

Elisabeth Svensson, professor i statistik, speciellt biostatistik, institutionen för ekonomi, statistik och informatik (ESI), Örebro universitet elisabeth.svensson@esi.oru.se

Vad är behandlingseffekt om patienten blev bättre men ingen vet hur mycket?

Analys av förändring när datamaterialet består av ordnade kategorier

II Vid utvärdering av möjlig effekt av en behandling, intervention eller annan åtgärd studeras samma individ minst två gånger, nämligen före och efter behandlingen i fråga. Datamaterialet kommer alltså att bestå av par av värden från varje individ, vilket måste beaktas vid valet av statistisk analysmetod. En annan viktig faktor, som är avgörande för valet av statistisk verktygslåda, är datamaterialets matematiska egenskaper (Fakta 1 och Fakta 2).

Variablers mätbarhet

Vilken eller vilka variabler påverkas av behandlingen, och hur är dessa operationellt definierade? Utvärdering av en behandling kan inbegripa olika typer av variabler, t ex sådana som operationellt definieras utifrån fysikens eller kemins lagar, såsom tryck, temperatur, längd och koncentration. Mätning av dessa variabler ger upphov till kontinuerliga, kvantitativa data även om man i vissa studier väljer att kategorisera datamaterialet till mätnivåer med begränsade matematiska egenskaper [1, 2].

En annan stor grupp av variabler representerar olika typer av kvaliteter, som normalt inte kan operationellt definieras utifrån standardiserade fysikaliska lagar utan utifrån subjektiva värderingar eller bedömningar. Det kan vara livs-, hälso- eller vårdkvalitet, förmåga, funktion, sjukdomsgrad eller någon annan mer eller mindre komplex egenskap som bedöms av expert, närstående person eller av studiepersonen själv. Datainsamlingen består då i att den aktuella variabeln klassificeras eller bedöms på en skala med ett antal kategorier. Skalkategorierna representerar ofta en ordning och kan ha mycket varierande utseende, från siffror, figurer, positioner på en linje till mer konkreta ord och beskrivande meningar.

Frestelsen att använda matematik

Oavsett utseendet består ett datamaterial från bedömningar på skalor av ordnade kategorier [1, 2]. Karakteristiskt för denna mätnivå, som brukar kallas för ordinal, är att datamaterialet har en ordningsstruktur men saknar egenskaperna storlek och avstånd.

En person som t ex bedömer sin oro till 4 på en 5-gradig skala är inte självklart dubbelt så orolig som en person med oro 2, och en ändrad bedömning från »mycket orolig« till »lite orolig« antyder minskad oro, men informationen innehåller ingen uppgift om storleken på förändringen, utan dataparen ger information endast om huruvida patienten har blivit bättre eller sämre, inte om förändringens storlek. En jämförelse av upprepade bedömningar på en skala ger informa-

Sammanfattat



Bedömningar på olika typer av skalor och bedömningsformulär ger upphov till ordnade kategoridata. Mätvärdena i denna typ av data representerar enbart en ordningsstruktur och saknar information om storlek och avstånd mellan kategorier.

Dessa begränsade matematiska egenskaper betyder att de mest välkända statistiska metoderna, såsom t-test och Wilcoxon's teckenrangtest, inte är användbara för analys av behandlingseffekt eller annan förändring, eftersom dessa test baseras på differenser mellan parade värden.

I artikeln presenteras klassiska metoder som är tillämpliga för analys av förändring i datamaterial som består av ordnade kategorier, som först måste dikotomiseras. Metoderna passar även för data som är dikotoma redan från början.

Klinisk forskning



Läs mer på www.lakartidningen.se

tion endast om förändringens riktning mot en högre, oförändrad eller lägre kategori, men information saknas om förändringens storlek [1-5]. Det innebär att addition och subtraktion inte är meningsfulla matematiska operationer på ordinaldata, och detta får påtagliga konsekvenser vid analys av förändring i bedömningar på skala. Dessa begränsade matematiska egenskaper hos ordinaldata kan tyckas besvärande, eftersom de vanligaste och mest välkända statistiska metoderna är baserade på matematiska beräkningar av summor, differenser och kvoter.

