

F6

Standardiseringsmetoder
Etiska regler och lagregler

Förra gången (F4-F5)...

- Lägesmått: aritmetiskt medelvärde (minst intervall), median (minst ordinal), typvärde (alla nivåer)
 - När vi vill beskriva *tyngdpunkten* i ett material
 - Observera – andel/proportion är ett specialfall av medelvärde
- Spridningsmått: varians (minst intervall), standardavvikelse (minst intervall), kvartilavstånd/avvikelse (minst ordinal), variationsvidd (minst intervall), variationskoefficient (minst intervall)
 - När vi vill beskriva *variationen* i ett material
- Vi ska kunna beräkna alla dessa mått samt veta när de är lämpliga

Förra gången (F4-F5), diagram

- Kvalitativa variabler
 - Stapeldiagram
 - Cirkeldiagram
- Kvantitativa variabler
 - Diskreta med fåtal värden
 - Stapeldiagram
 - Stolpdiaqram
 - Trappstegsdiagram (för kumulativ frekvens)
 - Kontinuerliga eller diskreta med många värden
 - Histogram
 - Frekvenspolygon
 - Summapolygon (för kumulativ frekvens)
 - Stambladdiagram (ingen klassindelning)
 - Lådagram (ingen klassindelning)
- Vi ska veta hur dessa diagram konstrueras, dvs kunna avläsa diagrammen.
 - Ex. Kan man utifrån diagram X veta hur många observationer som har värdet x?

Stambladdiagram

Stem-and-Leaf Display: längd

Stem-and-leaf of längd N = 27

Leaf Unit = 1,0

1	15	5
3	16	44
12	16	555888999
(11)	17	01112233444
4	17	58
2	18	02

Lådagram

Descriptive Statistics: längd

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
längd	27	0	170,33	1,06	5,53	155,00	168,00	171,00	174,00

Variable	Maximum
längd	182,00

Md=171

Q1=168

Q3=174

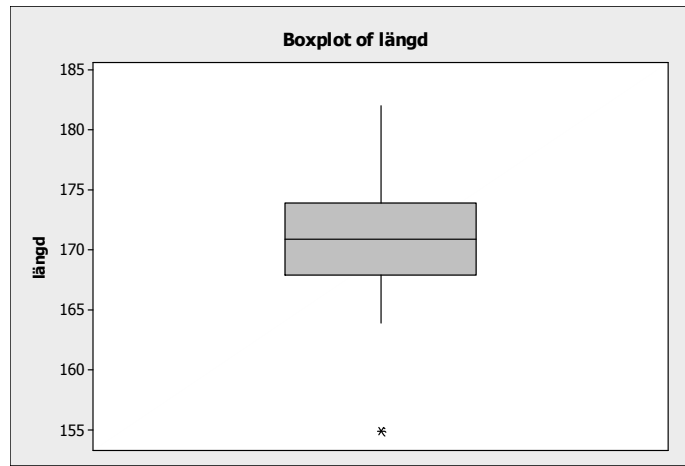
$D=1.5 \times (174-168)=9$

$D(\min)=168-9=159$

$D(\max)=174+9=183$

$X(\min)=155$

$X(\max)=182$



Standardiseringsmetoder

- Jämförelser mellan medelvärden för olika populationer kan vara missvisande
 - De anställda på företag A har högre medellön än de anställda på företag B
 - Företag A har fler med ledande ställning än företag A – spelar det någon roll?
 - Stad A har högre dödlighet än stad B
 - Stad A har fler äldre än stad B – spelar det någon roll?
- Standardvägning
 - Vi kan jämföra medelvärden i olika populationer sedan hänsyn tagits till kända faktorer
 - Ex. skillnader i andel med ledande ställning på ett företag, skillnader i åldersfördelning

Standardvägning

- Standardpopulationsmetoden
 - Vad skulle medelvärdet vara om fördelningen (ex. åldersfördelningen) såg ut som i en viss "standardpopulation"?
- Kapacitetsmetoden
 - En jämförelse mellan ett faktiskt tal och ett förväntat tal

