

Frågor till överlevnadskit för statistik

Det här är ett frågebatteri för att du enklare ska kunna arbeta med [Överlevnadskit för statistik](#) och använda det som ett konkret och praktiskt verktyg. Varje fråga i kitet kan delas upp i flera mindre, mer specifika frågor. Det här är ingen komplett lista över frågor. Du kanske kommer på utmärkta frågor att ställa som inte står med här. I så fall får du gärna höra av dig till charlotte[at]inclusiveinsights.se och berätta. Jag är nyfiken på hur just du arbetar med överlevnadskitet!

Om du vill se exempel på dålig statistik från verkligheten är du välkommen att kika in på mitt konto på [Facebook](#) eller [Instagram](#).

Kan du lita på källan?

Vem är källan? Statistik kommer från många olika källor. Från myndigheter kommer bland annat [Sveriges officiella statistik](#). Företag, ideella organisationer, forskare och enskilda individer tar också fram statistik. Ibland kommer statistiken från en mycket trovärdig källa. Ibland har någon helt enkelt hittat på den. Fundera på hur stor tilltro du bör sätta till den som har tagit fram statistiken.

Kan du anta att källan är opartisk? Den som presenterar statistiken kan ha vinning av att statistiken pekar i en viss riktning. Även en partisk källa kan presentera tillförlitlig och bra statistik, men det är alltid värt att tänka lite extra på vad den som presenterar statistiken tjänar på den bild som visas upp och om det påverkar kvaliteten på statistiken.

Vad säger andra källor? Ofta har flera olika källor tagit fram statistik som rör samma sak. En bra strategi är att titta på vad andra källor säger, till exempel andra forskare, andra politiker och andra organisationer.

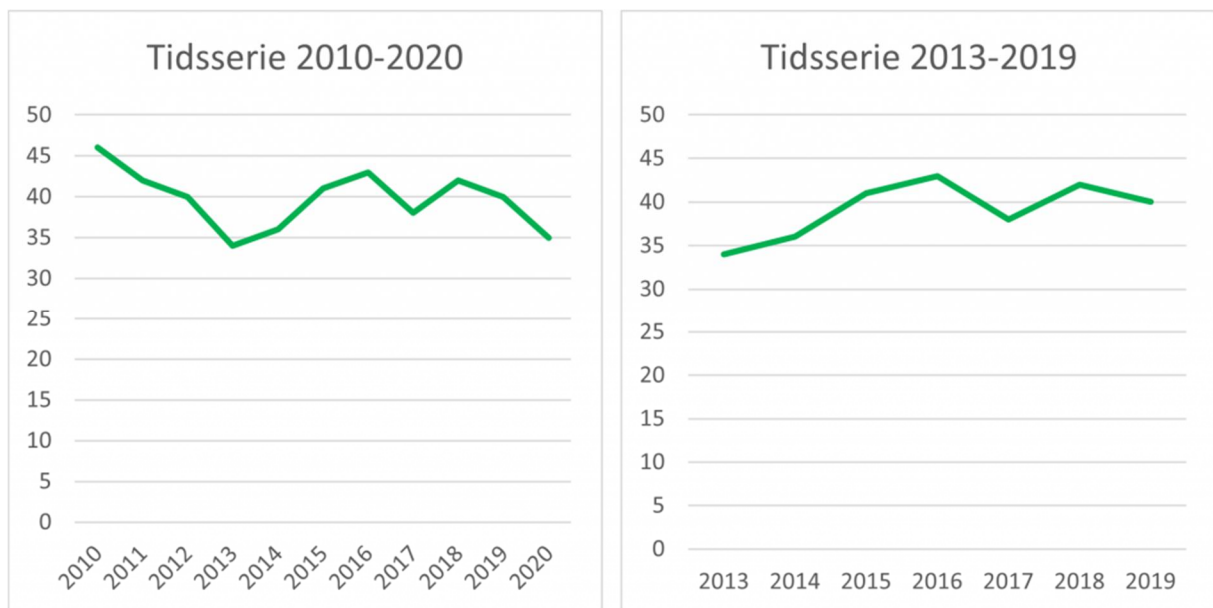
Varför har statistiken tagits fram?

