
Kryssvalidering – å analysere dataene på kryss og tvers

MEDISIN OG TALL

ARE HUGO PRIPP

apripp@ous-hf.no

Are Hugo Pripp er forsker og biostatistiker ved Oslo senter for biostatistikk og epidemiologi, Forskningsstøtteavdelingen, Oslo universitetssykehus. Han er professor II ved Fakultet for helsevitenskap, OsloMet – storbyuniversitetet.

Forfatteren har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

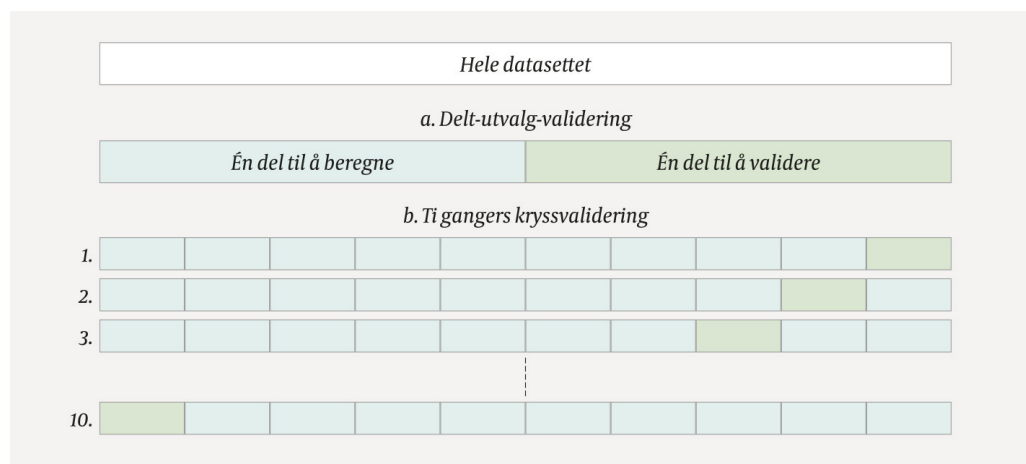
Ofte lønner det seg å undersøke ting på kryss og tvers, spesielt hvis noe synes å være for godt til å være sant. Kryssvalidering vurderer den statistiske modellen ved å prøve den på nytt i det samme datasettet.

Statistiske metoder skal presist analysere dataene i din studie. Vi bruker de samme dataene både til å beregne resultater og vurdere hvor god analysen er. Da kan den statistiske modellen bli for godt tilpasset akkurat våre data og dermed fungere dårlig i en gjentatt studie. For å være helt sikker burde vi gjentatt den samme studien mange ganger og repetert beregningene på nye data fra disse. Heldigvis finnes det statistiske knep som kan spare oss alt dette arbeidet. Vi kan bruke de samme dataene gjentatte ganger på en smart måte for å undersøke repeterbarheten (1). I jo større grad noe kan repeteres, desto mer kan vi stole på resultatet.

Kryssvalidering

Hvis vi vil bruke det samme datasettet til både en analyse og en vurdering av repeterbarheten, er en enkel løsning å dele dataene i to: den ene delen til den statistiske analysen og den andre delen til å undersøke om resultatene lar seg repetere med nye data. Dette er delt-utvalg-validering (*split-sample validation*)

(figur 1a). Metoden er enkel, men har sine ulemper. Vi får bare brukt halvparten av dataene til de statistiske beregningene. Hvordan dataene deles i to, kan lett påvirke resultatene, spesielt i små utvalg. En bedre strategi kan være å bruke mesteparten av dataene til den statistiske beregningen, og deretter prøve den ut på en gjenværende rest av «ubrukte» data. Dette kan vi så gjenta flere ganger, og hver gang ta ut en ny del av datasettet for å validere den statistiske analysen med «ubrukte» data. F.eks. kan vi dele utvalget i ti like store deler og bruke ni av delene til beregningen og den siste delen til å prøve den ut. Gjentar vi dette ti ganger, tar ut en ny del for hver gang og oppsummerer resultatet fra de ti repetisjonene, har vi utført ti gangers kryssvalidering (*10-fold cross-validation*). Kryssvalidering er en videreutvikling av delt-utvalg-validering, og gir mer stabile resultater (figur 1b). Den mest «ekstreme» varianten er ta-ut-én-kryssvalidering (*leave-one-out cross-validation*). Da tar vi ut én observasjon og utfører den statistiske beregningen på de resterende dataene. Deretter prøver vi ut beregningen på den ene utelatte observasjonen og gjentar det hele inntil alle observasjonene er tatt ut én gang (2).



