

Data-Analyse: Regressie Technieken

EXAMEN 2020-1

Vraag 1: je krijgt de output van een normale logistische regressie ($y=1$ student zit op kot; $y=0$ student zit niet op kot)

- Is er een verschil tussen het effect vrouwen en mannen (dummy: vrouw= 1; man = 0)?
=> antw => wald test of betrouwbaarheidsinterval
- bereken het effect van de rest van de Onafh variabelen
=> antw: odds
- log likelihood test uitvoeren
- we krijgen een classificatie tabel gegeven die deels is ingevuld en nog een Aantal gegevens. Vraag: vul de tabel aan en evalueer het totaal percentage correct geclassificeerde cases)
=> Antw in de bijlage staat de N (totaal aantal cases) heb je nodig voor de berekening.

Vraag 2: gegeven is een tabel over de keuze van huishoudens in turkije over de wijze hoe ze hun huis opwarmen

Gegeven: grote tabel met alleen verschillende "groepen " dummy's

- beschrijf waarom het een
- Voor 3 van deze groepen moet je de algemene tendens beschrijven
(=> je moet dus rekeningen houden met het feit dat het een dummy is en met het feit dat het een multinominale regressie is)
- in de output staat ook de IIA (independence of irrelevant alternatives)
=> Hier moet je beschrijven wat deze gegevens precies betekenen, wat het doel is van deze test en wat de nulhypothese is.

Vraag 3: een model

Gegeven: 4 outputs: 1 pooled model, 1 Fixed effects model, 1 random effects model en 1 beursch en pagan test

=> Uit de introductie kan je afleiden dat de gegevens uit de output betrekking hebben op een dif in dif analyse (wordt niet expliciet gezegd!)

- a. De test voor het verschil tussen een Fe- model en een RE-model is niet gegeven in de output. Hij vraagt wat deze test is, hoe hij werkt, nulhypothese en wat we verwachten wat het resultaat zal zijn op basis van de output
Antw. => hausman test volledig uitleggen

Keuzen tussen FE-model en RE-model baseren op de $cov(Xu, VV) = \dots$ => in het FE model staat dat deze niet gelijk is aan 0 + in het RE- model staat dat er wordt aangenomen dat het =0 is => hieruit kan je besluiten dat $cov(u, VV)$ wss verschilt van 0. => FE-model is beter.

- b. Alle andere testen die in het model staan ook beschrijven: nulhypothese geven, waarvoor de test dient, hoe ze werken, En wat we erover kunnen besluiten op basis van de output.
- c. In het FE- model is een variabele weg. Je moet duidelijk uitleggen waarom deze variabele er niet staat
=> Antw. Het is een tijd onafh variabele => worden opgenomen in de individueel specifieke constanten die voor alle cases worden geschat...
- d. Er wordt gevraagd wat de onderzoekers probeerden te meten en welke techniek ze hiervoor gebruikten en hoe deze werkt(dif in de analyse dus). Verder vraagt hij om het resultaat van deze analyse (de difference estimate) te beschrijven in de context van de gegeven output (dus gewoon het interactie-effect in de output interpreteren). Tot slot moet je ook de belangrijke voorwaarde uitleggen (gelijke trends)
- e. tot slot moet je nog een laatste variabele B1 beschrijven, deze variabele is in elke output hetzelfde. We moet beschrijven wat de reden hiervan zou kunnen zijn