

Att bestämma osäkerhet: användning för statistiken

Bättre osäker än felaktigt tvärsäker!

För att göra en bästa gissning behöver vi information, men hur uppskattar vi sedan hur osäker gissningen är? Detta diskuterar författaren med utgångspunkt i statistiska begrepp som variationsbredd, normalfördelning, standardfel och MOE – Margin of Error.

I en tid då tvärsäkra påståenden och övertygelse får allt större plats, är det desto viktigare att lyfta upp dess diametrala motsats, osäkerheten. Traditionellt är osäkerhetsbegreppet kopplat till negativa konnotationer just för att vi som människor föredrar det säkra och kända. Problemet är emellertid att det som tycks vara sant och säkert sällan är det, och värre än att vara osäker är att vara säker på något som är fel. Som den tidigare amerikanske försvarsministern Donald Rumsfeld uttryckte det:

There are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns – the ones we don't know we don't know.

Med andra ord, ett första steg i att nå bättre klarhet är att gå från okänd osäkerhet till känd osäkerhet. Det ämne som ansvarar för denna vitala transformation är statistiken och två syften är att:

- ◇ givet existerande information förse oss med en bästa gissning
- ◇ ge en uppskattning av hur osäker gissningen är.

Medan det första syftet någorlunda brukar komma in i matematikundervisningen genom allt från medelvärden till linjär regression, är det framförallt det andra syftet som fått en mycket underordnad roll. I exempelvis naturvetenskapliga laborationer är det inte helt ovanligt att göra upprepade mätningar och beräkna medelvärden, men betydligt mindre vanligt att försöka använda värdena till att uppskatta osäkerheten. Problemet med detta är att medelvärdes osäkerhet blir en okänd osäkerhet, och svår för elever att sätta i förhållande till andra källor.

Även i matematikundervisningen lyser detta osäkerhetsuppdrag i princip totalt med sin frånvaro. Viss statistik berörs kortfattat i Matematik 2, såsom normalfördelning och definition av standardavvikelse, men det blir ofta mer att de nämns och att några övningsuppgifter genomförs, snarare än att eleverna får med sig några verktyg för att hantera beräkningen av praktiska osäkerheter. Detta är något av ett missat tillfälle, för framförallt elever inom naturvetenskapligt program som ofta laborerar har ett stort behov av dessa tekniker.

Syftet med denna artikel är därför att visa hur matematiklärare, eller för den delen lärare i något annat ämne där data samlas, kan ge eleverna möjlighet att beräkna osäkerheter. Vi undervisar om detta varje år för eleverna inom *Unga Fysikers Världstävling* (IYPT) eftersom det krävs där att de kan uttala sig om osäkerheten i sina resultat.

Variationsbredd – en problematisk metod

Att diskutera statistisk osäkerhet generellt är så klart omöjligt, som ett av de mest spännande matematikområdena med en bredd från det praktiska till det abstrakta kan det knappast göras rättvisa på några sidor. Vi kommer därför att fokusera på en av de enklaste, men också vanligaste experimentella situationerna. En kvantitet x mäts upprepade gånger. Det kan vara spänning, tid, position, massa eller varför inte inkomst. Utifrån ett enskilt värde på x går det inte att säga särskilt mycket om dess noggrannhet, utan för att göra detta behövs upprepade observationer $x_1, x_2, x_3, \dots, x_N$. En inte ovanlig metod som en del lär ut är att skatta felet genom halva variationsbredden: $\frac{\max(x_n) - \min(x_n)}{2}$.

