

Bijlage 5 Methodologie

Methodologie

Om tot een betrouwbaar onderzoek te komen, moet de gekozen methodologie statistisch onderbouwd, toepasselijk en consistent zijn. Met name bij regressieanalyse spelen er vele assumpties waar rekening mee gehouden dient te worden bij het testen van de uitkomsten. Alle analyses in het onderzoek van Statisfact zijn op econometrische wijze verantwoord en daarom betrouwbaar.

Betrouwbaarheid gemiddelden

De gemiddelde gemeten waarde uit een steekproef van een onderzoek geeft informatie over de populatie die onderzocht wordt. Maar gemiddeldes kunnen ongeacht betrouwbaarheid altijd uitgerekend worden. De variatie (of eigenlijk, variantie) van de waarden kan fluctueren. Om te bepalen hoe betrouwbaar een gemeten gemiddelde is, wordt het betrouwbaarheidsinterval opgesteld. Standaard is dit op basis van 95%, wat wil zeggen dat er een 95% kans bestaat dat een waarde die in de populatie gemeten zou worden ook binnen dit interval valt. Hier kan echter afhankelijk van de parameters van het onderzoek, zoals respons en steekproefgrootte, ook voor een ander percentage gekozen worden.

Het aantal vrijheidsgraden binnen de steekproef geeft de t-waarde. Deze waarde is, vermenigvuldigd met de standaardfout, de afwijking (naar boven of beneden) die het steekproefgemiddelde kan hebben binnen het betrouwbaarheidsinterval. Met andere woorden, de werkelijke waarde in de populatie zal voor 95% zeker niet meer afwijken dan $[t\text{-waarde} * s.e.]$ van de gemeten gemiddelde waarde in de steekproef.

Doordat het betrouwbaarheidsinterval zou kunnen overlappen met meerdere antwoordklassen (zoals groen, oranje en rood), is het van belang om te controleren of deze klassenindeling niet gevoelig is voor bovengenoemde variatie.

Betrouwbaarheid regressies

Om causale verbanden te onderzoeken tussen variabelen onderling kunnen regressieanalyses uitgevoerd worden. Hiermee kunnen lineaire relaties onderzocht worden naar richting en grootte. De betrouwbaarheid van een verband kan naar significantie gemeten worden met de p-waarde, welke doorgaans op basis van 95% betrouwbaarheid getest wordt. Aan elk regressiemodel wordt een aantal assumpties gekoppeld en getest voor validiteit.

- Lineair in parameters:
De variabelen hangen samen met een lineair verband, uit te drukken met een constante, richtingscoëfficiënten en een residuwaarde.



- ⚡ Willekeurige steekproef:
Er moet sprake zijn van een willekeurige steekproef uit de populatie. Dit wordt geborgd door de onderzoeksopzet, waarbij ook het aantal waarnemingen gecontroleerd wordt.
- ⚡ Geen perfect lineair verband:
Tussen de verklarende variabelen onderling mag geen perfect lineair verband bestaan, zoals $X_1 = 1 - X_2$ en deze variabelen mogen niet constant zijn.
- ⚡ Geen of weinig multicollineariteit:
Tussen de verklarende variabelen onderling mag geen significant verband bestaan. Alle verklarende variabelen worden los als afhankelijke variabele getest in een regressie met de overige verklarende variabelen.
- ⚡ Homoskedasticiteit:
De variatie van residuwaarden in elke regressie moet een normale distributie vertonen, gecontroleerd met Q-Q plot of histogram.