Figur 1 a) Delt-utvalg-validering deler dataene i to – én del til beregningen og én til valideringen av den statistiske analysen. b) Ti gangers kryssvalidering deler dataene i ti – ni deler til beregningen og én til valideringen – og repeterer dette ti ganger med en ny del tatt ut til valideringen for hver gang.

Når bør vi bruke kryssvalidering?

Kryssvalidering, eller andre metoder for validering, er spesielt viktig for statistiske modeller om prognosen eller diagnosen til en pasient, såkalte prognostiske eller diagnostiske modeller. Framingham-risikoskåren, som estimerer tiårsrisiko for hjerte- og karsykdom til en person basert på alder, kjønn og ulike risikofaktorer, er et velkjent eksempel på en prognostisk modell (3). Den beste statistiske modellen er ikke nødvendigvis den som passer best til dataene som er brukt til å beregne modellen, men den modellen som fungerer best for andre eller fremtidige pasienter. Ved å utføre kryssvalidering undersøker vi repeterbarheten og om modellen fungerer for andre pasienter enn de i vårt datasett. Kryssvalidering eller andre metoder for intern validering, f.eks. bootstrapping (4), bør være et minimum (5). Noen ganger er resultatene for gode til å være sanne. Da er kryssvalidering spesielt viktig.

Intern validering er ikke ekstern validering

Selv om kryssvalidering er viktig for å evaluere repeterbarheten til en statistisk analyse, har den og andre metoder for intern validering sine begrensninger. I små utvalg gir den usikre resultater, som i mindre grad lar seg generalisere. Dessuten er det viktig med en viss størrelse på utvalget som skal validere modellen. Derfor er ta-ut-én-kryssvalidering ikke alltid anbefalt, da det kan gi et for optimistisk inntrykk av repeterbarheten (6). Generelt sett kan ikke kryssvalidering gjøre et lite utvalg stort, ei heller gjøre et ikke-representativt utvalg representativt. Enhver statistisk analyse gir informasjon om dataene, gitt deres begrensninger (7). En ekstern validering gjennom at andre reproducerer studien, er alltid den beste kvalitetssikringen.

LITTERATUR

1. Pripp AH. Et kritisk blikk på regresjonsmodeller. Tidsskr Nor Legeforen 2011; 131: 449. [PubMed][CrossRef]
2. James G, Witten D, Hastie T et al. Resampling methods. An introduction to statistical learning: with applications in R. New York, NY: Springer, 2013: 175–201.
3. Wilson PW, D'Agostino RB, Levy D et al. Prediction of coronary heart disease using risk factor categories. Circulation 1998; 97: 1837–47. [PubMed][CrossRef]
4. Skovlund E. Bootstrapping – å løfte seg selv etter håret? Tidsskr Nor Legeforen 2019; 139. doi: 10.4045/tidsskr.19.0413. [PubMed][CrossRef]
5. Steyerberg EW. Validation of prediction models. Clinical prediction models: A practical approach to development, validation, and updating. New York, NY: Springer International Publishing, 2019: 329–44.
6. Little MA, Varoquaux G, Saeb S et al. Using and understanding cross-validation strategies. Perspectives on Saeb et al. Gigascience 2017; 6: 1–6. [PubMed][CrossRef]
7. Morin K, Davis JL. Cross-validation: What is it and how is it used in regression? Commun Stat Theory Methods 2017; 46: 5238–51. [CrossRef]

Publisert: 28. september 2020. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.20.0154
Opphavsrett: © Tidsskriftet 2026 Lastet ned fra tidsskriftet.no 16. februar 2026.