

Inleidend College HOVO Gebruik en Misbruik Statistiek



Deze slides worden op de website gezet!

Docenten:

Richard Gill, Willem van Zwet, Peter Grünwald
Web: <http://www.math.leidenuniv.nl/~gill/hovo>

- 17-10: **dopingcontrole** in de sport (prosecutor's fallacy)
- 24-10: **psychologisch** onderzoek (frequentie-statistiek, smokkelen met "protocol")
- 31-10: **klimaatverandering**, de hockeystick grafiek (gerommel met data en methodes, publicatiebias)
- 7-11: **Lucia de Berk**: prosecutor's fallacy (o.a.)
- **Geneeskunde, epidemiologie**,
- 12-12: **forensisch DNA**: goed gebruik van statistiek

Menu 1e Uur

1. Een van de **Meest Gemaakte Fouten**
2. **Sally Clark**: Moordenaar of Pechvogel?
3. Analyse van de Fout via **Frequentistische** statistiek
 - Achterafstatistiek
 - **Publicatie bias**
4. Analyse van de Fout via **Bayesiaans** redeneren
 - Wat is een **kans**? Wat is een **voorwaardelijke kans**?
 - **Prosecutor's Fallacy**
5. **Begrippenkader**

Een van de meest gemaakte fouten

- Stel, er is iets heel onverwachts gebeurd; iets met een "hele kleine kans".
- Men redeneert vaak als volgt:
...er is iets gebeurd met een kleine kans, dus het is vast "geen toeval geweest"
– er moet haast wel iets achter zitten

Een van de meest gemaakte fouten

- Stel, er is iets heel onverwachts gebeurd; iets met een "hele kleine kans".
- Men redeneert vaak als volgt:
...er is iets gebeurd met een kleine kans, dus het is vast "geen toeval geweest"
– er moet haast wel iets achter zitten
- Ook hoogopgeleiden, **inclusief wetenschappers** zijn geneigd zo te redeneren

Maar deze redenering is verkeerd!
Dingen met hele kleine kans gebeuren gewoon!

Take Home Message Nr 1



- Gebeurtenissen met een astronomisch kleine kans gebeuren voortdurend. **Meestal is dat gewoon toeval!**

Take Home Message Nr 1



- Gebeurtenissen met een astronomisch kleine kans gebeuren voortdurend. **Meestal is dat gewoon toeval!**
- Als je statistiek **correct** toepast, komt er uit heel veel analyses "dit zou best wel eens toeval kunnen zijn"
- ...maar het is **heel heel gemakkelijk** kleine foutjes te maken die het vaak veel intuitievere of gewenste resultaat opleveren: **geen toeval**
- Veel **misbruik van statistiek** in rechtzaal, geneeskunde, sport, psychologie,... is hierop terug te voeren

Voorbeeld 1: Sally Clark (1964-2007)



- Sally Clark's eerste kind stierf drie maanden na de geboorte
 - SIDS (wiegendood), dacht men
- Sally's tweede kind stierf ook kort na de geboorte
 - Op het lijkje werden een aantal plekken gevonden die op fysiek geweld *konden* wijzen

Voorbeeld 1: Sally Clark (1964-2007)



- Sally Clark's eerste kind stierf drie maanden na de geboorte
 - SIDS (wiegendood), dacht men
- Sally's tweede kind stierf ook kort na de geboorte
 - Op het lijkje werden een aantal plekken gevonden die op fysiek geweld *konden wijzen*
- Sir Roy Meadow, beroemd specialist op het gebied van SIDS, betoogde
 - (1) "de kans dat twee kinderen in dezelfde familie aan SIDS overlijden is 1 op 73 miljoen"
 - (2) "DUS de zaak is uitermate verdacht."

"one sudden infant death in a family is a tragedy, two is suspicious and three is murder unless proven otherwise"



- Sir Roy betoogde
 - (1) "de kans dat twee kinderen in dezelfde familie aan SIDS overlijden is 1 op 73 miljoen"
 - (2) "DUS de zaak is uitermate verdacht."

"one sudden infant death in a family is a tragedy, two is suspicious and three is murder unless proven otherwise"
- Kritiek richtte zich vooral op getal 1/73 miljoen (was inderdaad incorrect), maar het woordje "DUS" is minstens zo verkeerd

- Sir Roy betoogde
 - (1) "de kans dat twee kinderen in dezelfde familie aan SIDS overlijden is 1 op 73 miljoen"
 - (2) ~~"DUS de zaak is uitermate verdacht."~~

~~"one sudden infant death in a family is a tragedy, two is suspicious and three is murder unless proven otherwise"~~



- Sir Roy betoogde
 - (1) "de kans dat twee kinderen in dezelfde familie aan SIDS overlijden is 1 op 73 miljoen"
 - (2) ~~"DUS de zaak is uitermate verdacht."~~
- Betere analyse: er zijn twee hypothesen.

