Föreläsning i Vetenskaplig metod 2008-11-08

Lärare: Sven Svensson, Östersund - Forskar bland annat i bemanningsbranschen om integrering av inhyrd personal. Koppling till konsultbranschen…

Böcker: Vetenskapsteori för nybörjare, Statistisk verktygslåda, Discovering statistics using SPSS (bra och instruktiv bok om SPSS)

Vetenskaplig historia – det blev inget sagt. Dock har jag hitta lite mer info på http://www.infovoice.se/fou/bok/10000025.htm.

Två olika vetenskapliga metoder:

Kvalitativ metod

- Ontologi
- Epistemologi

Kvalitativ metod = Sök kunskap genmom förståelse

Inte så intresserad av att generalisera. Arbetet sker genom intervjuer, observation, fokusgrupper, textanalys, fonetiska studier, diskursanalyser. Få studieobjekt som man drar slutsatser från.

Inrktad på ord, induktiv beslutsfattande, praktiska forskningsresultat.

Hermaneutiskt inrikad metod. ***Hermaneutik*** = Kunskap kan nås genom förståelse. Ex. en flykting som kommit från en krigshärd, posttraumatiskt syndrom. Det skulle vi kunna mäta, hjärnsubstanser, blodtryck osv. Det kan kopplas till flyktvägen och ge en bild. Men om vi vill ha en förståelse hur personen upplever sin situation och hur det är att fly och komma till ett nytt land. Det krävs samtal med fler personer där vi letar mönster. Detta är hermaneutikernas arbetssätt.

Konstruktivist – verkligheten skapas i interaktion med människor. Det gör att vi får svårt att motivera arbetet med enkäter eftersom det inte finns en allenarådande åsikt i någon fråga utan att skapa något i samband med människan.

Enligt wikipedia ”***Hermeneutik*** *är läran om tolkning och förståelse av diskurser. Det är en forskningsmetod där tolkningen är central. Ordet är efter grekiskans hermeneuo vilket betyder "tolka" efter guden Hermes. Förutom att vara en forskningsmetod kan hermeneutik också avse en existensiell filosofi som introducerades av Martin Heidegger.*”.

Kvantitativ metod

Det vi ska lägga tyngden på under dagen

- Variabeltyper
- Korstabeller
- Signifikansprövning

Positivistiskt inriktad. ***Positivism*** = Samhällsvetenskaperna ska kunna använda samma metoder som övriga grenar. Kunskap når vi genom det vi kan iaktta (empiri) och det vi kan räkna ut med logik, kvantifierbara fakta.

Objektivism – människan är ett objekt och styrs av strukturer. Oavsett hur strukturerna byggs upp så är tanken att det finns något större än människan. Jag kan ta fram en enkät och mäta attityden om kärnkraft, för att jag vet att det finns en sådan i samhället.

Enligt wikipedia ”***Positivism*** *(fr. positivisme), benämning på olika filosofiska riktningar som strävat efter att grunda tänkandet på "fakta", det vill säga kunskap som baseras på sinneserfarenhet. Genom empiriska studier försöker forskaren hitta egenskaper hos studieobjektet som återkommer också i andra fall och situationer. När man kartlagt ett tings regelbundenheter ger detta möjlighet att förutsäga, och ingripa i, ett skeende. Det som kan förutsägas (till exempel*

att jorden kommer att snurra ett varv runt solen under nästa år) är det som kan betraktas som kunskap. Det som inte kan vägas eller mätas betraktas som mindre intressant och man tenderar att se verkligheten som lineär.

Det mänskliga tänkandet anses enligt positivismen genomgått tre stadier: 1) Det teologiska stadiet där skeendena i världen ansågs beroende av gudar och andar. 2) Det metafysiska stadiet där skeendena i världen ansågs beroende av krafter och energier hos alla ting, till exempel hos atomer. 3) Det positiva stadiet där man insett att skeenderna i världen inte går, och inte behöver, förklaras.

Termen positivism myntades ursprungligen av Henri de Saint-Simon, men utvecklades främst av 1800-talssociologen Auguste Comte. En annan positivist var James Mill som uppfostrade sin son John Stuart Mill i positivistisk anda."

