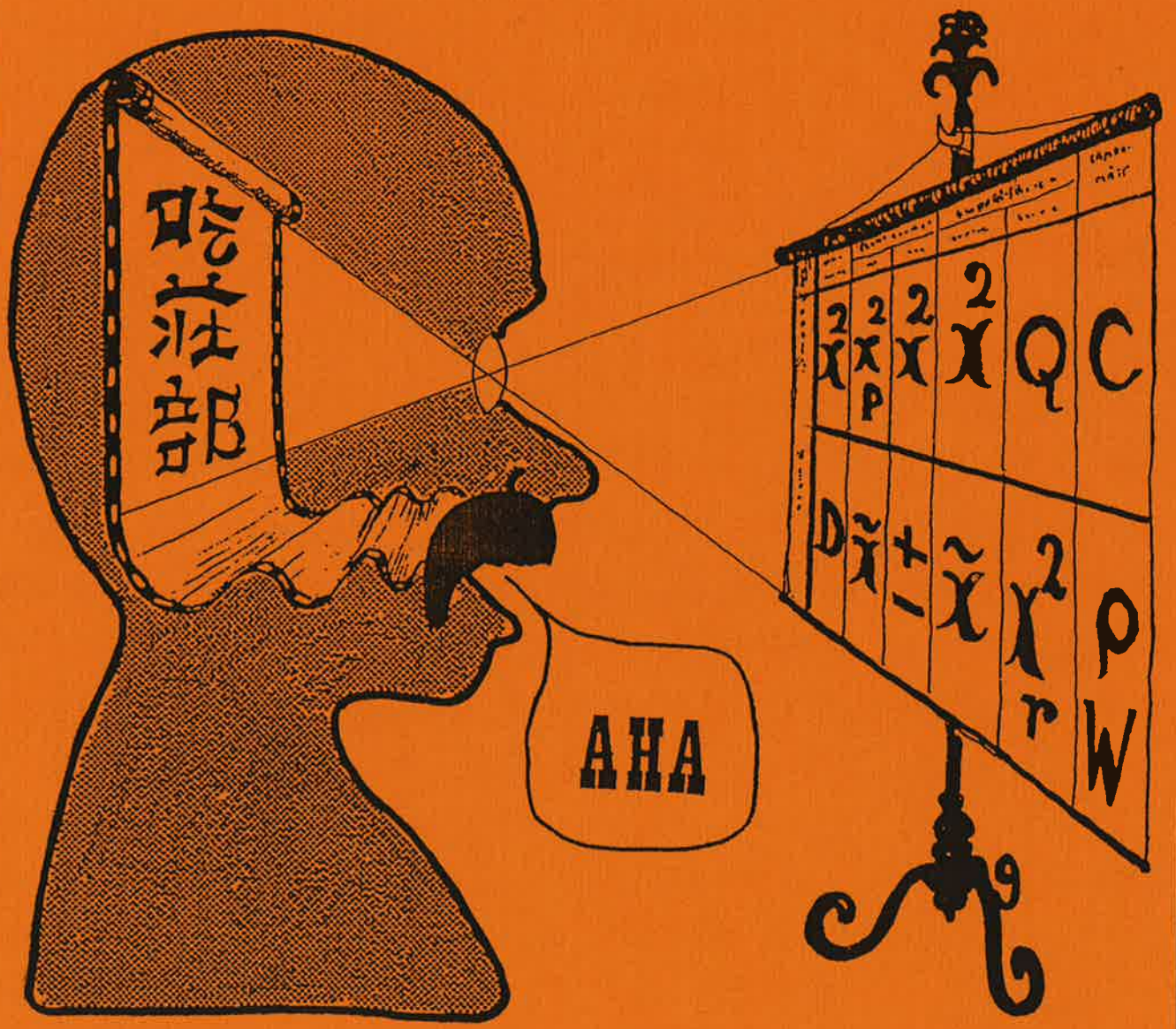
Kompendium i statistik

ICKE-PARAMETRISKA METODER



Sven Eric Reuterberg

INNEHÅLLSFÖRTECKNING

1. ATT VÄLJA STATISTISK METOD

1.1 Valet mellan parametriska och icke-parametriska metoder

Vid statistisk analys gäller det att finna en metod med vars hjälp man får ut så mycket och så säker information som möjligt ur de data som studeras. Hur mycket information man får ut hänger samman med metodens känslighet (termen känslighet är en fri översättning av den engelska termen "power") och känsligheten kan statistiskt definieras som

1 - sannolikheten för Typ II-fel.

Dvs. den anger metodens förmåga att ge utslag för de skillnader, som finns mellan två eller flera undersökningsgrupper. När man använder en metod med låg känslighet måste alltså skillnaderna vara större för att metoden skall ge utslag - dvs. nollhypotesen förkastas - jämfört med om vi har en metod med en hög känslighet. Med andra ord kan detta uttryckas så att en metods känslighet just avser dess förmåga att förkasta en falsk nollhypotes.

Vi angav ovan att den statistiska analysen också skall ge en säker eller tillförlitlig information. Graden av säkerhet påverkas bland annat av huruvida data är av sådan natur att de förutsättningar som den använda metoden ställer på data är uppfyllda. Några av de aspekter man har att ta hänsyn till i detta sammanhang är vilken skalnivå data uttrycks i, fördelningens form, i vissa fall också varianshomogenitet osv.

De statistiska metoder som har den högsta känsligheten och som således också ger mest information är de s.k. PARAMETRISKA metoderna. Som exempel på sådana metoder kan nämnas F-testet, t-testet och produktmomentkorrelation. Tyvärr är det så att denna grupp av metoder också ställer höga krav på egenskaperna hos data. Bland annat förutsätter de parametriska metoderna att data uttrycks lägst i intervallskala och de ställer dessutom vissa krav på fördelningens form.

Just på grund av att de parametriska metoderna har den högsta känsligheten väljer man i första hand en sådan metod under förutsätt-

ning att data uppfyller kraven för detta. Här är det emellertid befogat med en varning därför att i praktiken nöjer man sig ofta med att antaga att förutsättningarna för användning av en parametrisk metod är uppfyllda utan att närmare pröva detta. Skulle det sedan visa sig att en eller annan förutsättning saknas för den använda parametriska metoden innebär det att de slutsatser som den föranlett är mer eller mindre otillförlitliga. Oftast är det emellertid svårt att avgöra hur stora konsekvenser sådana fel får. Tyvärr är denna typ av fel inte alldeles ovanlig i beteendevetenskapliga undersökningar men det förhållande att en synd är vanlig gör ingalunda synden mindre.

Är man tveksam om huruvida kraven för en viss statistisk metod är uppfyllda bör man ta konsekvenserna av detta och i stället välja en metod med lägre känslighet men för vilken det inte råder någon tvekan om att data uppfyller kraven.

Sensmoral: det är bättre att nöja sig med något mindre men säker information än att få ut mycket men osäker information.

Det viktigaste vid valet av statistisk metod blir således att ta ställning till vilka egenskaper data har och därefter välja den metod som har den högsta känsligheten bland dem för vilka förutsättningarna är uppfyllda.

Antag nu att vi har en uppsättning data vars egenskaper är sådana att de parametriska metoderna är uteslutna. Vad gör man då?

Jo, förutom de parametriska metoderna finns det en grupp av metoder som kallas *ICKE-PARAMETRISKA* metoder. Dessa har visserligen en lägre grad av känslighet än de parametriska men å andra sidan ställer de lägre krav på datas egenskaper. Bland annat har de icke-parametriska metoderna inga eller i varje fall mycket låga krav på fördelningens utseende (ofta kallas icke-parametriska metoder av den anledningen fördelningsfria metoder). De kan också användas på data som uttrycks på en lägre skalnivå än intervallskala dvs. på nominal- eller ordinalskalenivå.

I och med att de icke-parametriska metoderna ställer lägre krav på data blir dessa metoder också mycket användbara - de kan ju teore-

tiskt sett användas även i sådana fall då data uppfyller de högre krav som ställs av de parametriska metoderna. Dock brukar man inte använda en icke-parametrisk metod i sådana fall där förutsättningarna är uppfyllda för användning av en parametrisk metod. Detta hänger samman med, som vi angivit ovan, att de icke-parametriska metoderna har lägre känslighet och därigenom ger mindre information än en tillämpbar parametrisk metod.

Vi har hittills diskuterat känsligheten som om denna enbart bestämdes av den statistiska metoden. Så är emellertid inte fallet. Känsligheten vid en statistisk analys påverkas också av undersökningsgruppens storlek (N) så att ju större N analysen baseras på desto högre blir känsligheten, vilket alltså innebär att vi kan kompensera en metods lägre känslighet genom att öka undersökningsgruppens storlek. Detta förändrar emellertid inte det förhållandet att vi i första hand väljer en parametrisk metod när så är möjligt eftersom vi då kan uppnå en känslighetsnivå som med en icke-parametrisk metod endast är möjlig att nå genom de ökade kostnader och arbetsinsatser som är förbundna med att öka gruppstorleken.

Det resonemang som förts hittills kan sammanfattas genom att göra en jämförelse mellan de parametriska och de icke-parametriska metoderna. Eftersom detta kompendium behandlar icke-parametriska metoder har vi valt att göra denna jämförelse utifrån dessa metoder och anger således vilka för- och nackdelar de icke-parametriska metoderna har i jämförelse med de parametriska.

Fördelar

1. De icke-parametriska metoderna ställer inga eller mycket låga krav på fördelningens form.
2. De kan användas på mycket små undersökningsgrupper.
3. De kan användas på grupper dragna från flera olika populationer utan speciella antaganden om dessa populationers fördelning.
4. Det finns icke-parametriska metoder, som är tillämpbara på data uttryckta på nominal- eller ordinalskalenivå.
5. De icke-parametriska metoderna är vanligtvis enklare att lära in och att använda än de parametriska.

Nackdelar

1. Om data har egenskaper, som tillåter användningen av en parametrisk metod har de icke-parametriska metoderna en lägre känslighet.

Ovanstående resonemang leder fram till följande tumregel för valet mellan en parametrisk och en icke-parametrisk metod:

Använd en parametrisk metod i de fall data uppfyller kraven för detta. Om det är tveksamt huruvida förutsättningar är uppfyllda för en parametrisk metod - välj då en icke-parametrisk.

1.2 Valet mellan olika icke-parametriska metoder

Nu finns det en hel mängd olika icke-parametriska metoder varför vi också har att göra ett val inom denna grupp. Vid detta val har man att ta ställning till följande tre frågor:

1) Vilken skala uttrycks data i?
2) Hur många uppsättningar av mätvärden har jag att studera: 1, 2 eller flera än 2 vilket brukar betecknas med k uppsättningar?
3) Är uppsättningarna av mätvärden oberoende eller beroende?

De olika skalnivåerna förutsätts vara kända, varför vi inte närmare behöver kommentera dem här.

När det gäller antalet uppsättningar av mätvärden kan det kanske synas märkligt att det förekommer fall där man studerar endast en uppsättning. Det är emellertid inte alldeles ovanligt i beteendevetenskapliga sammanhang att man har en enda uppsättning av värden, vars fördelning man vill jämföra med en känd t.ex. en teoretisk fördelning. Ett exempel på detta är att man tagit ut ett slumpmässigt stickprov av elever ur en viss årskurs. För att sedan få veta huruvida detta stickprov kan anses representativt med avseende på kunskapsnivå i ett visst ämne kan man jämföra dess betygsfördelning med den som anges av SÖ: 7 % 1-or, 24 % 2-or osv.

I detta fall har vi således en enda uppsättning av mätvärden (elevernas betygsfördelning) vilken jämförs med en teoretisk (den av SÖ angivna betygsfördelningen).

Försök finna andra exempel på undersökningssituationer där man har endast en empirisk uppsättning av mätvärden.

Vi har här av pedagogiska skäl gjort en avvikelse från den terminologi som används i engelskspråkig litteratur. I stället för att tala om antalet uppsättningar av mätvärden talar man där om antalet stickprov (sample). Termen stickprov får således representera såväl undersökningsgrupp som den uppsättning av mätvärden som studeras. Detta är emellertid mycket olyckligt i vissa sammanhang av den anledningen att antalet uppsättningar av mätvärden kan skilja sig från antalet undersökningsgrupper. Det kan t.ex. finnas undersökningar där man studerar en enda undersökningsgrupp men från denna får man flera uppsättningar av mätvärden. Låt oss ta ett exempel för att belysa detta:

Vi har utformat ett visst undervisningsavsnitt på fyra olika sätt enligt nedanstående:
A: enbart text utan någon form av illustration
B: text illustrerad med bilder
C: film + arbetsmaterial
D: programmerad undervisning

Efter det att alla elever i en klass fått genomgå samtliga fyra presentationssätt får de ange hur motiverande de upplevt vart och ett av dem.

Här har vi således en enda undersökningsgrupp (den studerade klassen) men denna grupp avger fyra uppsättningar av mätvärden (en uppsättning för vart och ett av presentationssätten).

Den tredje frågan vi ställde för att komma fram till en adekvat icke-parametrisk metod var om uppsättningarna av mätvärden är beroende eller oberoende. Även här avviker vi från den terminologi som används i engelskspråkig litteratur och som också är det gängse språkbruket hos oss. Man brukar nämligen tala om beroende och oberoende stickprov. I konsekvens och i klarhetens namn har vi emellertid valt att också i detta sammanhang tala om mätvärden i stället för att tala om stickprov.

En enkel tumregel, när det gäller att avgöra om mätvärdena skall anses vara beroende eller oberoende, är att vi har beroende mätvärden i två fall nämligen när de olika uppsättningarna av mätvärden härrör från en och samma undersökningsgrupp samt när mätvärdena kommer ifrån matchade grupper. I övriga fall betraktas mätvärdena som oberoende.

Det exempel vi angav ovan med en klass elever som fick ta del av ett undervisningsavsnitt presenterat på fyra olika sätt är ett exempel på beroende mätvärden. (Det är ju samma elever som avger alla uppsättningarna av mätvärden). Ett annat vanligt exempel på att en och samma undersökningsgrupp ger upphov till flera uppsättningar mätvärden är de s.k. före-efter-undersökningarna, där man först gör vissa mätningar på en grupp. Därefter får gruppen genomgå någon form av behandling varefter man upprepar mätningarna, vilka sedan jämförs med de första för att se om behandlingen haft någon effekt.

I vissa fall är det emellertid omöjligt att göra upprepade mätningar på en och samma grupp, varför man tvingas att arbeta med olika grupper. Det kan då vara ytterst viktigt att man får grupper som är så lika varandra som möjligt i sådana variabler som kan påverka den studerade variabeln. När grupperna tas ut på ett sådant sätt att varje individ i en grupp har en "tvilling" i den eller de andra grupperna talar man om att grupperna är matchade och uppsättningarna av mätvärden betraktas således som beroende.

Ett exempel på en sådan undersökningsuppläggning skulle vi ha om vi i exemplet med de fyra olika presentationssätten ovan i stället för motivationen vore intresserade av deras inlärningseffekt. Då skulle det självfallet vara omöjligt att använda en och samma grupp eftersom den får en viss inlärning vid varje presentation och tillskottens storlek kan påverkas av på vilken inlärningsnivå man startar. I stället skulle man här få ta ut fyra olika grupper - en grupp för varje presentationssätt. Dessa grupper måste emellertid vara så lika varandra som möjligt i t.ex. inlärningsförmåga och i förkunskaper inom ämnet för att man skall kunna renodla jämförelser till att gälla enbart presentationssätten. Därför måste de fyra grupperna tas ut på grundval av sådana variabler som av-

ser både inlärningsförmågan och förkunskaperna - man matchar grupperna utifrån dessa variabler.

Som vi angav är det alltså följande tre dimensioner vi har att ta hänsyn till vid valet av icke-parametrisk metod: skalnivå, antal uppsättningar av mätvärden, beroende eller oberoende mätvärden. På nästa sida har vi gjort ett urval av icke-parametriska metoder som är inplacerade i en översiktstabell uppbyggd på grundval av dessa tre dimensioner. Därtill har vi fogat vissa icke-parametriska korrelationsmetoder. Det är viktigt att observera att de metoder som tagits med i översiktstabellen och som vi kommer att behandla i kap. II och III endast utgör ett urval av icke-parametriska metoder. Då vårt syfte inte kan vara att ge en uttömmande beskrivning av samtliga icke-parametriska metoder har vi valt att göra urvalet så att vi presenterar minst en metod i varje cell i översiktstabellen, vilket också innebär att man med hjälp av de här redovisade metoderna kan klara de vanligast förekommande undersökningssituationer, där icke-parametriska metoder är tillämpbara.