SCHULD: Een moeder doodde haar 2 kinderen

ONSCHULD: 2 kinderen in 1 gezin sterven aan SIDS
- **Beide** hypothesen hebben extreem kleine kans

Je mag dus *niet* concluderen:
"de kans dat Sally schuldig is, is heel groot"

Take-Home Message Nr 2



- Verdediging haalde er een vermaard statisticus bij om uit te leggen dat de aanwijzing voor moord niet sterk was, maar de rechter vond het niet nodig de statisticus ter zitting te vragen, want hij vond: "this is not rocket science"
- Maar: **het is wel** "rocket science"

De Essentie van de Fout

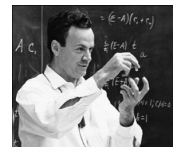
Stel, Jantje wint de loterij. Ik zeg: de kans dat Jantje zou winnen is zo klein (< 1 op 1 miljoen), dat het geen toeval kan zijn. Hij zal wel gefraudeerd hebben!

De Essentie van de Fout

Stel, Jantje wint de loterij. Ik zeg: de kans dat Jantje zou winnen is zo klein (< 1 op 1 miljoen), dat het geen toeval kan zijn. Hij zal wel gefraudeerd hebben!

Maar iemand *moet* de loterij winnen. Als Pietje had gewonnen, hadden we gezegd dat *dat* geen toeval was. Dat kan niet goed zijn!

Dingen met een kleine kans gebeuren gewoon



- Richard Feynman, een van de grootste natuurkundigen van de 20e eeuw:

"You know, the most amazing thing happened to me tonight. I was coming here, on the way to the lecture, and I came in through the parking lot. And you won't believe what happened. I saw a car with the license plate **ARW 357**. Can you imagine? **Of all the millions of license plates in the state, what was the chance that I would see that particular one tonight?** Amazing!"

Hoe dan wel statistiek bedrijven?

- Maar hoe kun je ooit valide statistische conclusies (bijv. "dit geneesmiddel is effectief") trekken als dingen met een hele kleine kans gewoon gebeuren?
- Statistici hebben twee methoden bedacht waarmee het toch kan:
 1. Frequentistische methode
 2. Bayesiaanse methode

Hoe dan wel statistiek bedrijven?

- Maar hoe kun je ooit valide statistische conclusies (bijv. "dit geneesmiddel is effectief") trekken als dingen met een hele kleine kans gewoon gebeuren?
- Statistici hebben twee methoden bedacht waarmee het toch kan:
 1. Frequentistische methode
 2. Bayesiaanse methode

(Er is een nog steeds niet geheel uitgewoede richtingensrijd over welke methode nu beter is. Dat is voorlopig niet van belang voor ons...het probleem zit 'm niet in welke methode men gebruikt, maar dat men methodes verkeerd gebruikt)

De Fout, Frequentistisch Bekeken

Stel, Jantje wint de loterij. Ik toets vervolgens twee hypothesen:

Hypothese H0: Er is niets aan de hand met Jantje vs.

Hypothese H1: Het is geen toeval
(bijv. Jantje fraudeert)

Ik zeg: de kans dat Jantje zou winnen is zo klein (< 1 op 1 miljoen), dat het geen toeval kan zijn

De Fout, Frequentistisch Bekeken

Stel, Jantje wint de loterij. Ik toets vervolgens twee hypothesen:

Hypothese H0: Er is niets aan de hand met Jantje vs.

Hypothese H1: Het is geen toeval
(bijv. Jantje fraudeert)

Ik zeg: de kans dat Jantje zou winnen is zo klein (< 1 op 1 miljoen), dat het geen toeval kan zijn

Maar iemand *moet* de loterij winnen. Als Pietje had gewonnen, hadden we gezegd dat *dat* geen toeval was. Dat kan niet goed zijn!

Correcte Frequentistische Statistiek

Jantje heeft de loterij gewonnen. Ik weet dat het best toeval kan zijn, maar ik vertrouw Jantje niet helemaal, want hij hing en hangt nog steeds met een verrekijker rond bij het gebouw van de staatsloterij. Daarom toets ik

Hypothese H0: Er is niets aan de hand met Jantje
vs.

