



Intraklassekorrelasjon

MEDISIN OG TALL

STIAN LYDERSEN

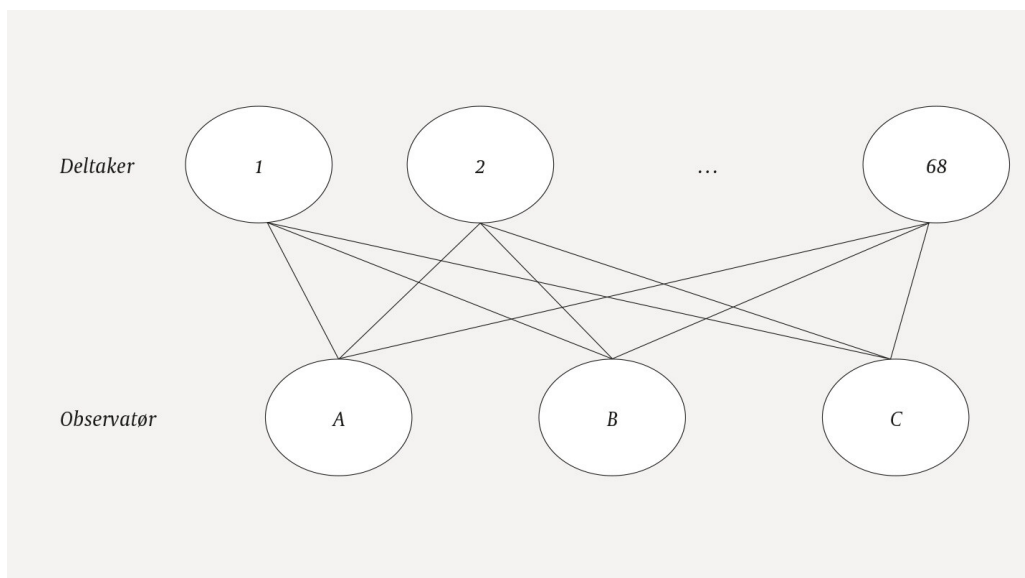
stian.lydersen@ntnu.no

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.

En viktig egenskap ved en målemetode er at den er reliabel, det vil si at den gir godt samsvar mellom gjentatte målinger på samme objekt. Intraklassekorrelasjonskoeffisienten er et mål på graden av samsvar mellom målinger på en kontinuerlig eller ordinal skala.

Hvis man skal klassifisere pasienter i kategorier, for eksempel etter hvorvidt pasienten har en bestemt diagnose eller ikke, vil Cohens kappa eller spesifikt samsvar være et relevant mål på samsvar (1, 2). Hvis pasientene derimot måles på en skala som er kontinuerlig eller ordinal, vil intraklassekorrelasjonskoeffisienten (*intraclass correlation coefficient*, ICC) være et relevant mål.

La oss ta utgangspunkt i et eksempel. Når man beveger den ene hånden, kan det oppstå ufrivillige bevegelser i den andre hånden. Slike speilbevegelser er vanlige hos små barn, og vil oftest opphøre før tiårsalderen. Hos barn med unilateral cerebral parese vil slike bevegelser ofte være mer tydelige, og dessuten bestå når barna blir eldre. Woods og Teuber-skalaen kan brukes til å kvantifisere graden av slike speilbevegelser. Skalaen går fra 0 til et teoretisk maksimum på 24, der en høy skår betyr høy grad av speilbevegelser. Magne og medarbeidere studerte grad av samsvar mellom observasjoner for denne skalaen, basert på videoopptak av 68 barn som ble skåret av tre observatører (3). Studiedesignet er illustrert i figur 1.



Figur 1 Studiedesign for beregning av samsvar mellom observasjoner i eksempelet.

