MINITAB

- ett ledande statistikpaket

GÖRAN ANDERSSON

Minitab kan beskrivas som ett bibliotek med datorprogram för tabulering, diagramritning, beräkningar av statistiska mått och utförande av statistiska analyser. Det finns mängder av engelskspråkig statistisk mjukvara av denna typ, men det finns numer även åtskilliga statistikpaket på svenska. *Göran Andersson* skriver här om sina erfarenheter från användning av Minitab.

Vid världskonferensen om statistikundervisning i Kanada 1986 (ICOTS II) tog jag som min huvuduppgift att försöka bevaka vad som hänt när det gäller statistikpaket lämpliga för undervisningsändamål. Mitt intryck blev att Minitab i sådana sammanhang är helt "outstanding" på den engelskspråkiga marknaden. Minitab är installerad i mer än 1 200 stordatorer, spridda över 40 länder. I Sverige finns Minitab vid samtliga universitets- och högskoleinstitutioner för tillämpad statistik.

Dessutom finns det minitabversioner för de flesta persondatorer på den internationella marknaden. Kostnaden för t ex IBM-PC-versionen understiger 10 000 kr, ett pris som kraftigt rabatteras vid samtidigt köp av flera kopior.

Klara — färdiga — kör

Minitab är ett interaktivt, konverserande programpaket. Man sitter vid sin dator/terminal. skriver en instruktion (=kommando) och skickar iväg den, varefter datorn efter några sekunder skriver ett svar. Därefter skickar man iväg nästa kommande och får strax ett nytt svar. Kommandospråket är enkelt och naturligt, vilket gör att det är lätt att komma ihåg och ibland också gissa språkets kommandon (TABLE, HISTOGRAM, PLOT, MEDIAN, AVERAGE etc). Naturligtvis medföljer manual men oftast klarar man sig genom att begära hjälp (HELP) av Minitab. När som helst kan man få upplysning (INFORMA-TION) om vad som är inmatat och eventuella mellanresultat. Endast de fyra första bokstäverna i kommandona behöver anges.

Alla som använt miniräknare vet att framräknade resultat kan sparas i ett eller flera minnen. Minitab har ett förhållandevis stort minnesutrymme till förfogande. Man kan nämligen spara sina beräkningar eller observationer i ett s k arbetsschema (eng worksheet) som omfattar t ex 100 kolumner med upp till 400 tal/observationsvärden i varje kolumn.

	C1	C2	C3	C100
rad 1	*	*	*	 *
rad 2	*	*	*	 *
rad 3	*	*	*	 *
rad 400	*	*	*	 *

Om man minskar antalet kolumner får man plats med fler rader i varje kolumn. Vid en intervjuundersökning av t ex 1 000 personer är det ingen svårighet att mata in observationsvärden för exempelvis 25 frågor. Varje intervjuad person motsvaras av en rad och dennes kodade svar placeras in på denna rad i kolumnerna C1 till C25. (Arbetsschemats utrymme varierar beroende på datorstorlek men 1 000 \times 25 inmatade observationsvärden är aldrig något problem.)

Exempel

Låt oss se hur Minitab fungerar i ett konkret exempel som gäller 77 nyfödda barn vid ett svenskt sjukhus. Vi har data över födelsemånad, kön, vikt (gram) och längd (cm) vid födelsen, graviditetens längd (veckor) och moderns ålder (år). För varje barn är det 6 uppgifter, vilka vi väljer att placera in i kolumnerna C1, C2, ..., C6. Om man ska göra tabeller, diagram och beräkningar är det lättare att överblicka resultaten om man namnger kolumnerna. Minitab konverserar som sagt på engelska, men det hindrar förstås inte att man döper kolumnerna med namn från vilket annat språk som helst, t ex det som tillskrivs äran och oss s k hjältar. Innan man ger ett Minitabkommando, i det här fallet NAME, måste man vänta på att den s k Minitabpromptern, MTB, skrivs ut på terminalen/datorskärmen. Denna klarsignal skriver Minitab själv ut efter det att ett kommando utförts. Namnen får omfatta högst 8 tecken. Så här går namngivningen till:

MTB>	NAME C1	'FÖDMÅNAD'
MTB>	NAME C2	'KÖN'
MTB>	NAME C3	'VIKT'
MTB>	NAME C4	'LÄNGD'
MTB>	NAME C5	'L-GRAVID'
MTB>	NAME C6	'M-ÅLDER'

Namnen får omfatta högst 8 tecken.

