Defensie Ondersteuningscommando
Ministerie van Defensie

Statistiek: college 9

Hypothesetoetsen

DOSCO

Dr. ir. Danny Blom
Nederlandse Defensie Academie
Faculteit Militaire Wetenschappen

Mei 2025

Vorige week

- Recap: populatie vs. steekproef
- Schatten van populatieparameters op basis van steekproefdata
- Betrouwbaarheidsintervallen
- Student's t-verdeling

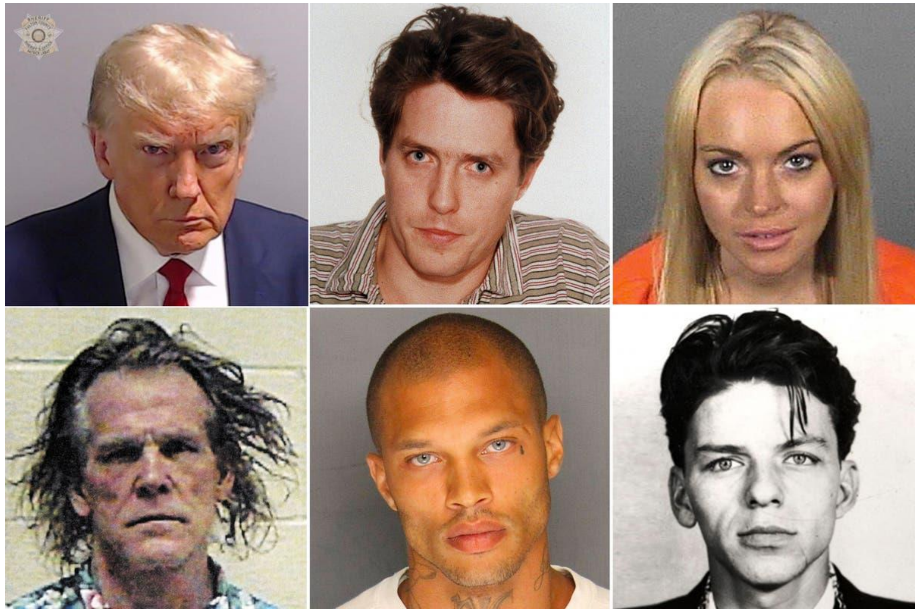
Leerdoelen

Aan het eind van dit college kunnen studenten:

- Het concept van hypothesetoetsing uitleggen in eigen woorden en de belangrijke begrippen kunnen benoemen
- Nul- en alternatieve hypothesen formuleren aan de hand van een casus.
- De verschillen tussen type-I en type-II fouten en hun interactie begrijpen
- Een z-toets en een t-toets uitvoeren en interpreteren.

Hypothesetoets: een rechter in een strafzaak

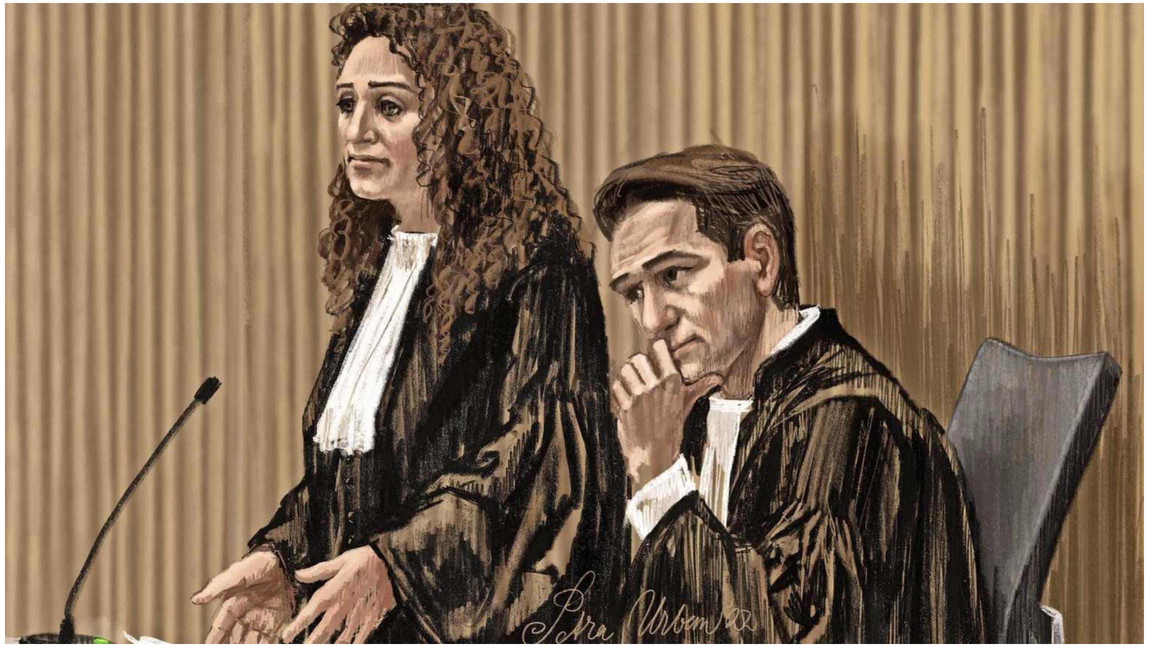
Stel je een rechter voor die in een strafzaak moet beslissen over het lot van een verdachte.



