



Beskrivande statistik och presentation

Stora datamängder måste sammanfattas och reduceras för att få en läsbar form. Denna text behandlar olika metoder för att redovisa, dvs. sammanfatta och presentera numeriska data. Kortfattat beskrivs centralmåttan median och medelvärde samt spridningsmått. Tips ges för presentation i text, tabeller, diagram och kartor.

Alla dessa metoder används allmänt när man vill presentera insamlade datamaterial på ett begripligt sätt för tänkta läsare. Vad man dessutom ofta märker när man arbetar med sammanställning av data för presentation, är att man hittar nya vägar att analysera sina data och även oväntade anomalier eller felaktigheter i materialet. De metoder för presentation av statistiken som beskrivs nedan, kan och bör användas även som arbetsredskap när man i rent analyssyfte penetrerar sina data. Skillnaden är att man vid presentation i skrift även måste tänka på läsbarhet och mer "estetiska värden".

Överblick och en första analys

I all statistisk redovisning är det lämpligt att göra en snabb analys av sina data. Det kan vara till en god hjälp att ha en redovisningsplan, "tabellplan", att utgå ifrån redan då data börjar droppa in, dvs. innan alla uppgifter finns registrerade.

Redovisningsplanen som görs vid planeringen av undersökningen ska visa t.ex. vilka sammanfattande värden som man anser är särskilt viktiga samt vilka variabler som man antar kan bli intressanta att redovisa grupperat, ihop med andra variabler, ihop med bakgrundsinformation som geografiska data eller dylikt. Naturligtvis kan denna tidiga plan även omfatta olika diagram eller kartor. När man ställs inför en ny, obekant och stor datamängd kan det i första början kännas svårt att veta hur man ska ta sig an detta material. Det är då man kan märka att en enkel analys i form av sorteringar av datamaterialet exempelvis i grupper och i storleksordning, samt enkla sambandsanalyser i punktdiagram och korstabeller

etc, kan ge förvånansvärt mycket. Även beräkningar av centralmått, spridningar och samband är enkla att få fram.

Mängder av siffror bör sorteras

Det är mycket lämpligt att på ett tidigt stadium i en pågående undersökning göra klart för sig i vilken form resultaten ska presenteras. Den analys man gör av insamlade data kan göras mer eller mindre "läsvänlig".

Till vissa intressenter kan stora tabeller med fördel överlämnas som råtabeller eller punktdiagram etc, exempelvis på diskett, CD-rom eller som epostbilaga. Det man ämnar presentera på papper kräver däremot genomarbetning. Man får resonera sig fram till om uppgifterna ska ges i löpande text, med enstaka siffror, i tabeller eller i form av diagram, kartor eller bild.

Det är dessutom viktigt att fel och osäkerheter i undersökningen beskrivs och diskuteras, t.ex. i form av en kvalitetsdeklaration.

En checklista

Urvalet av vilka uppgifter som ska presenteras beror på målgrupp. Tänk alltså på läsaren när du planerar din redovisning! Nedanstående tips kan kännas självklara men kan fungera som checklista:

- **Viktigt i sig - siffror.** Vissa data är intressanta att presentera i sig, det kan vara viktiga resultat eller siffror som kan fungera som intresseväckare. Ofta är det medelvärden eller totaler som bör få denna roll.
- **Störst - minst.** Att peka ut de högsta och lägsta värdena brukar vara ett bra tips om man vill väcka intresse för sina data.
- **Tio i topp/botten listor** är en väletablerad och framgångsrik variant av ovanstående.

- **Jämförelser med gränsvärden eller normvärden.** Om mätvärden eller undersökningsresultat ska jämföras med några gränsvärden är det självklart intressant att i diagrammet eller tabellen märka ut de värden som överskrider dessa gränsvärden.
- **Jämförelser i tiden.** Om man har återkommande undersökningar att redovisa är det naturligt att ställa samman data i en tidsserie. Det kan ske i tabeller eller i tidsseriediagram. Om trender kan utläsas ur tidsserierna kan de ritas in i diagrammet och kommenteras.
- **Jämförelser i rummet.** Kartor och kartogram är det naturligaste sättet att visa regionala skillnader.

Centralmått och spridningsmått

Dessa mått används ofta i presentationer. Det är viktigt att alltid ange vilken typ av centralmått och spridning som visas med lådagran eller ”lådor med antenner” eftersom det inte finns någon standard för dessa diagramtyper.

Aritmetiska medelvärden och standardavvikelse

Det centralmått (kallas även lägesmått) som oftast används och rekommenderas är det vanliga medelvärdet, som egentligen bör kallas aritmetiskt medelvärde. Det byggs upp av summan av de observerade värdena dividerat med antalet observationer. Om man har medelvärden för flera grupper så kan dessa jämföras, slås samman etc.

Aritmetiskt medelvärde är ett praktiskt mått genom att det enkelt kan bindas till spridningsmåttn standardavvikelse (ungefärligen genomsnittliga avvikelser från medelvärdet).

Medelvärden kan naturligtvis även användas vid beräkning av totaler, som ibland är lämpliga redovisningsmått. Det gäller speciellt för urvalsundersökningar om förhållandet mellan populationens och urvalets storlek är känt. Ett exempel är den totala belastningen på en recipient om man har medelhalten av en förorening i avloppsvatten. Denna medelhalt multiplicerat med ett (ev. skattat) totalflöde under en period ger den totala mängden av föroreningen under den givna perioden.

Median och kvartiler

Ett problem med det aritmetiska medelvärdet (μ) är att det ger stor tyngd åt extremvärden. Därför har median fått utbredd användning i exempelvis vattenkvalitetsredovisningar. Median innebär enkelt uttryckt att hälften av mätvärdena ligger under detta värde och hälften över. Det är lättförståeligt och ofta mer robust än μ . Det finns olika metoder att visa spridningen förknippad till median:

- **Variationsvidd**, differensen mellan högsta och lägsta värdet i det statistiska materialet
- **Kvartilavståndet**, differensen mellan tredje kvartilen (fjärdedelar av datamaterialet) och första kvartilen
- **Kvartilavvikelse**, kvartilavståndet dividerat med 2

Dessa mått kan grafiskt illustreras i lådagran

Trimmade medelvärden

Ett problem med alla dessa mått som bygger på kvartiler (fjärdedelar av datamaterialet) liksom med percentiler (100-delar) är att de är svåra att vidarebearbeta med klassiska statistiska metoder. Av denna orsak kan man ibland rekommendera medelvärden även för material där enstaka extrema värden förekommer. För att få robustare mått bör man i dessa fall överväga att göra en rensning av dessa extrema värden, s.k. uteliggare. Denna rensning dvs. att man tar bort vissa värden måste dock vara väl motiverad.

En användbar metod kan vara att rensa bort en godtyckligt vald andel av låga och höga värden i materialet (s.k. uteliggare) och beräkna medelvärdet för resten. Man bör alltid vid rensning av värden noga avväga om inte de extrema värdena är tecken på något allvarligt i materialet. De kan ju vara **alarmerande händelser** lika väl som felmätningar eller tryckfel. Om de är sanna mätvärden bör de självfallet inte rensas bort ur materialet!

Symmetriska fördelningar och skeva

När det statistiska materialet fördelar sig ungefärligen symmetriskt kring medelvärdet, är avståndet mellan medelvärdet och medianen litet. Om fördelningen liknar en klockformig kurva är den approximativt normalfördelad. Många av de meto-

Ett och samma material kan presenteras på olika sätt

Tabellen visar markanvändningen i hektar per län fördelat på 7 kategorier.

I stapeldiagrammet visas den procentuella fördelningen på markanvändning. Där kan man lätt se var i landet de mest jordbruks- och skogsbruksintensiva länen ligger. Likaså kan man lätt se att Malmöhus län och Gotland har förhållandevis lite inlandsvatten. Observera att länen står i syd-nordlig ordning nerifrån och upp, och inte i den konventionella ordningen med Stockholms län överst och Norrbotten längst ner.

