

Kvantitativ forskning

C2

Viktiga begrepp och univariat
analys

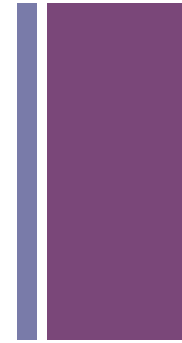
+ Delkursen mål

- Ni har grundläggande kunskaper över statistiska analyser (univariat, bivariat)
- Ni kan använda olika programvaror för olika statistiska analyser
- Ni ska kunna förstå/analysera/presentera olika sorts statistiska presentationsformer (olika statistiska mått, diagram, tabeller)



+ Statistik

- **Statistik** är en gren inom tillämpad matematik som sysslar med insamling, utvärdering, analys och presentation av data eller information
- Ordet "statistik" kommer från latin 'statisticus', med betydelsen 'statsman', 'politiker'
- → Försöker förbättra våra gissningar för att kunna styra bättre
- → statistiska samband är dock först matematiska samband som måste sedan översättas i sin betydelse för varje samband

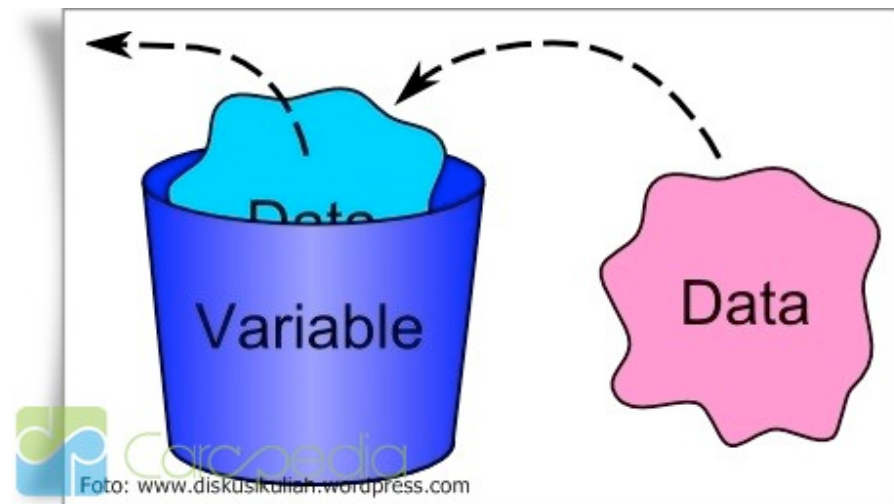




Genomgång: Variabler och variabeltyper

+ Variabel

- Är ett namngivet objekt som används för att representera ett okänt värde (platshållare), till exempel ett reellt tal.
- **Variabler** används i öppna utsagor. De kan anses stå i motsats till **konstanter** som är oföränderliga
 - $x * y = 10$
 - Ex. 15 åriga pojkars TV beteende. Vad är variabel och vad är konstant?



+ Variabel

- Definierade “**fack**” som fylls med innehåll (värden, *observationer*) från vissa *observationsenheter* (t.ex. grundskolelever)
 - variabel *ålder* (möjliga värde: 0-20)
 - *betyg* (möjliga värde: 1-5, A-F)
 - *attityd* (möjliga värde: Gillar inte (1) – Gillar mycket (5))
 - *utbildning* (ingen utbildning (1), gymnasium (2), högskolan (3))



+ Olika variabeltyper

- Olika observationer har olika karaktär, som måste beaktas → **observationsdata ser olika ut:**
- Man kan arbeta med **olika sorters skalor** som i sig innehåller **mer eller mindre** information samt kan användas för olika sorters analyser



+ Vad passar i de olika facken?
Variabler kan vara

Kvalitativa variabler

kan indelas i **kategorier**
Ex: kön, partitillhörighet

Kvantitativa variabler

(**kan räknas**)
Ex: pris, meritvärde

Diskreta variabler

skalstegen är fasta,
begränsad antal möjligheter
Ex: attityd, meritvärde

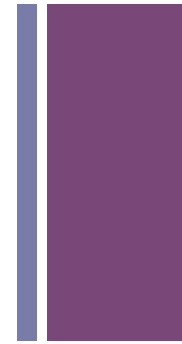
Kontinuerliga variabler

alla skalsteg är möjliga
Ex: pris, vikt, inkomst



+ Olika variabeltyper

- **Nominalskala:** frekvens, endast klassificering, antal fall per grupp, kan endast räknas
- **Ordinalskala:** rangordning, skalstegen olika långa, kan rangordnas <1.,2.,3.,4.>
- **Intervall:** kan mätas, kan adderas, ingen nollpunkt, jämna skalsteg (equidistans) (+/-)
- **kvotskala:** kan mätas, lika långa skalsteg, har absolut nollpunkt (+/-, 0)