Det är ett väldokumenterat faktum att många faller för frestelsen att hantera ordinaldata som om de vore kvantitativa data med matematiska egenskaper, såsom väl definierade

II Fakta 1

Statistiskt viktiga mätnivåer hos ett datamaterial

- Dikotoma data (binära data, två kategorier)
- Kategoridata utan ordningsstruktur (nominaldata)
- Kategoridata med ordningsstruktur (ordinaldata)
- Kvantitativa antalsdata (diskreta data)
- Kvantitativa, kontinuerliga data indelas i icke-negativa, intervallbegränsade eller obegränsade (intervalldata, kvotdata).

II Fakta 2

En skala

En skala (*scale, rating scale*) består av en uppsättning ordnade alternativ (kategorier) eller av en linje (visuell analog skala, VAS).

Mätningen/registreringen består av en bedömning (*assessment, judgement, rating*) på skalan.

När individen själv bedömer sitt tillstånd, t ex smärta, oro, välbefinnande, livskvalitet, används ofta uttrycket självskattning (*self-assessment*).

avstånd mellan kategorierna och känd standardiserad storlek [3, 4].

Syftet med artikeln är att ge exempel på klassiska statistiska metoder som är tillämpliga för analys av förändring i en variabel som bedöms på en skala. Valet av frågeställning och dess konsekvens på statistisk bearbetning men också på den information som de statistiska metoderna ger demonstreras på ett fiktivt datamaterial. Frågeställningarna avser beskrivning av förändring, hypotesprövning, intervallskattning. Jag har valt att utgå från en kortfattad resultatformulering och där efter redogöra för statistiska metoder och även kommentera hur datamaterialets mätnivå avgör val av statistisk metod.

Variabeln är »besvär«

Eftersom syftet är att visa på lämpliga statistiska verktyg för utvärdering av förändring i en egenskap/ett tillstånd som skattas på en skala kallar jag variabeln allmänt för »besvär«, vilket skulle kunna betyda illamående, klåda, smärta eller någon annan form av besvär. Antag att graden av besvär bedöms på en skala med sex ordnade verbala kategorier: inga besvär (A), mycket lätta besvär (B), måttliga besvär (C), påtagliga besvär (D), mycket stora besvär (E) och extrema besvär (F). Förändringen avser behandlingseffekt, och 20 individer har skattat sina besvär före och efter behandlingen.

Beskrivning av förändring

- Vid uppföljningen bedömde 14 av 20 patienter (70 procent) sina besvär till en lägre nivå och två (10 procent) till en högre nivå än vid bedömningen före behandlingen.

En allmän regel är att statistisk bearbetning av förändringsdata skall ta hänsyn till att varje individ representeras av två vär-

		Upplevda besvär före behandling						Totalt
		A	B	C	D	E	F	
Besvär efter behandling	F					1		1
	E				1			1
	D				1	1	1	3
	C			3	1		2	6
	B			2	1	1	1	5
	A		2	1			1	4
Totalt			2	6	4	3	5	20

Figur 1. Kontingenstabell över frekvensfördelningen av datapar som visar hur 20 individer bedömde sina besvär på en skala från A (inga besvär) till F (extrema besvär) före och efter en behandling.

II Fakta 3

Frekvensfördelning

En statistisk beskrivning av förändring i ordnade kategoridata kan baseras på antalet eller andelen (procent) patienter med högre, oförändrad eller lägre nivå vid uppföljningen.

Detta illustreras av frekvensfördelningen i en kvadratisk korsstabell (kontingenstabell) och med en eventuell komplettering med median- och kvartilkategorierna för respektive tillfälle.

