

Statistiska analyser

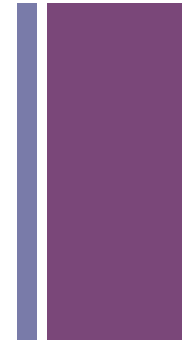
C2

Bivariat analys

Wieland Wermke

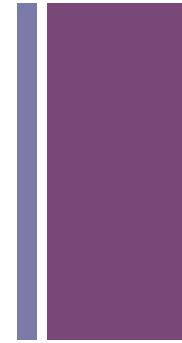
+ Bivariat analys

- Mål: Vi vill veta något om ett samband mellan två fenomen
- → inom kvantitativa strategier kan man undersöka detta genom att söka statistiska mönster mellan (minst) två variabler
- → Variabel: en egenskap som varierar i ett urval observationsenheter



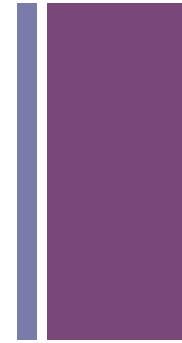
+ Bivariat analys

- Om man vet hur två variabler samvarierar, kan man, om man känner till värdet i ena variabeln, förutsäga värdet på den andra variabeln (med större säkerhet än om man bara gissar)
- <https://www.youtube.com/watch?v=VFjaBh12C6s>



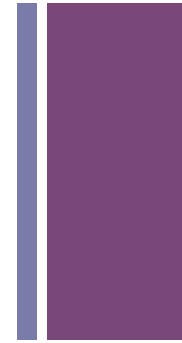
+ Bivariat analys

- En *ökning/minskning* av ett fenomen leder till en *ökning/minskning* av ett annat fenomen (*samvariation*), det kan vara:
 - **ensidiga samband:** t.ex. högre intelligens kan leda till bättre betyg i skolan, men inte tvärtom
 - **ömsesidiga samband:** en mer förtroendefull organisationskultur leder till bättre prestationer av medarbetare
 - **skensamband:** ju längre människor är, ju högre utbildning har de, här är det egentligen en annan variabel som verkar i bakgrunden



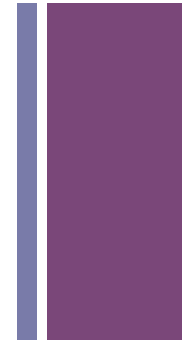
+ Samband – Kausalitet?

- När vi vill undersöka orsak-verkan, måste vi bestämma oss innan vi “räknar” vad som är
 - den **beroende** variabeln, dvs. den variabeln **som påverkas**
 - den **oberoende** variabeln, dvs. den variabeln **som påverkar**
- Men kan man sedan verkligen påstå att ett samband är kausalt?
 - Omvänt samband, ömsesidigt samband, slumpen, skensamband?
 - Statistiskt samband utan kausalitet, kausalitet utan statistiskt samband?





Hur kan man hävda kausalitet?



- Dessa steg behövs för att kunna uttala sig om kausalitet:
 - det finns ett samband i en riktning
 - det finns en **teori** att ett kausalt sammanhang är möjligt och **plausibelt** (före-efter, oberoende-beroende)
 - **Alla andra möjliga faktorer är kontrollerade/uteslutna**
 - Exempel: högre Intelligens → högre betyg (bara ett möjligt samband ang. betyg???)

+ Hur kan man hävda kausalitet?

- Dessa steg behövs för att kunna uttala sig om kausalitet

- 1) *A går i tiden före B.*
- 2) *A och B berör varandra i tid och rum.*
- 3) *Om A inträffar så är det nödvändigt att B inträffar.*

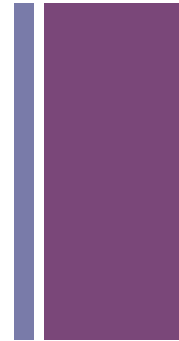
- En orsak är en nödvändig del av ett komplex av villkor vilka tillsammans är tillräckliga för verkan

+ Hur kan man hävda kausalitet?