Hypothese H1: Jantje fraudeert
met de toets "Jantje wint de *eerstvolgende* loterij"

Correcte Frequentistische Statistiek

Jantje heeft de loterij gewonnen. Ik weet dat het best toeval kan zijn, maar ik vertrouw Jantje niet helemaal, want hij hing en hangt nog steeds met een verrekijker rond bij het gebouw van de staatsloterij. Daarom toets ik

Hypothese H0: Er is niets aan de hand met Jantje
vs.

Hypothese H1: Jantje fraudeert
met de toets "Jantje wint de *eerstvolgende* loterij"

Als Jantje *nu weer* wint, dan zegt het wel degelijk wat, omdat we het van te voren voorspeld hebben!

Take Home Message Nr 3: **geen achteraf-frequentistische statistiek!**

- Frequentistische statistiek mag nooit gebruikt worden om een hypothese te toetsen op de data die de hypothese suggereerde. In plaats daarvan moet men *nieuwe* data verkrijgen, bijvoorbeeld door een experiment op te zetten
- Statistiek in rechtzaken (Lucia de Berk, Sally Clarke) is echter vrijwel altijd *achteraf* statistiek. Het is onethisch en soms zelfs onmogelijk om een experiment op te zetten om te kijken of het gevonden effect "herhaalbaar" is!
- Frequentistisch toetsen is dus meestal *niet toepasbaar* in dat soort zaken.

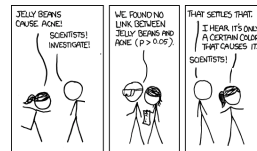
Frequentistisch Toetsen

- Vrijwel alle wetenschappelijke onderzoeken waarover u in de krant leest zijn gebaseerd op "nulhypothese toetsen", de standaard frequentistische methode
 - "aspirine helpt tegen hartinfarct", "met 3 koppen koffie op kun je beter opletten", "vleeseters gedragen zich onbeschoft", enz. enz.
- Wanneer goed uitgevoerd (dus bijv. niet "achteraf") zijn resultaten **in principe betrouwbaar**:
in slechts maximaal 5% van de onderzoeken zal de uitkomst zijn "er is iets aan de hand" terwijl er niets aan de hand is.
- Toch kan er van alles misgaan, bijvoorbeeld....

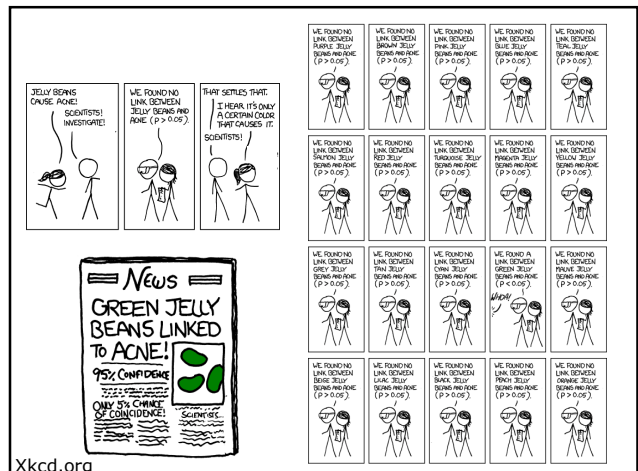
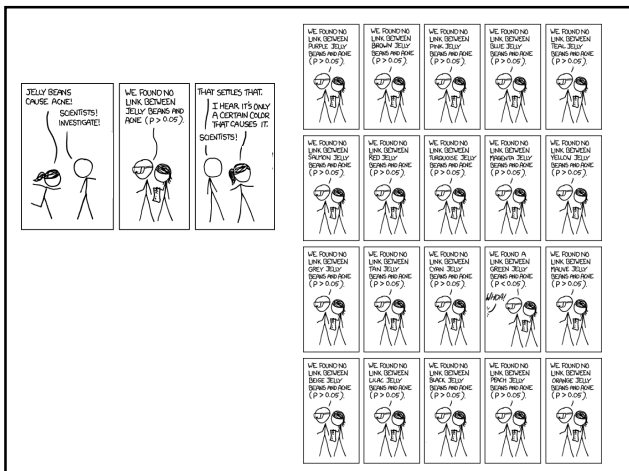
Publication Bias