SAMMANSTÄLLNING AV TILLÄMPNINGSOMRÅDEN FÖR VISSA ICKE-PARAMETRISKA TESTMETODER.

SKALA	1 UPPSÄTTNING MÄTVÄRDEN	2 UPPSÄTTNINGAR MÄTVÄRDEN		k UPPSÄTTNINGAR MÄTVÄRDEN		SAMBANDS-MÄTT
		oberoende	beroende	oberoende	beroende	
Nominal	1. Chi^2: Goodness of fit (2.3)	1. Chi^2: Tests of independence (2.4) 2. Chi^2: Fishers exakta sannolikhetstest (2.6)	1. McNemars test för signifikans av systematiska förändringar (2.7)	1. Chi^2:Tests of independence (2.4)	1. Cochrans Q-test (2.8)	1. Kontingenskoefficienten (C) (2.9)
Ordinal	1. Kolmogrov-Smirnov test för en uppsättning mätvärden (3.1)	1. Mediantestet (3.2)	1. Teckentestet (3.3)	1. Mediantestet (3.2)	1. Friedmans två-vägs variansanalys (3.4)	1. Spearmans rangkorrelation (p) (3.5) 2. Kendalls konkordanskoefficient (W) (3.6)

Anm: Siffrorna inom parentes utgör hänvisning till det kapitel och avsnitt i kompendiet, där metoden ifråga behandlas.

II. ICKE-PARAMETRISKA METODER FÖR DATA UTTRYCKTA I NOMINALSKALA

2.1 Inledning

Som framgår av översiktstabellen över icke-parametriska metoder kan olika varianter av chi^2-metoden användas då data är uttryckta i nominalskala och då vi har endast en uppsättning av mätvärden eller då vi har två eller flera datauppsättningar vilka är oberoende. Är datauppsättningarna däremot beroende har vi att välja andra metoder, vilka emellertid på olika sätt anknyter till chi^2. Även det sambandsmått som är tillämpbart på data i nominalskala, nämligen kontingenskoefficienten, hänger nära samman med chi^2.

Då de olika varianterna av chi^2 har vissa gemensamma drag har vi valt att inte strikt följa översiktstabellen i vår framställning utan vi kommer, att redogöra för dessa i ett sammanhang. Härigenom undviker vi bl.a. en mängd tyngande upprepningar.

2.2 Allmänt om chi^2

Utmärkande för chi^2 är att man med denna metod prövar *huruvida en empirisk eller observerad fördelning avviker från en förväntad fördelning*. Det är således skillnader mellan fördelningar som studeras.

Vi kan konkretisera detta genom att ta ett exempel:

Antag att vi vill pröva om en vanlig speltärning är "fixad". Om den inte är fixad bör slumpen verka så att vi får lika många 1-or, 2-or, 3-or, 4-or osv. Är den däremot fixad är det rimligt att antaga att ett visst värde (t.ex. 6-an) kommer upp flera gånger än slumpen bidrar med.

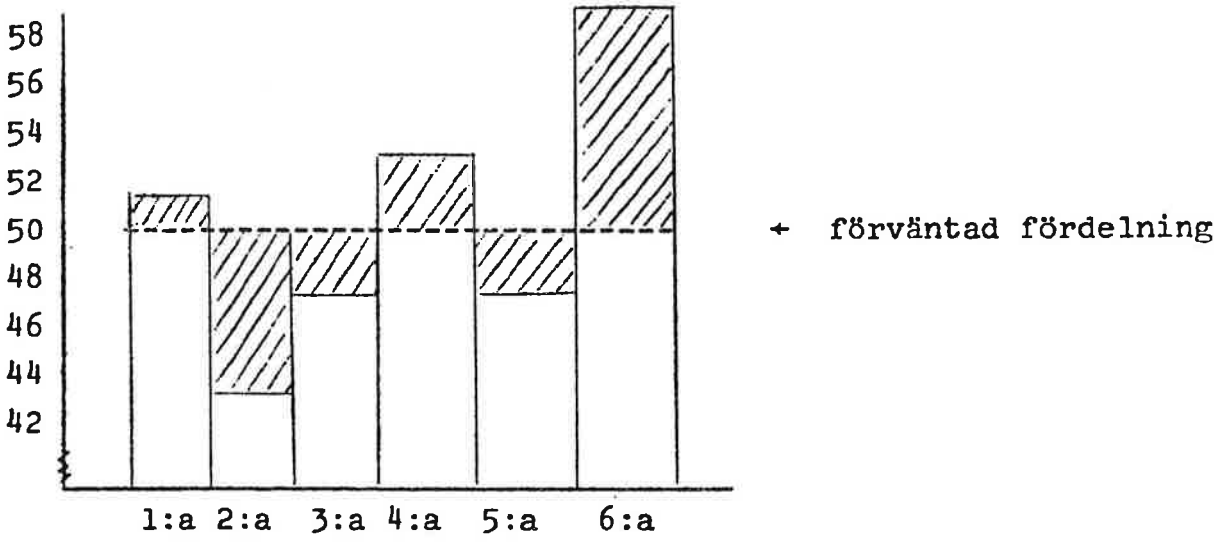
För att nu pröva huruvida den är fixad gör vi 300 kast med tärningen. Dessa ger följande fördelning:

etta	51
tvåa	43
trea	47
fyra	53
femma	47
sexa	59
N =	300

Tabellen ovan anger alltså den empiriska eller observerade fördelningen, vilken sedan skall jämföras med en förväntad , vilken i detta fall utgörs av en teoretisk fördelning. Denna teoretiska fördelning visar det utfall vi skulle få om kasten fördelade sig helt slumpmässigt på de sex olika utfallsmöjligheterna. Dvs. den teoretiska fördelningen beskriver det fall då det inte finns någon skillnad i frekvens mellan de sex utfallsmöjligheterna.

I vårt fall blir således den förväntade fördelningen att alla sex utfallsmöjligheterna skulle få frekvensen $\frac{300}{6} = 50$

Nu kan vi med hjälp av chi^2 studera huruvida den observerade och den förväntade (dvs. den teoretiska) fördelningen avviker från varandra och chi^2-värdet utgör ett sammanfattande mått på skillnaden mellan dessa två fördelningar. Om vi tar figuren nedan som utgångspunkt kan alltså chi^2 sägas vara ett sammanfattande mått på den streckade ytan som just anger skillnaden mellan den observerade och den förväntade fördelningen.



Vi angav ovan att chi^2 är ett sammanfattande mått på skillnaden mellan de två fördelningarna (den streckade ytan). Härav följer då att ju större skillnaden är desto högre värde antar chi^2. Detta framgår också av den statistiska definitionen av chi^2 (χ^2):

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

där: O = observerat värde
E = motsvarande förväntat värde

Formeln säger alltså att χ^2 är summan av de kvadrerade skillnaderna mellan de observerade värdena och motsvarande förväntade värden dividerat med respektive förväntat värde. Ju större skillnaden är mellan O och E desto högre värde måste alltså χ^2 få, som framgår av formeln.

Applicerar vi denna formel på det exempel vi tagit får vi följande tabell:

	Observerat (O)	Förväntat (E)	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
etta	51	50	1	1	0.02
tvåa	43	50	-7	49	0.98
trea	47	50	-3	9	0.18
fyra	53	50	3	9	0.18
femma	47	50	-3	9	0.18
sexa	59	50	9	81	1.62
N:	300	300	= 0		$\Sigma = 3.16 = \chi^2$

Innan χ^2-värdet beräknas är det lämpligt att lägga in ett par räknekontroller nämligen:

1) Totalen skall vara samma för observerade och förväntade värden. Givetvis kan i vissa fall smärre avvikelser förekomma p.g.a. avrundningsfel. Är avvikelserna stora föreligger däremot något räknefel.
2) Σ (O-E) skall alltid vara 0. Även här kan det förekomma små avrundningsfel, men större avvikelser från 0 innebär också räknefel.

I det förhållandet att summan av de okvadrerade avvikelserna alltid är noll har vi förklaringen till att dessa avvikelser måste kvadreras innan de summeras till χ^2. Detta förklarar också varför metoden kallas chi-*två*.

Vi återgår nu till vårt exempel med tärningen. Som vid andra statistiska analyser anger nollhypotesen att det inte finns någon skillnad mellan fördelningarna, medan den alternativa hypotesen anger att det föreligger en sådan skillnad. För att nu pröva huruvida nollhypotesen kan antagas eller förkastas måste vårt χ^2-värde

jämföras med de kritiska värden som finns angivna för olika signifikansnivåer i signifikanstabell 1 i slutet av kompendiet. Som signifikansnivå väljer vi 1 %-nivån. Nu inträffar emellertid den komplikationen att vi finner olika kritiska värden för olika antal frihetsgrader och innan vi går vidare måste vi utreda problemet med antal *frihetsgrader*.

Som vanligt gäller den regeln att antalet frihetsgrader anges av det antal värden som fritt kan variera. I vårt exempel har vi sammanlagt sex olika värden - ett för varje sida på tärningen. Totalen dvs. antalet kast är given. Hur många värden kan då variera? Jo, endast fem. Genom att totalen är given är det sjätte värdet låst - det måste vara lika med differensen mellan summan av de fem första värdena och totalen. Antalet frihetsgrader blir således fem, eftersom endast de fem första värdena kan variera fritt.

I vårt exempel har vi endast en kolumn av observerade värden och en med teoretiska värden, som utgör de förväntade, men i andra sammanhang, som vi kommer till senare, kan vi ha tabeller med flera kolumner av observerade värden. Ett exempel:

100 elever ur vardera socialgrupp I, II och III studerades med avseende på utbildningsvalet. (Detta var innan vi fick den nya, fina gymnasieskolan). Utfallet blev:

	I	II	III	Σ
gymnasium	80	50	25	155
fackskola	15	30	30	75
yrkesskola	5	15	25	45
ingen fortsatt utbildning	0	5	20	25
Σ	100	100	100	300

Vid beräkningen av antalet frihetsgrader utgår man alltid ifrån att såväl rad- som kolumn-summorna är givna. Börjar vi då med gymnasieväljarna så har vi tre socialgrupper men eftersom radsumman är given återstår det bara två värden som kan variera. Det tredje värdet måste bli resten, när summan av de två första värdena dragits från totalen. Vi förlorar således en frihetsgrad i raden för gymnasieväljarna. På samma sätt förloras en frihetsgrad

för vardera fackskole- och yrkesskoleväljarna. När vi sedan kommer till raden för dem som inte valt någon fortsatt utbildning inträffar en komplikation, nämligen den, att kolumnsummorna också är givna. Detta måste innebära att om vi känner till hur många av socialgrupp I-eleverna som valt gymnasium, fackskola respektive yrkesskola så måste antalet elever som inte valt någon fortsatt utbildning vara givet (dvs. resten). Vi förlorar således en frihetsgrad i kolumnen för socialgrupp I. På samma sätt förlorar vi också en frihetsgrad för vardera socialgrupp II och III.

Kontentan av detta blir att vi förlorar en frihetsgrad för varje rad(R) och en frihetsgrad för varje kolumn (C). Den generella formeln för antalet frihetsgrader (df) kan således skrivas:

$$df = (R-1)\ (C-1)$$

Vi går nu tillbaka till vårt tärningsexempel. För att förenkla vårt resonemang något kan vi säga att vi har två kolumner som jämförs med varandra, den för de observerade värdena och den för de teoretiska. Vi har samtidigt sex rader; en för varje tärningssida. Antalet frihetsgrader blir således:

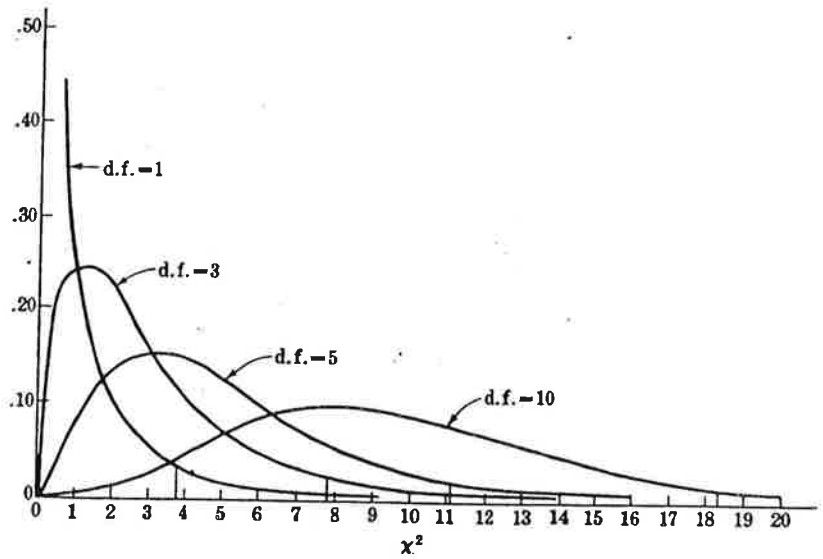
$$df = (6-1)\ (2-1) = 5.$$

För 5 frihetsgrader på 1 % signifikansnivå finner vi i χ^2-tabellen (signifikanstabell 1) ett kritiskt värde = 15.09, vilket skall jämföras med vårt χ^2-värde, som var 3.16. Vad anger då denna jämförelse? Jo, det funna värdet är lägre än tabellens kritiska värde. Vi har ju tidigare sagt att det funna χ^2-värdet är ett uttryck för skillnader och ökar när skillnaderna blir större vilket alltså innebär att de skillnader vi fick fram är mindre än de som det kritiska värdet ger uttryck för. Eftersom det kritiska värdet nu anger hur stora skillnaderna måste vara för att nollhypotesen skall förkastas är alltså den funna skillnaden för liten för att förkasta denna hypotes. Slutsatsen blir att det finns ingen statistiskt säkerställd differens mellan den empiriska fördelningen och den som vi utifrån slumpens inverkan skulle vänta oss. Det finns med andra ord inget belägg för att tärningen är fixad.

Vår framställning av jämförelsen mellan ett funnet och ett kritiskt χ^2-värde kan synas onödigt tillkrånglad. Tillkrånglandet har emellertid - paradoxalt nog - gjorts i de bästa avsikter nämligen att underlätta inlärning genom att leda fram till en förståelse av vad man gör, när man gör det.

Hela denna jämförelse kan emellertid sammanfattas i följande enkla tumregel: *För att förkasta nollhypotesen måste det funna χ^2-värdet vara minst lika högt som det kritiska.* (Tumregler är alldeles utmärkta om man bara kunde komma ihåg dem).

Varifrån kommer då alla kritiska χ^2-värden? Jo, från *CHI²-FÖRDELNINGEN*. Denna fördelning är en teoretisk fördelning precis på samma sätt som normalfördelningen och t-fördelningen m.fl. Egentligen utgör emellertid χ^2-fördelningen en hel grupp fördelningar med olika utseende beroende på antalet frihetsgrader, vilket framgår av nedanstående figur, där vi återger fördelningens form för några olika frihetsgrader.



Vi tänker oss att vi slumpmässigt drar ett oändligt antal z-värden från en normalfördelning och därefter kvadrerar vart och ett av dessa. Därefter lägger vi in det kvadrerade z-värdet på x-axeln och frekvenser av vart och ett av de erhållna kvadrerade z-värdena på y-axeln. Den fördelning vi då skulle få anger χ^2-fördelningen

för df = 1 i figuren ovan. Detta innebär att för df = 1 blir $\chi^2 \doteq z^2$, vilket framgår av att det kritiska χ^2-värdet på 5 %-nivån för df = 1 (3.84) är lika med kvadraten av det kritiska z-värdet på samma signifikansnivå: z = 1.96 ($1.96^2 \approx 3.84$). På samma sätt är det kritiska χ^2-värdet på 1 %-nivån för df = 1 (6.64) lika med normalfördelningens kritiska värde på 1 %-nivån taget i kvadrat ($2.58^2 \approx 6.64$).

Vi skall inte fördjupa oss i dessa tekniska resonemang utan det räcker med att konstatera att χ^2-fördelningen för df = 2 fås, genom att på samma sätt som förut dra 2 och 2 z-värden, kvadrera vart och ett av dessa och därefter summera de kvadrerade värdena och pricka av frekvenserna av dessa. För df = 3 drar man i stället 3 och 3 z-värden för df = 4 dras 4 och 4 z-värden osv. varefter man förfar på samma sätt som vid χ^2-fördelningen för df = 2.