*Även i figuren redovisas markanvändningskategorin Jordbruksmark länsvis. Denna diagramtyp är ytterligare ett exempel på **mellanting mellan karta och stapeldiagram**. Också här är länen i möjligaste mån uppställda från norr till söder. Höjden på respektive länsstapel är proportionell mot länets totala areal. Längden på stapeln visar den proportionella andelen av jordbruksarealen inom länet. Här är alltså längden på staplarna jämförbar med den vänstra, mörka delen av staplarna i stapeldiagrammet.*

Tabell 5 Markanvändningen omkring 1995. Ha

Län	Jordbruksmark	Skogsmark	Bebyggd mark och tillhörande mark ¹	Täcker ²	Myr	Berg, fjäll och övrig mark	Vatten	Totalt
Stockholms	116 074	321 500	93 500	1 100	11 000	105 830	29 540	678 540
Uppsala	179 943	401 500	34 900	250	17 000	65 320	19 380	718 290
Södermanlands	157 733	323 500	33 250	600	17 000	74 160	51 890	658 130
Östergötlands	265 635	601 500	45 950	550	22 000	120 570	106 680	1 162 880
Jönköpings	153 883	752 000	41 550	750	73 000	26 310	77 660	1 125 150
Kronobergs	91 391	649 500	29 700	650	55 000	19 550	96 560	942 350
Kalmar	187 885	734 500	41 475	900	25 000	127 330	49 470	1 166 560
Gotlands	99 203	123 000	14 000	550	10 000	67 260	3 240	317 250
Blekinge	52 322	194 500	19 950	300	3 000	24 060	10 920	305 050
Skåne	560 698	382 500	98 650	1 850	22 000	37 010	31 900	1 134 610
Hallands	143 284	286 500	35 400	650	32 000	47 600	25 660	571 090
Västra Götalands	607 056	1 300 500	157 600	2 300	108 000	218 700	140 160	2 534 320
Värmlands	150 634	1 296 000	47 750	550	152 000	111 670	178 470	1 937 070
Örebro	128 361	601 000	36 400	600	57 000	28 310	78 440	930 110
Västmanlands	140 243	397 500	29 650	550	33 000	29 230	30 160	660 330
Dalarnas	96 191	2 001 000	54 350	550	422 000	245 230	220 280	3 039 600
Gävleborgs	92 667	1 492 500	46 050	800	156 000	31 150	153 010	1 972 180
Västernorrlands	89 104	1 723 000	42 750	750	202 000	110 230	145 590	2 313 410
Jämtlands	72 540	2 645 500	35 300	450	851 000	1 340 550	475 360	5 419 700
Västerbottens	111 670	3 131 500	48 450	1 100	913 000	1 334 400	379 230	5 919 350
Norrbottens	66 816	4 044 000	49 650	800	1 764 000	3 965 800	697 520	10 588 590
Ej länsfördelat			85 000	15–20 000		37 000 ⁴		
De fyra största sjöarna ³							901 810	901 810
Hela riket	3 563 333	23 423 500	1 121 900	ca 35 000	4 945 000	8 004 690	3 902 930	44 996 350

1) Golfbanor, slalombackar och motionsspår är ej medräknade i bebyggd mark

2) Torvtäcker ej länsfördelade

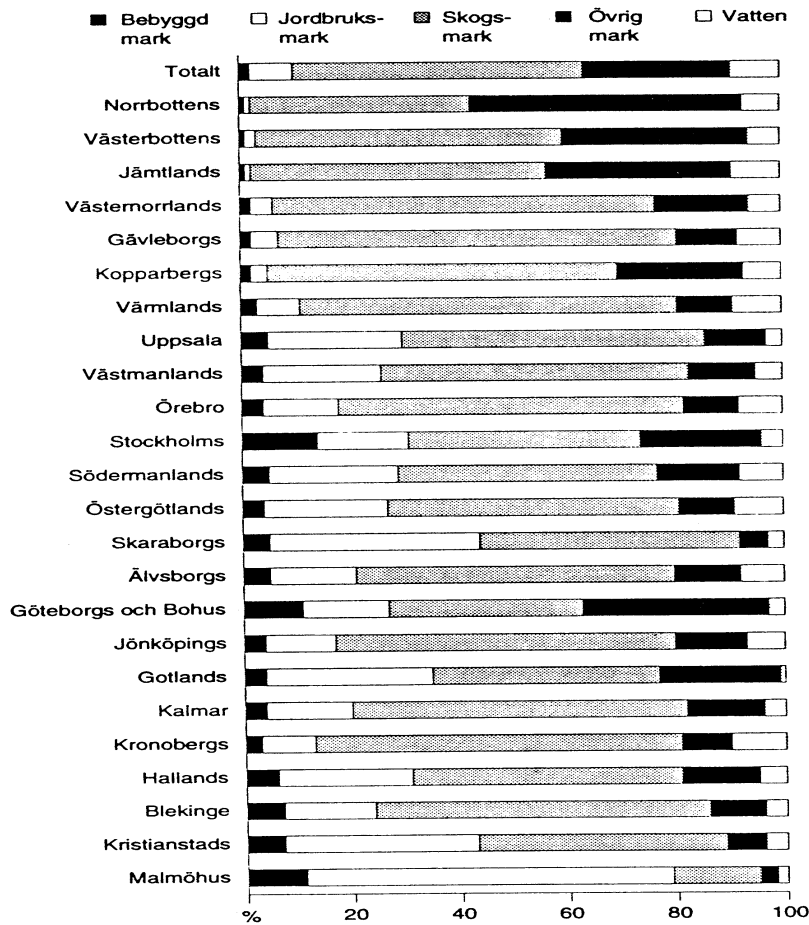
3) Vänern, Vättern, Mälaren och Hjälmaren

4) Golfbanor, slalombackar och motionsspår

Exempel 4.1a

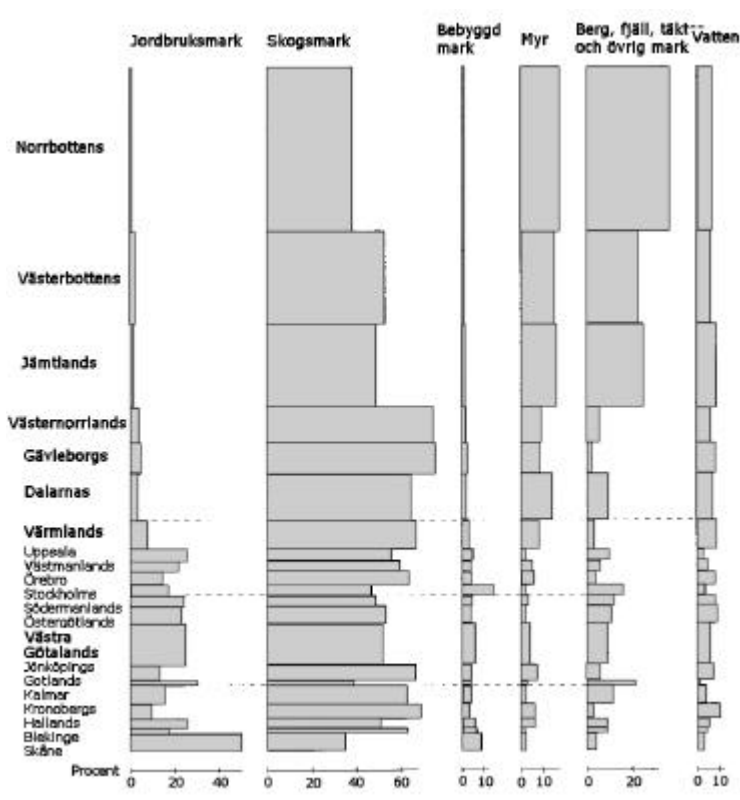
Från Markanvändningen i Sverige -tredje utg. tab 5 -länsstabell

Markanvändningen länsvis 1990, %.



