



Logistisk regresjon med mer enn to kategorier

MEDISIN OG TALL

STIAN LYDERSEN

stian.lydersen@ntnu.no

Stian Lydersen er dr.ing. og professor i medisinsk statistikk ved Regionalt kunnskapssenter for barn og unge – psykisk helse og barnevern (RKBU Midt-Norge) ved Institutt for psykisk helse, NTNU. Forfatteren har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

Når utfallsvariabelen har bare to nivå eller kategorier, kan man anvende vanlig binær logistisk regresjon. Når den har tre eller flere, kan man bruke andre varianter av logistisk regresjon.

La oss starte med et eksempel: Munthe-Kaas og medarbeidere (1) studerte en mulig sammenheng mellom skrøpeligheit (*frailty*) før hjerneslag og grad av kognitiv svikt tre måneder etter slag hos 598 pasienter. Den uavhengige variabelen, skrøpeligheit, ble målt med en indeks som i teorien går fra 0 til 1, der høyere verdi betyr mer skrøpeligheit. Pasientene hadde verdier fra 0 til 0,56 før slaget, med gjennomsnitt 0,14 og standardavvik 0,10. Den avhengige variabelen, grad av kognitiv svikt, er ordinal med tre kategorier: Tre måneder etter slaget hadde 268 (45 %) normal kognisjon, 172 (29 %) mild kognitiv svikt og 158 (26 %) demens.

Separate binære regresjonsmodeller

Når den avhengige variabelen er ordinal, som her, kan det være aktuelt å dikotomisere den og bruke binær logistisk regresjon. Det finnes to mulige skiller for dikotomisering av en variabel med tre kategorier (figur 1). Dersom vi legger skillet mellom demens og mild kognitiv svikt, får vi kategorien demens på den ene siden og den sammenslåtte kategorien mild svikt eller normal kognisjon på den andre. Den tilhørende oddsratioen per 0,10 økning i skrøpeligheit blir 3,09 (95 %-konfidensintervall (KI) 2,45 til 3,89, $p < 0,001$). Dersom vi legger skillet mellom normal kognisjon og mild kognitiv svikt, blir oddsratioen 2,29 (KI 1,83 til 2,87, $p < 0,001$).

	Separate binære		Ordinal		Multinomisk	
Demens		OR ₂ = 3,09		OR ₃ = 2,70		OR ₅ = 3,70
Mild kognitiv svikt	OR ₁ = 2,29		OR ₃ = 2,70		OR ₄ = 1,47	
Normal kognisjon						

Figur 1 Alternative logistiske regresjonsmodeller med oddsratio (OR) per 0,10 enhet økning i skrøpeligheidsindeks.

Ordinal logistisk regresjon

Alternativt kan man bruke ordinal logistisk regresjon. Det finnes flere versjoner (2), og den mest brukte kalles gjerne for proporsjonal odds-logistisk regresjon. Da gjør man som ovenfor en analyse for hvert mulige skille mellom kategoriene i den avhengige variabelen, men nå med den forenklete antakelsen at oddsratioen er den samme for hvert skille. I vårt eksempel får vi da en oddsratio på 2,70 (KI 2,23 til 3,27, $p < 0,001$).

Antakelsen om at oddsratioen er den samme på hvert skille, kalles proporsjonal odds-antakelsen. En hypotesetest for denne antakelsen i vårt eksempel gir en p -verdi på 0,014. Dette tyder på at antakelsen ikke er helt oppfylt. Ordinal logistisk regresjon kan likevel være egnet til å undersøke et forskningsspørsmål (3, s. 315–6). En hypotesetest om denne antakelsen kan dermed ha begrenset relevans. Munthe-Kaas og medarbeidere (1) valgte separate binære regresjonsmodeller i sin studie, ikke bare basert på denne hypotesetesten, men særlig for å undersøke om skrøpelighet var en sterkere prediktor for skillet mellom demens og mild kognitiv svikt enn for skillet mellom mild kognitiv svikt og normal kognisjon.

Multinomisk logistisk regresjon

Hvis kategoriene i den avhengige variabelen ikke er ordinale, kan man bruke multinomisk logistisk regresjon. Man velger da en referansekategori. I vårt eksempel er det naturlig å velge normal kognisjon som referanse. Vi får da en oddsratio for mild kognitiv svikt på 1,47 (KI 1,12 til 1,92, $p = 0,06$) og en oddsratio for demens på 3,70 (KI 2,81 til 4,87, $p < 0,001$), begge i forhold til normal kognisjon. Merk at oddsratioen for mild kognitiv svikt på 1,47 her blir lavere enn oddsratioen i binær logistisk regresjon på 2,29. Dette skyldes at i den binære logistiske regresjonen ser man på oddsratioen for kategoriene mild kognitiv svikt og demens samlet, mens man i multinomisk logistisk regresjon ser på mild kognitiv svikt alene. Tilsvarende ser vi at oddsratioen på 3,70 i multinomisk logistisk regresjon gjelder mot normal kognisjon, mens i binær logistisk regresjon er oddsratioen 3,09 lavere, da den er mot mild kognitiv svikt og normal kognisjon samlet.

Valg av modell

Hvis den avhengige variabelen ikke er ordinal, vil en multinomisk logistisk modell være mest aktuell. Hvis den avhengige variabelen er ordinal, kan man velge ordinal logistisk regresjon hvis proporsjonal odds-antakelsen gir en god approksimasjon til data, eller hvis det ikke er av praktisk interesse å skille mellom effekt på de forskjellige nivåene i den avhengige variabelen. En av fordelene med ordinal logistisk regresjon i forhold til separate binære regresjonsmodeller er at det blir færre ukjente størrelser, her oddsratioer, i modellen. Dette gir en betydelig forenkling av modellen, særlig hvis det er flere enn tre kategorier i den avhengige variabelen, eller hvis det er flere prediktorer.

De tre alternative regresjonsmodellene som er beskrevet her, har samme fortolkning av oddsratio som i binær logistisk regresjon: Oddsratioen er den relative økningen i odds for en enhets økning i prediktoren.

REFERENCES

1. Munthe-Kaas R, Aam S, Saltvedt I et al. Is Frailty Index a better predictor than pre-stroke modified Rankin Scale for neurocognitive outcomes 3-months post-stroke? *BMC Geriatr* 2022; 22: 139. [PubMed] [CrossRef]
2. Hosmer DW, Lemeshow S, Sturdivant RX. *Applied logistic regression*. 3. utg. Hoboken, NJ: Wiley, 2013.
3. Harrell FE. *Regression modeling strategies. With applications to linear models, logistic and ordinal regression, and survival analysis*. 2. utg. Cham: Springer, 2015.

Publisert: 23. juni 2022. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.21.0786

© Tidsskrift for Den norske legeförening 2023. Lastet ned fra tidsskriftet.no 20. januar 2023.