

# Statistisk klokskap

## — exempel för bl a den nya HSE-matematiken

GÖRAN ANDERSSON

Det är viktigt att skolans undervisning i statistik inte enbart tränar eleverna på att genomskåda "hur man ljugar med statistik". Minst lika viktigt är att ett positivt synsätt förmedlas. Att utbildningen poängterar statistikens roll som ovärderligt beslutsunderlag och ger tips om god statistik som — rätt tolkad — kan avslöja fördomar, ge insyn i hur samhället sköts och bli "murbräcka för reformer".

Sedd isolerad har den här artikeln kanske viss slagsida mot de kritiska aspekterna, något som dock mer än väl bör uppvägas av alla exempel på positiv statistik-användning i omgivande artiklar. *Göran Anderssons* bidrag innehåller ett antal övningsexempel för den lärare som huvudsakligen läser artikeln för sin egen fortbildning. Dessa exempel torde också kunna komma till användning i klassrummet. Dessutom finns under rubriken "elevuppgift" förslag till arbetsuppgifter kring frågeställningar som mestadels kräver att eleverna samlar in kompletterande data.

Merparten av innehållet berör den nya HSE-matematiken; de tre inledande avsnitten bör även lämpa sig för intresserade elever i åk 9.

### 1. Diagram som föder projekt?

I nästa avsnitt finns ett par tidningsdiagram som inte är särskilt lyckade. Det är heller ingalunda svårt att i pressen finna liknande diagram med vilseledande utformning.

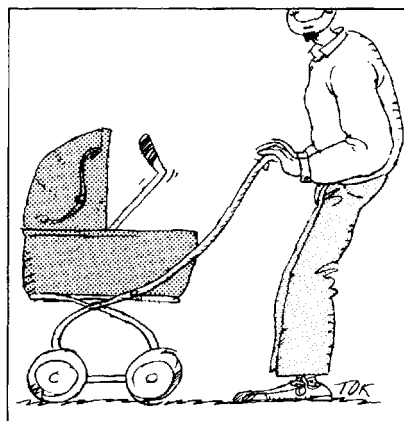
Samtidigt kan man på sätt och vis förstå dagspressens diagramkonstruktörer. En större dagstidning motsvarar 300—400 boksidor. Så i ett hushåll med ett par dagstidningar är det verkligen konkurrens om läsarnas uppmärksamhet och ett vanligt stoldiagram eller histogram är väl inte det första Medel-Johansson kastar sig över i frukostbrädskan. Nackdelen är dock att iveren att förse diagrammen med fantasifulle utsmyckningar alltför ofta får till följd att korrektheten blir lidande.

Men ibland blir resultatet bra. Diagrammet på nästa sida är så pass uppseendeväckande att det har chans att sälja sitt budskap, samtidigt som det är korrekt när det gäller skalor och proportioner.

Staplarna i histogrammet visar hur många procent av de 240 registrerade elitseriespelarna i ishockey säsongen 1984/85 som är födda respektive månad. Som framgår av bildtexten visar den streckade kurvan fördelningen efter födelsemånad för alla födslar i Sverige åren 1961—65, dvs de år då merparten av de här elitspelarna föddes. Vi ser att de spelare som lyckats så bra i sina ishockeyambitioner att de nått elitserien är klart överrepresenterade i födelsemånaderna jan—

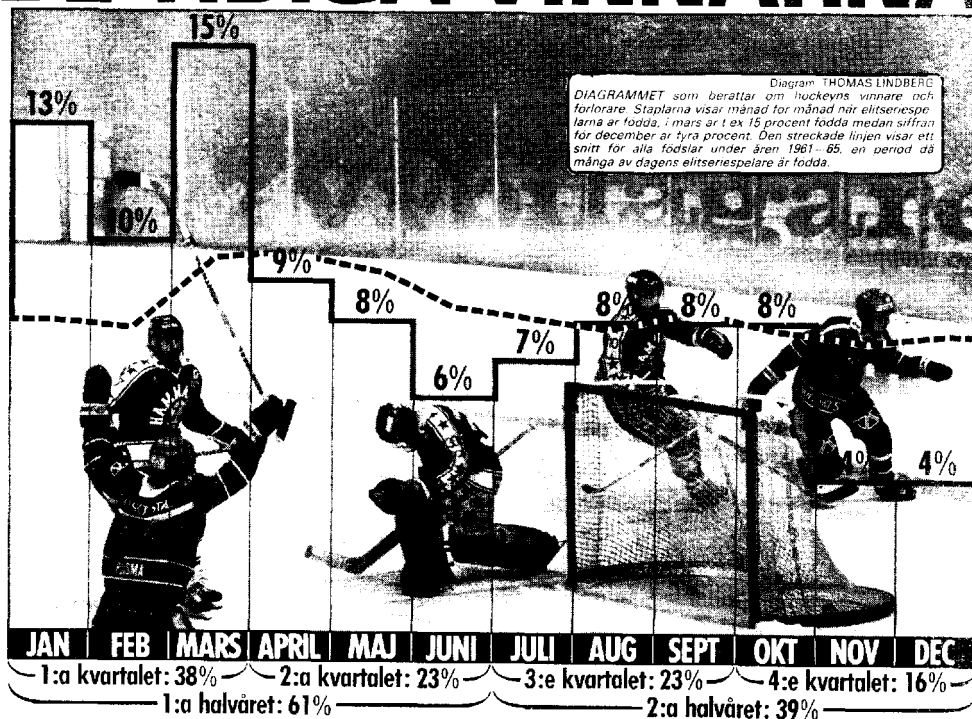
mars och i motsvarande mån underrepresenterade i det fjärde kvartalet. Ska man alltså vara född i början på året för att ha störst chans att bli en riktigt bra ishockeyspelare?

Den skeptiske kanske vill pröva om slumpen har spelat ett spratt. Skillnaden mellan de 240 spelarnas fördelning och den för samtliga födslar kanske beror på slumpen. Det är rätt enkelt att testa en sådan hypotes med hjälp av exempelvis ett s k chitvåtest<sup>1)</sup> (eller ännu hellre Kolmogorov-Smirnovs test). Ett sådant test visar att skillnaden



<sup>1)</sup> Se någon av titlarna i denna utgåvas avsnitt med litteraturtips.

# DE TIDIGA VINNARNÄ



Expressen 3/11 -84

är klart signifikant (statistiskt säkerställt). Enligt det senare av de båda testen är det mindre än 1 chans på 500 att skillnaden kan tillskrivas ren slump.

Det här diagrammet väcker ganska olustiga frågeställningar. Är förklaringen till snedfördelningen efter födelsemånad att utslagningen i barn- och ungdomsishockeyn främst drabbar dem som råkar vara födda i slutet på året? I ishockey, liksom i många andra idrotter, lär det ju vara så att tävlingsmatcher sker efter strikt åldersindelning. Att 10-åringar spelar rätt prestigefyllda matcher mot andra 10-åringar, medan 11-åringarna har sin egen serie etc. Tämigen självklart har den 10-åring som är född i januari större chans att hävda sig än den nästan ett år yngre födelseårskamrat som är född i december.

