

Hypotesetesting

Det skal mye til før en statistiker aksepterer en påstand om en sammenheng som en sannhet.

Forfattere

Kjetil Gundro Brurberg
forsker

Hugo Hammer
Førsteamanuensis

Sykepleien Forskning 2013 8(3)(267-269)
DOI: <https://doi.org/10.4220/sykepleienf.2013.0157>

I tidligere artikler har vi sett hvordan vi bruker konfidensintervall til å trekke slutninger om en populasjon basert på tilfeldige utvalg og hvordan konfidensintervall kan være til hjelp når vi sammenlikner grupper (1,2). I denne artikkelen skal vi se hvordan tilfeldige utvalg brukes til å bekrefte eller avkrefte hypoteser om en populasjon ved hjelp av en prosedyre vi kaller hypotesetesting.

Hva er en hypotese?

En hypotese er en påstand om en sammenheng. Hver og en av oss bombarderes med slike påstander hver eneste dag. Dagbladet kan påstå at en bestemt matvare gir økt kreftrisiko. En kollega kan påstå at ettermiddagskaffen bedrer konsentrasjonen, og en medisinprodusent påstå at den har laget en ny medisin som fungerer bedre enn medisinen som vanligvis brukes. Statistikere stiller seg generelt skeptiske til påstander, og det skal mye til før vi aksepterer slike påstander som en sannhet. Før vi konkluderer at medisinprodusentens påstand er rett vil vi være rimelig sikre på at vi ville sett større gjennomsnittlig forbedring hvis alle pasientene i Norge fikk den nye medisinen enn om de fikk den tradisjonelle medisinen. I likhet med konfidensintervall dreier hypotesetesting seg om å bruke utvalg til å trekke slutninger om en større populasjon.

Hvorfor hypotesetest?

Vi skal starte med et eksempel som viser gangen i en statistisk hypotesetest. Tabell 1 oppsummerer utvalgte resultater fra en studie som sammenlikner statiner (fluvastatin) mot placebo (3). Produsenten av dette legemiddelet kan for eksempel påstå at fluvastatin bidrar til større reduksjon i midjemål og totalkolesterol enn placebo. For å teste om produsentens påstand er rett får 42 tilfeldig utvalgte overvektige menn med høyt blodtrykk statiner mens 41 får narremedisin. Resultatene viser at gjennomsnittlig midjemål gikk ned med 1,6 cm blant pasienter som fikk statiner og 2,4 cm blant pasienter som fikk narremedisin. Gjennomsnittlig totalkolesterol endret seg lite hos pasientene som fikk narremedisin, men gikk ned med 0,6 mmol/l ved bruk av statiner. Her er det fristende å raskt konkludere med at den nye medisinen fungerer bedre enn narremedisin hvis hensikten er å redusere totalkolesterolet. Gode statistikere hopper sjeldent raskt til konklusjoner. De vil vite om de ville fått samme resultat hvis begge medisinene ble prøvd ut på alle overvektige menn med høyt blodtrykk i hele landet.

TABELL 1: Eksempel på resultattabell som viser gjennomsnittlig endring i midjemål og totalkolesterol etter ett års behandling med placebo eller fluvastatin. Talleksempelene er basert på data fra en norsk randomisert kontrollert studie (3).

Gjennomsnittlig endring fra start

Utfall	Statin (n=42)	Placebo (n=41)	Forskjell (95% KI)	P-verdi
Midjemål [cm]	-1,6	-2,4	0,8 (-1,8; 0,3)	0,16
Totalkolesterol [mmol/L]	-0,6	0,0	-0,6 (-0,4; -0,9)	<0,001

Hvilke svar gir en hypotesetest?

Resultatet av en hypotesetest er en p-verdi – et tall som alltid ligger mellom 0 og 1. Tabell 1 gjengir to p-verdier som er resultat av to ulike hypotesetester. Den første hypotesen forfatterne tester er om midjemålet endres mer ved statinbruk enn ved bruk av narremedisin, og vi ser at den aktuelle p-verdien er 0,16. Hvis vi velger å tro på resultatet, her vil det si å konkludere med at narremedisin gir noe større reduksjon i midjemål enn statiner, er det 16 prosent sjanse for at vi trekker en feil konklusjon. Den andre hypotesen som testes i tabell 1 er om kolesterolnivået endres mer ved statinbruk enn ved bruk av narremedisin, og da er p-verdien mindre enn 0,001. Hvis vi velger å tro på resultatet, her vil det si å konkludere med at statiner gir større kolesterolreduksjon enn narremedisin, så er det svært lite sannsynlig.

P-verdier og terskler

Mange lurer kanskje på hvorfor nettopp 0,05 ansees som en magisk grense der p-verdier som er mindre enn 0,05 får tommel opp mens p-verdier som er større enn 0,05 får tommel ned. De skarpe skillelinjene har historiske røtter. Hypotesetestingen ble utviklet på en tid der beregninger måtte gjøres for hånd, og for å spare tid og krefter var det mer hensiktsmessig å hente terskelverdier fra tabellverk enn å regne ut eksakte p-verdier. I dag har vi datamaskiner som beregner eksakte p-verdier, men mange velger likevel å holde fast ved bruken av terskelverdier. Det er imidlertid viktig å være klar over at det ikke er noe magisk ved slike terskelverdier. Når p-verdien er 0,049 sier vi at hypotesetesten leder til et statistisk signifikant funn, mens en p-verdi på 0,051 ikke er statistisk signifikant. I disse to tilfellene er sannsynligheten for å trekke feil konklusjon hvis vi velger å bekrefte hypotesen, henholdsvis 4,9 prosent og 5,1 prosent. Enten sjansen for å ta feil er 4,9 prosent eller 5,1 prosent gjør man lurt i å unngå skråsikre konklusjoner til fordel for videre undersøkelser.

Signifikans

Bekreftede hypoteser og statistisk signifikante forskjeller forteller ikke uten videre om et tiltak bør iverksettes. Generelt er det slik at studier som omfatter mange pasienter står bedre rustet til å finne statistisk signifikante forskjeller mellom grupper. Enkelte studier er imidlertid så godt utrustet at de klarer å påvise forskjeller som er så små at de er helt uten praktisk betydning. Det er derfor naturlig at man i en statistisk analyse ikke bare vurderer statistisk signifikans, men også relevansen til resultatet. Sagt på en annen måte: Når du bruker forskning i klinisk praksis er det viktig å veie statistisk signifikans og enkeltstående p-verdier opp mot klinisk skjønn. Resultatene i tabell 1 viser for eksempel at gjennomsnittspasienten kan vente at behandling med fluvastatin reduserer totalkolesterolet med 0,6 mmol/l. Hvorvidt dette er en endring som monner – om denne reduksjonen i totalkolesterol veier opp for kostnader og andre ulemper ved å ta medisinen – må helsepersonell og pasient vurdere i fellesskap. Hvis fordelene ved en bestemt behandling er stor nok til å bety en forskjell for den enkelte pasient sier vi gjerne at effekten er klinisk signifikant.

Referanser

1. Brurberg KG, Hammer HL. Hvorfor trenger vi statistikk? Sykepleien Forskning. 2013;8:78–82.
2. Brurberg KG, Hammer HL. Hvorfor sammenlikne statistisk? Sykepleien Forskning. 2013;8:174–7.
3. Hjelstuen A, Anderssen SA, Holme I, Seljeflot I, Klemsdal TO. Effect of lifestyle and/or statin treatment on soluble markers of aterosklerose in hypertensives. Scandinavian Cardiovascular Journal. 2007;41:313-20.