Onschuldpresumptie: *"een ieder tegen wie een vervolging is ingesteld, wordt voor onschuldig gehouden totdat zijn schuld in rechte is komen vast te staan."*

Metafoor: een rechter in een strafzaak

Welke stappen worden er in een strafzaak doorlopen om tot een vonnis te komen?



https://www.telegraaf.nl/images/1540x866/filters:format(jpeg):quality(50)/cdn-kiosk-api.telegraaf.nl/838afdf6-5b94-11ed-a6d7-0217670beecd.jpg

https://www.mensenrechten.nl/binaries/large/content/gallery/mensenrechten/content-afbeeldingen/toegang-tot-het-recht/rechter-slaat-met-voorzittershamer.jpg

Stap 1: definieer de nul- en alternatieve hypothese

In een strafzaak is de **nulhypothese**, oftewel het vermoeden van onschuld

$$H_0\text{: de verdachte is onschuldig}$$

Daartegenover staat de **alternatieve hypothese**, oftewel de reden is dat het proces plaatsvindt, de verdenking:

$$H_1\text{: de verdachte is schuldig}$$

Onschuldpresumptie: we gaan ervan uit dat de nulhypothese juist is tot er (genoeg) bewijs is verzameld voor de alternatieve hypothese.

Stap 2: bepalen van het significantieniveau

Bij een rechtzaak moet bepaald worden wat het “aanvaardbare risico op een onterechte veroordeling” is. Bij algemene hypothesetoetsen noemen we dit het **significantieniveau** α.

- Grens van “redelijke twijfel”

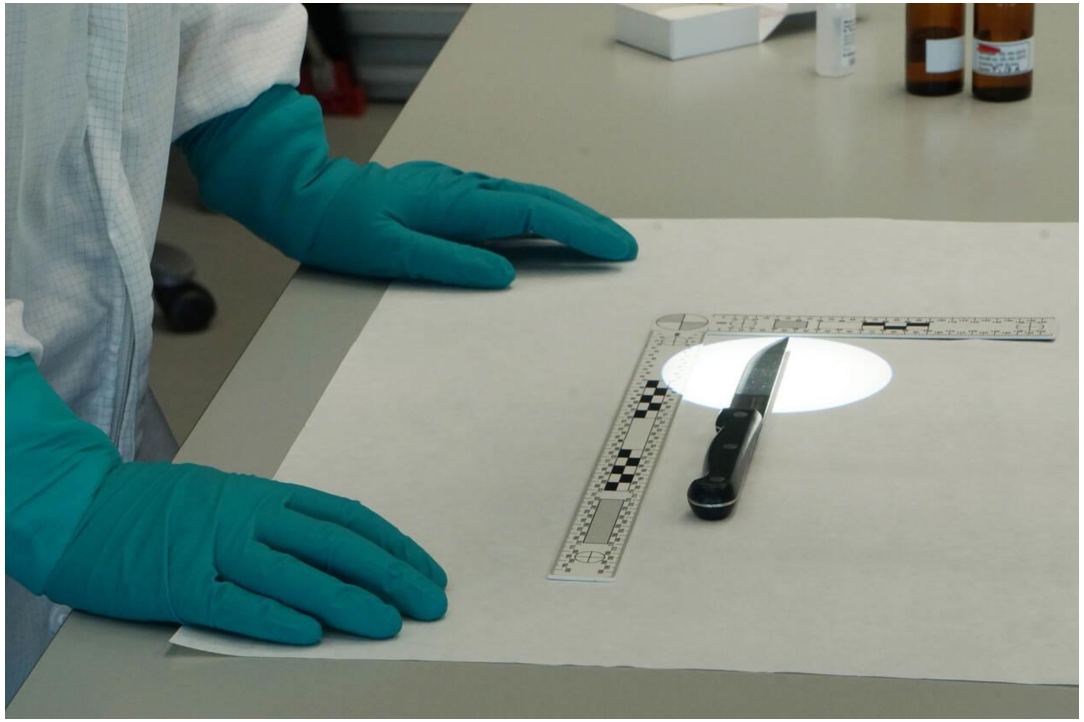
- Bepaalt de drempelwaarde van minimaal benodigd bewijs tegen de verdachte om tot veroordeling te komen.