Den. Datamaterialet, som innehåller information om förändringen i skattad besvärnivå mellan de båda tillfällena (före och efter behandling), består av par av kategorier, betecknade från A (inga besvär) till F (extrema besvär), och förändringsmönstrets utseende beskrivs lämpligen av frekvensfördelningen av hur dessa datapar fördelar sig på de 6 × 6 möjliga celler i en korstabell (kontingenstabell) som representerar de 36 möjliga kombinationerna av par av bedömningar på en skala med sex kategorier (Figur 1, Fakta 3).

Av frekvensfördelningen i kontingenstabellen (Figur 1) framgår att fyra patienter (20 procent) hade samma besvärnivå vid båda tillfällena. Dataparen i cellerna nedanför huvuddiagonalen representerar de 14 patienter (70 procent) som angav en lägre besvärnivå, och paren (D, E) och (E, F) representerar två patienter med en högre besvärnivå efter behandlingen. Frekvensfördelningen i tabellen beskriver det individuella förändringsmönstret, och de två frekvensfördelningarna på marginalerna ger kompletterande information om hur gruppens bedömningar fördelade sig på besvärsskalan vid respektive tillfälle. Av dessa kan vi konstatera att gruppens genomsnittliga besvärnivå har ändrats, eftersom mediankategorin var påtagliga besvär (D) före och måttliga besvär (C) efter behandlingen.

När medianen inte är entydig

För material som består av ett jämnt antal observationer av ordinaldata är medianen inte alltid entydigt bestämd, eftersom egenskapen avstånd saknas. Det innebär att medianen inte kan beräknas utifrån medelvärdet av de två mittersta observationerna i rangordning, som brukligt är för kvantitativa data. I vårt material av 20 individer var de två mittersta individernas värden lika, men vad är medianen då det inte är så,

II Fakta 4

Median

definieras matematiskt som det värde som är sådant att både sannolikheten att en observation har ett värde mindre än medianen och sannolikheten att en observation har ett värde större än medianen är högst 1/2 [6].

För ordinaldata med ett jämnt antal observationer där de mittersta värdena skiljer sig åt är inte medianen entydigt bestämd, utan båda kategorierna representerar medianen.

II Fakta 5

P-värde

är sannolikheten för att i en stickprovsstudie få det observerade resultatet eller ett ännu extremare när man utgår från att nollhypotesen är sann för den bakomliggande populationen av data.

II Fakta 6

Sannolikhetsfördelning

- Normalfördelningen (Gaussfördelningen) och χ^2 -fördelningen är två exempel på teoretiska sannolikhetsfördelningar som statistiska test kan utnyttja vid bestämning av P-värdet.
- Teststatistikor som under nollhypotesen följer den standardiserade normalfördelningen brukar betecknas med z , och $z = (\pm) 1,96$ betyder att det observerade stickprovsresultatet som är upphov till detta eller ännu extremare z -värden skulle kunna inträffa i 5 av 100 studier om nollhypotesen vore sann. Detta brukar leda till att nollhypotesen förkastas till förmån för forskarens hypotes, med risken $P = 0,05$ för ett felaktigt beslut.
- Ett antal test utnyttjar χ^2 -fördelningen, som är en sannolikhetsfördelning och som består av enbart positiva värden och vars fördelning beror av antalet frihetsgrader.

t ex om individerna nummer 10 och 11 i rangordning har bedömt sig till mycket lätta besvär (B) respektive måttliga besvär (C)? Svaret är att båda värdena representerar medianen [6]. Om det känns viktigt att ange en mediankategori skall man i förväg bestämma hur medianen skall definieras. Samma förfarande gäller för bestämning av kvartiler och andra percentiler. Av marginalfördelningen (Figur 1) för skattade besvär före behandling framgår att måttliga besvär (C) (individerna 5 och 6) utgör första kvartilen, men som tredje kvartil kan både kategori E och F användas, eftersom individerna 15 och 16 har olika värden. Ett förslag är då att använda den kategori som representerar det sämre tillståndet, extrema besvär (F), och att konsekvent göra så i studien (Fakta 4).