To tilfeldige effekter

Her tenker vi oss at videoopptakene av de 68 deltakerne representerer et tilfeldig utvalg fra en relevant populasjon. Enkelte deltakere vil vise høy grad av spillbevegelser og dermed få høye skårer, mens andre vil vise lav grad og få lavere skårer. Dette kan uttrykkes ved en varians mellom deltakerne. Videre kan det tenkes at noen observatører har en tendens til å gi systematisk høyere skårer enn andre observatører, hvilket kan uttrykkes ved en varians mellom observatørene. Og til slutt vil det være en variasjon som ikke kan forklares ut fra varians mellom deltakere eller mellom observatører, altså en residualvariens. Dette kalles en toveis tilfeldig-effekt-modell (*two-way random effect model*). Det er nemlig to tilfeldige effekter, uttrykt ved henholdsvis varians mellom deltakere og varians mellom observatører. Nærmere bestemt kalles dette en modell med kryssede faktorer, fordi én og samme observatør kan skåre flere deltakere (se figur 1).

Intraklassekorrelasjonskoeffisienten

I den aktuelle studien var gjennomsnittsskåren 7,30. Variansen mellom deltakerne var 33,44, variansen mellom observatørene 0,74 og residualvariansen 2,93. Totalvariansen er lik summen av disse, altså 37,11. Disse varianskomponentene kan brukes til å beregne forskjellige mål på samsvar mellom observasjoner ved hjelp av intraklassekorrelasjonskoeffisienten.

Korrelasjonen mellom skårer fra to observatører på samme deltaker er lik varians mellom deltakerne delt på totalvariansen, altså $33,44 / 37,11 = 0,90$ (4, s. 121). Denne intraklassekorrelasjonskoeffisienten er et mål på interobservatørsamsvar (*inter-rater reliability*). Korrelasjonen mellom to skårer fra samme observatør på samme deltaker, derimot, er lik summen av varians mellom deltakerne og varians mellom observatørene, delt på totalvariansen, altså $(33,44 + 0,74) / 37,11 = 0,92$ (4, s. 129). Denne intraklassekorrelasjonskoeffisienten er et mål på intraobservatørsamsvar (*intra-rater reliability*), altså hvis observatøren skårer samme deltaker på ny etter å ha glemt hvilken skår som ble gitt første gang.

Hvordan skal vi vurdere størrelsen på intraklassekorrelasjonskoeffisienten? Det finnes ingen generell standard for dette. Intraklassekorrelasjonskoeffisienten er en relativ størrelse og vil være større jo mer inhomogent utvalget er. Et forslag fra Koo og Li kan

imidlertid tjene som en retningslinje (5): $< 0,50$, $0,50-0,75$, $0,75-0,90$ og $0,90-1$ regnes som henholdsvis dårlig, moderat, godt og svært godt. Korrelasjonene på $0,90$ og $0,92$ i denne studien viser et godt eller svært godt samsvar.

Det finnes alternative beregningsmåter og definisjoner av intraklassekorrelasjonskoeffisienten. I de ovennevnte eksemplene ble det brukt en toveis tilfeldig-effekt-modell. En oversikt over alternativer finnes i Veierød og medarbeidere (6, s. 487). Notasjonen for alternativene er ikke entydig i litteraturen, så det er viktig å skrive i klartekst hvilken metode man bruker. I mange tilfeller vil en toveis tilfeldig-effekt-modell være egnet.

REFERENCES

1. Lydersen S. Cohens kappa – et mål på samsvar mellom observatører. Tidsskr Nor Legeforen 2018; 138: 467.
2. Lydersen S. Positivt og negativt samsvar. Tidsskr Nor Legeforen 2018; 138: 562.
3. Magne VA, Adde L, Hoare B et al. Assessment of mirror movements in children and adolescents with unilateral cerebral palsy: reliability of the Woods and Teuber scale. Dev Med Child Neurol 2021; 63: 7. [PubMed][CrossRef]
4. Gwet KL. Handbook of inter-rater reliability volume 2: analysis of quantitative ratings. 5. utg. Gaithersburg, MD: AgreeStat Analytics, 2021.
5. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016; 15: 155–63. [PubMed][CrossRef]
6. Veierød M, Lydersen S, Laake P. Medical statistics in clinical and epidemiological research. Oslo: Gyldendal akademisk, 2012.

Publisert: 18. juli 2022. Tidsskr Nor Legeforen. DOI: [10.4045/tidsskr.22.0041](https://doi.org/10.4045/tidsskr.22.0041)

© Tidsskrift for Den norske legeforening 2022. Lastet ned fra tidsskriftet.no 11. august 2022.