Två sätt att dra slutsatser

- Positivism: Deduktiva slutsatser = Härleda från logik, vi har en uppfattning om hur något funkar. Utifrån en teori tar vi in kunskap och ser om den överensstämmer. Först en teori och sedan en iakttagelse som bekräftar/fäller teorin. Som att åka till London och bekräfta att det regnar jämt :) Enligt wikipedia "*Deduktion är ett filosofiskt förfaringssätt för att härleda slutsatser från premisser. Utifrån ett antal premisser deducerar man en slutsats, exempel: "från A och B följer C". Deduktionen säger således ingenting om huruvida de ingående premisserna är sanna eller inte, bara att de kan sammankopplas till slutsatser, se exempel nedan. Det filosofiska begreppet deduktion är synonymt med matematikens begrepp bevis. Där kallas premisser axiom eller teorem och slutsatserna är också teorem.*"
- Hermeneutik: Induktiva slutsatser = Slutsatser från erfarenheter. Empiriska sanningar, sånt jag iakttar och sånt som kan vara representativt för fler/alla iakttagelser. Observationer och resultat ger teorier. Som turist är vi i ett land tre veckor och det regnar varje dag. Ett barns induktiva resonemang är att det alltid regnar i det här landet. Enligt wikipedia "*Induktion är ett filosofiskt förfaringssätt att härleda slutsatser från erfarenheter. Utifrån ett antal händelser inducerar man en sannolik slutsats. Ett exempel: "Solen har gått upp varje morgon hittills, alltså kommer den att gå upp imorgon också". En slutsats byggd på induktion kan anses ha bestämt sanningsvärde endast i fall där slutsatsen kan verifieras genom falsifikation. Inom naturvetenskap används metoden i kombination med deduktion och falsifikation. En slutsats baserad på induktion är alltså inte nödvändigtvis sann, eftersom endast verkan utan orsak är synlig. Ett exempel: "Den här tärningen har hittills endast visat sexor, alltså kommer nästa kast att resultera i en sexa". En invändning mot exemplet är dock att en större erfarenhet av tärningskast ger vid handen att alla tärningssiffror överlag är lika sannolika, vilket också kan ses som en induktiv slutsats. Trots att han medgav nödvändigheten av att använda oss av den i vardagslivet kritiserade och förkastade David Hume induktionen som en pålitlig kunskapskälla. Han visade bl.a. att de argument som kunde presenteras för den induktiva logikens giltighet tycktes förutsätta den, och att de därför var cirkulära. Projektet övergavs således framtills 1900-talets början då de logiska positivisterna genom sannolikhetslära försökte rättfärdiga den. Detta företag skulle Karl Popper komma att kritisera i sin bok Om forskningens logik.*".

Hermaneutik	**Positivism**
Induktion	Deduktion
Transperens	
Konstruktivism	Objektivism
	Reliabilitet och Validitet

Kvalitativ metod	Kvantitativ metod

Reliabilitet och validitet

Enligt wikipedia ”**Validitet** kan generellt sägas vara ett mått på hur väl man mäter det man vill mäta. Validiteten kan uttryckas som korrelationen mellan den teoretiska definitionen och den operationella definitionen.

Ett närliggande begrepp är **reliabilitet**, som uttrycker noggrannheten i mätningen. Ett valitt mått måste vara reliabelt, men det omvända gäller inte. Reliabiliteten för de två ingående variablerna sätter en övre gräns för vad validiteten kan vara, enligt:

$$r_{xy} \leq \sqrt{r_{xx} r_{yy}}$$

där r_{xy} är validiteten, och r_{xx} och r_{yy} är relibiliateten för de två variablerna.

Båda dessa är dragna **till den positivistiska läran.**”

Extern reliablitet

- Transperens – att vara tydlig med det du gjort, dina siffror och underlag. Visa upp
- Förförståelse – redogör för vad du påverkas av i din egen socialisering avseende fördomar, normer eller annat.

