 ISBN 978-91-639-6675-(

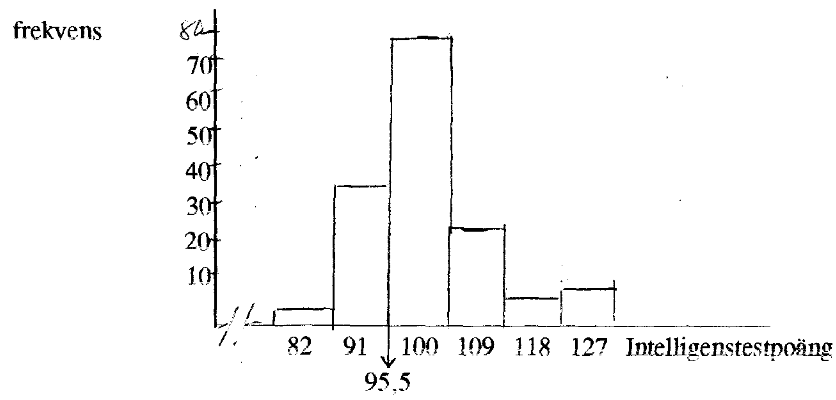
"Det finns inga precisa regler för hur klassindelningen skall göras, men man kan ha följande två konventioner i minnet:
(i) Antalet klasser bör vara mellan 5 och 15.
(ii) Man bör göra klassindelningen så att klassmitten blir ett heltal för enkelhets skull." (s.19). För överskådlighetens skull och för att man lättare kan jämföra t.ex. olika testomgångar, är det lämpligt att ha lika stora klasser, dvs. samma poängintervall för alla klasserna.

Grafisk framställning av klassindelat, approximativt kontinuerlig fördelning

Ni har ovan räknat ut klassmitten för alla klasser. Klassmitten för den första klassen:

78-86 poäng -	(klass 1) är 82
klassmitten	(klass 2) är 91
klassmitten	(klass 3) är 100
klassmitten	(klass 4) är 109
klassmitten	(klass 5) är 118
klassmitten	(klass 6) är 127.

"Man börjar lämpligen med att göra en skala för den klassindelade variabeln, där man har lika många skalsteg som klasser. På dessa skalsteg anges klassmitterna. Därefter ritar man symmetriskt omkring klassmitterna staplar som anger frekvensen inom varje klass." (s. 19). Se figur 2 nedan: Diagrammet kallas för ett *stapeldiagram* eller vanligen för ett *histogram*.



"Egentligen är det ytan på staplarna som anger frekvensen, men då staplarna har lika stor bas, har detta ingen betydelse. I detta fall är nämligen staplarnas höjd direkt proportionell mot ytan." (s. 19).
"Vi kan lägga märke till att gränserna mellan klasserna aldrig är ett heltal, utan ett heltal +0,5. Så t ex är gränsen mellan den andra klassen med klassmitt 91 och den tredje klassen med klassmitt 100, lika med 95,5. (Se figuren). Klassgränsen är sålunda ett värde som man aldrig erhåller i praktiken. Detta är en fördel då man därmed entydigt kan avgöra till vilken klass ett visst värde tillhör." (s.20).

 ISBN 978-91-639-6675-0

Klassindelning, ÖVNING

Uppgift 1:

a) Gör en frekvenstabell med absolut och kumulativ frekvens av följande testresultat:
3, 5, 9, 6, 4, 5, 7, 8, 2, 3, 2, 3, 6, 5, 4, 8, 6 ,7, 5, 1, 7, 5, 6, 4 ,3 ,6, 9
b) Dela in materialet i tre klasser. För en frekvenstabell. Bestäm klassgränser och klassmitten.

Uppgift 2:
Följande resultat på ett test har erhållits:

33, 37, 20, 38, 31, 23, 51, 40, 30, 43, 45, 38, 28, 48, 39, 37, 28, 47, 41, 17, 40, 48, 30, 38, 40, 27, 50, 34, 44, 42, 41

a) Gör en klassindelning av materialet och redovisa denna i en frekvenstabell.
b) Minska klassgränserna så att du får fler klasser. Hur påverkas fördelningen?

Uppgift 3:
Följande testpoäng erhölls för ett stickprov på 30 personer:
2, 11, 6, 6, 18, 1, 9, 2, 2, 15,
8, 16, 12, 11, 17, 3, 3, 5, 3, 7,
11, 9, 3, 16, 16, 16, 4, 9, 5, 7,

a) Gör en lämplig klassindelning av materialet och beskriv det i tabell och diagram.
b) Bestäm typvärdet, medianen och medelvärdet för datan
c) Uppvisar fördelningen någon snedhet?
d) Bestäm variationsvidden, variansen och standardavvikelsen för datan.

 ISBN 978-91-639-6675-0 7

Klassindelning : FACIT *till sid 22*

Uppgift 1: a) Gör en frekvenstabell med absolut och kumulativ frekvens av följande testresultat:
b) klassindela materialet i tre klasser. Bestäm klassmitten.
3, 5, 9, 6, 4, 5, 7, 8, 2, 3, 2, 3, 6, 5, 4, 8, 6 ,7, 5, 1, 7, 5, 6, 4 ,3 ,6, 9

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

X	f	kf
1	1	1
2	2	3
3	4	7
4	3	10
5	5	15
6	5	20
7	3	23
8	2	25
9	2	27

X	f	klassgränserna	klassmitten
(1-3)	7	0,5-3,5	2
(4-6)	13	3,5-6,5	5
(7-9)	7	6,5-9,5	8

Uppgift 2:Följande resultat på ett test har erhållits:
a) Gör en klassindelning av materialet och redovisa denna i en frekvenstabell.
b) Minska klassgränserna med 3 enheter så att du får fler klasser. Hur påverkas fördelningen?

(15-24)	17, 20, 23,
(25-34)	27, 28, 28, 30, 30, 31, 33, 34,
(35-44)	37, 37, 38, 38, 38, 39, 40, 40, 40, 41, 41, 42, 43, 44,
(45-54)	45, 47, 48, 48, 50, 51,

Ovan och nedan, två exempel på klassindelning.

(15-19)	17,
(20-24)	20, 23,
(25-29)	27, 28, 28,
(30-34)	30, 30, 31, 33, 34,
(35-39)	37, 37, 38, 38, 38, 39,
(40-44)	40, 40, 40, 41, 41, 42, 43, 44,
(45-49)	45, 47, 48, 48,
(50-54)	50, 51,

Uppgift 3: Följande testpoäng erhölls för ett stickprov på 30 personer:2, 11, 6, 6, 18, 1, 9, 2, 2, 15, 8, 16, 12, 11, 17, 3, 3, 5, 3, 7, 11, 9, 3, 16, 16, 16, 4, 9, 5, 7, ["överslag": 18/3= 6]

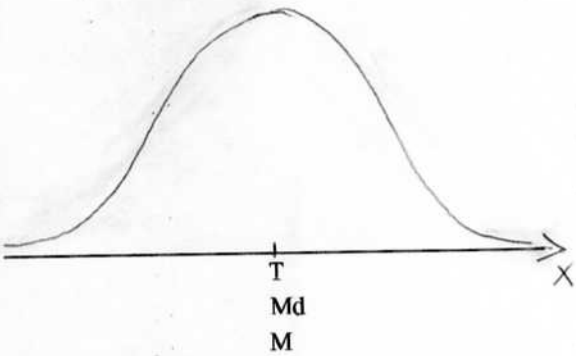
Klass	så här räknas dem	f	f/n
(1-6)	1, 2,2,2, 3, 3,3,3, 4, 5, 5, 6,6,	13	0.43
(7-12)	7,7, 8, 9,9,9, 11,11,11, 12,	10	0.33
(13-19)	15, 16, 16, 16, 16, 17, 18,	7	0.23

b) Bestäm typvärdet, T=16, medianen Md =7,5 och medelvärdet för datan, M= 8,43
c) Uppvisar fördelningen någon snedhet? positivt sned
d) Bestäm variationsvidden, variansen och standardavvikelsen för datan.R=17, $s^2 = 28,39$, s=5,33

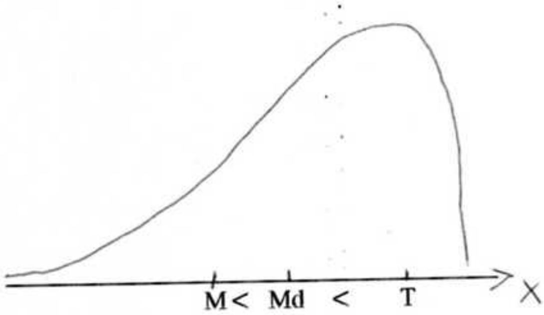
 ISBN 978-91-639-6675-0

Kvantitativa variabler
Fördelning och lägesmått

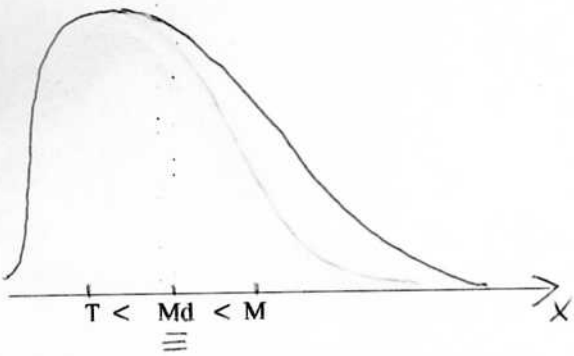
Vi har redan talat om fördelningar. Genom att presentera data i frekvenstabeller har vi fått fram frekvensfördelningen. Om man ritar fördelningen av observationerna som en kurva, kan kurvan variera i form. Nedan beskrivs fyra olika fördelningsformer och under varje kurva finns (på x-axeln som är fördelningens "bas") beteckningarna; T som står för *typvärdet* i fördelningen, Md som står för *medianen* i fördelningen och M som står för *medelvärdet* i fördelningen. Lägesmåtten T, Md och M är *statistiska storheter som sammanfattar någon egenskap hos variabelvärdena.* Lägesmåtten ger oss sammanfattningsmått på fördelningens centraltendens. Efter några ord om fördelningarna, skall vi titta på hur man räknar ut lägesmåtten.