Av mera central betydelse är emellertid det förhållandet att χ^2-värdet - tack vare att det är ett kvadrerat värde - alltid blir positivt. Härvidlag skiljer sig χ^2 således från z-testningen som kan ge såväl positiva som negativa värden. Det innebär således att vid z-testning har vi två svansar på fördelningen att ta hänsyn till: den positiva och den negativa och om testningen sker på 5 %-signifikansnivå innebär det en uppdelning så att, om den alternativa hypotesen ej är riktad, 2.5 % förläggs till vardera svansen. Är hypotesen däremot riktad vid signifikansprövning utifrån z förläggs alla 5 % till den ena svansen och det krävs således ett lägre z-värde för att nollhypotesen skall förkastas.

Tack vare att χ^2-värdet alltid är positivt måste emellertid denna signifikansprövning alltid ske med avseende på χ^2-fördelningens högra svans. Detta förhållande att χ^2 är ett ensvansat test har man emellertid tagit hänsyn till i tabellen över kritiska χ^2-värden och signifikansnivåerna är alltså de som anges för varje tabellkolumn, när alternativhypotesen inte är riktad. Det betyder att man vid en χ^2-testning bara får veta huruvida det finns några skillnader på den angivna signifikansnivån. Däremot säger denna signifikansprövning följaktligen inget om vilken riktning skillnaderna har. Av denna anledning är det också ovanligt att man arbetar med riktade hypoteser i samband med χ^2-testningar. För att tolka

resultaten och i samband därmed få veta vilken riktning de eventuella skillnaderna har måste man gå tillbaka till den ursprungliga tabellen.

Skulle det trots allt vara angeläget att arbeta med riktade alternativhypoteser är detta också möjligt vid χ^2-analys. Då förfar man på samma sätt som vid z-testningar: man gör helt enkelt den del av svansen, där H_o förkastas dubbelt så stor vilket innebär att det kritiska värdet för χ^2 på 5 %-nivån anges av de värden som i signifikanstabellen står för 10 %-nivån och det kritiska värdet på 1 %-nivån vid en riktad hypotes anges i denna tabell på samma sätt av värdet på 2 %-nivån.

2.3 Chi²: Goodness of fit

I det exempel vi arbetat med hittills och som rört tärningskasten har vi endast en uppsättning observerade värden nämligen de som blev resultatet av våra 300 kast. Denna fördelning jämfördes med en teoretisk fördelning dvs. en fördelning som vi inte fått fram empiriskt. Vi har visserligen två fördelningar som jämförs med varandra, men eftersom den ena är en teoretisk fördelning och inte en fördelning av uppmätta värden så räknas den inte, utan vid valet av metod betraktas ett sådant här fall som att vi har endast *en* uppsättning mätvärden. Hade vi varit riktigt noggranna skulle det i översiktstabellen hetat antal uppsättningar *empiriska* mätvärden.

Den chi²-metod som är tillämpbar i ett sådant här fall är den som kallas goodness of fit. En svensk översättning av det engelska ordet "fit" är "passa" eller "överensstämma". Denna metod avser alltså frågan huruvida en viss empirisk fördelning överensstämmer med en teoretisk fördelning och är således tillämpbar när vi har endast en uppsättning av empiriska mätvärden och data är uttryckta i nominalskala. Ett annat mycket vanligt exempel på användning av "goodness of fit" är när vi studerar överensstämmelsen i vissa variabler t.ex. socialgruppstillhörighet mellan ett urval försökspersoner och den population de representerar. Då är det inte ovanligt att man t.ex. känner till populationens fördelning på social-

grupper. Denna fördelning utgör då den teoretiska fördelningen och vi jämför undersökningsgruppens fördelning med denna teoretiska.

Försök finna några andra undersökningssituationer i vilka "goodness of fit" är tillämpbar.

Tillvägagångssättet vid "goodness of fit" är exakt det som vi angivit i samband med tärningsexemplet varför vi inte upprepar detta.

2.4 Chi²: Tests of independence

I de fall då vi har två eller flera uppsättningar av mätvärden, vilka är oberoende skall vi i stället välja den chi²-metod som kallas tests of independence. Det engelska ordet independence kan översättas med "oberoende" eller "oavhängighet". Man prövar således om två eller flera fördelningar kan anses vara oavhängiga av varandra eller med andra ord om de skiljer sig från varandra.

Vad som skiljer denna metod från goodness of fit är det sätt på vilket vi får fram de förväntade värdena (E). Vid goodness of fit erhölls ju E direkt utifrån en teoretisk fördelning eller utifrån en redan känd fördelning. När vi använder tests of independence får vi i stället räkna fram de förväntade värdena med hjälp av de observerade. Låt oss ta ett exempel för att belysa detta:

Vi vill studera manliga och kvinnliga sjuksystrars inställning till sitt arbete. 100 kvinnliga och 100 manliga systrar tillfrågas om huruvida de upplever sitt arbete positivt eller negativt. Utfallet blir:

		Systrar		
		Kvinnliga	Manliga	Σ
Inställning	Positiv	A 80	B 40	120
	Negativ	C 20	D 60	80
	Σ	100	100	200

Bokstäverna utgör beteckningar på tabellens fyra celler.

Finns det någon skillnad mellan de kvinnliga och de manliga systrarnas inställning? (Testningen görs på 1 %-nivån).

Första steget blir nu att ta reda på de förväntade värdena (E) för respektive cell. En viktig sak att komma ihåg är, att vid tests of independence beskriver de förväntade värdena alltid nollhypotesen - dvs. det utfall då det inte finns några skillnader mellan de studerade grupperna. För att vi skall få fram dessa E-värden måste vi följaktligen utgå ifrån rad- och kolumnsummorna - de enskilda cellerna inne i tabellen kan ju uppvisa stora skillnader och är således inte användbara - och vi börjar med att beräkna E-värdet för den cell som avser de positiva, kvinnliga systrarna (cell A):

Första steget är då att ta reda på hur stor andel av hela undersökningsgruppen som är kvinnor. Jo, det är $\frac{100}{200} = 50$ %. Alla i undersökningsgruppen är emellertid inte positiva, vilket framgår av summakolumnen, varför vi också måste beräkna andelen positiva i hela materialet, vilken är $\frac{120}{200} = 60$ %. För att nu få veta hur stor andel av kvinnorna som skulle vara positiva, om det inte fanns någon skillnad mellan de kvinnliga och de manliga systrarna, måste vi multiplicera 50 % med 60 % vilket är 30 %. Fanns det ingen skillnad mellan könen skulle alltså 30 % av hela undersökningsgruppen hamna i cell A vilket blir 30 % · 200 = 60 individer.

När vi, som i detta fall, har en 2×2-tabell skulle man kunna sluta beräkningen av E-värden här. Genom att man vet värdet i cell A kan man direkt få fram E-värdena för de övriga cellerna helt enkelt genom att beräkna differensen mellan det kända E-värdet och respektive rad- och kolumnsumma. Detta är ett exempel på vad vi tidigare sagt om frihetsgrader nämligen att vi förlorar 1 df på varje rad respektive kolumn. En sådan genväg är dock inte att rekommendera eftersom man då går miste om den kontroll som innebär att Σ (O-E) alltid är 0. Går man denna genväg och råkar räkna fel på det första förväntade värdet blir nämligen alla andra E-värden också fel men Σ (O-E) blir fortfarande 0.

Ett litet ord på vägen: Inte ens den mest raffinerade statistiska metod kan korrigera för räknefel.

Innan vi går vidare skall vi snabbt rekapitulera vad vi gjorde när vi beräknade E-värdet för cell A:

Först dividerade vi summan för kvinnokolumnen med totalen: $\frac{100}{200}$
därefter dividerade vi "positiv"-radens summa med totalen: $\frac{120}{200}$
och till slut multiplicerade vi dessa båda värden med varandra och med totalen. Alltså:

$$\frac{100}{200} \cdot \frac{120}{200} \cdot 200$$

Detta uttryck kan ju också skrivas:

$$\frac{100 \cdot 120}{200 \cdot 200} \cdot 200$$

Förkortar vi i detta uttryck får vi:

$$\frac{100 \cdot 120}{200}$$

Och vad är det? Jo, om vi går tillbaka till vår utgångstabell finner vi att vi helt enkelt har multiplicerat kolumnsumman för cell A (100) med denna cells radsumma (120) och dividerat med totalen (200). Detta innebär att vi har kommit fram till en genväg för beräkning av E-värden som i tumregelform kan uttryckas:

Multiplicera cellens radsumma med dess kolumnsumma och dividera därefter med totalen.

Applicerar vi nu denna tumregel på de övriga cellerna får vi följande förväntade värden

cell B (positiva manliga): $\frac{120 \cdot 100}{200} = 60$

cell C (negativa kvinnliga): $\frac{80 \cdot 100}{200} = 40$

cell D (positiva kvinnliga): $\frac{80 \cdot 100}{200} = 40$

Vi har nu 2 tabeller en för de observerade värdena och en för de förväntade värdena:

Observerade

	kvinnl.	manl.	Σ
Pos.	A 80	B 40	120
Neg.	C 20	D 60	80
Σ	100	100	200

Förväntade

	kvinnl.	manl.	Σ
Pos.	A 60	B 60	120
Neg.	C 40	D 40	80
Σ	100	100	200

Som framgår av tabellen för förväntade värden beskriver den det fall då det inte finns några skillnader dvs. nollhypotesen och detta är, som vi tidigare sagt, alltid förhållandet vid tests of independence.

Nu kan vi fortsätta beräkningarna på exakt samma sätt som vid goodness of fit:

	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
cell A	80-60= 20	400	$\frac{400}{60} = 6.67$
cell B	40-60= -20	400	$\frac{400}{60} = 6.67$
cell C	20-40= -20	400	$\frac{400}{40} = 10.00$
cell D	60-40= 20	400	$\frac{400}{40} = 10.00$
Σ	0		$\chi^2 = 33.34$

χ^2-värdet blir således 33.34, och detta värde skall jämföras med det kritiska värdet för 1 df på 1 %-nivån, vilket signifikanstabellen visar vara 6.64. Vårt funna χ^2-värde är alltså större än det kritiska värdet vilket innebär att det finns en skillnad mellan kvinnliga och manliga systrars inställning till sitt arbete. Däremot säger χ^2-värdet inget om vilken riktning dessa skillnader har dvs. huruvida de kvinnliga eller de manliga systrarna är mest positiva. För att få veta skillnadernas riktning kan man gå tillbaka till tabellen för observerade värden men ännu enklare är det att studera värdena för O-E. Av beräkningarna ovan framgår då att de observerade värdena överstiger de förväntade i cellerna A och D medan de observerade understiger de förväntade i cellerna B och

C. Detta innebär att vi har "för många" kvinnliga positiva och manliga negativa. Således är de kvinnliga systrarna mera positiva till sitt arbete än de manliga.

För att inte onödigt förlänga framställningen har vi valt att utgå från ett exempel som uttrycks i en 2×2 tabell. Tillvägagångssättet är emellertid exakt detsamma vid större tabeller. Vi får i det senare fallet bara flera celler att arbeta med och därmed också ett högre antal frihetsgrader.

För en 2×2-tabell kan χ^2 emellertid beräknas med hjälp av en alternativ formel, vilken har den fördelen att man endast går på de observerade värdena och således slipper att räkna fram de förväntade.

Vi utgår från vårt exempel med de kvinnliga och manliga sjuksköterskorna:

	kvinnlig	manlig	Σ
Positiv	A 80	B 40	A+B 120
Negativ	C 20	D 60	C+D 80
Σ	A+C 100	B+D 100	N 200

Formeln: $\chi^2 = \frac{N(A \cdot D - B \cdot C)^2}{(A+B)(C+D)(A+C)(B+D)}$

Vi utgår således från de fyra cellerna A-D. Dessa multipliceras med varandra korsvis och differensen mellan produkterna kvadreras. Detta värde multipliceras med totalen varefter vi dividerar med produkten av marginalsummorna.

I vårt exempel får vi:

$$\chi^2 = \frac{200(80 \cdot 60 - 40 \cdot 20)^2}{120 \cdot 80 \cdot 100 \cdot 100} = \frac{200(4800-800)^2}{120 \cdot 80 \cdot 100 \cdot 100} = \frac{200 \cdot 16000000}{120 \cdot 80 \cdot 100 \cdot 100} = 33.33$$

Som framgår av vårt exempel är denna alternativa formel en snabbare väg fram till χ^2-värdet än den vi angav först. Dock kan, som vi

angav ovan, denna alternativa formel endast användas på en 2×2-tabell.

2.5 Chi² och små förväntade värden

De kritiska värden som finns angivna i signifikanstabellen för chi² gäller under den förutsättningen att de förväntade värden, som vi arbetar med, inte är alltför låga. Vad som menas med uttrycket "alltför låga" råder det olika uppfattningar om. Ferguson (1966) anger som riktvärden att vid 1 df - dvs. vid en 2×2-tabell - får ingen förväntad frekvens understiga 5. Vid 2 eller flera frihetsgrader kan man vara mera liberal och han anger i detta fall som lägsta förväntad frekvens 2. Alla har emellertid inte samma uppfattning som Ferguson. Vid df = 1 hävdar t.ex. vissa att inget värde får understiga 10 och Siegel (1956) anger att för 2 frihetsgrader skall varje förväntat värde vara lägst 5 medan för fler än 2 frihetsgrader högst 20 % av de förväntade värdena får vara lägre än 5 och inget mindre än 1.

Vad gör man då om man trots allt har förväntade värden som understiger de lägsta tillåtna?

Har man df = 1 - dvs. en 2×2-tabell - och om man använder den alternativa χ^2-formeln - dvs. den som går direkt från de observerade värdena - kan man bygga in en korrektion direkt i formeln. Denna korrektion innebär att man reducerar differensen A·D-B·C (se formeln) med $\frac{N}{2}$ och vi får då följande fullständiga formel:

$$\chi^2 = \frac{N(|A\cdot D-B\cdot C|-\frac{N}{2})^2}{(A+B)(C+D)(A+C)(B+D)}$$

Uttrycket: |A·D-B·C| innebär att man tar den absoluta differensen dvs. man bryr sig inte om vilket tecken differensen har. Korrektionen innebär således att vi alltid reducerar differensen AD-BC med $\frac{N}{2}$ och eftersom detta står över bråkstrecket kommer således χ^2-värdet att reduceras genom korrektionen.

Har man däremot flera frihetsgrader än 1, kan man inte göra en sådan korrektion. Då får man i stället motverka de låga förvänta-

de värdena genom att slå samman svarskategorier och/eller grupper, vilket innebär att vi reducerar tabellens storlek och därmed också antalet frihetsgrader.

När man tvingas till sammanslagningar bör man i första hand sträva efter att föra ihop sådana svarskategorier eller grupper som har något gemensamt t.ex. tillhör en och samma övergripande kategori. Kan man inte finna något gemensamt för de olika svarskategorier respektive grupper som måste slås samman gör man sammanslagningen ändå och döper den till "övriga". När sedan resultatet av den statistiska analysen skall tolkas måste det alltid ske utifrån den tabell man fått fram *efter* sammanslagningen och på vilken analysen baseras.

En allmän regel i dessa sammanhang är emellertid att man skall göra så få sammanslagningar som möjligt. Sammanslagningarna leder ju till att vi får grövre och mer diffusa kategorier varigenom vi går miste om information.

2.6 Fishers exakta sannolikhetstest

Om vi har data i en 2×2-tabell men frekvenserna är så låga att chi^2 tests of independence är utesluten som analysmetod kan vi i stället använda oss av Fishers exakta sannolikhetstest. Denna metod leder direkt fram till en exakt sannolikhet för en differens och går således inte omvägen över χ^2 eller någon annan teoretisk fördelning.

Vi konkretiserar tillvägagångssättet genom att ta ett exempel. *Antag att vi har ett antal skolklasser som har kategoriserats som homogena respektive heterogena klasser. Dessa klasser har genomgått 8 olika kunskapsprov i vardera naturvetenskapliga och samhällsvetenskapliga ämnen.* Vi vill nu studera på 5 %-nivån om klassens homogenitet inverkar olika på kunskaperna i naturvetenskapliga och samhällsvetenskapliga ämnen. Med andra ord om det finns en interaktion mellan klasshomogenitet och ämnestyp.

