5. Statistische gevolgtrekking: schatten

Punt en intervalschattingen

Hoe gebruik je steekproef data voor het schatten van populatie parameters?

Met kwantitatieve variabelen schat je het populatie gemiddelde (bijv. hoeveel geld er gemiddeld is besteed aan medicijnen in 2011). Met categoriale variabelen schat je populatie proporties voor de categorieën (bijv. wie er wel of geen zorgverzekering heeft in 2011).

Er zijn twee typen parameterschattingen:

- Puntschatting (een getal dat de beste schatting is).
- Interval schatting (een interval rond een puntschatting, waarvan je denkt dat de populatie parameter er in valt).

Er is een verschil tussen een estimator en een estimate point (schatting), namelijk dat een estimator de manier is waarop je de schatting maakt (het schatten op zich), en een estimate point het getal dat eruit komt. Zo is je steekproef een estimator voor je populatie parameter, en is 0.73 (bijvoorbeeld) een estimate point voor je populatie gemiddelde. Dit betekent dat een bepaalde categorie op 73% wordt geschat. De korte naam van 'estimate point' is 'estimate'.

Puntschatting van parameters

Een goede schatting heeft een steekproefdistributie die 1) gecentreerd is rond de parameter, en 2) een zo klein mogelijke standaardfout heeft.

1: gecentreerd rond de parameter:

Een schatting is 'niet vertekend' (unbiased) wanneer de steekproefdistributie gecentreerd is rond de parameter. Helemaal natuurlijk wanneer het steekproefgemiddelde ook daadwerkelijk de populatieparameter is. Dus dan is $\bar{y}$ (steekproefgemiddelde) gelijk aan μ (populatiegemiddelde). $\bar{y}$ is dan een goede estimator voor μ.

Een schatting kan ook 'vertekend' (biased) zijn en dan is het steekproefgemiddelde geen goede schatting voor het populatiegemiddelde. Meestal zit het steekproefgemiddelde er dan onder, want de extremen in de steekproef kunnen nooit meer zijn dan die uit de populatie, alleen maar minder. Dus de distributie en variatie in de steekproef is dan kleiner, waardoor de steekproefvariatie de populatievariatie onderschat.

2: Een zo klein mogelijke standaardfout:

Een schatting is 'efficiënt' wanneer de standaardfout kleiner is dan die van andere schattingen (dit is de standaardfout van de estimators gemiddelde, mediaan etc. Dat zijn allemaal estimators).

Stel dat je een normale verdeling hebt. Bij een normale verdeling is de standaardfout van de mediaan altijd 25% groter dan de standaardfout van het gemiddelde. Het gemiddelde van de steekproef ligt dichterbij het gemiddelde van de populatie dan de steekproefmediaan. Het steekproefgemiddelde is dan een efficiëntere estimator dan de steekproefmediaan.

Samenvattend: een goede estimator is onpartijdig (unbiased; de steekproefdistributie is gecentreerd rond de parameter) en efficiënt (kleinste standaardfout).

Meestal gebruik je gewoon het steekproefgemiddelde als estimator voor het populatiegemiddelde, de steekproefstandaardafwijking als estimator voor de populatiestandaardafwijking, etc.

R.A. Fisher ontwikkelde de 'meest aannemelijke schatter'. Dit is een schattingsmethode die als schatting van een parameter die waarde kiest, waarvoor de aannemelijkheidsfunctie maximaal is. Hoe aannemelijk een parameterwaarde is, wordt gemeten aan de kans op het vinden van een steekproefuitkomst bij die waarde van de parameter.
Deze manier heeft drie voordelen, met name bij grote steekproeven: 1) ze zijn efficiënt: andere estimators hebben geen kleinere standaardfouten en liggen ook niet dichterbij de parameter, 2) ze zijn niet vertekend (minder vertekening wanneer de steekproef groter wordt), en 3) ze hebben meestal een normale steekproefverdeling.