Intern reliabilitet

Det finns mått för intern reliabilitet för att veta att frågorna är bra.

- Forskarlag – Inom forskarlaget måste vi känna oss förtrogna med att det insamlade materialet är säkert. Allt forskningsmaterial måste sparas i tio år.

Enligt Vetenskapsteori för nybörjare (Thurén,2008:26-27) beskrivs reliabilitet och validitet på ett väldigt bra sätt. Där framgår det att *Reliabilitet* innebär att en mätning är korrekt genomförd och att vi, till en hög sannolikhet, kan lita på materialet. *Validitet* innebär att man verkligen udnersökt det man ville undersöka och inget annat. Det där känner jag igen som att Reliabilitet är att mäta på rätt sätt, medan Validitet är att mäta rätt saker.

Statistik

Vi arbetade sedan resten av dagen åt kvantitativ metod och statistik. Allt skedde på tavlan och med olika exempel i stigande svårighetsskala. Jag tyckte att det var grymt svårt att hänga med men hoppas att våra böcker kommer att ge mer hjälp och insikt.

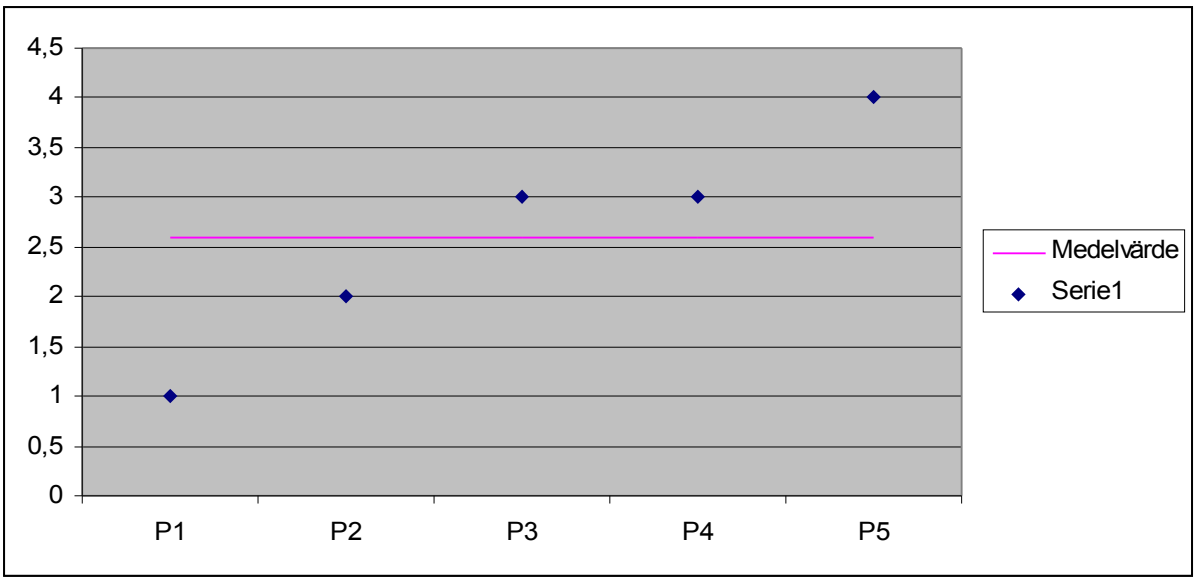
Medelvärde och standardavvikelse

Bygg en modell av verkligheten. Samla in information som tål statistiska test, och när vi sedan jämför verkligheten ska den passa in i det resultat vi fått ut.

Syskon hos mittuniversitets studenter

P1	P2	P3	P4	P5
1	2	3	3	4

Summan = 13
Medelvärde = 2,6



Summan av alla skillnader = -1,6 + -0,6 + 0,4 + 0,4 + 1,4 = 0!
Om vi kvadrerar alla diffar = $-1,6^2 + -0,6^2 + 0,4^2 + 0,4^2 + 1,4^2 =$ **5,2** ger sum of square error
Om vi delar det med N-1 får vi **1,3**
Vi drar roten ur får vi **1,14** vilket ger en standardavvikelse som kan relateras till medelvärdet på **2,6**.