- Lager risico → strengere eisen aan het bewijs, minder kans op onterechte veroordeling.

Stap 3: verzamelen van data

Om te kunnen oordelen over de (on)schuld van de verdachte, kan de rechter bewijsmateriaal laten verzamelen (**data**) in een dossier.

- Motieven
- Camerabeelden
- Forensisch sporenonderzoek
- (Gebrek aan) een alibi
- Getuigenverklaringen
- Psychologisch onderzoek van de verdachte
- Verklaringen van onafhankelijke experts
- …



https://magazines.forensischinstituut.nl/atnfi/2019/32/dna-sporenonderzoek-in-beeld

Stap 4: bepalen van de toetsingsgrootheid

Het dossier van de verzamelde data worden kwantitatief beoordeeld.
Dit resulteert in een **toetsingsgrootheid**: een kwantitatieve maat van hoe belastend het bewijs is tegen de verdachte, uitgaande van zijn onschuld.

Weging van verschillende soorten bewijsmateriaal:

- Is het bewijsmateriaal relevant?
- Hoe geloofwaardig is de bron / onderzoeksmethode (getuigen, DNA-sporen)?
- Is het bewijsmateriaal op rechtmatige wijze verkregen?

Stap 5: bepaal de conclusie en formuleer deze in de originele context

De rechter bekijkt aan de hand van de **toetsingsgrootheid** of de schuld van de verdachte voorbij “redelijke twijfel” is bewezen en spreek een vonnis (**conclusie**) uit .

Conclusie: één van twee opties:

1. de nulhypothese wordt **gehandhaafd**
   → de schuld is NIET voldoende bewezen

2. de nulhypothese wordt **verworpen**
   → de schuld is WEL voldoende bewezen

Merk op: de conclusie hoeft niet juist te zijn!!

https://images.rd.nl/fill/crop:4250:2657:sm/w:1020/plain/https://erdee-prod-bucket-s3-001.ams3.cdn.digitaloceanspaces.com/ANP_457987554_d029c8a487.jpg

Stappenplan hypothesetoetsen samengevat

Stap 1: definieer de nulhypothese H_0 en de alternatieve hypothese H_1

Stap 2: bepalen van het significantieniveau α

- Hoeveel risico wil je nemen met onterecht verwerpen van de nulhypothese?

Stap 3: verzamelen van data

Stap 4: bepalen van de toetsingsgrootheid

- Het significantieniveau bepaalt waar de grens ligt tussen handhaven / verwerpen van H_0

Stap 5: geef de conclusie van de hypothesetoets en formuleer deze in de originele context

Type-I en type-II fouten

Helaas zijn er soms verschillen tussen het vonnis van een rechter (**conclusie**) en de daadwerkelijke realiteit. Dit verstaan we onder gerechtelijke dwalingen:

Type-I fout: onterecht veroordeeld

Massale arrestaties in El Salvador

Type-II fout: onterecht vrijuit (tijdens carrière)