I vilket syfte? Ibland tas statistiken fram för att helt enkelt visa hur något ser ut, till exempel i forskningssyfte eller som underlag för politiska beslut. Ibland finns det andra syften. Till exempel kan en organisation ta fram och använda statistik för att argumentera för sin åsikt. Om den som presenterar statistiken har något att tjäna på att siffrorna ser ut som de gör är det bra att tänka ett varv extra.

Har källan något syfte som inte uttalas? Det är inte alltid den som producerar statistik är öppen med vad syftet är. Till exempel kan något som utger sig för att vara objektiv forskning i själva verket ha som syfte att främja försäljning av en produkt.

Lurar diagrammet ögat?

Varför börjar/slutar kurvan här? Om du tittar på en tidsserie (en kurva som visar hur något har förändrats över tid) bör du fundera på varför kurvan börjar och slutar just där. Ofta har början och slutet valts strategiskt för att få det att se ut som om något har utvecklats på ett visst sätt. Om du till exempel hävdar att det inte förekommer någon global uppvärmning kan du göra en tidsserie över temperaturer och välja ett ovanligt varmt år som början på din kurva. Nedan ser du två diagram som ger lite olika intryck eftersom olika start- och slutår har valts. Diagrammet till vänster ger intrycket av en sjunkande trend: slutåret är lägre än startåret. Diagrammet till höger ger snarare bilden av att trenden rör sig uppåt.

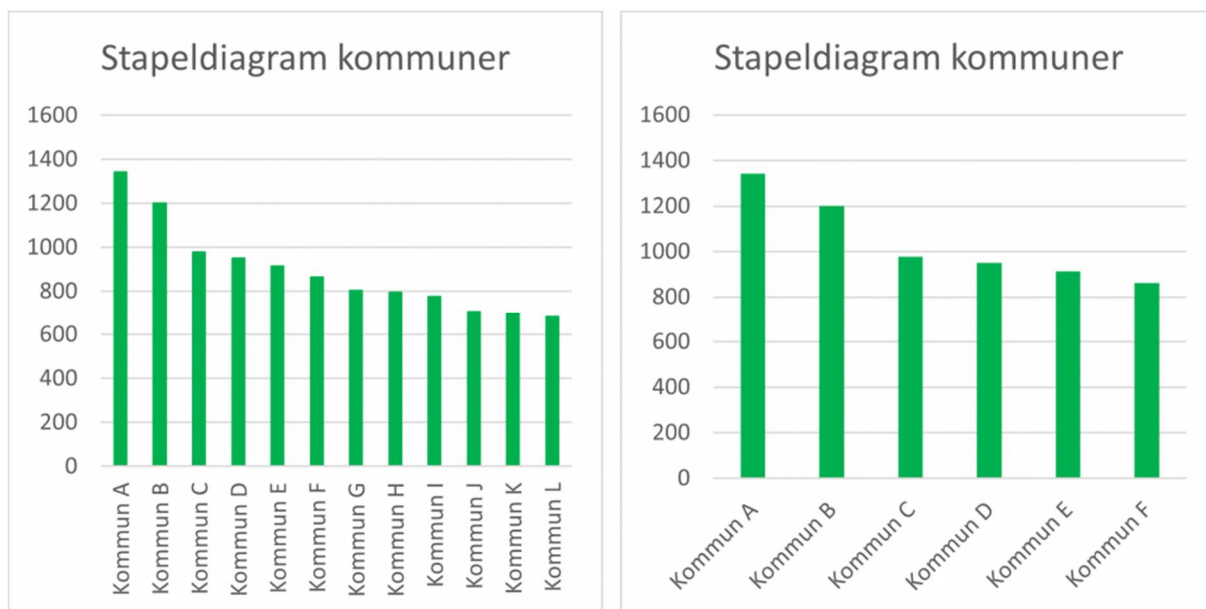


Figur 1. Två linjediagram som visar tidsserier med olika start- och slutår: det vänstra visar 2010-2020 och det högra 2013-2019.

