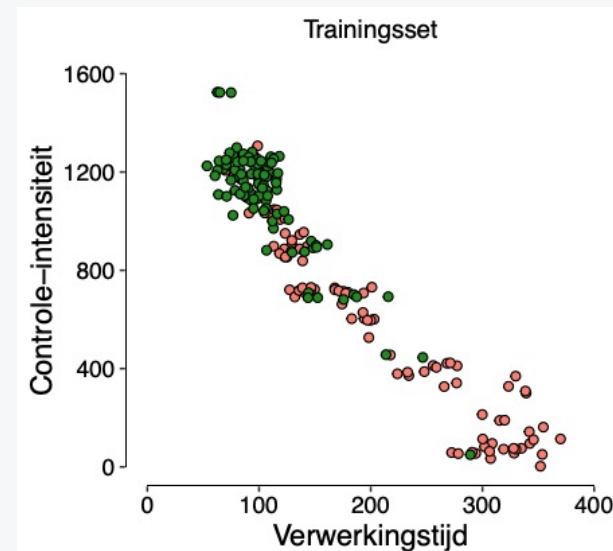
Test: 100

Totaal: 520



80 % = 420

20 % = 100

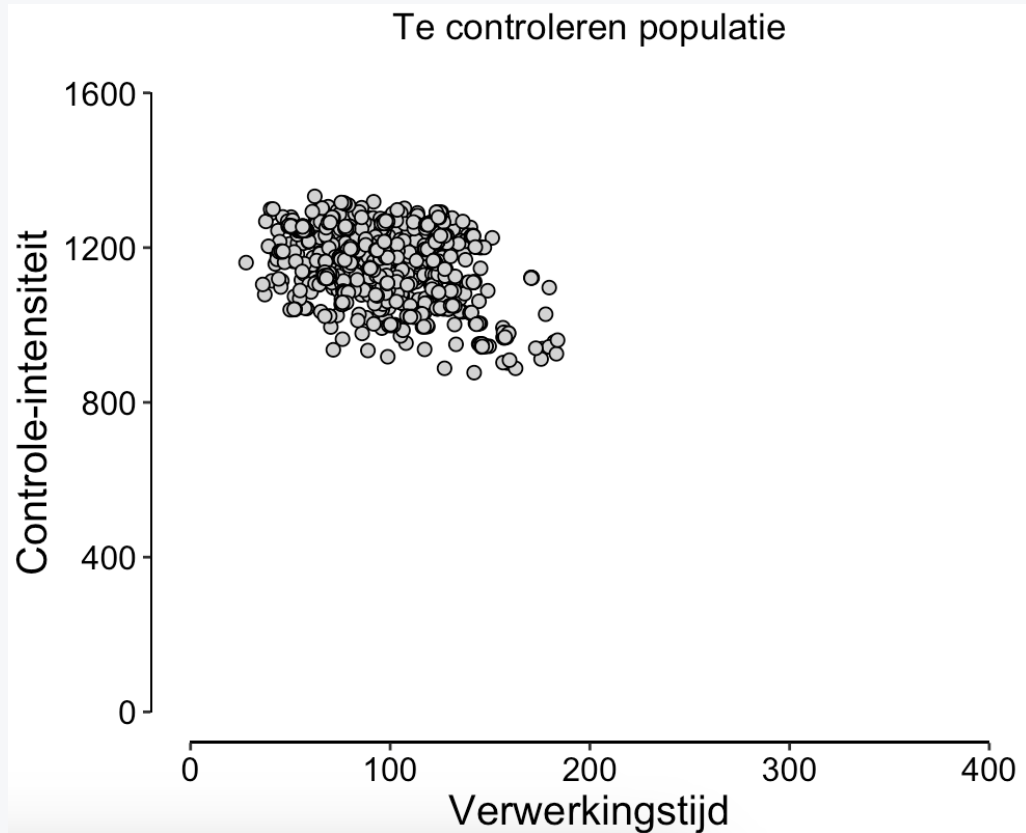


Model wordt geëvalueerd met de verwarringsmatrix

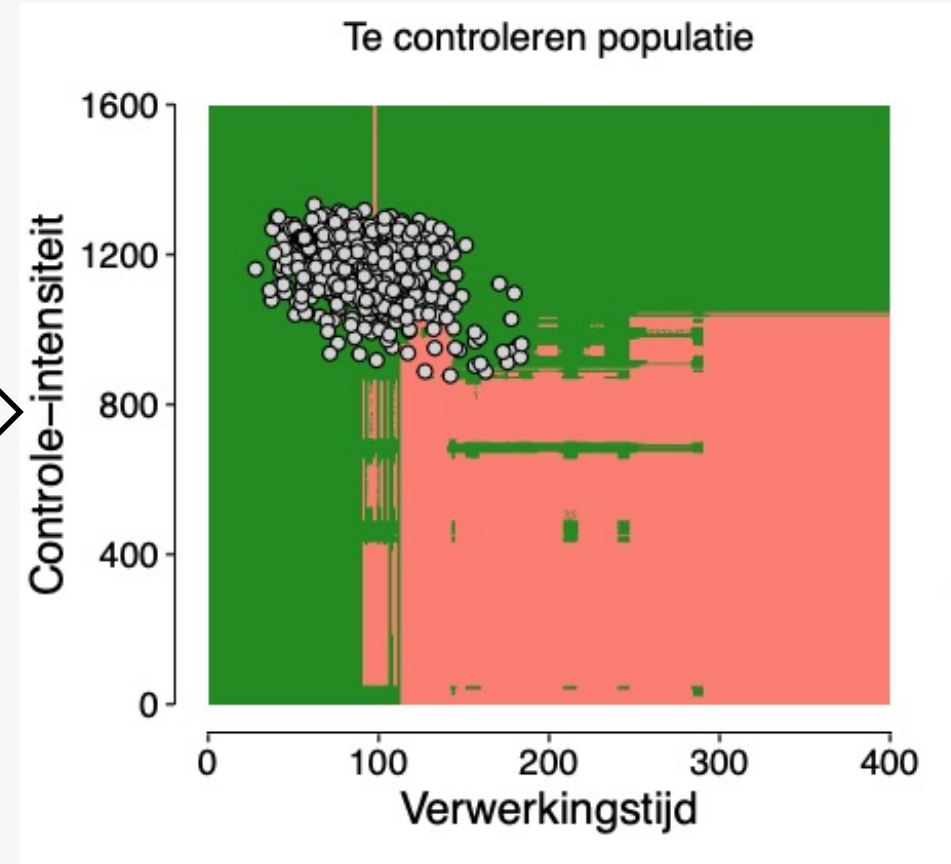
		Werkelijk	
		Fout	Niet fout
Voorspeld	Fout	11 (TP)	1 (FP)
	Niet fout	4 (FN)	84 (TN)

Bijv. nauwkeurigheid (*accuracy*) = $(TP+TN) / (TP+FN+TN+FP)$
 $(11 + 84) / (11 + 4 + 84 + 1) = 0,95$