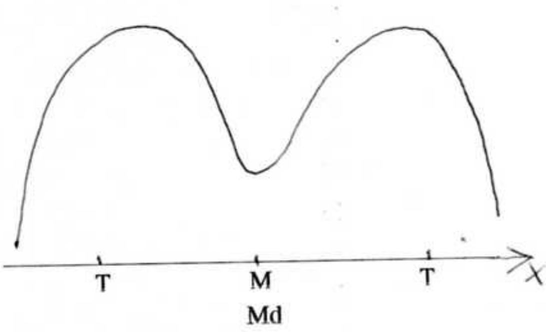
Normalfördelningen är symmetrisk och klockformad.
I normalfördelningen sammanfaller lägesmåtten M, Md och T.
I statistisk beräkning är normalfördelningen idealfallet.



En negativt sned fördelning har en "svans" som sträcker sig mot låga och negativa värden (efter nollpunkten origo i koordinatsystemet). Lägesmåtten hamnar på olika "platser".



I positivt sned fördelning gäller relationen M>Md>T.



En fördelning som har två toppar kallas för *bimodal*. I exemplet ovan är den bimodala fördelningen symmetrisk med två typvärden (T).

OBS! Vid snedfördelning, M missvisande!

 ISBN 978-91-639-6675-0

Lägesmått

Medelvärde

Frekvenstabeller och diagram ger oss en översikt över det statistiska materialet. Oftast vill man dock veta lite mera om materialet. Det vanligaste måttet på materialets centraltendens är det *aritmetiska medelvärdet* som är summan av alla observationer dividerad med antalet observationer. Medelvärdet förkortas M i stickprovet [ibland ser man också ett litet m, och med den grekiska bokstaven μ (my) i populationen]. Vi har lärt oss att storleken på stickprovet betecknas n, vår undersökningsvariabel betecknas X och den grekiska bokstaven stort sigma betyder "summan av". När vi vet dessa beteckningar kan vi skriva den aritmetiska formeln för medelvärdet:

$$M = \frac{\Sigma X}{n}$$

Exempel 1: Vi har fått 5 poäng, 7 poäng, 9 poäng och 10 poäng i ett test. Det betyder att n i det här fallet är 4 testresultat (n=4). Vi summerar testpoängen 5 + 7 + 9 + 10 och dividerar den erhållna summan med n=4 för att få fram medelvärdet på de i alla fyra test erhållna poängen.

$$M = \frac{5+7+9+10}{4} \qquad M = \frac{31}{4} \qquad M = 7,75$$

Svar: Medelvärdet för våra erhållna poäng i de fyra testen är sålunda 7,75.

Exempel 2:

Låt oss räkna ut medelvärdet för betygen i en stickprovsundersökning på barns avgångsbetyg i svenska; n = 50, (50 barn) med hjälp av frekvenstabellen nedan. Formeln för beräkning av medelvärdet i det här fallet är

$$M = \frac{\Sigma fX}{n}$$

Ur frekvenstabellen nedan kan vi läsa betygen och frekvensen barn som fått betygen. Vi kan räkna ut medelvärdet enklast genom att multiplicera varje betygsvärde med frekvensen (antalet barn som fått betyget) och summera siffrorna. Detta har gjorts till höger om frekvenstabellen.

$$\left(M = \frac{1\times10 + 2\times13 + 3\times17 + 4\times5 + 5\times5}{50}\right)$$

X betyg i svenska	f	
1	10	1 x 10 = 10.
2	13	2 x 13 = 26
3	17	3 x 17 = 51
4	5	4 x 5 = 20
5	5	5 x 5 = 25
		Σ = 132

M=132/50=2,64

Svar: Medelvärdet för barnens betyg i denna stickprovsundersökning på 50 barn är sålunda 2,64, dvs. M=2,64

 ISBN 978-91-639-6675-0

Lägesmått fortsättning,

Typvärde

Typvärdet (T) är det variabelvärde som har den största frekvensen. I vårt exempel beträffande barnens betyg på föregående sida, där vi erhöll medelvärdet 2,64, kan vi i frekvenstabellen utläsa typvärdet som är 3 eftersom flest antal barn har erhållit siffran 3 i betyg.
Har vi klassindelat material kan klassmitten i den klass som har den högsta frekvensen representera också typvärdet.

Median

Medianen (Md) delar ett rangordnat material mitt itu. Medianen är alltså det mittersta värdet i en fördelning. Hälften av observationerna kommer att ligga nedanför medianvärdet. Man delar helt enkelt fördelningen i två lika stora delar.

Har vi en variabel X med udda antal värden: 3, 6, 8, 13, 20, 35 och 39 är medianen (Md) det mittersta värdet, dvs 13.
Har vi en variabel X med jämnt antal värden: 12, 14, 15, 16, 18, 21 har vi inget entydigt mittvärde. Man brukar i detta fall definiera medianen genom att summera de två mittersta värdena och dela med 2. Medianen blir då (15 + 16) / 2. Md = 15,5.

Har vi klassindelat material kan vi endast uppskatta medianen. (Man kan bestämma medianen i klassindelat material som klassmitten i den klass som har det kumulativa frekvensvärdet närmast över halva antalet observationer, men detta är av underordnat intresse här).

ÖVNINGAR:

1. Räkna ut medelvärdet (M) för följande data på variabel X:

3,	4,	2,	5,
8,	5,	2,	6,
7,	2,	3,	3,
5,	3,	4,	3,

2. Uppvisar fördelningen någon snedhet?

3. Vad är medianen för fördelningen?

4. Vad är typvärdet för fördelningen?

(Facit: Fråga 1: M= ~~[illegible]~~ 4,06; Fråga 2. Ja, det är (närmast) en positivt sned fördelning; Fråga 3: Md=3,5; Fråga 4: T=3)

 ISBN 978-91-639-6675-0

Spridningsmått - variationsmått

Lägesmåtten ger oss ganska torftig information om inte den kompletteras med någon form av mått på hur spridda observationerna är i förhållande till lägesmåttet. Man räknar då ut den så kallade spridningen eller variationen. *I princip gäller att ju högre siffervärde ett variationsmått har, desto mer utspridda ligger observationerna.*

Variationsvidd eller range

Det kanske enklaste sättet att beskriva observationernas spridning är att beteckna *variationsvidden.* Variationsvidd kallas också för range (från engelskan) och den betecknas med ett stort R. Range anger det lägsta och det högsta värdet inom variationsområdet och definieras som skillnaden mellan det högsta observerade värdet och det lägsta observerade värdet, men då och då ser man i vetenskapliga artiklar att man - istället för att faktiskt räkna ut skillnaden - anger till exempel: "observationernas range ligger mellan 1 - 5 poäng." Variationsvidden brukar man ange som ett komplementärt spridningsmått.

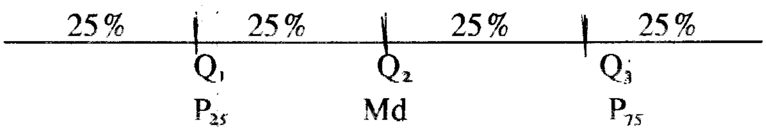
$$R = X_{max} - X_{min}$$

Kvartilavvikelsen

När man använder sig av medianen som lägesmått, brukar man använda sig av kvartilavvikelsen som spridningsmått. Kvartilavvikelsen anger hur långt uppåt - respektive nedåt - man måste gå från medianen för att finna de mittersta 50 procenten av observationerna. Kvartilavvikelsen definieras sålunda *såsom hälften av det område på variabeln inom vilket 50 procenten av fördelningens observationer ligger.*

Fördelningen brukar delas in i 4 kvartiler, som var och en innehåller 25 % av observationerna. Vissa läroböcker använder sig av Q_1 Q_2 och Q_3 för att på ett enkelt sätt beteckna gränserna mellan kvartilerna, medan andra läroböcker benämner kvartilen eller *percentilen* som det också kallas, P_{25}, och P_{75}. Denna beteckning används därför att 25 % av observationerna faller i den första percentilen nerifrån räknat som i P_{25}, och 75 % av observationerna nerifrån räknat (och 25 % av observationerna uppifrån räknat) markerar den övre kvartilens eller percentilens, P_{75}:s läge.

Som man lätt kan se från figuren nedan, motsvarar P-värdet i mitten av observationernas fördelning (dvs. P_{50} eller Q_2), medianens värde (Md).



Figur Q: *Illustrerar kvartilavvikelsens percentiler.*

 ISBN 978-91-639-6675-0

Hur man räknar ut kvartilavvikelsen.

