

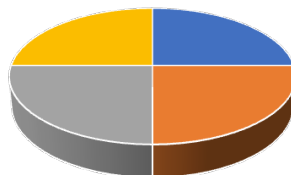
Kritisk granskning i statistik

Statistik bygger på relativt stora datainsamlingar vilket gör att statistiska undersökningar är lämpliga för att belysa samhällsfrågor. Denna artikel ger några exempel på vad kritisk hållning i statistisk läskunnighet kan vara från lågstadium till gymnasium med och utan digitala verktyg.

Iddo Gal, som är en tongivande statistikdidaktiker, menar att statistisk läskunnighet (literacitet) omfattar både kunskap och attityd. Kunskap avser kunskaper i både matematik och i det ämne som statistiken ska beskriva. Attityd avser att ställa frågor med en kritisk hållning.

3D-diagram är ofta olämpliga

I tidigare läroplaner, även utanför Sverige, har ett viktigt moment i statistik varit att genomsåda när ett i övrigt korrekt diagram ger ett förvrängt intryck, vilket i Lgr-69 gick under namnet "att ljuga med statistik". Ett typ-exempel på detta är 3D-grafik. Cirkeldiagrammet intill är kanhända visuellt vackert och geometriskt korrekt, men absolut inte informativt avseende dess statistiska budskap. Den geometriskt korrekta projektionen gör att föremål längre bort blir mindre, precis som en cykel på långt håll kan täcka lika stor del av synfältet som en insekt på mycket nära håll och detta är enkelt att illustrera experimentellt. I statistik ska diagram inte störa en korrekt förståelse utan istället stödja den. Därför är vanlig 2D-grafik att föredra.



Finns samband mellan biotop och populationsstorlek?

Kritisk granskning och läskunnighet i statistik är förstås mycket mer än att inte göra 3D-grafer. Följande övning kan skapa ett lärandetillfälle för kritisk hållning genom att man kan upptäcka ett samband mellan två slumpvariabler. Använd en rockring eller en från slöjden hopsnickrad mätruta. Läg den på en (a) nyklippt gräsmatta och räkna under en minut antalet insekter fördelade på (1) bin och humlor och (2) fjärilar i mätområdet. Räkna därefter även antalet blommor som är åtminstone en tumnagel stor. Ordningen är viktig – börja med att räkna insekterna eftersom de kanske skräms bort när blommorna räknas. Gör samma mätning, alltså samma tidslängd och storlek på arean, fast på en (b) oremsad gräsyta med diverse blommor såsom maskros, tusensköna, scilla eller vad som finns i gräsmattan. Gör samma mätning en tredje gång, fast i en (c) rabatt med flera sorters blommor. Observationerna ska göras under samma väderförhållande och då helst dagar med vackert väder när insekterna är aktiva. Redovisa detta i ett grupperat stapeldiagram. Den vågräta axeln ska visa kategorierna a–c {nyklippt; oremsad; rabatt} där



Scilla

även antalet blommor kan antecknas vid respektive kategori. Den lodräta axeln ska visa antalet insekter i två separata staplar (bin och humlor; fjärilar) för varje kategori av mätområde.

Kanske resultatet av denna undersökning blir att eleverna föreslår kommunen att man klipper halva parken åt gången, kanske remsor, så att insekterna inte behöver flytta så långt till nästa park som har maskrosorna kvar? Och även att ha diverse rabatter här och där med lite högre blommor eftersom fjärilar föredrar högre blommor framför marknära blommor.

Grundläggande försäkringsmatematik

Du ska köpa en tågbiljett Alvesta–Norrköping och finner att andraklass med återbetalningsskydd/ombokningsbar kostar 625 kr medan en biljett som ej kan ombokas kostar 475 kr. Det är klart att 475 kr är billigare, men tänk om du blir sjuk och inte kan resa. Då förlorar du 475 kr. Detta är faktiskt ett statistiskt problem som löses med försäkringsmatematik, exempelvis som i kalkylbladet nedan. Formeln i cell D5 visar skillnaden mellan att å ena sidan köpa 2 stycken resor (A5) utan återbetalningsskydd till priset i B2 och å andra sidan kostnaden för att köpa biljetter för priset i C2, men avboka en (A5-1).

