

Baggrundsnotat: Korrelationer

Korrelation er en statistisk beregning, som kan vise om, og hvor meget variable er relaterede. For eksempel er vægt og højde relateret – højere personer er ofte tungere end lavere personer. Forholdet er ikke perfekt. Personer med samme højde kan have forskellig vægt, og du kan formentlig hurtigt komme i tanker om nogen, du kender, hvor den laveste vejer mere end den højeste. Men ikke desto mindre er der en relation mellem højde og vægt, når der ses på gennemsnit. Korrelationen kan vise dig, hvor meget vægten afhænger af højden.

Udregn forskellige korrelationer i The Survey System

Der findes flere forskellige korrelationsberegninger. Statistikmodulet i The Survey System indeholder den mest almindelige beregningsform kaldet Pearson korrelation. Modulet indeholder også en variation af Pearson korrelationen, der hedder partiel korrelation. Sidstnævnte er nyttig, når du ser på forholdet mellem to variable, mens du fjerner effekten af en eller to andre variable.

Kun bestemte typer data kan korrelere

Som alle andre statistiske beregninger kan korrelation kun bruges til bestemte former for data. Korrelation er god til data, hvor tallene er udtryk for meningsfulde cifre, sædvanligvis mængder af forskellig slags. Korrelation kan ikke bruges til data med en kategorisk værdi, f.eks. køn, produktmærker eller yndlingsfarve. Skalaer er kontroversielle tilfælde. Tallene på en skala har betydning, men betydningen er ikke særlig præcis. Det er ikke ligesom med mængder eller antal. Ved antal (f.eks. kroner) er forskellen mellem 1 og 2 præcis det samme som forskellen mellem 2 og 3. På en skala er det ikke tilfældet. Du kan være sikker på at respondenterne mener at 2 er imellem 1 og 3, men du kan ikke være sikker på at de mener at 2 er præcis halvvejs mellem 1 og 3. Det gælder især, hvis du bruger betegnelser på din skala (du kan ikke antage at ”godt” er præcis halvvejs mellem ”perfekt” og ”nogenlunde”).

De fleste statistikere mener ikke, at man kan bruge korrelation til skalaer, fordi matematikken bag beregningen antager at forskellen mellem tallene er præcis lige stor. Ikke desto mindre er der mange forskere som bruger korrelationsberegning til skalaer, fordi resultatet som regel afspejler virkeligheden. Vi mener, at du kan bruge korrelationsberegninger til skalaer, men du skal gøre det med forsigtighed og kun lade korrelation være en indikator for en evt. sammenhæng. Hvis du arbejder med mængder og reelle tal vil korrelation give et meget mere sikkert tal.

Korrelationsanalysens resultater

Hovedresultatet af korrelationsberegningen kaldes korrelationskoefficienten (eller "r"). Den varierer fra -1 til $+1$. Jo nærmere r er til $+1$ eller -1 , jo tættere er sammenhængen mellem de to variable. Hvis r er tæt på 0, betyder det at der ikke er nogen sammenhæng mellem variablene. Hvis r er positiv betyder det, at hvis en variabel bliver større, så bliver den anden variabel det også. Hvis r er negativ, betyder det, at hvis en variabel bliver større, så bliver den anden mindre (kaldes også omvendt korrelation).

Korrelationskoefficienten i anden potens

Korrelationskoefficienter gengives normalt som $r =$ (en værdi mellem -1 og $+1$), men det gør det nemmere at forstå hvis de opløftes i anden potens. Kvadratet af koefficienten (eller r^2) er lig med den procent af variation i en variabel som er relateret til variation i den anden variabel. Efter at have opløftet r i anden potens, skal du ignorere decimaltegnet. Et r på .5 betyder at 25% af variationen er relateret ($.5 \times .5 = .25$). Et r på .7 betyder at 49% af variationen er relateret ($.7 \times .7 = .49$).

Signifikans

En korrelationsrapport kan også vise yderligere et resultat – statistisk signifikans. I dette tilfælde siger signifikansniveauet noget om, hvor sandsynligt det er, at korrelationerne skyldes tilfældigheder i form af fejl i stikprøveudvælgelsen. Hvis du arbejder med små stikprøvestørrelser skal du vælge et rapport format, der inkluderer signifikansniveauet. Formatet viser også stikprøvestørrelsen.

Pas på antagelser!

Når du arbejder med korrelationer er det vigtigt aldrig at antage at en ændring i en variabel medfører en ændring i en anden variabel. Salget af computere og løbesko er steget meget i de sidste par år og der er en høj korrelation mellem dem, men du kan ikke antage at købet af computere medfører at folk køber løbesko eller omvendt.

Pas på ikke-liniære forhold

En anden advarsel er, at Pearson korrelationsberegningen egner sig bedst til lineære forhold; hvis en variabel bliver større, så bliver den anden også større (eller mindre) direkte proportionalt. Beregningen arbejder ikke så godt med forhold, der ikke er lineære. Et ikke-lineært forhold er for eksempel forholdet mellem alder og behovet for pleje. De to variable relaterer til hinanden – men ikke lineært. Babyer og ældre mennesker har brug for mere pleje end f.eks. teenagere. Multipel regression (også inkluderet i statistik modulet) kan bruges til at analysere ikke-lineære forhold.