Kvantitativ strategi
Univariat analys 2

Wieland Wermke

Sammanfattande mått: centralmått

- Beroende på skalnivån finns det olika mått, som betecknar variablernas fördelning
- **Typvärde eller modalvärde**
  - talvärdet som förekommer mest
- **Median**
  - Värde som ligger i mitten av alla observationsvärden (när man sorterar dem efter sitt talvärde, i turordning), Md delar urvalet i två jämna grupper
  - 1 4 5 **7** 9 12 14
  - Om det är en jämt antal observationer, summerar man de två i mitten och delar summan med två
  - 1 4 5 **7** **9** 12 14 15 → Md = **8**
- **Medelvärde**

Formler

- Statistiska analyser/beräkningar sammanfattas oftast i matematiska formler
- Exempel: medelvärde:
  - Def. i vanliga ord: *Summera samtliga observationers variabelvärde och dela denna summa med antalet observationer*
- Fortfarande ganska lätt, dock finns det mer komplexa exempel
  - Därför beskrivs alla statistiska operationer i formler
  - Också bra, dem är lika på alla språk!

Formler (ex. medelvärde)

- Varje observation i urvalet betecknas med **i** som kan ta ett värde mellan 1 och **n** (antalet av observationer i urvalet)
- varje observation **i** har ett bestämt värde för en viss variabel x; vi har n stycken x, en för varje i: $\mathbf{x_i}$
- Alla $\mathbf{x_i}$ ska nu summeras, det betyder: $x_1 + x_2 + x_3 + \ldots + x_n$
  - Det uttrycks med ett sumtecken: grekiskt stor Sigma

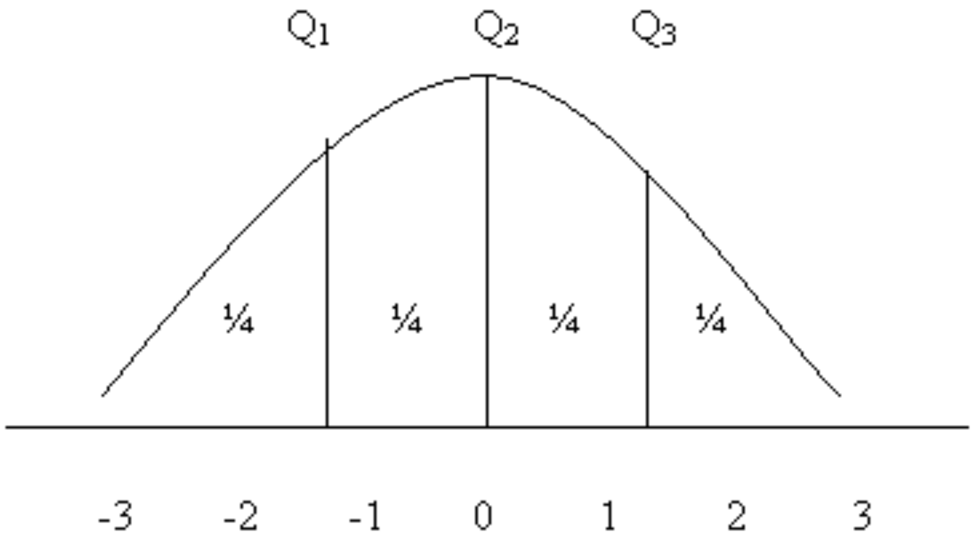
$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

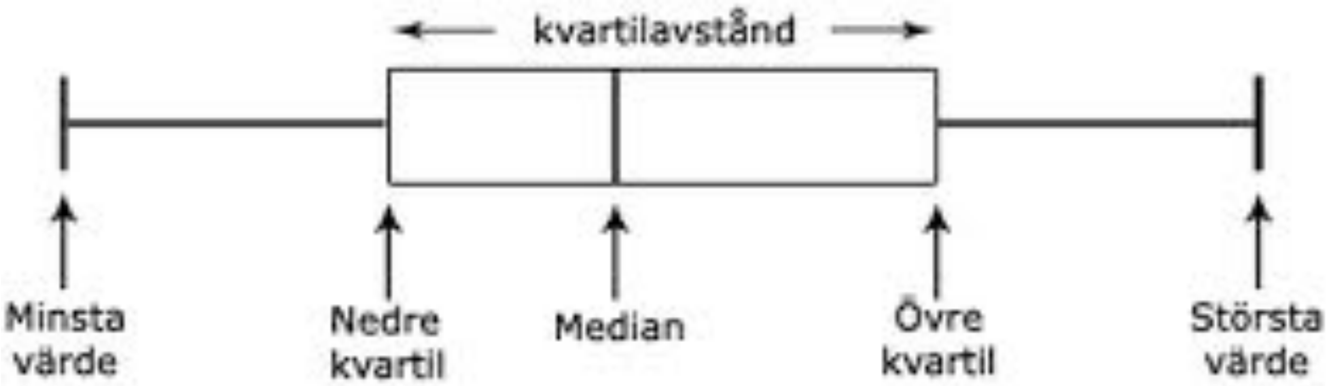
- Medelvärde av variabeln Summera alla värde x för observationen i som ligger mellan 1 och n och delar genom n
- 1 4 5 **7** **9** 12 14 15 = 8,375

