Khaled Ben Latief Jemai

Modellering av normalfördelad data

Normalfördelning möter elever i många situationer, men det är inte alltid så enkelt att förstå eller att undervisa om. Här belyser författaren några didaktiska aspekter av statistikundervisning och modellering med hjälp av Gausskurvan.

Modellering av normalfördelad data är vanligt inom statistik och dataanalys. Att undervisa om normalfördelad data har flera didaktiska konsekvenser, både när det gäller att förmedla själva begreppen och att tillämpa dem på verkliga situationer. Normalfördelning är ett kraftfullt verktyg för att förstå och förklara många olika typer av data men det kan inledningsvis vara utmanande att förstå. Därför är det viktigt att ge tydliga och begripliga förklaringar med exempel från det verkliga livet.

Normalfördelning används i stor utsträckning inom en mängd olika discipliner, däribland psykologi, ekonomi och biologi. Att lära ut relevans och tillämpning av normalfördelning kan motivera eleverna att bättre förstå begreppet och se värdet i att lära sig det. Att arbeta med normalfördelad data kan utveckla kritiskt tänkande, till exempel att bedöma om normalfördelning är en lämplig modell för en viss datamängd, och förstå effekten av observationer som avviker kraftigt från genomsnittet.

Två svårigheter är värda att speciellt ge akt på när man arbetar med normalfördelning:

- *Missförstånd*
  Vissa elever kan tro att alla datamängder följer en normalfördelning, vilket naturligtvis inte är fallet. Det är viktigt att visa eleverna exempel på data som inte är normalfördelad och att förklara konsekvenserna av att anta en normalfördelning när så inte är fallet.
- *Användning av teknik*
  Användning av statistisk programvara och grafer kan hjälpa eleverna att förstå och arbeta med normalfördelad data. Men det är även viktigt att lära sig tolka resultaten rätt, och att inse begränsningar och potentiella felkällor när tekniken används.

Gausskurva

Normalfördelning, även känd som Gaussfördelning eller Gausskurva (bell curve), är en kontinuerlig sannolikhetsfördelning som är väl lämpad för att beskriva många naturfenomen. Den är uppkallad efter den tyske matematikern Carl Friedrich Gauss, som bidrog till dess utveckling. En normalfördelning kännetecknas av två parametrar: medelvärde (μ) och standardavvikelse (σ). Medelvärdet representerar fördelningens centrala tendens, medan standardavvikelsen indikerar spridningen eller variansen av data. Vid modellering av normalfördelad data är det vanligt att anta att data är oberoende och identiskt fördelad. Detta innebär att varje datapunkt är oberoende av de andra och följer samma fördelning.

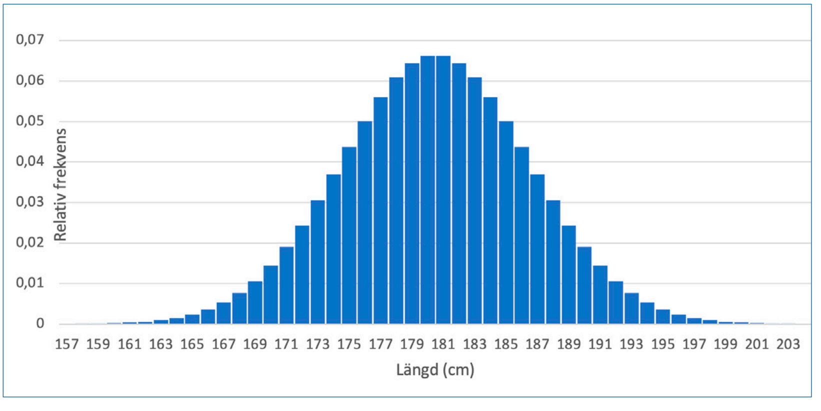
Gausskurvan definieras matematiskt av en funktion som innehåller de två parametrarna medelvärde (μ) och standardavvikelse (σ). Formeln för en normalfördelning är:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Några vanliga exempel på data som ofta antas följa en normalfördelning inkluderar exempelvis längd, vikt, intelligenskvot (IQ) och testresultat. Längd- och viktdata från en slumpmässig population antas normalt vara symmetriskt fördelad runt medelvärdet. IQ-poäng tenderar att vara normalfördelade, med medel-IQ satt till 100. Resultaten av standardiserade tester följer ofta en normalfördelning, där de flesta deltagare uppnår genomsnittliga resultat. Vidare antas felmarginaler för mätningar och uppskattningar vara normalfördelade.

