# Skillnaden mellan beskrivande statistik och inferentiell statistik

Beskrivande statistik handlar om att beskriva, visa och sammanfatta data. Det handlar om att göra ett datamaterial överskådligt och söka mönster i materialet. I beskrivande statistik studerar vi bara den data vi samlat in. Beskrivande statistik tillåter oss inte att dra slutsatser bortom den data vi samlat in och sammanställt om, till exempel, en population. Det är helt enkelt ett sätt att beskriva det datamaterial vi samlat in. Medelvärde, median, typvärde är tre lägesmått som används inom beskrivande statistik. Kvartil, variationsbredd och standardavvikelse är begrepp som används inom beskrivande statistik för att beskriva spridningen (variationen) i datamaterialet. Säg t.ex. att Skolhälsan på skolan Pluggis önskar få reda på alla (totalundersökning) sina elevers sömnvanor (F-9 skola). Populationen är Pluggis elever. När all data samlats in sammanställer skolhälsan materialet med verktyg från beskrivande statistik. Genom den beskrivande statistiken kan de urskilja mönster och tendenser. De kanske urskiljer att antalet sömntimmar minskar ju äldre eleverna blir, att det inte verkar vara någon större skillnad mellan pojkar och flickors sömnvanor, osv. I det här fallet klarar sig skolhälsan med beskrivande statistik i och med att de har tillgång till hela den population de utforskar.

Den beskrivande statistiken dominerar statistikundervisningen. Elever arbetar med att beräkna medelvärden, medianvärden, kvartiler och variationsbredd och åskådliggöra dessa värden i tabeller och diagram. Betydelsen av den beskrivande statistiken ska inte underskattas men den är begränsad till totalundersökningar. Oftast har vi dock inte tillgång till hela populationen utan bara av en begränsad mängd information av den genom stickprov. Det är den inferentiella statistiken som tillåter oss att dra generella slutsatser om en population baserat på stickprov. En förutsättning är dock att stickprovet är representativt för den population man vill dra en slutsats om. Om vi till exempel kan motivera att Pluggis är en skola som är representativ för svenska skolor så kan vi använda enkätsvaren från Pluggis som ett stickprov för att dra en slutsats (en inferens) om svenska elevers sömnvanor med hjälp av tekniker från inferentiell statistik.

Parameterestimering (parameteruppskattning) och hypotesprövning är två fall av inferentiell statistik. Medelvärdet är en parameter och Median är en annan. En parameterestimering i exemplet med elevers sömnvanor skulle då kunna vara att med Pluggis som ett representativt stickprov uppskatta medelvärdet av sömntimmar bland svenska F-9 elever. Vi kan inte bestämma medelvärdet, som vi kan göra i beskrivande statistik, då vi inte har tillgång till hela populationen. Vi kan endast uppskatta (estimera) medelvärdet. Vi använder sedan konfidensintervall för att bestämma de gränser inom vilka vi med en given säkerhet bedömer att det ”sanna” medelvärdet ligger, utifrån stickprovets medelvärde. I Modul 3 ska ni dock inte arbeta med parameterestimering utan med hypotesprövning. I hypotesprövning är man inte ute efter att uppskatta ett specifikt värde på en egenskap i populationen. I hypotesprövning är man ute efter att uppskatta om det föreligger en verklig skillnad mellan variabler i en population eller om den skillnad man observerar i en undersökning endast är resultatet av slump. Men, i att dra slutsats om en skillnad mellan variabler är verklig eller bara beror av slump blir medelvärdet centralt. Mer precist, i hypotesprövning används medelvärdet för en helt *likformig fördelning* (50-50 som referens. Vi återgår till exempel med elevers sömnvanor för att illustrera idéen med hypotesprövning.

Vi ser här Pluggis elever som ett stickprov av svenska elever. Det går att välja olika variabler. Som variabel skulle vi kunna välja årskurser och jämföra om det är någon verklig skillnad mellan olika årskursers sömnvanor. Men, vi kan också väl kön som variabel och då jämföra sömnvanor mellan flickor och pojkar. Det går att hitta fler variabler, som t.ex. betyg, socio-ekonomiskt bakgrund osv. Här väljer vi kön som variabel. Säg att det går *N* antal elever[[1]](#footnote-1) på Pluggis och att vi ber alla elever notera hur många timmar de sovit när de vaknar och att de gör det varje dag under ett år. För flickorna blir sömnmedelvärdet 7,9 h/natt och för pojkarna 8,1 h/natt. Baserat på detta stickprov är då frågan om vi kan säga att sömntiden mellan pojkar och flickor i Sverige (ålder F-9) skiljer sig eller inte. Vi prövar därför resultatet vi fick från Pluggis mot en hypotes! Vi prövar det mot hypotesen att det inte skulle finnas någon skillnad mellan flickor och pojkars sömnvanor, dvs. att fördelningen mellan flickor och pojkars sömnvanor är likformig i populationen. Med andra ord så frågar vi, ”Hur stor, eller liten, är sannolikheten att i ett stickprov om *N* deltagare få en skillnad på 0,2 h/natt mellan flickor och pojkar om det inte skulle finnas någon verklig skillnad (sömnmedelvärdet är samma) i den population som Pluggis är ett stickprov från?” Även om sömnmedelvärdet är samma så blir det alltid en skillnad när man tar ett stickprov. Frågan är bara om skillnaden är tillräckligt stor för att man ska börja tvivla på att hypotesen gäller eller om man ska förkasta den och dra slutsatsen att det finns en verklig skillnad i sömnvanor mellan pojkar och flickor. Vi utför inga beräkningar här men, om sannolikheten att under hypotesen likformig fördelning få en skillnad på 0,2 h/natt visar sig vara mindre än 5%[[2]](#footnote-2) tillåter konventionen att vi drar slutsatsen att det är en veklig skillnad (alltså, en skillnad som inte bara beror av slump) mellan flickor och pojkars sömnvanor bland svenska F-9 elever. Om det skulle visa sig att en visad skillnad i ett stickprov inte skulle vara så osannolikt att få från en likformig fördelning (att den sannolikheten är över 5%) så kan vi inte utesluta att skillnaden endast är ett resultat av slumpvariation och kan då inte dra slutsatsen (inferensen) att det är en skilland i sömntimmar mellan svenska F-9 flickor och pojkar.

1. Det är ungefär lika många flickor som pojkar på skolan. [↑](#footnote-ref-1)
2. En signifikansnivå på 5% är vanligt förekommande. [↑](#footnote-ref-2)