J. Ioannidis, PLoS Medicine 2005:

"minstens 30% van alle veelgeciteerde onderzoeksresultaten in de geneeskunde kloppen niet, d.w.z. ze zijn niet reproduceerbaar"



"minstens 30% van alle veelgeciteerde onderzoeksresultaten in de geneeskunde kloppen niet, d.w.z. ze zijn niet reproduceerbaar"



Xkcd.org

Take Home Message Nr 4

- De resultaten waar wij over lezen/horen in het nieuws
bijv. zout is slecht voor u, rode wijn is goed voor u, Zuidafrikaanse honden zijn racistisch, linkshandigen leven 8 jaar korter, enz. enz. zijn een **hele vreemde selectie**
- Opvallende/politiek gewenste resultaten halen het nieuws, "saai" resultaten doen dat niet: **publication bias**
- Gemiddeld over **alle** goed uitgevoerde onderzoeken wordt er slechts 5% een fout gemaakt; maar gemiddeld over goed uitgevoerde maar **opvallende** resultaten worden er vermoedelijk veel meer fouten gemaakt!
- Enige oplossing: belangwekkende resultaten moeten **gerepliceerd** worden. In bijv psychologie gebeurt dat bijna nooit!

Beperking van Frequentistisch Toetsen

- Statistiek in rechtzaken (Lucia de Berk, Sally Clarke) vrijwel altijd **achteraf** statistiek.
- Frequentistisch toetsen is dus meestal **niet toepasbaar** in dat soort zaken.
- Bayesiaans redeneren is **wel** enigszins bruikbaar in "achteraf" gevallen, maar heeft zo zijn eigen probleempjes...(het blijft "rocket science")

Menu

- "Er is iets bijzonders, met hele kleine kans gebeurd"
- Redenen voor foutieve conclusie
"er moet wel iets aan de hand zijn"
 - Onderbuikstatistiek van de krantenlezer/rechter/... ("waar rook is, is vuur")
 - Verkeerd toegepaste frequentie-statistiek
 1. "Achteraf Statistiek"
 2. Publicatie-Bias
- Prosecutor's Fallacy
 1. Wat is een (voorwaardelijke) kans?
 2. Regel van Bayes

Wat is een "Kans"?

- Het wiskundig symbool voor een kans is P of Pr
- Kansens zijn gedefinieerd op **uitkomsten** van een experiment. De som van de kansen op alle uitkomsten is altijd 1.
- Voorbeeld: een **eerlijke** dobbelsteen:
de **uitkomstenruimte** is $\{1, 2, 3, 4, 5, 6\}$
- De kansen zijn:

$$\Pr(X = 1) = \frac{1}{6}; \Pr(X = 2) = \frac{1}{6} \dots \Pr(X = 6) = \frac{1}{6}$$

Het Aanpassen van Kansen

1. Ik gooi een eerlijke dobbelsteen; ik zie de uitkomst (een getal tussen 1 en 6) maar u niet
2. Ik vertel u ofwel "de uitkomst is even" ofwel "de uitkomst is oneven"
3. Stel dat ik u vertel "de uitkomst is even". Wat is volgens u dan de kans dat er "4" is gegooid?

Het Aanpassen van Kansen

1. Ik gooi een eerlijke dobbelsteen; ik zie de uitkomst (een getal tussen 1 en 6) maar u niet
2. Ik vertel u ofwel "de uitkomst is even" ofwel "de uitkomst is oneven"
3. Stel dat ik u vertel "de uitkomst is even". Wat is volgens u dan de kans dat er "4" is gegooid?
 - U zegt: er zijn nog drie mogelijkheden over. Die hebben allemaal gelijke kans, dus: **de kans op "4" is nu 1/3**

Voorwaardelijke Kansen

- Eerst was de kans 1/6
- U past deze kans aan omdat u nieuwe informatie heeft; dit heet **conditioneren**
- De kans is nu 1/3 geworden
- We zeggen: "de **conditionele** (of "voorwaardelijke") kans op "X=4", **gegeven** dat "X is even", is 1/3"

Voorwaardelijke Kansen

- Eerst was de kans 1/6
- U past deze kans aan omdat u nieuwe informatie heeft; dit heet **conditioneren**
- De kans is nu 1/3 geworden
- We zeggen: "de **conditionele** kans op "X=4", **gegeven** dat "X is even", is 1/3"

$$\Pr(X = 4 \mid X \in \{2, 4, 6\}) = \frac{\Pr(X = 4)}{\Pr(X \in \{2, 4, 6\})} = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{3}$$