Kvartilavvikelsen brukar man definiera som avståndet mellan P_{25} och P_{75} delat med två. Formeln skrivs så här:

$$Q = \frac{P_{75} - P_{25}}{2}$$

eller något förenklat:

$$\left(Q = \frac{Q_3 - Q_1}{2} \right)$$

Först bestäms läget av den övre kvartilen eller 75:e percentilen, P_{75}, som anger var gränser går mellan de översta 25 procenterna av observationerna och de nedersta 75 procenten. Därefter fastställs läget av den undre kvartilen P_{25} som anger gränser mellan de nedersta 25 procenten av observationerna och de översta 75 procenten. Kvartilavvikelsen är hälften av skillnaden mellan dessa båda värden.

Medelavvikelsen

Ibland kan det vara behändigt att använda sig av *medelavvikelsen*, dvs. de observerade värdenas avvikelser från medelvärdet. Medelavvikelsen definieras $|X - M|$ där strecken om båda sidorna betyder absolutbeloppet av differensen, dvs. att vi skall bortse från tecknet för differensen X - M. Då medelavvikelsen har stora matematiska nackdelar behandlar vi den inte mer ingående.

Varians och standardavvikelse

När man använt sig av medelvärdet som lägesmått, räknas *standardavvikelsen* ut som spridningsmått. Det är det viktigaste variationsmåttet och den definieras som *den genomsnittliga avvikelsen från medelvärdet*. Formeln för standardavvikelsen (s) är :

$$s = \sqrt{\frac{\sum (X - M)^2}{n - 1}}$$

När man räknar ut standardavvikelsen börjar man med att definiera *variansen.* Variansen är den *genomsnittliga kvadrerade avvikelsen från medelvärdet.* För att räkna ut variansen kvadrerar man först observationernas avvikelser från medelvärdet: dvs. $(X - M)^2$

 ISBN 978-91-639-6675-0

Dessa avvikelser summeras och summan divideras med n - 1 för att erhålla variansen som betecknas s^2

$$s^2 = \frac{\Sigma(X-M)^2}{n-1}$$

Som ni säkert upptäckt, är standardavvikelsen (s) sålunda kvadratroten ur variansen (s^2) såsom formeln överst på sidan också visar. Sålunda kan man också skriva:

$$s = \sqrt{s^2}$$

Någon kanske undrar varför man dividerar med n - 1 och inte enbart med n. Det hänger samman med statistisk inferensteori. Ni kommer ihåg när vi diskuterade medelvärdet att det betecknas M i stickprovet och μ i **populationen.** På samma sätt finns det en varians (s^2) i stickprovet, och en varians som betecknas σ^2 (lilla **sigma i kvadrat) i populationen. Denna varians måste** uppskattas och för att på ett så korrekt sätt som möjligt **kunna uppskatta den,** använder vi oss av n - 1. Då förklaringen här är bristfällig, hänvisas den intresserade **läsaren till de många läroböckerna i ämnet för vidare studier.**

Om vi utgår ifrån en frekvenstabell när vi räknar ut variansen, kan vi skriva formeln:

$$s^2 = \frac{\Sigma f(X-M)^2}{n-1}$$

Standardavvikelsen i ovan fallet, får vi genom att dra kvadratroten ur variansen (s^2).

Exempel.

Vi har i en undersökning erhållit testresultaten 6 poäng, 7 poäng, 8 poäng, 6 poäng, och 5 poöng. Medelvärdet för våra poäng är sålunda: (6+7+8+6+5)/5 = 32/5 = M=6,4

För att räkna ut variansen subtraherar vi medelvärdet från varje x-värde enligt nedan:

1. poäng-Medelvärde (X-M)	2. När vi erhållit X-M, kvadrerar vi det för att erhålla enbart positiva värden $(X-M)^2$
6-6,4 = -.4	.16
7-6,4 = .6	.36
8-6,4 = 1,6	2,56
6-6,4 = -.4	.16
5-6,4 = -1,4	1,96

$\Sigma(X-M)^2$ = 3. $(X-M)^2$ summeras = 5,2, och summan divideras med n-1, dvs. 5-1=4; 5,2/4 = 1,3

s^2 =1,3

s är kvadratroten ur s^2, dvs. $s = \sqrt{s^2}$

s = 1,14

 ISBN 978-91-639-6675-0

ÖVNINGSUPPGIFTER
Range, kvartilavvikelse och median

1. Vad är Md och vad är kvartilavvikelse (Q) i följande observationsvärden?
9, 7, 9, 2, 4, 9, 5, 8, 8, 8, 5, 6, 7, 8, 8, 8, 8, 8, 9, 9
Rangordna observationerna enligt mallen nedan:

25 %		25 %		25 %		25%
2, 4, 5, 5, 6,	|	7, 7, 8, 8, 8,	|	8, 8, 8, 8 ,8,	|	9, 9, 9, 9, 9,
	Q_1				Q_3	

Antalet observationer: 20
Md = (8+8)/2 = 16/2 = 8. Markera ut medianen Md ovan.
25 % av 20 observationer = 5 observationer. Se på de rangordnade observationerna. Summera femte och sjätte observationen och dividera summan med 2. Det ger det ungefärliga Q_1 värdet 6,5.
75 % av observationerna är 15 observationer, 8+9=17/2 = ungefärligt Q_3 =8,5. Nu kan du räkna ut ett ungefärligt Q enligt formeln

$Q = \frac{Q_3 - Q_1}{2}$ $\left(\text{eller } \frac{P_{75} - P_{25}}{2}\right)$ Svar: Q = ____________

2. Om vi har fått resultaten X beträffande betyg i svenska, enligt frekvenstabellen nedan, vad är X-variabelns variationsvidd? Vilken bokstav betecknar variationsvidd?

X	f
1	10
2	13
3	17
4	5
5	5

M=2,64

Svar: ____________

3. Om vi har testresultaten 2, 2, 2, 2, 2, 2, 2, 2, 2, 6, 2, 2, 2, 2, 2, 2, och 20, vad är de observerade värdens range?

Svar: ____________

4. Varför kan range, eller variationsvidd ge en missvisande bild av fördelningen? (Leta i uppgift 3 ovan efter ledtråd)

5. Vad är medianen för X-variabeln som antar värdena 2, 6, 4, 8, 15, 17, 14, 19. 12. 11. 6. och 20?

 ISBN 978-91-639-6675-0

Övningsexempel, varians och standardavvikelse.

hämtad ur Roner Dochan & Calsson : Statistik – en introduktion

1. Räkna ut medelvärdet, variansen och standardavvikelsen för följande data.

X	f	X x f
36	4	
37	4	
38	2	
39	5	
40	2	
41	3	
	n=	$\sum$ fX =

Frekvenstabellen bredvid till vänster behövs för att räkna ut medelvärdet. Fyll i siffrorna!

Svar: $M = \frac{\sum fX}{n}$ $M =$

X	(X-M)	$(X-M)^2$
36		
36		
36		
36		
37		
37		
37		
37		
38		
38		
39		
39		
39		
39		
39		
40		
40		
41		
41		
41		

$\sum (X-M)^2 =$

Varians och standardavvikelse kan räknas på två sätt. Till vänster $s = \sqrt{\frac{\sum (X-M)^2}{n-1}}$ och

längst ner, räknar man efter formeln $s \sqrt{\frac{\sum f(X-M)^2}{n-1}}$

eller:

X	f	(X-M)	$(X-M)^2(\cdot f) =$	$f(X-M)^2$
36				
37				
38				
39				
40				
41				
			$\sum$	

$\sum f(X-M)^2 =$

$s^2 =$

$s = \sqrt{s^2} =$

$s =$

 ISBN 978-91-639-6675-0 F

Övningsexempel, varians och standardavvikelse. FACIT

Lösningen visas på två olika sätt! nedan.

1. Räkna ut medelvärdet, variansen och standardavvikelsen för följande data.

X	f	Xxf
36	4	144
37	4	148
38	2	76
39	5	195
40	2	80
41	3	123
n=	20	Σ 766

$$M = \frac{\Sigma fX}{n} = \quad = 38{,}3$$

(X-M)= (X - 38,3) i detta exempel.

X	(X-M)	$(X\text{-}M)^2$
36	-2,3	5,29
36	-2,3	5,29
36	-2,3	5,29
36	-2,3	5,29
37	-1,3	1,69
37	-1,3	1,69
37	-1,3	1,69
37	-1,3	1,69
38	-0,3	0,09
38	-0,3	0,09
39	0,7	0,49
39	0,7	0,49
39	0,7	0,49
39	0,7	0,49
39	0,7	0,49
40	1,7	2,89
40	1,7	2,89
41	2,7	7,29
41	2,7	7,29
41	2,7	7,29
		Σ 58,2

$$s^2 = \frac{\Sigma(X-M)^2}{n-1}$$

eller

$$s^2 = \frac{\Sigma f(X-M)^2}{n-1}.$$

$$\Sigma(X-M)^2 = 58{,}2$$

$$s^2 = \frac{58.2}{19} = 3.06$$

$$s = \sqrt{3.06} = 1.75$$

X	f	(X-M)	$(X\text{-}M)^2$	$f(X\text{-}M)^2$
36	4	-2,3	5,29	21,16
37	4	-1,3	1,69	6,76
38	2	-0,3	0,09	0,18
39	5	0,7	0,49	2,45
40	2	1,7	2,89	5,78
41	3	2,7	7,29	21,87
				Σ 58,2

$$\Sigma f(X\text{-}M)^2 = 58{,}2$$

$$s^2 = \frac{58.2}{19} = 3{,}06$$

$$s = \sqrt{s^2} = 1.75$$

$$s = 1{,}75$$

 ISBN 978-91-639-6675-0

ÖVNINGSUPPGIFTER dec 2005

I en barnpsykologisk undersökning tillfrågades 15 barn hur många lekkamrater de hade. Barnen svarade 5 lekkamrater, respektive 3, 1, 4, 4, 7, 2, 6, 4, 4, 2, 7, 2, 3 och 4 lekkamrater.

