Med statistiske verktøy kan vi estimere sammenhenger mellom variabler. For eksempel kan en t-test brukes for å si noe om forskjellen mellom to grupper, vi kan se på dette som en mulig sammenheng mellom gruppe og en utfallsvariabel. På en lignende måte kan en inndeling i flere grupper ha en sammenheng med en utfallsvariabel, noe som kan måles ved hjelp av en ANOVA-modell. En mer generell statistisk modell er regresjonsmodellen. Ved bruk av en regresjonsmodell kan vi måle sammenhengen mellom en eller flere uavhengige variabler og en avhengig variabel. Statistikken er full av modeller og metoder som måler sammenhenger.

Men hvordan kan vi tolke sammenhenger? I vitenskapshistorien finnes flere eksempler på hvordan man argumenterer for at vi ikke kan si noe om en kausalsammenheng basert på en statistisk sammenheng. «Correlation does not imply causation» er en vanlig frase som brukes for å avferda sammenhenger som betydningsløse. Vi kan finne støtte for en slik forståelse, hver morgen (i hvert fall på en hverdag, og i hvert fall om høsten) når vekkerklokken din ringer står du opp og lager en kopp kaffe. Når du setter deg ned for å drikke kaffen stiger solen over horisonten. Men, det at du drikker kaffe får ikke solen til å gå opp. Her finnes ingen kausalsammenheng.

Samtidig vet vi at noen ting påvirker andre ting, det finnes kausalsammenhenger. I et forsøk undersøker vi hvordan en ny medisin påvirker blodtrykket. En gruppe forsøksdeltakere blir tilfeldig inndelt i to grupper. Deltakere i den ene gruppen får den nye medisinen og deltakere i den andre gruppen får plasebo. Inndeling av deltakere i gruppene er skjult for deltakere og forskere. I etterkant av forsøket måler vi blodtrykk og kan konstatere at deltakere som fått medisinen har lavere blodtrykk sammenlignet med plasebogruppen. Forsøket inneholder noen komponenter som gjør at vi kan si at det finnes en kausal sammenheng mellom det å ta medisinen og det å få lavere blodtrykk. Hvis vi hadde muligheten å forandre på en variabel for å siden måle utfallet i en annen variabel, under kontrollerte forhold så kan vi si at det finnes (eller ikke) en kausal sammenheng. Det å forandre en variabel (f.eks. konsentrasjonen av et medikament i blodet) kan kalles for å gjøre en intervensjon.

I mange sammenhenger har vi ikke mulighet å forandre variabler ved å lage intervensjoner. Det kan for eksempel finnes etiske eller praktiske årsaker til dette. Men vi har fortsatt mulighet å undersøke kausale sammenhenger. Som i eksemplet med medisin for lavere blodtrykk over så krever det noen antagelser. I et forsøk hvor vi intervenerer kan antagelser inkludere at forsøksdeltakere blir tilfeldig tildelt en gruppe og at forskere ikke vet om hvilken pasient eller forsøksdeltakere som får den aktive behandlingen eller medisinen. Hvis disse antakelsene er sanne så kan vi med stor sikkerhet si at medisinen påvirket blodtrykket. I forskningsstudier hvor vi ikke har mulighet å intervenere har vi fortsatt mulighet å «kontrollere» for variabler som kan påvirke sammenhengen mellom to variabler som vi er interessert i. Vi kan identifisere variabler som vi bør kontrollere for ved hjelp av grafiske modeller. Disse modellene kan også gi innsikt i når vi ikke kan si noe om en kausalsammenheng

I denne modulen skal vi snakke om kausal inferens, hvordan kan vi si noe om hvordan en variabel påvirker en annen variabel.

## Eksperiment og kvasieksperiment

Et eksperiment kan for eksempel brukes til å sammenligne to typer av intervensjoner. Vi sammenligner disse intervensjonene i en definert gruppe. Denne gruppen, eller populasjonen, kan for eksempel være idrettsutøvere, og intervensjonene kan være to forskjellige treningsprogrammer hvor et program er et nytt program (N) og det andre er hva man vanligvis bruker, en slag kontrollprogram (K). Vi ønsker med vårt eksperiment å si noe om effekten av N på gruppen idrettsutøvere.

For å gjøre sammenligningen bestemmer vi en utfallsvariabel