↑
streep betekent: "gegeven dat"

The Prosecutor's Fallacy, vb. 1

- Wat is de kans dat een man langer dan 1m90 is, als gegeven is dat hij professioneel basketballer is?
– vrij groot

The Prosecutor's Fallacy, vb. 1

- Wat is de kans dat een man langer dan 1m90 is, als gegeven is dat hij professioneel basketballer is?
– vrij groot
- Wat is de kans dat een man professioneel basketballer is, als gegeven is dat hij langer dan 1m90 is?
– erg klein

The Prosecutor's Fallacy, vb. 1

- Wat is de kans dat een man langer dan 1m90 is, als gegeven is dat hij professioneel basketballer is?
– vrij groot
- Wat is de kans dat een man professioneel basketballer is, als gegeven is dat hij langer dan 1m90 is?
– erg klein
- Als wiskundige formule:

$$\Pr(A | B) \neq \Pr(B | A)$$

De kans op **A** gegeven **B** is niet hetzelfde als de kans op **B** gegeven **A**
A en B zijn gebeurtenissen

The Prosecutor's Fallacy, vb. 2

- De ELIZA aids test wordt in de VS gebruikt om bloeddonthaties te screenen.
- Als een donor aids heeft, is de testuitslag in 98% van de gevallen positief. Als een donor geen aids heeft, is de testuitslag in 93% van de gevallen negatief.
- Ans is bloeddonthor en test positief. Wat is de kans dat ze aids heeft?

The Prosecutor's Fallacy, vb. 2

- De ELIZA aids test wordt in de VS gebruikt om bloeddonoraties te screenen.
- Als een donor aids heeft, is de testuitslag in 98% van de gevallen positief. Als een donor geen aids heeft, is de testuitslag in 93% van de gevallen negatief.
- Ans is bloeddonor en test positief. Wat is de kans dat ze aids heeft? *Antwoord: ongeveer 3% (!!!)*

The Prosecutor's Fallacy, vb. 2

- De ELIZA aids test wordt in de VS gebruikt om bloeddonoraties te screenen.
- Als een donor aids heeft, is de testuitslag in 98% van de gevallen positief. Als een donor geen aids heeft, is de testuitslag in 93% van de gevallen negatief.
- Ans is bloeddonor en test positief. Wat is de kans dat ze aids heeft? *Antwoord: ongeveer 3% (!!!)*
- Reden: slechts 0.23% van Amerikaanse vrouwen heeft aids

The Prosecutor's Fallacy, vb. 2

- De ELIZA aids test wordt in de VS gebruikt om bloeddonoraties te screenen.
- Als een donor aids heeft, is de testuitslag in 98% van de gevallen positief. Als een donor geen aids heeft, is de testuitslag in 93% van de gevallen negatief.
- Ans is bloeddonor en test positief. Wat is de kans dat ze aids heeft?

$$3\% \Pr(\text{Ans heeft aids} \mid \text{test zegt "aids"}) \neq 98\% \Pr(\text{test zegt "aids"} \mid \text{Ans heeft aids})$$

The Prosecutor's Fallacy

- De "Prosecutor's Fallacy" is het verwisselen van de kans op de "hypothese" gegeven de "evidence" met de kans op de "evidence" gegeven de "hypothese"
- De *juiste* manier om deze twee aan elkaar te relateren is de formule van Bayes

De Stelling van Bayes

Posterior odds = likelihood ratio * prior odds

$$\frac{\Pr(A | B)}{\Pr(\bar{A} | B)} = \frac{\Pr(B | A)}{\Pr(B | \bar{A})} \cdot \frac{\Pr(A)}{\Pr(\bar{A})}$$

- A afkorting van "A gebeurt";
 - \bar{A} afkorting van "A gebeurt **niet**"
- Altijd geldt: $\Pr(\bar{A}) = 1 - \Pr(A)$

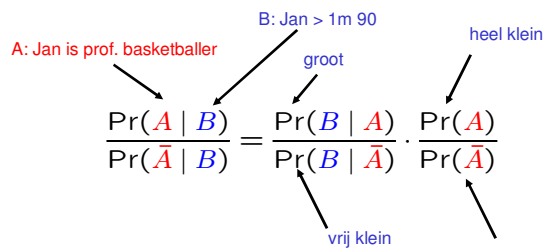
De Stelling van Bayes

A posteriori kans op A A priori kans op A

$$\frac{\Pr(A | B)}{\Pr(\bar{A} | B)} = \frac{\Pr(B | A)}{\Pr(B | \bar{A})} \cdot \frac{\Pr(A)}{\Pr(\bar{A})}$$