Hypotesprövning och P-värde

Vetenskapliga rapporter brukar inte innehålla enbart statistiska beskrivningar utan även ett eller flera P-värden [4, 7, 8]

(Fakta 5). Statistisk analys av datamaterialet i Figur 1 kan ge underlag till följande summariska resultatformulering.

- Efter behandlingen hade 14 patienter (70 procent) lägre, fyra oförändrad och två högre besvärnivå än före behandlingen. Enligt studien kan man dra slutsatsen att behandlingen kommer att medföra minskade besvär för majoriteten av patienterna i en relevant population ($P=0,006$).

Vilka statistiska analysmetoder ligger bakom resultatet och vad är egentligen det statistiska budskapet värt? Den generella slutsatsen baseras på statistisk hypotesprövning, i vilken forskarens hypotes ställs mot den statistiska nollhypotesen.

Enligt nollhypotesen förväntas det i en relevant population vara lika chans (50 procent) till förbättring som till försämring, vilket innebär att behandlingen saknar effekt i det långa loppet. Förutsatt att nollhypotesen vore sann, dvs att det inte finns någon behandlingseffekt, hur vanligt är det då att få det aktuella studieresultatet? Observera att inför analysen används endast de 16 patienter som uppvisar en förändring och att förändringen uttrycks dikotomt i ett av alternativen lägre respektive högre nivå efter behandling [4].

Den statistiska uppgiften blir att skatta sannolikheten att bland 16 patienter observera den aktuella obalansen (14 mot 2) eller en ännu extremare obalans i behandlingseffekt, när man under nollhypotesen förväntas observera lika många (8) med lägre som med högre nivå efter behandling. Denna sannolikhet kan skattas exakt med hjälp av binomialfördelningen, som är den sannolikhetsfördelning som beskriver en population av dikotoma data. Man kan också utnyttja att binomialfördelningen i vissa fall kan approximeras till Gaussfördelningen (normalfördelningen) och använda teckentestet för test av skillnad mellan parade proportioner (Fakta 6 och 7).

Förkasta eller inte förkasta nollhypotesen

Eftersom testet används på dikotoma data blir inverkan av varje observation på testvärdet stort, speciellt för små datamaterial. Därför bör teststatistikan (z -värdet) korrigeras inför bestämning av P-värdet [4, 9]. I vårt exempel är $z_c=2,75$, vilket antyder att det är högst osannolikt att vårt stickprovsresultat skulle kunna inträffa om nollhypotesen vore sann, eftersom detta värde motsvarar en sannolikhet på 0,6 procent ($P=0,006$) i normalfördelningen. Det innebär att det är ytterst osannolikt att minst 14 individer av 16 skulle uppvisa ändrad besvärnivå vid det andra mätillfället om behandlingen hade varit utan effekt.

Resultatet ger oss starka skäl (evidens) att förkasta nollhypotesen till förmån för mothypotesen, som säger att behandlingen i det långa loppet kommer att ha en besvärssänkande effekt på majoriteten i en relevant population. Det alternativa McNemars test ger $X^2 = 7,56$, vilket är ett ovanligt värde i χ^2 -tvåfördelningen för en frihetsgrad, $P=0,006$, alltså samma resultat som med teckentestet för skillnad mellan parade proportioner [4, 9].

Det bör påpekas att det finns flera olika test vars teststatistikor är χ^2 -fördelade under nollhypotesen. De olika testen för skillnad i kategorifördelning mellan två eller flera oberoende grupper kallas olyckligt nog χ^2 -test, och ofta betecknas även teststatistikan med χ^2 , som är beteckningen för den teoretiska sannolikhetsfördelningen. Vid analys av förändring skall ett test av skillnad mellan parade/beroende data användas, och valet är då antingen McNemars test eller teckentestet för skillnad mellan parade proportioner.

Statistiska test ger inte någon information om hur mycket bättre patientgruppen blir utan bara att stickprovsresultatet avviker starkt från ett förväntat resultat om nollhypotesen hade varit sann. Nollhypotesen är antingen sann eller inte. Vi

II Fakta 7

Teckentestet

för test av skillnad mellan parade proportioner kan skrivas

$$z = \frac{b - c}{\sqrt{b + c}}$$

där b och c är antal observationer som förändrats till det bättre respektive till det sämre vid uppföljningen.