Exempel 4.1b

Prop 1993/94:111 bil 2 sid 182.



Exempel 4.1c

Markanvändning omkring 1990. Procent av total areal. Höjden på respektive länsstapel är proportionell mot länets totala areal

der och formler som används vid statistisk analys bygger på antaganden om normalfördelning. I verkligheten förekommer det ofta att fördelningen är skev åt ettdera hållet. Transformation av data är ett sätt att komma tillrätta med problemet. Se vidare *analysmetoder och modeller*.

Tabeller

Dagens standardprogram för tabellframställning gör det enkelt att konstruera tabeller. För den som är intresserad av att grundligare studera tabelleringsprinciper finns mycket att hämta i läroböcker i deskriptiv statistik.

Sortering av material till tabeller

När man har ett stort datamaterial, exempelvis i form av siffror i datafiler måste man ordna dessa för att få en överblick av materialet. När väl materialet har granskats och kompletterats och upptäckta felaktigheter rättats till kan man för en första överblick göra en sammanställning av materialet i några olika tabeller. Det vanligaste är att man börjar med att sortera data i frekvenstabeller.

Har man kvantitativa data kan det vara lämpligt att först göra en sortering i storleksordning. På så sätt är det lätt att få en överblick av variabiliteten i materialet.

Om variabeln är kontinuerlig och antar många variabelvärden kan man konstruera klasser eller kategorier som är lämpliga för presentation av materialet. Man sorterar in data i dessa klasser och får på så sätt en redovisning över hur många observationer som ingår i varje kategori.

Oftast har man flera variabler i sitt material. Med hjälp av vanliga kalkylprogram för datatabeller kan man studera varje variabel för sig enligt det system som beskrivits ovan. Man kan också beräkna procentvärden, medelvärden etc. Ofta är man dock intresserad av att undersöka hur vissa variabler samvarierar. Om man studerar två variabler i taget och dessa variabler är dikotoma eller diskreta (d.v.s. kan anta två eller flera alternativa värden) går det att sortera in variabelparen i enkla korstabeller.

Redigering av tabeller

För presentation i publikationer bör man noga välja sina tabeller. Stora tabeller är svårlästa. Det gäller också flerdimensionella tabeller, d.v.s. då många variabler redovisas samtidigt exempelvis i underavdelningar. Man bör förenkla sina tabeller, ibland kan det vara bättre med två mindre tabeller än med en stor. Ofta kan en presentation i form av diagram vara en möjlighet.

Några tumregler

Tänk på läsaren när du redigerar dina tabeller. Tabeller ska ha:

- Rubrik
- Klara definitioner
- Lättförståelig ordning vad gäller kolumner och rader
- Angiven ledd för procentangivelser

Rubriken bör innehålla allt som behövs för att läsaren ska veta vilken information som finns i tabellen, och samtidigt vara mycket kort. Detta betyder att viss information om definitioner och dylikt lämpligen bör flyttas till en ingress, fotnot eller till faktaruta i närheten av tabellen. De enheter som används ska framgå antingen ur rubriken eller anges i tabellhuvudet eller i marginalen.

Avrundning

Kontrollera noggrannheten i siffror. Om man exempelvis redovisar cellvärden i form av tal med många decimaler, kan det ge en falsk övertro på säkerheten i tabellens värden. Dessutom är det ofta lättare att läsa en tabell med färre siffror. Det kan ofta vara bra att avrunda uppgifterna i tabellerna till två eller tre siffrors noggrannhet. Det innebär t.ex. att 27 486 kg avrundas till 27 000 eller 27 500 kg. Eller till 27 eller 27,5 ton. Om man i samma tabell även har förhållandevis små värden, kan även dessa avrundas till två eller tre siffror, exempelvis 3,567 kg till 3,6 eller 3,57 kg. Konventionella avrundningsprinciper, t.ex. att ange alla värden i ton (ev. med en decimal) hade i detta fall betytt att 3,567 kg blev 0 (noll). Det beror på omständigheterna vilken avrundningsmetod som är lämpligast.

Avrundningen ska naturligtvis inte ske för tidigt i tabellbearbetningarna. Det är lämpligt att göra det allra sist i redigeringskedet. Annars kan värdefull information försvinna inför analysarbetet.

Klassindelning

Om de variabler man ska sammanfatta i en tabell är kontinuerliga, måste materialet delas in i klasser. I detta läge måste man tänka på att:

- ha lagom många klasser
- klassgränserna bör kunna upplevas som naturliga
- i möjligaste mån använda "standardiserade" klassindelningar, för att underlätta jämförelser med andra dataserier

Även om det inte finns allmänt vedertagna standarder bör man försöka använda indelningar som går att jämkna ihop med andra publicerade serier.

Ibland kan det vara lämpligt att föra ihop långa tidsseriedata till flerårs-perioder. Väderleksstatistik publiceras av SMHI vanligen uppdelat på 12 tim-

mar, dygn, månad och år. I vissa sammanhang passar kvartal eller halvår bäst, i vissa växtsäsonger. Geografiska data kan förutom de administrativa regionerna och avrinningsområdena indelas i vegetationszoner. Kan man välja sin publiceringsform, bör man tänka på läsarens referensram och välja en för dem välkänd klassindelning.

Det kan påpekas att ofta anges klassgränserna så att "jämna" värden ges till de nedre klassgränserna och således finns i vänsterkolumnen. Normer på detta område finns inte, det är mest en smaksak hur man gör.

Procenttal i tabeller

Det underlättar för läsaren om man anger på vilken ledd en tabells procentvärden är beräknade. Lämpligen visar man det med att ange en 100%-kolumn i högra marginalen resp. med en 100%-rad nederst i tabellen. I tabellen högst upp ges virkesförrådet i skogskubikmetrar fördelat på trädslag och

Exempel 4.2

Miljöklasser (tillståndsbedömning) för vattendrag och sjöar enl. Bedömningsgrunder för miljöklasser, sjöar och vattendrag (NV rapport 4913).

Kvalitetsklass	Tot-P $\mu\text{g/l}$ maj-okt	Tot-N $\mu\text{g/l}$ maj-okt	TOC mg/l
1	<12	<300	<4
2	12,5 - 25	300 - 625	4 - 8
3	25 - 50	625 - 1250	8 - 12
4	50 - 100	1250 - 5000	12 - 16
5	>100	>5000	>15

Kvalitetsklass	Tot-P $\mu\text{g/l}$	Tot-N $\mu\text{g/l}$	COD mg/l
1	-7,5	-300,0	-5,0
2	7,6-15,0	300,1-450,0	5,1-10,0
3	15,1-25,0	450,1-750,0	10,1-15,0
4	25,1-50,0	750,1-1500,0	15,1-20,0
5	50,1-	1500,1-	20,1-

Det bör klart framgå, både för den som sammanställer tabellen och för läsaren, i vilken klass varje tänkbart värde hamnar. Om klassgränserna anges som i ovanstående exempel är det inte helt uppenbart att P-värdet 15,0 hamnar i klass 2, medan 15,1 hamnar i klass 3. Klassindelningen blir tydligare om klassgränserna för klass 1 anges med \leq och klass 5 med $>$ (det är så de ges i Naturvårdsverkets Allmänna råd 90:4). Ett annat sätt är att ange klassgränserna i de tre kolumnerna med en decimal överlag .