När det gäller naturvetenskapliga ämnen blir de homogena klasserna bäst i 6 av de 8 proven och de heterogena klasserna blir bäst i 1. Det återstående provet ger ingen skillnad.

I de samhällsvetenskapliga ämnena blir de heterogena klasserna bäst i 5 av proven och de homogena i 1. 2 prov ger ingen skillnad.

Vi får nu följande tabell; där vi också lagt in cellbeteckningarna.

	naturvet.	samhällsvet.	Σ
Homogena bäst	A 6	B 1	A+B 7
Heterogena bäst	C 1	D 5	C+D 6
Σ	A+C 7	B+D 6	N 13

Formeln för Fishers exakta sannolikhetstest är:

$$p = \frac{(A+B)!(C+D)!(A+C)!(B+D)!}{N!A!B!C!D!}$$

Utropstecknen anger att värdet skall tas i fakultet: 4! = 4·3·2·1= = 24. 0! är alltid 1.

Sätter vi in vårt resultat i tabellen ovan får vi följande:

$$p = \frac{7!6!7!6!}{13!6!1!1!5!} = \frac{7\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1\cdot7\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1}{13\cdot12\cdot11\cdot10\cdot9\cdot8\cdot7\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1\cdot1\cdot1\cdot5\cdot4\cdot3\cdot2\cdot1} =$$

$$= \frac{7}{286} = 0.024$$

De skillnader som tabellen ovan visar har alltså en sannolikhet av 0.024 att uppkomma av en slump. En statistisk analys som ger ett så lågt p-värde innebär ju normalt att nollhypotesen förkastas och att man antar den alternativa hypotesen. Det kan vi emellertid inte göra i detta fall eftersom den statistiska prövningen av nollhypotesen alltid ställer frågan: Vad är sannolikheten att av en slump få den uppkomna differensen *eller en större differens?*

Till skillnad från övriga statistiska metoder anger Fishers exakta sannolikhetstest bara sannolikheten för den uppkomna differensen men det erhållna p-värdet inkluderar däremot inte de mera extrema utfallen dvs. de utfall som ger större skillnad.

Om vi nu går tillbaka till ursprungstabellen ser vi att med bibehållna summavärden i tabellen finns det ett utfall som är mer extremt, än det undersökningen gav, nämligen:

	naturvet.	samhällsvet.	Σ
Homogena bäst	7	0	7
Heterogena bäst	0	6	6
Σ	7	6	13

Sannolikheten för detta mer extrema utfall får vi på samma sätt som föregående:

$$p = \frac{7!6!7!6!}{13!7!0!0!6!} =$$

$$= \frac{7\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1\cdot7\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1}{13\cdot12\cdot11\cdot10\cdot9\cdot8\cdot7\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1\cdot7\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1\cdot1\cdot1\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1} =$$

$$= 0.001$$

Vi har nu två olika sannolikheter: den för det resultat, som vår undersökning gav (p=0.024) och den för det mer extrema resultatet (p=0.001). Eftersom den statistiska prövningen, som vi tidigare sagt, alltid gäller den sammanlagda sannolikheten av den erhållna skillnaden och större skillnader måste dessa två p-värden summeras och vi får då p=0.024+0.001=0.025.

Resultatet av vår undersökning blir alltså att de kunskapsskillnader vi funnit mellan de båda klasstyperna i de naturvetenskapliga contra de samhällsvetenskapliga proven är så stora att dessa skillnader - eller större skillnader - har en mindre sannolikhet att uppkomma av en slump än 0.05 dvs. signifikansnivån. Vi måste således förkasta nollhypotesen. Vår slutsats blir att klassens homogenitet inverkar olika på kunskaperna i naturvetenskapliga och samhällsvetenskapliga ämnen såtillvida att homogena klasser är bättre i de naturvetenskapliga ämnena medan de heterogena klasserna uppvisar bättre kunskaper i samhällsvetenskapliga ämnen.

Hade vår första beräkning - den på de empiriskt erhållna värdena - givit ett p-värde som överstiger 0.05 hade vi givetvis kunnat avbryta beräkningarna redan på det stadiet och antagit nollhypotesen.

De metoder vi behandlat hittills har varit sådana som är tillämpbara när vi har endast en uppsättning mätvärden (X^2: goodness of fit), när vi har två oberoende uppsättningar (X^2: tests of independence respektive Fishers exakta sannolikhetstest) samt när vi har fler än två uppsättningar av oberoende mätvärden (X^2: tests of independence). Vi går i stället över till att behandla två metoder som är tillämpbara på beroende mätvärden med två respektive fler än två uppsättningar av mätvärden. Fortfarande gäller att data är uttryckta i nominalskala.

2.7 Mc Nemars test för signifikans av systematiska förändringar

Detta test är tillämpbart när vi har två uppsättningar av beroende mätvärden. Som vi angav i kap. I är mätvärdena beroende om de härrör från en och samma undersökningsgrupp som studerats vid två tillfällen eller om data avser matchade grupper.

En typisk tillämpning av Mc Nemars test är vid s.k. före-efter-undersökningar. Vi tar ett exempel:

50 nybörjare på C1 i pedagogik tillfrågades vid terminsstarten om sin inställning till statistik. 5 var positiva medan 45 var negativa. Efter att ha genomgått statistikkursen på C1:1 tillfrågades de ånyo. Nu var emellertid 40 positiva medan endast 10 var negativa. Utav de som var negativa innan kursen hade 38 ändrat uppfattning och blivit positiva och bland dem som var positiva före kursen hade 3 ändrat sig och blivit negativa. Har denna kurs påverkat de studerandes inställning till statistik? Vi prövar på 5 % signifikansnivå.

Data från undersökningen kan sammanställas i följande fyrfältstabell:

		Efter kursen: negativa	Efter kursen: positiva	
Före kursen	positiva	3	2	5
	negativa	7	38	45
		10	40	50

Metoden heter ju Mc Nemars test för signifikans av systematiska *förändringar*, vilket innebär att det är förändringarna som är av intresse här. I vårt exempel finner vi dem som förändrat sin inställning i cellerna A och D. Det är alltså dessa två värden som skall studeras.

Nollhypotesen är att kursen inte har påverkat de studerandes inställning till statistik på ett systematiskt sätt. Den alternativa hypotesen blir då att kursen haft en systematisk påverkan.

Om nollhypotesen är riktig bör det sålunda vara lika många studerande som ändrat sig från positiv till negativ inställning som de, vilka ändrat sig i motsatt riktning. Det betyder att frekvenserna i cell A och D skall vara lika höga.

Totalt har 3+38=41 studerande ändrat inställning. Vad vi skulle väntat oss enligt nollhypotesen är således att $\frac{3+38}{2} = 20.5$ studerande skulle ändrat sig åt varje håll.

Den statistiska prövningen innebär att vi skall pröva om de observerade värdena i cellerna A och D skiljer sig från de förväntade värden, som anges av nollhypotesen. Det är således samma problematik som vi tidigare mött i samband med χ^2 - och Mc Nemars test kan alltså betraktas som en variant av chi^2-metoderna.

Nu vet vi från tidigare att $\chi^2 = \Sigma \frac{(O-E)^2}{E}$ varvid E motsvarar det förväntade värdet enligt nollhypotesen (i vårt exempel: 20.5) och O är de empiriska värdena i cell A och D.

Sätter vi in bokstavsbeteckningarna får vi följande:

$$\chi^2 = \frac{(A-\frac{A+D}{2})^2}{\frac{A+D}{2}} + \frac{(D-\frac{A+D}{2})^2}{\frac{A+D}{2}}$$

Förkortar vi detta får vi:

$$\chi^2 = \frac{(A-D)^2}{A+D}$$

En sak att observera är att det i vår tabell råkar vara cellerna A och D som anger förändringen. Hade tabellen ställts upp på ett annat sätt hade i stället cellerna B och C varit de som angav förändring och då hade naturligtvis formeln blivit:

$$\chi^2 = \frac{(B-C)^2}{B+C}$$

Sätter vi in våra resultat i formeln får vi:

$$\chi^2 = \frac{(3-38)^2}{3+38} = \frac{1225}{41} = 29.88$$

Eftersom vi har en 2×2-tabell blir frihetsgraderna = 1 och av tabellen över kritiska χ^2-värden framgår det att på 5 %-nivån blir det kritiska värdet 3.84. Det föreligger således en klar skillnad mellan studenternas inställning till statistik efter kursen jämfört med före den. Av ursprungstabellen framgår nu att skillnaden innebär en förändring mot en mera positiv inställning.

Liksom vid chi^2 tests of independence måste vi här göra en korrektion om det förväntade värdet är lågt. Korrektionen innebär i detta fall att vi reducerar differensen mellan de observerade värdena med 1. Vi får då en formel enligt följande:

$$\chi^2 = \frac{(|A-D|-1)^2}{A+D}$$

2.8 Cochrans Q test

Mc Nemars test var alltså tillämpligt när vi har två uppsättningar av beroende mätvärden. Har vi flera än två uppsättningar (dvs. k uppsättningar) beroende mätvärden måste vi välja en annan metod nämligen Cochrans Q test vilket egentligen är en vidareutveckling av Mc Nemars test.

Den undersökningssituation då Q-testet är användbart innebär att vi har en undersökningsgrupp från vilken vi erhåller fler än två uppsättningar mätvärden. De enskilda mätvärdena förutsättes vara uttryckta i en dikotom skala dvs. de kan anta endera av två värden t.ex. rätt-fel, ja-nej osv.

I stället för att göra upprepade mätningar på en och samma undersökningsgrupp kan vi naturligtvis ha flera olika grupper, vilka är matchade, varvid vi gör en mätning på varje grupp.

Ett exempel: *Vi vill studera hur en grupp elevers inställning till ett visst undervisningsmaterial förändras allteftersom den tid de hållit på med materialet i fråga. Eleverna får ange sin inställning till materialet efter 5, 10 och 15 timmars arbete med det och deras inställning registreras som antingen positiv eller negativ - alltså en dikotom skala. Som signifikansnivå väljer vi 5 %.*

Resultatet blev:

	Genomförd arbetstid				
Elev	5 tim.	10 tim.	15 tim.	L	L^2
A	1	1	0	2	4
B	1	0	0	1	1
C	1	1	1	3	9
D	1	0	0	1	1
E	0	1	1	2	4
F	1	0	0	1	1
G	1	1	0	2	4
H	0	0	1	1	1
I	0	0	0	0	0
J	1	1	0	2	4
	G_5= 7	G_{10}= 5	G_{15}= 3	ΣL=15	ΣL^2=29

1 = positiv inställning
0 = negativ inställning

I tabellen ovan har vi också lagt in kolumnsummor (G_5, G_{10} och G_{15}) och radsummorna samt dessa i kvadrat: L respektive L^2. De senare skall också summeras, vilket gjorts i tabellen.

Vi har härvid valt att arbeta med de värden som anger positiv inställning dvs. 1-orna i tabellen. Självfallet hade vi lika väl kunnat utgå från 0-orna (negativ inställning) men då blir naturligtvis tolkningen av eventuella signifikanser den motsatta.

Vi har nu alla värden, så när som på två, vilka behövs för att räkna fram ett Q-värde. Det ena värdet, som saknas, är antalet uppsättningar av mätvärden, vilket i formeln nedan betecknas: "k". I vårt fall har vi 3 sådana uppsättningar ett för vardera 5, 10 och 15 timmars arbete (G_5, G_{10} och G_{15}). Det andra saknade värdet är $\overline{G}$, vilket är medelvärdet för de olika G-värdena. I vårt exempel:

$$\overline{G} = \frac{7+5+3}{3} = 5.$$

Vi kan således nu gå över till att beräkna Q-värdet vilket erhålls enligt följande formel:

$$Q = \frac{k(k-1)\Sigma(G-\overline{G})^2}{k\cdot\Sigma L-\Sigma L^2}$$

Vad vi prövar med Cochrans Q test är ju om det finns någon skillnad mellan de olika G-värdena. (I vårt exempel: om det finns någon skillnad mellan antalet positiva elever efter 5, 10 respektive 15 timmars arbete). Om man går till formeln ovan anges denna skillnad av uttrycket $\Sigma(G-\overline{G})^2$, vilket kan sägas vara en motsvarighet till mellangruppskvadratsumman i variansanalys. Ju större skillnaden är mellan de olika G-värdena desto högre blir uttrycket $\Sigma(G-\overline{G})^2$ och därmed får vi också ett högt Q-värde eftersom $\Sigma(G-\overline{G})^2$ står i täljaren.

Vi går tillbaka och beräknar Q för vårt exempel:

$$Q = \frac{3(3-1)\ [(7-5)^2+(5-5)^2+(3-5)^2]}{3\cdot 15-29} = \frac{3\cdot 2(4+0+4)}{3\cdot 15-29} = \frac{48}{16} = 3.00$$

Fördelningen av Q-värden följer chi^2-fördelningen för k-1 frihetsgrader och tolkningen är exakt densamma som vid en chi^2-testning. Vår Q-värde skall således jämföras med det kritiska värdet för chi^2 med 2 frihetsgrader på 5 %-nivån, vilket är 5.99. Då det funna Q-värdet är lägre än det kritiska chi^2-värdet måste alltså nollhypotesen antas och denna anger ju att det inte finns några skillnader. Slutsatsen för vårt exempel blir således att elevernas inställning till undervisningsmaterialet förändrades inte med den tid de arbetade med materialet.

2.9 Kontingenskoefficienten

Vi har nu gått igenom olika metoder för att signifikanspröva skillnader när data uttrycks i nominalskala. Av de metoder som finns upptagna i översiktstabellen (kap. I) återstår således bara sambandsmåttet, vilket på denna skalnivå är kontingenskoefficienten (C). Liksom de flesta övriga metoder som är tillämpbara på ordinalskalenivån har också C en nära anknytning till chi^2, vilket framgår av dess formel:

$$C = \sqrt{\frac{\chi^2}{N+\chi^2}}$$

C beräknas således utifrån ett erhållet χ^2-värde och det N som detta χ^2-värde är baserat på. Att N ingår i formeln är värt en mässa. Här kommer vi nämligen tillbaka till en problematik som diskuterades i kap. I i samband med metodernas känslighet. Den på kap. I välinläste minns då att en metods känslighet kan ökas genom att göra undersökningsgruppen större. Detta kan exemplifieras på ett enkelt sätt med hjälp av χ^2. Vi tar ett exempel.

Statens konsumentinstitut ville undersöka män och kvinnors inställning till att män använder nattskjorta. För att vidare studera om de eventuella könsskillnaderna är samma i stad och på landsbygd tas 2 undersökningsgrupper ut - en för vardera regionen. Resultatet blev:

Landsbygd

	pos.	neg.	Σ
män	10	20	30
kvinnor	10	10	20
	20	30	50

Stad

	pos.	neg.	Σ
män	30	60	90
kvinnor	30	30	60
	60	90	150

Eftersom vi har data i nominalskala och 2 oberoende uppsättningar värden väljer vi naturligtvis chi^2-tests of independence, varvid χ^2-värdena blir:

$$\chi^2_{\text{landsbygd}} = 1.39$$
$$\chi^2_{\text{stad}} = 4.17$$

Det kritiska värdet för χ^2 med 1 df och på 5 %-nivån är 3.84, varför vi drar den slutsatsen att det inte finns någon könsskillnad bland landsbygdsbefolkningen men väl bland stadsbor såtillvida att män är mer negativa till nattskjorta än kvinnor är i städer.

Denna slutsats är emellertid märklig för om vi går tillbaka till tabellerna finner vi att den relativa fördelningen av svar är exakt densamma. Enda skillnaden är att frekvenserna i stadstabellen är 3 gånger så höga som de i landsbygdstabellen. De olika utfallen av våra signifikansprövningar måste således bero på att vi har en 3 gånger så stor undersökningsgrupp i staden. Detta kan också uttryckas så att med den mindre undersökningsgruppen har χ^2-metoden för låg känslighet för att ge utslag för de befintliga skillnaderna. När gruppen görs större ökas metodens känslighet så att den förkastar den falska nollhypotesen.

χ^2-värdets beroende av N får också den konsekvensen att vi inte kan uttala oss om huruvida könsskillnaderna är större i den ena regionen än i den andra. För att göra en sådan jämförelse kan vi gå vidare utifrån χ^2-värdet och beräkna en kontingenskoefficient för vardera regionerna. C:s storlek påverkas nämligen inte av N.