Därefter är det dags att läsa in de observerade värdena i kolumnerna C1—C6, vilket sker så här med READ-kommandot (som synes nedan är Minitabpromptern DATA > efter inläsningskommandon).

MTB> READ C1-C6 38 28 DATA > 8601 1 3300 51 3060 42 19 DATA > 8601 1 51 2890 49 39 19 DATA> 8601 1 etc, etc.

Den första inmatningen ovan gäller en födsel i jan 86 av en pojke (koden för kön är här 1 =pojke, 2 = flicka) som vägde 3 300 gram och var 51 cm lång. Graviditeten hade varat i 38 veckor och den lyckliga modern var 28 år.

Tabeller

Innan man börjar analysera materialet är det bäst att kontrollera att data blivit korrekt inlästa. Vi ber därför om en utskrift.

MTB> PRINT C1-C6

Början på utprintningen ser ut så här.

RØ₩	FODMANAD	KÖN	VIKT	LANGD	L-GRAVID	M-ALDER
1	8601	1	3300	51	38	28
2	8601	1	3060	51	42	19
э	8601	1	2870	49	39	19
4	8601	1	3000	50	39	31
5	8601	1	3860	52	40	31
6	8601	1	3780	52	41	26
7	8601	1	3380	50	40	19
8	B601	2	3970	51	40	33
9	8601	2	2880	45	38	32

En första granskning av ett större datamaterial sker ofta med tabeller av olika slag. Låt oss se hur pojkarnas och flickornas födslar fördelar sig efter graviditetens längd genom att låta Minitab framställa följande korstabell.

MTB> TABLE 'L-GRAVID' 'KÖN'

ROWS:	L-GRAVID	COLUM	NS: KON
	1	2	ALL.
34	0	1	1
37	3	2	5
38 39	1 9	6 3	12
40 41	15 7	11 5	26 12
42	7	2	9
ALL	46	31	77
CELL	CONTENTS		
		COUNT	

Det är svårt att spåra eventuella skillnader mellan de båda fördelningarna eftersom det är så många fler pojkar än flickor i det här materialet. Den önskade jämförelsen sker lättare om vi tar fram de procentuella fördelningarna, vilket sker med underkommandot (SUBC) ROWPERCENT.

MTB> TABLE 'KÖN' 'L-GRAVID'; SUBC> ROWPERCENT.

ROWS:	KON	COLUMNS:	L~GRAVIE					
	34	36	37	38	39	40	41	42
1		4.35	6 52	2.17	19.57	32.61	15.22	15.22
2	3 23	3.23	6.45	19.35	7.68	35.48	16, 13	6 45
ALL	1.30	3.90	6.49	9.09	15.58	33.77	15.58	11.69
	43	ALL						
1	4.35	100.00						
2		100.00						
ALL	2.60	100.00						
CELL	CONTEN	TB						
		% OF	RON					

Nu ser man lättare att pojkfödslar är överrepresenterade i födslar efter lång graviditet (42 och 43 veckor). Om man vill, finns det möjlighet att få både procenttal och absoluta tal i samma tabell.

Diagram

Den kritiske läsaren kanske redan nu menar att Minitabs tabeller inte är särskilt eleganta. Jag håller gärna med om det. De här tabellerna lämpar sig sällan för direkt publicering utan bör huvudsakligen användas som arbetstabeller i ett första försök att få grepp om ett större datamaterial. Man lodar sig fram genom att göra korstabeller av olika kombinationer av variabler, prövar olika klassindelningar, ser efter om radprocent, kolumnprocent eller totalprocent ger tydligast information etc. På någon minut har man en uppsjö av tabeller, om man så önskar. För den slutliga redovisningen/rapporten väljer man sedan ut ett relativt fåtal Minitabtabeller som underlag för konstruktion av tabeller av mer konventionell design.

Vad som sagts om tabeller gäller i huvudsak också Minitabs diagramutskrifter. De är till stor hjälp när det gäller att snabbt och bekvämt få en överblick över datamaterialet. Men det är sällan de duger att klippa ut och bara klistra in i en rapport. Den s k dotplotten nedan, som visar flickornas och pojkarnas fördelning efter födelsevikt, är ett av få undantag. (Läsaren må fundera över om det månne är en inbyggd artighetsmekanism som gör att flickorna placerats överst trots att pojkarna var kodade som "l" . . .)