Dollartecknen för B\$2 och C\$2 betyder att radnummer två behålls när man kopierar formeln nedåt. Här kan eleverna jämföra vad som händer om de kopierar formeln nedåt utan att ha med dollartecknet. Med dollartecken framför kolumnbokstäverna går det att bevara kolumnen när man kopierar i sidled. Kalkylbladet visar att det är dyrare att köpa fyra resor som kan ombokas än fem resor som inte kan ombokas. Statistiskt betyder detta att om du behöver ställa in högst var femte resa, så är det faktiskt billigare att köpa resor som inte går att omboka.

	A	B	C	D
1		Utan återbetalning	Med återbetalning	
2	Pris för 1 resa	475	625	
3	Antal resor			
4	1	475	625	475
5	2	950	1250	=A5*B\$2-(A5-1)*C\$2
6	3	1425	1875	175
7	4	1900	2500	25
8	5	2375	3125	-125
9	6	2850	3750	-275

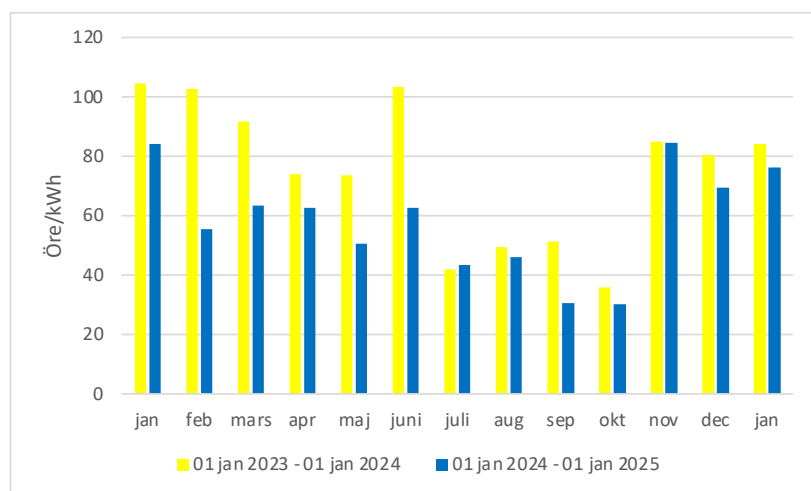
Liknande exempel finns för en del hemelektronik där man kan köpa till en drulleförsäkring om man håller läsk i bildskärmen (!) eller tappar den. Försäkringskostnaden ligger på runt 10 % av nyvärdet. Skulle man dessutom byta en fungerande skärm redan efter 5 år, så motsvarar försäkringskostnaden egentligen 20 % av värdet. Här är det alltså värt att fråga sig om det var minst fem år sedan man tog sönder sin bildskärm till datorn eller TV:n. I så fall är nog försäkringen onödig. Det är effektivare att ställa eller hänga upp skärmen på en plats där den inte riskerar att tippa eller få kaffe i sig.

En viktig aspekt för försäkringar mot förluster av små belopp är att administrationskostnaden ofta blir stor i förhållande till skadekostnaden. Däremot om det handlar om stora belopp, då blir försäkringen intressant. Ett exempel är en hemförsäkring mot brand. Erfarenheten säger att det är mycket liten risk att bostaden börjar brinna. Många hus runt omkring dig är säkert både

50 och 100 år gamla och har inte drabbats av brand och det kan man tycka är ett skäl att inte försäkra sig mot brand. Men, om olyckan ändå är där, så blir kostnaderna ofta stora. Även en släckt brand medför ofta kostnader för vattenskador, renovering av system för ventilering samt att måla om bostaden och köpa nya stoppade möbler då de gamla luktar sot. Det kostar många hundratusen och är en oförutsedd utgift man inte vill ha även om risken att få den är mycket liten.