Lance Armstrong

Type-I en type-II fouten

	Conclusie		
Werkelijkheid		H_0 is waar	H_0 is niet waar
	H_0 is waar (verdachte is onschuldig)	Verdachte wordt terecht vrijgesproken	Verdachte wordt onterecht veroordeeld (type-I fout)
	H_0 is niet waar (verdachte is schuldig)	Verdachte wordt onterecht vrijgesproken (type-II fout)	Verdachte wordt terecht veroordeeld

Doel: zoveel veel mogelijk juiste conclusies trekken (groene cellen)

Probleem: trade-off tussen type-I fouten en type-II fouten

Type-I en type-II fouten

Probleem: trade-off tussen type-I fouten en type-II fouten

	Conclusie		
		H_0 is waar	H_0 is niet waar
Werkelijkheid	H_0 is waar (verdachte is onschuldig)	Verdachte wordt terecht vrijgesproken	Verdachte wordt onterecht veroordeeld (type-I fout)
	H_0 is niet waar (verdachte is schuldig)	Verdachte wordt onterecht vrijgesproken (type-II fout)	Verdachte wordt terecht veroordeeld

- Altijd H_0 aannemen betekent geen type-I fout:
  - Niemand wordt veroordeeld → echte daders gaan altijd vrijuit ☹

- Altijd H_0 verwerpen betekent geen type-II fout:
  - Iedereen wordt veroordeeld → ook veel onschuldige mensen ☹

Gewetensvraag: welk type fout is ernstiger?

Onderscheidend vermogen

Het gekozen **significantieniveau** bepaalt de kans dat je een type-I fout maakt.

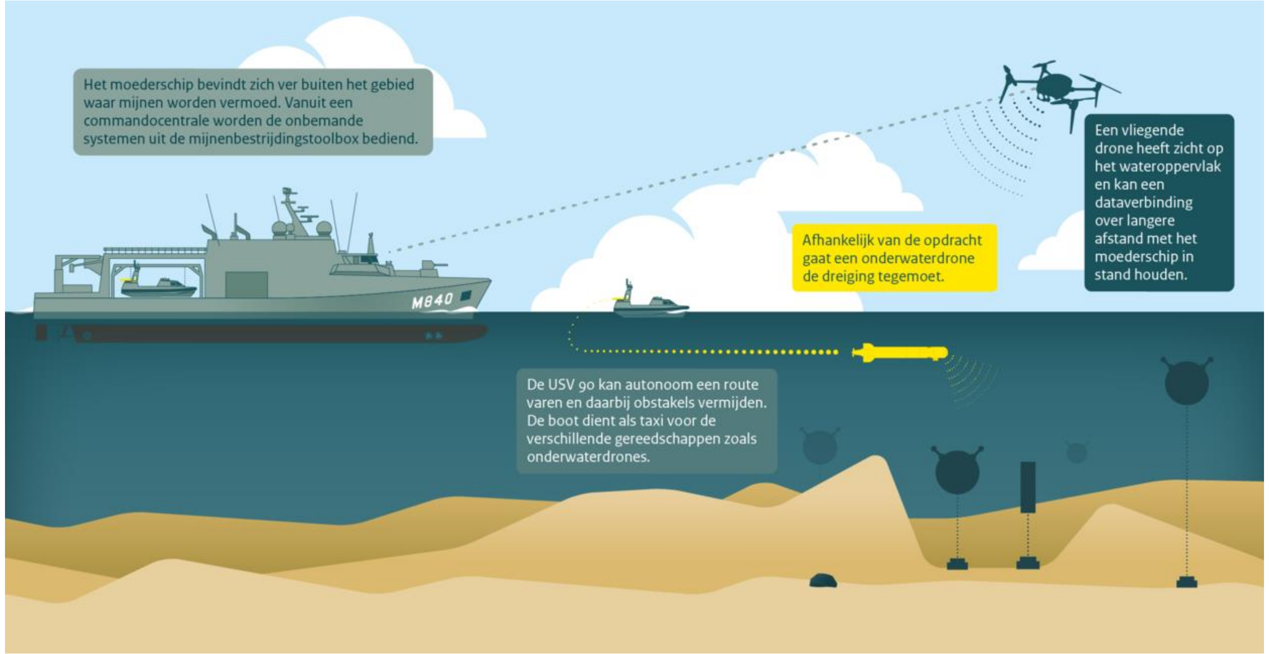
- Waar leg je de drempel van “redelijke twijfel” bij een strafzaak?

