

Et kritisk blikk på regresjonsmodeller

Regresjonsmodeller med mange variabler er mye brukt i medisinsk forskning. En metode for å vurdere modellen er å lage den for en tilfeldig gruppe av pasientene i datasettet og deretter prøve den ut på de andre – såkalt kryssvalidering. Metoden er enkel og nyttig, men brukes for lite innen medisinsk forskning.

■ Utsagnet «alle modeller er gale, men noen er nyttige» gjelder i høyeste grad for regresjonsanalyser på medisinske datasett. Bruken av statistiske programmer og ønsket om mer avanserte statistiske analyser gjør det fristende å lage store regresjonsmodeller med mange forklaringsvariabler. Et fundamentalt spørsmål blir likevel i hvor stor grad modellen forklarer det som observeres. Hvis den i liten grad forklarer de faktiske data, er nytteverdien begrenset. Da hjelper det lite at modellen statistisk-matematisk ikke er gal.

Metoden

Kryssvalidering er standardmetodikk innen fagfelter som biologi og kjemi samt ved industriell bruk av regresjonsmodeller. Prinsippet er enkelt og intuitivt – regresjonsmodellen estimeres for en gruppe

«Sammenhengen mellom hva den statistiske modellen finner og de faktiske, observerte resultatene, blir altså for god ved vanlig regresjonsanalyse uten validering»

pasienter og prøves deretter ut på en annen gruppe. Evnen til å predikere utfallet for pasientene vurderes dermed på et ubrukt datasett. I sin enkleste form deles data-materialet tilfeldig i to grupper.

Kryssvalidering generaliserer dette prinsippet. En mindre del av pasientene utelates fra datasettet og tester modellen som er laget på de resterende. Denne prosedyren gjentas flere ganger inntil alle pasienter har predikert utfall fra regresjonsmodeller der de ikke selv utgjør datagrunnlaget. Deretter analyseres sammenhengen mellom hva regresjonsmodellen predikerte og hva som faktisk ble observert. Korrelasjonen mellom de predikerte og de faktisk observerte verdier kan brukes som et mål på modellenes nytteverdi. Oppdelingen av gruppene kan gjøres på forskjellig vis. En klassisk tilnærming er å ta ut én pasient av datasettet, estimere regresjonsmodellen basert på de resterende pasientene og så predikere utfallet for denne ene utelatte pasienten. Dette gjentas på nytt inntil alle pasientene er blitt utelatt en gang. Metoden kalles utelat-en-om-gangen-kryssvalidering og er integrert i flere statistikkprogrammer (1).

Nytten av kryssvalidering

En regresjonsmodell bør gi økt forståelse og bidra til bedre medisinsk behandling. Da bør den være tolkbar for forskeren og noenlunde tilpasset de observerte data. Et grunnleggende fenomen for regresjonsmodeller, spesielt de med mange forklaringsvariabler, er en altfor «optimistisk» forklaring av dataene. Modellen tilpasser de spesifikke data i utvalget, men det er ikke dermed sagt den fungerer godt på andre data. En statistisk modell bør kunne brukes på andre og fremtidige data. Sammenhengen mellom hva den statistiske modellen finner og de faktiske, observerte resultatene, blir altså for god ved vanlig regresjonsanalyse uten validering. Regresjonsanalysen kan dermed være statistisk riktig, men lite klinisk nyttig (2).

En god statistisk modell bør være forståelig for dem som skal bruke den. Det

modellen på de resterende. Dette blir mye arbeid med mange modeller og grupper av pasienter å holde orden på. Dessverre er ikke kryssvalidering av regresjonsmodeller tilstrekkelig integrert i vanlige statistikkprogrammer som SPSS eller STATA. Programpakken R har funksjoner for kryssvalidering og kan lastes ned gratis fra Internett, men brukerterskelen er relativt høy og forutsetter skrevne kommandoer. Det finnes heldigvis enkelte brukervennlige program-pakker for regresjon med kryssvalidering. Et eksempel er Unscrambler, som er utviklet i Norge og brukes på verdensbasis innen industriell anvendelse av regresjonsmodeller på felter som kjemi, biologi, oljeindustri og matvareproduksjon (4). Unscrambler har også innebygd regresjonsmetoder som passer til situasjoner med flere variabler enn observasjoner, noe som ikke er sjelden for medisinske datasett.

Kryssvalidering bør først og fremst brukes ved utvikling av prognostiske modeller – dette gjelder enten man bruker lineær, logistisk eller Cox' regresjon. Den gir en validering av den statistiske analysen, men er ikke tilstrekkelig ved utvikling av prognostiske modeller til aktivt bruk i klinisk arbeid. Da er det nødvendig med en ekstern validering der modellen testes ut på et helt nytt datasett fra andre pasienter (5). Enhet for biostatistikk og epidemiologi ved Oslo universitetssykehus kan hjelpe forskere som ønsker å vite mer om kryssvalidering og utvikling av prognostiske modeller.

Are Hugo Pripp

are.hugo.pripp@oslo-universitetssykehus.no
Enhet for biostatistikk og epidemiologi
Oslo universitetssykehus
Postboks 4956 Nydalen 0424 Oslo

Oppgitte interessekonflikter: Ingen

Litteratur

1. Browne MW. Cross-validation methods. *J Math Psycho* 1999; 44: 108–32.
2. Altman DG, Royston, P. What do we mean by validating a prognostic model? *Statist Med* 2000; 19: 453–73.
3. Hjort NL, Claeskens G. Model selection and model averaging. *Cambridge series in statistical and probabilistic mathematics*. Cambridge: Cambridge University Press, 2008.
4. Camo. www.camo.com [22.11.2010].
5. Steyerberg EW. Validation of prediction models. I: Steyerberg EW. *Clinical prediction models. A practical approach to development, validation, and updating*. New York, NY: Springer, 2009: 299–311.

Mottatt 22.11. 2010, første revisjon innsendt 1.12. 2010, godkjent 9.12. 2010. Medisinsk redaktør Siri Lunde.