1. Beskriv materialet i en tabell med frekvens och relativ frekvens.

2. Beskriv materialet i en diagram (och vad heter diagrammet?)

3. Vad karaktäriserar typvärdet och vad är typvärdet för de erhållna svaren?

4. Vad är står Md för och hur skall man definiera det?

5. Vad är medianen för de erhållna svaren?

6. Hur räknar man ut medelvärdet?

7. Vad är medelvärdet för de erhållna svaren i ovan undersökning beträffande barnens (antal) lekkamrater? Är medelvärdet ett bra mått i det här fallet?

8. Vad är *range* för de erhållna värden?

9. Vilken typ av variabel är det frågan om i undersökningen? Kontinuerlig variabel eller diskret variabel (diskret variabel är bara ett annat ord för diskontinuerlig variabel) och vad innebär det att en variabel är diskret?

10. Vad innbär det att en variabel är kontinuerlig? Ge exempel på kontinuerliga variabler.

11. Vilka fyra huvudtyper av skalor finns det och vad är karaktäristiskt för skalorna?

 ISBN 978-91-639-6675-0

Statistisk beskrivning av en variabel; lägesmått, spridningsmått - sammanfattning

Låt oss anta att vi har gjort ett antal observationer på en variabel X.

(Observationerna ordnas lämpligen först i storleksordning om detta låter sig göras.)
1. En frekvenstabell ger oss en fingervisning om fördelningen av variabelvärden. Variabler med många värden och stor variationsvidd bör för överskådlighetens skull, delas in i klasser. Vid klassindelning vinner man i överskådlighet, men förlorar viss möjlighet att utläsa data; t.ex. "drunknar" enskilda värden i klassen.

3. En grafisk beskrivning av data, t.ex. i form av stolpdiagram för diskreta variabler och histogram (stapeldiagram) för (approximativt) kontinuerliga variabler, ger överskådlighet.

4. Fördelningens centraltendens karaktäriseras vanligen genom att man beräknar medelvärdet, men i vissa fall är det bättre att använda sig av medianen och/eller typvärdet.

5. Har vi använt oss av medelvärdet, beskrivs fördelningens variation med standardavvikelsen. Man kan också ange variationsvidden (range). Men om vi endast använt oss av medianen som mått på centraltendens, skall vi presentera spridningen med hjälp av kvartilavvikelsen.

Normering

När vi gör en undersökning av något slag, får vi sällan "användbara" poäng om vi skall jämföra resultaten mellan två olika variabler. De poäng vi erhåller när vi samlar in data kallas för ***råpoäng***. För att vi skall kunna jämföra resultaten mellan olika testskalor eller mätningar, måste skalorna ha liknande, kända egenskaper. Därför ***trasformeras*** skalorna på något sätt för att vi skall kunna arbeta vidare med dem. Kanske vill vi göra jämförelser mellan olika variabler - eller kanske addera två skalor? För att vi skall kunna göra jämförelser mellan olika mätningar eller addera skalor, måste skalorna ha samma spridning; annars väger de testpoäng som har störst spridning mest och vi får missvisande, oanvändbar data. Vi omvandlar sålunda variablernas observationsvärden till "samma sorts skala" som är bekvämare att arbeta med.
Vid transformeringar förändras en variabels medelvärde och standardavvikelse. *Normering* till Z är det vanligaste tillvägagångssättet. När vi transformerar vår variabel (X) till den *normerade variabeln Z som svarar mot X*, betyder det att vi kommer att få det kända medelvärdet 0, och standardavvikelsen 1 på skalan. Formeln för Z-normering är

$$Z = \frac{X - M}{s}$$

Z anger hur mycket observationsvärden på variabeln X avviker från medelvärdet uttryckt i standardavvikelser.

 ISBN 978-91-639-6675-0

Exempel på normering av variabelvärden

Låt oss säga att vi har observationsvärden på variabeln X och variabeln Y. Variabeln Y kan t.ex. bestå av ett antal individers poäng i en psykologisk test och variabeln X av individers betyg. Vi börjar med variabeln X och normerar den till Z_x variabel.

Anta att medelvärdet för råpoängen för variabel X är M = 2,6 och standardavvikelsen s = 1,43.

$$Z_x = \frac{X-M}{s}$$

[i vårt exempel - se nedan: $Z_x = \frac{X-2.6}{1.43}$]

1. Först räknar vi ut råavvikelsepoängen X-M som framgår av tabellen nedan. Råavvikelsepoängen talar om hur mycket, i råpoäng räknat, över eller under medelvärdet, som ett visst värde ligger. Nu är skalan normerad beträffande skalans nollpunkt, men vi måste också normera skalstegens längd.

2. Vi normerar skalstegens längd genom att tänka oss skalstegens längd lika med standardavvikelsen. Medelvärdet är noll, och den råavvikelsepoäng som ligger en standardavvikelseenhet ovanför medelvärdet är lika med värdet 1. Denna "transformering" av råavvikelsepoäng till *standardpoäng (standardiserade poäng)*, sker alltså genom att *dividera råavvikelsepoängen med standardavvikelsen i fördelningen*. Dessa poäng som man erhåller betecknas med Z.

Tabell över normering

råpoäng X	f	råavvikelsepoäng X-M	standardpoäng Z=(X-M)/s	omvandling med hjälp av en konstant Z=[10(X-M)/s]+50 *	
0	1	-2,6	-1,83	-18,3 + 50 =	31,7
1	1	-1,6	-1,12	-11,2 + 50 osv.	38,8
2	2	-0,6	-0,42		45,8
3	4	0,4	0,28		52,8
4	1	1,4	0,98		59,8
5	1	2,4	1,68		66,8

Poängen i tabellen ovan är alltså normerade så att de har medelvärdet 0 och standardavvikelsen 1. Standardpoängen betecknas med z. Råpoången 5, blir t.ex. till z-poäng 1,68 så här:

$$z_{x5} = \frac{5-2.6}{1.43} = \frac{2.4}{1.43} \approx 1.68$$

Det betyder att råpoängen 5 är lika med z poängen 1,68. Poängresultatet ligger på ett avstånd 1,68 gånger standardavvikelsen från medelvärdet (ungefär).

(* [I kolumnen satt i parentees längst till höger i tabellen ser man ett förfarande som man kan använda sig av för att slippa negativa värden i skalan och minska antalet decimaler. Man multiplicerar värden med en lämplig konstant, ex. 10, och adderar alla värden med en lämplig konstant, t.ex. 50, för att få en fördelning med - i det här fallet medelvärdet 50 och standradavvikelsen 10.])

 ISBN 978-91-639-6675-0

Nu är det din tur att öva dig:
Z-transformering av Y variabeln

Variabeln X på föregående sidan bestod av individers betyg. Variabeln Y kan t.ex. bestå av ett antal individers poäng i en psykologisk test. Din uppgift är att transformera Y-variabeln till z_y

Du minns formeln för normering från föregående sida: $Z_y = \frac{Y - M}{S}$

Följande poängvärden på Y har erhållits: 2, 0, 1, 3, 1, 2, 3, 5, 4, 3,

1. Fyll i frekvenserna i tabellen nedan.
2. Räkna ut medelvärdet.
3. Räkna ut standardavvikelsen.
4. Räkna ut råavvikelsepoängen Y-M.
5. Räkna ut standardpoängen (Y-M)/s.

Y	f	Yxf	Y-M	$(y-M)^2$	z= (Y-M)/s
0	1				
1	2				
2	2				
3	3				
4	1				
5	1				
Σ	n = 10	$\Sigma(Y \times f) =$		$\Sigma(y-M)^2$	

$M = \frac{\Sigma fY}{n} = \_\_\_ =$

$s^2 = \frac{\Sigma(Y-M)^2}{n-1} = \_\_\_ =$

$s = \sqrt{s^2} =$

 ISBN 978-91-639-6675-0

F

Nu är det din tur att öva dig:
Z-transformering av Y variabeln (FACIT)

Variabeln X på föregående sidan bestod av individers betyg. Variabeln Y kan t.ex. bestå av ett antal individers poäng i en psykologisk test. Din uppgift är att transformera Y-variabeln till z_y

Du minns formeln för normering från föregående sida: $Z = \frac{Y - M}{s}$

Följande poängvärden på Y har erhållits: 2, 0, 1, 3, 1, 2, 3, 5, 4, 3,

1. Fyll i frekvenserna i tabellen nedan.
2. Räkna ut medelvärdet.
3. Räkna ut standardavvikelsen.
4. Räkna ut råavvikelsepoängen Y-M.
5. Räkna ut standardpoängen (Y-M)/s.

