



STATISTIK

MATEMATIK AB

Vad menas med statistik?

Ursprungligen betecknade statistik den disciplin som ägnade sig åt att beskriva och kartlägga staten och dess invånare. Idag omfattar statistik i stort sett alla samhällets områden.

Med statistik menar man det vetenskapliga ämne som utvecklar och använder metoder för att insamla, bearbeta, beskriva och analysera siffermässiga data. Med hjälp av statistiska uppgifter och metoder kan man skaffa sig så tillförlitlig information att viktiga beslut kan fattas.

I vardagslivet kommer vi ofta i kontakt med statistik, inte minst inom media. Här presenteras ofta olika statistiska undersökningar, men också tabeller inom t.ex. fotboll.

Att ha kunskaper i statistik ger överblick över komplicerade siffersammanhang!

Viktigt att tänka på vid statistiska undersökningar

När man skall göra en statistisk undersökning utförs denna i tre steg. Först samlas allt material in, därefter bearbetas det och slutligen tolkas det. Nedan följer lite tips om vad man skall tänka på under arbetets gång!

Att samla in data (MaB)

När vi gör en statistisk undersökning har vi som målsättning att kartlägga en hel grupp av människor, t.ex. hur många procent av alla sextonåringar i Sverige som är rökare. Naturligtvis är det en omöjlighet att fråga precis varenda sextonåring i landet om de röker. Vi får istället göra ett *urval*, och utifrån dessa personers svar dra slutsatser om hela gruppen.

Hela gruppen som undersöks (i detta exempel alla sextonåringar i Sverige) kallas för en *population* och den grupp som väljs ut för undersökningen kallas *stickprov*. Gruppen som väljs ut skall vara representativ för hela populationen.

Hur väljer man då ut ett stickprov? Det finns tre olika tillvägagångssätt:

- *obundet slumpmässigt urval*, sker ex. genom lottning, alla i populationen har samma sannolikhet att väljas
- *systematiskt urval*, sker ex. genom att man väljer ut var tionde person på en lista, med slumpmässig start
- *stratifierat urval*, görs grupper utifrån det man tror speglar resultatet, t.ex.

könsuppdelning, slumpmässigt val med representativ procentsats ur varje grupp

Hur samlar man in data? De vanligaste sätten att samla in statistisk data är genom *intervjuer* eller *enkäter*. Innan man utför sin statistiska undersökning bör man tänka över vilken av dessa metoder man tror ger ett mest tillförlitligt resultat. Är det känsliga frågor som t.ex. sexualvanor, alkoholmissbruk eller kriminalitet som skall undersökas, finns det risk att de som svara inte gör det helt sanningsenligt. Kanske minskar denna risk om man gör undersökningen med hjälp av enkäter, så att de utvalda är helt anonyma.

Hur formulerar man frågor? Det finns en rad olika faktorer att tänka på då man ska formulera sina frågor till enkäten. Börja alltid med bakgrundsfrågor, ex. kön och ålder. I övriga frågor skiljer man mellan *fasta* svarsalternativ och *öppna*. Fördelen med fasta frågor är att de är lättare att bearbeta, medan fördelen med de öppna är att de ofta ger större möjlighet till utförligare, exaktare och mer spontana svar.

Tips att tänka på när man formulerar frågor:

- undvik att formulera ledande frågor
- se till att frågan endast kan tolkas på ett sätt
- ställ bara en fråga åt gången
- använd graderade svarsalternativ vid frågor om ”hur ofta”, eftersom begreppen ofta, sällan etc. har olika innebörd för olika människor
- ha med minst en öppen fråga, ex. övrigt att tillägga

Att sammanställa och bearbeta data

För att få en första överblick över sitt insamlade material kan det vara lämpligt att göra en *frekvenstabell*.

Detta kan göras stegvis på följande sätt:

1. Sätt upp de svarsalternativen, *variablerna*, som finns.
2. Läs av svaren och gör en markering i tabellen för respektive svar. Fortsätt tills alla observationer markerats.
3. Räkna ihop hur många markeringar varje svarsalternativ har. Detta kallas för *absoluta frekvensen*. Inför en frekvenskolumn.
4. Kontrollera att antalet markeringar är lika med det antal personer som svarat på frågan.
5. Inför ytterligare en kolumn där andelen beräknas i procent. Detta kallas *relativ frekvens*.

För att sedan få ett lättöverskådligt resultat använder man lämplig *diagramtyp*.

Nästa steg är att beräkna vissa statistiska storheter som i ett enda värde sammanfattar någon egenskap hos variabelvärdena. Man skiljer mellan

- *lägesmått* (typvärde, medelvärde, median)
- *spridningsmått* (variationsbredd, kvartilavstånd) (**MaB**)

Typvärdet är det variabelvärde (d.v.s. det svar) som har den största frekvensen, alltså som är vanligast.

Det mest använda lägesmättet är *medelvärde*. När man beräknar medelvärdet beräknar man först summan av alla variabler, och därefter dividerar man med antalet variabler.

$$\text{Medelvärdet} = \frac{\text{summan}}{\text{antalet}}$$

Om man har extremt låga eller höga svarsvärden sänker respektiver höjer de medelvärdet radikalt. Det kan då vara mer passande med ett annat lägesmått nämligen *median*. Medianen delar ett rangordnat material mitt itu. När man beräknar medianen skriver man upp alla svar man erhållit i storleksordning med det minsta först. Talet i mitten är medianen. Har vi ett jämt antal observationer är medianen värdet mitt emellan de två mittersta observationerna.

Precis som man kan mäta läget på olika sätt kan man också mäta spridningen med olika mått. Med *variationsbredd* menar man skillnaden mellan det största och minsta värdet. Detta är ett väldigt enkelt spridningsmått, ett bättre mått får man då man istället delar in datamängden i två delar, och sedan bestämmer medianen i de båda halvorna. Dessa kallas *kvartiler*. Man skiljer mellan nedre kvartil, mellersta kvartil (vanliga medianvärdet) och övre kvartil. Som ett mått på spridningen beräknar man avståndet mellan nedre och övre kvartil, detta kallas *kvartilavstånd*.

Att tolka data

Den viktigaste delen i ett statistiskt arbete är just att tolka resultatet, d.v.s. göra en analys av det man fått fram i undersökningen. Den kanske mest relevanta frågan är ifall resultatet man fått fram genom stickprov verkligen ger den rätta bilden av hela populationen.

Det finns ett antal felkällor man bör beakta:

- urvalsfel (är urvalet tillräckligt omfattande)
- bortfall (påverkar bortfallet resultatet)
- mätfel (felställda frågor)

Ett stort problem vid statistiska undersökningar är att man inte får in svar från alla tillfrågade. Om vi t.ex. skickar ut 100 enkäter och får 85 svar, har vi ett *bortfall* på 15%. Detta bortfall kan vi inte utan vidare ersätta med andra villiga försökspersoner, eftersom detta strider mot kravet på slumpmässigt urval. Bortfallet måste alltid redovisas och om möjligt analyseras. Det kan ju vara så att bortfallet på 15% medför att stickprovet inte längre är representativt för populationen. Om man t.ex. undersöker drogmissbruk, kan det vara så att de 15 som vägrat svara är just de som är narkomaner. Men det kan också vara så att bortfallet är slumpmässigt fördelat. I så fall påverkas inte resultatet.