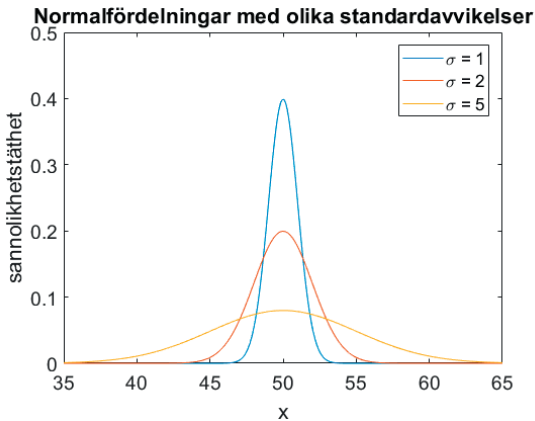
Intentionerna här är goda, variationsbredden beskriver den största observerade variationen och tanken är att det sanna värdet borde ligga där inom. Om man får vara lite negativ skulle man kunna säga att denna typ av feluppskattning lider av alla de brister som statistikteorin försöker lära ut. För att nämna några:

- ◇ En rimlig egenskap hos en felkattning är att då $N \rightarrow \infty$ så ska felet gå mot 0, men variationsbredden kommer alltid att vara en växande funktion när mer data samlas in. Det gör att den inte kan ge några råd om hur mycket mer data som ska samlas in.
- ◇ Variationsbredden utgår från spridningen i den data som har observerats. Det gör att den kritiskt underskattar felet i små datamängder, just de som vi ofta använder under laboration.
- ◇ Variationsbredden blandar samman ett *populationsmått*, spridningen hos alla datapunkter, med ett *parametermått*, osäkerheten i var medelvärde befinner sig.

Minimalistisk statistik – normalfördelning

Frågan är då om det faktiskt går att göra något bättre än variationsbredden med 3–5 observationer. Om vi har tillräckligt mycket data så behövs inga ytterligare antaganden för att fortsätta analysen (så kallad icke-parametrisk statistik), men eftersom vi är begränsade till något fåtal värden behövs även ett antagande om vilken typ av sannolikhetsfördelning som värdena följer.

Den mest minimala beskrivningen måste innehålla två parametrar, en som beskriver värdet utan mätfel (ofta betecknad μ), och en som beskriver storleken hos de experimentella störningarna. Normalfördelningen uppfyller dessa två villkor eftersom den har sitt medelvärde, median och typvärde i μ (en konsekvens av att den är symmetrisk och har en enda topp), och har parametern σ , kallad standardavvikelse för att beskriva spridningen i resultat.



Figur 1: Exempel på normalfördelningskurvor.

Det är viktigt att förstå att parametern σ är lika okänd som μ . Vi uppskattar μ med hjälp av medelvärdet, bland exempelvis fysiker betecknat som $\langle x \rangle$. På samma sätt kan vi uppskatta σ från den data vi har enligt formeln:

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^N (x_i - \langle x \rangle)^2}{N - 1}}$$

Hatten på $\hat{\sigma}$ är för att synliggöra att detta är en uppskattning. Denna är dock extremt osäker för små datamängder. Om $N=5$ och $\sigma=1$ så skulle upprepade experiment kunna ge dessa värden (genererade med slumpgenerator):

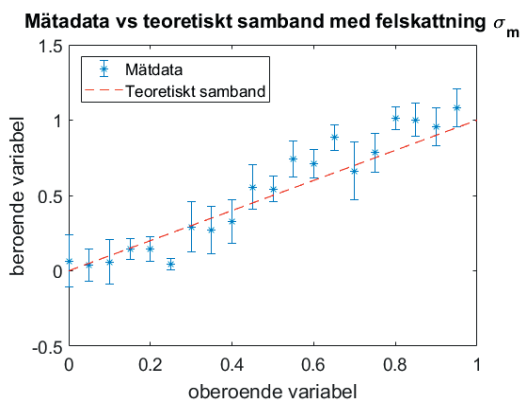
1,5219 2,0971 1,5968 0,8791 1,0322 0,7435 1,3608 1,3659 0,4874 0,6039

Notera att inget av värdena är särskilt nära $\sigma=1$, utan kan skilja sig åt ganska mycket från det sanna värdet och utan att det är något märkligt som pågår. Detta beror på att medan det är hyfsat enkelt att skatta centrum hos en fördelning, och se varomkring punkterna placerar sig, är det betydligt svårare att skatta en spridning. Det här gör också att många av de regler som lärs ut, såsom att införa att för två standardavvikelser finns 95% av datan, inte längre stämmer för den skattade standardavvikelsen. Det är också viktigt att komma ihåg att $\hat{\sigma}$ är ett mått på populationens spridning och säger egentligen inte så mycket om osäkerheten i medelvärdet.