Nu är väl inte ishockey det viktigaste här i livet, kanske inte ens tennis (kommer vi däremot in på fotboll tycker artikelförfattaren att det åtminstone börjar handla om väsentligheter . . .). Men det är ju inte tävling bara i idrottssammanhang. Hur är det i skolan? Klarar sig de som är födda i början på året bättre än senhöstbarnen? Varierar födelsemånadens eventuella betydelse mellan olika kategorier av skolämnen? Har födelsemånaden större betydelse bland pojkar än bland flickor?

Alla skolor bör ha massvis med data för analys av eventuella samband mellan födelsemånad/kvartal och studieframgång. Det bör också vara tämligen enkelt att avidentifiera detta material så att elevernas dataunderlag bara innehåller uppgift om kön, födelseår och -månad samt betyg i olika ämnen. Resultatet av undersökningen kanske blir ett argument för skolstart varje halvår.

Om Du skulle vilja handleda Dina elever i ett sådant här projektarbete är det ingalunda nödvändigt, men en viss fördel, om Du är någorlunda insatt i chitvåanalys. Testets idé är inte särskilt svårbegriplig (även om det skulle föra för långt att här beskriva metoden). Dessutom kan det vara en trevlig tillämpningsuppgift för duktiga elever att skriva ett datorprogram som beräknar den s k chitvåsumman för en korstabell. Å andra sidan torde de flesta programpaket på gymnasienivån för datorbearbetning av statistiska material innehålla denna facilitet.

För den som själv vill räkna på ishockeyexemplet ger vi här en sammanställning av elitseriospelarnas procentuella fördelning efter födelsemånad och motsvarande fördelning för samtliga födselar i Sverige 1961-65.

	Jan	Feb	Mars	Apr	Maj	Juni	Juli	Aug	Sep	Okt	Nov	Dec	S:a
De 240 ishockey-spelarna	13	10	15	9	8	6	7	8	8	8	4	4	100
Samtliga födselar 1961-65	8,1	7,8	9,6	9,7	9,3	8,3	8,2	7,9	8,2	7,8	7,5	7,6	100,0

## 2. Vilsledande diagram

Gymnasieläroböcker i matematik (och även böcker för högstadiet) innehåller oftast ett avsnitt betitlat "Att luras med statistik", vilket vanligen är liktydigt med "Att luras med diagram". Där exemplifieras manipulationer med skalförändringar och stympade axlar, och diagram av den här typen där man använt höjdpportionalitet i stället för areapportionalitet.

Uppfinningsrikedomen hos massmedias diagramkonstruktörer är imponerande men ibland får den alltför fritt spelrum på bekostnad av saktligheten.

Visst är det fyndigt att illustrera köttprisernas komponenter med de här trevliga djuren. Bonden får alltså inte hälften men i alla fall en väldigt fin bit av grisen och kossan, särskilt jämfört med öron, horn, tryne och mule som ändock skall åskådliggöra en tredjedel av vad bonden får.

# FÄLLDIN har krympt till hälften

Löget är nu bevisligt för både Ole Billsten och Thorsbjörn Fäll. Din Völjerna överger nu i allt snabbare takt de två mittenpartierna.

Om trenden i den senaste tidens opinionsmätningar håller i sig går centera och folkpartiet mot undergången som mågorlunda stora partier.

Partier som i slutet av 1973 hade 14,5 procent av rösterna i riksdagen har idag bara 14,5 procent.

Detta betyder att de två mittenpartierna har krympt till hälften.

Men det är inte bara de två mittenpartierna som krympt till hälften. Det är också de två stora partierna, centerpartiet och folkpartiet. De har också krympt till hälften. Det betyder att de två stora partierna har krympt till hälften.



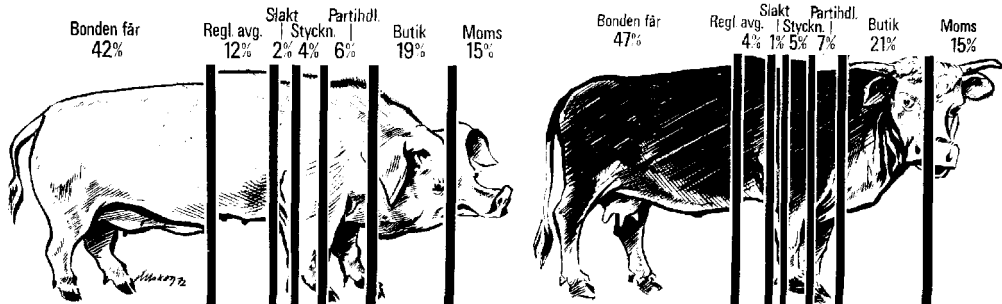
1980: 14,5 procent



1973: 29 procent

Aftonbladet 26/6 -80

# BONDEN FÅR INTE HÄLFTEN



Bonden som fött upp en gris får cirka 42 procent av det pris som vi får betala i butiken. Detta är ett genomsnitt. Att bondens inkomst inte blir större beror bl a på regleringsavgiften. Det är pengar till staten som just nu bl a används för att kompensera exportförlusterna på vår rekordstora spannmålssektor.

Slaktkostnaden är ganska låg, men slakteriet får samtidigt en inkomst på försäljning av hudar och andra organ som inte går till människoföda. Skavsavlingen som bl a sker hos brun och samford gör 4 procent. Partihandeln (ICA, ASK, KF) tar 7 procent. Butiken som får den färdigstykade och ofta paketerade varan tar 19 procent. Momsen är 15 procent.

Vi brukar räkna med att momsen är 17,65 procent. Men 15 procenta moms är det som som 17,65 procenta pålägg.

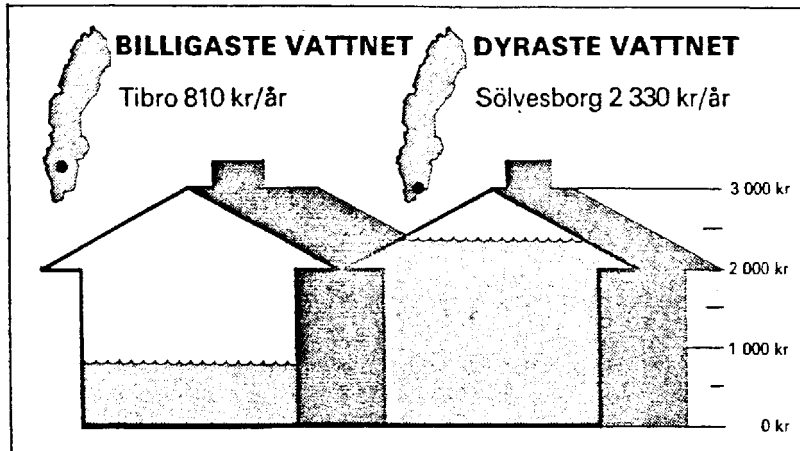
Bonden får en lite större procentdel andel av det slutliga priset när han levererar ett ungt djur till slakt. Detta om man jämför med en gris. Regelingsavgiften är lägre liksom slaktkostnaden. Det senare beror på att det vill slakt av en ko blir större förtjänst på hudar och andra organ som inte går till människoföda. Genom att dessa inkomster är procentuellt högre hos mjölköt jämfört med grisens blir det en lägre slaktkostnad. I svarta tal siffrorna tar sig själva.