Man kan alltid "downgrade", men inte inte "upgrade"

	<u>Nominalnivå</u>	<u>Ordinalnivå</u>	<u>Intervallnivå</u>	<u>Kvotnivå</u>
		rangordning	jämna steg	absolut nollpunkt
	kategorier (omvandlat till siffror)		"riktiga tal"	
Exempel	kön, etnisk bakgrund, typer	utbildning, sociekonomisk bakgrund, (betyg)	intelligenstest, testvärde, (attityd), (betyg)	vikt, längd, ålder, pris
Möjliga relationer	lika/olika	<i>utöver</i> : mindre/bättre; färre/mer; sämre/bättre	<i>utöver</i> : rum, distans mellan två entiteter (intervall) är kända och alltid lika långa; addera	<i>utöver</i> : relationer mellan två entiteter; dubbelt så mycket/halv så mycket
Möjliga statistiska mått	antal; frekvens (procent av helhet); Typvärde	<i>utöver</i> median; kvartilsavstånd	<i>utöver</i> : (aritmetrisk) medelvärde, standardavvikelse	<i>utöver</i> : (geometriskt) medelvärde

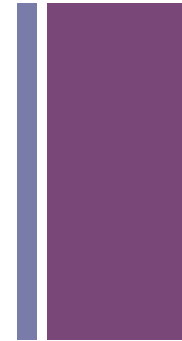


Genomgång: Olika mått

+ Sammanfattande mått: centralmått

- Beroende på skalnivån (nominal, ordinal, intervall) finns det olika mått, som betecknar variablernas fördelning
- Typvärde eller modalvärde (för nominal)
- Talvärdet som förekommer oftast
 - Ex. familjbakgrund: Tjänstmanfamilj: 120; **Arbetarfamilj: 110**
 - Högsta examen:
Högstadiet: 24; Studentexamen: 40; Kandidatexamen: 56

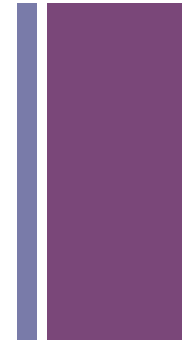
+ Median



- **För ordinalskalnivå**
- 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**

+ Exkurs: Formler

- Statistiska analyser/beräkningar sammanfattas oftast i matematiska formler
 - sparar tid, men måste också kunna förstås
 - lika i alla språk
- Exempel: medelvärde (centralmått för intervallnivå)?
- Def. i vanliga ord: **Summera samtliga observationers värde och dela denna summa med antalet observationer**



+ 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: **x_i**
- **Alla x_i ska nu summeras, det betyder: $x_1 + x_2 + x_3 + \dots + x_n$**
 - Det uttrycks med ett summatecken: grekiskt stor Sigma

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

- Medelvärde av variabeln Summera alla värde x för observationen i som ligger mellan 1 och n och delar genom n

- $1\ 4\ 5\ \mathbf{7}\ \mathbf{9}\ 12\ 14\ 15 = 8,375$

+ 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**: **x_i**
- Alla **x_i** ska nu summeras, det betyder: $x_1 + x_2 + x_3 + \dots + x_n$
 - Det uttrycks med ett sumtecken: grekiskt stor Sigma

$$\bar{x} = \frac{\sum x_i}{n}$$

■ Me
för
del
■ 1 4

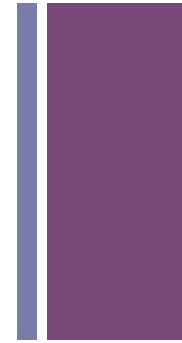
Förenklat:

$$\bar{x} = \frac{\sum x_i}{n}$$

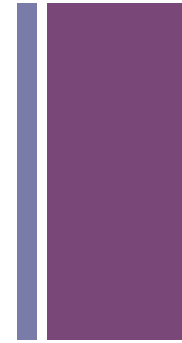
de **x**
n och

+ Sammanfattande mått: Spridningsmått

- Viktigt är också variabelns spridning
- **Utbredning** (*range*) → avståndet mellan minsta och största värdet
- 1 2 3 34 43 12: Range=42
- 4 99 1 2 3 4 101: Range= 100