Den här diagramtypen har jag inte sett tidigare i svensk statistiklitteratur. Varje individ motsvaras alltså av en punkt i diagrammet, och nog är väl det här ett bra sätt att snabbt få en uppfattning om både helheten och enskilda detaljer?!

Men å andra sidan: Är det stora och/eller många datamaterial kan det bli för mycket av detaljer och då är nog lådagrammen (eng boxplot) att föredra. Så här ser de ut i Minitabversion.



Hoppsan! Nu har pojkarna placerats överst. Minitab kanske har någon slags jämlikhetsmekanism i stället, ungefär som "Varannan dans damernas"!

Dåliga skämt åsido; lådagrammen visar tydligt att både medianen och de två kvartilvärdena är högre för pojkarnas vikter än för flickornas. Vill man ha numeriska mått på pojkarnas och flickornas viktfördelningar är det lätt ordnat. Med hjälp av kommandot CHOOSE plockar man ut pojkarnas vikter till en särskild kolumn, som man kallar t ex C13. Analogt placeras flickornas födelsevikter i exempelvis kolumn C23. Sedan är det dags att döpa dessa kolumner och kommendera DESCRIBE:

```
MTB> NAME C13 'POJKVIKT'

MTB> NAME C14 'TÖSVIKT'

('FLICKVIKT' innehåller 9 tecken och accep-

teras inte)

MTB> DESCRIBE 'POJKVIKT' 'TÖS-

VIKT'

N MEAN MEDIAN TRNEAN STDEV SEMEAN

N MEAN MEDIAN TRNEAN STDEV SEMEAN
```

40	1000. /	au/ 3. 0	0007.0	5/4. /	
31	3180. 3	3250. 0	3210.4	542.7	97.
MIN	MAX	Q 1	03		
1980.0	4770.0	3070.0	3675.0		
1300.0	4170.0	2680. 0	3460.0		
	31 MIN 1980.0 1300.0	31 3180.3 MIN MAX 1980.0 4770.0 1300.0 4170.0	31 3180, 3 3250, 0 31 3180, 3 3250, 0 MIN MAX Q1 1980, 0 4770, 0 3090, 0 1300, 0 4170, 0 2880, 0	40 3350.7 3350.7 3350.7 3370.3 31 3180.3 3250.4 3310.3 3270.4 MIN MAX 0.1 0.3 1980.0 4770.0 3090.0 3695.0 3460.0 1300.0 4170.0 2880.0 3460.0 3460.0	HIN MAX G1 G3 G

Ja, nog blev de här januaribebisarnas vikter omsorgsfullt beskrivna. Två av måtten bör kanske förklaras. TRMEAN är ett s k trimmat medelvärde. De 5 % högsta och de 5 % lägsta värdena utesluts vid beräkningen av TRMEAN, så t ex är inte den lilla flickan på 1 300 gram medräknad och inte heller bamsepojken på 4 770 gram. SEMEAN är standardavvikelsen för medelvärdet, dvs den ''vanliga'' standardavvikelsen dividerad med \sqrt{N} . Exempelvis gäller för flickornas vikter att 97,5 = 542,7/ $\sqrt{31}$. (1,96×SEME-AN är den s k felmarginalen. Lägger man denna felmarginal på ömse sidor om medelvärdet 3 180 erhålls ett intervall som med ca 95 % konfidens/ säkerhet täcker populationsmedelvärdet.) Låt oss för en stund intressera oss för de nyblivna mammorna. Här är ett stam-bladdiagram över deras ålder vid nedkomsten.

```
MTB > STEM-AND-LEAF 'M-ALDER'
```

```
Stem-and-leaf of M-ALDER N = 77
Leaf Unit = 1.0
```

4	1	77 77
8	2	1111
18	2	2223333333
24	2	444555
(15)	2	6666667777777777
38	2	888889999999
27	Э	00001111111111
13	з	22333
8	з	455
5	з	677
5	з	
2	4	0
1	4	
1	4	4

Mittenspalten anger tiotalsiffran och siffrorna på raderna till höger anger hur många år mamman har hunnit på respektive decennium. Överst har vi sålunda fyra 19-åriga mödrar och längst ner finner man att den äldsta är 44 år.

Tillbaka till de nyfödda flickorna. Låt oss undersöka hur pass starkt sambandet är mellan flickornas födelsevikter och graviditetens längd (i veckor). Viss information får man genom att "plotta" vikterna mot antal graviditetsveckor.