Vad säger siffrorna i diagrammet? Staplar i ett stapeldiagram och bitar i ett cirkeldiagram kan manipuleras och ge ett missvisande intryck. Titta därför inte bara

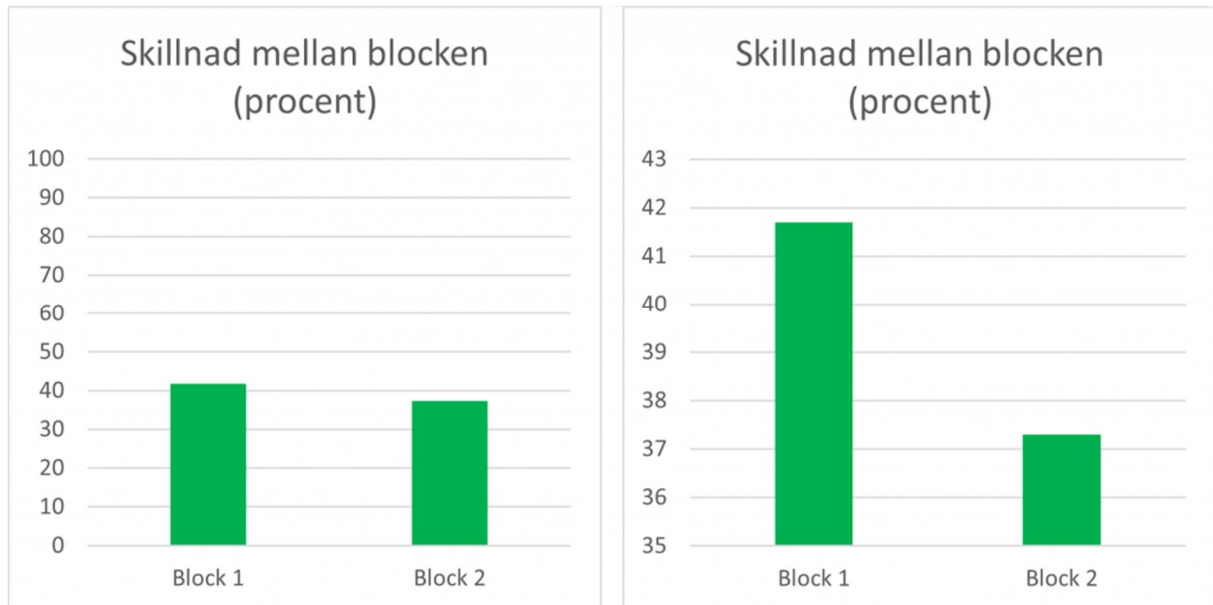
på vad diagrammet ger för intryck, utan uppmärksamma också siffrorna som står i diagrammet.

Finns det en "tappad svans"? Ibland kan delar av ett diagram ha kapats bort för att ge en viss bild. Tänk dig att vi har en region med tolv kommuner. Vi har statistik som visar en viss egenskap i kommunerna. Kommun A är bäst i regionen och har högst stapel. Kommun L är sämst. Du vill av någon anledning framställa kommun F i dålig dager. Det du kan göra då är att kapa bort svansen: alla kommuner som är sämre än F tas bort. Kommun F har gått från att befinna sig i mitten till att framstå som sämst i länet. Diagrammen nedan visar hur detta kan se ut. Givetvis går det att kapa i andra änden också. Om kommun A tas bort framstår genast B som bäst i regionen.



Figur 2. Två stapeldiagram. Det vänstra visar 12 kommuner och det högra 6 kommuner.

Hur ser skalan ut? Är axeln stympad? Genom att stympa en axel på ett diagram går det att ge ett intryck av att en skillnad är stor när den i själva verket är liten. Det går också att göra detta omvänt: genom att ha en skala med stora dimensioner kan stora skillnader göras i stort sett osynliga. Nedan ser du två diagram som visar precis samma siffror men ger väldigt olika intryck. Diagrammen visar skillnaden mellan två politiska block i en opinionsundersökning. Block 1 får mer stöd än Block 2. I diagrammet till vänster syns denna skillnad men det framgår att den är liten. I det högra diagrammet är skalan annorlunda. Block 1 ser nu ut att vara nästan tre gånger så stort som Block 2.