Jämför med en normalkurva där 2,6 är medel och 1,14 visar på den normala avvikelsen inom gruppen. Ett mått på hur väl medelvärdet beskriver ditt data för din grupp och dess sammansättning.

Men. All information är inte av samma kvalitet. Det går inte att räkna medel och std.avvikelse på allt material.

Beskrivande, deskriptiv statistik.

Variabler; kön, utbildning, lön… Hur beskriver vi en variabel: Spridning mm.
Variabler kan vara av olika kvalitet.

Univariat är en variabel åt gången.
Bivariat är två variabler åt gången. Inkomst ökar med större utbildning = två variabler.
Multivariat är fler variabler.

Skalningsnivåer

Skalningsnivåer påverkar också användningen av variabler, sätt att arbeta med statistiska verktyg och de slutsatser som kan dras. Se wikipedia ”*Mätskala är ett begrepp används inom beskrivande statistik för att klassificera den variabel som mäts.*”

Nominalskala är val av ett begränsat urval värden. Den klassificerar variabelns innehåll. Se wikipedia ”*Det man mäter kan endast delas in i grupper utan innebördes ordning. Exempel är kön, yrke och sjukdomstyp. Variabelns olika värden kan endast beskrivas med ord, t.ex man/kvinna, läkare/sjuksköterska/städare/säljare o.s.v. För en variabel mätt enligt denna skala är det möjligt att beräkna typvärde, men inte median eller aritmetiskt medelvärde.*”

Ordinalskala rangordnar variabelns värde. Innehåller per automatik nominalskalans beskrivning. Se wikipedia ***”Variabelns olika värden kan rangordnas, men det går inte att på något meningsfullt sätt ange skillnader eller avstånd mellan värdena. Exempel är utbildning som kan anta värdena grundskola/ gymnasium/ högskola/ universitet/ forskarutbildning. Det går att ordna värdena efter stigande utbildningsnivå, men man kan inte tilldela en viss utbildning något numeriskt värde. För en variabel mätt enligt denna skala är det möjligt att beräkna typvärde och median, men inte aritmetiskt medelvärde.”.***

Intervallskala är inte så vanlig, men ett exempel är temperatur. Vi mäter avstånd mellan punkter, men det går inte att se något inbördes förhållande. Se wikipedia ”*I detta fall kan det man mäter tilldelas ett numeriskt värde. Ett exempel är temperatur mätt i grader Celsius. Det är här meningsfullt att ange skillnaden mellan två mätvärden. 20 grader Celsius är 10 grader varmare än 10 grader Celsius och 30 grader är ytterligare 10 grader varmare. Däremot är det inte riktigt att påstå att det en dag är dubbelt så varmt som föregående dag eftersom nollpunkten är godtycklig och det finns negativa temperaturvärden. För variabler mätta enligt denna skala är det möjligt att beräkna både typvärde, median och aritmetiskt medelvärde.*”.

Kvotskala har en absolut nollpunkt. En person kan inte vara negativt lång. Det gör att vi kan jämföra två värden och se deras inbördes förhållande, dubbelt så mycket t.ex. Här går det att använda alla fyra räknesätten Se wikipedia ”*Det som mäts kan beskrivas med ett kontinuerligt varierande numeriskt värde, och det finns ett entydigt sätt att definiera ett nollvärde. Därmed kan man jämföra storleken mellan de olika värdena. Exempel är kroppslängd mätt i centimeter. En person kan vara dubbelt så lång som en annan. Även temperatur mätt i kelvin mäts enligt en kvotskala eftersom nollpunkten här är absolut och det inte finns några negativa temperaturvärden mätt i grader Kelvin. Multiplikation och division är endast meningsfulla för variabler som mäts enligt en kvotskala.*”.

Variabler som endast kan mätas enligt en nominalskala eller ordinalskala kallas med ett gemensamt namn **kvalitativa variabler** och de som kan mätas enligt en intervallskala, kvotskala eller absolutskala kallas **kvantitativa variabler**. För kvantitativa variabler är det, förutom lägesmått som medelvärde och median, också möjligt att beräkna spridningsmått som standardavvikelse och varians.