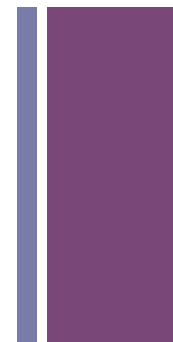
- Mycket **tankearbete** vad det gäller undersökningsdesign!
 - **Vilka variabler** måste samlas in **hur** för att kontrollera sambandet?
 - Hur ska data samlas in, vad är ett **passande urval**?
 - Kan man kanske använda **experiment** där man försöker utesluta alla påverkande variabler?
 - Vilken form av **randomisering** av undersökningsgrupperna?
 - **Stark teori och förankring i tidigare forskning!**
- Om du inte kan vara helt säkert på en kausal relation, bättre att då uttrycka att det finns "bara" ett **SAMBAND** (som också kan vara riktat)

+ En första ingång: korstabeller (nominal/ordinal)

- I en korstabell definierar den ena variabeln rader, den andra kolumner
- Korstabeller kan ha
 - **absoluta siffror**
 - **procentandel** (antal kategori/antal $n * 100$)
- Vid korstabeller med procent kan man på olika sätt bestämma sig, om man vill sammanfatta rader eller kolumner, eller båda



+ Korstabeller

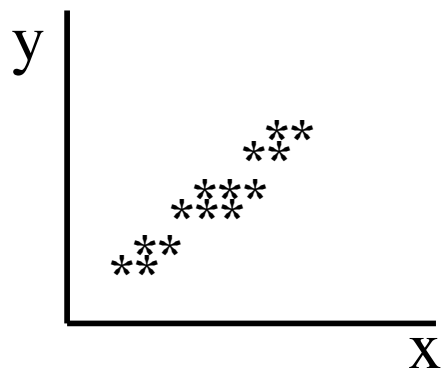


- Hur ser relationen mellan mattebetyg och tentamensångest ut?

Mattebetyg * Tentamensångest dikotomiserat Crosstabulation

			Tentamensångest dikotomiserat		Total
			ingen/lite tentamensångest	tentamensångest	
Mattebetyg	IG	Count	0	2	2
		% of Total	0.0%	4.0%	4.0%
	G	Count	0	24	24
		% of Total	0.0%	48.0%	48.0%
	VG	Count	6	10	16
		% of Total	12.0%	20.0%	32.0%
	MVG	Count	4	4	8
		% of Total	8.0%	8.0%	16.0%
Total	Count	10	40	50	
	% of Total	20.0%	80.0%	100.0%	

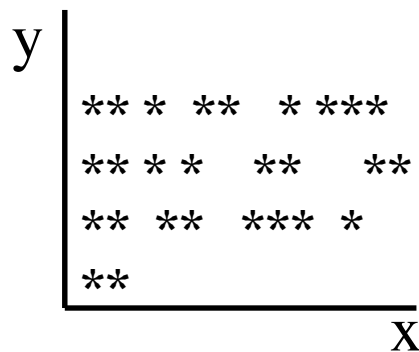
+ Olika samband i diagram



Positivt
(+ 1.00)

Ex. Positivt samband mellan provpoäng och betyg

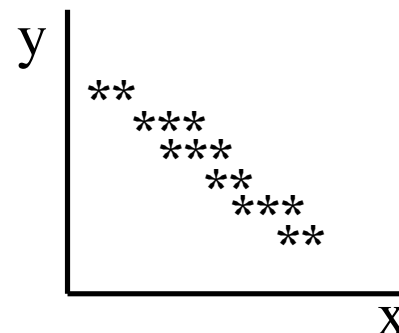
Ju högre X desto högre Y



Neutralt (noll)
(± 0.00)

Inget samband mellan har och betyg

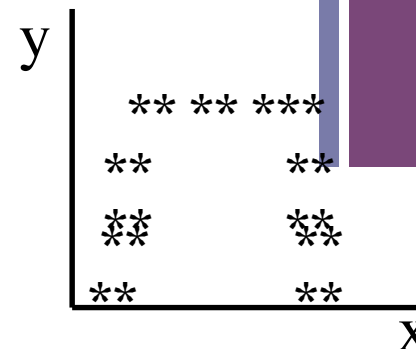
X har inget samband med Y



Negativt
(- 1.00)

Negativt samband mellan frånvaro och betyg

Ju högre X desto lägre Y



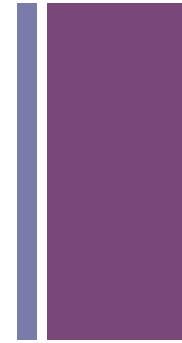
Non-linjärt
(±0.00)

Grad av njutning i samband med alkohol

Ju högre X desto högre Y till en viss nivå sedan lägre



Statistiska samband: Korrelation



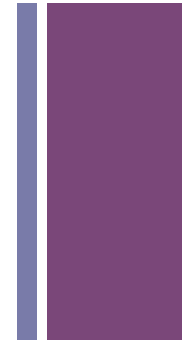
- ...om vi känner en undersökningsenhets ena egenskap kan vi också säga något om en annan egenskap hos undersökningsenheten (mer än om vi gissar)
- Med **vilken säkerhet** kan vi förutsäga ett samband = sambandets styrka
 - Om **osäkerheten** i vår förutsägelse är **hög** (mycket gissning), är **sambandets styrka låg**
 - Om **osäkerheten** i vår förutsägelse är **låg**, är **sambandets styrka hög**

+ Statistiska samband: Vad betyder det?