Figur 3. Två stapeldiagram som visar stöd i procent för två politiska block. Diagrammen har olika skalor.

Visar kartan populationsdensitet? Tänk dig att du tittar på en karta som visar hur stort stöd ett parti har i olika delar av ett land. Nästan hela kartan är täckt av partiets färg. Betyder det att partiet i fråga har ett enormt stöd? Nej, det behöver det inte göra. Det beror på hur tätbefolkade de markerade delarna är. Partiet kanske främst har stöd i glesbygden. Det intressanta i det här fallet är inte över hur stora landområden partiet har stöd utan hur många människor som stöttar partiet. Det är ju inte gräs och buskar som röstar. När du ser ett diagram i form av en karta, tänk på populationsdensiteten och om den kan spela roll för hur kartan ser ut.

Vilken information syns inte?

Handlar det om relativa risker? Säg att det är 0,002% risk att drabbas av en viss sjukdom. Om du äter minst en påse chips i veckan ökar risken med 300%! Det låter jättemycket. I själva verket är det en ökning från 0,002% till 0,006%. Risken att du drabbas är alltså fortfarande mycket liten. I medicinska sammanhang talas det ofta om sådana så kallade relativa risker: risken för att drabbas av sjukdomen om du äter chips i förhållande till risken om du inte äter chips. Det kan leda till väldigt stora procenttal som ser skräckinjagande ut. Tänk på att den faktiska förändringen bakom procenttalen kan vara liten.

Är det per capita eller totalt? Tänk dig att du har en lista över olika företag och hur många klagomål de får in till sin kundtjänst. I tabell 1 nedan ser vi att Företag 1 får flest klagomål. Men är det egentligen intressant vem som får flest? Företag 1 är kanske större än de andra företagen och har fler kunder. Det intressanta är rimligtvis hur många klagomål företagen får in per capita, till exempel per 100 000 kunder. I tabell 2 syns antalet klagomål per capita. Där ser vi att *antalet klagomål justerat för antal kunder* är högst för Företag 3. Vi kan se samma fenomen när vi tittar på befolkningstäthet. Självklart sker det fler cykelolyckor i Skåne än i Härjedalen. I Skåne finns det ju betydligt fler människor!

Företag	Antal klagomål
Företag 1	318
Företag 2	254
Företag 3	128

Tabell 1. Antal klagomål till tre olika företag.

Företag	Antal klagomål/100 000 kunder
Företag 1	487
Företag 2	214
Företag 3	99

Tabell 1. Antal klagomål per 100 000 kunder till tre olika företag.

Har det justerats för inflation? En vara kostade 50 kronor år 1960. Idag kostar den 400 kronor. Det är en ökning med 800%. Har företaget bakom varan verkligen varit så fräcka att de har höjt priset så mycket? Här är det viktigt att komma ihåg inflationen. 50 kronor år 1970 motsvarar 407 kronor i dagens penningvärde ([här finns en prisomräknare](#)). När det pratas om hur mycket dyrare något har blivit, tänk på att inflationen kan spela en stor roll i sammanhanget. Det går att justera statistik utifrån inflation så att den blir mer rättvisande.

Vid procent: hur ser de absoluta talen ut? Ett företag har 91,67% nöjda kunder. Det låter jättebra. Tänk dig nu att det bara är 12 kunder som har tillfrågats i undersökningen. 11 av dem var nöjda. Procentsatsen 91,67% stämmer. Problemet med procenttal är att de lätt ger ett intryck av att det har gjorts en stor och seriös

undersökning för att ta fram dem. Det behöver inte vara fallet. När så få personer tillfrågas är den så kallade statistiska felmarginalen stor. Slumpen i vilka som tillfrågas har en väldigt stor inverkan på resultatet. Hade företaget frågat 1000 personer hade det kanske blivit ett helt annat resultat än 91,67%.