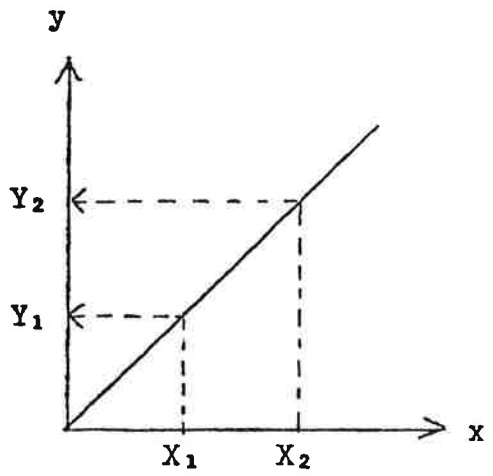
$$C_{landsbygd} = \sqrt{\frac{1.39}{50+1.39}} = \sqrt{0.027} = 0.164$$

$$C_{stad} = \sqrt{\frac{4.17}{150+4.17}} = \sqrt{0.027} = 0.164$$

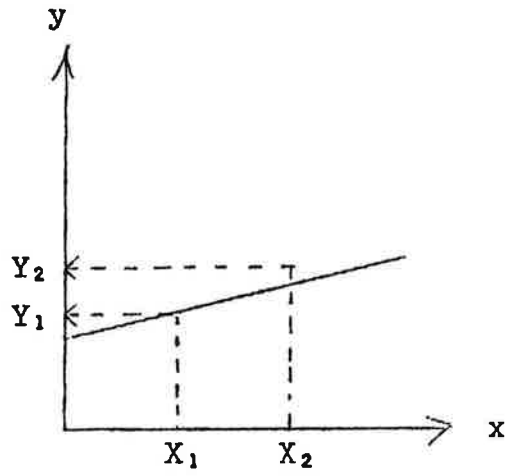
Kontingenskoefficienterna anger således att vi har ett lika högt samband mellan kön och inställning till nattskjorta i såväl stad som landsbygd. Att χ^2-värdena blev olika höga måste således bero på den varierande gruppstorleken och inte på att könsskillnaderna var olika stora i de två regionerna.

I vårt resonemang ovan kan det förefalla att vi varit inkonsekventa såtillvida att vi samtidigt talat om skillnader och samband. Så är det emellertid inte eftersom dessa två begrepp står för en och samma sak. Om vi tar produktmomentkorrelationen som exempel så anger ju en signifikant sådan korrelation att två olika x-värden motsvaras av 2 skilda värden i y-variabeln. Med andra ord re-

gressionslinjen lutar. Är produktmomentkorrelationen mycket låg har vi ju en i det närmaste vågrät regressionslinje, vilket innebär att olika x-värden motsvaras av i stort sett samma y-värde dvs. små skillnader i y-variabeln. Vi kan åskådliggöra detta med följande figurer:



En klar skillnad mellan Y_1 och Y_2: alltså en relativt hög korrelation.



En liten skillnad mellan Y_1 och Y_2 alltså en låg korrelation.

I vårt exempel med nattskjortorna motsvaras x-variabeln av kön och y-variabeln av inställning till nattskjortor. Det betyder att i vårt exempel är både x- och y-variabeln dikotoma men det principiella resonemanget blir fortfarande det samma som det vi förde ovan i anslutning till produktmomentkorrelationen. Alltså: finns det en skillnad mellan könen i deras inställning till nattskjortor så kan det lika väl uttryckas så att det finns ett samband mellan kön och inställning till nattskjortor.

Efter dessa utsvävningar i principiella resonemang begränsar vi oss till kontingenskoefficienten, som alltså kan sägas vara ett sambandsmått för variabler uttryckta i nominalskala. Den egenskapen att C kan användas när båda variablerna är i nominalskala har också phi-koefficienten men i motsats till phi kan C användas också för större tabeller än en 2×2-tabell. Eftersom C således alltid kan tillämpas i samband med chi^2-analyser oavsett antalet celler i tabellen får den ett stort användningsområde.

Signifikansprövningen av C är mycket enkel och sker via det i formeln ingående χ^2-värdet. Är detta värde signifikant är också C signifikant skilld från 0. För att anknyta till vårt resonemang ovan: är skillnaden signifikant är också sambandet signifikant. Däremot är det en mycket komplicerad process att signifikantpröva skillnaden mellan två C-värden varför vi inte tar upp detta här.

Det lägsta värde C kan anta är 0. Dvs. C är alltid positivt vilket är naturligt med hänsyn till att vi har data i nominalskala och att vi således inte kan ordna data i någon form av rangordning. Kontingenskoefficientens högsta möjliga värde är däremot beroende av ursprungstabellens storlek. När tabellen har lika många rader och kolumner kan C_{max} beräknas utifrån följande formel:

$$C_{max} = \sqrt{\frac{R-1}{R}} \quad (R= \text{antal rader resp. kolumner})$$

för en 2×2-tabell blir C_{max} således $\sqrt{\frac{1}{2}} = .707$ och för en 3×3-tabell blir C_{max} $\sqrt{\frac{2}{3}} = .816$.

Att C_{max} på detta sätt varierar med tabellens storlek medför att olika C inte blir direkt jämföra om de baseras på tabeller av olika storlek.

En annan begränsning hos C är att den förutsätter en χ^2-beräkning. Som vi tidigare angivit har χ^2 speciella krav på de förväntade frekvensernas storlek. Är inte dessa krav uppfyllda kan följaktligen inte heller C beräknas.

Slutligen har kontingenskoefficienten ytterligare en nackdel nämligen att dess värde inte är direkt jämförbart med andra korrelationsvärden.

III ICKE-PARAMETRISKA METODER FÖR DATA UTTRYCKTA I ORDINALSKALA

De metoder som redovisas i kap. II och som är tillämpbara på data i nominalskala kan visserligen användas också på data som uttrycks i ordinalskala. Emellertid brukar man inte välja dessa metoder på den högre skalnivån eftersom det finns speciella metoder med högre känslighet som kan tillämpas på denna typ av data.

3.1 Kolmogorov-Smirnov test för en uppsättning mätvärden

I avsnitt 2.3 redogjordes för chi^2-goodness of fit och vi angav där att detta test är tillämpligt när vi endast har en uppsättning av empiriska mätvärden och då data uttrycktes i nominalskala. Vid samma typ av problem men då vi har data i ordinalskala väljer vi i stället Kolmogorov-Smirnov-testet eftersom detta test har en betydligt högre känslighet än chi^2 goodness of fit.

Kolmogorov-Smirnov-testet prövar således huruvida en viss empirisk fördelning överensstämmer med en teoretisk fördelning. Härvid utgår man från de båda fördelningarnas kumulativa proportioner och beräknar differensen (D) mellan dessa på varje skalsteg. Det högsta D-värde som man då erhåller jämförs sedan med de värden som anges i signifikanstabell 2.

Nollhypotesen anger som vanligt att det inte finns någon skillnad mellan den empiriska och den teoretiska fördelningen dvs. att samtliga erhållna D-värden är så låga att de uppkommit av en slump. Om ett D-värde däremot är lika högt som eller överstiger det värde som anges i signifikanstabell 2 förkastas nollhypotesen och vi antar den alternativa hypotesen som anger att det finns skillnader mellan den empiriska och den teoretiska fördelningen.

Ett exempel:
Vi skall pröva ut ett undervisningspaket i matematik på en grupp elever i en viss årskurs. För att få en uppfattning om undersökningsgruppens representativitet vill vi bl.a. veta gruppens betygsfördelning i ämnet matematik, vilken är:

Betyg	N
5	12
4	25
3	44
2	14
1	5
Σ	100

Vi har här en enda uppsättning av empiriska mätvärden som nu skall jämföras med en teoretisk. Denna teoretiska fördelning får vi från SÖ:s rekommenderade betygsfördelning för en hel årskurs elever i Sverige och som anger att 7 % skall ha betyget 1, 24 %: betyget 2, 38 %: betyget 3, 24 %: betyget 4 och de 7 % bästa eleverna betyget 5.

Beräkningarna för Kolmogorov-Smirnov-testet är mycket enkla och innebär att vi först gör om dessa båda fördelningar till kumulativa proportioner varefter vi beräknar differensen mellan de båda fördelningarnas kumulativa proportioner på varje betygsnivå.

	Kumulativa proportioner		
Betyg	empirisk	teoretisk	D
5	1.00	1.00	0.00
4	0.88	0.93	-0.05
3	0.63	0.69	-0.06
2	0.19	0.31	-0.12
1	0.05	0.07	-0.02

Det högsta D-värdet har vi på betygsnivå 2, där D = 0.12. Detta värde skall då jämföras med det kritiska värdet i signifikanstabell 2 i slutet av kompendiet. Av denna tabell framgår att de kritiska värdena varierar med undersökningsgruppens storlek och om vi väljer 5 %-nivån anges som kritiskt värde: $\frac{1.36}{\sqrt{N}}$. I vårt fall är N = 100 varför det kritiska värdet blir $\frac{1.36}{\sqrt{100}} = 0.136$

Vårt högsta D-värde är emellertid lägre än det kritiska värdet, varför vi accepterar nollhypotesen. Slutsatsen måste således bli att vår empiriska fördelning inte avviker från den betygsfördel-

ning som rekommenderas av SÖ. Med andra ord kan vår undersöknings-grupp anses vara representativ med avseende på matematikbetyg för hela årskursens elever i Sverige.

Här kan det emellertid vara befogat med en varning. Vår slutsats är nämligen giltig enbart under den förutsättningen att den rekommenderade betygsfördelningen verkligen följs. Vi vet t.ex. från Svensson (1971) att det i verkligheten förekommer vissa avvikelser såtillvida att det finns en förskjutning mot de högre betygsnivåerna och följaktligen något färre elever på de lägre nivåerna än vad den teoretiska fördelningen anger. I vårt exempel är detta knappast några problem då vår empiriska fördelning visar en likadan avvikelse och således är mer lik den verkliga betygsfördelningen i Sverige än den som rekommenderades av SÖ. Eftersom vår fördelning inte visar någon statistisk säkerställd skillnad gentemot den rekommenderade fördelningen kan den följaktligen inte heller avvika från den verkliga betygsfördelningen för alla elever i årskursen.

3.2 Mediantestet

Mediantestet är ett mycket användbart test såtillvida att det kan tillämpas såväl när vi har 2 uppsättningar oberoende mätvärden som när vi har k sådana uppsättningar och när data uttrycks i ordinalskala. Det har också den fördelen att det i jämförelse med andra alternativa metoder är relativt lätthanterligt också när man arbetar med stora undersökningsgrupper, vilket ju är ganska vanligt i samband med pedagogiska undersökningar.

Den största nackdelen med mediantestet är att det har en relativt låg känslighet i jämförelse med t.ex. Mann-Whitney U-test (se Siegel 1956 sid. 116-127) för 2 uppsättningar oberoende mätvärden och i jämförelse med Kruskal-Wallis en-vägs variansanalys (se Siegel 1956 sid. 184-193) för k sådana uppsättningar. Vid relativt stora undersökninsgrupper torde mediantestet emellertid ha acceptabel känslighet (jfr vårt resonemang i kap. I om känslighet och N) men om undersökningsgrupperna däremot är små bör man i stället välja den metod av de ovan angivna som är tillämplig med hänsyn till undersökningsuppläggningen.

För att förenkla framställningen nöjer vi oss med att här redogöra för mediantestet och läsaren får själv vid behov hämta information om de båda alternativa metoderna i Siegels bok.

Tillvägagångssättet vid användning av mediantestet är exakt detsamma när vi har 2 uppsättningar mätvärden som då vi har flera uppsättningar. Skillnaden mellan dessa två fall är endast den att vi får beräkningstabeller med olika många celler.

Nollhypotesen vid mediantestet anger att undersökningsgrupperna har samma median medan den alternativa hypotesen anger att de har olika medianer. Eftersom medianen delar en fördelning i två lika stora delar innebär nollhypotesen att vi i samtliga undersökningsgrupper bör ha lika höga frekvenser över som under den gemensamma median vilken alltså enligt nollhypotesen skall vara samma som de enskilda undersökningsgruppernas medianer.

För att konkretisera framställningen beskriver vi tillvägagångssätet i anslutning till ett exempel:

Tre yrkesgrupper om vardera 50 personer får besvara en fråga rörande trivseln i arbetet varvid svaren ges i en femgradig skala:

	Grupp			
	A	B	C	Σ
Mycket bra	5	6	7	18
Bra	10	8	7	25
Varken bra eller dåligt	15	10	7	32 ← gemensam median
Ganska dåligt	12	12	15	39
Mycket dåligt	8	14	14	36
N	50	50	50	150

Tillsammans utgör de tre undersökningsgrupperna 150 individer, vilket innebär att den gemensamma medianen går där summakolumnen delas i två grupper om vardera 75 individer dvs. mellan kategorierna "varken bra eller dåligt" och "ganska dåligt". Efter att på detta sätt ha bestämt den gemensamma medianen har vi bara att räkna hur många i varje grupp som faller över resp. under denna median. Denna beräkning leder till en ny tabell:

	Grupp			
	A	B	C	Σ
Över medianen	30	24	21	75
Under medianen	20	26	29	75
N	50	50	50	150

Vi uttryckte tidigare nollhypotesen så att samtliga grupper har samma median vilken då, som vi tidigare sagt, måste överensstämma med den gemensamma medianen. Är denna hypotes riktig skall de tre grupperna följaktligen fördela sig så att vi i varje grupp får lika många individer över som under medianen. För att pröva detta gör vi slutligen en chi^2 tests of independence på det sätt som vi redovisat i föregående kapitel:

	Observerat:				Förväntat:			
	A	B	C	Σ	A	B	C	Σ
Över medianen	30	24	21	75	25	25	25	75
Under medianen	20	26	29	75	25	25	25	75
N	50	50	50	150	50	50	50	150

χ^2-värdet blir här 3.36, vilket skall jämföras med det kritiska värdet för 2 frihetsgrader på - låt oss säga 1 %-nivån - vilket är 9.21. Slutsatsen blir således att det inte finns någon statistiskt säkerställd differens mellan grupperna beträffande fördelningen omkring medianen. Grupperna kan således sägas ha samma median och därmed också samma trivsel i sitt arbete.

Tillvägagångssättet är exakt det samma om vi enbart har 2 uppsättningar mätvärden. Det enda som skiljer detta fall från det vi beskrivit ovan är att vi vid 2 uppsättningar av mätvärden får en 2×2-tabell till grund för den avslutande chi^2-analysen.

I vårt exempel råkade den gemensamma medianen falla exakt mellan två skalvärden, vilket emellertid inte alltid är fallet. Ofta händer det nämligen att den råkar hamna inne i en kategori. Får man ett sådant utfall har man att välja mellan två handlingsalternativ:

1) Om frekvenserna i denna svarskategori inte är alltför höga i förhållande till det totala antalet individer i undersökningen kan man helt enkelt utelämna kategorin i fråga vid den avslutande chi^2-analysen och göra den enbart på frekvenserna inom de övriga kategorierna.
2) Omfattar "mediankategorin" relativt höga frekvenser bör de däremot tas med vid chi^2-analysen för att inte känsligheten hos testmetoden skall bli alltför låg. I detta fall bildar de som faller ovanför mediankategorin en grupp medan de som faller inom denna svarskategori slås samman med dem som faller under medianen.

En sak att observera när det gäller mediantestet är att den avslutande chi^2-analysen ställer samma krav på de förväntade frekvensernas storlek som vi redogjorde för i samband med chi^2-metoderna (se avsnitt 2.5). Vidare beräknas frihetsgraderna utifrån den tabell, som ligger till grund för chi^2-analysen - dvs. på den tabell som innebär en uppdelning enbart på kategorierna "över" respektive "under" medianen.

3.3 Teckentestet

Teckentestet har fått sitt namn av att man vid detta utgår ifrån plus- och minustecken och inte ifrån kvantitativa data. Detta test används när vi har två uppsättningar av beroende mätvärden och när data uttrycks i ordinalskala. Den enda förutsättning som skall vara uppfylld är att de två mätvärdena inom varje matchat par eller från samma individ kan rangordnas i förhållande till varandra.

Metoden är mycket enkel att använda. Man har bara att fastställa riktningen på de skillnader som uppkommer inom varje par av mätvärden. Går skillnaden i den ena riktningen så markeras detta med ett plus medan man markerar ett minus för de par av mätvärden där skillnaden går i den andra riktningen. Finns det ingen skillnad mellan de två mätvärdena markeras detta fall med en nolla. Man har sedan bara att räkna antalet plus- och minustecken. Det tecken som har lägsta frekvens sätts sedan i relation till det totala antalet tecken vid själva signifikansprövningen.