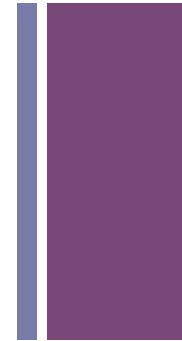
- ...vad betyder det att vi kan förutsäga ett värde genom kunskap av ett annat värde med 30% säkerhet?
 - ...vilket betyder att det är en risk att vi ligger fel med vår gissning i 70% av fallen?
- 1) **TÄNKÅ**: Teorier och hypoteser om möjliga samband
- 2) **RÄKNA**: Analysarbete utifrån insamlade data (lättaste delen)
- 3) **TÄNKÅ**: Tolka statistiska samband man får som resultat
 - Jämför olika samband, för att se sambandets betydelse

+ Relationer mellan nominala/ dikotoma kategorier

- Nominalmått (dikotoma data, distans mellan kategorier har ingen information, t.ex. ja/nej) utgår från **korstabeller** (oftast fyrfältare)
 - ...som innehåller antalet observationer (absolut eller i procent)
 - ...samt deras kombinationer t.ex. i form av randsumman (kolumn- och radsumma)
- Man relatera alla möjligheter till varandra och får ett mått (t.ex. **Phi**, **Cramer's V**, **Lambda**)



+ Korrelationsmått Phi



Variabel 2	Variabel 1		marginal-frekvens
	värde x	värde y	
värde q	a	b	a+b
värde r	c	d	c+d
<i>marg.frekv.</i>	a+c	b+d	

Gör en längre utbildning en bättre hälsa?

Exempel: phi=0.19

→ Ett svagt positivt samband

$$\varphi = \frac{22 * 35 - 33 * 10}{\sqrt{(33 + 22)(35 + 10)(33 + 35)(22 + 10)}}$$

$$\phi = \frac{bc - ad}{\sqrt{(a + b)(c + d)(a + c)(b + d)}}$$

hälsa	utbildning		marginal-frekvens
	kort	lång	
god	33	22	55
dålig	35	10	45
<i>marg.frekv.</i>	68	32	n=100

+ Exempel med SPSS, PSPP

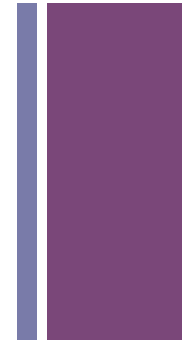
- *Hur stor är sambandet mellan studieämne (personal 0, ped 1) och motivation (inte motiverat 0, motiverat 1)?*

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.389	.019
	Cramer's V	.389	.019
N of Valid Cases		52	

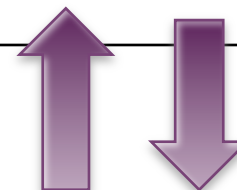
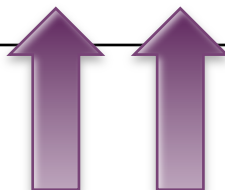
En signifikant (**men oriktat**) svagt positivt samband! Möjlig tolkning: pedagogerna är lite mer motiverade

+ Sambandsstyrka



- Förslag på en interpretation/tolkning:

0: ingen samband (korrelation) 0 – 0,2 mycket svagt samband 0,21 – 0,4 svagt samband 0,41 – 0,6 mellanstarkt samband 0,61 – 0,8 starkt samband 0,81 – 0,99 mycket starkt samband 1 perfekt samband	0 – -0,2 mycket svagt negativt samband -0,21 – -0,4 svagt negativt samband -0,41 – -0,6 mellanstarkt negativt samband -0,61 – -0,8 starkt samband -0,81 – -0,99 mycket starkt negativt samband -1 perfekt negativt samband
--	---



+ Ordinala Variabler: Parjämförelse

Upphandling av kompetensutveckling för avdelningen:

	Anbieter A 1	Anbieter B 2	Anbieter C 3	Anbieter D 4	Anbieter E 5	Anbieter F 6	Anbieter G 7
Anbieter A 1		0	2	2	1	0	2
Anbieter B 2	2		2	2	2	1	2
Anbieter C 3	0	0		0	0	0	0
Anbieter D 4	0	0	2		1	0	1
Anbieter E 5	1	0	2	1		1	2
Anbieter F 6	2	1	2	2	1		2
Anbieter G 7	0	0	2	1	0	0	

Summe	Rangfolge
7	3.
11	1.
0	6.
4	4.
7	3.
10	2.
3	5.

