



Når bør man velge en ikke-parametrisk metode?

MEDISIN OG TALL

EVA SKOVLUND

E-post: eva.skovlund@ntnu.no

Eva Skovlund (f. 1959) er professor i medisinsk statistikk ved Institutt for samfunnsmedisin, Norges teknisk-naturvitenskapelige universitet, og seniorforsker ved Nasjonalt folkehelseinstitutt. Forfatter har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

Når data ikke er normalfordelt, vil mange velge å gjøre statistisk analyse ved hjelp av en ikke-parametrisk metode. Selv om ikke-parametriske metoder er fordelingsfrie og robuste mot ekstremverdier, er det likevel ikke alltid en god idé å benytte disse i stedet for tradisjonelle parametriske metoder.

De fleste statistiske analyser er basert på bruk av parametriske metoder. T-test for sammenligning av gjennomsnitt og tradisjonell estimering av 95% konfidensintervall for gjennomsnitt er eksempler på dette.

En viktig forutsetning for at en t-test skal gi et riktig resultat er at observasjonene er uavhengige og tilnærmet normalfordelte, men forutsetningen om normalfordeling trenger ikke å være eksakt oppfylt. Gjennomsnittet av uavhengige observasjoner tenderer nemlig mot å være normalfordelt selv om fordelingen til enkeltobservasjonene ikke er det. Hvis vi har mange observasjoner, vil gjennomsnittet være tilnærmet normalfordelt selv om observasjonene stammer fra en svært skjev fordeling. Ofte vil 30 observasjoner være tilstrekkelig til at valg av en parametriske metode er uproblematisk.

Små utvalg og ikke-parametriske tester

I praksis kjenner vi ikke fordelingen til populasjonen som utvalget av data kommer fra, og vi må nøye oss med den informasjonen som ligger i observasjonene vi har.

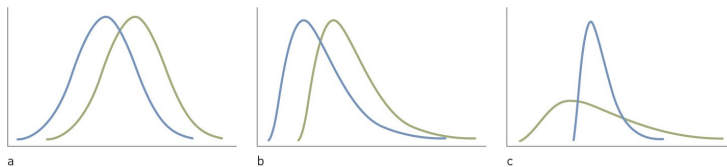
Et histogram kan gi en pekepinn om formen på fordelingen. Hvis den ser svært skjev ut (hvis det er én eller flere ekstreme verdier), eller hvis vi har så få observasjoner at det er vanskelig å si noe om formen på fordelingen, er det et alternativ å benytte en såkalt ikke-parametrisk metode. Eksempler på ikke-parametriske metoder som ofte benyttes i analyse av medisinske data, er Wilcoxon-Mann-Whitney-test og Wilcoxons test for paradata.

Ikke-parametriske metoder karakteriseres ved at man i stedet for å benytte de observerte verdiene direkte, benytter ordningsrekkefølgen til observasjonene og gir dem rangtall fra 1 (for laveste verdi) og oppover. Da vil ekstreme observasjoner ikke lenger bli like ekstreme, og det er ikke lenger noen forutsetninger knyttet til formen på fordelingen til selve observasjonene.

Ikke-parametriske metoder er såkalt fordelingsfrie, så hvorfor benytter vi ikke like gjerne alltid slike metoder i stedet for de parametriske? En årsak er at når observasjonene faktisk stammer fra en normalfordeling, vil en ikke-parametrisk test ha noe lavere teststyrke enn en parametrisk test, og vi vil trenge noen flere observasjoner for å avdekke sanne effekter eller sammenhenger.

Krav om lik spredning

En minst like viktig begrunnelse er at når vi ønsker å sammenligne to utvalg, krever en ikke-parametrisk metode at de to fordelingene som skal sammenlignes, har samme spredning og form (fig 1a og b). Denne antagelsen gjelder også for en t-test når vi sammenligner gjennomsnittet i to uavhengige grupper, men t-testen er noe mindre sårbar for forskjellig spredning enn en ikke-parametrisk test, i alle fall når de to utvalgene er omtrent like store (1). Den tradisjonelle t-testen tåler relativt godt avvik i størrelsesorden inntil to-tre ganger større standardavvik i den ene gruppen enn den andre. Mange statistikkpakker har for øvrig implementert en variant av t-testen (Welch test) som tar hensyn til at spredningen kan være ulik i de to gruppene. Vi kan altså relativt lett håndtere brudd på forutsetningen om lik spredning når vi benytter parametriske tester.



Figur 1 Ulike fordelinger: a) To normalfordelinger med forskjellig gjennomsnitt, men samme spredning, b) To skjeve fordelinger med forskjellig gjennomsnitt, men samme spredning og c) To skjeve fordelinger med samme gjennomsnitt, men forskjellig spredning

Transformasjon

Men hva hvis vi har få observasjoner og de to fordelingene begge er skjeve, men synes å ha forskjellig spredning som i figur 1c? Da vil verken t-test eller Wilcoxon-Mann-Whitney-test gi riktig p-verdi. Den mest nærliggende løsningen er å transformere data slik at fordelingen blir tilnærmet normal. For eksempel vil en logaritmetransformasjon ofte kunne gjøre en skjev fordeling med noen få høye verdier mer normalfordelt, og vi kan benytte en alminnelig t-test dersom kravet om rimelig lik spredning er oppfylt.

Er signifikanstest nødvendig?

Et enda viktigere spørsmål enn hvilken test vi bør velge, er kanskje om vi trenger en p-verdi i det hele tatt. Det gir vesentlig mer informasjon å estimere størrelsen av en effekt eller forskjell mellom to grupper med et tilhørende 95% konfidensintervall. Valg av metode følger samme retningslinjer som for signifikanstester: Med mindre antall observasjoner er veldig lite og datasettet inneholder enkelte ekstreme verdier, klarer vi oss stort sett fint med en tradisjonell parametrisk metode.

LITTERATUR:

1. Skovlund E, Fenstad GU. Should we always choose a nonparametric test when comparing two apparently nonnormal distributions? J Clin Epidemiol 2001; 54: 86 - 92.

Publisert: 4. september 2017. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.17.0219

© Tidsskrift for Den norske legeforening 2020. Lastet ned fra tidsskriftet.no