

Examen ATSAT Januari 2022

in totaal 18 W/V stellingen, de rest open vragen (voornamelijk berekeningen waar je enkel je einduitkomst moest schrijven op antwoordenblad, geen enkele open vraag waar je iets moet uitleggen of interpreteren, interpretatie van output gevraagd in W/V stellingen)

1) regressie

→ output van regressie model was gegeven gelijkaardig aan salary example met een parameter voor experience en het kwadraat van experience (allebei niet significant) (ln(salary) als afhankelijke variabele, experience, experience², age en education als onafhankelijke variabele)

- a. Vraag over glh. Je moest de correcte matrix notatie geven van L gegeven dat je wilt testen voor 2 parameters tegelijkertijd.

$L = \text{rbind}(c(0,0,1,0,0), c(0,0,0,1,0))$ 10x

- b. Daarna waren het 4 J/F stellingen over de interpretatie van de output

- “Stel we maken een model met enkel experience en age waarbij experience en age daar wel significant zijn. Dit is een goede reden om te zeggen dat in dit model (met experience²) multicollineariteit aanwezig is.” Waar x5 → moet niet noodzakelijk, indien experience heteroscedastisch was konden de p-waarden in het originele model fout zijn (waardoor na verwijdering van experience de p-waarden terug veranderen)
- interpretatie van een parameter (pas op afhankelijke variabele is ln(wage)!) Waar x7
- Stel dat je een model met wage als afhankelijke variabele schat en je wilt dit vergelijken met model in de opgave → methode op slide 9 van econometrie was gegeven (stelling was dus juist x3)
- het loon van iemand met bepaalde waarden voor elke verklarende variabele bedraagt 12.84 (pas op: afhankelijke variabele is ln(wage)) Fout x8 Juist x1

2) Het al dan niet deelnemen aan schaak sessies kunnen de testcores van bepaalde scholen verklaren. Het blijkt dat de scholen die meedoen aan de schaak sessies juist die scholen zijn met de beste scores (model: testcores = alpha +k beta*chess i)

→ 4 W/V stellingen:

- er is sprake van reverse causality hier (fout) x8 (Ik denk waar want hoge testcores (oorzaak) impliceert dat ze waarschijnlijk het schaakprogramma zullen hebben (gevolg), terwijl het hebben van een schaakprogramma (oorzaak), hogere testcores impliceert (gevolg)) (ik denk ook waar, ik ook dus x3?)
- beta blijft unbiased (onvertekend) (Fout x10) juist x1
- beta is inconsistent x7 consistent x6
- beta zal het effect van het deelnemen aan de schaak sessies overschatten (Waar x14) Fout x 1

3) Model waarbij de posterior kansen gegeven werden. Bereken een ontbrekende posteriori kans (0,0707) x6 en maak de classification tabel op basis van deze output

11	1	ik had	11	2	ik had ook een van deze 2 :)
2	7		1	7	

Bereken true positive rate (ik had 1) (5/9; x4) (% had ik) + % goed geklasseerd (17/21) x6 met cutoff value =0.7

W/V stellingen:

- als de cut-off stijgt, zal de sensitiviteit ofwel constant blijven ofwel stijgen maar zeker niet dalen (Fout x12) Juist x2
- observatie 4 en 5 vormen een concordant pair(waar x8) niet waar x2

4) Logistieke regressie output gegeven (heel gelijkaardig aan voorbeeld met accident/noaccident): nieuwe auto's die in dat jaar zijn uitgekomen zijn ofwel diesel ofwel benzine auto's → bepalen adhv cost, lengte, fiscal rate of een auto op benzine of diesel rijdt

- Geef de r code om de gegeven output te krijgen (was een multiple logistic model dus glm gebruiken) x6
- interpretatie van de output met W/V stellingen (een waarbij de odds ratio van een bepaalde variabele was gegeven) die odds ratio was juist x6 (-> stond 11,82 ofzo, maar ik kwam 11,83 uit, WAAR aangeduid, maar erbij geschreven)
- stel er is ook de mogelijkheid van elektrische auto's (hierdoor zijn er 3 response waarden in plaats van 2) welk model zou je nu schatten?
-> Discrete choice or conditional logit model x11
-> Multinomiaal logit model x5: denk deze want ging over characteristics van uw auto's (hp, lengte,...) ipv over characteristics van uw outcome (soorten benzine, anders zouden ze kost van uw benzine, diesel etc moeten geven ipv van de auto)-> idem

5) dit was een oefening op PCA. Er waren 3 eigenwaarden gegeven en dan nog de pattern loadings ξ_1 en ξ_2 na rotatie van X1-X6.

- a. bereken de verklarende variantie van ξ_1 voor rotatie 1.168 x4
- b. bereken de geïmpliceerde correlatie van X5 en X6 (1.168 dacht ik, huh? haha) (0.8455)
- c. bereken de communaliteit van X1 na rotatie (c = d) iets met 0,7 ofzo?
- d. bereken de communaliteit van X1 voor rotatie (c = d gee' verschil voor en na rotatie x7)

6) oefening op AFT model: tijd nodig tot iemand geneest van covid met als verklarende variabele gender, exercise (of je al dan niet sportief bent of niet, meer dan 3u per week sport of niet), age

- a. 2 juist of fout stellingen

- i. een reden om censored te zijn zou kunnen zijn dat de persoon nog niet genezen is op het einde van de studie of dat de persoon sterft tijdens de studie (Waar x11)
- ii. de tijd tot genezing is ongeveer 6 dagen korter bij iemand die veel sport vergeleken met iemand die niet veel sport (0.85 weken korter => fout, gedaald met 15% en niet met 85%) x6
- ~~b. vraag welk model dit zou zijn als het een logistische regressie model was~~
- c. heel hard de oefening van slide 41-43 volgen van AFT
 - i. aanduiden op grafiek waar $S_{\text{male}=0}(t)$, $S_{\text{male}=1}(t)$ en waar $S_{\text{age}=30}(t)$ en $S_{\text{age}=40}(t)$ zijn
→ idem slide 43
 - ii. $t_{0.50}$ berekenen voor een specifieke observatie (gegeven dat de residuals lognormaal verdeeld zijn) (6,12.. weken x7) (4,946 weken)
 - iii. $t_{0.50}$ aanduiden op grafiek van survival time (op dezelfde grafiek als i.)

7) cluster analyse

- a. output voor WALD gegeven en je moest nog een SSE berekenen (144) x10
- b. je moest de ordening van de cluster analyse geven (gebaseerd op de methode van Ward) (1-3-12-15) x6
echt man ik had hier gwn alle laagste SSE's opgeschreven met haar 100 lijntjes pfeh x5

8) Tukey over pauzes (variabele breaks = 0,1 of 2, variabele shift = day of night)

- 1: de variabele breaks is een stochastische regressor: False x6, True x1
- 2: op basis van de output van de tukey intervals kunnen we zeggen dat de dagshift 1 pauze nodig hebben (niet meer, niet minder), True x5, False x1
- 3: H_0 hypothese: True 2x, False x1