Exempel 4.3

Virkesförrådet fördelat på trädslag 1989-1993

Trädslag	Landsdel, milj. m ³ sk				
	N Norrland	S Norrland	Svealand	Götaland	Hela landet
Tall	285	264	304	247	1 101
Gran	175	357	306	416	1 254
Björk	80	74	60	75	289
Asp	4,2	6,9	13,2	11,7	36
Ek			3,6	23,9	28
Bok				17,3	17
Övr. lövträd	4,1	10,5	16,3	29,2	60
Torra+vindf.	14,2	16,4	10,9	9,4	51
Alla	562	730	715	830	2 836

Trädslag	Landsdel, %				
	N Norrland	S Norrland	Svealand	Götaland	Hela landet
Tall	26	24	28	22	100
Gran	14	28	24	33	100
Björk	28	26	21	26	100
Asp	12	19	37	33	100
Ek			13	87	100
Bok				100	100
Övr. lövträd	7	17	27	49	100
Torra+vindf.	28	32	21	18	100
Alla	20	26	25	29	100

Trädslag	Landsdel, %				
	N Norrland	S Norrland	Svealand	Götaland	Hela landet
Tall	51	36	43	30	39
Gran	31	49	43	50	44
Björk	14	10	8	9	10
Asp	1	1	2	1	1
Ek	0	0	1	3	1
Bok	0	0	0	2	1
Övr. lövträd	1	1	2	4	2
Torra+vindf.	3	2	2	1	2
Alla	100	100	100	100	100

Trädslag	Landsdel, %				
	N Norrland	S Norrland	Svealand	Götaland	Hela landet
Tall	10,05	9,31	10,72	8,71	38,82
Gran	6,17	12,59	10,79	14,67	44,22
Björk	2,82	2,61	2,12	2,64	10,19
Asp	0,15	0,24	0,47	0,41	1,27
Ek			0,13	0,84	0,97
Bok				0,61	0,61
Övr. lövträd	0,14	0,37	0,57	1,03	2,12
Torra+vindf.	0,50	0,58	0,38	0,33	1,79
Alla	19,82	25,74	25,21	29,27	100,00

Källa: Skogsdata 94, SLU

landsdel, den andra tabellen visar att ca 28% av tallförråden växer i Svealand, medan man ur den tredje tabellen kan läsa att virkesförrådet i Svealands skogar består till nästan 43% av tall.

Ibland kan det vara meningsfullt att ge cellvärdena i procent av totalen. I den nedersta tabellen kan man se att Svealands tallar står för 11 % av det totala virkesförrådet i Sverige. En tumregel är att inte kombinera flera procentfördelningar i samma tabell. Om man sätter in värdena från en av de visade procenttabellerna i grundtabellen kan tabellen göras tillräckligt överskådlig exempelvis om man anger procentvärdena i kursiv stil.

Diagram

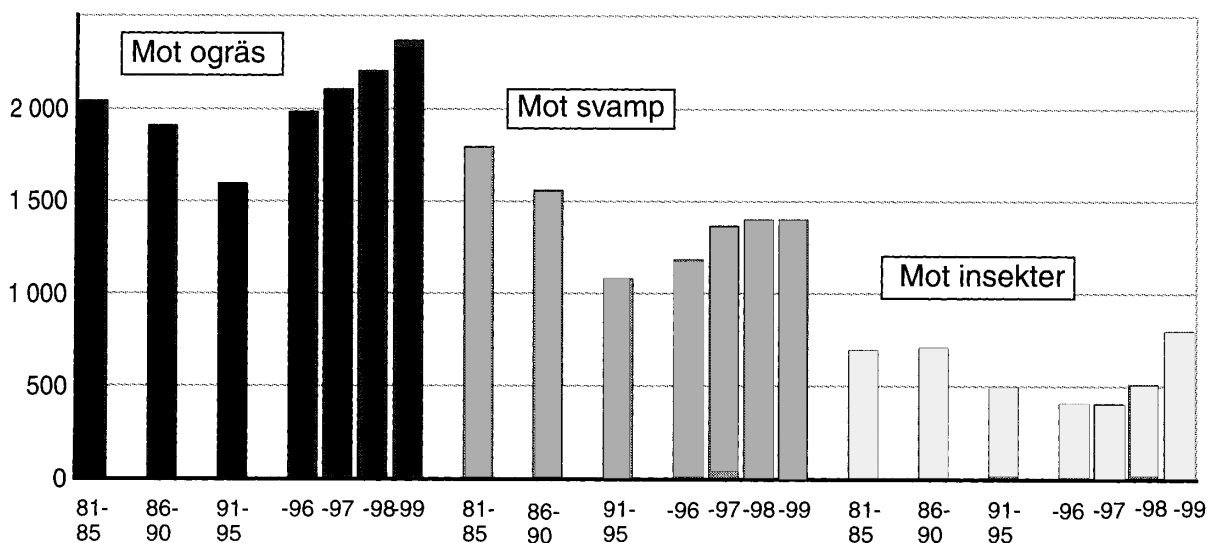
Att använda diagram kan vara ett effektivt sätt att lyfta fram det väsentliga ur stora datamaterial, att illustrera komplicerade samband och skapa överskådlighet. Bra diagram:

- **drar till sig uppmärksamheten och lockar därmed till studium.** Det förpliktigar till att använda diagram med förnuft. Man bör inte ägna sig åt att med diagram visa självklarheter eller förledas att utnyttja alltför osäkra data. Diagrammen bör vara välmotiverade. De framhäver betydelsefulla resultat.

Hektardoser av försålda bekämpningsmedel till jordbruket

Number of doses of pesticides in agriculture

Antal doser i tusental



Källa: Mi 31 SM 9901, Keml och SCB

Exempel 4.4

Hektardoser av försålda bekämpningsmedel till jordbruket. Ur *Naturmiljön i siffror (2000)*, sid 63.

- **är enkla och lättbegripliga.** Diagram är förträffliga när det gäller att åskådliggöra tendenser, samband, mönster, likheter och olikheter. Exempel på detta är när man vill visa om någonting har ökat eller minskat, om det finns skillnader mellan olika grupper, vilken grupp som är störst eller minst. Däremot kan det vara bättre att använda tabeller om man vill visa exakta numeriska värden. Exempelvis är siffror i en tabell överlägsna för att visa hur stor en halt eller mängd är.
- **kan vara lättare att förstå och minnas än siffror.** Vid konstruktion av diagram, och naturligtvis även av kartor, bör man eftersträva att hjälpa läsaren att avslöja fakta och samband som är svåråtkomliga i tabeller. Ett diagram kan vara en bra introduktion till materialet och locka till vidare studium även i mycket komplicerade system av tabellerade data.
- **kan vara utrymmesbesparande.** Välmotiverade och välritade diagram och kartor kan ofta ersätta långa förklarande texter. Ögat kan uppfatta mycket mer information i form av t.ex. en kurva, än motsvarande datamängd i siffror och text.

Några diagramtyper

Det finns många olika typer av diagram. Här ges en mycket kortfattad beskrivning av de vanligaste.

Stapeldiagram

Används vid diskret eller kvalitativ variabel för att visa antal, andelar eller andra relationstal för olika värden på variabeln. Med diskret variabel menar man att värdet kan anta endast vissa värden, exempelvis hela tal. Kvalitativa variabler kan vara kategorier, exempelvis grupper. Höjden på stapeln anger ofta en frekvens (absolut eller relativ), men kan också vara exempelvis en summa eller ett medelvärde. Dessa diagram kallas även för **stolpdiaqram** (engelska *bar charts*).

Histogram

När den variabel som ska illustreras är kontinuerlig och klassindelad kan histogram används i samma slags situationer som stapeldiagrammet, d.v.s. för att visa antal och andelar. Histogram liksom stapeldiagram ger en ofta lättillgänglig och bra visuell bild av spridningen i materialet.