Y	f	Yxf	Y-M	$(y-M)^2$	z= (Y-M)/s
0	1	0	0-2,4= -2,4	5,76	-2,4/1,4= -1,71
1	2	2	1-2,4=-1,4	1,96	-1,4/1,4= -1,0
2	2	4	2-2,4=-0,4	0,16	-0,4/1,4= -0,29
3	3	9	3-2,4= 0,6	0,36	0,6/1,4= 0,43
4	1	4	4-2,4=1,6	2,56	1,6/1,4 = 1,14
5	1	5	5-2,4= 2,6	6,76	2,6/1,4 = 1,86
	n = 10	Σ(Y x f)=24		$\Sigma(y-M)^2=17.56$	

$M = \frac{\Sigma f Y}{n} = \frac{24}{10} = 2,4$

$s^2 = \frac{\Sigma f (Y-M)^2}{n-1} = \frac{17,56}{9} = 1,95$

$s = \sqrt{s^2} = 1,40$

Vi har alltså medelvärdet M=2,4, och standardavvikelsen s=1,4 i råpoängen. Efter normering har vi Medelvärdet 0 och standardavvikelsen 1. Om vi nu omvandlar poängen med hjälp av konstanter på samma sätt som inom parenteesen på föregående sida när det gällde X-variabeln, detta för att minska antalet decimaler och bli av med de negativa talen, får vi en med X-variabeln jämförbar Y-variabel.

från föregående sida.

[10(Y-M)/s]+50	X variabelvärden	Yvariabelvärden
-17,1 + 50 = 39	31,7	39
-10,0 + 50 = 40	38,8	40
-2,9 + 50 = 47,1	45,8	47,1
4,3 + 50 = 54,3	52,8	54,3
11,4 + 50 = 61,4	59,8	61,4
18,6 + 50 = 68,6	66,8	68,6

FACIT TILL SID. 35.

 ISBN 978-91-639-6675-0

Att jämföra variabler - sambandsmått, och några ord om prediktion (linjär regression)

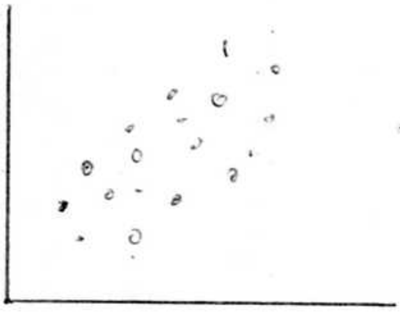
När vi har normerat våra variabelvärden på variablerna X och Y till Z_x och Z_y kan vi göra jämförelser mellan varibelvärdena. Finns det, t.ex. något samband mellan personers vikt och längd?

Den vanligaste metoden för att *söka efter samband mellan två variabler, och att beskriva styrkan i detta, eventuella samband*, är att räkna ut någon typ av korrelationskoefficient som betecknas med den lilla bokstaven r i ett stickprov och med den grekiska bokstaven ϱ (rå) i populationen.
Exempel på korrelationskoefficienter är Pearson's produktmomentkorrelationskoefficient och rangkorrelationskoefficient. Vilken typ av koefficient vi väljer beror på karaktären av vår data. Produktmomentkorrelationskoefficienten är kanske den mest använda korrelations-koefficienten.

Låt oss säga att vi har undersökt två variabler, X och Y, i ett stickprov. Vi tillämpar Z normering (se föregående sidor i kompendiet) för varje variabel för sig så att varje individ erhåller ett normerat X värde, dvs.ett Z_x värde, och ett normerat Y värde, dvs. ett Z_y värde.

$$Z_x = \frac{X - M_x}{s_x} \qquad Z_y = \frac{Y - M_y}{s_y}$$

Om vi bildar produkten av Z_x och Z_y, dvs. Z_xZ_y kommer alla de individer som ligger ovanför medelvärdet och alla de individer som ligger under medelvärdet i *båda* variablerna att erhålla *positiva* värden. Prickar vi in individernas resultat i ett *spridningsdiagram* som i exemplet till höger, ser vi att det föreligger ett *positivt* samband mellan variablerna.



Produkten av Z_xZ_y är negativ för alla individer som erhållit poäng under medelvärdet, antingen på variabeln X *eller* på Y, (ej på båda). Skulle detta vara fallet i vårt exempel och vi prickar in värdena i ett spridningsdiagram, får vi ett *negativt* samband mellan variablerna. Ett negativt samband kan se ut som i diagrammet till höger.

Summerar vi produkterna Z_xZ_y för alla individer och dela med n-1 får vi ett sammanfattningsmått. Vi kan plocka in definitionerna på Z (normering) i formeln för korrelationskoefficienten:

härledning:
$$\left[r_{xy} = \frac{\sum z_x z_y}{n-1} \right]$$

$$r_{xy} = \frac{\sum (X - M_x)(Y - M_y)}{(n-1)\, s_x s_y}$$

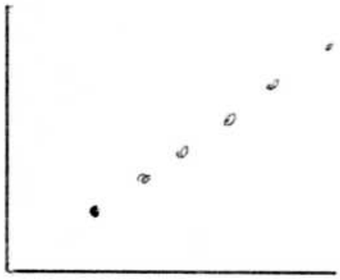
(obs! formeln kan också skrivas på annat sätt)

 ISBN 978-91-639-6675-0

Kännetecknande för produktmomentkorrelationskoefficienten är:

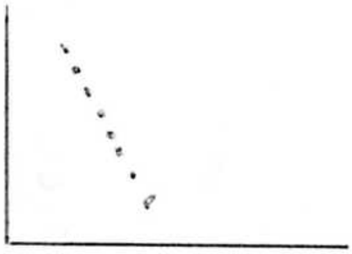
1. r_{xy} antar endast värden mellan -1 och +1.

2. Vid $r_{xy} = 1$ föreligger ett perfekt,
positivt,
linjärt
samband
mellan X och Y



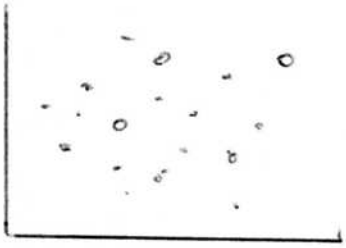
$r_{xy} = 1$

3. Om $r_{xy} = -1$ föreligger ett perfekt,
negativt,
linjärt
samband
mellan X och Y.



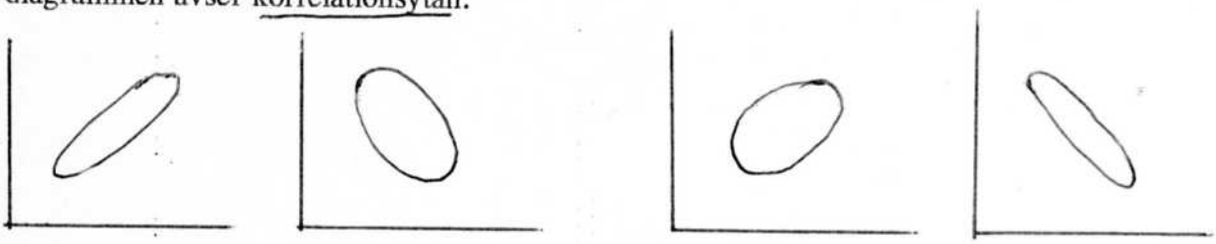
$r_{xy} = -1$

4. I de fall det **inte** finns något *linjärt* samband (det kan finnas *andra typer* av samband dock!)
är $r_{xy} = 0$
Då ligger observationerna på något sätt utspridda i spridningsdiagrammet: exempelvis så som de avbildats i diagrammet till höger.



$r_{xy} = 0$

Om spridningen ser ut såsom i de karikerade diagrammen nedan, kan r_{xy} **mycket grovt** tänkas ligga omkring de värden som anges under diagrammen. Ellipsytan i diagrammen avser korrelationsytan.



$r_{xy} \approx +.85$ $r_{xy} \approx -.60$ $r_{xy} \approx +.60$ $r_{xy} \approx -.85$

Några ord om linjär regression

Med hjälp av (normering och), korrelationskoefficienter och ekvationen för en rät linje kan vi bestämma en regressionslinje beträffande vår data på variablerna X och Y. Regressionslinjens ekvation kan vi sedan använda för prediktion (förutsägelser). Vi kanske vill predicera en individs utbildningsresultat (Y) med ledning av resultaten på test i förmågan att tillgodogöra sig undervisning (X).

 ISBN 978-91-639-6675-0

Statistisk inferens och hypotesprövning

Forskare som arbetar utifrån empiri arbetar ofta med hypoteser. Man kan då till exempel jämföra två populationer med avseende på skillnader i samplingsfördelningarnas medelvärden, eller kanske vill man jämföra skillnader i korrelationskoefficienterna för vissa variabler inom båda samplen.
Då det säkraste sättet att bevisa en hypotes innebär att man måste undersöka hela populationen, måste man av praktiska skäl oftast nöja sig med att pröva hypotesen på ett sampel som man drar ur populationen. En sådan statistisk prövning av hypotesen med ledning av ett sampel, leder inte till någon absolut säker slutsats. Det finns två huvudtyper av fel som slutsatserna kan vara behäftade med: Typ-1 fel och Typ-2 fel.

Typ 1 fel innebär att hypotesen om populationen är sann, men på basis av det undersökta samplet, drar man slutsatsen att hypotesen är falsk.
Typ 2 fel innebär att hypotesen är falsk, men påstås vara sann på basis av den statistiska analysen av samplet.