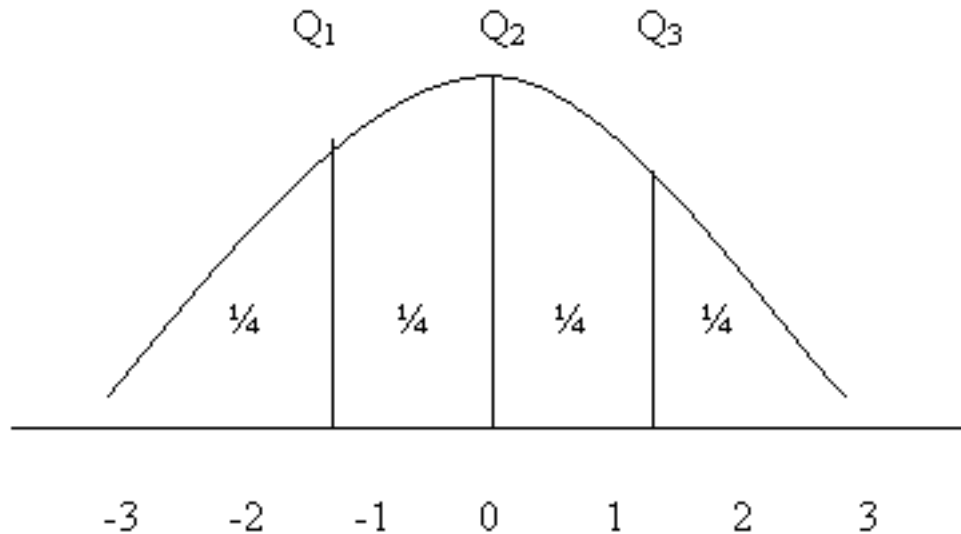


+ Spridningsmått: Kvartilsavstånd

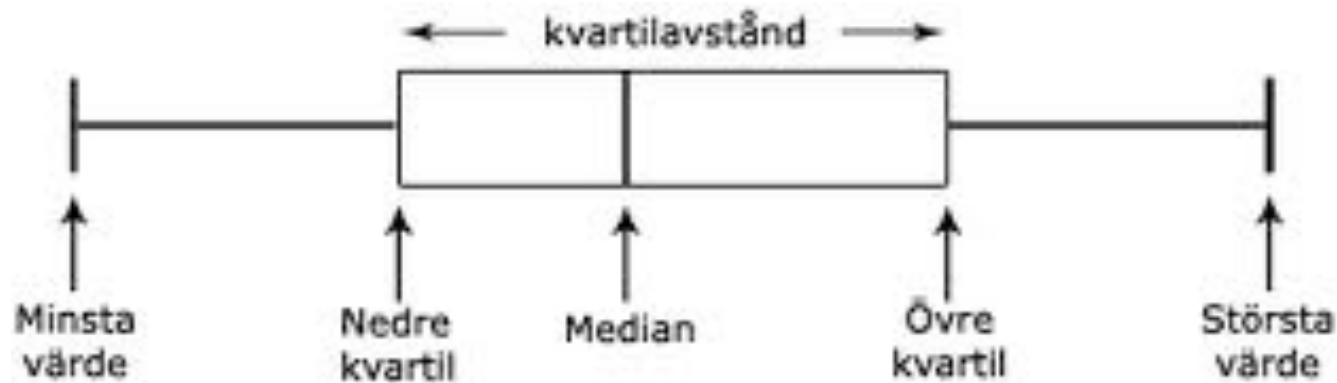


- Man delar fördelningen in i **fyra lika delar** efter att ha sorterat dem **efter sitt talvärde** (som man också gör för medianen):
- Q1=första 25 procentgräns (nedre kvartil), Q2=median, Q3=75 procent (övre kvartil)
- kvartilsavstånd= $Q3 - Q1$
- 1 2 3 4 5 6 7 8 9 10 11 12
 - Q1: 3 Q2: 6,5 Q3: 9 = 3 4 5 6 7 8 9

+ Kvartilsavstånd

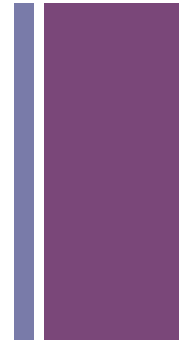


Vad är de centrala värde
Vad är kanske outliers?
I synnerhet viktigt, om
man har kontinuerliga
variabler