Expressen 4/10 -72

Ett ganska nytt diagram vill visa skillnaden mellan det dyraste och det billigaste kommunala vattnet i landet med vattennivån i dessa båda, typiskt vattenskadade hus.

Den vanartige gossen Ruda skulle kanske sammanfatta diagrammets information så här: *Ju dyrare vatten, desto mer fruktansvärda vattenskadador! Vid 3 000 kr/år når vattnet skorstenen!*

# Dyrt och billigt vatten



Teckning: BO PERSSON

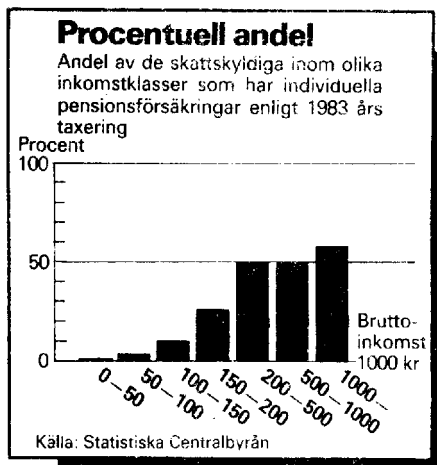
Hela 1520 kronor om året är skillnaden i vattenskaded mellan landets billigaste kommun (Tibro) och den dyraste (Sölvesborg).

DN 5/9 -84

### 3. Tämligen korrekt diagram fast i fel sammanhang

Här följer ett diagram i ett tidningsklipp (Dagens Nyheter 18/8 1986) och en text som kan användas som en övningsuppgift.

Exempel 1



Det citerade uttalandet kan också formuleras ungefär så här: "Merparten av de individuella pensionsförsäkringarna innehas av höginkomstagare." Tycker Du att uppgifterna i diagrammet stöder en sådan slutsats?

Diagrammet visar *andelen* (men inte *antalet*) pensionsförsäkringstagare i varje inkomstklass. Även om 60 % av miljoninkomsttagarna har försäkring betyder detta inte särskilt många personer. Det gör däremot 10 % av dem i det betydligt vanligare inkomstläget 100 000—150 000 kr.

Klippet härrör från en partsinlaga skriven av verkställande direktören i Svenska försäkringsbolag riksförbund. Som motvikt presenterar han

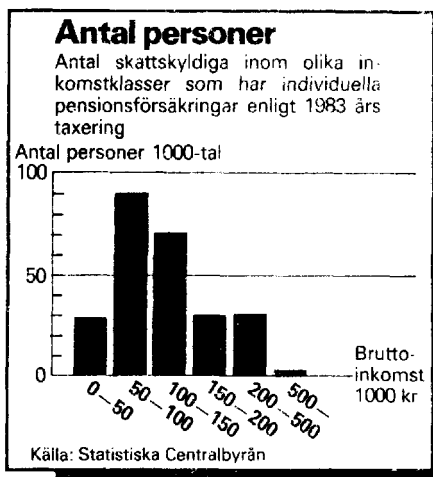
**R**ealrätteskatteskommittén har i sitt betänkande ett diagram enligt figur 1. Med hänvisning till detta uttalar man att "de individuella pensionsförsäkringarna i verkligheten är koncentrerade till höginkomstagare". Även LO-tidningen har i nr 25/26 förelästs att dra samma slutsats.

vidstående diagram över *antalet* pensionsförsäkringstagare i de olika inkomstklasserna och då ser man t ex att *antalet* pensionsförsäkrade som väntat är lågt i inkomstklasserna över halvmiljonen (eftersom det inte är särskilt många i riket som har så höga inkomster).

Nu rörde det sig som sagt om en partsinlaga. De som hade att fatta beslut i frågan bör ha varit mest betjänta av ett diagram utvisande *hur mycket pengar* från olika inkomstklasser som placerats i pensionsförsäkringar. Ett sådant diagram skulle förmodligen ha sett ut som ett mellanting av de två publicerade diagrammen.

#### Elevuppgift

Kontakta SCB, som åberopas som källa för de båda diagrammen i det här ärendet, och försök få fram statistik över hur de belopp som placerats i pensionsförsäkringar fördelar sig efter försäkringstagarnas inkomst. Presentera statistiken bl a i ett diagram konstruerat efter samma principer som de båda ovanstående.

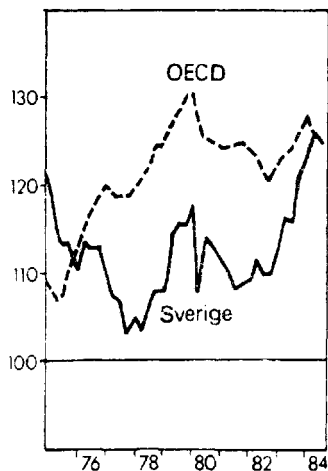


#### 4. Skäländring och byte av basår

De tre diagrammen som följer kommer från en sammanställning i Svenska Dagbladet (16 jan 1985) som Göran Albinsson Bruhner har gjort. Alla tre diagrammen vill visa samma sak, nämligen hur den svenska industriproduktionen har utvecklats i jämförelse med OECD under det senaste decenniet.

#### 1. Industriproduktion

Kvartalsvis. Index 1970=100



i all- är en allmänhet fås att bekänna vad som helst.

medkänsla. Alla profes- jag känner står främman- att manipulera- ikt att vilseföra- et en helt annan his- ven garvade yrkessta- medvetet och oavsikt- råka vinkla tabeller och

Ansvar för statistikens dåliga anseende bär alla de ombudsmän och politiker som misshandlar siffror.

Och det är med statistiska uppgifter som med människor. Torteras de tillräckligt länge kan de i

Regeringens finansplan är en oböjld partsinlaga. Så var det under de borgerliga regeringsåren och så är det nu.

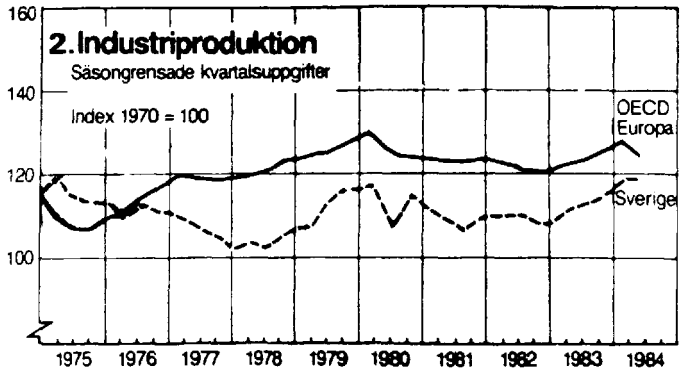
Därför är det uddlost att helt allmänt förebrå finansministern att han för sina ändamål hårdrar och vrider tillgänglig statistik. Frågan är dock om inte herr Feldt i åtminstone ett fall över-skrider den politiska anständighetens (mycket vida) gränser.