Exempel:
Attityden till dödsstraff
1=Tycker inte, 5=Tycker helt

Attityd	Frekvens	Relativt
5	2	3,8%
4	1	1,9%
3	10	18,9%
2	15	28,3%
1	25	47,2%
S:a	**53**	

Typvärde är det vanligaste värdet; **1**.
Medianvärdet är det värde som finns exakt i mitten av observationerna; **3**. Med ett ojämnt antal observationer tar vi det värde som hamnar mitt i, dvs värdet av variabel n+1/2 (dvs (5+1)/2 så vi tar värdet på plats 3. Med ett jämnt antal observation så tar vi (värdet av n+1/2) + (värdet av n+2/2) och delar dem med två.

Kvartilmått. I likhet med median finns det tre kvartilmått; Q1 Q2 Q3. Q1 är värdet på plats (n+1)/4. Q2 är värdet på plats (n+1)/2, dvs medianen. Q3 är värdet på plats 3(n+1)/4.

Median och kvartil är mått som vi kan använda på lägre skalnivåer, ordinalskala och nominalskala. Däremot gäller det att vara medveten om riskerna med dessa mått, även om de kan vara användbara.

För kvot- och intervallskalor finns fler matematiska modeller att använda.
Aritmetiskt medelvärde = summan av alla värden delat med antal observationer.
Variationsvidd = maxvärde – minvärde.

Standardavvikelse. Se wikipedia ”***Standardavvikelse,*** *ett mått på hur mycket de olika värdena i en population avviker från medelvärdet. Begreppet används inom statistik, laborationer och matematisk statistik. Standardavvikelsen (σ) är en egenskap hos en sannolikhetsfördelning och definieras som kvadratroten ur variansen för fördelningen:*

$$\sigma = \sqrt{Var(X)}$$„

Se exempel på skillnader mellan olika observationers standardavvikelser:

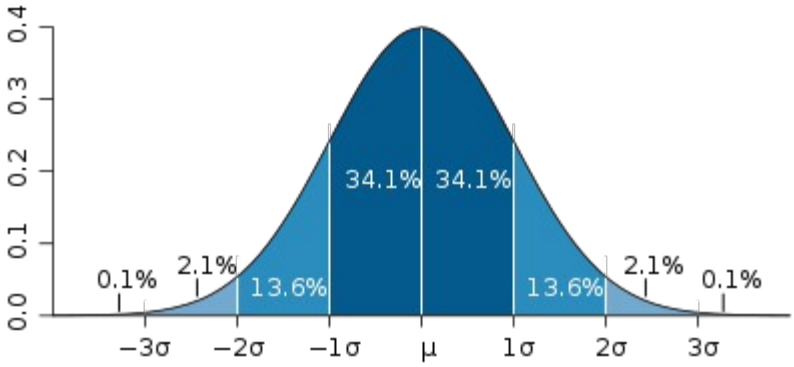
Månadslön i tusentals kronor

	Observation 1				Observation 2		
Person		Lön	$(x\text{-}xbar)^2$	Person		Lön	$(x\text{-}xbar)^2$
	1	19	64		1	12	2429,91
	2	20	49		2	18	1874,38
	3	21	36		3	24	1390,85
	4	22	25		4	30	979,32
	5	23	16		5	36	639,79
	6	24	9		6	42	372,26
	7	25	4		7	48	176,73
	8	26	1		8	54	53,20
	9	27	0		9	60	1,67
	10	28	1		10	66	22,15
	11	29	4		11	74	161,44
	12	30	9		12	80	349,91
	13	31	16		13	86	610,38
	14	32	25		14	94	1069,67
	15	33	36		15	100	1498,15
	16	34	49		16	106	1998,62
	17	35	64		17	112	2571,09
S:a		**459**	**408**	S:a		**1042**	**16199,53**
Antal		**17**		Antal		**17**	
Medel		**27**		Medel		**61,29**	
Std.avvikelse		**5,05**		Std.avvikelse		**31,82**	