- **Grundregel, ha lika breda klasser.** Eftersom variabelvärdena klassindelas får staplarna en bredd och *höjden gånger bredden (= ytan)* av de staplar som sålunda skapas får representera frekvensen per klass. Man bör göra samtliga klasser lika breda så att även staplarnas bredder blir lika. I så fall blir höjden proportionell mot frekvensen även i histogrammet såsom i stapeldiagrammet. Ibland kan det vara befogat med olikbredda klasser. Det kan bero på att materialet från början är indelat i olikbredda intervall, exempelvis definierade kvalitetsklasser. Eller att man saknar detaljinformation i vissa delar av materialet. Ett specialfall är då man har årsdata för vissa år, men enbart data för flerårsintervall för vissa tidsperioder. I dessa fall är det viktigt att påpeka att stapelhöjden inte anger frekvensen utan ger frekvensen/enheten på x-axeln.
- **Öppna klasser** bör man undvika i diagram, men ibland går det inte. I vissa undersökningar vet man t.ex. bara att delar av datavärdena ligger under en mätteknisk detektionsgräns. Någon riktigt bra metod att visa detta grafiskt finns inte, man får anpassa sig till de möjligheter som ritprogrammet ger, och göra en anmärkning i kolumnen.
- **Antal klasser.** Det kan vara svårt att veta hur många klasser man ska dela sitt datamaterial i.

Det betyder att vid $n=100$ ska man ha sju klasser och vid $n=1000$ tio klasser. Hur klasserna konstruerats får ibland stor betydelse för histogrammets utseende.

Grupperade och ackumulerade stapeldiagram

Jämförelser av några (få) variabler kan lämpligen visas med grupperade eller ackumulerade staplar. Skillnaden är att man i det första fallet ställer kategoriernas staplar bredvid varandra, i det andra ställs staplarna på varandra. I det övre diagrammet kan man lätt jämföra de sammanlagda frekvenserna.

Stapeldiagram och histogram kan även vridas 90° , och presenteras liggande. Det ger utrymme för förklarande text och är speciellt en fördel om man har långa texter eller många staplar. Ett exempel är diagrammet i exempel 4.1, som visar markanvändningen per län.

Ordningsföljden mellan staplarna

En väl vald ordningsföljd mellan kategorierna gör diagrammet bättre! Om man har en frihet att välja ordningsföljden, och inte är bunden av att ha den i en bestämd ordning exempelvis alfabetisk, numrerad klassindelning, eller som tidsföljd kan man rangordna staplarna. Sådana "tio-i-topp" diagram kan väcka intresse.

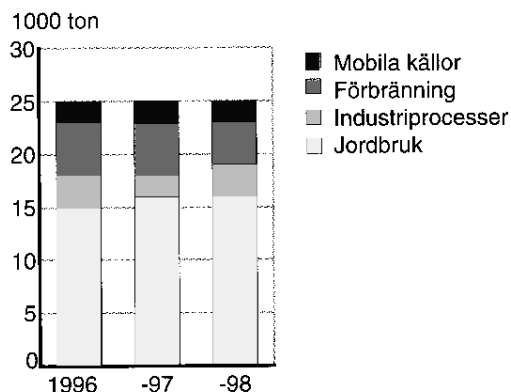
Ett annat sätt att ändra den "konventionella" ordningen på staplarna är exemplet i början av detta avsnitt där länen ges i syd-nordlig ordning. Nästan som en sverigekarta!

Cirkeldiagram

När man vill visa hur "ett helt" fördelar sig på grupper eller i delar är det lämpligt att visa det i form av en cirkel som delats i sektorer i de proportioner som grupperna eller delarna representerar. Det kallas ibland för paj- eller tårtdiagram.

Utsläpp av dikväveoxid 1996–98

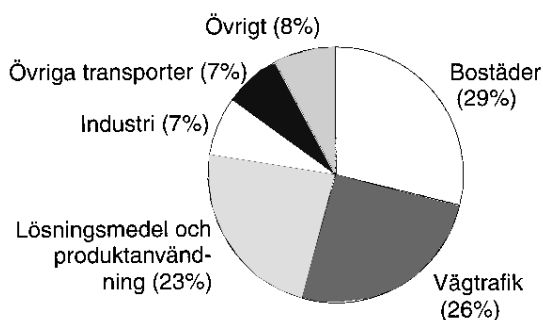
Emissions of N_2O , 1996–98



Källa: Mi 18 SM 9901, Naturvårdsverket och SCB

Utsläpp av flyktiga organiska ämnen från olika källor 1998

Emissions of NMVOC, 1998



Totalt ca 406 000 ton

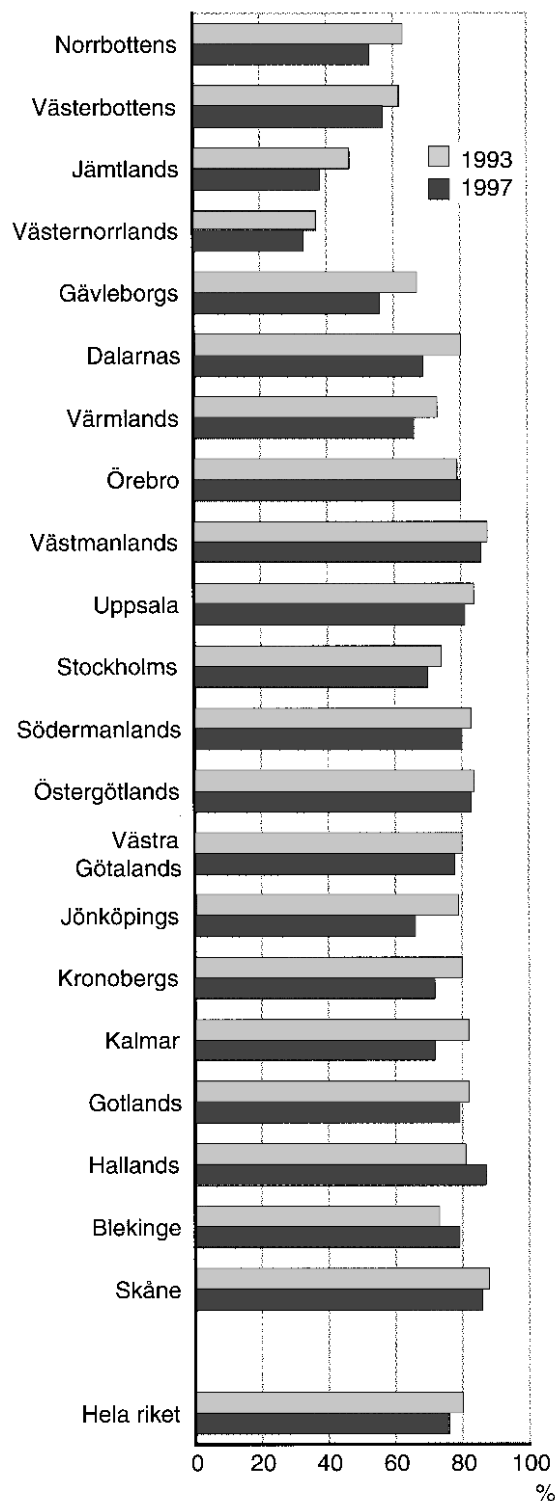
Källa: Mi 18 SM 9901, Naturvårdsverket och SCB

Exempel 4.5

Utsläpp av dikväveoxid 1996-1998. Utsläpp av flyktiga organiska ämnen från olika källor 1998. Ur *Naturmiljön i siffror (2000)*, sid 45.

Andel handelsgödslad¹ areal 1993 och 1997

Per cent arable land farmed using fertiliser, 1993 and 1997



1) Kväve från handelsgödsel

Källa: Na 30 SM 9402 och Na 30 SM 98 03, SCB

Exempel 4.6

Andel handelsgödslad areal 1993 och 1997.