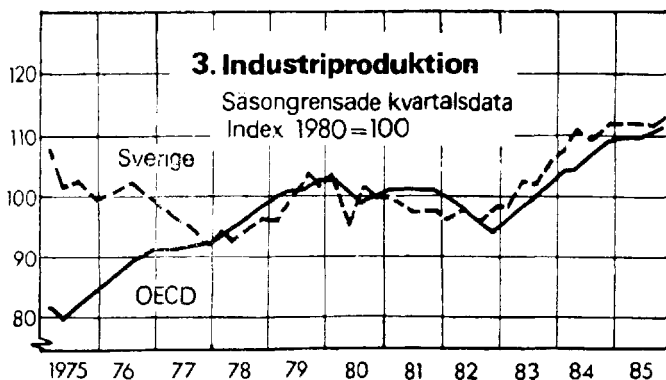
Jag syftar på ett diagram i årets Finansplan. Det visar hur den svenska industriproduktionens volym utvecklats i jämförel-

#### 2. Industriproduktion

Säsongrensade kvartalsuppgifter

Index 1970 = 100





Hur kan de tre diagrammen ge så olika intryck?

Skillnaden mellan diagram 1 och 2 är enbart effekten av att man i det högra har krympt y-axeln samtidigt som x-axeln tänjts ut. Det vänstra är hämtat från årets finansplan där finansministern (enligt Albinsson Bruhner) vill visa hur illa den svenska industriproduktionen utvecklats efter det borgerliga maktövertagandet 1976 i jämförelse med Västeuropas. Efter några år av socialdemokratiskt styre har dock Sverige kommit ifatt.

Enligt Albinsson Bruhner kunde det tredje diagrammet vara specialbeställt av det borgerliga blocket (i själva verket är det producerat av OECD-kansliet i Paris). Det framgår klart att Sverige sedan 1977 hållit jämna steg med hela OECD (alltså inkl expansiva Japan). En skillnad mellan diagram 2 och 3 är alltså att jämförelsen i diagram 2 sker mot europadelen av OECD. Men den väsentliga anledningen till att diagrammen ger olika intryck är att diagram 3 har 1980 som basår i stället för 1970.

Effekten av byte av basår för indexserier är något som inryms i den nya läroplanen där index fått större utrymme. I nästa avsnitt ges ytterligare ett drastiskt exempel på följden av basårsbyte.

#### Elevuppgift

Försök att i diagram 2 för varje år avläsa medelvärdet av de fyra kvartalsindextalerna för OECD. Uppskatta på samma sätt årsmedelvärdena av kvartalsindextalerna för Sverige. Räkna sedan om dessa båda indexserier så att de får 1980 som

basår. Rita in dessa nya indexserier i ett diagram och jämför med diagram 3. (Alternativt kan eleverna försöka att med hjälp av SCB skaffa fram originaldata.)

### 5. Exempel på deflatering och på effekten av byte av basår

#### Exempel 2

Beräkna indexserier med 1974 som basår över exporten i fasta priser för de båda länderna.

År	VÄSTTYSKLAND (DEM)		SVERIGE (SEK)	
	Export i miljarder (löp.pris)	Exportpris-index	Export i miljarder (löp.pris)	Exportpris-index
1974	230,6	80,7	70,5	176
75	221,6	83,8	72,0	190
76	256,6	87,0	80,2	201
77	273,6	88,5	84,2	212
78	284,9	89,9	98,2	224
79	314,5	94,1	118,2	251
80	350,3	100	130,8	282
81	397,0	105,8	144,7	307
82	427,8	110,4	168,0	343

#### Lösningsskiss

Även om det inte är nödvändigt för lösandet av uppgiften talar pedagogiska skäl för att man börjar med att räkna om de båda prisindexserierna så att de får samma basår, t ex 1974. För 1974, 1978 och 1982 får vi följande exportprisindextal:

	1974	1978	1982
Västtyskland	100	$\frac{89,9}{80,7} \cdot 100 = 111,4$	$\frac{110,4}{80,7} \cdot 100 = 136,8$
Sverige	100	$\frac{224}{176} \cdot 100 = 127,3$	$\frac{343}{176} \cdot 100 = 194,9$

Därefter räknar vi ut exporten i 1974 års priser så här:

	1974	1978	1982
Västtyskland	230,6	$\frac{284,9}{1,114} = 255,7$	$\frac{427,8}{1,368} = 312,7$
Sverige	70,5	$\frac{98,2}{1,273} = 77,1$	$\frac{168,0}{1,949} = 86,2$

Slutligen uttrycks ovanstående serier av exporten i 1974 års priser som två indexserier med 1974 som basår.

	1974	1978	1982
Västtyskland	100	$\frac{255,7}{230,6} \cdot 100 = 110,9$	$\frac{312,7}{230,6} \cdot 100 = 135,6$
Sverige	100	$\frac{77,1}{70,5} \cdot 100 = 109,4$	$\frac{86,2}{70,5} \cdot 100 = 122,3$

För samtliga år i tidsserien erhålls, efter avrundning till heltal, följande indexserier över exporten i fasta priser.

År	Västtyskland	Sverige
1974	100	100
75	93	95
76	103	100
77	108	99
78	111	109
79	117	118
80	123	116
81	131	118
82	136	122

Svar a):	År	Västtyskland	Sverige
	1977	100	100
	78	103	110
	79	108	119
	80	113	117
	81	121	119
	82	125	123

**SUCCE FÖR SVENSK EXPORTPOLITIK!  
SVENSK EXPORT ÖKAR T O M LIKA MYCKET SOM VÄSTTYSKLANDS.**

Under perioden 1974—82 har Västtysklands export i fasta priser (ungefärligen tolkbar som exportvolymen) ökad med 36 % medan motsvarande ökning för den svenska exporten var 22 %.

Svar b):	År	Västtyskland	Sverige
	1979	100	100
	80	105	98
	81	112	100
	82	116	103

#### Förslag till elevuppgifter

1. *Byte av basår.* Låt klassen arbeta i smågrupper om ca 3 elever i varje. Den ena hälften av grupperna får föreställa journalister för en tidning som vill få svensk exportutveckling att framstå som så gynnsam som möjligt i jämförelse med Västtyskland. Vilket basår bör man välja? Räkna om de båda indexserierna så att de får detta basår och formulera slutsatsen som en slagkraftig tidsningsrubrik.

Den andra hälften får en liknande uppgift men här gäller det att välja basår så att den svenska exportutvecklingen framstår som så dålig som möjligt.

**FIASKO FÖR BORGARNAS EXPORTÅTGÄRDER!  
SVENSKA ÖKNINGSTAKTEN INTE ENS EN FEMTEDEL AV VÄSTTYSKLANDS.**

2. Hur har det gått i fortsättningen? Leta fram senast möjliga uppgifter om de båda ländernas export och exportprisindex och förläng de båda tidigare beräknade indexserierna.

3. Konstruera ett datorprogram för att underlätta att av lösningen av problem av typ exempel 2.

## 6. Inflationstakten i några länder — övningsuppgifter på tillväxtfaktorn

När det gäller övningsuppgifter på tillväxtfaktorn är det helt onödigt med "fejlkade" data. I Statistisk årsbok finns åtskilligt med intresseväckande statistik att räkna på.

