

Methoden van het Wetenschappelijk Onderzoek

Zin en onzin van statistiek

Statistiek komt ernstig over... of niet

- "Deze tandpasta helpt tegen caries in 1 op 2 gevallen."
 - Het werd slechts geprobeerd op 4 personen.
- "De gemiddelde jonge informaticus verdient 1550 €."
 - Dit is een sample, men heeft echt niet alle jonge informaticussen bevroegd. Mensen liegen vaak over hun loon. En wat is de spreiding?
- "Windows XP is stabielere dan andere besturingssystemen."
 - Hoe meet je dat? Kan je dat sowieso objectief meten? Spelen hier geen "religieuze" kwesties mee? Wie waren je respondenten?
- "Knack wordt meer gelezen dan *Dag Allemaal*."
 - En toch wordt *Dag Allemaal* meer verkocht? Zouden respondenten soms liegen over wat ze lezen?
- "De effectiviteit van dit product werd wetenschappelijk getest."
 - Let op, er staat niet "wetenschappelijk bewezen"; waarschijnlijk helpt het product dus helemaal niet.

Waarom toch statistiek?

- Dient om onderzoek kwantitatief te onderbouwen.
- Dient om gegevens te "comprimeren".
- Twee soorten statistiek
 - Descriptieve statistiek
 - Inferentiële statistiek

"Statistiek is vervalend"

- Toch is statistiek de enige juiste manier om wetenschappelijk onderzoek correct te rapporteren.
- Een basiskennis van statistiek laat ook toe om wetenschappelijk onderzoek kritisch te **evalueren** en te **vergelijken**.

Het verzamelen van data

- Misschien wel het belangrijkste is het verzamelen van je data: **hier worden de meeste fouten (bewust of onbewust) gemaakt**.
- Voorbeeld
 - De NMBS die de stiptheid van de treinen nagaat.
 - De onveiligheidspoll van de federale regering.
 - Enquête naar seksuele ervaringen bij jongeren.
 - Enquête naar hygiëne.
- Die noemt men in de statistiek een "**bias**".
- Wees bewust van bias, en vermeld steeds bias bij je statistiek!

Verzamelen van data dmv. enquête

- Hoe wordt de data vergaard?
- Hoeveel deelnemers zijn er/welk percentage van de populatie is dit?
- Exclusiecriteria.
- Respons.
- Vertekent de niet-respons de resultaten?
- Hoe is de vragenlijst opgesteld?
- Betrouwbaarheid van de vragenlijst?
- Wie zijn de interviewers?
- Opleiding en opvolging van de interviewers?
- Uniformiteit van de omstandigheden van het interview?
- Externe factoren?

Gegevens verzamelen: methodes

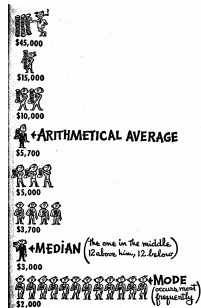
- Willekeurig verzamelen
 - Elk geval heeft een gelijke kans om opgenomen te worden in de statistiek.
 - Nodig: een welomtinnende populatie, een lijst met alle gevallen en een random generator.
- Systematisch verzamelen
 - Het eerste geval is willekeurig, de rest wordt volgens een bepaalde procedure gekozen. Bv. Start bij een willekeurig rolnr. en daarna increment met 10.
 - Vergroot niet echt een fout, behalve bij een vertekende populatie.
- Proportioneel gestratificeerd verzamelen
 - Per groep in de populatie neem je een aantal gevallen, gelijk aan de verhouding in de populatie.
- Disproportioneel gestratificeerd verzamelen.

Hoeveel samples?

- Hoeveel samples heb je nodig?
- Er is een vuistregel: **meer is beter**.
 - Natuurlijk is het verzamelen van gegevens beperkt door tijd, geld en ruimte.
- Vermeld altijd het aantal samples.
 - Dat is altijd het vergeten nummertje... Want zo kan je zelfs aantonen dat de kans dat je kop werpt 80% is.

Rapporteren van statistiek

- Het "gemiddelde", rekenkundig gemiddelde, mediaan, of modus?
 - Gemiddelde (mean): $\bar{x} = \frac{\sum x}{N}$
 - Mediaan: sorteren, er liggen evenveel waarden onder als boven de mediaan.
 - Modus: waarde welke het meest voorkomt (de populairste waarde).