Ett exempel:
15 kvinnor fick ange sin inställning till rökning respektive snusning. Sin inställning fick de markera i en 5-gradig skala, där 1 betydde "gräsligt obehagligt" och 5 betydde "toppen". Vi prövar nu på 5 %-nivån om dessa två njutningsmedel "föll kvinnorna på läppen" i olika hög grad.

Nollhypotesen är här att kvinnorna tycker lika illa eller bra om dessa två njutningsmedel.
Den alternativa hypotesen är då att de tycker bättre om det ena njutningsmedlet än det andra.

Resultatet blev:

	Rökn.	Snusn.		Rökn.	Snusn.
Ada	4	2	Ida	1	4
Beda	1	1	Josefina	2	3
Charlotta	3	5	Klara	1	5
Davida	3	4	Lydia	5	1
Ebba	1	2	Matilda	2	3
Frida	1	3	Natalia	4	1
Götilda	1	5	Ottilia	1	2
Hilma	2	2			

Om nollhypotesen är riktig - dvs. om kvinnor tycker lika bra om de två njutningsmedlen - skall antalet plustecken och antalet minustecken vara lika. Är däremot den alternativa hypotesen riktig skall de flesta skillnaderna inom varje par av skattningar gå i samma riktning vilket alltså innebär att vi får olika många av de två tecknen.

Vi markerar de fall där rökningen får en högst skattning med + och de fall där snusningen får högst skattning med -. De fall där vi får samma skattning för båda markeras med 0.

Resultaten ovan visar att vi får 3 plustecken och 10 minustecken. I två fall finns ingen skillnad mellan de två skattningarna. Vi har således sammanlagt 13 tecken, vilket markeras med N - nollorna uteslutes helt. Antalet plustecken är minst, varför detta antal markeras: x. Det är enbart dessa två värden vi använder oss av, när vi studerar signifikanstabell 3.

När vi nu går in i signifikanstabellen så markerar N den rad och x den kolumn som skall följas. Vi går således in i raden för 13 och kolumnen för 3. Det p-värde vi då får fram är 0.046. Som anges i samband med tabellen är dessa p-värden giltiga för de fall där man har en riktad hypotes. För vårt exempel skulle det innebära att vi i den alternativa hypotesen hade angivit vilket av de två njutningsmedlen kvinnorna tyckte bäst om. Så gjordes emellertid inte. För att få fram p-värdet för vårt fall - alltså med en icke-riktad alternativ hypotes måste tabellens värde dubblas vilket ger p = 0.092. Sannolikheten att av en slump få tre tecken av det ena slaget utav totalt 13 tecken är således ca. 9 fall av 100. Eftersom vi hade valt 5 % signifikansnivå måste nollhypotesen antas. Det finns med andra ord ingen skillnad mellan de två njutningsmedlen vid beträffar kvinnornas inställning till dem.

Tänk alltså på att de p-värden som anges i signifikanstabell 3 gäller för de fall där alternativhypotesen är riktad. Är denna hypotes ej riktad måste tabellens p-värden dubblas.

Det tillvägagångssätt vi angivit ovan gäller när N - antalet TECKEN - är 25 eller mindre. Har vi flera än 25 tecken kan vi i stället använda normalfördelningen enligt följande formel:

$$z = \frac{|D|-1}{\sqrt{N}}$$

där |D| står för det absoluta värdet av skillnaden mellan antalet plus- och minustecken. Att denna skillnad reduceras med 1 utgör en korrektionsfaktor.

Vi håller oss kvar vid det exempel vi hade ovan men antar att undersökningsgruppen fördubblas samt att de 15 nya kvinnorna svarade på samma sätt som de vi hade från början. Vi skulle då få totalt 26 tecken varav 6 plus och 20 minus. Eftersom vi nu har flera än 25 tecken går vi över normalfördelningen och vi får:

$$z = \frac{|(20-6)|-1}{\sqrt{26}} = \frac{13}{5.10} = 2.55$$

z överstiger således 1.96 vilket är det kritiska värdet på 5 %-nivån. Nollhypotesen förkastas och vi slår fast att kvinnorna hade

en mera positiv inställning till (här gäller det att veta vad plus och minus betecknar) snusningen än till rökningen.

3.4 Friedmans två-vägs variansanalys

Har vi fler än 2 uppsättningar av beroende mätvärden kan vi självfallet inte arbeta med skillnadernas riktningar. Vi måste i stället använda andra typer av mått. Det mått som Friedmans två-vägs variansanalys utgår ifrån är en alldeles speciell form av variationsmått.

Den undersökningsuppläggning på vilken denna metod är tillämplig innebär att vi har en undersökningsgrupp, som ger fler än två olika uppsättningar beroende mätvärden. Dessa mätvärden måste vidare uttryckas i ordinalskala så att de olika värden, som en individ avger, kan rangordnas i förhållande till varandra. Det är således inte de olika individerna som rangordnas i förhållande till varandra. I stället för att använda en enda undersökningsgrupp kan vi självfallet lika väl använda oss av flera matchade grupper. Metoden blir ändå densamma.

För att illustrera tillvägagångssättet tar vi ett exempel som vi tidigare utnyttjat i samband med att vi beskrev principerna för valet mellan olika icke-parametriska metoder:

Vi har utformat ett visst undervisningsavsnitt på fyra olika sätt:
a) enbart text utan någon form av illustration
b) text illustrerad med bilder
c) film + arbetsmaterial
d) programmerad undervisning.

Vi är nu intresserade av att studera om det finns någon skillnad mellan dessa olika presentationssätt beträffande förmåga att skapa motivation hos eleverna. För att inte trötthetseffekter och liknande skall inverka på utfallet använder vi oss av matchade grupper. Vi tar således ut fyra olika grupper om åtta elever vardera på ett sådant sätt att varje elev i en grupp har en "tvilling" i varje annan grupp - vi kan uttrycka det så att vi får åtta fyr-

lingsgrupper som inbördes är lika varandra i alla relevanta avseenden. Inom varje sådan grupp slumpas eleverna ut - en på varje presentationssätt.

Efter att ha gått igenom undervisningen får varje elev skatta den med avseende på dess motivationskapande förmåga i en skala från 1 till 7, där 7 är det mest positiva värdet. Eftersom Friedmans medod enbart arbetar med ranger måste vi utifrån skattningarna göra en rangordning mellan de olika presentationssätten. Skattningen och rangordningen presenteras i nedanstående tabeller:

Skattningar

	Metod a	b	c	d
Elev A	2	4	6	3
B	1	5	4	2
C	3	4	7	2
↑ D	1	2	3	1
N	E	2	5	3	3
↓ F	2	5	3	1
G	1	3	5	2
H	1	4	7	3
	←k→			

Rangordning

	Metod a	b	c	d
Elev A	4	2	1	3
B	4	1	2	3
C	3	2	1	4
D	3.5	2	1	3.5
E	4	1	2.5	2.5
F	3	1	2	4
G	4	2	1	3
H	4	2	1	3
R_j:	29.5	13	11.5	26

Vi har vid tabellerna ovan också angivit de olika beteckningar som ingår i formeln:

N anger antalet försökspersoner eller matchade enheter

k anger det antal mätvärden varje försöksperson/matchningsenhet avger

R_j anger rangsumman för respektive metod.

Dessa olika värden ingår nu i formeln på följande sätt:

$$\chi_r^2 = \frac{12}{N \cdot k(k+1)} \Sigma R_j^2 - 3N(k+1)$$

De olika R_j-värdena står angivna i tabellen ovan men av formeln framgår att dessa skall kvadreras och sedan summeras. Vi får då:

$R_a = 29.5$ $\quad R_a^2 = 870.25$

$R_b = 13$ $\quad R_b^2 = 169$

$R_c = 11.5$ $\quad R_c^2 = 132.25$

$R_d = 26$ $\quad R_d^2 = 676$

$\Sigma R_j^2 = 1847.50$

Det är just ΣR_j^2 som är det intressanta i formeln. De övriga komponenterna i den är ju redan givna när vi väl bestämt antalet försökspersoner/matchningsenheter och antalet "behandlingar". Vi skall därför något stanna vid uttrycket ΣR_j^2 .

Om vi går tillbaka till våra resultat ovan så ser vi av rangordningstabellen att ΣR_j är ett låst värde. Varje matchningsenhet anger ju 10 "rangordningspoäng" (1+2+3+4=10). Eftersom vi har 8 sådana enheter måste ΣR_j bli 80. Detta innebär att om en metod får ett högt R_j-värde måste det kompenseras av ett lågt R_j för en annan metod.

Låt oss nu anta att det inte fanns några som helst skillnader mellan våra fyra presentationssätt. Då skulle ju samtliga R_j bli lika höga dvs. $\frac{80}{4} = 20$. Kvadrerar vi dessa R_j-värden och sedan summerar dem får vi $\Sigma R_j^2 = 1600$.

Vi sätter in dessa värden i formeln:

$$\chi_r^2 = \frac{12}{8 \cdot 4 \cdot (4+1)} \cdot 1600 - 3 \cdot 8 \cdot (4+1), \text{ vilket ger } \chi_r^2 = 0$$

Uttrycken $\frac{12}{N \cdot k \cdot (k+1)}$ och $3\,N(k+1)$ fungerar således bara som korrektionsfaktorer i formeln så att $\chi_r^2 = 0$ när det inte finns några skillnader. I ett sådant fall antar också ΣR_j^2 sitt lägsta möjliga värde. Detta hänger samman med att R_j-värdena kvadreras - x^2-kurvan stiger ju allt brantare med ökande x-värden. Detta kan enkelt visas om vi gör en liten ändring i R_j-värdena.

Vi hade det exemplet att samtliga R_j-värden var 20. Då blev ΣR_j^2 = = 1600. Ökar vi ett av värdena till 21 måste detta kompenseras av

att ett annat värde sänks till 19. Vad händer då med ΣR_j^2 jo $20^2+20^2+21^2+19^2$ blir 400+400+441+361=1602. Det tillskott som 21^2 ger kompenseras alltså inte helt av den minskning som 19^2 utgör. Följden av detta måste bli att ju mera R_j-värdena varierar desto högre värde antar ΣR_j^2 , vilket följaktligen utgör ett variationsmått.

Räknar vi fram χ_r^2 för de resultat vi hade i vårt ursprungliga exempel får vi:

$$\chi_r^2 = \frac{12}{8 \cdot 4 \cdot 5} \cdot 1847.50 - 3 \cdot 8 \cdot 5 = 18.56$$

Eftersom χ_r^2 följer den vanliga chi^2-fördelningen går vi till signifikanstabell 1. Men först måste vi bestämma antalet frihetsgrader.

Det som prövas med denna metod är ju variationen mellan de fyra R_j-värdena. Som vi angivit ovan är det emellertid bara 3 av dessa som kan variera - ΣR_j är ju ett låst värde. Vi har således 3 frihetsgrader. Formeln för frihetsgrader vid Friedmans metod blir således df = k-1

Det kritiska χ^2-värdet på 1 %-nivån för df = 3 är 11.34 varför vi kan slå fast att de olika presentationssätten hade olika motivationsvärde.

I vårt fall blir tolkningen enkel. Av R_j-värdena framgår ju att metoderna b och c är klart mer motiverande än metoderna a och d. I andra fall kan det vara besvärligare att avgöra mellan vilka behandlingar det föreligger skillnader. Då kan man - på samma sätt som att en vanlig variansanalys följs av t-testningar - göra separata testningar mellan behandlingarna parvis - lämpligen med teckentestet.

3.5 Spearmans rangkorrelation

Vi har nu endast kvar att behandla en cell i översiktstabellen över icke parametriska metoder nämligen den som avser sambandsmått för data i ordinalskala. För denna typ av data har vi valt ut två olika korrelationsmetoder: en för beräkning av sambandet mellan två uppsättningar mätvärden - Spearmans rangkorrelation (ρ=rho) och en

för beräkning av samband mellan fler än två uppsättningar av mätvärden - Kendalls konkordanskoefficient (W), vilken vi skall redogöra för i nästa avsnitt.

Spearmans rangkorrelation är den icke-parametriska motsvarigheten till den parametriska produktmomentkorrelationen. Till skillnad från den sistnämnda metoden, som bl.a. kräver data i lägst intervallskala, kan Spearmans rangkorrelation användas i de fall, när vi har två uppsättningar mätvärden, vilka var för sig kan omvandlas till en rangskala. Man utgår nämligen ifrån ranger vid beräkningen av ρ.

Ett exempel:
Tolv personer skattades i en 7 gradig skala med avseende på grad av religiositet och benägenhet att svärja. Skattningen innebar att ju mera de svor och ju mera religiösa de var desto högre skattning fick de.

Vi vill nu veta om det finns något samband mellan religiositet och svordomsfrekvens.

Nedan redovisas dels skattningarna dels de rangordningar som baseras på dessa skattningar. Eftersom vi vill veta sambandet mellan två variabler i ordinalskala kan vi således beräkna ρ, vars formel är:

$$\rho = 1 - \frac{6\Sigma D^2}{N(N^2-1)}$$

N står för antalet individer. D står för skillnaden mellan de ranger en och samma individ får i de båda variablerna.

Eftersom formeln innehåller uttrycket ΣD^2 skall således denna skillnad kvadreras och därefter summeras över alla individer.

Utifrån rangerna måste vi således beräkna D och D^2 samt ΣD^2, vilket vi gjort i tabellen nedan.

	Skattningar			Ranger			
Ind.	Relig.	Svord.	Ind.	Relig.	Svord.	D	D^2
A	7	2	A	2	8.5	-6.5	42.25
B	5	3	B	5.5	6	-0.5	0.25
C	6	3	C	4	6	-2	4.00
D	1	7	D	11.5	1	10.5	110.25
E	7	1	E	2	11	-9	81.00
F	7	3	F	2	6	-4	16.00
G	2	5	G	9.5	3	6.5	42.25
H	1	1	H	11.5	11	0.5	0.25
I	3	4	I	8	4	4	16.00
J	4	2	J	7	8.5	-1.5	2.25
K	5	1	K	5.5	11	-5.5	30.25
L	2	6	L	9.5	2	7.5	56.25
						$\Sigma D=0.0$	$\Sigma D^2=401.00$

Sätter vi in värdena i formeln får vi följande resultat:

$$\rho = 1 - \frac{6 \cdot 401}{12(12^2-1)} = 1 - \frac{2406}{1716} = 1 - 1.40 = -0.40$$

Spearmans rangkorrelation kan variera från +1.00 via 0.00 till -1.00. Det högsta positiva sambandet (+1.00) får vi när $\Sigma D^2 = 0$ dvs. när det inte finns någon som helst skillnad mellan rangordningarna i de två variablerna. Det lägsta möjliga värdet får däremot ρ när ΣD^2 når sitt högsta värde, vilket innebär att vi har helt omvända rangordningar i de två variablerna.

Om vi nu går tillbaka till vårt exempel förefaller det ganska naturligt med hänsyn till de variabler som ingår att vi fått ett negativt samband. Däremot kunde man kanske väntat sig att det negativa sambandet skulle bli något starkare.

En anledning till att vi inte fått en högre negativ korrelation är att vi i båda variablerna har ganska många bundna ranger - dvs. vi har i flera fall individer som hamnat på samma rangnummer. De bundna rangerna får samma inverkan på rangkorrelationen som en låg varians har på produktmomentkorrelationen - de sänker sambandet. När det gäller Spearmans rangkorrelation kan man emellertid göra en

korrektion för de bundna rangernas inverkan, men denna korrektion förutsätter att vi använder en alternativ formel för beräkning av ρ:

$$\rho = \frac{\Sigma x^2+\Sigma y^2-\Sigma D^2}{2\sqrt{\Sigma x^2 \cdot \Sigma y^2}}$$

Där Σx^2 och $\Sigma y^2 = \frac{N(N^2-1)}{12}$ och ΣD^2 samma som vid den formel vi först presenterat.

För att visa att denna formel verkligen är alternativ till den först presenterade, som också är den vanligast förekommande, gör vi om beräkningen på vårt exempel, där alltså N = 12 och $\Sigma D^2 = 401$

Σx^2 och Σy^2 blir således $\frac{12(12^2-1)}{12} = 12^2 - 1 = 143$

vi får då $\rho = \frac{143+143-401}{2\sqrt{143 \cdot 143}} = \frac{-115}{286} = -0.40$

dvs. exakt samma värde som tidigare.