Kritisk granskning och datamining

Denna övning är på temat elpriser. Kärnkraft motiveras ofta som ett sätt att kapa pristoppar på el. Men är det så? Här kommer statistisk läskunnighet till uttryck både som kritisk granskning och datamining (datautvinning), som syftar på att undersöka större datamängder. En bakgrundsinformation är att ägarbolagen till svensk kärnkraft vill ha en prisgaranti på 80 öre/kWh för att inte vid nybyggnad av kärnkraft gå med förlust. Med denna bakgrundsinformation kan eleverna ta reda på elpriset med datamining. Två användbara databaser är Nordpool och Vattenfall. Nordpool ger elpriser i ett linjediagram med den för storhandlare lämpliga enheten (kilo-)kronor per MWh medan Vattenfall ger ett stapeldiagram med den välbekanta enheten öre/kWh. Figuren nedan jämför genomsnittspriset per månad 2023 och 2024.



Vad kan diagrammet berätta? Jo, kärnkraft kan kapa toppar på elpriset, men ... Redan idag finns det ett enkelt sätt att kapa toppar, nämligen genom fastprisavtal, som fungerar på precis samma sätt som en försäkring. Diagrammet visar att början på 2023 hade höga elpriser, precis som det även var under en stor del av 2022, och detta berodde på avvecklingen av rysk import av naturgas på grund av Rysslands krig mot Ukraina. Trots det hade de flesta månader 2023 ett medelpris på under eller nära 80 öre/kWh, vilket alltså är det pris som kärnkraftsbolagen vill ha som garanti. Med andra ord skulle kärnkraftsproducerad el gå med förlust och en not i sammanhanget är att det inte skulle synas på elräkningen men istället på skattsedeln. En komplikation med detta resonemang är dock att dagspriserna nu för tiden är höga på dagen när vi behöver elen. Kan vi istället styra om elanvändningen till natten? Ja, ett sätt för

privatpersoner är att ladda elbilen och värma vattnet i varmvattenberedaren på natten om man inte har fjärrvärme. Några företag kommer även att tillverka vätgas på natten för att använda vätgasen vid exempelvis tillverkning av det som kallas "grönt stål". När dessa ganska stora konsumtioner sker på natten så blir det mindre skillnad på dagspris och nattpris.

Kritisk granskning av spridningsmått

Kunskaper om statistiska spridningsmått är viktiga för att avslöja friserad data. Under covid-pandemin kunde man se att olika länder drabbades på mycket olika sätt. En del skillnader har demografiska orsaker såsom att länder med stor andel gamla hade högre andel dödsfall än andra länder. Smittspridningen främjades också av att ha en väl utbyggd kommunikation där människor reste, exempelvis till Italiens alper, och tog med smittan därifrån. Det fanns också politiska förklaringar, exempelvis att mitt i en smittvåg i Burma (Myanmar) upphörde journalister och vårdpersonal med att rapportera främst på grund av att många av dem fängslades under statskuppen 1 februari 2021. Det finns alltså flera bakomliggande orsaker till att statistiken för dödstalen ser olika ut. Det var dessutom så att många länder endast rapporterade dem som dog på sjukhus men inte på ålderdomshem och därför rapporterade dessa länder lägre dödlighet. En sak har dock alla dessa länder gemensamt: Variationen i antalet döda från dag till dag och säsong är ganska stor.

En övning om spridningsmått: Jämför formen på grafen för Belarus med annat valfritt land med ungefär samma folkmängd (10 ± 2 miljoner). Är det något särskilt som skiljer i formen?



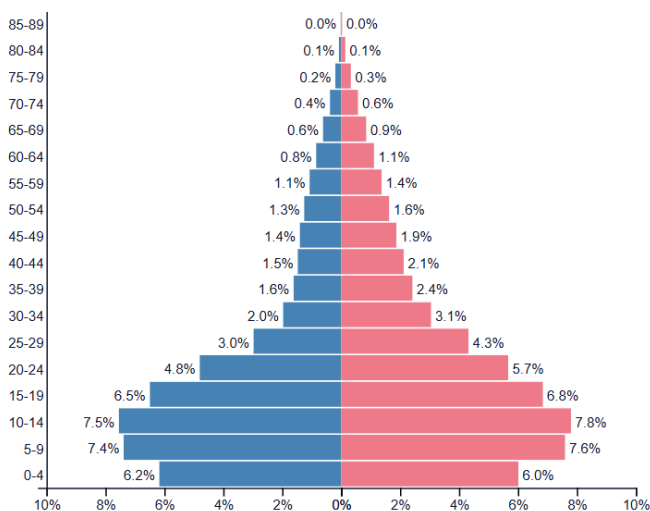
Antalet dödsfall i Covid-19 för Belarus (coronavirus.jhu.edu/region/belarus).