Hälsosofarliga och konsumenttillgängliga produkter fördelade på faroklass 1997

Products hazardous to health and obtainable by consumers, classified in terms of hazard category, 1997

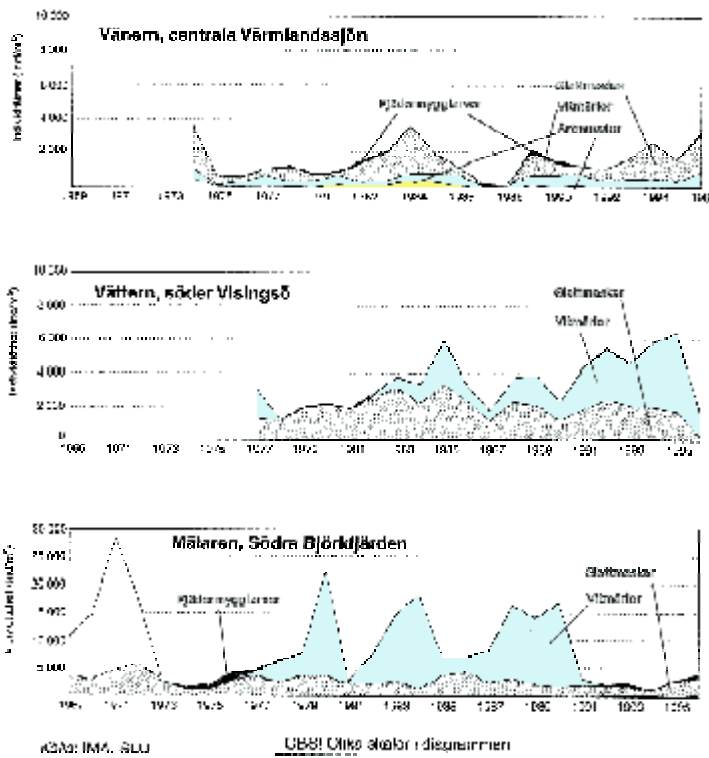
Faroklass/Produkttyp	Kvantitet, ton	Antal produkter
Giftig (T)		
Drivmedel (för förbränningsmotorer)	7 050 000	47
Uppvärmningsmedel	426 000	6
Explosivämnen	5 100	8
Appreturmedel	..	1
Frätande (C)		
Bindemedel (övriga)	58 400	8
Mur- och putsbruk	49 900	18
Syntesråvara (övriga)	14 100	2
Katalysatorer (övriga)	14 100	4
Irriterande (Xi)		
Bindemedel (övriga)	2 320 000	3
Drivmedel (för förbränningsmotorer)	151 000	6
Uppvärmningsmedel	54 500	1
Mur- och putsbruk	50 300	66
Hälsoskadlig (Xn)		
Drivmedel (för förbränningsmotorer)	3 510 000	60
Uppvärmningsmedel	1 370 000	22
Bränslen, brännolja (övriga)	603 000	16
Lösningsmedel	124 000	24
Måttligt hälsoskadlig (V)		
Drivmedel (för förbränningsmotorer)	728 000	7
Uppvärmningsmedel	21 500	4
Färger, lacker+Övriga färger och lacker	20 100	432
Spolarvätska	8 500	46

Källa: Mi 45 SM 99 01, Keml och SCB

Exempel 4.7

Miljö och hälsosofarliga kemiska produkterfördelat på farlighetsklass, 1997. Ur Naturmiljön i siffror (2000), sid 118.

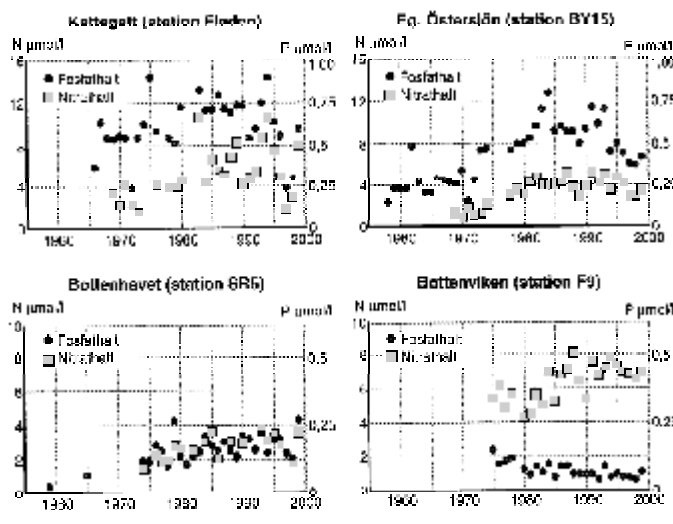
Djupbottenfaunans individtätet i augusti (Vänern och Vättern) och september/oktober (Mälaren) 1969–96
Numerical density of benthic fauna in August (Lakes Vänern and Vättern) and September/October (Lake Mälaren), 1969–96



Exempel 4.8

Djupbottenfaunans individtätet i augusti (Vänern och Vättern) och september/oktober (Mälaren) 1969-96. Ur Naturmiljön i siffror (2000), sid 153.

Fosfat- och nitrathalter i Västerhavet och Östersjön 1958–99
Phosphate and nitrate concentrations in seas west and east of Sweden, 1958–99



Källa: Svensk miljöövervakning/Dallic monitorings programme (DMF) samt Stahl

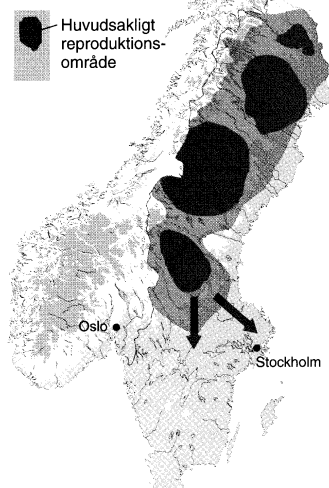
De halter som återges i diagrammen är medel värden för 0–5 m djup för månads- och års- (i Kattegatt och Fj. Östersjön) snittvärden november i Bottenhavet och i Bottenviken.

Exempel 4.9

Fosfat- och nitrathalter i Västerhavet och Östersjön 1958-99. Ur Naturmiljön i siffror (2000), sid 160.

Björn 1999

Antal djur: ca 1 000



Exempel 4.10

Rovdjurens utbredning 1999. Björn 1999. Ur Naturmiljön i siffror (2000), sid 188.

Cirkeldiagrammet är ett alternativ till stapeldiagram då man vill jämföra andelar. I exemplet visar cirkeln och stapeln samma fördelning.

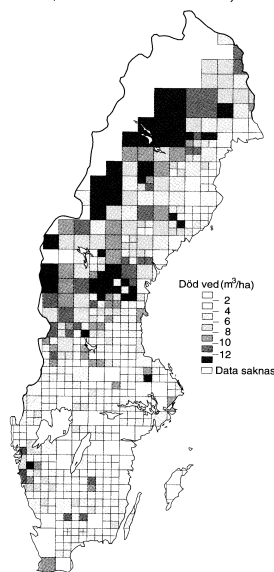
Frekvenspolygon

Frekvenspolygon kan användas som alternativ till histogram. I detta fall sammanbinds mittpunkterna på de översta sidorna av histogrammets staplar. Resultatet blir en kurva. Det bakomliggande histogrammet tas inte med! Om man har flera klassindelade fördelningar som man vill jämföra, kan det ofta vara enklare att göra det med frekvenspolygoner inritade i ett diagram, än med flera histogram.

Linjediagram

Tidsutveckling visas vanligen med *tidsserier* i form av kurvor, s.k. linjediagram. Ibland kan dock stapeldiagram vara ett alternativ. Stapeldiagram framhäver enskilda tidpunkter, och passar bäst om man ska jämföra ett fåtal tidpunkter. De är mindre lämpliga om man har flera serier i samma diagram. Linjediagram framhäver utvecklingsmönster och är lämpliga även då man har flera serier. De ger en klar bild av utvecklingen över tiden. Det ska vara lätt för ögat att följa var och en av flera kurvor om man vill jämföra flera variabler i samma diagram. Flera exempel finns i kap 5.4. Tidsserieanalys.