Teckentestet med kontinuitetskorrektur kan skrivas

$$z_c = \frac{b - c \pm 1}{\sqrt{b + c}}$$

och tecknet framför 1 väljs så att täljarens siffermässiga skillnad minskar, t ex $14 - 2 - 1 = 11$ eller $2 - 14 + 1 = -11$. Teststatistikan z alternativt z_c följer den standardiserade normalfördelningen.

McNemars test

är ett identiskt test för skillnad mellan parade proportioner

$$\chi^2 = \frac{(b - c \pm 1)^2}{b + c}$$

där teststatistikan χ^2 följer χ^2 -fördelningen med 1 frihetsgrad om nollhypotesen är sann.

har inte facit. Men med hjälp av en bra upplagd studie och statistisk metod kan vi få underlag till att fatta beslutet att förkasta eller inte förkasta nollhypotesen, och risken att förkasta en sann nollhypotes uttrycks i P-värdet. Observera att nollhypotesen »lika behandlingseffekt« inte kan testas med denna typ av hypotesprövning, utan det krävs en annan analysteknik för att bevisa sådan ekvivalens [10, 12].

Den statistiska felmarginalen, konfidensintervallet

Den stora andelen (70 procent) av gruppen som angav en lägre besvärnivå i förhållande till andelen som angav en högre (10 procent) utgör ett övertygande analysunderlag till att förkasta nollhypotesen och dra slutsatsen att behandlingen i genomsnitt kommer att ha en positiv inverkan på besvärnivån hos en relevant population ($P=0,006$). Men kommer – som i vårt stickprov – 60 procentenheter fler patienter i det långa loppet, dvs i populationen, att sänka sin besvärnivå efter behandlingen?

Av det statistiska testet kan vi vara säkra på att andelarna som sänker respektive höjer sin besvärnivå inte är lika. Med hjälp av en statistisk metod kan vi beräkna ett intervall, som skulle kunna kallas »den statistiska felmarginalen«, nämligen det 95-procentiga konfidensintervallet (95 procents CI) för skillnaden mellan parade proportioner [4, 7-12]. Resultatet kan då formuleras så här.

- Efter behandlingen hade 14 av 20 patienter (70 procent) lägre och två (10 procent) högre besvärnivå än före behandlingen (95 procents CI för skillnaden i procentenheter är 31–89, $P=0,006$) (Fakta 8, 9 och 10).

Vad ger konfidensintervallet för information?

Den statistiska metod som vi använder för att beräkna konfidensintervallet är sådan att på ett stort antal upprepade likadana studier skulle den andel av de beräknade intervall som anges av konfidensgraden täcka parametervärdet. Konfidensgraden är ofta 95 procent, vilket betyder att 95 av 100 inter-

II Fakta 8

Parameter

kallas ett statistiskt mått för en population och betecknas ofta med grekiska bokstäver, t ex populationens medelvärde (väntevärde) μ , standardavvikelse σ , andel π , skillnad i andel $\Delta\pi$ och korrelationskoefficient ρ .

Parametervärdet är ett fixt populationsvärde (en »sanning«), som skattas med hjälp av konfidensintervall för stickprovresultatet (intervallskattning).

II Fakta 9

Konfidensgränserna

konstrueras utifrån observerade data med en statistisk metod som – om den används på ett oändligt antal replikat av studien – skulle ge intervall som skulle täcka parametern med den sannolikhet som konfidensgraden anger, dvs 95 procent för 95-procentigt konfidensintervall.

II Fakta 10

Beräkning av 95-procentigt konfidensintervall

för skillnaden mellan parade proportioner $\Delta p \pm 1,96 \text{ SE}(\Delta p)$ utgår från skillnaden mellan andelarna individer med förändrade (minskade respektive ökade) besvär $\Delta p = (b - c)/n$.