Hur ser den långa trenden ut? Trender redovisas ofta i diagram. Det kan gälla till exempel försäljning, bensinpris eller temperatur. I många fall är det den långa trenden som är intressant. Ett diagram som visar hur temperaturen förändrats de senaste tio åren säger ingenting om någon global uppvärmning. Tio år är för kort tid för att visa detta. När du ser en sådan tidsserie som visar en trend, fundera på om den långa trenden kan se annorlunda ut. Fundera också på varför just dessa start- och slutpunkter har valts. De kanske har valts för att ge ett speciellt intryck.

Finns det grupper som har en annan utveckling? "Sveriges pensionärer har fått mer pengar med vår politik!" hävdar ett regeringsparti. Men det kan finnas väldigt stora skillnader inom gruppen pensionärer. Det är möjligt att medelvärdet för alla pensionärer totalt sett är bättre, men att till exempel de 20% som har lägst pension faktiskt har fått det sämre. Klyftorna har ökat. Men eftersom detta bara gäller 20% av pensionärerna "drunknar" deras siffror i medelvärdet och syns inte. Medelvärdet är ett väldigt trubbigt verktyg. De är bra på många sätt men kan dölja skillnader mellan grupper.

Ger medelvärdet en skev bild? Hur kan fördelningen tänkas se ut? Medelvärdet ger lätt ett intryck av att alla ligger på ungefär samma nivå. Så kan det vara men det behöver inte vara så. Om lönerna på ett företag är 30 000 kronor kan det betyda att alla tjänar ungefär så mycket. Det kan också betyda att vissa tjänar betydligt mer och vissa tjänar betydligt mindre. Statistik är vetenskapen som säger att om du håller ena handen i skällhett vatten och har den andra handen infryst i ett isblock så har du det i genomsnitt ganska skönt.

Hur har genomsnittet räknats ut (medelvärde, median)? Det finns 5 anställda på ett företag. De tjänar 20 000, 20 000, 20 000, 22 000 och 58 000 kronor. Medelvärdet är 28 000 kronor i månaden (summera alla lönerna och dela med 5!). Medelvärdet blir missvisande. I sådana här fall är det bättre att använda medianen, som är värdet i mitten när alla värden har rangordnats: 20 000 kronor. Det säger inte heller hela sanningen, men det lyckas pricka in vad de flesta på företaget tjänar. Om något presenteras som ett genomsnitt kan det vara ett medelvärde eller en median. Det kan vara stor skillnad mellan dessa. Kan något av måtten ha valts strategiskt för att ge ett visst intryck?

Hur har urvalet gjorts?

Vilka ingår i urvalet? Är det rimligt? Om det har gjorts en undersökning har den som genomför den på något sätt valt ut vilka som ska ingå. Se om du kan hitta information om hur detta urval har gjorts. Vilka som svarar på undersökningen avgör väldigt mycket hur resultatet blir.

Är urvalet slumpmässigt? Ofta har ett slumpmässigt urval gjorts. Det är bra eftersom ett icke-slumpmässigt urval inte kan ses som representativt för den grupp som undersökningen vill uttala sig om. Tänk dig att en tidning ställer en politisk fråga på sin hemsida som du som besökare kan svara på. Tidningen kan inte utifrån denna undersökning dra slutsatser som "35% av Sveriges befolkning tycker...". Det beror på att det inte är ett slumpmässigt urval av svenska befolkningen som besöker hemsidan och svarar på frågan. De som besöker tidningens sida kanske är mer benägna att ha en viss åsikt i frågan?

Finns det risk att urvalet är snett? Ett snett urval representerar inte den grupp som vi vill uttala oss om. Antingen har urvalet inte varit slumpmässigt, eller så har det slumpmässiga urvalet gjorts på ett sätt som inte är så bra. Om någon vill undersöka vad svenska folket tycker om ett lagförslag kan hen inte fråga människor som hen träffar på stan. Då får hen ju huvudsakligen med människor som bor/arbetar/studerar i denna stad och är ute vid just den tiden. Det är inte säkert att de är representativa för hela Sveriges befolkning, och då blir resultatet missvisande.