Figur 3.83 Geografisk fördelning av mängden död ved i skog (m³sk)
Geographical distribution of dead wood in forest (million cubic metres of timber)



Källa: Riksskogstaxeringen 1994–98, SLU

Exempel 4.11

Geografisk fördelning av mängden död ved i skog (m³sk). Ur *Naturmiljön i siffror (2000)*, sid 173.

Flödesdiagram

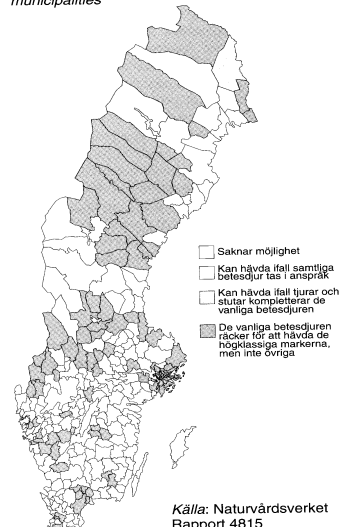
Om man vill visa statistik över storleken på olika bestånd och flöden mellan dem använder man sig av flödesdiagram. Exempelvis kan man visa olika typer av balanser, exempelvis energibalanser, där olika energislag "flödar" till olika slutanvändare.

Flödesdiagrammet kan ses som en vidareutveckling av stapeldiagrammet, eftersom storleken på variabelvärdena kan visas med bredden på staplar eller pilar.

Att visa statistik med hjälp av kartor

Statistiska kartor är det bästa sättet att visa variation, skillnader och likheter mellan olika geografiska områden, län, kommuner, avrinningsområden, klimatzoner eller andra. Ett alternativ är att indela kartan i rutnät eller med linjer som sammanbinder punkter med samma variabelvärden, isolinjer. Exempel på olika statistiska kartor ges nedan. Dessa är för enkelhets skull tagna ur *Naturmiljön i Sverige, SCB 2000*. Många bra kartor finns även i *Monitor* och *Nationalatlasen*. Man kan även vända sig till expertis inom området geografiska informationssystem.

Kommuners förutsättningar att hävda högklassiga naturbetesmarker
Potential for high-class natural pasture, by municipalities



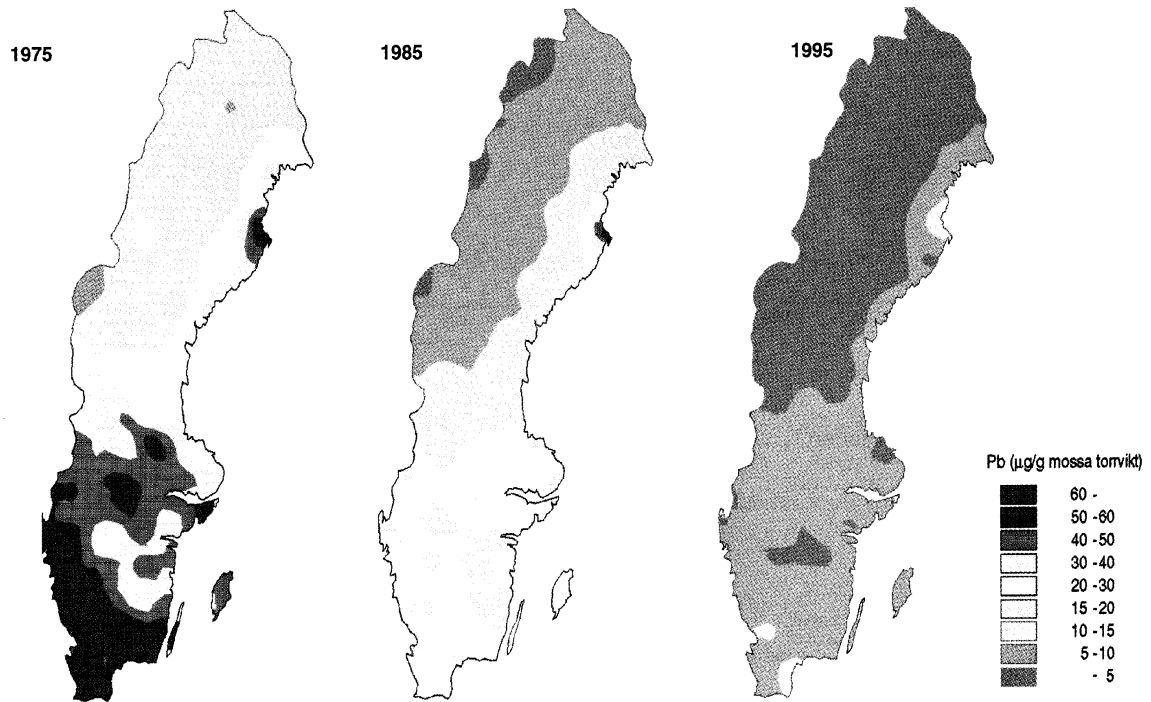
Källa: Naturvårdsverket Rapport 4815

Anmärkning: Som "vanliga betesdjur" räknas får, getter, kvigor och köttdjur. Till "samtliga betesdjur" räknas dessutom mjölkkor, tjurar och stutar.

Exempel 4.12

Kommuners förutsättningar att hävda högklassiga naturbetesmarker. Ur *Naturmiljön i siffror (2000)*, sid 180.

Figur 3.14 Bly i mossor 1975–95
Lead in moss, 1975–95

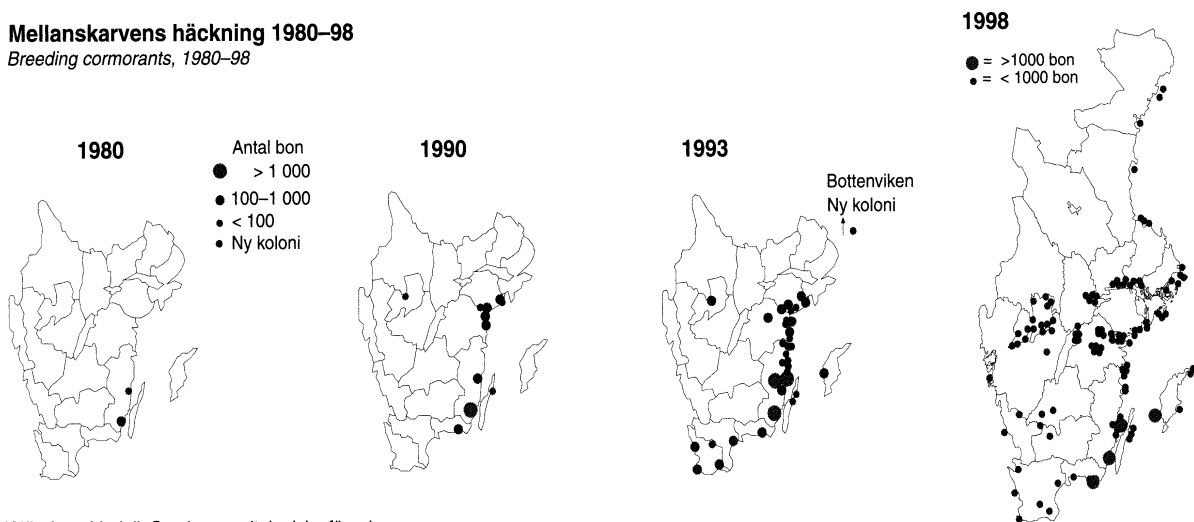


Källa: Nord 1996:37 samt Åke Rühling, Lunds universitet, via Datavärd IVL

Exempel 4.13

Bly i mossor 1975-95. Ur Naturmiljön i siffror(2000), sid 138.

