

Statistisk signifikans og p-verdi

Stein Atle Lie, statistiker, Institutt for klinisk odontologi, Det medisinske fakultet, Universitetet i Bergen, stein.lie@uib.no

I vitenskapelige artikler ser man ofte utsagn som «Forskjellen var signifikant». Ofte vil et slikt utsagn etterfølges av et tall – for eksempel slik «Forskjellen var signifikant ($p=0,042$)». Hva menes med dette? Og hva betyr tallet?

La oss si at det er gjennomført en studie der det sammenlignes lommedybde mellom røykere og ikke-røykere. La oss så si at gjennomsnittlig lommedybde er 3,8 mm hos røykerne, mens ikke-røykere har et gjennomsnitt på 3,3 mm. Altså har røykere 0,5 mm dypere lommer enn ikke-røykere. Klinisk kan man tenke at røyking er en faktor som har betydning for lommedybde og det er derfor fristende å si: «Ja, røykere har dypere lommer enn ikke-røykere».

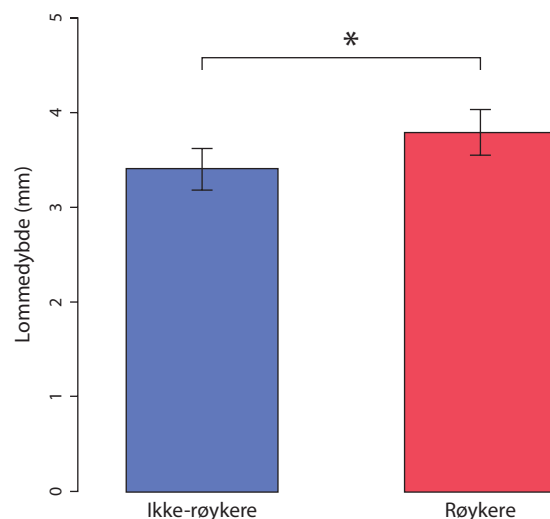
Så enkelt er det ikke. Om vi skal tro at den observerte forskjellen er en sann forskjell avhenger av flere faktorer. Både antall personer i studien, størrelsen på forskjellen vi observerer og hvor mye lommedybde varierer mellom de ulike personene har betydning. Statistikeren R.A. Fisher introduserte for omtrent 100 år siden det han kalte p-verdi (probability value) som et standardisert mål der alle disse momentene er tatt hensyn til. I tillegg er det viktig at deltakerne i utvalgene er tilfeldig plukket ut slik at andre faktorer som henger sammen med lommedybde (for eksempel alder), er likt fordelt mellom de to gruppene.

P-verdiene betyr ikke «sant/usant»

P-verdien er ikke et mål for om det er en sann/faktisk forskjell mellom to grupper, men er en standardisert størrelse for sjansen (sannsynligheten) for at forskjellen var mulig selv om gruppene egentlig er like. Løselig kan man si at om de to verdiene «egentlig» er like, så uttrykker p-verdien sjansen for at denne antagelsen er feil.

Uttrykket «Forskjellen var statistisk signifikant» betyr altså ikke at det er en sann forskjell. Det vil vi aldri kunne vite! Men, hvis p-verdien er liten vil man ha større grunn til å tro at antagelsen om at gruppene er like ikke er riktig. Ordet «signifikant» kan erstattes med «av betydning», men i forskning viser «statistisk signifikant» til p-verdien.

I odontologi og medisin sier man gjerne at når p-verdien er mindre enn 0,05 (5 %) da er p-verdien «liten» og forskjellen vi har observert er statistisk signifikant. Det er en utbredt misforståelse å tolke at forskjellen er sann hvis p-verdien er mindre enn denne satte grensen. Det er også like feil å si at gruppene er like når p-verdien er større enn 0,05. Beregningen av p-verdien er en matematisk standardisering der den observerte forskjellen, variasjon og antall observasjoner er tatt hensyn til. Begrepet man bruker sammen med p-verdi bør være «statistisk signifikant», ikke bare «signifikant». Når man så har vurdert om en forskjell er statistisk signifikant vil det være en medisinsk/odontologisk faglig vurdering om forskjellen har klinisk betydning.



En illustrasjon basert på reelle grunnlagsdata om lommedybde hos et utvalg røykere og ikke-røykere (Bunæs DF& al. Site-specific treatment outcome in smokers following non-surgical and surgical periodontal therapy. J Clin Periodontol 2015; 42: 933-42.). Søylen er gjennomsnitt med spredningsmål (standardfeil, standard error of the mean). Forskjellen mellom gruppene er statistisk signifikant og markeres ofte med én stjerne når p-verdien er mindre enn 0,05 (5%). I dette tilfellet var $p=0,028$. Bruken av stjerner er mer kompakt enn å skrive det faktiske tallet og benyttes ofte i figurer.