Det allra vanligaste sättet att pröva hypoteser är att undersöka skillnaderna mellan två medelvärden. Ett sådant exempel kan till exempel vara att man vill undersöka om det finns skillnader i medelvärden mellan män och kvinnor beträffande en variabel. Man tänker sig först då en *nollhypotes* som innebär att inga skillnader mellan gruppernas medelvärden föreligger. Nollhypotesen kan skrivas:

$$H_0 : \mu_1 - \mu_2 = 0$$

Om nollhypotesen är falsk, kan man anta en så kallad *alternativ hypotes* som skrivs:

$$H_1 : \mu_1 - \mu_2 \neq 0$$

Om man skall pröva nollhypotesen, undersöks skillnaderna i medelvärderna i båda grupperna. För att man skall kunna göra detta, måste man beräkna de båda samplens standardavvikelser. Man brukar också normera de observerade poängen till z-poäng och använda sig av olika statistiska test. Den säkerhet med vilken man presenterar resultatet, brukar man benämna *signifikansprövning*. Signifikans handlar om slumpens möjliga inverkan på resultatet, t.ex. på medelvärdesskillnaden i vårt exempeltest. Man brukar välja en lämplig signifikansnivå, t.ex. på 1 % eller 5 % ($p < .05$). En signifikansnivå på t.ex. 5 % säger att hypotesprövningen ger ett signifikant resultat på 5 % nivån, dvs. hypotesprövningen tyder på en verklig skillnad, men det finns fortfarande en viss spelmån för att resultatet som man har erhållit, enbart beror på slumpen.

Det finns otaliga olika statistiska förfaranden för att få fram så säkra slutsatser som möjligt. Man kan, till exempel använda sig av t-test, chi-två test, anova, etc. Då statistisken i sig är en vetenskap, lämnar vi här de olika förfaranden. Den som vill studera kan vända sig till litteraturlistan i slutet av kompendiet! Lycka till!

 ISBN 978-91-639-6675-0

"ETT LITET EXEMPEL UR VERKLIGHETEN"

Socialt Stöd och hörselhandikapp – en empirisk studie av Gia Korhonen

(abstracts) + std 9, 10, 12 och 13.

Syftet med studien var att se om det fanns samband mellan duration, handikappupplevelse och det sociala stöd som hörselskadade personer uppfattar sig ha. Avsikten var också att undersöka hur mycket stöd man uppfattar sig få och av vem, samt att se ifall det fanns könsskillnader i undersökningsgruppen. Tjugonio personer med hörselnedsättning - 15 kvinnor och 14 män - svarade på ett frågeformulär. Rådatan analyserades med deskriptiv statistik, och sambandsmått mellan variabler; Pearson korrelationer (r). Resultaten stödjer hypotesen, att ju mer stöd man uppfattar sig ha från sin familj, desto mer stöd uppfattar man också från sina vänner, ($p<.01$). Signifikant samband hittades dock varken mellan variablerna handikappupplevelse och socialt stöd, eller variablerna duration och socialt stöd. Vad beträffar könsskillnader, testades delgrupperna med t-test och chi-kvadrat och det uppdagades att kvinnor hade signifikant mer kontakt med familjen än männen ($p<.01$) och också att de rapporterade i större utsträckning svårare upplevelse av sina hörselproblem än männen ($p<.05$).

i samarbete med dr. Maud Ericsson-Mangold och dr. Anders Ringdahl

[I kompendiet; sidorna 45-49; att användas endast som exempel.]

Empiri 1 (abstr

 ISBN 978-91-639-6675-0

Tabell 1: Max. poäng, medelvärde, standardavvikelse och range för SSVH:s subskalor.

Subskala	max. poäng	M och Sd		range
A familjestöd	(30)	24.82	5.74,	8 - 30
B intimitet	(15)	09.76	3.56,	2 - 15
C tillhörighet	(15)	10.41	2.98,	3 - 15
D påtaglig hjälp	(15)	12.45	2.81,	2 - 15
E självkänsla	(15)	10.14	3.29,	3 - 15
* B till E	(60)	42.76	9.74,	22-58

B - E; summa stöd från vänner.

Variabeln stödet från familjen, (A), visade sig ha signifikant positivt samband med variabeln *B-E, r= .565, p<.01, vilket stöder hypotesen att ju mer stöd man har från sina närmaste, dvs. sin familj, desto bättre fungerar man även i andra relationer.

Tabell 2 visar sambanden mellan stödet från familjen, (A), och de olika subskalorna som mäter stödet från vänner, (*B-E). Sambandet mellan subskalan "intimitet", (B), och stödet från familjen, (A), var signifikant på 5%-nivån, (p<.05). Subskalan "tillhörighet", (C), och familjestöd, (A), var också signifikant på 5 %-nivån, (p<.05), och subskalan (E), som mäter självkänsla, och (A) var signifikant på 1 %-nivån, (p<.01).

Tabell 2: Korrelationsmatris för subskalorna på SSVH.

Correlation Matrix for Variables: $X_1 \dots X_6$

	famstöd	intim*	tillhör	påtag.hj	sj.känsla	B-Etot
famstöd	1					
intim	.455 *	1				
tillhör	.407 *	.455 *	1			
påtag.hj	.273	.471 *	.816 ***	1		
sj.känsla	.551 **	.485 *	.332	.043 *	1	
B-Etot	.565 **	.825 ***	.83 ***	.735 ***	.639 ***	1

p<.05=*, p<.01=**, p<.001=***

E.

Socialt stöd, (SSVH), handikappupplevelse, (GP), och duration, (SSVH)

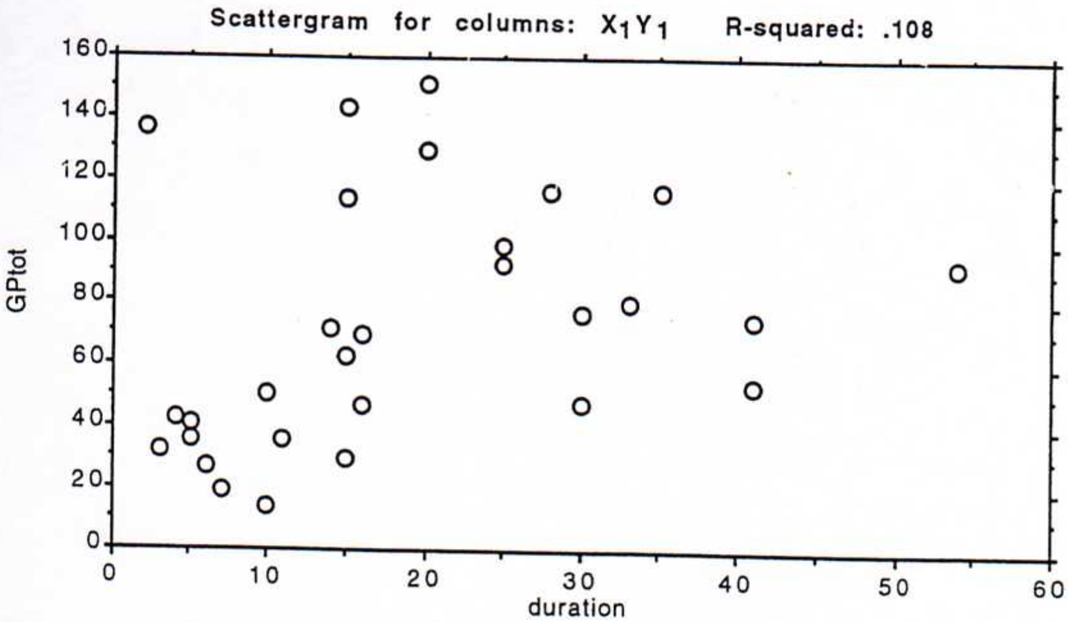
Medelvärdet för handikappupplevelsen mätt med Göteborgs-Profilen var för undersökningsgruppen, n=29, M=72.59+-39.62. och en range mellan 14 och 151 poäng.

Medelvärde, standardavvikelse och range för de olika delskalorna på Göteborgs-Profilen, GP1 t o m GP4, var följande:

GP1;	M=19.07+-10.19,	och en range mellan	6	och	45	poäng.
GP2;	M=16.41+- 9.42,	"	2	och	42	"
GP3;	M=19.83+-14.49,	"	1	och	48	"
GP4;	M=17.27+-15.38,	"	0	och	48	"

Som jag påpekat i inledningen, var det av intresse att se om det fanns samband mellan handikappupplevelsen hos personer med hörselnedsättning och det sociala stöd som de uppfattar sig ha. Inget samband hittades dock mellan totalpoängen för socialt stöd, mätt med SSVH, och totalpoängen för handikappupplevelse, mätt med Göteborgs-Profilen.

Mot förmodan uppdagades heller inget signifikant samband mellan variablerna duration, (SSVH), och totalpoängen för handikappupplevelse, (Göteborgs-Profilen). I figur 2 kan man se en tendens till positivt samband mellan handikappupplevelse och duration, (r=.33), dvs. ju längre tid man har varit hörselhandikappad, desto större tenderar handikappupplevelsen hos individerna i undersökningsgruppen vara.



Figur 2: En tendens till samband mellan variablerna handikappupplevelse, (GP), och duration, (SSVH), r=.33, ej signifikant.

ISBN 978-91-639-6675-0

Skillnader mellan delgrupper. 2; *de som hade barn och de som inte hade barn. (SSVH)*

Medelåldern för de som hade barn, n=22, var M=56.59+-6.68 år och för de som inte hade barn, n=7,M=37.71+-15.81, år.

Ett signifikant positivt resultat, t(27)=-2.09, p<.05, hittades mellan de som hade barn resp. de som inte hade barn och om man uppgav sig ha tillräckligt med vänner, (SSVH). De som hade barn, n=22, M=1.18+-.40, svarade i större utsträckning att de hade tillräckligt med vänner, än de barnlösa, n=7, M=1.57+-.54.

I X^2 -test med frekvenser fanns bland dem som inte hade barn proportionellt fler, dvs. 4 av 7 (se Tabell 4) som uppgav att de inte hade tillräckligt med vänner jämfört med 4 av 22 av de som hade barn. Detta var signifikant på 5% nivån, (p<.05).