Mellanskarvens häckning 1980–98
Breeding cormorants, 1980–98



Källa: Lars Lindell, Sveriges ornitologiska förening

Exempel 4.14

Mellanskarvens häckning 1980-98. Ur Naturmiljön i siffror (2000), sid 194.

tem, GIS, exempelvis redaktionen för Monitor eller SCB:s program för Regional planering och naturresurshushållning.

Ytmönsterkartor

Då skillnaden mellan regioner är små, men betydelsefulla är ytmönsterkartor ofta den lämpligaste presentationsformen. I tabeller och stapeldiagram är det svårt att se mönster om alla värden är nästan lika. Men till kartan kan man ju välja sin klassindelning så att de skillnader man vill påvisa verkligen kommer fram.

Kartogram-kartor med diagram och symboler

Ett kartogram är en underlagskarta, som används för lokalisering av de geografiska platser eller områden för vilka man vill visa variabelvärden.

Ofta ritas variabelvärdena in i kartan i form av bilder, symboler eller små diagram. Mängder kan exempelvis visas i form av olikstora staplar, eller cirkular. Andelar kan ges i form av cirkeldiagram etc.

Även flöden kan visas i kartor. Exempelvis kan bredden på linjen ange transportmängden på resp. vägsträcka.

Explorativ grafisk dataanalys

Jämförelse mellan två variabler i punktdiagram

När man ska undersöka och beskriva stora datamaterial kan det vara lämpligt att börja med att plotta materialet efter någon enkel princip. Ofta är observationerna i ett statistiskt material talpar. Det betyder att man har *två variabler* som kan tänkas samvariera. I detta fall kan man pricka in materialet i ett spridningsdiagram (punktdiagram, eng. scatterplot) innan man exempelvis beräknar regressionskoefficienten. Nedan beskrivs två enkla metoder att granska en variabel i taget.

Stam-bladdiagram

En metod att åskådliggöra stora datamaterial är att för varje variabel göra separata frekvens-

beräkningar. Det vanliga är att man, sedan man fått en tabell med frekvensvärden, ritar in dessa värden i ett histogram. Ett alternativ är att ställa samman data i ett *stam-bladdiagram*.

Ett **stam-bladdiagram konstrueras** genom att man ritar ett lodrätt streck och till vänster om detta anger exempelvis de heltalssiffror som förekommer i materialet. Detta bildar stammen i diagrammet. Sedan sätter man blad på stammen genom att efter respektive heltalssiffra ange den decimal-siffra som var och en av observationerna har.

Stam-bladdiagrammet kännetecknas av att det är enkelt och robust. Med robust menas här att stam-bladdiagrammet är okänsligt för eventuella felaktigheter i datamaterialet. I motsats till histogrammet, som det liknar efter en 90 graders vridning, är mera av den ursprungliga information med i stam-bladdiagrammet.

Lådagram

Lådagram (Boxplot) är en metod att illustrera spridningen i ett material. Det kan återge median, kvartilavståndet och extremvärden. Vanligen bestämmer man från ett stam-bladdiagram

- det mittersta värdet, **medianen** som åskådliggörs med mittstreck i lådan
- den **tredje kvartilen** (75:e percentilen) och den **första kvartilen** (25:e percentilen). Avståndet mellan dessa är kvartilavståndet och får bli lådans höjd
- **skevheten** belyses genom skillnaden i storlek mellan lådans delar som bildats av de tre måtten, median och tredje resp. första kvartilen.
- eventuella **extremvärden** kan uppmärksammas på olika sätt. Ett exempel är den s.k. Tukey's standard boxplot, där värden som ligger på längre avstånd från lådans kant än 1,5 gånger lådans höjd anges med • (punkt). Man kan även rita in linjer till det lägsta och det högsta värdet.

Layout av diagram

Datamaterialets karaktär bör bestämma vilken typ av diagram man ska använda. Därefter kan det vara bra att göra några olika skisser för att få en bild av hur det slutliga diagrammet ska se ut. Detta kan ske med hjälp av de dataprogram som man har till buds. Den moderna tekniken kan dock fresta en att visa för mycket på en gång. Man bör akta sig för

överlastade diagram, lika väl som för alltför detaljrika och omfattande tabeller.

Siffror i diagrammet?

Ett cirkeldiagram bör visa vilka sektorer som dominerar, vilka som är små etc. Om man tycker sig behöva ange exakta värden bör man göra det i en tabell. Möjligen kan man ha värden i diagrammet om man vill göra bådadera men inte har plats för såväl figur som tabell. Numera kan ett alternativ vara att hänvisa den som vill ha de exakta värdena bakom en figur till en internetlänk där varifrån data har hämtats, se t ex Naturvårdsverkets statistikpublikation *Miljötilståndet i skogen 1999*.

Stapeldiagram är ofta konstruerade för direkt avläsning på en skala. Här är det direkt olämpligt med siffror ovanför staplarna, eftersom proportionerna i diagrammet kan tyckas förändrade, ”ögat lägger till trycksvärtan” i siffrorna till staplarna!

Markeringar, mönster och skrafferingar

Kurvor

Om man har flera linjer i ett diagram använder man oftast olika markeringar för att skilja dem åt. Om en linje utgör en summa (en total) av vad de övriga linjerna visar kan det eventuellt vara bra att ge denna en mer markerad, ”tjockare” linje. Om man dessutom skriver de förklarande texterna vid linjerna slipper man ifrån en speciell teckelförklaring (legend).

Ytor

Tänk på gråskalan eller svartvitskalan. Mörka nyanser kan bli mycket dominant. Oftast är det bäst att överlag använda dämpade nyanser ur gråskalan. Dataprogrammen låter oss ofta välja mellan olika typer av prickar och randningar. Om man har många ytor bredvid varandra kan helhetsbilden bli rörig och i vissa fall leda till synvillor. Var observant!

Kartogram

Det ska vara enkla diagram om man har många på en karta.

Tredimensionella diagram

När datamaterialet består av tre variabler, som man vill beskriva samtidigt, kan man visa det i antingen tvådimensionella eller tredimensionella diagram.

Exempel på tvådimensionella diagram i detta sammanhang är stapeldiagram där två av variablerna finns i x- resp. y-axeln och den tredje motsvaras av olika skrafferingar.

Det finns också tredimensionella varianter av sådana diagram. Två variabler beskrivs av axlarna i ett bottenplan och den tredje av staplarnas höjd. Varianter av denna diagramtyp är perspektivritade kartor med staplar. Om dessa ritas med omsorg och staplarna inte inkräktar på varandra kan de vara mycket intresseväckande och ge en bra bild av samband. Men alltför ofta är dessa diagram svåra att avläsa, skalorna blir svåra att hitta och staplar kan gömma sig bakom varandra. En varning alltså för ett lättsinnigt bruk av de tredimensionella diagrammen! De kräver mycket arbete och urskillning för att bli tillräckligt bra.

Det kan även vara på sin plats att utfärda en varning för ”onödig tredje dimension”. Eftersom det med dagens teknik är så enkelt att rita diagram, och pynta dem på olika sätt, vrida på dem och lägga till ett djup i form av en tredje dimension, är det frestande att ibland använda sig av för mycket av dessa finesser. När data endast har två dimensioner, ger exempelvis en tredimensionell ”tårta” missvisande jämförelser.

Avbrutna och stympade skalor

Helst bör man undvika att bryta och stympa skalor eftersom det lätt leder till missuppfattningar. Det gäller speciellt den vertikala skalan, y-axeln. När man ändå anser sig vara tvungen att börja sin kurva vid någon annan punkt än vid 0, bör man markera detta.

Fördjupade studier

För fördjupning hänvisas till någon lärobok i Deskriptiv statistik eller till standardverk, som Tufte, E, R, *The visual display of quantitative information*; Tukey, J, *Exploratory Data Analysis*. På svenska finns Statistikens bilder - att skapa diagram av Wallgren et.al., SCB, 1994.