

Statistisk analys

Vid statistisk analys tillämpas specifika statistiska metoder på ett datamaterial. Här finns en rik arsenal av vetenskapliga metoder. En fallgrop är att göra mekaniska bearbetningar med hjälp av olika programvaror, utan tillräcklig kännedom om ämnes- och metodmässiga begränsningar. Området är mer krävande än att framställa beskrivande statistik. Bland annat bör man vid en urvalsundersökning i många fall beakta urvalsdesignen vid den statistiska analysen. Beräkningarna blir mer komplicerade vid analys än när man producerar beskrivande mått.

I det följande presenteras kort några metoder för statistisk analys:

Regressionsanalys

Regressionsanalys är en metod för analys av sambandet mellan en responsvariabel (beroende variabel) y och en eller flera förklarande x -variabler (oberoende variabler). Analysens syfte kan variera. Ibland beskrivs hur sambandet ser ut för att kunna förutsäga värden för den beroende variabeln med hjälp av värden på de oberoende variablerna. Alternativt söks efter en *modell* som så bra som möjligt beskriver ett samband. I analysen kan olika kombinationer av oberoende variabler samt olika matematiska modeller prövas.

Den enklaste regressionsanalysen är enkel och linjär. Med *enkel* menas att den endast har en oberoende variabel, och med *linjär* avses att materialet när det åskådliggörs i ett punktdiagram ansluter väl till en tänkt linje. Om det förekommer fler än en oberoende variabel kallas regressionen *multipl*, och om regressionen inte är linjär kallas den *icke-linjär* eller krökt. Ett slags icke-linjär regression är logistisk regression.

Regressionsanalys är samlingsnamnet för en generell typ av modeller och metoder, med en mångfald tillämpningar i empirisk vetenskap och teknik.

Variansanalys

Variansanalys är en metod för att analysera den totala variationen i ett datamaterial och dela upp den i ett antal komponenter som svarar mot intressanta variationsorsaker. En variansanalysmodell är i grunden en *linjär modell* där ett mätvärde kan skrivas som summan av olika effekter baserade på ett antal faktorer. Om en faktor har stor effekt på resultatet bidrar den till en stor del av variansen, medan en faktor med liten eller ingen effekt alls bidrar med en del som är liten eller av samma storleksordning som den rena slumpfaktorn.

De olika deltermerna kan sammanfattas i en variansanalystabell (ANOVA-tabell).

Klusteranalys

Vid klusteranalys delas objekt in i grupper (kluster) så att de objekt som liknar varandra eller ligger nära varandra med avseende på vissa egenskaper förs till samma grupp. Syftet med klusteranalysen är att bilda grupper av ad-hoc-karaktär, dvs. att klustren bildas för ett visst ändamål på befintliga data utan några föregående teoretiska antaganden.

Faktoranalys

Med faktoranalys förklaras samvariationen för ett antal uppmätta variabler via ett mindre antal bakomliggande tänkta variabler (faktorer). Det är en av flera metoder för multivariat analys av datamaterial. Faktoranalys förekommer främst inom beteende- och samhällsvetenskaperna.

Skillnaden mellan faktoranalys och klusteranalys för variabler är att faktoranalysen ska spegla den bakomliggande (teoretiska) strukturen mellan de ingående variablerna, medan klusteranalysen inte bygger på några bakomliggande antaganden, utan bildar kluster av variabler enbart utifrån hur det befintliga materialet ser ut.

Överlevnadsanalys

Överlevnadsanalys används för att studera olika objekts eller händelsers överlevnad. Ett exempel är att studera den tid det tar för en person som blivit arbetslös att åter få ett arbete. Överlevnadsanalys kan användas både vid s.k. observationsstudier, dvs. studier där man använder data "som ändå finns" (t.ex. i olika register), och vid experimentella studier, t.ex. kliniska prövningar. En vanlig frågeställning är hur vissa förklarande variabler påverkar överlevnadstiden; detta studeras med Cox-regression.

Tidsserieanalys (utöver säsongrensning)

Syftet med tidsserieanalys är att med hjälp av modellering beskriva och om möjligt förklara beroendet av tidsseriens komponenter vid olika tidpunkter. De modeller som man kan använda kan vara parametriska eller icke-parametriska, linjära eller icke-linjära respektive konstanta eller tidsvarierande.

Demografisk analys (demometri)

Demografisk analys är den gren av statistisk teori som ägnas åt metoder inom demografin för datainsamling, kontroll av datakvalitet och korrigerering av bristfälliga data samt, i synnerhet, åt analys av data om människors demografiska beteende.

Demografisk analys handlar om demografiska beteenden såsom barnafödande, dödsfall och migration samt deras koppling till samhällsutvecklingen. Analysen berör också befolkningskomposition, t.ex. åldersstruktur och familjesammansättning.

Ekonometrisk analys

Med ekonometrisk analys studeras empiriskt ekonomiska problem och förhållanden, ofta formulerade i matematisk form.