Tabell 4: Fördelningen av de som hade barn resp. de som inte hade barn och om de hade tillräckligt med vänner.

barn

Observed Frequency Table

		ja	nej	Totals:
tillräckligt med vänner	ja	18	3	21
	nej	4	4	8
	Totals:	22	7	29

Påtaglig hjälp, (D), mätt med SSVH, visade också signifikant positivt samband i t-test mellan de som hade barn och de utan barn, t(27)= 2,56, p<.05. De utan barn, n=7, M=10.29+-4.46, fick mindre med påtagligt stöd än de som hade barn, n=22, M=13.14+-1.67.

Skillnader mellan delgrupper. 3: *Könsskillnader*

I fråga om könsskillnader hade kvinnorna högre poäng på kontaktfrekvensen med familjen, mätt med SSVH, t(27)= 2.96, p<.01, n=15, M=5.33+-1.72, än männen, n=14, m=3.43+-1.74.

 ISBN 978-91-639-6675-0

Kvinnorna, $t(27)= 2.09$, $p<.05$, uppgav också att de upplevde svårare hörselproblem, mätt med SSVH, $n=15$, $M=2.60+-.63$, $p<.05$, än männen, $n=14$, $M=2.14+-.54$.

Det uppdagades en tendens till att kvinnor, $n=15$, $M=22.33+-10.63$, upplevde att de hade det besvärligare att höra tal, (GP1), $t(27)= 1.86$, $p=.07$, än männen, $n=14$, $M=15.57+-8.73$.

Män däremot, $n=14$, $M=13.07+-10.19$, hade fler manliga vänner än vad kvinnorna uppgav sig ha manliga vänner, $n=15$, $M=5.80+-7.60$. Detta var signifikant på 5 % nivån, $t(27)= -2.19$, $p<.05$.

Antaganden, att kvinnor skulle skatta högre på variabeln uppfattat socialt stöd, och intimitet-skalan och att män skulle skatta högre på variabeln påtaglig hjälp fick inget stöd av resultaten.

Sammanfattning av resultaten

* Ju mer stöd man har från sina närmaste, desto mera stöd tycker man sig också få också från andra människor.

* Hörselskadade människor som har barn upplever sig i större utsträckning ha tillräckligt med vänner och få mer påtaglig hjälp, än de som inte har barn.

* Det finns tecken på att ju mer upplevda hörselproblem man har desto mer stöd tycker man sig få från sina vänner, detta gäller speciellt mera påtaglig hjälp.

* Ensamlevande personer tenderar uppleva större hörselproblem och känna sig mer handikappade till följd av sin hörselskada, än sammanboende.

* Det finns tecken på att hörselskadade personer som lever ensamma har djupare relationer med sina vänner än de som är sammanboende har.

* Kvinnor rapporterade i större utsträckning svårare upplevelser av hörselproblem än män.

* Män har fler manliga vänner än kvinnorna.

* Av subskalorna på SSVH, fick påtaglig hjälp, (D), såsom förväntad, det högsta medelvärdet.

49

bilaga 5 (s.13)

ISBN 978-91-639-6675-0

Frågor baserade på
Forskningsmetodikens grunder av Patel & Davidsson
Grupparbete

1. Varför är matematik inte en *empirisk* vetenskap?

2. Nämn några sätt att söka preliminär kunskap om det forskningsområde/problemområde som du själv så småningom ämnar att studera? Varför är det bra att samla så mycket kunskap som möjligt om problemområdet som man är intresserad av?

3. Om du vill ha svar på frågorna:
   a) "Vilken är den bästa metoden för att lära någon att samarbeta bra?" och
   b) "Vilka underliggande eller outtalade regler styr ett bra samarbete?",
   vilka former av bearbetning och analys kommer då att behövas till respektive fråga?

4. Vad menas med att man *reducerar* den information som man *erhåller* från individer vid en undersökning som genomförs med ett frågeformulär? Kan man *expandera* information vid en undersökning?

5. Vad menas med en *fallstudie* och kan man generalisera resultaten från en fallstudie?

6. Vad menas med ett *fältexperiment*?

7. Finns det någon skillnad mellan en *förundersökning* och en *pilotstudie*, och vad består i så fall skillnaden utav?

8. Varför är det bra att föra *forskningsdagbok* ? (dvs. en dagbok över allt man gör beträffande sitt forskningsarbete).

9. Ge exempel på metoder att samla information som har *låg grad av standardisering* och *hög grad av strukturering*!

10. Vad kännetecknar intervjuer som har hög grad av standardisering och hög grad av strukturering i motsats till intervjuer som har låg grad av standardisering och låg grad av strukturering? Och hur ser frågeställningarna ut i de förstnämnda, jämfört med frågeställningarna i de sistnämnda?

11. Kan du ge exempel på vilken typ av frågor man bör undvika vid undersökningar?

12. Vilka problem medför det om man är en *okänd deltagande observatör* i något sammanhang?

13. Vad betyder *interobservatörreliabilitet* (eller "interbedömarreliabilitet")?
14. Beskriv kortfattat positivismen och hermeneutiken.

(Presentera)

Hälsningar
Oja

då många inte hade denna obligatoriska kursbok. [illegible]

Litteratur

Byström, Jan (1998). *Grundkurs i statistik.* Stockholm: Natur och Kultur.

Fhanér, Stig (1980). *Statistik - rätt upp och ner.* Stockholm: Almqvist & Wiksell.

Henrysson, Sten (1970). *Elementär statistik för psykologer, pedagoger m.fl.* Stockholm: Almqvist & Wiksell.

Röner Douhan, G. & Carlsson, M. (1992). *Statistik - en introduktion.* (s. 30-39) Stockholm: Runa förlag.

1991

Patel, R. & Davidsson, B. (?). *Forskningsmetodikens grunder.* Lund: Studentlitteratur.

*

Boken *Forskningsmetodikens grunder* av Patel och Davidsson finns tryckt ett flertal gånger och är en utmärkt introduktion till forskningsmetodik - LÄS DEN!!!

*

MATERIALET KOPLETTERAS MED
TIDNINGSARTIKLAR och DIAGRAM,
SAMT UTTDRAG UR FORSKNINGSARTIKEL!!!
Se Förord !

Beträffande etiska regler i samhällsforskningen; Här kan du lära dig mera:

www.wma.net/e/policy17-c_e.html
ovan adress gäller Helsingforsdeklarationen beträffande etiska regler för kliniska försök för utprovning av nya läkemedel

www.hsfr.se/humsam/index.asp
Det så kallade Humanistiskt-samhällsvetenskapliga forskningsrådet (hsfr) ingår vad som idag kallas för *Vetenskapsrådet* och ger bl. a. information om etiska regler beträffande forskning inom humaniora och samhällsvetenskaper.

För statistiker finns en yrkesetisk deklaration: www.cbs.nl/isi/ethics.htm

European Society for Opinion and Marketing Research (ESOMAR) publicerar etiska regler för marknadsundersökningar:
www.esomar.nl/codes_and_guidelines.html

 ISBN 978-91-639-6675-6

ÖVNINGSUPPGIFTER FACIT och REPETITION dec 2005

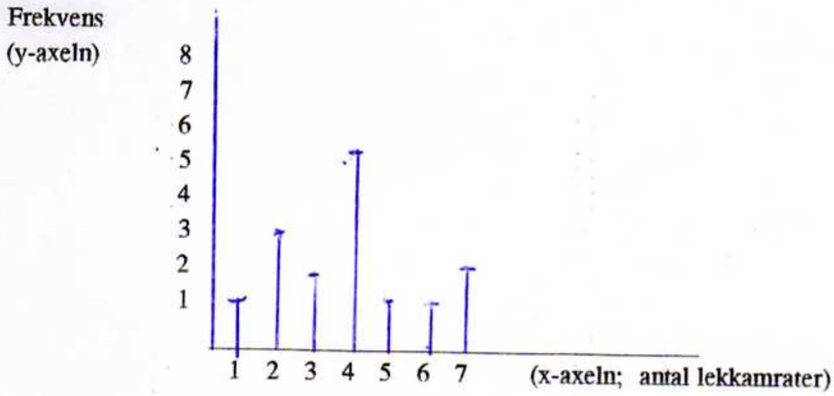
I en barnpsykologisk undersökning tillfrågades 15 barn hur många lekkamrater de hade. Barnen svarade 5 lekkamrater, respektive 3, 1, 4, 4, 7, 2, 6, 4, 4, 2, 7, 2, 3 och 4 lekkamrater (alltså; n=15).

1. Beskriv materialet i en tabell med frekvens och relativ frekvens.
En överskådlig beskrivning materialet får i om vi gör en *frekvenstabell* som anger antalet (frekvensen) lekkamrater som varje barn har. Tabellen kan se ut på följande sätt:

Antal lekkamrater (x)	Frekvens (f)	Relativ frekvens (f/n)
1	1	0,07
2	3	0,20
3	2	0,13
4	5	0,33
5	1	0,07
6	1	0,07
7	2	0,13
	n=15	1.0

Summan av frekvensen är ekvivalent med antalet observationer (n). I det här fallet 15. Summan av relativ frekvens skall alltid bli 1. Relativ frekvens kan man omräkna till procent, men för att procenttalet inte skall vara missvisande skall n vara minst 50.