Standardpopulationsmetoden

Personalkategori	AB LL		GAB	
	Antal personer	Medellön (tusen kr)	Antal personer	Medellön (tusen kr)
Ledande	5	38,6	13	35,5
Självständigt arbete	46	25,8	67	25,6
Kontorspersonal	75	19,2	31	18,8
Biträdespersonal	84	15,5	24	15,3

Se K & W sid 85

$$\bar{x}_{ABLL} = \frac{5 \times 38,6 + 46 \times 25,8 + 75 \times 19,2 + 84 \times 15,5}{210} = 19,628$$

$$\bar{x}_{GAB} = \frac{13 \times 35,5 + 67 \times 25,6 + 31 \times 18,8 + 24 \times 15,35}{135} = 23,161$$

$$\bar{x}_{GAB, ABLL} = \frac{5 \times 35,5 + 46 \times 25,6 + 75 \times 18,8 + 84 \times 15,3}{210} = 19,287$$

Standardpopulationsmetoden

- Kvoten mellan medelvärdet för AB LL och det standardiserade medelvärdet för GAB blir:

$$\frac{19628}{19287} = 1,02$$

- Lönenivån för AB LL ligger 2% över lönenivån för GAB när man tar hänsyn till att företagen har olika andelar i ledande ställning etc.
- AB LL har här använts som standardpopulation

Kapacitetsmetoden

- Område A: 1628 födda barn, 26506 fruktsamma kvinnor
 - Fruksamhet: $1628/26506=0,0614$
- Område B: 5480 födda barn, 91446 fruktsamma kvinnor
 - Fruksamhet: $5480/91446=0,0599$
- Enligt ovan är fruktsamheten högre i A än i B
 - Men det är olika åldersfördelningar i område A och B
 - Dock vet vi inte fruktsamheten per åldersgrupp i A och B

Kapacitetsmetoden

- Vi kan beräkna det *förväntade* antalet födda barn för båda områdena givet deras åldersfördelningar
- Förväntade antal födda barn i område A (se K & W sid 87):
 $3028 \times 0,0118 + 4976 \times 0,0902 + 4865 \times 0,1515 + 7362 \times 0,0529 + 6275 \times 0,0039 = 1636$
 - Kapacitetsindex: $1628/1636 = 99,5\%$
- Förväntade antal födda barn i område B:
 $10389 \times 0,0118 + 15983 \times 0,0902 + 17263 \times 0,1515 + 22470 \times 0,0529 + 25341 \times 0,0039 = 5467$
 - Kapacitetsindex: $5480/5467 = 100,2\%$

Ålder (år)	Antal kvinnor område A	Antal kvinnor område B	Fruktsamhet i riket (promille)
15-19	3028	10389	11,8
20-24	4976	15983	90,2
25-29	4865	17263	151,5
30-39	7362	22470	52,9
40-49	6275	25341	3,9

Kapacitetsmetoden

- I område A föds 0,5% färre barn än förväntat
- I område B föds 0,2% fler barn än förväntat.
 - Fruktsamheten är något högre i B än A om man tar hänsyn till åldersfördelning

Etiska regler och lagregler: skyddet för statistikuppgifter

- Statistiksekretessen
 - Ett undantag från offentlighetsprincipen
 - Omfattar uppgifter som samlas in av myndigheter för statistikproduktion
- Personuppgiftslagen (PuL)
 - Omfattar både myndigheter och privata företag
 - Personuppgifter får behandlas om registrerade lämnat informerat samtycke
 - Undantag för statistiska ändamål
 - Syfte: skydda mot kränkning av personliga integriteten
 - Datainspektionen: www.datainspektionen.se

Statistiska undersökningar

- Informerat samtycke
- För att minska bortfall: informera om lagligt skydd enligt PuL och SekrL, samt att inga enskilda svar kan identifieras
- Aidentifiering (efter bortfall påmints)

Etiska regler i samhällsforskning

- Vägning mellan forskningskrav samt individskyddskrav
- Humanistisk-samhällsvetenskapliga forskningsrådet (Se KD sid 358-359)
 - Ska informeras
 - Lämna samtycke utan påtryckningar
 - Informeras om möjlighet till att avbryta
 - Inget beroendeförhållande mellan forskare/uppgiftslämnare
 - Obehöriga ska inte kunna få tag på uppgifter.
Tystnadsplikt
 - Uppgifter får inte användas av myndigheter för påtryckning