Exempel med normalfördelad längd

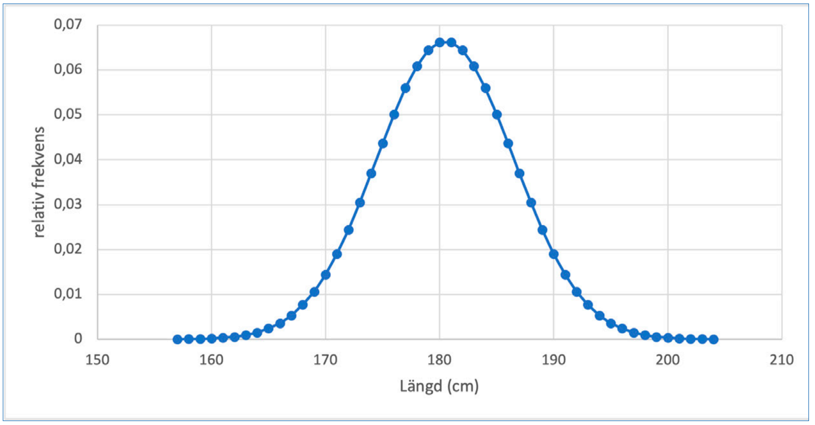
I följande exempel utgår vi från längden på 10 000 slumpmässigt utvalda män. Vi delar in männen i klasser med en klassbredd på 1 cm. Vi för in fördelningen i en tabell och använder den som underlag för att skapa ett histogram med en stapel för varje klass av data, där klasserna utgör intervall. Bredden på en stapel är lika med bredden på intervallet och arean på en stapel är lika med frekvensen av intervallet. Det är behändigt att införa såväl frekvens som relativ frekvens i tabellen. Skapar vi ett histogram utifrån den relativa frekvensen får vi att den totala arean av alla staplar är lika med 1.



Från histogram till kurva

Diskreta modeller representerar uppräkningsbar data, vanligtvis presenterade i form av histogram där varje stapel visar hur många gånger ett visst värde förekommer. Kontinuerliga modeller, å andra sidan, representerar data som kan mätas längs en kontinuerlig skala.

För att gå från en diskret modell till en kontinuerlig modell kan man föreställa sig att histogrammet "jämnas till". Istället för att ha separata staplar kan vi tänka oss en kurva med bästa möjliga passning till toppen på varje stapel. Vi får då en kurva som kan se ut på följande sätt:



Av ovanstående ser vi att histogrammet och grafen liknar varandra. Vi kan observera att vi får en symmetrisk kurva som faller mot sidorna i form av en kulle eller en klocka. Denna kurva kallas Gausskurva och är typisk för en normalfördelning.

Medellängd

I en normalfördelning kommer 68,25 % av alla datapunkter att ligga mellan plus en och minus en standardavvikelse från medelvärdet. I exemplet med männens längd kallar vi detta intervall för medellängden. För att kunna ange medellängden behöver vi beräkna både medelvärde och standardavvikelse. Medelvärdet $\overline{x}$ och standardavvikelsen σ beräknas med formlerna:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} = 180,5\,cm \qquad \sigma = \sqrt{\frac{\sum_{i=1}^{n} f_i \cdot (x_i - \overline{x})^2}{n}}$$

där x_i är värdet på observationen
f_i är observationsfrekvensen vid x_i
$\overline{x}$ är medelvärdet av frekvensfördelningen
n är det totala antalet observationer (summan av alla frekvenser).

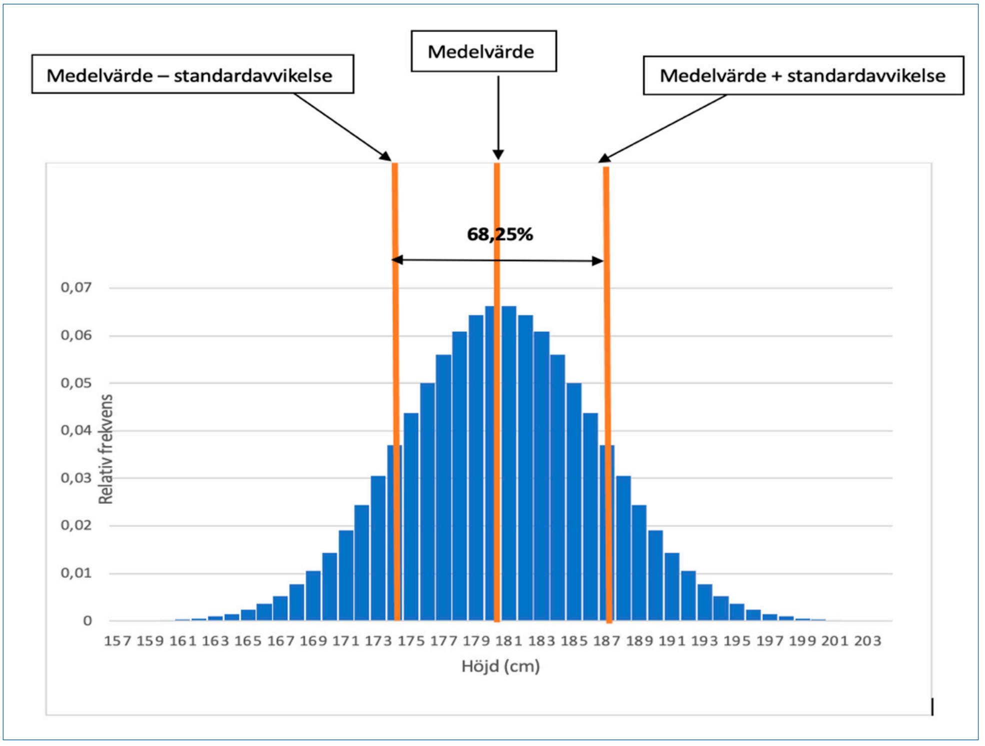
För att kunna beräkna standardavvikelse gör vi en tabell där vi fyller i de ingående värdena för varje längdintervall.