2. Beskriv materialet i en diagram (och vad heter diagrammet?) Diagrammet heter stolpdiagram. Det används vid diskreta variabler, (till skillnad från stapeldiagram eller *histogram* som används vid kontinuerliga variabler. Även stolpdiagrammet kan dock se ut som ett stapeldiagram, men då är det mellanrummet mellan staplarna som visar att variabeln är diskret). I ett koordinatsystem betecknar vanligen y-axeln frekvensen och x-axeln variabelvärdena.



3. Vad karaktäriserar typvärdet och vad är typvärdet för de erhållna svaren?
Typvärdet är det mest frekventa värdet i en fördelning. Typvärdet (T) undersökningen beträffande barnens lekkamrater är sålunda 4. (T=4).

4. Vad är står Md för och hur skall man definiera det?
Md står för median. Medianen betecknar det mittersta värdet i en fördelning. Det delar så att säga fördelningen i två lika stora delar.

 ISBN 978-91-639-6675-6

S

5. Vad är medianen för de erhållna svaren? Medianen för vårt exempel med barnens lekkamrater är sålunda också 4, (Md = 4).
Råkar vi ha jämt antal observationer, är det brukligt att ange de två mittersta värdens medelvärde som median. Hade de två mittersta värden till exempel varit 4 och 5, hade medelvärdet för dessa två varit M=(4+5)/2, *vilket hade gett medianen 4,5.* (Md=4.5)

6. Hur räknar man ut medelvärdet? [Medelvärdet förkortas M (ibland skrivs det m för stickprov; medan medelvärdet en population betecknas med μ)].

Medelvärdet är *summan av alla iobservationer delat med antalet observationer (n), dvs.* (M $=\sum$ X/n)

Om vi har gjort en frekvenstabell, är medelvärdet lätt att räkna ut med formeln: M =*f*X/n, vilket i klartext betyder: medelvärdet (M), är lika med (=), summan av ($\sum$), variabelvärdens frekvens (f,) gånger observationsvärden (X). Hur detta räknas ut kan du se vid nästa fråga nedan.

7. Vad är medelvärdet för de erhållna svaren i ovan undersökning beträffande barnens lekkamrater?

Antal lekkamrater (x)	Frekvens (f)	
1	1	1 x 1 +
2	3	2 x 3 +
3	2	3 x 2 +
4	5	4 x 5 +
5	1	5 x 1 +
6	1	6 x 1 +
7	2	7 x 2 =

$M = \frac{58}{15}$

Alltså $\frac{1x1 + 2x3 + 3x2 + 4x5 + 5x1 + 6x1 + 7x2}{n=15}$ M = 3,87

8. Vad är *range* för de erhållna värden? *Range* eller *variationsvidd* (betecknad R) *är ett spridningsmått som anger mätvärdens variationsområde genom att ange skillnaden mellan det högsta förekommande värdet och det lägsta värdet. I vårt exempel är* R (7-1)= 6.

9. Vilken typ av variabel är det frågan om i undersökningen? Kontinuerlig variabel eller diskret variabel (diskret variabel är bara ett annat ord för diskontinuerlig variabel) och vad innebär det att en variabel är diskret?
Här är det frågan om en diskret variabel, eftersom antalet lekkamrater inte kan anta andra än heltalsvärden. Det kan inte finnas "halva" lekkamrater.

10. Vad innebär det att en variabel är kontinuerlig? Ge exempel på kontinuerliga variabler. *En kontinuerlig variabel kan anta vilket värde som helst (inte bara heltal).* Exempel på kontinuerliga variabler är reaktionssnabbhet, tid, etc.

11. Vilka fyra huvudtyper av skalor finns det och vad är karaktäristiskt för skalorna?
Nominalskalan sorterar efter kvalitativa egenskap i klasser, ex. efter kön (man eller kvinna), religionstillhörighet (protestant eller muslim) osv. Ordinalskalan rangordnar, t. ex. som minst, mellan, störst...men säger inget om avståndet mellan rangordningen. Intervallskalan både rangordnar och anger avståndet mellan olika skalvärden, dvs. "stegen" mellan alla skalvärden är lika långa men det finns ingen absolut nollpunkt. Kvotskalan rangordnar, anger avståndet mellan olika skalvärden och har dessutom en absolut nollpunkt, dvs. det kan finnas observationer som inte har den egenskapen som man avser att mäta, t.ex antalet trafikoffer.

 ISBN 978-91-639-6675-0 BILAGA s. 1/4

HEMTENTAMEN (frågorna 1-18, 4 sidor)

1. Vad menas med validitet?

2. Vad menas med reliabilitet?

3. När man gör ett *stratifierat urval* får man två eller flera "*strata*" i stickprovet. Kan du beskriva *hur ett stickprov som har samplats med hjälp av ett stratifierat urval kan se ut*?

4. Varför är samplingen (eller det sättet med vilket vi väljer ut vårt stickprov med) vid statistisk inferens så viktigt?

5. Vi har diskuterat några tidningsartiklar baserade på statistik och nämnt flera *"saker" som kan orsaka, eller ligga till grund för, felaktigheter i det statistiska materialet som presenteras.* Några sådana "*saker*" eller "*aspekter*" skrevs upp på tavlan. Minns du någon/några?

6. Vi har talat om tre typer av diagram för presentationen av *kvalitativ* data. Kan du nämna någon typ/några typer av diagram som du skulle använda om du skulle *presentera* en *kvalitativ* variabel *av nominal karaktär*?

7. Om någon variabel presenteras med ett *histogram,* tror du att variabeln är

a) en diskret, kvantitativ variabel

b) en kontinuerlig, kvantitativ variabel Svar:____

 ISBN 978-91-639-6675-0 5, 4/4.

8. Vad är utmärkande för en diskret, kvantitativ variabel?

---

9. Rita en *trappstegskurva* över den *kumulativa frekvensen uttryckt i procent* med ledning av följande tabell:

Antal barn (x)	Frekvens (f)	Kumulativ f	Relativ frekvens (f/n)	(frekvens i procent)	kumulativ frekvens
0	15	15	0,30	30	30
1	10	25	0,20	20	50
2	13	38	0,26	26	76
3	6	44	0,12	12	88
4	3	47	0,06	6	94
5	3	50	0,06	6	100 %
	n=50		1,0	100%	

Svar:

 ISBN 978-91-639-6675-0

10. Teckna en *ungefärlig* skiss över hur *en positivt sned fördelning* kan se ut och markera med ett litet streck på fördelningens bas på den punkt där de tre lägesmåtten som vi har diskuterat är placerade.

11. Räkna ut medelvärdet för barnens (n=28) provresultat i matematik. Provresultaten (x) anges i frekvenstabellen nedan:

X	f	fX
10	3	30
9	5	45
8	7	56
7	7	49
6	4	24
5	2	10
	n=28	214 $=\sum fX$

Formeln för Medelvärdet i det här fallet är

$$M = \frac{\sum fX}{n}$$

12a. Kan du läsa av typvärdet i nedan presenterade poäng erhållna på ett test?

2, 2, 2, 3, 3, 3, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6, 7, 7, 7, 8, 8, 8,

Svar: T = ______.

12b. Vad är Medianen för poängserien ?

Svar: Md = ______.

12c. Vad är variationsvidden eller range för ovan poängserie?

Svar: R = ______.

13. Vilka andra spridningsmått känner du till utöver variationsvidden?

____

____

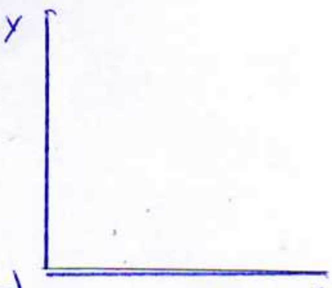
14. Om man har två variabler och har funnit ett samband dem emellan, beskrivs styrkan av detta samband med produktmomentkorrelationskoefficient. Vilka värden kan r anta?

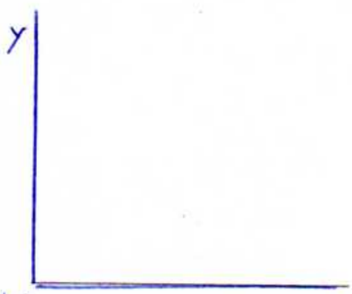
Svar:____

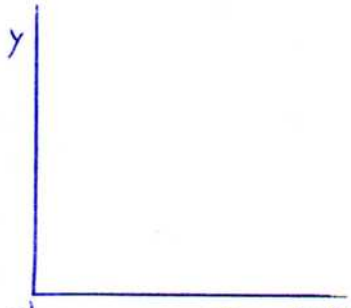
 ISBN 978-91-639-6675-0

15 Här har du tre stycken deluppgifter:
Teckna en skiss över ett *spridningsdiagram* över en produktmoment-korrelationskoefficient (r) som antar värdet

a) r = + 0.60
b) r = - 1.0
c) r = 0.85







a) r = +0,60, b) r = −1,0 , c) r = 0,85

16. Vad innebär Typ 1 fel vid hypotesprövning?

17. Vad innebär Typ 2 fel vid hypotesprövning?

Frivillig fråga

18. Räkna ut standardavvikelsen för barnens (n=28) provresultat i matematik.
Provresultaten anges i frekvenstabellen nedan:

X	f	fX
10	3	30
9	5	45
8	7	56
7	7	49
6	4	24
5	2	10
	n=28	214 = $\sum fX$

Formeln för Medelvärdet i det här fallet är

$$M = \frac{\sum fX}{n}$$

Formeln för standardavvikelsen är

$$s = \sqrt{\frac{\sum f(X-M)^2}{n-1}}$$

minns du? $\left(s = \sqrt{s^2}\right)$

(58) (4)