Het **onderscheidend vermogen** van een hypothesetoets is in dat geval gelijk aan $1 - \beta$, waarbij β de kans op een type-II fout is.

		Conclusie	
Werkelijkheid		H_0 is waar	H_0 is niet waar
	H_0 is waar (verdachte is onschuldig)	Terecht vrijgesproken (kans $1 - \alpha$)	Onterecht veroordeeld (type-I fout) (kans α)
	H_0 is niet waar (verdachte is schuldig)	Onterecht vrijgesproken (type-II fout) (kans β)	Terecht veroordeeld (kans $1 - \beta$)

Voorbeeld: besluitvorming in anti-submarine warfare (ASW)

Bij COMMIT is er een nieuw systeem ontworpen voor de USVs (unmanned surface vehicles) van de Koninklijke Marine voor het detecteren van zeemijnen. De Koninklijke Marine wil het systeem pas gebruiken als er voldaan wordt aan de operationele norm dat een zeemijn gemiddelde van meer dan 1250 meter afstand kan worden gedetecteerd. **Hoe kunnen we toetsen of hieraan wordt voldaan?**



Stappenplan hypothesetoetsen samengevat

Stap 1: definieer de nulhypothese H_0 en de alternatieve hypothese H_1

Stap 2: verzamelen van data

Stap 3: bepalen van het significantieniveau α

- Hoeveel risico wil je nemen met onterecht verwerpen van de nulhypothese?

Stap 4: bepalen van de toetsingsgrootheid

- Het significantieniveau bepaalt waar de grens ligt tussen handhaven / verwerpen van H_0

Stap 5: geef de conclusie van de hypothesetoets en formuleer deze in de originele context

Stap 1: definieer de nulhypothese en de alternatieve hypothese

Laat X de kansvariabele zijn die de afstand meet waarvandaan een willekeurige zeemijn kan worden gedetecteerd met het nieuwe systeem.

Nulhypothese H_0: de gemiddelde detectieafstand μ van een zeemijn met het nieuwe systeem voldoet niet aan de norm

$$\mu \leq \mu_0 = 1250$$

Alternatieve hypothese H_1: de gemiddelde detectieafstand μ van een zeemijn met het nieuwe systeem voldoet wel aan de norm:

$$\mu > \mu_0 = 1250$$

Stap 2: bepalen van het significantieniveau α

Hoeveel risico wil de Marine nemen met het onterecht verwerpen van de nulhypothese?

- Wat houdt het onterecht verwerpen van H_0 in?
  - Er wordt onterecht aangenomen dat het nieuwe systeem aan de norm voldoet.
  - Risico dat een zeemijn te laat wordt gedetecteerd!

Stel nu dat het significantieniveau α gekozen wordt op $\alpha = 0{,}01$

- 1% kans op onterecht verwerpen van H_0 als H_0 waar is.

Stap 3: verzamelen van data

Data: 50 experimenten met een USV die richting een zeemijn varen

- Op basis van verschillende hoeken van benadering, types zeemijnen, zeebodemprofielen, etc. → **representatieve steekproef!**
- Meten van de afstand waarop de zeemijn voor het eerst wordt gedetecteerd.
- Gemiddelde $\overline{x} = 1261$ meter en standaardafwijking $s = 42$ meter

Stap 4: bepalen van de toetsingsgrootheid

Theoretische steekproef:

- $X_1, X_2, \ldots, X_{50}$: metingen van detectieafstanden met gemiddelde μ / standaardafwijking σ
- We willen iets zeggen over de parameter μ
- **Centrale limietstelling:** onder H_0 geldt dat $\overline{X} \sim N\left(\mu = 1250; \frac{\sigma}{\sqrt{n}}\right)$

Toetsingsgrootheid:

- We bekijken het steekproefgemiddelde $\overline{X} \sim N\left(\mu = 1250; \frac{\sigma}{\sqrt{n}}\right)$
- Op basis van de aanname $\mu = 1250$ bekijken we wat "redelijke uitkomsten" voor $\overline{x}$ zijn.
- Hiervoor gebruiken we het gekozen significantieniveau α!