Intervalschatting

Een betrouwbaarheidsinterval is een intervalschatting voor een parameter. In dit interval zitten betrouwbare schattingen van de parameter. Je kijkt hiervoor naar de distributie van de steekproef, welke vaak een normale verdeling is. Wanneer je een betrouwbaarheidsinterval wilt met 95% zekerheid, valt de schatting van de parameter binnen twee standaardafwijkingen van het gemiddelde. In de praktijk vermenigvuldig je eerst de standaardfout met de z-waarde. De uitkomst tel je dan bij de puntschatting op en trek je van de puntschatting af. Je krijgt twee getallen, die samen het betrouwbaarheidsinterval vormen. Je kunt nu met 95% zekerheid zeggen dat een populatieparameter tussen deze twee getallen ligt. De z-waarde maal de standaardfout noem je ook wel de 'foutmarge' (margin of error).

Dus een intervalschatting is: de puntschatting ± de foutmarge.

Als je een intervalschatting wilt met 95% zekerheid, dan moet je een puntschatting ± een foutmarge doen (die foutmarge moet dan gelijk zijn aan twee standaardfouten).

Betrouwbaarheidsinterval voor een proportie

Nominale en ordinale variabelen zorgen voor categoriale data (Bijv. 'mee eens' – 'niet mee eens'). Als je hier uitspraken over wilt doen, kun je geen gemiddelden berekenen. Je gebruikt dan proporties of percentages. Een proportie valt tussen de 0 en de 1, en een percentage tussen de 0 en de 100.

De onbekende proportie van een populatie wordt aangeduid met het teken: π . De punt schatter van een populatie proportie is de 'steekproef proportie'. Hiermee schat je de populatie proportie. Je geeft de steekproef proportie aan met het teken: $\hat{\pi}$ (want het is een schatting van de daadwerkelijke proportie).

De z-waarde geeft aan hoe vaak de standaardfout vermenigvuldigd moet worden. Het zit zo in elkaar dat de kans onder een normale verdeling binnen z standaardfouten vanaf het gemiddelde gelijk is aan het betrouwbaarheidsniveau. Voor een betrouwbaarheidsniveau van 95% en 99% is z gelijk aan 1.96 en 2.58. De steekproefgrootte n moet zo groot zijn dat op zijn minst 15 observaties in de categorie van de behandelde proportie vallen en 15 buiten de categorie vallen, anders werkt het betrouwbaarheidsinterval niet.

Betrouwbaarheidsinterval voor populatie proportie $B.I. = \hat{\pi} \pm z \cdot (se)$ waarbij $se = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$

Hieruit kun je afleiden dat een grotere steekproef een meer accuraat betrouwbaarheidsinterval zal geven. Een grotere n zorgt voor een kleinere standaardfout, en een preciezer betrouwbaarheidsinterval. Meer specifiek: de steekproefgrootte moet verviervoudigen om de precisie te verdubbelen.

Betrouwbaarheidsinterval voor een gemiddelde

Ook hierbij heeft de betrouwbaarheidsinterval de vorm : puntschatting ± foutmarge. De foutmarge bestaat hier uit een t-waarde (in plaats van een z-waarde) maal de standaardfout.

De standaardfout wordt berekend door de standaarddeviatie van de steekproef (s) te delen door de wortel van de steekproefgrootte (n). De puntschatting is in dit geval het steekproefgemiddelde, aangeduid met het teken: $\bar{y}$.

Betrouwbaarheidsinterval voor populatie gemiddelde:

$$B.I. = \bar{y} \pm t_{.025} \cdot (se) \quad \text{met} \quad se = \frac{s}{\sqrt{n}}$$

Kenmerken t-verdeling

1. klok-vormig en symmetrisch vanaf het gemiddelde
2. Standaarddeviatie is iets groter dan 1. De precieze waarde ervan hangt af van de vrijheidsgraden (df). Deze worden berekend aan de hand van de vrijheidsgraden (n – 1).
3. Hoe groter de vrijheidsgraden (df), hoe meer de t-verdeling gaat lijken op een normaal verdeling. De verdeling wordt steeds puntiger. Bij df > 30 zijn ze bijna identiek.

Tabel: Schattingsmethoden voor gemiddelden en proporties.

Parameter	Puntschatting	Geschatte standaardfout	Betrouwbaarheidsinterval
Gemiddelde μ	$\bar{y}$	$se = \frac{s}{\sqrt{n}}$	$\bar{y} \pm t \cdot (se)$
Proportie π	$\hat{\pi}$	$se = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$	$\hat{\pi} \pm z \cdot (se)$