

Hvilken t-test er best?

MEDISIN OG TALL

STIAN LYDERSEN

E-post: stian.lydersen@ntnu.no

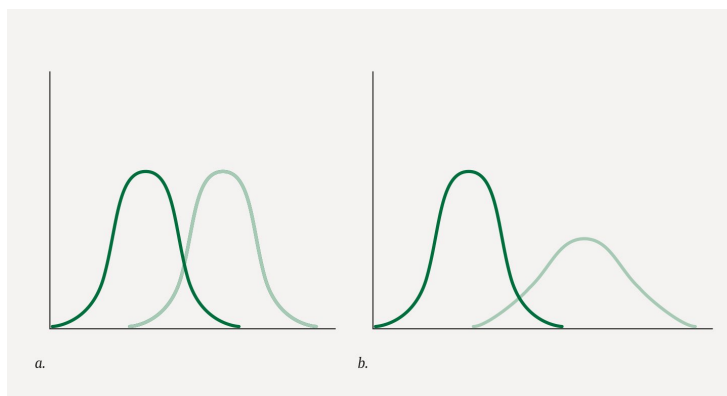
Stian Lydersen er dr.ing. og professor i medisinsk statistikk ved Regionalt kunnskapssenter for barn og unge – psykisk helse og barnevern (RKBU Midt-Norge) ved Institutt for psykisk helse, NTNU. Forfatteren har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

MORTEN WANG FAGERLAND

Morten Wang Fagerland er ph.d. og leder for Seksjon for biostatistikk og epidemiologi ved Oslo senter for biostatistikk og epidemiologi, Oslo universitetssykehus. Forfatteren har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

Det finnes to versjoner av t-testen for sammenlikning mellom to grupper. Students t-test er mest brukt, og den forutsetter at den underliggende variansen er lik i de to gruppene. En alternativ versjon, kalt Welchs test, forutsetter ikke lik varians. Hvordan skal vi velge den mest egnede versjonen?

Anta at du vil sammenlikne gjennomsnittene i to grupper, og at dataene er tilnærmet normalfordelt, slik at du kan bruke en t-test. Dersom du kan anta at den underliggende variansen er lik i de to gruppene, som illustrert i figur 1a, er den vanlige Students t-test å anbefale. Men dersom den underliggende variansen er ulik, som illustrert i figur 1b, vil ikke Students t-test holde signifikansnivået. Det vil si at sannsynligheten for type I-feil, altså å feilaktig konkludere at gjennomsnittene er forskjellige, med en slik prosedyre vil kunne bli vesentlig høyere enn det valgte signifikansnivået, som ofte settes til 5 %. Det finnes en egen versjon av t-testen, Welchs U-test, heretter bare kalt Welchs test, som er korrekt å bruke dersom variansen er ulik (1). Konfidensintervall for differansen mellom to gjennomsnitt finnes også i to versjoner basert på henholdsvis Students t-test og Welchs test. Problemet er at vi vanligvis ikke kjenner den underliggende «teoretiske» variansen. Hvordan skal vi i praksis velge den mest egnede versjonen?



Figur 1 a) To normalfordelinger med lik varians. b) To normalfordelinger med ulik varians.

Lik eller ulik varians?

En mulig fremgangsmåte kunne være å først teste en hypotese om at variansen er lik i de to gruppene. Levenes test og Fishers F-test er eksempler på slike tester. Deretter, dersom man forkaster hypotesen om lik varians, bruker man Welchs test – ellers bruker man Students t-test. En slik totrinns testprosedyre er ikke å anbefale, selv om den fremdeles omtales i enkelte lærebøker. Et vesentlig problem er at prosedyren ikke holder signifikansnivået (2).

Noen ganger kan man ha apriorisk kunnskap som tilsier at den underliggende variansen er lik i de to gruppene. Dette kan for eksempel være tilfelle når variansen skyldes måleusikkerhet ved den målemetoden som er brukt. Da er dette et godt grunnlag for å velge Students t-test. Det kan være fristende å anta at den underliggende variansen er lik dersom estimert varians ikke er veldig ulik, men ulempen er at estimatet kan avvike en del fra underliggende varians hvis utvalgene er små. Enkelte lærebøker, for eksempel Navidis *Statistics for Engineers and Scientists* (3), anbefaler å bruke Students t-test bare hvis man har apriorisk kunnskap som underbygger antakelsen om lik varians. Der påpekes det at dersom vi bruker Welchs test selv om den underliggende variansen er lik, så får vi tilnærmet samme svar som ved Students t-test. På den annen side vil vi risikere å få feil svar hvis vi bruker Students t-test når den underliggende variansen er ulik. En annen og viktig fordel med Welchs test er at den er mer robust enn Students t-test mot avvik fra antakelsen om normalfordelte data, som gjør at den kan anvendes i langt flere situasjoner (4).

Students t-test er forholdsvis robust mot ulik varians, med mindre det ene utvalget er vesentlig større enn det andre (5, 6). Enkelte forskere anbefaler derfor Students t-test når det observerte standardavviket i gruppen med størst standardavvik ikke er større enn for eksempel dobbelt så stort som standardavviket i gruppen med minst standardavvik (5), mens andre er skeptiske til å basere valget på observerte standardavvik (3).

Har Students t-test noen fordeler sammenliknet med Welchs test? Slik vi ser det, har den ikke så mange, men Students t-test har litt høyere teststyrke enn Welchs test ved lik utvalgsstørrelse. I motsetning til Welchs test er den også mulig å beregne for hånd, noe som gjør den velegnet til bruk i undervisning.

Konklusjon

Hvilken versjon av t-testen bør man velge i praksis? Hvis man har apriorisk kunnskap om lik varians, bør man velge Students t-test. Et trygt og enkelt svar kunne være å ellers alltid bruke Welchs test, men ulempen med dette er at Students t-test kan ha høyere teststyrke, spesielt i små utvalg og ved lik utvalgsstørrelse. En ting som er helt klart, er at det frarådes å basere valget på en totrinns testprosedyre der man starter med en hypotestetest for variansene.

LITTERATUR:

1. Welch BL. The significance of the difference between two means when the population variances are unequal. *Biometrika* 1938; 29: 13. [CrossRef]
2. Zimmerman DW. A note on preliminary tests of equality of variances. *Br J Math Stat Psychol* 2004; 57: 173–81. [PubMed][CrossRef]
3. Navidi W. *Statistics for engineers and scientists*. 5 utg. McGraw-Hill Education, 2020.
4. Fagerland MW, Sandvik L. Performance of five two-sample location tests for skewed distributions with unequal variances. *Contemp Clin Trials* 2009; 30: 490–6. [PubMed][CrossRef]
5. Skovlund E, Fenstad GU. Should we always choose a nonparametric test when comparing two apparently nonnormal distributions? *J Clin Epidemiol* 2001; 54: 86–92. [PubMed][CrossRef]

Publisert: 23. november 2020. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.20.0750
© Tidsskrift for Den norske legeforening 2020. Lastet ned fra tidsskriftet.no