MTB>PLOT 'TÖSVIKT' 'TÖSLGRAV'



Upplösningen är inte bättre än att en del observationer sammanfaller i diagrammet. Siffrorna 2 och 3 i diagrammet anger att två respektive tre observationer sammanfaller i denna "punkt". Eftersom varken x-axeln eller y-axeln har värdet 0 i skärningspunkten är det svårt att avgöra hur pass starkt sambandet är. Vi ber om korrelationskoefficienten — och får genast svar:



Koefficienten 0,564 anger — inte helt överraskande — att det råder ett ganska starkt samband mellan graviditetens längd och födelsevikten. Hur mycket betyder då en extra graviditetsvecka för födelsevikten? Ett ungefärligt svar kan man få ur ekvationen för den s k regressionslinjen (dvs den räta linje som — enligt vissa kriterier — "bäst" beskriver det linjära sambandet mellan de två variablerna).

MTB > REGR 'TOSVIKT' 1 'TOSLGRAV'

The regression equation is TOSVIKT = - 3604 + 173 TOSLGRAV

Tolkningen av koefficienten 173 är — för de här 31 flickfödslarna — att en extra graviditetsvecka i genomsnitt innebar en ca 170 gram tyngre flicka.

De exempel på Minitabanvändning jag visat har handlat om teoretiskt tämligen enkla tillämpningar. Tilläggas bör att Minitab dessutom har rutiner för simuleringar av olika slag och för något mer avancerade metoder som multipel regressionsanalys, variansanalys, hypotesprövning – även s k icke-parametriska text, tidsserieanalys m m. Vidare är det möjligt att bygga upp egna små program som kan sparas och senare användas vid behov.

En kommentar om didaktiken

På en del stadier blir resultaten genom ett Minitabkommando kanske lite väl lättillgängliga. Det är svårare att förstå innebörden av alla de mått som ramlar ut efter t ex DESCRIBE-kommandot, om man inte själv har räknat fram dem manuellt någon gång. Vid många observationer är dock detta väldigt mödosamt när man kommer in på exempelvis beräkning av korrelationskoefficienten.

I sådana fall har man genom Minitab också möjligheten att gå en medelväg. Genom kommandon som ADD, SUBT, MULT, DIVI, SUM etc kan eleverna stegvis beräkna de olika komponenterna i ett komplicerat mått. Låt mig visa hur beräkningen av medelvärde och standardavvikelse går till med huvudsaklig användning av LETkommandot.

Exempel

En parkeringsvakt i Lund noterade under 10 dagar antalet inrapporterade parkeringsförseelser. Han erhöll följande sammanställning:

Dag	Antal	Dag	Antal
1	104	6	102
2	106	7	104
3	105	8	100
4	97	9	96
5	100	10	103

Beräkna medelvärdet och standardavvikelsen för antalet inrapporterade parkeringsförseelser. Använd LET-kommandot. Lösning: SET C1 104 106 105 97 100 102 104 100 96 103 PRIN C1 (för att kontrollera att inläsningen är korrekt) LET K1 = SUM(C1)/10 LET K2 = SUM(C1-K1)**2)/9 LET K3 = K2**0.5 PRIN K1 K3

Minitab eller svenskt statistikpaket?

Minitab är trevligt och enkelt att arbeta med. Intrycket från universitetsutbildningen på elementär nivå är att intresset för statistiken stimuleras väsentligt när eleverna får denna möjlighet att pröva sig fram i stora material av data från verkligheten. Huruvida Minitab är att rekommendera också för svenska gymnasieskolor är dock svårt att svara på, jag har t ex inte sett särskilt mycket av de nyaste alternativen på svenska.

Hur som helst vågar jag rekommendera de lärare som har nära till en högskoleinstitution för tillämpad statistik att där försöka få provköra Minitab (varför inte höra sig för om en studiedag?). Efter att ha bekantat sig med detta — som jag uppfattar det — ledande statistikpaket för undervisningsändamål är det nog i alla fall lättare för gymnasieläraren att ställa upp kriterier inför valet mellan de svenska motsvarigheterna.

Litteratur. Jonsson, B: HELP MINITABHELP — Ett kompendium i grundläggande minitabanvändning. Statistiska inst, Uppsala universitet. Uppsala 1984.

PS. Den som vill ha mer information om priser, prestanda m m för Minitabversionen till ett speciellt PC-märke vänder sig lämpligen direkt till MINITAB, INC 3081 Enterprise Drive State Collage, PA 16801 USA. Tel: 814-238-3280