Vi lägger nu in korrektionen, som innebär att Σx^2 och Σy^2 reduceras med ΣT_x respektive ΣT_y.

$$\Sigma T = \Sigma \frac{t^3-t}{12}$$

där t anger det antal individer som har en och samma rang i respektive variabel.

Vi börjar med religiositet, som vi kallar x-variabeln. Då får vi:

$$\Sigma T_x = \Sigma \frac{t_x^3-t_x}{12}$$

I x-variabeln har vi först 3 individer som har rangnummer 2, vilket ger

$$\frac{3^3-3}{12} = \frac{24}{12} = 2$$

därefter har vi 2 individer som har rangnummer 5.5, vilket ger

$$\frac{2^3-2}{12} = \frac{6}{12} = 0.5$$

Vi har dessutom 2 individer på vardera rangsummorna 9.5 och 11.5 vilka vardera på samma sätt som ovan ger 0.5. Skriver vi ut det fullständiga värdet för ΣT_x får vi:

$$\Sigma T_x = \frac{3^3-3}{12} + \frac{2^3-2}{12} + \frac{2^3-2}{12} + \frac{2^3-2}{12} = 2 + 0.5 + 0.5 + 0.5 = 3.5$$

Korrektionen för bundna ranger innebär att uttrycket Σx^2 i formeln ovan skall reduceras med ΣT_x vilket ger 143 - 3.5 = 139.5.

På samma sätt som för x-variabeln beräknar vi korrektionsfaktorn för y-variabeln (svordomar) dvs. ΣT_y, vilken blir: (som vägledning sätter vi upp rangnumret inom parentes under uttrycket)

$$\Sigma T_y = \frac{2^3-2}{12} + \frac{3^3-3}{12} + \frac{3^3+3}{12} = 0.5 + 2 + 2 = 4.5$$

rangnr. (8.5) (6) (11)

På motsvarande sätt som tidigare skall Σy^2 reduceras med ΣT_y: 143 - 4.5 = 138.5

Efter korrektionen får vi följande värden i formeln för ρ :

$$\rho = \frac{139.5+138.5-401}{2\sqrt{139.5 \cdot 138.5}} = \frac{-123}{2 \cdot 139} = \frac{-123}{278} = -0.44$$

Korrektionen innebär således att det negativa sambandet ökade från -0.40 till -0.44.

Lika väl som vi kan signifikanstesta skillnader kan vi pröva om korrelationen är signifikant skild från 0 dvs. om det finns ett verkligt samband eller om det uppkommit av en slump.

I signifikanstabell 4 anges det lägsta möjliga värdet för ρ vid olika N för att korrelationen skall anses vara signifikant skild från 0. Det bör emellertid observeras att de värden som anges i tabellen avser riktade hypoteser dvs. man måste i förväg ha angivit om ρ skall bli positiv eller negativ.

Vi antar att vi från början gjort det rimliga antagandet att det skall finnas ett negativt samband mellan religiositet och benägenheten att svära.

Av signifikanstabellen framgår att ρ måste vara lägst 0.506 för att man skall anse sambandet vara signifikant på 5 %-nivån. Vårt värde var endast 0.44 efter korrektion för bundna ranger.

Slutsatsen för vårt exempel blir således att det inte finns något statistiskt säkerställt samband inom vår grupp mellan graden av religiositet och benägenheten att säga fula ord.

Signifikansprövningen kan också göras med hjälp av t-fördelningen under förutsättning att N är lägst 10. Denna prövning görs enligt nedanstående formel - vilken är identisk med den formel som gäller för signifikansprövning av produktmomentkorrelationen.

$$t = \rho \cdot \sqrt{\frac{N-2}{1-\rho^2}} \quad \text{med } N-2 \text{ frihetsgrader}$$

Som alternativ till Spearmans rangkorrelation finns Kendalls rangkorrelation, vilken emellertid inte är lika vanlig. Den som är intresserad av Kendalls metod hänvisas till Siegel (1956) sid. 213-223.

3.6 Kendalls konkordanskoefficient

Som framgick av föregående avsnitt utgör Spearmans rangkorrelation ett mått på överensstämmelsen mellan två gjorda rangordningar. I vissa fall kan det emellertid vara värdefullt att få ett sådant sammanfattande mått på tre eller flera rangordningar. Ett inte alldeles ovanligt exempel på detta är då man vill kontrollera den s.k. interbedömarreliabiliteten dvs. överensstämmelsen mellan ett antal bedömare som oberoende av varandra gjort en skattning t.ex. av ett antal barns aggressivitet. Man utgår härvid ifrån att ju bättre bedömarnas skattningar överensstämmer desto högre är reliabiliteten.

Ett sätt att få fram interbedömarreliabiliteten är naturligtvis att beräkna rangkorrelationer mellan alla möjliga par av bedömare och sedan räkna fram genomsnittet av de erhållna korrelationerna. Det skulle emellertid vara mycket arbetsamt och tidsödande i jämförelse med att direkt få ett sammanfattande mått på denna överensstämmelse. Ett sådant sammanfattande mått utgör Kendalls konkordanskoefficient (W).

I likhet med Spearmans rangkorrelation utgår Kendalls konkordanskoefficient ifrån data uttryckta i form av rangordningar och den antar sitt högsta värde (+1) när samtliga bedömare gjort exakt lika rangordningar. Däremot kan W inte antaga negativa värden - det är ju otänkbart att tre eller flera rangordningar är negativt korrelerade med varandra - varför W:s lägsta möjliga värde är 0.

W har också vissa likheter med Friedmans tvåvägs variansanalys såtillvida att den arbetar med variationen mellan olika rangsummor. De rangsummor som är aktuella vid Kendalls W är emellertid de bedömdas rangsummor.

Ett exempel:
Vid herrarnas fria åkning i konståknings-VM på skridsko bedömde 6 av domarna de 5 bästa konståkarna på följande sätt:

		Domare					
		1	2	3	4	5	6
Tävlande:	A	5.8	5.7	5.8	5.7	5.8	5.7
	B	5.9	5.8	5.9	5.9	5.9	5.8
	C	5.6	5.7	5.9	5.8	5.8	5.7
	D	5.9	5.9	5.8	5.9	6.0	5.9
	E	5.8	5.7	5.7	5.8	5.7	5.7

(Bedömningen sker i en skala från 0 till 6.0 poäng, där 6.0 är bästa möjliga resultat.)

Vi är nu intresserade av att veta graden av överensstämmelse mellan de 6 domarnas rangordning av de tävlande.

Första steget vid beräkningen av W är att vi överför poängen i tabellen ovan till ranger. Eftersom det är graden av överensstämmelse mellan domarnas bedömning som vi är ute efter måste rangordningen avse domarnas rangordning av de tävlande. (Ett vanligt fel är att man rangordnar bedömarna i stället för de bedömda).

I tabellen nedan har vi utifrån poängen satt upp varje domares rangordning av de tävlande. Vi har också *summerat rangerna för varje tävlande* i kolumnen R_j, varefter medelvärdet beräknats för dessa rangsummor ($\bar{R}_j$). Slutligen har för varje tävlande beräknats R_j-s avvikelse från $\bar{R}_j$ och denna avvikelse kvadrerats: $(R_j-\bar{R}_j)^2$.

Tävlande	Domare 1	2	3	4	5	6	R_j	$R_j-\bar{R}_j$	$(R_j-\bar{R}_j)^2$
A	3.5	4	3.5	5	3.5	4	23.5	5.5	30.25
B	1.5	2	1.5	1.5	2	2	10.5	-7.5	56.25
C	5	4	1.5	3.5	3.5	4	21.5	3.5	12.25
D	1.5	1	3.5	1.5	1	1	9.5	-8.5	72.25
E	3.5	4	5	3.5	5	4	25.0	7.0	49.00
							ΣR_j=90.0	Σ 0	Σ:220.00

$$\bar{R}_j = \frac{90.0}{5} = 18.0$$

Det intressanta värdet i beräkningarna ovan är $\Sigma(R_j-\bar{R}_j)^2$, vilket anger den kvadrerade avvikelsen för varje tävlandes rangsumma från rangsummornas medeltal - således ett variationsmått.

Detta variationsmått når sitt högsta värde när alla bedömarna har gjort en exakt lika rangordning av de bedömda.

Om vi hade fått ett sådant utfall i vårt exempel borde tävlande D fått rangnummer 1 av samtliga 6 domare dvs. $R_D = 6$. B borde genomgående fått rangnummer 2 - alltså $R_B = 12$ osv. Ett sådant utfall skulle gett följande R_j-värden:

	R_j	$R_j-\bar{R}_j$	$(R_j-\bar{R}_j)^2$
A	24	6	36
B	12	-6	36
C	18	0	0
D	6	-12	144
E	30	12	144
	$\Sigma R_j = 90$		Σ 360
	$\bar{R}_j = 18$		

I enlighet med Kendalls formel döper vi uttrycket $\Sigma(R_j-\bar{R}_j)^2$ till S, vars högsta värde enligt beräkningarna ovan blir 360, när vi har 6 domare som bedömer 5 försökspersoner. S_{max} kan också beräknas utifrån följande formel:

$$S_{max} = \frac{m^2(N^3-N)}{12}$$

där m anger antalet bedömare och N antalet bedömda.
Sätter vi in exemplets värden för m och N får vi:

$$\frac{6^2(5^3-5)}{12} = \frac{36 \cdot 120}{12} = 360$$

Vi kan således slå fast att S_{max} erhålls när alla bedömare gjort exakt lika rangordningar av de bedömda.

Kendalls W kan nu definieras som förhållandet mellan det verkliga S-värdet och S_{max}. Formelmässigt uttryckt:

$$W = \frac{S}{S_{max}}$$

Om vi i formeln ovan byter ut S_{max} mot den formel vi angivit ovan för detta värde erhåller vi

$$W = \frac{S}{\frac{m^2(N^3-N)}{12}} = \frac{12\ S}{m^2(N^3-N)}$$

Vi sätter nu in värdena för S, m och N utifrån vårt utgångsexempel och erhåller:

$$W = \frac{12 \cdot 220}{6^2(5^3-5)} = \frac{12 \cdot 220}{36 \cdot 120} = 0.61$$

Går vi tillbaka till rangordningstabellen ovan ser vi att det finns många bundna ranger. De bundna rangerna får samma effekt på Kendalls W som på Spearmans rangkorrelation dvs. W sänks. Vi kan emellertid även här göra en korrektion (T) vilken uttryckt i formel är:

$$\Sigma T = \frac{\Sigma(t^3-t)}{12}$$

där t anger det antal bedömda som av en viss bedömare fått en och samma rang.

Enklast sker beräkningen av T-värden så att de beräknas för varje bedömare för sig och sedan summeras över bedömarna.

Vi får:

Domare 1: $T_1 = \frac{(2^3-2)+(2^3-2)}{12} = 1$

2: $T_2 = \frac{3^3-3}{12} = 2$

3: $T_3 = \frac{(2^3-2)+(2^3-2)}{12} = 1$

4: $T_4 = \frac{(2^3-2)+(2^3-2)}{12} = 1$

5: $T_5 = \frac{2^3-2}{12} = 0.5$

6: $T_6 = \frac{3^3-3}{12} = 2$

Summerar vi T-värdena över domarna får vi: $\Sigma T = 7.5$

Detta värde skall sedan multipliceras med m och subtraheras från nämnaren.

Den fullständiga formeln för W inklusive korrektionen blir då

$$W = \frac{12 \cdot S}{m^2(N^3-N)-m\Sigma T}$$

Vårt exempel ger:

$$W = \frac{12 \cdot 220}{6^2(5^3-5)-6 \cdot 7.5} = 0.62$$

Som framgår har korrektionen för bundna ranger ganska liten betydelse såvida inte dessa är mycket frekventa.

Slutligen kan W signifikansprövas vilket sker med hjälp av χ^2 enligt formeln:

$\chi^2 = m(N-1) \cdot W$ med N-1 frihetsgrader.

Anledningen till att vi har N-1 frihetsgrader är att vi ju utgår från variationen mellan de bedömdas R_j-värden. Eftersom ΣR_j är given, när vi vet m och N, måste 1 av de N stycken R_j-värdena vara låst och vi får således df = N-1.

Vårt exempel ger:

$$\chi^2 = 6 \cdot 4 \cdot 0.62 = 14.88$$

Kritiskt χ^2-värde på 1 %-nivån med 4 frihetsgrader är 13.28. Vi kan således slå fast att domarna hade en på 1 %-nivån statistiskt säkerställd överensstämmelse i sina rangordningar av dessa 5 tävlande.

Kritiska värden för chi² (χ²)

df	Signifikansnivå													
	.99	.98	.95	.90	.80	.70	.50	.30	.20	.10	.05	.02	.01	.001
1	.00016	.00063	.0039	.016	.064	.15	.46	1.07	1.64	2.71	3.84	5.41	6.64	10.83
2	.02	.04	.10	.21	.45	.71	1.39	2.41	3.22	4.60	5.99	7.82	9.21	13.82
3	.12	.18	.35	.58	1.00	1.42	2.37	3.66	4.64	6.25	7.82	9.84	11.34	16.27
4	.30	.43	.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	11.67	13.28	18.46
5	.55	.75	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	13.39	15.09	20.52
6	.87	1.13	1.64	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	15.03	16.81	22.46
7	1.24	1.56	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	16.62	18.48	24.32
8	1.65	2.03	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	18.17	20.09	26.12
9	2.09	2.53	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	19.68	21.67	27.88
10	2.56	3.06	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	21.16	23.21	29.59
11	3.05	3.61	4.58	5.58	6.99	8.15	10.34	12.90	14.63	17.28	19.68	22.62	24.72	31.26
12	3.57	4.18	5.23	6.30	7.81	9.03	11.34	14.01	15.81	18.55	21.03	24.05	26.22	32.91
13	4.11	4.76	5.89	7.04	8.63	9.93	12.34	15.12	16.98	19.81	22.36	25.47	27.69	34.53
14	4.66	5.37	6.57	7.79	9.47	10.82	13.34	16.22	18.15	21.06	23.68	26.87	29.14	36.12
15	5.23	5.98	7.26	8.55	10.31	11.72	14.34	17.32	19.31	22.31	25.00	28.26	30.58	37.70
16	5.81	6.61	7.96	9.31	11.15	12.62	15.34	18.42	20.46	23.54	26.30	29.63	32.00	39.29
17	6.41	7.26	8.67	10.08	12.00	13.53	16.34	19.51	21.62	24.77	27.59	31.00	33.41	40.75
18	7.02	7.91	9.39	10.86	12.86	14.44	17.34	20.60	22.76	25.99	28.87	32.35	34.80	42.31
19	7.63	8.57	10.12	11.65	13.72	15.35	18.34	21.69	23.90	27.20	30.14	33.69	36.19	43.82
20	8.26	9.24	10.85	12.44	14.58	16.27	19.34	22.78	25.04	28.41	31.41	35.02	37.57	45.32
21	8.90	9.92	11.59	13.24	15.44	17.18	20.34	23.86	26.17	29.62	32.67	36.34	38.93	46.80
22	9.54	10.60	12.34	14.04	16.31	18.10	21.24	24.94	27.30	30.81	33.92	37.66	40.29	48.27
23	10.20	11.29	13.09	14.85	17.19	19.02	22.34	26.02	28.43	32.01	35.17	38.97	41.64	49.73
24	10.86	11.99	13.85	15.66	18.06	19.94	23.34	27.10	29.55	33.20	36.42	40.27	42.98	51.18
25	11.52	12.70	14.61	16.47	18.94	20.87	24.34	28.17	30.68	34.38	37.65	41.57	44.31	52.62
26	12.20	13.41	15.38	17.29	19.82	21.79	25.34	29.25	31.80	35.56	38.88	42.86	45.64	54.05
27	12.88	14.12	16.15	18.11	20.70	22.72	26.34	30.32	32.91	36.74	40.11	44.14	46.96	55.48
28	13.56	14.85	16.93	18.94	21.59	23.65	27.34	31.39	34.03	37.92	41.34	45.42	48.28	56.89
29	14.26	15.57	17.71	19.77	22.48	24.58	28.34	32.46	35.14	39.09	42.56	46.69	49.59	58.30
30	14.95	16.31	18.49	20.60	23.36	25.51	29.34	33.53	36.25	40.26	43.77	47.96	50.89	59.70

Kritiska värden för D i Kolmogorov-Smirnov-testet.

