

### Avsikt och matematikinnehåll

För att snabbt få överblick över ett statistiskt material, framförallt vid jämförelser mellan olika material, behöver man snabbritade figurer som framhäver viktiga egenskaper i materialet men utelämnar onödiga detaljer. *Lådagram*, eller box plot, är en modell som grafiskt åskådliggör fem viktiga storheter: det största värdet ( $X_{\max}$ ), det minsta värdet ( $X_{\min}$ ), medianen ( $Md$ ), nedre kvartilen ( $Q_1$ ) och övre kvartilen ( $Q_2$ ).

### Förkunskaper

Grundläggande kunskaper om statistiska begrepp.

### Material

Varje grupp behöver en arbetsbeskrivning, se sista sidan i detta dokument.

### Beskrivning

En framgångsrik "datadetektiv" måste på ett enkelt sätt kunna åskådliggöra ett siffermaterial. Vanligtvis görs det med tex stolpdiagram eller histogram, medelvärde och standardavvikelse. Under 70-talet lanserade John W Tukey EDA, vilket står för *exploratory data analysis*. EDA används för att söka ledtrådar och bevis i ett datamaterial.

Tukey föreslog att man grafiskt ska åskådliggöra fem storheter: det största värdet, det minsta värdet, medianen, nedre kvartilen och övre kvartilen. Ett sätt att illustrera dessa fem storheter är att använda "ihopvikning".

Antag att vi har följande 13 observationer:

53 52 54 56 50 56 50 55 50 59 49 52 50

Vi ordnar dem efter storlek

49 50 50 50 50 52 52 53 54 55 56 56 59

Sedan "viker vi av i mitten" och slutligen "viker vi upp" ändarna så att vi får bilden nedan:

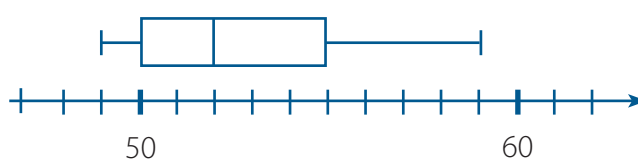
49					52					59
	50			52		53			56	
		50	50			54		56		
			50				55			
minsta		nedre		median		övre		största		
värde		kvartil				kvartil		värde		

Dessa värden kan åskådliggöras i en figur. Tukey kallar det för box plot. Översatt till svenska skulle det bli låd-diagram, vilket Andrejs Dunkels föreslog skulle dras ihop till lådagram.

Värdena från "ihopvikningen" på förra sidan kan sammanfattas:

$$X_{\min} = 49, \quad Q_1 = 50, \quad Md = 52, \quad Q_2 = 55, \quad X_{\max} = 59$$

Lådan i mitten har till uppgift att markera läget och spridningen av den mittersta hälften av observationerna. Lådans längd är *kvartilavståndet*, dvs från  $Q_1$  till  $Q_2$ . Genom att markera *medianen* får vi en känsla för hur snett materialet är. Oftast är de värden som avviker mest från "den stora massan" intressanta. Därför markeras alltid största och minsta värdet. Strecken ut från lådan kallas *morrhår*. Om de extrema värdena ligger på ett avstånd från lådan som är mer än 1,5 gånger lådans längd kallas de *uteliggare*.



## Introduktion

Innan eleverna börjar med uppgifterna på elevbladet kan de läsa igenom det och markera ord som de tycker är svåra eller inte känner igen. Hjälps åt att slå upp och diskutera orden. Samla dem sedan i en gemensam ordlista.

## Uppföljning

Efter genomförd aktivitet kan den gemensamma ordlistan som togs fram under introduktionen aktualiseras. Låt varje grupp skriva en redovisning av vad de har gjort och vad de har lärt sig.

## Variation

Denna aktivitet är ett komplement till Strävornaaktiviteten 4A4D *Stam-bladdiagram*.

## Ursprung

Aktiviteten har tidigare publicerats i NämnarenTEMA Uppslagsboken.

## Att läsa

Vännman, K. (1981). EDA – detektivarbete bland siffror med papper, penna och linjal. *Nämnaren* 1981:7, 42–52.

Emanuelsson, G. m fl (red). (1992). *Geometri och statistik*. Lund: Studentlitteratur.

Vännman, K. & Dunkels, A. (1984). *Boken om kreativ statistik med EDA*. Göteborg: Gothia.

## Några kommentarer

Naturligtvis har just 13 observationer valts för att ge en perfekt figur. Diskutera med eleverna vad som händer när det är ett annat antal observationer. Vad händer om det t ex är ett jämnt antal observationer?

Ett mer konkret sätt att "vika av i mitten" och "vika upp ändarna" är att skriva de storleksordnade observationerna jämnt fördelade på en pappersremsa och vika pappret på mitten, vika upp och sedan vika in kortändarna till mitten. Vad det handlar om är att hitta tre mittpunkter, "mitten på det hela" och sedan "mitten på de båda halvorna".