I Figur 1 är $b=14$, $c=2$ och totala antalet patienter $n = 20$.

Osäkerheten ges av medelfelet SE (standard error) för skillnad mellan parade proportioner

$$\text{SE}(\Delta p) = \frac{1}{n} \sqrt{b + c - \frac{(b - c)^2}{n}}$$

vall för upprepade likadana studier skulle täcka parametervärdet. Nu utförs endast en studie, och ett 95 procents konfidensintervall beräknas. Vi vet inte om just detta intervall är ett av dem som täcker parametern, men vi kan vara ganska säkra på att det gör det, och konfidensgraden uttrycker denna säkerhet. Så enligt den aktuella studien kan vi vara 95-procentigt säkra på att efter behandling av en relevant population förväntas mellan 31 och 89 procentenheter fler patienter få minskade besvär snarare än ökade besvär. Baserat på det aktuella stickprovresultatet uppskattas parametervärdet ($\Delta\pi$), dvs den »sanna skillnaden« i andelen patienter med förändrade besvär, till ett värde mellan 31 och 89 procentenheter.

Beräkning av konfidensintervallet för förändring

Beräkningen av det 95-procentiga konfidensintervallet för skillnaden mellan parade proportioner baseras på antalet patienter med skattad lägre besvärnivå ($b=14$) och med skattad högre nivå ($c=2$) och totala antalet patienter $n=20$ [4]. Me-

		Upplevda besvär före behandling						Totalt
		A	B	C	D	E	F	
Besvär efter behandling	F					5		5
	E				5			5
	D				5	5	5	15
	C			15	5		10	30
	B			10	5	5	5	25
	A		10	5			5	20
Totalt			10	30	20	15	25	100

Figur 2. Frekvensfördelningen över förändring i skattade besvär bedömda på en skala från A (inga besvär) till F (extrema besvär) före och efter en behandling.

delfelet för den procentuella skillnaden (0,60) är 0,148, vilket ger ett 95-procentigt konfidensintervall mellan 31 och 89 procentenheter. Konfidensintervallet täcker inte nollhypotesens skillnad mellan proportioner – som är 0 procent – vilket också verifieras av att nollhypotesen förkastades på starka grunder ($P=0,006$) vid hypotesprövningen. Däremot är konfidensintervallet brett, vilket betyder att det statistiska underlaget på 20 patienter är så litet att stickprovsresultatets 60 procentenheter i det långa loppet skulle kunna betyda allt från 31 till 89 procentenheter.

Storleken har betydelse

Det fiktiva exemplet består av par av subjektiva skattningar från 20 individer. Resultaten visar på statistiskt säkerställd förändring av besvärnivån, men konfidensintervallet är brett, vilket betyder att det är svårt att skapa ett generellt mått på behandlingens effekt för en relevant population, och det skulle vara statistiskt svårt att påvisa eventuella skillnader i behandlingseffekt mellan olika grupper.

Figur 2 visar samma förändringsmönster som Figur 1 men med ett fem gånger så stort material. Andelen patienter med lägre besvärnivå efter behandling är samma som tidigare, men underlaget ger starkare skäl att förkasta nollhypotesen och även smalare konfidensintervall (Tabell I).

Tabell I. Resultat av analyserna för samma grad av förändring men för olika stora stickprov.

	N=20	N=100
Antal (andel i procent) med lägre/högre besvärnivå efter behandling	14/2 (70/10)	70/10 (70/10)
Skillnad mellan parade andelar (procentenheter)	60	60
95-procentigt CI för skillnaden (procentenheter)	31–89	47–73
Teckentest z_c	2,75 ($P=0,006$)	6,60 ($P<0,0001$)

Val och konsekvens

Forskningens kvalitet bestäms av den svagaste länken i studien. Valet av mätmetod och studiedesign får konsekvenser för valet av statistiska metoder, så en eventuell kontakt med statistiker skall ske tidigt. Altman [4, 13] har granskat ett stort antal vetenskapliga artiklar och nämner bl a nedanstående grundläggande fel i analys och presentation:

- att analysera parade data som oparade
- att behandla upprepade observationer på individen som oberoende
- att inte ta hänsyn till ordnade kategorier
- att använda analysmetoder när förutsättningar för dem inte är uppfyllda
- att ge konfidensintervall som innehåller omöjliga värden
- att presentera resultat endast som ett P-värde.