Längd (x_i)	Antal (f_i) (frekvens)	$(x_i \cdot f_i)$	$(x_i - \overline{x})$	$(x_i - \overline{x})^2$	$f_i \cdot (x_i - \overline{x})^2$
157	0	0	-23,5	552,25	0
158	1	158	-22,5	506,25	506,25
159	1	159	-21,5	462,25	462,25
160	2	320	-20,5	420,25	840,5
161	4	644	-19,5	380,25	1521
162	5	810	-18,5	342,25	1711,25
163	10	1630	-17,5	306,25	3062,5
164	15	2460	-16,5	272,25	4083,75
165	24	3960	-15,5	240,25	5766
166	36	5976	-14,5	210,25	7569
167	53	8851	-13,5	182,25	9659,25
168	77	12936	-12,5	156,25	12031,25
⋮	⋮	⋮	⋮	⋮	⋮
201	2	402	20,5	420,25	840,5
202	1	202	21,5	462,25	462,25
203	1	203	22,5	506,25	506,25
204	0	0	23,5	552,25	0
Summa	10000	1805000			360636

Ur tabellen får vi fram följande:

medelvärde	180,5
varians	36,0636
standardavvikelse (σ)	6,01
standardavvikelse + medelvärde (avrundat)	186,51
standardavvikelse – medelvärde (avrundat)	174,50

Markeras värdena för medellängden, det vill säga medelvärde plus/minus en standardavvikelse, i vårt ursprungliga histogram får vi följande bild.



Varför syssla med modellering?

Matematisk modellering kan delas in i tre kategorier: modellering som innehåll, modellering som redskap och modellering som kritik.

- *Modellering som innehåll*
  betonar att modellering integreras som en grundläggande del av matematikkursen, med syftet att utveckla specifika färdigheter för att modellera verkliga situationer och främja praktisk förståelse och användning av modellering.
- *Modellering som redskap*
  betraktar modellering som ett pedagogiskt verktyg som kan göra matematiska begrepp mer tillgängliga och begripliga för elever, där det fungerar som ett medium för att illustrera och fördjupa matematiska idéer.
- *Modellering som kritik*
  uppmuntrar till kritisk reflektion kring användning och implementering av matematik i samhället, och en djupare analys av matematiska modeller i olika samhällssammanhang, med det övergripande syftet att odla en kritisk förståelse för matematikens roll i samhället.

Sammanfattning

Den här artikeln finner sin plats i *modellering som innehåll*, där huvudmålet är att lära ut specifika färdigheter relaterade till förståelsen och användningen av statistiska begrepp, inklusive normalfördelning. Genom att utforska modellering av normalfördelad data visades betydelsen av begreppet och hur ofta det förekommer i statistiska beräkningar. Dess närvaro i en mängd olika discipliner gör det till ett viktigt verktyg för dataanalys.

Att undervisa om normalfördelning är inte utan utmaningar, särskilt när det gäller vanliga missförstånd kring dess användning och tolkning av data. Det är extra viktigt att betona och med konkreta exempel illustrera att alla datamängder inte följer en normalfördelning.

Avslutningsvis riktas en uppmuntran till lärare och elever att ytterligare utforska detta rika och mångfacetterade område, med en påminnelse om att förståelse av normalfördelning inte bara är en nyckelkunskap i statistik, utan också en utgångspunkt för en djupare förståelse för världen omkring oss.

LITTERATUR

Barbosa, J. C. (2006). Mathematical modelling in classroom: A socio-critical and discursive perspective. *ZDM, 38*, 293–301.

Hana, G. M. (2013). *Matematiske byggesteine*r. Caspar.

Johnson, N., Kotz, S. & Balakrishnan, N. (1987). Normal distributions. *Continuous univariate distributions, 1*, 156–157.

Lavröd, J. (2019). *Att bestämma osäkerhet: användning för statistiken.* Nämnaren 2019:1.