

## Statistiek: betrouwbaarheidsintervallen

Havo 5, wiskunde A, schooljaar 17/18

### Algemeen

Bepaal of je variabelen kwalitatief of kwantitatief zijn. Op basis daarvan beslis je welke van de onderstaande typen betrouwbaarheidsinterval gaat gebruiken.

$$\text{soorten variabelen} \begin{cases} \textit{kwalitatief} \\ \textit{kwantitatief} \begin{cases} \textit{discreet} \\ \textit{continu} \end{cases} \end{cases}$$

Als je een 95%-betrouwbaarheidsinterval opstelt, dan zeg je eigenlijk:

“Stel ik zou mijn onderzoek 100 maal herhalen, dan verwacht ik dat de werkelijke waarde ( $p$  of  $\bar{X}$ ) minstens 95 keer binnen het (voor ieder onderzoek apart opgestelde) 95% betrouwbaarheidsinterval ligt.”

### Betrouwbaarheidsinterval bij kwalitatieve variabelen

Als je onderzoek hebt gedaan naar kwalitatieve eigenschappen, dan zijn je waarnemingen uit te drukken in woorden. Denk aan geslacht: “man/vrouw”, profiel: “CM, EM, NG, NT”, haarkleur: “rood, blond, bruin, zwart, grijs” etc. Dus zijn de gegeven getallen de frequenties per categorie. Bijvoorbeeld bij een steekproef is het aantal mannen = 300 en het aantal vrouwen = 400, de totale steekproefomvang (totale frequentie)  $n = 700$ .

In dit soort gevallen kun je alleen de steekproefproportie en de standaardafwijking bepalen met de volgende formules:

- $\hat{p} = \frac{\textit{frequentie van de betreffende categorie}}{n = \textit{steekproefomvang}}$
- $\sigma = \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$

Er geldt nu dat:

- Het 68%-betrouwbaarheidsinterval =  $[\hat{p} - \sigma; \hat{p} + \sigma]$
- Het 95%-betrouwbaarheidsinterval =  $[\hat{p} - 2\sigma; \hat{p} + 2\sigma]$

Let op: rond tussendoor niet af en in je antwoord op drie decimalen, tenzij dit anders in de vraag vermeld is.

Deze intervallen bestaan altijd uit getallen die tussen 0 en 1 liggen! Het zijn getallen die (met de gegeven betrouwbaarheid) aangeven wat de te verwachten verhouding van een bepaalde categorie in de populatie of steekproef zal zijn.

### Betrouwbaarheidsinterval bij kwantitatieve variabelen

Op het moment dat je onderzoek betrekking heeft op kwantitatieve eigenschappen, dan zijn je waarnemingen zelf juist getallen. Denk aan lengte, gewicht, salaris, etc. Je waarnemingen hebben daarom een gemiddelde en een standaardafwijking uitgedrukt in een eenheid, bv. cm, kg, euro. Let op: dit is anders dan bij kwalitatieve variabelen, daar kun je geen eenheid achter de steekproefproportie of de standaardafwijking zetten.

Bij een steekproef met kwantitatieve variabelen spreken we dus van:

- het steekproefgemiddelde  $\bar{X}$ , met de eenheid van de betreffende variabele
- de steekproefstandaardafwijking  $S$ , met de eenheid van de betreffende variabele
- de steekproeflengte  $n$

Er geldt nu dat:

- Het 68%-betrouwbaarheidsinterval =  $\left[ \bar{X} - \frac{S}{\sqrt{n}}; \bar{X} + \frac{S}{\sqrt{n}} \right]$
- Het 95%-betrouwbaarheidsinterval =  $\left[ \bar{X} - 2 \cdot \frac{S}{\sqrt{n}}; \bar{X} + 2 \cdot \frac{S}{\sqrt{n}} \right]$

Let op: rond tussendoor niet af en in je antwoord op één decimaal meer dan de gegevens, tenzij dit anders in de vraag vermeld is.

Deze intervallen bestaan altijd uit getallen die qua grootte in de buurt van het steekproefgemiddelde liggen. Het zijn getallen die (met de gegeven betrouwbaarheid) aangeven wat het te verwachten gemiddelde in de populatie of de volgende steekproef zal zijn.