Stap 5: geef de conclusie en formuleer deze in de originele context

Methode 1: het kritieke gebied

- Hoger steekproefgemiddelde → waarschijnlijker dat aan de norm wordt voldaan.
- Het **kritieke gebied** (interval van onwaarschijnlijke uitkomsten voor $\overline{x}$ als H_0 waar is):

$$[g, \infty) = \left[\mu_0 + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}; \infty\right) = \left[1250 + 2{,}3263 \cdot \frac{42}{\sqrt{50}}; \infty\right) = [1263{,}82; \infty)$$

Aangezien het geobserveerde steekproefgemiddelde $\overline{x} = 1261$, ligt deze NIET in het **kritieke gebied** → H_0 wordt niet verworpen.

Er is onvoldoende bewijs om aan te nemen dat het nieuwe systeem een gemiddelde detectieafstand voor zeemijnen heeft die wel aan de norm voldoet.

Stap 5: geef de conclusie en formuleer deze in de originele context

Methode 2: de p-waarde

- Hoe groot is de kans op een nog **extremere waarde** voor het steekproefgemiddelde?
- Verwerp H_0 als $p < \alpha$ (kans op extremere waarde is kleiner dan α)

$$p = P\left(\overline{X} \geq \overline{x} = 1261\right) = \text{normalcdf}\left(a = 1261; b = 10^{99}; \mu = 1250; \sigma = 42/\sqrt{50}\right)$$
$$\approx 0{,}0320$$

Aangezien $p > \alpha = 0{,}01$, wordt de nulhypothese H_0 niet verworpen!

Er is onvoldoende bewijs om aan te nemen dat het nieuwe systeem een gemiddelde detectieafstand voor zeemijnen heeft die wel aan de norm voldoet.

Type-I en type-II fouten

	Conclusie		
		H_0 is waar	H_0 is niet waar
Werkelijkheid	H_0 is waar (systeem voldoet NIET aan de operationele norm)	Nieuwe sonarsysteem wordt op waarde geschat	Type-I fout
	H_0 is niet waar (sonarsysteem voldoet WEL aan de operationele norm)	Type-II fout	Nieuwe sonarsysteem wordt op waarde geschat

Doel: zoveel mogelijk juiste beslissingen (groene cellen)

Probleem: trade-off tussen type-I fouten en type-II fouten → **welke is erger?**

Type-I en type-II fouten

Stel nu dat de echte gemiddelde detectieafstand gelijk is aan $\boldsymbol{\mu = 1265}$ meter (dus $\boldsymbol{H_1}$ is waar). Wat is dan het onderscheidend vermogen van deze hypothesetoets $\boldsymbol{1-\beta}$?

- We hebben eerder het kritieke gebied bepaald, namelijk $[1263{,}82; \infty)$.
- Indien het werkelijke gemiddelde $\mu = 1265$ is, dan is het onderscheidend vermogen gelijk aan de kans $1-\beta$ dat de nulhypothese **terecht** wordt verworpen.
- Dit is hetzelfde als de kans dat $\overline{X} \sim N(\mu = 1265; \sigma = \frac{42}{\sqrt{50}})$ een waarde in het kritieke gebied aanneemt:

$$P\left(\overline{X} \geq 1263{,}82\right) = \text{normalcdf}\left(a = 1263{,}82;\ b = 10^{99}; \mu = 1265; \sigma = 42/\sqrt{50}\right)$$
$$\approx 0{,}5787$$

Interactieve plot: hypothesetoetsen

https://interactive-hypothesetoets.streamlit.app

Samenvatting

- Het proces van een hypothesetoets
- Effect van type-I en type-II fouten
- Onderscheidend vermogen van een hypothesetoets

Huiswerk:

- Lezen van A. Buijs: hoofdstuk 9.1 (blz. 281-289), 9.3 (blz. 293-297)
- Opdrachten:
  - Hoofdstuk 9: m1-m6, 9.3, 9.5, 9.7, 9.11, 9.18, 9.23

Volgende les: de chi-kwadraatverdeling