Sammanfattande mått: Spridningsmått

- Viktigt är också variabelns spridning
- **Utbredning** (*range*) → avståndet mellan minsta och största värdet
- Kvartilsavstånd:
  - man indelar fördelningen i fyra lika delar efter att ha sorterat dem efter sitt talvärde (som man också gör för medianen): Q1=första 25 procentgräns, Q2=median, Q3=75 procent
  - kvartilsavstånd= Q3-Q1
  - Man får utesluta outliers etc.

Kvartilsavstånd

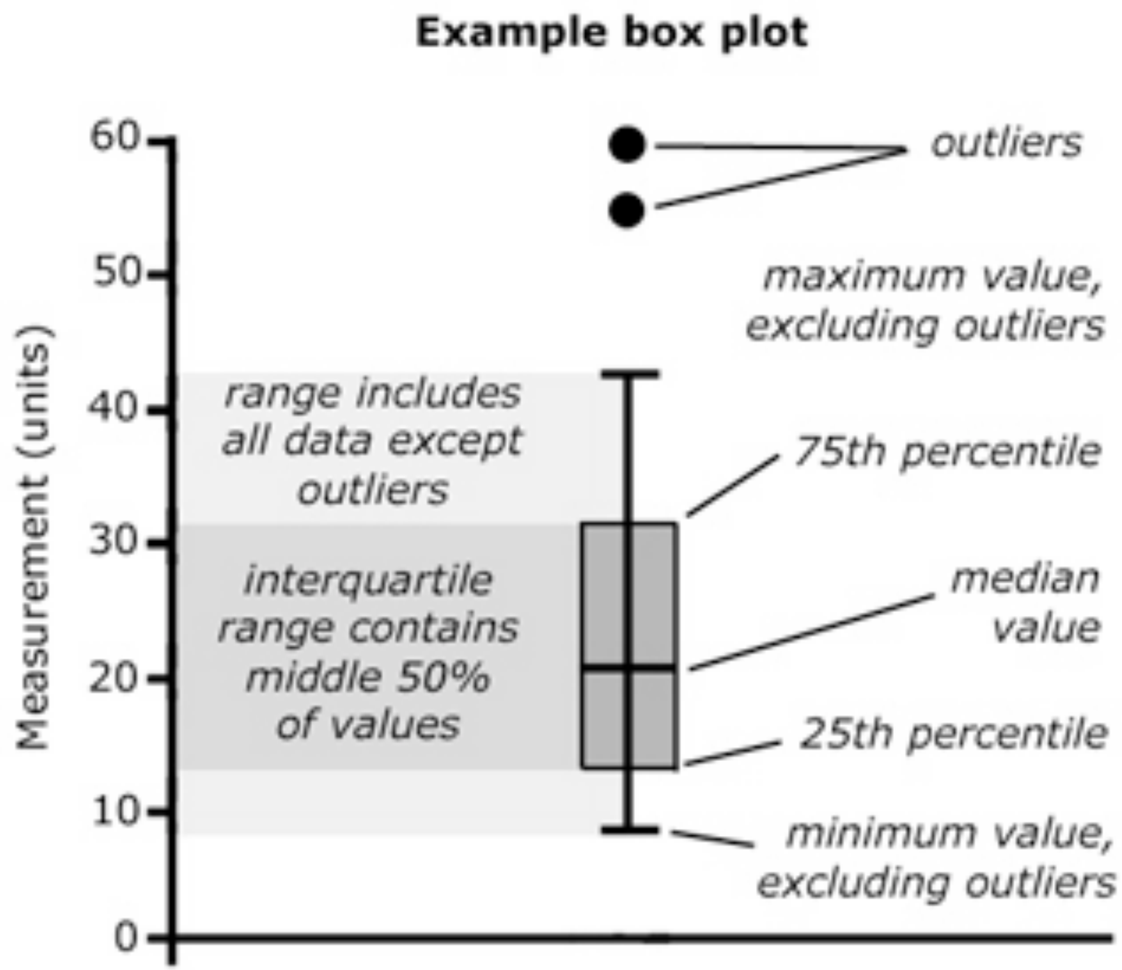




Varför är spridningsmått så viktiga?