Getraind model maakt voorspelling voor populatie

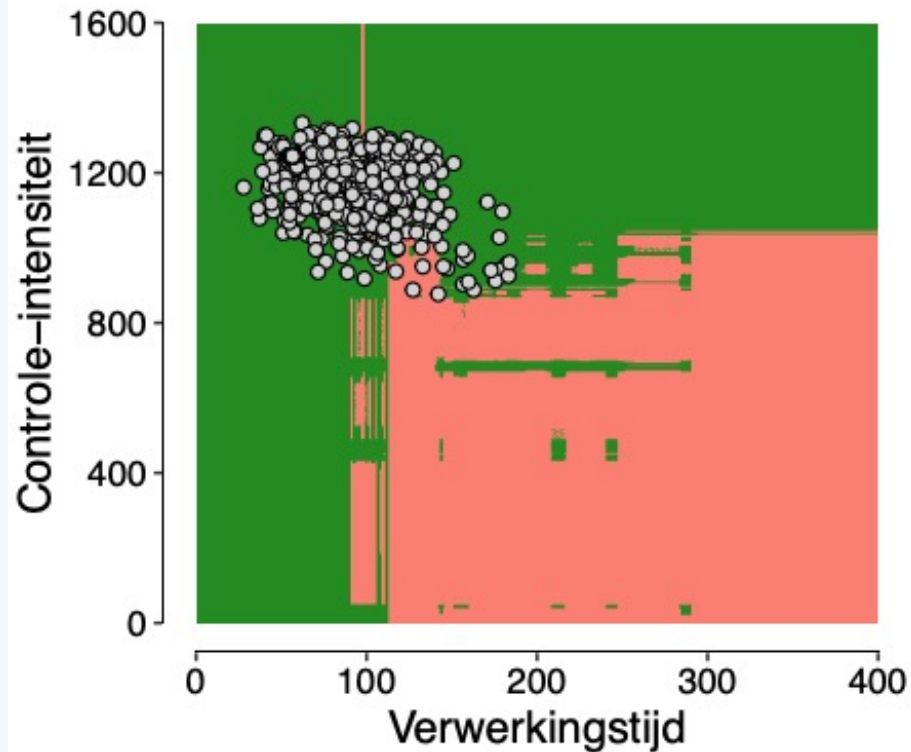


Voorspellen

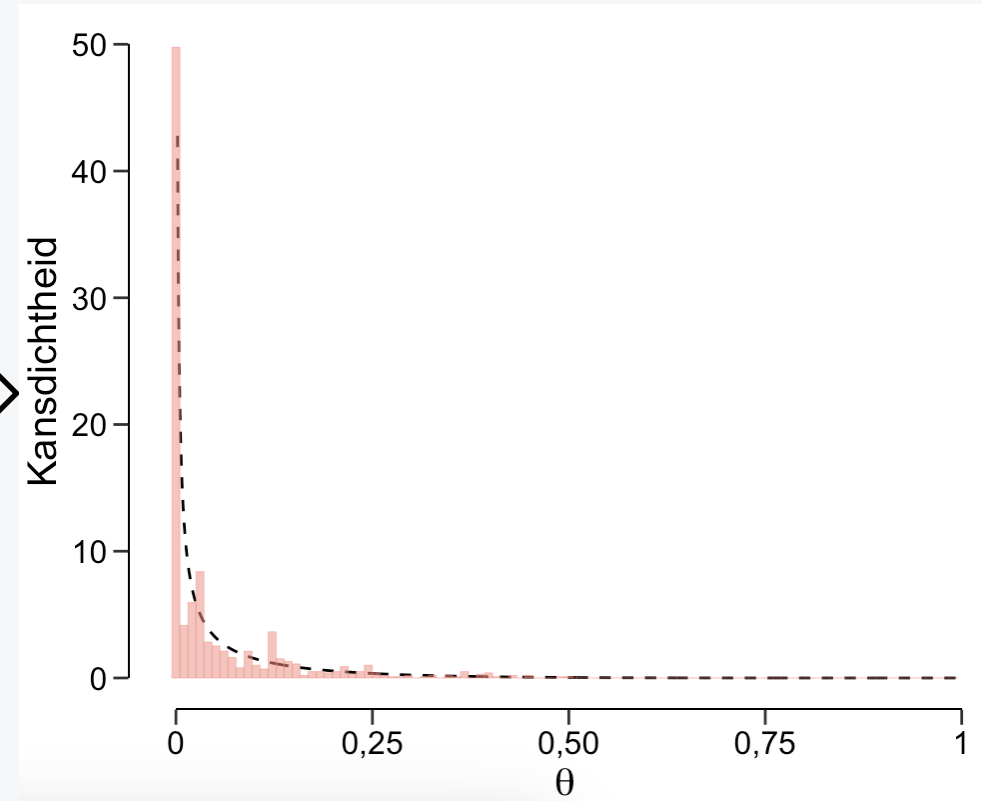


Relatief kleine voorspelde foutkansen kan gezien worden als positieve voorinformatie