Exempel: "att upphandla"

2 = bättre

1 = lika

0 = sämre

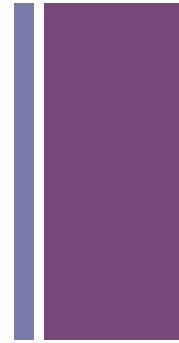
Samma logik som en Allsvenska-Tabell!!

+ Ordinala variabler: att jämföra rangordningar

- Eller man rangordnar alla x och alla y och sedan jämför rangordningen (man ställer två lister bredvid varandra) med varandra
 - Många överensstämmelser: **små/små och stor/stor** = stort samband (konkordant), eller små/stor och stor/små (diskordant)
- **Lista 1:** 12345678 **Lista 2:** 12345678 (samband= 1)
- **Lista 1:** 12345678 **Lista 2:** 87654321 (samband=-1)

+ Sambandsmått mellan **ordinala kategorier**

- Alla metoder har vissa nackdelar som man hittar i alla metodböcker eller på internet
- Det finns många olika mått: Kendall's tau a,b,c , Guttman's gamma, Yule's Q
- Som för sin del också ha olika fördelar och nackdelar



+ Exempel (SPSS, PSPP)

- Finns det ett samband mellan ointresse i statistik och betyget i matte?
 - *Betyg och ointresse behandlas som ordinalskalvariabler*

Symmetric Measures

	Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.	
Ordinal by Ordinal	Kendall's tau-b	-.358	.108	-3.218	.001
	Kendall's tau-c	-.311	.097	-3.218	.001
	Gamma	-.541	.157	-3.218	.001
N of Valid Cases		50			

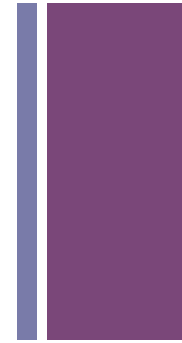
a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

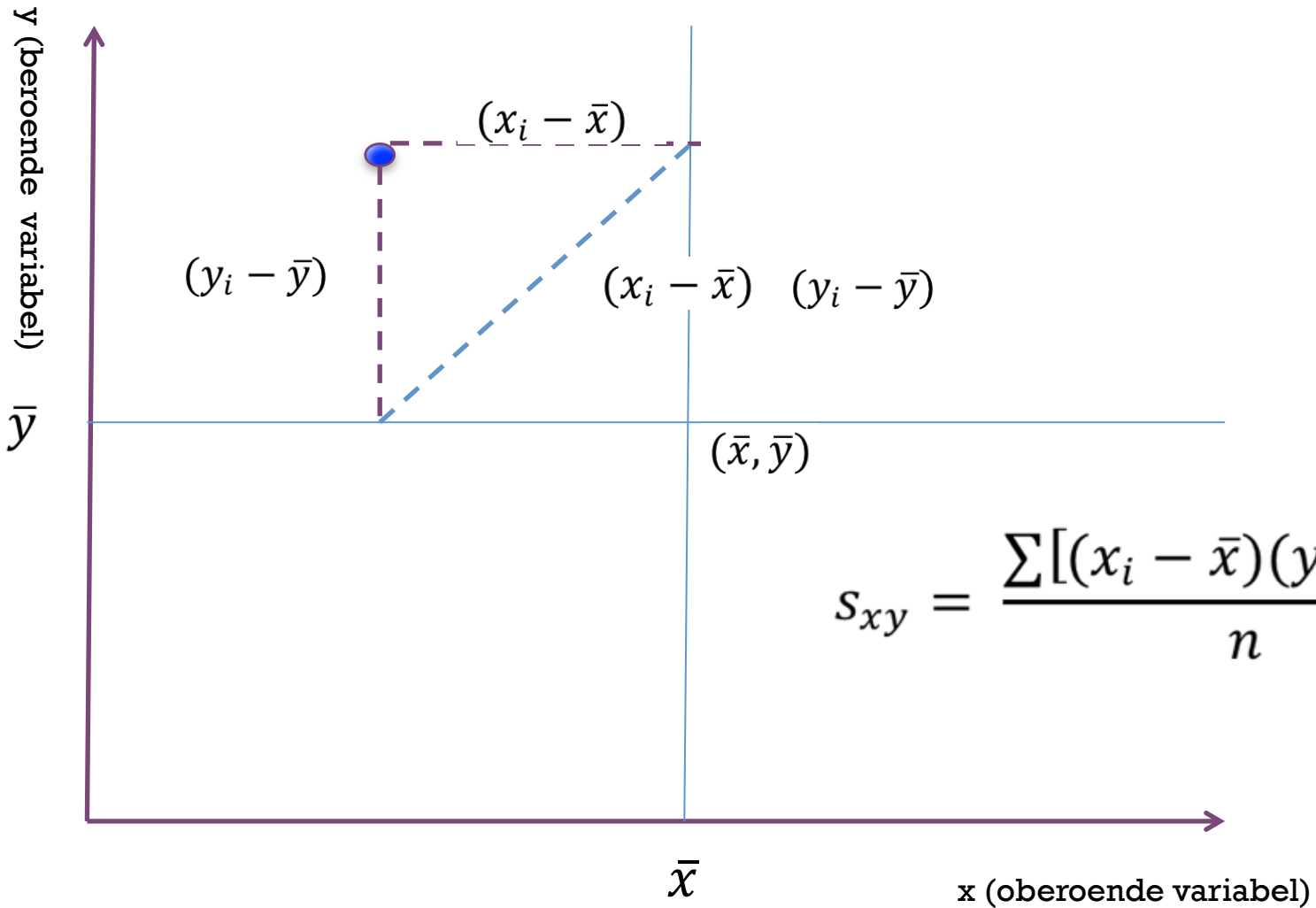
Ett signifikant, men inte riktat svag/mellan negativ samband. Möjlig tolkning: med bättre betyg, växande intresse

