



Pearsons eller Spearmans korrelasjonskoeffisienter

MEDISIN OG TALL

ARE HUGO PRIPP

E-post: apripp@ous-hf.no

Are Hugo Pripp (f. 1971) er forsker og biostatistiker ved Oslo senter for biostatistikk og epidemiologi, Forskningsstøtteavdelingen, Oslo universitetssykehus, og professor II ved Fakultet for helsefag, OsloMet – storbyuniversitetet.

Forfatter har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

Korrelasjonskoeffisienter er blant de mest kjente og brukte statistiske verktøyene, men valg av egnet korrelasjonskoeffisient og tolkningen kan by på utfordringer.

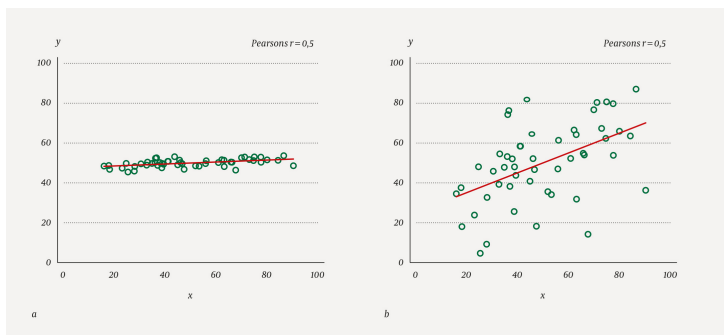
En påfallende likhet mellom medisinske syndromer og statistiske estimatorer er at de gjerne får navn etter personen som oppdaget dem, så også for korrelasjonskoeffisienter. Karl Pearson (1857–1936) og Charles Spearman (1863–1945) var to ledende forskere i statistisk metodikk, foruten betydelige vitenskapelige bidragsytere i biologi, arvelære og psykologi (1). Pearsons og Spearmans korrelasjonskoeffisienter er antageligvis de to mest brukte korrelasjonskoeffisientene i medisinsk forskning, men hva er forskjellen og likheten mellom dem?

Pearsons korrelasjonskoeffisient

Korrelasjonskoeffisienter tar verdier på en enhetsfri skala fra -1 til $+1$ som uttrykk for henholdsvis negativ og positiv korrelasjon, der verdien 0 angir ingen korrelasjon.

Pearsons korrelasjonskoeffisient måler styrken av den lineære sammenhengen mellom to variabler. Hvis man plotter variablenes verdier og tilpasser en rett linje mellom punktene, angir korrelasjonskoeffisienten hvor nært punktene er til linjen. Hvis alle punktene ligger på en rett linje, blir korrelasjonen «perfekt» med en korrelasjonskoeffisient lik enten -1 eller $+1$. Hvis det ikke er en lineær sammenheng, blir korrelasjonskoeffisienten lik 0 . Dette estimatet er viktig, nyttig og mye brukt, men gir likevel begrenset statistisk informasjon.

Korrelasjonskoeffisienten forteller oss ikke den nøyaktige tallmessige lineære sammenhengen, for eksempel stigningsgraden til den rette linjen mellom punktene (2). Dette er illustrert i figur 1. La oss tenke oss 50 observasjoner av to variabler, x og y , som kan ta verdier fra 0 til 100 . Korrelasjonene i plottene a og b er begge nøyaktig $0,5$ ($p < 0,001$), men stigningstallet og den tallmessige sammenhengen er forskjellige. En korrelasjonskoeffisient alene gjør det ikke mulig å predikere en forventet verdi til den ene variabelen basert på en observert verdi fra den andre. Da er det nødvendig med en regresjonsanalyse.



Figur 1 Plottene i a) og b) viser at to datasett med ulik tallmessig sammenheng kan gi lik korrelasjon

En signifikantest av korrelasjonskoeffisienten er, hvis ikke annet er spesifisert, en test av om korrelasjonen er lik 0. En signifikant korrelasjonskoeffisient er ingen garanti for en relevant assosiasjon. Dette henger sammen med antall observasjoner. I et datasett med 100 observasjoner trenger verdien til en signifikant korrelasjon på 0,05-nivå ikke å være større enn 0,2. Ved 500 observasjoner er signifikante korrelasjonskoeffisienter ikke nødvendigvis større enn 0,09. Da forklarer den lineære sammenhengen mellom de to variablene mindre enn 1 % av variasjonen i dataene.

For mest valid statistisk anvendelse, tolkning og signifikantesting bør de parvise observasjonene av to variabler være kontinuerlige, ha en lineær sammenheng, ikke ha data med betydelig avvik fra sammenhengen (såkalte uteliggere eller «outliers») og følge en bivariat normalfordeling.

Spearman's (rang)korrelasjonskoeffisient

Spearman's korrelasjonskoeffisient, ofte forkortet til Spearman's rho, er statistisk relatert til Pearson's korrelasjonskoeffisient, men basert på verdiene til den relative rangeringen av observasjonene og ikke de observerte verdiene. Dette medfører at Spearman's korrelasjonskoeffisient er spesielt godt egnet for analyse av ordinale variabler, ikke-normalfordelte variabler og/eller variabler med en eller flere avvikende data (uteliggere). Det er ingen krav om en lineær sammenheng mellom de faktisk observerte dataene, så lenge rangeringsverdiene gir en lineær sammenheng (3). For å undersøke om Spearman's korrelasjonskoeffisient er egnet, er det enkelt å regne ut rangeringsverdiene ved hjelp av statistiske programvarer.

Kritisk vurdering

Man bør være varsom med å «screene» etter korrelasjoner, noe som er godt illustrert i en populærvitenskapelig bok om «falske» korrelasjoner (4). Et generelt råd er at en korrelasjonsanalyse begynner med deskriptiv statistikk og avsluttes med en regresjonsanalyse. Til slutt er det viktig kritisk å vurdere om korrelasjonen gir mening.

LITTERATUR:

1. Rodgers JL, Nicewander WA. 13 ways to look at the correlation-coefficient. Am Stat 1988; 42: 59 - 66. [CrossRef]
 2. Sedgwick P. Pearson's correlation coefficient. BMJ 2012; 344: 2.
 3. Sedgwick P. Spearman's rank correlation coefficient. BMJ 2014; 349: g7327. [PubMed][CrossRef]
 4. Vigen T. Spurious correlations. New York, NY: Hachette Books, 2015.
-