Signifikansnivå

N	.20	.15	.10	.05	.01
1	.900	.925	.950	.975	.995
2	.684	.726	.776	.842	.929
3	.565	.597	.642	.708	.828
4	.494	.525	.564	.624	.733
5	.446	.474	.510	.565	.669
6	.410	.436	.470	.521	.618
7	.381	.405	.438	.486	.577
8	.358	.381	.411	.457	.543
9	.339	.360	.388	.432	.514
10	.322	.342	.368	.410	.490
11	.307	.326	.352	.391	.468
12	.295	.313	.338	.375	.450
13	.284	.302	.325	.361	.433
14	.274	.292	.314	.349	.418
15	.266	.283	.304	.338	.404
16	.258	.274	.295	.328	.392
17	.250	.266	.286	.318	.381
18	.244	.259	.278	.309	.371
19	.237	.252	.272	.301	.363
20	.231	.246	.264	.294	.356
25	.21	.22	.24	.27	.32
30	.19	.20	.22	.24	.29
35	.18	.19	.21	.23	.27
Over 35	$\frac{1.07}{\sqrt{N}}$	$\frac{1.14}{\sqrt{N}}$	$\frac{1.22}{\sqrt{N}}$	$\frac{1.36}{\sqrt{N}}$	$\frac{1.63}{\sqrt{N}}$

Sannolikheten för x (det minst frekventa tecknet) i Teckentestet

Nedan angivna sannolikheter (p) gäller för en riktad alternativhypotes. Är denna inte riktad skall p-värdet fördubblas.

N \ x	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
5	031	188	500	812	969	†										
6	016	109	344	656	891	984	†									
7	008	062	227	500	773	938	992	†								
8	004	035	145	363	637	855	965	996	†							
9	002	020	090	254	500	746	910	980	998	†						
10	001	011	055	172	377	623	828	945	989	999	†					
11		006	033	113	274	500	726	887	967	994	†	†				
12		003	019	073	194	387	613	806	927	981	997	†	†			
13		002	011	046	133	291	500	709	867	954	989	998	†	†		
14		001	006	029	090	212	395	605	788	910	971	994	999	†	†	
15			004	018	059	151	304	500	696	849	941	982	996	†	†	†
16			002	011	038	105	227	402	598	773	895	962	989	998	†	†
17			001	006	025	072	166	315	500	685	834	928	975	994	999	†
18			001	004	015	048	119	240	407	593	760	881	952	985	996	999
19				002	010	032	084	180	324	500	676	820	916	968	990	998
20				001	006	021	058	132	252	412	588	748	868	942	979	994
21				001	004	013	039	095	192	332	500	668	808	905	961	987
22					002	008	026	067	143	262	416	584	738	857	933	974
23					001	005	017	047	105	202	339	500	661	798	895	953
24					001	003	011	032	076	154	271	419	581	729	846	924
25						002	007	022	054	115	212	345	500	655	788	885

Anm. 1) † anger att p = 1

2) I tabellen har decimalkommat uteslutits, vilket innebär att t.ex. sannolikheten (p) för att få x = 5 vid N = 25 är 0.002.

Kritiska värden för ρ (Spearmans rangkorrelationskoefficient).

Nedan angivna ρ-värden gäller för en riktad alternativhypotes.

N.	Signifikansnivå	
	.05	.01
4	1.000	
5	.900	1.000
6	.829	.943
7	.714	.893
8	.643	.833
9	.600	.783
10	.564	.746
12	.506	.712
14	.456	.645
16	.425	.601
18	.399	.564
20	.377	.534
22	.359	.508
24	.343	.485
26	.329	.465
28	.317	.448
30	.306	.432

LITTERATURREFERENSER

Ferguson,G.A. Statistical Analysis in Psychology and Education (2:nd edition). New York, 1966.

Siegel,S. Nonparametric Statistics. New York, 1956.

Svensson,A. Relative Achievement. School performance in relation to intelligence, sex and home environment. Stockholm, 1971.

ÖVNINGSUPPGIFTER

1. Vilket av följande påståenden är riktigt beträffande en jämförelse mellan parametriska och icke-parametriska metoder?
   a) de icke-parametriska metoderna har högre känslighet än de parametriska
   b) de parametriska metoderna ställer högre krav på fördelningen hos den studerande variabeln än de icke-parametriska
   c) såväl parametriska metoder som icke-parametriska metoder kräver att data uttrycks lägst i intervallskala
   d) endast parametriska metoder ger möjlighet till korrelationsstudier

2. Ett antal elever har efter att ha matchats med avseende på begåvning utsatts för fyra olika undervisningsmetoder. Du vill sedan pröva huruvida undervisningsmetoderna har olika effekt. Detta är exempel på en undersökning med:
   a) 1 uppsättning oberoende mätvärden
   b) 1 uppsättning beroende mätvärden
   c) k uppsättningar oberoende mätvärden
   d) k uppsättningar beroende mätvärden

3.

Individ	Verbal förmåga	Kaffedrickningskapacitet
Vilhelmina	1	8
Fredrika	2	7
Dorotea	3	6
Desideria	4	5
Evelina	5	4
Alexandra	6	3
Gabriella	7	2
Ottilia	8	1

I tabellen ovan anges rangordningen bland åtta personer dels i verbal förmåga, dels i kaffedrickningskapacitet. Om man räknar ut sambandet mellan färdigheterna i de båda aktiviteterna erhålles följande koefficient:
   a) +1.00
   b) +0.50
   c) -0.50
   d) -1.00

4. I en internationell drinkblandartävling deltar tio barmästare. Barmästarnas skicklighet bedöms av sex kvalificerade avsmakare, vilka var och en rangordnar de tio tävlande. För att få ett mått på domarnas samstämmighet använder man sig av följande koefficient:
   a) Spearmans rangkorrekationskoefficient
   b) Kontingenskoefficienten
   c) Kruskal-Wallis korrelationskoefficient
   d) Kendalls konkordanskoefficient

5. Tre grupper om vardera 200 elever från tre olika skolformer får besvara en fråga rörande trivseln i skolan, varvid svaren fördelar sig på följande svarsalternativ: "mycket bra", "bra", "varken bra eller dåligt", "ganska dåligt", "mycket dåligt". Vi vill nu studera om det finns några skillnader mellan grupperna, Hur många frihetsgrader baseras den avslutande statistiska analysen på?
   a) 2
   b) 4
   c) 6
   d) 8

6. I en undersökning tillfrågades 100 män och 100 kvinnor om de ansåg att man borde höja skatten på alkoholhaltiga drycker. Bland männen var 30 positiva och 70 negativa till förslaget. Bland kvinnorna var 70 positiva och 30 negativa. För att undersöka om det finns någon signifikant könsdifferens görs en chi^2-analys, varvid man får följande värde:
   a) 4
   b) 8
   c) 16
   d) 32

7. Majken, Signe och Alma är storkonsumenter av en viss vara. Av denna vara förekommer följande märken: Göteborgs prima fint, Röda lacket, Ljunglöfs etta, General, Karlskrona och Smoke less. Fröknarna avsmakar samtliga märken och rangordnar dem efter mustighet. För att få ett mått på samstämmigheten använder man sig av följande koefficient:
   a) Spearmans rangkorrelationskoefficient
   b) Kontingenskoefficienten
   c) Kruskal-Wallis korrelationskoefficient
   d) Kendalls konkordanskoefficient

8. En observatör har bedömt 15 pojkar och 20 flickor med avseende på skolanpassning i en 9-gradig skala. Observatören vill sedan studera om det finns någon skillnad mellan pojkar och flickor beträffande skolanpassning. Vilken metod bör han då använda?
   a) Friedmans två-vägs variansanalys
   b) Fishers exakta sannolikhetstest
   c) Mediantestet
   d) Kolomogorov-Smirnovs test

9. I en undersökning tillfrågades 100 män i Grönköping och 100 män i Svedala om de använde pyjamas eller nattskjorta. Av männen i Grönköping använde hälften nattskjorta och hälften pyjamas. I Svedala använde 40 nattskjorta och 60 pyjamas. För att undersöka om det finns några skillnader mellan männen på de båda orterna med avseende på deras nattliga vanor gör Du en chi^2-analys och får då följande värde:
   a) 2.02
   b) 3.33
   c) 4.28
   d) 5.71

10.

Individ	Höjdhopp	Längdhopp
Adam	1	2
Bertil	2	3
Cesar	3	1
David	4	6
Erik	5	5
Fredrik	6	4

I tabellen ovan anges rangordningen bland sex personer dels i höjdhopp, dels i längdhopp. Om man räknar ut sambandet mellan färdigheten i de båda idrottsgrenarna erhålles följande koefficient:

a) 0.40
b) 0.50
c) 0.60
d) 0.70

11. 100 f.d. elever från var och en av fackskolans ekonomiska grenar fick ange grenvalets lämplighet i förhållande till sina aktuella arbeten. Resultatet blev:

	Gren			
Alternativ	Ka	Di	Ad	Esp
Mycket lämpligt	22	15	14	21
Ganska lämpligt	50	25	24	42
Varken lämpligt eller olämpligt	19	37	37	28
Ganska olämpligt	3	8	13	6
Mycket olämpligt	1	4	7	2
Ej svar	5	11	5	1
N	100	100	100	100

Pröva på 5 %-nivån huruvida det finns någon skillnad mellan grenarna beträffande grenvalets lämplighet.

12. Vid en drinkblandartävling fick en jury bestående av 100 utvalda män avsmaka 8 drinkar av varierande styrka och därefter ange vilken de tyckte bäst om. Resultatet visas i nedanstående tabell, där drinkarna är rangordnade med avseende på styrka.

Drink	N
Snarker (starkast)	19
Gäsper	17
Trötter	13
Toker	12
Glader	11
Kloker	10
Tyster	9
Butter (svagast)	9

Pröva på 1 %-nivån om rösterna fördelat sig slumpmässigt på de 8 drinkarna.

13. I en enkätundersökning bland skolelever i Göteborg uttogs ett slumpmässigt stickprov bestående av 1000 elever. Av dessa erhölls svar från 800 elever vilka fördelade sig på socialgrupp enligt nedanstående:
Socialgrupp I: 110; Socialgrupp II: 320; Socialgrupp III: 370.

Från offentlig statistik vet man att socialgruppsfördelningen i Göteborg är:

I: 12 %; II: 40 %; III: 48 %

Pröva på 5 %-nivån om socialgruppsfördelningen bland de svarande kan anses representativ för Göteborg.

14. Vid anställning av sekreterare i ett bolag uttogs en jury bestående av en manlig och en kvinnlig bedömare som skulle göra en sammanfattande skattning av de 10 sökandes kvalifikationer i olika avseenden. Skattningen gjordes i en 7-gradig skala där 7 innebär "mycket lämplig" och 1 "mycket olämplig".

Sökande	Manlig bedömare	Kvinnlig bedömare
A	7	2
B	7	2
C	5	2
D	4	4
E	4	4
F	4	7
G	3	5
H	2	1
I	2	6
J	1	3

a) Vilket samband finns mellan bedömarnas skattningar? (Korrigera för bundna ranger)
b) Är sambandet signifikant på 5 %-nivån?

15. I sin doktorsavhandling fann en känd svensk, vid namn Sven, att det förelåg ett visst samband mellan studentbetyg i gymnastik och giftermålsfrekvens. En kvinnlig gymnastikförening, som ville utnyttja detta i reklamsyfte vågade inte helt lita på Svens utsaga, varför man beslöt att göra en jämförelse mellan 50 gifta och 50 ogifta kvinnors gymnastikbetyg. Resultatet blev:

Betyg	Gifta kvinnor	Ogifta kvinnor
A	3	2
a	10	6
AB	17	11
Ba	10	15
B	8	11
BC	1	3
C	1	2
	50	50

Pröva på 5 %-nivån om det fanns någon skillnad mellan gifta och ogifta kvinnors gymnastikbetyg. (Betygen betraktas som uttryckta i ordinalskala).

16. 50 svenskar och 50 spanjorer tillfrågades om de känner sig mest attraherade av blondiner eller mörkhåriga kvinnor. Av svenskarna svarade 20 och av spanjorerna svarade 30 att de föredrog blondiner. De övriga föredrog mörkhåriga kvinnor. Pröva på 5 %-nivån om det är riktigt att spanjorer verkligen är mer attraherade av blondiner än svenska män är.

17. 5 bilförare har provkört 5 olika märken och rangordnat dessa med avseende på väghållningsförmåga. Bilarna har sedan skattats i en 5-gradig skala, där 8 anger bästa och 1 anger sämsta värde. Följande skattningar erhölls:

		Bilmärke				
		A	B	C	D	E
Förare	I	6	2	4	3	3
"	II	5	4	3	1	2
"	III	8	6	4	2	5
"	IV	5	5	4	3	2
"	V	6	2	3	2	2

a) Pröva på 5 %-nivån om det föreligger någon skillnad mellan bilmärkena.

b) Finns det någon statistisk säkerställd överensstämmelse mellan bilförarnas bedömningar? Pröva detta på 5 %-nivån.

18. 30 föräldrar fick ange om de tyckte att skolans sexualundervisning kunde anses vara lämplig. 18 svarade "ja" och 12 "nej". Efter att ha fått en grundlig information om denna undervisning fick de på nytt besvara denna fråga. Utav de som först svarat ja hade 3 ändrat uppfattning och svarade nu nej. Bland de som först svarat nej hade 9 ändrat uppfattning och svarade efter informationen ja. Har informationen haft någon effekt? (Pröva på 5 %-nivån).

19. I ett välskrivningsprov rangordnades 8 pojkar och 8 flickor efter handstilens kvalitet. Resultatet blev:

1. Greta	7. Frans	12. Frida
2. Artur	8. Frideborg	13. Egon
3. Stina	9. Berta	14. Albin
4. Valborg	10. Sune	15. Melker
5. Sixten	11. Helga	16. Pär
6. Ebba		

a) Pröva på 5 %-nivån om flickorna har bättre handstil än pojkarna.

20. En grupp manliga och en grupp kvinnliga studenter fick ange sin inställning till undervisning i fönsterlösa lokaler, dels innan de haft någon sådan undervisning dels efter. Resultatet blev:

Manliga

		Efter pos.	Efter neg.
före	pos.	7	12
	neg.	2	13

Kvinnliga

		Efter pos.	Efter neg.
före	pos.	11	22
	neg.	8	20

a) Pröva på 5 %-nivån om de manliga respektive kvinnliga studenterna ändrat uppfattning efter att ha fått erfarenhet av denna typ av lokaler

b) Finns det någon tendens till att den ena gruppen ändrat sig mera än den andra gruppen?

21. 10 fyrabarnsmammor fick efter varje förlossning ange om de kunde tänka sig använda p-piller. Resultatet blev:

	1:a barnet	2:a barnet	3:e barnet	4:e barnet
Ada	nej	nej	nej	nej
Beda	nej	ja	ja	ja
Clara	nej	nej	ja	ja
Doris	ja	ja	ja	ja
Emma	nej	ja	nej	ja
Fina	nej	nej	ja	ja
Greta	ja	nej	ja	ja
Hilma	nej	ja	ja	ja
Ina	ja	ja	ja	ja
Josefina	nej	nej	nej	nej

Pröva på 5%-nivån om mammorna hade ändrat inställning till p-piller.

22. 15 män fick ange sin inställning till äktenskapet dels innan de gifte sig dels efter att ha varit gifta i 2 år. Inställningen skattades i en skala från 5=toppen till 1=gräsligt obehagligt. Resultatet blev:

	före	efter		före	efter		före	efter
A	4	3	F	2	4	K	4	3
B	3	5	G	3	2	L	2	1
C	3	4	H	2	1	M	3	1
D	3	2	I	3	3	N	1	1
E	5	1	J	4	3	O	2	5

Pröva på 5 %-nivån om deras inställning hade förändrats.

23. På Grönköpings KK föddes det under en vecka följande antal barn:

	pojkar	flickor		pojkar	flickor
Måndag	10	7	Fredag	7	9
Tisdag	15	9	Lördag	12	10
Onsdag	7	12	Söndag	8	6
Torsdag	4	6			

a) Föreligger det någon skillnad mellan de olika veckodagarna beträffande antalet nedkomster? (Pröva på 5 %-nivån).

b) Föreligger det någon skillnad mellan pojkar och flickor beträffande "födelsedag"? (Pröva på 5 %-nivån).