- de uttrycker kvalitet av utsagor över medelvärdet
  - en hög standardavvikelse ifrågasätter en utsaga över genomsnittet
- För att kunna genomföra vissa analyser är det ibland bättre att utesluta vissa *outliers* (som kan analyseras extra eller tyda på fel svar)
- Bloxplots är en bra instrument för att se sådana

+ Boxplotdiagram



Värden som ligger längre ifrån boxen än 1,5 gånger avståndet mellan de yttre kvartilerna betraktas som utliggare outliers).

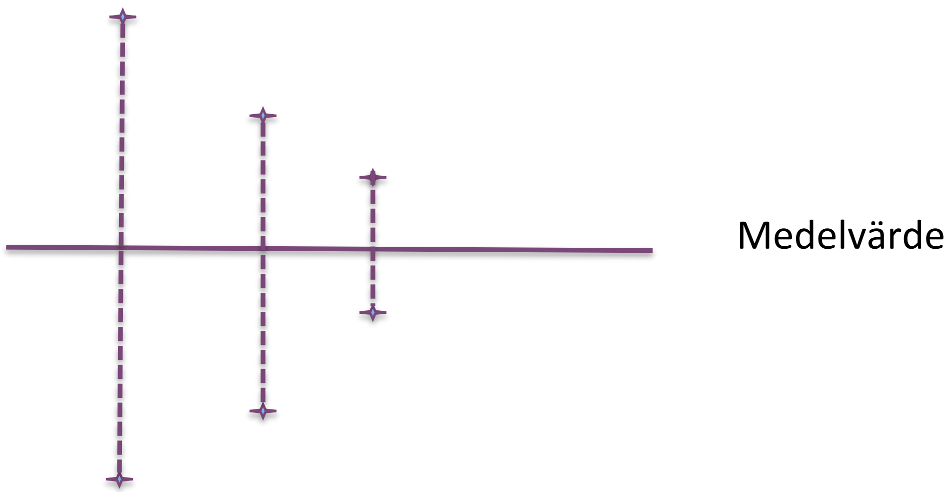
Varians och standardavvikelse

- Det absolut vanligaste förkommande spridningsmåttet är **standardavvikelsen**
- Första steget är att beräkna den så kallade **variansen**
- Båda är nyckelmått i nästan alla kvantitativa analyser
- Medelvärde och standardavvikelse ska anges i alla studier som använder sig av intervallskaladata!

+ Varians

- Hur långt ifrån variabelns medelvärdet befinner sig varje observation?

- Vi vill ha ett mått som uttrycker det **genomsnittliga avståndet från medelvärdet**, som kan visa hur stor variansen är
  - ju mindre variansen, ju större koncentration runt medelvärdet



Varians

- Problem: det går inte att bara summera, eftersom det bara blir noll (annars hade vi inget medelvärde)
  - → kvadrera differens mellan medelvärdet och det observerade värdet (minus försvinner), sedan summera
- Problem: när man delar genom antalet, där blir variansen för liten
  - dra av **ett** från antalet innan du delar summan

Varians

- Variansen visar därmed
  - ju större skillnaderna mellan medelvärdet och observationerna är, ju större är variansen
  - ju mer observationer ju mindre blir variansen (ju säkrare är medelvärdet)

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

Standardavvikelse

- Eftersom variansen inte har samma dimension/enhet (t.ex. cm blir cm^2) som medelvärdet (pga. kvadraten), drar vi kvadratroten, så har både centralmått (medelvärde) samt spridningsmått (standardavvikelse) samma mått igen

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

Standardavvikelse, exempel

Observation	Värde		
1	1	$(1 - 2) = -1$	$(-1)^2 = 1$
2	1	$(1 - 2) = -1$	$(-1)^2 = 1$
3	1	$(1 - 2) = -1$	$(-1)^2 = 1$
4	3	$(3 - 2) = 1$	$(1)^2 = 1$
5	4	$(4 - 2) = 2$	$(2)^2 = 4$
Summa	10 **Medelvärde** 10/5 = 2		8 **Variansen** 8/(5-1) = 8/4 = 2 **Standardavvikelsen** Roten ur 2 = 1,4

Vilka mått användas för vilken skalnivå?

	CENTRALMÅTT			SPRIDNINGSMÅTT		
	Typvärde (modalvärde)	Median	Medel-värde	Variations-vidd	Kvartil-avvikelse	Standard-avvikelse/ MAD
Nominalskala	+*)					
Ordinalskala	+	+	?	+	+	?
Intervall-/kvotskala skev fördelning	(+)	+	(+)	(+)	+	(+)
Intervall-/kvotskala normal fördelning	(+)	(+)	+	(+)	(+)	+

***) ange även typvärdets andel (t ex i % = modalprocent) av alla – plus det totala antalet enheter (n)**

(+) = Tillåtet med sällan förstahandsalternativet
? = Tänkbart om kategorierna är fler och om de på ett meningsfullt sätt kan åsättas talvärden