Standardfel – ett osäkerhetsmått med konceptuella brister

För att ta reda på osäkerheten i medelvärdet går det att dra nytta av att det för normalfördelningen är så att även medelvärdet är normalfördelat, med samma μ och en standardavvikelse $\sigma_m = \sigma/\sqrt{N}$, ibland kallad standardfelet. Då N ökar minskar σ_m som väntat. Dock är avtagandet väldigt långsamt, om man vill halvera felet för en undersökning gjord med $N=5$, krävs $N=20$. Om man vill minska felet med en hel storleksordning krävs $N=500$. Det är en god lärdom att även om medelvärdet är bättre än de enskilda observationerna, så förtar det inte vikten av god experimentdesign, eftersom det ofta är betydligt effektivare att minska σ . En skattning för σ_m är $\hat{\sigma}_m = \hat{\sigma}/\sqrt{N}$. Det är vanligt att se (särskilt bland fysiker) att $\hat{\sigma}_m$ då används som uppskattning för felet. Även detta är en dålig idé av flera anledningar:

- ◇ Som räkneexemplet ovan visade, är $\sigma_m \approx \hat{\sigma}_m$ en dålig approximation för små datamängder och alltid en sämre approximation än $\mu \approx \langle x \rangle$.
- ◇ Betrakta figur 2. Skulle du säga att det för varje mätpunkt tycks råda överensstämmelse mellan teori och experiment? Mellan 0,6 och 0,8 skulle man snarare vilja hävda att det tycks finnas någon form av avvikelse. Detta är en ren illusion, datan är en rät linje med pålagt normalfördelat brus. Problemet är att med $N=5$ är sannolikheten bara 63% att felgränserna faktiskt överlappar det sanna värdet. Det gör det svårt att föra ett resonemang kring felens storlek.
- ◇ Det finns en godtycklighet i valet av $\hat{\sigma}_m$, det är en skattning av en parameter i fördelningen, men har inget direkt samband till det faktiska felet. Eftersom sannolikheten att det sanna svaret befinner sig i intervallet ändras när N ändras (50% för $N=2$ till 68,3% för $N=\infty$), är det inte heller ett konsekvent mått.



Figur 2: Stämmer sambandet? Det finns en stor risk att se avvikelser här som egentligen inte finns om man använder standardfel.

Hur gör man då? – MOE

Problemet med de tidigare presenterade osäkerhetsmåten är att de börjar i fel ände, nämligen att en intressant formel har presenterats, och sedan har vi analyserat dess egenskaper. Mer rimligt är att börja definiera osäkerhetsbegreppet och sedan leta upp en formel som passar till det.

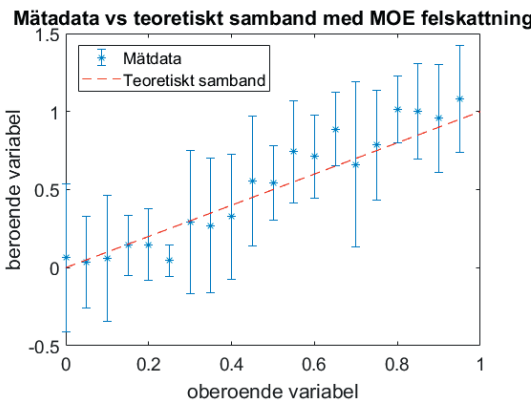
Ett ofta använt kriterium är att introducera ett så kallat sannolikhetsintervall där det är 95% säkert att rätt svar går att finna (den exakta tolkningen skiljer sig något mellan olika skolor av statistiker men spelar här ingen roll). Det betyder att i genomsnitt kommer var tjugonde felskattning att avvika, vilket gör att det går att tillfoga ganska hög tillförlitlighet till de flesta av uppskattningarna.

För en symmetrisk fördelning som normalfördelningen innebär det ett intervall centrat kring medelvärdet av mätvärdena och med en längd som på engelska kallas *Margin of Error*, MOE (ett begrepp som populariserats av bland annat Geoff Cumming), dvs felmätt. Detta är vårt hett eftersökta osäkerhetsmått. Det gäller att MOE är proportionellt mot $\hat{\sigma}/\sqrt{N}$, vilket också är väntat, men med en proportionalitetskonstant som beror på N .