För personen med 24k (observation 6 och 3) är z-transformeringen

(x-xbar)/s = (24-27)/5,05		(x-xbar)/s = (24-61,3)/31,82	
z=	-0,59	z=	-1,17

Statistiker har räknat ut ett antal fördelningar för att kunna analysera värden i stora urval. Z-fördelningen är en av de enklaste och helt normalfördelad. Medelvärdet är 0 och alla värden runt 0 är jämnt fördelad med hälften av värdena på den positiva sidan och hälften på den negativa sidan, och det är alltid en standardavvikelse som är 1 (z1 och –z1). Inom spannet från det negativa s (-z1) till det positiva s (z1) finns 68,26% av alla observationer. Om vi tittar mellan –z2 och z2 ligger 95,44% av alla värden. Med –z3 och z3 ges >99%.



Den centrala gränsvärdessatsen säger att om jag går ut och mäter längden på alla människor jag möter så kommer jag alltid att få en normalfördelning. Den kommer alltid att lägga sig normalfördelat runt medelvärdet. Det säger sig självt för att de längder vi mätt placerar sig på en kurva som de varit med och byggt upp.

Att jobba med z2 som ger 95,44% av populationen är besvärligt. Därför arbetar vi oftare med z1,96 och –z1,96 som ger 95%. Det ger en möjlig felmarginal på 5% i våra mätningar.

Om vi mätt upp ett medelvärde i längdmätningen. Hur troligt är det att vårt medelvärde och vårt underlag är representativt och finns på riktigt? Det är viktigt att ta reda på.

Exempel 1
En äldreomsorg med 1000 anställda. Mycket sjukskrivningar. Chefen tror att det beror på hög stressnivå. Dör att se om kommunens personal är mer stressad än personal i alämnhet. Vi mäter blodtryck och jämför med genomsnittsblodtrycket på en normalperson.
Personalchefen har format en hypotes om att blodtrycket är kopplat till stressnivån. Vi ska nu hypotespröva och bevisa detta. Vi arbetar dock tvärtom och försöker göra en falsifiering och motbevisa tesen.
H_0 är Kommunens blodtryck=Nationens blodtryck, dvs Kommunen-nationen=0
H_1 är Kommunens blodtryck ≠ Nationens blodtryck, dvs Kommunen-Nationen≠ 0

Om vi använder t-fördelningen och får ett värde som är extremt högt vet vi att vi är illa ute.
Med underlag där vi har färre än 30 personer kan vi använda t-fördelningar.
Nu jämför vi stickproven från verklighetens medelvärde som är 139!
SE=Standarderror=s/rot(n)=2,69
Vi får ett t-värde på 3,35 för differensen mellan 148 och 139.
Det jämför vi med vår kritiska nivå som vi slår upp via frihetsgrad 8 (antal värden -1) och en 5%-felmarginal ger 2,306. Det innebär att vårt värde landar i felmarginalen och vi har bevisat att det finns en, till 95% sannolikhet, skillnad mellan gruppen och allmänheten.

Exempel 2
Vi mäter 100 individer/studenter, lön per år. Vi vill testa att det medelvärde vi får fram med ett konfidensintervall. Med ett antal kronor +/- från medelvärdet hur säker jag är.
xbar=49700
s=3000

n=100
StandardError = SE = s/rot(n) = 3000/rot(100) = 300 kronor
(Egentligen bör vi använda verklighetens standardavvikelse som beteckans sigma, och då är SE=sigma/rot(n) men detta värde vet vi ju inte så därför använder vi provets medelvärde)

Xbar+-SE*z med en 95%ig säkerhet kräver z=1,96, 99% ett z=2,58 och 99.9% är z=3,29

Vårt konfidensintervall blir
95%: 49700 +- 300*1,96 = 47900+-588 inom vilket område verklighetens medelvärde ligger
99%: 49700 +- 300*2,58 = 47900+-774 inom vilket område verkligheten befinner sig
99.9%: 49700 +- 300*3,29 = 47900+-987 inom vilket område verkligheten finns