I detta avseende avviker Belarus kraftigt från alla andra länder av rent matematiska skäl. Dels saknar dödstalen i stort sett säsongvariationer, men matematiskt märkligare är att antalet döda i Belarus låg mycket konstant på en av de tre plattåerna 5 ± 1 , 10 ± 1 eller 15 ± 1 personer per dag. Inget annat land (med ungefär samma folkmängd) har ett sådant mönster. Det är lite som att upprepat kasta många tärningar och få väldigt många 2, 4 och 6 men endast få 1, 3 och 5. Slutsatsen av covid-statistiken från Belarus är att den inte kan vara annat än kraftigt friserad. Hur den friserades eller rapporterades kanske vi aldrig får reda på. Det kan vara så att statistiken bygger på endast en avdelning med först 5 och sedan 10 respektive 15 platser för döende, men det är bara en spekulation. Oavsett orsak så blir slutsatsen att statistiken inte är pålitlig för landet, även om just den rapporterade statistiken kanske stämmer för en utvald vårdavdelning.

Kritisk granskning av befolkningspyramider

Detta är en riktigt otäck historia. På sajten <https://www.populationpyramid.net/cambodia/1980/> kan man se att nativiteten sjönk kraftigt ungefär 1980 genom ett stort hack i befolkningspyramiden. Där syns även en svacka i ett linjediagram för Kambodjas totala folkmängd. Förklaringen står visserligen i historieböckerna, men rent

matematiskt kan man avläsa det på följande sätt. Vanligen föds det något fler pojkar än flickor. Samtidigt har pojkar något högre dödlighet i alla åldrar, vilket ger ett kvinnoöverskott i högre åldrar. Detta kvinnoöverskott utvecklar sig sakta över tid, men för Kambodja utvecklas detta mycket plötsligt och särskilt i åldrarna 15–50 år, vilket man ser om man jämför befolkningspyramiderna för 1980 (bilden) och 1970. Sådana snabba förändringar betyder i regel bara en sak: våld. Spår av detta kan man även se i exempelvis Finland och Tyskland för dem som var minst 15 år från 1940.



Kambodjas befolkningspyramid 1980 som visar fördelning mellan män (blått) och kvinnor (rött).

Statistik och politisk aritmetik

Eftersom statistik per definition kräver en någorlunda stor datainsamling hamnar man per automatik ofta utanför sin egen privata sfär. Med andra ord: statistiska frågor handlar ofta om samhället och därmed ofta om ekonomiska och politiska frågor. Ett historiskt exempel på detta är att Hermann Conring jämförde *stater* avseende naturtillgångar och näringsliv. Han var rådgivare 1650–1669 till drottning Kristina, Karl X och XI. Conrings efterföljare Gottfried Achenwall gav därför Conrings vetenskap namnet *statistik*. Ännu ett exempel på att statistik ofta berör samhällsfrågor är att John Graunt under 1600-talet kallade det för *politisk aritmetik*. Sensmoralen blir att kunskaper i statistik och sannolikhetslära är viktigt vid beslutsfattande som berör samhället och att det kan användas för att granska samhället.

LITTERATUR OCH LÄNKAR

- Sjöström, O. (2002). *Svensk statistikhistoria: en undanskynd kritisk tradition*. Gidlund.
- Tukey, J. W. (1980). We need both exploratory and confirmatory. *The American Statistician*, 34(1), 23–25.
- Nordpool: <https://data.nordpoolgroup.com/>
- Vattenfall: <https://www.vattenfall.se/elavtal/elpriser/timpris/>