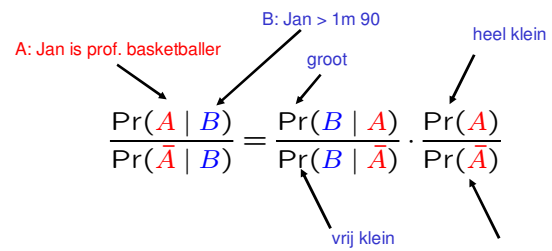
- A afkorting van "A gebeurt";
 - \bar{A} afkorting van "A gebeurt **niet**"
- Altijd geldt: $\Pr(\bar{A}) = 1 - \Pr(A)$

De Stelling van Bayes, basketbal



- A afkorting van "A gebeurt";
- \bar{A} afkorting van "A gebeurt **niet**"

De Stelling van Bayes, basketbal



vrij klein = groot * heel klein

De Stelling van Bayes, **aidstest**

$$\frac{\Pr(A | B)}{\Pr(\bar{A} | B)} = \frac{\Pr(B | A)}{\Pr(B | \bar{A})} \cdot \frac{\Pr(A)}{\Pr(\bar{A})}$$

A: Ans heeft aids B: test zegt "Ans heeft aids"
 0.98: bijna 1 0.0023: HEEL klein
 0.07: vrij klein bijna 1

- A afkorting van "A gebeurt";
- \bar{A} afkorting van "A gebeurt niet"

$\Pr(\bar{A}) = 1 - \Pr(A)$

De Stelling van Bayes, **aidstest**

$$\frac{\Pr(A | B)}{\Pr(\bar{A} | B)} = \frac{\Pr(B | A)}{\Pr(B | \bar{A})} \cdot \frac{\Pr(A)}{\Pr(\bar{A})}$$

A: Ans heeft aids B: test zegt "Ans heeft aids"
 bijna 1 HEEL klein
 vrij klein bijna 1

vrij klein = groot * heel klein

De Stelling van Bayes, **Sally Clark**

$$\frac{\Pr(A | B)}{\Pr(\bar{A} | B)} = \frac{\Pr(B | A)}{\Pr(B | \bar{A})} \cdot \frac{\Pr(A)}{\Pr(\bar{A})}$$

A: Sally schuldig B: Sally's 2 babies sterven HEEL klein
 1 heel klein Vrijwel 1

Conclusie Prosecutor's Fallacy

- De "prosecutor's fallacy" is een redeneerfout die in sommige contexten **evident** is (basketbal), maar in andere contexten regelmatig gemaakt wordt (aidstest) en **niet "gespot" wordt** (Claudia Pechstein)
 - In **natuurlijke taal** is het verschil tussen $\Pr(A | B)$ en $\Pr(B | A)$ vaak niet goed uit te drukken
- De redeneerfout "als iets met kleine kans gebeurt, is het geen toeval" is in feite dezelfde redeneerfout als "the prosecutor's fallacy".

Take Home Message Nr 5: Prosecutor's Fallacy

- In **natuurlijke taal** is het verschil tussen

$$\Pr(A | B)$$
en

$$\Pr(B | A)$$
vaak niet goed uit te drukken
- Dat gaat vaak mis in rechtzaken maar ook in de manier waarop journalisten en krantenlezers statistische resultaten interpreteren!

Menu

- “Er is iets bijzonders, met hele kleine kans gebeurd”
- Redenen voor foutieve conclusie “er moet wel iets aan de hand zijn”
 - De onderbuikstatistiek van de krantenlezer (“waar rook is, is vuur”)
 - Verkeerd toegepaste frequentie-statistiek
 1. “Achteraf Statistiek”
 2. Publicatie-Bias
 - **Prosecutor's Fallacy**
 1. **Wat is een (voorwaardelijke) kans?**
 2. **Regel van Bayes**

Conclusie 1e Uur

- Theorie
 - Frequentistische en Bayesiaanse methoden
 - Kansen
 - Voorwaardelijke Kansen (aanpassen van kansen bij nieuwe informatie)
 - Regel van Bayes
- Misbruik van Statistiek:
 - Kleine kans betekent *niet* “het was geen toeval”
 - Prosecutor's Fallacy
 - Achteraf Statistiek
 - Publication Bias
- **Goed, Nuttig** gebruik van Statistiek: 2e Uur!