Exempel 3
Jämförelser mellan flickor och pojkars matteprov i skolan, räknat i poäng. Tesen är att pojkar är bättre på matematik.
H_0 : Differensen är 0, dvs Pojkar xbar = Flickor xbar
H_1 : Differensen är inte 0, dvs Pojkar xbar ≠ Flickor xbar

Pojkar
xbar=52,2857 i medel
s^2=152,5714 i varians
n=7

Flickor
xbar=50,25
s^2=112,2143
n=8

$t=d/SE_d$
t=pojkar xbar – flickor xbar / estimat av medelfelet
t=52,2857 – 50,25 / medelfelet = 2,0357 / medelfelet

Estimat av medelfelet = $rot(s^2 * (n_p + n_f)/(n_p * n_f))$

Eftersom vi har både pojkar och flickor måste variansen gälla båda grupperna.
$s^2=((n_p – 1)*s^2_p + (n_f – 1)*s^2_f) / (n_p+n_f-2)$

Det ger t = 2,0357 / rot(((7-1)*152,5714 + (8-1)*112,2143)/(7+8-2) * rot((7+8)/7*8))
t = 0,3438
Skumt?? Jag fick 0,087!?

Med nivå 5% och med 13 frihetsgrader (7+8-2) krävs, hämtat ur tabellen, 2,16 i t-värde för att bevisa en avvikelse och det gör att 0,3438 är alldeles för lite! Ingen skillnad mellan tjejer och killar bevisade.

Exempel 4
Sverige består av 40% arbetarklass, 50% medelklass och 10% överklass.
Om vi tar ett stickprov på förhållandet i universitetsvärlden. Är medel och överklass överrepresenterad?

Med chi-fördelning (χ) kan vi avgöra om vi har en snedfördelning bland studenterna. Den fungerar på nominal- och ordinalskalor. Om vårt chi2 överstiger det kritiska chi2 från tabellen visar det på att det finns en skillnad på riktigt.

Klasstillhörighet på universitetsstudenter

n=	200					
		Expected	**Observed**	**O-E**	**$(O-E)^2$**	**(O-E)2/E**
40%	Arbetarklass	80	30	-50	2500	31,25
50%	Medelklass	100	120	20	400	4
10%	Överklass	20	50	30	900	45
					Summa	**80,25** =chi2

chi2 = χ2 = summan av (O1 – E1)2 / E = 80,25
Frihetsgraderna på chi2 är (kategorierna -1) = 2
Läs ur tabellen för 95% sannolikhet 5,991 vilket gör att vi är låångt över och bevisar tesen.
Med 99,9% får vi 13,816 som innebär att vi fortfarande är långt över och kan sägas ha bevisat vår uppfattning.

Exempel 5
Undersök om det finns en koppling mellan urvalet av löneskillnader mellan män och kvinnor och se om det kan anses stämma för populationen i stort.

Snedfördelning av lön mellan kön

n= 170

Observed

	Män		**Kvinnor**		
0-15 tkr	4	2,35%	19	11,18%	**23**
16-25 tkr	93	54,71%	54	31,76%	**147**
	97		**73**		**170**

Expected

Män	**Kvinnor**
13,1	9,9
83,9	63,1

O	**E**	**O-E**	**$(O-E)^2$**	**(O-E)2/E**
4	13,1	-9,1	83,2	6,3
93	83,9	9,1	83,2	1,0
19	9,9	9,1	83,2	8,4
54	63,1	-9,1	83,2	1,3
			S:a	**17,1**

df = (r-1)*(k-1) = 1
Vi har ett chi2 på 17,1 och ska överskrida 3,841 för 95% eller 10,828 för 99,9% vilket innebär att vi kan anse det bevisat att vår test om att kvinnor har lägre lön än män.