Exemplet nedan visar att den som önskar klaga på inflationstakten i Sverige kan få vatten på sin kvarn. Men det är inte heller svårt att hitta länder som drabbats av betydligt högre inflation.

### Exempel 3

Tabellen visar konsumentprisindex (motsvarande) i några länder åren 1978, 1980 och 1984.

	KPI (1970 $\cong$ 100)		
	1978	1980	1984
Schweiz	151	163	194
Sverige	198	242	346
Tjeckoslovakien	105	112	121
Storbritannien	270	361	481
Israel	786	3236	179 919
Chile	232 773	419 601	842 169

Källa: Statistisk årsbok 1982 och 1986

Beräkna den årliga genomsnittliga inflationen för var och en av dessa länder under perioderna

a) 1970—1978 b) 1978—1980 c) 1980—1984

Lösningar på några av uppgifterna:

$$\text{Tjeckoslovakien } 1970\text{—}1978: \left(\frac{105}{100}\right)^{\frac{1}{8}} = 1,0061 \text{ dvs } 0,6 \%$$

$$\text{Chile } 1970\text{—}1978: \left(\frac{232\ 773}{100}\right)^{\frac{1}{8}} = 2,636 \text{ dvs } 163,6 \%$$

$$\text{Sverige } 1978\text{—}1980: \left(\frac{242}{198}\right)^{\frac{1}{2}} = 1,106 \text{ dvs } 10,6 \%$$

$$\text{Israel } 1980\text{—}1984: \left(\frac{179\ 919}{3\ 236}\right)^{\frac{1}{4}} = 2,731 \text{ dvs } 173,1 \%$$

### Elevuppgifter

a) Vad har hänt sedan 1984 med inflationen i länderna i exemplet. Sök i senaste upplagan av Statistisk årsbok eller i Monthly Bulletin of Statistics (FN). Även SCBs upplysningstjänst bör kunna hjälpa till.

b) Ta reda på om skolan eller kommunens bibliotek har en någorlunda aktuell årgång av FNs Statistical Yearbook. Leta upp tre utomeuropeiska länder som har haft låg inflation under de fem senaste åren och tre med hög inflation. Genomför motsvarande uppgift för länder i Europa. Gör en jämförande beskrivning i tabeller och diagram av inflationens utveckling i de utvalda länderna; bl a ska Du redovisa den genomsnittliga årliga inflationen.

Vad känner Du till om de politiska systemen i de utvalda länderna?

## 7. Problem med standardvägning

Enligt de lärare jag pratat med tycks standardvägning vara det svåraste att "sälja" till eleverna av de nya statistikmomenten för SE2—3. Kanske kan det vara till hjälp för någon lärare om jag visar hur jag brukar gå tillväga. Låt oss utgå från följande två övningsuppgifter.

### Exempel 4

Materialet och idén till det här exemplet kommer från en "klassisk" lärobok i statistik, nämligen Erland Hofsten: *Praktisk statistik*.

Enligt 1936 års partiella folkräkning var medelantalet barn per bestående äktenskap 3,16 bland gods- och hemmansägare och 2,77 bland jordbruksarbetare. Kan detta stämma? Har inte historien och litteraturen lärt oss att det vimlade av barn i statarfamiljerna?

En tänkbar snedvridande faktor är äktenskapets "ålder". Kanske gods- och hemmansägarfamiljerna i genomsnitt har existerat längre och att de därför har hunnit skaffa sig fler barn. Materialet medgav en uppdelning av familjerna efter äktenskapets varaktighet:

Äkten- skapets varak- tighet	Medelantal barn bland		Antal äkten- skap i riket
	gods/hem- mansägare	jordbruks- arbetare	
0—4	0,92	1,07	45 603
5—9	1,86	2,11	39 902
10—14	2,63	2,86	33 956
15—19	3,29	3,74	32 034
20—35	4,34	4,80	63 867
			214 662

Som synes blir nu bilden den motsatta. I varje "åldersklass" är medelantalet barn högre bland jordbruksarbetarna. Att totalmedelvärdet trots detta blir högre bland gods- och hemmansägarna måste bero på att dessa är överrepresenterade i högre åldrarna med flest barn.

Gör en jämförelse mellan gods-/hemmansägare och jordbruksarbetare som eliminerar inverkan av olikheten i fördelningen över varaktighetsklasser. Glöm inte att formulera slutsatsen av Dina beräkningar i ord.

### Exempel 5

En industri framställer bl a tre olika typer (A, B och C) av en viss detalj. Produktionen av dessa sker vid två separata gjutavdelningar (avd I och avd II). De färdiga enheterna går till kontrollavdelningen där de misslyckade utsorteras vid en röntgenkontroll.



Summa	Antal producerade enheter			Antal utsorterade enheter		
	Avd I	Avd II	Summa	Avd I	Avd II	Summa
A	3 000	2 000	5 000	210	180	390
B	2 000	4 000	6 000	40	160	200
C	1 000	4 000	5 000	80	240	320
Summa	6 000	10 000	16 000	330	580	910

Genomför den önskade jämförelsen. Formulera slutsatsen i ord.

Man vill göra en jämförelse mellan avd I och avd II avseende kassationsfrekvensen. (Kassationsfrekvens = relativa frekvensen misslyckade enheter.) Man vet emellertid att kassationsfrekvensen ej är densamma för de tre olika typerna (A, B och C) och vill därför ta hänsyn till att avdelningarna skiljer sig åt beträffande produktionsfördelningen. I nedanstående tabell redovisas antalet producerade och antalet utsorterade enheter under år 1982 fördelade efter typ och produktionsställe.

När eleverna bekantat sig med standardvägningsmetoden genom att räkna ett par övningsuppgifter, brukar jag göra en sammanfattning av bakgrund och lösningsmetod. Jag använder då begreppen *studerad variabel* (dvs den variabel man uttalar sig om), *jämförda grupper* ("grupper" har här en mycket vid betydelse — det kan som i exempel 4 vara grupper av individer, men också ett par företag eller ett par år som jämförs), *snedvridande faktor* och *tänkbara standardpopulationer*. För exemplen 4 och 5 blir sammanställningen så här:

<i>Exempel 4</i>	
Studerad variabel:	Medelantal barn per äktenskap
Jämförda grupper:	Gods-/hemmansägare och jordbruksarbetare
Snedvridande faktor:	Äktenskapets varaktighet
Tänkbara standardpopulationer:	Antal äktenskap i hela riket

<i>Exempel 5</i>	
	Kassationsfrekvensen (dvs <i>relativa</i> frekvensen misslyckade enheter)
	Avdelning I och II
	Produktionsinriktningen
	Produktionen i avd I eller i avd II eller den sammanlagda produktionen i de båda avdelningarna.

Med tänkbara standardpopulationer menas alltså de populationer för vilka man har data tillgängliga om fördelningen efter den snedvridande faktorn.