Teckentestet, McNemars test och konfidensintervallen för proportioner är typiska testmetoder för dikotoma data, dvs för datamaterial som är grupperade i två kategorier [1, 4, 9]. På grund av de begränsade matematiska egenskaperna hos ordinaldata är det dessa klassiska metoder som är lämpliga för analys av förändring i ordinaldata, eftersom förändring i ordinaldata inte kan definieras av skillnad mellan värden före och efter. Konsekvensen är att den mera utförliga information som finns i data som består av fler än två ordnade kategorier går förlorad. Om forskarens frågeställning och syfte överensstämmer med ovan nämnda metoder är det dessa som är relevanta för analys av förändring i ordinaldata. Det finns dock mycket mer information att hämta ur ett datamaterial som består av par av ordnade bedömningar för att utvärdera förändring, vilket kommer att belysas i en kommande artikel. ➔

kvartsannons

Andra vanliga statistiska test av förändring är det icke-parametriska Wilcoxon's teckenrangtest och det parametriska Students parade t-test. Varför är inte dessa test tillämpliga på analys av förändring av ordnade kategoridata? Wilcoxon's teckenrangtest baseras på att differenser mellan individernas mätvärden rangordnas, men beräkning av differenser av ordinaldata är inte relevant, eftersom begreppet avstånd inte är definierat. Students t-test ställer stora matematiska krav på datamaterialet, som består av kvantitativa förändringsdata. För kvantitativa data är det vanligt att förändring definieras som skillnaden mellan individens båda värden eller som kvoten mellan dem. Det viktigaste kravet är att datamaterialet, som utgörs av differenser eller logaritmerade kvoter, skall kunna antas tillhöra en population normalfördelade eller approximativt normalfördelade data. Inget av villkoren kan uppfyllas av subjektiva bedömningar som görs på skalor av olika slag [1, 4, 14].

Meningslöst att synda i statistik

Om man trots allt frestas att använda dessa test skall man tänka på att resultaten gäller endast om förutsättningarna för testen är uppfyllda, eftersom testen utnyttjar dessa förutsättningar i sannolikhetsberäkningarna. Testresultaten är giltiga endast under förutsättning att de krav på datamaterialet som analyserna utnyttjar är uppfyllda, därför är det meningslöst att synda i statistik!

*

Potentiella bindningar eller jävsförhållanden: Inga uppgivna.

Referenser

1. Siegel S, Castellan NJ. Nonparametric statistics for the behavioral sciences. 2nd ed. New York: McGraw Hill; 1988.
2. Svensson E. Val och konsekvens: mättnivån avgör den statistiska verktygslådan. Läkartidningen 2005;102:1331-7.

3. Agresti A. Categorical data analysis. New York: John Wiley and Sons; 1990.
4. Altman DG. Practical statistics for medical research. London: Chapman & Hall, 1991.
5. Dunn G, Everitt B. Clinical biostatistics. An introduction to evidence-based medicine. London: Edward Arnold; 1995.
6. Kruskal WH. Ordinal measures of association. Journal of the American Statistical Association 1958;53:814-61.
7. Taube A. Det borde inte vara stjärnor. Läkartidningen 1985;82:2422-4.
8. Taube A. Varför alltid sätta P för signifikanta resultat? Läkartidningen 1990;87:960-2.
9. Swinscow TD, Campbell M. Statistics at square one. 10th ed. London: BMJ Books; 2003.
10. Feinstein A. Beyond statistics: what is really important in medicine? Cleveland Clinic Journal of Medicine 1997;64(3):127-8.
11. Altman DG, Bland JM. Absence of evidence is not evidence of absence. BMJ 1995;311:485.
12. Newcombe RG, Altman DG. Proportions and their differences. In: Altman DG, Machin D, Bryant TN, Gardner MJ, editors. Statistics with confidence. 2nd ed. Bristol: BMJ Books; 2000. p. 45-57.
13. Altman DG. Statistical reviewing for medical journals. Statistics in Medicine 1998;17:2661-74.
14. Svensson E. Guidelines to statistical evaluation of data from rating scales and questionnaires. J Rehabil Med 2001;33:47-8.