Hur har undersökningen genomförts?

Var kommer siffrorna från (register, enkäter, experiment)? Många siffror kommer från Sveriges officiella statistik. Dessa siffror finns ofta tillgängliga på myndigheters hemsidor. Mycket forskning baseras på siffror från Sveriges officiella statistik och denna statistik håller hög kvalitet. Statistik kan också komma från en enkätundersökning. Dessa undersökningar kan ha hög eller låg kvalitet. Detsamma gäller experiment: de kan vara bra eller dåligt genomförda.

Hur många har intervjuats/svarat på enkäten/deltagit i experimentet? Om det är väldigt få som har undersökts bör det ringa en varningsklocka. Ju färre individer som ingår, desto högre är den så kallade statistiska felmarginalen. En hög statistisk felmarginal betyder att slumpen har en stor inverkan på varför resultatet ser ut som den gör och att resultatet därför har en stor osäkerhet.

Hur stort är bortfallet? Det är farligt att utgå från att de som väljer att inte svara på en undersökning har samma åsikter som dem som väljer att svara. Om en slumpmässig urvalsundersökning har gjorts, se om du kan hitta information om hur stort bortfallet är.

När är personerna tillfrågade/undersökningen gjord? Människors åsikter är inget statistiskt som inte förändras över tid. Olika händelser kan påverka hur människor svarar i en undersökning. I samband med vissa händelser kan människor vara mer eller mindre benägna att ha en viss åsikt, så vad som nyss presenterats på nyheterna kan avgöra om människor är för eller emot ett visst lagförslag eller nöjda som kunder hos ett företag.

Hur är frågan formulerad? Det är viktigt att titta noga på frågan och fundera på varför den är formulerad som den är. Kan den vara formulerad just så för att styra svaren i en viss riktning? En frågas formulering har stor påverkan på hur människor svarar.

Är frågan känslig? Känsliga frågor som handlar om till exempel personliga alkoholvanor, sexualitet och brottslighet ökar risken för att människor inte är ärliga. De kanske skäms eller är rädda att deras svar inte är anonyma.

Är frågan ledande? En ledande fråga försöker styra svaret i en viss riktning. "Tycker du precis som många andra svenskar att hundskatt ska införas?", "Visst vore det väl bra med en sjätte semestervecka?" och "Du har väl förtroende för vår finansminister?" är extrema exempel på ledande frågor. En ledande fråga kan också föregås av information som leder i en viss riktning: "23 procent av Sveriges kvinnor har avstått från att gå ut av rädsla för att utsättas för hot eller våld. Tycker du att straffen för våldsbrott ska skärpas?".

Vilka svarsalternativ finns? Det är viktigt att alla tänkbara alternativ finns med och att svaren inte överlappar varandra. Det ska vara klart och tydligt för den som svarar. Ibland kan det vara svårt att svara på en fråga om de enda alternativen är ja och nej. På en enkel fråga som "Är rött en fin färg?" kan det finnas många som har en åsikt mellan ja och nej, till exempel: "ja, ganska, men grönt är finare" eller "nej, men vinrött är helt okej".

Hur tolkas resultatet?

Vilka definitioner har använts? Genom att definiera saker på olika sätt går det att komma till olika slutsatser. Flera partier kan till exempel hävda att just deras parti för den bästa HBTQ-politiken. Skillnaden är att de definierar "bästa HBTQ-politiken" på olika sätt. För vad är det egentligen som ska mätas? Hur partiet har röstat i HBTQ-frågor i riksdagen? Vad som står i partiprogrammet? Hur partiets kandidater har svarat i en enkät att de tycker? Fundera på om definitionen som används är rimlig.