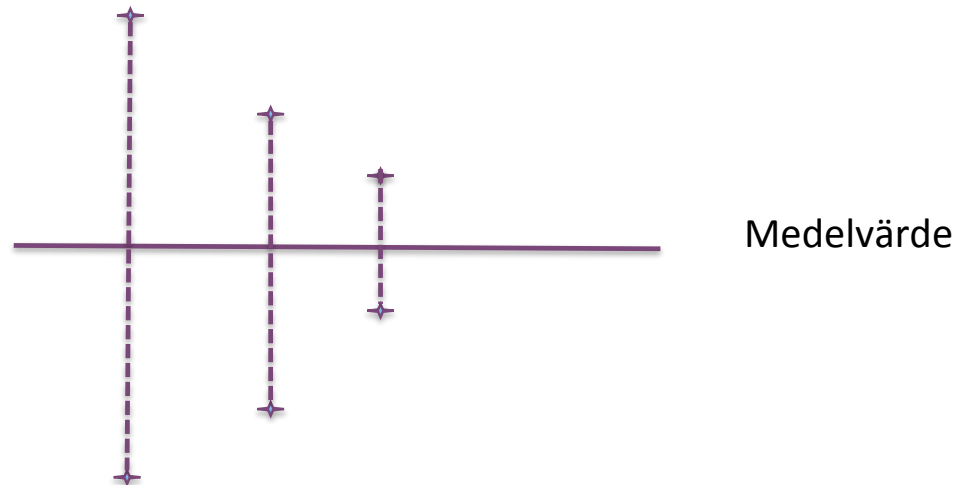
+ Varians och standardavvikelse

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



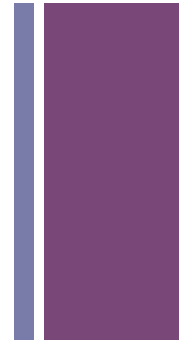
+ Varians

- Hur långt ifrån variabelns medelvärde 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 alla värde
- Problem: när man delar genom antalet, där blir variansen för liten
 - dra av **ett** från antalet innan du delar summan (bara acceptera detta och lita på matematikerna)



+ 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

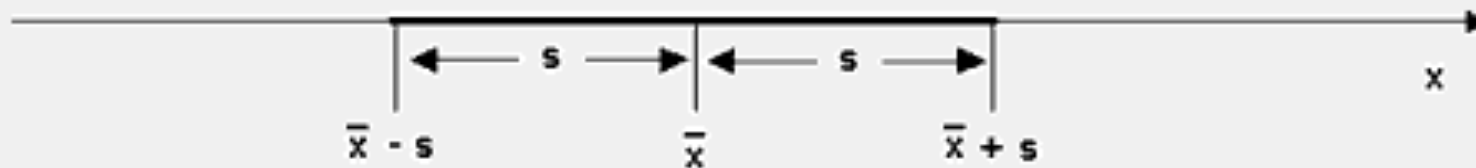
- Eftersom variansen inte har samma dimension/enhet (t.ex. cm blir cm²) som medelvärdet (pga. kvadraten), drar vi kvadratroten, så har både centralmått (medelvärde) samt spridningsmått (standardavvikelse) samma mått igen

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

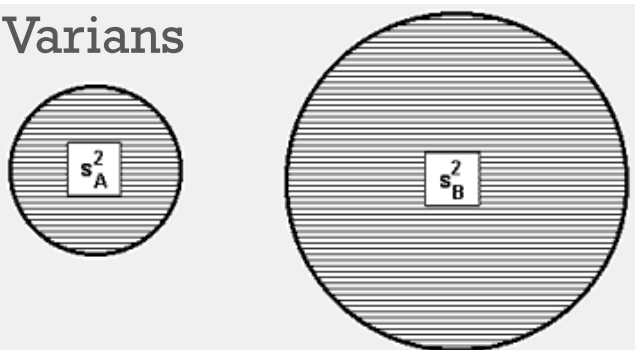
+ Hur kan spridningsmått tolkas

- Enklast: säger hur homogen ett urval, och hur tillförlitligt ett medelvärde är (mycket viktig för kontinuerliga variabler). Det behövs naturligtvis något att jämföra med

Standardavvikelse



Varians



<https://www.youtube.com/watch?v=pGfwj4GrUIA>

+ Standardavvikelse

- Med medelvärde de två viktigaste mått för statistiska analyser på intervallnivå!
- $\ominus \pm sd$ = Medelvärde plus/minus standardavvikelse
- Antagningspoäng PIVA programmet $\ominus 15 \pm 1$
 - = Medelvärde 15 med en genomsnittlig spridning av 1 poäng (14-16)
- Det betyder naturligtvis också att vissa värde kan ligga mer än 1 sd från medelvärdet (t.ex. 2 poäng)
- Standardavvikelse och medelvärde utgör normalfördelningen



+ Olika skalnivåer, olika mått

Centralmått				Spridningsmått		
	Typvärde	Median	Medelvärde	Variationsvidd	Kvartilsavstånd	Standardavvikelse
Nominal	ja	nej	nej	ja, dock meningslös	nej	nej
Ordinal	ja	ja	(nej), dock vissa skalor kan omvandlas till intervall: betyg, attityd	ja	ja	(nej), se medelvärde
Intervall/ kvot	ja, används dock mycket sällan	ja, används om urvalet är mycket litet	ja	ja, används dock mycket sällan	ja, används dock mycket sällan	ja