I exempel 4 finns det således data för endast en tänkbar standardpopulation, antalet äktenskap i hela riket. Med denna standardpopulations fördelning efter den snedvridande faktorn (äktenskapets varaktighet) som vikter blir de standardvägda medeltalen för

GoH:

$$\frac{45\,603 \times 0,92 + 39\,902 \times 1,86 + 33\,956 \times 2,63 + \dots}{214\,662} = 2,73$$

J-arb:

$$\frac{45\,603 \times 1,07 + 39\,902 \times 2,11 + 33\,956 \times 2,86 + \dots}{214\,662} = 3,05$$

Slutsatsen blir att, om man eliminerar inverkan av olikheten i fördelningen över varaktighetsklasser, medelantalet barn per äktenskap är ca 0,3 högre bland jordbruksarbetarna än bland gods- och hemmansägarna.

Det är viktigt att eleverna förstår att *standardvägning endast kan utföras på medelvärden och*

*relativtal* (vanligen proportioner, procent eller promille). Om inte dessa medelvärden eller relativtal är givna måste de alltså beräknas före själva standardvägningen. I exempel 4 får vi följande relativa frekvenser av utsorterade enheter:

Typ	Avd I	Avd II
A	0,07	0,09 (= 180/2000)
B	0,02	0,04
C	0,08	0,06

Med den sammanlagda produktionen i de båda avdelningarna som standardpopulation får vi följande standardvägda (relativa) kassationsfrekvenser:

Avd I:

$$\frac{5\,000 \times 0,07 + 6\,000 \times 0,02 + 5\,000 \times 0,08}{16\,000} = 0,054$$

Avd II:

$$\frac{5\,000 \times 0,09 + 6\,000 \times 0,04 + 5\,000 \times 0,06}{16\,000} = 0,062$$

Efter att ha eliminerat effekten av olika produktionsinriktning (genom att för både avd I och II

göra en vägning med samma fördelning efter produktionsinriktning) finner vi att kassationsfrekvensen är högst i avdelning II.

I exempel 5 fanns tre möjliga standardpopulationer (var och en av de båda avdelningarna samt summan av dem). Det är kanske för mycket begärt att eleverna ska kunna värdera olika standardpopulationer. Annars kan sägas att, av dessa tre alternativ, summan brukar anses vara den mest neutrala standardpopulationen. I andra situationer kan det finnas data för hela riket, (som i exempel 4) eller för branschen som helhet. I allmänhet är även dessa alternativ lämpliga som standardpopulationer.

## 8. Sannolikhetsurval

Så gott som dagligen redovisas i massmedia resultat från stickprovsundersökningar av högst skiftande kvalitet. För att kunna bedöma sådana resultat är det viktigt att eleverna har viss kännedom om de vanligastes metoderna för *s k sannolikhetsurval*, dvs de metoder som medger beräkning av den statistiska felmarginalen för en skattning baserad på stickprovet. Glädjande nog har de läroböcker enligt den nya läroplanen för SE-matematiken, som jag granskat, haft rätt utförliga avsnitt om dessa metoder.

Sannolikhetsurval innebär att varje enhet i populationen har en känd sannolikhet att komma med i stickprovet. Vid *obundet slumpmässigt urval (OSU)* av  $t$  ex  $n = 200$  enheter från en population bestående av  $N = 3\ 000$  enheter har varje populationsenhet sannolikheten  $200/3\ 000$ , dvs  $1/15$ , att bli uttagen till stickprovet. Ett OSU ur ett numererat register kan tillgå så att en dator helt slumpmässigt producerar siffror som motsvarar nummer i registret (på samma sätt som sker vid penninglotteriets dragningar). *Urvalsram* är den tekniska termen för det register från vilket urvalet dras.

Sannolikhetsurval behöver dock inte innebära att alla populationsenheter har samma sannolikhet. Om populationen utgörs av företag kan det finnas skäl att ge storföretag större sannolikhet än småföretag. Principen med sannolikhetsurval är alltså inte den att alla populationsenheter ska ha *samma* sannolikhet, utan det räcker med att urvals sannolikheten är känd och att den beaktas när man gör skattningar från stickprovet till populationen.

En metod för sannolikhetsurval, som ofta används i riksomfattande intervjuundersökningar, är *s k clusterurval*. Ett sådant urval kan gå till så att man i ett första steg lottar ut 70 kommuner och i varje kommun intervjuar ett antal invånare (även detta urvalssteg skall baseras på slumpmässigt urval). IMUs *s k basurval* består av just 70 kommuner/församlingar från vilka man i ett nästa steg utlottar intervjupersonerna ur befolk-

ningsregistren. Sifo arbetar på ungefär samma sätt. Basurvalet omfattar 380 postnummerområden och från dessa har man slumpmässigt valt ut intervjupersonerna ur befolkningsregistren.

Clusterurval ger i allmänhet sämre precision än OSU vid samma antal intervjupersoner, säg 500. Denna nackdel kompenseras ofta vid besöksintervjuer av att koncentrationen till ett begränsat antal kommuner/församlingar/postnummerområden minskar resekostnaderna per intervju så att antalet intervjuer vid given totalkostnad för undersökningen kan öka. Intervjukostnaden för ett OSU om 500 personer utspridda över hela landet kan på så sätt komma att motsvara kostnaden för 1 000 personer vid clusterurval.

Sannolikhetsurval är en förutsättning för att man ska kunna beräkna den statistiska felmarginalen för en skattning baserad på stickprovet.

Undersökningen, som illustreras på den gode TOKS teckning nedan, är nog inte OSU, knappast någon annan metod för sannolikhetsurval heller, hur slumpmässigt urvalet än må verka. Tillvägagångssättet är mer likt *s k kvoturval*, en metod som bland statistiker åtnjuter ett stabilt dåligt rykte. Utgångspunkten för urvalet är populationens fördelning över kön, ålder, socialgrupp och kanske någon annan variabel. Målsättningen vid urvalsproceduren är att få samma fördelning över dessa variabler också i urvalet. Så långt är metoden statistisk oantastlig. Det betänkliga är att intervjuarna sedan har full frihet vid urvalet av individerna till dessa kvotgrupper. Intervjuarna frestas då att undvika personer som ser ilska ut, är sjaskigt klädda eller verkar ha bråttom. Inte heller går de gärna in i villor, bevakade av hundar. Denna urvalsmekanism medför att urvalet får en systematisk snedvridning av okänd storlek.



## 9. Stickprov ur svarsbortfallet

I den nya matematiken för SE2—3 ingår ett avsnitt om konsekvenserna av svarsbortfall. Enligt läroplanen ska man också belysa möjligheterna att genom ett stickprov ur bortfallsgruppen få en avsevärt säkrare skattning. Här är ett sådant exempel.

## Enkät i Eskilstunaområdet:

# Hälften inom Metall positiva till arbetet

### Exempel 6A

Rubriken baseras på de 2 550 som svarat av 7 500 tillfrågade. Svartsbortfallet var alltså 66%! Beräkna proportionen "positiva till arbetet" bland samtliga 7 500 tillfrågade *under antagande* att

a) alla b) ingen c) 30 % d) 70 %  
i bortfallsgruppen är positiv till arbetet.

### Exempel 6B

Antag att man gjorde en bortfallsuppföljning genom besöks- eller telefonintervjuer med ett obundet slumpmässigt urval om  $n=400$  ur bortfallsgruppen. Antag vidare att andelen positiva i detta stickprov blev 20%. Beräkna ett 95%-konfidsintervall för andelen positiva bland de  $N=4\,950$  i bortfallsgruppen.