Jämförs äpplen och päron? Äpplen innehåller 56 kalorier per 100 gram och päron 282 kalorier per 100 gram. Ja, alltså, det gäller för färska äpplen och torkade päron. Är det rimligt att jämföra dem såhär? Nej! Det kan vara förrädiskt att jämföra. Att jämföra mellan till exempel år och länder kan bli missvisande. Definitioner som används kan se annorlunda ut. Olika brott kanske inte klassificeras på samma sätt i olika länder. Det har kanske skett en lagändring som ändrat definitionen och gör att jämförelser mellan år blir fel.

Ligger resultatet inom felmarginalen? Den så kallade statistiska felmarginalen är den osäkerhet som alltid ingår när en slumpmässig urvalsundersökning genomförs. Om ett politiskt parti får 5,6% i en opinionsundersökning, och den statistiska felmarginalen är 0,9% med ett konfidensintervall på 95%, ska det läsas såhär: det är 95% sannolikhet att partiet får mellan 4,7% och 6,5%. Om vi mäter hur mycket stödet för ett parti har förändrats, och förändringen ligger innanför felmarginalen, kan vi inte säga att vi har skett en förändring. Då är sannolikheten för stor att det som ser ut som en förändring är slumpens inverkan.

Finns det ett mörkertal? För att bedöma hur många människor som använder narkotika kan vi inte titta på statistik över dömda för narkotikabrott. Många som använder narkotika grips aldrig för det, och då döms de heller inte. Detta kallas för ett mörkertal. Mörkertal är vanliga till exempel när det rör sig om brottslighet.

Gäller resultatet för en annan tid, plats och grupp? Om en undersökning har genomförts på boende i Skövde kommun 2016 är det svårt att generalisera till andra tider och platser. Kanske hade resultatet blivit ett helt annat i Jokkmokks kommun 2020? Ibland kan vi se att för långtgående slutsatser dras utifrån resultat som rör en specifik tid, plats och grupp. När detta görs, fundera på om det verkar rimligt att det går att generalisera på det sättet.

Vilka slutsatser dras? En enkät innehåller frågan "Har du känt dig nedstämd någon gång de senaste två veckorna?". 34% svarar ja på frågan. Nyheten blir: "En tredjedel av svenskarna är deprimerade!". Det har dragits en slutsats av svaren som det inte finns belägg för. När en slutsats dras, fundera på om den verkar rimlig och om det finns belägg för den i den undersökning som genomförts.

Finns det ett samband?

Är detta ett samband eller ett skensamband? Att två saker händer samtidigt betyder inte att den ena har orsakat den andra. Vi ser en ökning av antalet autismdiagnoser och vi ser att människor skaffar allt större tv-apparater. Leder stora tv-skärmar till autism? Eller är autistiska människor mer benägna att köpa stora tv-apparater? Självklart inte. Som samhälle har vi blivit allt bättre på att diagnosticera autism och den tekniska utvecklingen går också framåt, vilket ger större tv-apparater. [Här finns en hemsida som samlar absurda så kallade skensamband.](#)

Finns det en bakomliggande förklaring som inte syns? Ibland kanske A inte orsakas av B, eller B av A, utan det hela beror på en mellanliggande variabel Z. Visste du att människor med stora fötter är mer intelligenta än människor med små fötter? Den mellanliggande variabeln är ålder. När barn växer upp ökar deras intelligens och också storleken på deras fötter. Sambandet gäller alltså bara för barn, inte vuxna. Nu, när vi känner till den mellanliggande variabeln, verkar det till synes märkliga sambandet genast rimligt.

Åt vilket håll går orsakssambandet? Vad är hönan och vad är ägget? Statistik kan visa att det finns ett samband men inte åt vilket håll sambandet går. Om du har statistik som visar att kvinnor har lägre lön än män svarar siffrorna inte på om a) att vara kvinna leder till att du får lägre lön, eller b) om lägre lön leder till att du blir kvinna.

I det här exemplet faller det på sin egen orimlighet att det skulle vara det sistnämnda. Men det är inte alltid så uppenbart. Säg att det finns en koppling mellan att vara deprimerad och ha dålig kondition. Leder dålig kondition till depression? Eller är det så att deprimerade människor inte motionerar i lika stor utsträckning som andra?