=artikeln är referentgranskad

SUMMARY

Assessments on scales and questionnaires generate ordered categorical data. Characteristic feature is the ordered structure with lack of information regarding size and inter-categorical distances. These limited mathematical properties of ordered categorical data mean that the most common statistical methods for evaluation of change, the Student's paired t-test and the Wilcoxon sign rank test, are not appropriate to use as they are based on differences. This paper presents classical methods useful for dichotomous data but they are appropriate for dichotomised ordered categorical data as well.

Elisabeth Svensson

Correspondence: Elisabeth Svensson, Institutionen för ekonomi, statistik och informatik (ESI), Örebro universitet, SE-701 82 Örebro, Sweden
elisabeth.svensson@esi.oru.se

Medicinhistorisk paus

»Det är perioder av hunger som försakar epidemier«

II Rubrikens citat är ett utomordentligt framsynt konstaterande av en framstående arabisk medicinsk författare, Ibn Khaldun. Han föddes 1332 i Tunis, tjänade flera furstar under den tid när muslimerna kontrollerade även Spanien.

I det nuvarande Algeriet drog han sig undan för att skriva sin omfattande världshistoria, där digerdöden upptar en stor del av hans reflektioner. Denna katastrof hade svept över arabvärlden, och Ibn Khaldun egen familj hade drabbats svårt.

Stora truppflyttningar bidrog, enligt Ibn Khaldun, till digerdödens explosiva utveckling, liksom en förtätad befolkning och utdunstningar i luften. Han hade därmed en förbluffande modern syn på smittans förutsättningar.

Han förespråkade tomma och isolerade områden mellan byar och städer för att ge luften fritt spelrum, en tidig form av karantänstänkande. Hårtill kom sedan andra, mer filosofiska, komponenter, som t ex att »pesten kommer som ett straff från ovan när staten börjar förtrycka sina medborgare«.

Ibn Khaldun avled i Kairo 1406.

Stephan Rössner
professor, Karolinska
Universitetssjukhuset Huddinge

Medicinhistorisk paus

»Sömnen är en härlig uppfinning!«

II Den ryske författaren Michail Bulgakov (1891-1940) läste medicin i Kiev, tjänstgjorde som krigskirurg vid ryska fronten under första världskriget och arbetade sedan vid ett landsortssjukhus i Nikolskoje.

Hans mest kända verk är »Mästaren och Margarita«. Denna konstfullt sam-

manvävda berättelse, liksom det mesta av hans litterära produktion, publicerades dock först efter hans död.

Rubrikens citat är hämtat ur novellsamlingen »En ung läkares anteckningar« från 1919, där Bulgakov berättar om sina första kliniska lärospån. Primitiva förhållanden, otillräckliga resurser, dåligt hälsotillstånd i befolkningen gjorde hans liv som ung läkare till ett krävande slit dygnet runt, men med tillfällena till vila som en njutningsfull belöning.

Bulgakov har i en av novellerna bl a beskrivit det morfinberoende han själv drabbades av efter en misslyckad injektion, som gav plågsamma smärtor. Under 1920-talet lämnade han läkaryrket, levde med svårigheter på sitt författarskap, skrev pjäser som Stalin först tillät, men blev sedan en »ickeperson«. Han dog, liksom fadern, i komplikationerna till sina cystnjurar.

Stephan Rössner
professor, Karolinska
Universitetssjukhuset Huddinge