A-uppgiften i detta exempel har jag utförligt behandlat i Nämnaren 1982/83:1 och jag återger här bara den numeriska lösningen.

Andelen positiva bland samtliga 7 500 tillfrågade fås enligt formeln

$$P = \frac{2\,550 \times s + 4\,950 \times b}{7\,500}$$

a)  $P_{\max} = \frac{2\,550 \times 50 + 4\,950 \times 100}{7\,500} = 83\%$

b)  $P_{\min} = \frac{2\,550 \times 50 + 4\,950 \times 0}{7\,500} = 17\%$

c)  $P = \frac{2\,550 \times 50 + 4\,950 \times 30}{7\,500} = 36,8\%$

d)  $P = \frac{2\,550 \times 50 + 4\,950 \times 70}{7\,500} = 63,2\%$

Observera att differensen mellan högsta och lägsta tänkbara andel positiva,  $P_{\max} - P_{\min}$ , är 66 procentenheter, dvs samma procenttal som bortfallet utgjorde av samtliga tillfrågade. Man kan ganska lätt visa att detta faktum gäller generellt.

Så till B-uppgiften. Ett 95%-konfidsintervall erhålles enligt

$$p \pm 1,96 \sqrt{\frac{p(100-p)}{n} \left(1 - \frac{n}{N}\right)}$$

där  $p$  = andelen positiva (i %) i stickprovet, dvs  $p=20$  i exemplet

$n$  = antalet tillfrågade i stickprovet, dvs  $n=400$

$N$  = antalet i den population från vilket stickprovet dragits, dvs  $N=4\,950$ .

Vi erhåller således konfidsintervall

$$20 \pm 1,96 \sqrt{\frac{20 \times 80}{400} \left(1 - \frac{400}{4950}\right)} = 20 \pm 3,8$$

Tolkningen är att intervallet  $(20 \pm 3,8)\%$  täcker den andel positiva som skulle erhållits vid totalundersökning av de 4 950 i bortfallsgruppen.

Genom att sätta in av detta resultat i den i A-uppgiften använda formeln för beräkning av andelen positiva bland samtliga 7 500 i populationen erhåller vi följande skattning av andelen positiva:

$$\frac{2\,550 \times 50 + 4\,950 \times (20 \pm 3,8)}{7\,500} = 30,2 \pm 2,5$$

Notera den väsentliga informationsökningen genom detta stickprov. I A-uppgiften (dvs före bortfallsuppföljningen) var vi mycket osäkra på nivån vad gäller andelen osäkra. Olika gissningar gav resultat som skilde flera tiotals procentenheter. Genom stickprovet om  $n=400$  har vi en ganska säker uppfattning om storleksordningen på det sökta procenttalet (ca 30%). Dessutom vet vi att felmarginalen för denna skattning är blott 2,5 %-enheter.

Om Du tar upp detta exempel i undervisningen, visa gärna också vad som händer med felmarginalen om man nöjer sig med ett stickprov om blott  $n=100$  ur bortfallsgruppen. Om man bortser från faktorn  $\sqrt{(1 - n/N)}$  i formeln för konfidsintervall — denna faktor ligger ju ganska nära 1 både för  $n=400$  och  $n=100$  i det här exemplet — varierar konfidsintervallens längd med  $\sqrt{1/n}$  (givet samma procenttal  $p$  i stickprovet). Felmarginalen i den sammanvägda skattningen blir alltså  $\pm 5$  procentenheter om  $p=20$  procent baseras på ett stickprov om  $n=100$ . Eller med andra ord: När stickprovet minskar med faktorn 4, ökar felmarginalen (endast) med faktorn 2.



Sådana här fakta är viktiga att känna till när man värderar resultaten av enkätundersökningar presenterade i t ex massmedia. Vid stort svarsbortfall kan man alltså genom en stickprovsundersökning av blott något hundratal i bortfallsgruppen erhålla en drastiskt ökad säkerhet i uttalanden om hela populationen. Det är således en oskickligt genomförd undersökning om undersökningsledningen accepterat ett bortfall på t ex 50 % eller mer utan att komplettera med en stickprovsvis gjord bortfallsuppföljning. När det gäller att dra slutsatser till hela populationen är resultaten av sådana undersökningar vanligen tämligen värdelösa. Och nämns ingenting alls om bortfallet i redovisningen av en enkätundersökning skall man vara ytterligt skeptisk. I praktiken blir det alltid ett svarsbortfall i enkätundersökningar och nämns inget därom i rapporten tyder

det på att rapportören antingen är oseriös eller inte ens förstått problemet.

Vill Du ha ytterligare ett praktikfall som underlag för konstruktion av övningsuppgifter av samma typ som exempel 6A och 6B följer här ett klipp från 14 okt 1986. "Varannan ägare av ny bil missnöjd" påstår rubriken. Förmodligen åsyftas då de 55 % som enligt artikeln inte klarat sig undan "större fel under det första året med den nya bilen". Resultaten baseras på "en utfrågning av knappt 8 000 nybilsägare". Vad som inte alls framgår av denna tidnings artikel var att antalet tillfrågade nybilsägare uppgick till ca 50 000. Av dessa svarade 7 859 — svarsbortfallet var alltså ca 84 %!

Till saken hör att den här undersökningen kommenterades (vanligen okritiskt) i stort uppslagna referat i samtliga större dagstidningar.

## Varannan ägare av ny bil missnöjd



Minst fel hade Toyota Corolla enligt utfrågningen av nya bilägare.



Nissan Stanza hade också nöjda ägare.

Att köpa en ny bil är ingen garanti för att slippa problem. Det framgår med all önskvärd tydlighet av en rapport från Votormännens Riksförbund (M).

I en utfrågning av knappt 8 000 ägare till nya bilar av 1985 års modell är det bara knappt hälften av ägarna (45 procent) som klarat sig undan större fel under det första året med den nya bilen.

Med större fel avses sådana som tvingat ägaren till verkstad eller som orsakat bilbrändning.

Mindre fel, sådana som ägarna kunnat laga själva eller kunnat vänta med till nästa service tillfälle, har bara en tredjedel (35 procent) undgått!

Möjligen har bilarna, genom snittligt sett, ändå blivit bättre. I M:s förra undersökning, en enkät med ägare till m/84 som då gått upp till ett år, var det bara 35 resp 31 procent som klarat sig undan större resp mindre fel på sina nya bilar.

Fiat, Ford och Volvo gör bilarna som ägarna har problem att starta

Var tredje ägare (33 procent) anser lackproblemen som ett "större fel" på sin bil (ytterligare 25 procent tar upp det i kolumnen "mindre fel").

Släp	Modellbilen (3 procent) är också
Pugeot 305	7 %
Renault 5	6 %
Fiat Ritmo	6 %
Ford Escort	5 %
Toyota Tercel	4 %
Honda Civic	4 %

Fiat Ritmo ledde ligan av missnöjda ägare.



### Nöjdaste ägare

Japanska och (dyra) tyska bilar ger nybilsägarna minst anledning till klagomål, enligt redovisningen i M:s rapport.