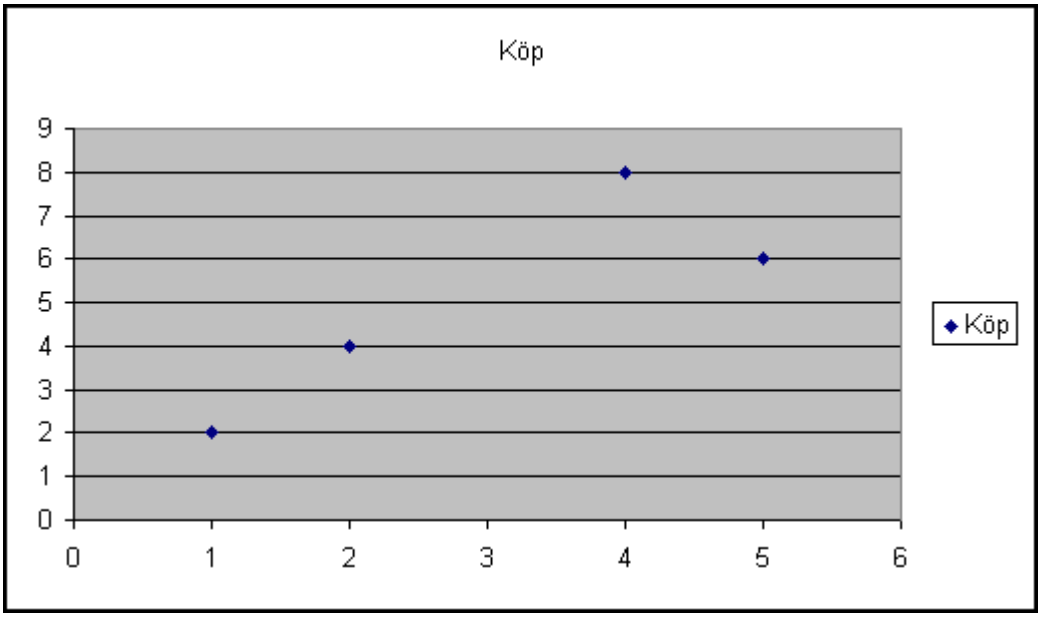
Pearsons produktmomentkorrelationskoefficient
Två variabler samarbetar med varandra. Exempel, ett cigarettföretag gör reklam för sina cigaretter. Går det att finna ett samband mellan hur många gånger en person tittar på reklamen och hur mycket cigaretter den personen köper. Riktningen är i det första fallet irrelevant.

Korrelationskoeefficienten r går mellan -1 och 1.
Den räknas ut som summan av alla (x – xbar)*(y – ybar) / ((n – 1)*s_x*s_y

Samband mellan reklam och köp av cigaretter

	Antal tittar	$(x\text{-xbar})^2$	Köp	$(x\text{-xbar})^2$	x-xbar	y-ybar	(x-xbar)*(y-ybar)
	1	4	2	9	-2	-3	6
	2	1	4	1	-1	-1	1
	5	4	6	1	2	1	2
	4	1	8	9	1	3	3
S:a	**12**	**10**	**20**	**20**			**12**
Antal	**4**		**4**				
Medel	**3**		**5**				
Std.avvikelse	**1,83**		**2,58**				

r= summan (x-xbar)*(y-ybar) / (N-1)*s_x*s_y= 0,848528



Med r=0,85 kan vi se en stark korrelation mellan antalet reklam vi tittar på och antalet cigaretter vi köper. Med ett r som är 1 finns en exakt koppling, och med ett r som är 0 finns inget samband alls.

Slutsats

Bestäm vilken skalnivå jag ska arbeta på. Normalfördelat data beter sig på sätt som vi kan använda z- t- och chi-fördelningar.

Not1. Jag tar inget ansvar för dessa anteckningar. Jag är en enkel student och kan ha missuppfattat hela föreläsningen, så använd materialet på ditt eget ansvar. Om du hittar nåt galet, mejla mig på peterA1967@hotmail.com :)

Not2. xbar skall läsas som ”medelvärdet av x” och jag skulle vilja skriva det som ett x med ett streck över men jag har inte hittat det tecknet i word ;) I övrigt är det nog många matematiska tecken som jag skulle vilja skriva på rätt sätt, men hey, detta är snabb-anteckningar…

Thurén Torsten (2008). *Vetenskapsteori för nybörjare*. Liber