

Beskrivande statistik

19 okt. 2026, 4 min läsning

UTKAST

#matematik #statistik #metod

Kurs: S0008M

Beskrivande statistik sammanfattar ett datamaterial med några få tal och diagram: var värdena ligger, hur utspridda de är och vilken form fördelningen har. Det är det första steget i varje analys och ämnet för laboration 1. Här modelleras ingenting ännu; allt räknas direkt på de observerade värdena.

1. Lägesmått och spridningsmått

Lägesmått säger *var* värdena ligger, **spridningsmått** *hur utspridda* de är.

Lägesmått	Spridningsmått
medelvärde $\bar{x} = \frac{1}{n} \sum x_i$	standardavvikelse s
median	variationsbredd, max – min
	kvartilavstånd

Standardavvikelsen räknas med nämnaren $n - 1$, samma s som i **Punktskattning**:

$$s = \sqrt{\frac{1}{n - 1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Medianen påverkas inte av enstaka extrema värden, medelvärdet gör det. Ligger de nära varandra är materialet ungefär symmetriskt.

2. Median och kvartiler enligt Tukey

Sortera värdena. **Djupet** är hur långt in från kanten ett värde ligger i den sorterade listan.

$$\text{medianens djup} = \frac{n+1}{2}, \quad \text{kvartilernas djup} = \frac{[\text{medianens djup}] + 1}{2}$$

Nedre kvartilen Q_1 räknas från botten, övre kvartilen Q_3 från toppen. Slutar ett djup på ,5 tas medelvärde av de två värdena runt det.

$$\text{kvartilavstånd} = Q_3 - Q_1$$

Andra program räknar annorlunda

Minitab och räknaren interpolerar kvartilerna på andra sätt och kan ge något andra värden än Tukeys djup. I laboration 1 gav Minitab $Q_1 = 6,825$ mot Tukeys 6,85. Skillnaden är liten och ändrar sällan slutsatsen.

Motsvarande begrepp för en stokastisk variabel finns i [Median och kvantiler](#).

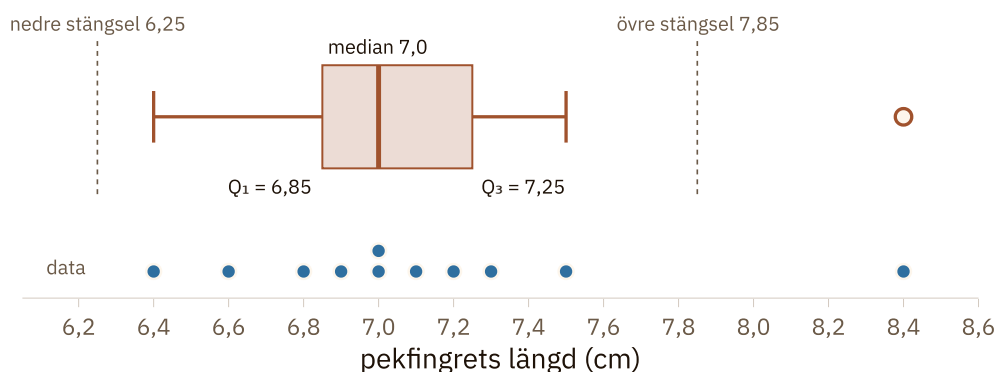
3. Stängsel, uteliggare och lådagram

Stängslen ligger 1,5 kvartilavstånd utanför lådan:

$$\text{nedre stängsel} = Q_1 - 1,5(Q_3 - Q_1), \quad \text{övre stängsel} = Q_3 + 1,5(Q_3 - Q_1)$$

Värden utanför stängslen är **uteliggare**.

I ett **lådagram** går lådan från Q_1 till Q_3 med ett streck vid medianen. Morrhåren går ut till det mest extrema värdet som fortfarande ligger *innanför* stängslen, och uteliggarna ritas som egna punkter. Själva stängslen ritas inte ut.



Figuren hör till exemplet nedan. Stängslen är utritade streckade för att visa var de ligger; i ett vanligt lådagram syns de inte.

☰ Exempel – Tukeys mått för elva värden

Elva pekfingerlängder (cm), sorterade:

6,4 6,6 6,8 6,9 7,0 7,0 7,1 7,2 7,3 7,5 8,4

Bestäm median, kvartiler, kvartilavstånd, stängsel och uteliggare, och ange var morrhåren slutar. Beräkna också $\bar{x}$ och s .

Lösning

Median. Djupet är $(11 + 1)/2 = 6$; det sjätte värdet är 7,0.

Kvartiler. Djupet är $(6 + 1)/2 = 3,5$, alltså medel av tredje och fjärde värdet från respektive ände:

$$Q_1 = \frac{6,8 + 6,9}{2} = \text{6,85}, \quad Q_3 = \frac{7,3 + 7,2}{2} = \text{7,25}, \quad Q_3 - Q_1 = 0,4$$

Stängsel. $1,5 \cdot 0,4 = 0,6$:

$$6,85 - 0,6 = \text{6,25}, \quad 7,25 + 0,6 = \text{7,85}$$

8,4 ligger ovanför övre stängslet och är en **uteliggare**. Morrhåren går till 6,4 och 7,5.

Medelvärde och standardavvikelse.

$$\bar{x} = \frac{78,2}{11} \approx \text{7,11}, \quad s \approx \text{0,53}$$

Medelvärdet ligger över medianen, dragen uppåt av uteliggaren.

På räknaren. One-Variable Statistics ger $\bar{x}$, s_x och **MedianX**. Kvartilerna **Q1X** och **Q3X** räknas med en annan metod och blir 6,8 och 7,3 här; använd Tukeys djup när uppgiften kräver det.

Ej kontrollerat mot facit – två oberoende lösningar stämmer överens.

4. Stam-bladdiagram och histogram

Ett **stam-bladdiagram** sorterar data för hand och visar formen utan att tappa några värden. För 7,3 är 7 stammen och 3 bladet. Blir det för få stammar delas varje stam på flera rader.

Ett **histogram** delar in värdena i klasser och ritar en stapel per klass. För få klasser döljer formen, för många ger ett taggigt diagram. En tumregel är ungefär $\sqrt{n}$ klasser, och klassbredden bör vara en multipel av mätnoggrannheten så att inga klasser blir tomma av avrundningsskäl.

Snedhet. En lång svans åt höger betyder att fördelningen är **högerskev** (positivt sned), en lång svans åt vänster att den är **vänsterskev**.

5. Två stickprov eller par

Samma skillnad som i inferensdelen gäller redan här. Jämförs två olika grupper, till exempel män och kvinnor, är det **två stickprov** och vi ritar ett lådagram per grupp. Mäts samma person två gånger, till exempel dominant och icke-dominant hand, är det **stickprov i par**: räkna ut differensen per person och beskriv differenserna.

Vanliga fel

- **Blanda läges- och spridningsmått.** Kvartilerna själva är lägesmått; det är *avståndet* mellan dem som mäter spridning.
- **Morrhår till stängslen.** Morrhåren slutar vid det mest extrema *observerade* värdet innanför stängslen, inte vid stängslet.
- **s_x i stället för s_x .** Stickprovets standardavvikelse har nämnaren $n - 1$.
- **Jämföra par som två grupper.** Två mätningar per person är inte oberoende; titta på differenserna.

Se även

- Median och kvantiler
- Varians och standardavvikelse
- Punktskattning
- Två stickprov
- Stickprov i par
- S0008M Lab 1 ifylld

Resurser

- [Khan Academy: Summarizing quantitative data](#)
 - [Wikipedia: Box plot](#)
-
-