"Laga svåra problem" kryssar i medeltal knappt varannan (45 procent) av nybilsägarna för i redovisningen. Bästa utfallen har de här modellerna:

Toyota Corolla	71 %
Nissan Stanza	68 %
Honda Accord	67 %
Toyota Camry	66 %
BMW 316-323	66 %
BMW 516-528	63 %
Nissan Cherry	63 %
Nissan Micra	63 %
Saabur 1900	62 %
Toyota Starlet	61 %
Volvo 740 GLE	60 %

Lancia Prisma hamnade som tvåa i problemklassen.



## 10. Exempel för introduktion av kritisk granskning

### Exempel 7

Kommentera kritiskt — med ledning av informationen i artikeln — allmängiltigheten i rubriken "En av tre har sömnproblem".

Undersökningen gäller endast

- 45—65-åringar ("de åldrar då bruket av sömntabletter är vanligast")
- som bor i Tierp!

Huruvida *en av tre* ens gäller för denna begränsade population beror på hur att "ha sömnproblem" definieras. Om "problem med insomnitet" inräknas blir andelen betydligt högre (drygt två tredjedelar av kvinnorna och drygt hälften av männen hade detta problem enligt sista stycket).

När eleverna övat på detta exempel är det nog inte så svårt för dem att komma med liknande invändningar mot Dagens Nyheters "avslöjande rapport om svenskarna."

## En av tre har sömnproblem

Ungefär var tredje person har någon form av sömnproblem. Bland kvinnor är problemen avsevärt vanligare än bland män.

Det visar den första undersökningen om sömnproblem som gjorts i skandinavien.

Bakom undersökningen står forskare vid sömnlaboratoriet vid psykiatriska kliniken på Akademiska sjukhuset i Uppsala.

Undersökningen gjordes i form av en enkät med 65 frågor om

sömnproblem som skickades till 400 personer i Tierpsområdet. Samtliga utvalda var i åldrarna 45—65 år, de åldrar då bruket av sömntabletter är vanligast.

Svaren visar att var fjärde kvinna har stora problem med för litet sömn. Var sjunde man har samma problem.

Drygt två tredjedelar av kvinnorna uppgav vidare att de hade problem med insomnitet. Var sjunde hade stora problem. Motsvarande siffror för männen var drygt hälften som hade problem, medan var tjugonde hade stora problem med att somna.

(TT Uppsala)

# Åtta av tio äter medicin

Av GUN LEANDER

**I åtta av tio svenska hushåll pågår någon form av mediciner. Det visar en ny studie som avslöjar hur det ser ut i de svenska hemmens läkemedels-gömmor.**

Att vi är ett medicinerande folk har stått klart länge. Men hittills har bara två svenska studier gjorts av hur mycket genomsnittsbefolkningen får i sig och de har båda gällt enbart äldre personer.

Den här nya undersökningen är gjord av distriktssköterskor som besökt människor i alla åldrar i deras hem inom Uppsala kommun.

Många människor ruvar över sina läkemedelsförråd som de dyrbaraste skatter, visar studien, som beskrivs i senaste numret av Läkartidningen.

Tre av fyra svenska kvinnor har just nu en läkemedelskur på gång. Åtminstone om man skall överföra resultatet från den här studien till hela svenska folket. Bland männen är pillerätandet

inte lika utbrett, men tre av fem medicinerar.

## Mest äldre

Sex distriktssköterskor i Uppsala bad att få titta på de här förråden i samband med hembesök hos 116 personer i totalt sett 63 hushåll. Åldersfördelningen kom att domineras av äldre, men här finns även många barn och ungdomar.

hushåll hade mer än fem olika

Det var enbart receptbelagda läkemedel som undersöktes.

Sköterskorna tog även reda på hur många oanvända mediciner som stod och samlade damm i hyllorna. Då visade det sig att många protesterade särskilt de äldre. De ville inte visa sina oanvända mediciner ens för välbekanta distriktssköterskor. Tio oanvända läkemedel. I ett fall hittade man 17 olika öppnade förpackningar av olika mediciner. Den vanligaste hylvärmarren var magmedicin, som många kan tänka sig att använda senare "vid behov".

DN 24/7 -85

När det gäller urvalsundersökningar är det naturligt att fråga sig *hur* urvalet gått till. Här är det sex distriktssköterskor som gjort hembesök, något som väl vanligen bara sker hos personer som är eller nyligen varit sjuka. Undersökningen omfattar endast 116 personer med en dominans av äldre. Att från detta lilla och snedvridna urval dra slutsatser om "svenskarna" är förstås befängt.

Någon kanske tycker att massmedias behandling av statistiskt material inte är mycket att bråka om ("alla vet ju att det som står i tidningarna ofta är fel . . ."). Jag tror dock att man i så fall underskattar massmedias genomslagskraft. De flesta tidningar görs av proffs med påföljd att slagkraftiga rubriker och artikelfragment ofta har en förmåga att etsa sig fast.

För nog är det väl den här DN-artikeln som sex veckor senare dykt upp i finansministerns huvud när Expressen bett om "Feldts besked" om bl a svenska folkets badrumsskåp och "inköp av allehanda mediciner".

## **Köper vi för mycket medicin?**

Däremot tycker Kjell-Olof Feldt att det är dags att se över kostnaderna för svenska folkets inköp av allehanda mediciner.

— I varje svenskt hem lär det vara så att badrumsskåpet är proppat med gammal medicin.

Feldt tror att läkarna skriver ut på tok för stora mängder medicin och att det beror på att läkarna inte har något kostnadsansvar.

Så landets läkare får nog bereda sig på att smaka en besk medicin om finansministern heter Kjell-Olof Feldt också efter valet om en vecka.

## 11. Checklista för kritisk granskning av statistiska undersökningar

I den här artikeln har blandats exempel på god och dålig användning av statistik. Så görs också i massmedia, men där finns sällan några etiketter som anger kvaliteten. Därför tycker jag det är viktigt att skolan bl a genom matematikundervisningen tränar eleverna att sovra bland den statistiska informationen. Ett av målen för denna träning kan vara att få eleverna att uppleva det som naturligt att ställa följande frågor innan de accepterar ett med statistiska data underbyggt påstående — och innan de själva formulerar en egen slutsats:

Vem påstår? Varför påstår denne? Vilken definition har använts? Finns det andra definitioner som ger annat resultat?

Vilken är urvalsramen? Är det ett sannolikhetsurval? Hur stort är urvalet? Den statistiska felmarginalen? Hur stort var bortfallet?

Finns det snedvridande faktorer?

Exakt vilken fråga ställdes? Var besvarades frågorna? Anonymt? I hemmet, skolan eller på jobbet? Föreligger risk för grupptryck eller intervjuarpåverkan?

Finns det andra (ev opublicerade) undersökningar på området?

Har man förväxlat orsak och verkan?

När genomfördes undersökningen? Inträffade vid den tidpunkten någon händelse som tillfälligt kan ha påverkat opinionen?

**PÅSTÅENDEN SOM INTE BASERAS PÅ STATISTIK BÖR MAN VARA ÄNNU MER FRÅGANDE INFÖR . . .**