+ Samband mellan två intervallskalvariabler

- Ett sambandsmått mellan två intervallskalvariabler är **Pearsons r** , som bygger på relationer mellan variablers **varians**
 - Variansen anger hur observationsenheter koncentreras runt variabelns **medelvärde**
- Relationen kallas **kovarians**
 - Vilket visar, för varje observation och båda variabler, **avståndet mellan observationen och medelvärdet**
 - Varje observation har i ett **koordinatsystem** därmed en viss position



+ Kovarians S_{xy}



+ Pearsons r

- Kovariansen – liksom variansen – är beroende av variabelns mått/dimension (se varians-standardavvikelse)
 - **det betyder man inte kan jämföra varianser med olika dimensioner**
 - **därför standardiseras det även här (för att få värde mellan -1 och 1)**
 - man delar kovariansen med produkten av båda variablernas standardavvikelses produkt

$$r_{xy} = \frac{S_{xy}}{S_x S_y}$$

+ Exempel

- Finns det ett samband mellan tentamensångest och tolkningsångest ang. statistiska data?

Correlations

		Tentamensångest	Tolkningsångest
Tentamensångest	Pearson Correlation	1	.714**
	Sig. (2-tailed)		.000
	N	52	52
Tolkningsångest	Pearson Correlation	.714**	1
	Sig. (2-tailed)	.000	
	N	52	52

** . Correlation is significant at the 0.01 level (2-tailed).

Ett signifikant och starkt positivt samband

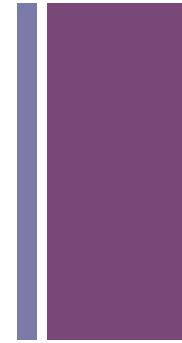
Har du tentamensångest så finns det en stor chans att du också har tolkningsångest

+ Analys av olika skalnivåer

y variabel			
	nominal	ordinal	intervall/kvot
nominal	Korstabeller Chitvå (signifikans); Phi, Lambda (korrelation)	Korstabeller Spearman's Rho (korrelation)	Stapeldiagram Variansanalys (Anova: f-test, t test) Eta (variansförklaring)
ordinal		Korstabeller Spearman's Rho Gamma, Kendall's Tau (korrelation)	Spearman's Rho (korrelation)
intervall/kvot			Scatterplot, Pearsons r Regressionsanalys (bivariat, multivariat) Faktoranalys
x variabel			

+ Samband mellan variabler med olika skalnivåer

- Vad gör man om man söker en samband mellan två variabler med olika skalnivå?
- **Man använder då måttet för den “lägre” skala**
 - Man förlorar dock informationer
 - Resultatet är inte helt precis
- Ordinal/Intervall behandlas som ordinal/ordinal
- Nominal/Ordinal behandlas som nominal/nominal



+ Parametriska och icke parametriska sambandsmått

- "Parameter" standardavvikelse, medelvärde och normalfördelning
 1. **nominal, ordinal** har inget medelvärde och standardavvikelse, följer inte normalfördelningen
 2. **intervall, kvot** har medelvärde, standardavvikelse och följer normalfördelning
- 1: behöver *icke* **parametriska** metoder
 - kan alltid användas, men är inte de mest precis metoder
- 2: behöver **parametriska** metoder