Te controleren populatie

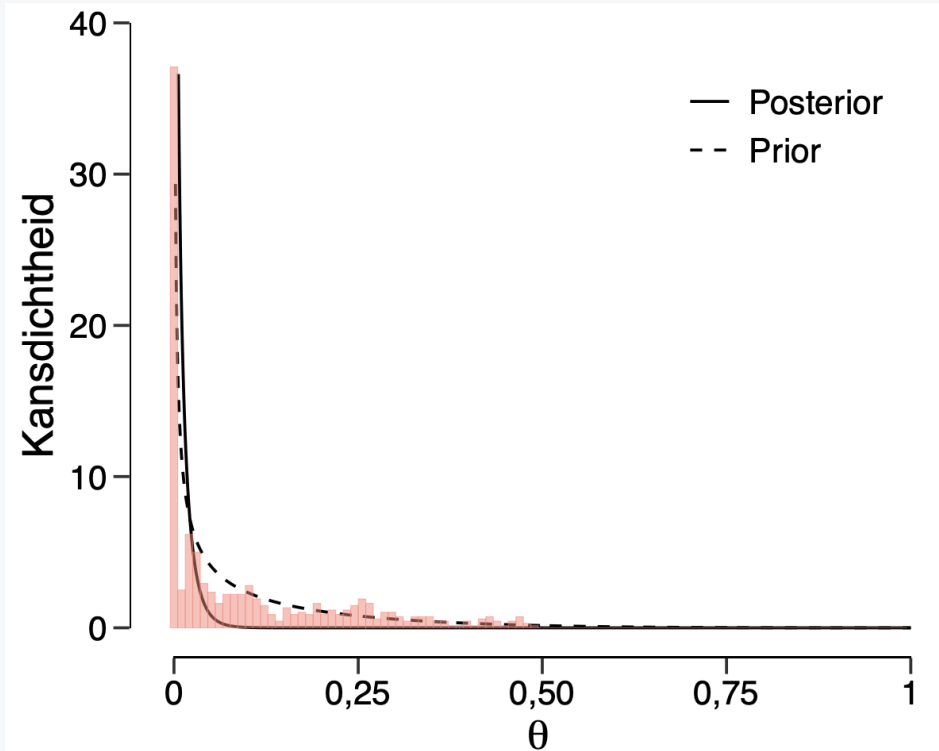


Kansuitspraak



Prior-verdeling: $\text{beta}(0,26; 5,26)$

De prior obv voorspelling reduceert de steekproef



Steekproef:

37 ipv 80 geldeenheden nodig om 95% bovengrens onder 3.75% te krijgen.

Tabel 1. Planningssamenvatting

	Waarde
Uitvoeringsmaterialiteit	0.0375
Auditrisico	0.0500
Toelaatbare fouten	0.0000
Minimale steekproefomvang ^a	37

Hoe veel nodig als er ook nog gesteund kan worden op de interne beheersing?

Samenvatting - III

1. Steekproeven worden bij de Belastingdienst **Bayesiaans** geëvalueerd
2. De 'klassieke steekproef' gebruikt een **conservatieve prior-verdeling**
3. **Positieve voorinformatie** maakt de aanname dat lage foutfracties waarschijnlijker zijn dan grote foutfracties aannemelijk
4. Alle positieve voorinformatie (bijv. IB) komt in de **prior-verdeling** terecht
5. Dat zorgt voor een **reductie** in de steekproef
6. **Data-analyse** kan hier ook voor ingezet worden

Wat voor soorten data-analyse bestaan er?

3 Hoofdtoepassingen van machine learning onderverdeeld in 2 categorieën:



1. *Regressie (begeleid)*: Voorspellen van een numerieke variabele



2. *Classificatie (begeleid)*: Voorspellen van een categorische variabele



3. *Clusteren (onbegeleid)*: Structuur in de gegevens vinden (zie MAB-artikel)

Wat is de *take-home message*?

- 1. Data-analyse maakt steekproeven efficiënter en leuk,**
- 2. op een statistisch verantwoorde manier,**
- 3. binnen het statistisch raamwerk gebruikt door de Belastingdienst.**

Meer te weten komen? Lees het artikel!

MAB MAANDBLAD VOOR
ACCOUNTANCY EN
BEDRIJFSECONOMIE

Maandblad voor Accountancy en Bedrijfseconomie 98(4) (2024): 131–143
DOI 10.5117/mab.98.124167

Research Article



Toepassing van data-analyse om de steekproef te rationaliseren

Koen Derks, Lotte Mensink, Jacques de Swart, Ruud Wetzels

Received 28 March 2024 | **Accepted** 29 August 2024 | **Published** 18 September 2024

<http://mab-online.nl/article/124167/>