Maat van dispersie

- Als je gemiddelden rapporteert, moet er steeds een maat van dispersie bij!
- Een goede maat van dispersie houdt rekening met
 - Alle gegevens.
 - Beschrijft de gemiddelde afwijking van de scores ten opzichte van het gemiddelde.
 - Neemt toe als de heterogeniteit van de gegevens toeneemt.
- Uitersten van de gegevens.
- Gemiddelde deviate: $GD = \frac{\sum |x_i - \bar{x}|}{N}$
- Variatie: $V = \frac{\sum (x_i - \bar{x})^2}{N}$

Standaard afwijking

$$s = \sqrt{V} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N}}$$

De "semiattached figure"

- Als je iets niet echt statistisch kan aantonen, toon dan iets aan wat er op lijkt.
 - Bv. In een streng laboratoriumonderzoek is aangetoond dat dit middel tegen verkoudheden gemiddeld $56\% \pm 7\%$ van de ziektekiemen doodde in 2.2 ± 0.5 dagen tijd.
 - "Vier maal meer verkeersdoden om 7h dan om 19h"
- Het niet vergelijken van je data met een referentie.
 - Bv. "26% meer sap!!!"

Correlatie

- Een krachtig gereedschap in statistiek het berekenen van correlatie (met correlatiematrix, regressiecoëfficiënten, ...)
- Echter correlatie wordt al te vaak misbruikt om causale verbanden te leggen tussen twee factoren.
 - Sommige klinken aannemelijk: "kinderen van ouders met laag inkomen hebben slechtere studieresultaten".
 - Maar je kan echt alles correleren als je wil: "leesvaardigheid hangt af van de schoenmaat"
- Als er een causal verband is, dan is een correlatie.
- Een causaal verband werkt in twee richtingen, soms moeilijk te onderscheiden in welke richting.

Reïfificatie

- Reïfificatie = "verdingelijking"
- Een complex fenomeen wordt in een numerieke voorstelling gegoten, waarbij alle nuancering verloren gaat.
- Bv.
 - IQ
 - Temperatuur

Afronden van getallen

- Als je een getal geeft, rond dan steeds netjes af.
 - Vuistregel: twee significante cijfers van de standaard deviatie gemiddelde = 98.2346321 st.dev. = 0.00342879
wordt 98.2346 ± 0.0034

Aan de slag met statistiek

- Als je statistiek wenst te gebruiken voor het rapporteren van onderzoek.
 - Gebruik dan de statistische methodes van gelijkaardig onderzoek, dit laat vergelijking toe.
 - Zoek een goed rect-toe-recht-aan boek (vaak te vinden bij "statistically challenged fields").

- Benjamin Disraeli: "There are three kinds of lies: lies, damn lies, and statistics."

Mieke Vogels wacht op e-mailtjes van mannen

De Standard, 21/06/2010



BRUSSEL -- De Vlaamse minister van Gelijke Kansen, Mieke Vogels, vraagt dat 'oude' en 'nieuwe' mannen haar mailen en schrijven. Zij doet dit naar aanleiding van de VUB-studie waaruit blijkt dat de nieuwe man niet bestaat.

„Mannen hebben wel eens de neiging te denken dat ze veel doen in het huis houden, maar de realiteit bewijst vaak het tegendeel”, zegt minister Vogels. „Maar de studie is een steekproef en de vlakken de goede uitzonderingen weg. Er zijn wel mannen die deeltijds gaan werken en de huishoudelijke taken en de zorg om de kinderen gelijk verdelen met hun partner. Die worden nu een beetje ontmoeid door de studie. Met mijn actie wil ik onder meer de verhalen van die moderne mannen laten zien.”

Ze vindt het wel positief dat de studie aantoont dat zowel mannen als vrouwen meer *kwaaftijd* besteden aan de opvoeding en zorg voor de kinderen.

„Het onderzoek bevestigt dat een ruim maatschappelijk debat over de combinatie van arbeid, gezin en zorg nodig is. Hierin wil ik een voortrekkersrol spelen. Ik wil iedereen bij het begin van zijn loopbaan drie rugzakjes meegeven met drie verschillende kredieten waaruit hij of zij kan putten: een krediet voor zorg (om bijvoorbeeld voor kinderen te zorgen), een krediet voor opleiding en een krediet voor zijn eigen dromen.”

„Ik wil hiervoor een sensibiliseringscampagne opzetten en vraag aan alle mannen ideeën en voorstellen aan te brengen die ik hiervoor kan gebruiken.”