Denna proportionalitetskonstant beräknades för första gången 1908 av William S Gosset i hans arbete med att få fram bra kvalitet på ölen hos Guinness (Salsburg, 2002) och kallas för t -faktor. I exempelvis excel beräknas den med funktionen `TINV` som tar två argument. Det första är sannolikheten att missa det rätta värdet, dvs 0,05 och det andra är hur många så kallade frihetsgrader som experimentet har, alltså antal mätdata minus antal skattade parametrar (förutom standardavvikelsen själv), dvs i detta fall $N - 1$. Om $N = 5$ skriver man `=TINV(0.05;4)`, vilket ger 2,78, så i detta fall är $MOE = 2,78 \hat{\sigma}_m$. Idén här är att eftersom $\hat{\sigma}_m$ har en inbyggd osäkerhet, kompenserar t -faktorn för denna genom att vi gör intervallet lite större. Den generella formeln blir då i exempelvis excel:

$$MOE = TINV(0.05; Antal\ mätvärden - 1) \frac{STDAV.S(Data)}{\sqrt{Antal\ mätvärden}}$$

Om vi visar samma samband som vi såg ovan, så får vi en mer rättvisande bild av mätosäkerheten:



Figur 3: En mer rättvisande bild av mätosäkerheten med MOE.

Notera att i detta specifika fall var det bara ett enda värde som avvek, dvs $1/20$ -regeln stämde precis, om än av mer tur än skicklighet. Det är också värt att notera att om man gör olika mätningar där man har anledning att tro att den underliggande standardavvikelsen är lika så ska man så klart utnyttja detta för att minska t -faktorn.

Reflektioner

När jag presenterar denna metod för elever brukar de snabbt komma in i den. Likaså brukar de flesta tycka att formens tre delar, standardavvikelsen, $1/\sqrt{N}$ och t -faktorn alla spelar var sin naturlig roll i berättelsen om hur vi får fram mätosäkerheten. Det brukar inte ta särskilt lång tid innan eleverna ropar till varandra vad deras MOE är, och eftersom excel även tillåter att lägga in felstaplar, kan de få beräkningarna i grafisk form.

Däremot möter ibland formeln en del skepsis från lärarhåll, som menar att det inte går att undervisa gymnasieelever sådant här, ”de har ju inte lärt sig tillräckligt med matematik än, det kommer på högskolan”. Visst det är ett argument, men med logiken att allt eleverna själva inte direkt kan beräkna inte ska användas, då får man sluta med linjär regression också, för hur många av eleverna är det som kan anpassa linjer eller polynom till data med minsta kvadratmetod? Eller får de för den skull använda datorn? De vet ju inte hur assemblerkoden fungerar. Nu travestera jag här på invändningen, men min poäng är att det ofta går väldigt bra att undervisa något som eleverna inte förstår i detalj hur det uppkommit, bara man har en tydlig konceptuell förklaring.

Andra invändningar jag hört är att det ”inte hinns med att göra upprepade mätningar”. Det kan säkert ligga en del i det, men om du laborerar med en hel klass, sammanställ då alla elevernas resultat och använd dem som utgångspunkt för beräkningarna, det är en bra lärdom att visa att experimentellt arbete sällan utförs i isolering.

I sammanfattning så är min och mina kollegors erfarenhet från experimentellt arbete med gymnasieelever att det går att föra mycket intressanta diskussioner så snart vi har MOE på plats, såsom att jämföra med teori, eller de uppskattningar man själv skulle göra av felet. Tvärt emot att det känns jobbigt, så kan osäkerhetsbegreppet bli intressant när det går från en okänd okänd till känd okänd.

LITTERATUR

- Cumming, G. (2012). *Understanding the new statistics*. Taylor & Francis Group.
- Lavröd, J. *Kan gymnasieelever bedriva forskning?* Nämnaren 2018:1.
- Salsburg, D. (2002). *The lady tasting tea*. Owl Books.